SpaceX Rocket Data Analysis

Vignes KV

August 4, 2024

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion
  - Findings & Implications
- Conclusion
- Appendix

EXECUTIVE SUMMARY

- SpaceX is the most established private satellite company in the world
- As SpaceY, our goal is to out-compete them by first understanding their methodologies of space launches throughout various rocket launch locations
- There are several reasons SpaceX is very successful as a rocket company
  - Multiple stages to launch rocket (first stage is to send rocket to space and land rocket successfully, and second stage helps bring payload to orbit)
  - They reuse part of the rocket launch to save dramatically on costs
  - Other companies spend ~$180 million per rocket launch, but SpaceX only spends about $60 million per rocket launch
- By predicting whether the first stage of the rocket will land, given certain factors such as rocket type, payload weight, and orbit, we can effectively determine whether SpaceX satellites will be successful, and use those factors to our advantage at SpaceY.

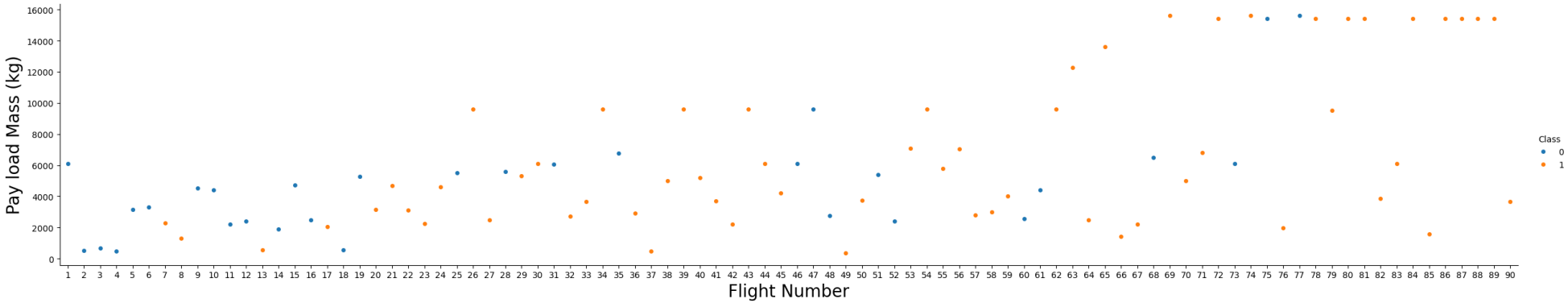
INTRODUCTION

- Satellites have been launched by rockets since the Soviet Union launched Sputnik in 1969
- Most of the work has diverged from the public sector (such as USA's NASA, and other nation's governments) into the private sector as they can better prioritize profitability and other private interests
- Several companies have started working on satellite launches, such as Globalstar, Jeff Bezo's Blue Origin, Sir Richard Branson's Virgin Galactic, but probably the most well-known and successful of them all is Elon Musk's SpaceX, which was last valued at ~$200 billion.
- SpaceX's success has come from their multi-stage launches, which enable them to reuse rockets after launching them so that they can continue using the same parts for multiple launches after they have deployed the satellites into space.
- By understanding and determining whether satellite launches are successful based upon multiple factors such as payload amount, rocket type, and other factors, we can effectively determine whether the rocket can be successfully landed and reused in future missions.

