IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Jacob Candelaria
Link to Git Repository
13 Feb 2024

Outline

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

Executive Summary

Objective: Utilize machine learning to accurately predict SpaceX Falcon 9 first-stage landing successes to enhance cost efficiency and planning.

Findings:

- Analysis identified key factors influencing landing success: payload mass, orbit type, and launch site.
- The Decision Tree model outperformed other models with a 94% accuracy, demonstrating high reliability in predicting landing outcomes.

Impact:

- Accurate predictions enable SpaceX to optimize launch operations, manage risks, and potentially reduce costs by improving rocket reusability.
- Insights support operational and engineering enhancements to increase future landing success rates.
- Conclusion: Machine learning offers a transformative approach for SpaceX to advance launch efficiency and sustainability, reinforcing its industry leadership in space exploration.

Introduction

- **Project Background and Context:**
- SpaceX, pioneering in space technology, aims to reduce space travel costs by reusing rocket components.
- Falcon 9's first-stage landing success is crucial for the reusability aspect, significantly impacting overall launch costs.
- Historical launch data provides insights into factors influencing landing outcomes, offering a foundation for predictive analysis.

- **Problems You Want to Find Answers to:**
- What factors most significantly influence the success of Falcon 9 first-stage landings?
- Can machine learning models accurately predict first-stage landing successes based on historical data?
- Among various predictive models, which one provides the highest accuracy in forecasting landing outcomes?
- How can predictive insights inform operational improvements and strategic decisions to enhance landing success rates and cost-efficiency?

Section 1

Methodology

Methodology – Executive Summary

Data Collection Methodology:

- Utilized SpaceX API and augmented datasets with key launch parameters including payload mass, orbit type, and landing outcomes.

Data Wrangling:

- Cleansed and structured raw data for analysis, handling missing values and standardizing formats.

Data Processing:

- Applied feature engineering to enhance the dataset, creating additional variables like launch site proximity factors.
- Standardized numerical data to prepare for machine learning model input.

Exploratory Data Analysis (EDA):

- Employed visualization tools and SQL queries to uncover patterns and correlations in the launch data.
- Identified significant factors potentially influencing landing success.

Interactive Visual Analytics:

- Utilized Folium for geospatial mapping of launch sites and their success rates.
- Developed a Plotly Dash application for dynamic, user-driven exploration of the data.

Predictive Analysis Using Classification Models:

- Built multiple classification models including Logistic Regression, SVM, Decision Tree, and KNN to predict landing success.
- Tuned models using GridSearchCV to identify optimal hyperparameters.

Model Evaluation:

- Assessed model performance using accuracy, confusion matrices, and classification reports to select the best-performing model for predicting landing outcomes.

Data Collection

Data Sources:

- Primary dataset retrieved from SpaceX API, including comprehensive launch records with geographical coordinates, payload mass, orbit, launch site, and landing outcomes.

Process Overview:

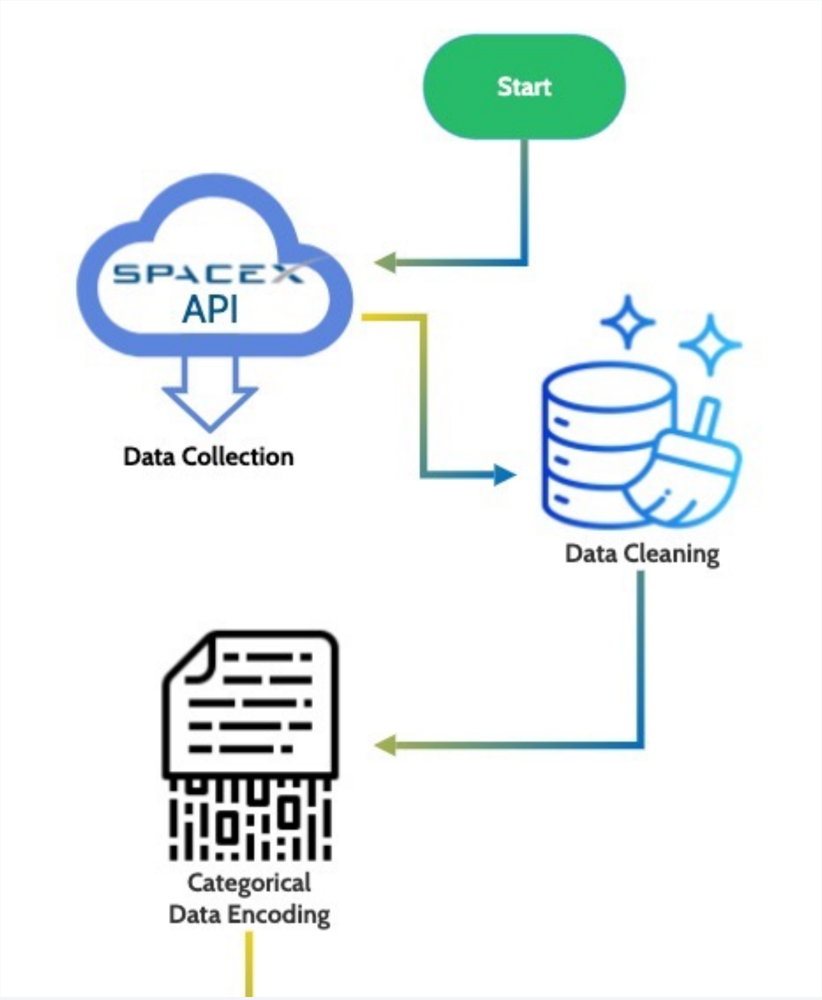
1. **API Retrieval:** Direct download of the SpaceX launch dataset via SpaceX API, featuring pre-included geographical coordinates for launch sites and detailed launch attributes.
2. **Data Cleaning:** Implementation of cleaning procedures to handle missing values, remove duplicates, and correct inconsistencies across the dataset.

Feature Encoding:

1. **Binary Encoding:** Applied to categorical data requiring binary representation for machine learning model compatibility.
2. **One-Hot Encoding:** Utilized for nominal categorical variables to transform them into a machine-readable format without introducing ordinality.

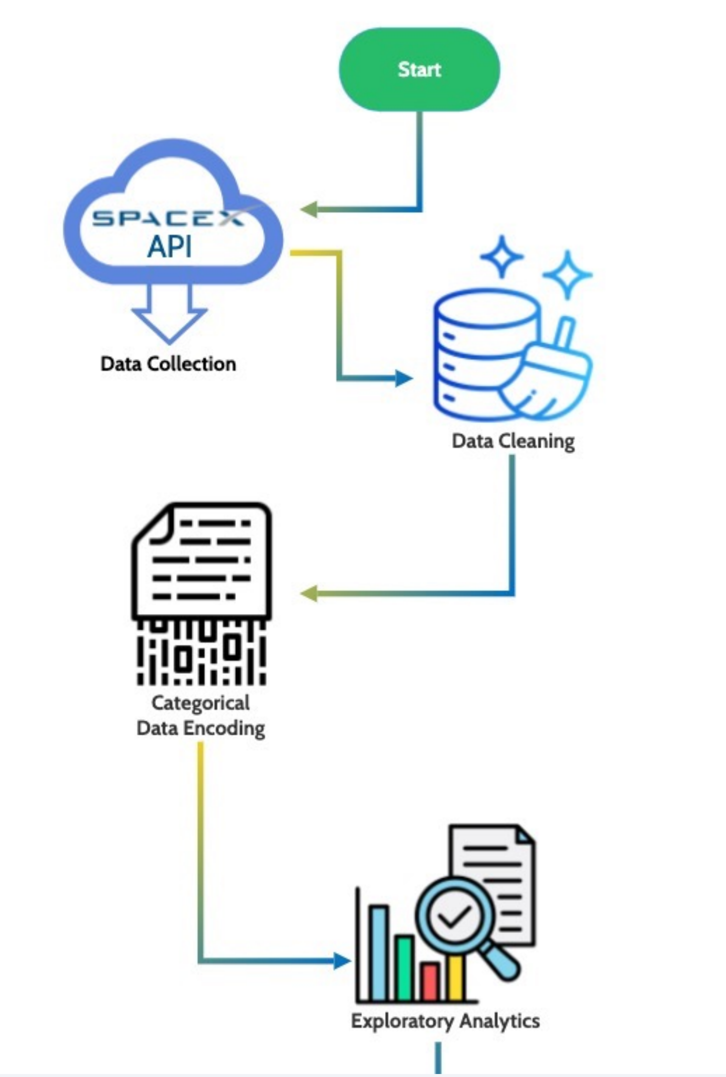
Key Phrases:

- "Automated data retrieval"
- "Data enrichment for depth of analysis"
- "Comprehensive dataset creation"



Data Collection – SpaceX API

- **API Data Retrieval:** Utilized the SpaceX API to gather comprehensive launch records.
- **Data Transformation:** Conducted thorough data cleaning and encoded categorical variables to prepare for analysis.
- **Insightful Analytics:** Performed exploratory data analysis to extract meaningful patterns and relationships.
- **GitHub Repository:** Access detailed code and outcomes in the Jupyter Notebook here.



Data Collection - Scraping

Objective:

- Extract historical launch records of Falcon 9 and Falcon Heavy from Wikipedia to analyze SpaceX launch success rates.

Process:

- Utilized BeautifulSoup to parse HTML content from the Wikipedia page.
- Extracted launch records from multiple tables containing detailed information about each launch.
- Cleaned and processed data to format dates, payload mass, and other key attributes for analysis.

Outcome:

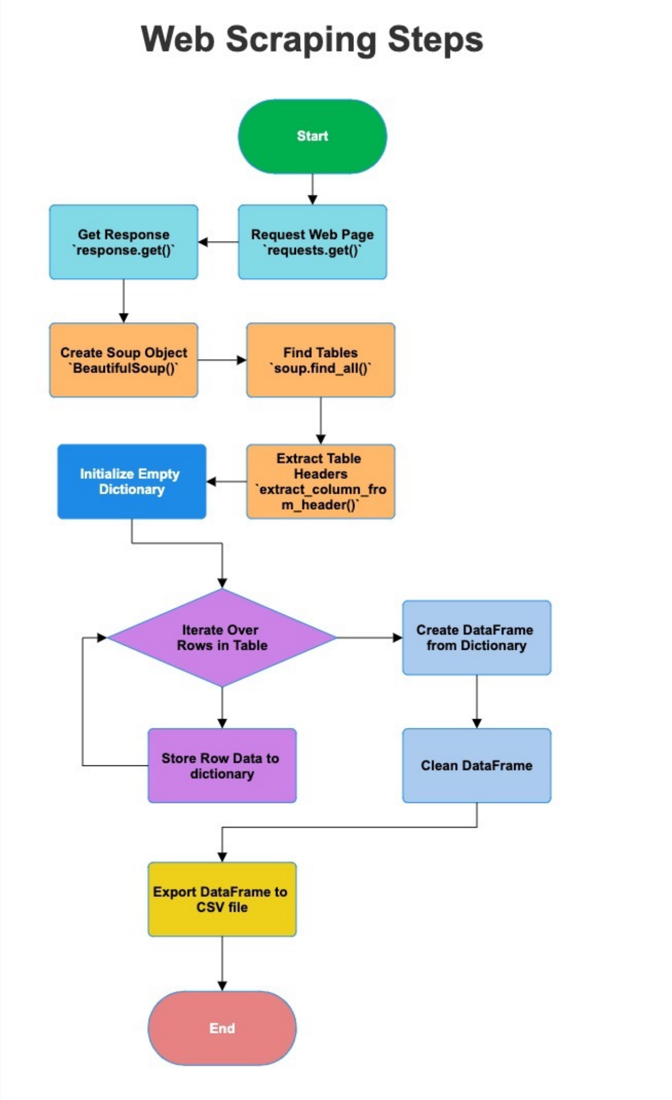
- Successfully compiled a dataset with over 90 Falcon 9 and Falcon Heavy launches, including details like launch date, payload mass, launch site, booster version, and launch outcome.
- This dataset served as a foundational element for further exploratory data analysis and machine learning modeling to predict launch success.

