IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Roman Hulyk
01.08.2025

Outline

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

Executive Summary

- This project analyzes SpaceX Falcon 9 launch data to predict first-stage landing success. We collected data from SpaceX API and Wikipedia, cleaned and merged it, and explored patterns through visualizations and SQL queries.
- We built an interactive dashboard and trained several classification models to predict mission success based on payload, launch site, and booster version.
  Insights from this analysis can help optimize future launch decisions and reduce mission costs.

Introduction

- SpaceX aims to reduce launch costs by reusing rocket components.
  This project investigates factors influencing Falcon 9's first-stage landing success.
  We seek to answer:
- What factors influence landing outcomes?
- Can we predict landing success using mission features?

Section 1

Methodology

Methodology

Executive Summary

- SpaceX API used to extract mission data
- Wikipedia scraping provided additional booster information
- Datasets were cleaned and merged for analysis
- EDA performed with plots and SQL queries
- Folium maps explored geographic patterns
- Dashboard enabled real-time user interaction
- ML models predicted binary landing outcomes (success/failure)

Data Collection

- We collected SpaceX mission data using two main sources:
- REST API: SpaceX API was used to retrieve detailed mission launch records in JSON format
- Web Scraping: Wikipedia pages were scraped using BeautifulSoup to obtain booster and payload details
- Data from both sources was parsed, normalized, and converted into pandas DataFrames.
- The combined dataset was used for analysis, visualization, and modeling.

Data Collection – SpaceX API

- We collected the launch data from the SpaceX public REST API using a Python GET request via the requests library. The returned JSON data was normalized and converted into a pandas DataFrame for further processing.
- The collected data included fields such as mission name, launch date, rocket type, payload mass, landing success, and launch site.
- The process was encapsulated into reusable functions for flexibility and efficiency.
- The data acquisition code is available on GitHub for peer review and reproducibility.
- https://github.com/lorumo/SpaceY/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

GET request to SpaceX API

↓

Receive JSON response

↓

Normalize JSON into tabular format

↓

Convert to pandas DataFrame

↓

Save for further analysis

Data Collection - Scraping

- Collected Falcon 9 launch data from Wikipedia
  using requests and BeautifulSoup.
  Parsed the HTML table and converted it to a pandas DataFrame.
  Extracted fields include launch date, site, booster version, payload, and outcome.
  Code and results are available on GitHub.
- https://github.com/lorumo/SpaceY/blob/main/jupyter-labs-webscraping.ipynb

Send GET request to Wikipedia page

↓

Parse HTML content using BeautifulSoup

↓

Locate the Falcon 9 Launch Records table

↓

Extract headers and rows

↓

Convert to pandas DataFrame

Data Wrangling

- Cleaned and merged API and web-scraped data.
  Converted string and date formats, handled missing values.
  Created new columns such as "Class" for supervised learning.
- GitHub notebook includes full data prep workflow.
- https://github.com/lorumo/SpaceY/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Created scatter and bar plots to analyze launch trends.
  Explored relationships between payload mass, flight number, orbit type, and success.
- Visualizations supported feature selection for modeling.
- GitHub includes all charts and code.
- https://github.com/lorumo/SpaceY/blob/main/edadataviz.ipynb

EDA with SQL

- Used SQL queries to explore launch records.
- Retrieved launch sites, payload statistics, landing outcomes, booster types, and trends.
- Enabled structured insights into mission success factors.
- Full notebook and query results are available on GitHub.
- https://github.com/lorumo/SpaceY/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Built interactive maps using Folium to visualize launch sites and outcomes.
- Added markers, popups, and color-coded results.
- Helped explore spatial proximity to features like highways and coastlines.
- All map code are on GitHub.
- https://github.com/lorumo/SpaceY/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Built an interactive dashboard using Plotly Dash.
- Included dropdown for launch site, payload range slider, pie chart, and scatter plot.
- Users can explore success rates by site, payload, and booster version.
- Full code and screenshots are available on GitHub.
- https://github.com/lorumo/SpaceY/blob/main/spacex-dash-app.py

Predictive Analysis (Classification)

- Trained classification models (Logistic Regression, SVM, Decision Tree, KNN) to predict landing success.
- Tuned hyperparameters and evaluated models using accuracy score and confusion matrix.
- Selected the best-performing model based on metrics.
- Full code and evaluation results available on GitHub.
- https://github.com/lorumo/SpaceY/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

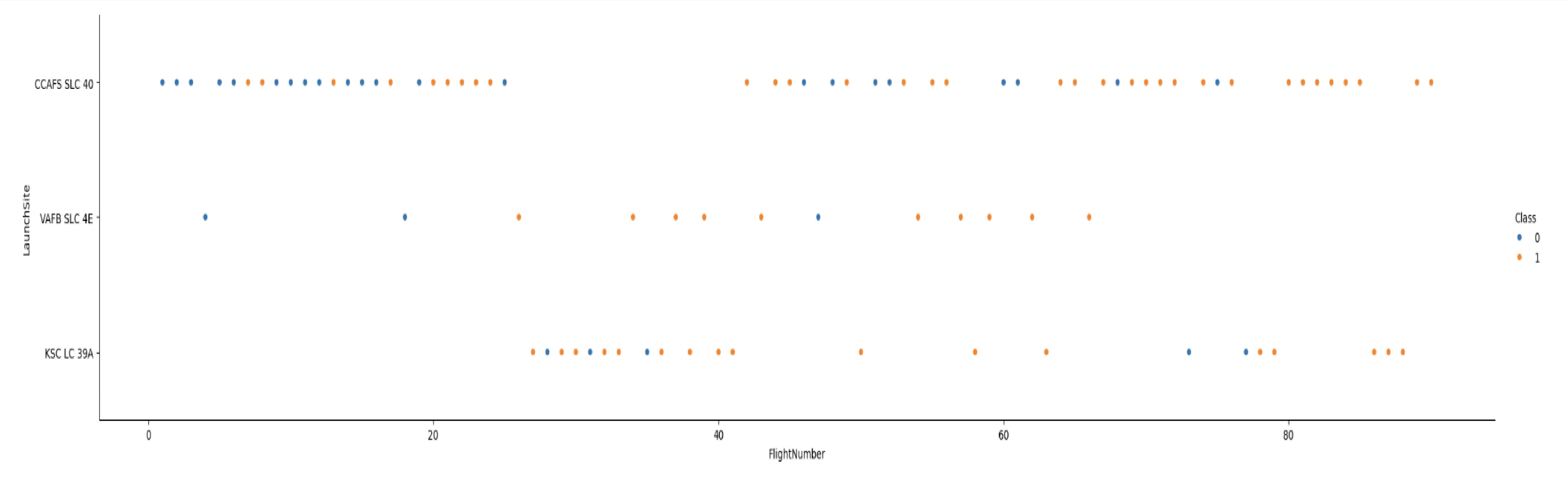
Results

- This section presents the key findings and visual insights obtained from the analysis.
  It includes:
- Exploratory visualizations
- SQL query results
- Interactive map screenshots
- Dashboard outputs
- Predictive model performance

