

Dibujo de un animal

Descripción generada automáticamente con confianza baja

Proyecto N° 1

ANÁLISIS PREDICCIONES DE VUELO

Integrantes:

López, Valeria

Peña, Zunida

Profesor:

Tobar, José Miguel

Cátedra:

Data Analytics

**ÍNDICE**

[**Introducción 3**](#_Toc85641852)

1. [Descripción del Negocio 4](#_Toc85641852)
2. [Descripción del Dataset desde el punto de vista de negocio y técnico 7](#_Toc85641852)
3. [Supuestos Utilizados 9](#_Toc85641852)
4. [Limpieza o transformaciones aplicadas 10](#_Toc85641852)
5. Features Relevantes 12
6. Análisis Exploratorio - EDA 13
7. Preguntas de Negocio 17
8. Problemas y Conclusiones 22

[**Referencias 23**](#_Toc85641852)

[**Anexos 24**](#_Toc85641852)

**INTRODUCCIÓN**

*«Nuestro destino de viaje nunca es un lugar, sino una nueva forma de ver las cosas».*

*Henry Miller*

El transporte aéreo ha ido evolucionando constantemente, experimentando un gran crecimiento ya sea gracias al aumento tanto de las economías como el de las personas, un aumento en las ofertas de vuelos y la existencia de tickets aéreos a bajo precio, “ajustándose al bolsillo” de los consumidores.

No obstante, el acelerado desarrollo de la industria aérea ha conllevado a un aumento en los números de vuelos disponibles, congestionando los aeropuertos. Asimismo, a lo anteriormente indicado se le adiciona el aumento de las medidas de seguridad que se aplican en los aeropuertos y en los propios aviones, generando que, en conjunto, existan retrasos en los vuelos, o bien, la cancelación de los mismos.

Las consecuencias tanto de los retrasos como las cancelaciones de los vuelos se traducen en pérdidas en dinero y tiempo para las compañías que pertenecen a la industria del transporte aéreo, pero esto también ha perjudicado a los consumidores, específicamente en relación con la calidad del servicio que han recibido.

Ante la situación descrita es que se realiza un análisis al *dataset* relacionado a las predicciones de vuelo para el mes de enero en Estados Unidos durante el año 2020, el cual nos permitirá predecir las razones o factores de los retrasos de los vuelos en el aeropuerto de destino.

1. **DESCRIPCIÓN DEL NEGOCIO**

Antes de definir el modelo de negocios del problema en cuestión, primero es importante definir el concepto de “modelo de negocios”. Este concepto posee múltiples definiciones, pero en esta situación se hace referencia a una definición más simplificada de éste, la cual lo define como “ una herramienta previa al plan de negocio, cuyo objetivo es permitir conocer con claridad el tipo de negocio que se va a crear e introducir en el mercado, a quién va dirigido, cómo se va a vender y cómo se van a conseguir los ingresos.” (Peiró, 2017).

Osterwalder, Pignes y Tucci (citado por Btesh, F., 2010) presentan los elementos principales que componen los modelos de negocios en general, y es en base a estos que se determinará el cómo crear un modelo de negocio para aquellas compañías que forman parte del sector del transporte aéreo, ya que el problema identificado está relacionado a este tema.

Por lo tanto, los elementos claves que debe comprender el modelo de negocio, en este caso, de una aerolínea son: el nivel de servicio, la estrategia del precio (único o variable), costes (fijos, variables del vuelo y variables de los clientes), servicio al cliente (‘timing’ y retrasos), marketing estratégico (nivel de posicionamiento genérico y particular, nivel del valor) y la estrategia de vuelo (‘hub-and-spoke’ y vuelos directos). (Osterwalder, 1996; Hirschler, 2004 y Rego, 2007, citado por Btesh, F., 2010). Los factores anteriormente señalados se detallan en la siguiente tabla:

Tabla 1: Factores esenciales de un modelo de negocio para empresas que operan en el sector del transporte aéreo.