Tools & Techniques:

- Python for scripting the web scraping process.
- Requests library to make HTTP requests to retrieve web page content.
- BeautifulSoup for parsing HTML and extracting relevant information.

Significance:

- Web scraping enabled the collection of a comprehensive dataset that is not readily available in a structured form, facilitating a deeper analysis of factors influencing launch success.
- GitHub URL for Web Scraping Notebook: [click here]
- Web scraping is a crucial step in data collection, enabling us to gather detailed launch records for SpaceX's Falcon missions, laying the groundwork for predictive analytics.



Data Wrangling

- Data was sourced from the SpaceX API, which provided comprehensive launch details.
- Cleaning involved handling missing values, removing duplicates, and filtering relevant features.
- Categorical variables were encoded using One-Hot Encoding to convert them into a format suitable for machine learning algorithms.
- The payload mass, which had missing and inconsistent formats, was standardized to numerical values.
- Timestamps were parsed to separate date and time, allowing for time-series analysis.

Key Phrases for Flowchart:

- Data Extraction from API
- Data Cleaning and Preprocessing
- Categorical Data Encoding
- Normalization of Numerical Values
- Parsing and Splitting Timestamps
- Feature Selection for Model Training

EDA with Data Visualization

Launch Success Rate Over Time:

- Purpose: To analyze trends in launch success over the years, identifying improvements or patterns in SpaceX operations.
- Chart: Line graph showing the proportion of successful launches each year.

Payload Mass vs. Success Rate:

- Purpose: To understand the relationship between the payload mass of the rocket and its success rate, indicating if heavier payloads influence launch outcomes.
- Chart: Scatter plot with a trend line highlighting the correlation between payload mass and success rate.

Launch Site Effectiveness:

- Purpose: To compare the success rates across different SpaceX launch sites, determining the most reliable launch location.
- Chart: Bar graph comparing the success rates of launches from each SpaceX site.

Booster Version Success:

- Purpose: To evaluate how different versions of Falcon 9 boosters impact launch success, identifying the most efficient booster models.
- Chart: Histogram showing the success rate distribution across different Falcon 9 booster versions.

- GitHub URL for EDA Notebook: click here
- These visualizations provide insights into factors that influence Falcon 9 launch success, aiding in making informed decisions for future missions.

EDA with SQL

Overview:

- Performed SQL queries to explore SpaceX launch data, focusing on launch success rates, payload masses, and other key factors.

SQL Queries Summary:

- Launch Success Rate by Year: Queried the total number of launches and successful launches per year to calculate annual success rates.
- Average Payload Mass by Orbit Type: Calculated the average payload mass for different orbit types to understand the distribution of mission objectives.
- Launch Success Rate by Launch Site: Analyzed launch success rates for each launch site to identify the most reliable launch locations.
- Booster Version Frequency: Counted the frequency of each booster version used to highlight the most utilized versions in missions.
- Correlation between Payload Mass and Launch Success: Investigated the relationship between payload mass and launch success to detect any potential influence of payload mass on mission outcomes.

Tools & Techniques:

- Utilized SQLite for database management and execution of SQL queries.
- Applied pandas for handling query results and further analysis in Python.

Insights Derived:

- Identified trends in launch success rates over the years, suggesting improvements in technology and operations.
- Observed that certain orbit types are associated with higher payload masses, likely reflecting mission requirements.
- Recognized variability in success rates across different launch sites, indicating potential factors affecting launch outcomes.

Significance:

- The SQL-based EDA provided foundational insights into SpaceX's launch history, informing subsequent predictive modeling and analysis phases.
- GitHub URL for EDA with SQL Notebook: click here

- SQL-based exploratory data analysis offered critical insights into the SpaceX launch dataset, laying the groundwork for in-depth analysis and predictive modeling.

Build an Interactive Map with Folium

Overview:

- Utilized Folium library to create interactive maps highlighting SpaceX launch sites, their proximity to key geographical features, and launch outcomes.

Folium Map Objects Created:

- **Launch Sites Markers**: Placed markers on the map for each SpaceX launch site to provide a quick reference to their locations.
- **Success/Failure Circles**: Added colored circles around launch sites, with green indicating successful launches and red indicating failures, to visually represent launch outcomes.
- **Proximity Lines**: Drew lines from launch sites to nearby points of interest, such as coastlines or cities, to evaluate the strategic positioning of each site.
- **Information Popups**: Integrated popups with additional details such as launch site names, number of launches, and success rates, offering users more context upon interaction.

Reasons for Adding Objects:

- **Geospatial Insights**: To explore the spatial relationship between launch sites and their geographic environment, aiding in understanding potential factors influencing launch success.
- **Launch Outcome Visualization**: To provide an immediate visual representation of the success rates associated with each launch site, facilitating a more intuitive analysis.
- **Interactive Exploration**: To enable interactive engagement with the map, allowing users to explore and discover insights about SpaceX launch activities in a more dynamic manner.

Tools & Techniques:

- Employed Folium for map creation and customization, leveraging its capabilities to add various visual elements and interactivity features.
- Integrated data from SpaceX API and other sources to populate the map with accurate and relevant information.

Significance:

- The interactive map serves as a powerful tool for visualizing SpaceX's launch history and geospatial data, enhancing understanding of launch site selection and operational success factors.
- GitHub URL for Interactive Map with Folium Notebook: click here

- The Folium-based interactive map offers a comprehensive visualization of SpaceX launch sites, outcomes, and their geographical contexts, enriching the analysis with spatial insights.

Build a Dashboard with Plotly Dash

Overview:

- Developed an interactive dashboard using Plotly Dash to visualize SpaceX Falcon 9 launch data, enabling dynamic exploration of launch records and outcomes.

Key Features of the Dashboard:

- **Launch Site Selection Dropdown**: Users can filter the dashboard view by selecting a specific SpaceX launch site or opt to view all sites.
- **Success Rate Pie Chart**: Visualizes the proportion of successful launches versus failed attempts, offering insights into launch reliability.
- **Payload Mass vs. Outcome Scatter Plot**: Displays the relationship between payload mass and launch success, highlighting how payload size may impact mission outcomes.

Interactive Components:

- **Range Slider for Payload Mass**: Allows users to refine the scatter plot data based on payload mass range, facilitating targeted analysis.
- **Hover Data and Tooltips**: Provides additional details about each launch when hovering over data points in plots, enriching the user's interactive experience.

Reasons for Dashboard Elements:

- **Strategic Insights**: The dashboard's interactive elements were designed to provide stakeholders and enthusiasts with comprehensive insights into SpaceX launch performance, emphasizing factors like launch site effectiveness and payload capacities.
- **Data Exploration Flexibility**: By incorporating interactive filters and detailed visualizations, the dashboard empowers users to explore the dataset from various angles, fostering a deeper understanding of the intricacies involved in space launch operations.
- **Enhanced User Engagement**: The use of Plotly Dash for creating interactive, web-based visualizations ensures an engaging and accessible platform for users to interact with complex datasets intuitively.

Tools & Techniques:

- Utilized Plotly Dash for its powerful capabilities in building interactive, web-based dashboards that can handle real-time data updates and user interactions seamlessly.
- Integrated SpaceX launch data to construct meaningful visualizations that highlight key aspects of launch history and success metrics.

Significance:

- The dashboard serves as a versatile tool for analyzing SpaceX launch data, offering stakeholders valuable insights into performance trends and facilitating data-driven decision-making processes.
- GitHub URL for Plotly Dash Dashboard Script: click here
- This Plotly Dash dashboard exemplifies how interactive data visualization can enhance understanding of space launch data, offering users a hands-on approach to analyzing SpaceX's launch history.

Predictive Analysis (Classification)

Objective:

Develop a predictive model to forecast the outcome of Falcon 9 first stage landings based on launch data, enhancing decision-making for future launches.

Model Building and Evaluation:

- Data Preprocessing: Standardized features to scale the data, ensuring uniformity for model training.
- Model Selection: Explored multiple classification algorithms, including Logistic Regression, Support Vector Machine (SVM), Decision Trees, and K-Nearest Neighbors (KNN).
- Hyperparameter Tuning: Utilized GridSearchCV for each model to identify optimal hyperparameters, maximizing model accuracy.

Performance Evaluation:

- Accuracy Assessment: Evaluated models using accuracy scores on test data to compare predictive performance.
- Confusion Matrix: Analyzed true positives, false positives, true negatives, and false negatives to understand model strengths and weaknesses.

Model Improvement:

- Feature Engineering: Investigated additional features and transformations to enhance model predictability.
- Cross-Validation: Implemented cross-validation techniques to ensure model robustness and generalizability.

Best Performing Model:

- Decision Tree Classifier emerged as the top-performing model with superior accuracy and interpretability.
- Reasons for Selection: The model demonstrated high predictive accuracy and was effective in capturing the complexities of the dataset, making it well-suited for predicting landing outcomes.

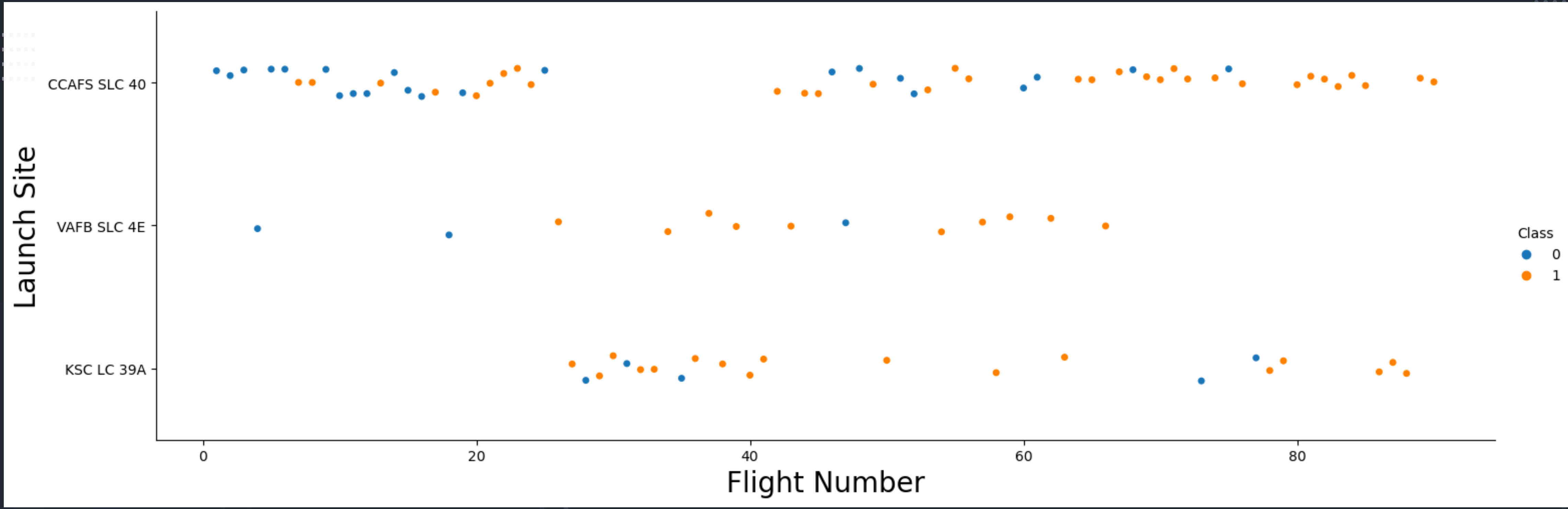
Implementation and Insights:

- The selected model provides actionable insights into factors influencing successful landings, aiding in future mission planning and booster design improvements.

