IBM Developer
SKILLS NETWORK

SpaceX Falcone 9 launch Analysis

Nutukurthi Mani Prem Gowtham
13 Dec 2025

Outline

- Abstract
- Introduction
- Aim and Scope of the Study
- Methodologies
- Results
- Conclusion
- Appendix

Abstract

Falcon 9 is a partially reusable launch vehicle developed by SpaceX, known for its high reliability and advanced booster recovery system. This report presents an analytical report of Falcon 9 launch data obtained from the official SpaceX REST API and additional details collected through Wikipedia web scraping.

The study explores historical launch patterns, booster version performance, and the role of launch sites in mission success. A machine learning classifier is developed to predict whether a Falcon 9 booster will land successfully based on mission parameters such as payload mass, orbit type, booster version, and launch pad.

The findings highlight key factors influencing landing outcomes and demonstrate the effectiveness of machine learning in supporting mission analysis and aerospace decision-making.

Introduction

SpaceX's Falcon 9 rocket has pioneered cost-effective spaceflight through reusable booster technology. Understanding the factors that contribute to successful landings is essential for optimizing mission performance and improving rocket engineering.

This report analyzes Falcon 9 missions using dataset compiled from two sources:

- SpaceX open REST API
- Wikipedia launch history pages scraped using Python

The dataset covers mission parameters, booster versions, payload characteristics, launch sites, and landing results. The report includes exploratory data analysis (EDA), visualization, geospatial mapping, and machine learning classification.

Github link: https://github.com/NMPGowtham/SpaceX-Falcon-9-Analysis

Aim and Scope of the Study

- **Problem Statement :** How can mission data obtained from the SpaceX API and Wikipedia be analyzed to identify what influences Falcon 9 landing success, and can machine learning accurately classify whether the booster will land?
- Objectives:
  - Extract and combine Falcon 9 launch data from SpaceX API and Wikipedia.
  - Analyze launch success trends and booster performance.
  - Study the impact of payload, orbit, and launch site on landing outcomes.
  - Build a machine learning model to classify landing success vs failure.
  - Identify the best-performing classifier using hyperparameter tuning.

Methodologies

- Data Collection :
  - Called the SpaceX API using Python (requests module).
  - Scraped Wikipedia Falcon 9 missions using BeautifulSoup and pandas.read_html().
  - Cleaned inconsistent orbit labels, missing payloads, and booster version variations.
  - Merged API and scraped tables using flight number.
- Exploratory Data Analysis (EDA) :
  - Launch success distribution
  - Payload mass distribution
  - Booster version vs landing outcome
  - Launch site performance
  - Folium maps to visualize launch coordinates and success markers
  - Distance computation for geographic analysis

Data Wrangling

Data wrangling is a crucial step in preparing the raw datasets obtained from the SpaceX REST API and Wikipedia web-scraping for meaningful analysis. Since the collected data came from multiple sources with varying formats, several preprocessing operations were required to transform the data into a structured and machine-learning–ready format.

Data Collection

Web scraping data from a table in SpaceX's Wikipedia entry and API requests from the SpaceX REST API were used in the data collection process. To obtain comprehensive information about the launches for a more in-depth analysis, we had to employ both of these data collection techniques.

Data Columns are obtained by using SpaceX REST API:

FlightNumber, Date, BoosterVersion, PayloadMass, Orbit, LaunchSite, Outcome, Flights, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude

Data Columns are obtained by using Wikipedia Web Scraping:

Flight No., Launch site, Payload, PayloadMass, Orbit, Customer, Launch outcome, Version Booster, Booster landing, Date, Time

Data Collection - Steps

1. SpaceX Api Request :

Initiate API Request → Fetch Launch data → Convert JSON into DataFrame → Store Data Locally

2. Web Scrapping :

Extract HTML Table → Parse with BeautifulSoup → Convert to DataFrame → Store Data Locally

Data Cleaning

The raw data contained missing values, inconsistencies, and redundant fields. Key cleaning steps included:

- Handling missing numerical values (e.g., payload mass) using mean/median imputation.
- Removing irrelevant text fields not used for modeling.
- Standardizing categorical labels (e.g., "Success", "Failure", “Partial Failure”).
- Converting date columns into proper datetime formats.
- Dropping duplicated rows where scraped data overlapped with API entries.

