IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Tayla Ivory-Cousins
30/01/2023
[taylaivory/IBM-Captone-Project (github.com)]

Outline

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

Executive Summary

This project uses data collected from SpaceX and Wikipedia to analyze past Falcon 9 launches and predict the successful landing of the first stage.

- Data was collected using web scrapping and REST API queries
- Exploratory Data Analysis (EDA) was completed using Python and SQL
- Interactive Visual Analytics and Dashboard were created using Folium and DASH
- Machine Learning models were used to predict the success of landing

Introduction

SpaceX is a key player in the commercial space age, their ability to reuse the first stage allows for the rocket launches to be relatively inexpensive.

Determining if the first stage will land influences the cost of the launch ($62 million compared to their competitors of upwards of $165 million)

Prediction of success is therefore crucial, to do this we much answer the following questions:

- What trends can be seen from the previous landings?
- What factors influence the success of first stage landing?
- How can we at SpaceY use this information to determine our position in the commercial space age?

Section 1

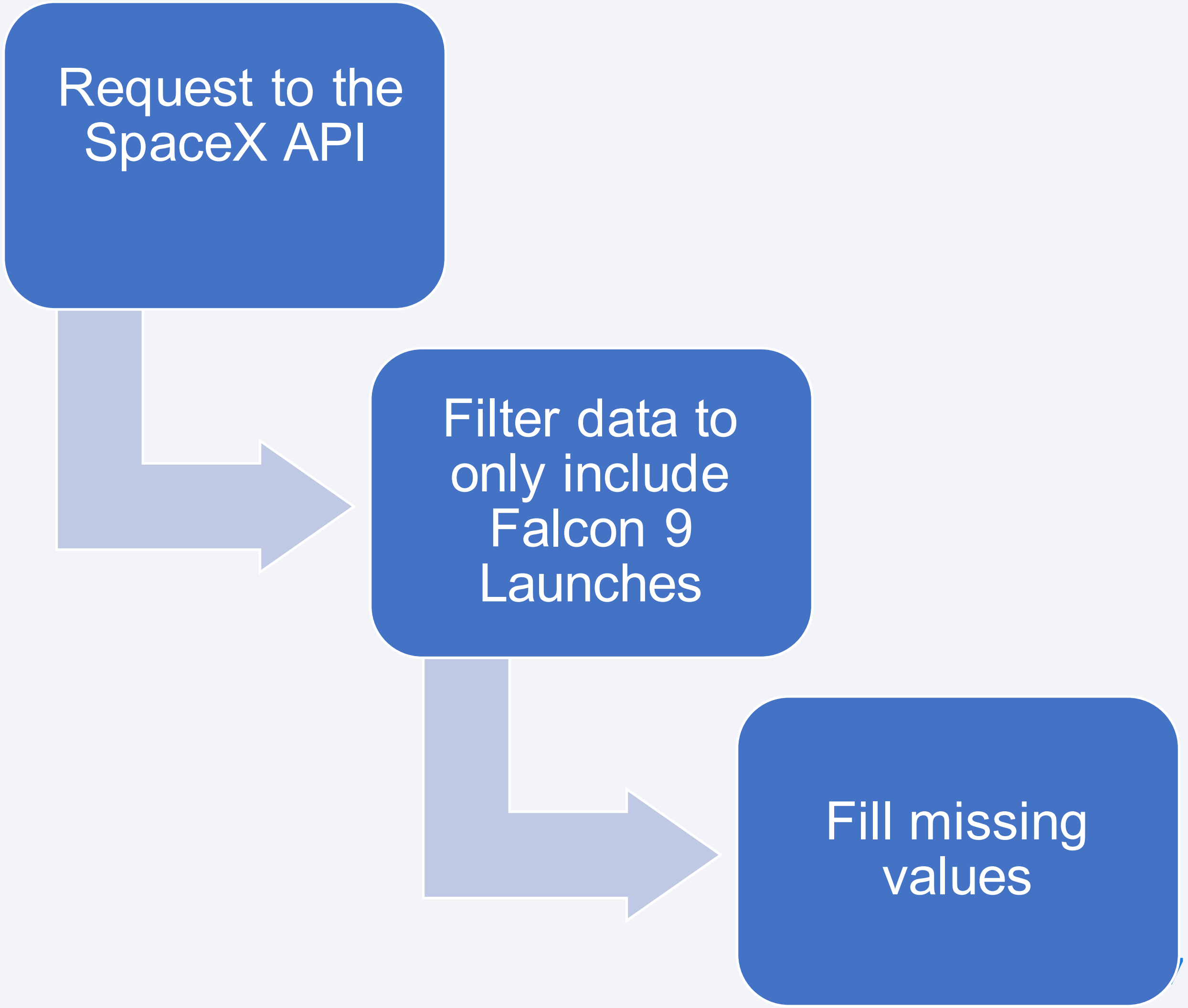
Methodology

Methodology

- Data collection methodology:
  - Data regarding SpaceX was obtained from 2 sources for this project
  - SpaceX API https://api.spacexdata.com/v4/rockets/
  - WebScraping (List of Falcon 9 and Falcon Heavy launches – Wikipedia)
- Perform data wrangling
  - The data was processed by creating a classification variable representing the landing outcome (0 or 1)
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - The data was first normalized, divided into training and test data sets and used to evaluate four machine learning models and find the accuracy of each model using the optimal parameters

