IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Alexey Tyurin
1/4/2024

Outline

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

Executive Summary

- Summary of methodologies

The project encompasses a range of methodologies, including machine learning for SpaceX predictions, exploratory data analysis (EDA), data visualization techniques, SQL database interactions, and the development of a Dash app for visualizing SpaceX data. Additionally, the project makes use of datasets for analysis and model training.

- Summary of all results

The results encompass the comprehensive outcomes of the project, specifically detailing the findings from exploratory data analysis, aspects of interactive analytics, and insights from predictive analysis.

Introduction

- Project background and context

The project’s background and context revolve around predicting the successful landing of SpaceX’s Falcon 9 rocket’s first stage.

A significant aspect of SpaceX's cost advantage, offering launches at $62 million compared to competitors' $165 million, lies in reusing this first stage. Accurately predicting its landing success is key to determining launch costs, providing crucial data for competitive bidding in the aerospace industry. The project involves data collection and formatting from an API, analyzing both successful and controlled unsuccessful landings, like those in oceans, planned by SpaceX for various operational reasons.

- Problems you want to find answers

In the context of this project, the problems we aim to address are focused on determining the key factors that influence the successful landing of SpaceX's Falcon 9 first stage, assessing the reliability and efficiency of the Falcon 9 launch system, and predicting the cost-effectiveness of SpaceX launches for potential competitors. These problems involve analyzing and interpreting complex aerospace data to gain insights that could impact strategic decisions in the space launch industry.

Introduction

- Falcon 9's Launch Success, Booster Landings, and Payload Capacity

The Falcon 9 rocket family from SpaceX has an impressive record of launches and landings. Over 14 years, Falcon 9 rockets have been launched 296 times, achieving a 99.3% full mission success rate. Specifically, the Falcon 9 Block 5 version has had 231 missions, all of which were successful. In 2022, Falcon 9 set a new record with 60 launches in a single calendar year, surpassing the previous record held by the Soyuz-U rocket.

In terms of landing success, Falcon 9 first-stage boosters have successfully landed in 261 of 272 attempts, which is about 96% success rate. The Falcon 9 Block 5 version boasts an even higher success rate with 98.3% of its 240 landing attempts being successful.

The Falcon 9 rocket uses several versions, including v1.0, v1.1, and Full Thrust, leading up to the latest Block 5 variant. Each version has seen improvements in terms of fuel capacity, engine power, and landing capabilities. For example, the v1.0 and v1.1 versions have specific impulse ratings at sea level of 275 s and 282 s respectively, while the Full Thrust version, which includes Block 5, has further optimized these specifications.

The Falcon 9 rocket is not only a medium-lift launch vehicle but can also function as a heavy-lift launch vehicle when needed. It is notable for being the first commercial rocket to launch humans into orbit and is currently the only U.S. rocket certified for transporting humans to the International Space Station (ISS). The rocket's heaviest payloads to geostationary transfer orbit (GTO) include the Intelsat 35e at 6,761 kg and the Telstar 19V at 7,075 kg.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:

  The project employs two main data collection methods: utilizing the SpaceX REST API for structured launch data and web scraping from Wikipedia for supplementary historical and contextual information about Falcon 9 launches.

- Perform data wrangling

  Data wrangling for this project entails the extraction and structuring of Falcon 9 launch records. It includes thorough data preparation—resolving missing values and applying preprocessing techniques such as One-Hot Encoding and Standardization—to optimize the dataset for machine learning applications.

Methodology (cont.)

Executive Summary

- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

  Perform predictive analysis by selecting, training, tuning, and valuating classification models on our well-prepared dataset. Choose the best-performing model for the problem. If necessary, deploy it for real-world predictions, ensuring continuous monitoring and updates for optimal performance.

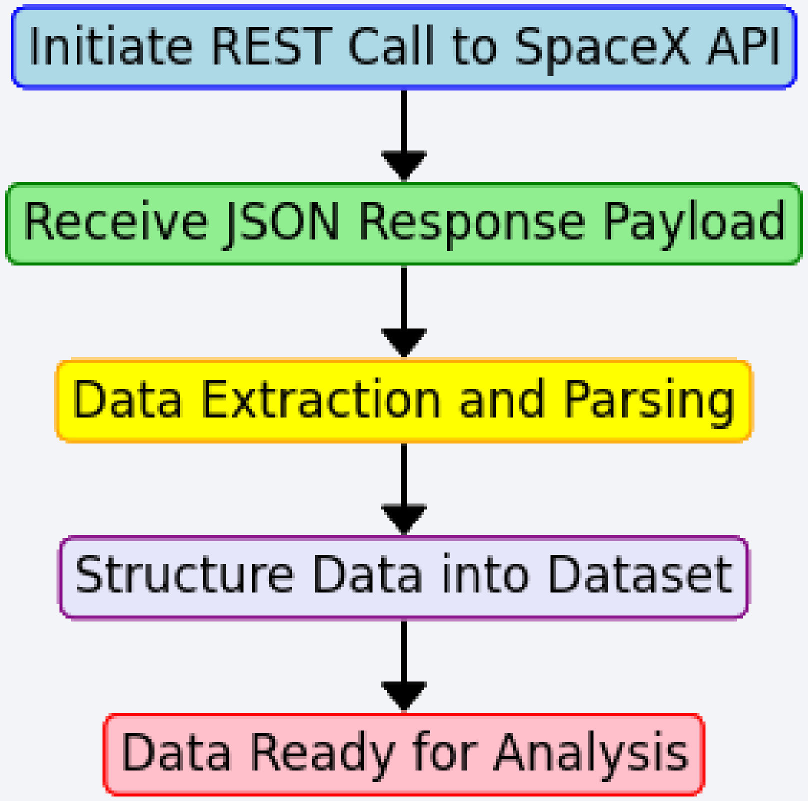
Data Collection

The data collection methodology for the project includes two primary approaches:

- First, we access SpaceX's REST API to gather structured data directly related to Falcon 9 launches. This data includes details about each mission, the first stage landing outcomes, and other relevant technical specifications.
- Second, we employ web scraping techniques on Wikipedia to extract additional contextual information and historical data about Falcon 9 launches and SpaceX's operations. This approach complements our data collection, potentially offering a broader view and deeper insights for our analysis.

Data Collection – SpaceX API

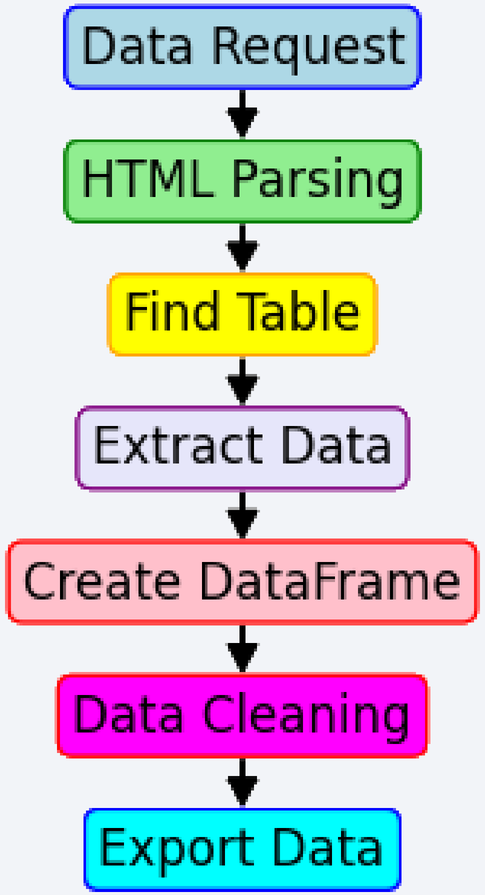
- Initiate REST Call to SpaceX API:
  This is the first step where a request is sent to the SpaceX API. The request specifies the endpoint and any necessary parameters or headers
- Receive JSON Response Payload:
  Once the API receives the request, it responds with data in JSON format. This response contains the information requested from the API
- Data Extraction and Parsing:
  The JSON payload is processed to extract relevant information. This involves parsing the JSON structure to isolate the specific data elements related to SpaceX launches
- Structure Data into Dataset:
  The extracted information is then structured into a format suitable for analysis. This could be a tabular dataset, a database, or any other structured data format
- Data Ready for Analysis:
  The final step involves having the data in a format that is ready for analysis. This could involve additional steps like data cleaning or transformation, depending on the specific requirements of the analysis



GitHub URL: [Collecting the data]

Data Collection - Scraping

- Data Request
  Use Python's requests library to fetch the HTML content of the Wikipedia page containing Falcon 9 launch records
- HTML Parsing
  Utilize BeautifulSoup to parse the fetched HTML content
- Table Extraction
  Identify and extract the relevant HTML table containing launch data
- Data Processing
  Apply helper functions to extract specific details from the table cells, such as date, time, booster version, and landing status
- DataFrame Creation
  Convert the extracted data into a structured format using Pandas DF
- Data Cleaning and Formatting
  Perform necessary data cleaning and formatting tasks to ensure data quality and consistency
- Exporting Data
  Save the processed data to a CSV file for further analysis or use in subsequent tasks

