IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Oyedeji Olaoluwa Victor
25/09/2024

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

Methodologies

- **Data Collection:** Utilized SpaceX REST API and web scraping from Wikipedia to gather launch data.
- **Data Wrangling:** Cleaned and processed data, calculated launch statistics, and created new labels.
- **Exploratory Data Analysis (EDA):** Used visualization tools like Folium and Plotly Dash to analyze launch patterns and trends.
- **Predictive Analysis:** Built and tuned classification models to predict first stage reuse and evaluate model performance.

Results

- **EDA Findings:** Identified trends in launch sites, payloads, and success rates.
- **Predictive Analysis:** The decision tree classifier achieved the highest accuracy in predicting launch outcomes.
- **Interactive Analytics:** Developed dashboards to visualize launch success rates and payload relationships.
- **Launch Success Trends:** Noted an overall increase in SpaceX's launch success rates over time.

Introduction

Project Background and Context

The commercial space industry is booming, with companies like Virgin Galactic, Rocket Lab, Blue Origin, and SpaceX leading the charge. SpaceX, in particular, has achieved remarkable feats: delivering cargo to the International Space Station, launching Starlink satellites for global internet access, and sending humans into space. A key factor in their success is their relatively inexpensive launch costs. Their Falcon 9 rocket, advertised at $62 million, significantly undercuts competitors whose launches can cost upwards of $165 million. This cost advantage hinges on SpaceX's ability to reuse the first stage of the Falcon 9, dramatically reducing launch expenses.

Problems we want to find answers to

As data scientists working for a new rocket company, Space Y, we aim to compete with SpaceX in the commercial launch market. To achieve this, we need to determine the following:

- Pricing: Can we determine a competitive launch price for Space Y by analysing SpaceX's data?
- First Stage Reuse Prediction: Instead of relying solely on complex engineering calculations, can we train a machine learning model using publicly available data to predict whether SpaceX will attempt to reuse the first stage of their Falcon 9 rocket in a specific launch?

By answering these questions, we can develop effective strategies to compete with SpaceX and offer cost-efficient launch solutions for our clients.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

Our data collection process involved two primary methods:

1. **API Extraction:** We leveraged the SpaceX REST API to retrieve structured data on past launches.
2. **Web Scraping:** We scraped relevant information from SpaceX Wiki pages using Python's BeautifulSoup library.

API Extraction

Request to the SpaceX API → Clean the requested data

Web Scraping

Extract a Falcon 9 launch records HTML table from Wikipedia → Parse the table and convert it into a Pandas data frame

Data Collection – SpaceX API

Request and parse the SpaceX launch data using the GET request

- response.status_code
- data=pd.json_normalize(response.json())
- data.head()

Filter the dataframe to only include Falcon 9 launches

- data_falcon9=data[data['BoosterVersion']!='Falcon 1']
- data_falcon9.loc[:,'Flight Number']=list(range(1,data_falcon9.shape[0]+1))

GitHub URL: https://github.com/Ola-Vic22/IBM-Data-Science-Capstone-Project/blob/f0905856ba8df04b3e9594a62cc8c1d9c80f6340/jupyter-labs-spacex-data-collection-api-v2.ipynb

Data Collection - Scraping

Request the Falcon9 Launch Wiki page from its URL

- data=requests.get(static_url)
- soup=BeautifulSoup(data.text,'html.parser')
- soup.title

Extract all column/variable names from the HTML table header

- html_tables=soup.find_all('table')
- first_launch_table=html_tables
- print(first_launch_table)

Create a data frame by parsing the launch HTML tables

- df=pd.DataFrame({key: pd.Series(value)forkey,valueinlaunch_dict.items()})

GitHub URL: https://github.com/Ola-Vic22/IBM-Data-Science-Capstone-Project/blob/f0905856ba8df04b3e9594a62cc8c1d9c80f6340/jupyter-labs-webscraping.ipynb

Data Wrangling

Calculate the number of launches on each site

- df['LaunchSite'].value_counts()

Calculate the number and occurrence of each orbit

- df['Orbit'].value_counts()

Calculate the number and occurrence of mission outcome of the orbits

- landing_outcomes=df['Outcome'].value_counts()
- fori,outcomeinenumerate(landing_outcomes.keys()):print(i,outcome)
- bad_outcomes=set(landing_outcomes.keys()[[1,3,5,6,7]])bad_outcomes

Create a landing outcome label from Outcome column

- landing_class=df['Outcome'].apply(lambdax:0ifxinbad_outcomeselse1)
- df['Class']=landing_classdf[['Class']].head(8)

GitHub URL: https://github.com/Ola-Vic22/IBM-Data-Science-Capstone-Project/blob/f0905856ba8df04b3e9594a62cc8c1d9c80f6340/labs-jupyter-spacex-Data%20wrangling-v2.ipynb

EDA with Data Visualization

Scatterplots:

- **Flight Number vs. Launch Site:** This chart shows how the launch site changes over time. It can identify patterns in launch site selection and potential trends.
- **Payload vs. Launch Site:** This chart shows the relationship between the weight of the payload and the launch site. It can be used to determine if certain launch sites are more suitable for heavier payloads.
- **Flight Number vs. Orbit Type:** This chart shows how the orbit type changes over time. It can identify trends in orbit selection and potential relationships with other factors.
- **Payload vs. Orbit Type:** This chart shows the relationship between the weight of the payload and the orbit type. It can be used to determine if certain orbit types are more suitable for heavier payloads.

Bar Chart:

- **Success Rate vs. Orbit Type:** This chart compares the success rates of different orbit types. It can identify which orbit types are more challenging or risky.

Lineplot:

- **Launch Success Yearly Trend:** This chart shows the success rate of launches over time. It can identify trends in launch success and potential improvements or challenges.

GitHub URL: https://github.com/Ola-Vic22/IBM-Data-Science-Capstone-Project/blob/f0905856ba8df04b3e9594a62cc8c1d9c80f6340/jupyter-labs-eda-dataviz-v2.ipynb

EDA with SQL

SQL Queries Performed:

- **Retrieve distinct launch sites:** SELECT DISTINCT Launch_Site FROM SPACEXTBL
- **Retrieve initial rows from CCA launch site:** SELECT * FROM SPACEXTBL WHERE Launch_Site LIKE 'CCA%' LIMIT 5
- **Calculate total payload mass for NASA (CRS) customers:** SELECT SUM(PAYLOAD_MASS__KG_) AS Total_Payload FROM SPACEXTBL WHERE Customer LIKE 'NASA (CRS)'
- **Calculate average payload mass for F9 v1.1 booster version:** SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE Booster_Version LIKE 'F9 v1.1'
- **Retrieve the earliest launch date:** SELECT MIN(Date) FROM SPACEXTBL
- **Retrieve booster versions for payloads between 4000 and 6000 kg:** SELECT DISTINCT Booster_Version FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000 AND Mission_Outcome LIKE 'Success'
- **Count mission outcomes:** SELECT Mission_Outcome, COUNT(Mission_Outcome) AS Total_Outcomes FROM SPACEXTBL GROUP BY Mission_Outcome
- **Retrieve booster version with maximum payload mass:** SELECT Booster_Version FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL)
- **Retrieve launch data for failed drone ship landings in 2015:** SELECT SUBSTR(Date, 6, 2) AS month, Landing_Outcome, Booster_Version, Launch_Site FROM SPACEXTBL WHERE SUBSTR(Date, 0, 5) = '2015' AND Landing_Outcome = 'Failure (drone ship)'
- **Count landing outcomes within a specific date range:** SELECT Landing_Outcome, COUNT(*) AS Outcome_Count FROM SPACEXTBL WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome ORDER BY Outcome_Count DESC

GitHub URL: https://github.com/Ola-Vic22/IBM-Data-Science-Capstone-Project/blob/ff118da9a45e40975aefedd003c964e8c9737926/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

The map objects I created and added to the folium map are:

- **folium.Circle and folium.Marker:** These were used to represent launch sites on the map.
  - **Circle:** Used to highlight the launch site location with a defined area.
  - **Marker:** Used to pinpoint the exact launch site location within the highlighted circle.
- **folium.Marker (within marker_cluster):** These were used to represent each launch from the spacex_df dataframe. This likely created a cluster of markers for all launches, potentially overwhelming the map if there were many launches.
- **MousePosition:** This tool allows the user to interact with the map by displaying the coordinates of any point they hover over. This helps users explore the map and find specific locations (like railways) without needing to know their exact coordinates beforehand.
- **PolyLine:** This was used to draw a line between a specific launch site and a chosen point on the coastline. This likely aimed to visualize the launch trajectory or path towards the target location.

