IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Edit Guntermann

21.10.2025

Outline

- Executive Summary
- Introduction
- Project Objectives
- Methodology
- Results
- Conclusion

Executive Summary

- This project analyzes **SpaceX Falcon 9 launch records** to determine the factors influencing **first-stage landing success**.
- Multiple data sources were used, including **SpaceX REST API**, **web scraping from Wikipedia**, and **CSV datasets**.
- Performed **data wrangling**, **exploratory data analysis (EDA)**, **SQL analysis**, **interactive geospatial analysis**, and **dashboard development**.
- Built and evaluated **multiple machine learning models** to predict landing outcomes.
- **Decision Tree Classifier** achieved the highest accuracy on test data (**94%**), demonstrating strong predictive potential.

Introduction

- SpaceX has revolutionized space travel by **reusing Falcon 9 first-stage boosters**, significantly reducing launch costs.
- The goal of this project is to explore **which factors affect landing success** and **predict future mission outcomes**.

Introduction

Research Questions:

- What features correlate with successful booster landings?
- Which launch sites demonstrate highest success rates?
- Can we predict landing outcomes based on mission characteristics?

Business Value:

- Improving landing prediction helps **optimize mission planning** and **reduce operational risks**.

Project Objectives

- Collect and prepare historical SpaceX launch data.
- Perform **EDA using visualization and SQL queries**.
- Build **interactive analytics tools** using Folium Maps and Plotly Dash.
- Develop and compare **classification models**:
  - Logistic Regression
  - Support Vector Machine (SVM)
  - Decision Tree
  - K-Nearest Neighbors (KNN)
- Identify the **best model** for predicting landing success.
- Deliver **actionable insights** and recommendations.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Retrieved launch data from SpaceX REST API
  - Performed web scraping from Wikipedia
- Perform data wrangling
  - Cleaned, standardized and combined datasets
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Built classification models to predict landing outcomes, compared models and derived conclusions

Data Collection

Data was collected from multiple sources:

- **SpaceX REST API**

  Historical Falcon 9 launch data programmatically retrieved

- **Web Scraping (Wikipedia)**

  Retrieved mission details like booster versions and customer payload data

- **SpaceX Launch CSV File**

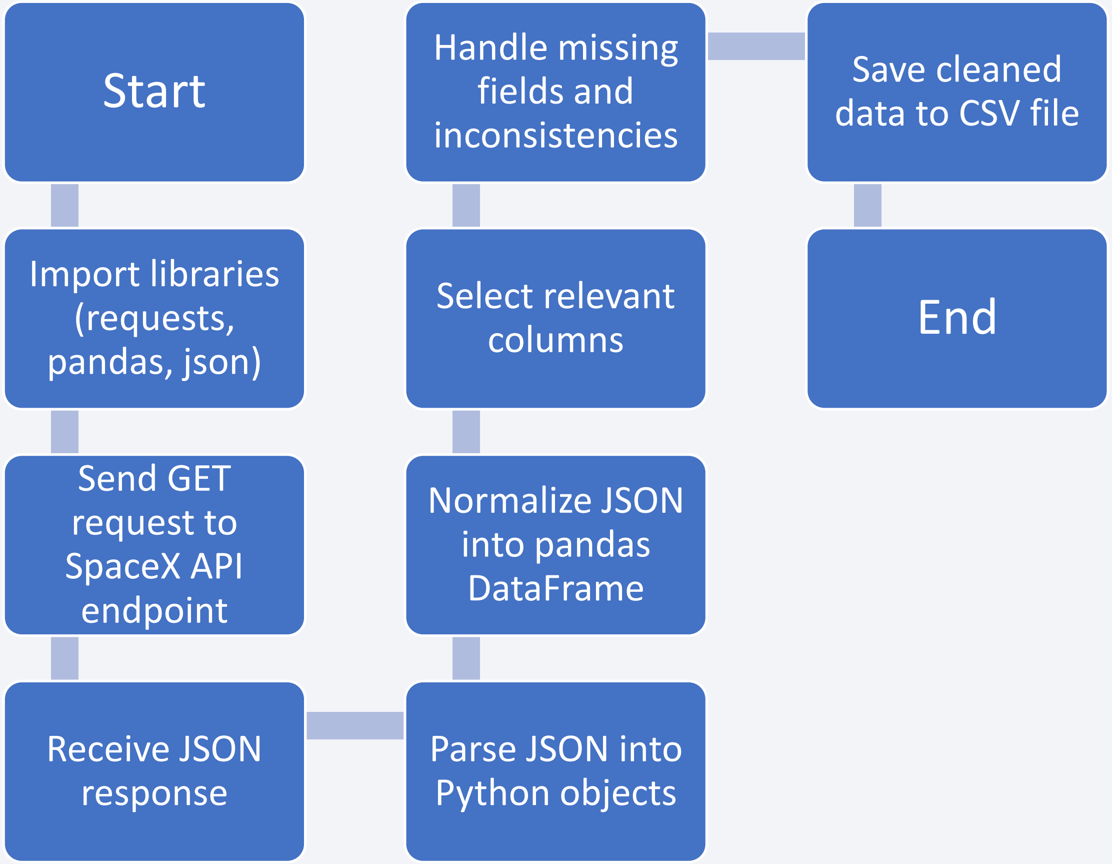
  Used in dashboard development and payload analysis

- Total dataset size: **90+ launches** with **multiple mission attributes**

Data Collection – SpaceX API

Used Python requests library to access **SpaceX REST API** endpoints

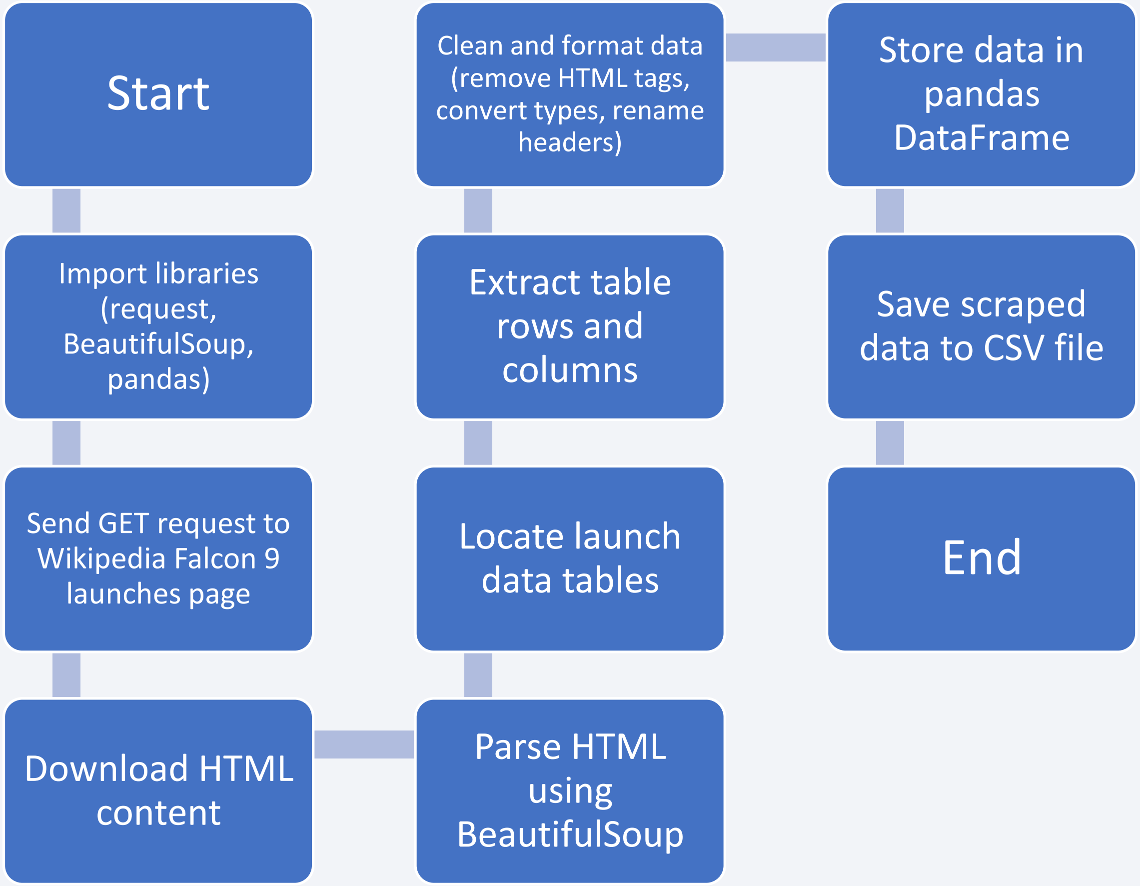
- Extracted data fields:
  - Flight Number
  - Launch Date
  - Launch Site
  - Orbit Type
  - Payload Mass
  - Landing Outcome
- Data converted to pandas DataFrame
- GitHub Notebook: [SpaceX Data Collection API]



