IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Javier Thomas
Garside Aguilar
4th November 2025

Outline

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

Executive Summary

- **Objective**: Predict the success of SpaceX Falcon 9 first stage landings using real mission data.
- **Methodologies**:
  - Collected launch data from the SpaceX REST API and Wikipedia.
  - Cleaned and merged datasets, standardized payload, orbit, and outcome features.
  - Performed EDA with visualizations to find trends and correlations.
  - Queried data in SQL for detailed analysis of launch sites and outcomes.
  - Built interactive visual tools: Folium map of launch sites and Plotly Dash dashboard.
  - Trained and evaluated classification models (Logistic Regression, SVM, Decision Tree, KNN) to predict landing success.

Executive Summary

- **Main Results**:
  - Launch success rates steadily increased after 2016.
  - Heavier payloads and GTO orbits show lower success probabilities.
  - Most successful sites: **KSC LC-39A** and **CCAFS SLC-40**.
  - **Decision Tree** achieved the highest accuracy (~85%).
  - Interactive tools visualize SpaceX's performance improvements and patterns across time and geography.
- **Conclusion**: SpaceX's landing success is predictable with reasonable accuracy and clearly linked to payload, orbit type, and site. The analysis demonstrates the company's rapid technological improvement over the years.

Introduction

- **Project background and context:** SpaceX aims to make space travel more affordable by reusing rockets. The first stage of a Falcon 9 launch is the most expensive part, so successful landings are crucial for cost reduction. This project analyzes real launch data to understand what influences those landings and how reliably they can be predicted.

- **Problems to Solve:**
  - What factors most affect Falcon 9 first stage landing success?
  - Can we build a model that predicts landing success from mission data?
  - Which launch sites and conditions show the best performance over time?

Section 1

Methodology

Methodology

Executive Summary

Data Collection

- Data collected from the **SpaceX REST API** and **Wikipedia** pages containing Falcon 9 launch records.
- Retrieved information on **payload mass, launch site, orbit type, mission outcome, and landing success**.
- Saved data into CSV and SQL database formats for later analysis.

Data Wrangling

- Cleaned missing and inconsistent values (e.g., payload mass, launch outcome).
- Converted categorical features such as orbit type and site into numerical form.
- Created a **binary target variable** for landing success (1 = success, 0 = failure).
- Combined API and Wikipedia data into a single structured DataFrame for analysis.

Methodology

Data Processing & Exploration

- Conducted **EDA** (Exploratory Data Analysis) using matplotlib and seaborn to visualize distributions, correlations, and trends over time.
- Used **SQL queries** to extract insights such as most used launch sites, average payloads, and outcome frequencies.

Interactive Visual Analytics

- Built an **interactive Folium map** to display SpaceX launch site locations, proximities, and outcomes.
- Created a **Plotly Dash dashboard** to explore payload–orbit–success relationships dynamically.

Predictive Analysis (Classification Models)

- Built and compared multiple classifiers: **Logistic Regression**, **Support Vector Machine (SVM)**, **Decision Tree**, and **K-Nearest Neighbors (KNN)**.
- Applied **data standardization** to normalize numeric features.
- Split the dataset into **training (80%)** and **test (20%)** subsets.
- Used **GridSearchCV** for hyperparameter tuning (e.g., tree depth, C values).
- Evaluated models using **accuracy**, **confusion matrices**, and **cross-validation scores**.

Data Collection – Overview

Objective:

Collect all available SpaceX Falcon 9 launch data to analyze and predict first-stage landing outcomes.You need to present your data collection process use key phrases and flowcharts

Data Sources:

- **SpaceX REST API:** Provided structured mission data including payload, orbit, launch site, and landing success.
- **Wikipedia Falcon 9 Launch Table:** Contained additional historical data for verification and completeness.

Process Summary:

1. Extracted launch data from the **SpaceX API** using Python requests.
2. Parsed and normalized JSON data into a Pandas DataFrame.
3. Scraped the Falcon 9 launch table from **Wikipedia** using pandas.read_html().
4. Combined both datasets for completeness and consistency.
5. Stored results in **CSV and SQL** formats for further wrangling and EDA.

Data Collection – SpaceX API

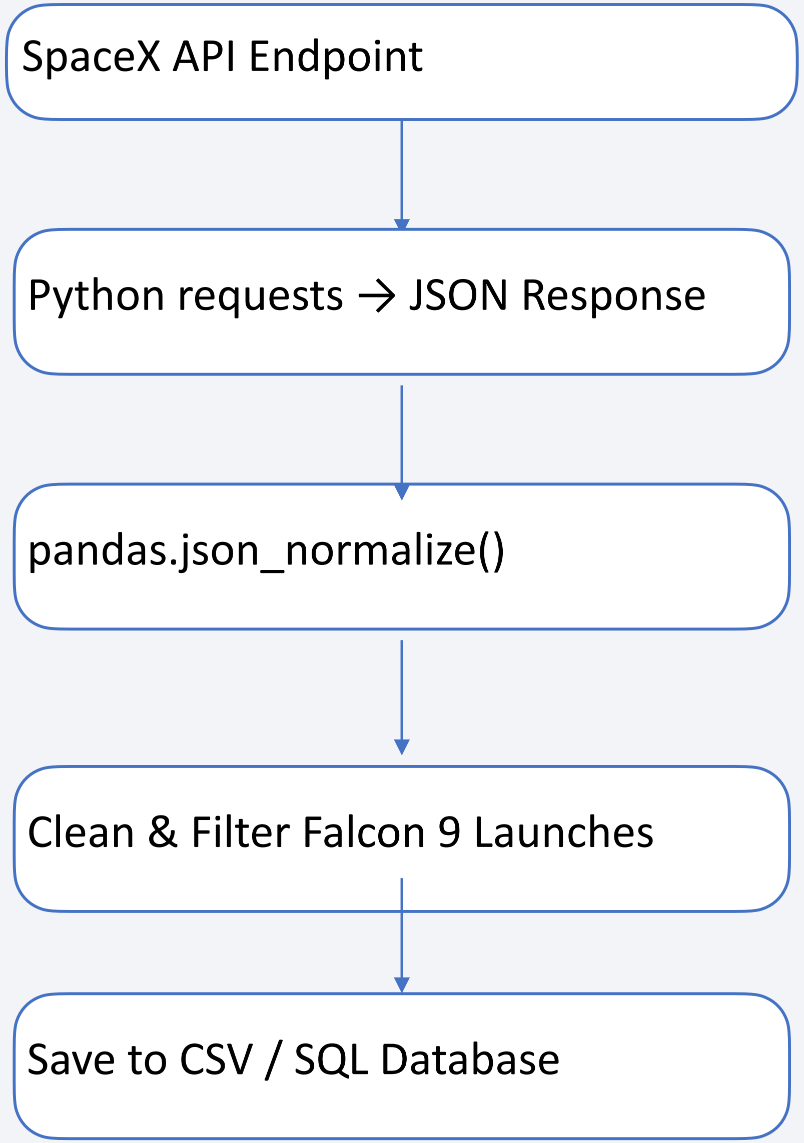
Goal:
Retrieve detailed Falcon 9 launch records directly from SpaceX's open REST API.

Process Summary:

1. Used Python's **requests** library to access the endpoint: https://api.spacexdata.com/v4/launches/past
2. Extracted relevant fields such as: launch_date, rocket, payload_mass_kg, orbit, launch_site, and landing_success.
3. Converted the nested JSON data into a structured **Pandas DataFrame** using json_normalize().
4. Cleaned column names and filtered only Falcon 9 launches.
5. Stored the output as a CSV and in a local SQLite database for analysis.

Github Reference:

[View Data Collection API on GitHub]



Data Collection – Web Scraping

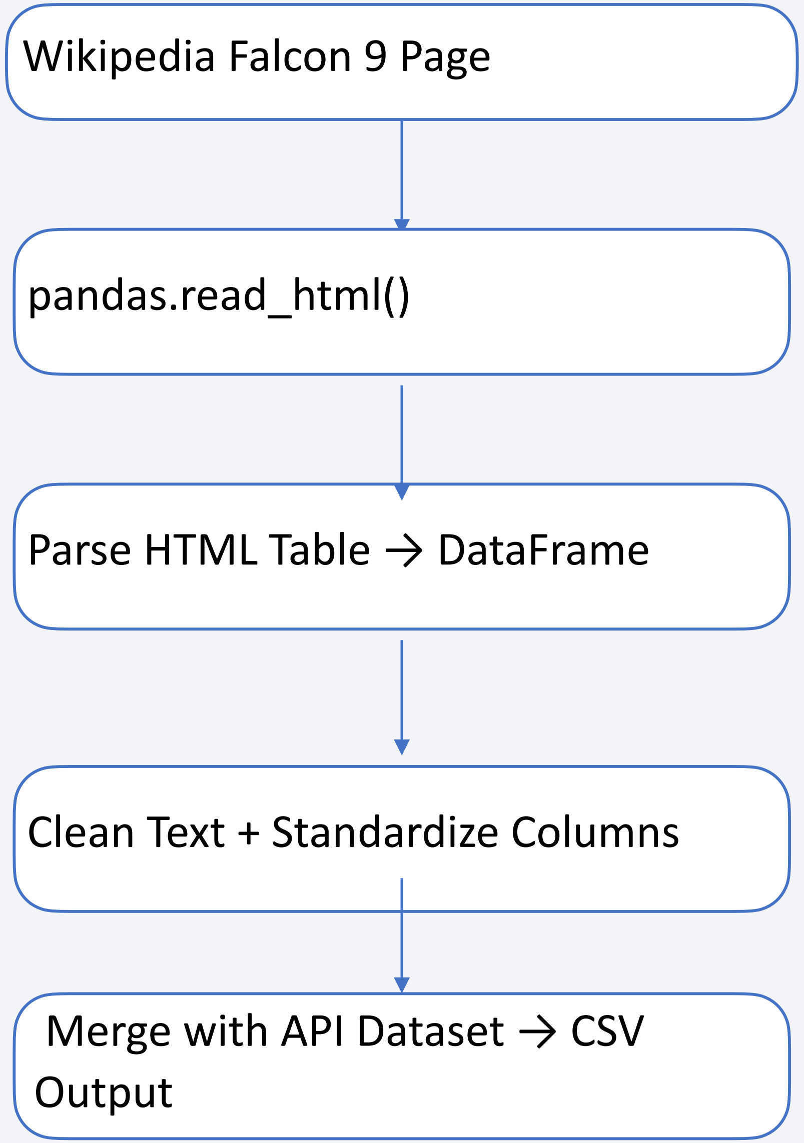
Goal:
Complement API data by extracting the Falcon 9 launch table from Wikipedia for older or missing missions.