Reasons for adding these objects:

- **Markers and Circles:** These provide a clear visual representation of launch sites on the map, helping users understand their geographical distribution.
- **Marker cluster:** While helpful for large datasets, it might be better to implement filtering or zoom-dependent marker display to avoid overwhelming the map with too many markers.
- **MousePosition:** This enhances user interactivity by allowing them to explore the map and find specific locations of interest.
- **PolyLine:** This helps visualize launch trajectories or paths from launch sites to target locations.

GitHub URL: https://github.com/Ola-Vic22/IBM-Data-Science-Capstone-Project/blob/f0905856ba8df04b3e9594a62cc8c1d9c80f6340/lab-jupyter-launch-site-location-v2.ipynb

Build a Dashboard with Plotly Dash

	Description	Function in Dashboard
Pie Chart	This chart shows the success rate of launches from different launch sites. The user can select a launch site from a dropdown menu, and the chart will update to display the success rate for that specific site.	This chart provides a clear and concise overview of launch success rates for different launch sites. The interactivity allows users to focus on specific sites of interest and compare their performance.
Scatter Plot	This chart shows the relationship between payload mass and launch success. The user can adjust a range slider to filter the data based on payload mass, and the chart will update to display the scatter plot for the selected payload range.	This chart helps visualize the relationship between payload mass and launch success. The interactive range slider enables users to explore how launch success varies across different payload ranges.

GitHub URL: https://github.com/Ola-Vic22/IBM-Data-Science-Capstone-Project/blob/f0905856ba8df04b3e9594a62cc8c1d9c80f6340/Interactive-dashboard-with-plotly-dash.ipynb

Predictive Analysis (Classification)

Data Preparation

Feature Engineering: Create a new column representing the target class (dependent variable).

Standardization: Standardize the features (independent variables) using StandardScaler to ensure all features are on a similar scale.

Splitting Data: Split the data into training and testing sets using train_test_split to train the model and evaluate its performance on unseen data.

Model Selection and Hyperparameter Tuning

Model Selection: Choose multiple classification models to evaluate, such as Logistic Regression, Support Vector Machine (SVM), etc.

Hyperparameter Tuning (for each model)

Grid Search: Define a dictionary containing potential values for different hyperparameters of the chosen model.

GridSearchCV: Use GridSearchCV to efficiently search for the best hyperparameter combination within the defined grid.

Fitting: Fit the model with the hyperparameter combinations on the training data.

Evaluation: Evaluate the performance of each hyperparameter combination using cross-validation (cv=10 in this case) and select the one with the best score (e.g., accuracy).

Model Evaluation

Test Accuracy: Calculate the accuracy of the final model with the best hyperparameters on the held-out test data using the score method.

Confusion Matrix: Generate a confusion matrix to visualize the model's performance in terms of True Positives, False Positives, True Negatives, and False Negatives.

Model Selection (among all models)

Compare Results: Compare the test accuracy and other relevant metrics (e.g., F1-score) achieved by each tuned model.

Best Performing Model: Choose the model with the best overall performance on the test data as your final classification model.

GitHub URL: https://github.com/Ola-Vic22/IBM-Data-Science-Capstone-Project/blob/f0905856ba8df04b3e9594a62cc8c1d9c80f6340/SpaceX-Machine-Learning-Prediction-Part-5-v1.ipynb

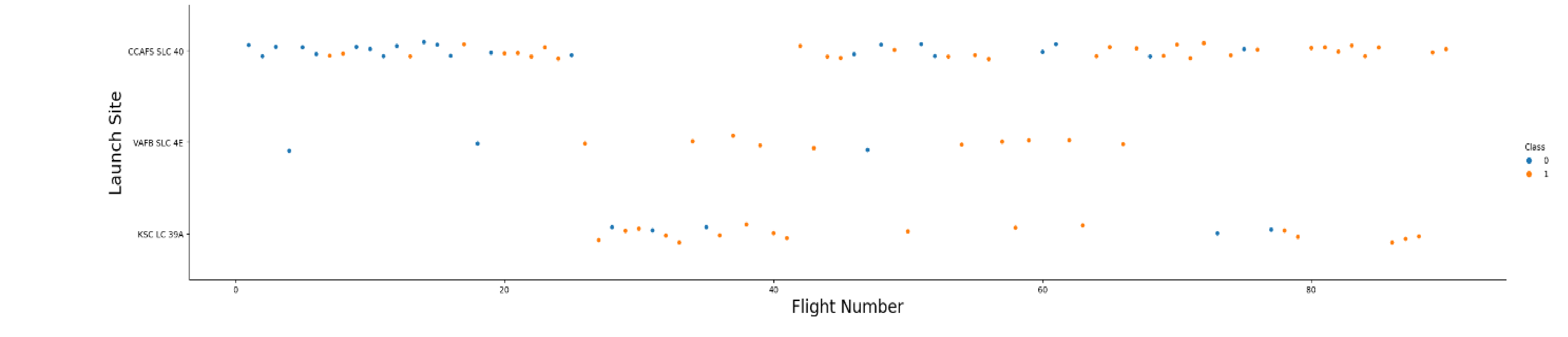
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site



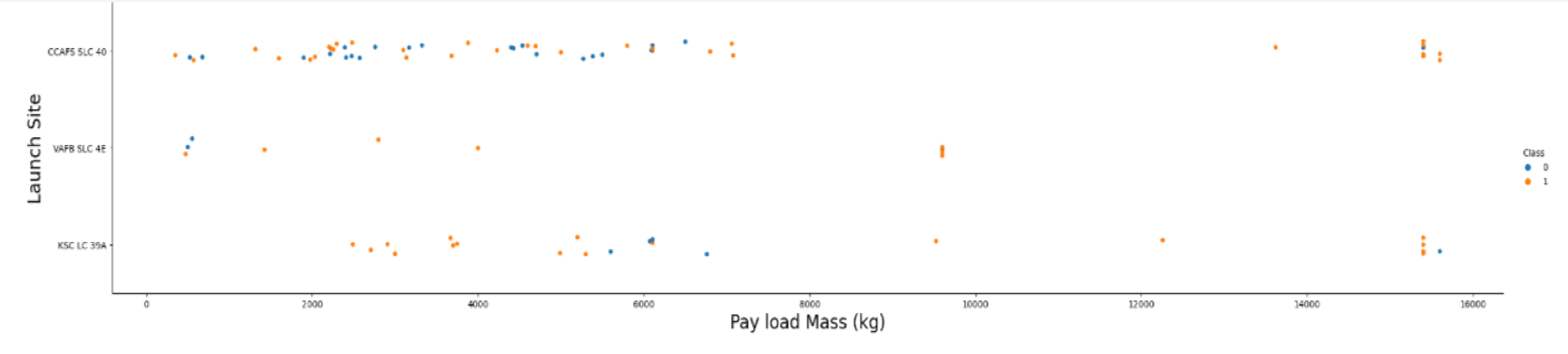
The scatter plot shows the relationship between SpaceX launch site and flight number. Each dot in the plot represents a single launch, with the x-axis representing the flight number and the y-axis representing the launch site. The different colors of the dots indicate the outcome of the launch, with blue representing successful launches and orange representing failed launches.

From the plot, we can observe the following:

- **Launch Frequency:** The number of launches per launch site appears to be relatively consistent, with a few clusters of launches at specific flight numbers.
- **Success Rates:** While the plot does not explicitly show success rates, it can be inferred that some launch sites have a higher proportion of successful launches than others based on the number of blue dots compared to orange dots.
- **Launch Trends:** There seems to be a general trend of increasing flight numbers over time, suggesting that SpaceX has been launching more frequently in recent years.

Overall, the scatter plot provides a visual representation of SpaceX's launch history, showing the distribution of launches across different sites and the outcomes of those launches.

