Cairo University
Faculty of Engineering

Credit Hours system

Egypt Domestic Demand Forecasting

Airtransport system analysis

Presented to : *Dr. Mohamed Lotfy*
Eng. Mariam Amin

Name	Code
Youssef Fawzy	1180023
Abdallah Mohamed	1190418
Muhammad Atef	1190343

Contents

List of figures

List of tables

Abstract

This report presents the results of a study on demand forecasting for domestic flights in Egypt. Using data from flight trackers, we estimated the demand for each flight leg and every possible itinerary using an itinerary builder. We also calculated the utility function and probability of each itinerary to better understand customer preferences. Additionally, we analysed market concentration using the QSI and HHI indices for each airline and route. Our findings provide valuable insights into customer behaviour and preferences, which can help airlines and other stakeholders in the domestic flights market to make more informed decisions about pricing, scheduling, and other important factors.

Introduction

The domestic flights market in Egypt is a rapidly growing and competitive industry. With more than 17 million passengers traveling domestically each year [1], airlines and other stakeholders need to understand customer behaviour and preferences in order to make informed decisions about pricing, scheduling, and other factors. This report aims to provide such insights by presenting the results of a study on demand forecasting for domestic flights in Egypt.

To estimate demand, we collected data from flight trackers that provided information on the number of passengers traveling between different cities and airports in Egypt. Using an itinerary builder, we estimated the demand for each flight leg and every possible itinerary. We also calculated the utility function and probability of each itinerary to better understand customer preferences.

Additionally, we analysed market concentration using the QSI and HHI indices for each airline and route. This allowed us to identify areas where competition may be particularly intense or where there may be opportunities for growth.

Overall, our study provides valuable insights into demand forecasting for domestic flights in Egypt.The rest of this report will provide a more detailed analysis of our methods, results, and implications.

Data collection

The first step of the process was to find all scheduled domestic routes in Egypt. Using FlightRadar24.com, we were able to find all routes, their departure and arrival time, the aircraft used, and the airline operating them. To do this, 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.

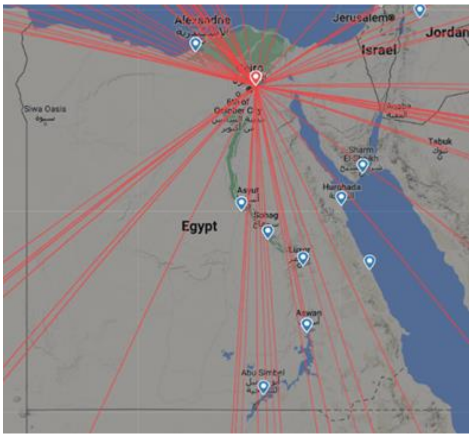

Fig. [1]

FlightRadar24 is a website and mobile application that tracks real-time flight information for commercial and private aircraft around the world. It uses Automatic Dependent Surveillance-Broadcast (ADS-B) technology to receive information about the flight's position, altitude, speed, and other flight data from the aircraft's transponder. FlightRadar24 then aggregates this data and displays it on a map, allowing users to track the flight's progress, view flight information, and receive alerts about flight status changes. FlightRadar24 is widely used by aviation enthusiasts, aviation professionals, and air traffic controllers to monitor flight activity and manage air traffic

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

Fig. [2]

The first step was to collect all domestic flight legs available which were as follows:

Origin	Destination	Origin	Destination
ABS	ASW	HMB	CAI
ASW	ABS	HRG	CAI
ASW	CAI	HRG	LXR
ASW	LXR	HRG	SPX
ATZ	CAI	HRG	SSH
ATZ	HMB	LXR	CAI
CAI	ASW	LXR	HRG
CAI	ATZ	LXR	SSH
CAI	HMB	RMF	CAI
CAI	HRG	SPX	HRG
CAI	LXR	SPX	SSH
CAI	RMF	SSH	CAI
CAI	SSH	SSH	HRG
HBE	HMB	SSH	LXR
HMB	ATZ	SSH	SPX

Table [1]

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

Sample of collected data

	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
5	MS	ABS	11:00	ASW	0h45min	Jet	B738	162	216

Table [2]

