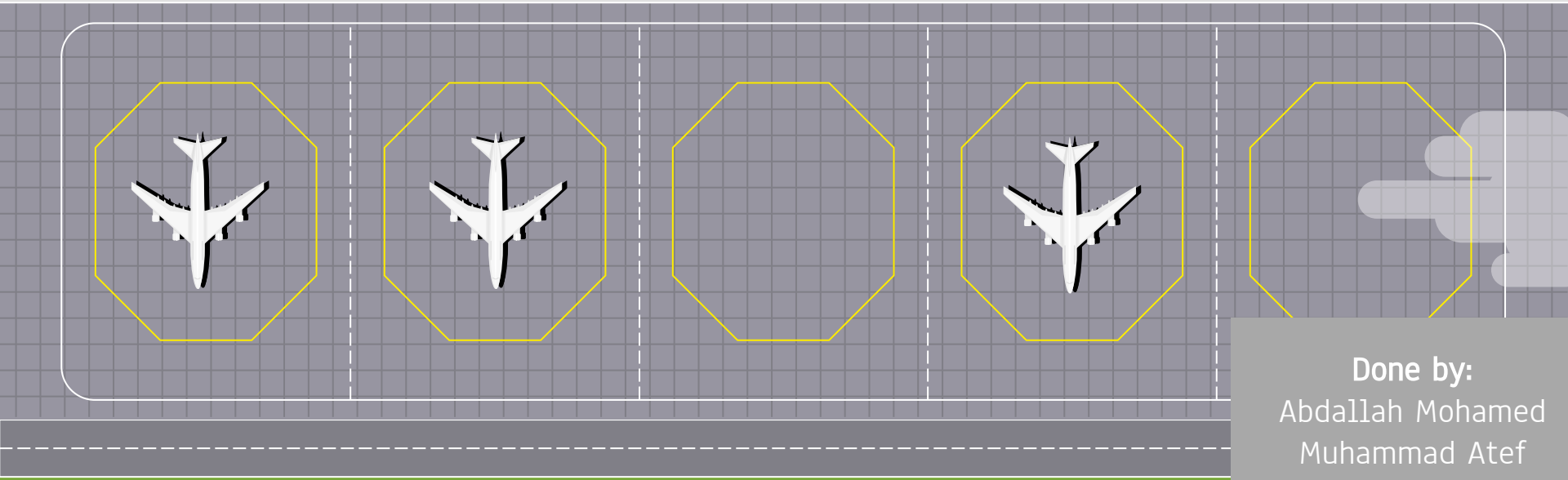
Itinerary Builder



Done by:
Abdallah Mohamed
Muhammad Atef
Youssef Fawzy

Table of contents

01 Overview

02 Data Collection

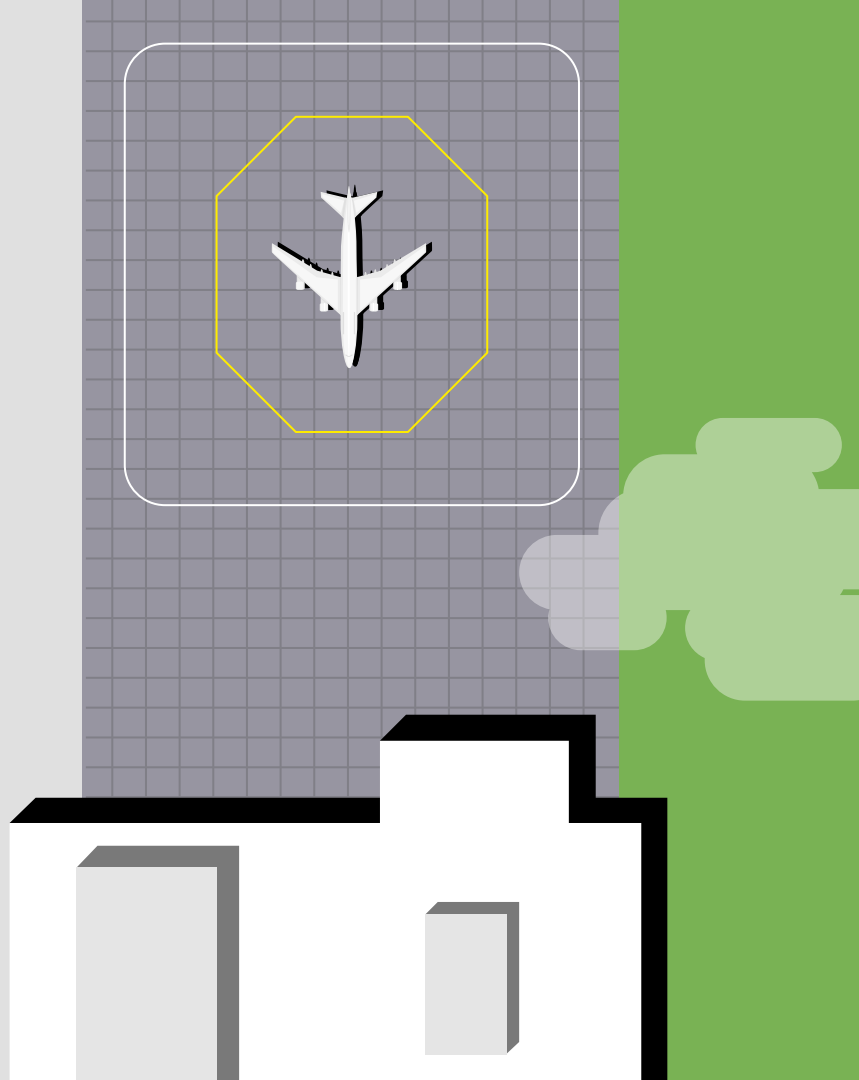
03 Itinerary Builder

04 Market Demand Estimation

05 Regression Coefficients

06 Conclusion

Overview

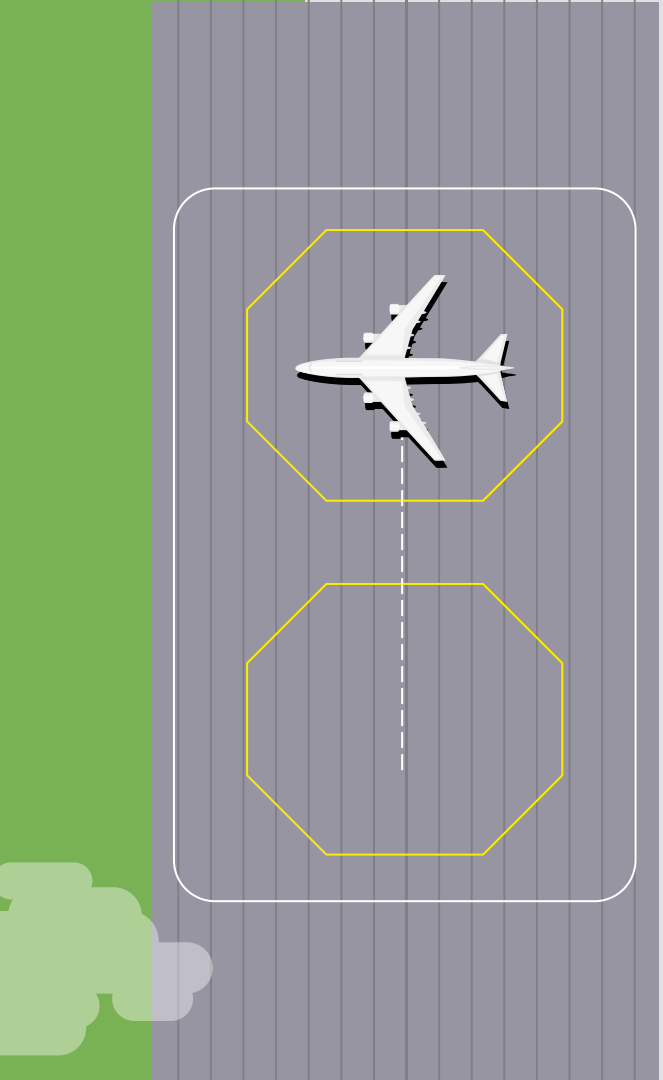


Overview

- **Understand customer behaviour** and preferences in order to make informed decisions about pricing, scheduling to **increase Revenue**
- **Collected data** from flight trackers that provided information on the flights traveling between different cities and airports in Egypt
- Using an **itinerary builder**, different itineraries for each city pair

Overview

- **Estimated demand** between every city pairs
- Calculated the **utility function** and probability of each itinerary to better understand customer preferences
- Analysed **market concentration** using the QSI and HHI indices for each airline and route



Data Collection

Data Collection

- Using FlightRadar24.com, we were able to find all routes, their departure and arrival time, the aircraft used, and the airline operating them