|  |  |
| --- | --- |
| **FACTOR** | **DESCRIPCIÓN** |
| Nivel de Servicio | La organización debe tener determinado a que sujeto se pretende dar valor, esto considerando la visión de la compañía. |
| Estrategia de Precio | * **Two-Tired Pricing:** La empresa ofrece al cliente un precio único de pasaje. Para atraerlos ocasionalmente ofrecen ciertos números de asientos a precios económicos. * **Peak-Load Pricing:** Los niveles de demanda cambian según la temporada alta o baja y horarios. Esta estrategia propone la adaptación de precios en relación con el momento del viaje. |
| Costos | * **Fijos**: Gatos mensuales de la empresa los cuales debe afrontar independientemente del nivel de actividad que esta tenga. * **Variables vinculados a vuelos:** Relacionados directamente con la actividad de volar, y no se tendrían si la compañía no realizara viajes. (gastos en combustible, pilotos y tripulación, y tasas por permanencia en aeropuertos) * **Variables vinculados con el cliente:** Relacionados con la cantidad de personas que se trasladan. (gastos de servicios que se brindan dentro del avión y gastos de reserva) |
| Servicio al Cliente | * **Timing:** Grado de eficiencia que la empresa alcanza en relación con los factores de puntualidad y demoras entre vuelos y conexiones. * **Tiempo de demora de un vuelo:** Tiempo que tarda un viaje programado entre el sitio de origen y el de destino. |
| Marketing Estratégico | * **Nivel de posicionamiento genérico:** Lugar que la empresa desea ocupar en la mente del consumidor al cual están destinados sus servicios. * **Nivel de posicionamiento particular:** La empresa analiza el cómo quiere ser reconocida por sus clientes objetivos. * **Nivel de valor:** Se considera el valor agregado que la empresa le brinda a sus consumidores y el valor que este último percibe sobre la organización. |
| Estrategia de Vuelo | * **“Hub-and-spoke”:** Aeropuerto pivote, que recibe un gran número de vuelos para redistribuirlos a sus destinos finales. * **Vuelos Directos:** Trasladar a los pasajeros de manera directa a su destino final sin utilizar aeropuertos de pivote. |

Fuente: ¿Qué distingue a una aerolínea de bajo costo en un mercado emergente, de una aerolínea de bajo costo en uno desarrollado? (Btesh, 2010)

Es a través de estos factores que se genera la competencia entre empresas ligadas al sector de transporte aéreo ya que éstas toman en consideración ciertos elementos más que otros, lo que genera que se diferencien entre ellas (Peiró, 2017)

1. **Tipos de Servicios de aerolíneas**

* *Aerolíneas tradicionales*

Las aerolíneas tradicionales se caracterizan por ofrecer una serie de servicios a los pasajeros, los cuales ya vienen incluidos en el precio del pasaje. Este tipo de aerolíneas ofrecen a sus clientes rutas de larga distancia, con escalas, transoceánicas, etc. En la mayoría de los países, este tipo de aerolíneas son de carácter mixto, es decir, tiene una parte accionaria privada y otra remanente estatal. Al ser “líneas de bandera”, su objetivo principal es unir a un país con el resto del mundo, por lo que la mayoría de los viajes son de carácter internacional.

Este tipo de aerolíneas contempla asientos para tres tipos de clase: turista, *business* o ejecutiva, y primera clase, siendo esta ultima la de mayor prestigio. Además, posee una estrategia de vuelo *hub-and-spoke*, es decir, permite a sus pasajeros a viajar a cualquier parte del mundo y aterrizar en los principales aeropuertos.

* *Aerolínea regional*

Este tipo de aerolíneas centran sus esfuerzos en segmentos más pequeños. Lo anterior conlleva a que comprendan un número limitado de rutas y que satisfagan a un grupo especifico de consumidores, es decir, proporciona transporte a comunidades que no cuentan con una gran demanda para atraer una aerolínea tradicional.

Cabe destacar que, en algunas situaciones, las grandes empresas aéreas realizan convenios con las empresas aéreas regionales. Es decir, contratan los servicios de aerolíneas regionales con el propósito de cubrir rutas de menor volumen, logrando conectar pequeñas ciudades con los principales aeropuertos de las aerolíneas tradicionales.

* *Aerolínea Segmentada / Destinos Específicos*

Al igual que las aerolíneas regionales, este tipo de aerolíneas centran sus esfuerzos y actividades en un grupo especifico de personas. Esta se considera de destino especifico ya que su principal interés es abarcar, tal como se indica anteriormente, a consumidores particulares y en regiones geográficas específicas, es decir, desea satisfacer las necesidades de un grupo en particular.

* Aerolíneas de Bajo Costo o *Low-Cost Carriers*

Este tipo de aerolíneas una ofrece una tarifa más económica a cambio de eliminar varios de los servicios que reciben los pasajeros de aerolíneas tradicionales, o bien cobran estos servicios de manera adicional. A diferencia de las aerolíneas anteriores, se prioriza la venta de pasajes mediante vía online y no existe la asignación de asientos, es decir, estos se van ocupando en base al orden de llegada de los pasajeros. La idea de estas aerolíneas es entregar un servicio simple y dar valor al consumidor. Es por lo anterior que este tipo de compañías deben trabajar en hacer lo simple de manera eficiente y lograr competitividad.

En Chile, el ticket de avión más barato se encuentra bordeando los $10.000, y el más caro bordea los $365.000[[1]](#footnote-1).

