Airline Data Challenge

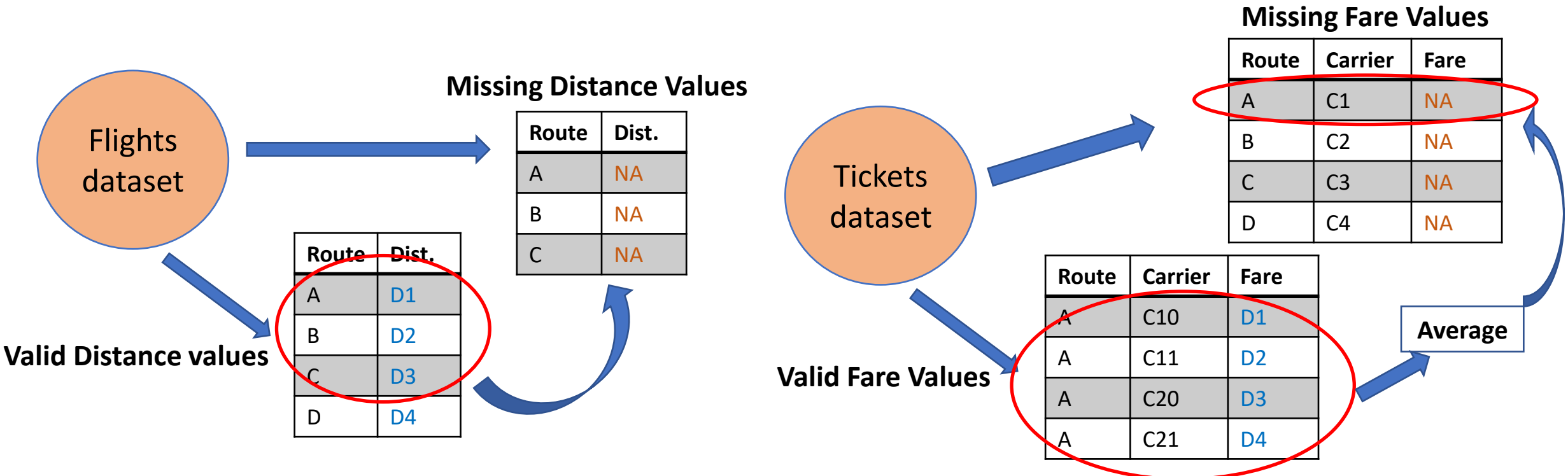
By: Haile Endeshaw

Objectives

- Find the top-10 busiest round-trip routes
- Find the top-10 most profitable round-trip routes, excluding aircraft capital cost
- Identify 5 round-trip routes for investment
- Determine required number of flights to breakeven

