SpaceX Falcon 9 first Landing Prediction

Lianping Wu
8-8-23

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

- Summary of methodologies
  - Basic data wrangling and formatting
  - Web scrap launch records with BeautifulSoup
  - Exploratory Data Analysis
  - Launch Sites locations analysis with Folium
  - Machine learning Prediction

- Summary of methodologies
  - EDA

    Launch success rate is increasing

    ES-L1, GEO, HEO, SSO are the most successful orbit types
  - Data visualization: most launch sites are near the equator and close to the coast
  - All models performed similarly

Introduction

SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. Therefore if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

Methodology

Data collection and data wrangling methodology

Collect and make sure the data is in the correct format from an API.

- To do some basic data wrangling and forming
  - Request to the SpaceX API
  - Clean the requested data

- Export the data to a csv file for the next section:

https://github.com/lianpingWu/IBM-data-science/blob/main/1-spacex-data-collection-api.ipynb

https://github.com/lianpingWu/IBM-data-science/blob/main/2-webscraping.ipynb

Methodology

EDA and interactive visual analytics methodology

Convert those outcomes into Training Labels with `1` means the booster successfully landed `0` means it was unsuccessful.

- Exploratory Data Analysis
- Determine Training Labels

- Export the data to a csv file for the next section:

https://github.com/lianpingWu/IBM-data-science/blob/main/3-spacex-Data%20wrangling.ipynb

Methodology

EDA with visualization results

Perform Exploratory Data Analysis and Feature Engineering

- Exploratory Data Analysis
- Preparing Data Feature Engineering

- Export the data to a csv file for the next section:

https://github.com/lianpingWu/IBM-data-science/blob/main/5-eda-dataviz.ipynb

Methodology

Interactive map with Folium results

- **The launch success rate may depend on many factors. may also depend on the location and proximities of a launch site. Finding an optimal location for building a launch site certainly involves many factors and hopefully we could discover some of the factors by analyzing the existing launch site locations.**

- Mark all launch sites on a map
- Mark the success/failed launches for each site on the map
- Calculate the distances between a launch site to its proximities

After completed the above tasks, you should be able to find some geographical patterns about launch sites.

Methodology

Build a Dashboard with Plotly Dash

- Allow users to select specific launch sites using a dropdown
- Allow users to select payload mass range using slider
- Pie chart displaying ratio of successful/failed launches
- Rendering scatterplot showing success rate vs. payload mass separated out by booster version

https://github.com/lianpingWu/IBM-data-science/blob/main/7-dash_interactivity.py

Methodology

Predictive analysis (classification)

- **Perform exploratory Data Analysis and determine Training Labels***
  - create a column for the class*
  - Standardize the data*
  - Split into training data and test data\-Find best Hyperparameter for SVM, Classification Trees and Logistic Regression*
  - Find the method performs best using test data

https://github.com/lianpingWu/IBM-data-science/blob/main/8-Machine_Learning_Prediction_Part_5.ipynb

Results

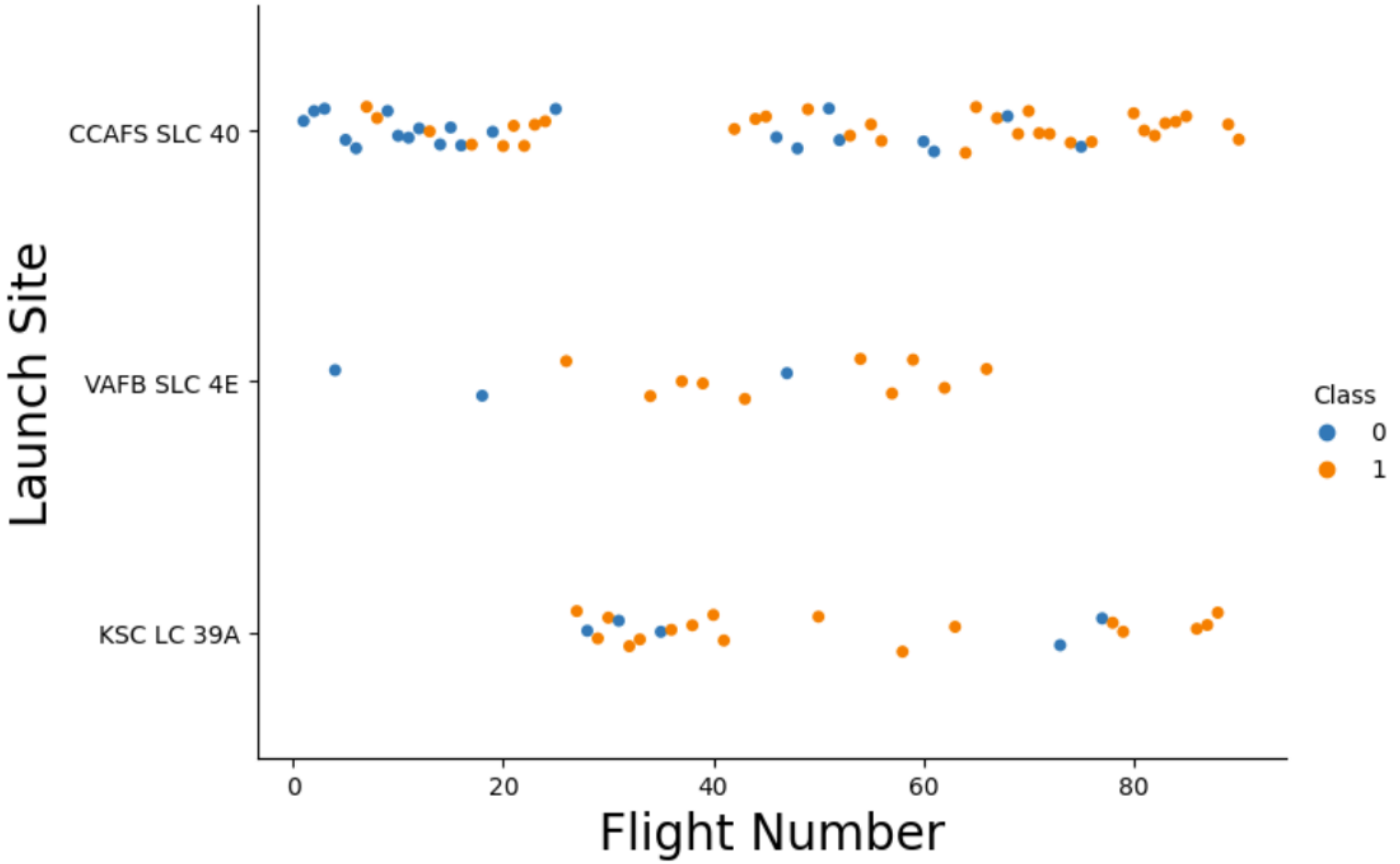
Exploratory data analysis results

- Launch success is steadily improving
- KSC-LC39A has the highest success rate of all the sites (~77%)
- Orbits ES-L1, GEO, HEO, SSO have 100% success rate

