19ZO02-SOCIAL AND ECONOMIC NETWORK ANALYSIS

PROJECT REPORT

Team Members
19z227 Monirhithikka S P
19z240 Samyuktha Sreekanth
19z243 Sarayu Miththira V C

BACHELOR OF ENGINEERING

Branch: COMPUTER SCIENCE AND ENGINEERING

Of Anna University

PSG College of TechnologyCoimbatore –

641004

PROBLEM STATEMENT:

Visualization of Network graph of Airline system to illustrate varying viewpoints in the mapping of air travel.

A lot of factors affect the overall price of airline tickets, including the airline, the date of travel, source, destination, route, duration, and so on. Each provider seems to have its own unique set of regulations and methods for determining pricing. Recent breakthroughs in Artificial Intelligence (AI) and Machine Learning (ML) allow for the inference of such principles as well as the modelling of price volatility. While we might hear about those aspects the most, the customer experience is not about just the flight itself. It's everything from purchasing reasonable tickets to quality airline availability and thoughtful bookings. It's important that customers have an excellent experience every time they travel.

DATASET DESCRIPTION:

- This datasets are extracted from the Kaggle database.
- Dataset Attributes:
    1. Airline
    2. Airline ID
    3. Source airport ID
    4. Source airport
    5. Destination airport ID
    6. Destination airport
    7. Date_of_Journey
    8. Source
    9. Destination
    10. Price

TOOLS USED:

- Gephi:
  Gephi is a tool for data analysts and scientists keen to explore and understand graphs.
- Python:
  We have used the Python Language for the coding part because of its User-friendly Data Structures.
- Google collab:
  Google Collab is particularly well suited to machine learning, data analysis, and education since it enables anyone to develop and run arbitrary Python code through the internet. The cloud-based, open-source Jupyter Notebook, which provides free computing resources, is essentially another name for Google Collab. Python code may be written and run through a browser using Google Collab.
- Networkx:
  The Python programming language's Network X package is used to create, modify, and research the structure, dynamics, and functionalities of complex graph networks.

CHALLENGES FACED:

- Combining datasets in order to make all the necessary analysis available together.
- Understanding of Gephi and networkx.
- Understanding the necessity of various libraries required for making the analysis.

CONTRIBUTION OF TEAM MEMBERS:

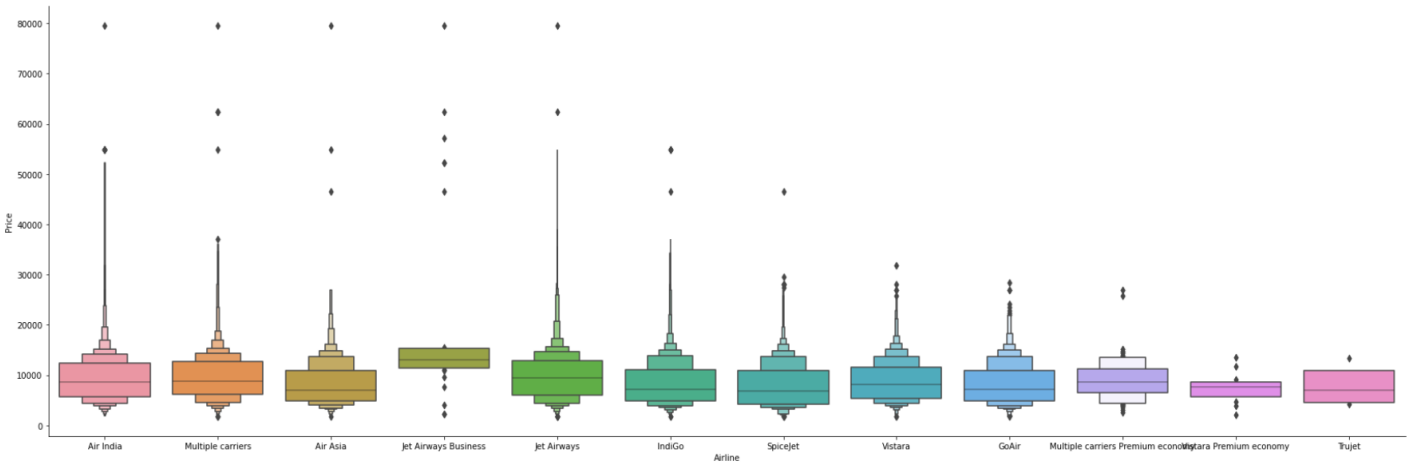
- Exploratory Data Analysis
- Fair price prediction
- Clustering Classification of air traffic based on K- Means Clustering
- Project Report Formulation

All four categories were equally worked on by the 3 team members .

