

WINNING SPACE RACE WITH DATA SCIENCE

Alejandro Montenegro López

November 30, 2024



OUTLINE

- 1. Executive Summary (pg. 3)**
- 2. Introduction (pg. 4)**
- 3. Methodology (pg. 5)**
- 4. Results (pg 5.)**
- 5. Conclusion (pg. 45)**
- 6. Appendix (pg. 46)**

EXECUTIVE OVERVIEW

Summary of Methodologies

This project employs a structured approach to predict the success of first-stage rocket landings, leveraging data science to drive insights and optimize operations. The methodology begins with collecting publicly available data on past launches, followed by cleaning and preprocessing to ensure quality and compatibility for analysis. Key factors influencing landing success are identified through exploratory data analysis, and machine learning models are developed to predict outcomes with high accuracy. The predictive models are evaluated and visualized with interactive dashboards for actionable insights. These tools will support Space Y's strategic decision-making, cost optimization, and competitive positioning in the commercial space market.

Summary of All Results

SpaceX has established itself as a leader in space exploration by achieving consistent advancements in technology, operational efficiency, and launch reliability. The company has demonstrated steady improvements, particularly with newer booster models, achieving higher success rates even with heavier payloads. Launch site performance reveals KSC LC-39A as the most reliable, while early challenges at CCAFS SLC-40 have been overcome. Orbital reliability varies, with some orbits achieving perfect success rates while others face challenges like GTO and SO. Strategic placement of coastal launch sites enhances efficiency and safety, while predictive models like Decision Trees, with an accuracy of 88.9%, optimize operational planning and reliability, setting new benchmarks for sustainable and cost-effective space travel.

INTRODUCTION

Background & Context

The commercial space industry is experiencing rapid growth and innovation, with reusability of rocket components, particularly the first stage, emerging as a transformative factor in reducing launch costs and enhancing competitiveness.

Reusable rockets have redefined the economics of space travel, allowing companies to offer more affordable and sustainable launch services. At Space Y, our mission is to position ourselves as a strong competitor in this dynamic market by delivering reliable, cost-efficient rocket launches. To achieve this, we must harness the power of data science to predict the success of first-stage landings, which are critical to controlling operational costs, optimizing resource use, and ensuring the reusability of key rocket components.

Problems We Are Trying to Find Answers To

This project seeks to answer key questions: What factors influence first-stage landing success? How can we reliably predict outcomes using machine learning? What insights can guide cost optimization and competitive positioning? By addressing these challenges, we aim to build a data-driven foundation for innovation and growth in the commercial space market.



SECTION 1

METHODOLOGY

METHODOLOGY

Executive Summary

- **Data Collection Methodology:** Data was collected from the SpaceX REST API and Wikipedia through web scraping, creating a comprehensive dataset of launch details and outcomes.
- **Perform Data Wrangling:** Landing outcomes were standardized into binary labels to enable clear training labels for supervised machine learning models.
- **Perform Exploratory Data Analysis (EDA) using Visualization and SQL:** SQL and visualizations revealed patterns in landing success rates based on variables like launch site and payload mass.
- **Perform Interactive Visual Analytics using Folium and Plotly Dash:** Geospatial maps analyzed launch site locations and outcomes, uncovering location-based trends that inform site optimization.
- **Perform Predictive Analysis Using Classification Models:** Machine learning models were trained and optimized to predict landing success, with evaluation using accuracy and other performance metrics.

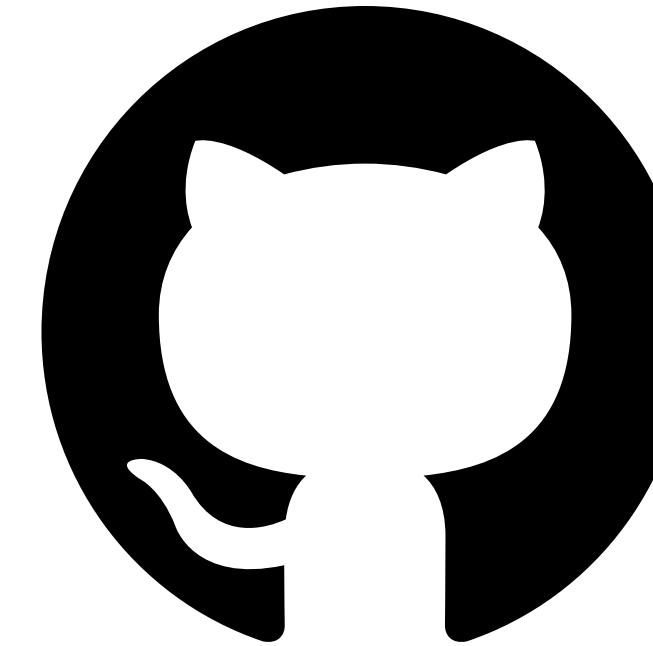
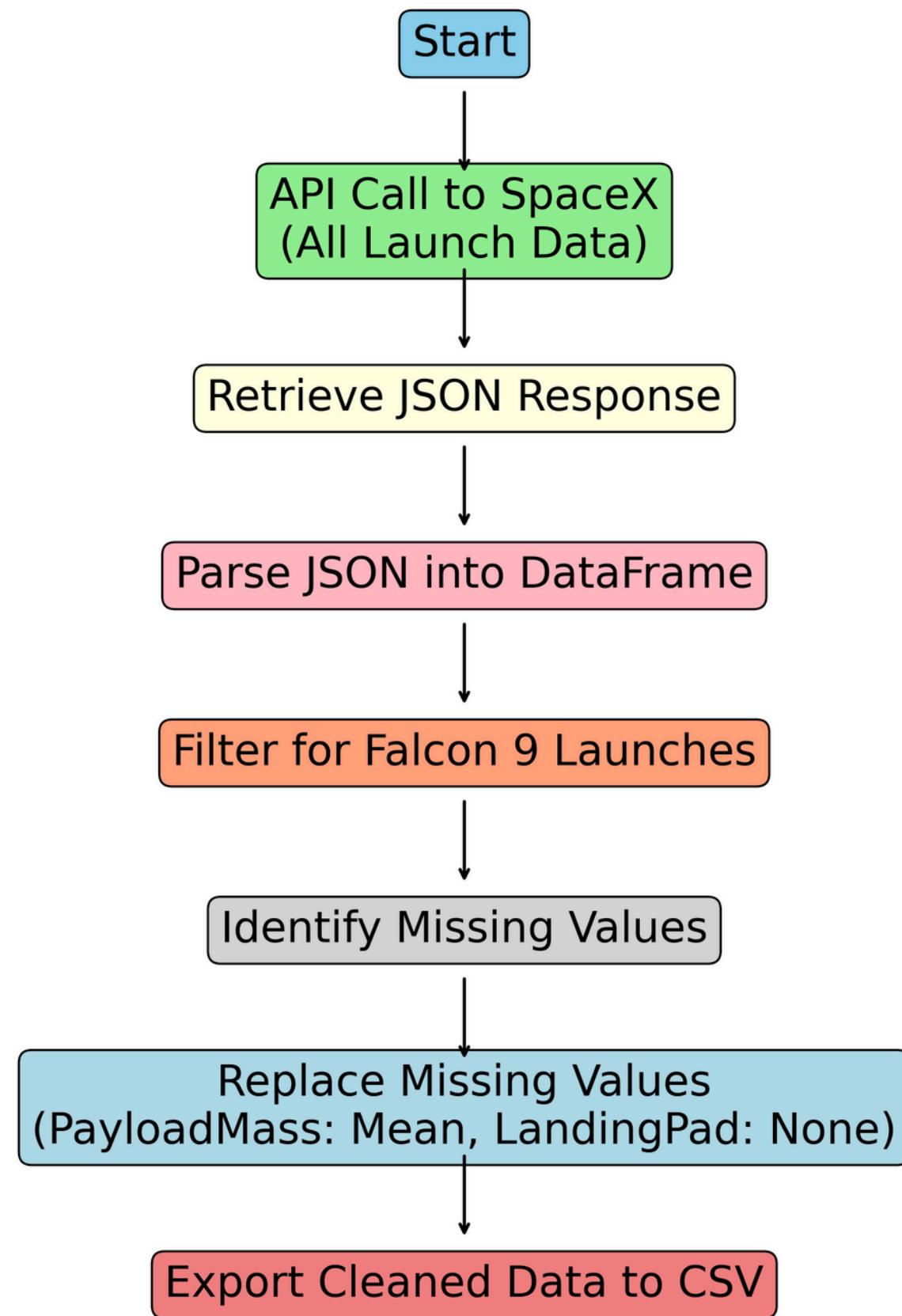
DATA COLLECTION

Data was collected from two primary sources: the SpaceX REST API and web scraping of the Wikipedia page titled List of Falcon 9 and Falcon Heavy launches. The SpaceX API provided comprehensive information about each launch, including rocket details, payload specifications, launch conditions, and landing outcomes. Web scraping complemented this by gathering historical Falcon 9 launch records, ensuring a rich and reliable dataset for analysis.

Steps:

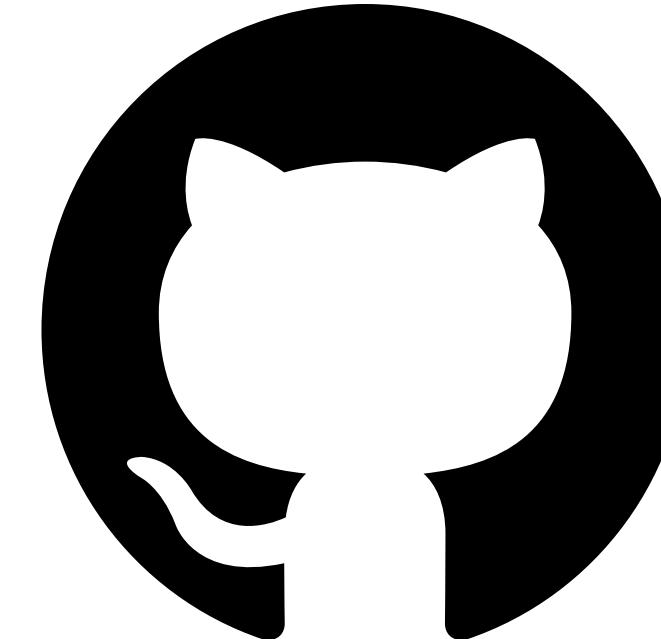
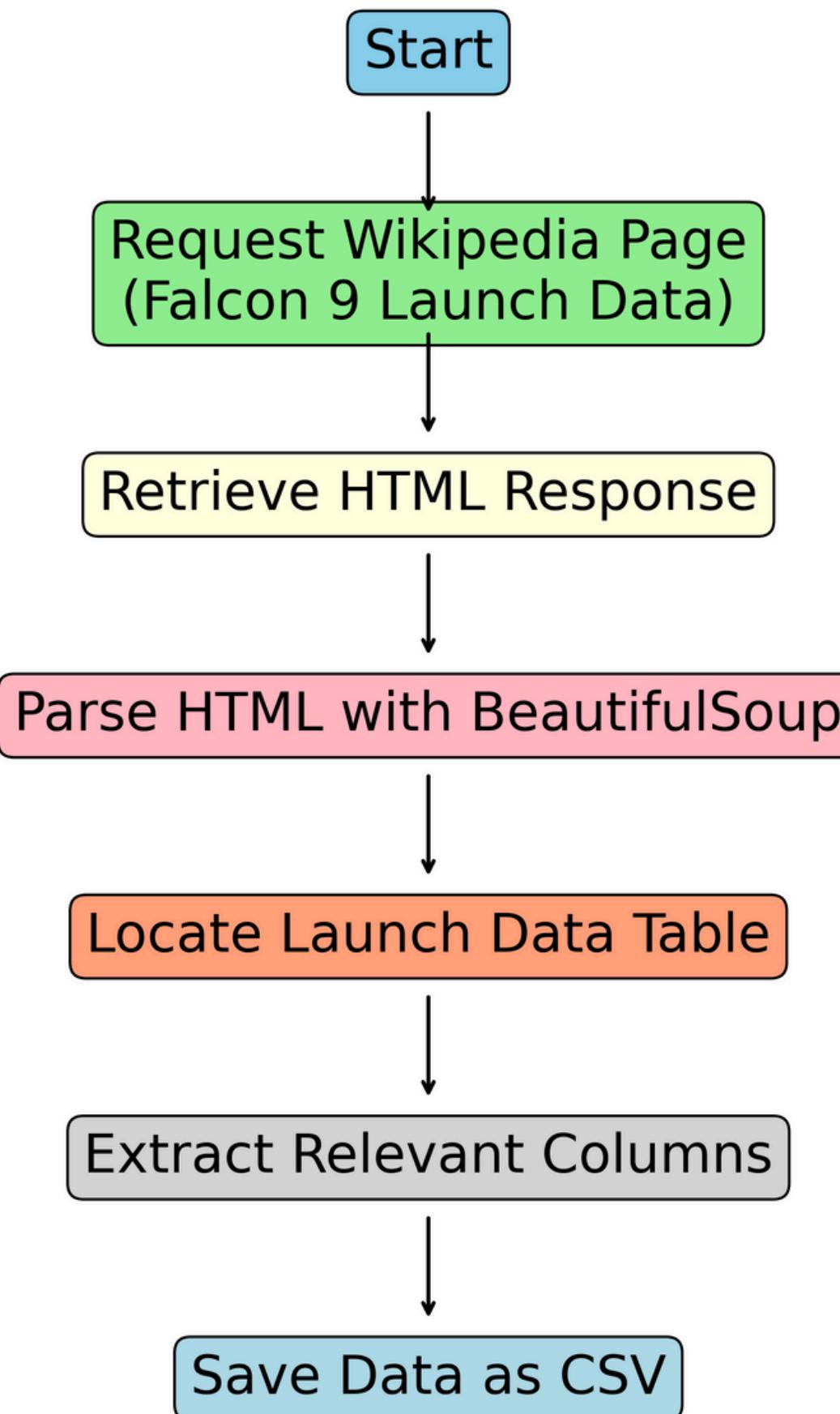
1. Import libraries and dataset
2. Explore dataset structure
3. Filter for Falcon 9 launches
4. Identify missing values
5. Fill missing PayloadMass with mean
6. Retain None in LandingPad for unused pads
7. Validate cleaned data
8. Export to CSV

