# Winning the Space Race with Data Science

Predicting Falcon 9 First Stage Landings

Shamsaddin Amanov

1 January 2026

# Executive Summary

**Goal:** Predict successful landings of the Falcon 9 first stage to estimate launch costs for SpaceY (a competitor).

**Approach:** Utilized the full Data Science lifecycle: API collection, SQL analysis, Interactive Dashboards, and Machine Learning.

**Key Result:** We successfully built a **Decision Tree** model that predicts landing outcomes with **88.9% accuracy**.

# Project Background

SpaceX saves millions by reusing the first stage of its rockets.

A successful landing equals a lower launch cost (~$62M vs $165M).

**Problem Statement:** Can we use historical launch data to predict if a rocket will land successfully before it even launches?

# Methodology

**Data Collection:** SpaceX REST API & Wikipedia Scraping.

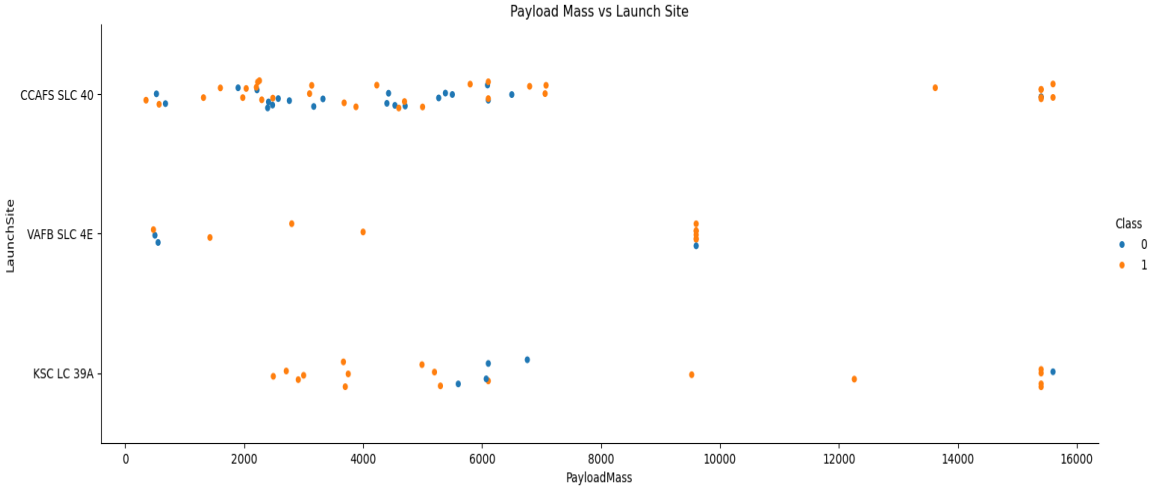
**Data Wrangling:** Pandas for cleaning (handling missing payload data).

**Exploratory Data Analysis (EDA):** SQL for querying and Seaborn for visualization.

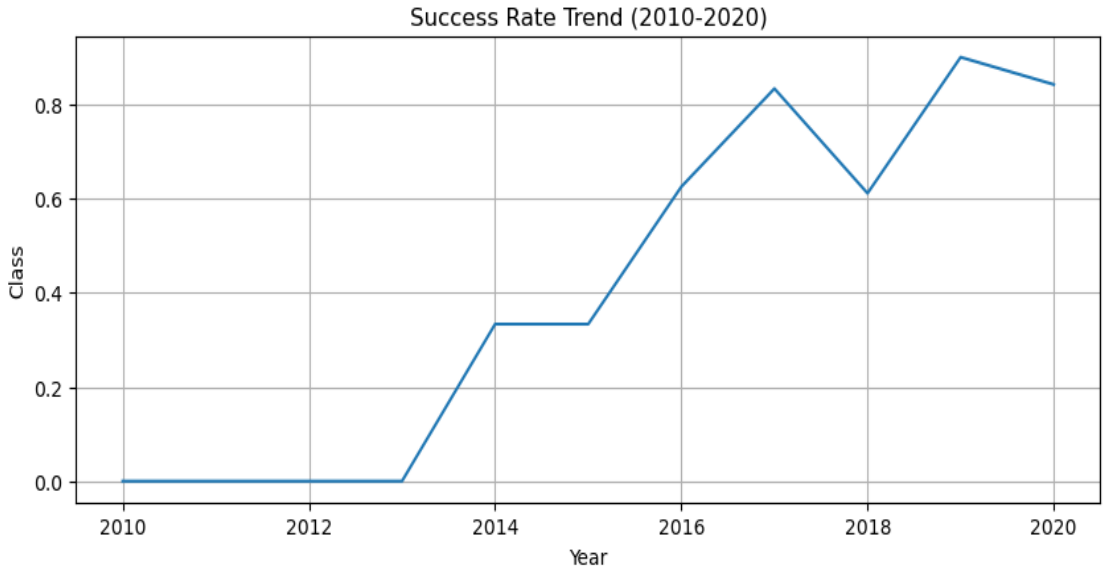
**Interactive Analytics:** Folium for geospatial maps and Plotly Dash for user-driven dashboards.

