IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Name: SOHEIL SANATI MOUTABAN
Date: 07 AUGUST 2025

Outline

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

Executive Summary

Summary of Methodologies

1. Data Collection:

- Accessing SpaceX *API* and *Web Scraping* on SpaceX Wikipedia records

2. Data Cleaning & Preparation:

- Executing data cleansing and formatting to ensure accuracy and consistency.
- Loaded and managed datasets in *IBM Db2*, writing and optimizing *SQL queries* for efficient retrieval and analysis.
- Performed comprehensive *Exploratory Data Analysis* (*EDA*) to surface insights and guide downstream modeling.

3. Feature Engineering:

- Engineered additional predictive variables to enrich the dataset.
- Standardized and scaled features to ensure consistent, model-ready inputs.

4. Interactive Visualizations:

- Developed interactive geospatial maps in *Folium* to display launch locations and corresponding success rates.
- Designed and deployed a dynamic analytics dashboard with *Plotly Dash* for real-time exploration of mission performance data.

5. Model Building & Evaluation:

- Constructed and benchmarked classification models using *SVM*, *Decision Trees*, and *k-NN*.
- Optimized hyperparameters through exhaustive *GridSearchCV* searches to maximize performance.
- Validated model effectiveness on a held-out test set, reporting accuracy and other key evaluation metrics.

Summary of Results

1. Data Insights:

- Pinpointed key drivers of Falcon 9 landing success and mapped geographic patterns and success rates.

2. Model Performance:

- *Decision Tree* led with 94.4 % accuracy; *SVM* and *k-NN* followed at 83.3 %.

3. Key Findings:

- Launch site and payload mass drive landing success, with the Decision Tree model emerging as the strongest predictor.

Introduction

Project Background and Context:

This capstone project focuses on forecasting the likelihood that SpaceX's Falcon 9 first stage will land successfully. Because SpaceX can reuse this stage, it markets launches at prices far below those of other providers.

By accurately predicting landing outcomes, we can better estimate launch costs and supply rival companies with strategic insights.

Problems Addressed:

- Which variables have the greatest impact on the Falcon 9 first stage chance of landing safely?
- How can machine-learning techniques be applied to generate precise landing predictions?
- Which machine-learning model delivers the most accurate forecasts of landing success?

Section 1

Methodology

Methodology

Executive Summary: Built an end-to-end pipeline that ingests SpaceX launch data, cleans and explores it, visualizes insights interactively, and trains machine-learning models to predict Falcon 9 first-stage landing success.

Data Collection: Pulled comprehensive launch records—dates, sites, payloads, outcomes—directly from the SpaceX API.

Data Wrangling: Fixed missing values, unified formats, and engineered extra features to enrich the dataset.

Exploratory Data Analysis (EDA): Charted success rates, payloads, and launch sites with *Matplotlib* / *Seaborn*, and ran targeted *SQL queries* for deeper insights.

Predictive Modeling: Trained *Logistic Regression*, *SVM*, *k-NN*, and *Decision Trees*; tuned with *GridSearchCV* and picked the top model by test accuracy.

Interactive Visuals: Built *Folium* maps to display launch sites and outcomes, and a *Plotly Dash* app with dropdowns and sliders to explore success rates across payload ranges.

Data Collection

First Step - API

SpaceX API Request → Fetching Launch Data → Save Data locally → Output → Raw JSON / CSV

Second Step - Scrape

Wikipedia HTML table → Parse with BeautifulSoup → DataFrame → Output → Clean table

Third Step - Integrate

Data → Merge Datasets → Data Integration → Output → Unified Dataset

Data Collection – SpaceX API

Step 1: Initiate API Request

- Initiating ***requests.get()*** function toward SpaceX API at ***/v4/launches***

  Result: Raw JSON

Step 2: Parse API Response

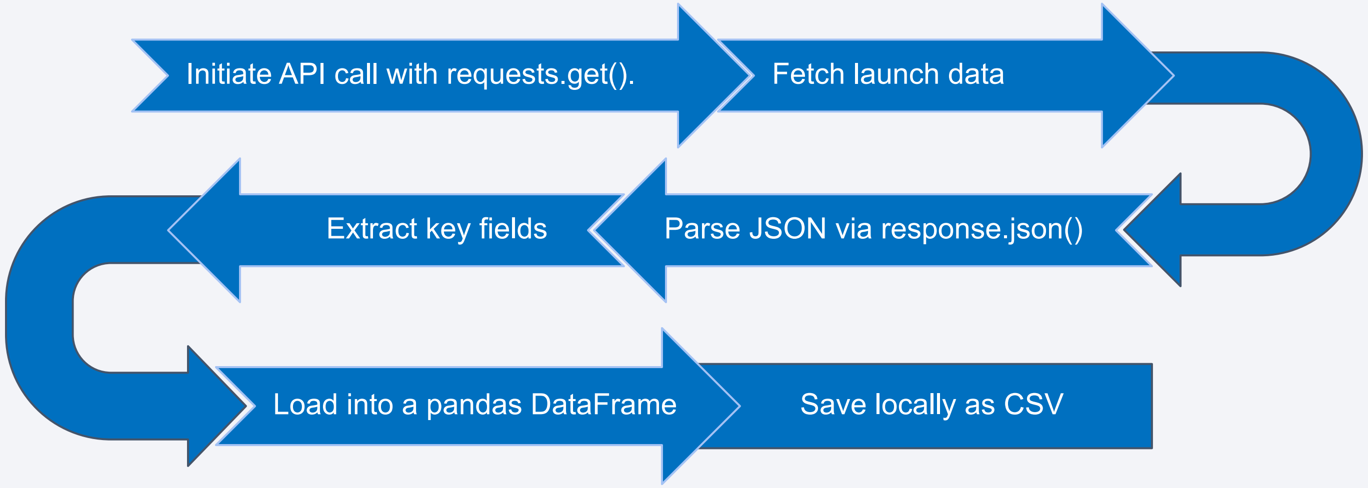
- ***response.json()*** → key fields (date, site, payload, outcome)

  Result: Clean dictionary *[launch date, launch site, payload mass, rocket type, outcome.]*

Step 3: Store Data Locally

- ***pandas.DataFrame → CSV / DB***

  Result: Dataframe ready for analysis



Github URL: [1. jupyter-labs-spacex-data-collection-api.ipynb]

Data Collection - Scraping

Step 1: Fetch HTML Content

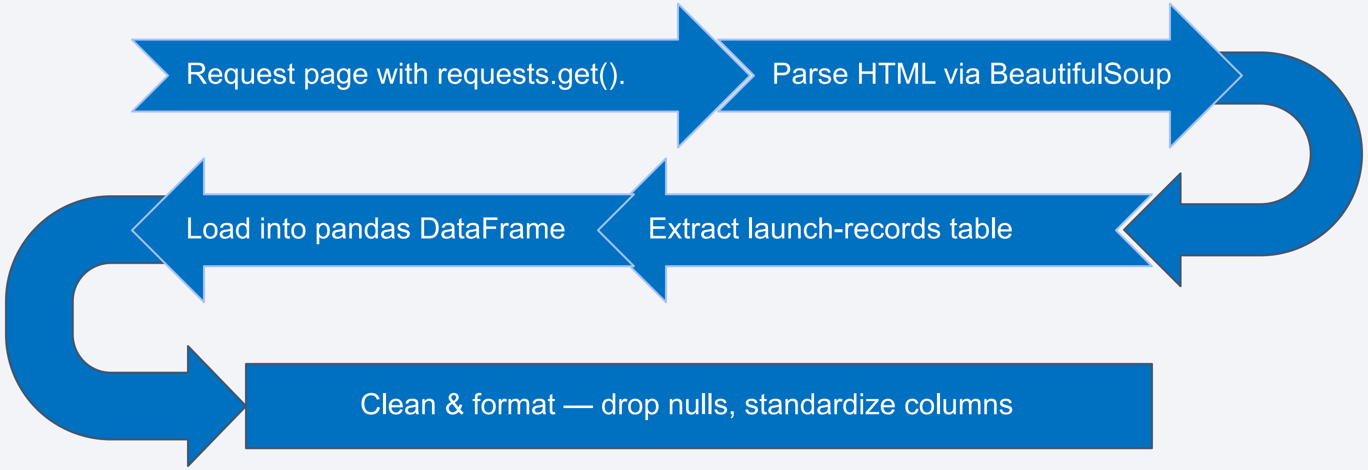
- *requests.get()* the Wikipedia page on Falcon 9 / Heavy launches

Step 2: Parse HTML & Extract content

- Use BeautifulSoup to locate the launch-records table and pull it out

Step 3: Convert to DataFrame & Clean

- Load the table into pandas, tidy up columns, and enforce consistent formats.



Github URL: [2. jupyter-labs-webscraping.ipynb]

Data Wrangling

Overview:

Data wrangling entails cleansing, converting, and arranging raw information into a structured format that is ready for analysis.

- **Step 1: Data Cleaning**
  - Detect and either replace or eliminate missing entries in the dataset.
  - Apply suitable imputation methods, or discard rows/columns that contain excessive gaps.

- **Step 2: Data Transformation**
  - Change data types to correct formats (e.g., datetime, numeric).
  - Harmonize text fields (e.g., lowercase, trim whitespace).
  - Derive new attributes from existing ones (e.g., pull the year from a date).
  - Normalize or scale numerical variables to maintain consistency.

