IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Dr. Addam Brown, PhD
December, 2024

Outline

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

Executive Summary

Summary of methodologies

Data Collection

Data was collected from the SpaceX API, web scraping, and CSV files containing historical SpaceX launch data.

Data Wrangling

Data was cleaned and transformed to ensure consistency and usability for analysis.

Exploratory Data Analysis (EDA)

Visualizations and SQL queries were performed to explore and understand the data.

Interactive Visual Analytics

Interactive maps and dashboards were created using Folium and Plotly Dash.

Predictive Analysis

Classification models were built, tuned, and evaluated to predict the success of future launches.

Summary of all results

Launch Outcomes

KSC LC-39A has the highest success rate and the largest number of successful launches.

Payload Analysis

Payload range of 2880.0 to 3840.0 kg has the highest success rate.

Payload range of 960.0 to 1920.0 kg has the lowest success rate.

Booster Versions

Booster version B5 has the highest launch success rate.

Predictive Models

Classification models showed high accuracy in predicting launch success outcomes, with one model standing out as the most accurate.

Introduction

Project background and context

SpaceX has launched numerous missions, and analyzing these launches can provide insights into factors contributing to successful missions.

This project aims to explore SpaceX launch data, visualize key metrics, and build predictive models to forecast future launch success.

Problems needing answers

Which launch sites have the highest success rates?

What payload ranges are most likely to result in successful launches?

Which booster versions are most reliable?

Can we predict the success of future launches based on historical data?

Section 1

Methodology

Methodology

Executive Summary

Data collection methodology

Data was collected using various methods including API calls to the SpaceX API, web scraping for additional data, and reading CSV files containing historical SpaceX launch data.

Data wrangling

Data was processed to clean and transform it into a usable format. This involved handling missing values, correcting data types, and combining different data sources.

EDA using visualization and SQL

EDA was conducted using visualizations and SQL queries to understand the distribution and relationships in the data.

Interactive visual analytics using Folium and Plotly Dash:

Interactive maps and dashboards were created using Folium and Plotly Dash to visualize launch site locations and success rates.

Performed predictive analysis using classification models:

Classification models were built, tuned, and evaluated to predict the success of future SpaceX launches.

Data Collection

Describe how data sets were collected:

Data was collected from multiple sources including the SpaceX API, web scraping, and CSV files.

Key phrases and flowcharts:

Key phrases: API calls, web scraping, CSV files, data cleaning, data transformation.

Flowchart: Data Source → API Calls / Web Scraping / CSV → Data Cleaning → Data Transformation → Usable Data.

Data Collection – SpaceX API

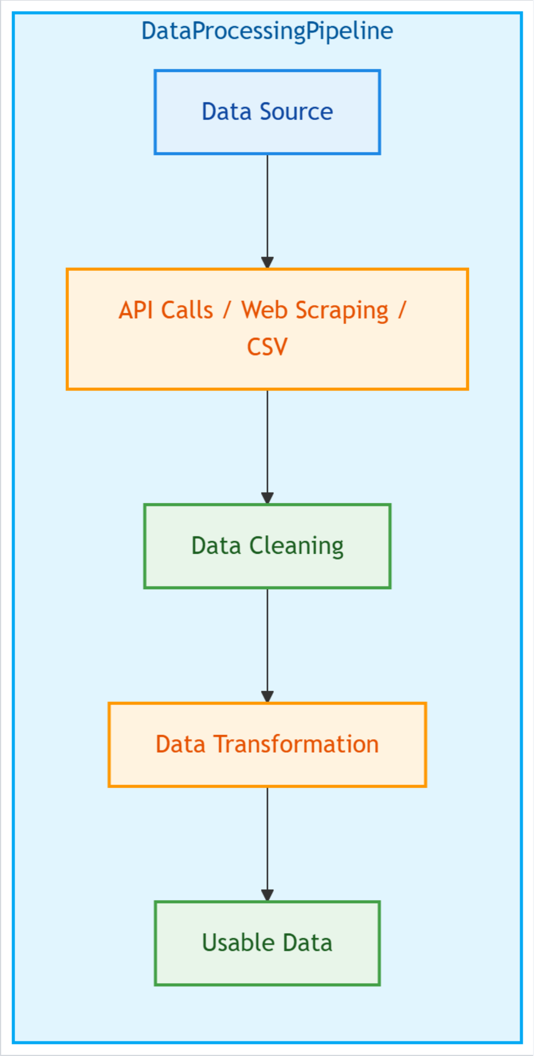
Data collection with SpaceX REST calls

Key phrases: SpaceX API, REST calls, JSON response, data parsing.

Flowchart: SpaceX API → REST Call → JSON Response → Data Parsing → Usable Data.

GitHub URL for the SpaceX API calls Jupyter Notebook

GitHub URL of SpaceX API Calls Notebook



Data Collection - Scraping

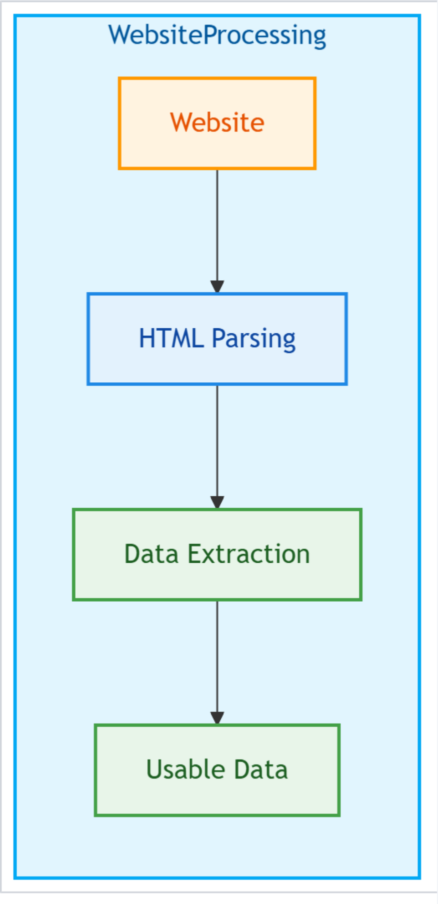
Web scraping process key phrases

Key phrases: web scraping, BeautifulSoup, HTML parsing, data extraction.

Flowchart: Website → HTML Parsing → Data Extraction → Usable Data.

GitHub URL web scraping notebook

GitHub URL of Web Scraping Notebook



Data Wrangling

Data Processing

Data was cleaned to remove null values and correct data types. It was transformed to create new features and combined with other datasets for a comprehensive analysis.

Data wrangling process

Key phrases: data cleaning, data transformation, feature engineering.

Flowchart: Raw Data → Data Cleaning → Data Transformation → Feature Engineering → Processed Data.

GitHub URL data wrangling notebooks

GitHub URL of Data Wrangling Notebook

EDA with Data Visualization

Charts and reasons

Scatter plots, bar charts, line charts, and pie charts were used to visualize the relationship between different features, distribution of data, and trends over time.

GitHub URL EDA with data visualization

GitHub URL of EDA with Data Visualizations Notebook

EDA with SQL

SQL queries performed

Queries were performed to count launches per site, calculate success rates, find payload ranges, and extract specific data points related to launch outcomes.

GitHub URL EDA with SQL notebook

GitHub URL of EDA with SQL Notebook

Build an Interactive Map with Folium

Map objects created and added to folium map

Markers for launch sites, circles for impact zones, and lines showing proximity to facilities.

Reasons for objects

These objects help visualize the geographical distribution of launch sites and their proximity to key facilities for a comprehensive understanding of launch logistics and safety.

GitHub URL interactive map with objects

GitHub URL of Folium Map Notebook

Build a Dashboard with Plotly Dash

Plots/graphs and interactions on the dashboard

Pie charts, scatter plots, line charts, and interactive sliders to filter data based on user input.

Reason for plots and interactions:

These visualizations and interactions allow users to explore the data dynamically, uncover patterns, and draw insights on launch success factors.

GitHub URL Plotly Dash notebook:

GitHub URL of Plotly Dash Notebook

Predictive Analysis (Classification)

Predictive classification model build, evaluation, and fine tuning

Built multiple classification models, tuned hyperparameters, evaluated using cross-validation, and selected the best performing model based on accuracy.

Model development process with key phrases and flowchart

Key phrases: model building, hyperparameter tuning, cross-validation, model evaluation.

Flowchart: Data Preparation → Model Building → Hyperparameter Tuning → Cross-Validation → Model Evaluation → Best Model Selection.

GitHub URL of predictive analysis notebook:

GitHub URL of Predictive Analysis Notebook

Results

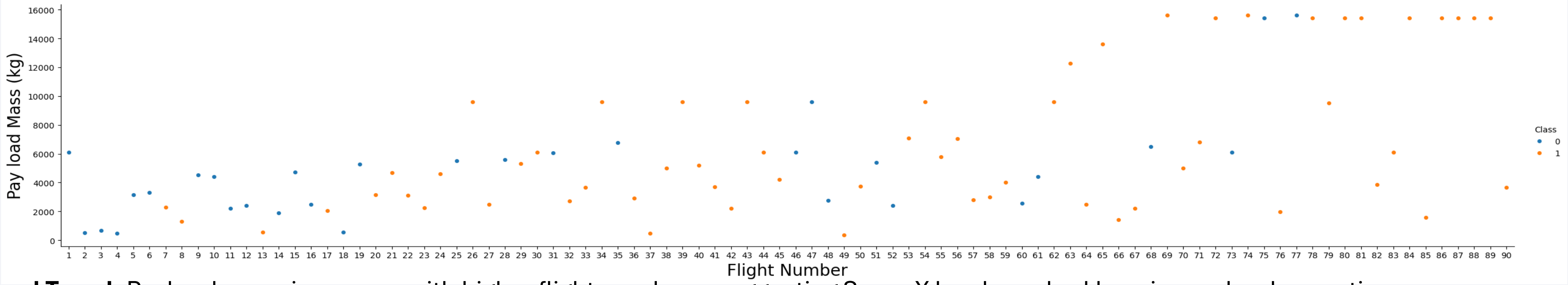
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

Payload vs. Flight Number Overall Trend

The plot differentiates between two classes: Class 0 (blue points) and Class 1 (orange points). This can represent different launch outcomes, such as failures and successes.



