

Bases de Datos - Proyecto 1

Davis Alvarez - 15842

Febrero 28, 2019

Introducción

Históricamente, se conoce que el primer avión que emprendió vuelo exitosamente fue construido por los hermanos Wright. Hoy en día, los aviones se han vuelto algo cotidiano y que ha generado una competencia entre las aerolíneas existentes. Es por esta razón que se ha planteado la pregunta de investigación **¿qué aerolínea se desempeñó mejor durante el 2017-2018?** . La cual busca identificar las características que puede tener una aerolínea para que sea considerada preferible sobre las otras.

Base de datos

La base de datos utilizada para esta investigación se obtuvo en la página web de BTS, la cual contiene información sobre los vuelos comerciales realizados las aerolíneas más grandes en los Estados Unidos. La información que se puede encontrar en este set de datos abarca desde 1987 a 2018, en donde se detalla información sobre los vuelos retrasados, los puntuales, los cancelados, los desviados, entre otro.

Con ayuda del DBI, el cual es una Interfaz de Base de Datos que permite comunicarse con una base de datos SQLite, PostgreSQL o MySQL de forma sencilla y efectiva, se creó la conexión hacia la Base de Datos de las Aerolíneas.

```
cnm <- dbConnect(RSQLite::SQLite(), "DB/ontime-2017-2018.db")
```

La base de datos está estructurada de manera relacional, esta cuenta con 20 tablas:

Tablas en la DB
airline_id
airport
airport_id
airport_seq_id
cancellation
carrier_history
city_market_id
deparrblk
distance_group_250
diversions
months
ontime
ontime_delay_groups
quarters
state_abr_aviation
state_fips
unique_carriers
weekdays
world_area_codes
yesno_resp

Como parte de la comprensión de la Base de Datos de los Vuelos, se realizó un diagrama de Entidad Relación con las características principales. El cual puede ser accedido este enlace.

Análisis Exploratorio

Con el fin de responder la pregunta de investigación, se planteó indagar más sobre las aerolíneas del set de datos. Por esta razón se plantearon interrogantes adicionales que respondieran aspectos específicos sobre el comportamiento de las compañías aéreas. Cada una de estas preguntas se planteó con el fin de contribuir en la solución de la pregunta de investigación.

Pregunta 1

¿Qué aerolínea realizó mayor cantidad de vuelos durante el 2017 y 2018?

Con esta pregunta se buscó identificar las aerolíneas que presentaron mayor actividad durante los últimos dos años, es decir en 2017 y 2018, con el fin de identificar cuáles pueden ser las compañías más grandes en los Estados Unidos.

```
SELECT airline_id.Code, airline_id.Description, count(ontime.Flights)/1000 AS Vuelos
FROM ontime
INNER JOIN airline_id ON ontime.DOT_ID_Reporting_Airline = airline_id.Code
GROUP BY airline_id.Code
ORDER BY Vuelos DESC
```

Table 2: Resultado de la Consulta:

Code	Description	Vuelos
19393	Southwest Airlines Co.: WN	2588
19790	Delta Air Lines Inc.: DL	1811
19805	American Airlines Inc.: AA	1748
20304	SkyWest Airlines Inc.: OO	1427
19977	United Air Lines Inc.: UA	1163
20409	JetBlue Airways: B6	581
20366	ExpressJet Airlines Inc.: EV	533
19930	Alaska Airlines Inc.: AS	412
20416	Spirit Air Lines: NK	319
20452	Republic Airline: YX	290
20398	Envoy Air: MQ	270
20397	PSA Airlines Inc.: OH	255
20363	Endeavor Air Inc.: 9E	225
20436	Frontier Airlines Inc.: F9	214
20378	Mesa Airlines Inc.: YV	196
19690	Hawaiian Airlines Inc.: HA	157
21171	Virgin America: VX	89
20368	Allegiant Air: G4	88

En el resultado obtenido al ejecutar la Consulta, encontramos que *Southwest Airlines Co.* es la Aerolínea con más vuelos realizados en el periodo de 2017 al 2018, con una diferencia de más de 700 mil vuelos ante *Delta Air Lines Inc.*, quien es la segunda en realizar más vuelos en este periodo. Con los resultados se pudo observar que solo se tienen registros de 18 compañías aéreas en la Base de Datos de Vuelos. Apesar de tener registradas 1000 aerolíneas en la tabla de *airline_id*.

Pregunta 2

¿Qué aerolínea ha tenido menos vuelos cancelados?