En relación con los destinos más populares, se pueden encontrar los siguientes: Santiago (23% del total de búsquedas a Chile), Puerto Montt (11%), Punta Arenas (10%), La Serena (9%), Calama (7%) e Iquique (7%).[[2]](#footnote-2)

1. **DESCRIPCIÓN DEL DATA SET DESDE EL PUNTO DE VISTA DE NEGOCIO Y TÉCNICO**

El dataset a analizar para esta investigación es ***Fligh Prediction,*** el cual contiene una descripción de todos los vuelos realizados en los Estados Unidos, siendo estos más de 600.000.

Las variables que componen el dataset son las que se indican a continuación:

1. ***DAY\_OF\_MONTH:*** Variable cuantitativa que indica el día del mes, siendo estos del 1 al 31, dependiendo del mes.
2. ***DAY\_OF\_WEEK***: Variable cuantitativa que indica el día de la semana, tomando valores del 1 al 7 (Domingo = 1 y Sábado = 7)
3. ***OP\_UNIQUE\_CARRIER***: Variable cualitativa que indica el código del transportista. Cuando el mismo código ha sido utilizado por varios transportistas, se utiliza un sufijo numérico para los usuarios anteriores, por ejemplo, PA, PA(1), PA(2).
4. ***OP\_CARRIER\_AIRLINE\_ID***: Variable cuantitativa que señala el número de identificación asignado por el Departamento de Transporte de EE.UU. para identificar una línea aérea única (compañía aérea). Una aerolínea (transportista) única se define como aquella que tiene y reporta bajo el mismo certificado del DOT, independientemente de su código, nombre o compañía/corporación.
5. ***OP\_CARRIER***: Variable cualitativa que indica el código asignado por la IATA y utilizado habitualmente para identificar a una compañía aérea. Como el mismo código puede haber sido asignado a diferentes transportistas a lo largo del tiempo, el código no siempre es único.
6. ***TAIL\_NUM***: Variable cualitativa que indica el número de Cola
7. ***OP\_CARRIER\_FL\_NUM:*** Variable cuantitativa que indica el número de vuelo.
8. ***ORIGIN\_AIRPORT\_ID***: Variable cuantitativa que señala el Aeropuerto de origen, ID de aeropuerto. Número de identificación asignado por el Departamento de Transporte de Estados Unidos para identificar un único aeropuerto.
9. ***ORIGIN\_AIRPORT\_SEQ\_ID***: Variable cuantitativa que indica el Aeropuerto de origen, ID de secuencia del aeropuerto. Número de identificación asignado por el Departamento de Transporte de los Estados Unidos para identificar un único aeropuerto en un momento determinado.
10. ***ORIGIN***: Variable cualitativa que señala el Aeropuerto de Origen
11. ***DEST\_AIRPORT\_ID***: Variable cuantitativa que indica el Aeropuerto de destino, ID de aeropuerto. Un número de identificación asignado por el Departamento de Transporte de Estados Unidos para identificar un único aeropuerto.
12. ***DEST\_AIRPORT\_SEQ\_ID***: Variable cuantitativa que indica el Aeropuerto de destino, ID de secuencia de aeropuerto. Un número de identificación asignado por el Departamento de Transporte de los Estados Unidos para identificar un aeropuerto único en un momento determinado.
13. ***DEST***: Variable cualitativa que señala el Aeropuerto de Destino
14. ***DEP\_TIME***: Variable cualitativa que señala la Hora de salida real (hora local: hhmm)
15. ***DEP\_DEL15***: Variable categórica del tipo numérico que es un Indicador de retraso en la salida, 15 minutos o más (1=Sí, 0=No)
16. ***DEP\_TIME\_BLK***: Variable cualitativa que indica el Bloque horario de salida, intervalos horarios
17. ***ARR\_TIME***: Variable cualitativa que indica la Hora de llegada real (hora local: hhmm)
18. ***ARR\_DEL15***: Variable categórica del tipo numérico que representa un Indicador de retraso de llegada, 15 minutos o más (1=Sí, 0=No)
19. ***CANCELLED***: Variable cuantitativa que es un Indicador de vuelo cancelado (1=Sí, 0=No)
20. ***DIVERTED***: Variable categórica del tipo numérico que representa a un Indicador de vuelo desviado (1=Sí, 0=No)
21. ***DISTANCED***: Variable cuantitativa que señala la Distancia entre aeropuertos (millas)
22. ***FLIGHT\_TIME***: Variable construida a partir de la diferencia entre el tiempo exacto de arribo (ARR\_TIME) y el tiempo exacto de la salida del avión desde el origen (DEP\_TIME). Contiene datos del tipo cuantitativo.
23. ***FLIGHT\_ROUT***: Variable construida a partir de la unión de la variable ORIGIN (aeropuerto de origen) y DEST (aeropuerto de destino). Contiene datos cualitativos y señala la ruta del vuelo.
24. ***SPEED***: Variable construida a partir de la división entre la distancia (DISTANCE) y el tiempo de vuelo (FLIGHT\_TIME). Contiene datos cuantitativos y señala la velocidad a la que se dirigía el avión.
25. **SUPUESTOS UTILIZADOS**

