IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

M G Thamizh Valavan
1 Jan 2026

Outline

Executive Summary

Introduction

Methodology

Results

Conclusion

Appendix

Executive Summary

Methodologies

- Collected Falcon 9 launch data from the **SpaceX REST API** and **Wikipedia web scraping**.
- Performed **data wrangling**
- Conducted **EDA using visualizations and SQL**
- Built **interactive analytics** using **Folium maps** and a **Plotly Dash dashboard**.
- Developed and evaluated **classification models** (Logistic Regression, SVM, Decision Tree, KNN).

Summary of All Results

- Landing success improves with flight experience (flight number).
- Payload mass, orbit type, and launch site strongly affect landing outcomes.
- LEO and ISS orbits show higher success rates.
- Best classification models achieved ~83% accuracy, with Decision Tree and SVM performing best.
- Results confirm that Falcon 9 landing success is predictable using historical mission data.

Introduction

Project Background and Context

- SpaceX's Falcon 9 rocket is designed for **first-stage reusability**, significantly reducing launch costs.
- Successful recovery of the first stage depends on multiple **mission, orbital, and operational factors**.
- Understanding these factors through **data-driven analysis** is critical for improving mission planning and reliability.
- This project applies **data science and machine learning techniques** to analyze historical Falcon 9 launch data.

Problems You Want to Find Answers

- What factors most strongly influence first-stage landing success?
- How do payload mass, orbit type, launch site, and flight number affect outcomes?
- Are there observable trends in launch success over time?
- Can landing success be accurately predicted using classification models based on historical data?

Section 1

Methodology

Methodology

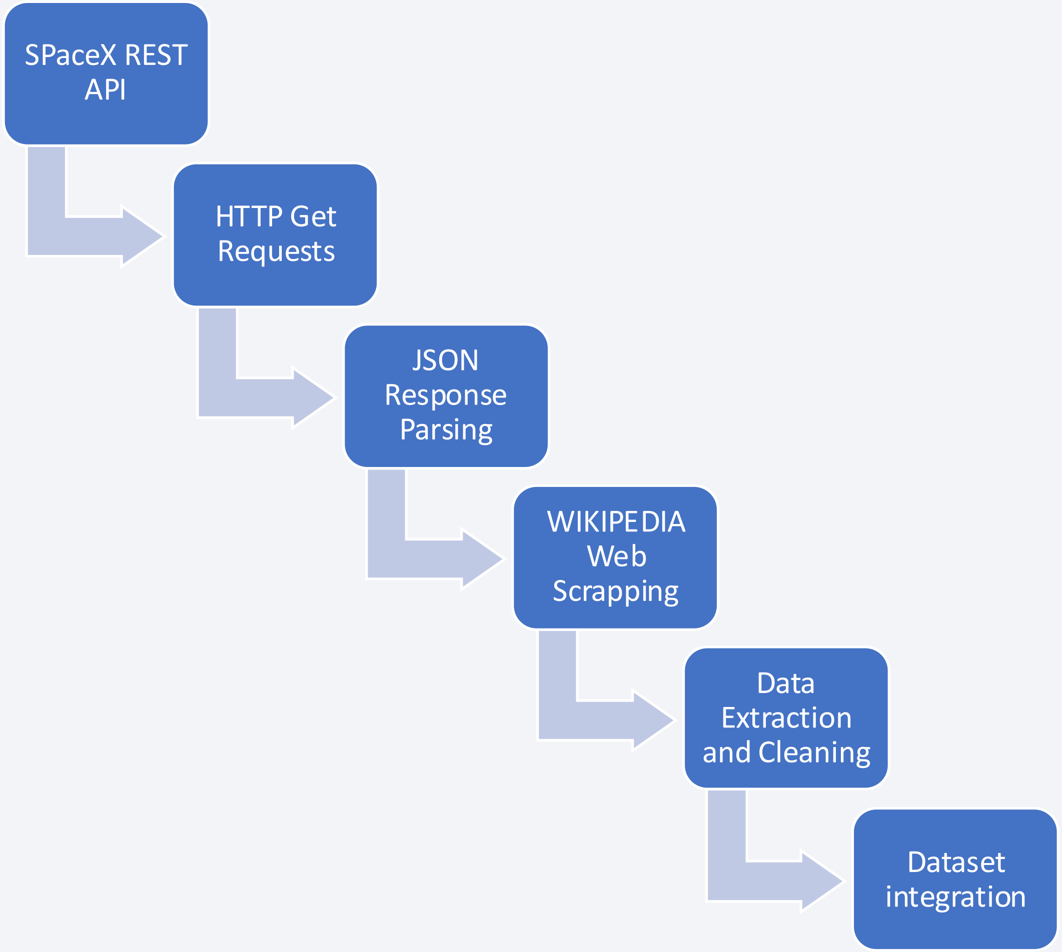
Executive Summary

- Data collection methodology:
  - Collected Falcon 9 launch data via **SpaceX REST API** and **Wikipedia web scraping**.
- Perform data wrangling
  - Performed **data wrangling and preprocessing**, including filtering, missing-value handling, and feature engineering.
- Data analysis (EDA) using visualization and SQL
- Built **interactive analytics** using **Folium maps** and a **Plotly Dash dashboard**
- Perform predictive analysis using classification models
  - Logistic Regression, SVM, Decision Tree and KNN

Data Collection

How the Data Sets Were Collected

- •Retrieved structured launch data using the **SpaceX REST API**.
- •Extracted mission details including **rocket, payload, orbit, launch site, and landing outcome**.
- •Performed **web scraping on Wikipedia** to supplement historical Falcon 9 launch records.
- •Combined API and scraped data to ensure **data completeness and consistency**.
- •Stored collected data in **structured tabular format** for downstream analysis.



Data Collection – SpaceX API

- **SpaceX API Data Collection**
- Collected launch data using **SpaceX REST API endpoints**.
- Retrieved structured data via **HTTP GET requests**.
- Parsed **JSON responses** to extract:
  - Launch date and flight number
  - Rocket and booster details
  - Payload mass and orbit
  - Launch site and landing outcome
- Converted extracted fields into a **structured Pandas DataFrame** for analysis.

- Add the GitHub URL of the completed SpaceX API calls notebook (must include completed code cell and outcome cell), as an external reference and peer-review purpose

SpaceX REST API

↓

HTTP GET Requests

↓

JSON Responses

↓

Parse & Extract Fields

↓

Structured Falcon 9 Dataset

Data Collection - Scraping

- **Web Scraping Process**
- Scraped Falcon 9 launch records from **Wikipedia** using **BeautifulSoup**.
- Sent HTTP requests to retrieve **HTML pages**.
- Parsed HTML tables to extract **launch dates, booster versions, payloads, and landing outcomes**.
- Cleaned and standardized scraped data for consistency.
- Integrated scraped data with API data to improve **historical coverage and completeness**.
- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose

HTTP Request

↓

Wikipedia Web Page

↓

HTML Response

↓

BeautifulSoup Parsing

↓

Table Extraction

Data Wrangling

How the Data Were Processed

- Filtered the dataset to include **Falcon 9 launches only**.
- Removed irrelevant and redundant columns.
- Handled missing values (e.g., **payload mass imputed using mean**).
- Standardized variable names and data types.
- Converted categorical variables into **numerical features**.
- Created a **binary target variable** representing landing success.

Raw Datasets
↓
Data Cleaning
↓
Missing Value Handling
↓
Feature Engineering
↓
Categorical Encoding
↓
Final Structured Dataset

Add the GitHub URL of your completed data wrangling related notebooks, as an external reference and peer-review purpose

EDA with Data Visualization

Charts Used and Purpose

- **Scatter plots** (e.g., payload vs. flight number, payload vs. launch site) to examine relationships between mission variables and landing success.
- **Bar charts** to compare **success rates across orbit types and launch sites**.
- **Line charts** to analyze **trends in launch success over time**.
- **Categorical plots** to assess the impact of **reusability features** (grid fins, legs, reuse) on landing outcomes.
- **Heatmaps** to identify correlations between numerical features and landing success.

Add the GitHub URL of your completed EDA with data visualization notebook, as an external reference and peer-review purpose

Why These Charts Were Used

- To visually detect **patterns, trends, and anomalies** in launch data.
- To compare **categorical variables** influencing landing outcomes.
- To support and validate findings from **SQL-based analysis**.
- To guide **feature selection** for predictive modeling.

EDA with SQL

- Retrieved **unique launch site names** from the dataset.
- Filtered launch records based on **launch site patterns** (e.g., sites beginning with “CCA”).
- Calculated **total and average payload mass** by booster version.
- Identified the **first successful ground landing date**.
- Extracted records of **successful drone ship landings** within specific payload ranges.
- Computed **counts of successful and failed missions**.
- Identified **boosters carrying the maximum payload mass.**
- Analyzed **failed landing outcomes in drone ship for the year 2015**.
- Ranked **landing outcomes by frequency** within a given date range.

- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose

Build an Interactive Map with Folium

- **Map Objects Created**
- **Markers** to display the geographic locations of SpaceX launch sites.
- **Color-coded markers** to distinguish successful and failed landing outcomes.
- **Circles** to represent launch site proximity zones.
- **Lines (polylines)** to calculate and visualize distances between launch sites and nearby features such as coastlines.
- **Pop-ups and tooltips** to display launch site details interactively.
- Explain why you added those objects

Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

Why These Objects Were Added

•To provide a **geospatial overview** of Falcon 9 launch activity.

•To visually compare **landing success and failure by location**.

•To analyze the **impact of geographic proximity** (coastline, infrastructure) on launch and landing operations.

•To enable **interactive exploration** of spatial patterns not evident in static charts.

Build a Dashboard with Plotly Dash

- **Plots and Interactions Added**
- **Pie chart** showing launch success vs. failure across all launch sites.
- **Pie chart** highlighting success ratio for a selected launch site.
- **Scatter plot** of payload mass vs. launch outcome.
- **Dropdown menu** to select launch sites dynamically.
- **Range slider** to filter payload mass interactively.
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Why These Plots and Interactions Were Added

•To enable **interactive exploration** of launch success patterns.

•To compare **success rates across launch sites** dynamically.

•To analyze how **payload mass influences landing outcomes**.

•To allow users to **filter and drill down** into specific mission scenarios.

Predictive Analysis (Classification)

- **Model Development Process**
- Defined landing outcome as a **binary classification problem**.
- Selected features based on **EDA and domain relevance**.
- Split data into **training and testing sets**.
- Built multiple models: **Logistic Regression, SVM, Decision Tree, KNN**.
- Tuned hyperparameters using **GridSearchCV** and cross-validation.
- Evaluated models using **test accuracy and confusion matrices**.
- Selected the **best-performing model** based on overall accuracy and error balance.
- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose

Prepared Dataset
↓
Feature Selection
↓
Train–Test Split
↓
Model Training
↓
Hyperparameter Tuning
↓
Model Evaluation
↓
Best Classification Model

Results

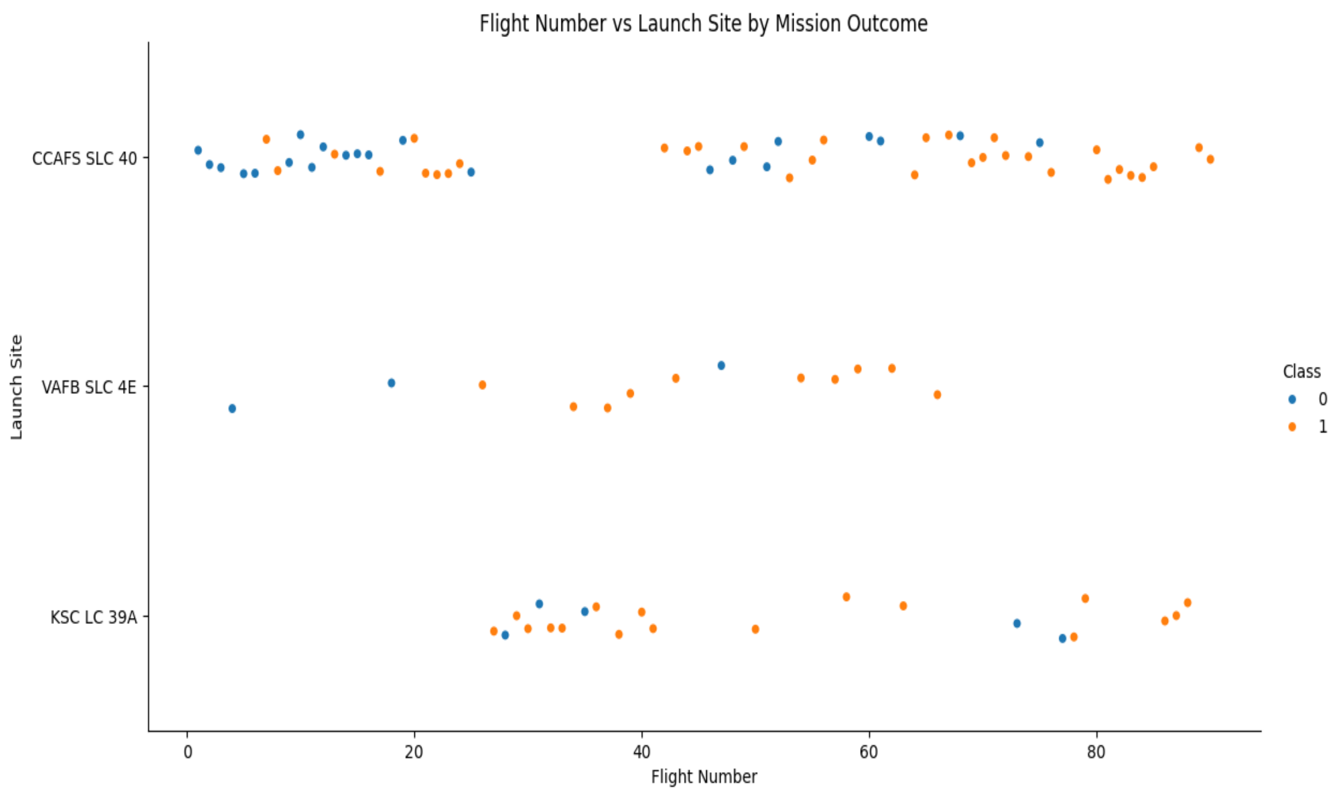
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

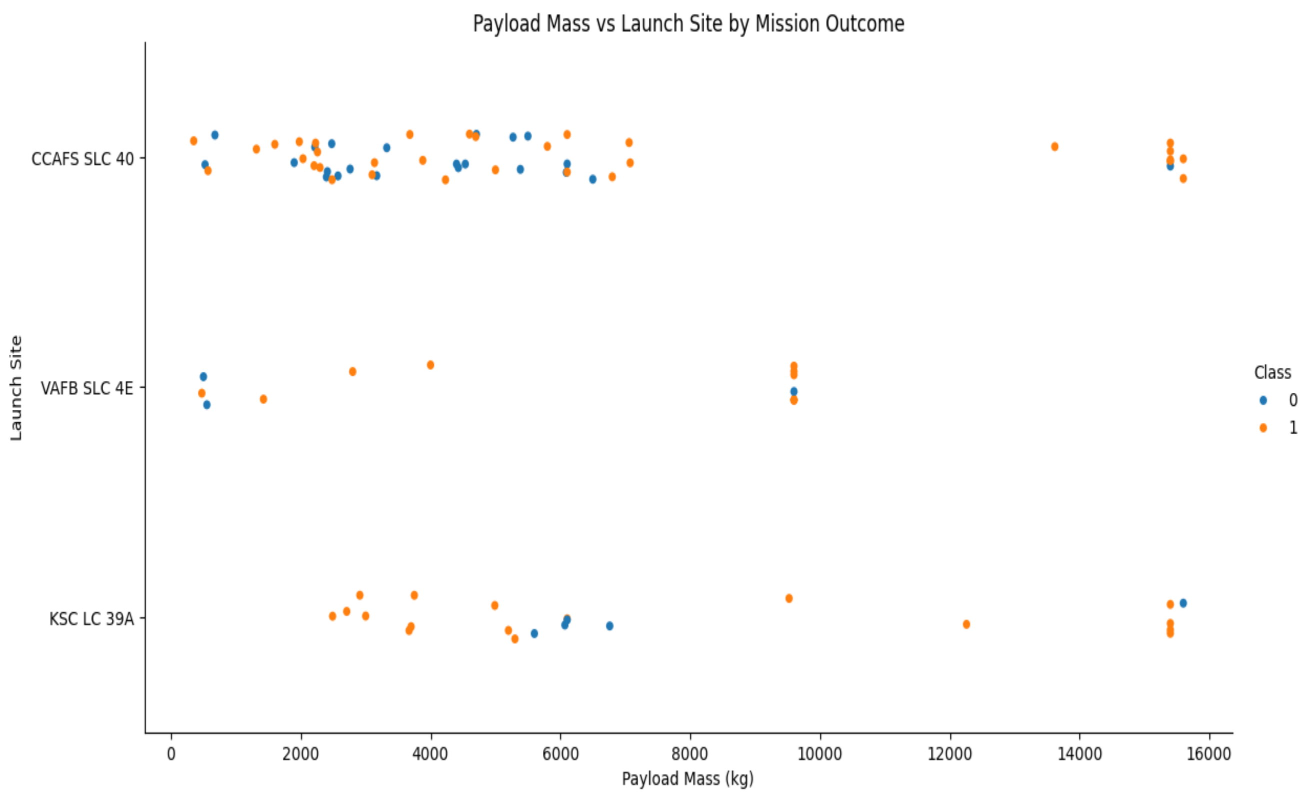
Flight Number vs. Launch Site

- Early missions (low flight numbers) show a **higher concentration of failures** across all launch sites.
- As the **flight number increases**, the number of successful landings increases significantly.
- This trend is visible across **CCAFS SLC 40, VAFB SLC 4E, and KSC LC 39A**, indicating improved operational reliability over time.
- Later missions from **KSC LC 39A and CCAFS SLC 40** show a higher density of successful outcomes.
- The pattern suggests a strong **learning effect**, where accumulated launch experience improves landing success.



