IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Md Amanullah Nayeem
7/31/2024

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

- In this report, We have determined the price of each launch by gathering information about Space X and creating dashboards.
- We also trained a machine learning model and use public information to predict if SpaceX will reuse the first stage.
- From the various methods, we found Decision Tree get the more accurate prediction.

Introduction

- The commercial space age is here, companies are making space travel affordable for everyone.
- One reason SpaceX can do this is the rocket launches are relatively inexpensive because SpaceX can reuse the first stage.
- Our Company, Space Y that would like to compete with SpaceX.
- We need to determine the price of each launch.
- We also need to determine if SpaceX will reuse the first stage.

Section 1

Methodology

Methodology

- Data collection methodology:
  - We have used SpaceX REST API, which give us data about launches, the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome.
  - The SpaceX REST API endpoints starts with api.spacexdata.com/v4/.
  - We will use this URL to target a specific endpoint of the API to get past launch data.
  - We will perform a get request using the requests library to obtain the launch data, which we will use to get the data from the API.
  - This result can be viewed by calling the .json() method.
  - To convert this JSON to a dataframe, we have used the json_normalize function.
  - Falcon 9 Launch data obtained from web scraping related Wiki pages.

Methodology

- Data collection methodology:
  - We have used the Python BeautifulSoup package to web scrape some HTML tables that contain valuable Falcon 9 launch records.
  - Then have parsed the data from those tables and convert them into a Pandas data frame for further visualization and analysis.
  - This raw data transformed into a clean dataset which provides meaningful data on the situation we are trying to address:
    - Wrangling Data using an API,
    - Sampling Data, and
    - Dealing with Nulls.

Methodology

- Data collection methodology:
  - In some of the columns, we have an identification number.
  - We have used the API again targeting another endpoint to gather specific data for each ID number.
  - The data will be stored in lists and will be used to create our dataset.
  - Launch data includes data for the Falcon 1 booster, whereas we only want falcon 9.
  - We have filter/sample the data to remove Falcon 1 launches.

Methodology

- Perform data wrangling
  - Some of the attributes:
    - Flight Number
    - Date
    - Booster version
    - Payload mass Orbit
    - Launch Site
    - Outcome : This is the status of the first stage Flights
    - Grid Fins: These help with landing Reused
    - Legs: used in landing pad
    - Block
    - Reused count
    - Serial
    - Longitude and latitude of launch

Methodology

- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash

Methodology

- Perform predictive analysis using classification models
  - We have build a machine learning pipeline to predict if the first stage of the Falcon 9 lands successfully.
  - This will include:
    - Preprocessing: Allowing to standardize our data
    - Train_test_split: Allowing to split data into training and testing data
  - Train the model and perform Grid Search to find the hyperparameters that allow a given algorithm to perform best.
  - Using the best hyperparameter values, we will determine the model with the best accuracy using the training data.

Methodology

- Perform predictive analysis using classification models
  - We have test using:
    - Logistic Regression
    - Support Vector machines
    - Decision Tree Classifier
    - K-nearest neighbors.

Data Collection

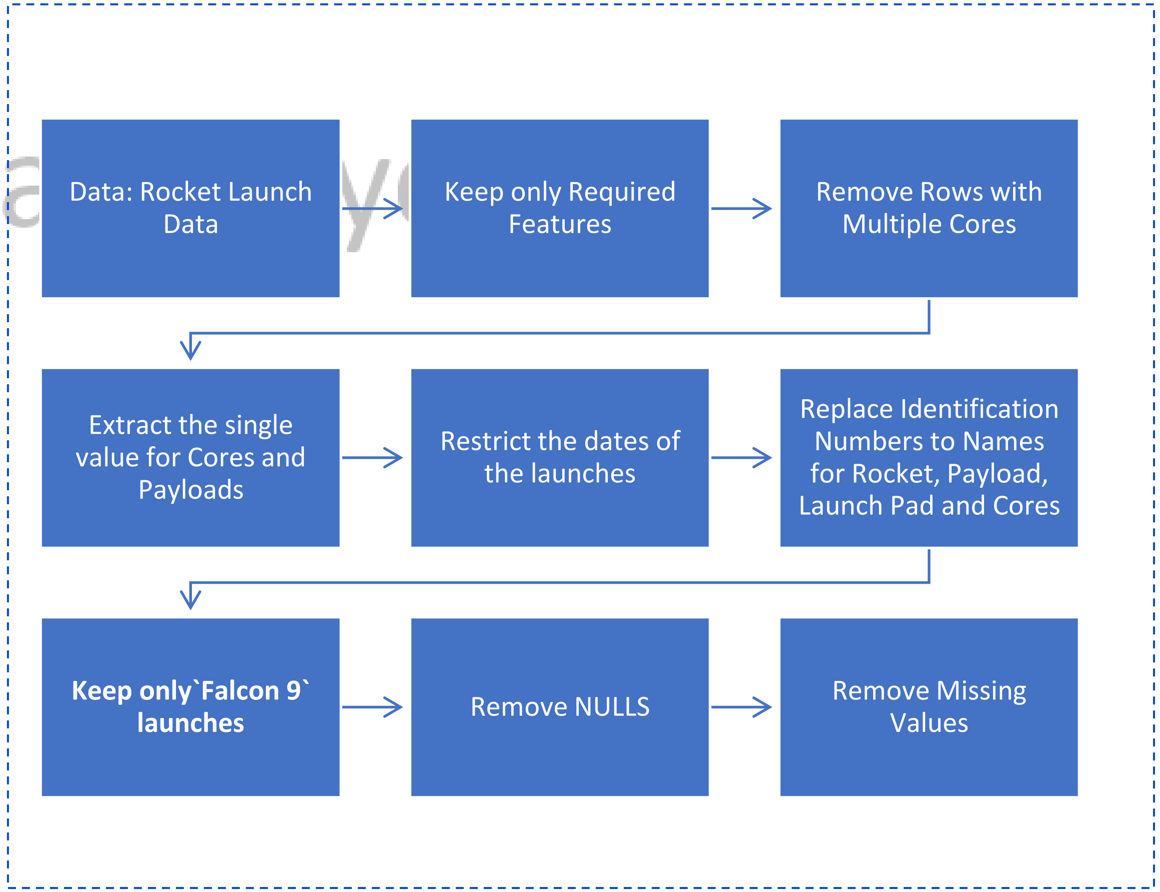
- We have defined a series of helper functions that will help us use the API to extract information using identification numbers in the launch data:
  - Rocket: https://api.spacexdata.com/v4/rockets/
  - Launch Pad: https://api.spacexdata.com/v4/launchpads/
  - Payloads: https://api.spacexdata.com/v4/payloads/
  - Cores: https://api.spacexdata.com/v4/cores
  - Rocket Launch Data: https://api.spacexdata.com/v4/launches/past with only the below columns:
    - 'rocket', 'payloads', 'launchpad', 'cores', 'flight_number', 'date_utc'

Data Collection

- Removed rows with multiple cores
- Restricted the dates of the launches
- Filter the Data Frame to only include `Falcon 9` launches

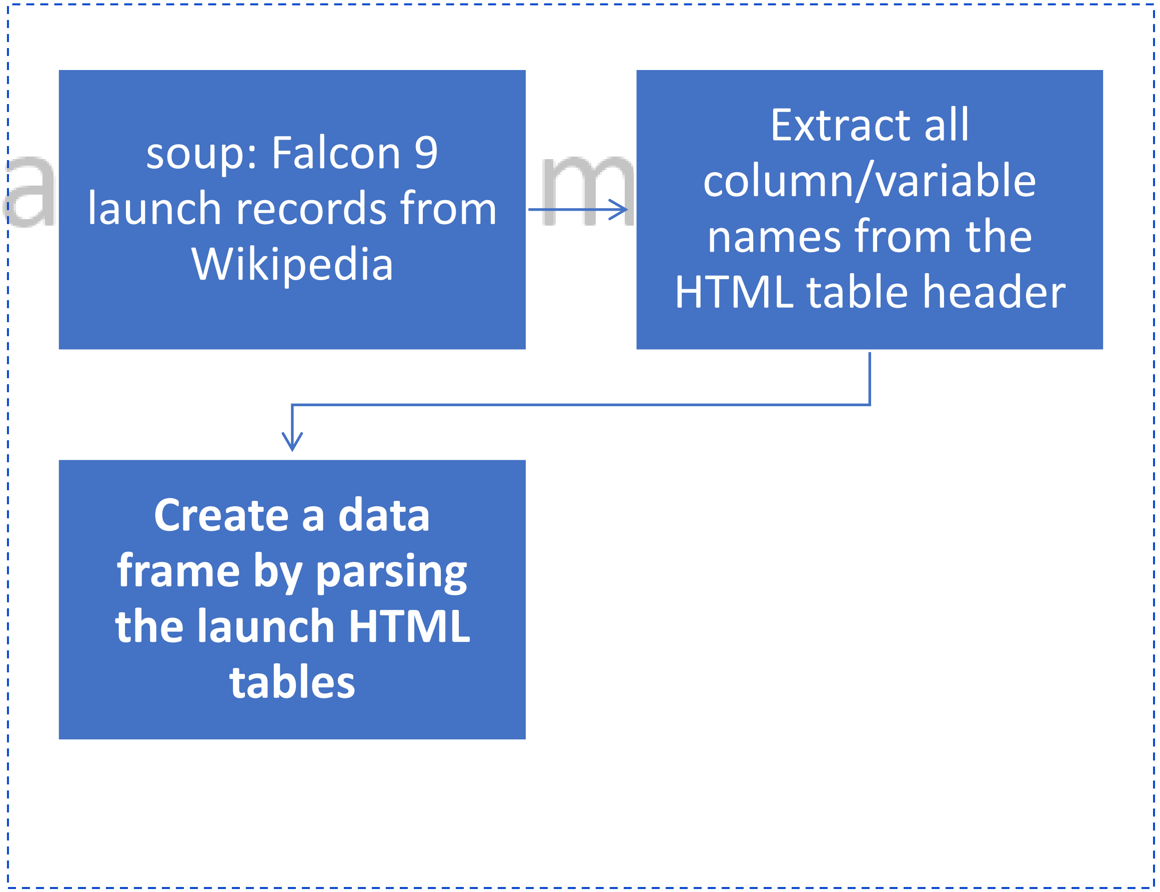
Data Collection – SpaceX API

- GitHub URL of the completed SpaceX API calls notebook:
- https://github.com/nym2010/Applied-Data-Science-Capstone/blob/2479d34100a407b3e1ca82fa04dd39822c548b4b/jupyter-labs-spacex-data-collection-api.ipynb