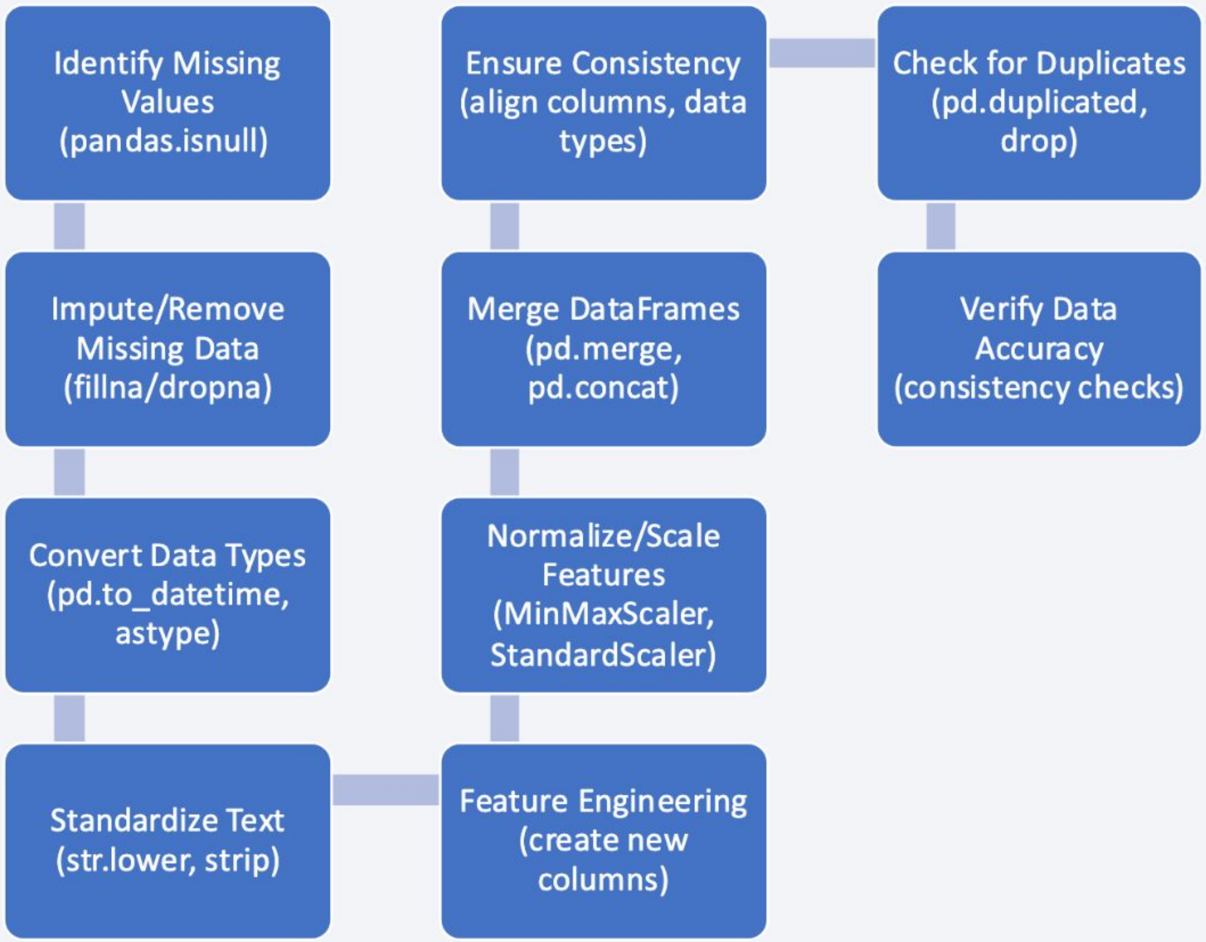
Data Wrangling

- **Step 3: Data Integration**
  - Combine data gathered from separate sources (API, scraping) into one unified dataset.
  - Harmonize column names and data formats so they match across all inputs.

- **Step 4: Data Validation**
  - Check for duplicate records and remove them.
  - Verify the accuracy and consistency of data entries.

Github URL: 3. labs-jupyter-spacex-Data wrangling.ipynb

Data Wrangling - FlowChart



EDA with Data Visualization

Overview: Exploratory Data Analysis (EDA) focuses on visually examining and summarizing the key characteristics of a dataset. The objective is to grasp distributions, detect patterns, and uncover relationships among variables.

Github URL: [5. jupyter-labs-eda-dataviz.ipynb]

1. **Histograms:**
   - **Purpose:** Display the distribution of quantitative variables such as landing success rates, payload mass, and flight count.
   - **Why:** Clarifies dispersion and central tendency, identifies outliers, and reveals data skewness.

2. **Bar Charts:**
   - **Purpose:** Contrast categorical variables like launch outcomes (success/failure) across different groups such as launch sites or rocket types.
   - **Why:** Provides a clear side-by-side view of frequencies or proportions within categories, highlighting notable trends.

3. **Line Charts:**
   - **Purpose:** Illustrate metrics over time, for instance the annual success rate of Falcon 9 launches.
   - **Why:** Exposes temporal patterns and aids in interpreting performance shifts or changes across specific periods.

4. **Scatter Plots:**
   - **Purpose:** Employed to probe the relationship between two quantitative variables—for example, payload mass versus launch success.
   - **Why:** Uncovers correlations or dependencies, visually showing how changes in one metric correspond to shifts in the other.

5. **Heatmaps:**
   - **Purpose:** Used to present correlation matrices spanning multiple numerical variables.
   - **Why:** Highlights strong positive or negative associations between variables, supporting feature selection and revealing multicollinearity.

6. **Box Plots:**
   - **Purpose:** Utilized to display the distribution of numerical data through quartiles and medians.
   - **Why:** Illustrates spread, skewness, and outliers, enabling side-by-side comparison of distributions across different categories.

EDA with SQL

Aggregate Queries:

- Computed the overall tally of launches.
- Tallied both successful and unsuccessful missions.
- Derived success percentages by launch site and rocket class.

Join Queries:

- Linked launch tables with auxiliary information (e.g., rocket specs).
- Merged multiple datasets to enable holistic analysis.

Filtering Queries:

- Isolated records based on outcome (success / failure).
- Applied conditions to retrieve launches by date or rocket configuration.

Sorting Queries:

- Ordered results to spotlight trends and anomalies.
- Listed launches chronologically or by success rate for deeper review

Subqueries:

- Nested statements to compute secondary metrics (e.g., average payload mass per site).
- Leveraged subqueries for focused analysis within large tables.

Github URL:
4. jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

Map Objects Created

Markers:

- Placed markers to indicate launch sites on the map.
- Each marker represents a specific geographical location where SpaceX launches have occurred.

Circles:

- Added circles around launch sites to visually represent proximity zones.
- Circles help visualize the areas around launch sites that might influence operational decisions.

Lines:

- Drew lines to connect launch sites with their proximities or other relevant locations.
- Lines provide spatial context and connections between different points of interest related to launches.

Reasons for Adding Objects

Markers:

- To pinpoint exact launch locations for spatial reference.
- Helps users identify where SpaceX has conducted launches geographically.

Circles:

- Illustrates the potential impact zones around launch sites.
- Provides a visual representation of safety perimeters or operational boundaries.

Lines:

- Shows connections or relationships between launch sites and relevant features.
- Enhances understanding of spatial relationships and dependencies.

Github URL: [6. lab_jupyter_launch_site_location.ipynb]

Build a Dashboard with Plotly Dash

Plots/Graphs Added

Success Pie Chart:

- Illustrates the split between successful and failed launches.
- Provides a snapshot of overall success rate and performance trends.

Success–Payload Scatter Plot:

- Reveals how payload mass relates to launch success.
- Lets users investigate how varying payload weights influence mission outcomes.

Interactions Added

Launch Site Dropdown:

- Allows users to choose specific launch sites for analysis.
- Supports filtering and focused exploration by geographic location.

Payload Range Slider:

- Enables dynamic adjustment of payload-mass ranges.
- Adds flexibility when examining launch success across different payload masses.

Github URL: spacex_dash_app.py

Predictive Analysis (Classification)

1. **Data Preprocessing:**

   • Standardized all features so each variable contributes equally.

   • Split the dataset into training and test subsets to validate models.

2. **Model Selection:**

   • Evaluated several classifiers—SVM, Decision Trees, and K-Nearest Neighbors (KNN).

   • Selected algorithms suited to a binary-classification problem and project objectives.

3. **Hyperparameter Tuning:**

   • Applied GridSearchCV to systematically locate the best hyperparameters.

   • Tuned key parameters such as C (SVM), max_depth (Decision Trees), and n_neighbors (KNN).

4. **Model Evaluation:**

   • Tested models with cross-validation to ensure robustness and generalizability.

   • Measured performance using accuracy, precision, recall, and F1-score.

5. **Improvement Iterations:**

   • Iteratively refined models based on insights from validation metrics.

   • Further adjusted hyperparameters to enhance predictive accuracy and reliability.

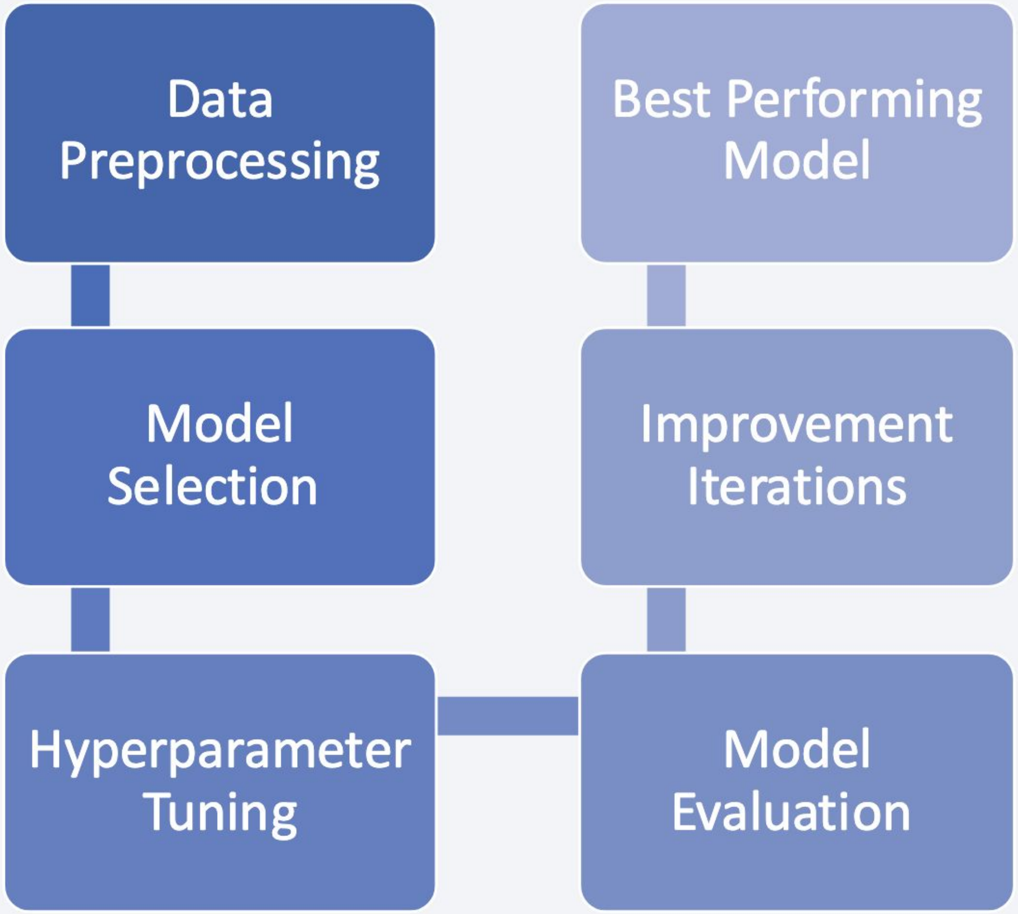
6. **Selection of Best Performing Model:**

   • Identified the model with the top test-set accuracy as the winner.

