

## Airline Use Case

### Problem Statement

- Top 10 busiest round-trip routes in terms of number of round-trip flights
- Top 10 most profitable round-trip routes
- 5 route recommendations for investment
- Required trips to reach ROI for the 5 investment aircrafts
- KPIs to be tracked

### Scope

- 1 metadata file
- 3 data files

### Assumptions

- Airport\_Codes.csv is incomplete dimension data is will be used as-is. Missing information will be considered normal.
- Tickets.csv data sample size will be considered source of truth in the ticket pricing stand-point. Extending the sample size will skew the output and is expected.
- Flights.csv is transactional data and data type inference will be based solely on the metadata file.
- Airport costs are fixed. Anything that isn't large or medium will be assigned a zero-value to avoid data type issues.
- Operational costs per unit (mile, min) are fixed
- Aircraft costs are fixed at \$90M and will be the basis for ROI. Depreciation and inflation will not be considered.
- Flight data between airports are static. Reroutes, layovers, and emergency incidents are not considered.

### Approach

- Create a python scrip for data transformation and plotting
- Generate proposal data for investments
- Create visuals as needed per problem

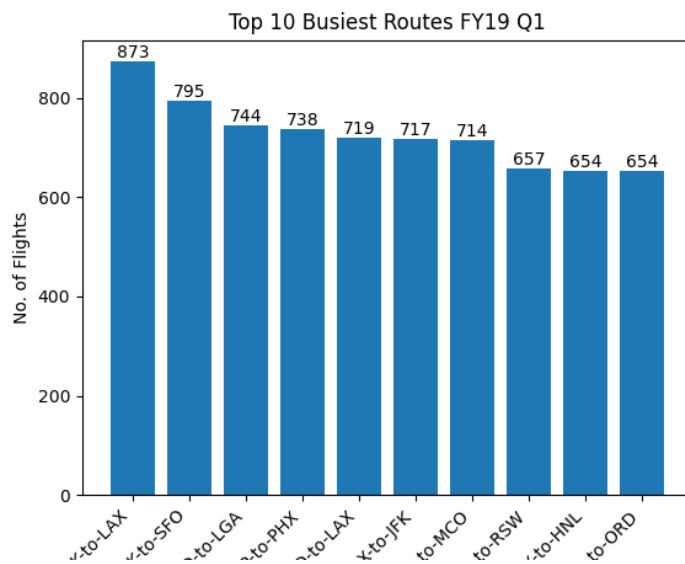
### Solution

- Files (in zipped folder)
  - *data\_usecase.ipynb* – main code, created in Jupyter Notebook browser version
  - *data\_usecase\_sourcecode.py* – backup python script
  - *profit.csv* – dataset for top 10 profitable routes
  - *investment.csv* – dataset for 5 investment routes