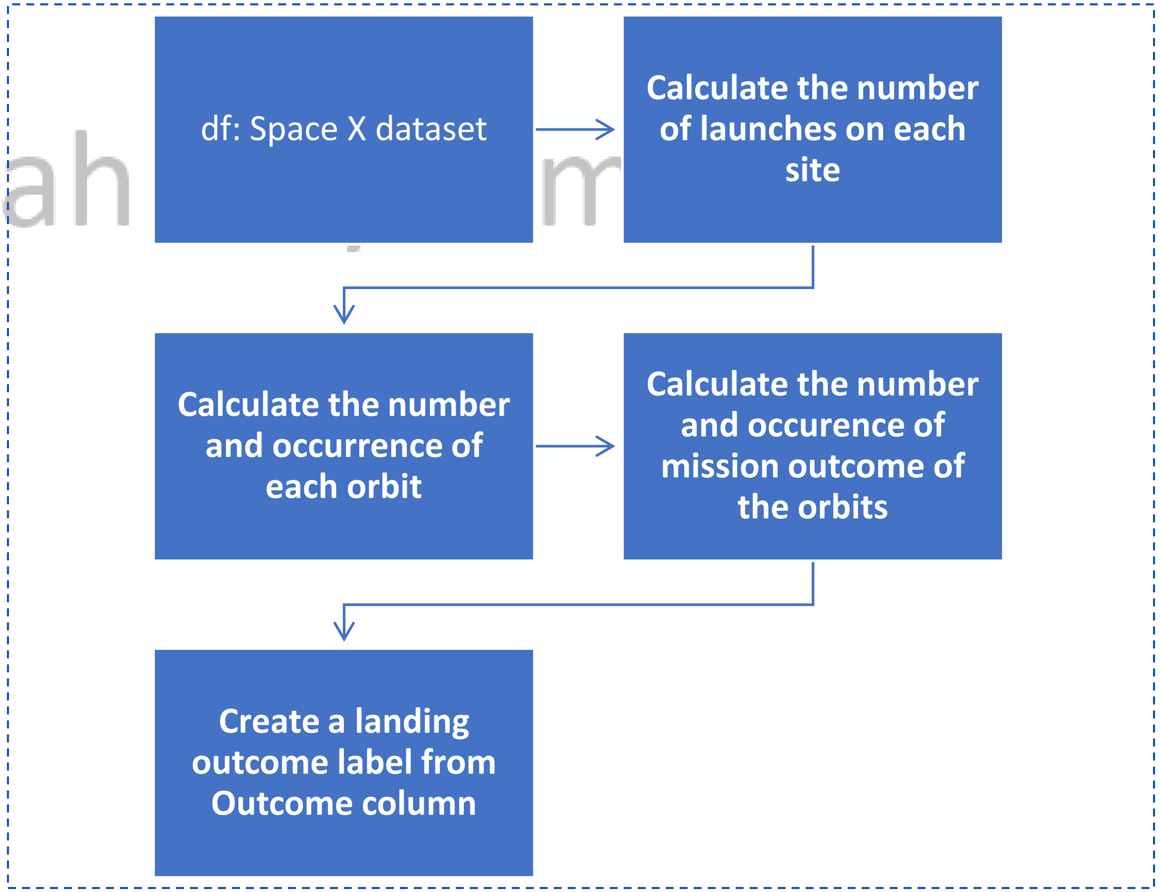
Data Collection - Scraping

- GitHub URL of the completed web scraping notebook:
- https://github.com/nym2010/Applied-Data-Science-Capstone/blob/2479d34100a407b3e1ca82fa04dd39822c548b4b/jupyter-labs-webscraping.ipynb



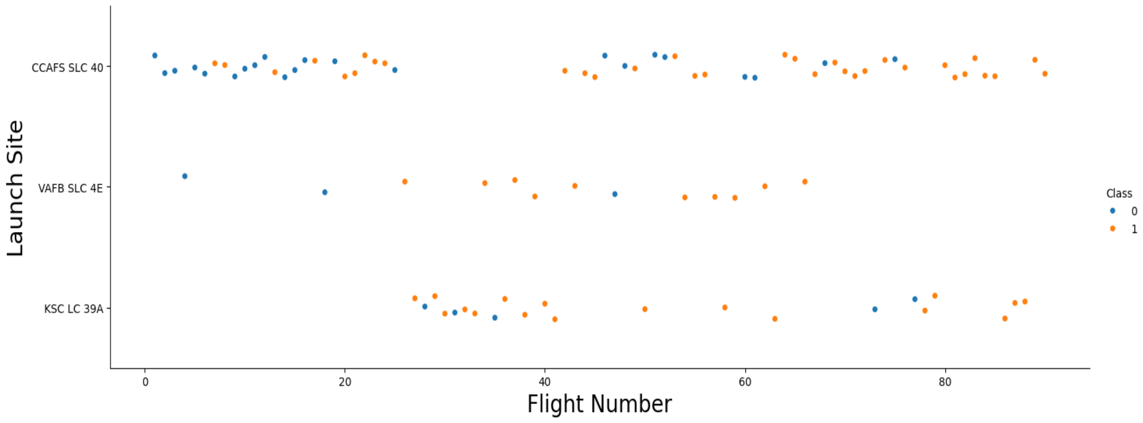
Data Wrangling

- GitHub URL of the completed web scraping notebook:
- https://github.com/nym2010/Applied-Data-Science-Capstone/blob/2479d34100a407b3e1ca82fa04dd39822c548b4b/labs-jupyter-spacex-Data%20wrangling.ipynb

