

Airline Passenger Satisfaction Prediction And Analysis

B07702011 陳韋傑、B06504104 石子仙、B07701222 莊啟宏、B07702060 李宇軒

● 產業背景與任務描述

○ 動機/產業背景

預測並提升航空服務業中顧客的滿意度，有以下兩個好處

■ 培養忠誠顧客，讓顧客再次消費

- 留住一個舊顧客的成本只有開發一個新顧客的 $\frac{1}{6}$ (Thomas et al., 2004)
- 國內兩項研究顯示，航空業顧客滿意度高對於購後行為傾向有正面影響
(一、葉晶雯講師, 2003; 二、湯玲郎副教授等三人, 2021)

■ 積累好評，達到正向口碑宣傳，提升企業形象

- 航空業競爭的產品相似且不易區隔，好的企業形象為顧客重要的決策依據

○ 任務描述

■ 探究是否能藉由旅客資料與問卷結果準確預測旅客滿意與否：

以整體資料建構預測模型，透過各項指標檢視模型成效

■ 了解實施問卷調查是否有助於預測旅客滿意與否：

比較以旅客資料建構的預測模型與以問卷結果建構的預測模型

■ 找出不同客群在意的重要指標，藉此提出經營策略建議：

以不同客群的資料建立預測模型、比較重要變數的異同

● 資料收集與探索

○ 資料集來源

此資料集為 kaggle 上 "Airline Passenger Satisfaction"。網址為：

<https://www.kaggle.com/teejmahal20/airline-passenger-satisfaction>

○ 欄位說明：

■ 自變數：

- 旅客與旅程資料：

欄位名稱	說明	資料內容
id	旅客編號	整數
Gender	性別	Male (男性); Female (女性)

Customer Type	旅客忠誠分類	Loyal Customer (忠誠顧客); Disloyal Customer (非忠誠顧客)
Age	年齡	整數
Type Of Travel	旅行目的	Business travel (商務旅行); Personal Travel (非商務旅行)
Class	艙位分級	Eco (經濟艙); Eco Plus (舒適經濟艙); Business (商務艙)
Flight Distance	飛行距離	整數
Departure Delay in Minutes	起飛延誤時間	整數, 以分鐘計
Arrival Delay in Minutes	抵達延誤時間	整數, 以分鐘計

- 問卷結果：

欄位名稱	說明	資料內容
Inflight wifi service	機上Wifi	
Departure/Arrival time convenient	航班時間便利性	
Ease of Online booking	線上訂票便利性	
Gate location	登機門位置	
Food and drink	餐飲	
Online boarding	線上報到	
Seat comfort	座位	
Inflight entertainment	機上娛樂	
On-board service	登機服務	
Leg room service	腿部空間	

乘客搭乘該航班時對於各項服務的問卷調查，每項皆為0至5分

Baggage handling	行李處理	
Checkin service	登機服務	
Inflight service	機上服務	
Cleanliness	整潔度	

■ 應變數：

- 旅客滿意分類：

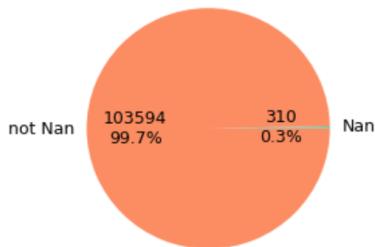
欄位名稱	說明	資料內容
satisfaction	旅客滿意分類	satisfied (滿意); neutral or dissatisfied (中立或不滿意)

