

TREBALL DE FINAL DE GRAU

Anàlisi per a la Gestió Eficient de l'Equipatge de Cabina en l'Aviació Comercial

Grau en Gestió Aeronàutica

Curs 2019 - 2020

Mar Purull Polo

Berta Quirante Saderra

Dirigit per: **Dr. Angel A. Juan**

Sabadell, 02 de juliol del 2020

El **Dr. Angel Alejandro Juan**, professor de l'Escola d'Enginyeria de la UAB,

CERTIFICA:

Que el treball al que correspon la present memòria ha estat realitzat sota la seva direcció
per:

Mar Purull Polo i Berta Quirante Saderra

I per a que consti firma la present.

Sabadell, juliol de **2020**



Signat: **Dr. Angel Alejandro Juan**

FULL DE RESUM – TREBALL FI DE GRAU DE L'ESCOLA D'ENGINYERIA

Títol del Treball Fi de Grau: Anàlisi per a la Gestió Eficient de l'Equipatge de Cabina en l'Aviació Comercial	
Autor[a]: Mar Purull Polo i Berta Quirante Saderra	Data: Juliol 2020
Tutor[a]/s[es]: Dr. Angel Alejandro Juan	
Titulació: Grau en Gestió Aeronàutica	
Paraules clau	
<ul style="list-style-type: none"> • Català: optimització, equipatge de mà, regressió logística, clúster, aerolínia, metodologia. • Castellà: optimización, equipaje de mano, regresión logística, clúster, aerolínea, metodología. • Anglès: optimization, hand baggage, logistic regression, cluster, airline, methodology. 	
Resum del Treball Fi de Grau	
<ul style="list-style-type: none"> • Català: En aquest projecte hem desenvolupat una metodologia aplicable a totes les companyies aèries per solucionar el problema d'excés d'equipatge de mà en l'aviació comercial. Per fer-ho possible hem dissenyat un model de regressió logística per mesurar quina és la probabilitat que un passatger facturi l'equipatge de mà, i un <i>cluster analysis</i> per tal de poder donar les directrius sobre com elaborar les noves polítiques relacionades amb l'equipatge de mà en funció del tipus de passatger, i arribar a aconseguir que aquest/a, per la seva pròpia voluntat, decideixi facturar l'equipatge de mà o, directament, no portar-ne. • Castellà: En este proyecto hemos desarrollado una metodología aplicable a todas las compañías aéreas para solucionar el problema de exceso de equipaje de mano en la aviación comercial. Para hacerlo posible hemos diseñado un modelo de regresión logística para medir cuál es la probabilidad de que un pasajero facture el equipaje de mano, y un <i>cluster analysis</i>, para poder dar las directrices de cómo elaborar las nuevas políticas relacionadas con el equipaje de mano en función del tipo de pasajero, y llegar a lograr que éste/a, por decisión propia, decida facturar el equipaje de mano o, directamente, no llevarlo. • Anglès: In this project we have developed a methodology applicable to all airlines to solve the problem of excess of hand luggage in commercial aviation. To make this possible, we have designed a logistic regression model to measure the probability of a passenger to check-in the hand luggage, and a cluster analysis, in order to give the guidelines on how to create new policies related with hand luggage based on the type of passenger, to encourage the passenger to check-in the hand luggage or not to carry it. 	

ÍNDEX

SECCIÓ 1 INTRODUCCIÓ.....	7
1.1. RESUM.....	8
1.2. OBJECTIUS GLOBALS I PARCIALS DEL PROJECTE	12
1.3. ESTAT DE L'ART	13
1.4. METODOLOGIA	19
1.5. PLANIFICACIÓ TEMPORAL DEL PROJECTE	20
SECCIÓ 2 BASE TEÒRICA I EVOLUCIÓ DEL PROBLEMA.....	21
2.1. EFECTES DE L'EQUIPATGE DE MÀ	22
2.1.1. Econòmics.....	22
2.1.2. Ambientals.....	24
2.1.3. Operacionals.....	26
2.2. INICIATIVES QUE S'HAN DUT A TERME FINS EL MOMENT.....	30
2.3. POLÍTIQUES D'EMBARCAMENT	35
2.4. ESTUDI DEL PASSATGER	41
2.4.1. Elecció del tipus d'equipatge	41
2.4.2. <i>Customer journey</i>	44
2.4.3. <i>Social dilemma theory</i>	44
SECCIÓ 3 METODOLOGIA DE RESOLUCIÓ I EXECUCIÓ	51
3.1. DESCRIPCIÓ DETALLADA DEL PROBLEMA	52
3.2. <i>INPUTS</i>	53
3.3. IMPLEMENTACIÓ I EXPLICACIÓ DEL CODI.....	58
3.3.1. Model de regressió logística	58

3.3.2. <i>Cluster analysis</i>	65
3.3.2.1. <i>K-means</i>	66
3.3.2.2. <i>Agglomerative clustering</i>	69
3.3.2.2.1. <i>Ward</i>	69
3.3.2.2.2. <i>Complete</i>	70
SECCIÓ 4 ANÀLISI DE RESULTATS	72
4.1. INTRODUCCIÓ	73
4.2. ANÀLISI DELS RESULTATS DEL MODEL DE REGRESSIÓ LOGÍSTICA	73
4.3. ANÀLISI DELS RESULTATS DEL CLUSTER ANALYSIS	76
SECCIÓ 5 CONCLUSIONS I TREBALL FUTUR	77
5.1. CONCLUSIONS	78
5.2. TREBALL FUTUR	82
SECCIÓ 6 REFERÈNCIES BIBLIOGRÀFIQUES	84
SECCIÓ 7 ANNEX	90
7.1. DADES <i>RANDOM</i> GENERADES	91
7.2. CODI DEL PROGRAMARI	92

Agraïments

Aquest projecte no hagués estat possible sense el suport que hem obtingut des del primer moment per part de, principalment, dues persones:

Primer de tot, del nostre tutor de treball, el Dr. Angel Alejandro Juan, que sense la confiança transmesa des del primer moment no hagués estat possible realitzar aquest projecte.

Al mateix temps, agrair a la Mariem Gandouz la dedicació i el suport tècnic en l'anàlisi de les dades en les seccions 3 i 4 durant el desenvolupament del treball.

Gràcies a tots dos per ajudar-nos en cada una de les fases d'aquest treball aportant opinions, idees i recolzament. I sobretot per fer-nos créixer tant personalment com acadèmicament, en un àmbit pel qual teníem molta curiositat.

SECCIÓ 1

INTRODUCCIÓ

1.1. RESUM

Durant l'última dècada, les polítiques d'equipatge en les diferents companyies aèries han patit grans canvis. Anys enrere, quan s'adquiria un bitllet d'avió, ja fos de curt, mitjà o llarg radi, en la majoria dels casos l'equipatge de mà i una maleta facturada, estaven inclosos. Viatjar era considerat un luxe, on només gent amb un cert poder adquisitiu podia accedir-hi, mentre que actualment gairebé tothom que vulgui ho pot fer. Amb l'aparició de les companyies *low-cost*, que han afectat majoritàriament al curt i mitjà radi, i l'alta competència que van suposar a les anomenades aerolínies de bandera (en el que a preus es refereix), la solució de moltes va ser reduir serveis per aconseguir abaratir el preu del bitllet i poder fer front a aquesta nova tendència de viatjar. És en aquest punt on es generen noves tarifes aèries, ja que fins al moment les dues que predominaven eren la tarifa de classe turista i primera classe, on les diferències entre una i l'altra, es trobaven principalment als serveis. Amb l'aparició de les noves tarifes, en les quals pots arribar a trobar varis subtipus dins de la mateixa, l'equipatge tant de bodega i sobretot de cabina, passen a tenir un paper important per tal de donar la possibilitat al passatger d'escollir a què està disposat a renunciar per adquirir el bitllet el més econòmic possible.

En els seus inicis, totes les companyies *low cost* tenien incloses en totes les seves tarifes l'equipatge de mà, (que pot variar entre les mides 55x35x25cm a 56x45x25cm), tal i com es pot observar a la taula 1. En la majoria dels casos també s'inclouïa una peça de menors dimensions, com un maletí, motxilla o bossa.

POLÍTICA EQUIPATGE DE MÀ			
AEROLÍNEA	MIDA PERMESA	NÚMERO DE PECES	TIPUS
Air France (Air France, s.f.)	55 x 35 x 25	1 peça + 1 accessori	BANDERA ¹
British Airways (British Airways, s.f.)	56 x 45 x 25	1 peça + 1 accessori	BANDERA
Iberia (Iberia, s.f.)	56 x 45 x 25	1 peça + 1 accessori	BANDERA
Vueling (Vueling, s.f.)	55 x 40 x 20	1 peça + 1 accessori	HÍBRIDA ²
Lufthansa (Lufthansa, s.f.)	55 x 40 x 23	1 peça	BANDERA
Tap Portugal (TAP Portugal, s.f.)	55 x 40 x 20	1 peça + 1 accessori	HÍBRIDA
Air Europa (Air Europa, s.f.)	55 x 35 x 20	1 peça + 1 accessori	HÍBRIDA
Norwegian (Norwegian, s.f.)	30 x 20 x 38	1 peça petita	LOW COST ³
Easyjet (Easyjet, s.f.)	56 x 45 x 25	1 peça	LOW COST
Volotea (Volotea, s.f.)	55 x 40 x 20	1 peça + 1 accessori	LOW COST
Ryanair (Ryanair, s.f.)	40 x 20 x 25	1 peça petita	LOW COST

Taula 1 – Política de l'equipatge de mà

Actualment, com podem observar a la taula 1, totes les aerolínies inclouen almenys un equipatge de mà sense cost addicional en totes les seves tarifes, tot i que algunes d'elles en tarifes superiors permeten una peça extra. Observem que les companyies *low cost*, en general (tret de Volotea), només ofereixen en totes les seves tarifes el transport d'una única peça, mentre que en les conegudes de bandera i híbrides el passatger pot transportar una peça acompanyada d'un accessori. Destacar però, que dins de les *low*

¹ Antigament es denominava així a les aerolínies gestionades per l'estat. Actualment també s'utilitza per aquelles companyies que estaven gestionades per l'estat, però que ara estan privatitzades.

² Aquelles companyies que es troben a mig camí entre les de bandera i les *low cost*.

³ Aerolínea que ofereix tarifes més econòmiques respecte les tradicionals, a canvi d'eliminar molts dels serveis que ofereixen les de bandera.

cost, trobem dues companyies (Norwegian i Ryanair), que només ofereixen una peça amb unes dimensions menors a una maleta de cabina estàndard.

En canvi, l'equipatge de majors dimensions per facturar a bodega, en les tarifes més baixes de totes les companyies, té un cost addicional en els vols de curt i mitjà radi, a diferència de fa uns anys enrere, abans d'aparèixer les *low cost*, on no hi havia cap càrrec suplementari.

És per això que molts usuaris avui dia opten per l'equipatge de cabina, causant un problema d'espai i considerables retards durant l'embarcament dels avions, ja que aquests no estan configurats de manera que càpiguen tantes maletes de mà com capacitat total de passatgers.

El model d'aeronau més utilitzat per les aerolínies en els vols de curt i mitjà radi és l'Airbus 320 i l'Airbus 321, amb una capacitat de 104 i 113 maletes de cabina respectivament.

- En el cas del A320 més utilitzat, la seva capacitat de passatgers, és de 180, per tant, en un vol en el que s'arribés a la capacitat màxima, si tots els passatgers portessin una peça d'equipatge de mà, s'haurien de retirar 76 maletes de la cabina per ser reubicades a bodega.
- D'altra banda, l'A321 més comú, té una capacitat de 220 passatgers, causant per tant la retirada de 106 maletes en el cas que es donés un vol amb capacitat màxima en el que tots els passatgers portessin maleta de mà.

A conseqüència d'això, en els últims anys, les diferents aerolínies han adoptat estratègies per agilitzar el procés d'embarcament, retirada, etiquetat i reubicació a la bodega, de les maletes que no tenen espai a la cabina. Algunes d'aquestes estratègies han sigut:

- Facturar l'equipatge de mà de forma gratuïta en el mostrador de facturació, sempre i quan compleixi amb les mesures i pes indicats.
- Cobrar un extra per l'equipatge de mà.
- Reduir el número de peces que pot portar cada passatger.
- Un cop contades les maletes que tindrien espai a la cabina, començar a etiquetar les que hauran de ser transportades a bodega abans del procés d'embarcament.

Tot i així, basant-nos en la nostre experiència, tant com a passatgeres i com a treballadores dins l'aeroport, diríem que cap d'aquestes estratègies ha acabat de funcionar, ja que a dia d'avui la problemàtica amb l'equipatge de mà segueix present.

1.2. OBJECTIUS GLOBALS I PARCIAIS DEL PROJECTE

Objectius globals

L'objectiu principal del projecte és crear una metodologia per tal d'analitzar els diferents grups de passatgers de qualsevol companyia aèria obtenir les directrius per elaborar polítiques personalitzades entorn l'equipatge de mà, segons les característiques de cada grup de passatgers.

Com ja s'ha esmentat, assolint aquest objectiu aconseguirem tenir un model d'anàlisi estàndard per a qualsevol companyia aèria, independentment de les seves característiques.

Objectius parcials

Per tal d'arribar a l'objectiu global, seran necessaris una sèrie d'objectius parcials, per definir una metodologia el més acurada possible. Aquests objectius són:

1. Definir quines són les àrees del problema: Estudiar el problema plantejat per obtenir els seus orígens i així facilitar la cerca de la millor solució possible.
2. Realitzar un estudi de les diferents polítiques actuals de les diferents aerolínies i quins són els resultats d'aquestes: Ens servirà per analitzar de quina manera han enfocada aquesta problemàtica al llarg dels anys i per què.
3. Recollir les dades necessàries i classificar-les, per tal d'identificar quins són els paràmetres que afecten al problema, i poder dissenyar una metodologia entorn a aquests paràmetres. Utilitzarem eines d'anàlisi de dades per poder extreure aquesta informació i classificar-la.
4. Una vegada obtinguts els resultats d'aquesta metodologia, donar directrius per solucionar el problema.
5. Com a objectiu personal, aprendre més sobre la branca d'anàlisi de dades.

1.3. ESTAT DE L'ART

En aquesta primera cerca no ens hem centrat únicament en la recerca d'articles, tesis i estudis relacionats de forma directa amb la problemàtica plantejada respecte l'excés d'equipatge de mà en un vol, ja que fins al moment no hi ha un gran ventall de publicacions al respecte i, bona part d'elles són confidencials. És per això que ho hem enfocat de manera més oberta, per veure quin és el paper que té en general l'equipatge de mà en un vol i, a partir d'aquí, poder analitzar els diferents elements que interfereixen en el problema i que anteriorment podíem no estar contemplat.

Avui en dia moltes companyies low cost posen en pràctica el desglossament del preu del bitllet, per tal de reduir costos i augmentar ingressos. Ofereixen un preu estàndard pel bitllet bàsic sense cap servei extra i implementen taxes per l'augment d'aquests.

(Wong, Zhang, Hui, & Leung, 2009) estudia la capacitat de la bodega en relació a l'equipatge del passatger i la càrrega i com això afecta econòmicament. En el procés d'anàlisi, observa que quan la companyia implementa taxes per l'equipatge de mà s'assegura que el passatger no porti equipatge de més i així redueix en costos.

(Choo & Dresner, 2018) després de realitzar l'estudi, conclou que les companyies que implementen taxes generen un ingrés extra, són sostenibles en el mercat i, un factor molt important, milloren el rendiment de les operacions (puntualitat, manipulació d'equipatges, consum de combustible, etc). També esmenta la importància de no globalitzar el problema, sinó que és necessari definir-lo i estructurar-lo, ja que remarca que no tots els vols tenen les mateixes característiques, ni totes les companyies la mateixa visió.

El model desenvolupat per (Scotti & Dresner, The impact of baggage fees on passenger demand on US air routes, 2015) demostra que un augment en les taxes d'equipatge té una repercussió negativa en la demanda dels clients (aquest cas es centra en EEUU), però té menys repercussió que si s'augmenta el preu del bitllet genèric amb tots els serveis inclosos. A més a més, també argumenta que la implementació de taxes per la

maleta és un resultat socialment òptim, ja que el passatger que decideix no portar maleta, s'estalvia uns diners obté un preu de bitllet més econòmic, i l'aerolínia disminueix el cost d'operacions al no processar la maleta. L'article de (Scotti, Dresner, & Martini, 2016) que analitza la relació que hi ha entre les taxes, el rendiment operatiu i la satisfacció del client, també es veu reflectit en aquesta última conclusió.

En un estudi realitzat per (Bijker J. , 2016) on l'objectiu principal és reduir l'equipatge de mà, argumenta que el temps, la seguretat i el cost són els elements principals que fan que un passatger es decideixi per l'equipatge de mà i no l'equipatge facturat. És en aquest punt on comenta que les possibles solucions a la problemàtica de l'equipatge de mà que s'han posat en pràctica fins al moment en altres companyies aèries, han tingut en la major part un impacte negatiu en l'experiència del client i en el resultats esperats per la companyia. Per aquesta raó, arriba a la conclusió que la millor manera per reduir l'equipatge de mà és proposar noves polítiques, enfocades al client, sense costos addicionals per aquest, on tot i tenint l'opció de poder portar l'equipatge a cabina, aquest decideixi facturar-lo o inclús no portar-ne.

Modificar l'actitud del passatger és un factor clau a l'hora d'implementar una nova política. Al llarg de la història de l'aviació, hi ha hagut grans canvis en el funcionament de les companyies aèries i grans modificacions que afectaven directament al passatger. (Dennis, 2007), per exemple, ens mostra la reducció de la qualitat de servei que van haver d'implementar les companyies de bandera per tal de reduir costos per ser competitives amb les *low cost*. El passatger es va haver d'adaptar al nou servei, i va haver d'entendre que si es volia, per exemple, menú a bord en segons quins vols (un exemple seria en vols de curt radi), havia de contractar el servei a part o pagar una categoria superior en el preu del bitllet, ja que antigament el menú i els aperitius entraven dins del preu del bitllet, sense cap cost addicional.

(Székely, Weinmann, & Brocke, 2016) a partir de la *Nudge Theory* , realitzen un estudi que busca fomentar el pagament extra (optatiu) per part dels passatgers, en el moment de comprar el bitllet d'avió, per així mitigar la petjada ecològica del seu vol en termes de diòxid de carboni alliberat en aquest.

En una recerca inicial queda demostrat que, a través dels sistemes d'informació, es pot conduir a la gent a actuar d'una forma més responsable amb el medi ambient (Baeriswyl, Staake, & Loock, 2011), (Graml, Loock, Baeriswyl, & Staake, 2011), (Loock, Staake, & Landwehr, 2011), (Loock, Landwehr, Staake, Fleisch, & Pentland, 2012) i (Loock, Staake, & Thiesse, 2013)). Cosa que els porta a pensar que els S.I. tenen un alt potencial per influenciar de forma positiva en la decisió sobre el pagament compensatori, per l'alliberament de CO2 en el vol seleccionat en el moment d'efectuar la compra del bitllet.

Hi ha 3 mecanismes bàsics per canviar el comportament d'una activitat econòmica: alterar els preus (per incrementar o decrementar l'opció d'escollir una altra alternativa), proporcionar informació, o aplicar restriccions. Tot i així es considera que aquests mecanismes són massa bàsics i poc efectius en moltes situacions (Johnson, y otros, 2012). Estudis en psicologia han demostrat que les persones actuen amb racionalitat limitada (Simon, 1955), i sovint es prenen decisions menys òptimes degut a la falta d'informació, atenció, autocontrol (Bailes & Hoy, 2014), (Goldsmith & Cyboran, 2013), i adopció heurístiques (Mullainathan & Thaler, 2000).

Les decisions de les persones estan influenciades també en funció de com es presenten les opcions d'aquestes, ja que no existeix una manera neutral per fer-ho. (Johnson, y otros, 2012).

Al final de l'estudi es conclou que canviant el valor econòmic del pagament extra que surt per defecte en forma de compensació per la petjada ecològica del bitllet, els valors més alts per defecte incrementen la donació i viceversa. Per tant, a diferència del que s'hagués pogut pensar en un principi, un bon disseny per encaminar la presa de decisions mitjançant els sistemes d'informació té un impacte molt positiu en els resultats.

(Armando & Katalina, 2019) fan un estudi que parla de l'aplicació del *digital nudging* en el procés de compra pel qual passa una persona quan vol adquirir un bé o servei, ja que una aplicació important d'aquesta metodologia és la de dissenyar el procés de compra enfocat a obtenir els resultats esperats.

(Thaler & Sunstein, 2003), descriuen el *nudging* com un mètode que, tot i preservant la lliure elecció, autoritza a les institucions públiques i privades a persuadir a les persones cap a una direcció concreta per tal d'aconseguir el benefici propi. Es pot definir per tant el *digital nudging* com a *nudging* que compta amb el suport de sistemes d'informació en un entorn digital.

A través de diverses cerques entre diferents autors, (Lemon & Verhoef, 2016), proposen un model genèric del procés de compra del consumidor, que es divideix en tres escenaris: pre-compra, compra, i post-compra. Un bon disseny d'aquests 3 escenaris pot encaminar al client a prendre la decisió més favorable per l'empresa.

Per desenvolupar *digital nudges*, (Schneider, Weinmann, & Vom Brocke, 2018), va desenvolupar un marc de referència que inclou: definició de l'objectiu, entendre i comprendre a l'usuari, disseny del *nudge*, i test del *nudge*.

(Budd, 2013) porta a terme un estudi sobre quins són els mètodes de transport que utilitza la gent per desplaçar-se fins l'aeroport, i quines són les mesures que es podrien adoptar per tal de disminuir el transport privat i fomentar el transport públic.

Per realitzar l'estudi parteixen de dues teories: la Teoria dels Valors de (Schwartz, 1977) i la Teoria de la Conducta Planificada de (Ajzen, 1991).

La teoria dels valors argumenta que la conducta ve determinada per les intencions i percebuda pel control envers una acció. Són accions que requereixen sacrificis individuals per aconseguir el bé d'altres. (Abrahamse, Steg, & Vlek, 2009).

Aquesta teoria té una gran importància en el procés de compra, ja que aquest pot ser vist com un conjunt de decisions morals. Aquest procés, basant-se en el concepte original de (Schwartz, 1977), ve determinat per: la preocupació de les conseqüències (percepció sobre la possibilitat d'aparició de conseqüències negatives si no s'actua d'una determinada manera), i mesura de la responsabilitat (definit com una tendència general envers la negació de responsabilitat).

Durant l'estudi (Budd, 2013), troba que la preocupació sobre els impactes generats per l'ús del transport privat va connectat als sentiments de culpa que, alhora, van lligats com

a conseqüència a l'augment de sentiments positius cap a la obligació moral d'utilitzar el transport públic.

