US FLIGHTS ANALYSIS WITH PYTHON

Group - 7

DATA SUMMARY

- US domestic flights dataset for the year 2017
- Source – US Bureau of Transportation
- Contains approx. 6 million records with 70 columns (~2 GB)

DATA CLEANING

- Merging data of each month together
- Replacing null values with appropriate values
- Changing datetimes to proper format (e.g. 1020 → 10:20:00)
- Merging data with geographical coordinates

DATA DESCRIPTION

- DEP_DELAY & ARR_DELAY (Departure & Arrival Delay in mins.)

Negative delay implies early departure/arrival.

- CRS_ELAPSED_TIME (Computerized Reservations System) :

Elapsed Time = Airtime + Taxi In + Taxi Out

- TAXI_IN :Time elapsed between wheels on and gate arrival at the destination airport

- TAXI_OUT :Time elapsed between departure from origin airport gate and wheels off

INSIGHT 1

Would you prefer to choose a particular airline if you knew 45% of its flights are delayed?

Probably not!?

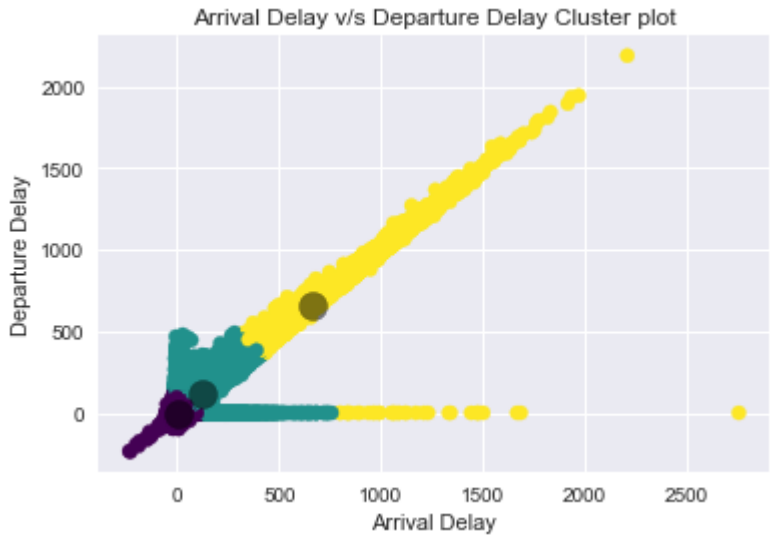
INSIGHT 1

Using clustering, we found out that:

- Even though some flights have delayed departures, they still arrive on time or in many cases even before the scheduled arrival time!

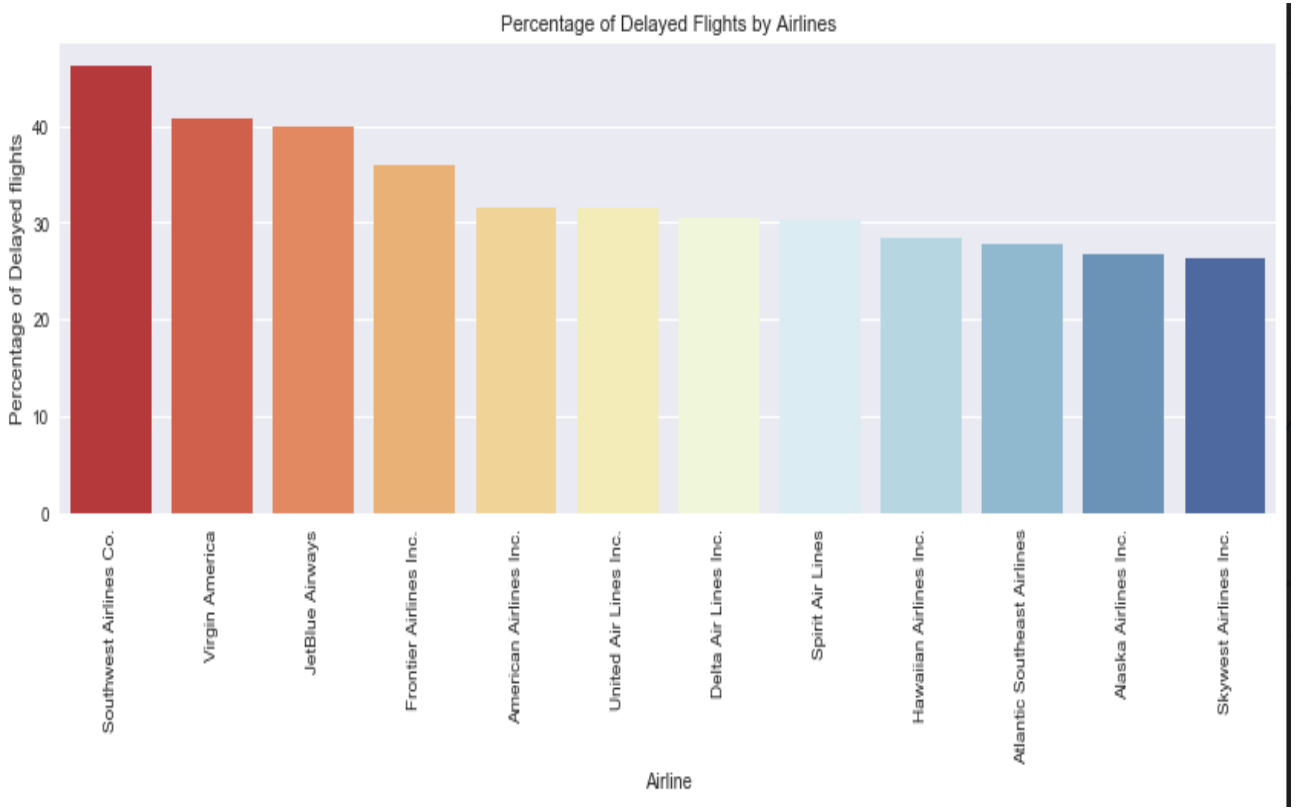
- Reason?