GitHub URL for Predictive Analysis Notebook: click here

Section 2

Insights drawn from EDA



Flight Number vs. Launch Site

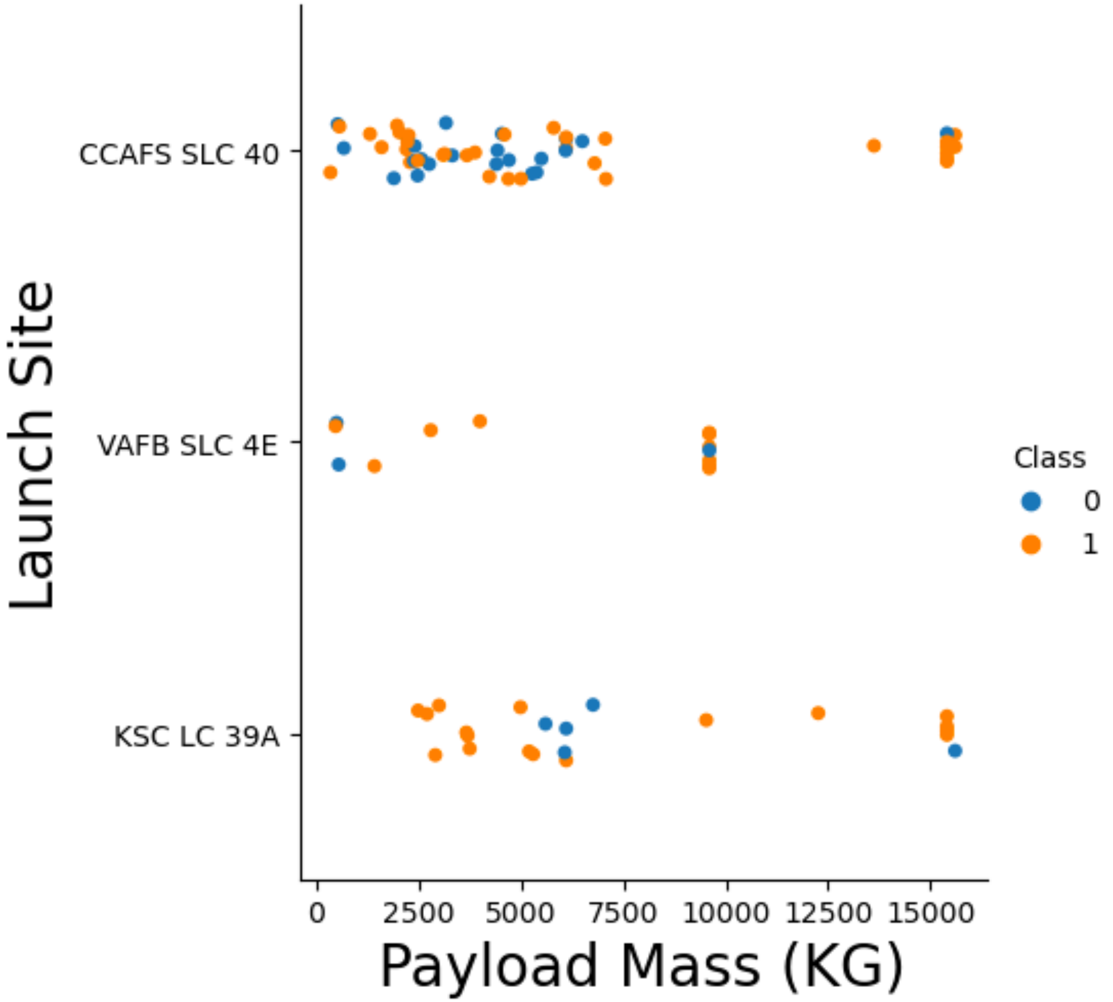
This scatter plot visualizes the success rate of SpaceX launches (Class 1: orange, Class 0: blue) at various launch sites over time. As flight numbers increase, a trend towards higher success rates is observed.

Results – EDA

Payload vs. Launch Site

This scatter plot shows the outcomes of each flight based on payload mass (kg) and launch site.

- VAFB-SLC 4E has no rockets launched with a payload that exceeds 10,000 kg.
- CCAFS SLC 40 appears to have more successful missions with a payload exceeding 10,000 kg.
- KSC LC 39A appears to have the higher mission success rate with payloads ranging from 2500 kg to ~15,000 kg.
- CCAFS SLC 40 appears to have the most rocket launches. This could be an indicator of being the location where rocket testing started.



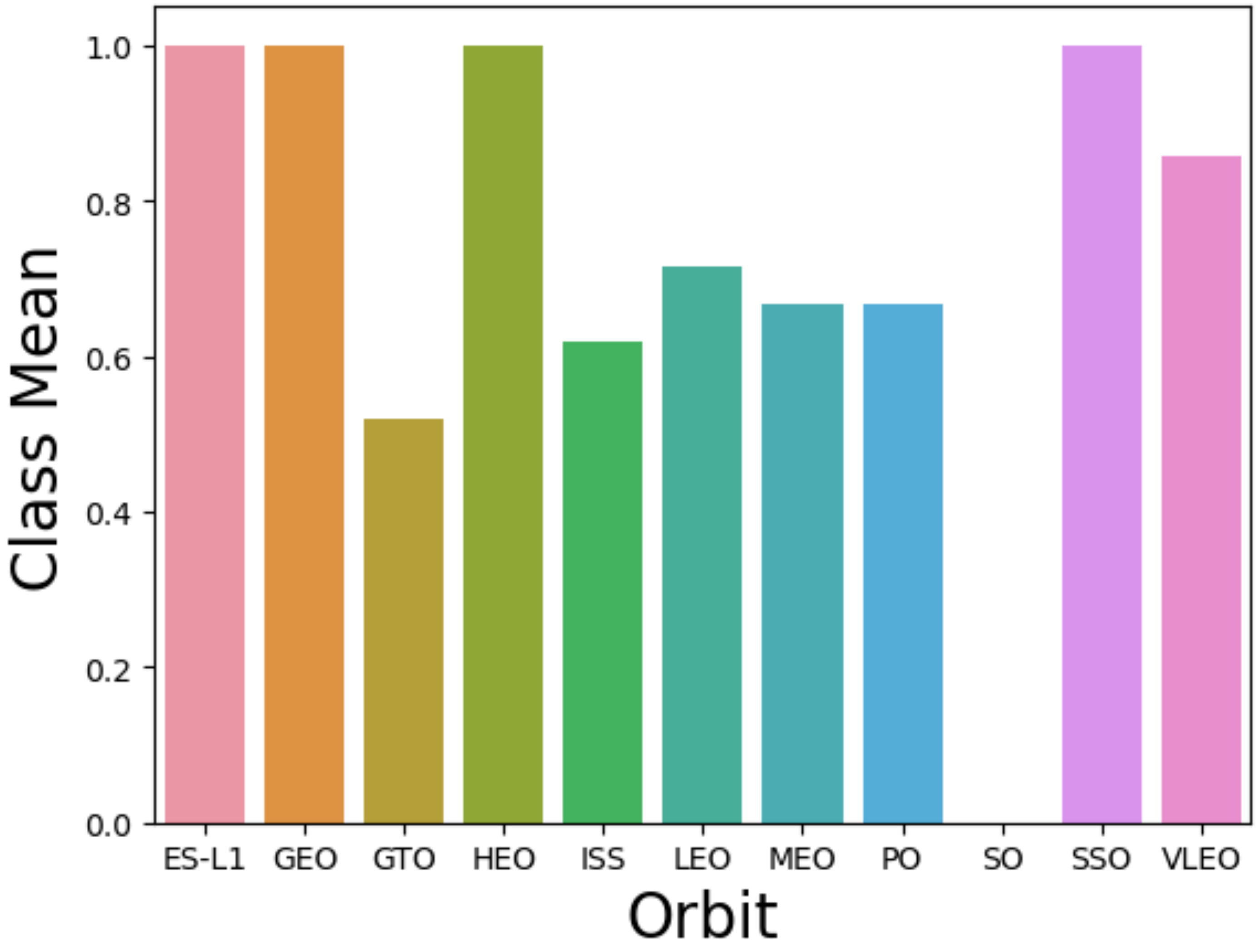
Results – EDA

Success Rate vs. Orbit Type

This bar chart displays the mean success rate of each flight based on type of Orbit.