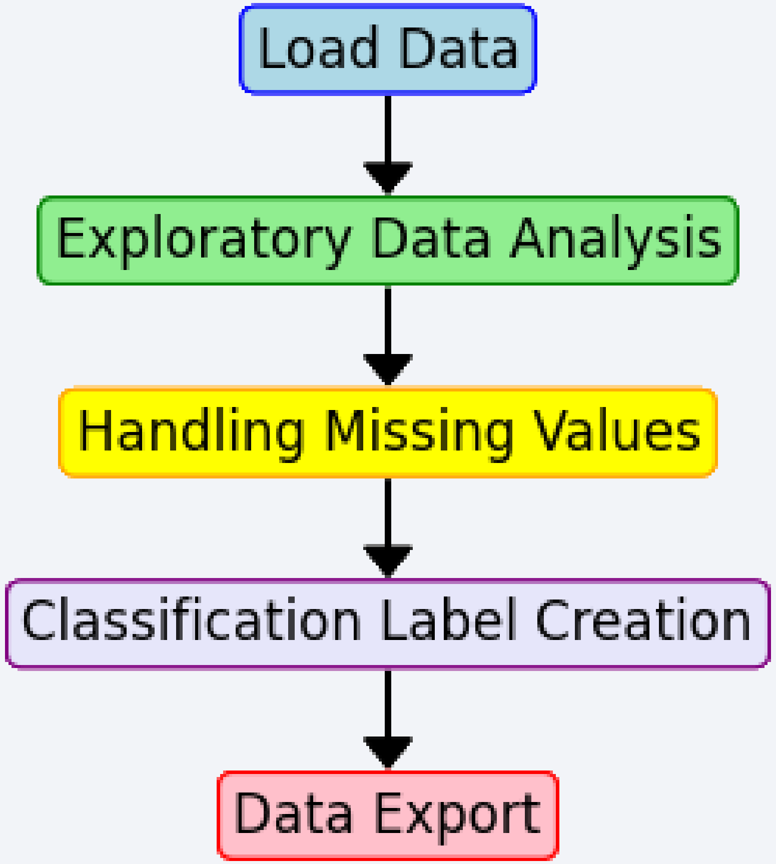


GitHub URL: Web scraping

Data Wrangling

Data wrangling involved several key steps:

- **Load Data**
  Import the SpaceX Falcon 9 launch dataset using Pandas.
- **Exploratory Data Analysis**
  Analyze various aspects like launch sites, orbits, and outcomes
- **Handling Missing Values**
  Identify and address missing data in the dataset
- **Classification Label Creation**
  Convert launch outcomes into binary training labels, where '1' represents a successful landing and '0' an unsuccessful one
- **Data Export**
  Save the processed data to a CSV file for further use

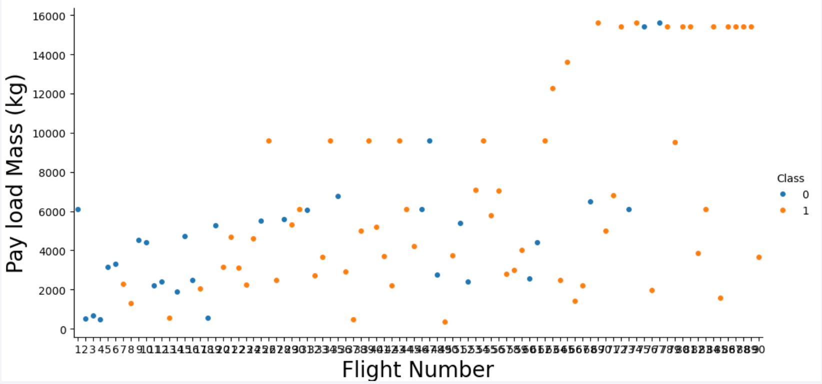


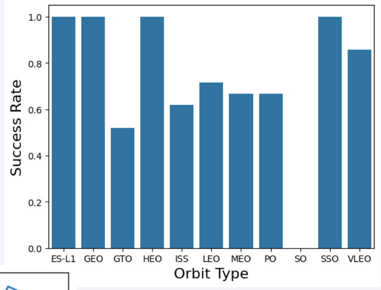
GitHub URL: [Data wrangling]

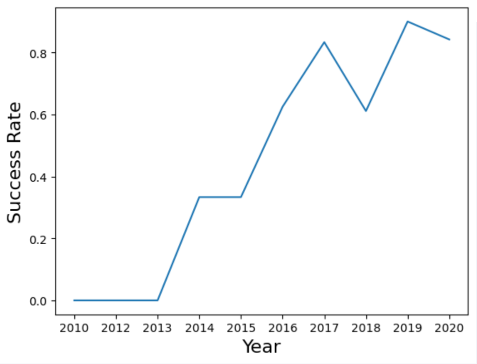
EDA with Data Visualization

Each chart was chosen to explore specific aspects of the data, helping to understand the factors influencing Falcon 9's landing success:

- **Scatter Plot (Flight Number vs. Payload Mass):** This chart was used to examine the relationship between the flight number and payload mass, overlaying launch outcomes to understand how these factors affect landing success.
- **Categorical Plot (Flight Number vs. Launch Site):** This visualized the relationship between flight numbers and different launch sites, with hues indicating the class or outcome of the launch.
- **Scatter Plot (Payload vs. Launch Site):** This chart helped in assessing if there's any correlation between the payload mass and the launch site used, also considering the launch outcome.
- **Bar Chart (Success Rate by Orbit Type):** This plot was utilized to compare the success rates across different orbit types.
- **Scatter Plot (Flight Number vs. Orbit Type):** This chart analyzed the relationship between flight numbers and orbit types, observing if the number of flights affects success in different orbits.
- **Scatter Plot (Payload vs. Orbit Type):** This visual helped in understanding the relationship between payload mass and orbit types, particularly looking at how heavy payloads affect landing success in different orbits.
- **Line Chart (Yearly Success Trend):** A line chart showing the trend in launch success rates over the years, based on the average annual success rate.







GitHub URL: Exploring and Preparing Data

EDA with SQL

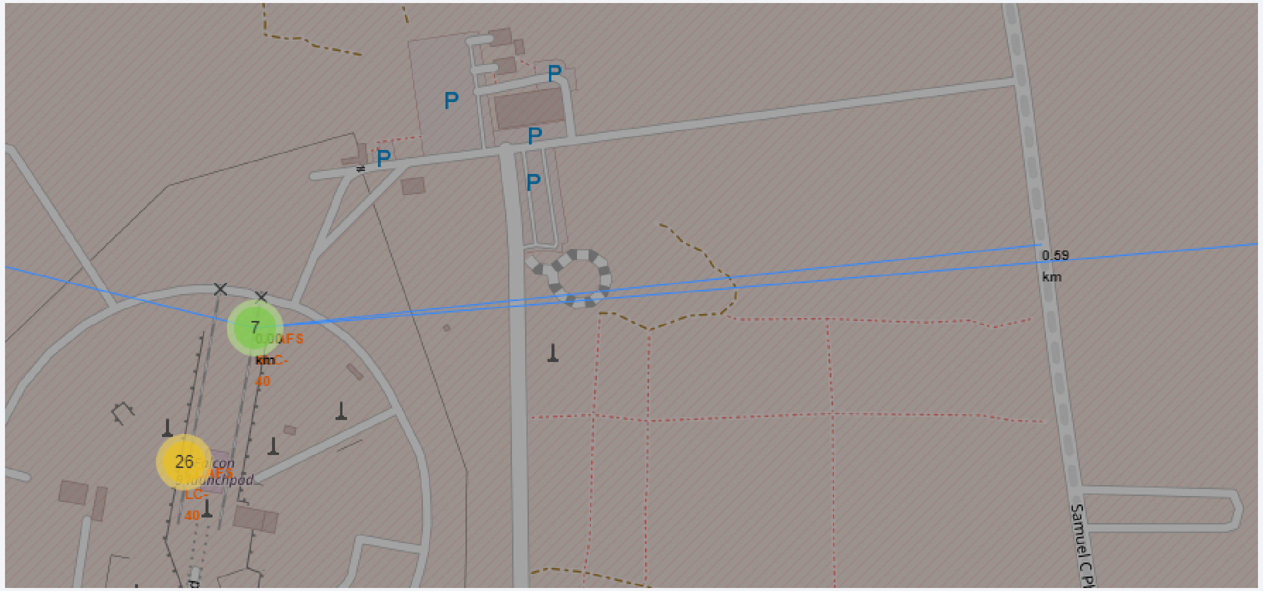
The SQL queries performed:

- Identified unique SpaceX launch sites using a SELECT DISTINCT query
- Displayed records for launch sites starting with 'CCA' using a WHERE clause with LIKE
- Calculated total payload mass for NASA (CRS) missions using SUM and WHERE
- Found average payload mass for Falcon 9 version 1.1 using AVG and WHERE
- Determined the date of the first successful ground pad landing using MIN and WHERE
- Listed boosters successful in drone ship landings with payload masses 4000-6000 kg
- Counted total number of successful and failed missions using GROUP BY
- Identified boosters carrying the maximum payload using a subquery
- Listed monthly failure outcomes in drone ships for 2015 using CASE and WHERE
- Ranked landing outcomes between two dates in descending order using RANK and GROUP BY.

GitHub URL: SQL Notebook

Build an Interactive Map with Folium

We created and added various objects to a Folium map for enhanced visual analytics of SpaceX launch sites:



- **Markers and Circles**

  Added for each launch site to visually represent their geographic locations.

- **Color-coded Markers**

  Indicated success (green) and failure (red) of launches at each site.

- **PolyLines**

  Drew lines between launch sites and key proximities like coastlines, cities, railways, and highways to visually assess their distances.

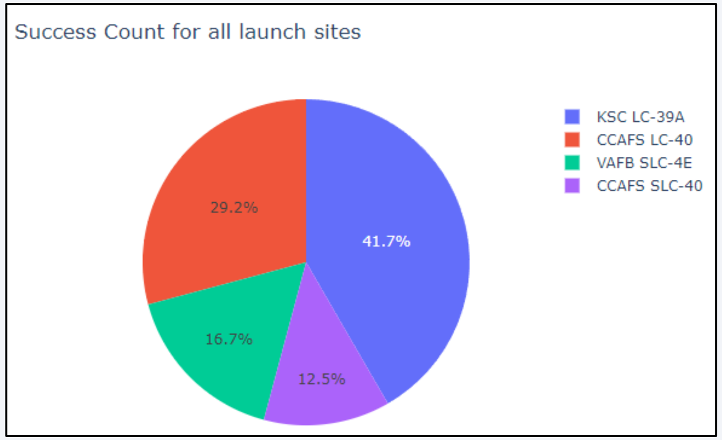
These objects were added to help identify geographical patterns and infrastructural proximities of launch sites, contributing to understanding factors influencing site selection and launch success.