Per una altra banda tenim la teoria de la conducta planificada, que diu que el comportament ve predominant per conviccions morals individuals ja existents, i són aquestes les que guien a les persones a actuar d'una forma o altra. Aquesta està marcada per consideracions personals i va en busca dels majors beneficis per un mateix. (Bamberg, Fujii, Friman, & Gärling, 2011) assumeix que, en el cas d'existir diferents alternatives, l'elecció d'una es basa en la fortalesa d'intencions a l'hora de portar a terme cada una d'elles. La probabilitat que una persona adopti un cert comportament es considera que ve determinada per 3 elements: actitud (sentiments d'aprovació o desaprovació envers una acció), normes subjectives (sentiments d'aprovació o desaprovació per part d'un grup envers un comportament en concret), i percepció de control conductual (es refereix a la percepció que té l'individu sobre la possibilitat de poder adoptar una conducta o no).

En el context del procés de compra, aquests termes van lligats a aspectes pràctics del viatge com el cost, el temps, la flexibilitat, etc.

En un experiment realitzat per (Anable & Gatersleben, 2005) es determina que, per desplaçaments de caràcter laboral, es tenen més en compte factors instrumentals com els comentats anteriorment. Mentre que per desplaçaments d'oci, els usuaris tenen en compte en parts iguals factors instrumentals i afectius tals com la comoditat, control, estrès, etc.

S'ha detectat també que els patrons diaris de comportament, respecte el transport, es repeteixen al llarg del temps, i és per això que en alguns estudis també s'incorpora la mesura dels hàbits. (Gärling & Axhausen, 2003).

Una de les hipòtesis que es planteja (Budd, 2013) en el seu estudi és quan utilitzar cada una d'aquestes teories. Per tal de poder decidir de la manera més acurada possible, divideix el mercat en diferents segments, basant-se en les característiques del viatge i les actituds dels passatgers, mitjançant un *cluster analysis*.

D'aquesta manera serà possible identificar quins perfils de passatgers tenen més potencial per canviar la seva conducta envers el transport públic i aplicar la teoria que millor s'ajusti per crear les millors polítiques per cada grup.

Aquesta primera introducció ens ajuda a començar analitzar els factors rellevants que haurem de tenir en compte a l'hora de crear el nostre model analític per tal d'aconseguir una gestió eficient de l'equipatge de mà.

1.4. METODOLOGIA

Després de desenvolupar la nostra idea i marcar uns objectius parcials, és necessari establir una metodologia per tal d'assolir l'objectiu final d'aquest projecte.

Inicialment s'elaborarà un esquema estructural del treball per tenir la idea ben definida. A més a més, es farà una recerca d'informació per ampliar la nostra base teòrica i poder fer un estudi de tota la informació obtinguda. Tota aquesta recerca ens ajudarà a poder definir millor la nostra solució del problema. Per poder aconseguir tota la informació, accedirem a fonts electròniques públiques, articles i revistes digitals, la biblioteca de la Universitat Autònoma de Barcelona i documents proporcionats per la universitat.

Un cop desenvolupada tota la nostra recerca, passarem a desenvolupar la metodologia del nostre projecte per arribar a l'objectiu final. Desenvolupant un estudi del passatger per conèixer millor el seu comportament. Per poder generar l'estudi del passatger, realitzarem un *cluster analysis* desenvolupat en l'entorn web de *Jupyter* mitjançant *Python*. A més a més, realitzarem un model de regressió per analitzar la probabilitat que un passatger facturi l'equipatge de mà.

A partir dels diferents grups de passatgers es determinaran quines característiques han de seguir les polítiques d'equipatge de mà per aconseguir que el major nombre de passatgers facturi aquest equipatge.

Per portar a terme el procés esmentat, tindrem l'ajuda del nostre tutor de projecte, que ens servirà per adquirir nous coneixements sobre el tema, a més a més dels coneixements ja adquirits per part de les autores d'aquest estudi al llarg de la carrera i en concret de l'assignatura impartida pel tutor.

Per acabar, esmentar que durant el desenvolupament del projecte es portarà a terme una sèrie de tutories, entregues i reunions de seguiment amb el tutor, per supervisar cada pas, adquirir mesures correctives per poder millorar el nostre projecte i arribar a l'objectiu final.

1.5. PLANIFICACIÓ TEMPORAL DEL PROJECTE

A continuació es presenta una planificació de les diferents tasques del projecte (Taula 2), amb la seva corresponent data d'inici, data final i duració. Aquesta planificació és de gran importància, ja que volem realitzar una bona estructuració i organització del treball. Tenint en compte que sempre poden sorgir desviacions i imprevistos, sempre es seguiran aquestes dates de referència per tal d'assolir l'objectiu final.

NOM DE LA TASCA	INICI	FI	DURACIÓ
Realització del TFG	10 - 02 - 2020	02 - 07 - 2020	150 dies
Realització del pla de treball	10 - 02 - 2020	17 - 02 - 2020	7 dies
Revisió de l'estat d'art	10 - 02 - 2020	06 - 03 - 2020	25 dies
Recopilació de la informació	10 - 02 - 2020	20 - 02 - 2020	10 dies
Revisió de la informació	20 - 02 - 2020	06 - 03 - 2020	15 dies
Realització de la memòria	07 - 03 - 2020	25 - 06 - 2020	110 dies
Primera part - teòrica	07 - 03 - 2020	06 - 04 - 2020	30 dies
Bases teòriques i definició del problema	07 - 03 - 2020	06 - 04 - 2020	30 dies
Segona part - pràctica	06 - 04 - 2020	25 - 06 - 2020	80 dies
Anàlisi de les dades	06 - 04 - 2020	06 - 05 - 2020	30 dies
Modelització del estudi i resultats	06 - 05 - 2020	20 - 06 - 2020	45 dies
Realització del model	06 - 05 - 2020	05 - 06 - 2020	30 dies
Verificació i validació del model	05 - 06 - 2020	15 - 06 - 2020	10 dies
Obtenció de resultats	15 - 06 - 2020	17 - 06 - 2020	2 dies
Avaluació de resultats	17 - 06 - 2020	20 - 06 - 2020	3 dies
Extracció de conclusions	20 - 06 - 2020	25 - 06 - 2020	5 dies
Revisió de la memòria final	26 - 06 - 2020	01 - 07 - 2020	5 dies
Entrega de la memòria final	02 - 07 - 2020	02 - 07 - 2020	1 dia
Preparació de la presentació	03 - 07 - 2020	19 - 07 - 2020	16 dies
Presentació davant del tribunal	20 - 07 - 2020	24 - 07 - 2020	4 dies

Taula 2 – Planificació temporal del projecte

SECCIÓ 2

BASE TEÒRICA I EVOLUCIÓ DEL PROBLEMA

2.1. EFECTES DE L'EQUIPATGE DE MÀ

En aquest apartat es veuran els efectes que té l'equipatge de mà en tres aspectes diferents: econòmic, ambiental i operacional, plasmant la importància i les conseqüències positives que genera el fet de reduir aquest equipatge al màxim.

2.1.1. Econòmics

A conseqüència de la gran competència que tenen entre elles, les companyies aèries donen una gran importància a les seves despeses econòmiques per poder reduir-les i ser més competitives. A les figures següents es presenten diversos costos que han de suportar les aerolínies. La Figura 1 fa referència als costos de les companyies aèries europees i la Figura 2 a les americanes (Kolukisa, 2011).

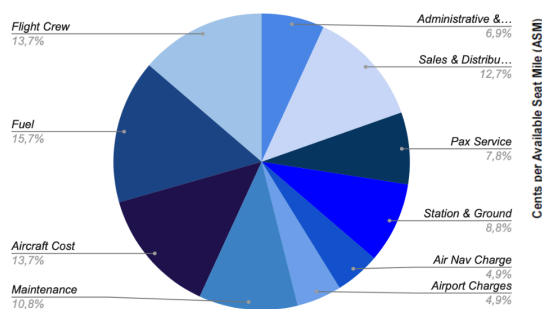


Figura 1 – Distribució dels costos de les companyies aèries Europees
Font: Adaptació (Smith, 2004)

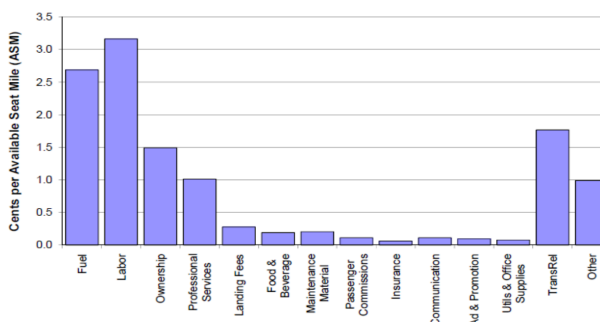


Figura 2 – Distribució dels costos de les companyies Americanes
Font: (Association, 2009)

Com podem observar en les dues figures, el cost del combustible és un dels costos més elevats per a les companyies tant europees com americanes i, per tant, té un paper molt important a l'hora de gestionar els costos d'aquestes.

El mercat del petroli és altament variable a causa dels pocs productors que hi ha al mercat. Això provoca que les companyies aèries utilitzin instruments i contractes financers amb asseguradores com a part de les seves estratègies de gestió de riscos per reduir els preus del combustible i així disminuir l'exposició d'una companyia a canvis inesperats del preu. (Lim & Hong, 2013)

El combustible que ha de portar un avió, depèn del pes de l'avió i el recorregut de la ruta que efectuarà, entre d'altres. És per aquesta raó, que l'equipatge de mà té un impacte directe sobre cost del combustible, ja que com més pes porti un avió, aquest tindrà una proporció de combustible més elevat.

Degut a que cada companyia aèria utilitza instruments i contractes financers amb asseguradores diferents, no es pot establir un preu de combustible genèric per observar l'impacte econòmic global. Tot i així, Eurocontrol emet informes amb estadístiques i dades i l'informe del juny del 2019 esmenta que el preu del petroli en el primer semestre del 2019 estava de mitjana a 58€ el barril. A la Figura 3, es mostren els índexs del petroli per barril i del fuel per tona des de l'any 2004 fins al juny del 2019. IATA en aquest informe preveia que la factura de combustible global de la indústria aèria seria de 180€ mil milions (\$206) en el 2019, dels quals en el 2018 van ser 158€ mil milions (\$180) dels quals van suposar un 23,5% de les despeses d'exploració. (Eurocontrol, 2019)



Figura 3 – Preus del petroli i del fuel a Europa
Font: (Eurocontrol, Industry Monitor, 2019)

Un estudi de la KLM determina que cada quilo addicional de pes a l'equipatge de mà sumava uns 1,3 eurocent⁴ de cost de combustible en un vol europeu i 15 cèntims en un vol internacional. Per tant, podem veure que el pes de l'equipatge té un efecte relativament reduït en l'ús de combustible si ho observem individualment, però si ho observem globalment calculant la gran quantitat de passatgers transportats, una reducció d'equipatge pot tenir un impacte significatiu.

⁴ 1 eurocent = 0,01€

A més a més, s'ha de tenir en compte que l'equipatge de mà no només augmenta el consum del combustible, sinó que també ocasiona retards en l'embarcament i, per tant, en la sortida del vol. Per pal·liar aquests retards es genera la necessitat de contractar personal addicional per agafar les bosses a la porta d'embarcament i baixar-les a la bodega de l'avió per fer front als problemes d'ocupació de l'equipatge de mà. Tots aquests factors comporten un cost molt més elevat. En el mateix estudi es calcula que el cost d'aquests agents addicionals s'estima internament al voltant d'un milió d'euros a l'any. (Bijker j. , 2016)

En els següents apartats, es seguiran fent mencions a costos econòmics en relació als problemes ambientals i operacionals.

2.1.2. Ambientals

El consum del combustible pot ser una bona indicació sobre l'impacte ambiental ja que, en general, és el que té un nivell d'emissions més elevat. Per calcular l'impacte ambiental, un dels factors que es pot utilitzar és l'impacte sobre el canvi climàtic. Per calcular-ho, el combustible consumit es converteix en emissions i aquestes es converteixen en la empremta sostenible.

Dins d'un article sobre les implicacions mediambientals que tenen les aerolínies quan trien la mida de l'aeronau, s'esmenta que l'impacte dels avions en el canvi climàtic està relacionat principalment amb les emissions de CO₂, NO_x i H₂O, de les quals ens interessen el CO₂ i l'H₂O ja que són les emissions que estan directament relacionades amb el consum de combustible i, per tant, es poden estimar de forma precisa. Per a les emissions de CO₂, s'utilitza una proporció de 3,15kg per cada kg de combustible (Sutkus, Baughcum, & DuBois, 2001) (Dings, Wit, Leurs, Davidson, & Fransen, 2003), en canvi, per a les emissions de H₂O s'utilitza 1,237 kg per cada kg de combustible (European Environmental Agency, 2005). A la taula 3 es presenten els costos del canvi climàtic de l'operació de diferents avions que operen un recorregut de 500nm. (Givoni & Rietveld, 2009)

Aeronau (seients)	30	8.3	Total
	Euro/tona	Euro/tona	
	CO2	H2O	
A320 - 200 (150)	2,49	0,27	2,76
A320 - 200 (164)	2,28	0,25	2,53
B747 - 400 (524)	2,34	0,25	2,59
B747 - 400D (568)	2,16	0,23	2,39
A319 - 100 (124)	2,76	0,3	3,06
A330 - 300 (335)	2,34	0,25	2,59

Taula 3 – Cost en sostenibilitat a partir d'un vol de 500 nm (Euro/seient).
Font: Adaptació (Givoni & Rietveld, 2009)

Com podem observar l'equipatge de mà comporta efectes ambientals negatius, ja que el pes de l'equipatge de mà comporta una quantitat de combustible superior i, a conseqüència, es generen més emissions.

Segons l'estudi de la KLM, cada 100 kg addicionals de pes s'afegeixen de mitjana, en un vol europeu, 3,5 kg de combustible i 11,08 kg d'emissions de CO₂ i, en un vol intercontinental 32,5 kg de combustible i 102,14 kg de CO₂. Respecte a la factorització de l'emissió d'altres gasos com el diòxid de nitrogen, l'òxid nítric i el vapor d'aigua, es proposa un factor de ponderació d'emissions d'1,7 segons l'article per Azar i Johansson 2012. Això suposaria l'equivalent a 173,6 kg de CO₂ per a un vol intercontinental.

L'estudi també esmenta que si el pes mitjà dels equipatges de mà dels passatgers es reduís en 1 kg, la reducció total anual de CO₂ ascendiria a més de 10.000 tones. Aquesta reducció és equivalent aproximadament a la quantitat de CO₂ produït en un any per 1.500 llars. La reducció del cost del combustible seria d'aproximadament 1,5 milions d'euros, aquests càlculs s'efectuen mitjançant la taula 4 proporcionada pel mateix article. (Bijker J. , 2016)

	Europa	Amsterdam - Roma	Intercontinental	Amsterdam - Los Angeles
Fuel per passatger	0,45 kg / 100 km		0,41 kg / 100 km	
CO ₂ per passatger	1,42 kg / 100 km		1,29 kg / 100 km	
Distància mitjana	780 km	1,297 km	7,918 km	8,977 km
Mitjana de fuel per passatger	3,5 kg	5,48 kg	32,5 kg	36,81 kg
Cost del fuel	1,30€	2,17€	12,00€	13,59€
CO ₂ per passatger	11,08 kg	18,27 kg	102,14 kg	115,92 kg

Taula 4 – Consum del combustible i emissions del diòxid de carboni
Font: Adaptació (Bijker J. , 2016)

2.1.3. Operacionals

Un dels factors més importants en les companyies aèries a l'hora de retallar costos sense comprometre la qualitat i augmentant el rendiment és amb el *turnaround*⁵. Si no s'apliquen mesures per reduir-lo, algunes empreses es troben en situacions en les quals acaben amb beneficis molt baixos o pèrdues molt elevades. (More & Sharma, 2014)

Quan un avió quan es troba a terra ferma, els ingressos que genera són nuls i, per tant, genera una despesa a la companyia, ja que aquesta paga un preu per minut destinada a l'aeroport per els serveis que està rebent. L'estalvi d'uns quants segons durant l'embarcament de passatgers suposa una diferència en el rendiment, tant econòmicament com en el registre de puntualitat. (DeVries, 2009)

Si els retards es transformen en costos econòmics, es calcula que un operador europeu posicionat entre els deu primers, assumirà els costos de retard anuals entre 100 i 400 milions d'euros en l'any 2000, la qual cosa mostra la importància d'una gestió correcta en les operacions (Niehues, y otros, 2001). A més a més, les companyies aèries han de suportar sancions força considerables quan els vols es retarden. Qantas, la companyia australiana, estima que la millora de l'1% de la puntualitat de les operacions aporta un benefici addicional de 15 milions de dòlars en un any (Guo, 2005). En el mateix article on s'esmenta tota aquesta informació, es mostra un gràfic (Figura 4 (Niehues, y otros, 2001)) on es pot observar que les companyies aèries puntuals semblen ser més rendibles i es tradueix en una millora potencial de beneficis de 4 a 16 milions d'euros, depenent de la mida de la companyia aèria.

⁵ *Turnaround*: es defineix com el total de temps que l'avió està a terra, és a dir, el temps necessari per descarregar un avió després de la seva arribada a la porta i preparar-lo per a la sortida de nou.

Anàlisi per a la Gestió Eficient de l'Equipatge de Cabina en l'Aviació Comercial

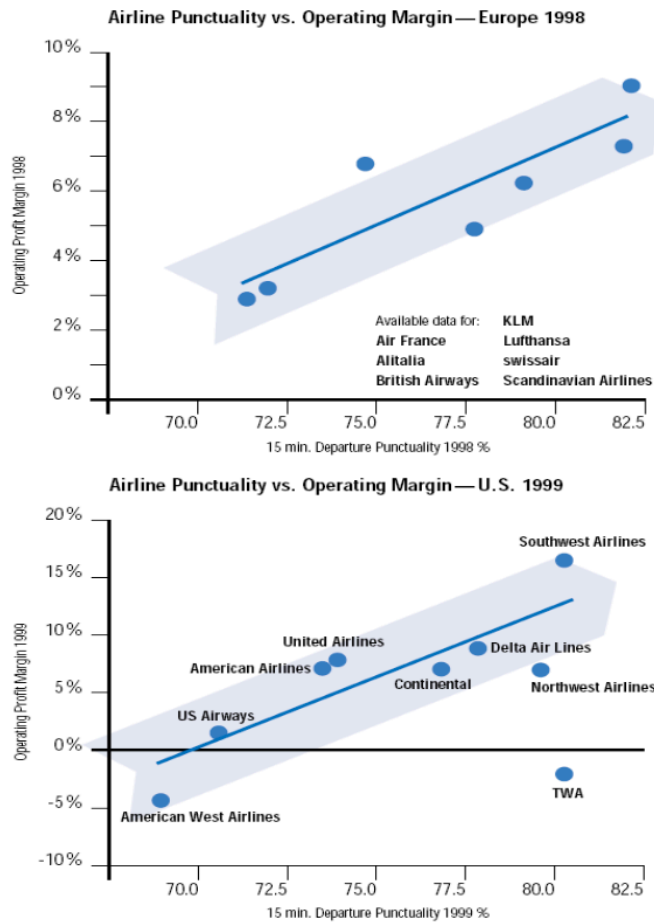


Figura 4 – Puntualitat d’una companyia vs. Marge operatiu
Font: (Niehues, y otros, 2001)

A més a més, l'article parla de les causes que generen retards en els vols.

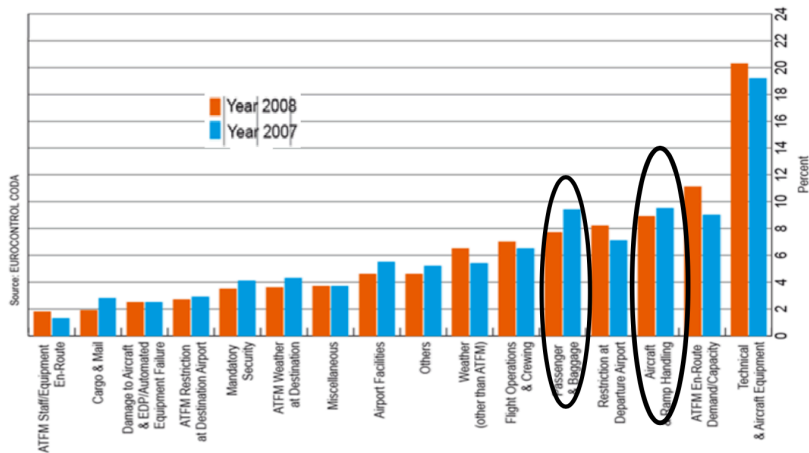


Figura 5 – Causes dels retards agrupades per codi IATA a Europa
Font: (Eurocontrol, 2008)

Tal i com es mostra a la Figura 5, van aparèixer dos motius de retard molt importants en el moment de procedir al *turnaround* en el 2008 i el 2007, que són el maneig de la rampa

i *handling* i la manipulació dels passatgers i els equipatges (Eurocontrol, 2008). (Kolukisa, 2011)

Avui en dia, encara esdevenen un problema per a les companyies. Com es pot observar a la Figura 6, on es mostra un gràfic actualitzat (Juny 2011, Juliol 2012) dels factors que ocasionen retards. La manipulació dels passatgers i els equipatges segueix tenint un índex alt de retards. (Eurocontrol, 2012)

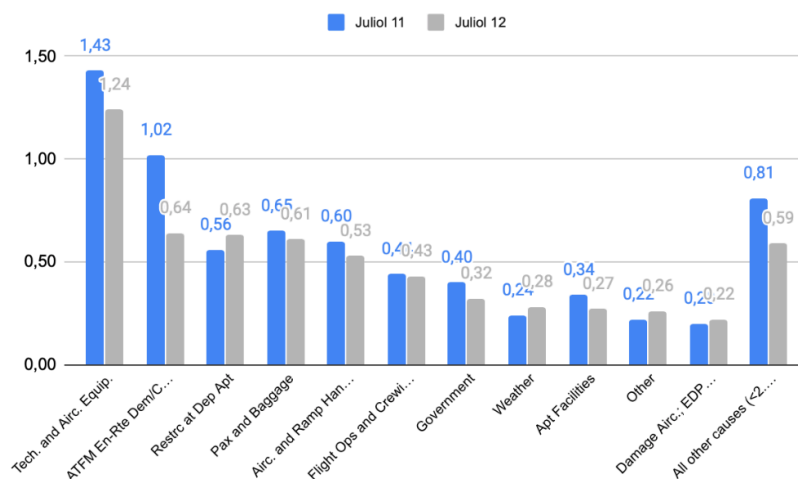


Figura 6 – Retard de totes les causes primàries i reaccionàries, mitjançant codi IATA (%)
Font: Adaptació (Eurocontrol, 2012)

L'article (Wu & Caves, 2002) esmenta que les estimacions de pèrdues monetàries relacionades amb el retard d'una companyia aèria oscil·len entre els \$45 i els \$150 per minut.