Data Collection - Scraping

Used **BeautifulSoup** to scrape Falcon 9 mission data from Wikipedia

- Extracted additional features:
  - Booster Version
  - Launch Outcome
  - Payload Type
- Cleaned HTML table structure and converted to DataFrame
- GitHub Notebook: *Web Scraping Falcon 9 Launch Data*



Data Wrangling

Performed data cleaning and preprocessing:

- Handled **missing values** and replaced unknown payloads
- Standardized **landing outcome labels**
- Removed **irrelevant columns**
- Created **binary target column**: Class
  - 1 = Successful landing
  - 0 = Failed landing
- Merged all datasets into a single analytical table
- GitHub Notebook: *Data Wrangling and Preprocessing*

EDA with Data Visualization

Explored relationships between key mission variables using **Matplotlib, Seaborn**, and **Plotly**

- Plotted scatter, bar, and line charts to:
    - Analyze launch trends over time
    - Compare launch site performance
    - Examine payload effects on success
    - Understand relationships with orbit type
- Visual analysis helped identify:
    - Launch sites differ in reliability
    - Higher payloads correlate with better missions
    - Success rates improved over the years
- GitHub Notebook: *Exploratory Data Analysis with Visualization*

EDA with SQL

Queried SpaceX launch dataset stored in a SQL database

- Used SQL queries to:
  - Summarize launch statistics
  - Filter mission outcomes by year, site, and orbit
  - Identify highest and lowest performing missions
  - Rank landing outcomes
- SQL helped perform structured analysis efficiently
- SQL skills used: SELECT, COUNT, GROUP BY, WHERE, ORDER BY, JOIN, LIKE, HAVING
- GitHub Notebook: *EDA with SQL Queries*

Build an Interactive Map with Folium

Created geospatial visualizations using Folium library

- Added:
    - Markers for each launch site
    - Color-coded circles based on landing success
    - Distance markers to highways, railways, and coastlines
- Mapped mission geography for insight into location impact
- Focused on proximity analysis of launch site environments
- GitHub Notebook: *Interactive Map Visualization with Folium*

Build a Dashboard with Plotly Dash

Built an interactive analytics dashboard for launch insights

- Included:
    - Launch site filter dropdown
    - Success pie chart
    - Payload range slider
    - Scatter plot for payload vs success
- Enabled data exploration without coding
- Shows mission performance patterns interactively
- Python Script: *spacex-dash-app.py (hosted on GitHub)*

Predictive Analysis (Classification)

Built supervised learning models to predict landing success

- Models used:
  - Logistic Regression
  - Support Vector Machine
  - Decision Tree
  - K-Nearest Neighbors
- Performed:
  - Train-Test split
  - Feature scaling
  - Hyperparameter tuning with GridSearchCV
  - Model evaluation using accuracy and confusion matrix
- **Best-performing model: Decision Tree (94% test accuracy)**
- GitHub Notebook: *Predictive Analysis – Classification*

Results

The analysis produced insights across three major areas:

- **Exploratory Data Analysis (EDA):**
  - Identified key trends across launch sites, payloads, orbit types, and mission timelines
  - Found mission factors most correlated with landing success
- **Interactive Analytics:**
  - Built dynamic tools using Folium maps and a Plotly Dash dashboard for deeper data exploration
  - Enabled geographic and site-based analysis with interactive visualizations
- **Predictive Analysis:**
  - Developed and compared machine learning models to predict landing success
  - Achieved highest accuracy using Decision Tree Classifier (94%)

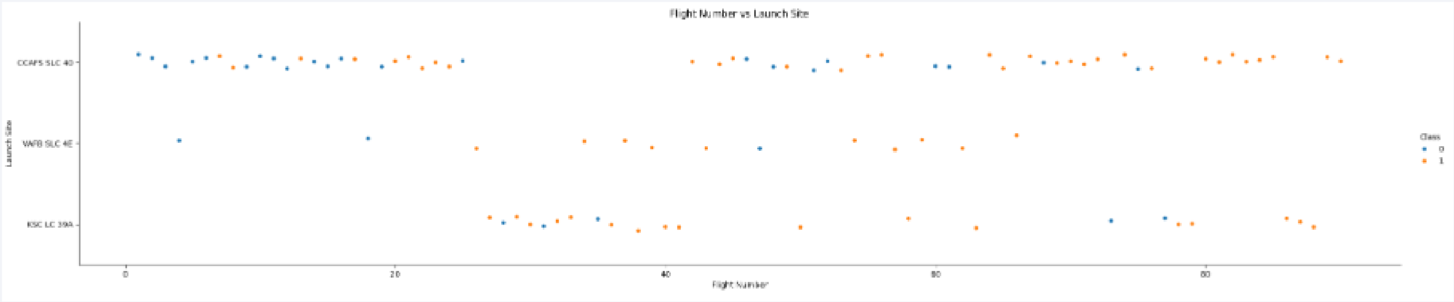
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

This scatter plot shows the relationship between flight number and launch site.

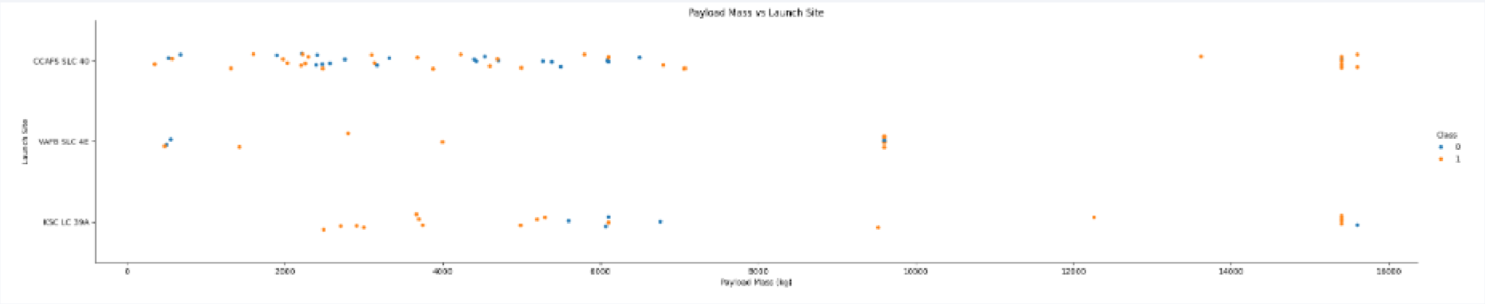
- Each dot represents a mission.
- Insights:
  - KSC LC-39A and CCAFS SLC-40 have the highest mission counts.
  - VAFB SLC-4E has fewer launches but a higher proportion of successful missions.
  - Launch frequency has increased over time, shown by higher flight numbers concentrated in specific sites.
- Conclusion: Certain launch sites show operational priority and reliability over others.