Results

Exploratory data analysis results

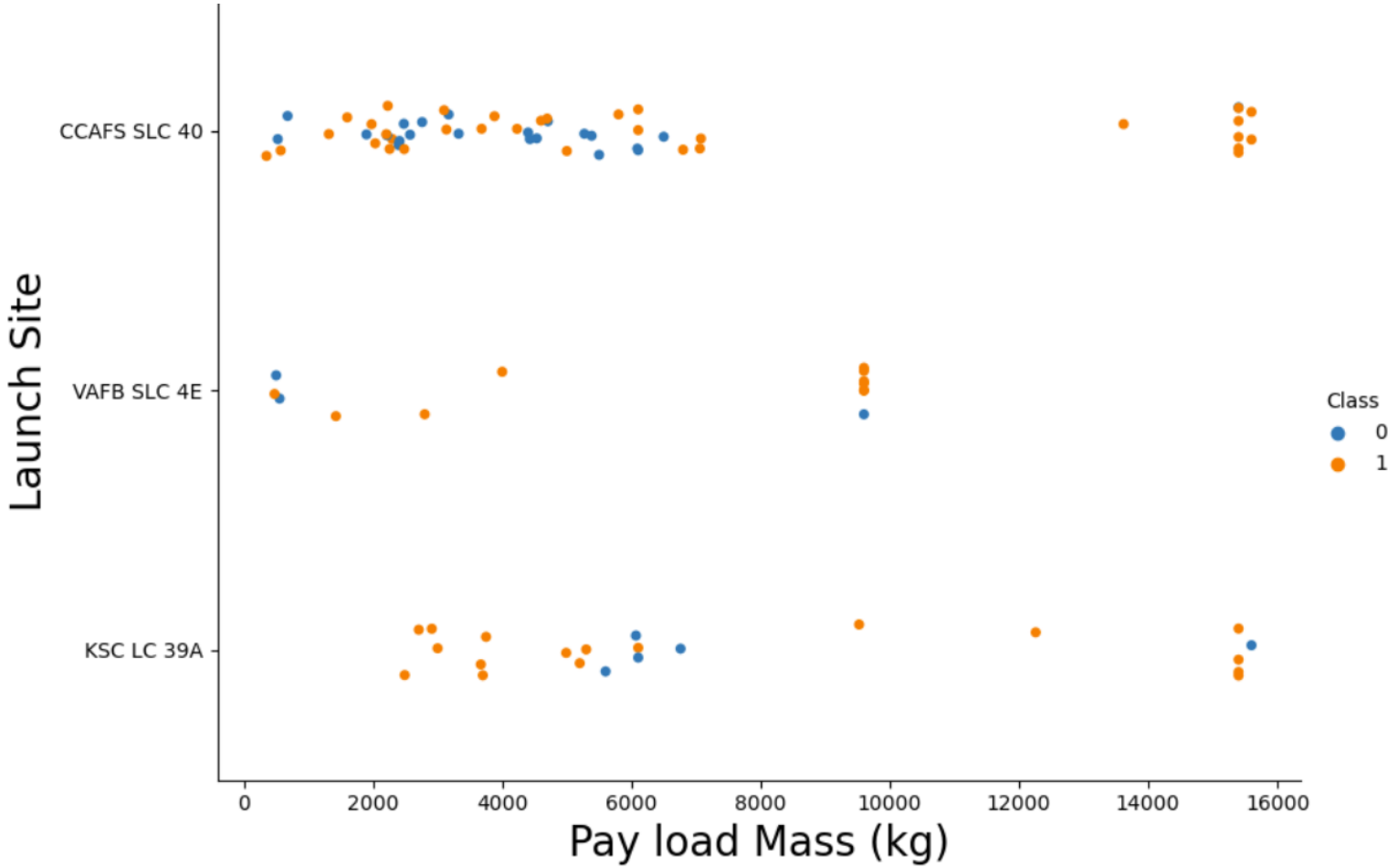
Flight Number vs. LaunchSite



Results

Exploratory data analysis results

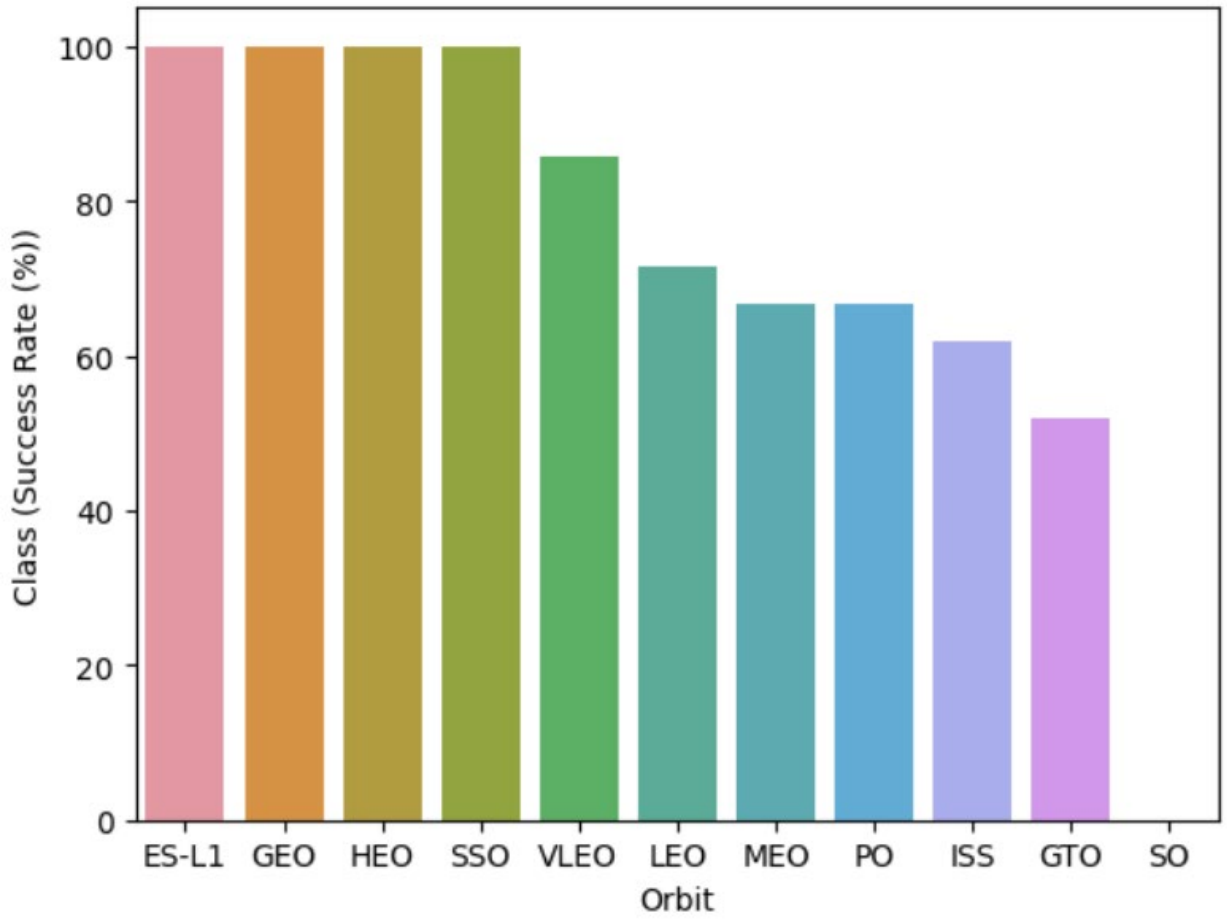
Payload vs. LaunchSite



Results

Exploratory data analysis results

