IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Gabin DJOMO NKUIGWA
09, Nov 2022

Outline

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

Executive Summary

To carry out this project, we borrowed the following road

- Project understanding
  - Know the Goal
  - Know the motivation
- Summary of methodologies
  - Data (*Collection / Understanding / Preparation*)
  - Location Launch
  - Machine Learning Models
- Summary of all results
  - Data Wrangling
  - Launch site map location
  - Static and interactive visualization
  - Machine Learning Prediction

Introduction

- In this study, the objective was to predict whether the Falcon 9 first stage will land successfully
- If we can determine if the first stage will land, we can determine the cost of a launch
- In this study, we will approach many concepts that we will see below
- We will present our research to you and we will discuss the findings

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

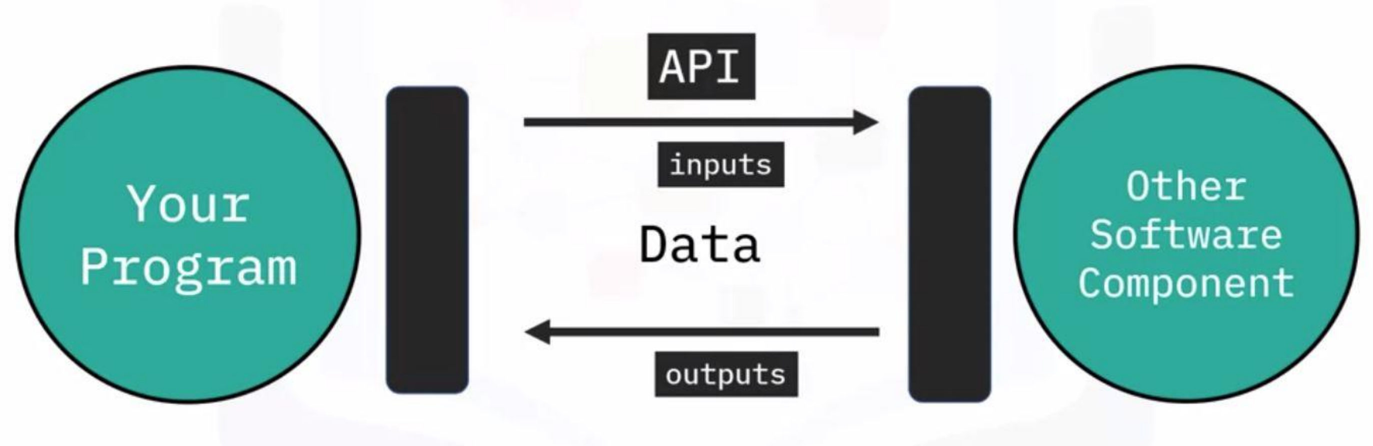
Data Collection

- The data is taken from a csv file placed in a remote server or through their official website
- We use an REST API which will connect to the remote platform to allow us to extract the data
  - First, our work was done on Jupyter Notebook code editor
  - We used its built-in API which provided us with an interface between the editor and remote data sources.
  - We connected to the data and extracted it
  - They are stored in a data frame

Data Collection – SpaceX API

- Our program use an API integrated to communicate with the remote server to get the file historical launch data and load it into the data frame, to handle late.

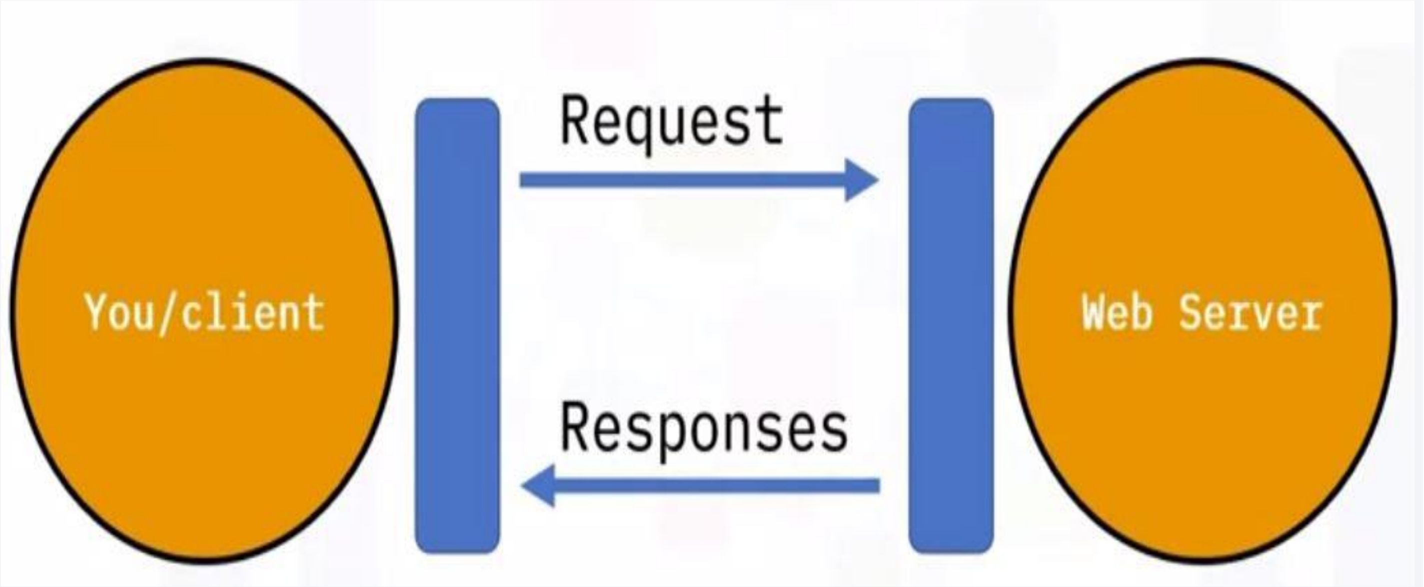
- This is the GitHub URL of the completed SpaceX API calls notebook https://github.com/DJOMOgabin/IBM-Project-Capstone/blob/master/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- Our program use an WebScraping Library (*Beautifulsoup*) to communicate with web pages to get the launch data and load it into the data frame, to handle late.