General Trend: Payload mass increases with higher flight numbers, suggesting SpaceX has launched heavier payloads over time.

Class Distribution: More successful launches occur with higher payload masses, indicating higher success rates for heavier flights.

Payload Range and Success: Successful flights are mostly in the 4,000-10,000 kg range, suggesting this is optimal. Lighter payloads show mixed outcomes.

Flight Number and Learning Curve: Higher flight numbers correlate with more successful launches, suggesting improved success rates over time.

Outliers:

- A few high payloads (above 10,000 kg) are successful, indicating significant launches.
- Some failed lower payload flights warrant further analysis to understand failure factors.

Summary: The scatter plot shows a positive trend in success rates over time and identifies an optimal payload range for successful launches. Further analysis can provide insights into factors influencing launch outcomes.

Flight Number vs. Launch Site

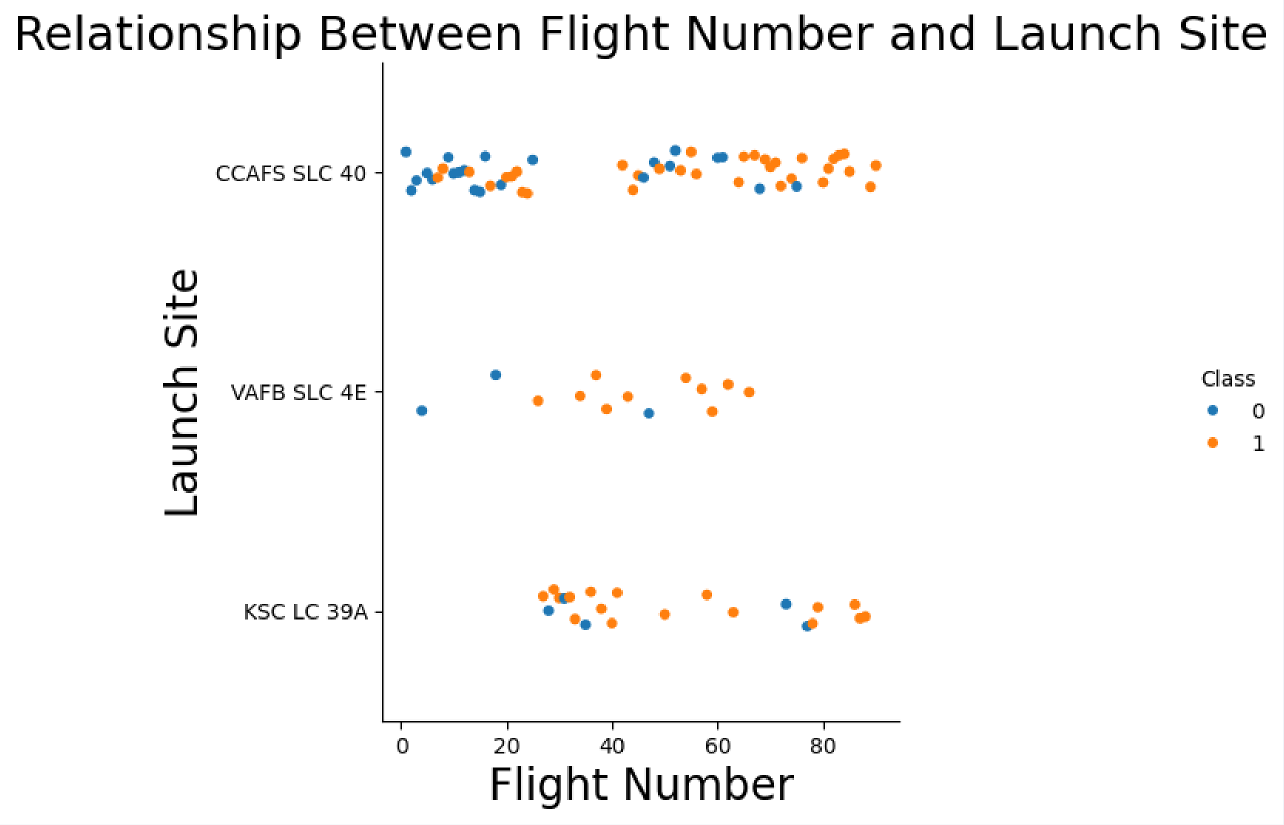
General Trend: The distribution of flight numbers is spread across three main launch sites: CCAFS SLC 40, VAFB SLC 4E, and KSC LC 39A.

Class Distribution: KSC LC 39A appears to have a higher concentration of successful launches (Class 1).

Launch Site Usage: KSC LC 39A shows a consistent pattern of usage across a wide range of flight numbers while VAFB SLC 4E has fewer flights compared to the other two sites, indicating less frequent use.

Success Rates: There is a noticeable concentration of successful launches (Class 1, orange points) at KSC LC 39A. CCAFS SLC 40 has a mix of successes and failures, suggesting variability in launch outcomes at this site.

Summary: This scatter plot provides insights into the relationship between flight numbers and launch sites, with a notable success rate at KSC LC 39A. Further analysis could explore the specific factors contributing to the success at KSC LC 39A and the variability at CCAFS SLC 40.



Payload vs. Launch Site

General Trend: The payload masses are distributed across three main launch sites: CCAFS SLC 40, VAFB SLC 4E, and KSC LC 39A, with varying payload masses.

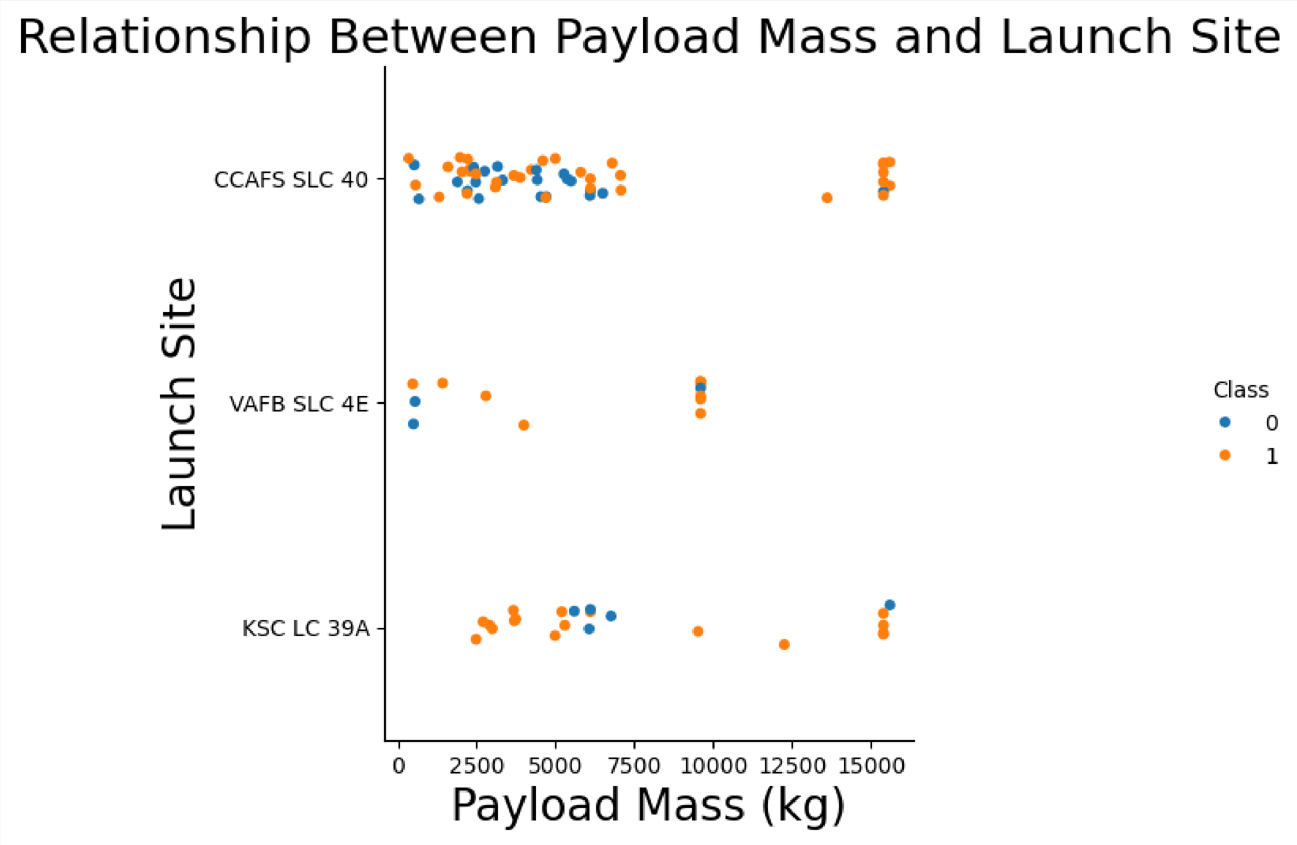
Class Distribution: There appears to be a higher concentration of successful launches (Class 1) at KSC LC 39A and CCAFS SLC 40.

Payload Range and Success: Both successful and failed launches are observed across this range, with no specific payload range showing a markedly higher success rate.

Launch Site Usage: KSC LC 39A and CCAFS SLC 40 show a consistent pattern of usage across a wide range of payload masses while VAFB SLC 4E has fewer data points, indicating less frequent use or specific mission types with varied payloads.