Para desarrollar este trabajo, se indican a continuación ciertos parámetros que son excluidos de este estudio:

* *Mercado antes de pandemia:* Debido a las condiciones sanitarias que actualmente se está viviendo en todo el mundo, el mercado de las aerolíneas ha sido uno de los más afectados, es decir, gracias a las restricciones impuestas por cada gobierno, la movilidad de las personas se ha visto reducida, generando una disminución en la compra de tickets aéreos. No obstante, el “efecto pandemia” no se considera al realizar la investigación, ya que, durante el mes de enero del 2020, en Estados Unidos aún no existían restricciones relacionadas a esta situación.

Cabe recalcar que el *dataset* está compuesto por observaciones para el mes de enero del 2020 y es de origen estadounidense.

* *Datos perdidos:* Al encontrarse con datos nulos (NA) que seguramente se perdieron o no se midieron, supondremos que todas maneras la falta de ellos no afectará nuestro análisis y se podrá utilizar el *dataset* para los objetivos impuestos.
* *Días de la Semana*: Para este trabajo que supondrá que el primer día de la semana (1) será el día Domingo en los Estados Unidos, esto de acuerdo a la información que se buscó en la web.

1. **LIMPIEZA O TRANSFORMACIONES APLICADAS**

Al importar el dataset *“flight-prediction”* a Rstudio, esta contenía 22 variables y 607.346 observaciones, se procedió a analizar la existencia de valores perdidos o vacíos por cada variable que compone la base de datos. Como resultado se obtuvo que, para las variables *DEP\_TIME*, *DEP\_DEL15*, *ARR\_TIME*, *ARR\_DEL15,* y una última columna compuesta por “NA”, existían valores vacíos. Por lo tanto, se procedió a eliminarlos, trabajando, de esta forma, con 599.268 observaciones y 21 variables.

Asimismo, se agregó una nueva columna representando la variable *FLIGHT\_TIME*. Esta cuenta con la duración exacta del vuelo desde el aeropuerto de origen hasta el aeropuerto de destino. La variable “tiempo de vuelo” se incluye con el propósito de analizar cuáles son los trayectos que demandan mayor cantidad de horas y viceversa.

También, se agregó otra columna que representa la variable *FLIGHT\_ROUT,* la cual nos indicala ruta del vuelo determinada por el aeropuerto de origen y el aeropuerto de destino. Esta variable se implementa con el propósito de analizar y determinar cuáles son las rutas más populares y demandadas.

Por consiguiente, se crea una columna que contempla la variable *SPEED.* Esta señala la velocidad a la que se dirigía el vuelo hacia su destino, su propósito es analizar cuan rápido se desplazan los vuelos.

Finalmente se comenzó a trabajar con un *dataset* compuesto por 24 variables y 599.268 observaciones.

1. **FEATURES RELEVANTES**

Como se indica anteriormente, el *dataset* está compuesto por 23 variables, de las cuales algunas tienen más relevancia que otras. Las variables que más destacan son: OP\_CARRIER, ORIGIN, DEST, DEP\_DEL15, ARR\_DEL15, CANCELLED, DISTANCE, FLIGHT\_TIME, FLIGHT\_ROUT y TAIL\_NUM:

* **OP\_CARRIER.** Esta variable indica la compañía aérea que realiza el vuelo. Es importante debido a que su análisis nos muestra cuales son las aerolíneas más demandadas durante el mes de enero de 2020 en los Estados Unidos.
* **ORIGIN**. Esta variable indica el aeropuerto de origen del vuelo. Es importante puesto que al analizarla nos indica cuales son los aeropuertos con más vuelos de salida, además nos ayuda a crear la variable *FLIGHT\_ROUT* que determina la ruta del vuelo.
* **DEST.** Esta variable indica el aeropuerto de destino del vuelo. Es importante puesto que al observarla nos señala cuales son los aeropuertos con más vuelos de llegada, además nos ayuda a crear la variable *FLIGHT\_ROUT* que determina la ruta del vuelo.
* **DEP\_DEL15.** Esta variable indica si existe retraso de 15 minutos en la salida del vuelo. Es importante porque nos muestra que aerolíneas son puntuales al momento de despegar.
* **ARR\_DEL15**. Esta variable indica si existe retraso de 15 minutos en el arribo del vuelo en el aeropuerto de destino. Es importante porque nos muestra que aerolíneas son puntuales al momento de aterrizar.
* **CANCELLED**. Esta variable representa la cantidad de vuelos que fueron cancelados durante el mes de enero del 2020 en Estados Unidos. Es relevante puesto nos da a conocer las aerolíneas que tienen problemas con sus vuelos.
* **DISTANCE.** Esta variable indica la distancia (millas) entre el aeropuerto de origen con el aeropuerto de destino. Es relevante puesto que nos indicará la cantidad de vuelos que recorren una mayor distancia.
* **FLIGHT\_TIME.** Esta variable indica el tiempo total del vuelo desde el despegue del avión en el aeropuerto de origen hasta el aterrizaje en el aeropuerto de destino.
* **FLIGHT\_ROUT.** Esta variables nos indica la ruta del vuelo señalando el aeropuerto de origen y el aeropuerto de destino.
* **TAIL\_NUM.** Esta variable identifica a los aviones que realizan ciertos recorridos, lo cual puede ayudar a determinar cuáles de ellos han realizado más viajes, entre otros.

