Data Visualization and Delay Prediction of the Copa Airlines Flight Network

Sarthak Sharma
Master of Science (Computational Science) student,
Department of Scientific Computing,
Florida State University.

Contents

4. Flight Delay Prediction Using Machine Learning

5. Conclusions and Future Work

Appendix

1. Introduction

This research project carried out jointly between Copa Airlines and Florida State University. Its primary aim is to improve (if possible, by automating) some or all aspects of flight planning / scheduling of Copa Airlines, so that delays in departure and arrival of Copa Airlines flights can be reduced, and using artificial intelligence / machine learning to predict future delays, given the data of past delays. The ultimate goals are to improve passenger experience and increase profits of Copa Airlines.

1.1. Flight's time-related considerations

On-time performance is important for any for-profit organization selling services to customers. For airlines, this service is the flight, which is supposed to transport its passengers from the origin airport to the departure airport at expected times, without any delays or unexpected events. Avoiding delays becomes even more critical when many passengers are transferring from one flight to another, as delays can easily cascade downstream in a flight network. Events causing flight delays may broadly be classified in two categories: those that are external events and hence outside the control of the airlines and airports (for example, natural disasters like earthquakes, storms / bad weather, government restrictions, terrorist attacks, etc.), and those that are internal events and hence may be under the control of airlines and airports (for example, late arrival of crew members, aircraft and airport maintenance, waiting on arriving flights that supply ("feed") passengers to departing flights, etc.

In the following two sections, we consider a few important time metrics used to measure delays, followed by a few definitions of delay-related variables.

1.1.1. Important times / events in a flight:

The course of a flight has four important events, whose times are described by special terminology. First, there is OUT, when the boarding of passengers and luggage is completed, and the aircraft doors are closed, at the flight's origin airport (that is, the aircraft "disconnects" from the origin airport's terminal). Taxiing takes place, followed by the second event: OFF, when the aircraft flies up from the runway at the flight's origin airport. After the flight is almost completed, the third event is ON, when the aircraft lands on the runway at the flight's destination airport. Finally, the fourth event is IN, when the aircraft connects to a terminal at the flight's destination airport, and doors open, allowing passengers to deboard and luggage to be taken out by staff.

OFF and ON times are used mostly for analysis of taxi time, runway congestion, etc. Flight delay is calculated using the other two times: OUT and IN. All the four quantities are usually measured in minutes (maybe also seconds in some cases). The time can be either the local time at the airports' locations, or a universal time like UTC. When comparing times across different time zones, it is important to use the same time standard. UTC is Coordinated Universal Time, the primary time standard by which the world regulates clocks and time. It is within about 1 second of mean solar time at 0° longitude (at the IERS Reference Meridian as the currently used prime meridian) and is not adjusted for daylight saving time. [1]

1.1.2. Delays of a flight:

Delay of an event is the difference between actual and scheduled times of that event. Delays can be positive, negative, or zero. For flights, delays are usually measured in minutes. There are two types of delays for a flight: departure and arrival delays.

Departure delays are measured as the difference between the actual and scheduled times of departure.

$$D_{dep} = O_{act} - O_{sch} \tag{1.1}$$

where, $D_{departure}$ is the flight's departure delay, O_{actual} is the flight's actual OUT time, $O_{scheduled}$ is the flight's scheduled OUT time. Departure delay may be caused by bad weather at origin airport, arrival delay of feeder flight (from which passengers are obtained), certain circumstances at origin airport, technical glitches in aircraft, etc.

Arrival delay: This is defined as

$$D_{arr} = I_{act} - I_{sch} \tag{1.2}$$

where, $D_{arrival}$ is the flight's arrival delay, I_{actual} is the flight's actual IN time, $I_{scheduled}$ is the flight's scheduled IN time. Arrival delay may be caused by bad weather on flight route (causing alternate, longer route) or at destination airport (causing wait before landing), significant departure delay of the same flight, certain circumstances during flight (e.g. medical or other emergency of passenger or crew, necessitating temporary halt of aircraft at intermediate airport), busy or temporarily unavailable runway at destination airport, causing wait before landing, technical glitches in aircraft, etc.

1.2. Copa Airlines

Copa Airlines (code CM) is the national airline of the Republic of Panama, located between the North and South American continents. Panama City is the capital of the Republic of Panama, and its airport: Tocumen International Airport (code PTY) is the hub of Copa Airlines.

1.2.1. Advantage: strategic location of hub airport

The strategic location of PTY airport makes it ideal as a connecting hub between many destinations, especially inter-continental flights between North America and South America, as shown in figure 1.1. PTY airport is also origin or destination for many flights to and from other parts of the world.

Figure 1.1.: Map showing part of flight network of Copa Airlines in 2016. Flights with PTY airport as either origin or destination. [2]

1.2.2. Types of flights in Copa Airlines

As noted earlier, PTY airport is the hub of Copa Airlines. It is the airport where CM aircraft may be ordinarily kept for servicing and repair, and when they are not in use for long duration. Due to the geographically advantageous location of PTY airport, almost all flights of Copa Airlines involve PTY airport as either origin or destination. Thus, there are two main types of flights in Copa Airlines: **inbound flights** (those that have PTY airport as *destination*) and **outbound flights** (those that have PTY airport as *origin*).

Most passengers of Copa Airlines are “transit” passengers, that is, they transfer from inbound flights to outbound flights at PTY airport. Only a small fraction of Copa Airlines passengers may have PTY airport as their origin or destination. Thus, most inbound flights are **feeders** to outbound flights, that is, they “feed” passengers to outbound flights.

1.2.3. Challenge: preventing delay propagation in flight network

Along with the advantage of location, there is a challenge for Copa Airlines: possibility of flight delays propagating in the flight network. Large arrival delay of an inbound flight may cause departure delays in the outbound flights that are fed by it. Thus, if a feeder (inbound) flight has a significant arrival delay, and there is insufficient available connection time (e.g. less than 30 minutes) between flights, then officials at Copa Airlines are faced with the challenge of choosing one of the following two options: either delaying departure of outbound flights fed by the delayed inbound flight, where available connection times are insufficient, to allow the delayed connecting passengers to board the outbound flights; or not delaying departure of outbound flights, and instead transferring the delayed connecting passengers to next available outbound flights if space is available on them. (If the wait time is large, then this may involve Copa Airlines having to pay for stay of those passengers in hotels near PTY airport.)

Clearly, both are difficult choices that Copa Airlines would like to avoid. But some delays may be out of control of airline and airport officials, and hence inevitable. Thus, to improve passenger experience and increase profits, it may help to have an automated system that algorithmically decides which option to choose. There may even be a computational system (perhaps involving artificial intelligence / machine learning) that can predict (giving probability) of future delays, based on data of past delays.

1.2.4. Subsidiary of Copa Airlines

The parent company of Copa Airlines (Copa Holdings, S.A.) owns the subsidiary Copa Airlines Colombia (corporate name of Aero República S.A.). [3] [4] In 2016, it was newly known as "Wingo". Its code is *CM**. [5]

1.3. Resources required in the commercial aviation industry

Successful operations of commercial flights (of all airlines at all airports) requires a few basic resources that are mentioned next.

First, of course, we need the **aircraft**. These must be functioning properly. Most commercial aircraft are manufactured by Boeing, Airbus, and their subsidiaries, and sold to airlines. Sizes and capacities can vary. Most aircraft of Copa Airlines are of Boeing and Embraer. Then, we need **airports**, from where aircraft can take of and land. Time slots and physical space must be available for the aircraft's take-off at the origin airport, landing at the destination airport, and stationary stay (parking) when not in use. Copa Airlines has PTY airport at Panama City as its "hub".

The people in a flight consist of crew and passengers. The **flight crew** consists of pilots (for flying aircraft) and flight attendants (to ensure comfort, safety, and rule compliance of passengers). **Passengers** are the revenue sources for airlines and airports. In a flight, there can be one or more classes of passengers. Common types are: first class or business class passengers (denoted as "C" class in the data) who pay higher airfare and have more comfort / luxury in flight and at airport, and economy class passengers (denoted as "Y" class in the data) who pay lower airfare and get lesser comfort / luxury.

When the aircraft is not in use, the airline's **engineers / technical staff** perform routine checks of the aircraft. If required, repairs and replacements are performed. Flight take-offs and landings are monitored by airport's control tower. The aircraft's propulsion comes from its engines, which are powered by **aviation fuel**. If required, this fuel is filled in the aircraft at the origin airport, before the flight starts. Aviation fuel may be largest expenditure for airline operations. The airline's and airport's **ground staff** handle luggage transfer, passenger boarding and deboarding, etc. **Security** is maintained by security staff by thorough checking of all passengers, crew, luggage, etc. at all entry points. Security goals include preventing flight hijacking or bomb implanting by terrorists / militants and ensuring general safety and security of all. **Financial resources (capital)** are required for the airline to handle all operations, including possible compensation of passengers in unfortunate or unanticipated events like cancellations or large delays of flights, lost luggage, etc. Finally, there are **other resources** including cleaning and repair staff for maintenance of airport; uninterrupted electrical power supply at airport for 24x7 flight operations, etc.

1.4. Banks at PTY airport

There are six "banks" for aircraft at PTY airport, where CM aircraft may arrive or depart. These are: B1, B2, B3, B4, B5, B6. Some CM aircraft may be "off-bank" (that is, not being assigned to any of these 6 banks).

The CM flights are categorized in to banks at PTY airport based on their arrival or departure time at PTY airport, and their origin or destination. The reason for organizing flights in to banks is to minimize aircraft congestion at the airport. Flights typically arrive and depart in "waves" of time slots during the 24 hours of a day.

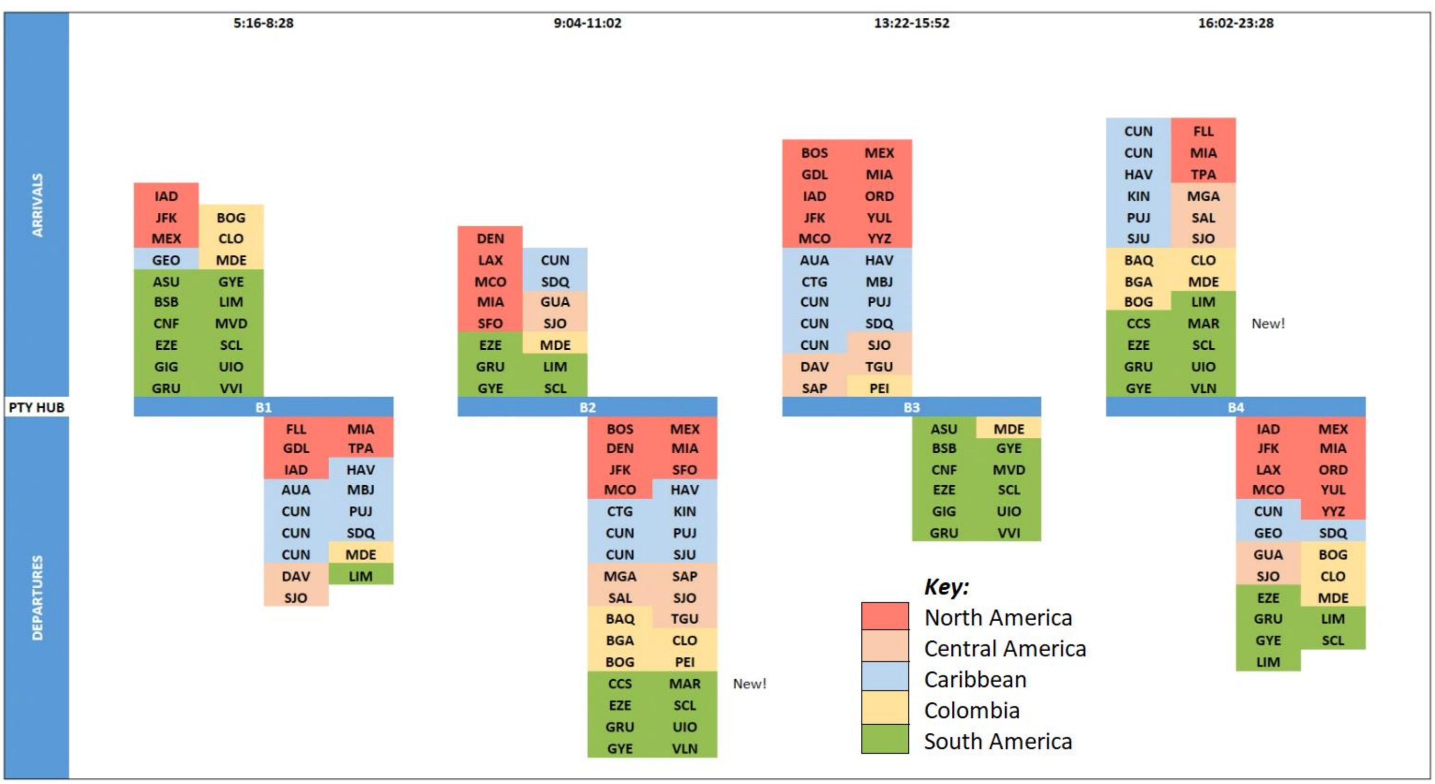


Figure 1.2: Example 4-bank hub structure of Copa Airlines at PTY airport in February 2021.

1.5. Flights considered for analysis

In this work, the flights considered for analysis are the ordinary Copa Airlines flights satisfying the following four conditions: airline code should be *CM* which denotes Copa Airlines (not *CM** or any other code); origin airport must be different from destination airport; PTY airport must be as either origin or destination; and there should be no missing or negative values in required fields, in the data provided by Copa Airlines.

1.6. Flight's unique ID

A flight's unique ID is a sequence of 24 characters that is the concatenation of the following five strings. First ten characters are of the scheduled departure date (Zulu) in yyyy-mm-dd format. Next three characters denote the origin code. Next three characters denote the destination code. Then, there are five characters denoting the flight's scheduled departure time (Zulu) in 24 hours format hh:mm. Finally, the last three characters are of the flight number. No two flights will have the same combination of these five quantities, even over a long period of several years. Hence, each flight will have a unique flight ID. A flight's unique ID is used for cleaning and filtering data from the input data files.

2. Delay Probability Analysis

As mentioned in the introduction, there are two types of flight delays: departure delay and arrival delay. These are defined in equations 1.1 and 1.2 of chapter 1. A flight delay can be positive (flight is *late* in that event), or zero (flight is *on time* in that event), or negative (flight is *early* in that event).

2.1. Preliminary analysis

In a data cleaning phase, we removed outliers, which we defined as positive or negative delays with magnitude greater than 60 minutes. When an aircraft lands at a destination city, a negative delay represents an early arrival. If a departure has negative delay, then it means that the aircraft leaves the gate earlier than scheduled. We start by presenting some preliminary analysis related to the average delay per origin-destination (OD) pair. [6]

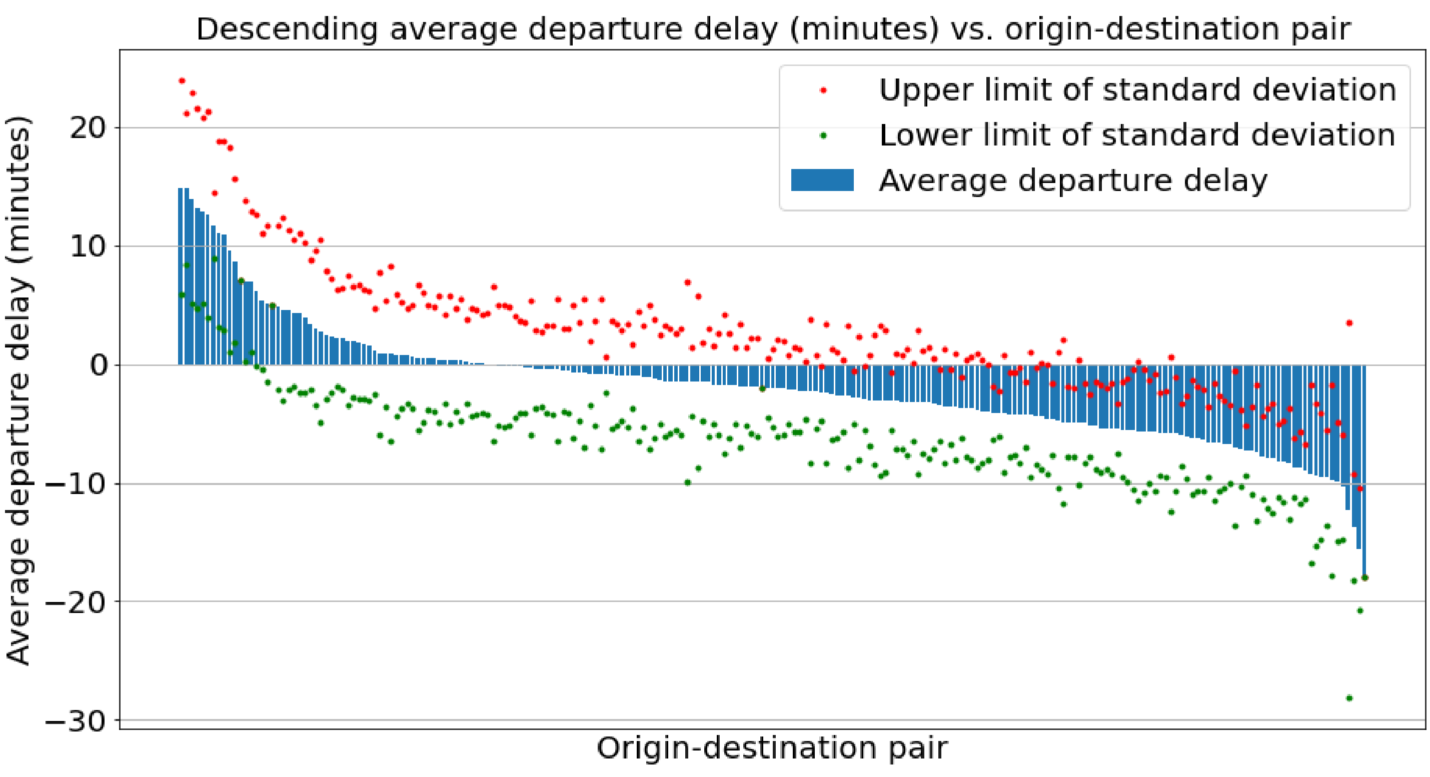


Figure 2.1.: Average departure delay for different origin-destination pairs.

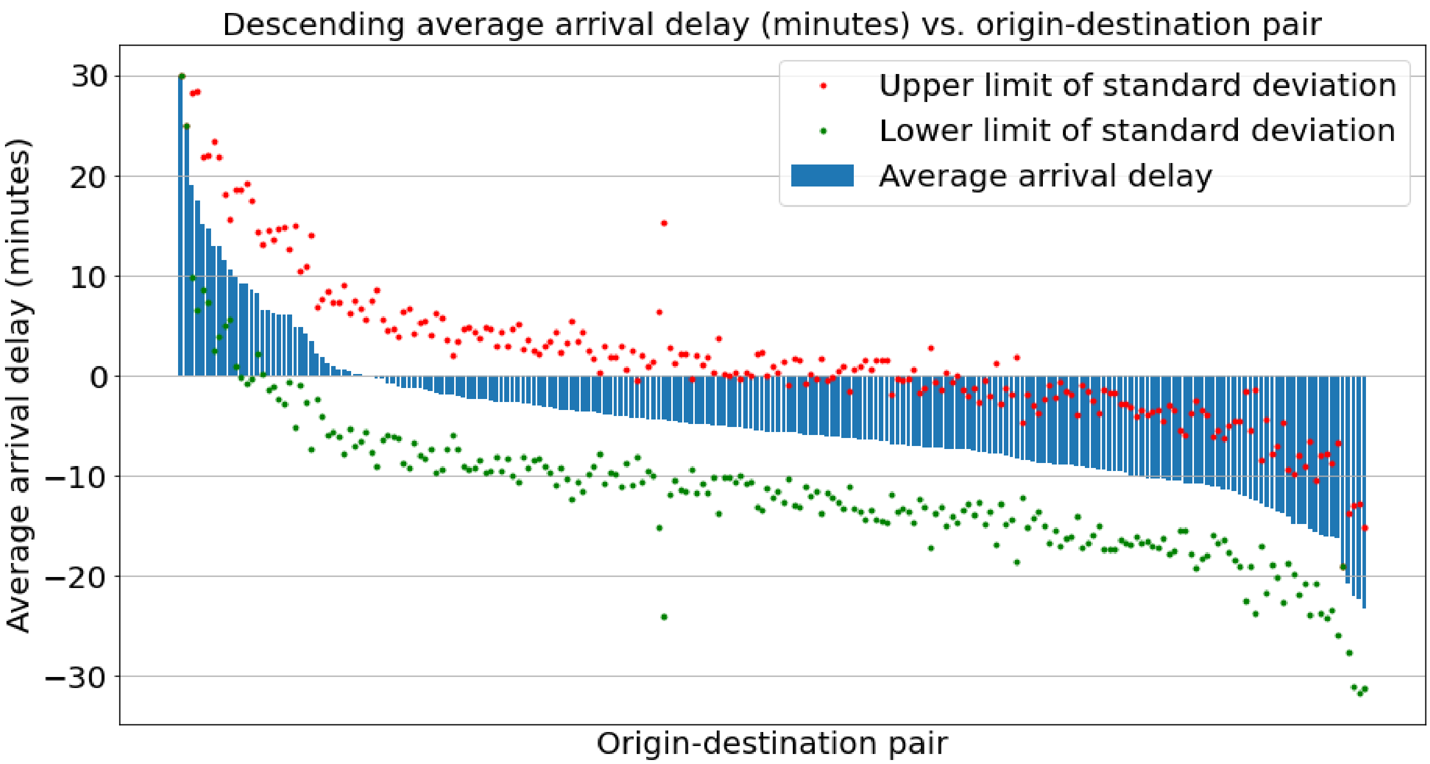


Figure 2.2.: Average arrival delay for different origin-destination pairs.

