Winning Space Race with Data Science

Davide Corbo
02.01.2024

Outline

EXECUTIVE SUMMARY

INTRODUCTION

METHODOLOGY

RESULTS

CONCLUSION

APPENDIX

Executive Summary

This work provides an in-depth analysis of SpaceX rocket launches, focusing on the critical factors that contribute to their cost efficiency and success. SpaceX, a leader in the aerospace industry, advertises Falcon 9 rocket launches at a significantly lower cost of $62 million per mission, compared to over $165 million charged by other providers. A major factor behind this cost reduction is the reusability of the first stage of the rocket, a technological breakthrough that minimizes material waste and production expenses.

This report aims to explore the determinants of successful first-stage landings, as they are pivotal to cost reduction. By leveraging historical launch data, we analyze key variables such as payload mass, launch sites, booster versions, and mission outcomes to develop predictive models. These models can assist in determining the likelihood of a first-stage landing, thereby offering valuable insights into the potential cost of future launches.

In addition to understanding the economic implications of SpaceX's reusability strategy, this analysis highlights broader trends in the aerospace sector, such as the impact of innovation on reducing barriers to space exploration. Ultimately, this report serves as a foundation for further research into optimizing rocket launches and fostering sustainable advancements in space technology.

In this analysis, we focus on understanding the factors that contribute to the success of Falcon 9's first stage landing. By examining historical data on launch outcomes, payload mass, booster versions, and launch sites, we aim to uncover patterns that could provide insights into the likelihood of a successful landing.

Key factors considered include the landing location—whether it's a drone ship in the ocean or a ground pad—and the payload mass carried by the rockets. Additionally, we investigate the role of the booster version in landing success, as newer versions may have improved reusability and landing precision.

Through this analysis, we not only explore the cost-saving potential of reusing rocket stages but also provide actionable insights that could inform future SpaceX missions, potentially contributing to further reductions in launch costs. Our findings suggest that certain launch conditions, such as specific booster versions and payload mass ranges, have a statistically significant impact on landing success, which could help optimize mission planning for future launches.

Introduction

- SpaceX, founded by Elon Musk in 2002, has revolutionized space travel with its innovative approach to rocket launches, particularly with the Falcon 9 series. The company’s main selling point is the reusability of its rockets, particularly the first stage, which can be recovered and reused, drastically reducing the cost of launches. Falcon 9 launches typically cost $62 million per mission, compared to over $165 million for other providers. This cost reduction has positioned SpaceX as a major player in both government and commercial space exploration.
- This project aims to analyze SpaceX's launch data, focusing on the key factor that influences the overall cost: the successful landing of the first stage. Specifically, it seeks to predict the likelihood of a successful landing, which directly impacts the cost efficiency of a mission. By exploring factors such as launch site, booster version, payload mass, and launch outcome, the goal is to build a data-driven model that could offer valuable insights into launch outcomes, including the probability of the first-stage booster successfully landing.
- The primary problem we aim to address is: What factors contribute to the successful landing of the Falcon 9 first stage, and can we predict the success of future landings based on past data? By answering this, we can better understand how reusability affects launch economics, providing insights into how SpaceX could further optimize costs and improve mission efficiency.

Section 1

Methodology

Methodology - Summary

DATA COLLECTION

- Rocket launch data was obtained from the SpaceX API at: https://api.spacexdata.com/v4/launches/past
- A static response object was used for this project: https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json
- The data was filtered to include only Falcon 9 launches, excluding Falcon 1 records.
- Web scraping was performed using BeautifulSoup to extract Falcon 9 launch records from HTML tables on Wikipedia. The table was parsed and converted into a Pandas DataFrame.

Methodology - Summary

DATA WRANGLING

- After collecting the data and importing it into a data frame, we cleaned the data and replaced missing values with mean values.
- We calculated the number of launches on each site
- We calculated the number and occurrence of each orbit (each launch aims to a dedicated orbit
- We calculated the number and occurrence of each orbit
- We calculated the number and occurrence of mission outcome of the orbits

Methodology - Summary

EXPLORATORY DATA ANALYSIS (EDA)

- We performed an exploratory analysis to identify patterns and trends in the data.
- Visualizations were created to explore relationships between variables, such as launch sites, payload mass, and mission outcomes.
- SQL queries were used to extract insights from the dataset.

Methodology - Summary

INTERACTIVE VISUAL ANALYTICS

- Dashboards and interactive maps were created using Folium and Plotly Dash to provide insights into launch sites, orbits, and payload trends.

Methodology - Summary

PREDICTIVE ANALYSIS USING CLASSIFICATION MODELS

- Data was standardized and split into training and test sets.
- Hyperparameters were optimized for K-Nearest Neighbor, Support Vector Machine (SVM), Classification Tree, and Logistic Regression.
- All these 4 models reached a good level of accuracy on test data.

Data Collection

After importing the necessaries libraries, we defined a series of helper functions that helped us use the API to extract information using identification numbers in the launch data. These functions allowed us to fetch specific details about rockets, launch sites, payloads, and core stages directly from the SpaceX API. For instance, we used the rocket ID to retrieve the booster version, the launchpad ID to obtain geographic coordinates and site names, and payload IDs to gather information on mass and orbit. By automating these processes, we efficiently enriched the dataset with additional critical attributes, ensuring a comprehensive analysis of Falcon 9 launches.

Data Collection – SpaceX API

Libraries and Setup: Imported libraries like requests and pandas to manage API calls and process data.

Helper Functions: Defined custom functions to retrieve additional details from the SpaceX API:

getBoosterVersion: Fetches booster versions using rocket IDs.

getLaunchSite: Retrieves geographic details and site names from launchpad IDs.

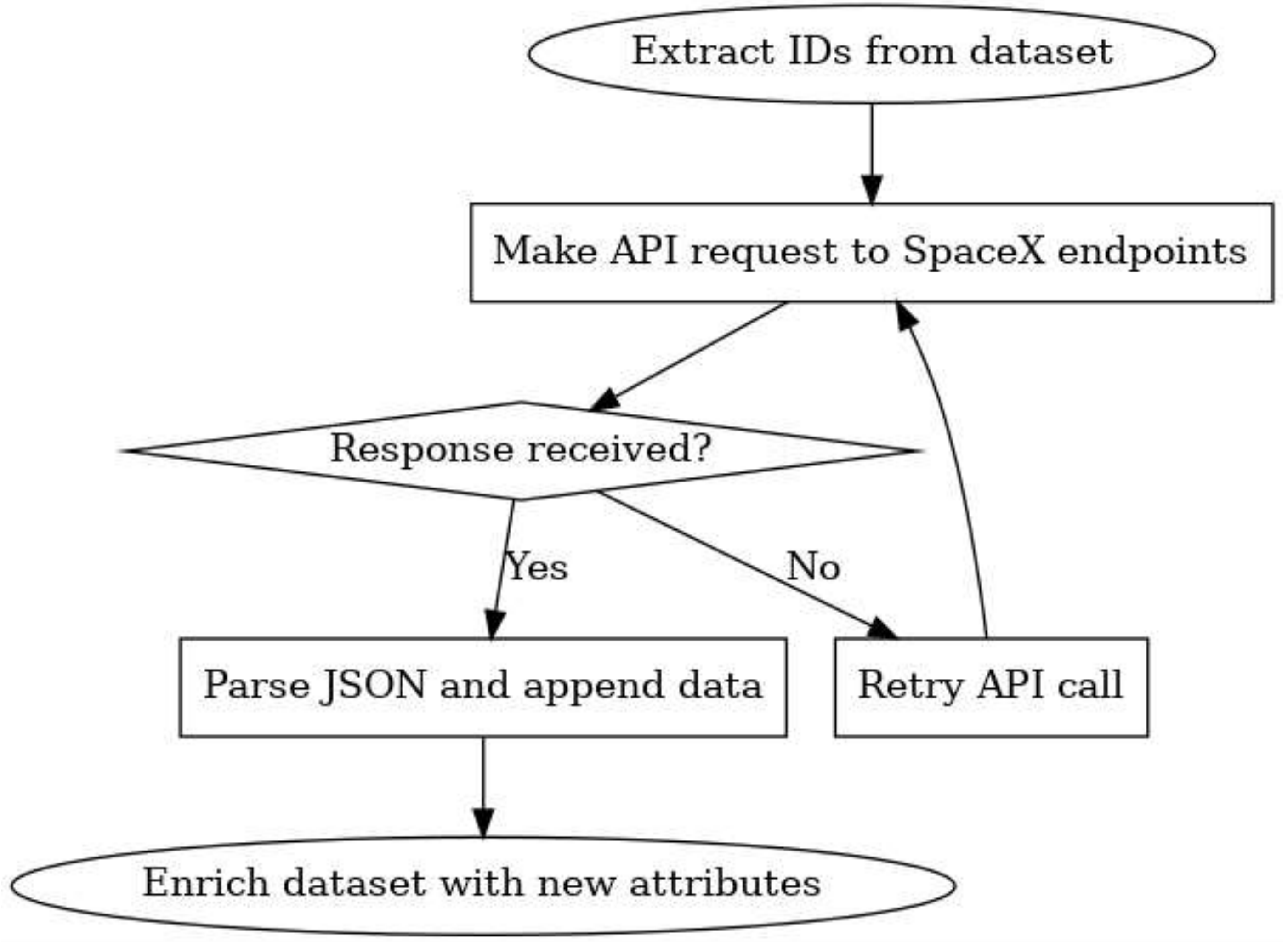
getPayloadData: Collects payload mass and orbit information from payload IDs.

getCoreData: Extracts core-related data such as block, reuse count, and landing outcomes.

Integration: Enriched the dataset by appending API responses, ensuring detailed attributes for each Falcon 9 launch.

GitHub Link: https://github.com/iddadobroc/SpaceX-Falcon-9-Analysis/blob/8246b3091c53bbef631e3401cfc016fe25df2561/Lab%201%20-%20Collecting%20the%20data.ipynb

Data Collection – SpaceX API Flowing Chart



Data Collection – Scraping

Objective: Collected Falcon 9 launch records from Wikipedia to enrich the dataset with comprehensive historical data.

Tools Used: Utilized BeautifulSoup from Python's bs4 library to parse and extract data from HTML tables.

Fetched Falcon 9 launch records table using requests to get the HTML source.

Parsed the HTML to locate the relevant table containing launch details.

Converted the table data into a Pandas DataFrame for analysis and integration.

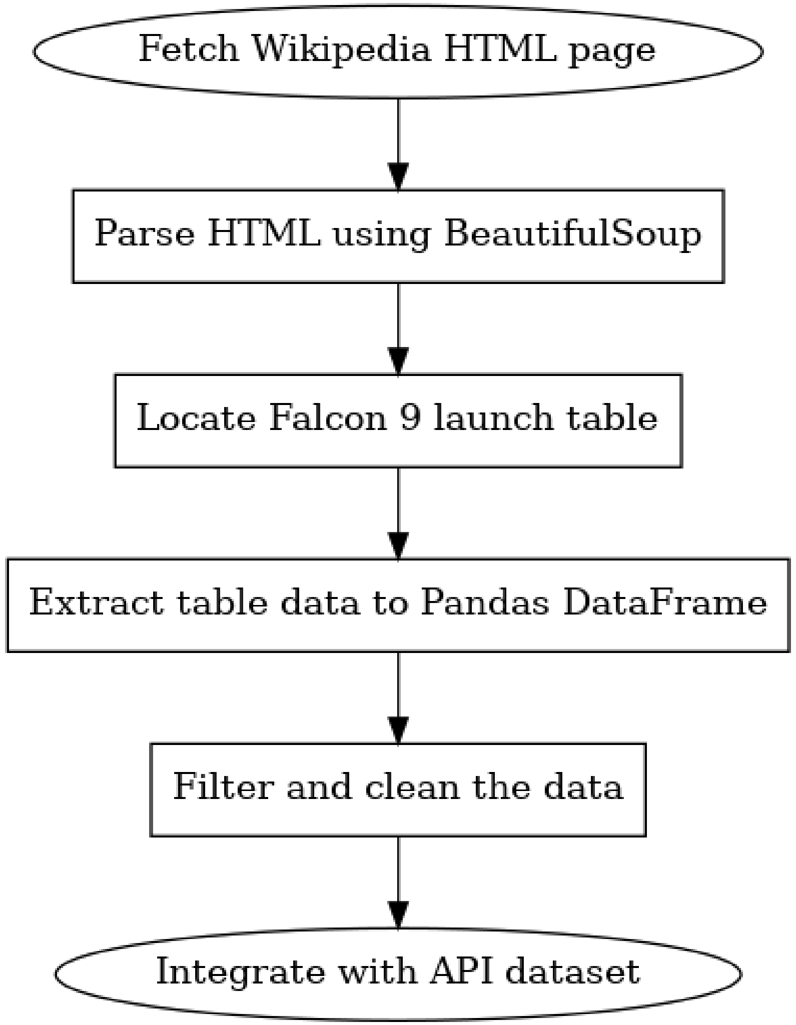
Removed non-Falcon 9 launches.

Standardized columns and handled missing data for accurate analysis.

Integration: Combined the scraped data with API-enriched datasets for a complete overview of Falcon 9 launches.

GitHub Link: https://github.com/iddadobroc/SpaceX-Falcon-9-Analysis/blob/8246b3091c53bbef631e3401cfc016fe25df2561/Web%20scraping%20Falcon%209%20and%20Falcon%20Heavy%20Launches%20Records%20from%20Wikipedia.ipynb

Data Collection – Scraping Flowing Chart



Data Wrangling

Objective: Cleaned and prepared the collected data for analysis, ensuring completeness, consistency, and usability.

→ **Handling Missing Values**: Identified and filled missing values in numeric columns using mean or median.

→ Categorical missing values were filled with the most frequent category or marked as "Unknown".

→ Removed non-relevant records, keeping only Falcon 9 launch data.

→ Filtered out launches with incomplete or inconsistent data.

→ Renamed columns for clarity and consistency.

→ Converted date and time columns to datetime format for easier manipulation.

