Logo, company name

Description automatically generated

Rutuja Pohare

Data visualization for Flight Delay  
and cancellation Analysis for

USA domestic airlines for Year 2015

**Course – Data Visualization DATA230**

**Date – December 10th, 2022**

**Professor - Andrew Bond**

Table of content:

Section A: Introduction

ABSTRACT

1. Purpose of this document
2. Intended Audience
3. Scope
4. Problem statement
5. Design Approach

Section B: Data source and Usage

1. Data source used for this project
2. Data analysis tool (TABLEAU)
3. Datasets
4. Joins and Relationships between datasets

Section C: Data Wrangling

1. Mathematical calculations to plot graphs
2. Sigmoid function

Section D: Abbreviations

1. Airports
2. Airlines

Section E: Insights

1. Arrival, Departure delay, year2015.
2. Top (more delay, n) and bottom (less delay, n).
3. Origin to Destination with respect to average arrival delay.
4. Origin to Destination with respect to average departure delay
5. Weekly Airline performance
6. Reasons for Airline delay
7. Monthly airlines cancellation for year 2015

Section F: Use cases

Section G: Summary

Section H: Reference

Section A: Introduction

**Abstract:**

Flight delays and cancellation are inevitable and they often lead to concern in passengers as well as profit loss of the airlines and airports. An accurate estimation of flight delays and cancellation is critical for airlines because the results can be applied to increase customer satisfaction and income of airline agencies. This paper proposed several drawn insights on a dataset containing flight information from 2015

## 

## Purpose of this document

## Flight delays are one of the most common but unpleasant experiences that people dread having. Every year, a lot of flights get delayed which involves some cost both for the airline and the passenger in different ways. The passenger’s time and money get affected and at the same time, the airline’s reputation is at stake. Delay is treated as one of the most remembered performance indicators of the airline. There could be some reasons which are inevitable such as weather conditions, air trafficking or any unforeseen event; but there also could be some reasons which can be dealt with by improving the process. Therefore, statistics of the flight delays become a crucial factor in understanding the flight’s performance.

## This study presents the analysis driven from flight delay and cancellation data for the United States domestic airlines for the year 2015. This study analyses the variety of factors responsible for and associated with flight delays and cancellation for different airlines.

## Intended Audience

The target audience for this study are people such as passengers, airport officials, crew members, etc. who are directly related to the airline industry and the people such as data analysts who are interested in the analysis of flight datasets.

## 

## Scope

The goal of our visualization is to have an interactive and intuitive

Dashboard showing delays and cancellation for the USA domestic

airlines for the year 2015. Travel is an essential part of business life.

an easily interpretable visualization of delay and cancellation can be

useful to understand and plan travel. Our Dashboard shows delays

and cancellation by the airport origin, destination, carriers, month

and day of the week. The visuals are linked and can be filtered and

highlighted to cater to different viewers

**Problem Statement.**

Thousands of Americans experience delayed or cancelled flights each year as airlines and airports in the US struggle with surging demand, seasonal storms, smoke from wildfires, worker shortages and mechanical failures among other causes of flight delays and cancellations. According to U.S. Department of Transportation, on average, airlines delayed 27 percent of flights in June 2015 and cancelled 2 percent. These numbers have significantly increased exponentially.

# ****Design Approach****

Once the vision of our study was clear, we performed a literature review by reading different research papers and articles to understand how the airplane dataset has been already used to create meaningful visualizations. After performing the literature reviews and carefully studying the available dataset, we made some sketches which could be turned into visualizations by using visualization tools. Different types of sketches were created to observe the trends about the delays based on delay types and airplane carriers. Having the sketches ready, we implemented our designs on the visualization tool called Tableau wherein we focused on creating visualizations that were simple and easily readable.

**Section B: Data source and its usage.**

**Data source used for this project:**

* 2015 flights data with delay and cancellation
* Domestic flights in USA
* Downloaded from Kaggle
* <URL:><https://www.kaggle.com/datasets/usdot/flight-delays>

This data deals with the data on the flights taken from the U.S Department of Transportation’s (DOT Bureau of transportation statistics. It tracks the on-time performance of domestic flights operated by different air carriers over different airports.

**Data Analysis Tool:**

* TABLEAU

**Dataset used for this project:**

Csv Files considered for this project:  
  
1). Flights.csv – (31 columns)  
2). Airport.csv – (4 columns)  
3). Airplane.csv – (2 columns)  
  
These three csv filles are connected to each other by 3 different means as follows. And different operations are performed on it separately to analyze the data source precisely:

In a relational database system like Access, you often need to extract information from more than one table at a time. This can be accomplished by using an [**JOIN**](https://learn.microsoft.com/en-us/office/vba/access/concepts/structured-query-language/join-microsoft-access-sql-reserved-word) statement, which enables you to retrieve records from tables that have defined relationships, whether they are one-to-one, one-to-many, or many-to-many.

1). Inner joins:

Diagram

Description automatically generated

The [**INNER JOIN**](https://learn.microsoft.com/en-us/office/vba/api/overview/access), also known as an equi-join, is the most commonly used type of join. This join is used to retrieve rows from two or more tables by matching a field value that is common between the tables. The fields you join on must have similar data types.

The INNER JOIN creates a new result table by combining column values of two tables (table1 and table2) based upon the join-predicate. The query compares each row of table1 with each row of table2 to find all pairs of rows which satisfy the join-predicate.

When the join-predicate is satisfied, column values for each matched pair of rows of A and B are combined into a result row.

**Representation of inner join**

It is used in this project to help to connect the airport csv file to flight csv file in order to analyses the data precisely and extract some useful information.

For my project I have connected the flight csv dataset with the airport csv dataset. For the analysis it was required that the flight table should be first connected in the form of union join to the flight csv with the column name table name, this column has two domains ‘flights.csv’ and ‘flight.csv1’. Using the table name column, I have created a route identifier column known as origin\_route\_destination, in order to connected the origin and destination airport name and to plot the path between origin and the destination. Again, by using the Table column I have created another column known as Route\_order. As it is connected to the table column then domain flight.csv is route 1 and flight.csv1 is route 2. Same goes for the Route\_location where flight.csv is for the origin airports and flight.csv1 is for the destination airport.

Graphical user interface, text, application

Description automatically generated

Now in order to connect the flight dataset to the airport dataset we have to use the column which have created on the flight.csv named as Route location with the airport.csv dataset column using the IATA code using the INNER JOIN function. As you can see in the below screenshot from the TABLEAU workbook software.

Graphical user interface, application, Word

Description automatically generated

Final screenshot after the connection has been made successfully.