1. **ANÁLISIS EXPLORATORIO - EDA**

A partir de la data completa y ya realizada su limpieza, se puede destacar la siguiente información para los Estados Unidos durante el mes de enero:

* Las variables del tipo alfanuméricas tales como: *OP\_UNIQUE\_CARRIER,* *OP\_CARRIER\_AIRLINE\_ID, OP\_CARRIER\_FL\_NUM, ORIGIN\_AIRPORT\_ID, ORIGIN\_AIRPORT\_SEQ\_ID, DEST\_AIRPORT\_ID, DEST\_AIRPORT\_SEQ\_ID* no resultan relevantes para este estudio ya que estas variables solo proporcionan identificación ya sea de los vuelos en sí, como de los aeropuertos de origen y de destino, por lo que estas no nos ayudarán a responder a las preguntas que con posterioridad se planteen.
* Al comenzar la evaluación de la base de datos, se determina que los días de la semana que cuentan con una mayor cantidad de vuelos son los días: martes, miércoles y jueves. (véase Anexo 1, Ilustración 1).
* De un total de 351 Aeropuertos de origen en los Estados Unidos, los tres que cuentan con más viajes de salida son: Aeropuerto Internacional Hartsfield-Jackson en Atlanta (ATL), Aeropuerto Internacional O’Hare en Chicago (ORD) y Aeropuerto Internacional de Dallas-Fort Worth en Dallas (DFW). (véase Anexo 2, Ilustración 2). En cuanto al aeropuerto que cuenta con menos vuelos de origen, este es el aeropuerto de Cabo Cod (COD), el cual solo contó con un solo vuelo de origen. La cantidad de vuelos por cada aeropuerto se indican a continuación:

*Tabla 2: Ranking de Aeropuertos con más vuelos de origen*

|  |  |
| --- | --- |
| Aeropuerto | Cantidad de Vuelos |
| ATL | 32.024 |
| ORD | 24.766 |
| DFW | 23.818 |

Fuente: Elaboración Propia

* Con respecto a los Aeropuertos de destino, existen 350 disponibles, siendo los más populares: Aeropuerto Internacional Hartsfield-Jackson en Atlanta (ATL), Aeropuerto Internacional O’Hare en Chicago (ORD) y Aeropuerto Internacional de Dallas-Fort Worth en Dallas (DFW). (véase Anexo 3, Ilustración 3). En cuanto al aeropuerto que cuenta con menos vuelos de llegada, este es el aeropuerto de Quincy, el cual contó con solo 9 vuelos de destino. La cantidad de vuelos por cada aeropuerto se indican a continuación:

Tabla 3: Ranking de Aeropuertos con más vuelos de llegada

|  |  |
| --- | --- |
| Aeropuerto | Cantidad de Vuelos |
| ATL | 32.062 |
| ORD | 24.720 |
| DFW | 23.742 |

Fuente: Elaboración Propia

* En consideración a los horarios, el bloque en el cual generalmente existen más vuelos son los que salen antes de las 18 horas, mientras que se registran menos vuelos entre las 23:00 – 23:59 horas. A continuación, se presentan los bloques de horarios que registran mayor cantidad de vuelos:

*Tabla 4: Ranking de Bloque de horario con más vuelos de origen*

|  |  |
| --- | --- |
| Bloque | Cantidad de Vuelos |
| 06:00 – 06:59 | 43.742 |
| 07:00 – 07:59 | 40.284 |
| 17:00 – 17:59 | 39.192 |
| 08:00 – 08:59 | 39.021 |
| 11:00 – 11:59 | 37.700 |
| 12:00 – 12:59 | 37.622 |