Aircrafts capacity

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

Flow chart

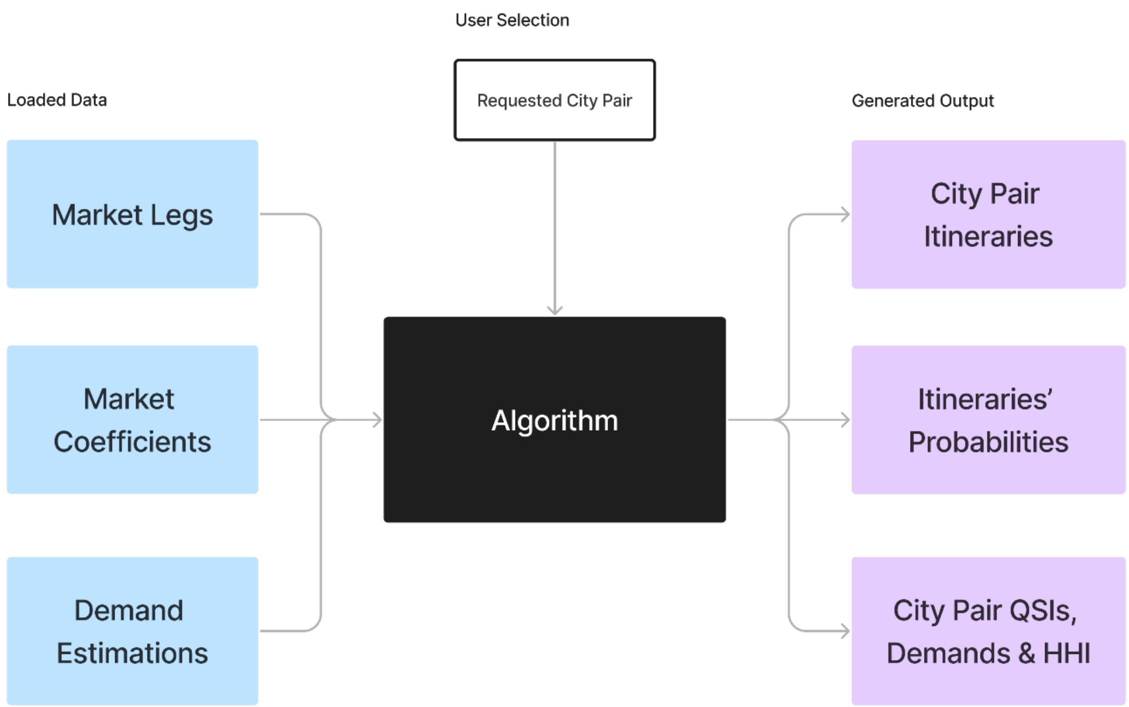

Fig. [3]
Overall system showing inputs [Market legs, Market coefficients, Demand estimations] and outputs [City pair itineraries, Iternaries probabilities, City pair QSIs and HHI]