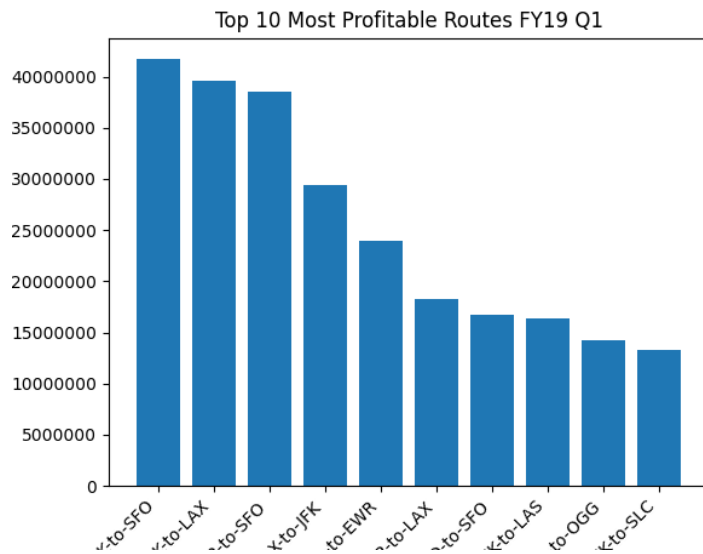
- Schema created for solution

| Field              | Source            | Description   |
|--------------------|-------------------|---|
| ROUTE              | n/a               | String label to identify origin and destination combination                 |
| ORIGIN             | Tickets.csv       | IATA Code of departure airport  |
| DESTINATION        | Tickets.csv       | IATA Code of arrival airport  |
| ROUNDTrip          | Tickets.csv       | Total round-trip flights per route  |
| AVG_FARE           | Tickets.csv       | Average ticket fare fee for all round-trip flights going that route         |
| CANCELLED          | Flights.csv       | Total cancelled flights per route   |
| DELTA              | n/a               | Total completed flights per route   |
| AIRPORT_FEE        | Airport_Codes.csv | Fixed airport fee based on size of origin and destination airports          |
| OVERHEAD_COST      | Flights.csv       | Total cost based on distance (miles) of each route                          |
| OPERATION_COST     | Flights.csv       | Overhead cost plus average delay penalties per route                        |
| AVG_OCCUPANCY_RATE | Flights.csv       | Average passengers (out of 200) per route                                   |
| REVENUE            | n/a               | revenue per flight <passengers> x <fare> - <airport fee> - <operation cost> |
| PROFIT             | n/a               | potential revenue based on foot traffic per route                           |
| ROI                | n/a               | total flights to break-even profit and capital (USD 90M)                    |

Problem 1: Top 10 busiest round-trip routes in terms of number of round-trip flights



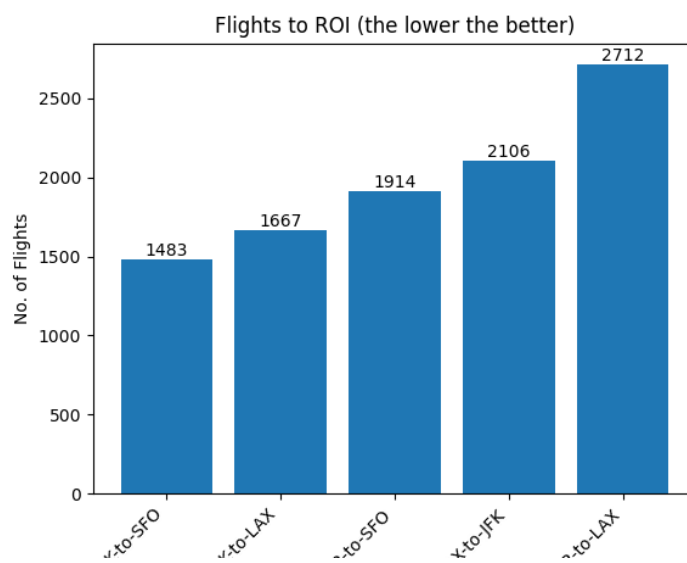
Problem 2: Top 10 most profitable round-trip routes



Problem 3: 5 route recommendations for investment

- JFK-to-SFO, JFK-to-LAX, EWR-to-SFO, LAX-to-JFK, EWR-to-LAX
- The 5 recommended routes were selected based on these factors
  - There are in the top 10 busiest and most profitable routes
  - Only 4 airports to focus on (JFK, LAX, SFO, EWR)
  - 3 of the routes will reach ROI faster

Problem 4: Required trips to reach ROI for the 5 investment aircrafts



Problem 5: KPIs to be tracked

- **No. of flights** – The routes were selected because of high foot traffic which means minimal dips in profit due to off-season market.
- **Total distance travelled** – Since each aircraft has an average lifespan base on mileage, capital cost may spike before ROI is reached.
- **Variable operational cost** – Higher foot traffic can also mean delays. Penalties can add up so making efforts to address this help.