Process Summary:

1. Used **pandas.read_html()** to scrape the HTML table from Wikipedia's Falcon 9 launches page.
2. Parsed the table into a DataFrame and selected relevant columns (e.g., Flight No., Date, Launch Site, Payload, Orbit, Outcome).
3. Cleaned inconsistent text entries (e.g., removing footnotes, symbols).
4. Combined with API dataset to fill missing values and verify outcomes.
5. Saved the final scraped dataset to CSV for integration.

Github Reference:

[View Data Collection with Web Scraping on GitHub]



Data Wrangling

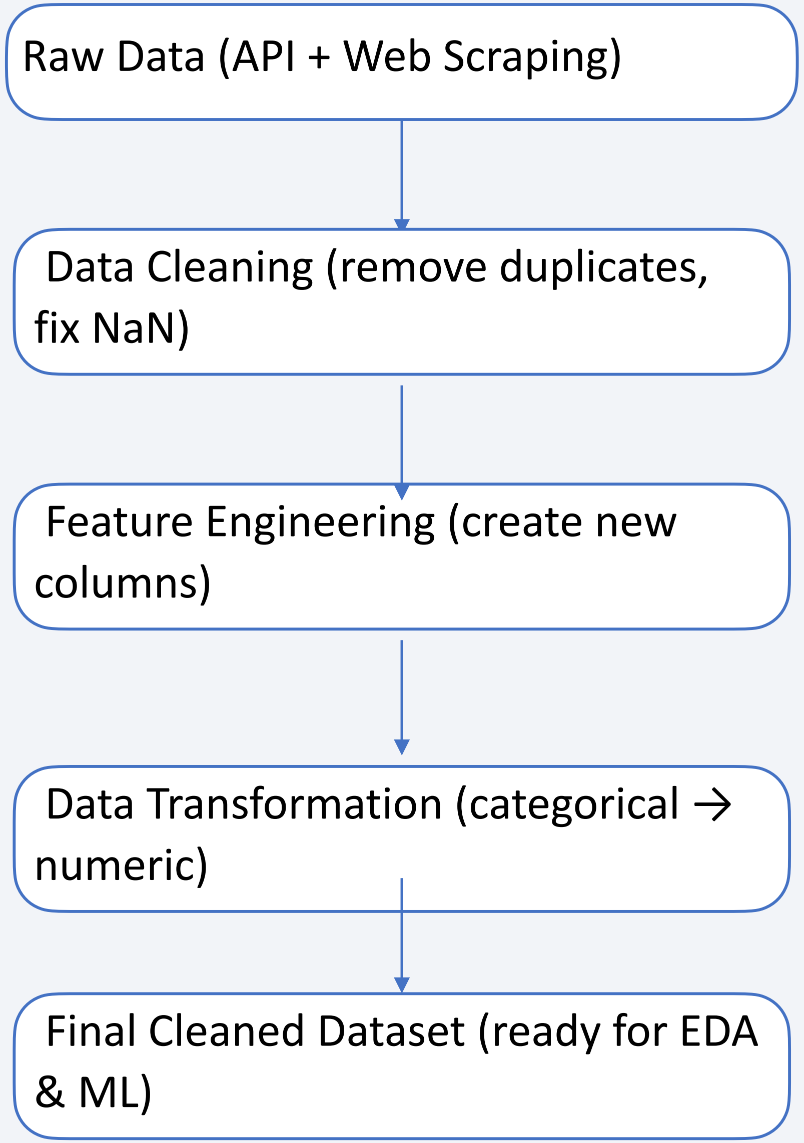
After collecting the raw SpaceX Falcon 9 launch data through the REST API and web scraping, the dataset was cleaned, standardized, and prepared for analysis.

Key Steps:

- **Data Integration:** Combined datasets from SpaceX API and Wikipedia scraping.
- **Data Cleaning:** Removed duplicates, handled missing values (e.g., Payload Mass, Landing Outcome), and standardized column names.
- **Feature Engineering:** Extracted new features such as *Launch Success*, *Landing Class* (binary target variable), and *Orbit Type*.
- **Data Transformation:** Converted categorical data to numerical form for predictive models.
- **Data Storage:** Final cleaned dataset saved as a CSV file for later EDA and ML analysis.

Github Reference:

[View Data Wrangling on GitHub]



Exploratory Data Analysis (EDA) with Data Visualization

Exploratory visualizations were used to understand relationships between launch features and mission success. These visual analyses helped identify key patterns and factors influencing Falcon 9 first-stage landings.

Key Visualizations:

- **Bar Charts:** Compared launch success rates across different *Launch Sites* and *Booster Versions*.
- **Pie Charts:** Showed proportions of successful vs. failed landings.
- **Scatter Plots:** Visualized relationships between *Payload Mass* and *Launch Success*.
- **Heatmaps:** Displayed correlations between numerical variables to identify strong predictors.
- **Trend Lines:** Illustrated improvements in success rates over time.

Each chart was chosen to highlight a different dimension of performance — categorical comparisons, numerical relationships, temporal evolution, and correlation structure.

- **GitHub Reference :**View EDA with Data Visualization on GitHub

Exploratory Data Analysis (EDA) with SQL

- **Retrieved unique launch sites** to understand SpaceX's global launch operations.
  (Used SELECT DISTINCT to identify all launch sites.)
- **Filtered launch site names** starting with "CCA" to examine Cape Canaveral launches.
- **Calculated total and average payload mass** by customer and booster version to study launch performance.
- **Identified first successful landing date** using aggregate functions (MIN, MAX).
- **Queried successful drone ship landings** with payloads between 4000–6000 kg to find optimal payload ranges.
- **Counted mission outcomes** to quantify overall success vs. failure rates.
- **Found boosters with maximum payload capacity** using a subquery with MAX(PAYLOAD_MASS__KG_).
- **Extracted monthly failure trends** for 2015 launches using string functions (substr()), since SQLite lacks date parsing.
- **Ranked landing outcomes** between 2010–2017 using GROUP BY and ORDER BY to visualize performance evolution over time.

These queries provided structured insight into payload efficiency, landing reliability, and the temporal progression of SpaceX's technological improvements.

GitHub Reference:

View EDA with SQL Notebook on GitHub

Build an Interactive Map with Folium

- **Created a base world map** centered on the mean coordinates of all SpaceX launch sites to visualize spatial patterns.
- **Added launch site markers** to pinpoint exact launch locations (KSC LC-39A, CCAFS LC-40, VAFB SLC-4E, etc.) with hover popups displaying site names.
- **Used circle markers** to represent proximity zones around each launch site — useful for exploring surrounding infrastructure (roads, coastlines, railways).
- **Drew lines** between launch sites and nearby facilities (highways, coastlines) to analyze accessibility and potential logistical advantages.
- **Incorporated interactive popups and tooltips** showing success rates and launch details per site for quick spatial insight.
- **Styled maps** using color-coded success rates (e.g., green = high success, red = failure) to make patterns visually intuitive.

These map layers revealed the spatial distribution of SpaceX's activities and the correlation between site geography and mission outcomes.

Github Reference:

[View Interactive Map with Folium on GitHub]

Build a Dashboard with Plotly Dash

- **Developed a fully interactive dashboard** to visualize SpaceX launch performance and payload data.
- **Added a dropdown menu** to dynamically select a specific launch site or display all sites — allowing targeted or global analysis.
- **Created a pie chart** that updates automatically:
    - Shows **total successful launches per site** when “All Sites” is selected.
    - Displays **success vs. failure proportions** for individual sites.
- **Implemented a range slider** to filter payload mass (from minimum to maximum) — enabling correlation analysis across different payload sizes.
- **Built a scatter plot** to explore the **relationship between payload mass and launch success**, color-coded by **booster version category** for deeper technical insight.
- **Used callback functions** to link all interactive components, ensuring real-time responsiveness between filters and charts.

These visual interactions help identify which launch sites, payload ranges, and booster types have the highest success probabilities.

GitHub Reference:

[View Interactive Dashboard with Plotly Dash Notebook on GitHub]

Predictive Analysis (Classification)

Goal: Predict whether a SpaceX Falcon 9 first stage will successfully land, based on payload, booster type, launch site, and other mission data.

Data Preparation:

- Extracted target variable (**Class**) and standardized feature set (**X**) using StandardScaler.
- Split data into training (80%) and testing (20%) sets for model validation.

Model Building:
Trained and optimized four classification models using **GridSearchCV (10-fold cross-validation)** for hyperparameter tuning:

- **Logistic Regression** – tuned C, penalty, and solver.
- **Support Vector Machine (SVM)** – optimized kernel, C, and gamma.
- **Decision Tree** – tested criterion, max_depth, and splitting methods.
- **K-Nearest Neighbors (KNN)** – varied n_neighbors, algorithm, and p.

