

**Informe del tp 1**

# **Trabajo Practico**

**nº1**

**Analisis Exploratorio de la  
aerolínea**



Integrantes	Padrón
Marianela Fernanda Gareca Janko	109606
Beltran Malbrán	110036

## Hoja de ruta

Hoja de ruta

Link a Drive con los materiales:

Hipótesis 1

Archivos usados

Desarrollo

Conclusión

Hipótesis 2:

Archivos usados

Desarrollo:

Conclusión

Hipótesis 3:

Archivos usados

Desarrolló:

Conclusión:

Hipótesis 4:

Archivos usados

Desarrollo:

Conclusión:

Hipótesis 5:

Archivos usados

Desarrollo:

Conclusión

Hipótesis 6:

Archivos usados

Desarrollo:

Conclusión

## Link a Drive con los materiales:

[https://drive.google.com/drive/folders/1yzugR3R3JMIwtenu3K76ExPgWaO4LZ82?usp=drive\\_link](https://drive.google.com/drive/folders/1yzugR3R3JMIwtenu3K76ExPgWaO4LZ82?usp=drive_link)

## Hipótesis 1

**¿Existe una relación entre la satisfacción del pasajero y la elección de servicios específicos durante el vuelo? ¿Cómo esta relación varía entre diferentes**

**regiones geográficas?**

## **Archivos usados**

Para el análisis exploratorio de datos de esta hipótesis utilizamos los archivos de `airline_data.csv` y `costumer_airways_data.csv`.

## **Desarrollo**

Para este análisis, se hicieron modificaciones a los archivos para su uso correcto. Se eliminaron las filas duplicadas, se recategorizaron los tipos de datos para optimizar el uso de memoria, mejorando así su velocidad de procesamiento. Se filtraron aquellos países que no alcanzaban el umbral mínimo de encuestas para ser representativos, es decir, aquellos que contaban con menos del 75% del total de encuestas realizadas. Además, solo se tomaron en cuenta las encuestas que tenían la reserva completa.

Una vez hecha la limpiezas de datos correspondientes se analizaron los cinco países con el mejor promedio de puntaje. Se puede ver que en estos países a mayor promedio de puntaje las preferencia incrementan, esto es un indicador de que aunque las preferencias varían entre las diferentes categorías de servicios, hay una tendencia hacia puntuaciones más altas en relación con ciertas preferencias, como se muestra en la figura 1.

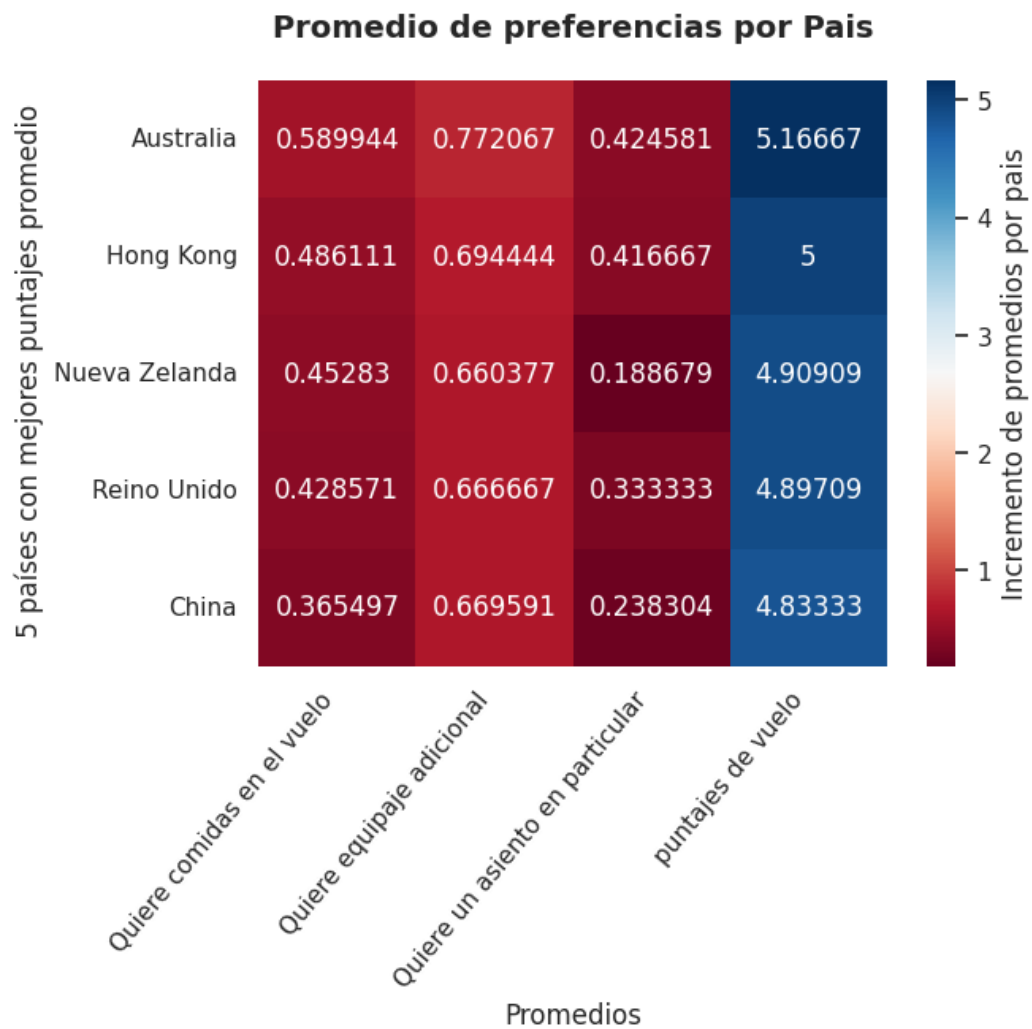


Figura 1— Relación de promedio de preferencias y puntajes de los 5 países con mejor promedio.