Payload vs. Launch Site

Scatter plot analyzes Payload Mass (kg) vs Launch Site.

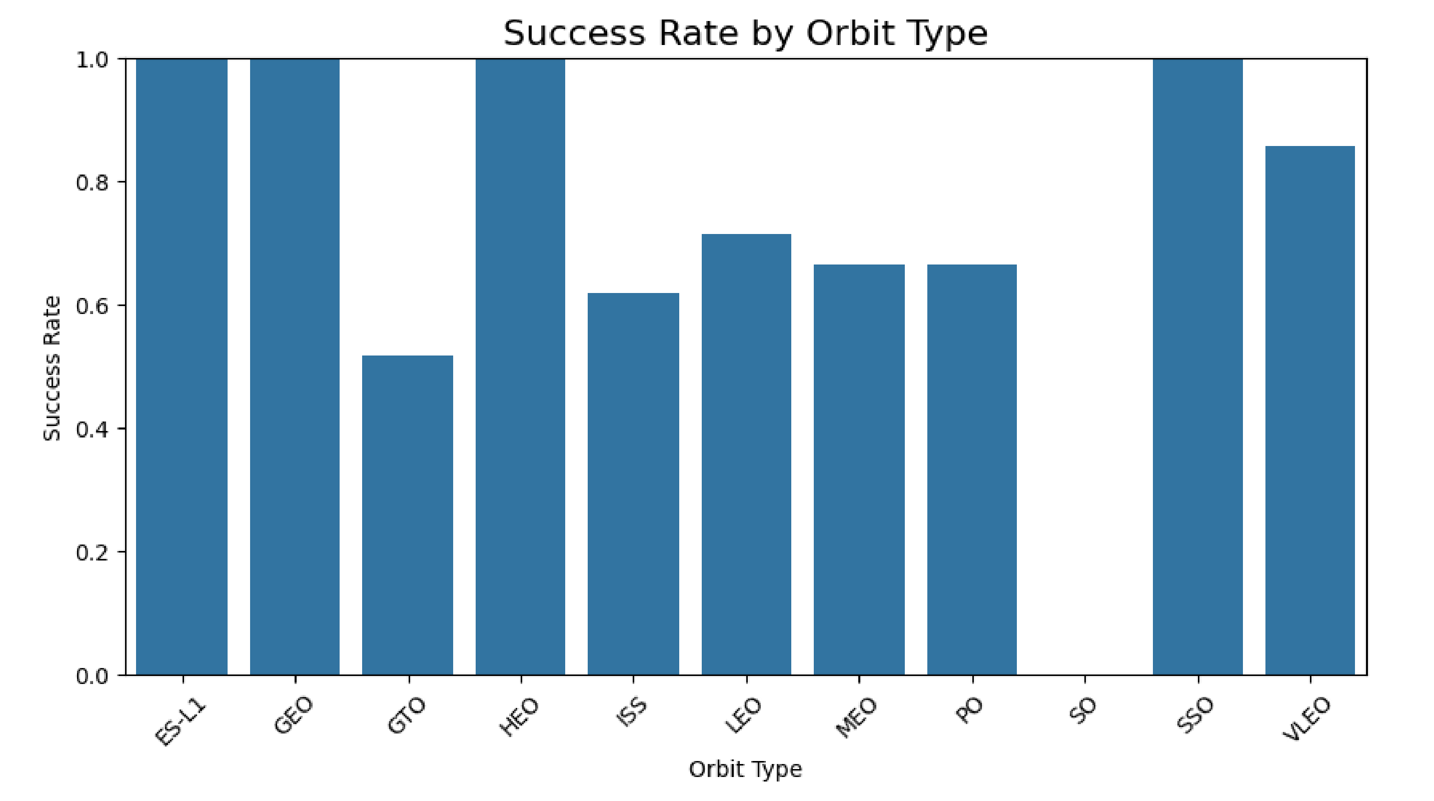
- Findings:
  - KSC LC-39A handled heaviest payloads, supporting large commercial satellites.
  - VAFB SLC-4E missions mostly had medium payloads.
  - CCAFS SLC-40/LC-40 hosted a mix of light to medium-size payload launches.
- Conclusion: Payload capacity utilization varies by site, suggesting site specialization.



Success Rate vs. Orbit Type

Bar chart compares landing success rates for different orbit types.

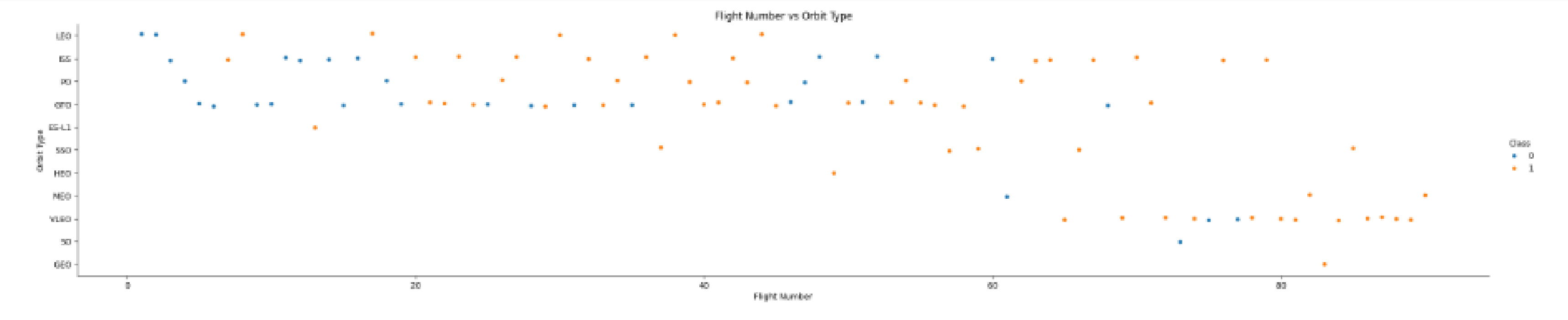
- Observations:
    - Highest success rates occur with ES-L1, GEO, HEO and SSO missions.
    - GTO (Geostationary Transfer Orbit) missions show lowest success rates, likely due to higher mission complexity.
    - Other orbits like LEO/MEO are less frequent and exhibit more variability.
- Conclusion: Mission orbit type influences landing success probability.



Flight Number vs. Orbit Type

Scatter plot shows orbit types used over time.

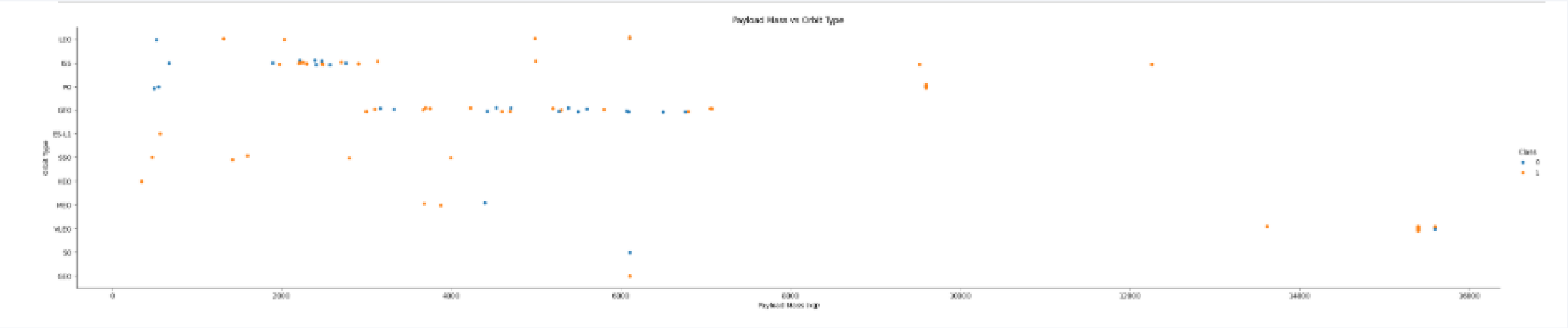
- Observations:
  - Early missions focused on LEO orbits.
  - Gradually expanded to include GTO and ISS resupply missions.
  - Increased mission diversity reflects growing mission confidence and reuse capability.
- Conclusion: SpaceX has expanded mission capability over time.



