IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Hoo Yan Peng
20th October 2024

Outline

Executive Summary

Methodologies:

1. Data Collection: Used SpaceX API and web scraping for launch data.
2. Data Wrangling: Cleaned data, addressed missing values, and aggregated metrics.
3. Exploratory Data Analysis (EDA): Utilized visualizations and SQL queries for insights.
4. Interactive Analytics: Implemented Folium and Plotly Dash for maps and dashboards.
5. Predictive Analysis: Built and evaluated classification models.

Results:

1. EDA Insights: Increased launch activity; improved success rates.
2. Interactive Analytics: Insights on launch distributions and success trends.
3. Model Performance: SVM achieved highest accuracy (83.33%).

Introduction

Project Background and Context

SpaceX has revolutionized the aerospace industry with its Falcon 9 rocket, offering competitive launch costs—$62 million compared to over $165 million from other providers—primarily due to the reusability of its first stage. Significant milestones include being the first private company to return a spacecraft from low Earth orbit in December 2010. This innovation has made space access more economical and attracted global attention.

Analysis Objectives

This analysis aims to address key challenges in predicting the successful landing of the Falcon 9's first stage by examining critical factors such as payload mass, orbit type, and launch site characteristics.

Our goals are to:

- Predict Landing Success: Assess factors affecting successful landings.
- Identify Critical Factors: Analyze influences on launch outcomes.
- Provide Insights: Offer actionable insights for competitors bidding against SpaceX

Section 1

Methodology

Methodology

Executive Summary

1. Data Collection:
    - via SpaceX API and Web Scraping.

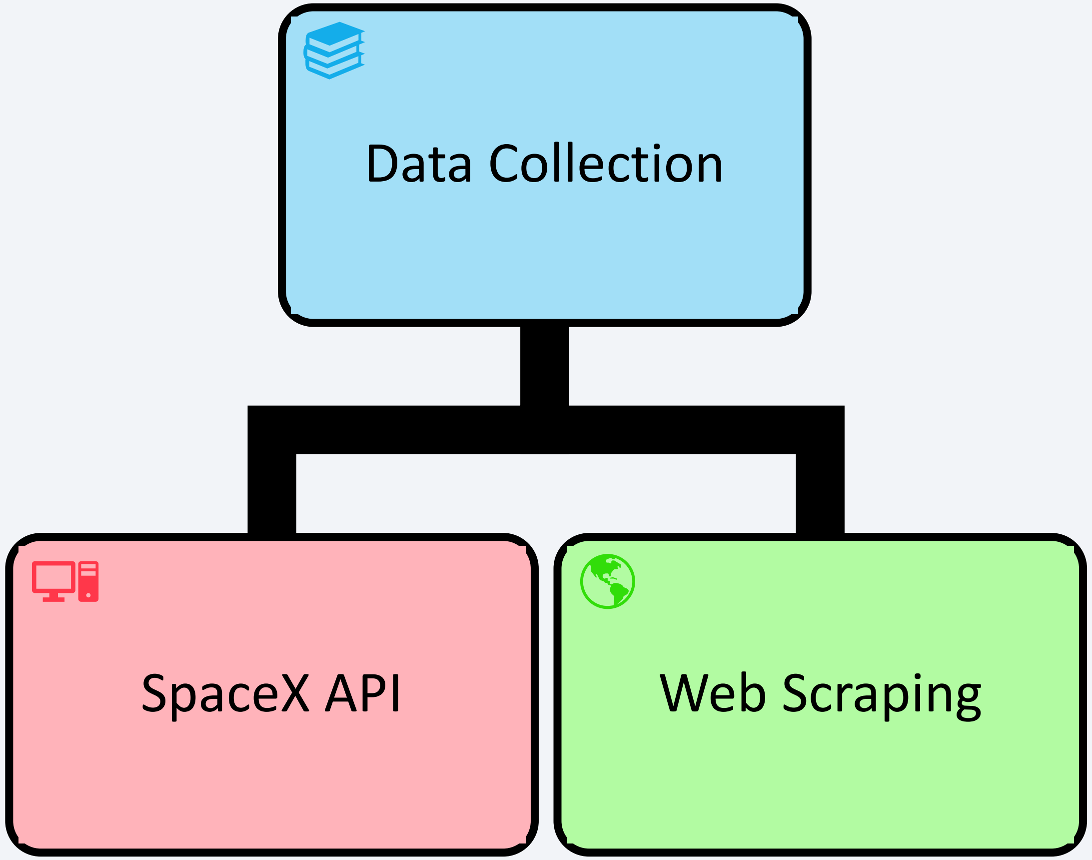
2. Data Wrangling
    - Cleaned and transformed data, addressed missing values, and aggregated key metrics.

3. Exploratory Data Analysis (EDA)
    - Used visualizations and SQL queries for insights.

4. Interactive Visual Analytics:
    - Implemented Folium for maps and Plotly Dash for interactive dashboards.

5. Predictive Analysis:
    - Built and tuned classification models; evaluated using accuracy and confusion matrices.

Data Collection

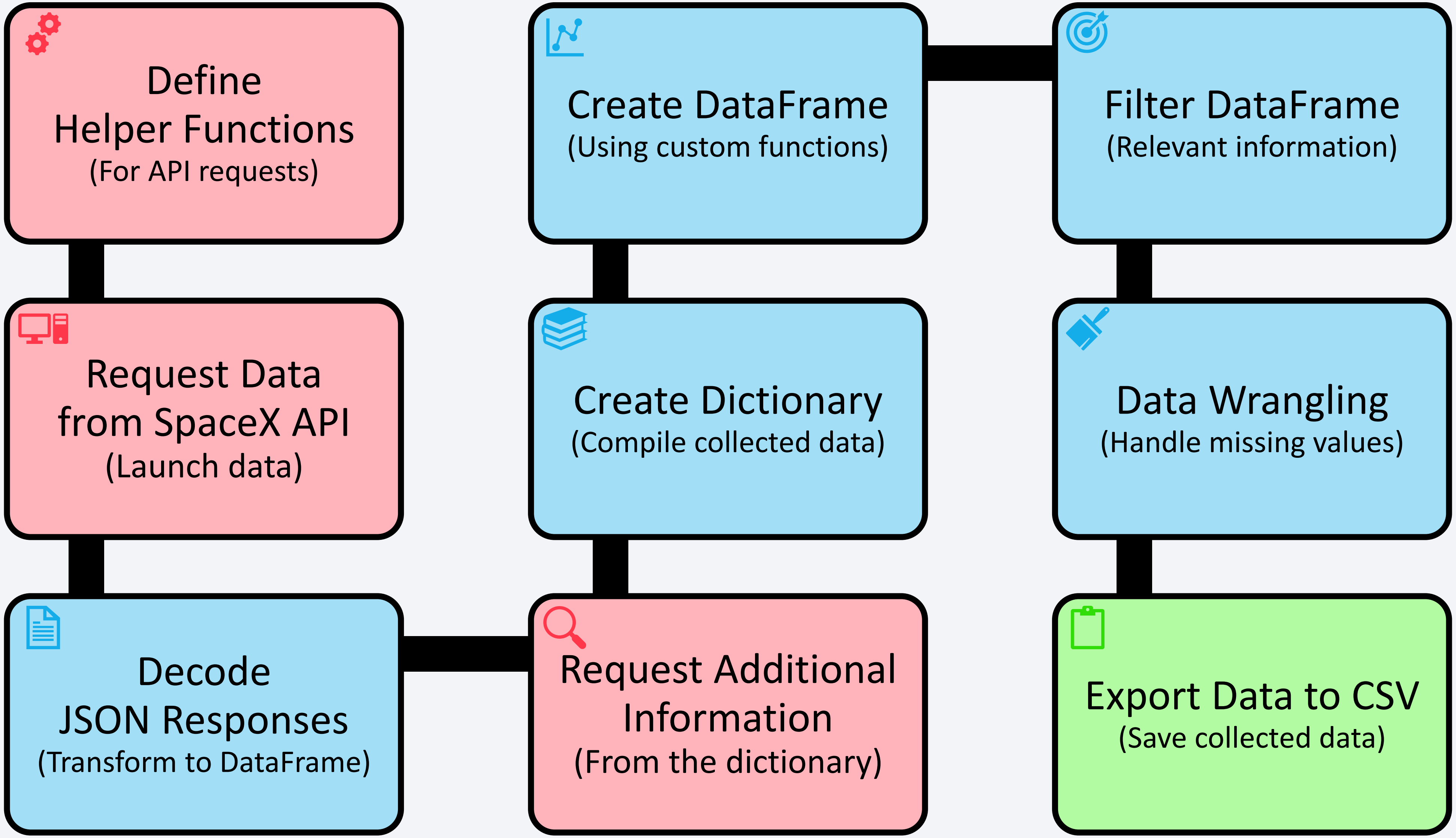


1. [SpaceX API]
   - **Efficient Retrieval:** Used the SpaceX API to fetch structured launch data.
   - **Data Processing:** Decoded responses and organized into a DataFrame for analysis.

