Winning Space Race with Data Science

Francisco Alex Mares Solano
07-04-2024

Outline

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

Executive Summary

In this project, the data collection process involved sourcing historical launch data from SpaceX's API.

Methodology

- Data Sources: Utilized SpaceX's API.
- Data Points: Collected details on launch date, rocket type, payload, launch site, and landing outcome.
- Automation: Employed automated scripts to ensure data consistency and accuracy.

Results

- **Collected details** on launch date, rocket type, payload, launch site, and landing outcome.
- Mapped **geographic data** to visualize launch and landing sites, providing a spatial perspective on the data.
- **Interactive dashboard** to display findings and facilitate user interaction with the data.
- **Built and evaluated AI models** to predict landing success, achieving high accuracy and providing actionable insights for future launches.

Introduction

Background and context.

- SpaceX, founded by Elon Musk, focuses on reducing space transportation costs and enabling Mars colonization.
- The project aims to analyze SpaceX's historical launch data to identify factors that contribute to successful rocket landings using data analysis and machine learning.

Problems I want to find answers

- What factors influence rocket landing success?
- How can we predict future landing success?
- What trends and patterns exist in the launch data?
- How can visualizations enhance data understanding?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

Data Sources:

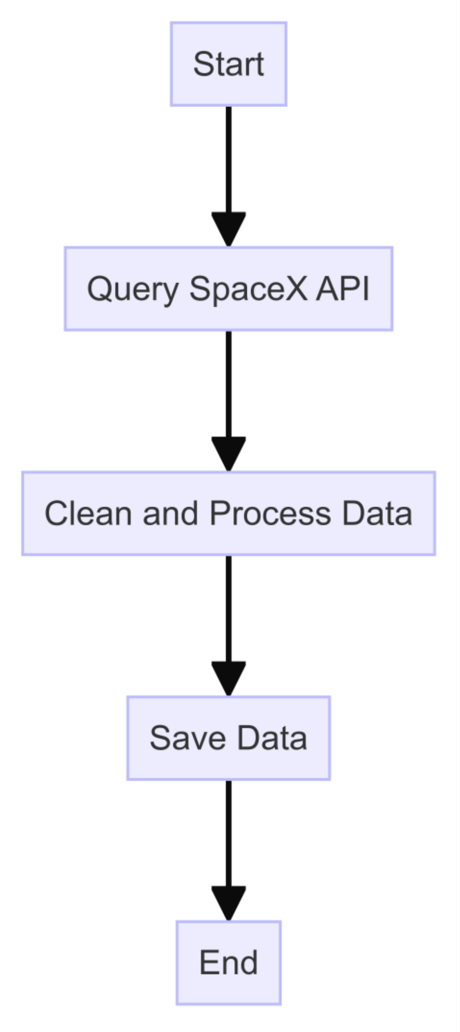
- SpaceX API: Collected historical launch data.
- Web Scraping: Extracted launch tables for Falcon 9 and Falcon Heavy.

Key Steps:

1. API Data Collection:
    - Queried SpaceX API for launch details.
    - Gathered data on launch date, rocket type, payload, launch site, and landing outcome.
2. Web Scraping:
    - Scraped launch tables for Falcon 9 and Falcon Heavy.
    - Supplemented API data with additional details.

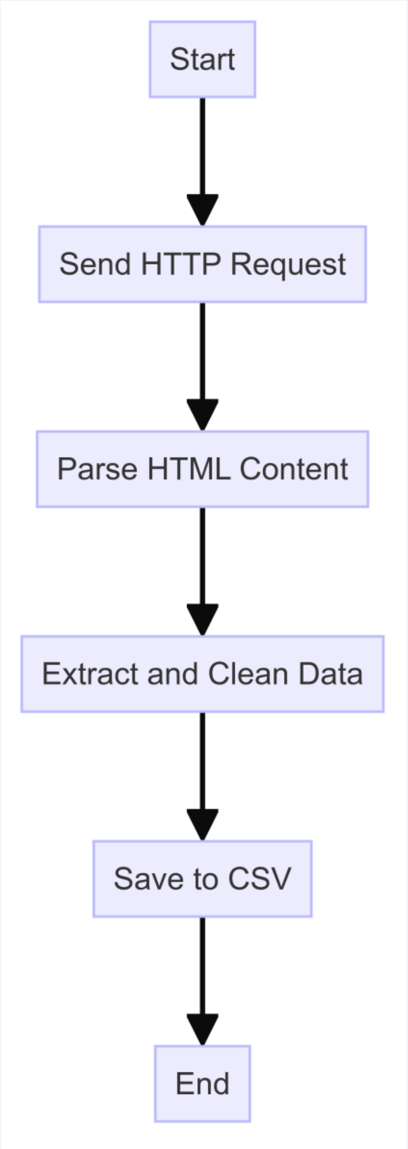
Data Collection – SpaceX API

https://github.com/alxmares/-SpaceX-Launch-Analysis-and-Landing-Prediction/blob/main/01-Collecting_data.ipynb



Data Collection - Scraping

https://github.com/alxmares/-SpaceX-Launch-Analysis-and-Landing-Prediction/blob/main/02-WebScraping.ipynb



Data Wrangling

https://github.com/alxmares/-SpaceX-Launch-Analysis-and-Landing-Prediction/blob/main/03-DataWrangling.ipynb

EDA with Data Visualization

- Categorical Plot: Payload Mass vs. Flight Number
- Categorical Plot: Launch Site vs. Flight Number
- Categorical Plot: Launch Site vs. Payload Mass
- Bar Plot: Success Rate of Each Orbit
- Scatter Plot: Flight Number vs. Orbit
- Scatter Plot: Payload Mass vs. Orbit
- Line Plot: Success Rate Over Time

https://github.com/alxmares/-SpaceX-Launch-Analysis-and-Landing-Prediction/blob/main/05-EDA2.ipynb

EDA with SQL

- CREATE TABLE SPACEXTABLE AS SELECT * FROM SPACEXTBL WHERE Date IS NOT NULL
- SELECT DISTINCT Launch_Site FROM SPACEXTABLE
- SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE 'CCA%' LIMIT 5
- SELECT AVG(PAYLOAD_MASS__KG_) WHERE Booster_Version LIKE 'F9 v1.0%'
- SELECT * WHERE date = (SELECT MIN(date) FROM SPACEXTABLE)
- SELECT Payload WHERE Landing_Outcome = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000
- SELECT DISTINCT(Mission_Outcome) FROM SPACEXTABLE
- SELECT substr(Date, 6, 2), landing_outcome, booster_version, launch_site WHERE substr(Date, 0, 5) = '2015' AND landing_outcome = 'Failure (drone ship)'
- SELECT landing_outcome, COUNT(*) AS outcome_count WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY landing_outcome ORDER BY outcome_count DESC

https://github.com/alxmares/-SpaceX-Launch-Analysis-and-Landing-Prediction/blob/main/04-EDA_SQL.ipynb

Build an Interactive Map with Folium

Objects Added:

- Markers:
    - NASA Johnson Space Center
    - SpaceX Launch Sites
- Circles:
    - NASA Johnson Space Center
    - SpaceX Launch Sites
- Lines:
    - Distances to points of interest (coastline, city, railway, highway)

Purpose:

- Highlight key locations and visualize spatial relationships.
- Aid in analysis of distances and geographic context.

https://github.com/alxmares/-SpaceX-Launch-Analysis-and-Landing-Prediction/blob/main/06-Maps.ipynb