Payload vs. Orbit Type

Scatter plot of Payload Mass vs Orbit Type.

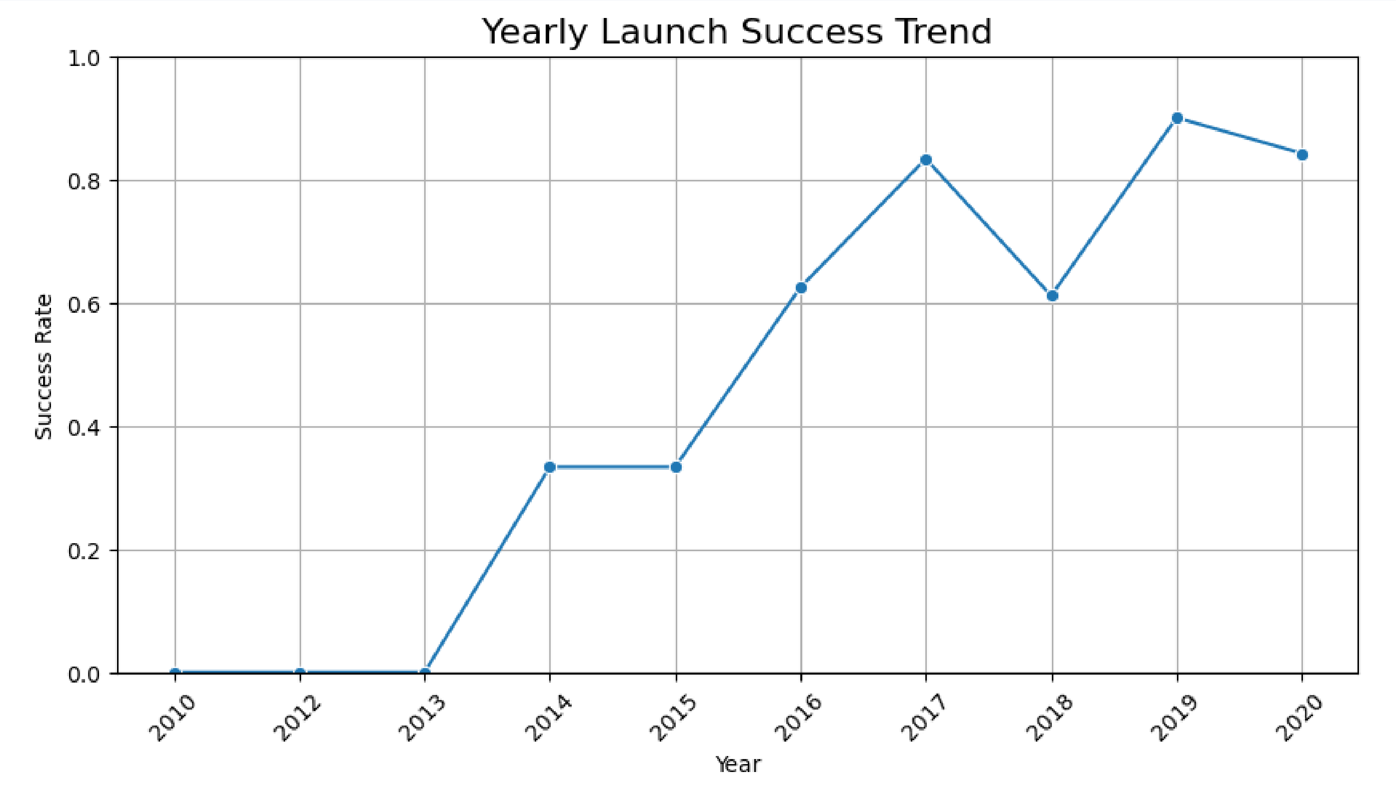
- Observations:
  - GTO missions carry heavier payloads.
  - Polar/Sun-synchronous orbits (SSO) carry lighter scientific and Earth observation payloads.
  - LEO missions show the widest range of payload sizes.
- Conclusion: Payload mass is strongly linked with orbit selection.



Launch Success Yearly Trend

Line chart shows average landing success rate per year.

- Observations:
  - Noticeable improvement from 2013 onward.
  - Success rate crossed 80% by 2017, coinciding with booster reuse improvements.
  - Fewer failures after 2016 due to system refinements.
- Conclusion: SpaceX landing success improved significantly with iterative engineering and booster reuse strategy.



All Launch Site Names

Query: Find all unique launch site names

SQL Purpose:

- Identify distinct launch sites in the dataset.

Insight:

- There are **4 unique launch sites** used by SpaceX for Falcon 9 launches.
- These are:
  - KSC LC-39A
  - VAFB SLC-4E
  - CCAFS LC-40
  - CCAFS SLC-40

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

Conclusion: SpaceX conducts operations primarily from NASA Kennedy Space Center and Air Force Stations in California and Florida.

Launch Site Names Begin with 'CCA'

Query: Find 5 records where launch sites begin with CCA%

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

SQL Purpose:

- Filter launch records by site prefix.

Insight:

- Query returns missions launched from Cape Canaveral Air Force Station (CCAFS).
- Both **CCAFS LC-40** and **CCAFS SLC-40** are included.

Conclusion: These sites host many early Falcon 9 missions and continue as major launch hubs.

Total Payload Mass

Total_Payload_Mass
45596

Query: Calculate the total payload mass of missions from NASA

Insight:

- NASA missions contributed a significant total payload mass.
- NASA used Falcon 9 for ISS resupply missions (CRS program).

Conclusion: Strong long-term collaboration exists between SpaceX and NASA.

Average Payload Mass by F9 v1.1

Query: Calculate average payload mass for booster version F9 v1.1

Avg_Payload_Mass
2928.4

Insight:

- F9 v1.1 carried mid-weight payloads as the early workhorse booster.
- It was used during transition to reusable technology.

Conclusion: F9 v1.1 was an important step before Falcon 9 FT (Full Thrust) upgrades.

First Successful Ground Landing Date

Query: Find the earliest date of a successful ground landing (success AND landing_pad)

First_Ground_Pad_Landing_Date
2015-12-22

Insight:

- The first successful ground landing occurred in 2015.
- This marked the first reuse breakthrough for SpaceX and commercial space history.

Conclusion: Ground landing success paved the way for booster reusability.

Successful Drone Ship Landing with Payload between 4000 and 6000

Query: List booster names that successfully landed on drone ships and had payload mass between 4000 and 6000 kg

Booster_Version	PAYLOAD_MASS__KG_	Landing_Outcome
F9 FT B1022	4696	Success (drone ship)
F9 FT B1026	4600	Success (drone ship)
F9 FT B1021.2	5300	Success (drone ship)
F9 FT B1031.2	5200	Success (drone ship)

Insight:

- Identifies reusable boosters capable of heavy missions and drone ship return.
- These missions typically could not return to land due to distance/velocity.
- Highlights Falcon 9 Full Thrust (FT) boosters as top performers.

Conclusion: Drone ship landings enable high-energy missions while maintaining reusability.

Total Number of Successful and Failure Mission Outcomes

Mission_Outcome	Total_Count
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Query: Count total successful vs failed landing outcomes

Insight:

- Dataset shows more successes than failures over time.
- Failure rate decreased significantly after iterative design improvements.
- Demonstrates SpaceX learning curve and reliability growth.