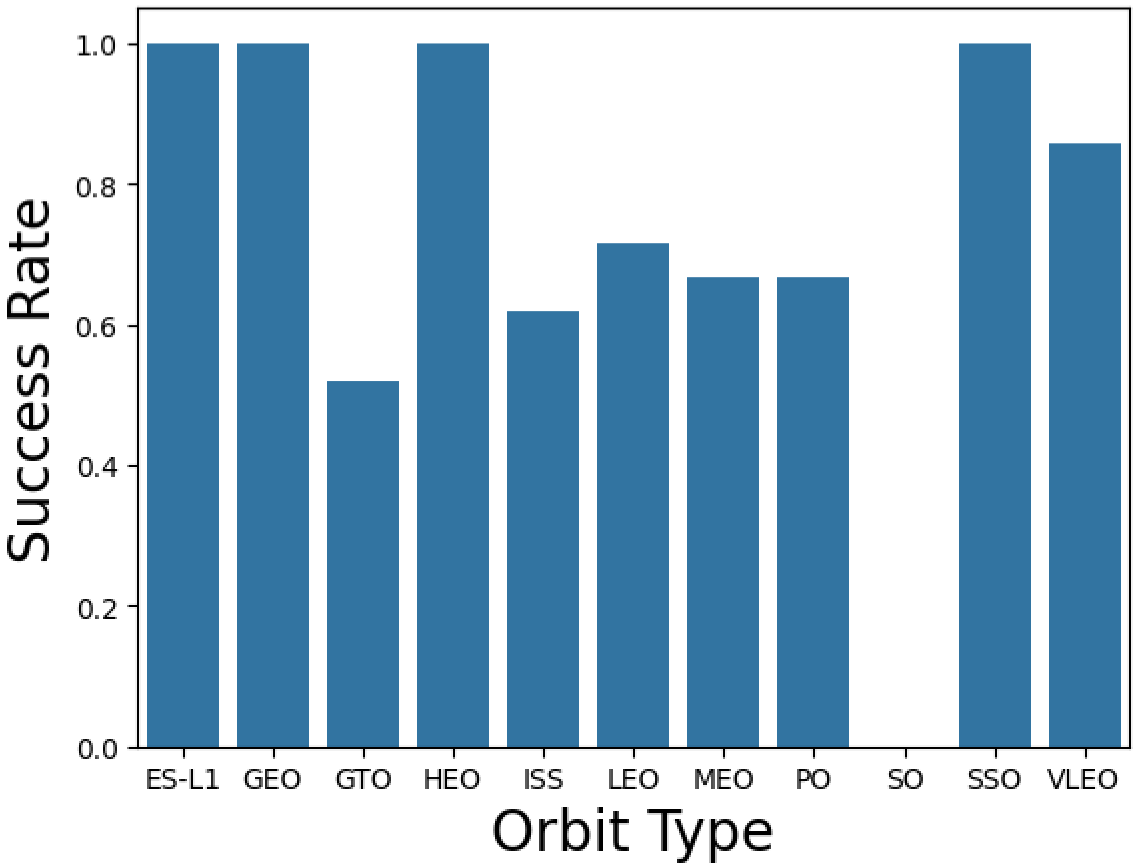


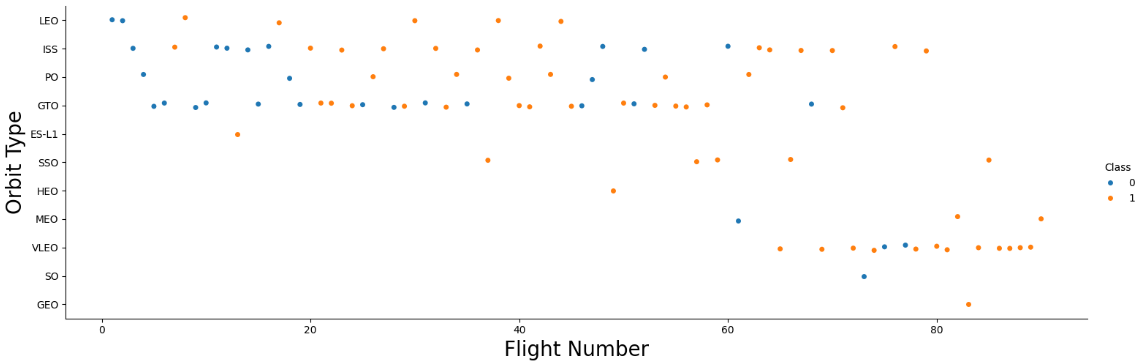
EDA with Data Visualization

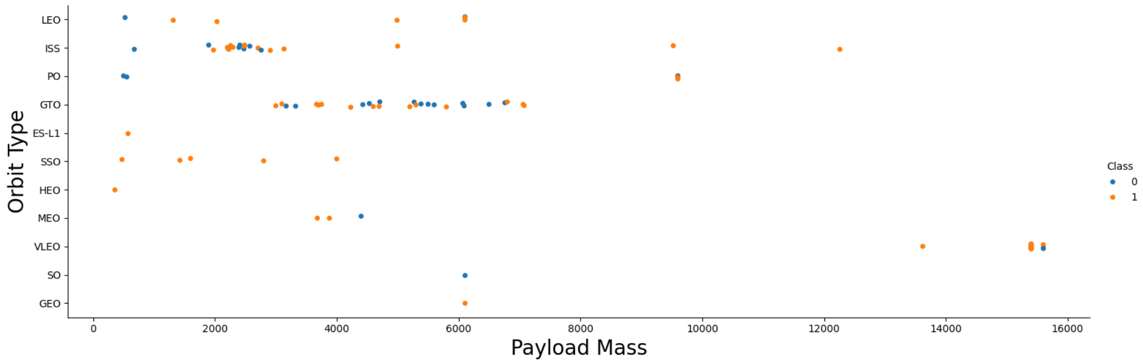
- GitHub URL of the completed web scraping notebook:
- https://github.com/nym2010/Applied-Data-Science-Capstone/blob/2479d34100a407b3e1ca82fa04dd39822c548b4b/jupyter-labs-eda-dataviz.ipynb



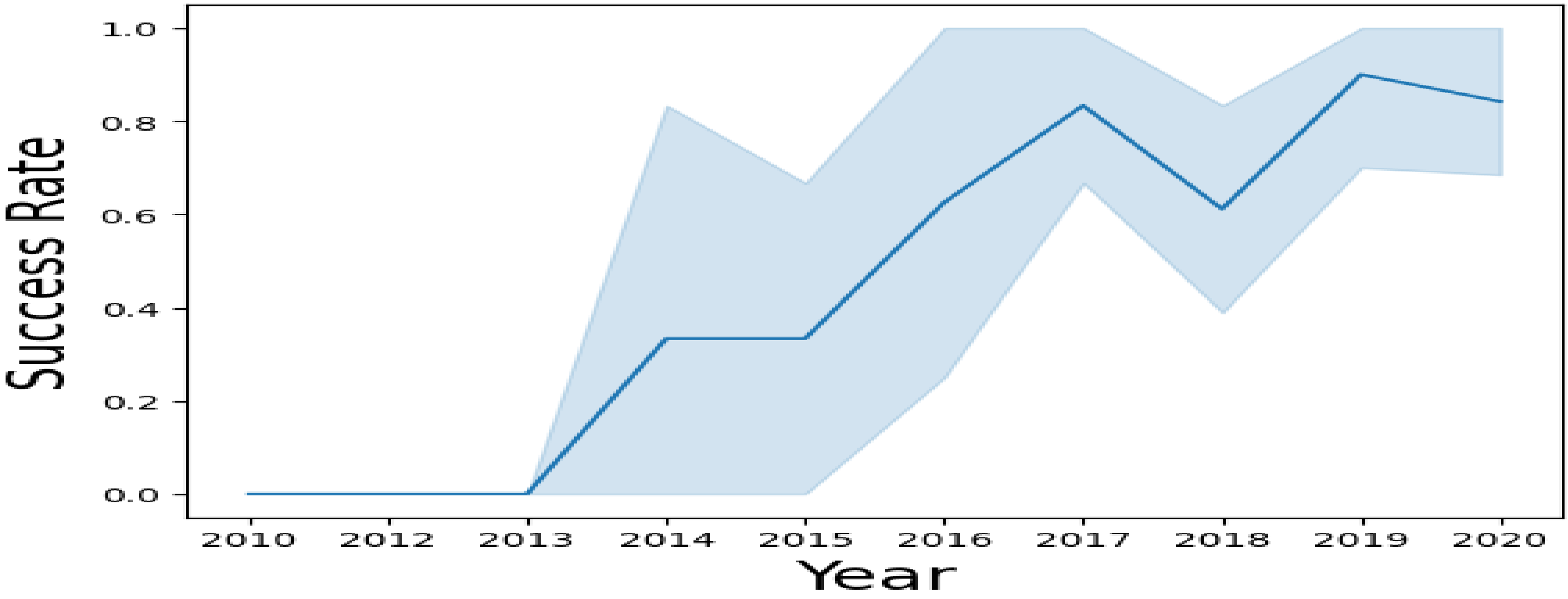
EDA with Data Visualization







EDA with Data Visualization



EDA with SQL

- GitHub URL of the completed web scraping notebook:
- https://github.com/nym2010/Applied-Data-Science-Capstone/blob/2479d34100a407b3e1ca82fa04dd39822c548b4b/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Map objects created:
  - Map: Create a map with all Launch Sites
  - Markers: Create an icon as a text label
  - Circles: Add a highlighted circle area with a text label on Launch Sites
  - Marker Cluster: To simplify a map containing many markers having the same coordinate
  - Line: To understand the distance between any location and Launch Sites
- Explain why you added those objects
- GitHub URL of your completed interactive map with Folium map :

https://github.com/nym2010/Applied-Data-Science-Capstone/blob/2479d34100a407b3e1ca82fa04dd39822c548b4b/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

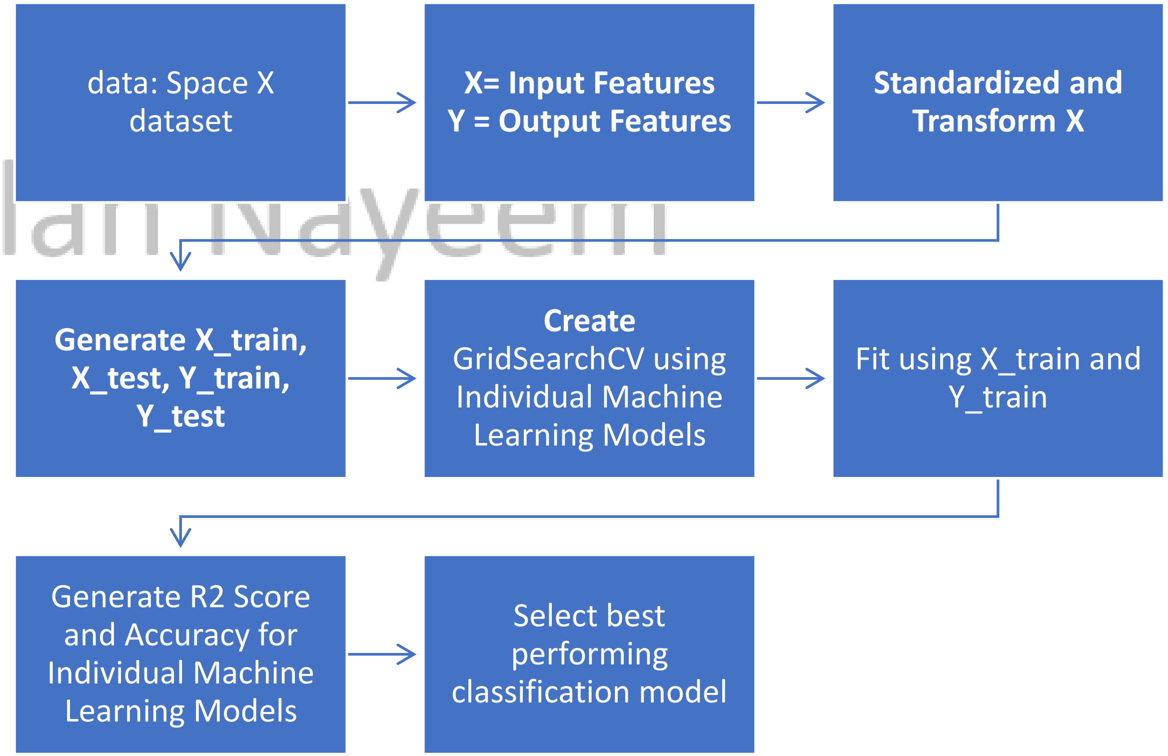
- Here we have created plots and interactions in the Dashboard:
  - Total Success Rates based for:
    - All Launch Sites
    - Any one of the Launch Sites
  - Correlation Between Payload and Success for:
    - All Launch Sites
    - Any one of the Launch Sites