Una gran quantitat d'equipatge de mà ocasiona més temps a l'hora de procedir a l'embarcament, ja que ocupa un temps significatiu a l'hora de guardar l'equipatge en el compartiment i també quan no hi ha suficientment espai en aquest, ja que s'ha de buscar lloc o baixar les maletes a bodega. És per aquesta raó que genera un impacte negatiu dins les operacions de la companyia aèria i pot provocar retards.

Una de les causes que fan que hi hagi més equipatge de mà és que les companyies aèries augmentin les tarifes de l'equipatge de bodega, fet que genera la disminució de la velocitat d'embarcament. (Nyquist & McFadden, 2008) demostren que la diferència en

els temps d'embarcament quan els passatgers tenen dues maletes en comptes de zero és gairebé el 60%. (Ball, y otros, 2010) estimen que el cost total dels retards de les companyies aèries el 2007 va ser de 29.000 milions de dòlars només als Estats Units. (Milne & R.Kelly, 2013)

Dels 9.440 vols que operava la KLM el gener del 2016, en 108 es van ocasionar retards provocats per les maletes de mà, aproximadament un 1%. El retard total d'aquests vols va ser superior a 9 hores, segons dades proporcionades per Ricardo Kallendorf (KLM) el 18 de març de 2016 a l'article (Bijker J. , 2016). En el mateix article, no només s'esmenta la conseqüència econòmica que genera una companyia aèria quan tenen vols amb retards, sinó que també mostra greus conseqüències per a les rutes on han patit retards. Degut a que els vols de la KLM que operaven a Nova Delhi (Índia) patien retards molt sovint, moltes vegades a causa dels problemes d'embarcament causats per les maletes de mà, la companyia va estar a punt de perdre els seus *slots* en aquell aeroport. Per tal de solucionar aquell problema i mantenir els *slots*, en aquells vols es va posar en marxa un equip especial amb agents a la porta d'embarcament per tal d'agilitzar el procés de les maletes de mà.

2.2. INICIATIVES QUE S'HAN DUT A TERME FINS EL MOMENT

Les diferents restriccions de mides, pes i número de peces en el que es refereix a l'equipatge de mà, pot variar en funció de la companyia aèria, tarifa contractada i dimensió de l'avió. Tot i així, IATA6 recomana unes dimensions generals que no haurien d'excedir els 56x18x25 cm, incloent les rodes, butxaques i nanses de la maleta. La majoria d'aerolínies també apliquen limitacions de pes, les quals comencen sobre els 5 kg. (IATA, 2020)

IATA Cabin OK

L'any 2012 IATA amb el conjunt d'aerolínies que la conformen i amb els fabricants de maletes, van anunciar una nova iniciativa per tal d'optimitzar l'espai per a l'equipatge de mà a la cabina de l'avió. Aquesta iniciativa proposa unes dimensions específiques per l'equipatge de mà de 55x35x20, on teòricament, en un avió de 120 seients o més, tots els passatgers tindrien espai per la seva maleta, inclús en vols en els que s'haguessin venut la totalitat dels seients. De tota manera aquesta iniciativa no és d'obligat compliment, i en molts casos, les dimensions proposades són inferiors a les de moltes companyies.

Per identificar les maletes que compleixen amb les mides proposades, aquestes porten un distintiu únic per tal de que el personal de l'aerolínia identifiqui ràpidament de que es tracta d'una maleta IATA Cabin OK. Els passatger que la seva maleta tingui aquest distintiu tenen alts beneficis en companyies que hagin adoptat aquesta mesura, ja que la seva maleta tindrà prioritat per ser transportada a la cabina de l'avió en cas de que el volum de maletes superi la capacitat d'aquesta. (IATA, 2015) (IATA, 2015).

⁶ IATA és l'Associació Internacional del Transport Aeri. La seva missió és representar, liderar i servir a la indústria aeronàutica a través de l'estandardització d'aquesta. (IATA, Vision and Mission, s.f.)

Cost extra per l'equipatge de mà

Hi ha diverses companyies que per tal de controlar la problemàtica de l'equipatge de mà, han imposat un cost addicional per aquest. Amb aquesta mesura, les aerolínies poden saber amb antelació la quantitat de passatgers que portaran equipatge de mà a cada vol, així com garantir que aquest pugui ser transportat a la cabina de l'avió i no a la bodega. Algunes d'aquestes companyies, com és el cas de la nord americana Spirit (SPIRIT, 2017), al·leguen que una altra de les raons per les que han decidit cobrar per l'equipatge de mà és per poder mantenir unes tarifes baixes.

Des de novembre de 2018 si viatges amb Ryanair (RYANAIR, 2019) i Wizz-Air (WIZZ-AIR, 2018), l'equipatge de mà no està inclòs. Pots portar de forma gratuïta un bulto de petites dimensions de 40x20x25cm (en el cas de Ryanair) i 40x30x23cm (en el cas de Wizz-Air) que ha de poder ser emmagatzemat sota del seient davanter. En el cas de que vulguis dur una maleta addicional de majors dimensions, de fins 55x40x20cm (en el cas de Ryanair) i 55x40x23cm (en el cas de Wizz-Air), has de pagar el que ells anomenen *Priority* (que inclou prioritat per embarcar i el transport de l'equipatge de mà a la cabina de l'avió). En el cas de que no paguis aquest extra i et presentis a la porta de l'avió amb un bulto addicional, és obligatori pagar una penalització.

Spirit Airlines, des de l'any 2017 aplica una política semblant a la esmentada anteriorment amb Ryanair i Wizz-Air. Amb la diferència de les dimensions, que en aquest cas serien, una maleta de menors dimensions inclosa en tots els seus bitllets de fins a 45x35x20cm i una maleta addicional amb un cost extra de fins a 56x46x25cm. A més el fet de pagar per la maleta de mà no t'assegura que la puguis transportar a la cabina de l'avió.

Delta Early Valet

Delta Airlines consta d'un servei anomenat Delta Early Valet, en tots els seus *hubs* nacionals, incloent Atlanta, Seattle, Minneapolis/St.Paul i New York-JFK. Disponible en els avions de fuselatge estret⁷ en determinades rutes, i orientat a famílies que viatgen amb nens petits, passatgers amb necessitats especials, gent gran i clients amb un alt nivell de fidelització. Consisteix en que el personal de terra, uns 45 minuts abans de la sortida de l'avió, s'encarrega d'emmagatzemar l'equipatge als compartiments corresponents. Per fer-ho les maletes estan etiquetades amb el número de seient del passatger.

S'ha pogut comprovar que aplicant aquesta iniciativa hi un estalvi d'un minut i mig en el temps mitjà d'embarcament, en el cas de que es porti a terme amb 6-9 maletes. A més a més, la satisfacció dels clients augmenta un 8% en comparació amb els vols on aquest servei no està disponible. (Flenoury, 2015)

EasyJet App

La companyia EasyJet té una aplicació a través de la qual pots mesurar les dimensions del teu equipatge de mà. En el cas de que la maleta no s'ajusti a les mesures indicades, dins la mateixa aplicació pots comprar una maleta facturada, o sinó canviar la maleta per una de menors dimensions. (EASYJET, s.f.)

Door to door baggage

El *door to door baggage*, consisteix en enviar un equipatge d'un punt a un altre, en el qual el client, no s'hagi de preocupar de cap tràmit, tals com portar-lo fins l'aeroport, facturar-lo i recollir-lo a la cinta d'equipatges. Sinó que el client rep directament la maleta en el punt escollit de destí. Aquesta alternativa està disponible a gran part del món, però té un cost elevat i has de preparar la maleta amb més dies d'antelació que si ho fessis de forma tradicional. I a més a més, és una opció escollida majoritàriament per transportar equipatges de dimensions superiors a l'equipatge de mà, per tant no té un efecte notable sobre aquest. (Send My Bag, s.f.) (My Baggage, s.f.)

⁷ Són aquells avions on el fuselatge de cabina és menor als 4 metres i els seients es distribueixen en un únic passadís central. Són normalment utilitzat en rutes de curt radi. (Real Academia de Ingeniería, s.f.)

Distribuïdors de maletes

També hi ha nombrosos portals de venda de maletes online com físics en els quals disposen de mesuradors i indicacions a les etiquetes de la maleta, sobre quines companyies són les que accepten les dimensions de la maleta en qüestió. Tot i ser una bona iniciativa, no es pot considerar 100% fiable, a no ser que estigui avalada per la companyia aèria, ja que les dimensions poden variar d'un any per un altre i quedar desfasades.

Level

La companyia Level durant el procés de facturació comprova i etiqueta totes les maletes de mà conforme són acceptades a la cabina de l'avió, fins exhaurir les etiquetes, ja que n'hi ha tantes com espais disponibles a la cabina. En el cas de que no s'exhaureixin durant el procés de facturació, aquestes etiquetes es porten a la porta d'embarcament, i abans d'iniciar-lo s'acaben de col·locar a les maletes de mà dels passatger que no han passat pel mostrador de facturació. Si durant aquest procés queden maletes per etiquetar i ja no queden etiquetes, significa que les maletes restants hauran de viatjar a la bodega de l'avió. D'aquesta manera es pot fer una previsió molt més concreta de quantes maletes s'hauran de baixar a la bodega i de si és millor la col·laboració d'un o dos agents per agilitzar el procés. Tanmateix en vols molt plens en el moment de realitzar la facturació s'ofereix la possibilitat de facturar l'equipatge de mà de forma gratuïta.

KLM

Una de les altres alternatives aplicades per la companyia KLM són unes etiquetes que es col·loquen a l'equipatge de mà. Una etiqueta de color groc per la maleta de majors dimensions en la que s'especifica que no està garantit que aquesta pugui ser transportada a la cabina i que pot ser necessari baixar-la a bodega una vegada a la porta d'embarcament, sense cost. I una etiqueta de color blau pel bulto de menors dimensions indicant que aquest té garantit el seu transport a cabina, sempre i quan es col·loqui sota del seient davanter. (Bijker J. , 2016)

Vueling

Vueling té la possibilitat de facturar l'equipatge de mà en el moment de la facturació per 4€, sempre i quan aquest compleixi amb les mesures i pes reglamentaris. D'aquesta manera els passatgers no es tenen que preocupar de carregar la maleta fins la porta d'embarcament. Tot i així aquesta opció només és possible en el mostrador i no a les màquines d'autofacturació, obligant per tant als passatger el procés de facturació en dos parts. Tanmateix, a la porta d'embarcament Vueling també dona la possibilitat de facturar la maleta de mà una vegada a la porta d'embarcament abans de començar a embarcar, oferint a canvi l'embarcament prioritari.

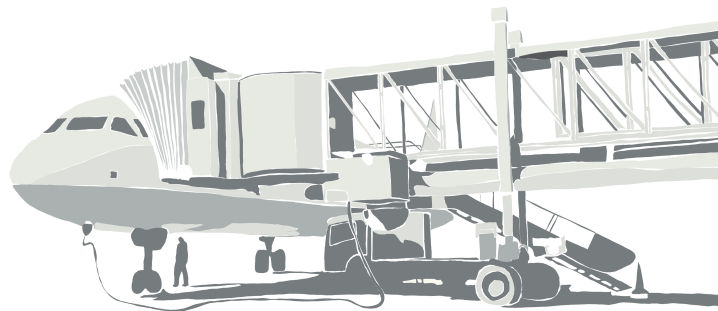
Tenim companyies que posen restriccions i companyies que busquen alternatives, tot i que algunes alternatives busquen eliminar el retràs causat per les maletes, no disminueixen la quantitat d'aquestes.

2.3. POLÍTIQUES D'EMBARCAMENT

Per poder procedir a l'embarcament per accedir a dins de l'avió, les companyies aèries utilitzen dues maneres diferents per fer-ho. Una d'elles és quan l'avió està en remot i s'accedeix des de la terminal mitjançant l'autobús i les escales per entrar dins de l'avió (il·lustració 1) i l'altre és mitjançant la passarel·la, o més conegut com *finger*, connectat directament des de la terminal (il·lustració 2).



Il·lustració 1 – Avió en remot
Autor: Ramon Illan Salas



Il·lustració 2- Avió amb el finger
Autor: Ramon Illan Salas

Ens centrarem amb l'embarcament mitjançant el *finger* ja que és el que genera més temps per embarcar perquè només es pot accedir a l'avió per la porta davantera. Normalment amb l'avió en remot les dues portes estan operatives i si és així l'embarcament s'efectua més ràpidament.

Les estratègies més utilitzades per embarcar són:

a. Back-to-front

L'embarcament *Back-to-front* és l'estratègia més utilitzada entre les companyies aèries. L'embarcament s'efectua primer mitjançant el grup prioritari (número 1) i seguidament s'embarquen els passatgers dividits en grups (en aquest cas 5 grups) en un ordre de posterior fins al davant de l'avió (números 2,3,4,5,6) tal i com es representa a la Figura 7.

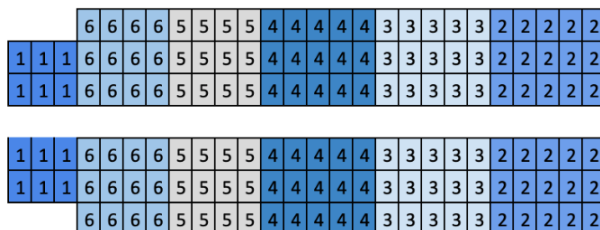


Figura 7 – Estratègia d'embarcament back-to-front
Font: Adaptació (Mas, 2012)

Aquesta estratègia provoca que sigui molt fàcil d'entendre i d'efectuar-la per als passatgers. Un dels principals avantatges per a les famílies i per els grups és que poden efectuar l'embarcament junts. Tot i així, no és que sigui l'estratègia més eficient. El principal problema que genera són les interferències que obtenen els passatgers en el moment de dipositar la maleta de mà en el compartiment i seure en el seient corresponent.

b. Random

Aquesta estratègia consisteix en entrar a l'avió aleatòriament. Els passatgers tenen assignat el seu seient i com s'observa a la Figura 8, primer de tot entrarien els passatgers prioritaris (1). Un cop han embarcat els prioritaris, la resta de passatgers (2) entren a l'avió en cap ordre predefinit, és a dir amb un procés de FIFO (*first-in first-out*).

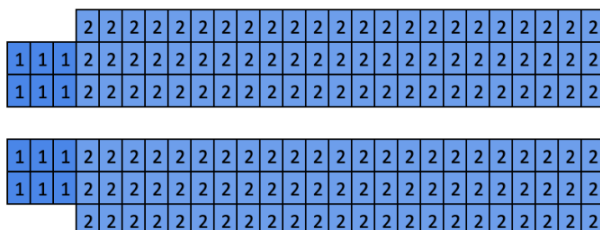


Figura 8 – Estratègia d'embarcament Random
Font: Adaptació (Mas, 2012)

davant de tot (3), després aniria el grup (4) posicionat casi al final i conclouríem l'embarcament alternant els grups (5 i 6).

				3	3	3	3	3	5	5	5	5	6	6	6	6	4	4	4	4	2	2	2	2
1	1	1	3	3	3	3	3	5	5	5	5	6	6	6	6	4	4	4	4	2	2	2	2	
1	1	1	3	3	3	3	5	5	5	5	6	6	6	6	4	4	4	4	2	2	2	2		

Figura 10 - Estratègia d'embarcament Rotating zone
Font: Adaptació (Mas, 2012)

Amb aquesta política aconseguies embarcar dos grups diferents en el mateix moment. Quan hi ha interferències en el grup (2) per exemple, la cua pot ocupar bona part del passadís de l'avió, però gràcies a que el següent grup és el de davant de tot (3), aquests passatgers poden començar a seure en el seu seient i així agilitzar processos.

e. No assigned seat

Aquesta és una estratègia poc usual. Consisteix a fer un embarcament on els seients no estiguin assignats, és a dir el passatger pot seure a qualsevol lloc. L'avantatge d'aquesta política és que obligues al passatger a ser puntual a la porta d'embarcament, ja que com més ràpid entri el passatger a l'avió, més possibilitats de seure a on ell vulgui. El nombre d'interferències es poden minimitzar, ja que les persones ocupen qualsevol lloc que estigui disponible. Però aquesta política comporta molt mal estar als passatgers i no genera una bona satisfacció al client.

f. Individual

L'embarcament *individual* consisteix en una combinació de les estratègies *back-to-front* i *outside-inside boarding*. És considera una de les estratègies d'embarcament més ràpides, ja que agafa les parts més positives de cada una de la combinació d'estratègies esmentades.

Per tal de tenir una visió general de la optimitat de les diferents estratègies d'embarcament i com afecte l'equipatge de mà a aquestes polítiques, observarem l'estudi desenvolupat per (Schultz, 2018).

Aquest estudi només té en compte les polítiques de *back-to-front*, *random*, *outside-inside boarding* i *individual*, però ens servirà per veure el temps que generen les polítiques a l'hora d'embarcar i les conseqüències de l'equipatge de mà. El temps que utilitza per l'embarcament de qualsevol estratègia a l'atzar s'utilitza com a referència un valor del 100%. Podem observar a la Figura 11 el temps mitjà d'embarcament que es mostra en funció a una taxa de conformitat creixent. Evidentment, l'estratègia *random* no es pot veure afectada pels canvis en la taxa de conformitat. L'estratègia *back-to-front* mostra un temps d'embarcament decreixent al principi, però després d'un temps d'embarcament mínim (entre 95% i 75% de la taxa de conformitat), l'embarcament augmenta de nou. Per a les estratègies *outside-inside boarding* i *individual* millora progressivament i augmenta la conformitat.

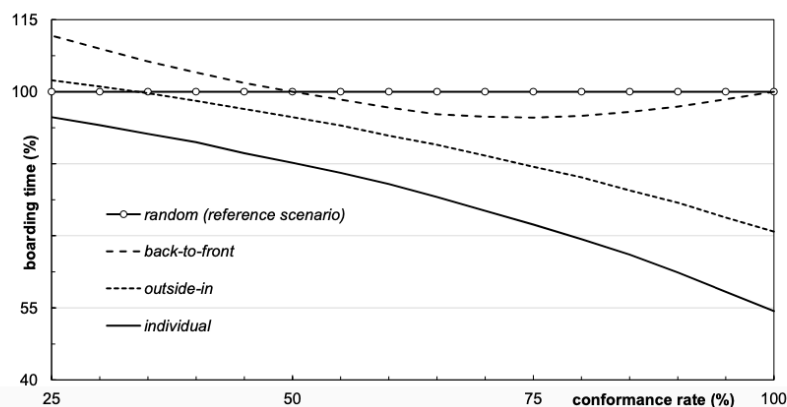


Figura 11 – Temps d'embarcament de diferents estratègies d'embarcament amb un ritme de conformitat creixent.

Font: (Schultz, 2018)

Com ja hem esmentat en punts anteriors, una gran quantitat d'equipatge de mà provoca que augmenti el temps d'embarcament en qualsevol política a causa de compartiments que es queden bloquejats i moviments addicionals dels passatgers a contra flux en el passadís de l'aeronau per tal de buscar espais lliures per deixar el seu equipatge. Per tant, si s'aconsegueix reduir l'equipatge de mà, provocaria un estalvi potencial de temps d'embarcament. A la Figura 12, es mostra una reducció del temps de l'embarcament si el nombre d'equipatge de mà es redueix d'una mitjana d'1,5 a 0 articles per passatger. En el plantejament inicial s'assumeix almenys una peça d'equipatge per passatger i les probabilitats per a una peça, dues i tres són del 60%, 30% i 10%, respectivament, i es pren com a base per reduir l'equipatge de mà. Com es mostra a la figura, totes les polítiques d'embarcament generen un embarcament més ràpid, concretament, el

temps d'embarcament es redueix en un 34%, un 33%, un 25% i un 12% a les estratègies per *random*, *back-to-front*, *outside-inside boarding* i *individual*, assumint que els passatgers no portin equipatge de mà.

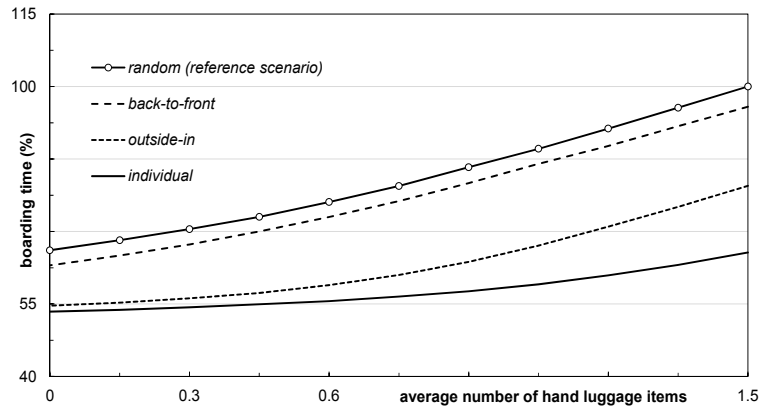


Figura 12 – Temps d'embarcament de diferents estratègies d'embarcament amb disminució de l'equipatge de mà
Font: (Schultz, 2018)

En conclusió, independentment de la política d'embarcament que utilitzi una companyia aèria, obtenir una reducció d'equipatge de mà a la porta d'embarcament, ajuda a la companyia a efectuar els embarcaments molt més ràpid i en conseqüència ajudar a retallar costos, ser més puntuals i obtenir majors beneficis. Tot i així, com podem observar en les dues figures anteriors, la política de *back-to-front* és la única que acaba sent creixent en les dues situacions.

2.4. ESTUDI DEL PASSATGER

2.4.1. Elecció del tipus d'equipatge

A través de l'eina "Hand baggage prediction tool" (Bijker J. , 2016) durant el juliol de 2015 es va preguntar a 8.577 passatgers de la KLM si portarien una peça d'equipatge de mà gran (de les que no caben a sota del seient davanter i han de ser guardades en els compartiments superiors). Els resultats obtinguts els van creuar amb la informació dels passatgers, observant que: Un 54,9% dels passatgers sí que porten un bulto de mà gran. Com es pot observar a la Figura 13, un 61% del 27% de passatgers que no facturen equipatge porten un bulto de mà gran. Un 54% factura una maleta, fent que la xifra dels passatgers que porten equipatge de mà gran es redueixi a un 49% d'aquests. Tot i així en els casos de passatgers que facturen dues maletes (15%), el percentatge de passatgers que porten equipatge de cabina gran augmenta al 64%.

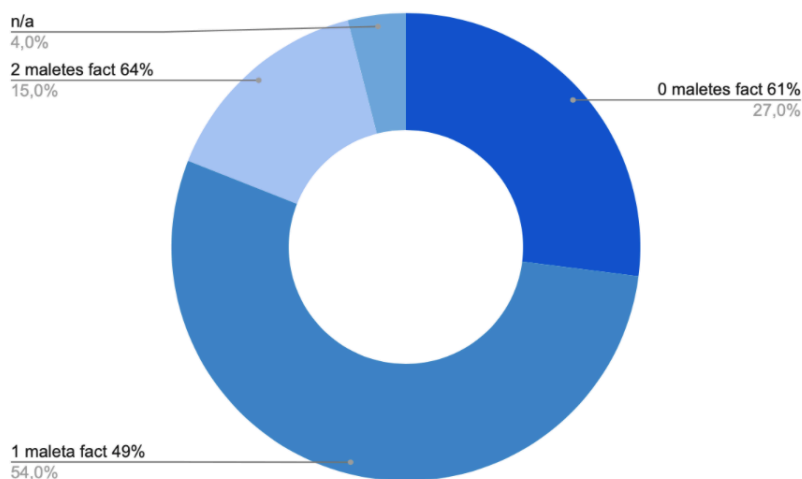


Figura 13 – % de passatgers amb maleta de mà gran x número de maletes facturades

Font: Adaptació (Bijker J. , 2016)

Per tant podem arribar a la conclusió que el fet de facturar equipatge no és determinant a l'hora de decidir si portar equipatge de mà o no.