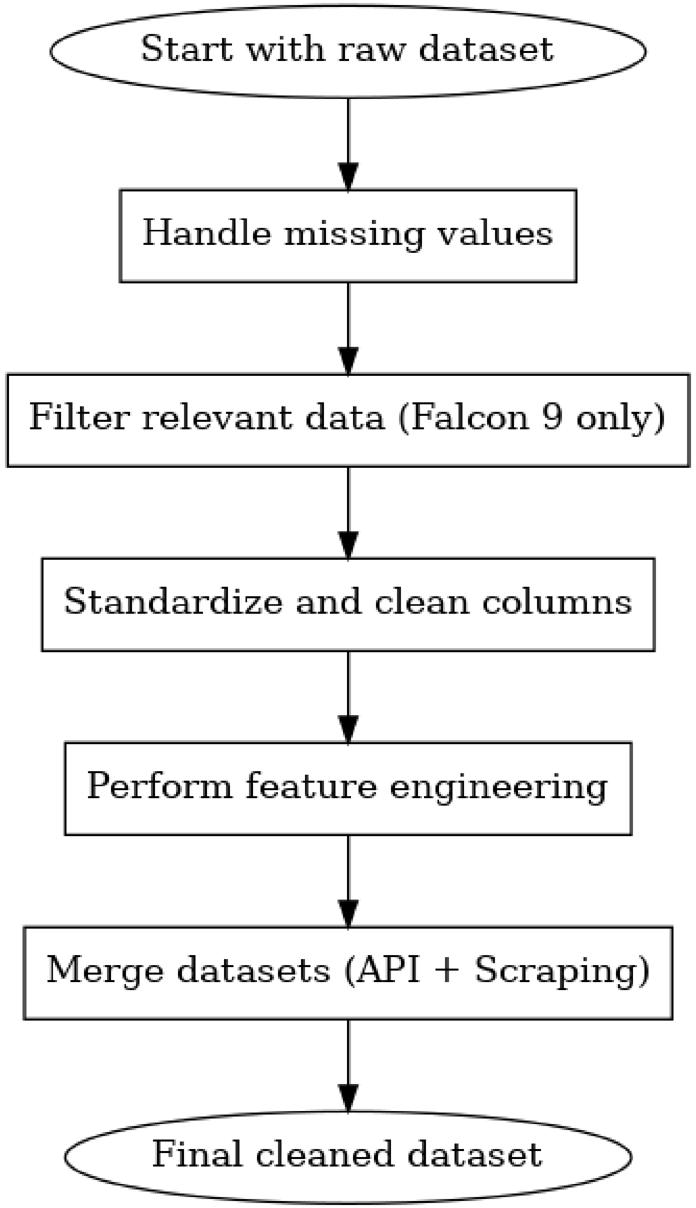
→ Created new attributes like Landing Success, combining core data fields.

→ Merged API-enriched and scraped datasets to create a unified dataframe.

→ **Outcome**: The dataset was ready for exploratory analysis and predictive modelling, with key attributes like booster versions, payload mass, launch sites, and outcomes accurately prepared.

→ **GitHub Link**: https://github.com/iddadobroc/SpaceX-Falcon-9-Analysis/blob/8246b3091c53bbef631e3401cfc016fe25df2561/Lab%202%20-%20Data%20wrangling.ipynb

Data Wrangling Flowing Chart



EDA with Data Visualization

- **Objective**: To explore trends and relationships in the dataset and uncover insights about Falcon 9 launches.
- **Key Visualizations:**
  - **Scatter Plot**: Flight Number vs. Launch Site (examine the distribution of launches across different sites and observe patterns related to launch frequency).
  - **Scatter Plot**: Payload Mass vs. Launch Outcome (identify how payload mass affects the success of first-stage landings):
  - **Bar Chart**: Success Rate by Orbit Type (analyze the success rate of launches based on orbit types).
  - **Line Chart**: Yearly Average Success Rate (observe trends in launch success over time and identify improvements in technology or processes).
  - **Pie Chart**: Launch Success by Site (visualize the success ratio at each launch site and highlight the most reliable locations).
- **Outcome**: These visualizations provided actionable insights, such as identifying payload mass ranges that increase the likelihood of success and launch sites with the highest success ratios.
- **GitHub Link**: https://github.com/iddadobroc/SpaceX-Falcon-9-Analysis/blob/56f88a8fdcbc18e73d645d76216b92f81b739a9c/EDA%20with%20Data%20Visualization.ipynb

EDA with SQL

- **Key Queries Performed**:
  - **Unique Launch Sites:**
    - **Query**: Identified all unique launch site names.
    - **Purpose**: To understand the distribution and variety of launch locations.
  - **Payload Statistics:**
    - **Query**: Calculated the total and average payload mass for specific booster versions.
    - **Purpose**: To evaluate the efficiency and capabilities of different boosters.
  - **Mission Outcomes:**
    - **Query**: Counted the total number of successful and failed missions.
    - **Purpose**: To assess the overall success rate of Falcon 9 launches.
  - **Landing Outcome Trends:**
    - **Query**: Ranked landing outcomes (e.g., success on ground pads, failure on drone ships) within specific date ranges.
    - **Purpose**: To identify patterns in landing success over time.
  - **Specific Payload Analysis:**
    - **Query**: Listed boosters that successfully landed on drone ships with payload mass between 4000 and 6000 kg.
    - **Purpose**: To examine specific payload ranges associated with success.
  - **Historical Success:**
    - **Query**: Retrieved the date of the first successful ground pad landing.
    - **Purpose**: To highlight key milestones in SpaceX's history.
- **Outcome**: These queries provided critical insights into launch site performance, payload efficiency, and mission success trends.
- **GitHub Link**: https://github.com/iddadobroc/SpaceX-Falcon-9-Analysis/blob/a1e0d36ee460f747ec106df9ae55cafa819947ea/SQL%20Notebook%20for%20Peer%20Assignment.ipynb

Build an Interactive Map with Folium

- **Map Objects Created:**
  - **Markers for Launch Sites**: Added markers for each Falcon 9 launch site with popup labels indicating the site name and coordinates (provide a quick visual reference for the location and distribution of SpaceX's launch sites).
  - **Circles Indicating Proximity**: Placed circles around each launch site with a radius to indicate proximity to nearby infrastructure, such as highways and coastlines (analyze accessibility and logistical advantages of the sites).
  - **Color-Coded Outcomes**: Used color-coded markers to represent the success or failure of launches from each site (visually correlate launch outcomes with geographic factors).
  - **Lines Connecting Launch Sites to Landing Pads**: Drew lines from launch sites to landing pads for selected missions (illustrate the relationship between launch and landing locations).
- **Outcome**: The interactive map allowed for a comprehensive geographic analysis, highlighting trends in launch site success and identifying potential logistical advantages.
- GitHub Link: https://nbviewer.org/github/iddadobroc/SpaceX-Falcon-9-Analysis/blob/8d16328591234a91127f6b41b238504507939486/Interactive%20Visual%20Analytics%20with%20Folium.ipynb

Build a Dashboard with Plotly Dash

Summary of Plots/Graphs and Interactions

Pie Chart: Total Successful Launches by Site

- **Interaction**:
  - Dropdown menu to select a specific launch site or view aggregated data for all sites.
  - The chart updates dynamically based on the selected site.
- **Purpose**:
  - Highlights the distribution of successful missions across various launch sites.
  - Provides an immediate overview of site-specific performance.

Scatter Plot: Payload vs. Success

- **Interaction**:
  - Slider to filter data based on the payload mass range.
  - Dropdown menu to select a specific launch site.
  - The chart dynamically updates based on applied filters.
- **Purpose**:
  - Analyzes the relationship between payload mass and launch success.
  - Displays the impact of booster version on the probability of success.

Why These Plots and Interactions Were Added

Pie Chart:

- Offers a clear visualization of total successful launches and allows for comparisons between different launch sites.
- Helps identify the most reliable or highest-performing sites.

Scatter Plot:

- Provides detailed insights into the correlation between payload, booster versions, and launch outcomes.
- Helps identify payload limits or specific patterns influencing success.

GitHub Link:

- **https://github.com/iddadobroc/SpaceX-Falcon-9-Analysis/blob/a00a0b725299f24003794e23a6f1cca130a2ad26/spacex_dash_app.py**

Predictive Analysis (Classification)

Model Development Process

Data Preparation:

- Extracted the target variable (Y) from the Class column.
- Standardized the feature matrix (X) using StandardScaler.

Data Splitting:

- Split the dataset into training (80%) and test (20%) sets using train_test_split.

Model Selection and Optimization:

- Evaluated four classification models:
  - Logistic Regression
  - Support Vector Machine (SVM)
  - Decision Tree
  - K-Nearest Neighbors (KNN)
- Used GridSearchCV with cross-validation (cv=10) to tune hyperparameters for each model.

Evaluation:

- Measured accuracy, precision, and recall for each model.
- Used the confusion matrix to analyze errors (e.g., false positives).

Results

- Best Parameters Found:
  - Logistic Regression: {'C': 0.01, 'penalty': 'l2', 'solver': 'lbfgs'}
  - SVM: {'C': 1.0, 'gamma': 0.031, 'kernel': 'sigmoid'}
  - Decision Tree: {'criterion': 'gini', 'max_depth': 16, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 2, 'splitter': 'random'}
  - KNN: {'algorithm': 'auto', 'n_neighbors': 10, 'p': 1}

Test Accuracy:

- Logistic Regression: 83.3%
- SVM: 83.3%
- Decision Tree: 83.3%
- KNN: 83.3%

Conclusion:

- All models performed equally well with identical accuracy scores.
- Model selection can depend on interpretability or computational efficiency.

GitHub Link:

- https://github.com/iddadobroc/SpaceX-Falcon-9-Analysis/blob/a00a0b725299f24003794e23a6f1cca130a2ad26/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

Exploratory Data Analysis Results

Success Rate by Orbit Type:

- The success rate varies across orbit types, with some orbits (e.g., ES-L1, GEO, HEO) achieving a 100% success rate (due to fewer launches), while others (e.g., GTO) have a lower success rate.
- Insights can help SpaceX focus on improving performance for less successful orbits.

Flight Number vs Launch Site:

- The scatter plot shows the distribution of flight numbers across different launch sites.
- Launches are concentrated around specific sites (e.g., CCAFS SLC 40 and KSC LC 39A).

Flight Number vs Orbit Type:

- Different orbit types are distributed across flight numbers, showing SpaceX's growing versatility over time.

Flight Number vs Payload Mass:

- Payload mass increases with flight numbers, indicating advancements in launch capabilities.

Payload Mass vs Launch Site:

- Payload mass varies significantly across launch sites, with some sites handling heavier payloads than others.

Payload Mass vs Orbit Type:

- Specific orbit types correspond to higher payload masses, reflecting unique mission requirements.

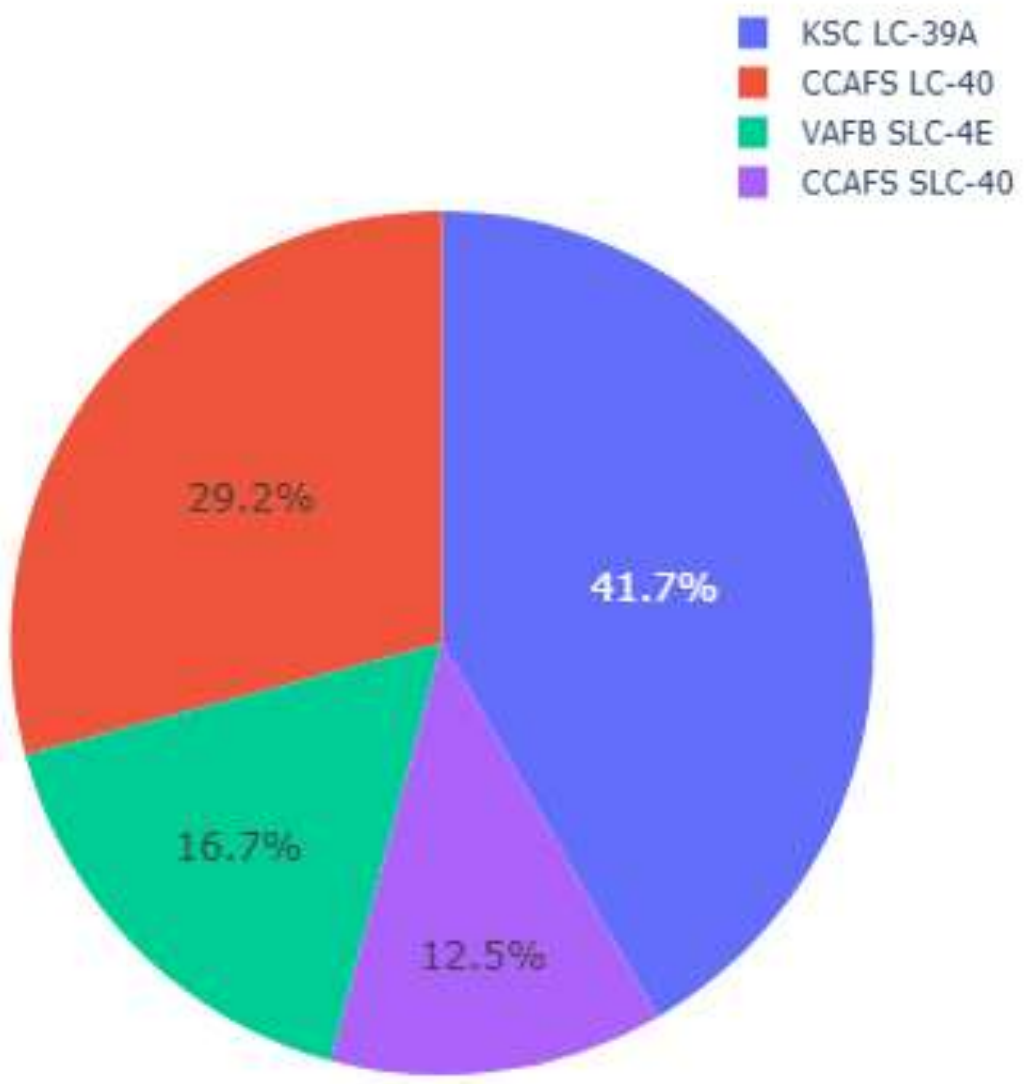
Yearly Launch Success Rate Trend:

- SpaceX's yearly success rate shows a consistent improvement over time, demonstrating technological advancements and operational efficiency.

Interactive Analytics Demo in Screenshots

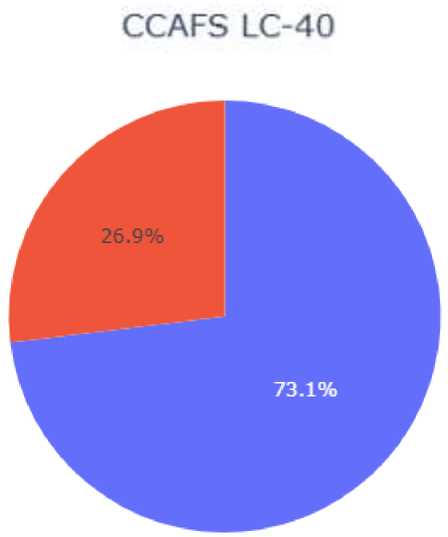
Total Success Launches by Site

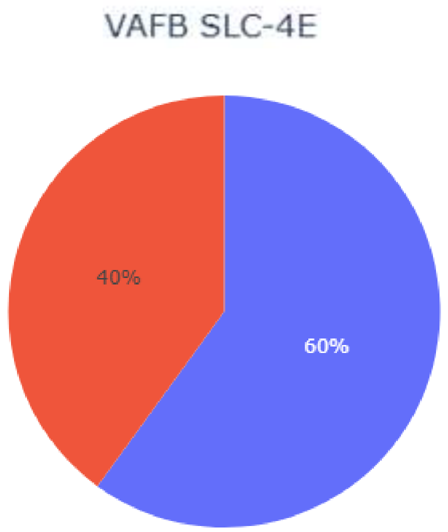
A pie chart showing the distribution of total successful launches across all launch sites. It highlights the contribution of each site to SpaceX's overall success.