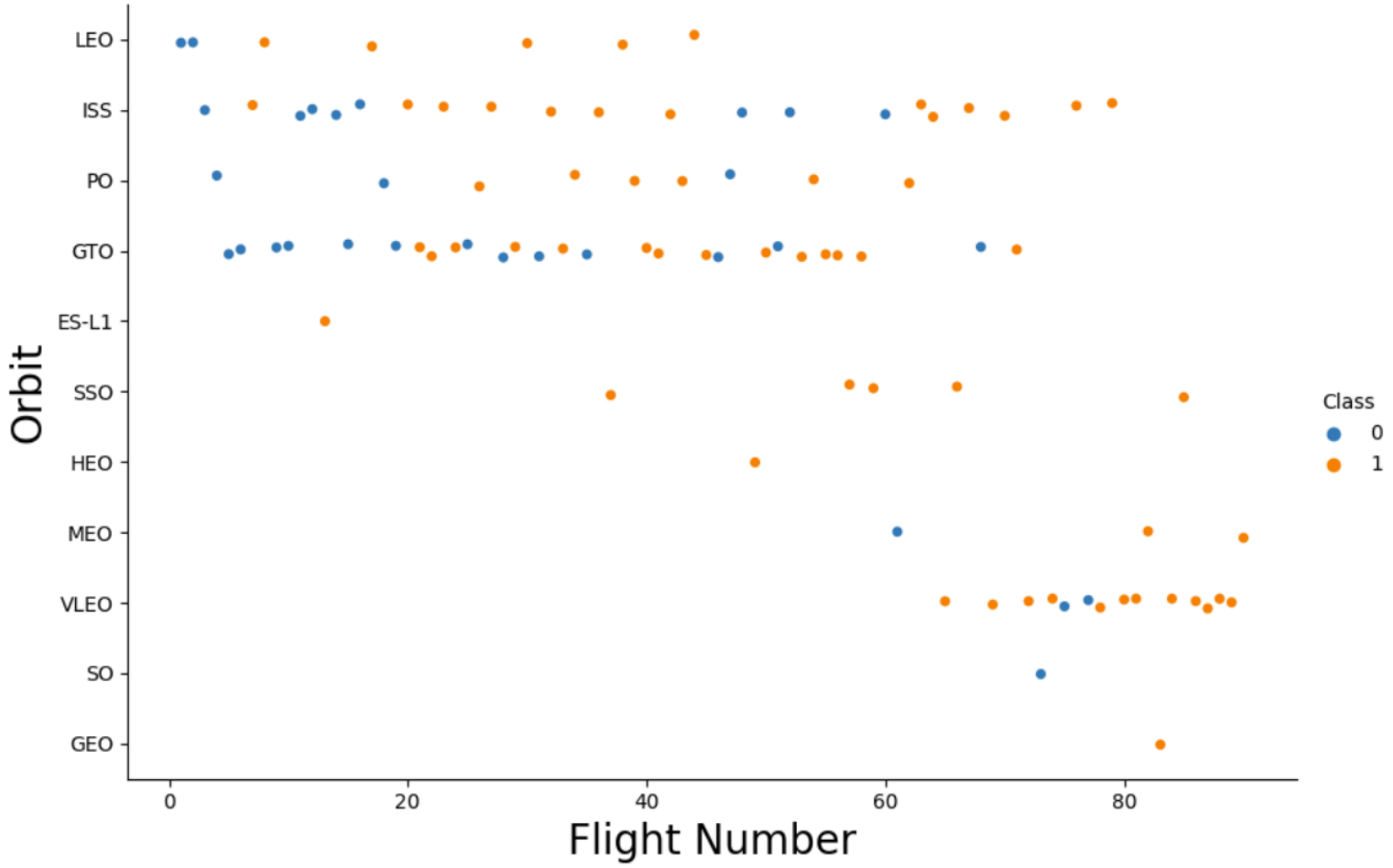
Success Rate vs. Orbit Type



Results

Exploratory data analysis results

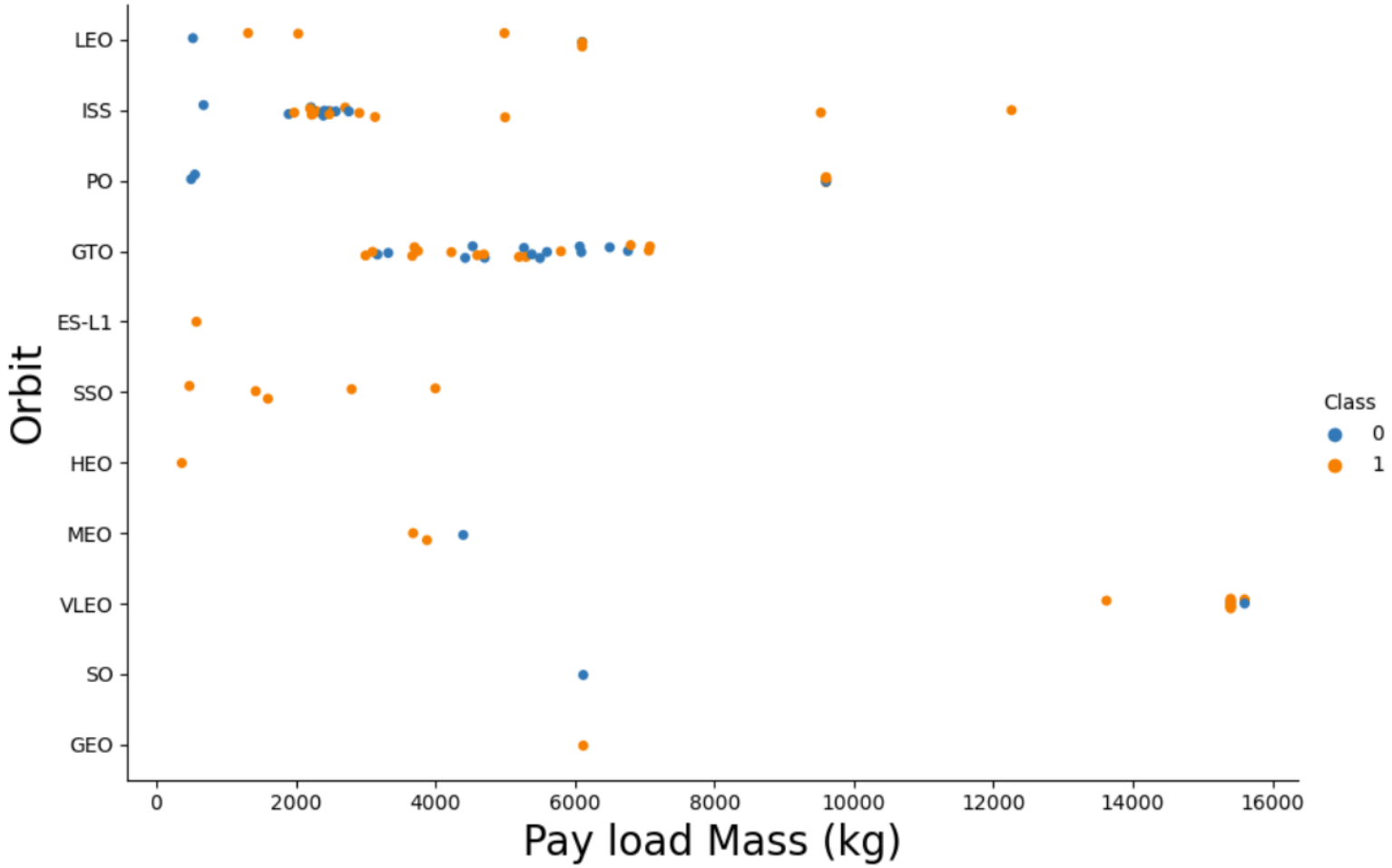
Flight Number vs. Orbit Type



Results

Exploratory data analysis results

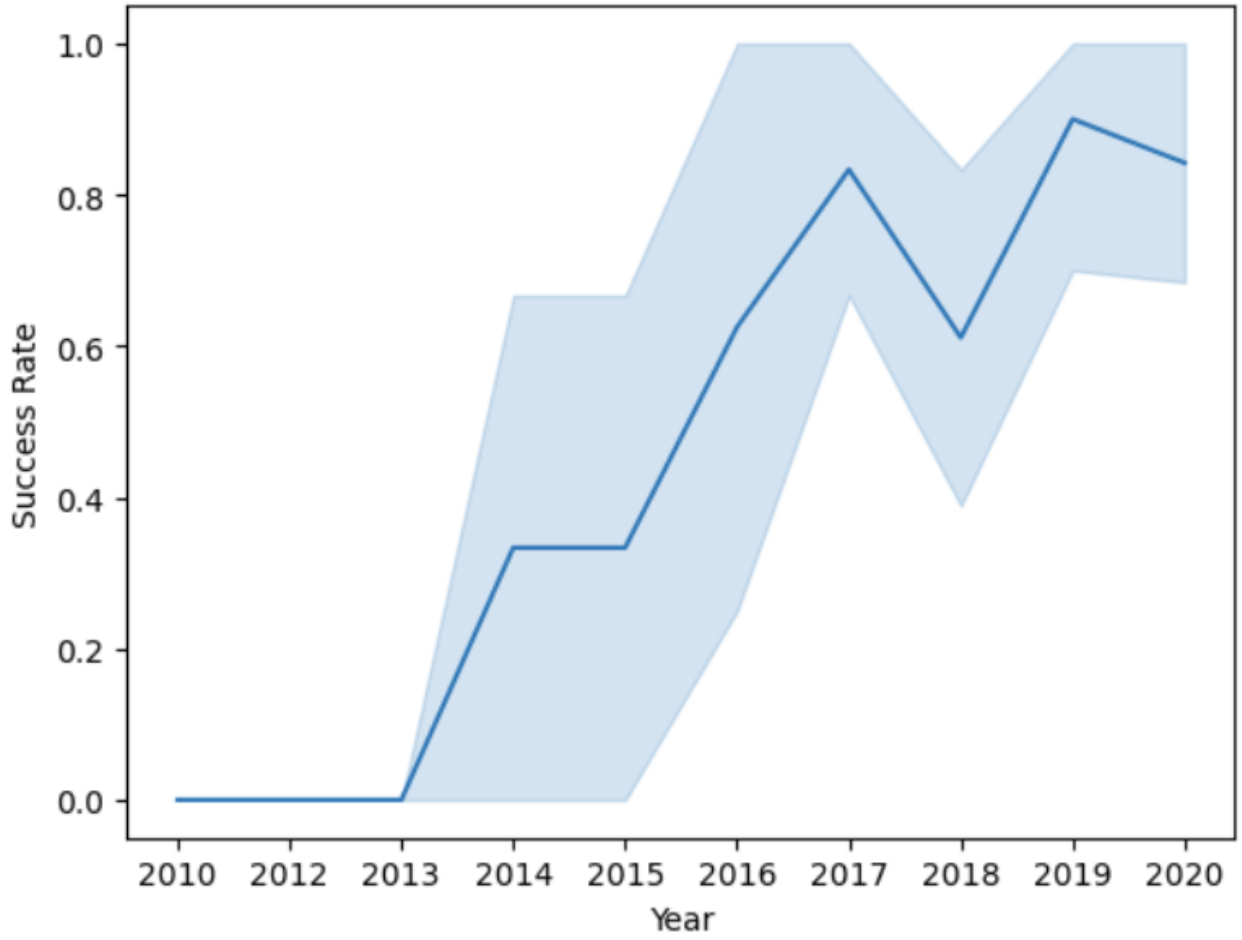