Data Cleaning

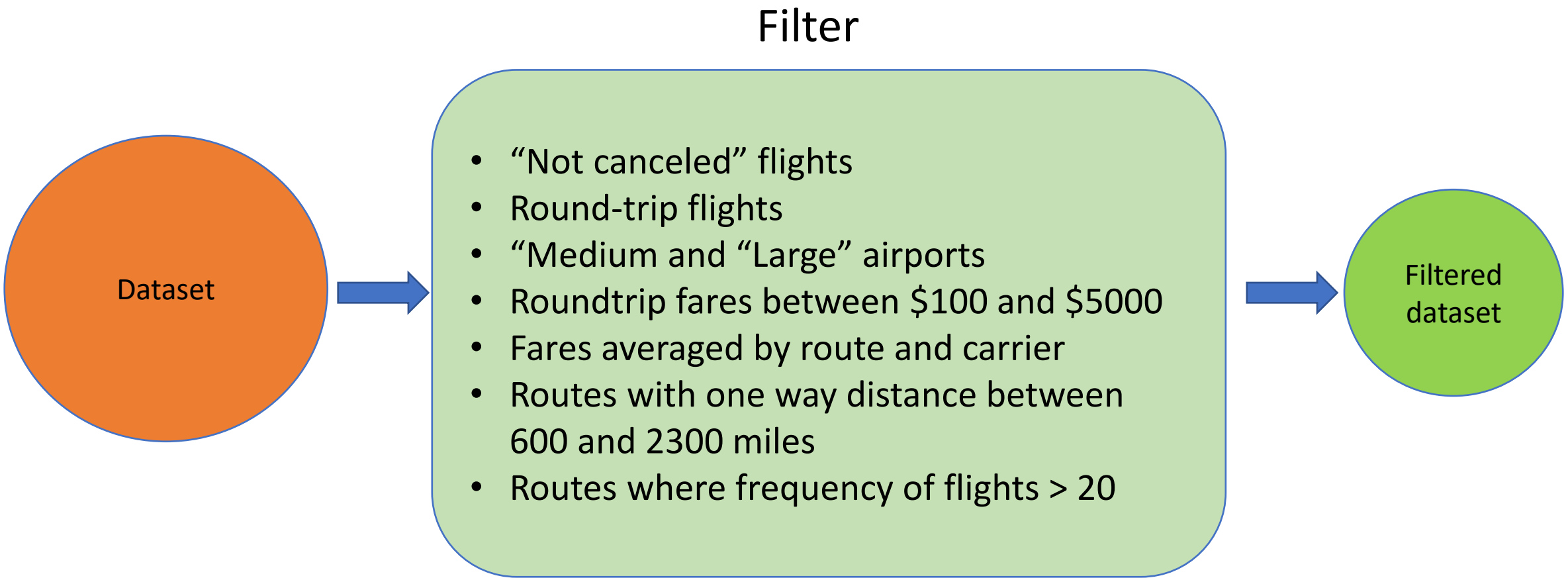
- Missing Data
  - find missing data from other entries
    - *Distance* – from other same route entries
    - *Occupancy rate* – from same route entries, averaged by carrier
    - *Fare* – from same route entries, averaged by carrier



