SpaceX Data Science Analysis

A. Bolzonello

25th August 2023

Contents

EXECUTIVE SUMMARY

The following report determines the success rate and factors of a successful SpaceX rocket launch by using data science models and tools to conduct the following steps to reach our results :

- Data collection
- Exploratory data analysis
- Visualizations; and
- Predictive analytics

INTRODUCTION

SpaceX's cost-effective Falcon 9 launches, priced at $62 million due to reusability, stand out against competitors' $165 million rates. Their 2010 achievement of recovering a spacecraft from low-earth orbit is historic. This capstone project aims to predict Falcon 9 first-stage landings, crucial for competitors, using available data.

By the end of this report, we will answer the following questions:

- How payload mass, launch site, number of flights, and orbits affect first-stage landing success
- Rate of successful landings over time
- Best predictive model for successful landing

Methodologies

- Data collection
  - SpaceX API
  - Falcon 9 historical launch records via Wikipedia and (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)
- Data Wrangling
  - Prepare the data by cleaning and transforming it to uncover data patterns.
- Perform exploratory data analysis (EDA)
  - Used visualization and SQL, to determine: total payload, payload range for successful launches, and total number of successful and failed outcomes
- Create visual analytics using Folium and Plotly Dash
  - To explore factors to contributed to launch site success rates and to visualize launch sites that were the most successful.
- Predictive analysis using classification models
  - To determine the best predictive model for landing outcomes we used logistic regression, support vector machine (SVM), decision tree and K-nearest neighbor (KNN)

Data Collection and Data Wrangling

Data Collection - API

1. Request data from SpaceX API (rocket launch data)
2. Decode response and create data frame
3. Filter data frame to contain only Falcon 9 launches
4. Replace missing values with average Payload Mass

Data Collection – Web Scrapping

5. Request data (Falcon 9 launch data) from Wikipedia
6. Create BeautifulSoup object from HTML response
7. Extract all column names for HTML tables
8. Create data frame through parsing the launch HTML tables

Data Wrangling

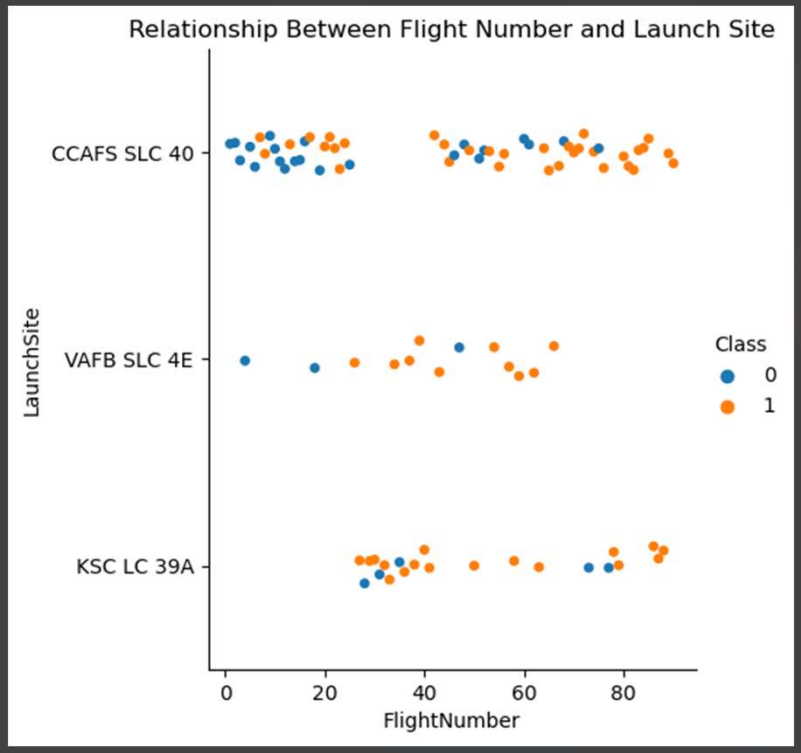
Summaries launch sites by site, orbit and mission outcome

Convert outcomes into 1 for a successful landing and 0 for an unsuccessful landing