- This is the GitHub URL of the completed web scraping notebook https://github.com/DJOMOgabin/IBM-Project-Capstone/blob/master/jupyter-labs-webscraping.ipynb



Data Wrangling

- Data cleaning is done by pruning all unnecessary data
- We identify and mitigate the impact of missing data
- After do this, we Exploratory Data Analysis (*EDA*)
- We determinate training labels
- We normalize data

- This is the Github link of the completed Wrangling Notebook https://github.com/DJOMOgabin/IBM-Project-Capstone/blob/master/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- The plots that are presented are as follows :
  - *We plot FlightNumber vs PayloadMass in scatter plot to see how these variables would affect the launch outcome*
  - *We plot FlightNumber vs LaunchSite in scatter plot to see how these variables would affect the launch outcome*
  - *We plot PayloadMass vs LaunchSite in scatter plot to see how these variables would affect the launch outcome*
  - *We plot Orbit vs Class to visually check if there are any relationship between these variables*
  - *Etc…*
- Visualize the launch success yearly trend

https://github.com/DJOMOgabin/IBM-Project-Capstone/blob/master/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- SQL Queries
  - *total payload mass carried by boosters launched by NASA (CRS)*
  - *average payload mass carried by booster version F9 v1.1*
  - *the date when the first succesful landing outcome in ground pad was acheived*
  - *the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000*
  - *the total number of successful and failure mission outcomes*
  - *the names of the booster_versions which have carried the maximum payload mass*

- https://github.com/DJOMOgabin/IBM-Project-Capstone/blob/master/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- map object we created:
  - We have located all launch sites on the map
  - We added a circle around it to circumscribe the launch site
  - We have added markers to highlight all the launches that have been made on this site, the color differentiates if the launch was a success or a failure
  - Lines have been added to represent the between one point and another

https://github.com/DJOMOgabin/IBM-Project-Capstone/blob/master/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- We plot:
  - Total success launches all sites
  - Success rate launches for each site
  - Correlation between Payload and Success for all sites in Range of Payload Mass
  - Correlation between Payload and Success for each site

https://github.com/DJOMOgabin/IBM-Project-Capstone/blob/master/spacex_dash_app.py

Predictive Analysis (Classification)

- Our Model development process are as follow:
  - Enumerate different algorithm
  - Use the GridSearchCV with different parameter to determinate the best parameter for each algorithm
  - Split data in train data and test data
  - Fit each algorithm with train data
  - Calculate the accuracy of each algorithm with train data
  - Plot the confusion matrix to have an overview for each algorithm

https://github.com/DJOMOgabin/IBM-Project-Capstone/blob/master/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

