Analysis of SpaceX Landings

Derjean Grant

03/05/2023

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion
- Conclusion

EXECUTIVE SUMMARY

Commercial Space Travel will be an alluring new product.

SpaceX has distinguished itself among competitors by:

1. Being the first to offer the service in 2020
2. Using technology that allows lower rocket costs
3. Partnering with NASA

This technology is their rockets ability to recover the 1st Stage after launching which indicates a successful launch.

The success rate of SpaceX launchings have been trending upwards.

INTRODUCTION

- Commercializing space travel is viable only with reduced cost. Rocket prices are easily in the region of $160 Million USD
- SpaceX has managed for their flagship rocket, the Falcon 9 to cost around $60 Million USD
- The major cost reduction is possible through the rocket's ability to recover the First Stage once successfully launched.
- This report seeks to evaluate the success of the Falcon 9 launches by:
    I. Observing the success of Falcon 9 launches
    II. Predicting the success of Falcon 9 launches

METHODOLOGY
Data Sources

There were two data sources:
SpaceX Website & Wikipedia

Data was collected from the SpaceX website: 'https://api.spacexdata.com/v4/launches/past' using an application programming interface (API).

Falcon 9 being the focus of our study, the data required filtering to limit the results.

Missing data points for the Pay Load Mass KG was encountered and replaced with the average value.

METHODOLOGY
Data Sources

There were two data sources:
SpaceX Website & Wikipedia

SpaceX data from the Wikipedia website: 'https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches' was collected via Webscraping.

The table containing details of Falcon 9 launches was extracted from the webpage. There were no missing values.

METHODOLOGY- Data Sources

- There was an additional data source collected as an Excel file. This file was analyzed using SQL from which several insights into SpaceX launches were garnered.
- Further explorations of the Excel file was done using Python to investigate the relationship among features in the data using Bar charts and scatterplots
- More processing of the data was done to convert dummy variables to categorical variables. Successful launches labelled as 1 and unsuccessful launches a 0 in a new column "Class"

METHODOLOGY

Having collected the data

Preparing &Pre-processing the data for analysis

Two analytical methods were applied

- Interactive Dashboard
- Machine Learning Classfication

METHODOLOGY- Analytics

- The Interactive dashboard was created for a real- time depiction of the relationship between variables of the data and to show the distribution of successful launches
- To predict successful SpaceX launches, several machine learning classification algorithms were applied to the data:
  - Logistic Regression
  - Support Vector Machine
  - KNN
  - Decision Tree Classifier

METHODOLOGY- Analytics

For all the classification techniques applied to the data:

- The target variable for our analysis is the success of launches which was transformed into a categorical variable 'Class' with values: 0 & 1
  - Success labelled 1
  - Failure labelled 0
- The data was standardized to ensure a consistent value range
- The data was split into a test and training set
- Models were trained on the training set and the model parameters were tuned for optimization using GridSearchCV, and then tested on the test set.

RESULTS: Exploratory Data Analysis in SQL

EDA in SQL revealed the following:

1. SpaceX has four launch stations:
   - CCAFS LC-40
   - VAFB SLC-4E
   - KSC LC-39A
   - CCAFS SLC-40
2. There are different versions of the Falcon 9(F9) identified by its Booster Versions, for which have different Payload Mass carrying capacities:
   - F9 v1.1 has an average payload mass of 2928.4KG

RESULTS: Exploratory Data Analysis in SQL

EDA in SQL revealed the following:

3. The maximum payload mass is 15,600KG which can be carriend by the versions:

F9 B5 B1048.4, F9 B5 B1049.4, F9 B5 B1051.3,

F9 B5 B1056.4, F9 B5 B1048.5, F9 B5 B1051.4,

F9 B5 B1049.5, F9 B5 B1060.2, F9 B5 B1058.3,

F9 B5 B1051.6, F9 B5 B1060.3, F9 B5 B1049.7

4. The earliest date for a successful launch was:

First_Successful_Groundpad_Landing
01-05-2017

RESULTS: Exploratory Data Analysis in SQL

EDA in SQL revealed the following:

5. F9 Boosters which have success in drone ship launches and have payload mass greater than 4000KG but less than 6000KG:

Booster_Version	PAYLOAD_MASS__KG_
F9 FT B1022	4696
F9 FT B1026	4600
F9 FT B1021.2	5300
F9 FT B1031.2	5200

RESULTS: Exploratory Data Analysis in SQL

EDA in SQL revealed the following:

6. The number of successful landing outcomes between 2016 and 2018:

Date	Landing _Outcome	Number_of_Succesful_ Landings
18-07-2016	Success (ground pad)	6
08-04-2016	Success (drone ship)	8
07-08-2018	Success	20

RESULTS: Exploratory Data Analysis in SQL

EDA in SQL revealed the following:

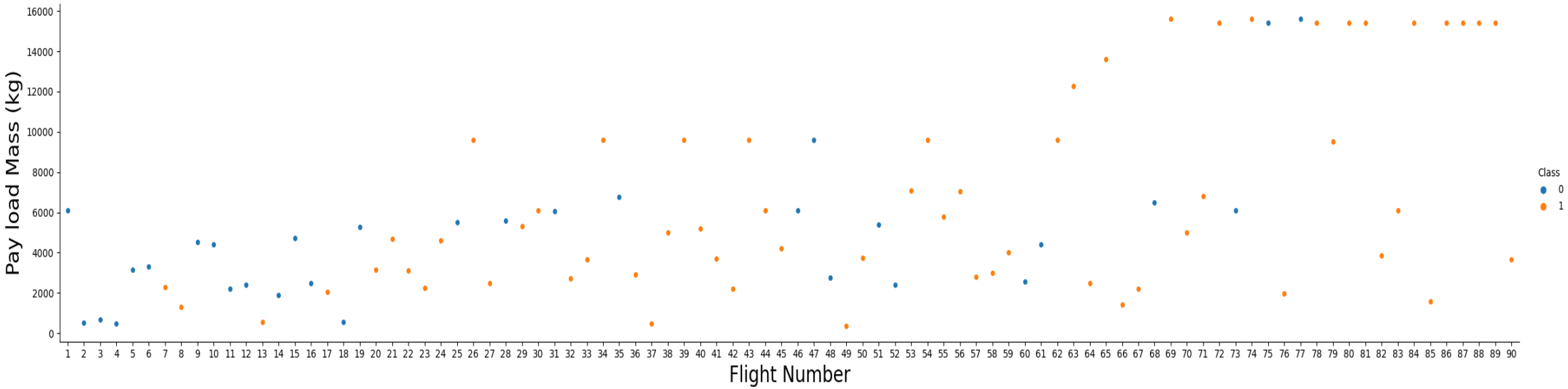
7. Total number of successful and failure mission outcomes.

