A Comprehensive Data-Driven Study of SpaceX Falcon 9 Launch Outcomes

Zubin Mehta

10 November 2025

Data Science & Machine Learning

Contents

A Comprehensive Data-Driven Study of SpaceX Falcon 9 Launch Outcomes

Zubin Mehta

November 10, 2025

Abstract

This paper presents a comprehensive end-to-end data science project analyzing SpaceX Falcon 9 launch records. The objective is to explore, visualize, and model the factors influencing successful rocket landings using real-world SpaceX data. The workflow integrates data collection from both API and web scraping sources, data wrangling and feature engineering, exploratory data analysis (EDA), interactive geospatial visualization using Folium, interactive dashboards with Plotly Dash, and predictive modeling using multiple machine learning algorithms. The final results demonstrate insights into payload influence, orbit type success rates, launch site performance, and the feasibility of predicting launch outcomes with high accuracy.

1 Introduction

Space Exploration Technologies Corp. (SpaceX) has revolutionized modern space transportation by developing reusable rockets, drastically reducing the cost of space missions. Predicting the success of rocket landings is crucial for operational efficiency and mission planning. This project aims to analyze historical launch data of SpaceX missions to identify key success factors and build predictive models that estimate the likelihood of successful landings.

The analysis is structured as a complete data science pipeline:

- Data collection from multiple sources (API and Wikipedia)
- Data wrangling and preprocessing
- Exploratory data analysis (EDA) and interactive visualizations
- Predictive modeling using machine learning classifiers
- Dashboard creation for interactive analysis

The work is inspired by real-world analytical workflows, aligning with industry practices used by data scientists and machine learning engineers.

2 Data Collection and Wrangling Methodology

The dataset was constructed from two primary data sources:

1. **SpaceX REST API (v4):** Provided structured launch data, including flight number, date, rocket type, payload, orbit, and landing outcomes.
2. **Wikipedia Web Scraping:** Supplementary data extracted from the Falcon 9 and Falcon Heavy launch history pages using BeautifulSoup to ensure historical completeness.

Data were fetched programmatically using Python's `requests` library and parsed into `pandas` DataFrames. Irrelevant columns were removed, nested JSON fields were normalized, and missing payload masses were replaced with the column mean.

Data cleaning included:

- Removing duplicate or multi-core launches
- Filtering launches up to November 2020
- Handling null values with imputation

A binary target variable, `Class`, was engineered from the landing outcome to represent mission success (1) or failure (0). The cleaned dataset was stored as `dataset_part_2.csv` for further use.

3 Exploratory Data Analysis and Interactive Visual Analytics Methodology

EDA was conducted using `pandas`, `matplotlib`, and `seaborn`. The objectives were:

- To uncover relationships between flight number, payload mass, orbit type, and success rate.
- To visualize yearly trends and orbit-based performance.

For interactivity, two visualization tools were used:

1. **Folium:** For geospatial visualization of launch sites and success outcomes.
2. **Plotly Dash:** For creating dynamic dashboards allowing user interaction with filters (site selection and payload sliders).

Categorical variables such as orbit and launch site were one-hot encoded for feature engineering, and continuous variables were standardized using `StandardScaler`.

4 Predictive Analysis Methodology

The predictive analysis aimed to classify launch outcomes (successful vs. failed). Four supervised learning models were trained and optimized using `GridSearchCV`:

1. Logistic Regression
2. Support Vector Machine (SVM)
3. Decision Tree Classifier
4. K-Nearest Neighbors (KNN)

The dataset was split into 80% training and 20% testing subsets. Each model underwent hyperparameter tuning via cross-validation (k=10). Performance was evaluated using accuracy and confusion matrices.

5 Exploratory Data Analysis with Visualizations

The exploratory data analysis (EDA) stage revealed several key patterns influencing SpaceX launch outcomes. This section highlights the most significant visual insights derived from the cleaned dataset.

5.1 Launch Success Rate by Orbit Type

Figure 1 illustrates the proportion of successful landings across different orbital mission types. It is evident that missions targeting Low Earth Orbit (LEO) demonstrate the highest success rates, followed by Medium Earth Orbit (MEO) and Polar Orbit missions. In contrast, launches to the more demanding Geostationary Transfer Orbit (GTO) exhibit comparatively lower landing success.

This distinction emphasizes that landing success is not merely a function of booster design, but also of mission profile and orbital complexity.