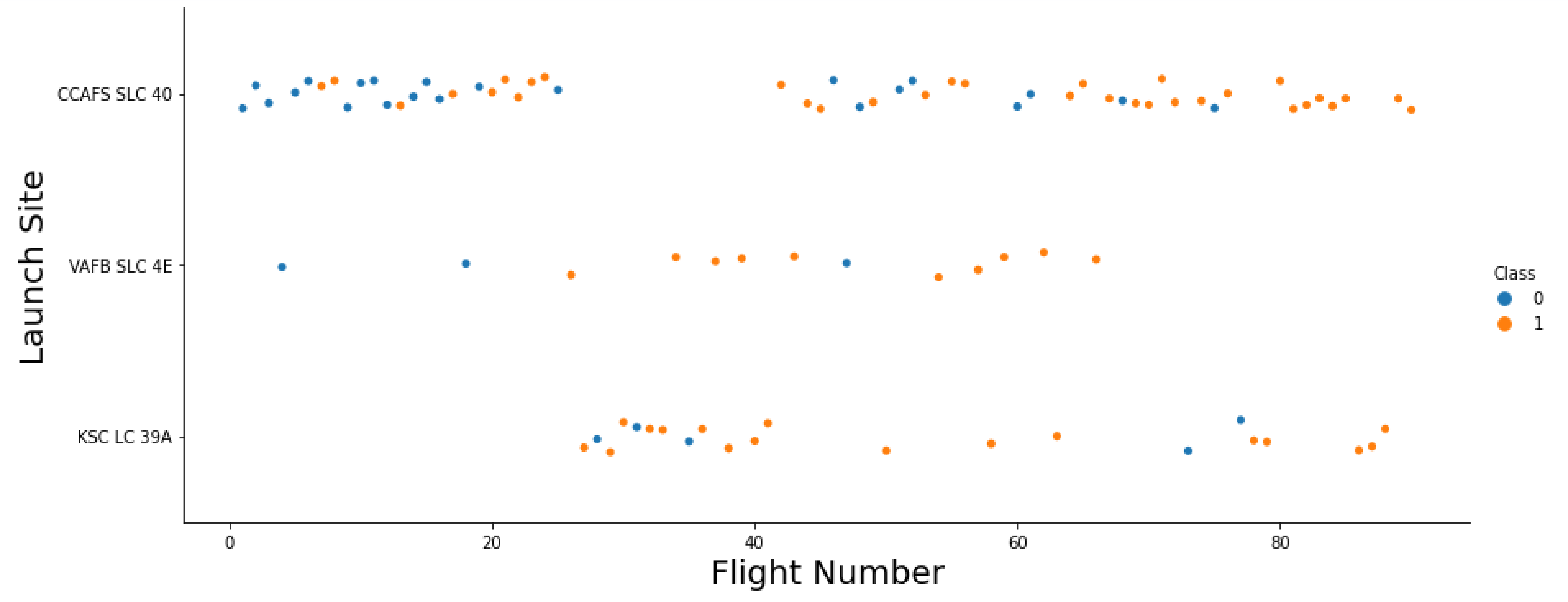
Results

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

Section 2

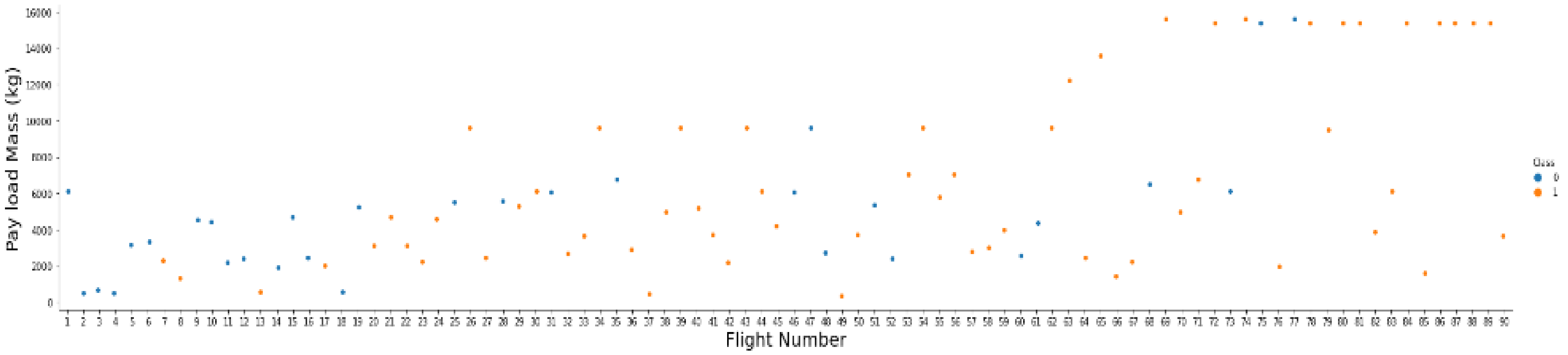
Insights drawn from EDA

Flight Number vs. Launch Site



We observe that for each site, the higher the FlightNumber, the higher our chance of success.