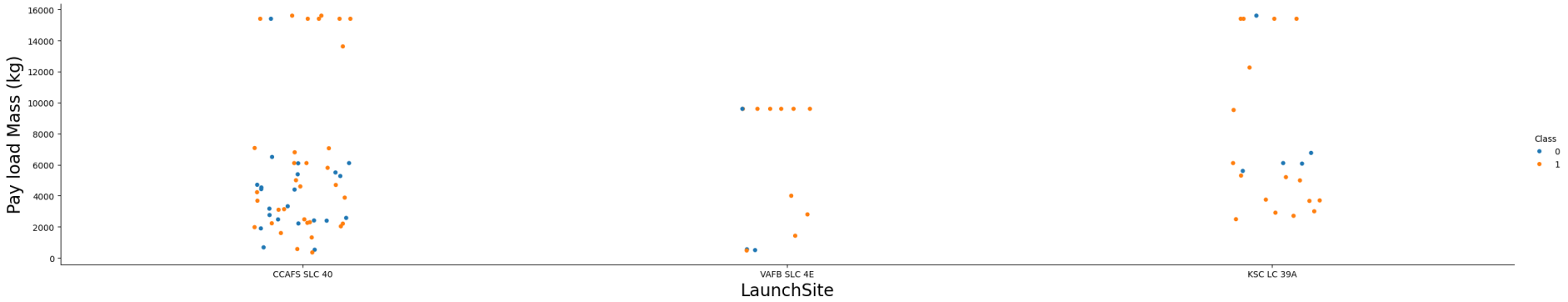
METHODOLOGY

- We first source data from SpaceX's public REST API: https://github.com/r-spacex/SpaceX-API
  - On this dataset, we will perform Exploratory Data Analysis (EDA) to understand the overall structure of the dataset as well as general statistics
- We will then perform data scraping from Wikipedia's historical table of Falcon 9 launches: https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
  - Using Beautiful Soup, we will again perform EDA to understand the structure and statistics of this dataset
- We will also utilize SQL and Folium with EDA to visually determine certain aspects of each rocket launch
- We will create an interactive visual dashboard using Plotly and Dash to determine the success rate of each rocket launch given the payload and rocket type, as well as other statistics
- Finally, we will utilize several machine learning models (such as Decision Trees, kNNs, and SVMs) to determine whether we can predict successful rocket launches given the above dataset

SPACEX EDA RESULTS

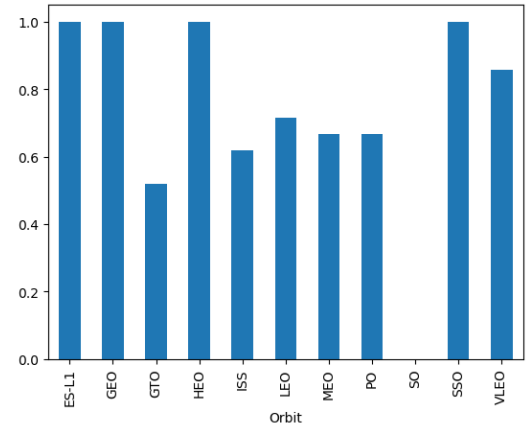


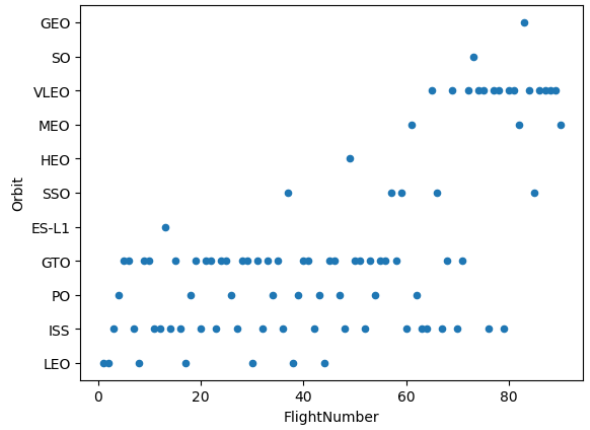
From the figure above, we can see that the majority of flights are successful, particularly those that happened more recently and with higher payloads.

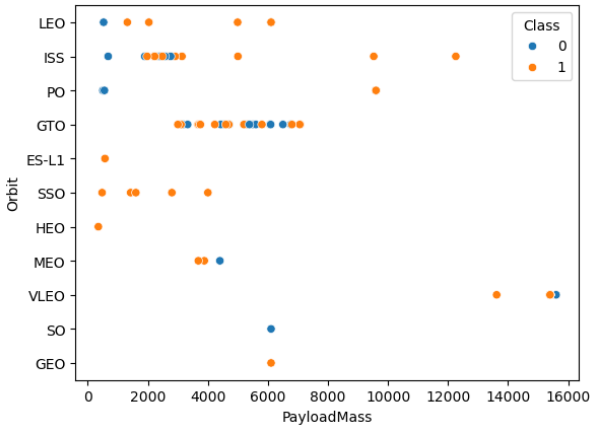


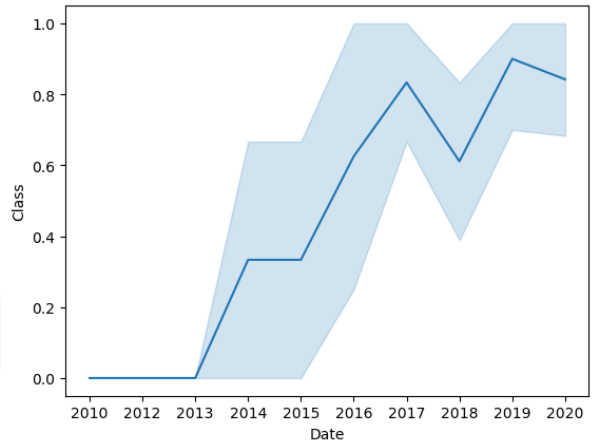
Comparing the launch sites, we can see that RSC LC 39A and VAFB SLC 4E are the more successful launch sites.