As can be seen in the plot displayed in figures 2.1 and 2.2, the average delays in flight departures and arrivals are quite variable. Extreme values are probably for OD pairs with very few flights. The standard deviation of departure and arrival delays is usually around 10 to 20 minutes. From these plots, we deduce that most flight routes of Copa Airlines have negative average delays (in both departure and arrival). Also, a larger fraction of flights arrive early, compared to flights departing early. This is understandable, as pilots may want to depart early from the origin so that they can minimize the probability of arriving late at the destination. However, they can not depart too early from the origin, as then there is risk of some passengers missing their flights. Also, if a flight departed a little late, then the lost time can be recovered by increasing the aircraft's average velocity during flight. Hence, a flight departing a few minutes late from the origin may still arrive on time at the destination. But if a flight departs from the origin very late, then it will probably arrive late at the destination.

Next, we plot the average delays of CM flights at different banks at PTY airports. There are six banks at PTY airport: B1, B2, B3, B4, B5, B6. Aircraft are assigned to banks depending on their origin airport (in case of flights arriving at PTY airport) or their destination airport (in case of flights departing from PTY airport), and the time of the day when they arrive at or depart from PTY airport. Some aircraft may not be assigned to any bank; these are known as "off-bank" flights.

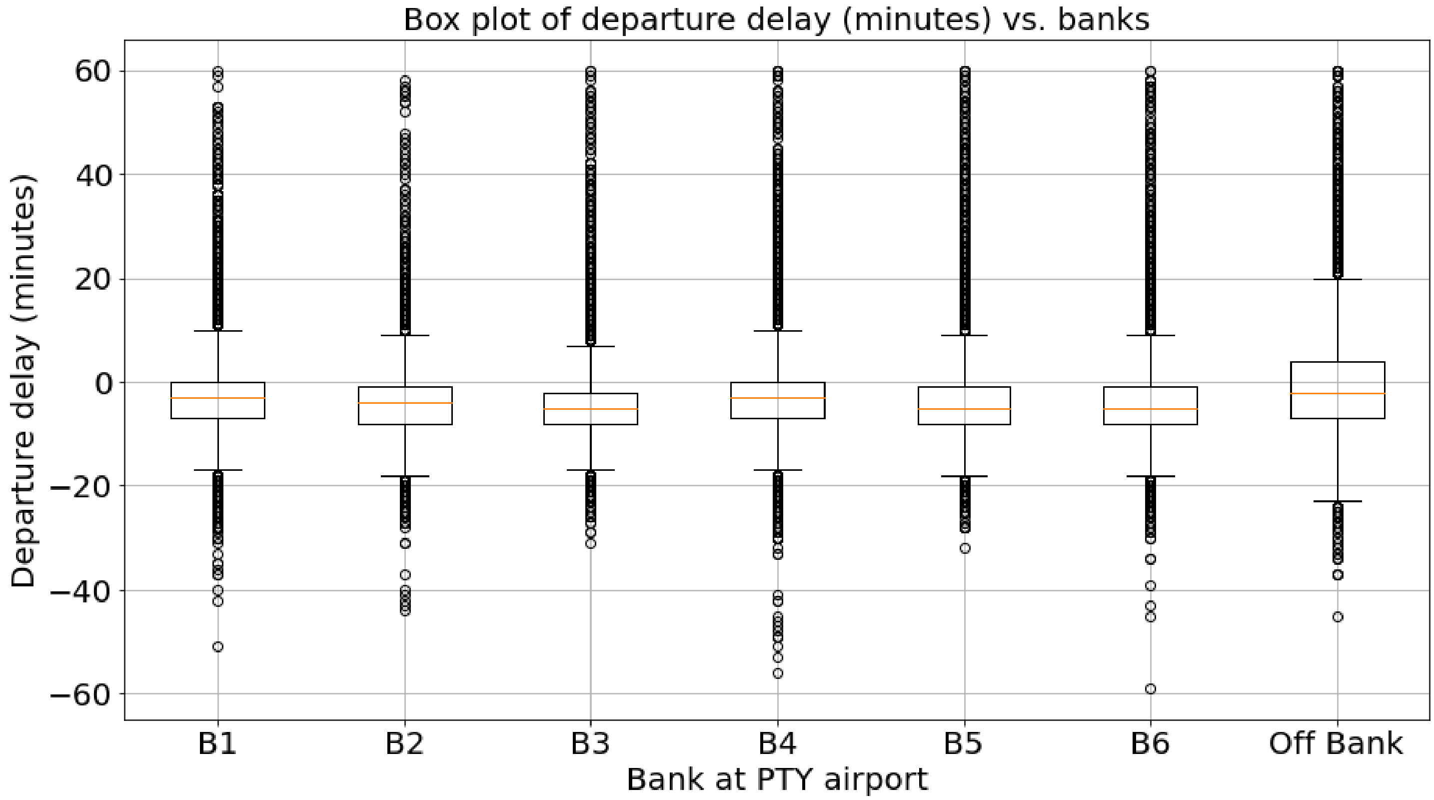


Figure 2.3.: Copa Airlines average departure delays for different banks at PTY airport.

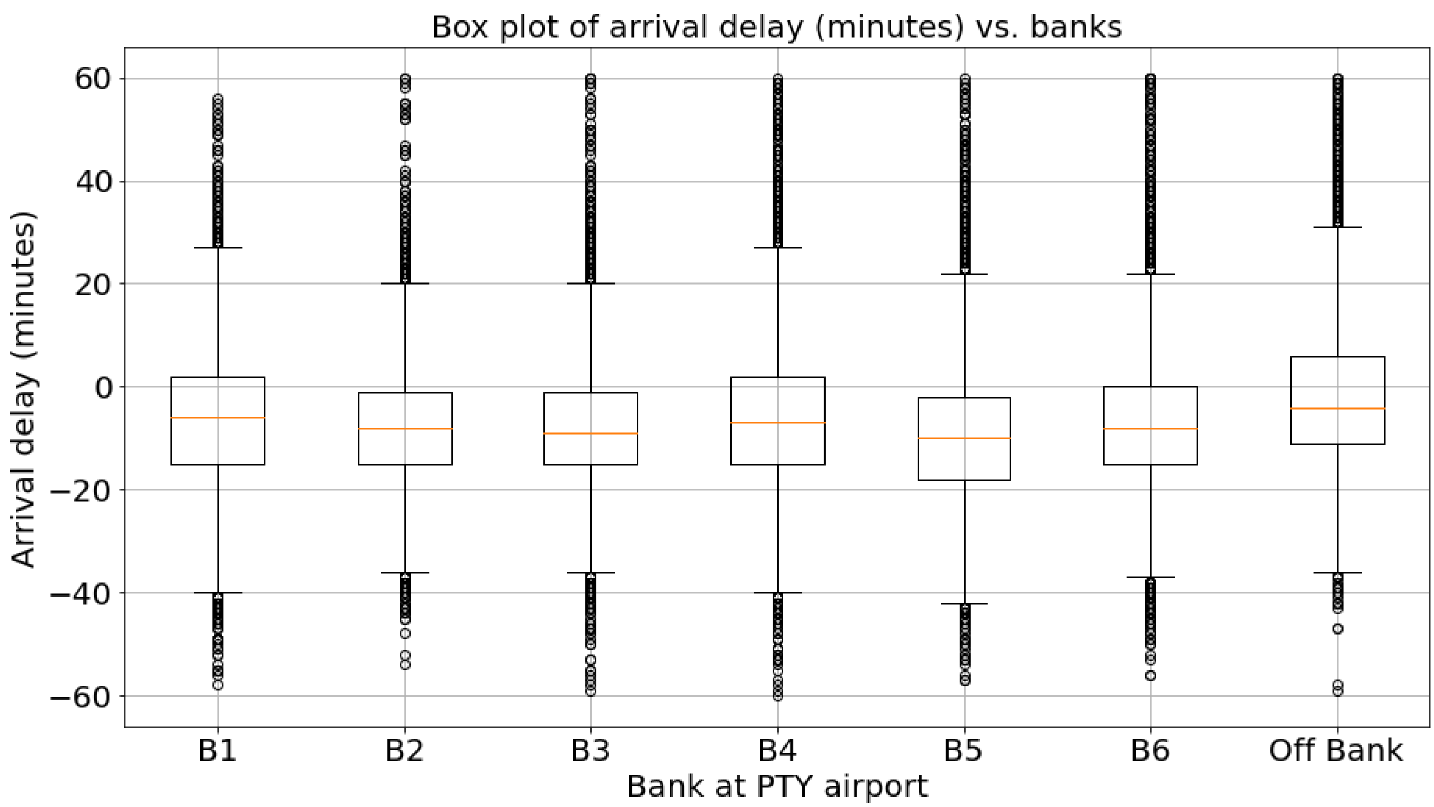


Figure 2.4.: Copa Airlines average arrival delays for different banks at PTY airport.

While average departure delay can be widely varied (and positive for banks B1, B4, and B6), the average arrival delay is small and negative (i.e., flights arriving early) for all banks. Only for off-bank flights, both average departure delay and average arrival delay are large.

2.2. Delay probabilities and their correlation, if any, with flight duration.

Rather than predict the exact delay a plane will incur during its trip, insufficient data suggests that delays be binned into intervals. We have analyzed (at [7]) analyzed the average delays (of both, departure and arrival) according to the following five delay categories, by a loop: 1 to 15 minutes, 16 to 30 minutes, 31 to 45 minutes, 46 to 60 minutes, and 61 minutes or more. Thus, there are ten analyses (five of average departure delays, and five of average arrival delays).

For each unique origin-destination (OD) pair, we calculated its delay probability in a particular delay category as:

$$P_C \quad = \quad \frac{N_C}{N_{OD}} \tag{2.1}$$

where, P_C is the probability of a flight being in a delay category, N_C is the number of flights of that origin-destination pair in that delay category, N_{OD} is the total number of flights of that origin-destination pair, **provided that the total number of flights of that OD is not less than a certain value (for example, 30)** since if there are too few flights, then the calculated probability may not be reliable. Then, we sorted the average delays in descending order, and plotted them with the average flight duration for those origin-destination pairs. This way, we may be able to see correlation, if any, between average flight duration and average delay (of departure or arrival). Input data is provided by Copa Airlines. Plots of only some of the delay categories are shown next, as they are all similar.

For each delay interval, we also calculate the Pearson's correlation coefficient between delay probability and flight duration, to see if these two quantities are correlated in any way. Pearson's correlation coefficient is defined as:

$$r \quad = \quad \frac{\sum_{i=1}^{n}[(x_i - \bar{x})\,(y_i - \bar{y})]}{\sqrt{[\,\sum_{i=1}^{n}\{(x_i - \bar{x})^2\}\,]\;[\,\sum_{i=1}^{n}\{(y_i - \bar{y})^2\}\,]}} \tag{2.2}$$

where, x_i and y_i are the data samples of the two quantities (x and y), n is the number of data samples in each quantity (must be same for both x and y), $\bar{x}$ is average of all x_i, and $\bar{y}$ is average of all y_i. Pearson's correlation coefficient is a real, dimensionless number between -1 and +1 (that is, $r \in [-1, +1]$). Correlation coefficient of -1 denotes perfect negative correlation, that is, as x increases, y decreases linearly. Correlation coefficient of +1 denotes perfect positive correlation, that is, as x increases, y also increases linearly. Correlation coefficient of 0 denotes that the two quantities are not correlated in any way.

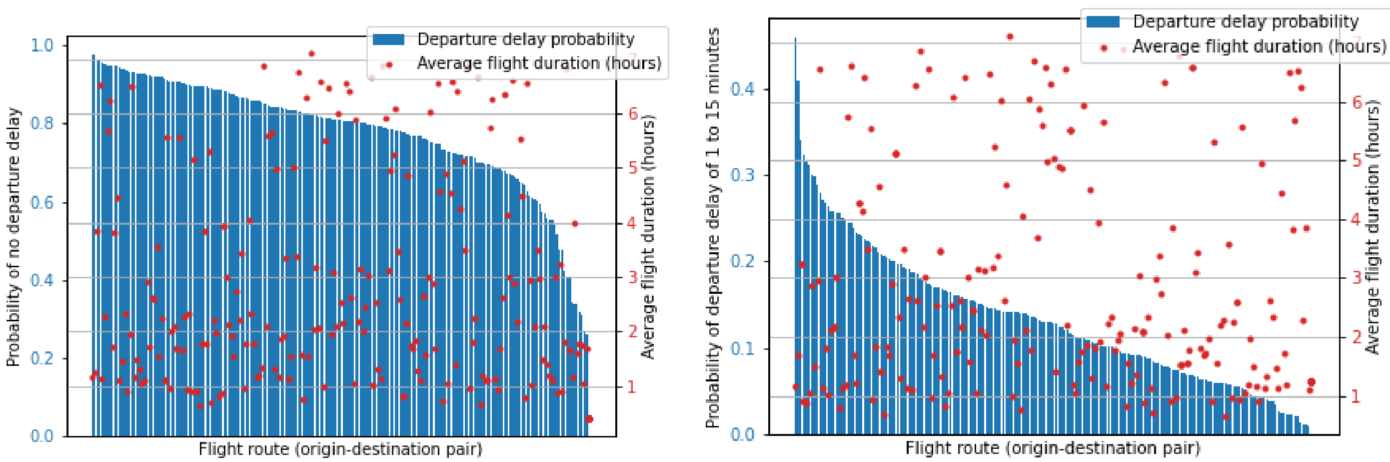
For both, average departure delay analysis and average arrival delay analysis, the plots for delay categories of 1 to 15 minutes and 16 to 30 minutes are shown in this chapter, and plots of other delay categories (31 to 45 minutes, 46 to 60 minutes, and more than 60 minutes) are displayed in the appendix.

2.2.1. Average departure delays

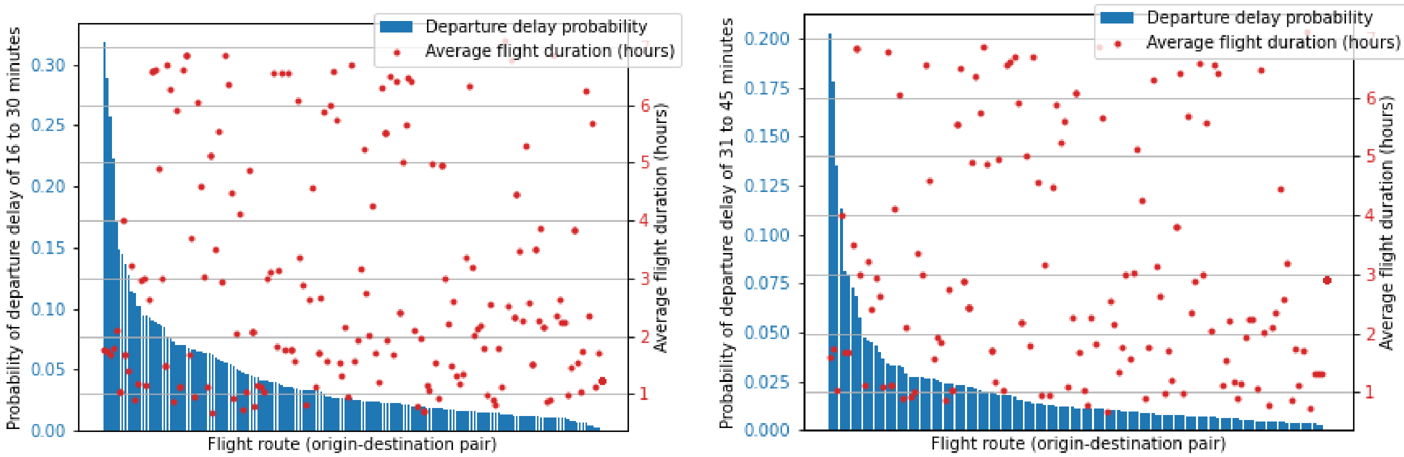
For each 4eparture delay category, the Pearson correlation coefficient between departure delay probability and flight duration are displayed in table 2.1.

Table 2.1: Correlation coefficient (between departure delay probabilities and average flight durations) and fraction of flights, for various departure delay categories.

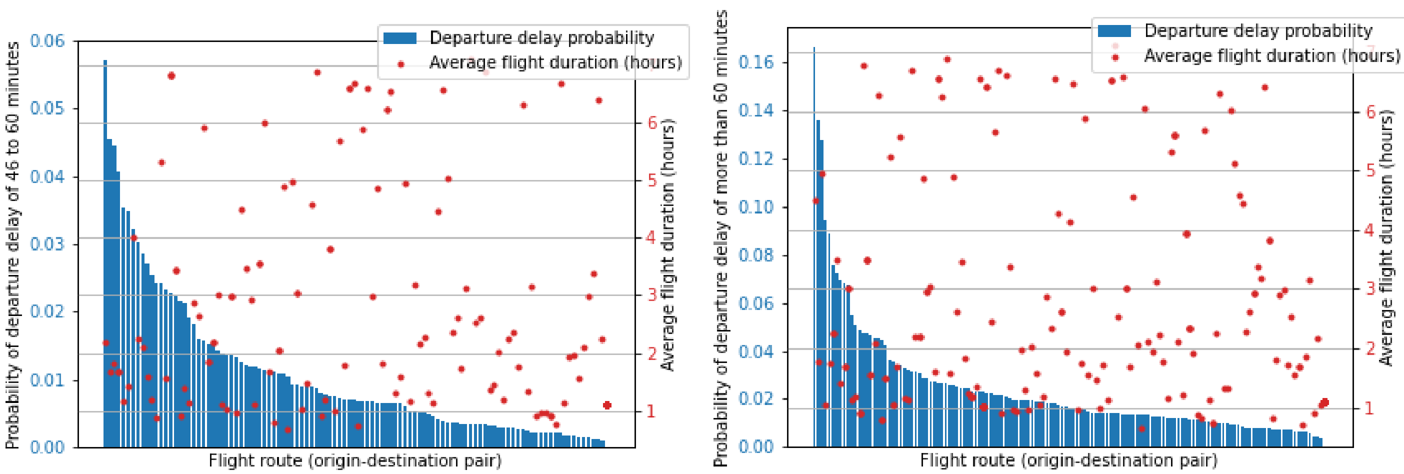
Departure delay category (minutes)	**Correlation coefficient between departure delay probabilities and average flight durations**	**Fraction of flights in this departure delay category**
≤ 0	0.17098	83.6 %
1 to 15	0.07274	11.5 %
16 to 30	0.00012	2.3 %
31 to 45	0.01290	0.8 %
46 to 60	-0.12173	0.3 %
≥ 61	0.02813	1.3 %



Figures 2.5 and 2.6: Departure delay probabilities and average flight durations, for different flight routes. Average departure delay: Left: none. Right: between 1 and 15 minutes.



Figures 2.7 and 2.8: Departure delay probabilities and average flight durations, for different flight routes. Average departure delay: Left: between 16 and 30 minutes. Right: between 31 and 45 minutes.



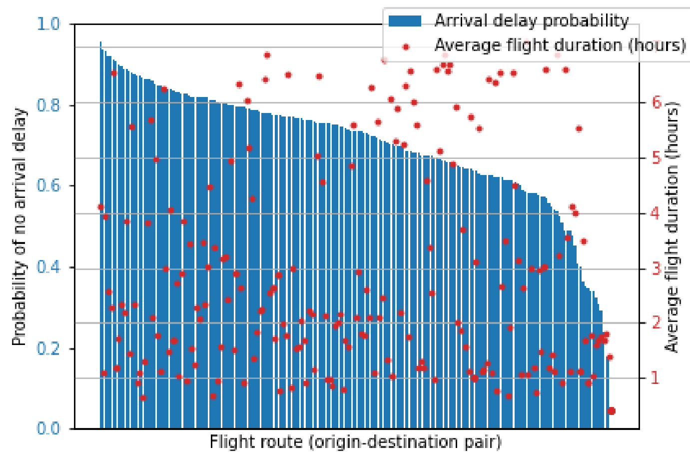
Figures 2.9 and 2.10: Departure delay probabilities and average flight durations, for different flight routes. Average departure delay: Left: between 46 and 60 minutes. Right: more than 60 minutes.

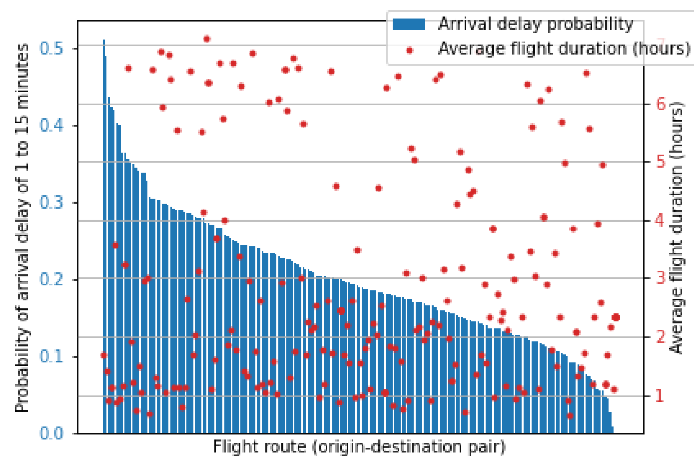
2.2.2. Average arrival delays

We now consider the number of flights with departure and arrival delay prob [1-15 min] less than 10 percent.[Count the number of OD pairs with departure prob in [1-15] less than 10% and do the same for arrival prob, and then give the numbers. Even better: How many OD pairs have a departure delay [1-15 min] between [0-10%], [10%-20%], [20%-30%], [30%-40%], [40%-50%]. Do this for departure and arrival, and then create a single bar plot, which has 5 sets of two bars: one bar for arrival and one bar for departure.

