IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

PRINCE MARWA
09/11/2025

Outline

Executive Summary

This project presents a full-cycle analysis of SpaceX Falcon 9 launches, aimed at predicting first-stage landing success, a critical factor in estimating launch cost and planning reusable missions. The workflow spans data acquisition via the SpaceX REST API and web scraping, followed by data wrangling, SQL-based exploratory analysis, geospatial mapping, dashboard development, and predictive modeling. Each phase contributes to a robust pipeline that transforms raw launch data into actionable insights for SpaceY.

While cost was not modeled directly, it is tightly linked to first-stage recovery: successful landings enable booster reuse, significantly reducing launch expenses. By predicting recovery likelihood based on payload mass, launch site, and booster configuration, this project provides a foundation for estimating cost scenarios and optimizing launch strategies. The Decision Tree model achieved 94.4% test accuracy and offers interpretable predictions, while geospatial analysis highlighted CCAFS LC-40 as a favorable site due to its proximity and historical performance. The interactive dashboard empowers stakeholders to explore launch outcomes dynamically, enhancing transparency and decision-making.

Executive Summary

Recommendations
Use the predictive model to estimate recovery likelihood and inform cost planning
Prioritize launch sites with strong historical success and favorable proximity metrics
Consider payload mass thresholds when planning reusable missions
Extend the dataset with weather, booster condition, and time-series data for future iterations

Impact
This project demonstrates how technical rigor and visual storytelling can be combined to support strategic aerospace decisions. It equips SpaceY with tools to reduce launch costs, improve mission planning, and scale reusable launch operations with confidence.

Introduction

Project background and context

SpaceX has revolutionized space travel by developing reusable rocket technology, significantly reducing launch costs. The Falcon 9 rocket's first stage is designed to return and land after launch which is a critical step toward full reusability.

While many Falcon 9 missions have successfully landed, others have failed due to factors like payload weight, orbit type, launch site, and booster configuration. Understanding these patterns can help improve future mission planning and risk assessment.

Problems we want to find answers to:

What factors most influence the success of a Falcon 9 first stage landing?
Can we predict landing success using historical launch data?
Which launch sites and mission types are most reliable?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:

Data Collection Methodology

Collected launch records using the SpaceX API

Scraped historical data from Wikipedia using BeautifulSoup

- Data processing/ wrangling

Data Wrangling

Merged datasets and cleaned missing values

Standardized formats and created binary labels for landing outcomes

Perform exploratory data analysis (EDA) using visualizations and SQL

- **Exploratory Data Analysis (EDA)**
- Used pandas and SQL to explore trends in payload mass, orbit types, and launch outcomes
- Used SQL and pandas for pattern discovery

Methodology

Interactive Visual Analytics

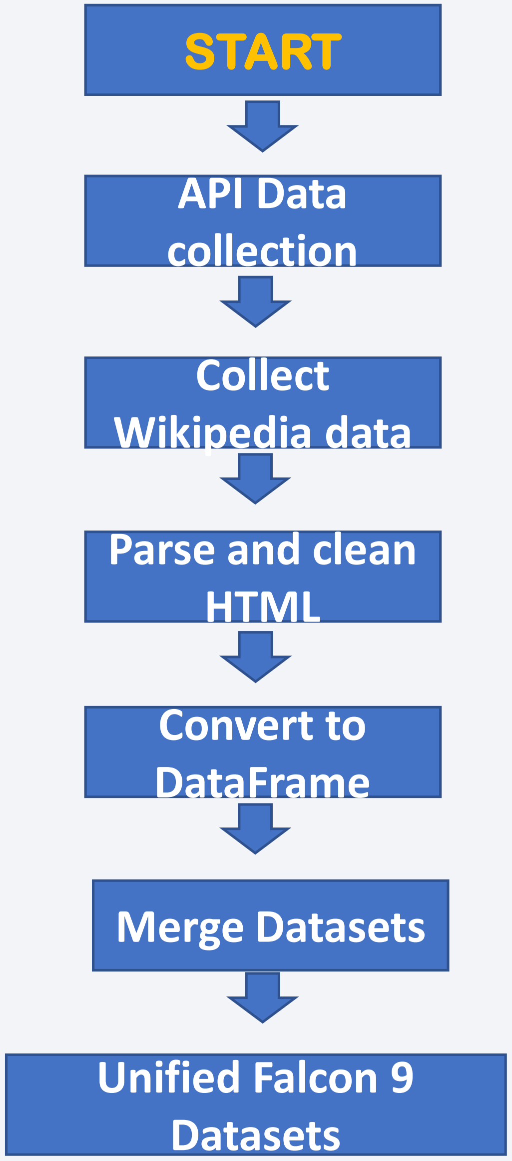
Mapped launch sites and landing zones using Folium
Developed a responsive dashboard using Plotly Dash
Enabled dynamic filtering by launch site and payload mass
Visualized launch outcomes with real-time updates and intuitive layout
Designed for stakeholder usability, supporting exploratory insight generation

Predictive Analysis

Framed the problem as a binary classification: landing success vs. failure
Trained multiple models: Logistic Regression, SVM, Decision Tree, KNN
Tuned hyper parameters using GridSearchCV
Evaluated models using accuracy score and confusion matrix

Selected Decision Tree as final model (94.4% test accuracy) for its interpretability and performance

Data Collection

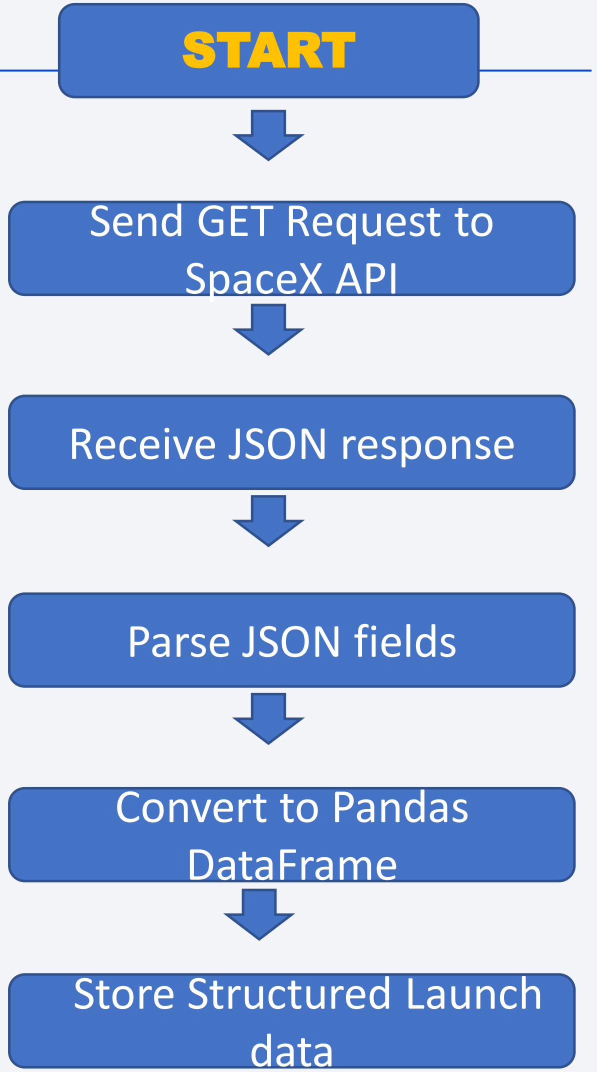


To build a reliable dataset for launch analysis, we combined two complementary data acquisition methods: API integration and web scraping. The SpaceX REST API provided structured launch metadata, while web scraping filled gaps in site details and booster outcomes. This dual-source strategy ensured both breadth and depth, laying a strong foundation for downstream wrangling, analysis, and modeling.

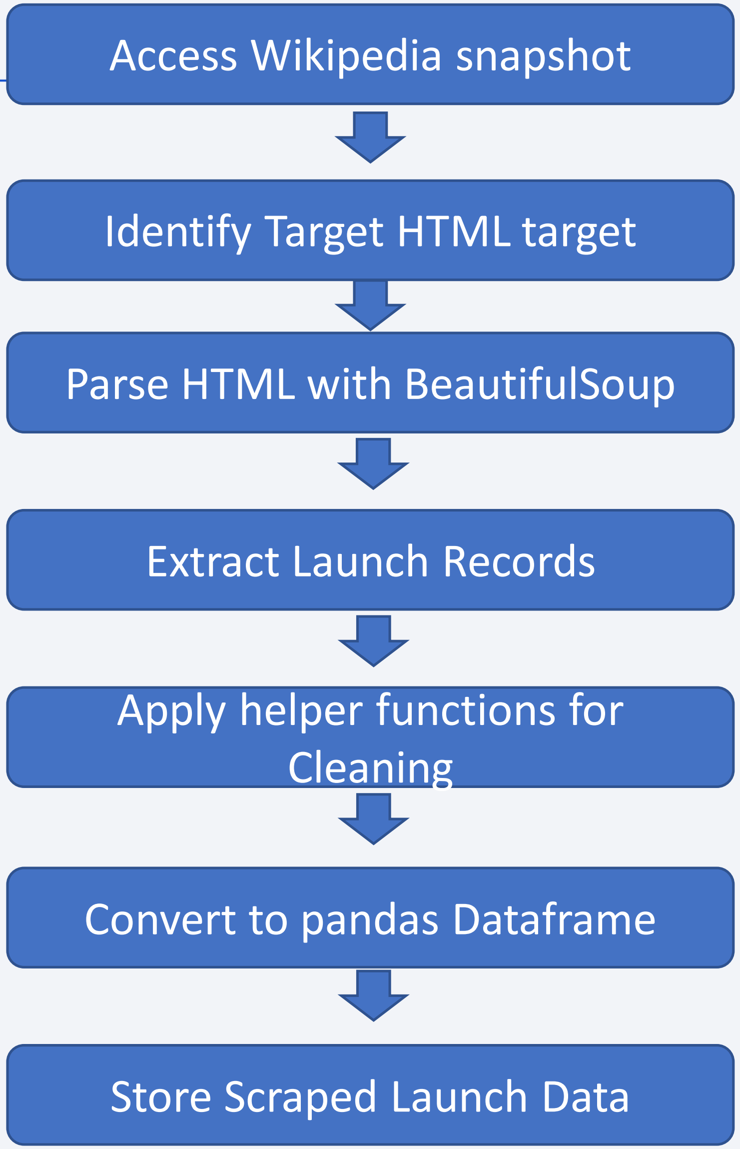
Data Collection – SpaceX API

Step 1: Send GET Request to SpaceX API
Initiated a RESTful API call using Python's requests.get() method to access launch data from SpaceX's public endpoint.
Step 2: Receive JSON Response
The API returned structured launch data in JSON format, containing nested fields like flight number, payload mass, orbit, and launch outcome.
Step 3: Parse JSON Fields
Used Python to extract relevant fields from the JSON object. Flattened nested structures and selected key variables for analysis.
Step 4: Convert to pandas DataFrame
Transformed the parsed JSON into a tabular format using pd.DataFrame(), enabling easy manipulation and filtering.
Step 5: Store Structured Launch Data
Saved the DataFrame for downstream tasks like wrangling, visualization, and modeling. This dataset became the foundation for the rest of the analysis.