Build a Dashboard with Plotly Dash

- GitHub URL of our completed Plotly Dash lab:
  - https://github.com/nym2010/Applied-Data-Science-Capstone/blob/main/All_Sites_Launch_Dashboard_1.png
  - https://github.com/nym2010/Applied-Data-Science-Capstone/blob/main/All_Sites_Launch_Ranger_1.png
  - https://github.com/nym2010/Applied-Data-Science-Capstone/blob/main/Specific_Sites_Launch_Dashboard_1.png
  - https://github.com/nym2010/Applied-Data-Science-Capstone/blob/main/Specific_Sites_Launch_Ranger_1.png

Predictive Analysis (Classification)

- We have test using:
  - Logistic Regression
  - Support Vector machines
  - Decision Tree Classifier
  - K-nearest neighbors
- We have used R2 Score and Accuracy to find the best performing classification model.
- GitHub URL of completed predictive analysis lab: https://github.com/nym2010/Applied-Data-Science-Capstone/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

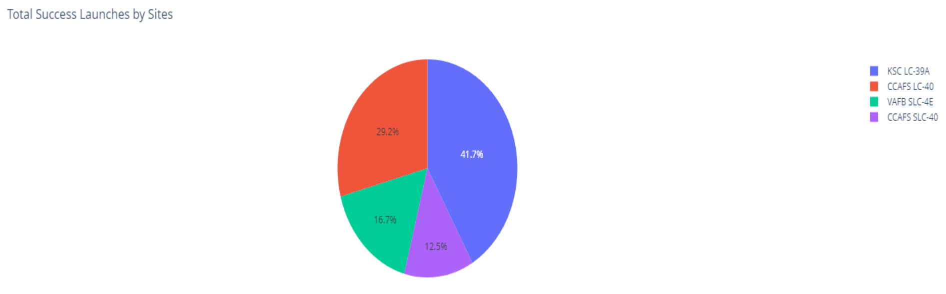


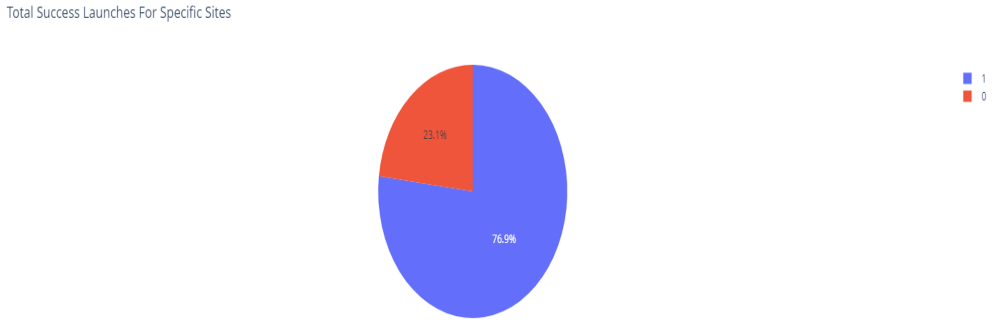
Results

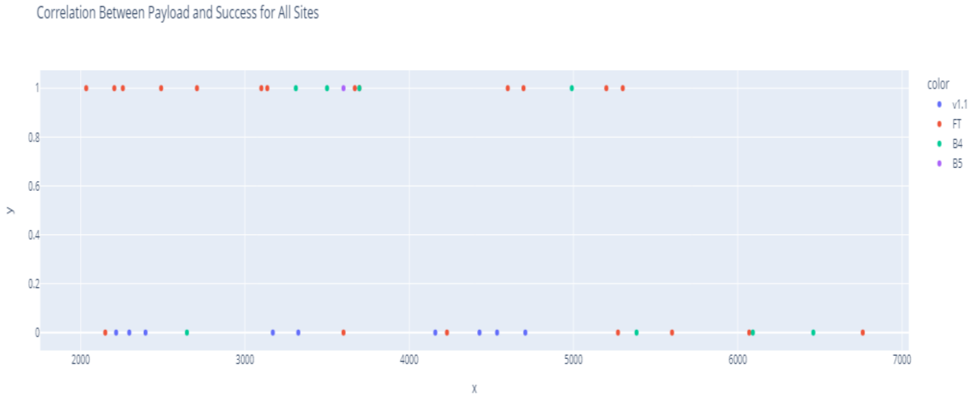
- Exploratory data analysis results
  - When Flight number increases, the first stage is more likely to land successfully.
  - While more massive the payload, the less likely the first stage will return.
  - ES-L1, GEO, HEO and SSO have higher success rate.
  - For VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).
  - LEO orbit the Success appears related to the number of flights but no relationship between flight number when in GTO orbit.
  - With heavy payloads the successful landing rate are more for Polar, LEO and ISS., but not for GTO.
  - Success rate since 2013 kept increasing till 2017 (stable in 2014) and after 2015 it started increasing.

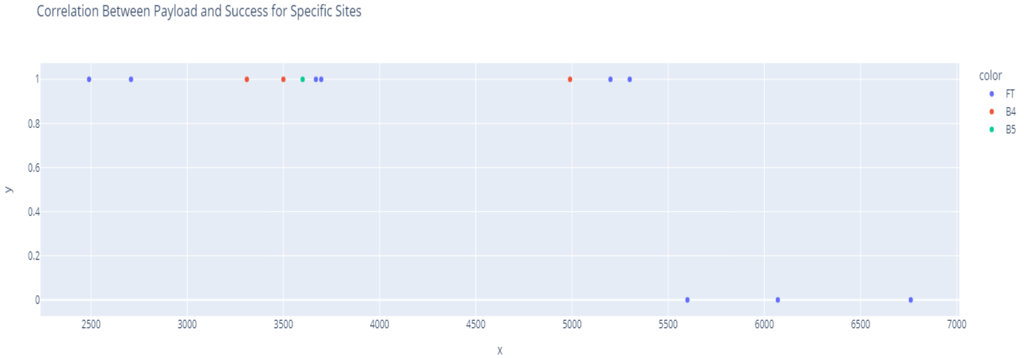
Results

- Interactive analytics demo in screenshots









Results

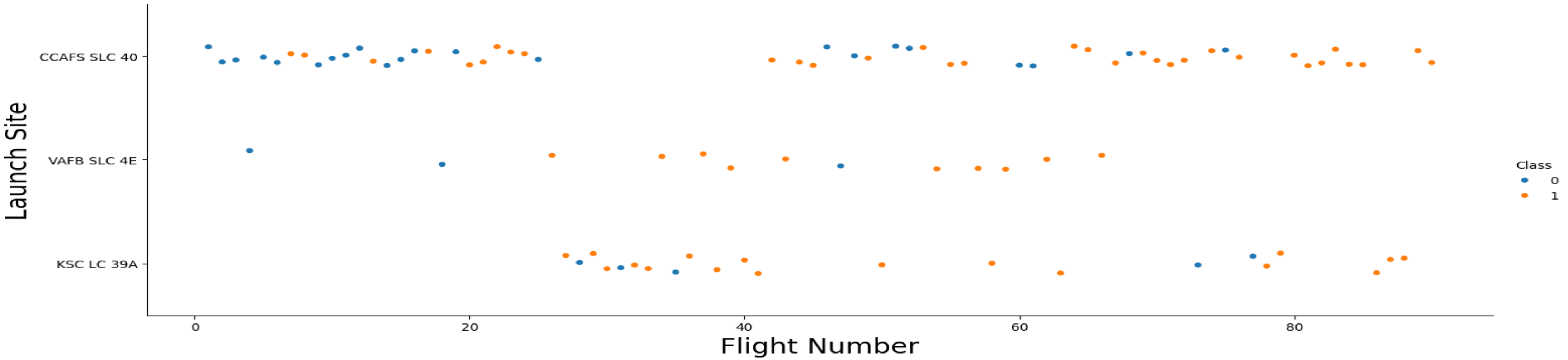
- Predictive analysis results

Method	R-Square	Accuracy
Decision_Tree	0.944444	0.875
KNN	0.833333	0.848214
Support Vector	0.833333	0.848214
Logistic	0.833333	0.846429

Section 2

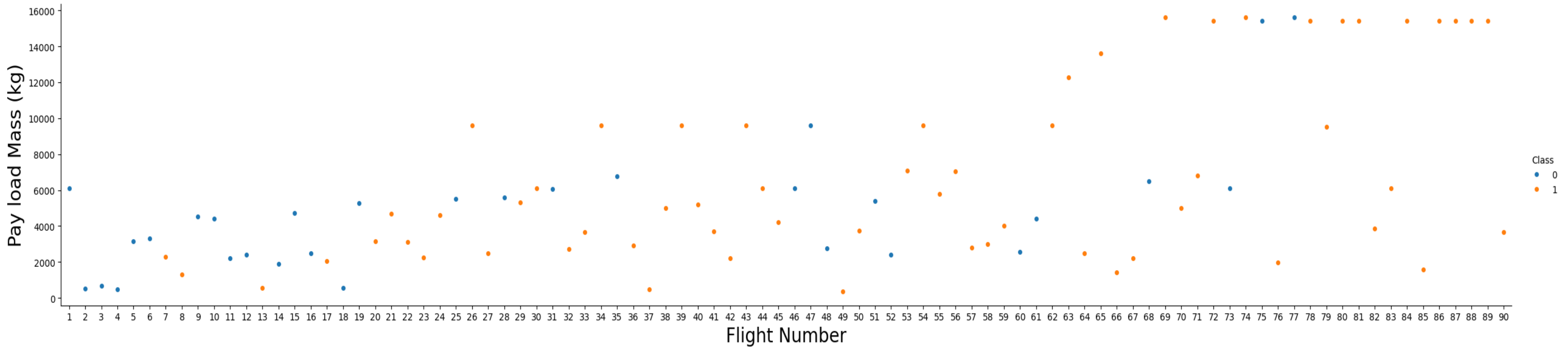
Insights drawn from EDA

Flight Number vs. Launch Site



- When Flight number increases, the first stage is more likely to land successfully.