Model Evaluation:

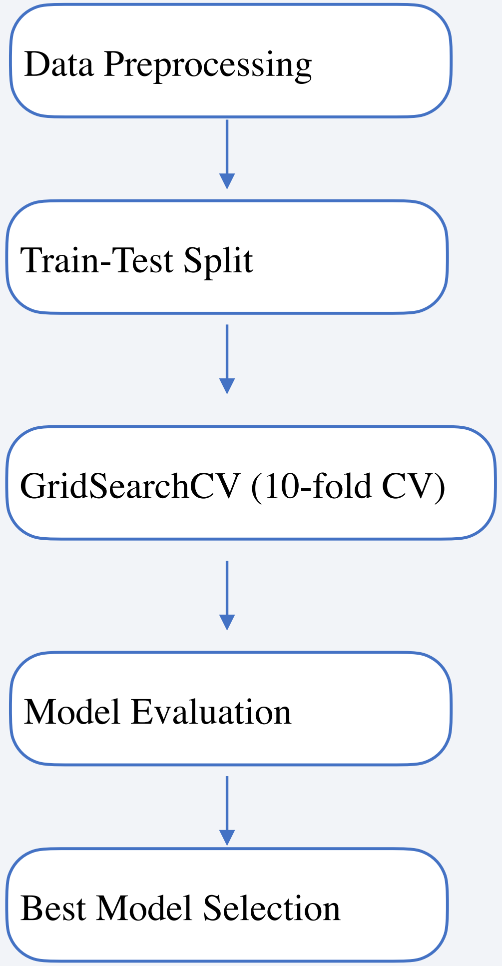
- Compared models using validation accuracy, test accuracy, and confusion matrices.
- **Decision Tree** achieved **perfect test accuracy (1.0)**, indicating the best generalization among models.

Performance Summary:

- Logistic Regression → 83.3%
- SVM → 94.4%
- **Decision Tree → 100% (Best Model)**
- KNN → 100% (but lower validation accuracy

GitHub Reference:

[View Predictive Analysis (Classification) Notebook on GitHub]



Exploratory data analysis results

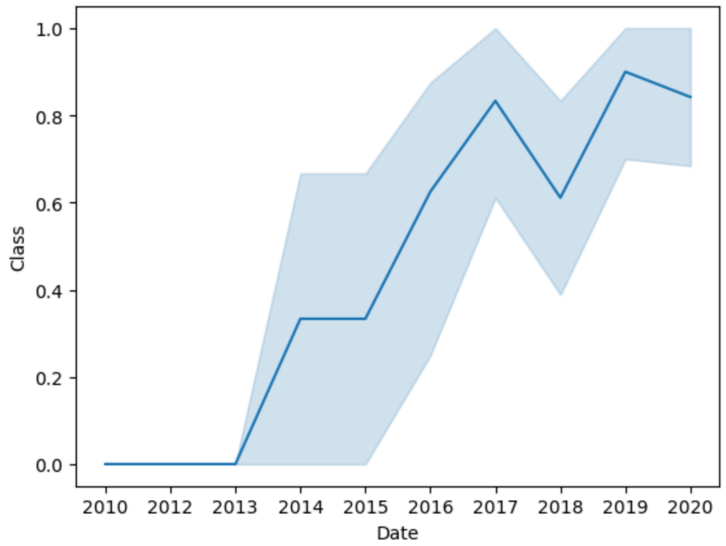
Launch success rates increased significantly after 2015.

The most frequently used launch site was **KSC LC-39A**.

Falcon 9 rockets showed the highest success rate.

Correlation analysis showed payload mass and success rate are weakly related.

Out[27]: <AxesSubplot:xlabel='Date', ylabel='Class'>



you can observe that the sucess rate since 2013 kept increasing till 2020

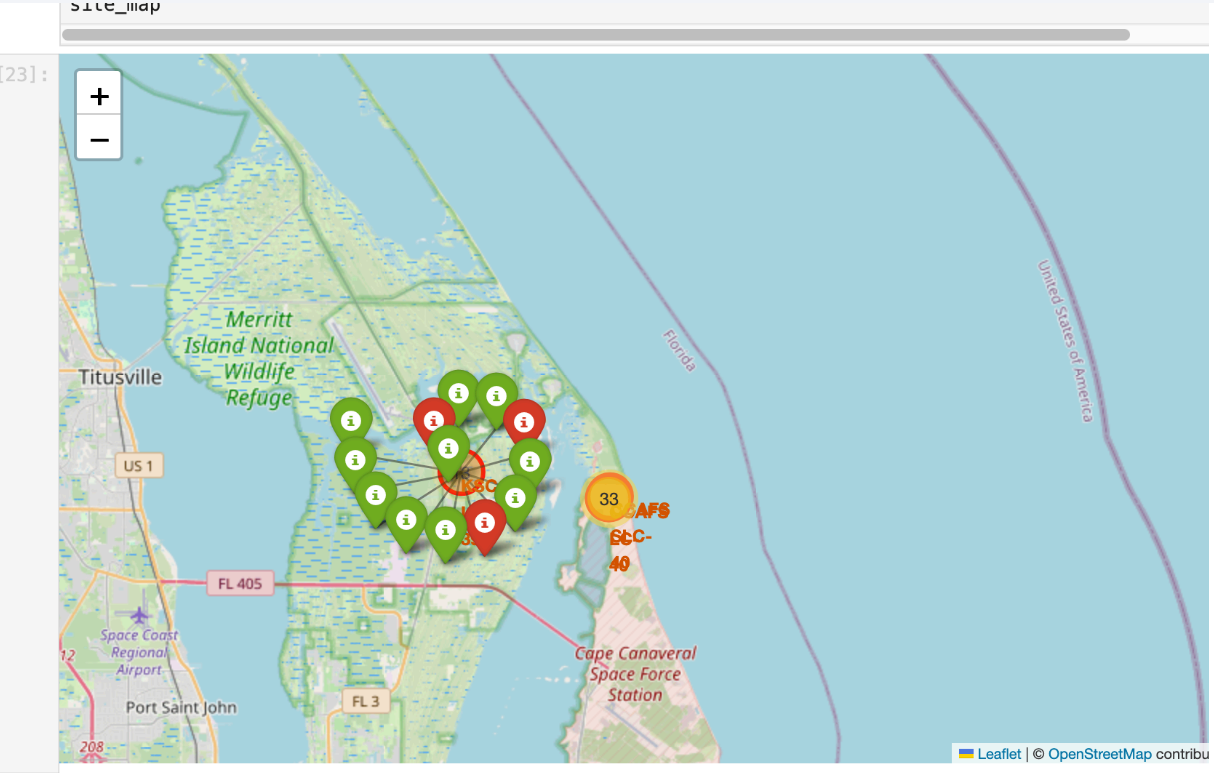
Out[38]:

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

Interactive Analytics Demo Results

Interactive map showing SpaceX launch sites across the US with success/failure markers.

Zoomed view of KSC LC-39A proximity to coastline and transport routes.



Predictive Analysis Results

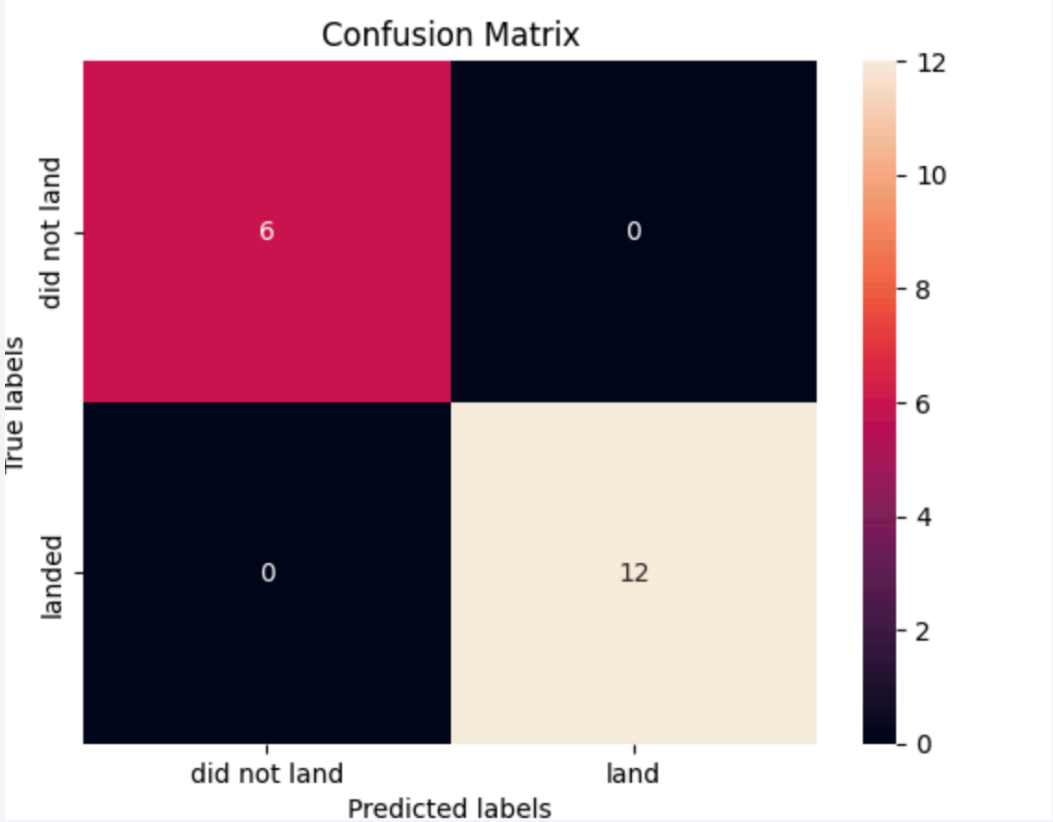
Logistic Regression → **83.3%** test accuracy