2. [Web Scraping]
   - **Data Extraction:** Employed BeautifulSoup to scrape historical launch records from Wikipedia.
   - **Transformation:** Converted raw HTML tables into a structured format for insights.

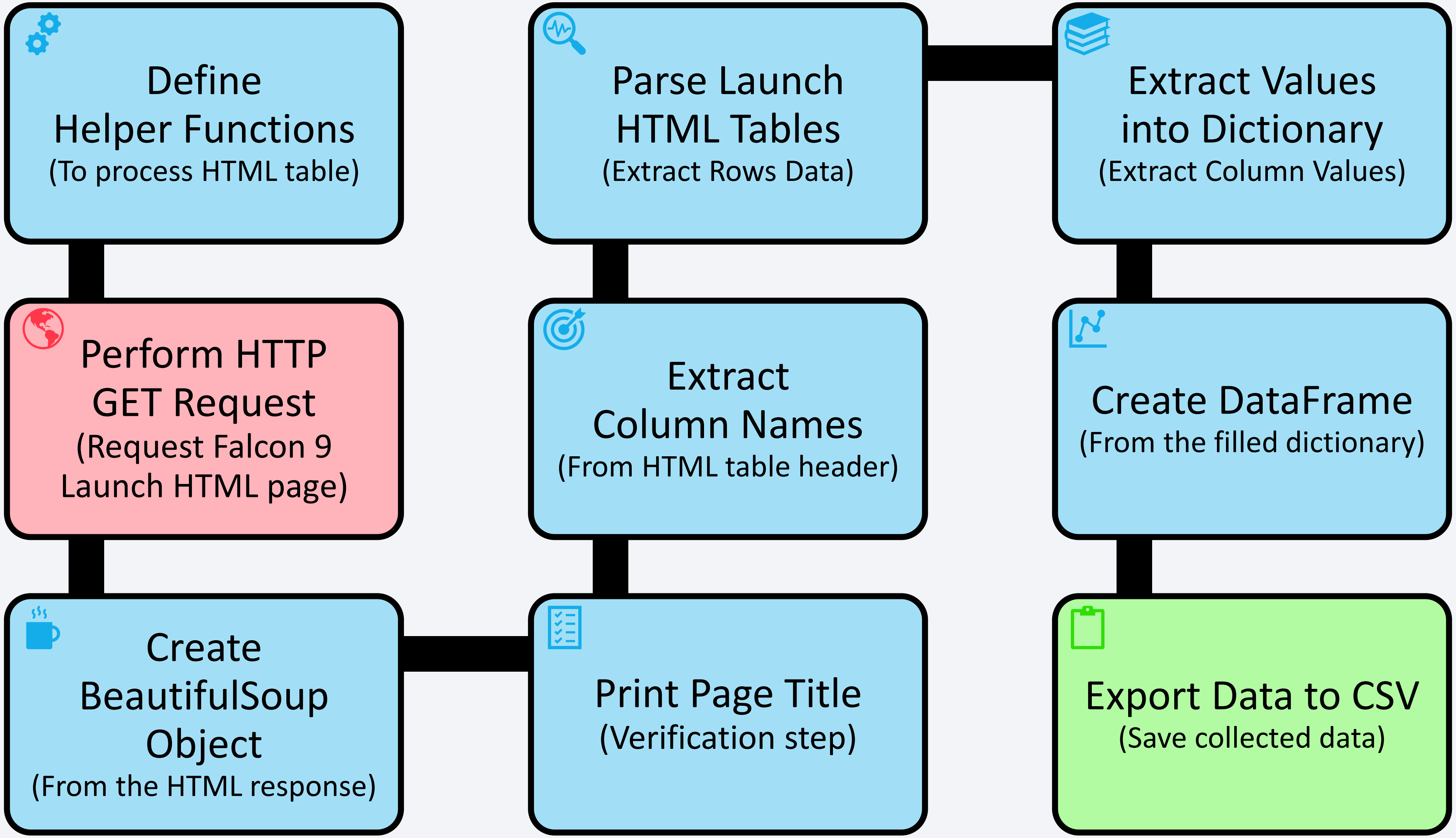
Data Collection – SpaceX API

[GitHub URL]



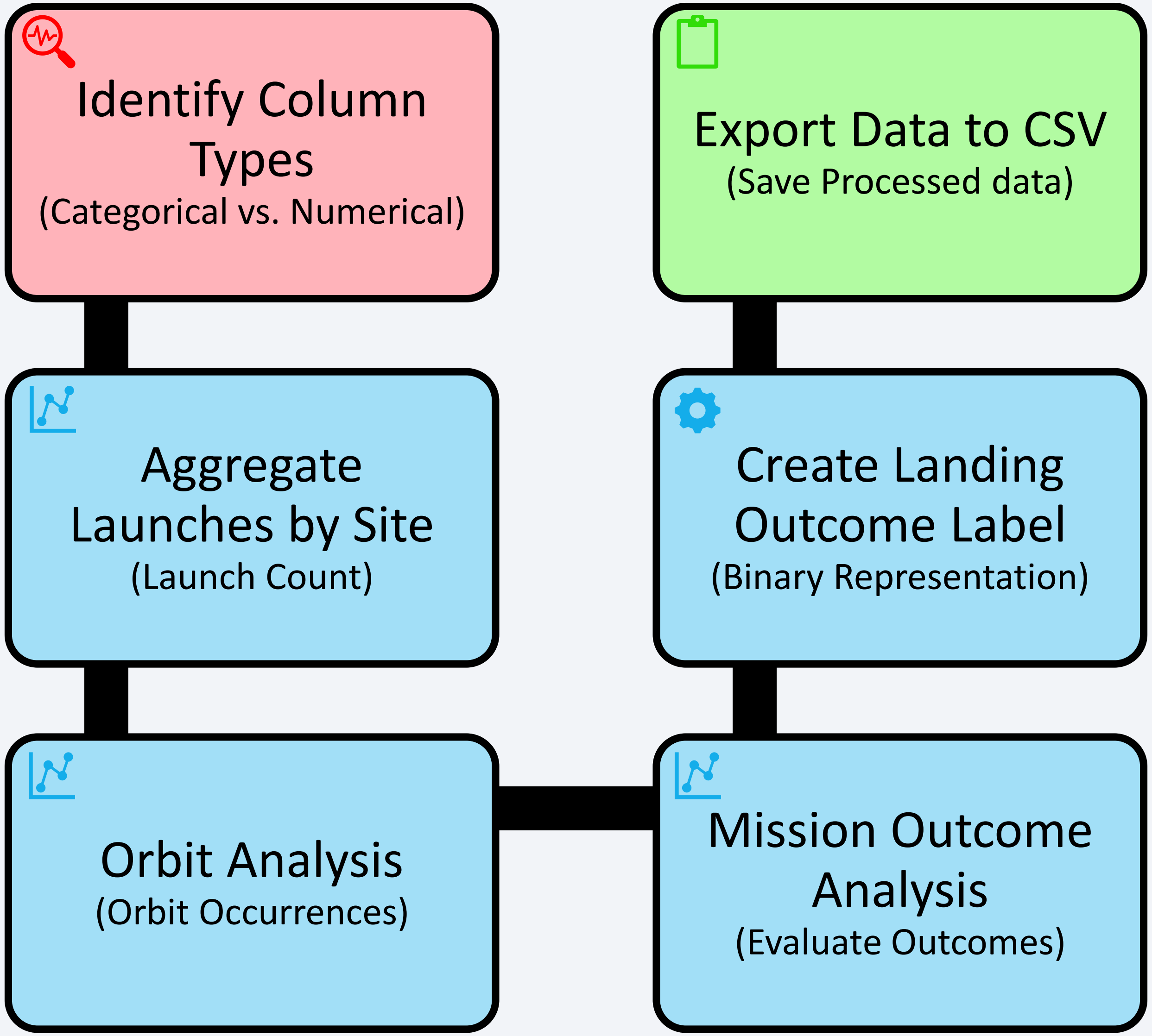
Data Collection - Scraping

[GitHub URL]



Data Wrangling

GitHub URL



EDA with Data Visualization

GitHub URL

1. **Scatterplot**:

   Used to visualize the relationship between two continuous variables and identify patterns or trends within the data.