JupyterNotebook task link: spacex-falcon9-landing-prediction/notebooks/jupyter-labs-spacex-data-collection-api.ipynb at main · princem601/spacex-falcon9-landing-prediction



Data Collection - Scraping



Step 1: Access Wikipedia Snapshot
Used a static version of the Wikipedia page to ensure consistency and avoid future edits affecting the dataset.

Step 2: Identify Target HTML Table
Located the third table on the page, which contained structured launch records including booster version, payload mass, and landing status.

Step 3: Parse HTML with BeautifulSoup
Used Python's BeautifulSoup library to parse the HTML content and isolate table rows and cells.

Step 4: Extract Launch Records
Iterated through table rows to extract relevant fields. Handled nested tags and inconsistent formatting.

Step 5: Apply Helper Functions
Created custom functions to clean and standardize extracted data, including unit conversion and missing value handling.

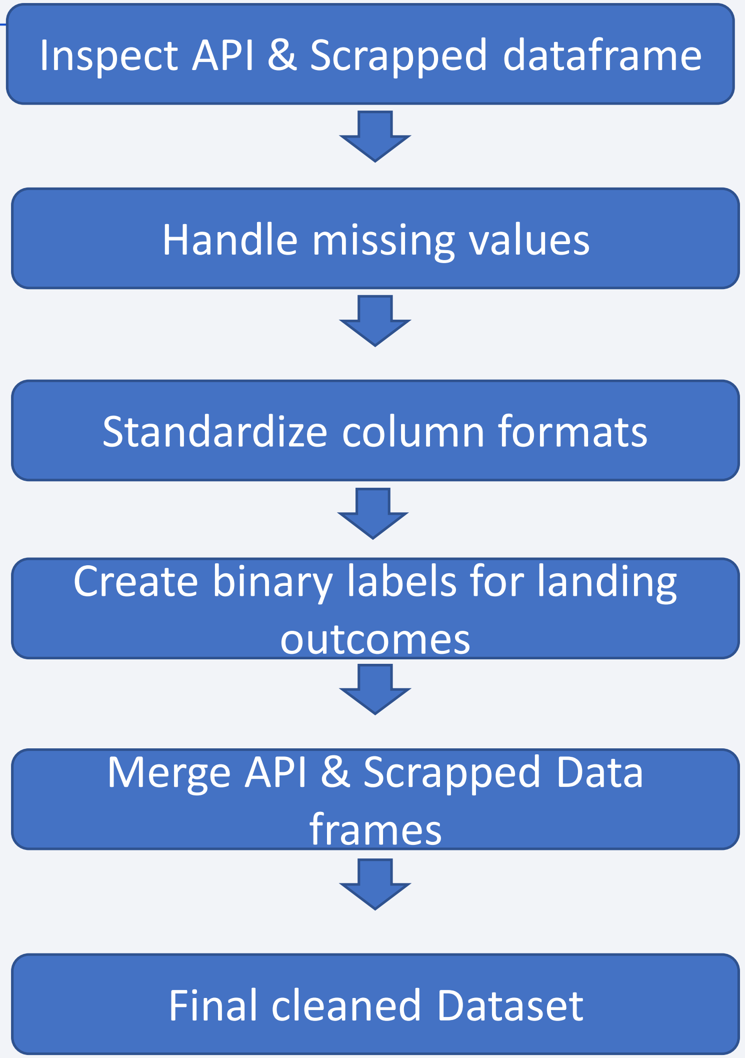
Step 6: Convert to pandas DataFrame
Structured the cleaned data into a DataFrame for merging with API data and further analysis.

JupyterNotebook task link: [spacex-falcon9-landing-prediction/notebooks/jupyter-labs-webscraping.ipynb at main · princem601/spacex-falcon9-landing-prediction]

Data Wrangling

Step 1: Inspect DataFrames
Reviewed both datasets for structure, column names, and completeness. Identified inconsistencies and missing fields.
Step 2: Handle Missing values
Used pandas to fill or drop missing entries. Applied logic to preserve key records while ensuring data integrity.
Step 3: Standardize Formats
Aligned column names, units (e.g., payload mass), and categorical values across both datasets.
Step 4: Create Binary Labels
Generated a new column Class to represent landing success:
1 for successful landing
0 for failure or no attempt
Step 5: Merge DatasetsJoined the cleaned API and scraped DataFrames on shared keys (e.g., flight number). Ensured no duplication or loss of records.
Step 6: Final OutputProduced a unified DataFrame with consistent formatting, complete records, and engineered features for modeling.

Jupyter Notebooks task link: [spacex-falcon9-landing-prediction/notebooks/labs-jupyter-spacex-Data wrangling.ipynb at main · princem601/spacex-falcon9-landing-prediction]()



EDA with Data Visualization

Objective
To explore key relationships in the Falcon 9 launch dataset using targeted visualizations.

Charts Plotted and Why

1. **Scatter Plot: Flight Number vs Payload Mass (with Class overlay)** *Why*: To examine how launch experience and payload size affect success. *Insight*: Higher flight numbers correlate with more successful landings, even for heavier payloads.
2. **Scatter Plot: Flight Number vs Launch Site** *Why*: To assess how different launch sites perform over time. *Insight*: Some sites show increasing success with more launches.
3. **Scatter Plot**: Payload Mass vs Launch Site Why: To compare payload capacity across launch sites. Insight: VAFB-SLC rarely handles heavy payloads (>10,000 kg).
4. **Bar Chart:** Success Rate by Orbit Type Why: To evaluate which orbit types have higher success rates. Insight: LEO and ISS orbits show stronger performance than GTO.
5. **Scatter Plot:** Flight Number vs Orbit Type Why: To explore how launch experience affects success across orbit categories. Insight: LEO shows improvement with flight number; GTO does not.
6. **Scatter Plot:** Payload Mass vs Orbit Type Why: To analyze how payload weight influences success across various orbit types. Insight: Polar, LEO, and ISS handle heavy payloads with higher success.
7. **Line Chart:** Yearly Launch Success Trend Why: To visualize how SpaceX's success rate evolved over time. Insight: Success rate steadily increased from 2013 to 2020.

Jupyter Notebooks link to tasks: [spacex-falcon9-landing-prediction/notebooks/edadataviz.ipynb at main · princem601/spacex-falcon9-landing-prediction]

EDA with SQL

We used SQL to explore deeper patterns in the cleaned dataset. We queried launch outcomes, orbit types, payload ranges, and booster reuse to validate trends and prepare modeling inputs. This step helped us confirm visual insights with structured logic and isolate high-impact features for prediction.

Summary of SQL Queries Performed

•**Query 1: SELECT * FROM dataset** *Purpose*: Inspect the full table and verify schema, column types, and completeness.
•**Query 2: COUNT success vs failure (Class column)** *Purpose*: Calculate overall launch success rate.
•**Query 3: GROUP BY LaunchSite → AVG(Class)** Purpose: Compare success rates across different launch sites.
•**Query 4: GROUP BY Orbit → AVG(Class)** Purpose: Identify which orbit types are most reliable.
•**Query 5: WHERE PayloadMass BETWEEN 4000 AND 6000** Purpose: Focus on mid-range payloads and analyze their outcomes.
•**Query 6: Filter by BoosterVersion and Outcome** Purpose: Explore how booster type affects landing success.
•**Query 7: GROUP BY Year → AVG(Class)** Purpose: Track SpaceX's performance improvement over time.
•**Query 8: WHERE Reused = 1 → Compare success** Purpose: Evaluate the impact of booster reuse on mission success.
•**Query 9: Filter by Orbit and PayloadMass range** Purpose: Isolate high-risk or high-performance mission profiles.
•**Query 10: Create temporary table for modeling inputs** Purpose: Prepare a clean subset of features for machine learning.

Jupyter notebook task link: spacex-falcon9-landing-prediction/notebooks/jupyter-labs-eda-sql-coursera_sqllite (1).ipynb at main · princem601/spacex-falcon9-landing-prediction

Build an Interactive Map with Folium

We enhanced our Folium map by visualizing SpaceX launch sites and their proximity to nearby coastlines. This geospatial layer added context to launch trajectories and helped simulate directional mapping.

Map Objects Added
Marker (folium.Marker) Purpose: To identify each launch site and coastline point. Details: Each marker included a popup with site name and a tooltip for clarity.
Circle (folium.Circle) Purpose: To highlight launch site areas and emphasize activity zones. Details: Circle radius was scaled to represent launch frequency or success rate.
PolyLine (folium.PolyLine) Purpose: To draw a line between the launch site and its nearest coastline point. Details: Simulated trajectory direction and spatial relationship.
Distance Label (folium.Marker with DivIcon) Purpose: To display the calculated distance between launch site and coastline. Details: Styled HTML label added at the coastline point for visual clarity.
Tile Layer (folium.TileLayer) Purpose: To improve map readability and aesthetics. Details: A clean base layer was used to keep focus on launch data.

GITHUB Jupyter notebook link to task: spacex-falcon9-landing-prediction/notebooks/lab_jupyter_launch_site_location (1).ipynb at main · princem601/spacex-falcon9-landing-prediction

Build a Dashboard with Plotly Dash

We built an interactive dashboard using Plotly Dash to explore SpaceX launch records. The dashboard allowed users to filter data by launch site and payload range, revealing patterns in success rates and payload performance.

