Airline Delay Analysis Presentation

Justin, Hanzian, Gideon

Overview

This project has been inherited from The Airline Domain. In Airline, if you want to travel or anyone who wants to travel, he must book the flight from one place to another.

There are number of factors, which can impact the flight journey like Weather, flight departure time, boarding gate time and departure time etc. Keeping these factors in mind, we can decide that particular aircraft can be landed or arrive on time or not or how much it will be delay.

Every airline has their flight history past journey, which can help them in predicting future flight delay. We can implement a machine learning model, which will help us in the prediction of a flight delay.

The Motivation behind this project is to optimization of network operation, ground staff management and passenger.

Understanding the problems

Flight delays are a common occurrence in the airline industry, causing significant inconvenience and frustration for passengers. Despite efforts to improve airline operations and reduce delays, they continue to impact the travel experience and cost the industry billions of dollars each year. An analysis of flight delays could involve examining various factors that contribute to delays, such as:

- Weather conditions: Extreme weather conditions such as storms, heavy rain, snow, and fog can cause flight delays or cancellations.
- Air traffic congestion: Busy airports and crowded airspace can cause delays in takeoff and landing.
- Technical issues: Technical issues with aircraft or equipment can cause delays, as well as maintenance or repair work.
- Crew issues: Delays can occur if there is a shortage of pilots or flight attendants, or if they are delayed or unavailable due to sickness, injury, or other reasons.
- Security issues: Security checks and procedures can cause delays, especially during high-security situations.

Project objective

Analyzing flight delay data can provide insights into patterns and trends, such as the busiest times of year or the most common reasons for delays. This information can help airlines and airport operators to develop strategies for managing delays and improving the overall travel experience for passengers. It can also inform policy decisions related to air travel, such as regulations related to airport capacity or airline operations.

Questions to Answer

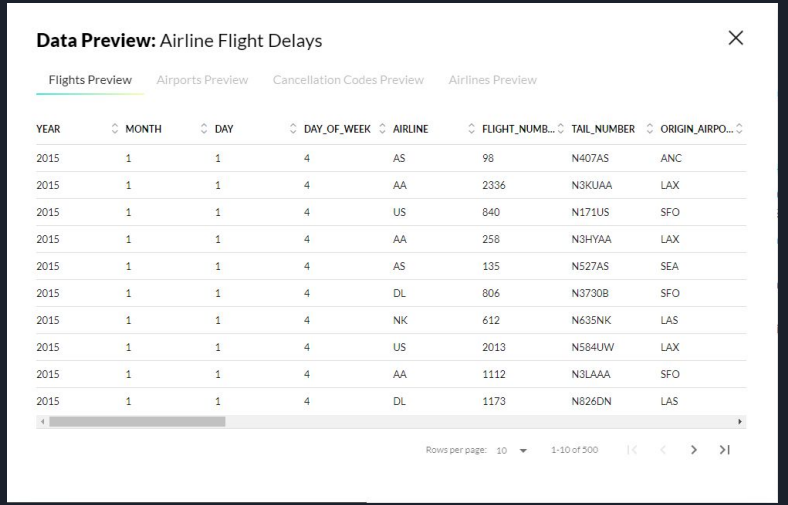
- How does the overall flight volume vary by month? By day of week?
- What percentage of flights in experienced a departure delay in 2015? Among those flights, what was the average delay time, in minutes?
- How does the % of delayed flights vary throughout the year?
- How many flights were cancelled in 2015? What % of cancellations were due to weather? What % were due to the Airline/Carrier?
- Which airlines seem to be most and least reliable, in terms of on-time departure?