DATA COLLECTION – SPACEX API



[https://github.com/Ale-Monte/Data-
Science-Capstone/blob/main/1-
spacex-data-collection-api.ipynb](https://github.com/Ale-Monte/Data-Science-Capstone/blob/main/1-spacex-data-collection-api.ipynb)

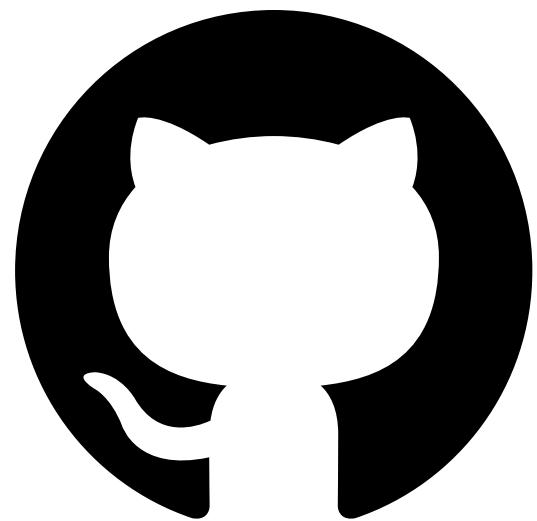
DATA COLLECTION – WEBSRAPING



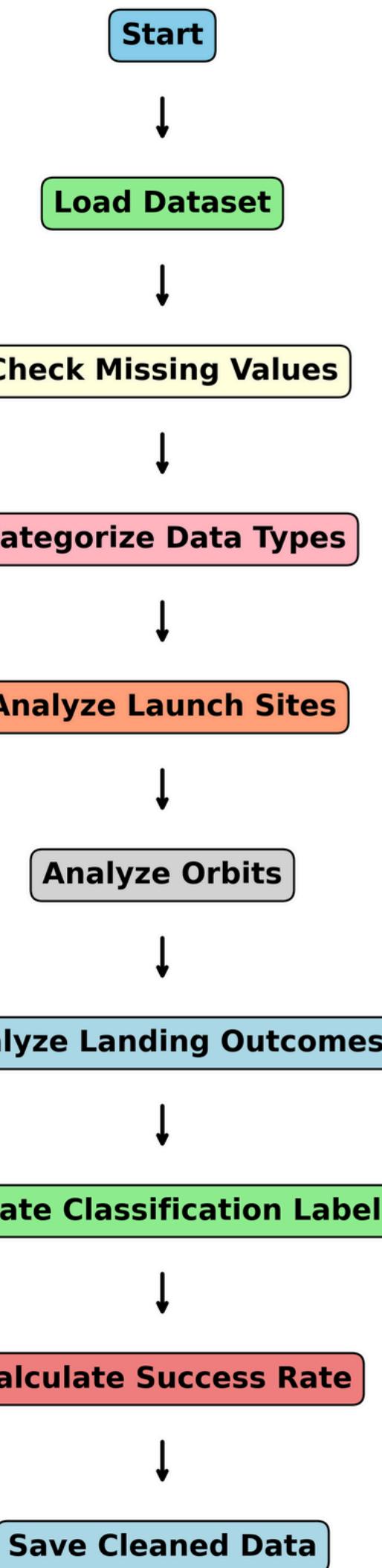
[https://github.com/Ale-Monte/Data-
Science-Capstone/blob/main/2-
spacex-websraping.ipynb](https://github.com/Ale-Monte/Data-Science-Capstone/blob/main/2-spacex-websraping.ipynb)

DATA WRANGLING

The dataset was cleaned and prepared for analysis by standardizing outcomes into binary labels for supervised learning models. Landing outcomes such as True Ocean, True RTLS, and True ASDS (successful landings) were labeled as 1, while unsuccessful outcomes (False Ocean, False RTLS, and False ASDS) were labeled as 0. This transformation enabled the identification of clear training labels.



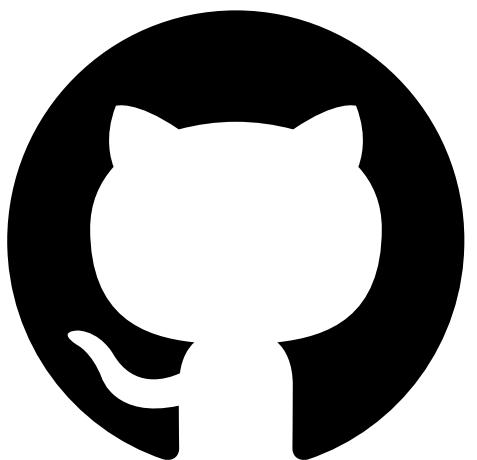
<https://github.com/Ale-Monte/Data-Science-Capstone/blob/main/3%20-spacex-data-wrangling.ipynb>



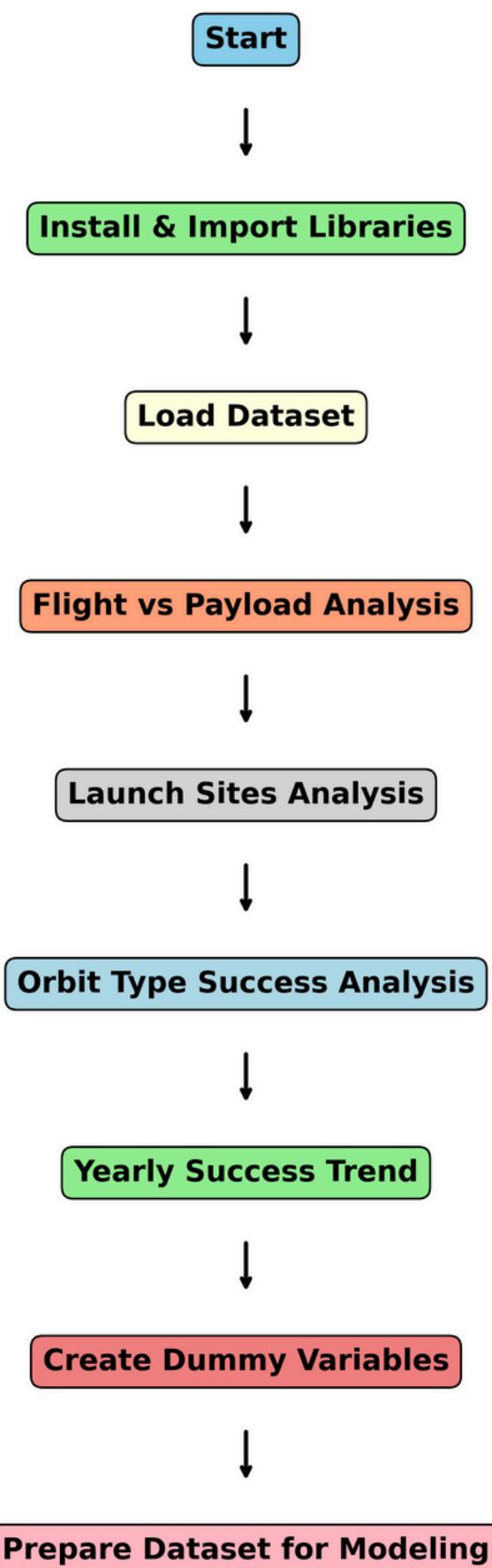
EDA WITH DATA VISUALIZATION

Using Pandas and Matplotlib, visualizations were created to uncover trends and relationships in the data. For instance, scatter plots and histograms highlighted how variables like payload mass, launch site, and mission type correlated with landing success.

Specifically, the following were graphed: Flight Number vs. Launch Site, success rate of each orbit type, flight number vs. orbit type, payload vs. orbit type and



<https://github.com/Ale-Monte/Data-Science-Capstone/blob/main/4-eda-dataviz.ipynb>

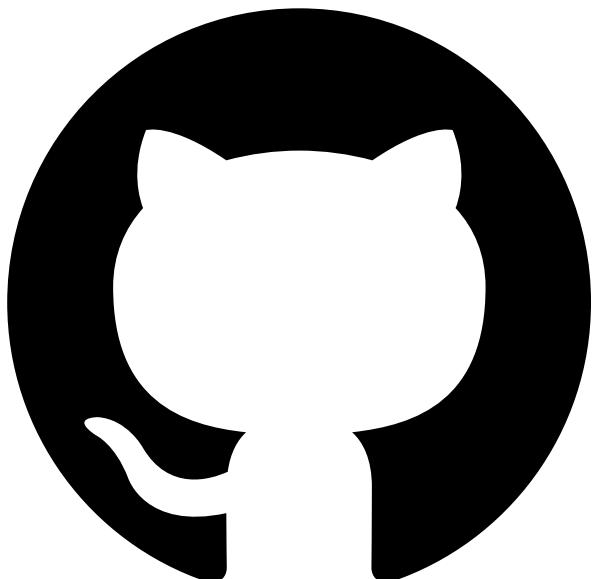


EDA WITH SQL

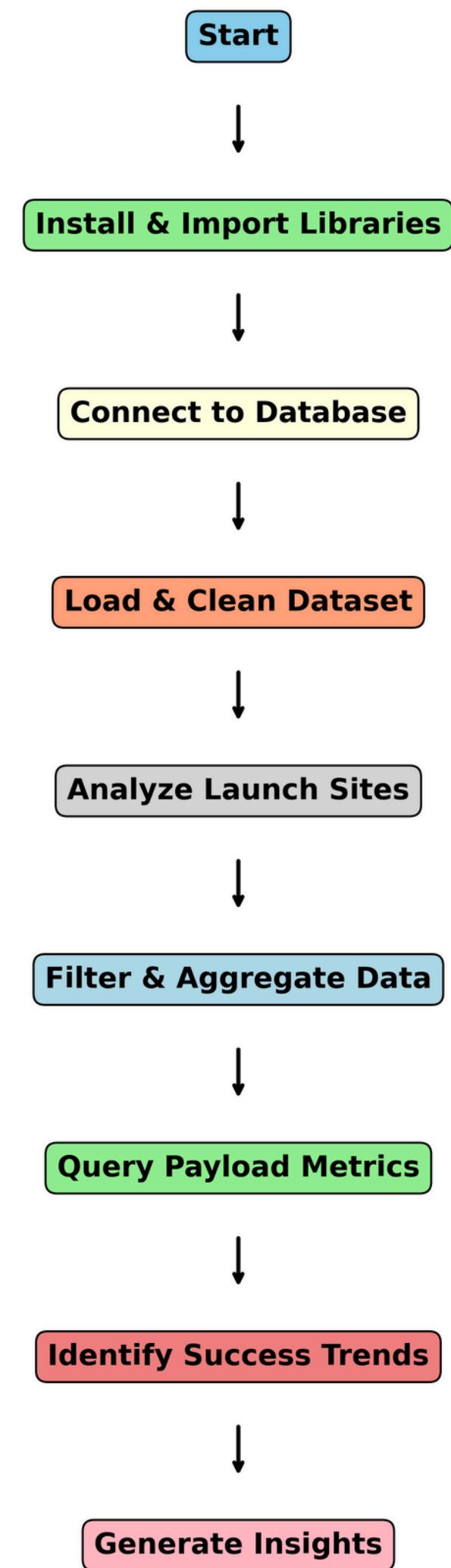
Using SQL, patterns in landing success rates were analyzed.