The following Orbits have the most success (>= 80%):

- ES-L1: 100%
- GEO: 100%
- HEO: 100%
- SSO: 100%
- VLEO: ~85%

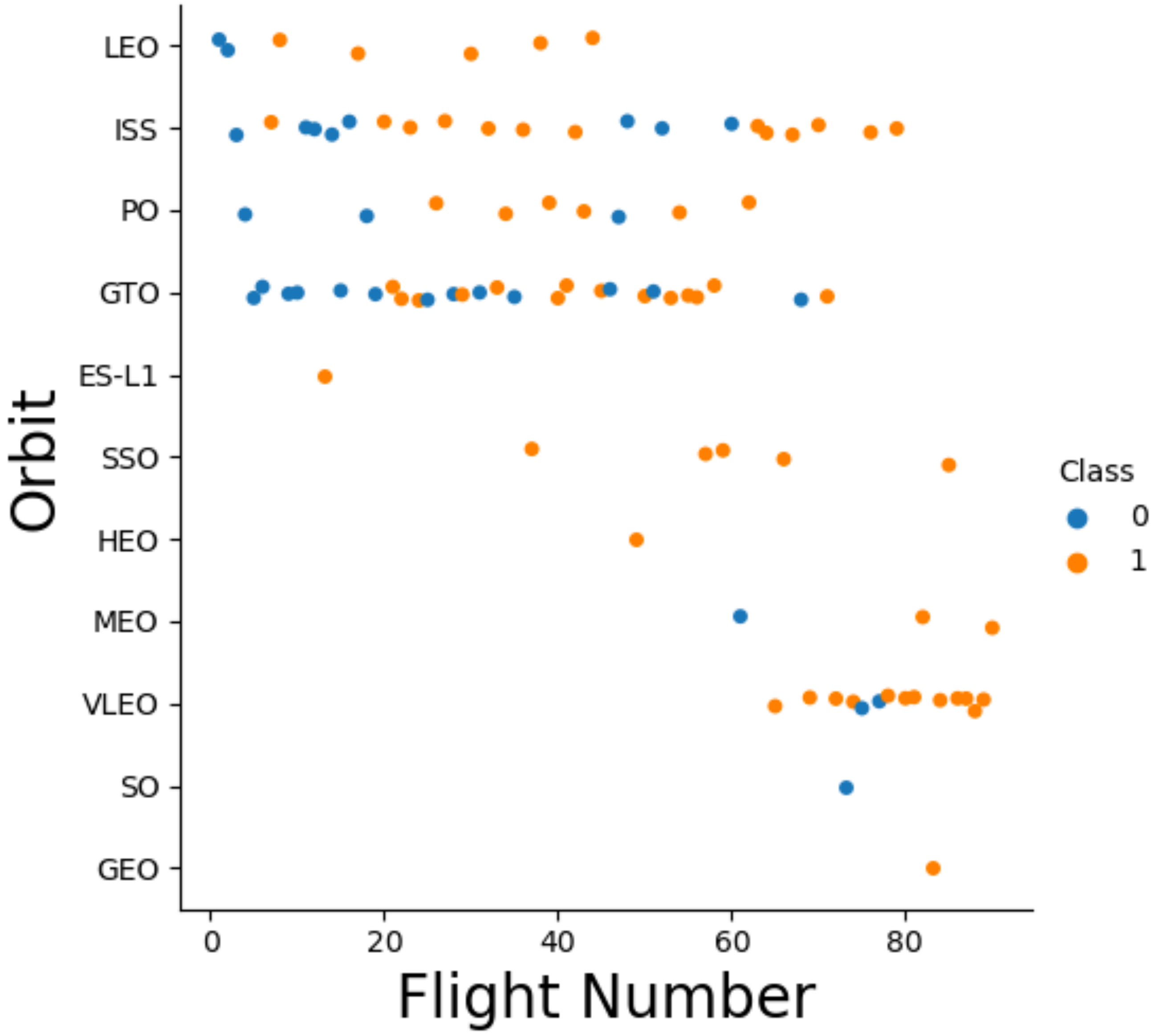


Results – EDA

Flight Number vs. Orbit Type

This line chart displays the mean success rate of SpaceX over a 10-year period.

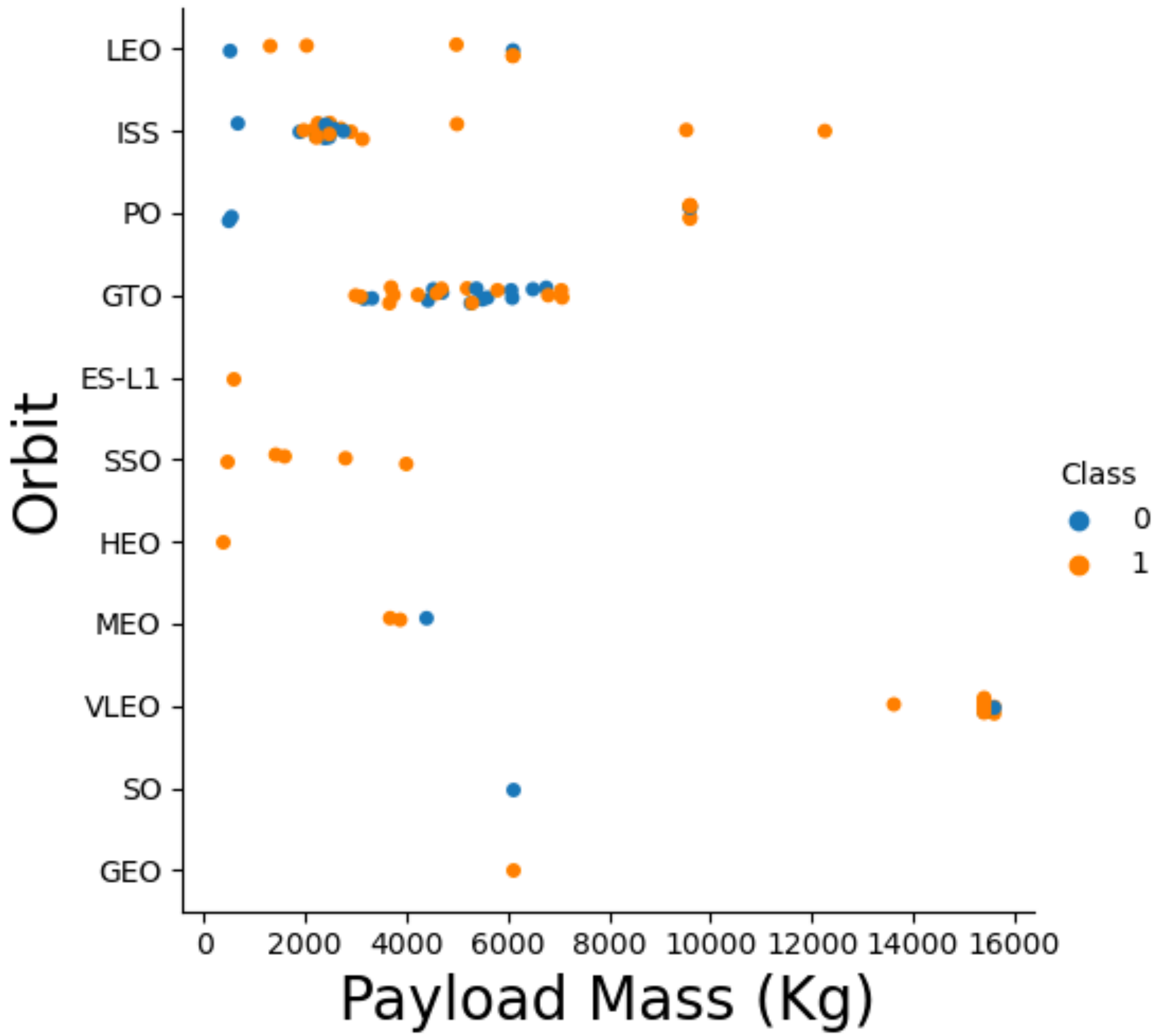
- 2010 – 2013 saw a continued 0% mean success rate
- 2014 saw a significant leap in the success rate, resulting to ~35%. The success rate remained stable through 2015.
- There was another significant leap in successful missions between 2016 and 2017.
- 2018 saw about a 20% decrease in success rate before rebounding in 2019.



Results – EDA
Payload vs. Orbit Type

This scatter plot displays the successful and failed missions by payload mass (kg) and type of Orbit.

- Polar, LEO and ISS display a higher success rate with heavy payloads.
- GTO appears to have an equal number of successful and failed landings.
- ES-L1, SSO, HEO, and GEO appear to heave a 100% success rate.

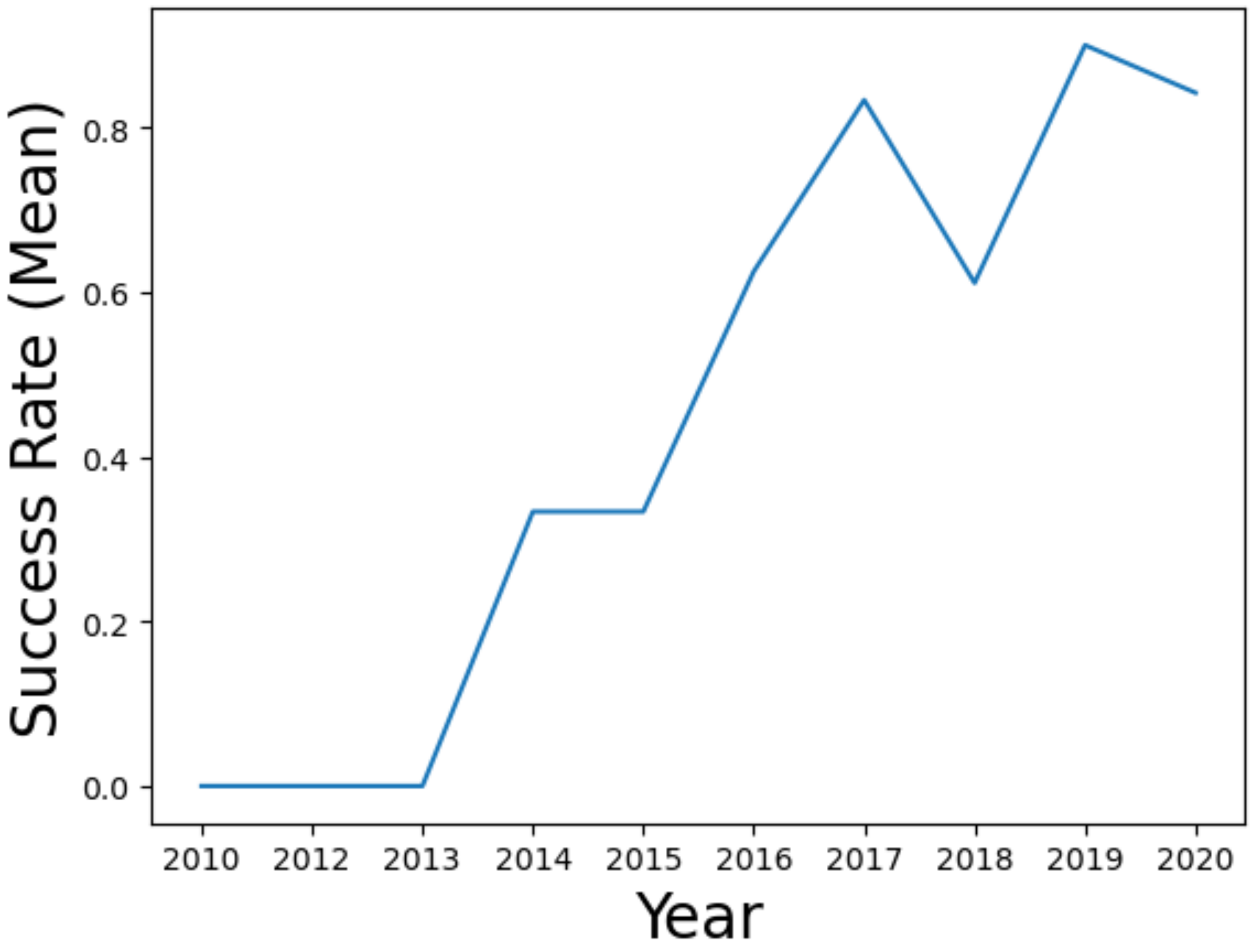


Results – EDA

Launch Success Yearly Trend

This line chart displays the mean success rate of SpaceX over a 10-year period.

- 2010 – 2013 saw a continued 0% mean success rate
- 2014 saw a significant leap in the success rate, resulting to ~35%. The success rate remained stable through 2015.
- There was another significant leap in successful missions between 2016 and 2017.
- 2018 saw about a 20% decrease in success rate before rebounding in 2019.