SPACEX EDA RESULTS









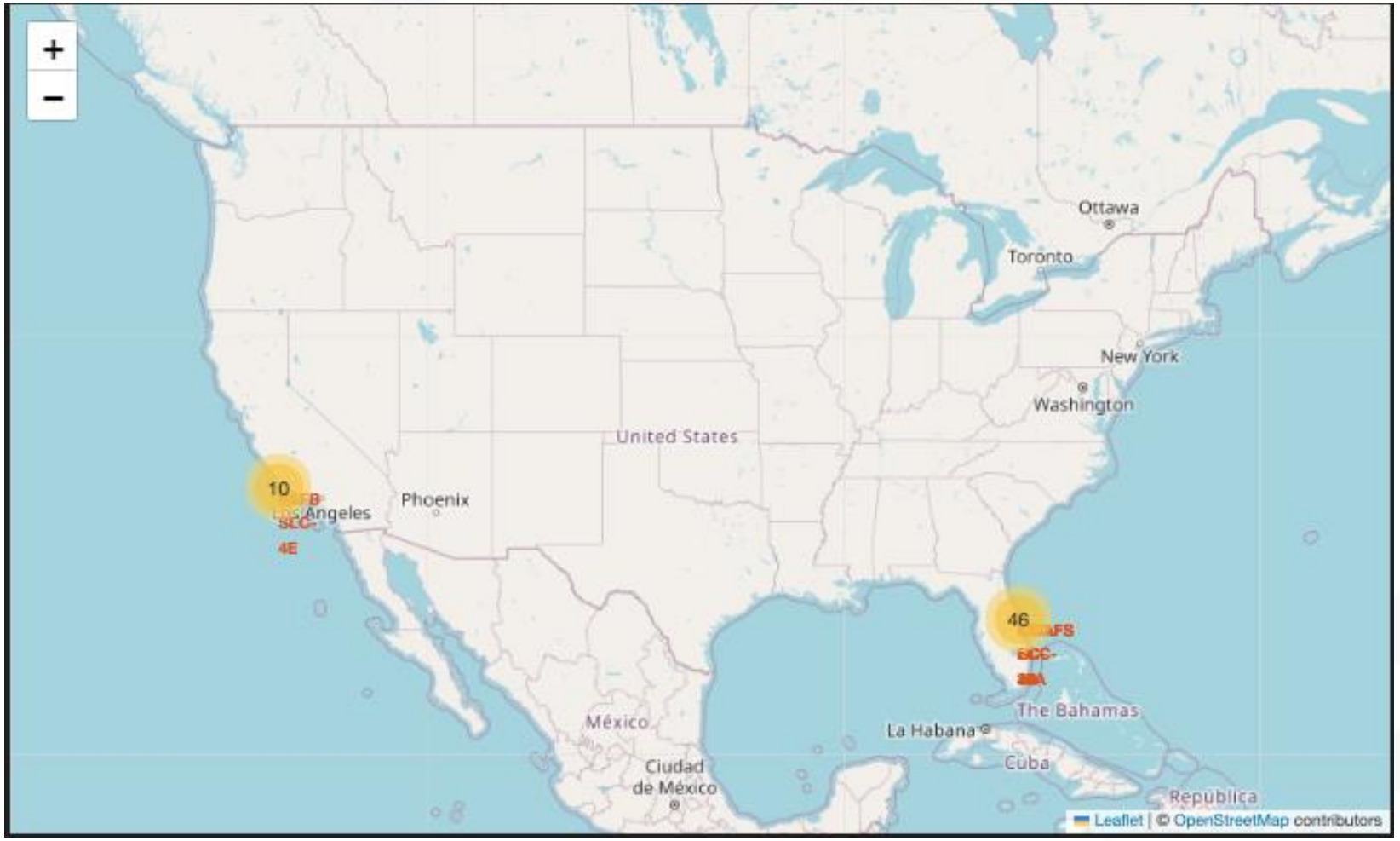
- Top Left: This figure compares the different forms of orbit with the success rate of each mission. From the image, we can see that ES-L1, GEO, HEO, and SSO orbit patterns have the highest success rate of 100%.
- Top Right: This image depicts the chronological order of each flight and the type of orbit that it wants to achieve. As you can see, the focus seems to be shifting from GTO and ISS, towards orbits like VLEO.
- Bottom Left: This graph compares the payload mass of each rocket with the anticipated orbit, as well as the success of each mission. As we can see, most missions are generally successfully, with generally lower payload missions being unsuccessful.
- Bottom right: This graph depicts the success rate of all mission in all given years. As we can see from the trendline, SpaceX has had more successful missions over the years.