El conocer cual compañía aérea tiene mayor cantidad de vuelos cancelados puede ser útil para saber cuál puede ser la más confiable. Por esta razón se buscó identificar cual dé es la que presento mayor cantidad de vuelos cancelados en los últimos 2 años.

```
SELECT airline_id.Code , airline_id.Description , SUM(ontime.Cancelled) as Cancelados
FROM ontime
    INNER JOIN airline_id ON ontime.DOT_ID_Reporting_Airline = airline_id.Code
GROUP BY airline_id.Code
ORDER BY Cancelados ASC;
```

Table 3: Resultado de la Consulta:

Code	Description	Cancelados
19690	Hawaiian Airlines Inc.: HA	484
20368	Allegiant Air: G4	707
21171	Virgin America: VX	1430
20436	Frontier Airlines Inc.: F9	3264
19930	Alaska Airlines Inc.: AS	3316
20378	Mesa Airlines Inc.: YV	5098
20363	Endeavor Air Inc.: 9E	6239
20416	Spirit Air Lines: NK	6948
20452	Republic Airline: YX	9857
20398	Envoy Air: MQ	10083
20397	PSA Airlines Inc.: OH	10851
19977	United Air Lines Inc.: UA	11080
19790	Delta Air Lines Inc.: DL	12517
20409	JetBlue Airways: B6	14687
20366	ExpressJet Airlines Inc.: EV	15633
20304	SkyWest Airlines Inc.: OO	20215
19805	American Airlines Inc.: AA	26131
19393	Southwest Airlines Co.: WN	36200

En los resultados de esta Consulta podemos observar que *Southwest Airlines Co.* fue la compañía que presento mayor cantidad de vuelos cancelados durante el periodo seleccionado y que *Hawaiian Airlines Inc.* es la que menos vuelos cancelo. Pero la cantidad de vuelos cancelados puede no ser tan importante como la probabilidad de que un vuelo sea cancelado. Por lo que se decidió realizar otra Consulta para obtener el porcentaje de vuelos cancelados.

```
SELECT airline_id.Code , airline_id.Description , SUM(ontime.Cancelled)/sum(ontime.Flights) as Porcenta
FROM ontime
    INNER JOIN airline_id ON ontime.DOT_ID_Reporting_Airline = airline_id.Code
GROUP BY airline_id.Code
ORDER BY PorcentajeCancelados ASC;
```

Table 4: Resultado de la Consulta:

Code	Description	PorcentajeCancelados
19690	Hawaiian Airlines Inc.: HA	0.0030733
19790	Delta Air Lines Inc.: DL	0.0069094
20368	Allegiant Air: G4	0.0079843

Code	Description	PorcentajeCancelados
19930	Alaska Airlines Inc.: AS	0.0080479
19977	United Air Lines Inc.: UA	0.0095270
19393	Southwest Airlines Co.: WN	0.0139856
20304	SkyWest Airlines Inc.: OO	0.0141648
19805	American Airlines Inc.: AA	0.0149488
20436	Frontier Airlines Inc.: F9	0.0152394
21171	Virgin America: VX	0.0160084
20416	Spirit Air Lines: NK	0.0217518
20409	JetBlue Airways: B6	0.0252450
20378	Mesa Airlines Inc.: YV	0.0259358
20363	Endeavor Air Inc.: 9E	0.0277009
20366	ExpressJet Airlines Inc.: EV	0.0293115
20452	Republic Airline: YX	0.0339369
20398	Envoy Air: MQ	0.0372885
20397	PSA Airlines Inc.: OH	0.0424870

En este caso, se puede observar que *Hawaiian Airlines Inc.* sigue siendo la compañía que menor porcentaje de vuelos cancelados tiene, pero en el caso de la compañía que tiene mayor porcentaje de vuelos cancelados es *PSA Airlines Inc.* con el 4% de vuelos cancelados en los últimos 2 años.

Pregunta 3

¿Qué aerolínea presenta menor cantidad de retraso antes del abordaje?

