SpaceX Analysis

Jeremy Raff

February 5, 2022

OUTLINE

EXECUTIVE SUMMARY

- This study outlines analyzes SpaceX data to determine the most predictive variables for launch site success
- Findings include the following:
  - Successful SpaceX launches come from 2014 or more recent, with most success coming after 2017
  - Payloads between 2,000-4,000kg are most successful
  - Recovery via drone ship is most effective

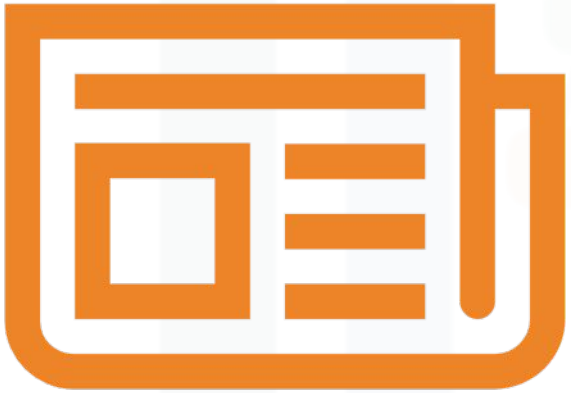
- This study develops a machine learning model that has a 88.7% accuracy rate. While not perfect, this is does provide a strong predictive value.

-

INTRODUCTION

- SpaceX has conducted over 100 launches since their inception
- These launches has a variety of successes, but they’ve managed some high profile wins
- Most recently, SpaceX sent astronauts to the edge of space in a commercial flight
- This study aims to determine what characteristics lead to successful SpaceX flights and how to predict launch success

METHODOLOGY: Data Collection & Wrangling



- Data collected from SpaceX API
- Select 15 different variables and place into dataframe
- Additional variables scraped using BeautifulSoup from SpaceX wikipedia page and converted into pandas dataframe
- Missing variables addressed in the dataset

METHODOLOGY: Data Collection & Wrangling

- There are 8 different outcomes each of which can be determined as success or failures
- Each are coded into 0 or 1 as a dummy variable for further exploratory data analysis
  - 1: Successful
  - 0: Unsuccessful

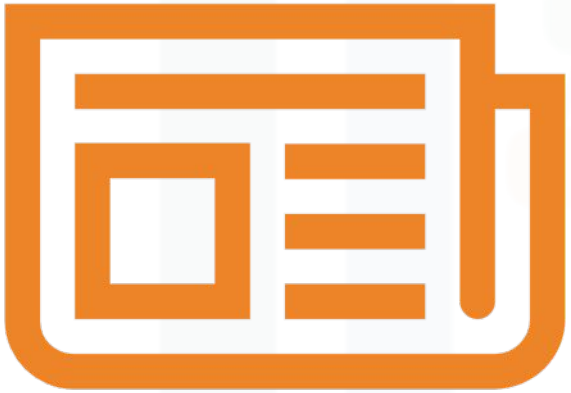
METHODOLOGY: EDA & Interactive Analytics

- Analysis of the most common orbits, launch sites, and success rates
- Results coded to identify the success rate of launches
- Utilized seaborn library to visualize distribution of data

METHODOLOGY: EDA with SQL

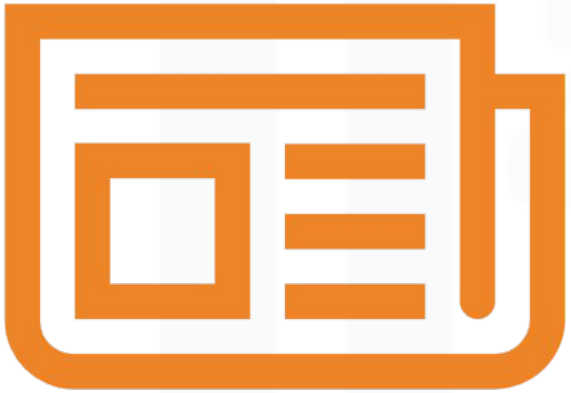
- SQL is used to run queries to better understand the data. Relationships are evaluated between the following variables:
  - Launch site
  - Payload
  - Mission Outcome
  - Launch Date
  - Booster Type

