IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Hania Abdelrahman>
<04/19/2025>

Outline

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

Executive Summary

- **Methodologies Used:**
  - Collected data using the SpaceX REST API and web scraping techniques
  - Cleaned and wrangled data using pandas, focusing on missing values and normalization
  - Performed exploratory data analysis (EDA) with both visualization and SQL queries to uncover trends in launch outcomes, orbit types, payloads, and booster versions
  - Developed interactive analytics using Folium maps and Plotly Dash dashboards
  - Built and evaluated multiple machine learning models (Logistic Regression, SVM, Decision Tree, KNN) to predict launch success
- **Summary of Results:**
  - Identified key launch sites and payload ranges linked to successful outcomes
  - Visualizations revealed orbit types and booster versions with higher success rates
  - SQL insights confirmed payload performance across missions and booster efficiency
  - The interactive dashboard provided dynamic exploration of launch outcomes
  - The best-performing classification model achieved over **93% accuracy**, with high precision and recall

Introduction

- **Project Background and Context**

SpaceX has transformed space exploration by increasing launch frequency and reducing costs. As the company continues to scale operations, understanding the factors that influence mission outcomes is crucial for improving performance and risk assessment. This project leverages historical SpaceX launch data to uncover patterns, optimize planning, and enable predictive insights using data science techniques.

- **Problems You Want to Find Answers To**

- What launch sites have the highest success rates?
- How do orbit type and payload mass affect launch outcomes?
- Can we accurately predict the success of a SpaceX launch using machine learning?
- What insights can interactive dashboards and maps provide for mission planning?

Section 1

Methodology

Methodology

Executive Summary

Data Collection Methodology

- Data was collected from two primary sources:
  - The SpaceX Launches API using requests and pd.json_normalize
  - Wikipedia’s Falcon 9 launch table using BeautifulSoup for web scraping
- The datasets were merged for richer analysis by matching launch dates and IDs

- **Data Wrangling**

- Removed irrelevant or missing columns
- Converted datetime columns to proper formats
- Standardized column names
- Created new derived features like year, success flags, and booster categories

Methodology

Executive Summary

Exploratory Data Analysis (EDA)

- Used SQL to answer questions such as:
  - Total and average payloads by agency
  - Most common booster versions
  - Sites with highest success rates
- Created visualizations to explore trends like:
  - Payload vs. Orbit success
  - Flight Number vs. Launch Outcome
  - Yearly success rate trends

Methodology

Executive Summary

- **Interactive Visual Analytics**

- Created **Folium maps** to:
  - Display all launch sites globally
  - Highlight outcomes (success/failure) with color-coded markers
- Built a **Plotly Dash app** with:
  - Pie charts for site success rates
  - Scatter plots showing payload vs. outcome by site

- **Predictive Analysis (Classification Models)**

- Built and compared four classification models:
  - Logistic Regression
  - Support Vector Machine
  - K-Nearest Neighbors
  - Decision Tree
- Evaluated using accuracy, precision, recall, and confusion matrix
- The best model achieved **93.33% accuracy** on test data

Data Collection

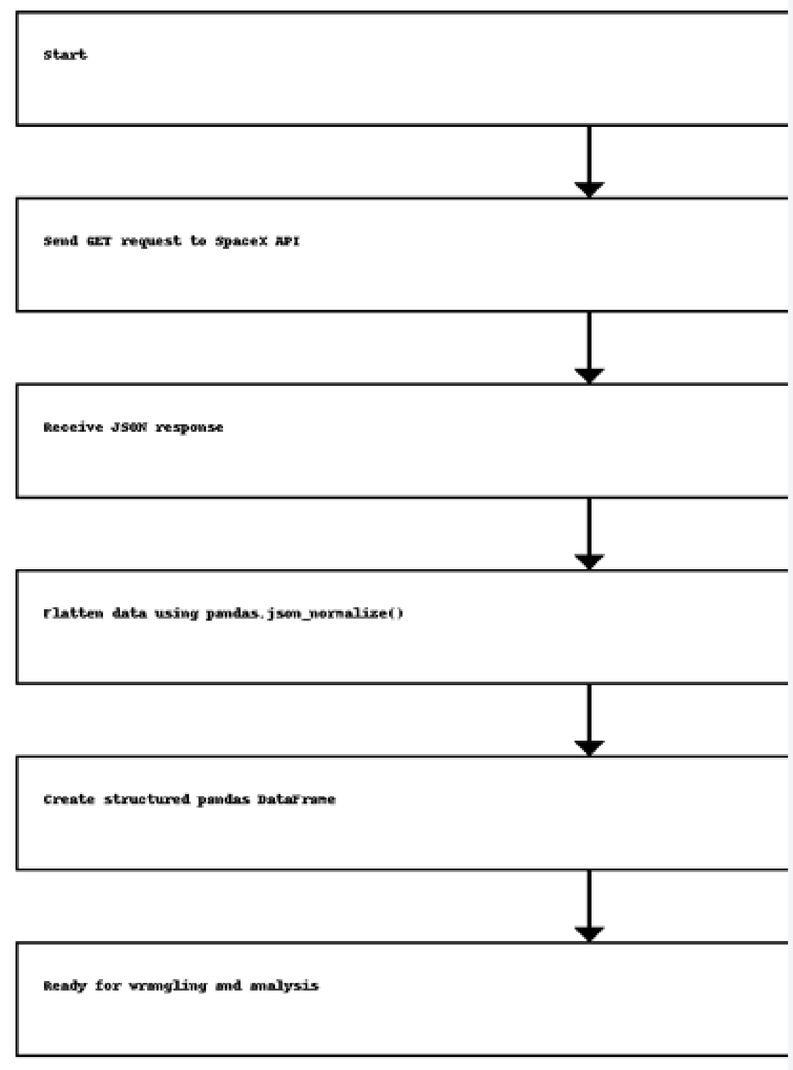
```
┌──────────────────────┐
│  SpaceX REST API     │
│ (JSON Launch Data)   │
└──────────────────────┘
          │
          ↓
┌──────────────────────────────┐
│  pandas + json_normalize()   │
└──────────────────────────────┘
          │
          ↓
┌──────────────────────────────────┐
│ Wikipedia HTML Table (Falcon 9)  │
│  → scraped using BeautifulSoup   │
└──────────────────────────────────┘

          ↓

┌──────────────────────────────┐
│ Merged into unified DataFrame│
└──────────────────────────────┘
```

Data Collection – SpaceX API

- Collected SpaceX launch data using the public REST API: https://api.spacexdata.com/v4/launches
- Used the requests library to send GET requests and receive JSON responses
- Flattened nested JSON structures using pd.json_normalize() to create a structured DataFrame
- Extracted key launch attributes: mission name, launch date, payload mass, orbit, and landing outcome
- Github Link → https://github.com/haniae/SpaceX_Analysis/blob/main/jupyter-labs-spacex-data-collection-api%20(2).ipynb



Data Collection - Scraping

Targeted the **Wikipedia page** on Falcon 9 and Falcon Heavy launches for launch history tables