Feature Engineering

Several new variables were created to support visualization and machine learning tasks:

- **Landing Outcome (Class):** Converted into binary labels (1 = landed successfully, 0 = failed).
- **Booster Version Category:** Extracted major version identifiers from booster names .
- **Launch Site Coordinates:** Added latitude and longitude for mapping.

Exploratory Data Analysis

Key findings from EDA:

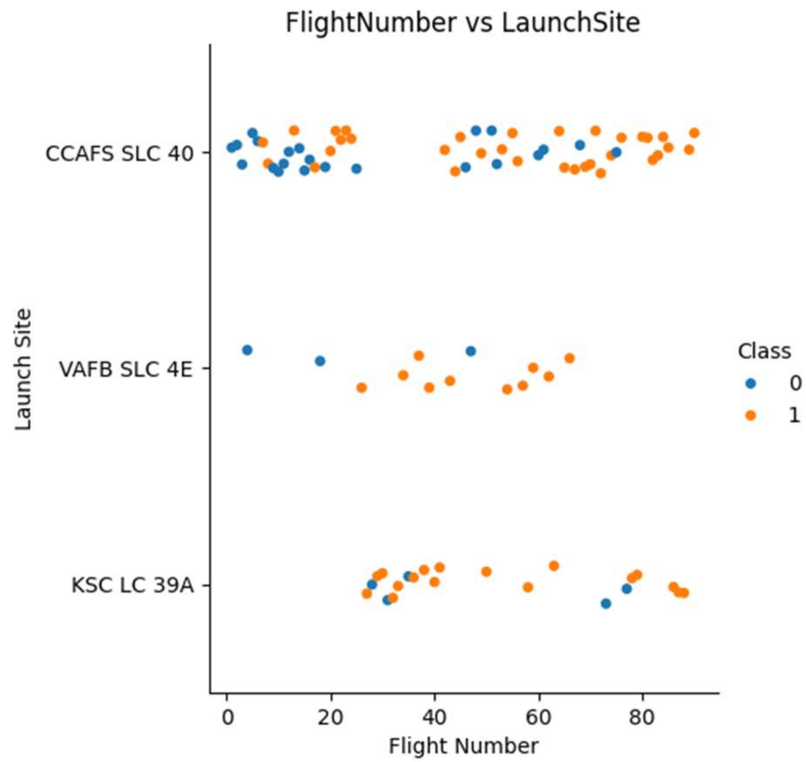
- Launch sites such as KSC LC-39A have the highest success frequency.
- Heavier payloads influence landing difficulty, especially for GTO missions.
- ES-L1, GEO, HEO, SSO missions have significantly higher landing success compared to high-energy orbits.
- Folium maps visualize clusters of successful and failed landings geographically.