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

RESULTS:
Exploratory Data Analysis with Visualization

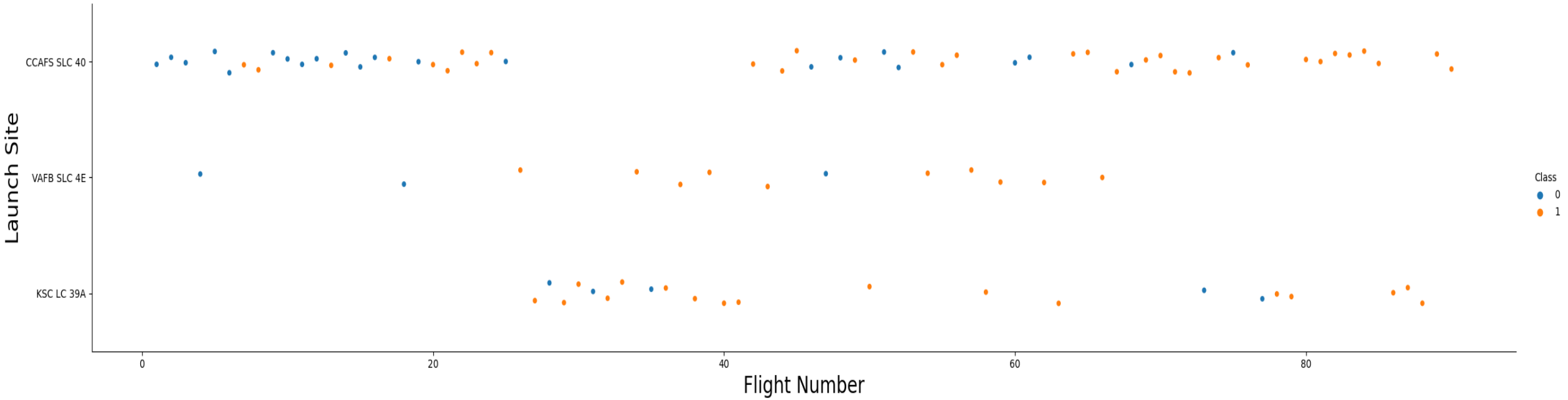
Using Python to evaluate the relationship among variables: The following were of interest:

- Flight Number and Payload Mass
- Launch Site and Flight Number
- Launch Site and Payload Mass
- Success Rate and Orbit
- Flight Number and Orbit
- Orbit and Payload Mass



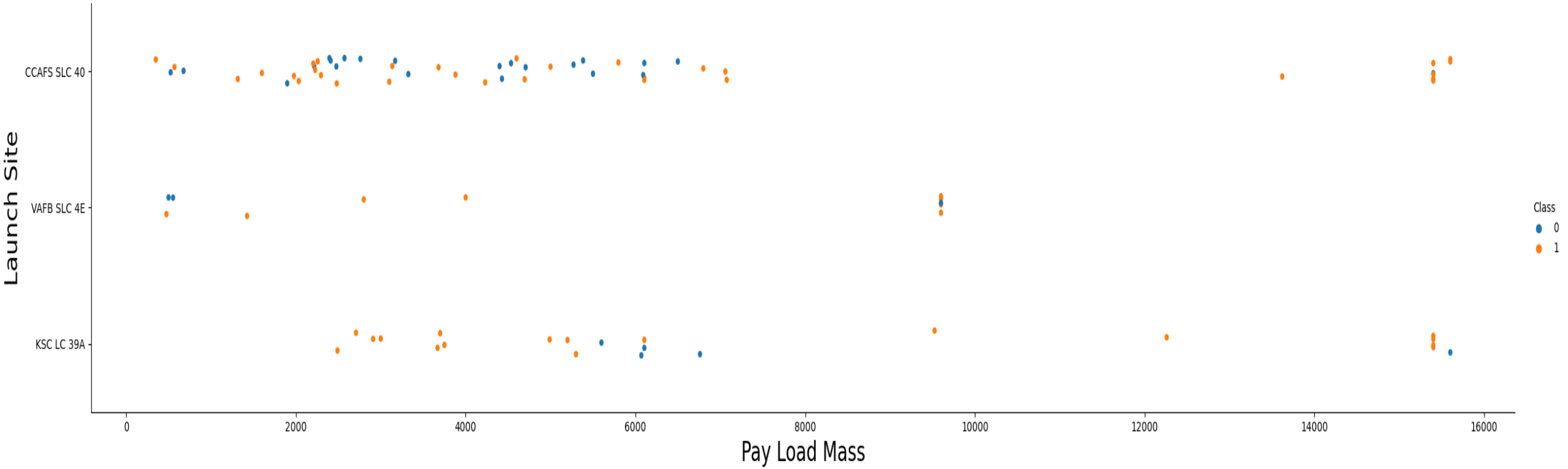
RESULTS: Exploratory Data Analysis with Visualization

