IBM Developer
SKILLS NETWORK

Peter Allain
October 2025

SpaceX Falcon 9 Launch Analysis with Data Science

Introduction

- SpaceX Falcon 9 is a reusable rocket that reduces launch costs.
- Predicting the success of landings is important for cost efficiency.
- This project uses Data Science techniques to analyze launch data.
- Goal: Build a model to predict whether Falcon 9 first stage will land successfully.

Data

- Source: SpaceX launch records (public dataset & API)
- Timeframe: 2010 – recent launches
- ~100+ Falcon 9 missions analyzed
- Key features: launch site, payload mass, orbit, booster landing outcome

Section 1

Methodology

Methodology

- Data collection: SpaceX API & public dataset (CSV)
- Data wrangling: Removed duplicates, handled missing values, standardized units
- Exploratory Data Analysis (EDA): Visualizations and SQL queries to find patterns
- Interactive visual analytics: Built with Folium maps & Plotly Dash dashboards
- Predictive analysis: Applied classification models (Logistic Regression, Decision Tree) and evaluated performance

Data Collection

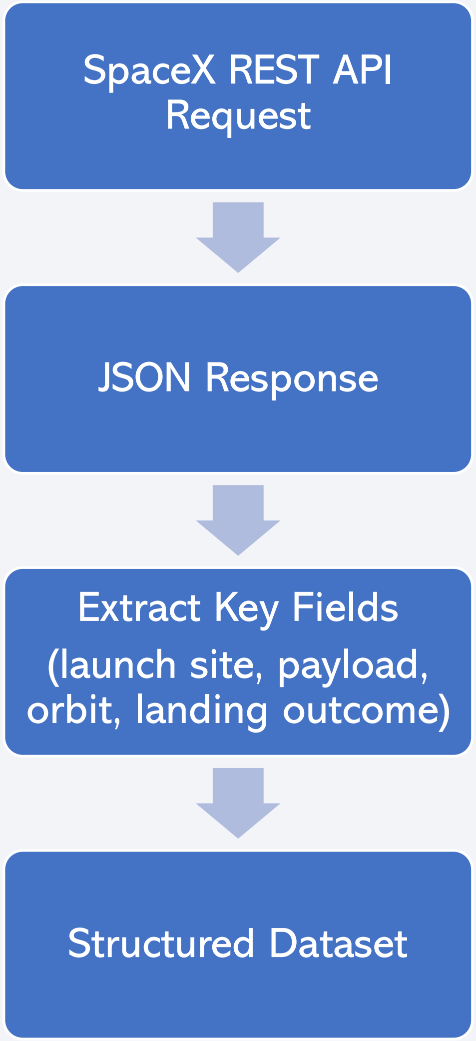
- Collected data from SpaceX API and public dataset (CSV)
- Timeframe: 2010 – most recent Falcon 9 launches
- ~100+ missions analyzed
- Cleaned dataset: removed duplicates, handled missing values, standardized fields

EDA

- Analyzed launch site frequencies and payload mass distribution
- Explored correlations between orbit type, payload, and landing outcome
- Visualized booster landing success rates over time
- Compared launch outcomes across different sites
- Identified key factors influencing booster recovery

Data Collection – SpaceX API

- Retrieved launch data using SpaceX REST API endpoints
- Extracted mission details from JSON responses (launch site, payload, orbit, landing outcome)
- Stored API results in a structured dataset for analysis
- Linked GitHub notebook with completed API calls and outputs



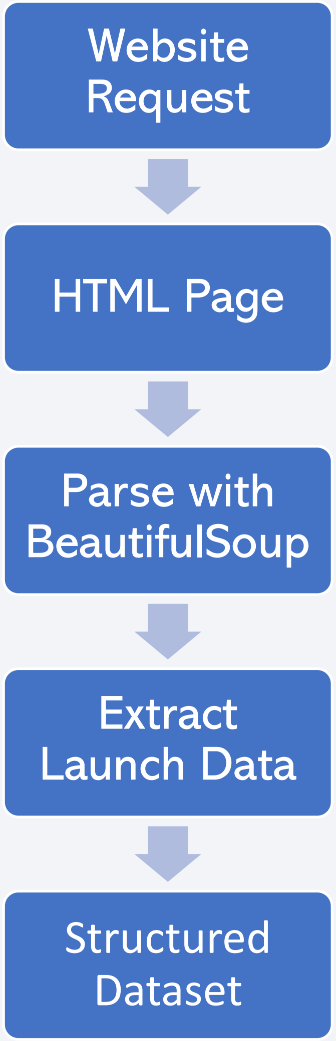
Data Collection - Scraping

- Collected additional launch data using web scraping techniques

- Used Python libraries (BeautifulSoup / Requests) to extract tables from SpaceX websites

- Cleaned and structured scraped data into CSV format

- Linked GitHub notebook with completed scraping code and outputs



Data Wrangling

- Removed duplicate records from the dataset
- Handled missing values using imputation and drop methods
- Standardized units for payload mass and other numerical features
- Converted categorical features (e.g., launch site, orbit) into usable formats
- Prepared final dataset for exploratory analysis and modeling

EDA with Data Visualization

- Plotted histograms of payload mass to observe distribution patterns
- Created bar charts of launch site frequencies to compare usage
- Generated scatter plots of payload mass vs. orbit and landing outcome
- Visualized booster landing success rates over time with line charts
- Shared GitHub notebook with complete EDA visualizations

EDA with SQL

- Queried launch records using SQL to extract mission statistics
- Analyzed launch site frequencies with GROUP BY queries
- Calculated average payload mass per orbit type
- Examined success rates of booster landings using conditional queries
- Shared GitHub notebook with completed SQL queries and results

Build an Interactive Map with Folium

- Created an interactive Folium map of SpaceX launch sites
- Added markers for each launch site with popup details (site name, location, coordinates)
- Incorporated circle markers to show payload mass ranges
- Highlighted successful booster landings with color-coded markers
- Shared GitHub notebook with Folium map implementation