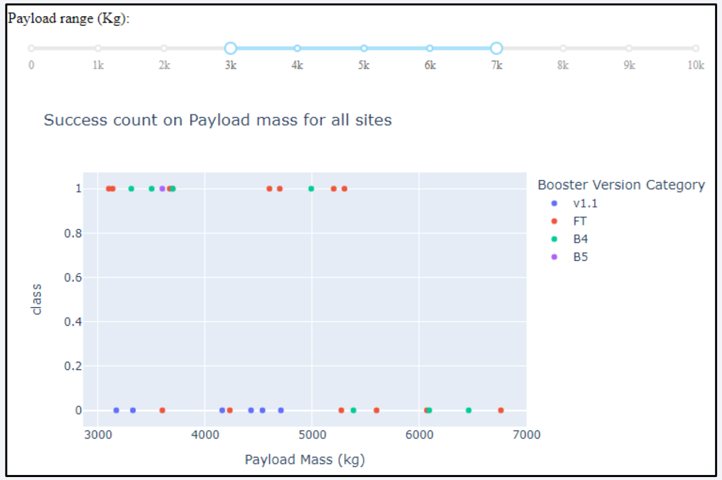
GitHub URL: Launch Sites Locations Analysis with Folium

Build a Dashboard with Plotly Dash

In the SpaceX Launch Records Dashboard:

- **Dropdown Menu for Launch Site Selection**
  Allows users to filter data based on a specific launch site or view all sites together. This interaction enables customized data exploration.
- **Pie Chart for Launch Success**
  Visualizes the success rate of launches, either for all sites combined or for a specific site selected from the dropdown. This helps in understanding the success trends at different sites.
- **Payload Range Slider**
  Allows users filter the data based on the payload mass range, enhancing the analysis of how payload affects launch success.
- **Scatter Plot for Payload vs. Success**
  Shows the correlation between payload mass and launch success, with color coding for booster versions. This plot is responsive to both site selection and payload range, offering deeper insights into factors affecting launch success.



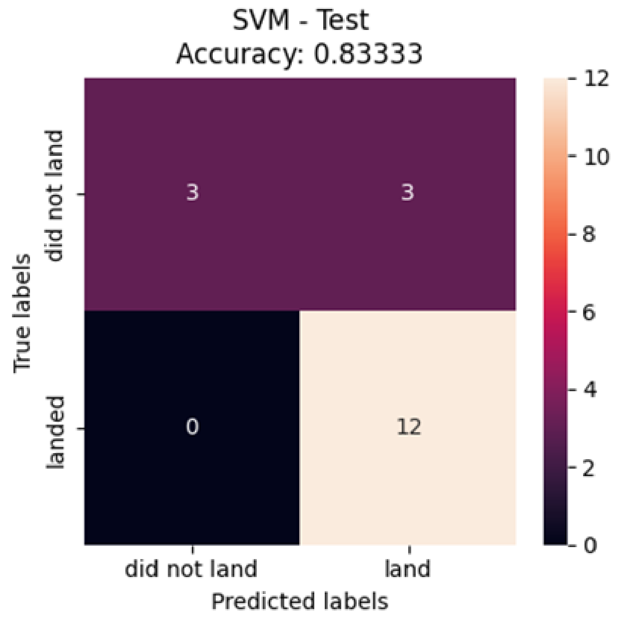


GitHub URL: [Dashboard with Plotly Dash]

Predictive Analysis (Classification)

In building the classification model, the process involved:

- **Data Loading and Preprocessing**
  - Loaded SpaceX dataset
  - Created a target array from the 'Class' column, and standardized features
- **Model Selection and Hyperparameter Tuning**
  - Utilized Logistic Regression, SVM, Decision Trees, and KNN models
  - Performed hyperparameter tuning with GridSearchCV
- **Model Training and Evaluation**
  - Split data into training and testing sets
  - Trained models on training data
  - Evaluated using accuracy and confusion matrices
- **Best Model Identification**
  - Compared model performances on test data to identify the best-performing model
  - Considering accuracy and confusion matrix results.



GitHub URL: [Machine Learning Prediction]

Results

- **Exploratory data analysis results**

The exploratory data analysis (EDA) for the SpaceX Falcon 9 first stage landing prediction involved examining various factors that could influence landing success. Key elements analyzed included payload mass, launch site, booster version, orbit type, and flight number. The analysis aimed to uncover trends and patterns, such as correlations between payload mass and landing outcomes, the impact of different launch sites on success rates, and how specific booster versions or orbit types might affect the likelihood of successful landings. This EDA provided crucial insights for building more accurate predictive models.

In the Exploratory Data Analysis for SpaceX Falcon 9 landings:

- **Payload Mass**. Trends suggest lighter payloads generally have better success rates than heavier ones.
- **Booster Version Performance.** Different booster versions have varying impacts on landing outcomes.
- **Orbit Type Correlation.** Specific orbits like GEO, HEO, SSO, and ES L1 are linked to higher success rates
- **Improvement Over Time.** Increasing success rate over time, indicating advancements and experience gain.
- **Launch Site Effectiveness.** KSC LC 39A shows the highest success rates among launch sites.

Results

- Predictive analysis results

The predictive analysis for the SpaceX Falcon 9 first stage landing prediction involved using machine learning models such as Logistic Regression, Support Vector Machine (SVM), Decision Trees, and K-Nearest Neighbors (KNN). Each model was trained on standardized features and hyperparameters were optimized using GridSearchCV. The models' performances were evaluated based on accuracy and confusion matrices. Decision Tree was identified as the best-performing model, demonstrating slightly better accuracy and generalization capabilities on test data compared to the other methods.

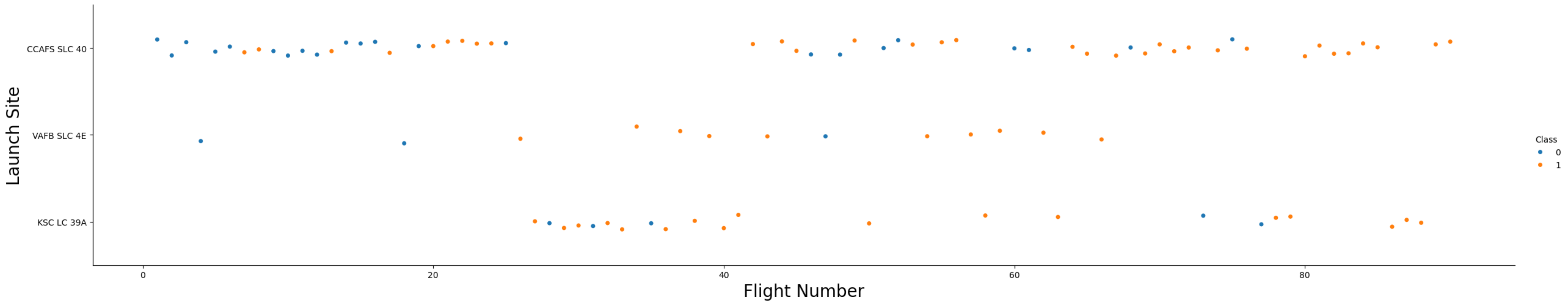
In the predictive analysis for SpaceX Falcon 9 landings:

- **Logistic Regression.** Evaluated with various regularization parameters.
- **SVM.** Tested with different kernels like linear, rbf, poly, sigmoid, and varying 'C' and 'gamma' values.
- **Decision Trees.** Assessed using various depths, criteria (gini, entropy), and other tree-specific params.
- **K-Nearest Neighbors.** Examined with different numbers of neighbors and distance metrics.
- **The Decision Tree** model, after hyperparameter tuning, outperformed others in accuracy on test data, indicating superior prediction capability for this specific task.