Section 2

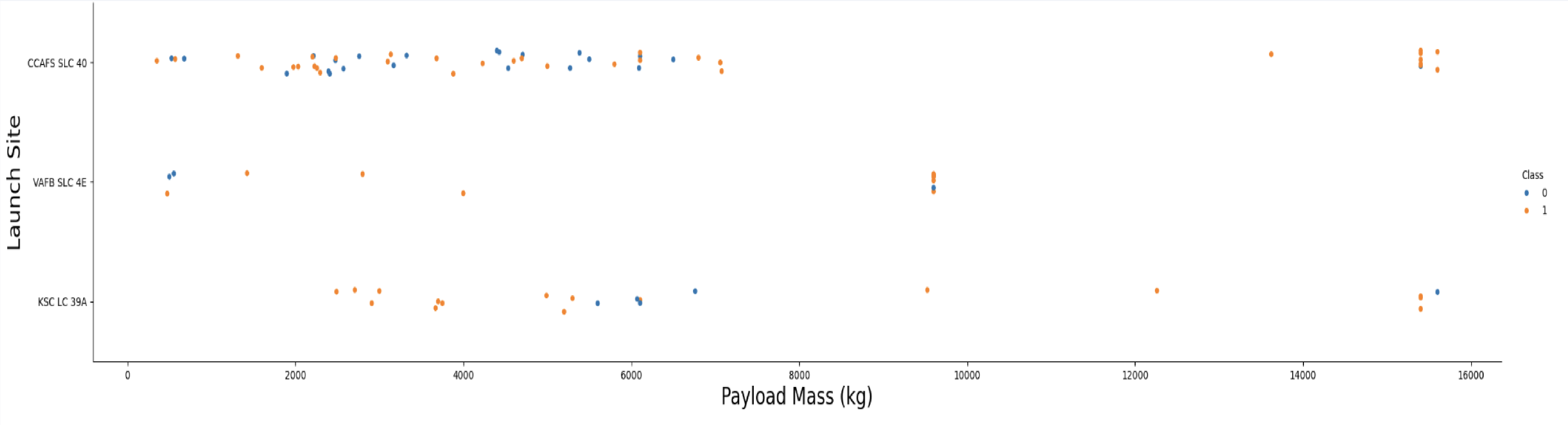
Insights drawn from EDA

Flight Number vs. Launch Site



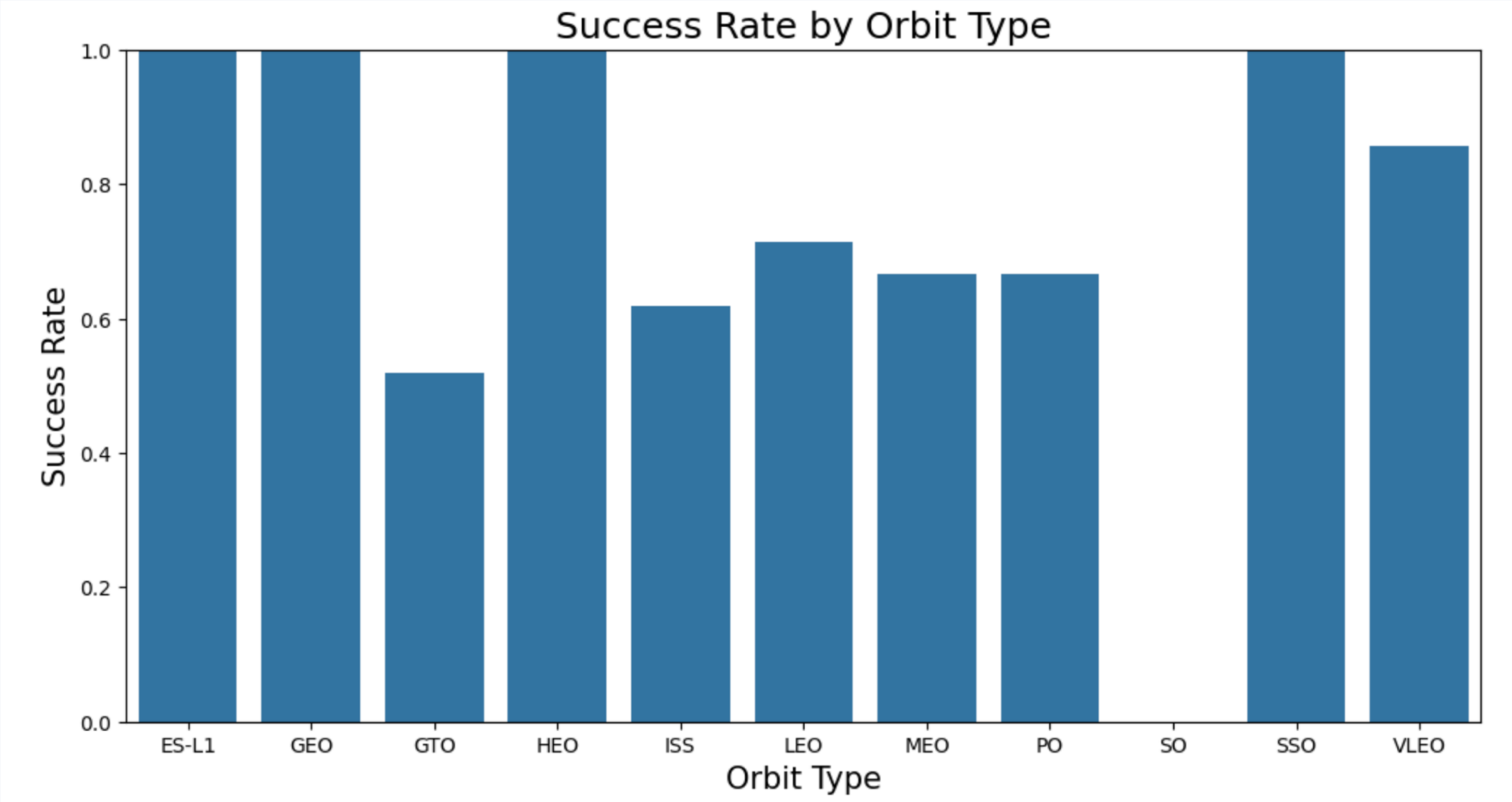
- More recent flights (higher flight numbers) show increased success across all sites.
- CCAFS SLC 40 and KSC LC 39A have the most launches.
- Early missions had more failures