A screenshot of a computer

Description automatically generated

2). union joins:

Chart

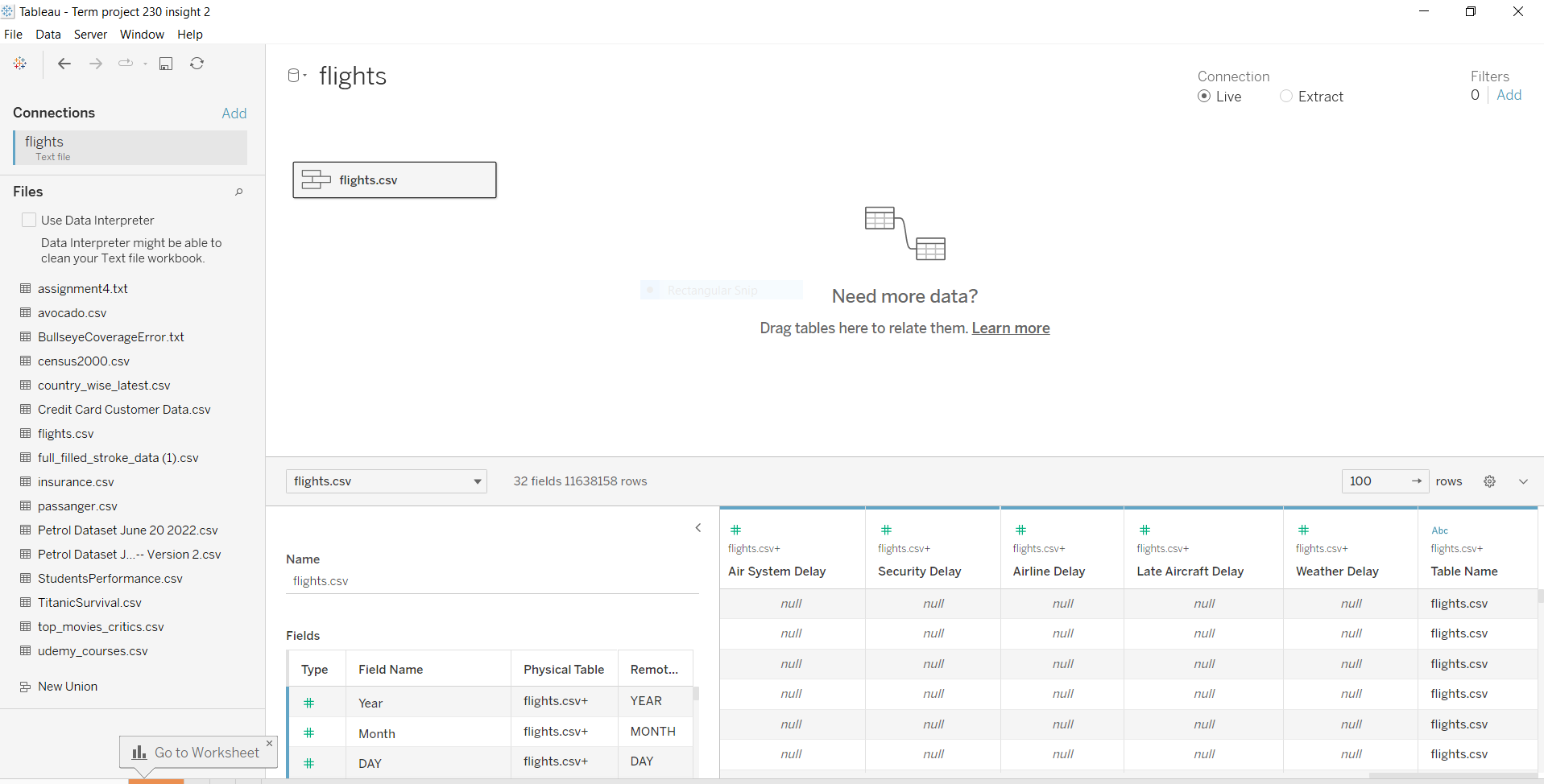
Description automatically generated

UNION combines the results of two or more queries into a single result set that includes all the rows that belong to all queries in the union. By using JOINs, you can retrieve data from two or more tables based on logical relationships between the tables. Joins indicate how SQL should use data from one table to select the rows in another table.

**Representation of union join**

It is used in this project to help to connect the flights.csv file to flight.csv file in order to analyses the data precisely and extract some useful information.

Here I have stacked the flights.csv file on top of the flight.csv file in order to create a union function among them. So, we get a new column named as Table Name. We are using the union function on the same flight.csv file because we want to extract the flights.csv data a route1 and flights.csv1 data as route2.



3). Many to many relationships:

Diagram

Description automatically generated

A many-to-many database relationship is a relationship between two database tables where a record in each table can reference several records in the other table. For example, in a blog, a table for posts can have a many-to-many relationship with a table for storing authors. Each post can have many authors, and each author can write many posts. Therefore, there is a many-to-many relationship between posts and authors. For another example, in a social media application, each post may have many hashtags, and each hashtag may have many posts.

**Representation of many to many relationship**

It used in this project to help to connect the airport csv file to flight csv file in order to analyses the data precisely and extract some useful information.

We have created many-to-many relationships between flight.csv, airport.csv and airplane.csv.

So, that we can extract the connected data between these three files simultaneously. In Addition, we can combine this information and can draw a meaningful insight which can help in proper decision making.

Graphical user interface

Description automatically generated

**Section C: Data Wrangling.**

**Data wrangling:**

Data wrangling is a term often used to describe the early stages of the data analytics process. It involves transforming and mapping data from one format into another. The aim is to make data more accessible for things like business analytics or machine learning. The data wrangling process can involve a variety of tasks. These include things like data collection, exploratory analysis, data cleansing, creating data structures, and storage.

Diagram

Description automatically generated

Data wrangling is time-consuming. In fact, it can take up to about 80% of a data analyst’s time. This is partly because the process is fluid, i.e., there aren’t always clear steps to follow from start to finish. However, it’s also because the process is iterative, and the activities involved are labor-intensive. What you need to do depends on things like the source (or sources) of the data, their quality, your organization’s data architecture, and what you intend to do with the data once you’ve finished wrangling it.

For this project a certain number of calculated fields have been created for the deep understanding of the respective datasets. As discussed above, a few calculated fields have been created in order to create joins between two different datasets. Calculated fields come under the criteria of data wrangling. Pictures of the created columns are as follows.

A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated with medium confidence

Furthermore, more data wrangling has been performed in order to create a SANKEY CHART which will represent the relationship between the months and the airline. The connection between these two fields is calculated by the number of cancelled flights.