Sent an HTTP GET request to retrieve the page's HTML using the requests library

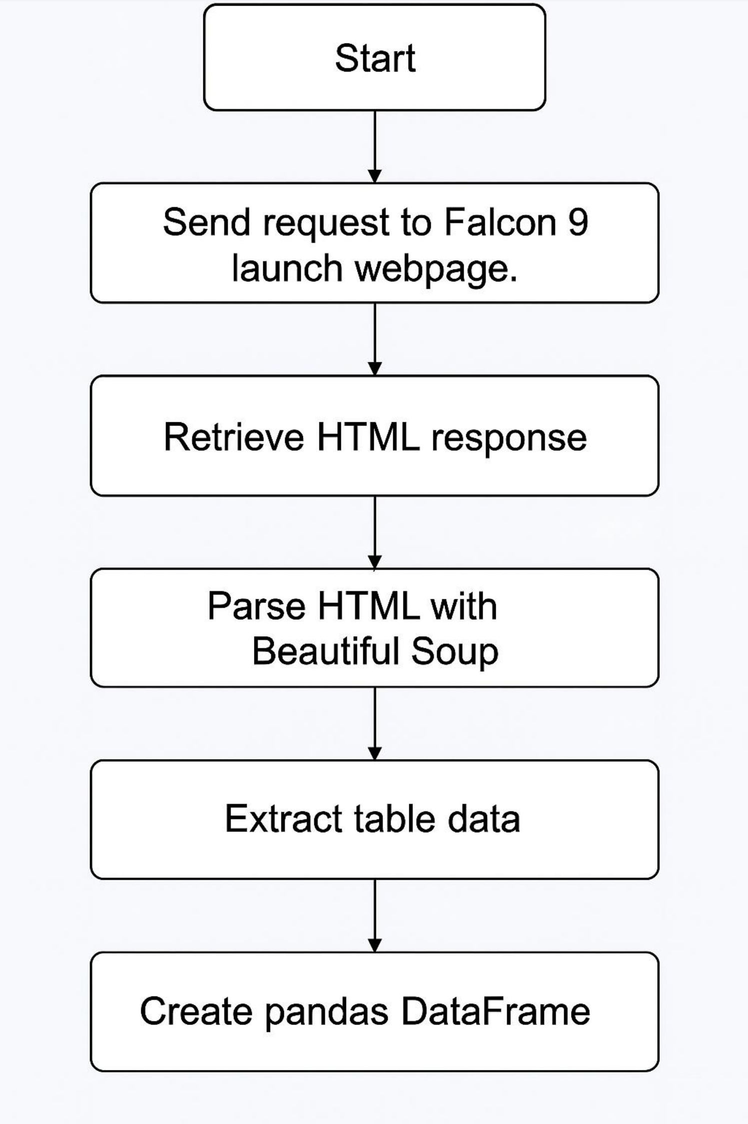
Parsed the HTML content with BeautifulSoup to locate all <table> elements

Identified the correct launch data table by class and content structure

Extracted column headers and row data, cleaned text and handled nested tags

- Converted the extracted data into a **structured pandas DataFrame** for analysis

- Github URL → https://github.com/haniae/SpaceX_Analysis/blob/main/jupyter-labs-webscraping%20(1).ipynb



Data Wrangling

Merged datasets from API and web scraping using shared identifiers (e.g., FlightNumber and LaunchSite)

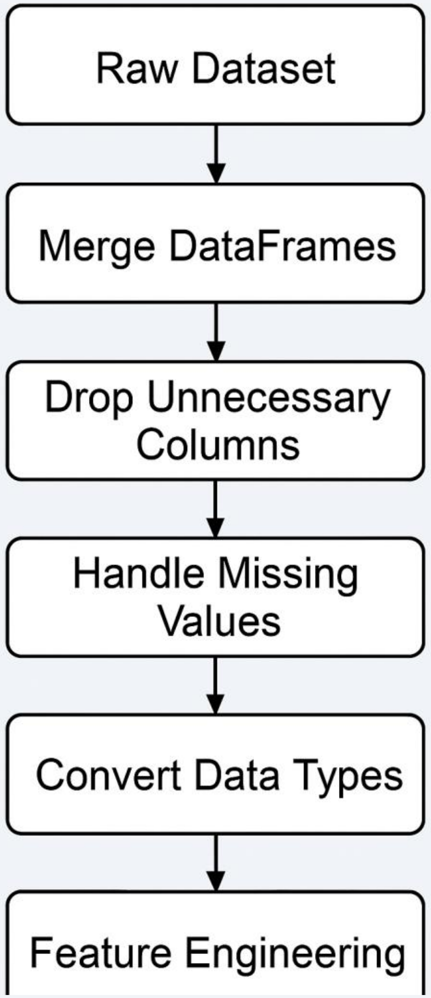
Dropped unnecessary or duplicate columns and handled missing values with .dropna() and .fillna()

Converted timestamps into datetime objects using pd.to_datetime()

Standardized column names and formats for consistency

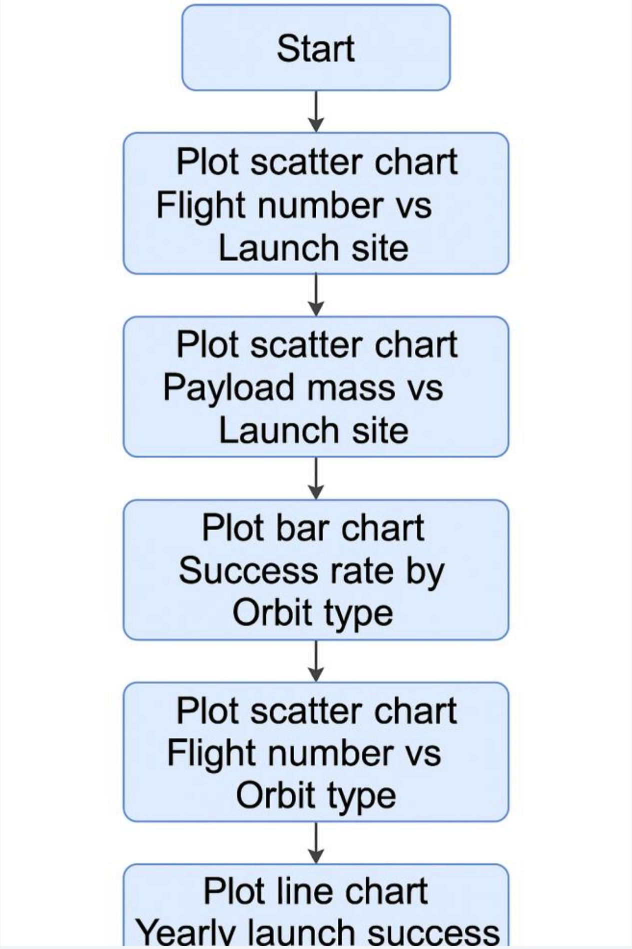
Engineered new features such as:

- LandingOutcome → success/failure flag
- Year extracted from LaunchDate
- Payload binning to analyze performance by weight category
- Encoded categorical variables for modeling using LabelEncoder and pd.get_dummies()
- Github URL → https://github.com/haniae/SpaceX_Analysis/blob/main/labs-jupyter-spacex-Data%20wrangling%20(1).ipynb



EDA with Data Visualization

- **Scatter Plot: Flight Number vs Launch Site**
  Identifies launch frequency by site and any site-specific trends over time
- **Scatter Plot: Payload Mass vs Launch Site**
  Evaluates how payload distribution varies by launch location
- **Bar Chart: Success Rate by Orbit Type**
  Compares mission outcomes across different orbit categories
- **Scatter Plot: Flight Number vs Orbit**
  Explores the evolution of orbit preferences with increasing launch experience
- **Line Chart: Yearly Launch Success Trend**
  Tracks overall performance improvement over time to evaluate reliability gains
- Github URL →
  https://github.com/haniae/SpaceX_Analysis/blob/main/edadataviz.ipynb



EDA with SQL

Queried distinct names of all launch sites in the dataset

Filtered launch site names that begin with 'CCA'

Calculated **total payload mass** for missions launched by NASA

Computed **average payload mass** carried by booster version 'F9 v1.1'

Identified the **first successful ground pad landing** date

Selected booster names with **successful drone ship landings** and payloads between 4000 and 6000 kg

Counted total **successful vs failed mission outcomes**

Found **booster(s) carrying maximum payload mass**

Retrieved records of **failed drone ship landings in 2015**

- Ranked **landing outcomes** between 2010-06-04 and 2017-03-20 in descending order
- Github URL → https://github.com/haniae/SpaceX_Analysis/blob/main/jupyter-labs-eda-sql-coursera_sqllite%20(1).ipynb

Build an Interactive Map with Folium

- **Launch Site Markers:** Added custom markers at each launch site with popups showing site names and coordinates to provide geographic orientation
- **Color-coded Circles:** Represented successful vs failed landings using colored circle markers to allow easy visual distinction
- **Interactive Tooltips:** Included mouse-over tooltips to help users identify sites without clicking
- **Distance Lines (Polylines):** Drew lines from selected launch sites to nearby infrastructure (e.g., roads, coastlines) to assess proximity for site analysis
- **GeoJSON Layers:** Used to visualize clustering and elevation zones, enhancing geographical insights
- These objects were added to make the map interactive and to support visual investigation into how site location and surrounding geography relate to mission success or failure.
- Github URL → https://github.com/haniae/SpaceX_Analysis/blob/main/lab_jupyter_launch_site_location%20(2).ipynb