Conocer las causas de retraso antes del abordaje del avión es importante debido a que el menor tiempo de espera es sinónimo de eficiencia del personal y por ende mejor organización. Por esta razón se planteó identificar las aerolíneas que presentan menor tiempo en las causas de retraso como la seguridad, NAS, cargador y aeronave. Es importante resaltar que no se tomó en cuenta los retrasos causados por el clima, ya que es algo que no se puede controlar y es afectado directamente por cada región.

```
SELECT airline_id.Code , airline_id.Description, (security + carrier + nas + aircraft) as Retraso
FROM (
    SELECT DOT_ID_Reporting_Airline as idC, AVG(SecurityDelay) as security,
        AVG(CarrierDelay) as carrier, AVG(NasDelay) as nas,
        AVG(LateAircraftDelay) as aircraft
    FROM ontime
    GROUP BY DOT_ID_Reporting_Airline
)
INNER JOIN airline_id ON idC = airline_id.Code
ORDER BY Retraso ASC;
```

Table 5: Resultado de la Consulta:

Code	Description	Retraso
19690	Hawaiian Airlines Inc.: HA	4.383109
19930	Alaska Airlines Inc.: AS	7.575510
19790	Delta Air Lines Inc.: DL	8.659029
19393	Southwest Airlines Co.: WN	9.415346
20398	Envoy Air: MQ	10.569509
20452	Republic Airline: YX	11.076633
19805	American Airlines Inc.: AA	11.210230
19977	United Air Lines Inc.: UA	11.864688

Code	Description	Retraso
20397	PSA Airlines Inc.: OH	12.689701
20363	Endeavor Air Inc.: 9E	12.957785
20378	Mesa Airlines Inc.: YV	13.010602
20304	SkyWest Airlines Inc.: OO	13.333810
20416	Spirit Air Lines: NK	13.373853
21171	Virgin America: VX	15.065131
20368	Allegiant Air: G4	15.630081
20366	ExpressJet Airlines Inc.: EV	15.955263
20436	Frontier Airlines Inc.: F9	17.652033
20409	JetBlue Airways: B6	18.145215

Para obtener el resultado de esta Consulta, primero se calculó la cantidad promedio de retraso por cada una de las 4 causas de retraso elegidas en cada aerolínea. Luego se sumó los 4 promedios para obtener la cantidad promedio de retraso en un vuelo. El resultado indicó que la compañía que presenta mayor promedio de retraso por vuelo es *JetBlue Airways* y la compañía que presenta menos retraso es *Hawaiian Airlines Inc.*.

Pregunta 4

¿Qué aerolínea presenta menor cantidad de retraso después del abordaje?

Ahora que se tiene conocimiento sobre el retraso de las compañías aéreas antes de abordar el avión, se procedió a distinguir cuáles son las aerolíneas con menor retraso después de abordaje. Para calcular este retraso se tomó en cuenta el retraso de despegue y de aterrizaje de cada vuelo.

Table 6: Resultado de la Consulta:

Code	Description	Retraso
19690	Hawaiian Airlines Inc.: HA	2.173207
19930	Alaska Airlines Inc.: AS	2.450461
19790	Delta Air Lines Inc.: DL	8.376851
20452	Republic Airline: YX	10.442818
20398	Envoy Air: MQ	12.724639
19977	United Air Lines Inc.: UA	13.514787
19805	American Airlines Inc.: AA	13.948517
20363	Endeavor Air Inc.: 9E	14.910424
19393	Southwest Airlines Co.: WN	15.390914
20416	Spirit Air Lines: NK	15.904678
20304	SkyWest Airlines Inc.: OO	17.091607
20378	Mesa Airlines Inc.: YV	19.142825
20397	PSA Airlines Inc.: OH	20.175285
20366	ExpressJet Airlines Inc.: EV	20.833553
21171	Virgin America: VX	20.922208
20368	Allegiant Air: G4	23.419361
20436	Frontier Airlines Inc.: F9	26.426877
20409	JetBlue Airways: B6	27.091052

Para obtener el resultado de esta Consulta, primero se calculó la cantidad promedio de retraso del despegue y del aterrizaje. Posteriormente se sumó promedios para obtener la cantidad promedio de retraso. El resultado indicó que la compañía que presenta mayor promedio de retraso por vuelo es *JetBlue Airways* y la compañía

que presenta menos retraso es *Hawaiian Airlines Inc.*.

Pregunta 5

¿Qué aerolínea mejor promedio de vuelos puntuales?