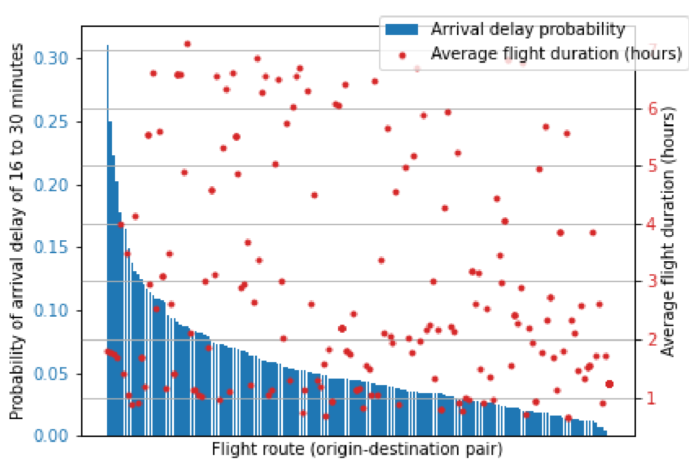
Table 2.2: Correlation coefficient between arrival delay probabilities and average flight durations, for various arrival delay categories.

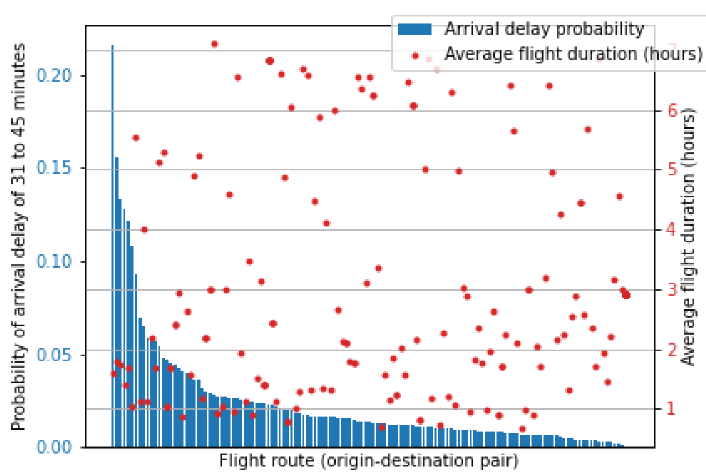
Arrival delay category (minutes)	Correlation coefficient between arrival delay probabilities and average flight durations	Fraction of flights in this arrival delay category
≤ 0	0.16510	75.2 %
1 to 15	0.07465	18.5 %
16 to 30	0.08033	3.7 %
31 to 45	-0.05864	0.9 %
46 to 60	-0.06731	0.4 %
≥ 61	0.02198	1.3 %



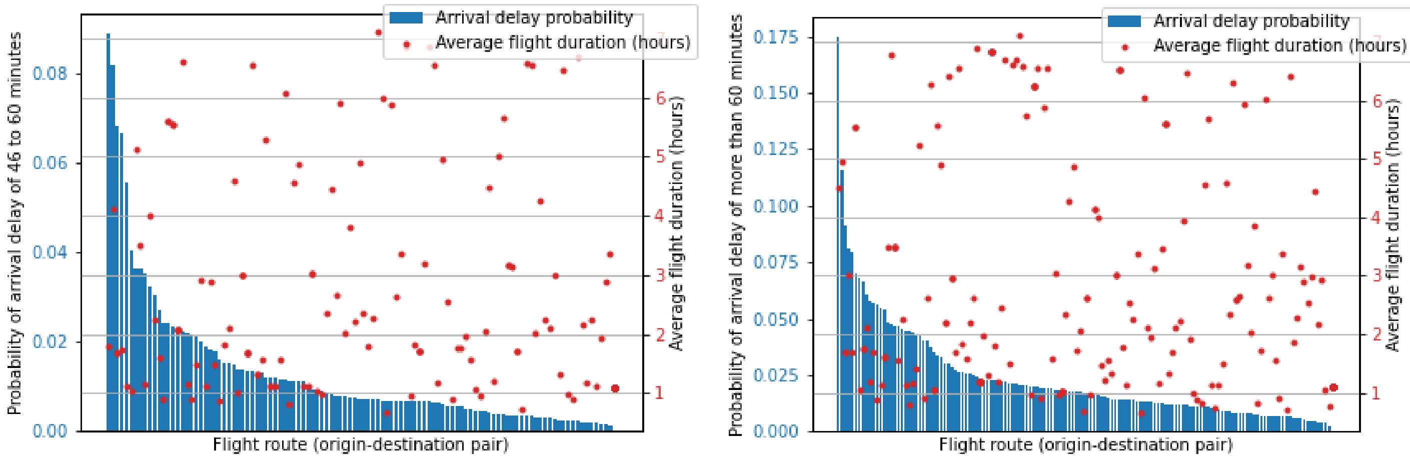


Figures 2.11 and 2.12: Arrival delay probabilities and average flight durations, for different flight routes.
Average arrival delay: Left: none. Right: between 1 and 15 minutes.





Figures 2.13 and 2.14: Arrival delay probabilities and average flight durations, for different flight routes.
Average arrival delay: Left: between 16 and 30 minutes. Right: between 31 and 45 minutes.



Figures 2.15 and 2.16: Arrival delay probabilities and average flight durations, for different flight routes. Average arrival delay: Left: between 46 and 60 minutes. Right: more than 60 minutes.

For both departure and arrival delays, the combined plots of delay probability and average flight durations (each plot for a different delay category) do not seem to suggest any strong correlation between delay probability and flight duration. These two quantities seem to be independent of each other. (If they were correlated, then flight duration should show some trend – either consistently decreasing or consistently increasing, as delay probability decreases. But we see that it is spread randomly in all the plots, as delay probability decreases.)

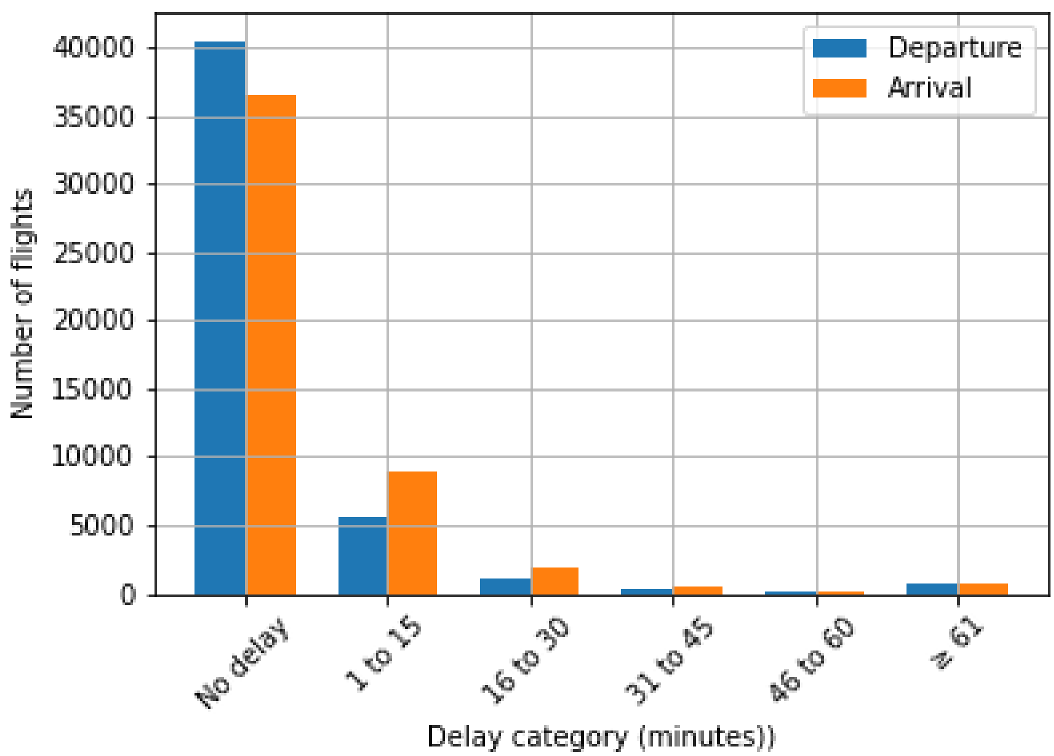


Figure 2.17: Distribution of flights in different delay categories of departure and arrival.

3. Data Visualization

3.1. Visualization of scheduled departure and arrival times with respect to different banks at PTY airport.

As a visualization exercise, we created plots of scheduled departure and arrival times in Zulu (GMT), colored according to the flight's bank at PTY airport. [8].

Figures 3.1 and 3.2 show the various flights plotted according to their scheduled departure time in Zulu time (GMT) and color-coded according to their banks at PTY airport. In figure 3.2, the flights are sorted in ascending order of their scheduled departure times. Similarly, figures 3.3 and 3.4 show the flights plotted according to their scheduled arrival times in Zulu time (GMT) and color-coded according to their banks at PTY airport. These figures are plotted to show if there is any relationship between flights' scheduled arrival or departure times and their banks at the PTY airport. We observe how the different banks at PTY correspond to different times of the day when flights are scheduled to arrive at or depart from PTY, which is the hub airport of Copa Airlines. (Local time at PTY is five hours behind GMT. So, if it is 6 pm local time at 0^o longitude, then it will be 1 pm local time at PTY.)

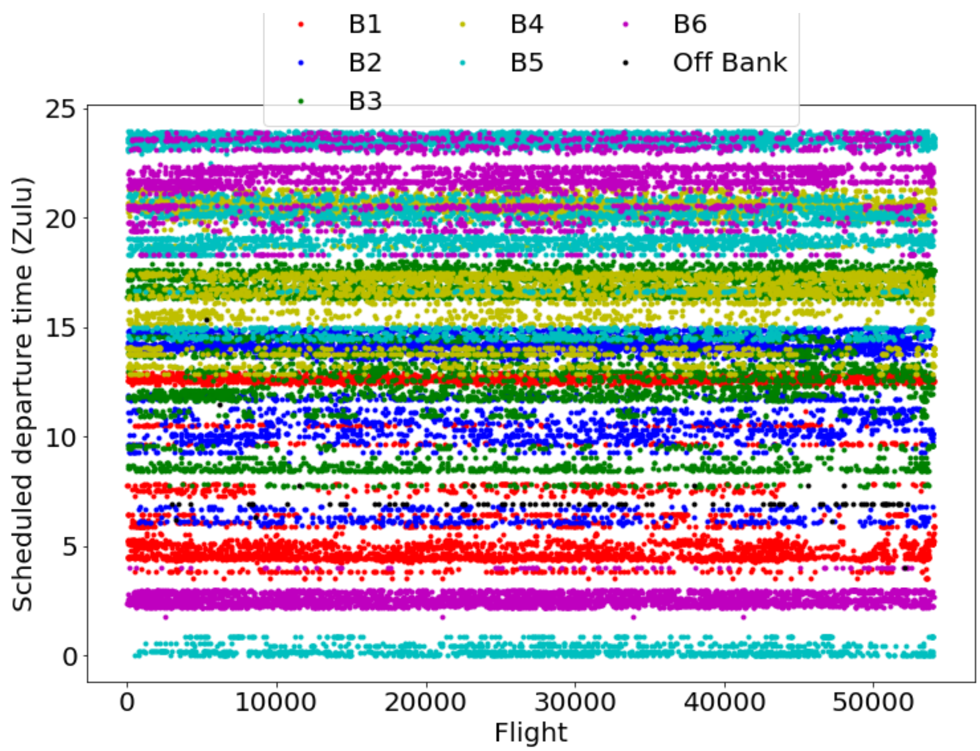


*Figure 3.1.: Scheduled **departure** times (Zulu) of Copa Airlines flights for different banks at PTY airport.*

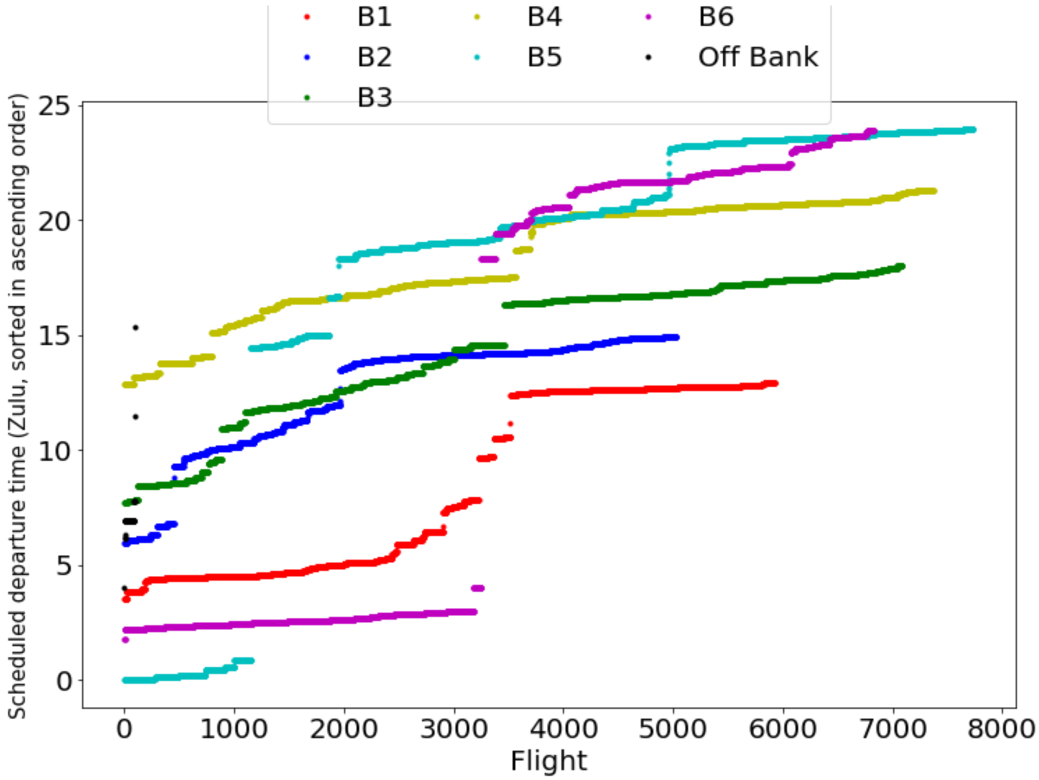


*Figure 3.2.: Scheduled **departure** times (Zulu) of Copa Airlines flights, **sorted** in ascending order, for different banks at PTY airport.*

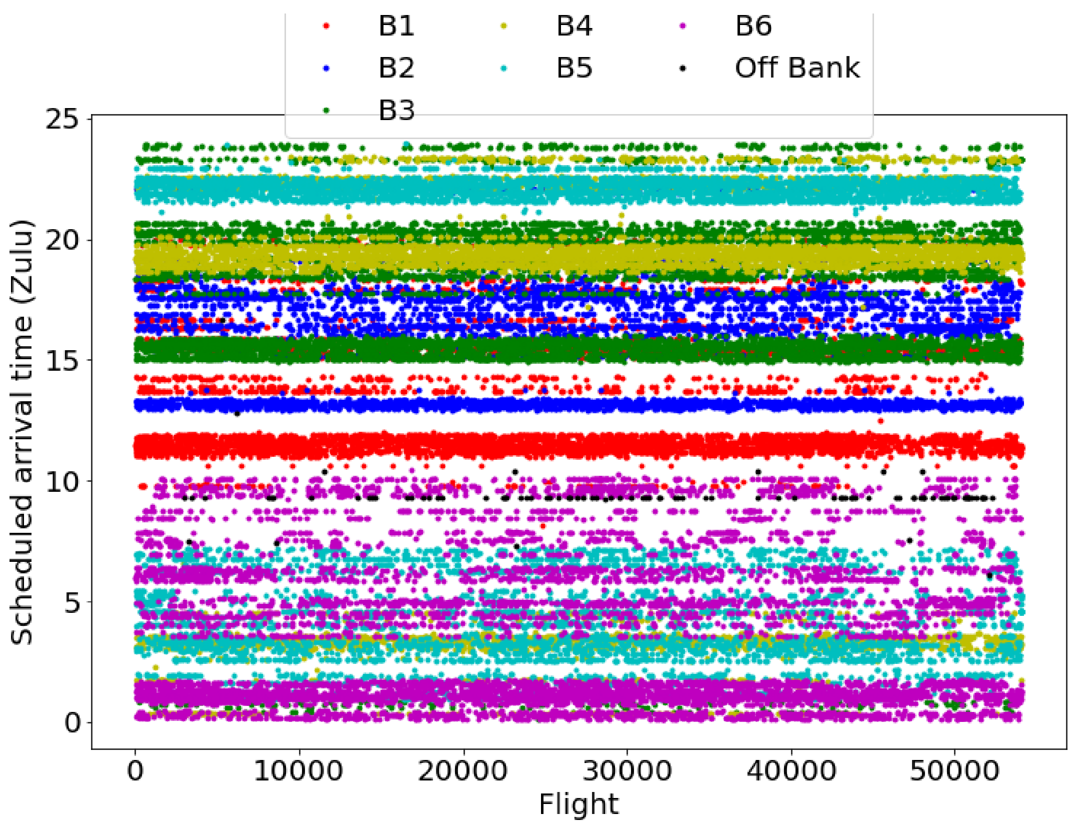


*Figure 3.3.: Scheduled **arrival** times (Zulu) of Copa Airlines flights for different banks at PTY airport.*

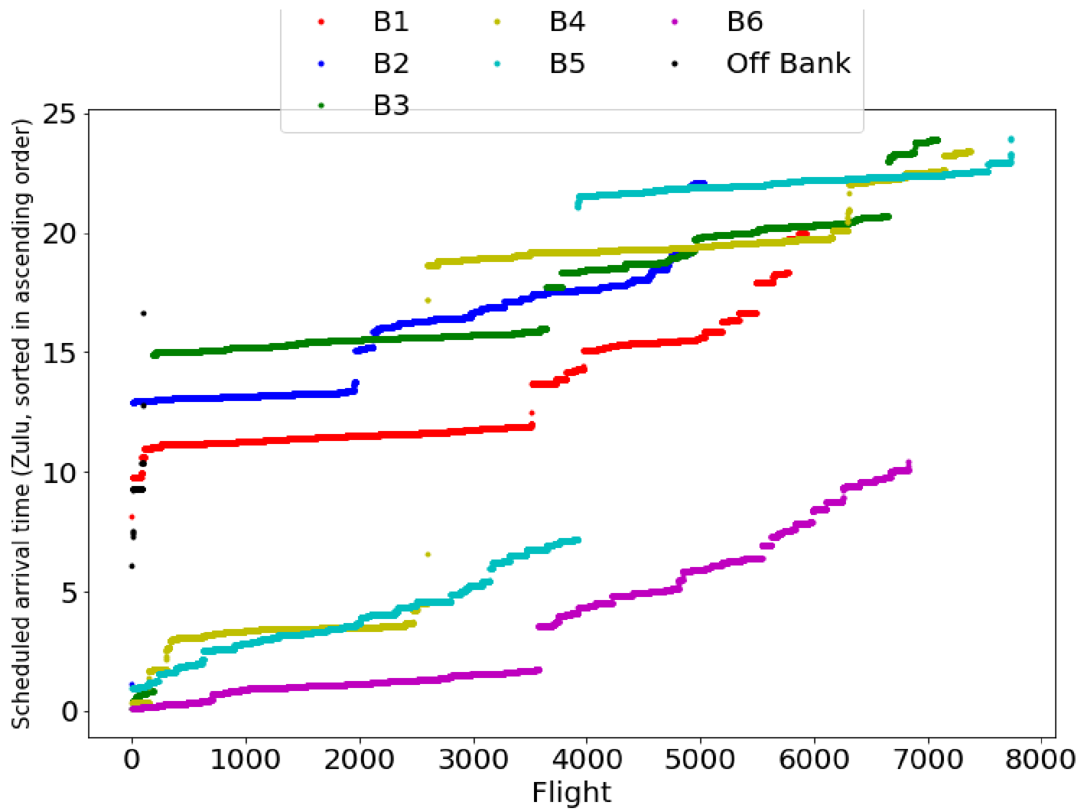


Figure 3.4.: Scheduled ***arrival*** *times (Zulu) of Copa Airlines flights. The flights on the horizontal axis are* ***sorted*** *in ascending order of arrival time, separately for each banks at PTY. .*

From figures 3.1 to 3.4, we see that, as expected, the different banks correspond to different times of the day when flights arrive at and depart from PTY airport.

3.2. Approaches to visualization of flight networks

A "network", when represented for visualization, is depicted as a graph, built from nodes and edges. A node connects to one or more to more nodes, via links. A link connects two nodes. More precisely, a graph *G(V,E)* is defined as a set *V* of *N* nodes v_i and a set *E* of edges $e_{i,j}$ between nodes *i* and *j.*

Two common paradigms to visualize flight networks is the use of either airports or flights as nodes [citations]. In the former case, edges represent the flight between two airports [citations]. In the latter case, edges represent the transfer of resources (passengers, crew, luggage) as links (edges) between nodes. Other methods may be possible. Deciding which method to select depends on how we want to use the available data. We discuss these two methods next in some detail.

1.1.1. Airports as nodes, and flights as links between nodes.

This seems the more intuitive of the 2 methods, as it is closer to the real-world representation (nodes are physical locations and flights are the links between those locations). It is what most people may first think of, when visualizing flight networks. Example schematic is shown in figure 3.5.