Dashboard Components and RationaleLaunch Site Dropdown (dcc.Dropdown)

Purpose: Enabled users to select a specific launch site or view all sites. Why: Facilitated site-level comparisons and focused analysis of individual launch locations.

Success Pie Chart (dcc.Graph) Purpose: Displayed total successful launches per site or success vs failure for a selected site. Why: Provided a quick visual summary of mission reliability across different launch pads.

Payload Range Slider (dcc.RangeSlider) Purpose: Allowed users to filter launches by payload mass.
Why: Helped isolate payload bands and examine their impact on launch outcomes.

Payload vs Outcome Scatter Plot (dcc.Graph) Purpose: Showed the relationship between payload mass and launch success, colored by booster version. Why: Revealed trends in payload performance and booster reliability across missions.

Call back Functions (@app.callback) Purpose: Linked user inputs (dropdown and slider) to dynamic plot updates. Why: Enabled real-time interactivity and responsive data exploration.

Github url to task:[spacex-falcon9-landing-prediction/notebooks/spacex-dash-app.py at main · princem601/spacex-falcon9-landing-prediction]

Predictive Analysis (Classification)

We began with exploratory data analysis (EDA) to understand the SpaceX launch dataset. We identified key features like payload mass, launch site, and booster version that could influence landing success.

To prepare for classification, we created a new column called Class, which labeled each launch as either successful (1) or not (0). This became our target variable.
Next, we standardized the feature data using StandardScaler to ensure all inputs were on the same scale. This step was crucial for models like SVM and KNN that are sensitive to feature magnitude.
We split the data into training and test sets, using an 80/20 split with random state = 2. This gave us 18 test samples to evaluate final model performance.
We trained four classification models ; Logistic Regression, Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors (KNN). Each model was tuned using GridSearchCV with 10-fold cross-validation to find the best hyper parameters.
After training, we evaluated each model using validation accuracy, test accuracy, and confusion matrices. This helped us understand not just how accurate each model was, but also how they handled false positives and false negatives.
The Decision Tree model stood out with the highest test accuracy of **94.4%** and the lowest number of false positives. It correctly predicted 12 successful landings and only misclassified one non-landing.
We finalized the Decision Tree as our best-performing model, balancing accuracy, interpretability, and generalization.
GiTHUB url to lab task: spacex-falcon9-landing-prediction/notebooks/SpaceX_Machine Learning Prediction_Part_5.ipynb at main · princem601/spacex-falcon9-landing-prediction

Results

Exploratory Data Analysis Results

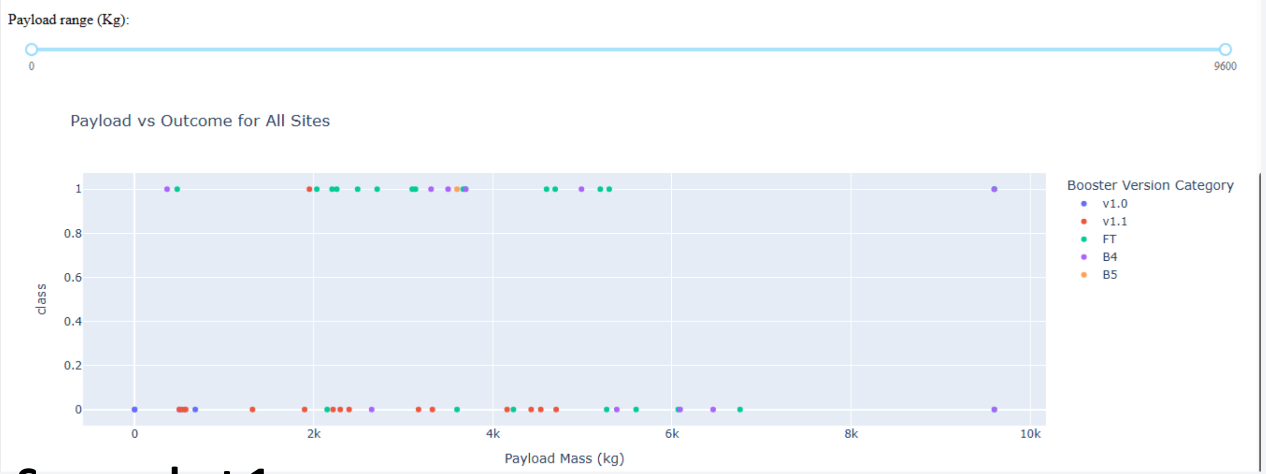
Key Insights from the Dataset

Launch Success Distribution: Majority of launches were successful (landed), but failures were present across multiple sites and payloads.
Launch Site Comparison: Certain sites (e.g., CCAFS LC-40, VAFB SLC-4E) had lower success rates compared to KSC LC-39A.
Payload Mass vs. Success : Heavier payloads showed a slightly lower success rate. This trend informed the payload slider range in the dashboard (final range: 0–5000 kg).
Booster Version Impact:
Newer booster versions (e.g., FT) had higher landing success rates.

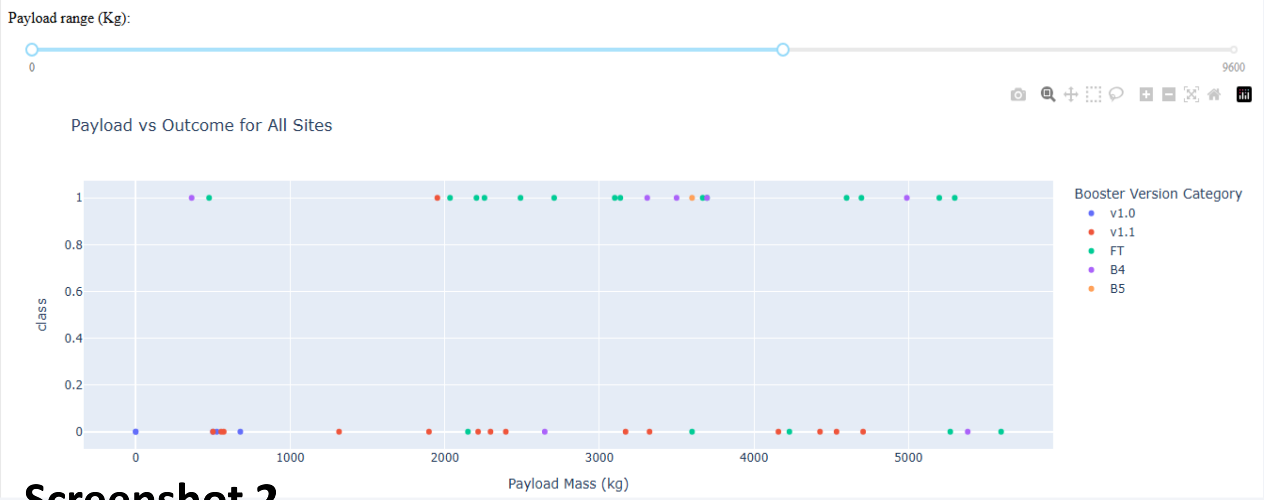
Results

Interactive analytics demo results

To validate dashboard interactivity, we captured phased screenshots showing how user inputs affect visualizations:



Screenshot 1

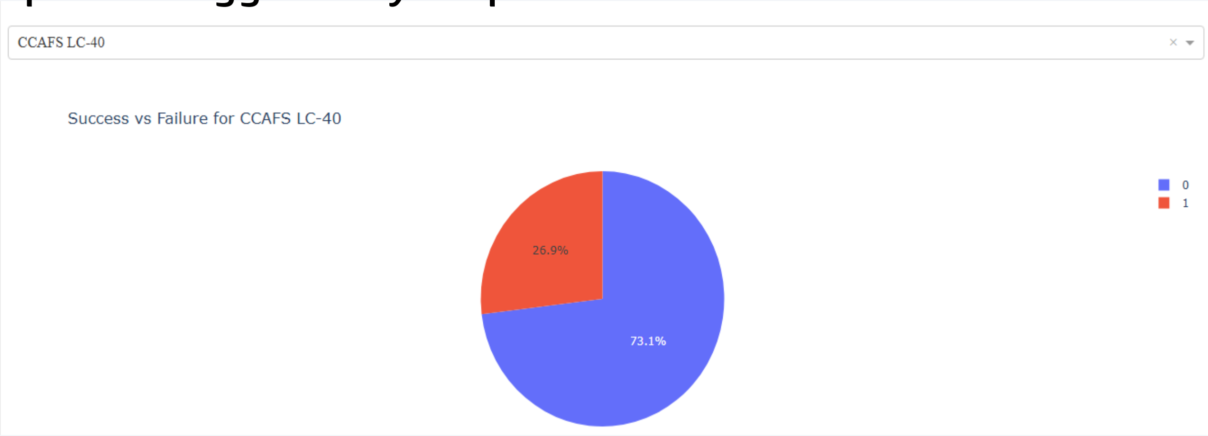


Screenshot 2

Payload Range Exploration:
Screenshot 1: Payload range extended to 9600 kg — scatter plot shows broader distribution.
Screenshot 2: Final range narrowed to 0–5000 kg — clearer separation of landing outcomes.

Launch Site Filter: Screenshot 3 shows how selecting a different site updates both the pie chart and scatter plot.

Call back Confirmation: Each screenshot reflects real-time updates triggered by dropdowns and sliders.



Screenshot 3

Results

Predictive analysis results

Classification Model Comparison

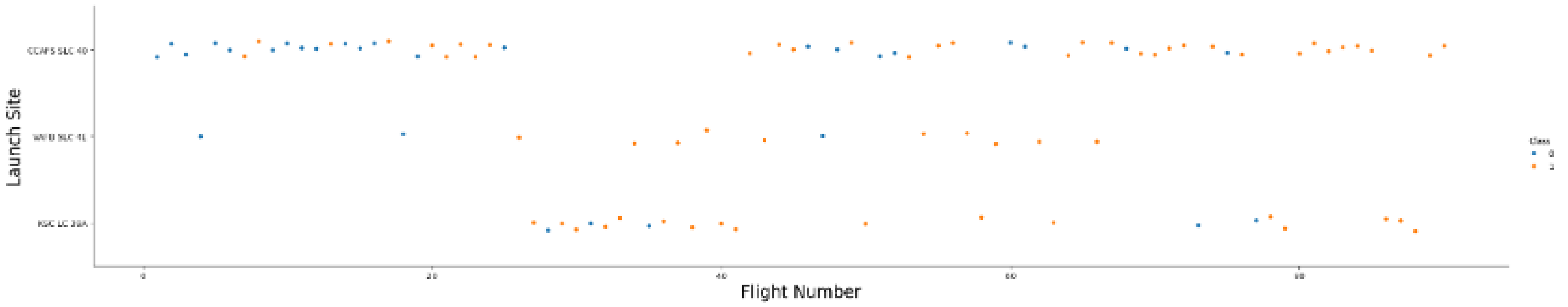
We trained and tuned four classifiers to predict SpaceX booster landing outcomes. The Decision Tree Classifier achieved the highest test accuracy (94.4%) and lowest false positive rate.

