IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Uhenna Nwajideobi
09/10/2023

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection through API
  - Data Collection with Web Scraping
  - Data Wrangling using Pandas
  - Exploratory Data Analysis with SQL
  - Exploratory Data Analysis with Data Visualization
  - Interactive Visual Analysis

- Summary of all results
  - Exploratory Data Analysis result
  - Interactive Analytics (visualization fig)
  - Predictive Analytics result

Introduction

- Project background and context
  - Space X 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 Space X can reuse the first stage. Therefore if we can determine if the first stage will land, we can determine and control the cost of a launch. This information can be used if an alternate company wants to bid against space X for a rocket launch. We wish to create a machine learning pipeline to predict if the first stage will land given the data from space X api.
- Problems you want to find answers
  - What factors determines if the first stage will land successfully.
  - Interactions of the acquired data in determining successful landing.
  - Can we predict a successful landing

Section 1

Methodology

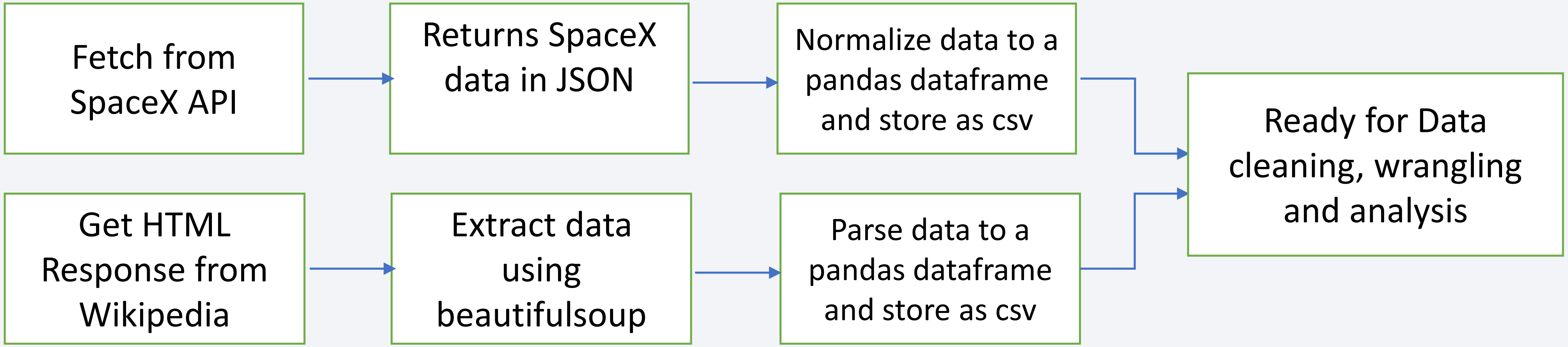
Methodology

Executive Summary

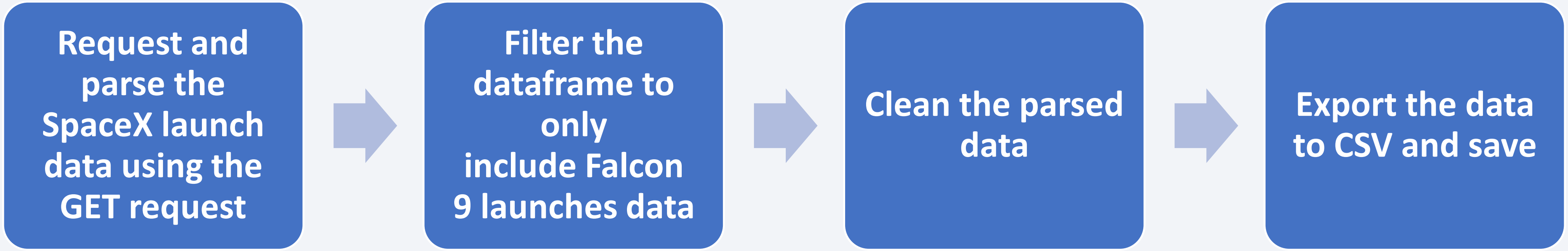
- Data collection methodology:
  - Data was collected by calling and querying Space X api, and web-scraping wikipedia pages.
- Perform data wrangling
  - Cleaning, encoding, presenting and storing acquired data using python libs
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- Description of how data sets were collected.
- Data was collected by making request to Space X API and web scrapping of Wikipedia pages using request libs and BeautifulSoap module.
- You need to present your data collection process use key phrases and flowcharts



Data Collection – SpaceX API



GitHub URL of the completed SpaceX API calls:

https://github.com/uchemangajs/Space-X_Falcon9_First_Stage_Landing_Prediction/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

Request the Falcon9 Launch Wiki page from its URL → Extract Data from response using BeautifulSoup Module → **Create a data frame by parsing the launch HTML tables** → Export the data frame to CSV and save

- Github url: https://github.com/uchemangajs/Space-X_Falcon9_First_Stage_Landing_Prediction/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

Data Examination: check for null values, data types → **Calculate the number of launches on each site** → **Calculate the number and occurrence of each orbit** → **Create a landing outcome label from Outcome column** → Export data frame to CSV and save

- https://github.com/uchemangajs/Space-X_Falcon9_First_Stage_Landing_Prediction/blob/main/spacex-data_wrangling_jupyterlite.jupyterlite.ipynb

EDA with Data Visualization

- Summary of what charts were plotted and why we used those charts
  - The following chats were plotted:
  - Categorical Plot
  - Scatter plot
  - Bar plot
  - Line Plot

These plots were used to check the interactions between select variables

- GitHub URL: https://github.com/uchemangajs/Space-X_Falcon9_First_Stage_Landing_Prediction/commit/90b18bac728237980e010a17aafa8e87b579a1af

EDA with SQL

- Summary of the SQL queries you performed
  - Display the names of the unique launch sites in the space mission
  - Display the names of the unique launch sites in the space mission
  - Display the total payload mass carried by boosters launched by NASA (CRS)
  - List the date when the first succesful landing outcome in ground pad was acheived.
  - List the names of the boosters which have success in drone ship and have payload mass between 4000 and 6000
  - List the total number of successful and failure mission outcomes
  - List the names of the booster_versions which have carried the maximum payload mass. Use a subquery
  - List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
  - 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.

GitHub URL:https://github.com/uchemangajs/Space-X_Falcon9_First_Stage_Landing_Prediction/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Summary of map objects:
  - Markers: show information about an area of interest on a map
  - Cluster: Helps group markers together where multiple exist in close proximity.
  - Circle: Helps to highlight an area of interest.
  - Line: Helps to show distance between two locations.

The map objects were added to describe and highlight characteristics of launch locations.