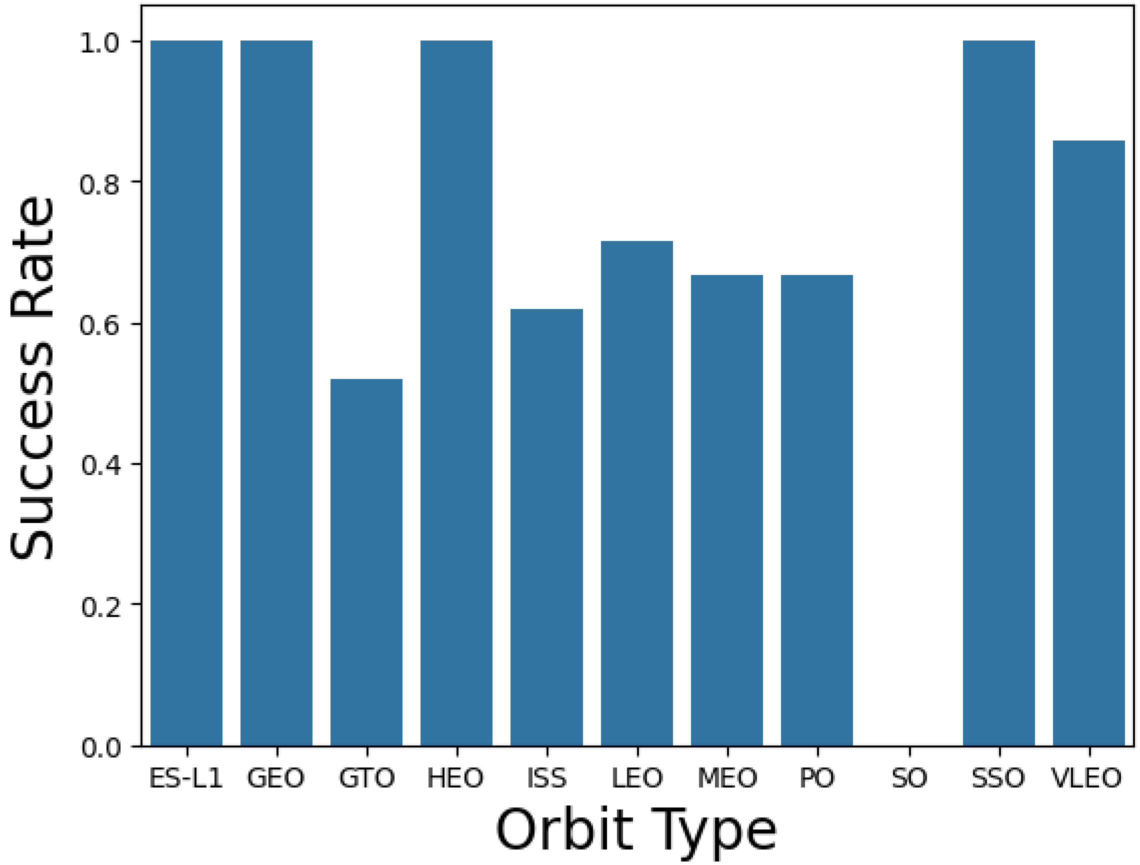
Payload vs. Launch Site



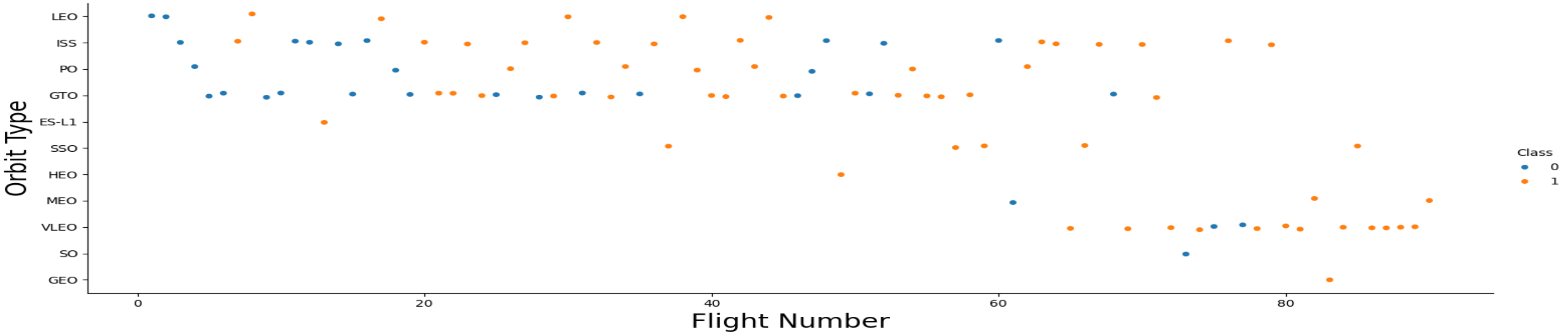
- While more massive the payload, the less likely the first stage will return.

Success Rate vs. Orbit Type

- ES-L1, GEO, HEO and SSO have higher success rate.

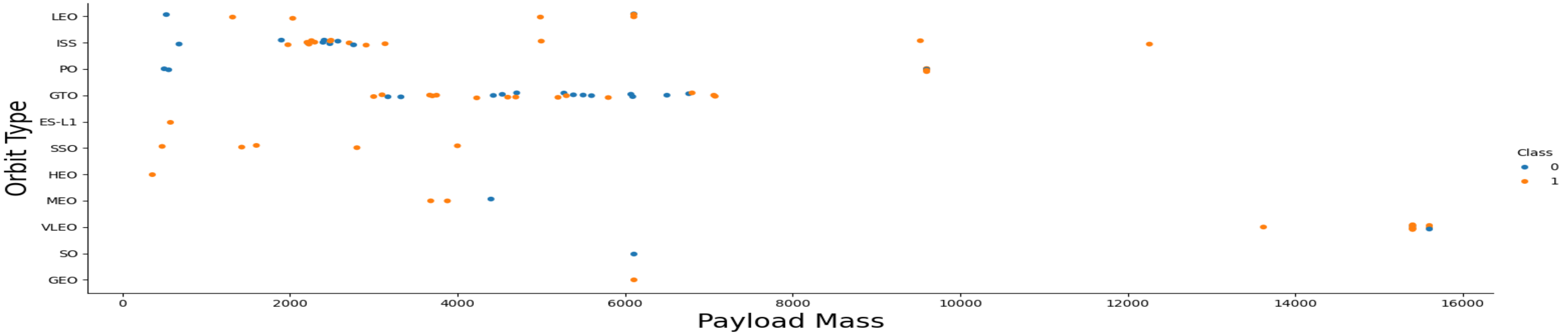


Flight Number vs. Orbit Type



- LEO orbit the Success appears related to the number of flights but no relationship between flight number when in GTO orbit.

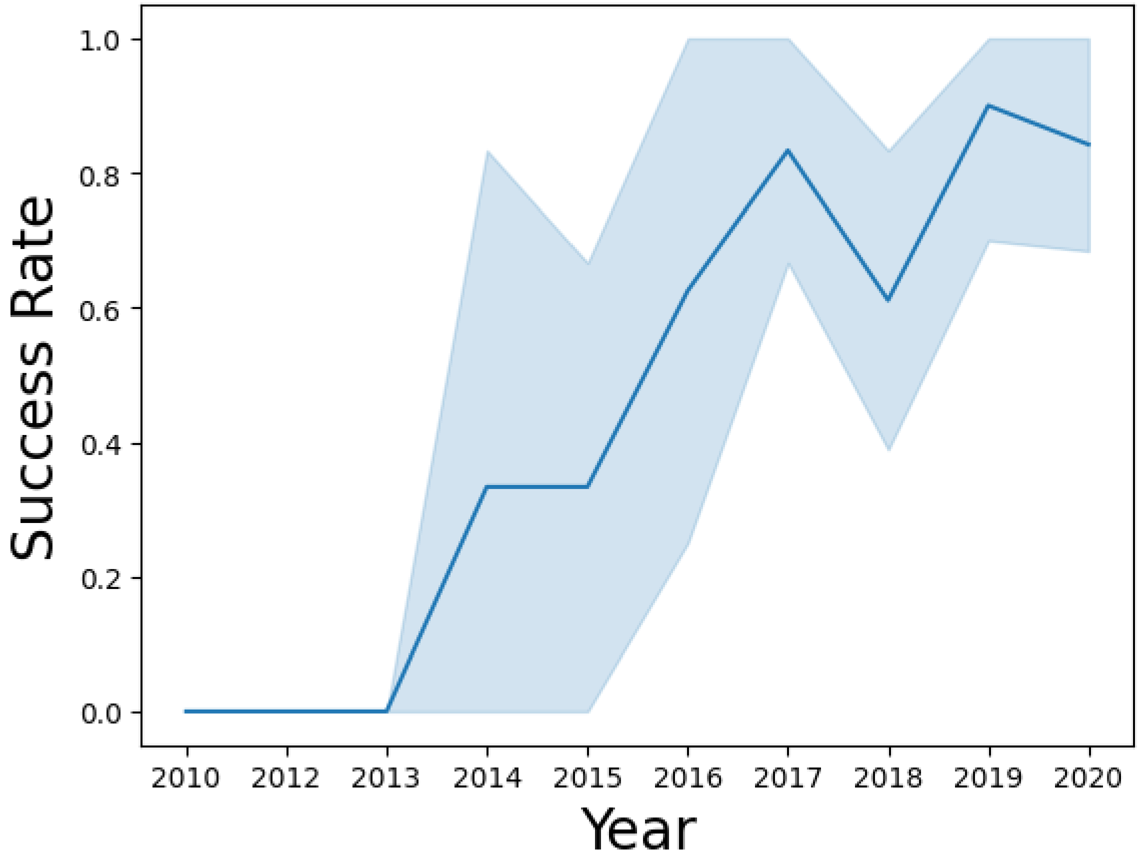
Payload vs. Orbit Type



- With heavy payloads the successful landing rate are more for Polar, LEO and ISS., but not for GTO.