La puntualidad es una cualidad importante cuando se trata de viajes importantes. Por esta razón se decidió buscar cuales aerolíneas presentan el mayor porcentaje de vuelos que aterrizaron en el tiempo esperado o antes.

```
SELECT airline_id.Code , airline_id.Description , (puntuales/vuelosT) as PorcentajePuntuales
FROM ( SELECT DOT_ID_Reporting_Airline as idC, SUM(Flights) as puntuales
      FROM ontime
      WHERE ArrDelay<1
      GROUP BY DOT_ID_Reporting_Airline
      ORDER BY DOT_ID_Reporting_Airline ASC
    )
INNER JOIN airline_id ON idC = airline_id.Code
INNER JOIN (SELECT DOT_ID_Reporting_Airline as idAir, count(Flights) as vuelosT
            FROM ontime
            group by DOT_ID_Reporting_Airline
            ORDER BY DOT_ID_Reporting_Airline ASC) vTotales on vTotales.idAir = airline_id.Code
GROUP BY airline_id.Code
ORDER BY PorcentajePuntuales DESC;
```

Table 7: Resultado de la Consulta:

Code	Description	PorcentajePuntuales
19790	Delta Air Lines Inc.: DL	0.7191252
19977	United Air Lines Inc.: UA	0.6706302
20363	Endeavor Air Inc.: 9E	0.6574034
20416	Spirit Air Lines: NK	0.6497768
20452	Republic Airline: YX	0.6464464
19690	Hawaiian Airlines Inc.: HA	0.6452111
20304	SkyWest Airlines Inc.: OO	0.6410603
20366	ExpressJet Airlines Inc.: EV	0.6367852
19930	Alaska Airlines Inc.: AS	0.6344225
19805	American Airlines Inc.: AA	0.6204140
20398	Envoy Air: MQ	0.6058246
20378	Mesa Airlines Inc.: YV	0.6053866
20368	Allegiant Air: G4	0.6009328
19393	Southwest Airlines Co.: WN	0.5930774
20436	Frontier Airlines Inc.: F9	0.5831236
20409	JetBlue Airways: B6	0.5695729
20397	PSA Airlines Inc.: OH	0.5669392
21171	Virgin America: VX	0.5283898

En los resultados podemos observar que *Delta Air Lines Inc.* aterriza el 71% de sus vuelos en el tiempo previsto, mientras que *Virgin America* es la alinea que tiene menor porcentaje de vuelos aterrizados puntalmente con el 52% de vuelos a tiempo. También es importante resaltar que muchas de las compañías de aviones presentan aterrizar alrededor del 60% de sus vuelos a tiempo y que muchos retrasos en el aterrizaje pueden ser no significativos. Por lo que se recomienda realizar un análisis de la cantidad promedio de retraso en el aterrizaje por cada aerolínea.

Pregunta 6

¿Qué aerolínea visita más aeropuertos?

La cantidad de aeropuertos visitados por una aerolínea nos dice la cantidad de terreno que cubre cada aerolínea, lo que se traduce en más opciones de vuelos para distintos destinos. Por esta razón se planteó encontrar cual es la aerolínea que visita más aeropuertos dentro de los Estados Unidos.

```
SELECT airline_id.Description, COUNT(idCompa) as CantidadVisitados
FROM(SELECT DISTINCT DOT_ID_Reporting_Airline AS idCompa, airport_id.Code as idAir
      FROM ontime
      INNER JOIN airport_id ON airport_id.Code=ontime.OriginAirportId)
INNER JOIN airline_id ON idCompa = airline_id.Code
GROUP BY idCompa
ORDER BY CantidadVisitados DESC;
```

Table 8: Resultado de la Consulta:

Description	CantidadVisitados
SkyWest Airlines Inc.: OO	268
ExpressJet Airlines Inc.: EV	176
Delta Air Lines Inc.: DL	158
Envoy Air: MQ	157
Endeavor Air Inc.: 9E	136
Allegiant Air: G4	126
Mesa Airlines Inc.: YV	121
United Air Lines Inc.: UA	117
American Airlines Inc.: AA	111
Republic Airline: YX	105
PSA Airlines Inc.: OH	101
Frontier Airlines Inc.: F9	97
Southwest Airlines Co.: WN	88
Alaska Airlines Inc.: AS	75
JetBlue Airways: B6	70
Spirit Air Lines: NK	43
Virgin America: VX	28
Hawaiian Airlines Inc.: HA	18

Podemos observar que *SkyWest Airlines Inc.* es la aerolínea que realiza vuelos desde la mayor cantidad de destinos, con 268 lugares diferentes. También, se puede observar que *Hawaiian Airlines Inc.* solamente hace viajes desde 18 aeropuertos diferentes.

Pregunta 7

¿Qué aerolínea ha tenido menor promedio de vuelos desviados?