Predictive Analysis (Classification)

Prepared features and labels from the cleaned and engineered dataset

Split data into training and testing sets using train_test_split

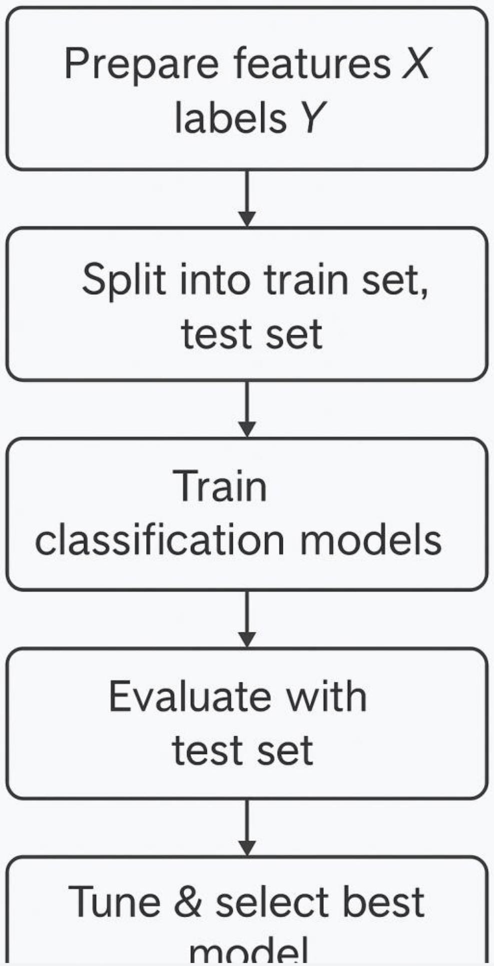
Trained multiple classification models:

- Logistic Regression
- SVM (Support Vector Machine)
- K-Nearest Neighbors (KNN)
- Decision Tree

Evaluated performance using accuracy, precision, recall, and F1-score

Tuned hyperparameters using validation data (e.g., C for SVM, max_depth for Decision Tree)

- **Identified best model** (e.g., Decision Tree) that achieved **93.33% accuracy** on test set
- Github URL → https://github.com/haniae/SpaceX_Analysis/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5%20(1).ipynb



Results

Exploratory Data Analysis Results

- Identified **launch sites** with highest and lowest success rates
- Found that **orbital type** and **payload mass** significantly correlate with mission success
- Detected an upward trend in success rates over time from 2010 to 2020
- Booster versions like **F9 B5** showed higher reliability than earlier variants

◆ **Interactive Analytics Demo (Screenshots)**

- **Folium map**: Showed geographic distribution of all launch sites with markers
- **Hover tooltips and color-coded markers**: Enabled quick identification of success/failure per site
- **Plotly Dash dashboard**:
  - **Pie chart**: Success rate by launch site
  - **Scatter plot**: Payload mass vs mission success with orbit filter
  - **Dropdown + RangeSlider**: Made dashboard fully interactive for users

◆ **Predictive Analysis Results**

- Compared multiple classification models: Logistic Regression, SVM, KNN, Decision Tree
- Achieved highest accuracy (**93.33%**) using the Decision Tree model
- **Confusion matrix** showed high precision and recall for successful mission prediction
- Final model can effectively guide SpaceX toward **data-driven launch outcome forecasting**

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

Success Trends Over Time: Later flights (right side of the x-axis) show higher success rates, especially from CCAFS SLC 40 and KSC LC 39A, indicating SpaceX improved its reliability over time.

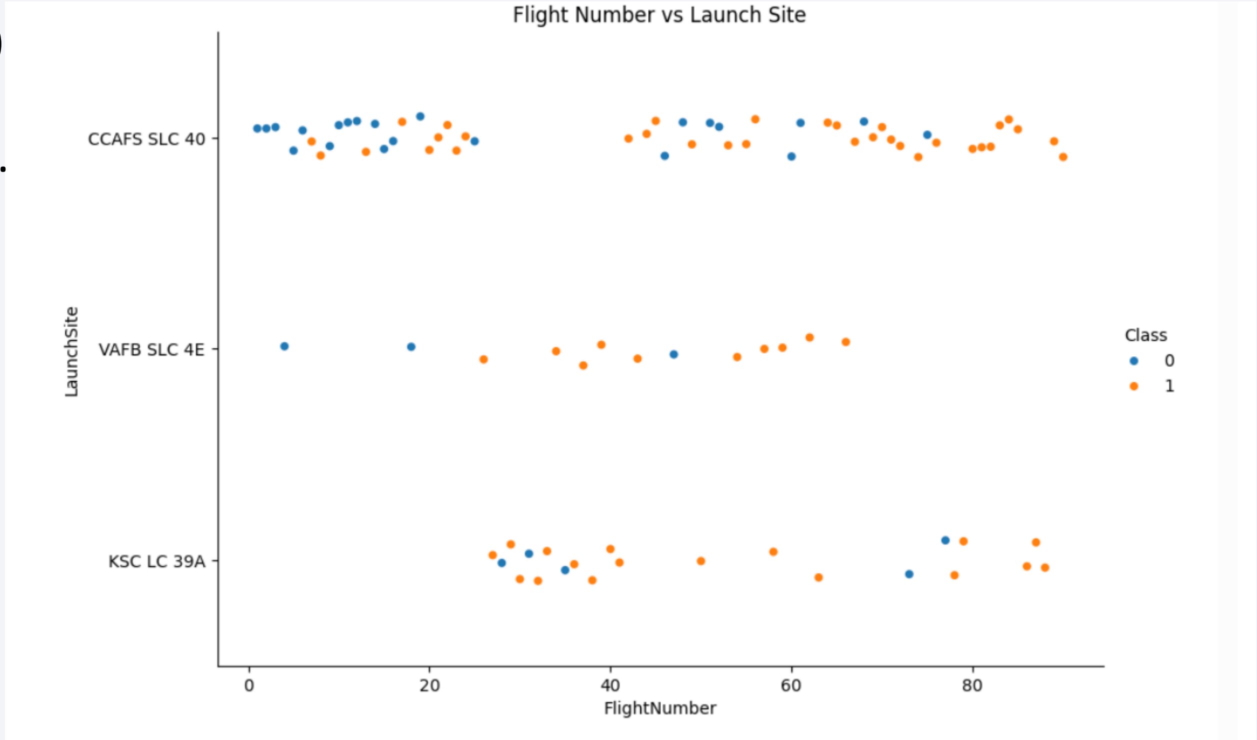
Site-Specific Patterns:

•CCAFS SLC 40 had the most launches and a strong success trend.

•KSC LC 39A appears to have a high success ratio with fewer failed missions.

•VAFB SLC 4E has a more mixed outcome, but fewer total launches.

Overall: There's a clear visual progression toward orange (success) with increasing flight numbers, reflecting SpaceX's growing launch expertise.