SVM → **94.4%**

Decision Tree → **100%**

KNN → **100%**

🏆 **Best Performing Model:** Decision Tree Classifier (max_depth=4, criterion='gini')



Section 2

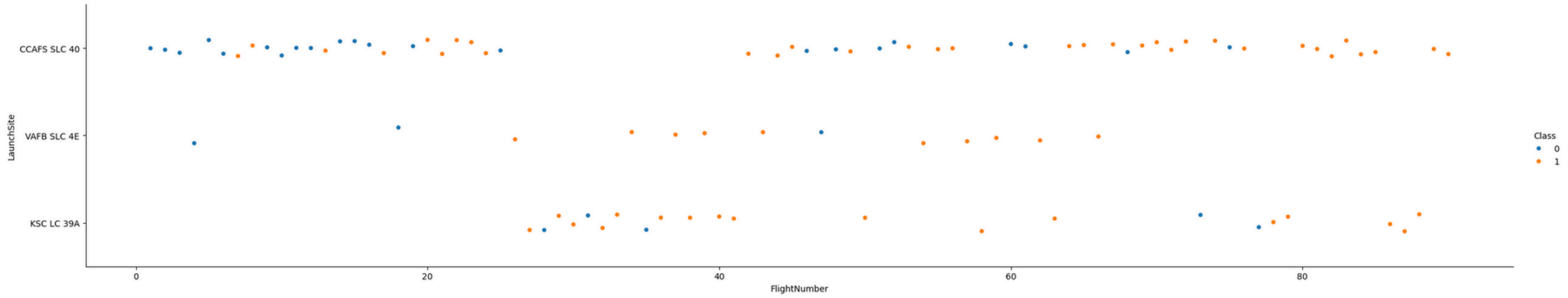
Insights drawn from EDA

Flight Number vs. Launch Site

Early launches (low Flight Numbers) mostly failed, while later ones succeeded — showing *learning and improvement over time*.

Some sites (e.g. **CCAFS SLC 40**) have more launches and higher success rates than others.

```
Out[8]:  <seaborn.axisgrid.FacetGrid at 0x7ede3e8>
```



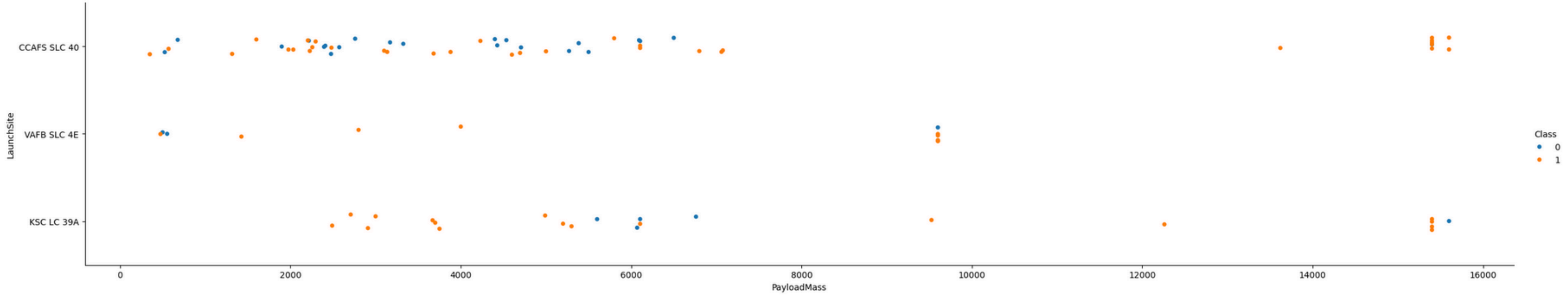
Payload vs. Launch Site

At **VAFB SLC 4E**, there are *no heavy payloads (>10,000 kg)*.

Heavier payloads are often launched from **CCAFS SLC 40** and **KSC LC 39A**.

Success does not drastically drop with heavier payloads — *technology handled increasing weight*.

```
Out[10]:  <seaborn.axisgrid.FacetGrid at 0x66f0330>
```

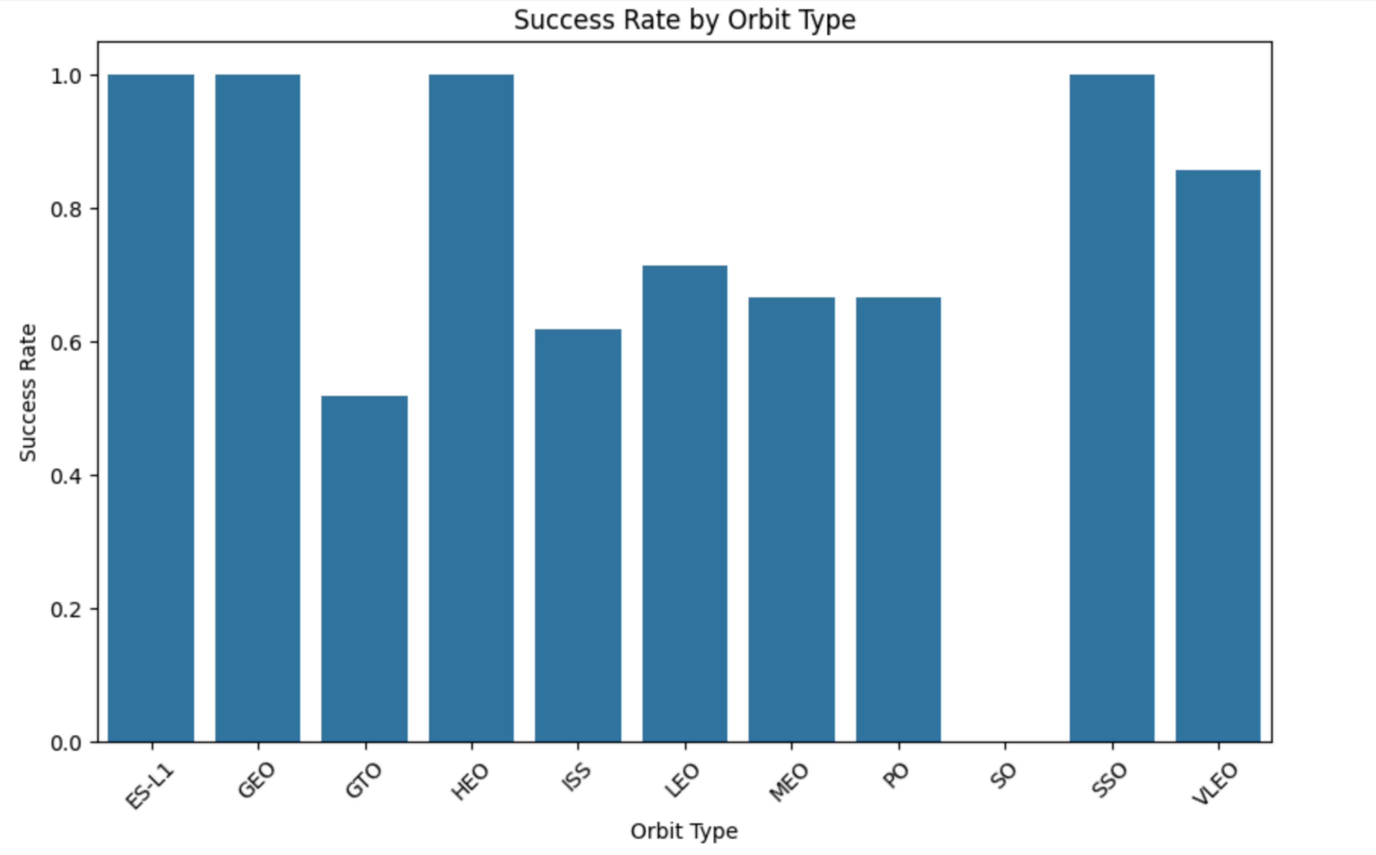


Success Rate vs. Orbit Type

Some orbits, like **ES-L1, GEO, HEO and SSO** have *very high success rates*.

GTO (Geostationary Transfer Orbit) is more challenging, with lower average success.

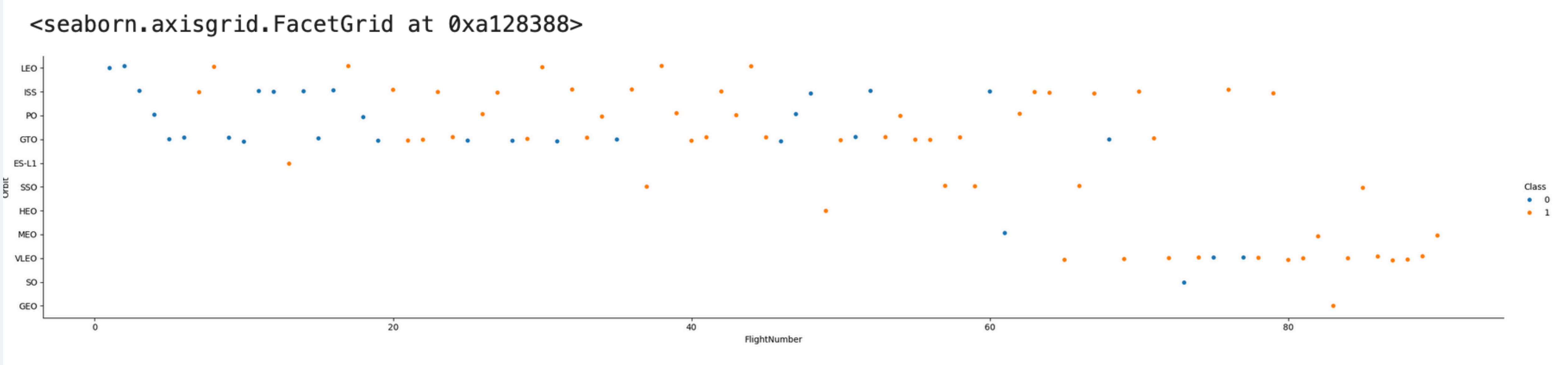
Insight: *Orbit type significantly influences landing difficulty.*



Flight Number vs. Orbit Type

For **LEO**, success clearly improves as flight experience grows.