Data sets

Records for 5,000,000+ commercial airline flights in 2015, compiled for the U.S. DOT Air Travel Consumer Report. Each record represents a single flight, including the airline name, flight number, origin/destination airport and flight distance, as well as scheduled/actual departure and arrival times.

https://www.mavenanalytics.io/data-playground?search=flight%20delays

- flights.csv
- airports.csv
- airline.csv
- cancellation_codes.csv



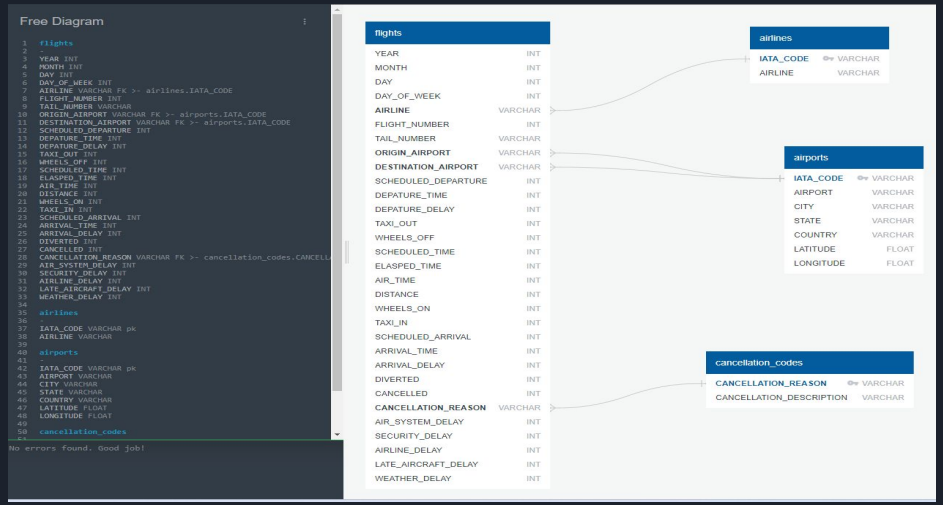
Data Preview: Airline Flight Delays

Flights Preview | Airports Preview | Cancellation Codes Preview | Airlines Preview

YEAR	MONTH	DAY	DAY_OF_WEEK	AIRLINE	FLIGHT_NUMB...	TAIL_NUMBER	ORIGIN_AIRPO...
2015	1	1	4	AS	98	N407AS	ANC
2015	1	1	4	AA	2336	N3KUAA	LAX
2015	1	1	4	US	840	N171US	SFO
2015	1	1	4	AA	258	N3HYAA	LAX
2015	1	1	4	AS	135	N527AS	SEA
2015	1	1	4	DL	806	N3730B	SFO
2015	1	1	4	NK	612	N635NK	LAS
2015	1	1	4	US	2013	N584UW	LAX
2015	1	1	4	AA	1112	N3LAAA	SFO
2015	1	1	4	DL	1173	N826DN	LAS

Rows per page: 10 1-10 of 500

Database & ETL



We used a PostgreSQL relational database for data storage. Using a relational database allows us to creating connections between the different tables and help answer any of the questions we may have with the data. Additionally, the database can be connected to using Pandas to perform ETL processes.

Data Output | Messages | Notifications

	year integer	month integer	day integer	day_of_week integer	airline character varying	flight_number character varying	tail_number character varying	origin_airport character varying	destination_airport character varying	scheduled_departure double precision	departure_time double precision	departure_delay double precision	taxi_out double precision	wheels_off double precision
1	2015	1	1	4	AS	98	N407AS	ANC	SEA	5	2354	-11	21	
2	2015	1	1	4	AA	2336	N3KUAA	LAX	PBI	10	2	-8	12	
3	2015	1	1	4	US	840	N171US	SFO	CLT	20	18	-2	16	
4	2015	1	1	4	AA	258	N3HYAA	LAX	MIA	20	15	-5	15	
5	2015	1	1	4	AS	135	N527AS	SEA	ANC	25	24	-1	11	
6	2015	1	1	4	DL	806	N3730B	SFO	MSP	25	20	-5	18	
7	2015	1	1	4	NK	612	N635NK	LAS	MSP	25	19	-6	11	
8	2015	1	1	4	US	2013	N584UW	LAX	CLT	30	44	14	13	
9	2015	1	1	4	AA	1112	N3LAAA	SFO	DFW	30	19	-11	17	
10	2015	1	1	4	DL	1173	N826DN	LAS	ATL	30	33	3	12	
11	2015	1	1	4	DL	2336	N958DN	DEN	ATL	30	24	-6	12	
12	2015	1	1	4	AA	1674	N853AA	LAS	MIA	35	27	-8	21	
13	2015	1	1	4	DL	1434	N547US	LAX	MSP	35	35	0	18	
14	2015	1	1	4	DL	2324	N3751B	SLC	ATL	40	34	-6	18	
15	2015	1	1	4	DL	2440	N651DL	SEA	MSP	40	39	-1	28	1
16	2015	1	1	4	AS	108	N309AS	ANC	SEA	45	41	-4	17	
17	2015	1	1	4	DL	1560	N3743H	ANC	SEA	45	31	-14	25	
18	2015	1	1	4	UA	1197	N78448	SFO	IAH	48	42	-6	11	
19	2015	1	1	4	AS	122	N413AS	ANC	PDX	50	46	-4	11	