- GitHub URL: https://github.com/uchemangajs/Space-X_Falcon9_First_Stage_Landing_Prediction/blob/main/Folio_launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

- Summary of what plots/graphs and interactions we have added to a dashboard
  - Pie Chat
  - Scatter plot

Pie chat shows ratio and proportions while scatter plot shows relations between select variables and outcomes.

- GitHub URL: https://github.com/uchemangajs/Space-X_Falcon9_First_Stage_Landing_Prediction/blob/main/spacex_dash_app.py

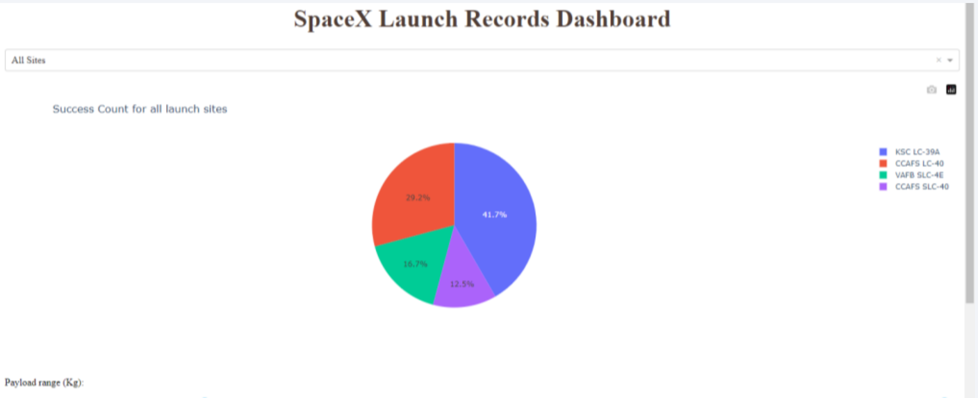
Predictive Analysis (Classification)

1. Create the dependent variable Y, standardize and transform X variables
2. Split X and Y data into training and test data
3. Build Model: LRM, SVM, Decision Tree and KNN
4. Evaluate the results of the various model using the method score and others
5. Build a table of results and choose the best performance

GitHub URL: https://github.com/uchemangajs/Space-X_Falcon9_First_Stage_Landing_Prediction/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

- Exploratory data analysis results
- Significant relationship was found between landing outcome and FlightNumber, PayloadMass, Orbit, LaunchSite, Flights, GridFins among others.
- Interactive analytics demo in screenshots

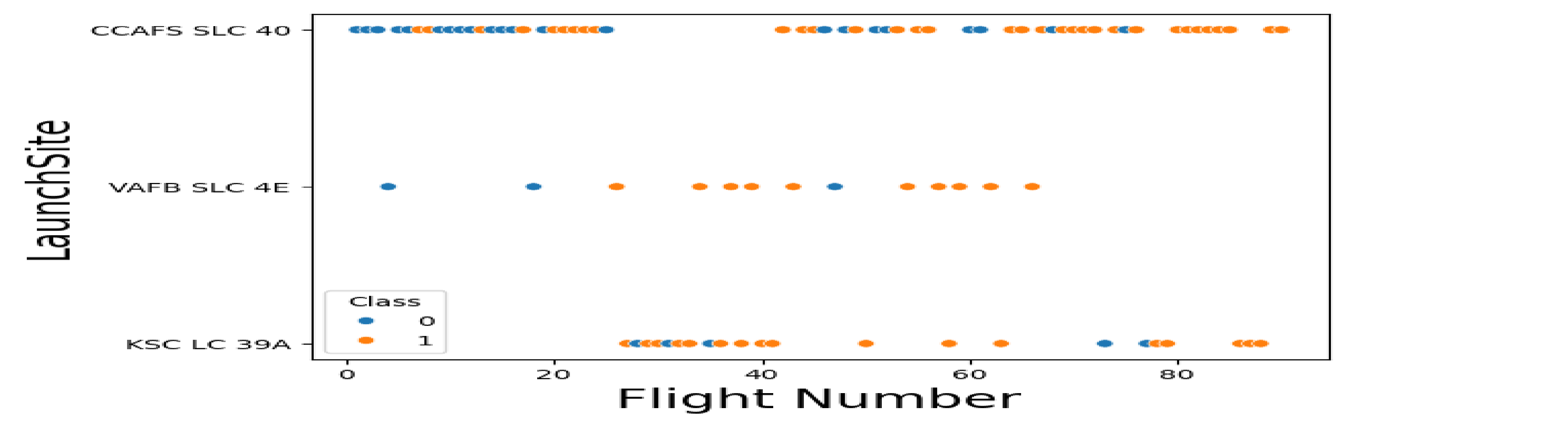


- Predictive analysis results
- Decision tree classifier has the highest accuracy score at 0.875