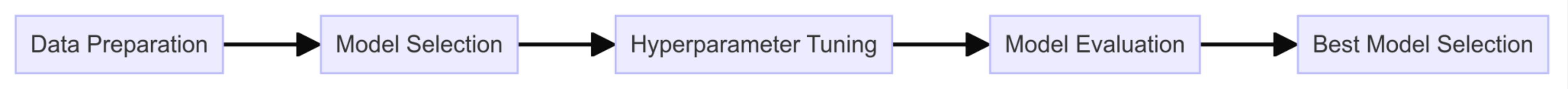
Build a Dashboard with Plotly Dash

- Pie Chart:
  - Displays: Success rates of SpaceX launches.
  - Interactions: Updates based on selected launch site.
  - Purpose: Visualize the proportion of successful launches to assess overall performance.
- Scatter Plot:
  - Displays: Relationship between payload mass and launch outcomes.
  - Interactions: Updates based on selected launch site and payload range.
  - Purpose: Analyze the impact of payload mass on launch success and compare booster versions.

https://github.com/alxmares/-SpaceX-Launch-Analysis-and-Landing-Prediction/blob/main/07-Dashboard.py

Predictive Analysis (Classification)

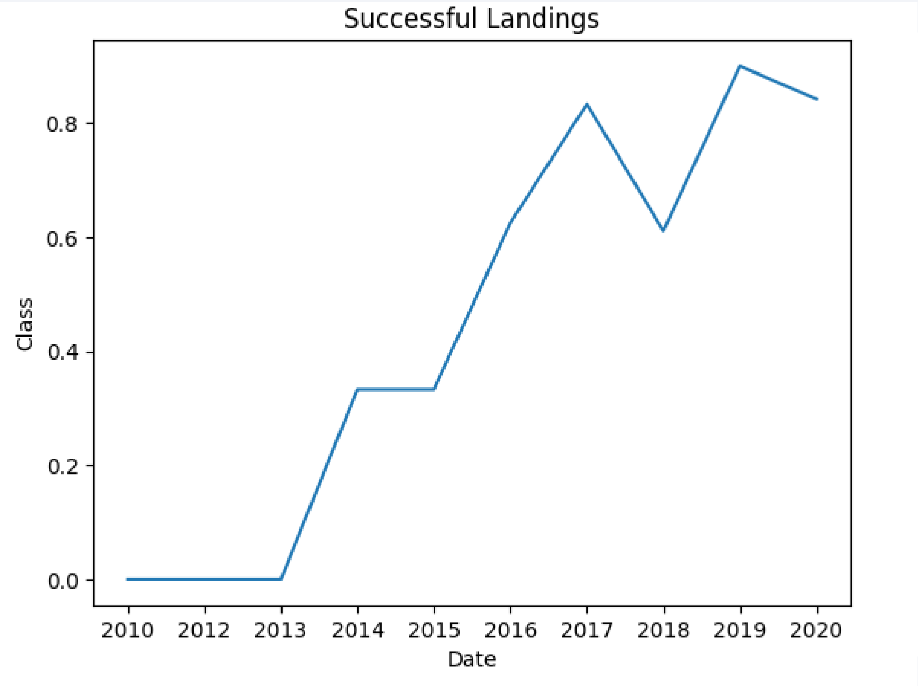
- Model Selection and Training:
  - Used GridSearchCV for hyperparameter tuning.
  - Evaluated Logistic Regression, SVM, Decision Tree, and KNN models.
- Model Evaluation:
  - Compared accuracy scores and confusion matrices.
  - Selected the best-performing model based on test accuracy.

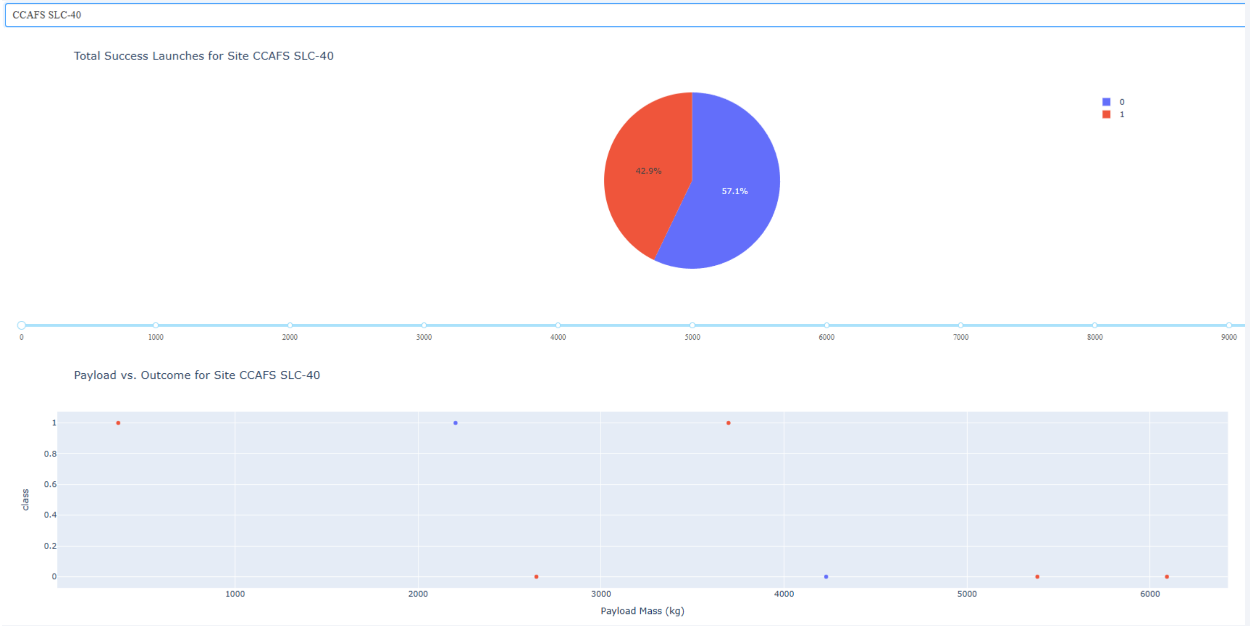


https://github.com/alxmares/-SpaceX-Launch-Analysis-and-Landing-Prediction/blob/main/08-ML_to_predict_land.ipynb

Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results: 83.33%

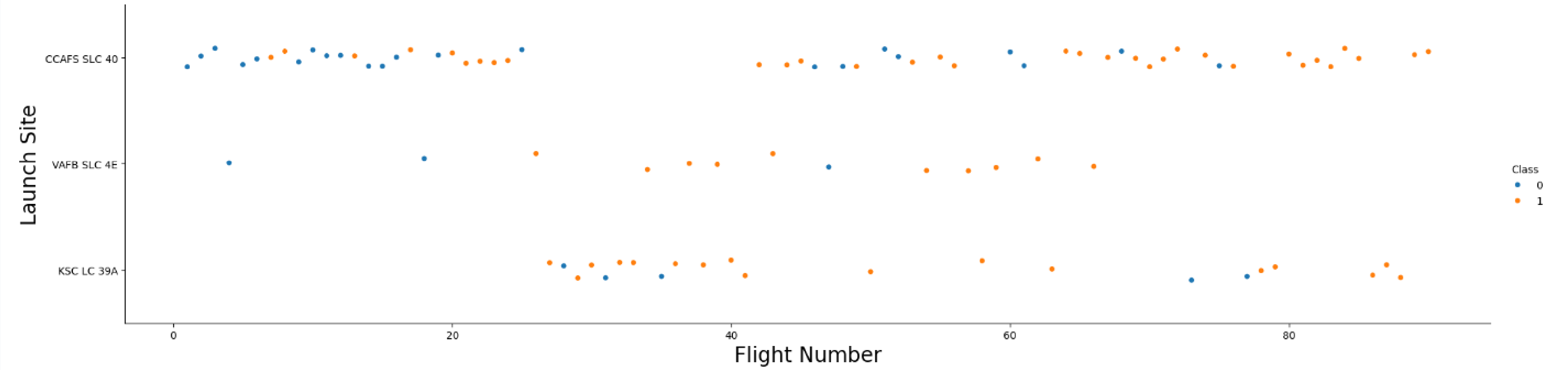




Section 2

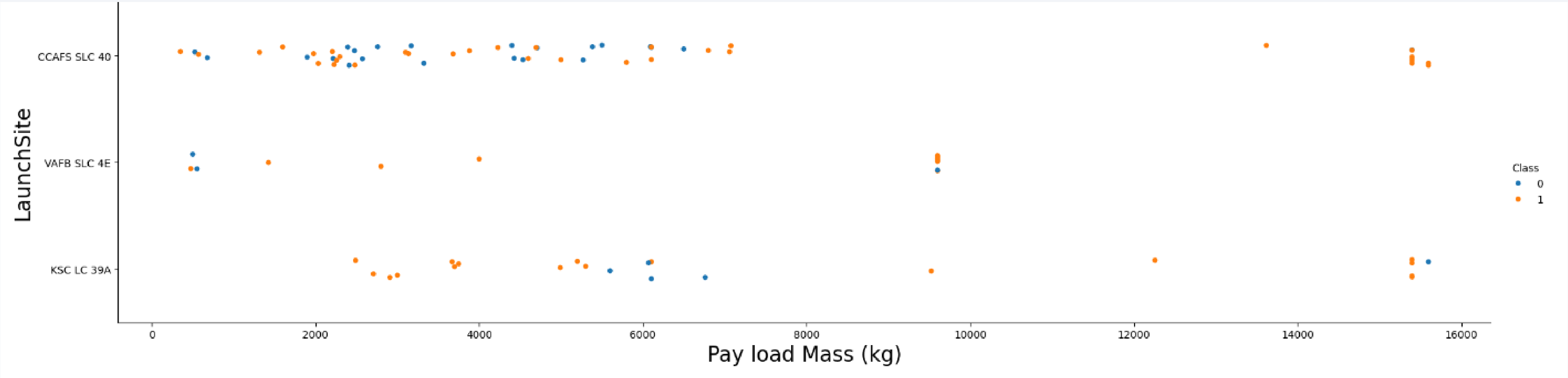
Insights drawn from EDA

Flight Number vs. Launch Site



- Launch Sites:
  - CCAFS SLC 40: High success rate, more consistent performance.
  - VAFB SLC 4E: Fewer launches, mixed success.
  - KSC LC 39A: Higher number of launches, moderate success rate.
- Trends: Later flights tend to have higher success rates, indicating improvements over time.

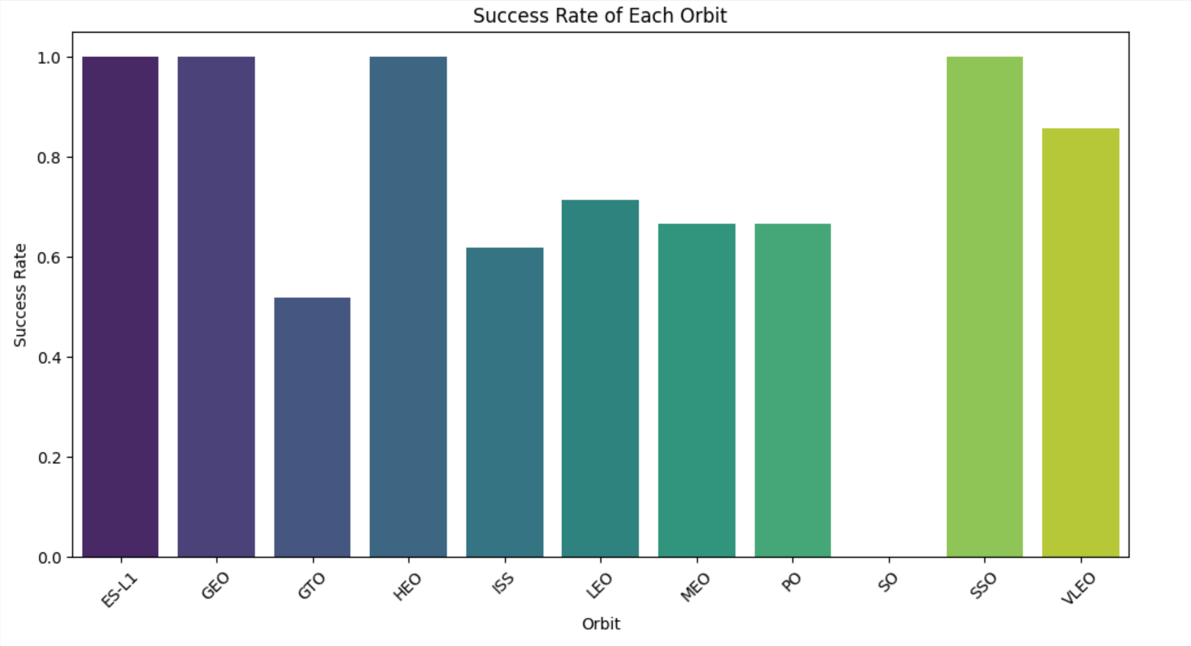
Payload vs. Launch Site



- Launch Sites:
  - CCAFS SLC 40: Handles a wide range of payload masses with a relatively high success rate.
  - VAFB SLC 4E: Fewer data points, moderate success with varying payload masses.
  - KSC LC 39A: Manages high payload masses with mixed success.
- Trends: Higher payload masses tend to have a mixed success rate, but success is not solely dependent on payload mass.