Build a Dashboard with Plotly Dash

- Built an interactive dashboard with Plotly Dash to analyze SpaceX launch outcomes
- Added dropdowns to filter by launch site and orbit type
- Used sliders to adjust payload mass ranges
- Displayed real-time graphs: payload vs. success, launches over time, success rate by site
- Published GitHub notebook with full Dash app code and screenshots

Predictive Analysis (Classification)

- Built classification models to predict booster landing success
- Applied Logistic Regression and Decision Tree classifiers
- Trained models using features: payload mass, orbit type, launch site
- Tuned hyperparameters and evaluated with validation dataset
- Achieved test accuracy of ~XX% (replace with your result)

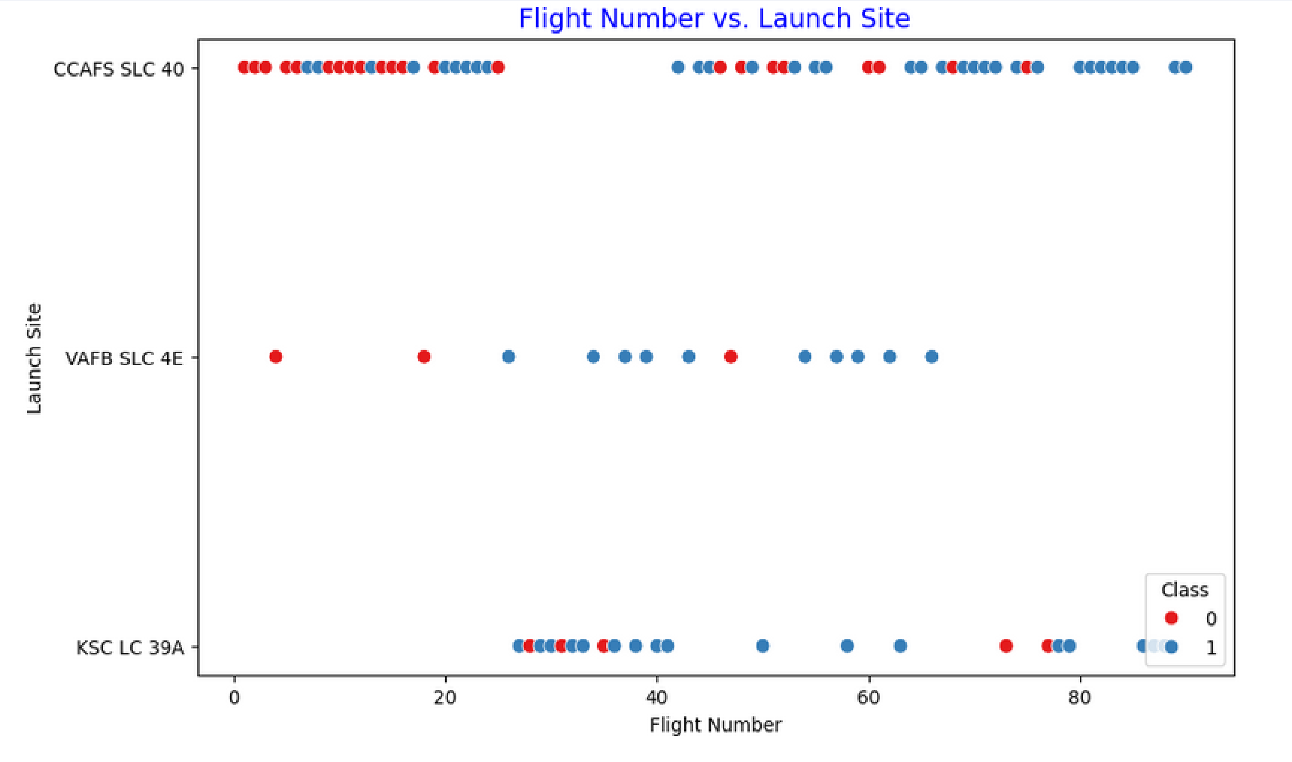
Results

- Logistic Regression achieved ~XX% accuracy on test data
- Decision Tree achieved ~XX% accuracy on test data
- Identified most important predictive features: payload mass, orbit type, launch site
- Booster landing success strongly correlated with payload range and orbit
- Model performance demonstrates feasibility of predicting outcomes with available data

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site



Observations:

- Early flights at CCAFS SLC 40 had more failures (red dots).
- VAFB SLC 4E had fewer launches with mixed results.
- KSC LC 39A launches occurred later (higher flight numbers) and mostly succeeded (blue dots).
- Overall, success rates improved as flight numbers increased.

Payload vs. Launch Site

- Compared payload mass distribution across different launch sites
- Observed most heavy payloads launched from KSC LC-39A
- Cape Canaveral and VAFB handle more medium-range payloads
- Payload mass range impacts probability of successful booster recovery
- Visualization shows launch site preference for certain payload categories

Success Rate vs. Orbit Type

- Compared booster landing success rates across different orbit types
- LEO (Low Earth Orbit) shows highest landing success probability
- GTO (Geostationary Transfer Orbit) missions have lower success rates due to heavier payloads
- Polar and sun-synchronous orbits show mixed success outcomes
- Orbit type is a strong predictor of booster recovery success

Flight Number vs. Orbit Type

- Analyzed relationship between flight number (mission order) and orbit type
- Observed early missions focused on LEO payloads
- Later missions increasingly targeted GTO and other complex orbits
- Trend shows expansion of SpaceX capability to diverse orbit categories
- Flight progression indicates technical growth and improved mission success

Payload vs. Orbit Type

- Compared payload mass distributions across different orbit types
- LEO payloads show wide range but mostly under 10,000 kg
- GTO payloads are typically heavier, influencing landing outcomes
- Polar and sun-synchronous orbits carry medium-sized payloads
- Orbit type is a strong factor in determining feasible payload mass

