OCTOBER 28, 2021

PROJECT 2

DASC 5300-001

TEAM 09
ALISHBAH FAHAD (1001924185)
KSHITIZ KHADKA (1001990191)

Table of Contents

1. Objective

The objective of this project was to analyzing the airlines data, identifying the airline assigned to our team, determining the top hubs for our airline and conduct a deep analysis to predict what can be the next hub for our airline.

2. Data Set

We were provided a large dataset containing the airlines indicating the flights operated by that airline.

3. Process

- Clear understanding of the goal
- Obtaining the data
- Cleaning and exploring the data
- Generating the network graph
- Identifying the airline
- Predictive analysis to find next hub

Clear understanding of the goal: The first step was to understand the project requirements and the goals we needed to meet.

Obtaining the data: Since the data was provided to us, we were able to download the data and use it instead of collecting it on our own. But we have collected data for our predictive analysis as it was not provided to us.

Cleaning the data:

- Since the data that was provided to us was in a .txt format, we changed it to a csv format for our convenience.
- Then we have dropped the rows with newline character "\N".
- Furthermore, we extracted the data with the airline assigned to us for our analysis. According to the given parameters our airline data should comprise of 289 rows but we get 287 rows because we have dropped the rows with newline character.

Generating the network graph: For visualization of connection between source and destination airports we have plotted network graph by using networkx library.

Identifying the airline: By determining the top hubs we find out our airline. From the network graph and top hubs, we saw that ID 2188 was the center for the network and its occurrence was

119 which was highest among all other airlines. Then we saw code for the corresponding ID which was **“DXB”** giving us our airline **“United Arab Emirates”**.