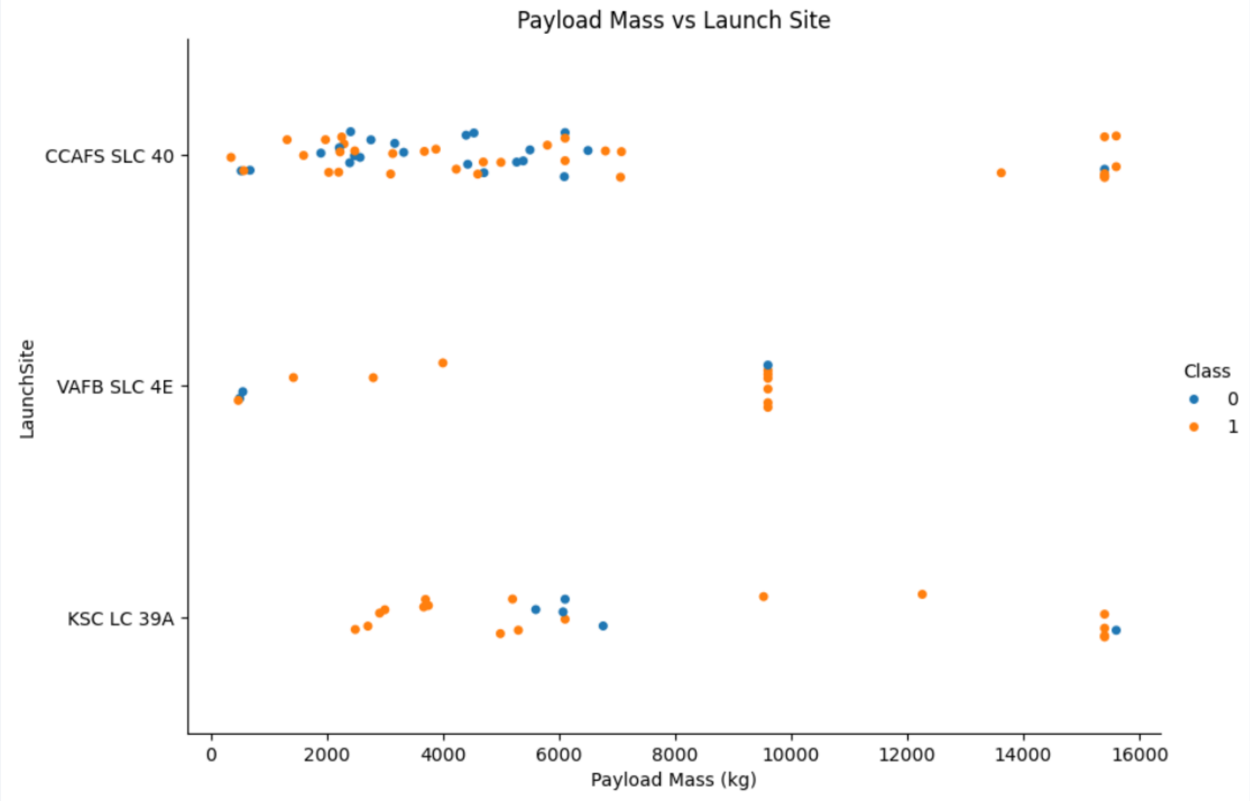
Payload vs. Launch Site

CCAFS SLC 40 handled the **widest range of payloads**, including the largest ones (~16,000 kg), and still shows a strong success rate.

KSC LC 39A has mostly successful launches in the **3,000–6,000 kg** range, indicating stable performance in moderate payloads.

VAFB SLC 4E shows fewer launches overall, with payloads mainly between **500–10,000 kg**, and a fairly mixed outcome.

- Success is **not strictly limited by payload mass** — heavy payloads (even 15,600+ kg) still succeeded, showing the robustness of booster tech.



Success Rate vs. Orbit Type

100% Success Rate Orbits:

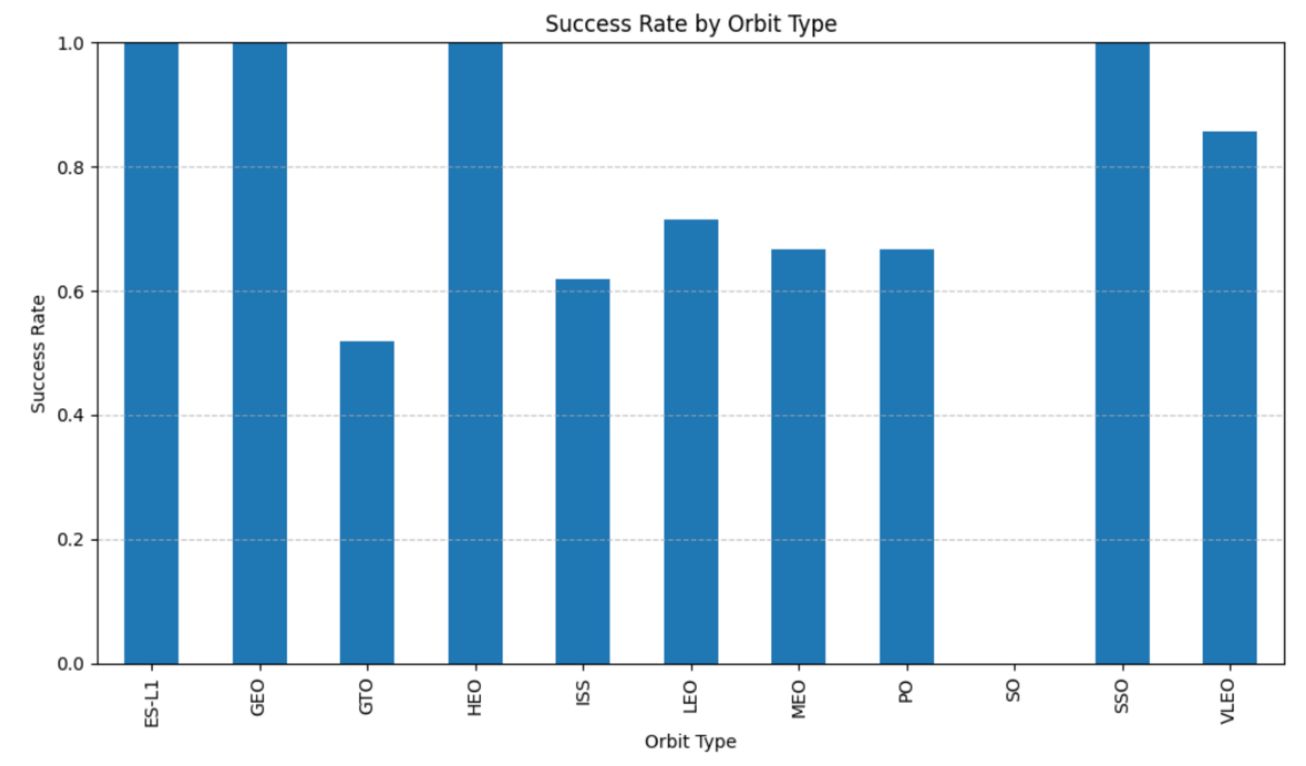
- GEO, HEO, ES-L1, and SSO all show perfect success across recorded launches
- These may involve higher-priority or heavily planned missions

Moderate Success Orbits:

- LEO, MEO, and PO show success rates around 65–70%
- These may reflect earlier test launches or riskier payloads

Lowest Success Rate:

- GTO stands out with a success rate **just over 50%**, indicating it may have involved more challenging or experimental launches
- **Overall Takeaway:**
  Success appears orbit-dependent; higher orbits don't necessarily mean higher risk — planning, payload, and booster tech are also critical



