SPACEX FALCON-9 ANALYSIS

Khushi Murthy | IBM’s Data Science Capstone Project | June 2024
[GitHub Link]

Executive Summary:

This is a guided project for the IBM Data Science capstone, which predicts if the SpaceX Falcon 9 first stage will properly land. To conduct this study, I used 6 main steps: Data Collection, Data Wrangling, Data Formatting, Data Analysis, Data Visualization, and Machine Learning Predictions.

In this report, I present graphs of success rate vs. various parameters (orbit type, payload, year), and compare different machine learning algorithms to optimize accuracy.

Introduction:

In this capstone project, our goal is to predict the successful landing of the Falcon 9 first stage. SpaceX offers Falcon 9 rocket launches for $62 million, whereas other providers charge upwards of $165 million per launch. The significant cost savings stem from SpaceX's ability to reuse the first stage. By predicting the likelihood of a successful first-stage landing, we can estimate launch costs more accurately.

The aim of this project is to accurately predict whether a launch will land based on different parameters. This data analysis will immensely help opposing companies in the launch market to decide whether to bid against the Falcon 9 launch.

Methodology:

Libraries / Modules used throughout the project: pandas, numpy , datetime, sys, requests, BeautifulSoup, imb_db, seaboard, matplot, folium, dash, dash_html_components, plotly.express, sklearn

Data Collection, Wrangling, and Formatting:

The Data Collection stage includes 3 parts: requesting the SpaceX API, filtering the data set to only include data we want to use, and web scraping from Wikipedia. To do this we imported the Pandas and NumPy libraries, and created helper functions to append launchpads, payloads, and core data. Next, we filtered the data to only include flight number, date, payload mass, orbit type, launch site, outcome, for falcon 9 launches.

The web scraping portion aimed to extract information from Wikipedia using Beautiful Soup objects. I first requested the data, then extracted column and variables, and lastly created a data frame by parsing launch HTML tables.

EDA with Data Visualization:

The main goal of this EDA section is to prepare data feature engineering. Using pandas reading functions, I read the SpaceX dataset and printed a summary. I then plotted different graphs to visualize effects of different parameters: Launch Site vs. Flight Number, Launch Site vs. Payload Mass, Success Rate vs. Orbit, and success for yearly trend.

To prepare for features engineering, I selected the features that will be used in success prediction in the future module.

EDA with SQL:

To more easily analyze the data set, I uploaded the dataset into the IBM cloud using DB2 UI. Using SQL queries, I analyzed the data. I found the amount of unique launch sites, total payload mass carried by boosters, average mass, and number of different types of landings. The SQL statements I used included LIMIT, LIKE, AS, FROM, WHERE, MIN(), BETWEEN, YEAR(), COUNT(), SUM(), AVG(), and SELECT.

Building an Interactive Map with Folium

To create a visual representation of the launch sites and successes, I used the folium library. First, I marked all launch sites on a map. To do this, I downloaded the dataset, selected relevant sub-columns, and set an initial start location as the NASA Johnson Space Center at Houston, Texas. I then used the folium.circle to create a highlighted circle on specific coordinates and launch sites. I used MarkerCluster objects to cluster data points to signify a place with multiple launches. To allow a hovering mouse to display coordinates, I used MousePosition to get the coordinate (Lat, Long) for a mouse over on the map. Using different NumPy functions, I found the distance each launch site was from highways, railroads, and cities, to see where most launch sites are.

Build a Dashboard with Plotly Dash

Using Python and HTML, I created a dashboard application that contains input components to display data in a user-friendly way. I implemented a dropdown menu to enable launch site selection, added callback functions to create a success pie chart and a scatter plot.

Results / Analysis:

Dashboard Displays:

Figure A: This is a display of the main page of the dashboard. It shows a pie chart of different types of launches. You can control the site location using the drop down menu as seen in Figure B.