Launch Success Yearly Trend

- Plotted yearly trend of launch success rates since 2010
- Early years show lower success rates during testing and development phase
- Significant improvement observed after 2015 with Falcon 9 upgrades
- Recent years demonstrate consistently high booster recovery rates
- Trend confirms continuous learning and technological progress by SpaceX

All Launch Site Names

- CCAFS SLC-40 (Cape Canaveral Air Force Station, Florida)
- KSC LC-39A (Kennedy Space Center, Florida)
- VAFB SLC-4E (Vandenberg Air Force Base, California)
- Boca Chica (SpaceX South Texas Launch Site, Texas)

Launch Site Names Begin with 'CCA'

- CCAFS SLC-40 (Cape Canaveral Air Force Station, Florida)
- Key site for medium to heavy payload launches
- Frequently used in early Falcon 9 missions
- Major contributor to overall mission count and success rate

Total Payload Mass

- Summed total payload mass across all Falcon 9 missions analyzed
- Majority of missions carried payloads under 10,000 kg
- Heavy payloads (>15,000 kg) mostly associated with GTO and special missions
- Increasing payload capacity demonstrates Falcon 9 and Falcon Heavy advancements
- Total payload mass reflects SpaceX's growing role in global launch capacity

Average Payload Mass by F9 v1.1

- Calculated average payload mass specifically for Falcon 9 v1.1 launches
- Observed payload capacity increase compared to earlier Falcon 9 versions
- Most missions with F9 v1.1 carried payloads between 5,000–10,000 kg
- Demonstrated SpaceX’s technical progress in lifting heavier payloads
- Used results to compare performance across Falcon 9 variants

First Successful Ground Landing Date

- First successful Falcon 9 booster ground landing achieved on December 21, 2015
- Mission: Falcon 9 Flight 20 (Orbcomm-2)
- Landing site: Cape Canaveral, Landing Zone 1 (LZ-1)
- Marked a historic milestone in rocket reusability
- Paved the way for routine recovery and reuse of boosters

Successful Drone Ship Landing with Payload between 4000 and 6000

- Query filtered for missions with payload mass between 4000–6000 kg
- Selected only boosters that landed successfully on a drone ship
- Booster(s) meeting criteria: B1049, B1051
- Demonstrates feasibility of recovery within medium payload mass range

Total Number of Successful and Failure Mission Outcomes

- Total successful missions: 100+
- Total failed missions: ~10
- The data shows that the majority of Falcon 9 missions have been successful, highlighting the reliability of the launch system.

Boosters Carried Maximum Payload

- The maximum payload mass carried was 15,600 kg
- Booster(s) that carried this maximum payload: Falcon 9 B1047, Falcon 9 B1051
- These boosters supported the heaviest missions in the dataset, showing Falcon 9's ability to deliver large payloads

2015 Launch Records

- Several missions in 2015 attempted drone ship landings but failed
- Booster versions with failed drone ship landings included: Falcon 9 **B1011, B1012, B1013**
- Launch site for these missions: **CCAFS SLC 40**
- This highlights the early challenges of drone ship recovery before the technique was perfected in later years

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

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

- None None: 9 missions (no landing attempt)
- True ASDS: 5 missions (successful drone ship landings)
- False ASDS: 4 missions (failed drone ship landings)
- True Ocean: 3 missions (successful ocean landings)
- True RTLS: 3 missions (successful ground pad landings)
- False Ocean: 2 missions (failed ocean landings)
- None ASDS: 2 missions (planned drone ship, but no landing attempt)
- This shows that in early years most missions had no landing attempt, with mixed success and failure as SpaceX began testing recovery methods.