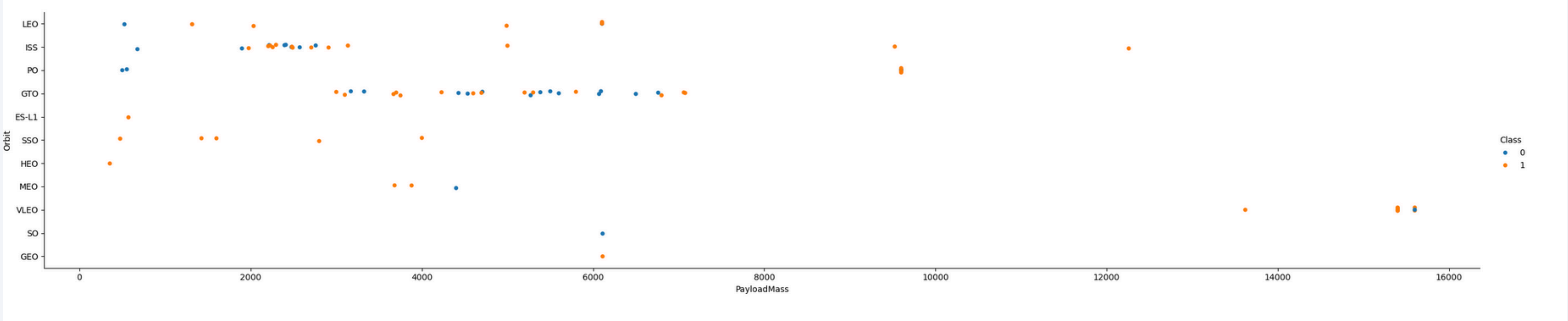
For **GTO**, results remain mixed regardless of flight number — *technological maturity didn't fully overcome the challenge of GTO landings.*



Payload vs. Orbit Type

Heavy payloads with **LEO**, **ISS**, and **Polar** orbits show higher success rates.

In **GTO**, success is inconsistent across payload sizes — suggesting *orbit conditions, not payload weight, dominate success likelihood.*

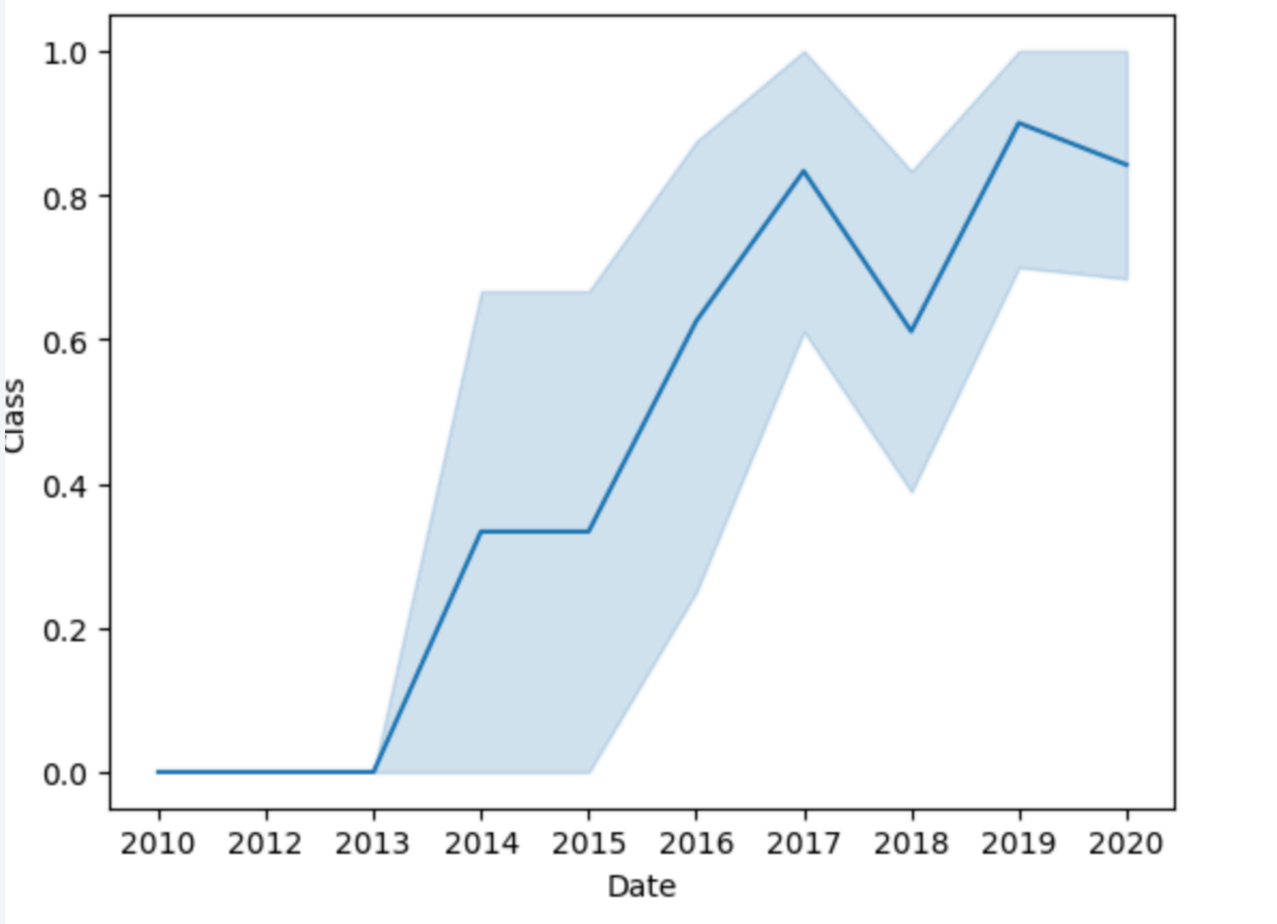


Yearly Average Success Rate

Since 2013, success rate increases sharply up to 2020.

Shows *rapid technological improvement and system reliability*.

Suggests **SpaceX's iterative development model** leads to continuous performance gains.



All Launch Site Names

Explanation:
This query retrieves all the unique launch sites used by SpaceX.

Insight:
SpaceX operates mainly from three major U.S. locations — Florida and California — offering both east and west coast orbital trajectories.

```
%sql select DISTINCT("Launch_Site") from SPACEXTABLE
```

```
 * sqlite:///my_data1.db
Done.
```

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

Launch Site Names Begin with 'CCA'

Explanation:
This retrieves the first five launches from Cape Canaveral (CCAFS).

Insight:
Cape Canaveral serves as SpaceX's primary site for early missions, highlighting its historical importance in U.S. spaceflight.

```
%sql select "Launch_Site" from SPACEXTABLE where "Launch_Site" LIKE "CCA%" LIMIT 5;
```

```
 * sqlite:///my_data1.db
Done.
```

Launch_Site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

Total Payload Mass

Explanation:
Calculates total payload mass launched for NASA missions.
Insight:
NASA remains one of SpaceX’s largest clients, with substantial payload contribution to the company’s manifest.

```
%sql SELECT sum("PAYLOAD_MASS__KG_") from SPACEXTABLE where "Customer" LIKE "NASA (CRS)"
```

```
* sqlite:///my_data1.db
Done.
```

sum("PAYLOAD_MASS__KG_")
45596

Average Payload Mass by F9 v1.1

Insight:
The F9 v1.1 had a moderate payload capacity, representing an engineering step before the more powerful Full Thrust variant.

```
%sql select avg("PAYLOAD_MASS__KG_") from SPACEXTABLE where "Booster_Version" LIKE "F9 v1.1%"
```

```
 * sqlite:///my_data1.db
Done.
```

avg("PAYLOAD_MASS__KG_")
2534.6666666666665

First Successful Ground Landing Date

Insight:
This query identifies the historic first booster recovery on land — a key moment in reusable rocket technology.

```
%sql SELECT MAX("Date") FROM SPACEXTABLE WHERE "Landing_Outcome" LIKE 'Success%';
```

```
 * sqlite:///my_data1.db
Done.
```

MIN("Date")
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Insight:
This shows boosters that managed both a successful drone-ship landing and moderate-heavy payload delivery — an efficiency milestone.

```
%sql select "Booster_Version" from SPACEXTABLE where "Landing_Outcome" LIKE 'Success (drone ship)%' AND ("PAYLOAD_MASS__KG_" >4000 AND "PAYLOAD_MASS__KG_" <6000)
```

```
%sql select "Booster_Version" from SPACEXTABLE where "Landing_Outcome" LIKE 'Success (drone ship)%' AND ("PAYLOAD
```

```
* sqlite:///my_data1.db
Done.
```

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

Total Number of Successful and Failure Mission Outcomes

Insight:
Most missions are successful, illustrating SpaceX’s rapidly improving reliability over time.