Para verificar si la hipótesis está respaldada por los datos, analizamos los cinco países con el puntaje promedio más bajo. En estos se observó que la elección del servicio de elección de un asiento en particular fue mayor que en los casos de países con mayor puntaje. Esto podría indicar que existe alguna relación entre la satisfacción del cliente y ciertas preferencias específicas, como las de obtener equipaje adicional o comer en el vuelo,

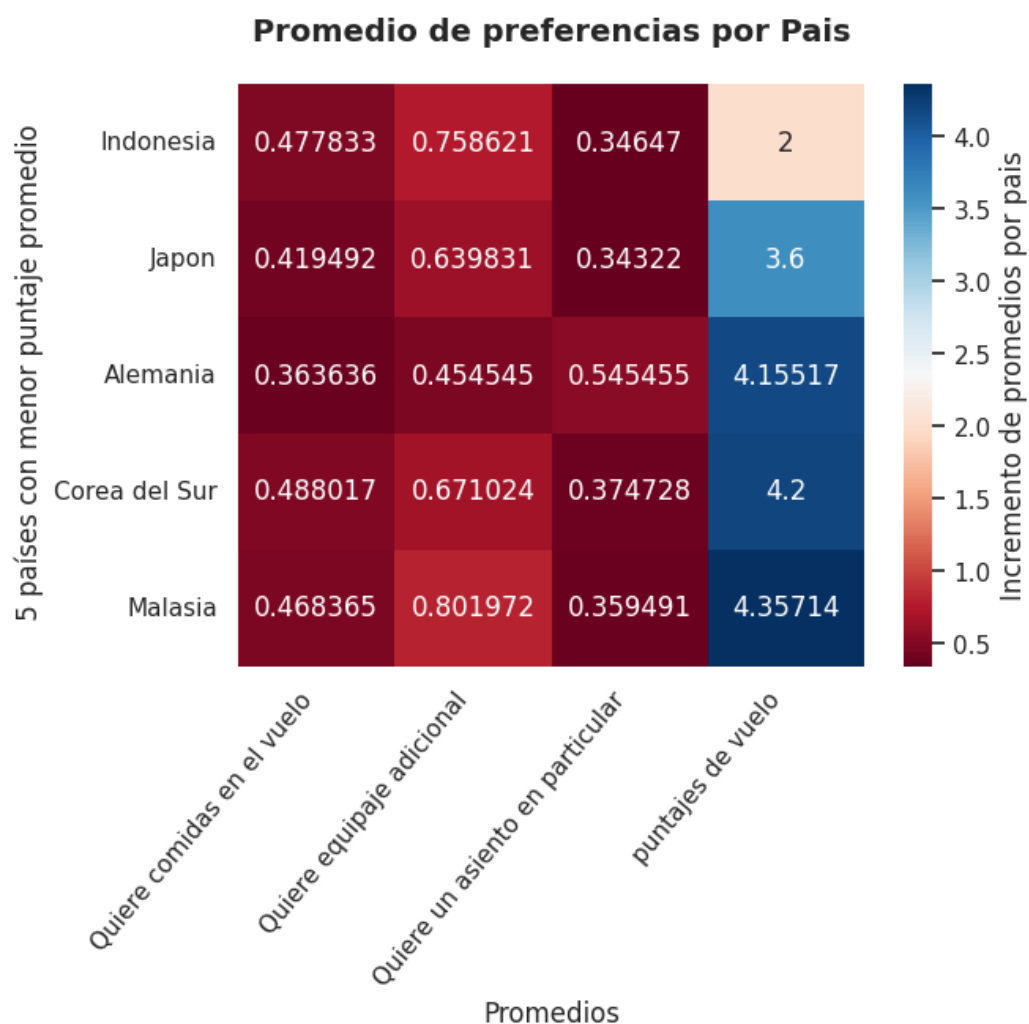


Figura 2 — Relación de promedio de preferencias y puntajes de los 5 países con mejor promedio.

## Conclusión

Una vez analizadas las posibles relaciones entre los puntajes y los servicios disponibles durante los vuelos pasamos a realizar un analisis geoespacial de la distribución de los puntajes y de los países con mayor encuestas para poder visualizar si existe alguna relación entre los países con mayores encuestas y algún rango de puntaje en especifico.

