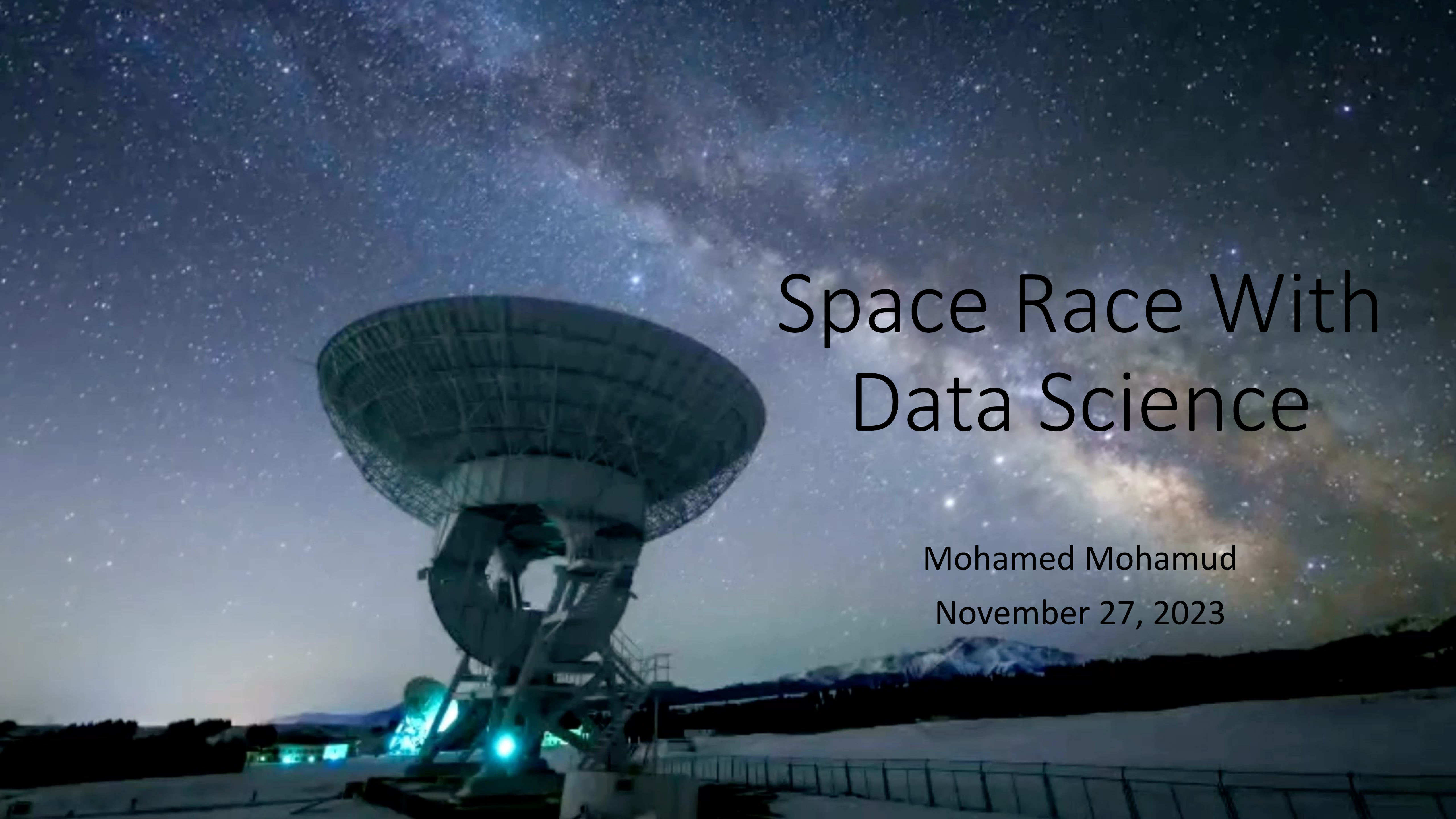


Space Race With Data Science

Mohamed Mohamud

November 27, 2023

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusions
- Appendix

Executive Summary

Summary of Methodology

- Data Collection
- Data Wrangling
- Exploratory Data Analysis with Data Visualization
- Exploratory Data Analysis with SQL
- Building A Dashboard with Plotly-Dash
- Predictive analysis-Classification

Summary of all results

- Exploratory Data Analysis results
- Interactive analytics demo in Screenshots
- Predictive analysis

Introduction

Project background and context

SpaceX is the most successful company of the commercial space age, making space travel affordable. The company advertises Falcon 9 rocket launches on its website, costing 62 million dollars; other providers cost upwards of 165 million dollars each; much of the savings is because SpaceX can reuse the first stage. Therefore, if the first stage will land, we can determine the cost of a launch. We will predict if SpaceX will reuse the first stage based on public information and machine learning models.

Questions to be answered

- How do variables such as payload mass, launch site, number of flights, and orbits affect the success of the first stage landing?
- Does the rate of successful landings increase over the years?
- What is the best algorithm that can be used for binary classification in this case?

Methodology

Data Collection

The data collection process involved a combination of API requests from SpaceX REST API and Web Scraping data from a table in SpaceX's Wikipedia entry.

Data Columns are obtained by using SpaceX REST API:

FlightNumber, Date, Booster Version, PayloadMass, Orbit, LaunchSite, Outcome, Flights, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude

Data Columns are obtained by using Wikipedia Web Scraping:

Flight No., Launch site, Payload, PayloadMass, Orbit, Customer, Launch outcome, Version Booster, Booster landing, Date, Time

Data Collection –Space X API

REQUESTING ROCKET LAUNCH DATA FROM SPACEX API

EXPORTING THE DATA TO CSV

DECODING THE RESPONSE CONTENT USINGSON AND TURNING IT INTO A DATAFRAME USING JSON NORMALIZE()

REQUESTING NEEDED INFORMATION ABOUT THE LAUNCHES FROM SPACEX API BY APPLYING CUSTOM FUNCTIONS

CONSTRUCTING DATA, WE HAVE CIBTAINED INTO A DICTIONARY

REPLACING MISSING VALUES OF PAYLOAD MASS COLUMN WITH A CALCULATED MEAN() FOR THIS COLUMN

FILTERING THE DATAFRAME ONLY TO INCLUDE FALCON 9 LAUNCHES

CREATING A DATAFRAME FROM THE DICTIONARY

Data Collection- Web Scrapping

Requesting	Requesting Falcon 9 launch data from Wikipedia
Creating	Creating a BeautifulSoup object from the HTML response
Extracting	Extracting all column names from the HTML table header
Collecting	Collecting the data by parsing HTML tables
Exporting	Exporting the data to CSV
Creating	Creating a dataframe from the dictionary
Constructing	Constructing data we have obtained into a dictionary

Data Wrangling

- The data set contains various instances where the booster failed to land successfully. In some cases, the landing was attempted but failed due to an accident. For instance, "True Ocean" indicates that the mission outcome was a successful landing in a specific region of the ocean, while "False Ocean" means that the mission outcome was an unsuccessful landing in a specific region of the ocean. Similarly, "True RTLS" indicates that the mission outcome was a successful landing on a ground pad, while "False RTLS" means that the mission outcome was an unsuccessful landing on a ground pad. Lastly, "True ASDS" indicates that the mission outcome was a successful landing on a drone ship, while "False ASDS" means that the mission outcome was an unsuccessful landing on a drone ship.
- Labels with “1” means the booster successfully landed, and "O" means it was unsuccessful.