Success Rates: There is a noticeable concentration of successful launches (Class 1, orange points) at KSC LC 39A and CCAFS SLC 40 which suggests that these two sites have a relatively higher success rate compared to VAFB SLC 4E.

Summary: This scatter plot provides insights into the relationship between payload mass and launch sites, highlighting the success rates and usage patterns of different sites. KSC LC 39A and CCAFS SLC 40 appear to be the most frequently used and successful launch sites for a wide range of payload masses, while VAFB SLC 4E shows less consistent use.



Success Rate vs. Orbit Type

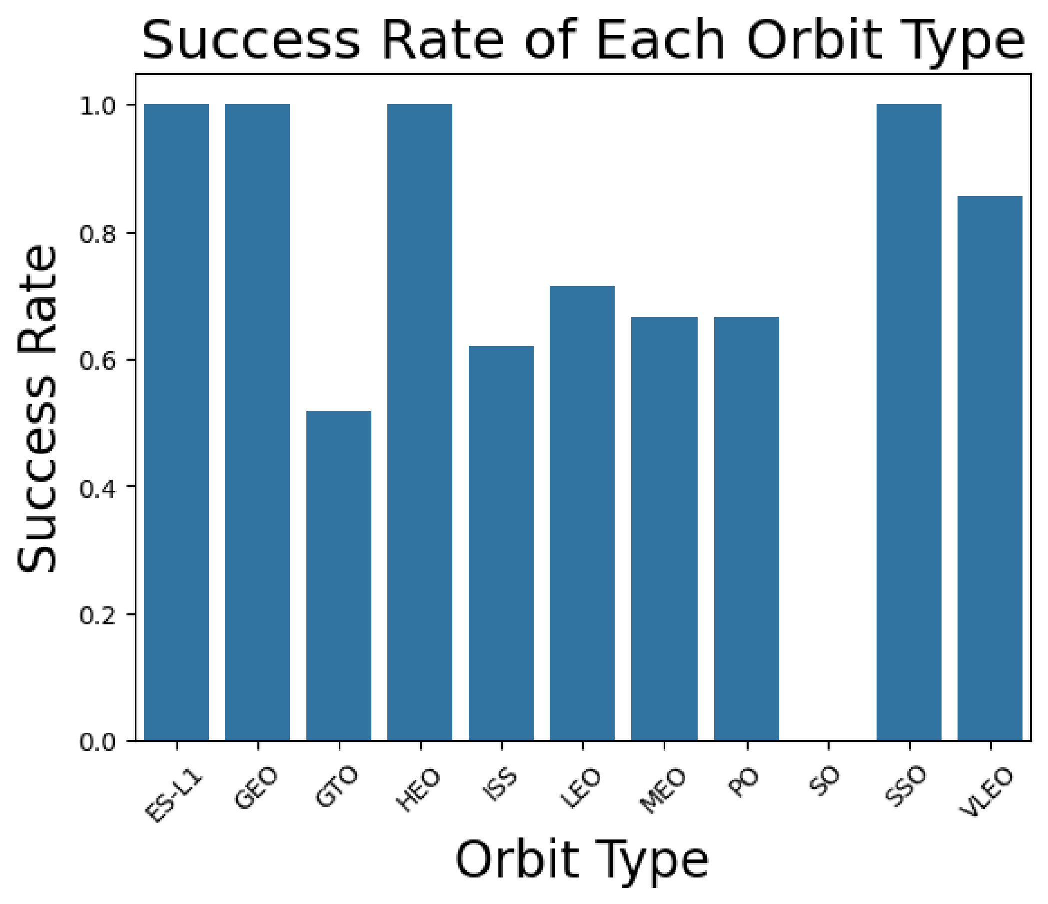
Highest Success Rates: The orbit types **ES-L1**, **GEO**, and **SSO** have a 100% success rate, indicating that all missions targeting these orbits have been successful.

High Success Rates: HEO (Highly Elliptical Orbit) has a very high success rate of approximately 95% while **VLEO** (Very Low Earth Orbit) and **SO** (Solar Orbit) also have high success rates of approximately 85% and 90%, respectively.

Moderate Success Rates: LEO (Low Earth Orbit), **MEO** (Medium Earth Orbit), and **PO** (Polar Orbit) have moderate success rates of around 75%.

Lower Success Rates: ISS (International Space Station) orbit has a lower success rate of approximately 65% but **GTO** (Geostationary Transfer Orbit) has the lowest success rate at around 50%, suggesting challenges or higher risks associated with missions to this orbit.

Summary: The chart clearly shows that ES-L1, GEO, and SSO orbits have the highest reliability, with all missions to these orbits being successful. GTO and ISS orbits present more challenges, with GTO having the lowest success rate. Understanding these success rates can help in planning future missions and identifying areas for improvement.



Flight Number vs. Orbit Type

General Trend: The plot shows a wide distribution of flight numbers across various orbit types, indicating that SpaceX has launched missions to multiple types of orbits with varying frequencies.

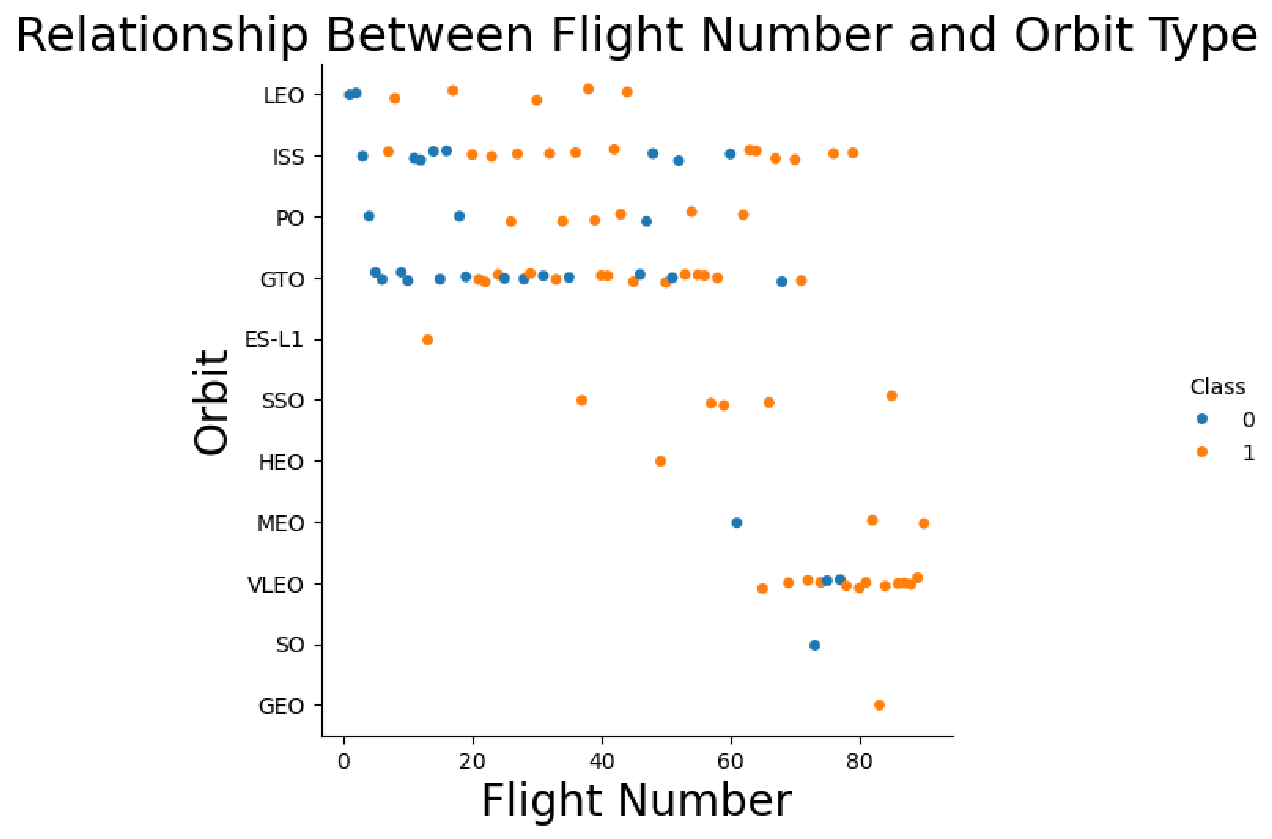
Class Distribution: There is a higher concentration of successful launches (Class 1) in certain orbit types such as LEO, ISS, and GTO.

Popular Orbits: LEO (Low Earth Orbit) has the highest number of launches, followed by ISS (International Space Station) and GTO (Geostationary Transfer Orbit) suggesting that these orbits are more frequently targeted by SpaceX missions.

Success Rates by Orbit: LEO, ISS, and GTO orbits have a higher number of successful launches (Class 1), indicating a potentially higher reliability or more successful missions to these orbits. There are fewer data points for orbits like ES-L1, SSO, HEO, MEO, VLEO, SO, and GEO, but the successful launches in these orbits indicate specialized missions with a generally high success rate.

Flight Number Distribution: Flight numbers are well-distributed across all orbit types, with no particular orbit type being limited to a specific range of flight numbers suggesting that SpaceX's missions to various orbits have been consistent over time.

Summary: The scatter plot provides a clear visualization of the relationship between flight numbers and orbit types, showing that certain orbits like LEO, ISS, and GTO are more frequently targeted and have a higher number of successful launches. The distribution of flight numbers across different orbit types indicates a consistent mission strategy over time. Further analysis could explore the specific factors contributing to success in these frequently targeted orbits.



Payload vs. Orbit Type

Payload Distribution: Payloads for LEO (Low Earth Orbit) and ISS (International Space Station) generally have lower masses, mostly below 6000 kg. GTO (Geostationary Transfer Orbit) has a wide range of payload masses, from around 2000 kg to over 16000 kg.