SPACEX SQL RESULTS

From the SQL EDA, the following results have been determined:

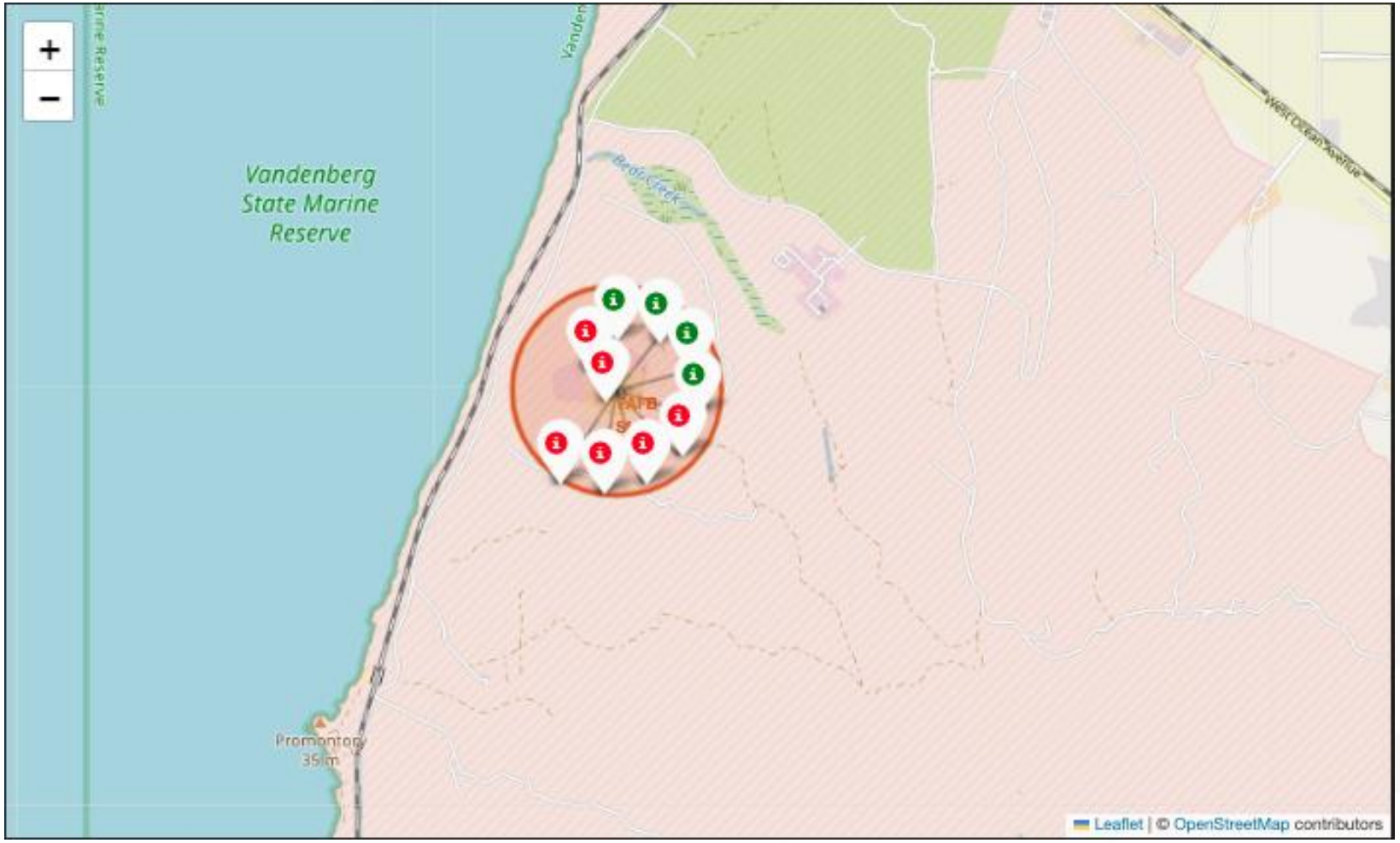
- There are four launch sites that SpaceX has launched from: CCAFS LC-40, KSC LC-39A, KSC LC-39A.
- The total payload mass carried by boosters launched by NASA was 45596 kg.
- Average payload mass carried by the F9 v1.1 booster was about 2534.67 kg.
- The earliest successful landing on a ground pad was on June 4, 2010.
- There were four boosters that have had success in drone ship landings and carried payloads between 4 and 6 thousand kilograms:
  - -F9 FT B1022
  - F9 FT B1026
  - F9 FT B1021.2
  - F9 FT B1031.2
- There were 100 successful mission outcomes and only 1 failure mission outcomes.