Launch Success Yearly Trend

- Success rate since 2013 kept increasing till 2017 (stable in 2014) and after 2015 it started increasing.



All Launch Site Names

- All Launch Sites are:

Launch Site	Lat	Long
CCAFS LC-40	28.562302	-80.577356
CCAFS SLC-40	28.563197	-80.57682
KSC LC-39A	28.573255	-80.646895
VAFB SLC-4E	34.632834	-120.610745

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with `CCA`

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
0:00:00	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0:00:00	LEO	SpaceX	Success	Failure (parachute)
0:00:00	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0:00:00	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
0:00:00	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	0:00:00	LEO (ISS)	NASA (COTS)	Success	No attempt
0:00:00	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	0:00:00	LEO (ISS)	NASA (CRS)	Success	No attempt
0:00:00	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	0:00:00	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- Total payload carried by boosters from NASA

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1

SUM(PAYLOAD_MASS__KG_)
2534.666667

First Successful Ground Landing Date

- Date of the first successful landing outcome on ground pad

MIN(Date)
12/22/2015

Successful Drone Ship Landing with Payload between 4000 and 6000

- Names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes

Mission_Outcome	COUNT(*)
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- Booster which have carried the maximum payload mass

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

- Failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
6/4/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
12/8/2010	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)
5/22/2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
10/8/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
3/1/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

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

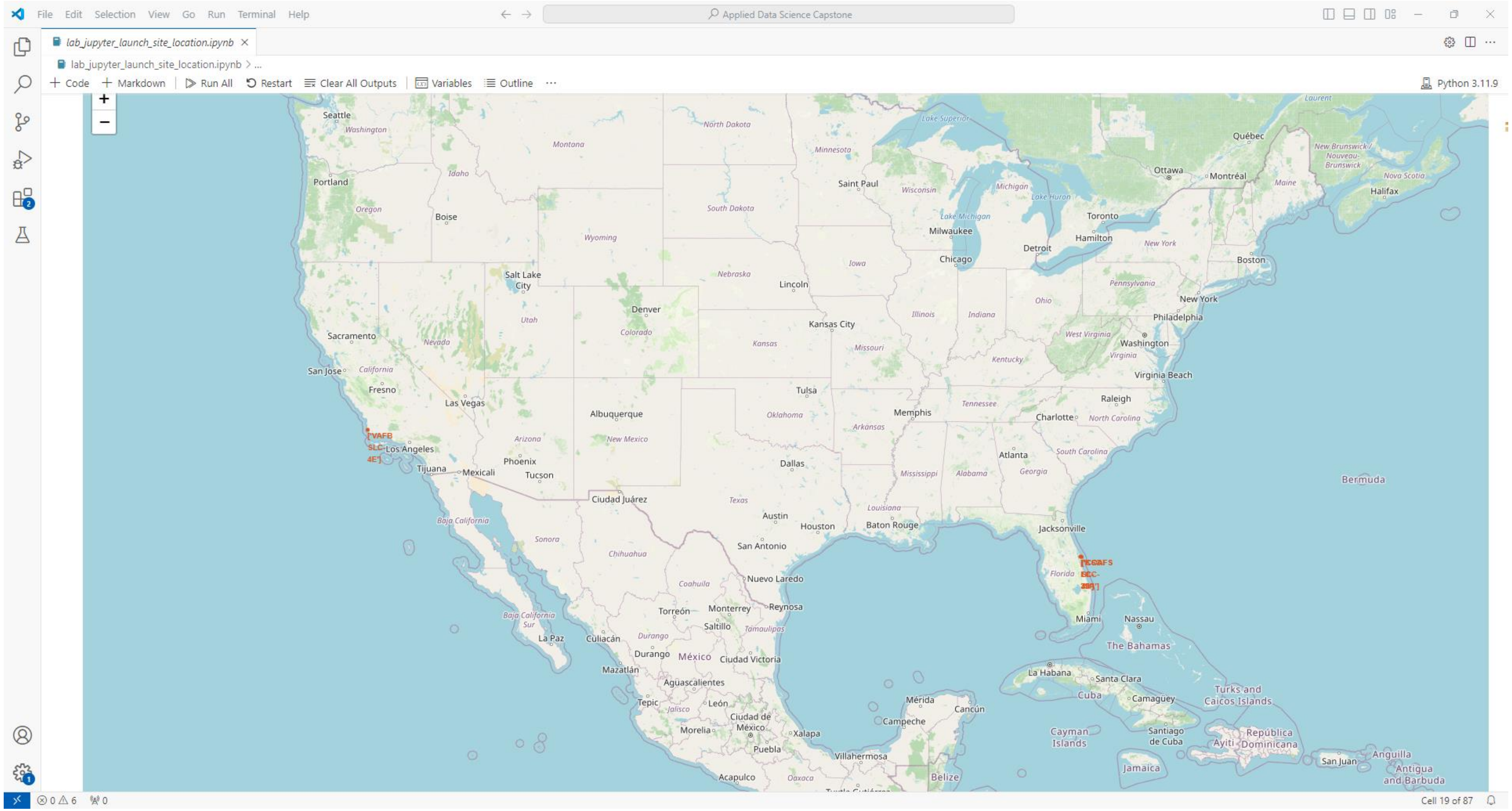
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