- Relationship between Flight Number and Payload Mass



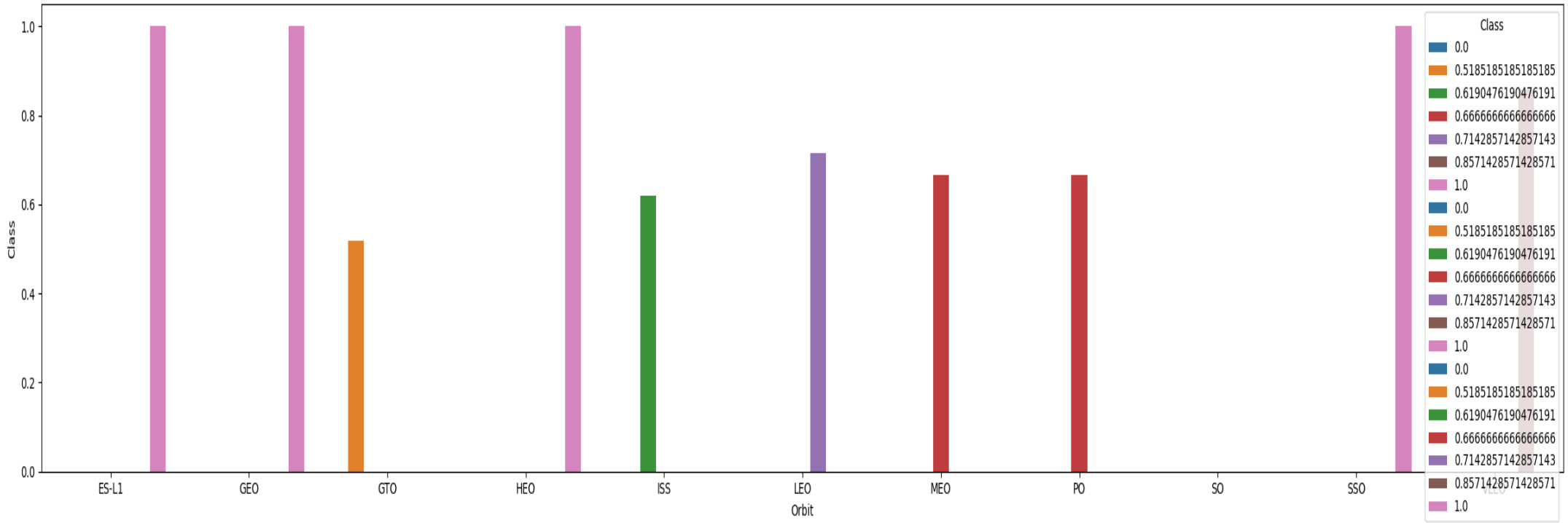
RESULTS: Exploratory Data Analysis with Visualization

- Relationship between Launch Site and Flight Number



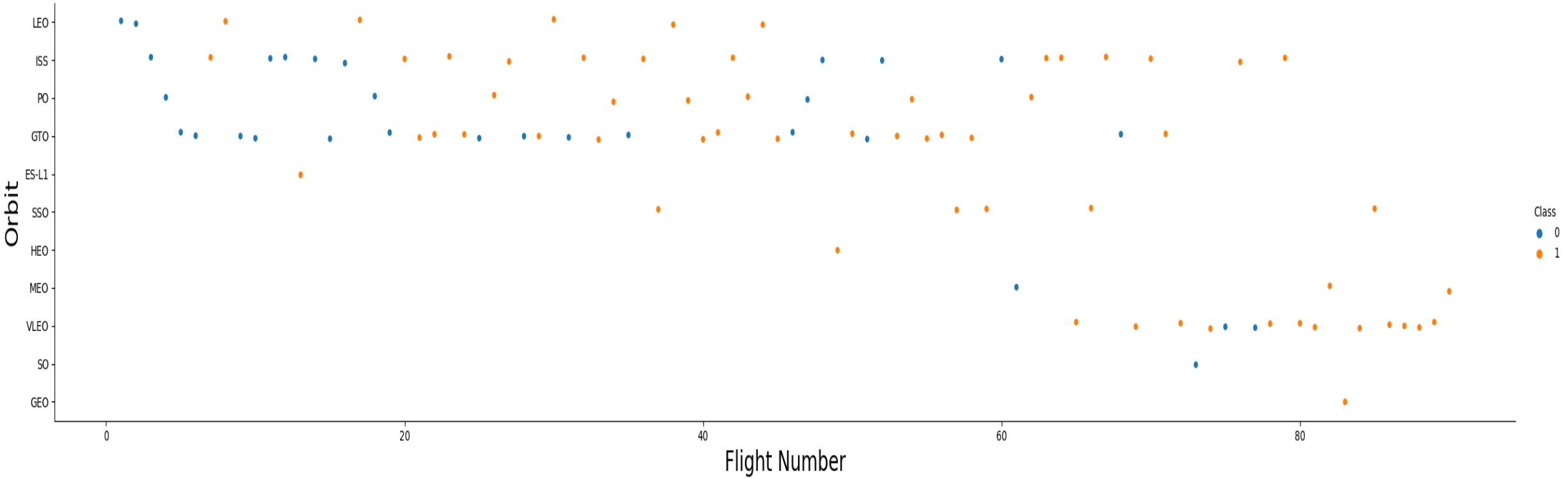
RESULTS: Exploratory Data Analysis with Visualization

- **Relationship between Launch Site and Payload Mass**



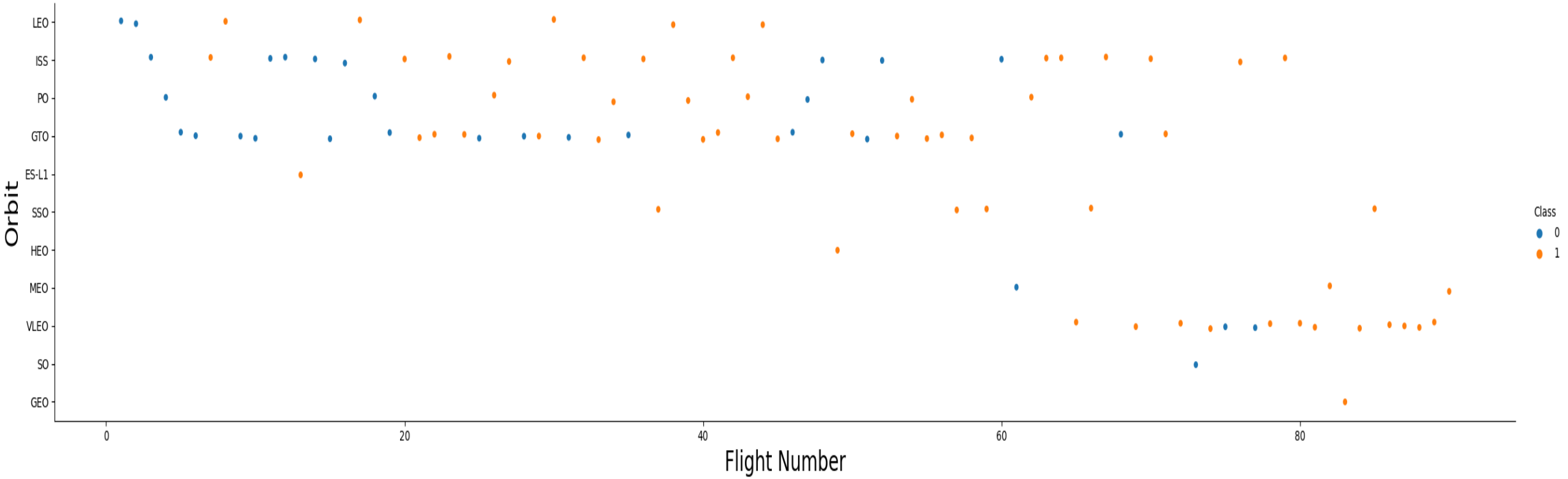
RESULTS: Exploratory Data Analysis with Visualization