Section 2

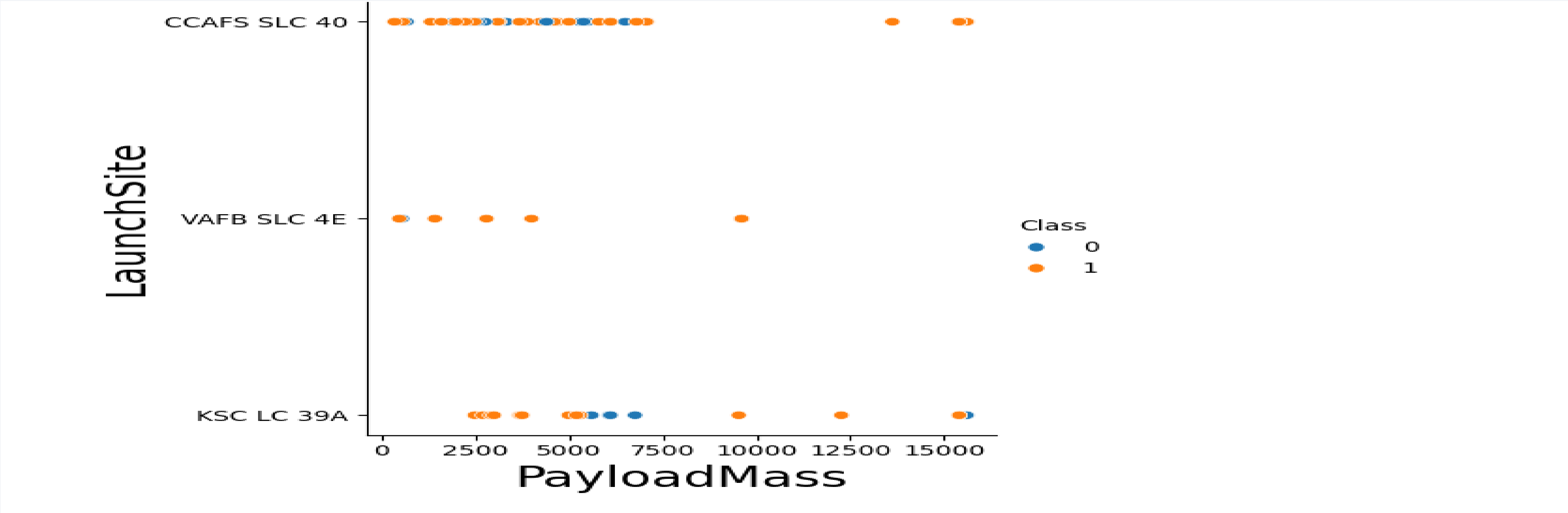
Insights drawn from EDA

Flight Number vs. Launch Site



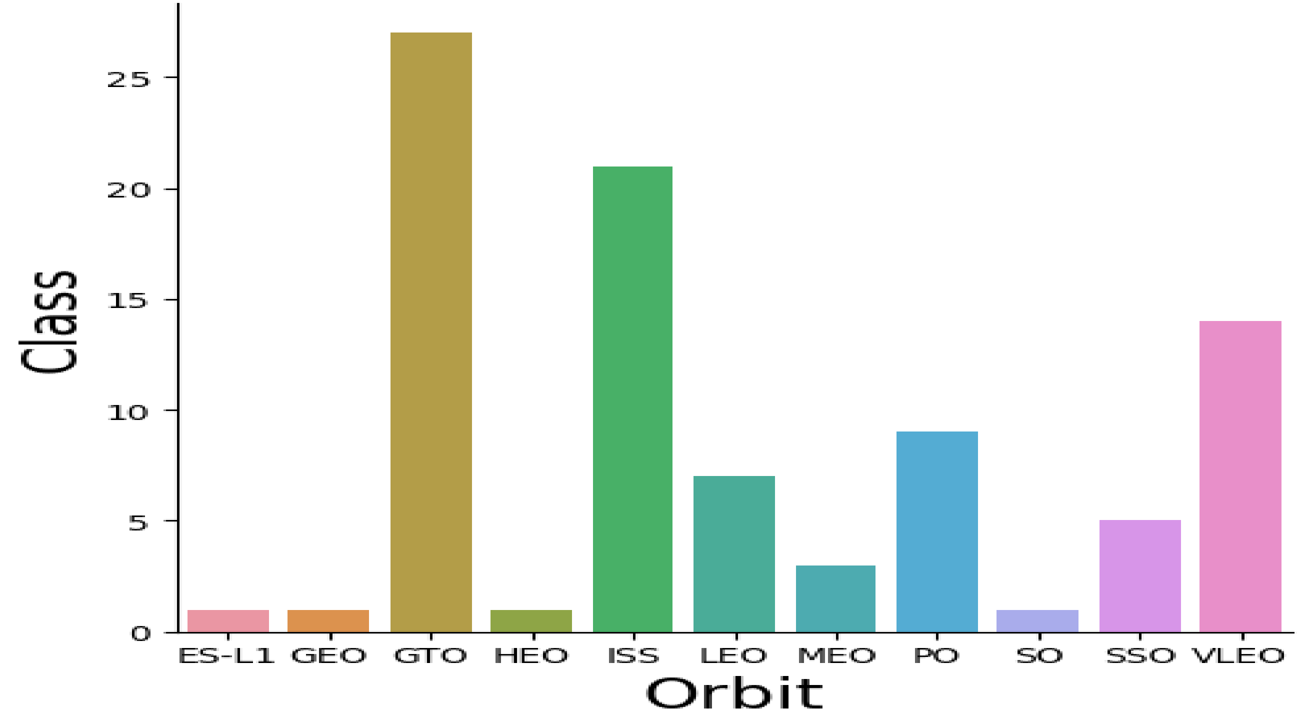
The scatter plot shows that at higher flight_number, more successful landing occured

Payload vs. Launch Site



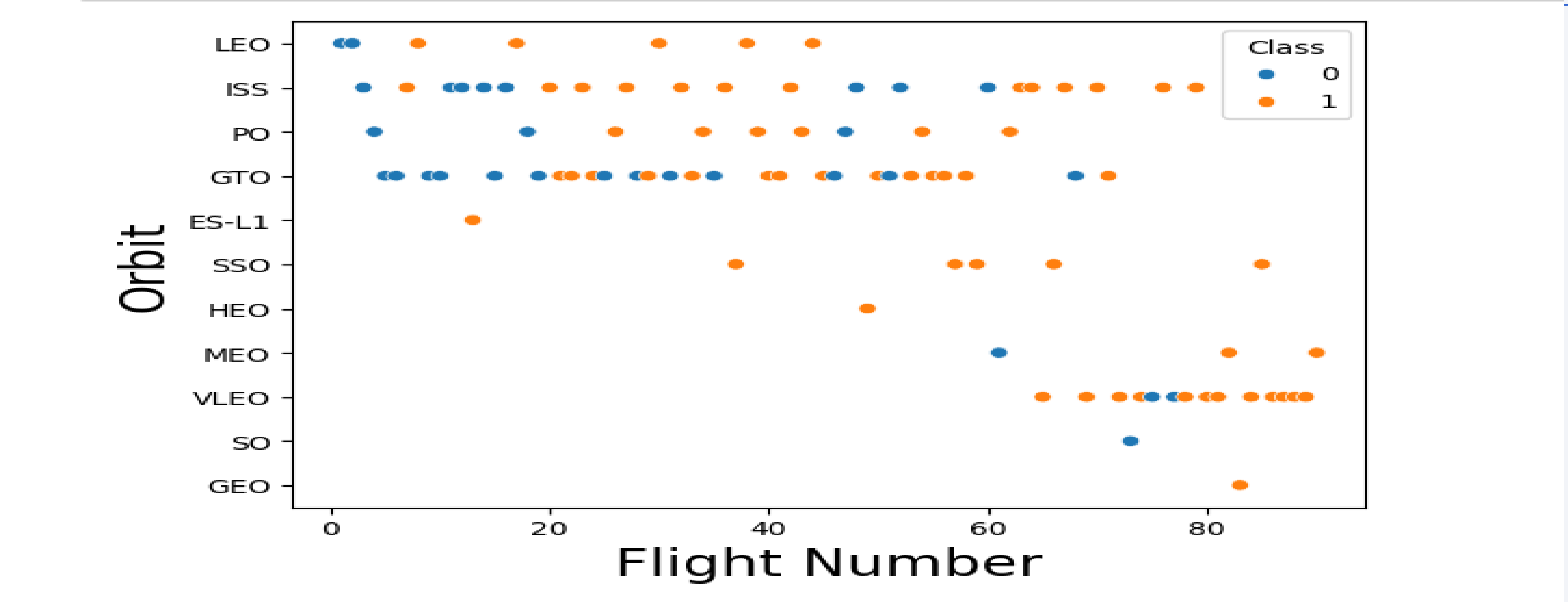
As the payload increases, success in landing of the first stage increases