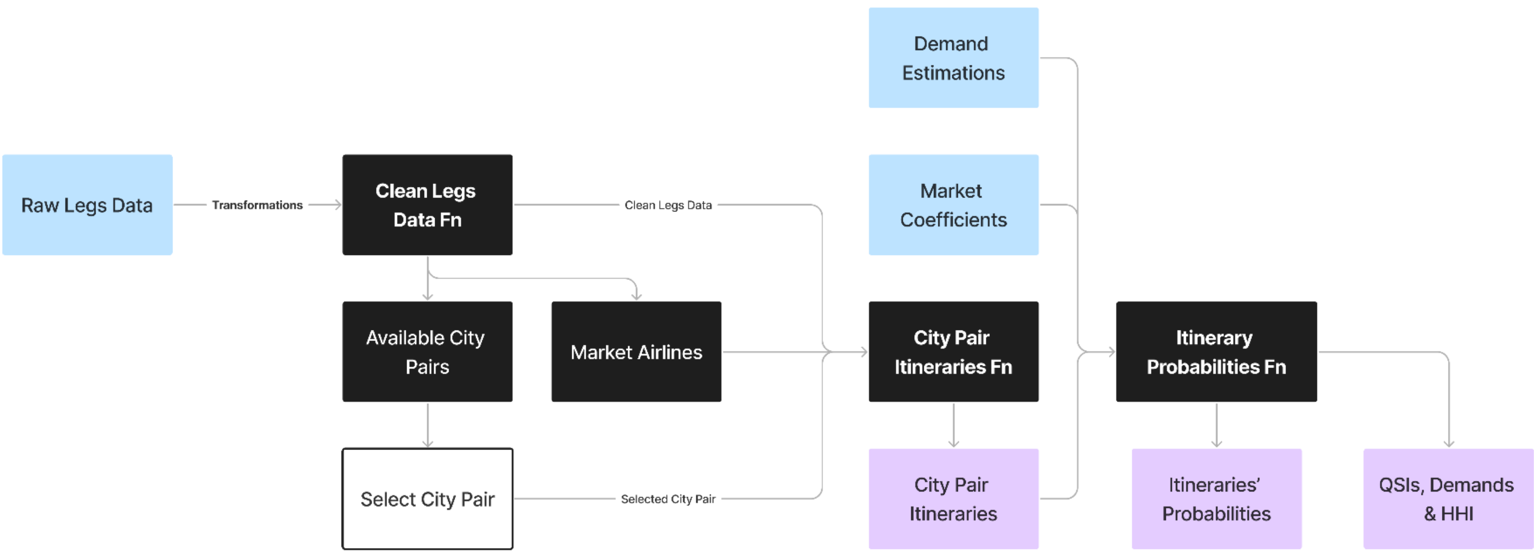

Fig. [4]
Detailed system More detailed flow chart showing the different functions used in black boxes