EDA with Visualization

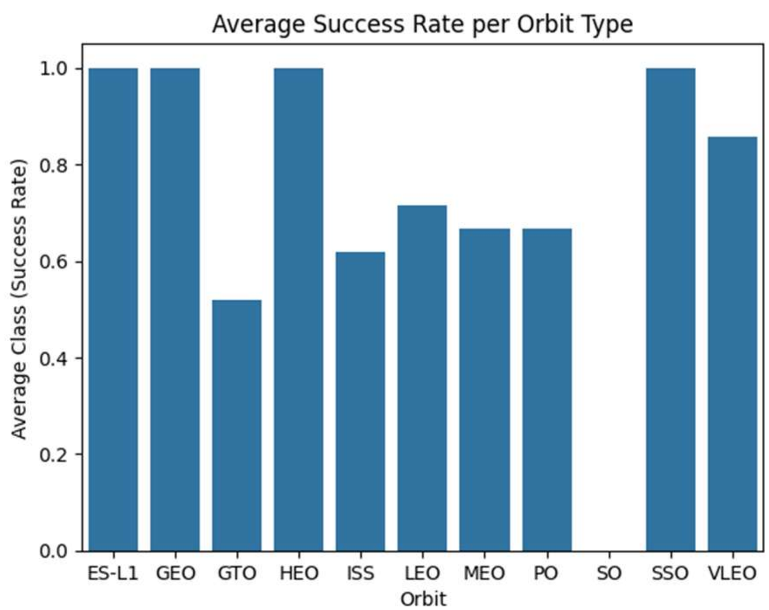
- To understand the behavior, performance, and landing success of Falcon 9 missions, Exploratory Data Analysis (EDA) was conducted on the dataset obtained from the SpaceX API and Wikipedia scraping. The dataset contains mission details such as payload mass, launch site, orbit type, booster version, and landing outcomes.
- The visualizations described below were generated using Matplotlib, Seaborn, and Folium.

Flight Number Vs Launch Site

- As we can see most early flights resulted in **landing failures** (Class = 0 , blue dots).
- Over time The success rate of flights increased in landing (Class = 1, Orange dots).
- The most used launch Site is SLC 40.
- LC 39A has the most success rate all time.



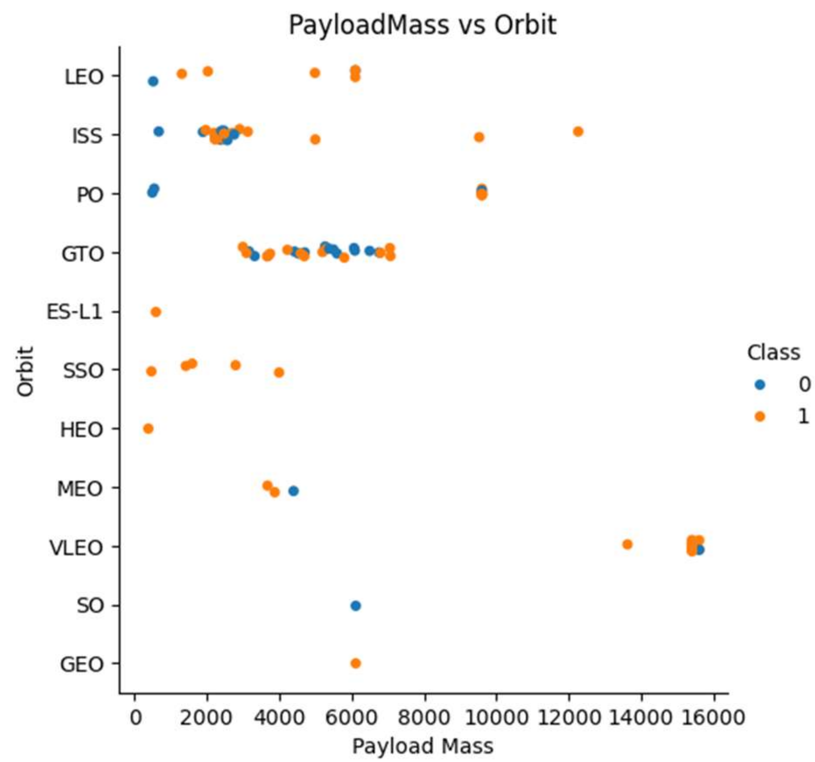
Success Rate Vs Orbit Type



- The ES-L1, GEO, HEO,SSO has the highest success rate of 100% .
- Poor Success rate of landing is seen on SO as only one mission was conducted. Suggesting the mission involves high difficulty and challenging to do.