Fuente: Elaboración Propia

* Con respecto a la cantidad de vuelos diarios durante el mes de enero, existe una gran concentración que indica que diariamente se realizan más de 17.000 vuelos cada día (véase Anexo 4, Ilustración 4), siendo los días 06, 09 y 31 de enero en los cuales se realizaron más de 20.500 viajes. (véase Anexo 5, Ilustración 5).
* Se generaron más de 5700 rutas que se realizaron en Estados Unidos el mes de enero, siendo las más populares: desde el Aeropuerto Internacional de los Ángeles hacia el Aeropuerto Internacional de San Francisco (LAX-SFO), desde el Aeropuerto de San Francisco hacia el Aeropuerto de Los Ángeles (SFO-LAX) y desde el Aeropuerto Internacional McCarran hacia el Aeropuerto Internacional de Los Ángeles (LAS-LAX). (véase Anexo 6, Ilustración 6).
* Para el mes de enero en los Estados Unidos, se registraron retrasos durante el despegue y el arribo del vuelo, siendo los retrasos de llegada mayores en cantidad que los retrasos de salida. A continuación, se presenta la cantidad de retrasos que se generaron durante los vuelos:

Tabla 5: Número de vuelos que tuvieron retraso

|  |  |  |
| --- | --- | --- |
|  | En la Salida | En el Arribo |
| Retrasos | 81.447 | 82.285 |
| % del total | 15,73 | 15,92 |

Fuente: Elaboración Propia

* En relación con la distancia la mayoría de los vuelos recorren entre 214 a 733 millas, lo cual se presenta con más detalle a continuación:

Tabla 6: Ranking de distancias mayormente recorridas por los vuelos

|  |  |
| --- | --- |
| Distancia (millas) | Cantidad de Vuelos |
| 337 | 3.826 |
| 296 | 2.823 |
| 399 | 2.763 |
| 733 | 2.447 |
| 214 | 2.363 |
| 404 | 2.356 |

Fuente: Elaboración Propia

* Existen 17 compañías aéreas que realizaron viajes dentro de los Estados Unidos en el mes de enero de 2020, siendo las más demandadas por los usurarios: Southwest Airlines (WN), Delta Air Lines (DL), American Airlines (AA). (véase Anexo 7, Ilustración 7). A continuación, se presenta con más detalle las compañías mayormente demandadas:

Tabla 7: Ranking de compañías aéreas con más vuelos

|  |  |
| --- | --- |
| Aerolínea | Cantidad de Vuelos |
| WN | 107.708 |
| DL | 79.928 |
| AA | 75.472 |

Fuente: Elaboración Propia

* Ninguno de los vuelos fue cancelado durante enero del 2020.
* Se identificaron 5443 aviones, de los cuales 8 de ellos han realizado más de 290 viajes. No obstante, los tres aviones que han realizado más recorridos se identifican de la siguiente forma: N488HA, N490HA y N494HA (véase Anexo 8, Ilustración 8).
* Se observó que, durante el mes de enero del 2020, los viajes tendieron a caer cada una semana aproximadamente. Este patrón se observó tres veces en el mes, específicamente los días 11, 18 y 25 de enero (véase Anexo 9, Ilustración 9).

1. **PREGUNTAS DE NEGOCIO**

Ya descompuesta la información y anteriormente haber realizado una visualización de ésta, se llega a la conclusión de que el mayor problema que tienen las aerolíneas son los retrasos., es por ello que se realizan las siguientes preguntas de negocio con el objetivo de aumentar la productividad en este sector:

1. **¿Qué tipos de aerolíneas están prefiriendo los consumidores para realizar sus viajes?**

Actualmente, tanto en Chile como en el mundo, ha existido una fuerte demanda por las aerolíneas *“low-cost”* ya que otorgan la posibilidad a muchas más personas para que puedan volar de un lugar a otro, tanto por turismo o como reemplazo de otros medios de transporte tales como autos y buses. Unos de los principales beneficios del sistema *low-cost* está el significativo ahorro para el pasajero, pero no sólo los pasajeros se han visto beneficiados, las agencias de viajes tradicionales también han sido beneficiadas por este fenómeno.

1. **¿****Existe evidencia que respalde que a medida que los vuelos se hacen más largos existe más probabilidad de atraso?**

Desde un punto de vista técnico, se podría decir que los viajes más largos presentarían una mayor probabilidad de atrasarse ya que estos, dependen, entre otros múltiples factores, de las condiciones mecánicas del avión y del clima de la ciudad de destino.

En el Gráfico 1, cada punto representa un vuelo y el tamaño del punto indica la cantidad de vuelos que recorrieron ciertas millas. Mientras que la línea azulada representa cómo se comportan los atrasos a medida que el número de millas va en aumento, observando, así, un alza de éstos cuando los vuelos recorren más de 900 millas.

Por lo tanto, sí existiría evidencia, en este conjunto de datos, que a medida que los vuelos se hacen más largos existiría mayor probabilidad de retraso de éste.