   - Flight Number vs. Launch Site
   - Payload Mass vs. Launch Site
   - Flight Number vs. Orbit Type
   - Payload Mass vs. Orbit Type

2. **Bar Chart**:

   Used to compare categorical data and visualize differences in quantities across categories.

   - Success Rate by Orbit Type

3. **Line Chart**:

   Used to show trends over time by connecting data points with a continuous line.

   - Launch Success Trend Over Years

EDA with SQL

GitHub URL

Summary of SQL Queries:

- Display unique launch sites.
- Display 5 records of launch sites starting with 'CCA'.
- Calculate total payload mass for NASA (CRS) missions.
- Find average payload mass for F9 v1.1 boosters.
- List the date of the first successful ground pad landing.
- Identify boosters with successful drone ship landings (4000 < mass < 6000).
- Count total successful and failed mission outcomes.
- List booster versions with maximum payload (using subquery).
- Show month names and failed drone ship landings for 2015.
- Rank landing outcomes between 2010-06-04 and 2017-03-20.

Build an Interactive Map with Folium

GitHub URL

Summary of Map Objects

1. Highlighted Circle:
   - folium.Circle at NASA Johnson Space Center.

2. Launch Sites:
   - folium.Circle and folium.Marker to mark launch sites and explore proximity to the Equator and coastline.

3. Launch Outcomes:
   - Color-coded markers (green for success, red for failure) to show success rates with clusters for overlapping coordinates.

4. Mouse Position:
   - MousePosition to display coordinates for proximity measurements.

5. Distance Lines:
   - PolyLine connecting launch sites to nearby features (coastline, city, railway, highway).

Build a Dashboard with Plotly Dash

GitHub URL

Summary of Plots and Interactions

1. Launch Site Drop-down:
   - Allows users to select a launch site, with options for all sites or specific locations, enabling focused analysis.

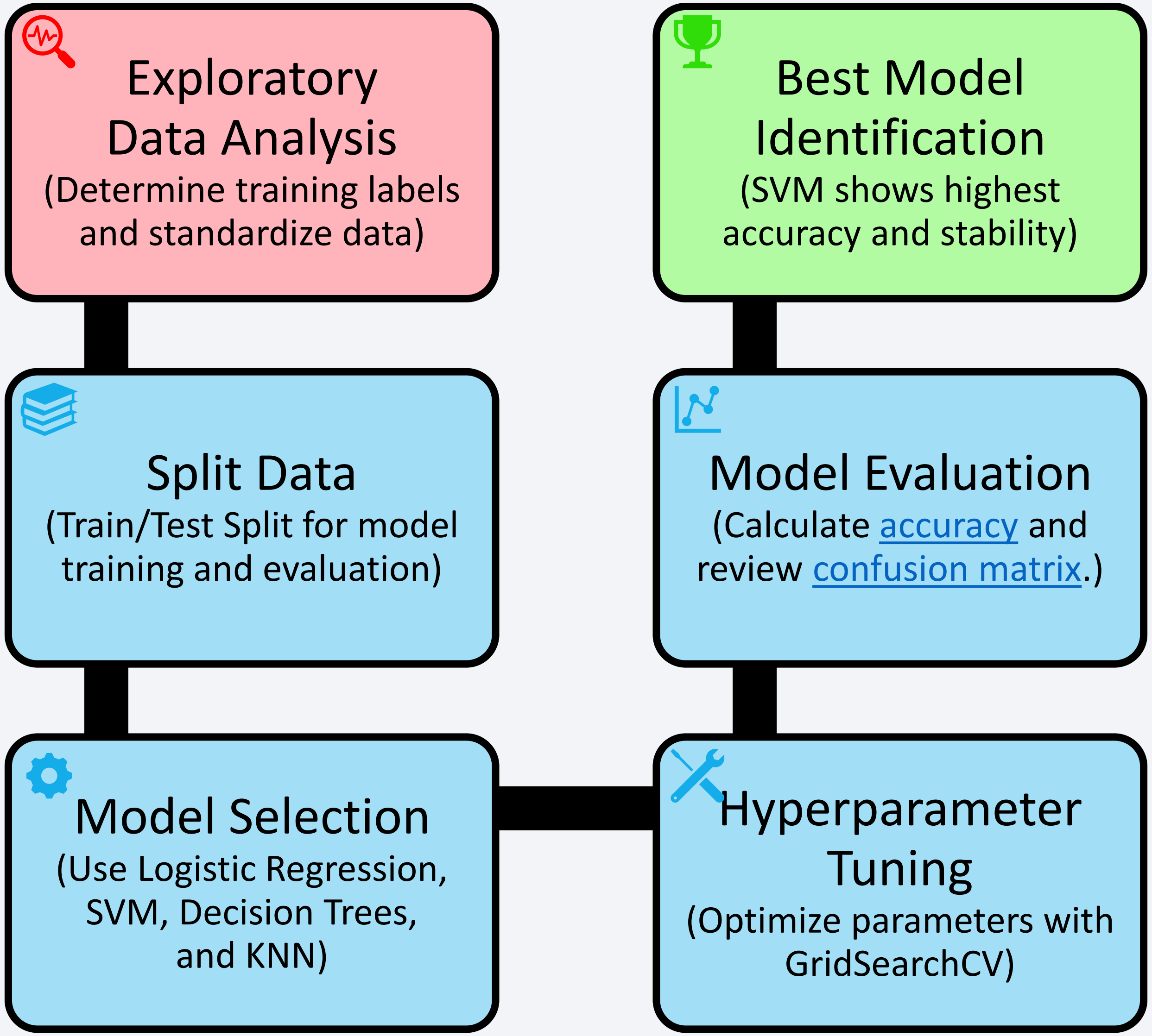
2. Success Pie Chart:
   - Displays total successful launches by site; updates based on the selected site to compare success vs. failure.

3. Payload Range Slider:
   - Enables users to filter data based on payload mass, refining the analysis of launch success.

4. Success-Payload Scatter Chart:
   - Visualizes correlation between payload mass and launch success; updates based on selected site and payload range.

Predictive Analysis (Classification)

GitHub URL



Results

1. **Exploratory Data Analysis (EDA)**
    - Launch Activity: Most Falcon 9 launches at CCAFS SLC-40; increased activity from other sites after Flight 25.
    - Success Rates: Improved reliability over time.
    - Payload Management: CCAFS SLC-40 and KSC LC-39A handle diverse payloads; VAFB SLC-4E focuses on lighter missions.
    - Orbit Success: ES-L1, GEO, HEO, SSO orbits at 100% success; SO orbit at 0%.
    - Drone Ship Landings: F9 FT boosters (payload mass 4,000-6,000 kg) show high success rates, demonstrating versatility.

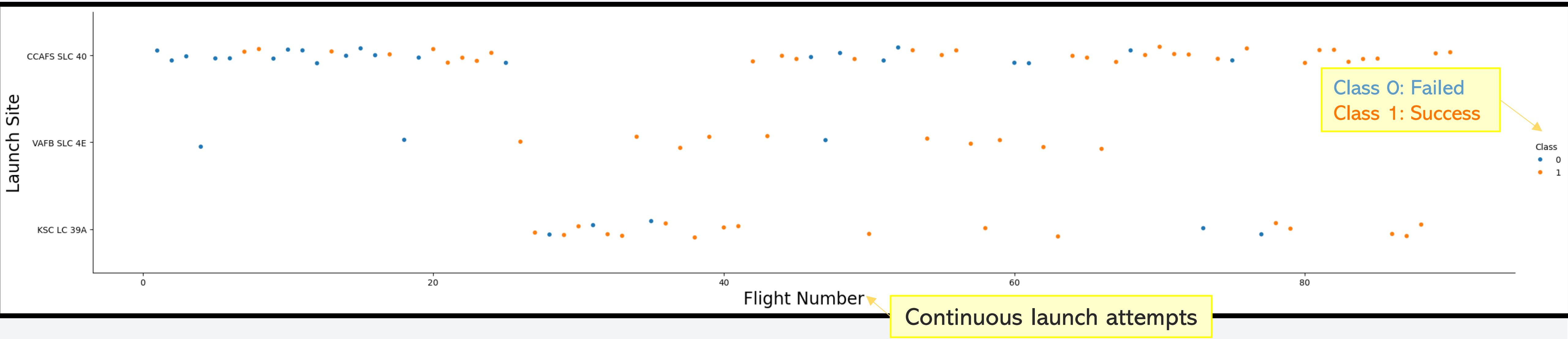
2. **Interactive Analytics**
    - Insights: Geographic launch site distributions and success trends.