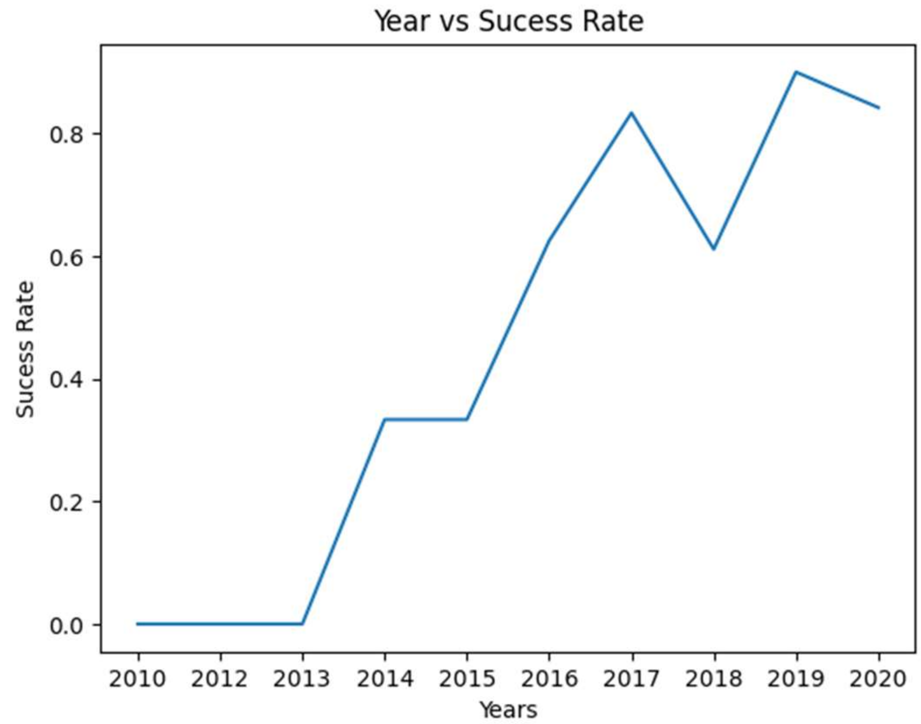
Pay Load mass Vs Orbit

- The effect of pay load mass is very high.
- As the Payload mass and distance increasing the success rate of the mission is decreasing.
- Here at Very Low Earth Orbit (VLEO) has highest payload mass with highest Success rate.



Launch success Yearly Trend

- The Annual launch success rate has shown a significant improvement from 2013 onwards , reaching over 80% by 2020.
- Despite a drop of success rate in 2018, the overall trend indicates increasing success of the Falcon 9.



EDA using SQL

Exploratory Data Analysis (EDA) was performed using SQL to understand the structure, trends, and patterns in the SpaceX Falcon 9 dataset. The SQL-based EDA helped in quickly summarizing launch outcomes, identifying frequently used launch sites, inspecting missing values, and understanding relationships between key variables such as payload mass, orbit type, and landing success.

- Github-link : https://github.com/NMPGowtham/SpaceX-Falcon-9-Analysis/blob/main/Notebooks/jupyter-labs-eda-sql-coursera_sqllite.ipynb

EDA using SQL

All Launch Site Names

```
%sql select distinct Launch_Site from SPACEXTABLE;
```

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

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

Launch Pad names beginning with CCA

```
%sql select * from SPACEXTABLE where Launch_Site like "CCA%" limit 5;
```