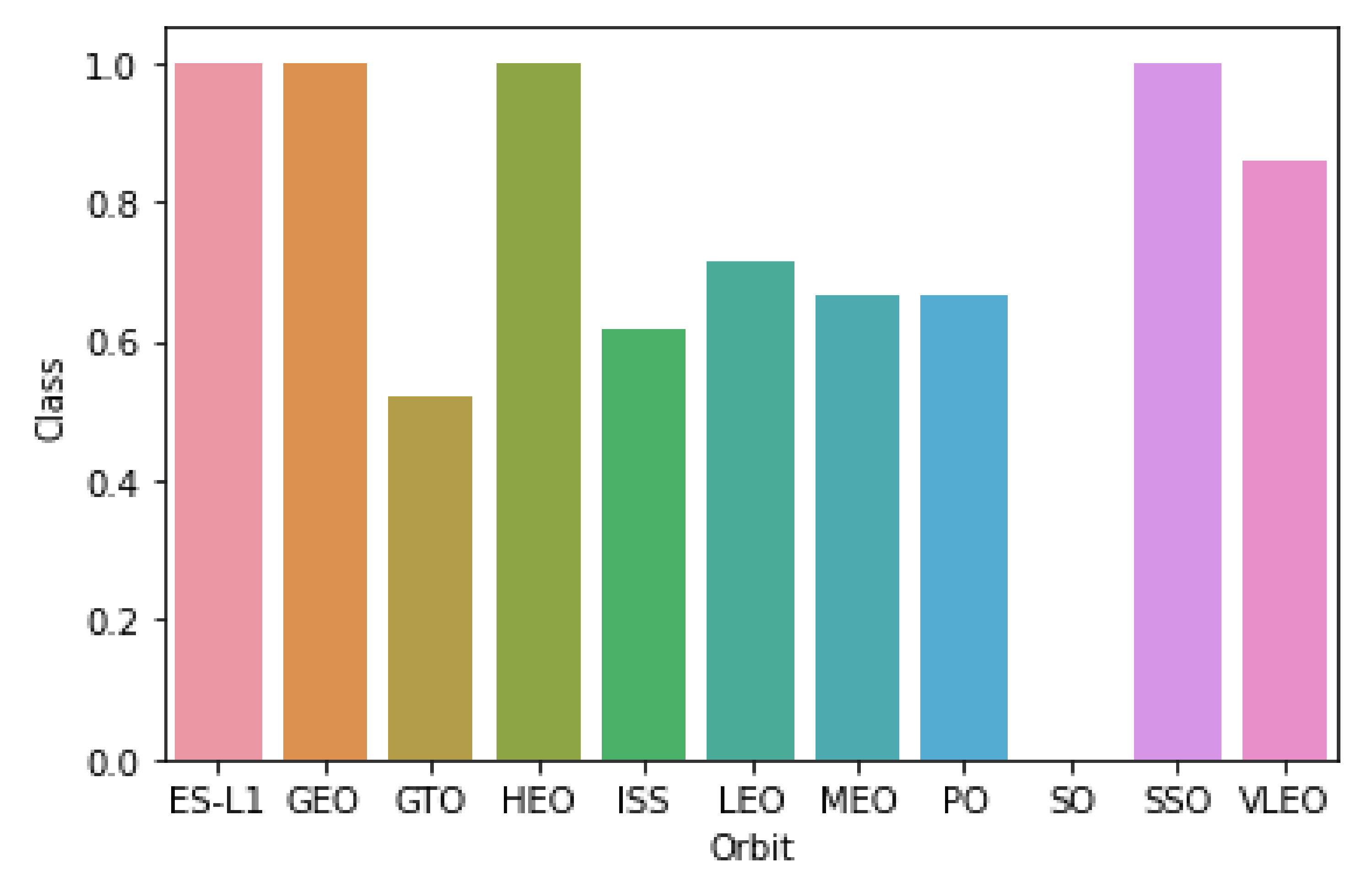
Payload vs. Launch Site



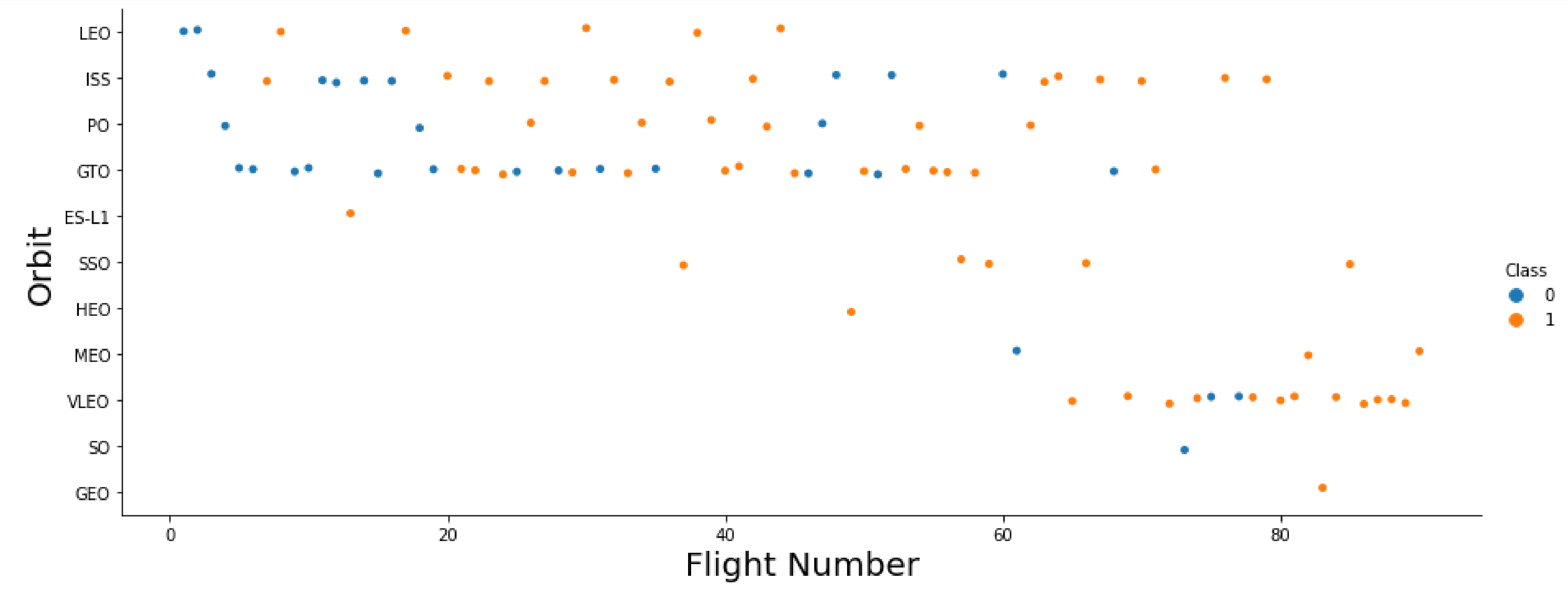
We see that different launch sites have different success rates. CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%