For instance, the success rates varied by launch site.

Combining features provided deeper insights, such as observing the effects of payload mass. This analysis informed the selection of attributes like launch number, payload mass, and launch site for machine learning models.

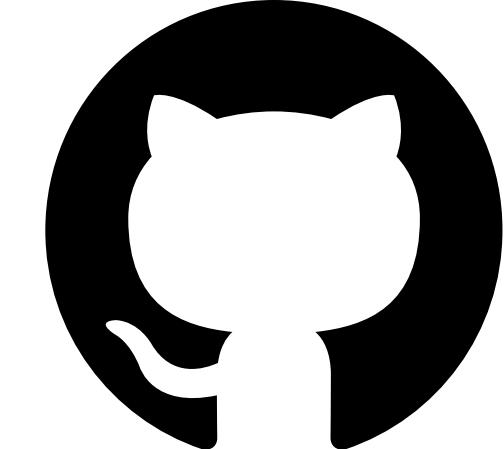
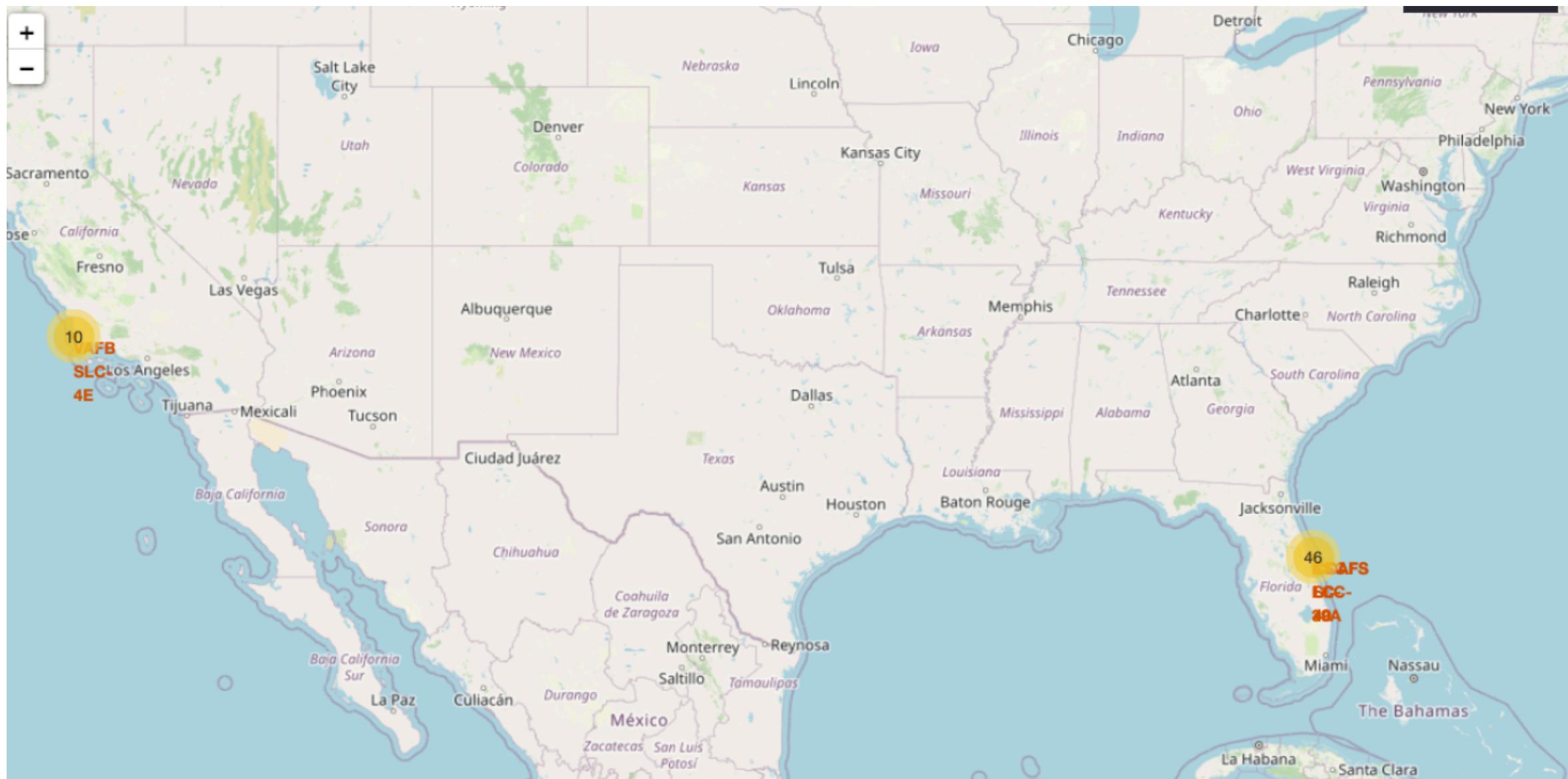


<https://github.com/Ale-Monte/Data-Science-Capstone/blob/main/5-edu-sql-sqlite.ipynb>



INTERACTIVE MAP WITH FOLIUM

Geospatial patterns were analyzed using Folium. Maps were created to mark all launch sites, overlaying the success or failure of each launch with markers. Proximity analyses calculated distances between launch sites and nearby geographical features with lines to explore how location influences success rates. This interactive mapping revealed insights into the relationship between site location and launch outcomes, supporting decisions on optimal launch site placement.

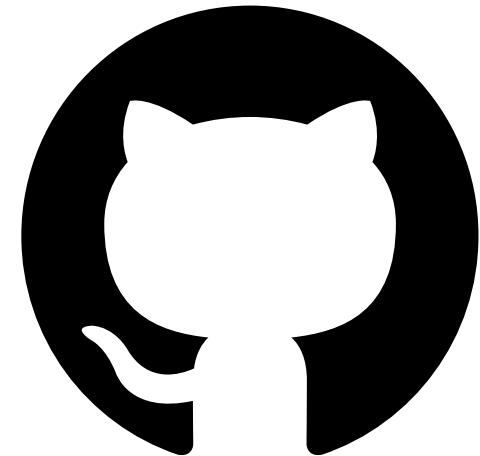


<https://github.com/Ale-Monte/Data-Science-Capstone/blob/main/6-launch-site-location.ipynb>

BUILD A DASHBOARD WITH PLOTLY DASH

The dashboard includes a pie chart to display success rates, either aggregated across all launch sites or showing success vs. failure for a specific site, depending on the dropdown selection. A scatter plot visualizes the relationship between payload mass and launch success, color-coded by booster version category, enabling detailed trend analysis. Interactivity is introduced through a dropdown menu to select specific launch sites and a range slider to filter data by payload mass, dynamically updating the scatter plot for more tailored exploration.

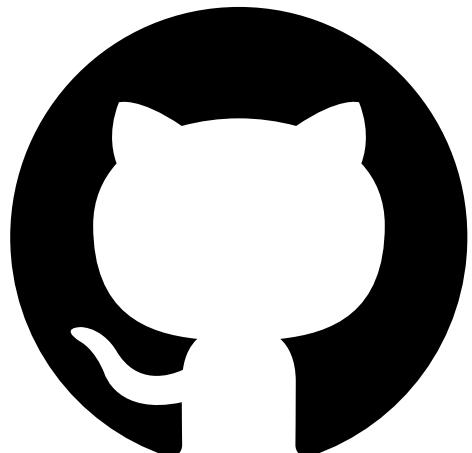
Why these were added? These visualizations and interactions were added to provide both a high-level overview and detailed insights into SpaceX's launch data. The pie chart offers a quick summary of success rates, useful for comparing sites at a glance. The scatter plot allows for deeper analysis of how payload mass and booster categories influence outcomes. Interactivity via the dropdown and slider enhances user engagement, enabling customized exploration to uncover patterns or anomalies specific to sites or payload ranges. This combination ensures the dashboard is both informative and user-friendly.



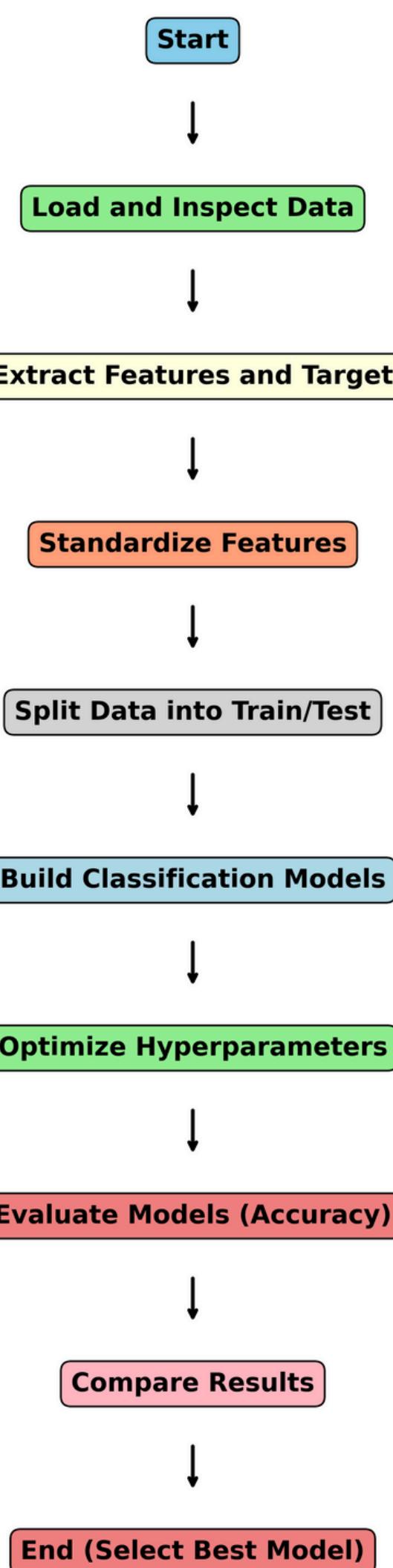
https://github.com/Ale-Monte/Data-Science-Capstone/blob/main/7-spacex_dash_app.py

PREDICTIVE ANALYSIS

The model development process began with loading and preparing the data, including feature extraction, target selection, and standardization to ensure consistency across variables. The data was split into training and testing subsets, with 80% used for training and 20% for testing. Four classification models, logistic regression, support vector machine, decision tree, and k-nearest neighbors, were developed and optimized by tuning their hyperparameters using cross-validation. The models were evaluated on the test set based on accuracy and confusion matrix analysis. Finally, the models were compared, and the most accurate and reliable one was selected for robust and generalizable predictions.



<https://github.com/Ale-Monte/Data-Science-Capstone/blob/main/8-spaceX-Machine-Learning-Prediction.ipynb>



RESULTS

EDA (SQL)

1. Unique Launch Sites:

- a. CCAFS LC-40, VAFB SLC-4E, KSC LC-39A, CCAFS SLC-40

2. Total Payload Mass by NASA (CRS):

- a. 45,596 kg

3. Average Payload Mass (F9 v1.1):

- a. 2534.67 kg

4. First Successful Ground Pad Landing:

- a. 2010-06-04

5. Mission Outcomes:

- a. Successful: 100
- b. Failure (in flight): 1

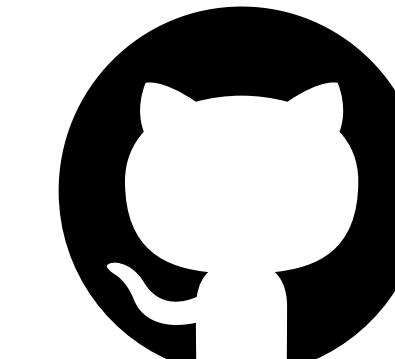
6. Failures in Drone Ship Landings (2015):

- a. January: F9 v1.1 B1012 (SpaceX CRS-5)
- b. April: F9 v1.1 B1015 (SpaceX CRS-6)

Visualization:

Upon analyzing the Payload vs. Launch Site scatter plot, it is evident that at the VAFB-SLC launch site, no rockets were launched with heavy payloads exceeding 10,000 kg. Success rates across various orbits reveal that ES-L1, GEO, HEO, and SSO achieved a perfect 100% success rate, while the SO orbit recorded the lowest success rate at 0%. In the LEO orbit, success appears to correlate with the number of flights; however, this relationship is not observed in the GTO orbit, where flight frequency does not seem to influence success. For heavy payloads, the Polar, LEO, and ISS orbits exhibit a higher positive landing rate or successful landings. Conversely, the GTO orbit presents a mixed scenario, with both successful and unsuccessful landings occurring at similar rates.