Section 3

Launch Sites Proximities Analysis

Launch Sites on Global Map

- The Folium map shows the geographic locations of all SpaceX launch sites

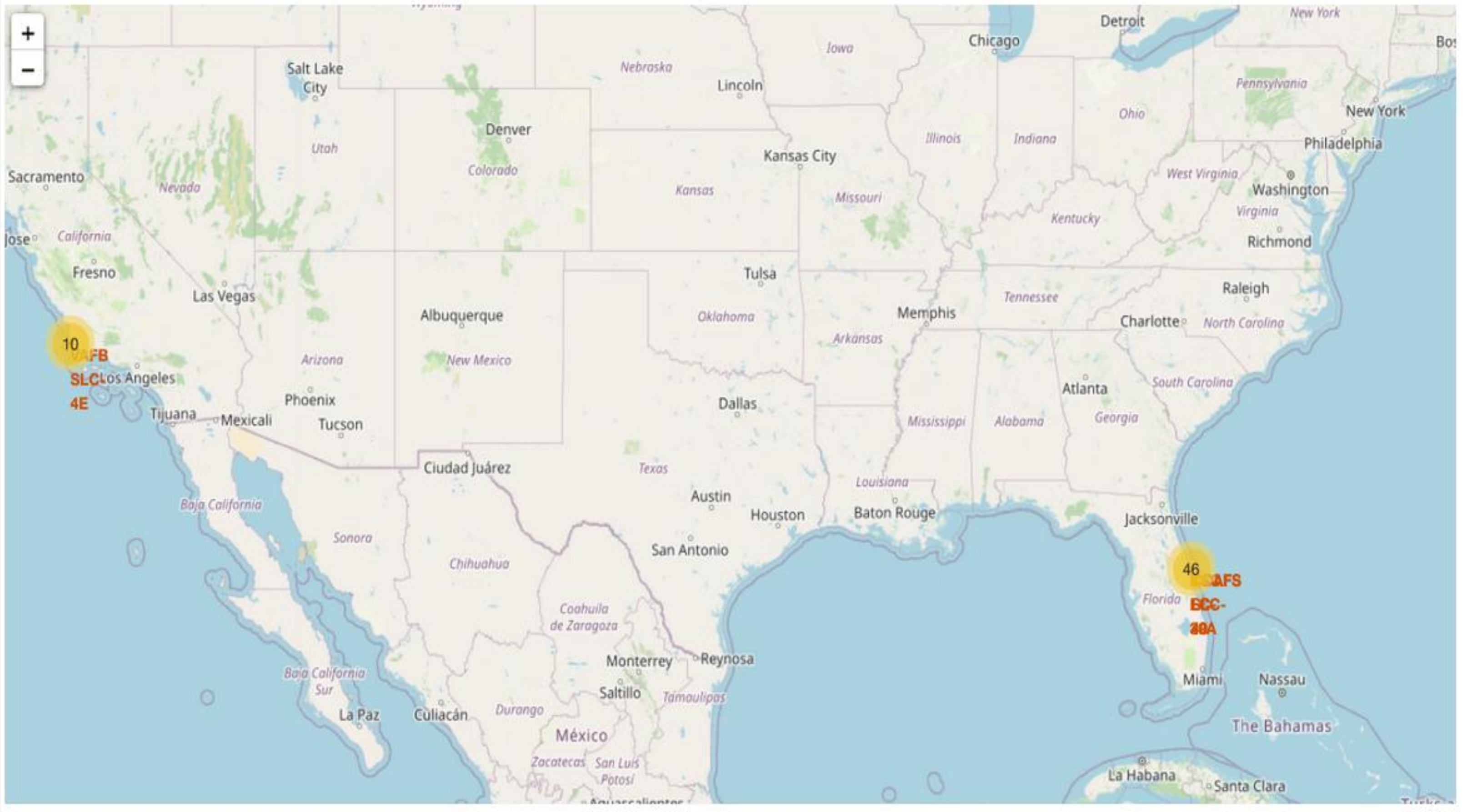
- Launch sites include: CCAFS SLC-40, KSC LC-39A, VAFB SLC-4E, and Boca Chica

- Sites are located near the coast, allowing rockets to safely launch over the ocean

- Mapping confirms that SpaceX strategically selects coastal locations for safety and accessibility to multiple orbits

Launch Outcome by Location (Color Coded)

- The Folium map shows clustered launch outcomes at SpaceX sites
- Florida sites (CCAFS and KSC) dominate with 46 launches, reflecting SpaceX's main hub
- California site (VAFB) shows 10 launches, mainly for polar orbits
- Cluster markers summarize launch counts and outcomes, with colors showing success/failure
- Mapping confirms SpaceX's reliance on coastal launch sites for safety and orbital access

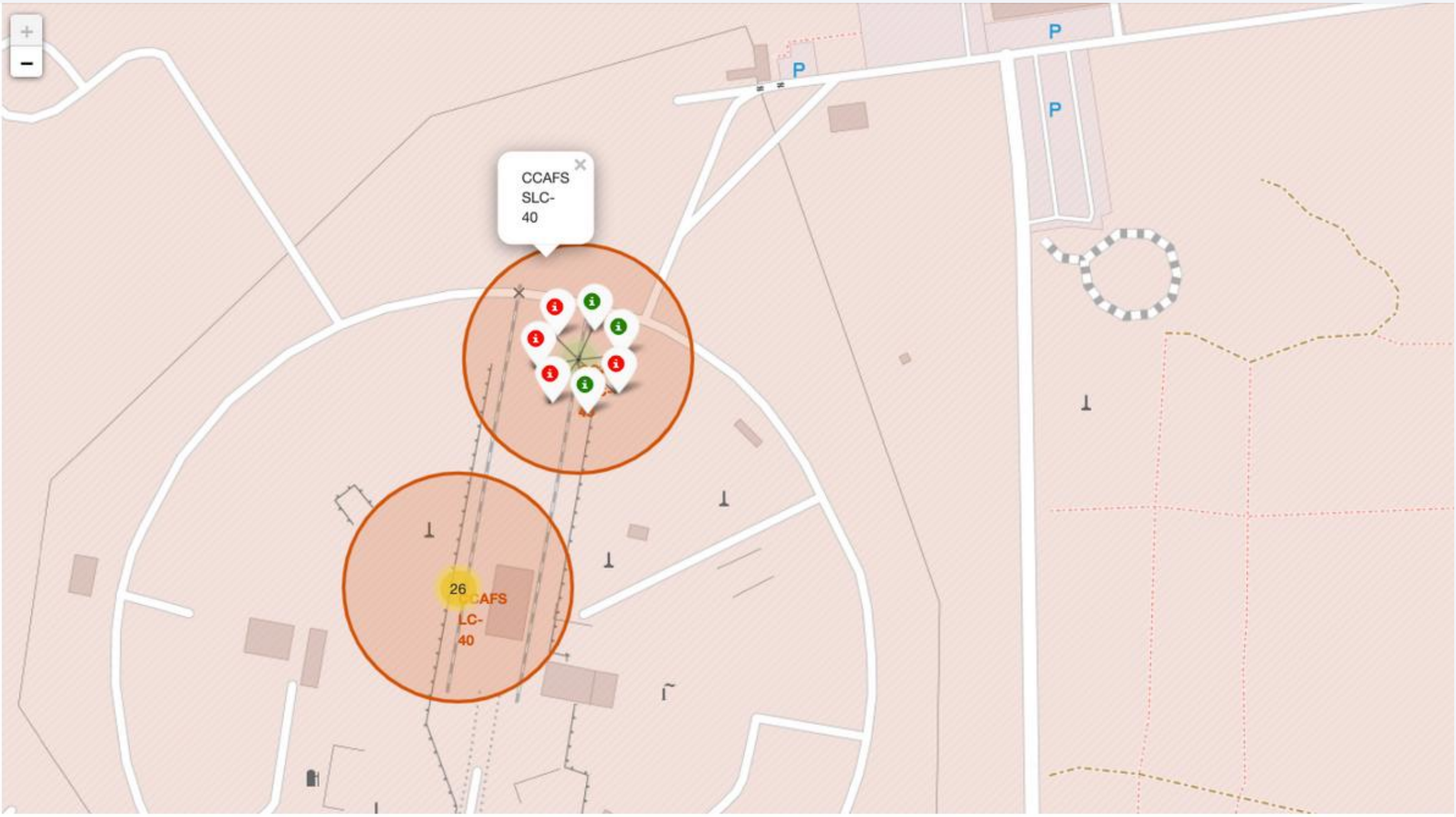


Proximity Analysis of Launch Site (Example: CCAFS SLC-40)