Section 2

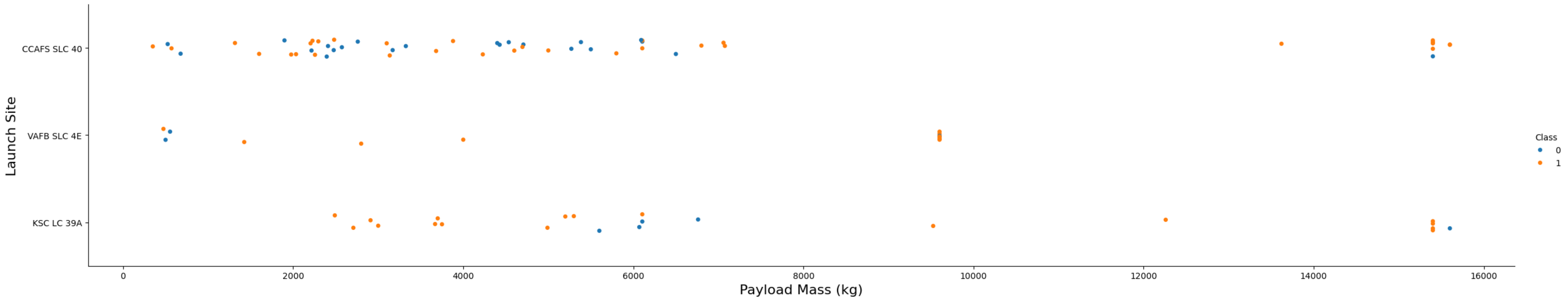
Insights drawn from EDA

Flight Number vs. Launch Site



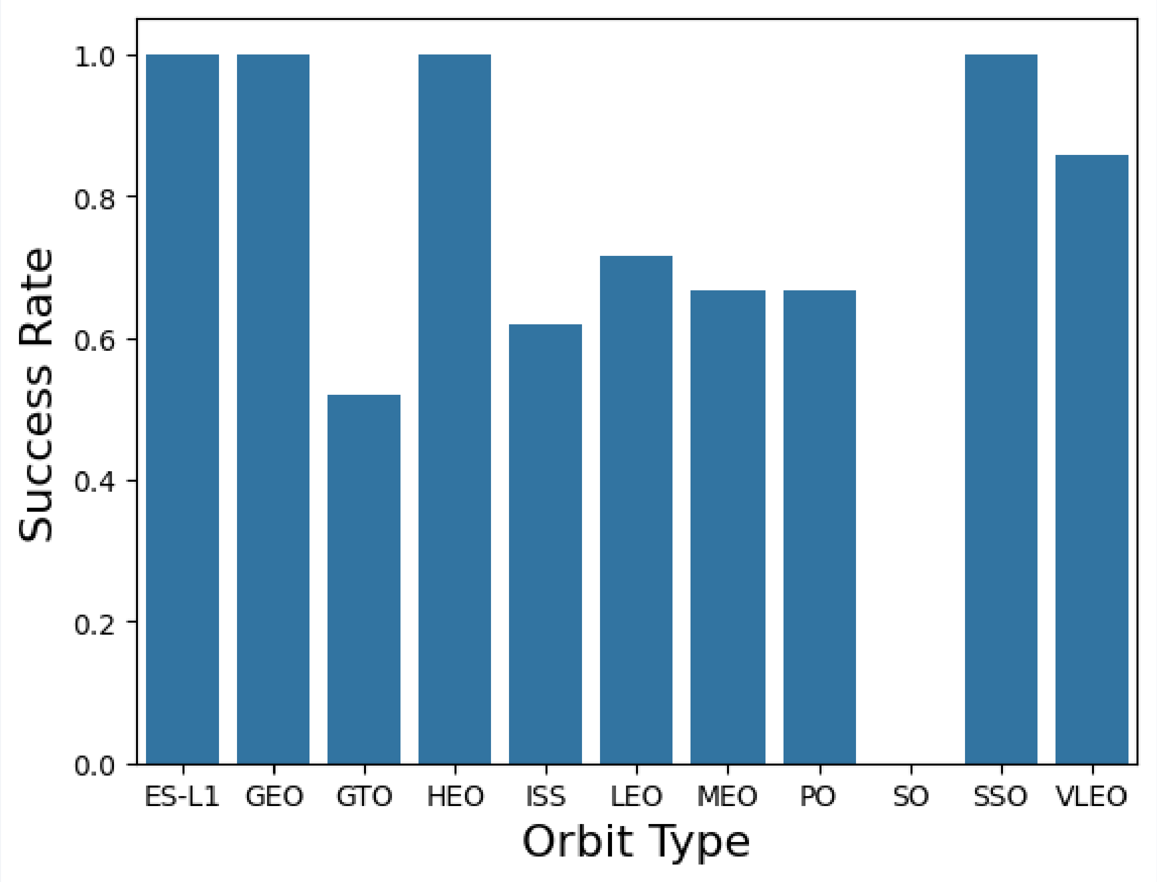
- **Launch Frequency.** The density of data points at the CCAFS SLC 40 (Cape Canaveral Air Force Station) level is visibly higher than at the other two sites. This suggests that the Cape Canaveral Air Force Station's Space Launch Complex 40 has been used for a greater number of launches compared to "VAFB SLC 4E" and "KSC LC 39A".

Payload vs. Launch Site



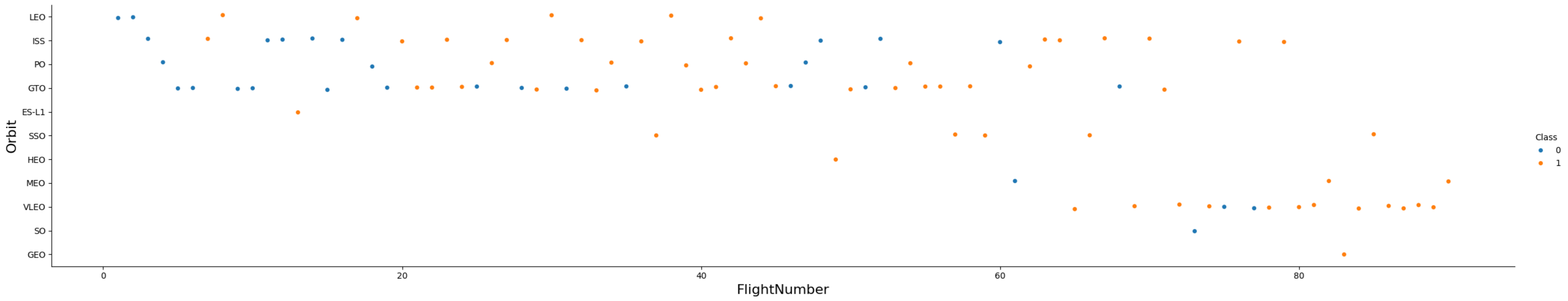
- **Launch Sites:**
  - **CCAFS SLC 40.** Wide payload mass range, both success and failure, used for diverse missions.
  - **VAFB SLC 4E.** Fewer launches, varied payload masses, mixed success outcomes.
  - **KSC LC 39A.** Supports wide payload mass range, shows both success and failure, like CCAFS SLC 40.
- **Payload Mass vs. Success:**
  - No clear relationship between payload mass and landing success; successful landings across all masses.
- **Heavier Payloads:**
  - Fewer data points for very heavy payloads (>10,000 kg), suggesting fewer common launches; successful landings at high payload masses.
- **Site-Specific Trends:**
  - All launch sites have both successful and unsuccessful launches across various payload masses
  - No exclusive mass range for successful landings at any site.

Success Rate vs. Orbit Type



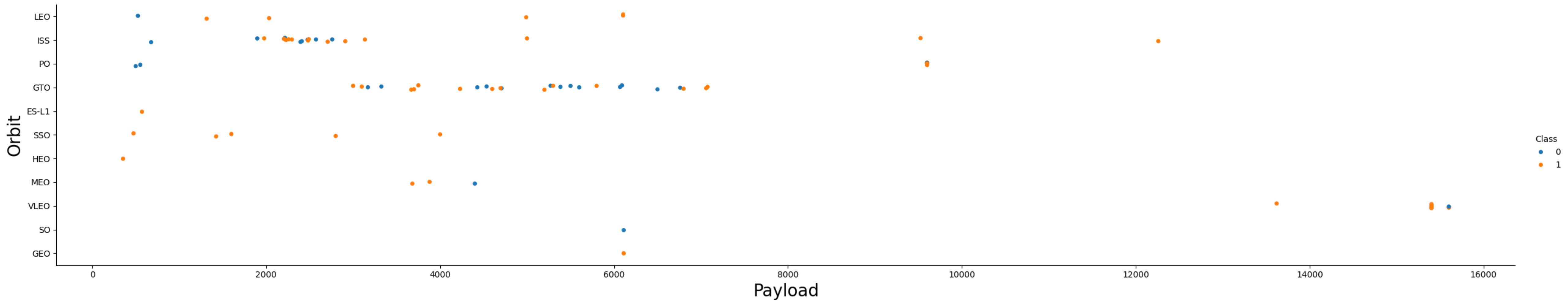
- **High Success Orbits (100% and Above 70%):** GEO, HEO, ES-L1, and combined SSO/SO have high success rates, with GEO, HEO, ES-L1 having a perfect record from a single attempt each, and SSO/SO showing a strong success rate of 83.33% over 6 attempts.
- **Moderately Successful Orbits (60-70%)** ISS has a moderate success rate of 61.90% with 21 attempts, indicating a somewhat challenging recovery process. MEO and PO are also moderately successful, with success rates of 66.67% across 3 and 9 attempts, respectively.
- **Varied Success Orbits (Below 60% to Above 70%):** VLEO exhibits a high success rate of 85.71% from 14 attempts, while LEO has a success rate of 71.43% with 7 attempts. GTO, with the highest number of attempts (27), shows a lower success rate of 51.85%, indicating it as the most challenging orbit for recoveries.

Flight Number vs. Orbit Type



- **High Success Orbits**
  Certain orbits like ES-L1, GEO, HEO, and SSO have shown a high success rate in the data provided, suggesting these orbits may be more conducive to successful landings, possibly due to the trajectory and energy requirements that allow for more controlled landings.
- **Moderately Successful Orbits**
  Orbits such as ISS, MEO, and PO have moderate success rates, indicating a fair reliability in landings but also room for improvement. These orbits may present specific challenges, such as varying re-entry dynamics or operational constraints that could affect landing outcomes.
- **Varied Success in GTO**
  - The GTO orbit has a notably lower success rate, which may highlight the technical challenges associated with the high velocity and energy requirements for reaching geostationary transfer orbits and the subsequent return trip for the first stage.
- **Trend Toward VLEO**
  The scatter plot suggests a trend of increasing launches to Very Low Earth Orbit (VLEO) in more recent flights, which also show a high success rate for landings. This could indicate a strategic move by SpaceX to focus on VLEO missions, potentially due to market demand for such orbits or because of the favorable conditions for first-stage recovery.