3. **Predictive Analysis Results**
    - Model Accuracy: Logistic Regression, SVM, KNN at 83.33%; Decision Tree at 78.89%.
    - Top Model: SVM shows highest performance and stability.

Section 2

Insights drawn from EDA

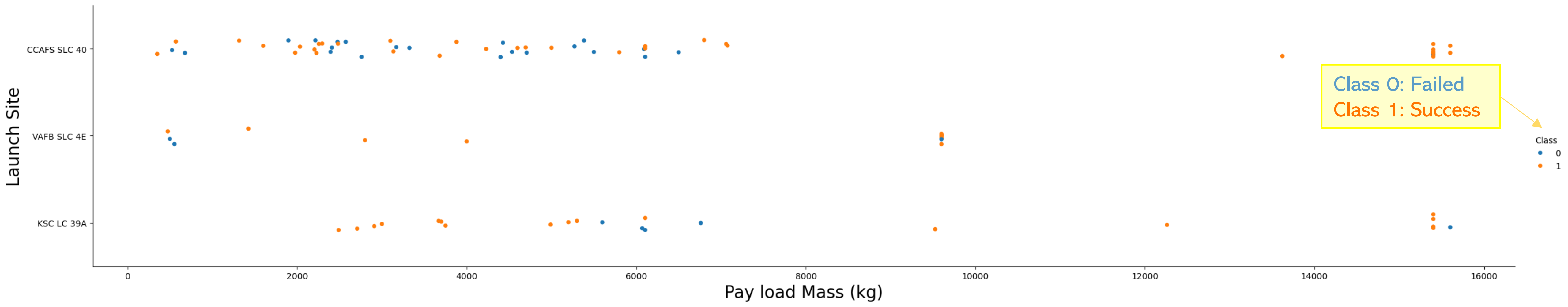
Flight Number vs. Launch Site



Explanation:

- Most Falcon 9 launches occurred at CCAFS SLC 40, with significant activity at this site. After Flight Number 25, other sites became more active.
- Success (Class 1) increases with Flight Number, indicating that early attempts showed variability, while later launches reflect consistent improvement due to ongoing testing.

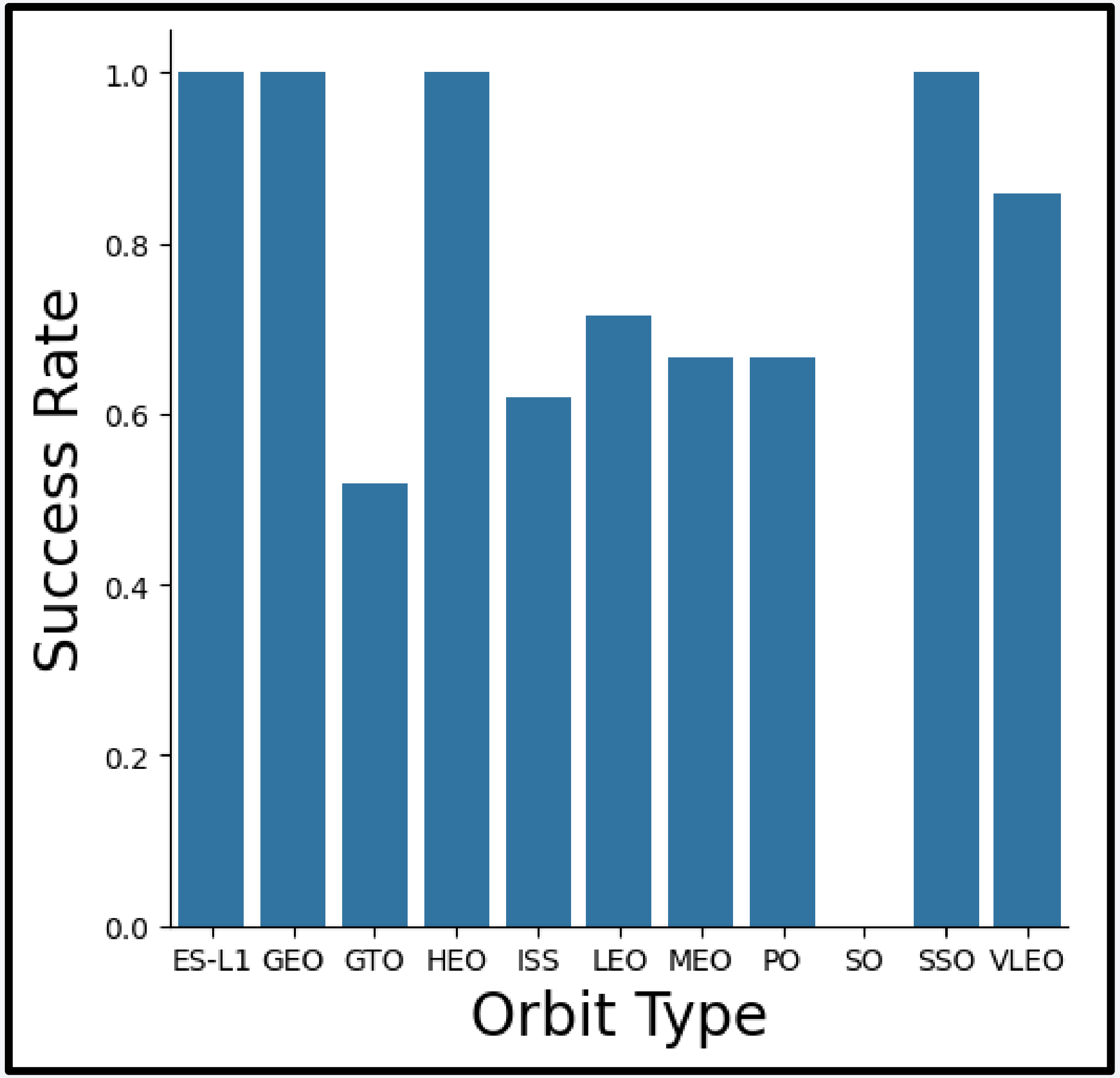
Payload vs. Launch Site



Explanation:

- CCAFS SLC 40 & KSC LC 39A supporting a broader range of payloads, effectively managing both lighter and heavier missions.
- VAFB SLC 4E has no rockets launched for heavy payload masses (greater than 10,000 kg), indicating a focus on lighter mission profiles.

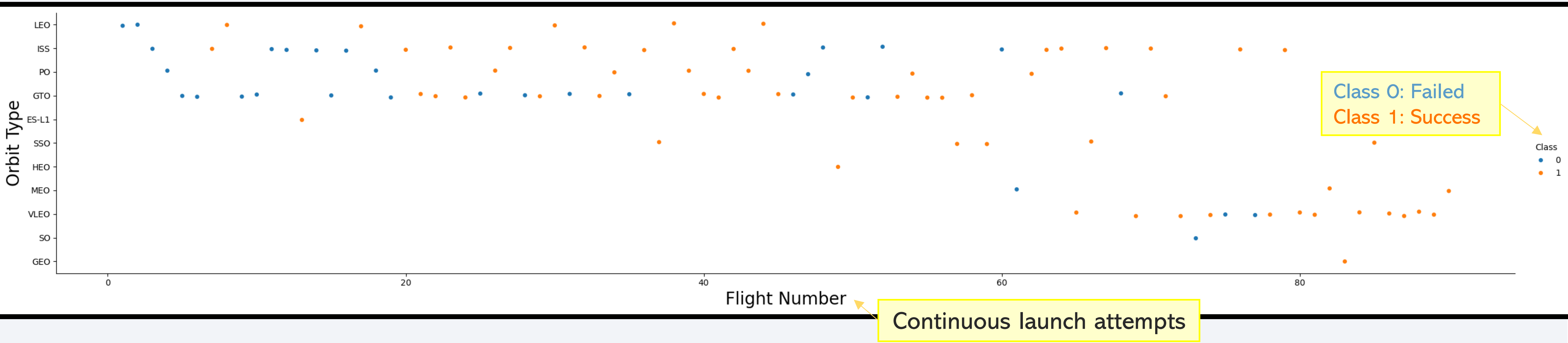
Success Rate vs. Orbit Type



Explanation:

- Orbits like ES-L1, GEO, HEO, and SSO have perfect success rates (100%), reflecting strong reliability.
- The SO orbit shows a 0% success rate, indicating significant issues that require further investigation.
- Developing tailored strategies is crucial for improving performance in orbits with lower success rates.