Payload vs. Launch Site



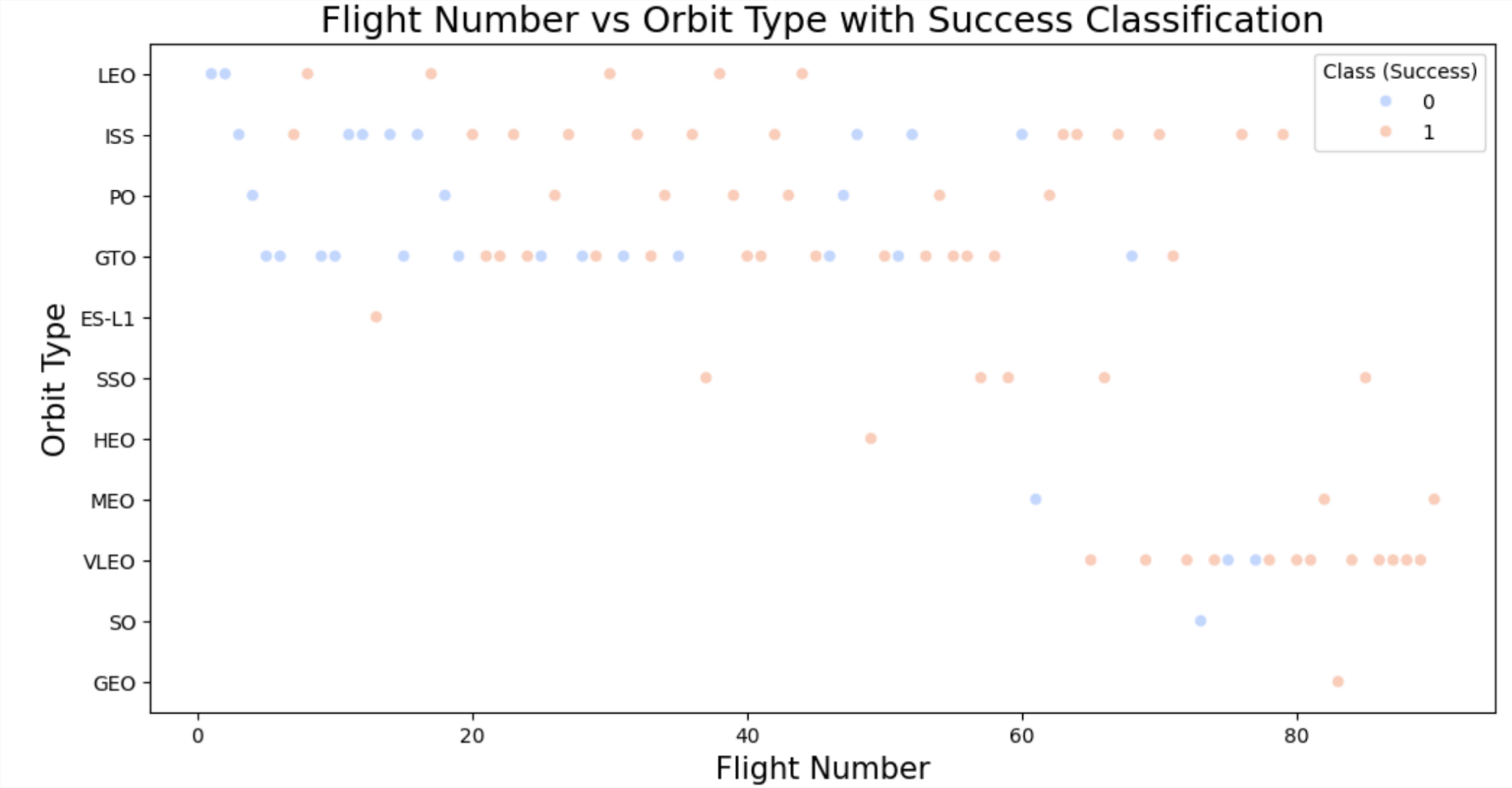
- Most launches carried payloads between 2000–6000 kg.
- All sites show both successes and failures.
- Heavier payloads still resulted in successful landings.

Success Rate vs. Orbit Type



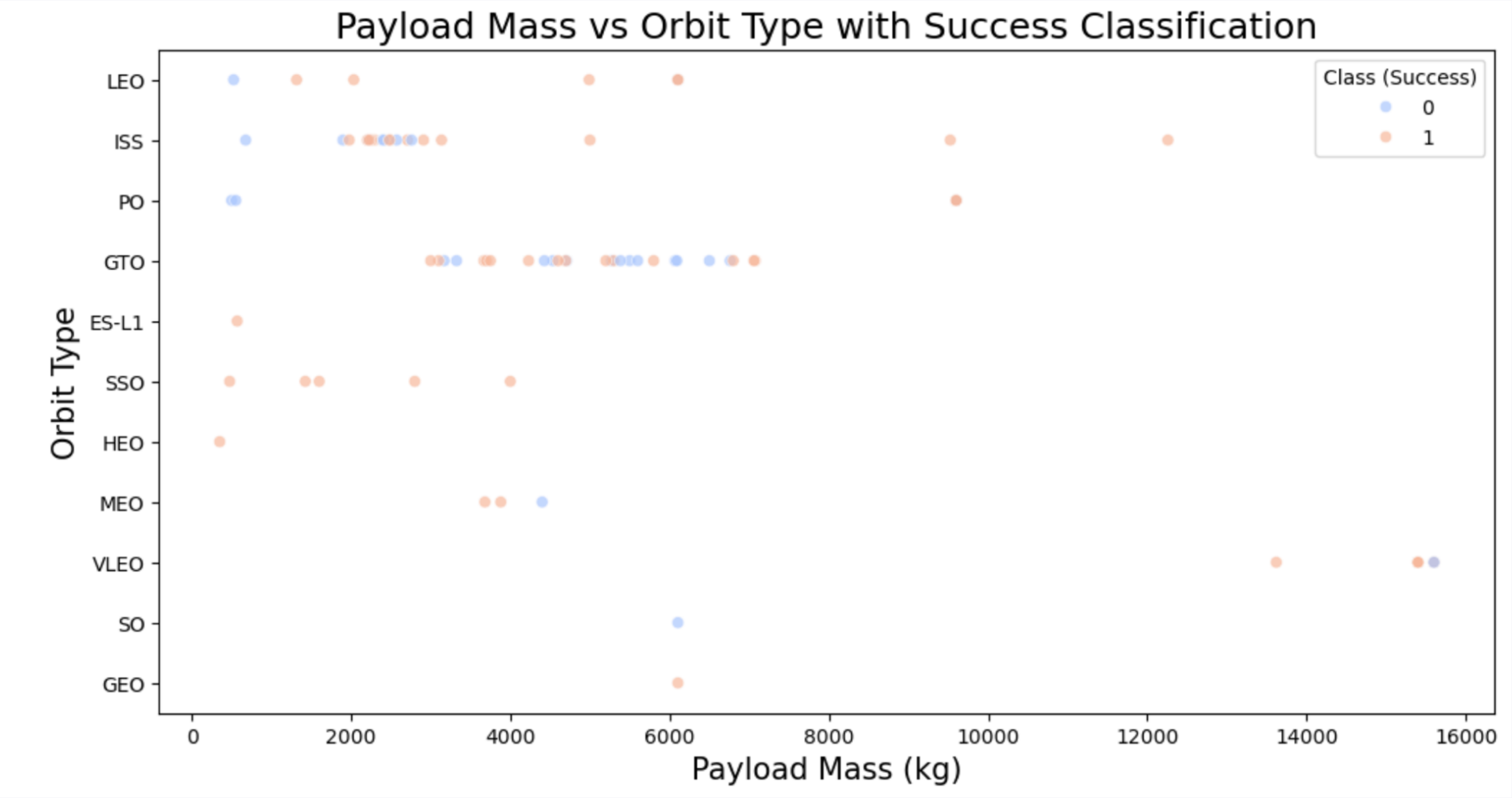
- Orbits like GEO, HEO, SSO, and ES-L1 show 100% success rate.
- GTO has the lowest success rate.
- Orbit type appears to impact landing outcome.

Flight Number vs. Orbit Type



- Later flights show improved success across all orbit types.
- GTO and ISS have the most mixed outcomes.
- Some orbits (SSO, HEO) show mostly successful launches.