**Predictive Modeling:** Logistic Regression, SVM, Decision Tree, and KNN.

# EDA with SQL

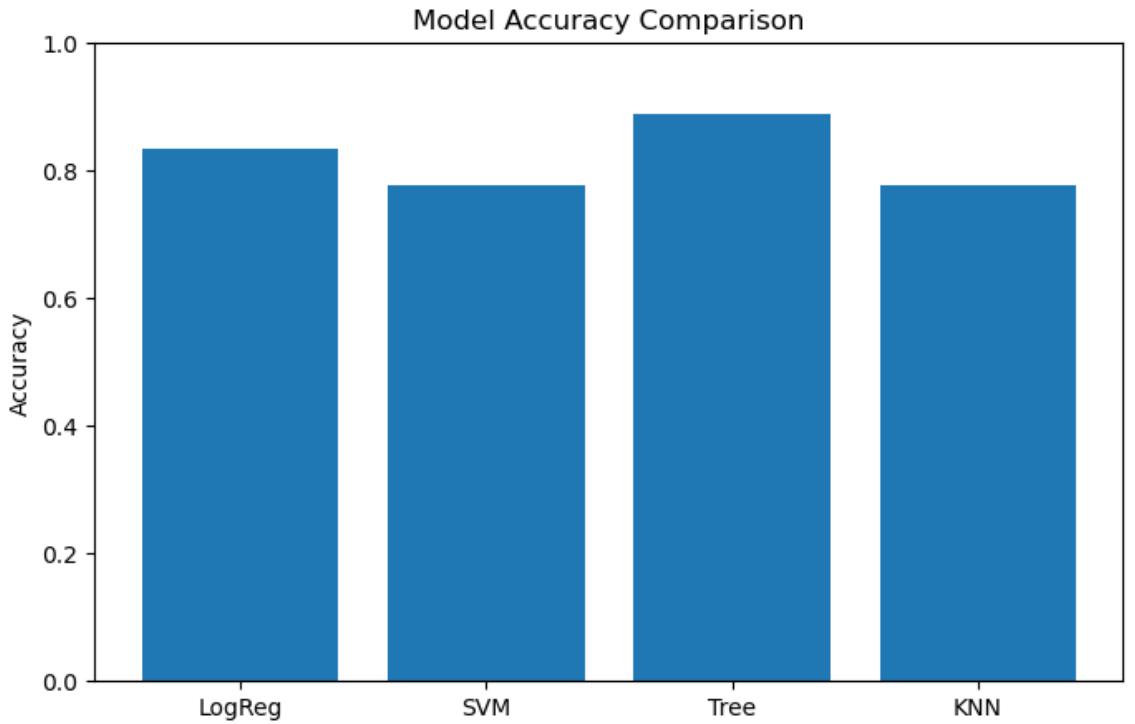




- SQL Analysis: Landing Trends
  - Querying the database revealed that "No Attempt" was common in early years.
  - **Insight:** The success rate of "Drone Ship" landings has significantly increased as technology matured.
  - Success rates (indicated by orange dots) clearly increase as the Flight Number increases.
  - **Insight:** The KSC LC-39A launch site shows the highest density of successful launches.