SPACEX FOLIUM RESULTS



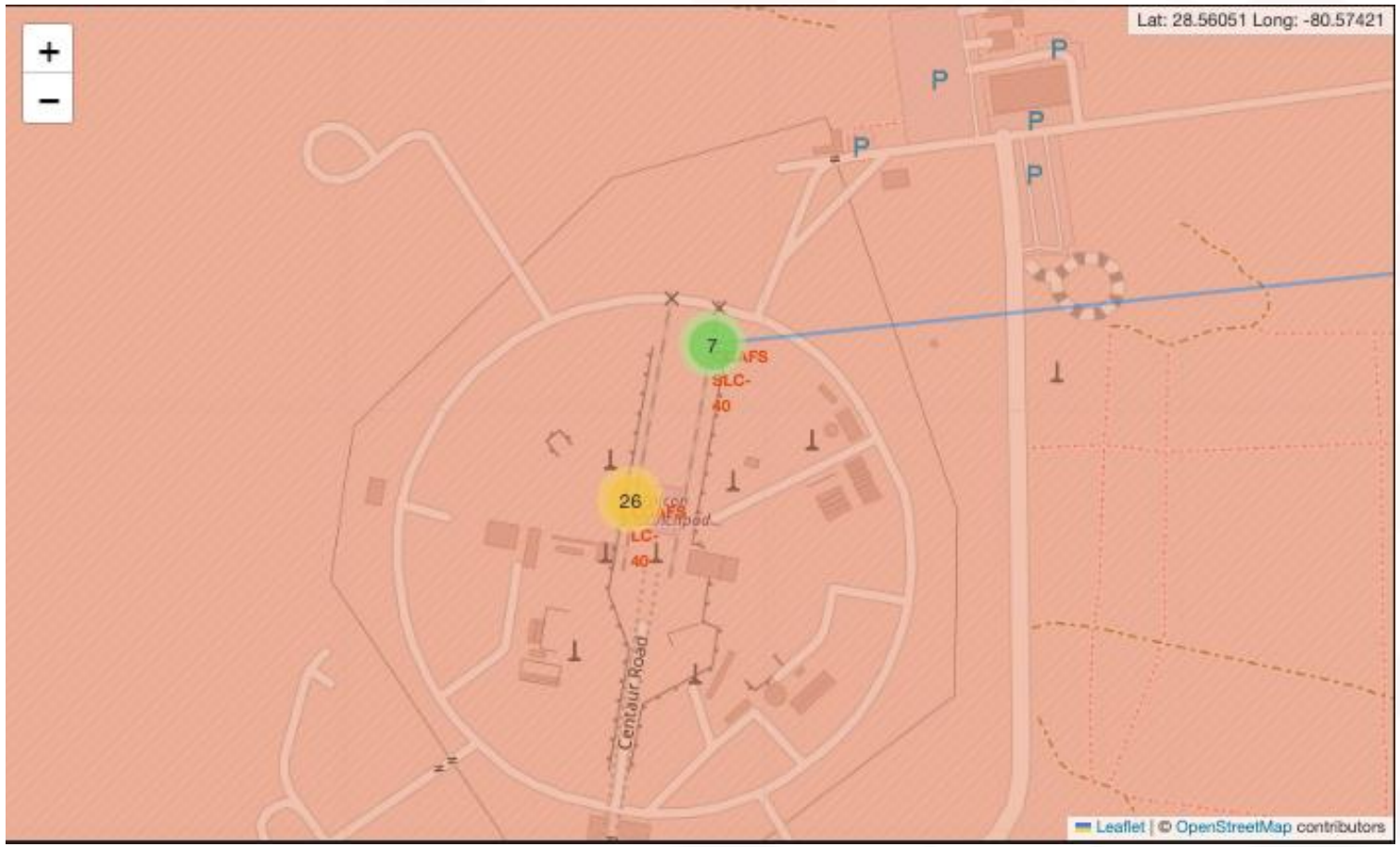
After plotting all of the SpaceX locations on Folium, the results look like the image to your left.

SPACEX FOLIUM RESULTS



Clicking on the VAFB SLC-4E region for example allows us to see the number of successful (represented by green) and unsuccessful (represented by red) markers.