Data Wrangling

Perform an exploratory data analysis to determine the training labels.

Calculate the number of launches that have taken place on each site, as well as the number and occurrence of each orbit.

Additionally, calculate the number and occurrence of mission outcomes per orbit type.

Create a landing outcome label from the Outcome column and then export the data to a CSV file.

EDA with Data Visualization

- The following charts were created:
  - 1. Flight Number vs. Payload Mass
  - 2. Flight Number vs. Launch Site
  - 3. Payload Mass vs. Launch Site
  - 4. Orbit Type vs. Success Rate
  - 5. Flight Number vs. Orbit Type
  - 6. Payload Mass vs. Orbit Type
  - 7. Success Rate Yearly Trend Scatter plots are used to show the relationship between variables.
- If there is a relationship between variables, it can be used in a machine-learning model. Bar charts, on the other hand, are used to compare different categories.
- They help to show the relationship between the categories being compared and the measured value.
- Line charts show trends in data over time, also known as time series.

EDA with SQL

The following SQL queries were performed:

- Names of the unique launch sites in the space mission.
- Top 5 launch sites whose name begins with the string 'CCA'.
- Total payload mass carried by boosters launched by NASA (CRS).
- Average payload mass carried by booster version F9 v1.1.
- Date when the first successful landing outcome in the ground pad was achieved.
- Total number of successful and failed mission outcomes.
- Names of the booster versions which have carried the maximum payload mass.
- Failed landing outcomes in drone ships, their booster versions, and launch site names for 2015.
- Rank of the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the dates 2010-06-04 and 2017-03-20.

Interactive map with Folium

- Folium Maps used Markers, circles, lines, and marker clusters to represent different information.
  - Markers signify specific points like launch sites.
  - Circles show highlighted areas around certain coordinates, such as NASA Johnson Space Center.
  - Marker clusters indicate groups of events in each coordinate, like all the launches that happened in a particular launch site.
  - Lines are used to indicate the distance between two coordinates.

Dashboard with Ploty Dash

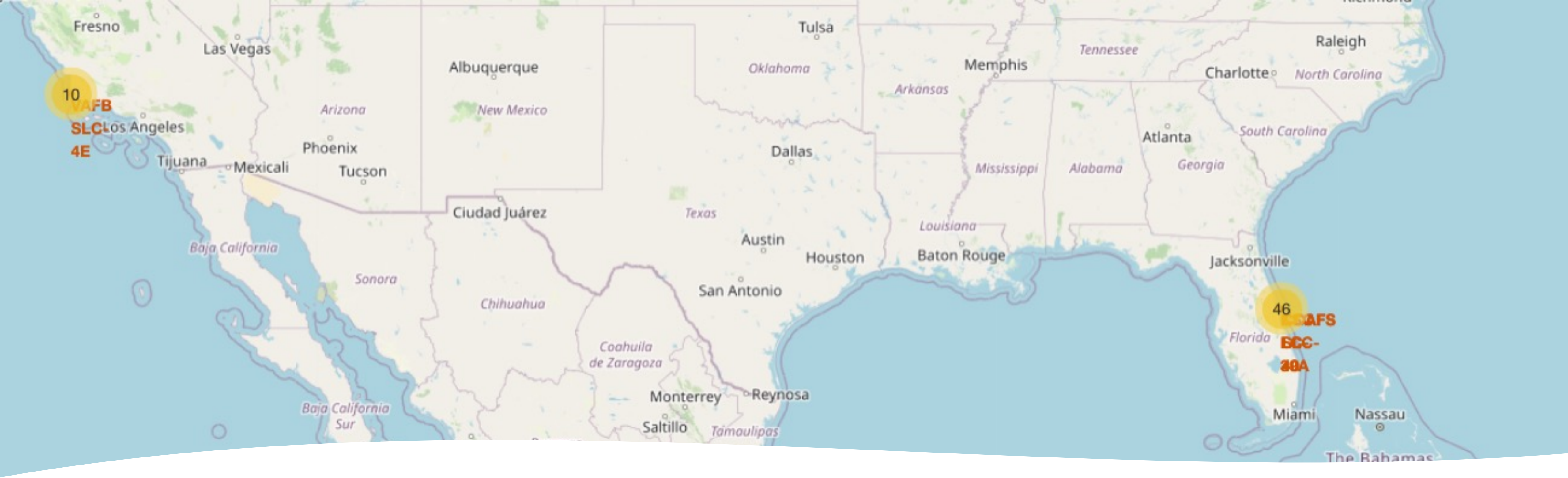
- The data was visualized using a set of graphs and plots, including the percentage of launches by site and the payload range.
- These visualizations helped to quickly identify the relationship between the payloads and launch sites, making it easier to determine the best place to launch based on the payload.

Predictive Analysis

- • Four classification models were compared: logistic regression, support vector machine, decision tree, and k nearest neighbors.
- Data preparation and standardization
- Test of each model with combinations
- Comparison results

Results

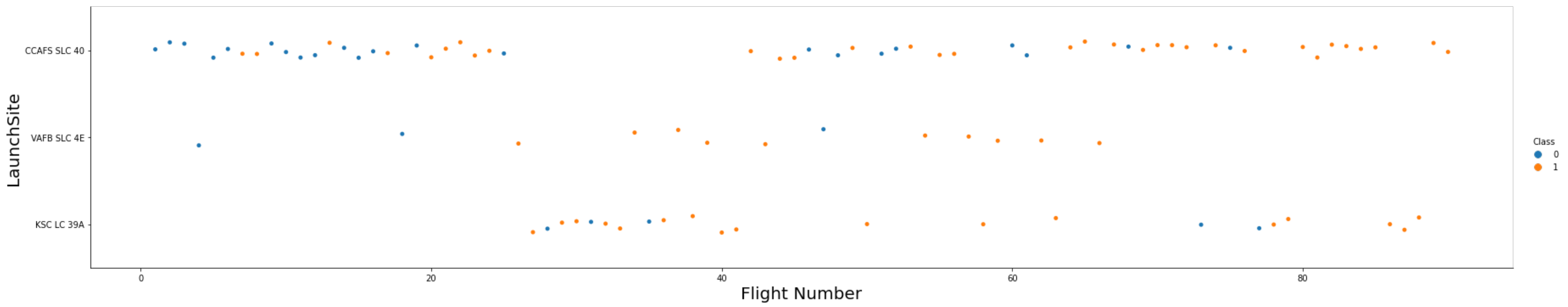
- Here are the key findings from the exploratory data analysis:
- - SpaceX has used four different launch sites.
- - The initial launches were conducted for SpaceX itself and NASA.
- - The average payload capacity of the F9 v1.1 booster is 2,928 kg.
- - It took five years after the first launch for SpaceX to achieve its first successful landing.
- - Many versions of the Falcon 9 booster have been successful at landing on drone ships with payloads above the average.
- - Nearly 100% of their missions have been successful.
- - In 2015, two booster versions, F9 v1.1 B1012 and F9 v1.1 B1015, failed to land on drone ships.
- - The number of successful landing outcomes has improved over the years.