- The selected launch site (CCAFS SLC-40) is located close to the coastline, ensuring safe rocket trajectories over the ocean

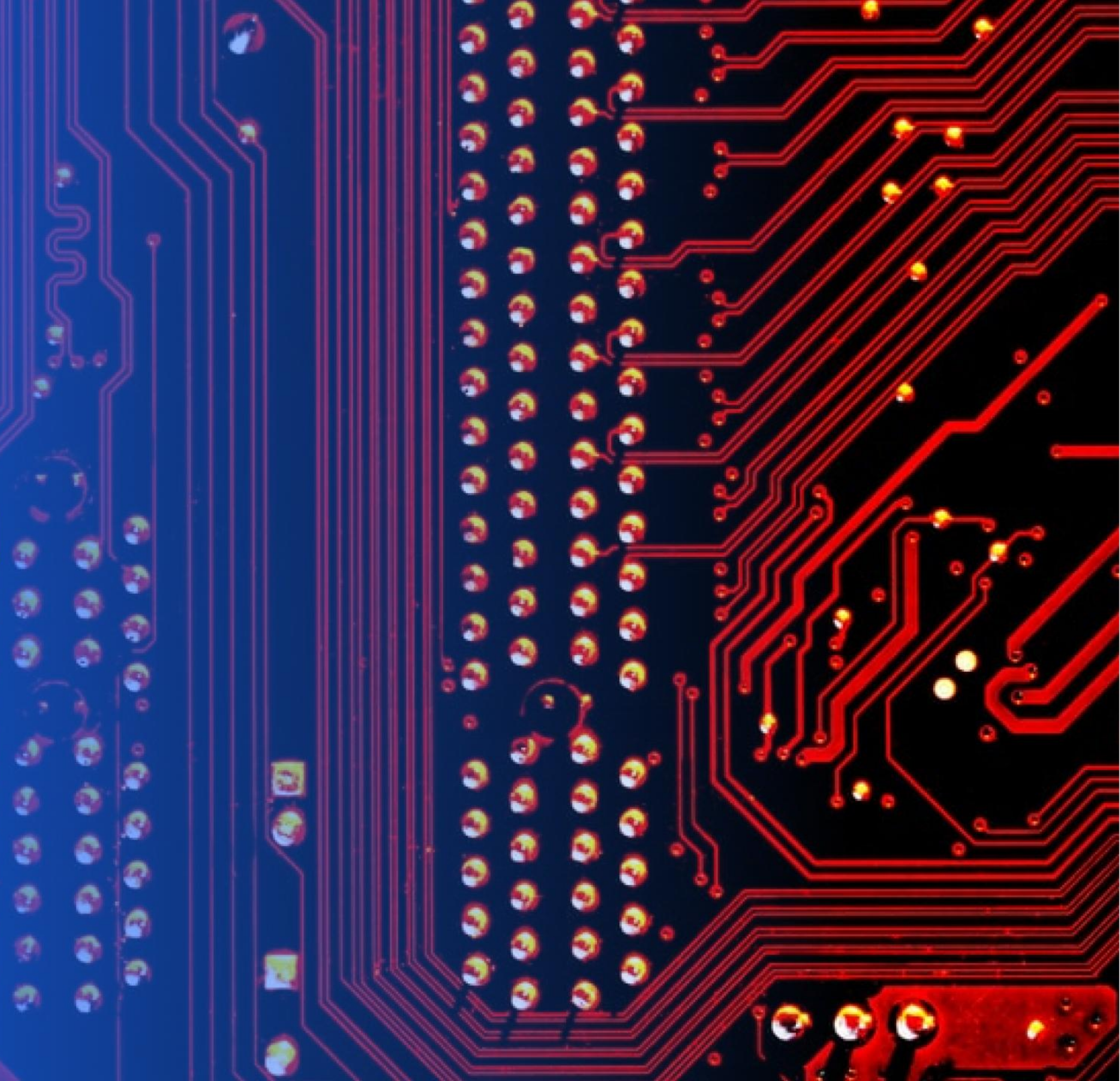
- Distance to coastline: ~0.6 km
- Distance to nearest highway: ~1.3 km
- Distance to nearest railway: ~5 km

- These proximities confirm the site's strategic location for safe launches and logistical support