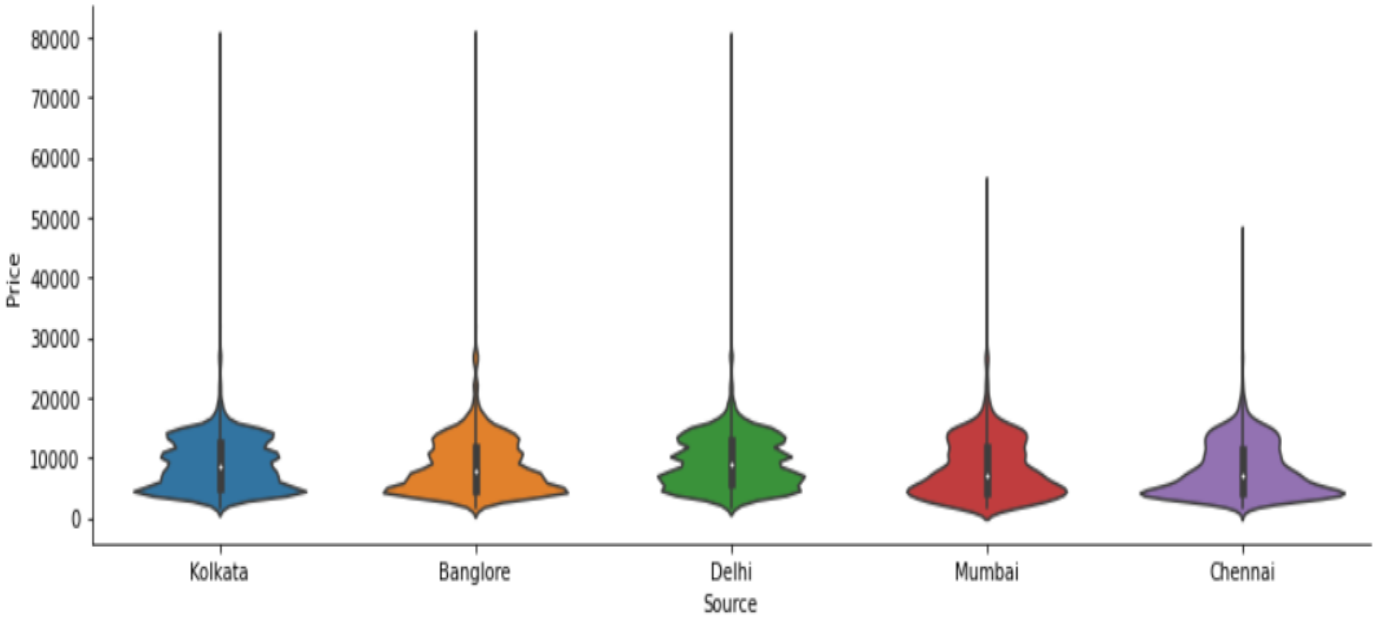
ANNEXURE 1:

Link to google collab -
https://colab.research.google.com/drive/14ESplM1WbcKGWscMiePjNiSbCPT3dEN8#scrollTo=r5w-jqO9oXh4