Conclusion: Landing success improves with technology maturity and mission experience

Boosters Carried Maximum Payload

Query: List booster(s) that carried the maximum payload mass

Insight:

- Heaviest payloads were carried by later Falcon 9 versions, especially Block 5.
- Capable of commercial satellite launches to GTO.
- Reinforces increased thrust and reliability in newer boosters.

Conclusion: Booster upgrades directly improve payload capability.

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

Query: Get failed drone ship landings in 2015 including booster version and launch site

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

Insight:

- Early drone ship recovery missions had mixed success.
- Most failures linked to engine re-entry burns and landing leg issues.
- All failures occurred during testing and development phases.

Conclusion: Failure data was crucial in improving drone ship landing technology.

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

Query: Rank landing outcomes between 2010-06-04 and 2017-03-20 in descending order

Insight:

- Successful drone ship and successful ground pad landings dominate ranking.
- Unsuccessful outcomes decline significantly over time.
- Validates steady improvement in Falcon 9 landing capabilities.

Landing_Outcome	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

Conclusion: SpaceX landing operations show consistent progress and optimization.

Section 3

Launch Sites Proximities Analysis

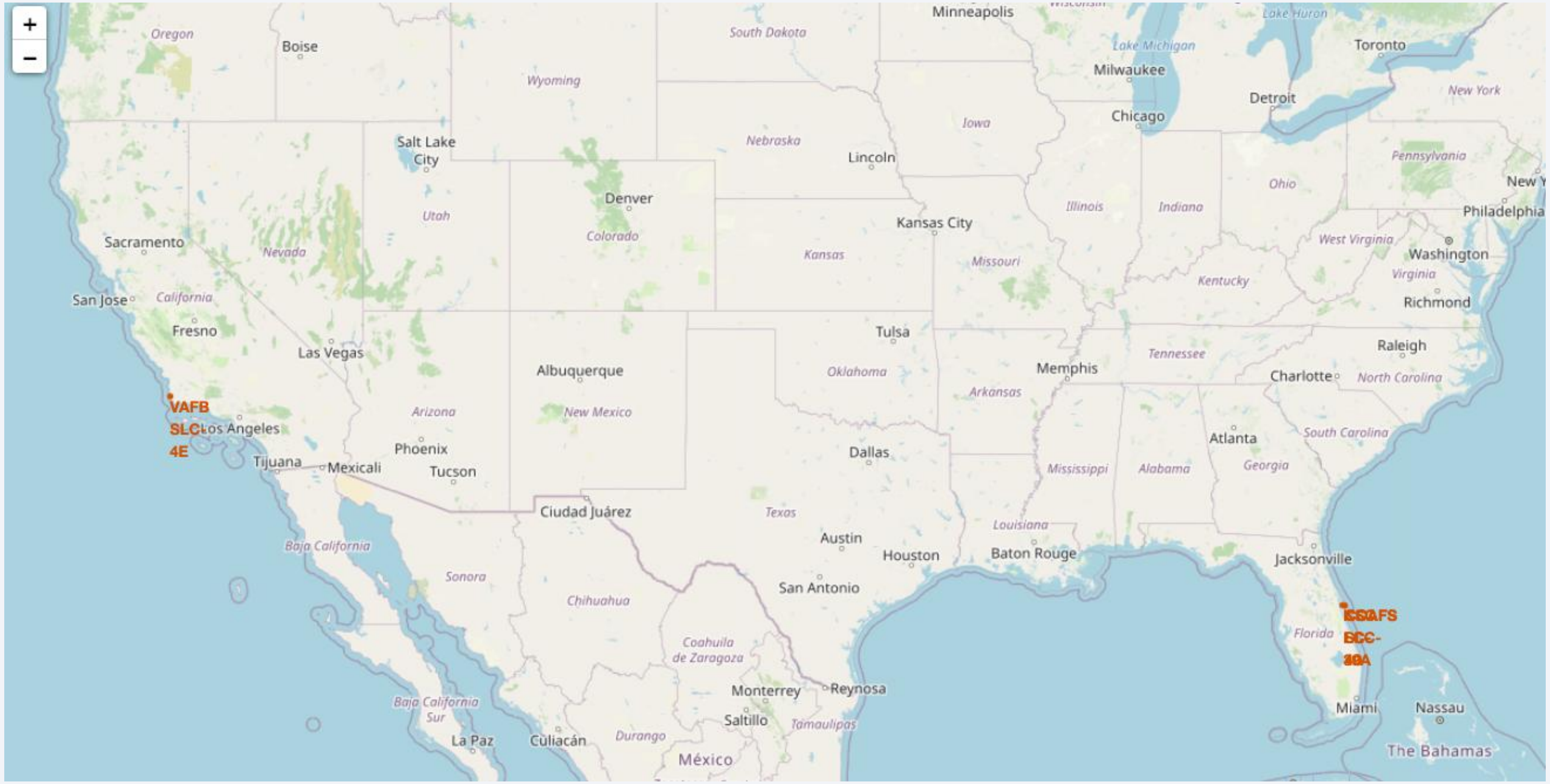
Global Falcon 9 Launch Sites

Description:

- This interactive Folium map displays the geographic locations of SpaceX launch sites.
- Markers were added for:
  - KSC LC-39A (Florida, USA)
  - CCAFS LC-40 (Florida, USA)
  - CCAFS SLC-40 (Florida, USA)
  - VAFB SLC-4E (California, USA)
- Map centered using average latitude and longitude of launch sites