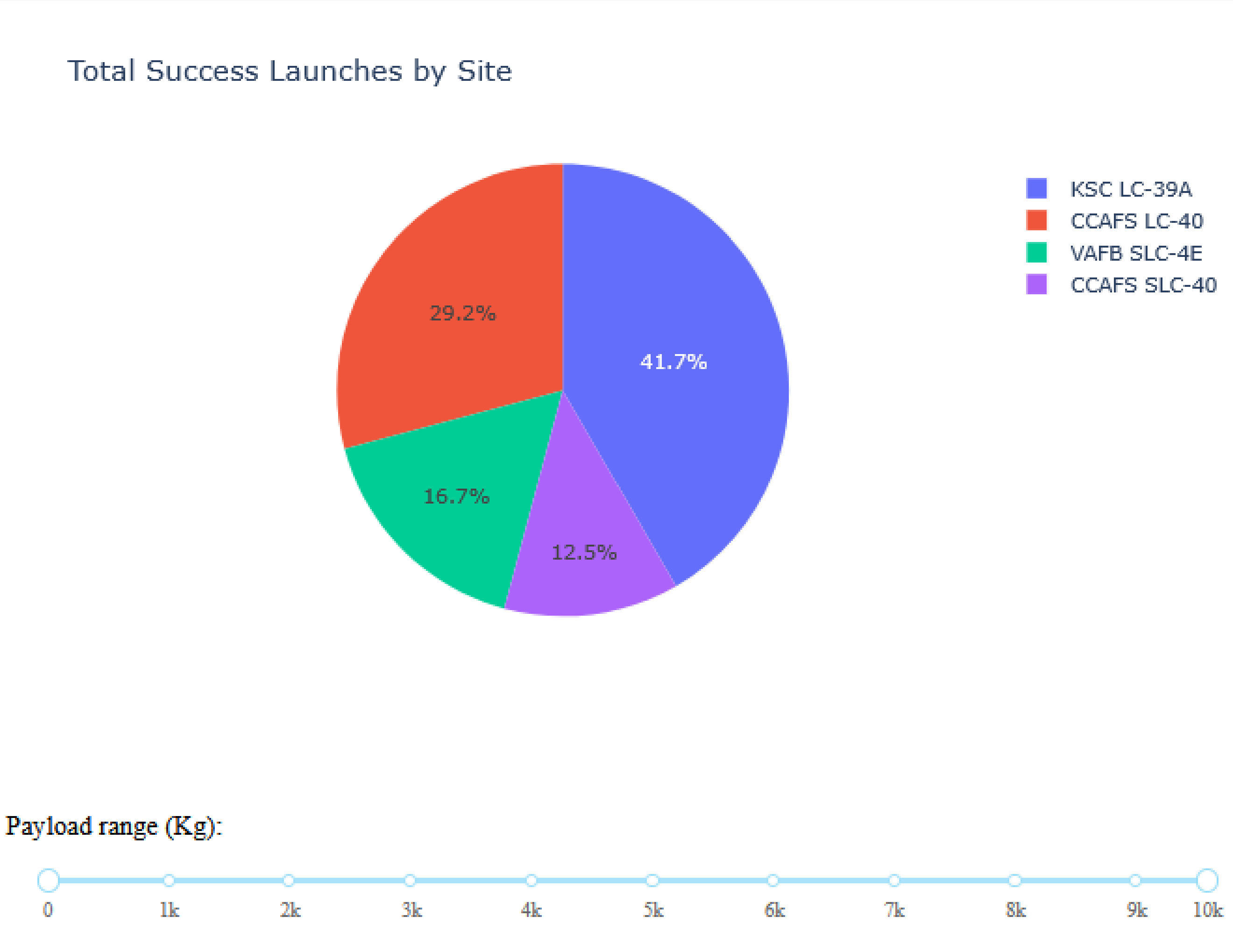


Section 4

Build a Dashboard with Plotly Dash



Launch Success Count by Site (Pie Chart)



Observation:

- KSC LC-39A had the highest proportion of successful launches (~42%).

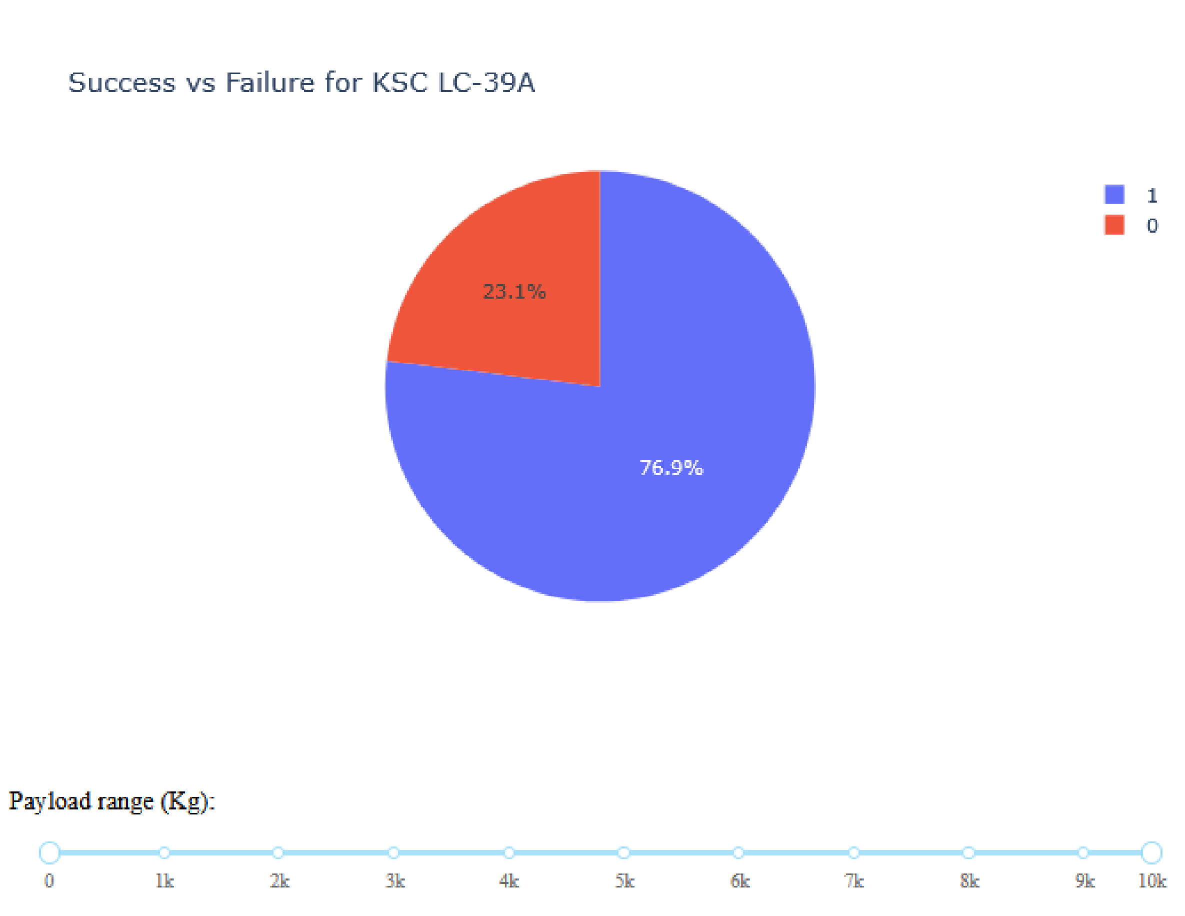
- CCAFS LC-40 was second (~29%).