All Launch Site Names

Query:

```
SELECT DISTINCT launch_site FROM
spacextable;
```

Description:

This SQL query retrieves unique launch site names from the 'spacextable' without any duplicates.

Query Output

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

Launch Site Names Begin with 'CCA'

Query:

```
SELECT * FROM spacextable
WHERE launch_site LIKE 'CCA%’
LIMIT 5;
```

Description:
This SQL query selects all columns from the 'spacextable' table for rows where the 'launch_site' begins with 'CCA'. It limits the result to only the first 5 records.

Query Output

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

Total Payload Mass

Query:

```
SELECT customer,
SUM(payload_mass_kg) AS 'Sum of Payload Mass (KG)’
FROM spacextable
WHERE customer = "NASA (CRS)";
```

Description:
This SQL query filters the records to only include those where the 'customer' column has the value "NASA (CRS)". It then sums up the 'payload_mass_kg' for these filtered records, giving a total payload mass for NASA Commercial Resupply Services missions. The result will display the customer's name alongside the calculated sum of the payload mass.

Query Output

customer	Sum of Payload Mass (KG)
NASA (CRS)	45596

Average Payload Mass by F9 v1.1

Query:

```
SELECT Booster_Version, AVG(payload_mass_kg) AS
'Average Payload Mass (KG)'
FROM spacextable
WHERE Booster_Version = 'F9 v1.1’;
```

Description:

This SQL query calculates the average payload mass in kilograms for all the launches that used the 'F9 v1.1' version of the Falcon 9 booster. It selects records from 'spacextable' where the 'Booster_Version' is exactly 'F9 v1.1', computes the average value of the 'payload_mass_kg' column for these records, and labels this average as 'Average Payload Mass (KG)'. The result will display the booster version alongside the calculated average payload mass for that specific version.

Query Output

booster_version	Average Payload Mass (KG)
F9 v1.1	2928.4

First Successful Ground Landing Date

Query:

```
SELECT MIN(date) AS "First Successful Ground Pad
Landing"
FROM spacextable
WHERE landing_outcome = 'Success (ground pad)’;
```

Description:
This SQL query identifies the earliest date of a successful Falcon 9 booster landing on a ground pad. It searches the 'spacextable' for entries where the 'landing_outcome' column matches the condition 'Success (ground pad)', and then finds the minimum value in the 'date' column among these entries. The resulting date is labeled as "First Successful Ground Pad Landing" and represents the first time SpaceX achieved a successful land-based recovery of their Falcon 9 rocket's first stage.

Query Output

booster_version	Average Payload Mass (KG)
F9 v1.1	2928.4

Successful Drone Ship Landing with Payload between 4000 kg and 6000 kg

Query:

```
SELECT booster_version, payload_mass_kg,
landing_outcome
FROM spacextable
WHERE landing_outcome = 'Success (drone ship)'
AND payload_mass_kg > 4000 AND payload_mass_kg < 6000;
```

Description:
This SQL query retrieves information about Falcon 9 booster missions that successfully landed on a drone ship and had a payload mass between 4000 and 6000 kilograms. Specifically, it selects the booster version, payload mass in kilograms, and the landing outcome from the 'spacextable'. The query filters the data to include only those records where the 'landing_outcome' is 'Success (drone ship)' and the 'payload_mass_kg' falls within the specified range, indicating a successful sea-based landing of the rocket's first stage with a moderately heavy payload.

Query Output

booster_version	payload_mass_kg	landing_outcome
F9 FT B1022	4696	Success (drone ship)
F9 FT B1026	4600	Success (drone ship)
F9 FT B1021.2	5300	Success (drone ship)
F9 FT B1031.2	5200	Success (drone ship)

Total Number of Successful and Failure Mission Outcomes

Query:

```
SELECT mission_outcome, COUNT(*) AS 'Outcome Totals'
FROM spacextable GROUP BY mission_outcome;
```

Description:
This SQL query is designed to count the number of occurrences of each mission outcome in the 'spacextable'. It groups the records by the 'mission_outcome' field and then counts the number of records that correspond to each unique mission outcome, labeling this count as 'Outcome Totals'. Essentially, it provides a summary of the total number of missions that resulted in each type of mission outcome.

Query Output

mission_outcome	Outcome Totals
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

Query:

```
SELECT DISTINCT booster_version AS 'Booster Version',
payload_mass_kg AS 'Max Payload Mass (KG)'
FROM spacextable WHERE payload_mass_kg = (SELECT
MAX(payload_mass_kg) FROM spacextable)
ORDER BY booster_version;
```

Description:
This SQL query identifies the booster versions along with the maximum payload mass they have carried. It selects distinct entries for booster versions and their corresponding payload mass, but only where the payload mass matches the maximum payload mass recorded in the 'spacextable'. Essentially, it finds which booster versions were used to carry the heaviest payloads, listing each unique booster version that has achieved this maximum payload mass. The results are ordered by the booster version to provide a clear, organized list of these top-performing boosters in terms of payload capacity.

Query Output

Booster Version	Max Payload Mass (KG)
F9 B5 B1048.4	15600
F9 B5 B1048.5	15600
F9 B5 B1049.4	15600
F9 B5 B1049.5	15600
F9 B5 B1049.7	15600
F9 B5 B1051.3	15600
F9 B5 B1051.4	15600
F9 B5 B1051.6	15600
F9 B5 B1056.4	15600
F9 B5 B1058.3	15600
F9 B5 B1060.2	15600
F9 B5 B1060.3	15600

2015 Launch Records

Query:

```
SELECT CASE
   WHEN substr(date, 6, 2) = '01' THEN 'January'
   WHEN substr(date, 6, 2) = '02' THEN 'February'
   WHEN substr(date, 6, 2) = '03' THEN 'March'
   WHEN substr(date, 6, 2) = '04' THEN 'April'
   WHEN substr(date, 6, 2) = '05' THEN 'May'
   WHEN substr(date, 6, 2) = '06' THEN 'June'
   WHEN substr(date, 6, 2) = '07' THEN 'July'
   WHEN substr(date, 6, 2) = '08' THEN 'August'
   WHEN substr(date, 6, 2) = '09' THEN 'September'
   WHEN substr(date, 6, 2) = '10' THEN 'October'
   WHEN substr(date, 6, 2) = '11' THEN 'November'
   WHEN substr(date, 6, 2) = '12' THEN 'December'
END AS 'Month',
landing_outcome, booster_version, launch_site
FROM spacextable
WHERE landing_outcome = 'Failure (drone ship)' AND substr(date, 0, 5) = '2015’;
```

Query Output

Month	landing_outcome	booster_version	launch_site
January	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
April	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

Description:

This SQL query extracts and displays the month names, landing outcomes, booster versions, and launch sites for launches that resulted in a 'Failure (drone ship)' landing outcome within the year 2015. It uses a CASE statement to convert the month number extracted from the date column into the corresponding month name. The query filters the records to include only those where the landing outcome was a failure at a drone ship and the year is 2015, showcasing the specific failures for that year to analyze and perhaps understand the challenges faced during drone ship landings within that time frame.

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

Query:

```
SELECT landing_outcome, COUNT(*) AS 'Landing Count’
FROM spacextable
WHERE date BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY landing_outcome
ORDER BY COUNT(*) DESC;
```

Description:
This SQL query calculates and displays the count of different landing outcomes for SpaceX launches that occurred between June 4, 2010, and March 20, 2017. It groups the results by the type of landing outcome, such as successful landings, failures, or any other specified outcomes in the data. The query then orders the results in descending order based on the count of each landing outcome, enabling an analysis of which landing outcomes were most common during this period. This can provide insights into the reliability and performance of SpaceX's landing operations over these years.

Query Output

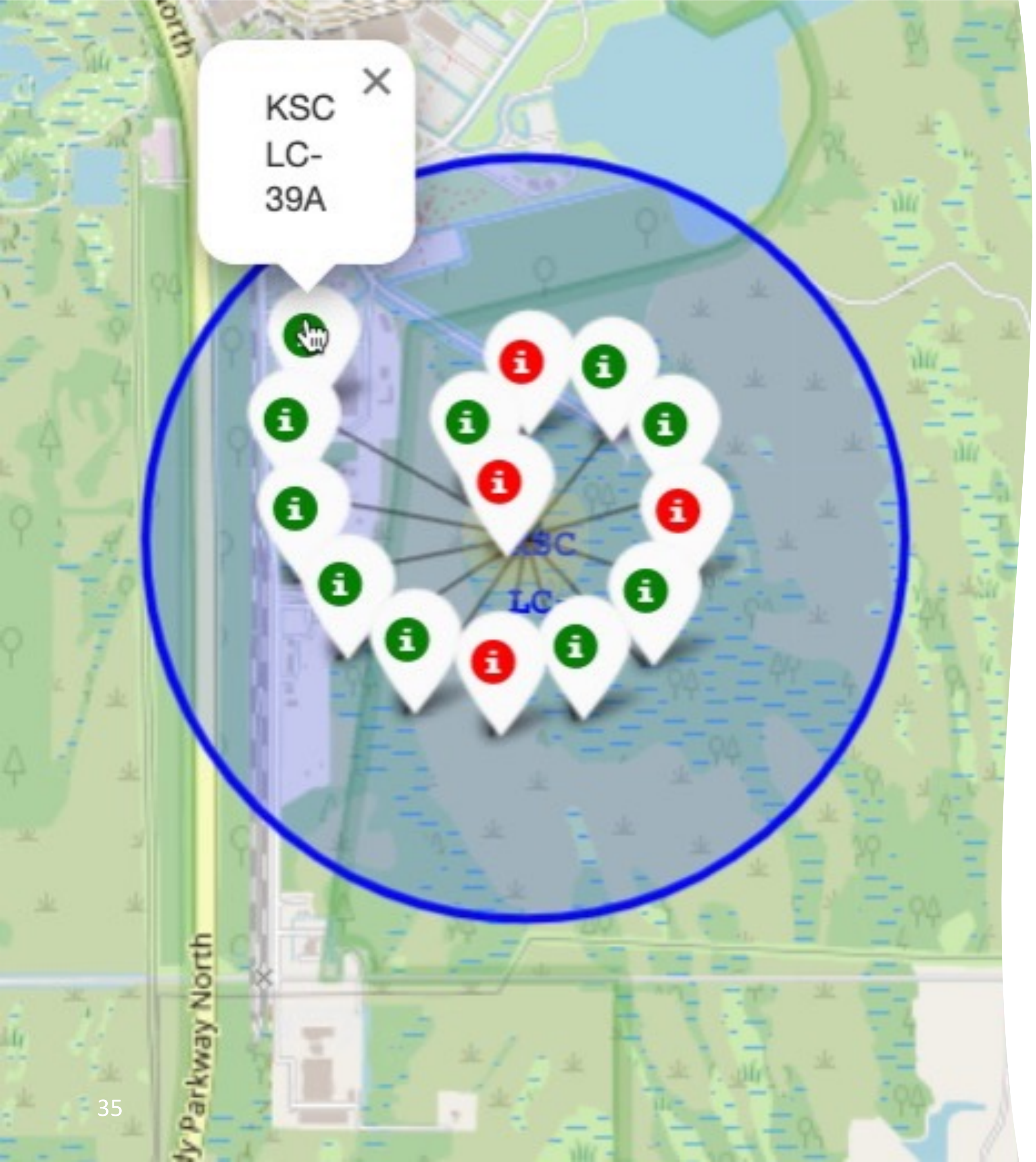
landing_outcome	Landing Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