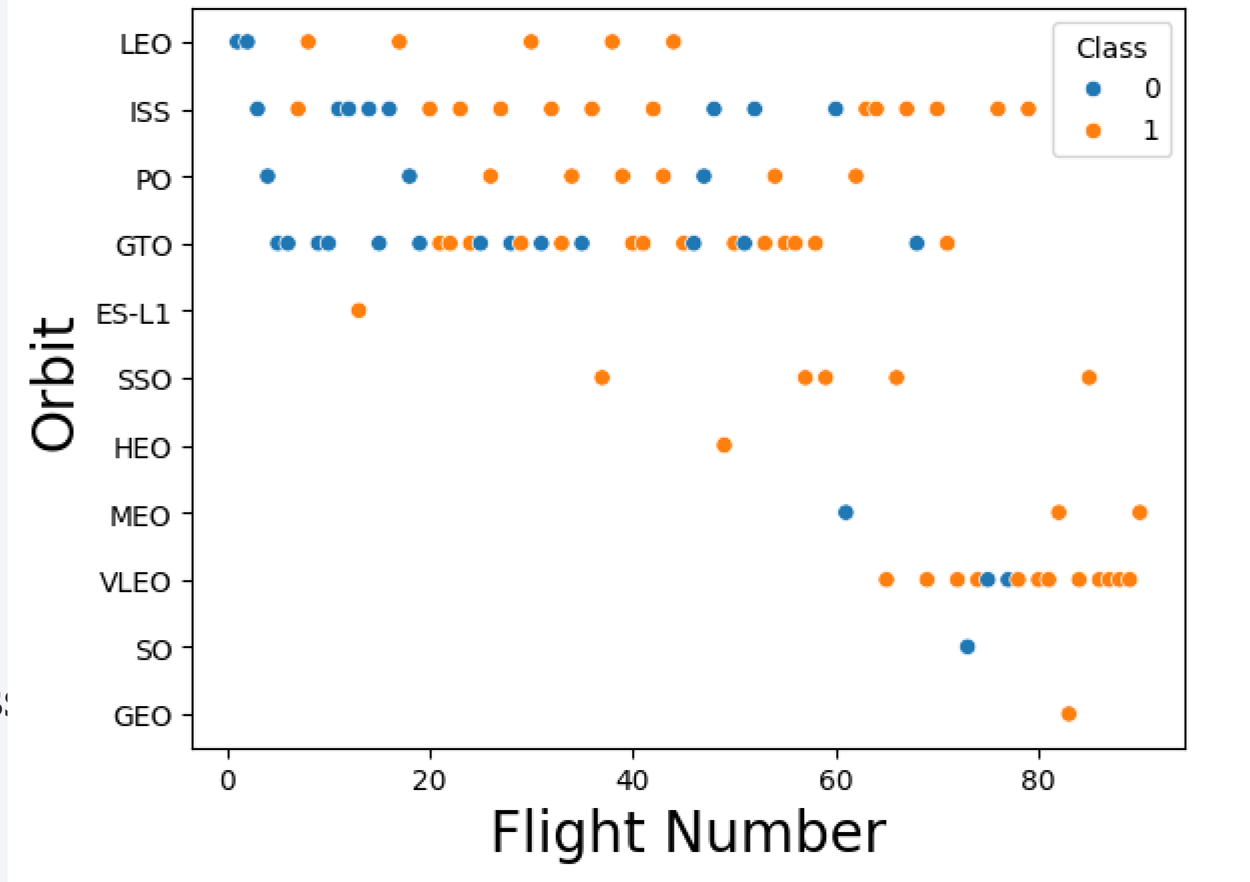
Success Rate vs. Orbit Type

- High Success Rates:
  - ES-L1, GEO, HEO: Near 100% success rates.
  - SSO: Very high success rate.
- Moderate Success Rates:
  - LEO, MEO, PO, VLEO: Success rates around 70-80%.
- Lower Success Rate:
  - GTO: Success rate around 50%. for the success rate of each orbit type



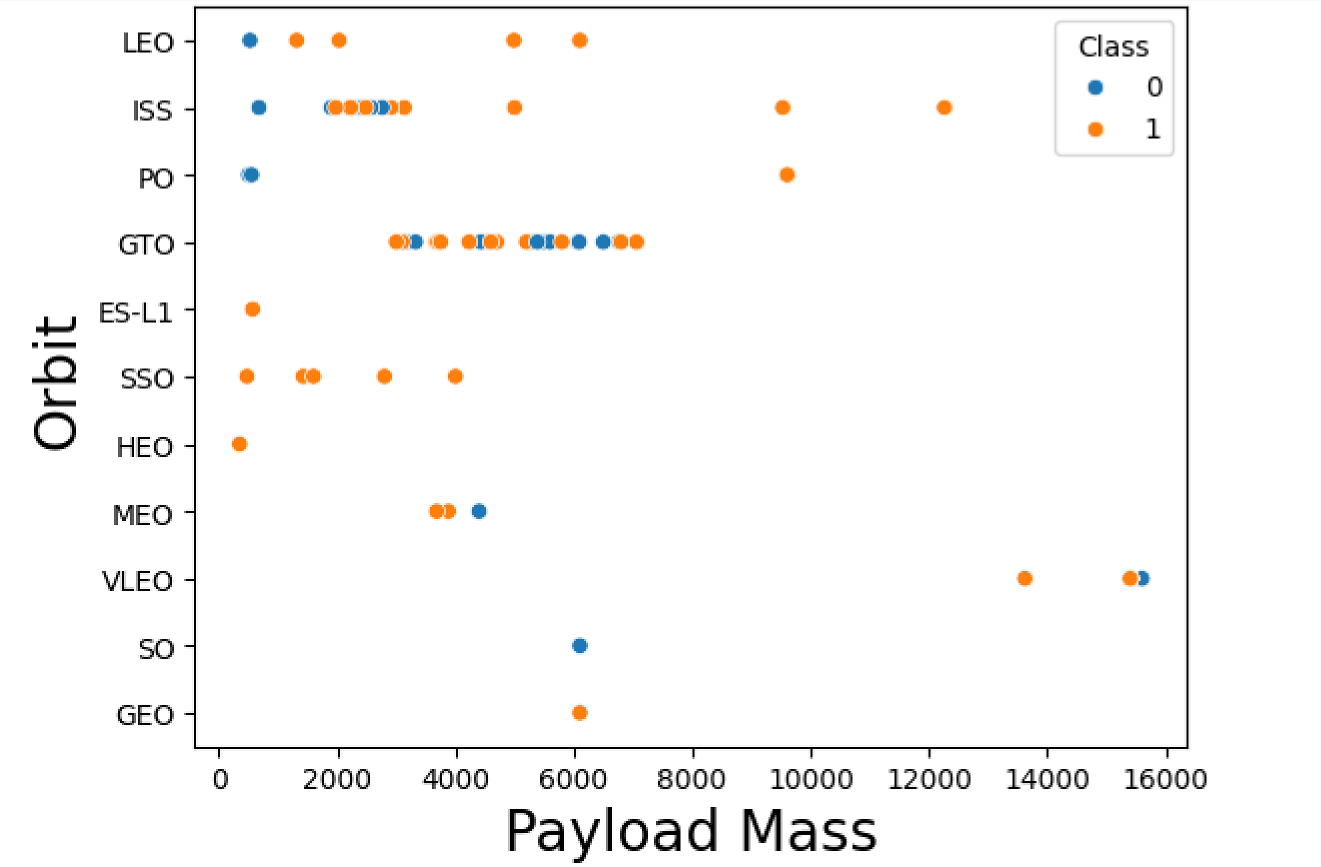
Flight Number vs. Orbit Type

- LEO (Low Earth Orbit): High number of launches with a mix of successes and failures.
- ISS, PO: Moderate number of launches, both successful and unsuccessful.
- GTO, HEO: Fewer launches but high success rates.
- Trend: Recent flights (higher flight numbers) show more consistent success across different orbits.