Model	Validation Accuracy	Test Accuracy	False Positives	Best Parameters
Logistic Regression	84.6%	83.3%	3	C=0.01, penalty=l2
SVM	84.8%	83.3%	3	C=1.0, gamma=0.0316, kernel=sigmoid
Decision Tree	**87.5%**	**94.4%**	**1**	max_depth=4, splitter=random
KNN	84.8%	83.3%	3	N_neighbors=10, p=1

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site



This scatter plot visualizes SpaceX launches over time across three sites: CCAFS LC-40, VAFB SLC-4E, and KSC LC-39A. Each point represents a flight, with the x-axis showing flight number and the y-axis showing launch site. Points are color-coded by landing outcome; orange for success and blue for failure.

The plot reveals several patterns:

Most launches occurred at CCAFS LC-40 and KSC LC-39A, while VAFB SLC-4E had fewer.

Success rates improved over time, especially at high-frequency sites.

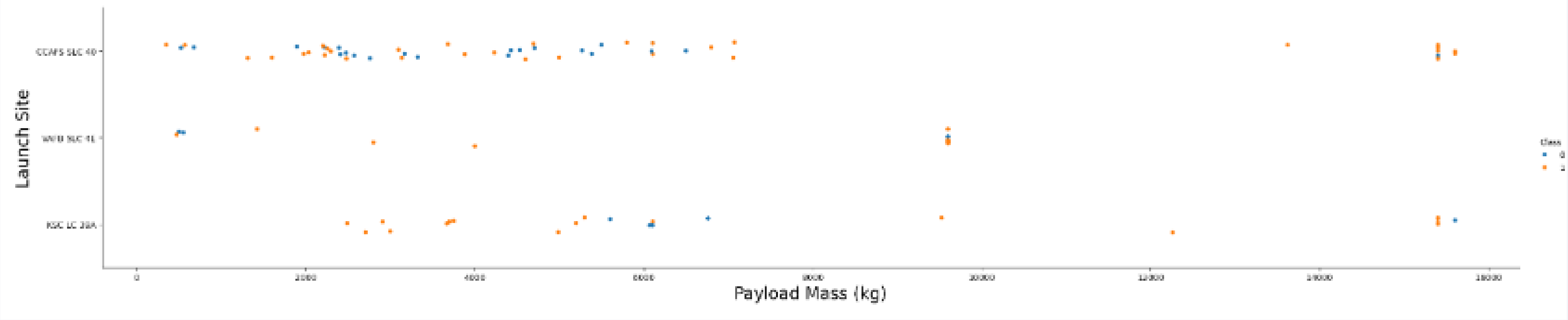
KSC LC-39A shows a higher concentration of successful landings.

VAFB SLC-4E had more failures, suggesting lower reliability.

Launches cluster in bursts, hinting at site-specific operational cycles.

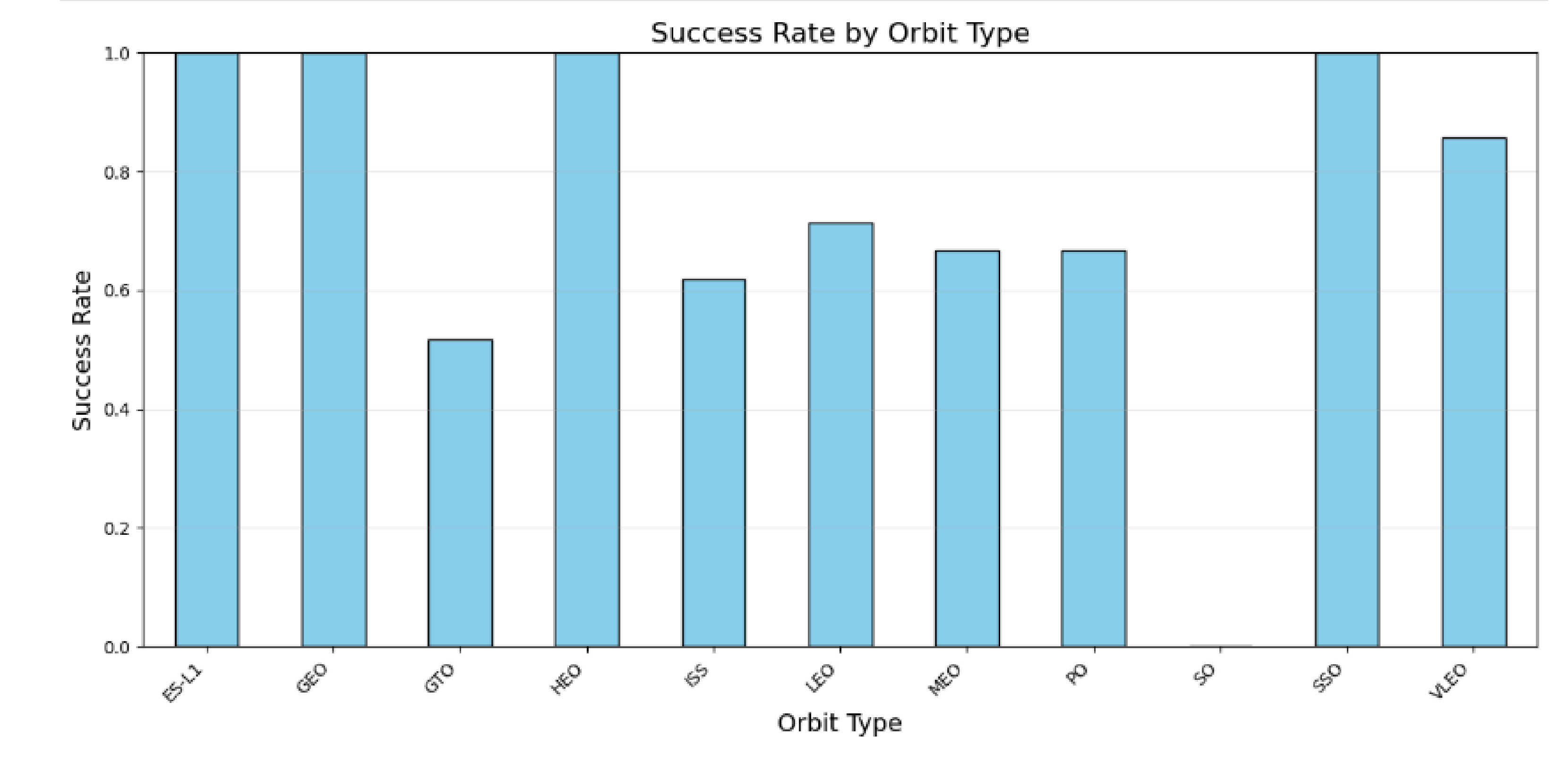
These insights helped guide feature selection and informed the classification modeling phase.

Payload vs. Launch Site



This scatter plot shows how payload mass varies across three launch sites: CCAFS LC-40, CCAFS SLC-40, and KSC LC-39A.
Each point represents a launch, with payload mass on the x-axis and launch site on the y-axis. Points are color-coded by landing outcome; orange for success and blue for failure.
Key insights:
KSC LC-39A handled the heaviest payloads, often with successful landings.
CCAFS LC-40 and SLC-40 launched lighter payloads with mixed outcomes.
Higher payloads were more likely to succeed at KSC LC-39A, suggesting site-specific capacity or mission type.
Launch site appears to influence both payload handling and landing success.

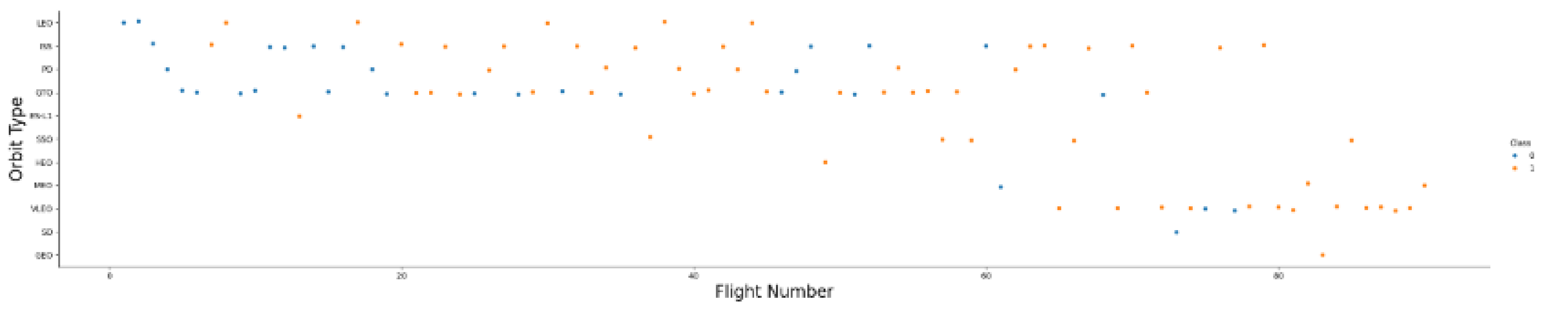
Success Rate vs. Orbit Type



This bar chart compares the landing success rates across different orbit types. Each bar represents the proportion of successful booster landings for missions targeting a specific orbit.

Key insights:
EL-L1, GEO, and VLEO achieved a perfect success rate (100%).
LEO and SS also performed well, with success rates above 85%.
GTO had the lowest success rate (~60%), indicating higher mission complexity or risk.
Orbit type influences landing outcomes, making it a valuable feature for prediction.

Flight Number vs. Orbit Type



This scatter plot shows how orbit types vary across SpaceX launches over time. Each point represents a flight, with flight number on the x-axis and orbit type on the y-axis. Points are color-coded by landing outcome: orange for success and blue for failure.