Success Rate vs. Orbit Type



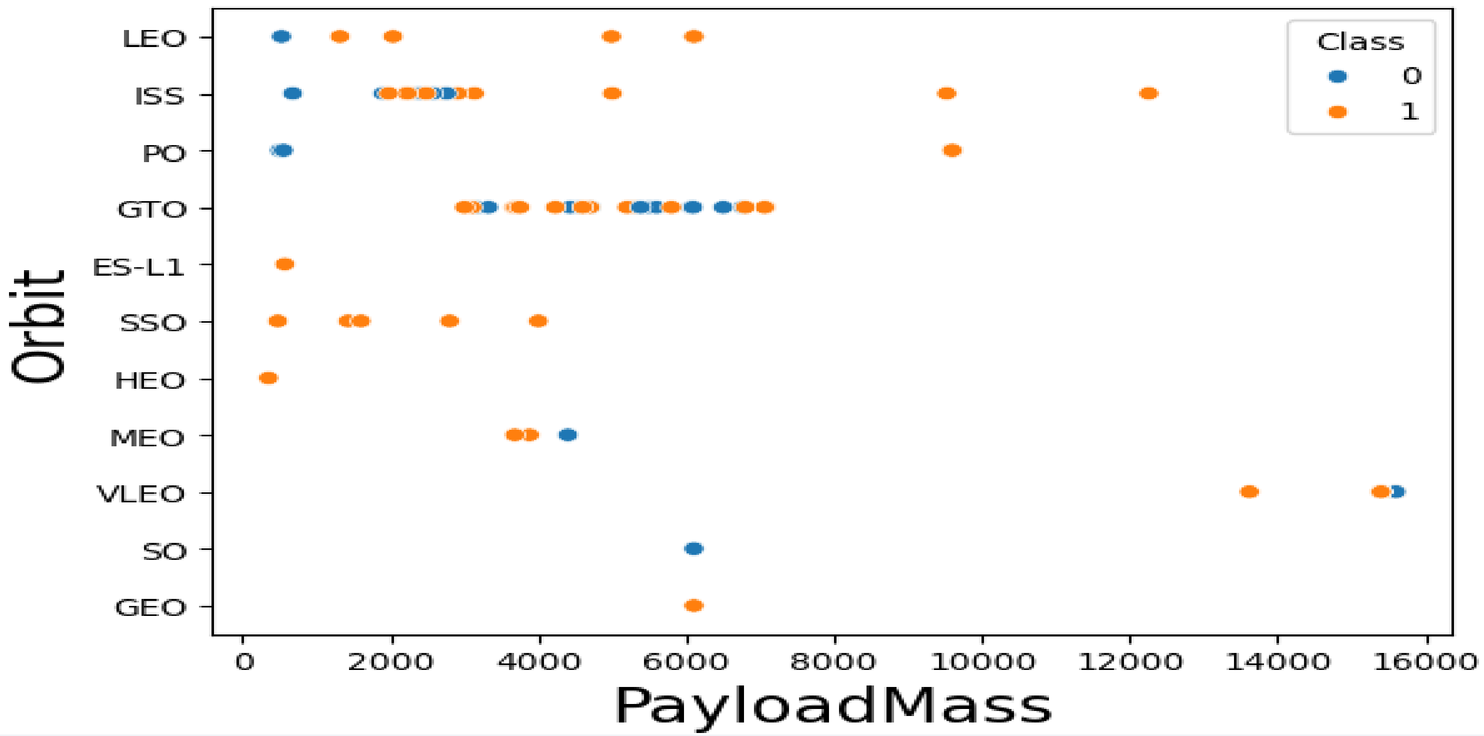
- GTO Orbit has the highest success rate

Flight Number vs. Orbit Type



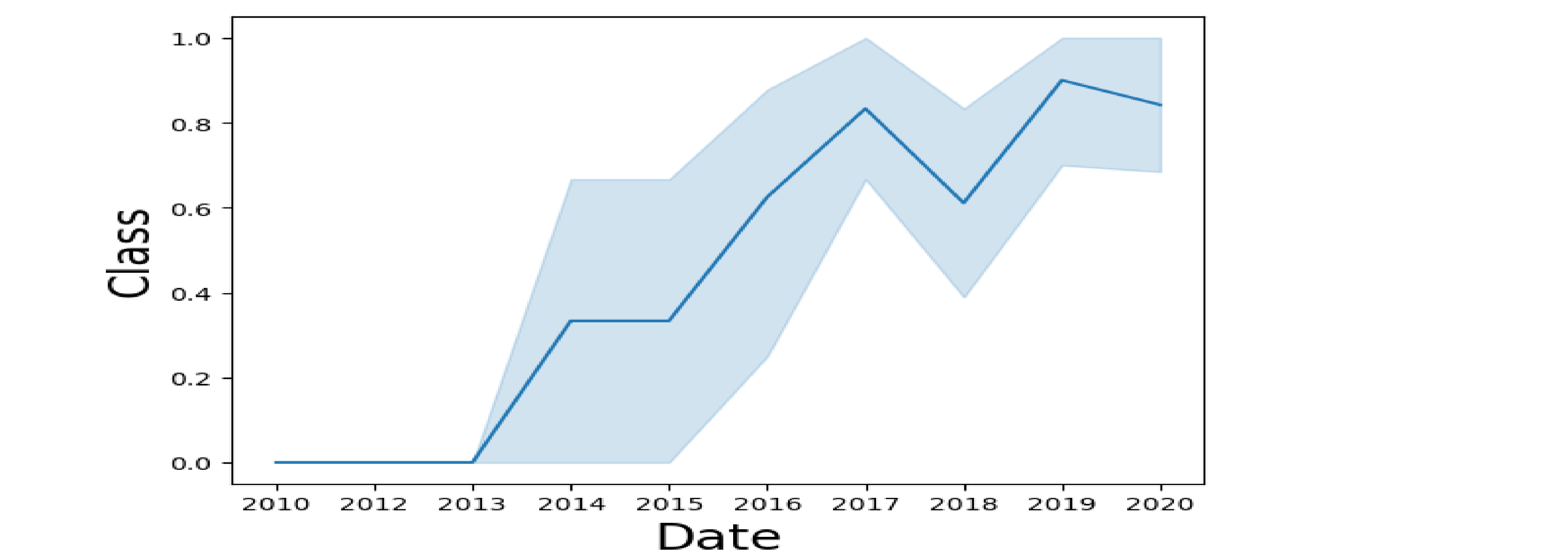
Apart from LEO orbit which shows positive relation with Flight Number, there seems not to be distinctive relation.