Payload vs. Launch Site



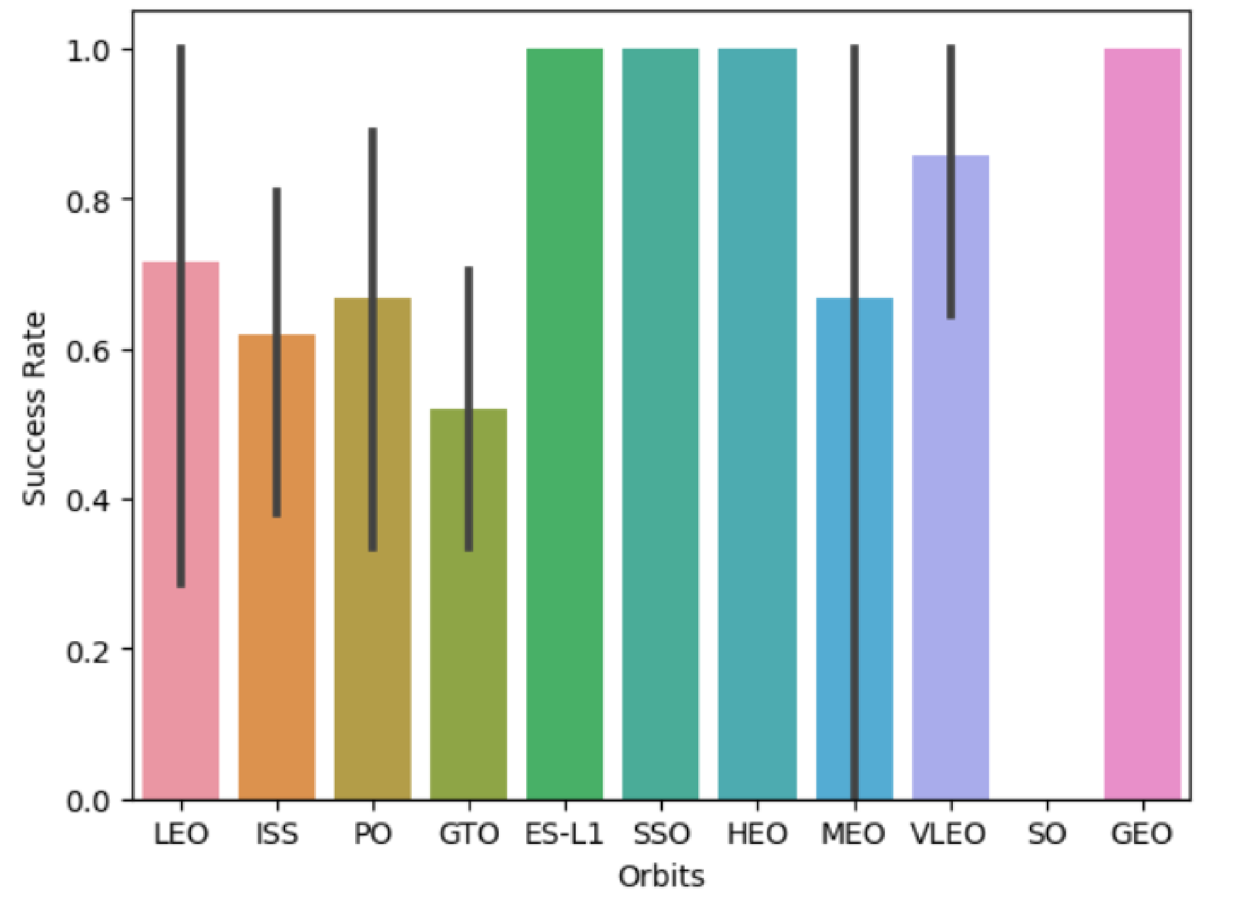
The scatterplot shows the relationship between the launch sites and payload mass for SpaceX launches. Each dot in the plot represents a single launch, with the x-axis representing the payload mass in kilograms and the y-axis representing the launch site. The different colors of the dots indicate the outcome of the launch, with blue representing successful launches and orange representing failed launches.

From the plot, we can observe the following:

- **Payload Mass Distribution:** The payload mass for launches varies significantly across different sites. Some sites, such as CCAFS SLC-4N, have a wider range of payload masses compared to others, such as VAFB SLC-4E.
- **Launch Success:** The plot does not explicitly show success rates, but it can be inferred that some launch sites have a higher proportion of successful launches with heavier payloads compared to others. For example, CCAFS SLC-4N seems to have a higher concentration of blue dots (successful launches) in the higher payload mass range.
- **Launch Trends:** There appears to be a general trend of increasing payload masses over time, especially for CCAFS SLC-4N, suggesting that SpaceX has been launching heavier payloads in recent years.

Overall, the scatterplot provides a visual representation of SpaceX's launch history, showing the distribution of payload masses across different sites and the outcomes of those launches.

Success Rate vs. Orbit Type



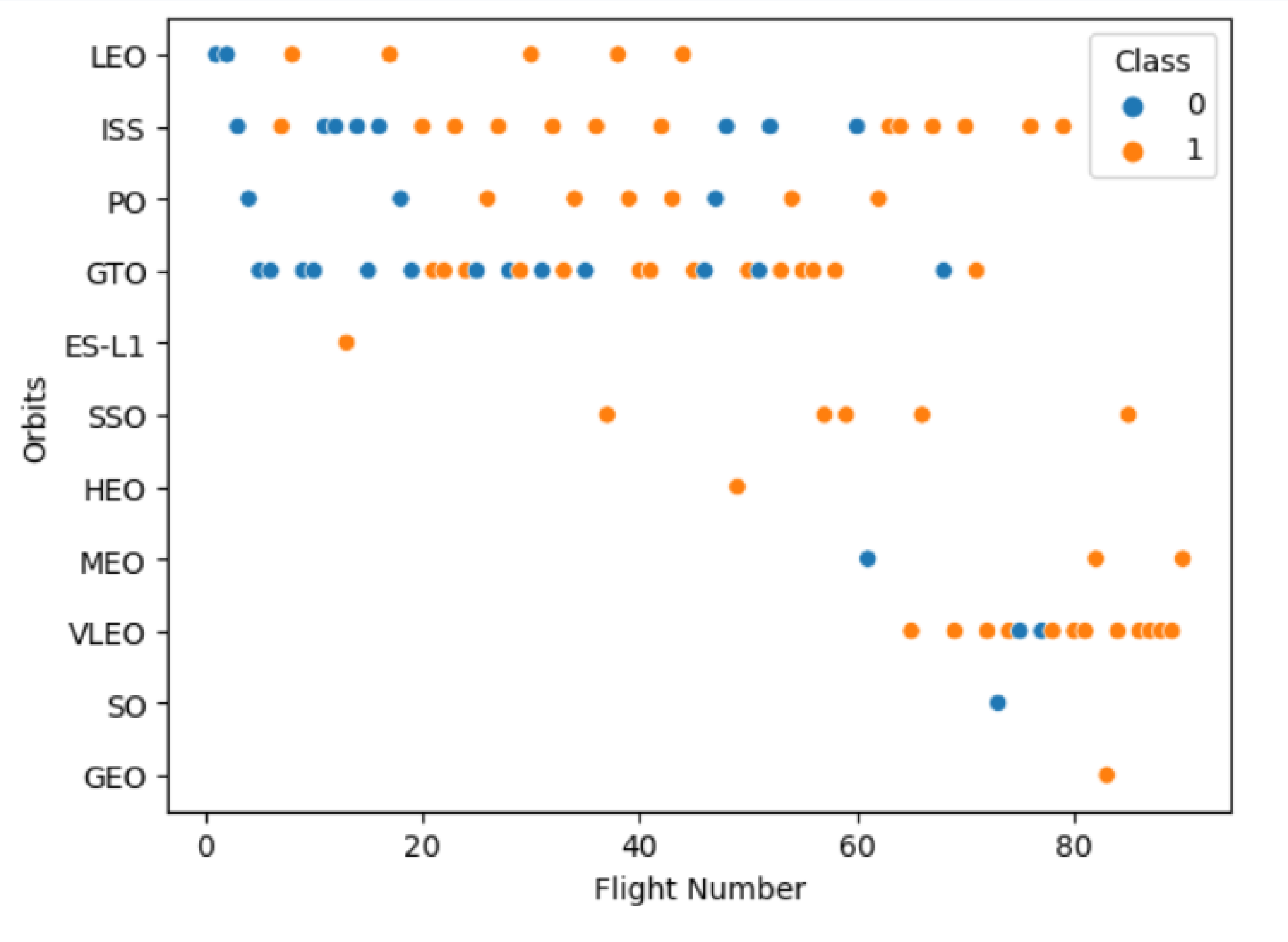
The barplot shows the success rate of SpaceX launches for different orbits. The x-axis represents the different orbits, while the y-axis represents the success rate, ranging from 0 to 1. Each bar represents the average success rate for a particular orbit, with error bars indicating the variability in success rates.

From the plot, we can observe the following:

- **Orbit Success Rates:** The success rates for different orbits vary significantly. Some orbits, such as GEO, have very high success rates close to 1, while others, such as PO, have lower success rates around 0.6.
- **Orbit Variability:** The error bars for some orbits are relatively large, indicating a significant amount of variability in success rates within those orbits. Other orbits like LEO and GEO have smaller error bars, suggesting more consistent success rates.
- **Orbit Trends:** There is no clear trend between orbit type and success rate. Some orbits with higher altitudes have lower success rates, while others with lower altitudes have higher success rates.