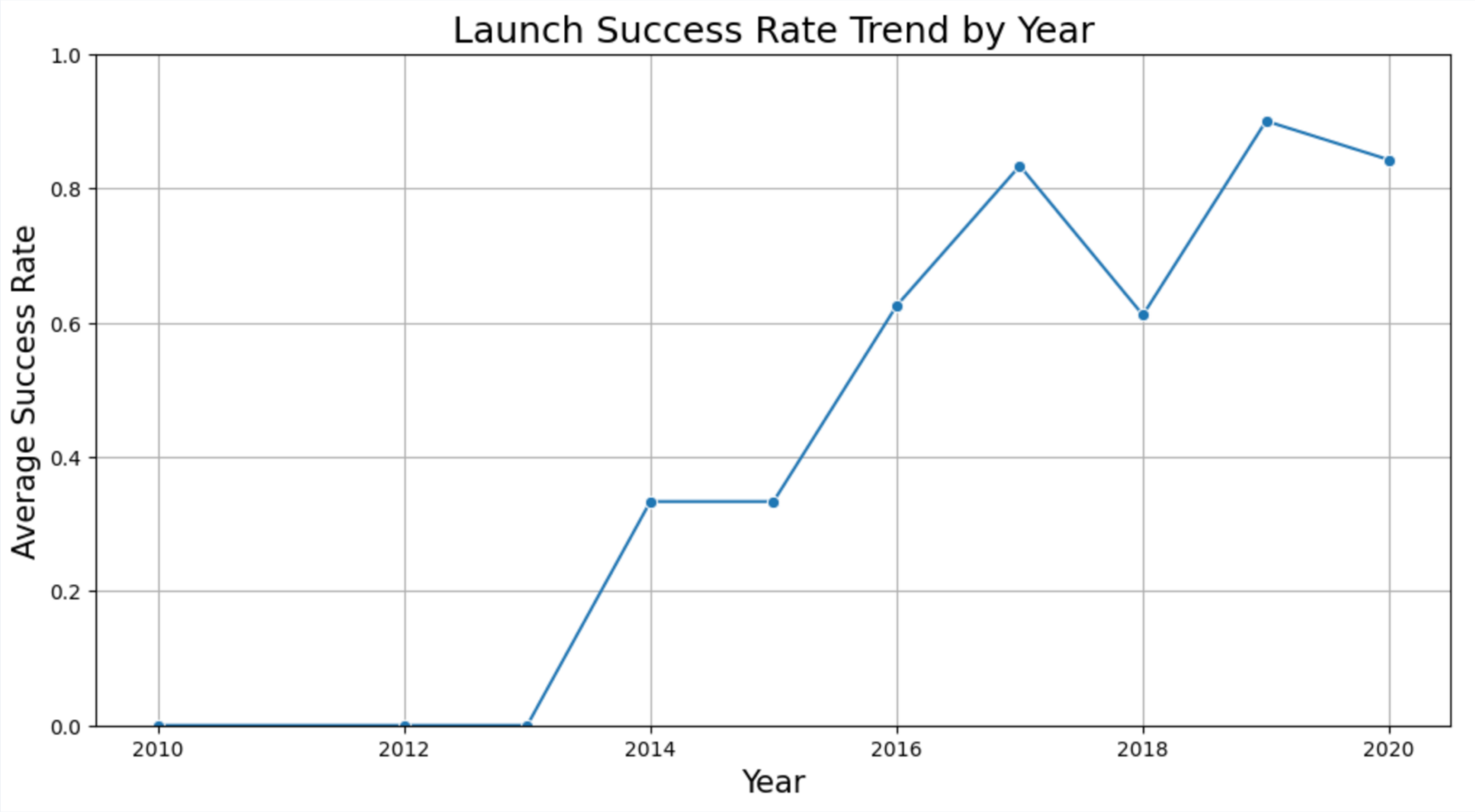
Payload vs. Orbit Type



- Show a scatter point of payload vs. orbit type
- Show the screenshot of the scatter plot with explanations

Launch Success Yearly Trend

- Success rate improved significantly after 2015.
- Peak success in 2019; trend remains high afterward.
- Indicates strong learning curve and technology improvement.



All Launch Site Names

- Retrieved distinct launch site names using SQL.
  Found 4 unique launch sites used in SpaceX missions:
- CCAFS LC-40
- CCAFS SLC-40
- VAFB SLC-4E
- KSC LC-39A

Launch Site Names Begin with 'CCA'

- Queried 5 records where launch site names start with 'CCA'.
  All records are from CCAFS LC-40, showing early Falcon 9 missions (2010–2013). These launches were mostly successful missions, but early landing attempts failed or were not attempted.

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

- Used SQL to calculate total payload mass delivered for NASA (CRS) missions.
- The result shows a total of **48,213 kg** carried by Falcon 9 boosters.
- This highlights the scale of NASA–SpaceX collaboration in cargo delivery.

Average Payload Mass by F9 v1.1

- Queried the average payload mass for booster version **F9 v1.1.**
  The result shows an average of **2,928.4 kg** per mission.
  This reflects the typical capacity handled during early Falcon 9 upgrades.

First Successful Ground Landing Date

- Retrieved the date of the first successful landing on a ground pad.
- The landing occurred on **December 22, 2015.**
- This marked a major milestone in SpaceX's reusability efforts.

Successful Drone Ship Landing with Payload between 4000 and 6000

- Queried boosters that successfully landed on a drone ship and carried payloads between 4000–6000 kg.
  Found the following booster versions:
- F9 FT B1022
- F9 FT B1026
- F9 FT B1021.2
- F9 FT B1031.2
- These results highlight reliable performance of the F9 FT class in medium-payload missions.

Total Number of Successful and Failure Mission Outcomes

- Queried total counts of mission outcomes. Results show:
- Success: 99 (including “Success” and “Success (payload status unclear)”)
- Failure (in flight): 1
- Overall success rate is extremely high, showing SpaceX’s operational reliability.

Boosters Carried Maximum Payload

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

- Retrieved booster versions that carried the maximum payload mass of 15,600 kg. Found 12 missions using F9 B5 series boosters, including:

- These results show that F9 B5 boosters were used for the heaviest missions.

2015 Launch Records

- Queried 2015 missions with Failure (drone ship) landing outcomes.
  Found 2 such missions, both from CCAFS LC-40 using F9 v1.1 boosters:

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

- Ranked landing outcomes by frequency between June 4, 2010 and March 20, 2017.
- Most missions during this period did not attempt landing. Drone ship attempts had mixed results, with both successes and failures.