Dels diferents mètodes per fer la facturació, l'online és la que més utilitzen els passatgers que decideixen portar una peça gran a la cabina amb un 64%. Cosa que ens porta a pensar que és un perfil de públic en el que haurem d'orientar gran part de les polítiques que proposem.

En el mateix any es va portar a terme una enquesta online a 82 persones aleatòries, de les quals un 51% eren estudiants i un 76% menors de 34 anys. Com podem veure a la Figura 14, dels enquestats un 43% va dir que viatjava entre 1-2 cops per any, un 34% entre 3 i 12 cops i el 16% restant ho fa esporàdicament.

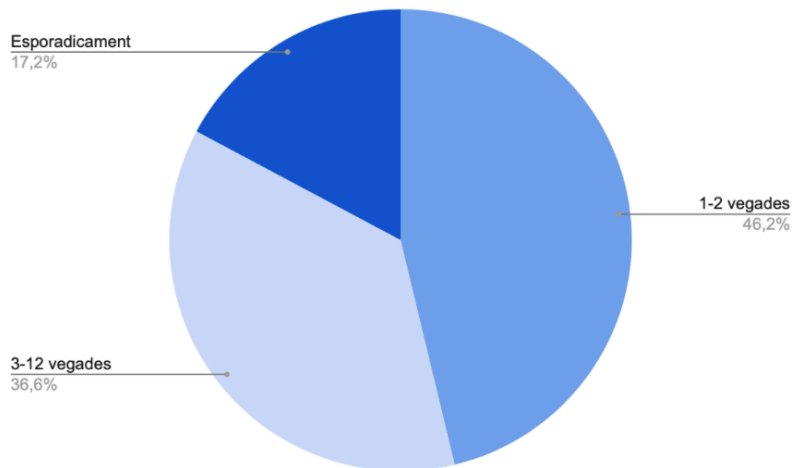


Figura 14- Número de viatges per any
Font: Adaptació (Bijker J. , 2016)

A la Figura 15 es pot observar que només un 2% dels enquestats factura la totalitat del seu equipatge envers a un 39% que només porta equipatge de mà. Un 20% en canvi fa una combinació entre els dos, mentre que el 39% restant al·lega que depèn de varies circumstàncies.

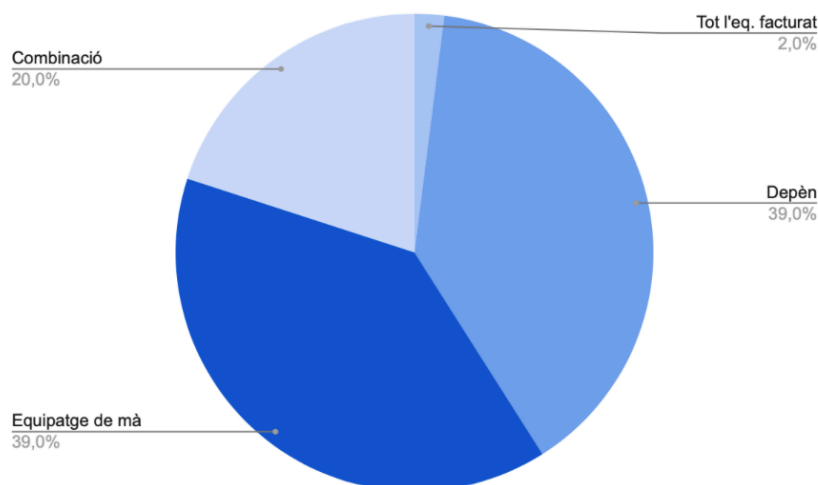


Figura 15- % de passatgers segons el tipus d'equipatge i la seva facturació
Font: Adaptació (Bijker J. , 2016)

Els factors que més vegades es van repetir a l'hora d'escollir quin tipus d'equipatge portar són: la durada del viatge, com més llarg sigui el viatge més equipatge s'acostuma a necessitar i en conseqüència l'opció de facturar es converteix en la més adient. També

hi influeixen factors com la finalitat del viatge, cost del bitllet així com de la facturació d'equipatge i duració del vol.

En el cas de portar equipatge amb unes dimensions acceptades per ser transportat a cabina, les principals raons per no facturar-lo són el temps, la seguretat i el cost.

Temps

El temps es refereix en el temps extra que es necessita per facturar-lo, ja que una persona que només porti equipatge de mà i tingui la facturació prèviament feta des de casa no haurà de passar pel mostrador o per les màquines d'autofacturació. A més del temps que cal esperar una vegada arribes al destí per tal de recollir l'equipatge de la cinta.

Seguretat

Seguretat en el sentit de que un passatger si porta la maleta amb ell en tot moment, aquest té la certesa de que les seves pertinences no corren cap tipus de perill de ser extraviades.

Cost

Finalment el cost extra que avui en dia suposa la facturació de l'equipatge és un determinant de pes a l'hora de decidir si facturar-lo o no, quan es té la possibilitat de portar-lo a cabina de forma gratuïta.

Es va preguntar també si totes les coses que posava el passatger a la maleta, finalment les utilitzava al llarg del viatge. Un 52% va respondre que no envers a un 48% que sí. Aquestes coses que acabaven no sent utilitzades són un 38% roba, 12% llibres, kit amb medicaments de primera necessitat, elements de higiene personal i accessoris electrònics. La majoria d'aquests elements són transportats en l'equipatge de mà, de manera que es podrien estudiar alternatives per evitar o si més no reduir la necessitat de transportar certs elements, com per exemple en el cas dels llibres, oferint llibres a bord entre altres.

2.4.2. *Customer journey*

Un 50% dels enquestats va contestar que decideix si facturar maletes abans de comprar el bitllet, a diferència d'un 30% que ho fa durant el procés de compra. Un 10% ho fa durant el *check-in online* mentre un 10% ho fa just abans de viatjar. Les diferents fases per les que passa un client durant el *customer journey* són (Avram, 2017):

El procés de compra és la primera fase. Engloba la part de planificació del viatge, i cerca de possibles opcions per arribar al destí desitjat, així com la comparació amb altres aerolínies i mètodes de transport.

D'acord amb Travelport (Booth, 2016), dos dies abans del vol seria considerat la segona fase, en la que un 30% dels passatgers realitzen compres com equipatge per facturar, serveis de menjar a bord, etc.

La tercera fase seria la de l'arribada a l'aeroport. En aquesta fase les aerolínies aprofiten per aplicar ofertes d'últim minut, així com ofertes per fer més agradable l'estada del passatger en el aeroport, ja que aquesta pot resultar estressant.

Durant el vol tindriem la quarta fase, en la que un bon servei a la cabina és clau.

Una vegada en el destí seguiríem amb la cinquena fase, que englobaria el moment de recollir l'equipatge en el cas de que aquest hagués sigut facturat i la sortida de l'aeroport fins a la destinació final. Tot i que en el cas de tenir un vol en connexió, tornariem a la fase 3.

Finalment la última fase seria una vegada a la destinació final, la qual és un punt estratègic per crear fidelitat amb el passatger.

Les millors fases en les que podríem involucrar polítiques per reduir l'equipatge de mà i animar als passatgers a que el facturin seria a la primera, segona, tercera i cinquena fase.

2.4.3. *Social dilemma theory*

En el moment de decidir si facturar l'equipatge de mà o no, un passatger es troba davant d'un dilema social, en el que preguntes tals com l'impacte que tindrà prendre una decisió o una altra, què faran els altres passatgers, què en puc treure jo a canvi, si valdrà l'esforç, entre d'altres, apareixen en el context. Les respostes a aquestes preguntes així

com saber identificar quina és la resposta segons el tipus de passatger i dilema plantejat, és molt important.

Un dilema social és una situació en la que hi ha un conflicte d'interessos entre prendre la decisió més beneficiosa per un mateix o aquella que beneficia a un major número de persones.

Tot dilema social comparteix una estructura. Cada una de les persones involucrades en aquest tenen dues opcions: cooperar (decisió que beneficia més al col·lectiu) o desertar (decisió que beneficia més a l'individu). Alguns d'aquests dilemes socials tenen estratègies dominants. En aquests casos desertar és considerada l'estratègia dominant, és a dir, aquella que produeix el major benefici per l'individu, independentment del que qualsevol dels altres implicats en el dilema faci. Tot i així les diferents estratègies dominants (desertar) tenen un punt d'intersecció en el que es produeix un equilibri deficient. On si tots els implicats en el dilema trien desertar, el resultat serà pitjor que si tots els individus haguessin escollit estratègies no dominants, és a dir cooperar.

Els dilemes socials es poden classificar en dos grups, els *Two-Person Dilemmas*, que són aquells que involucren a dues persones i els *N-Person Dilemmas*, aquells que involucren a més de dues persones. Per modelar els diferents tipus de dilemes s'utilitzen diferents històries metafòriques. On segons l'equilibri de *Nash*, que és un concepte de la teoria dels jocs, el resultat més òptim és aquell en el que cap jugador té l'incentiu de canviar la seva estratègia una vegada coneguda la dels demés. Com a mínim existeix un equilibri de *Nash* per tots els jocs finits.

En els *N-Person Dilemmas*, les accions d'un, no són necessàriament conegudes pels demés, aquestes poden ser anònimes, en canvi, en els *Two-Person Dilemmas* cada persona sap com ha actuat l'altre.

El cost que imposa un, quan pren la decisió de desertar, és a dir la més beneficiosa per ell, en els *N-Person Dilemmas* es reparteix entre el grup, mentre que en els *Two-Person Dilemmas* el cost recau al complet sobre l'altre.

Finalment, en els *Two-Person Dilemmas* cada persona té un control significatiu sobre quins poden ser els possibles resultats del comportament de l'altre, mentre que en els *N-Person Dilemmas* es té un control molt petit.

Per modelar els *Two-Person Dilemmas*, s'utilitzen principalment 3 històries metafòriques:

Prisoner's Dilemma (desenvolupat per Merrill Flood i Melvin Dresher i formalitzat per Albert Tucker)

En aquest dilema hi ha dos presoners en els quals se'ls ofereix la possibilitat de callar (cooperar) o de confessar (desertar). Els possibles resultats són: Si els dos confessen aniran 6 anys a la presó cada un (CC), si el presoner 1 confessa i el presoner 2 no ho fa, el presoner 1 anirà 10 anys a la presó i el presoner 2 quedarà en total llibertat (DC) i viceversa (CD), i finalment si els dos presoners queden callats, aniran els dos a la presó, però només 1 any (DD), on $DC > CC > DD > CD$.

En aquest dilema hi ha una estratègia dominant, sigui quina sigui l'opció que l'individu contrari faci, sempre serà millor desertar. Tot i així les diferents estratègies dominants (desertar) tenen un punt d'intersecció en el que es produeix un equilibri deficient. És a dir, si tots els implicats en el dilema trien desertar, el resultat serà pitjor que si tots els individus haguessin escollit estratègies no dominants, és a dir cooperar.

Assurance game

Normalment representat pel joc *Stag Hunt* de Jean-Jacques Rousseau en el que es planteja a dos caçadors l'opció de caçar un conill, la qual cosa pot fer de forma individual o caçar un cérvol entre els dos. Si cacen entre els dos el cérvol, tindran més carn per cada un que si cacen un conill o dos. A més han de prendre la decisió sense saber com actuarà l'altre persona, i una vegada presa ja no es pot canviar. Els possibles resultats són: Que el caçador 1 i el caçador 2 decideixen cooperar l'un amb l'altre per caçar el cérvol (CC), que el caçador 2 decideixi cooperar i caçar un cérvol, però el caçador 1 decideixi caçar el conill, aconseguint per tant 2 conills el caçador 1 i res el caçador 2 (DC) i viceversa (CD), o que el caçador 1 i 2 decideixin caçar un conill cada un (DD), on $CC > DC > DD > CD$.

En aquest joc els implicats estarien disposats a cooperar sempre i quan l'altre també ho estigués. Es tendeix a pensar que en aquest joc no hi ha dilema i que la resposta lògica seria cooperar, però lo cert és que cooperar no és una estratègia dominant, ja que la forma en la que decideixi actuar l'altre individu farà variar si el millor resultat és cooperar o desertar. Per tant, aquest joc té dos equilibris: La cooperació mútua (CC), que és un equilibri òptim i desertar mútuament (DD), que és un equilibri deficient. Aquest model és dels que s'ajusta més a la majoria de dilemes socials en els que ens trobem dia a dia.

Chicken game

En aquest joc tenim a dos individus que condueixen un cotxe l'un contra l'altre. El conductor que es desvia primer és considerat un "gallina" i perd el respecte de la gent, mentre que el conductor contrari guanya fama per la seva valentia. Els possibles resultats són: El conductor 1 i 2 decideixen cooperar i desviar-se, per tant es converteixen els dos en gallines (CC), que el conductor 2 decideixi cooperar i desviar-se, mentre que el conductor 1 no ho fa, aconseguint per tant la victòria pel conductor 1 i convertint-se en gallina el conductor 2 (DC) i viceversa (CD). O que tant el conductor 1 com el 2 decideixin no frenar cap dels dos i es morin (DD), on $DC > CC > CD > DD$. Hi ha dos equilibris: Desertar de forma unilateral o cooperar de forma unilateral, és a dir, fer el contrari del que farà l'altre.

En aquest joc tampoc hi ha una estratègia dominant, ja que el comportament del contrari és determinant per escollir el comportament propi que proporcioni el millor resultat.

Els *N-person dilemmas*, es poden classificar en funció del cost i beneficis que tenen les decisions que pren cada individu. Distingim entre:

Social fence (Public goods Dilemma)

La persona s'enfronta a un cost immediat que genera un benefici compartit per tot el col·lectiu. D'aquesta forma l'individu té un incentiu per evitar el cost, però si tots els altres ho fan també el resultat a llarg plaç és pitjor que si tots s'haguessin repartit el

cost. La recerca dins d'aquesta categoria s'ha dut a terme a través de la metàfora de *The provision of public goods* (provisió de béns públics).

Un bé públic és un recurs del qual tothom es pot beneficiar, tot i no haver col·laborat en proveir-lo. Estan caracteritzats per ser no rivals, és a dir l'ús del bé per part d'una persona no disminueix l'ús que li pugui donar una altra. Existeix la temptació de no contribuir en la creació o manteniment del bé, són els anomenats *free-riders*. Si tots els individus actuessin d'aquesta forma el bé no existiria i el resultat seria el pitjor. Hi ha una altre situació que pot comportar aquest resultat i és la de sentir que s'està col·laborant en una causa perduda, on el que un fa no té repercussió en el resultat final.

Social trap (Tragedy of the commons)

L'individu està temptat per un benefici immediat el qual produeix un cost compartit per tot el col·lectiu. Si tots decideixen actuar d'aquesta forma, el resultat és el conegut *tragedy of the commons*, a través del qual s'ha conduït la recerca dins d'aquesta categoria.

Aquest dilema va ser descrit per Lloyd l'any 1932 i es va fer famós per Hardin l'any 1968. Hardin descriu a un grup de pastors, amb una vaca cada un, que tenen accés a una parcel·la comuna per alimentar-les. En aquesta parcel·la hi ha aliment suficient per a que cada una de les vaques es mantingui sana. Si algun dels pastors enlloc d'una vaca en porta dues, tindrà major benefici, però el cost resultant de l'acció, serà menys aliment per totes les vaques i per tant el benefici serà individual però el cost repartit entre tots. Si tots els pastors actuessin pensant en el benefici propi el resultat seria que no hi hauria aliment suficient per cap de les vaques. A diferència del *Public goods dilemma*, aquest sí que és rival, ja que el bé que consumeixes deixa d'estar disponible per l'altre.

El problema d'aquest tipus de dilemes, és que els possibles resultats del comportament d'un individu pot afectar la situació de l'altre, sense haver-hi un acord entre les dues parts.

Aquests dilemes socials es poden classificar en 3 grups en funció del grau d'egoisme amb el que l'individu actua i en funció de la possibilitat de canvi de l'estructura de la situació.

Solucions

Trobem solucions motivacionals, que són aquelles en les que els individus no són 100% egoistes i tenen en compte l'impacte que tindrà la seva decisió cap els altres. Les solucions estratègiques van enfocades a individus que prenen les decisions de forma egoista sense tenir en compte l'impacte que tindran en els demés. Cap de les dos categories contempen un canvi en l'estructura de la situació. Mentre que les solucions estructurals són aquelles que sí contempen canvis, i dins d'aquestes tenim dos subgrups. Les solucions estructurals dels individus que contempen l'impacte que aquesta tindrà en els demés i les que no.

- Les solucions més adequades pel primer tipus d'individus i enfocades en la problemàtica de l'equipatge de mà són la comunicació i el formar part d'un grup. Quan s'estableix comunicació entre els individus, la cooperació entre aquests creix. Per tant, tenir transparència amb les diferents polítiques d'equipatge de mà, així com amb els impactes de l'excés d'aquest en els vols (retards, procés d'embarcament lent...), podria ajudar a reduir-lo. Segons diversos estudis, els humans necessitem sentir que formem part d'un col·lectiu, ens agrada sentir-nos identificats amb altres que comparteixin les nostres idees. És per això que crear polítiques individualitzades, englobant a un col·lectiu de persones, fent sentir al passatger per exemple que forma part d'aquell grup i que la solució proposada és la que els passatgers d'aquell grup faran, pot motivar a l'individu a realitzar-la.
- En el segon tipus les millors solucions són les recíproques, aquelles en les que pel fet de realitzar una certa acció, n'obtindràs alguna cosa a canvi en un futur. Tot i així en el tema que estem explorant, les raons de més pes per portar maleta de mà són factors a curt termini, per tant les polítiques que es puguin crear per cada grup, hauran de tenir en compte aquest element. El fet de saber les conseqüències sobre l'acció que realitzaràs acompanyat de si existeix un vincle amb la persona o col·lectiu que en patirà els efectes, també influeix el comportament de la persona. Resulta important també per aquest perfil de persones, saber que les accions que s'estan portant a terme tenen un propòsit i

generen un benefici, així com saber que els demés també les estan portant a terme. Al igual que el primer tipus, el formar part d'un grup resulta motivador, però en aquest cas no té el mateix pes que en el primer. (Kollock, 1998)

SECCIÓ 3

METODOLOGIA DE RESOLUCIÓ I EXECUCIÓ

3.1. DESCRIPCIÓ DETALLADA DEL PROBLEMA

Una vegada feta i revisada la base teòrica i havent observat quines han sigut les diferents metodologies per tal de solucionar el problema plantejat fins el moment, i quins elements les han fet fallar, en aquesta part de treball, intentarem implementar la nostra metodologia pròpia. Aquesta és de caràcter universal, és a dir, qualsevol companyia aèria la podrà fer servir amb les seves dades, independentment dels diferents elements que la caracteritzen, tals com la localització, el tipus de vols que oferten, etc.

La nostra metodologia es basa en dos procediments. Un model de regressió logística i un *cluster analysis*.

Amb el model de regressió logística busquem quina és la probabilitat de que una persona facturi la maleta de mà a partir d'altres variables ja conegudes per la companyia. Per portar-ho a terme hem creat un model de regressió logística, ja que aquest estudia com una variable dependent o variable resposta està relacionada amb dos o més variables independents o predictores.

L'objectiu del *cluster analysis* és segmentar els clients de la companyia aèria en subgrups que comparteixin unes característiques similars, amb la finalitat de poder elaborar polítiques d'equipatge enfocades al tipus de client de la companyia, d'una forma més personalitzada.

3.2. INPUTS

Per desenvolupar la nostre metodologia hem elaborat una enquesta de 17 preguntes per la qual hem generat 100 respostes de forma aleatòria (adjuntades a l'annex). Amb la informació de les 17 preguntes hem pogut generar 22 variables tal i com es mostra a la Figura 16. Les 100 observacions les hem generat de forma aleatòria ja que com hem comentat, el que ens interessa buscar és una metodologia universal, no una solució per una companyia en concret. Tot i així les hem generat seguint una certa lògica per tal de poder fer les dades més realistes i poder exemplificar el resultat de la forma més aproximada a la realitat possible.

Dataset info		Variables types	
Number of variables	22	CAT	10
Number of observations	100	NUM	7
Missing cells	0 (0.0%)	BOOL	5
Duplicate rows	0 (0.0%)		
Total size in memory	17.3 KiB		
Average record size in memory	177.3 B		

Figura 16 – Generació de les dades

L'enquesta està classificada en 4 seccions:

Secció A (p1-p2): Preguntes de caràcter personal que proporcionen la següent informació: gènere i rang d'edat.

Secció B (p3-p14): Informació del vol que ens proporciona la següent informació: origen i destí; mes en el que es realitza el viatge; dia de la setmana; finalitat del viatge; duració del vol; duració del viatge; número de persones que participen en el viatge; si viatja amb nens de 0-6 anys; si viatja amb algú de mobilitat reduïda; número i tipo de maletes amb les que viatja de forma individual; si es factura o es porta a la cabina l'equipatge de mà; número de viatge en els últims 12 mesos.

Secció C (p15-p16): Preguntes de l'actitud del passatger, que proporcionen la següent informació: influència del nostre entorn sobre el nostre comportament i grau d'acord i desacord en diferents afirmacions utilitzant l'escala de *Likert*.

Secció D (p17): Pregunta sobre l'actitud del passatger a l'hora de planejar el viatge en un futur, que proporciona la següent informació: facilitat per optar per cada una de les opcions de transportar l'equipatge de mà mitjançant l'escala de *Likert*.