4. Network Graph

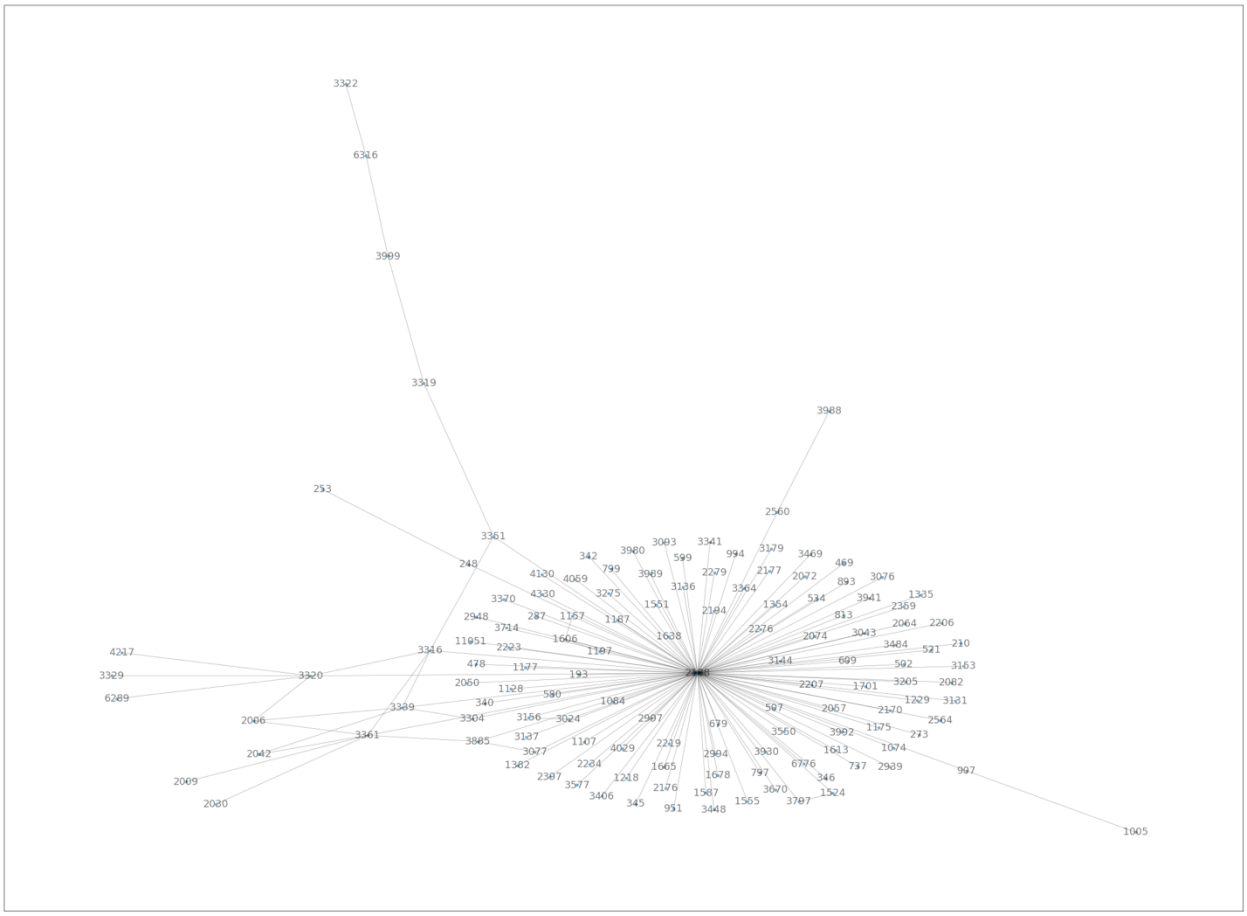

Figure 1 *Network Graph for our Airline*

Figure 1 shows the network graph having nodes and edges. Nodes represents the IDs for destination airports and the edges represents connection between the source and the destination airports.

5. Graph Characteristics

Number of Nodes: 135

Number of Edges: 148

Density: 0.01636263128800442

Number of connected components: 1

Connected components: [135]

Diameter: 7

Minimum degree: 1

Maximum degree: 120

Average degree: 2.1925925925925926

Standard Deviation degree:10.255319912703115

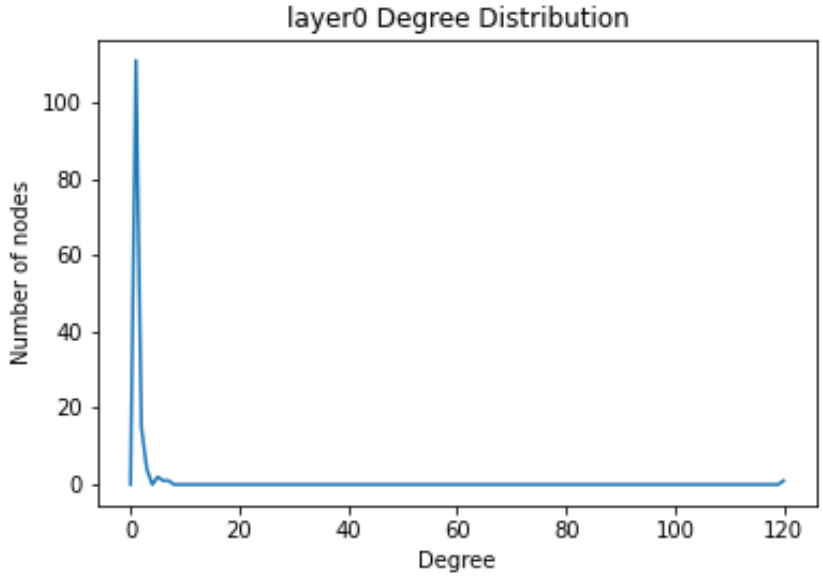

Figure 2 *Degree Distribution*

6. Identifying top hubs

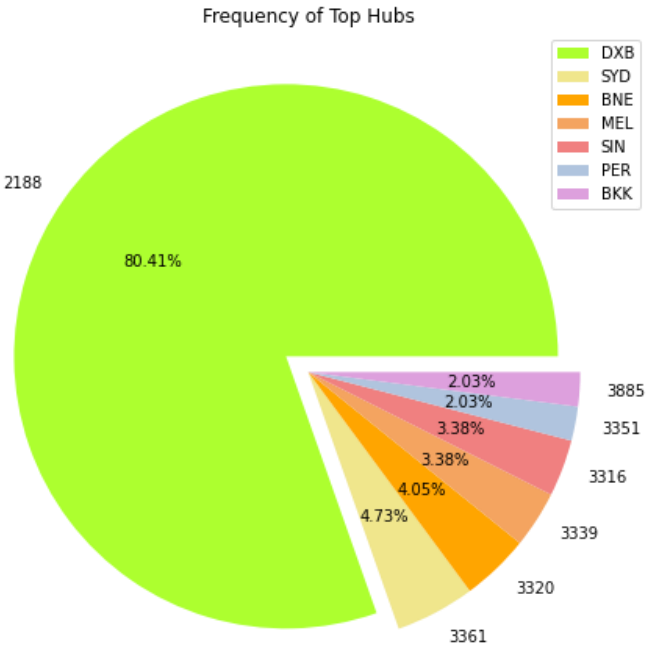

Figure 3 *Top Hubs*

The figure 3 represents the top 7 hubs with the highest frequency of the top 7 hubs.