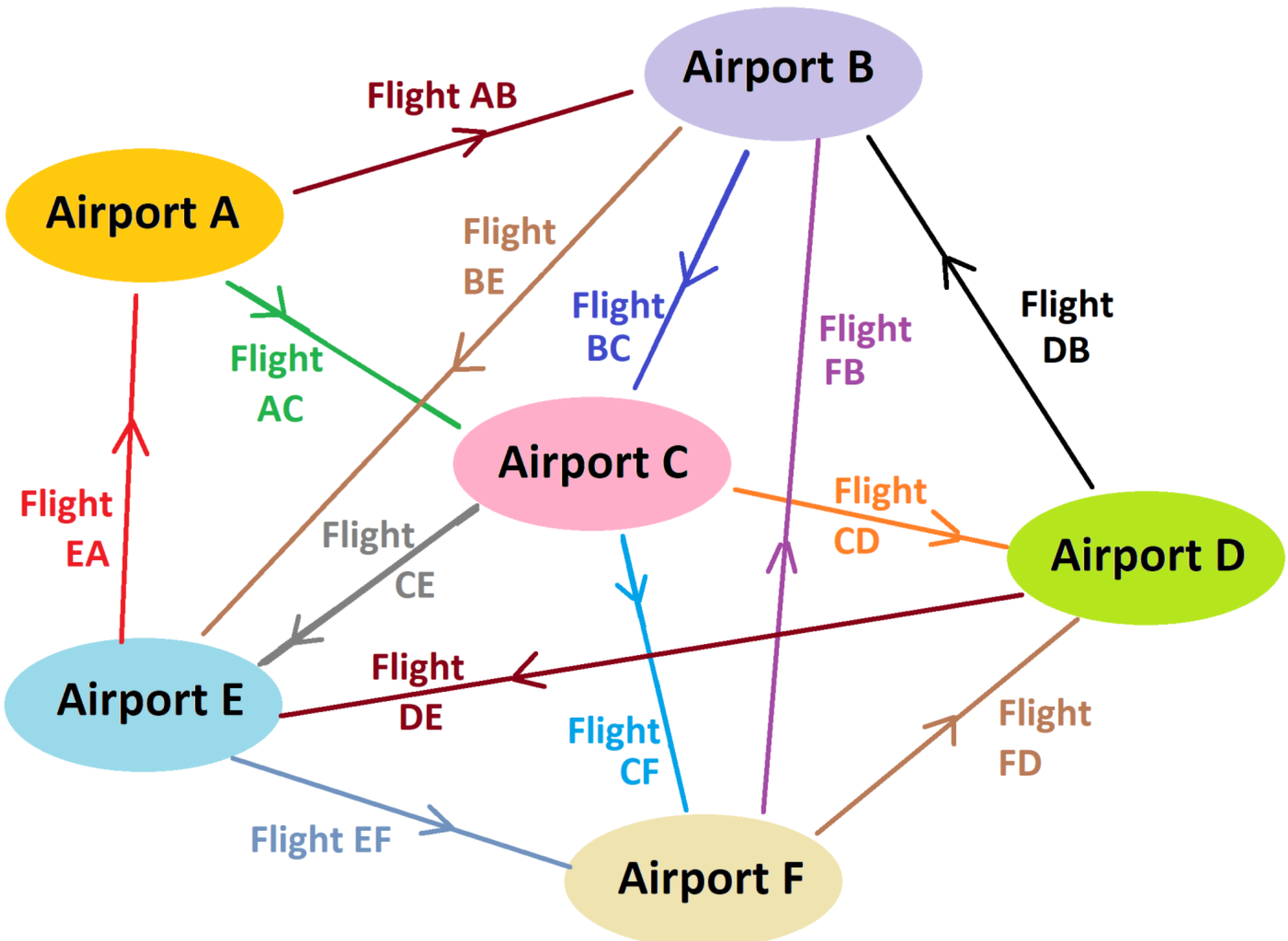


Figure 3.5.: Example schematic of flight network visualization where airports are nodes and flights are edges (links) between nodes.

1.1.2. Flights as nodes, and transfer of resources as links between nodes.

Here, flights are the nodes of the network, and transfer of resources (passengers, crew, luggage) as links (edges) between nodes. This seems a more unconventional way of representing a flight network, as it is not exactly analogous to the real world. But it may be useful when there are many connections (transfers) between flights. Example schematic is shown in figure 3.6.

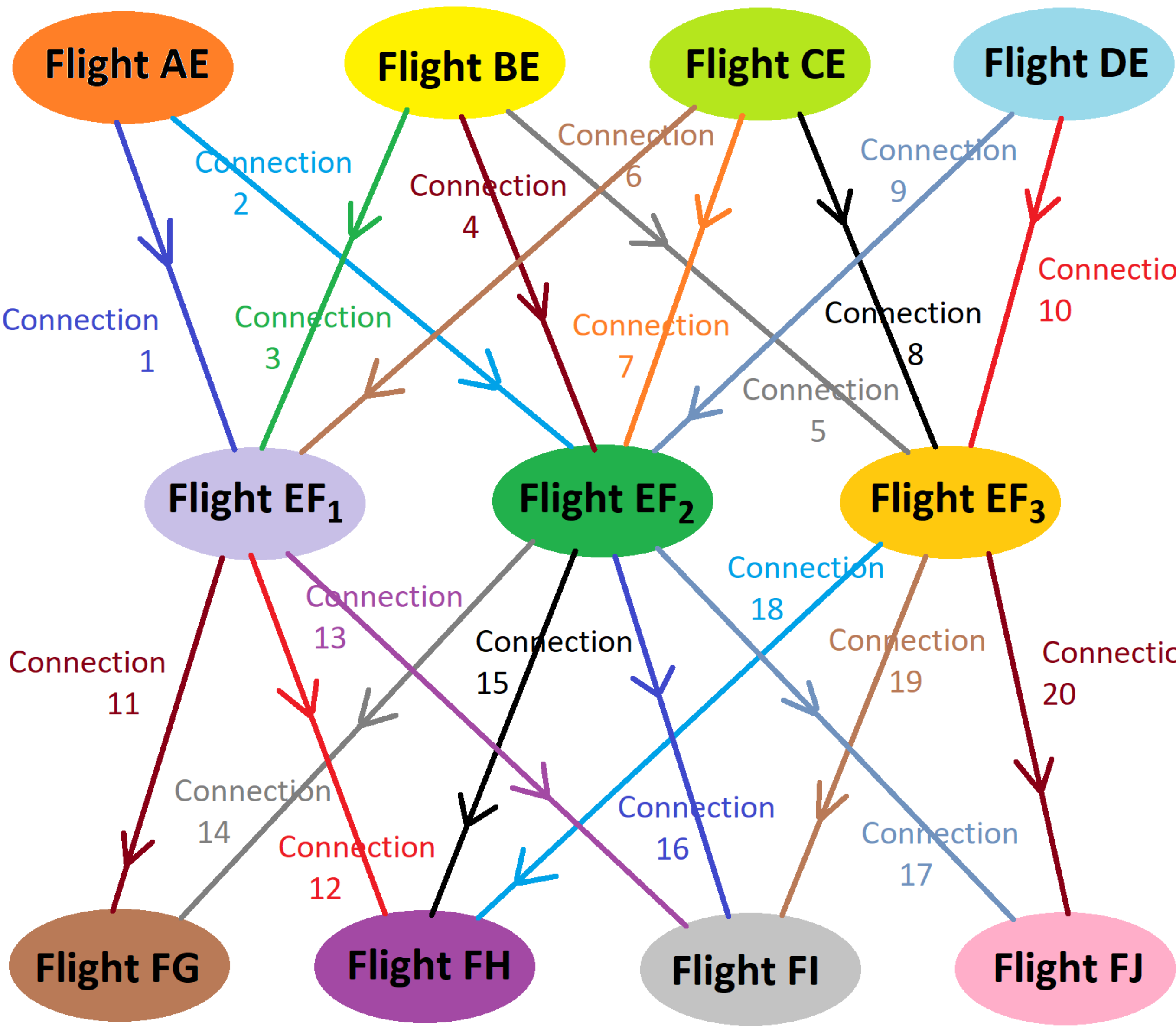


Figure 3.6.: Example schematic of flight network visualization where flights are nodes and transfer of resources (passengers, luggage, crew) are edges (links) between nodes.

1.1.3. Method selected for our analysis

In our work of visualizing the flight network of Copa Airlines, we are using the second method [flights as nodes, and connections between those flights as edges (links) between nodes]. Reasons include the following. For a flight network like that of Copa Airlines, where a particular airport (e.g., PTY) is the hub (i.e., almost all flights have that airport as either origin or destination), it makes more sense to visualize flights as nodes. Having airports as nodes will just create a hub-and-spoke visualization for every visualization case, where airports other than the hub would be connected to a single node (the hub) via a single edge. Having flight connections as edges may reveal important data about: transfers of resources (passengers, crew, cargo, etc.) between different flights; aircraft maintenance between two consecutive flights; etc. Above may be difficult using the first method (airports as nodes and flights as edges).

3.3. Computational tools considered for flight network visualization

We explored several visualization tools, such as Plotly [9], Altair [10], Streamlit [11], React [12], Vue [13], etc. for visualization of flight networks. We have described our work based on Plotly and NetworkX [14], with the support of Pandas [15] and NumPy [16] for cleaning and filtering tasks. (Data cleaning and filtering include: not considering flights with the same origin and destination, as they are experimental flights rather than commercial flights; avoiding flights with missing values of the important attributes like departure delay and arrival delay; etc.)

3.4. Random flights and connections

As a first exercise in flight network visualization, I tried selecting some random flights from the data provided by Copa Airlines, and visualize their network by having flights as nodes and passenger transfers as edges between nodes. Jupyter notebook is at GitHub location [17]. Packages used for data extraction and cleaning include NumPy and Pandas, and for data visualization include Plotly.graph_objects and NetworkX.

Properties of the created graph include: nodes representing flights, with color red for inbound flights (to PTY), and green for outbound flights (from PTY). The node size is directly proportional to flight duration (but can be some other quantity). Edges represent transfer of passengers from inbound (feeder) flights to outbound (non-feeder) flights. Their color is directly proportional to number of passengers transferred (dark for more passenger transfer, light for less passenger transfer). Thickness is constant for all edges, but can be changed to vary according to some parameter.

Two possible layouts of nodes are either random or circular.

Output visualization examples:

This type of visualization is helpful to see the flight connections with large number of passenger transfers. In the visualization with the random arrangement of nodes, there is not much clarity. But when the nodes are arranged in a circular manner, there is better visual appeal and clarity. We also see that there are some nodes that are not connected to any other node in the figure. This is because there are no connections from those flights to any other flight in the set of selected flights. But there may be connections to other flights not included in the set of selected flights.

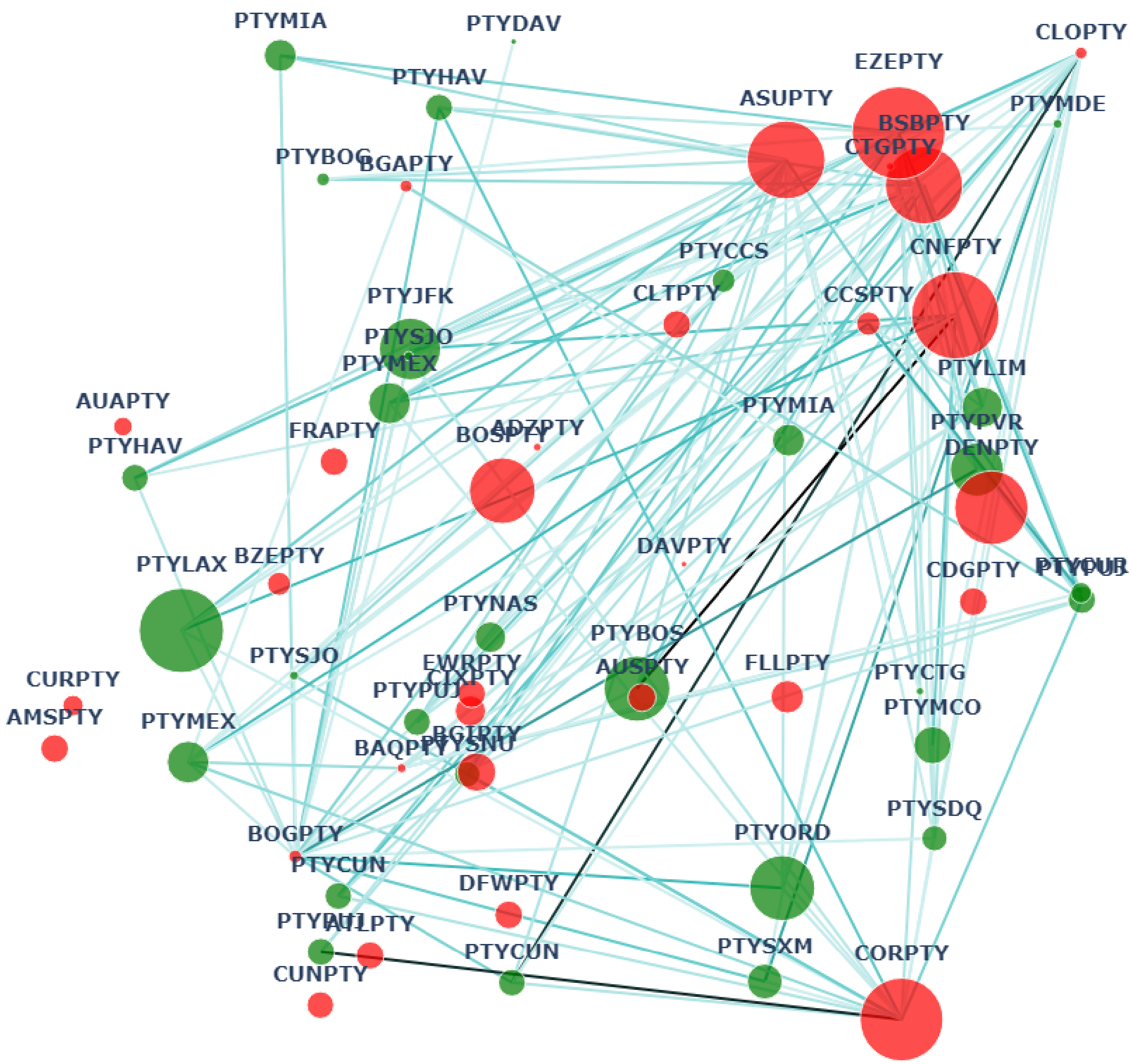


*Figure 3.7.: Flight network visualization using Plotly and Networkx, with **random** layout of nodes.*

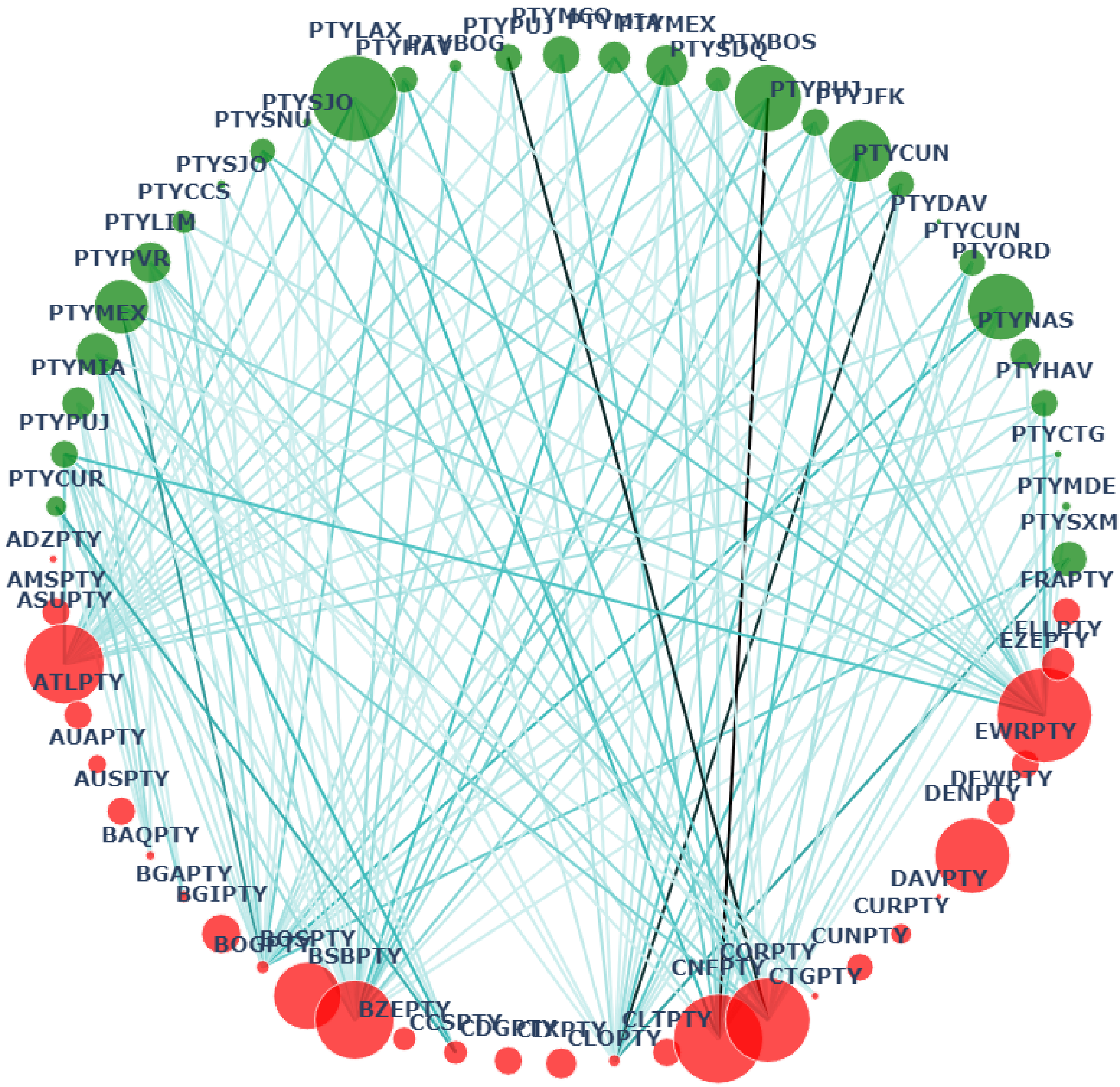


Figure 3.8.: Flight network visualization using Plotly and Networkx, with ***circular*** *layout of nodes.*

3.5. Tree showing one inbound flight, outbound flights fed by it, and return flights of same aircraft to PTY

A tree visualization of a single inbound flight, and outbound flights fed by it, may help in identifying, analyzing, and reducing delay propagation in flight network. The Jupyter notebook [18] uses the libraries *Pandas* and *NumPy* for data extraction and cleaning, and *plotly.graph_objects* and *igraph* for data visualization.

In the generated tree, the nodes denote flights, with their color indicating the arrival delay categories (can be changed to departure delay, if required). There are seven categories (most of 15-minute delay intervals): no delay, 1 to 15 minutes delay, 16 to 30 minutes delay, 31 to 45 minutes delay, 46 to 60 minutes delay, 61 or more minutes delay, and data not available. The flight's origin-destination code of six characters is displayed on node.

The tree's structure has the nodes arranged in three horizontal rows or levels: the top level for feeder (inbound) flight, the middle level for outbound flights getting passengers from feeder flight above, and the bottom level for return (inbound) flights of the same aircraft (tail) as the outbound flights above. Passenger transfer direction is from top to bottom.

Hover information display (through tool-tip) is enabled on the generated tree. When the cursor is made to hover over a node, the following data of the flight is displayed in a box whose color is same as node color (depending on delay category, as shown in legend). The 11 items displayed in tool-tips are: origin-destination code, aircraft tail number, departure delay in minutes, arrival delay in minutes, flight number, flight duration in hours, bank code at PTY airport, scheduled departure date (in Zulu time), scheduled departure time (Zulu), scheduled arrival date (in Zulu time), and scheduled arrival time (Zulu).

Output visualization examples: Example of output visualization is shown in figure 3.9. Hover information (tooltips) are shown in figure 3.10.

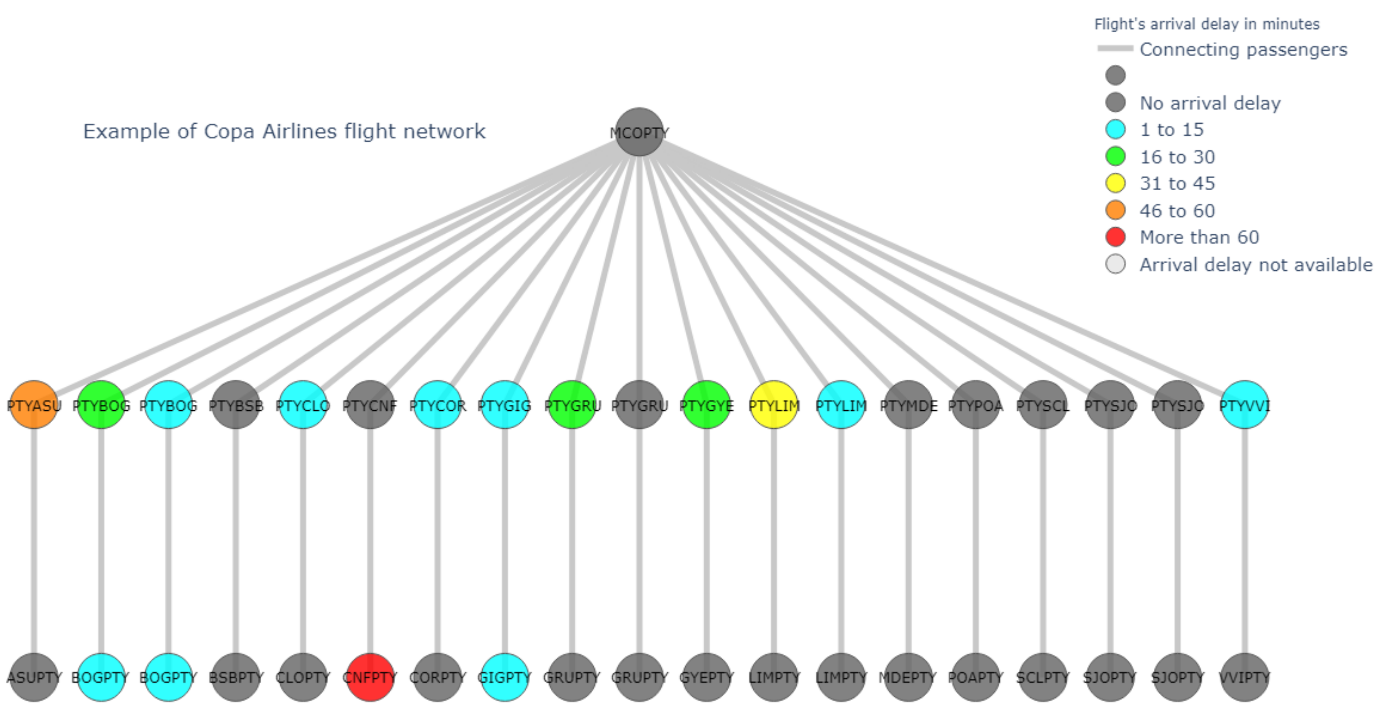


Figure 3.9.: Example of output visualization, for Copa Airlines flight network.