Create landing outcome table

Results

Flight Number vs. Launch Site

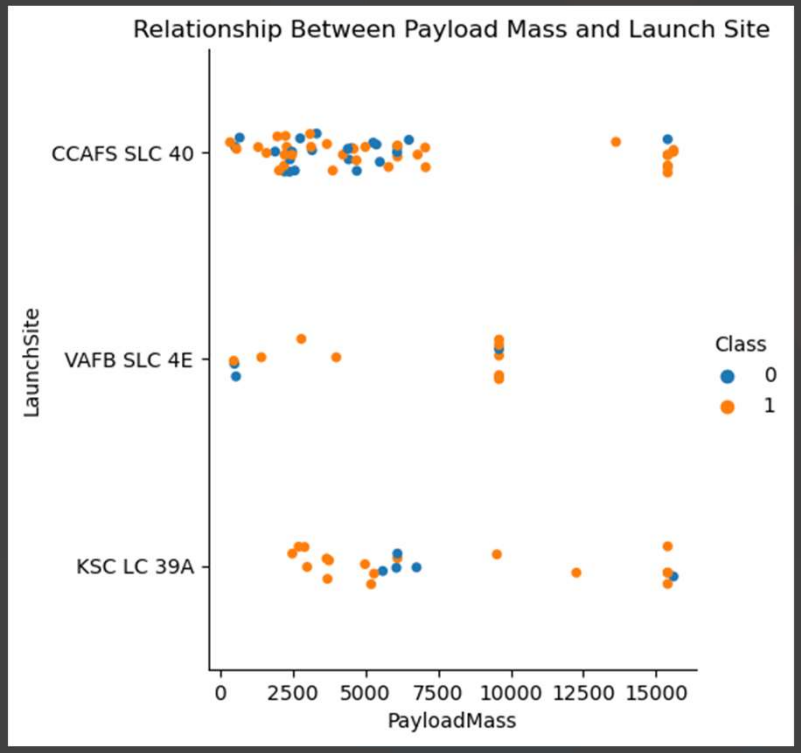


Observations:
VAFB SLC 4E and KSC LC 39A have higher success rates

The most frequent launches came from CCAFS SLC 40 launch site

Later flight in orange has a higher success rate than earlier flights in blue.

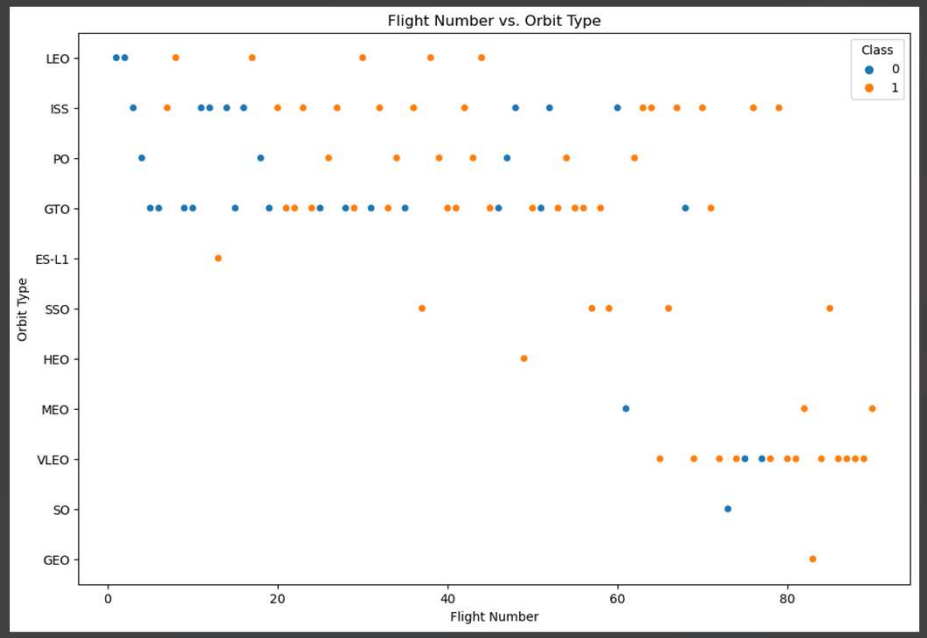
Payload vs Launch Site



Observations:

- Typically, the higher the payload mass (kg), the higher the success rate
- Most launches have a payload of less than 7500kg , however, launches over 7500 were a majority successful
- KSC LC 39A has a 100% success rate for launches less than 5,500 kg