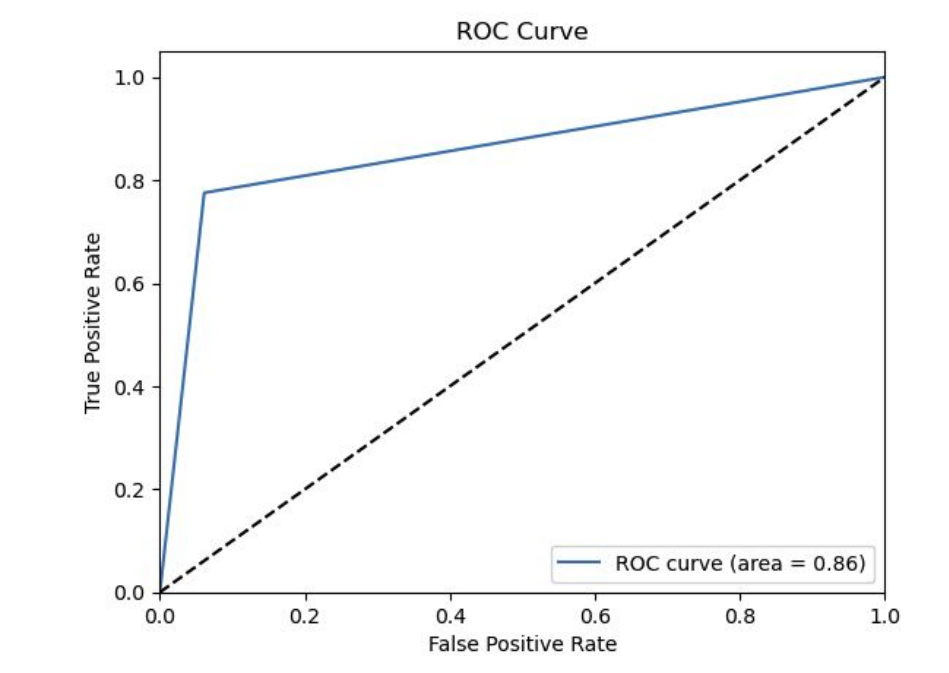
Machine Learning

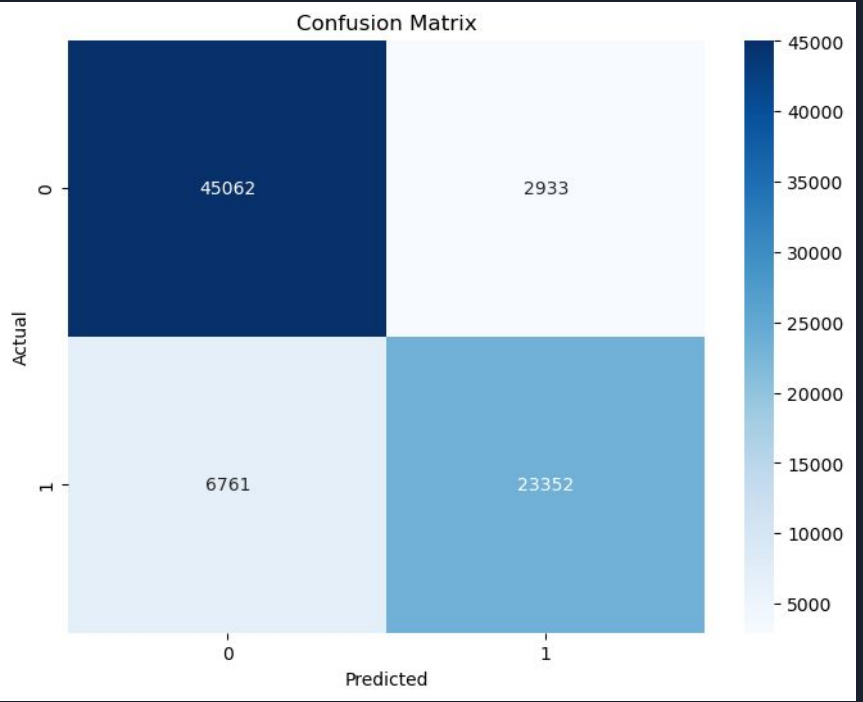
Prediction Analysis

- Logistic regression model was trained and tested using a train-test split of the data. The model achieved an accuracy score of 0.87, which means that it correctly classified 87% of the observations in the test set. The recall score is 0.89, which means that the model correctly identified 89% of the positive cases in the test set. The precision score is 0.75, which means that when the model predicted a positive case, it was correct 75% of the time.
- Decision tree classifier model achieved an accuracy score of 0.83, which means that it correctly classified 83% of the observations in the test set.
- Random forest classifier model was trained and tested using a train-test split of the data, similar to the previous models. The model achieved an accuracy score of 0.88, which means that it correctly classified 88% of the observations in the test set. The recall score is 0.89, which means that the model correctly identified 89% of the positive cases in the test set. The precision score is 0.78, which means that when the model predicted a positive case, it was correct 78% of the time.