A continuació es pot observar l'enquesta:

SECCIÓ A – PREGUNTES DE CARÀCTER PERSONAL

1. Selecciona el vostre gènere:

- a. Home
- b. Dona

2. Amb quin rang d'edat pertany?

- a. 15 – 24
- b. 25 – 34
- c. 35 – 44
- d. 45 – 54
- e. 55 – 64
- f. + 65

SECCIÓ B – INFORMACIÓ DEL VOL

3. Quin és el teu origen i el teu destí?

- a. Origen
- b. Destí

4. A quin mes estàs realitzant el viatge?

- | | |
|-----------|-------------|
| a. Gener | g. Juliol |
| b. Febrer | h. Agost |
| c. Març | i. Setembre |
| d. Abril | j. Octubre |
| e. Maig | k. Novembre |
| f. Juny | l. Desembre |

5. A quin dia de la setmana estàs viatjant?

- a. Dilluns
- b. Dimarts
- c. Dimecres
- d. Dijous
- e. Divendres
- f. Dissabte
- g. Diumenge

6. Quina és la finalitat del viatge?

Oci		Negoci	
Vacances: Tarifa pagada per separat	1	Desplaçament cap a la feina	7
Vacances: Pack, tot inclòs	2	Assistir per negoci intern de l'empresa	8
Visitar amics i familiars	3	Conferència o fira	9
Event cultural / esportiu	4	Reunió amb clients o proveïdors	10
Anar a viure fora	5	Educació universitària o escolar	11
Altres (escriure-ho)	6	Altres (escriure-ho)	12

7. Quina és la duració del vol?

8. Quina és la duració del viatge?

9. Incloent-te a tu, quantes persones realitzareu el viatge?

(Si viatges sol, anar directament a la pregunta 10)

10. Viatges amb nens de 0 a 6 anys?

- a. Si (Quants?)
- b. No

11. Viatges amb algú de mobilitat reduïda?

- a. Si
- b. No

12. Individualment, amb quantes maletes viatges? (Contestar en funció de les mides, independentment de si es factura la maleta en qüestió o no)

- a. Maletes grans
- b. Maletes de mà (aquelles que compleixen amb les mides per ser transportades a la cabina de l'avió)

13. D'aquestes maletes de mà, quantes factures i quantes les portes amb tu a la cabina?

- a. Facturades
- b. Cabina

14. Quantes vegades has viatjat en els últims 12 mesos?

SECCIÓ C – ACTITUD ENVERS EL PROBLEMA

15. Marca amb quin grau estàs d'acord o amb desacord amb aquestes afirmacions:

Grau de les afirmacions	5	4	3	2	1
	Molt d'acord	D'acord	Neutre	En desacord	Molt en desacord
a) Si sapiguessis que portar maleta de mà ocasiona retards en la sortida del vol deixaries de portar-la?	5	4	3	2	1
b) Si obtinguessis beneficis per facturar l'equipatge de mà, el facturaries? (Per exemple prioritat per embarcar, un descompte pel següent vol...)	5	4	3	2	1
c) Si sapiguessis que el pes de la maleta de mà ocasiona impactes negatius (econòmics, mediambientals...) reduiries el pes?	5	4	3	2	1

16. En l'hipotètic cas de fer el viatge acompanyat, escolliries la mateixa política d'equipatge de mà que la persona/es que t'acompanya/en?

- a. Sempre
- b. Quasi sempre
- c. A vegades
- d. Rares vegades
- e. Mai

SECCIÓ D – ACTITUD A L'HORA DE PLANEJAR EL VIATGE

17. Com de fàcil seria per a tu optar per les següents opcions en relació a l'equipatge de mà?

Grau de les afirmacions	5	4	3	2	1
	Molt fàcil	Fàcil	Neutre	Difícil	Molt difícil
Cabina	5	4	3	2	1
Facturar al check-in	5	4	3	2	1
Facturar a la porta d'embarc	5	4	3	2	1
No portar equipatge de mà	5	4	3	2	1

MOLTES GRÀCIES PEL TEU TEMPS

3.3. IMPLEMENTACIÓ I EXPLICACIÓ DEL CODI

Una vegada explicat en què consisteixen les dos possibles solucions a implementar per tal de resoldre el problema així com esmentats els diferents inputs que necessitem per desenvolupar la solució, en aquesta secció explicarem de forma detallada quins són els passos que hem seguit per dissenyar cada mètode.

3.3.1. Model de regressió logística

Com ja hem esmentat anteriorment, hem utilitzat un model de regressió de logística. Les variables que utilitzem per portar-lo a terme són les diferents preguntes de l'enquesta, que podeu observar a l'annex.

La variable dependent és si el passatger factura la maleta de mà. Aquesta variable correspon a la pregunta 13 de l'enquesta. Les 21 variables restants corresponen a les variables independents.

Un cop executat el model amb les variables esmentades, ens trobem que les nostres dades contenen una proporció alta de multicolinealitat. Una de les condicions per construir un model de regressió logística és que les variables independents, com ja indica el nom, no depenguin les unes de les altres. Per tal d'observar les variables que estan altament relacionades entre elles fem la matriu de correlació mitjançant *Pearson*. Com es pot observar a la matriu tots els valors que s'aproximen als intervals $[-0.7, -1]$ o $[0.7, 1]$ indiquen que tenen un grau alt de correlació (Figura 17).

Anàlisi per a la Gestió Eficient de l'Equipatge de Cabina en l'Aviació Comercial

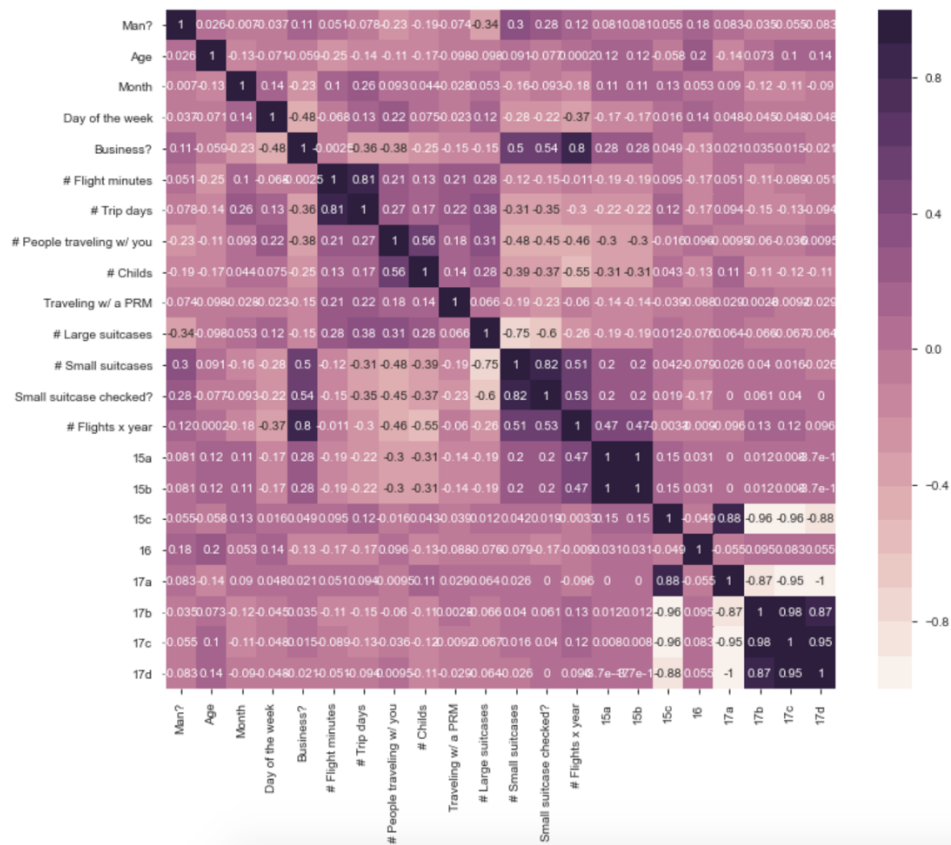


Figura 17 – Matriu de correlació (Pearson)

Una forma de mesurar la multicolinealitat és amb el factor d'inflació de la variància (VIF), que calcula quant incrementa la variància d'un coeficient de regressió estimat si les variables predictores estan correlades entre elles. Un VIF entre 5 i 10 indica una alta correlació que pot causar problemes en el nostre model. Si el VIF és superior a 10 pots assumir que els coeficients de regressió estan mal estimats a causa de la multicolinealitat.

Per calcular-ho fem servir la següent funció que es pot observar a la Figura 18.

Warning: Maximum number of iterations has been exceeded.
 Current function value: 0.167336
 Iterations: 35

Results: Logit

=====						
Model:		Logit		Pseudo R-squared:		0.759
Dependent Variable:		Small suitcase checked?		AIC:		51.4672
Date:		2020-06-09 17:20		BIC:		74.9138
No. Observations:		100		Log-Likelihood:		-16.734
Df Model:		8		LL-Null:		-69.315
Df Residuals:		91		LLR p-value:		3.7468e-19
Converged:		0.0000		Scale:		1.0000
No. Iterations:		35.0000		=====		
	Coef.	Std.Err.	z	P> z	[0.025	0.975]

Man?	1.3269	1.0725	1.2372	0.2160	-0.7752	3.4290
Month	0.1897	0.1256	1.5098	0.1311	-0.0566	0.4359
Day of the week	0.5845	0.3047	1.9181	0.0551	-0.0128	1.1817
Business?	28.6670	49053.6553	0.0006	0.9995	-96114.7307	96172.0647
# Flight minutes	-0.0023	0.0019	-1.2339	0.2172	-0.0059	0.0013
# Childs	-12.6586	40196.9828	-0.0003	0.9997	-78797.2972	78771.9800
Traveling w/ a PRM	-24.8625	229502.7890	-0.0001	0.9999	-449842.0632	449792.3382
# Large suitcases	-26.9404	49053.6553	-0.0005	0.9996	-96170.3380	96116.4573
16	-0.8531	0.3755	-2.2720	0.0231	-1.5891	-0.1172
=====						

Figura 20 – Resultats mitjançant VIF

Seguint amb el problema de la multicolinealitat, decidim reduir encara més el número de variables. Diversos mètodes per fer-ho possible són: *Filter Method*, *Forward Regression*, *Backward Elimination*, *Stepwise* i *Best-Subsets Regression*.

El primer mètode que provem és el *Filter Method*. En el que com es pot observar a la Figura 21, estem dient que ens mostri les variables en les que la seva *R-squared* sigui major a 0,5 per tal d'eliminar-les del model.

```
#Correlation with output variable
cor=df.corr()
cor_target = abs(cor["Small suitcase checked?"])
#Selecting highly correlated features
relevant_features = cor_target[cor_target>0.5]
relevant_features
```

Figura 21 – Codi Filter Method

Les variables majors de 0,5 són:

<i>Business?</i>	0,540548
<i>#Large suitcases</i>	0,604165
<i>#Small suitcases</i>	0,816497
<i>Small suitcases check?</i>	1,000000
<i>#Flights x year</i>	0,526460

A continuació, com es pot veure a la Figura 22, implementem el model amb les variables que hem obtingut, i tal i com s'observa a la Figura 23 obtenim una *R-squared* de 0,652 i uns *P-valors* dos dels quals són molt elevats ja que s'aproximen a 1, i un de 0,0073. Referint-nos a la *R-squared* podria ser un model acceptable però degut als dos *P-valors* tan elevats ens podria causar problemes.

```
#Implementing the model
logit_model=sm.Logit(df['Small suitcase checked?'],df.drop(['Small suitcase checked?'], axis=1))
result=logit_model.fit()
print(result.summary2())
```

Figura 22 – Codi implementació del model

```
Warning: Maximum number of iterations has been exceeded.
Current function value: 0.241518
Iterations: 35

Results: Logit
=====
Model:                Logit                Pseudo R-squared: 0.652
Dependent Variable:  Small suitcase checked?  AIC:                54.3036
Date:                2020-06-09 17:21             BIC:                62.1191
No. Observations:   100                    Log-Likelihood:    -24.152
Df Model:           2                        LL-Null:           -69.315
Df Residuals:      97                       LLR p-value:       2.4322e-20
Converged:          0.0000                    Scale:             1.0000
No. Iterations:    35.0000

-----
                Coef.   Std.Err.   z     P>|z|   [0.025   0.975]
-----
Business?           14.8811  240.8870  0.0618  0.9507  -457.2489  487.0110
# Large suitcases  -14.5135  240.8860 -0.0603  0.9520  -486.6414  457.6145
# Small suitcases   1.1148   0.4154  2.6840  0.0073   0.3007   1.9289
=====
```

Figura 23 – Resultat mitjançant Filter Method

Un altre mètode que provem és el de *Step Forward Regression*. Aquest mètode consisteix en que a cada iteració es pot afegir una única variable i les anteriors no es poden eliminar. Aquest procés continua fins que no es pot aplicar cap millora afegint una variable.

A partir de les funcions adjuntes a la Figura 24, aquesta ens retorna que les variables més importants i que hem de tenir en compte, acord amb el *Forward Selection Method* són *#childs* i *#small suitcases*.

```
#Build step forward feature selection
sfs1= sfs(logreg, k_features=2, forward= True, floating=False,scoring='r2',verbose=2,cv=5)
#Perform SFS
sfs1= sfs1.fit(X_train, y_train)
```

Figura 24 – Codi selecció de funcions mitjançant Step Forward

A continuació com es pot veure a la Figura 25 implementem el model amb les variables que hem obtingut. Com s'observa a la Figura 26 obtenim una *R-squared* acceptable, però el *P-valor* de *#childs* és molt elevat.

```
#Implementing the model
logit_model=sm.Logit(df['Small suitcase checked?'],df.drop(['Small suitcase checked?'], axis=1))
result=logit_model.fit()
print(result.summary2())
```

Figura 25 – Codi implementació model Step Forward

```
Warning: Maximum number of iterations has been exceeded.
Current function value: 0.432204
Iterations: 35
```

```
Results: Logit
=====
Model:                Logit                Pseudo R-squared:    0.376
Dependent Variable:   Small suitcase checked?  AIC:                 90.4407
Date:                 2020-06-09 17:22      BIC:                 95.6511
No. Observations:    100                    Log-Likelihood:      -43.220
Df Model:              1                        LL-Null:             -69.315
Df Residuals:         98                        LLR p-value:         5.0415e-13
Converged:            0.0000                          Scale:               1.0000
No. Iterations:       35.0000

-----
              Coef.      Std.Err.      z      P>|z|      [0.025      0.975]
-----
# Childs      -26.8523  225827.8648  -0.0001  0.9999  -442641.3341  442587.6295
# Small suitcases  1.7148    0.3621    4.7358  0.0000     1.0051     2.4245
=====
```

Figura 26 – Resultats mitjançant Step Forward

És per això que decidim provar el mètode de *Backward Elimination*. Aquest procés comença amb un model que inclou totes les variables predictores possibles. A cada iteració agafa la variable amb el *P-valor* > *alpha-to-remove*, tal i com es pot observar a la Figura 27, i s'elimina del model per poder calcular de nou els *P-valors*. Aquest és un procés repetitiu fins que aconseguim obtenir un model fiable on no hi podem implementar cap millora a través de l'eliminació de variables.

```

while (len(cols)>0):
    p= []
    x_1 = x[cols]
    x_1 = sm.add_constant(x_1)
    model = sm.OLS(y,x_1).fit()
    p = pd.Series(model.pvalues.values[1:],index = cols)
    pmax = max(p)
    feature_with_p_max = p.idxmax()
    if(pmax>0.05):
        cols.remove(feature_with_p_max)
    else:
        break
selected_features_BE = cols
print(selected_features_BE)

```

Figura 27 – Codi implementació mètode Backward Elimination

A partir d'aquest mètode obtenim que les millors variables per incloure en el nostre model són *age*, *#trip days* i *#small Suitcases*.

A continuació, com es pot veure a la Figura 28 implementem el model amb les variables que hem obtingut, i tal i com s'observa a la Figura 29 obtenim una *R-squared* i uns *P-valors* acceptables.

```

#Implementing the model
logit_model=sm.Logit(df['Small suitcase checked?'],df.drop(['Small suitcase checked?'], axis=1))
result=logit_model.fit()
print(result.summary2())

```

Figura 28 – Codi implementació model Backward Elimination

Optimization terminated successfully.

Current function value: 0.207277

Iterations 9

Results: Logit

=====						
Model:		Logit	Pseudo R-squared: 0.701			
Dependent Variable:		Small suitcase checked?	AIC:	47.4554		
Date:		2020-06-09 20:44	BIC:	55.2709		
No. Observations:		100	Log-Likelihood:	-20.728		
Df Model:		2	LL-Null:	-69.315		
Df Residuals:		97	LLR p-value:	7.9237e-22		
Converged:		1.0000	Scale:	1.0000		
No. Iterations:		9.0000				

	Coef.	Std.Err.	z	P> z	[0.025	0.975]

Age	-1.0723	0.3455	-3.1034	0.0019	-1.7496	-0.3951
# Trip days	-0.4972	0.1819	-2.7337	0.0063	-0.8537	-0.1407
# Small suitcases	8.4687	2.1395	3.9583	0.0001	4.2754	12.6620
=====						

Figura 29 – Resultats mitjançant Backward Elimination

3.3.2. Cluster analysis

Per realitzar el *cluster analysis* hem utilitzat l'algoritme *K-means* i l'algoritme d'*Agglomerative Clustering*, per decidir quina és la millor forma d'agrupar les dades.

Preparació de les dades:

Per classificar els passatgers en grups segons el seu comportament hem utilitzat les preguntes de la secció C i D (P15 - P17) que corresponen amb les preguntes de l'escala de *Likert* (de l'1 al 5). En total obtenim 8 variables que ocupen un espai de 8 dimensions i utilitzem l'anàlisi de components principals (PCA) per tal de poder transformar aquestes variables en un espai de dues dimensions i així visualitzar el resultat del *cluster analysis*. Al passar la informació de les 8 variables d'un espai de 8 dimensions a un espai de dos dimensions, és impossible que alguna informació no es perdi. En el nostre cas aconseguim preservar el 80% de la variança.

Exploració de les dades:

Mitjançant el codi observat a la Figura 30 trobem que la mitjana de les variables que hem seleccionat és la que es pot observar a la Figura 31, sent la mitjana total de 2,9274999999999998.

```
#Data Exploration

#Q15-Q17 are all Likert-type, meaning that the values are taken from {1,2,3,4,5}
X_questions = df.iloc[:,14:22]
question_means = X_questions.mean(axis = 0)
grand_mean = question_means.mean()
print(question_means)
print(grand_mean)
```

Figura 30 – Codi mitjana de les variables

15a	3.50
15b	3.00
15c	3.18
16	2.10
17a	4.34
17b	2.99
17c	2.65
17d	1.66

Figura 31 – Mitjana de les variables

3.3.2.1. K-means

L'algorisme *K-means* agrupa les dades d'un conjunt de n observacions en k grups (determinats prèviament per la persona que realitza el clúster) on cada observació forma part del grup més proper a la mitjana. En el nostre cas, no tenim el número de grups en els que es pot segmentar la nostra mostra. Per saber-ho, utilitzem el mètode d'*Elbow* per seleccionar el número òptim de clústers (k), utilitzant el codi que es mostra a la Figura 32.

```
#Training Models : K-Means
distortions = []
K_to_try = range(1, 6)

for i in K_to_try:
    model = KMeans(
        n_clusters=i,
        init='k-means++',
        # n_init=10,
        # max_iter=300,
        # n_jobs=-1,
        random_state=1)
    model.fit(X_pca)
    distortions.append(model.inertia_)

plt.plot(K_to_try, distortions, marker='o')
plt.xlabel('Number of Clusters (k)')
plt.ylabel('Distortion')
plt.show()
```

Figura 32 – Codi mètode Elbow

Com es pot observar a la Figura 33, l'*Elbow* el localitzem a $k=3$.

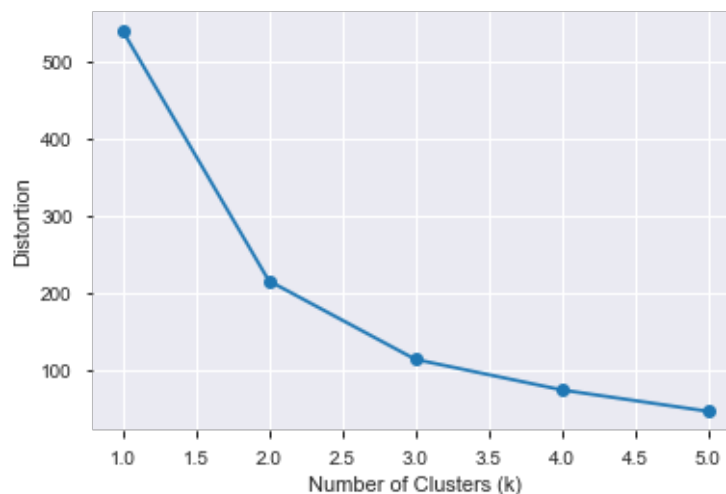


Figura 33 – Gràfic resultat mitjançant Elbow

És per això que entrenem *K-means* amb 3 clústers tal i com es mostra a la Figura 34 i visualitzem els resultats en un gràfic de dispersió (Figura 35). Obtenim 43 persones que formen part del grup 1, 34 persones del grup 2 i 23 persones del grup 3.

```
# use the best K from elbow method
model = KMeans(
    n_clusters=3,
    init='k-means++',
    # n_init=10,
    # max_iter=300,
    # n_jobs=-1,
    random_state=1)

model = model.fit(X_pca)

y = model.predict(X_pca)
```

Figura 34 – K-means mitjançant k=3

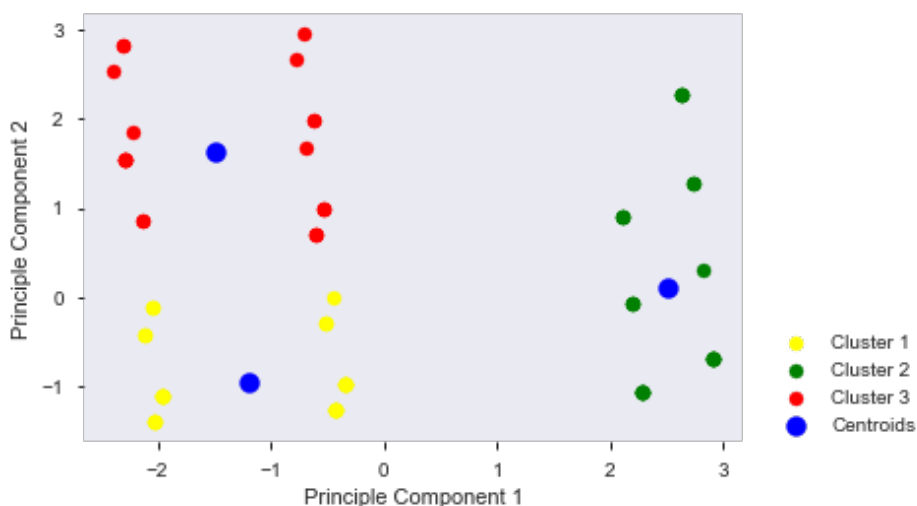


Figura 35 – Gràfic de dispersió K-means

Utilitzem el *Silhouette analysis* per estudiar la separació entre els diferents clústers. D'aquesta manera es pot mesurar la distància en les que els punts d'un clúster estan respecte els d'un clúster veí. Aquesta mesura té un rang que va de [-1,1] on els valors que es troben a prop de +1 indiquen que la mostra està lluny del clúster veí. Si el valor és 0, indica que la mostra està en el límit entre un clúster i l'altre. I un resultat negatiu indica que la mostra pot haver estat assignada en el clúster incorrecte. En el nostre cas, com es pot observar a la Figura 36 , el *Silhouette analysis* dona 0,6.