Additionally, the success rate trend since 2013 indicates a consistent increase, with stability observed in 2014. Notably, after 2015, the success rate experienced a marked improvement, peaking in 2017.



<https://github.com/Ale-Monte/Data-Science-Capstone>

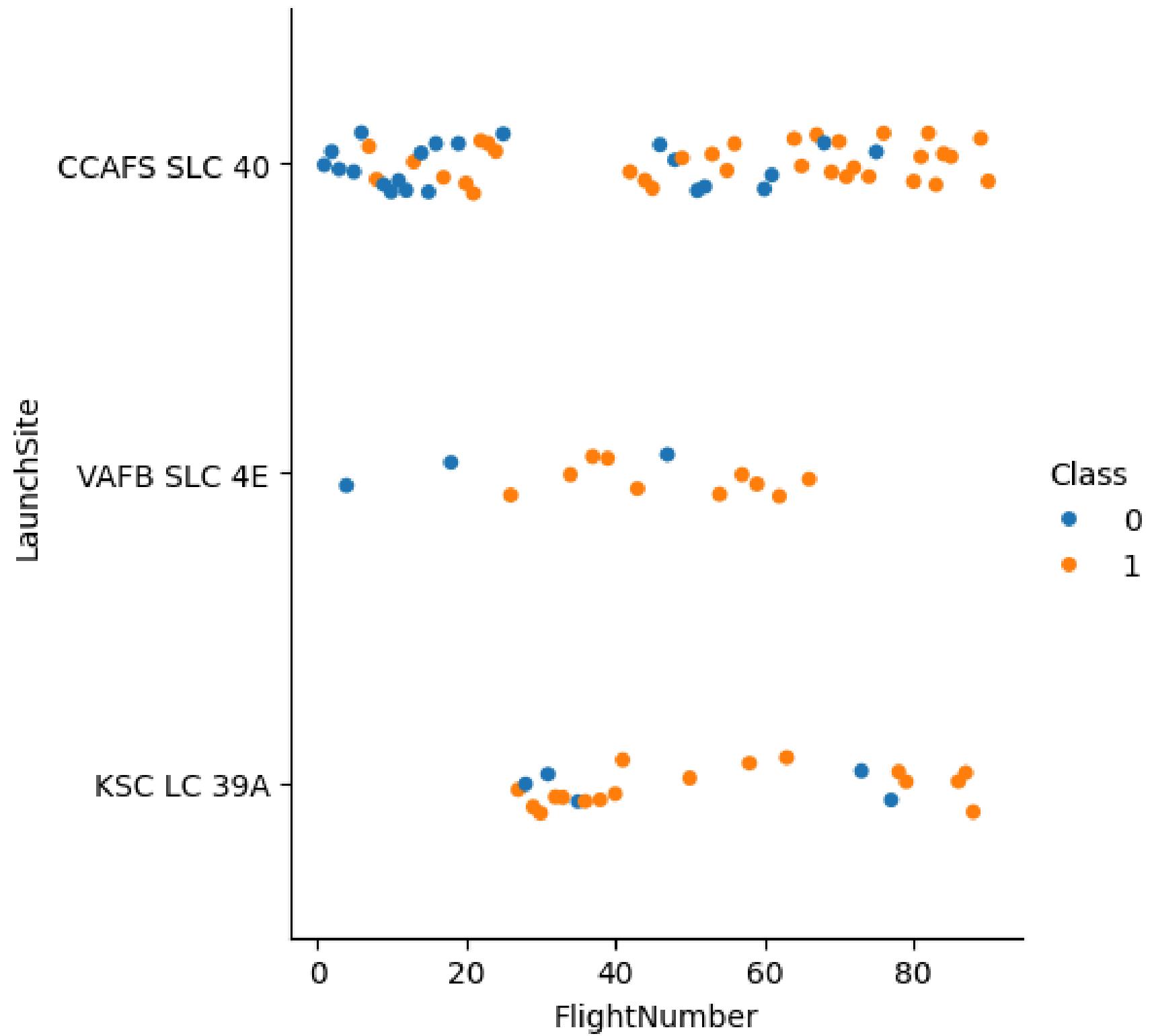
A dark blue background image of a space shuttle launching from a launch pad, with a large plume of white smoke and fire billowing from its engines. A thick diagonal graphic element runs from the bottom right towards the top left, composed of three nested triangles in shades of blue and white.

SECTION 2

INSIGHTS DRAWN

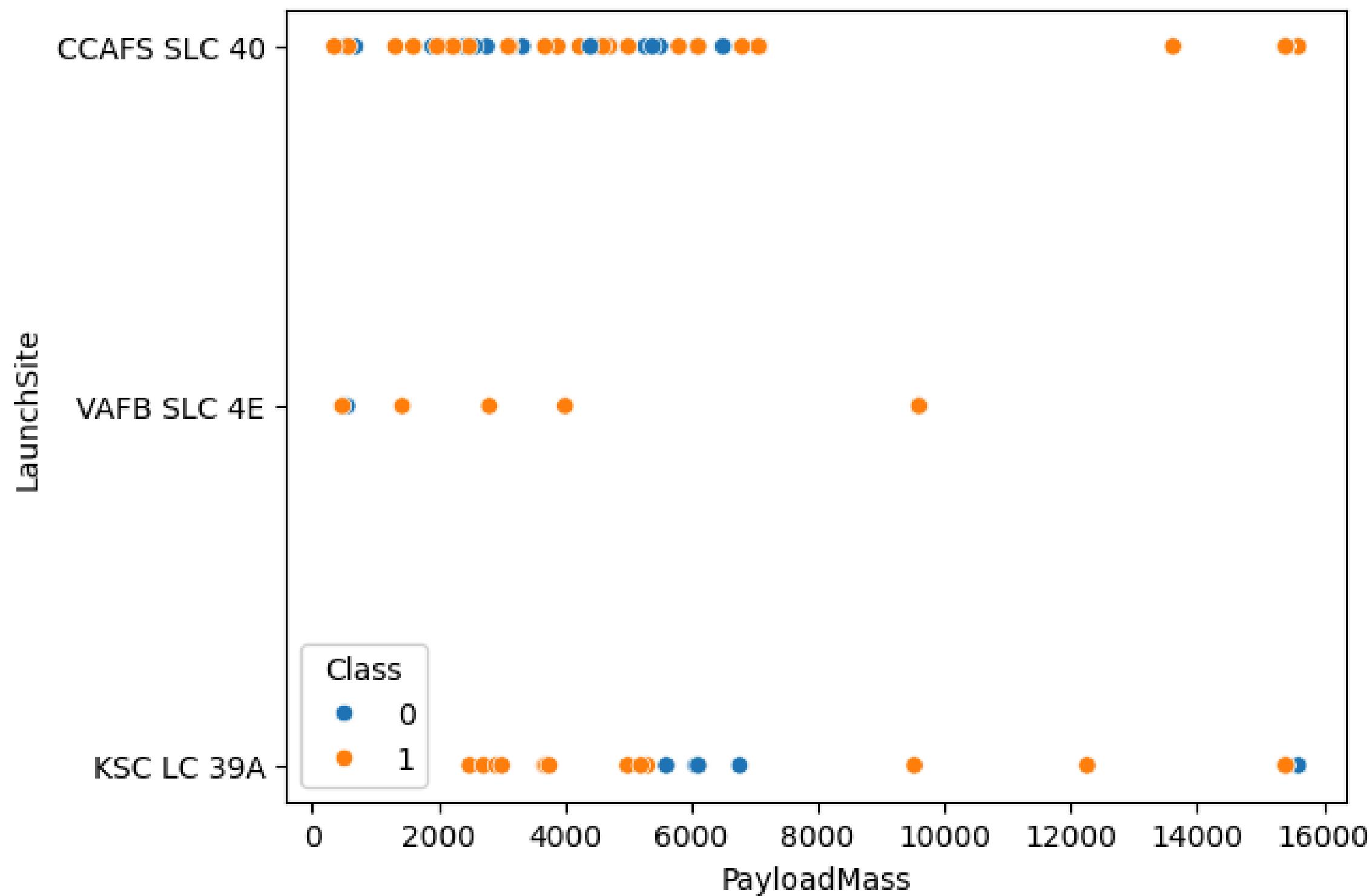
FROM EDA

FLIGHT NUMBER VS. LAUNCH SITE



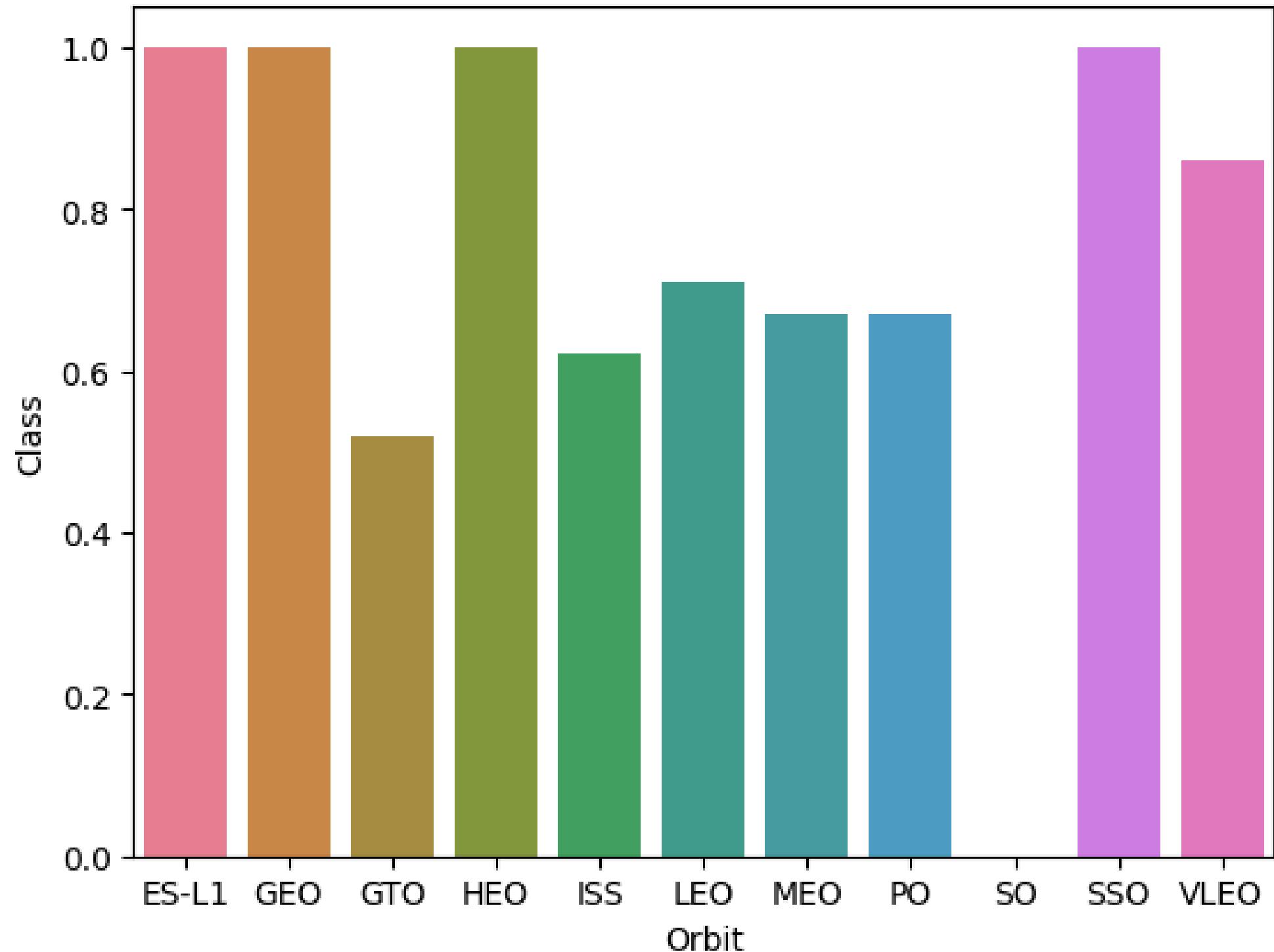
This graph illustrates the relationship between the flight number and launch site for SpaceX rocket launches, with mission outcomes categorized by success (orange, Class 1) or failure (blue, Class 0). The y-axis lists the three launch sites: CCAFS SLC 40, VAFB SLC 4E, and KSC LC 39A, while the x-axis shows the chronological flight numbers. The data reveals that launches occurred at all three sites, but CCAFS SLC 40 has a higher density of launches. More importantly, a trend emerges where later flights show more successes, indicating improvements in SpaceX's launch reliability over time. The graph also highlights variations in outcomes across sites, with all three showing a mix of successes and failures, though success rates appear to increase as SpaceX gained more experience.