Results

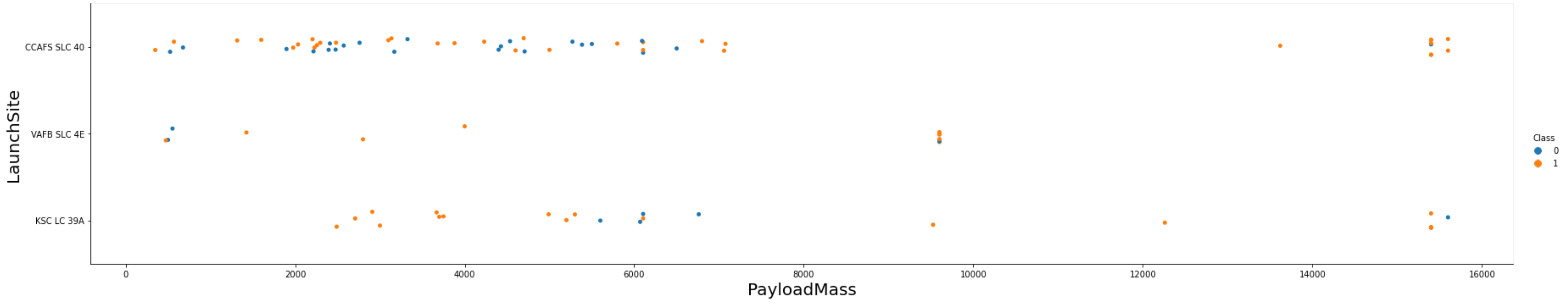
- Using interactive analytics, it was possible to identify that launch sites used to be in safe places, near the sea and had an excellent logistic infrastructure.
- Most launches happen at east-cost launch sites.
- Predictive Analysis showed that the Decision Tree Classifier is the best model to predict successful landings, having an accuracy of over 87% and an accuracy for test data of over 94%.

Insights From EDA Visualization



Flight Number vs Launched Site

- Based on the data presented in the graph, it is evident that the CCAF5 SLC 40 launch site has the highest success rate for recent launches. The VAFB SLC 4E site comes in second place, followed by KSC LC 39A in third place. Additionally, the overall success rate has shown improvement over time.

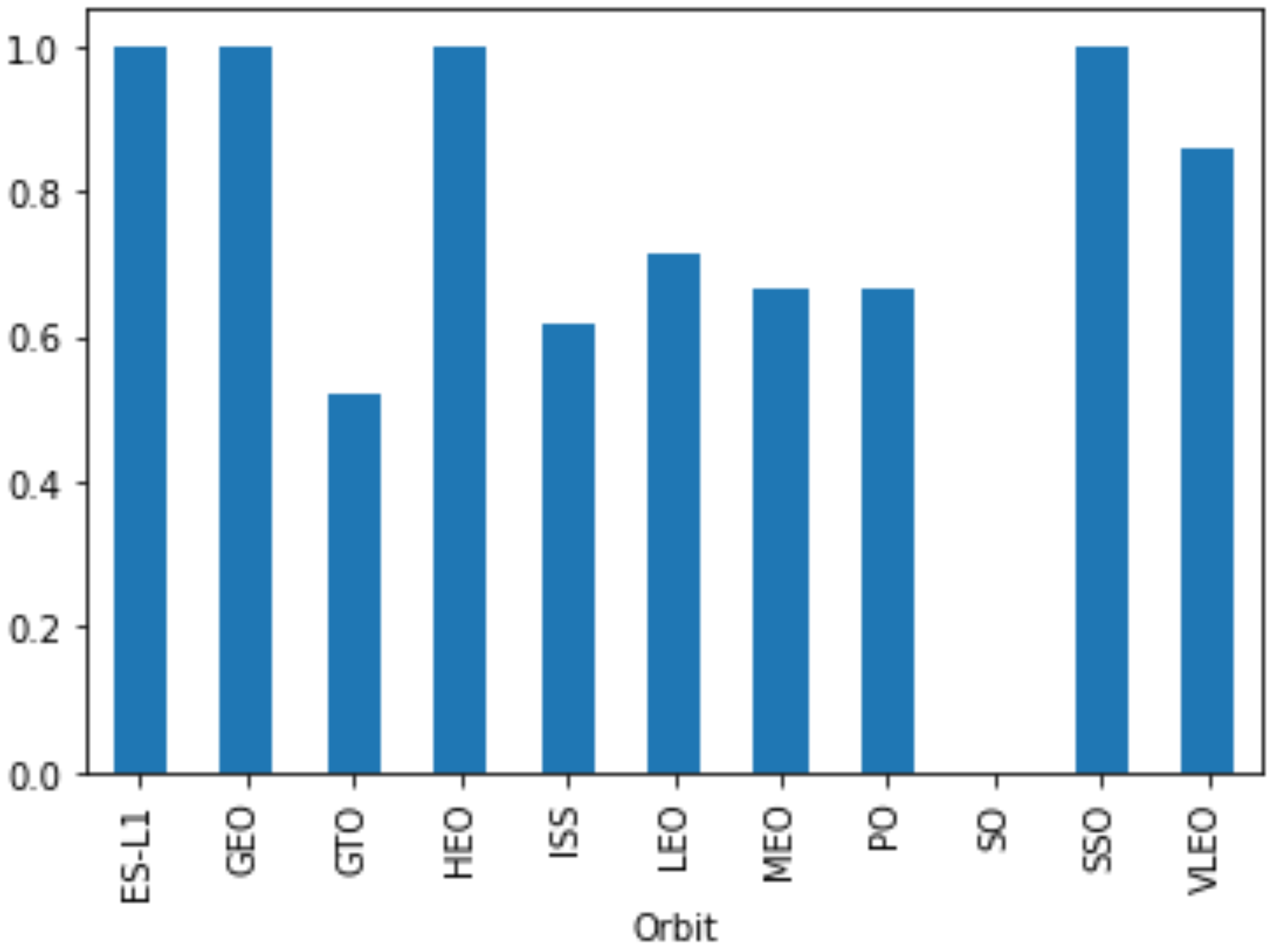


Payload with Launch Site

- Payloads weighing over 9,000kg, roughly the weight of a school bus, have a high success rate. Payloads over 12,000kg can only be launched from CCAFS SLC 40 and KSC LC 39A sites.

