

# Linie lotnicze Projekt 2022

Milosz Ksiazek

21 05 2022

Projekt przedstawiający polecenia wraz z wygenerowanymi tabelami z analizy bazy danych dotyczących opóźnień połączeń lotniczych w USA w lipcu 2017 r.

## Zadanie 1

Jakie było średnie opóźnienie przylotu?

```
SELECT avg(arr_delay_new) AS "avg_delay"
FROM "Flight_delays";
```

Table 1: 1 records

avg_delay
15.91152

## Zadanie 2

Jakie było maksymalne opóźnienie przylotu?

```
SELECT max(arr_delay_new) AS "max_delay" FROM "Flight_delays";
```

Table 2: 1 records

max_delay
1895

## Zadanie 3

Który lot miał największe opóźnienie przylotu?

[przewoźnik, miasto wylotu, miasto przylotu, data lotu, opóźnienie]

```
SELECT carrier, origin_city_name, dest_city_name, fl_date, arr_delay_new
FROM "Flight_delays"
WHERE arr_delay_new = (SELECT max(arr_delay_new) FROM "Flight_delays");
```

Table 3: 1 records

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

## Zadanie 4

Które dni tygodnia są najgorsze do podróżowania?

[tabela zawierająca dla każdego dnia tygodnia średni czas opóźnienia]

```
SELECT weekday_name, avg(arr_delay_new) AS average
FROM "Flight_delays" f
INNER JOIN "Weekdays" d ON f.day_of_week = d.weekday_id
GROUP BY weekday_name
ORDER BY average desc;
```

Table 4: 7 records

weekday_name	average
Friday	20.80747
Monday	18.04801
Wednesday	16.10514
Thursday	15.64696
Saturday	15.21876
Tuesday	12.88056
Sunday	12.77606

## Zadanie 5

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

[tabela zawierająca nazwę przewoźnika oraz średnie opóźnienie z jego wszystkich lotów]

```
SELECT airline_name, avg_delay from
(SELECT airline_id , avg(arr_delay_new) AS avg_delay
FROM "Flight_delays"
GROUP BY airline_id ) AS all_avg
INNER JOIN "Airlines" USING (airline_id)
WHERE airline_id IN
(SELECT DISTINCT(airline_id) FROM "Flight_delays" WHERE origin = 'SFO')
ORDER BY avg_delay DESC;
```

Table 5: Displaying records 1 - 10

airline_name	avg_delay
JetBlue Airways: B6	28.841148
Frontier Airlines Inc.: F9	18.980300
American Airlines Inc.: AA	18.375314
United Air Lines Inc.: UA	16.950403
SkyWest Airlines Inc.: OO	16.808273
Virgin America: VX	13.964467
Southwest Airlines Co.: WN	13.823983
Delta Air Lines Inc.: DL	12.258788
Alaska Airlines Inc.: AS	7.453927
Hawaiian Airlines Inc.: HA	4.202719

## Zadanie 6

Jaka część linii lotniczych ma regularne opóźnienia, tj. jej lot ma średnio co najmniej 10 min. opóźnienia? [tylko linie lotnicze występujące w tabeli Flight\_delays]

```
SELECT cast(A.number_of_carriers_10 AS float)/CAST(B.number_of_carriers AS float)
AS late_proportion
FROM
  (SELECT count(*) number_of_carriers_10 FROM (
SELECT carrier, avg(arr_delay_new) AS average
FROM "Flight_delays"
GROUP BY carrier
HAVING avg(arr_delay_new) >= 10) AS aa ) A,
  (SELECT count(*) number_of_carriers
FROM (SELECT DISTINCT(carrier)
FROM "Flight_delays") AS bb) B;
```

Table 6: 1 records

late_proportion
0.8333333

## Zadanie 7

Jak opóźnienia wylotów wpływają na opóźnienia przylotów? [współczynnik korelacji Pearsona między czasem opóźnienia wylotów a czasem opóźnienia przylotów]

```
SELECT CORR(arr_delay_new, dep_delay_new) as Pearsons_r
FROM "Flight_delays" ;
```

Table 7: 1 records

pearsons_r
0.9763465

```
SELECT (avg(arr_delay_new * dep_delay_new) - (avg(arr_delay_new) *
avg(dep_delay_new))) / (stddev(arr_delay_new) * stddev(dep_delay_new))
AS Pearsons_r
FROM "Flight_delays";
```