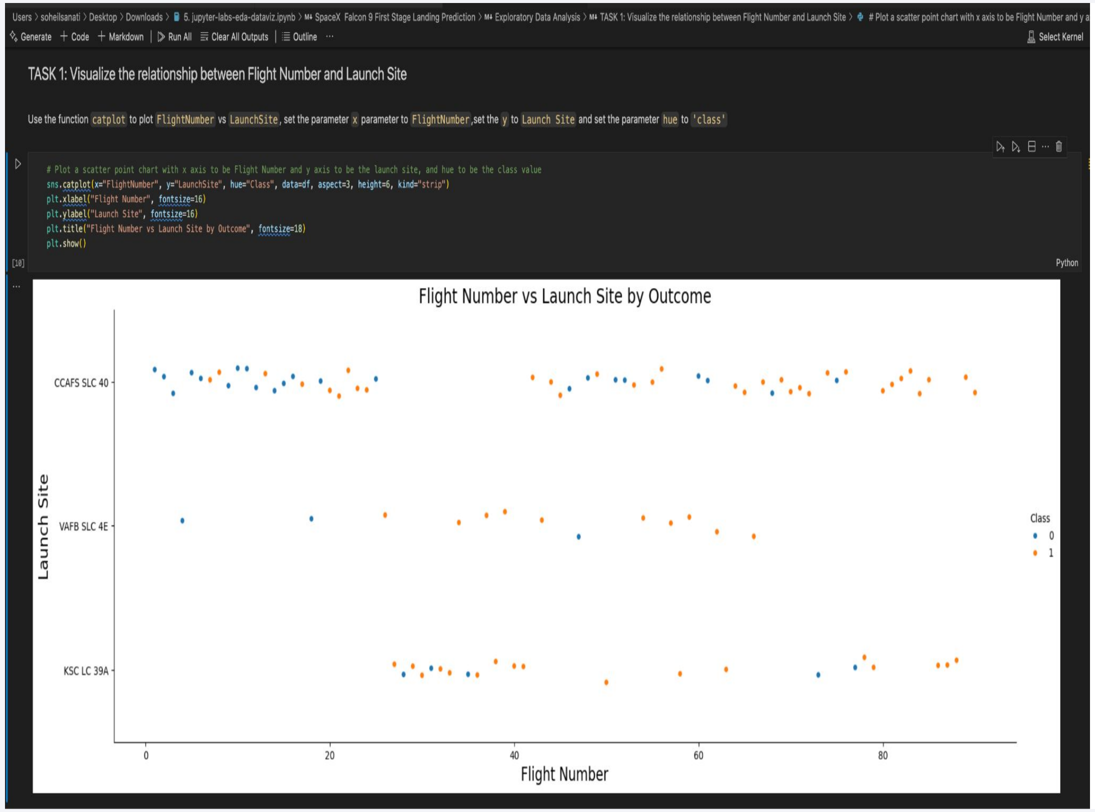
   • Weighed both training and test results to prevent overfitting and ensure practical use.

Github URL: [7. SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb]

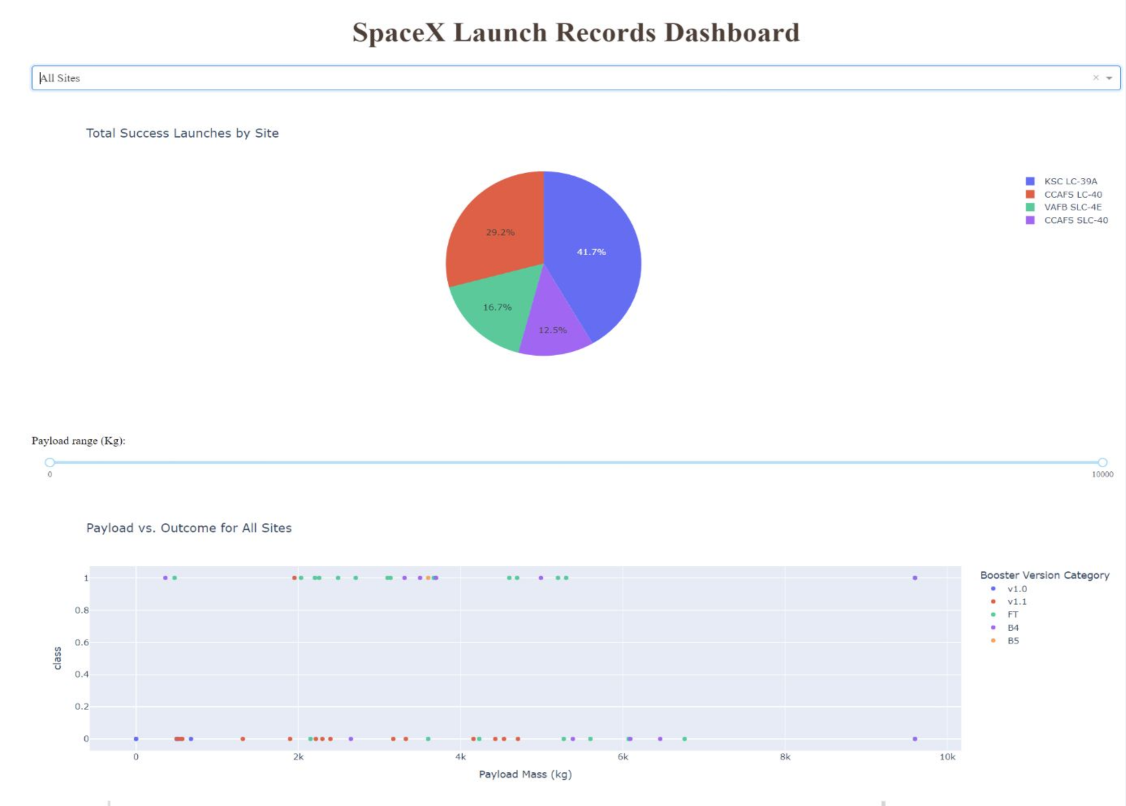
Predictive Analysis (Classification)



Results



TASK 1: Visualize the relationship between Flight Number and Launch Site
Use the function catplot to plot FlightNumber vs LaunchSite, set the parameter x parameter to FlightNumber,set the y to Launch Site and set the parameter hue to 'class'
# Plot a scatter point chart with x axis to be Flight Number and y axis to be the launch site, and hue to be the class value
sns.catplot(x="FlightNumber", y="LaunchSite", hue="Class", data=df, aspect=3, height=6, kind="strip")
plt.xlabel("Flight Number", fontsize=16)
plt.ylabel("Launch Site", fontsize=16)
plt.title("Flight Number vs Launch Site by Outcome", fontsize=18)
plt.show()
Flight Number vs Launch Site by Outcome
CCAFS SLC 40
VAFB SLC 4E
KSC LC 39A
Launch Site
Flight Number
Class
0
1



SpaceX Launch Records Dashboard
All Sites
Total Success Launches by Site
41.7%
29.2%
16.7%
12.5%
KSC LC-39A
CCAFS LC-40
VAFB SLC-4E
CCAFS SLC-40
Payload range (Kg):
0
10000
Payload vs. Outcome for All Sites
class
Payload Mass (kg)
Booster Version Category
v1.0
v1.1
FT
B4
B5

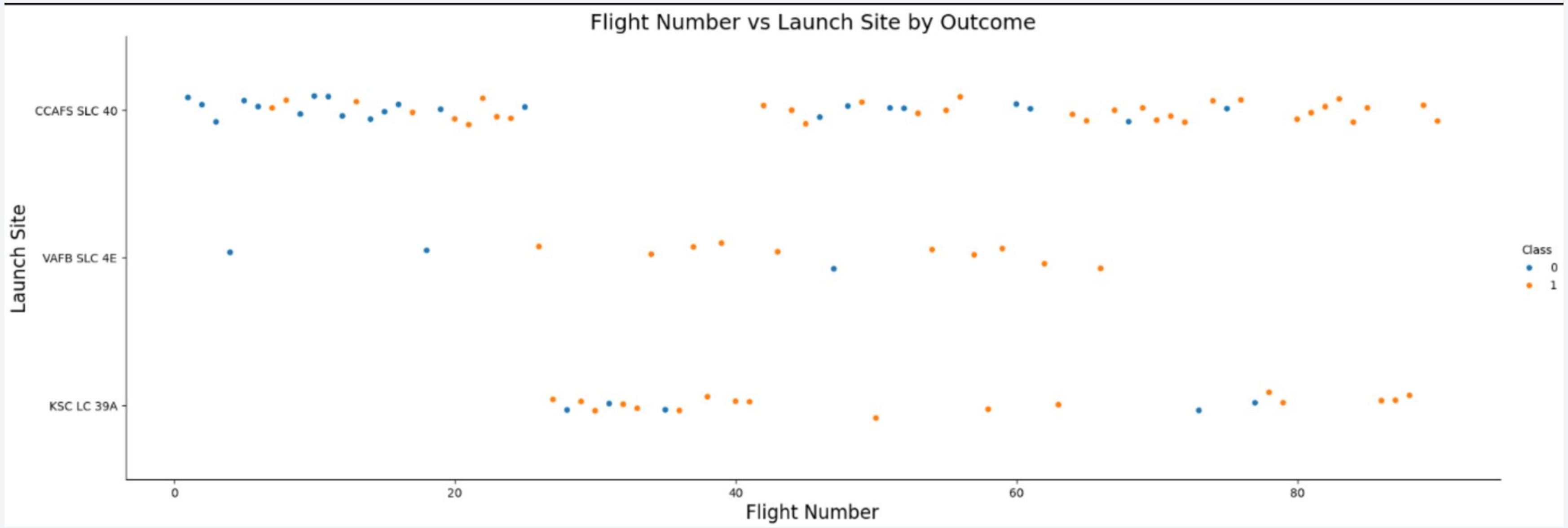
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

Varied Results at Primary Launch Venues: CCAFS SLC-40 and KSC LC-39A record both successful (orange) and unsuccessful (blue) landings, suggesting that influences beyond the pad location itself affect landing outcomes.

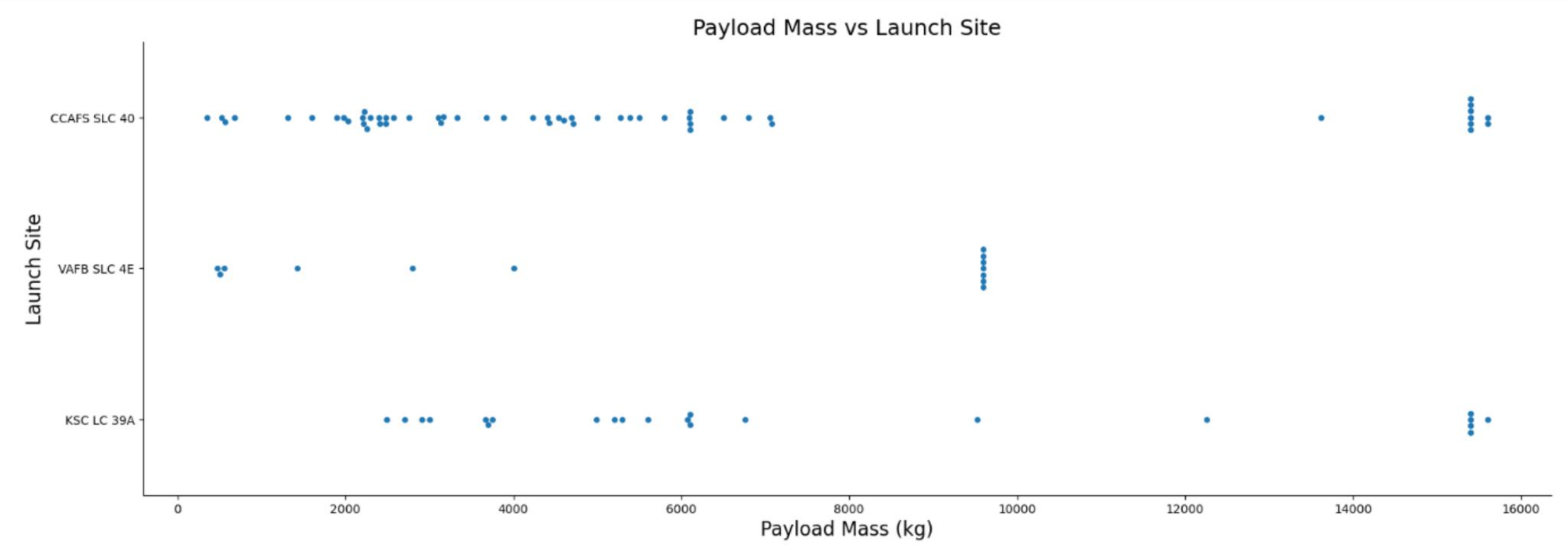
Stable Launch Volume by Flight Number: Missions span a broad range of flight indices at every site, indicating steady activity over time with no clear upward or downward trend in landing success.