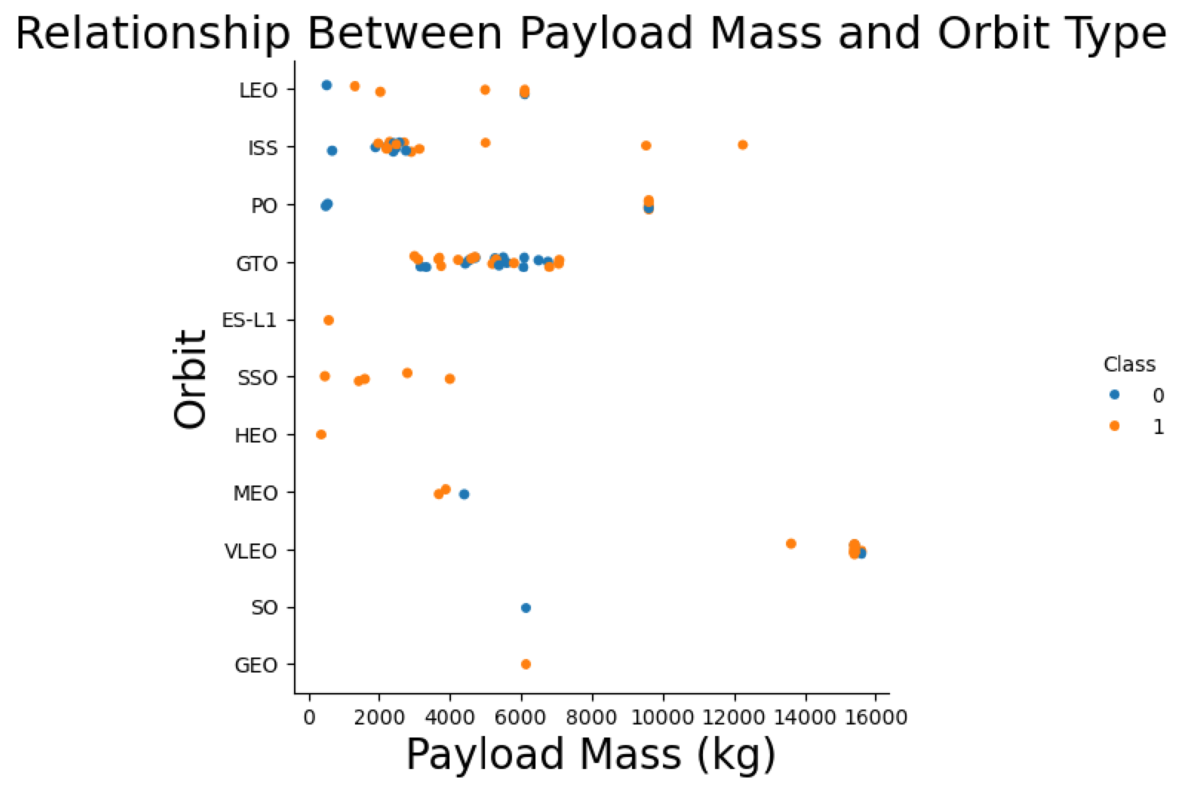
Popular Orbits: LEO, ISS, and GTO appear to be the most commonly targeted orbits, with the highest number of data points while SSO (Sun-Synchronous Orbit) also sees a significant number of payloads, mostly below 6000 kg.

Class Distribution: No clear dominance of one class over the other in any specific orbit type, indicating that both successful and unsuccessful launches are spread across different orbits.

Less Common Orbits: There are fewer data points for orbits like ES-L1, HEO, MEO, VLEO, SO, and GEO, indicating these orbits are less commonly used or have fewer recorded payloads.

Success Rates: Both Class 0 and Class 1 payloads are present in the commonly used orbits like LEO, ISS, and GTO, suggesting variability in launch outcomes for these orbits.

Summary: This scatter plot provides insights into the relationship between payload mass and orbit type, highlighting the commonality of certain orbits and the variability in launch outcomes. The plot shows that LEO, ISS, and GTO are frequently targeted orbits with a wide range of payload masses, while other orbits see fewer missions. Understanding the distribution and success rates across different orbits can aid in planning future missions.



Launch Success Yearly Trend

Initial Growth: From 2010 to 2013, the success rate remained at 0.0, indicating that there were no successful launches during these years.

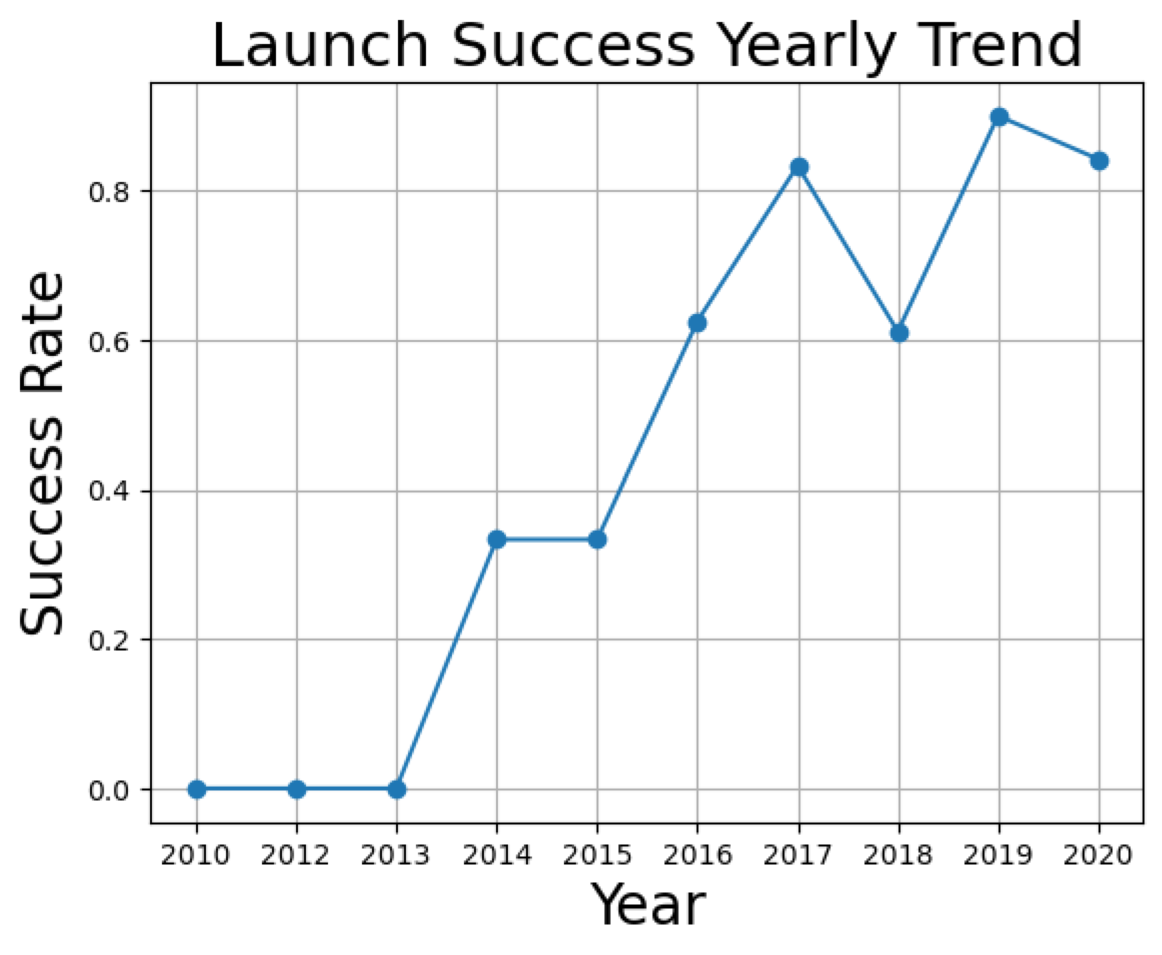
Steady Increase: In 2014, the success rate started to increase, reaching approximately 0.2. This trend continued into 2015, with the success rate rising to around 0.4.

Significant Improvement: By 2016, the success rate saw a notable increase, reaching about 0.6. The peak of the success rate occurred in 2017, reaching approximately 0.85, indicating a significant improvement in launch success.

Variability: In 2018, there was a drop in the success rate to around 0.6, suggesting some variability or challenges faced during this period. However, the success rate rebounded in 2019 to approximately 0.8.

Recent Stability: In 2020, the success rate slightly decreased to around 0.75, maintaining a relatively high level of success compared to earlier years.

Summary: This line plot illustrates the overall trend of increasing launch success rates over the decade, with notable improvements starting from 2014 and peaking in 2017. While there was some variability in 2018, the general trend shows a positive trajectory, indicating that SpaceX has become more successful in its launch attempts over time. This trend reflects improvements in technology, processes, and experience.



All Launch Site Names

SQL Query:

```
unique_launch_sites = df['Launch_Site'].unique()

print("Unique Launch Sites:")

print(unique_launch_sites)
```

Explanation:

This query provides a list of all unique launch site names.