Overall, the barplot provides a visual representation of SpaceX's launch success rates for different orbits, highlighting the variability in success rates across different orbit types.

Flight Number vs. Orbit Type



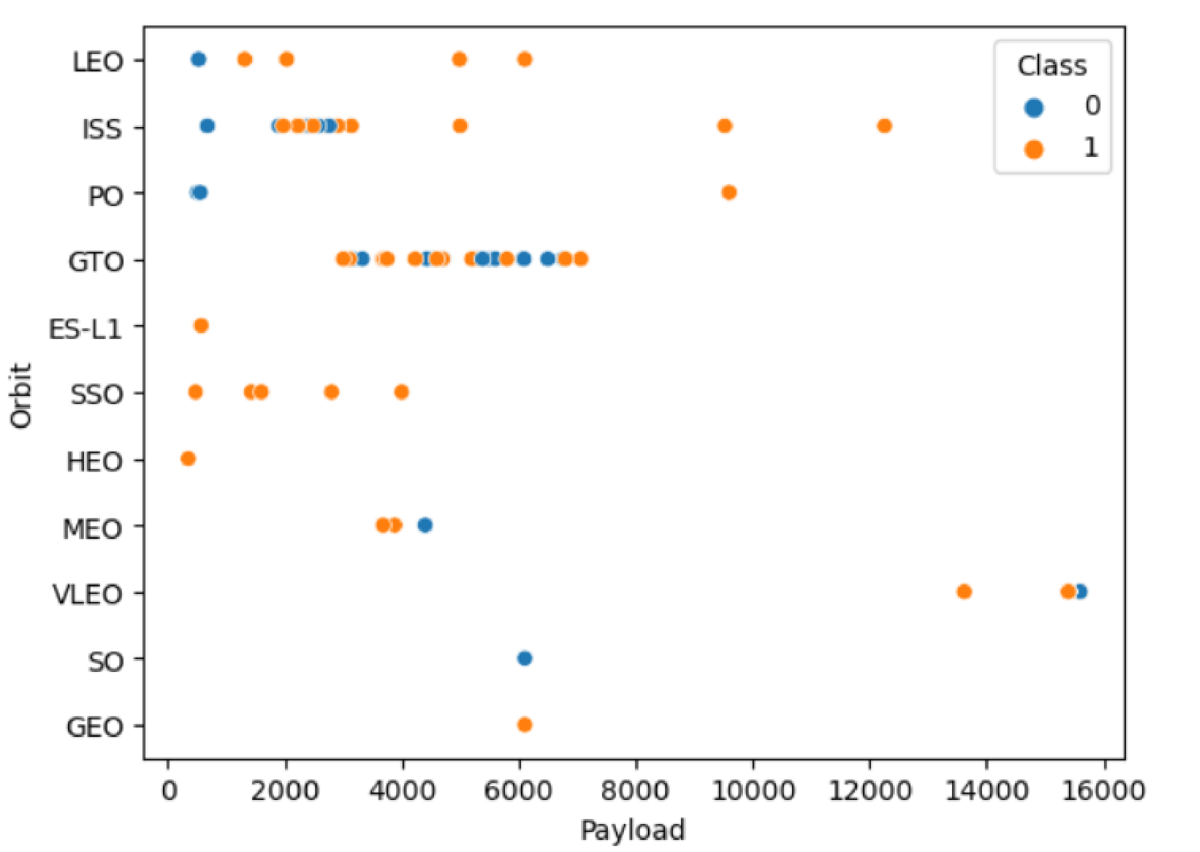
The scatterplot shows the relationship between orbit and flight number for SpaceX launches, with different colors representing the class of the launch.

Here's a breakdown of what we can observe from the plot:

- **Orbit Distribution:** The plot shows that SpaceX has launched missions to a variety of orbits, with some orbits, such as LEO and GTO, having a higher number of launches compared to others.
- **Flight Number Progression:** The x-axis shows the flight number, indicating the chronological order of the launches. We can see that SpaceX has been launching missions to different orbits over time.
- **Class Differentiation:** The different colors of the dots represent the class of the launch. While the specific meaning of the classes is not provided, it's likely that they represent different types of missions or payloads.
- **Orbit-Specific Trends:** Some orbits may show specific patterns or trends in terms of the class of launches. For example, we might observe that certain orbits are primarily used for commercial missions (one class), while others are used for government or scientific missions (another class).

Overall, the scatterplot provides a visual representation of SpaceX's launch history, showing the distribution of launches across different orbits and the types of missions associated with each orbit.

Payload vs. Orbit Type



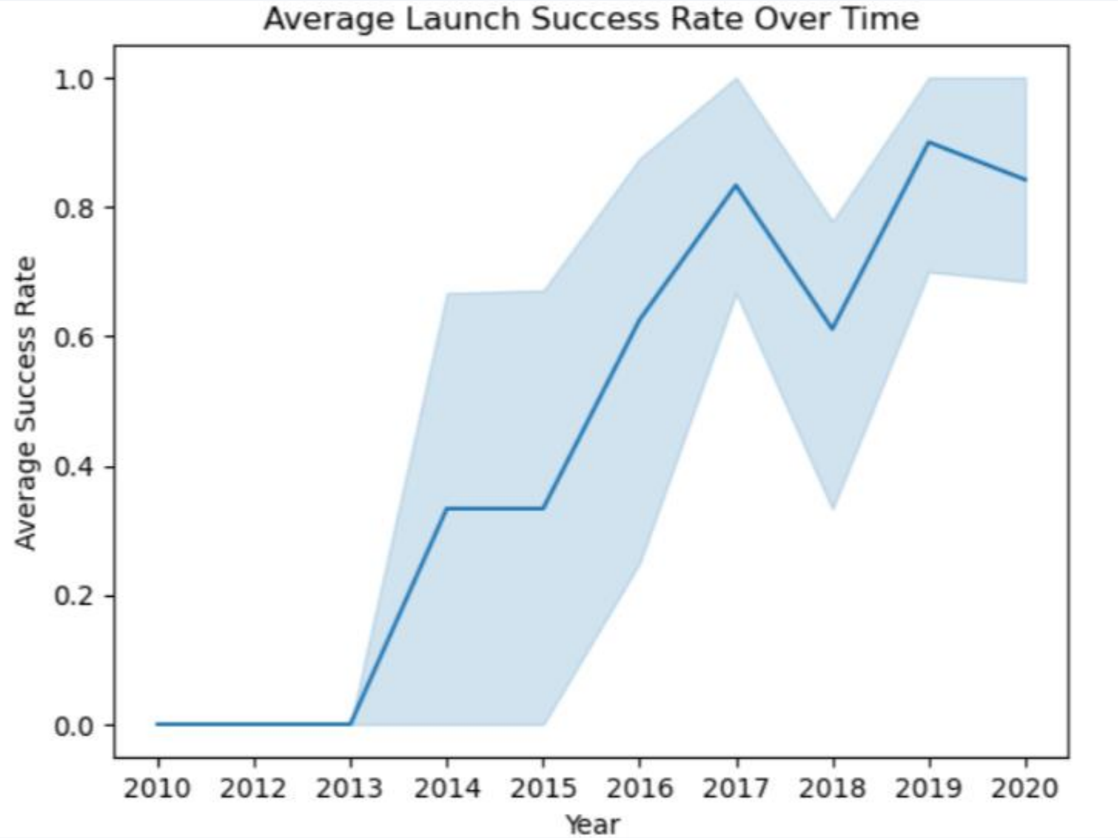
The scatterplot shows the relationship between orbit and payload mass for SpaceX launches, with different colors representing the class of the launch.

Here's a breakdown of what we can observe from the plot:

- **Orbit Distribution:** The plot shows that SpaceX has launched missions to a variety of orbits, with some orbits, such as LEO and GTO, having a higher number of launches compared to others.
- **Payload Mass Range:** The x-axis represents the payload mass, and we can see that there's a wide range of payload masses for different launches. Some orbits, like GEO, have a higher concentration of launches with heavier payloads.
- **Class Differentiation:** The different colors of the dots represent the class of the launch. While the specific meaning of the classes is not provided, it's likely that they represent different types of missions or payloads.
- **Orbit-Specific Trends:** Some orbits may show specific patterns or trends in terms of the class of launches and payload mass. For example, we might observe that certain orbits are primarily used for commercial missions (one class) with heavier payloads, while others are used for government or scientific missions (another class) with lighter payloads.

Overall, the scatterplot provides a visual representation of SpaceX's launch history, showing the distribution of payload masses across different orbits and the types of missions associated with each orbit.

Launch Success Yearly Trend



The lineplot shows the average launch success rate for SpaceX over time. The x-axis represents the year, while the y-axis represents the average success rate, ranging from 0 to 1. The blue line shows the trend in average success rate over the years, while the shaded area represents the confidence interval or standard deviation around the average.