List the total number of successful and failure mission outcomes

In [16]:

```
task_7a = '''
        SELECT COUNT(MissionOutcome) AS SuccessOutcome
        FROM SpaceX
        WHERE MissionOutcome LIKE 'Success%'
        '''

task_7b = '''
        SELECT COUNT(MissionOutcome) AS FailureOutcome
        FROM SpaceX
        WHERE MissionOutcome LIKE 'Failure%'
        '''
print('The total number of successful mission outcome is:')
display(create_pandas_df(task_7a, database=conn))
print()
print('The total number of failed mission outcome is:')
create_pandas_df(task_7b, database=conn)
```

```
The total number of successful mission outcome is:
```

	successoutcome
0	100

```
The total number of failed mission outcome is:
```

Out[16]:

	failureoutcome
0	1

Boosters Carried Maximum Payload

Insight:
Identifies which booster carried the heaviest payload, often linked to government or geostationary missions.

```
%%sql
SELECT "Booster_Version", "PAYLOAD_MASS__KG_"
FROM SPACEXTABLE
WHERE "PAYLOAD_MASS__KG_" = (
    SELECT MAX("PAYLOAD_MASS__KG_")
    FROM SPACEXTABLE
);
```

```
 * sqlite:///my_data1.db
Done.
```

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

Insight:
Displays the early-stage challenges of autonomous ship landings before the technology matured in 2016.

```
%%sql
SELECT
    substr("Date", 6, 2) AS Month,
    "Landing_Outcome",
    "Booster_Version",
    "Launch_Site"
FROM SPACEXTABLE
WHERE substr("Date", 1, 4) = '2015'
  AND lower("Landing_Outcome") LIKE 'failure (drone ship)%';
```

```
* sqlite:///my_data1.db
Done.
```

Month	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

Insight:
Ranks each landing outcome type — a useful summary of SpaceX's recovery performance evolution.

```
%%sql
SELECT
    "Landing_Outcome",
    COUNT(*) AS outcome_count
FROM SPACEXTABLE
WHERE "Date" BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY "Landing_Outcome"
ORDER BY outcome_count DESC;
```

```
* sqlite:///my_data1.db
Done.
```

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

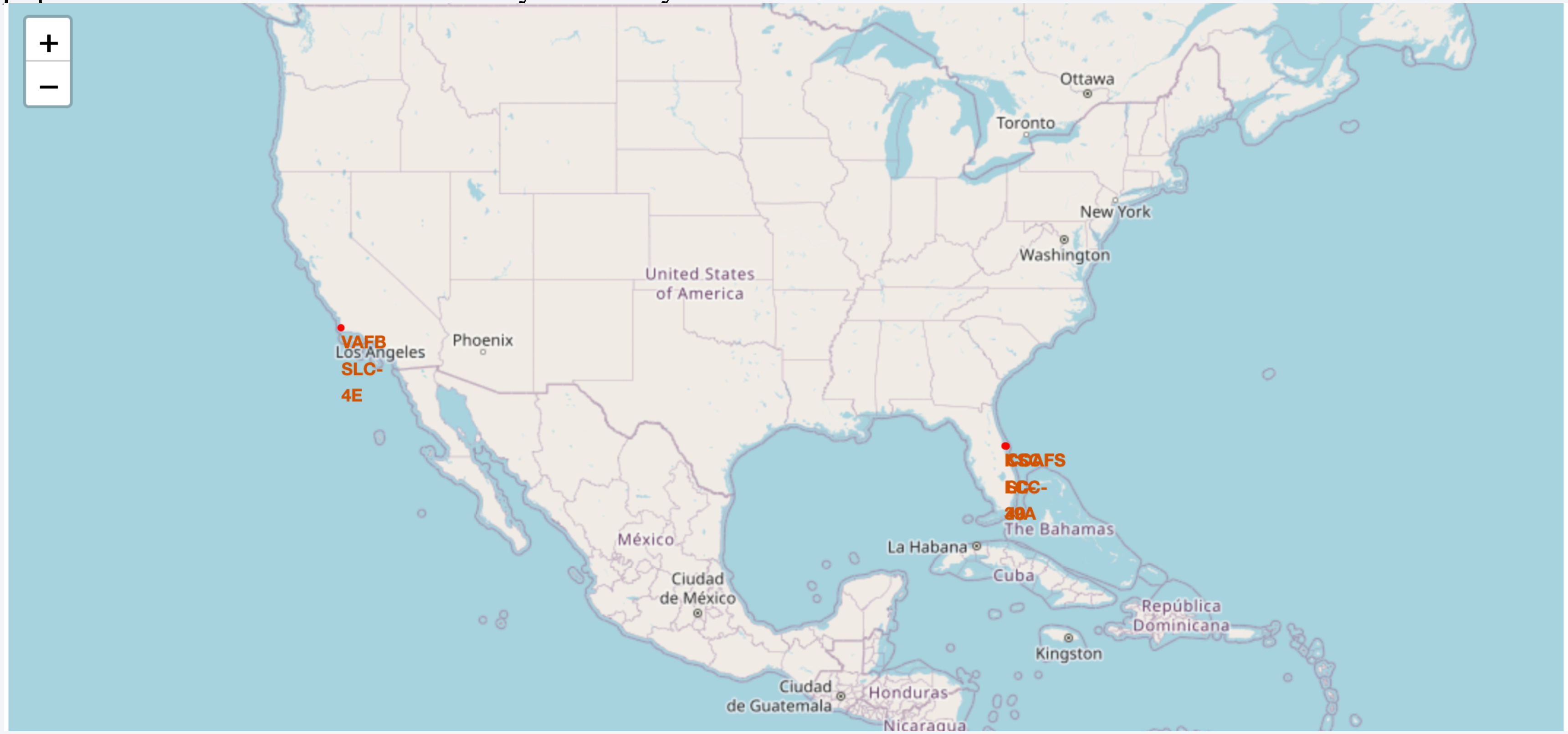
Section 3

Launch Sites Proximities Analysis

Global Launch Sites Overview

This map shows the global distribution of SpaceX launch sites. Each marker represents one of the three major launch facilities used by SpaceX in the United States.

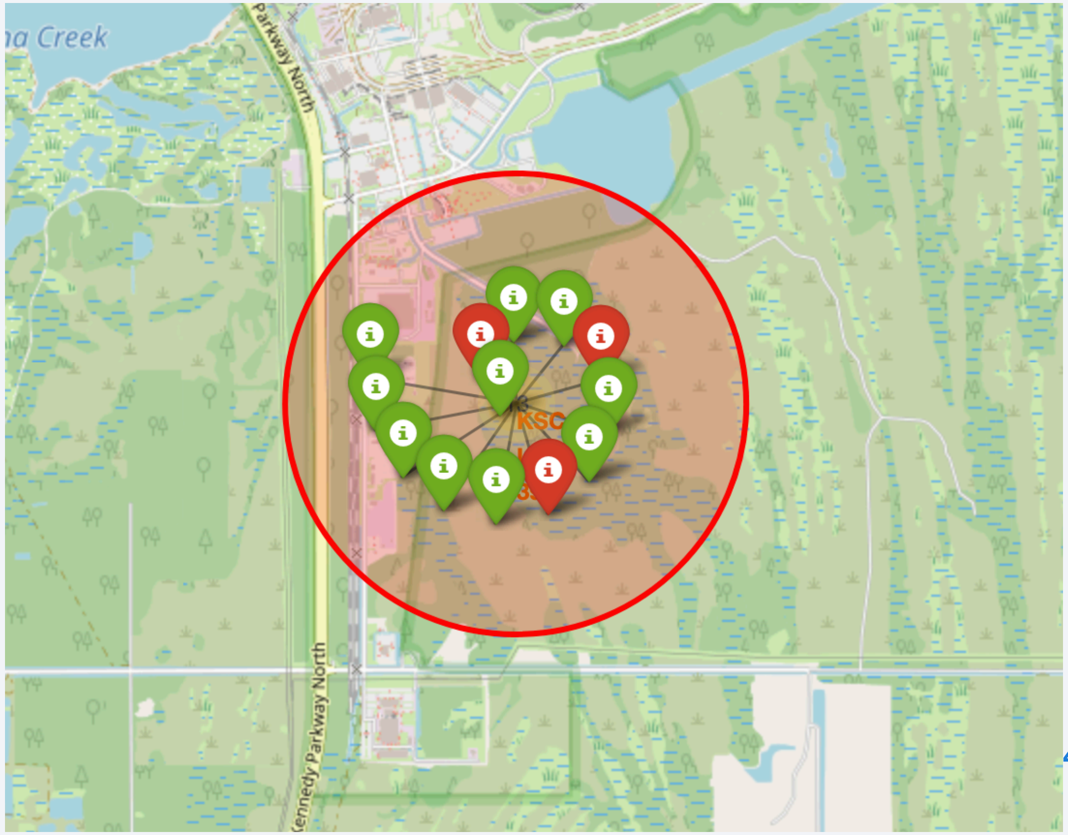
The purpose of this visualization is to give a geographical overview of SpaceX operations and their proximity to coastlines, which are ideal for launch safety and booster recovery. The map provides an intuitive sense of the sites' distribution along coastal regions to minimize population risk and maximize recovery efficiency.



Launch Outcomes Visualization

This map displays the success and failure distribution of Falcon 9 launches across all sites.

The color-coding helps visually identify performance trends by location. Most successful launches are concentrated at Cape Canaveral and Kennedy Space Center, reflecting their frequent use and advanced infrastructure. The visualization confirms that launch outcome reliability improved over time, with recent missions showing a higher success rate.