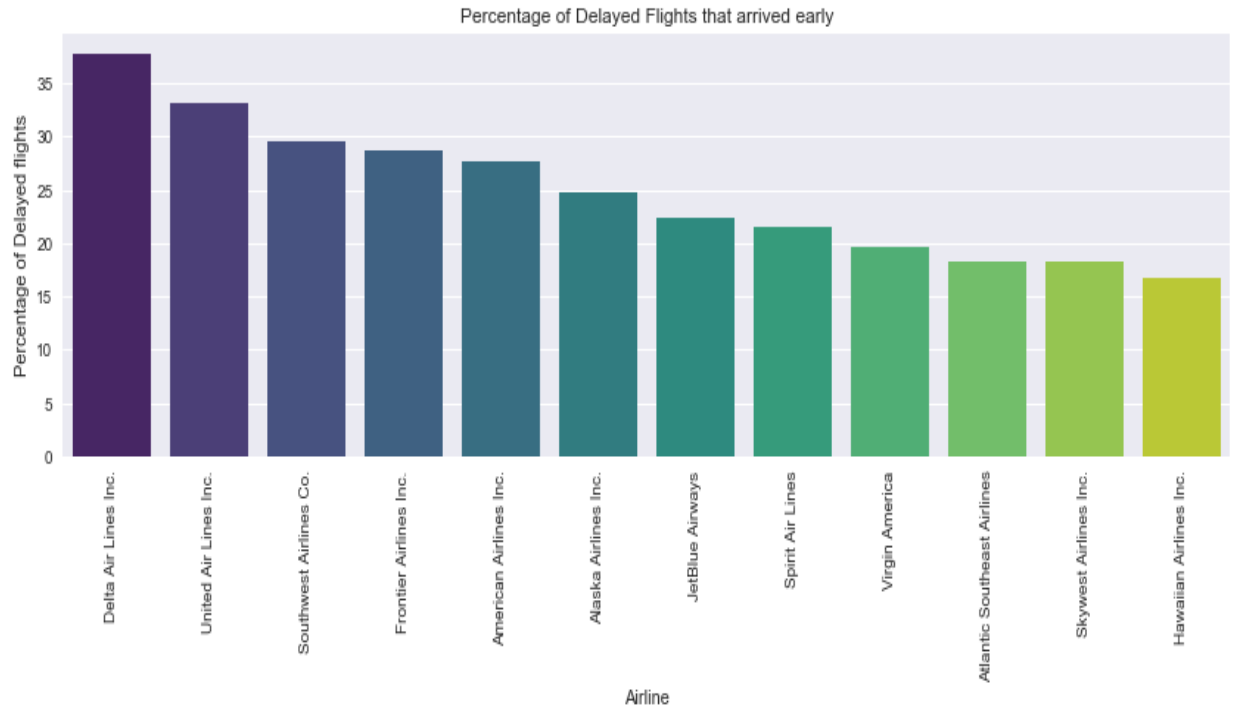
→CRS time adjustment



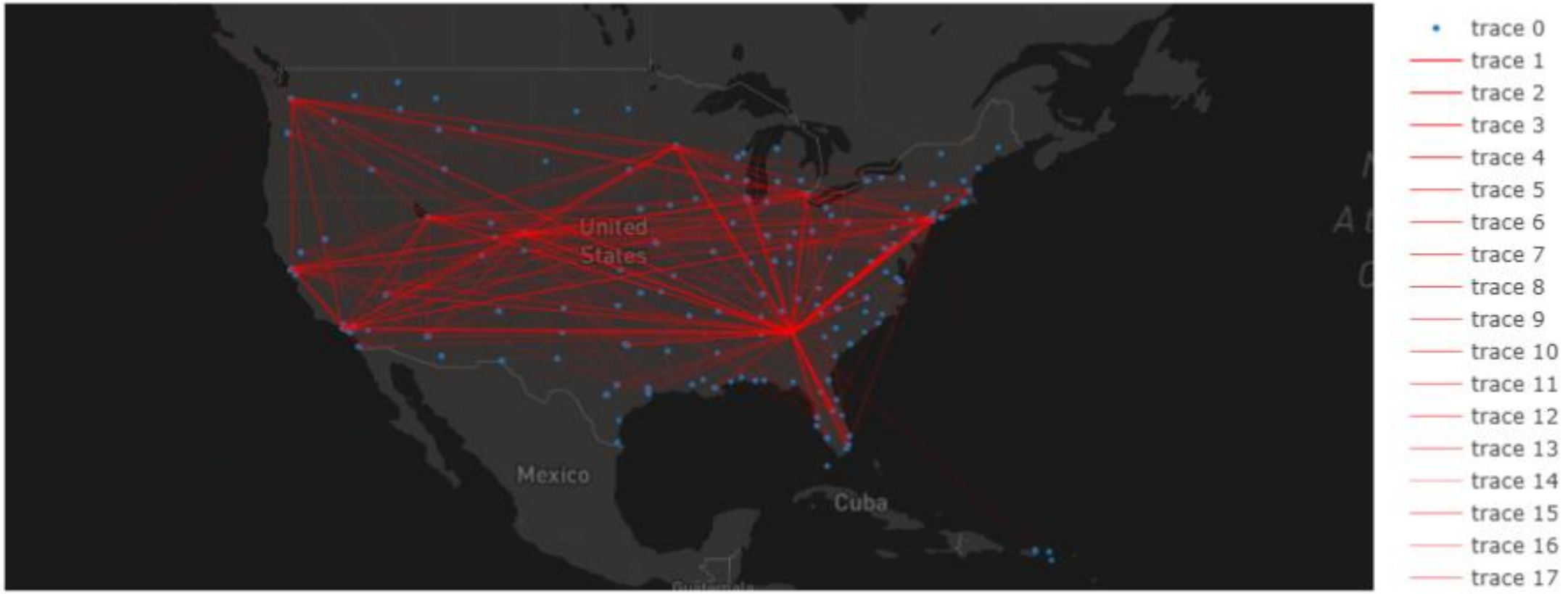
cluster	DEP_DELAY	ARR_DELAY
0.0	1.849170	-3.701989
1.0	121.278625	119.624648
2.0	659.669278	651.941197

INSIGHT 1


Percentage of Delayed Flights by Airlines
Percentage of Delayed flights
0
10
20
30
40
Southwest Airlines Co.
Virgin America
JetBlue Airways
Frontier Airlines Inc.
American Airlines Inc.
United Air Lines Inc.
Delta Air Lines Inc.
Spirit Air Lines
Hawaiian Airlines Inc.
Atlantic Southeast Airlines
Alaska Airlines Inc.
Skywest Airlines Inc.
Airline