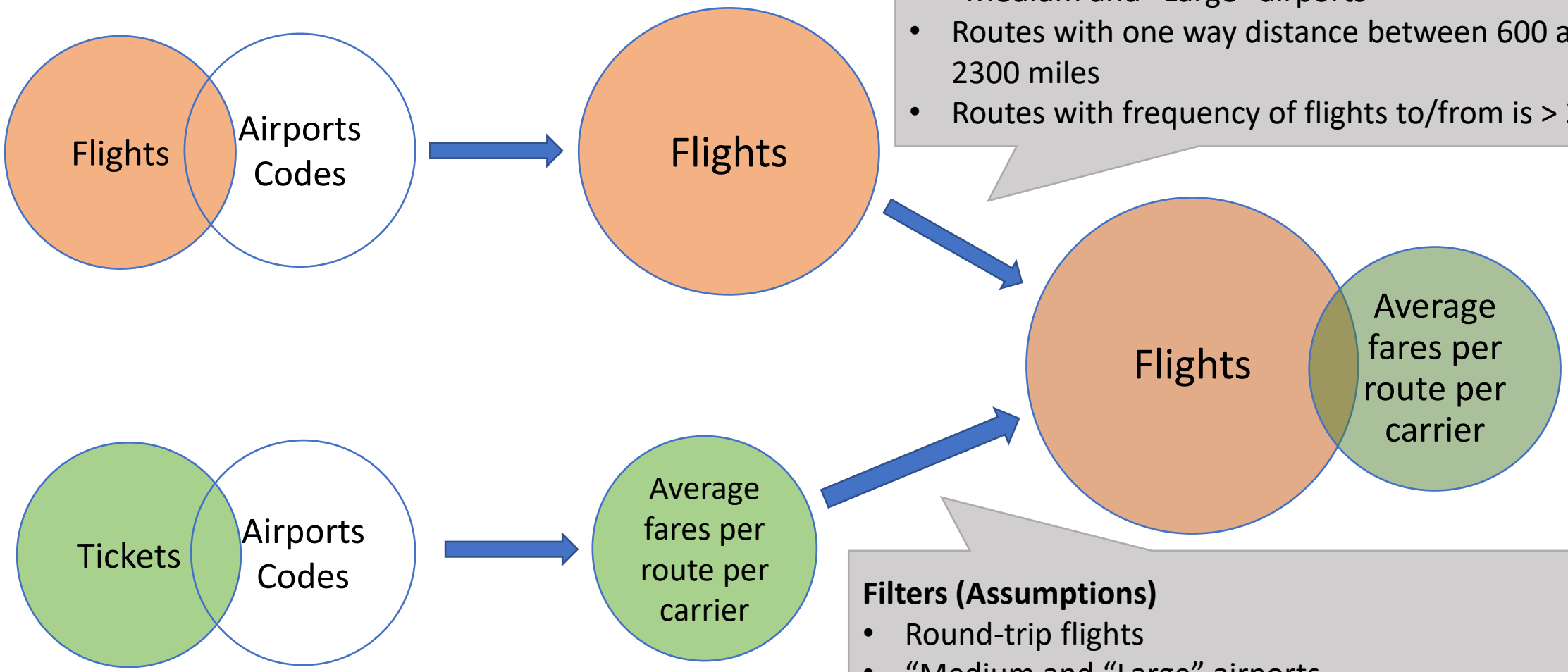
Data Cleaning

- Wrong Data types
  - Conversion of character strings into numeric
- Erroneous data (e.g. negative entries of distance)
  - Find correct value from other entries
  - Eliminate
- Other
  - Double distance data for roundtrip analysis

Assumptions



Joining Datasets



Results: **Busiest Routes**

origin_IATA_code <chr>	dest_IATA_code <chr>	num_routes <int>
SFO	LAX	4176
LAX	SFO	4164
ORD	LGA	3580
LGA	ORD	3576
LAX	LAS	3257
LAS	LAX	3254
LAX	JFK	3162
JFK	LAX	3158
LAX	SEA	2502
SEA	LAX	2497

Results: **Cost & Revenue per Roundtrip**

- Cost
  - *O&M*: fuel, oil, maintenance, and crew ($8/mile)
  - *Airport use*: fixed origin and dest. airport use cost ($5,000 or $10,000)
  - *Delay*: arrival and dep. delays ($75/min for delay > 15 min)
  - *Misc*: insurance, depreciation, etc. ($1.18/mile)