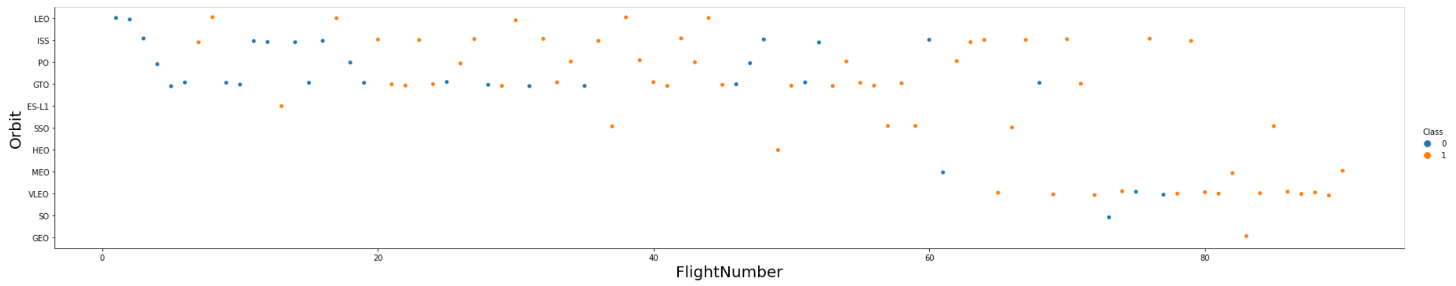
Success rates vs orbit

- The most successful orbits are ES-L1, GEO, HEO, and SSO, followed by VLEO (above 80%) and LFO (above 70%).



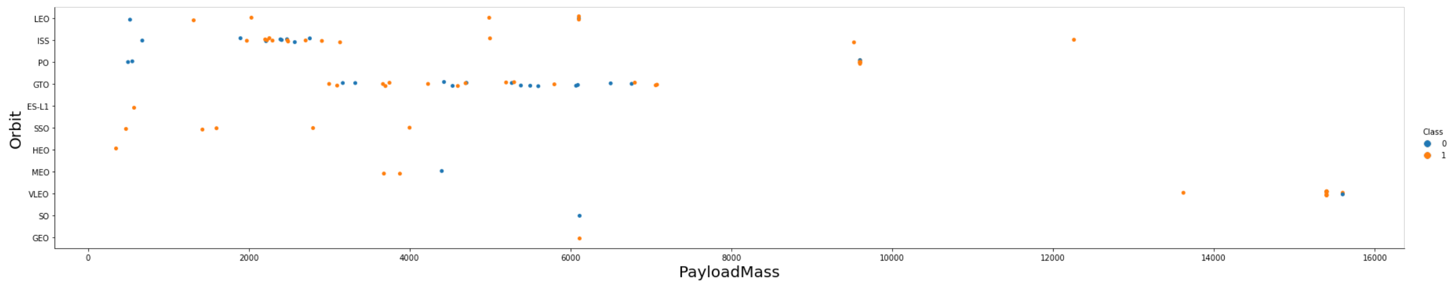
Flight Number vs Orbit Type

- Success rate improved over time to all orbits;
- VLEO orbit seems to be a new business opportunity due to the recent increase in its frequency.



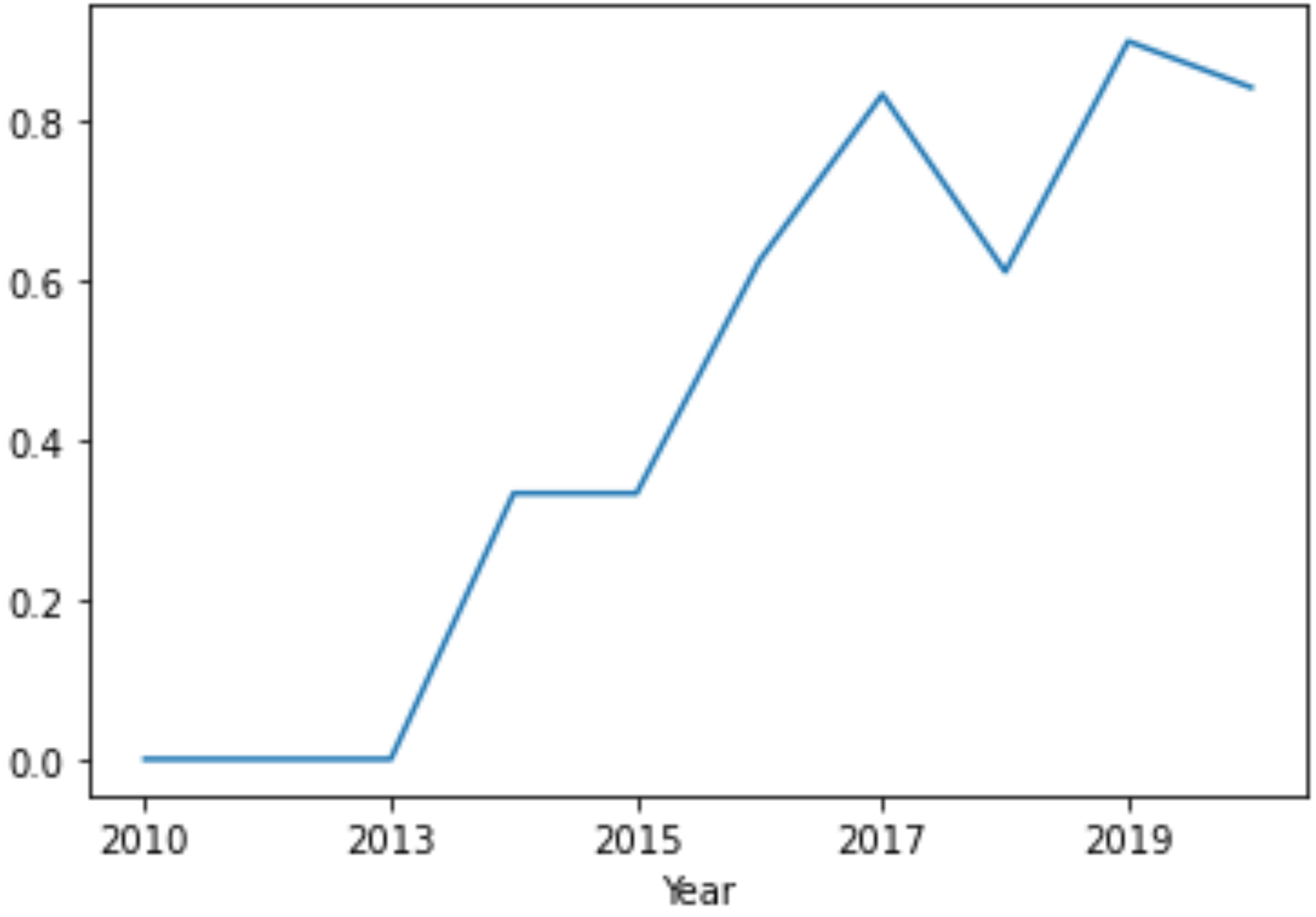
Payload Mass vs Orbit Type

- It appears that there is no correlation between payload and success rate when launching to the GTO orbit. The ISS orbit, on the other hand, has a wide range of payload options and a good success rate. It's worth noting that there are only a few launches to the SO and GEO orbits



Land Success Yearly Trend

- The success rate started increasing in 2013 and continued until 2020.
- The first three years were a period of adjustments and improvement of technology.