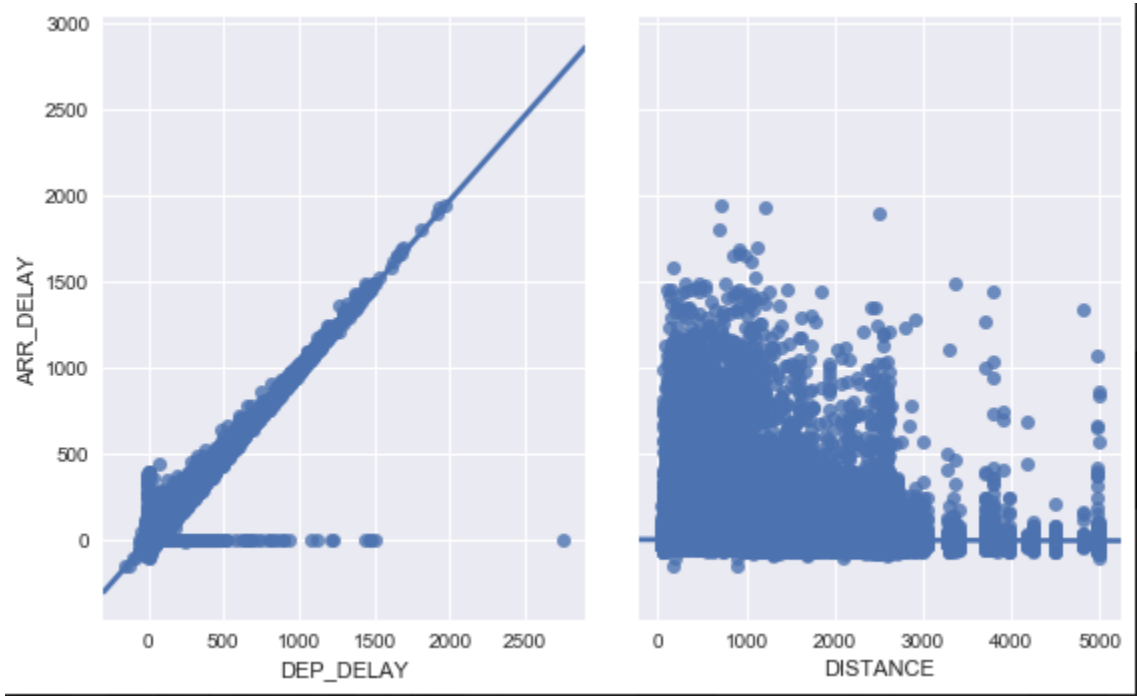

Percentage of Delayed Flights that arrived early
Percentage of Delayed flights
0
5
10
15
20
25
30
35
Delta Air Lines Inc.
United Air Lines Inc.
Southwest Airlines Co.
Frontier Airlines Inc.
American Airlines Inc.
Alaska Airlines Inc.
JetBlue Airways
Spirit Air Lines
Virgin America
Atlantic Southeast Airlines
Skywest Airlines Inc.
Hawaiian Airlines Inc.
Airline

Southwest and Delta Airlines Delayed but Early Arrival Flights



INSIGHT 2

- Predicting arrival delays using regression models.