Objective: Visualize the global footprint of Falcon 9 operations

Global Falcon 9 Launch Sites



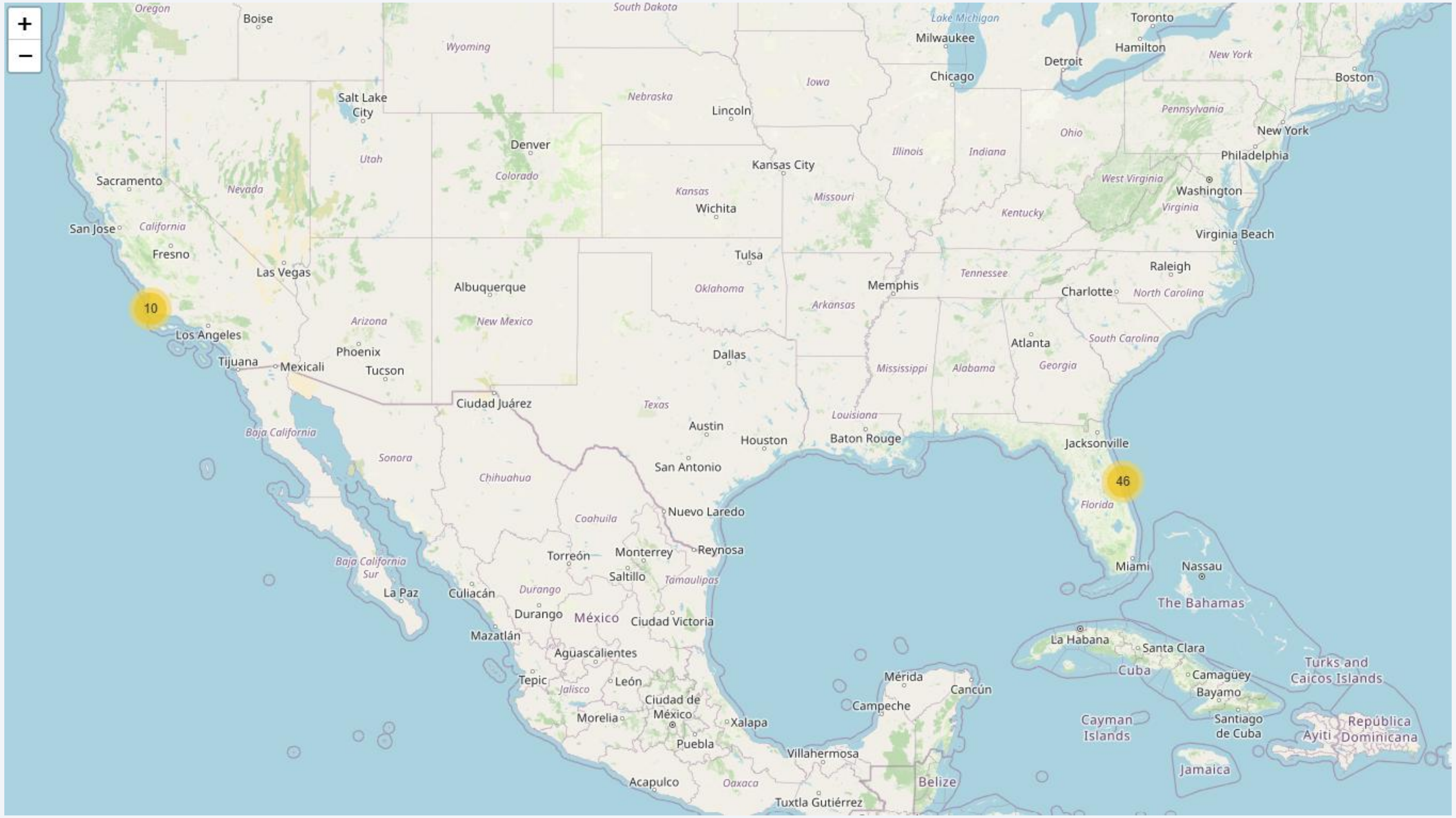
Landing Outcome Visualization

Description:

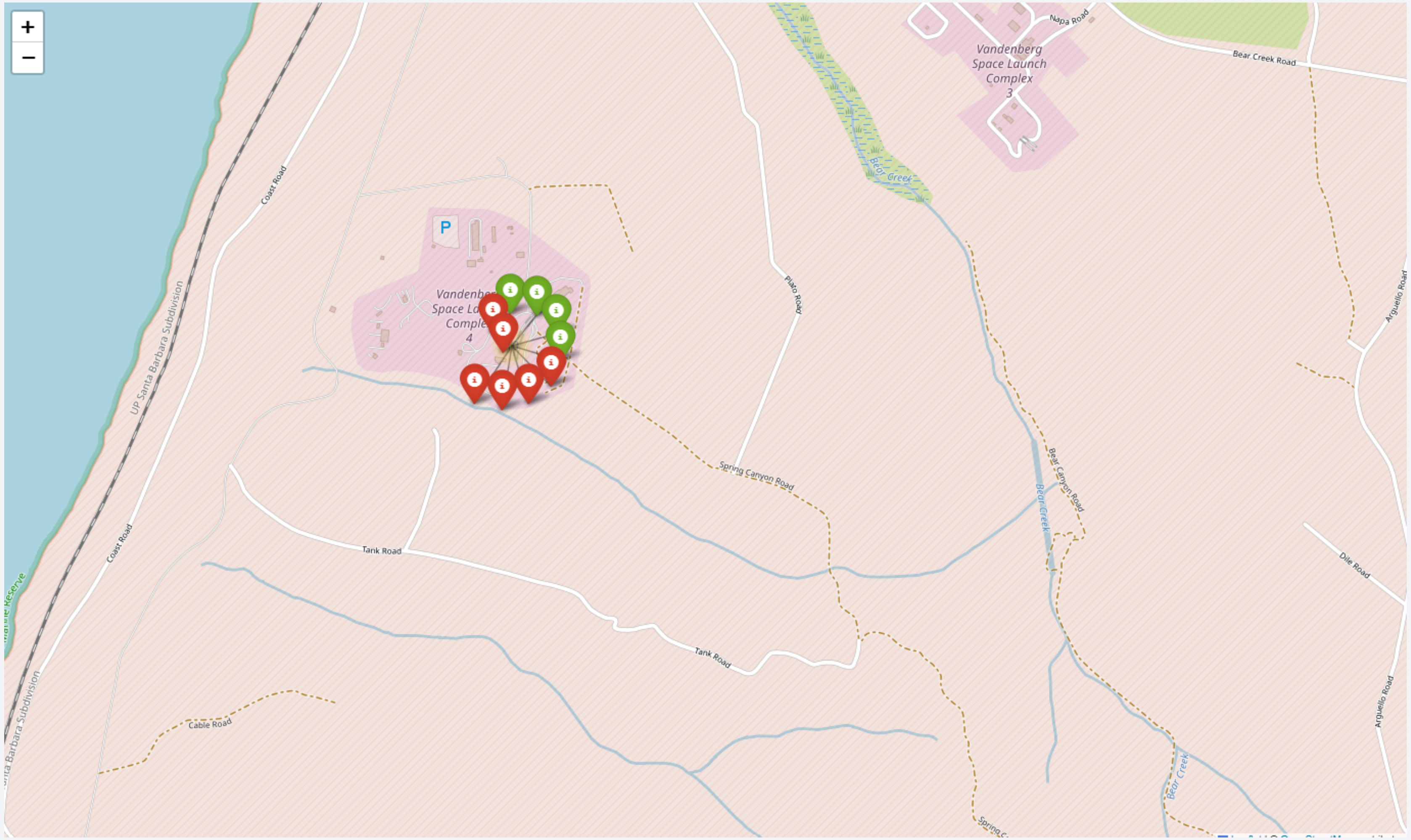
- Enhanced the map with color-coded markers:
  - 🟢 Green = Successful landing
  - 🔴 Red = Failed landing
  - ⚪ Grey = No landing attempt
- Tooltip and popup added for each marker with launch metadata

Purpose: Show landing performance patterns across launch locations

Landing Outcome Visualization



Landing Outcome Visualization



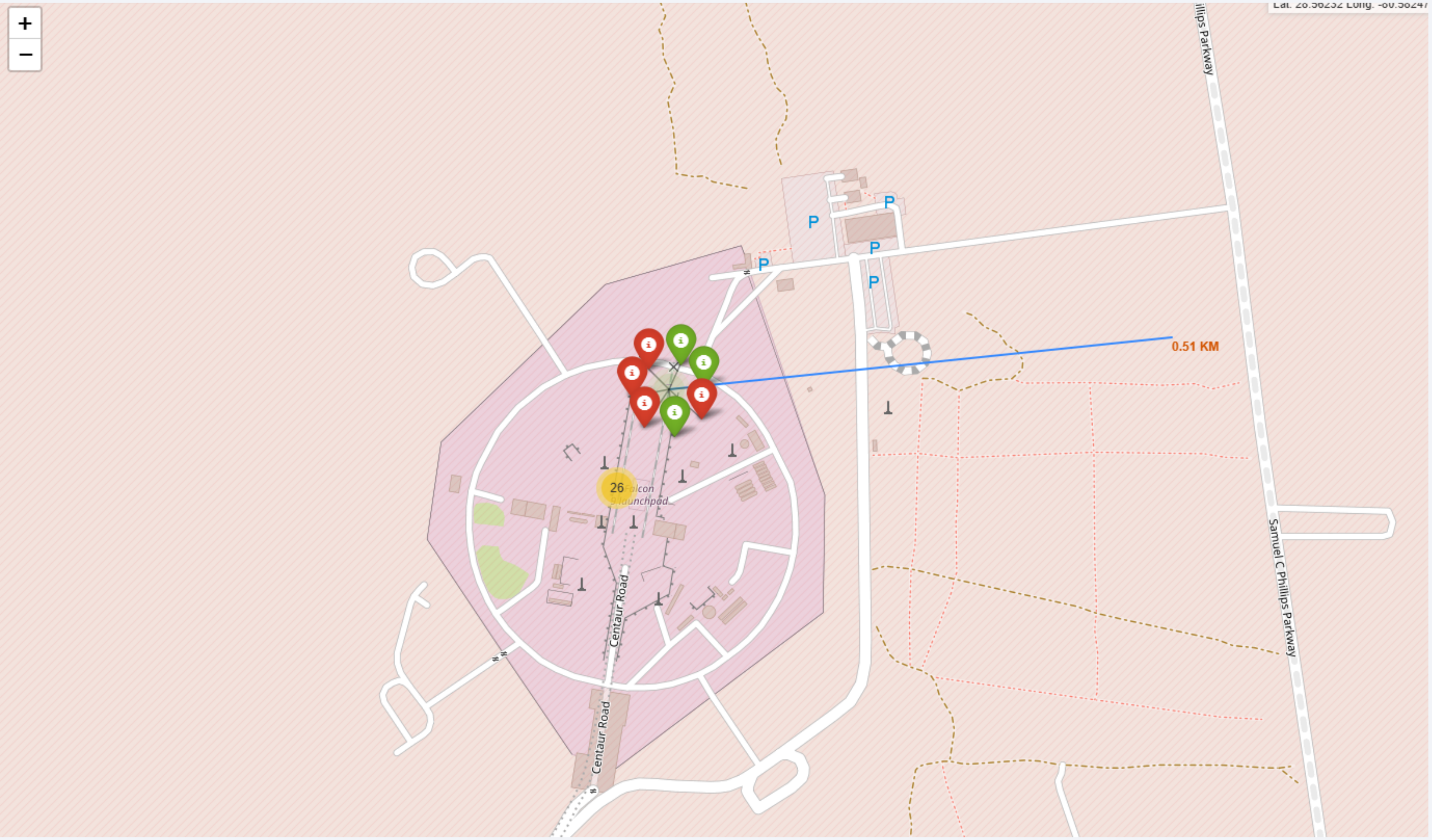
Proximity Analysis of Launch Site Environment