METHODOLOGY: EDA with Folium



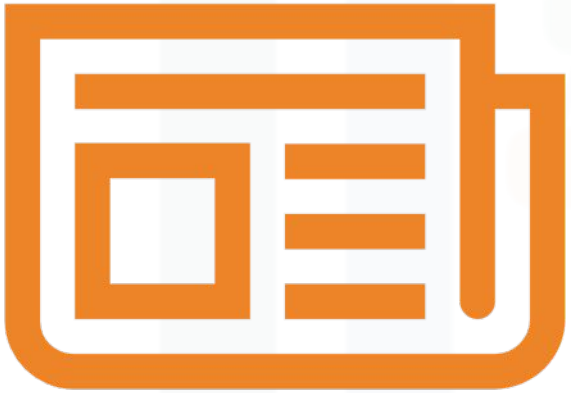
- The python package *Folium* is used to visualize and map some of the key variables.
- These are mapped to determine proximity to cities, coast lines, highways, and railways

METHODOLOGY: Plotly Dash



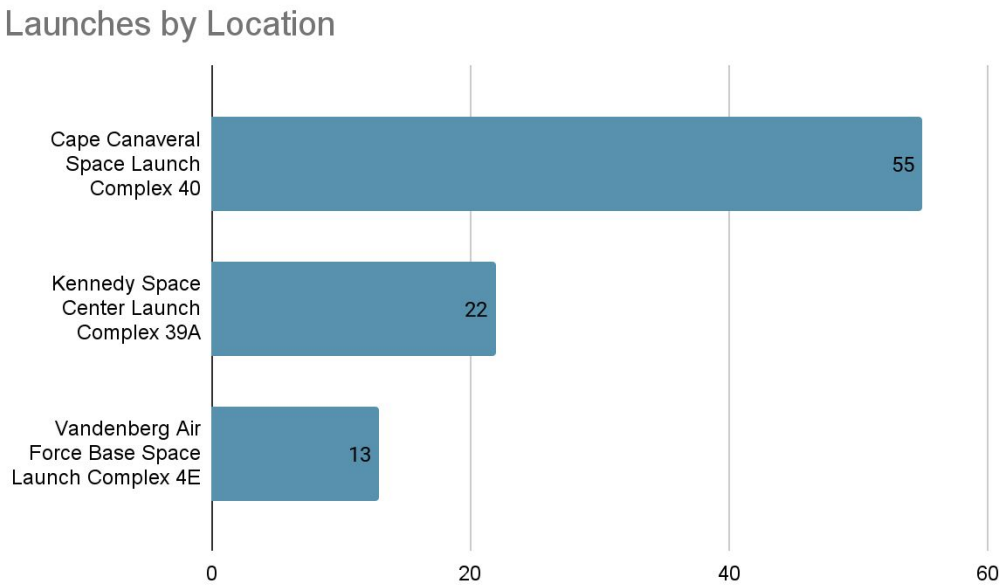
- Plotly Dash is utilized to create an interactive dashboard.
- Pie chart used showing successful and unsuccessful launches
- A scatterplot maps recovery outcome vs. payload mass
- Each visual can be altered by the end user for analysis