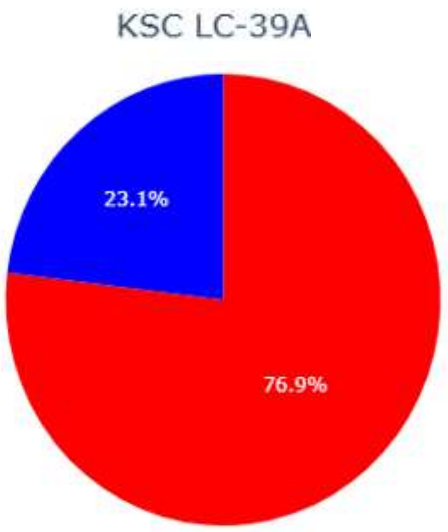


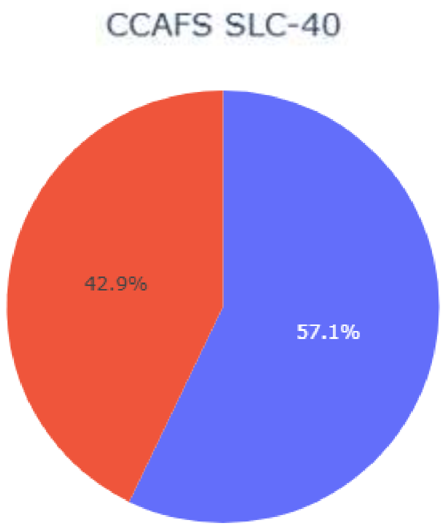
Interactive Analytics Demo in Screenshots

Site-specific Success and Failure Rates









Separate pie charts for each site showcasing the success and failure rates for launches from these locations.

Success

Failure

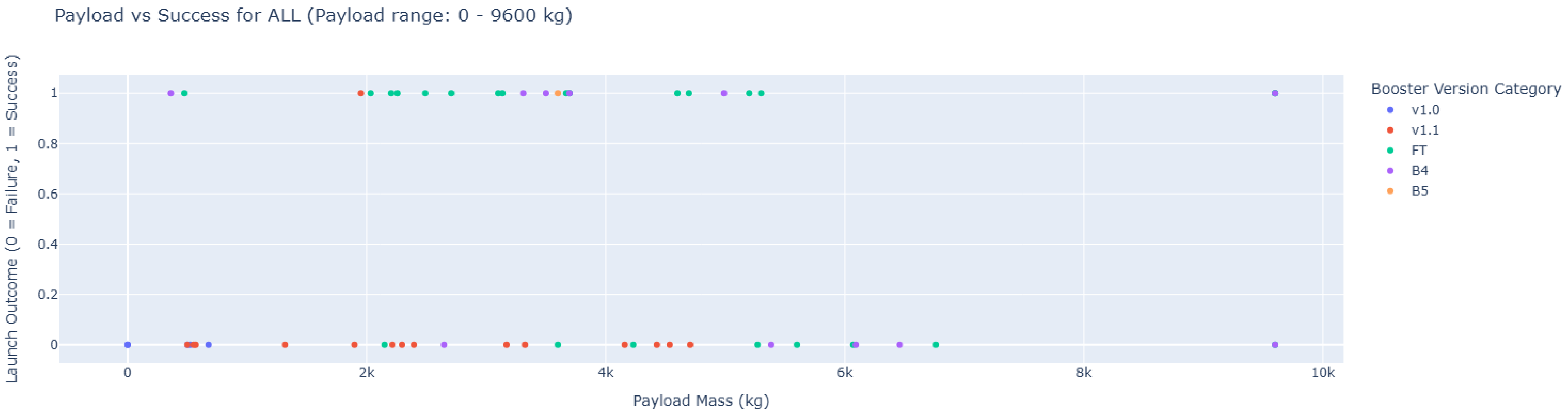
Interactive Analytics Demo in Screenshots

Payload vs Success Correlation:

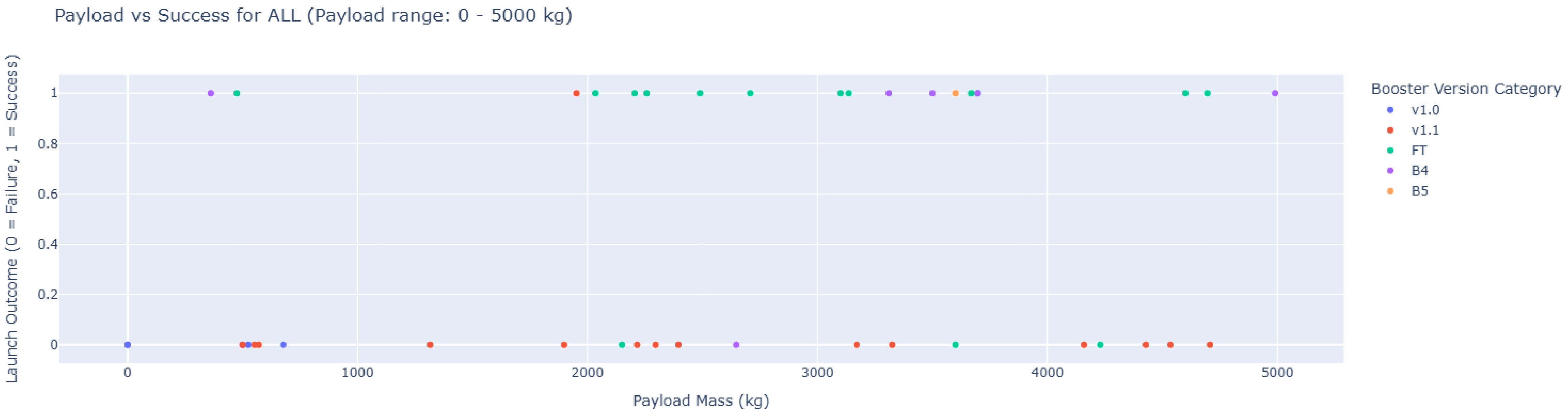
Scatter plots analyzing the relationship between payload mass and launch outcomes.

Different payload ranges (e.g., 0-5000 kg, 5000-10000 kg) are analyzed to understand how payload mass impacts success rates.

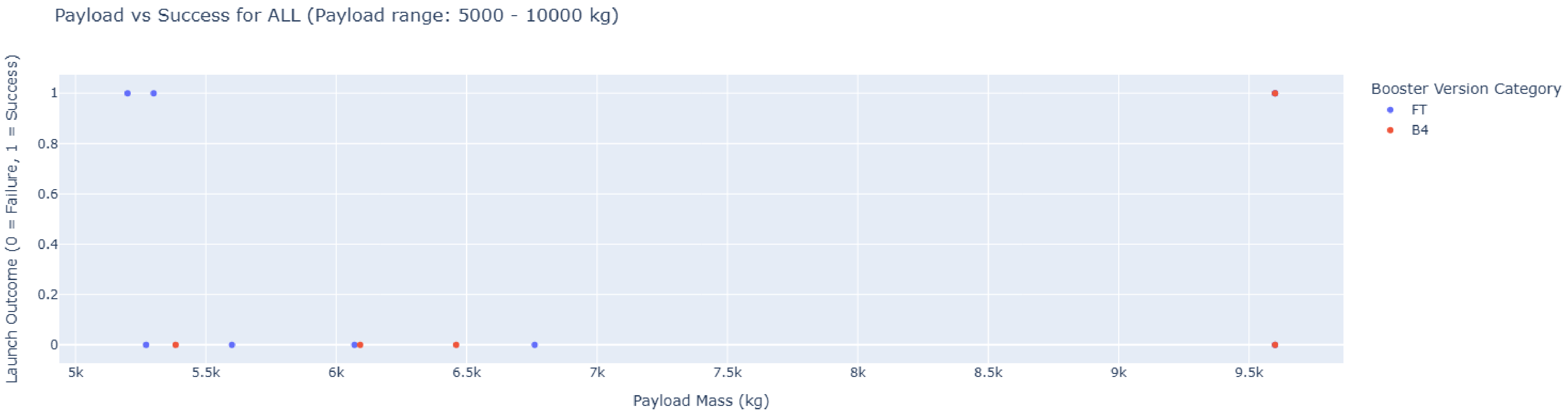
Scatter Plot: Payload vs Success (All Sites)



Scatter Plot: Payload vs Success (0–5000 kg)



Scatter Plot: Payload vs Success (5000–10,000 kg)



Predictive Analysis Results

- **Models Evaluated**:
  - Four classification models were developed and tested: Logistic Regression, Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors (KNN).
- **Model Performance**:
  - All models achieved an accuracy of 83.3% on the test dataset.
  - The results were validated using cross-validation (cv=10) and confusion matrices.
- **Insights**:
  - Logistic Regression with tuned hyperparameters showed strong interpretability and consistent results.
  - SVM and Decision Tree models provided flexibility with different hyperparameter configurations.
  - KNN demonstrated competitive performance but required careful tuning of the number of neighbors.
- **Conclusion**:
  - Any of the models can be used for prediction due to identical accuracy scores.
  - Future work can focus on larger datasets or ensemble methods for further improvement.

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

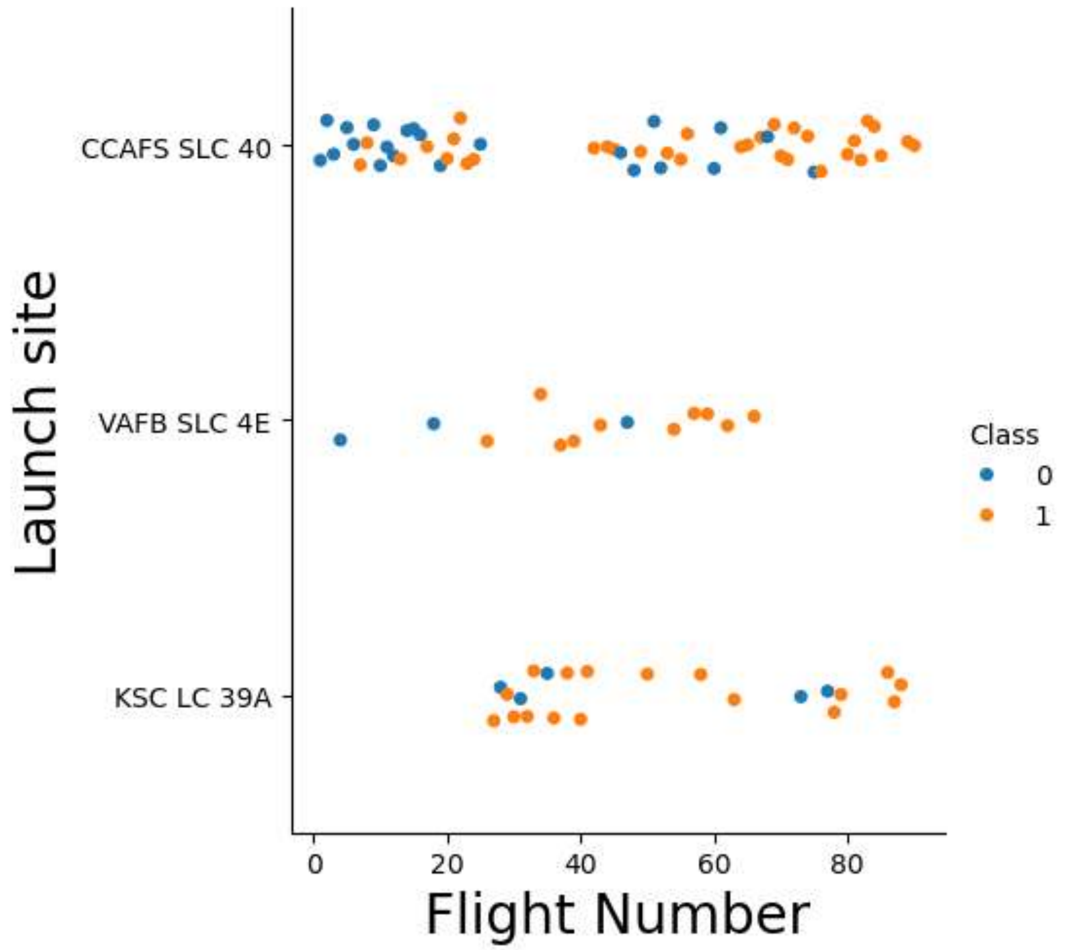
The scatter plot showcases the relationship between flight numbers and their respective launch sites. Each point represents a single launch, color-coded by its outcome:

- **Orange (1)**: Represents successful launches.
- **Blue (0)**: Represents failed launches.

Key Insights:

- **KSC LC-39A** has a high density of successful launches compared to other sites, indicating consistent performance as flight numbers increase.
- **CCAFS LC-40** shows a mix of successes and failures, particularly in earlier flight numbers, but demonstrates improvement in outcomes over time.
- **VAFB SLC-4E** has relatively fewer launches, with successes becoming more frequent in later flights.

This scatter plot reveals trends in launch success by location, highlighting improvement in reliability as SpaceX gained experience.