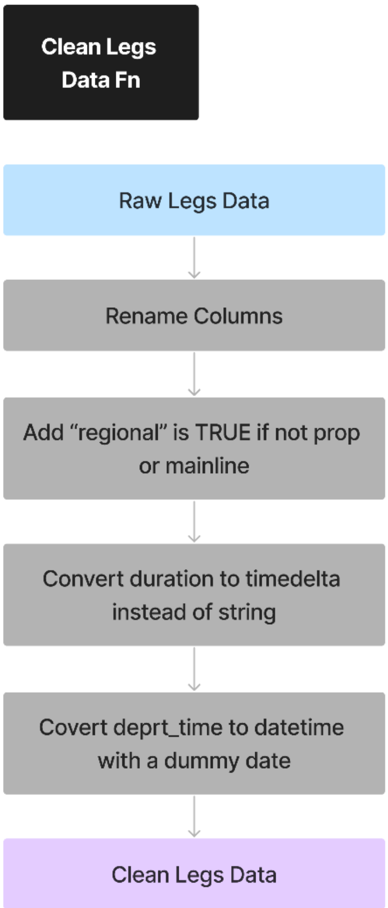

Fig. [5]
Cleaning data Detailed algorithm of Clean Data function

`

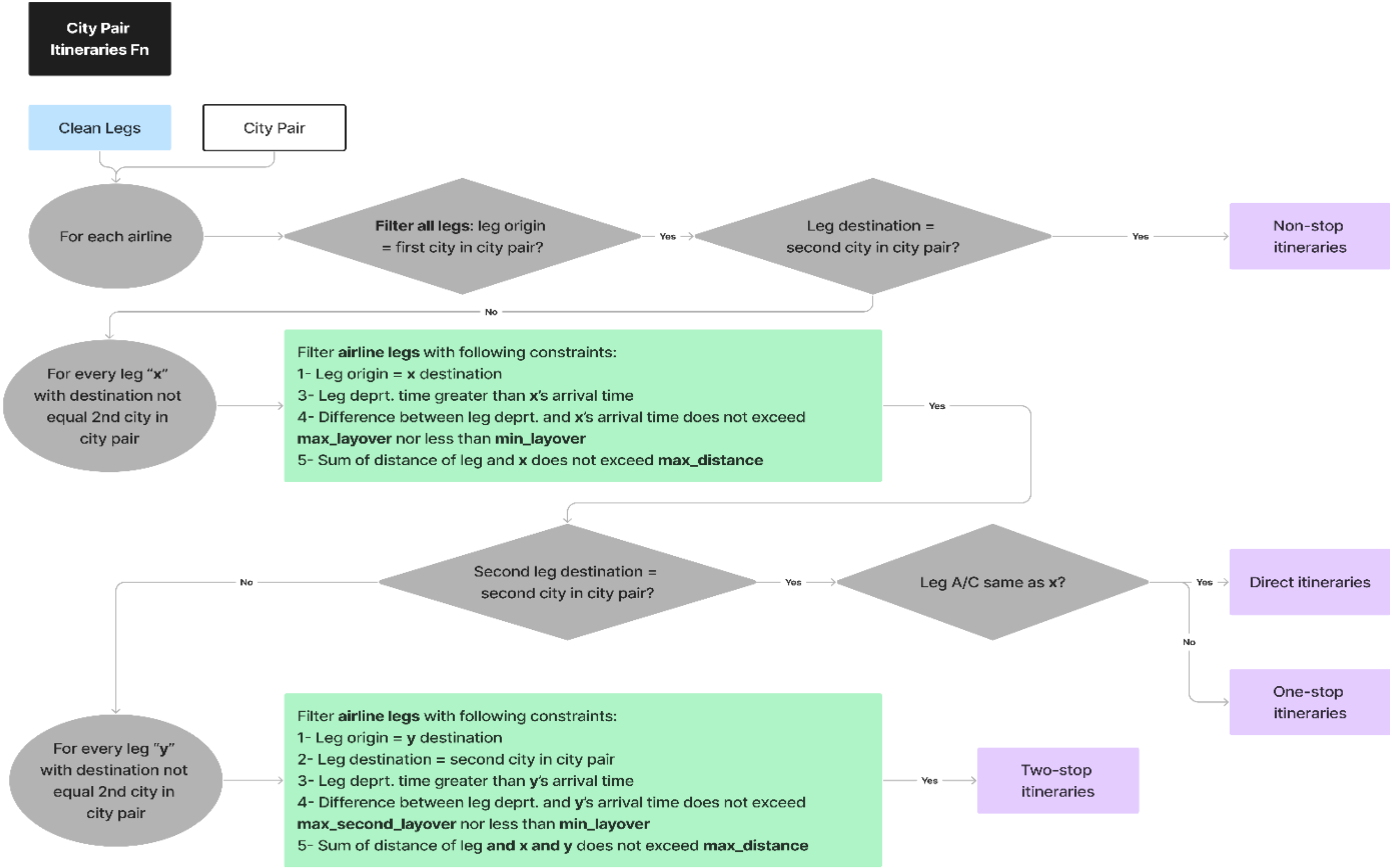

Fig. [6]

City Pair Itineraries Fn Showing detailed algorithm used in this function where for each city pair/market it outputs itineraries with different services

`

Itinerary builder

The benefits of using an itinerary builder are numerous. Firstly, it enables airlines to forecast demand more accurately, which helps them optimize their schedules and reduce operating costs. Secondly, it provides valuable data that can be used to adjust fares and revenue management strategies to maximize revenue. Finally, it gives airlines a competitive advantage by providing insights into passenger preferences and market trends, allowing them to provide a better customer experience and gain market share. In summary, an itinerary builder is a powerful tool that allows airlines to make data-driven decisions, optimize their operations, and increase revenue.

Conditions

The following are conditions used to build itineraries for each city pair:

- 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

Where the maximum layover is longest amount of time that a passenger can spend between connecting flights at an intermediate airport.

Min layover is the minimum time required for a passenger to get off their arriving flight, collect their luggage, go through security checks and get onboard the connecting aircraft, and it can be as short as 30 minutes [3]

Default max distance and max distance factor are assumed as 1200 km and 1.7 respectively to eliminate the generation of unfeasible itineraries for ex: CAI->ABS->SSH.

Regional jet capacity is defined as 110 [4] which is used as an exploratory variable.

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. However, the flight legs data of different markets can be automatically fetched by the means of a web scraper or with integration with an appropriate API.

2. Market coefficients:
   Those are the coefficients used in the utility function to generate the itinerary score for each itinerary. The coefficients are generated using a multinomial logistic regression model fed with flight transactions' data including all the required independent variables. However, due to scarcity of said data, in the scope of this project the coefficients utilized were the East-East US model coefficients generated by Abdelghany and Abdelghany in their text (Abdelghany & Abdelghany, 2019).

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. The algorithm is designed such that if the demand is unavailable it is assumed zero and the rest of the outputs are generated as they're not dependent on it.

The flight legs are used to get a list of all cities in the market, all possible city pairs, all airlines in the market and thus to generate all possible flight itineraries. Once all possible itineraries are generated, the coefficients are used to calculate the probability of each itinerary in a certain city pair. Then, the itinerary probabilities are used to calculate the QSI of each airline in that city pair. Finally, the QSIs of the airlines can calculate the HHI score of the chosen city pair's market. Hence, 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

Implementation: Technology & File Structure

The algorithm is implemented using Python and the user interface is implemented using the Flask framework and HTML/CSS. Pandas & Numpy libraries are used to allow all data to be manipulated as dataframes and arrays which allows for vectorized functions in order to increase execution speed.