1. CCAFS SLC-40 (Cape Canaveral Air Force Station Space Launch Complex 40)
2. VAFB SLC-4E (Vandenberg Air Force Base Space Launch Complex 4E)
3. KSC LC-39A (Kennedy Space Center Launch Complex 39A)
4. CCAFS LC-40 (Cape Canaveral Air Force Station Launch Complex 40)

Launch Site Names Begin with 'CCA'

SQL Query:

```
cca_launch_sites = df[df['Launch_Site'].str.startswith('CCA')]

print("5 Records where Launch Sites begin with 'CCA':")

print(cca_launch_sites.head(5))
```

```
5 Records where Launch Sites begin with 'CCA':
         Date Time (UTC) Booster_Version  Launch_Site  \
0  2010-06-04   18:45:00  F9 v1.0  B0003  CCAFS LC-40
1  2010-12-08   15:43:00  F9 v1.0  B0004  CCAFS LC-40
2  2012-05-22    7:44:00  F9 v1.0  B0005  CCAFS LC-40
3  2012-10-08    0:35:00  F9 v1.0  B0006  CCAFS LC-40
4  2013-03-01   15:10:00  F9 v1.0  B0007  CCAFS LC-40

                                                Payload  PAYLOAD_MASS__KG_  \
0                 Dragon Spacecraft Qualification Unit                  0
1  Dragon demo flight C1, two CubeSats, barrel of...                  0
2                                  Dragon demo flight C2                525
3                                         SpaceX CRS-1                500
4                                         SpaceX CRS-2                677

       Orbit         Customer Mission_Outcome      Landing_Outcome
0        LEO           SpaceX         Success  Failure (parachute)
1  LEO (ISS)  NASA (COTS) NRO         Success  Failure (parachute)
2  LEO (ISS)      NASA (COTS)         Success           No attempt
3  LEO (ISS)       NASA (CRS)         Success           No attempt
4  LEO (ISS)       NASA (CRS)         Success           No attempt
```

Explanation:

- This query lists 5 records where launch sites begin with `CCA` to identify specific sites.

Total Payload Mass From NASA

SQL Query:

```
nasa_crs_payload_mass = df[df['Booster_Version'].str.contains('NASA (CRS)', regex=False)]['PAYLOAD_MASS__KG_'].sum()

print(f"Total Payload Mass carried by boosters launched by NASA (CRS): {nasa_crs_payload_mass}")
```

Output: `Total Payload Mass carried by boosters launched by NASA (CRS): 0`

Explanation:

- This query calculates the total payload carried by boosters from NASA.

Average Payload Mass by F9 v1.1

SQL Query:

```
avg_payload_mass_f9v1_1 = df[df['Booster_Version'] == 'F9 v1.1']['PAYLOAD_MASS__KG_'].mean()

print(f"Average Payload Mass carried by booster version F9 v1.1: {avg_payload_mass_f9v1_1}")
```

Output:

```
Average Payload Mass carried by booster version F9 v1.1: 2928.4
```

Explanation:

- This query calculates the average payload mass carried by booster version F9 v1.1.

First Successful Ground Landing Date

SQL Query:

```
first_successful_ground_pad_landing = df[df['Landing_Outcome'] == 'Success (ground pad)']['Date'].min()

print(f"Date of first successful landing outcome on the ground pad: {first_successful_ground_pad_landing}")
```

Output: `Date of first successful landing outcome on the ground pad: 2015-12-22`

Explanation:

- This query finds the first successful ground landing.

Successful Drone Ship Landing with Payload between 4000 and 6000

SQL Query:

```
successful_drone_ship_boosters = df[(df['Landing_Outcome'] == 'Success (drone ship)') & (df['PAYLOAD_MASS__KG_'] > 4000) & (df['PAYLOAD_MASS__KG_'] <
6000)]['Booster_Version’]

print("Names of boosters with successful landing on drone ship and payload mass between 4000 and 6000:")

print(successful_drone_ship_boosters.unique())
```

Output:

```
Names of boosters with successful landing on drone ship and payload mass between 4000 and 6000:
['F9 FT B1022' 'F9 FT B1026' 'F9 FT  B1021.2' 'F9 FT  B1031.2']
```

Explanation:

- This query calculates the average payload mass carried by booster version F9 v1.1.

Total Number of Successful and Failure Mission Outcomes

SQL Query:

```
mission_outcomes = df['Landing_Outcome'].value_counts()

print("Total number of successful and failure mission outcomes:")

print(mission_outcomes)
```

```
Total number of successful and failure mission outcomes:
Landing_Outcome
Success                   38
No attempt                21
Success (drone ship)      14
Success (ground pad)       9
Controlled (ocean)         5
Failure (drone ship)       5
Failure                    3
Uncontrolled (ocean)       2
Failure (parachute)        2
Precluded (drone ship)     1
No attempt                 1
```

Explanation:

- This query calculates the total number of successful and failed mission outcomes.

Boosters Carried Maximum Payload

SQL Query:

```
max_payload_mass = df['PAYLOAD_MASS__KG_'].max()boosters_max_payload_mass = df[df['PAYLOAD_MASS__KG_'] == max_payload_mass]['Booster_Version’]

print("Names of booster versions which have carried the maximum payload mass:")

print(boosters_max_payload_mass.unique())
```

Output:

```
Names of booster versions which have carried the maximum payload mass:
['F9 B5 B1048.4' 'F9 B5 B1049.4' 'F9 B5 B1051.3' 'F9 B5 B1056.4'
 'F9 B5 B1048.5' 'F9 B5 B1051.4' 'F9 B5 B1049.5' 'F9 B5 B1060.2 '
 'F9 B5 B1058.3 ' 'F9 B5 B1051.6' 'F9 B5 B1060.3' 'F9 B5 B1049.7 ']
```

Explanation:

- This query lists the names of the booster that carried the maximum payload mass.

2015 Launch Records

SQL Query:

```
df['Date'] = pd.to_datetime(df['Date’])

df['Month'] = df['Date'].dt.month_name()

records_2015 = df[(df['Date'].dt.year == 2015) & (df['Landing_Outcome'] == 'Failure (drone ship)')][['Month', 'Landing_Outcome', 'Booster_Version', 'Launch_Site’]]

print("Records for 2015 with failure landing outcomes on drone ship:")

print(records_2015)
```

Outcome:

```
Records for 2015 with failure landing outcomes on drone ship:
      Month       Landing_Outcome Booster_Version  Launch_Site
13  January  Failure (drone ship)   F9 v1.1 B1012  CCAFS LC-40
16    April  Failure (drone ship)   F9 v1.1 B1015  CCAFS LC-40
```

Explanation:

- This query lists the failed landing outcomes on drone ships, their booster versions, and launch site names for the year 2015.

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

SQL Query:

```
landing_outcomes_ranked = df[(df['Date'] >= '2010-06-04') & (df['Date'] <= '2017-03-20')]['Landing_Outcome'].value_counts().reset_index()

landing_outcomes_ranked.columns = ['Landing_Outcome', 'Count’]

landing_outcomes_ranked = landing_outcomes_ranked.sort_values(by='Count', ascending=False)

print("Rank of landing outcomes between 2010-06-04 and 2017-03-20:")

print(landing_outcomes_ranked)
```

```
Rank of landing outcomes between 2010-06-04 and 2017-03-20:
         Landing_Outcome  Count
0             No attempt     10
1    Failure (drone ship)     5
2    Success (drone ship)     5
3     Controlled (ocean)      3
4   Success (ground pad)      3
5     Failure (parachute)     2
6   Uncontrolled (ocean)      2
7  Precluded (drone ship)     1
```

Explanation:

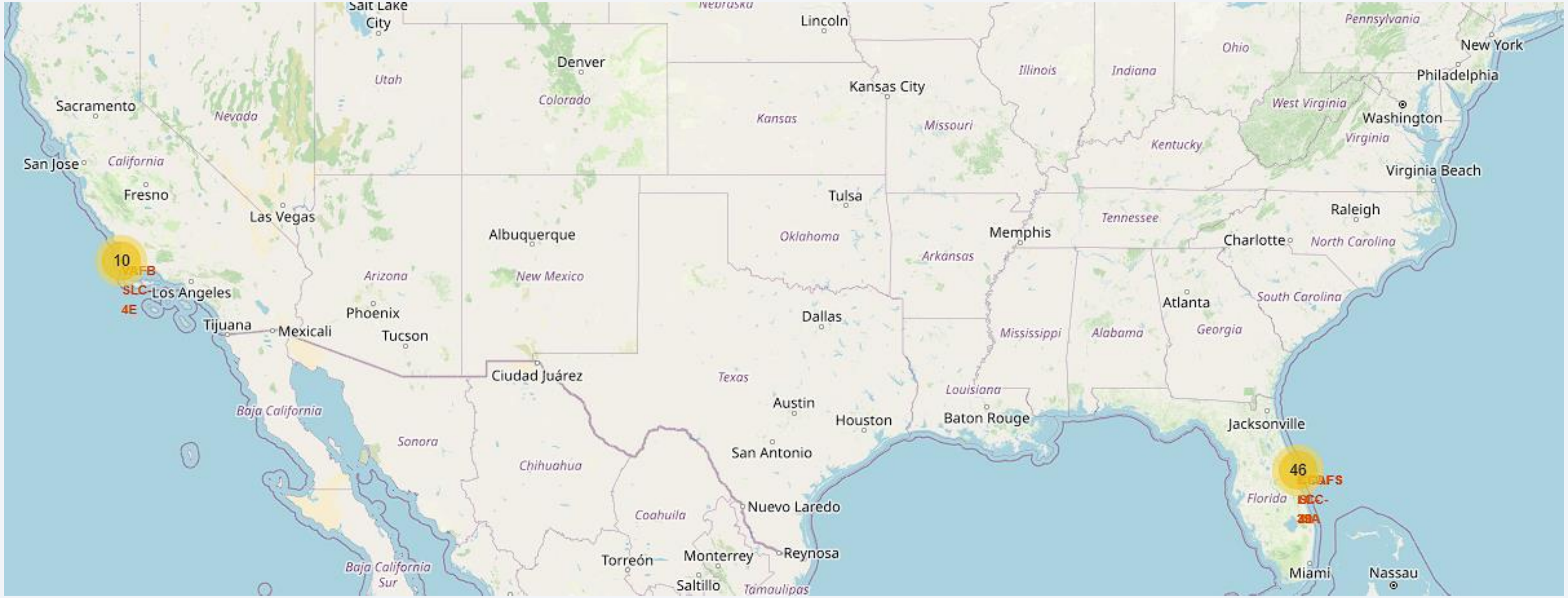
- This query lists the failed landing outcomes on drone ships, their booster versions, and launch site names for the year 2015.