Payload vs. Launch Site

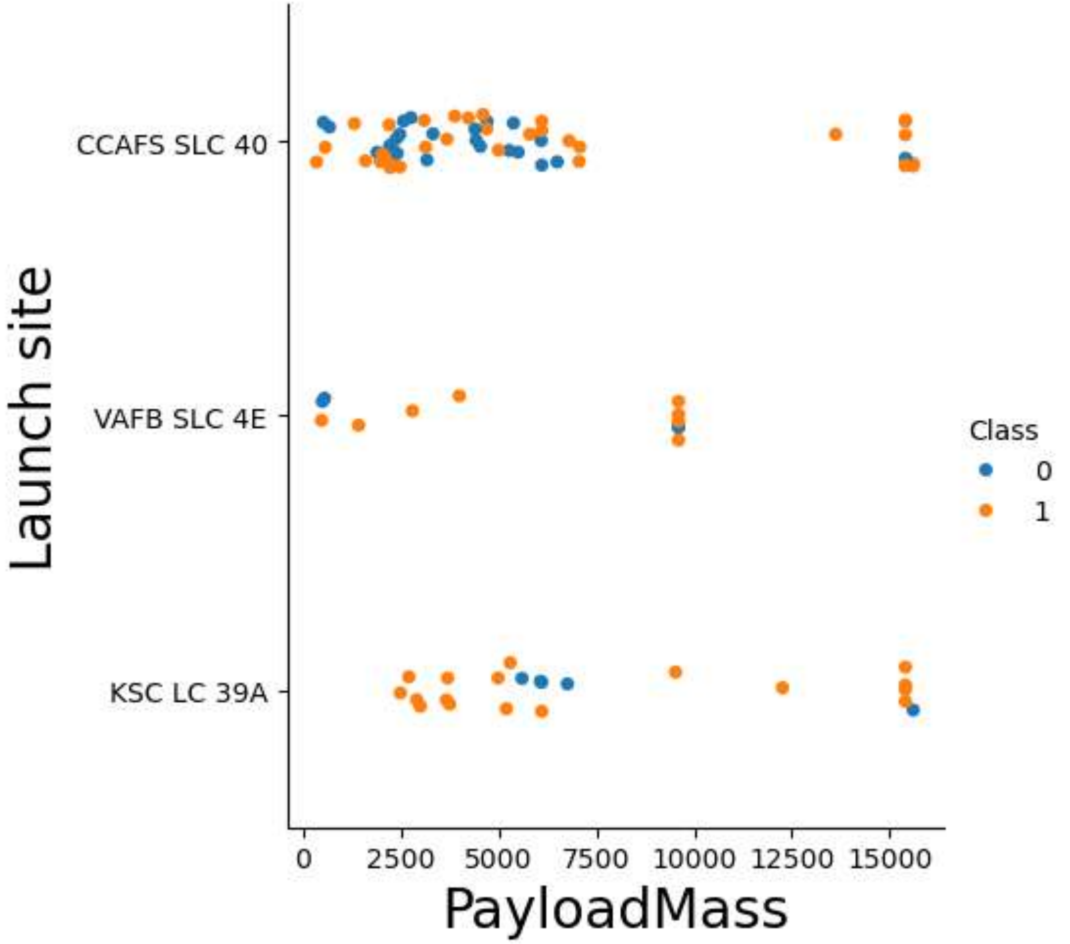
The scatter plot illustrates the relationship between payload mass and launch outcomes across different launch sites. Each point corresponds to a launch, with color-coded outcomes:

- **Orange (1)**: Successful launches.
- **Blue (0)**: Failed launches.

Key Insights:

- **CCAFS SLC-40** and **KSC LC-39A** handle a broader range of payloads, from light to heavy, with a noticeable increase in successful launches for payloads under 10,000 kg.
- **VAFB SLC-4E** primarily manages heavier payloads, with a mix of successes and failures.
- Successful launches appear more clustered within specific payload ranges, indicating a potential sweet spot for reliability.

This plot emphasizes the distribution of payload capacities across sites, showing patterns of reliability for different payload masses.



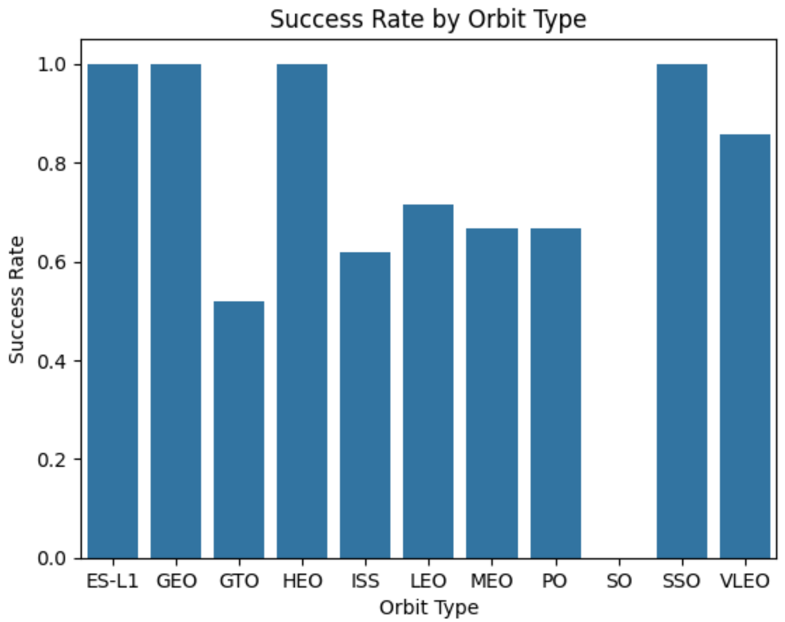
Success Rate vs. Orbit Type & Flight Number vs. Orbit Type

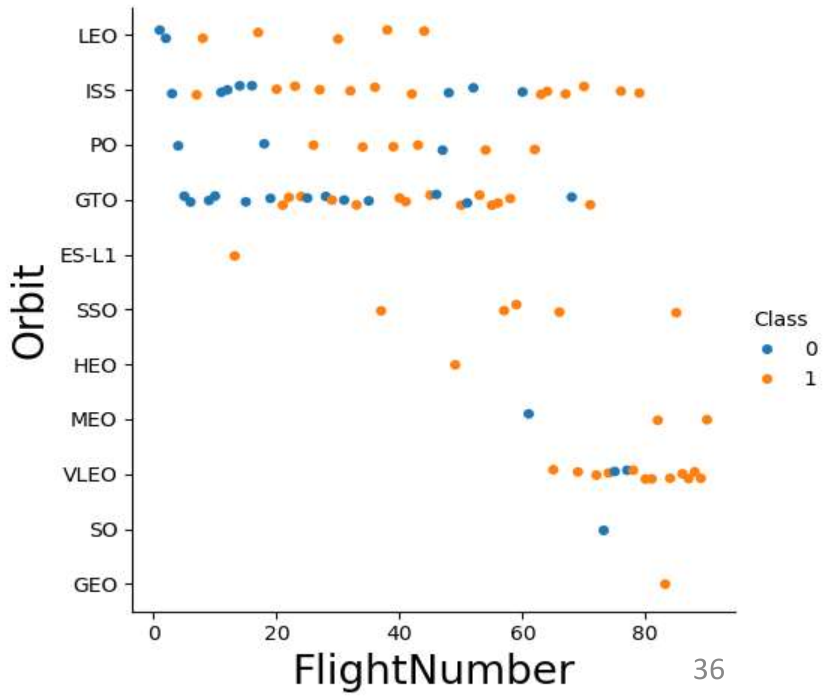
- **Bar Chart: Success Rate by Orbit Type**

The bar chart illustrates the success rate of SpaceX launches for different orbit types. Orbits such as ES-L1, GEO, HEO, and SSO achieved a 100% success rate, highlighting their reliability for successful launches, although the number of launches on these orbits is significantly lower than others as GTO, that conversely displayed a relatively lower success rate, suggesting potential challenges or complexities associated with this orbit type.

- **Scatter Plot: Flight Number vs. Orbit Type**

The scatter plot shows the relationship between flight numbers, orbit types, and launch outcomes (success or failure). Each point is colored based on the launch outcome, with orange representing success and blue indicating failure. It can be observed that certain orbits, such as VLEO and SSO, tend to have a higher density of successful launches, regardless of the flight number. This pattern underscores the influence of orbit type on mission success rates.



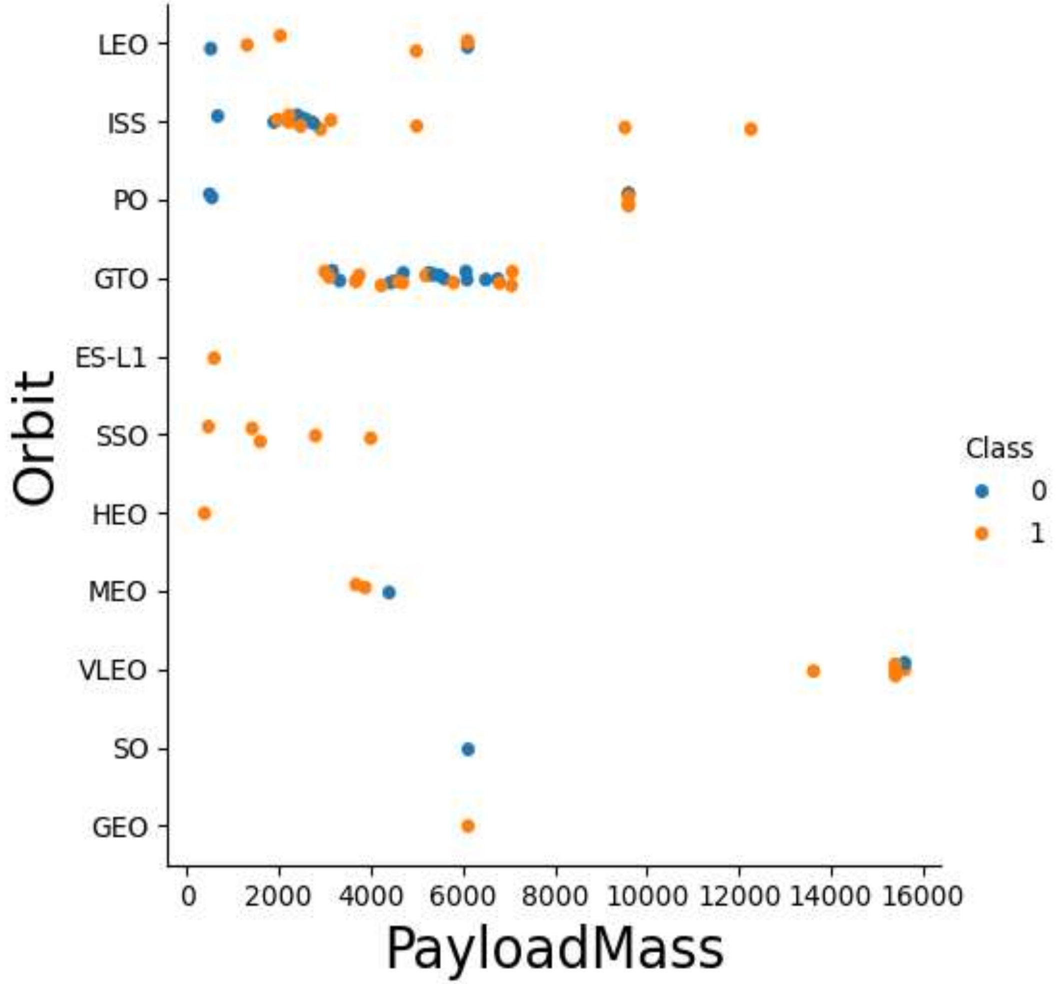


Payload vs. Orbit Type

The scatter plot visualizes the relationship between payload mass and orbit type, with the outcomes (success or failure) indicated by color. It reveals that certain orbit types, such as GTO and SSO, are chosen within a certain range of payload mass. In the case of GTO orbit, it seems payload doesn't affect the success or failure of the mission.

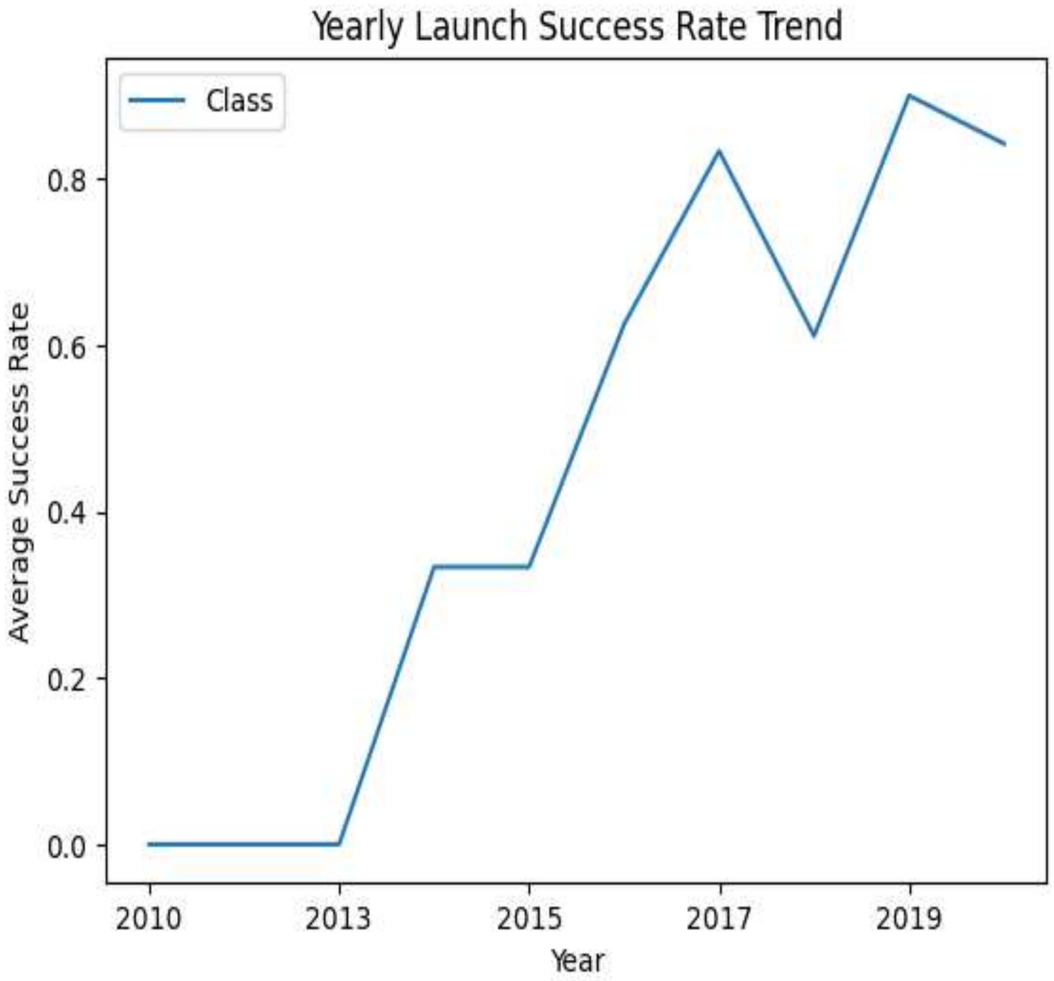
Additionally, orbits like ISS demonstrate a broader range of payload mass, suggesting its suitability for different type of launches and mission requirements. Furthermore, it seems that heavier payloads have a higher success rate on ISS orbit.

The data highlights how payload mass and orbit type collectively influence mission success.



Launch Success Yearly Trend

The line chart illustrates the yearly trend in SpaceX's launch success rate. Starting from 2010, the success rate remained at zero for the initial years but began to increase significantly from 2013 onward. The trend shows consistent improvement over time, reaching a peak around 2018-2019. This upward trajectory indicates advancements in technology, processes, and mission execution, reflecting SpaceX's growing reliability and expertise in achieving successful launches.