- **Relationship between Success Rate and Orbit**



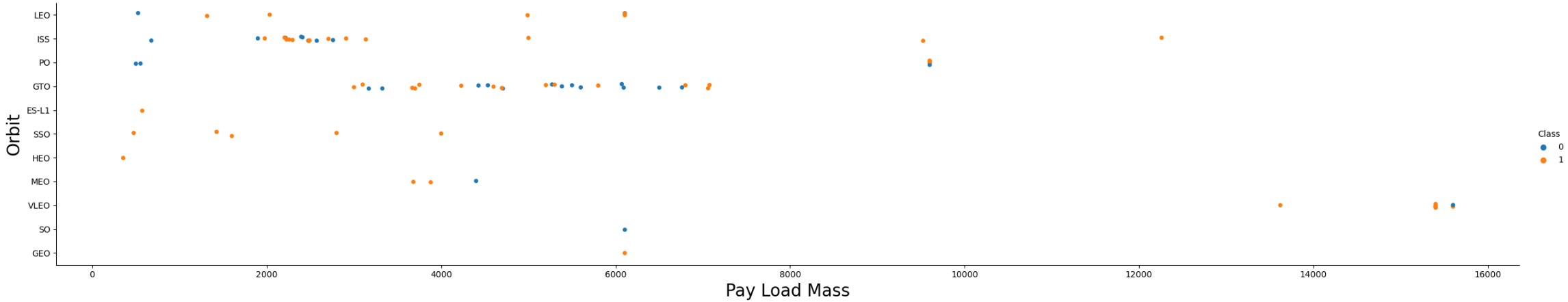
RESULTS: Exploratory Data Analysis with Visualization

- **Relationship between Flight Number and Orbit**



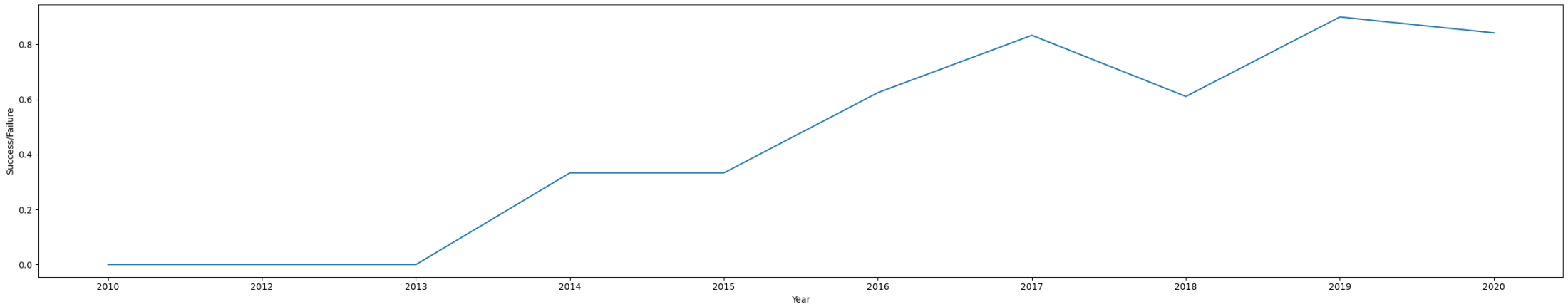
RESULTS: Exploratory Data Analysis with Visualization

- **Relationship between Flight Number and Orbit**



RESULTS: Exploratory Data Analysis with Visualization

- Relationship between Orbit and Payload Mass



RESULTS: Exploratory Data Analysis with Visualization

- Success Rate of Launches over the years

RESULTS: Exploratory Data Analysis with Visualization

- Interactive showing three launch sites created using Folium

RESULTS: Exploratory Data Analysis with Visualization

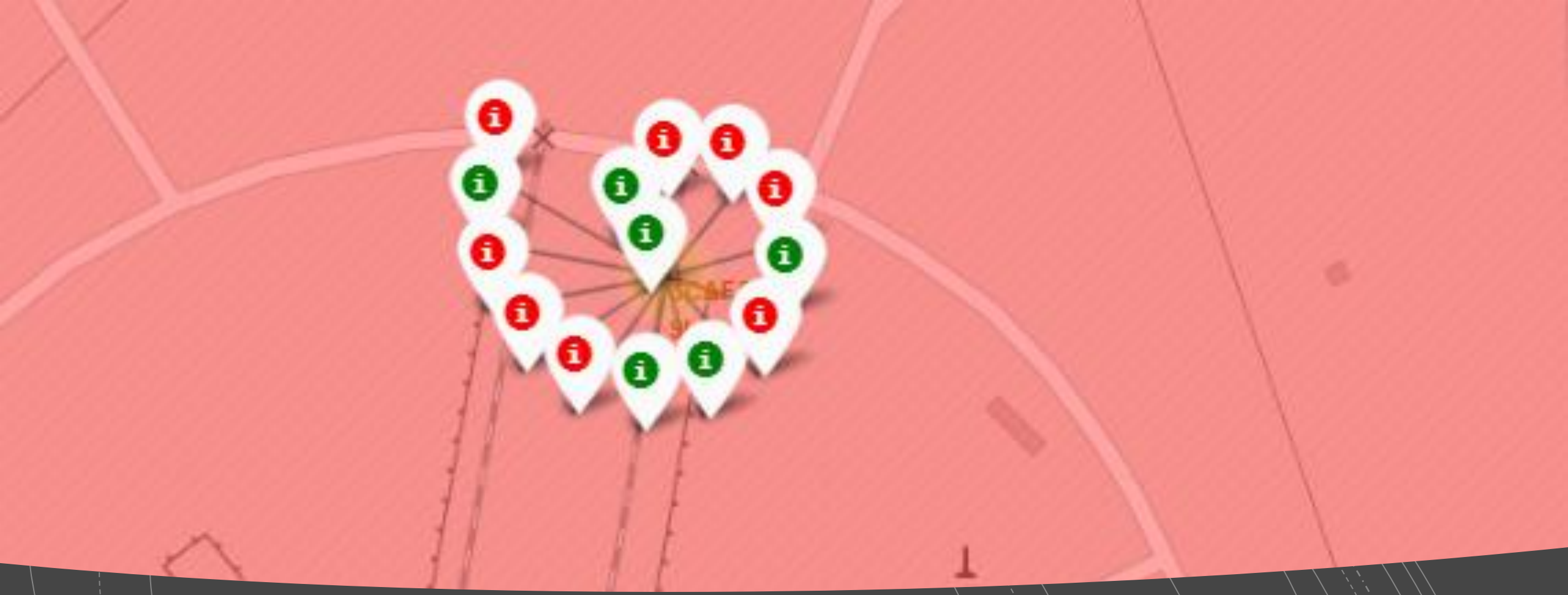
- Interactive showing one launch site created using Folium