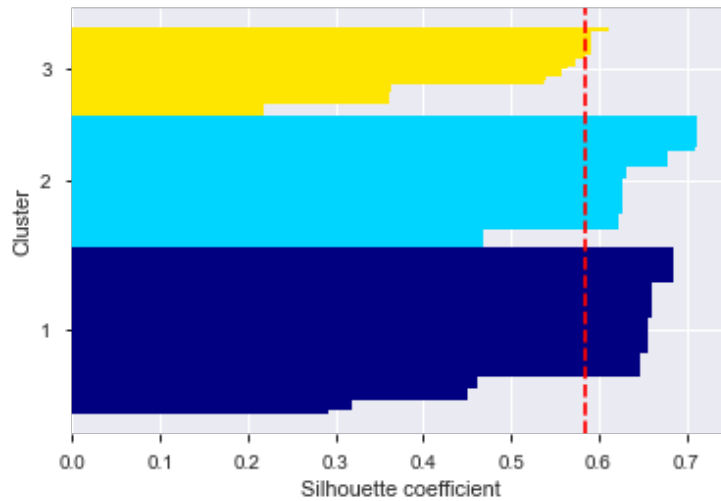


Figura 36 – Silhouette Analysis

La resposta mitjana de cada clúster per tal de poder observar la diferència entre ells és la següent (Figura 37).

```
Mean cluster 1 : 2.5839793281653733,STD :0.18638714151762784  
Mean cluster 2 : 2.5555555555555555,STD :0.2868876552746234  
Mean cluster 3 : 3.0917874396135265,STD :0.19434246651396977
```

Figura 37 – Resposta mitjana dels diferents clústers (K-means)

3.3.2.2. *Agglomerative clustering*

Aquest algoritme inicialment considera cada resposta de forma individual i a cada iteració que fa ajunta els dos clústers més similars i els combina tot creant-ne un de nou. Aquest procés es va repetint fins que tots els punts estan units en un únic clúster donant lloc a un dendrograma. Existeixen diversos mètodes de *cluster agglomeration*.

3.3.2.2.1. *Ward*

El primer que utilitzem és el mètode de *Ward*. Aquest minimitza la variança total dins del clúster. A cada iteració la parella de clústers amb una distància entre clústers menors és combinada. A la Figura 38 es pot observar pel mètode de *Ward* el millor número de $k=2$.

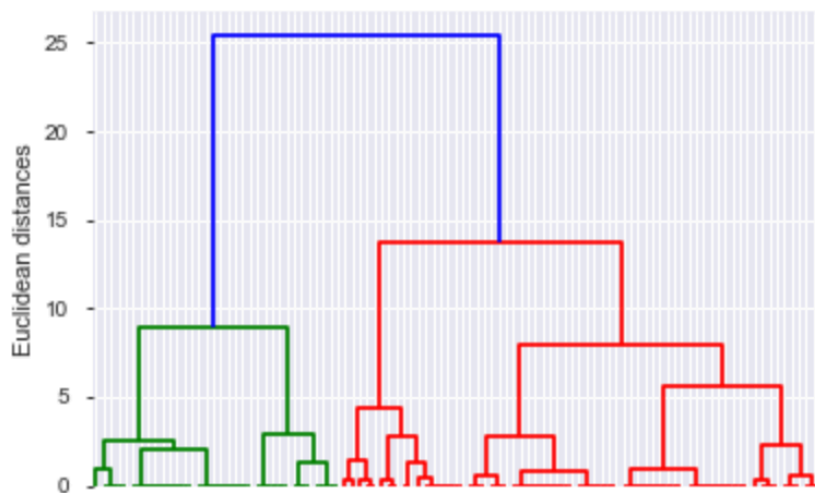


Figura 38 – Mètode de Ward

A la Figura 39 següent es pot observar el resultat en una gràfica de dispersió. En el primer clúster tenim a 66 persones i en el segon 34.

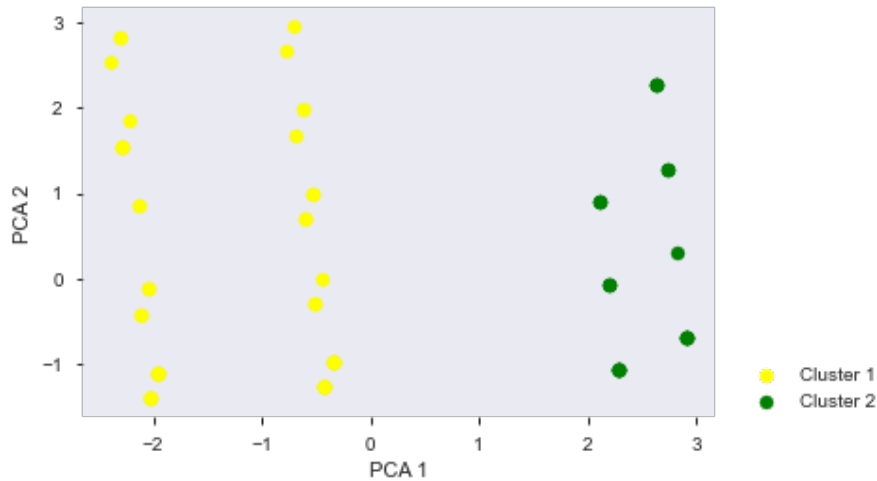


Figura 39 – Gràfica de dispersió de Ward

La resposta mitjana de cada clúster per tal de poder observar la diferència entre ells és la següent (Figura 40).

```
Mean of cluster 1 : 2.6835016835016825,STD :0.23245044544755156
Mean of cluster 2 : 2.5555555555555555,STD :0.2868876552746234
```

Figura 40 - Resposta mitjana dels diferents clústers (Ward)

3.3.2.2.2. Complete

En aquest mètode es mesura la distància màxima entre parelles de punts de cada clúster. Agrupa conjunts completament veïns i busca grups compactes. Com podem observar en el dendrograma (Figura 41) pel mètode *Complete* el millor número de $k = 3$.

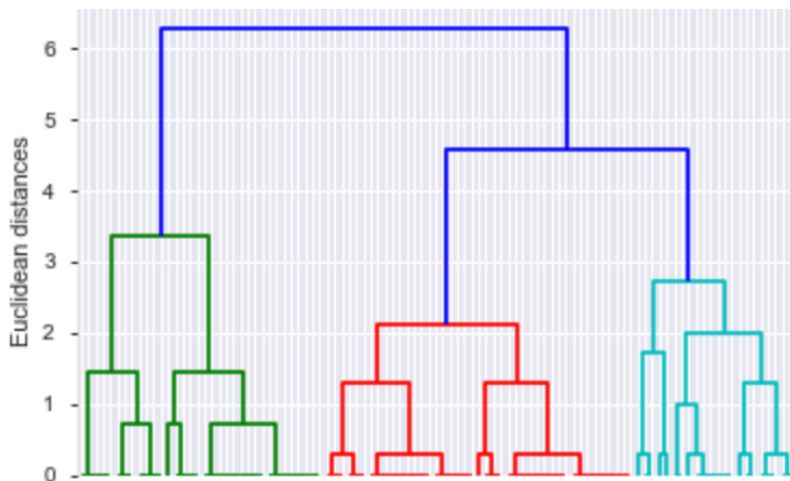


Figura 41 – Mètode Complete

A la Figura 42 següent es pot observar el resultat en una gràfica de dispersió. En el primer clúster tenim a 43 persones, en el segon 34 i el tercer 23.

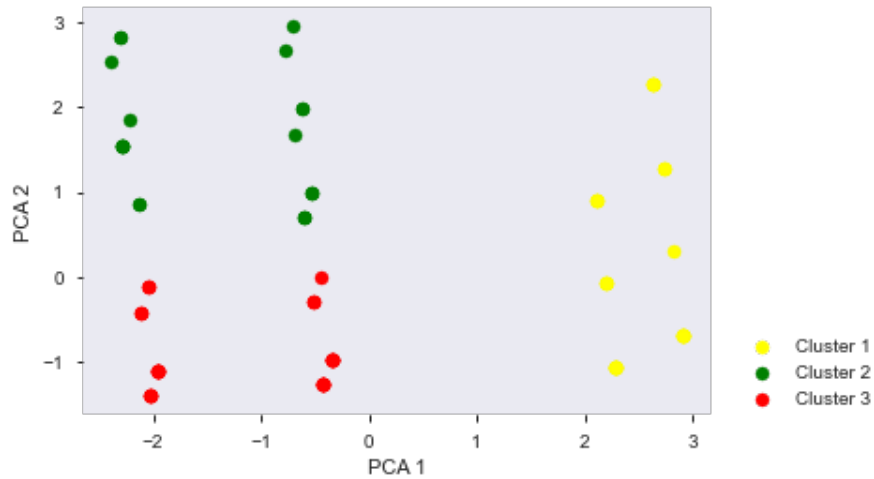


Figura 42 – Gràfica de dispersió de Complete

La resposta mitjana de cada clúster per tal de poder observar la diferència entre ells és la següent (Figura 43).

```
Mean cluster 1 : 2.4444444444444438,STD :0.2868876552746234  
Mean cluster 2 : 2.980676328502415,STD :0.19434246651396972  
Mean cluster 3 : 2.806201550387597,STD :0.18638714151762806
```

Figura 43 - Resposta mitjana dels diferents clústers (Complete)

SECCIÓ 4

ANÀLISI DE RESULTATS

4.1. INTRODUCCIÓ

Després d'executar el model de regressió logística i el *cluster analysis*, en aquesta secció analitzarem i extraurem la informació més significativa dels resultats.

4.2. ANÀLISI DELS RESULTATS DEL MODEL DE REGRESSIÓ LOGÍSTICA

El model que ens ha donat millors resultats és en el que apliquem el Backward Elimination method. Recordem que les variables d'aquest model són *age*, *#trip days* i *#small suitcases*, on els *P-valors* i la *R-squared* ens proporcionen el model més òptim de tots.

Per comprovar el correcte funcionament del model, aquest s'ha d'entrenar i testar amb les diferents dades de la mostra, per veure si obtenim uns bons resultats. En aquest cas l'hem entrenat amb un 80% de les dades i l'hem posat a prova amb el 20% restant. Mitjançant això, com es pot observar a la Figura 44, obtenim que l'exactitud (*accuracy*) del model és del 90%, és a dir, el model encerta 9 de cada 10 casos. Per un altre banda, tenim una precisió del 81%, per tant significa que de totes les vegades que el model prediu que un passatger facturarà maleta, un 81% de les vegades ho farà. Finalment, si observem l'exhaustivitat (*recall*) veiem que aquesta és del 100%, és a dir, que el nostre model és capaç d'identificar tots els passatgers que facturarien l'equipatge de mà.

```

Accuracy of logistic regression classifier on test set: 0.90
      precision    recall  f1-score   support

0         1.00      0.82      0.90         11
1         0.82      1.00      0.90          9

 accuracy
macro avg      0.91      0.91      0.90         20
weighted avg   0.92      0.90      0.90         20

Accuracy: 0.9
Precision: 0.8181818181818182
Recall: 1.0
    
```

Figura 44 – Precisió de la regressió logística

Amb el *Backward Elimination* hem aconseguit els millors resultats. Per acabar d'avaluar el model fem una *Confusion Matrix* tal i com es pot observar a la Figura 45. Tal i com es pot observar, obtenim 9+9 prediccions correctes i 0+2 prediccions incorrectes.

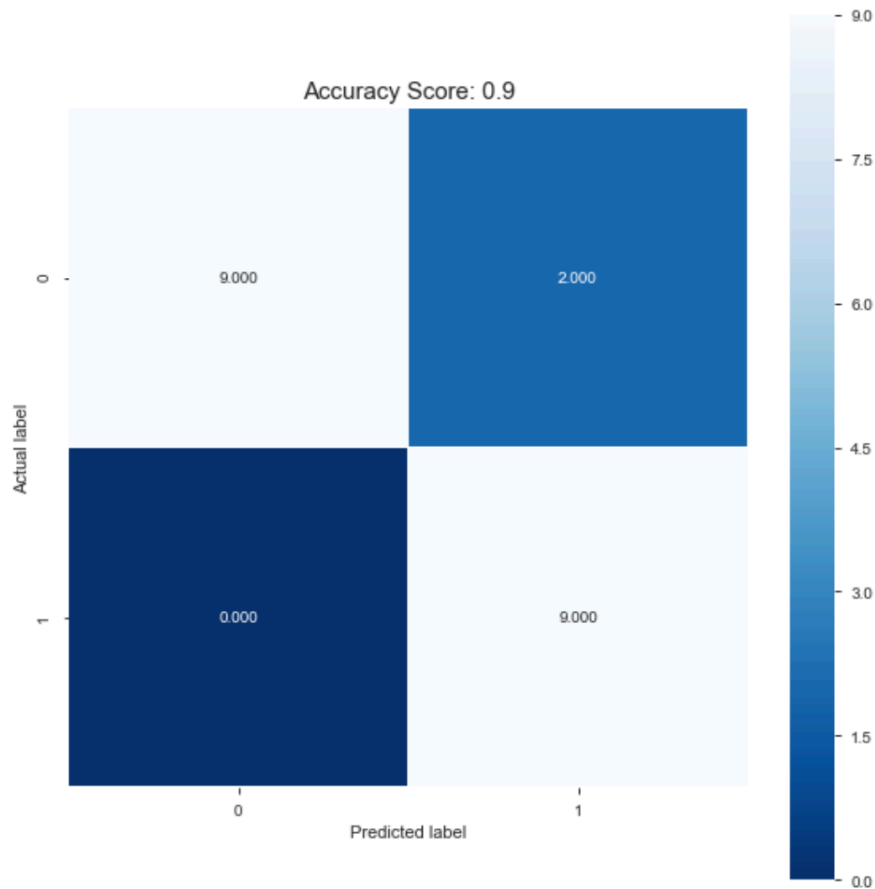


Figura 45 – Confusion Matrix

Una altre eina que s'utilitza és la corba ROC (*Receiver Operating Characteristic*). Com es pot veure a la Figura 46, la línia discontinua representa la corba de ROC d'un classificador purament aleatori. Un bon classificador es manté tan lluny com sigui possible d'aquesta línia, és a dir, a la cantonada superior esquerra tal i com es pot observar.

El *AUC score* del nostre cas és 0.91. Un *AUC score* de 1 és un perfecte classificador i menor a 0,5 significaria que no és un bon classificador. Arribar a 1 és quasi impossible, és per això que 0.91 és un molt bon *AUC score*.

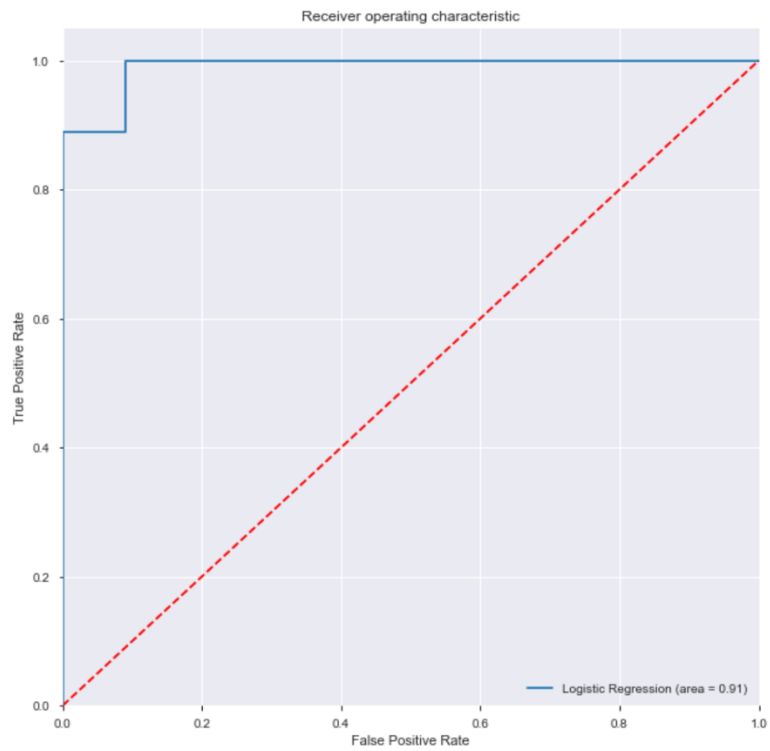


Figura 46 - Corba ROC (Receiver Operating Characteristic)

4.3. ANÀLISI DELS RESULTATS DEL CLUSTER ANALYSIS

Un cop observats el resultat dels 3 clústers, decidim que el millor algoritme correspon al *Agglomerative Cluster* amb el mètode de *Complete Linkage*. Recordem que les mitjanes de cada *clúster* són:

Clúster 1: 2.4444444444444438

Clúster 2: 2.980676328502415

Clúster 3: 2.806201550387597

Hem seleccionat aquest perquè és el que ens proporciona les millors mitjanes per cada clúster. Amb els altres mètodes les mitjanes de cada clúster són molt semblants. Això ens pot ocasionar dificultats a l'hora de classificar els diferents tipus de passatgers, ja que ens podem trobar amb casos on no quedi clar a quin grup pertany el passatger degut a la diferència tan petita entre les diferents mitjanes dels altres clústers.

Ens trobem amb tres tipus de passatgers. Els passatgers que s'aproximen més a la mitjana del clúster 1, són aquells que la seva mentalitat és més tancada (*closemind*) i per tant seran menys influenciables. En conseqüència els passatgers en els que les seves respostes s'aproximen més al valor de la mitjana del clúster 2, que correspon amb el valor més alt, són aquells passatgers que la seva mentalitat tendeix a estar més oberta (*openmind*) davant dels canvis i són més influenciables. Finalment els passatgers que es troben més pròxims a la mitjana del clúster 3, són aquells més neutrals (*neutrals*) i que per tant amb les tècniques adequades podem aconseguir que es decantin cap a on a la companyia li interessen.

Podem observar que un 43% dels passatgers són del clúster 1, un 34% del clúster 2 i 23% del clúster 3.

SECCIÓ 5
CONCLUSIONS I
TREBALL FUTUR

5.1. CONCLUSIONS

Durant aquest projecte hem dut a terme un estudi detallat del problema sobre la gestió eficient de l'equipatge de mà en l'aviació comercial. Hem elaborat, dissenyat, implementat i validat una metodologia de resolució innovadora basada en un model de regressió logística i un *cluster analysis*.

Per assolir l'objectiu global del projecte, primer de tot hem fet una descripció detallada del problema, seguida d'una cerca exhaustiva de les diferents aproximacions realitzades sobre el problema fins el dia d'avui.

A continuació, hem realitzat el marc teòric sobre el que hem basat la nostra metodologia.

Posteriorment, s'ha descrit quina ha sigut la metodologia a seguir. En aquesta metodologia, s'han detallat quines dades són necessàries per dur-la a terme, així com els processos i el codi utilitzat per desenvolupar cada model.

Per comprovar el correcte funcionament de la metodologia, s'han analitzat els models obtinguts amb una sèrie de tests a partir de les dades generades a través de les enquestes.

Finalment, una vegada validats i analitzats els resultats, s'han pogut extreure les següents conclusions.

Pel que fa referència al model de regressió logística, podem afirmar que el nostre model encerta 9 de 10 vegades si el passatger facturarà l'equipatge de mà o el portarà a la cabina de l'avió a partir de les variables dependents: *age*, *#trip days* i *#small suitcases*. Partim de la premissa que les companyies aèries tenen accés a aquesta informació sobre el passatger. En el cas de la variable *age*, qualsevol companyia aèria quan un passatger realitza una reserva, està obligat a donar la seva data de naixement. En canvi, la variable *#trip days*, només estarà disponible per la companyia quan el passatger realitzi tant

l'anada com la tornada amb ells i pressuposant que la tornada sigui el final del viatge. Per una altra banda, les dades de la variable *#small suitcases* no es poden extreure directament del procés clàssic de reserva, és per això que s'hauria de buscar la manera de poder extreure aquesta informació amb la màxima antelació possible. Tal cosa la veiem factible, ja que com hem esmentat en el punt 2.4.2., un 80% dels passatgers decideix si facturar maletes abans o durant el procés de compra, com a conseqüència de decidir el tipus d'equipatge que portarà. És en aquest procés de compra que es podria introduir la pregunta per saber la variable *#small suitcases*.

Amb els resultats del model, les companyies aèries podran determinar per a cada vol, com ja hem esmentat, amb una exactitud del 90%, quants passatgers facturaran l'equipatge de mà i quants el portaran a la cabina de l'avió. Amb aquesta informació les aerolínies podran prendre mesures per pal·liar els efectes negatius derivats de l'equipatge de mà en els vols on la proporció d'aquest a la cabina sigui més alta.

Per una altra banda, mitjançant el desenvolupament del *cluster analysis*, hem pogut classificar els passatgers de la mostra en tres grups: *openmind*, *neutral* i *closemind*.

A partir de l'anàlisi que hem efectuat en el punt 2.4.3. en el que parlem del *Social Dilemma Theory*, hem determinat quines són les accions estratègiques a prendre per cada clúster, per aconseguir que els passatgers facturin o no portin equipatge de mà.

Closemind - Clúster 1:

Aquest tipus de passatgers es caracteritzen per ser egoistes, poc oberts al canvi i tenir poc en compte l'impacte de les seves accions en la resta. Les millors solucions a aplicar per aquest tipus de passatgers són:

- Polítiques recíproques a través de les quals el passatger n'obté algun benefici, preferiblement immediat o a curt termini.
- Mostrar al passatger l'impacte positiu i la importància que tindrà el fet que ell no porti o facturi l'equipatge de mà, fent-li veure que la seva participació és important per la companyia.
- Polítiques en les que el passatger senti que forma part d'un grup.

Exemples

A canvi de facturar l'equipatge de mà o no portar-ne:

- Oferir embarcament prioritari.
- Oferir la facturació en un mostrador prioritari.
- Oferir un *snack* o refresc durant el vol.
- En el procés de reserva, informar al passatger de l'estalvi en terme de temps que tindrà en el procés d'embarcament i en el control de seguretat, l'acció de no portar o facturar l'equipatge de mà.

Openmind - Clúster 2:

Quan parlem de passatgers *openmind*, ens referim a passatgers que no són 100% egoistes, tenen en compte l'impacte que tindrà la seva decisió en la resta i estan oberts a canviar el seu comportament. Les millors accions que es poden implementar per aquests passatgers són:

- Millorar la percepció que tenen sobre la facturació de l'equipatge de mà promovent les alternatives de no portar-lo a cabina.
- Remarcar els aspectes positius de facturar o no portar equipatge de mà i mostrar com aquesta decisió genera un impacte positiu.
- Per aquest grup de passatgers, també es poden implementar polítiques en les quals es sentin que formen part d'un grup.

Exemples

- Facturar l'equipatge de mà el moment abans d'entrar a la cabina, assegurant al passatger que de la porta de l'avió l'equipatge va directament a bodega.
- Enviar una nota informativa després del vol agraint i esmentant l'impacte positiu que s'ha generat per no portar o facturar l'equipatge de mà.
- Esmentar una política on el nom d'aquesta faci referència a un grup en concret (negoci, oci, famílies, jubilats, joves, etc).