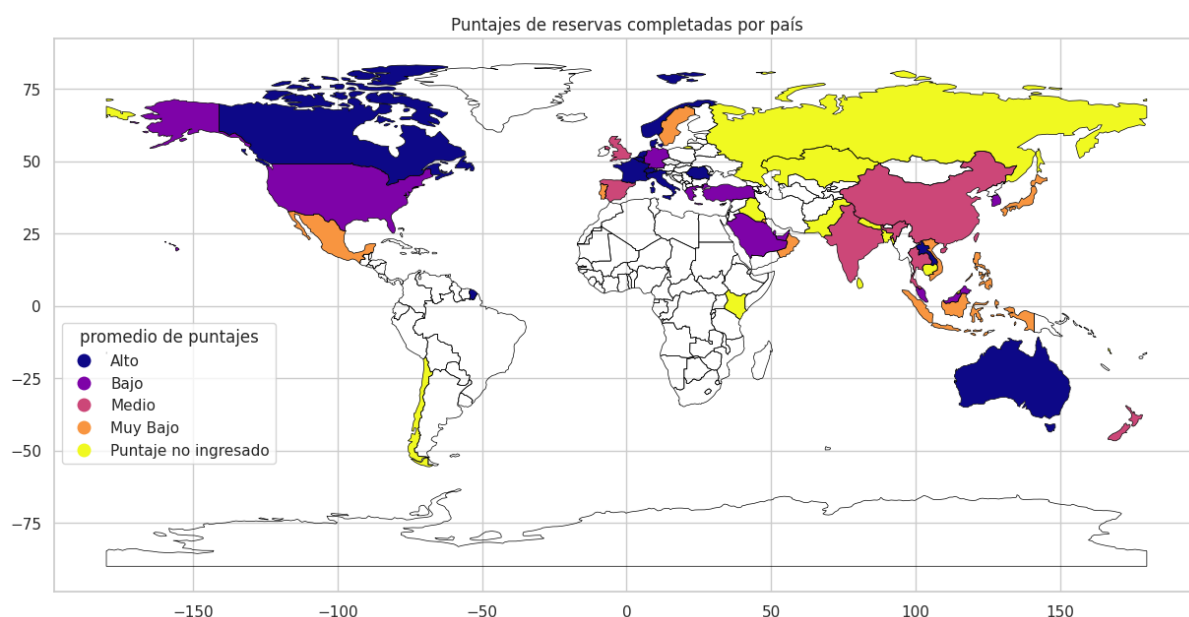


Figura 3- puntajes de los países que usaron la aerolínea

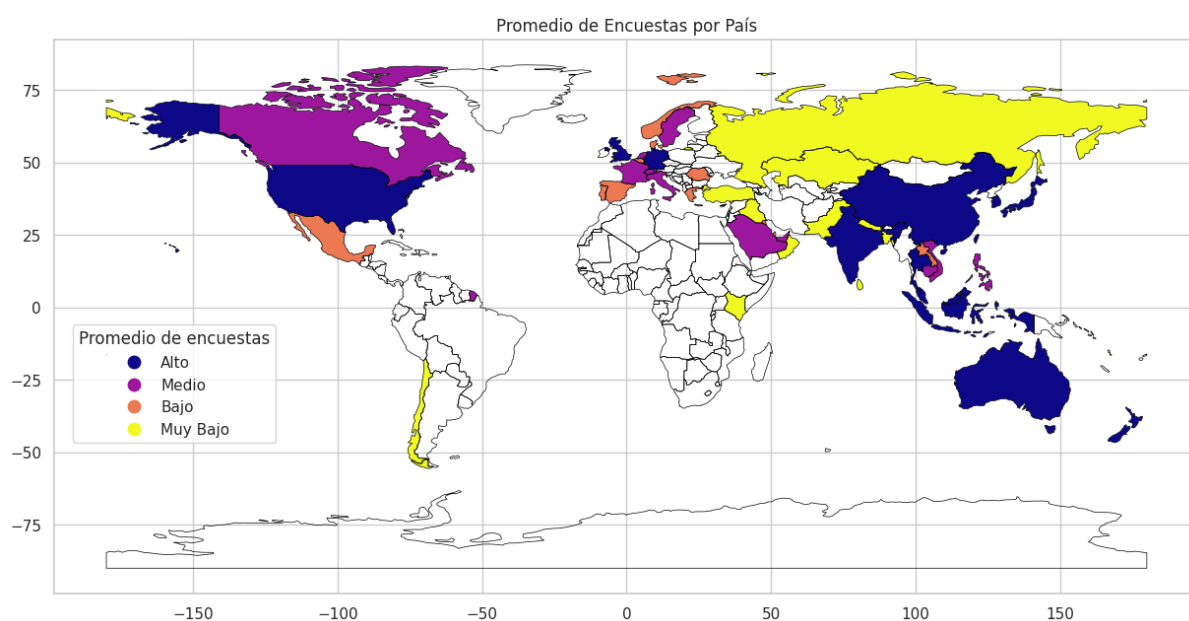


Figura 4- Promedio de encuestas por país.

La figura 3 y 4 nos permiten visualizar una clara relación entre los rangos de encuestas ingresadas y el puntaje en cada país. En primer lugar se puede ver que los puntajes no ingresados están relacionados casi en su totalidad con los países con un promedio de encuestas ingresadas muy bajo. La mayoría los países que tuvieron un promedio de puntaje medio y alto presentan un porcentaje de encuestas ingresadas superior o igual al 50% del total, es decir el rango medio y alto, a excepción de los países pertenecientes a la región de Asia oriental los

cuales a pesar de tener un alto porcentaje de encuestas realizadas su valor de puntajes oscila entre muy bajo y medio.