Payload vs. Orbit Type

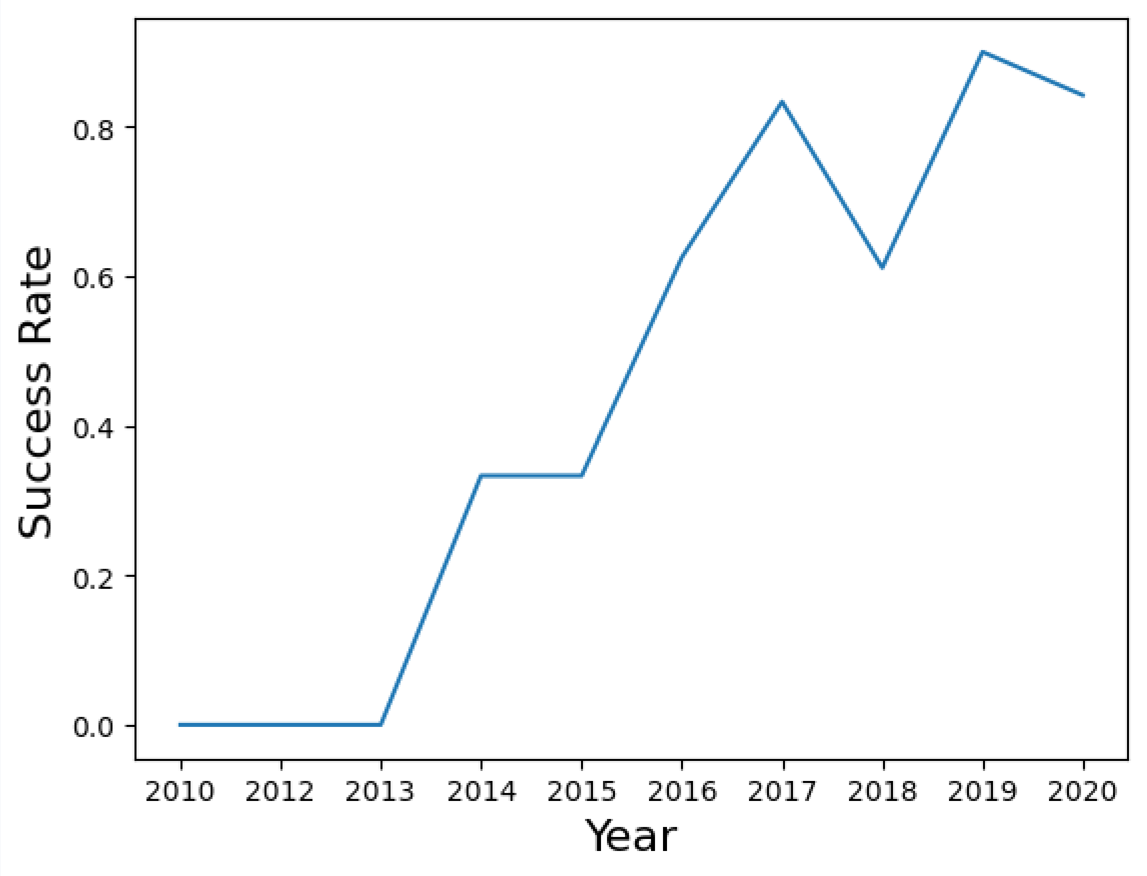


- **Payload Mass Impact**
  The plot suggests that successful landings occur across a wide range of payload masses. However, there is a noticeable presence of unsuccessful landings at higher payload masses, particularly in orbits like GTO, which might indicate that heavier payloads could affect the success of landings.

- **Orbit-Specific Patterns**
  Certain orbits like GTO and LEO have a mix of successful and unsuccessful landings across various payload masses. In contrast, some orbits such as ES-L1, SSO, and HEO show only successful landings within the displayed payload range.

- **Trends in Payload and Orbit**
  While the plot shows successful landings at the extremes of the payload mass range, most of the unsuccessful landings seem to cluster in the mid-range of the payload mass, particularly for GTO. This could suggest that there is a payload mass range that is more challenging for landing success.

Launch Success Yearly Trend



Overall, the trend from 2010 to 2020 is one of significant improvement, with initial successes starting in 2014 and a rapid rise to high levels of success by 2020:

- **Initial Phase (2010-2013).** No successful outcomes were recorded in this period.
- **Initial Success (2014-2015).** The success rate rose to 33.33%, marking the start of successful outcomes.
- Significant Improvement (2016): Success rate improved markedly to 62.5%.
- **Peak Performance (2017).** The success rate peaked at 83%.
- **Slight Decline (2018).** A minor decrease in the success rate to 61.11%, indicating some setbacks.
- **Strong Recovery (2019).** A high success rate of 90%, showing a strong rebound.
- **Stabilization (2020).** The success rate stabilized at a high level of 84.21%, suggesting sustained performance after the previous year's peak.

All Launch Site Names

```
%sql SELECT DISTINCT Launch_Site FROM SPACEXTBL
```

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

Launch Site Names Begin with 'CCA'

```
%sql SELECT * FROM SPACEXTBL WHERE Launch_Site LIKE 'CCA%' LIMIT 5
```

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

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) AS NASA_CRS_PAYLOAD_MASS_KG \
    FROM SPACEXTBL \
    WHERE Customer = 'NASA (CRS)'
```

NASA_CRS_PAYLOAD_MASS_KG
45596

Average Payload Mass by F9 v1.1

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) AS F9_1_1_PAYLOAD_MASS_KG \
    FROM SPACEXTBL \
    WHERE Booster_Version LIKE 'F9 v1.1%'
```

F9_1_1_PAYLOAD_MASS_KG
2534.6666666666665

First Successful Ground Landing Date

```
%sql SELECT MIN(Date) AS MIN_DATE \
    FROM SPACEXTBL WHERE Landing_Outcome = 'Success (ground pad)'
```

MIN_DATE
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql SELECT Booster_Version \
    FROM SPACEXTBL \
    WHERE Landing_Outcome = 'Success (drone ship)' \
          AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000
```

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 SELECT Mission_Outcome, COUNT(*) AS Total_Number \
    FROM SPACEXTBL \
    GROUP BY Mission_Outcome
```

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

Boosters Carried Maximum Payload

```
%sql SELECT Booster_Version \
    FROM SPACEXTBL \
    WHERE PAYLOAD_MASS__KG_ = \
            (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL)
```

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

2015 Launch Records

List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

(List the records which will display the month names, failure landing_outcomes in drone ship, booster versions, launch_site for the months in year 2015.)

```
%sql \
SELECT (CASE strftime('%m', Date) \
        WHEN '01' THEN 'January' WHEN '02' THEN 'February' WHEN '03' THEN 'March' \
        WHEN '04' THEN 'April' WHEN '05' THEN 'May' WHEN '06' THEN 'June' \
        WHEN '07' THEN 'July' WHEN '08' THEN 'August' WHEN '09' THEN 'September' \
        WHEN '10' THEN 'October' WHEN '11' THEN 'November' WHEN '12' THEN 'December' END) as Month_Name, \
        Landing_Outcome, Booster_Version, Launch_Site \
FROM SPACEXTBL \
WHERE Landing_Outcome = 'Failure (drone ship)' AND substr(Date, 0, 5) = '2015'
```

Month_Name	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

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

```
%sql \
SELECT RANK () OVER(ORDER BY COUNT(*) DESC) AS Rank, \
       Landing_Outcome, COUNT(*) AS Count \
FROM SPACEXTBL \
WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' \
GROUP BY Landing_Outcome
```

Rank	Landing_Outcome	Count
1	No attempt	10
2	Success (drone ship)	5
2	Failure (drone ship)	5
4	Success (ground pad)	3
4	Controlled (ocean)	3
6	Uncontrolled (ocean)	2
6	Failure (parachute)	2
8	Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

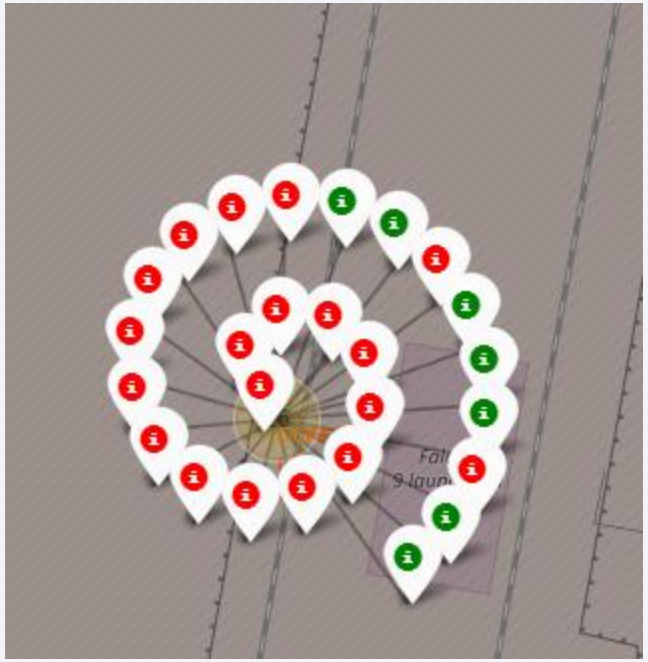
Map with marked launch sites

The map highlights two important launch sites:

- The **VAFB** SLC 4E in California is strategically chosen for polar orbit launches, emphasizing the safety considerations due to its coastal location.
- The **CCAFS** sites in Florida, including CCAFS LC-40, CCAFS SLC-40, and KSC LC-39A, are ideal for equatorial orbits. Their proximity to the equator allows for more efficient launches to geostationary and other equatorial orbits, taking advantage of the Earth's rotational velocity.