All Lunch Sites names

launch_site
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E

Site Names With CCA

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mi
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	

Total Payload Mass

total_payload
111268

Average Payload Mass by F9 v1.1

avg_payload
2928

First Successful Ground Landing Date

FIRST_SUCCESS_GP

2015-12-22

Success Droneship Landing

- booster_version

F9 FT B1021.

2F9 FT B1031.

2F9 FT B1022

F9 FT B1026

Total Number of Successful and Failure Mission Outcomes

- mission_outcome
- Failure (in flight)1
- Success99
- Success (payload status unclear)1

Booster carried

booster_version
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

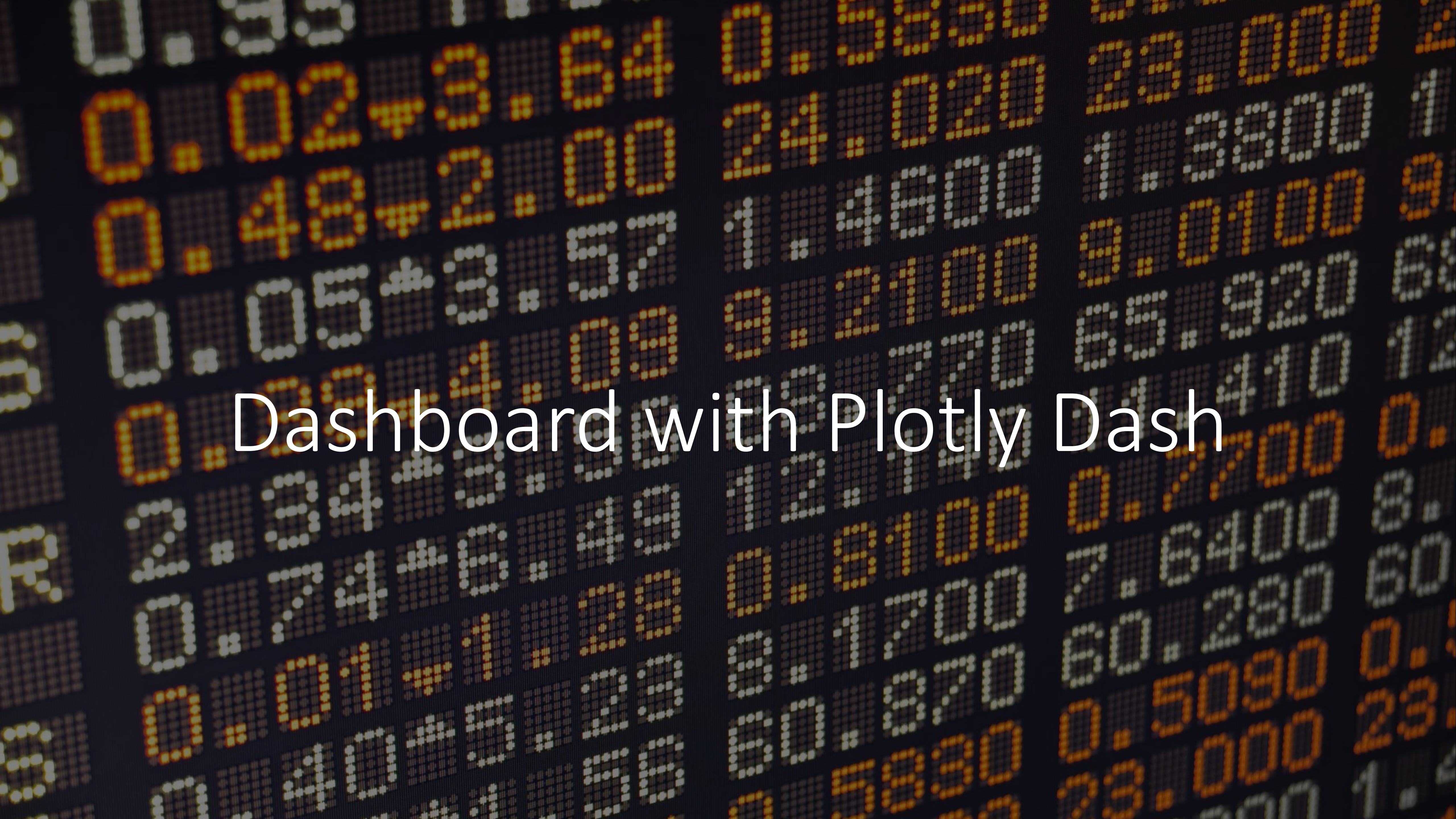
2015 Lunch Records

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

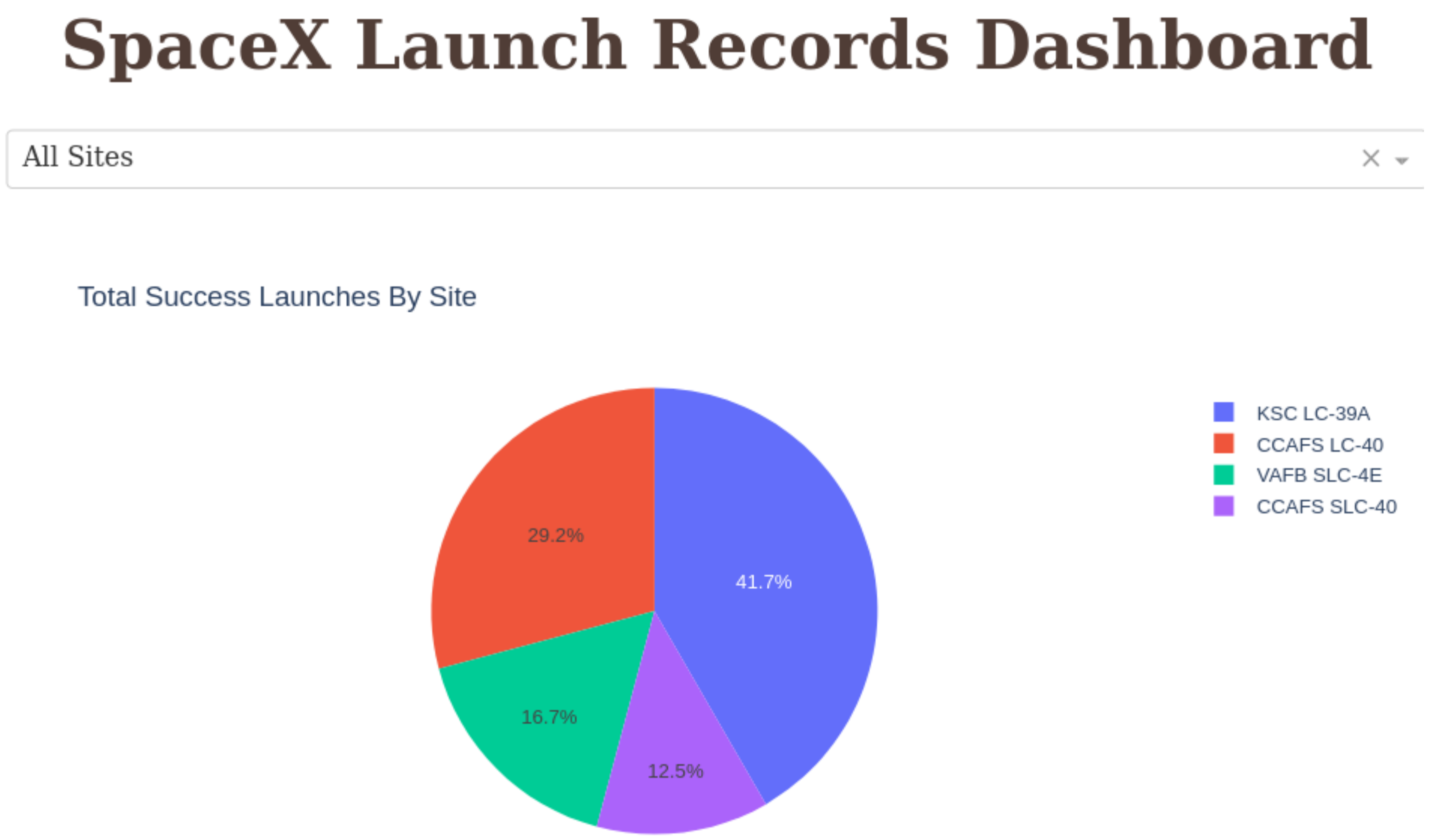
Rank Landing

landing__outcome	qty
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Dashboard with Plotly Dash