## Hipótesis 2:

***Cuáles son las rutas más elegidas? ¿De qué nacionalidad son los viajeros que las toman? ¿Existe alguna relación con los puntajes de estos países?***

### Archivos usados

Para el análisis exploratorio de datos de esta hipótesis utilizamos los archivos de customer\_airways\_data.csv y cleaned\_reviews.csv

### Desarrollo:

En este análisis, buscamos determinar si existe alguna relación entre las rutas más populares y la nacionalidad de los viajeros, así como si estas rutas están vinculadas a los países con mayor puntaje. Para obtener esta información, filtramos los datos para incluir solo las encuestas que indicaban reservas completas de los viajeros verificados, esta selección se hizo para asegurar un análisis preciso y representativo. Identificamos la ruta con más viajeros en cada país y seleccionamos las diez más populares con el fin de ver las rutas mas influyentes para la aerolínea y analizar la posible existencia de relación entre las rutas y las regiones con mayores puntajes, las cuales se pueden observar en la figura 3.

### Conclusión

De este análisis podemos concluir que, tal como se ve en la figura 5, la ruta que va de Nueva Zelanda a Malasia (AKLKUL) es la más elegida por los originarios de Malasia, Reino Unido y de Estados Unidos de América siendo esta la ruta la más elegida de la aerolínea por una notable diferencia. Si bien la ruta más elegida es la recién mencionada los viajeros que utilizan esta ruta no son todos de la misma región geográfica. La mayor cantidad de personas que eligen esta ruta son originarios de Malasia, en segundo lugar vienen los británicos, los cuales transitan por este tramo casi la mitad de veces que los malasios, y en último lugar los estadounidenses son los que también utilizan dicha ruta. La segunda ruta más elegida es la que va de Vietnam a Australia (SGNSYD) la cual la toman exclusivamente los originarios de Australia.

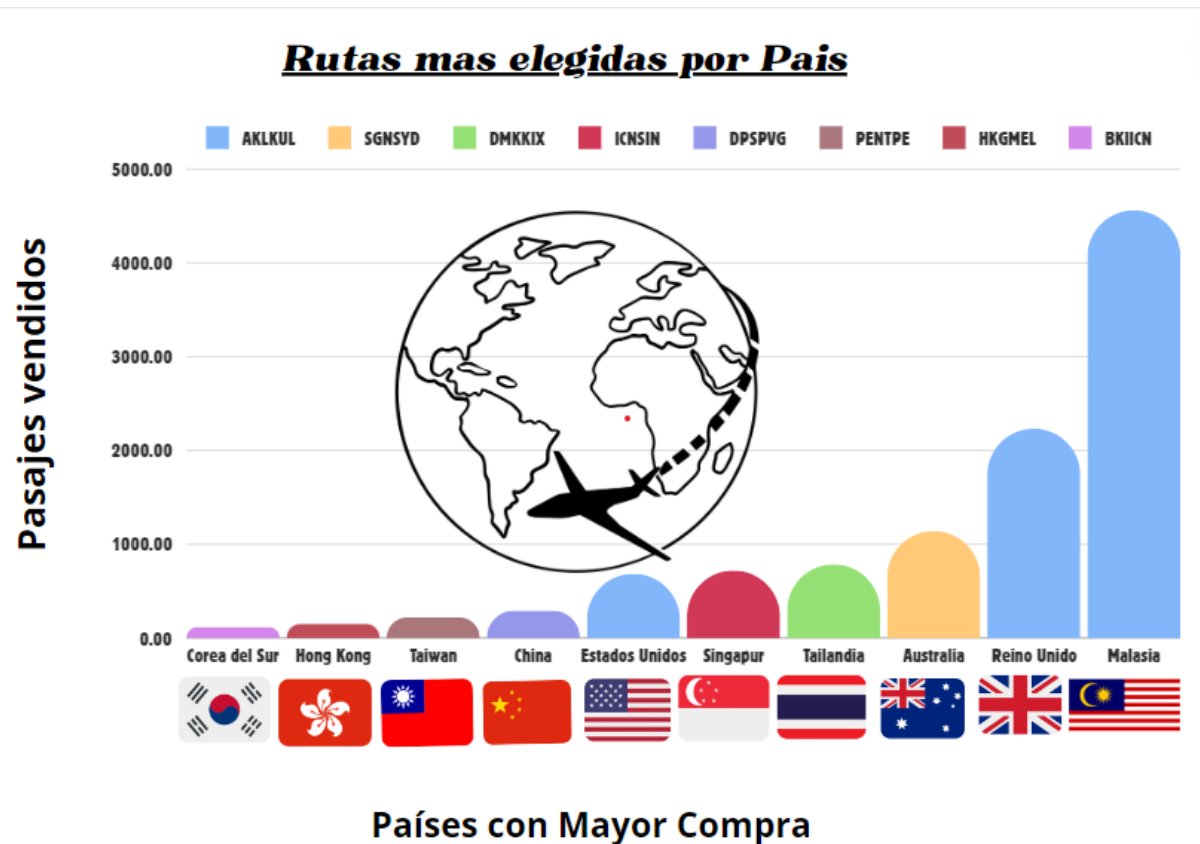


Figura 5— Relación del país de origen y la ruta que tomaron.