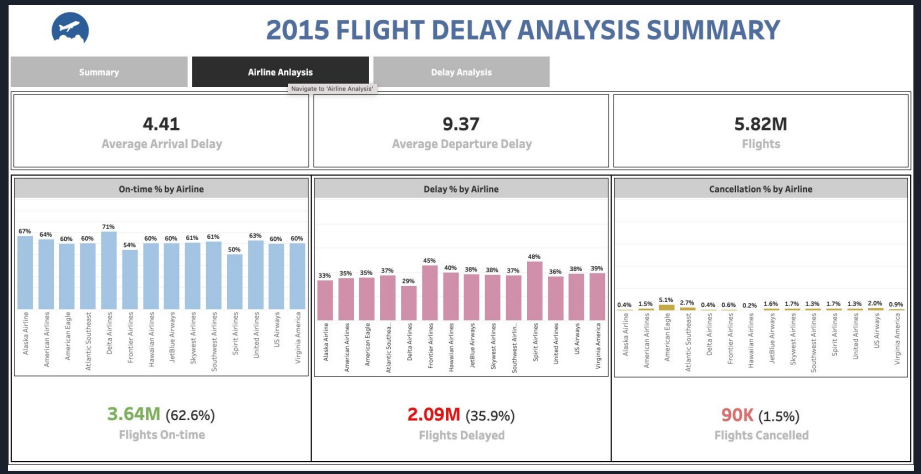
Machine Learning

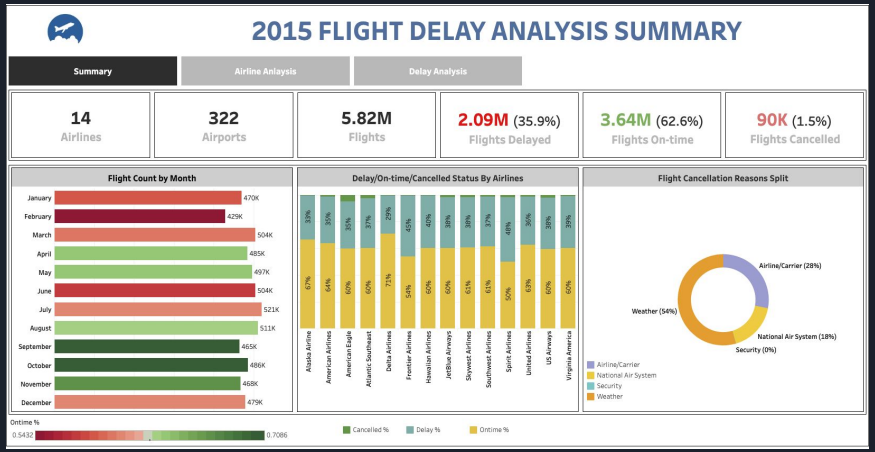
Visualizations

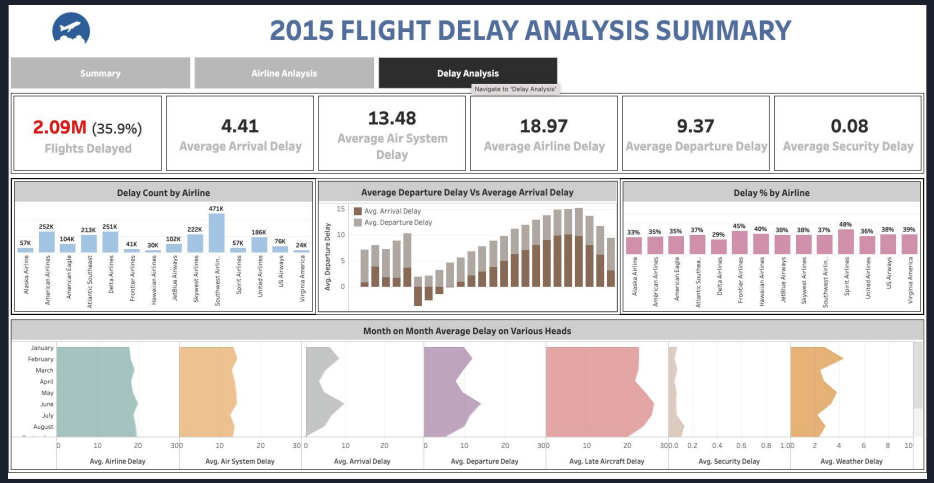




Data Visualization







We used Tableau to create a dashboard to visually present the data we have chosen. This will allows users to better understand the questions we were trying to answer and allow they to filter the data in a way that they can understand.

Conclusion

In conclusion, the data provided suggests that flight delays and cancellations are a common occurrence in the airline industry, with weather and airline operational issues being among the main contributing factors. While airlines themselves are often the primary cause of delays, air traffic control systems can also play a significant role.

When looking at specific airlines, it appears that Southwest and Spirit Airlines are among the most frequently delayed carriers, while Delta and Alaska Airlines tend to have the best on-time performance. American Eagle has the highest rate of cancellations, while Hawaiian Airlines has the lowest.

While this information can be helpful for travelers when making decisions about which airline to book with, it's important to remember that delays and cancellations can happen with any carrier. It's always a good idea to plan ahead and be prepared for potential disruptions, such as by booking flexible tickets and having backup travel plans in place.

Recommendations

1. Book flights during the months of September and October, as they have a higher likelihood of being on time compared to other months.
2. When booking with airlines such as Southwest and Spirit Airlines, be prepared for potential delays, and consider booking a flexible ticket to allow for changes if needed.
3. Consider booking with airlines such as Delta and Alaska Airlines, as they tend to have better on-time performance.
4. When traveling during months with a higher likelihood of delays (such as February, June, January, and March), plan ahead and be prepared for potential disruptions.
5. When booking flights, look for airlines with flexible change and cancellation policies, which can allow you to adjust your plans if needed.
6. Monitor your flight status and sign up for alerts from the airline, so you can be notified of any changes or delays.
7. If your flight is delayed or canceled, be proactive and reach out to the airline for assistance, such as rebooking on a different flight or receiving compensation for the inconvenience.

By following these recommendations, travelers can increase their chances of having a smooth and stress-free travel experience, even in the face of potential delays and disruptions.

Thank you!

We are now going to open the floor up for any questions anyone may have.