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

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

Average payload mass carried by booster version F9 v1.1

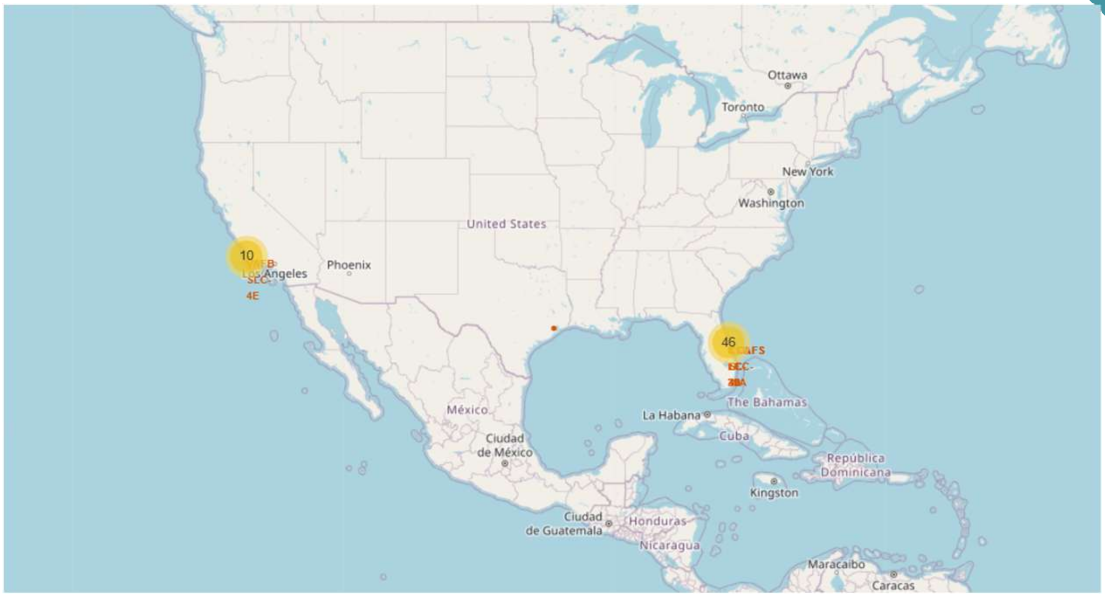
```
%sql select AVG(PAYLOAD_MASS__KG_) from SPACEXTABLE where Booster_Version = "F9 v1.1"
```

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

AVG(PAYLOAD_MASS__KG_)
2928.4

Launch Sites Proximities Analysis

- All the Launch Sites are marked in the map.
- Not All sites are close to equator.
- All the Launch Sites are close to coastal area.
- CCAFS LC-40, CCAFS SLC-40 and KSC LC-39A are near the coast in Florida
- VAFB SLC-4E is near the coast of California
- Github link: https://github.com/NMPGowtham/SpaceX-Falcon-9-Analysis/blob/main/Notebooks/lab_jupyter_launch_site_location.ipynb

