Proj. analityczny: linie lotnicze

Adrianna Hornicka

28 maja 2019

Raport dotyczy opóźnień połączeń lotniczych w USA w lipcu 2017 r. Uwzglednia dane udostępnione przez Departament Transportu Stanów Zjednoczonych. Sa to m.in.:

- linie lotnicze,
- lotniska,
- miasta wylotu i przylotu,
- dokladna data,
- czas opoznienia

1.Jakie bylo średnie opóźnienie przylotu?

```
delayavg <-  DBI::dbGetQuery(con, "SELECT avg(arr_delay_new) as 'Avg delay'
                             FROM Flight_delays ")
delayavg
```

```
##   Avg delay
## 1  15.91152
```

2.Jakie było maksymalne opóźnienie przylotu?

```
delaymax <- DBI::dbGetQuery(con, "SELECT max(arr_delay_new) as 'Max delay'
                            FROM Flight_delays")
delaymax
```

```
##   Max delay
## 1      1895
```

3.Który lot mial najwieksze opóźnienie przylotu?

```
delayflight <- DBI::dbGetQuery(con, "SELECT carrier, origin_city_name, dest_city_name, fl_date, arr_dela
delayflight
```

```
##   carrier origin_city_name  dest_city_name    fl_date arr_delay_new
## 1      AA         Kona, HI Los Angeles, CA 2017-07-26          1895
```

4.Które dni tygodnia są najgorsze do podróżowania?

```
delayweek <- DBI::dbGetQuery(con, "SELECT weekday_name, avg(arr_delay_new) as'avg_delay' FROM Weekdays .
delayweek
```

```
##   weekday_name avg_delay
## 1       Friday  20.80747
## 2       Monday  18.04801
## 3     Saturday  15.21876
```

```
## 4       Sunday  12.77606
## 5     Thursday  15.64696
## 6      Tuesday  12.88056
## 7    Wednesday  16.10514
```

5.Które linie lotnicze latające z San Francisco (SFO) mają najmniejsze opóźnienia przylotu?

```
sfodelay <- DBI::dbGetQuery(con, "SELECT A.airline_name,
       (SELECT avg(F.arr_delay_new) AS 'avg_delay'
        FROM Flight_delays F
               JOIN Airlines A2 ON A2.airline_id = F.airline_id
        WHERE A.airline_name = A2.airline_name GROUP BY A2.airline_name) AS 'time'
FROM Flight_delays F JOIN Airlines A on F.airline_id = A.airline_id
WHERE F.origin LIKE 'SFO'
GROUP BY A.airline_name
ORDER BY 'time' DESC;")
sfodelay
```

```
##                  airline_name      time
## 1         JetBlue Airways: B6 28.841148
## 2  Frontier Airlines Inc.: F9 18.980300
## 3  American Airlines Inc.: AA 18.375314
## 4   United Air Lines Inc.: UA 16.950403
## 5   SkyWest Airlines Inc.: OO 16.808273
## 6          Virgin America: VX 13.964467
## 7  Southwest Airlines Co.: WN 13.823983
## 8    Delta Air Lines Inc.: DL 12.258788
## 9    Alaska Airlines Inc.: AS  7.453928
## 10 Hawaiian Airlines Inc.: HA  4.202719
```

6.Jaka część linii lotniczych ma regularne opóźnienia, tj. jej lot ma średnio co najmniej 10 min. opóźnienia?

```
regulardelays <- DBI::dbGetQuery(con, "SELECT
(SELECT count(*)
from (select airline_id from flight_delays
      group by airline_id
      having avg(arr_delay_new) >= 10 ) as tab) /
CONVERT(float, (select count(distinct airline_id) from Flight_delays) )as 'late proportion'")

regulardelays
```

```
##   late proportion
## 1       0.8333333
```

7.Jak opóźnienia wylotów wpływają na opóźnienia przylotów?

```
pearson <- DBI::dbGetQuery(con, "SELECT ((SUM(dep_delay_new*arr_delay_new)-(SUM(dep_delay_new)*
SUM(arr_delay_new))/COUNT(*)))/(SQRT(SUM(dep_delay_new*dep_delay_new)-
```

```
(SUM(dep_delay_new)*SUM(dep_delay_new))/COUNT(*))*
SQRT(SUM(arr_delay_new*arr_delay_new)-
(SUM(arr_delay_new)*SUM(arr_delay_new))/COUNT(*))) as 'Pearson r'
FROM Flight_delays")
pearson
```

```
##   Pearson r
## 1 0.9717058
```

8.Która linia lotnicza miała największy wzrost (różnica) średniego opóźnienia przylotów w ostatnim tygodniu miesiąca, tj. między 1-23 a 24-31 lipca?

```
delaylast <- DBI::dbGetQuery(con, "WITH firstaverage AS (SELECT A.airline_name, avg(F.arr_delay_new) as
FROM Flight_delays F JOIN Airlines A on F.airline_id = A.airline_id
WHERE F.day_of_month BETWEEN 1 AND 23
GROUP BY A.airline_name),
secondaverage AS (SELECT A.airline_name, avg(F.arr_delay_new) as 'delay_increase2'
from Flight_delays F join Airlines A on F.airline_id = A.airline_id
WHERE F.day_of_month BETWEEN 24 AND 31
GROUP BY A.airline_name)

SELECT top 1 firstaverage.airline_name, (secondaverage.delay_increase2 - firstaverage.delay_increase1)
FROM firstaverage JOIN secondaverage on firstaverage.airline_name = secondaverage.airline_name
order by 'Delay_increase' DESC")
delaylast
```

```
##                 airline_name Delay_increase
## 1 Southwest Airlines Co.: WN       0.584763
```

9.Które linie lotnicze latają zarówno na trasie SFO › PDX (Portland), jak i SFO › EUG (Eugene)?

```
portlandeugene <- DBI::dbGetQuery(con, "(SELECT distinct A.airline_name
FROM Airlines A join Flight_delays F ON A.airline_id = F.airline_id
WHERE F.origin = 'SFO' AND F.dest = 'EUG')
INTERSECT
(SELECT distinct A.airline_name
FROM Airlines A join Flight_delays F ON A.airline_id = F.airline_id
WHERE F.origin = 'SFO' AND F.dest = 'PDX')")
portlandeugene
```

```
##                 airline_name
## 1 United Air Lines Inc.: UA
## 2 SkyWest Airlines Inc.: OO
```

10.Jak najszybciej dostać się z Chicago do Stanfordu, zakładając wylot po 14:00 czasu lokalnego?

```
timelocal <- DBI::dbGetQuery(con, "SELECT origin 'Origin', dest 'Desitnation', avg(arr_delay_new)
'Delay'
FROM Flight_delays
WHERE origin IN ('ORD', 'MDW') AND dest IN ('SJC', 'OAK', 'SFO')
AND crs_dep_time > 1400
GROUP BY origin, dest
ORDER BY 'Delay' DESC")
timelocal
```

```
##   Origin Desitnation    Delay
## 1    ORD         SFO 22.19253
## 2    MDW         SFO 19.85714
## 3    MDW         SJC 17.20000
## 4    ORD         SJC 14.81111
## 5    MDW         OAK 12.12903
```