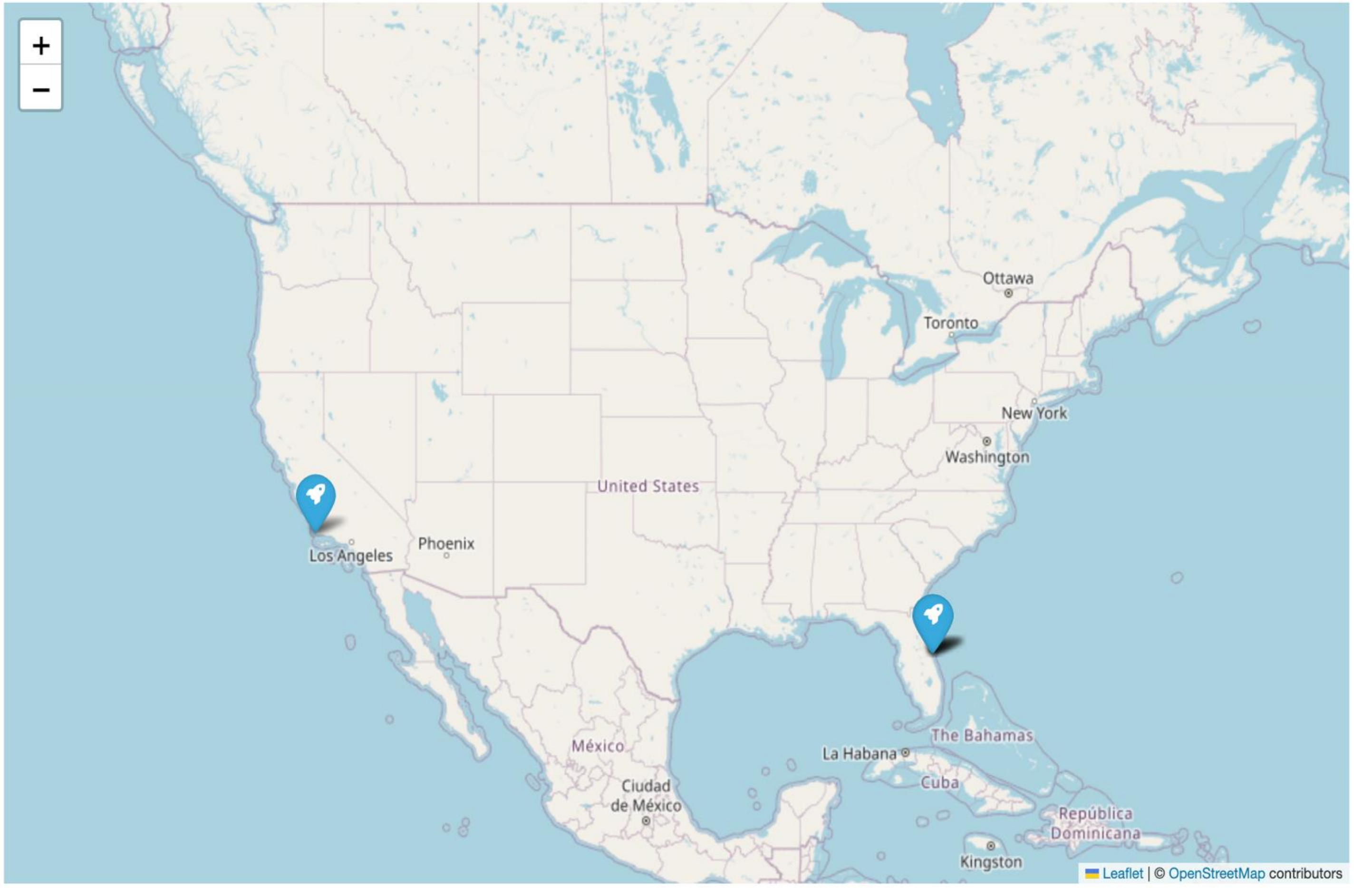
Landing_Outcome	Outcome_Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

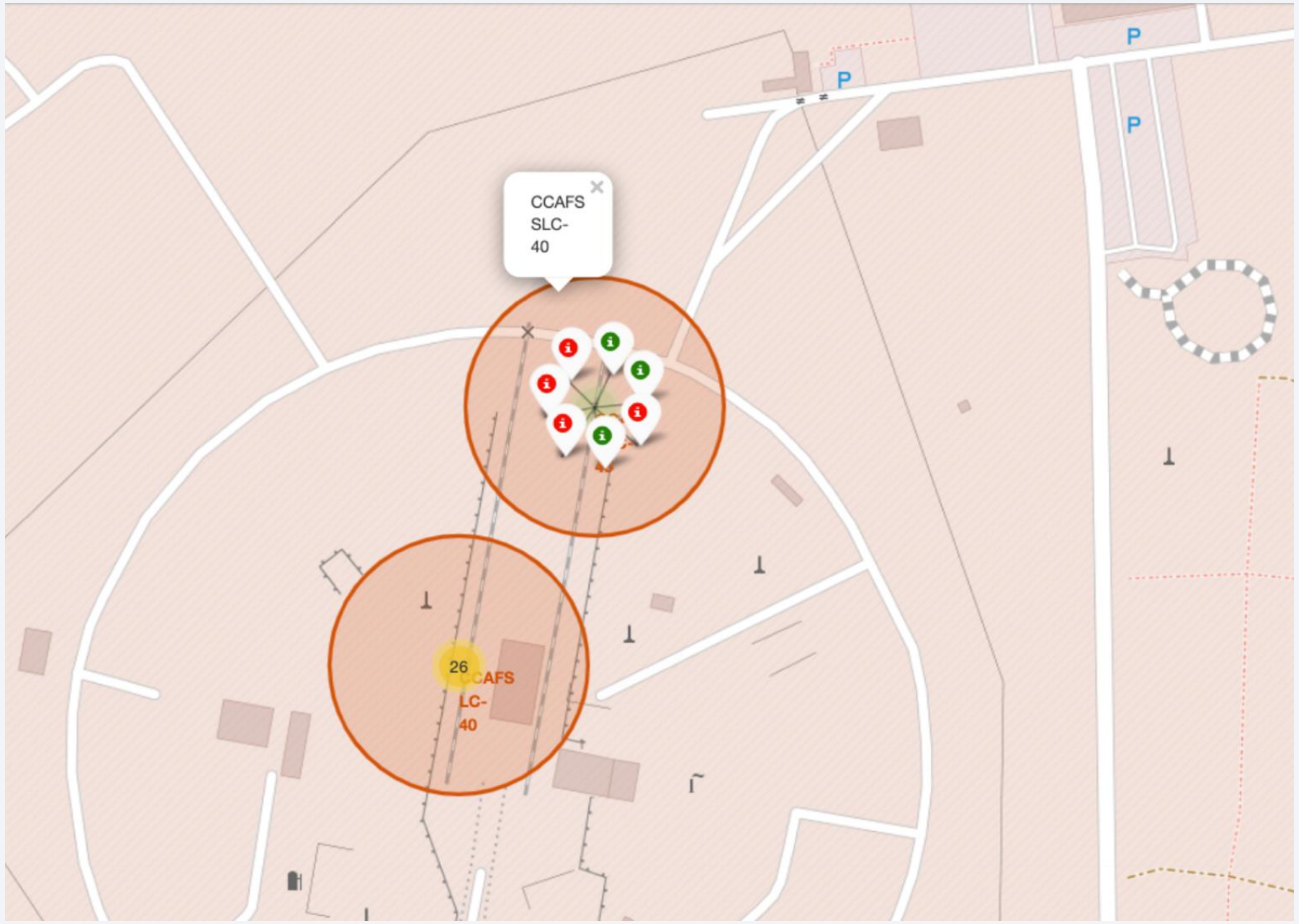
Launch Sites Proximities Analysis

Launch Site Locations in the United States

- This interactive map displays the geographical locations of SpaceX launch sites.
  Markers are placed at:
- Cape Canaveral, Florida
- Vandenberg Air Force Base, California
- These sites are key operational hubs for Falcon 9 and Falcon Heavy launches.
  The map provides a clear spatial overview of SpaceX's U.S. launch infrastructure.