All Launch Site Names

The SQL query **SELECT DISTINCT Launch_Site FROM SPACEXTABLE;** retrieves the unique names of all launch sites from the dataset. This helps identify the various locations used by SpaceX for its launches.

The unique launch sites identified are:

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

These launch sites represent SpaceX's key operational locations, showcasing their diverse geographic strategy for accommodating different mission profiles.

Launch Site Names Begin with 'CCA'

The SQL query used to find records where the launch site names begin with "CCA" is:

SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE 'CCA%' LIMIT 5;

This query retrieves the first 5 records from the dataset where the Launch_Site starts with the prefix "CCA". The % wildcard is used to match any number of characters following "CCA".

The results show that **Cape Canaveral Air Force Station (CCAFS)** is one of the main launch sites of SpaceX. This indicates SpaceX's significant operational presence at this location, utilizing different launch pads like **LC-40** and **SLC-40** for their missions.

[7]:

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

SUM (PAYLOAD_MASS__KG_)
45596

Total Payload Mass

The SQL query used to calculate the total payload mass carried by boosters for NASA is:

SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE Customer = 'NASA (CRS)';

This query calculates the sum of the PAYLOAD_MASS__KG_ column, filtering only those rows where the Customer is "NASA (CRS)", which refers to NASA's Commercial Resupply Services missions.

This result highlights the significant payload mass carried by SpaceX boosters for NASA's resupply missions to the International Space Station (ISS). It demonstrates SpaceX's role in fulfilling NASA's requirements for reliable and efficient cargo transportation to space.

Average Payload Mass by F9 v1.1

AVG (PAYLOAD_MASS__KG_)
2928.4

The SQL query used to calculate the average payload mass carried by the booster version F9 v1.1 is:

SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE Booster_Version = 'F9 v1.1’;

This query calculates the average (AVG) of the PAYLOAD_MASS__KG_ column, filtering only the rows where the Booster_Version is 'F9 v1.1'.

This result represents the average mass of payloads delivered by the F9 v1.1 booster version. It reflects the typical capability of this specific booster version in transporting payloads to various orbits during its operational lifetime. This information is crucial for evaluating performance and planning future missions.

First Successful Ground Landing Date

The SQL query used to determine the date of the first successful landing outcome on a ground pad is:

SELECT MIN(DATE) FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (ground pad)’;

This query retrieves the earliest (MIN) date from the DATE column, filtering the records where the Landing_Outcome is 'Success (ground pad)’.

This date represents the first time SpaceX achieved a successful landing of a booster on a ground pad. This milestone was crucial in SpaceX's journey toward reusable rockets, significantly reducing costs and paving the way for sustainable space exploration.

MIN(DATE)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

The SQL query used to list the boosters with successful drone ship landings and payloads between 4000 and 6000 kg is:

SELECT Booster_Version FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000;

This query filters records where:

- Landing_Outcome is 'Success (drone ship)', indicating a successful drone ship landing.
- PAYLOAD_MASS__KG_ is greater than 4000 and less than 6000.

The boosters listed here achieved successful drone ship landings while carrying payloads in the specified range. These launches highlight the precision and reliability of SpaceX's Falcon 9 in handling substantial payloads and achieving successful recoveries at sea.

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

SQL Query:

SELECT Mission_Outcome, COUNT(*) FROM SPACEXTABLE GROUP BY Mission_Outcome;

The query calculates the total number of occurrences for each mission outcome by grouping the data based on the Mission_Outcome column and counting the entries for each unique outcome. The result provides a breakdown of how many missions were successful and how many failed, offering insight into SpaceX's historical mission performance.

Mission_Outcome	COUNT(*)
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

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

Boosters Carried Maximum Payload

SQL Query:

SELECT Booster_Version FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE);

This query identifies the booster(s) that carried the maximum payload mass by first determining the highest payload mass using a subquery (SELECT MAX(PAYLOAD_MASS__KG_)). It then retrieves the Booster_Version associated with this maximum payload value from the SPACEXTABLE. The result highlights the most capable booster(s) in terms of payload capacity in SpaceX's history.

2015 Launch Records

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

SQL Query:

SELECT substr(Date, 6, 2) AS Month, Landing_Outcome, Booster_Version, Launch_Site FROM SPACEXTABLE WHERE substr(Date, 0, 5) = '2015' AND Landing_Outcome = 'Failure (drone ship)';

This query retrieves the records for failed landing outcomes on a drone ship in the year 2015. It uses the substr function to filter rows where the Date column starts with '2015' and where the Landing_Outcome is 'Failure (drone ship)'. The query also selects specific columns: Month (extracted from the date), Landing_Outcome, Booster_Version, and Launch_Site. The result highlights the boosters, launch sites, and months associated with these failed landings.

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

SQL Query:

```
SELECT Landing_Outcome, COUNT(*) FROM SPACEXTABLE WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome ORDER BY COUNT(*) DESC;
```

This query calculates the total count of each type of Landing_Outcome within the date range 2010-06-04 and 2017-03-20. It groups the results by Landing_Outcome, and orders them in descending order based on their count. This helps identify the most frequent landing outcomes during this time frame. The result ranks outcomes by their frequency of occurrence.

Landing_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

Section 3

Launch Sites Proximities Analysis

Launch Sites Geographical Distribution

The map highlights the geographical distribution of SpaceX launch sites. Key observations include:

- **Major concentration in Florida:** Three out of the four launch sites are located in Florida, demonstrating its strategic importance for SpaceX operations. Two of these sites are at Cape Canaveral (CCAFS LC-40 and CCAFS SLC-40), and one is located at the Kennedy Space Center (KSC LC-39A). Florida's proximity to the equator allows for efficient launches into a variety of orbital inclinations.
- **California site:** The fourth launch site, VAFB SLC-4E, is located at Vandenberg Space Force Base in California. This site specializes in polar and sun-synchronous orbit launches due to its advantageous location on the West Coast.

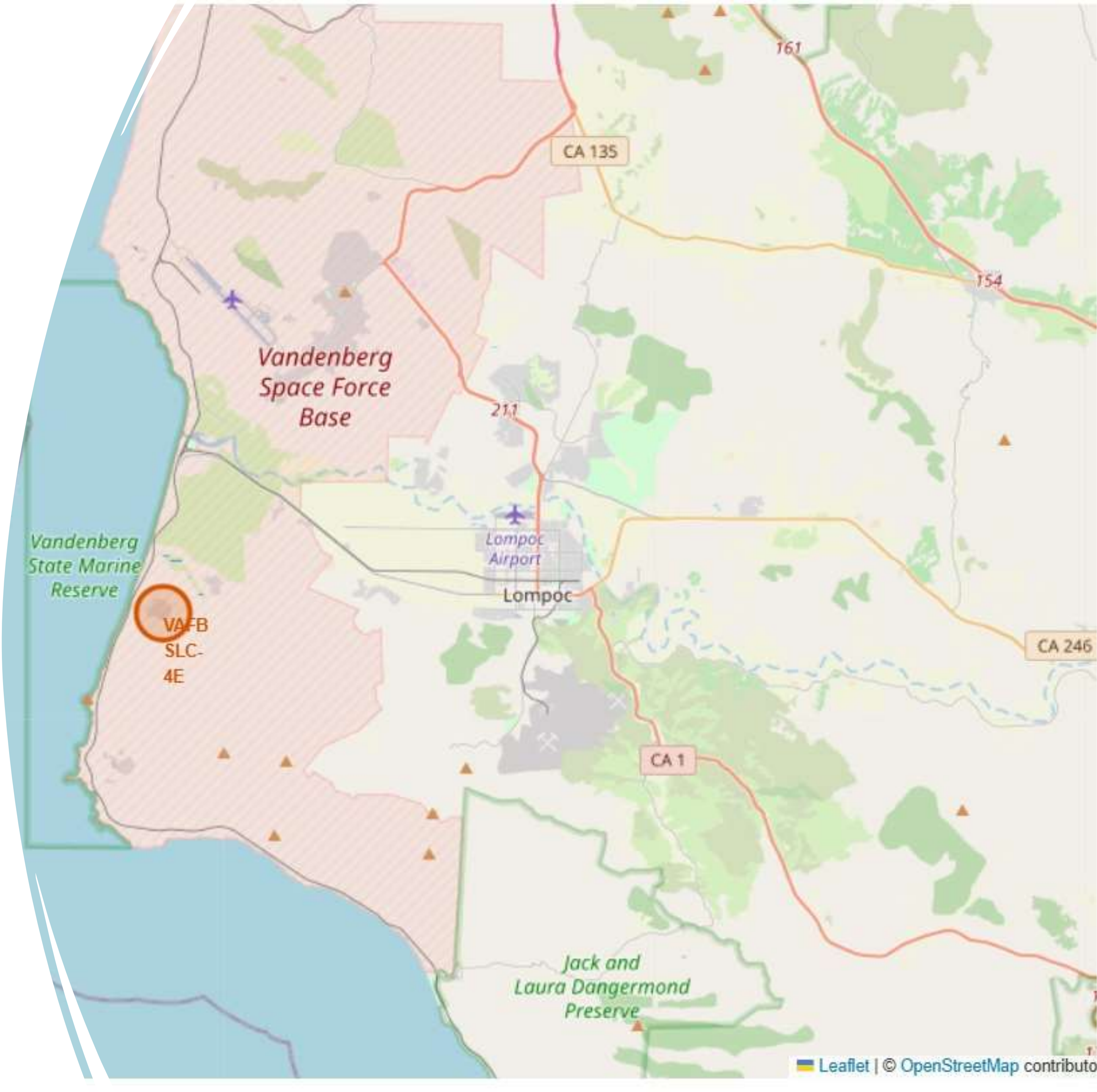
These proximities emphasize the strategic positioning of SpaceX's launch sites for diverse mission profiles.

Florida Launch Sites

As we can see, in few kilometers we have 3 launch sites: 2 are located in Cape Canaveral (CCAFS LC-40 and CCAFS SLC-40) and 1 in Kennedy Space Center (KSC LC-39A)

California Launch Site

In California, we only have one Launch Site: Vandenberg Space Launch Complex 4 (VAFB SLC-4E)



Launch Outcome Analysis by Site

The four screenshots illustrate the launch outcomes color-coded by site on the folium map:

- **KSC LC-39A**: Most markers are green, indicating a higher success rate. This makes KSC LC-39A a reliable site for launches.
- **VAFB SLC-4E**: Displays a good balance between successful and failed launches. The site appears critical for missions requiring polar orbits due to its geographical location.
- **CCAFS LC-40**: Launch outcomes are predominantly marked in red, indicating a higher rate of failed outcomes compared to successes, which are marked in green. This suggests challenges in this site's operational history.
- **CCAFS SLC-40**: Shows a relatively balanced mix of green and red markers, suggesting a moderate success rate in recent launches.

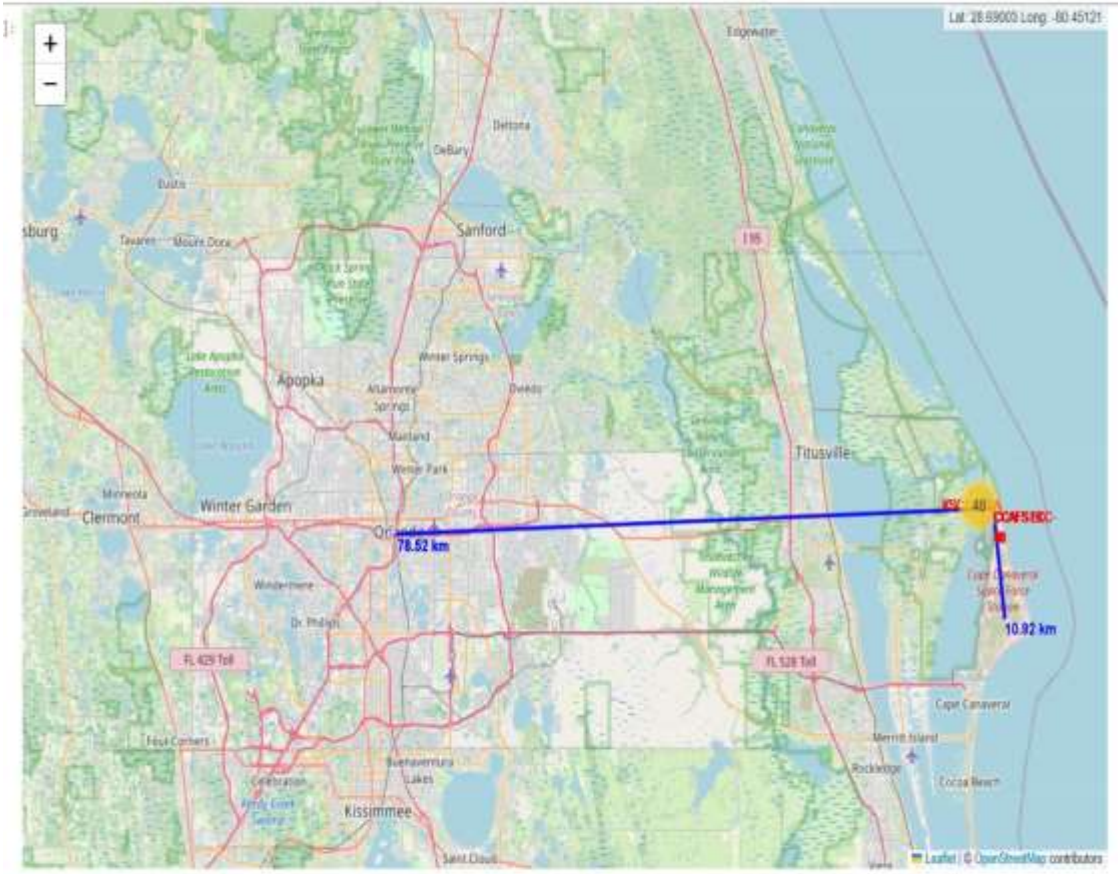
The analysis emphasizes how launch outcomes vary by site, with Florida launch sites (CCAFS and KSC) contributing significantly to both successes and failures, while Vandenberg complements the SpaceX network for specific mission types.