Key insights:

Orbit types like LEO, GTO, and SSO appear frequently across a wide range of flight numbers.

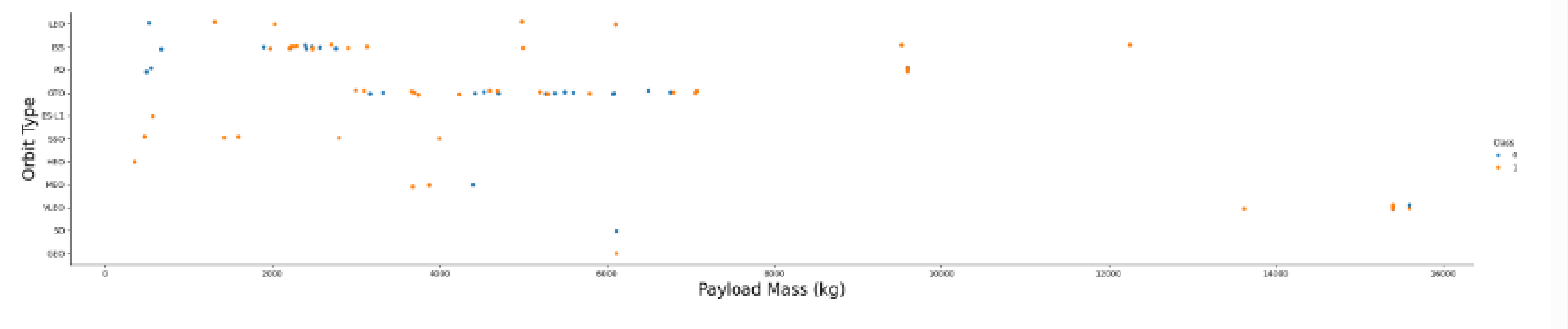
Success rates improved over time, especially for common orbits like LEO and SSO.

Less frequent orbits (e.g., GEO, EL-L1) are associated with later flights and higher success rates.

The distribution suggests that mission type (orbit) influences landing outcome and evolves with launch experience.

These trends helped identify orbit type as a meaningful feature for classification modeling.

Payload vs. Orbit Type



This scatter plot displays how payload mass varies across different orbit types. Each point represents a launch, with payload mass on the x-axis and orbit type on the y-axis. Points are color-coded by landing outcome: orange for success and blue for failure.

Key insights:

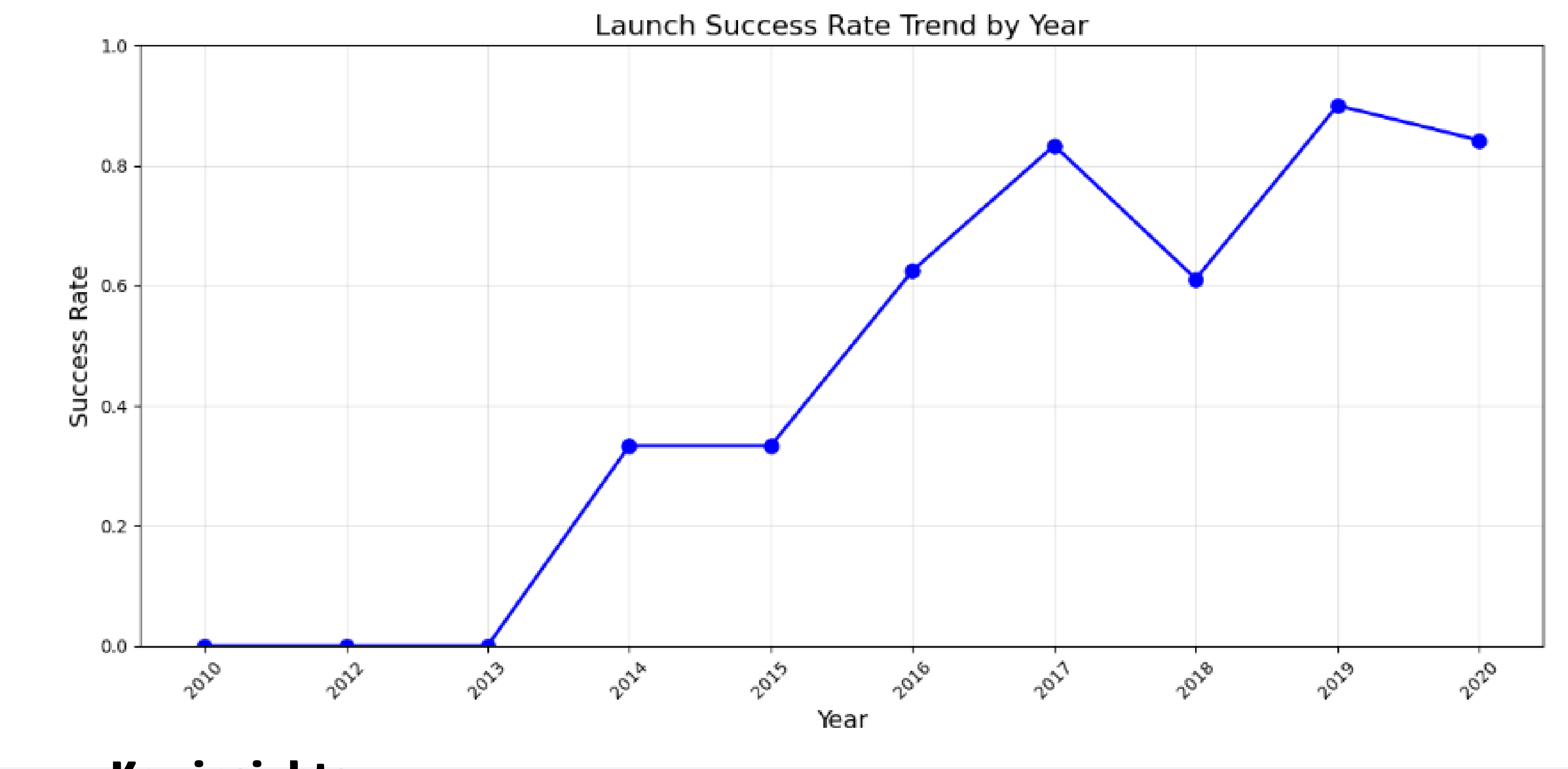
Heavier payloads are mostly associated with GTO and GEO missions.

Lighter payloads are common in LEO, SSO, and Polar orbits.

Successful landings are more frequent with lighter payloads and lower-energy orbits.

Orbit type and payload mass together influence landing outcomes, making both valuable features for prediction.

Launch Success Yearly Trend



This line chart tracks SpaceX's booster landing success rates from 2010 to 2020. The x-axis shows the year, and the y-axis shows the proportion of successful landings.

Key insights:

No successful landings occurred from 2010 to 2012.

Success rates began rising in 2013 and remained steady through 2015.

A sharp increase occurred from 2016 onward, peaking at 100% in 2018.

Slight dips in 2019 and 2020 suggest operational challenges or mission complexity.

Overall, the trend reflects major improvements in landing technology and mission reliability.

All Launch Site Names

We queried the dataset to extract all unique launch site names used in SpaceX missions. These sites represent the physical locations from which Falcon 9 rockets were launched.
Query used;

```
%sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE;
```

Query Result:

The dataset includes the following distinct launch sites:
CCAFS LC-40
40CCAFS SLC-40
40KSC LC-39A
39AVAFB SLC-4E

Explanation:
These launch sites vary in frequency and performance. CCAFS LC-40 and KSC LC-39A are the most active, while VAFB SLC-4E has fewer launches and lower success rates.
This information was used to analyze site-specific trends and select features for modeling.

Launch Site Names Begin with 'CCA'

We used a SQL query to filter launch site names that start with 'CCA', These sites played a central role in early and mid-phase launches, contributing significantly to the dataset’s volume and success trends.

Query used:

```
%sql SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE 'CCA%' LIMIT 5;
```

Query Result:

The launch sites beginning with “CCA” in the SpaceX dataset are:

CCAFS LC-40

CCAFS SLC-40

These sites are among the most frequently used by SpaceX for Falcon 9 launches.

Total Payload Mass

Using SQL, we filtered the dataset to include only launches where the customer was 'NASA (CRS)‘

query used:

```
%SELECT SUM(PAYLOAD_MASS__KG_) as Total_Payload_Mass FROM SPACEXTABLE WHERE Customer = 'NASA (CRS)';
```

Result:

The total payload mass carried by NASA boosters is **45,596 kg.**

Explanation:

This query aggregates the payload mass for all launches contracted by NASA under the Commercial Resupply Services (CRS) program. It helped quantify NASA’s launch volume and supported our customer-based analysis during the EDA phase.

Average Payload Mass by F9 v1.1

Using SQL, we calculated the average payload mass for launches using the booster version 'F9 v1.1'.
Query Used:

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) as Average_Payload_Mass FROM SPACEXTABLE WHERE
Booster_Version = 'F9 v1.1';
```

Result:
The average payload mass carried by F9 v1.1 boosters is **2,928.4 kg**.

Explanation:
This query filtered the dataset by booster version and computed the mean payload mass. The result reflects the typical cargo weight handled by F9 v1.1 during its operational period, helping us compare performance across booster generations.

First Successful Ground Landing Date

To Identify the date of the first successful booster landing on a ground pad we used the query

```
%sql SELECT MIN(Date) as First_Successful_Ground_Landing FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (ground pad)';
```

Result:
December 22, 2015

Explanation:
This query filters the dataset for successful landings on ground pads, sorts the results by date in ascending order, and returns the earliest one.

Successful Drone Ship Landing with Payload between 4000 and 6000

To Identify launches with successful drone ship landings where the payload mass was between 4000 and 6000 kg we used the query :

```
%sql SELECT Booster_Version FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000;
```

Result:
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Explanation:
This query filtered the dataset by landing outcome and payload range, then returned the unique booster versions that met both criteria. These boosters demonstrated reliable performance under mid-weight payload conditions and contributed to our feature selection for modeling.

Total Number of Successful and Failure Mission Outcomes

To Calculate the total number of successful and failed mission outcomes we used the query :