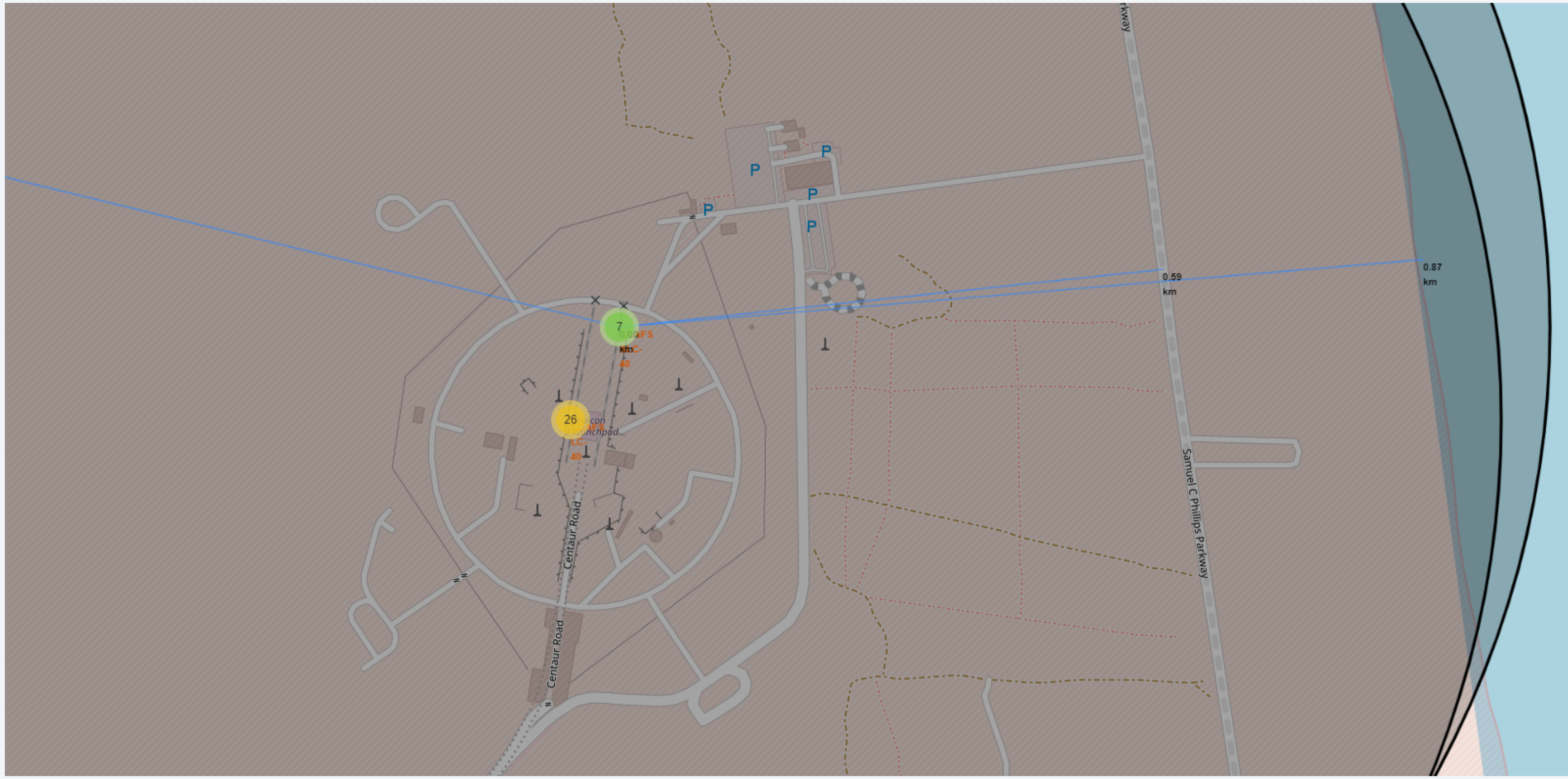
Launch Sites with Successful / Failed Indicators

- **Launch Site Labels.** Labels identify specific launch pads
- **Quantity Indicators.** Numerical labels on markers denote the number of launches



- **Color Codes.** Green markers indicate successful landings, while red markers show unsuccessful ones.
- **Launch Activity.** The density of markers suggests the frequency of launches at each site.
- **Success Rate Visualization.** The mix of green and red markers offers a quick visual reference to the success rate at the launch site.

Calculating Proximity Distances from a Launch Site



Nearby Features:

- The closest highway is approx. 0.59 km away from the launch site.
- The distance to the coastline is about 0.87 km, indicating the launch site's proximity to the ocean, which is typical for safety and trajectory planning in rocket launches.

The map provides crucial information for logistical planning and risk assessment, considering the proximity of infrastructure which can be vital for launch preparations and emergency response strategies.

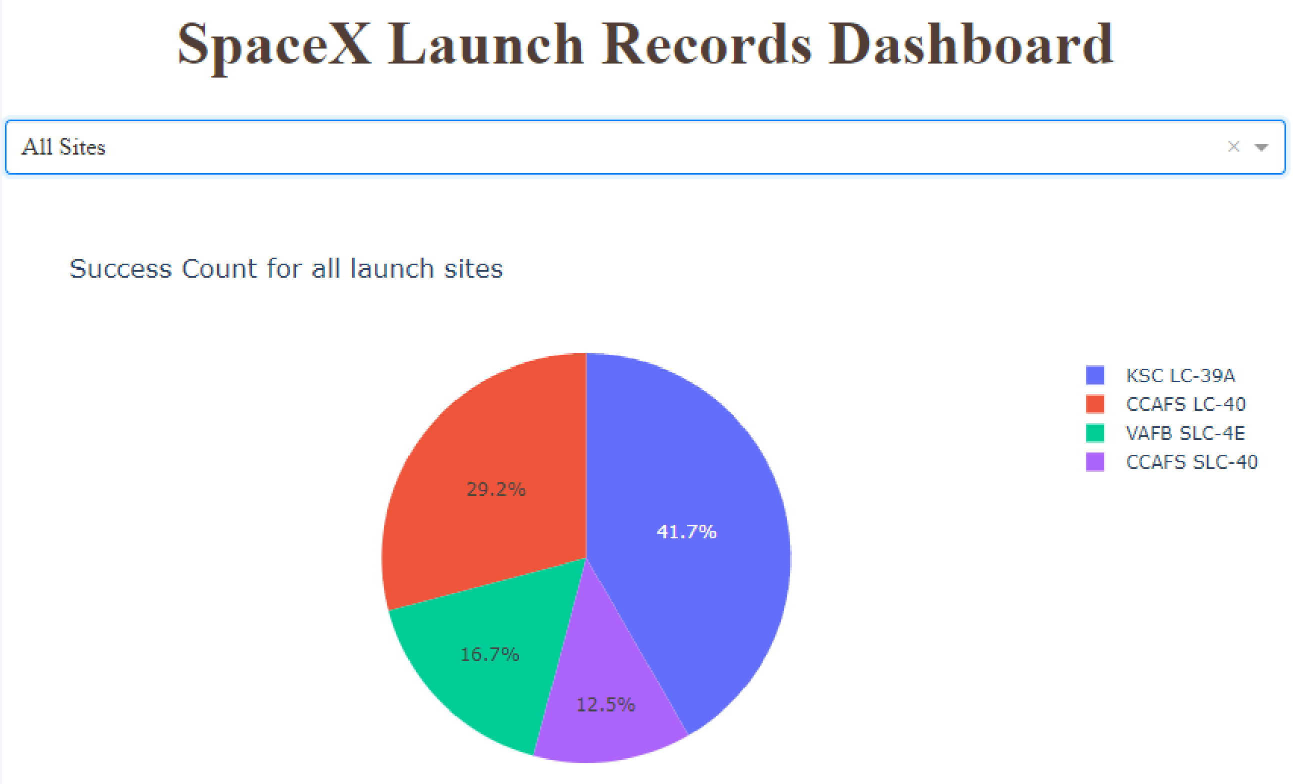
The screenshot displays a folium map with a focus on a particular launch site and its proximities:

- **Proximity Lines.** Blue lines extend from the launch site markers to various points, representing the calculated distances to nearby features such as highways and coastlines.
- **Distance Measurements.** Text labels on the lines (e.g., "0.59 km" and "0.87 km") show the measured distances from the launch site to those features.

Section 4

Build a Dashboard with Plotly Dash

Successful Launches by Site: A Pie Chart Analysis



Dashboard displaying a pie chart that shows the success count for all SpaceX launch sites. Here are the key elements:

- **Dashboard Functionality.** Dropdown menu "All Sites" - the dashboard allows users to filter the displayed data by individual launch sites or view aggregated data for all sites.
- **Data Interpretation.** The pie chart indicates that KSC LC-39A is the most successful launch site in terms of the number of successful launches, followed by CCAFS LC-40. VAFB SLC-4E and CCAFS SLC-40 have fewer successful launches comparatively.

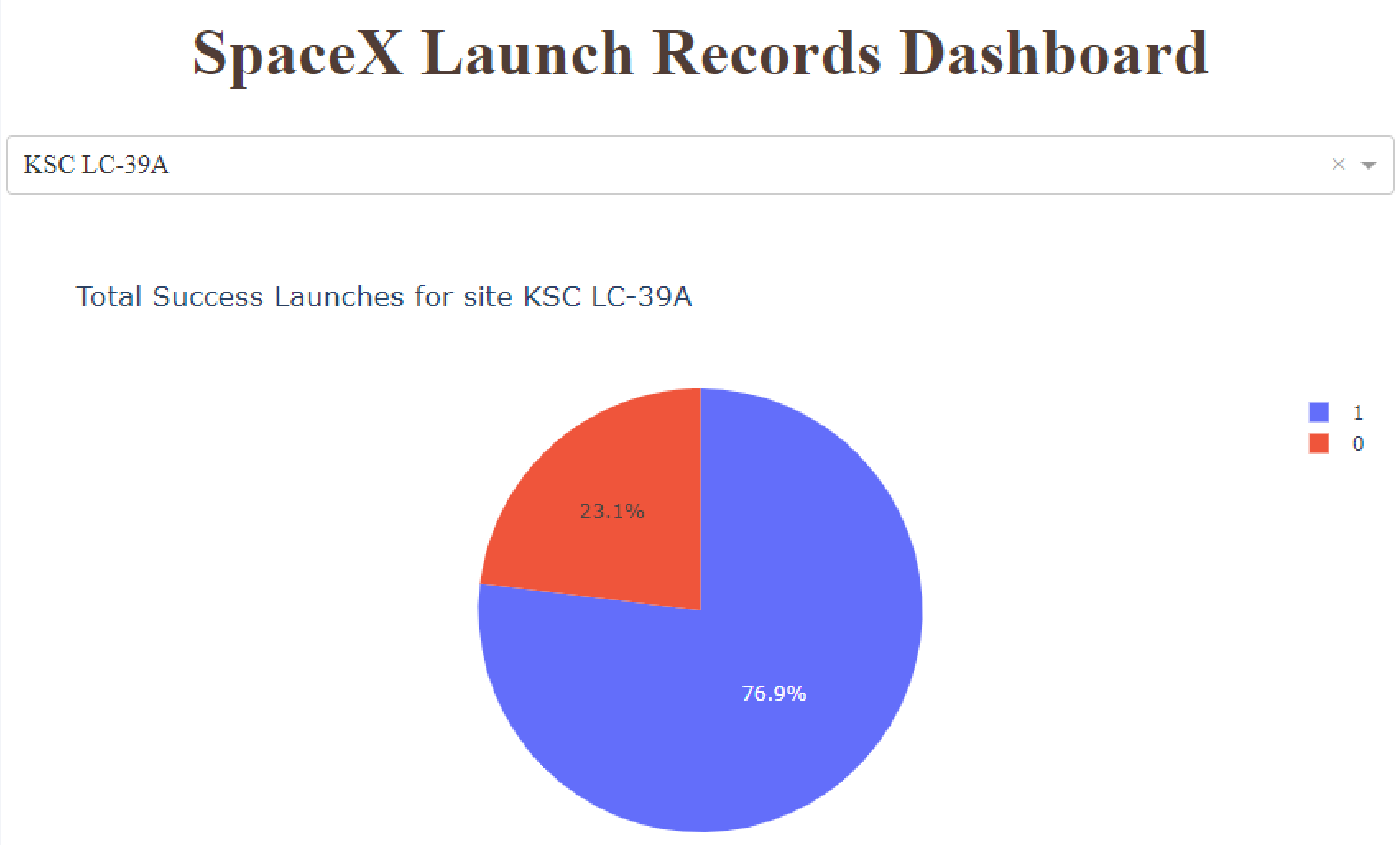
The pie chart effectively communicates the distribution of successful launches among the different SpaceX launch sites, which can be useful for analyzing site performance and planning future launch operations.