Success Rate vs. Orbit Type

- We have average launch success following different orbit

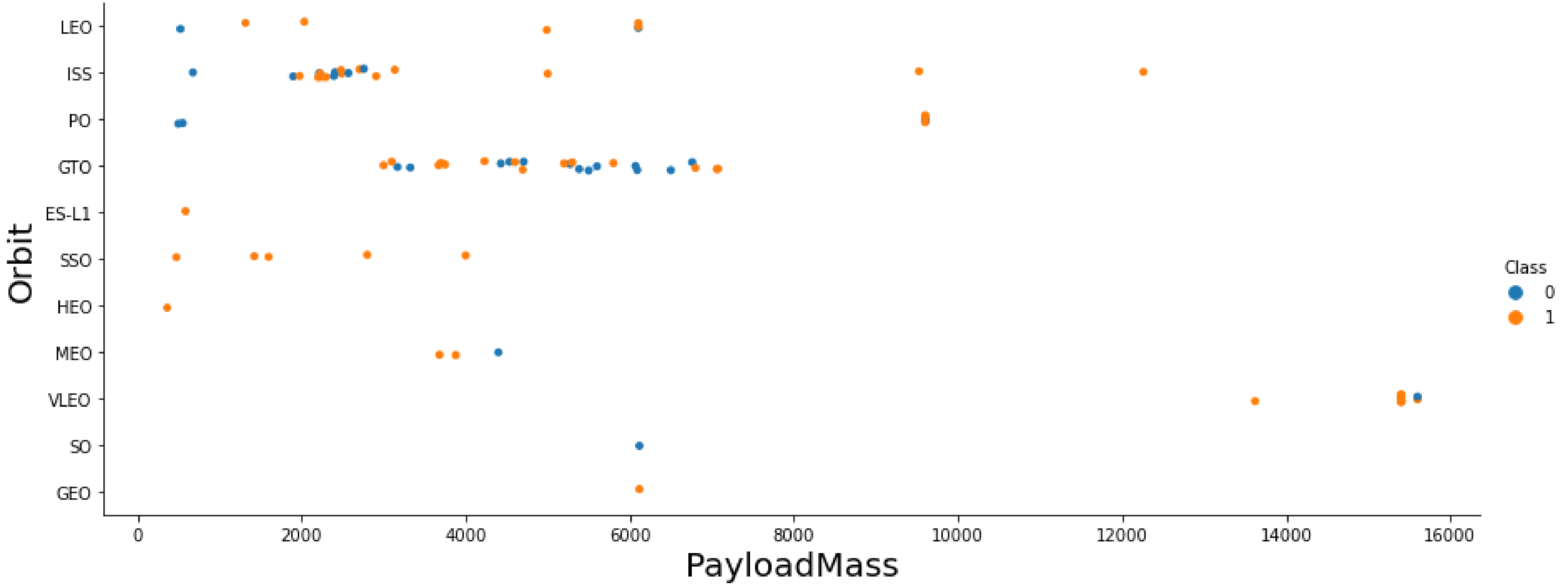


Flight Number vs. Orbit Type



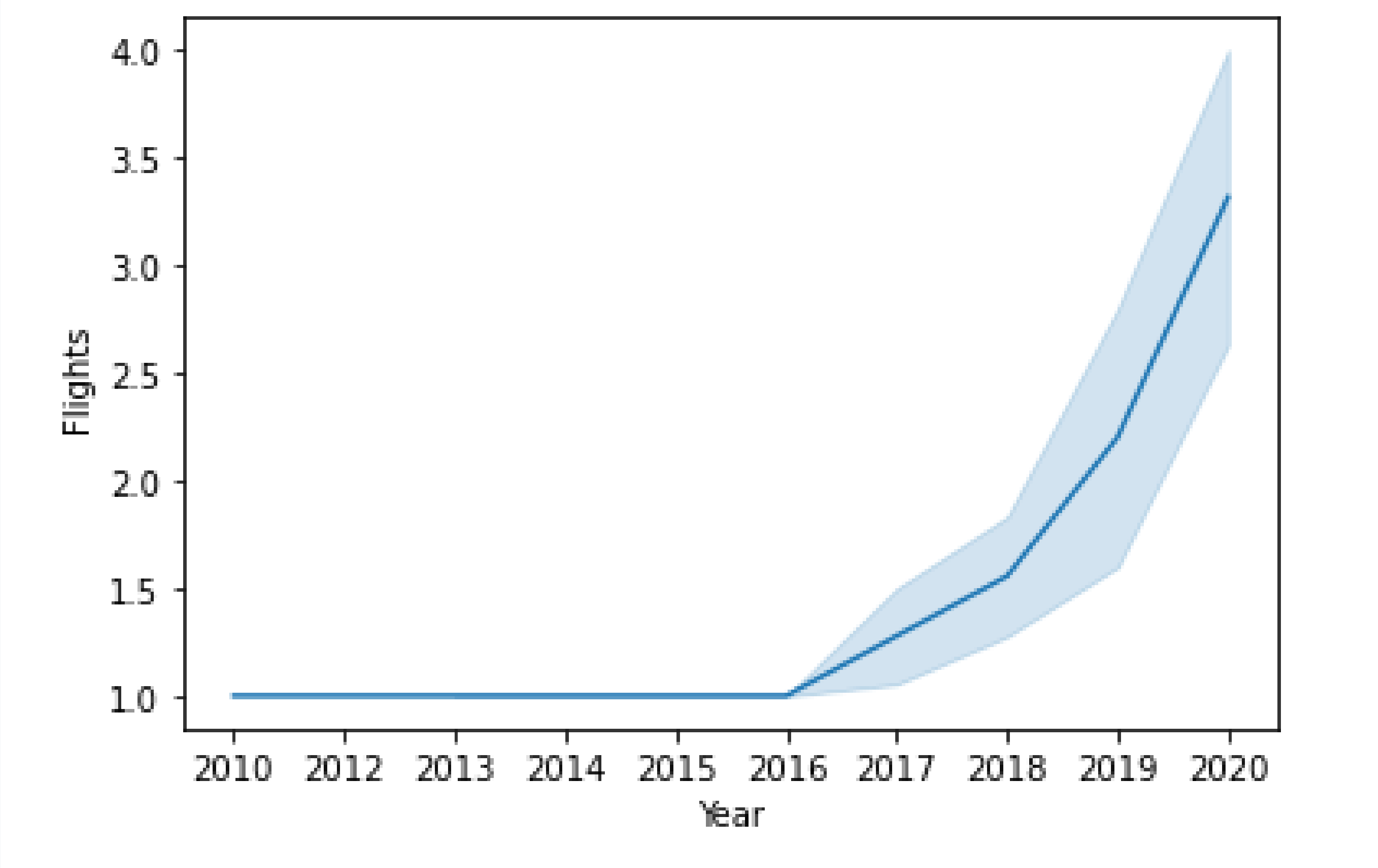
- For all orbits, the higher the FlightNumber, the better the result

Payload vs. Orbit Type



Launch Success Yearly Trend

The number of Flights has increased over time. and we experienced strong growth from 2016



All Launch Site Names

- **Launch_Site:**
  - CCAFS LC-40
  - VAFB SLC-4E
  - KSC LC-39A
  - CCAFS SLC-40

- Query: select distinct "Launch_Site" from spacextbl

Launch Site Names Begin with 'CCA'

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

Query: select * from spacextbl where "Launch_Site" like "CCA%" limit 5

Total Payload Mass

Total_payload: 45596

Query: select sum("PAYLOAD_MASS__KG_") total_payload from spacextbl where Customer='NASA (CRS)'

Average Payload Mass by F9 v1.1

Average: 2534.66

Query: select avg("PAYLOAD_MASS__KG_") avg from spacextbl where "Booster_Version" like 'F9 v1.1%'

First Successful Ground Landing Date

Date: 01-05-2017

Query: select min(date) first from spacextbl where "Landing _Outcome" like 'Success (ground pad)%'

Successful Drone Ship Landing with Payload between 4000 and 6000

Query: select Booster_Version from spacextbl where "Mission_Outcome"="Success" and "PAYLOAD_MASS__KG_" between 4000 and 6000

Booster_Version
F9 v1.1
F9 v1.1 B1011
F9 v1.1 B1014
F9 v1.1 B1016
F9 FT B1020
F9 FT B1022
F9 FT B1026
F9 FT B1030
F9 FT B1021.2
F9 FT B1032.1
F9 B4 B1040.1
F9 FT B1031.2
F9 FT B1032.2
F9 B4 B1040.2
F9 B5 B1046.2
F9 B5 B1047.2
F9 B5 B1046.3
F9 B5 B1048.3
F9 B5 B1051.2
F9 B5B1060.1
F9 B5 B1058.2
F9 B5B1062.1

Total Number of Successful and Failure Mission Outcomes

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

Query: select "Mission_Outcome", count(date) nbr from spacextbl group by "Mission_Outcome"

Boosters Carried Maximum Payload

Query: select "Booster_Version" from spacextbl where "PAYLOAD_MASS__KG_" in (select max("PAYLOAD_MASS__KG_") from spacextbl)

Booster_Version
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

2015 Launch Records

month	Booster_Version	Launch_Site
01	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40

Query: select substr(Date, 4,2) month, "Booster_Version", "Launch_Site" from spacextbl where substr(Date,7,4)='2015' and "Landing _Outcome" = "Failure (drone ship)"