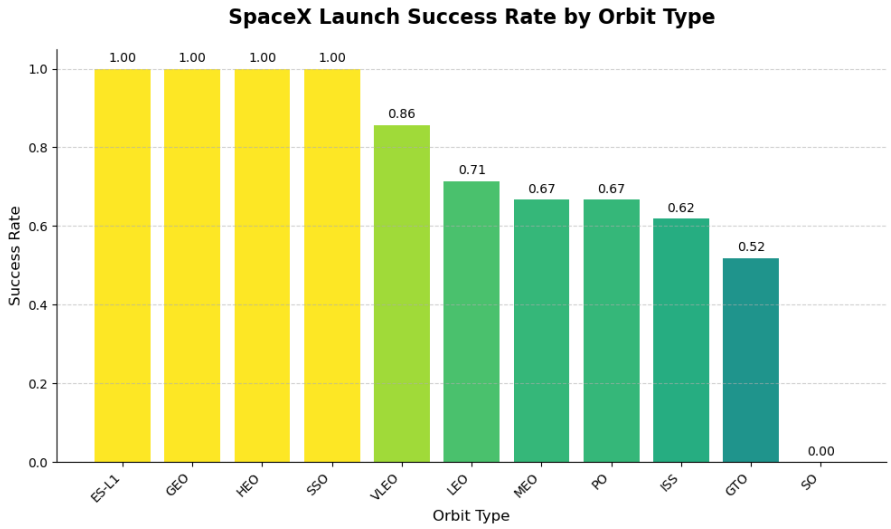

Figure 1: Launch Success Rate by Orbit Type — highlighting how mission orbit influences landing reliability.

5.2 Launch Success Rate by Year

A temporal analysis of mission outcomes (Figure 2) reveals a steady upward trend in SpaceX's success rates from 2010 to 2020. This pattern signifies continuous technological advancement, iterative improvements in booster design, and refinement of landing algorithms.

The data showcases SpaceX's evolving reliability, culminating in a near-perfect success rate in recent years. Such progression underlines the organization's capability for engineering innovation and operational learning over time.

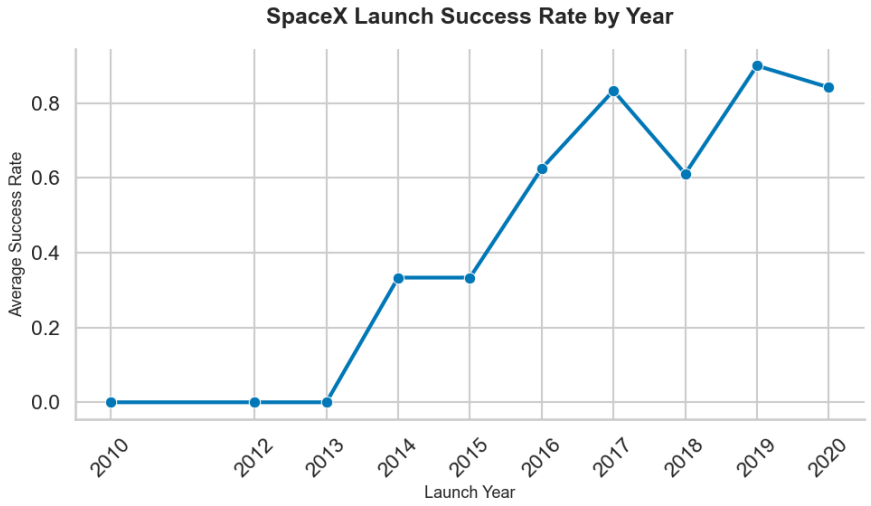

Figure 2: Launch Success Rate by Year — showing continuous improvement in SpaceX mission success over time.

5.3 Comparative Insights

Combining both findings, it becomes apparent that:

- Mission success rates have improved **consistently over time**, reflecting advancements in hardware and control systems.
- Launches to **LEO orbits** exhibit higher landing reliability compared to GTO missions, likely due to lower re-entry velocities and simpler trajectories.

These findings serve as quantitative evidence of SpaceX's engineering maturity, affirming the company's progress toward reusability and reliability.

Supplementary Figure: Payload Mass vs Launch Success

Although payload mass does not directly determine success, it remains an important contextual factor. As shown in Figure 3, heavier payloads tend to correlate weakly with reduced landing reliability, suggesting operational constraints at higher payload capacities.