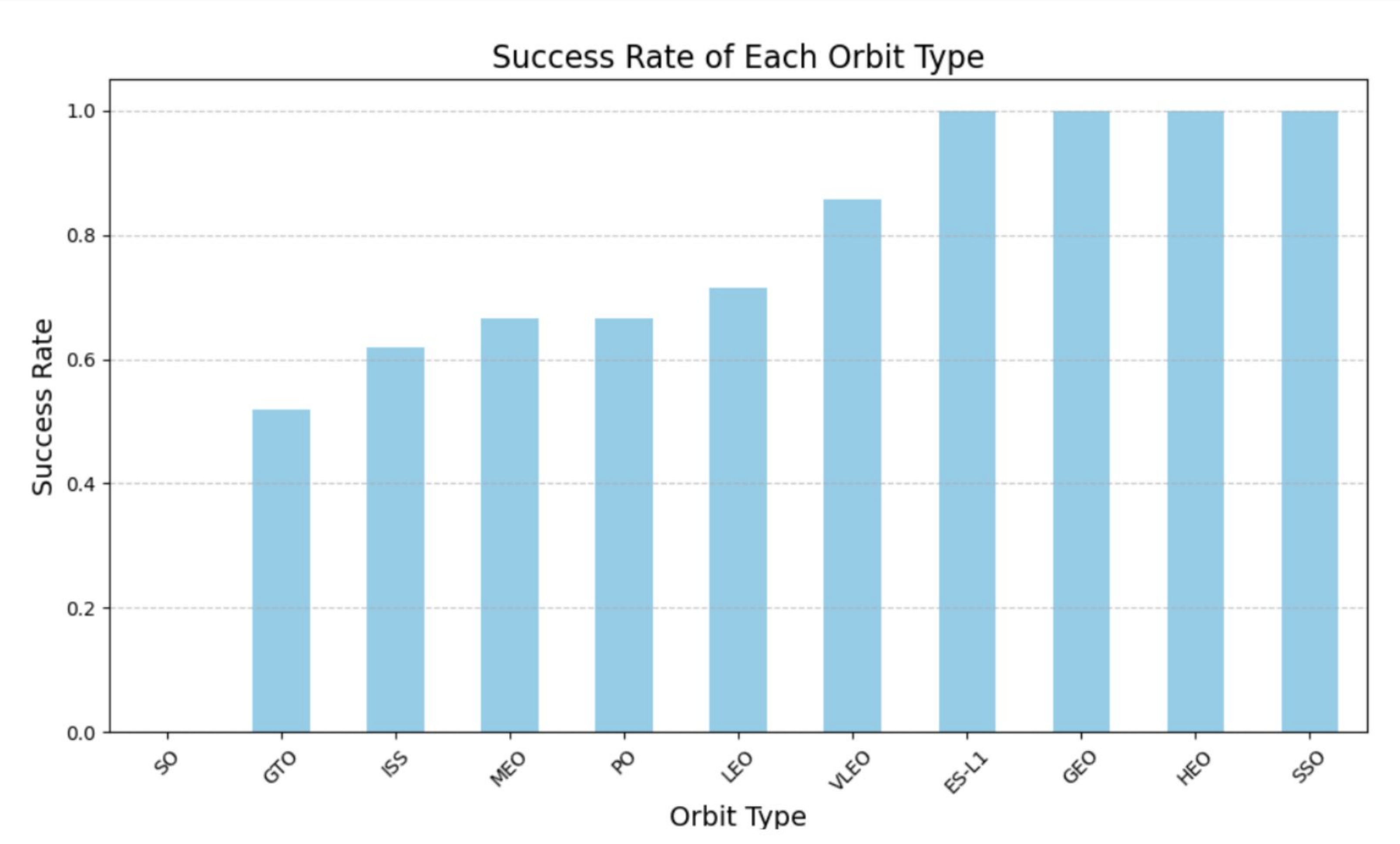
Payload vs. Launch Site

Payload Spread by Site: CCAFS SLC-40 missions mostly carry less than 10 t, whereas VAFB SLC-4E and KSC LC-39A support a broader mass band—signaling more diverse mission profiles at those pads.

Heavy-Lift Hub: KSC LC-39A repeatedly launches payloads above 15 t, underscoring its role as the preferred site for high-capacity missions.



Success Rate vs. Orbit Type



High Success Rates: Missions to VLEO, ES-L1, GEO, HEO, and SSO orbits have achieved a 100 % landing success rate, highlighting these destinations as highly reliable for first-stage recovery.

Lower Success Rate for GTO: Launches bound for GTO exhibit a notably reduced success rate compared with other orbit types, indicating greater operational challenges or complexities.

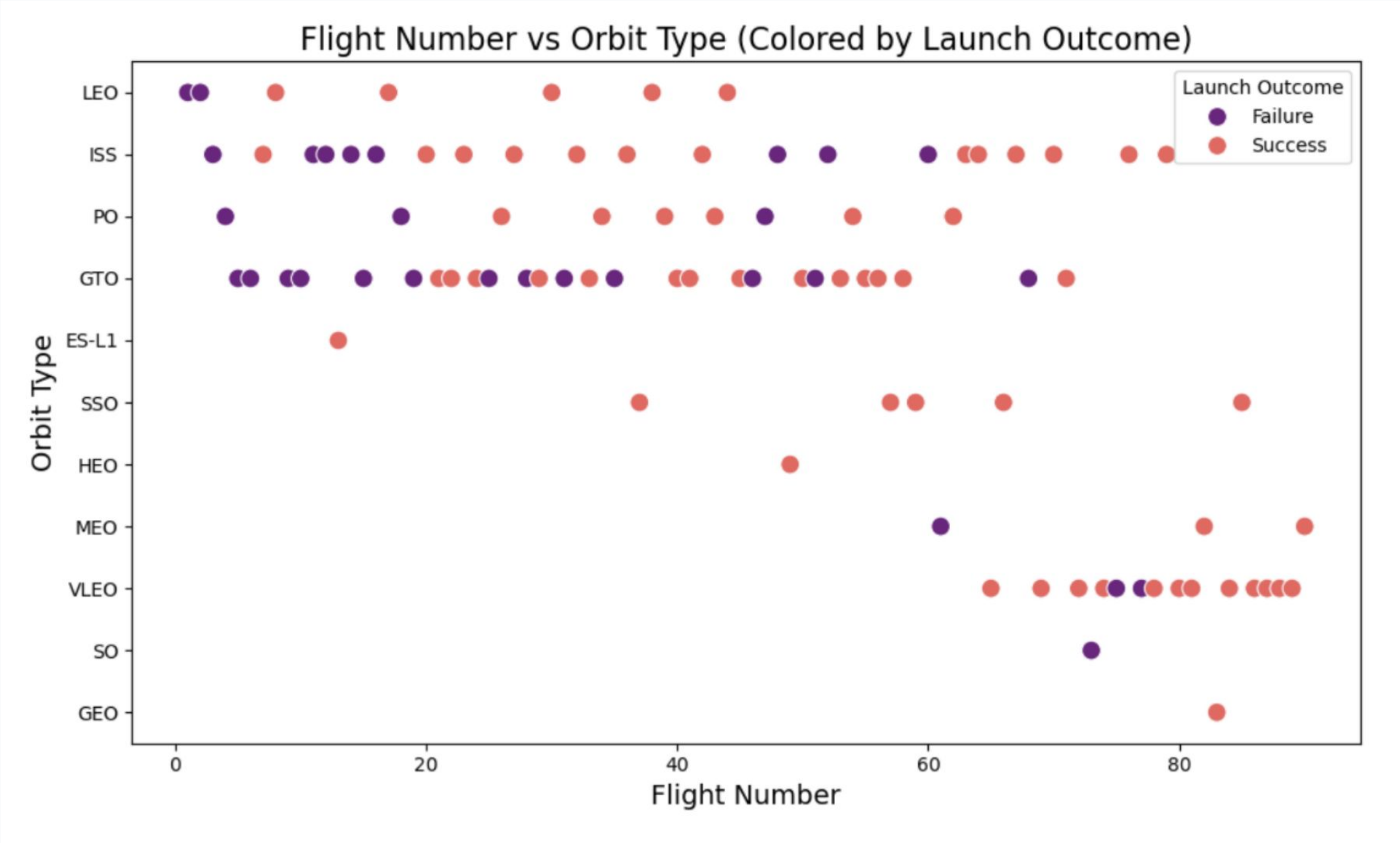
Flight Number vs. Orbit Type

Improved Success With Experience:

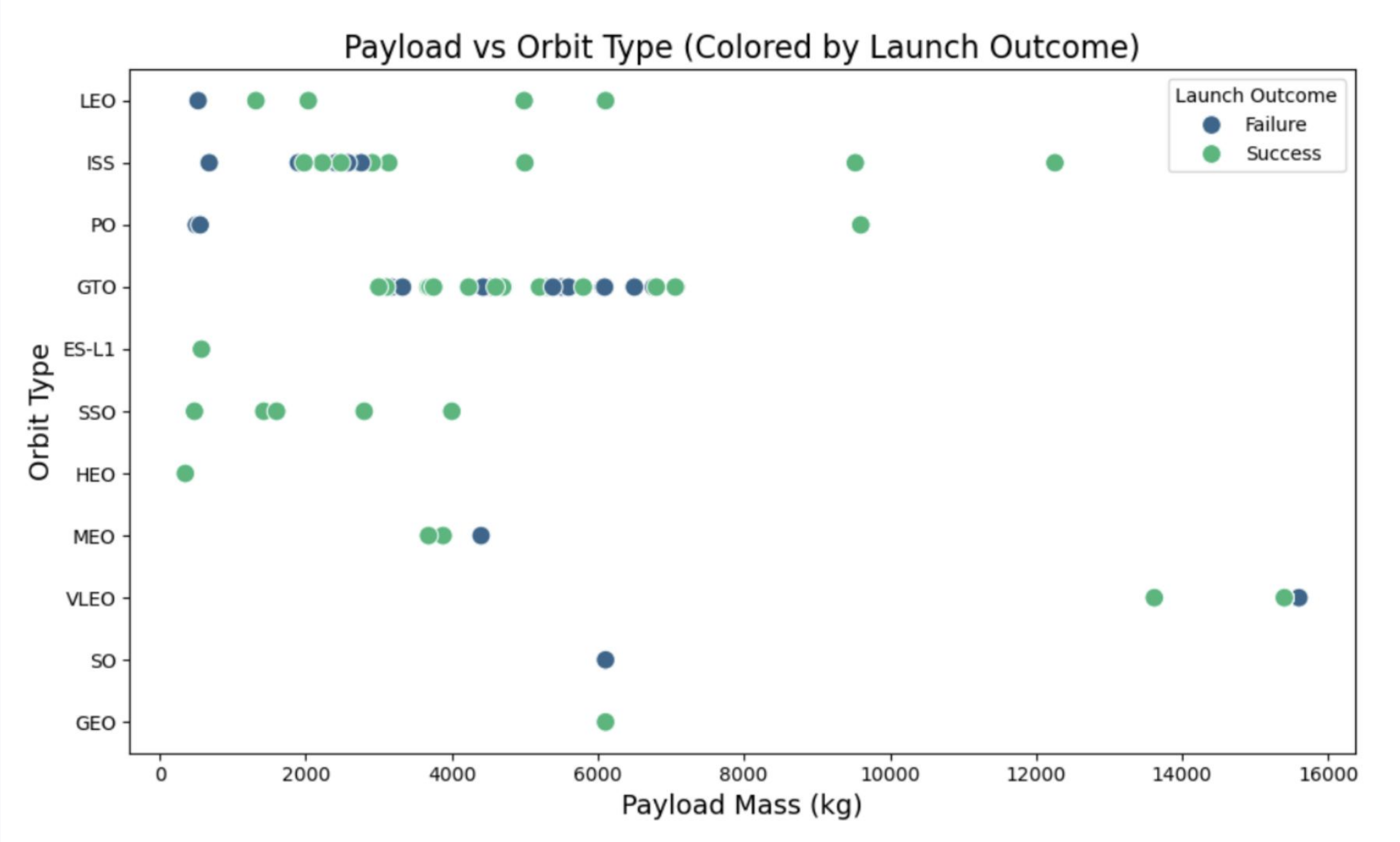
Falcon 9's landing success climbs sharply on later flight numbers, showing that accrued know-how and iterative refinements drive better outcomes.