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

- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

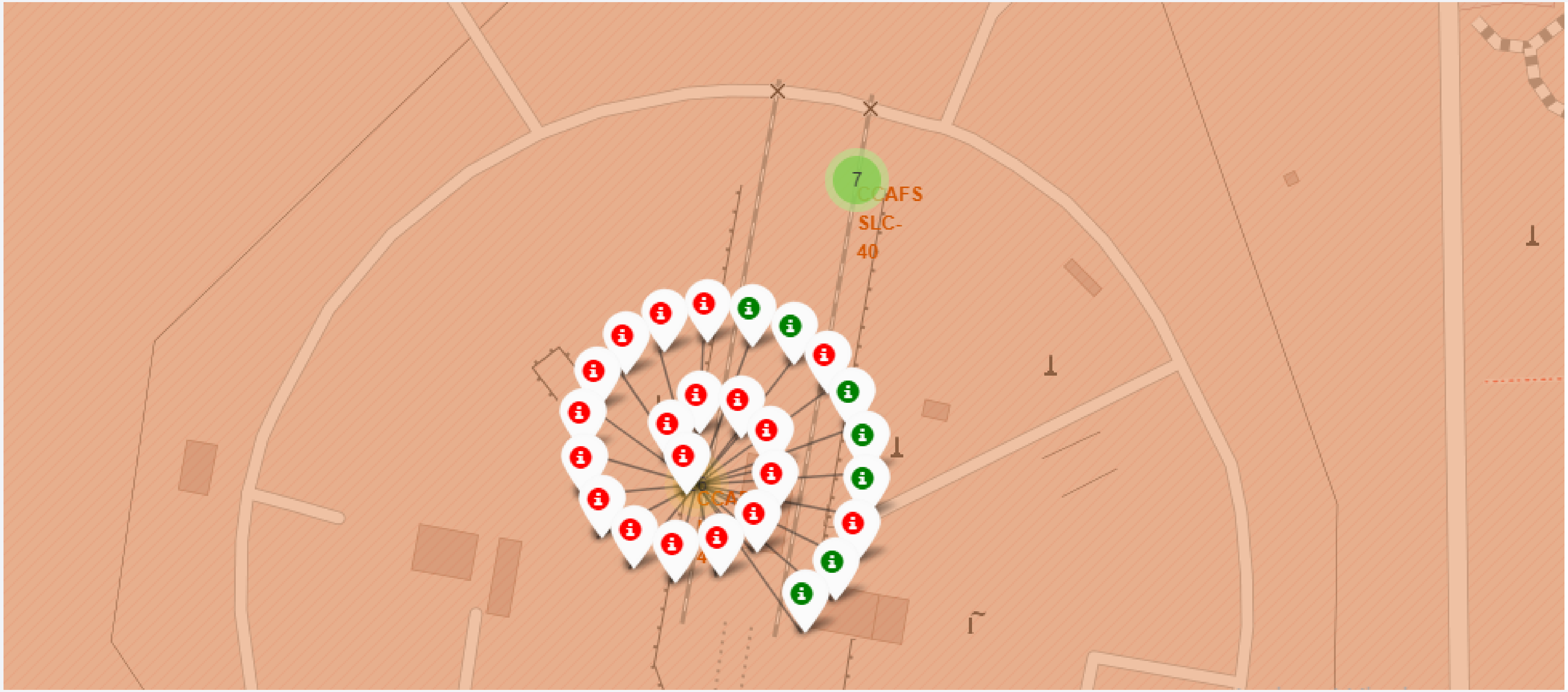
- Present your query result with a short explanation here

Section 3

Launch Sites Proximities Analysis

Mark all Launch Sites on the map

Mark the success/failed launches for each site on the map



Calculate the distances between a launch site to its proximities

- Replace <Folium map screenshot 3> title with an appropriate title

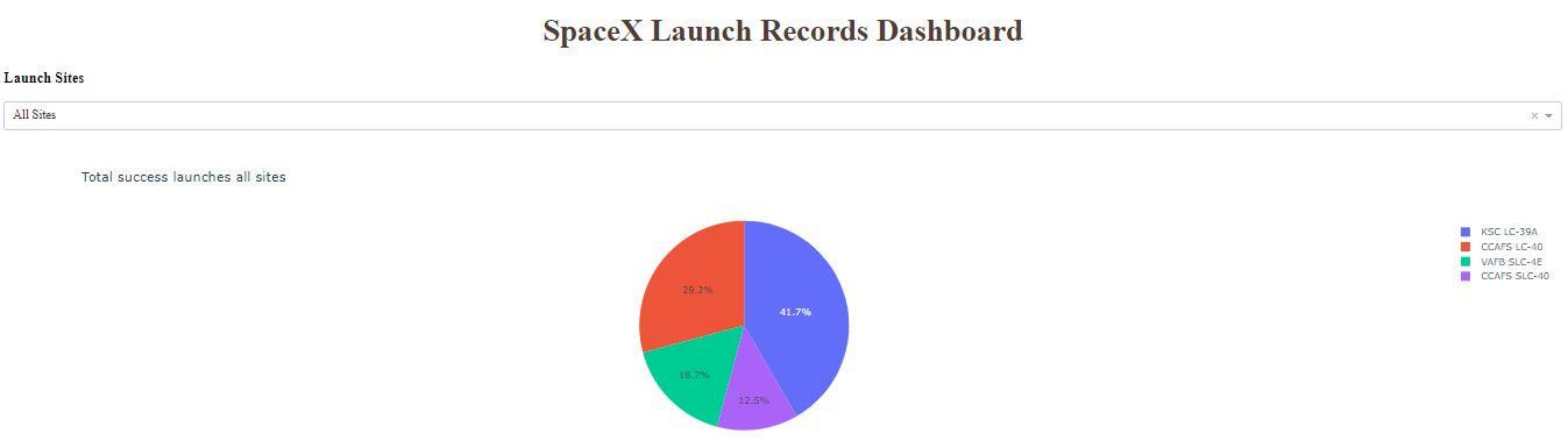
- Explore the generated folium map and show the screenshot of a selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed