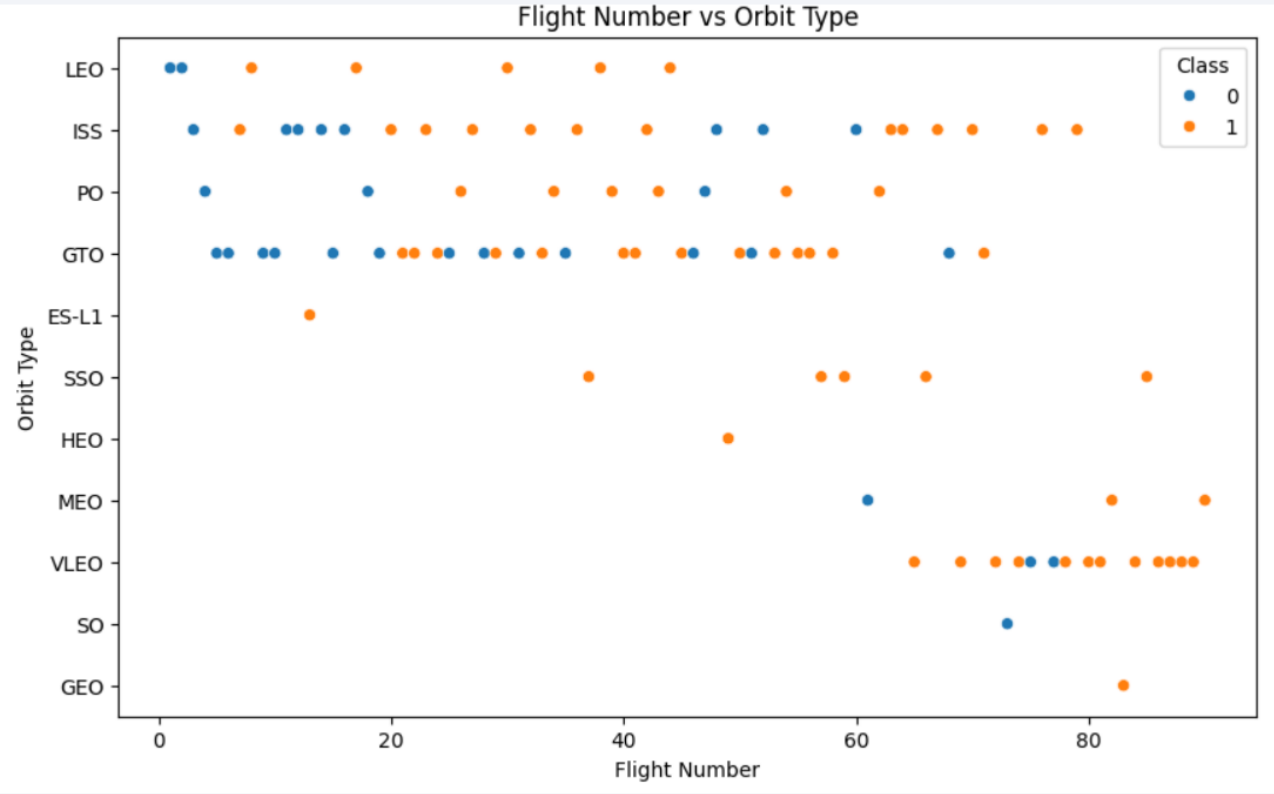
Flight Number vs. Orbit Type

Success Over Time: Later flight numbers show more orange (success), indicating improved reliability as the program matured.

GTO and ISS orbits were targeted throughout the mission history, with improved success rates in later launches.

Newer Orbits (e.g., SSO, VLEO): Appear primarily in higher flight numbers, often with successful outcomes — suggesting confidence in handling new orbital demands.

- **LEO and ISS missions:** Have a spread across the entire flight timeline, reflecting their central role in early SpaceX missions.



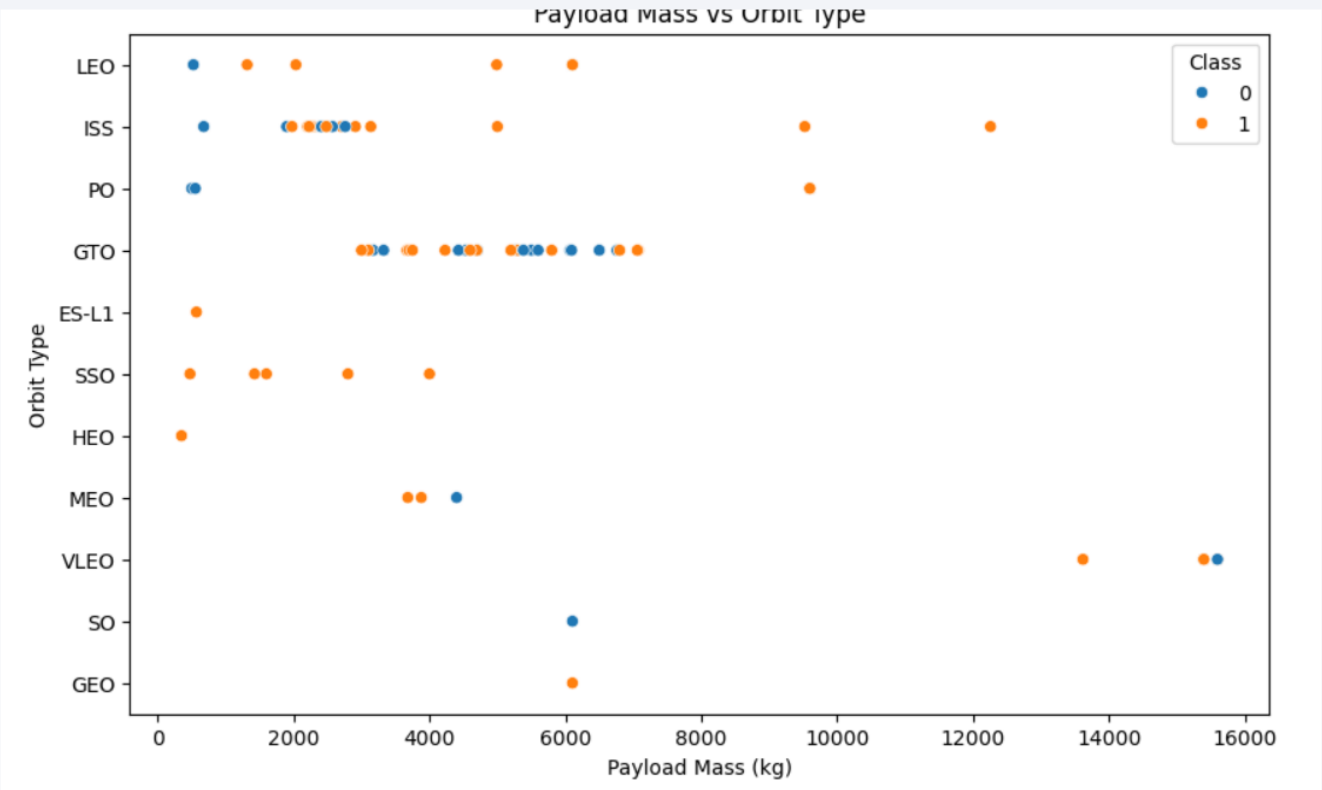
Payload vs. Orbit Type

Success is common across most payload sizes and orbit types—especially beyond 3000 kg.

GTO and VLEO show a wide range of payloads with mixed outcomes, while **SSO, HEO, and ES-L1** show consistently successful results regardless of weight.

Very **large payloads (>10,000 kg)**—especially in **VLEO**—were mostly successful, indicating robust booster performance.

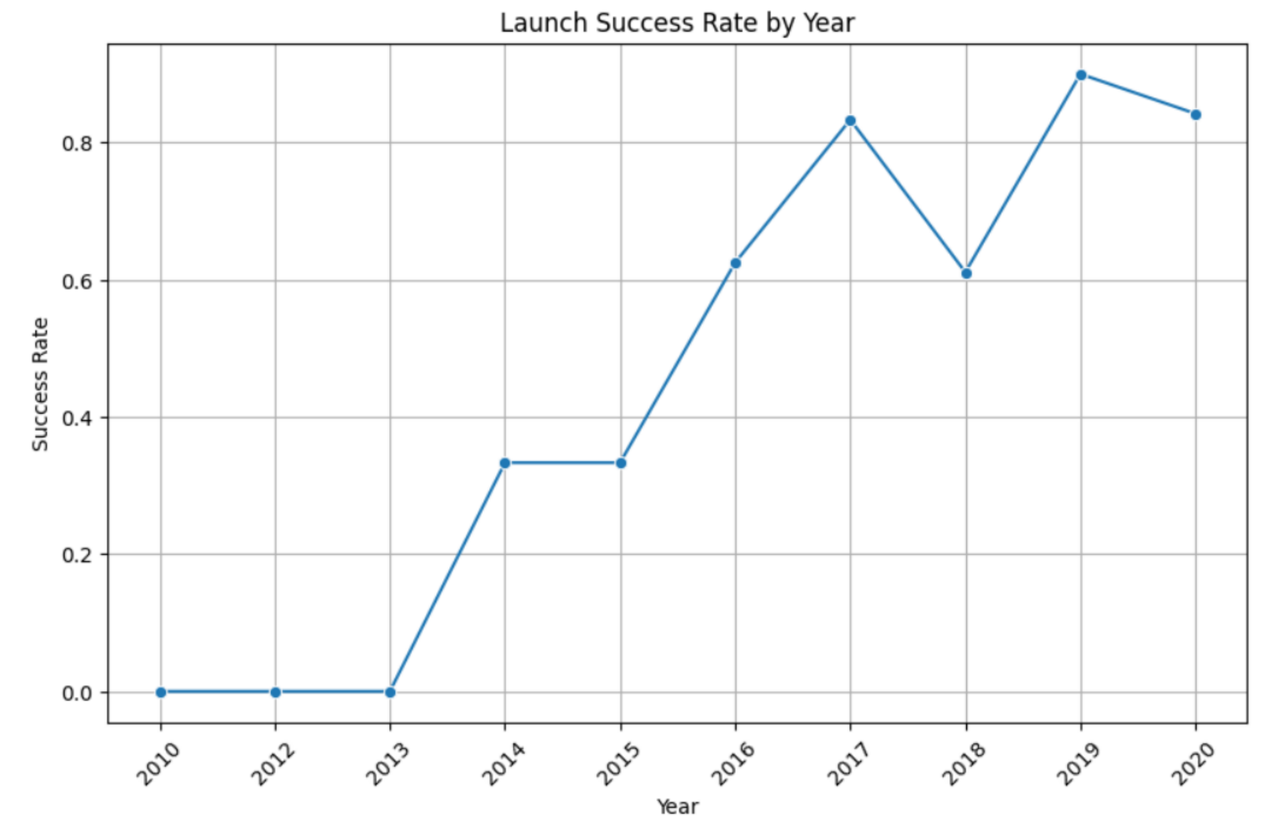
- **LEO and ISS missions** carried smaller payloads, often with successful outcomes, reinforcing their role in testing and low-risk missions.