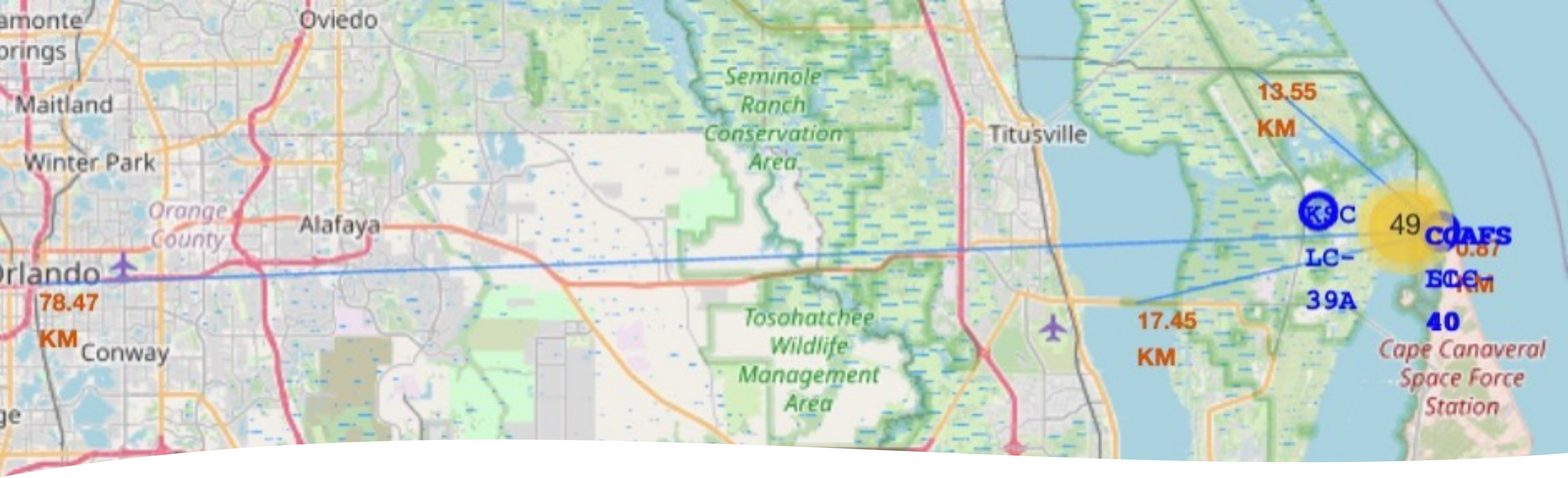
Geographic Distribution of SpaceX Launch Sites and Outcomes

The map is displaying the geographic locations of SpaceX launch sites. Each marker represents a different launch site, with their coordinates determined by latitude and longitude values. The color of each marker icon indicates the success (green) or failure (red) of launches from that site, as determined by the `marker_color` attribute in the data. The use of a marker cluster suggests that multiple launches have occurred at each site, allowing for a cleaner visualization that can be interactively explored to reveal individual launch details. The map itself is interactive, enabling zoom and pan functions for detailed examination.



Launch Outcomes at Kennedy Space Center LC-39A

- This map shows a cluster of markers indicating the launch outcomes of Falcon 9 rockets at the Kennedy Space Center Launch Complex 39A (KSC LC-39A). The markers represent individual launches, with different colors indicating the success or failure of each launch's first stage landing. A circle is drawn around the cluster, representing the boundary of the launch complex.



Launch Site Proximities

- City Proximity: The launch site is situated approximately 78.47 km from a major city, ensuring safety and minimizing urban disruptions.
- Railway Proximity: A railway is located roughly 17.45 km from the site, facilitating transport of materials and equipment.
- Highway Proximity: A major highway runs close to the site, about 13.55 km away, offering accessible routes for logistical support.
- Coastline Proximity: The launch site is positioned near the coastline, at a distance of 17.45 km, which is advantageous for safety and launch trajectory flexibility.

Section 4

Build a Dashboard with Plotly Dash

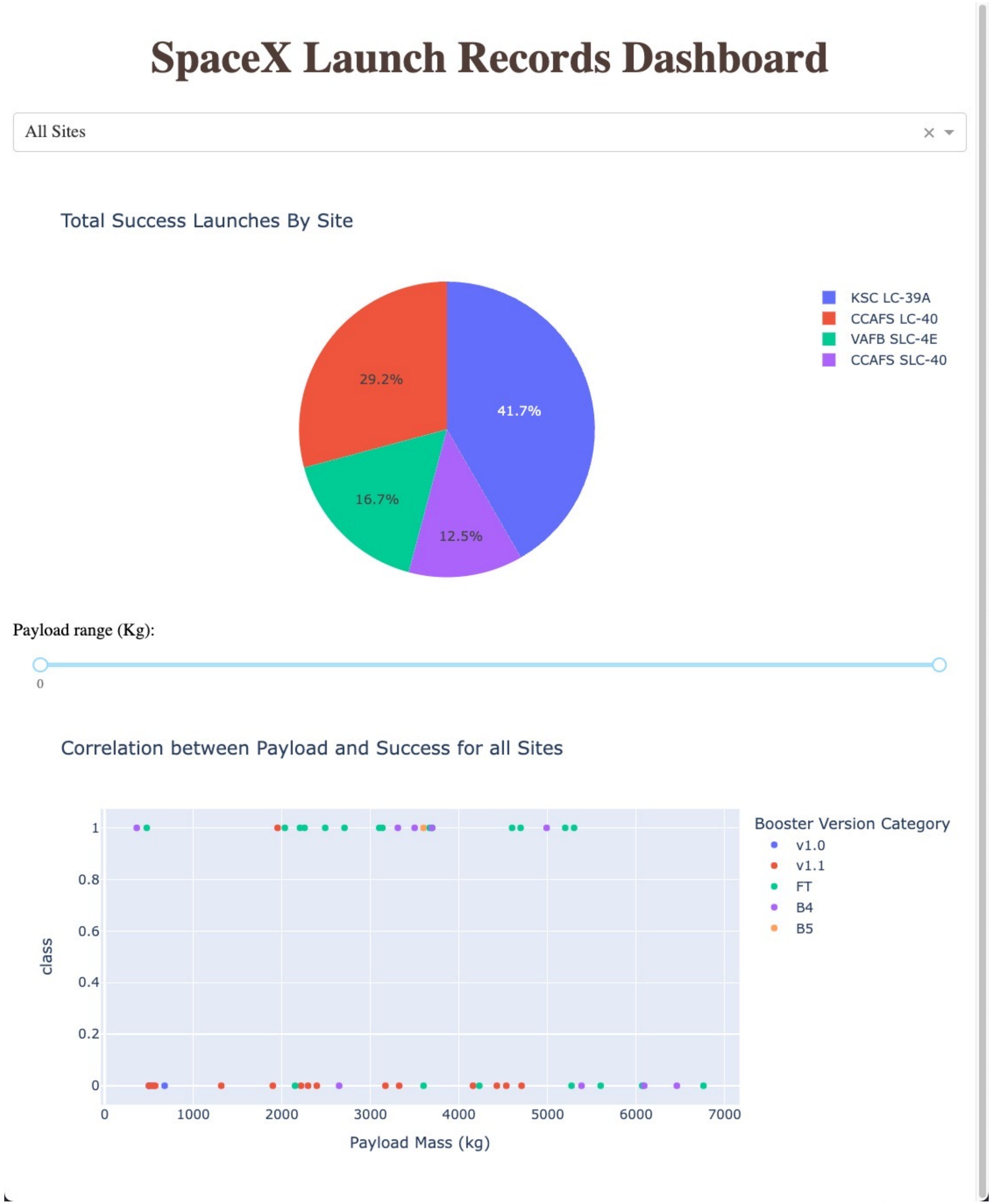
Launch Records Dashboard Overview

Total Success Launches by Site (Pie Chart):

- The pie chart illustrates the percentage distribution of total successful launches among the SpaceX sites.
- KSC LC-39A has the highest success rate at 41.7%, indicating it is the most utilized site for successful launches.
- CCAFS SLC-40 follows with 29.2%, VAFB SLC-4E at 16.7%, and lastly, CCAFS LC-40 at 12.5%.

Correlation between Payload and Success (Scatter Plot):

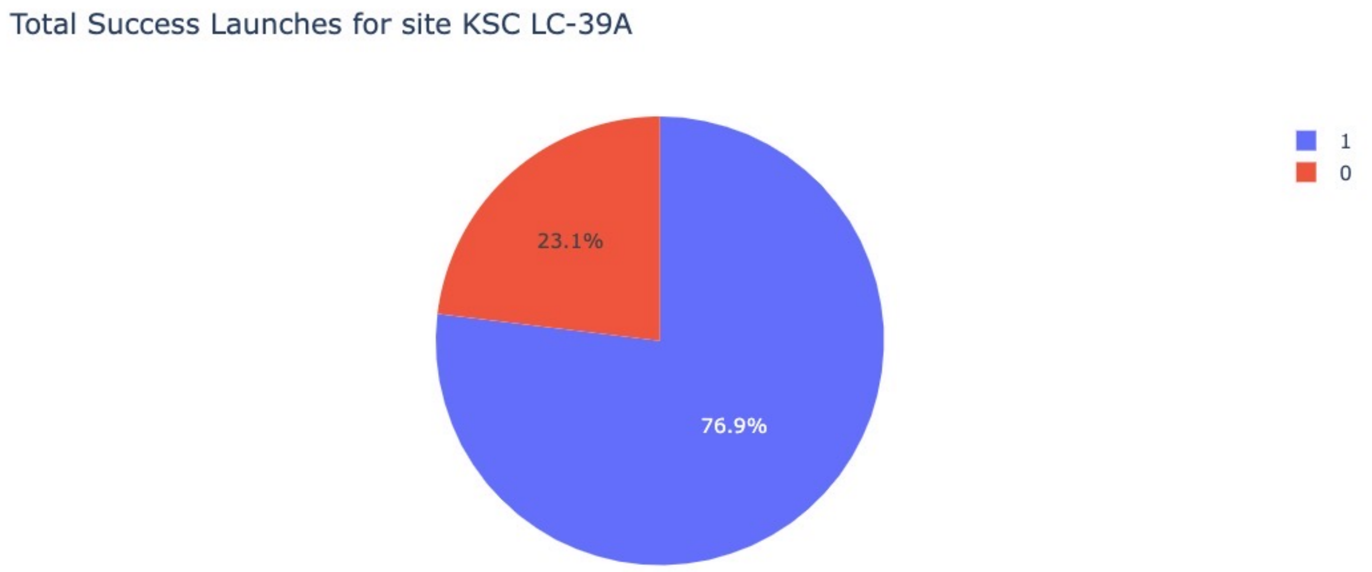
- The scatter plot below shows the correlation between payload mass (in kg) and launch success for all sites.
- Each point represents a launch, categorized by the booster version used, with different colors indicating different versions like v1.0, v1.1, FT, B4, and B5.
- There's no clear upward or downward trend, suggesting that within the range of payloads presented, there isn't a strong correlation between payload mass and the success of the launches. The success rate appears consistent across various payload masses.