Section 3

Launch Sites Proximities Analysis

Global Map

Oceanic Coastal Launch Sites Overview

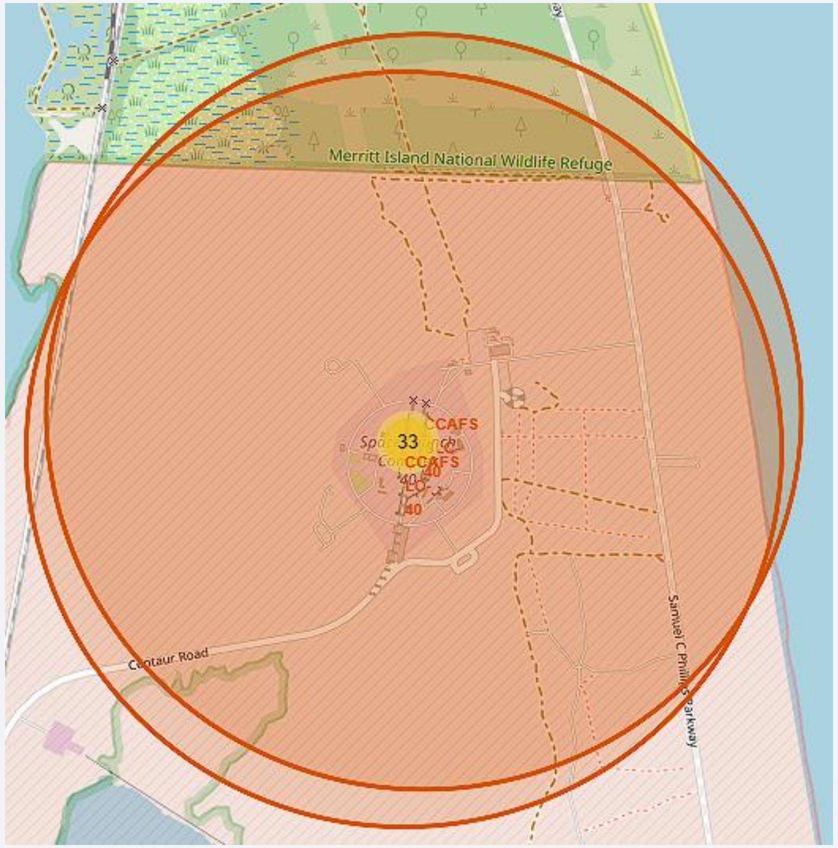


Important elements include the distribution of launch sites globally and the success rate visualization through color-coded markers.

Findings highlight the concentration of launch sites in specific regions and their relative success rates.

Cape Canaveral Launch Pad

Launch Outcomes CCAFS – LC 40



Observations:

Successful launches are marked in green, showing some successes for this site, though the overall success rate is around 27%, indicating challenges that have been faced.

Cape Canaveral Launch Pad

CCAF LC-40 Proximity to Important Locations

East Coastline: 1.43 km

West Coastline: 1.05 km

Railway: 0.83 km

Highway: 0.75 km

Cape Canaveral: 19.20 km

Observations:

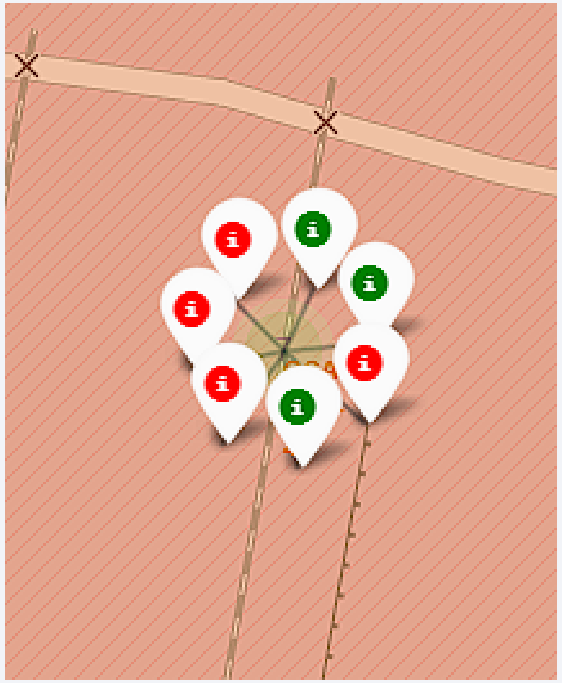
The launch site is proximate to a major highway, ensuring easy transport access.

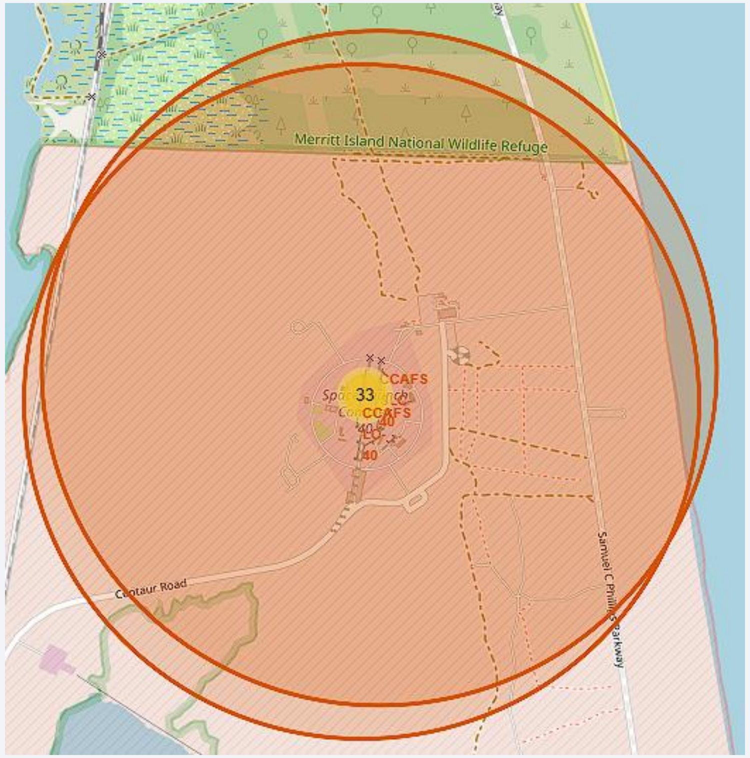
A railway line is nearby, facilitating the transportation of heavy equipment.

The coastline is relatively close, providing an emergency landing area for failed launches.

Cape Canaveral Sub-Launch Pad

Launch Outcomes CCAF – SLC 40





Observations:

- The map displays seven launch attempts at a specific site with successes in green and failures in red. This results in a success rate of approximately 43% for this site.
- Despite the relatively modest success rate, achieving reliable outcomes is possible with further optimizations and improvements.

Cape Canaveral Sub-Launch Pad

CCAF LC-40 Proximity to Important Locations

East Coastline: 0.87 km
West Coastline: 1.25 km
Railway: 0.93 km
Highway: 0.78 km
Cape Canaveral: 19.26 km

Observations:

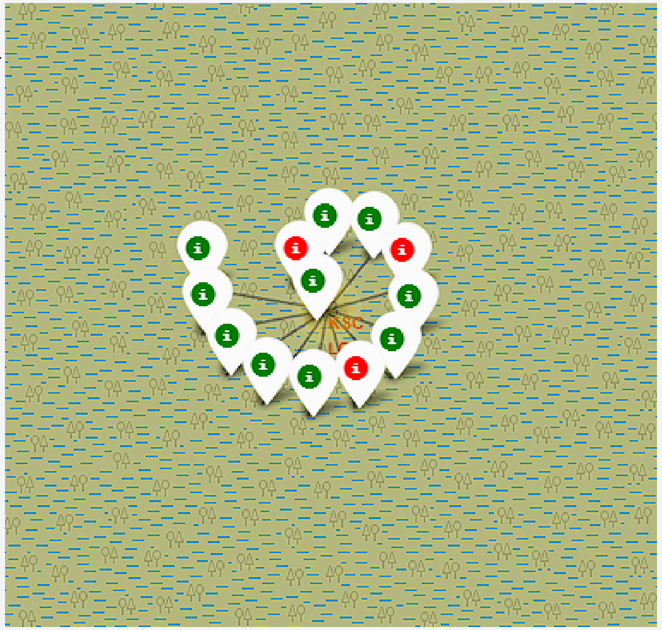
The launch site is proximate to a major highway, ensuring easy transport access.

A railway line is nearby, facilitating the transportation of heavy equipment.

The coastline is relatively close, providing an emergency landing area for failed launches.

Kennedy Space Center Launch Pad

Launch Outcomes for KSC LC-39A



Observations:

At the KSC LC-39A launch site, a total of 13 launch attempts have been recorded. Among these, 10 are marked in green, indicating successful launches, and 3 are marked in red, indicating failed launches. This results in a success rate of approximately 77%.

The high success rate of 77% at this launch site is indicative of the overall reliability and effectiveness of the launch operations conducted here.