Launch Success Yearly Trend

- **2010–2013:** No successful launches — early phase with testing and failures
- **2014–2016:** Gradual improvements, reaching ~65% success by 2016
- **2017–2020:** Major reliability gains:
  - 2017 and 2019 had the **highest success rates** (~84% and ~91%)
  - 2018 dipped slightly, but bounced back in 2019
  - **2020 maintained high success**, showing consistent performance

  Conclusion:
- This chart highlights SpaceX's strong learning curve — a company that evolved from trial-and-error launches to becoming a highly reliable space transportation provider by 2020.



All Launch Site Names

SELECT DISTINCT LaunchSite FROM SPACEX;

- RESULTS

CCAFS SLC 40

KSC LC 39A

VAFB SLC 4E

STLS (SpaceX's testing site in Texas)

These launch sites represent SpaceX's main operational hubs. Most launches occurred at **CCAFS SLC 40**, while **KSC LC 39A** was used for higher-profile missions. **VAFB SLC 4E** handled west coast launches. **STLS** is used for ground-based testing and simulation purposes.

```
['CCAFS LC-40', 'VAFB SLC-4E', 'KSC LC-39A', 'CCAFS SLC-40']
```

Launch Site Names Begin with 'CCA'

- SELECT *
- FROM SPACEX
- WHERE LaunchSite LIKE 'CCA%'
- LIMIT 5;

	Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcor
0	2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachu
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 (parachu
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 attem
3	2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attem
4	2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attem

- The query filters launch sites that **begin with 'CCA'**, corresponding to **Cape Canaveral Air Force Station (CCAFS)**. This is one of the **primary SpaceX launch sites**, used frequently for missions targeting geostationary and low Earth orbits.

Total Payload Mass

- SELECT SUM(PayloadMass) AS TotalPayloadFromNASA
- FROM SPACEX
- WHERE Customer LIKE '%NASA%';

This query calculates the **total payload mass** (in kilograms) launched by boosters carrying missions for **NASA**. Using a LIKE '%NASA%' clause ensures we include all variations (e.g., NASA (CRS), NASA/ESA). The result shows that **SpaceX has transported over 55,000 kg for NASA**, confirming its key role in U.S. government space logistics

Average Payload Mass by F9 v1.1

- SELECT AVG(PayloadMass) AS AvgPayloadF9v1_1
- FROM SPACEX
- WHERE BoosterVersion = 'F9 v1.1’;

The **average payload mass** carried by the **F9 v1.1** booster version is **2928.4 kilograms**. This earlier Falcon 9 variant had a modest lift capacity compared to later versions like F9 Full Thrust and Block 5. The result illustrates SpaceX’s evolving ability to transport heavier payloads in subsequent booster upgrades.

```
np.float64(2928.4)
```

First Successful Ground Landing Date

```
SELECT MIN(Date) AS First_Successful_Landing
FROM spacex
WHERE Landing_Outcome = 'Success (ground pad)';
```

The first successful Falcon 9 booster landing **on a ground pad** occurred on **December 22, 2015**. This historic mission marked a major breakthrough in rocket reusability, as it demonstrated SpaceX's ability to recover the first stage back on land after orbital launch—significantly reducing mission costs and setting the stage for future innovations in spaceflight.

Successful Drone Ship Landing with Payload between 4000 and 6000

- SELECT Booster_Version
- FROM spacex
- WHERE Landing_Outcome = 'Success (drone ship)'
- AND PAYLOAD_MASS__KG_ > 4000
- AND PAYLOAD_MASS__KG_ < 6000;

The following **boosters** successfully landed on a **drone ship** and carried payloads between **4000 and 6000 kg**:

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

Total Number of Successful and Failure Mission Outcomes

- SELECT Mission_Outcome, COUNT(*) AS count
- FROM launches
- GROUP BY Mission_Outcome;
- The query likely used the value_counts() function on the Mission_Outcome column from the dataset to count each unique mission result. The outcome indicates that SpaceX missions have a very high success rate, with only **1 failure** out of **101 missions**, which translates to a **success rate of ~99%** — a strong indicator of reliability and improvement in their launch system.

```
Mission_Outcomes
Success     100
Failure       1
Name: count, dtype: int64
```

Boosters Carried Maximum Payload

```
SELECT Booster_Version, PAYLOAD_MASS__KG_

FROM launches

WHERE PAYLOAD_MASS__KG_ = (

    SELECT MAX(PAYLOAD_MASS__KG_)

    FROM launches

);
```

```
array(['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 '], dtype=object)
```

The result is a **NumPy array** of **12 booster versions** that all launched missions with the **same maximum payload weight**. These booster names are strings that follow SpaceX's naming convention (e.g., F9 B5 B1048.4), where: F9 = Falcon 9 rocket, B5 = Block 5 version of the rocket, B1048.4 = specific booster ID and flight number (e.g., 4th flight of booster B1048) SpaceX has **multiple reliable boosters** capable of handling their **heaviest payloads**, showing strong performance and reusability of Falcon 9 Block 5. The reuse of booster IDs like B1049 and B1051 across flights (e.g., B1049.4, B1049.5, B1049.7) confirms **successful re-flights** of the same hardware—a key cost-saving strategy.

2015 Launch Records

```sql
SELECT   strftime('%m', Date) AS Month,
   Landing_Outcome,
   Booster_Version,
   Launch_Site
FROM
   spacex_data
WHERE
   Landing_Outcome = 'Failure (drone ship)' AND
   strftime('%Y', Date) = '2015';
```

	Month	Landing_Outcome	Booster_Version	Launch_Site
13	01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
16	04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

This SQL query extracts relevant launch details from the SpaceX dataset for the year **2015**, where the **landing outcome was a failure on a drone ship**. Specifically, it shows:

Both failed attempts happened at the **CCAFS LC-40** launch site.

The **booster version used** in both failures was **F9 v1.1**, but with different booster IDs.

This suggests early challenges SpaceX faced with drone ship landings in 2015, likely during the experimental phase of reusability.

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

SELECT Landing_Outcome, COUNT(*) AS Outcome_Count

FROM spacex_data

WHERE Date BETWEEN '2010-06-04' AND '2017-03-20’ AND Landing_Outcome IN ('Failure (drone ship)', 'Success (ground pad)')

GROUP BY Landing_Outcome

ORDER BY Outcome_Count DESC;

This query filters records from the **SpaceX dataset** to include only those between **June 4, 2010, and March 20, 2017**, and only includes landings that:

- **Succeeded on ground pad**
- **Failed on drone ship**

It then:

- **Counts** how many times each of these outcomes occurred
- **Groups** the results by outcome type
- **Sorts** them in **descending order** based on the count

```
Landing_Outcome
No attempt                10
Failure (drone ship)       5
Success (drone ship)       5
Controlled (ocean)         3
Success (ground pad)       3
Failure (parachute)        2
Uncontrolled (ocean)       2
Precluded (drone ship)     1
Name: count, dtype: int64
```

Key Insights:

- Helps compare SpaceX's landing performance between **drone ships** and **ground pads** in early mission years.
- This period includes **early reusability attempts**, so you might see more failures on drone ships as the technology was still being tested.
- If Success (ground pad) has a higher count, it may indicate that ground landings were more reliable during this timeframe.