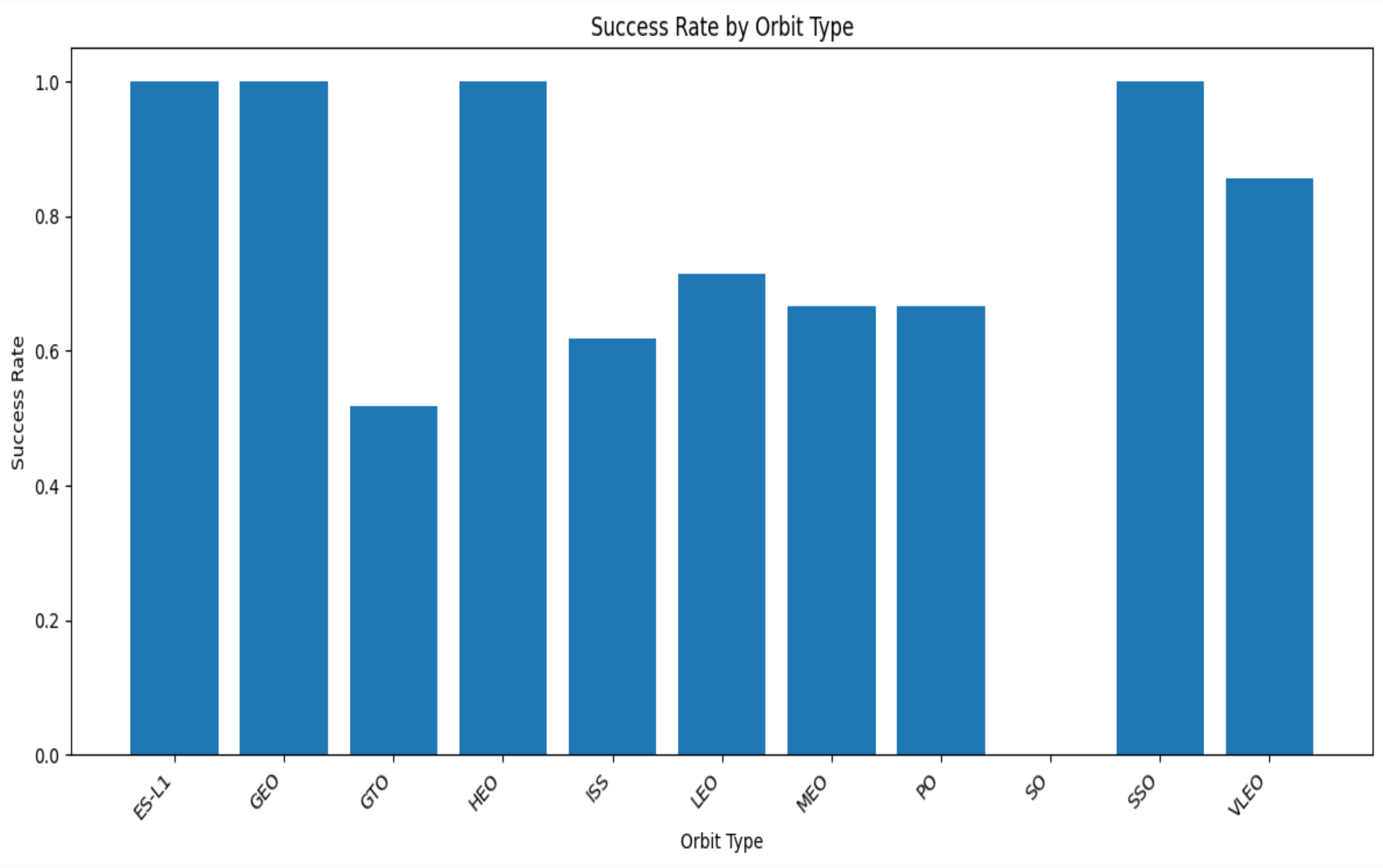
Payload vs. Launch Site

- Successful landings occur across a **wide range of payload masses**, but lighter payloads show a higher concentration of success.
- **Higher payload masses** are associated with fewer successful landings, particularly in earlier missions.
- **CCAFS SLC 40** supports the widest payload range with a high number of successful outcomes.
- **KSC LC 39A** shows improved success at higher payload masses in later missions.
- **VAFB SLC 4E** has fewer launches and a more limited payload range, making trends less pronounced.



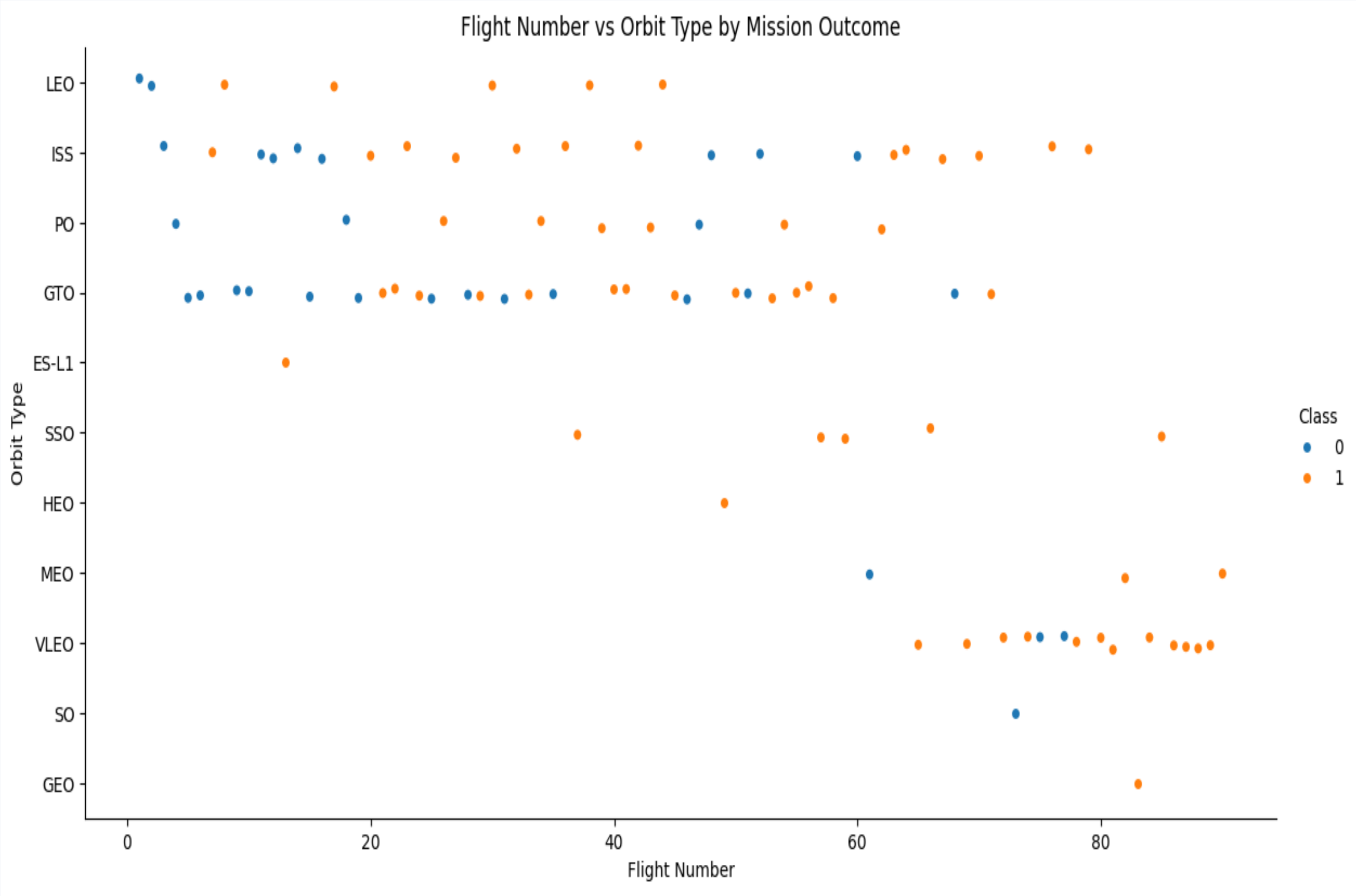
Success Rate vs. Orbit Type

- **LEO and ISS missions** show consistently high landing success rates.
- **GTO missions** have noticeably lower success rates, reflecting higher mission complexity and energy requirements.
- **SSO and HEO orbits** also demonstrate high success rates, though with fewer missions.
- Medium Earth and polar orbits show **moderate success**, indicating varying mission difficulty.
- Orbit type is a **strong operational factor** influencing first-stage landing success.