RESULTS: Exploratory Data Analysis with Visualization

- Interactive map showing all launch sites created using Folium

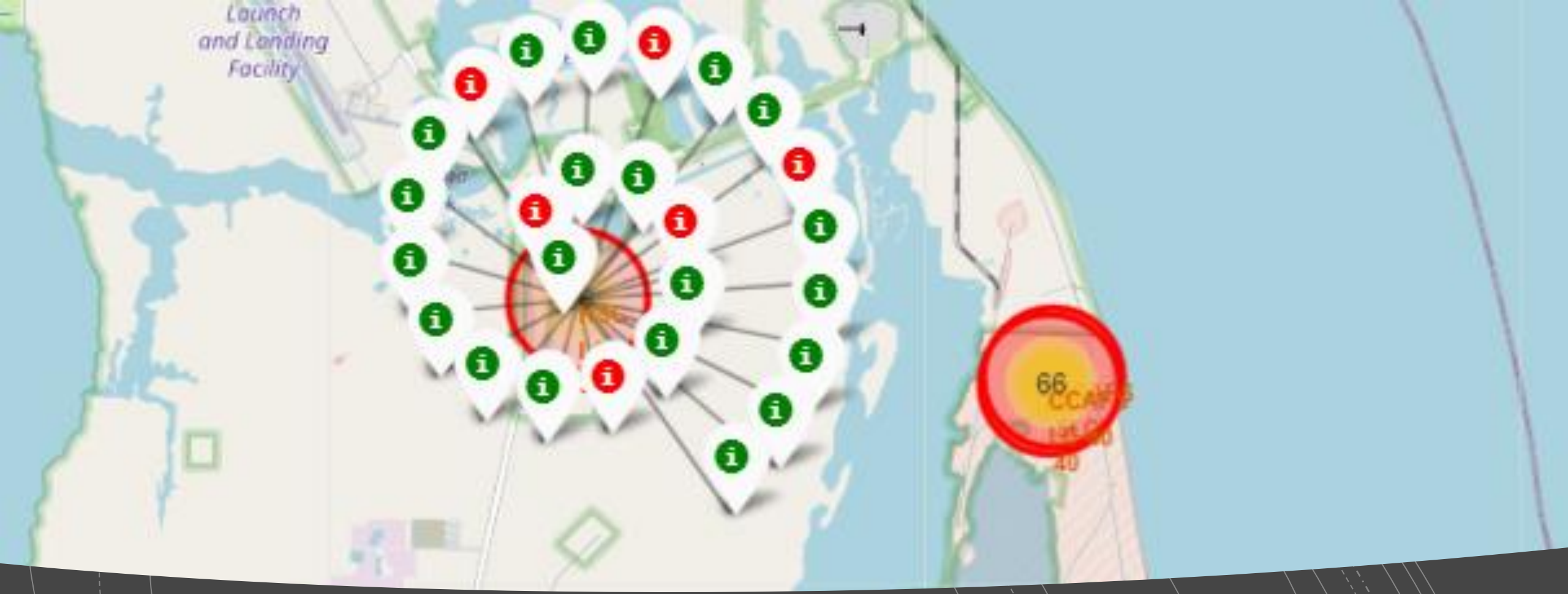
RESULTS: Exploratory Data Analysis with Visualization

- Launch site CCAFS LC-40 has 14 successful launches of 56.



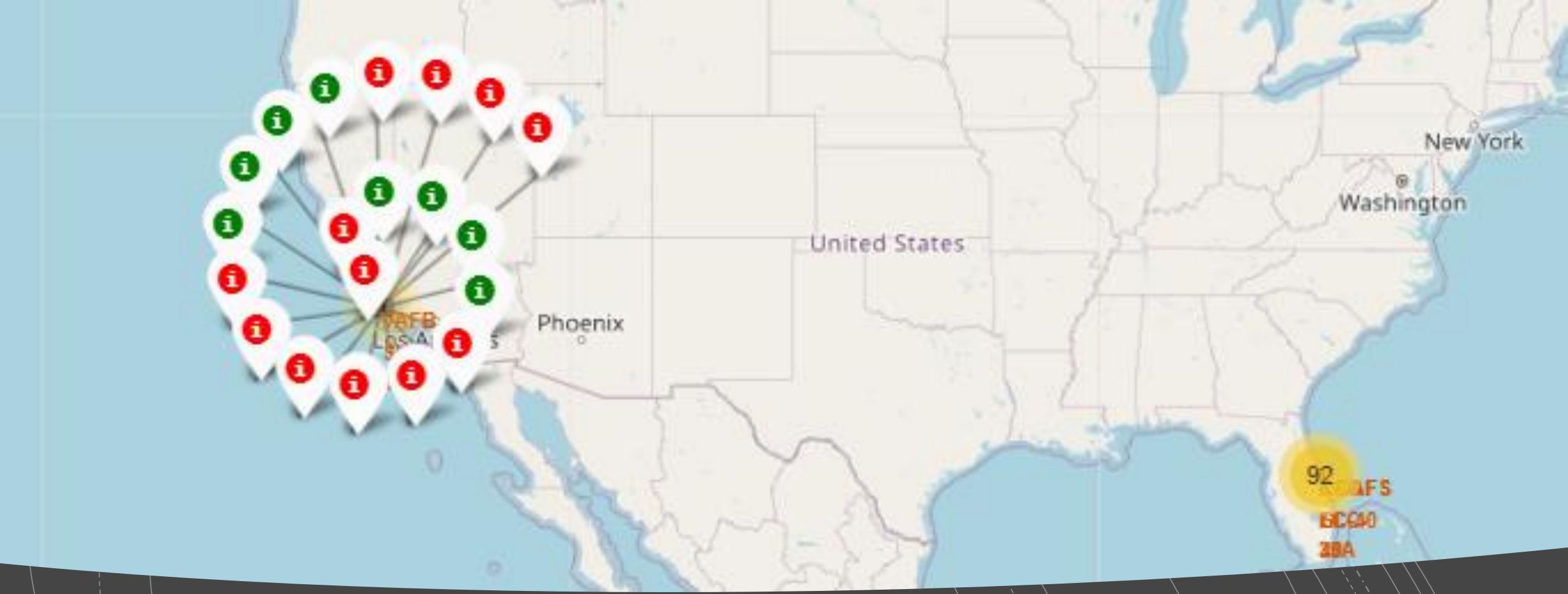
RESULTS: Exploratory Data Analysis with Visualization