Pie Chart Segments. Each segment represents the proportion of successful launches from a specific SpaceX launch site:

CCAFS SLC-40 (Cape Canaveral Air Force Station Space Launch Complex 40): **purple**, making up 12.5% of successful launches.
CCAFS LC-40 (Cape Canaveral Air Force Station Launch Complex 40): Colored **red**, this represents 29.2% of the successes.

VAFB SLC-4E (Vandenberg Air Force Base Space Launch Complex 4E): Shown in **green**, accounts for 16.7% of the total successful launches.
KSC LC-39A (Kennedy Space Center Launch Complex 39A): The largest segment, colored **blue**, accounts for 41.7% of the successful launches.

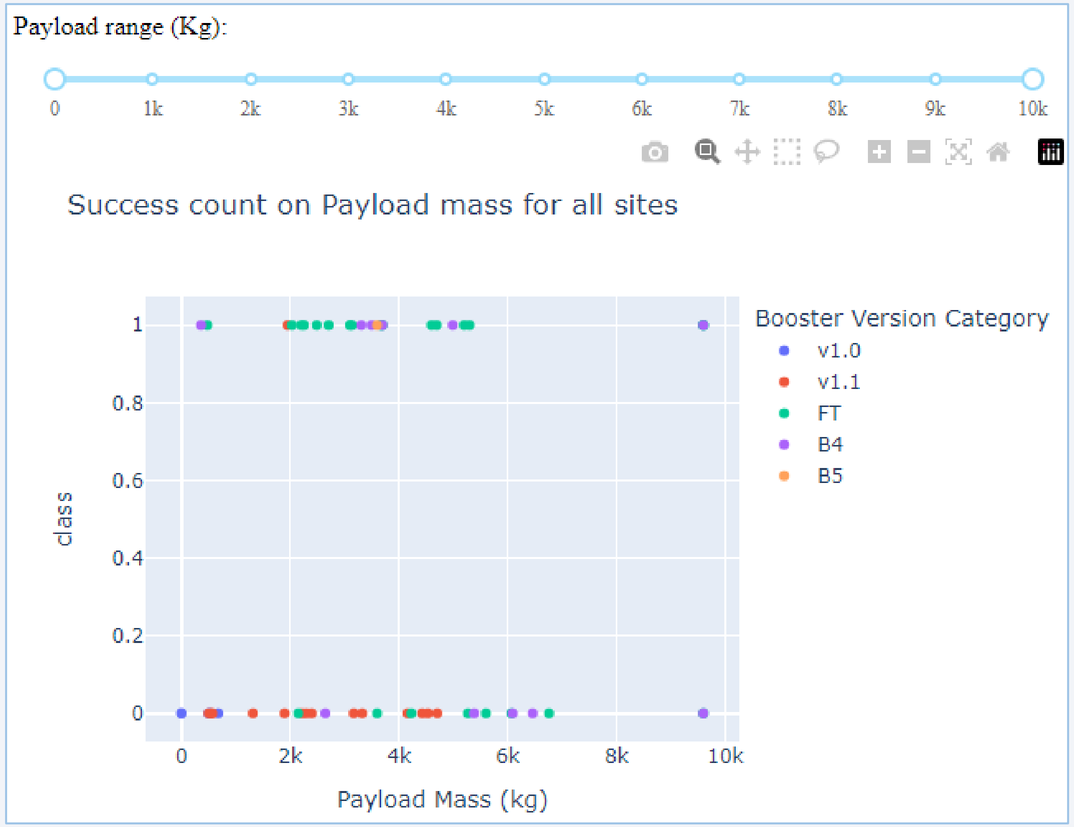
KSC LC-39A Launch Success Ratio

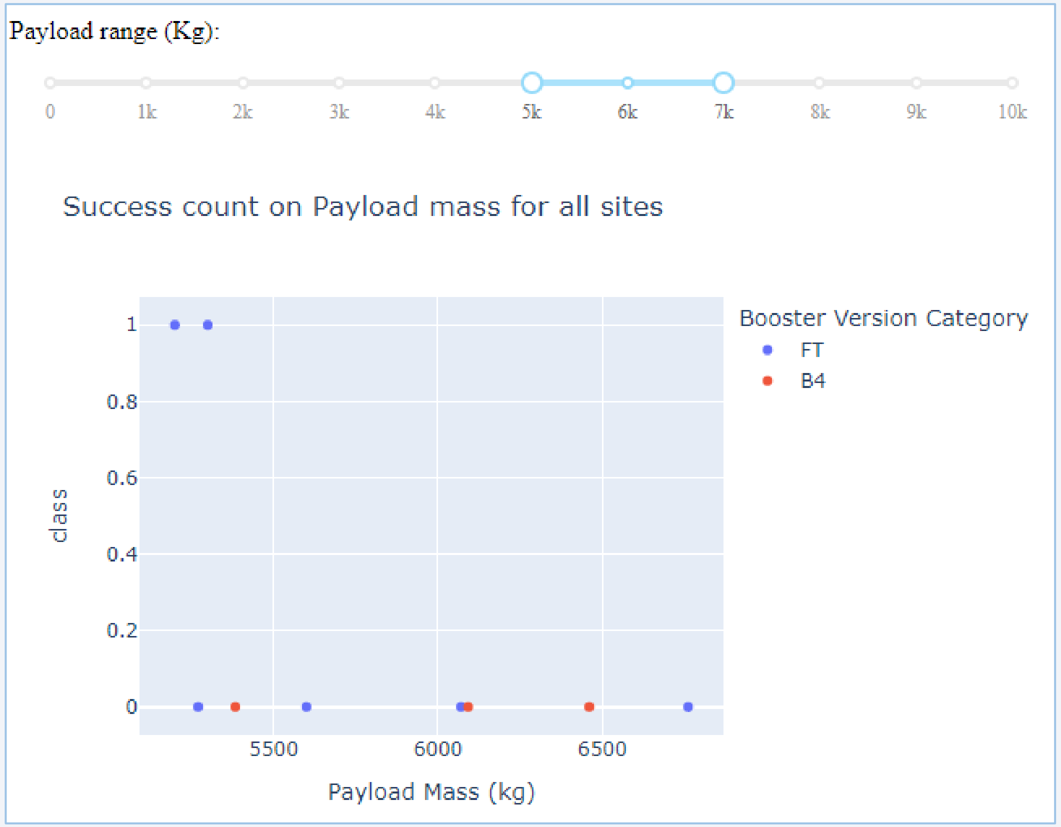


- **Pie Chart.** Class "1" (shown in blue) represents successful landings and class "0" (shown in red) represents unsuccessful ones.
- **Success Ratio.** The blue segment covers a majority of the chart, specifically 76.9%, indicating that a significant majority of landings started at KSC LC-39A have been successful.
- **Unsuccessful Launches.** The red segment, which represents unsuccessful launches, makes up 23.1% of the pie chart.
- **Findings.** KSC LC-39A has a high success rate with over three-quarters of landings being successful.

This pie chart is a visual tool used for quickly assessing the success rate of launches from KSC LC-39A and is helpful for decision-making or presenting to stakeholders interested in launch success metrics.