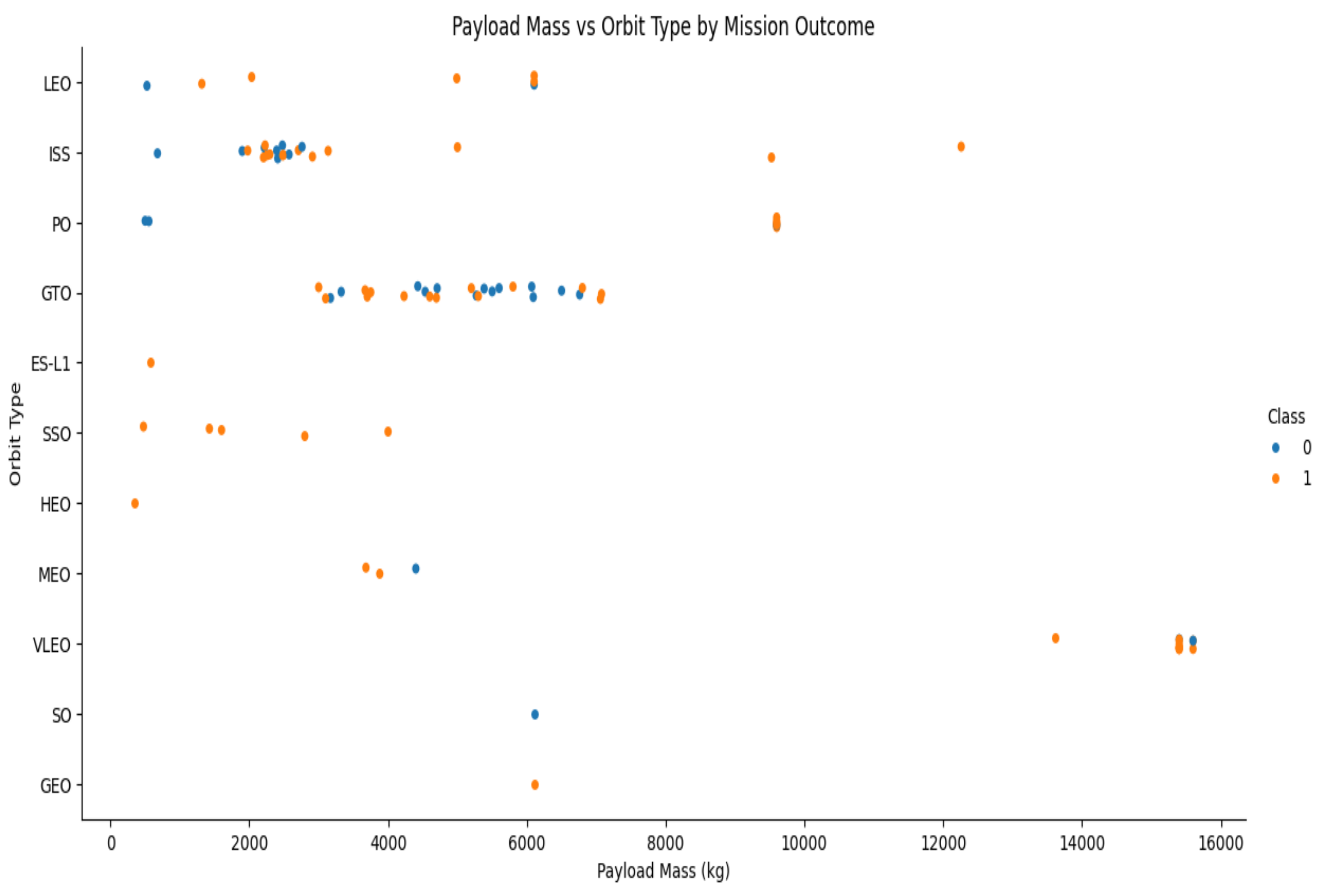
Flight Number vs. Orbit Type

- Early missions across most orbit types show a **higher proportion of failures**, especially for complex orbits such as **GTO**.
- As the **flight number increases**, successful landings become more frequent across nearly all orbit types.
- **LEO and ISS missions** exhibit consistently high success rates as experience increases.
- **GTO missions** improve over time but still show more variability compared to lower-energy orbits.
- Later missions to **VLEO and SSO** are predominantly successful, reflecting operational maturity.



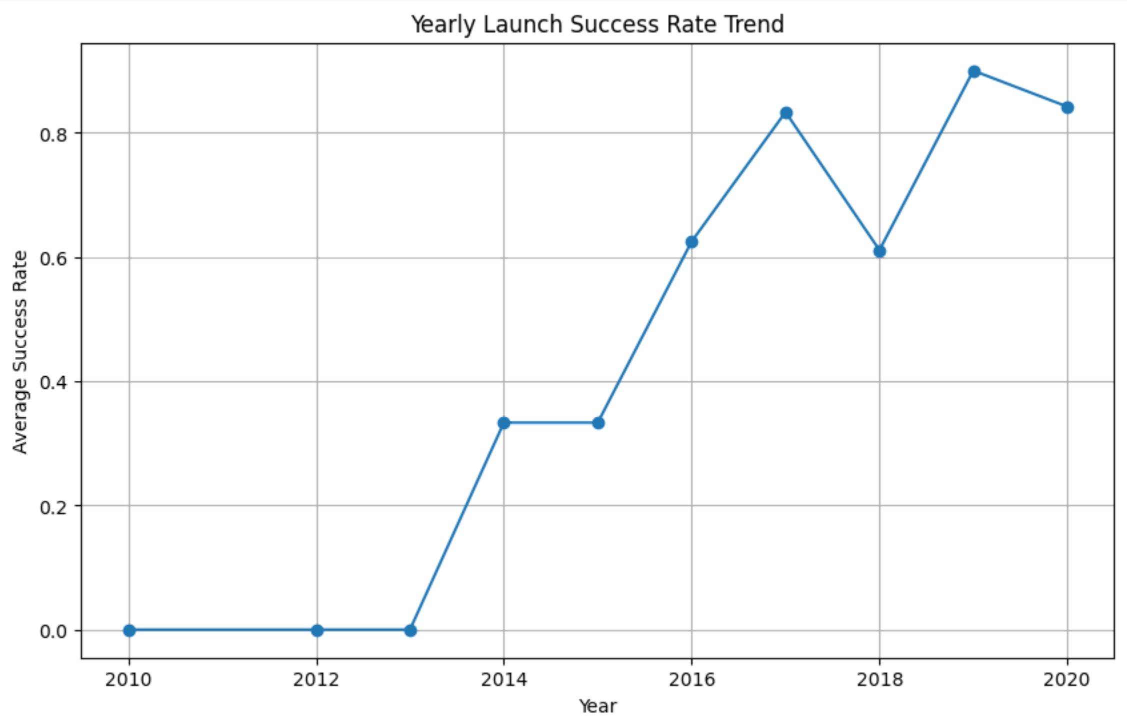
Payload vs. Orbit Type

- **Lower-energy orbits** such as **LEO and ISS** support successful landings across a wide range of payload masses.
- **GTO missions** cluster around medium payload masses and show **greater variability** in landing outcomes.
- **Very high payload masses** are primarily associated with **VLEO missions**, where most launches are successful, reflecting improved booster performance.
- **High-energy orbits** generally exhibit fewer successful landings at higher payload masses.
- Payload mass interacts with orbit type, making **orbit-specific analysis** essential.



Launch Success Yearly Trend

- Early years (2010–2013) show **very low or zero success rates**, reflecting early-stage development.
- From **2014 onwards**, the success rate increases sharply, indicating rapid technological and operational improvements.
- A noticeable dip around **2018** suggests occasional mission complexity or operational challenges.
- Peak success rates are observed in **2019 and 2020**, exceeding **80%**, demonstrating operational maturity.
- Overall, the trend shows **consistent long-term improvement** in Falcon 9 landing success.



All Launch Site Names

- The SQL query identified the following **unique SpaceX Falcon 9 launch sites**:
- **CCAFS SLC 40**
- **KSC LC 39A**
- **VAFB SLC 4E**

- The query returns **four records**, but they represent **three unique physical launch sites**.
- **“CCAFS LC-40”** and **“CCAFS SLC-40”** refer to the **same launch complex**, with naming differences arising from data source inconsistencies.
- **KSC LC-39A** and **VAFB SLC-4E** are distinct launch sites supporting different mission profiles.
- This result highlights the need for **data standardization during data wrangling** to ensure accurate site-level analysis.