**SANKEY CHART**.

Diagram

Description automatically generated

The Sankey chart in the tableau is a great diagram. This diagram will show the flow and relationship between two different elements. These two different elements are called nodes and relationship or connection between two different elements called links. When we have two different dimensions (it is qualitative data or description of data) and want to know the relationship between them or mapping then the Sankey chart is the best way to use. Three major points in the Sankey chart are Target, Source, and Size. These three parameters will generate the Sankey chart.

So, in order to create a SANKEY CHART, first we need to understand the **sigmoid function** which the key function is used to build the SANKEY CHART.

**Sigmoid function:**

A Sigmoid function is a mathematical function which has a characteristic S-shaped curve. There are a number of common sigmoid functions, such as the **logistic function**, the **hyperbolic tangent**, and the **arctangent.**

In machine learning, the term sigmoid function is normally used to refer specifically to the logistic function, also called the logistic sigmoid function.

All sigmoid functions have the property that they map the entire number line into a small range such as between 0 and 1, or -1 and 1, so one use of a sigmoid function is to convert a real value into one that can be interpreted as a probability. *One of the most widely used sigmoid functions is the logistic function, which maps any real value to the range (0, 1). Note the characteristic S-shape which gave sigmoid functions their name (from the Greek letter sigma).*

Sigmoid functions have become popular in [deep learning](https://deepai.org/machine-learning-glossary-and-terms/deep-learning) because they can be used as an [activation function](https://deepai.org/machine-learning-glossary-and-terms/activation-function) in an artificial [neural network](https://deepai.org/machine-learning-glossary-and-terms/neural-network). They were inspired by the activation potential in biological neural networks.

Sigmoid functions are also useful for many machine learning applications where a real number needs to be converted to a probability. A sigmoid function placed as the last layer of a machine learning model can serve to convert the model's output into a probability score, which can be easier to work with and interpret.

Sigmoid functions are an important part of a [logistic regression](https://deepai.org/machine-learning-glossary-and-terms/logistic-regression) model. Logistic regression is a modification of [linear regression](https://deepai.org/machine-learning-glossary-and-terms/linear-regression) for two-class classification, and converts one or more real-valued inputs into a probability, such as the probability that a customer will purchase a product. The final stage of a logistic regression model is often set to the logistic function, which allows the model to output a probability.

Chart, line chart

Description automatically generated

Calculating the Sigmoid Function

Example Calculation of Logistic Sigmoid Function

Taking the logistic sigmoid function, we can evaluate the value of the function at several key points to understand the function's form.

At x = 0, the logistic sigmoid function evaluates to:

Diagram, schematic

Description automatically generated

This is useful for the interpretation of the sigmoid as a probability in a logistic regression model, because it shows that a zero input results in an output of 0.5, indicating equal probabilities of both classes.

**Properties and Identities of Sigmoid Function**

The graph of sigmoid function is an S-shaped curve as shown by the green line in the graph below. The figure also shows the graph of the derivative in pink color. The expression for the derivative, along with some important properties are shown on the right.

Chart

Description automatically generated

## Sigmoid As An Activation Function In Neural Networks.

The sigmoid function is used as an activation function in neural networks. Just to review what is an activation function, the figure below shows the role of an activation function in one layer of a neural network. A weighted sum of inputs is passed through an activation function and this output serves as an input to the next layer.

Diagram

Description automatically generated

After understanding the basic concept of sigmoid function.

We can create the Sankey chart at TABLEAU.

* Creating the TOPPAD field

Transfer the data between the interval of 0 and 1.

First we have to create a TOPPAD in order to divide the table structure into 1 by 49 parts. Which will form two poles between the connection. As you can see below:

Graphical user interface, text, application, email

Description automatically generated

* Creating the padded field in order to create the bins.

Graphical user interface

Description automatically generated

**T field to create what we will consider on the x axis**

- Make the second calculated field as given below image.

- These two calculations are telling the tableau were to put a line. - These calculations are graphically meaningful. The Region on left side and category on the right side, draw the connection between these two categorial.1 and 49 is for drawing a curve.

Graphical user interface, application

Description automatically generated

* Secondly, we have used the calculated padded field to the create the 49 distinct points which can be done by using the t field which we have create.

Graphical user interface, text, application, email

Description automatically generated

Create Rank1 and Rank2 calculated field for the curve.

We have calculated the rank 1 and rank 2 by using the sum of diverted flights. As it can be seen from the below flight.

R2 is created the same way as Rank 1

Graphical user interface, text, application, email

Description automatically generated

This is how we created a field named sigmoid in tableau with the use of mathematical formula as discussed above.

Graphical user interface, text, application, email

Description automatically generated

CREATING the CURVE:

Graphical user interface, text, application, email

Description automatically generated

Following are the steps taken to create the Sankey chart for flight cancelled analysis:

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

A picture containing diagram

Description automatically generated

Graphical user interface

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

This is all data wrangling has been done to create the Sankey chart.

The detailed description for and analysis from this chart will be discussed further.

**Section D: Abbreviations.**

Before going further some important abbreviations need to be considered and understood.

1). Airports:

|  |  |  |
| --- | --- | --- |
| IATA\_CODE | AIRPORT | CITY |
| ABE | Lehigh Valley International Airport | Allentown |
| ABI | Abilene Regional Airport | Abilene |
| ABQ | Albuquerque International Sun port | Albuquerque |
| ABR | Aberdeen Regional Airport | Aberdeen |
| ABY | Southwest Georgia Regional Airport | Albany |
| ACK | Nantucket Memorial Airport | Nantucket |
| ACT | Waco Regional Airport | Waco |
| ACV | Arcata Airport | Arcata/Eureka |
| ACY | Atlantic City International Airport | Atlantic City |
| ADK | Adak Airport | Adak |
| ADQ | Kodiak Airport | Kodiak |
| AEX | Alexandria International Airport | Alexandria |
| AGS | Augusta Regional AirportÂ (Bush Field) | Augusta |
| AKN | King Salmon Airport | King Salmon |
| ALB | Albany International Airport | Albany |
| ALO | Waterloo Regional Airport | Waterloo |
| AMA | Rick Husband Amarillo International Airport | Amarillo |
| ANC | Ted Stevens Anchorage International Airport | Anchorage |
| APN | Alpena County Regional Airport | Alpena |
| ASE | Aspen-Pitkin County Airport | Aspen |
| ATL | Hartsfield-Jackson Atlanta International Airport | Atlanta |
| ATW | Appleton International Airport | Appleton |
| AUS | Austin-Bergstrom International Airport | Austin |
| AVL | Asheville Regional Airport | Asheville |
| AVP | Wilkes-Barre/Scranton International Airport | Wilkes-Barre/Scranton |
| AZO | Kalamazoo/Battle Creek International Airport | Kalamazoo |
| BDL | Bradley International Airport | Windsor Locks |
| BET | Bethel Airport | Bethel |
| BFL | Meadows Field | Bakersfield |
| BGM | Greater Binghamton Airport | Binghamton |
| BGR | Bangor International Airport | Bangor |
| BHM | Birmingham-Shuttlesworth International Airport | Birmingham |
| BIL | Billings Logan International Airport | Billings |
| BIS | Bismarck Municipal Airport | Bismarck |
| BJI | Bemidji Regional Airport | Bemidji |
| BLI | Bellingham International Airport | Bellingham |
| BMI | Central Illinois Regional Airport at Bloomington-Normal | Bloomington |
| BNA | Nashville International Airport | Nashville |
| BOI | Boise AirportÂ (Boise Air Terminal) | Boise |
| BOS | Gen. Edward Lawrence Logan International Airport | Boston |
| BPT | Jack Brooks Regional AirportÂ (Southeast Texas Regional) | Beaumont/Port Arthur |
| BQK | Brunswick Golden Isles Airport | Brunswick |
| BQN | Rafael HernÃ¡ndez Airport | Aguadilla |
| BRD | Brainerd Lakes Regional Airport | Brainerd |
| BRO | Brownsville/South Padre Island International Airport | Brownsville |
| BRW | Wiley Post-Will Rogers Memorial Airport | Barrow |
| BTM | Bert Mooney Airport | Butte |
| BTR | Baton Rouge Metropolitan Airport | Baton Rouge |
| BTV | Burlington International Airport | Burlington |
| BUF | Buffalo Niagara International Airport | Buffalo |
| BUR | Bob Hope AirportÂ (Hollywood Burbank Airport) | Burbank |
| BWI | Baltimore-Washington International Airport | Baltimore |
| BZN | Bozeman Yellowstone International AirportÂ (Gallatin Field Airport) | Bozeman |
| CAE | Columbia Metropolitan Airport | Columbia |
| CAK | Akron-Canton Regional Airport | Akron |
| CDC | Cedar City Regional Airport | Cedar City |
| CDV | Merle K. (Mudhole) Smith Airport | Cordova |
| CEC | Del Norte County AirportÂ (Jack McNamara Field) | Crescent City |
| CHA | Chattanooga Metropolitan AirportÂ (Lovell Field) | Chattanooga |
| CHO | Charlottesville-Albemarle Airport | Charlottesville |
| CHS | Charleston International Airport/Charleston AFB | Charleston |
| CID | The Eastern Iowa Airport | Cedar Rapids |
| CIU | Chippewa County International Airport | Sault Ste. Marie |
| CLD | McClellan-Palomar Airport | San Diego |
| CLE | Cleveland Hopkins International Airport | Cleveland |
| CLL | Easterwood Airport | College Station |
| CLT | Charlotte Douglas International Airport | Charlotte |
| CMH | Port Columbus International Airport | Columbus |
| CMI | University of Illinois - Willard Airport | Champaign/Urbana |
| CMX | Houghton County Memorial Airport | Hancock |
| CNY | Canyonlands Field | Moab |
| COD | Yellowstone Regional Airport | Cody |
| COS | City of Colorado Springs Municipal Airport | Colorado Springs |
| COU | Columbia Regional Airport | Columbia |
| CPR | Natrona County International Airport | Casper |
| CRP | Corpus Christi International Airport | Corpus Christi |
| CRW | Yeager Airport | Charleston |
| CSG | Columbus Metropolitan Airport | Columbus |
| CVG | Cincinnati/Northern Kentucky International Airport | Covington |
| CWA | Central Wisconsin Airport | Mosinee |
| DAB | Daytona Beach International Airport | Daytona Beach |
| DAL | Dallas Love Field | Dallas |
| DAY | James M. Cox Dayton International Airport | Dayton |
| DBQ | Dubuque Regional Airport | Dubuque |
| DCA | Ronald Reagan Washington National Airport | Arlington |
| DEN | Denver International Airport | Denver |
| DFW | Dallas/Fort Worth International Airport | Dallas-Fort Worth |
| DHN | Dothan Regional Airport | Dothan |
| DIK | Dickinson Theodore Roosevelt Regional Airport | Dickinson |
| DLG | Dillingham Airport | Dillingham |
| DLH | Duluth International Airport | Duluth |
| DRO | Durango-La Plata County Airport | Durango |
| DSM | Des Moines International Airport | Des Moines |
| DTW | Detroit Metropolitan Airport | Detroit |
| DVL | Devils Lake Regional Airport | Devils Lake |
| EAU | Chippewa Valley Regional Airport | Eau Claire |
| ECP | Northwest Florida Beaches International Airport | Panama City |
| EGE | Eagle County Regional Airport | Eagle |
| EKO | Elko Regional Airport | Elko |
| ELM | Elmira/Corning Regional Airport | Elmira |
| ELP | El Paso International Airport | El Paso |
| ERI | Erie International Airport | Erie |
| ESC | Delta County Airport | Escanaba |
| EUG | Eugene AirportÂ (Mahlon Sweet Field) | Eugene |
| EVV | Evansville Regional Airport | Evansville |
| EWN | Coastal Carolina Regional AirportÂ (Craven County Regional) | New Bern |
| EWR | Newark Liberty International Airport | Newark |
| EYW | Key West International Airport | Key West |