Payload vs. Orbit Type



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

Launch Success Yearly Trend



We can observe that the success rate since 2013 kept increasing till 2020

All Launch Site Names

Launch Site Names
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

- This are the launch sites names from which the different launches were made

Launch Site Names Begin with 'CCA'

Display 5 records where launch sites begin with the string 'CCA'

In [11]:
```
%sql SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE 'CCA%' LIMIT 5
```

```
 * sqlite:///my_data1.db
Done.
```

Out[11]:

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

5 records where launch sites begin with `CCA`

- The ‘LIKE %’ statement was used here

Total Payload Mass

- Total payload carried by boosters from NASA

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

- Date of the first successful landing outcome on ground pad

MIN(DATE)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

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

Booster_Version	Payload
F9 FT B1022	JCSAT-14
F9 FT B1026	JCSAT-16
F9 FT B1021.2	SES-10
F9 FT B1031.2	SES-11 / EchoStar 105

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes

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

Boosters Carried Maximum Payload

- Names of the 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

- List of the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

Date	Booster_Version	Launch_Site	Landing_Outcome
2015-10-01	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

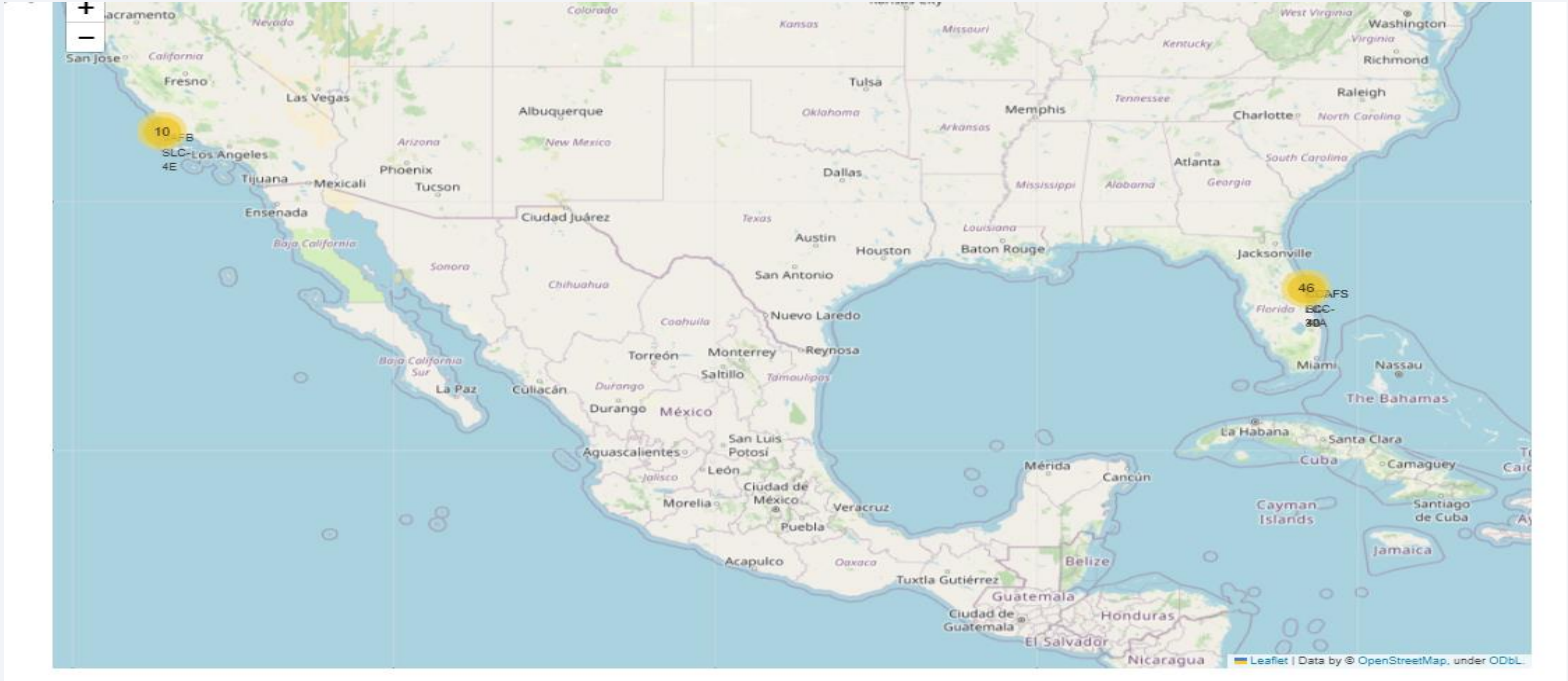
- Rank of count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

Landing_Outcome	count_outcomes	Date
No attempt	10	2012-05-22
Success (ground pad)	5	2015-12-22
Success (drone ship)	5	2016-08-04
Failure (drone ship)	5	2015-10-01
Controlled (ocean)	3	2014-04-18
Uncontrolled (ocean)	2	2013-09-29
Precluded (drone ship)	1	2015-06-28
Failure (parachute)	1	2010-08-12