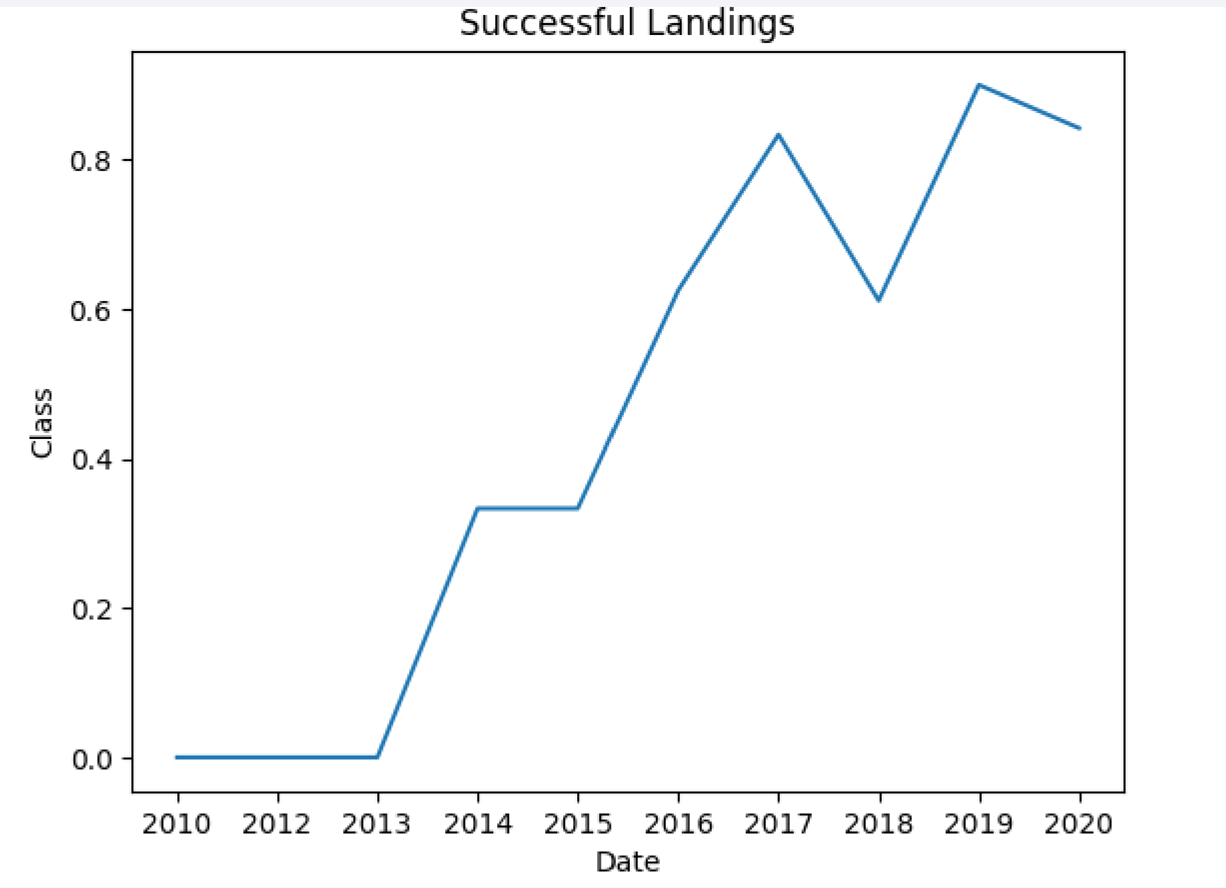
Payload vs. Orbit Type

- LEO: Handles a wide range of payload masses with a relatively high success rate.
- ISS and GTO: High concentration of launches around specific payload masses, mixed success rates.
- High Payloads (>10,000 kg): Few launches, success not guaranteed.
- Trend: Success is generally not dependent solely on payload mass, as there are successful and unsuccessful launches across various masses for different orbits.



Launch Success Yearly Trend

Marked improvement in the success rate of landings over the years, indicating advancements in technology and processes by SpaceX.



All Launch Site Names

The unique launch sites are:

- CCAFS LC-40
- VAFB SLC-4
- EKSC LC-39
- ACCAFS SLC-40

Launch Site Names Begin with 'CCA'

These records detail early SpaceX launches, all from the CCAFS LC-40 site, showing initial mission successes with varied landing outcomes.

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

Total Payload Mass:

- Sum: 619,967 kg

This query calculates the total payload mass for all SpaceX launches in the table.

Average Payload Mass by F9 v1.1

Average Payload Mass for Booster Version 'F9 v1.0’:

- 340.4 kg

This represents the average payload mass for SpaceX launches using the 'F9 v1.0' booster version.

First Successful Ground Landing Date

Earliest Launch Record:

- Date: 2010-06-04
- Time (UTC): 18:45:00
- Booster Version: F9 v1.0 B0003
- Launch Site: CCAFS LC-40
- Payload: Dragon Spacecraft Qualification Unit
- Payload Mass: 0 kg
- Orbit: LEO
- Mission Outcome: Success
- Landing Outcome: Failure (parachute)

Successful Drone Ship Landing with Payload between 4000 and 6000

Payloads with Successful Drone Ship Landings (4000 kg < Mass < 6000 kg):

- JCSAT-14
- JCSAT-16
- SES-10
- SES-11 / EchoStar 105

Total Number of Successful and Failure Mission Outcomes

Mission Outcomes:

- Success: 98 occurrences
- Failure (in flight): 1 occurrence

This shows the count of missions that were either successful or failed in flight, highlighting a high success rate for SpaceX launches.

Boosters Carried Maximum Payload

Booster Versions with 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

2015 Launch Records

January:

- Booster Version: F9 v1.1 B1012
- Launch Site: CCAFS LC-40

April:

- Booster Version: F9 v1.1 B1015
- Launch Site: CCAFS LC-40

These records show specific months in 2015 when SpaceX experienced drone ship landing failures, highlighting challenges faced early in the year with the F9 v1.1 booster version at the CCAFS LC-40 launch site.

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

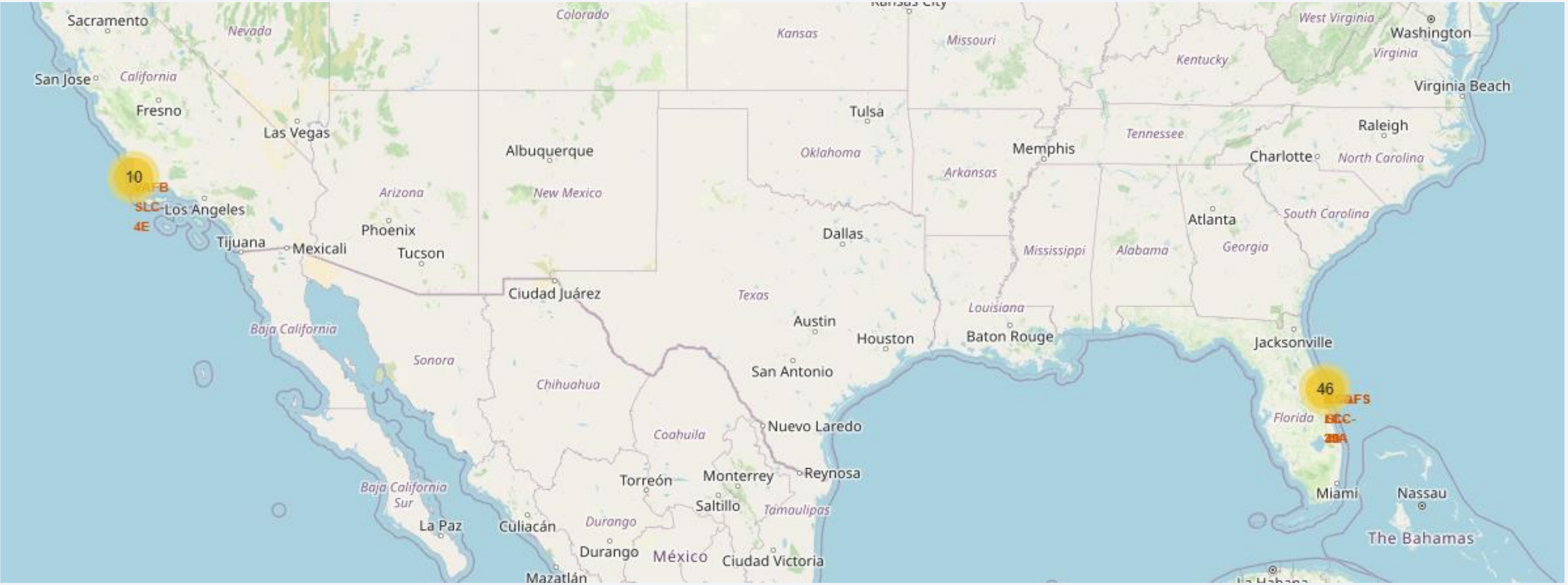
- No attempt: 10
- Success (drone ship): 5
- Failure (drone ship): 5
- Success (ground pad): 3
- Controlled (ocean): 3
- Uncontrolled (ocean): 2
- Failure (parachute): 2
- Precluded (drone ship): 1

This shows the count of each landing outcome for SpaceX launches within the specified date range, with "No attempt" being the most frequent outcome.