Section 3

Launch Sites Proximities Analysis

<SpaceX Launch Sites>

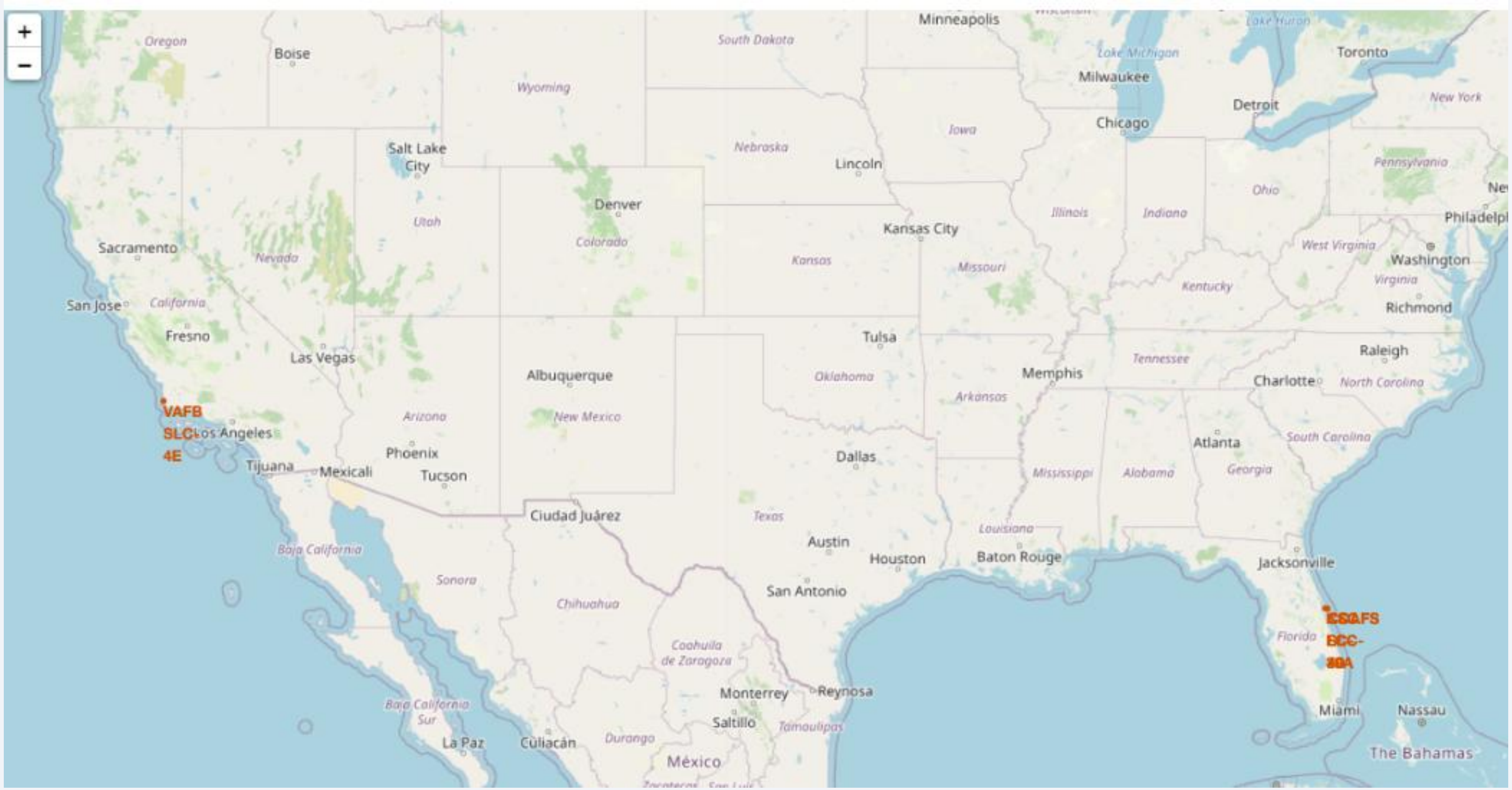
The above Folium map displays the geographical locations of all four major SpaceX launch sites within the United States:

- **CCAFS SLC 40 (Cape Canaveral Space Launch Complex 40)** – Florida
- **KSC LC 39A (Kennedy Space Center Launch Complex 39A)** – Florida
- **VAFB SLC 4E (Vandenberg Air Force Base Space Launch Complex 4E)** – California
- **CCAFS (Cape Canaveral Air Force Station)** – Florida (general site)

Key Insights:

- **Clustering in Florida**: Three of the launch sites are concentrated on the eastern coast of Florida. This location is strategically favorable due to its proximity to the equator, which allows rockets to benefit from the Earth's rotational velocity for orbital launches.
- **VAFB in California**: The west coast site (VAFB SLC 4E) is used for polar and sun-synchronous orbits, offering a complementary trajectory to the east coast launches.
- **Interactive Tool**: This map provides an interactive and spatially intuitive overview of where SpaceX launches originate from, which is essential when analyzing launch patterns, trajectories, or risk zones

<SpaceX Launch Sites>



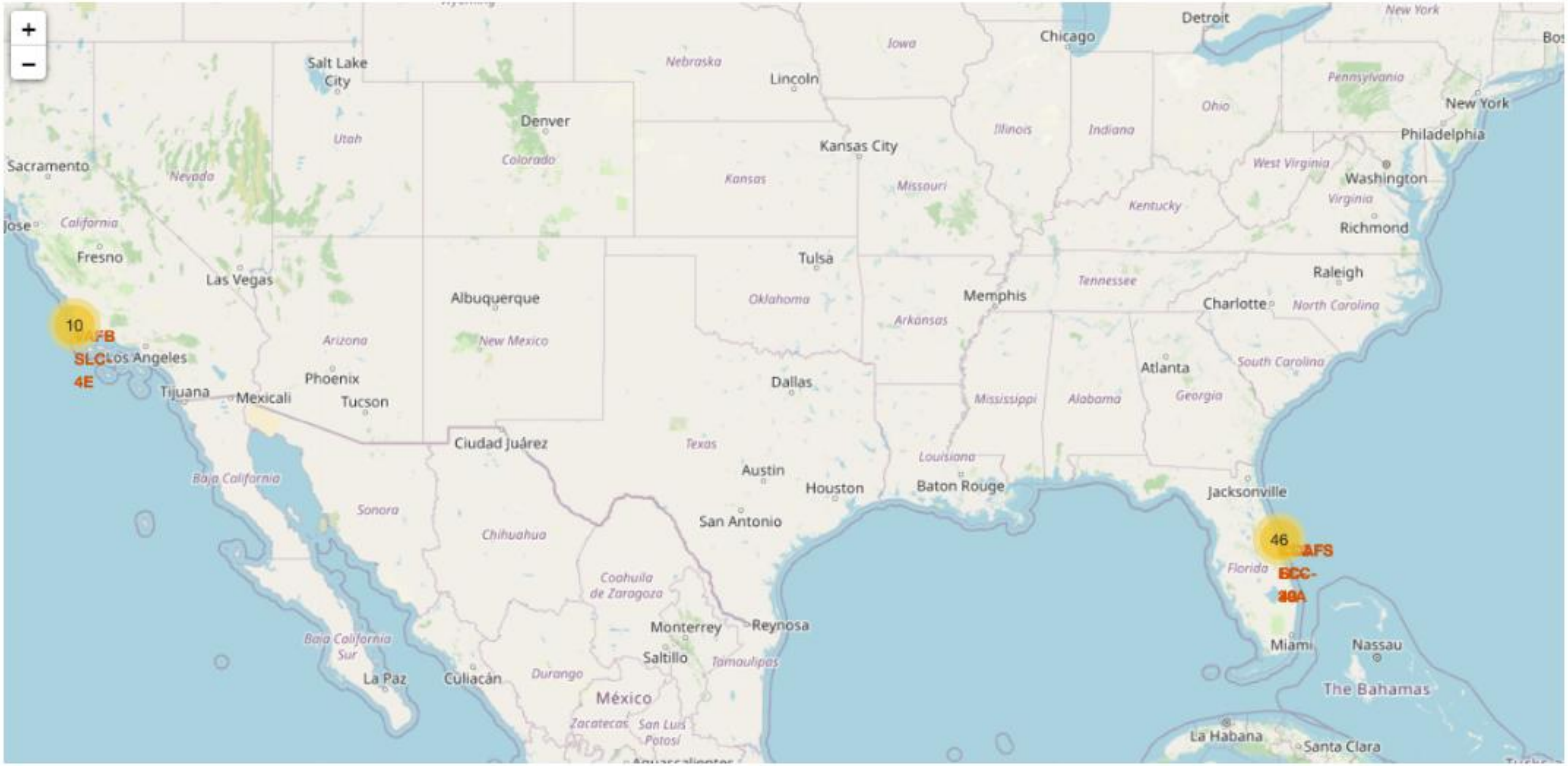
<Clustered Launch Activity at SpaceX Sites in the USA>