Payload vs. Launch Success: A Scatter Plot Analysis



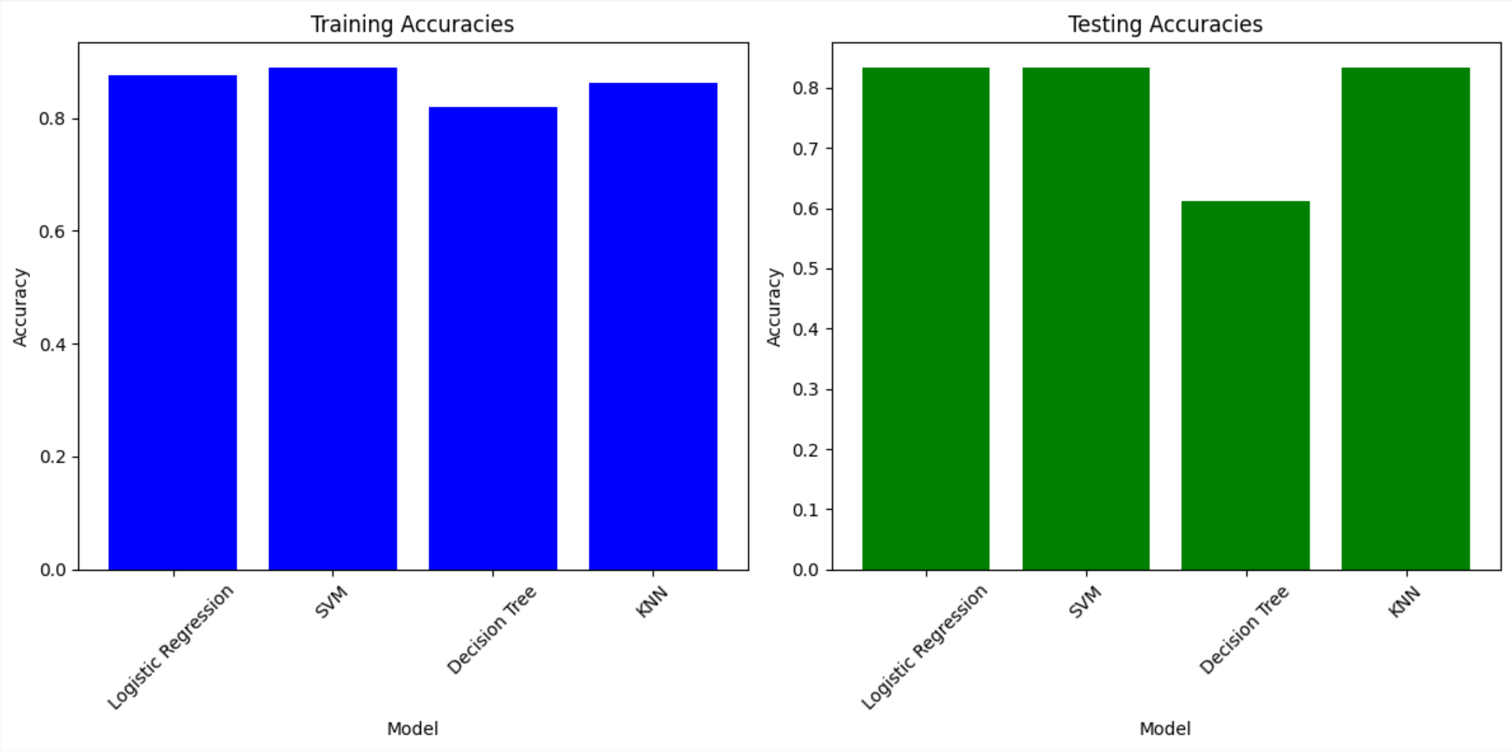


- **High Success Across Payloads.** SpaceX demonstrates a consistently high success rate across a broad range of payload masses, as indicated by most data points achieving a 'class' level of 1, which signifies successful launches.
- **Booster Version Success.** The Full Thrust (FT) version of the Falcon 9 booster exhibits a notably high success rate across the selected payload range. Some failures have been observed in payloads exceeding 5,000 kg, as noted in the second image.
- These observations imply that SpaceX's launch success is high across various payloads, with the FT booster version showing strength when carrying payloads below 5,000 kg.

Section 5

Predictive Analysis (Classification)

Classification Accuracy



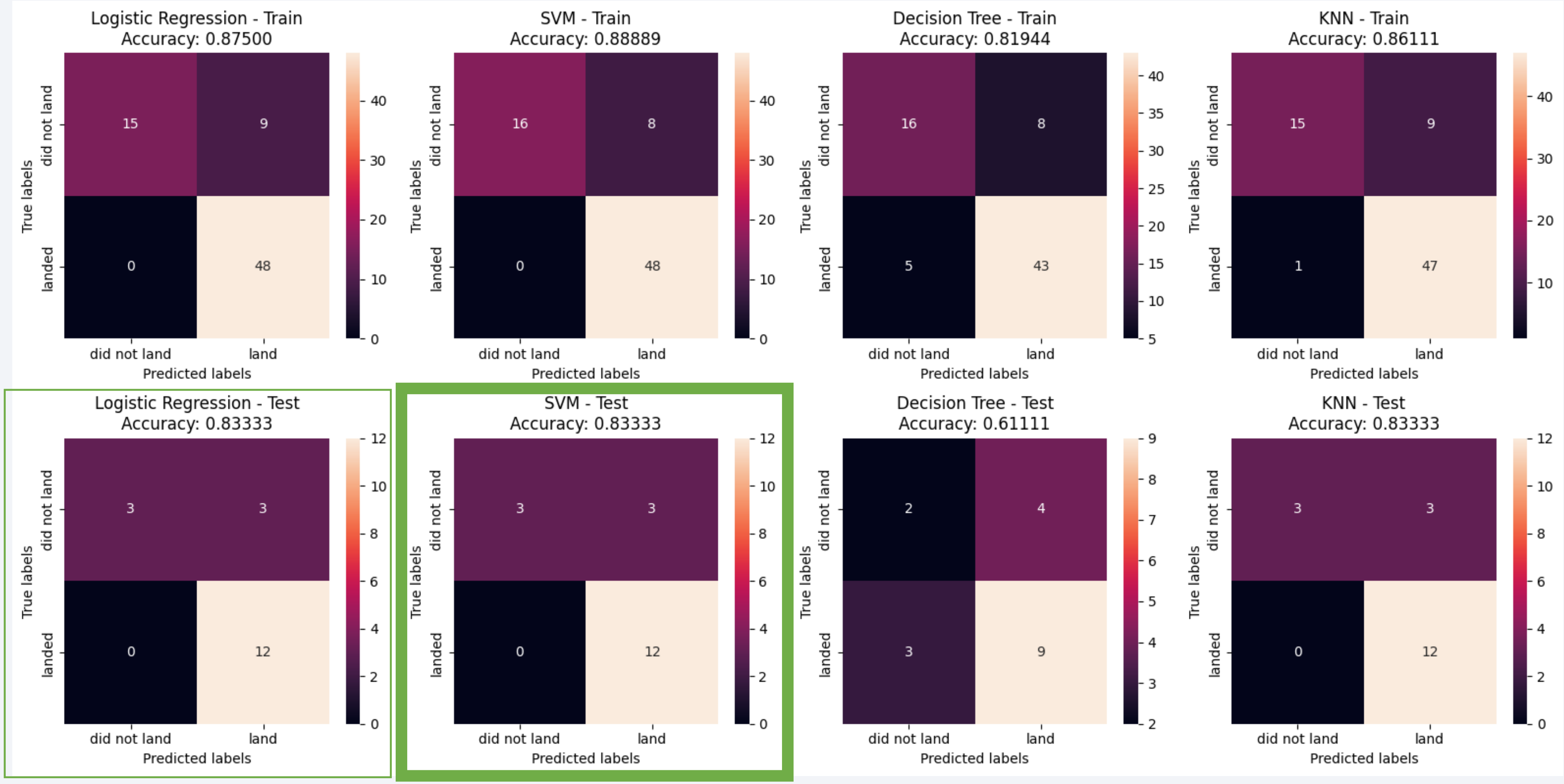
Training Metrics:

Model	Accuracy	Recall	Precision	ROC AUC
Logistic Regression	0.8750	1.0000	0.8421	0.9618
SVM	0.8889	1.0000	0.8571	0.9123
Decision Tree	0.8194	0.8958	0.8431	0.8867
KNN	0.8611	0.9792	0.8393	0.9023

Testing Metrics:

Model	Accuracy	Recall	Precision	ROC AUC
Logistic Regression	0.8333	1.0000	0.8000	0.8889
SVM	0.8333	1.0000	0.8000	0.9583
Decision Tree	0.6111	0.7500	0.6923	0.7500
KNN	0.8333	1.0000	0.8000	0.8958

Confusion Matrix



Predictive Analysis (Classification) - Recommendations

- Among the **Logistic Regression**, **SVM**, and **KNN** models, all showed equal **Accuracy** on the test data.
- **SVM** exhibits the best discriminative ability with the highest ROC AUC score, suggesting it might be the best at predicting landing success.
- **Logistic Regression** offers interpretability, which can be crucial for understanding and improving launch outcomes.
- **Recommendation**: While all three models share high accuracy, SVM is preferred for its overall performance, with Logistic Regression as a secondary option for its explanatory power.

Conclusions

- **Model Selection for Prediction.** The SVM (Support Vector Machine) model is the preferred choice for predicting the success of SpaceX launches due to its superior performance metrics. Logistic Regression is identified as a secondary option, valued for its ability to provide clear explanations for its predictions, which can be important for understanding the factors influencing launch outcomes.

- **Payload Weight Influence.** Lighter payloads have been associated with higher success rates compared to heavier payloads. This could be due to a variety of factors including the challenges of launching heavier payloads or the increased complexity of missions that typically carry more weight.

Conclusions

- **Improvement Over Time.** There's a positive correlation between time and the success rates of SpaceX launches. This suggests a learning curve effect, where SpaceX is continuously improving its launch processes, leading to an expectation that they will eventually perfect their launch methodology.

- **Launch Site Success.** Among the launch sites, KSC LC 39A (Kennedy Space Center Launch Complex 39A) has recorded the most successful launches, indicating it as a particularly reliable site for launch operations.

- **Orbit-Specific Success Rates.** The orbits GEO (Geostationary Earth Orbit), HEO (Highly Elliptical Orbit), SSO (Sun-Synchronous Orbit), and ES L1 (Earth-Sun Lagrange Point 1) are noted for having the best success rates. This could be reflective of SpaceX's expertise and experience in targeting these specific orbits or the inherent mission profiles that suit these orbits.

Appendix

- You can find all related Python code snippets, SQL queries, charts, Notebook outputs, and data sets for this project on GitHub at the following address:

https://github.com/6760525/ads

Thank you!