Orbit-Specific Progress:

Early GTO and ISS missions saw mixed results, but recent launches to these orbits record far higher success rates, reflecting advances in mission planning and execution.



Payload vs. Orbit Type



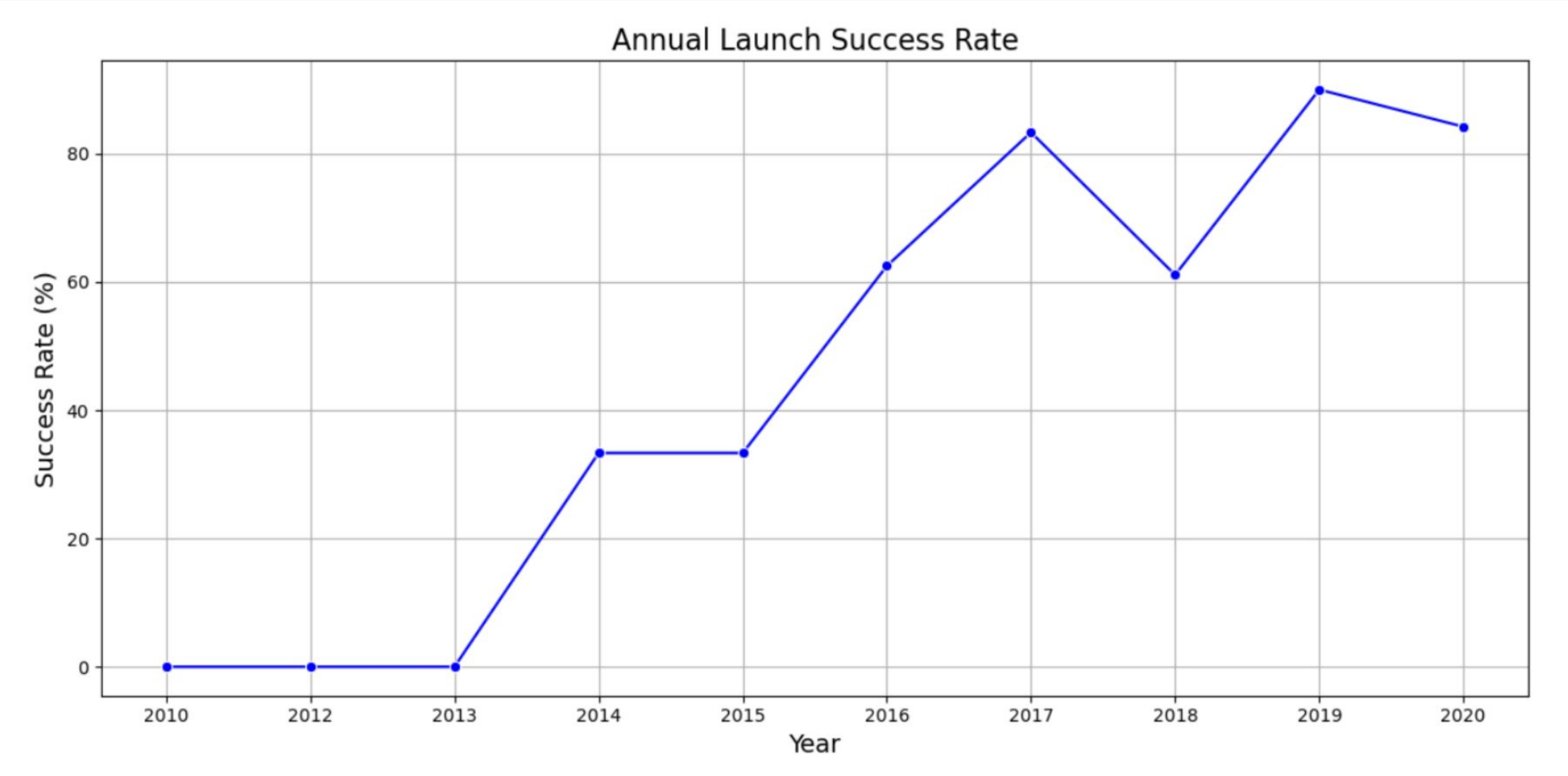
- Landing success is high across all orbit classes, with payloads under 6 000 kg enjoying the greatest success frequency.

- Payloads exceeding 10 000 kg show a mix of wins and misses, pointing the added challenges of heavier-lift missions.

Launch Success Yearly Trend

• Yearly launch success has risen sharply since 2013, topping 80 % by 2020.

• A brief dip in 2018 aside, Falcon 9 missions show a clear, long-term uptick in reliability.



All Launch Site Names

Task 1

Display the names of the unique launch sites in the space mission

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

```
 * sqlite:///my_data1.db
Done.
```

[21]:

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

Launch Site Names Begin with 'CCA'

Task 2

Display 5 records where launch sites begin with the string 'CCA'

[26]:
```
%sql SELECT * FROM SPACEXTABLE WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5;
```

```
 * sqlite:///my_data1.db
Done.
```

[26]:

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

Task 3

Display the total payload mass carried by boosters launched by NASA (CRS)

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

[30] Python

```
 * sqlite:///my_data1.db
Done.
```

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

Task 4

Display average payload mass carried by booster version F9 v1.1

[34]
```
%sql SELECT AVG("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "Booster_Version" = 'F9 v1.1';
```
Python

```
 * sqlite:///my_data1.db
Done.
```

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

Task 5

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

```
[36]  %sql SELECT MIN("Date") FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (ground pad)';
```

Python

```
 * sqlite:///my_data1.db
Done.
```

MIN(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Task 6

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

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

[38] Python

```
* sqlite:///my_data1.db
Done.
```

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

Total Number of Successful and Failure Mission Outcomes

Task 7

List the total number of successful and failure mission outcomes

```
%sql SELECT "Mission_Outcome", COUNT(*) AS "Total" FROM SPACEXTABLE WHERE "Mission_Outcome" IN ('Success', 'Failure') GROUP BY "Mission_Outcome";
```

[40] Python

```
* sqlite:///my_data1.db
Done.
```

Mission_Outcome	Total
Success	98

Boosters Carried Maximum Payload

Task 8

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

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

[42] Python

```
 * sqlite:///my_data1.db
Done.
```

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

2015 Launch Records

Task 9

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.

```
%%sql

SELECT
    CASE
        WHEN substr("Date", 6, 2) = '01' THEN 'January'
        WHEN substr("Date", 6, 2) = '02' THEN 'February'
        WHEN substr("Date", 6, 2) = '03' THEN 'March'
        WHEN substr("Date", 6, 2) = '04' THEN 'April'
        WHEN substr("Date", 6, 2) = '05' THEN 'May'
        WHEN substr("Date", 6, 2) = '06' THEN 'June'
        WHEN substr("Date", 6, 2) = '07' THEN 'July'
        WHEN substr("Date", 6, 2) = '08' THEN 'August'
        WHEN substr("Date", 6, 2) = '09' THEN 'September'
        WHEN substr("Date", 6, 2) = '10' THEN 'October'
        WHEN substr("Date", 6, 2) = '11' THEN 'November'
        WHEN substr("Date", 6, 2) = '12' THEN 'December'
        ELSE 'Unknown'
    END AS "Month_Name",
    "Mission_Outcome",
    "Booster_Version",
    "Launch_Site"
FROM
    SPACEXTABLE
WHERE
    substr("Date", 0, 5) = '2015';
```

[69] Python

```
* sqlite:///my_data1.db
Done.
```

Month_Name	Mission_Outcome	Booster_Version	Launch_Site
January	Success	F9 v1.1 B1012	CCAFS LC-40
February	Success	F9 v1.1 B1013	CCAFS LC-40
March	Success	F9 v1.1 B1014	CCAFS LC-40
April	Success	F9 v1.1 B1015	CCAFS LC-40
April	Success	F9 v1.1 B1016	CCAFS LC-40
June	Failure (in flight)	F9 v1.1 B1018	CCAFS LC-40
December	Success	F9 FT B1019	CCAFS LC-40

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

Task 10

Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.

```
%%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;
```