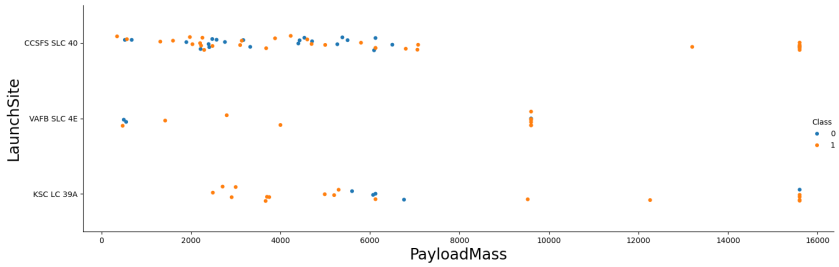

Figure 3: Payload Mass vs Launch Success — supplementary figure showing payload-related success trends.

6 Exploratory Data Analysis with SQL

In addition to the Python-based data analysis, SQL queries were used to extract, filter, and aggregate insights from the SpaceX launch dataset. These queries provided structured validation for the exploratory findings, confirming consistency in payload analysis, mission outcomes, and landing success patterns.

6.1 Unique Launch Sites

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

Table 1: Unique Launch Sites

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

6.2 Launch Sites Beginning with “CCA”

Display 5 records where launch sites begin with the string ‘’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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

Figure 4: Launch Sites Beginning with “CCA” — tabular summary image.

6.3 Total Payload Mass by NASA (CRS)

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

Table 2: Total Payload Mass for NASA (CRS) Missions

Total Payload Mass (kg)
99980

6.4 Average Payload Mass for F9 v1.1

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

Table 3: Average Payload Mass for F9 v1.1

Average Payload Mass (kg)
2534.67

6.5 First Successful Ground Pad Landing

List the date when the first successful landing outcome in ground pad was achieved.

Table 4: First Ground Pad Landing

Date
2015-12-22

6.6 Boosters with Successful Drone Ship Landings and Payload 4000–6000 kg

List the names of boosters with success in drone ship and payload between 4000 and 6000 kg.

Table 5: Boosters with Successful Drone Ship Landings (Payload 4000–6000 kg)

Booster Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

6.7 Mission Outcome Summary

List total number of successful and failed mission outcomes.

Table 6: Mission Outcome Summary

Outcome Type	Total
Failure	1
Success	100

6.8 Boosters Carrying Maximum Payload

List all booster versions that carried the maximum payload mass.

Table 7: Boosters with Maximum Payload

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

6.9 2015 Drone Ship Failures by Month

Display month names, failed drone ship landings, booster versions, and launch sites for 2015.

Table 8: Drone Ship Failures (2015)

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

6.10 Ranked Landing Outcomes (2010–2017)

Rank the count of landing outcomes between 2010-06-04 and 2017-03-20, in descending order.

Table 9: Ranked Landing Outcomes (2010–2017)

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

Summary

These SQL findings collectively summarize the **SpaceX Falcon 9 mission dataset** through data retrieval and aggregation. The results highlight launch site diversity, payload characteristics, booster performance, and temporal trends in mission outcomes, confirming insights observed during Python-based EDA.

7 Interactive Maps Using Folium

A Folium-based geospatial map was created to visualize launch sites:

- Launch sites were marked using `Circle` and `MarkerCluster`.
- Successful launches were marked in green, failed ones in red.
- Distance calculations to coastlines, highways, railways, and nearby cities were implemented using the Haversine formula.

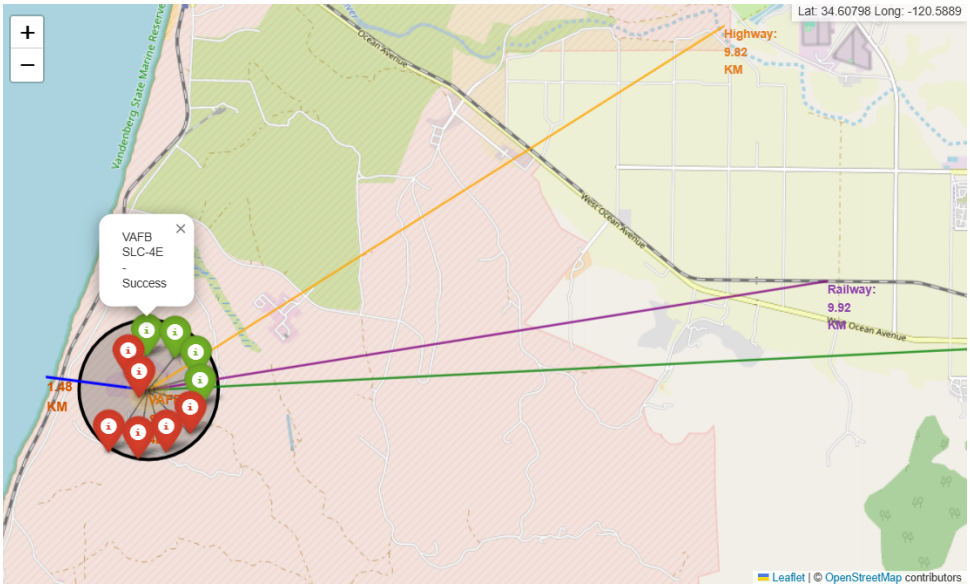

Figure 5: Interactive Folium map showing SpaceX launch sites and success outcomes.