- Explain the important elements and findings on the screenshot

Section 4

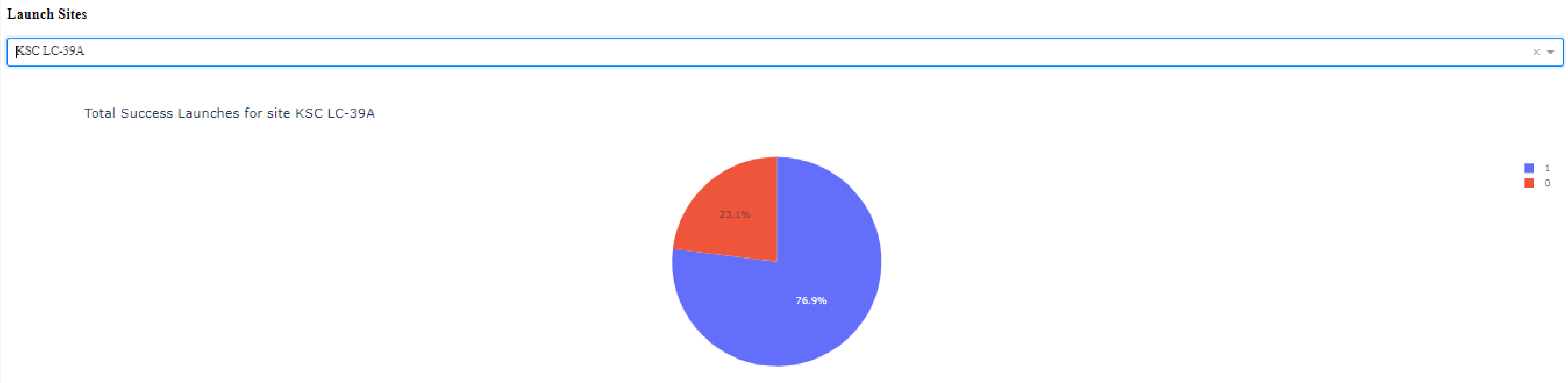
Build a Dashboard with Plotly Dash

Total success launches all sites



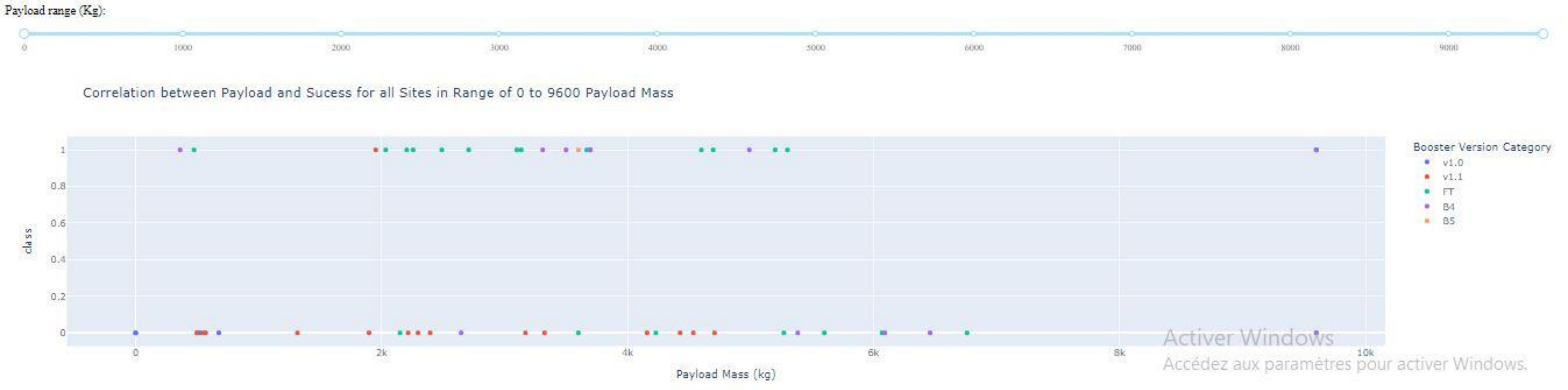
- We can see that the one with the best rate is KSC LC-39A