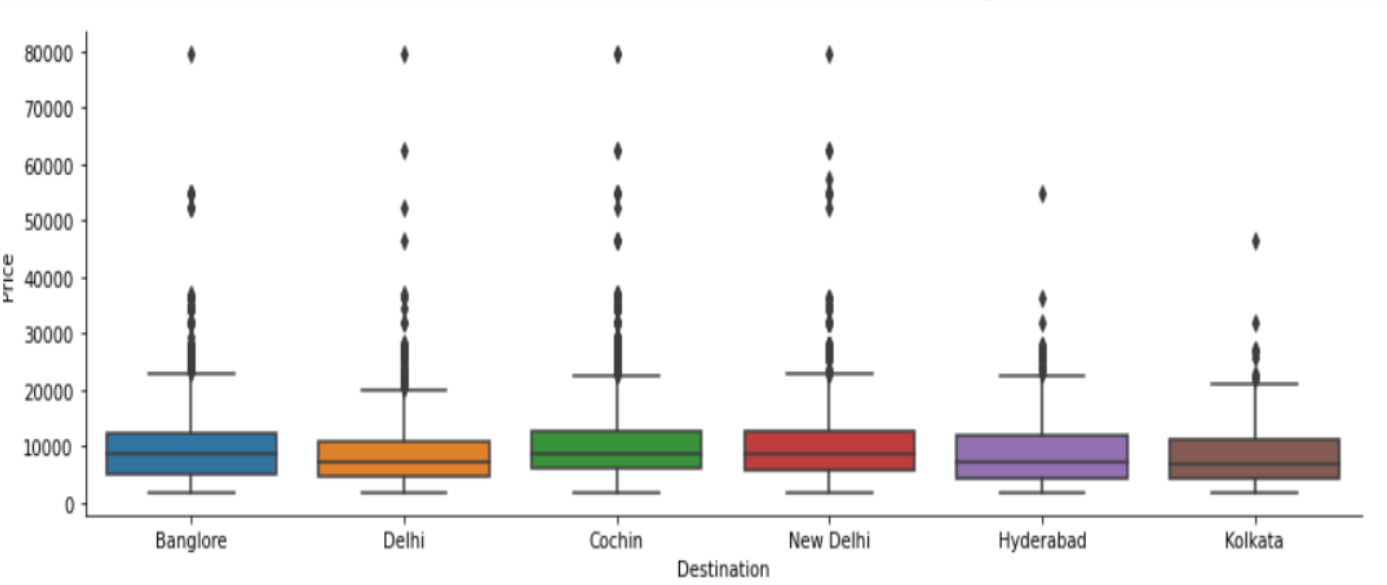
ANNEXURE 2:



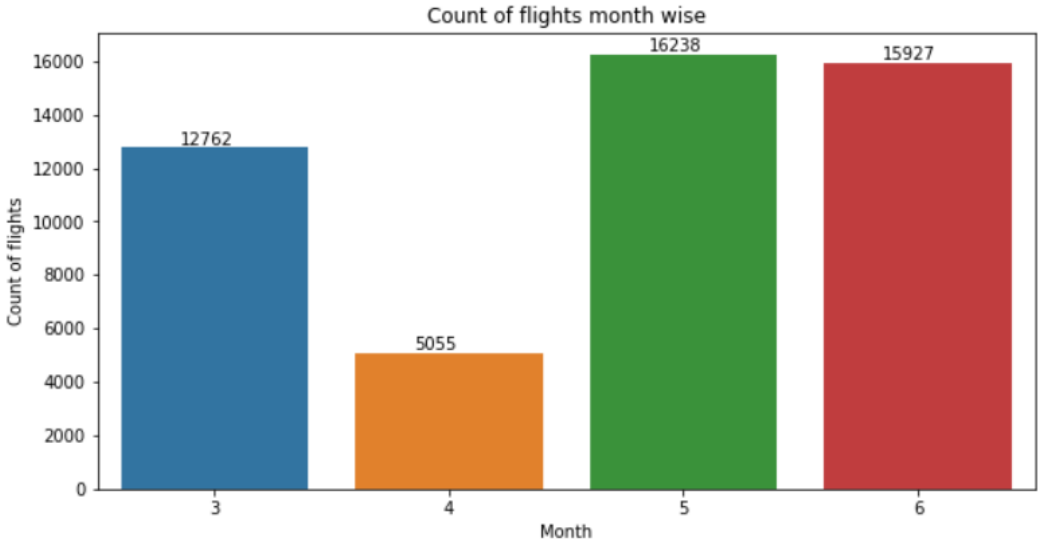
- The price of jet airways is high(many outlier) while the rest of the airways there's not much changes that can be observed.