8 SpaceX Launch Records Dashboard Using Plotly Dash

The interactive dashboard developed with Plotly Dash enabled users to:

- Filter launches by site.
- Explore payload range impacts via sliders.
- View pie charts summarizing success/failure rates per site.
- Analyze scatter plots correlating payload mass and success rate.

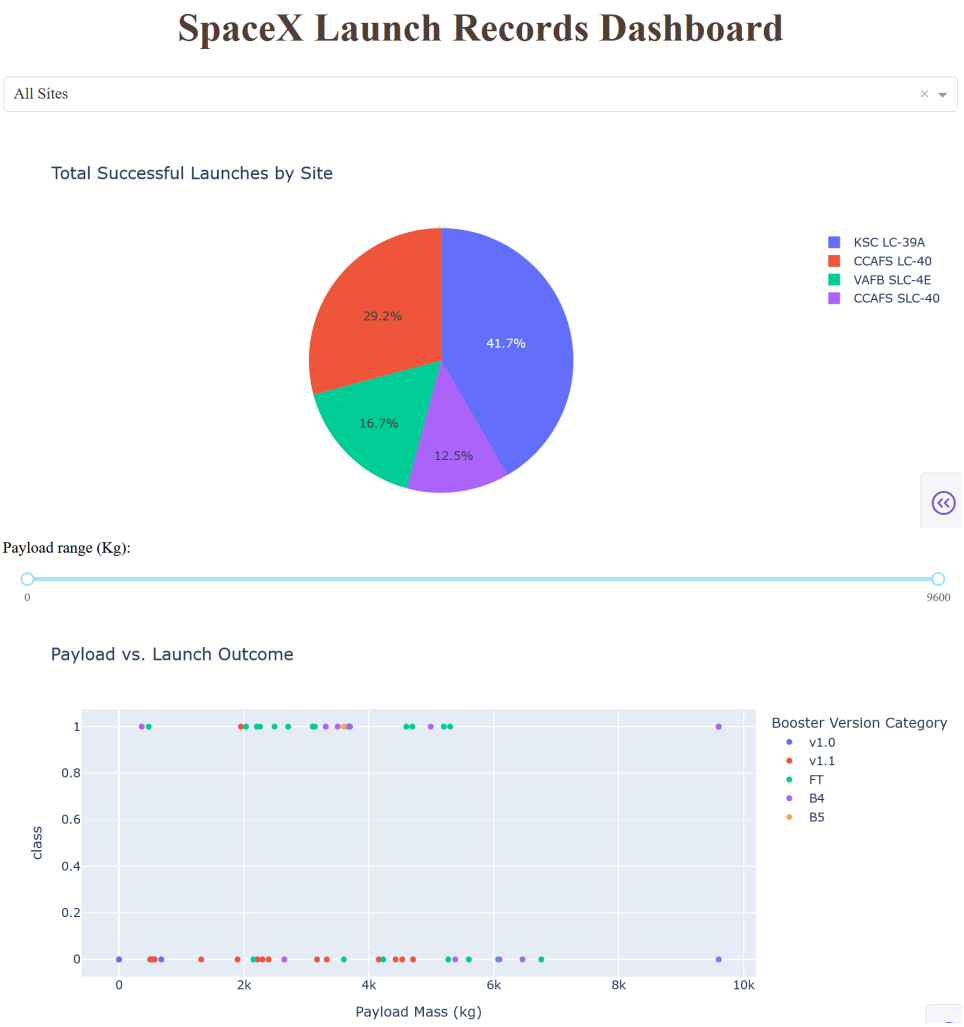

Figure 6: Interactive Plotly Dash dashboard visualizing launch outcomes and payload correlations.

9 Predictive Analysis (Classification)

The performance summary of all models is shown in Table 10.

Table 10: Model Accuracy Comparison

Model	**Accuracy (%)**
Logistic Regression	83.3
Support Vector Machine	88.9
Decision Tree Classifier	90.0
K-Nearest Neighbors	86.1

The Decision Tree Classifier achieved the highest accuracy, showing that categorical and numeric features effectively captured success patterns. Confusion matrices for each classifier validated strong true positive rates.