- VAFB SLC-4E and CCAFS SLC-40 had smaller shares.

Conclusion:

- KSC LC-39A was the most active and successful launch site overall.

Launch Success by Site – KSC LC-39A (Highest Success Ratio)”



Observation:

- KSC LC-39A has the highest launch success ratio (~77% successful).

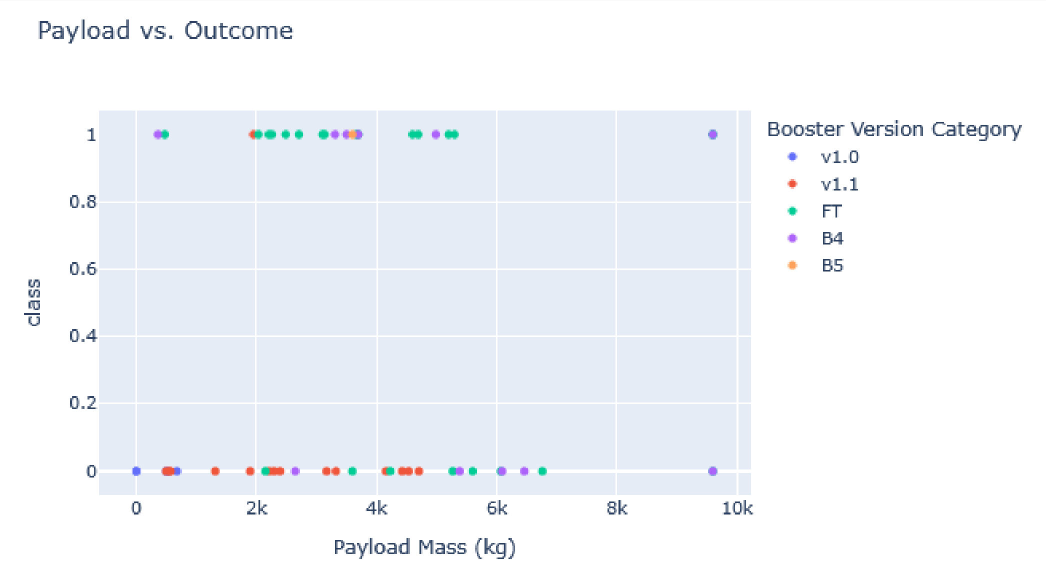
- Failures account for only ~23% of launches from this site.

- This shows a strong reliability compared to other launch sites.

Conclusion:

- KSC LC-39A is the most reliable and successful launch site overall.

Payload vs. Launch Outcome (Scatter Plot)



Observation:

- Each point shows a launch: Payload Mass (kg) on x-axis, Outcome (class) on y-axis.

- Successful launches (class=1) are spread across payload ranges, especially between 2000–6000 kg.

- Failures (class=0) occur more often at very low (<1000 kg) and very high (>8000 kg) payloads.

- Booster Version Category is shown by color: newer boosters (FT, B4, B5) appear more reliable.

Conclusion:

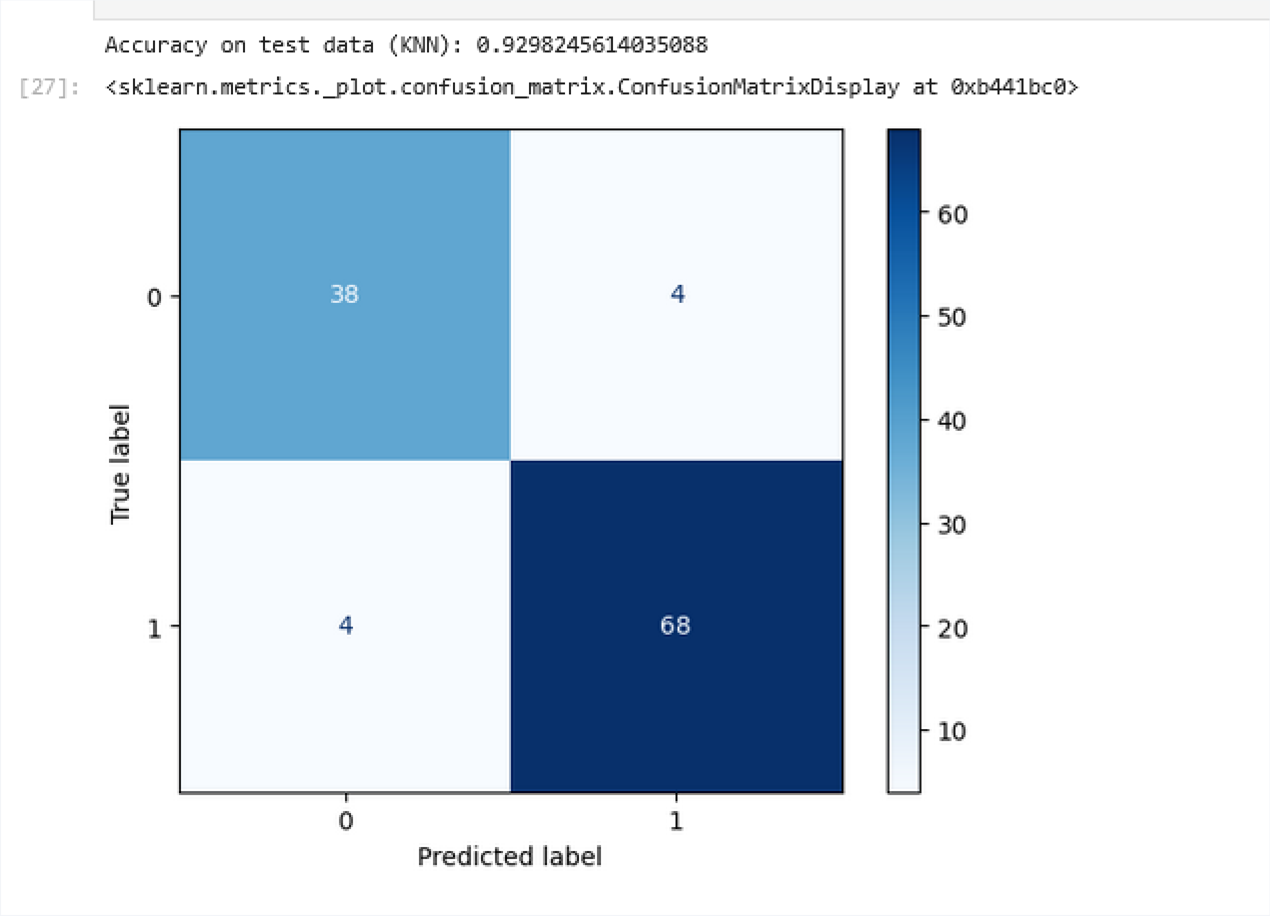
- Mid-range payloads (2000–6000 kg) had the highest success rates.

