IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Dmitriy Fisch
07/23/2022

https://github.com/schahmatist2/IBM-Capstone-Project/

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

- **Summary of Methodologies**

The goal of this project is the successful estimation of cost of the space launch. The important factor that contributes to the cost is the reusability of the first stage. We collected publicly available data (using both API and web scrapping) to determine how successful SpaceX is in landing their first stages. After cleaning our data we used multiple visualizations and EDA analysis to determine main factors that contributed to the first stage landing success. We also built predictive models and estimated the chance for a first stage to safely land depending on the conditions and features.

- **Summary of Results**

We found that multiple factors affect the success rate of the launch including how recent is the flight, how heavy is the load, and what launching sites are used. Using the collected data and predictive modeling we reached 83.3% percent of accuracy in determining the probability of success. Those findings can help a commercial space company to design their own flights or invest into SpaceX.

Introduction

- In a commercial space age the cost of a space flight is an important consideration. The coast of a multi-stage flight depends a lot on the chances that the first stage can be safely landed. The pioneer of this method is SpaceX, and there is a publicly available data with information how successful SpaceX is in landing their first stages.

- If we can use publicly available information to determine the factors that contributed to successful landings of the first stage, we can predict the chances those landings for the future flights. We can use data science and machine learning to address our problem. The meaningful findings can be extremely useful in investing into SpaceX as well as for helping other commercial companies to plan their own flights.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Beautiful Soup for extracting data from Wiki sources
  - SpeceX Api was used to get data directly from SpaceX
- Data wrangling:
  - Data was cleaned, null values were filled with mean values of removed when appropriate.
- Exploratory data analysis (EDA) using visualization and SQL
- Interactive visual analytics using Folium and Plotly Dash
- Predictive analysis using classification models
  - Four classification models were built using scikit-learn (Logistic Regression, KNN, SVM, and Random Trees)
  - Hyperparameters were tuned using GridSearch
  - Accuracy_score and confusion matrix were used to evaluate the models

Data Collection

Sources:

- Wikipedia
- SpaceX API

Data Collection process:

- Information about 90 launches (since 2012) were downloaded
- Beatiful Soup and Spacex API were used to get the data
- Data was loaded to pandas dataframe and saved in .csv files.

Features:

The following features were extracted:
- Booster Version
- Payload Mass
- Orbit
- Launch
- Site
- Longitude/Latitude
- Outcome

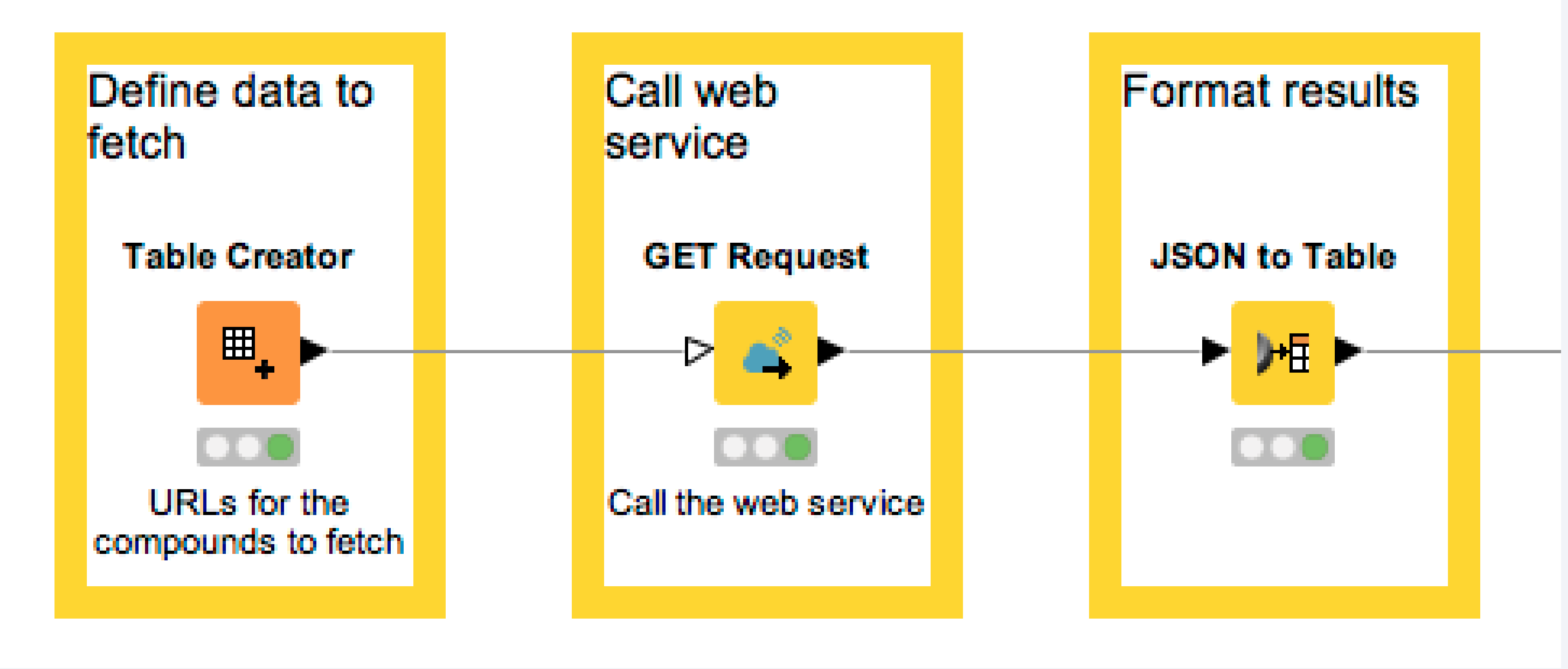
... and others

Data Collection – SpaceX API

- Requested and parsed the SpaceX launch data using the GET request
- Loaded data to dataframe
- Filtered the dataframe to only include Falcon 9 launches
- Imputed Missing Values

For more information please check:

SpaceX API Data Collection Notebook



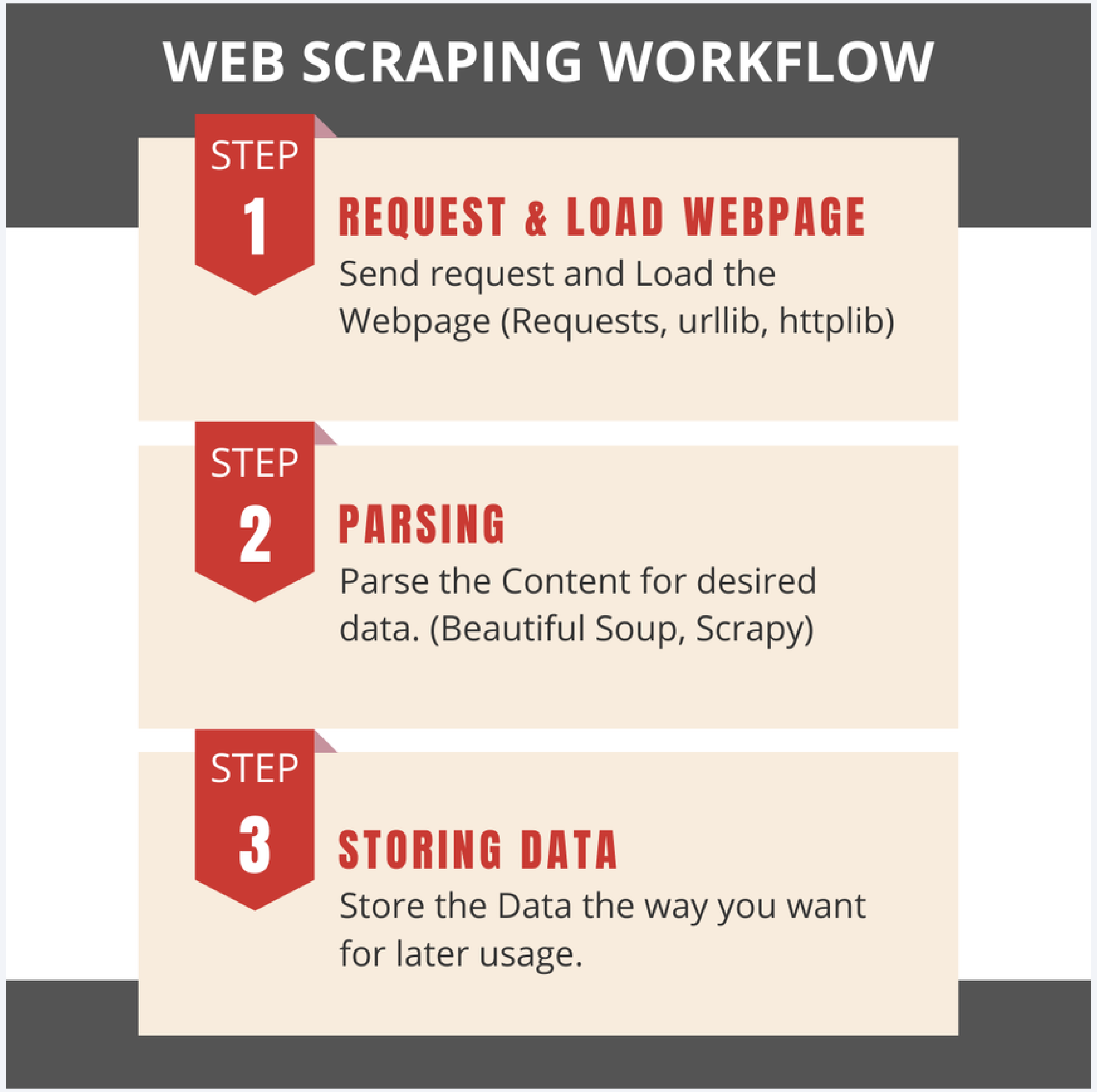
Data Collection - Scraping

- Requested the Falcon9 Launch Wiki page from: https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