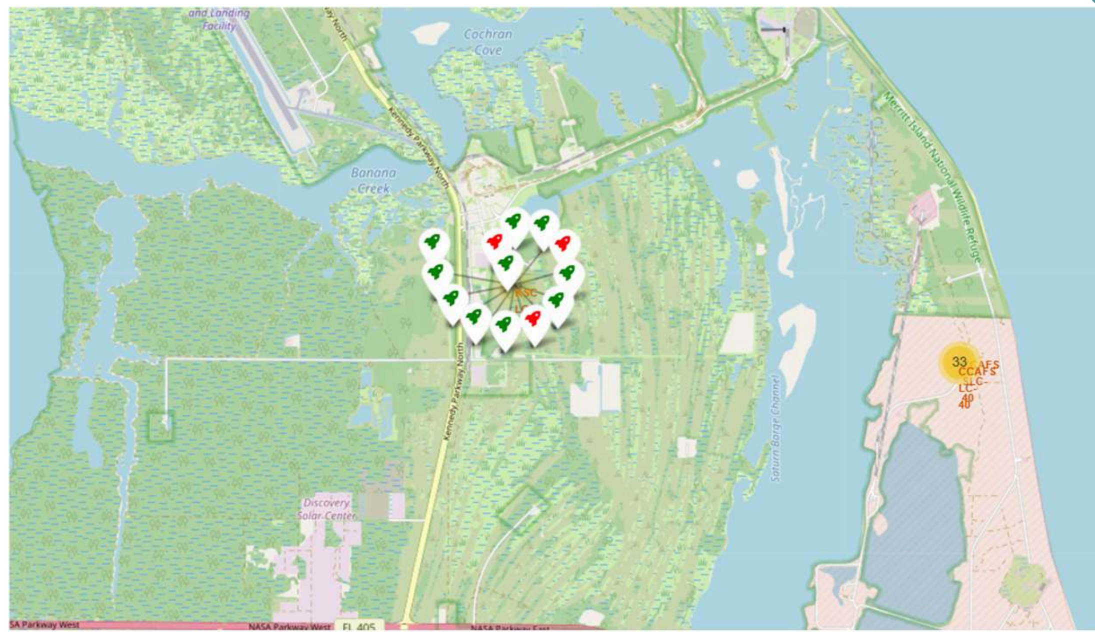


Color-labeled launch Records on the Map

❖ From the color-labeled markers we should be able to easily identify which launch sites have relatively high success rates.

- Green Marker = Successful Launch
- Red Marker = Failed Launch

❖ Launch Site KSC LC-39A has a very high Success Rate



Distance from the launch site CCAFS SLC-40 to its proximities

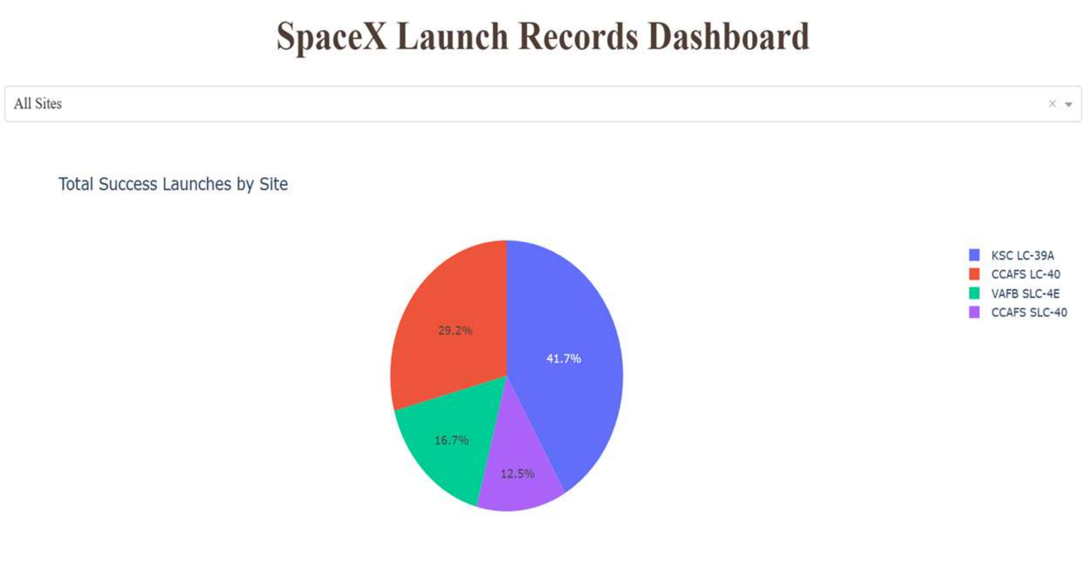
- From the Map we can see distance between CCAFS SLC-40 Launch Site and Florida's coastline.
- The Calculated Distance is about 0.90KM towards East.
- This Close proximity to the coastline is typical for launch sites to facilitate safe recovery operations, ensuring minimal risk to populated areas.