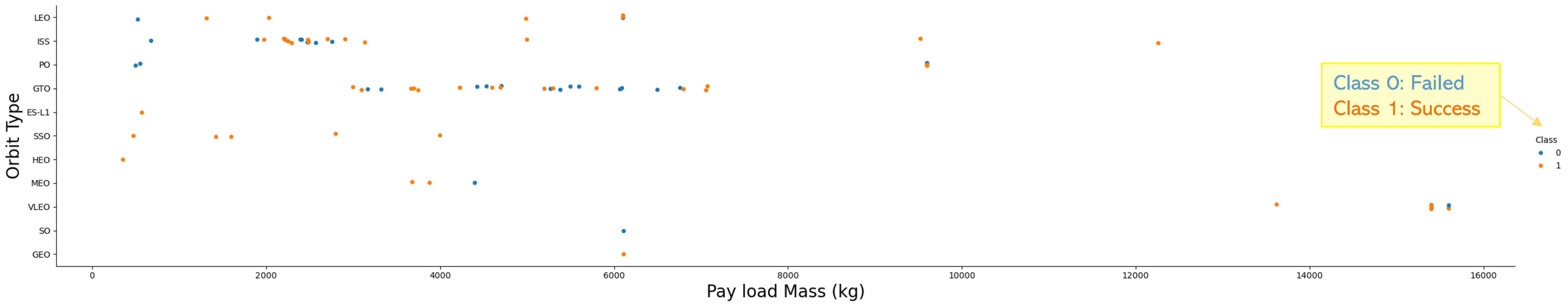
Flight Number vs. Orbit Type



Explanation:

- In the LEO orbit, success seems to be related to the number of flights.
- Conversely, in the GTO orbit, there appears to be no relationship between flight number and success.

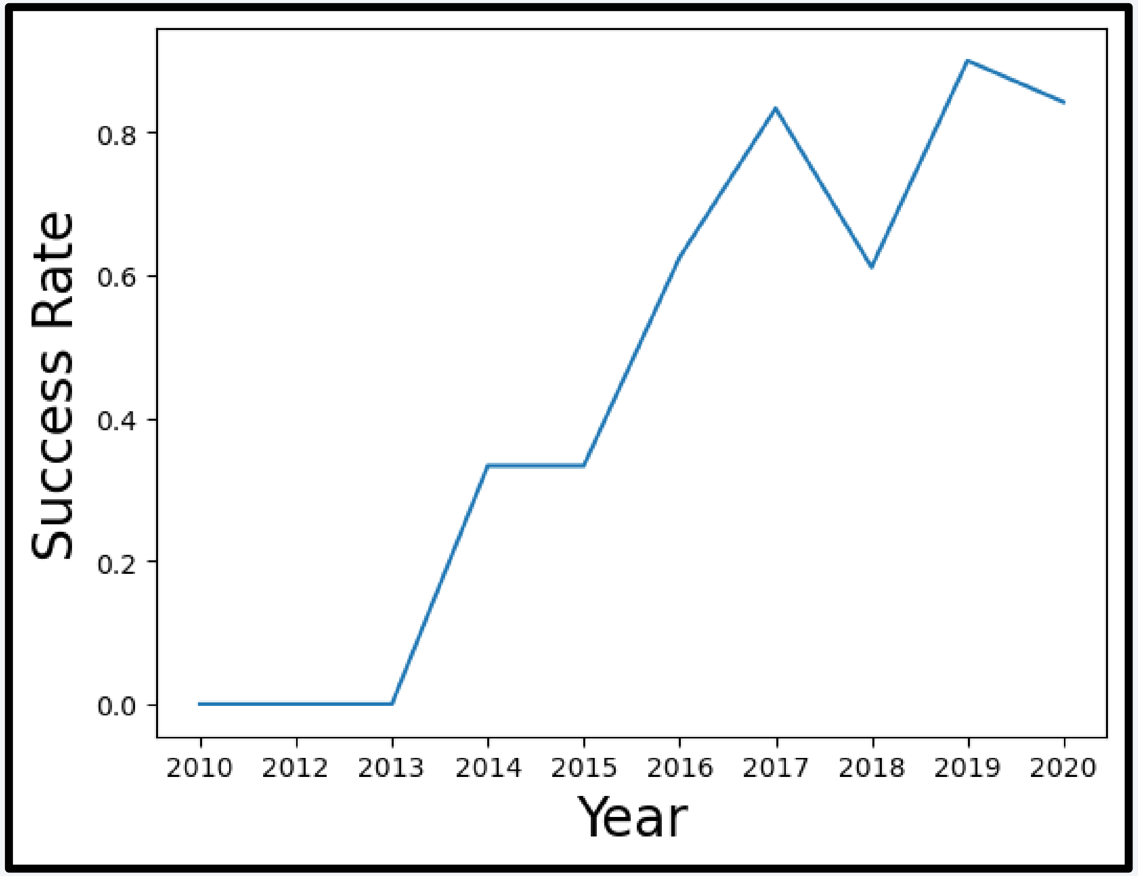
Payload vs. Orbit Type



Explanation:

- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.
- However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.

Launch Success Yearly Trend



Explanation:

- Since 2013, the success rate kept increasing till 2020.
- A significant rise begins in 2015, with rates exceeding 80% by 2017.
- This trend reflects a positive trajectory in launch success rates from 2013 to 2020, showcasing advancements in technology and mission execution.

All Launch Site Names

Unique Launch Sites Identified:

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

Explanation:

- The analysis highlights the geographical diversity of launch sites utilized for space missions.
- Notably, CCAFS LC-40 and CCAFS SLC-40 are different designations for the same site, indicating its versatility for various mission types.

Launch Site Names Begin with 'CCA'

Five records where launch sites begin with `CCA`:

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

Explanation:

- All five records feature the F9 v1.0 booster version, which was exclusively launched from CCAFS LC-40.
- This booster version is specifically focused on LEO (Low Earth Orbit) missions.

Total Payload Mass

Total payload mass carried by boosters launched by NASA (CRS):

total_payload_mass
48213

Explanation:

- This total underscores the crucial role of SpaceX in transporting supplies, equipment, and scientific payloads to the International Space Station (ISS).

Average Payload Mass by F9 v1.1

Average payload mass carried by the Falcon 9 v1.1 booster version:

average_payload_mass
2534.6666666666665

Explanation:

- This figure is derived from multiple launches using the F9 v1.1 variant, highlighting its capability to support a range of payloads effectively.

First Successful Ground Landing Date

Date of the first successful landing outcome on ground pad:

first_succesful_date
2015-12-22

Explanation:

- This milestone marked a significant advancement in SpaceX's landing technology, demonstrating the feasibility of reusing rocket components and enhancing cost efficiency in space missions.

Successful Drone Ship Landing with Payload between 4000 and 6000

Boosters with Successful Drone Ship Landings (Payload Mass: 4000 - 6000 kg):

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

Explanation:

- This demonstrates the versatility and reliability of the Falcon 9 FT boosters in successfully managing a range of payload capacities during maritime recovery operations.

Total Number of Successful and Failure Mission Outcomes

Total number of successful and failure mission outcome:

Mission_Outcome	total_number
Failure	1
Success	100

Explanation:

- This highlights the reliability and effectiveness of SpaceX's launch capabilities, with a failure rate of only about 1%.

Boosters Carried Maximum Payload

Booster which have carried the maximum payload mass:

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

Explanation:

- This demonstrates the capability of the Falcon 9 B5 boosters to handle substantial payloads, enhancing their versatility for various missions.

2015 Launch Records

Failure Landing Outcomes on Drone Ships in 2015:

Month	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

Explanation:

- In 2015, both recorded failures of drone ship landings involved the Falcon 9 v1.1 variant.
- This pattern underscores the challenges faced by this booster during maritime recovery operations, highlighting areas for improvement in SpaceX's landing technology.

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

Count of landing outcome:

Landing_Outcome	count(Landing_Outcome)
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

Explanation:

- The ranking of landing outcomes between June 4, 2010, and March 20, 2017, shows that "No attempt" was the most common outcome, with 10 occurrences.
- This was followed by "Success (drone ship)" and "Failure (drone ship)," each with 5 counts.
- The data illustrates the evolving nature of SpaceX's recovery operations, indicating a mix of successes and challenges during this period.