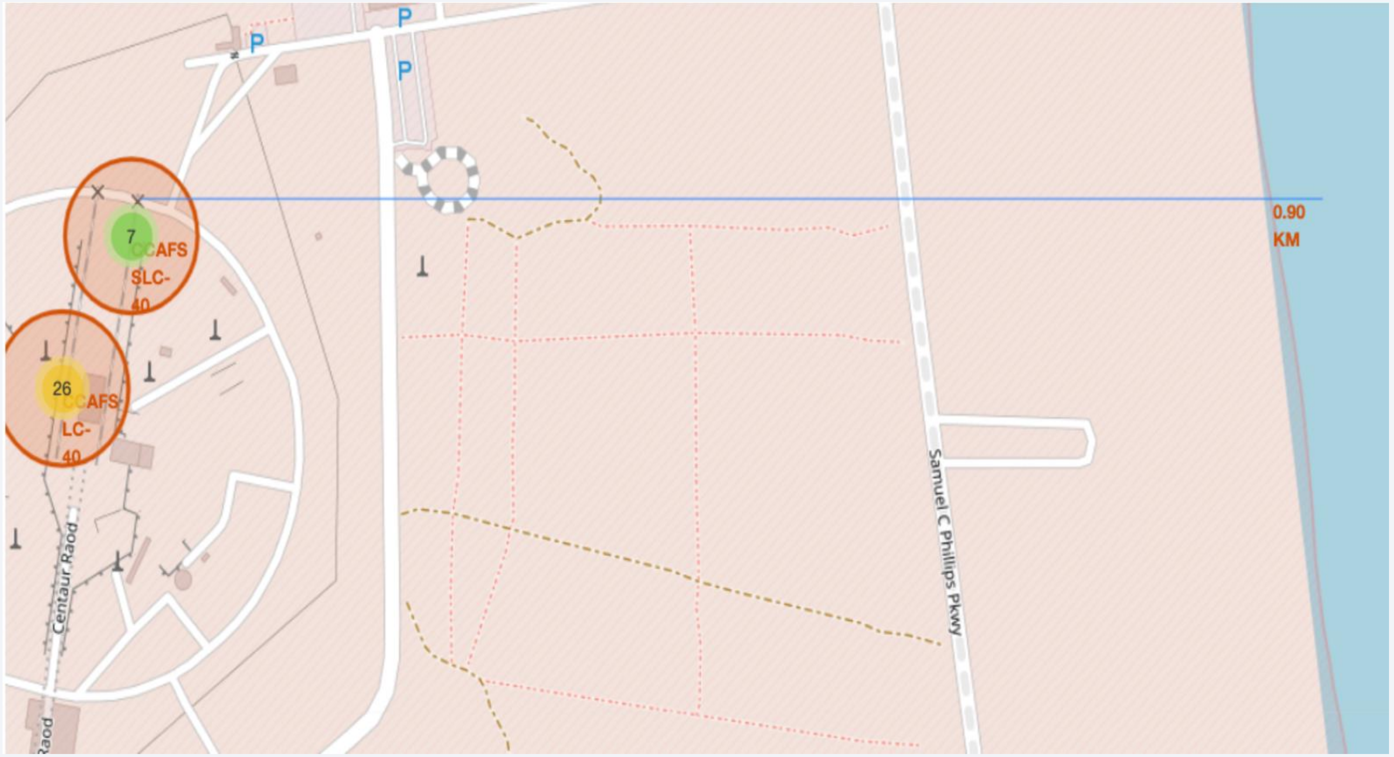


Launch Outcomes by Marker Color at CCAFS SLC-40



- This map visualizes launch outcomes at CCAFS SLC-40 using colored markers:
- Green = successful landing
- Red = failed landing
- The grouping shows multiple attempts from this site, with several successful landings over time.
- This visual helps identify historical performance at a specific launch pad.

Proximity to Coastline and Infrastructure near CCAFS SLC-40



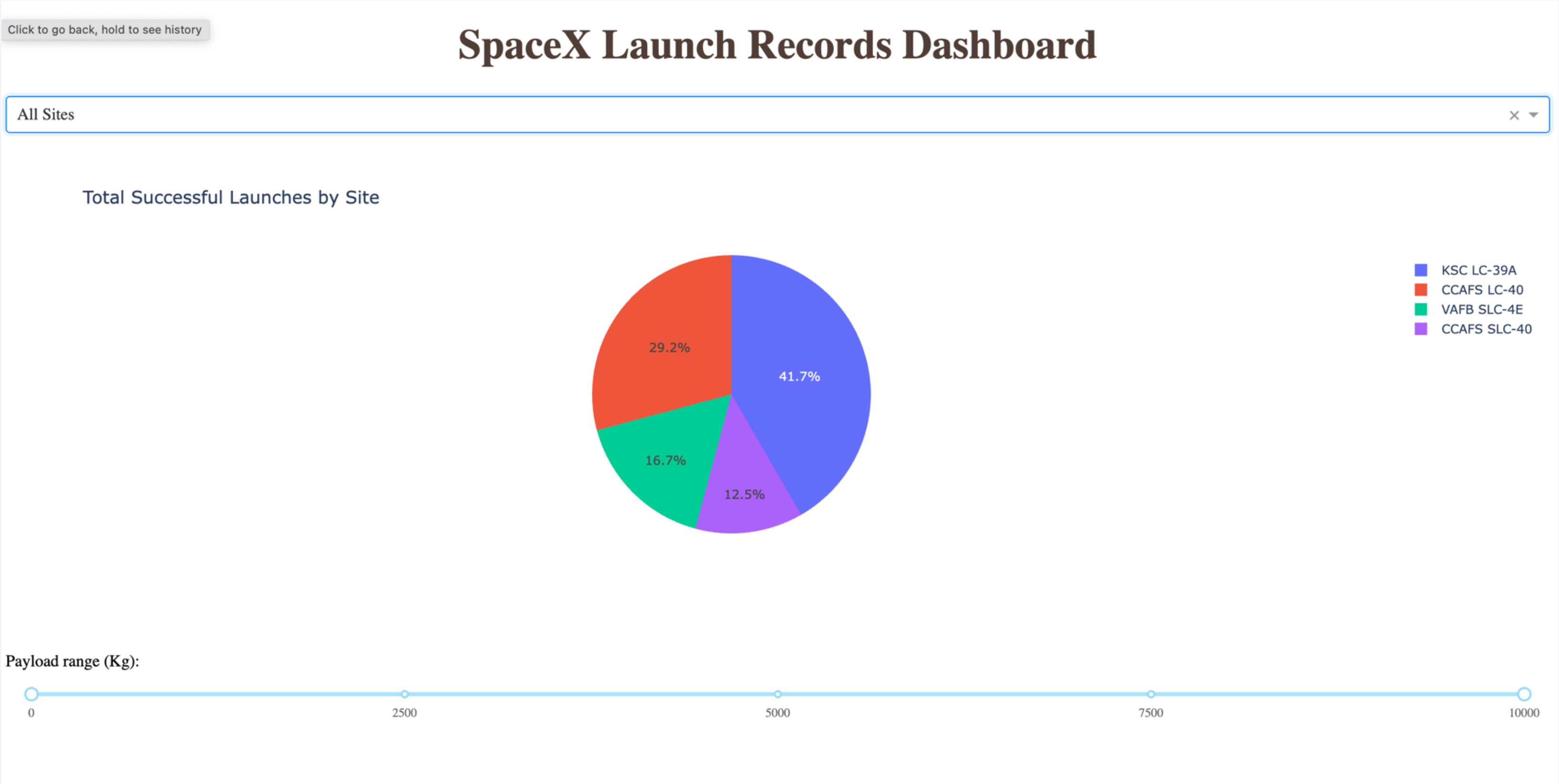
- This map highlights the spatial proximity of CCAFS SLC-40 to key surroundings:
- Distance to coastline is ~0.90 km
- Nearby road infrastructure includes Samuel C Phillips Parkway
- Understanding such proximity is crucial for launch safety, logistics, and recovery planning.