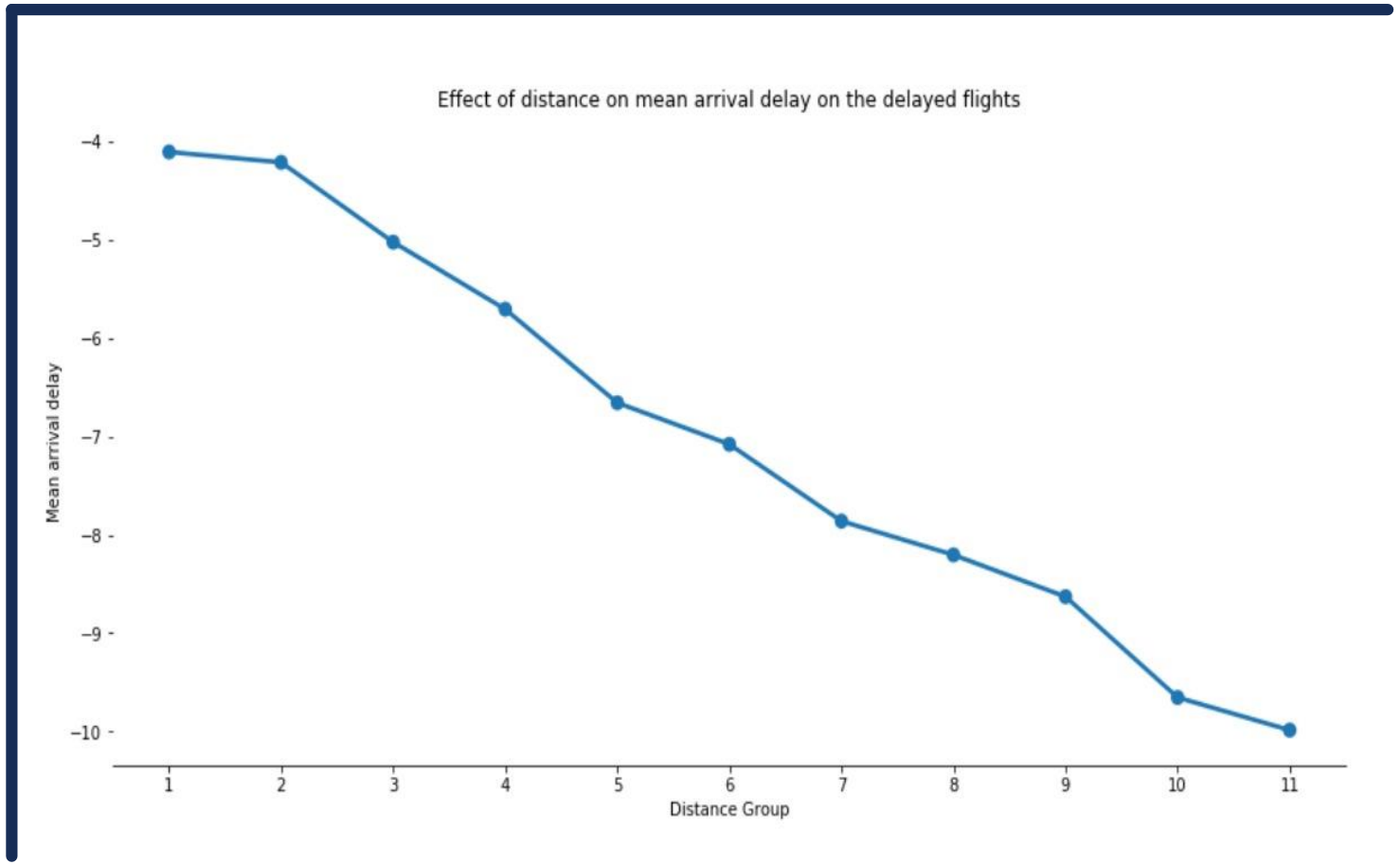
Steps overview:

1)Used Lasso to find important attributes for prediction.
2)Trained and predicted delays using Lasso Regression Model. (MAD=4.33 mins)
3)Trained and predicted delays using Linear Regression Models. (MAD=3.80 mins)
4)Used Linear Regression Model to predict 2018 quarter 1, delays. (MAD = 3.885 mins)