Section 3

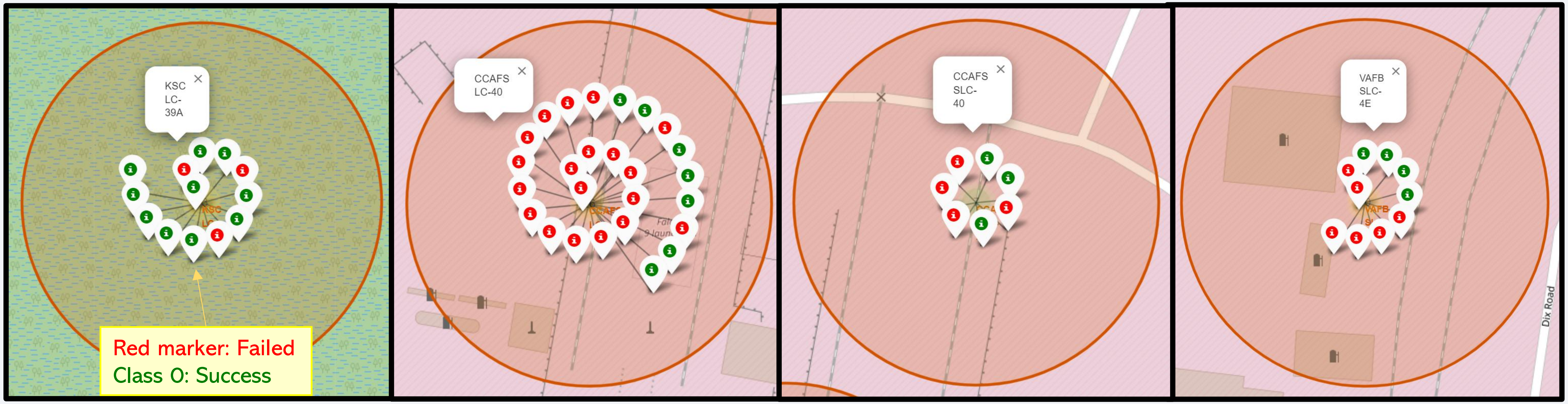
Launch Sites Proximities Analysis

Launch Site Analysis

Findings:

- The launch sites are relatively close to the Equator, allowing them to better utilize Earth's rotational speed, reducing fuel needs and increasing payload capacity.
- All sites are near the coast, ensuring safer launches by directing trajectories over the ocean, minimizing risks to populated areas.

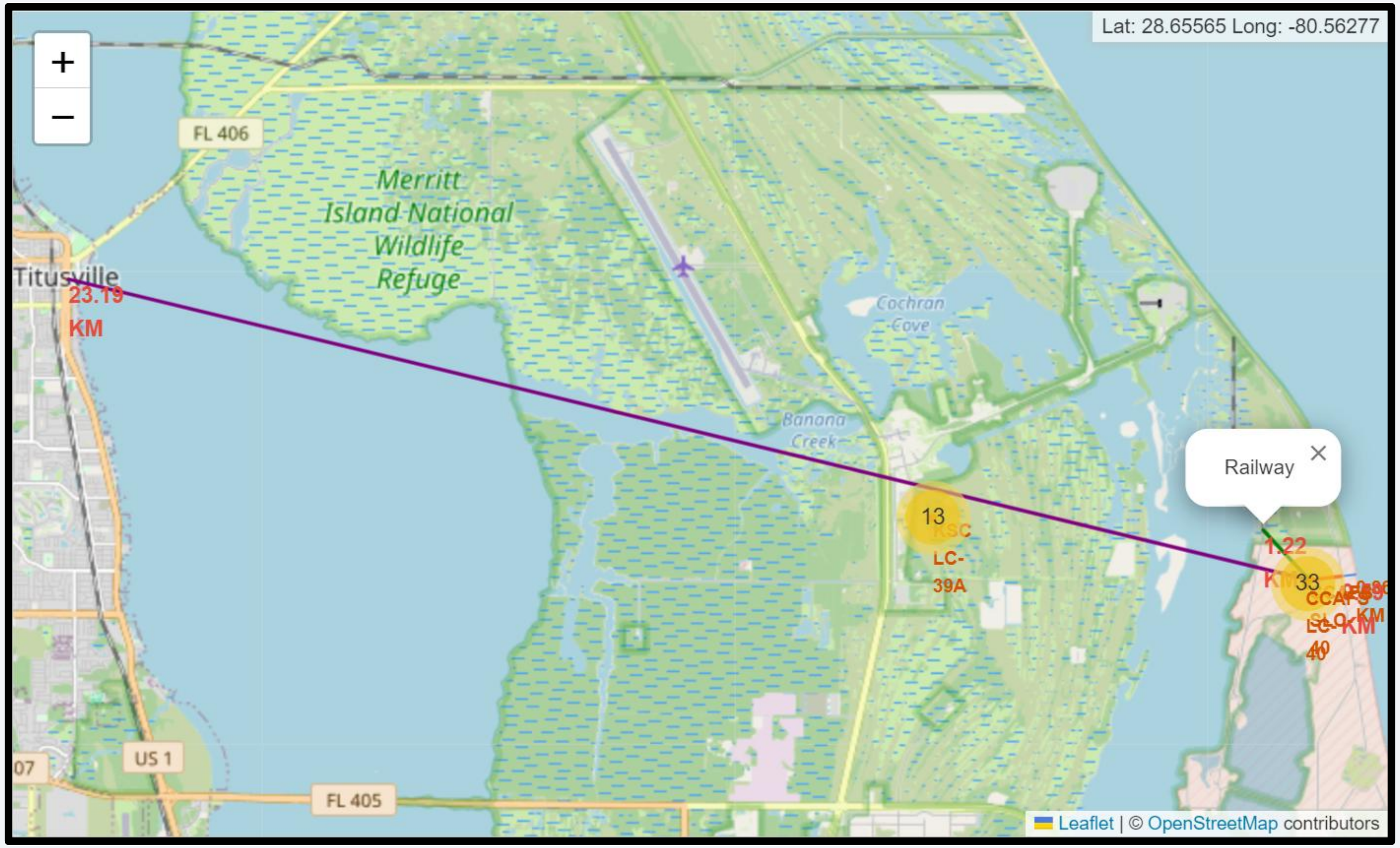
Launch Success and Failure Rates by Site



Findings:

- KSC LC-39A is the most reliable site with the highest success rate.
- CCAFS LC-40 has a high failure rate that needs further investigation.
- CCAFS SLC-40 and VAFB SLC-4E require monitoring for consistent performance.

Proximity Analysis of CCAFS SLC-40 Launch Site



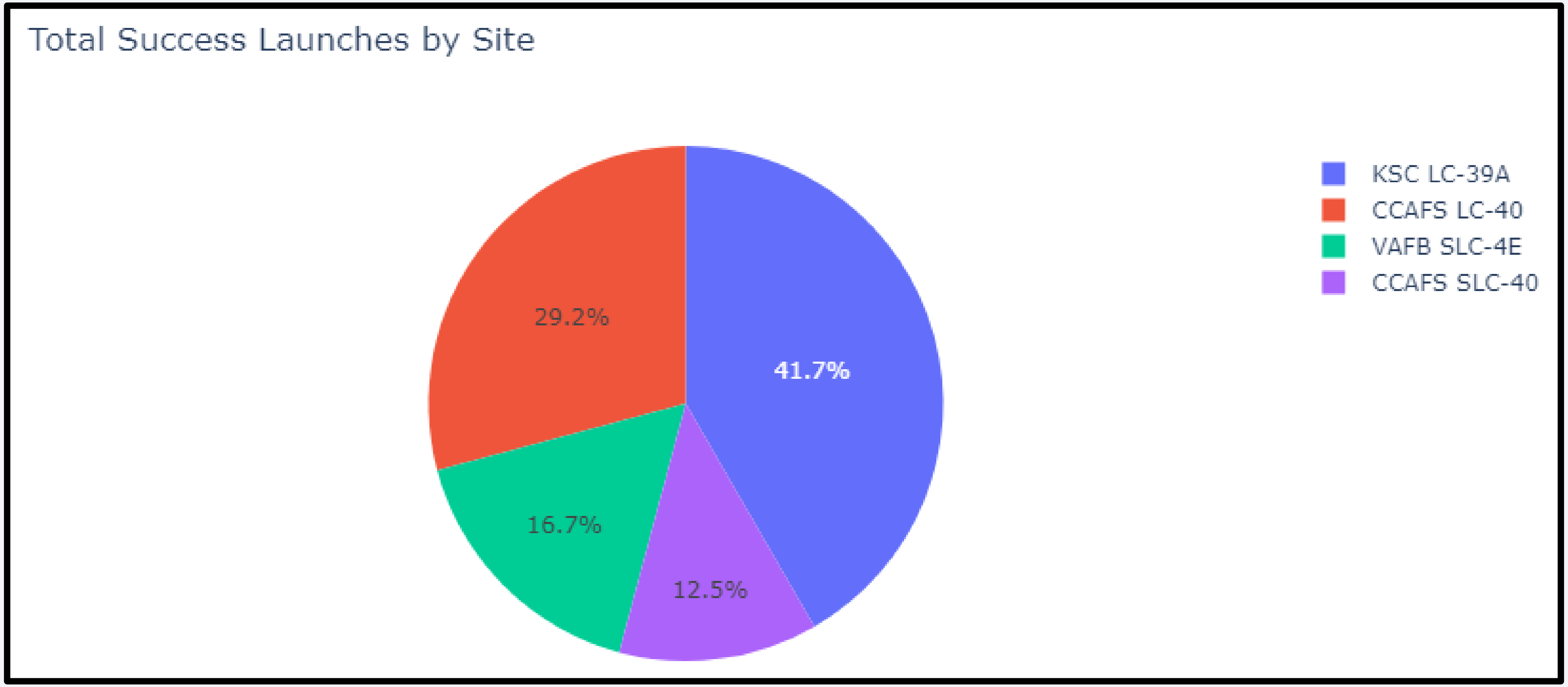
Findings:

- Close distance to Railway (1.22 km) and Highway (0.59 km) – Provides good access for logistical support and facilitates efficient transport of materials and personnel to the site.
- Close distance to Coastline (0.86 km) – Ensures safe launch trajectories over water.
- Far distance to City (23.19 km) – Maintains a safe distance from urban populations, enhancing public safety.

Section 4

Build a Dashboard with Plotly Dash

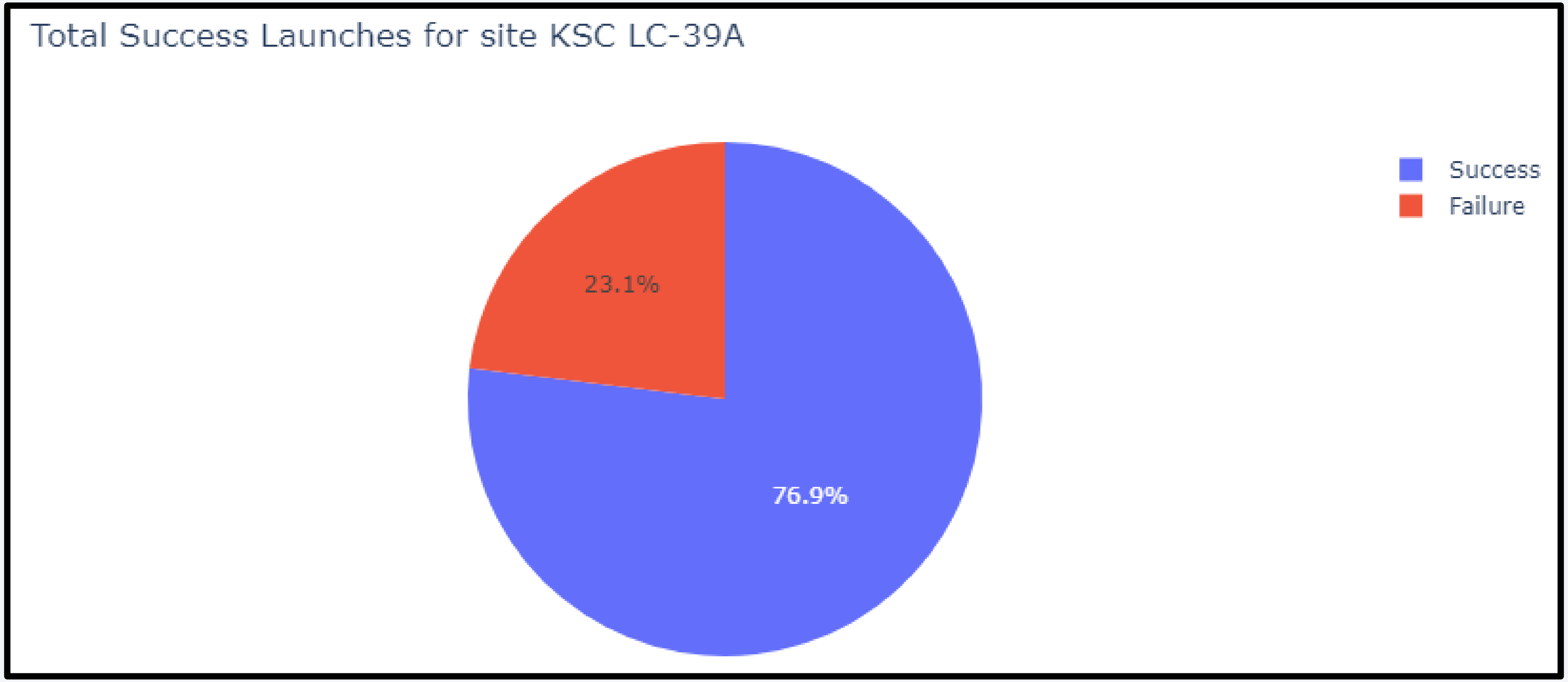
Launch Success Count for All Sites



Findings:

- KSC LC-39A accounts for the largest share of successful launches at 41.7%.

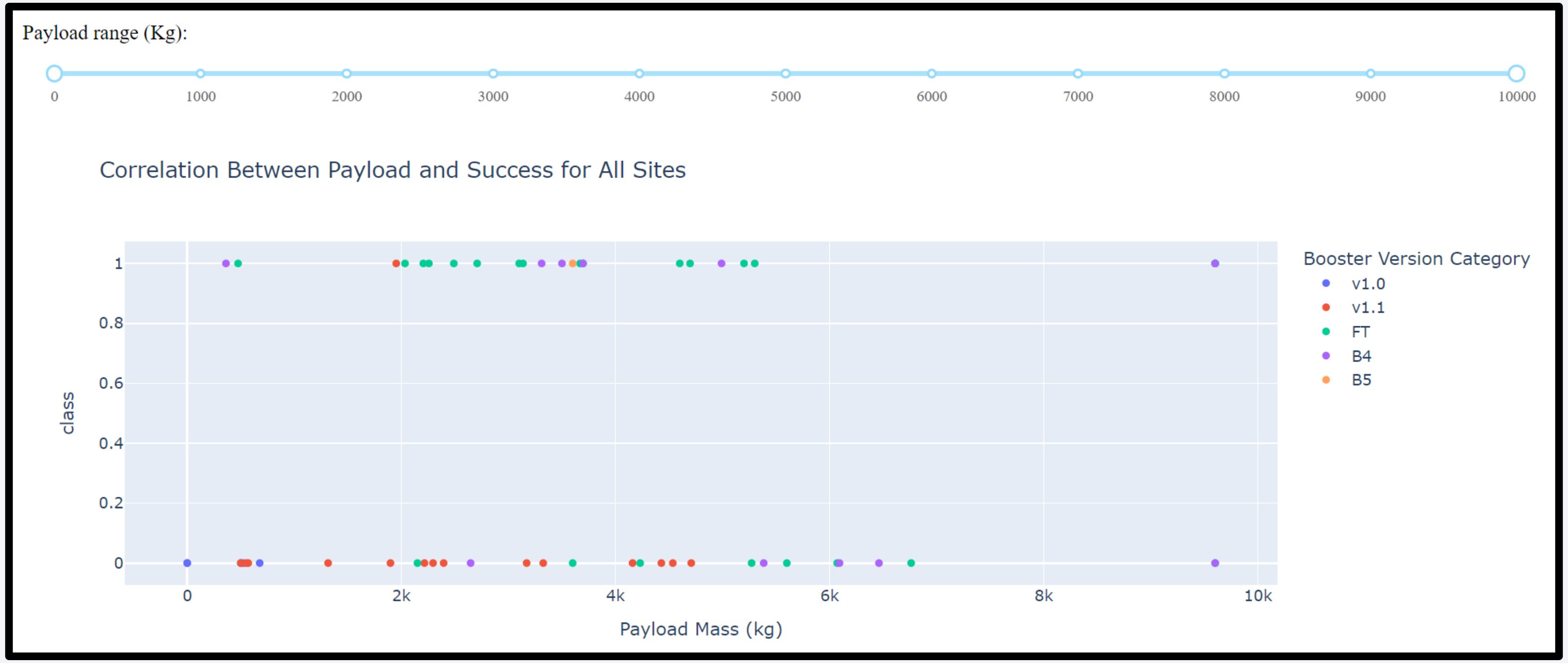
Launch Site with Highest Success Ratio



Findings:

- KSC LC-39A boasts the highest launch success rate at 76.9%.