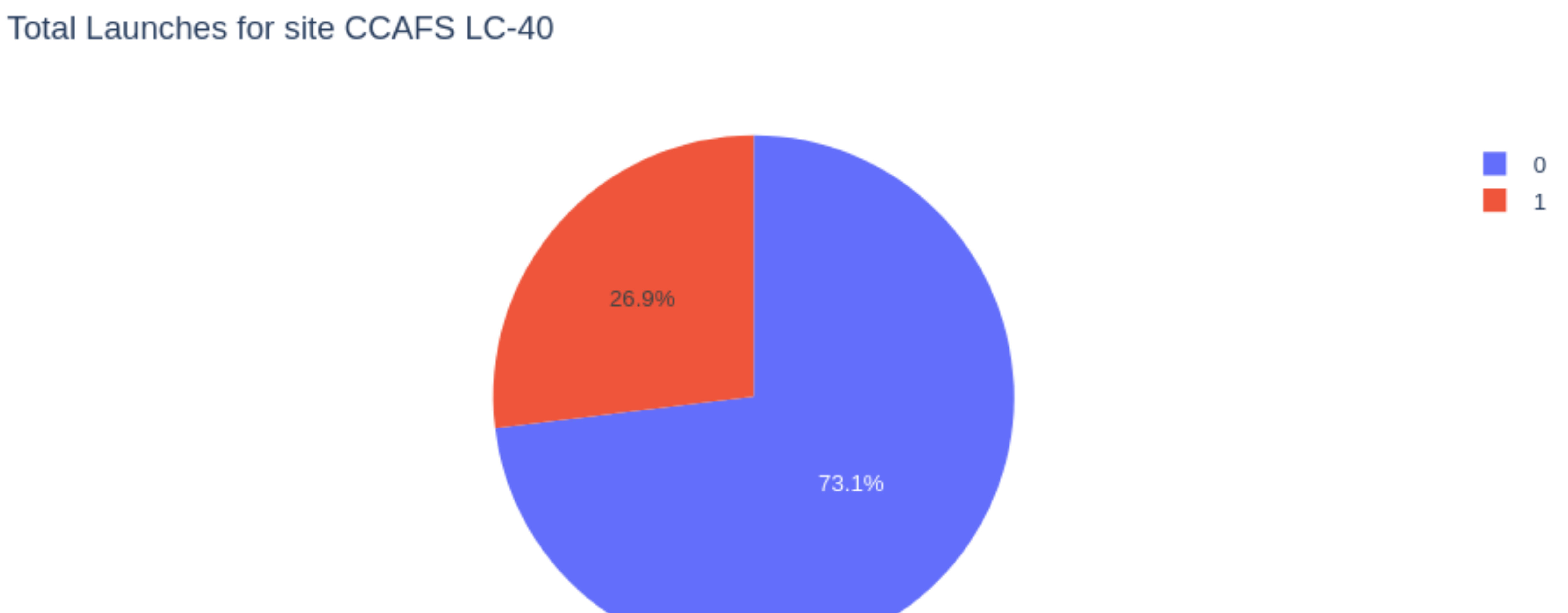


SpaceX Launch Records Dhashboard

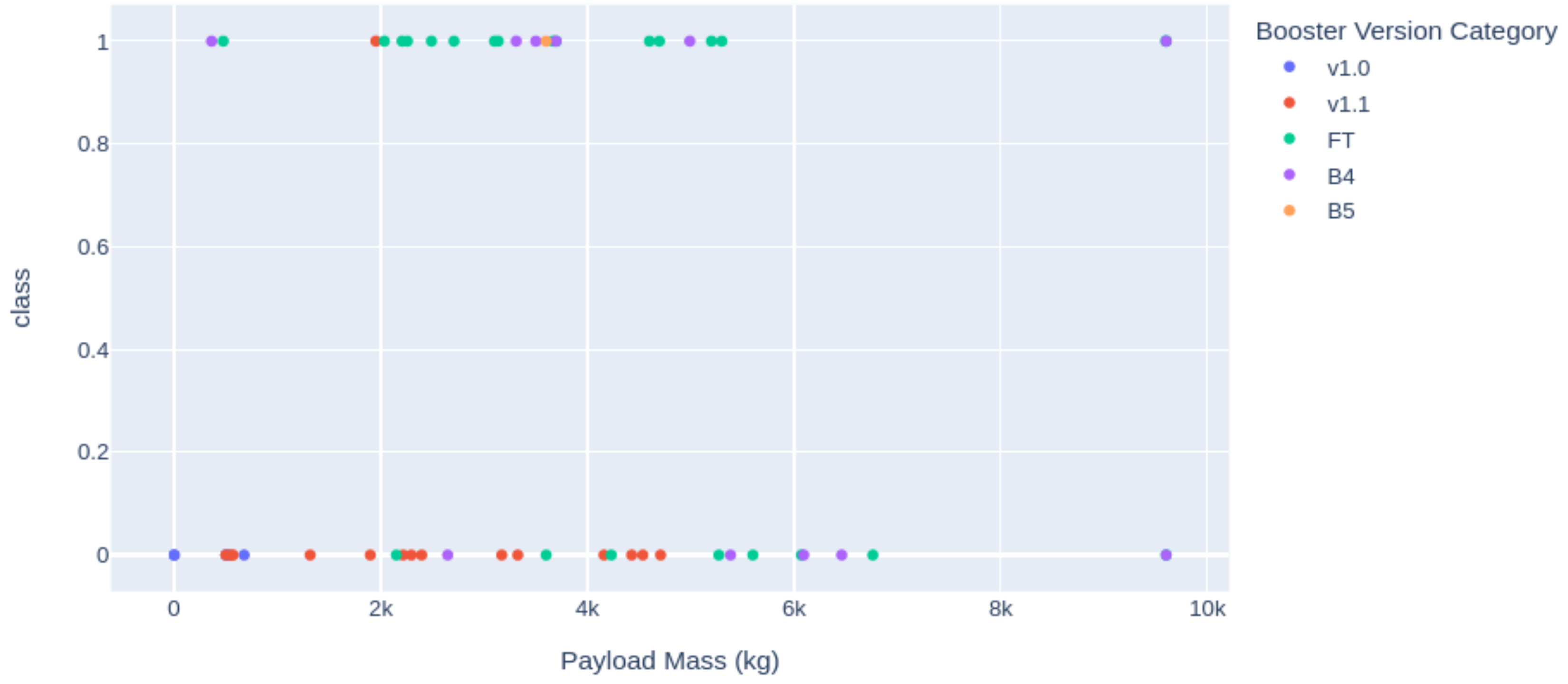