● 資料探索與前處理

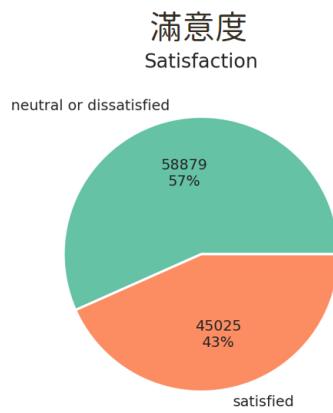
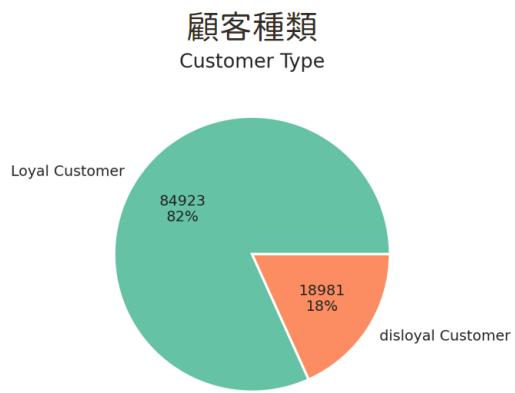
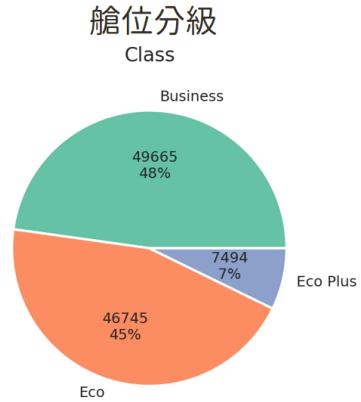
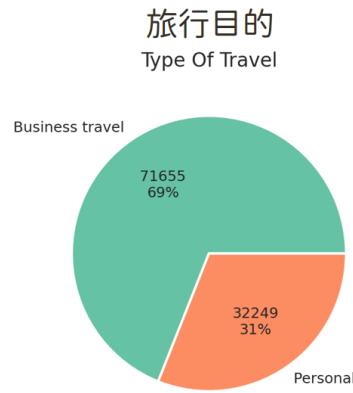
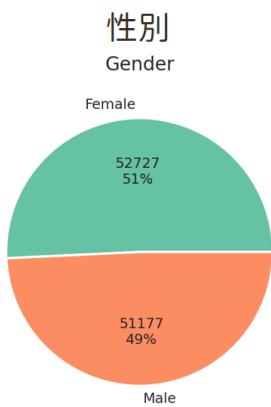
- 資料探索：

首先我們先查看資料中是否有NA。只有Arrival Delay in Minutes (抵達延誤時間) 有，占整體資料約0.3%。

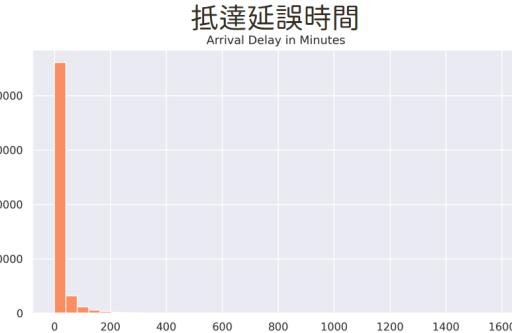
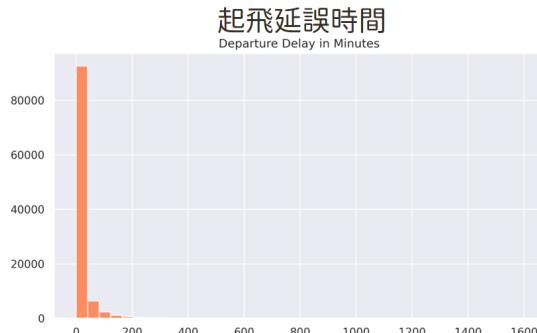
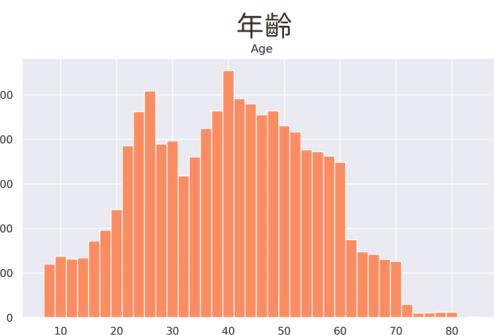
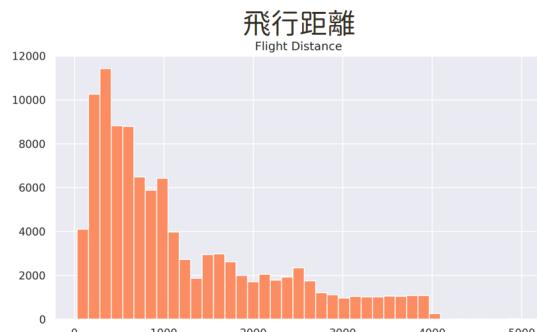
Nan (Arrival Delay in Minutes)