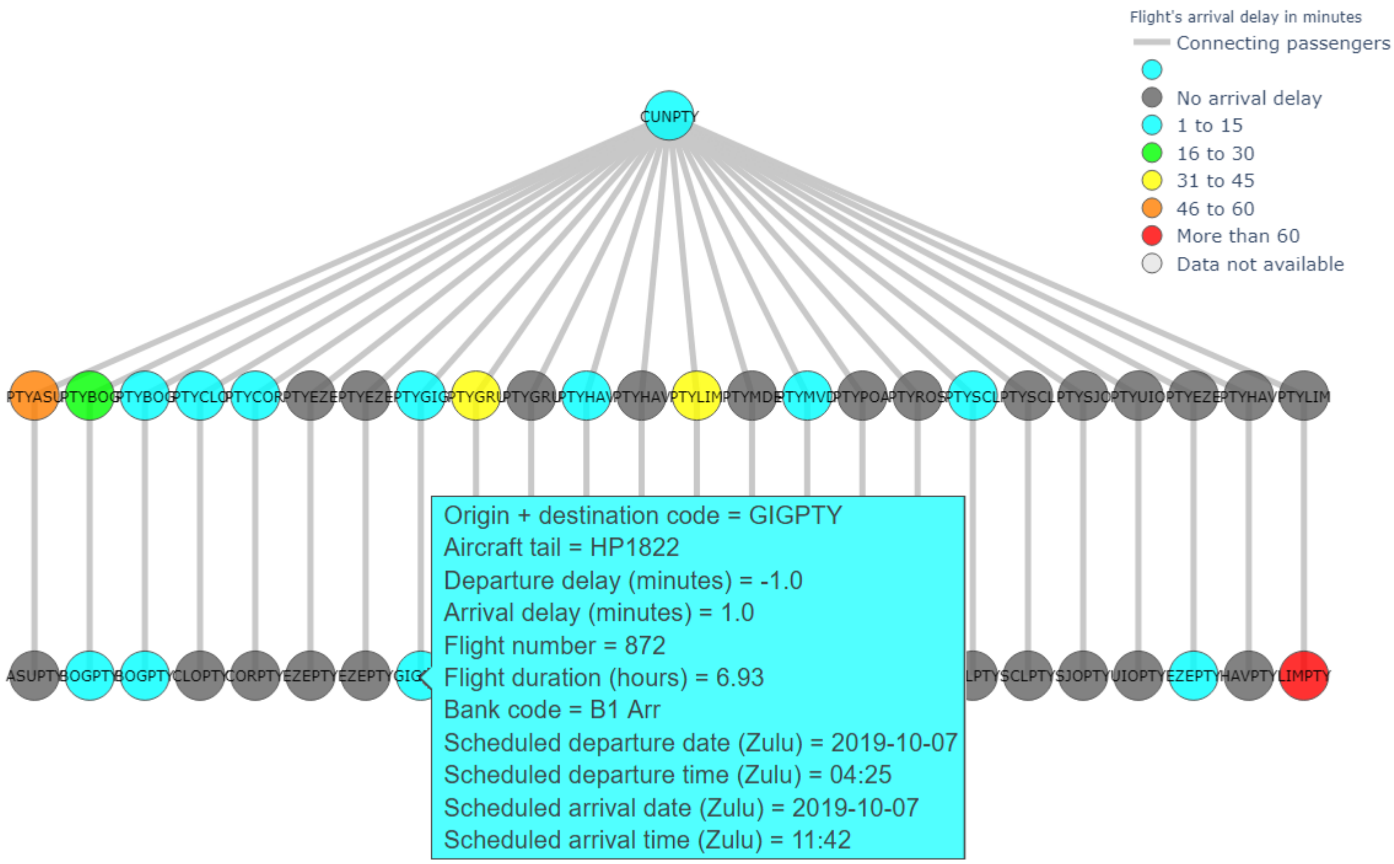


Figure 3.10.: Colors of delay categories, and tool-tips (hover information).

3.6. Displaying multiple feeders of a selected outbound flight

In the previous visualization, a single feeder flight is considered for the top level. But for any outbound flight in the middle level, the delay depends on many factors, including all the feeder flights (not just one). So, it may be better to select one outbound flight, and show all its feeder flights above it, as shown in schematic in figure 3.11. Thus, in this visualization, there are four horizonal rows or levels of flight nodes: the first level (top), where all the feeder flights for the chosen outbound flight; the second level, which displays the chosen outbound flight; the third level, showing the return (inbound) flight of the same aircraft (that is, same tail number) as the outbound flight above; and finally, the fourth level (bottom), showing the outbound flights fed by the return (inbound) flight above. Again, passenger transfer direction is from top to bottom.

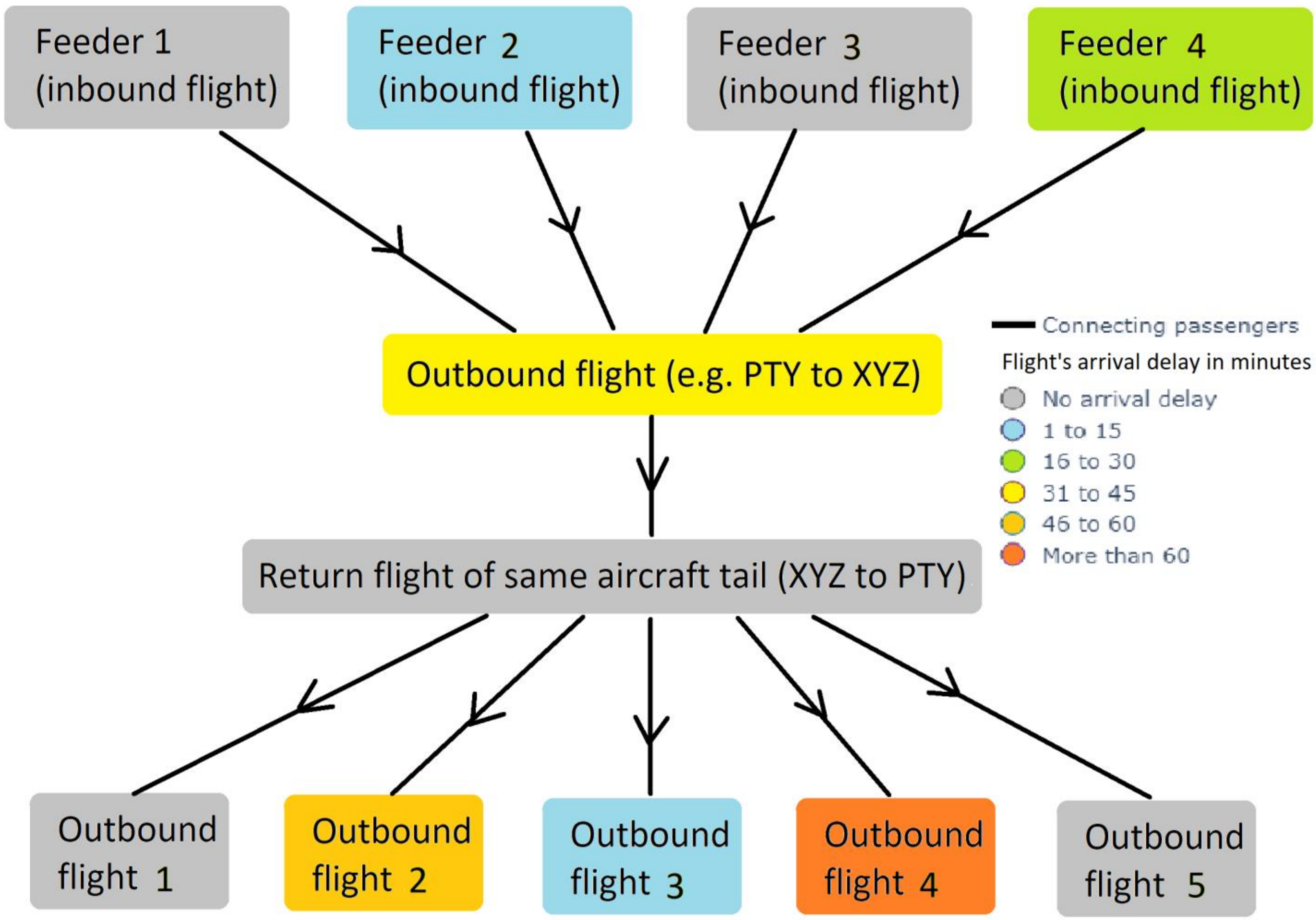


Figure 3.11.: Schematic of flight network visualization displaying all feeder flights of a selected outbound flight, and its return flight, and outbound flights fed by that return flight.

The Jupyter notebook [19] takes the input data and uses libraries *Pandas* and *NumPy* for data extraction and cleaning, and *plotly.graph_objects* and *igraph* for data visualization. As before, nodes represent flights, and their color depends on category of arrival delay (can be changed to departure delay if required). Each flight's origin-destination code is displayed on node. Selection of outbound light of second level is done through a variable that stores a random integer typed by user, which is used as row number in input data file to select the outbound flight. Thus, to work correctly, the integer typed must be less than the number of rows in the input data fileof flights.

An addition could be showing the different types and quantities of resource transfer (for example, number of passengers transferred, crew transfer, aircraft transfer, total luggage items or mass transferred, etc.) when we hover over the links between the nodes.

Output visualization examples:

Examples of output visualization and hover information (tool-tips) are shown in figures below. If the return (inbound) flight's complete data is not available in the input data files, then data of the outbound flights fed by the return flight can not be obtained. Hence, the fourth level of nodes is not generated, as shown in figure 3.13.

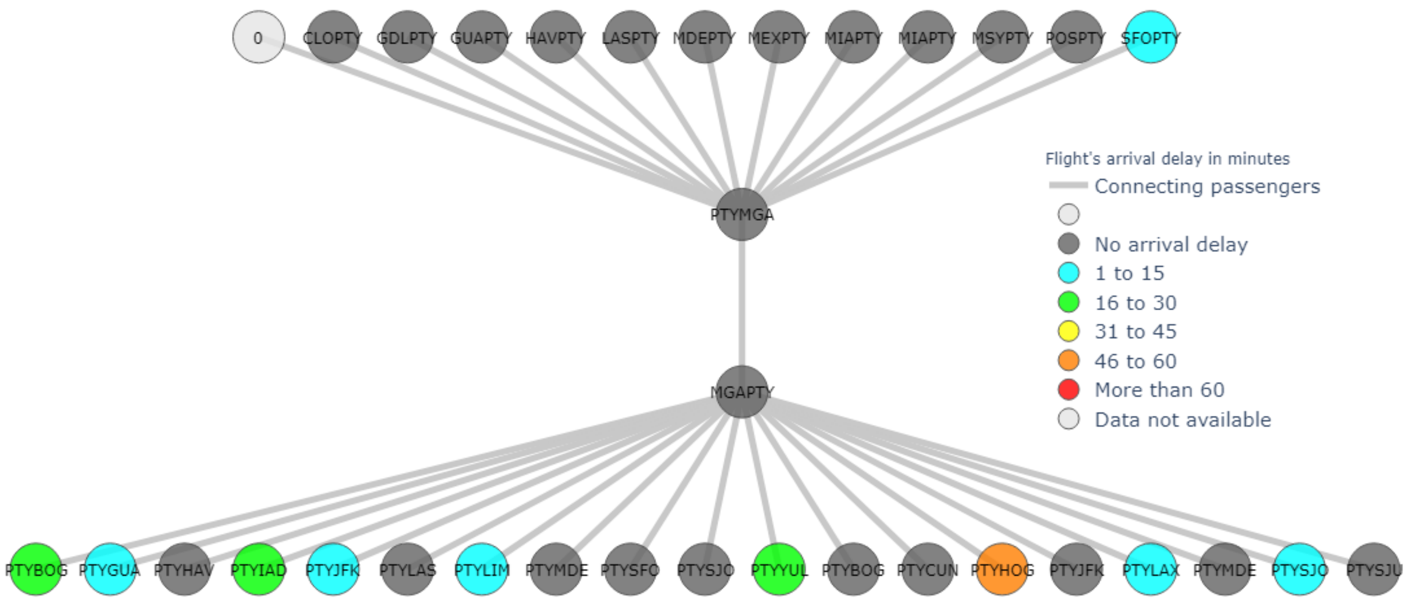


Figure 3.12: Example flight visualization generated.

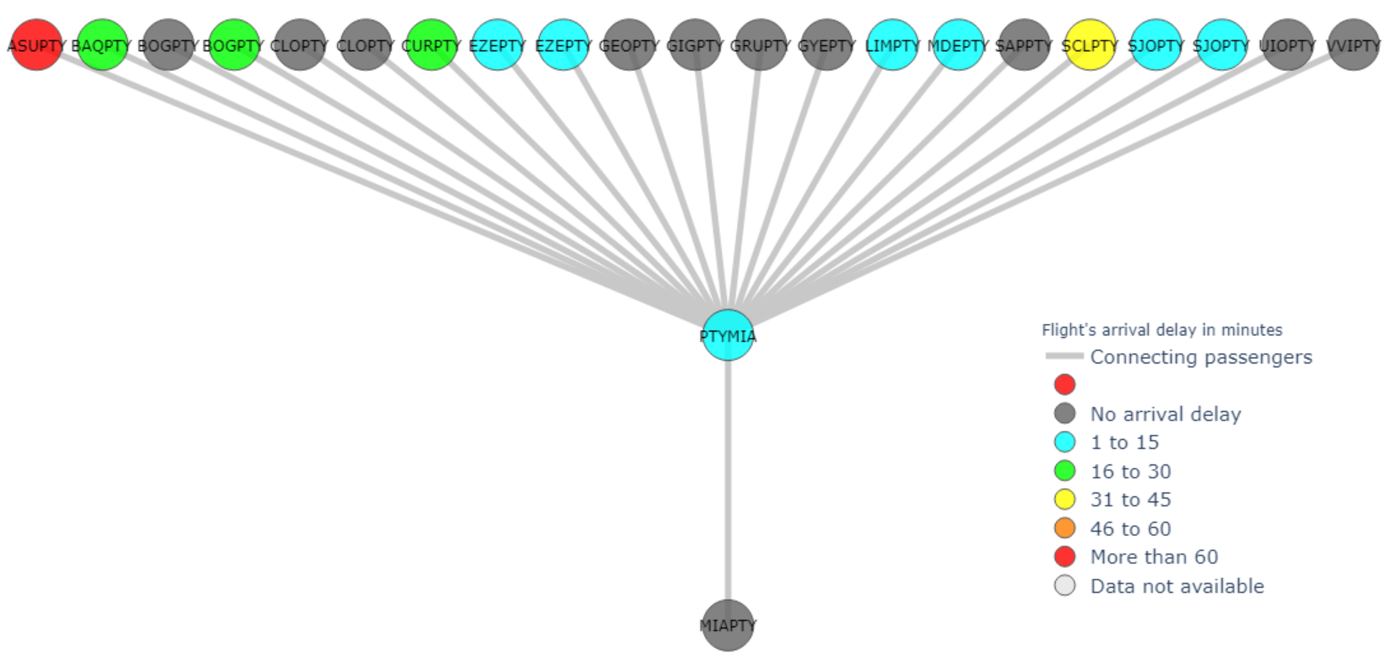


Figure 3.13: Example of flight visualization where the return (inbound) flight's all data. is not available in the input data files. Hence, data of the outbound flights fed by the return flight can not be obtained, and the fourth level of nodes is not generated.

As before, there is hover information (tooltips) which shows data of a flight when mouse cursor hovers over node of that flight, as shown in figure 3.14. tool-tip box is of same color as node, which depends on flight's category of arrival delay (can be changed to departure delay).

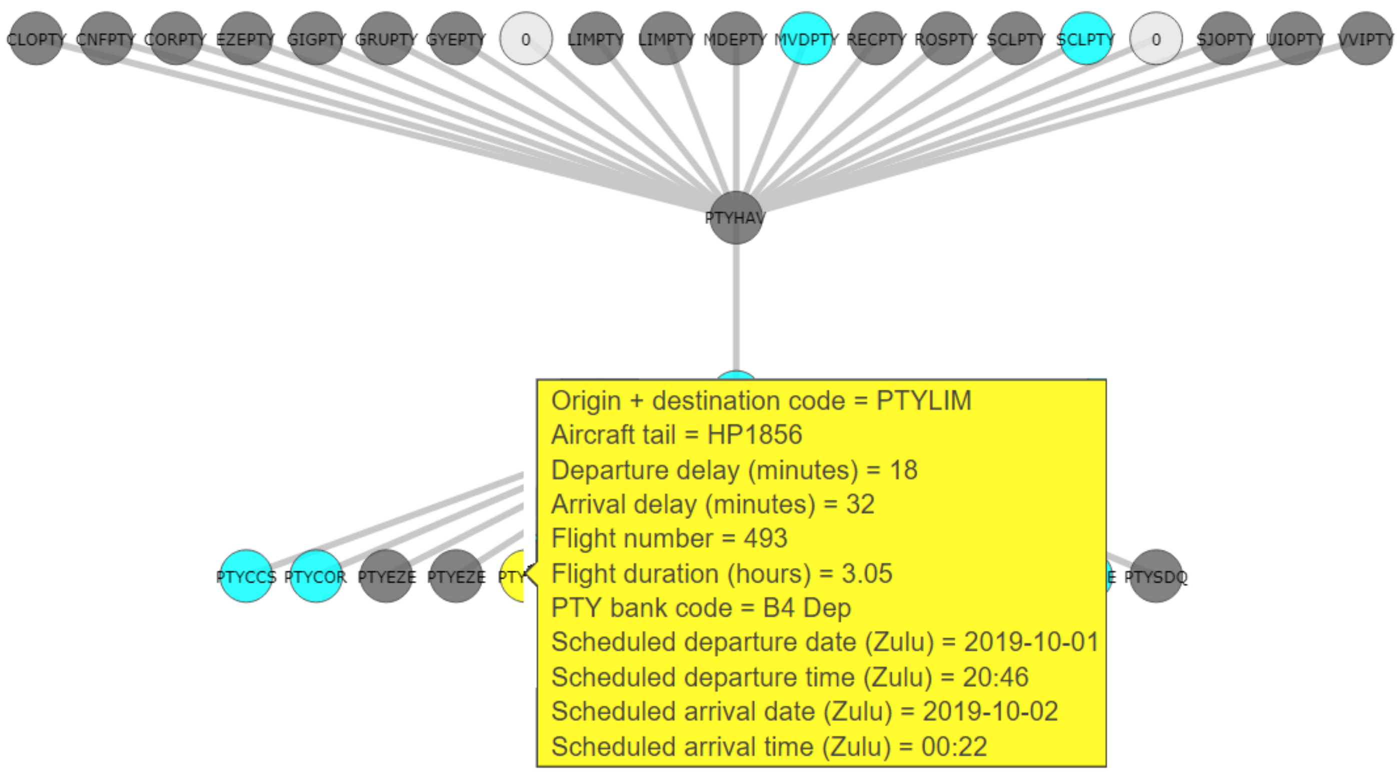


Figure 3.14: Example of hover information (tooltips), as seen in Jupyter notebook.

3.6.1. Incorporating widget (dropdown menu) to select the outbound flight of second level

This is an improvement where the outbound flight of second level is selected by a user-friendly widget of dropdown menu that displays a list of all outbound flights' unique flight IDs (collection of all unique values from column *id_nf* of input data file "*bookings_ids+pax_date.csv*").

The Jupyter notebook [20] uses the libraries *Pandas* and *NumPy* for data extraction and cleaning, *plotly.graph_objects* and *igraph* for data visualization, and *ipywidgets.widgets* for including dropdown menu widget. Remaining details are mostly similar to above.

The next three images show screenshots of the visualization outputs generated when the Jupyter Notebook executes successfully. Figure 3.15 shows the dropdown menu widget (for selecting outbound flight of second level). Figure 3.17 shows the hover information (tool-tips), that gives data about a flight when the cursor is hovered over the node of that flight.