Neutral - Clúster 3:

En aquest cas, aquests tipus de passatgers són aquells que no tenen una idea predefinida, ja sigui perquè no han viatjat mai, perquè no tenen la suficient informació

de les diferents opcions o perquè cap de les opcions els acaba de convèncer. És un grup que s'ha de tenir molt en compte, ja que el fet de proposar una política o una altra, serà determinant en l'elecció final del passatger. És gent indecisa, que amb les estratègies de persuasió adequades es decantarà cap a on a la companyia li interessi, que en aquest cas seria que la persona no portes o facturés l'equipatge de mà.

Per aquest tipus de passatgers lo ideal seria fer un mix de polítiques del clúster 1 i 2, a més a més de donar informació molt detallada sobre els beneficis de cada política.

El model de regressió logística s'haurà d'actualitzar cada vegada que la companyia introdueixi algun canvi en les seves polítiques que afectin al comportament dels passatgers, ja que si les polítiques introduïdes generen aquests canvis, el model antic quedarà desfasat i les prediccions no seran acurades. De la mateixa manera es pot fer servir per mirar si el comportament dels passatgers del model actualitzat ha variat en comparació al model anterior, abans d'aplicar els canvis en les polítiques i així poder valorar el funcionament de les polítiques i si aquestes estan generant els canvis esperats per la companyia aèria en la presa de decisions dels passatgers.

Si les companyies segueixen aquestes directrius, podran elaborar unes bones polítiques per aconseguir augmentar la proporció de passatgers que facturaran o no portaran l'equipatge de mà. Tal i com es mostra a l'estudi del punt 2.3., si s'assoleix aquest objectiu el procés d'embarcament s'efectuarà més ràpidament, disminuint per tant els retards que es deriven del paper que té l'equipatge de mà en l'embarcament i, en conseqüència, les pèrdues econòmiques: entrar a l'avió i situar la maleta en el compartiment superior o el procés de retirada de les maletes que no tenen espai a cabina i han de ser transportades a bodega.

5.2. TREBALL FUTUR

En el nostre projecte hem presentat una metodologia per solucionar la problemàtica de l'equipatge de mà en els vols comercials. Gràcies a aquesta metodologia hem aconseguit que qualsevol companyia aèria amb les seves dades pugui executar el nostre model i seguir les directrius per elaborar les polítiques que millor s'adaptin als seus tipus de passatgers.

Com tot treball, en els resultats d'aquesta primera versió li podríem aplicar millores per desenvolupar més a fons el nostre estudi. Aquestes millores les anomenarem a continuació.

1. Aplicar la metodologia a un cas real per comprovar el correcte funcionament d'aquesta a la realitat. Per fer-ho lo ideal seria que una companyia aèria ens donés accés a la seva base de dades.
2. El nostre model dona per suposat que la companyia té accés a la informació de les tres variables independents del model de regressió logística, quan en un cas real, no és així. Una línia futura de treball seria buscar la forma d'incloure l'extracció de la informació de les variables en el procés de reserva, sense que el passatger sentís que s'està violant la seva intimitat.
3. Buscar polítiques concretes de com procedir amb les maletes que arriben a la porta d'embarcament i finalment no tenen espai a cabina.
4. A partir del model de regressió logística a través del qual sabem la quantitat de passatgers que portaran la maleta a la cabina, analitzar i decidir mitjançant les diferents polítiques esmentades en el apartat 2.3, quina serà la millor política a implementar segons la quantitat d'equipatge de mà.

5. Considerar realitzar un nou *cluster analysis* amb més variables per veure com varien els resultats d'un mètode a un altre i si existeix una forma de classificar els passatgers més completa.

6. Fer un estudi sobre a partir de quina quantitat de passatgers que no porten l'equipatge de mà a la cabina o simplement no porten equipatge de mà, gràcies a les polítiques implementades, milloraria de forma significativa el temps d'embarcament, per poder quantificar l'impacte amb valors econòmics i de temps.

SECCIÓ 6
REFERÈNCIES
BIBLIOGRÀFIQUES

- Abrahamse, W., Steg, L., & Vlek, C. (2009). Factors influencing car use for commuting and the intention to reduce it: A question of self-interest or morality? *Transportation Research Part F*.
- Air Europa. (sense data). www.aireuropa.com. Recollit de Air Europa: <https://www.aireuropa.com/es/vuelos/equipaje>
- Air France. (sense data). www.airfrance.es. Recollit de Air France: <https://www.airfrance.es/ES/es/common/voyage-en-avion/preparation-voyage/bagages-avion.htm>
- Ajzen. (1991). The Theory of planned Behaviour. *Organizational Behaviour and human Process*.
- Anable, J., & Gatersleben, B. (2005). All work and no play? The role of instrumental and affective factors in work and leisure journeys by different travel modes. *Transportation Research Part A*.
- Armando, S., & Katalina, S.-S. (2019). Application of Digital Nudging in Customer Journeys - A Systematic Literature Review. *Twenty fifth Americas Conference on Information Systems* .
- Association, A. T. (2009). ATA Quarterly Cost Index: U.S. Passenger Airlines. www.airlines.org.
- Avram, B. (2017). Ancillaries in the Aviation Industry. *Expert Journal of Marketing*.
- Baeriswyl, M., Staake, T., & Loock, C. (2011). The effects of user identity and sanctions in online communities on real-world behavior. *Proceedings of the International Conference on Information Systems*.
- Bailes, L., & Hoy, W. (2014). Designing school contexts for success: paternalism or libertarianism? *International Journal of Educational Management*.
- Ball, M., Barnhart, C., Dresner, M., Hansen, M., Neels, K., Odoni, A., . . . Zou, B. (2010). Total delay Impact Study. *Report prepared for the Federal Aviation Administration*.
- Bamberg, S., Fujii, S., Friman, M., & Gärling, T. (2011). Behaviour theory and soft transport policy options. *Transport Policy*.
- Bijker, j. (2016). Hand baggage. A passenger-centric approach to decrease the amount of hand baggage. *Faculty of Industrial Design Engineering*.
- Booth. (2016). How Mobile is empowering airlines to drive ancillary revenues. *Travelport Digital Magazine*.
- British Airways. (sense data). www.britishairways.com. Recollit de British Airways: <https://www.britishairways.com/es-es/information/baggage-essentials>
- Budd, T. (2013). An examination of passenger surface access travel behaviour. *Loughborough University*.
- Choo, W., & Dresner, M. E. (2018). The impact of a baggage fee waiver on airline choice: Evidence from the Washington-Baltimore region. *Elsevier*.

- Dennis, N. (2007). End of the free lunch? The responses of traditional European airlines to the low-cost carrier threat. *University of Westminster*.
- DeVries, P. D. (2009). Airline passenger information systems and process improvements. *University of Houston-Downtown*.
- Dings, J., Wit, R., Leurs, B., Davidson, M., & Franssen, W. (2003). External Cost of Aviation. *Federal Environmental Agency*.
- EASYJET. (sense data). *Aplicación para móvil de easyJet*. Recollit de <https://www.easyjet.com/es/mobile-app>
- Easyjet. (sense data). *www.easyjet.com*. Recollit de Easyjet: <http://www.easyjet.com/es/ayuda/equipaje/equipaje-de-mano-y-equipaje-facturado>
- Eurocontrol. (2008). *Delays to Air Transport in Europe*. Recollit de <https://extranet.eurocontrol.int>.
- Eurocontrol. (2012). *Delays to Air Transport in Europe*. Recollit de <https://www.eurocontrol.int/publication/delays-air-transport-europe-july-2012>
- Eurocontrol. (2019). *Industry Monitor*. Recollit de <https://www.eurocontrol.int/publication/industry-monitor-june-2019>
- European Environmental Agency. (2005). EMEP/CORINAIR Emission Inventory Guidebook. Group 8: Other Mobile Sources and Machinery. *Air Traffic European Environmental Agency, EEA*.
- Flenoury, A. (2015). *Early Valet Service Debut at US Hubs*. Recollit de DELTA: <https://news.delta.com/early-valet-service-debuts-us-hubs>
- Gärling, T., & Axhausen, K. (2003). Introduction: Habitual travel choice. *Transportation*.
- Givoni, M., & Rietveld, P. (2009). The environmental implications of airlines' choice of aircraft size. *Journal of Air Transport Management*.
- Goldsmith, C., & Cyboran, S. (2013). Reducing suboptimal employee decisions can build the business case for employee benefits. *Benefits Q*.
- Graml, T., Looock, C., Baeriswyl, M., & Staake, T. (2011). Improving residential energy consumption at large using persuasive systems. *Proceedings of the European Conference on Information Systems*.
- Guo, Y. (2005). Decision Support System for Airline Crew Recovery. *University of Paderborn*.
- IATA. (2015). *Airlines to Address Carry-on Bag Dilemma*. Recollit de <https://www.iata.org/en/pressroom/pr/2015-06-09-02/>
- IATA. (2015). *IATA Clarifies Cabin Ok Initiative*. Recollit de <https://www.iata.org/en/pressroom/pr/2015-06-12-01/>
- IATA. (2020). *Passenger Baggage Information*. Recollit de <https://www.iata.org/en/programs/ops-infra/baggage/check-bag/>
- IATA. (sense data). *Vision and Mission*. Recollit de <https://www.iata.org/en/about/mission/>

- Iberia. (sense data). www.iberia.com. Recollit de Iberia:
<https://www.iberia.com/es/equipatges/de-ma/>
- Johnson, E. J., Shu, S. B., Dellaert, B. G., Fox, C., Goldstein, D. G., Häubl, G., . . . Schkade, D. (2012). Beyond nudges: Tools of a choice architecture. *Marketing Letters*.
- Kollock, P. (1998). Social Dilemmas: The anatomy of cooperation. *Department of Sociology, University of California at Los Angeles*.
- Kolukisa, A. (2011). Evaluating Aircraft Turnaround Process in the Framework of Airport Design and Airline Behaviour. *Universidade do Porto*.
- Lemon, K., & Verhoef, P. (2016). Understanding Customer Experience Throughout the Customer Journey. *Journal of Marketing* , 69 - 96.
- Lim, S. H., & Hong, Y. (2013). Fuel Hedging and Airline operating costs. *Journal of Air Transport Management* .
- Loock, C., Landwehr, J., Staake, T., Fleisch, E., & Pentland, A. (2012). The Influence of Reference Frame and Population Density on the Effectiveness of Social Normative Feedback on Electricity Consumption. *Proceedings of the International Conference on Information Systems*.
- Loock, C., Staake, T., & Landwehr. (2011). Green IS Design and Energy Conservation: An Empirical Investigation of Social Normative Feedback. *Proceedings of the International Conference on Information Systems*.
- Loock, C., Staake, T., & Thiesse, F. (2013). Motivating Energy-Efficient Behavior with Green IS: An Investigation of Goal Setting and the Role of Defaults. *MIS Quarterly*.
- Lufthansa. (sense data). www.lufthansa.com. Recollit de Lufthansa:
<https://www.lufthansa.com/ve/es/equipaje-resumen>
- Mas, M. S. (2012). Aircraft Boarding Strategies: a study of the airline boarding problem. *Universitat Autònoma de Barcelona*.
- Milne, R., & R.Kelly, A. (2013). A new method for boarding passengers onto an airplane. *Elsevier*.
- More, D., & Sharma, R. (2014). The turnaround time of an aircraft: a competitive weapon for an airline company. *Indian Institute of Management Calcutta*.
- Mullainathan, S., & Thaler, R. (2000). Behavioral economics. *National Bureau of Economic Research*.
- My Baggage. (sense data). *Luggage Shipping - Sending Ahead Boxes and Suitcases*.
Recollit de <https://www.mybaggage.com/shipping/>
- Niehues, A., Belin, S., Hansson, T., Hauser, R., Mostajo, M., & Richter, J. (2001). How Airlines Can Improve On-time Performance. *Booz-Allen and Hamilton*.
- Norwegian. (sense data). www.norwegian.com. Recollit de Norwegian:
<https://www.norwegian.com/es/informacion-sobre-el-viaje/equipaje/>
- Nyquist, D., & McFadden, K. (2008). A study of the airline boarding problem. *J. Air Trans. Manage.*

- Real Academia de Ingeniería. (sense data). *Aeronáutica y Astronáutica*. Recollit de <http://diccionario.raing.es/es/lema/fuselaje-estrecho>
- RYANAIR. (2019). *We've Updated Our Cabin Bag Policy*. Recollit de <https://www.ryanair.com/gb/en/plan-trip/flying-with-us/baggage-policy>
- Ryanair. (sense data). *www.ryanair.com*. Recollit de Ryanair: <https://www.ryanair.com/es/es/informacin-util/equipaje/politica-de-equipaje-ryanair>
- Schneider, C., Weinmann, M., & Vom Brocke, J. (2018). Digital Nudging: Guiding Online User Choices through Interface Design. *Communications of the ACM*.
- Schultz, M. (2018). Fast Aircraft Turnaround Enabled by Reliable Passenger Boarding. *Aerospace*.
- Schwartz, S. (1977). Normative influences on altruism. *Advances in Experimental Psychology, Academic Press*.
- Scotti, D., & Dresner, M. (2015). The impact of baggage fees on passenger demand on US air routes. *Elsevier*.
- Scotti, D., Dresner, M., & Martini, G. (2016). Baggage fees, operational performance and customer satisfaction in the US air transport industry. *Elsevier*.
- Send My Bag. (sense data). *Send My Bag*. Recollit de <https://www.sendmybag.com>
- Simon, H. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*.
- Smith, C. (2004). Airline Operating Costs - The Variations. http://www.sh-e.com/presentations/csmith_081204.pdf.
- SPIRIT. (2017). *What are the sizes and weight limits for bags?* . Recollit de <https://customersupport.spirit.com/en-us/category/article/KA-01143>
- Sutkus, D., Baughcum, S., & DuBois, D. (2001). Scheduled Civil Aircraft Emission Inventories for 1999: Database Development and Analysis. *NASA/CR-2001-211216, Washington DC*.
- Székely, N., Weinmann, M., & Brocke, J. V. (2016). Nudging people to pay CO2 offsets – The effect of anchors in flight booking processes. *AIS Electronic Library*.
- TAP Portugal. (sense data). *www.flytap.com*. Recollit de TAP Portugal: <https://www.flytap.com/es-es/equipaje>
- Thaler, R., & Sunstein, C. (2003). Libertarian Paternalism. *American Economic Review*, 175 - 179.
- Volotea. (sense data). *www.volotea.com*. Recollit de Volotea: <https://www.volotea.com/es/equipaje/>
- Vueling. (sense data). *www.vueling.com*. Recollit de Vueling: <https://www.vueling.com/ca/serveis-vueling/prepara-el-teu-viatge/maletes/equipatge-de-ma>
- WIZZ-AIR. (2018). *Información y Servicios*. Recollit de <https://wizzair.com/es-es/informacion-y-servicios/informacion-del-viaje/equipaje>

- Wong, W. H., Zhang, A., Hui, Y. V., & Leung, L. C. (2009). Optimal Baggage - Limit Policy: Airline Passenger and Cargo Allocation. *Transportation science*.
- Wu, C.-L., & Caves, R. E. (2002). Modelling of aircraft rotation in a múltiple airport environment. *Pergamon*.

SECCIÓ 7
ANNEX

7.1. DADES RANDOM GENERADES

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20					
	Man?	Age	Origin	Destination	Month	Day of the week	Business?	# Flight minutes	# Trip days	# People w/ you	# Childs	Traveling w/ a PRM	# Large suitcases	# Small suitcases	Small suitcase checked?	# Flights x year	15a	15b	15c	16	17a	17b	17c	17d	
Persons	1	0	3	bcn	dbx	11	dissable	1	377	8	3	0	0	1	1	3	4	4	5	1	5	2	1	1	
	2	0	4	bcn	hnd	2	dilluns	1	762	8	1	0	0	1	1	0	3	3	2	4	1	5	2	1	1
	3	0	3	bcn	mad	10	dimecres	1	76	1	1	0	0	0	1	0	3	4	4	2	4	4	4	4	2
	4	0	4	bcn	bio	3	dimarts	1	72	1	1	0	0	0	1	1	3	3	2	2	2	4	4	4	2
	5	0	4	bcn	bio	1	dilluns	1	72	1	1	0	0	0	1	1	3	4	4	3	1	4	3	3	2
	6	0	2	bcn	mad	5	dijous	1	78	1	1	0	0	0	1	1	3	3	2	3	1	4	3	3	2
	7	0	5	bcn	mad	11	dilluns	1	76	2	1	0	0	1	1	1	3	4	4	2	1	4	4	4	2
	8	0	5	bcn	ams	2	dimecres	1	136	5	1	0	0	1	1	1	3	4	4	5	1	5	2	1	1
	9	0	2	bcn	ory	1	dimarts	1	83	2	1	0	0	1	1	1	3	4	4	3	1	4	3	3	2
	10	0	4	bcn	lgw	4	dilluns	1	126	3	1	0	0	1	1	1	3	4	4	5	1	5	2	1	1
	11	0	3	bcn	ams	12	dimarts	1	136	3	1	0	0	1	1	1	3	4	4	2	2	4	4	4	2
	12	0	4	bcn	dps	8	dissable	0	916	15	1	0	0	2	0	0	2	3	2	3	1	4	3	3	2
	13	0	3	bcn	mex	12	dimecres	0	680	10	2	1	0	1	0	0	2	4	4	5	1	5	2	1	1
	14	0	2	bcn	eze	12	dilluns	0	767	12	3	1	1	1	0	0	2	3	2	3	3	4	3	3	2
	15	0	3	bcn	jfk	4	dissable	0	449	7	10	0	0	1	0	0	2	3	2	3	5	4	3	3	2
	16	0	1	bcn	jfk	7	dimecres	0	449	7	3	0	1	1	0	0	3	4	4	2	1	4	4	4	2
	17	0	1	bcn	foo	4	dijous	0	96	4	5	3	0	1	0	0	1	3	2	4	1	5	2	1	1
	18	0	2	bcn	mah	7	divendres	0	37	4	4	2	0	0	1	0	1	3	2	2	1	4	4	4	2
	19	0	4	bcn	rak	11	dimecres	0	135	5	4	2	0	1	0	0	1	3	2	4	2	5	2	1	1
	20	0	4	bcn	ory	5	divendres	0	83	4	3	1	0	2	0	0	2	3	2	2	4	4	4	4	2
	21	0	6	bcn	lbz	5	dimarts	0	44	4	5	3	0	1	0	0	1	3	2	3	1	4	3	3	2
	22	0	2	bcn	rak	9	dimecres	0	135	5	3	1	0	2	0	0	1	3	2	4	1	5	2	1	1
	23	0	3	bcn	foo	3	dissable	0	96	4	2	1	0	1	0	0	1	3	2	3	4	4	3	3	2
	24	0	5	bcn	cph	10	dilluns	0	163	4	3	1	0	1	0	0	1	3	2	3	1	4	3	3	2
	25	0	1	bcn	pmi	9	dijous	0	32	4	2	0	0	1	0	0	1	3	2	4	1	5	2	1	1
	26	0	4	bcn	mah	9	dissable	0	37	4	2	0	1	1	0	0	2	3	2	2	1	4	4	4	2
	27	0	4	bcn	lgw	12	dissable	0	126	4	2	0	0	1	0	0	2	4	4	3	1	4	3	3	2
	28	0	1	bcn	lbz	9	divendres	0	44	3	2	0	0	1	0	0	3	4	4	5	4	5	2	1	1
	29	0	3	bcn	rak	6	dilluns	0	135	6	2	0	0	2	0	0	1	3	2	2	1	4	4	4	2
	30	0	5	bcn	lgw	3	dimarts	0	126	6	2	0	0	1	0	0	2	3	2	3	1	4	3	3	2
	31	0	1	bcn	pmi	8	dijous	0	32	4	2	0	0	1	0	0	2	3	2	4	1	5	2	1	1
	32	0	3	bcn	lbz	6	divendres	0	44	3	2	0	0	1	0	0	2	3	2	3	2	4	4	4	2
	33	0	4	bcn	foo	3	dimecres	0	96	5	2	0	0	1	0	0	2	4	4	2	1	4	4	4	2
	34	0	4	bcn	cph	4	dimarts	0	163	4	2	0	0	2	0	0	2	4	4	2	2	4	4	4	2
	35	0	5	bcn	pmi	10	dimecres	0	32	4	2	0	0	1	0	1	2	4	4	5	1	5	2	1	1
	36	0	5	bcn	lbz	10	dijous	0	44	4	2	0	0	1	0	0	2	3	2	3	3	4	3	3	2
	37	0	6	bcn	foo	11	dilluns	0	96	5	2	0	0	0	1	0	2	4	4	3	1	4	3	3	2
	38	0	4	bcn	mah	7	dijous	0	37	4	2	0	0	1	0	0	2	4	4	2	3	4	4	4	2
	39	0	4	bcn	rak	6	dumenge	0	135	6	1	0	0	0	1	0	2	3	2	2	2	4	4	4	2
	40	0	5	bcn	cph	4	dimecres	0	163	5	1	0	0	0	1	0	2	3	2	3	1	4	3	3	2
	41	0	3	bcn	mah	8	divendres	0	37	6	1	0	0	2	0	0	2	3	2	4	2	5	2	1	1
	42	0	6	bcn	pmi	3	dumenge	0	32	5	1	0	0	1	0	0	2	4	4	3	1	4	3	3	2
	43	0	6	bcn	lbz	5	dimarts	0	44	4	1	0	0	0	1	0	2	4	4	2	3	4	4	4	2
	44	0	4	bcn	foo	1	dijous	0	96	4	1	0	0	0	1	0	2	4	4	3	5	4	3	3	2
	45	1	2	bcn	dbx	10	dilluns	1	377	10	1	0	0	1	1	0	3	4	4	5	1	5	2	1	1
	46	1	4	bcn	sin	4	dimarts	1	796	10	1	0	0	1	1	0	2	3	2	2	2	4	4	4	2
	47	1	5	bcn	hnd	1	dilluns	1	762	7	2	0	0	1	1	1	3	3	2	3	1	4	3	3	2
	48	1	2	bcn	dbx	2	dimecres	1	377	4	1	0	0	1	1	1	3	3	2	2	1	4	4	4	2
	49	1	3	bcn	mad	2	dilluns	1	76	1	1	0	0	0	1	1	3	3	2	4	3	5	2	1	1
	50	1	3	bcn	bio	10	dimarts	1	72	1	1	0	0	0	1	1	3	4	4	3	1	4	3	3	2
Persons	51	1	4	bcn	ams	4	dimecres	1	136	1	1	0	0	1	1	3	4	4	5	4	5	2	1	1	
	52	1	4	bcn	mad	1	dijous	1	76	1	1	0	0	1	1	1	3	4	4	2	1	4	4	4	2
	53	1	3	bcn	ory	3	dilluns	1	83	1	1	0	0	0	1	1	3	4	4	5	1	5	2	1	1
	54	1	3	bcn	bio	5	dimarts	1	72	1	1	0	0	0	1	1	3	3	2	3	1	4	3	3	2
	55	1	4	bcn	ory	6	dilluns	1	1	2	2	0	0	0	1	1	3	4	4	2	5	4	4	4	2
	56	1	3	bcn	lgw	2	dimecres	1	126	2	1	0	0	1	1	1	3	4	4	2	2	4	4	4	2
	57	1	5	bcn	mad	11	dimarts	1	76	2	1	0	0	0	1	1	3	4	4	5	4	5	2	1	1
	58	1	5	bcn	bio	5	dilluns	1	72	2	1	0	0	1	1	1	3	2	3	2	4	3	3	3	2
	59	1	2	bcn	ory	10	dimecres	1	83	3	1	0	0	0	1	1	3	4	4	5	3	5	2	1	1
	60	1	3	bcn	ams	2	dilluns	1	136	2	1	0	0	1	0	0	3	4	4	5	4	5	2	1	1
	61	1	4	bcn	ams	6	dimarts	1	136	6	1	0	0	0	1	1	3	4	4	2	1	4	4	4	2
	62	1	4	bcn	ory	5	dimarts	1	83	3	1	0	0	0	1	1	3	4	4	3	1	4	3	3	2
	63	1	3	bcn	bio	12	dimecres	1	72	2	1	0	0	0	1	1	3	4	4	2	1	4	4	4	2
	64	1	3	bcn	lgw	9	dilluns	1	126	3	1	0	0	0	1	1	3	4	4	2	1	4	4	4	2
	65	1	1	bcn	dps	9	dilluns	0	916	11	1	0	0	2	0	0	2	4	4	5	1	5	2	1	1
	66	1	1	bcn	mex	11	dijous	0	689	9	2	0	0	1	1	0	2	3	2	3	1	4	3	3	2
	67	1	2	bcn	eze	12	dissable	0	767	10	4	2	0	2	0	0	1	3	2	4	2	5	2	1	1
	68	1	4	bcn	lax	7	divendres	0	707	7	3	0	0	0	1	0	2	3	2	3	2	4	4	3	2
	69	1	5	bcn	jfk	12	dilluns	0	449	7	4	0	0	1	0	0	2	4	4	3	4	4	3	3	2
	70	1	2	bcn	jfk	9	divendres	0	449	7	3	1	0	1	0	0	1	3	2	2	4	4	4	4	2
	71	1	2	bcn	mia	2	dijous	0	594	7	4	2	1	1	0	0	1	3	2	4	1	5	2	1	1
	72	1	3	bcn	foo	6	dimecres	0	96	5	2	1	0	1	0	0	2	4	4	3	1	4	3	3	2
	73	1	3	bcn	lgw	4	divendres	0	126	4	2	1	0	1	0	0	1	3	2	2	1	4	4	4	2
	74	1	5	bcn	lbz	6	divendres	0	44	3	2	0	0	1	1	0	2	3	2	4	1				