Payload vs. Orbit Type



Results

Exploratory data analysis results

Launch Success Yearly Trend



Results

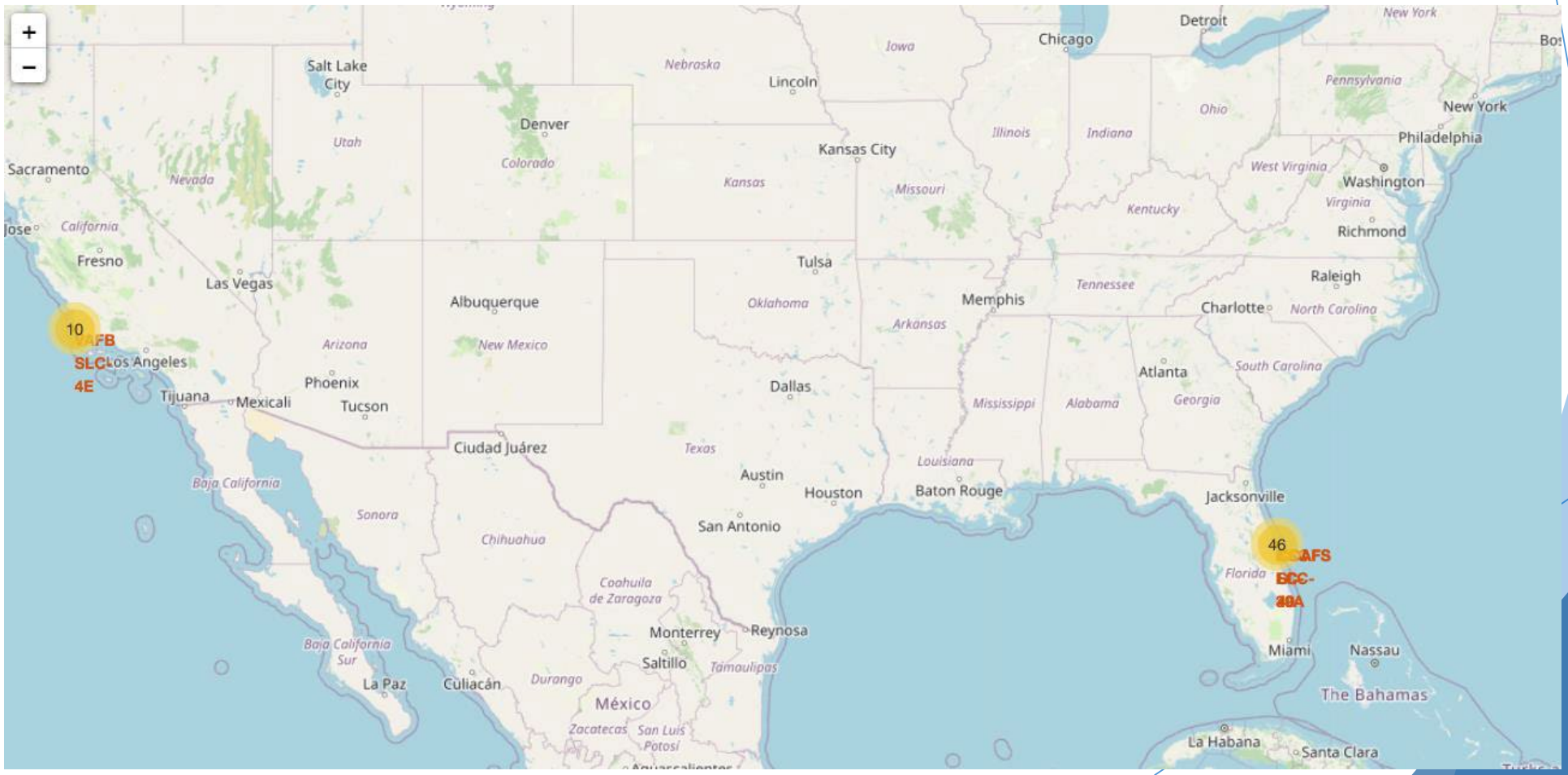
Interactive analytics demo in screenshots

- Most sites are near the equator, close to the coast, and