Total Success launches for site KSC LC-39A



- We observe that this site has a success rate of more than 75%

Correlation between Payload and Success

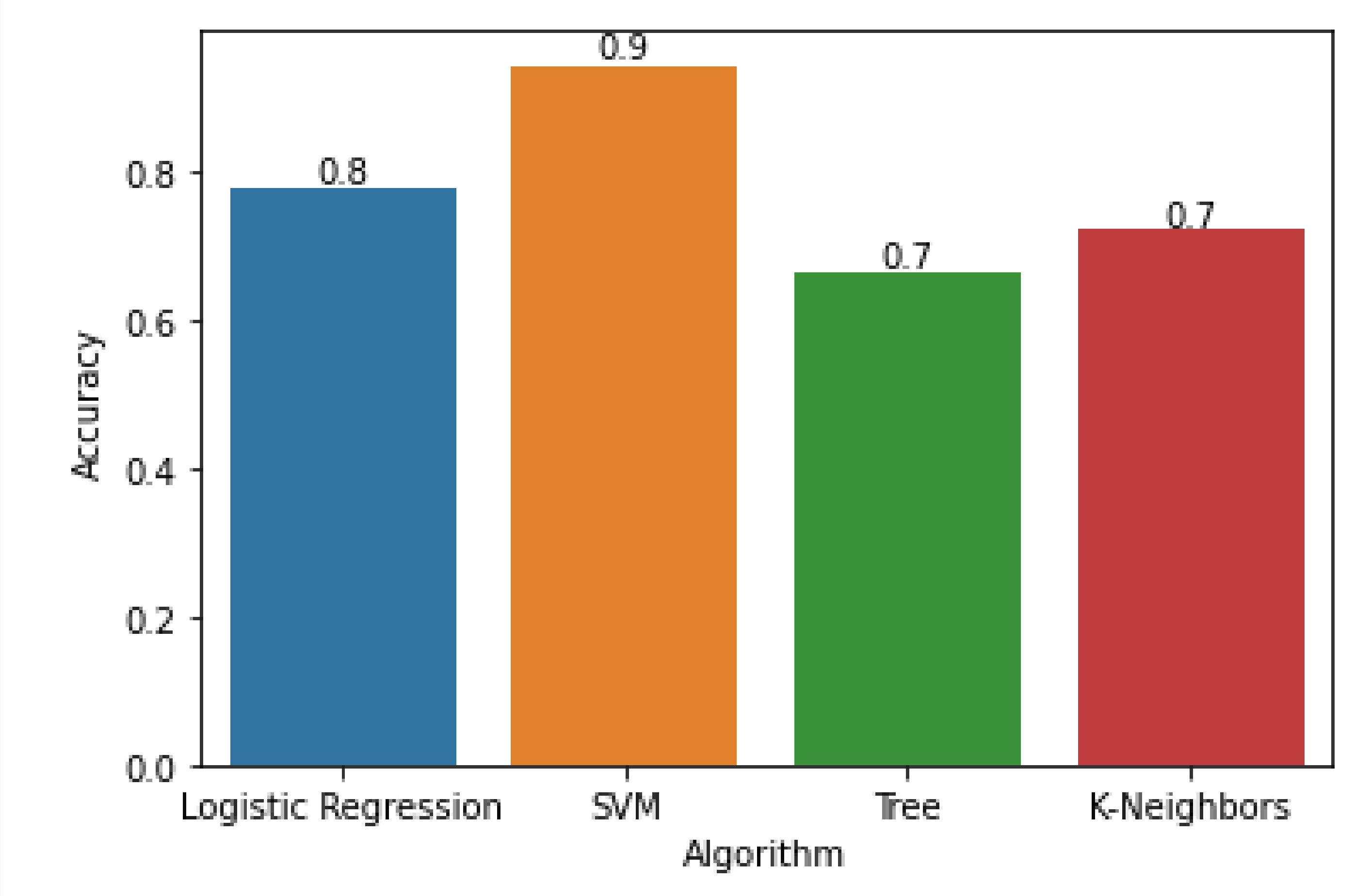


Section 5

Predictive Analysis (Classification)

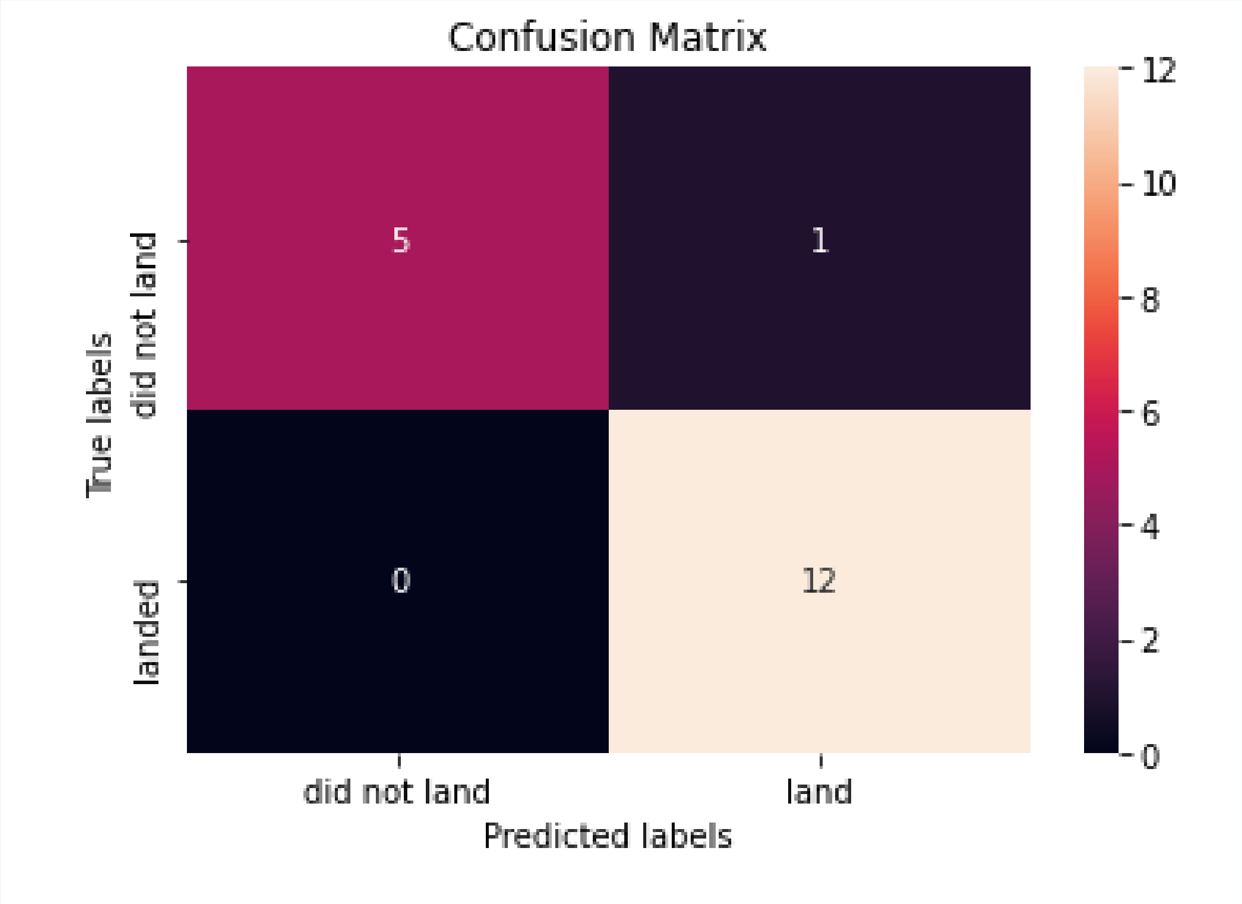
Classification Accuracy

According to the graph opposite, we can easily deduce that Support Vector Machine has better accuracy than other algorithms.



Confusion Matrix

- The algorithm has a better prediction when the operation is a success, but on the other hand, the prediction is worse when it is a failure



Conclusions

- Our purpose was to predict whether the Falcon 9 first stage will land successfully
- we answered it following this route plan:
  - Data Collection (*Through REST API*)
  - Data Wrangling
  - Exploratory Data Analysis through SQL Queries, Visualization
  - Overview different site of launch on the map
  - Built an interactive dashboard for more overview
  - Built, fit and test several prediction model and find the best
- We can conclude that launch success depends on several variables and the reuse of certain parts for launches is possible.

Appendix

Thank you!