Kennedy Space Center Launch Pad

KSC LC-39A Proximity to Important Facilities

East Coastline: 3.91 km
West Coastline: 6.88 km
Railway: 0.36 km
Highway: 0.36 km
Merritt Island: 21.92 km

Observations:

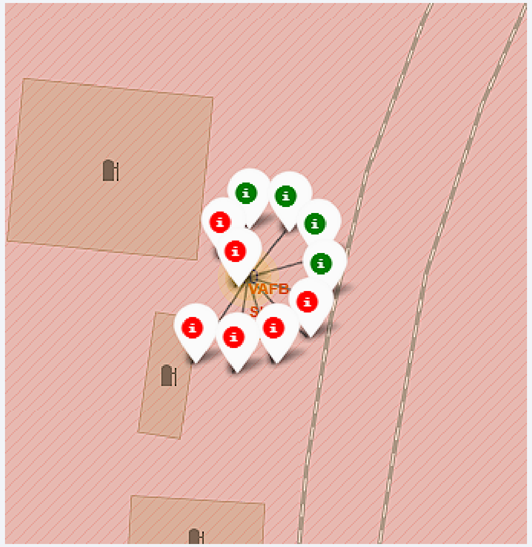
Located near major highways for efficient transport.

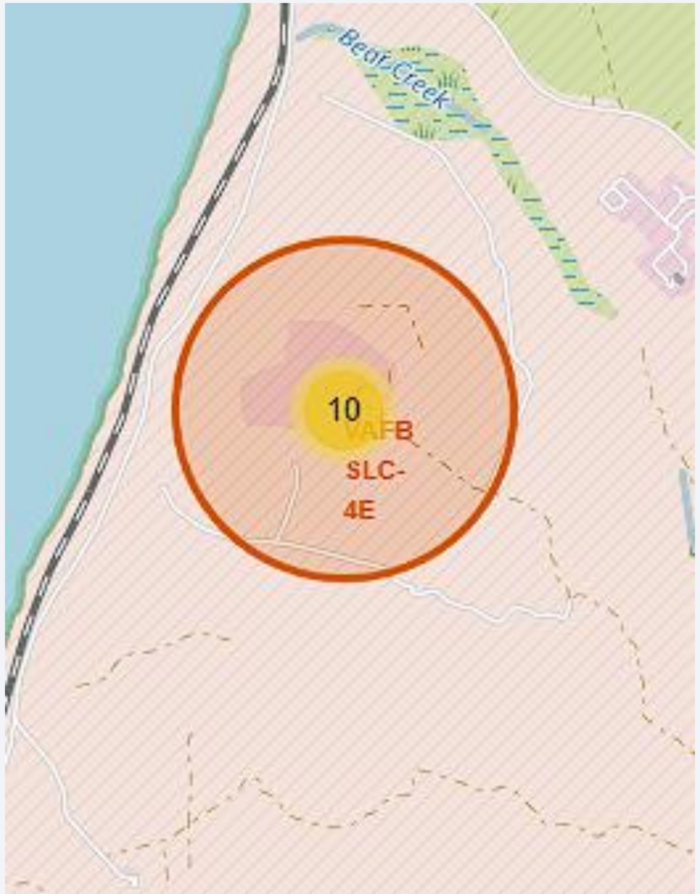
Close to railway lines for the transportation of heavy equipment.

Proximity to the coastline provides a safe landing area for aborted launches.

Vandenberg Air Force Base Sub-Launch Pad

Launch Outcomes for VAFB SLC-4E



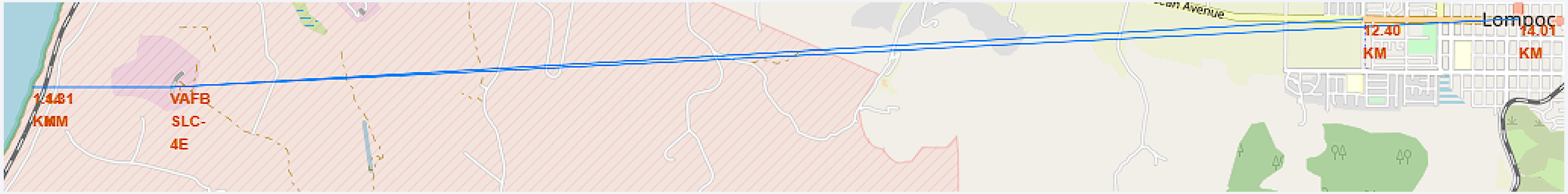


Observations:

The map of the VAFB SLC-4E launch site shows a total of 10 launch attempts. 4 launches are marked in green, indicating successful launches showing a success rate of 40%.

The success rate at VAFB SLC-4E is 40%, highlights the challenges faced at this site, as well as the importance of ongoing improvements to increase the reliability and success of future launches from this site.

Vandenberg Air Force Base Sub-Launch Pad



VAFB SLC-4E Proximity to Important Facilities

West Coastline: 1.44 km

Railway: 1.31 km

Highway: 12.40 km

Lompoc: 14.01 km

Observations:

Proximity to major highways facilitates the transport of materials and personnel.

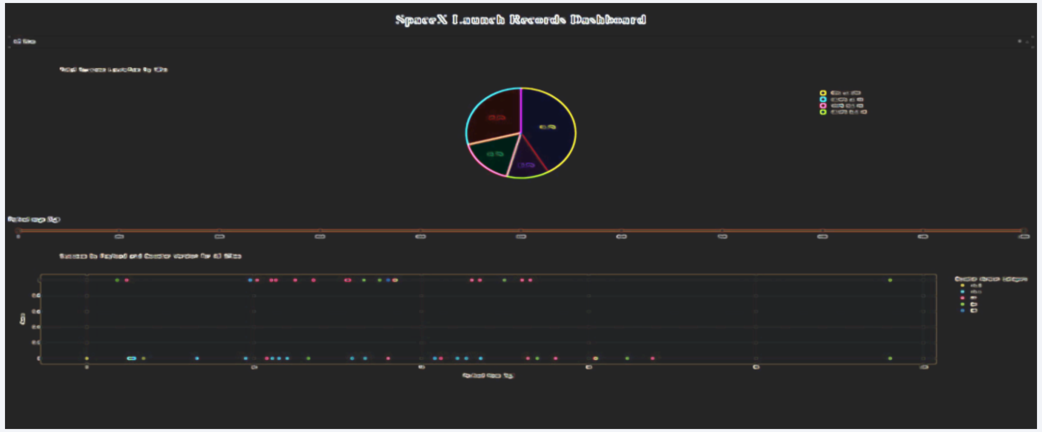
Nearby railway lines aid in the movement of heavy launch components.

The close coastline ensures safety during potential launch anomalies.

Section 4

Build a Dashboard with Plotly Dash

SpaceX Launch Records Dashboard



Dropdown Menu:
Function: Select a launch site to filter data.
Options: All Sites, CCAFS LC-40, CCAFS SLC-40, KSC LC-39A, VAFB SLC-4E.
Usage: Choose a site to see specific launch outcomes.

Payload Range Slider:
Function: Adjust the range of payload mass for analysis.
Range: 0 to 10,000 kg.
Usage: Move the slider to set the minimum and maximum payload mass.

Success Pie Chart:

Function: Visualizes total success launches by site.

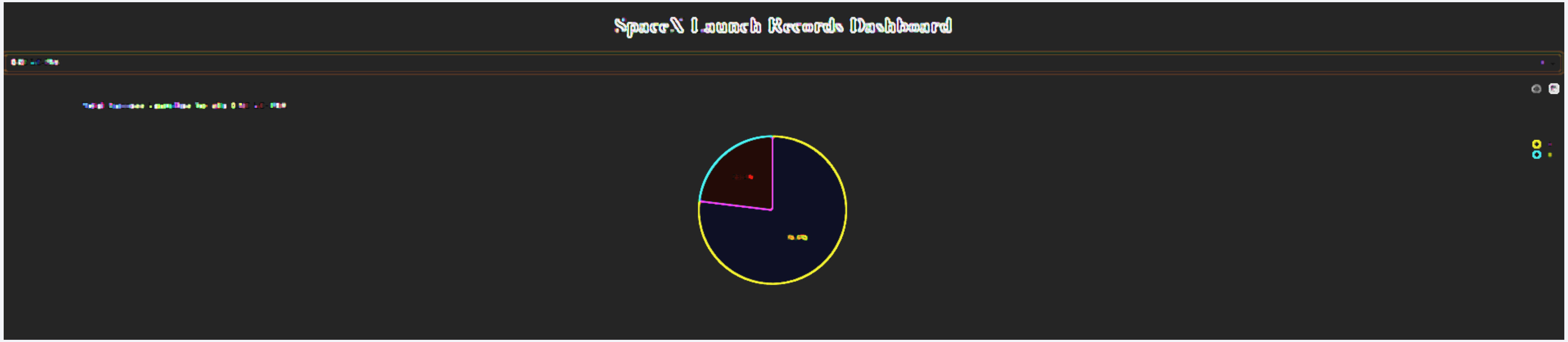
Interaction: Updates based on selected site from the dropdown.

Success by Payload Scatter Chart:

Function: Shows the relationship between payload mass and launch success.

Interaction: Filters based on payload range set by the slider and selected site.

Most Successful Launch Site



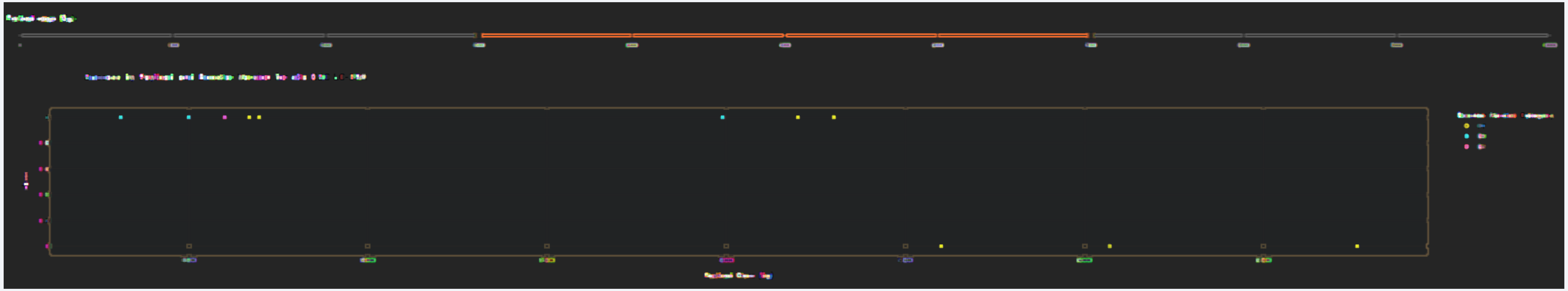
Launch Success Rates for All Sites

By selecting the KSC LC-39A, this pie chart only shows the count of successful launches for the selected launch site.

