SpaceX Launch Success

LAUNCH EVALUATIONS FOR THE FALCON9

Table of Contents

Executive Summary

Methodologies

Data Collection with the Spacex API and also with webscrapping.

Data Wrangling to facilitate clear and efficient data exploration.

Data Exploration of successful launch factors through attribute selection.

Data Analysis of results with SQL to quantify the amount each attribute influences the success rate.

Data Visualization with folium and plotly to visualize attribute statistics in a geographical and comparative manner.

Model Building to make predictions with: Support Vector Machines, Logistic Regression, K- Nearest Neighbor, and Decision Trees.

Results

Exploratory Data Analysis and Visualization
Per launch site, success rate improved with flight number.
Low Earth Orbit seems to have a higher success rate with heavier payloads.

Visualization with Folium
Launch sites are close to ocean coasts.

Model Evaluation
The decision tree model performed the best, but all the tested models had scores relatively close to each other.

Introduction

This project aims to predict the potential success of a rocket launch based of selected attributes from a data set collected from internet sources. The project focuses on the rocket launches of the SpaceX company and aims to evaluate data visually and predictably. Relevant attributes are explored and selected through the data ETL (Extract, Transform, and Load) process.

Data Exploration:

Which attributes affect launch success, ie. Launch site, payload mass, and orbit.

Visualize data to analyze the geographical impact and the correlation of success with time.

Methodology

Data Collection – REST API and webscrapping

Data Wrangling – Attribute selection, missing value correction, and one hot encoding for binary values.

Data Exploration – SQL and visualization with matplot lib and seaborn.

Data Visualization – Folium and Plotly Dash for geographical and interactive data presentations.

Model Building – evaluate various models with the clean data to select the best predictor.

Data Collection

REST API data collection from the SpacX website.

- With get method from the requests lib, receive JSON.
- Translate the response into a data frame for pandas.
- With df info collect data for rocket, payloads, launchpad, and cores
- Filter the df down to selected attributes.

Data Collection

With Webscrapping

- With Beautiful Soup, collect additional information, specifically for Falcon 9 to add to the data frame.
- With the requests lib receive an html object, which can be parsed with the html.parser in a Beautiful Soup function call.
- The information is put into a dictionary, which can then be added to the data frame.

Data Wrangling

Missing Values - Add the mean of values in PayloadMass to the missing values in that same column.

Calculate - Number of launches per launch site, occurrences per orbit type, landing outcomes per True Ocean, False Ocean, True RTLS (ground pad), False RTLS, True ASDS (drone ship), False ASDS. None ASDS and None None represent a failure to land.

Create data frame column with the calculated outcomes.

EDA - Visualization

Visual Representations, with Seaborn, of data correlations:

- All charts identify success and failure with binary coloring.
- Flight Number and Payload Mass
- Flight Number and Launch Site
- Payload Mass and Launch Site
- Payload Mass and Orbit Type
- Orbit and Flight Number
- Success over time

Bar charts and Scatter Plots helped visually identify strong correlations.

EDA - SQL

SQL queries identify relevant data:

- Avg and sum of various payload masses, ie. for Nasa and Falcon 9 rockets.
- Select min and max parameters for selected attributes.
- Count functions allow a numeric representation for successful and failed missions.
- With the success count information, landing methods by drone ship and ground pad can ranked.

Folium Map

The Folium Map allows for a geographical representation of selected data:\

- Markers allow clear identity of locations such as the NASA Johnson Space Center.
- Marker Groups identify important locations where multiple data points are related. For instance, launches from certain launch sites.
- Colors serve as delimiters between success and failure.
- After sites are evaluated, there geographical information may be quickly assessed. For instance, finding that Launch Sites are close to Ocean coast lines.