Section 3

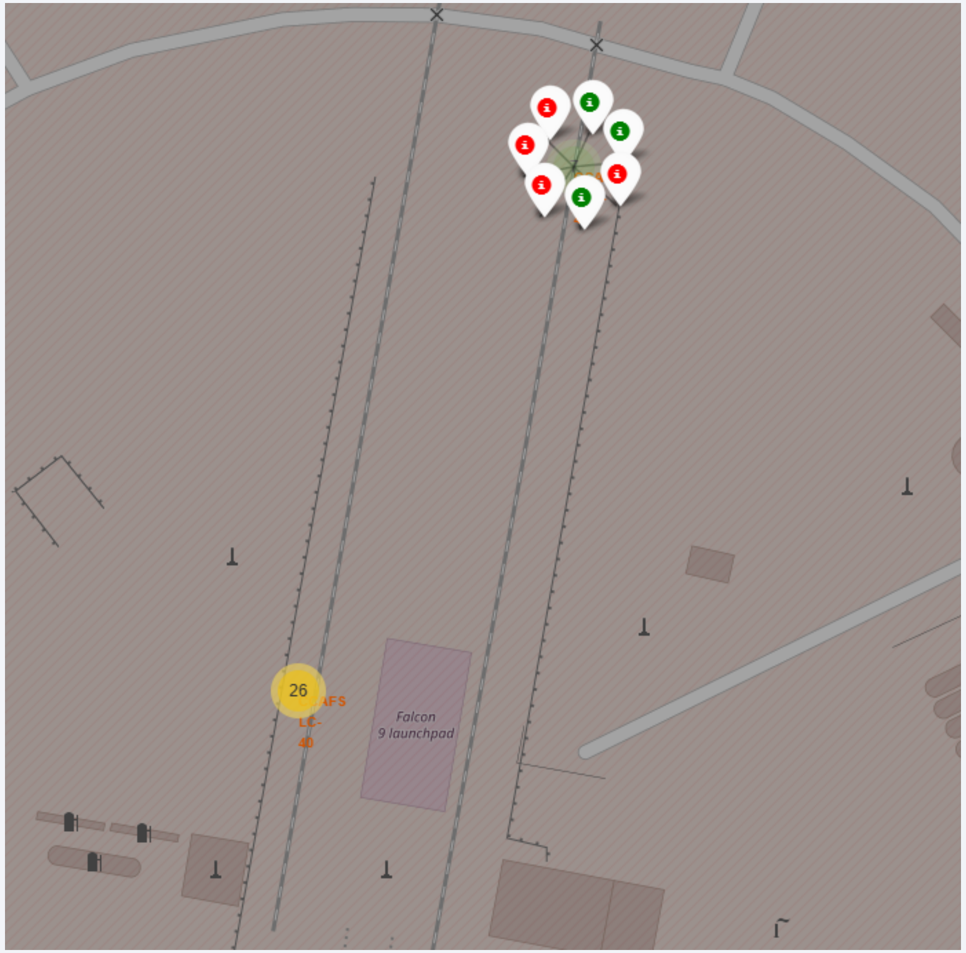
Launch Sites Proximities Analysis

SpaceX Launch Sites and Outcomes



This map visualizes SpaceX launch sites, highlighting the number and success rate of launches at each location.

Detailed View of Launch Sites



Markers:

- Green: Successful launches.
- Red: Unsuccessful launches.

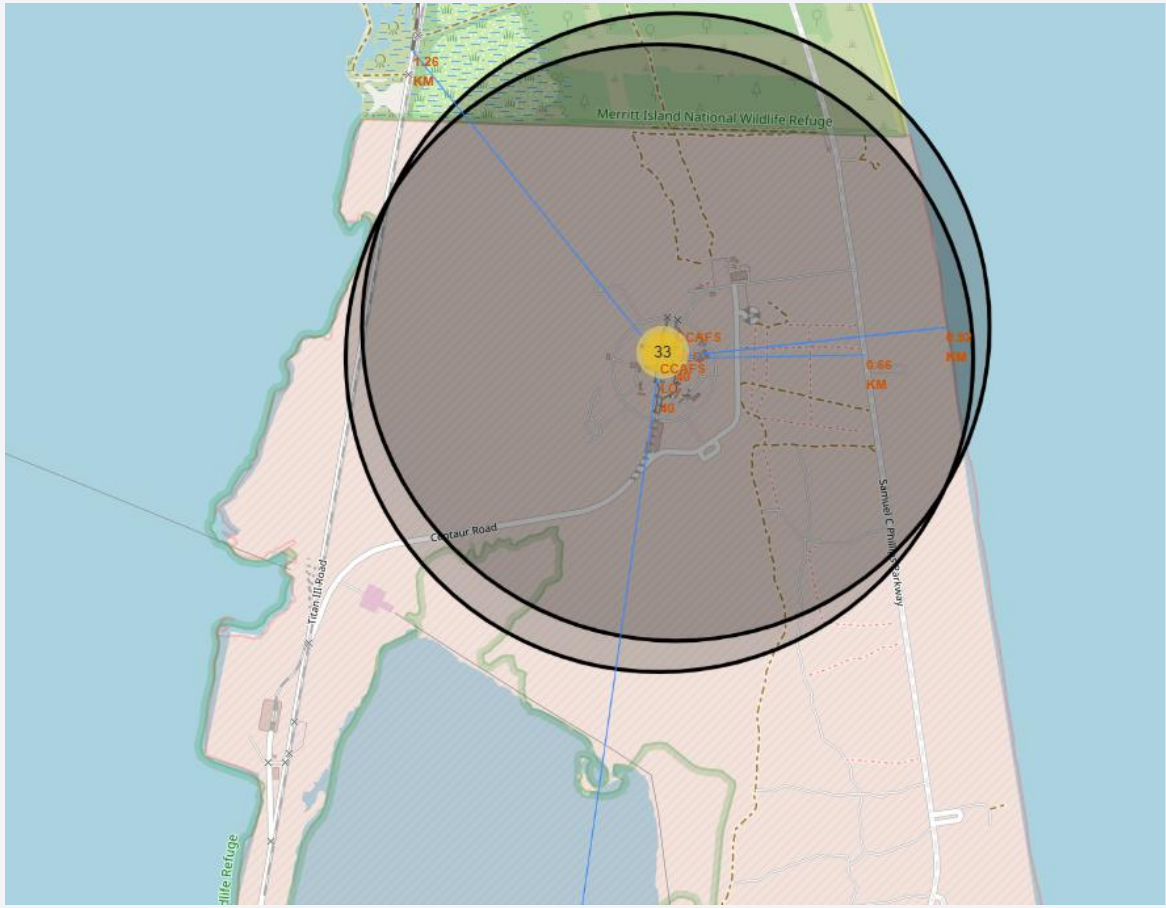
Clusters: Markers are clustered to show launch site density

Distances from Launch Site to Points of Interest

Distance Markers: Show distances from the launch site to key points.

- Coastline: 0.92 KM
- Closest City: 26 KM
- Railway: 0.66 KM
- Highway: 0.68 KM

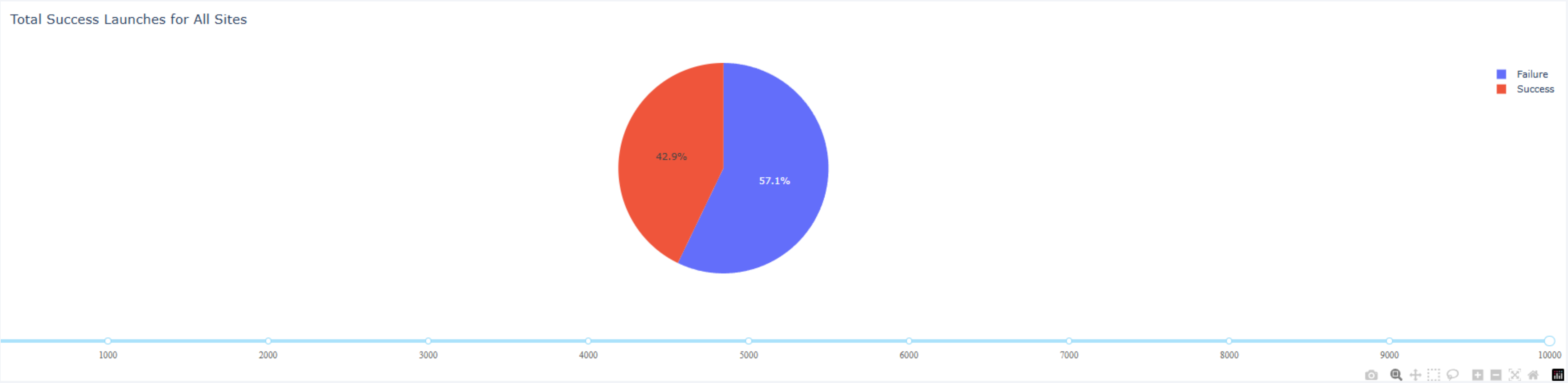
Lines: Connect the launch site to points of interest, indicating exact distances.