Launch Site Names Begin with 'CCA'

- The SQL query filters records where the **launch site name starts with "CCA"**, corresponding to **Cape Canaveral Air Force Station**.
- The result displays **five representative Falcon 9 launches** from this location.
- These early missions primarily targeted **LEO and ISS orbits**.
- Most missions show **successful mission outcomes**, though early landings often resulted in **parachute failure or no landing attempt**.
- This confirms that **CCAFS LC-40** was the **primary launch site** during early Falcon 9 operations.

	Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
0	2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
1	2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of...	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2	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
3	2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
4	2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

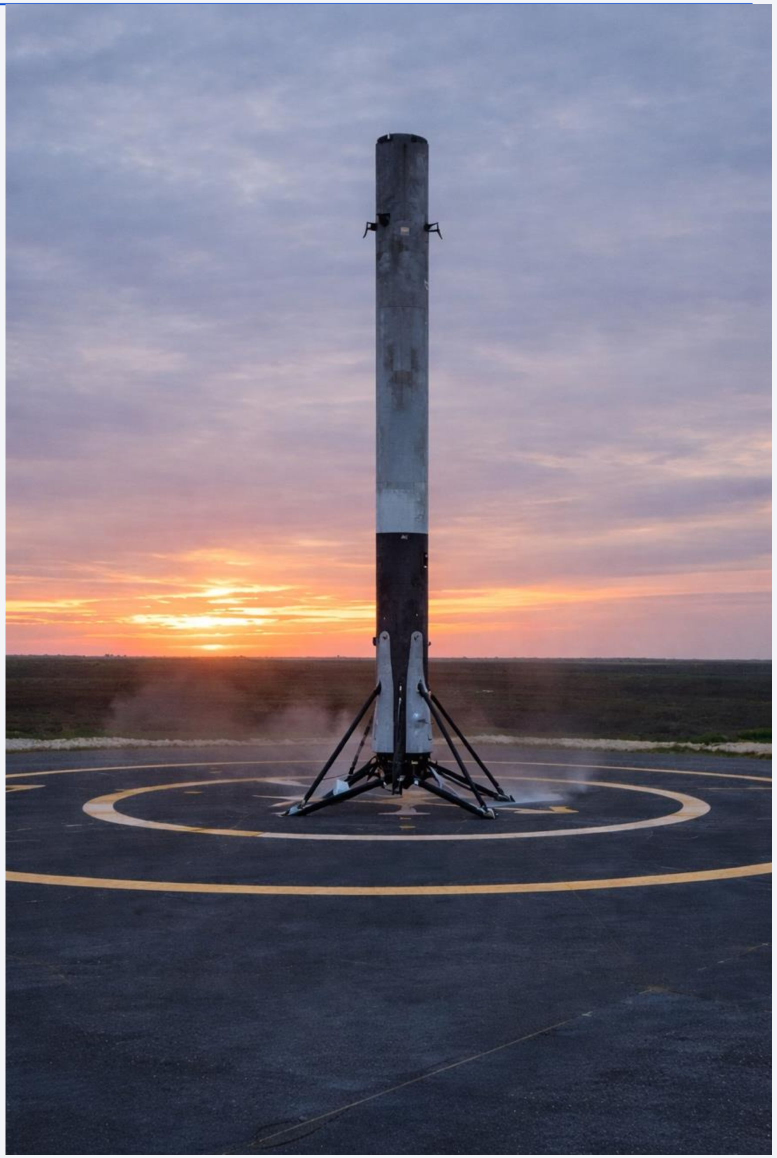
- **Total Payload Mass: 45,596 kg**
- The SQL query calculates the **sum of payload mass** for missions where the **customer is NASA**.
- This value represents the **cumulative payload capacity delivered by Falcon 9 for NASA missions** during the analyzed period.
- The result highlights SpaceX's significant role in **supporting NASA missions**, particularly to **LEO and ISS orbits**.
- These missions generally show **high mission and landing success rates**, contributing to operational reliability.

Average Payload Mass by F9 v1.1

- **Average Payload Mass: 2,928.4 kg**
- The SQL query calculates the **mean payload mass** for missions flown using **Falcon 9 booster version v1.1**.
- This average reflects the **typical payload capacity** handled during the early operational phase of Falcon 9.
- Compared to later booster versions, **F9 v1.1 carried relatively lighter payloads**, indicating progressive improvements in booster capability over time.
- This result provides a useful **baseline for comparing payload capacity across booster versions**.

First Successful Ground Landing Date

- **Date: 2015-12-22**
- The SQL query identifies the **earliest date** on which a **successful ground pad landing** was achieved.
- This milestone marks the **first fully successful recovery of a Falcon 9 first stage on land**.
- The result represents a **major technological breakthrough** in SpaceX's reusability program.
- Subsequent missions show a **significant increase in landing success rates** following this event.
-



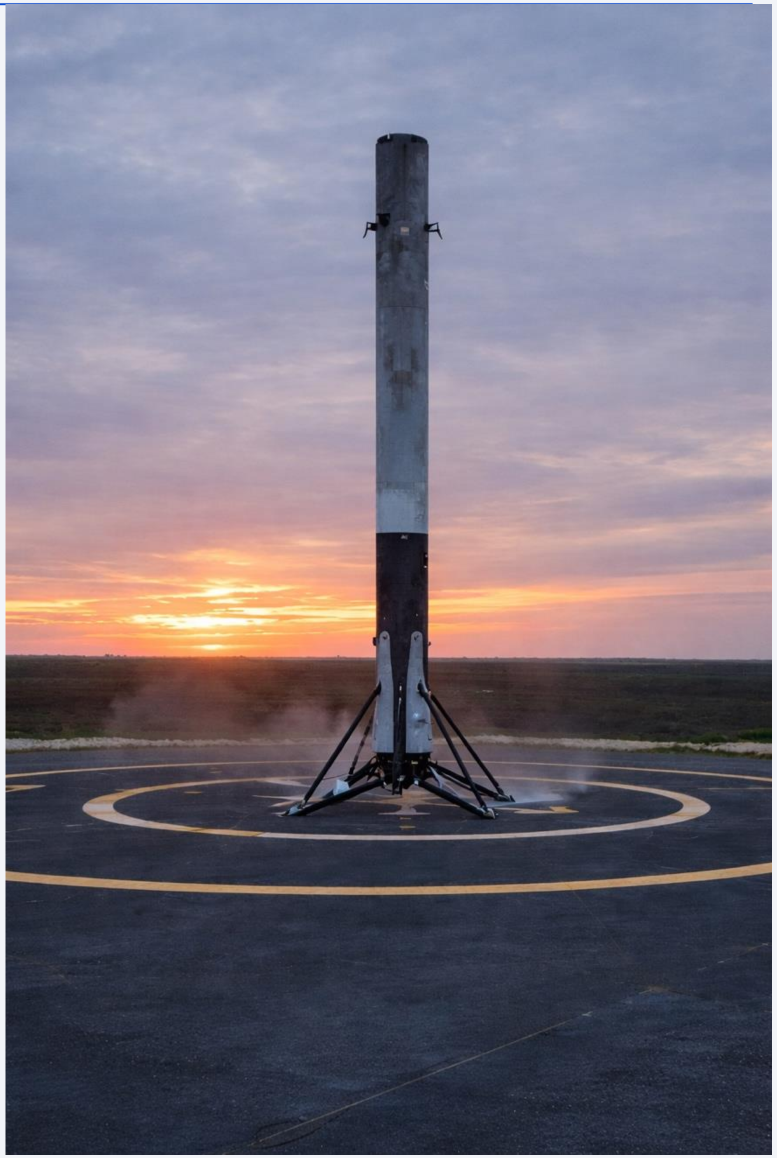
Successful Drone Ship Landing with Payload between 4000 and 6000

The following booster versions successfully landed on a **drone ship** while carrying payloads between **4000 kg and 6000 kg**:

- **F9 FT B1022**
- **F9 FT B1026**
- **F9 FT B1021.2**
- **F9 FT B1031.2**

Explanation

- The SQL query filters launches based on **successful drone ship landings** and a **payload mass range of 4000–6000 kg**.
- These results indicate that **Falcon 9 Full Thrust (FT)** boosters were capable of reliable recovery even under **moderate-to-heavy payload conditions**.
- The presence of **reused boosters** (e.g., B1021.2, B1031.2) highlights SpaceX's progress in **booster reusability and operational consistency**.
- This analysis demonstrates that **payload mass within this range does not prevent successful drone ship recovery**.
- mass greater than 4000 but less than 6000



Total Number of Successful and Failure Mission Outcomes

Success: 98

Failure (in flight): 1

Success (payload status unclear): 1

Success (with trailing space): 1

The results show that **successful missions overwhelmingly dominate** the dataset.

Minor variations such as **“Success ” (with trailing space)** and **“Success (payload status unclear)”** arise from **inconsistent labeling** in raw data.