From the plot, we can observe the following:

- **Increasing Success Rate:** Overall, the trend of the line shows a significant increase in the average launch success rate over time. This indicates that SpaceX has been improving its launch technology and procedures.
- **Variability:** The shaded area around the line shows that there is some variability in success rates from year to year. While the overall trend is upward, there are years with slightly lower or higher success rates.
- **Specific Years:** We can identify specific years with particularly high or low success rates by looking at the peaks and troughs of the line. For example, the year 2017 seems to have a relatively high success rate, while the year 2014 seems to have a lower success rate.

Overall, the lineplot provides a visual representation of SpaceX's progress in improving launch reliability over time.

All Launch Site Names

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

Below is the list of unique launch sites:

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

Launch Site Names Begin with 'CCA'

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
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	Success	Failure (parachute)
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

The provided data includes information on SpaceX launches, including the date, time, booster version, launch site, payload, payload mass, orbit, customer, mission outcome, and landing outcome. We can use this data to analyse various aspects of SpaceX's launch history

Total Payload Mass

Total_Payload
45596

The total payload carried by boosters from NASA is **45596** kilograms.

Average Payload Mass by F9 v1.1

avg(PAYLOAD_MASS__KG_)
2928.4

The average payload mass carried by booster version F9 v1.1 is **2928.4** kilograms.

First Successful Ground Landing Date

min(Date)
2010-06-04

The date of the first successful landing outcome on the ground pad is **2010-06-04**. This marked a significant milestone for SpaceX as it demonstrated the feasibility of reusable rockets, which could potentially reduce the cost of spaceflight.

Successful Drone Ship Landing with Payload between 4000 and 6000

The following boosters have successfully landed on a drone ship and had a payload mass greater than 4000 but less than 6000 kilograms:

- F9 B5 B1056.3
- F9 B5 B1049.4
- F9 B5 B1046.4
- F9 B5 B1051.3
- F9 B5 B1056.4
- F9 B5 B1048.5
- F9 B5 B1051.4
- F9 B5B1058.1
- F9 B5 B1049.5
- F9 B5 B1059.3
- F9 B5 B1051.5
- F9 B5 B1049.6
- F9 B5 B1060.2
- F9 B5 B1058.3
- F9 B5 B1051.6
- F9 B5 B1060.3
- F9 B5B1061.1
- F9 B5 B1049.7

These boosters were all part of SpaceX's Falcon 9 rocket family and were used for various missions, including satellite launches and resupply missions to the International Space Station.

Total Number of Successful and Failure Mission Outcomes

Mission_Outcome	Total_Outcomes
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Of the mission outcomes listed, there were 99 successful missions and 1 failed mission.

Boosters Carried Maximum Payload

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

The following boosters have carried the maximum payload mass:

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

These boosters were all part of SpaceX's Falcon 9 rocket family and were used for various missions, including satellite launches and resupply missions to the International Space Station.

2015 Launch Records

The following table lists the failed landing outcomes on drone ships, their booster versions, and launch site names in the year 2015:

Month	Landing Outcome	Booster Version	Launch Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

These were two of the early attempts by SpaceX to land a rocket booster on a drone ship. Both attempts were unsuccessful, but these failures were crucial in developing the technology that would eventually lead to successful drone ship landings.

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

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

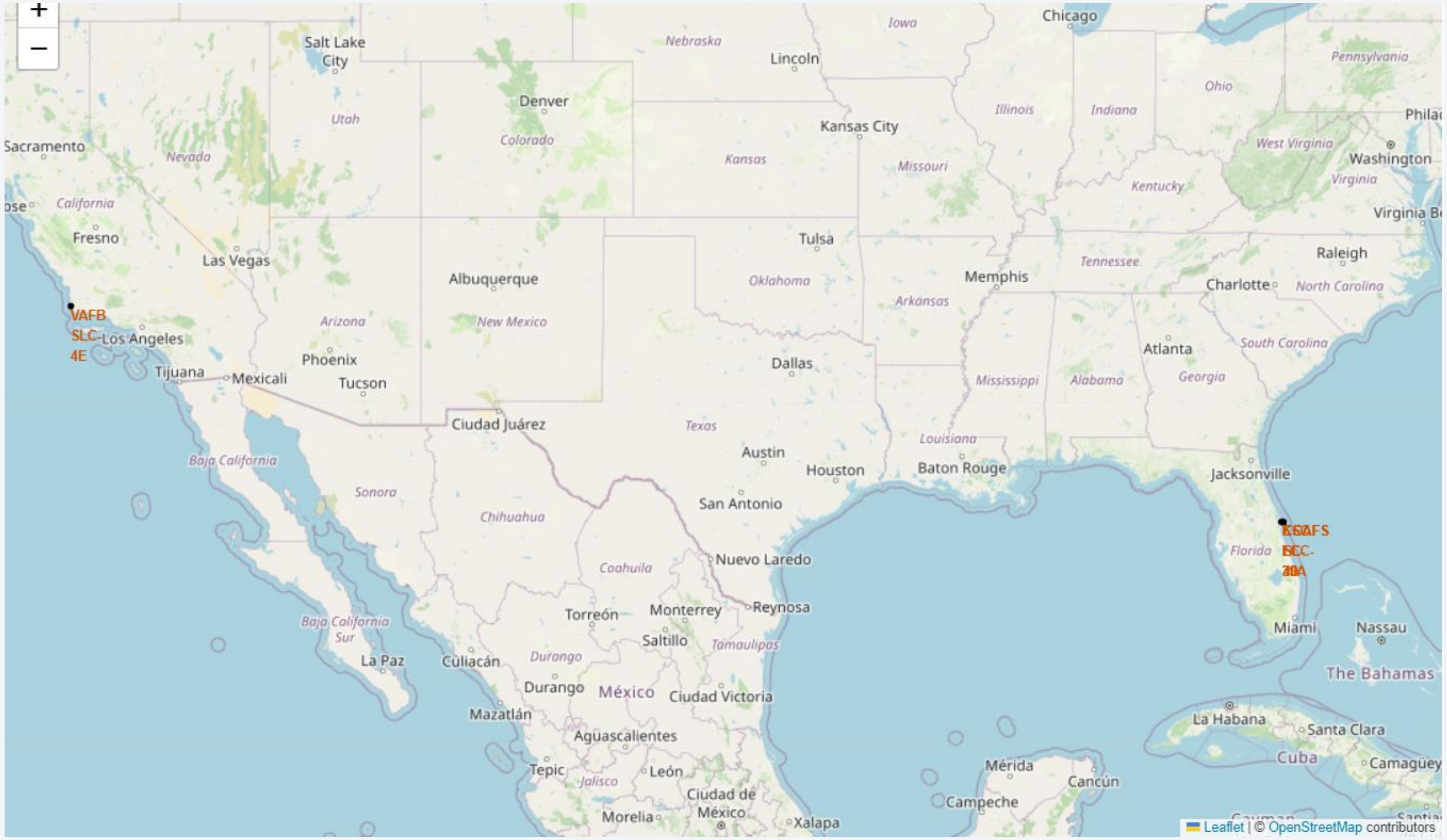
The table ranks the count of landing outcomes between the date 2010-06-04 and 2017-03-20 in descending order.

This data shows that during the time period specified, the most common landing outcome was "No attempt," followed by "Success (drone ship)" and "Failure (drone ship)."