- FlightRadar24 is a website and mobile application that tracks real-time flight information for commercial and private aircraft around the world
- A certain day was chosen so that our model can be made based on true flights done in that single day. We obtained all the flight data for Friday, March 17th as Friday usually has the most domestic flights per week
- For a given origin airport, the entry on Flight Radar showed the flight number, the destination, its departure time, and the aircraft used:

Assiut Airport ATZ / HEAT - 344 km	MS186	6:45 AM B738	-	-	6:45 AM 320	-	-	-	-
Aswan International Airport ASW / HESN - 699 km	MS5	-	-	-	-	-	5:10 PM 320	-	-
	MS80	10:05 PM 738	10:05 PM 738	10:05 PM 738	10:05 PM 320	10:05 PM 738	10:05 PM 738	10:05 PM 32N	10:20 PM 738
	MS82	6:15 AM A320	6:15 AM 738	6:15 AM 738	6:15 AM 320	6:15 AM 738	6:15 AM 32N	6:15 AM 223	6:30 AM 223

Data Collection

- The first step was to collect all domestic flight legs available
- Next, each instance of each flight leg on the chosen day was recorded. The main criteria we were concerned with were:
  - Operating Airline
  - Origin Airport
  - Destination Airport
  - Departure Time
  - Duration
  - Engine type (Jet or Propeller)
  - Aircraft Type
  - Aircraft Capacity
  - Flight leg distance