Launch Outcomes for KSC LC-39A

The pie chart presents the distribution of launch outcomes for the KSC LC-39A site:

- The blue section, representing successful launches (classified as '1'), constitutes the majority with 76.9%.
- The red section, representing failed launches (classified as '0'), accounts for 23.1% of the launches at this site.



Launch Outcomes for KSC LC-39A

The scatter plot visualizes the correlation between payload mass (in kilograms) and launch success at the CCAFS LC-40 launch site. Each point represents a launch attempt, with the vertical axis (`class`) indicating success (1) or failure (0), and the horizontal axis showing the payload mass.

Key observations:

- The majority of successful launches (marked in green) appear across a wide range of payload masses.
- Launch failures (marked in red) are more frequent at lower payload masses.
- The booster version categories (v1.0, v1.1, and FT) are color-coded, but there does not seem to be a clear pattern linking booster versions to success or failure.

The data suggests that payload mass may not be a determining factor for success since successful launches are spread across the entire range of payload masses. However, there is a visible concentration of failed launches with lower payload masses, which might warrant further investigation into other factors that could contribute to launch outcomes.

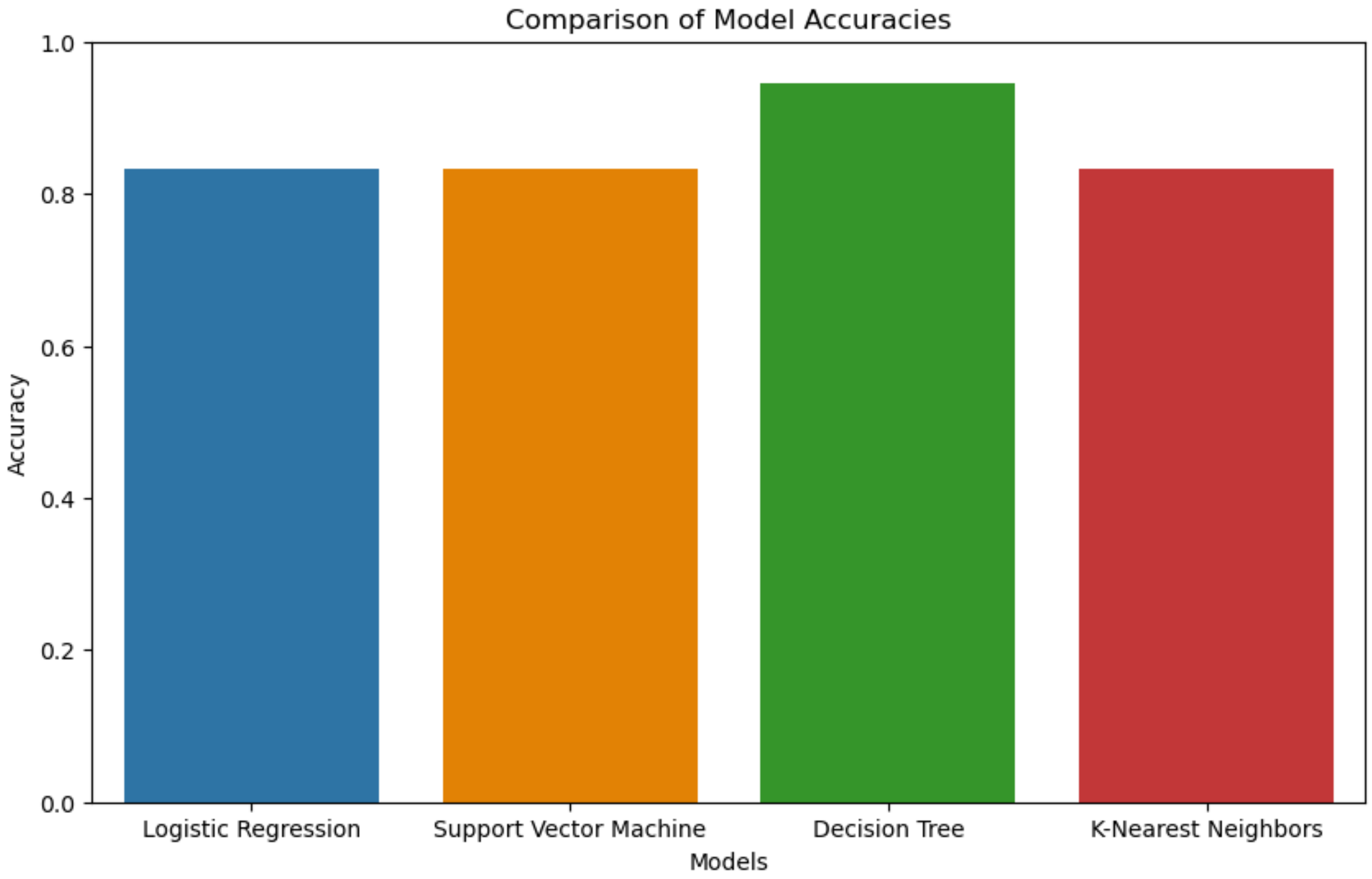
Section 5

Predictive Analysis (Classification)

Classification Accuracy

The bar chart provides a visual comparison of the accuracy of four different machine learning models used for predicting the success of Falcon 9 first stage landings.

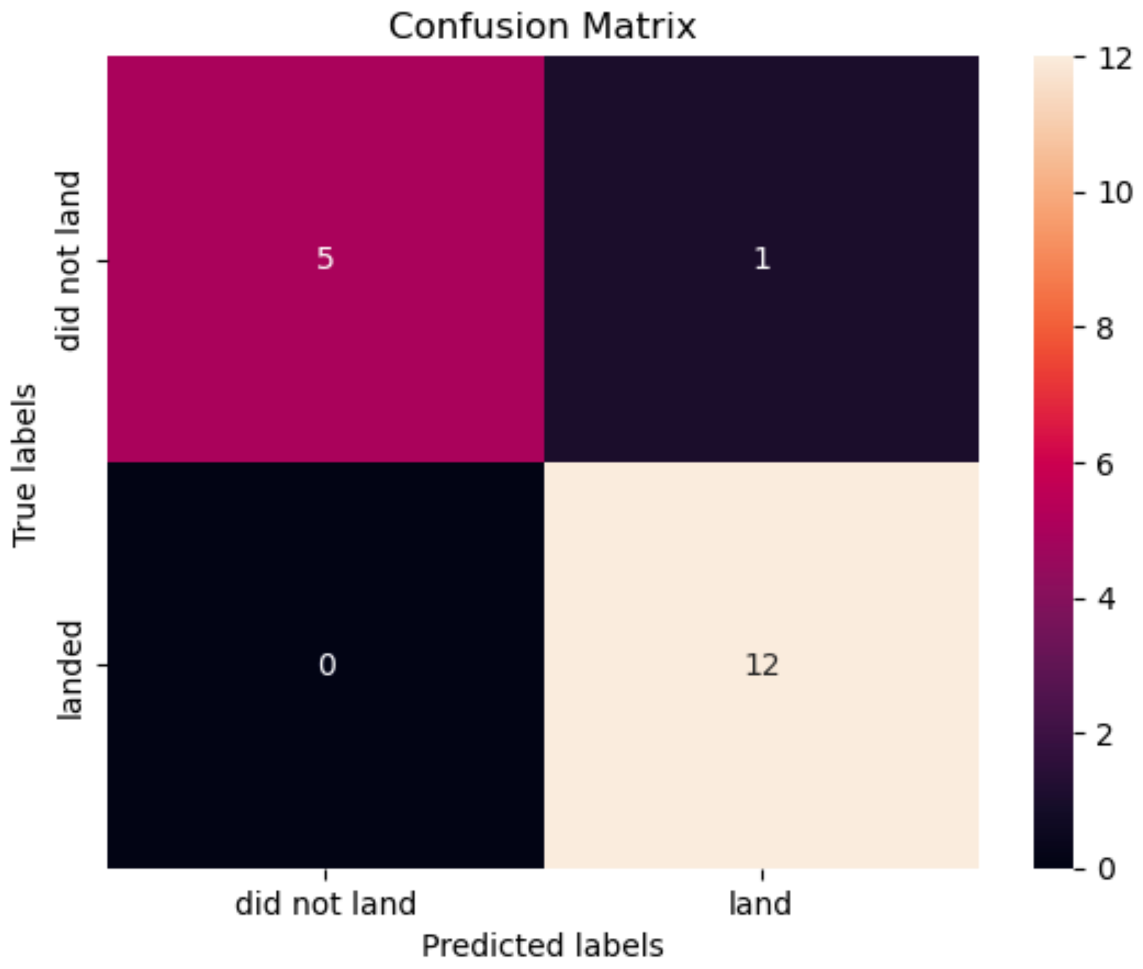
- Logistic Regression and Support Vector Machine models exhibit comparable performance, with Decision Tree showing a slight improvement in accuracy.
- The K-Nearest Neighbors model demonstrates similar levels of accuracy, indicating that all models perform well, but the Decision Tree model appears to have a slight edge.
- This suggests that, for this specific dataset and prediction task, the Decision Tree model might be the best choice among the four evaluated models.



Confusion Matrix

The confusion matrix displays the performance of the Decision Tree model in predicting Falcon 9 first stage landings.

- The model accurately predicted 12 landings (true positives) and correctly identified 5 instances where the landing did not occur (true negatives).
- However, there is 1 false positive, where the model predicted a landing that did not actually happen.
- There are no false negatives, indicating the model did not miss any successful landings.
- Overall, the model shows strong predictive power with high true positive and true negative rates, with very few misclassifications.



Conclusions

- The analysis of SpaceX Falcon 9 launch data revealed key insights into the factors influencing launch success and booster landing outcomes.
- Machine learning models, particularly the Decision Tree classifier, demonstrated high accuracy in predicting first-stage landings, underscoring the potential of data-driven approaches in enhancing launch strategies.
- Interactive visual analytics using Folium and Plotly Dash provided an intuitive understanding of launch site proximities and success rates, further aiding in the assessment of launch logistics.
- The exploratory data analysis (EDA) established clear correlations between payload mass, orbit type, and launch outcomes, enabling a better prediction of launch success probabilities.
- Overall, the project successfully utilized various data science techniques to distill actionable insights from the Falcon 9 launch data, offering a comprehensive view of the variables that contribute to mission success.