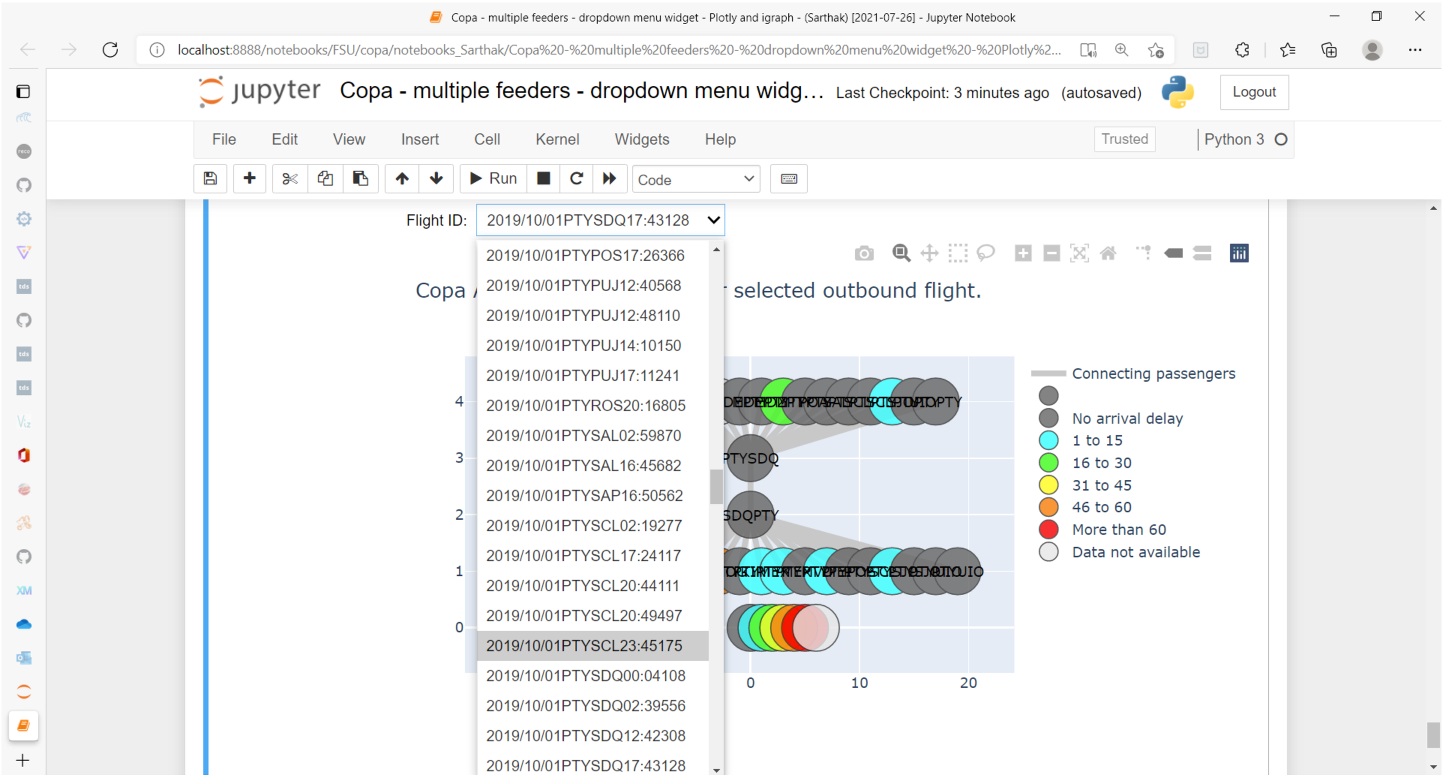


Figure 3.15: Screenshot of program in the Jupyter notebook. Dropdown menu widget (for selecting outbound flight of second level), as seen in Jupyter notebook.

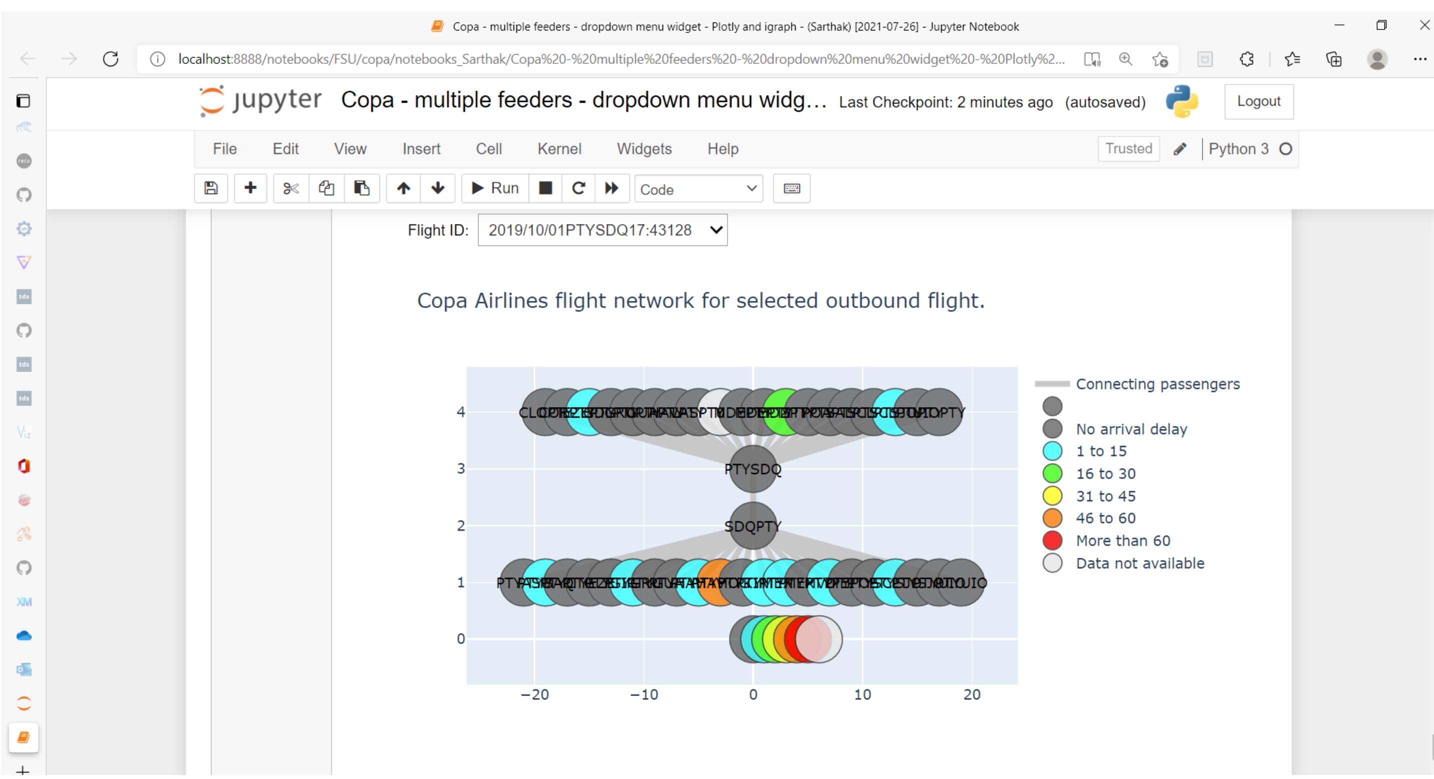


Figure 3.16: Flight network visualization, as seen in Jupyter notebook. Typical visualization output without the dropdown menu or tool-tip is shown here.

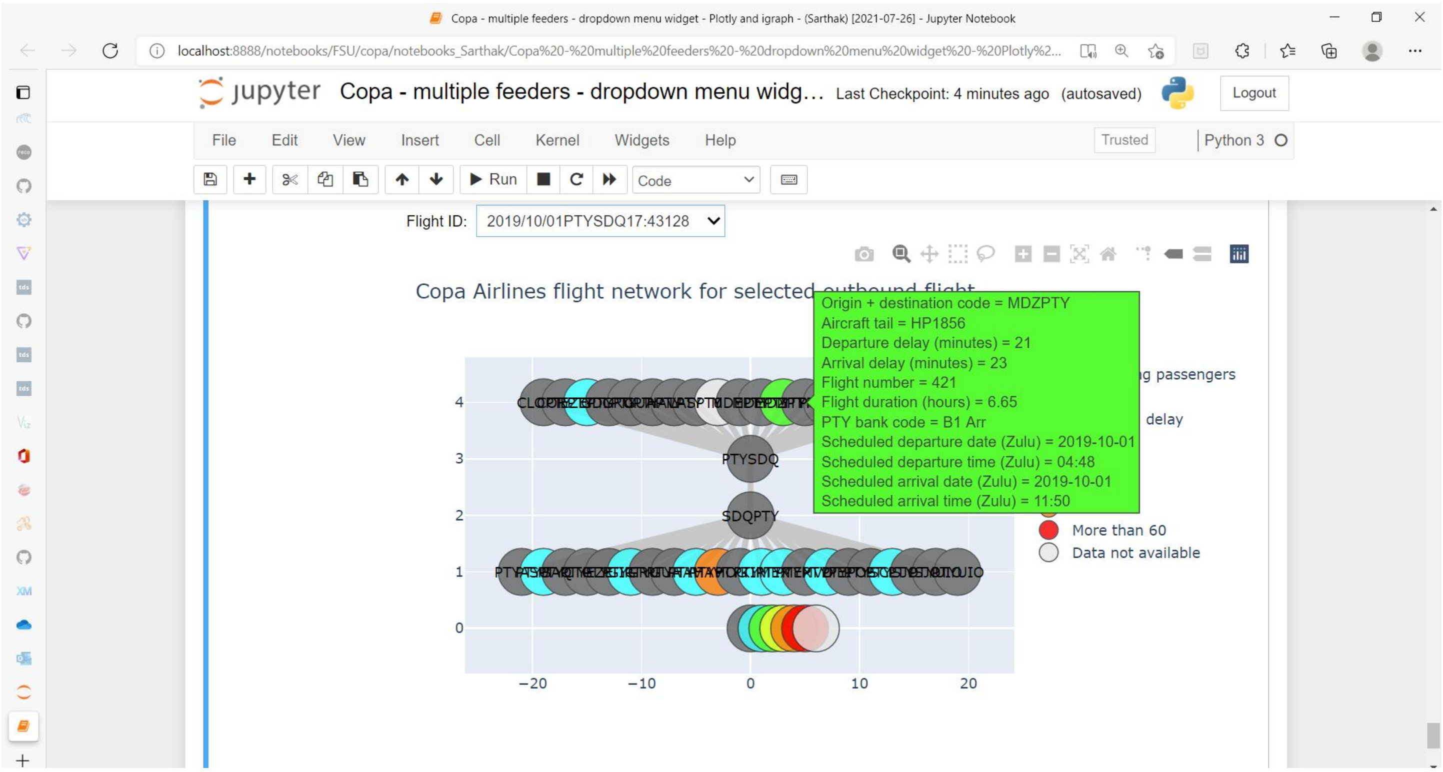


Figure 3.17: Hover information (tool-tips), as seen in Jupyter notebook. Tool-tips box is of same color as flight node, which depends on flight's category of arrival delay (can be changed to departure delay), as shown in legend.

4. Flight Delay Prediction Using Machine Learning

4.1. Introduction

Flight delay prediction is an important and popular research topic in data mining. Over the course of many years of research, sophisticated machine learning algorithms have successfully applied classification, regression, and neural network models to predict information critical to the economic growth of the airlines industry. Many airlines use the resources of artificial intelligence and machine learning to improve customer experience, reduce costs, and make proactive scheduling decisions in response to predicted delays. Such airlines include Delta, Southwest Airlines, and Air France. Copa Airlines, however, has not yet taken advantage of machine learning algorithms and the many applications they have in flight delay prediction. As of now, it performs all flight scheduling manually. Therefore, the goal of our research was to use machine learning to forecast flight delays, improve flight planning / scheduling, improve passenger experience, and increase profits for Copa Airlines.

Almost all Copa Airlines flights involve PTY airport as either origin or destination. Prediction of flights' arrival and departure delays may benefit Copa Airlines in many ways. Let there be a selection of flights inbound to PTY airport. At PTY airport, some passengers may transfer from these inbound flights to an outbound Copa Airlines flight. If one or more of these inbound flights has an arrival delay, then Copa Airlines staff decide (as of now, manually) whether to postpone the departure of that shared outbound flight. Conflicting interests must be weighed. Flight delays may propagate to connected flights and potentially affect a larger number of passengers later. In some cases, it may be better to delay the outbound flight so that connecting passengers of delayed feeder flights do not miss it. In other cases, it may be more profitable to Copa Airlines if the outbound flight departs PTY airport on time without the delayed passengers, and these delayed passengers get seats on the next available flight. When making this decision, the added cost for passenger overnight residence in hotels must also be considered. Flight delay forecasting may help to automate this decision in a data-driven way.

4.2. Data exploration

The dataset that we used is for Copa Airlines flights before the COVID-19 pandemic started. Figures 4.1 through 4.5 are plotted for initial data exploration. We want to see how the flights are distributed across the different delay categories in departure and arrival. Also, if there is any correlation between departure delays and arrival delays of different flights. Are there any trends worth noticing?

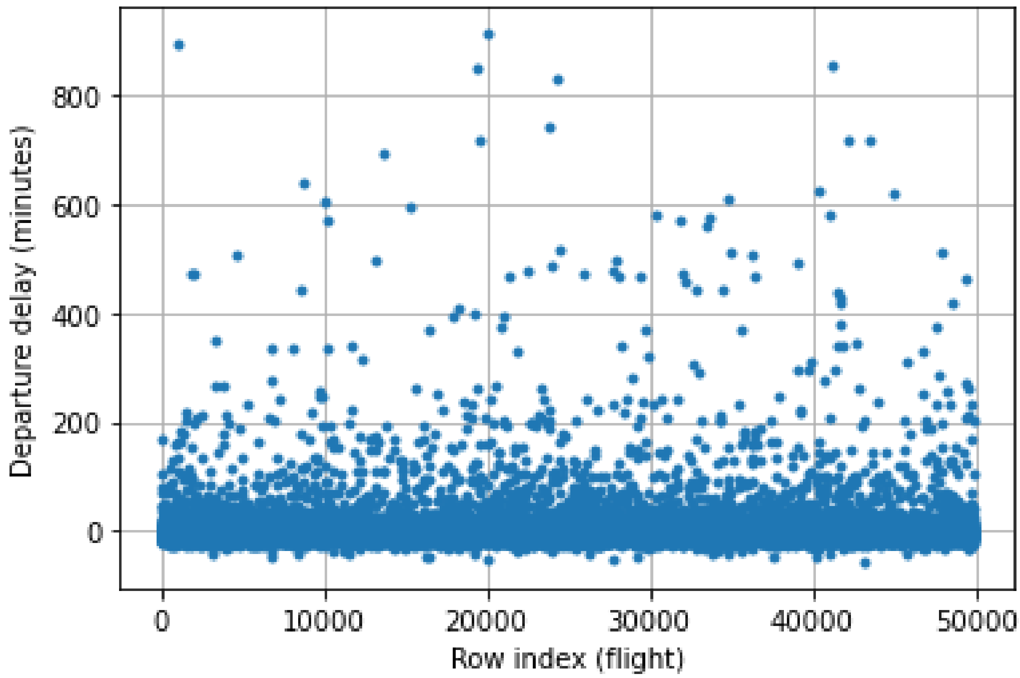


Figure 4.1: Scatter plot of departure delays of some Copa Airlines flights.

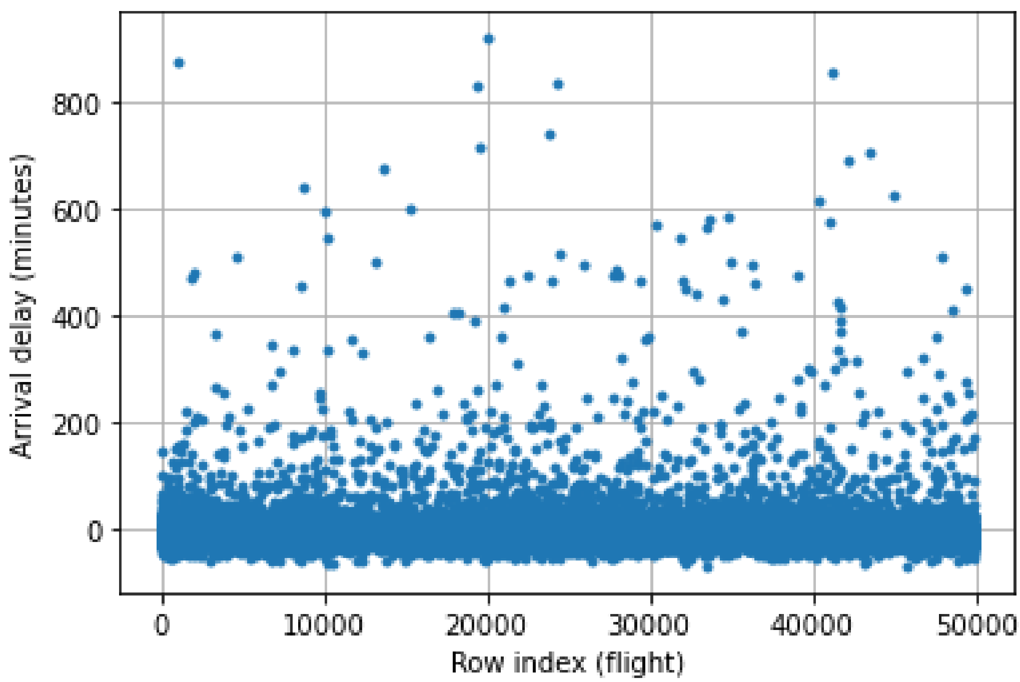


Figure 4.2: Scatter plot of arrival delays of some Copa Airlines flights.

As can be seen in the next scatter plot, departure delays and arrival delays are strongly correlated. Their correlation coefficient is found to be about 0.9234. This strong correlation is expected, as flights that depart late by *n* minutes are likely to arrive late by a similar number of minutes.

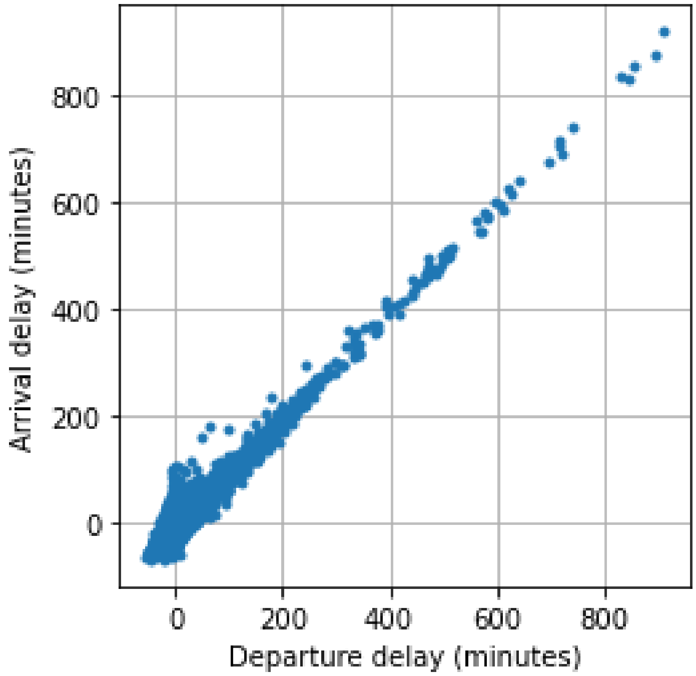


Figure 4.3: Scatter plot of showing some correlation between departure delays and arrival delays.

Next, we see the distribution of flights in different delay categories.

Delay category (minutes)	Fraction of flights in this departure delay category	Fraction of flights in this arrival delay category
≤ 0	83.6 %	75.2 %
1 to 15	11.5 %	18.5 %
16 to 30	2.3 %	3.7 %
31 to 45	0.8 %	0.9 %
46 to 60	0.3 %	0.4 %
≥ 61	1.3 %	1.3 %

We display bar charts and pie charts to show distribution of Copa Airlines flights in these delay categories for departure and arrival:

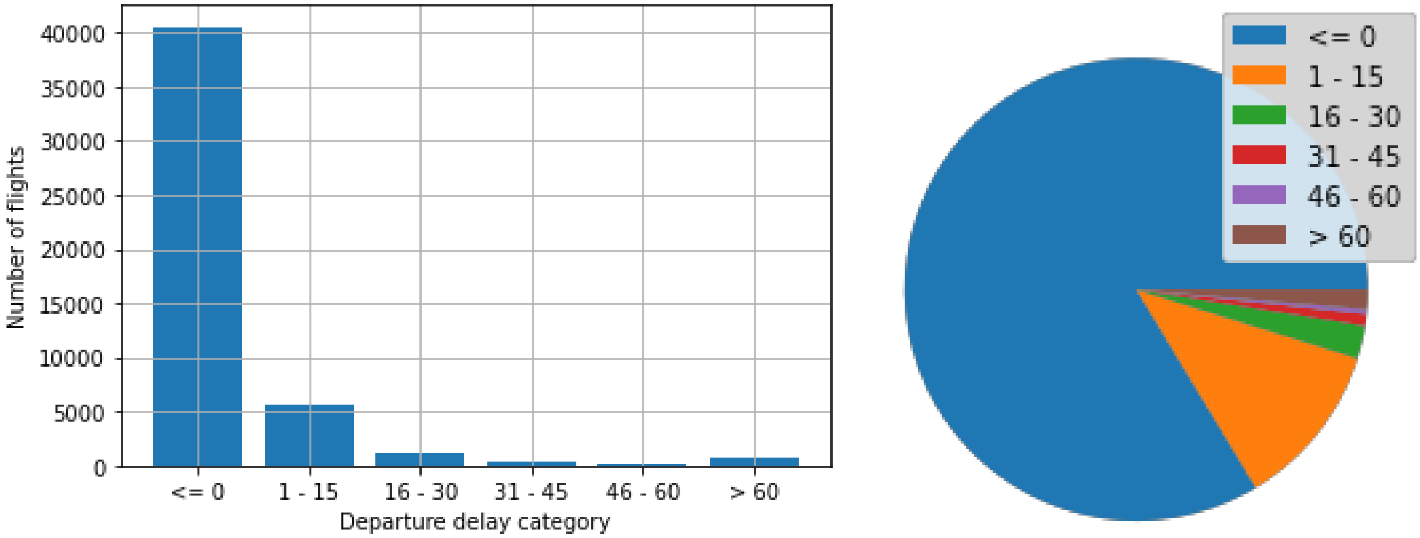


*Figure 4.4: Bar chart and pie chart showing distribution of flights in **departure** delay categories.*

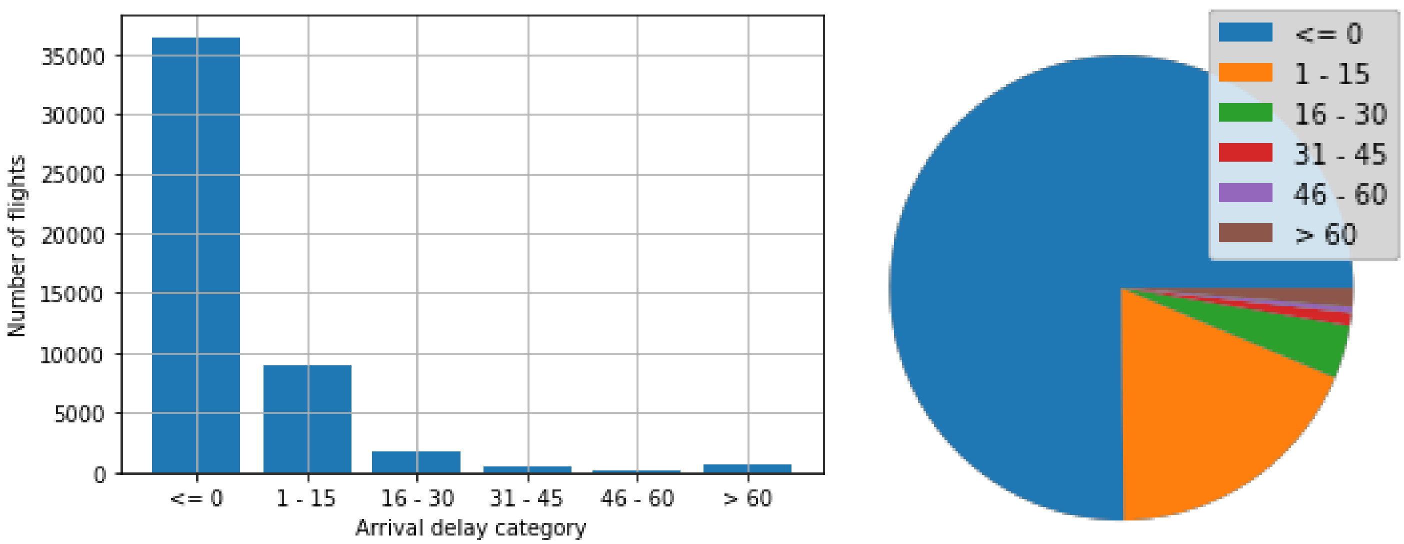


*Figure 4.5: Bar chart and pie chart showing distribution of flights in **arrival** delay categories.*

As can be seen, in both departure and arrival, most flights do not have delay. A few are in the 1 to 15 minutes delay category, and very few flights have delays of more than 15 minutes. Hence, it is a skewed data-set.