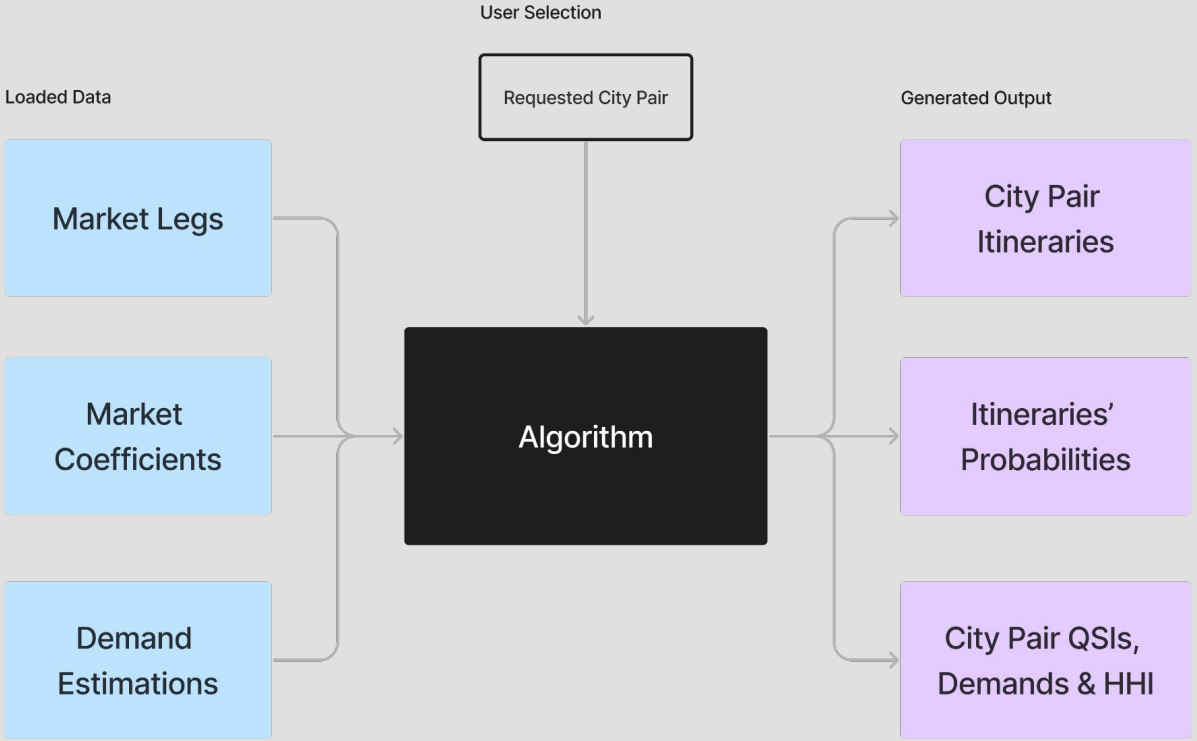
Data Collection

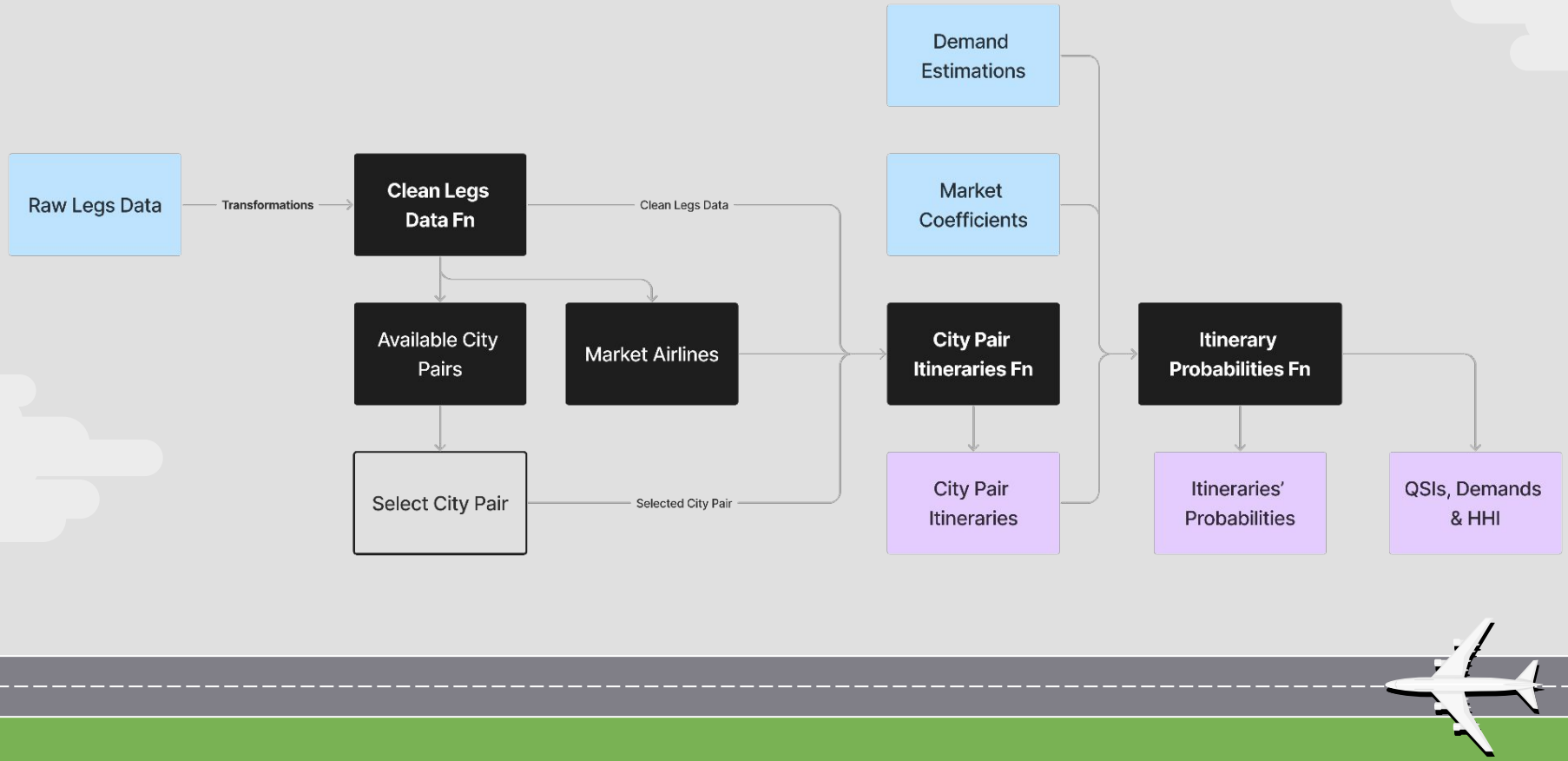
- To help with demand calculation later on, the capacity of each aircraft used was assumed:
  - Boeing 737-800: 162 seats
  - Boeing 737-400: 188 seats
  - Airbus A320: 150 seats
  - Embraer E190: 160 seats
  - ATR 72-600: 70 seats

Data Collection

id	airline_iata	origin_iata	deprt_time	dest_iata	duration	engine	ac_type	capacity	distance
1	FT	CAI	13:00	SSH	1h00min	Jet	B738	162	376
2	FT	SSH	13:50	CAI	1h00min	Jet	B738	162	376
3	MS	ABS	8:30	ASW	0h45min	Jet	B738	162	216
4	MS	ABS	10:05	ASW	0h45min	Jet	B738	162	216

Itinerary Builder


User Selection
Requested City Pair
Loaded Data
Market Legs
Market Coefficients
Demand Estimations
Algorithm
Generated Output
City Pair Itineraries
Itineraries' Probabilities
City Pair QSIs, Demands & HHI


Demand Estimations
Raw Legs Data
Transformations
Clean Legs Data Fn
Clean Legs Data
Market Coefficients
Available City Pairs
Market Airlines
City Pair Itineraries Fn
Itinerary Probabilities Fn
Select City Pair
Selected City Pair
City Pair Itineraries
Itineraries' Probabilities
QSIs, Demands & HHI