far away from pieces of human infrastructure

Results

Interactive analytics demo in screenshots

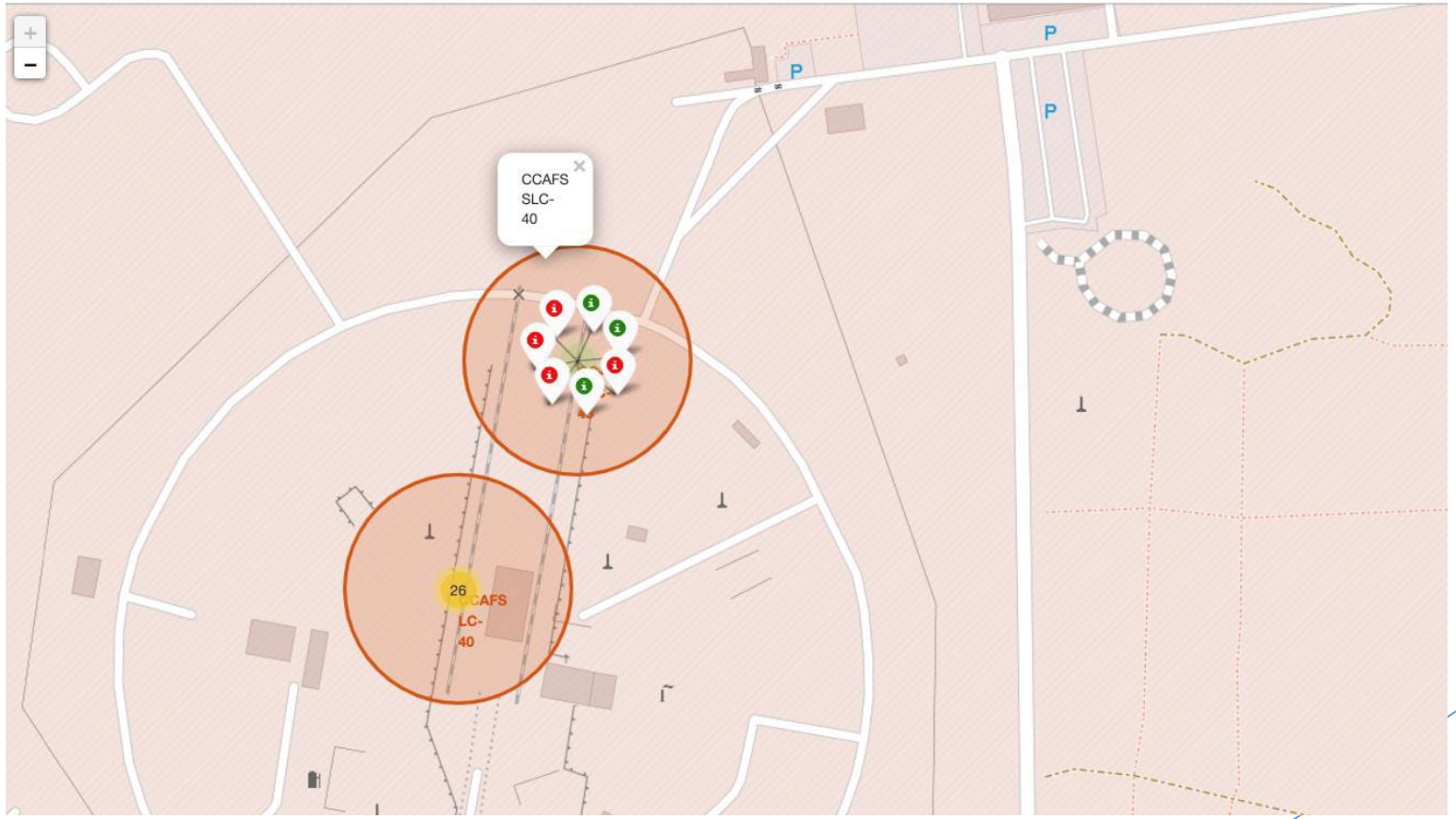
The launch sites are near coasts, which makes sense. It’s better to have debris potentially fall into the ocean than on land.