Proximity of Cape Canaveral Launch Sites to Major Infrastructures

This screenshot highlights the strategic location of Cape Canaveral Launch Sites in Florida, showcasing their proximity to essential infrastructures:

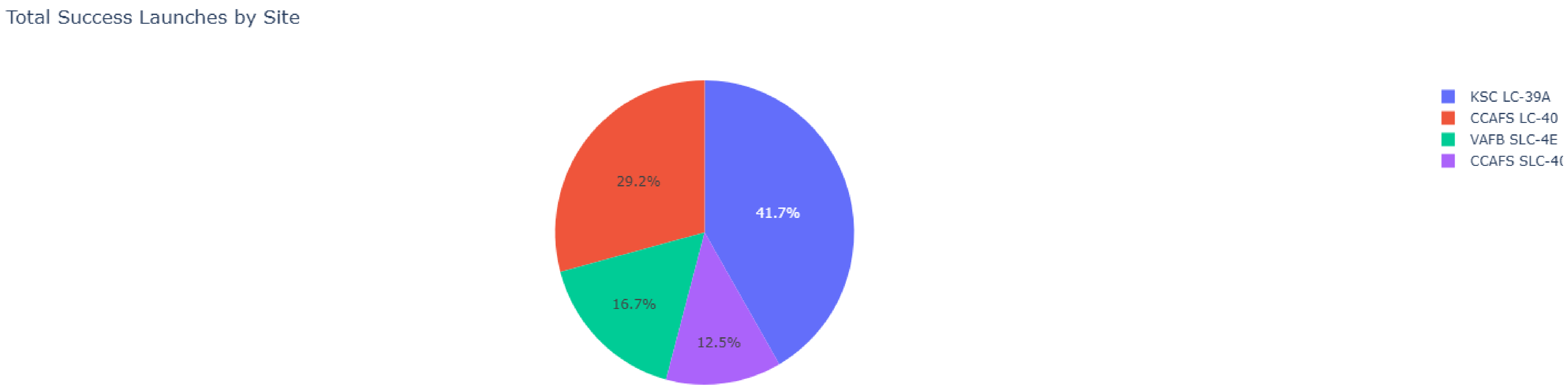
- **Distance to Orlando**: The map measures a straight-line distance of approximately 78.52 km from the Cape Canaveral Space Force Station to Orlando, a major urban center and transportation hub in Florida.
- **Proximity to Railways and Highways**: The launch sites are conveniently close to railway lines and highways, facilitating the transportation of heavy equipment and materials essential for launch operations.
- **Coastline Accessibility**: The sites are located near the Atlantic Ocean coastline, enabling the safe recovery of reusable rocket components like boosters.

The map underscores the deliberate selection of launch site locations to balance operational efficiency with accessibility to urban resources and logistical infrastructure.



Section 4

Build a Dashboard with Plotly Dash

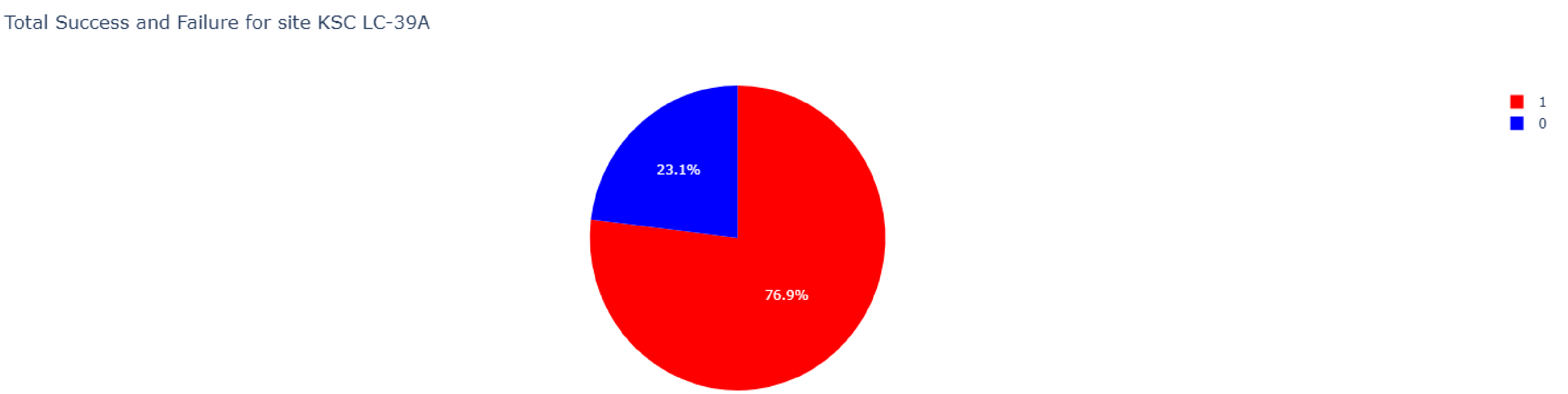


Interactive Dashboard: Launch Success Count for all Sites

This slide highlights the interactive dashboard built using Plotly Dash, showcasing the distribution of total successful launches across all SpaceX launch sites in a dynamic pie chart.

- **Insights:** The chart demonstrates that KSC LC-39A has the highest proportion of successful launches (41.7%), followed by CCAFS LC-40 (29.2%). This provides a clear view of each site's contribution to SpaceX's overall success.
- **Practical Application:** This dashboard enables stakeholders to explore data interactively, facilitating targeted decision-making based on visual insights.
- **Integration:** The pie chart is connected to other visualizations within the dashboard, offering a comprehensive perspective of SpaceX's launch data.

The interactive nature of this tool makes it ideal for presentations, data exploration, and real-time analysis.



Launch Success Ratio: KSC LC-39A

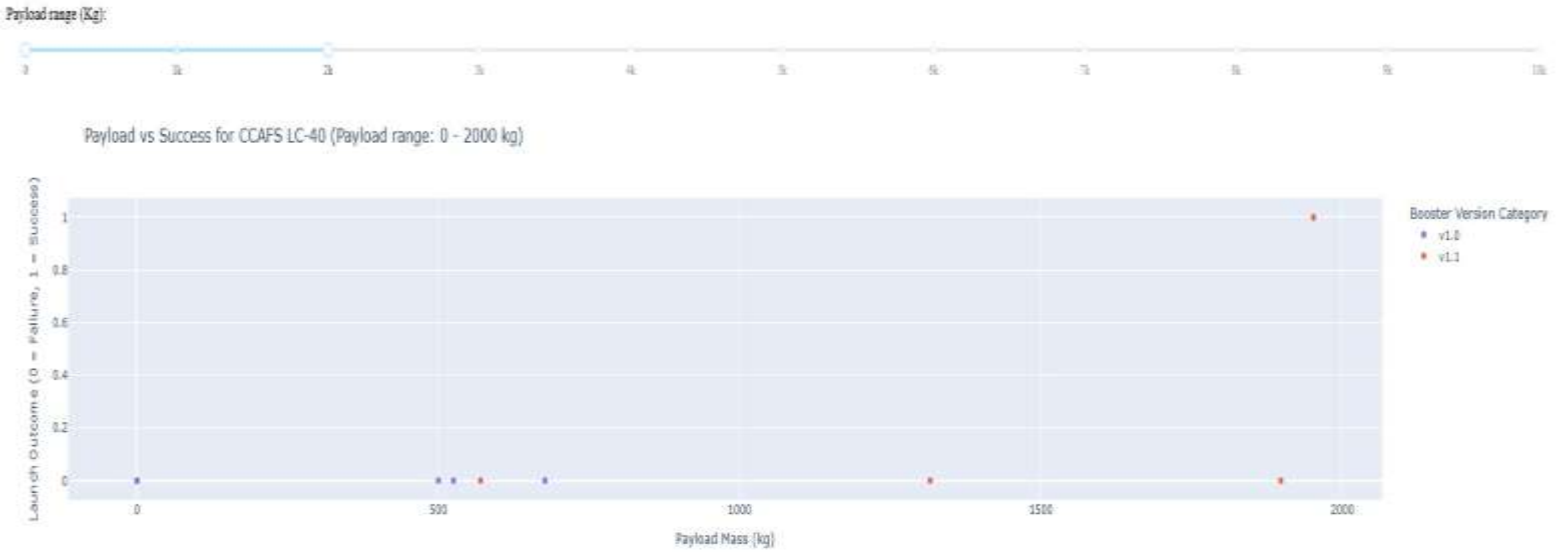
This slide displays a pie chart focusing on the KSC LC-39A launch site, which boasts the highest success ratio among all SpaceX launch sites.

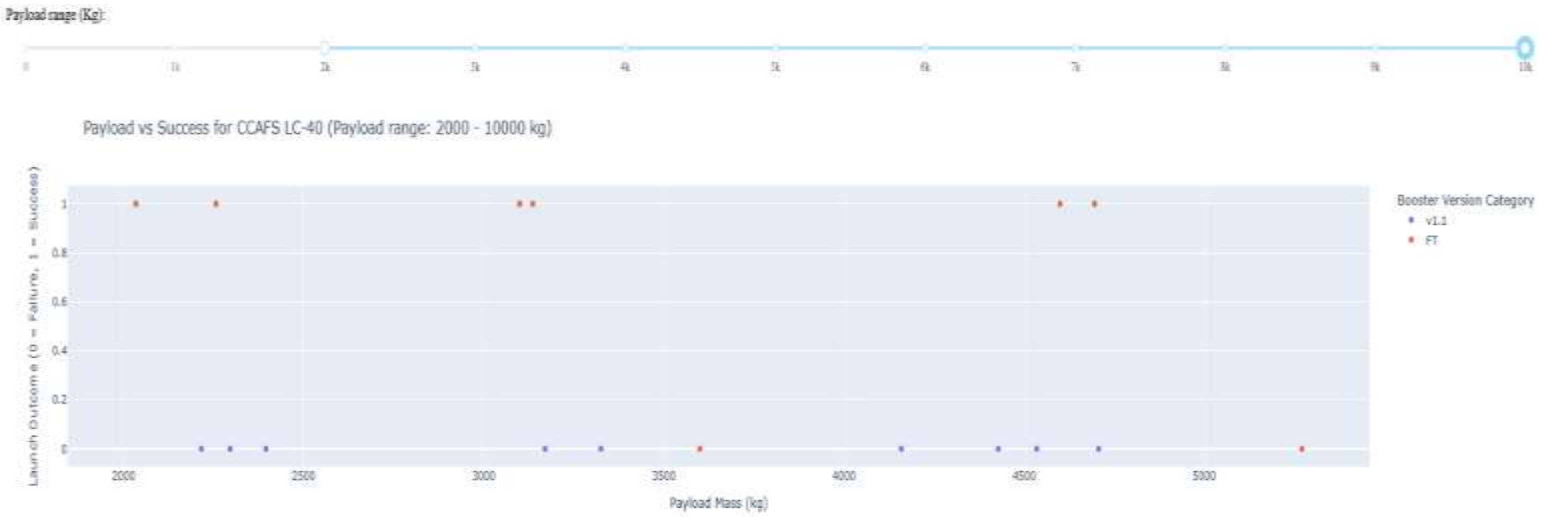
- **Visual Breakdown:** The chart shows 76.9% of launches as successful (red) and 23.1% as failures (blue).
- **Significance:** This high success ratio emphasizes the reliability and efficiency of the KSC LC-39A site.
- **Analysis:** This site has consistently achieved better outcomes compared to other locations, making it a crucial asset for SpaceX's operations.
- **Impact:** Such insights are vital for stakeholders to prioritize resource allocation and strategic planning at KSC LC-39A.

This pie chart reinforces the importance of focusing on high-performing sites for future launches.

Analysis for Each Launch Site Based on Payload Ranges and Booster Versions

CCAFS LC-40





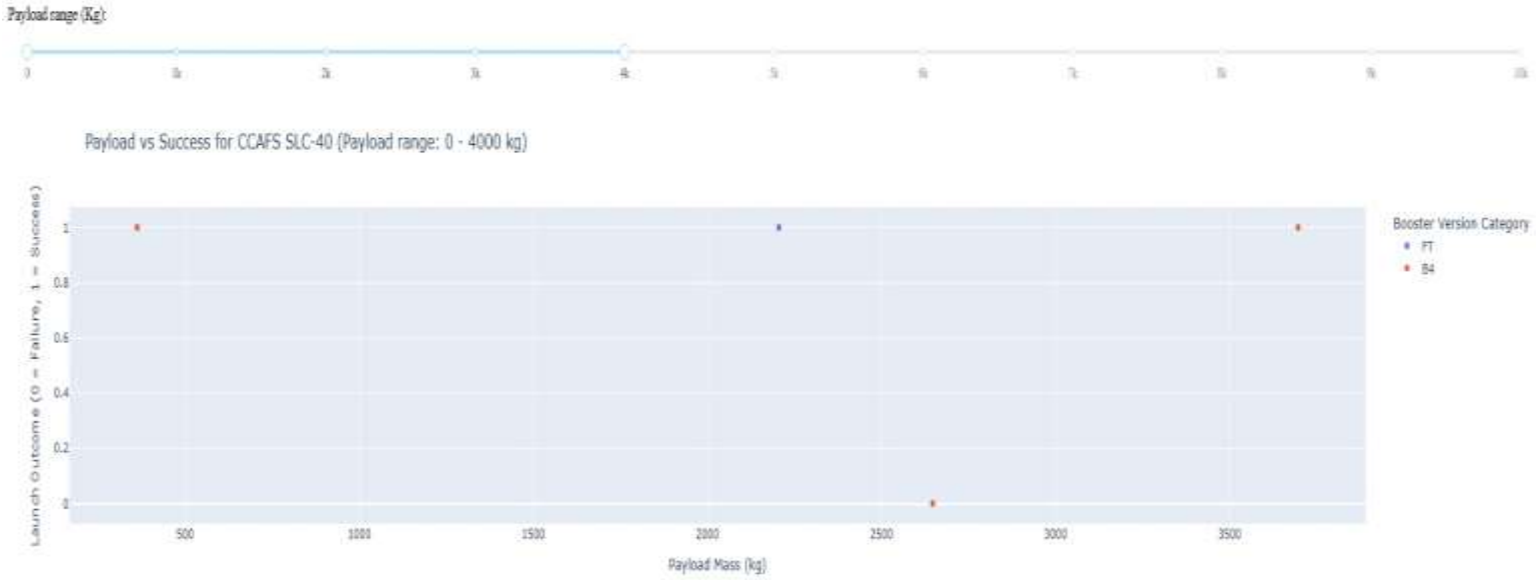
Payload Range 0-2000 kg:

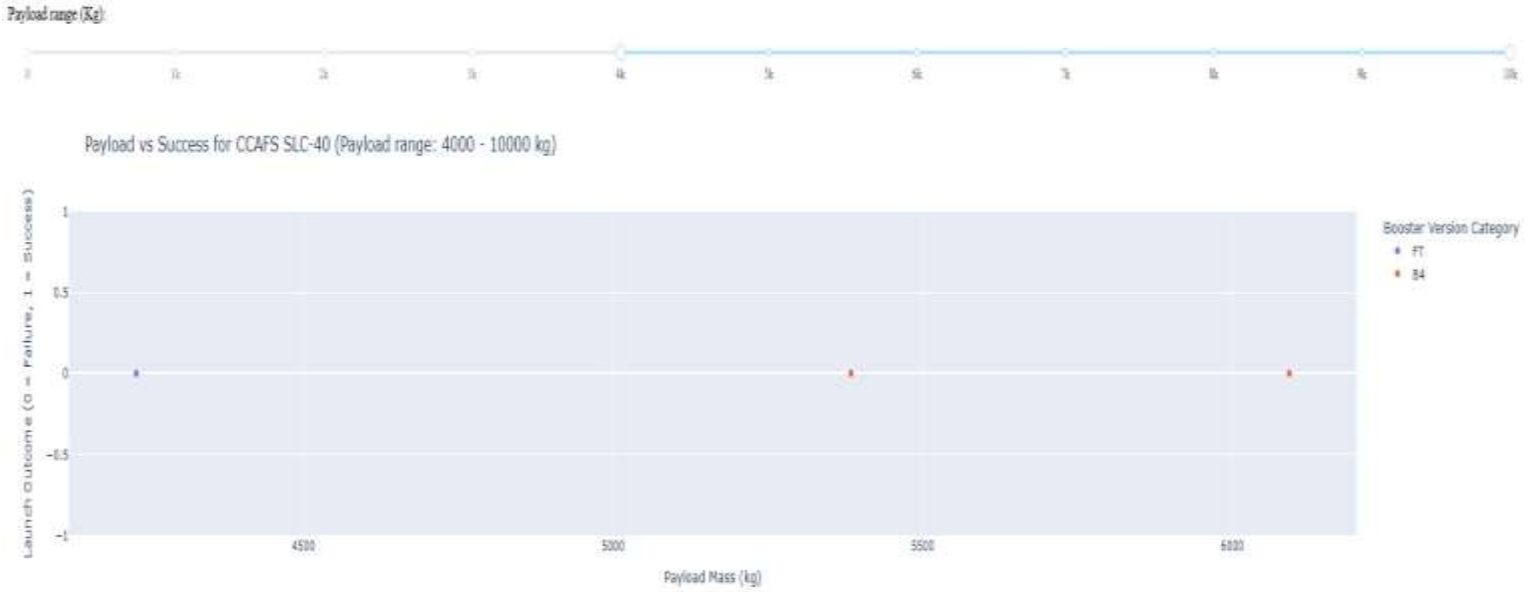
- Most outcomes were failures with Booster Versions **v1.0** and **v1.1**. These versions struggled with lower payloads.
- Minimal successful launches were observed, indicating lower efficiency for this payload range.