Clean Legs Data Fn

Raw Legs Data

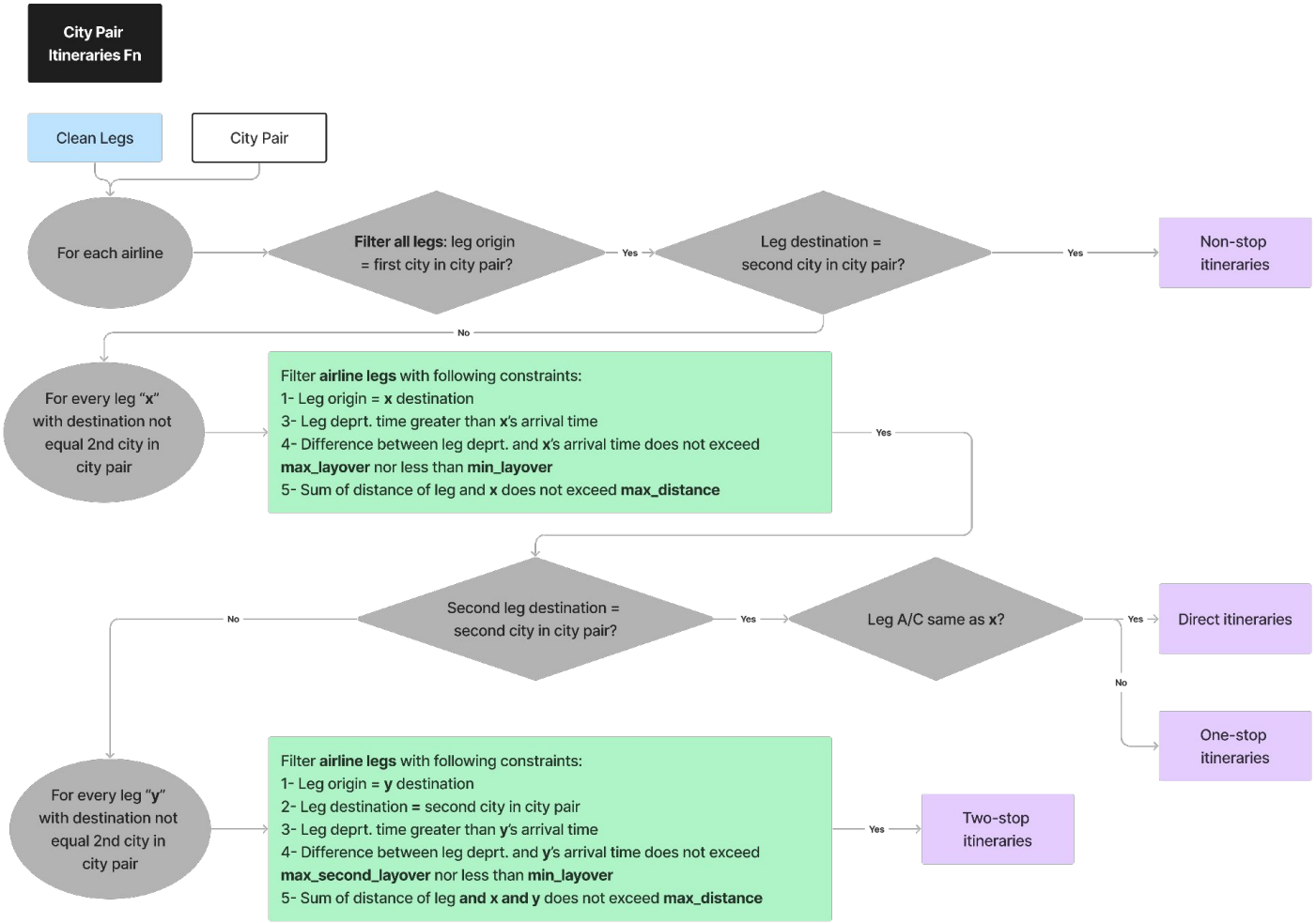
Rename Columns

Add “regional” is TRUE if not prop or mainline

Convert duration to timedelta instead of string

Covert deprt_time to datetime with a dummy date

Clean Legs Data


City Pair Itineraries Fn
Clean Legs
City Pair
For each airline
Filter all legs: leg origin = first city in city pair?
Yes
Leg destination = second city in city pair?
Yes
Non-stop itineraries
No
For every leg "x" with destination not equal 2nd city in city pair
Filter airline legs with following constraints:
1- Leg origin = x destination
3- Leg deprt. time greater than x's arrival time
4- Difference between leg deprt. and x's arrival time does not exceed max_layover nor less than min_layover
5- Sum of distance of leg and x does not exceed max_distance
Yes
Second leg destination = second city in city pair?
No
Yes
Leg A/C same as x?
Yes
Direct itineraries
No
One-stop itineraries
For every leg "y" with destination not equal 2nd city in city pair
Filter airline legs with following constraints:
1- Leg origin = y destination
2- Leg destination = second city in city pair
3- Leg deprt. time greater than y's arrival time
4- Difference between leg deprt. and y's arrival time does not exceed max_second_layover nor less than min_layover
5- Sum of distance of leg and x and y does not exceed max_distance
Yes
Two-stop itineraries

General Constraints

The following are constraints utilized for filtering all possible itineraries:

- max_first_layover = 240 min
- max_second_layover = 240 min
- min_layover = 30 min
- default_max_distance = 1200 km
- max_distance_factor = 1.7
- regional_jet_capacity = 110