- Launch site CCAFS SLC-40 has 6 successful launches of 14.



RESULTS: Exploratory Data Analysis with Visualization

- Launch site KSC LC- 39A has 20 successful launches of 26.



RESULTS: Exploratory Data Analysis with Visualization

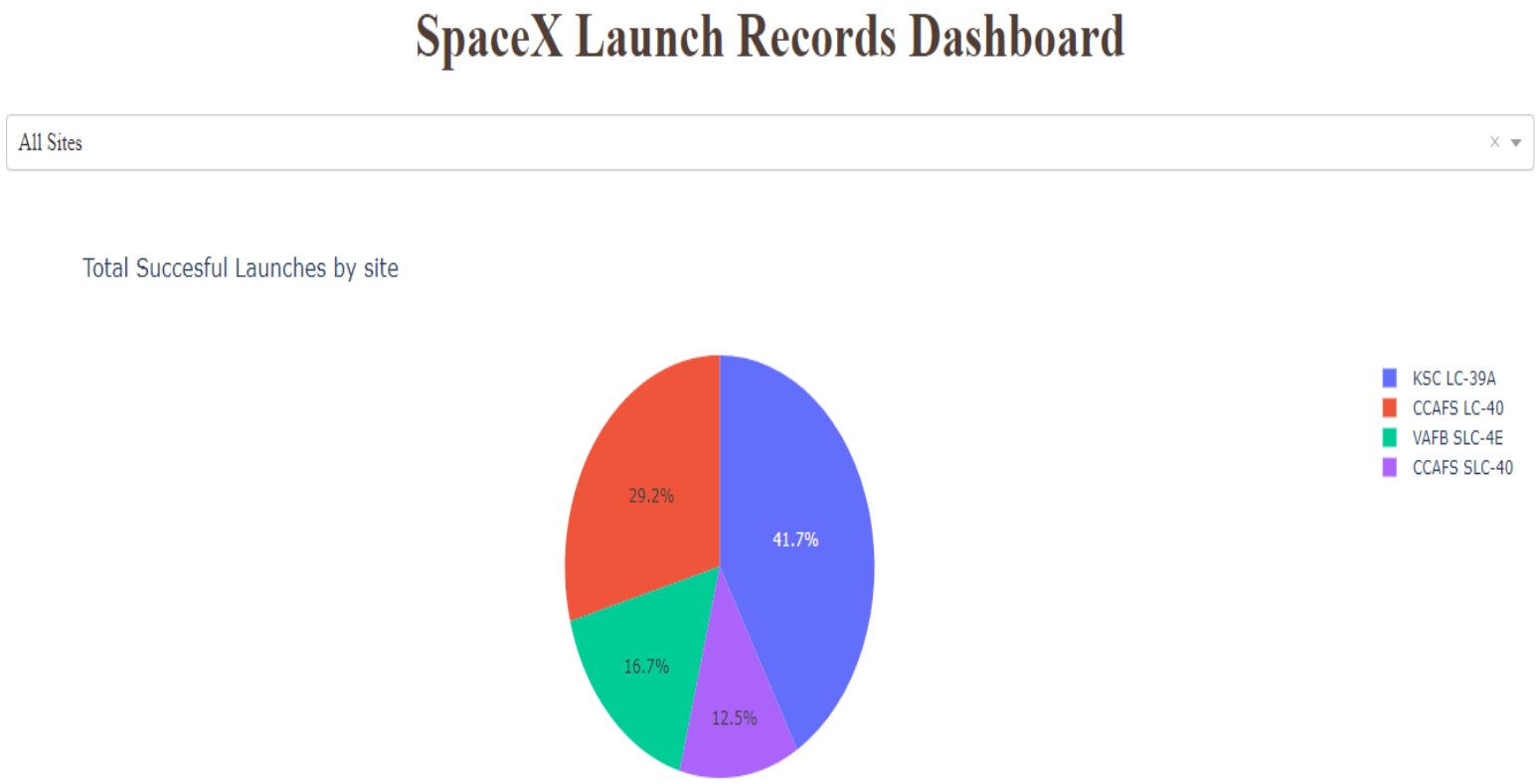
- Launch site VAFB SLC-4E has 8 successful launches of 20.