Section 4

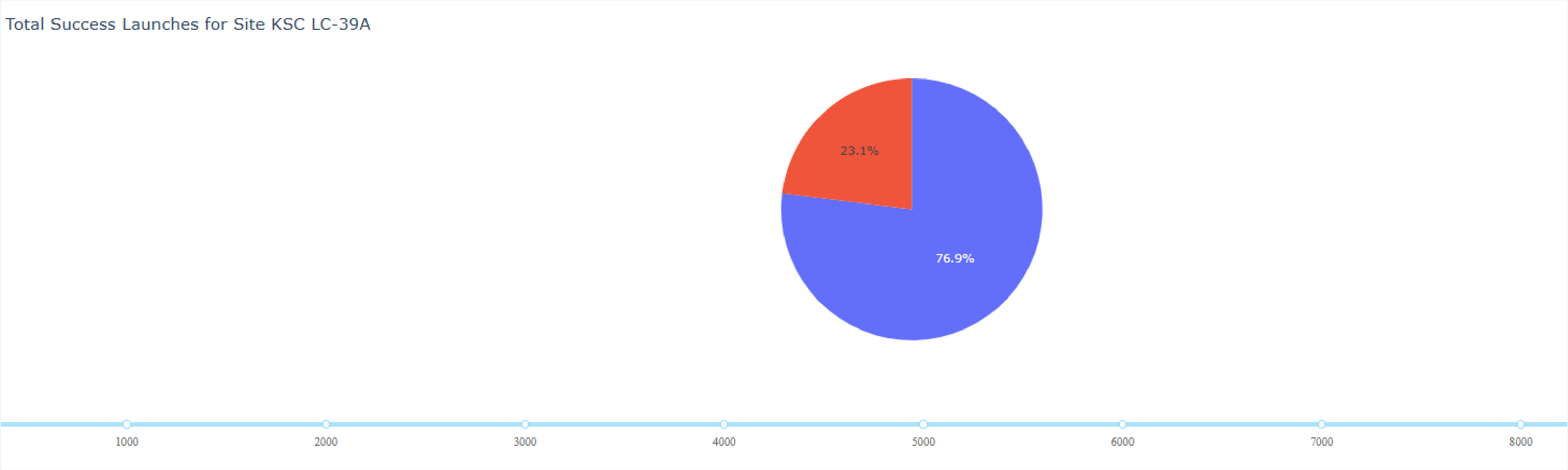
Build a Dashboard with Plotly Dash

Launch Success Rate for All Sites



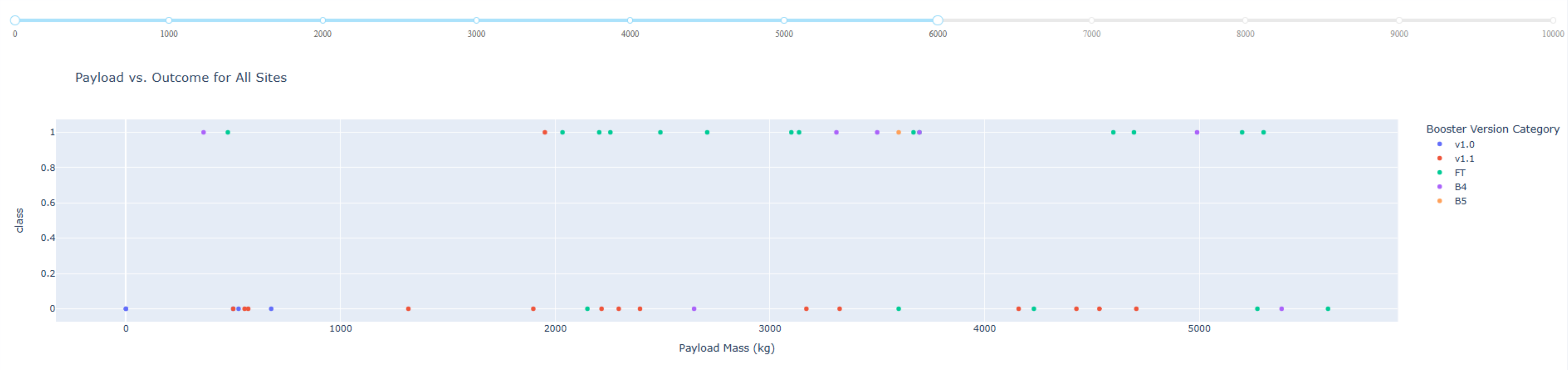
The pie chart shows 57.1% successful launches and 42.9% failures across all sites.

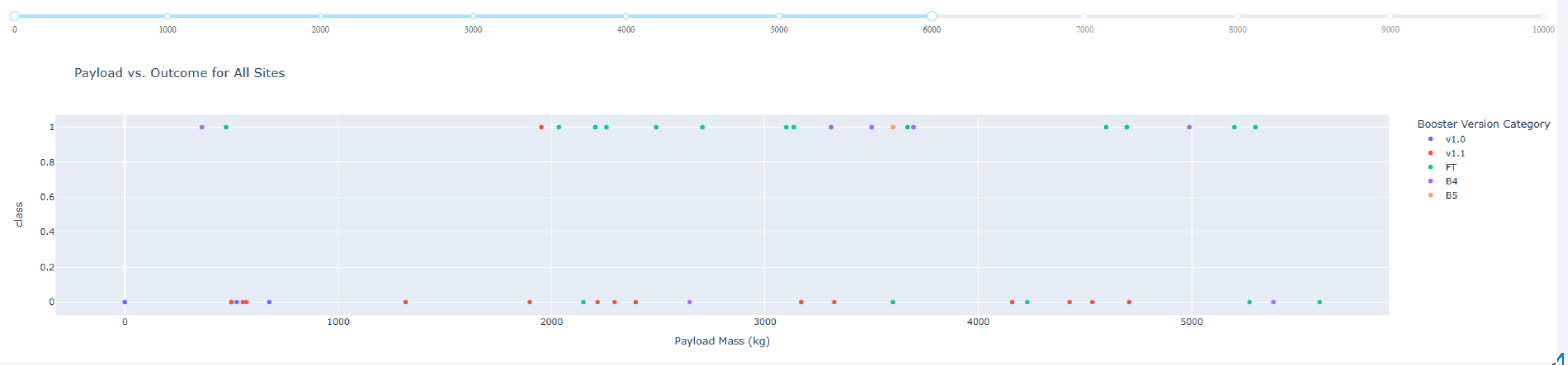
Highest launch success ratio: KSC LC-39A



The pie chart shows 76.9% successful launches and 23.1% failures at KSC LC-39A.