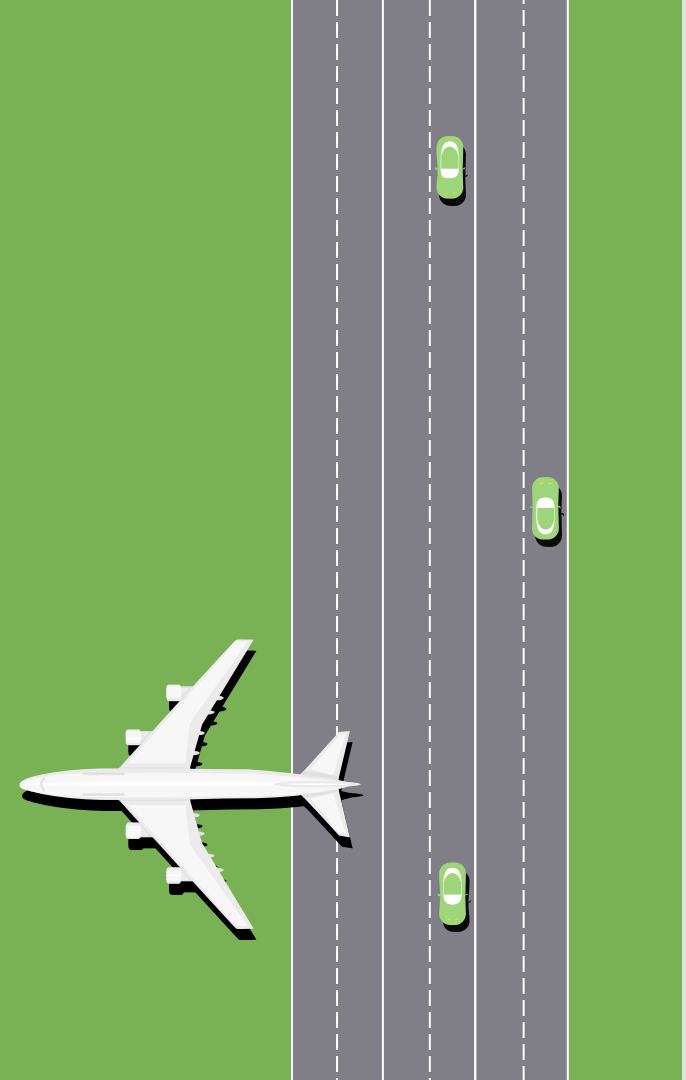
Technology

- **Python**: Pandas, Numpy
- **Flask** framework & HTML/CSS
- The main two functions used to generate the required output are ***get_city_pair_itineraries*** and ***get_city_pair_probabilities***.

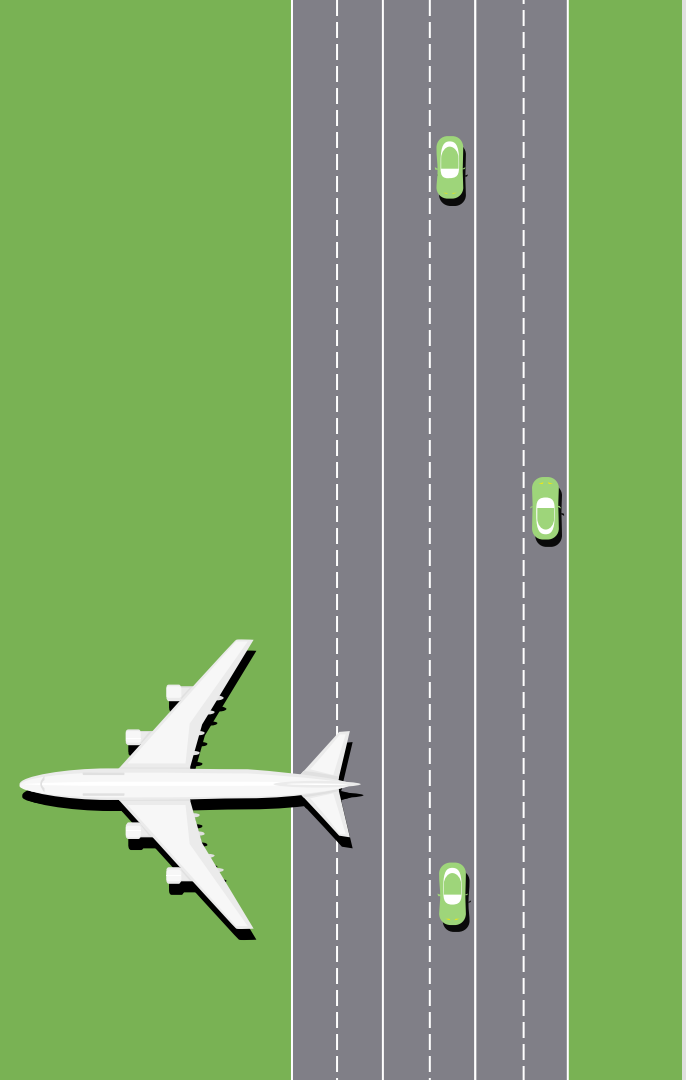
Technology

- *get_city_pair_itineraries*
  - Inputs:
    - routes_dataframe, city_pair_selected, list_of_airlines
  - Outputs:
    - dataframe containing all possible itineraries across the city pair