7.2. CODI DEL PROGRAMARI

```
[ ] import pandas as pd
from sklearn.linear_model import LogisticRegression
import pandas_profiling
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report
from sklearn import metrics
import matplotlib.pyplot as plt
import numpy as np
from statsmodels.stats.outliers_influence import variance_inflation_factor
from sklearn.metrics import roc_auc_score
from sklearn.metrics import roc_curve
import statsmodels.api as sm
from sklearn.decomposition import PCA
import collections
import scipy.cluster.hierarchy as hier
from sklearn.cluster import AgglomerativeClustering
from sklearn.metrics import silhouette_samples
from matplotlib import cm
from sklearn.cluster import DBSCAN
from sklearn.cluster import KMeans

[ ] df= pd.read_csv('Dades_Full 1.csv',sep=';',decimal=",")

#with pd.option_context('display.max_rows', None, 'display.max_columns', None):
    #print(df['Day of the week'])
print(df.shape)
df.columns
df.dtypes
df.head()
df['Small suitcase checked?'].value_counts()
df.groupby('Small suitcase checked?').mean()

[ ] #Transforming Nominal Attributes_ Day of the Week

from sklearn.preprocessing import LabelEncoder
gle = LabelEncoder()
#genre_labels = gle.fit_transform(df['Day of the week'])
#genre_mappings = {index: label for index, label in
#                  enumerate(gle.classes_)}
#genre_mappings

[ ] reate dummy variables based on day of the week
= pd.read_csv('Dades_Full 1.csv',sep=';',decimal=",")
= df.drop(['Origin','Destination'], axis=1)
transforming Nominal Attributes_ Day of the Week
['Day of the week']= df['Day of the week'].replace(['dilluns','dimarts','dimecres','dijous','divendres','dissabte','diumenge'], ['0','1','2','3','4','5','6'])
['Day of the week']= df['Day of the week'].replace(['divendres'], ['4'])
['Day of the week']= df['Day of the week'].astype(int)
```

Problema de multicolinealitat:

```
[ ] profile = pandas_profiling.ProfileReport(df,title='Flights Report', html={'style':{'full_width':True}})
profile.to_notebook_iframe()

[ ] #Pearson correlation
#Correlation matrix
plt.figure(figsize=(12,10))
ax=sns.heatmap(df.corr(), annot=True, cmap=sns.cubehelix_palette(20, light=0.95, dark=0.15))
ax.xaxis.tick_top
plt.savefig('correlation matrix')
plt.show()
```

Factor d'inflació de la variança (VIF):

```
[ ] # This checks VIF values and then drops variables whose VIF is more than 5

def calculate_vif(X, thresh=5.0):
    variables = list(range(X.shape[1]))
    dropped = True
    while dropped:
        dropped = False
        vif = [variance_inflation_factor(X.iloc[:, variables].values, ix)
              for ix in range(X.iloc[:, variables].shape[1])]
        maxloc = vif.index(max(vif))
        if max(vif) > thresh:
            print('dropping \'' + X.iloc[:, variables].columns[maxloc] +
                  '\' at index: ' + str(maxloc))
            del variables[maxloc]
            dropped = True

    print('Remaining variables:')
    print(X.columns[variables])
    return X.iloc[:, variables]

[ ] calculate_vif(df)
#X_train
#new_df.drop('Small suitcase checked?',axis=1)

[ ] # According to the Variance Inflation Factor(VIF) selection features method
# These are the final features given by VIF
new_df= df.drop(['Age',' Trip days',' People traveling w/ you',' Small suitcases',' Flights x year','15a','15b','15c','17a','17b','17c','17d'], axis=1)

#Implementing the model
logit_model=sm.Logit(new_df['Small suitcase checked?'],new_df.drop(['Small suitcase checked?'], axis=1))
result=logit_model.fit()
print(result.summary2())

[ ] #Test Train Split
#Logistic Regression Model Fitting

X_train, X_test, y_train, y_test = train_test_split(new_df.drop('Small suitcase checked?',axis=1),
                                                    new_df['Small suitcase checked?'], test_size=0.20,
                                                    random_state=0)

logreg = LogisticRegression(max_iter=370)
logreg.fit(X_train, y_train)
```

Anàlisi per a la Gestió Eficient de l'Equipatge de Cabina en l'Aviació Comercial

```
[ ] #Predicting the test set results and calculating the accuracy
#Keep this simple and use accuracy as our metric
#(fraction of correct predictions): correct predictions / total number of data points

#score = logmodel.score(X_test, y_test)
#print(score)
y_pred = logreg.predict(X_test)
print('Accuracy of logistic regression classifier on test set: {:.2f}'.format(logreg.score(X_test, y_test)))

[ ] #Evaluation
print(classification_report(y_test, y_pred))
#The support is the number of occurrences of each class in y_test.

print("Accuracy:",metrics.accuracy_score(y_test, y_pred))
print("Precision:",metrics.precision_score(y_test, y_pred))
print("Recall:",metrics.recall_score(y_test, y_pred))
```

Filter Method:

```
[ ] #1. Filter Method

#Correlation with output variable
cor=df.corr()
cor_target = abs(cor["Small suitcase checked?"])
#Selecting highly correlated features
relevant_features = cor_target[cor_target>0.5]
relevant_features

[ ] print(df[["Business?","Large suitcases"]].corr())
print(df[["Small suitcases","Business?"]].corr())
print(df[["Flights x year","Business?"]].corr())

[ ] df= pd.read_csv('Dades_Full 1.csv',sep=';',decimal=",")
df= df.drop(['Flights x year',axis=1])
df=df[['Business?','Large suitcases', 'Small suitcases', 'Small suitcase checked?']]

#Implementing the model
logit_model=sm.Logit(df[['Small suitcase checked?']],df.drop(['Small suitcase checked?'], axis=1))
result=logit_model.fit()
print(result.summary2())

[ ] #Test Train Split
#Logistic Regression Model Fitting

X_train, X_test, y_train, y_test = train_test_split(df.drop('Small suitcase checked?',axis=1),
df[['Small suitcase checked?']], test_size=0.20,
random_state=0)

logreg = LogisticRegression()
logreg.fit(X_train, y_train)

y_pred = logreg.predict(X_test)
print('Accuracy of logistic regression classifier on test set: {:.2f}'.format(logreg.score(X_test, y_test)))

#Evaluation
print(classification_report(y_test, y_pred))
#The support is the number of occurrences of each class in y_test.

print("Accuracy:",metrics.accuracy_score(y_test, y_pred))
print("Precision:",metrics.precision_score(y_test, y_pred))
print("Recall:",metrics.recall_score(y_test, y_pred))
```

Step Forward Regression:

```
[ ] from mlxtend.feature_selection import SequentialFeatureSelector as sfs
logreg = LogisticRegression(max_iter=1000)

X_train, X_test, y_train, y_test = train_test_split(df.drop('Small suitcase checked?',axis=1),
df[['Small suitcase checked?']], test_size=0.20,
random_state=0)

#Build step forward feature selection
sfs1= sfs(logreg, k_features=2, forward= True, floating=False,scoring='r2',verbose=2,cv=5)
#Perform SFS
sfs1= sfs1.fit(X_train, y_train)

[ ] feat_cols= list(sfs1.k_feature_idx_)
print(feat_cols)
df.drop('Small suitcase checked?',axis=1).columns
#So most important features according to the forward selection method is '# Childs', '# Small suitcases'

[ ] df= pd.read_csv('Dades_Full 1.csv',sep=';',decimal=",")
df=df[['Childs', '# Small suitcases', 'Small suitcase checked?']]

#Implementing the model
logit_model=sm.Logit(df[['Small suitcase checked?']],df.drop(['Small suitcase checked?'], axis=1))
result=logit_model.fit()
print(result.summary2())
```

Anàlisi per a la Gestió Eficient de l'Equipatge de Cabina en l'Aviació Comercial

```
[ ] #Test Train Split
#Logistic Regression Model Fitting

X_train, X_test, y_train, y_test = train_test_split(df.drop('Small suitcase checked?',axis=1),
                                                  df['Small suitcase checked?'], test_size=0.20,
                                                  random_state=0)

logreg = LogisticRegression()
logreg.fit(X_train, y_train)

y_pred = logreg.predict(X_test)
print('Accuracy of logistic regression classifier on test set: {:.2f}'.format(logreg.score(X_test, y_test)))

#Evaluation
print(classification_report(y_test, y_pred))
#The support is the number of occurrences of each class in y_test.

print("Accuracy:",metrics.accuracy_score(y_test, y_pred))
print("Precision:",metrics.precision_score(y_test, y_pred))
print("Recall:",metrics.recall_score(y_test, y_pred))
```

Backward Elimination:

```
[ ] x = df.drop('Small suitcase checked?',axis=1)
y = df['Small suitcase checked?']
cols = list(x.columns)
pmax = 1
while (len(cols)>0):
    p = []
    x_1 = x[cols]
    x_1 = sm.add_constant(x_1)
    model = sm.OLS(y,x_1).fit()
    p = pd.Series(model.pvalues.values[1:],index = cols)
    pmax = max(p)
    feature_with_p_max = p.idxmax()
    if(pmax>0.05):
        cols.remove(feature_with_p_max)
    else:
        break
selected_features_BE = cols
print(selected_features_BE)
```

```
[ ] df= pd.read_csv('Dades_Full 1.csv', sep=';', decimal=',')
df=df[['Age', '# Trip days', '# Small suitcases', 'Small suitcase checked?']]

#Implementing the model
logit_model=sm.Logit(df['Small suitcase checked?'],df.drop(['Small suitcase checked?'], axis=1))
result=logit_model.fit()
print(result.summary2())
```

```
[ ] #Test Train Split
#Logistic Regression Model Fitting

X_train, X_test, y_train, y_test = train_test_split(df.drop('Small suitcase checked?',axis=1),
                                                  df['Small suitcase checked?'], test_size=0.20,
                                                  random_state=0)

logreg = LogisticRegression()
logreg.fit(X_train, y_train)

y_pred = logreg.predict(X_test)
print('Accuracy of logistic regression classifier on test set: {:.2f}'.format(logreg.score(X_test, y_test)))

#Evaluation
print(classification_report(y_test, y_pred))
#The support is the number of occurrences of each class in y_test.

print("Accuracy:",metrics.accuracy_score(y_test, y_pred))
print("Precision:",metrics.precision_score(y_test, y_pred))
print("Recall:",metrics.recall_score(y_test, y_pred))
```

```
[ ] #Confusion Matrix
from sklearn.metrics import confusion_matrix
confusion_matrix = confusion_matrix(y_test, y_pred)
print(confusion_matrix)

#The result is telling us that we have 9+9 correct predictions and 0+2 incorrect predictions.
```

```
[ ] #Visualizing Confusion Matrix using Heatmap
plt.figure(figsize=(8,8))
sns.heatmap(confusion_matrix, annot=True, fmt=".3f", linewidths=.5, square = True, cmap = 'Blues_r');
plt.tight_layout()
plt.ylabel('Actual label');
plt.xlabel('Predicted label');
plt.title('Confusion matrix', y=1.1)
all_sample_title = 'Accuracy Score: {}'.format(logreg.score(X_test, y_test))
plt.title(all_sample_title, size = 15);
```

```
[ ] logit_roc_auc = roc_auc_score(y_test, logreg.predict(X_test))
fpr, tpr, thresholds = roc_curve(y_test, logreg.predict_proba(X_test)[:,1])
plt.figure(figsize=(10,10))
plt.plot(fpr, tpr, label='Logistic Regression (area = %0.2f)' % logit_roc_auc)
plt.plot([0, 1], [0, 1], 'r--')
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('Receiver operating characteristic')
plt.legend(loc='lower right')
#plt.savefig('Log_ROC')
plt.show()
```

Clustering k-means:

```
[ ] df.shape
df.describe()

[ ] #Data Exploration

#Q15-Q17 are all Likert-type, meaning that the values are taken from {1,2,3,4,5}
X_questions = df.iloc[:,14:22]
question_means = X_questions.mean(axis = 0)
grand_mean = question_means.mean()
print(question_means)
print(grand_mean)

[ ] #X=df
X=df.iloc[:,14:22]
#X= X.drop(['15a', '15b', '15c', '16'],axis=1)
# reduce to 2 dimensions
pca = PCA(n_components = 2, random_state=1)
X_pca = pca.fit_transform(X)

print('Explained Variance Ratio : ' + str(pca.explained_variance_ratio_.cumsum()[1]))

[ ] #Training Models : K-Means
distortions = []
K_to_try = range(1, 6)

for i in K_to_try:
    model = KMeans(
        n_clusters=i,
        init='k-means++',
        # n_init=10,
        # max_iter=300,
        # n_jobs=-1,
        random_state=1)
    model.fit(X_pca)
    distortions.append(model.inertia_)

plt.plot(K_to_try, distortions, marker='o')
plt.xlabel('Number of Clusters (k)')
plt.ylabel('Distortion')
plt.show()

[ ] # use the best K from elbow method
model = KMeans(
    n_clusters=3,
    init='k-means++',
    # n_init=10,
    # max_iter=300,
    # n_jobs=-1,
    random_state=1)

model = model.fit(X_pca)
y = model.predict(X_pca)

[ ] plt.scatter(X_pca[:, 0], X_pca[:, 1], s=50, c = 'yellow', label = 'Cluster 1')
plt.scatter(X_pca[:, 0], X_pca[:, 1], s=50, c = 'green', label = 'Cluster 2')
plt.scatter(X_pca[:, 0], X_pca[:, 1], s = 50, c = 'red', label = 'Cluster 3')
plt.scatter(model.cluster_centers_[0], model.cluster_centers_[1], s = 100, c = 'blue', label = 'Centroids')

plt.xlabel('Principle Component 1')
plt.ylabel('Principle Component 2')
plt.legend(loc=(1.04, 0))
plt.grid()
plt.show()

print('K Means Result : ')
print(collections.Counter(y))

[ ] cluster_labels = np.unique(y)
n_clusters = cluster_labels.shape[0]
silhouette_vals = silhouette_samples(X_pca, y, metric='euclidean')

y_ax_lower, y_ax_upper = 0, 0
yticks = []
for i, c in enumerate(cluster_labels):
    c_silhouette_vals = silhouette_vals[y == c]
    c_silhouette_vals.sort()
    y_ax_upper += len(c_silhouette_vals)
    color = cm.jet(float(i) / n_clusters)
    plt.barh(range(y_ax_lower, y_ax_upper),
             c_silhouette_vals,
             height=1.0,
             edgecolor='none',
             color=color)
    yticks.append((y_ax_lower + y_ax_upper) / 2.)
    y_ax_lower += len(c_silhouette_vals)
silhouette_avg = np.mean(silhouette_vals)

plt.axvline(silhouette_avg, color="red", linestyle="--")
plt.yticks(yticks, cluster_labels + 1)
plt.ylabel('Cluster')
plt.xlabel('Silhouette coefficient')
plt.show()

[ ] #Without applying PCA

# use best k from Elbow method
model_k = KMeans(
    n_clusters=3,
    init='k-means++',
    # n_init=10,
    # max_iter=300,
    # n_jobs=-1,
    random_state=1)

# fit with X instead of X_pca
model_k = model_k.fit(X)

y_final = model_k.predict(X)

print('Final K Means Result (no PCA) : ')
print(collections.Counter(y_final))
```

```
[ ] y_final = pd.DataFrame(y_final, columns=['cluster'])

raw_result = pd.concat([X, y_final], axis=1)

y = pd.DataFrame(y, columns=['cluster'])
raw_result_pca = pd.concat([X, y], axis=1)

mean_by_student_1 = raw_result[raw_result['cluster']==0].iloc[:, 0:28].mean(axis = 1)
mean_by_student_2 = raw_result[raw_result['cluster']==1].iloc[:, 0:28].mean(axis = 1)
mean_by_student_3 = raw_result[raw_result['cluster']==2].iloc[:, 0:28].mean(axis = 1)

print('Mean cluster 1 : ' + str(mean_by_student_1.mean()) + ',STD : ' + str(mean_by_student_1.std()))
print('Mean cluster 2 : ' + str(mean_by_student_2.mean()) + ',STD : ' + str(mean_by_student_2.std()))
print('Mean cluster 3 : ' + str(mean_by_student_3.mean()) + ',STD : ' + str(mean_by_student_3.std()))
```

Agglomerative Clustering:

```
[ ] # Ward
```

```
dendrogram = hier.dendrogram(hier.linkage(X_pca, method = 'ward'))
plt.title('Dendrogram')
plt.xlabel('questions')
plt.ylabel('Euclidean distances')
plt.show()
```

```
[ ] model = AgglomerativeClustering(n_clusters = 2,
    affinity='euclidean',
    linkage='ward')
y = model.fit_predict(X_pca)

plt.scatter(X_pcaly == 0, X_pcaly == 0, 1], s = 50, c = 'yellow', label = 'Cluster 1')
plt.scatter(X_pcaly == 1, 0], X_pcaly == 1, 1], s = 50, c = 'green', label = 'Cluster 2')
plt.title('Clusters of passengers')
plt.xlabel('PCA 1')
plt.ylabel('PCA 2')
plt.legend(loc=(1.04, 0))
plt.grid()
plt.show()

print('AgglomerativeClustering w/ Ward Result : ')
print(collections.Counter(y))
```

```
[ ] y = pd.DataFrame(y, columns=['cluster'])
raw_result = pd.concat([X, y], axis=1)

# raw_result.to_csv('YOUR_PATH/output/raw_result.csv', index=False)
# raw_result_pca.to_csv('YOUR_PATH/output/raw_result_pca.csv', index=False)

mean_by_student_1 = raw_result[raw_result['cluster']==0].iloc[:, 0:28].mean(axis = 1)
mean_by_student_2 = raw_result[raw_result['cluster']==1].iloc[:, 0:28].mean(axis = 1)

print('Mean of cluster 1 : ' + str(mean_by_student_1.mean()) + ',STD : ' + str(mean_by_student_1.std()))
print('Mean of cluster 2 : ' + str(mean_by_student_2.mean()) + ',STD : ' + str(mean_by_student_2.std()))
```

```
[ ] #Complete
dendrogram = hier.dendrogram(hier.linkage(X_pca, method = 'complete'))
plt.title('Dendrogram')
plt.xlabel('questions')
plt.ylabel('Euclidean distances')
plt.show()
```

```
[ ] model = AgglomerativeClustering(n_clusters =3,
    affinity='euclidean',
    linkage='complete')
y = model.fit_predict(X_pca)

plt.scatter(X_pcaly == 0, 0], X_pcaly == 0, 1], s = 50, c = 'yellow', label = 'Cluster 1')
plt.scatter(X_pcaly == 1, 0], X_pcaly == 1, 1], s = 50, c = 'green', label = 'Cluster 2')
plt.scatter(X_pcaly == 2, 0], X_pcaly == 2, 1], s = 50, c = 'red', label = 'Cluster 3')
plt.title('Clusters of passengers')
plt.xlabel('PCA 1')
plt.ylabel('PCA 2')
plt.legend(loc=(1.04, 0))
plt.grid()
plt.show()

print('AgglomerativeClustering c/ Complete Result : ')
print(collections.Counter(y))
```

```
[ ] y = pd.DataFrame(y, columns=['cluster'])
raw_result = pd.concat([X, y], axis=1)

# raw_result.to_csv('YOUR_PATH/output/raw_result.csv', index=False)
# raw_result_pca.to_csv('YOUR_PATH/output/raw_result_pca.csv', index=False)

mean_by_student_1 = raw_result[raw_result['cluster']==0].iloc[:, 0:28].mean(axis = 1)
mean_by_student_2 = raw_result[raw_result['cluster']==1].iloc[:, 0:28].mean(axis = 1)
mean_by_student_3 = raw_result[raw_result['cluster']==2].iloc[:, 0:28].mean(axis = 1)

print('Mean cluster 1 : ' + str(mean_by_student_1.mean()) + ',STD : ' + str(mean_by_student_1.std()))
print('Mean cluster 2 : ' + str(mean_by_student_2.mean()) + ',STD : ' + str(mean_by_student_2.std()))
print('Mean cluster 3 : ' + str(mean_by_student_3.mean()) + ',STD : ' + str(mean_by_student_3.std()))
```