Description:

- Used circles and distance lines to measure:
  - Distance to nearest highway/road
  - Distance to coastline
  - Distance to railroad
- Example analysis:
  - VAFB SLC-4E is located close to coast → ideal for safe launch over ocean
  - CCAFS sites near roads → easy logistics for booster transport

Purpose: Understand environmental and safety factors affecting launch site selection

Proximity Analysis of Launch Site Environment



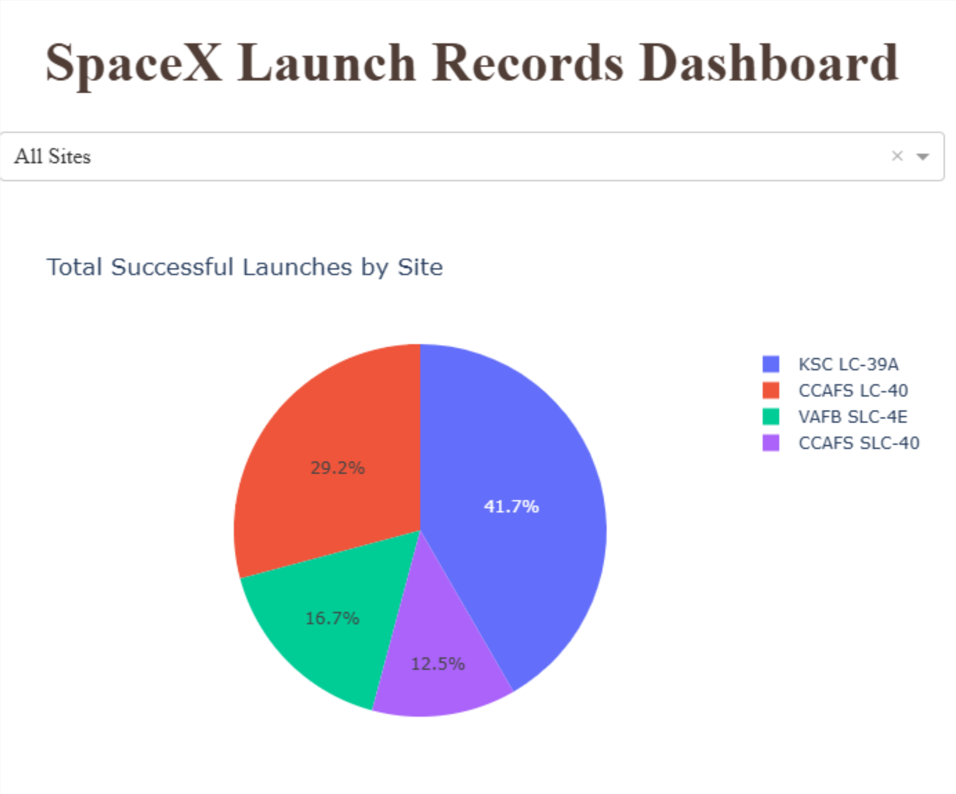
Section 4

Build a Dashboard with Plotly Dash

Launch Success Distribution Across Sites

This dashboard view shows total successful launches for each SpaceX site

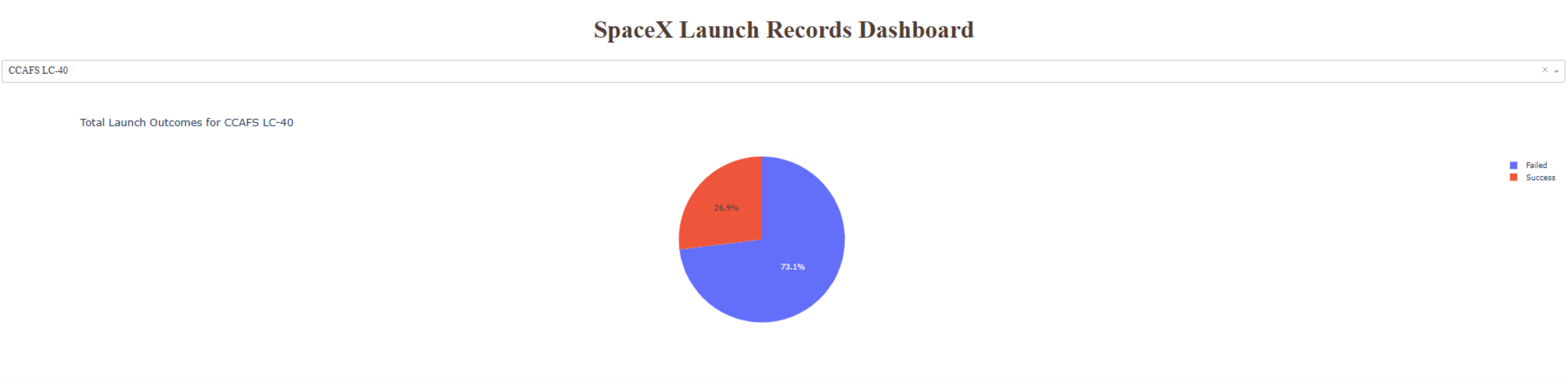
- Insights:
  - KSC LC-39A has the highest number of successful launches
  - CCAFS SLC-40 also has strong mission performance
  - VAFB SLC-4E has fewer launches but high success ratio
- Purpose: Understand site reliability and mission density



Highest Success Rate Launch Site

The dashboard was filtered to analyze each site individually

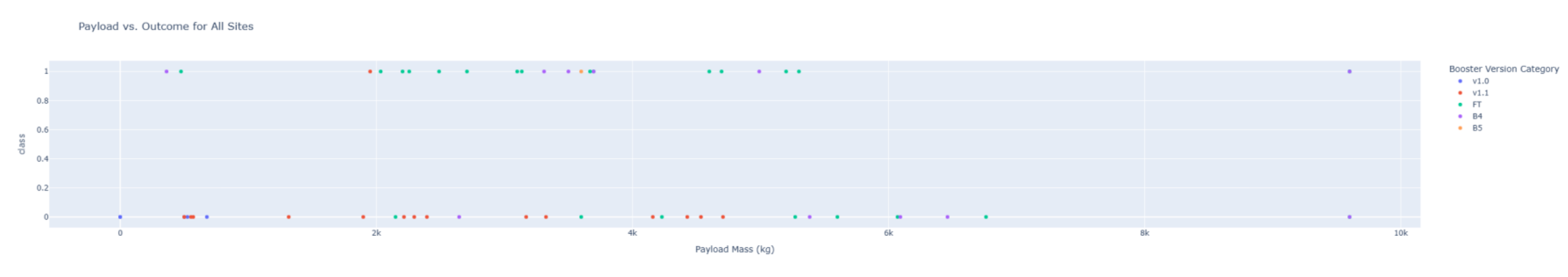
- Insight:
  - VAFB SLC-4E shows the highest success rate percentage
  - Although it has fewer missions, it demonstrates consistent performance
- Confirms earlier EDA finding using interactive analysis



Payload vs Landing Success by Booster Version

The scatter plot explores correlation between payload mass and landing success

- Observations:
  - Payloads between 2000–4000 kg show highest success rate
  - Heavier missions often use advanced boosters like Block 5
  - Booster Version FT shows highest reuse and success
- The payload slider allows interactive filtering for deeper insight



Section 5

Predictive Analysis (Classification)

Predictive Modeling Workflow

Goal: Predict the success of Falcon 9 first-stage landings using machine learning.