| FAI | Fairbanks International Airport | Fairbanks |
| FAR | Hector International Airport | Fargo |
| FAT | Fresno Yosemite International Airport | Fresno |
| FAY | Fayetteville Regional Airport | Fayetteville |
| FCA | Glacier Park International Airport | Kalispell |
| FLG | Flagstaff Pulliam Airport | Flagstaff |
| FLL | Fort Lauderdale-Hollywood International Airport | Ft. Lauderdale |
| FNT | Bishop International Airport | Flint |
| FSD | Sioux Falls Regional Airport | Sioux Falls |
| FSM | Fort Smith Regional Airport | Fort Smith |
| FWA | Fort Wayne International Airport | Fort Wayne |
| GCC | Gillette-Campbell County Airport | Gillette |
| GCK | Garden City Regional Airport | Garden City |
| GEG | Spokane International Airport | Spokane |
| GFK | Grand Forks International Airport | Grand Forks |
| GGG | East Texas Regional Airport | Longview |
| GJT | Grand Junction Regional AirportÂ (Walker Field) | Grand Junction |
| GNV | Gainesville Regional Airport | Gainesville |
| GPT | Gulfport-Biloxi International Airport | Gulfport-Biloxi |
| GRB | Green Bay-Austin Straubel International Airport | Green Bay |
| GRI | Central Nebraska Regional Airport | Grand Island |
| GRK | Killeen-Fort Hood Regional Airport | Killeen |
| GRR | Gerald R. Ford International Airport | Grand Rapids |
| GSO | Piedmont Triad International Airport | Greensboro |
| GSP | Greenville-Spartanburg International Airport | Greer |
| GST | Gustavus Airport | Gustavus |
| GTF | Great Falls International Airport | Great Falls |
| GTR | Golden Triangle Regional Airport | Columbus-Starkville-West Point |
| GUC | Gunnison-Crested Butte Regional Airport | Gunnison |
| GUM | Guam International Airport | Agana |
| HDN | Yampa Valley AirportÂ (Yampa Valley Regional) | Hayden |
| HIB | Range Regional AirportÂ (Chisholm-Hibbing Airport) | Hibbing |
| HLN | Helena Regional Airport | Helena |
| HNL | Honolulu International Airport | Honolulu |
| HOB | Lea County Regional Airport | Hobbs |
| HOU | William P. Hobby Airport | Houston |
| HPN | Westchester County Airport | White Plains |
| HRL | Valley International Airport | Harlingen |
| HSV | Huntsville International Airport | Huntsville |
| HYA | Barnstable Municipal Airport | Hyannis |
| HYS | Hays Regional Airport | Hays |
| IAD | Washington Dulles International Airport | Chantilly |
| IAG | Niagara Falls International Airport | Niagara Falls |
| IAH | George Bush Intercontinental Airport | Houston |
| ICT | Wichita Dwight D. Eisenhower National AirportÂ (Wichita Mid-Continent Airport) | Wichita |
| IDA | Idaho Falls Regional Airport | Idaho Falls |
| ILG | Wilmington Airport | Wilmington |
| ILM | Wilmington International Airport | Wilmington |
| IMT | Ford Airport | Iron Mountain/Kingsford |
| IND | Indianapolis International Airport | Indianapolis |
| INL | Falls International Airport | International Falls |
| ISN | Sloulin Field International Airport | Williston |
| ISP | Long Island MacArthur Airport | Islip |
| ITH | Ithaca Tompkins Regional Airport | Ithaca |
| ITO | Hilo International Airport | Hilo |
| JAC | Jackson Hole Airport | Jackson |
| JAN | Jackson-Evers International Airport | Jackson |
| JAX | Jacksonville International Airport | Jacksonville |
| JFK | John F. Kennedy International AirportÂ (New York International Airport) | New York |
| JLN | Joplin Regional Airport | Joplin |
| JMS | Jamestown Regional Airport | Jamestown |
| JNU | Juneau International Airport | Juneau |
| KOA | Kona International Airport at Keahole | Kailua/Kona |
| KTN | Ketchikan International Airport | Ketchikan |
| LAN | Capital Region International Airport (Lansing Capital City) | Lansing |
| LAR | Laramie Regional Airport | Laramie |
| LAS | McCarran International Airport | Las Vegas |
| LAW | Lawton-Fort Sill Regional Airport | Lawton |
| LAX | Los Angeles International Airport | Los Angeles |
| LBB | Lubbock Preston Smith International Airport | Lubbock |
| LBE | Arnold Palmer Regional Airport | Latrobe |
| LCH | Lake Charles Regional Airport | Lake Charles |
| LEX | Blue Grass Airport | Lexington |
| LFT | Lafayette Regional Airport | Lafayette |
| LGA | LaGuardia Airport (Marine Air Terminal) | New York |
| LGB | Long Beach AirportÂ (Daugherty Field) | Long Beach |
| LIH | Lihue Airport | Lihue |
| LIT | Bill and Hillary Clinton National AirportÂ (Adams Field) | Little Rock |
| LNK | Lincoln AirportÂ (Lincoln Municipal) | Lincoln |
| LRD | Laredo International Airport | Laredo |
| LSE | La Crosse Regional Airport | La Crosse |
| LWS | Lewiston-Nez Perce County Airport | Lewiston |
| MAF | Midland International Airport | Midland |
| MBS | MBS International Airport | Saginaw |
| MCI | Kansas City International Airport | Kansas City |
| MCO | Orlando International Airport | Orlando |
| MDT | Harrisburg International Airport | Harrisburg |
| MDW | Chicago Midway International Airport | Chicago |
| MEI | Meridian Regional Airport | Meridian |
| MEM | Memphis International Airport | Memphis |
| MFE | McAllen-Miller International AirportÂ (McAllen Miller International) | McAllen |
| MFR | Rogue Valley International Airport | Medford |
| MGM | Montgomery Regional Airport | Montgomery |
| MHK | Manhattan Regional Airport | Manhattan |
| MHT | Manchester-Boston Regional Airport | Manchester |
| MIA | Miami International Airport | Miami |
| MKE | General Mitchell International Airport | Milwaukee |
| MKG | Muskegon County Airport | Muskegon |
| MLB | Melbourne International Airport | Melbourne |
| MLI | Quad City International Airport | Moline |
| MLU | Monroe Regional Airport | Monroe |
| MMH | Mammoth Yosemite Airport | Mammoth Lakes |
| MOB | Mobile Regional Airport | Mobile |
| MOT | Minot International Airport | Minot |
| MQT | Sawyer International Airport | Marquette |
| MRY | Monterey Regional AirportÂ (Monterey Peninsula Airport) | Monterey |
| MSN | Dane County Regional Airport | Madison |
| MSO | Missoula International Airport | Missoula |