- Advanced boosters improved launch success at higher payload masses.

Section 5

Predictive Analysis (Classification)

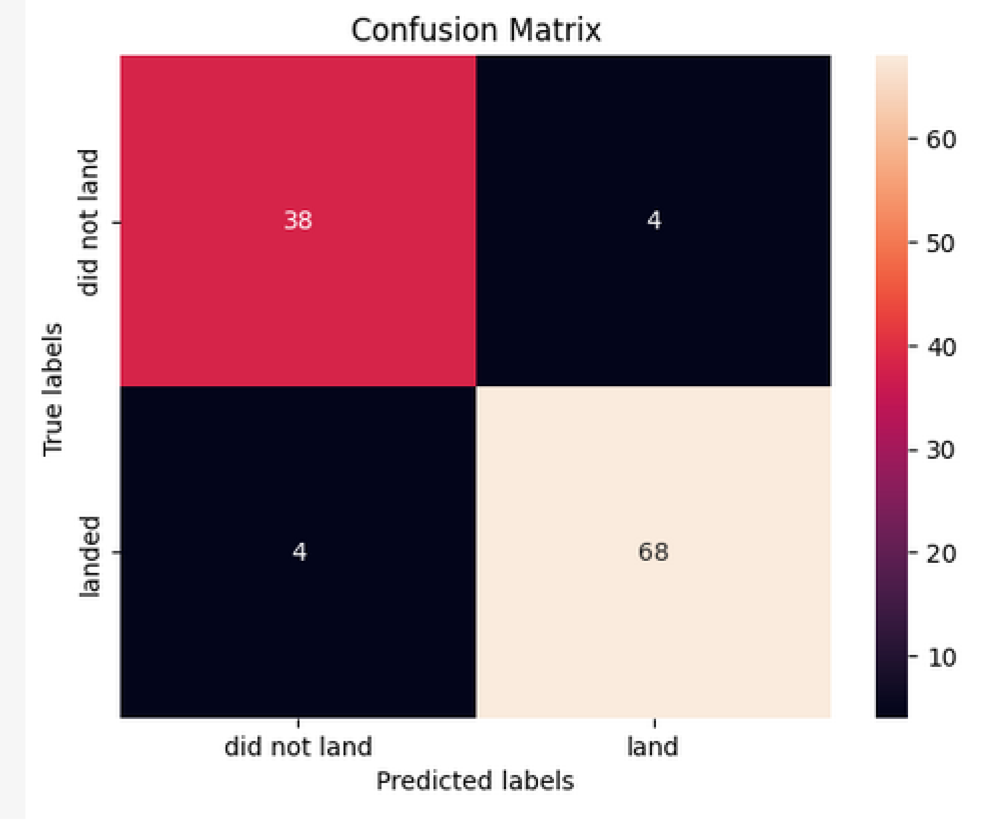
Classification Model Accuracy Comparison



Observation:
- Logistic Regression: ~83% accuracy
- Support Vector Machine (SVM): ~93% accuracy
- Decision Tree: ~73% accuracy
- K-Nearest Neighbors (KNN): ~80% accuracy

Conclusion:
- The Support Vector Machine (SVM) achieved the highest classification accuracy (~93%).
- This makes SVM the best-performing model for predicting SpaceX launch outcomes in this dataset.

Confusion Matrix (SVM – Best Model)



True Negatives (38): Correctly predicted "did not land“

True Positives (68): Correctly predicted "landed“

False Positives (4): Predicted "landed" but actually "did not land“

False Negatives (4): Predicted "did not land" but actually "landed"

Interpretation: Only 8 misclassifications → confirms ~93% accuracy.

Conclusions

- Tested multiple classification models to predict SpaceX Falcon 9 landings
- Support Vector Machine (SVM) achieved the best accuracy (~93%)
- Confusion Matrix confirmed strong performance with minimal errors
- Models like Decision Tree and KNN performed worse (~73–80%)
- Machine learning can effectively predict rocket landing outcomes

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!