[81] Python

```
 * sqlite:///my_data1.db
Done.
```

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

Section 3

Launch Sites Proximities Analysis

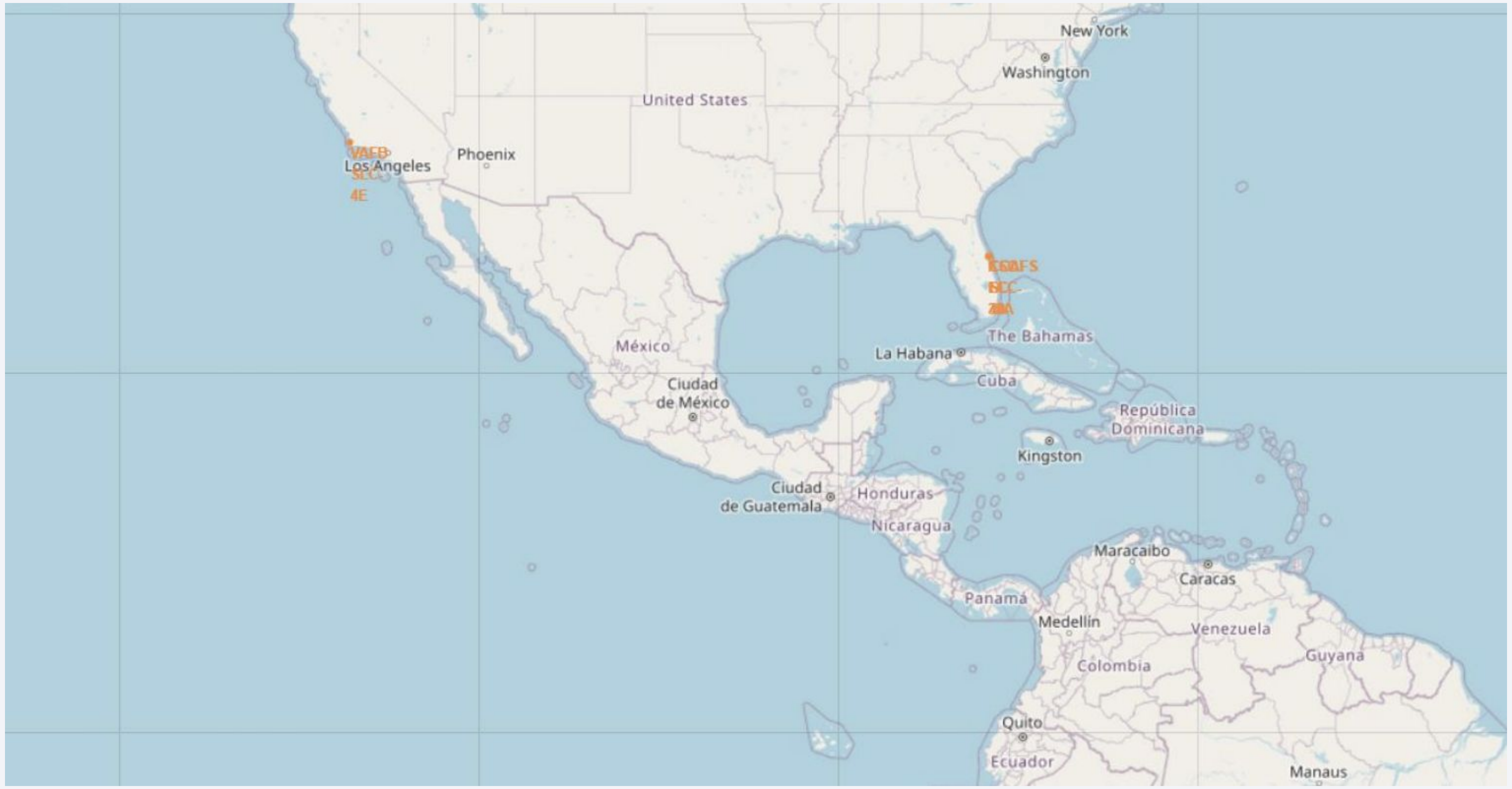
<Folium Map Screenshot 1>

1. Are all launch facilities situated near the Equator?

- Not every launch complex lies close to the Equator.
- For instance, Vandenberg Space Force Base's SLC-4E sits at roughly 34.6° N latitude—considerably farther from the Equator than the Florida pads.

2. Are all launch facilities positioned right by the shoreline?

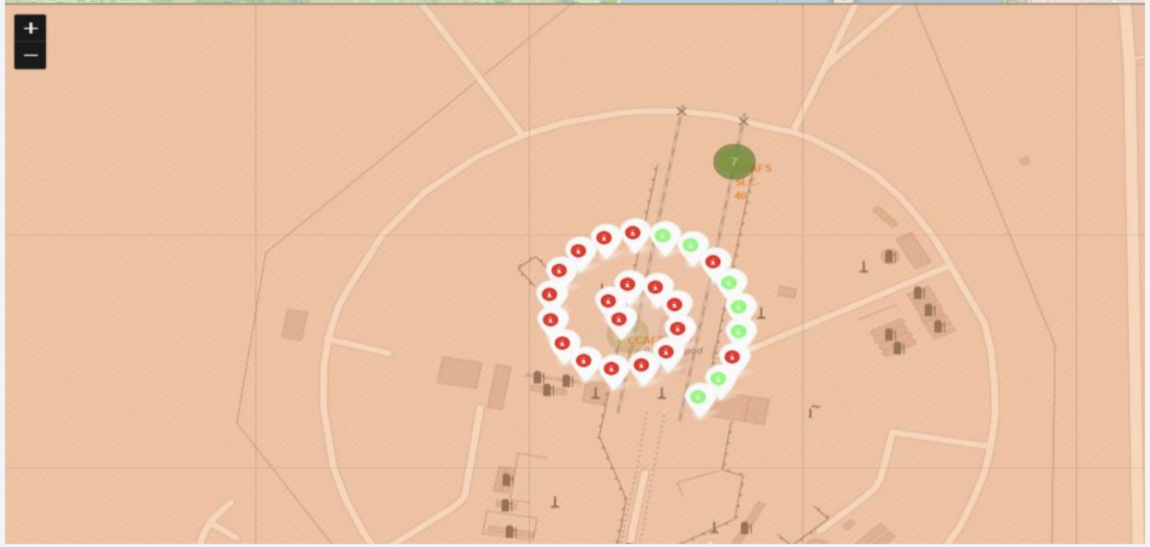
- Yes—each of the listed pads is located on or very near the coast.
- Cape Canaveral's LC-40 and SLC-40, along with Kennedy Space Center's LC-39A, border the Atlantic coast in Florida.
- Likewise, Vandenberg's SLC-4E is situated on California's Pacific coastline.



<Folium Map Screenshot 2>

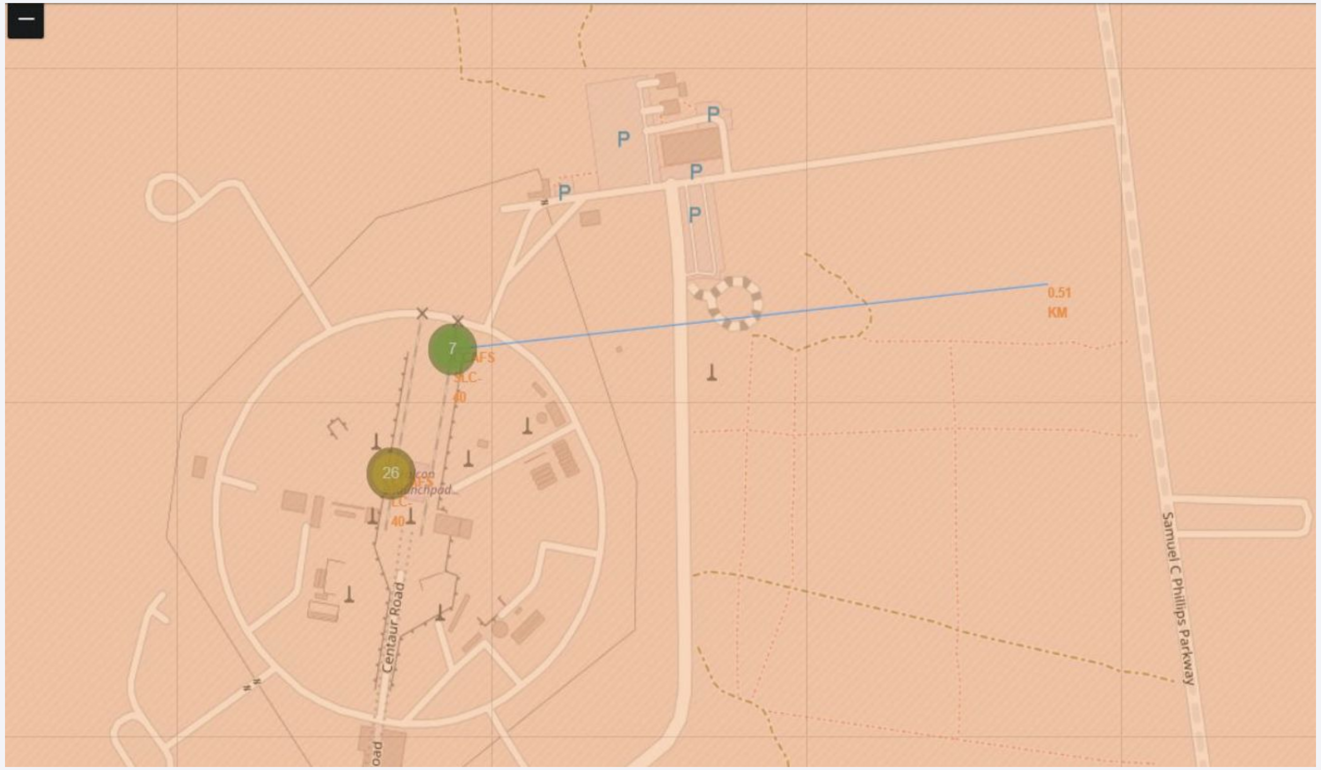
- **The upgraded map, which now employs clustered pins, lets you browse and study SpaceX launch records far more effectively.** By bunching markers together, the display stays tidy even with dozens of launches, revealing patterns that would be lost in a crowded scatter of individual points. Coupling these clusters with colour-coded symbols and information-rich pop-ups gives you quick, intuitive insight into where—and how—SpaceX has flown.

- **Consider the sample view of CCAFS LC-40:** out of 26 plotted launches, the map shows **19 red pins and 7 green pins**. This palette offers an instant visual cue: green may mark successes while red flags anomalies, or you can map it to any other categorical split you wish. Either way, the colour scheme delivers immediate feedback on launch outcomes and makes it easy to gauge performance at a glance for each pad.



<Folium Map Screenshot 3>

The figure depicts the straight-line gap between Cape Canaveral Space Force Station's SLC-40 pad and the Atlantic shoreline. A bold PolyLine links the two points, and the accompanying label shows a measured distance of about **0.51 km**. This tight coastal placement is deliberate: launching over open water offers a clear flight corridor, enables safer booster recovery, and shields nearby communities from potential launch-related debris.




```python
coastline_lat = 28.56367
coastline_lon = -80.57163

# Example launch site coordinates (replace with actual launch site coordinates)
launch_site_lat = launch_sites_df.loc[launch_sites_df['Launch Site'] == 'CCAFS SLC-40', 'Lat'].values[0]
launch_site_lon = launch_sites_df.loc[launch_sites_df['Launch Site'] == 'CCAFS SLC-40', 'Long'].values[0]

# Calculate distance using the calculate_distance function
distance_coastline = calculate_distance(launch_site_lat, launch_site_lon, coastline_lat, coastline_lon)

print(f"Distance from launch site to closest coastline: {distance_coastline} km")
```

[43] Python

```
... Distance from launch site to closest coastline: 0.5097439631188213 km
```