Building a Dashboard using Dash by Plotly

- A fully interactive dashboard was developed using Plotly Dash to visualize SpaceX Falcon 9 launch performance, landing outcomes, and mission characteristics. Dash was chosen because it allows the creation of responsive, web-based dashboards entirely in Python without requiring frontend frameworks.
- The dashboard serves as the central interface for exploring launch locations, payload effects, and mission success trends in a user-friendly manner.
- Github link : https://github.com/NMPGowtham/SpaceX-Falcon-9-Analysis/blob/main/Dashboard/spacex_dash_app.py

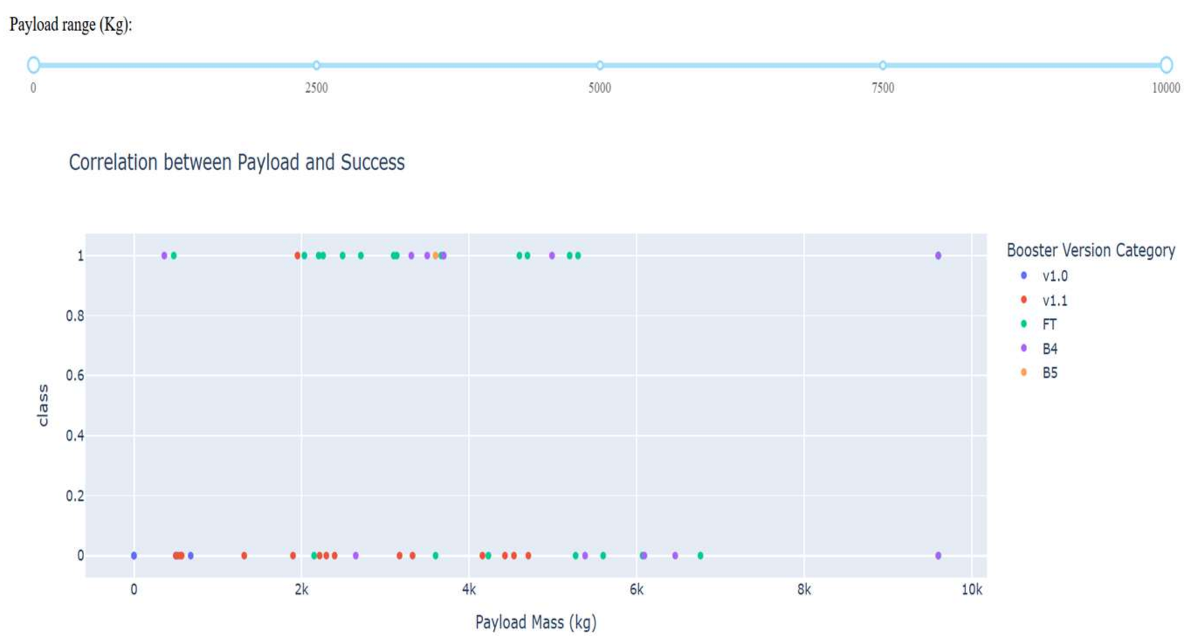
Launch Success rate for All Sites

- Key Findings :
  - KSC LC-39A : 41.7%
  - CCAFS LC-40 : 29.2%
  - VAFB SLC-4E : 16.7%
  - CCAFS SLC-40 : 12.5%
- The KSC LC-39A has the highest success rate make it more reliable site for Launching.



Booster Version Performance

- Booster Version "FT" appears to be the most frequently used and has a high success rate across various payload masses.
- Booster version "v1.0" has fewer launches and may required further analysis to understand its performance.
- Overall, booster versions do not show a clear trend that higher payload masses correlated with lower success rate.



Predictive Analysis

The goal of predictive analysis in this study is to build a machine learning model that can **classify whether a Falcon 9 mission will successfully land** (Class = 1) or result in a failure (Class = 0). By analyzing historical mission features such as payload mass, orbit type, launch site, and booster characteristics, the model can estimate future landing outcomes.