Section 3

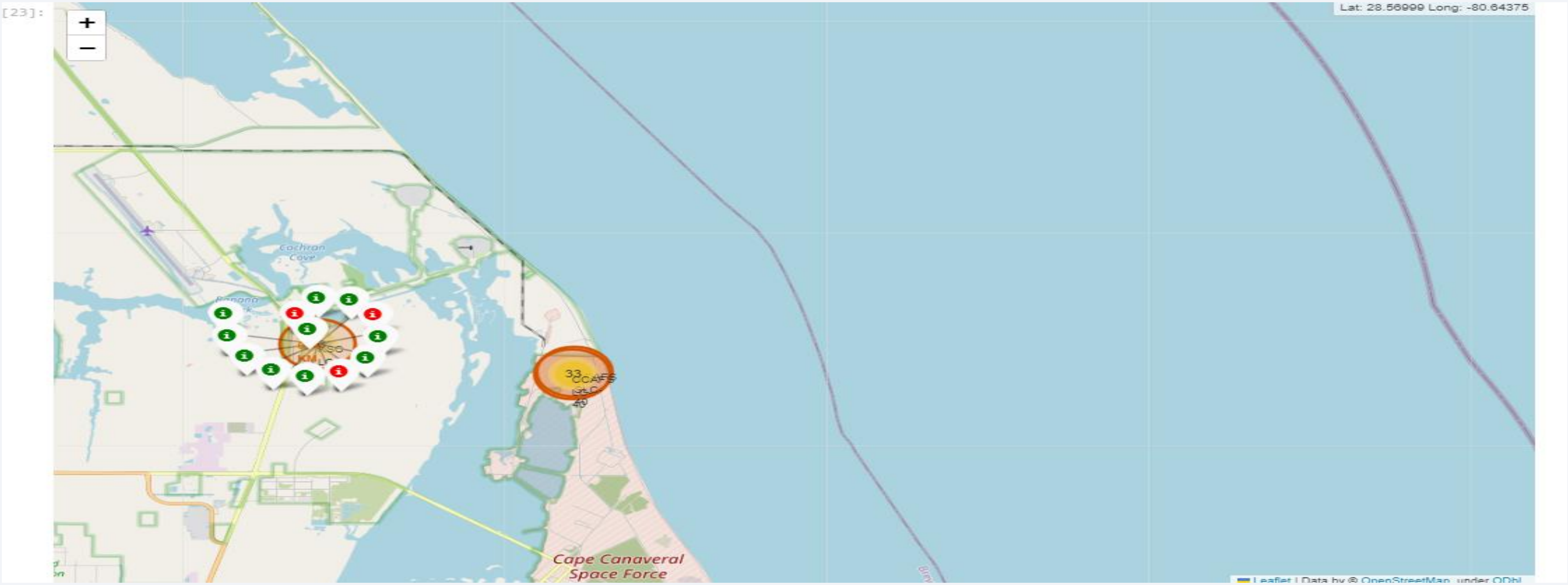
Launch Sites Proximities Analysis

ALL LAUCH SITE LOCATION MARKERS



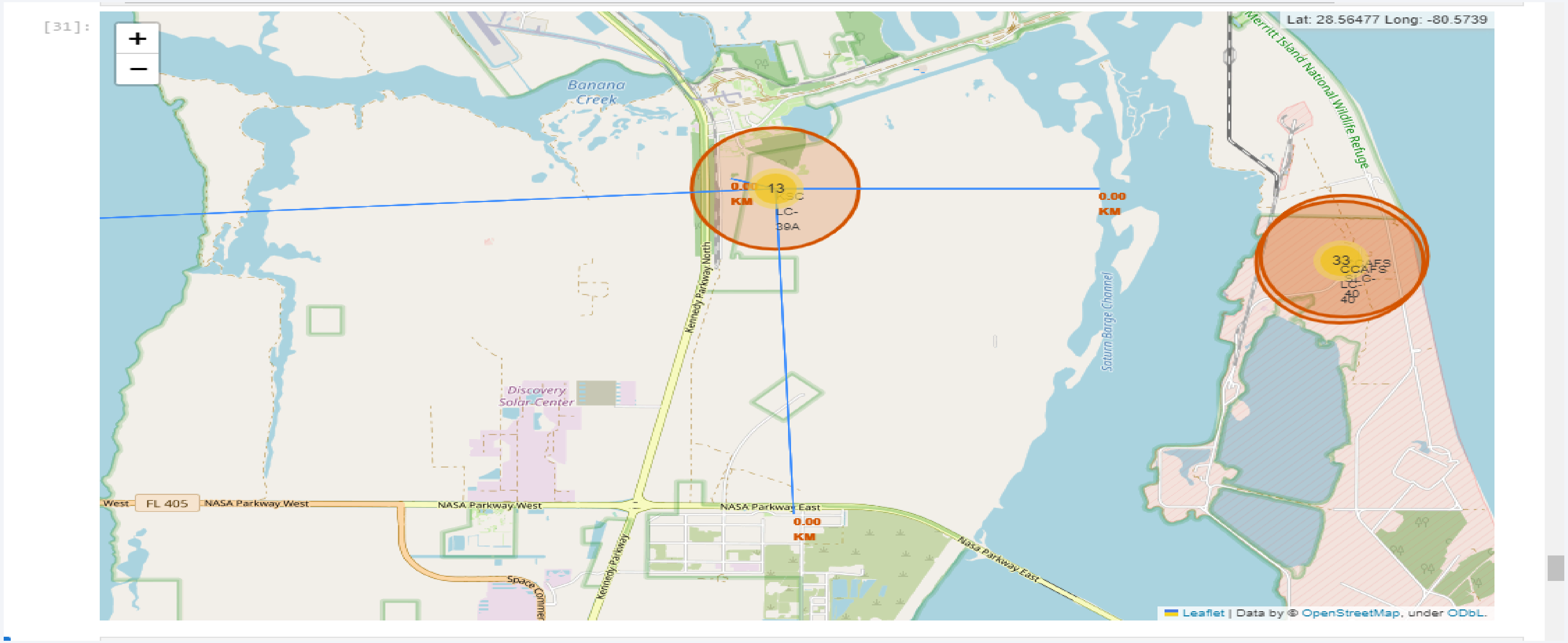
- The launch sites are depicted by a yellow marker on the east and of the map.