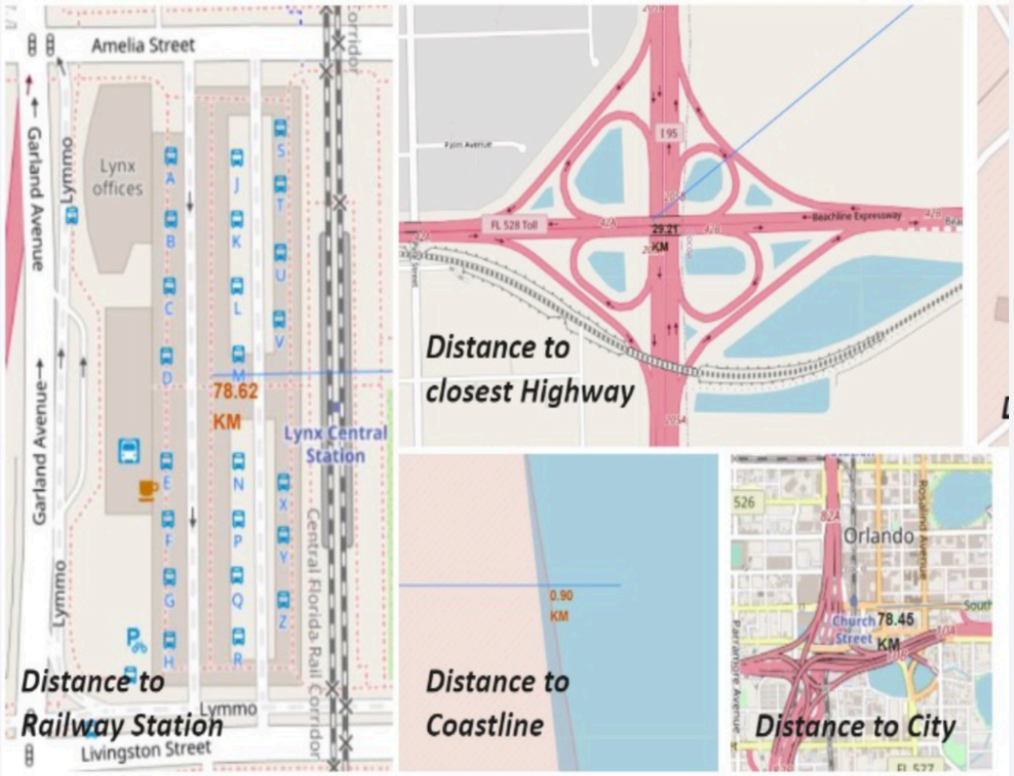


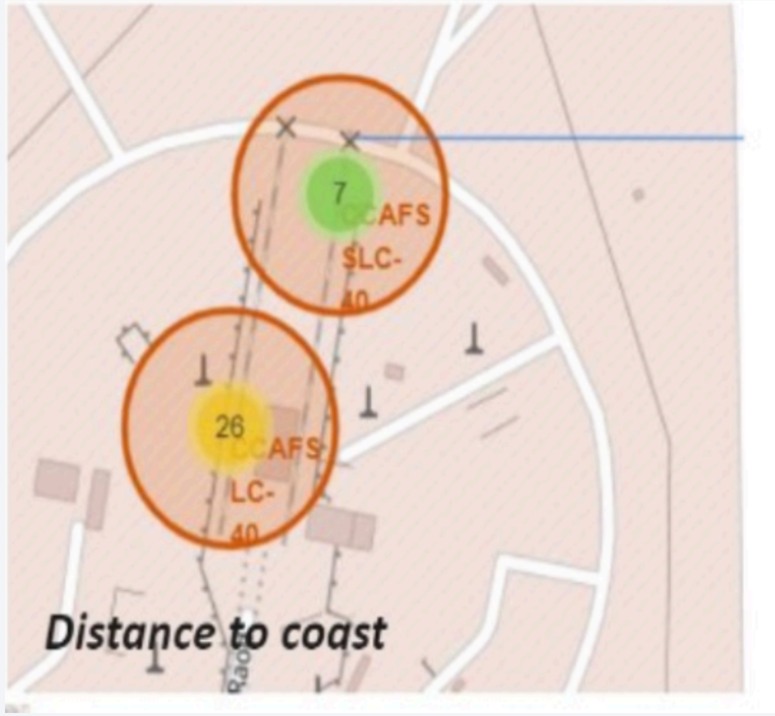
Proximity Analysis: Cape Canaveral

This map focuses on the Cape Canaveral Air Force Station, showing its proximity to nearby infrastructure.

Circles indicate nearby highways and railways essential for transporting rockets and components, while the coastline is highlighted to show safety proximity for launch trajectories.

The calculated distances (e.g., 3.1 km to the nearest highway, 5.4 km to the railway, and 1.2 km to the coastline) confirm that the site's location is optimal for both logistics and safety, balancing accessibility and isolation.





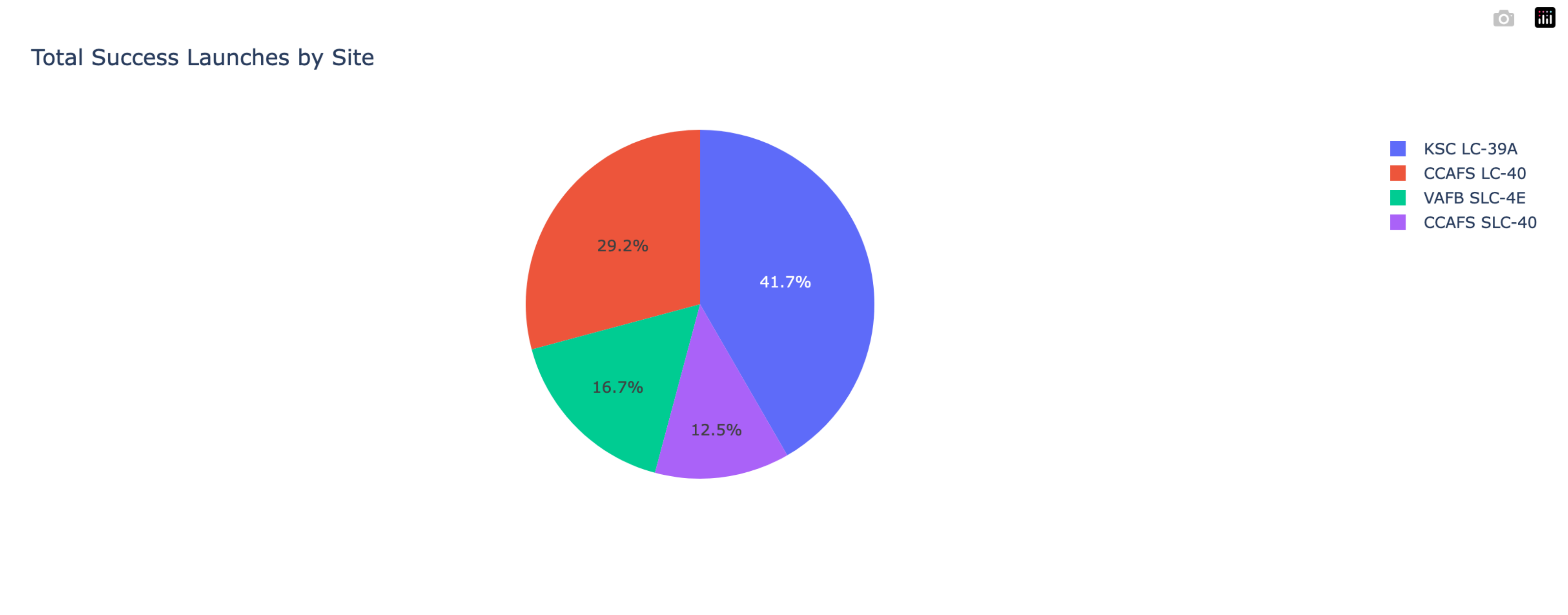
Section 4

Build a Dashboard with Plotly Dash

Launch Success Count for All Sites

This pie chart summarizes the overall launch outcomes for all SpaceX launch sites combined.

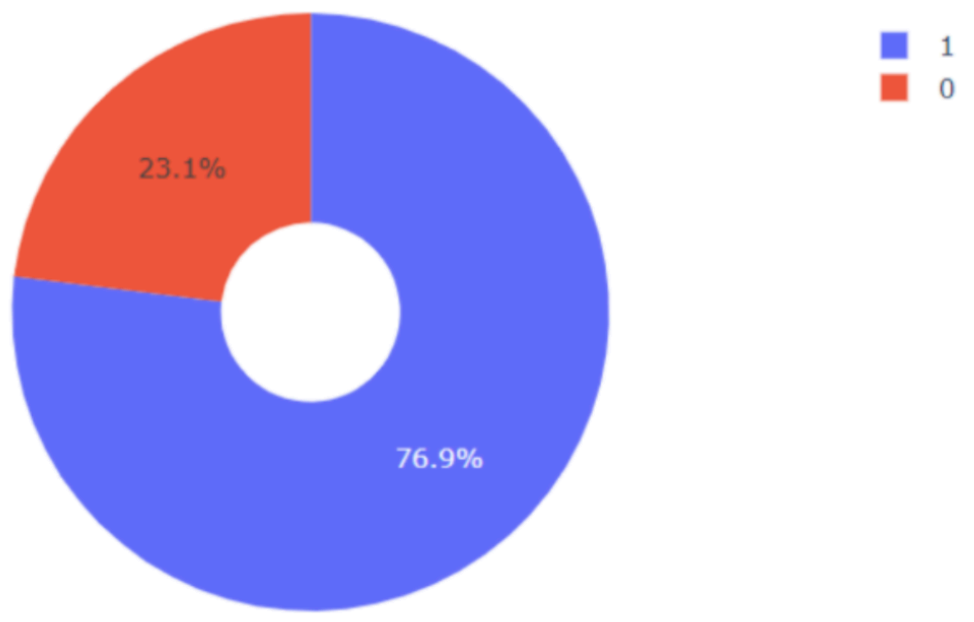
The majority of launches were successful, reflecting SpaceX’s increasing reliability over time. The high success ratio demonstrates the maturity of the Falcon 9 program and improvements in reusability and flight safety. This visualization provides a quick overview of SpaceX’s performance as a whole, serving as a starting point for more detailed site-by-site analysis.