Section 3

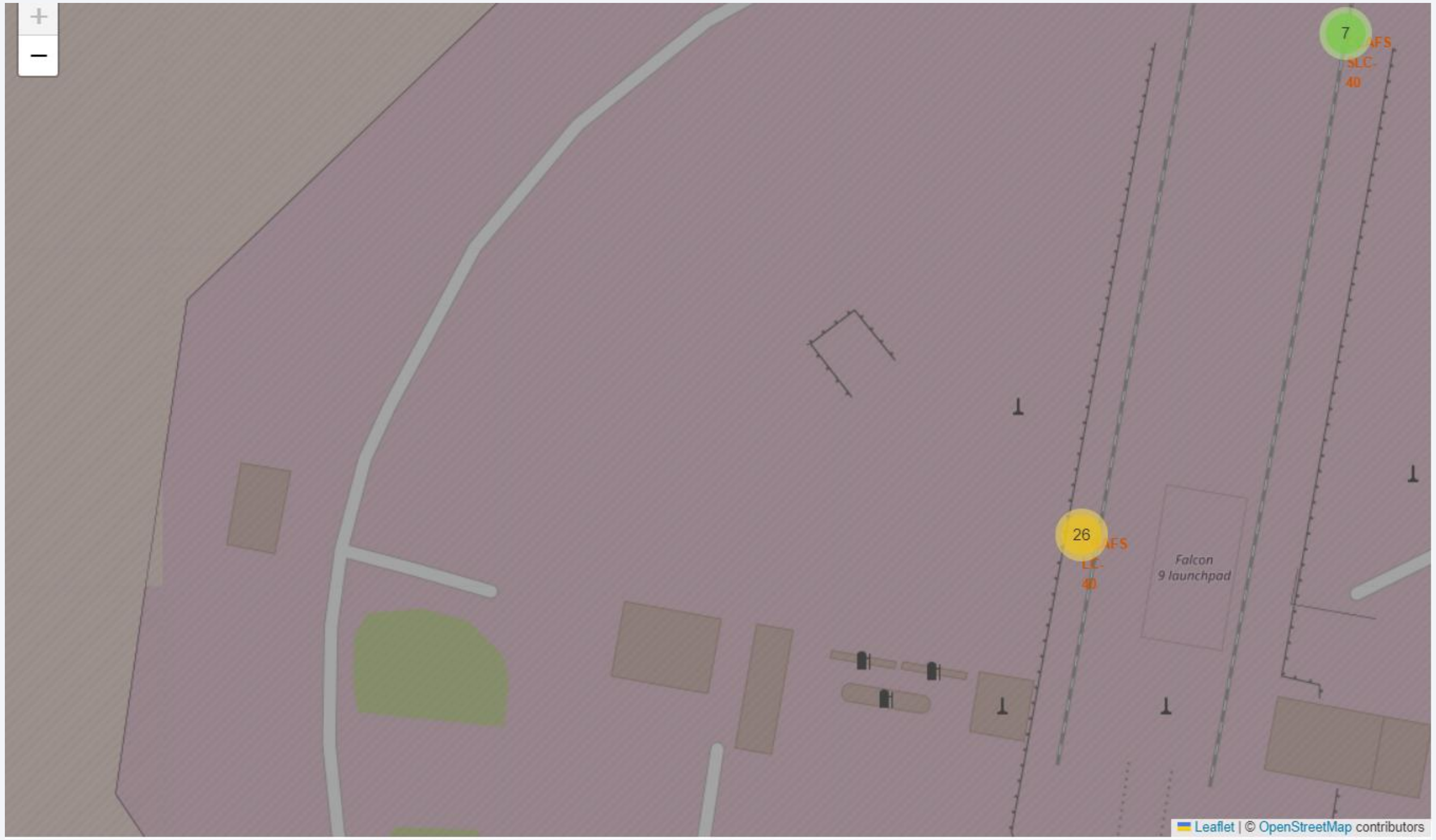
Launch Sites Proximities Analysis

Launch Site Locations



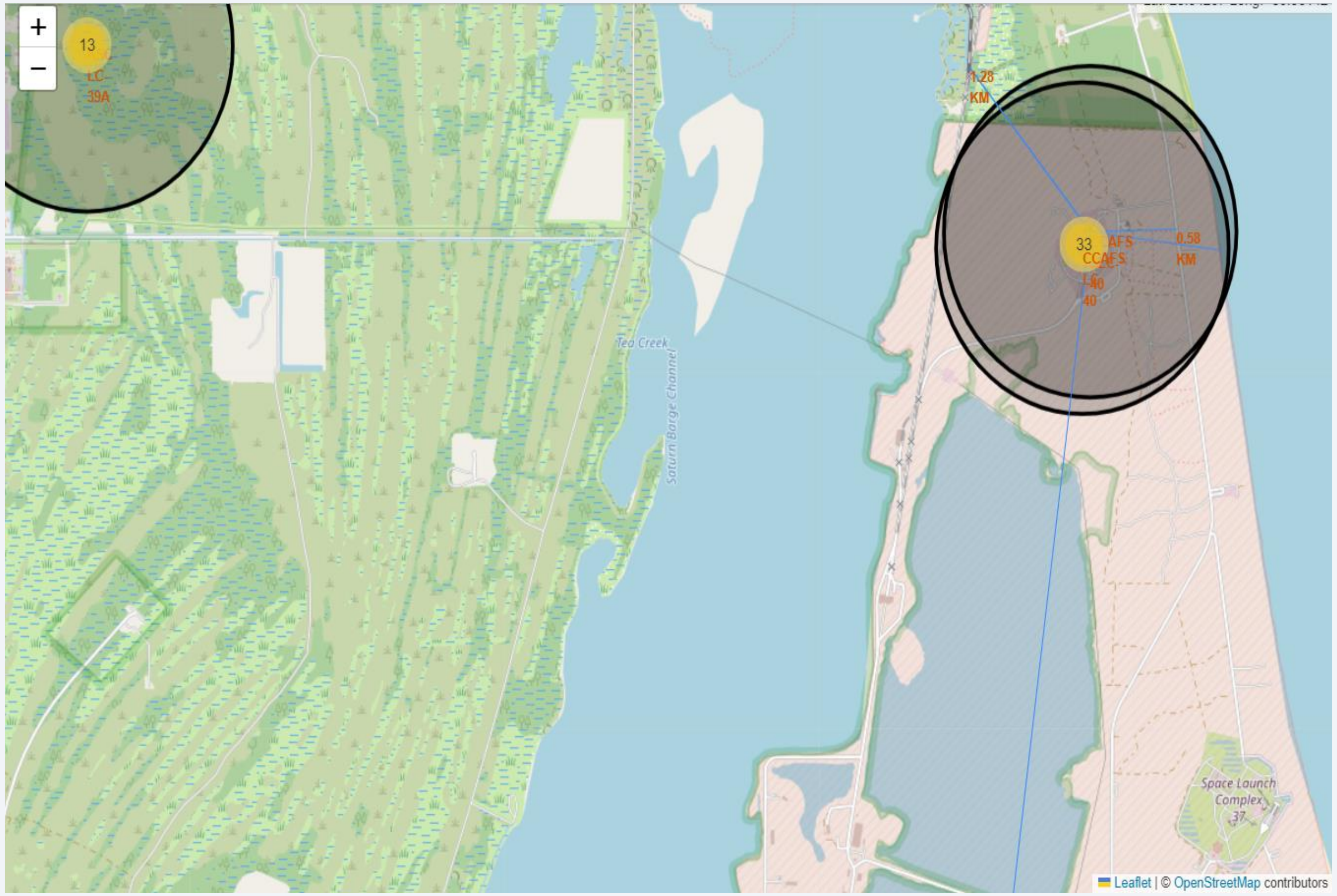
The key elements on this map are the launch sites highlighted in red text accompanied by a dot on the map.

Colour Labelled Outcomes on Map



The launch sites for successful launches are indicated by the 7 in a green marker cluster