Launch Outcomes on the Map



- Green colors represent success and red represent failure

<Folium Map Screenshot 3>

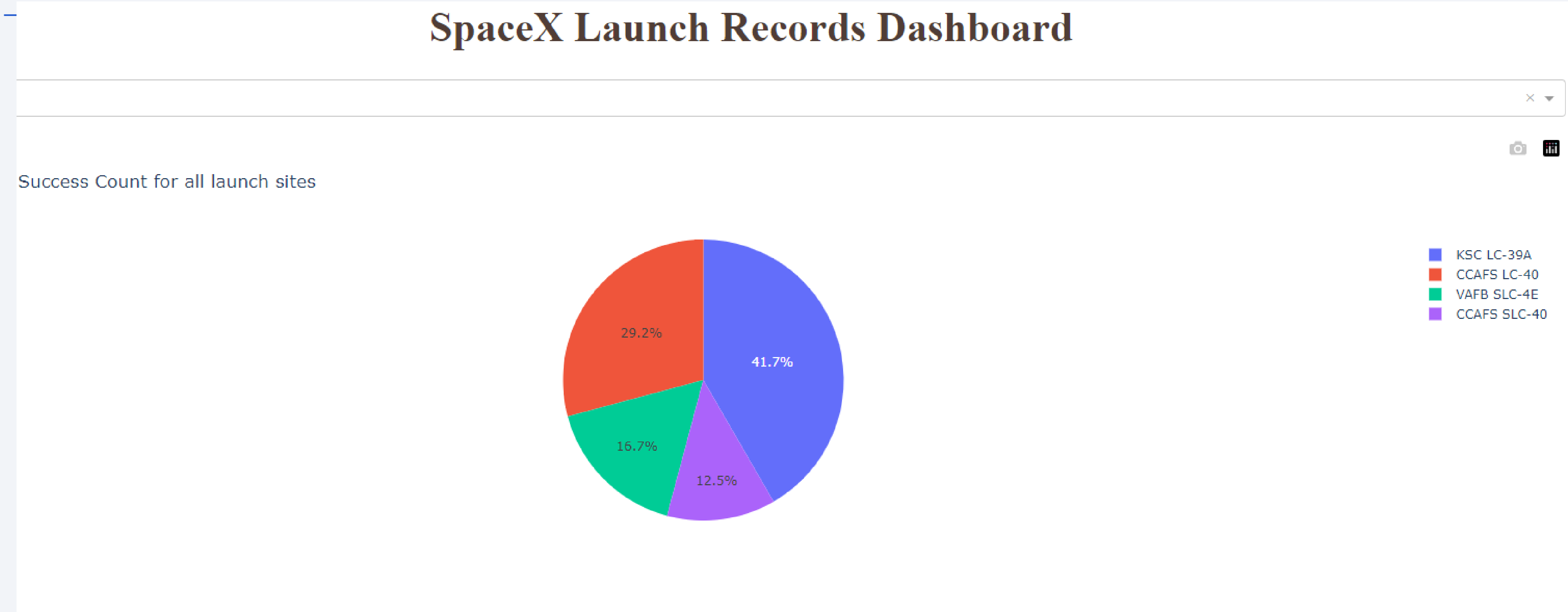


- The distances calculated are less than 1km each

Section 4

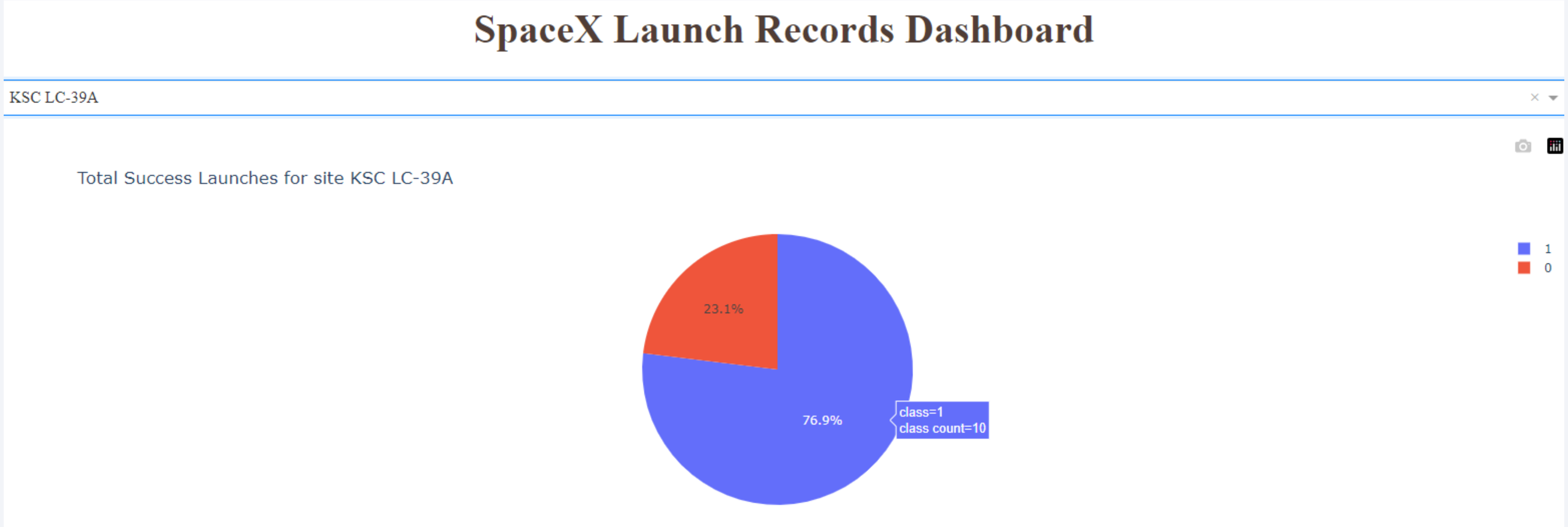
Build a Dashboard with Plotly Dash

launch success count for all sites, in a piechart



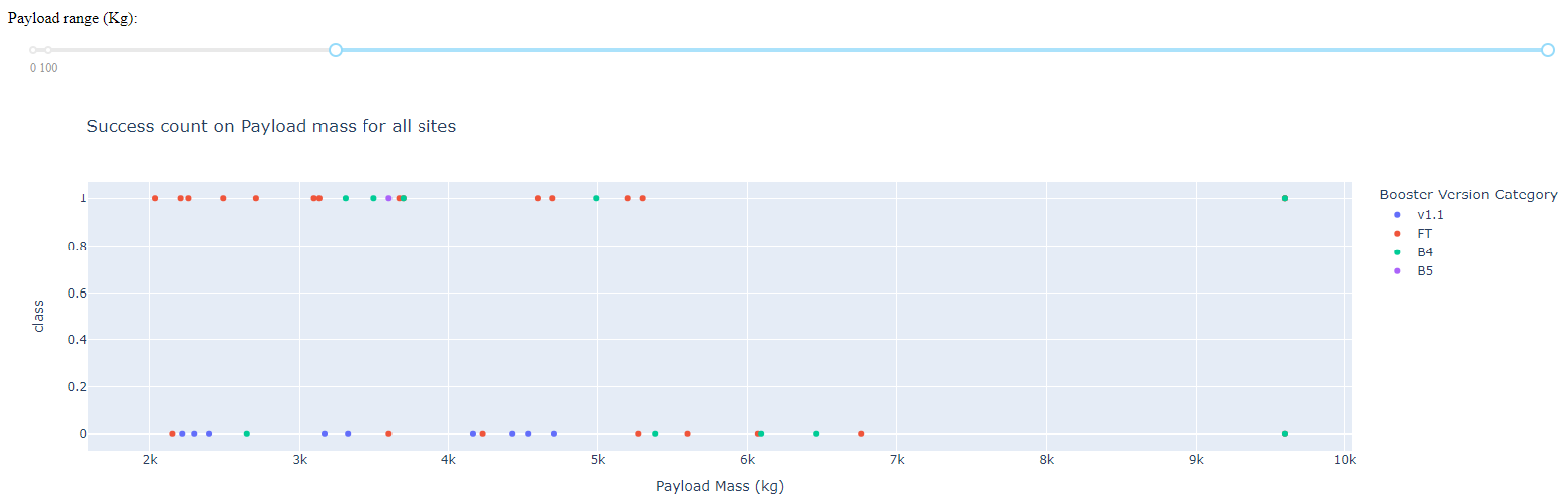
The launch site with the highest success rate is KSC LC-39A , accounting for 41% of all successful lauches.