Results

Interactive analytics demo in screenshots

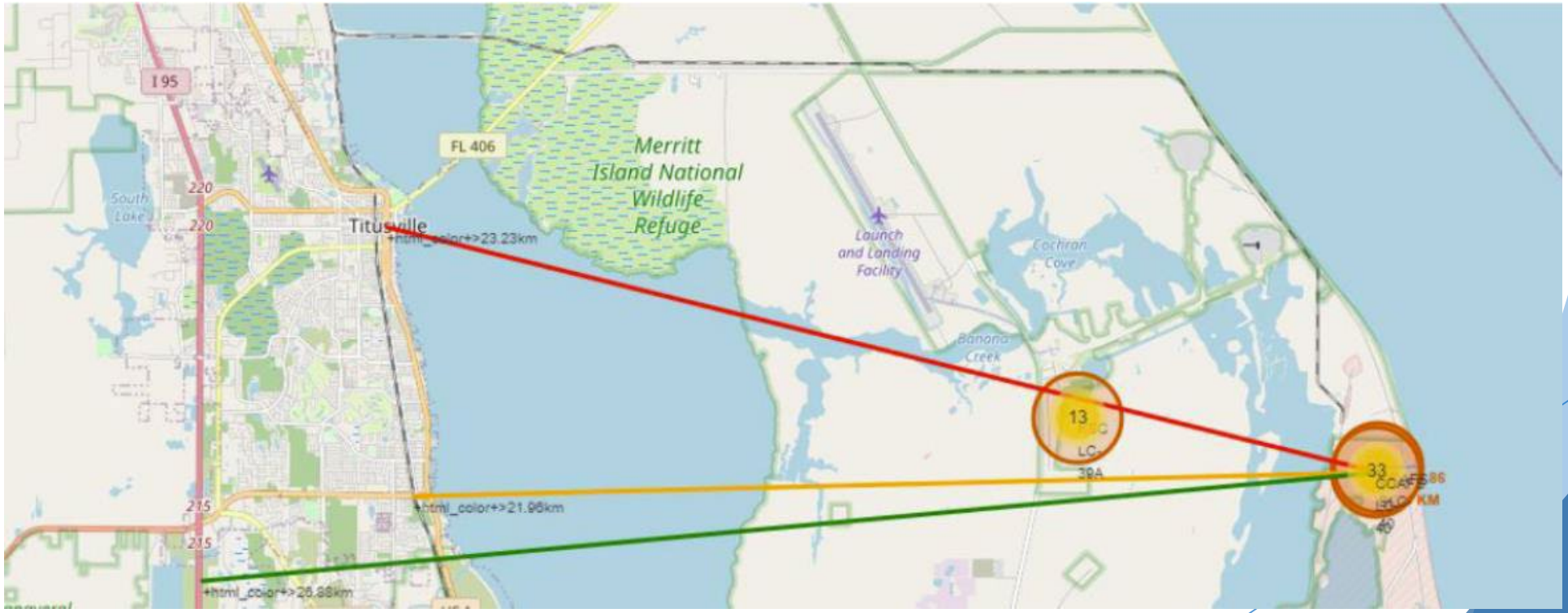
- Green markers denote successful launches, red markers denote unsuccessful launches.
- Site CCAFS SLC-40 has a success rate of 3/7 (42.9%)



Results

Interactive analytics demo in screenshots

- 86 km from nearest coastline;
- 21.96 km from nearest railway
- 23.23 km from nearest city
- 26.88 km from nearest highway