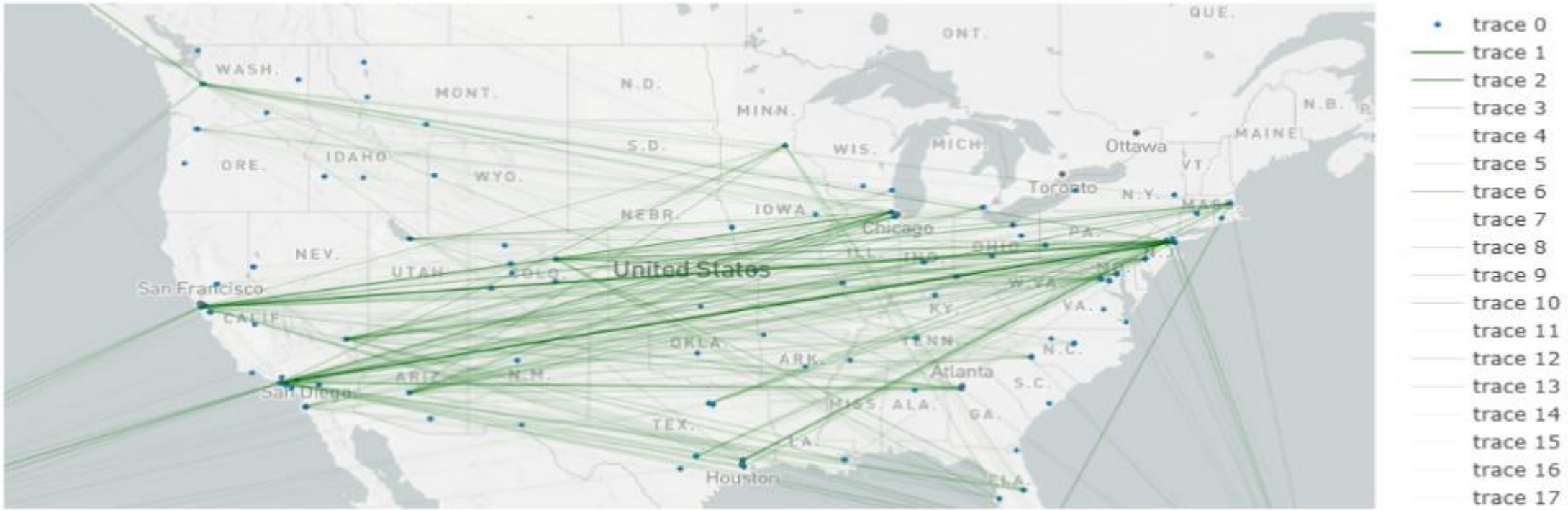
INSIGHT 3

- Shorter distance flights have comparatively more delays than longer distance flights!

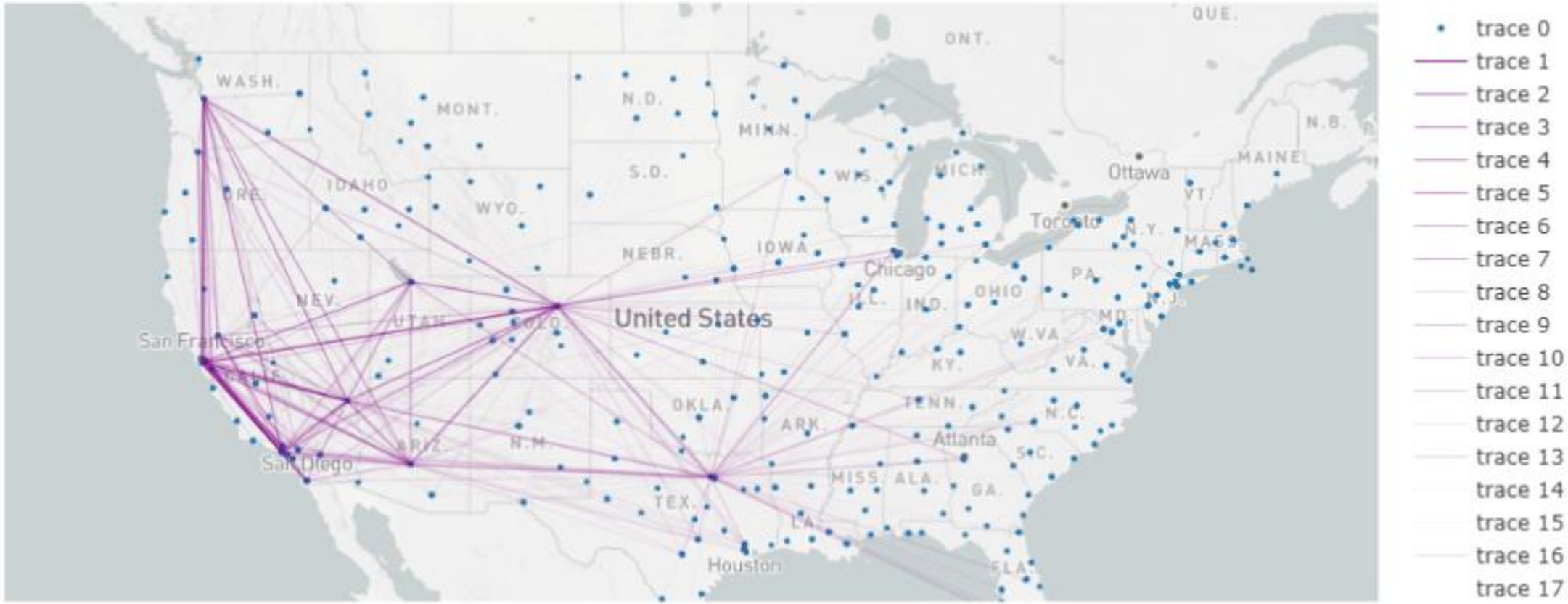
cluster	DEP_DELAY	ARR_DELAY	DISTANCE
0.0	9.109806	5.611746	293.301731
1.0	9.554211	2.171681	1624.293096
2.0	10.101546	3.947220	1021.161403
3.0	9.214863	3.974665	631.592727
4.0	11.406771	3.380153	2440.436203