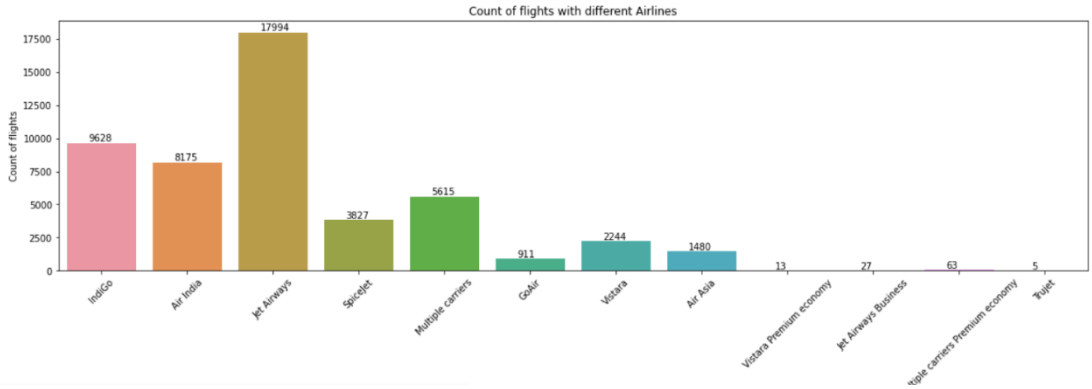
- Kolkata, Bangalore and Delhi as source has more outliers compared to Chennai and Mumbai. A person boarding from these three locations will have a significant rise in ticket price compared to the other two.



- Cochin, Bangalore and New Delhi as destination has more outliers compared to Hyderabad and Kolkata. The price to travel to and for from Bangalore is pretty high compared to every other place. Using these inference the customers can choose which path can be chosen especially for the ones who have been doing regular travel in such a way that their expenses can be mediated.

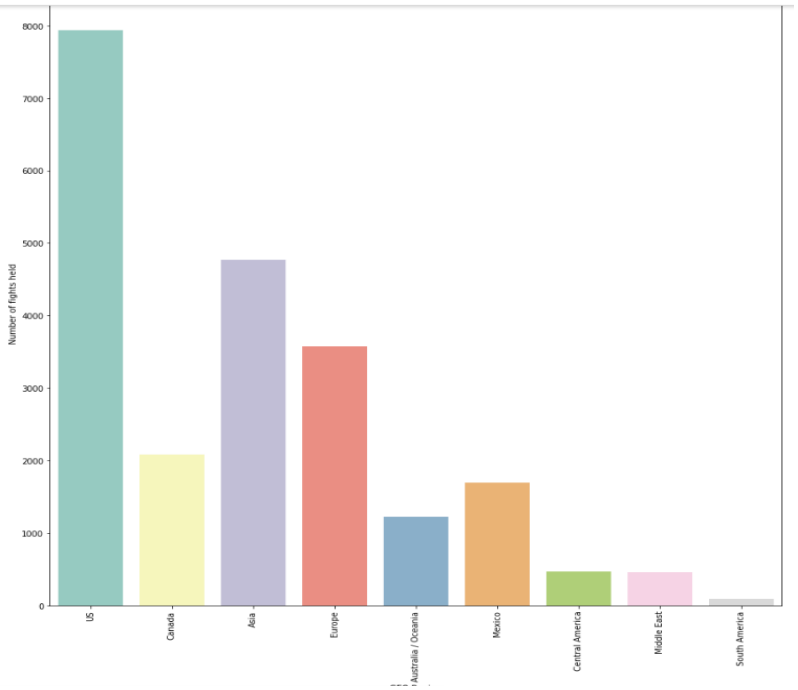


- During the month of may there is more demand and in April the demand is less. Based on this the customers can book their tickets easily during April.

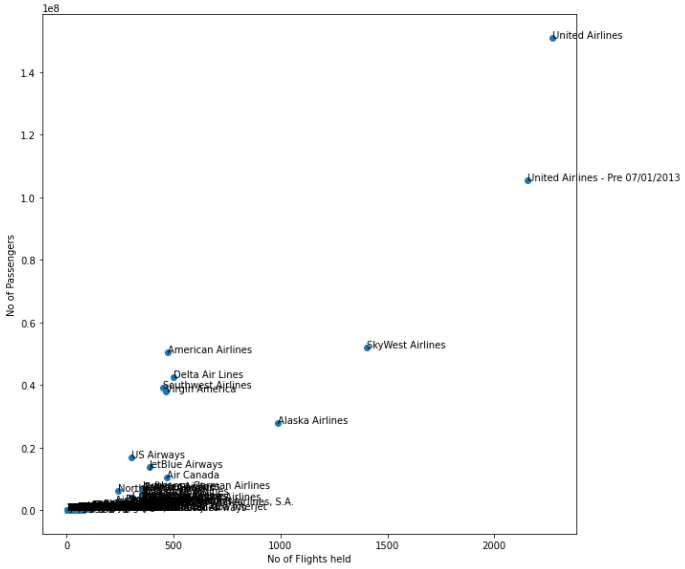


- Jet-airways(economy) is the most boarded flight from which we can infer that it is the most favorable airlines and provides the best customer services.