$$Total\ Cost = 2 \times Distance \times (O\&M + Misc) + Airport\_Use + Delay$$

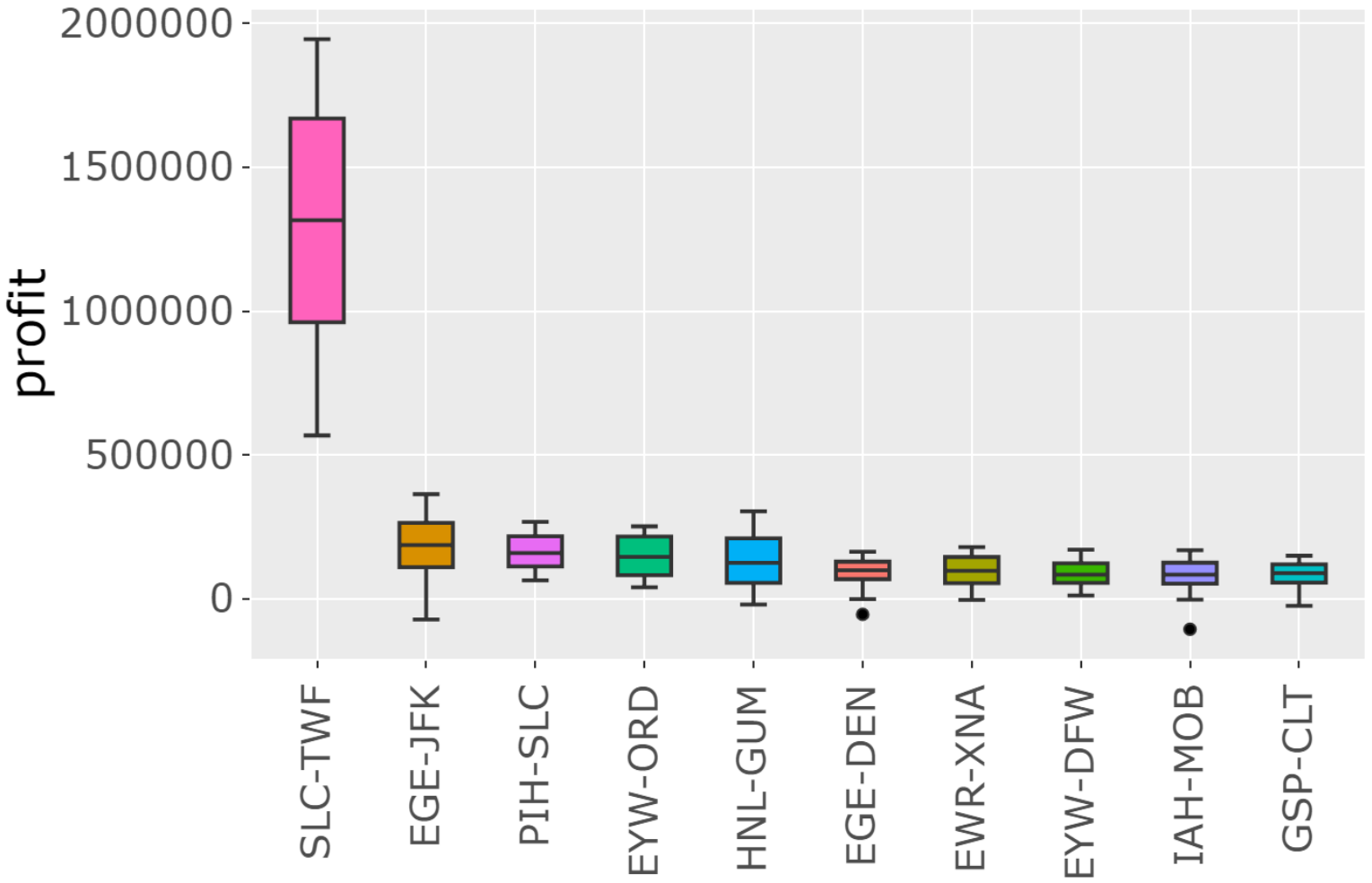
- Revenue
  - *Fare*: roundtrip
  - *Baggage fee:* fixed $70
  - *Occupancy rate* (*OR*)
  - *Capacity*: aircraft capacity = 200

$$Revenue = Capacity\ \times OR \times \left(Fare + \frac{1}{2} Baggage\_Fee\right)$$