Section 4

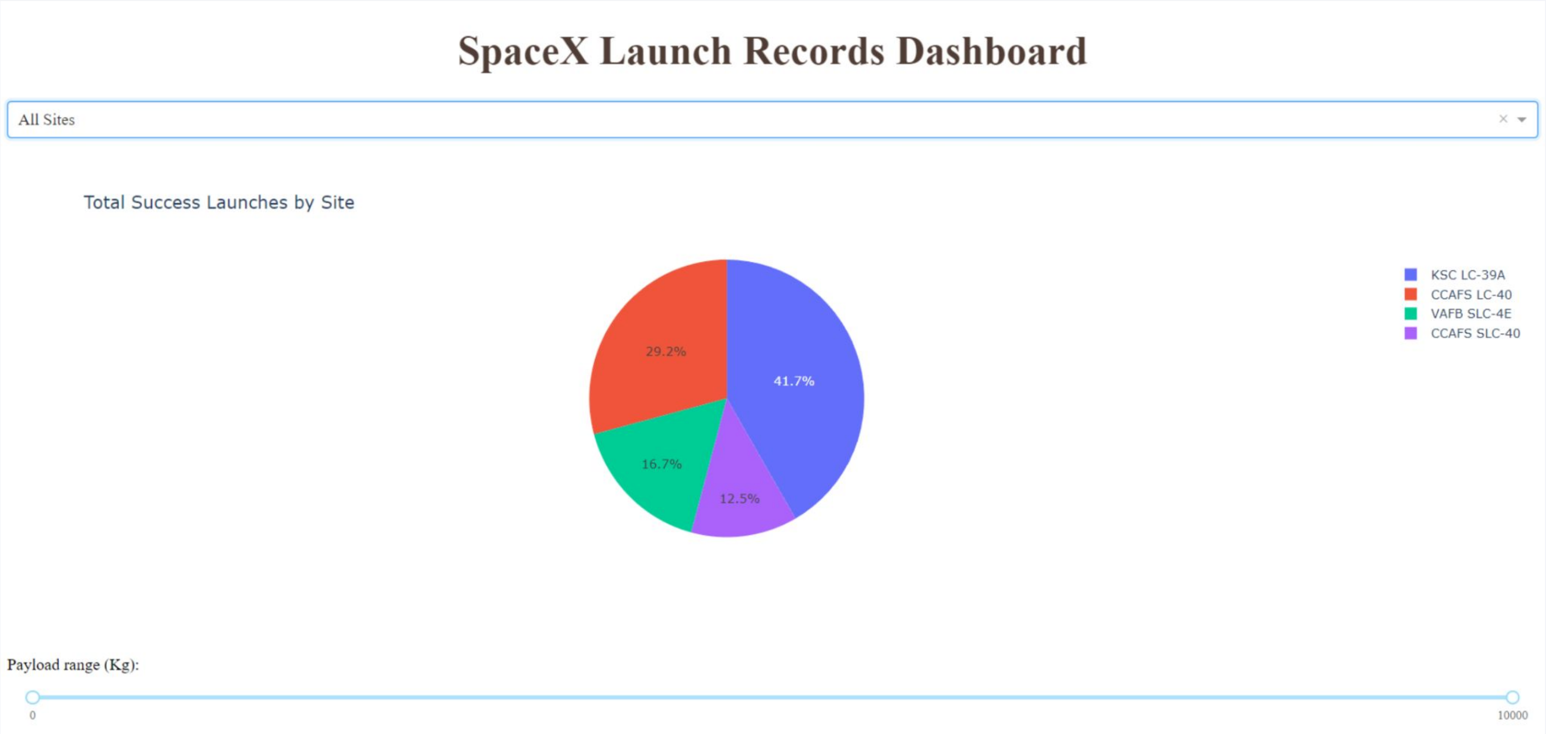
Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>

Principal Take-aways

- **CCAFS LC-40:** 29.2 % of all recorded successes
- **CCAFS SLC-40:** 12.5 %
- **VAFB SLC-4E:** 16.7 %
- **KSC LC-39A:** 41.7 %

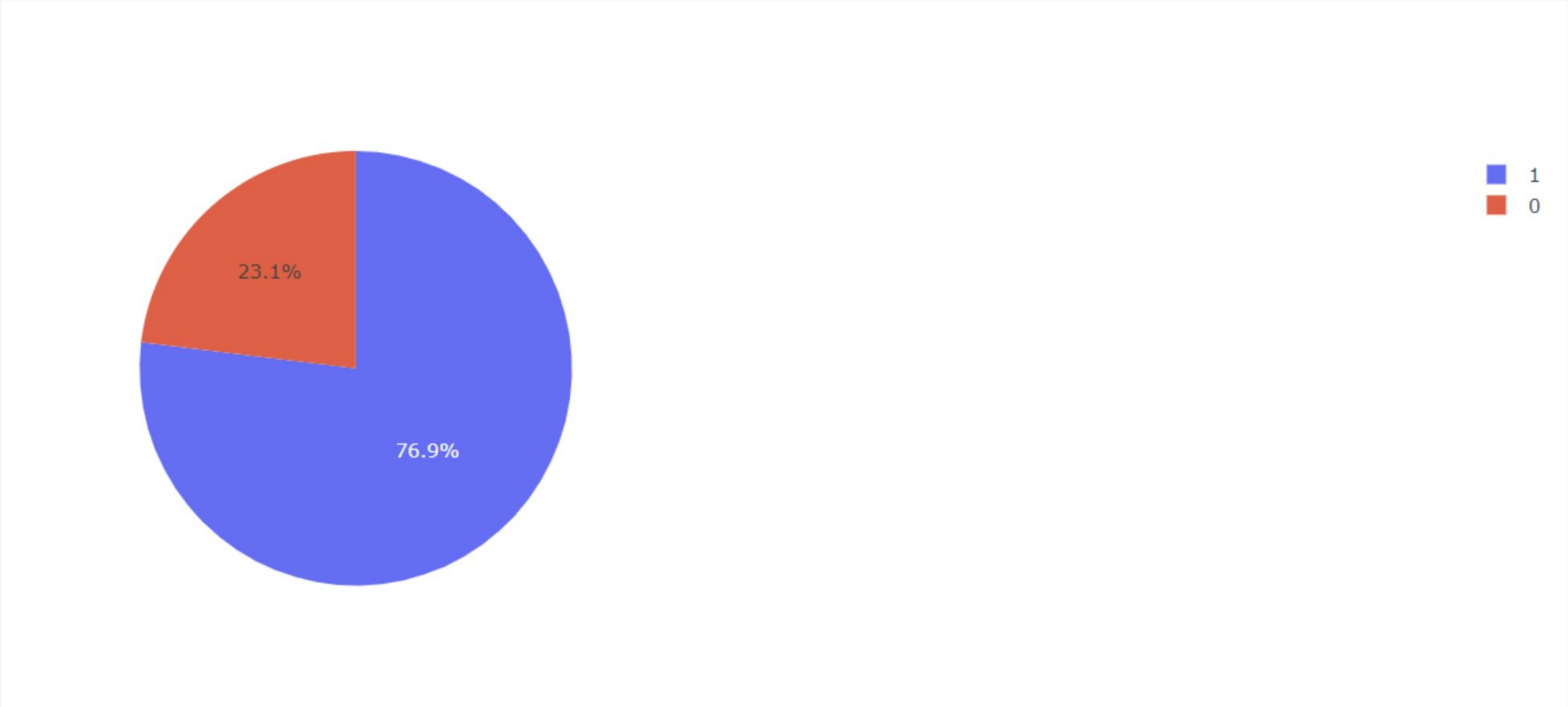
With more than two-fifths of the successful launches credited to it, **Kennedy Space Center's LC-39A stands out as SpaceX's most dependable pad to date.**



Launch Success Count for all sites (in a pie chart)

<Dashboard Screenshot 2>

Pie chart for the launch site with highest launch success ratio



Key Insights for KSC LC-39A

- The site's performance record strongly affirms its reliability: **76.9 %** of launches fall into **Class 1 (successful)**, while only **23.1 %** are **Class 0 (unsuccessful)**.
- This three-to-one ratio of successes to failures underscores LC-39A's effectiveness as SpaceX's most dependable launch complex.

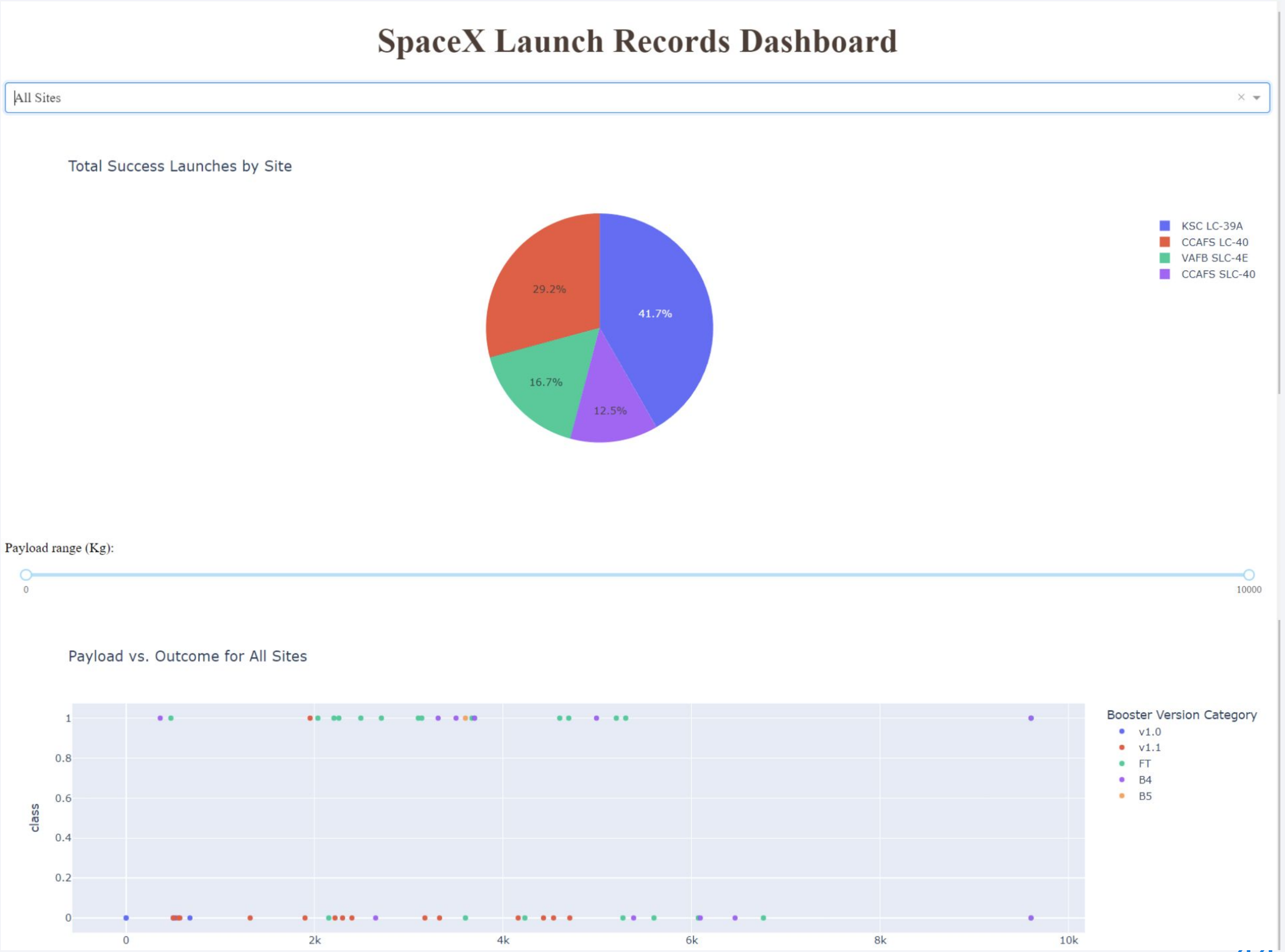
<Dashboard Screenshot 3>

Launch-Site Reliability Snapshot

- **CCAFS LC-40 tops the chart**, accounting for **43.7 % of all documented successes**—making it the most dependable pad in this comparison.
- **KSC LC-39A, VAFB SLC-4E, and CCAFS SLC-40** post lower success fractions, revealing that launch outcomes vary noticeably from one site to another.