Plotly Dash

The Dash board allows for interactive representations of data:

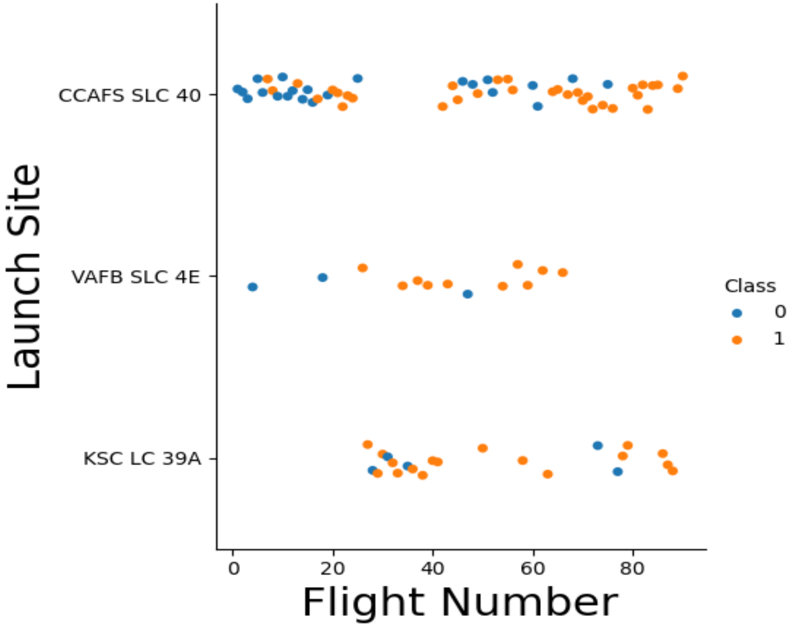
- Launch sites could be selected by user as input.
- Output allows the user to see success for selected input as a pie chart.
- Another user selected input, Payload Mass, may be selected with a slider.
- The payload Mass input will Output a scatter chart.

Model Evaluation

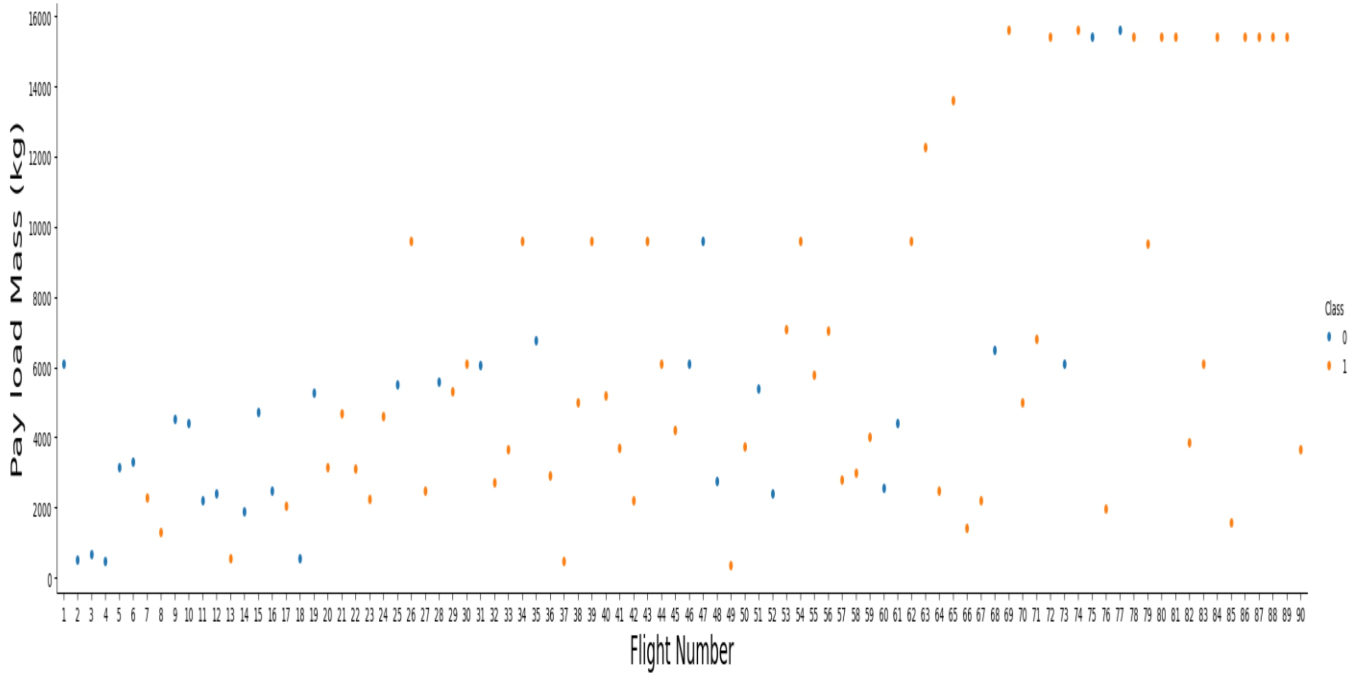
After Data was scaled, then split into testing and training sets, four models were trained and evaluated with the data.