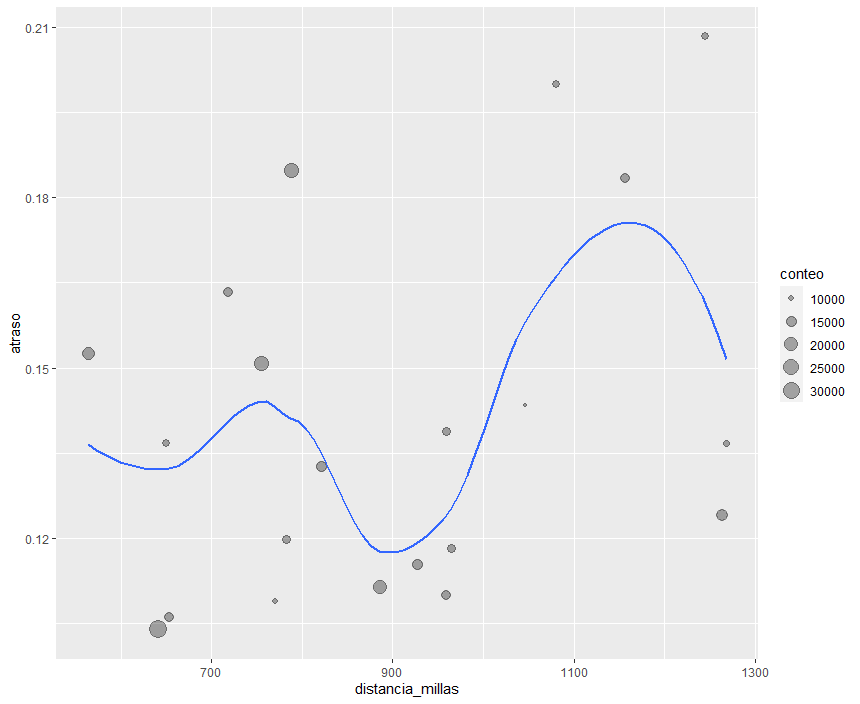
Cabe recalcar que el gráfico se realizó con aquellos aeropuertos de origen que presentaron más de 9000 vuelos durante el mes de enero.

Gráfico 1: Probabilidad de atraso en relación con la distancia recorrida (millas)

1. **¿****Los retrasos dependen de factores temporales, es decir, la hora y el día de salida de un vuelo, es determinante en los retrasos?**

De acuerdo con el Gráfico 2, la mayor concentración de retrasos de vuelos comprende los días: miércoles, jueves y viernes, siendo los dos primeros antes mencionados los días en los cuales se realizan una mayor cantidad de viajes (véase Anexo1, Ilustración1). Con respecto a este punto podría inferirse que la mayoría de las personas que realizan viajes dentro de Estados Unidos durante los días anteriormente mencionados los podrían realizar debido a motivos laborales.

Con respecto al bloque de horario en los cuales se registran más atrasos, este corresponde al número 13 (17:00 – 17:59), para luego caer drásticamente hasta el último bloque (23:00 – 23:59) en el cual se alcanza un menor número de atrasos (Gráfico 3). El bloque de horario 13 se definiría como la “segunda ola” de mayor cantidad viajes, superada por el bloque 2 (06:00 – 06:59), siendo este último en los cuales se realizan más viajes durante enero de 2020 en Estados Unidos. (véase Anexo 9, Ilustración 9).

De esto se desprende que las personas tienden a realizar viajes a un horario más temprano o más tardío (uno de los dos extremos) ya que es durante la tarde (desde las 14:00 hasta las 17:59) donde se registra un alza en los atrasos de los vuelos.

Por lo tanto, los retrasos, para este conjunto de datos, podrían depender tanto del día como del bloque de horario en los cuales se realicen.

Gráfico, Gráfico de barras

Descripción generada automáticamente

Gráfico 2

Fuente: Elaboración Propia

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Fuente: Elaboración Propia

Gráfico 3

1. **¿Las rutas más populares (mayor número de vuelos) tienen mayor tasa de retraso?**

Como se indica en el apartado anterior, las rutas más populares son: desde el Aeropuerto Internacional de los Ángeles hacia el Aeropuerto Internacional de San Francisco (LAX-SFO), desde el Aeropuerto de San Francisco hacia el Aeropuerto de Los Ángeles (SFO-LAX) y desde el Aeropuerto Internacional McCarran hacia el Aeropuerto Internacional de Los Ángeles (LAS-LAX). (véase Anexo 6, Ilustración 6).

A través del Gráfico 4, se observa que la ruta que registró más atrasos es SFO – LAX, seguida por LAX – SFO, las cuales se encuentran dentro de las tres rutas preferidas.

Por lo tanto, según los datos obtenidos, las rutas más populares sí podrían estar afectas a tener una tasa mayor de retraso.

Gráfico, Gráfico de barras

Descripción generada automáticamente

Fuente: Elaboración Propia

Gráfico 4

1. **PROBLEMAS Y CONCLUSIONES**

A continuación, se destacan algunos problemas y conclusiones en relación con el análisis de las predicciones de vuelos.