- Extracted all column/variable names from the HTML table header

- Created a data frame by parsing the launch HTML tables

For details please check:
[Data Collection with Web Scraping Notebook]



Data Wrangling

EDA Included:

- Calculated the number of launches on each site
- Calculated the number and occurrence of each orbit
- Calculated the number and occurrence of mission outcome per orbit type
- Created a landing outcome label (0 for failed or 1 for success) from Outcome column

For more information please check:
Data Wrangling EDA Notebook

EDA with Data Visualization

The following correlations were explored and visualized:

- The relationship between Flight Number and Launch Site
- The relationship between Payload and Launch Site
- The relationship between success rate of each orbit type
- The relationship between FlightNumber and Orbit type
- The relationship between Payload and Orbit type
- The launch success yearly trend

For more details please check: EDA with Data Visualization Notebook

EDA with SQL

- Got the names of the unique launch sites in the space mission
- Checked records where launch sites begin with the string 'CCA'
- Checked the total payload mass carried by boosters launched by NASA (CRS)
- Checked average payload mass carried by booster version F9 v1.1
- Listed the date when the first succesful landing outcome in ground pad was acheived.
- Listed the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Listed the total number of successful and failure mission outcomes
- Listed the names of the booster_versions which have carried the maximum payload mass.
- Listed the records which will display the month names, failure landing_outcomes in drone ship, booster versions, launch_site for the months in year 2015.
- Ranked the count successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

For more details please check: EDA with SQL Notebook

Interactive Maps with Folium

- In order to get better understanding of data we marked the launch sites (used Clusters and Markers to visualized) on a map
- We determined the distance between launching sites and infrastructure/coastlines objects
- We also marked which launches were successful and which were not.

For more details please check Interactive Map with Folium Notebook

Dashboard with Plotly Dash

- For better Data understanding we created an intuitive dashboard to display percentage of successful launches for each launches sites
- We also visualized how payload mass contributed to the success of the launch

For more details please check: python script that generates the Dashboard

Predictive Analysis (Classification)

Four classification models were built using scikit-learn:
- Logistic Regression
- KNN
- SVM
- Random Trees

- Hyperparameters were tuned using GridSearch

- Accuracy_score and confusion matrix were used to evaluate the models

- Most of the models achieved an Accuracy of 83.3% to predicting if a launch is going to be Successful or not

For more details please check: Machine Learning Prediction Notebook

Results

The following sections include:

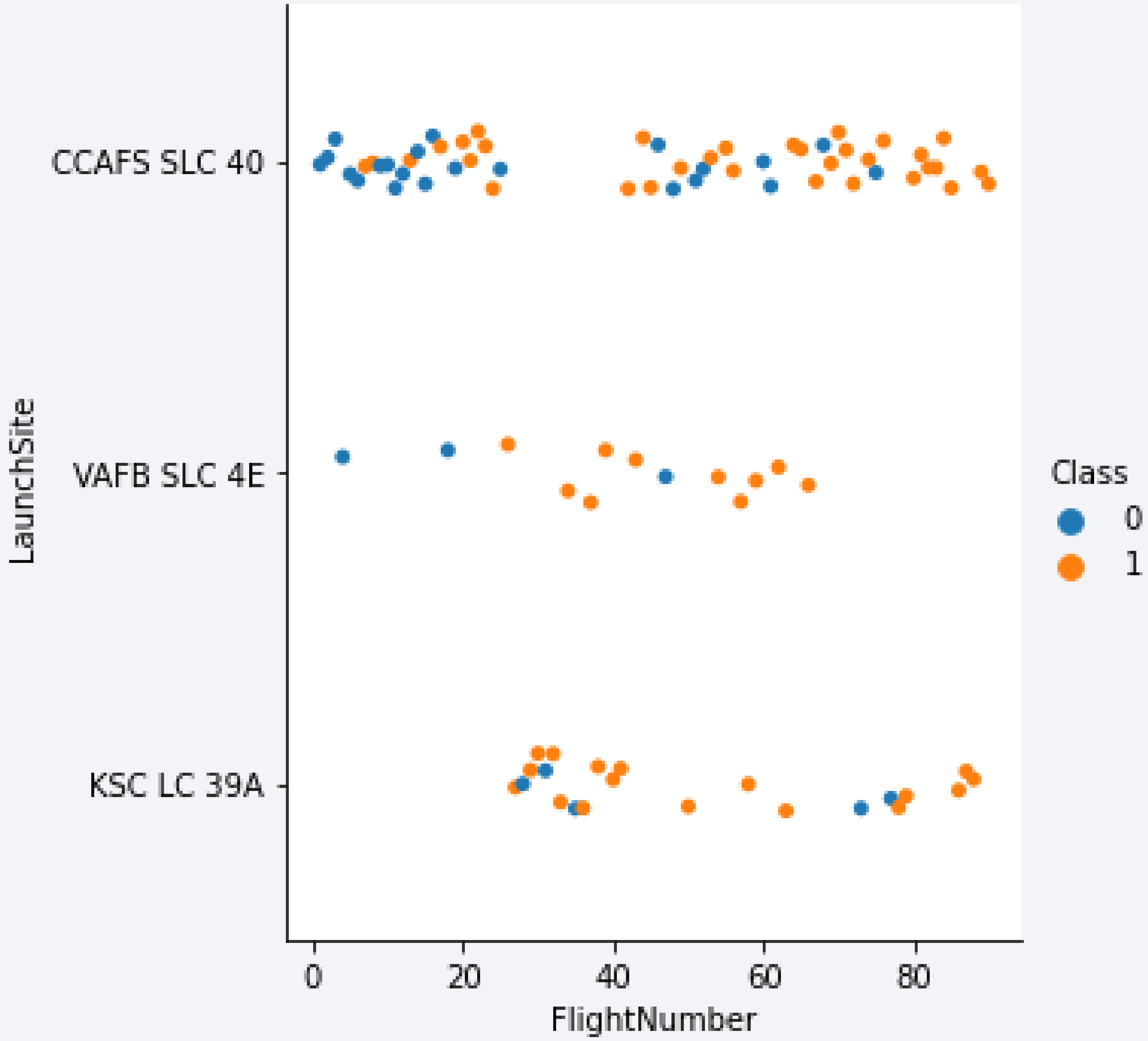
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

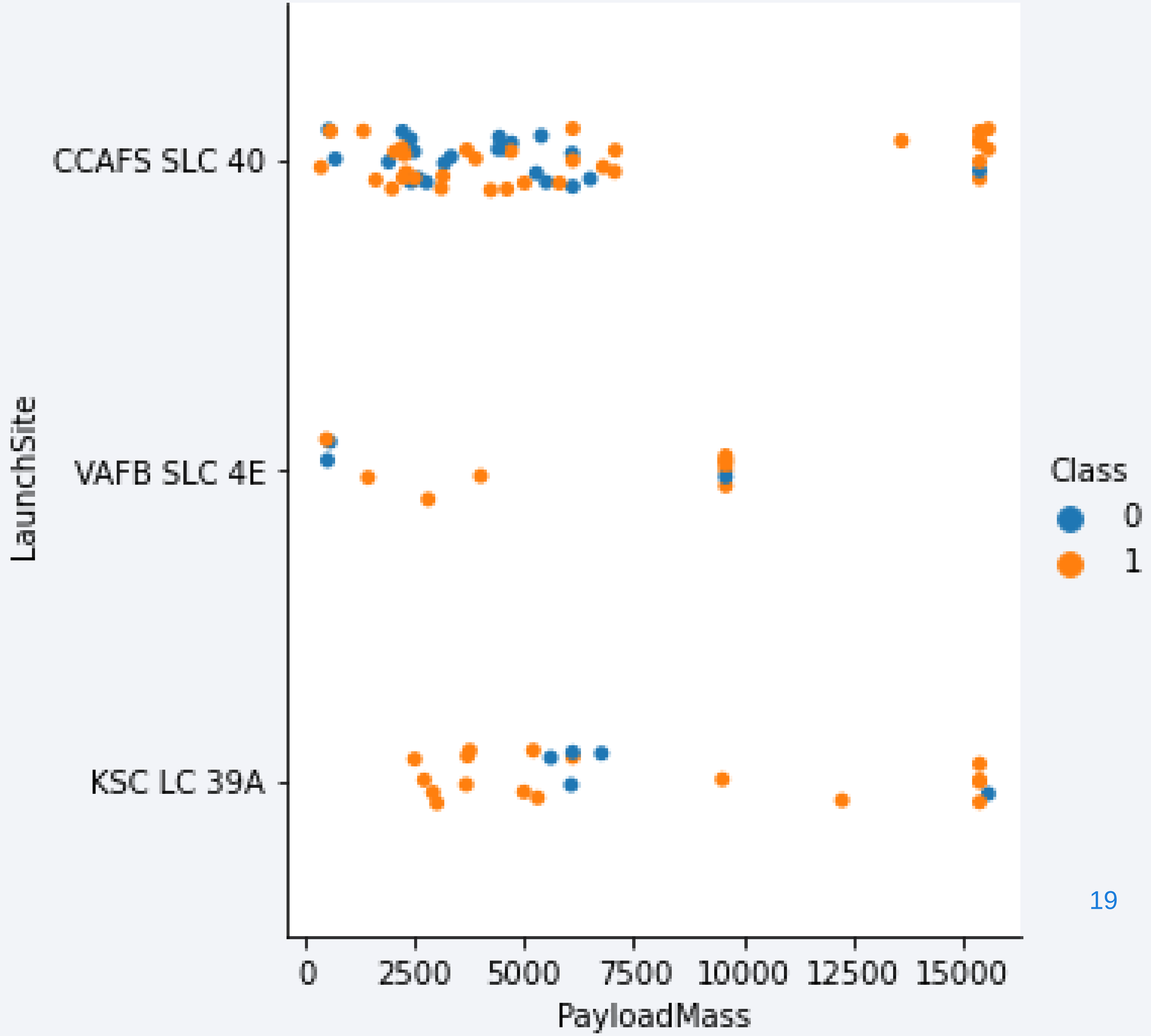
- We can see, that later (larger) Flight Numbers tend to be more successful than earlier flights for all 3 launching Sites:



Payload vs. Launch Site

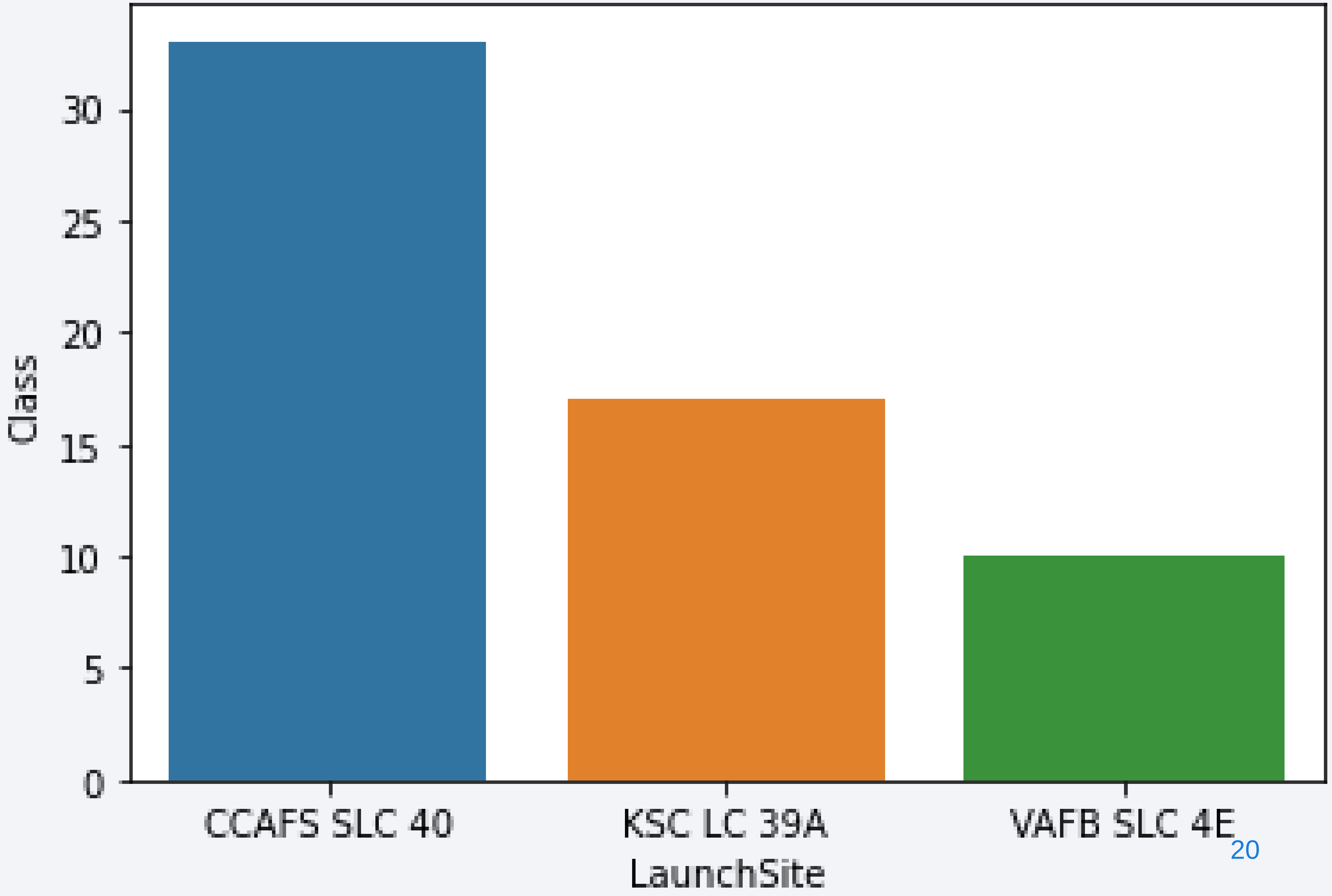
- For the VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000)