- Profit

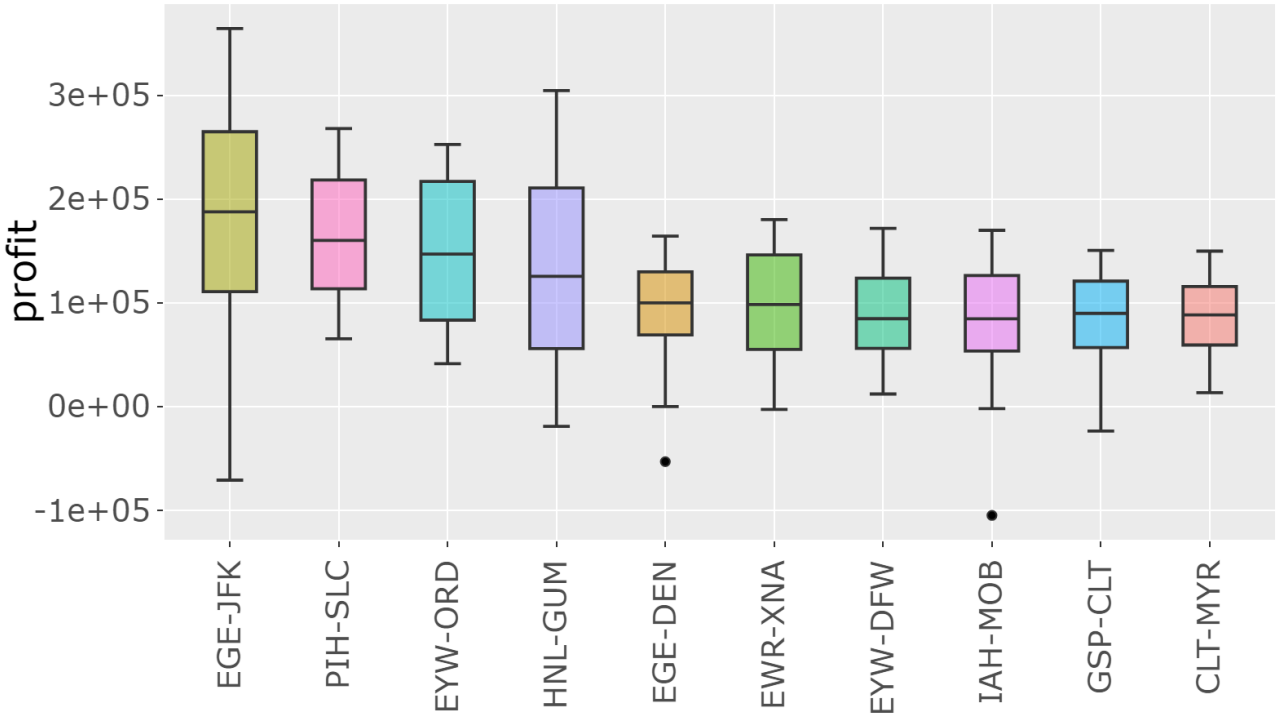
$$Profit = Revenue\ - Cost$$

Results: **Most Profitable Routes**

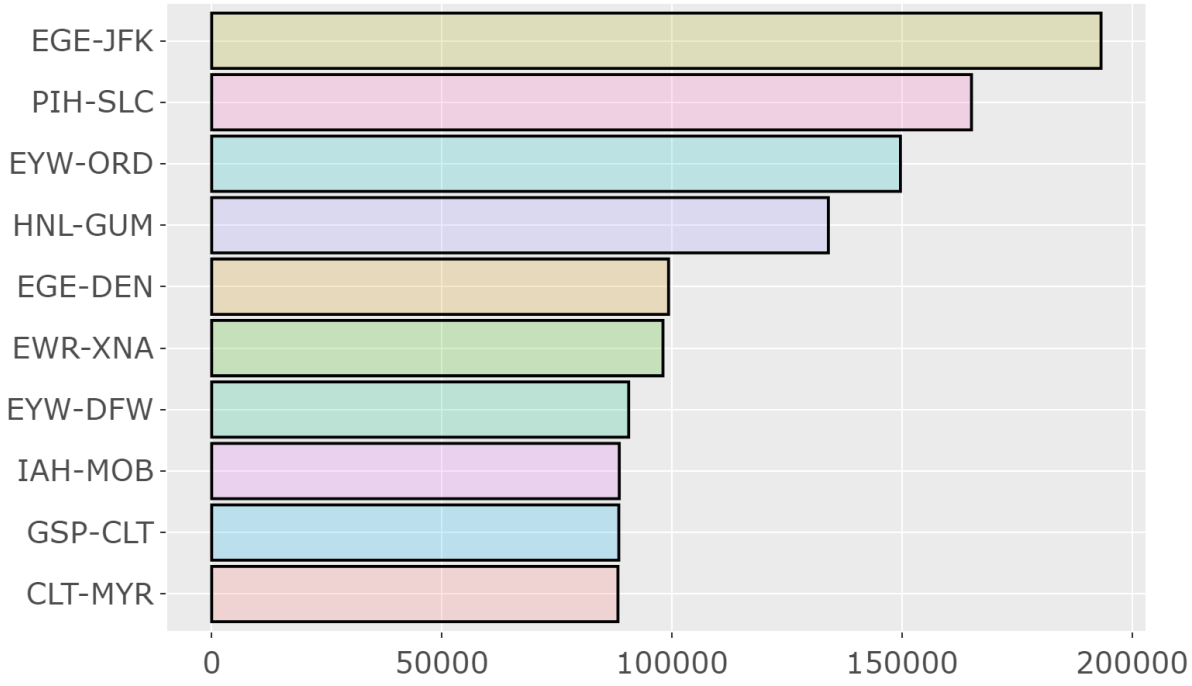


- **The first route has significantly higher profit**
  - Resulting from ***incorrect fare value***