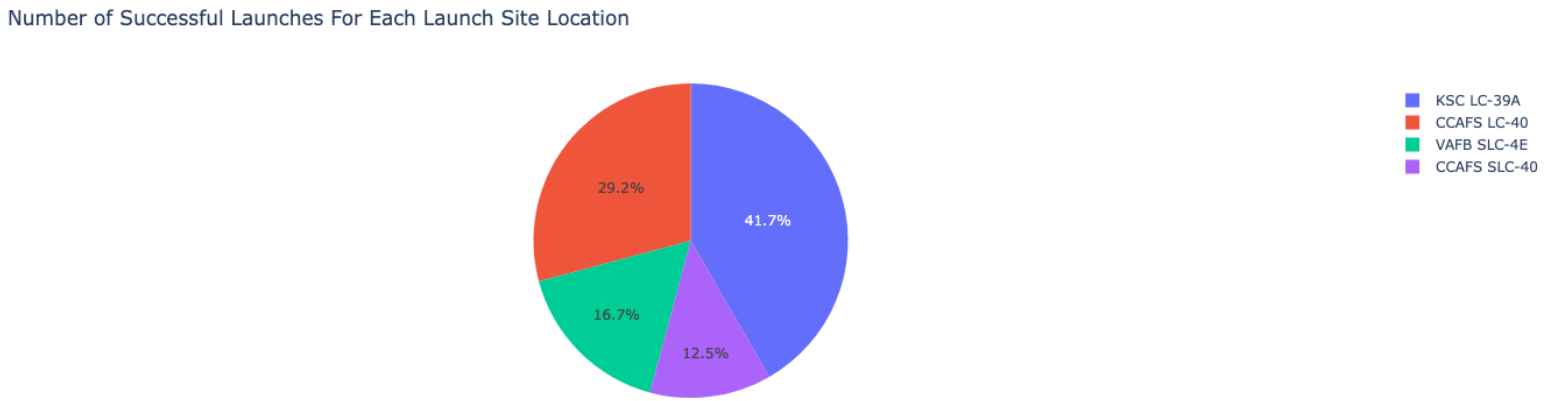
SPACEX FOLIUM RESULTS



Similar example with the two launch locations in Florida as well. Unfortunately, I was not able to get the PolyLine functionality working with my system, as you can see with the stray blue line on the map. However, I can make these initial impressions:

- Launch locations seem to be within a mile of a highway, and about two miles from a railroad. However, they seem to be pretty far from cities, which makes sense for safety reasons.

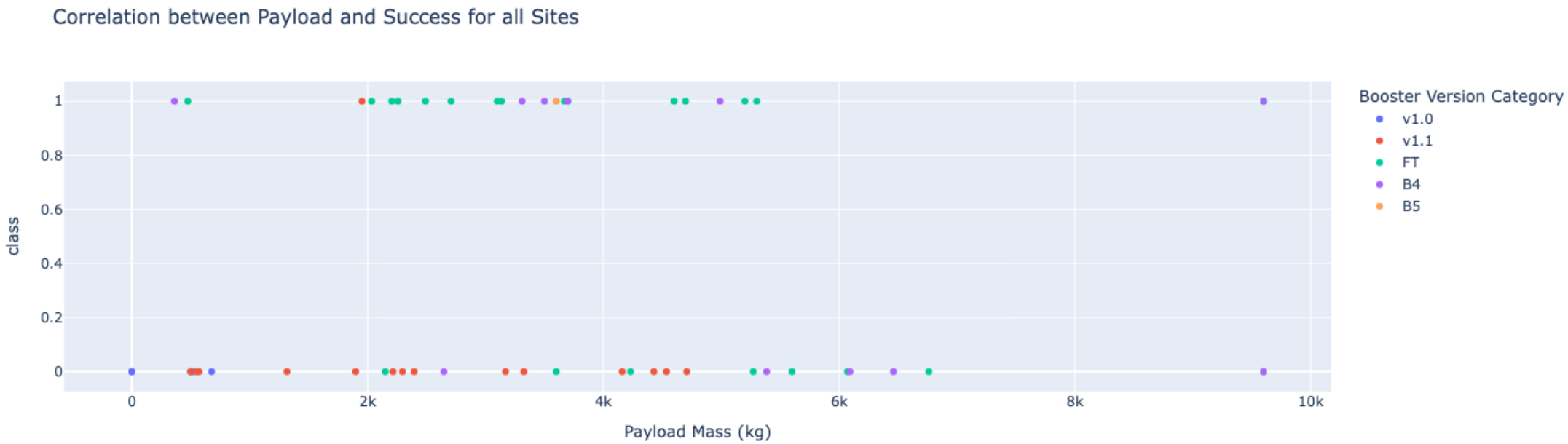
SPACEX DASH RESULTS



From this pie chart, we can clearly see that the KSC LC-39A location has the most successful launches out of all locations, followed by the VAFB SLC-4E location.