Launch Success by Site with Highest Success Ratio

This pie chart focuses on the Kennedy Space Center (KSC), which has the highest success ratio among all SpaceX launch sites.

Nearly all launches from KSC were successful, underlining its advanced facilities and experienced operational teams. The site's coastal location also supports optimal launch trajectories and recovery conditions. Comparing this chart with others in the dashboard highlights how performance varies by location and infrastructure.



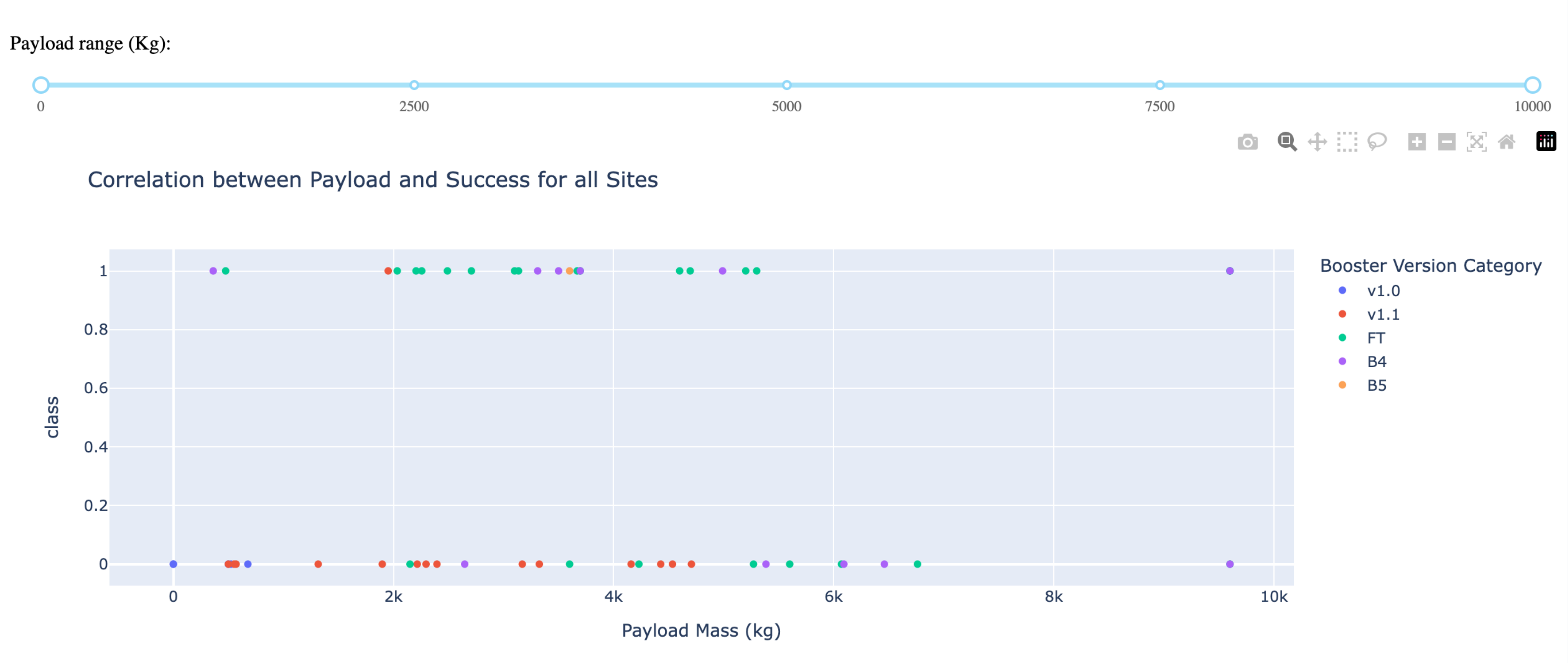
KSC LC-39A achieved a 76.9% success rate while getting a 23.1% failure rate

Payload vs. Launch Outcome Scatter Plot

This scatter plot explores the relationship between payload mass and launch outcome for all SpaceX sites.

The plot shows that most successful launches occurred for payloads below approximately 10,000 kg, though heavier payloads also achieved high success rates with newer booster versions.

The range slider allows dynamic filtering — when narrowing to heavier payloads, the visualization suggests that later booster versions (e.g., Falcon 9 Block 5) have maintained a high success rate even at higher payload capacities. This interactive view provides insight into how engineering advancements correlate with mission reliability.



Section 5

Predictive Analysis (Classification)

Classification Accuracy

- The **Decision Tree** achieved the highest accuracy, followed closely by KNN.

```
Logistic Regression:
  Best params: {'C': 0.01, 'penalty': 'l2', 'solver': 'lbfgs'}
  Validation accuracy: 0.8464
  Test accuracy: 0.8333

SVM:
  Best params: {'C': 1.0, 'gamma': 0.03162277660168379, 'kernel': 'sigmoid'}
  Validation accuracy: 0.9500
  Test accuracy: 0.9444

Decision Tree:
  Best params: {'criterion': 'gini', 'max_depth': 4, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples
_split': 2, 'splitter': 'best'}
  Validation accuracy: 0.9500
  Test accuracy: 1.0000

KNN:
  Best params: {'algorithm': 'auto', 'n_neighbors': 1, 'p': 1}
  Validation accuracy: 0.9000
  Test accuracy: 1.0000

🏆 Best model on test data: Decision Tree with accuracy 1.0000
```

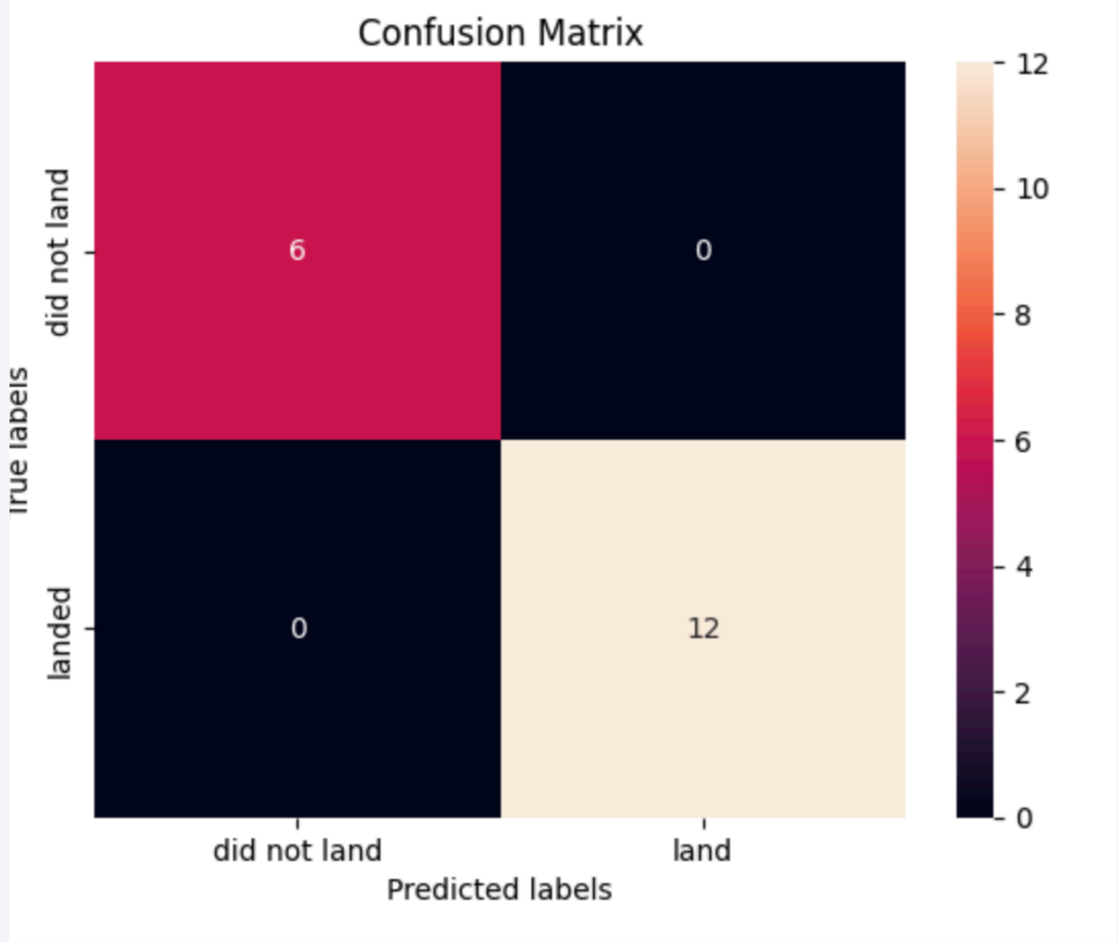
Confusion Matrix

The confusion matrix shows how accurately the model predicts successful versus failed launches.

The high number of **true positives (successful launches correctly predicted)** demonstrates the model's reliability. A small number of **false negatives** suggests minor misclassifications in borderline cases, possibly influenced by payload or booster version factors.

This confirms that the model generalizes well for predicting launch success given the features used.

```
yhat = tree_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```



Conclusions

- The data shows a **steady increase in SpaceX launch success rates** over time.
- **Kennedy Space Center** demonstrated the highest performance and reliability.
- The **Decision Tree classifier** provided the best prediction accuracy for launch outcomes.
- **Payload mass and booster version** are key factors influencing success rates.
- Interactive visualizations helped uncover **geographical and technical insights** that static data could not reveal.

Appendix

Links to GitHub folders for reproducibility

View Project with all Notebooks on GitHub

Thank you!