Results: **Most Profitable Routes**



Arranged boxplots of profits in descending order



Most profitable routes by average profit

Results: **# of Flights to Breakeven**

- Number of roundtrip flights to breakeven, $N_{r\_flights}$

$$N_{r\_flights} = \frac{Aircraft\_Capital\_Cost}{Average\_Profit\_Per\_Route}$$

route <chr>	ave_profit <dbl>	n_breakeven_routes <dbl>
EGE-JFK	193260	466
PIH-SLC	165108	546
EYW-ORD	149671	602
HNL-GUM	134005	672
EGE-DEN	99279	907
EWR-XNA	98066	918
EYW-DFW	90615	994
IAH-MOB	88566	1017
GSP-CLT	88471	1018
CLT-MYR	88280	1020

Results: **Key Performance Indicators**

Fare

Distance

Occupancy rate

Number of flights

Recommendations

- **Top-five routes for investment**
  - **EGE-JFK**, (*Eagle county regional, Gypsum, CO - JFK Airport, Queens, NY*)
  - **PIH-SLC**, (*Pocatello regional, Pocatello, ID - Salt lake city Intl, UT*)
  - **EIY-ORD**, (*Key West Intl, KeyWest FL - Chigaco OHare, Chicago, IL*)
  - **HNL-GUM**, (*Inouye Intl, Honolulu, HI - Won Pat Intl, Guam*)
  - **EGE-DEN** (*Eagle county regional, Gypsum, CO - Denver Intl, Denver, CO*)

Future Work

- Number of flights to breakeven

route <chr>	ave_profit <dbl>	n_breakeven_routes <dbl>
EGE-JFK	193260	466
PIH-SLC	165108	546
EYW-ORD	149671	602
HNL-GUM	134005	672
EGE-DEN	99279	907
EWR-XNA	98066	918
EYW-DFW	90615	994
IAH-MOB	88566	1017
GSP-CLT	88471	1018
CLT-MYR	88280	1020

- However, how much ***time*** does it take to breakeven?

Future Work

1. Annual profitability provides better prediction of profitability, rather than profit per route.
2. Time value of money is needed to accurately determine number of flights to breakeven.

Future Work: Annual Profitability

- *Annual profitability* rather than *profit per route*.
- Factors affecting annual profitability
  - Number of flights per year
  - Fare
  - Occupancy rate

route <chr>	ave_profit <dbl>	n_breakeven_routes <dbl>
EGE-JFK	193260	466
PIH-SLC	165108	546
EYW-ORD	149671	602
HNL-GUM	134005	672

$$\text{Occupancy\_rate} = \boldsymbol{f}(\textit{fare, number\_of\_flights})$$

$$\textbf{Annual profit} \sim \boldsymbol{f}(\textit{fare, number\_of\_flights, occupancy\_rate})$$

- Determine the optimum fare and # of flights per year for a maximum profit.

Result: Number of flights/year and fare needed for maximum profit

Future Work: Time Value of Money

- *Without considering time value of money*
  - Number of roundtrip flights to breakeven, $N_{r\_flights}$

$$N_{r\_flights} = \frac{Aircraft\_Capital\_Cost}{Average\_Profit\_Per\_Route}$$

- *Considering time value of money*

$$NPV = -\,aircraft\_capital\_cost + \sum_{t=1}^{N_f} \frac{C_t}{(1+r)^t} = 0$$

- ***Result**: Number of roundtrip flights, N_f, required for NPV = 0. (i.e. breakeven)*

Thank You!