Airlines

Miawka26

31.03.2022

a. What was the average arrival delay?

```
SELECT AVG(arr_delay_new) AS avg_delay
FROM "Flight_delays"
```

Table 1: 1 records

avg_delay
15.91152

b. What was the maximum arrival delay?

```
SELECT MAX(arr_delay_new) AS max_delay
FROM "Flight_delays"
```

Table 2: 1 records

max_delay
1895

c. Which flight had the largest arrival delay?

```
SELECT carrier,
       origin_city_name,
       dest_city_name,
       fl_date,
       arr_delay_new
FROM "Flight_delays"
order BY COALESCE(arr_delay_new, 0.0) DESC
LIMIT 1
```

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

d. Which days of the week are the worst for travel?

```
SELECT w.weekday_name AS weekday_name,
       AVG(arr_delay_new) AS avg_delay
FROM "Flight_delays" f
```

```
  INNER JOIN "Weekdays" w
    ON f.day_of_week=w.weekday_id
GROUP BY w.weekday_name
ORDER BY AVG(arr_delay_new) desc
```

Table 4: 7 records

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

e. Which airline flying from San Francisco (SFO) have the lowest arrival delays?

```
SELECT a.airline_name,
       AVG(f.arr_delay_new) AS avg_delay
FROM "Flight_delays" f
    INNER JOIN "Airlines" a
        ON f.airline_id=a.airline_id
WHERE f.airline_id IN
    (SELECT airline_id
    FROM "Flight_delays"
    WHERE origin='SFO')
GROUP BY a.airline_name
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

f. What proportion of airlines has regular delays, i.e., their flight has an average delay of at least 10 minutes?

```
SELECT count(DISTINCT f.airline_id)*1.0/count(DISTINCT f1.airline_id)*1.0 AS late_proportion
FROM "Flight_delays" f
  INNER JOIN "Flight_delays" f1
    ON f.fl_num = f1.fl_num
WHERE f.airline_id IN
```

```
  (SELECT airline_id
  FROM "Flight_delays"
  GROUP BY airline_id
  HAVING avg(arr_delay_new)>10)
```

Table 6: 1 records

late_proportion
0.8333333

g. How do departure delays affect arrival delays? [Pearson correlation coefficient between departure delay time and arrival delay time].

```
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 7: 1 records

Pearsons r
0.9717058

h. Which airline had the largest increase (difference) in average arrival delay during the last week of the month, i.e., between July 1-23 and July 24-31?

```
SELECT a.airline_name AS "airline_name",
       AVG(f2.arr_delay_new)-AVG(f1.arr_delay_new) AS "delay_increase"
FROM "Flight_delays" f1
  INNER JOIN "Flight_delays" f2
    ON f1.airline_id=f2.airline_id
  INNER JOIN "Airlines" a
    ON f1.airline_id=a.airline_id
WHERE (f1.DAY_of_month<=23) AND (f2.DAY_of_month>=24)
GROUP BY a.airline_name
ORDER BY AVG(f2.arr_delay_new)-AVG(f1.arr_delay_new) DESC
LIMIT 1
```

Table 8: 1 records

airline_name	delay_increase
Southwest Airlines Co.: WN	0.584763

i. Which airline flies both SFO → PDX (Portland) and SFO → EUG (Eugene)?

```
SELECT DISTINCT(a.airline_name)
FROM "Flight_delays" f
  INNER JOIN "Airlines" a
    ON f.airline_id=a.airline_id
WHERE f.origin='SFO' AND  f.dest='PDX' AND f.airline_id IN
```

```
  (SELECT airline_id
  FROM "Flight_delays"
  WHERE origin='SFO' AND  dest='EUG')
```

Table 9: 2 records

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

j. What is the fastest way to get from Chicago to Stanford, assuming departure after 2:00 pm local time?

```
SELECT origin,
       dest,
       AVG(arr_delay_new)
FROM "Flight_delays"
WHERE (origin IN ('MDW','ORD')) AND (dest IN ('SFO','SJC','OAK')) AND (crs_dep_time>1400)
GROUP BY origin,
         dest
ORDER BY AVG(arr_delay_new) DESC
```

Table 10: 5 records

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