Esta información indica que la mayoría de las personas que utilizan la aerolínea viajan de Nueva Zelanda a Malasia, siendo originarias de este último. Esto permite ampliar la información proporcionada por la hipótesis 1 ya que aunque los viajeros originarios de Nueva Zelanda no expresaron disconformidad con la aerolínea como se observa en la figura 1, los viajeros malasios y estadounidenses, a pesar de ser de los viajeros mas frecuentes como muestra la figura 4, han mostrado su insatisfacción con sus viajes tal como se puede visualizar en la figura 3.

## Hipótesis 3:

**¿Cómo afecta el medio de compra de pasajes a los puntajes de los clientes en los diez países con el mayor volumen de compra?**

### Archivos usados

Para el análisis exploratorio de datos de esta hipótesis utilizamos los archivos de customer\_airways\_data.csv y cleaned\_reviews.csv

### Desarrolló:



Para empezar este analisis realizamos una selección de datos para evaluar solo aquellas encuestas que presentaban su reserva completa con el fin de quedarnos con los usuarios que utilizan, y mejor representan, la aerolínea en cuestión. Una vez hecho esto, y teniendo en cuenta las conclusiones de la hipótesis 2, decidimos explorar mas en profundidad los diez países que registraron el mayor volumen de compra de pasajes.

Para empezar este analisis realizamos una selección de datos para evaluar solo aquellas encuestas que presentaban su reserva completa con el fin de quedarnos con los usuarios que utilizan, y mejor representan, la aerolínea en cuestión. Una vez hecho esto, y teniendo en cuenta las conclusiones de la hipótesis 2, decidimos explorar mas en profundidad los diez países que registraron el mayor volumen de compra de pasajes. Seguimos utilizando estos países con el fin de comprender por que a pesar de continuar eligiendo la aerolínea solo uno de ellos, Australia, mostro un nivel alto de satisfacción.

En este nuevo analisis investigaremos si el proceso de compra del pasaje influye de alguna manera en la variación de puntajes, para poder realizar esto empezamos viendo cual fue el medio de compra mas usado en todo el mundo, esto nos permitió ver que la mayor parte de clientes de la aerolínea utiliza el medio de compra catalogado como "Internet" para reservar su pasaje y el resto utiliza la opción de "Celular" tal como se ve en la Figura 6.

## Ventas totales por medios de compra

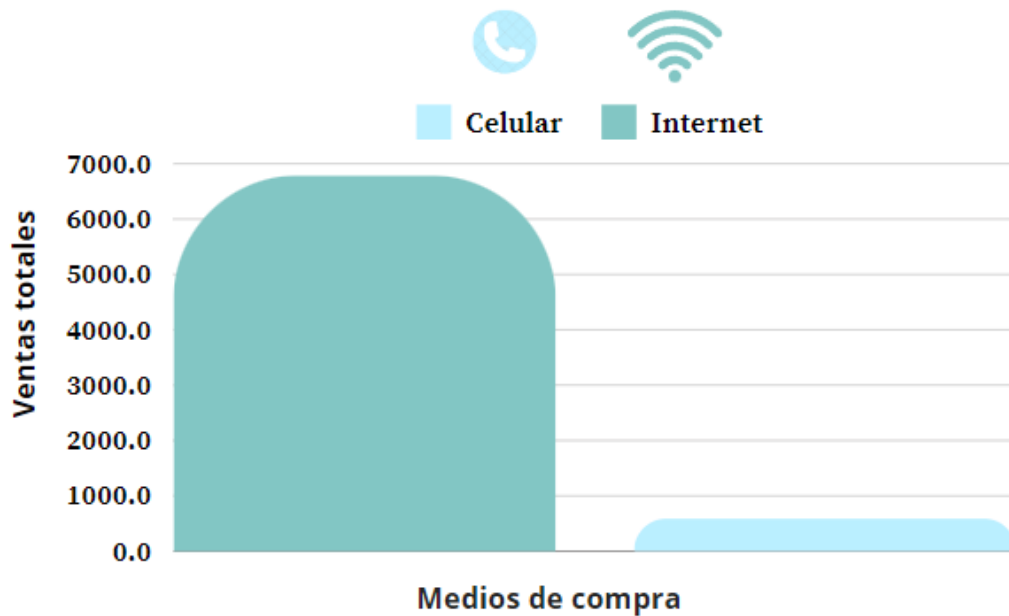


Figura 6 — Reservas totales según el medio de compra.

### **Conclusión:**

Teniendo estos datos nos interesa conocer la preferencia de los diez países mencionados anteriormente a la hora de elegir un medio para realizar una reserva en la aerolínea, y ver si esto influye en su distribución de pasajes. Este análisis nos va a permitir ver si existe algún tipo de relación entre un medio de compra en particular y los países con un promedio de satisfacción bajo. Tal como se ve en la figura 7 podemos ver que no hay una relación directa entre la elección de un medio de compra en específico y su relación en puntajes, sin embargo se descubrió que China, Taiwán y Malasia elijen realizar su reserva de pasaje mediante el celular. Esto es llamativo ya que, tal como se ve en la figura 6, es algo que no suele suceder. Otro dato a tener en cuenta es que Reino Unido no presenta ninguna reserva realizada a través del celular.

## Promedio de medios de compra por país



Figura 7 - Promedios de reserva de los 10 países que mas representativos por celular e internet.

## Hipótesis 4:

***¿Cómo afectan los días y horas de vuelo a la satisfacción de los viajeros en los países con un alto volumen de compra de pasajes?***

### Archivos usados

Para el análisis exploratorio de datos de esta hipótesis utilizamos los archivos de customer\_airways\_data.csv.