Launch Site Proximities

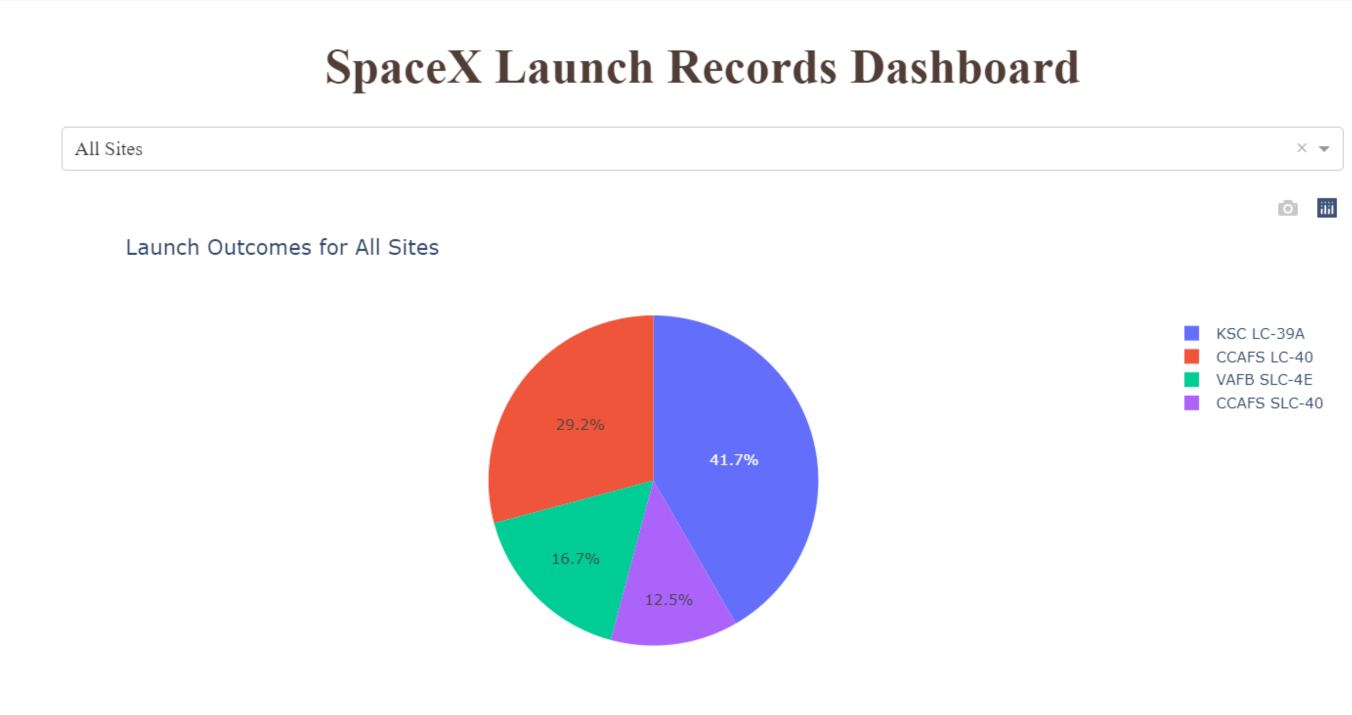


The blue lines on the map are Polylines indicating the proximity of our chosen launch site to other key areas (railway, highway, coastline) with the distance calculated and displayed

Section 4

Build a Dashboard with Plotly Dash

Launch Success for All Sites



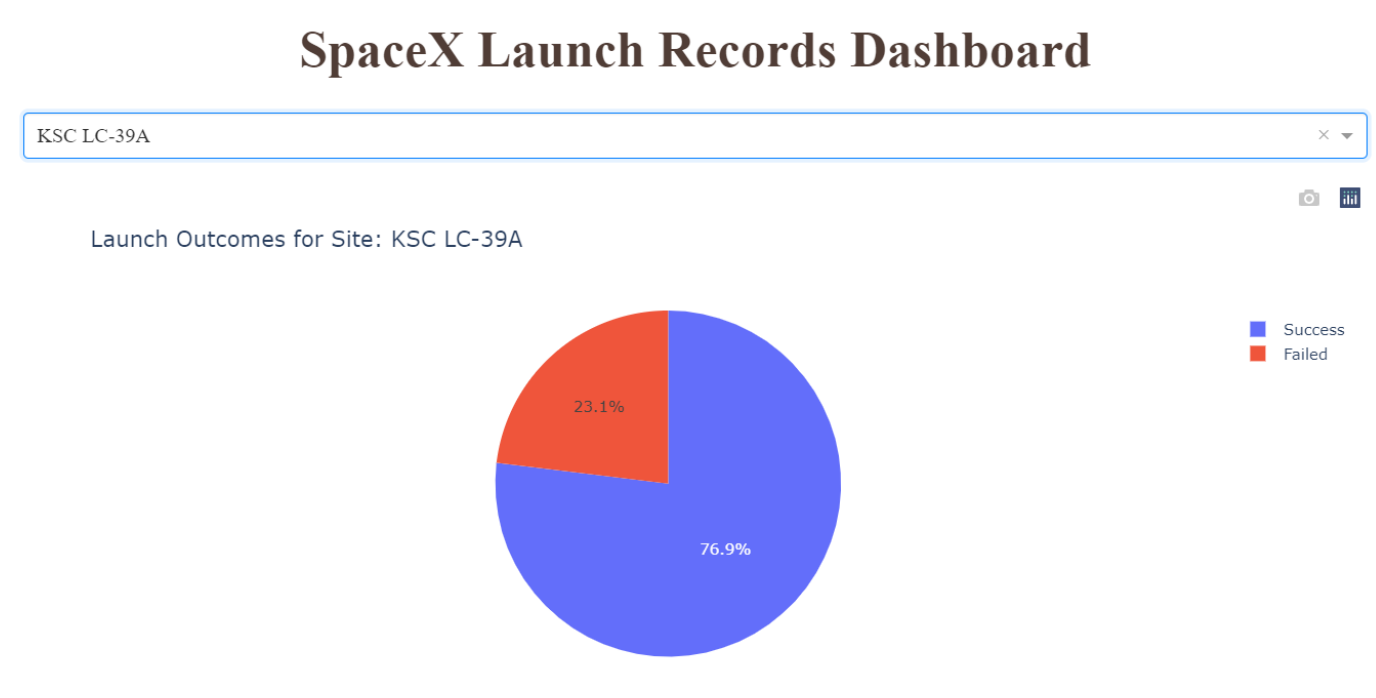
The pie chart is divided into four segments, each representing a different launch site: KSC LC-39A, CCAFS LC-40, VAFB SLC-4E, and CCAFS SLC-40. The percentage of successful launches for each site is displayed on the chart.

Here is a breakdown of the launch outcomes for each site:

- KSC LC-39A: 41.7%
- CCAFS LC-40: 29.2%
- VAFB SLC-4E: 16.7%
- CCAFS SLC-40: 12.5%

Overall, the chart shows that KSC LC-39A has the highest success rate among all SpaceX launch sites.

Launch Site with Highest Launch Success Ratio



The screenshot shows a pie chart representing the launch outcomes for the SpaceX launch site KSC LC-39A. The chart is divided into two segments, one for successful launches and one for failed launches.

Here's a breakdown of the launch outcomes for KSC LC-39A:

- **Success:** 76.9%
- **Failed:** 23.1%

The chart shows that KSC LC-39A has a relatively high success rate, with almost 80% of launches resulting in success.

Payload vs. Launch Outcome by Launch Site (ALL)

The scatter plots visualize the relationship between payload mass and launch outcome for different SpaceX launch sites. Each dot represents a single launch, with the x-axis representing payload mass and the y-axis indicating the launch outcome (success or failure).

The key findings from the analysis of the Payload vs. Launch Outcome scatter plots are:

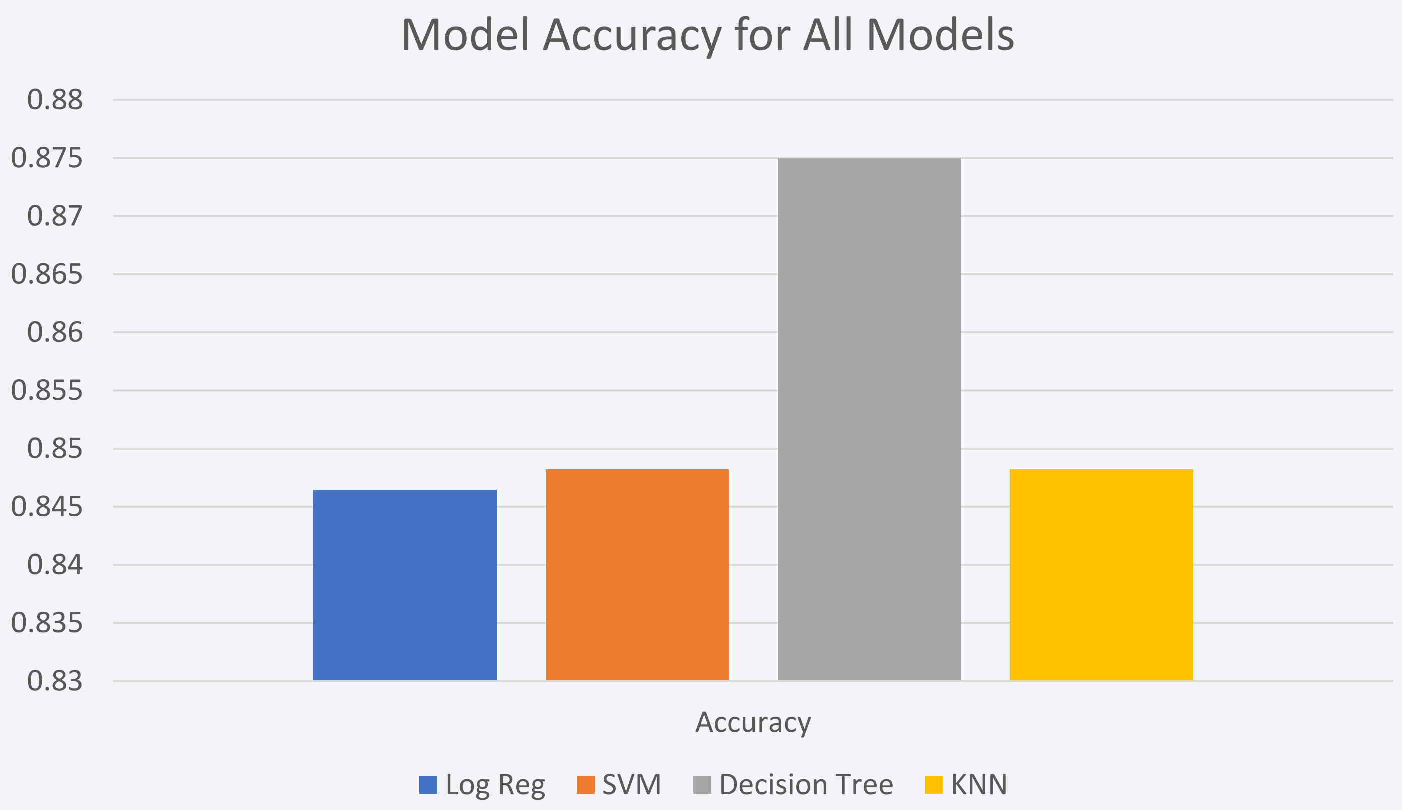
- **Payload Range and Success Rate:** There is a general trend that launches with higher payload masses tend to have a higher success rate, especially for certain launch sites.
- **Launch Site Performance:** VAFB SLC-4E consistently demonstrates high success rates across a wide range of payload masses, making it the top-performing launch site. KSC LC-39A also shows strong performance, especially for heavier payloads. CCAFS LC-40 exhibits mixed results, with some launches achieving high success rates but others experiencing failures.
- **Booster Version Impact:** While booster versions are not explicitly shown in the provided screenshots, previous analysis suggests that certain booster versions may have slightly different performance characteristics in terms of payload capacity and reliability.