*Sobre problemas y acotaciones:*

* Existieron problemas para analizar la variable creada SPEED ya que no se pudo sumar las velocidades que existieron cada día y con ellas obtener la velocidad promedio que hubo cada día del mes de enero.

*Conclusiones:*

* Se registraron más vuelos realizados los días: martes, miércoles y jueves.
* Los destinos más visitados en Estados Unidos son Atlanta, Denver y Dallas.
* Cada día se registran más de 17.000 vuelos realizados en un día.
* Las rutas más populares son: Los Ángeles – San Francisco (y viceversa) y Las Vegas – Los Ángeles.
* A medida que los vuelos se hacen más largos estos tendrían la probabilidad de presentar atrasos.
* Los días de la semana y el bloque de horario podrían ser factores que condicionen el atraso de los vuelos.
* Los vuelos más populares podrían tener una mayor tasa de retraso.

# **REFERENCIAS**

Btesh, F. (2010). *Repositorio Digital San Andrés.* Obtenido de https://repositorio.udesa.edu.ar/jspui/bitstream/10908/245/1/%5bP%5d%5bW%5d%20T.L.%20Adm.%20Francis%20Btesh.pdf

Herrera, S., & Spera, A. (s.f.). *Portal de America.* Obtenido de https://www.portaldeamerica.com/images/documentos/Modelos\_de\_negocio\_en\_la\_Aviacion\_Comercial.pdf

Peiró, R. (4 de Agosto de 2017). *Economipedia*. Obtenido de https://economipedia.com/definiciones/modelo-de-negocio.html

Wikipedia. (2021). *Anexo: Aeropuertos más importantes de Estados Unidos*. Obtenido de https://es.wikipedia.org/wiki/Anexo:Aeropuertos\_m%C3%A1s\_importantes\_de\_Estados\_Unidos

Edestinos. (2021). *Anexo: Ranking de Aerolineas de Estados Unidos*. Obtenido de

https://www.edestinos.cl/aerolineas/nacionales/co/us/estadounidenses

**ANEXOS**

Gráfico, Gráfico de barras

Descripción generada automáticamente

Ilustración 1: Anexo 1

Ilustración 2: Anexo 2

Gráfico, Gráfico de barras

Descripción generada automáticamente

Gráfico, Gráfico de barras

Descripción generada automáticamente

Ilustración 3: Anexo 3

Gráfico, Histograma

Descripción generada automáticamente

Ilustración 4: Anexo 4

Gráfico, Gráfico de barras

Descripción generada automáticamenteGráfico, Gráfico de barras

Descripción generada automáticamente

Ilustración 5: Anexo 5

Ilustración 6: Anexo 6

Gráfico, Gráfico de barras

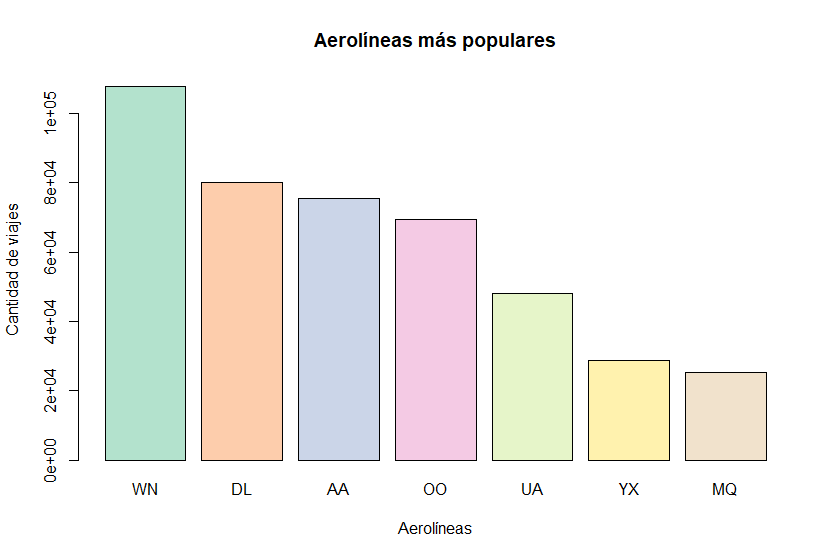
Descripción generada automáticamente

Ilustración 7: Anexo 7

Ilustración 8: Anexo 8

Gráfico, Gráfico de barras, Histograma

Descripción generada automáticamente

Ilustración 9: Anexo 9

Ilustración 9: Anexo 9

Gráfico, Gráfico de líneas, Histograma

Descripción generada automáticamente

1. Ambas informaciones se encuentran en <https://www.kayak.cl/vuelos/Chile-CL0/Chile-CL0> [↑](#footnote-ref-1)
2. [↑](#footnote-ref-2)