Payload Range 2000-10000 kg:

- The Booster Version **FT** demonstrated higher success rates in the 2000-5000 kg range.
- Failures were still present but significantly reduced compared to lighter payloads, showing the capability of **FT** boosters for moderate payloads.

CCAFS SLC-40





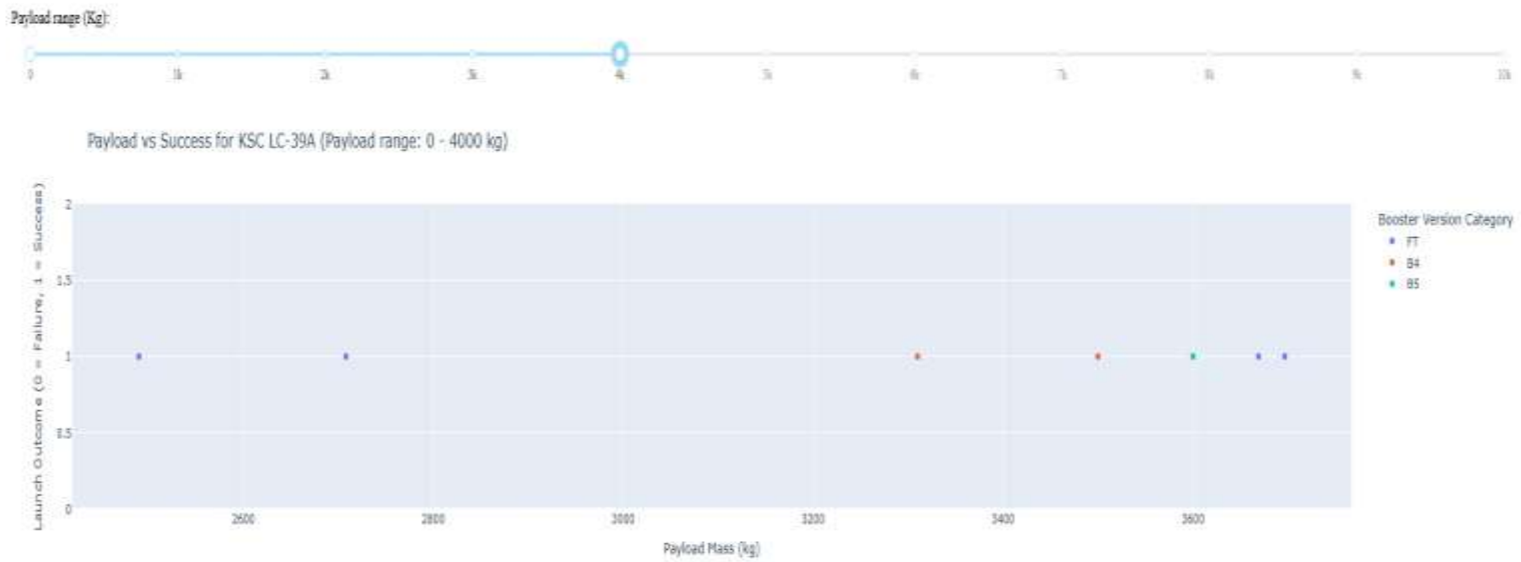
Payload Range 0-4000 kg:

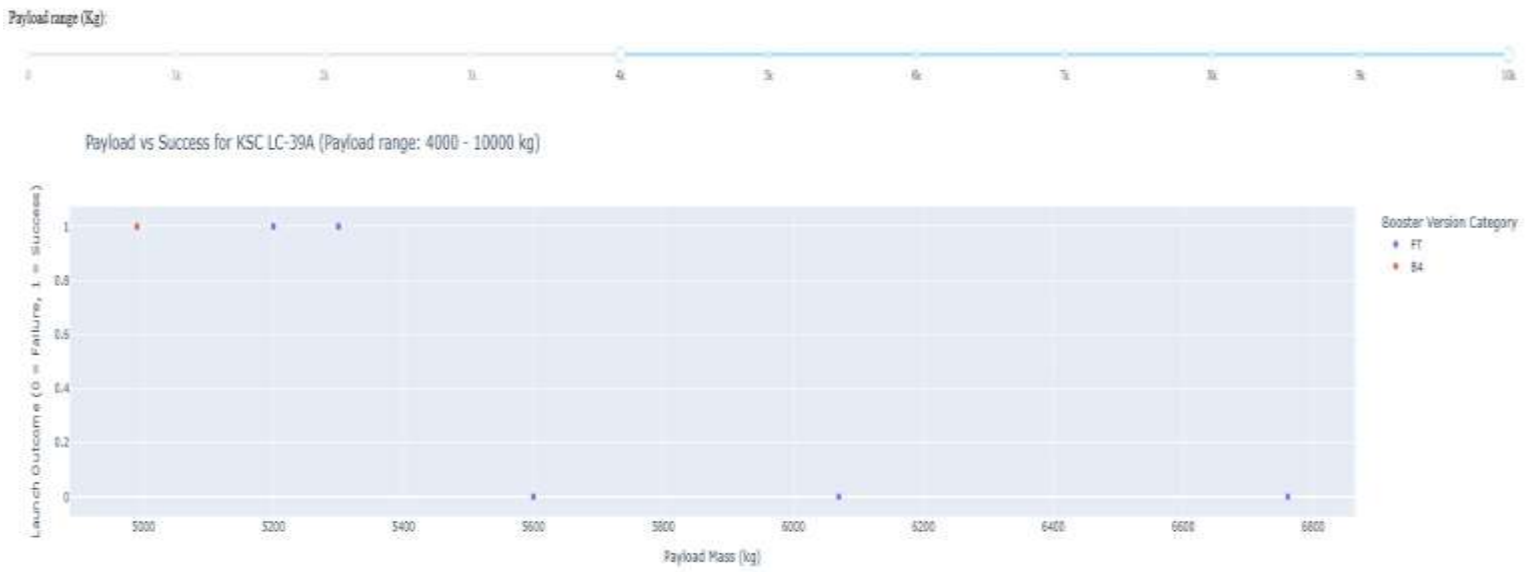
- 3 out of 4 launches in this range were successful, indicating the reliability of FT and B4 boosters in this payload range.

Payload Range 4000-10000 kg:

- 0 out of 3 launches in this range were a success, indicating several issues with heavier payloads even with advanced boosters as FT and B4.

KSC LC-39A





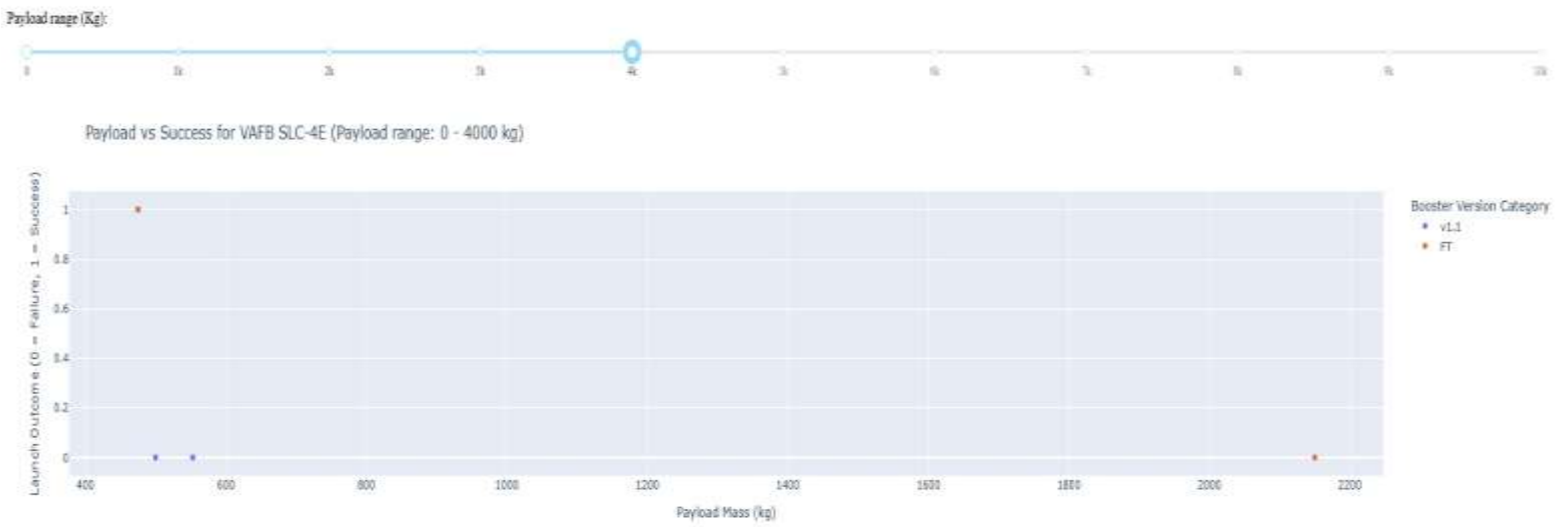
Payload Range 0-4000 kg:

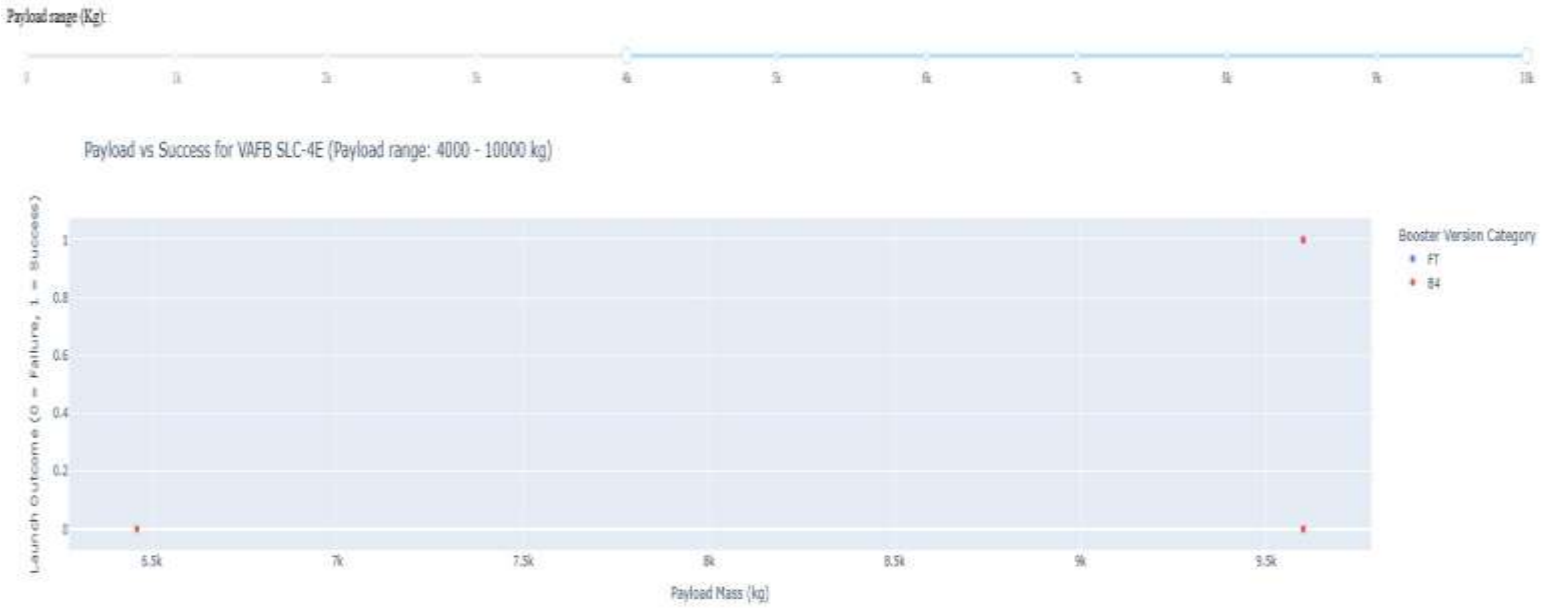
- FT, B4 and B5 boosters displayed a perfect success rate for payloads within this range.

Payload Range 4000-10000 kg:

- Launches were a success until 5600 kg payloads.
- The only failed missions from this launch site were with payloads heavier than 5600 kg, indicating that heavy payload remain a difficult challenge to face even from the most successful launch site.

VAFB SLC-4E





Payload Range 0-4000 kg:

- The only successful launch in this range was with an extremely light FT booster (475 kg), indicating the possibility of successful missions even with very light payloads.

Payload Range 4000-10000 kg:

- 1 out of 3 launches was a success, with the heaviest payload analyzed in this work (9600 kg), indicating that VAFB can be a reliable site for mission with heavier payloads.