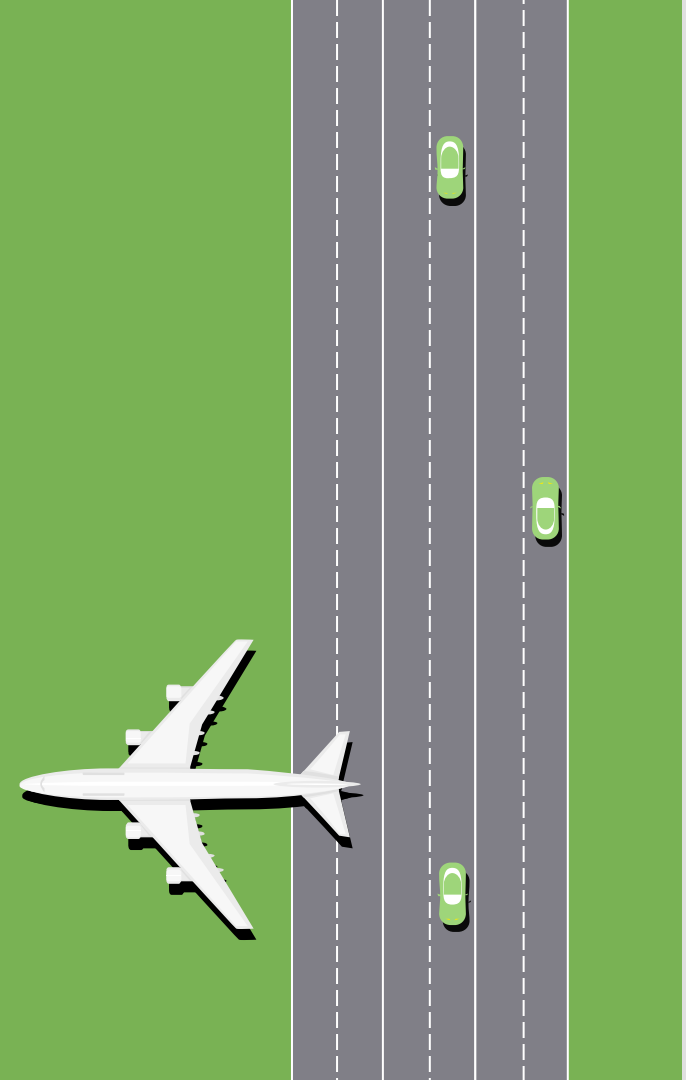
- *get_city_pair_probabilities*
  - Inputs:
    - city_pair_itns, coefficients, demand
  - Outputs
    - itn_score, itn_probability, airlines_qsi, hhi_score



Web App



Data Samples



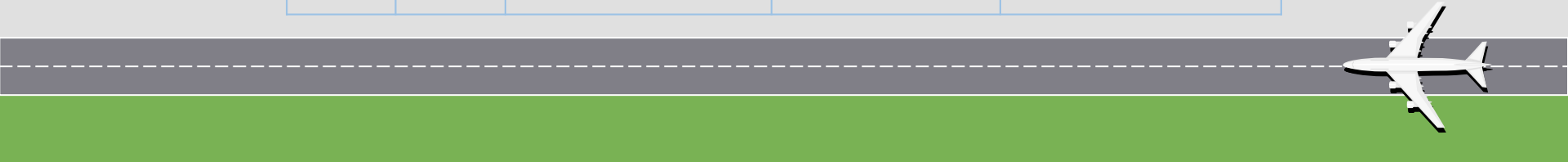
Market Demand Estimation

Estimation Using Supply

- Since the aircraft used in each flight is known, using the capacity assumptions mentioned in the data collection section, we can calculate the number of seats available, or the supply, on each flight leg per day.
- Based on historical data, the average load factor (i.e., the percentage of seats occupied) on domestic flights is around 70-80%. Using this and supply of seats we can estimate the number of passengers on each flight, assuming a load factor of 75%.

Estimation Using Supply

From	To	Number of fights	Seats Available	After Load Factor
ASW	CAI	34	4964	3723
ATZ	HMB	1	162	122
CAI	ASW	35	4918	3689
LXR	CAI	24	2880	2160

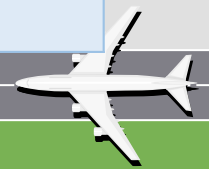


Estimation Using Population

- Population of each city with an airport is obtained
- Since most of these cities are popular tourist destinations as well, the number of tourists per year is also needed, because our model is assumed to be repeated daily, this number is divided by 365 for an average of tourists per day
- Estimating the fraction of the general population that would travel from a given airport each day, assumed at 0.05%
- Estimating fraction of tourists that would take a domestic flight per day, assumed at 0.25 flights per day

Estimation Using Population

City	Population	Population after factoring	Tourists per Year	Tourists per Day	Estimated Total Demand
ASW	1568000	784	500000	1370	1127
ATZ	4472000	2236	0	0	2236
CAI	22183000	11092	2000000	5479	12462
HBE	5588000	2794	10000	27	2801



Estimation Using Population

- This gives us an estimation of how many people are going to travel out of each airport per day. In order to translate this into demand on each flight leg, we need more information.
- Obtain a ratio for each flight leg to multiply by the demand for passengers to travel from a certain origin
- To illustrate, we know that 80 domestic flights leave Cairo International Airport per day. We also know that there are 9 flights from Cairo to Hurghada per day. Therefore, we can assume that the number of passengers going from Cairo to Hurghada in that day is 9/40 multiplied by the estimated number of passengers travelling from Cairo airport.
- The final total estimated demand for each flight leg to be used in analysis is the average of the two methods.