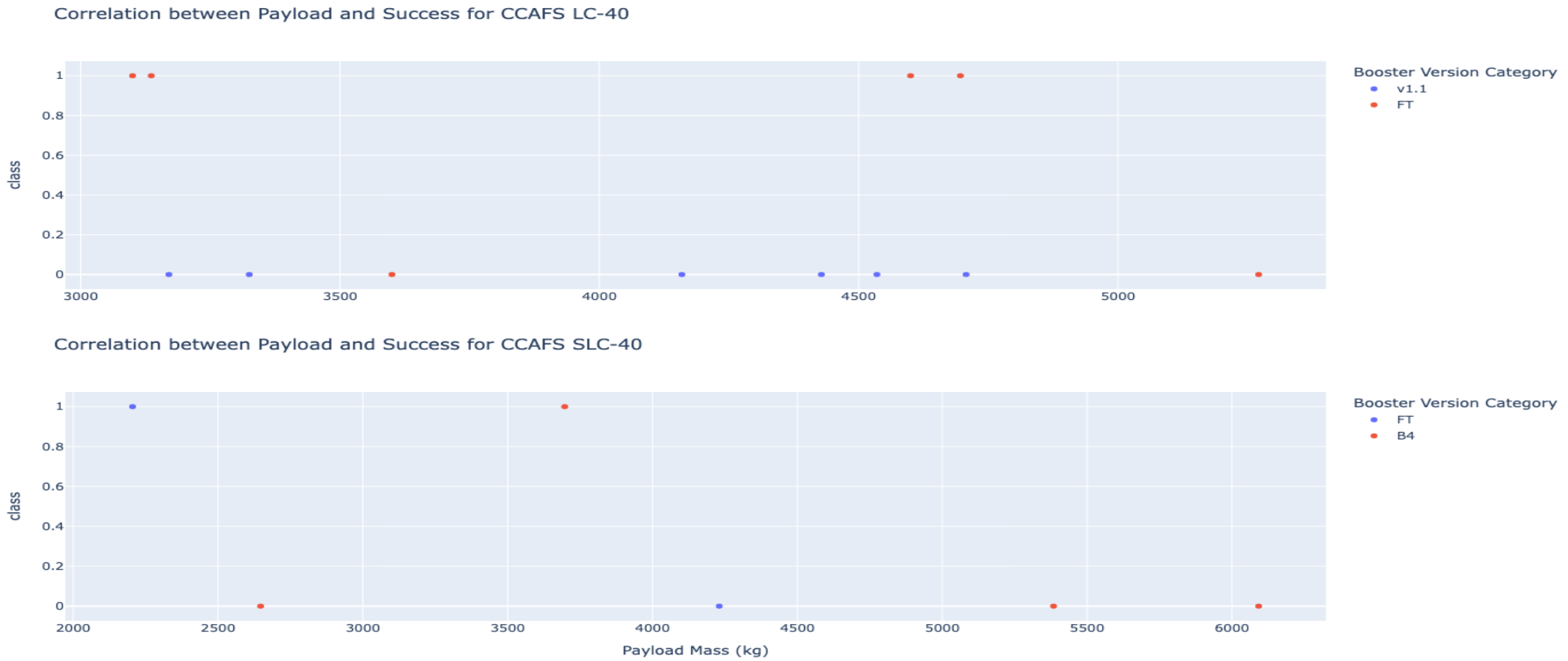
After generating pie charts for each location, we can see that the KSC LC-39A has the highest landing success rate out of all sites at 76.9%, followed closely by CCAFS LC-40 at 73.1% success rate, VAFB SLC-4E at 60%, and CCAFS SLC-40 at 57.1%.