PAYLOAD VS. LAUNCH SITE



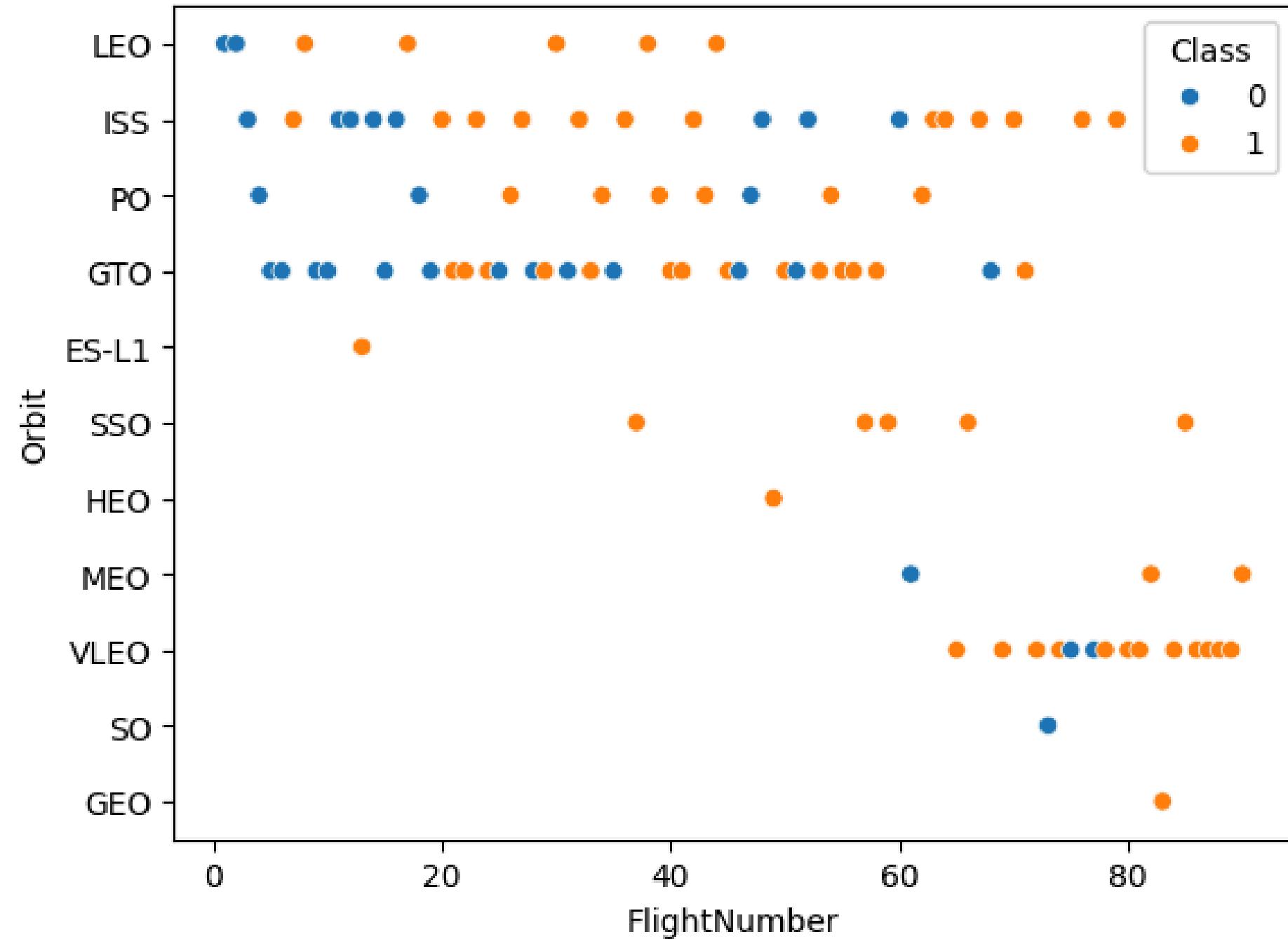
CCAFS SLC 40 and KSC LC 39A handle a wide range of payloads, with heavier payloads (above 10,000 kg) predominantly resulting in successful launches. VAFB SLC 4E primarily launches lighter payloads, which also show a high success rate. Success becomes more frequent as payload mass increases beyond 7,000 kg, particularly at KSC LC 39A and CCAFS SLC 40, indicating SpaceX's growing reliability in handling heavy payloads. This suggests improved performance and technology over time, especially for large-scale payload missions.

SUCCESS RATE VS. ORBIT TYPE



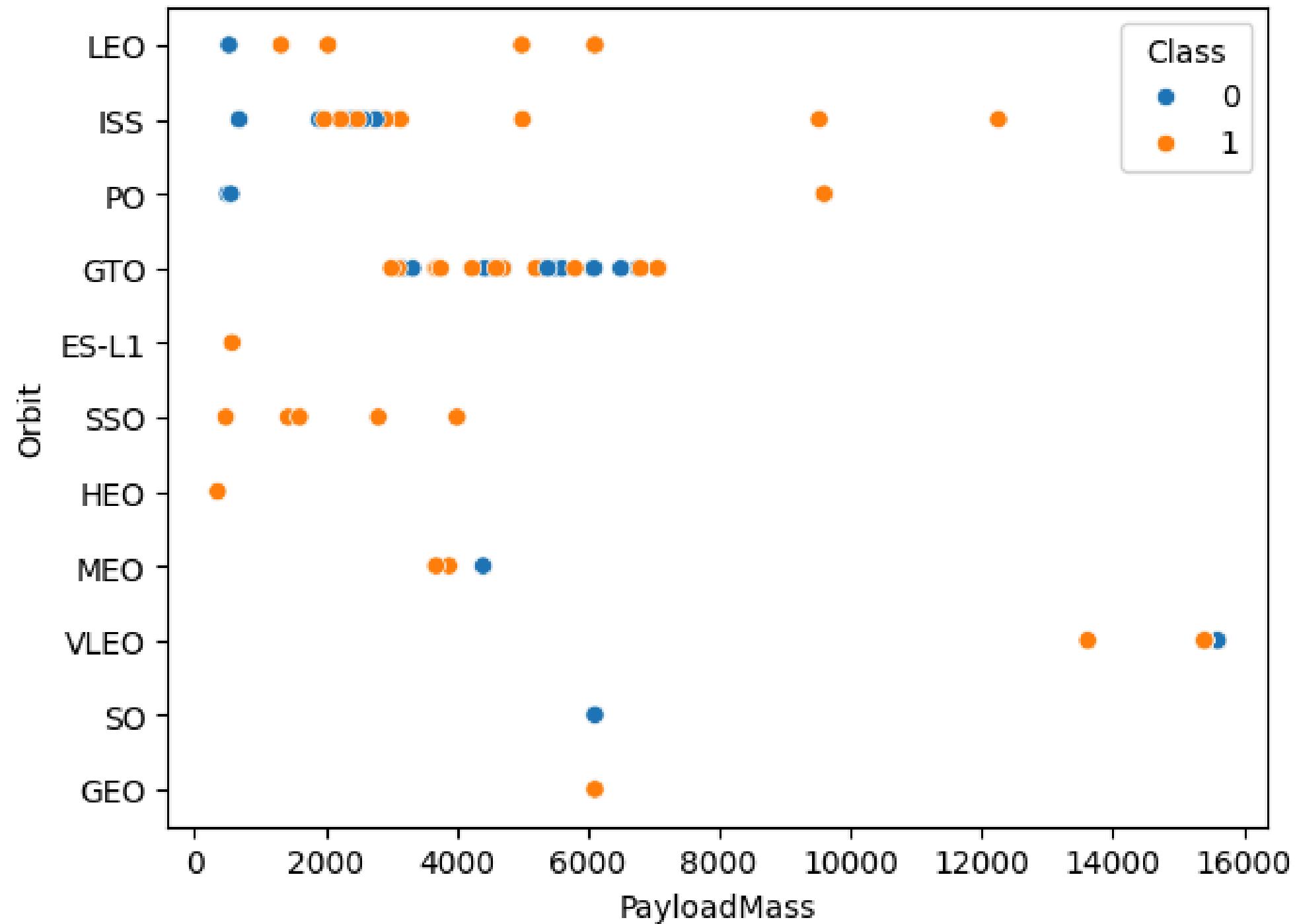
Orbits ES-L1, GEO, HEO and SSO show a 100% success rate, reflecting consistently reliable missions for these destinations. In contrast, SO has a 0% success rate, while LEO, MEO, and PO exhibit moderate success rates around 60%. This data underscores SpaceX's strong performance across most orbital types, with occasional challenges in specific orbits like SO, which may involve more complex mission parameters or environmental factors.

FLIGHT NUMBER VS. ORBIT TYPE



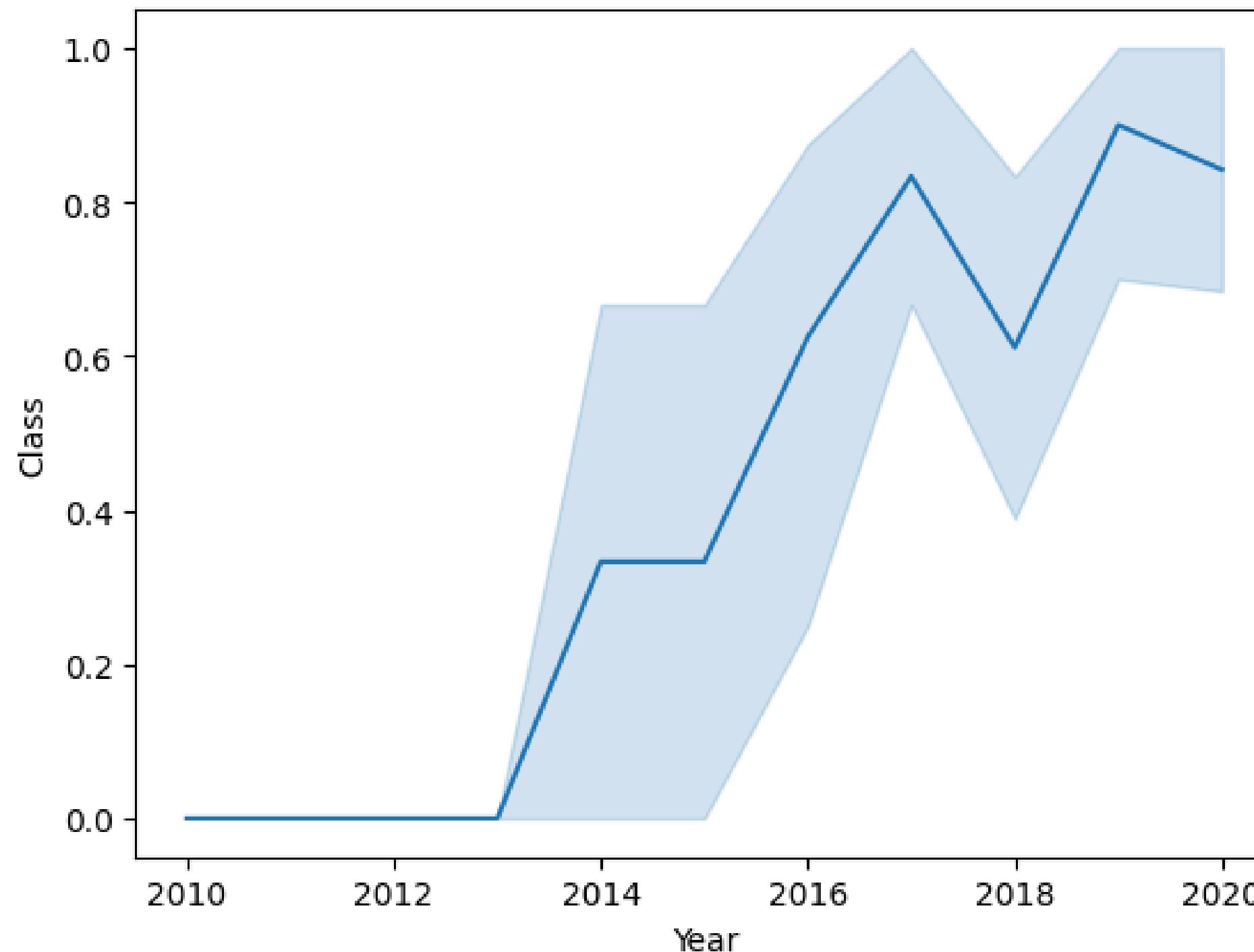
The data reveals that earlier flights exhibit a mix of successes and failures across most orbits. However, in general, as flight numbers increase, the proportion of successful launches grows like seen on the LEO orbit, indicating improved reliability and mission success over time. In contrast, orbits like GTO display a mix of outcomes, reflecting varying levels of mission complexity where there seems to be no relationship between flight number and success.