7. Predictive Analysis

Based on the study of both the data given to us and collected by us, we can see that the predicted airport hubs don't match. From the data given to us, Bangkok is likely to be the next hub but from the data collected by us, London – Heathrow has the highest chance to be picked as the next hub. That brings us to a conclusion that London – Heathrow airport will be the next hub for Emirates. The below chart and clarifications explain why London-Heathrow will be the next hub.

We collected the data from the official website of Emirates of all 138 destinations that Emirates flies to and from. The dataset contains all the flights until October, 2021. The records were collected manually by ourselves consisting of weekly flights that flies to each destination.

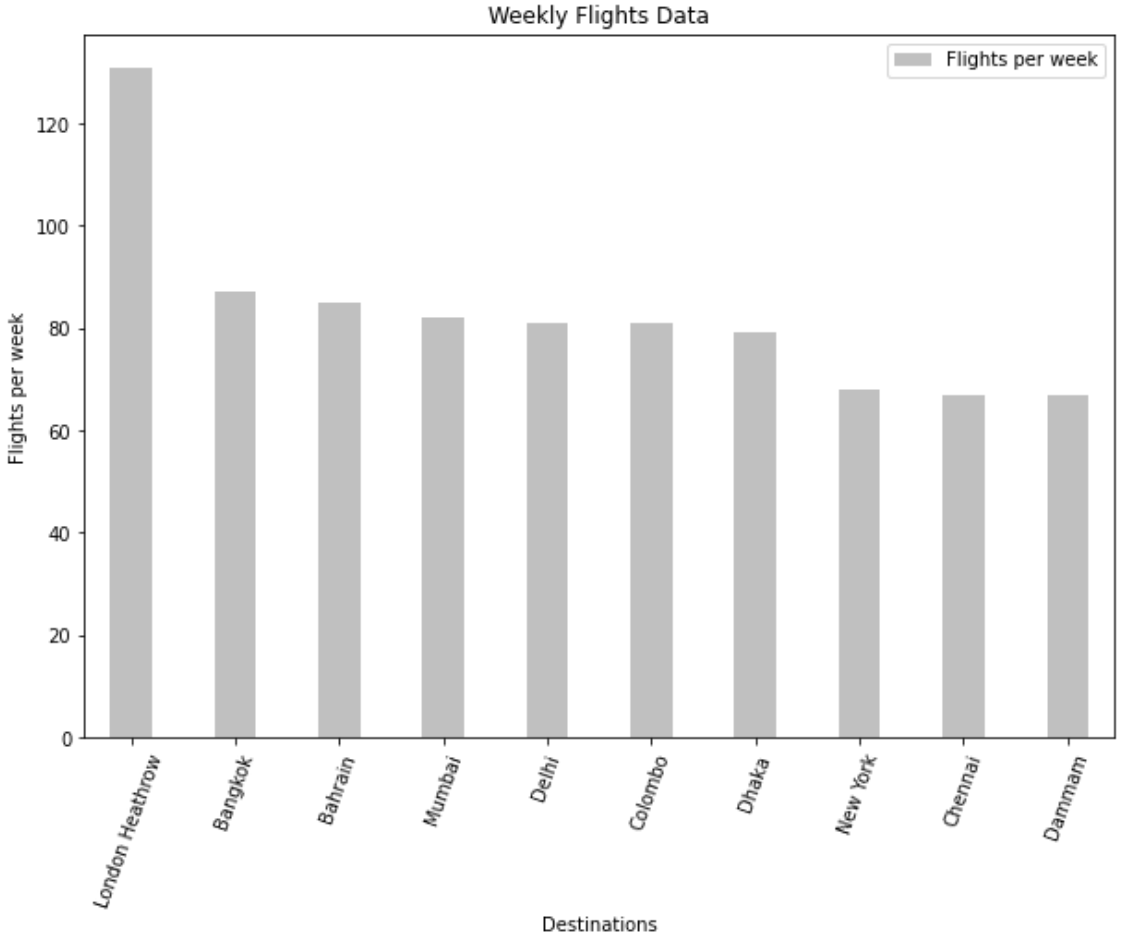

Figure 4 *Weekly Flights for Emirates*

From the above graph we can see that the highest number of Emirates flights flown to and from is London – Heathrow airport.

Evaluating the information collected from the above graph, we conducted a deeper analysis on the top most airport based on the city it is located, the demographic, the traffic of those airport routes and airport with most transfers to make a prediction.