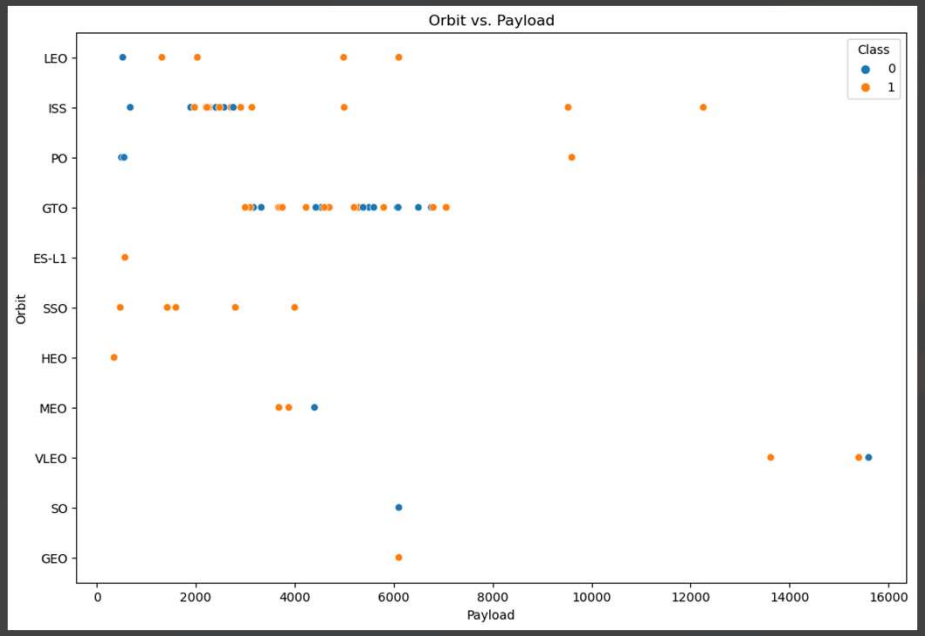
Flight Number vs. Orbit



Observations:

- Success rate increases with the number of flights for each orbit except for GTO orbit

Payload vs. Orbit



Observations:

- Heavier payload mass for PO, LEO and ISS orbits results in higher landing success

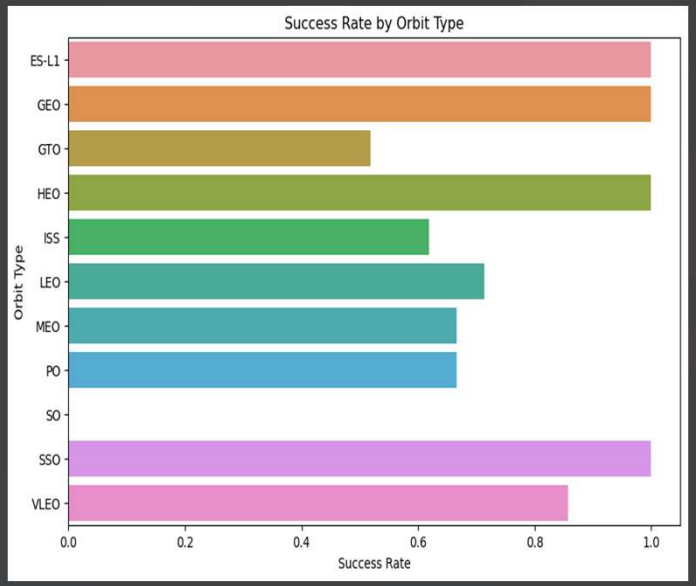
Success Rate

Success rate by Orbit type

ES-L1, GEO, HEO and SSO
100% Success Rate

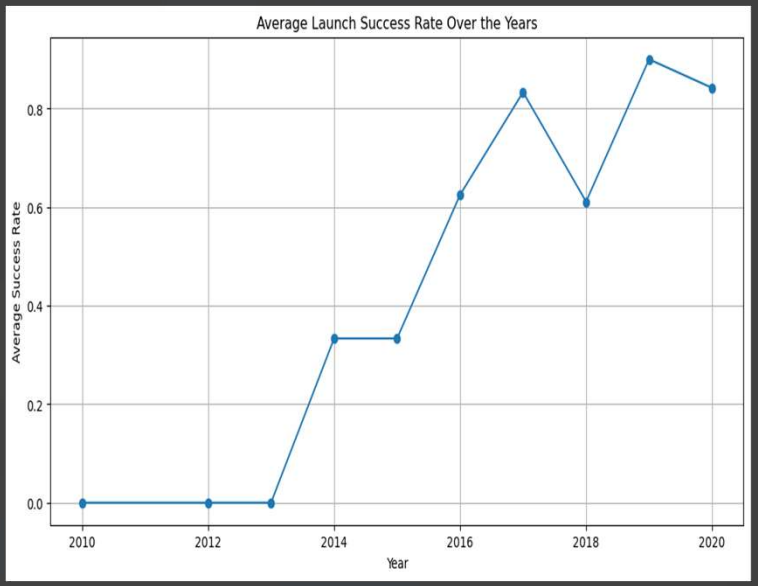
GTO, ISS, LEO, MEO, PO
50%-80% Success Rate