- The models evaluated with GridSearchCV were:
- Logistic Regression
- Support Vector Machine
- Decision Tree
- K- Nearest Neighbor
- The scores for the decision tree illustrated that it is likely the best model for making predictions of launch success.

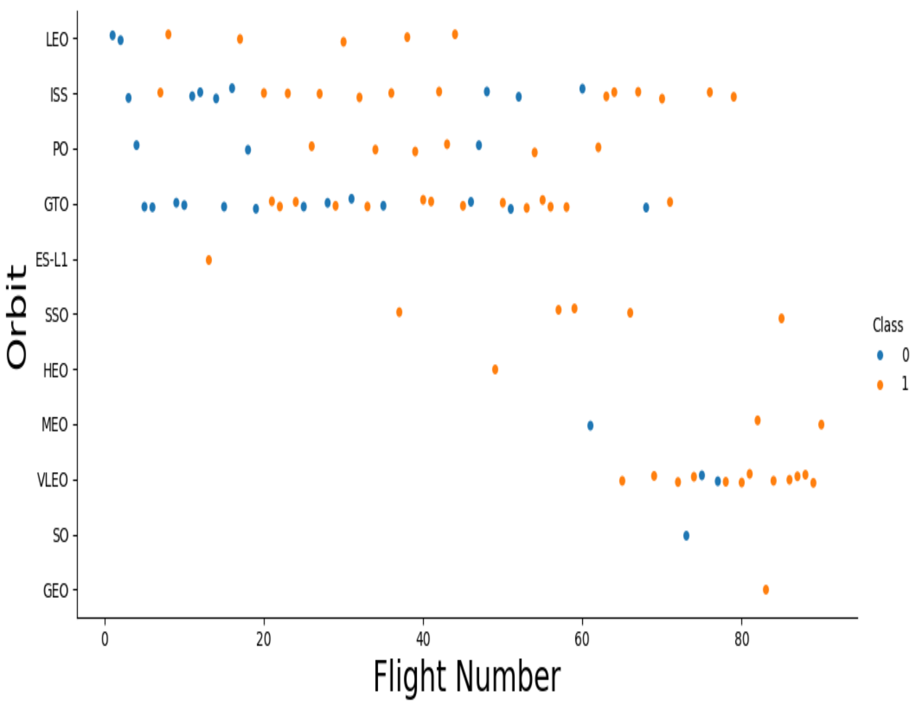
Results

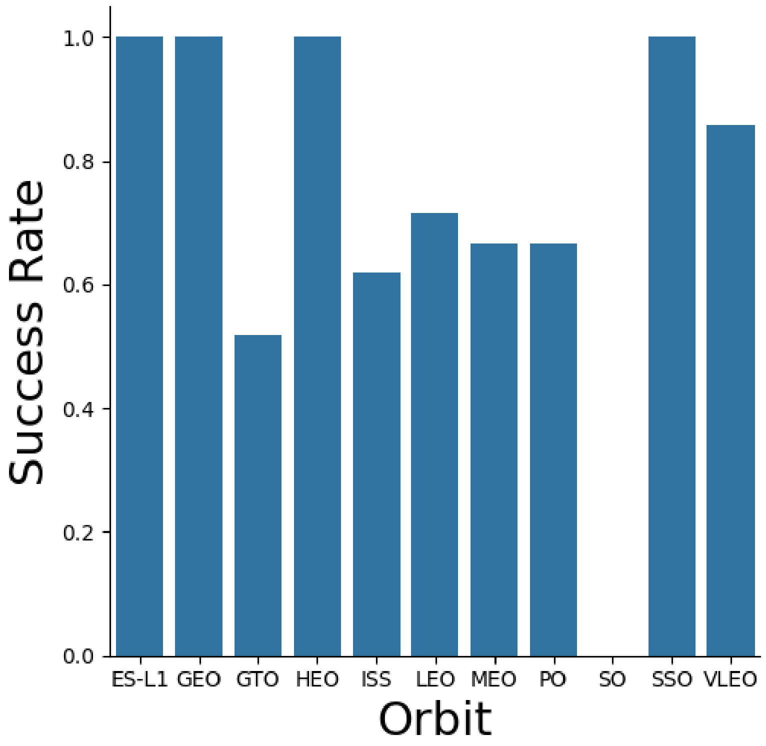


Success rate increases over time regardless of launch site.