- For KSC LC 39A all launches have payload greater than 2500



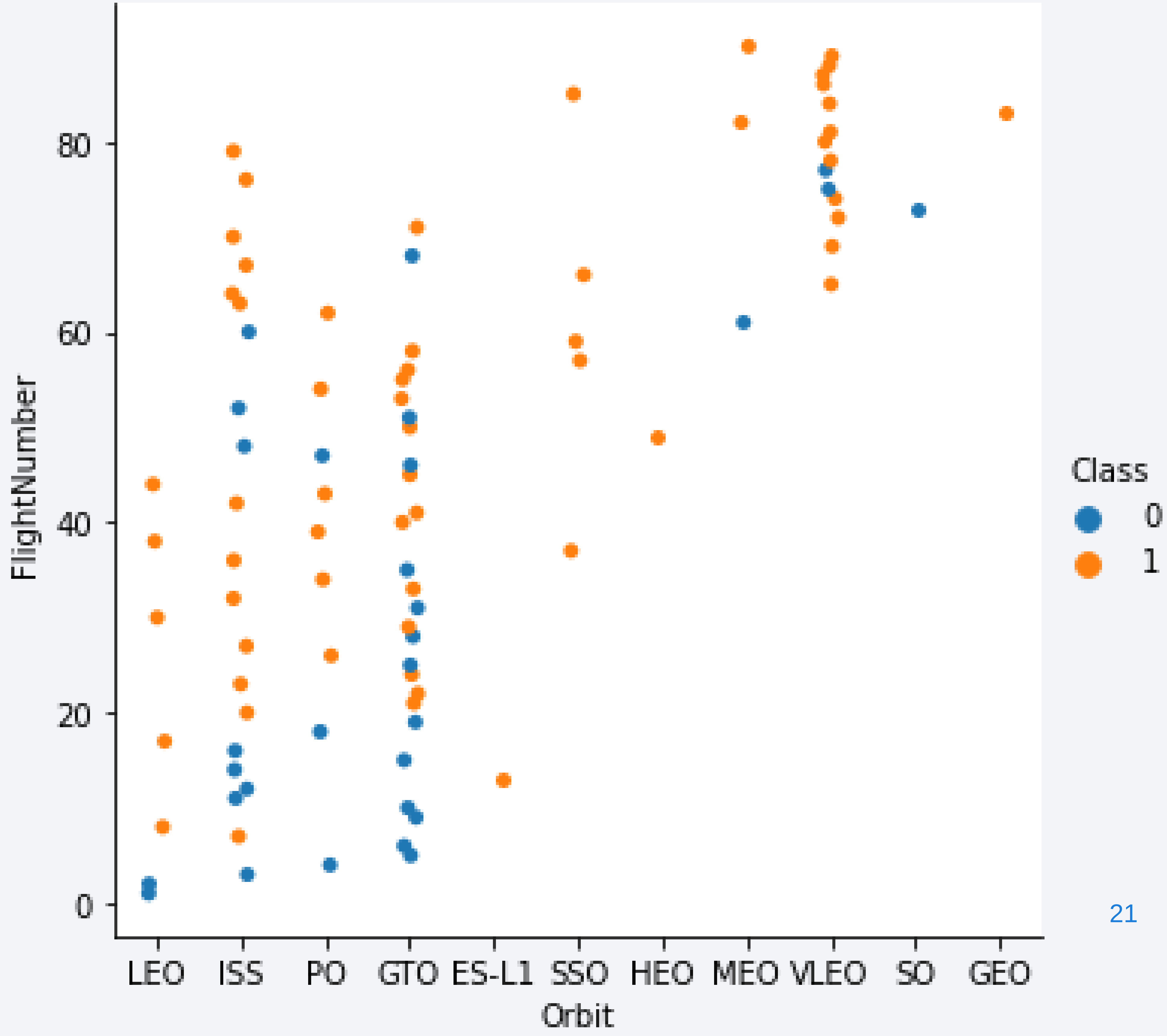
Success Rate vs. Orbit Type

- CCAFS has the most successful launches followed by KSC LC and VAFB



Flight Number vs. Orbit Type

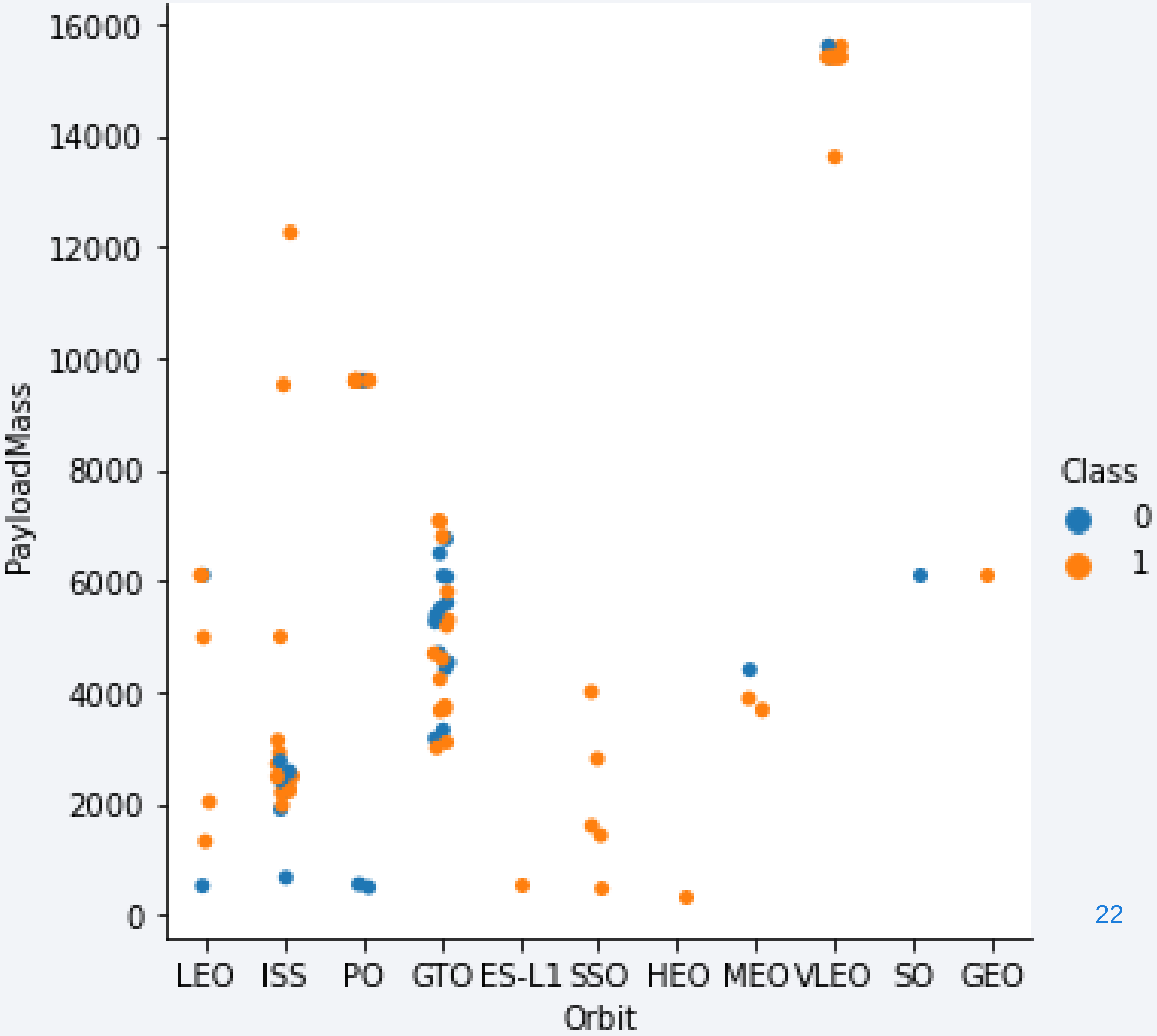
- In the LEO orbit the Success appears related to the number of flights
- There seems to be no relationship between flight number when in GTO orbit.



Payload vs. Orbit Type

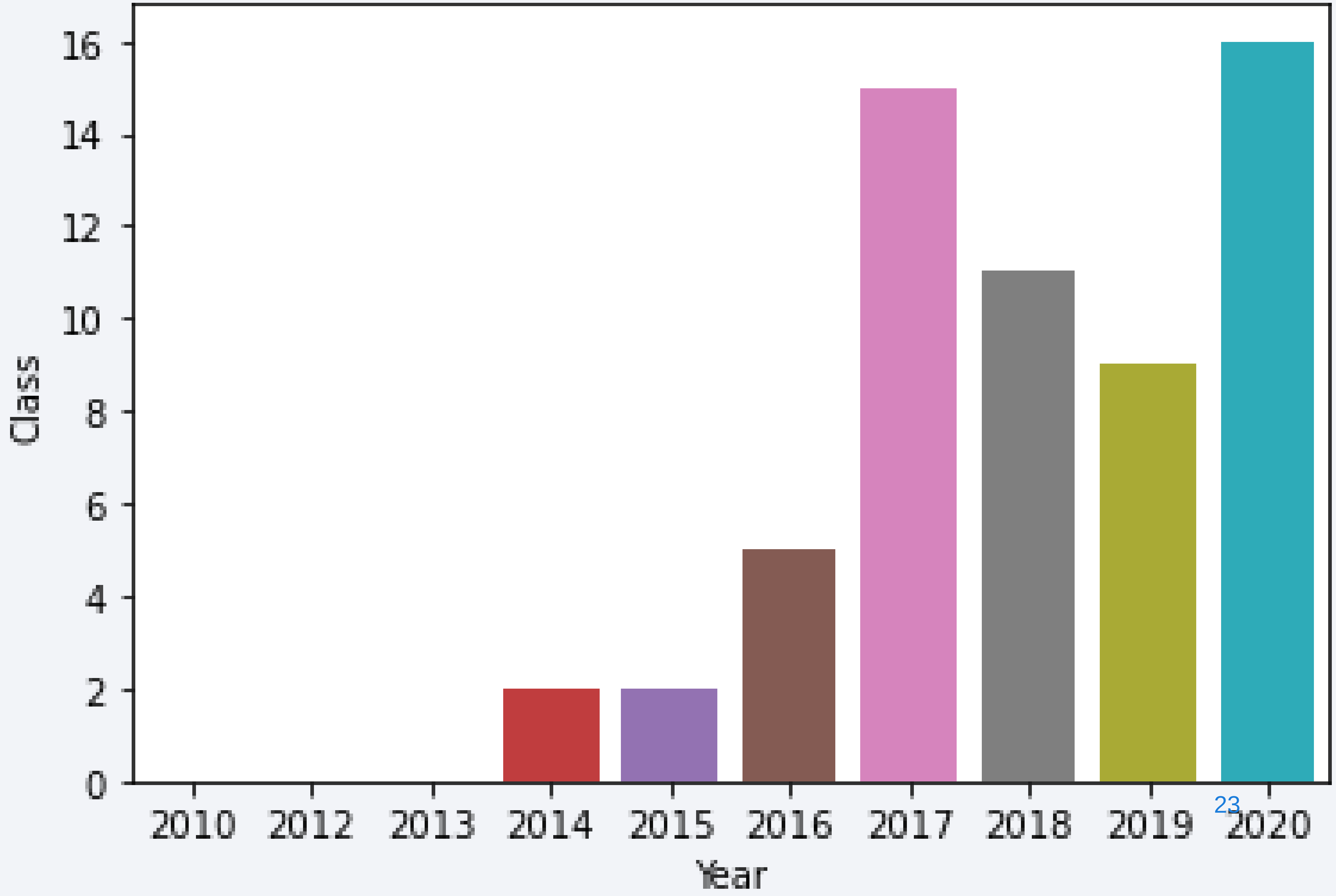
- With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS

- However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.



Launch Success Yearly Trend

- The sucess rate since 2013 kept increasing till 2020



All Launch Site Names

The following unique launch sites were identified:

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

Launch Site Names Begin with 'CCA'

- Exploring CCAFS flights:

7]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- total payload mass carried by boosters launched by NASA (CRS):

Out[17]:

sum(PAYLOAD_MASS__KG_)
48213

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1:

avg(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

- The first successful landing outcome on ground pad:

Out[62]:

min(Date)
01-05-2017

Successful Drone Ship Landing with Payload between 4000 and 6000

- Boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
06-05-2016	05:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
14-08-2016	05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
30-03-2017	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
11-10-2017	22:53:00	F9 FT B1031.2	KSC LC-39A	SES-11 / EchoStar 105	5200	GTO	SES EchoStar	Success	Success (drone ship)

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes:

substr("Mission_Outcome",1,7)	count(*)
Failure	1
Success	100

Boosters Carried Maximum Payload

- Boosters which have carried the maximum payload mass:

Booster_Version
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

2015 Launch Records

- The failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015:

substr(Date, 4, 2)	Booster_Version	Launch_Site	Landing _Outcome
01	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