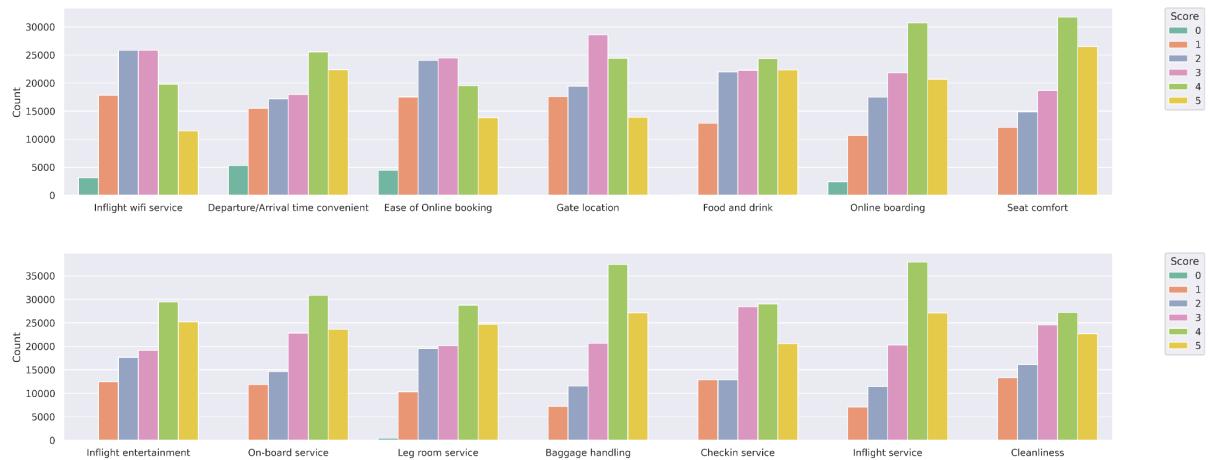
再來視覺化呈現各欄位，性別大概是一半一半；商業旅行站大多數；而經濟艙、商務艙差不多，頭等艙較少；顧客種類多為忠誠客戶，且大部分滿意度為中立或不滿意。



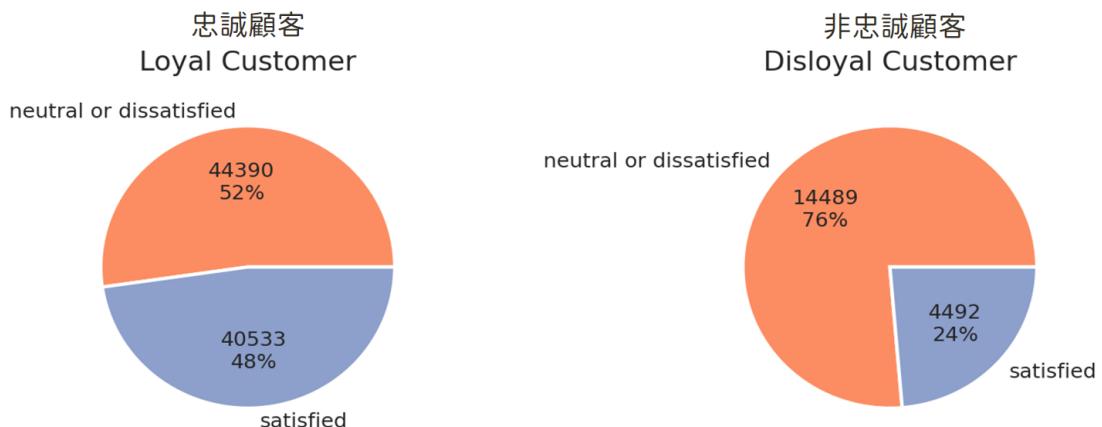
再來看飛行距離、年齡、起飛延誤時間與抵達延誤時間，在年齡上接近常態分佈。大部分的航班飛行距離都偏短，並且起飛延誤時間都不是很久。