| MSP | Minneapolis-Saint Paul International Airport | Minneapolis |
| MSY | Louis Armstrong New Orleans International Airport | New Orleans |
| MTJ | Montrose Regional Airport | Montrose |
| MVY | Martha's Vineyard Airport | Marthas Vineyard |
| MYR | Myrtle Beach International Airport | Myrtle Beach |
| OAJ | Albert J. Ellis Airport | Jacksonville |
| OAK | Oakland International Airport | Oakland |
| OGG | Kahului Airport | Kahului |
| OKC | Will Rogers World Airport | Oklahoma City |
| OMA | Eppley Airfield | Omaha |
| OME | Nome Airport | Nome |
| ONT | Ontario International Airport | Ontario |
| ORD | Chicago O'Hare International Airport | Chicago |
| ORF | Norfolk International Airport | Norfolk |
| ORH | Worcester Regional Airport | Worcester |
| OTH | Southwest Oregon Regional AirportÂ (North Bend Municipal) | North Bend |
| OTZ | Ralph Wien Memorial Airport | Kotzebue |
| PAH | Barkley Regional Airport | Paducah |
| PBG | Plattsburgh International Airport | Plattsburgh |
| PBI | Palm Beach International Airport | West Palm Beach |
| PDX | Portland International Airport | Portland |
| PHF | Newport News/Williamsburg International Airport | Newport News |
| PHL | Philadelphia International Airport | Philadelphia |
| PHX | Phoenix Sky Harbor International Airport | Phoenix |
| PIA | General Wayne A. Downing Peoria International Airport | Peoria |
| PIB | Hattiesburg-Laurel Regional Airport | Hattiesburg-Laurel |
| PIH | Pocatello Regional Airport | Pocatello |
| PIT | Pittsburgh International Airport | Pittsburgh |
| PLN | Pellston Regional Airport of Emmet County | Pellston |
| PNS | Pensacola International AirportÂ (Pensacola Gulf Coast Regional Airport) | Pensacola |
| PPG | Pago Pago International Airport (Tafuna Airport) | Pago Pago |
| PSC | Tri-Cities Airport | Pasco |
| PSE | Mercedita Airport | Ponce |
| PSG | Petersburg James A. Johnson Airport | Petersburg |
| PSP | Palm Springs International Airport | Palm Springs |
| PUB | Pueblo Memorial Airport | Pueblo |
| PVD | Theodore Francis Green State Airport | Providence |
| PWM | Portland International Jetport | Portland |
| RAP | Rapid City Regional Airport | Rapid City |
| RDD | Redding Municipal Airport | Redding |
| RDM | Redmond Municipal AirportÂ (Roberts Field) | Redmond |
| RDU | Raleigh-Durham International Airport | Raleigh |
| RHI | Rhinelander-Oneida County Airport | Rhinelander |
| RIC | Richmond International Airport | Richmond |
| RKS | Rock Springs-Sweetwater County Airport | Rock Springs |
| RNO | Reno/Tahoe International Airport | Reno |
| ROA | Roanoke Regional AirportÂ (Woodrum Field) | Roanoke |
| ROC | Greater Rochester International Airport | Rochester |
| ROW | Roswell International Air Center | Roswell |
| RST | Rochester International Airport | Rochester |
| RSW | Southwest Florida International Airport | Ft. Myers |
| SAF | Santa Fe Municipal Airport | Santa Fe |
| SAN | San Diego International AirportÂ (Lindbergh Field) | San Diego |
| SAT | San Antonio International Airport | San Antonio |
| SAV | Savannah/Hilton Head International Airport | Savannah |
| SBA | Santa Barbara Municipal AirportÂ (Santa Barbara Airport) | Santa Barbara |
| SBN | South Bend International AirportÂ (South Bend Regional) | South Bend |
| SBP | San Luis Obispo County Regional AirportÂ (McChesney Field) | San Luis Obispo |
| SCC | Deadhorse AirportÂ (Prudhoe Bay Airport) | Deadhorse |
| SCE | University Park Airport | State College |
| SDF | Louisville International AirportÂ (Standiford Field) | Louisville |
| SEA | Seattle-Tacoma International Airport | Seattle |
| SFO | San Francisco International Airport | San Francisco |
| SGF | Springfield-Branson National Airport | Springfield |
| SGU | St. George Regional Airport | St George |
| SHV | Shreveport Regional Airport | Shreveport |
| SIT | Sitka Rocky Gutierrez Airport | Sitka |
| SJC | Norman Y. Mineta San JosÃ© International Airport | San Jose |
| SJT | San Angelo Regional AirportÂ (Mathis Field) | San Angelo |
| SJU | Luis MuÃ±oz MarÃ­n International Airport | San Juan |
| SLC | Salt Lake City International Airport | Salt Lake City |
| SMF | Sacramento International Airport | Sacramento |
| SMX | Santa Maria Public AirportÂ (Capt G. Allan Hancock Field) | Santa Maria |
| SNA | John Wayne AirportÂ (Orange County Airport) | Santa Ana |
| SPI | Abraham Lincoln Capital Airport | Springfield |
| SPS | Wichita Falls Municipal Airport/Sheppard AFB | Wichita Falls |
| SRQ | Sarasota-Bradenton International Airport | Sarasota |
| STC | St. Cloud Regional Airport | St Cloud |
| STL | St. Louis International Airport at Lambert Field | St Louis |
| STT | Cyril E. King Airport | Charlotte Amalie |
| STX | Henry E. Rohlsen Airport | Christiansted |
| SUN | Friedman Memorial Airport | Hailey |
| SUX | Sioux Gateway Airport | Sioux City |
| SWF | Stewart International Airport | Newburgh |
| SYR | Syracuse Hancock International Airport | Syracuse |
| TLH | Tallahassee International Airport | Tallahassee |
| TOL | Toledo Express Airport | Toledo |
| TPA | Tampa International Airport | Tampa |
| TRI | Tri-Cities Regional Airport | Bristol |
| TTN | Trenton Mercer Airport | Trenton |
| TUL | Tulsa International Airport | Tulsa |
| TUS | Tucson International Airport | Tucson |
| TVC | Cherry Capital Airport | Traverse City |
| TWF | Magic Valley Regional AirportÂ (Joslin Field) | Twin Falls |
| TXK | Texarkana Regional AirportÂ (Webb Field) | Texarkana |
| TYR | Tyler Pounds Regional Airport | Tyler |
| TYS | McGhee Tyson Airport | Knoxville |
| UST | Northeast Florida Regional AirportÂ (St. Augustine Airport) | St. Augustine |
| VEL | Valdez Airport | Vernal |
| VLD | Valdosta Regional Airport | Valdosta |
| VPS | Destin-Fort Walton Beach Airport/Eglin AFB | Valparaiso |
| WRG | Wrangell Airport | Wrangell |
| WYS | Westerly State Airport | West Yellowstone |
| XNA | Northwest Arkansas Regional Airport | Fayetteville/Springdale/Rogers |
| YAK | Yakutat Airport | Yakutat |
| YUM | Yuma International Airport | Yuma |