The file structure tends to the main algorithm and the accompanying user interface as follows:

Itinerary_builder.py	Main algorithm script
app.py	User interface app script
templates/home.html	Homepage of app
static/style.css	App styling sheet

Implementation Details

The main two functions used to generate the required output are ***get_city_pair_itineraries*** and ***get_city_pair_probabilities***.

get_city_pair_itineraries:

This function takes as input: *routes_dataframe, city_pair_selected, list_of_airlines* and generates a dataframe containing all possible itineraries across the city pair as an output.

The function begins by looping through each of the airlines in the provided list. For each airline, the function filters the routes dataframe to get the airline's routes. Then, the routes dataframe is filtered for routes that begin at the chosen city pair's origin and end at the chosen city pair's destination. For routes that achieve all three conditions, those would be considered as non-stop itineraries. Then, the itinerary properties are calculated, such as whether the itinerary has a propeller aircraft or not, whether it has a regional jet or not and the arrival time. Afterwards, the non-stop itineraries are appended to a resultant dataframe.

Once done, the program loops again over the original airline's routes dataframe. This time around, it enters another loop over each route that did not qualify as a non-stop route. For each of those routes, the original airline's routes dataframe is filtered again according to the conditions that would allow a route to be paired with the current route selected in the loop to become its second leg for a direct or one-stop itinerary. Once those routes are determined, the itinerary properties are calculated then the itineraries are appended to the resultant dataframe.

Similarly, for the third time, the program loops over other routes that did not qualify as a non-stop itinerary nor a first leg of a direct or one-stop itinerary. For each of those routes, the original airline's routes dataframe is again filtered according to the conditions that would allow a route to become a second leg in a two-stop itinerary. Once those are determined, the program loops again over each first leg & second leg pair and filters the original airline's routes dataframe once again according to the conditions that would allow a route to become a third leg in the two-stop itinerary. Once those are determined, the itinerary properties are calculated and the itineraries are appended to the resultant dataframe.

Finally, the resultant dataframe is returned.

get_city_pair_probabilities:

This function takes as input: *city_pair_itns, coefficients, demand* and returns three outputs: *city_pair_itns, airlines_qsi, hhi_score.*

The function takes the city pair itineraries calculated in ***get_city_pair_itineraries*** function as input and then begins to manipulate the dataframe to produce the required independent variables to calculate the score of each itinerary. Even though the information is previously calculated and present in the itineraries dataframe, such as departure time or itinerary duration as a datetime object and a timedelta object respectively, those information need to be listed in a format that matches the model used to generate the market coefficients that will be utilized. For example, the model that generated the utilized coefficients had the day divided into 1-hour intervals where each interval is represented by a separate independent dummy variable, and its value is true if the itinerary's departure time lies within the interval. Therefore, such transformations are performed and all required independent variables are generated.

Afterwards, the utility function is applied to calculate the score of each itinerary. Then, the exponentials are calculated and used to get the probabilities.

The resultant dataframe is then grouped by the airlines to get the airlines' probabilities and QSI scores. Once calculated, the airlines' probabilities are used to get the market HHI score.

Other utility functions

Other utility functions were also created to aid with the code and modularize the algorithm for ease of debugging and future expansion. For example, functions were created to clean the routes dataframe, get cities in a list, get all possible city pairs in a list, loop over the city pair itineraries function to generate a CSV file with all possible itineraries in all possible city pairs, get the demand of a certain city pair and so on.

`

airline_iata	origin_iata	deprt_time	dest_iata	duration	prop	ac_type	capacity	distance	price_usd	regional
SM	ABS	01/01/2023 11:40	HMB	0 days 06:20:00	1	ATR	70	1176	55	0
SM	ABS	01/01/2023 11:40	HMB	0 days 06:20:00	0	ATR	70	1176	55	1
MS	ASW	01/01/2023 06:50	ABS	0 days 00:45:00	0	B738	162	216	55	0
MS	ASW	01/01/2023 07:15	ABS	0 days 00:45:00	0	B738	162	216	55	0
MS	ASW	01/01/2023 09:45	ABS	0 days 00:45:00	0	B738	162	216	55	0