Payload vs. Launch Outcome Analysis



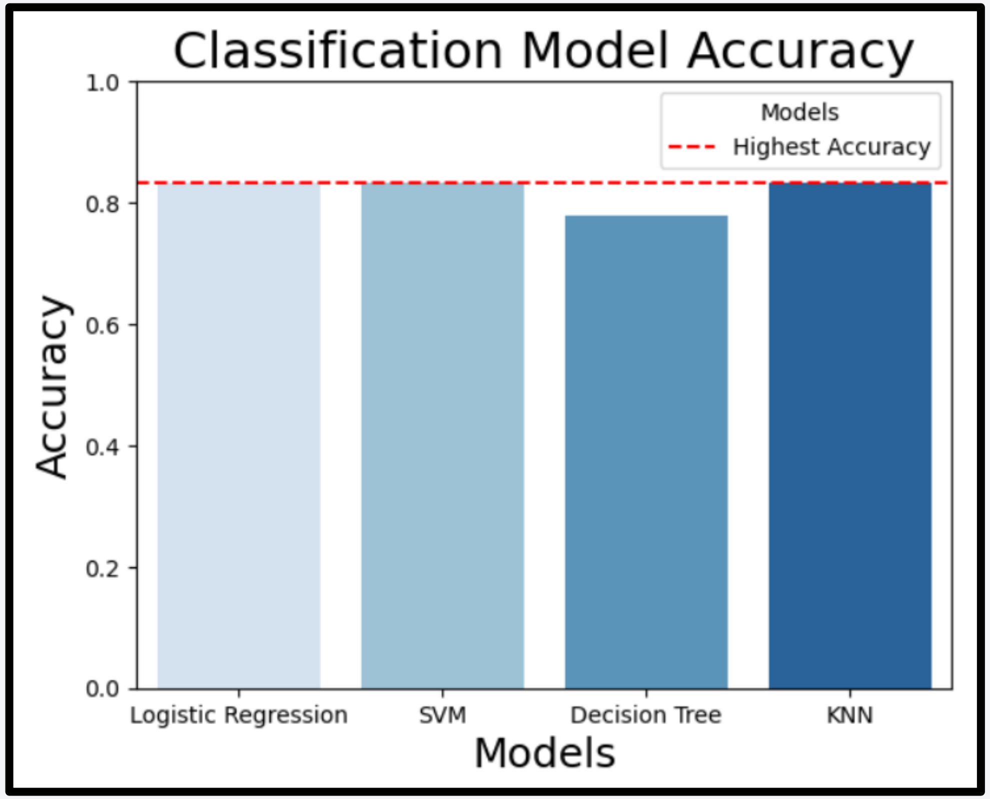
Findings:

- Payload range 2,000 kg – 4,000 kg shows the highest launch success rate.
- Payload range 6,000 kg – 8,000 kg indicates the lowest launch success rate.
- F9 Booster version FT has the highest overall launch success rate.

Section 5

Predictive Analysis (Classification)

Classification Accuracy



cross-validation:

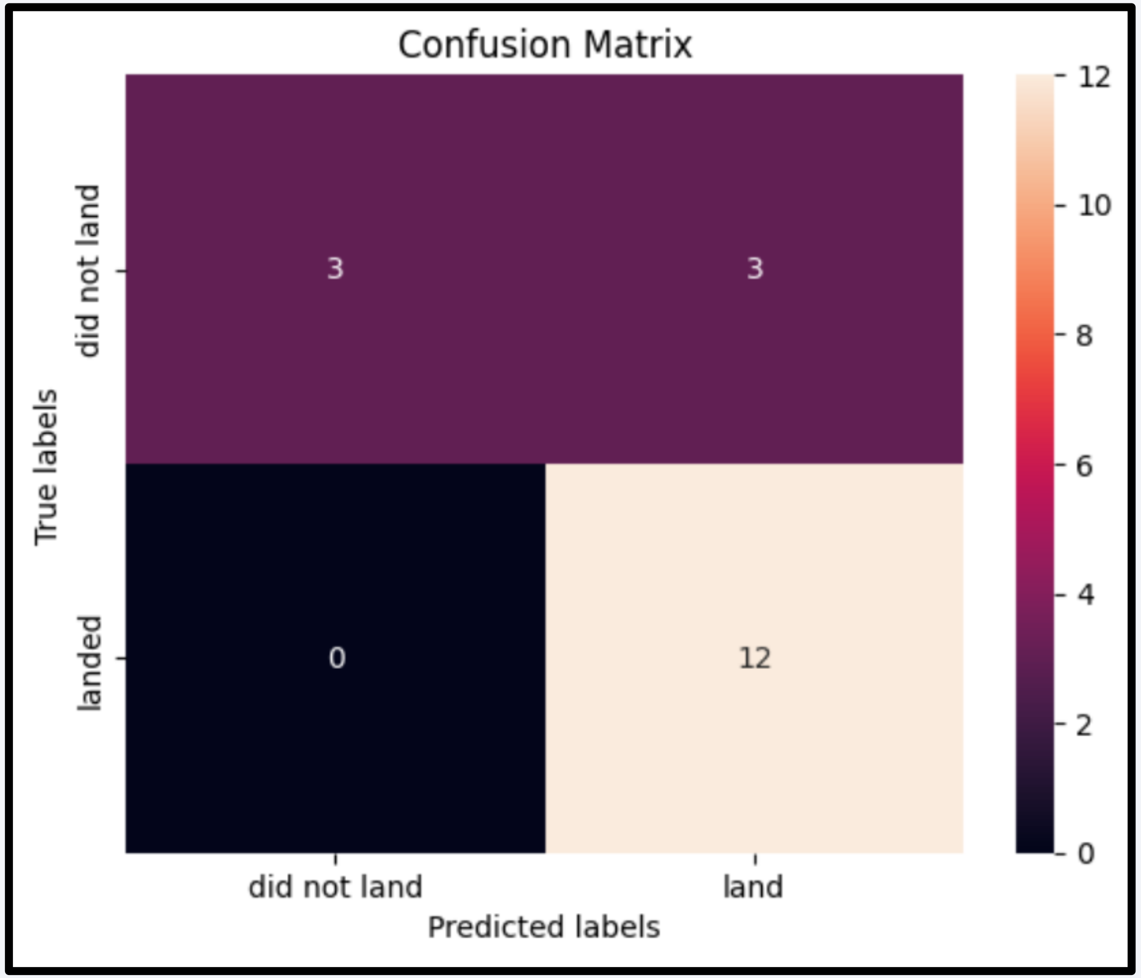
	Mean Accuracy	Std Dev
Logistic Regression	0.800000	0.129577
SVM	0.833333	0.113855
Decision Tree	0.755556	0.155556
KNN	0.766667	0.135628

Findings:

- Logistic Regression, SVM, and KNN: 83.33% (highest accuracy)
- Decision Tree:

  78.89% (performs poorly)
- Limited data affects reliability; SVM shows the best performance with higher accuracy and stability based on cross-validation.

Confusion Matrix for SVM



Findings:

- True Positives:

  12 instances correctly predicted as landed.

- False Positives:

  3 instances incorrectly predicted as landed (true label is not landed).

- The main issue is the presence of false positives, which may impact the reliability of predictions.

Conclusions

1. **Model Performance**:
   - Logistic Regression, SVM, and KNN reached 83.33% accuracy; SVM is the most stable.

2. **Launch Site Insights**:
   - CCAFS SLC-40 and KSC LC-39A are key sites; KSC LC-39A has the highest success rate (76.9%). Monitor sites with higher failure rates.

3. **Payload Success Rates**:
   - Payloads of 2,000 kg to 4,000 kg correlate with the highest success; The F9 FT boosters (4,000-6,000 kg) also demonstrate high success in drone ship landings, while heavier payloads face more challenges.

4. **Trends in Success**:
   - Notable improvement in launch success rates from 2013 to 2020 due to technological advancements.

5. **Reliability Challenges**:
   - High success rates, but false positives in predictions need further investigation for improved reliability.

Appendix

Python Notebooks:

Data Collection (SpaceX API)

Data Collection (Webscraping)

Data Wrangling

EDA (SQL Queries)

EDA (Visualization)

Launch Site Location (Folium)

Machine Learning (Prediction)

Dashboards:

Plotly Dash

Data Sets:

Dataset Part 1

Dataset Part 2

Dataset Part 3

Thank you!