```
%sql SELECT Mission_Outcome, COUNT(*) as Total_Count FROM SPACEXTABLE GROUP BY Mission_Outcome;
```

Result:

Mission Outcome	Total Count
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Explanation:

This query grouped all missions by outcome and counted how many times each occurred. The results show that 99 missions were successful, with only one failure and one mission with unclear payload status. This confirms SpaceX's high reliability rate and supports our classification model's emphasis on success prediction.

Overall, the data shows a strong success rate with only one recorded in-flight failure.

Boosters Carried Maximum Payload

To identify booster versions that carried the heaviest payloads we used the query;

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

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:

All boosters listed belong to the **Falcon 9 Block 5 (F9 B5)** family, known for high thrust and reusability. Their repeated appearance reflects consistent performance at maximum payload capacity, making them key contributors to SpaceX's heavy-lift missions.

2015 Launch Records

To retrieve launch records from the year 2015, including landing outcome, booster version, and launch site we used the query;

```
%sql SELECT substr(Date, 6, 2) as Month, Landing_Outcome, Booster_Version, Launch_Site FROM SPACEXTABLE WHERE substr(Date, 0, 5) = '2015' AND Landing_Outcome = 'Failure (drone ship)';
```

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

Explanation:
This query filtered launches by year and returned key details for each. In 2015, both recorded missions attempted drone ship landings but failed. These early Falcon 9 v1.1 boosters were part of SpaceX's initial efforts to develop reusable launch technology.

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

To Count and rank all landing outcomes for launches between June 4, 2010 and March 20, 2017 we used the query;
%sql SELECT Landing_Outcome, COUNT(*) as Count FROM SPACEXTABLE WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome ORDER BY COUNT(*) DESC;

Result Table:

Landing 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

Explanation:
This query ranks all landing outcomes within the specified date range. The most frequent outcome was **"No attempt"**, reflecting early missions before reusable landing technology was implemented. The data also shows a balanced number of successes and failures on drone ships, highlighting SpaceX's iterative progress toward reliable booster recovery.

Section 3

Launch Sites Proximities Analysis

Launch site distribution- map overview

We created an interactive Folium map to visualize the geographic distribution of SpaceX launch sites. Using coordinates from the dataset, we plotted markers for each site and added popup labels for identification.

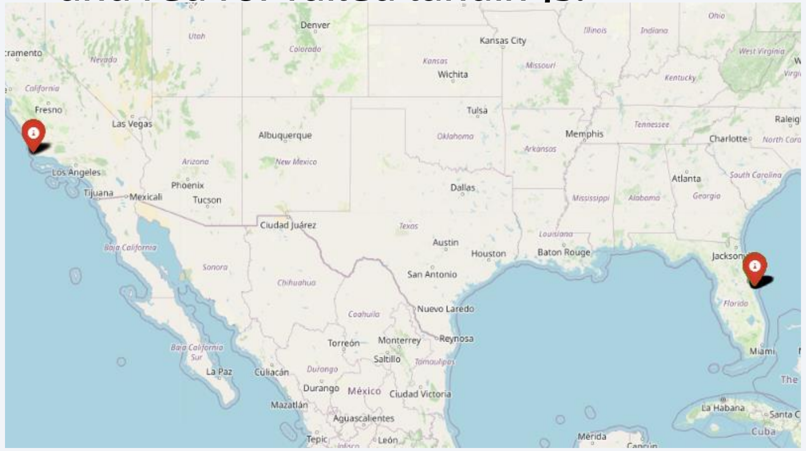
Launch Sites Displayed:
- **CCAFS LC-40**
- **KSC LC-39A**
- **VAFB SLC-4E**

Key Insights:

- All sites are located near coastlines, supporting safe launch trajectories and recovery operations
- The two Florida sites are geographically close and may appear clustered on the map
- Florida sites benefit from proximity to NASA infrastructure, enhancing operational efficiency
- VAFB's west coast location enables access to polar orbits, expanding mission capabilities

Launch outcomes by location

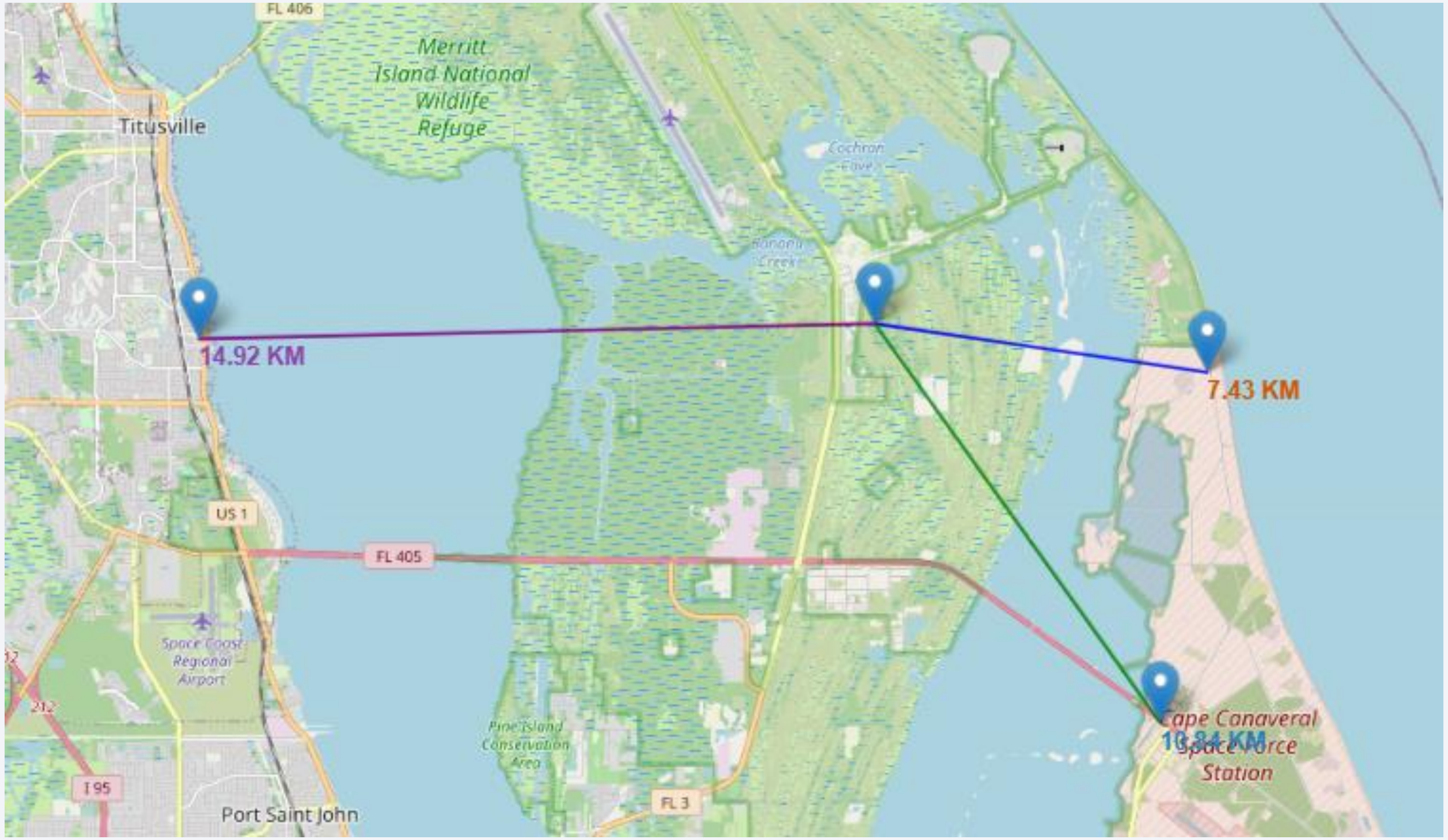
We created a Folium map to visualize SpaceX launch outcomes by location, using launch site coordinates from the dataset. Each marker is color-coded to represent the result of the launch it corresponds to: green for successful landings and red for failed landings.



Marker Distribution:
- **VAFB SLC-4E** (left side of map): only red marker (no successful landings)
- **CCAFS LC-40**: both red and green markers (mixed outcomes)
- **KSC LC-39A**: only green marker (successful landings)

Key Insights:
- Florida sites show a progression from early failures to successful landings
- VAFB launches had no successful landings during the analyzed period
- The visual clustering of Florida sites reflects operational density and outcome diversity

Launch site proximity to infrastructure & coastline



We selected the **CCAFS LC-40** launch site to analyze its proximity to nearby geographic features using the Haversine formula. The map displays the launch site alongside its closest coastline point, NASA Parkway West (major road), and a nearby residential area. Each location is connected to the launch site with a color-coded line, and distance labels are shown directly on the map using styled markers.