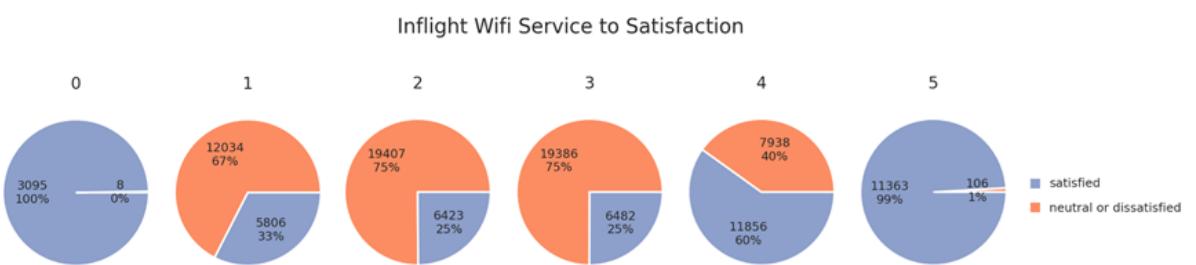
接下來看問卷資料，大部分的分數都落在三到四分，而少數欄位出現零分。



除了單純呈現各分布，我們對資料進行近一步的分析。在忠誠與非忠誠顧客中，滿意度有明顯的差異。非忠誠顧客滿意的比例小於忠誠顧客非常多。



而我們也對問卷欄位進行上述的分析，發現分數為0的顧客都是滿意，不太符合常理。後來到原資料集討論區看發現那個0應為NA，所以在前處理的部分，我們也將評分為0欄位視為NA處理。



- 資料前處理：

在前處理的部分，我們將問卷評分為0的資料轉換為NA，再drop所有NA值（有NA欄位包含抵達延誤時間、各項問卷資料）。之後對欄位進行Normalization，以進行後續模型預測。

- 模型預測與資料分析

- 方法描述

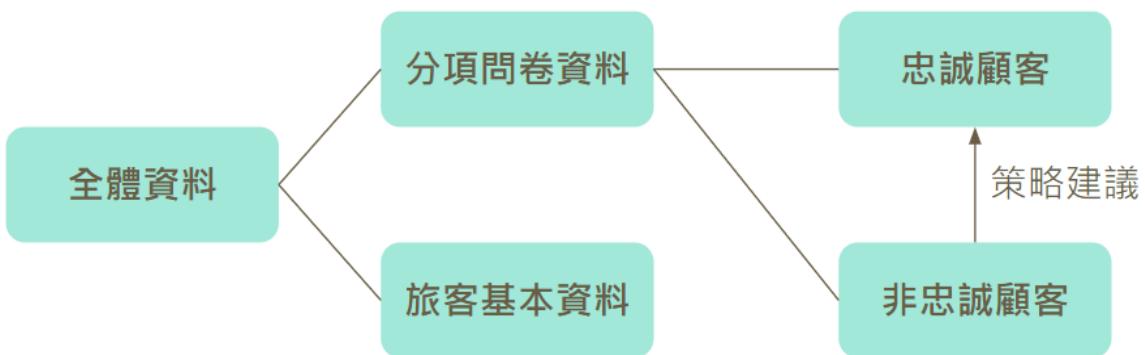
我們這邊使用多種分類模型來對資料進行預測與分析，包含：

- KNN
- Gradient Boosting Machine
- Random Forest
- XGBoost
- Decision Tree

最後會再將以上Model做Ensemble，並且使用Accuracy, Precision, Recall, F1-score等指標來評量模型成效。我們的任務有以下兩個：

- 預測顧客對於整體服務是否滿意
- 使用投票方式，分析重要變數，並給出具體策略建議

我們的分析架構大致如下：



我們會先對全體資料進行預測，並將全體資料拆解成分項問卷資料與旅客基本資料，以了解問卷對於預測滿意與否的幫助。最後再將問卷資料拆解成忠誠顧客與非忠誠顧客，了解其中的重要變數以後，便可以針對結果進行分析與策略建議。

- 模型預測與資料分析

- 預測:全體資料

Model / Metric	Accuracy	Precision	Recall	F1-score
KNN	0.9383	0.9637	0.8901	0.9254
Gradient Boosting	0.9604	0.9695	0.9373	0.9532
Random Forest	0.9623	0.9721	0.9394	0.9555
XGBoost	0.9649	0.9730	0.9425	0.9575
Decision Tree	0.9432	0.9412	0.9258	0.9334
Ensemble	0.9627	0.9729	0.9396	0.9560
Average	0.9553	0.9654	0.9291	0.9468

可以看到全體資料的預測表現在各個指標表現的都還不錯，都可以達到0.9。

- 預測:旅客基本資料

Model / Metric	Accuracy	Precision	Recall	F1-score
KNN	0.7905	0.7614	0.7473	0.7543
Gradient Boosting	0.8065	0.7614	0.8013	0.7809
Random Forest	0.7730	0.7370	0.7348	0.7359
XGBoost	0.8061	0.7627	0.7975	0.7797
Decision Tree	0.8039	0.7555	0.8049	0.7794
Ensemble	0.8055	0.7629	0.7950	0.7786

Average	0.7976	0.7568	0.7801	0.7681
---------	--------	--------	--------	--------