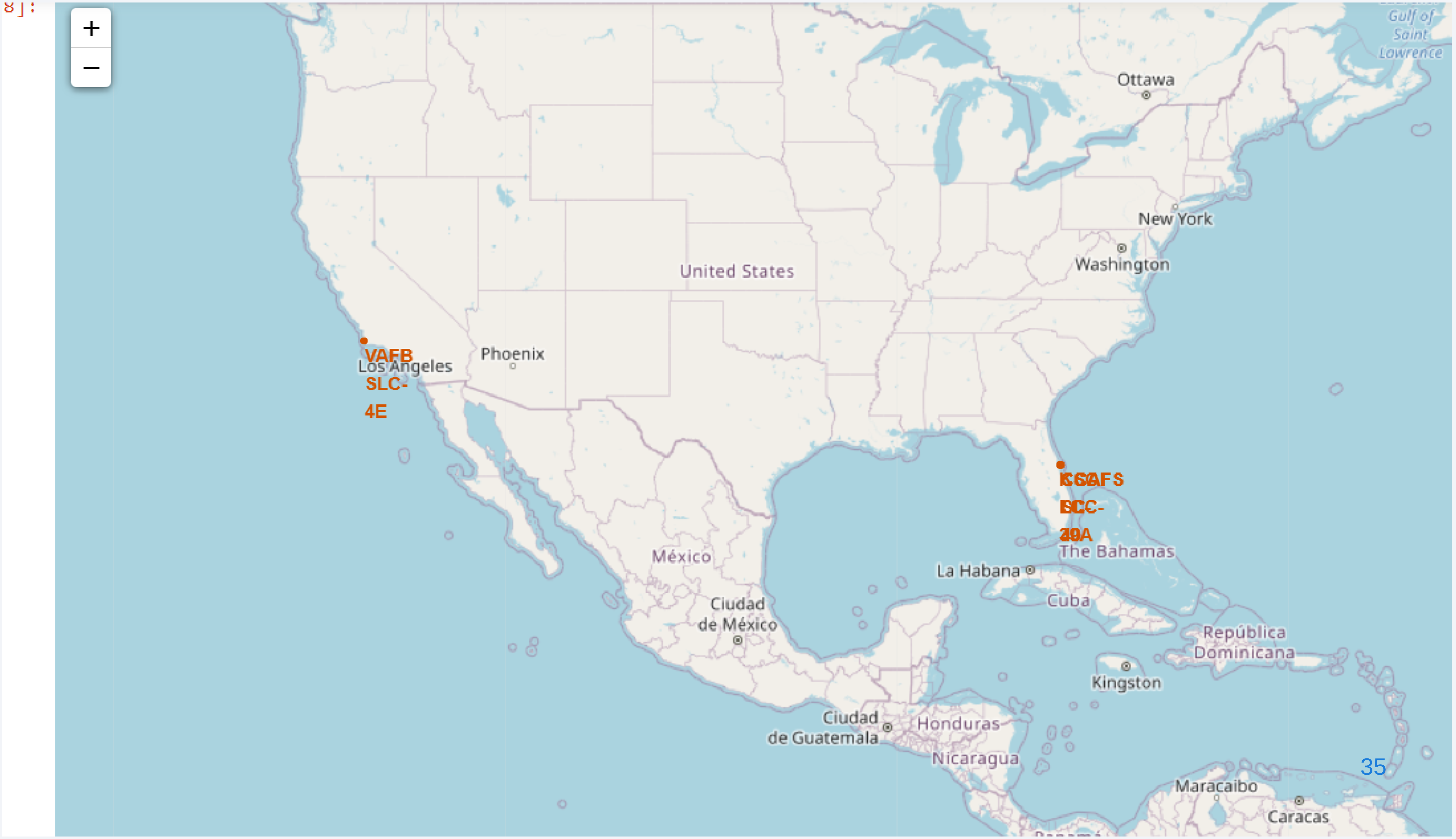
Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
19-02-2017	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
14-01-2017	17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600	Polar LEO	Iridium Communications	Success	Success (drone ship)
14-08-2016	05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
18-07-2016	04:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
27-05-2016	21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success	Success (drone ship)
06-05-2016	05:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
08-04-2016	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136	LEO (ISS)	NASA (CRS)	Success	Success (drone ship)
22-12-2015	01:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