Key Findings:

•**Coastline:** 7.43 km — supports safe launch and recovery operations

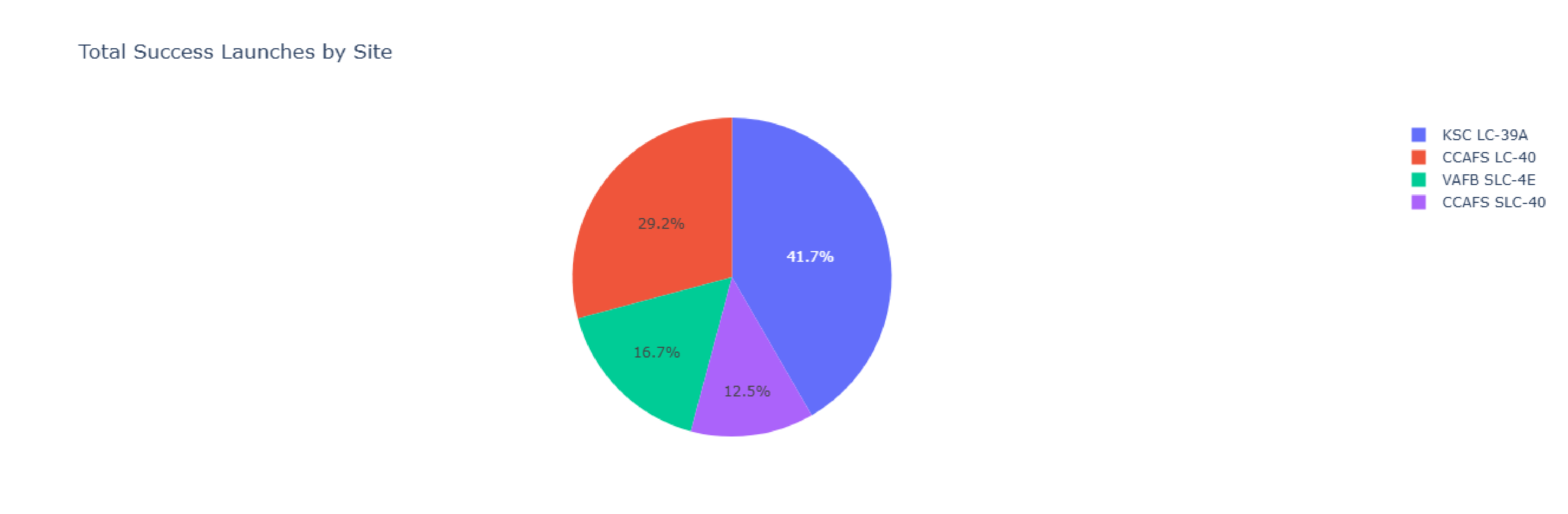
•**NASA Parkway West:** 10.10 km — enables efficient transport and logistics

•**Residential area:** 14.93 km — provides a safe buffer from populated zones

Section 4

Build a Dashboard with Plotly Dash

Launch success distribution across all sights



This pie chart visualizes the proportion of successful SpaceX launches across four major launch sites. Each segment represents the percentage of total successful missions attributed to a specific location.

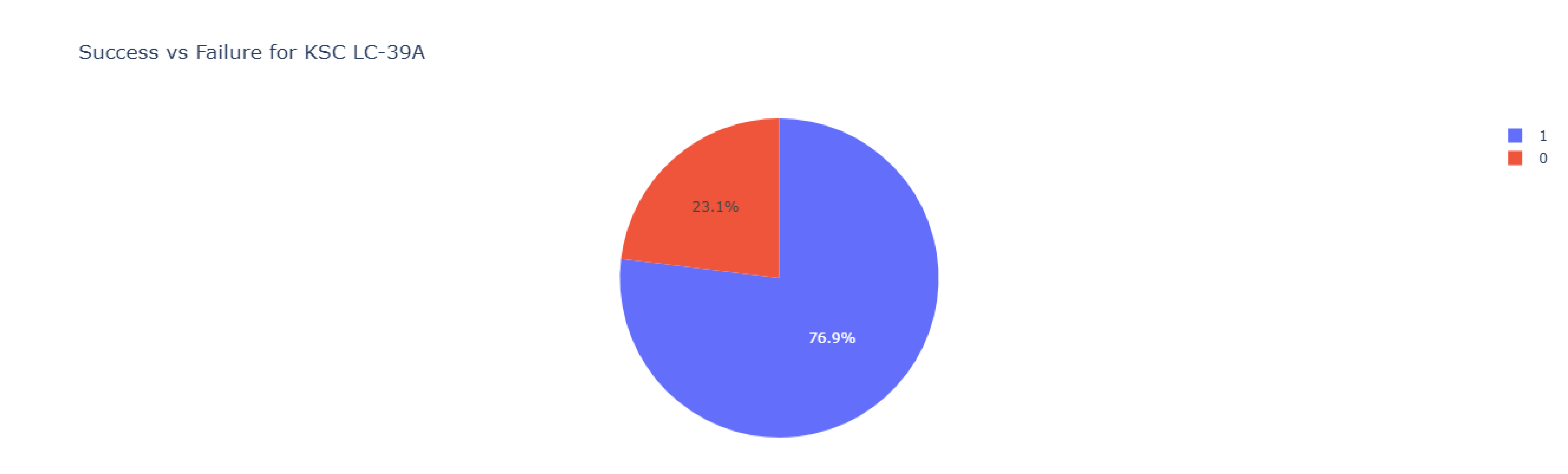
Key Findings

KSC LC-39A accounts for the largest share of successful launches (41.7%), indicating its central role in SpaceX operations.

CCAFS LC-40 follows with 29.2%, showing strong performance and frequent usage.

VAFB SLC-4E and CCAFS SLC-40 contribute smaller shares (16.7% and 12.5%, respectively), reflecting more specialized or less frequent launch activity. The chart highlights how launch success is distributed geographically, offering insight into site utilization and reliability.

Highest launch success ratio- KSC LC-39A

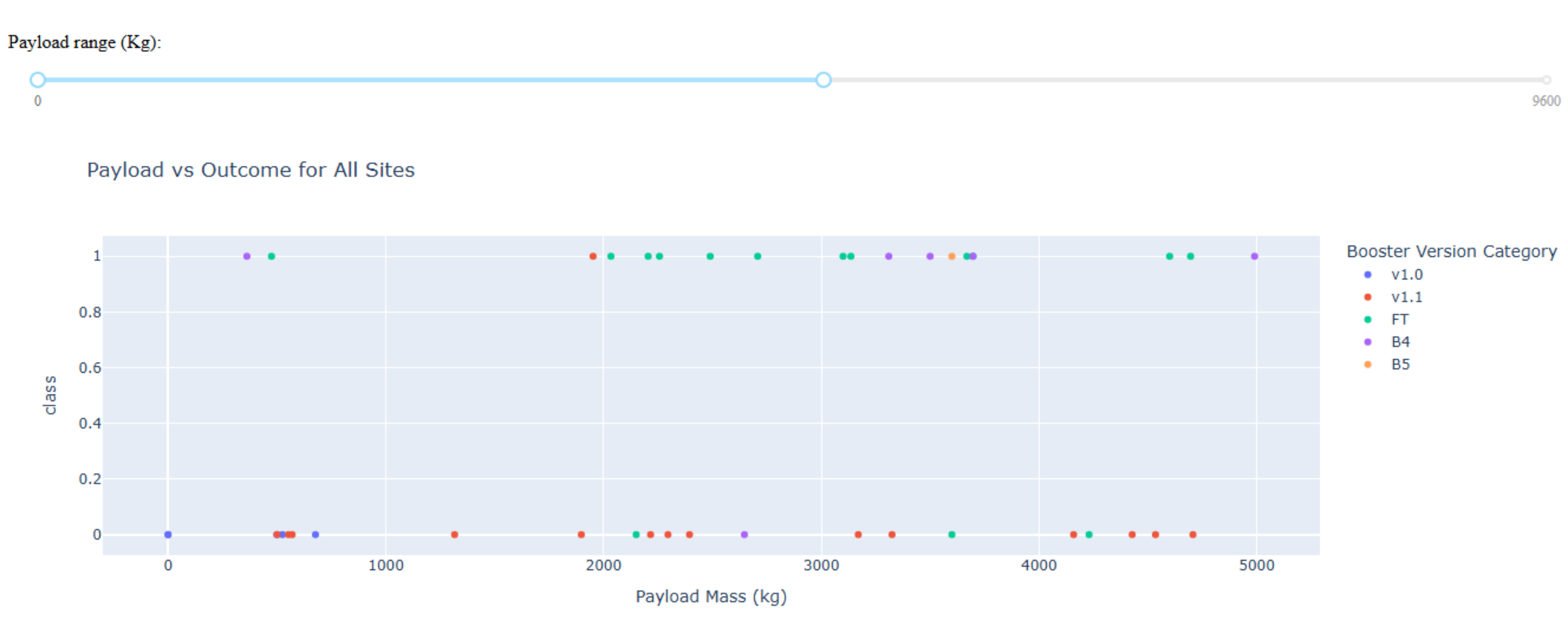


This pie chart presents the launch outcomes for KSC LC-39A, SpaceX's most successful launch site. The chart is divided into two segments: successful launches (blue) and failed launches (red), with percentage labels indicating the distribution.

Key Findings

76.9% of launches from KSC LC-39A were successful, confirming its status as the most reliable site in the dataset. Only 23.1% of launches failed, which is lower than the failure rates observed at other sites. The chart provides a clear visual of the site's operational performance, reinforcing its strategic importance.

Payload vs Launch Outcome — Filtered Range (0–5000 kg)



This scatter plot displays the relationship between payload mass and launch success, filtered to include only launches with payloads between 0 and 5000 kg. Each point represents a launch, with color indicating the booster version and vertical position showing success (1) or failure (0).

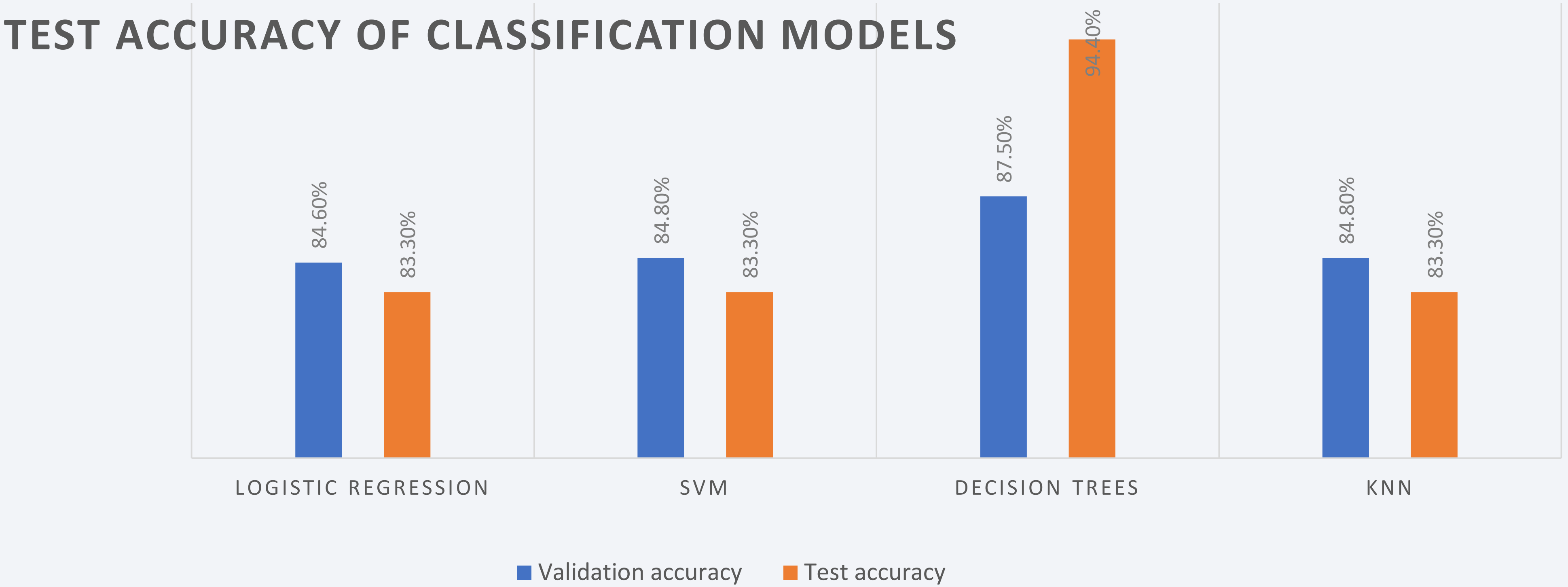
Key Findings

Most launches within this range were successful, especially those using newer booster versions. Booster Version Category continues to show strong correlation with success, suggesting technological improvements over time. The filtered view reveals that lighter payloads tend to have higher success rates, though some failures still occur across the range.

Section 5

Predictive Analysis (Classification)

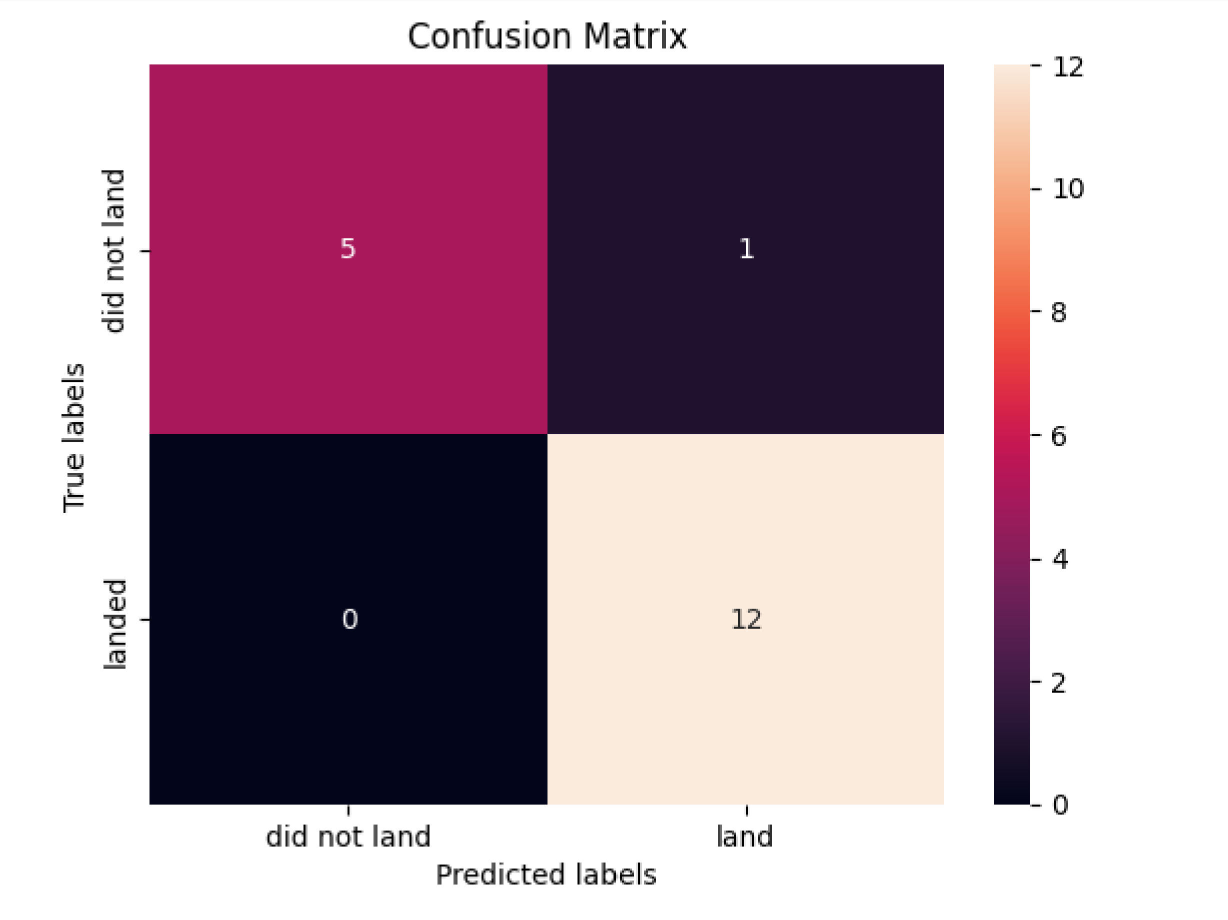
Classification Accuracy



Highest Accuracy

Decision Tree Classifier Achieved the highest test accuracy: 94.4% outperformed all other models in both validation and test phases.

Confusion Matrix



This matrix shows the performance of the Decision Tree Classifier on the test set:

True Positives (**12**): Boosters correctly predicted to have landed
True Negatives (**5**): Boosters correctly predicted not to have landed
False Positives (**1**): One booster incorrectly predicted to have landed
False Negatives (**0**): No missed landings

The model demonstrates high precision and perfect recall for successful landings, confirming its reliability and making it the best-performing classifier.

Conclusions

Key Takeaways from the Classification Analysis

EDA revealed key patterns: Launch site, payload mass, and booster version significantly influenced landing success, guiding feature selection.
Interactive dashboard validated insights: Dropdowns and sliders allowed users to explore launch outcomes dynamically, with phased screenshots documenting interactivity.
Four classification models were built: and tuned Logistic Regression, SVM, Decision Tree, and KNN were optimized using GridSearchCV and evaluated on a test set.
Decision Tree outperformed all others: Achieved the highest test accuracy (94.4%) and lowest false positive rate, making it the most reliable predictor.
Confusion matrix confirmed model reliability: The Decision Tree correctly identified all successful landings and misclassified only one non-landing, showing strong precision and recall.

Appendix

Appendix A: Data Collection

This section documents the full data acquisition process for the SpaceX launch analysis project. It includes both API integration and web scraping to build a comprehensive dataset.