Long Distance Delayed Flights



Short Distance Delayed Flights

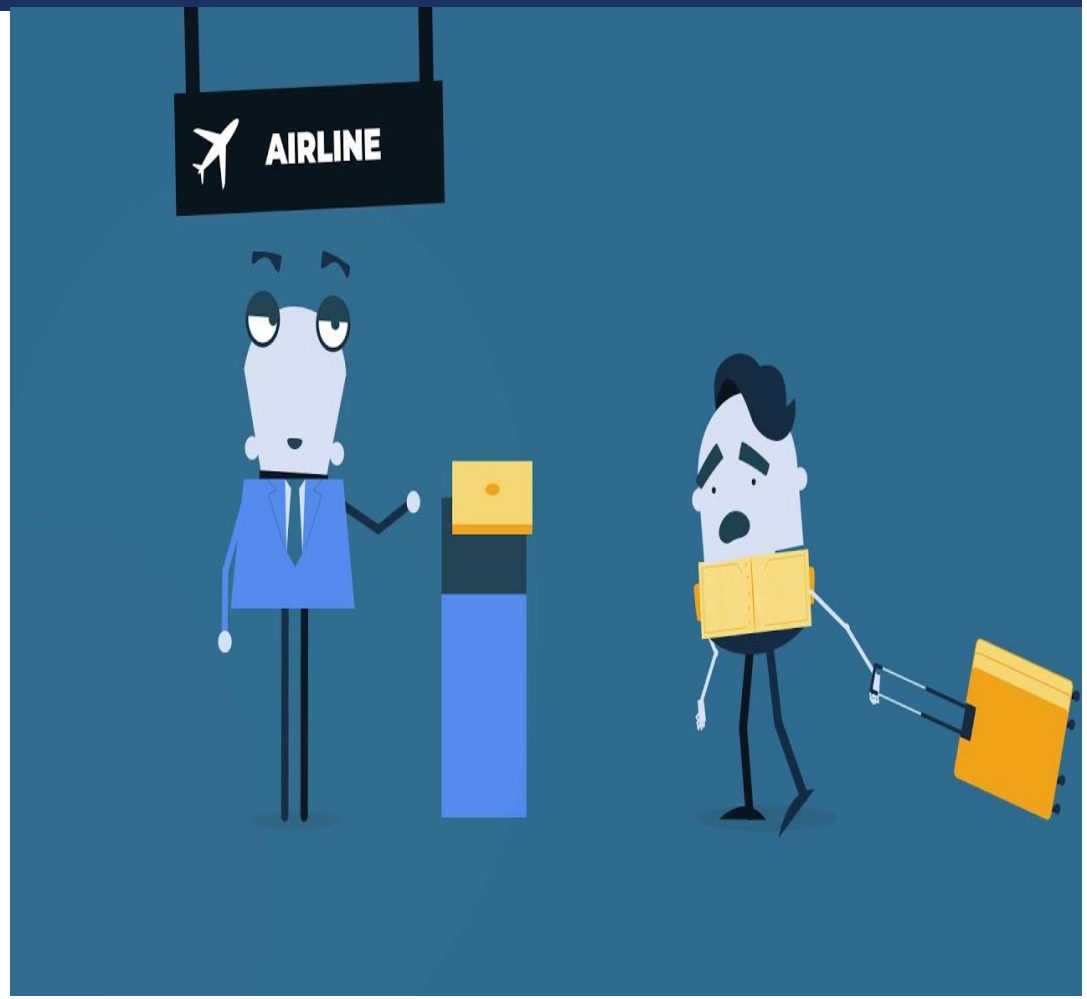


JUST FOR FUN!

- Flights are delayed more in summers than monsoon or winters!

Causes → Low air density makes it difficult for planes to take off.

- Flights are delayed more during weekdays than weekends.



THANK YOU!