These variants represent **successful missions** and were **standardized during data wrangling**.

Only **one true mission failure** occurred, indicating **high overall Falcon 9 mission reliability**.

This distribution confirms that Falcon 9 has achieved **operational maturity**, supporting its reusability strategy.

Boosters Carried Maximum Payload

S. No.	Booster Version
1	F9 B5 B1048.4
2	F9 B5 B1049.4
3	F9 B5 B1051.3
4	F9 B5 B1056.4
5	F9 B5 B1048.5
6	F9 B5 B1051.4
7	F9 B5 B1049.5
8	F9 B5 B1060.2
9	F9 B5 B1058.3
10	F9 B5 B1051.6
11	F9 B5 B1060.3
12	F9 B5 B1049.7

These boosters correspond to missions carrying the **maximum payload mass** in the dataset and are all **Falcon 9 Block 5 (B5)** variants, highlighting their enhanced payload capability and reusability.

2015 Launch Records

- The query lists **Falcon 9 launches in 2015** where the **landing outcome on a drone ship was a failure**.
- Results include the **booster version**, **launch site**, and **landing outcome** for each failed attempt.
- The year **2015 represents an early phase** of SpaceX's drone ship landing experiments.
- Multiple drone ship landing attempts **failed during this period**, reflecting technological and operational challenges.
- Most failures occurred before the **first successful ground pad landing in December 2015**.
- These records highlight the **learning phase** that preceded the significant improvements in landing success in later years.

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

Rank	Landing Outcome	Count
1	No attempt	10
2	Success (drone ship)	5
3	Failure (drone ship)	5
4	Success (ground pad)	3
5	Controlled (ocean)	3
6	Uncontrolled (ocean)	2
7	Failure (parachute)	2
8	Precluded (drone ship)	1

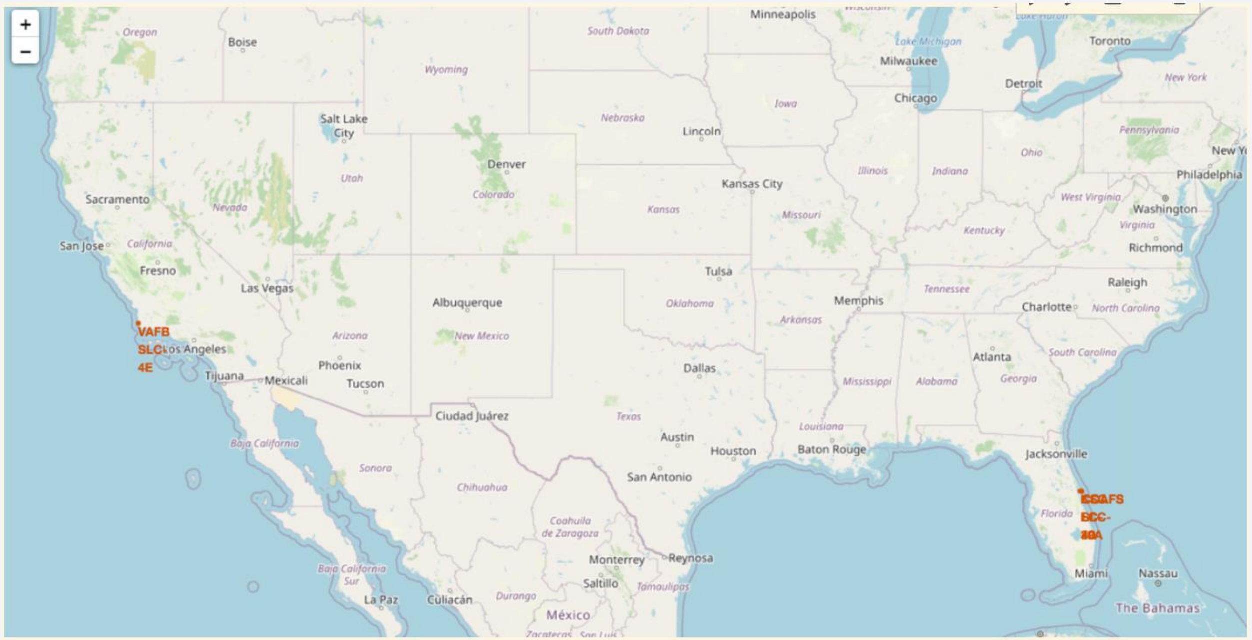
- The most frequent outcome during this period was **“No attempt”**, reflecting early missions where landing recovery was not yet pursued.
- **Drone ship landings** show an equal number of successes and failures, highlighting the **experimental phase** of autonomous sea landings.
- **Ground pad successes** begin to appear but are fewer, as land landings were introduced later.
- Ocean landings (controlled and uncontrolled) indicate **intermediate recovery strategies** used before consistent pad or drone ship landings.
- Overall, the ranking illustrates the **evolution of SpaceX’s landing strategy**, progressing from no attempts to increasingly successful recoveries.

Section 3

Launch Sites Proximities Analysis

Distribution of SpaceX Falcon 9

- The map displays **location markers** for all SpaceX Falcon 9 launch sites worldwide.
- Each marker represents a **unique launch site**, positioned using latitude and longitude coordinates.
- The global view highlights the **geographical spread** of SpaceX launch operations.
- SpaceX launch activities are concentrated in the **United States**, with sites on both the **East Coast (Florida)** and **West Coast (California)**.
- **CCAFS LC-40** and **KSC LC-39A** are located on the eastern coast, supporting missions to **LEO, ISS, and GTO**.
- **VAFB SLC-4E**, located on the western coast, primarily supports **polar and sun-synchronous orbits**.
- The spatial distribution reflects how **orbit requirements influence launch site selection**.

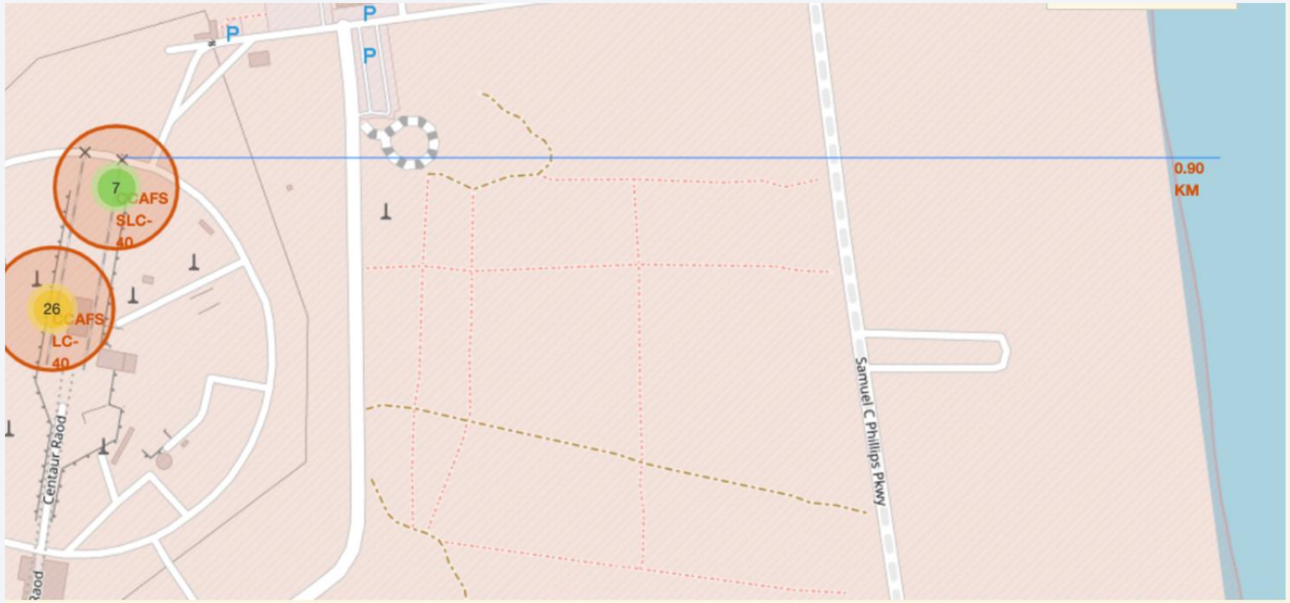


Launch Outcomes by Location

- Launch sites are shown using **color-coded markers** based on landing outcome.
- **Green markers** represent successful landings, while **red markers** indicate failed landings.
- Marker pop-ups provide **launch site details and outcome information**.
- The map allows visual comparison of outcomes across locations.
- **Key Findings**
  - Later launches show a higher concentration of **successful landings**, particularly at **CCAFS LC-40** and **KSC LC-39A**.
  - **VAFB SLC-4E** exhibits fewer launches, with outcomes reflecting mission-specific orbit constraints.
  - The spatial distribution highlights how **location and mission profile influence landing success**.
  - Color-coding makes success and failure patterns **immediately interpretable**.