SO
0% Success Rate



Yearly Trend

Over the past 10 years the average launch rates have continued to increase significantly with a drop in 2018



EDA with SQL - Findings

Names of launch sites:

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

Average payload mass carried by booster version F9 v1.1

AVG(PAYLOAD_MASS__KG_)
2928.4

Total number of successful and failure mission outcomes

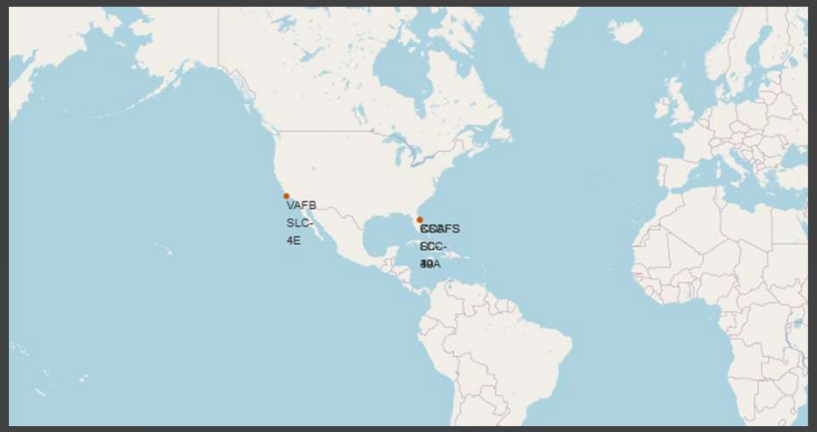
Mission_Outcome	total_number
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

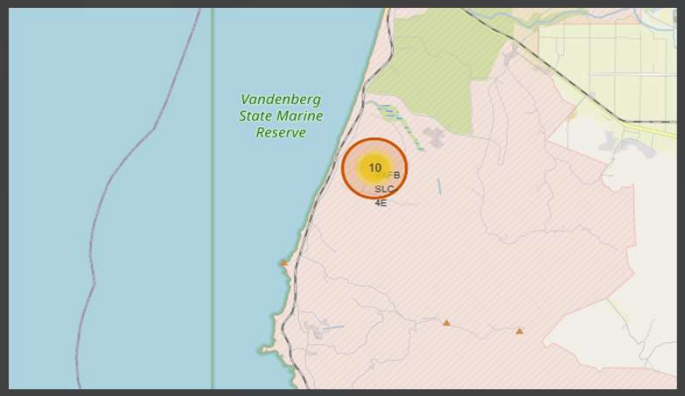
Launch Sites with Folium

Observations:

All launch sites are located on the equator

All launch site are in close proximity to coastlines

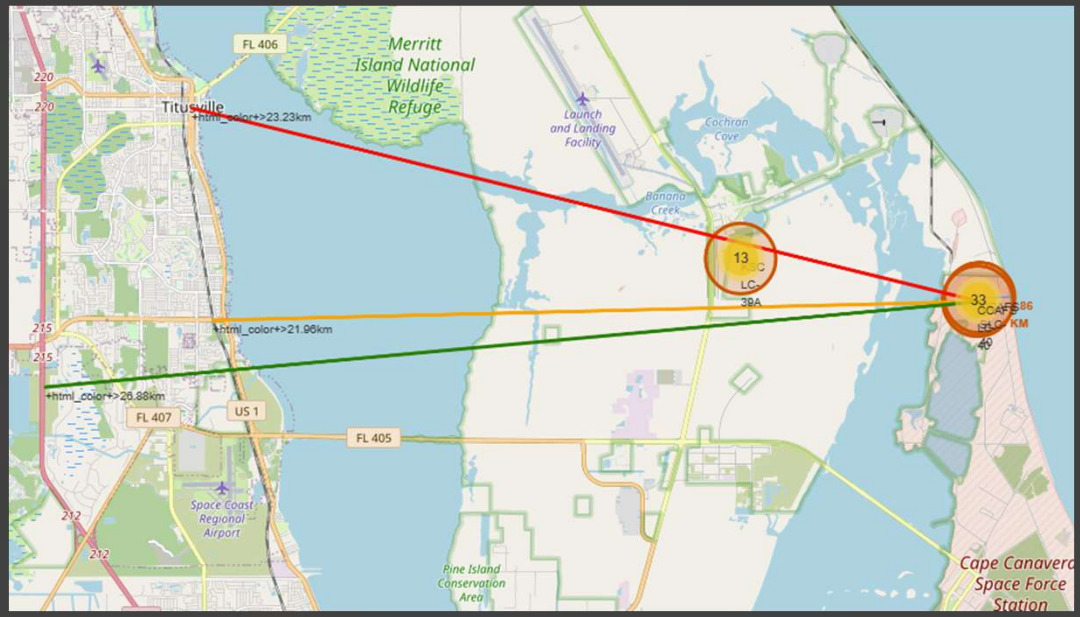




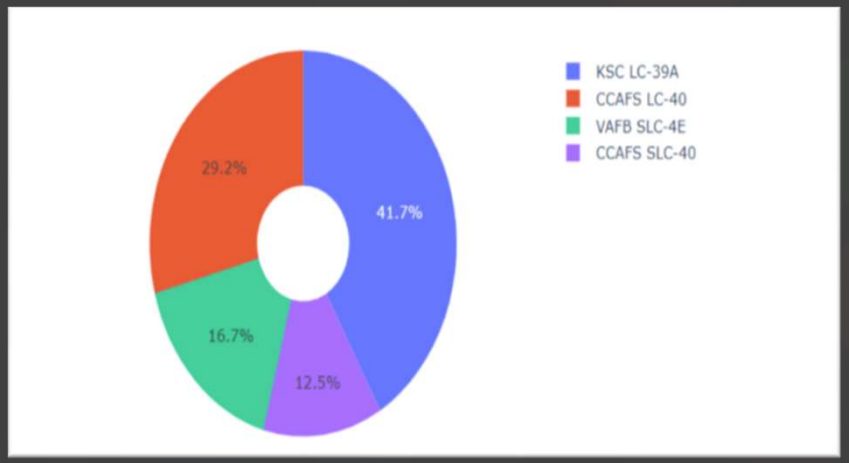
Launch Sites with Folium cont.

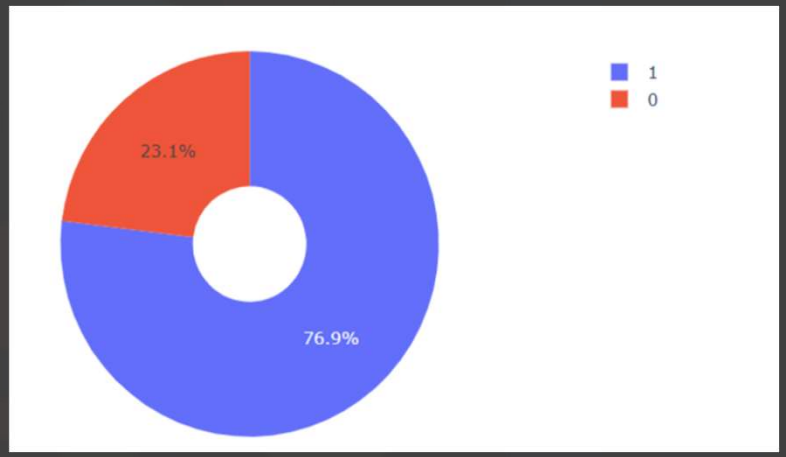
```
City Distance 23.234752126023245
Railway Distance 21.961465676043673
Highway Distance 26.88038569681492
Coastline Distance 0.8627671182499878
```

Launch sites are located far from cities to be safe but close enough to roads, rails and ports for the transport of people and material



Launch Site with Plotly





KSC LC-39A have the most successful launches at a 76.9% success rate

Machine Learning Prediction

All the models had the same accuracy scores possible due to the small sample size of 18

	ML Method	Accuracy Score (%)
0	Support Vector Machine	83.333333
1	Logistic Regression	83.333333
2	K Nearest Neighbour	83.333333
3	Decision Tree	83.333333

The Decision Tree slightly outperformed when we used GridSearchCV to determine the highest score (at 0.8892857142857145)

Summary

Conclusion

- Component Success
  - Launch Site: KSC LC-39A had the highest success rate.
  - Orbits: ES-L1, GEO, HEO, SSO, VLEO had the highest success rates.
- Successful landings over time
  - Launch success has improved over time from 2015 after the first successful outcome.
- Best predictive model for successful landing.
  - The Decision tree classifier

Further Research

A large set would have determined a better predictive model.