INTERACTIVE DASHBOARD

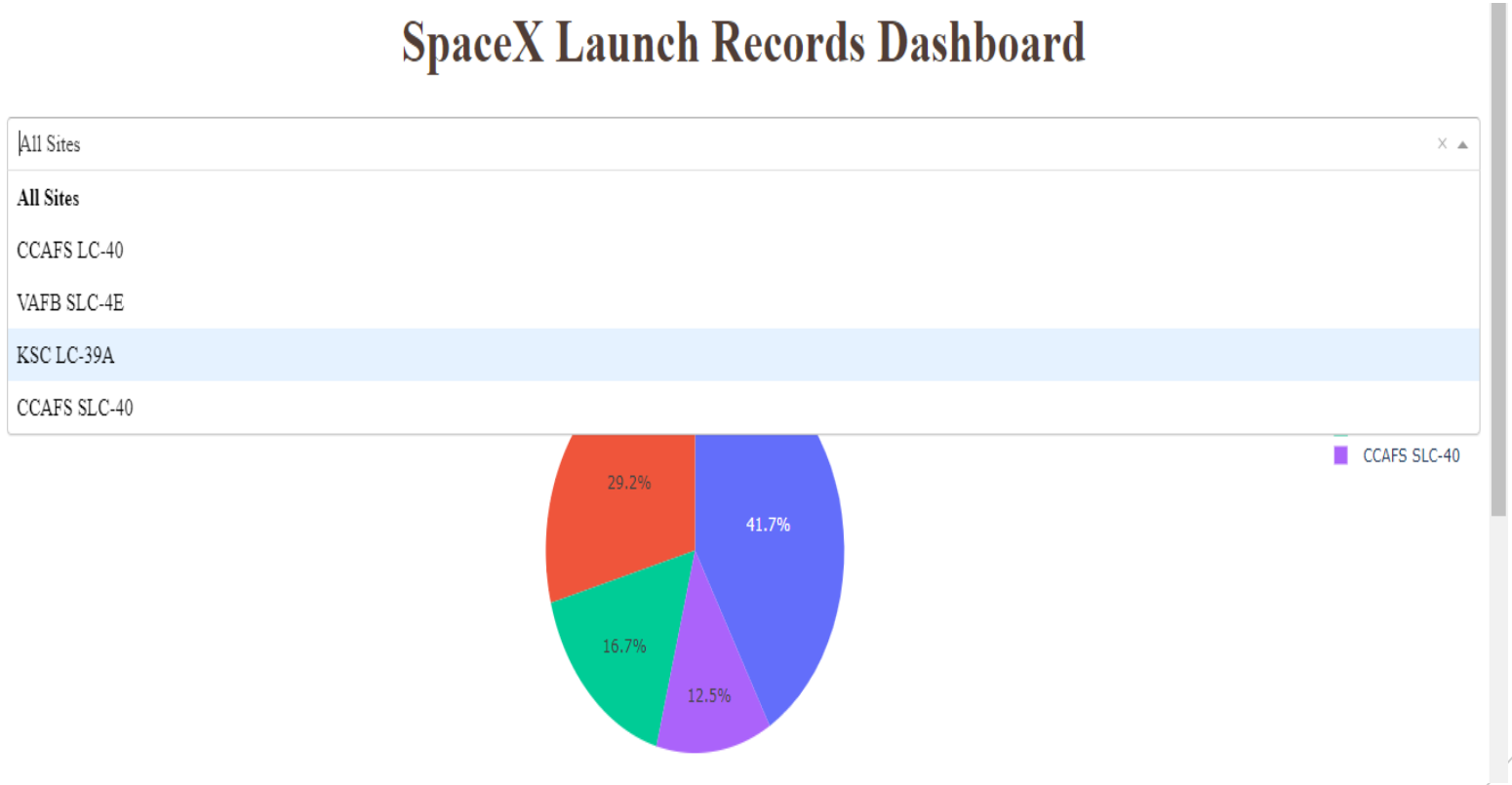
Link for the dashboard created using Plotly and Dash in Python

<https://djgrantel-8050.theiadocker-2-labs-prod-theiak8s-4-tor01.proxy.cognitiveclass.ai/>