### Desarrollo:

Si bien la hipótesis 3 no nos presento datos que respalden el motivo de por que los países con mas compra de pasajes no presentan una alta tasa de satisfacción, si nos permitió descartar que no esta relacionado al medio de compra que utilizan dichos países. Teniendo en cuenta esto decidimos ver si la poca cantidad de puntajes altos presentes estaban relacionadas a la hora y día en la cual se realizaron estos vuelos. Tomando como parámetro que los días y las horas laborables resultan inconveniente para el viajero analizaremos el numero de pasajeros por día, tal como esta indicado en la figura 8, y por hora, como muestra la figura 9. Para poder realizar este analisis conservamos las encuestas que

completaron su reserva y sumamos todos los pasajeros de la aerolínea para luego dividirlos por día y hora de partida.

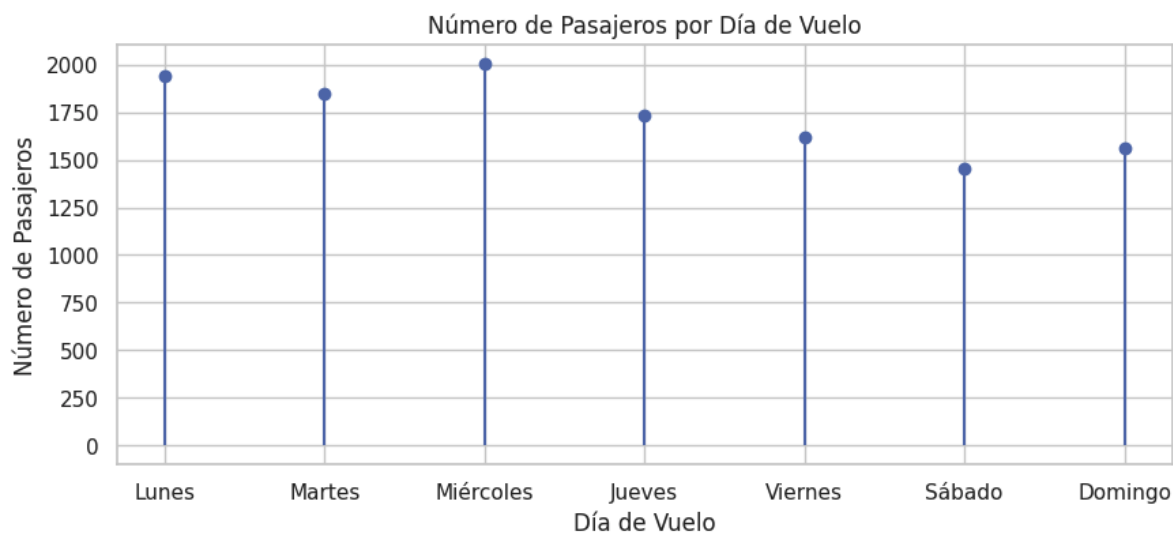


Figura 8 - Flujo de pasajeros durante las días de la semana.

Tal como se visualiza en la figura 8, el mayor flujo de pasajeros ocurre el día miércoles, en segundo lugar el lunes y en tercer lugar el martes. Esto indica que la mayor cantidad de pasajeros viajan en días laborables lo cual sugiere que los bajos puntajes pueden estar relacionados a esto. Para respaldar esta información analizaremos las horas, tomando el parámetro de horas 09:00 AM A 17:00 PM como horas laborables, y por lo tanto inconvenientes para el cliente.

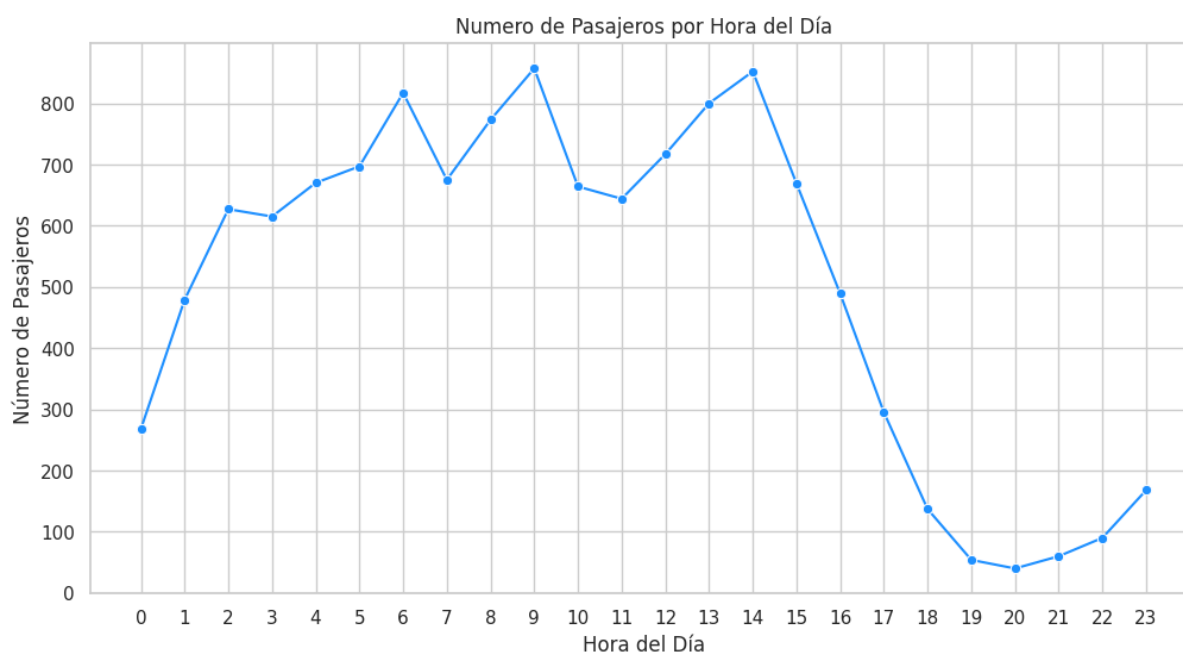


Figura 9 - Flujo de pasajeros durante las horas del día.