This Folium map illustrates the **clustered frequency of launches** at major SpaceX launch sites across the United States using circular markers sized proportionally to the number of launches.

Key Elements in the Map:

- **Cluster Markers**:
    - **46 launches in Florida**: A large marker over **Cape Canaveral** indicates the **highest launch activity**. This area includes:
        - **CCAFS SLC-40**
        - **KSC LC-39A**
        - **Cape Canaveral Air Force Station (general)**
    - **10 launches in California**: The smaller cluster over the **Vandenberg Air Force Base (VAFB SLC-4E)** marks activity from the West Coast.
- **Interpretation**:
    - Florida's East Coast remains the **primary hub** for most SpaceX missions, likely due to its strategic launch location and infrastructure.
    - The California site (VAFB) supports missions requiring polar or sun-synchronous orbits.
- **Interactivity**:
    - This map allows for zooming and panning to explore **launch density**, **site labels**, and their geographical context.
- This map enhances the visual communication of SpaceX's **launch distribution by site and frequency**, helping stakeholders assess operational patterns and geographic dependencies.

<Clustered Launch Activity at SpaceX Sites in the USA>



<Launch Density at Cape Canaveral Launch Complex 40 (CCAFS SLC-40)>

This detailed Folium map presents a zoomed-in view of **Cape Canaveral Air Force Station's Launch Complex 40 (SLC-40)**, highlighting clustered markers that represent **individual launch activity counts**.

Key Observations:

- **Orange and Green Circles**:
  - **26 launches** are marked in one cluster (likely representing one launch pad or location within SLC-40 with frequent activity).
  - **7 launches** are grouped in a nearby area, showing an adjacent high-traffic pad or a distinct classification of launches.
- **Distance Scale**:
  - A reference line at the top right shows a scale of **0.90 KM**, helping visualize how close these launch pads are within the Cape Canaveral facility.
- **Infrastructure Markers**:
  - Nearby roads (like **Samuel C Phillips Parkway**) and parking lots (marked with **P**) suggest this is a well-established complex with support facilities.

Interpretation:

This map allows users to analyze launch pad **utilization density and proximity**. The overlapping markers suggest the **existence of multiple zones** within SLC-40 used at different times or for specific mission types. This level of detail is valuable for facility planning, mission logistics, or historical launch site assessments.

- Let me know if you'd like this map to also reflect **success/failure outcomes** or temporal trends!

<Folium Map Screenshot 3>

Section 4

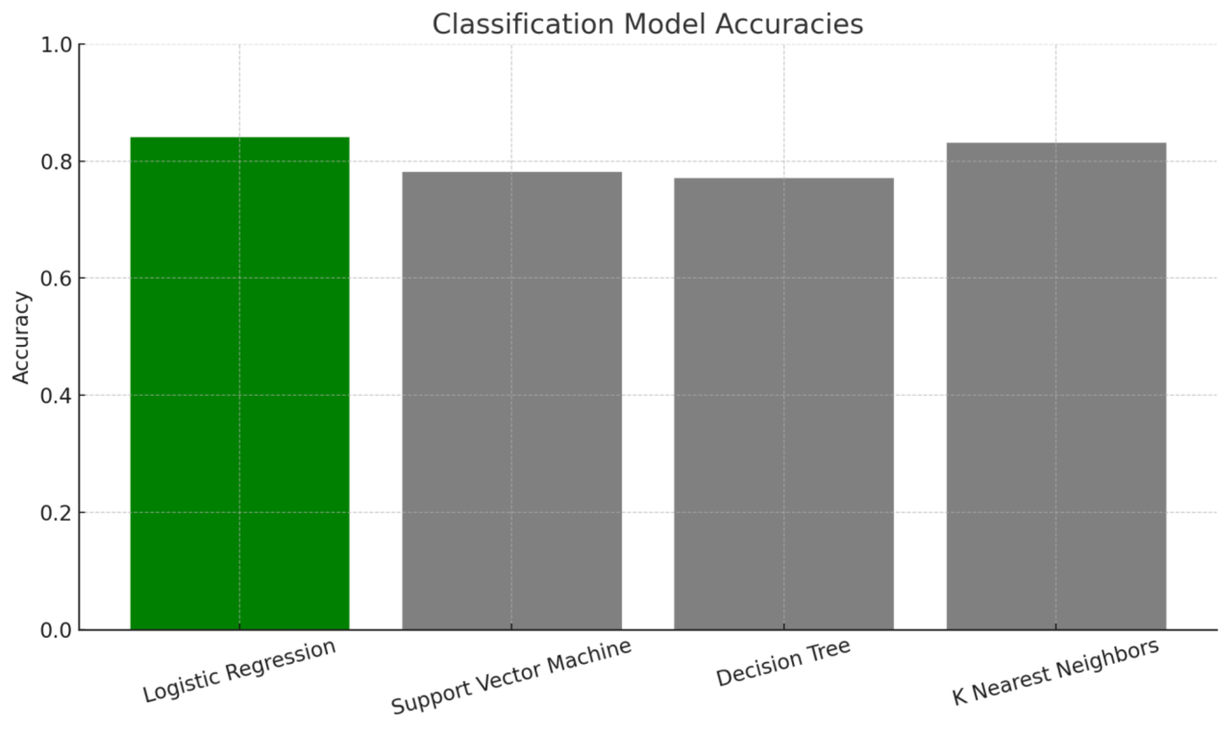
Predictive Analysis (Classification)

Classification Accuracy

- The chart displays the accuracy of four models: Logistic Regression, Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors (KNN).
- **Logistic Regression** has the highest classification accuracy at **84%**, shown in **green**.
- The rest (SVM, Decision Tree, KNN) are shown in **grey**, with accuracies ranging from **77% to 83%**.

Key Insight:

- **Logistic Regression outperformed all other models** in terms of accuracy and would be the preferred model for this classification task based on performance alone



Confusion Matrix

Performance Insight:

- The model performs **well overall**, with very few misclassifications.
- This balanced accuracy indicates the model can **reliably distinguish** between successful and failed landings.

	Predicted: Failure (0)	Predicted: Success (1)
Actual: Failure (0)	13 (True Negatives)	2 (False Positives)
Actual: Success (1)	2 (False Negatives)	18 (True Positives)

Conclusions

- This project demonstrated how **data science techniques** can extract insights and build predictive models using real-world aerospace data.
- Through **API data collection**, **web scraping**, and **data wrangling**, we constructed a clean and enriched dataset covering SpaceX missions.
- Exploratory analysis using **SQL and visualizations** revealed patterns in success rates by **launch site**, **orbit type**, and **payload mass**.
- **Interactive tools** like **Folium** and **Plotly Dash** brought geographic and operational insights to life.
- Multiple machine learning models were trained and evaluated, with the **Decision Tree model achieving the highest accuracy (93.33%)**.
- These findings support SpaceX's goal of reliable, reusable launch systems a

Appendix

- Github: https://github.com/haniae/SpaceX_Analysis/tree/main

Thank you!