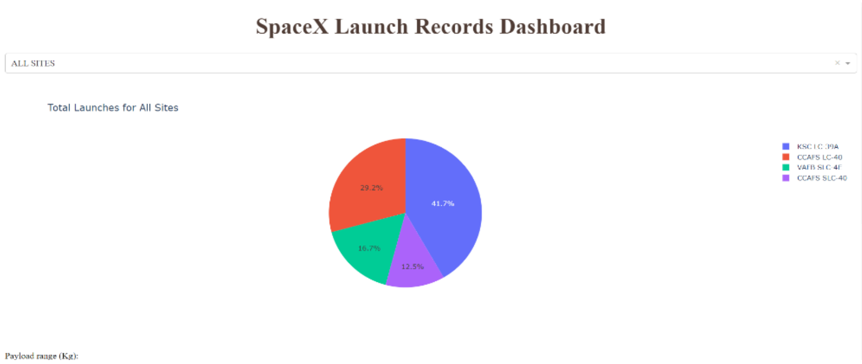


Figure B:

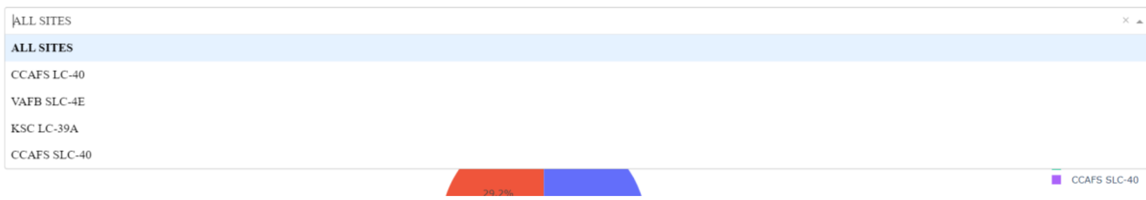
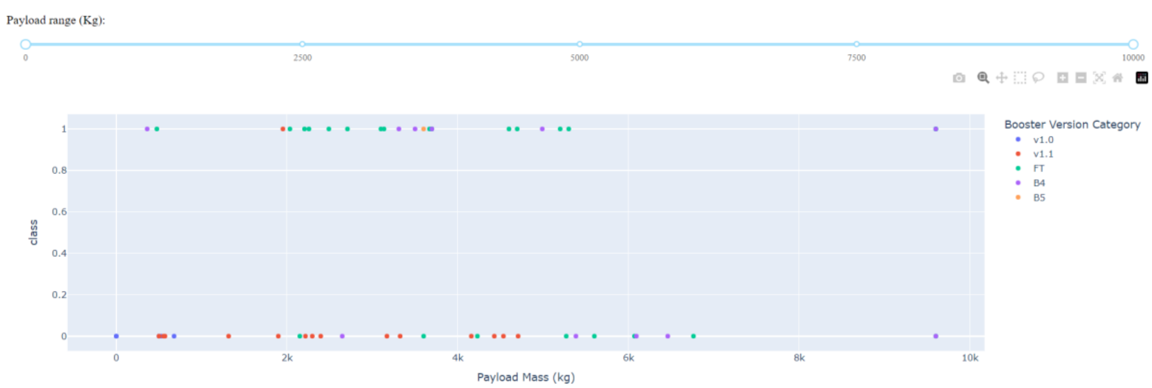


Figure C: Shows the Class Type vs Payload Mass for each Booster Version.



Insights Drawn From EDA:

The following attachments are the graphs described earlier.