Section 3

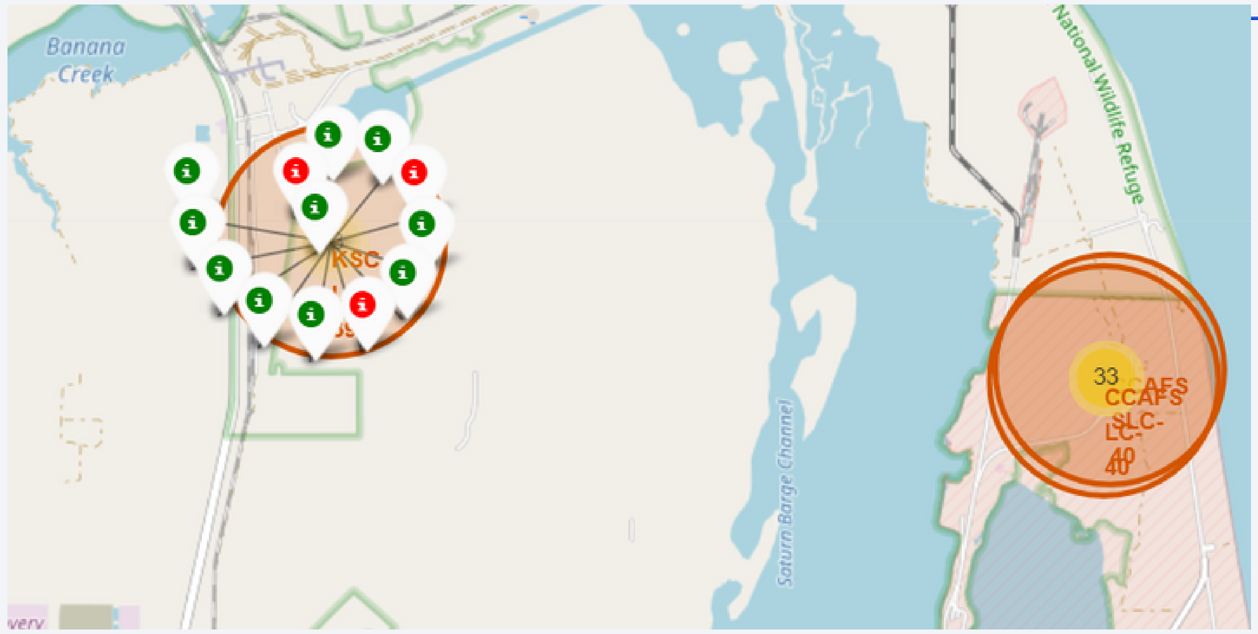
Launch Sites Proximities Analysis

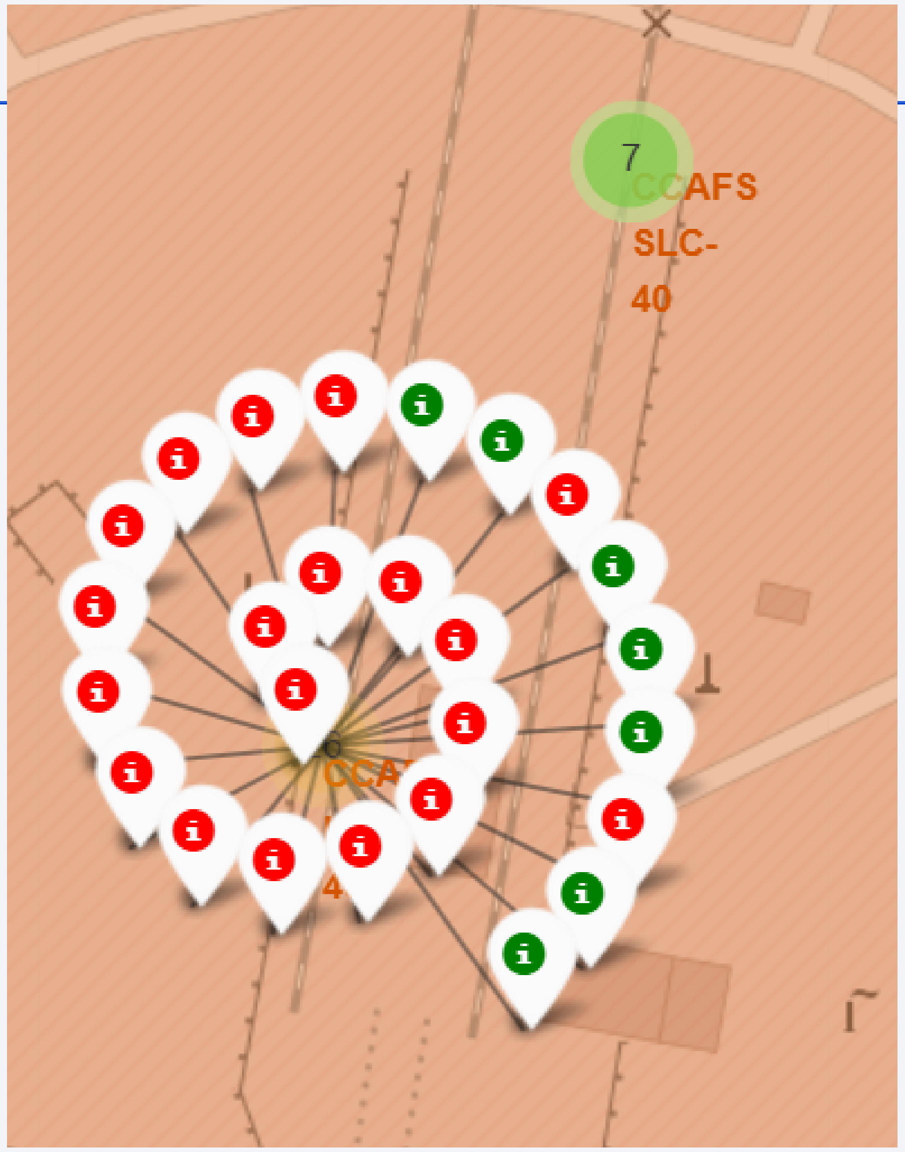
Launch sites Locations

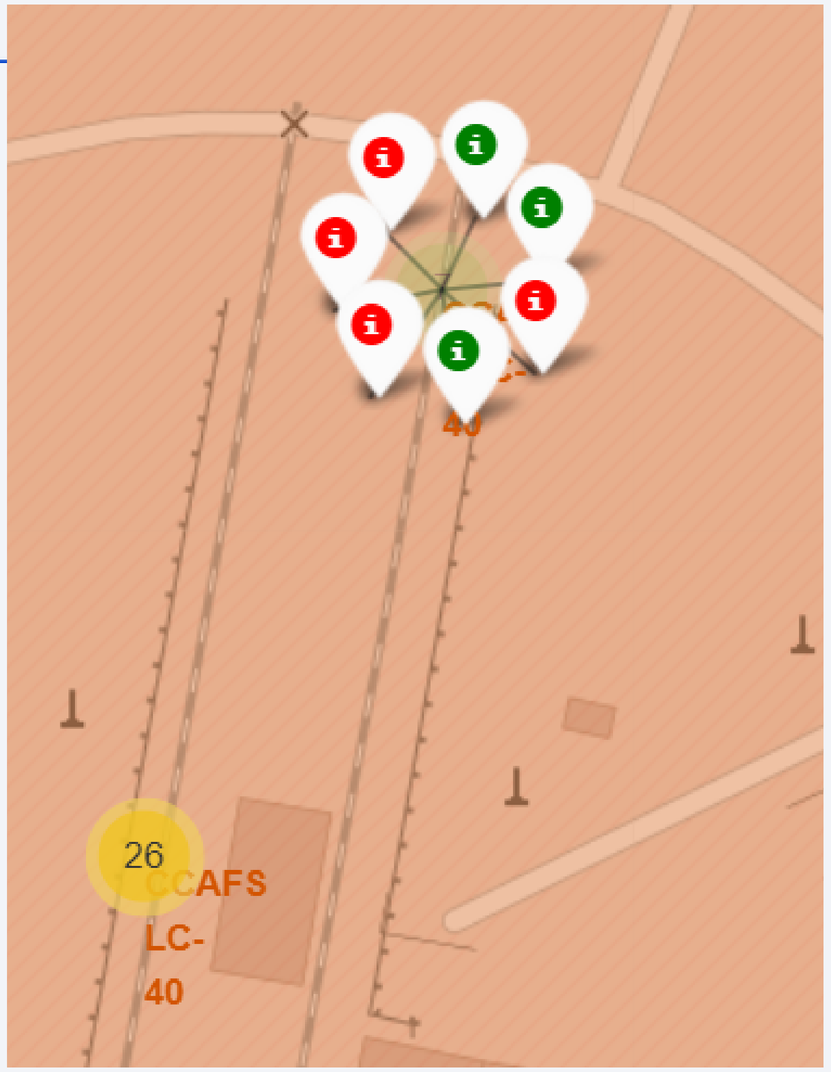
- All launch locations are close to coastlines (West and East):

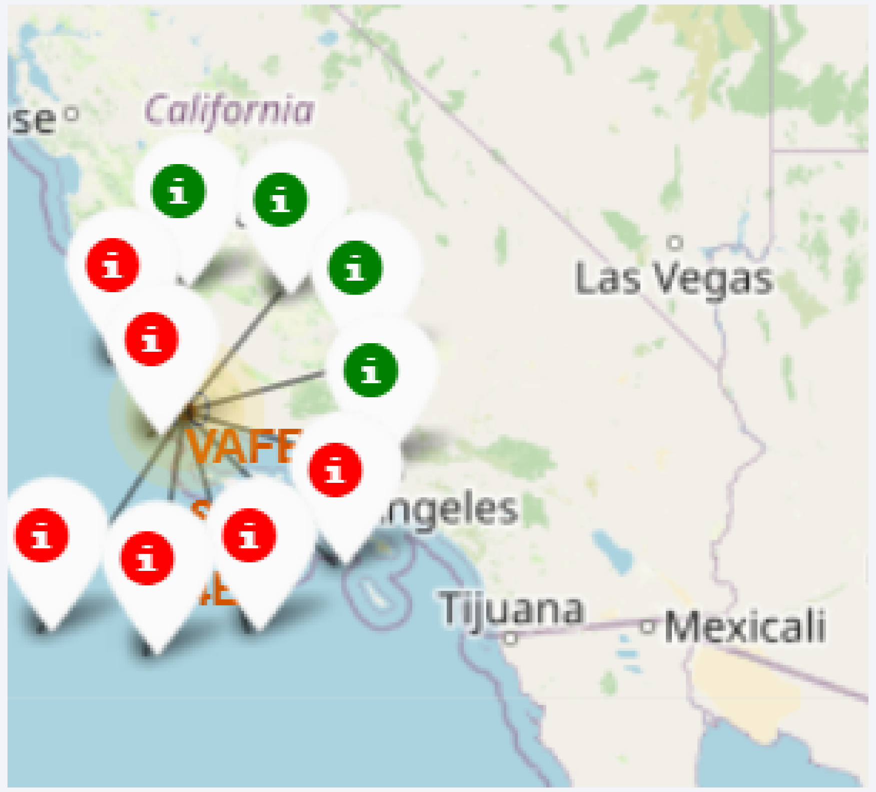


Visualization of Launch Sites with different Outcomes









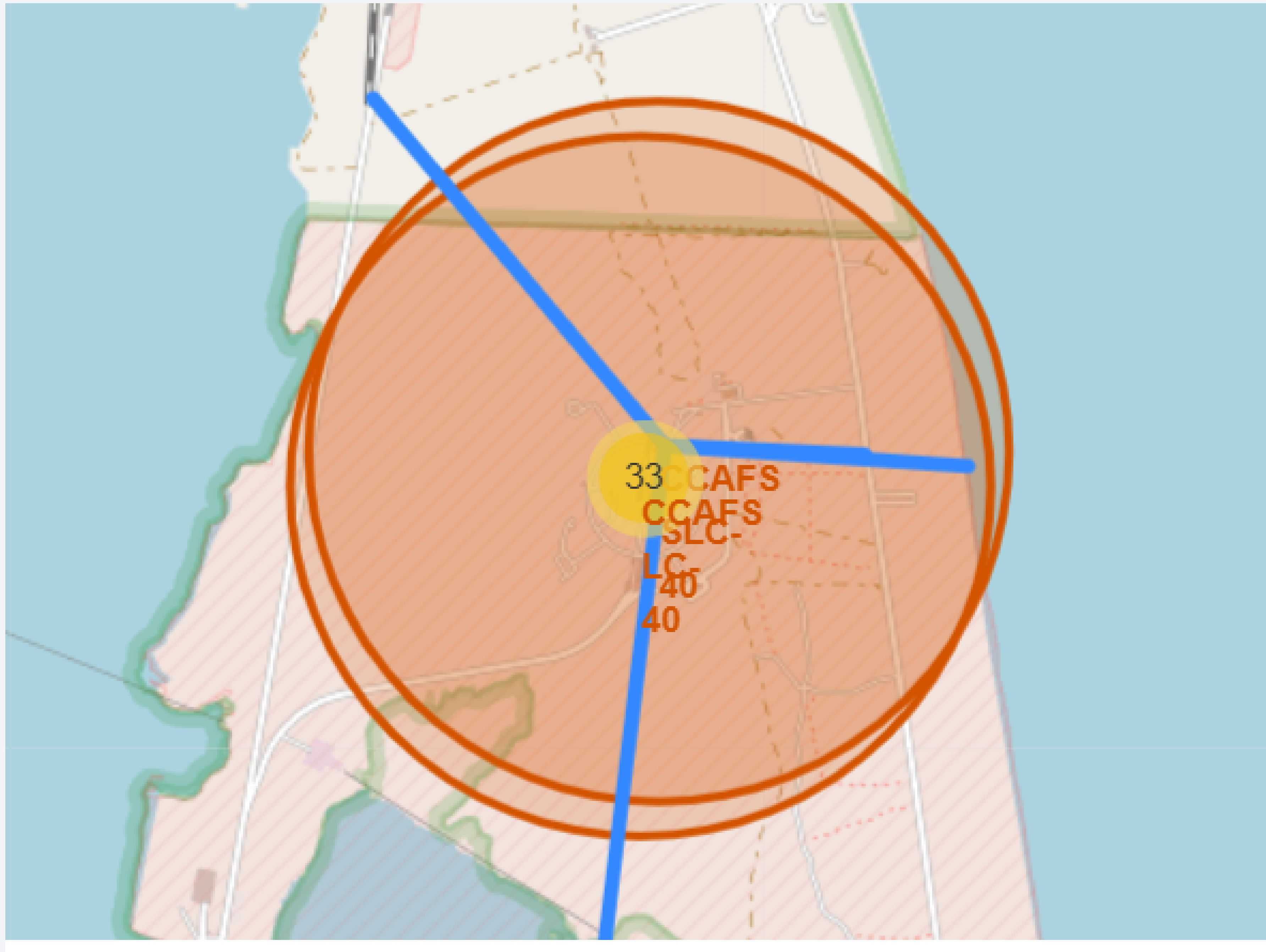
Distances from LaunchSites to infrastructure and coastlines