Section 4

Build a Dashboard with Plotly Dash

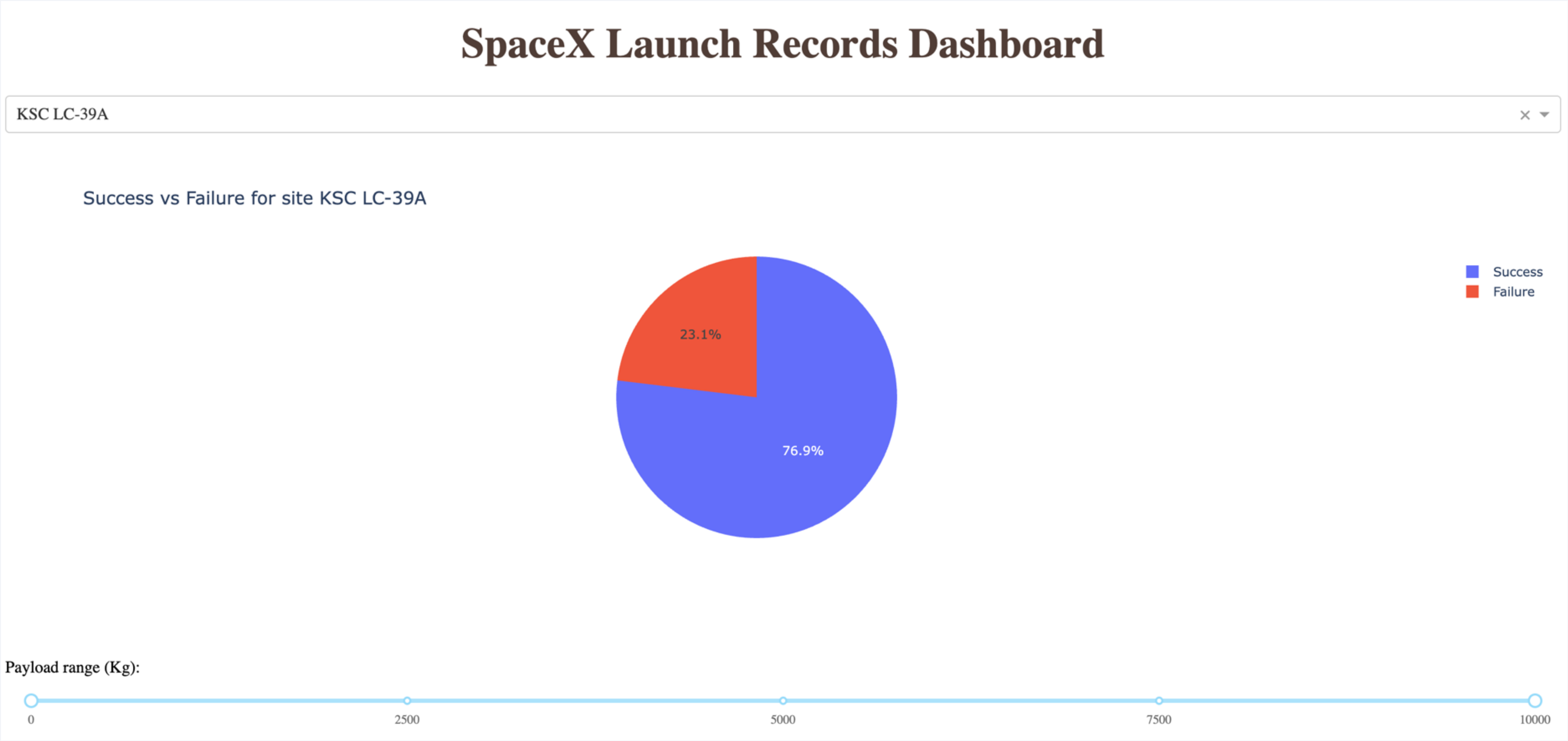
Launch Success Count by Site (All Sites)

- This pie chart shows the distribution of successful launches across all launch sites.
- KSC LC-39A had the most successes (41.7%), followed by CCAFS LC-40 (29.2%).
- The chart gives a quick overview of launch site performance.

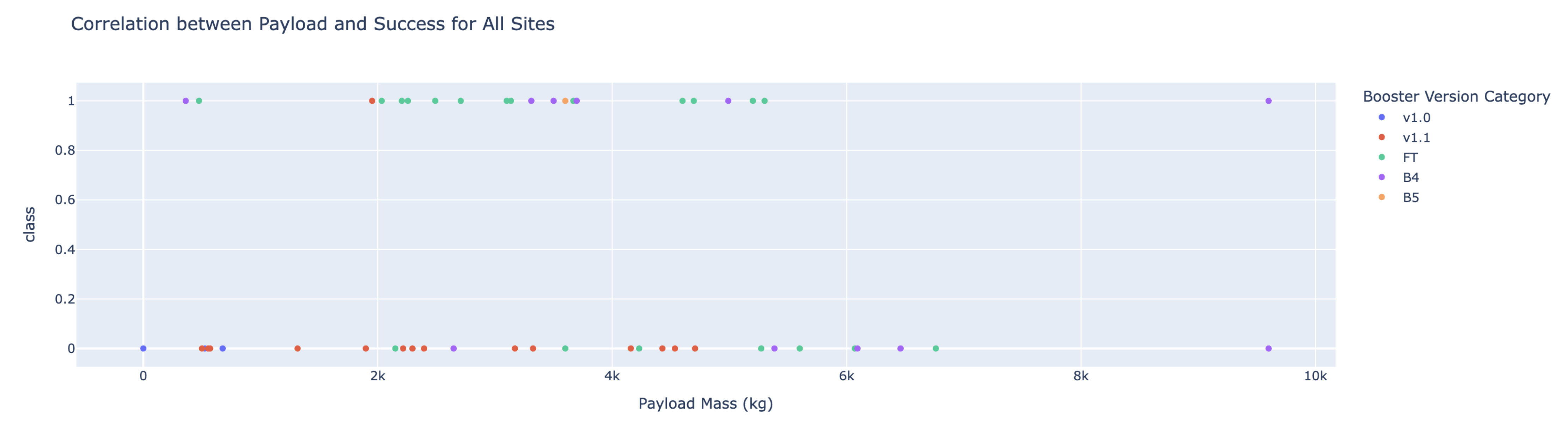


Success vs Failure for KSC LC-39A

- This pie chart shows the launch outcomes for KSC LC-39A, one of SpaceX's most active sites.
- 76.9% of launches were successful,
  while 23.1% failed.
- The high success rate highlights strong operational performance at this location.



Payload vs. Launch Outcome by Booster Version (All Sites)