可以看到僅用旅客基本資料進行預測的話，各個指標表現略差，落在0.8上下。

■ 預測：分項問卷資料

Model / Metric	Accuracy	Precision	Recall	F1-score
KNN	0.9268	0.9463	0.8799	0.9119
Gradient Boosting	0.9420	0.9477	0.9157	0.9314
Random Forest	0.9464	0.9520	0.9219	0.9367
XGBoost	0.9485	0.9562	0.9227	0.9391
Decision Tree	0.9249	0.9314	0.8910	0.9108
Ensemble	0.9470	0.9550	0.9202	0.9373
Average	0.9393	0.9481	0.9086	0.9279

可以看到若僅使用分項問卷資料，雖結果稍比全體資料差，但表現依然相當不錯，各項指標也有達到0.9以上。

透過以上三個預測結果，我們大致可以歸納出以下幾點：

- 全體資料滿意度預測之 Accuracy, Precision, Recall, F1 皆達到 0.9 以上
- KNN 表現多較差、Boosting 與 Ensemble 表現多不錯
- 問卷資訊對了解顧客滿意度的確有幫助

在進行完預測以後，我們接著想透過後四個模型找出來的重要變數進行分析。因此我們先將客群分為忠誠顧客與非忠誠顧客，再對找出來的重要變數進行分析與研究。

■ 重要變數:分項問卷資料 - 忠誠顧客

Features (Not in order)	Votes	Description
Leg room service	4	Satisfaction level of Leg room service
Inflight wifi service	4	Satisfaction level of the inflight wifi service
Departure/Arrival time convenient	4	Satisfaction level of Departure/Arrival time convenient
Gate location	4	Satisfaction level of Gate location
On-board service	4	Satisfaction level of On-board service
Inflight entertainment	4	Satisfaction level of inflight entertainment
Online boarding	4	Satisfaction level of online boarding

■ 重要變數:分項問卷資料 - 非忠誠顧客

Features (Not in order)	Votes	Description
Leg room service	4	Satisfaction level of Leg room service
Inflight wifi service	4	Satisfaction level of the inflight wifi service
Departure/Arrival time convenient	4	Satisfaction level of Departure/Arrival time convenient
Gate location	4	Satisfaction level of Gate location
On-board service	4	Satisfaction level of On-board service
Baggage handling	4	Satisfaction level of baggage handling
Checkin service	4	Satisfaction level of Check-in service
Inflight service	4	Satisfaction level of inflight service

透過以上兩個分析結果，我們大致可以歸納出以下幾點：

- 對所有顧客而言，座位空間、機上 Wi-fi、航程時間方便性、登機門位置、機上服務皆是影響滿意度的重要因子
- 對忠誠顧客而言，較重視機上娛樂與線上登機方便性，推測因常搭乘，故較注重搭乘體驗與訂位方便性
- 對非忠誠顧客而言，較重視行李托運、報到服務與航程中服務，推測多是新客，較注重整體服務品質與流程順暢度

● 策略建議

- 策略建議對象：美國各大航空公司，由於在此資料集中並未指明資料來源的航空公司，因此在此策略建議中，我們的策略建議對象為美國各大航空公司
- 策略目標：透過提升顧客滿意度，將非忠誠顧客轉為忠誠顧客，進而提升航空公司的營收
- 策略建議的重要因子：我們從模型中，挑選出在忠誠顧客與非忠誠顧客中相同的重要因子，有以下四項
 - 機上服務(On-board service)
 - 機上網路(Infight wi-fi service)
 - 登機門位置(Gate Location)
 - 座椅空間(Leg Room Service)

我們針對以上四項重要因子，找尋可能會讓顧客感到不滿意的問題，並提供相對應的策略建議：

- 機上服務(On-board service)
 1. 問題描述：服務種類過少
 2. 建議：提升服務多樣性，可提供機上Wi-fi、機上娛樂（如：電視、電影、遊戲、音樂）、機上餐食（如：正餐、點心、飲品、酒水）、購物產品（如：周邊商品、Outlet）、個人化服務
- 機上網路(Infight wi-fi service)
 1. 問題描述：若未提供機上網路可能會降低顧客滿意度，根據2021 TPG美國最佳航空公司報告，前7名的航空公司都有提供機上網路，而8-10名的航空公司則沒有提供機上網路