## Conclusión:

La siguiente figura, referenciada como figura 10, nos permite ver que los días hábiles tienen un mayor flujo de pasajeros que los no hábiles. Esto nos indica que los viajeros, especialmente aquellos originarios de países con un volumen alto de compra de pasajes y que constituyen la mayor parte de los clientes de la aerolínea, viajan los días laborables y en horas laborables lo cual podría significar un quiebre en su rutina, lo cual se ve reflejado en los puntajes.



Figura 10 - Cantidad de pasajeros durante la semana y las horas del día.

## Hipótesis 5:

**¿ El motivo del viaje, determinado por la época del año y el destino, en los puntajes de los clientes de la aerolínea?**

### Archivos usados

Para el análisis exploratorio de datos de esta hipótesis utilizamos los archivos de `customer_airways_data.csv`. y `cleaned_reviews.csv`

## Desarrollo:

La figura 8 nos permite ver que el día Miércoles fue el día de mayor concurrencia de pasajeros, sumado a esto, como se evidencia en la figura 10, los días laborables suelen ser mucho mas concurridos que los no laborables. Esto sugiere que los clientes de la aerolínea se ven perjudicados por estos horarios ya que son días de trabajo, aunque existe la posibilidad de que los viajes correspondan a periodos de vacaciones. Para poder comprobar esto dividimos el año en cuatro trimestres y elegimos los 5 destinos mas visitados. Si el viaje sucede en un trimestre correspondiente a la época con el clima mas agradable en el lugar de llegada se asumirá que no fue con fines laborables, en caso contrario se asumirá que si lo fue.

A fin de ver los valores mas representativos para la aerolínea se seleccionaron los 5 destinos mas visitados dentro del periodo de años, 2018-2023, dados por la

catedra. Esto nos permitirá ver si estos destinos están relacionados con los países con mayor cantidad de compra de pasajes y si el trimestre en el cual se viajó explica los bajos puntajes de dichos países.

## Conclusión

Tal como se puede visualizar en la figura 11, Australia es uno de los destinos mas elegidos,, principalmente a Sídney y a Perth. La elección de Australia como destino mas visitado esta respaldada por la figura 3, la cual lo cataloga como un país con un alto índice de satisfacción según los clientes. Un dato interesante a notar es que todos los países presentes en la figura 10 presentan un índice de promedios de puntaje que oscila entre "medio" y "alto".

Respecto a si los clientes suelen viajar en periodos vacacionales la información presentada indica que hay una mayor cantidad de viajeros, especialmente hacia Australia, durante el tercer trimestre del año en el cual es invierno en esa región, considerado como el mejor clima para conocer dicho país. Por lo cual se podría suponer que su alto puntaje esta relacionado a que la mayor parte de sus arribos sucedieron en esa época.

### Los 5 destinos mas visitados por trimestres

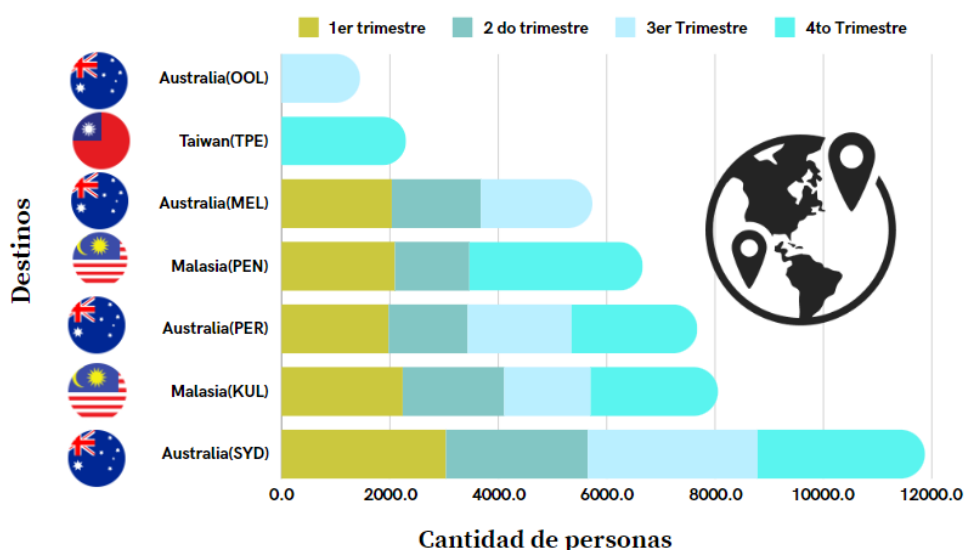


Figura 11 - Destinos mas elegidos según los distintos trimestres del año entre (2018-2023).