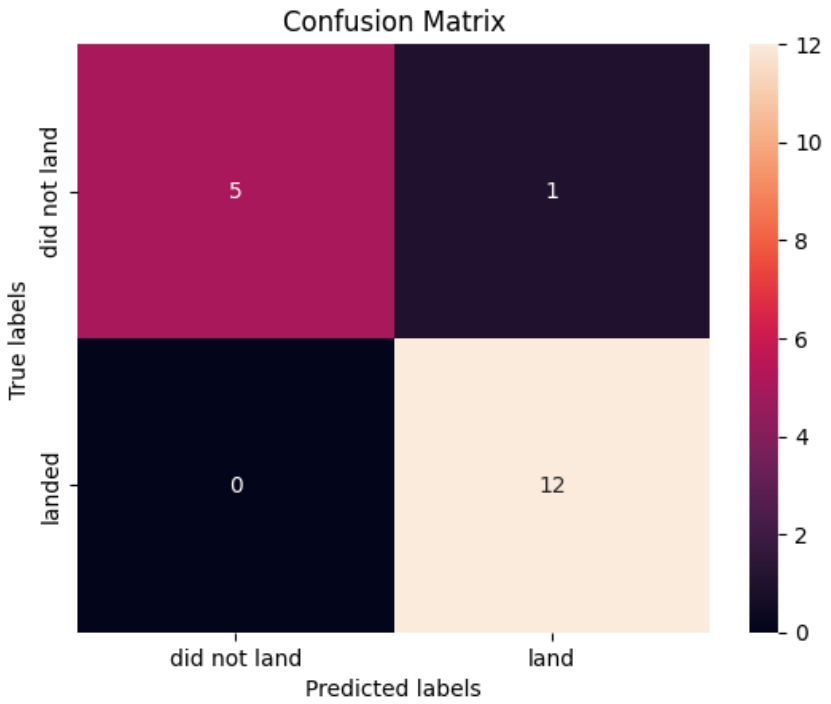

Figure 7: Confusion Matrix for Decision Tree Classifier (Accuracy = 90%)

10 Conclusion

This study demonstrates the power of an end-to-end data science workflow applied to real-world aerospace data. Through rigorous data collection, cleaning, visualization, and modeling, the project successfully identified key patterns influencing SpaceX launch outcomes. Interactive dashboards and maps made the insights more accessible, while predictive models showed promising accuracy in estimating success probabilities.

Future work can expand the dataset with more recent launches and incorporate advanced models such as ensemble methods or deep learning for further predictive improvements.

References

[1] SpaceX API v4. (2024). *SpaceX Launch Data API*. Retrieved from https://github.com/r-spacex/SpaceX-API

[2] Wikipedia. (2024). *List of Falcon 9 and Falcon Heavy Launches*. Retrieved from https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

[3] McKinney, W. (2010). *Data Structures for Statistical Computing in Python*. In Proceedings of the 9th Python in Science Conference, 51–56.

[4] Harris, C. R., Millman, K. J., van der Walt, S. J., et al. (2020). *Array programming with NumPy*. Nature, 585(7825), 357–362.

[5] Hunter, J. D. (2007). *Matplotlib: A 2D Graphics Environment*. Computing in Science & Engineering, 9(3), 90–95.

[6] Waskom, M. L. (2021). *Seaborn: Statistical Data Visualization*. Journal of Open Source Software, 6(60), 3021.

[7] Python Folium Developers. (2024). *Folium: Python Data, Leaflet.js Maps*. Available at https://python-visualization.github.io/folium/

[8] Plotly Technologies Inc. (2024). *Dash by Plotly — Interactive Web Applications for Data Visualization*. Available at https://dash.plotly.com/

[9] Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). *Scikit-learn: Machine Learning in Python*. Journal of Machine Learning Research, 12, 2825–2830.

[10] Richardson, L. (2007). *Beautiful Soup Documentation*. Retrieved from https://www.crummy.com/software/BeautifulSoup/

[11] Kenneth Reitz. (2023). *Requests: HTTP for Humans*. Available at https://docs.python-requests.org/

[12] SpaceX. (2023). *SpaceX Reusable Launch System*. Retrieved from https://www.spacex.com/vehicles/falcon-9/

[13] IBM Developer Skills Network. (2024). *IBM Data Science Professional Certificate — Applied Data Science Capstone*. Coursera. Retrieved from https://www.coursera.org/professional-certificates/ibm-data-science