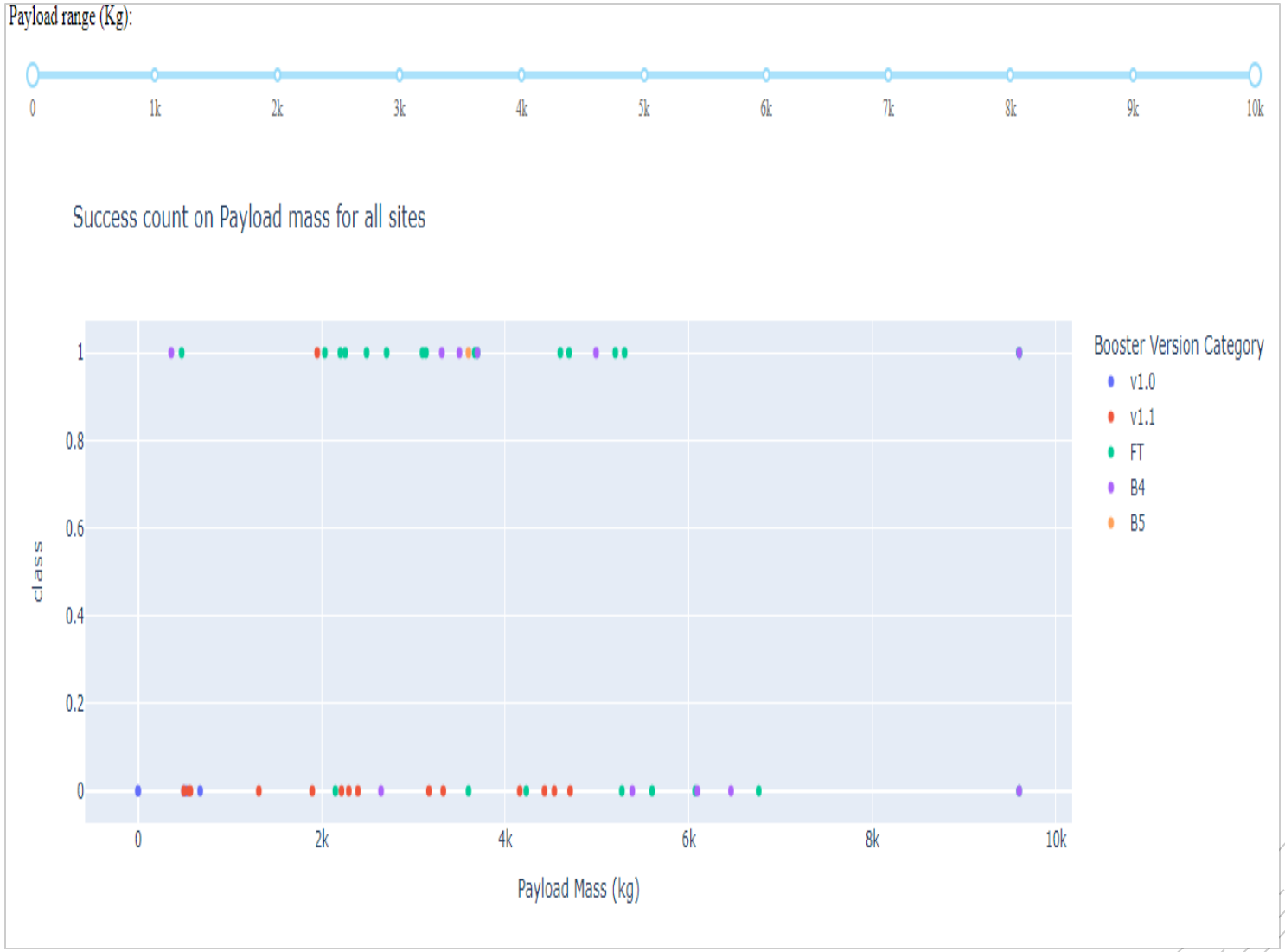
DASHBOARD TAB 1



DASHBOARD TAB 2



DASHBOARD TAB 3



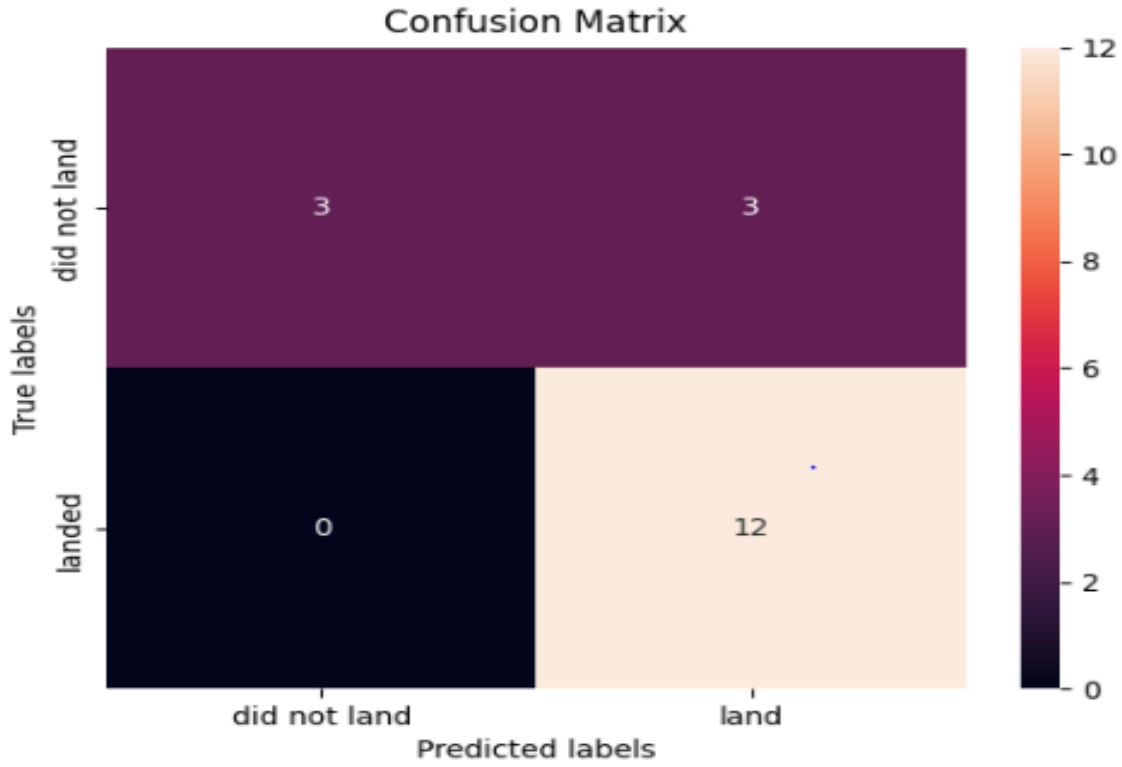
Results- Classification Model Parameters

The parameters for each classification model was selected using the GridSearchCV in Python:

1. Logistic Regression: `{'C': 0.01, 'penalty': 'l2', 'solver': 'lbfgs'}`
2. SVM: `{'C': 1.0, 'gamma': 0.03162277660168379, 'kernel': 'sigmoid'}`
3. Decision Tree: `{'criterion': 'entropy', 'max_depth': 4, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 5, 'splitter': 'random'}`
4. KNN: (best parameters) `{'algorithm': 'auto', 'n_neighbors': 10, 'p': 1}`

Results- Classification Model Parameters

The accuracy of each classification model was evaluated using the condusion matrix, all of which yielded the same result:



	Log Regression Training Score	Log Regression Test Score	Support Vector Training Score	Support Vector Test Score	Decision Training Score	Decision Tree Test Score	KNN Training Score	KNN Test Score
1	0.846429	0.833333	0.848214	0.833333	0.889286	0.777778	0.848214	0.833333

Results- Classification Model Metrics

Training and Test Score results for each model

DISCUSSION

- In observing the successful launches of SpaceX, the KSC LC- 39A had the largest number of successful launches and the highest success rate among the launch sites
- The most frequent payload range for launches is 0KG-6000KG and both successful and unsuccessful launches occur in this range.
- To distinguish the outcome of the launches in this payload range we consider the relationship between Payload Mass and Flight number where flight numbers over 40 exhibit more successful launches for all sites.

DISCUSSION

- Exploring other factors for successful launches it was found that the Orbit type is of significance as the following Orbits: HEO, GEO, SSO, ES-L1 exhibited more successful launches on average.
- The booster version is also of importance for the Falcon 9, as the FT booster version had the highest launch success rate among booster versions.
- All the models fit on the data had similar accuracy rates for the test and training set
- Likewise, the confusion matrix was the same for each model displaying a strong ability to predict landed launches but less effective with unsuccessful launches.

CONCLUSION

- By increasing the number of flights, SpaceX can anticipate more successful launches
- SpaceX can continue to reduce costs with more successful launches given the performance of the models to predict them
- SpaceX should focus on carrying out launches at certain sites with certain orbit types.