Overall, the analysis indicates a positive correlation between payload mass and launch success, with certain launch sites and potentially booster versions demonstrating superior performance in handling heavier payloads.

Section 5

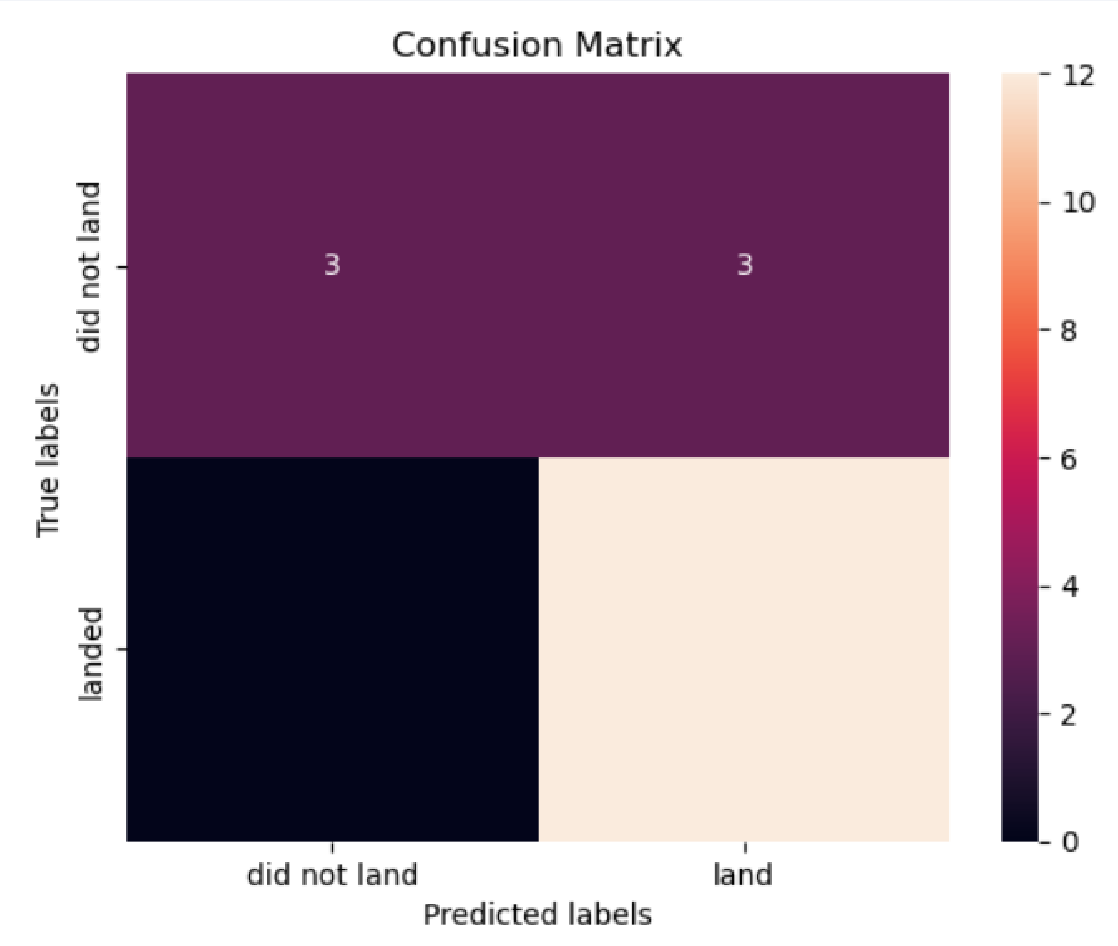
Predictive Analysis (Classification)

Classification Accuracy



From our bar chart, we can see that the decision tree classifier model surpasses the other models by a considerable margin of accuracy

Confusion Matrix



True Positive - 12 (True label is landed, the Predicted label is also landed)

False Positive - 3 (True label is not landed, the Predicted label is landed)

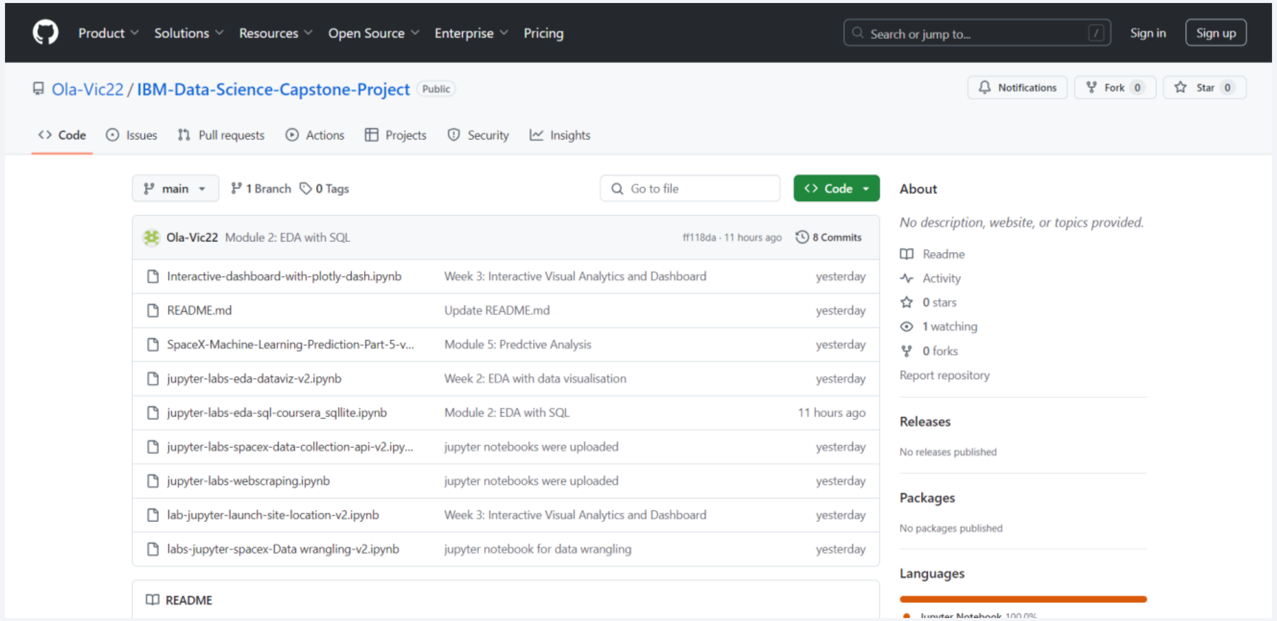
Conclusions

- Competitive Pricing: By analyzing SpaceX's data, we can determine a competitive launch price for Space Y.
- First Stage Reuse Prediction: Machine learning models can predict SpaceX's first stage reuse attempts, aiding cost reduction strategies.
- Data-Driven Insights: Our methodologies, including data collection, wrangling, and predictive analysis, provide actionable insights for Space Y.
- Future Strategies: Leveraging these insights, Space Y can develop effective strategies to compete in the commercial launch market.

Appendix

All the assets created during this project can be found at:

https://github.com/Ola-Vic22/IBM-Data-Science-Capstone-Project.git



Thank you!