Payload vs. Outcome for All Sites



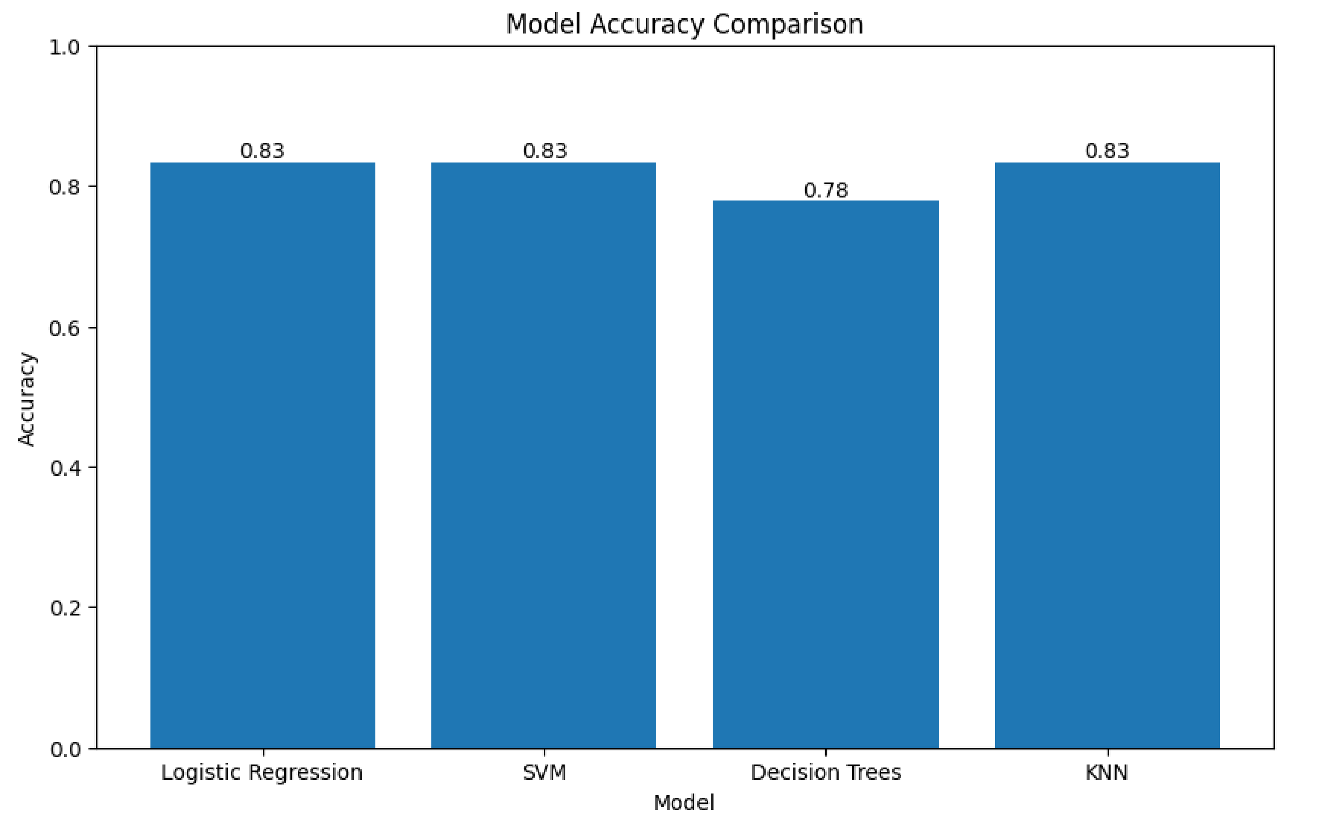


Section 5

Predictive Analysis (Classification)

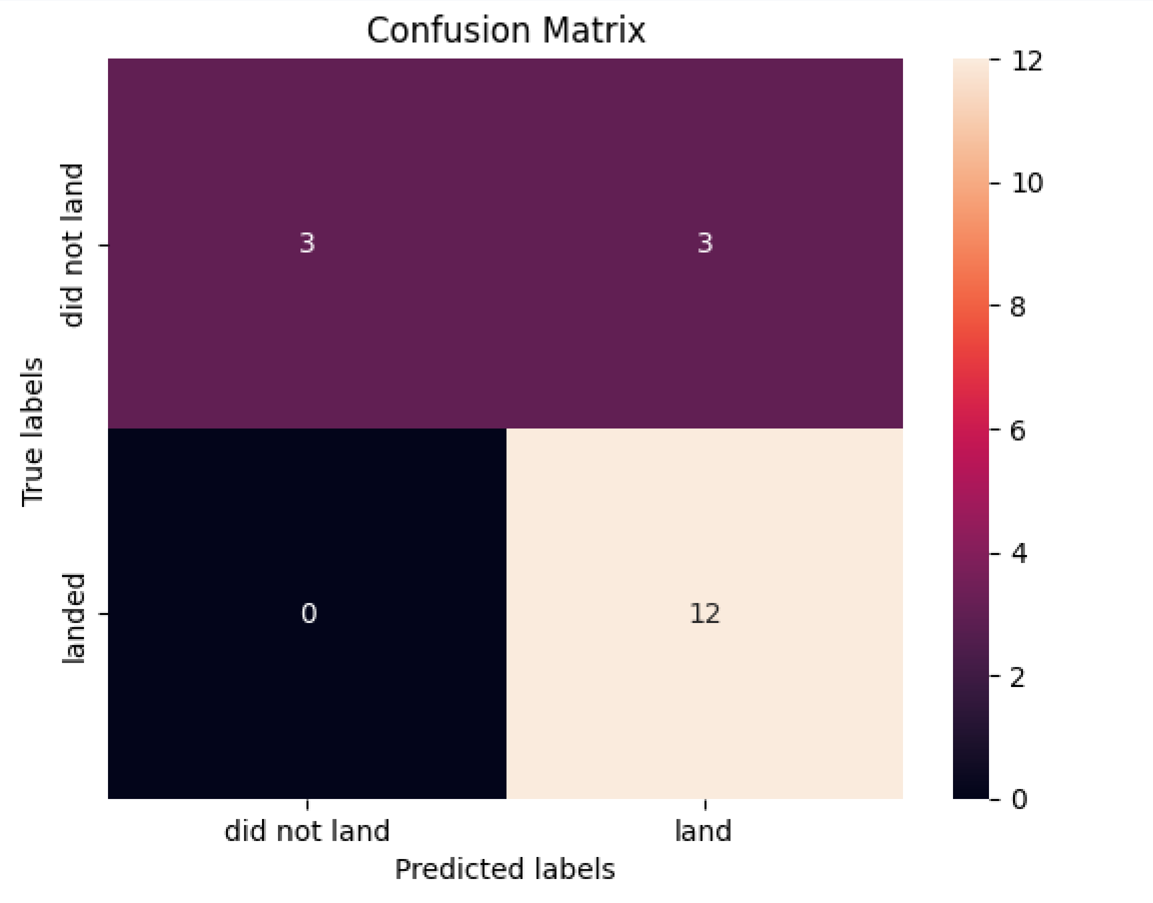
Classification Accuracy

This bar chart shows that Logistic Regression, SVM, and KNN models have the highest accuracy, each at 83%, while Decision Trees have a slightly lower accuracy at 78%.



Confusion Matrix

This confusion matrix indicates that the SVM model correctly predicts most landings, with some errors in predicting non-landings.



Conclusions

Data Visualization:

- Created visualizations to understand launch success rates and payload outcomes, identifying trends and patterns across sites and payload ranges.

Model Accuracy:

- Logistic Regression, SVM, and KNN models achieved the highest accuracy (83%), with Decision Trees slightly lower at 78%.
- SVM model showed strong performance with 12 true positives, 3 false positives, and no false negatives.

Launch Site Insights:

- KSC LC-39A had the highest launch success rate at 76.9%, indicating its reliability.
- Scatter plots revealed successful launches across a range of payload masses, showing how payload mass influences outcomes.
- Implemented interactive dashboards for dynamic data exploration, enabling detailed analysis of launch outcomes and payload influences.

Appendix

Data URL:

- https://api.spacexdata.com/v4/rockets/
- https://api.spacexdata.com/v4/launchpads/
- https://api.spacexdata.com/v4/payloads/
- https://api.spacexdata.com/v4/cores/
- https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json

Thank you!