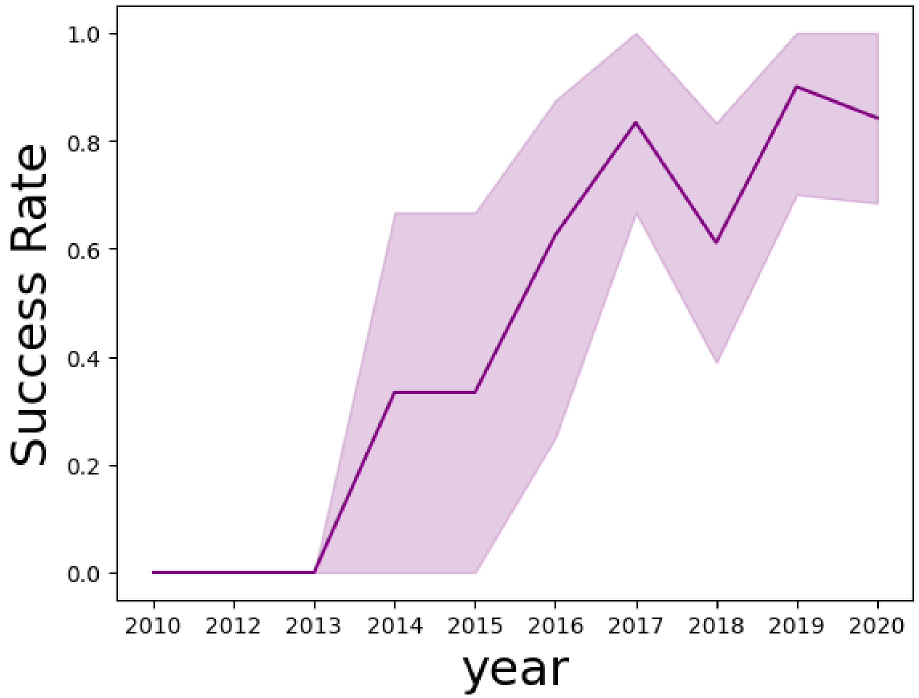
Results cont





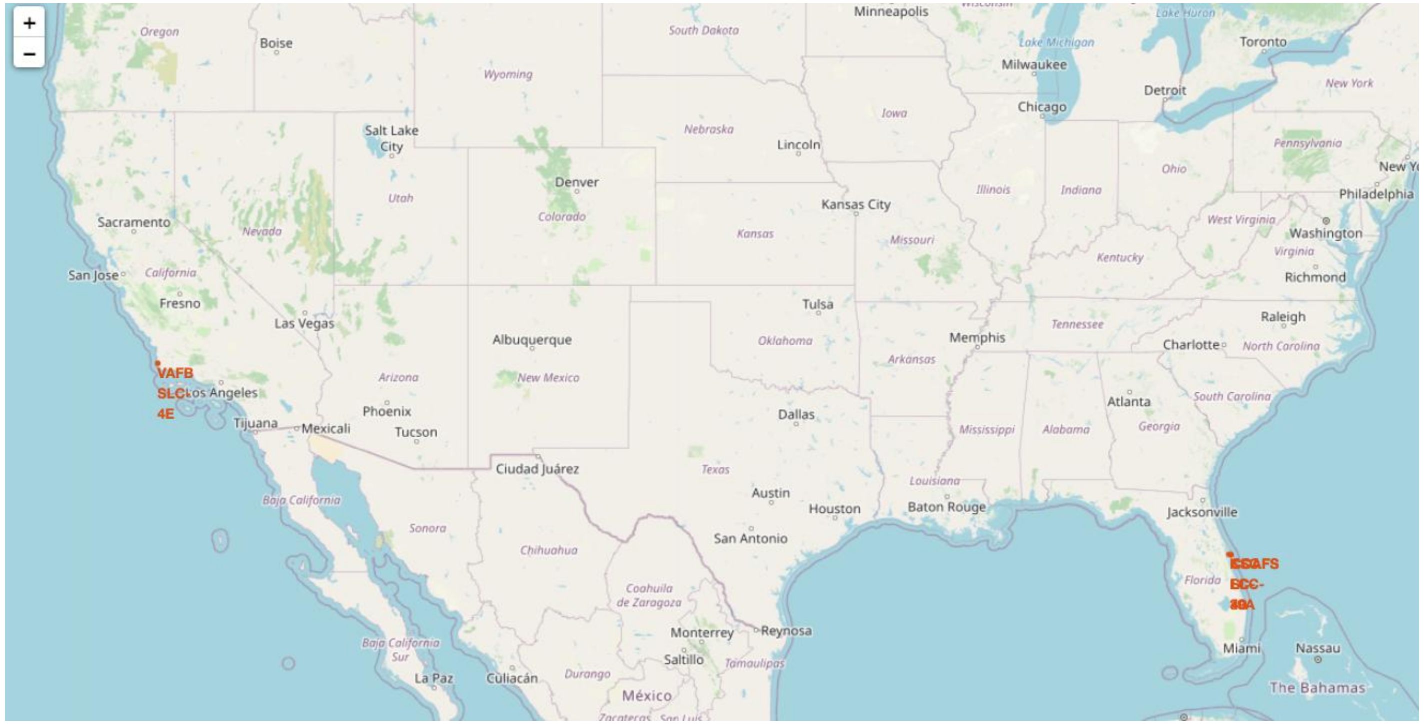
GTO orbit (geosynchronous orbit, which is in tune with the earths rotation. Great for weather satellites) has one of the lowest success rates. SO (Sun orbit) has one failed mission.

Results cont



Success rate increased greatly from 2013 to 2017 and then dipped and seemed to plateau. Although a plateau is fine nearest a 100 percent success rate, it isn't quiet there. Identifying factors that influence failed missions and reducing their weight, may be beneficial for future successes.

Results cont



Launch sites are located near Ocean coastlines and away from large cities.

Conclusion

The data suggests that although success rate has increased over time, it still has experienced significant failures and improvements would allow for a success rate closer and more consistently towards 100 percent. Finding the issues with launches at GT Orbit that have a lower success rate should be identified.