Explanation:

The chart reveals the ratio of successful launches, indicating its reliability and performance for KSC LC-39A.

Payload Launch Success and Failures



Payload vs. Launch Scatter Plot

By moving the slider to the desired payload range and using the drop down this scatter plot shows success based on payload.

Explanation:

The plot helps identify which payload ranges or booster versions have the highest success rates, providing insights into the optimal payloads for successful launches.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

Why Logistic Regression is Chosen?

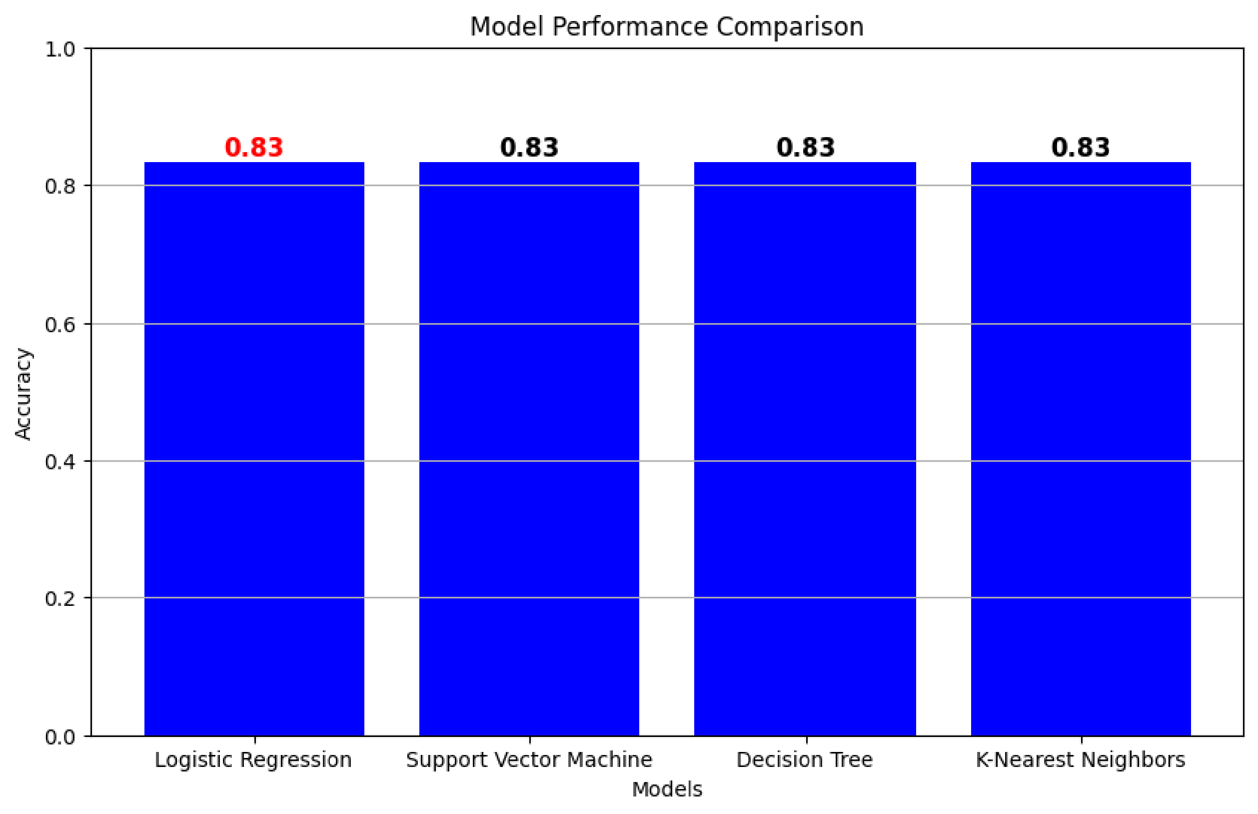
Interpretability: Clear and easy to understand; provides probability outputs.

Training Time & Complexity: Fast to train; requires minimal hyperparameter tuning.

Scalability: Efficiently handles large datasets and high-dimensional data.

Robustness to Overfitting: Less prone to overfitting with regularization.

Probabilistic Outputs: Offers reliable probability scores for decision-making.



Confusion Matrix

Logistic Regression Model

The model has an overall accuracy of 83.33%, with 80% precision for "landed" predictions and 100% recall for identifying actual "landed" instances.
However, it has a 50% recall rate for "did not land" instances, indicating challenges in accurately predicting this class.

Best Parameters:
{'C': 0.01, 'penalty': 'l2', 'solver': 'lbfgs'}

Training Accuracy:
0.8464285714285713

Test Accuracy:
0.8333333333333334

Explanation of the Confusion Matrix:

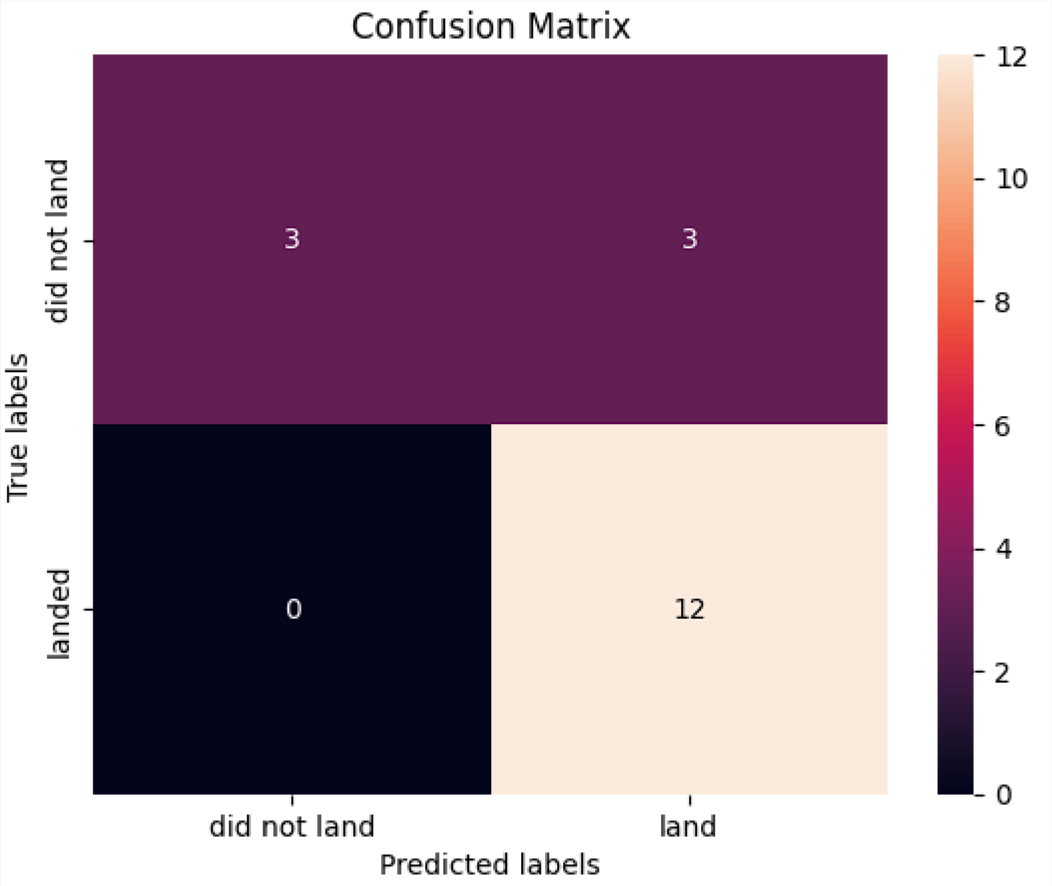
Accuracy (83.33%)

Precision (Landed) (80%)

Recall (Landed) (100%)

Precision (Did Not Land) (100%)

Recall (Did Not Land) (50%)



Conclusions

1. **KSC LC-39A Success Rate:**
   - KSC LC-39A stands out as the most successful launch site, with a 77% success rate. This indicates its reliability and effectiveness in conducting successful launches compared to other sites.
2. **Payload Range:**
   - The optimal payload range for successful launches is between 2880.0 and 3840.0 kg. Launches within this range have shown higher success rates, making it a crucial factor for mission planning.
3. **Booster Version Reliability:**
   - Booster version B5 has proven to be the most reliable, with the highest success rate. This reinforces the need to prioritize the use of this booster version for future missions to improve the likelihood of success.
4. **Model Accuracy:**
   - Predictive models demonstrated high accuracy in forecasting launch success, with Logistic Regression highlighted for its simplicity and interpretability. The model's training accuracy is 84.64%, and test accuracy is 83.33%, aiding in future mission planning.
5. **Confusion Matrix Insights:**
   - The confusion matrix reveals an accuracy of 83.33%. Precision for "landed" predictions is 80%, with a perfect recall of 100%. For "did not land," the precision is also 100%, but the recall is 50%, indicating room for improvement in predicting this class.

Appendix

Jupyter Notebooks with Outputs and Resources:

GitHub Repository Full Collection

API Data Collection

Webscraping Data Collection

Data Wrangling

EDA with SQL

EDA Visualizations with SQL

Launch Site Folium Maps

SpaceX Dash App

SpaceX Machine Learning Prediction

Thank you!