Landing_Outcome	COUNT(*)
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

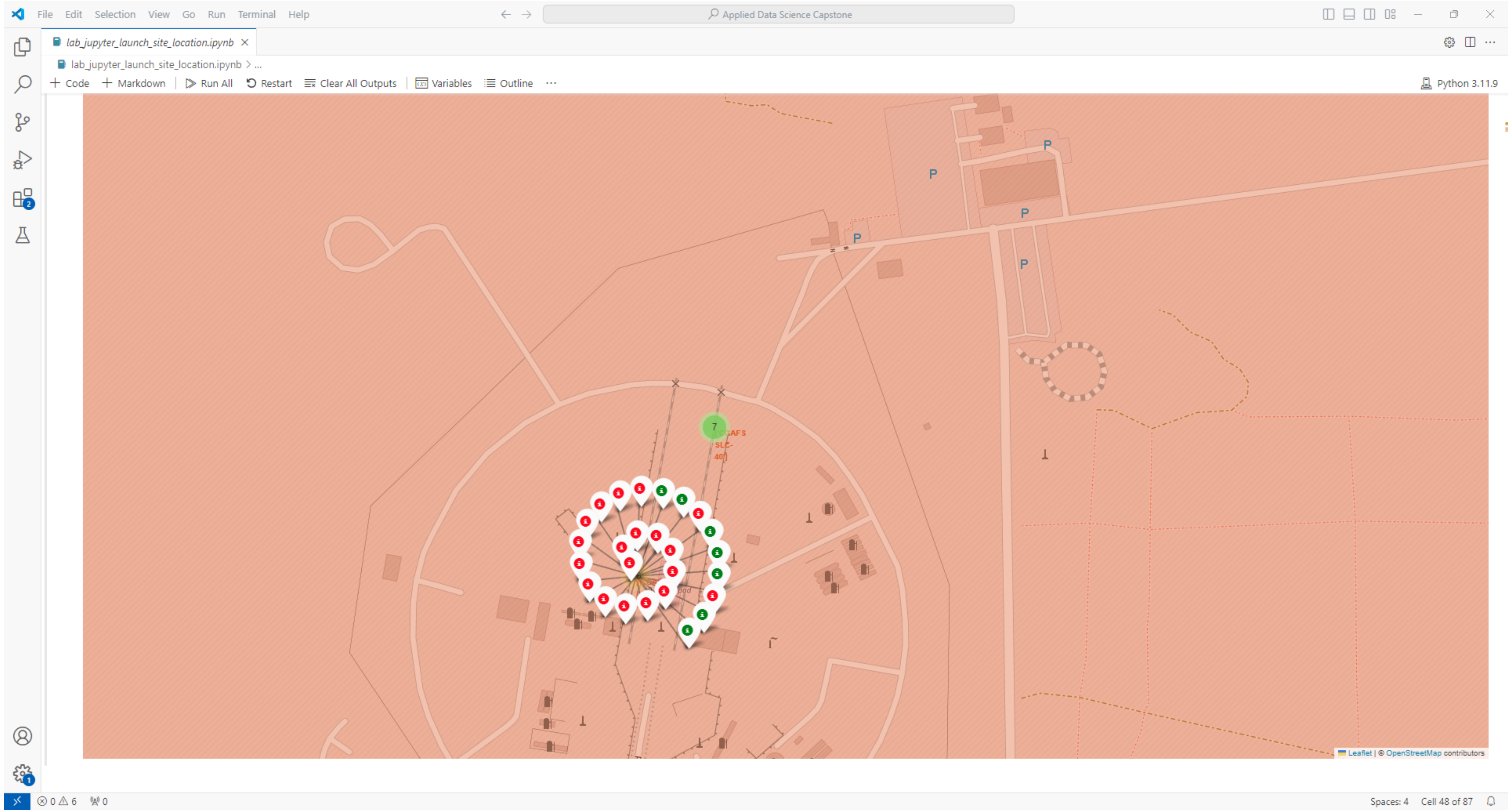
Section 3

Launch Sites Proximities Analysis

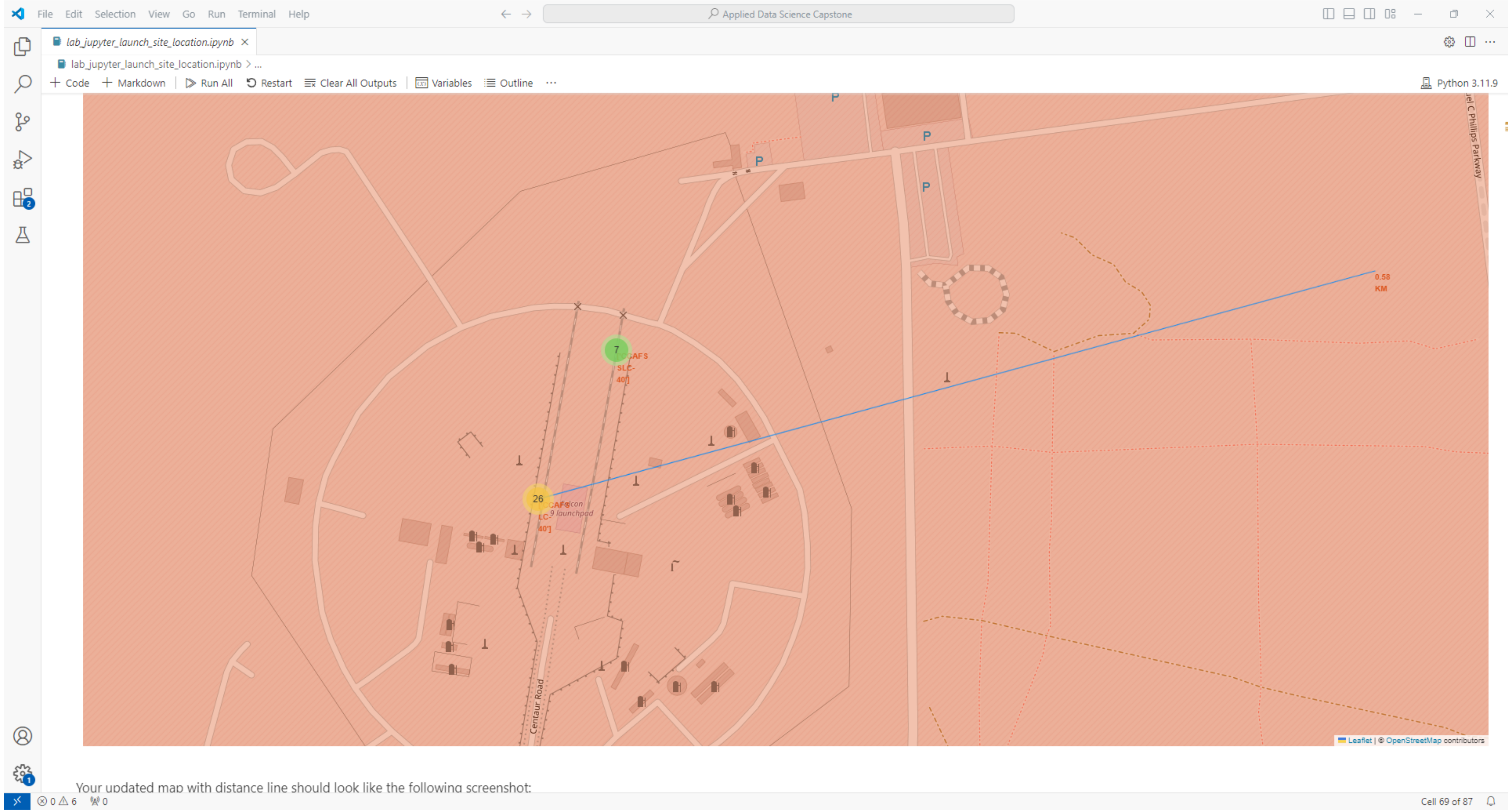
All Launch Sites' Location Markers



Color-labeled Launch Outcomes on the Map



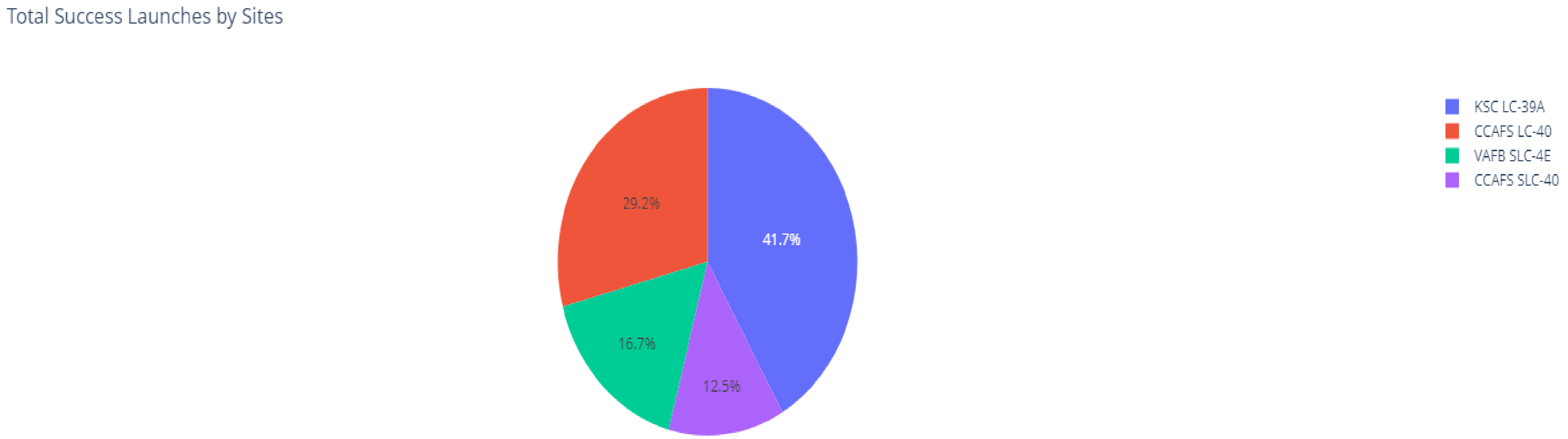
Selected launch site to its proximities with distance calculated and displayed