- Data Preparation:
  - Feature scaling with StandardScaler
  - Train-test split (80/20)
- Models developed and tuned using GridSearchCV (cv=10):
  - Logistic Regression
  - Support Vector Machine (SVM)
  - Decision Tree
  - K-Nearest Neighbors (KNN)
- Evaluation metrics:
  - Accuracy Score
  - Confusion Matrix

Classification Accuracy

Model test accuracies:

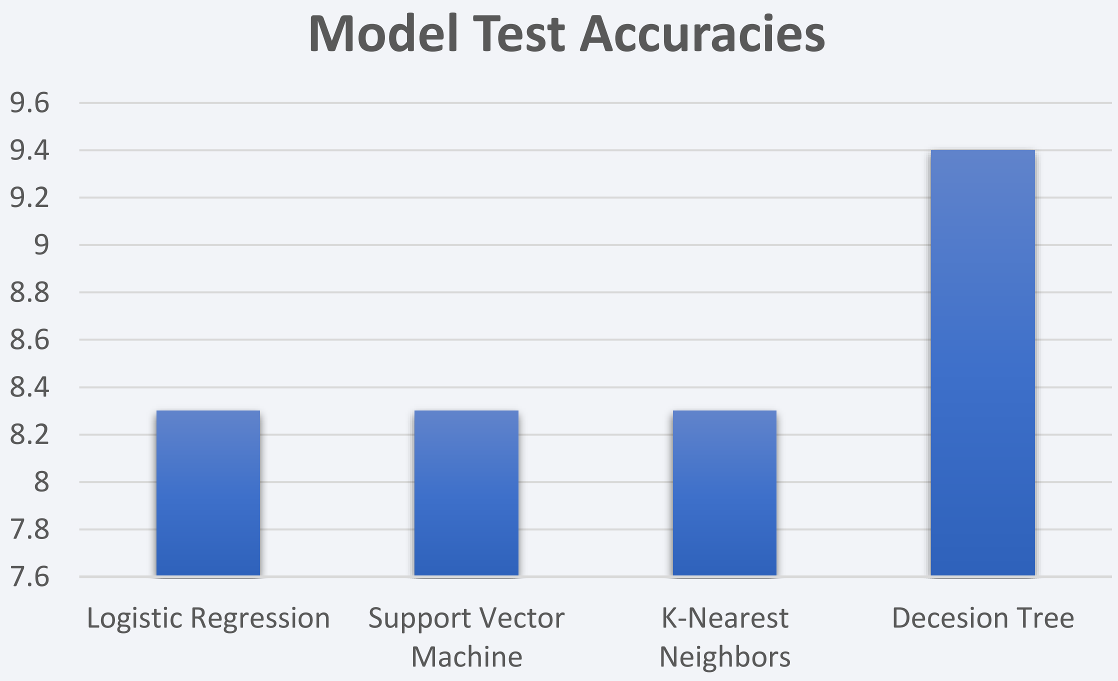
- Logistic Regression → 0.83
- Support Vector Machine → 0.83
- K-Nearest Neighbors → 0.83
- Decision Tree → 0.94

Decision Tree Classifier achieved the highest accuracy (94%)

Insights:

- Tree-based models captured non-linear relationships
- Performance consistent across cross-validation folds

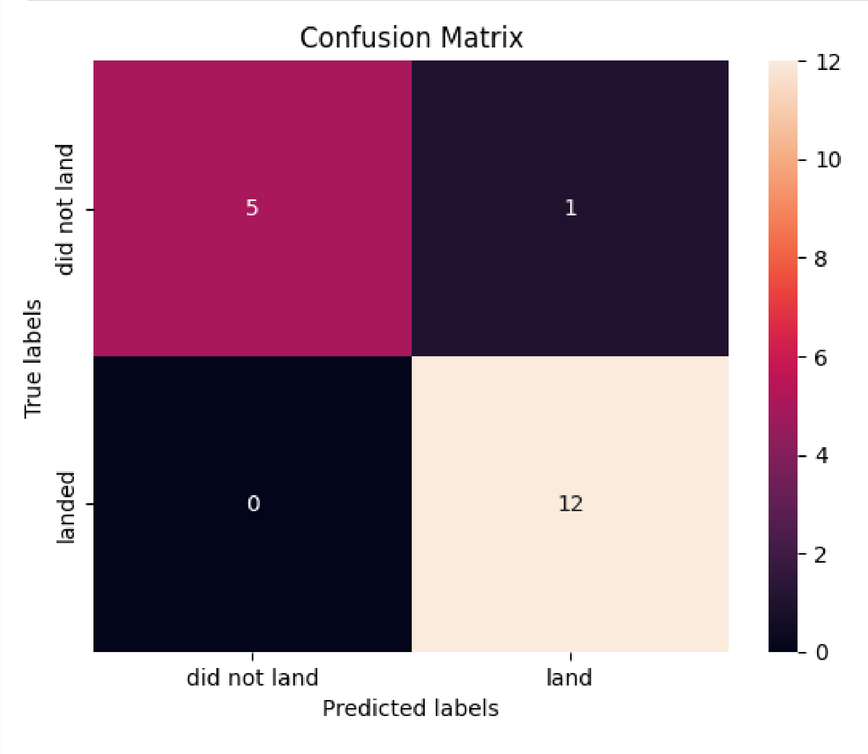
Conclusion: Decision Tree is the optimal model for predicting landing outcomes.



Confusion Matrix

The confusion matrix shows Decision Tree predictions on test data

- Metrics:
  - True Positives: Correctly predicted successful landings
  - True Negatives: Correctly predicted failed landings
  - Few misclassifications observed
- Interpretation:
  - The model is highly accurate and generalizes well
  - Suitable for predicting future Falcon 9 landing success



Key Findings and Insights

Data collection via REST API and web scraping successfully gathered launch details and outcomes for analysis.

Exploratory Data Analysis (EDA) revealed:

- Success rates varied across sites — KSC LC-39A had the most launches.
- Payload mass and booster version significantly influenced success.

Interactive visualizations (Folium + Dash) provided dynamic insights into:

- Geographic patterns of launch success.
- Payload and site-level success distributions.

Machine learning identified the Decision Tree Classifier as the most accurate model (94%).

Recommendations & Future Work

Operational Insight: Decision Tree model can be integrated into mission planning to predict landing success pre-launch.

Engineering Optimization: Focus on boosters with high reliability (e.g., Block 5) to maintain strong success rates.

Data Enrichment: Incorporate additional features such as weather data, booster reuse count, and engine thrust levels.

Future Work: Extend analysis with deep learning and time-series forecasting to anticipate anomalies in upcoming launches.

Overall Impact: Data-driven insights can help reduce costs and risks in future Falcon 9 missions.

Thank you!