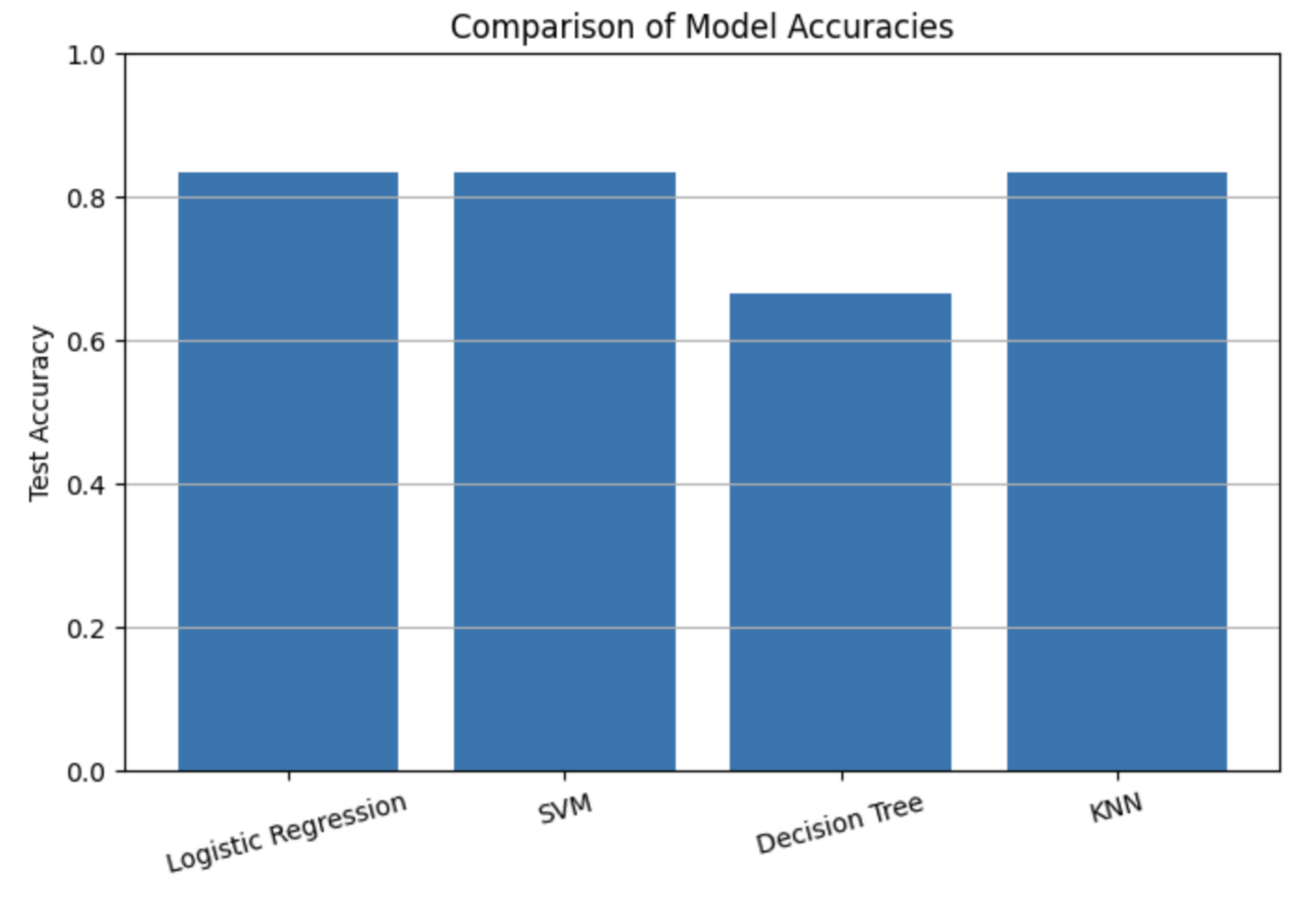
- This scatter plot visualizes launch outcomes based on payload mass and booster version.
- Successes (class 1) are more common in the 2000–6000 kg range.
- Booster versions FT and B5 appear frequently in successful missions.

Section 5

Predictive Analysis (Classification)

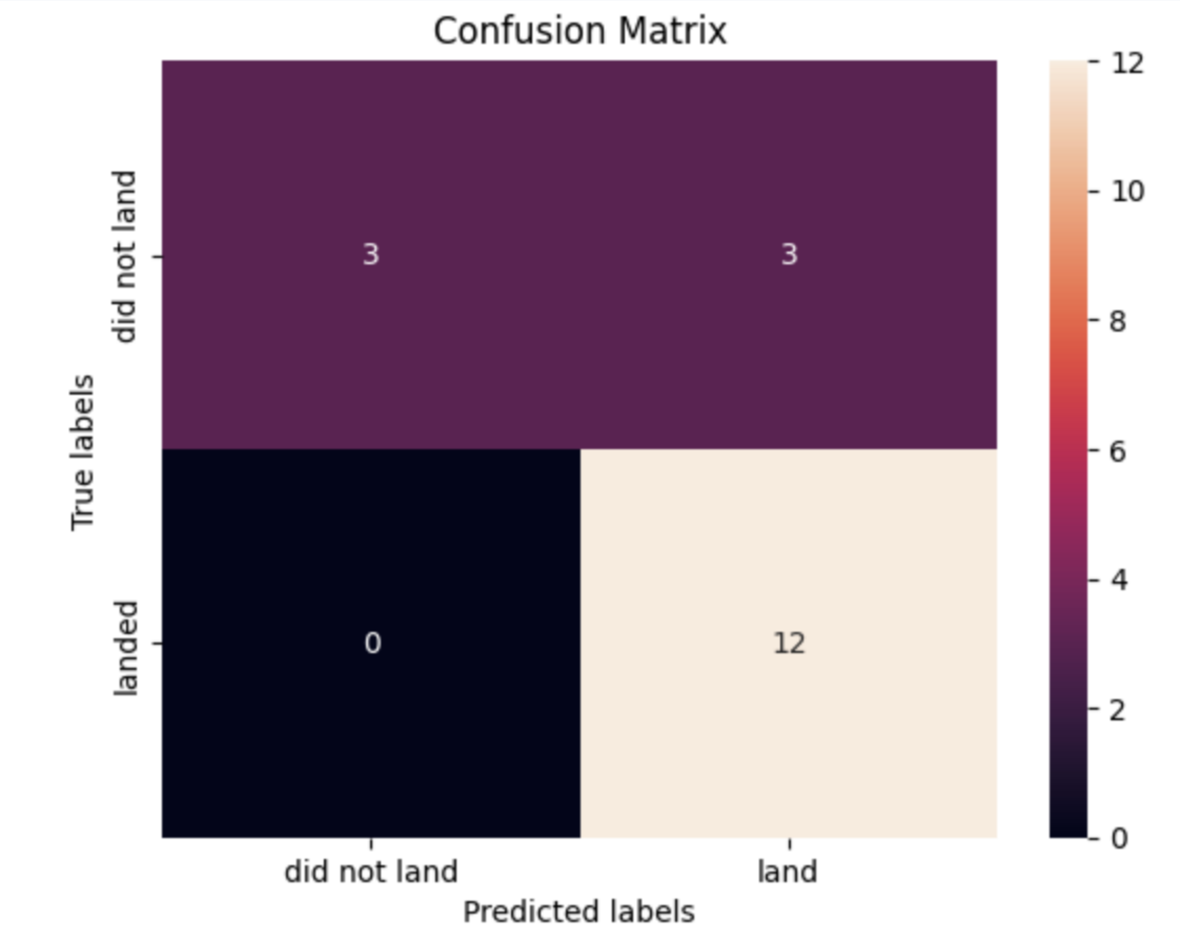
Classification Accuracy

- This bar chart compares test accuracies of four classification models.
- Logistic Regression, SVM, and KNN all achieved top performance (~84%)
- Decision Tree lagged behind with ~68% accuracy
- Overall, multiple models performed well, with Logistic Regression, SVM, and KNN being equally effective.



Confusion Matrix

- The top 3 models — Logistic Regression, SVM, and KNN — achieved identical performance.
- All three produced the same confusion matrix:



This indicates that these models made the same classification decisions, with identical strengths and errors.

Conclusions

- Collected SpaceX launch data via API and web scraping
- Cleaned, visualized, and explored data using Pandas, SQL, and Folium
- Built interactive dashboard with Plotly Dash to explore payload and site patterns
- Trained classification models to predict landing success
- Logistic Regression, SVM, and KNN achieved highest accuracy (~84%)
- Data showed launch site, payload, and booster type strongly influence success
- The analysis helps understand what factors lead to successful Falcon 9 landings and can support future launch planning.

Appendix

- All important resources are available in the project's GitHub repository, including notebooks, SQL queries, visualizations, dashboard code, and model results.
- GitHub:
- https://github.com/lorumo/SpaceY

Thank you!