Results

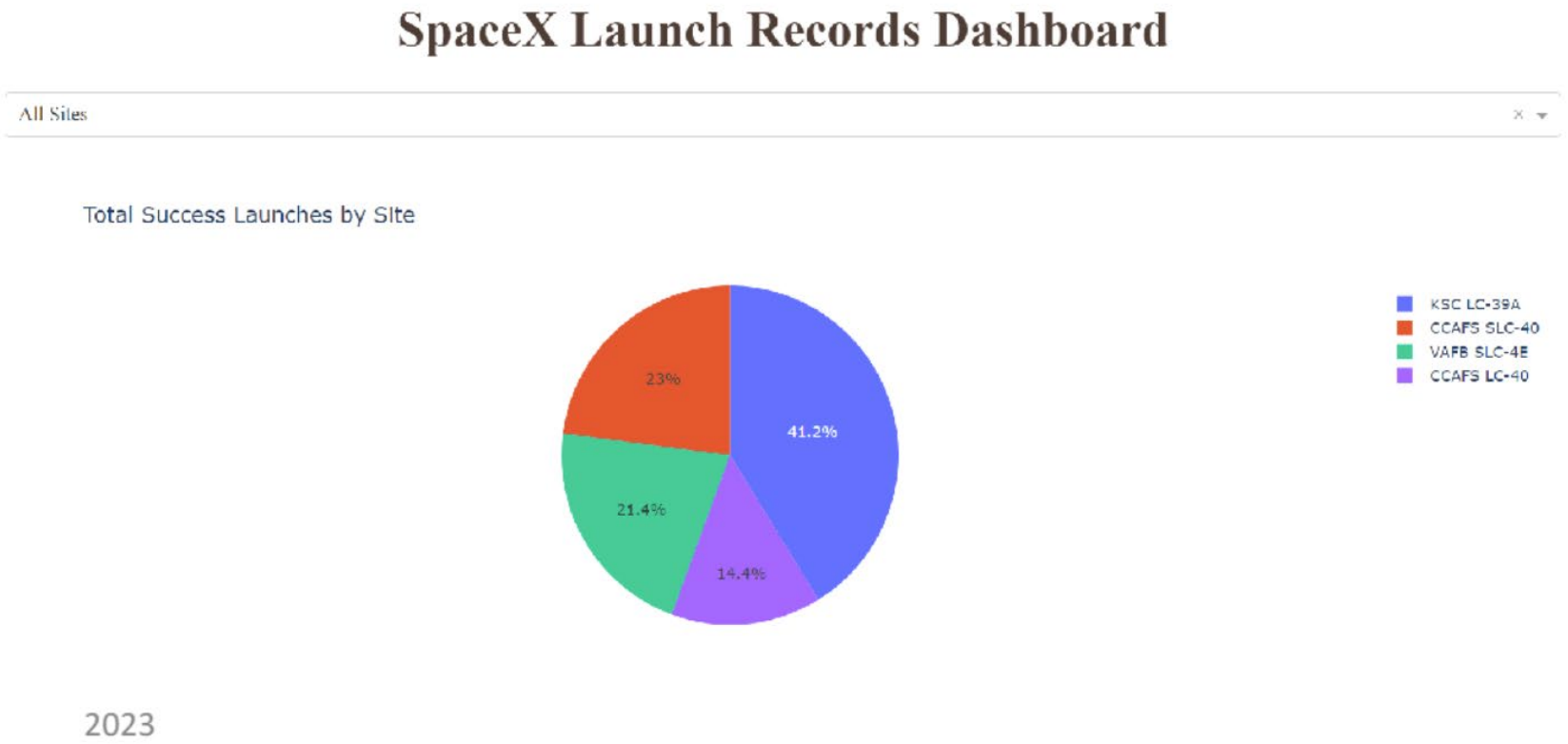
Predictive analysis results

- All models have equal performance

Results

Dashboard plot

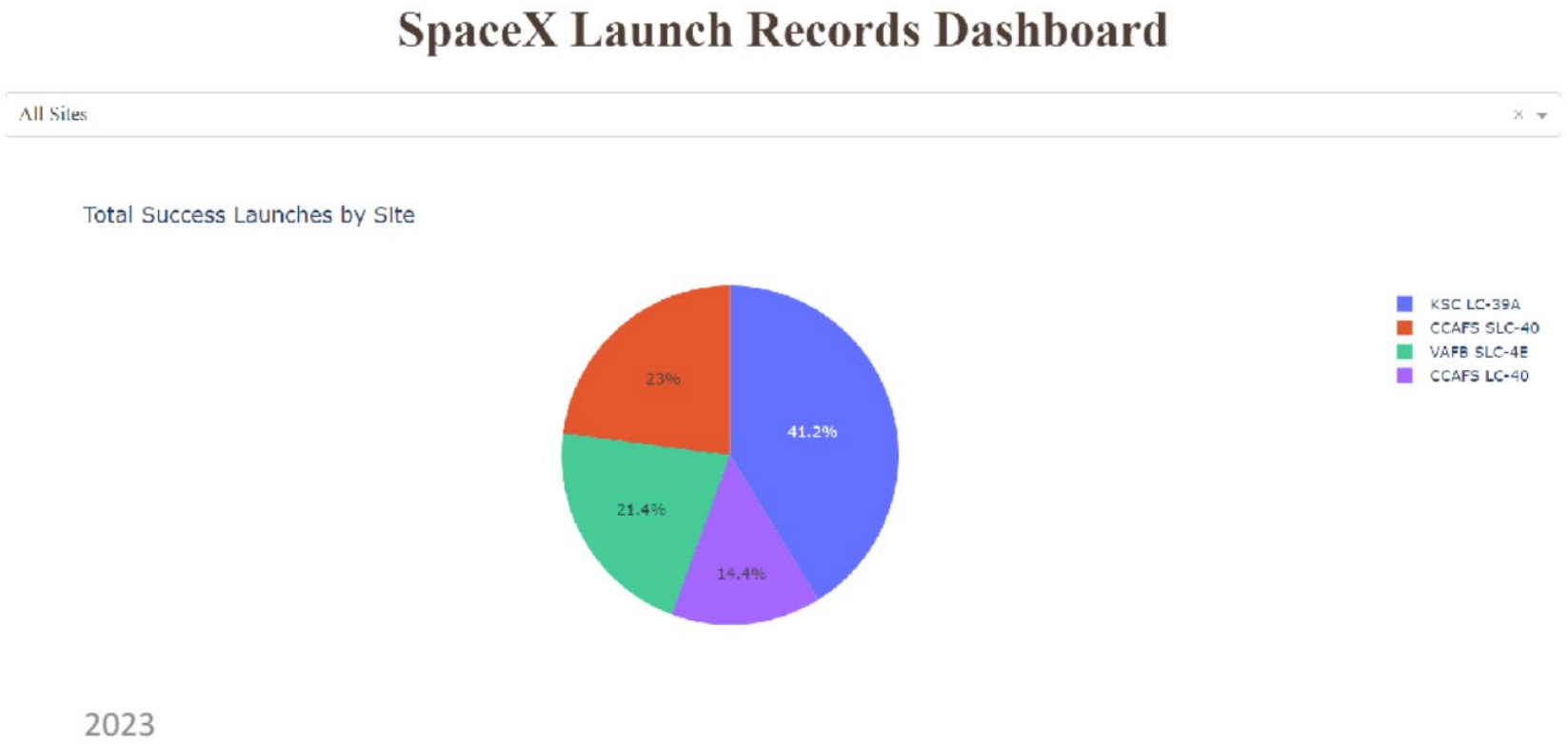
KSC LC39A has produced the most successful launches comparatively (~41%)