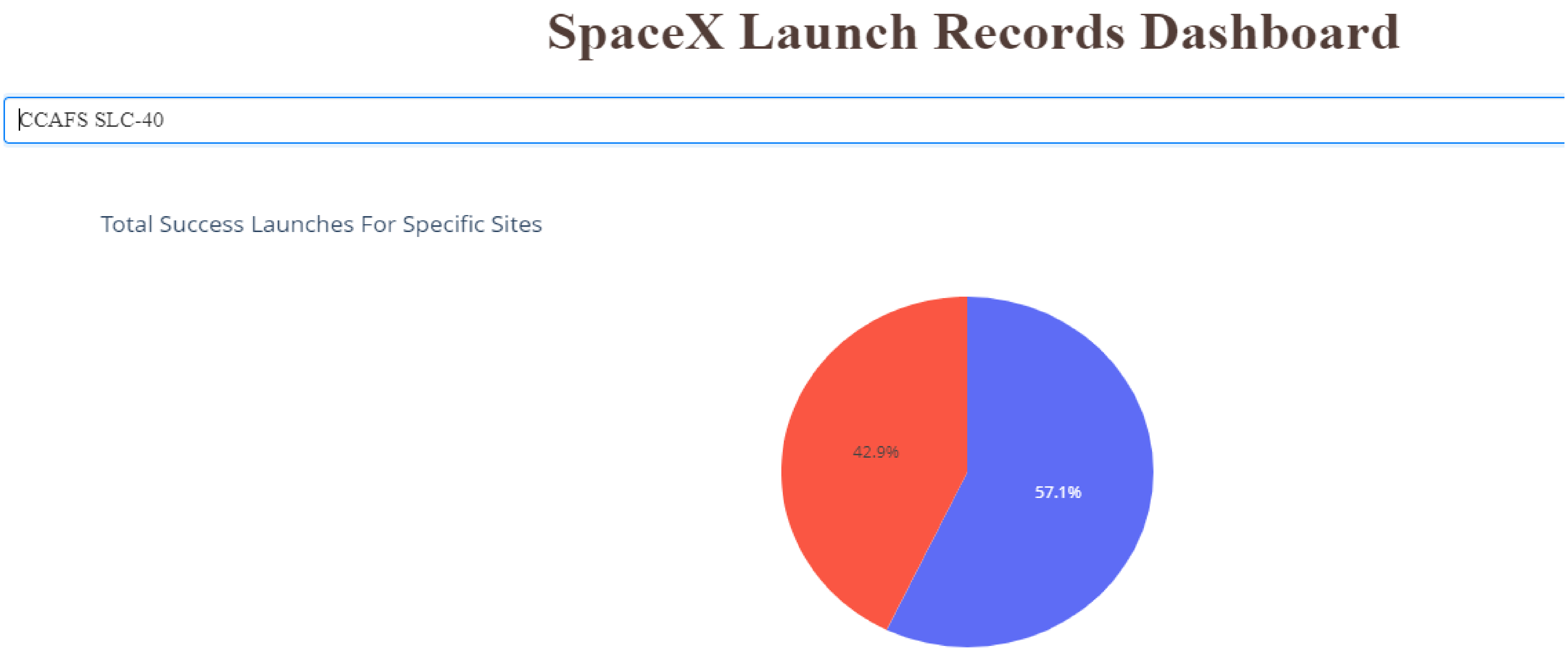
Section 4

Build a Dashboard with Plotly Dash

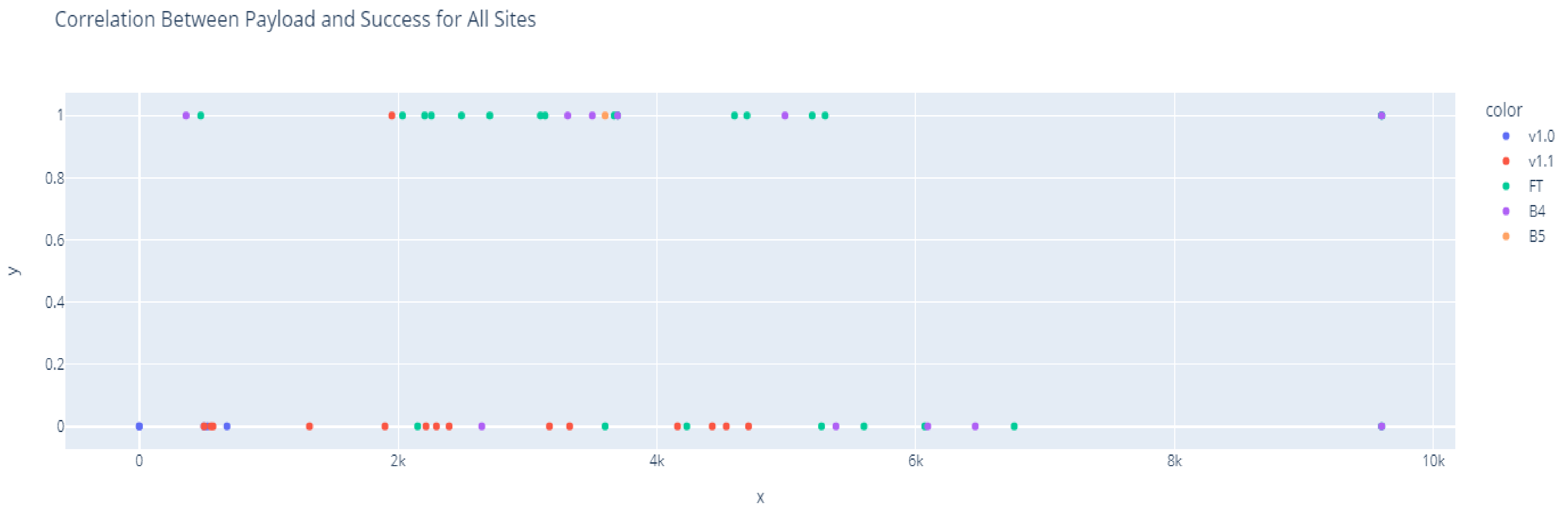
Launch Success Count for all Sites



Launch Site with Highest Launch Success Ratio



Payload vs. Launch Outcome scatter plot for all sites



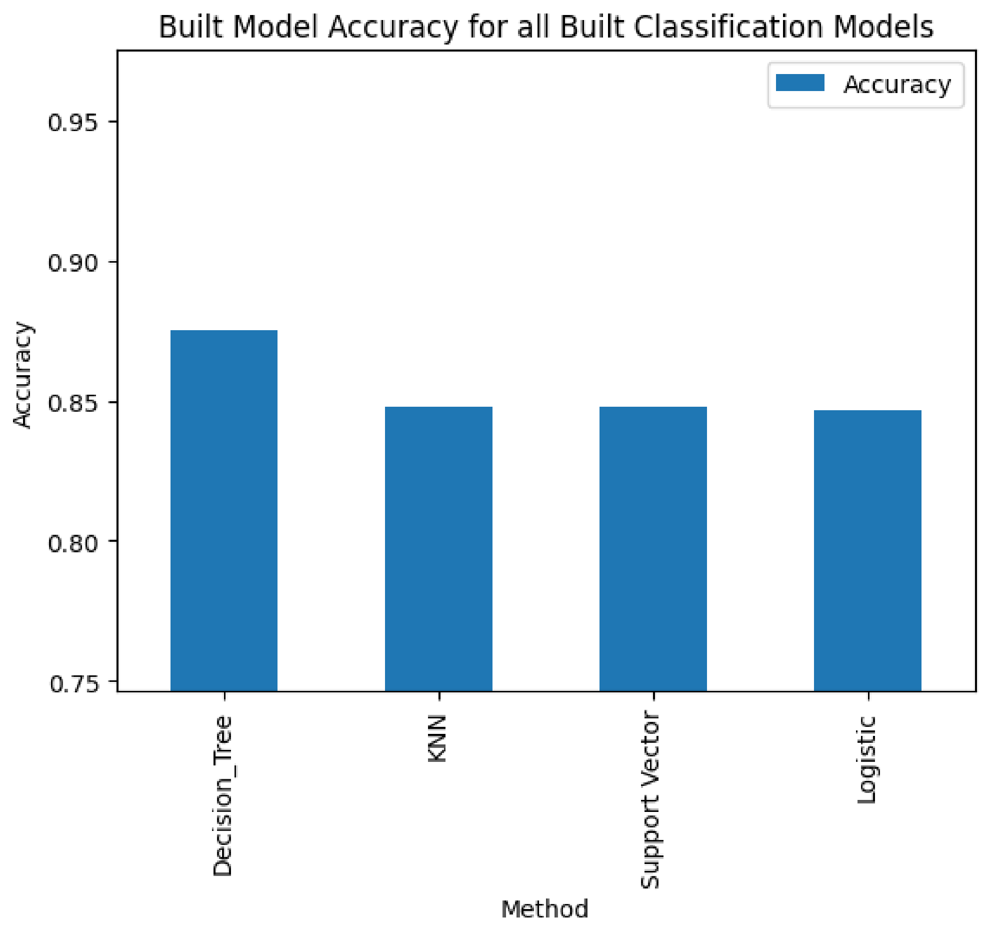
Section 5

Predictive Analysis (Classification)

Classification Accuracy

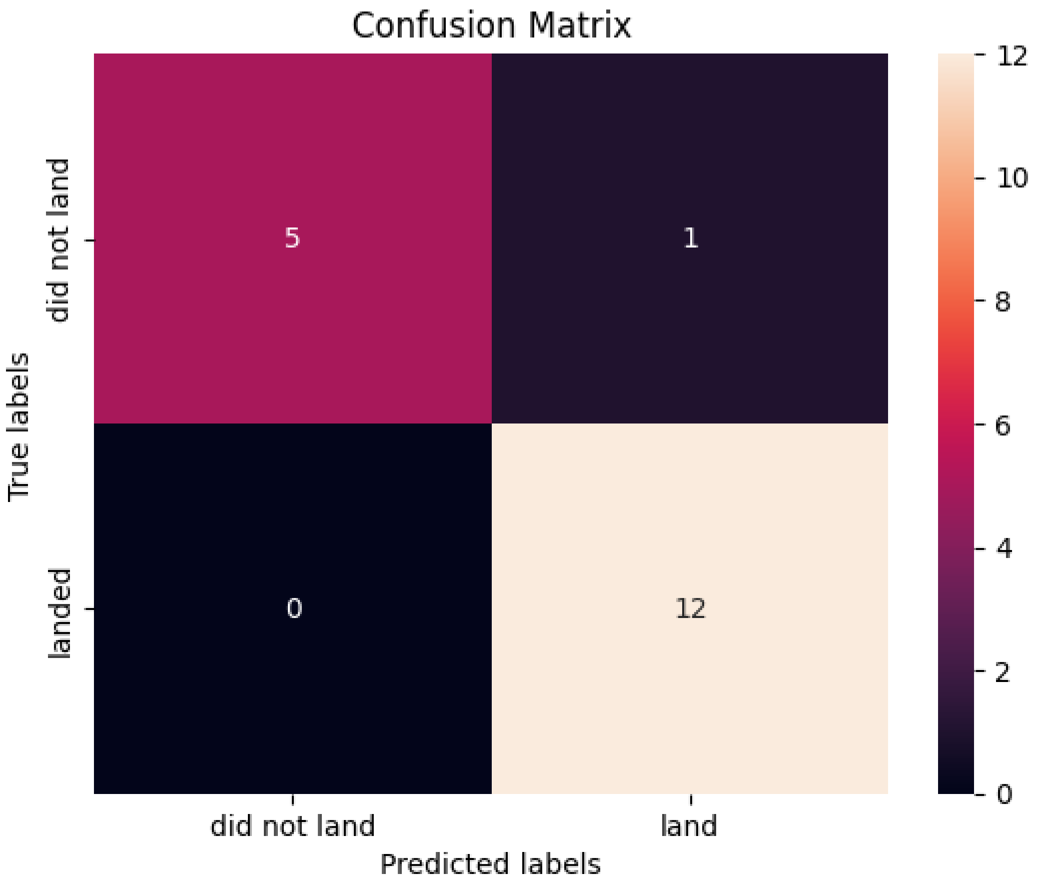
- Built Model Accuracy for all Built Classification Models

- Decision Tree has the highest Accuracy



Confusion Matrix

- Confusion matrix of the best performing model : Decision Tree
- As shown, the False Positive is the lowest, we can say Decision Tree is the best performing method



Conclusions

- When Flight number increases, the first stage is more likely to land successfully.
- While more massive the payload, the less likely the first stage will return.
- ES-L1, GEO, HEO and SSO have higher success rate.
- For VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).
- LEO orbit the Success appears related to the number of flights but no relationship between flight number when in GTO orbit.
- With heavy payloads the successful landing rate are more for Polar, LEO and ISS., but not for GTO.
- Success rate since 2013 kept increasing till 2017 (stable in 2014) and after 2015 it started increasing.

Thank you!