Tasks Completed:
Connected to the SpaceX REST API to retrieve structured launch data
Queried endpoints for booster, payload, core, and launch site metadata
Supplemented with web scraping to enrich missing or incomplete fields
Combined and cleaned the data for downstream wrangling and analysis

Relevant Notebooks:

API Data Collection Notebook: spacex-falcon9-landing-prediction/notebooks/jupyter-labs-spacex-data-collection-api.ipynb at main · princem601/spacex-falcon9-landing-prediction

Web Scraping Data Collection Notebook: spacex-falcon9-landing-prediction/notebooks/jupyter-labs-webscraping.ipynb at main · princem601/spacex-falcon9-landing-prediction

Appendix

Appendix B: Data Wrangling

This section covers the preprocessing steps applied to the raw SpaceX launch data to prepare it for analysis and modeling.

Tasks Completed:
Filtered the dataset to include only Falcon 9 launches
Created binary classification labels to indicate first-stage landing success
Handled missing values across payload, booster, and landing outcome fields
Standardized feature formats for modeling and dashboard integration

Relevant Notebooks:

Data Wrangling Notebook:spacex-falcon9-landing-prediction/notebooks/labs-jupyter-spacex-Data wrangling.ipynb at main · princem601/spacex-falcon9-landing-prediction

Appendix

Appendix C: Data Wrangling

This section focuses on extracting structured insights from the cleaned SpaceX launch dataset using SQL queries and visualizations.

Tasks Completed:
Connected the dataset to a SQL engine for structured querying
Filtered launches by site, payload mass, and success status
Aggregated metrics to assess launch performance across sites
Visualized payload distributions, success rates, and site comparisons using Seaborn and Matplotlib

Relevant Notebooks:

SQL-Based EDA Notebook:spacex-falcon9-landing-prediction/notebooks/jupyter-labs-eda-sql-coursera_sqllite (1).ipynb at main · princem601/spacex-falcon9-landing-prediction

EDA Visualization Notebook:spacex-falcon9-landing-prediction/notebooks/edadataviz.ipynb at main · princem601/spacex-falcon9-landing-prediction

Appendix

Appendix D: Interactive visual analytics with folium (geospatial analysis)

This section documents the mapping and proximity calculations performed to assess the spatial context of SpaceX launch sites.

Tasks Completed:

Mapped major launch sites (KSC LC-39A, CCAFS LC-40, VAFB SLC-4E) using FoliumVerified site coordinates manually via Google Maps
Added interactive markers and popups for each site
Conducted proximity analysis for CCAFS LC-40 using the Haversine formula
Calculated distances to coastline, NASA Parkway West, and residential zones
Annotated maps with color-coded lines and distance labels

Relevant Notebooks: spacex-falcon9-landing-prediction/notebooks/lab_jupyter_launch_site_location (1).ipynb at main · princem601/spacex-falcon9-landing-prediction

Appendix

Appendix E: Interactive dashboard with Plotly Dash

This section highlights the development of an interactive dashboard to explore SpaceX launch outcomes and payload dynamics.

Tasks Completed:
Built a Plotly Dash application for interactive data exploration
Enabled dynamic filtering by launch site and payload mass range
Visualized launch success rates and payload distributions in real time
Integrated call back functions to update graphs based on user input
Designed layout for clarity, responsiveness, and stakeholder usability

Relevant Notebook: spacex-falcon9-landing-prediction/notebooks/spacex-dash-app.py at main · princem601/spacex-falcon9-landing-prediction

Appendix

Appendix F : Machine Learning (Predictive Modeling)

This section details the machine learning workflow used to predict whether the Falcon 9 first stage will successfully land — a key factor in estimating launch cost.

Tasks Completed:

Selected relevant features including payload mass, booster version, and launch site
Split the dataset into training and test sets
Trained four classification models:
Logistic Regression, SVM, Decision Tree, and KNN
Applied GridSearchCV for hyper parameter tuning,
code snippet;

```
grid = GridSearchCV(DecisionTreeClassifier(), param_grid, cv=5)
grid.fit(X_train, y_train)
```

Evaluated models using accuracy scores and confusion matrices
Identified Decision Tree as the best-performing model with 94.4% test accuracy

Relevant Notebook: [spacex-falcon9-landing-prediction/notebooks/SpaceX_Machine Learning Prediction_Part_5.ipynb at main · princem601/spacex-falcon9-landing-prediction]

Thank you!