Figure D: Payload Mass vs. Flight Number

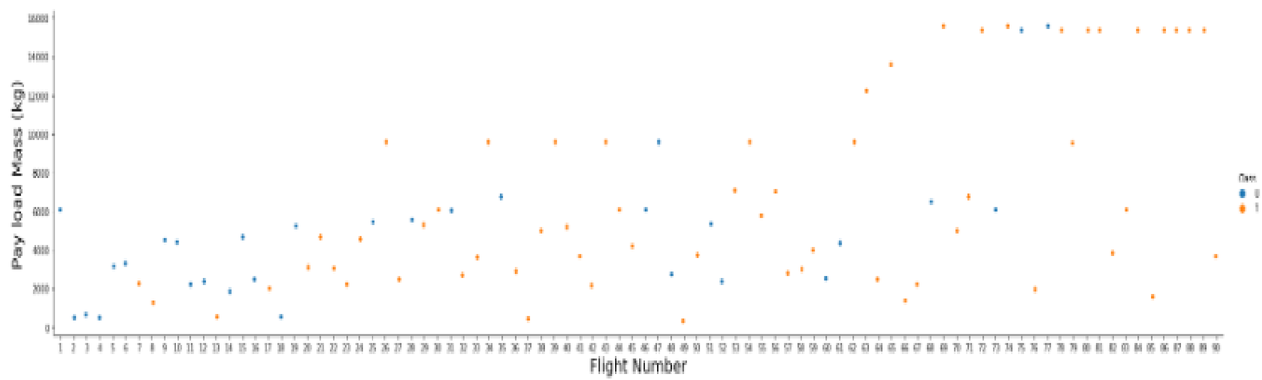


Figure E: Payload Mass vs. Flight Number

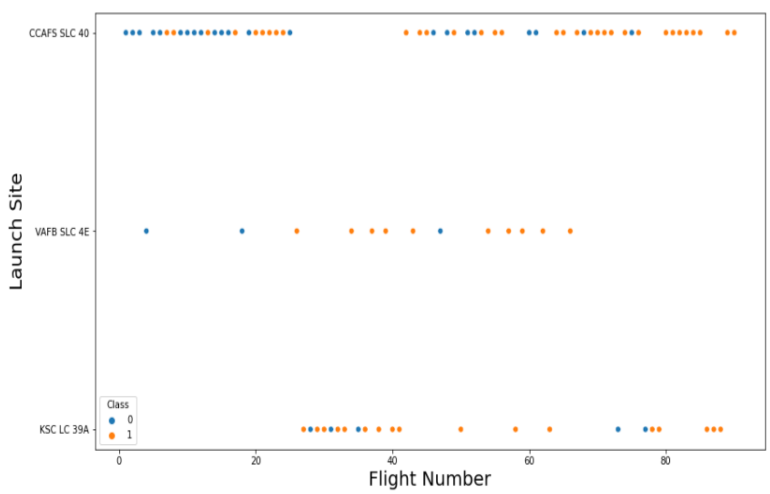


Figure F: Launch Site vs. Payload Mass

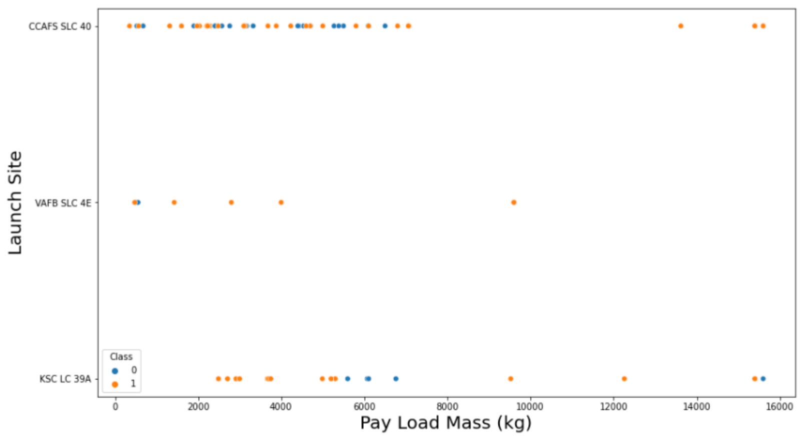


Figure G: Orbit vs. Flight Number

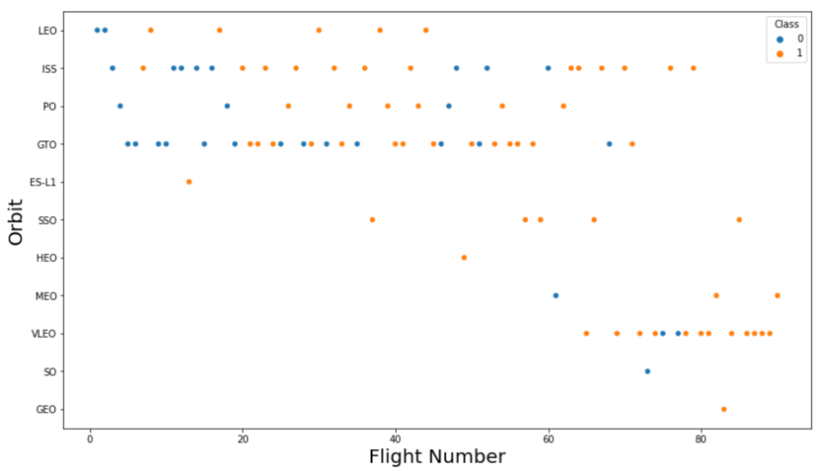


Figure H: Orbit vs. Payload Mass

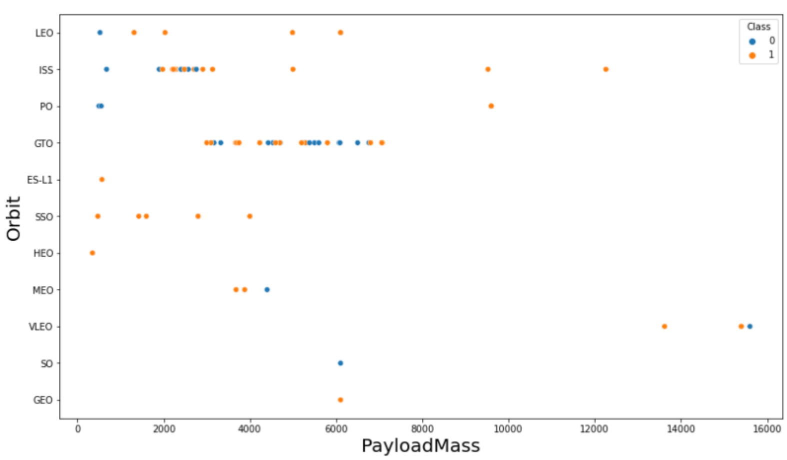


Figure I: Success vs. Year

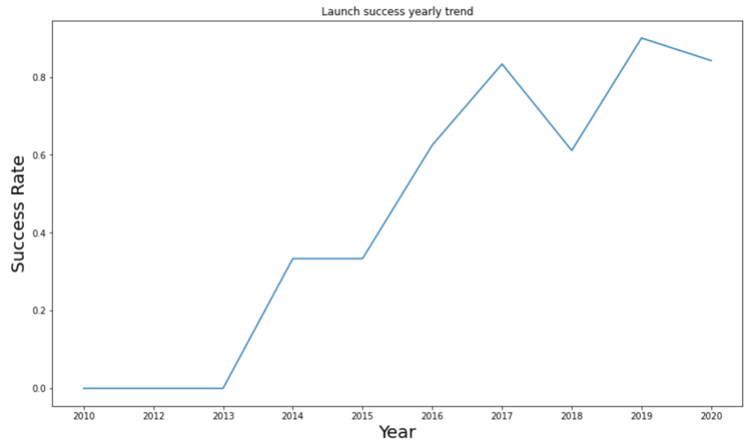
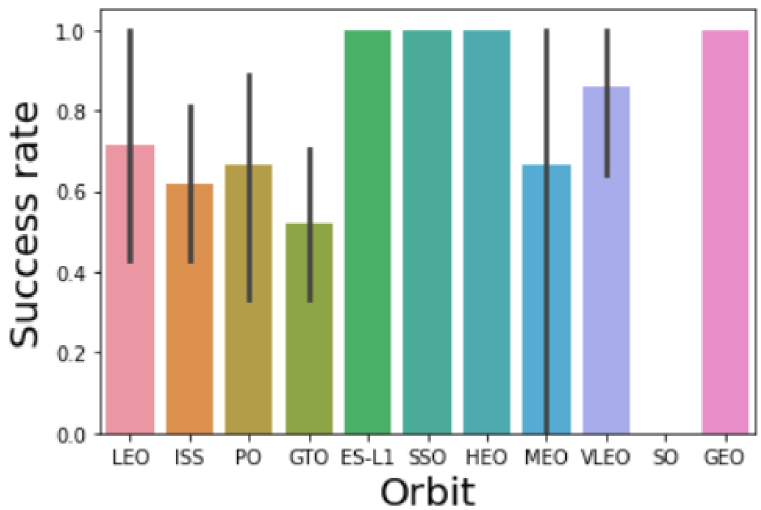


Figure G: Success Rate vs. Orbit



Launch Sites:

The Overall Launch Sites and Outcomes are as follows:

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

landing__outcome	landing_count
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

Predictive Analysis (ML) :

Using EDA, I was able to see the different effects of launch heuristics on success rate. Using Machine Learning for Predictive Analysis, I am able to more precisely predict the outcome of launches. The overall steps I used was creating a column for the class (using numpy), standardizing the data, and splitting into training data and test data (using train_test_split).

In this section, I found that the biggest problem was false positives. However, by selecting the best hyperparameters on decision tree classifiers, the accuracy of test data and outcomes was 83.33%. The Decision Tree Algorithm had the highest accuracy out of all 4 machine learning algorithms.

Conclusion:

Space X Falcon 9 launches are important to analyze to find how to reduce costs, how opposing bidders can invest, and how to improve launches.

In this project, I analyzed the effect of various factors (launch site location, mass, orbit type, etc.) and their effect on the success of landing.

I created graphs, tables, and deployed machine learning algorithms to prove my findings.