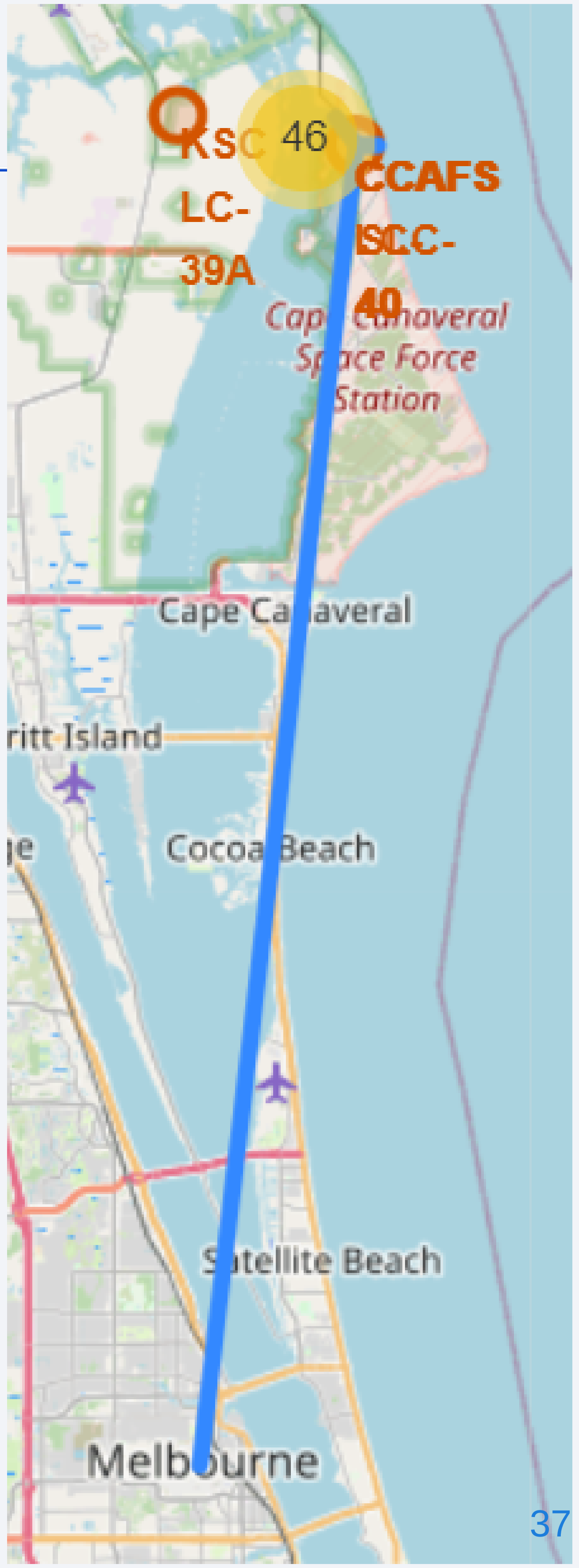
Close distances to:

- Highways
- Railroads
- Coastlines

More than 50 miles distance from the cities:

```
1 distance_city
51.360985044341156
```