Appendix

Visualize Decision Tree Outcomes

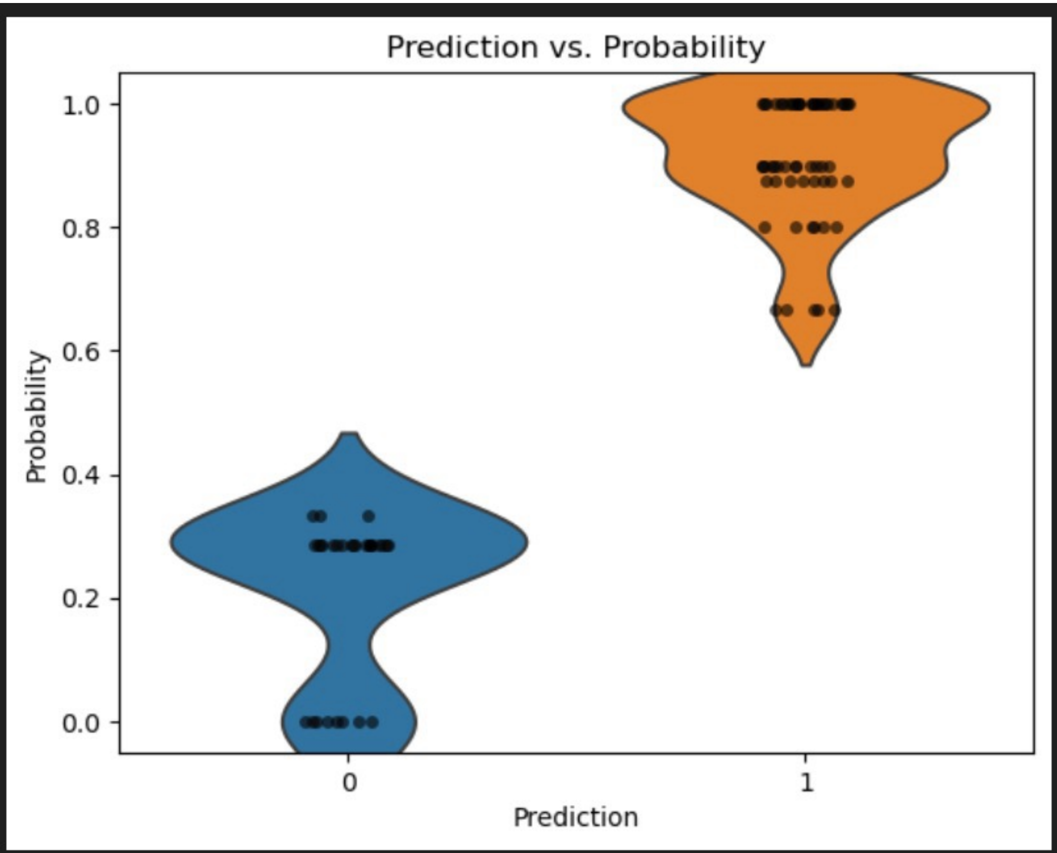
```python
# Prepare data for visualizing the predictions.
# Run the model on a copy of the wide dataset (X)
# Calculate Preditions and Probabilities
# Generate predictions
predictions = tree_cv.predict(X)

# Generate probabilities for the positive class
probabilities = tree_cv.predict_proba(X)[:, 1]

flight_numbers = data['FlightNumber'].values

# Create a DataFrame for predictions and probabilities with FlightNumber
predictions_df = pd.DataFrame({
    'FlightNumber': flight_numbers,
    'Prediction': predictions,
    'Probability': probabilities
})
```

[38] ✓ 0.0s Python

```python
#Merge the predictions back into the original data DataFrame
tree_df = pd.merge(data, predictions_df, on='FlightNumber', how='left')
```

[39] ✓ 0.0s Python

Appendix



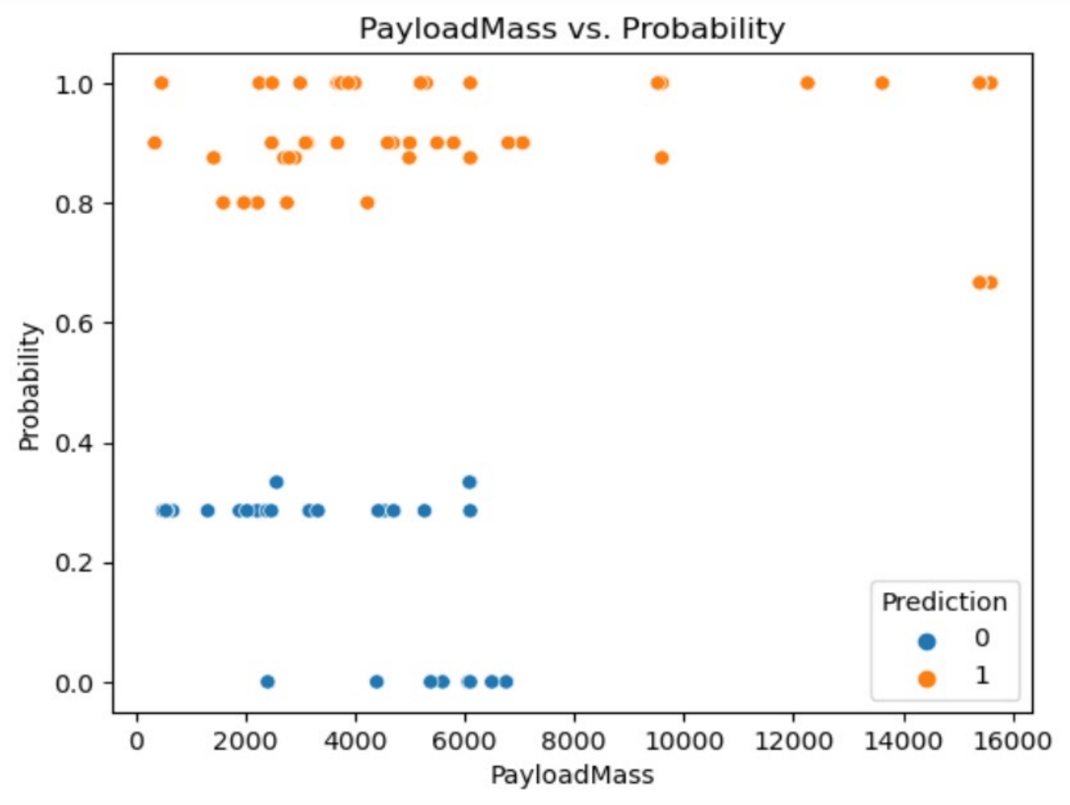
Notes:

The violin plot demonstrates that when the model predicts a successful landing (Prediction 1), it does so with high confidence, as evidenced by a dense concentration of data points at probability levels around 80% and higher. The shape of the violin indicates that most of these predictions are clustered near the probability of 1, which suggests a strong certainty in these outcomes.

Conversely, for predictions of unsuccessful landings (Prediction 0), the model's probabilities are concentrated around 30%, with the bulk of these predictions being closer to 0. This again shows a high level of confidence in these predictions, albeit the confidence is in the negative outcome.

The "confidence" in this context refers to the model's assigned probability to its predictions, with probabilities closer to 1 or 0 indicating stronger confidence in the respective predicted outcomes.

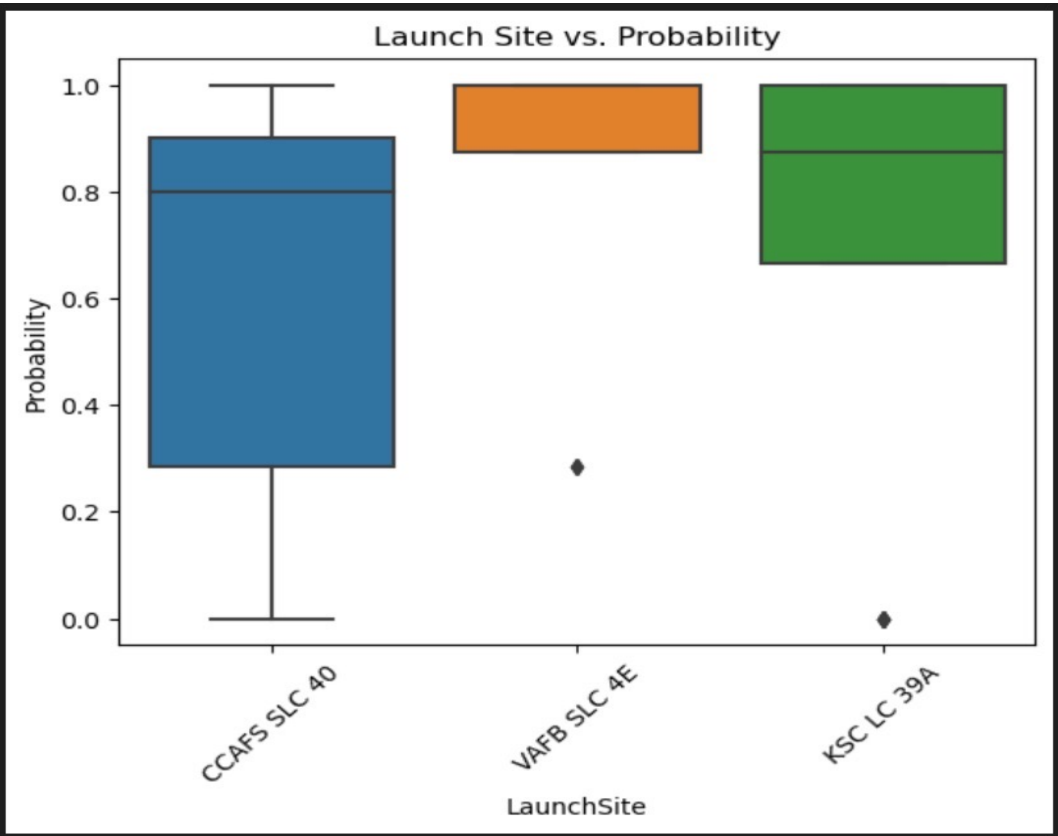
Appendix



Notes:

The scatter plot exploring the relationship between payload mass and the probability of a successful landing reveals a nuanced pattern. While heavier payloads uniformly correlate with a high probability of successful landings, indicating consistent model confidence across all launches with higher masses, the data for lighter payloads present a more varied picture. Within this lower mass range, there is a blend of both successful and unsuccessful predicted outcomes, with a tendency towards predicting success. Notably, the instances of predicted failures are predominantly situated at the lower end of the mass spectrum. This pattern could suggest that during the initial phases of the Falcon 9 launches, SpaceX was still refining their landing techniques, which is reflected in the model's predictions of a greater likelihood of unsuccessful landings for these early, lighter payloads. Over time, as expertise and technology advanced, the model's predictions shift towards anticipating success, regardless of the payload mass, which could be indicative of improved reliability and a maturation of SpaceX's landing capabilities.

Appendix



Notes:

The box and whisker plot illustrates a comparative analysis of the model's predicted probabilities for successful landings at various launch sites. VAFB SLC 4E stands out with a notably higher median probability, suggesting that, according to the model, this site has a more consistent track record of successful landings. The data points for VAFB SLC 4E are tightly clustered, indicating less variation in launch outcomes and a strong model confidence in successful landings.

In contrast, CCAFS SLC 40 and KSC LC 39A exhibit a wider interquartile range, implying a greater variability in the model's predictions for these sites. Despite this variability, both sites display median probabilities exceeding 80%, denoting that the model generally predicts successful landings more often than not. This high median probability is indicative of a solid success rate, although the broader spread of data points suggests that there are still a number of launches for which the model reserves judgment or predicts a lower chance of success. Overall, the model's predictions mirror an optimistic outlook for landing success at these sites, with a recognition of the inherent complexities and varying confidence levels associated with individual launches.

Thank you!