METHODOLOGY: Predictive Analysis



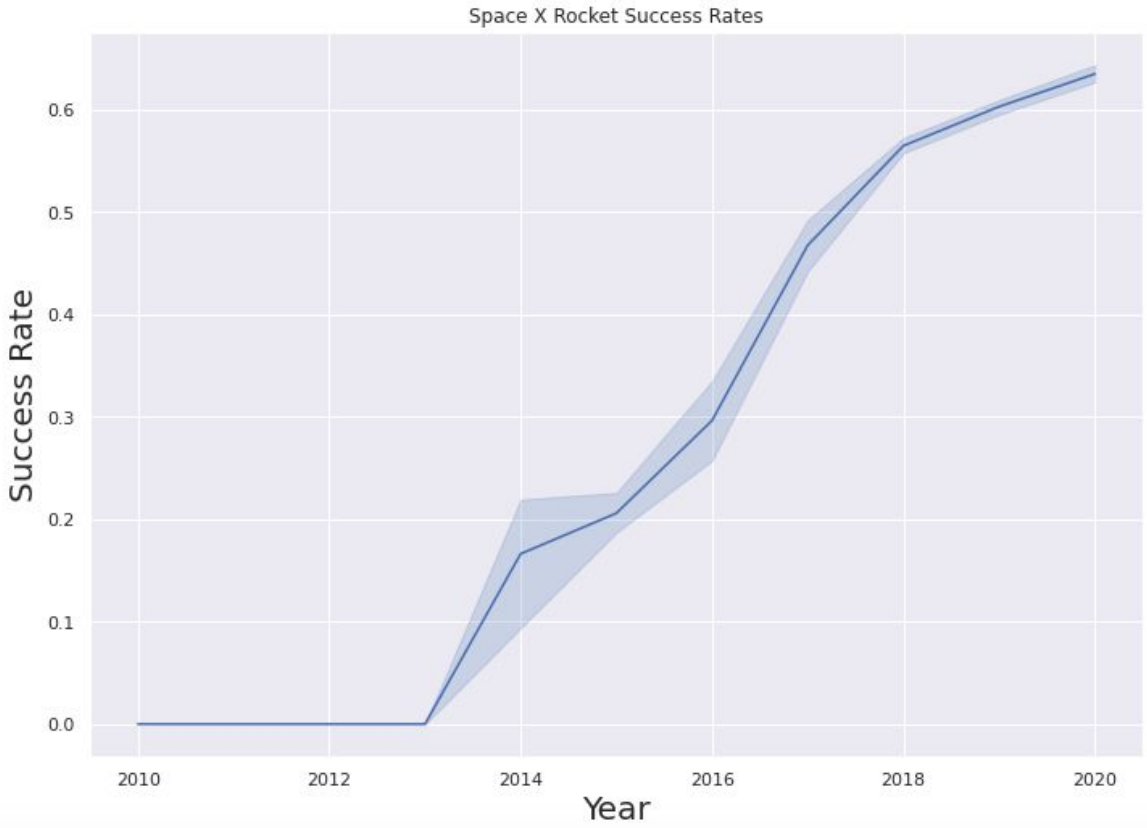
- Four machine learning models are applied to best predict launch success:
  - Logistic Regression
  - K-Nearest Neighbor (KNN)
  - Decision Trees
  - Support Vector Machine (SVM)
- Each method splits the data into testing and training sets
- Models are evaluated to determine which is most accurate

Results: EDA & Interactive Analytics



- CCAFS SLC 40 is identified as the most common launch site
- GTO Orbit was the most common
- Launches, in total, have a 66.6% success rate

Results: EDA & Interactive Analytics



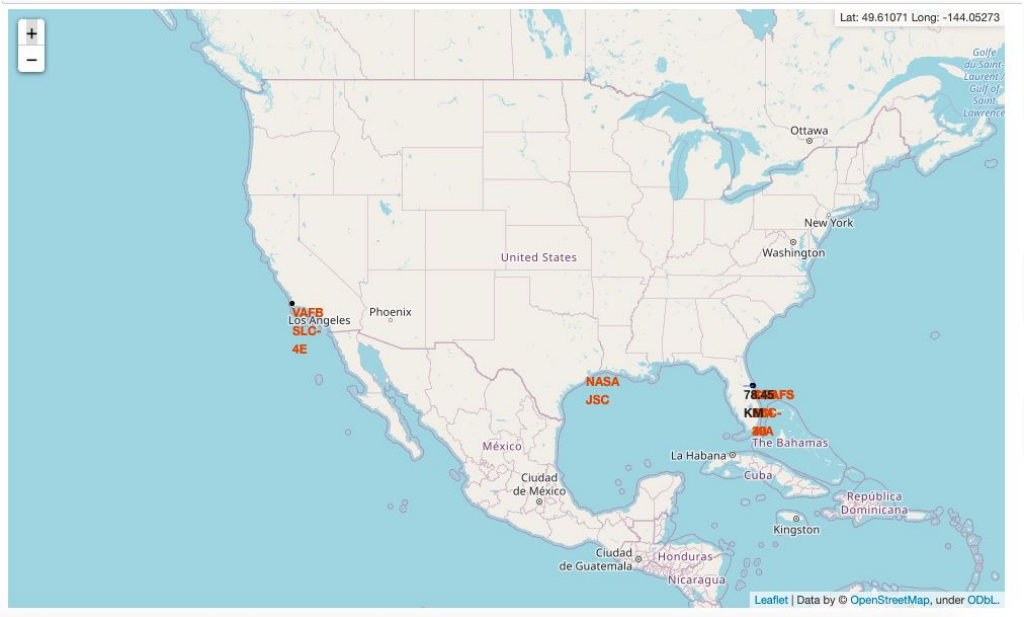
- VAFB SLC 4E has the best launch rate success
- The most successful orbits by mean are GEO, HEO, SSO, and ES-L1
- Heavier payloads are more successful for Polar, LEO, and ISS
- GTO has no discernable difference
- The launch success rate has steadily improved which is to be expected given improvements in technology (see figure to the left). There is a significant improvement since 2014
- Drop ships appear to have the best success at recovery