Pie chart for the launch site with highest launch success ratio



This station has 77% successful Launches

Payload vs. Launch Outcome scatter plot for all sites

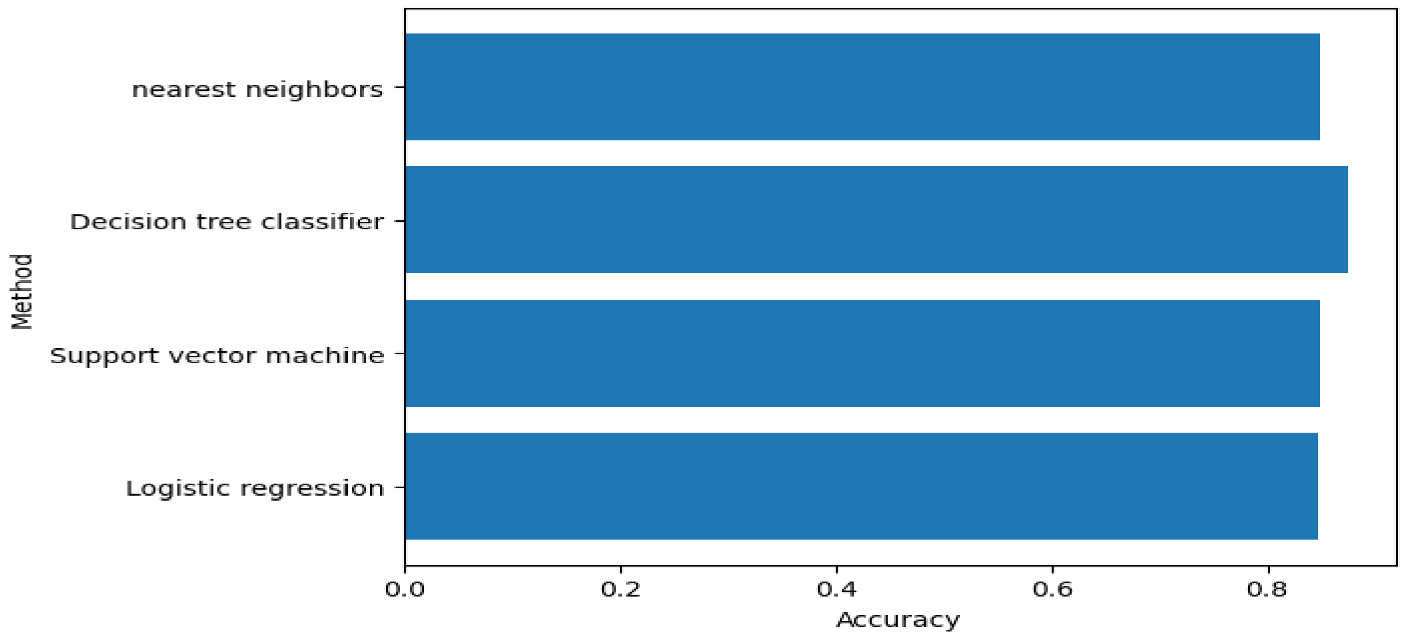


Section 5

Predictive Analysis (Classification)

Classification Accuracy

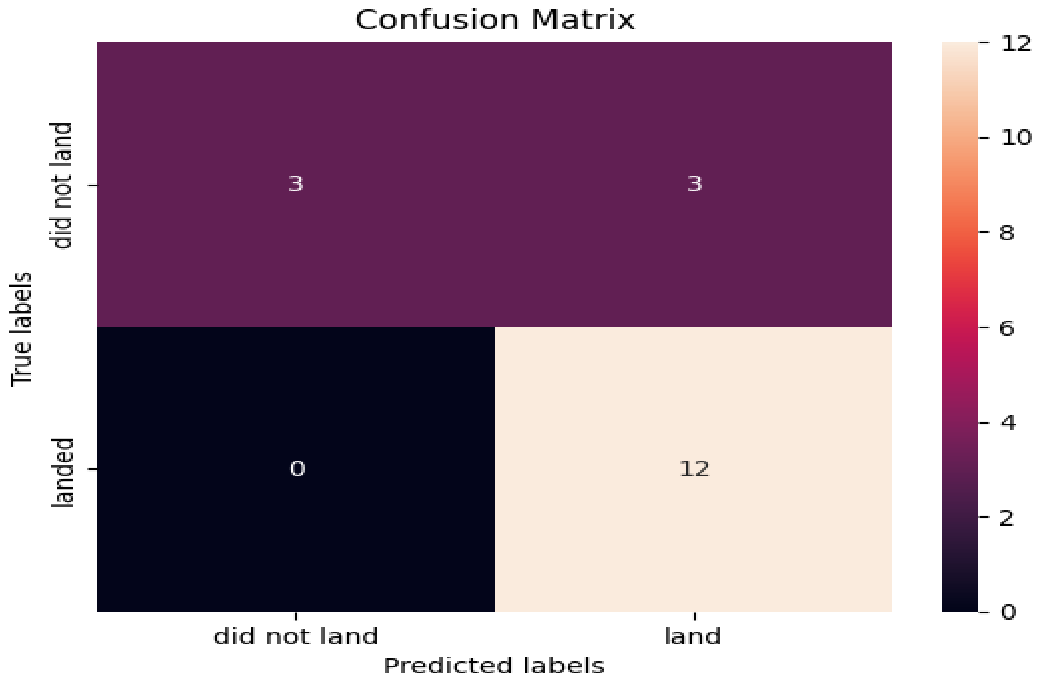
```
import matplotlib.pyplot as plt

plt.barh(result_df['Methods'], result_df['Accuracy'])
plt.xlabel('Accuracy')
plt.ylabel('Method')
plt.show()
```



Decision Tree Classifier model is the best performing model

Confusion Matrix

```
[42]: yhat = tree_cv.predict(X_test)
      plot_confusion_matrix(Y_test,yhat)
```



There are 3 true positives and 12 True Negatives

Conclusions

- Using Decision Tree Classifier model, we have about 87.5% accuracy in prediction the outcome of launches.
- The launch site with the highest success rate is KSC LC-39A, This station has 77% successful launches
- We see that as the flight number increases, the first stage is more likely to land successfully.
- From the insight in our analysis, we can safely predict success landing of the first stage, hence able to manage the overall cost and compete with spaceX by reusing it.
-

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!