# Predictive Analysis Results

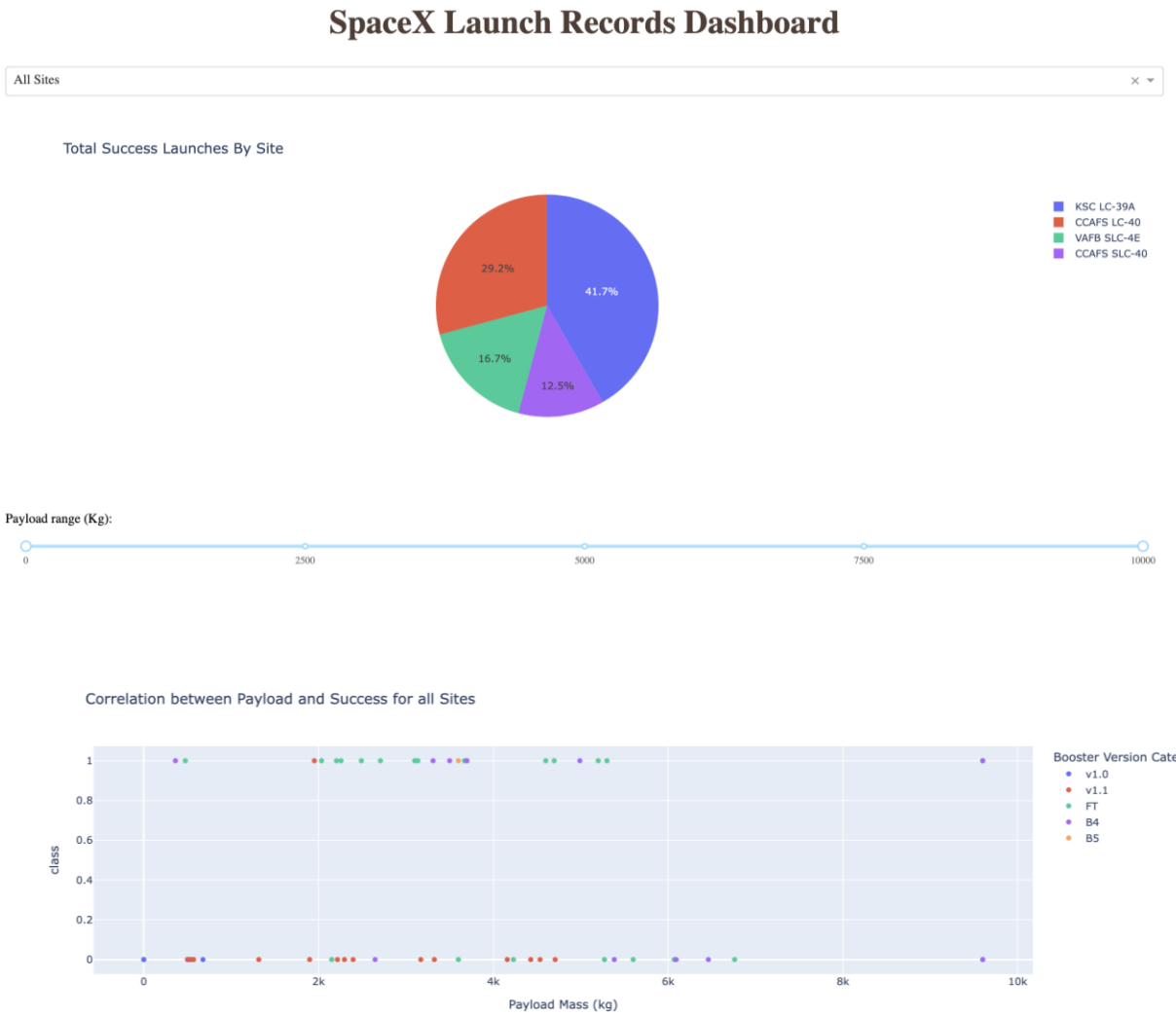
- We tested four classification models: Logistic Regression, SVM, Decision Tree, and KNN.
- **Winner:** The **Decision Tree** model achieved the highest accuracy (**88.9%**) on the test data.
- All models performed well, confirming the data contains strong predictive signals.



# Dashboard

- Dynamic Launch Dashboard

- Built a dashboard to allow stakeholders to filter performance by Site and Payload.
- **Insight:** The Pie Chart updates dynamically to show that KSC LC-39A has the highest success rate among all sites.



# Conclusion

- **Conclusion:** We can confidently predict landing success. The dataset shows a clear trend of improvement over time.
- **Recommendation:** SpaceY should focus its competitive bids on missions with Payloads < 5,000kg, where SpaceX has the most consistent success.
- **Future Work:** Incorporate weather data (wind speed/humidity) to further improve model precision.