Results

Dashboard plot

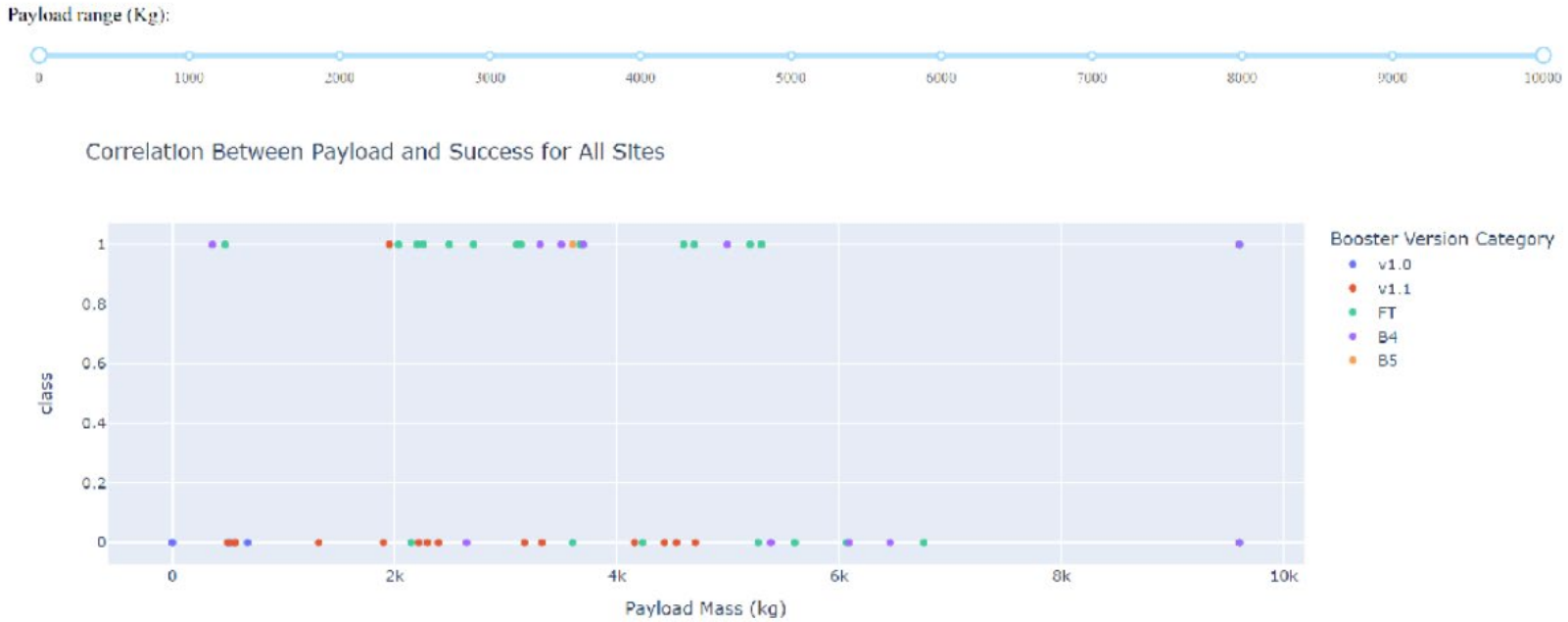
KSC LC39A has produced the most successful launches comparatively (~41%)



Results

Dashboard plot

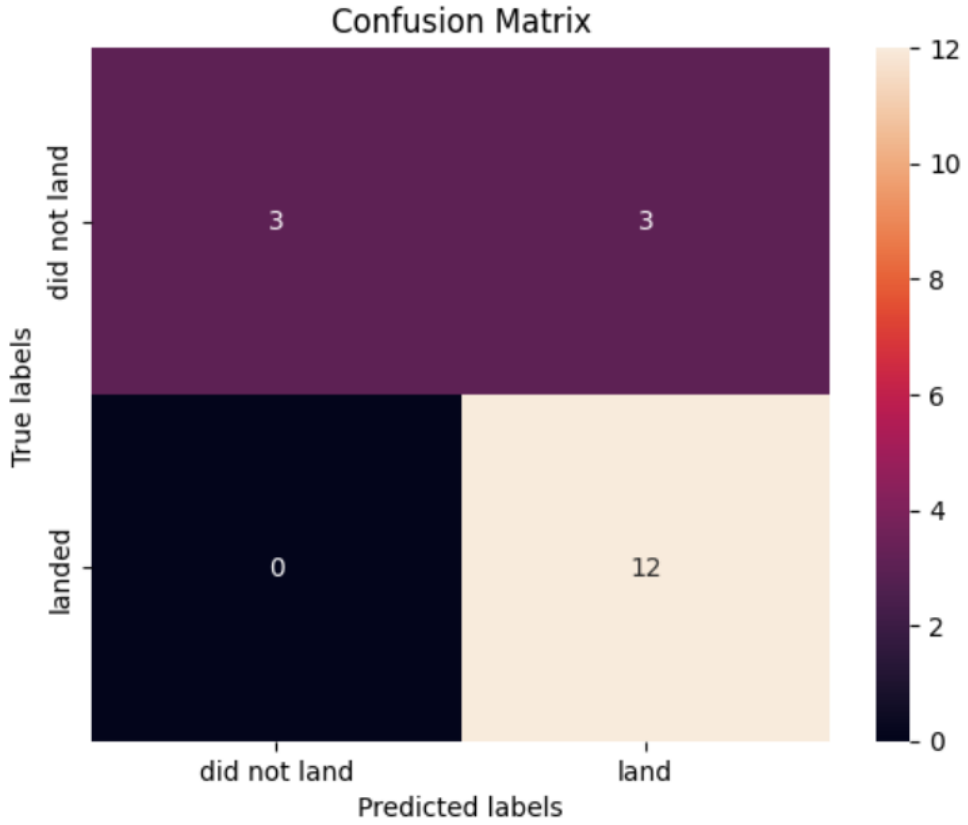
- Payloads between 2000 and 5000kg have the highest success rates.
- (1 indicates success, 0 indicates failure)



Results

Predictive analysis results

All four models had the exact same confusion matrix.
There were no false negative errors, only false positives.



All models have equal accuracy

Method	Test Data Accuracy
Logistic_Reg	0.833333
SVM	0.833333
Decision Tree	0.833333
KNN	0.833333

Conclusion

- **Exploratory data analysis results**
  - Launch success is steadily improving
  - KSC-LC39A has the highest success rate of all the sites (~77%)
  - Orbits ES-L1, GEO, HEO, SSO have 100% success rate

- **Interactive analytics demo in screenshots**
  - Most sites are near the equator, close to the coast, and far away from pieces of human infrastructure.
  - Launches with payloads below 2000kg and above 5000kg have very little chance of succeeding.

- **Predictive analysis results**
  - All models have equal performance

Thank you!