Overall Findings

1. **Best Booster Versions**:

- **FT** boosters consistently outperformed others, especially in the range below 6000 kg.
- **B4** and **B5** boosters also demonstrated high reliability, especially for KSC LC-39A.
- **V1.0** and **V1.1** boosters show the lowest success rate, also because they were the firsts boosters made by SpaceX.

2. **Payload Range Effect**:

- **Lighter payloads (<2000 kg)** were more prone to failures, particularly with older boosters like v1.0 and v1.1.
- **Moderate payloads (2000-4000 kg)** achieved higher success rates, highlighting advancements in booster technology.
- **Heavy payloads (4000-10000 kg)** have not a good enough success rate, indicating that further improvements must be done for this range of payload.

3. **Site Efficiency**:

- **KSC LC-39A** emerged as the most efficient site, benefiting from advanced booster versions and infrastructure.
- **CAAFS SLC-40** has a good success rate for payloads up to 4000 kg, while it has a 0 % success rate for heavier payload, indicating that this site is probably not adequate for heavy payload launches.
- **CCAFS LC-40** have a good success rate with FT boosters regardless of payloads, while it has a low success rate with v1.0 and v1.1 boosters, indicating that the version of booster highly influences the success of the mission.
- **VAFB SLC-4E** has lower launches than other sites, indicating that this site is used for specific missions. The only successful missions from this site were with FT and B4 boosters, showing that they are more reliable than older boosters like v1.1. It is important to underline that this site was used for the rockets with the heaviest payloads (9600 kg), indicating its suitability for these type of missions.

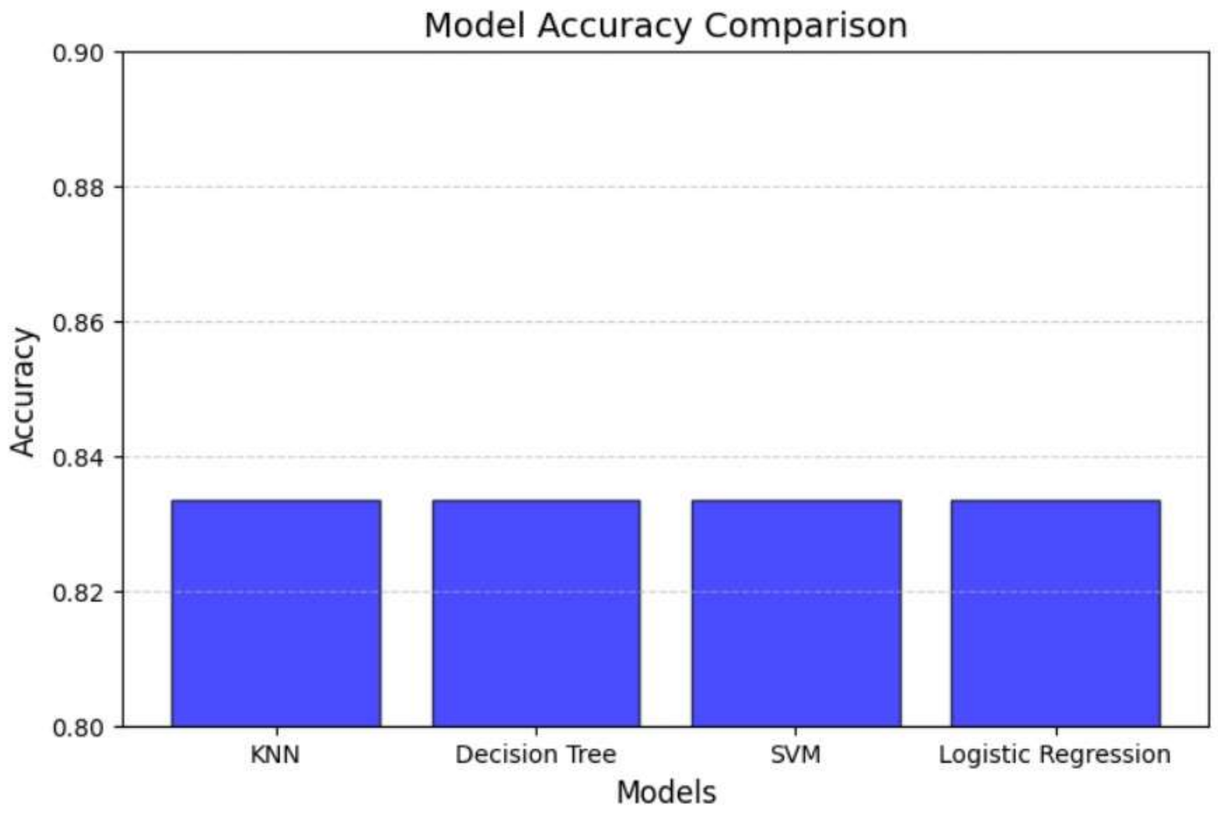
This analysis underscores the importance of pairing the right payload with the appropriate booster and launch site to optimize mission outcomes.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

All the models (KNN, Decision Tree, SVM and Logistic Regression) share the same accuracy score on test data: 0.8333333333333334



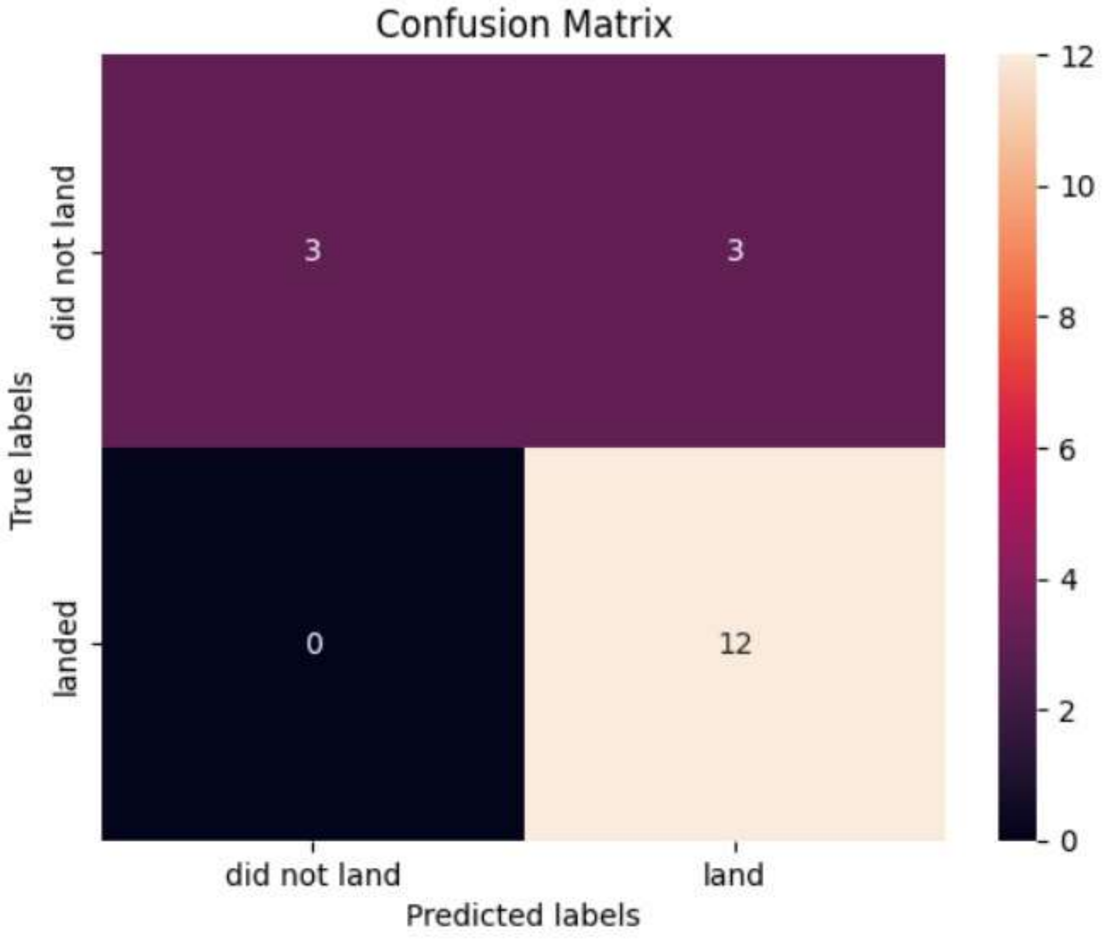
Confusion Matrix

As with the previous bar chart, all the models share the same confusion matrix.

18 test data were examined by the models. 12 were successfully landed launches, 6 were launches that failed to land.

All the models have the same structure:

- True Positive - 12 (True label is landed, Predicted label is also landed)
- False Positive - 3 (True label is not landed, Predicted label is landed)



Conclusions

Strategic Importance of Launch Sites

- The geographical analysis revealed that SpaceX's launch sites are strategically positioned near critical infrastructures, such as railways, highways, and urban centers, ensuring logistical efficiency.
- Among the sites analyzed, **KSC LC-39A** demonstrated the highest success rate, making it a reliable choice for high-stakes missions.
- Launch sites in Cape Canaveral, such as **CCAFS LC-40** and **SLC-40**, contributed significantly to overall success rates, underscoring their importance to SpaceX's operations.

Payload and Booster Version Insights

- The analysis highlighted a correlation between payload mass, booster versions, and launch outcomes. Success rates were higher for payloads within specific ranges and for boosters like **FT, B4** and **B5**.
- **CCAFS LC-40** showed a strong performance for medium payload ranges, while **KSC LC-39A** excelled across broader payload distributions.
- These insights suggest the need to optimize payload configurations and booster selections to maximize mission success rates.

Machine Learning Models for Outcome Prediction

- Four classification models (KNN, Decision Tree, SVM, Logistic Regression) were built to predict launch outcomes. All models achieved an accuracy of **83.3%**, demonstrating their potential for enhancing mission planning.
- The high accuracy across models indicates that the available data is well-suited for outcome prediction, but further feature engineering and data enrichment could lead to even better results.

Operational and Strategic Recommendations

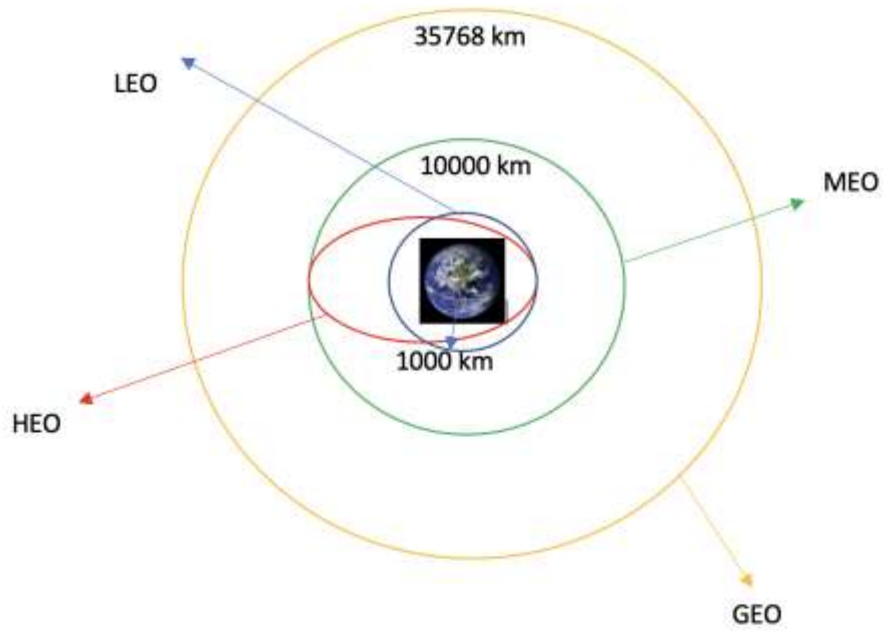
- **Focus on High-Performing Sites**: Prioritize launches from **KSC LC-39A** for critical missions due to its consistently high success rate.
- **Payload Optimization**: Refine payload strategies to target the ranges with historically higher success rates, particularly leveraging **FT boosters** for medium-to-high payload masses.
- **Predictive Analysis for Risk Mitigation**: Utilize classification models in operational planning to assess the likelihood of mission success, enabling proactive risk management.
- **Continuous Model Improvement**: Invest in expanding the dataset with more features (e.g., weather data, launch timing) to improve predictive accuracy and derive actionable insights.

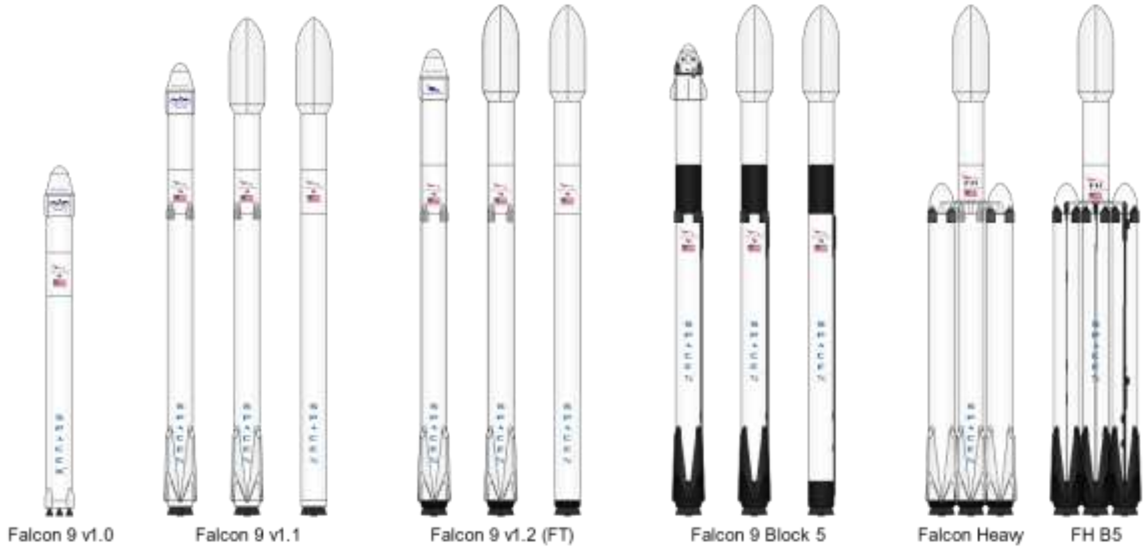
Broader Implications for SpaceX

- The analyses underline the critical role of data-driven decision-making in optimizing launch operations.
- By leveraging machine learning and analytical tools, SpaceX can not only enhance operational efficiency but also maintain its competitive edge in the rapidly growing space exploration industry.

Appendix

In this final slide, you can see two important images that didn't find a previous space during this work: the types of orbit that affect the choice of launch site, booster version and payload, and the different types of rockets analyzed.





Thank you!