Problem Definition

- This is a **binary classification problem,** where the target variable is:

    1. Class = 1 → Successful Landing
    2. Class = 0 → Failed Landing

Predictive Analysis (Feature used)

- Numerical Features:
  - FlightNumber
  - PayloadMass
  - Block
  - ReusedCount
  - Flights
  - GridFins (0/1)
  - Legs (0/1)
  - Reused (0/1)

- Categorical Features (OneHotEncoded):
  - Orbit
  - LaunchSite
  - LandingPad
  - BoosterVersion
- Target Feature:
  - Class

Model Implementation

Multiple classification models were built and evaluated:

Models Used:

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

Comparison of Model Performance

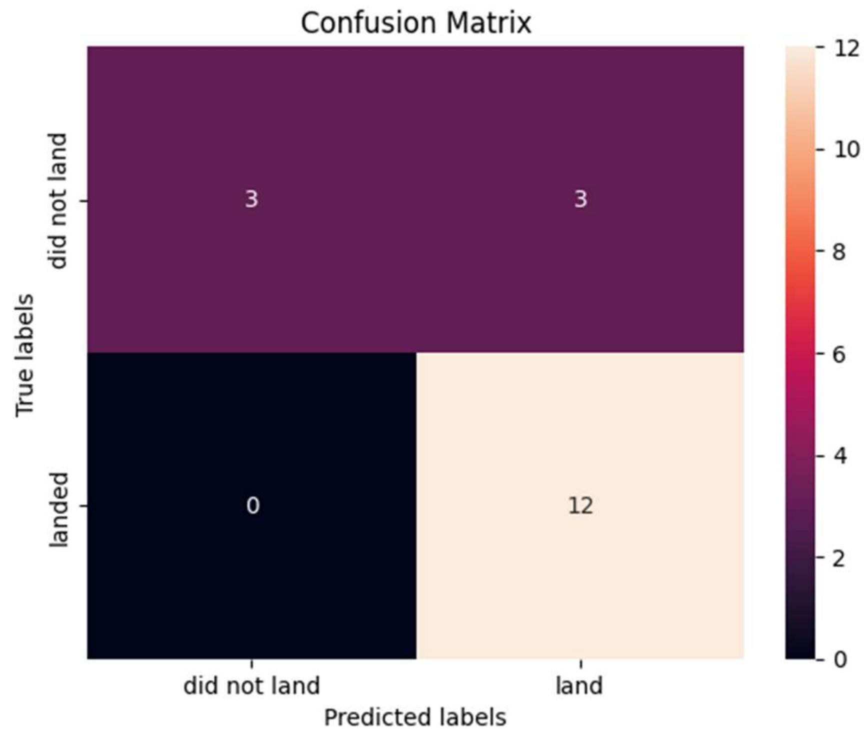
❖ An accuracy of 83.33% reflects that the models can correctly classify 5 out of 6 landing outcomes, making them reasonably reliable given the size and nature of the dataset. Further improvements could be achieved through feature engineering, hyperparameter tuning, and expanding the dataset with more historical launches.

Model	Accuracy	Best Hyperparameters
Logistic Regression	0.83	Lbfgs solver
SVM	0.83	Sigmoid Kernel
Decision Tree	0.83	Max depth 10
KNN	0.83	N neighbors 10

Confusion matrix

A confusion matrix is a fundamental evaluation tool used to assess the performance of binary classification models, such as the machine learning models built in this project to predict whether a Falcon 9 booster successfully lands (Class = 1) or fails to land (Class = 0).

The confusion matrix summarizes model predictions by comparing actual outcomes with predicted outcomes, providing deeper insights beyond simple accuracy.



Conclusion

This project provides an end-to-end analytical study of SpaceX Falcon 9 launches, focusing on the factors influencing booster landing success and failure. By collecting data from the SpaceX API and Wikipedia scraping, followed by extensive data wrangling, SQL-based EDA, visualization, and machine learning modeling, the report offers a comprehensive exploration of launch performance trends.

The exploratory data analysis revealed clear patterns:

- Landing success rates steadily increased over the years as booster design, reuse capability, and engineering refinements improved.
- Specific launch sites, especially KSC LC-39A and CCAFS SLC-40, showed higher success probabilities.
- Technical attributes such as GridFins, Legs, Block version, and booster reuse count played a significant role in determining landing outcomes.

Thank you