2) Airplanes:

|  |  |  |  |
| --- | --- | --- | --- |
| IATA\_CODE | AIRLINE |  |  |
| UA | United Air Lines | |  |
| AA | American Airlines Inc. | | |
| US | US Airways Inc. | |  |
| F9 | Frontier Airlines | |  |
| B6 | JetBlue Airways | |  |
| OO | SkyWest Airlines | |  |
| AS | Alaska Airlines Inc. | |  |
| NK | Spirit Air Lines | |  |
| WN | Southwest Airlines Co. | | |
| DL | Delta Air Lines Inc. | |  |
| EV | Atlantic Southeast Airlines | | |
| HA | Hawaiian Airlines Inc. | | |
| MQ | American Eagle Airlines Inc. | | |
| VX | Virgin America | |  |

**SECTION E: Insights from the data source**

**Insight**: Arrival and departure delay for each Airline

In this insight we can see which airline is best and which is not to travel by within USA based on the arrival and departure delay.

We have considered airlines on X axis and all the delay values on Y axis in order to plot the bar chart.

We have concluded that WN- southwest airline is not the good airline if we consider the arrival airline and departure delays.

Secondly, we can conclude that HA- Hawaiian airline has the lowest delays.

The screenshot:Chart, bar chart

Description automatically generated

Furthermore, analysis of airline based on the not considering the top (n) not good airline and considering the bottom (n) airline based on the more and less delay caused by it.

In this, top (n) and bottom (n) analysis we have considered two reasons for this delay, the first is diverted delay and weather delay.

And from that as well we have concluded that WN- southwest airline has the more diverted delays and weather delays in comparison to the all the other airlines.

VX-virgin America has the least diverted delay and weather delay.

So, it will be a good option to consider the VX airline and avoid the WN airline if you want to avoid diverted weather delay.

The Screenshot is as follows:

Chart, bar chart

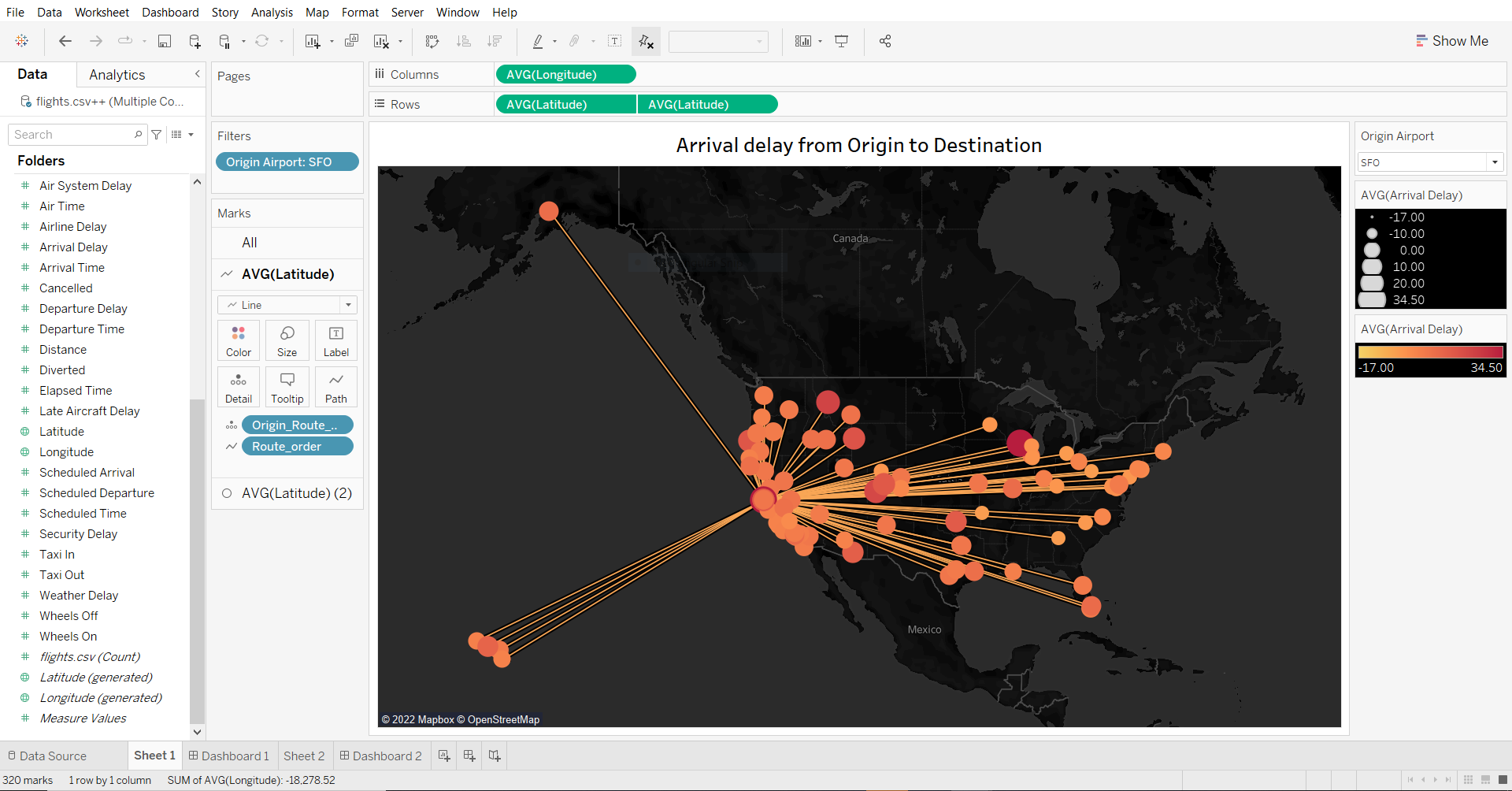
Description automatically generated

**Destination map. (Origin to destination with average arrival delay)**

**Insight**: Through this map we can analyze few things:

1. We can figure out what all destination airports the airline travels to from a single selected origin.
2. Secondly, as the origin and destination has been defined on map, we can analyze the average arrival delay of the flight for each airport from a selected origin airport.
3. And finally, we can see which airport had the most avg arrival delay for a specific selected origin airport.

This is how it was created using TABLEAU.



This is the final Dashboard representation of destination map (ORIGIN to DESTINATION).

Diagram

Description automatically generated

Here, in this dashboard we have selected the origin airport as SFO (San Francisco international Airport). And we can see that it represents all the airports the SFO airport is connected to.

It also shows the avg arrival delay for each of the airport.

Considering the big dot on the graph represents the average arrival delay. The darker the dot is the more the delay is.

So, if the origin airport is SFO the highest average arrival delay is for

the SFO\_MSN (Dane County Regional Airport, Madison) airport is for 34.50 mins.

Least average arrival delay is for the SFO\_ATL

(Hartsfield-Jackson Atlanta International Airport, Atlanta) is -3.32 min(minus sign indicates that the flight was 3.32 mins before the expected the avg arrival delay for that airport)

**Destination Map ( Origin to Destination with avg departure delay).**

**Insight**: Through this map we can analyze few things:

It analyzes all the flights coming from different destinations

to a specific origin. And analyze their average departure

delay or find travel path between destination and origin

airport.

This is how it is created on the tableau.

Graphical user interface, application, website, map

Description automatically generated

Dashboard representation of (origin to Destination average departure delay):

Map

Description automatically generated

From this above graph we can conclude that if we select the origin airport as SFO. The highest average departure delay is for the (Missoula International Airport, MISSOULA) MSO\_SFO airport which is 23.26 min.