4.3. Methodology and implementation

In the simplest case of flight delay prediction for departure and arrival, we can calculate the average departure delay and average arrival delay for each flight route (origin + destination pair). This average departure or arrival delay may be stated as the predicted departure or arrival delay for that flight route. Though simple, such a prediction may not be very accurate because it considers only space (different flight routes), and not time. Hence, this simple model may not be very useful.

Flight delays may depend on various factors, including spatial consideration (flight route) and temporal considerations (time of the day, day of the week, season of the year, etc.). hence, the model accuracy may be improved by considering not only the flight route, but also the scheduled departure and arrival times, day of the week for departure and arrival, month or season of the year, etc. also, we may consider other factors like distance to be covered in the flight, weather patterns at that time, and even geo-political considerations that may affect flight paths and delays.

As a next step to improve delay prediction, we can try to use a simple artificial neural network like a multi-layer perceptron, that takes flight parameters as input, and tries to predict flight delay (either delay category in 15-minute intervals, which would be a classification problem, or the number of minutes delays, which would be a regression problem).

First, we try to use an artificial neural network to predict the categories of departure delays and arrival delays (classification problem). We use a neural network (multi-layer perceptron), with linear combinations of weights for artificial synapses, and ReLU (rectified linear unit) activation functions for artificial neurons.

The multi-layer perceptron is implemented using PyTorch [21], with three hidden layers, each having 100 artificial neurons. Synapses are modeled using linear combinations of weight functions, and activation functions use ReLU. The number of input points equals the number of features selected ($n_i = 28$), and the number of output points equals the number of delay categories ($n_o = 6$). The selected features include scheduled times and dates of departure and arrival, origin and destination airports, aircraft tail numbers, flight numbers, available and scheduled rotation times (duration between two consecutive flights using the same aircraft), etc. Loss function used is the cross-entropy loss function. The neural network is trained for ten epochs, in each type of analysis (predicting categories of departure delays and arrival delays). Training and testing are implemented as functions that are called in each epoch. A batch size of 100 flights is used.

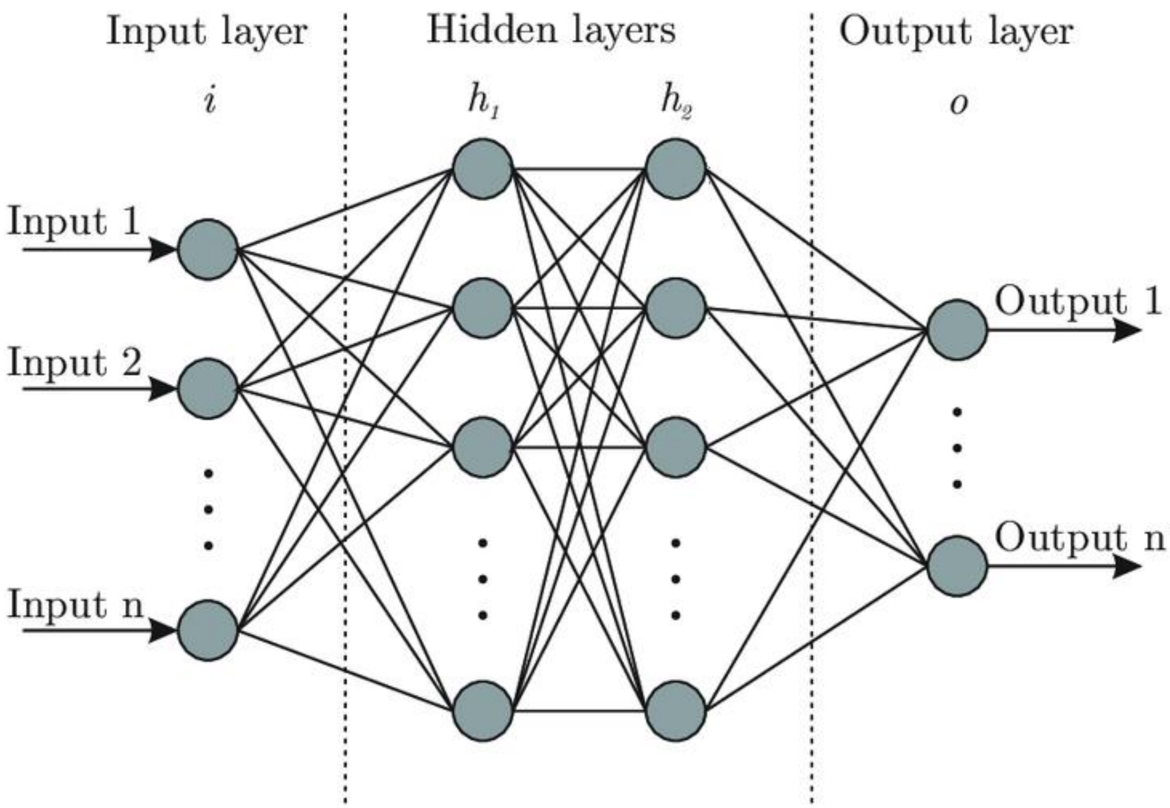


Figure: 4.6: Schematic of artificial neural network with an input layer, two hidden layers, and an output layer.

4.4. Results and discussion

As the data is imbalanced (most of the flights do not have delays in departure or arrival), the multi-layer perceptron was predicting "no delay" category for all flights. Though the accuracy may seem good initially (equal to the percentage of flights that do not have delay: about 84 % for departure, and 75 % for arrival), more improvement is needed, because a classification model that always predicts a single class is not much use, even if the predicted class is seen in majority of labels.

To improve, we are considering using Graph Neural Networks to predict departure and arrival flight delays (either predicting delay category, which will be a classification problem, or predicting the number of minutes delays, which will be a regression problem). This is work that is ongoing, and we hope to complete it soon.

5. Conclusions and Future Work

5.1. Conclusions

This has been a research project that successfully developed some computational tools to visualize part of Copa Airline's flight network. We are also attempting to use machine learning to predict flight delays in departure and arrival. Flight network visualization and flight delay prediction can be valuable tools to help in the flight network management and flight scheduling. The analysis may also be extended to other types of networks, for example, road traffic networks, data transfer networks, etc. Future work is described below.

5.2. Future work

Future work may involve using Artificial Intelligence to automate some or all aspects of the flight planning / scheduling process, to have better flight management and better overall passenger experience. For example, as of now, if an outbound Copa flight's inbound feeder flights have delayed arrivals at PTY airport, then Copa Airlines staff manually decide whether to postpone the departure of that outbound flight (to prevent the delayed inbound passengers from missing that outbound flight). If the decision is "yes, departure of that outbound flight should be postponed", then staff must also decide by how much time they should postpone that departure. Too much delay can propagate downstream in the flight network. If the decision is "no, departure of that outbound flight should not be postponed", then Copa officials may have to make alternative arrangements for the delayed passengers who will miss their next flight. (For example, putting these passengers on the next available flight with space to the same destination, and possibly providing these passengers accommodation in hotels near PTY airport if that flight will depart after a long time. If there is no space available on the next flight to the same destination, then the airline may have to see if space can be created by persuading some passengers of that flight to move to a later flight, in exchange for flight credit, partial refund, or some other gift.) The airline may have to balance conflicting interests like cost reduction, on-time performance, and customer satisfaction. It may help to automate these decisions in a data-driven and AI-enabled way.

For this analysis and design problem, it may be better to use data of "ordinary" times before the COVID19 pandemic, since data of times during the COVID19 pandemic (March 2020 to present) may not represent usual flight traffic. During the COVID19 pandemic, the global airline industry experienced significant negative circumstances (cancellations of most flights) due to lockdowns and travel restrictions enforced by governments world-wide, to stop the spread of the SARS-CoV-2 virus.

Appendices

A.1. Data description

The file "FSU_fully_cleaned.csv" has flight data. There are 49862 rows and 76 columns. Each row has data of a particular flight. It is data before the COVID-19 pandemic started. Below are explanations for some important columns in the file “FSU_fully_cleaned.csv”.

- **AIRLINE_DESIGNATOR_CD**: CM for Copa Airlines. CM* for Wingo / Copa Airlines Columbia (subsidiary of Copa Airlines).
- **ARR_DELAY_MINUTES**: Flight's arrival delay in minutes.
- **BANK OVERALL**: Shows the flight's "bank" at PTY airport. PTY airport has different "banks" where aircraft may be kept: B1, B2, B3, B4, B5, B6. Some aircraft may not be assigned a bank.
- **BANK_CD**: Shows the flight's bank, and whether the flight arrived at ("Arr") or departed from ("Dep") that bank.
- **CAPACITY_C_CNT**: Number of first class and business class seats available on that flight.
- **CAPACITY_Y_CNT**: Number of economy class seats available on that flight.
- **CAPACITY_CNT**: Total number of seats available on that flight (= CAPACITY_C_CNT + CAPACITY_Y_CNT).
- **COUNT_ARR00**: 1 if flight has arrival delay of 0 minutes or less. 0 otherwise.
- **COUNT_ARR14**: 1 if flight has arrival delay of 14 minutes or less. 0 otherwise.
- **COUNT_ARR15**: 1 if flight has arrival delay of 15 minutes or less. 0 otherwise.
- **COUNT_ARR90**: 1 if flight has arrival delay of 90 minutes or less. 0 otherwise.
- **COUNT_CANCELLATIONS**: 1 if flight was cancelled. 0 if flight was not cancelled.
- **COUNT_DEP00**: 1 if flight has departure delay of 0 minutes or less. 0 otherwise.
- **COUNT_DEP05**: 1 if flight has departure delay of 5 minutes or less. 0 otherwise.
- **COUNT_DEP15**: 1 if flight has departure delay of 15 minutes or less. 0 otherwise.
- **COUNT_DEP80**: 1 if flight has departure delay of 80 minutes or less. 0 otherwise.
- **CREW_CNT**: Number of crew members on that flight.
- **DEP_DELAY_MINUTES**: Flight's departure delay in minutes.
- **DEST_CD**: 3-letter code of the destination airport (e.g. “PTY” for Panama City airport, “MIA” for Miami International Airport, etc.). Airport’s full name can be searched on the internet.
- **FLT_ACTUAL_HR**: Flight duration in hours.
- **FLT_NUM**: Flight number.
- **IN_DTMZ**: “IN” date time Zulu (UTC: 0° longitude time and date when aircraft connects to arrival terminal) expressed as number of nano-seconds since some chosen reference time in the past (may be mid-night of January 1, 1970).

- **OD**: Flight's origin + destination code of 6 letters. First 3 letters show origin airport, and last 3 letters show destination airport. (E.g. "PTYMIA" for flight departing from Panama City airport and arriving at Miami airport.) Airport's full name can be searched on the internet.
- **OFF_DTMZ**: "OFF" date time Zulu (UTC: 0° longitude time and date when aircraft takes off from departure runway) expressed as number of nano-seconds since some chosen reference time in the past (may be mid-night of January 1, 1970).
- **ON_DTMZ**: "ON" date time Zulu (UTC: 0° longitude time and date when aircraft lands on arrival runway) expressed as number of nano-seconds since some chosen reference time in the past (may be mid-night of January 1, 1970).
- **ORIG_CD**: 3-letter code of the origin airport (e.g. "PTY" for Panama City airport, "MIA" for Miami International Airport, etc.). Airport's full name can be searched on the internet.
- **OUT_DTMZ**: "OUT" date time Zulu (UTC: 0° longitude time and date when aircraft disconnects from departure terminal) expressed as number of nano-seconds since some chosen reference time in the past (may be mid-night of January 1, 1970).
- **ROTATION_AVAILABLE_TM**: Available time for rotation of aircraft (i.e., available time between arrival of previous flight and departure of next flight using the same aircraft).
- **ROTATION_PLANNED_TM**: Planned time for rotation of aircraft (i.e., planned time between arrival of previous flight and departure of next flight using the same aircraft).
- **ROTATION_REAL_TM**: Actual time for rotation of aircraft (i.e., actual time between arrival of previous flight and departure of next flight using the same aircraft).
- **SCH_ARR_DTL**: Scheduled arrival date in local time of arrival airport.
- **SCH_ARR_DTML_PTY**: Scheduled arrival date and time in local time of PTY airport, expressed as number of nano-seconds from some chosen reference time in the past (may be mid-night of January 1, 1970).
- **SCH_ARR_DTMZ**: Scheduled arrival date and time in Zulu (UTC: 0° longitude time), expressed as number of nano-seconds from some chosen reference time in the past (may be mid-night of January 1, 1970).
- **SCH_ARR_DTZ**: Scheduled arrival date in Zulu (UTC: 0° longitude time).
- **SCH_ARR_TML_PTY**: Scheduled arrival time in local time of PTY airport (5 hours behind Zulu time).
- **SCH_ARR_TMZ**: Scheduled arrival time in Zulu (UTC: 0° longitude time).
- **SCH_DEP_DTL**: Scheduled departure date in local time of departure airport.
- **SCH_DEP_DTML_PTY**: Scheduled departure date and time in local time of PTY airport, expressed as number of nano-seconds from some chosen reference time in the past (may be mid-night of January 1, 1970).
- **SCH_DEP_DTMZ**: Scheduled departure date and time in Zulu (UTC: 0° longitude time), expressed as number of nano-seconds from some chosen reference time in the past (may be mid-night of January 1, 1970).
- **SCH_DEP_DTZ**: Scheduled departure date in Zulu (UTC: 0° longitude time).
- **SCH_DEP_TML_PTY**: Scheduled departure time in local time of PTY airport (5 hours behind Zulu time).
- **SCH_DEP_TMZ**: Scheduled arrival time in Zulu (UTC: 0° longitude time).
- **TAIL**: Tail number (identifier) of aircraft. In an airline, every aircraft has a different tail number.

- **TAXI_IN_MINUTES**: Taxi time in minutes, at arrival (i.e., time between aircraft landing on destination runway, and then connecting to destination terminal).
- **TAXI_OUT_MINUTES**: Taxi time in minutes, at departure (i.e., time between aircraft disconnecting from origin terminal, and then taking off from origin runway).
- **AC_OWNER_CARRIER_CD**: Code of airline that owns the aircraft used for that flight ("CM" for Copa Airlines).
- **AIRCRAFT_TYPE_ICAO**: Aircraft type (by International Civil Aviation Organization). E.g. "B738" for Boeing 738, "E190" for Embraer 190, etc.
- **IN**: Time of aircraft connecting to destination terminal.
- **ON**: Time of aircraft landing on destination runway.
- **OUT**: Time of aircraft disconnecting from origin terminal.
- **OFF**: Time of aircraft taking off from origin runway.

A.2. Delay analysis of flights

Table A-1: Average departure delays in descending order, for various origin-destination pairs.

Rank	Origin-Destination pair	Average departure delay (minutes)	Standard deviation of delays (minutes)	Number of flights
1	SJOBOG	14.878	18.119	41
2	VVCBOG	14.75	12.784	8
3	AUABOG	13.955	17.79	22
4	ADZCTG	13.09	16.777	67
5	CURBOG	12.895	15.684	38
6	CLOADZ	12.583	17.456	72
7	PTYBLB	11.625	5.499	8
8	ADZBOG	10.976	15.676	82
9	MEXBOG	10.847	15.931	59
10	ADZCLO	9.603	17.282	73
11	BOGSJO	8.674	13.767	43
12	MEXGUA	7	0	1
13	RECPTY	6.977	13.666	43
14	BAQADZ	6.938	11.917	81
15	BOGAUA	6.174	12.696	23
16	SJOSJO	6.167	21.98	6
17	BLBMDE	5.28	11.381	118
18	MDEBLB	5.017	13.155	118
19	YYZYUL	5	0	1
20	BOGCCS	4.763	13.839	80
21	BLBHAV	4.6	15.448	35
22	BLBBOG	4.544	13.485	149
23	BOGHAV	4.304	12.317	69
24	BGIPTY	4.256	13.361	43
25	PTYPTY	4.225	14.582	71
26	ADZBAQ	3.901	12.537	81
27	BOGCTG	3.321	10.989	215
28	CUNBOG	3.029	12.953	69
29	STIBOG	2.75	15.352	4
30	BOGGYE	2.422	10.72	45
31	BOGMEX	2.367	9.695	60
32	GYEBOG	2.178	8.239	45
33	YULPTY	2.122	8.567	82
34	BOGBLB	1.947	10.918	152

35	PTYPOA	1.884	9.404	146
36	CTGADZ	1.819	9.654	72
37	SXMPTY	1.655	9.206	58
38	BOGSDQ	1.508	9.266	61
39	YYZPTY	1.073	7.327	109
40	SDQBOG	0.852	13.618	61
41	BOGADZ	0.835	8.937	85
42	MSYPTY	0.831	14.725	77
43	PTYCOR	0.735	10.187	151
44	PTYROS	0.716	9.034	141
45	PTYDAV	0.686	7.946	287
46	PTYBSB	0.568	8.758	146
47	ORDPTY	0.53	12.118	268
48	PTYVVI	0.52	11.03	200
49	YULYUL	0.5	3.5	2
50	PTYTPA	0.481	8.805	108
51	PTYCCS	0.401	8.695	444
52	PTYPBM	0.4	10.74	45
53	PTYJFK	0.398	7.461	447
54	GIGPTY	0.385	10.812	278
55	PTYSSA	0.378	8.701	45
56	BLBCLO	0.341	10.239	82
57	CTGCLO	0.163	7.084	43
58	CNFPTY	0.118	9.054	127
59	PTYGRU	0.083	8.81	767
60	PTYFLL	0.024	8.278	126
61	TPATPA	0	0	1
62	PTYGEO	0	8.509	145
63	HAVBOG	-0.015	13.022	68
64	BOGPUJ	-0.156	10.155	64
65	SJOTGU	-0.168	10.229	149
66	CTGBOG	-0.186	10.101	215
67	PTYHOG	-0.2	8.56	65
68	PTYVLN	-0.238	7.795	151
69	PTYREC	-0.31	7.667	42
70	SFOPTY	-0.329	11.276	280
71	PTYMAR	-0.438	6.609	112
72	PTYCNF	-0.438	6.295	130
73	PTYMDZ	-0.477	7.304	86
74	PTYCIX	-0.477	7.503	44
75	IADPTY	-0.51	11.987	335
76	PTYMCO	-0.533	7.065	711