flight_path	city_pair	arrive_time	flight_legs	service_lvl	first_layover	second_layover	first_transfer_city	second_transfer_city
ABS -> ASW -> CAI -> HMB	ABS_HMB	01/01/2023 18:00	[99, 103, 128]	3	0 days 00:40:00	0 days 01:40:00	ASW	CAI
ABS -> ASW -> CAI -> HMB	ABS_HMB	01/01/2023 18:00	[99, 104, 128]	3	0 days 01:40:00	0 days 01:15:00	ASW	CAI
ASW -> ABS	ASW_ABS	01/01/2023 07:35	[7]	0	0 days 00:00:00	0 days 00:00:00		
ASW -> ABS	ASW_ABS	01/01/2023 08:00	[8]	0	0 days 00:00:00	0 days 00:00:00		
ASW -> ABS	ASW_ABS	01/01/2023 10:30	[9]	0	0 days 00:00:00	0 days 00:00:00		

Table [3]

`

Markets demand

One of the most important factors that airlines must take into consideration when examining the possibility of building a new itinerary is the demand on each flight leg. Since sales data for each airline and route is not publicly available, the demand must be estimated. In our analysis, we estimated demand using two methods.

Method 1: Estimation using Population

For this method, the 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.

Next, we must estimate the fraction of the general population that would travel from a given airport each day. Using data concerning the average income of an Egyptian citizen, it is assumed that 0.05% of any given city's population would travel out of its airport on a domestic flight per day.

Tourists, however, are more likely to travel by air than the average citizen, in order to save time on their trips. On average, a tourist stays in Egypt for 8 days, between at least 2 cities. Assuming that they travel between these 2 cities by air, spending 4 days in each city, we get 0.25 flights per day per tourist.

The final demand estimation is a summation of these values, and is shown in the table below:

City	Population	Population after factoring	Tourists per Year	Tourists per Day	Estimated Total Demand
ABS	2600	1	60000	164	42
ASW	1568000	784	500000	1370	1127
ATZ	4472000	2236	0	0	2236
CAI	22183000	11092	2000000	5479	12462
HBE	5588000	2794	10000	27	2801
HMB	5063000	2532	0	0	2532
HRG	3200000	1600	1000000	2740	2285
LXR	1328000	664	500000	1370	1007
RMF	10000	5	10000	27	12
SPX	9200000	4600	500000	1370	4943
SSH	73000	37	750000	2055	551

Table [4]

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.

Using the spreadsheet containing all domestic flights obtained in the data collection section of the report, we can find the number of domestic flights leaving a given airport each day. We also know how many times a certain flight leg is repeated per day. Using these values we 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.

An example of this is given below:

From	To	After Load Factor	Number of flights from origin	Number of Flights	Ratio	Origin Estimated Total Demand	Estimated Demand
ABS	ASW	539	5	5	1	42	42
ASW	ABS	660	41	6	0.14634	1127	165
CAI	ASW	3689	80	35	0.4375	12462	5452
CAI	ATZ	113	80	1	0.0125	12462	156
CAI	LXR	1917	80	22	0.275	12462	3427
HMB	CAI	234	3	2	0.6666	2532	1688
HRG	CAI	800	10	7	0.7	2285	1600
LXR	SSH	53	26	1	0.0384	1007	39
RMF	CAI	122	1	1	1	12	12

Table [5]

Method 2: 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%.

An example of this is given below:

From	To	Number of fights	Seats Available	After Load Factor
ABS	ASW	5	718	539
ASW	ABS	6	880	660
ASW	CAI	34	4964	3723
ATZ	HMB	1	162	122
CAI	ASW	35	4918	3689
CAI	ATZ	1	150	113
CAI	HMB	2	312	234
LXR	CAI	24	2880	2160
LXR	HRG	1	150	113
LXR	SSH	1	70	53
RMF	CAI	1	162	122

Table [6]

Finally, the final total estimated demand for each flight leg to be used in analysis is the average of the two methods.