PAYLOAD VS. ORBIT TYPE



Heavier payloads, primarily directed to orbits like LEO and ISS, tend to be more successful, as seen by the predominance of orange points in those regions. Certain orbits, such as SSO exhibit consistently successful outcomes across all payload masses. However, for orbits such as GTO there is no distinguishable pattern of success.

LAUNCH SUCCESS YEARLY TREND



The data reveals a clear upward trend in success rates starting from 2013, with steady improvements through 2016. After reaching a high success rate in 2017, there is a slight dip in 2018, followed by recovery in subsequent years. This graph demonstrates SpaceX's growing reliability and technological advancements over the decade, reflecting a consistent increase in mission success as the company refined its processes and technologies.

ALL LAUNCH SITE NAMES

These launch sites collectively represent critical infrastructure for SpaceX's diverse range of missions, enabling launches into various orbital trajectories from both coasts of the United States:

1. CCAFS LC-40
2. VAFB SLC-4E
3. KSC LC-39A
4. CCAFS SLC-40

LAUNCH SITE NAMES BEGIN WITH 'CCA'

This dataset provides details of SpaceX rocket launches from CCAFS LC-40. Information included: Launch dates, rocket and payloads, payload mass, orbit type, mission customers and mission outcomes.

- 1.('2010-06-04', '18:45:00', 'F9 v1.0 B0003', 'CCAFS LC-40', 'Dragon Spacecraft Qualification Unit', 0, 'LEO', 'SpaceX', 'Success', 'Failure (parachute)')
- 2.('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)')
- 3.('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')
- 4.('2012-10-08', '0:35:00', 'F9 v1.0 B0006', 'CCAFS LC-40', 'SpaceX CRS-1', 500, 'LEO (ISS)', 'NASA (CRS)', 'Success', 'No attempt')
- 5.('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

The result shows that the total payload mass delivered by SpaceX for missions contracted by NASA is 45,596 kilograms. This value represents the cumulative payload mass carried by SpaceX under the CRS program, which includes multiple resupply missions to the International Space Station (ISS).

45,596 kg

AVERAGE PAYLOAD MASS BY F9 V1.1

The result indicates that the average payload mass delivered by SpaceX launches using the Falcon 9 v1.1 booster version is approximately 2,534.67 kilograms. This value represents the mean payload mass for all missions conducted with this specific booster version, reflecting its payload capacity and mission requirements.

2534.67 kg

FIRST SUCCESSFUL GROUND LANDING DATE

The result shows that the earliest successful mission recorded in the SpaceX database occurred on **June 4, 2010**. This was likely SpaceX's first successful launch of the Falcon 9 rocket, marking a significant milestone in the company's history.

2010-06-04

SUCCESSFUL DRONE SHIP LANDING WITH PAYLOAD BETWEEN 4000 AND 6000

The result lists the booster versions used in SpaceX missions that were successful and carried payloads weighing between 4,000 kg and 6,000 kg. The boosters include various iterations of the Falcon 9 v1.1, Falcon 9 FT (Full Thrust), and Falcon 9 Block 4 and Block 5 models.

.....

F9 v1.1	F9 FT B1030	F9 B5 B1046.2
F9 v1.1 B1011	F9 FT B1021.2	F9 B5 B1047.2
F9 v1.1 B1014	F9 FT B1032.1	F9 B5 B1048.3
F9 v1.1 B1016	F9 B4 B1040.1	F9 B5 B1051.2
F9 FT B1020	F9 FT B1031.2	F9 B5B1060.1
F9 FT B1022	F9 FT B1032.2	F9 B5 B1058.2
F9 FT B1026	F9 B4 B1040.2	F9 B5B1062.1

TOTAL NUMBER OF SUCCESSFUL AND FAILURE MISSION OUTCOMES

The query results show that there were 100 successful missions and 1 failure during flight. This demonstrates that SpaceX has an overwhelmingly high success rate, with very few in-flight failures recorded in the dataset.

('Success', 100),

('Failure (in flight)', 1)

BOOSTERS CARRIED MAXIMUM PAYLOAD

The result lists the booster versions used for SpaceX missions that carried the maximum payload mass recorded in the dataset. All the boosters are of the Falcon 9 Block 5 version, this indicates that the Falcon 9 Block 5, being the most advanced and capable version of the Falcon 9 series, has been used to handle SpaceX's heaviest payloads, reflecting its superior payload capacity and reusability.

('F9 B5 B1048.4',), ('F9 B5 B1049.5',),

('F9 B5 B1049.4',), ('F9 B5 B1060.2 ',),

('F9 B5 B1051.3',), ('F9 B5 B1058.3 ',),

('F9 B5 B1056.4',), ('F9 B5 B1051.6',),

('F9 B5 B1048.5',), ('F9 B5 B1060.3',),

('F9 B5 B1051.4',), ('F9 B5 B1049.7 ',)

2015 LAUNCH RECORDS

This result shows two launches in 2015 where the boosters failed to successfully land on the drone ship, both using the Falcon 9 v1.1 booster and launching from CCAFS LC-40. This reflects early challenges SpaceX faced while developing their landing and reusability technology.

('01', 'Failure (drone ship)', 'F9 v1.1 B1012', 'CCAFS LC-40')

('04', 'Failure (drone ship)', 'F9 v1.1 B1015', 'CCAFS LC-40')

RANK LANDING OUTCOMES BETWEEN 2010-06-04 AND 2017-03-20

Between 2010-06-04 and 2017-03-20, SpaceX's landing outcomes show significant progress in recovery technology. The most frequent outcome was "No attempt" (10 times) in the early years, reflecting initial challenges. Successful recoveries improved over time, with "Success (drone ship)" occurring 5 times after 2016-04-08. Early parachute recovery failures (2 times) highlight SpaceX's transition to advanced landing technologies, achieving consistent success by 2017.

- 1.('Success (drone ship)', '2016-04-08', 5),
- 2.('Success (ground pad)', '2015-12-22', 3),
- 3.('Precluded (drone ship)', '2015-06-28', 1),
- 4.('Failure (drone ship)', '2015-01-10', 5),
- 5.('Controlled (ocean)', '2014-04-18', 3),
- 6.('Uncontrolled (ocean)', '2013-09-29', 2),
- 7.('No attempt', '2012-05-22', 10),
- 8.('Failure (parachute)', '2010-06-04', 2)

SECTION 3

LAUNCH SITES

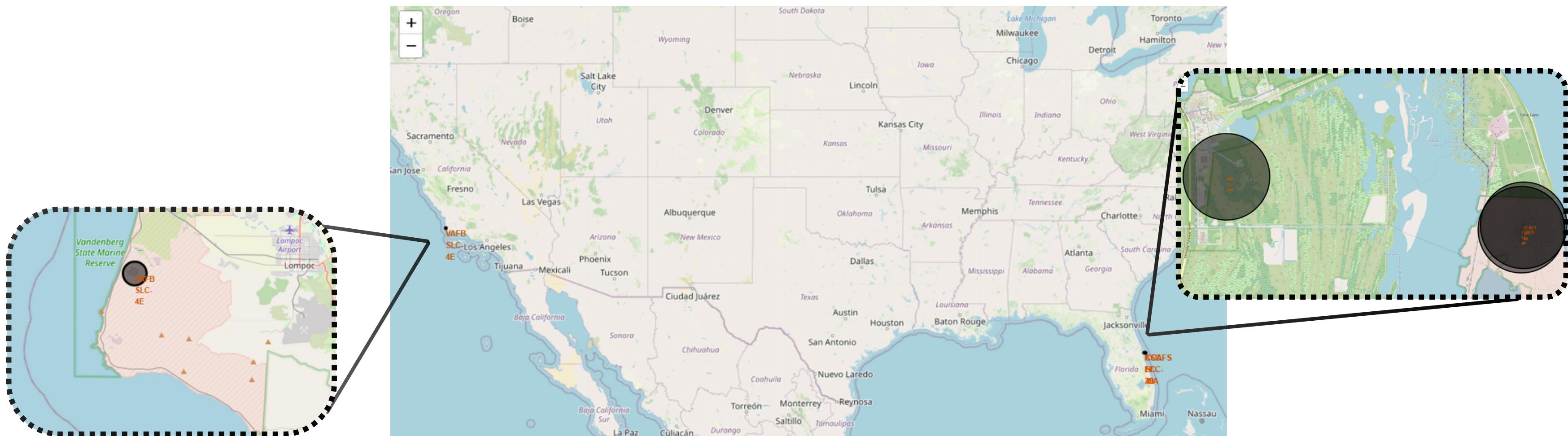
PROXIMITIES

ANALYSIS



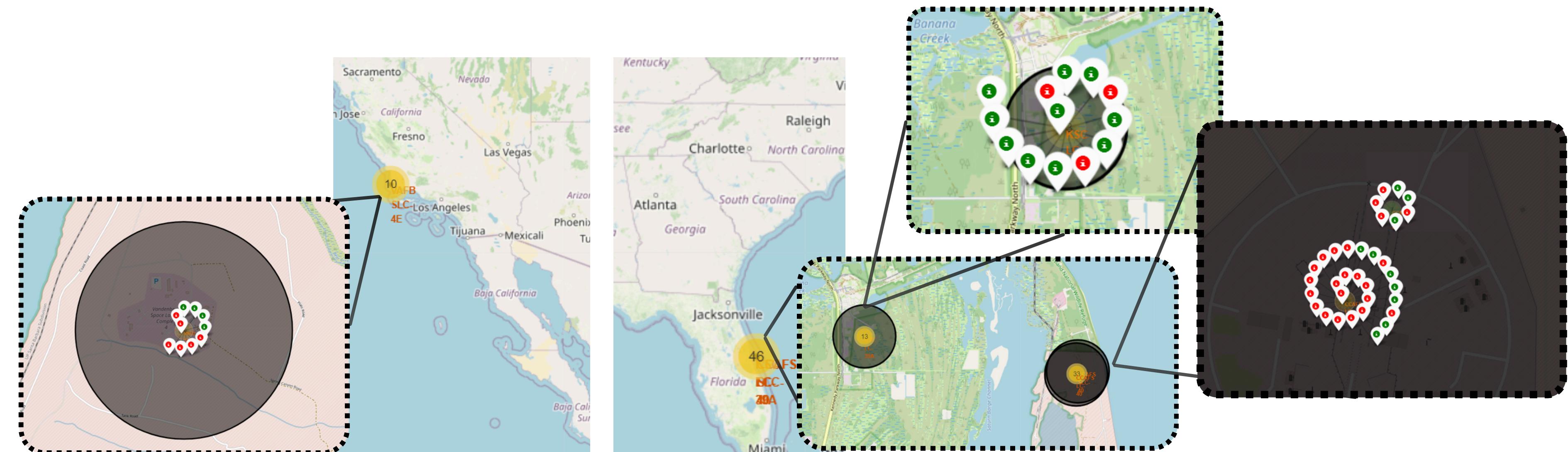
ALL LAUNCH SITES

The map highlights the locations of all the launch sites, strategically positioned along the coasts and in southern regions. This placement leverages the Earth's faster rotational speed near the equator, enhancing launch efficiency, and ensures proximity to the oceans, providing a safer trajectory by minimizing risks to populated areas in case of launch anomalies.