77	PTYIAD	-0.572	7.063	332
78	PTYBGI	-0.659	8.304	44
79	BOGUIO	-0.659	11.257	88
80	BOGCUR	-0.78	12.573	41
81	PTYSXM	-0.793	5.432	58
82	CTGBLB	-0.809	8.832	89
83	LASPTY	-0.898	12.665	127
84	BOGVVC	-0.9	3.113	10
85	PTYGYE	-0.911	9.152	704
86	PTYORD	-0.964	8.585	274
87	PTYYYZ	-0.991	7.587	108
88	PTYHAV	-0.991	8.656	985
89	SLASLA	-1	0	1
90	PTYNAS	-1.014	5.43	72
91	JFKPTY	-1.043	10.857	440
92	PTYGIG	-1.08	8.517	289
93	TGUSJO	-1.095	12.188	148
94	BLBCTG	-1.302	9.994	86
95	PTYMIA	-1.355	7.529	866
96	BOGCUN	-1.449	9.411	69
97	LAXPTY	-1.451	8.89	430
98	PTYLIM	-1.469	8.173	848
99	PTYMVD	-1.49	9.035	298
100	HAVBLB	-1.5	16.835	34
101	PTYDEN	-1.518	5.748	83
102	NASPTY	-1.521	14.461	73
103	PTYMGA	-1.522	6.65	414
104	PTYSCL	-1.552	9.133	717
105	PTYSFO	-1.716	6.585	285
106	PTYEZE	-1.724	8.495	449
107	MVDPTY	-1.724	11.644	294
108	PTYMSY	-1.812	8.819	80
109	PTYSAP	-1.836	6.365	140
110	CLOBLB	-1.854	10.45	82
111	PTYCTG	-1.883	6.592	453
112	BOSPTY	-1.887	8.056	160
113	MIAPTY	-1.968	8.265	866
114	BGABGA	-2	0	1
115	PTYTGU	-2.007	5.003	149
116	PTYLAX	-2.019	6.524	426
117	PTYSJO	-2.025	8.24	1324
118	PTYUIO	-2.061	7.82	624

119	PTYMEX	-2.189	5.769	755
120	PTYGUA	-2.202	7.113	598
121	PTYBOG	-2.248	7.013	1005
122	PTYSTI	-2.253	4.796	79
123	DAVPTY	-2.272	12.054	287
124	PTYAUA	-2.356	6.239	163
125	SSAPTY	-2.465	11.69	43
126	PTYPVR	-2.465	4.541	43
127	PTYYUL	-2.523	7.626	86
128	PTYPUJ	-2.653	7.269	685
129	PTYCUR	-2.661	6.011	127
130	PUJBOG	-2.719	11.954	64
131	PTYPAP	-2.839	4.571	87
132	UIOBOG	-2.872	10.435	86
133	PTYLAS	-2.929	5.43	127
134	PTYSDQ	-3.056	7.612	658
135	DENPTY	-3.073	10.897	82
136	TGUPTY	-3.08	12.609	150
137	TPAPTY	-3.094	11.967	106
138	PTYSNU	-3.131	4.815	84
139	PTYCLO	-3.133	8.083	565
140	PTYASU	-3.197	7.884	269
141	BOGPTY	-3.203	9.028	1011
142	PTYCUN	-3.232	6.686	1013
143	CIXPTY	-3.244	12.074	45
144	SCLPTY	-3.25	8.662	720
145	PTYBOS	-3.28	9.456	143
146	PTYPEI	-3.37	7.701	211
147	PTYADZ	-3.471	5.954	138
148	LIMPTY	-3.566	9.677	853
149	PEIPTY	-3.597	6.353	211
150	SLAPTY	-3.643	8.944	42
151	PTYKIN	-3.69	5.129	84
152	MAOPTY	-3.753	8.119	81
153	GRUPTY	-3.754	8.641	767
154	HOGPTY	-3.938	9.577	65
155	MEXPTY	-4.003	8.735	757
156	PTYSJU	-4.026	8.052	274
157	PTYMBJ	-4.137	4.506	131
158	PTYBZE	-4.163	3.839	43
159	EZEPTY	-4.197	9.901	446
160	PTYMDE	-4.254	7.152	856

161	PAPPTY	-4.267	6.967	86
162	MCOPTY	-4.287	8.008	712
163	PTYSAL	-4.295	5.57	285
164	POAPTY	-4.305	10.566	141
165	PTYGDL	-4.397	8.231	131
166	CLOCTG	-4.419	8.885	43
167	GUASJO	-4.693	9.197	300
168	PTYBAQ	-4.695	6.026	266
169	GUAMGA	-4.752	11.521	109
170	PBMPTY	-4.867	13.824	45
171	BSBPTY	-4.867	6.022	143
172	STIPTY	-4.911	5.781	79
173	AUAPTY	-4.933	10.653	163
174	PTYPOS	-4.956	6.744	248
175	PTYMTY	-5.165	5.264	85
176	BAQPTY	-5.19	7.458	269
177	SDQPTY	-5.422	7.413	650
178	MBJPTY	-5.431	6.85	130
179	MDEPTY	-5.457	7.582	865
180	PTYSLA	-5.465	4.299	43
181	GYEPTY	-5.537	7.944	708
182	BZEPTY	-5.548	8.68	42
183	FLLPTY	-5.552	10.191	125
184	VLNPTY	-5.673	11.635	153
185	SAPPTY	-5.676	10.413	139
186	PTYMAO	-5.706	8.718	85
187	SJOGUA	-5.754	9.929	297
188	SJUPTY	-5.904	7.08	272
189	SJOMGA	-5.908	7.216	152
190	CCSPTY	-5.909	13.059	449
191	SJOPTY	-5.912	9.657	1311
192	KINPTY	-5.964	5.29	83
193	CUNPTY	-6.176	7.025	1018
194	HAVPTY	-6.182	9.619	985
195	POSPTY	-6.274	8.781	248
196	CTGPTY	-6.398	8.537	450
197	PTYBGA	-6.602	5.876	83
198	GUAPTY	-6.605	9.823	597
199	MGASJO	-6.68	7.963	150
200	CURPTY	-6.734	7.351	128
201	SALPTY	-6.818	6.6	280
202	MARMAR	-7	0	1

203	GEOPTY	-7.062	13.056	144
204	CLOPTY	-7.11	6.402	566
205	BGAPTY	-7.296	4.093	81
206	MGAGUA	-7.459	11.444	109
207	UIOPTY	-7.875	6.986	624
208	VVIPTY	-7.966	8.409	237
209	MGAPTY	-7.993	9.264	415
210	MTYPTY	-8.186	6.148	86
211	PVRPTY	-8.256	6.864	43
212	GDLPTY	-8.405	9.394	131
213	BOGSTI	-8.75	5.019	4
214	LIRPTY	-8.786	6.057	42
215	PTYLIR	-9.047	4.539	43
216	ADZPTY	-9.25	14.972	140
217	MDZPTY	-9.345	11.879	87
218	MARPTY	-9.482	10.606	112
219	CORPTY	-9.547	8.034	148
220	CCSBOG	-9.78	15.997	82
221	ASUPTY	-9.875	10.007	273
222	PUJPTY	-10.365	8.849	688
223	BLBPTY	-12.286	31.631	7
224	ROSPTY	-13.716	8.987	141
225	SNUPTY	-15.565	10.2	85
226	BOGMDE	-18	0	1

Table A-2: Average arrival delays in descending order, for various origin-destination pairs.

Rank	OD pair	Average arrival delay (minutes)	Standard deviation of arrival delays (minutes)	Number of flights
1	YYZYUL	25	0	1
2	BOGSJO	19.071	18.437	42
3	AUABOG	17.478	21.978	23
4	BOGVVC	15.2	13.235	10
5	VVCBOG	14.625	14.628	8
6	SJOBOG	13	20.948	42
7	CLOADZ	12.871	17.906	70
8	BOGGYE	11.578	13.017	45
9	PTYBLB	10.625	9.937	8
10	ADZCLO	9.764	17.528	72
11	ADZBOG	9.222	18.634	81
12	CURBOG	9.2	20.026	40
13	ADZCTG	8.623	17.776	69
14	BOGADZ	8.229	12.105	83
15	MDEBLB	6.624	13.042	117
16	BOGAUA	6.522	15.871	23
17	BAQADZ	6.222	14.64	81
18	CTGADZ	6.127	16.985	71
19	MEXBOG	6.066	17.645	61
20	BLBMDE	6.008	13.331	119
21	BOGMEX	4.85	20.109	60
22	MAOPTY	4.765	11.513	81
23	ADZBAQ	4.16	13.53	81
24	YULPTY	3.362	21.437	80
25	BOGSTI	2.25	9.121	4
26	BGIPTY	1.818	11.794	44
27	BOGBLB	1.28	14.313	150
28	GYEBOG	0.891	12.898	46
29	PTYREC	0.69	13.428	42
30	PTYSSA	0.556	16.865	45
31	SJOTGU	0.43	11.539	149
32	BLBBOG	0.23	14.506	148
33	CLOBLB	0.085	13.266	82
34	CTGCLO	0	11.117	43
35	BOGCCS	-0.025	15.103	80
36	MSYPTY	-0.253	17.618	79
37	PTYFLL	-0.344	11.948	125

38	PTYMAR	-0.741	10.507	112
39	PTYDAV	-0.756	10.765	287
40	PTYPEI	-1.152	10.234	210
41	BOGCUN	-1.186	15.227	70
42	HAVBOG	-1.191	15.93	68
43	BLBCLO	-1.244	10.886	82
44	PTYLIM	-1.321	13.313	848
45	PTYGIG	-1.453	13.73	287
46	SXMPTY	-1.621	11.31	58
47	PTYSFO	-1.777	15.904	283
48	CUNBOG	-1.824	15.091	68
49	PTYCLO	-1.842	10.95	565
50	PTYPAP	-1.92	7.918	87
51	PEIPTY	-1.986	10.83	212
52	PTYMVD	-2.154	13.817	299
53	BOGCTG	-2.264	14.165	216
54	PBMPTY	-2.356	13.58	45
55	DAVPTY	-2.382	12.195	285
56	PTYPOA	-2.39	14.41	146
57	TGUSJO	-2.419	14.279	148
58	CURPTY	-2.606	11.048	127
59	PTYMAO	-2.607	13.939	84
60	PTYGYE	-2.618	11.235	705
61	BOGSDQ	-2.661	14.749	62
62	NASPTY	-2.694	15.79	72
63	PTYSAP	-2.757	10.739	140
64	RECPTY	-2.791	12.742	43
65	PTYVLN	-2.974	10.82	152
66	PTYCIX	-3	10.438	44
67	PTYNAS	-3.042	12.113	72
68	CTGBOG	-3.102	13.031	215
69	BOGHAV	-3.261	15.295	69
70	PTYBOG	-3.485	11.546	1001
71	UIOBOG	-3.488	13.616	86
72	SDQBOG	-3.508	17.783	63
73	GUASJO	-3.582	14.052	299
74	TGUPTY	-3.613	15.866	150
75	PTYCNF	-3.636	12.377	129
76	CTGBLB	-3.652	10.739	89
77	PTYBAQ	-3.733	8.122	266
78	PTYROS	-3.879	13.759	141
79	BLBCTG	-3.919	11.557	86

80	PTYUIO	-3.998	11.674	624
81	PTYBSB	-4.076	14.039	144
82	PTYCUR	-4.087	9.304	127
83	AUAPTY	-4.233	13.526	163
84	PTYSTI	-4.253	7.683	79
85	PTYBGI	-4.273	12.635	44
86	PTYCTG	-4.299	10.431	455
87	PTYCCS	-4.3	11.501	443
88	YYZPTY	-4.361	21.583	108
89	BLBPTY	-4.375	39.306	8
90	PTYGRU	-4.52	14.68	766
91	CLOCTG	-4.558	11.682	43
92	BOGPUJ	-4.609	13.585	64
93	PTYMEX	-4.677	13.744	756
94	PTYSAL	-4.839	9.014	285
95	PTYEZE	-4.846	13.751	449
96	PTYSJO	-4.848	11.839	1327
97	BOGPTY	-4.861	13.581	1007
98	PTYMGA	-4.918	10.463	413
99	IADPTY	-4.964	17.555	335
100	MDEPTY	-5.014	10.378	863
101	PAPPTY	-5.116	10.123	86
102	PTYMDE	-5.134	10.99	858
103	PTYAUA	-5.135	9.769	163
104	PTYKIN	-5.274	11.093	84
105	PTYSXM	-5.328	10.593	58
106	SSAPTY	-5.405	15.311	42
107	LAXPTY	-5.536	15.623	431
108	PTYHOG	-5.569	11.285	65
109	MBJPTY	-5.6	13.072	130
110	PTYPUJ	-5.61	11.873	685
111	PTYVVI	-5.616	13.896	198
112	SJOMGA	-5.638	9.271	152
113	PTYSCL	-5.665	14.667	713
114	SLAPTY	-5.81	14.71	42
115	PTYADZ	-5.87	10.319	138
116	PTYSDQ	-5.874	12.184	660
117	PTYHAV	-5.885	11.235	985
118	BZEPTY	-6	15.339	44
119	PTYGUA	-6.129	11.218	596
120	PTYTPA	-6.13	11.987	108
121	PTYSLA	-6.14	13.075	43

122	LASPTY	-6.23	14.184	126
123	PTYTGU	-6.282	9.545	149
124	PTYGEO	-6.315	13.854	146
125	BOGCUR	-6.35	14.547	40
126	SFOPTY	-6.396	16.046	280
127	BOGUIO	-6.425	14.023	87
128	PTYMDZ	-6.437	15.897	87
129	MVDPTY	-6.483	15.966	294
130	BLBHAV	-6.571	16.242	35
131	PTYMBJ	-6.87	9.846	131
132	SAPPTY	-6.906	13.238	138
133	HOGPTY	-6.908	12.759	65
134	PTYASU	-6.993	13.266	270
135	SCLPTY	-7.001	15.217	721
136	BAQPTY	-7.048	10.576	269
137	PTYCUN	-7.117	11.839	1014
138	ORDPTY	-7.181	19.907	265
139	PTYBZE	-7.205	13.194	44
140	CUNPTY	-7.229	11.733	1013
141	GIGPTY	-7.281	15.233	281
142	SJOGUA	-7.299	13.33	298
143	POSPTY	-7.369	14.721	249
144	PTYMIA	-7.404	12.074	868
145	PTYBGA	-7.405	10.722	84
146	GUAPTY	-7.534	12.711	595
147	BGAPTY	-7.634	10.001	82
148	CIXPTY	-7.644	14.484	45
149	CTGPTY	-7.762	11.595	449
150	PTYDEN	-7.78	18.076	82
151	MGASJO	-7.82	9.928	150
152	CNFPTY	-8.031	13.713	128
153	MGAGUA	-8.092	12.451	109
154	DENPTY	-8.329	20.531	82
155	PTYSNU	-8.429	7.344	84
156	MIAPTY	-8.456	13.32	867
157	SJUPTY	-8.614	11.254	272
158	CLOPTY	-8.657	9.734	566
159	SJOPTY	-8.663	12.526	1314
160	PTYORD	-8.81	15.675	274
161	PTYIAD	-8.843	13.337	331
162	PTYCOR	-8.854	16.327	151
163	CCSPTY	-8.895	14.629	447

164	VLNPTY	-8.941	14.25	153
165	FLLPTY	-9	10.061	124
166	PTYPBM	-9.089	16.241	45
167	PTYYUL	-9.163	15.122	86
168	SDQPTY	-9.209	13.288	654
169	GYEPTY	-9.293	11.239	707
170	PTYLAS	-9.354	15.923	127
171	ADZPTY	-9.525	15.694	139
172	PTYJFK	-9.534	15.619	446
173	VVIPTY	-9.58	13.677	238
174	TPAPTY	-9.72	13.838	107
175	MEXPTY	-9.982	13.722	757
176	PTYMCO	-10.041	11.933	712
177	GUAMGA	-10.064	13.231	109
178	PTYGDL	-10.221	12.724	131
179	LIMPTY	-10.256	13.404	853
180	HAVPTY	-10.3	13.651	986
181	MGAPTY	-10.344	11.705	413
182	GRUPTY	-10.405	14.861	768
183	POAPTY	-10.44	14.157	141
184	STIPTY	-10.481	9.997	79
185	KINPTY	-10.735	9.527	83
186	PTYMSY	-10.753	14.158	81
187	EZEPTY	-10.785	16.647	447
188	GDLPTY	-10.788	14.794	132
189	PTYYYZ	-10.981	14.024	108
190	UIOPTY	-11.047	9.816	623
191	PVRPTY	-11.095	11.174	42
192	PTYLIR	-11.333	10.239	42
193	MCOPTY	-11.339	12.655	713
194	PTYSJU	-11.493	13.838	274
195	MARPTY	-11.811	14.578	111
196	JFKPTY	-12.027	20.909	445
197	STIBOG	-12.25	13.442	4
198	BOSPTY	-12.528	22.354	159
199	SALPTY	-12.738	8.625	279
200	CCSBOG	-13.049	17.254	82
201	PTYPOS	-13.327	11.112	248
202	MTYPTY	-13.529	13.071	87
203	PTYLAX	-13.671	18.048	428
204	LIRPTY	-14.048	9.335	42
205	PTYPVR	-14.814	9.919	43

206	PUJPTY	-14.832	13.908	690
207	BSBPTY	-14.874	11.803	143
208	GEOPTY	-15.221	17.4	145
209	PTYMTY	-15.659	10.272	85
210	CORPTY	-15.859	15.779	149
211	PUJBOG	-16	16.248	64
212	ASUPTY	-16.062	14.597	273
213	MDZPTY	-16.276	19.157	87
214	BOGMDE	-19	0	1
215	SNUPTY	-20.714	13.891	84
216	PTYBOS	-21.965	18.134	144
217	HAVBLB	-22.235	18.871	34
218	ROSPTY	-23.207	16.029	140

References

[1] "Coordinated Universal Time," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Coordinated_Universal_Time.

[2] Caribbean Avenue, "Copa Airlines nearly quadruples in size since 2006.," 11 July 2016. [Online]. Available: https://www.caribbeanavenue.com/aviation/viewtopic.php?t=22450.

[3] "Copa Holdings," [Online]. Available: https://en.wikipedia.org/wiki/Copa_Holdings.

[4] "Copa Airlines Colombia," [Online]. Available: https://en.wikipedia.org/wiki/Copa_Airlines_Colombia.

[5] "Wingo (airline)," [Online]. Available: https://en.wikipedia.org/wiki/Wingo_(airline).

[6] S. Sharma, "Jupyter Notebook: Copa_Airlines_delay_analysis_(Python_programs)_Sarthak_Sharma.ipynb," [Online]. Available: https://github.com/engineer-scientist/Copa-Airlines-research-project/blob/main/delay_analysis/Copa%20Airlines%20-%20delay%20analysis%20(Python%20programs)%20%5BSarthak%20Sharma%5D.ipynb .

[7] S. Sharma, "Jupyter Notebook: Delay_probabilities_(Copa-FSU-02172021_data)_Sarthak_Sharma_[2021-04-05].ipynb," [Online]. Available: https://github.com/engineer-scientist/Copa-Airlines-research-project/blob/main/delay_analysis/Delay%20probabilities%20(Copa-FSU-02172021%20data)%20%5BSarthak%20Sharma%2C%202021-04-05%5D.ipynb.

[8] S. Sharma, "Jupyter Notebook: Visualization-Copa-FSU-02172021_data_Sarthak_[2021-05-14].ipynb," [Online]. Available: https://github.com/engineer-scientist/Copa-Airlines-research-project/blob/main/visualization/Visualization%20-%20Copa-FSU-02172021%20data%20%5BSarthak%2C%202021-05-14%5D.ipynb.

[9] Dash Enterprise, "Plotly," Plotly, [Online]. Available: https://plotly.com/. [Accessed March 2021].

[10] Altair Engineering, "Altair," Altair Engineering, [Online]. Available: https://www.altair.com/. [Accessed March 2021].

[11] Streamlit, "Streamlit," Streamlit, [Online]. Available: https://streamlit.io/. [Accessed March 2021].

[12] "React: a JavaScript library for building user interfaces," React, [Online]. Available: https://reactjs.org/. [Accessed March 2021].

[13] Vue, "Vue: The Progressive JavaScript Framework," [Online]. Available: https://vuejs.org/. [Accessed March 2021].

[14] NetworkX, "NetworkX: Network Analysis in Python," NetworkX, [Online]. Available: https://networkx.org/. [Accessed March 2021].

[15] PyData, "Pandas," NumFOCUS, [Online]. Available: https://pandas.pydata.org/. [Accessed February 2021].

[16] NumPy, "NumPy: The fundamental package for scientific computing with Python," NumPy, [Online]. Available: https://numpy.org/. [Accessed January 2021].

[17] S. Sharma, "Jupyter Notebook: Copa_flight_network_visualization_using_Plotly_and_Networkx_(Sarthak)_[2021-06-05].ipynb," [Online]. Available: https://github.com/engineer-scientist/Copa-Airlines-research-project/blob/main/visualization/Copa%20flight%20network%202-D%20visualization%20using%20Plotly%20and%20Networkx%20(Sarthak)%20%5B2021-06-05%5D.ipynb.

[18] S. Sharma, "Jupyter Notebook: Copa_flights_tree-plot_using_Plotly_(Sarthak)_[2021-06-24].ipynb," [Online]. Available: https://github.com/engineer-scientist/Copa-Airlines-research-project/blob/main/visualization/Copa%20flights%20tree-plot%20using%20Plotly%20(Sarthak)%20%5B2021-06-24%5D.ipynb.

[19] S. Sharma, "Jupyter Notebook: Copa_flights_network-multiple_feeders-Plotly_and_igraph_(Sarthak)_[2021-07-09].ipynb," [Online]. Available: https://github.com/engineer-scientist/Copa-Airlines-research-project/blob/main/visualization/Copa%20flights%20network%20-%20multiple%20feeders%20-%20Plotly%20and%20igraph%20(Sarthak)%20%5B2021-07-09%5D.ipynb.

[20] S. Sharma, "Jupyter Notebook: Copa-multiple_feeders-dropdown_menu_widget-Plotly_and_igraph-(Sarthak)_[2021-07-26].ipynb," [Online]. Available: https://github.com/engineer-scientist/Copa-Airlines-research-project/blob/main/visualization/Copa%20-%20multiple%20feeders%20-%20dropdown%20menu%20widget%20-%20Plotly%20and%20igraph%20-%20(Sarthak)%20%5B2021-07-26%5D%20.ipynb.

[21] "PyTorch," PyTorch, [Online]. Available: https://pytorch.org/.

[22] S. Sharma, "Porgrams: Copa Airlines research project," GitHub, [Online]. Available: https://github.com/engineer-scientist/Copa-Airlines-research-project.