Table 8: 1 records

pearsons_r
0.9737081

```
SELECT ((SUM(arr_delay_new * dep_delay_new) - (SUM(arr_delay_new)
* SUM(dep_delay_new)) / COUNT(*)))
/ ((SQRT(SUM(arr_delay_new * arr_delay_new) - (SUM(arr_delay_new)
* SUM (arr_delay_new)) / COUNT(*))
* SQRT(SUM(dep_delay_new * dep_delay_new) - (SUM(dep_delay_new)
* SUM(dep_delay_new)) / COUNT(*)))
AS Pearsons_r
```

```
FROM "Flight_delays";
```

Table 9: 1 records

pearsons_r
0.9717058

## Zadanie 8

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

[nazwa przewoźnika oraz wzrost]

```
SELECT * FROM
(SELECT airline_name , delay_increase FROM "Airlines"
INNER JOIN
(
SELECT l.airline_id, average_last - average_first AS delay_increase
FROM
(SELECT airline_id, avg(arr_delay_new) AS average_first
FROM "Flight_delays"
WHERE day_of_month <= 23
GROUP BY airline_id) AS f
INNER JOIN
(SELECT airline_id, avg(arr_delay_new) AS average_last
FROM "Flight_delays"
WHERE day_of_month >= 24
GROUP BY airline_id) AS l
ON f.airline_id = l.airline_id
ORDER BY delay_increase DESC) AS ids
USING (airline_id)) AS list
WHERE list.delay_increase =
(SELECT max(delay_increase)
FROM (SELECT airline_name, delay_increase FROM "Airlines"
INNER JOIN
(
SELECT l.airline_id, average_last - average_first AS delay_increase
FROM
(SELECT airline_id, avg(arr_delay_new) AS average_first
FROM "Flight_delays"
WHERE day_of_month <= 23
GROUP BY airline_id) AS f
INNER JOIN
(SELECT airline_id, avg(arr_delay_new) AS average_last
FROM "Flight_delays"
WHERE day_of_month >= 24
GROUP BY airline_id) AS l
ON f.airline_id = l.airline_id
ORDER BY delay_increase DESC) AS ids
USING (airline_id)) AS wh);
```

Table 10: 1 records

airline_name	delay_increase
Southwest Airlines Co.: WN	0.584763

## Zadanie 9

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

```
SELECT airline_name FROM "Airlines"
INNER join
(SELECT distinct(one.airline_id) from
(SELECT airline_id, origin, dest
FROM "Flight_delays"
WHERE origin = 'SFO' AND(dest = 'PDX' OR dest = 'EUG')) AS one
INNER JOIN
(SELECT DISTINCT airline_id, origin, dest
FROM "Flight_delays"
WHERE origin = 'SFO' AND(dest = 'PDX' OR dest = 'EUG')) AS two
ON one.airline_id = two.airline_id
AND one.dest <> two.dest) AS ids
USING (airline_id);
```

Table 11: 2 records

airline_name
United Air Lines Inc.: UA
SkyWest Airlines Inc.: OO

## Zadanie 10

Jak najszybciej dostać się z Chicago do Stanfordu, zakładając wylot po 14:00 czasu lokalnego? [tabela zawierająca jako miejsce wylotu Midway (MDW) lub O'Hare (ORD), jako miejsce przylotu San Francisco (SFO), San Jose (SJC) lub Oakland (OAK) oraz średni czas opóźnienia przylotu dla wylotów po 14:00 czasu lokalnego (atrybut crs\_dep\_time); wyniki pogrupowane po miejscu wylotu i przylotu, posortowane malejąco]

```
SELECT origin, dest, avg(arr_delay_new) AS avg_delay
FROM "Flight_delays"
WHERE (origin = 'MDW' OR origin = 'ORD')
AND (dest = 'SFO' OR dest = 'OAK' OR dest = 'SJC')
AND crs_dep_time > 1400
GROUP BY origin, dest
ORDER BY avg_delay DESC;
```

Table 12: 5 records

origin	dest	avg_delay
ORD	SFO	22.19253
MDW	SFO	19.85714
MDW	SJC	17.20000
ORD	SJC	14.81111
MDW	OAK	12.12903