RESULTS: EDA with SQL

- Total payload mass carried by NASA booster is 45,596
- Booster FV9 v1.1 carried an average of 2,928.4 payload mass
- The first successful launchpad landing was on December 22, 2015
- Boosters with drone ship success with payload mass between 4,000 and 6,000 are B1022, B1026, B1021.2, and B1031.2
- 101 missions have been completed
- Drone ship landing failures in 2015 include
  - JANUARY Failure (drone ship) F9 v1.1 B1012 CCAFS LC-40
  - APRIL Failure (drone ship) F9 v1.1 B1015 CCAFS LC-40
- Successful launches between 2010 and 2017 are 2015-12-22, 2016-04-08, 2016-05-06, 2016-05-27, 2016-07-18, 2016-08-14, 2017-01-14, and 2017-02-19

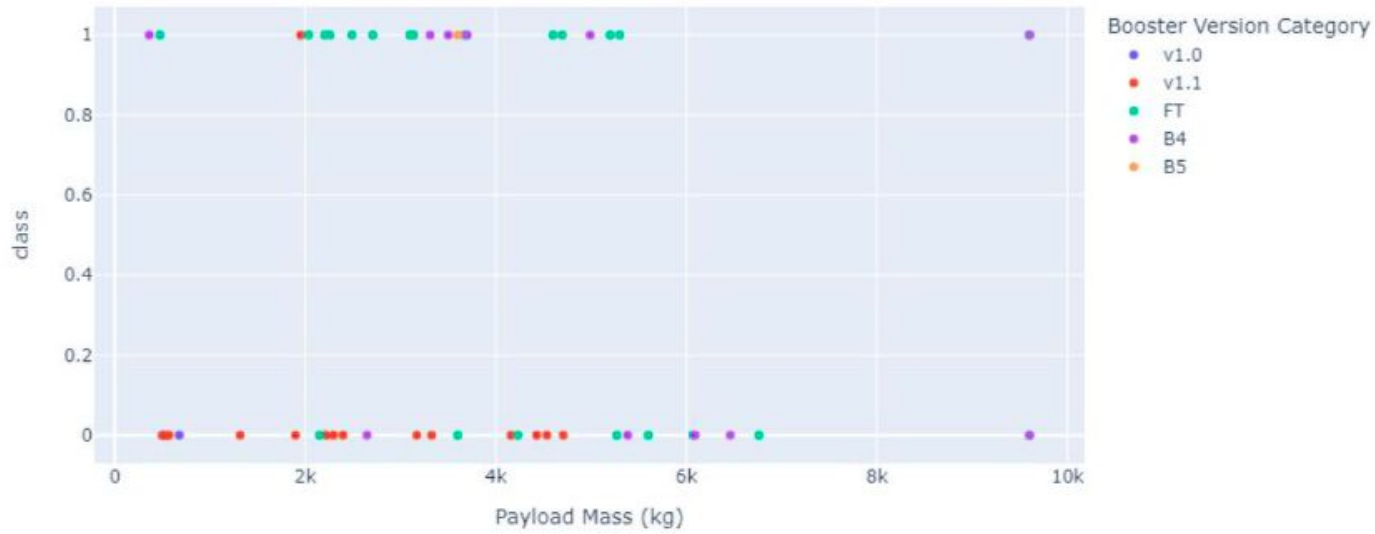
Results: Interactive Map with Folium

- All launch sites are relatively close to the equator and very close to the coast
- Launch sites remain close to railways but still remote
- They have highways nearby but not too close
- They are located by the coastline but maintain a significant distance from cities

