Universitetet
i Stavanger

Finding Flight Delay Trends

DAT500 Project Group 17

Bhakti Prabhu & Stephan F. W. Brandasu

Faculty of Science and Technology

Why?

the use case

As a traveller:

1. Not miss any important meetings/events/functions.
2. Pre-plan journey
3. Have idea of buffer time while flight booking

As an airline:

1. know when to increase workforce.
2. Opportunity to improve over competitors.

What?

the dataset

Sourced from the United States Bureau of Transportation statistics.
The dataset is a total of 5GB for 3 years of data

```
FL_DATE 1/1/2022 12:00:00 AM
OP_UNIQUE_CARRIER 9E
OP_CARRIER_AIRLINE_ID 20363
OP_CARRIER 9E
TAIL_NUM N138EV
OP_CARRIER_FL_NUM 4732
ORIGIN_AIRPORT_ID 10135
ORIGIN_AIRPORT_SEQ_ID 1013506
ORIGIN_CITY_MARKET_ID 30135
ORIGIN ABE
DEST_AIRPORT_ID 11433
DEST_AIRPORT_SEQ_ID 1143302
DEST_CITY_MARKET_ID 31295
DEST DTW
ARR_DELAY -15.00
ARR_DELAY_NEW 0.00
CANCELLED 0.00
CARRIER_DELAY
WEATHER_DELAY
NAS_DELAY
SECURITY_DELAY
LATE_AIRCRAFT_DELAY
```

Raw data → structure → information

Map Reduce

Spark

focus - ability to merge add data to existing table

Note: delay is measured in minutes and cancelled is 0 for no and 1 for yes
Note: some fields don't always contain data, this was just the first entry from the dataset

https://www.transtats.bts.gov/

How?

high level plan

1. Finding High probability of flight delays in a period
    - Which days in a month
    - Which month in a year
    - Which quarter in a year
2. Find Trend in Flight Delays from year to year