SPACEX DASH RESULTS

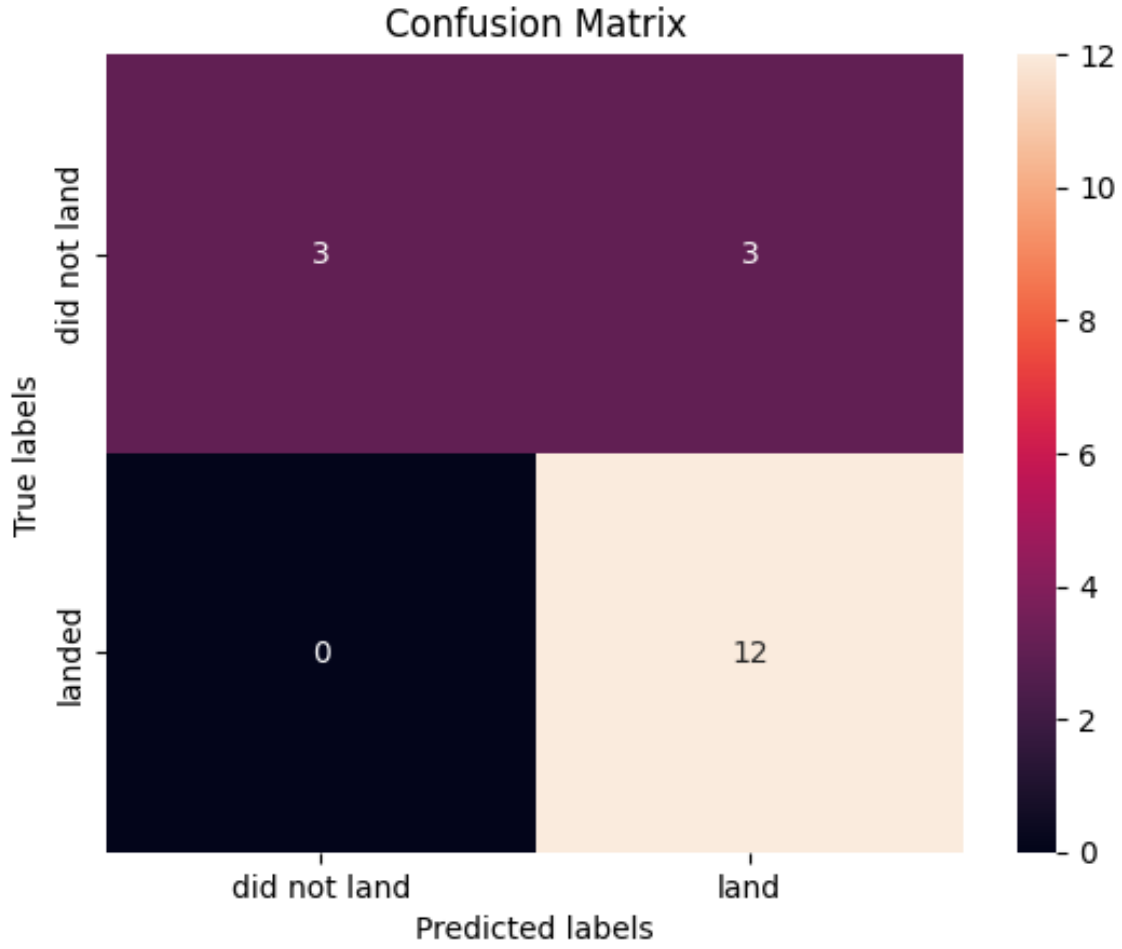


From this graph, we can compare the success of each rocket launch with the booster version category and its payload mass. As we can see generally FT missions are successful, and v1.1 missions are failures. With the help of Plotly and Dash, we can adjust the payload range and specific launch sites as well (next slide).

SPACEX DASH RESULTS



SPACEX PREDICTIVE ANALYSIS RESULTS



After analyzing this data, we have considered implemented several machine learning models to figure out which one is best at predicting successful landings given the payload mass, flights, orbit, number of grid fins, and other such attributes. The models that we have considered training along with their accuracies are the following:

- Logistic Regression (training: 0.846, testing: 0.833)
- SVM (training: 0.848, training: 0.833)
- Decision Trees (training: 0.877, testing: 0.833)
- K-nearest neighbors (training: 0.848, testing: 0.833)

Since all models performed the same on the testing data, the confusion matrices of all models can be found to the LHS. However,

CONCLUSION

- SpaceX is a very successful rocket-launching company due to many factors, including:
  - Their one-of-a-kind rocket reusability launching technique, which saves dramatically on costs
  - An overall 67% successful rate and a success rate that is over 80% in the year 2020 with a positive trendline
  - Their unique locations near the Atlantic and Pacific coastlines that are also near NASA government facilities
- Many space launches are going from a GTO or VSS orbit to ones like VLEO due to a higher success rate.
- The most successful launch site is also the same place with the most successful launching percentage, that being the KSC LC-39A launch site.
- In terms of predictive analysis, an SVM model seems to predict successful launches the best with an overall training accuracy of 88% and testing accuracy of 83%.
- In order for SpaceY to beat SpaceX, we suggest utilizing the many techniques that SpaceX currently uses that enabled them to be the world's most renowned rocket launching company, but even better.

REFERENCES

- GitHub repository: https://github.com/vignes-12/spacex-rocket-analysis
- SpaceX Open-source REST API: https://github.com/r-spacex/SpaceX-API
- Wikipedia List of Falcon 9 and Falcon Heavy Launches: https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

THANK YOU!