Identificar la cantidad de vuelos que fueron desviados es importante, ya que, de esta manera se puede observar si una aerolínea cambia los destinos de sus vuelos recurrentemente. Por esta razón se buscó identificar la compañía de aviones que tiene mayor promedio de vuelos desviados.

```
SELECT airline_id.Code , airline_id.Description , SUM(ontime.Diverted)/sum(ontime.Flights) as Porcentaje
FROM ontime
      INNER JOIN airline_id ON ontime.DOT_ID_Reporting_Airline = airline_id.Code
```

```
GROUP BY airline_id.Code
ORDER BY PorcentajeDesviados DESC;
```

Table 9: Resultado de la Consulta:

Code	Description	PorcentajeDesviados
21171	Virgin America: VX	0.0036495
20304	SkyWest Airlines Inc.: OO	0.0031279
20366	ExpressJet Airlines Inc.: EV	0.0029906
20397	PSA Airlines Inc.: OH	0.0029405
20398	Envoy Air: MQ	0.0028624
20368	Allegiant Air: G4	0.0026426
20409	JetBlue Airways: B6	0.0025285
19977	United Air Lines Inc.: UA	0.0025176
19930	Alaska Airlines Inc.: AS	0.0024780
20378	Mesa Airlines Inc.: YV	0.0024115
19805	American Airlines Inc.: AA	0.0023873
20452	Republic Airline: YX	0.0023343
20363	Endeavor Air Inc.: 9E	0.0022422
19790	Delta Air Lines Inc.: DL	0.0020220
19393	Southwest Airlines Co.: WN	0.0019000
20416	Spirit Air Lines: NK	0.0015747
20436	Frontier Airlines Inc.: F9	0.0014894
19690	Hawaiian Airlines Inc.: HA	0.0011175

El resultado de esta Consulta indicó que la cantidad de vuelos desviados en general es poca para todas las compañías. Tomando en cuenta que estos desvíos pueden ser causados por el clima, puede que este porcentaje no sea muy importante para tomar en cuenta en la resolución de la pregunta de investigación.

Pregunta 8

¿Qué aerolínea tuvo mayor Distancia recorrida durante la temporada alta?

```
SELECT airline_id.Code, airline_id.Description, CAST(Promedio_Distancia AS INTEGER) AS Distancia
FROM
    (SELECT ontime.DOT_ID_Reporting_Airline as idAero, sum(ontime.Distance)/1000000 as Promedio_Distancia
    FROM ontime
    LEFT JOIN months ON ontime.Month = months.Code
    WHERE months.Description = 'January' OR months.Description = 'February' OR months.Description = 'July'
    OR months.Description = 'August' OR months.Description = 'December'
    GROUP BY ontime.DOT_ID_Reporting_Airline)
LEFT JOIN airline_id ON idAero = airline_id.Code
ORDER BY Distancia DESC;
```

Table 10: Resultado de la Consulta:

Code	Description	Distancia
19393	Southwest Airlines Co.: WN	668
19805	American Airlines Inc.: AA	631
19790	Delta Air Lines Inc.: DL	558
19977	United Air Lines Inc.: UA	480
20304	SkyWest Airlines Inc.: OO	237

Code	Description	Distancia
20409	JetBlue Airways: B6	215
19930	Alaska Airlines Inc.: AS	179
20416	Spirit Air Lines: NK	116
20366	ExpressJet Airlines Inc.: EV	87
20436	Frontier Airlines Inc.: F9	80
20452	Republic Airline: YX	62
21171	Virgin America: VX	52
20398	Envoy Air: MQ	42
20363	Endeavor Air Inc.: 9E	38
20378	Mesa Airlines Inc.: YV	38
19690	Hawaiian Airlines Inc.: HA	35
20397	PSA Airlines Inc.: OH	35
20368	Allegiant Air: G4	28

Analisis de resultados

Para realizar la comparación entre las compañías aéreas en cada una de las Consultas se realizó una Tabla de Comparación, en la cual se colocó cada aerolínea según su resultado. De esta manera se pudo observar su comportamiento, respecto a la consistencia de sus resultados comparadas contra las otras líneas aéreas.

Debido a que, en la Tabla de Comparación se puede notar que *Delta Air Lines Inc.* siempre se encuentra en el en los puestos más altos y que en las categorías más importantes está en los primeros lugares, se recomienda realizar la inversión en esta compañía. Por lo que la respuesta a la pregunta de investigación **¿qué aerolínea se desempeñó mejor durante el 2017-2018?** es *Delta Air Lines Inc.*.