SpaceX Launch Records Dashboard

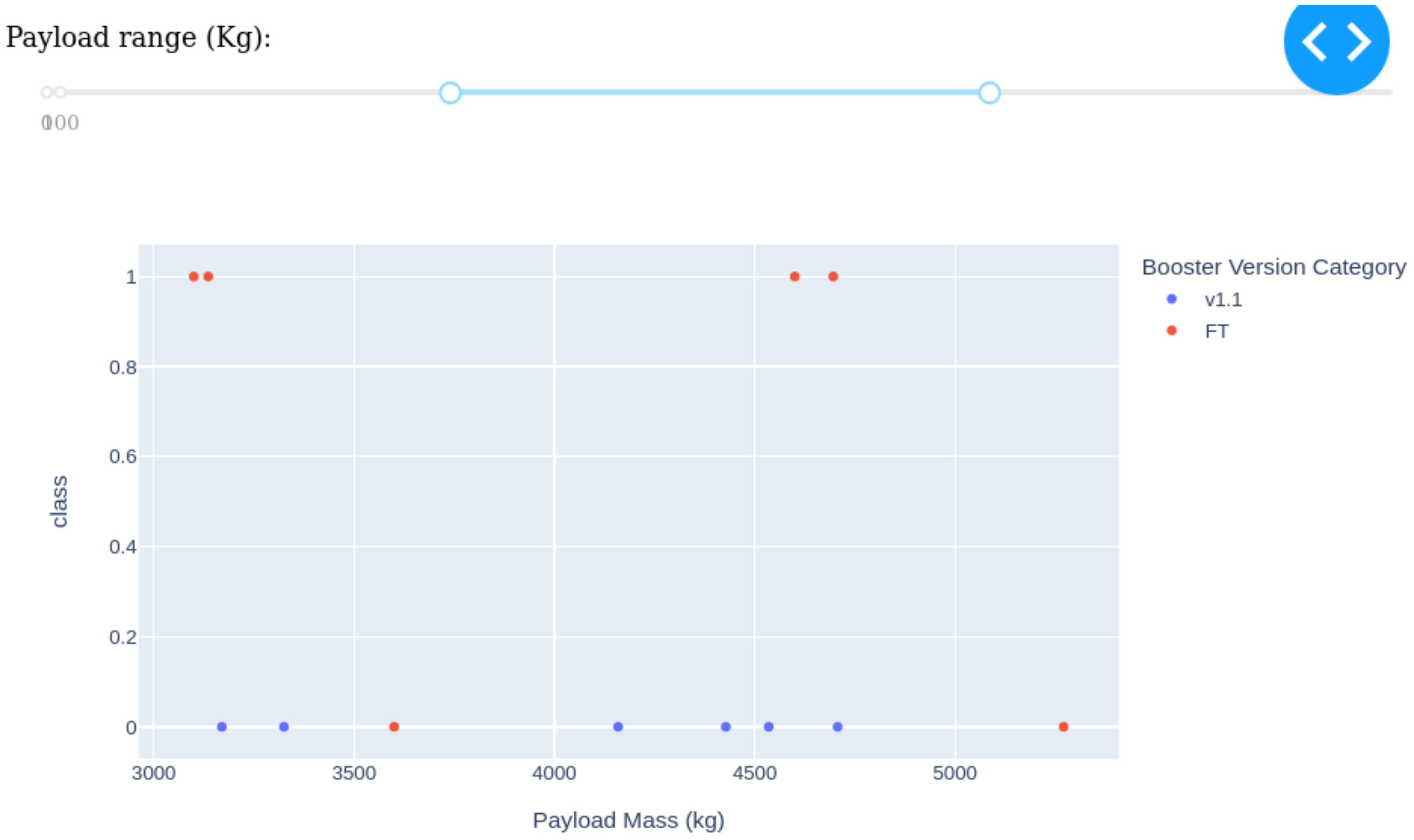
CCAFS LC-40



Payload range (Kg):