Booster Version Highlights

- **Falcon 9 "FT"** is the fleet workhorse: it flies most often **and** maintains a consistently high success rate across a wide span of payload masses.
- **Falcon 9 "v1.0"** appears less frequently in the manifest; its smaller data set warrants deeper study before drawing firm conclusions about reliability.
- Looking across all versions, **there's no strong evidence that heftier payloads drive success rates downward**—performance remains broadly stable regardless of mass.



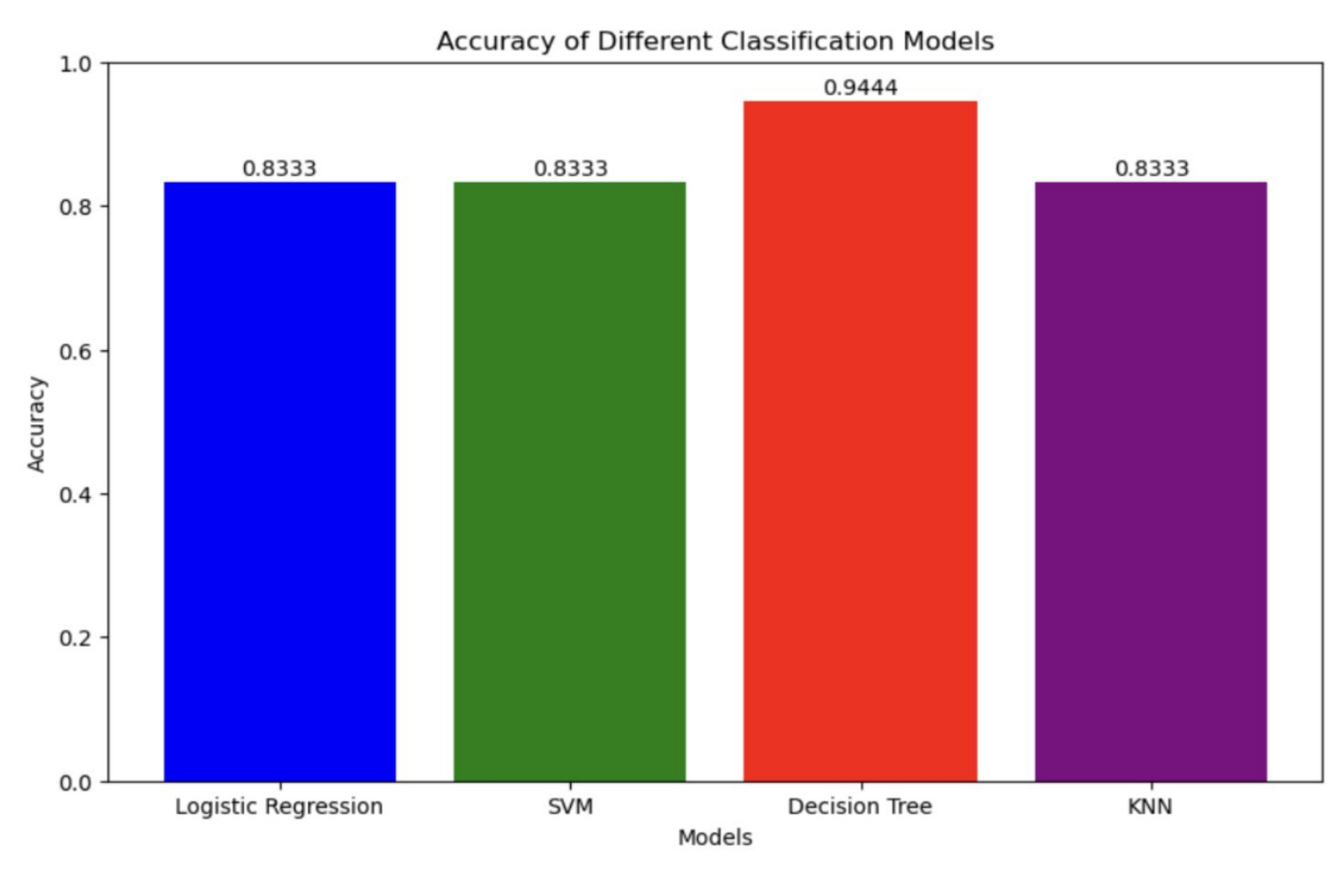
Section 5

Predictive Analysis (Classification)

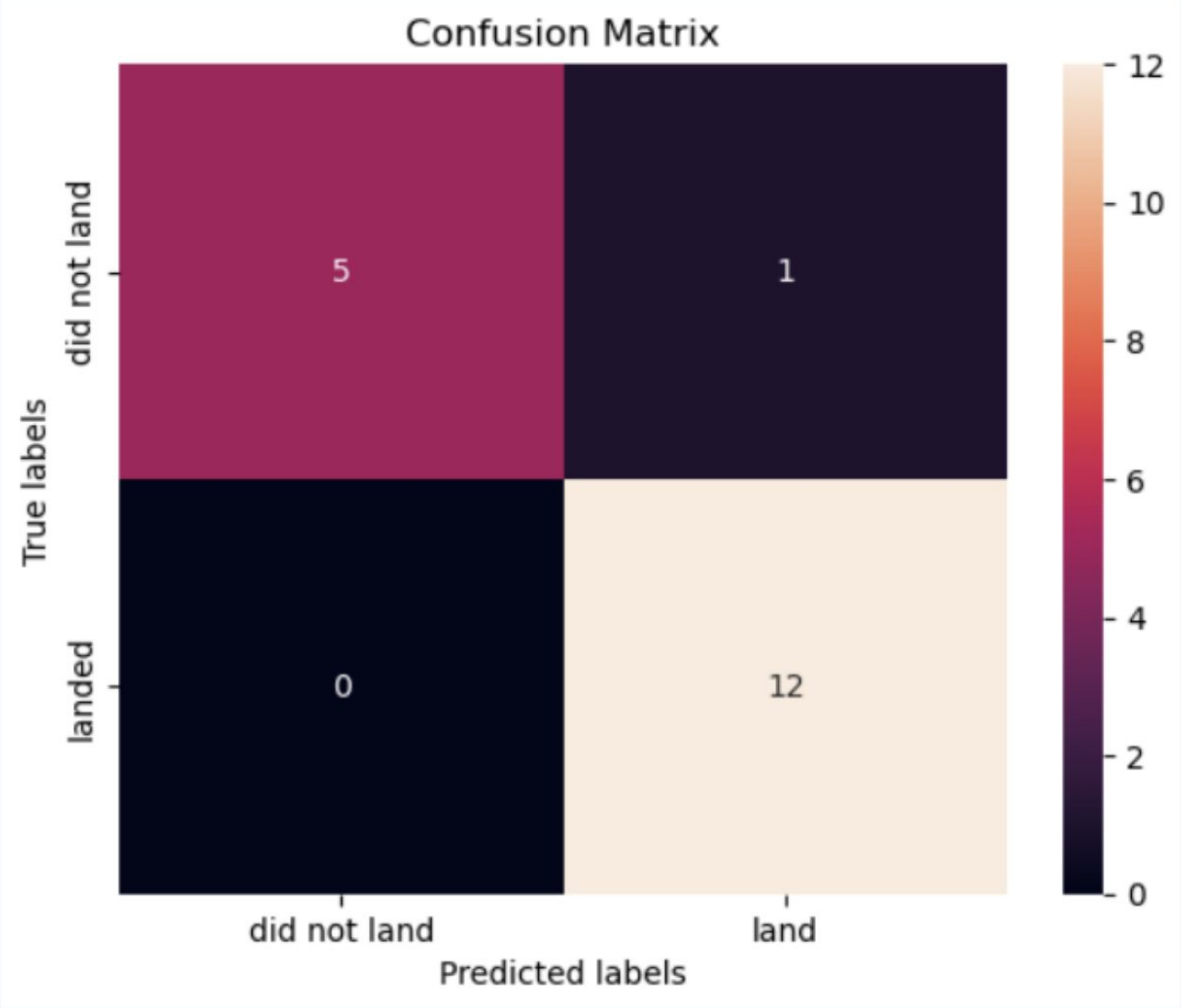
Classification Accuracy

Model-Comparison Take-Away

- **Decision Tree:** Delivers the top test-set score, clocking in at **0.9444 accuracy**.
- **Logistic Regression, Support Vector Machine, and K-Nearest Neighbors:** Each level out at **0.8333 accuracy**.



Confusion Matrix



Explanation & Insights

- **High predictive accuracy (94.44 %)** — The classifier delivers a robust overall score, supported by a strong tally of both true positives and true negatives. This shows it can reliably distinguish between successful and unsuccessful Falcon 9 first-stage landings.

- **Zero false-negative rate** — Not a single actual landing was missed. In practice, that means mission controllers never underestimate readiness or safety requirements, because every real success is correctly flagged.

- **Minimal false positives (1 case)** — A lone instance of predicting a landing that did not occur is far less disruptive than the reverse. Over-preparation is manageable; under-preparation is not—so this error profile is acceptable for operational use.

- **Balanced, success-leaning performance** — The model errs slightly on the side of predicting success, a bias well aligned with aerospace priorities where ensuring recovery is more valuable than dismissing a viable landing opportunity. This balance supports accurate cost estimation, scheduling, and resource planning.

Conclusions

Point 1 – Site performance:

Our evaluation pinpoints **CCAFS LC-40** as the front-runner, responsible for **43.7 %** of all successful flights. This dominance implies that the pad's local conditions, infrastructure, or operating procedures provide an especially favorable environment for mission success.

Point 2 – Booster reliability:

The scatter-plot review shows the **Falcon 9 "FT"** variant achieving consistently high success across a wide range of payload masses, highlighting its robustness relative to other boosters. Future campaigns could leverage this reliability advantage by prioritizing the FT configuration.

Point 3 – Payload mass vs. outcome:

We detected **no systematic link** between heavier payloads and diminished success rates. Instead, factors such as launch-site characteristics and booster type exert a far greater influence on mission outcomes.

Point 4 – Interactive visual analytics:

Leveraging **Folium** for geospatial maps and **Plotly Dash** for dynamic dashboards gave us a powerful window into SpaceX's launch patterns—both geographic and operational. These interactive tools let stakeholders drill down from continent-level views to individual pads, filter by booster version or payload mass, and watch trends emerge in real time. By turning raw numbers into intuitive visuals, the platform equips decision-makers with the comprehensive context they need to plan future missions confidently and efficiently.

In summary, the combination of predictive modeling and rich, interactive visualization has spotlighted the principal drivers behind SpaceX launch outcomes and delivered a scalable framework for ongoing evaluation. The takeaways—from site and booster performance to the negligible effect of payload mass—can guide refined launch strategies and support the continued advancement of reusable-rocket technology.

Appendix

- No Appendix

Thank you!