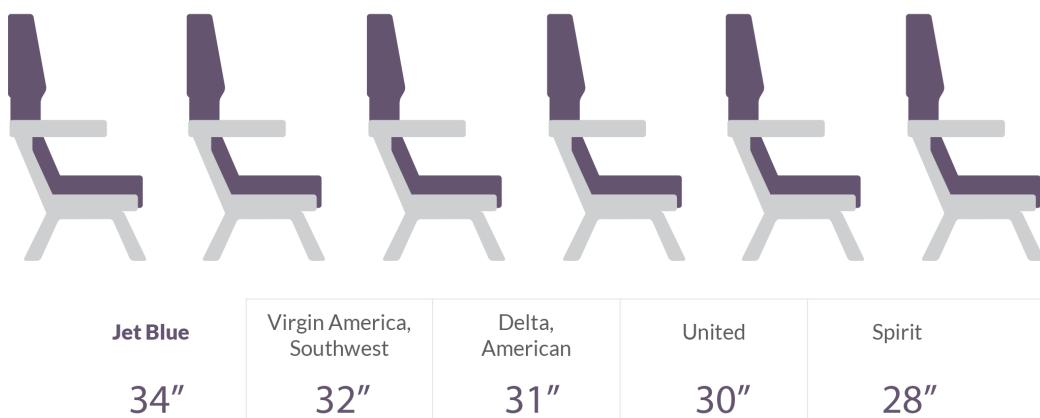
2. 建議:除了建議各大航空公司都應該提供機上網路之外，可以盡興方案的區分。
 - a. 基本方案:較為便宜，有限制流量與速度
 - b. 進階方案:價格較高，無限制流量與速度
 - c. 訂閱制方案：提供給時常搭乘飛機的顧客，可在一個月內無限量的使用

- 登機門位置(Gate Location)

1. 問題描述:在高峰時段，不同登機門會有顧客流量不平衡的現象，某些登機門人很多造成堵塞，而其他登機門則人流量低
2. 建議:利用Airport Gate Assignment(AGA模型)最大限度地減少不同登機門的乘客數量差異
 - a. 目標函式： $\min \sum_{k \in K} \sum_l (G_{lkt})^2 + M^2 \sum_{ij} y_{ij}$
 - b. 變數定義：
 - i. G : 在時間點t時使用第k個內部登機門前往第i個登機門的乘客數量
 - ii. y : 當航班i使用登機門j時，y為1，否則為0
 - iii. M : 一個很大的數字

- 座椅空間(Leg Room Service)

1. 問題描述:若座椅空間過小，顧客的腳無法自然伸展，可能會降低顧客的旅程滿意度，但座椅空間過大又會減少座位數量
2. 建議:參考美國各大航空公司的座椅空間，建議座椅空間在30、31英吋左右
3. 參考來源:可以看見下圖中在2021 TPG美國最佳航空公司排名第三的United座椅空間為30英吋，第一名的Delta座椅空間為31吋，然而Spirit的座椅空間為28英吋，在排名上卻來到了第八名



● 結論

- 預測效果佳：我們使用了多種模型以及投票的方式來預測航空公司的顧客滿意度，各個模型在準確率等各項指標都達到了0.9以上，我們認為這個資料集中，變數的分布對於顧客滿意與否本身就有一定程度的區分程度，因此各個模型的預測結果都很高。
- 問卷對於預測顧客滿意度有幫助：在預測顧客滿意度時，僅使用問卷資料就能很好的預測顧客滿意度，預測準確率等各項指標在0.9以上。
- 針對模型篩選出的重要變數進行策略建議：對於美國各大航空公司，我們針對模型篩選出忠誠與非忠誠顧客共同的四項重要變數進行策略建議，包括機上服務、機上網路、登機門位置、座椅空間。透過提升滿意度，預期讓非忠誠顧客轉為忠誠顧客，提升美國各大航空公司的營收。

● 參考資料

1. 國內研究一：葉晶雯 <http://aca.cust.edu.tw/pub/journal/29/29-19.pdf>
2. 國內研究二：湯玲郎 <https://www.iot.gov.tw/cp-86-205628-ebdd0-1.html>
3. TPG special report: Best US airlines of 2021 :
<https://thepointsguy.com/news/tpg-2021-best-us-airlines/>
4. Airport Gate Assignment for Improving Terminals' Internal Gate Efficiency, Lee et al., 2016 :
https://www.researchgate.net/publication/319092802_Airport_Gate_Assignment_for_Improving_Terminals'_Internal_Gate_Efficiency