Launch Site Proximity Analysis (Infrastructure & Coastline)

- The map focuses on a **selected launch site** with its surrounding infrastructure.
- **Markers** indicate the launch site and nearby features such as **railways, highways, and coastline**.
- **Lines (polylines)** are drawn from the launch site to each nearby feature.
- **Distance labels** display the calculated distance between the launch site and each feature.
- **Key Findings**
  - Launch sites are located **close to coastlines**, enabling safe flight paths over the ocean.
  - Proximity to **highways and railways** supports efficient transportation of rocket components and equipment.
  - The spatial layout reflects **logistical and safety considerations** in launch site selection.
  - Distance calculations help explain why certain locations are **operationally optimal** for frequent launches and recoveries.

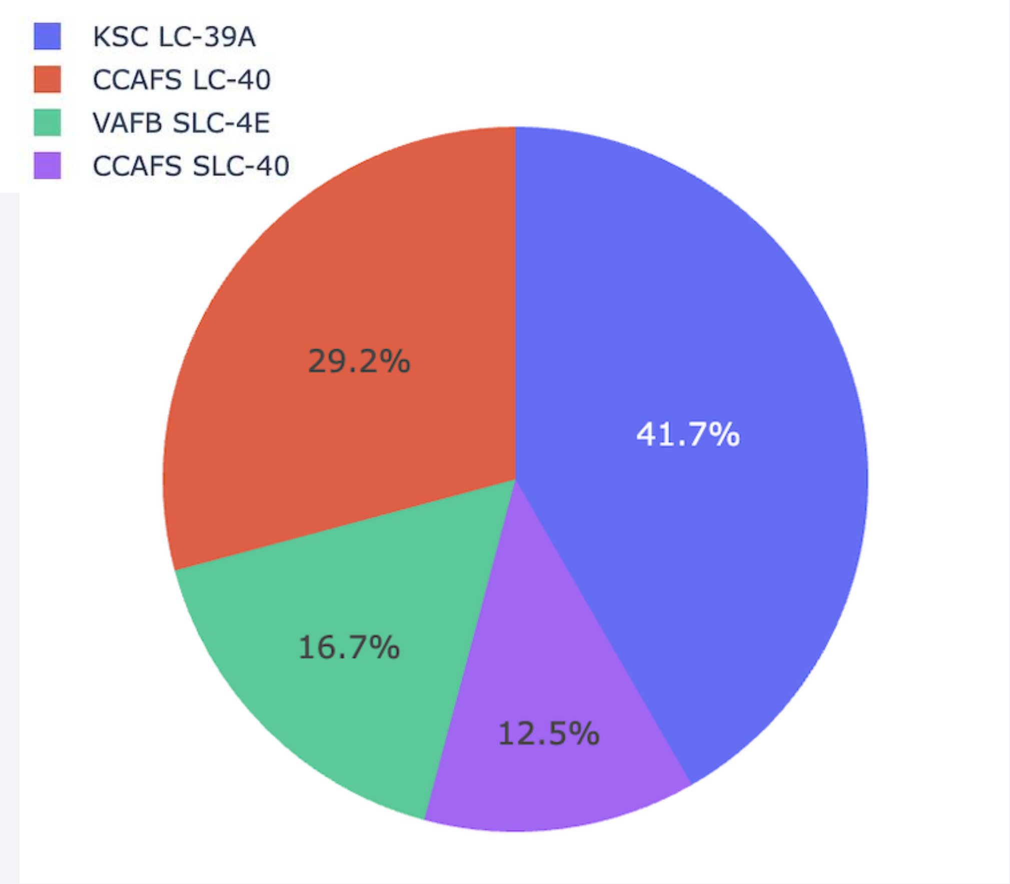


Section 4

Build a Dashboard with Plotly Dash

Launch Success Distribution Across All Launch Sites

- **Explanation of Dashboard Elements**
- The dashboard displays a **pie chart** summarizing **launch success counts** across all SpaceX launch sites.
- Each slice represents the **proportion of successful launches** contributed by a launch site.
- Interactive hover information shows **exact counts and percentages**.
- **Key Findings**
  - **CCAFS LC-40** contributes the largest share of successful launches, reflecting its role as the primary launch site.
  - **KSC LC-39A** also shows a high success contribution, especially in later missions.
  - **VAFB SLC-4E** accounts for a smaller portion due to fewer launches and specialized orbit missions.
  - The pie chart provides a **quick comparative overview** of site-wise performance.

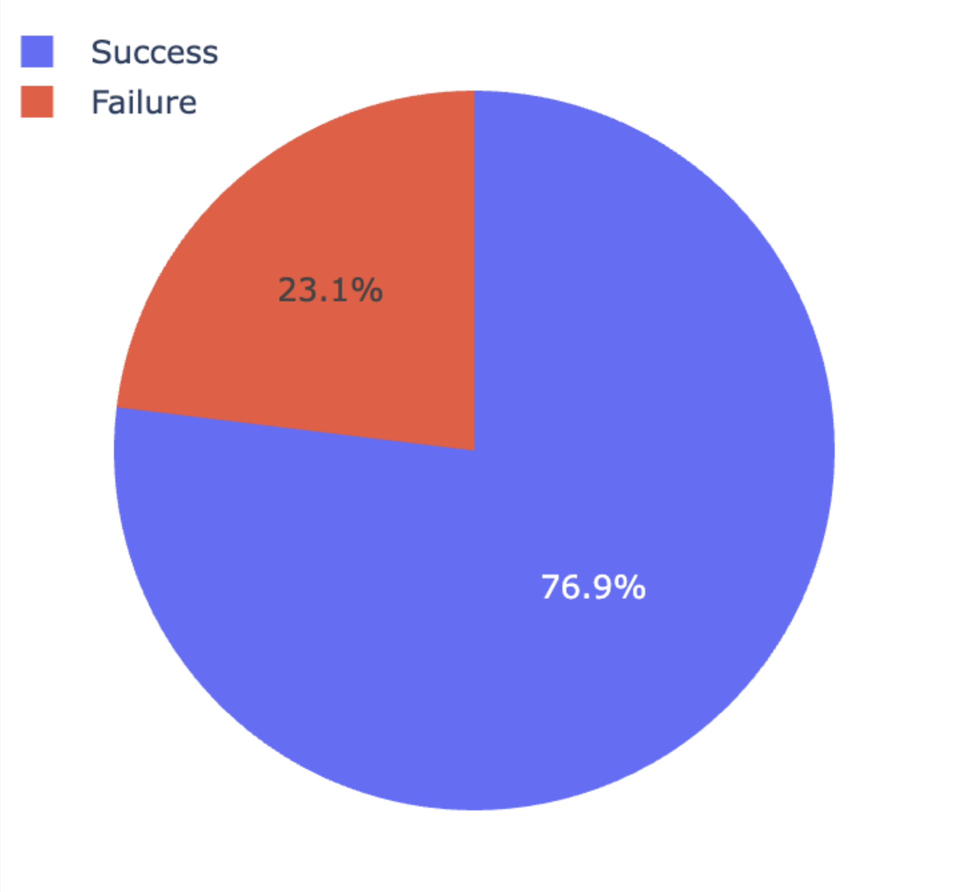


Success Rate of Lauches (at maximum successful site)

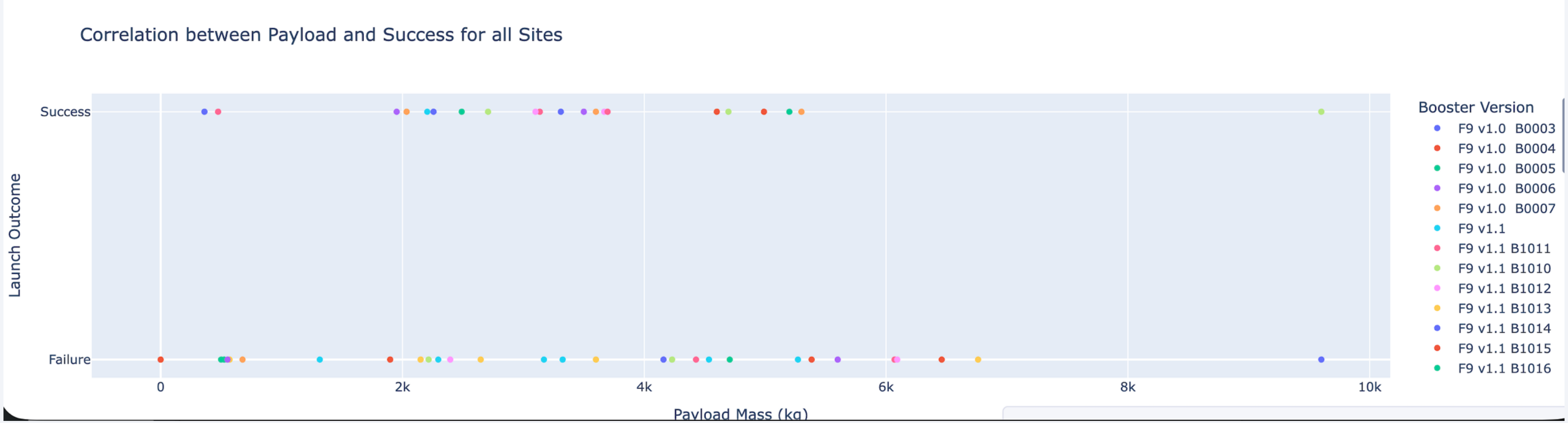
Explanation & Key Findings

- The pie chart shows the **launch success ratio for KSC LC-39A**, which has the **highest success rate among all launch sites**.
- **Successful launches account for ~76.9%**, while **failures make up ~23.1%** of total launches from this site.
- This high success proportion highlights **KSC LC-39A as SpaceX's most reliable launch location**, likely due to:
  - Advanced ground infrastructure
  - Operational maturity in later Falcon 9 missions
  - Improved booster reusability and landing precision over time