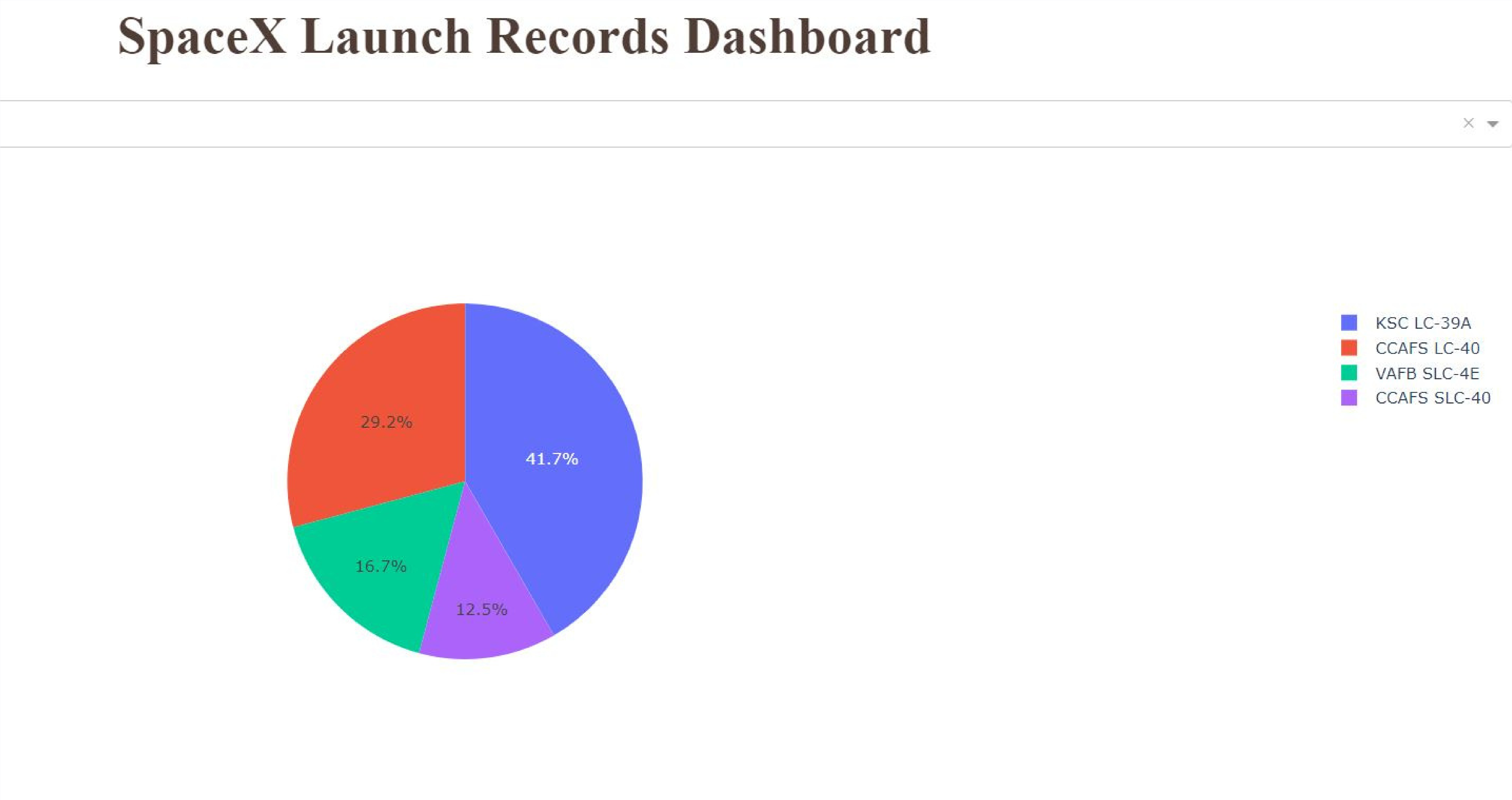




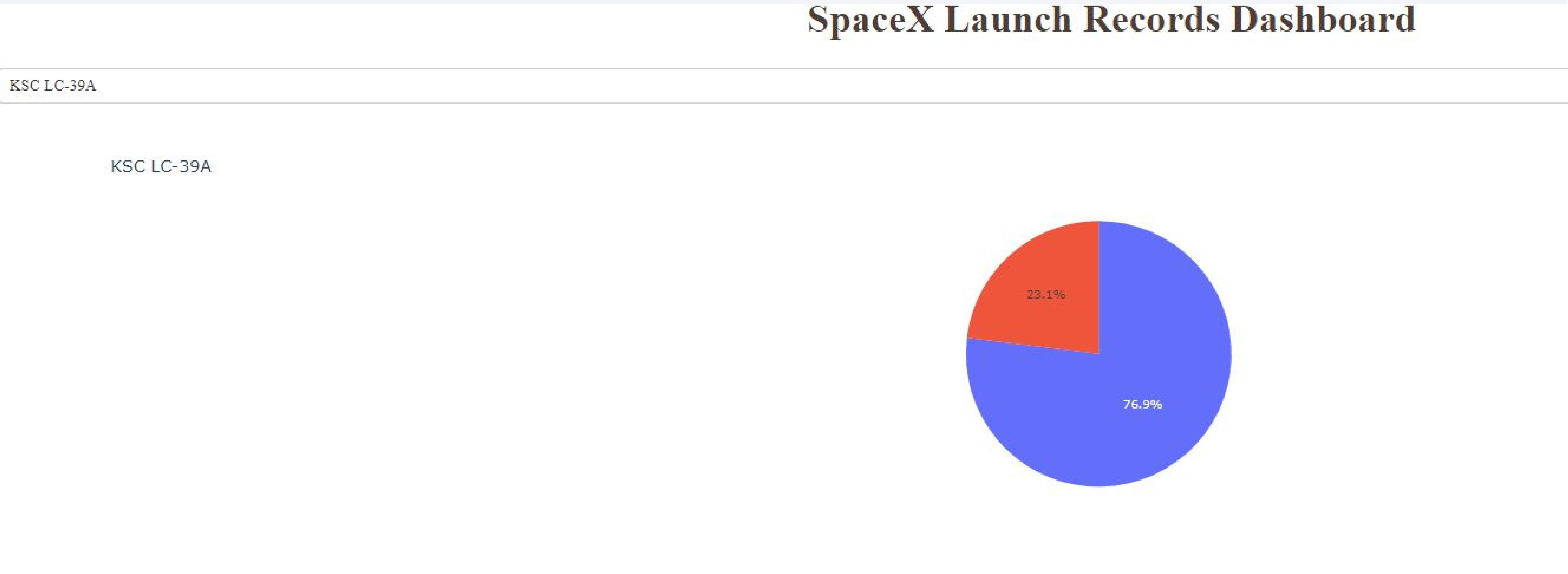
Section 4

Build a Dashboard with Plotly Dash

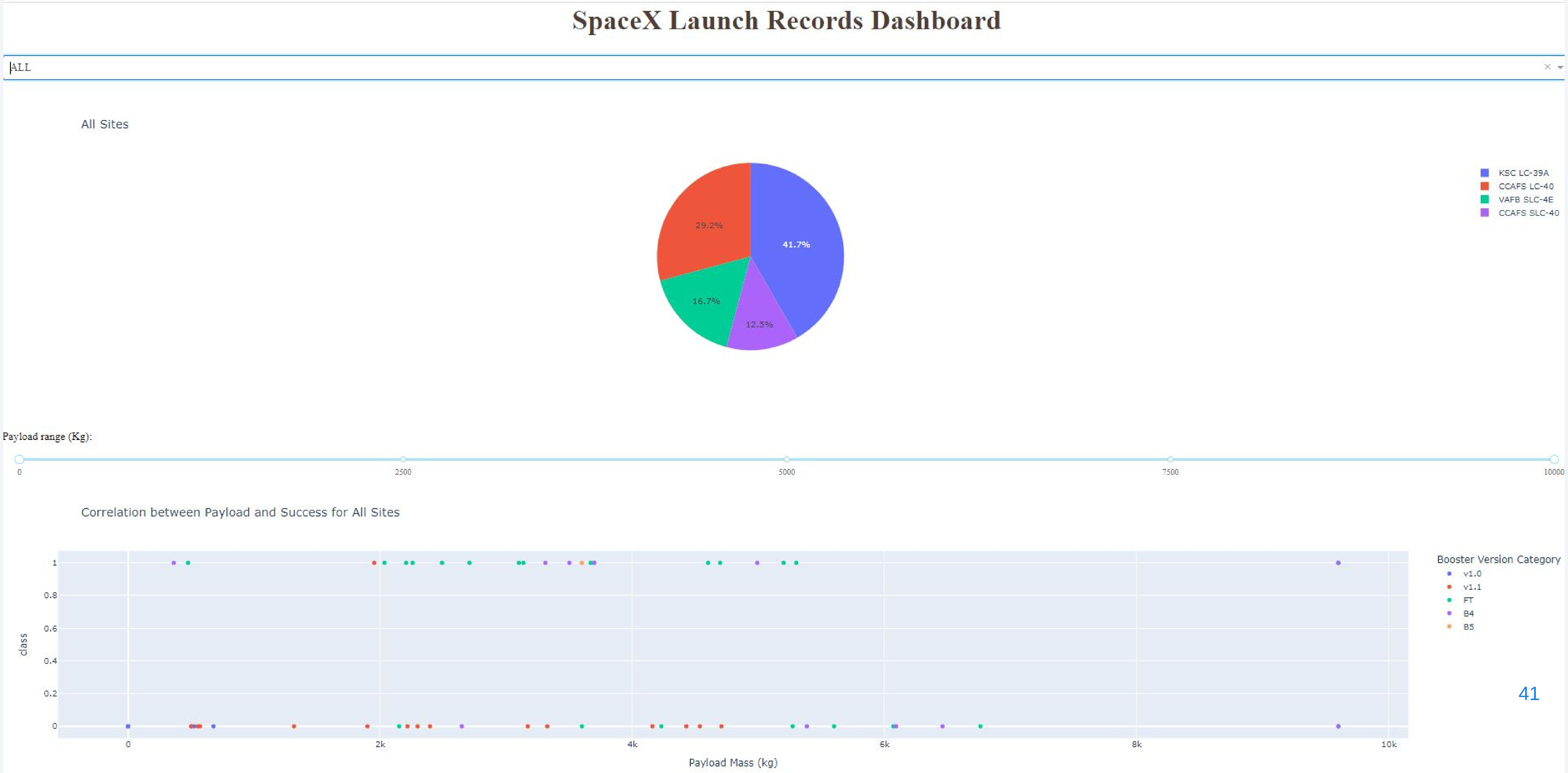
Launch success count for all the launching sites



KSC Launch Site Info:



Various Payload Launches Statistics:

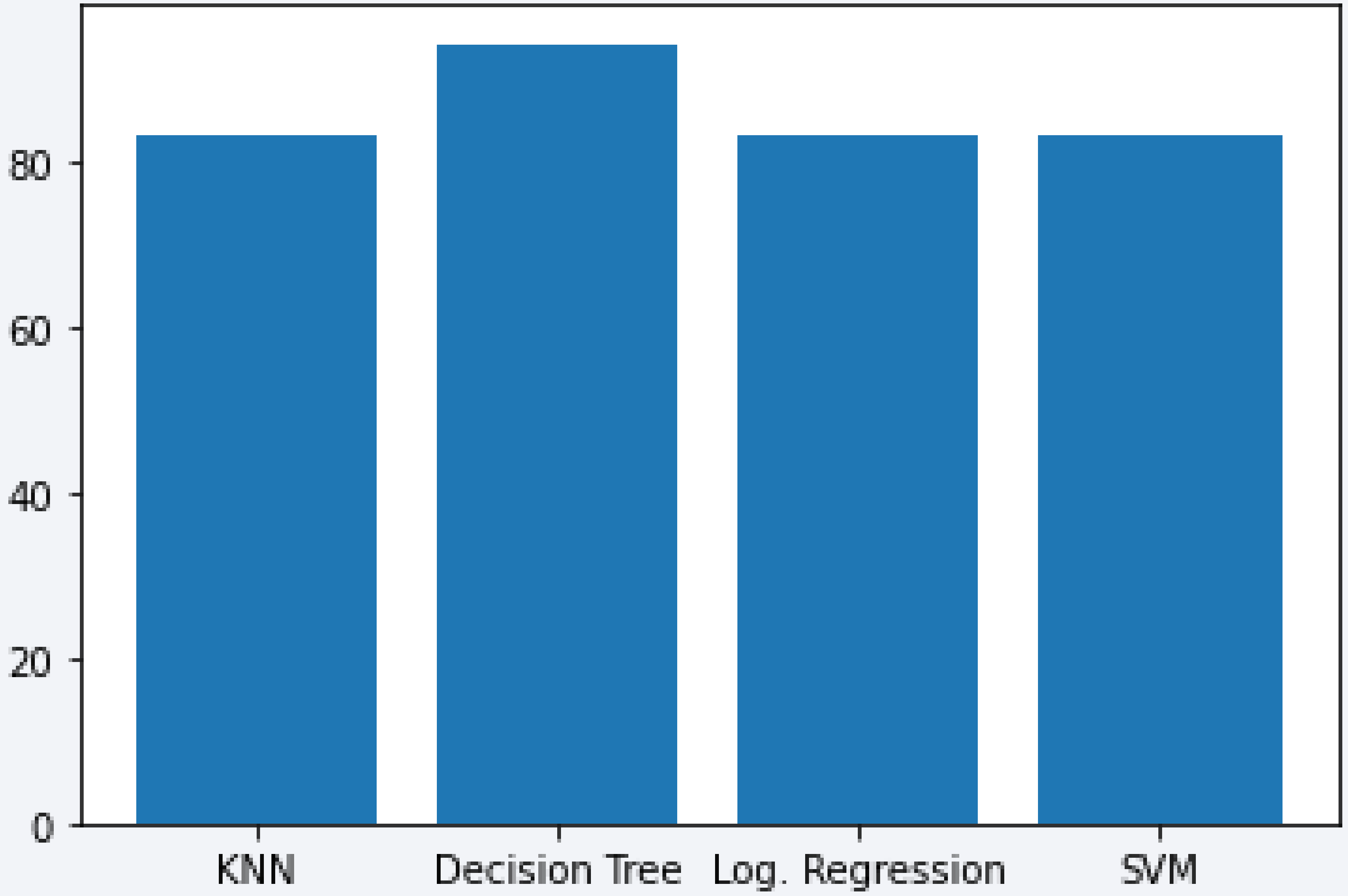


Section 5

Predictive Analysis (Classification)

Classification Accuracy

- KNN: 83.3
- Decision Tree: 94.4
- Logistic Regression: 83.3
- SVM: 83.3



Confusion Matrix for the best performing Decision Tree model:

- Only one False Positive results, all others classified correctly

Conclusions

- More recent launches are more successful
- The most succesful launching site is KSC-LC-39
- Heavier loads introduce a higher risk
- The best model Decision Tree can has 94.4% accuracy for out-of-sample data

Appendix

- Additional wikipedia information about launches

Thank you!