- Using these plots we can infer the range of ticket pricing , Customers can easily navigate between expensive and relatively cheap flights saving them considerable amount of money and time .

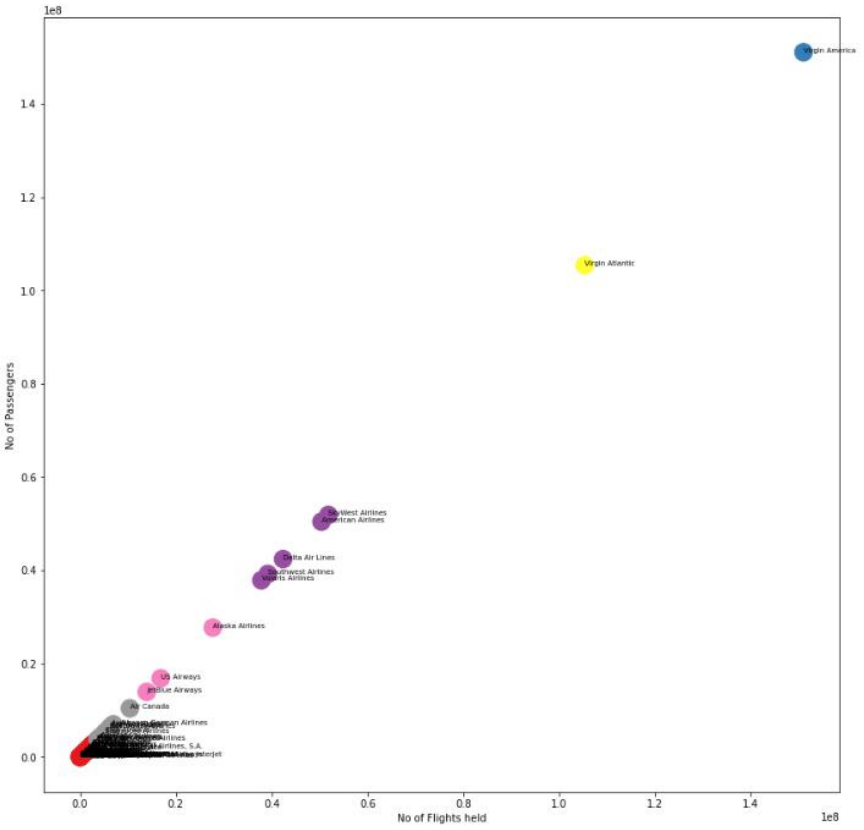


- United Airlines -Pre can be considered as a potential outlier. From this plot we can see that the number of flight held by US is considerably high compared to the others . The aviation industry for this particular geographical region is blooming and considerable amount of travel persists .



- From this plot we can infer that more airlines are clustered at the bottom right which w.r.t number of flights held and passengers boarding shows that these flights are among the low tiered airlines , and has a medium range of favor from citizens considering the quality . US has more number of flights compared

to the other regions which appear at top right of the plot consisting of bigger flights with better service and landing more crowd than the others .



- We can see that American Airlines and SkyWest airlines hold many flights and transport many people, so we can infer that maybe this airlines have planes with bigger capacity or people use these airlines more than other .We can thus make an assumption on the quality of travel and customer satisfaction of these airlines compared to others .

REFERENCES:

[1] https://www.researchgate.net/publication/335936877_A_Framework_for_Airfare_Price_Prediction_A_Machine_Learning_Approach

[2] https://www.ijraset.com/research-paper/flight-fare-prediction-system-using-ml

[3] https://arxiv.org/pdf/2102.03711.pdf

[4] https://www.sciencedirect.com/science/article/pii/S131915781830884X

[5] https://medium.com/mlearning-ai/flight-fare-predictions-b1150fdc45a3

[6] https://www.ijraset.com/research-paper/flight-price-prediction

[7] https://www.academia.edu/46805163/Flight_Ticket_Price_Predicting_With_the_Use_of_Machine_Learning

[8] https://www.ijstr.org/final-print/dec2019/Predicting-The-Price-Of-A-Flight-Ticket-With-The-Use-Of-Machine-Learning-Algorithms.pdf