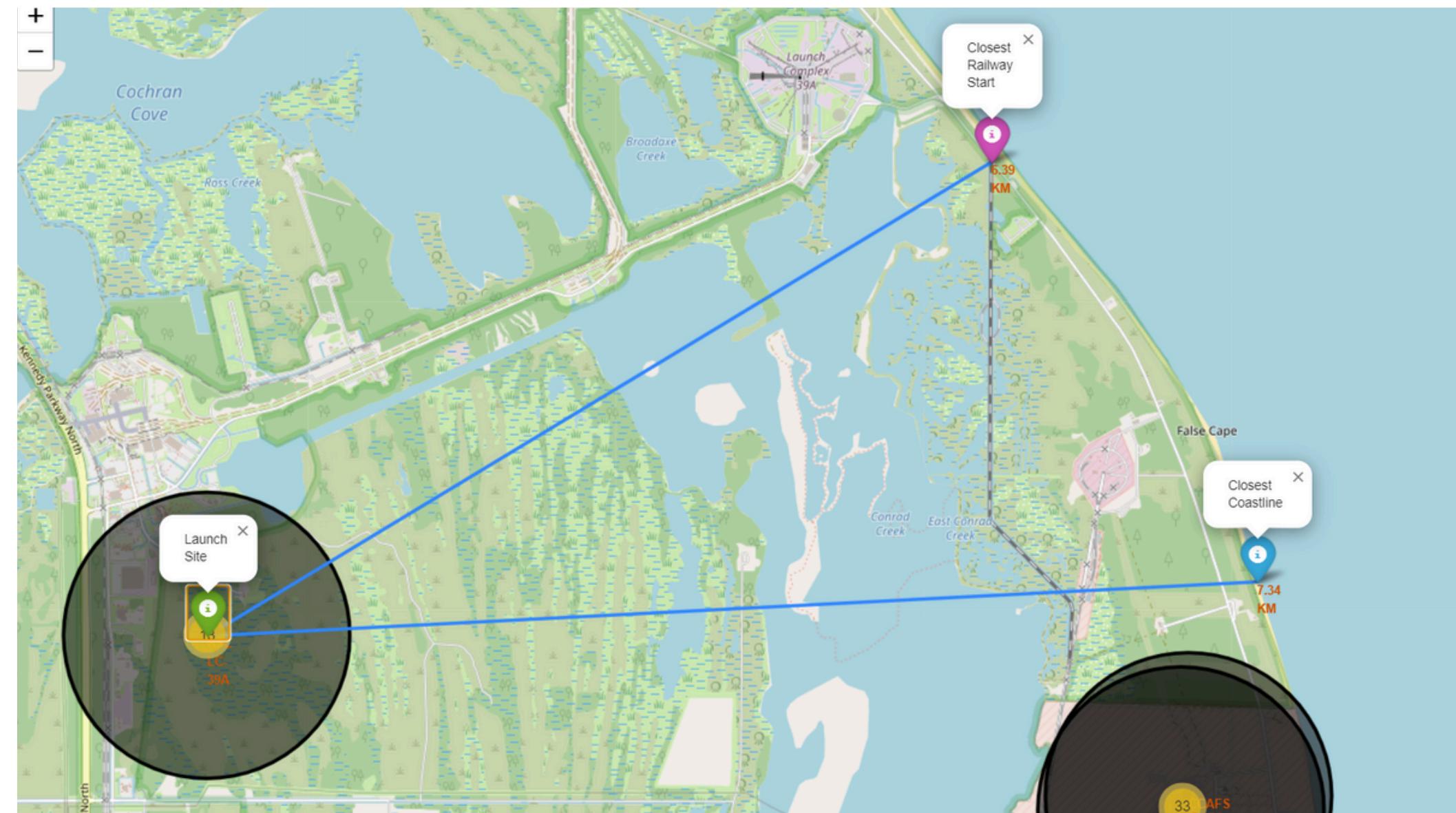
ALL LAUNCH SITES

This map provides a detailed breakdown of the successes and failures across various launch sites. On the west coast, VAFB SLC-4E accounts for 10 launches, with 6 failures and 4 successes, primarily serving polar orbit missions. On the east coast, there are three active launch sites in Florida: CCAFS LC-40, KSC LC-39A, and CCAFS SLC-40, collectively hosting 46 launches. CCAFS LC-40 has conducted 26 launches, with 19 failures and 7 successes, marking it as the site with the most failures. KSC LC-39A, on the other hand, boasts the highest success rate, with 13 launches resulting in 10 successes and only 3 failures. CCAFS SLC-40, hosting 7 launches, contributes to the overall activity in Florida. These figures highlight the distribution of launches and the varying success rates across SpaceX's launch infrastructure, with the east coast playing a dominant role in mission activity and KSC LC-39A emerging as the most reliable site.



DISTANCES TO COASTLINE AND RAILWAY

This map illustrates the proximity of the launch site KSC LC-39A to the nearest coastline and the start of the railway system. The distance to the coastline is approximately 7.34 km, while the distance to the start of the railway is about 6.39 km, as shown on the map. These distances highlight the strategic positioning of the launch site, ensuring a safe buffer zone from the coast for public safety during launches, while maintaining close access to rail transport for efficient logistics and transportation of heavy rocket components or payloads.



SECTION 4

BUILD A

DASHBOARD

WITH PLOTLY

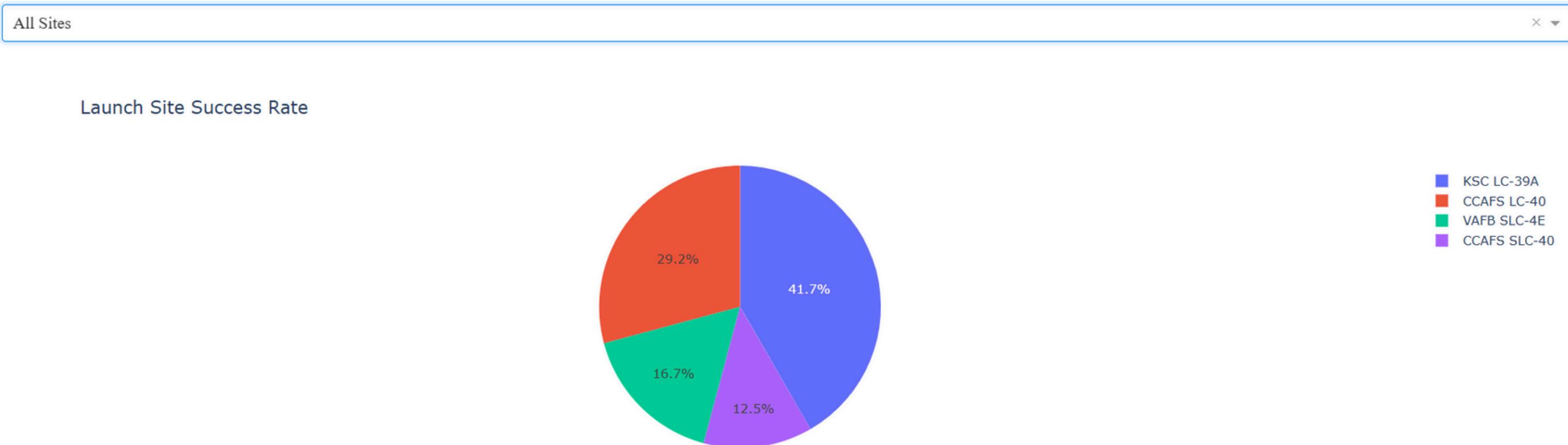
DASH



SUCCESS DISTRIBUTION BY LAUNCH SITE

This dashboard uses a pie chart to illustrate the distribution of successful launches across all launch sites. The site with the highest proportion of successes is KSC LC-39A, accounting for 41.7% of the total successful launches. In contrast, CCAFS SLC-40 has the lowest contribution, with only 12.5% of the successes. This visualization highlights the significant performance disparity between the launch sites, with KSC LC-39A standing out as the most reliable and productive site for successful missions, while CCAFS SLC-40 lags behind in terms of successful outcomes.

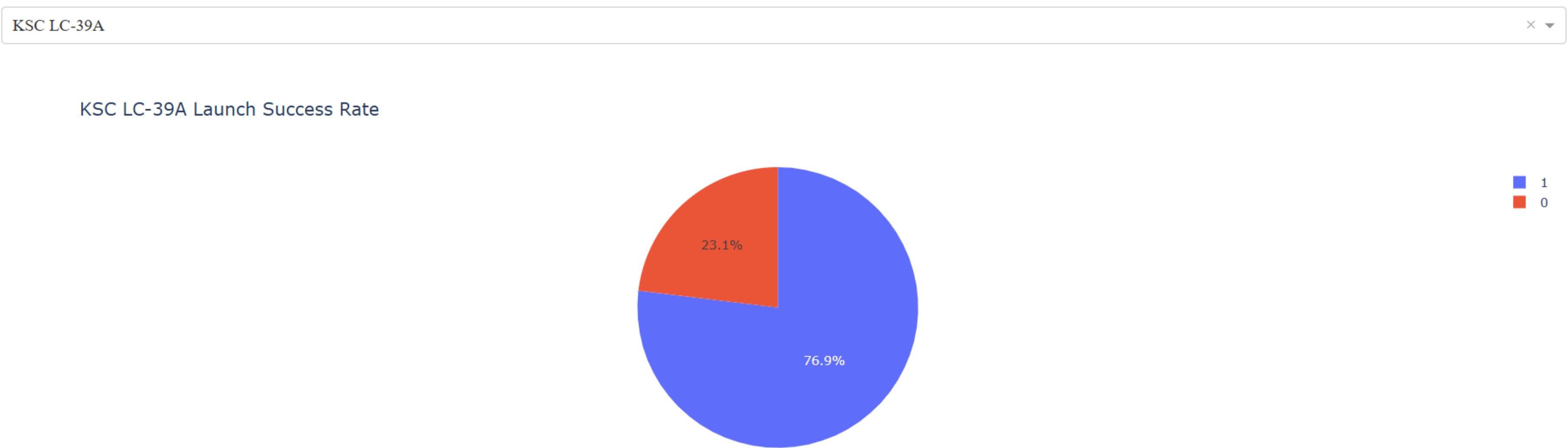
SpaceX Launch Records Dashboard



SUCCESS RATIO OF KSC LC-39A

The launch site KSC LC-39A, previously identified in the dashboard as having the highest distribution of successes, also boasts the highest success rate, as shown in the pie chart below, with an impressive success rate of 76.9%. This emphasizes its reliability and efficiency as SpaceX's most consistently successful launch site.

SpaceX Launch Records Dashboard



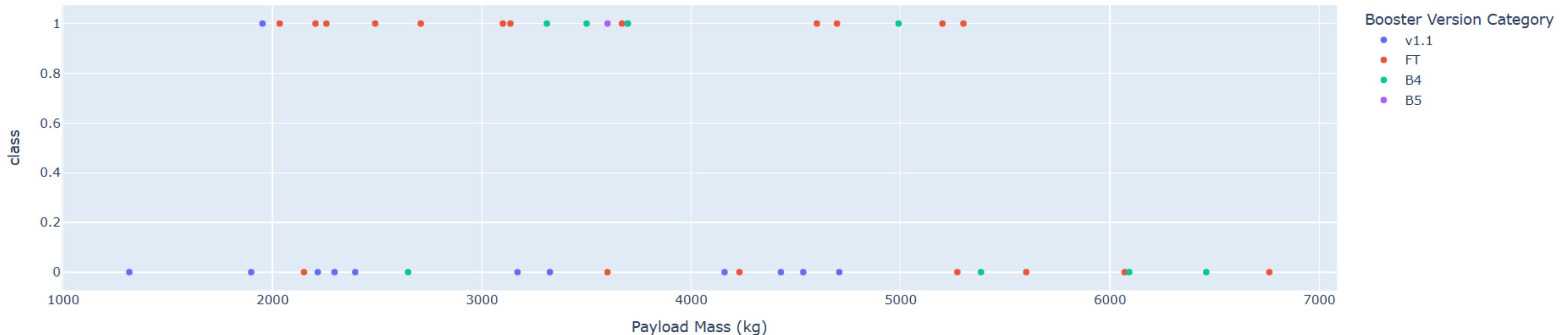
SCATTER PLOT PAYLOAD VS. SUCCESS

This graph showcases SpaceX's advancement in mission reliability across various booster versions. A noticeable cluster between the 2,000 to 4,000 kg payload range shows a higher concentration of successful launches. Earlier booster versions, such as v1.1, faced challenges with both lower and medium payloads, resulting in a mix of successes and failures. In contrast, newer versions like FT, B4, and B5 demonstrate significantly higher success rates, even as payload mass increases. This trend reflects SpaceX's technological progress and growing ability to handle a wider range of payloads with improved reliability and efficiency.