Data-driven insights

Using flask library in Python a web app is designed, that can have city pair as user input ,for example:

Select City Pair

From: LXR **To:** SSH

Get Probabilities

Outputs:

Market HHI Score: **0.962**

Market Airlines' QSIs

Airline	Probability	QSI
MS	**1.94%**	**0.019359471339865198**
SM	**98.06%**	**0.9806405286601348**

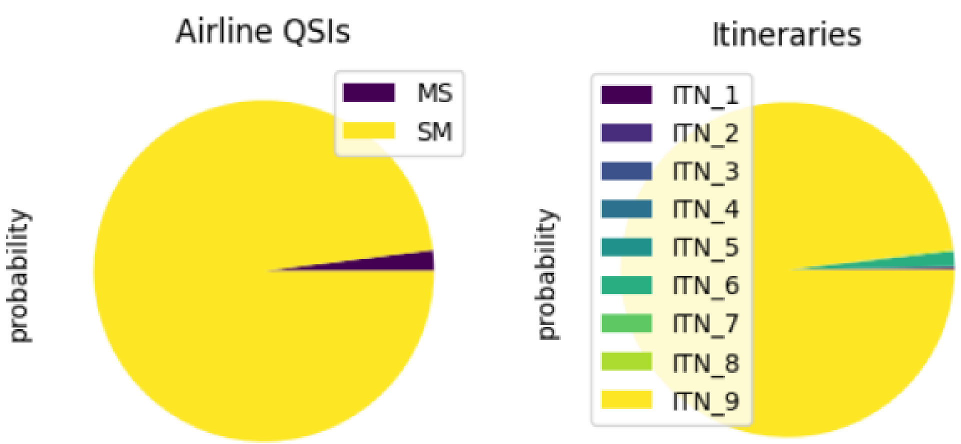

Fig. [7]

Itinerary Scoring

deprt_time	prop	capacity	distance	regional	flight_path	arrive_time	service_lvl	first_layover	second_layover	
2023-01-01 06:50:00	0	162	883	0	LXR -> CAI -> SSH	2023-01-01 10:30:00	1	0 days 01:30:00	0 days 00:00:00	-
2023-01-01 08:40:00	0	162	883	0	LXR -> CAI -> SSH	2023-01-01 14:40:00	1	0 days 03:50:00	0 days 00:00:00	-1
2023-01-01 09:10:00	0	162	883	0	LXR -> CAI -> SSH	2023-01-01 14:40:00	1	0 days 03:20:00	0 days 00:00:00	-[illegible]
2023-01-01 10:00:00	0	162	883	0	LXR -> CAI -> SSH	2023-01-01 14:40:00	1	0 days 02:30:00	0 days 00:00:00	-
2023-01-01 15:25:00	0	162	883	0	LXR -> CAI -> SSH	2023-01-01 20:50:00	1	0 days 03:15:00	0 days 00:00:00	-[illegible]

Table [7]

Regression Coefficients

Found some US data of passengers including their flight origin and destination airport besides departure time, example of the data:

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

Table [8]

The following betas at different hours in the day were obtained.

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

Table [9]

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.

Conclusion

In conclusion, this project aimed to forecast demand for domestic flights in Egypt and generate all possible itineraries using an itinerary builder. Python was used to develop a code that estimates the demand for each flight itinerary, and Flask was used to create a small web app that allows users to input their desired city pair and obtain the corresponding itineraries. Our findings suggest that the demand for domestic flights in Egypt is highest between certain city pairs. The itinerary builder proved to be a useful tool for generating all possible itineraries while accounting for user preferences such as maximum layover time and number of stops. However, there is potential for further improvements such as expanding the tool to include international routes and integrating it with airline and travel agency booking systems.

References

[1] *Egypt statistical yearbook 2020* (2020) *Arab Development Portal*. Available at: https://arabdevelopmentportal.com/news/egypt-statistical-yearbook-2020-0 (Accessed: March 20, 2023).

[2] Flightradar24 (no date) *Live flight tracker - real-time flight tracker map*, *Flightradar24*. Available at: http://www.flightradar24.com/about. (Accessed: March 20, 2023).

[3] Qubein, R. (2015) *The difference between a stopover and layover, and why you should care*, *USA Today*. Gannett Satellite Information Network. Available at: https://www.usatoday.com/story/travel/roadwarriorvoices/2015/08/18/the-difference-between-a-stopover-and-layover-and-why-you-should-care/83274870/ (Accessed: March 20, 2023).

[4] Hayward, J. and Garbuno, D.M. (2023) *The battle of the regional aircraft - what aircraft is best?*, *Simple Flying*. Available at: https://simpleflying.com/regional-aircraft-battle/ (Accessed: March 20, 2023).

[5] Abdelghany, A. and Abdelghany, K. (2016) *Modeling applications in the airline industry*. Florence: Taylor and Francis.