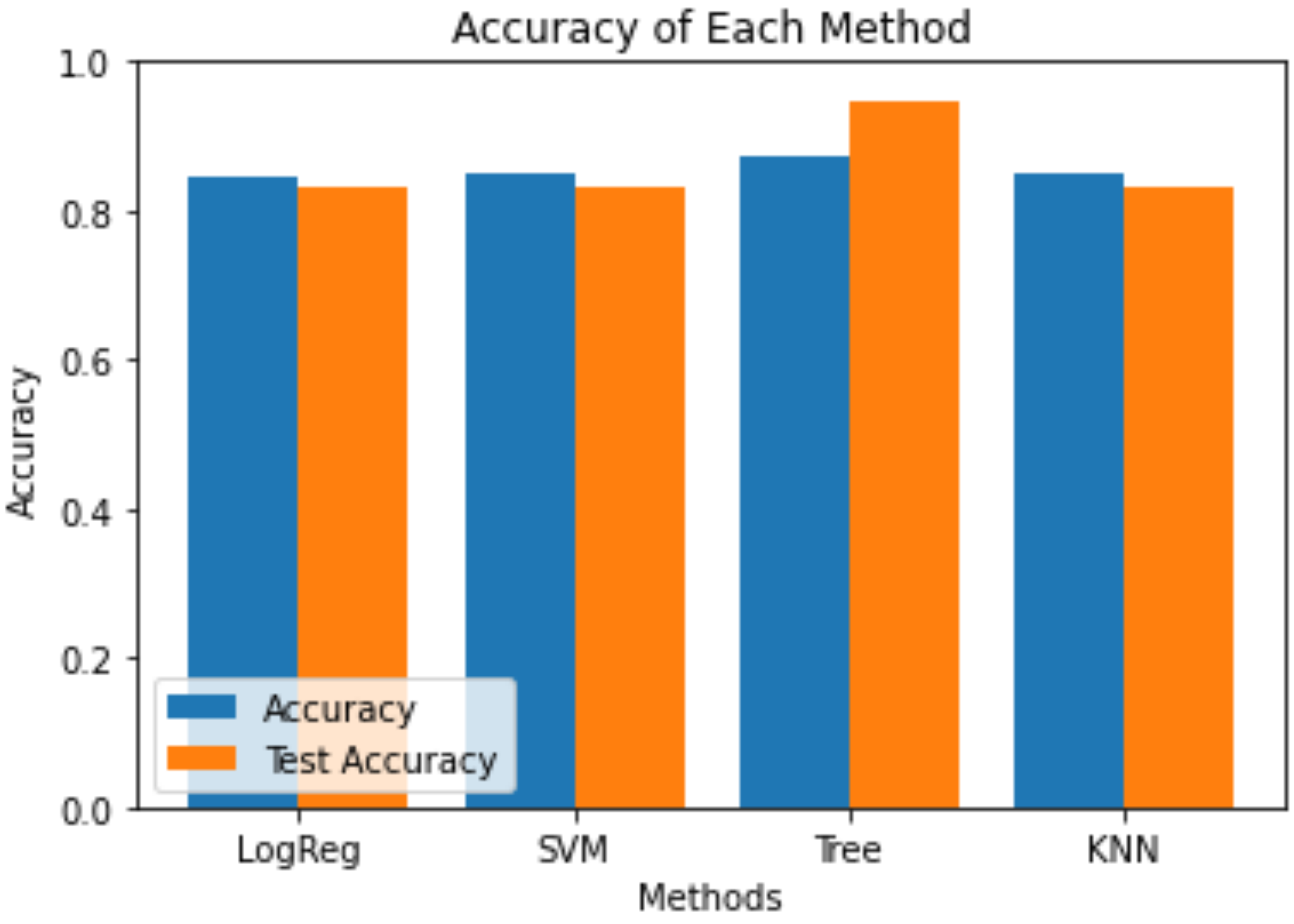


Payload vs launch Outcome

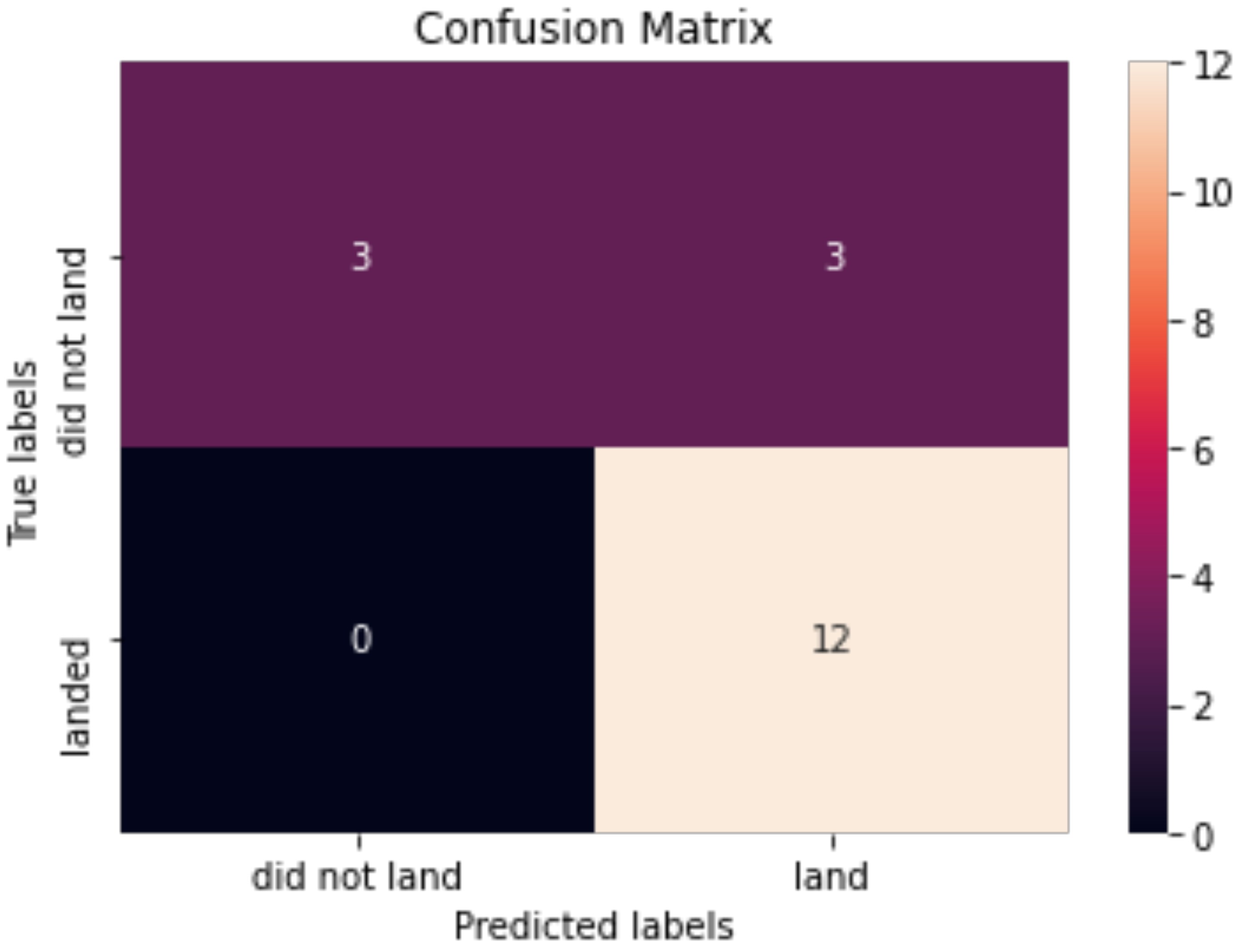


Predictive Analysis

Classification Accuracy



Confusion Matrix of Decision Tree Classfier



Conclusion

- Several data sources were analyzed and conclusions were refined throughout the process. After careful consideration, it was determined that KSC LC-39A is the best launch site.
- Additionally, launches above 7,000kg were found to be less risky. While most mission outcomes are successful, it seems that successful landing outcomes improve over time in accordance with the evolution of processes and rockets.
- It was also discovered that the Decision Tree Classifier can be utilized to predict successful landings, which can ultimately lead to increased profits.

THANK YOU

Let us connect on each other's LinkedIn for networking connections and improve our skills in Data Science

LinkedIn: https://www.linkedin.com/in/mohamed-2-mohamud/

GitHub: https://github.com/mmohamud25/IBM-Applied-Data-Science-Capstone