Payload range (Kg):



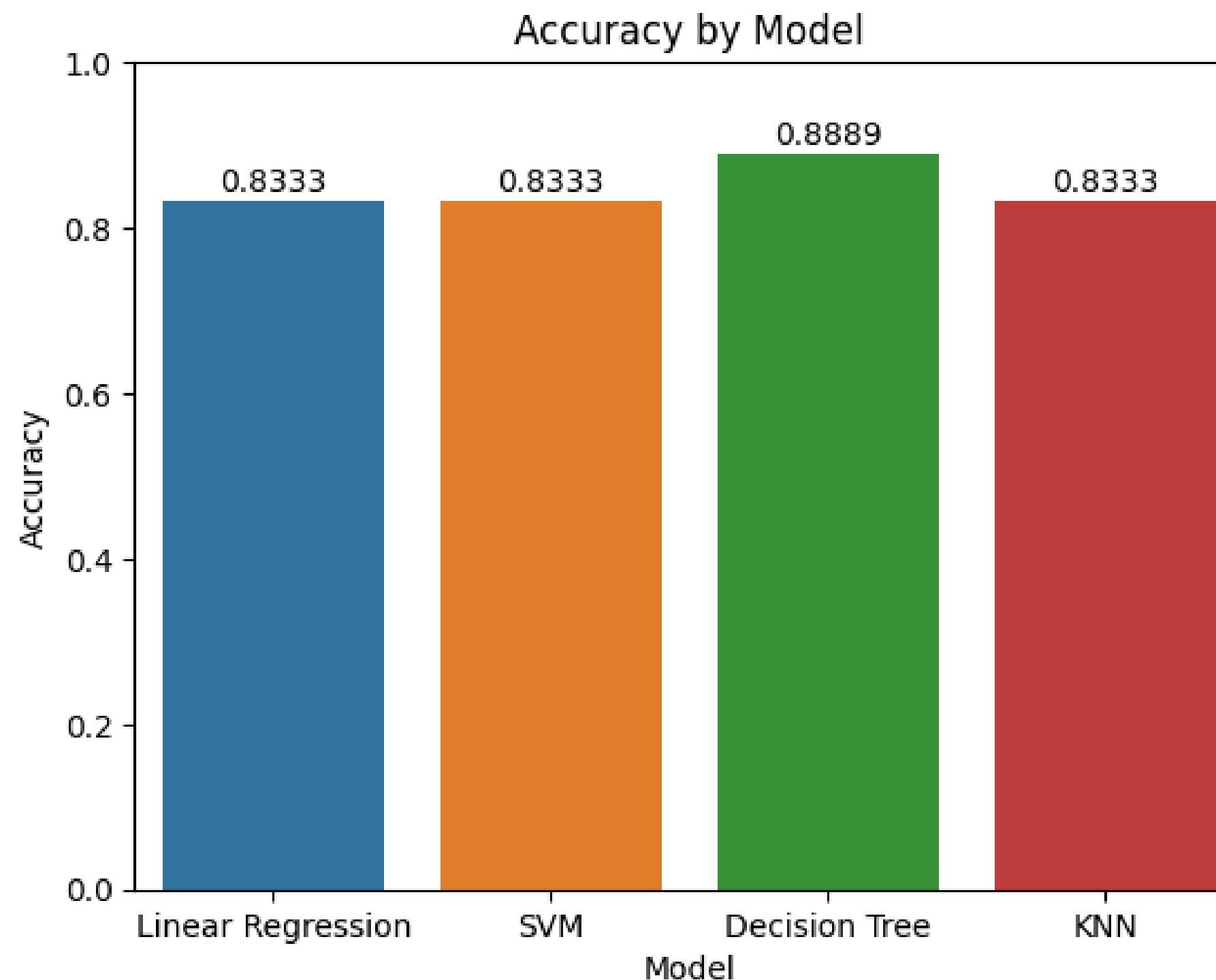
ALL Scatter Plot Payload vs. Success



SECTION 5 PREDICTIVE ANALYSIS (CLASSIFICATION)

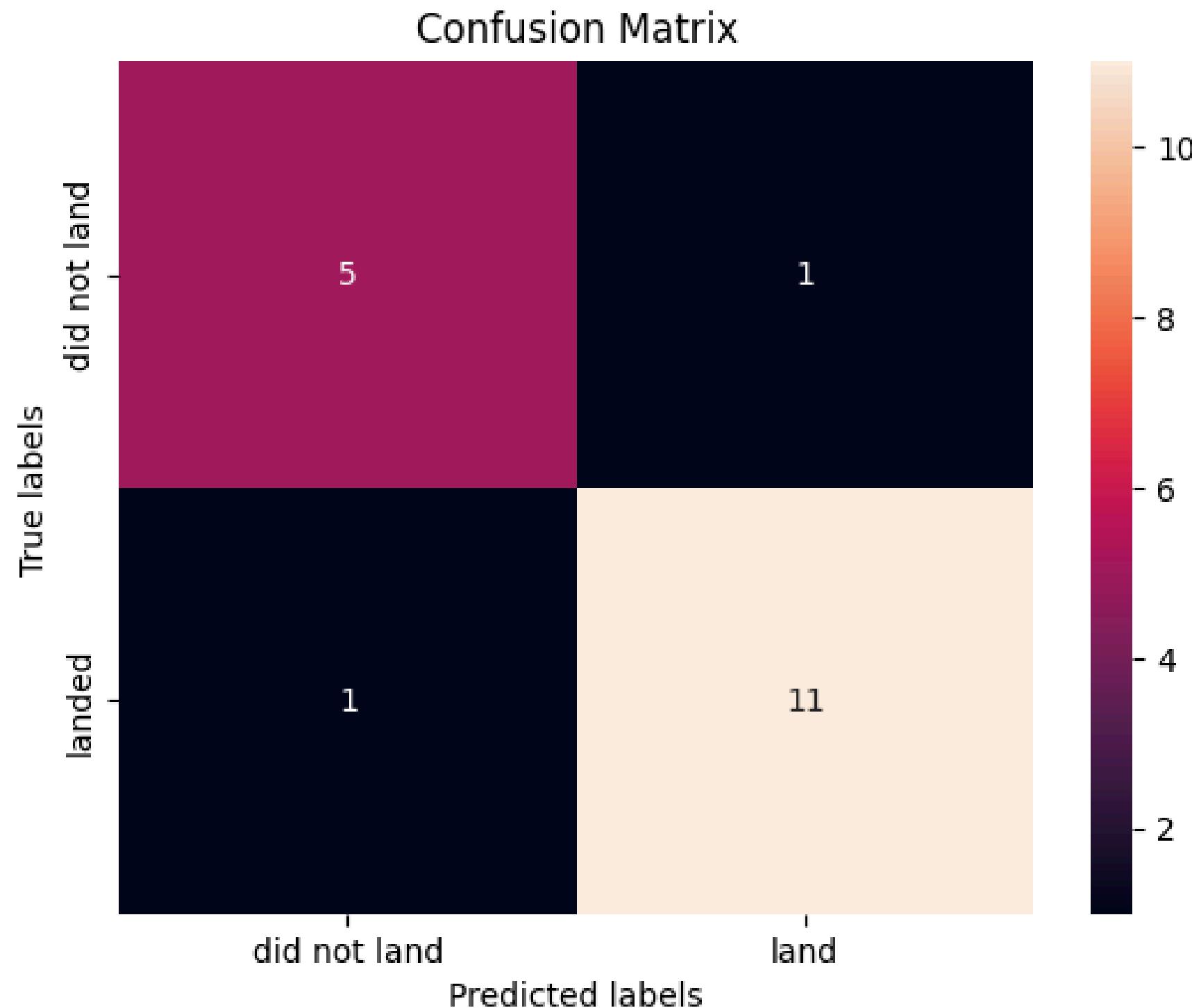


CLASSIFICATION ACCURACY



This graph illustrates the accuracy of different classification models in predicting whether the first stage of a Falcon 9 rocket will successfully return, based on relevant features. Four models were evaluated: Linear Regression, Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors (KNN). Linear Regression, SVM, and KNN achieved identical accuracies of 83.3% (0.833 repeating), while the Decision Tree emerged as the most effective model with an accuracy of 88.9% (0.8888 repeating). This comparison highlights the Decision Tree's superior performance in this specific classification task.

CONFUSION MATRIX



This is the confusion matrix for the best-performing model, the Decision Tree. The matrix shows 11 true positives, correctly predicting successful rocket landings, and 5 true negatives, accurately identifying instances where the rocket did not return. In contrast, the model made only 1 false positive, incorrectly predicting a rocket would land when it didn't, and 1 false negative, failing to predict a successful landing. These results highlight the model's strong predictive accuracy, with minimal misclassifications.

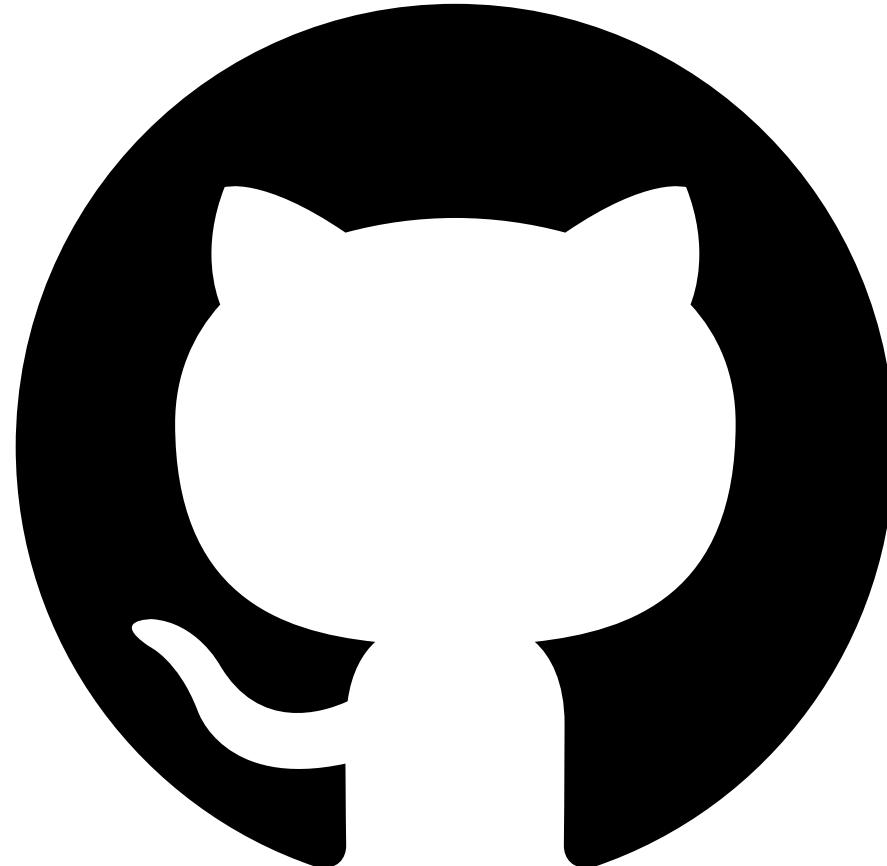
CONCLUSIONS

- 1. Launch Site Performance:** SpaceX's success rates vary across launch sites, with KSC LC-39A emerging as the most reliable site, achieving the highest success rate (76.9%) and accounting for the largest proportion of successful missions (41.7%). Conversely, CCAFS SLC-40 experienced the highest number of failures, reflecting early challenges that SpaceX has since overcome.
- 2. Orbital Reliability Trends:** Specific orbits, such as ES-L1, GEO, HEO, and SSO, show a 100% success rate, indicating consistent reliability. In contrast, orbits like GTO and SO face challenges, with GTO exhibiting a mixed pattern and SO failing completely.
- 3. Continuous Improvement in Reliability and Technology:** SpaceX has demonstrated steady growth in launch success rates, driven by significant technological advancements. Since 2013, newer booster versions like FT, B4, and B5 have achieved higher success rates, even with heavier payloads, reflecting improvements in design, engineering, and operational processes. While earlier models struggled with reliability, later iterations consistently increased success rates, underscoring SpaceX's ability to innovate and refine its technology over time.
- 4. Predictive Modeling:** Decision Tree emerged as the most accurate model for predicting Falcon 9 first-stage landing success, achieving an accuracy of 88.9%. The confusion matrix confirms the model's robustness, with only minimal misclassifications, further validating the feasibility of predictive analytics in space operations.
- 5. Strategic Launch Site Placement:** SpaceX's launch sites are strategically located along coastal regions to leverage the Earth's rotational speed and ensure safety, which supports efficient and reliable operations.

SpaceX's consistent advancements in technology, operational efficiency, and booster reliability have positioned it as a leader in space exploration. With a growing capacity to handle larger payloads and more complex missions, SpaceX is not only improving its success rates but also setting new benchmarks for cost-effective and sustainable space travel. By utilizing predictive models like Decision Trees, which demonstrated high accuracy in classifying rocket landing outcomes, we can enhance our own capabilities to optimize launch reliability and determine cost of launch.

APPENDIX

The GitHub repository linked below contains all the notebooks and scripts utilized in this data science project.



[GITHUB.COM/ALE-MONTE/DATA-SCIENCE-CAPSTONE](https://github.com/Ale-Monte/Data-Science-Capstone)



THANK YOU

Alejandro Montenegro López

November 30, 2024