Total Success Launches for site KSC LC-39A



Payload Mass vs. Launch Outcome Across All Sites



Explanation

- The dashboard displays an **interactive scatter plot** of **payload mass vs. launch outcome**.
- A **range slider** allows dynamic filtering of payload mass.
- Points are color-coded by **launch success and failure**.
- Data from **all launch sites** are displayed simultaneously for comparison.

Key Findings

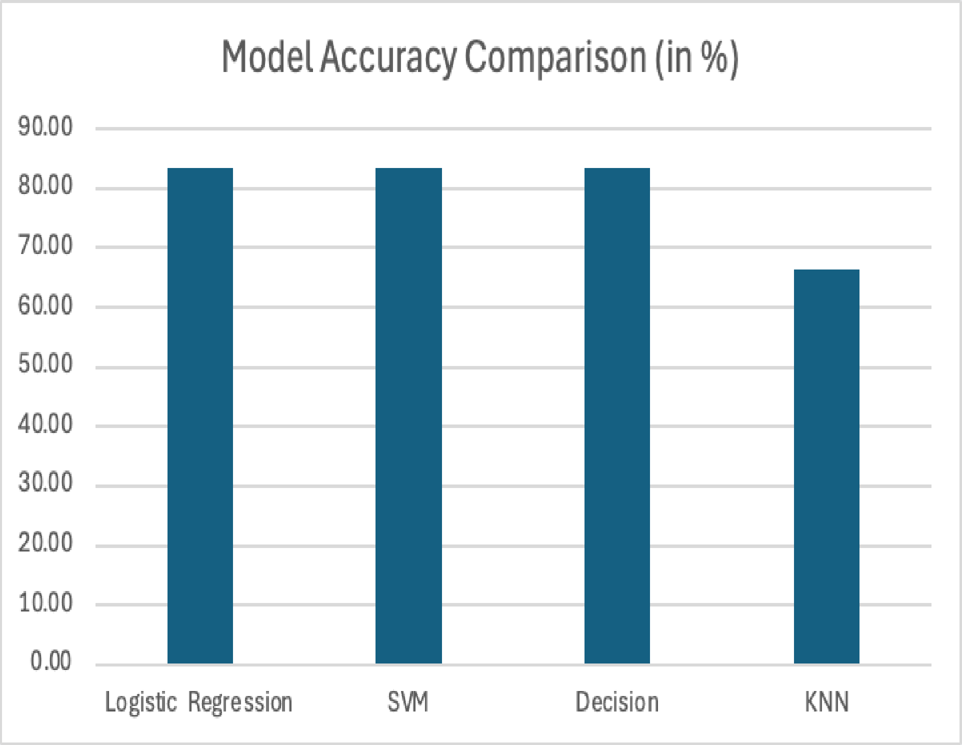
- **Lower to medium payload ranges** show a higher concentration of successful launches.
- At **very high payload masses**, success rates remain high for newer booster versions.
- **Falcon 9 Block 5 boosters** demonstrate the strongest performance across most payload ranges.
- Filtering by payload range reveals how **mission complexity increases with payload mass**, impacting outcomes.

Section 5

Predictive Analysis (Classification)

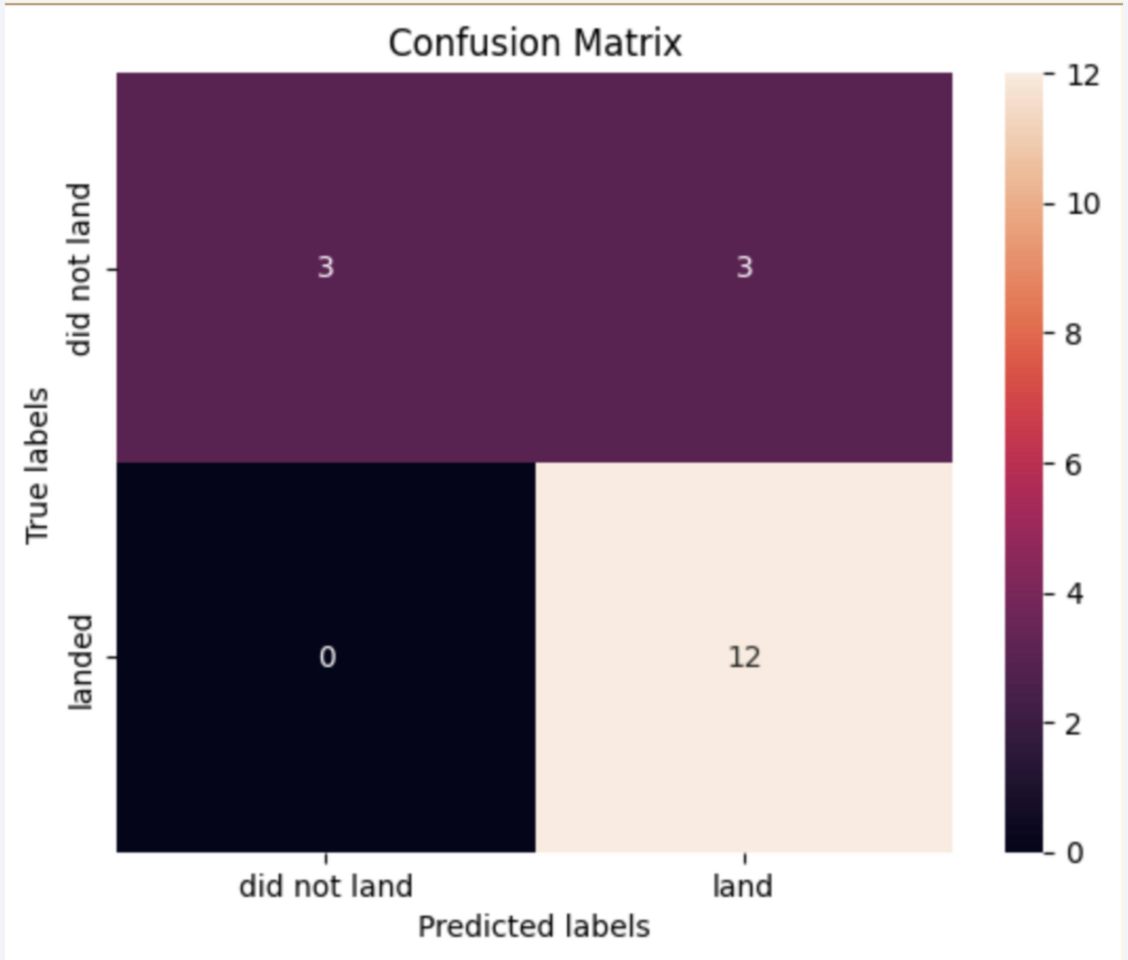
Classification Accuracy

- **Logistic Regression, Decision Tree and SVM** achieved the **same classification accuracy (~83%)**.
- Overall, all models performed well, indicating that landing success is **predictable from mission features**.



Confusion Matrix

- **Model Used: Logistic Regression**
- **Explanation**
  - **True Positives (TP):** Correctly predicted successful landings.
  - **True Negatives (TN):** Correctly predicted unsuccessful landings.
  - **False Positives (FP):** Predicted success when the landing failed.
  - **False Negatives (FN):** Predicted failure when the landing was successful.



Conclusions

- Falcon 9 landing success has **improved significantly over time**, driven by accumulated flight experience and technological refinements.
- **Payload mass, orbit type, launch site, and flight number** are key factors influencing landing outcomes.
- Exploratory and geospatial analyses revealed clear **spatial and operational patterns** across launch sites.
- Interactive dashboards and maps enabled **deeper insight and dynamic exploration** beyond static analysis.
- **Decision Tree and SVM models** achieved the best predictive performance (~83% accuracy), demonstrating that landing success can be **reliably predicted using historical data**.

Thank you!