## Hipótesis 6:

**¿La verificación y la duración de los vuelos tiene algun impacto en la satisfacción de los clientes en los países con mayor compra de pasajes?,**

## Archivos usados

Para el análisis exploratorio de datos de esta hipótesis utilizamos los archivos de customer\_airways\_data.csv. y cleaned\_reviews.csv

## Desarrollo:

Si bien uno de los motivos de la falta de altos puntajes en los países con mayor compra es la época del año en la cual se viaja no podemos suponer que es el único. Por este motivo analizaremos si la la cantidad de viajeros verificados y sus respectivas horas de vuelo nos brinda mas información.

Para empezar este analisis realizamos un promedio de los puntajes en los usuarios verificados y los no verificados en los diez países con mayor compra de pasajes tal como se ve en la figura 12 y 13 respectivamente. Estos graficos nos permiten ver que los clientes verificados y no verificados presentan, usualmente, puntajes entre 4 y 5.5. Un dato a resaltar es que los pasajeros verificados muestran un rango de puntajes mas compacto, entre 2 a 5, mientras que los no verificados presentan todo tipo de puntajes, desde un 1 a Malasia a un 9 a Taiwán. Esto podría indicar que los pasajeros verificados suelen ser mas críticos y constructivos a la hora de dar una respuesta que los que no lo están.

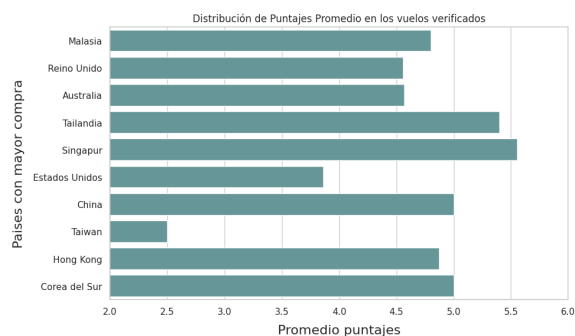


Figura 12 - Relación entre los puntajes promedios de los usuarios verificados y los países con mayor compra

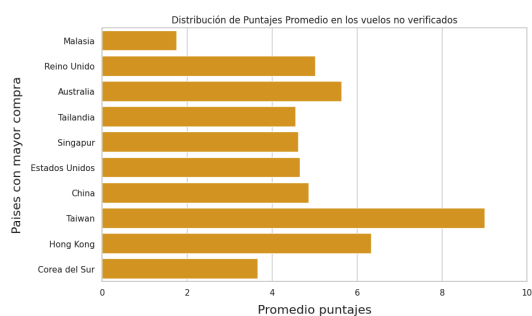


Figura 13 - Relación entre los puntajes promedios de los usuarios no verificados y los países con mayor compra

## Conclusión

Una vez hecho esto categorizamos las horas de vuelo en cuatro categorías distintas, las cuales van desde muy bajo hasta alto. Estas categorías se toman a partir de los cuartiles de un promedio de las horas de vuelos presentes en la figura 9, siendo muy bajo todo lo menor e igual al primer cuartil y alto todo lo superior al tercero.

Relación entre horas de vuelo y puntaje en los vuelos verificados de los países con mayor compra

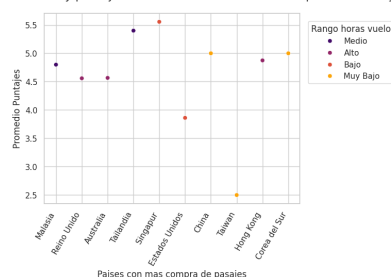


Figura 14 - Relación entre las horas y puntajes de los vuelos verificados

Relación entre horas de vuelo y puntaje en los vuelos no verificados de los países con mayor compra

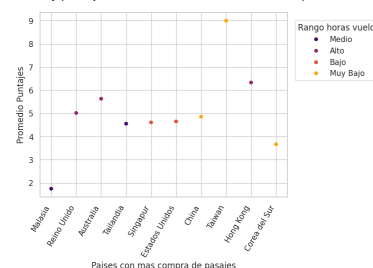


Figura 15— Relación de los países con mayor compra con vuelos no verificados con respecto al promedio puntaje y el rango de horas de vuelo.

Tal como podemos ver en la figura 14, los puntajes promedio están dispersos entre los rangos de horas de vuelo, es importante notar que los rangos muy bajo y bajo incluyen los puntajes más altos, China y Singapur, lo que podría sugerir que vuelos con menos horas de vuelo están relacionados con una alta satisfacción. Sin embargo también hay un puntaje, Taiwán con una calificación de dos y medio, el cual no permite afirmar esto, los países en el rango "Alto" tienen puntajes de aproximadamente cinco, esto podría indicar que, a pesar de la mayor duración de los vuelos, la satisfacción de los usuarios no decae. En la figura 15 sucede algo bastante similar, con la salvedad que los puntajes presentes en los rangos muy bajo y bajo presentan una polarización de los puntajes mucho mas pronunciada variando desde 3 a 9.

Teniendo en cuenta toda la información presentada este analisis nos indica que a pesar de esperarse que los vuelos más cortos siempre tengan mejores evaluaciones los datos muestran que un alto puntaje puede estar relacionado a esto pero que no es exclusivo a esta condición.