On the other hand, the lowest average departure delay for the

(Yampa Valley Airport (Yampa Valley Regional, Hayden) HDN\_SFO is -4.37min which means the avg departure delay is not actually delay but those flights are -4.37 min earlier.

**Weekly Airline performance**

**Insight:** From this Airline we can determine the airline delay by the days of the week considering the airline, avg departure delay and find the number of diverted flights and number of flights for specific airline

From the screenshot below we can determine that for the MQ airline the avg departure delay is 9. As well as the lowest airline delay is on Wednesday and highest airline delay is on month. Similarly, we can find out from the correlation graph between avg arrival delay and departure delay and plotting the trend line considering all the airports we can analyze which airline is going through which all airports.

Chart, line chart

Description automatically generated

**Reasons for airline delay with respect to each airline.**

Insight: Using Bar chart between multiple reasons for airline delay and airline we can predict which airline to be travel by to avoid the reasons which is causing those delays.

Chart, waterfall chart

Description automatically generated

From the above chart we can analyze each reason with respect to each airline.

The Screenshot provided below will show the actual numbers.

Table

Description automatically generated

**Flight cancellation analysis.**

**Insight**: Sankey chart represents - Airlines and their Cancelled number of Flights based on Month

After creating all the required calculations and plotting of the above component together we can see a very interesting model of Sankey chart which has been created. And this will help in analyzing the relation between the left-side source component with the right-side target component and all the sigmoid create curves will display the relation between source and target.

So, from the below chart we can analyze that for each month for year 2015 (which is source), flights got cancelled with respect to different airline (which is target), how much percent of all the flight for each respective airline got cancelled as well as how many percentages of overall flight got cancelled for with respective to each airline.

So, by looking at the below Sankey chart we can conclude that,

WN-Southwest Airline highest cancelled flight with (% of total count of cancelled long tables(across ) is 21.70%) and (% of total cancelled long tables(across ) is 17%).

On the other hand, HA- Hawaiian Airlines has the lowest cancelled flight with (% of total count of cancelled long tables (across ) is 1.31%) and (% of total cancelled long tables(across ) is 0.18%).

Considering the Months, February of 2015 was the highest month where the flights got cancelled – 23.47% (% by consider all the Airlines).And April was the month where there was least flight got cancelled 5.17% (% by consider all the Airlines).

Diagram

Description automatically generated

**Conclusion for Sankey chart**:

We have seen what Sankey Chart is In Tableau and how to create it with two different dimensions. Whenever we need to map data with different dimensions Sankey chart is the best-suited way. It helps us show the mapping source and target with the help of the size. Here in this screenshot as above we have created the Sankey chart between two dimension called as month and all American airlines.

**Section F: Use case**

1. This analysis is used in real time to find out all the airlines’ performance based on delays they cause for different reasons and the cancellation numbers.
2. Using the correlation graph we can also determine the relation between the average delays and the airports. Through which we can find which airline flies through which all airports. And this is helpful to find which airline to use in order to go to any specific city.
3. We can find the best weekday to book the airline to avoid the delays.

**Section G: Summar****y**

**Summary for each insight:**  
  
1). From the above analysis based on the plotted and calculated charts, we can **select the best airline** to travel with and by considering the reasons such as best weekdays and avoiding reasons for multiple delays.  
  
2). We can determine which airlines are going from specific origin to the different destination and plot their avg arrival delay. On the other hand, we can plot the inverse of this by plotting the avg departure delay from different destination to a specific origin.  
  
3). We have calculated the cancelled flights between the months of 2015 and airlines by using the “SANKEY CHART”. The visualization shows what accounted for the flight cancellation in a particular month. Finally, it gives the percentage contribution of all airlines for the overall delays in a year divided into months.

# ****Section H: References****

1. Anderson Ong, M. (2017) “Data Visualization of Flight delays with Tableau”. Medium, 2017

2. Ball M., Barnhart C., Dresner M., Hansen M., Neels K., Odoni A., Peterson E., Sherry L., Trani A., Zou B. (2010). Total Delay Impact Study — A Comprehensive Assessment of the Costs and Impacts of Flight Delay in the United States. Final Report — October, 2010. Accessed from website: <http://scholar.rhsmith.umd.edu/sites/default/files/mball/files/total_delay_0.pdf?m=1467366850>

3. Dey, T., Phillips, D. and Steele, P. (2011). A Graphical Tool to Visualize Predicted Minimum Delay Flights. Journal of Computational and Graphical Statistics, 20(2), pp.294–297.

4. Gopalakrishnan K., Balakrishnan H. (2017). “A Comparative Analysis of Models for Predicting Delays in Air Traffic Networks.” Air Traffic Management Research and Development Seminar, June 2017, Seattle, Washington, USA, ATM Seminar, June 2017

5. Gupta P., Dwivedi A., Agrawal A. (2016). “An Analysis of US Domestic Flight delays using SAS Enterprise Miner”. South Central SAS Users Group Educational Forum, 2016

6. Mainero A., Schmidt T. and Sugarman H. (2013). Heat Mapping and Predicting Flight Delays and Their Propagations in a Real-World Air Traffic Simulation. December 2013. Retrieved from Stanford University. Access website — <http://snap.stanford.edu/class/cs224w-2013/projects2013/cs224w-056-final.pdf.>

7. Michael T. Crotty. (2014). “Visualizing More Than Twenty Years of Flight Data for the Raleigh-Durham International Airport.” SAS Institute, USA, 2014

8. Rebollo Jose J. and Balakrishnan H. (2012). “A Network-Based Model for Predicting Air Traffic Delays.” 5th International Conference on Research in Air Transportation (ICRAT 2012)

9. Rodríguez-Sanz, Á., Gómez Comendador, F., Arnaldo Valdés, R., Cordero García, J. and Bagamanova, M. (2018). Uncertainty Management at the Airport Transit View. Aerospace, 5(2), p.59 Received: 28 January 2018; Accepted: 21 May 2018; Published: 1 June 2018.

10. Sternberg A., Soares J., Carvalho D., Ogasawara E. (2017). A Review on Flight Delay Prediction. CEFET/RJ Rio de Janeiro, Brazil, November 6, 2017.

11. Understanding the Reporting of Causes of Flight Delays and Cancellations | Bureau of Transportation Statistics. [online] Available at: <https://www.bts.gov/topics/airlines-and-airports/understanding-reporting-causes-flight-delays-and-cancellations>.

12. Wicklin, R. (2009). “An Analysis of Airline Delays with SAS/IML Studio”. SAS Institute Inc., 2009

13. [www.educba.com](http://www.educba.com)

14. [www.sjsu.com](http://www.sjsu.com)