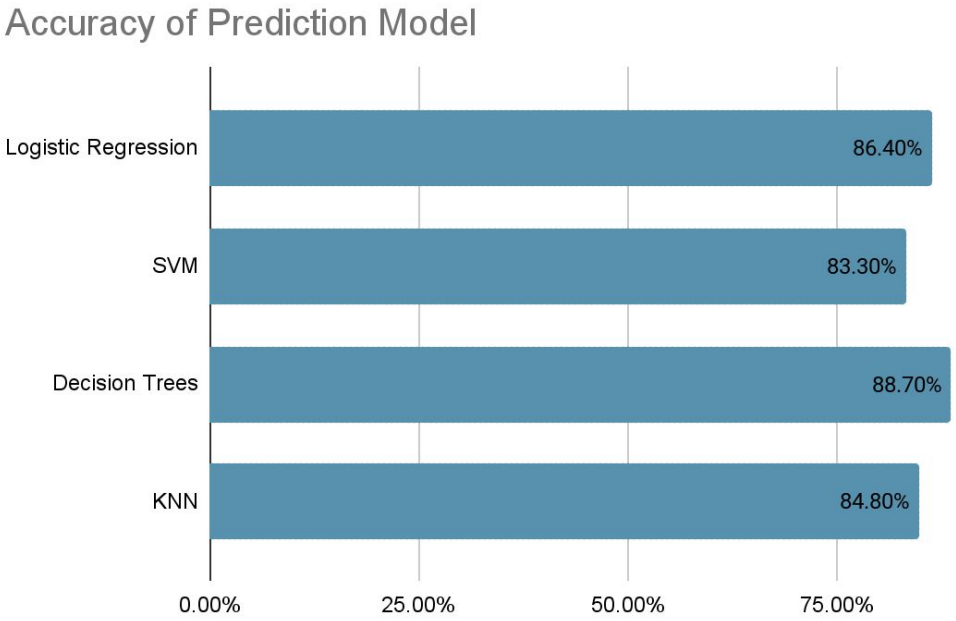


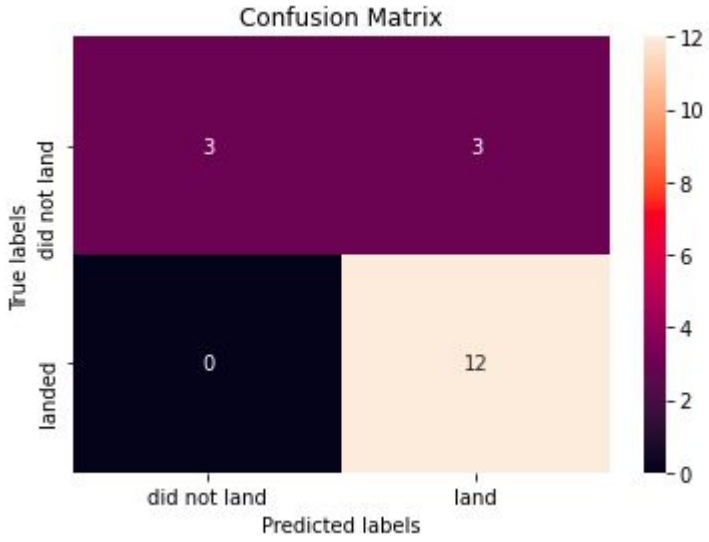
Results: Plotly Dashboard

- CCFAS LC-40 has the most launches, but it’s success percentage is not the highest
- KSC LC-39A has the best success rate of all launch sites
- The 2000-4000kg payload range appears to be the most successful
- The FT booster appears to be the most successful



Results: Predictive Analysis





- The models offer similar results with decision trees providing the strongest accuracy at 88.7%
- The confusion matrix below outlines the decision tree results
- While largely accurate, it does incorrectly assume 3 will land that do not.

IBM

SKILLS NETWORK

CONCLUSION

- Successful SpaceX launches come from 2014 or more recent, with most success coming after 2017
- Payloads between 2,000-4,000kg are most successful
- Recovery via drone ship is most effective
- Our selected model has a 88.7% accuracy rate. While not perfect, this is does provide a strong predictive value.