Estimation Using Population

From	To	After Load Factor	Number of flights from origin	Number of Flights	Ratio	Origin Estimated Total Demand	Estimated Demand
CAI	ASW	3689	80	35	0.4375	12462	5452
CAI	ATZ	113	80	1	0.0125	12462	156
HRG	CAI	800	10	7	0.7	2285	1600

Regression Coefficients

Regression Coefficients

ORIGINAIRPORTCODE	DESTAIRPORTCODE	DEPTIME	Market	Market ID
ABQ	DAL	1425	ABQ_DAL	1
ABQ	IAH	1136	ABQ_IAH	2
ABQ	MCI	1338	ABQ_MCI	3
ABQ	LAS	1925	ABQ_LAS	4
ABQ	IAH	1453	ABQ_LBB	5
ABQ	LBB	1637	ABQ_DFW	6
ABQ	DFW	1817	ABQ_CVG	7
ABQ	CVG	1520	ABQ_ELP	8

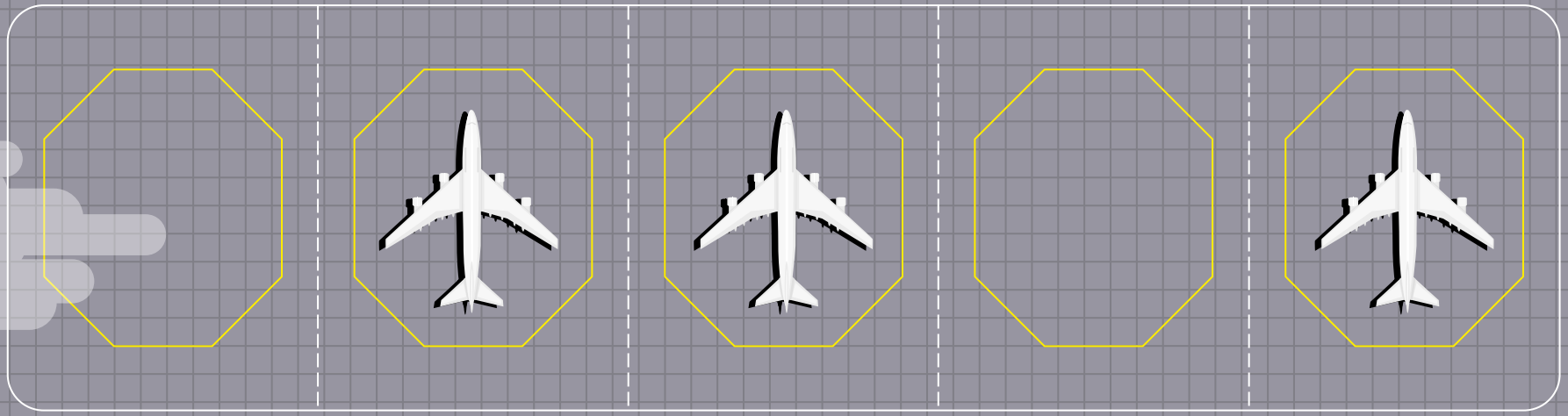
Multinomial logistic regression using **sklearn**

Time	Betas
Midnight-5Am	-0.03025
5-6 AM	-0.19321
6-7 AM	-0.01278
7-8 AM	-0.03137
8-9 AM	0.242108
9-10 AM	0.105657
10-11 AM	0.173145
11-12 Noon	0.069667
12-1 PM	0.119791
1-2 PM	0.177996
2-3 PM	0.065702
3-4 PM	0.155365
4-5 PM	0.109184
5-6 PM	0.197573
6-7 PM	0.024266
7-8 PM	0.244358
8-9 PM	0.221075
9-10 PM	0.404148
10-Midnight	0.313265

Conclusion

Future Work

- **Expand** the itinerary builder to include international routes and more airlines.
- Incorporate **additional factors** into the demand forecasting model, such as seasonality, holidays, and special events.
- Conduct a sensitivity analysis to test the impact of different input parameters on the output itineraries.
- Explore **machine learning techniques** to improve the accuracy of the demand forecasting model and the efficiency of the itinerary builder.
- **Collaborate** with airlines and travel agencies to integrate the itinerary builder into their booking systems.



Thank you!

Any Questions?

Data Sample

airline_iata	origin_iata	deprt_time	dest_iata	duration	prop	ac_type	capacity	distance	regional
SM	ABS	01/01/2023 11:40	HMB	0 days 06:20:00	1	ATR	70	1176	0
SM	ABS	01/01/2023 11:40	HMB	0 days 06:20:00	0	ATR	70	1176	1
MS	ASW	01/01/2023 06:50	ABS	0 days 00:45:00	0	B738	162	216	0
MS	ASW	01/01/2023 07:15	ABS	0 days 00:45:00	0	B738	162	216	0

Algorithm and Implementation

The algorithm developed takes the following as inputs:

1. Flight legs (routes) in market:
   In the case of this project, the data utilized was the collected domestic Egyptian market data.

2. Market coefficients:
   Those are the coefficients used in the utility function to generate the itinerary score for each itinerary.

3. Estimated demands:
   Estimated demand across each city pair is inputted to calculate final demands of each itinerary and airline in the city pair market.

Algorithm and Implementation

The algorithm output in summary is:

1. All possible city pair itineraries
2. Each itinerary's probability and demand in this city pair
3. Each airline serving this city pair's QSI and demand in this market
4. HHI of the market