Top 10 Busiest International Airports by Seats

Calculated using international frequency only

2021 Ranking	Airport Code	Airport Name	Seats	2019 Ranking
1	DXB	Dubai	2,728,723	1
2	AMS	Amsterdam	2,479,152	3
3	FRA	Frankfurt	2,187,008	7
4	LHR	London Heathrow	2,182,740	2
5	IST	Istanbul	2,117,541	9
6	CDG	Paris Charles de Gaulle	1,927,621	5
7	DOH	Doha	1,612,187	13
8	MAD	Madrid Adolfo Suarez-Barajas	1,303,785	14
9	AYT	Antalya	1,202,375	32
10	VIE	Vienna	1,128,228	22

Figure 5 *Top 10 Busiest International Airports by Seats*

From the table posted in OAG website, we can see that Dubai is the busiest international airport by seats. Since Dubai is already a hub for Emirates, we analyzed the second busiest airport and whether Emirates flies there or not.

The article posted by “EAS BARCELONA” states that London-Heathrow airport is the busiest in Europe.

Top 10 Busiest International Routes - Global

Route	Route Name	Seats
AYT-SVO	Antalya-Moscow Sheremetyevo	270,402
DXB-RUH	Dubai -Riyadh	240,938
DXB-LHR	Dubai -London Heathrow	217,668
DME-SIP	Moscow Domodedovo-Simferopol	206,853
CAI-JED	Cairo-Jeddah	204,536
JFK-LHR	New York JFK-London Heathrow	165,223
MCO-SJU	Orlando-San Juan	163,296
AYT-VKO	Antalya-Moscow Vnukovo	152,393
BAH-DXB	Bahrain-Dubai	145,542
DXB-JED	Dubai -Jeddah	138,898

Figure 6 *Top 10 Busiest International Routes*

From the data posted by OAG in their website, we can see the top 10 busiest international routes for October 2021. From the table it can be seen that the route from Dubai – London Heathrow is the third busiest route. An article posted in "The Telegraph" states that London Heathrow Airport is the world's second busiest airport for international transfers after Dubai International Airport. Since the current hub for Emirates which is located in Dubai falls in Asia, Emirates might seek to expand their reach to Europe more by making the **London-Heathrow** airport as its next hub.

8. File Description

The list of files we have created are:

1. DASC5300_Proj2_Fall21_team_09_.html: PYDOC file showing the documentation of our codes.
2. DASC5300_Proj2_Fall21_team_09_.py: This documented file consists of all the codes used to create the graphs and the csv files.
3. cleaned data.csv: CSV file which was cleaned by dropping the rows with newline character.
4. Airlines for Analysis.csv: CSV file for our airline analysis.
5. input.txt: Input file comprising of IDs for source airport and destination airport.

6. output.txt: Output file comprising the characteristics of graph.
7. Flights.csv: Weekly flights of 138 destination of Emirates airline flying to and from.

9. Division of Labor

We worked on the project individually in the beginning. Since the requirement of the project was different than the last time, we worked on its side by side collecting the data from the web and had countless constructive discussions for the analysis part. We spent at least 2 hours a day since we were assigned the project.

10. Problems encountered and how we handled them

The 3 problems that we encountered are:

1. Confusion regarding Part e: For ‘e’ present in the list of things we need to do, it’s not clear about the goal. We were confused if we needed to predict the next hub based on the dataset given to us or can predict the next hub using data that is in the internet to predict.
2. Vague requirements for the top hubs: The project assignment don’t state if we have to only state the top hubs or create a graph. We as a team decided to create a bar chart representing the top hubs and a pie chart representing the top hubs.
3. New concept: Since this assignment was for mapping graph which was a new concept to us, it took us time to figure out the main objective of the project and what was required for us to do.

References

1. *Fly better across the world*. Emirates. (n.d.). Retrieved October 28, 2021, from https://www.emirates.com/us/english/.
2. Worldwide, O. A. G. A. (n.d.). *Busiest airports in the world*. OAG. Retrieved October 29, 2021, from https://www.oag.com/busiest-airports-world.
3. Boreadas, J. (2019, October 2). *The world's most transited airports*. EAS Barcelona. Retrieved October 29, 2021, from https://easbcn.com/en/the-worlds-most-transited-airports/.
4. Worldwide, O. A. G. A. (n.d.). *Busiest routes right now*. Busiest Routes Right Now. Retrieved October 29, 2021, from https://www.oag.com/busiest-routes-right-now.