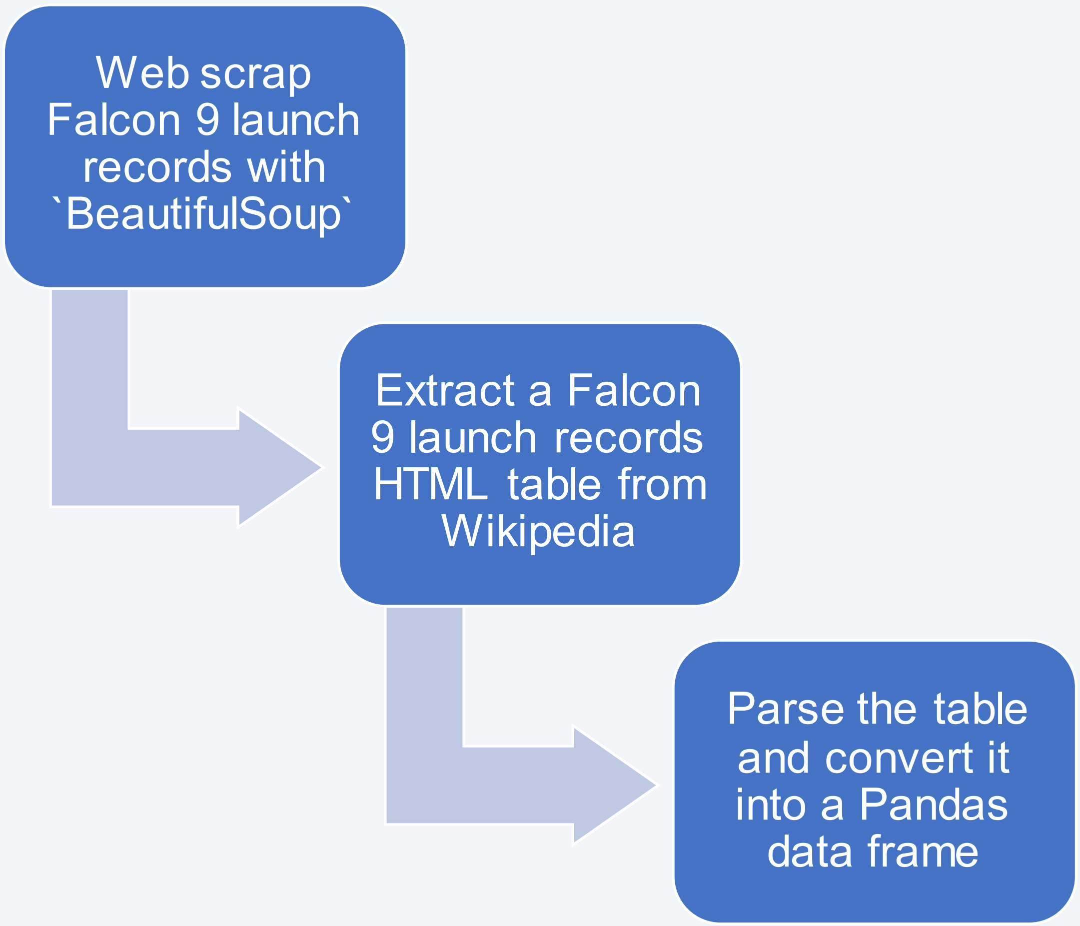
Data Collection – SpaceX API

- The API has data on other rocket launches however we only focused on Flacon9 launches therefore we filtered the data to only include them
- Missing values are filled with the mean value of the corresponding column
- IBM-Captone-Project/Data_Collection_API_Lab.ipynb at main · taylaivory/IBM-Captone-Project (github.com)



Data Collection - Scraping

- Data was scraped from List of Falcon 9 and Falcon Heavy launches – Wikipedia)
- This data was only for Falcon 9 launches so didn't need to filter for this
- IBM-Captone-Project/Web_scraping_Wikipedia_Lab.ipynb at main · taylaivory/IBM-Captone-Project (github.com)



Data Wrangling

- Exploratory Data Analysis (EDA) was performed on the dataset
- Summarization of the data was completed
  - Launches per site
  - Occurrences of each orbit
  - Occurrence of mission outcome per orbit type
- Then, the classification variable was created- the landing outcome label either 0 (failure) or 1 (success)
- IBM-Captone-Project/Data_Wrangling_Lab.ipynb at main · taylaivory/IBM-Captone-Project (github.com)

EDA with Data Visualization

- Multiple charts were created to visualize the data and understand the relationship between the features and how they influenced the launding outcome
- Scatter plots, bar charts and line charts were utilized
- IBM-Captone-Project/EDA_Visualisation_Lab.ipynb at main · taylaivory/IBM-Captone-Project (github.com)

EDA with SQL

- Using SQL we queried the data to help explore the collected data
- These included:
  - The names of the unique launch sites
  - Successful Drone Ship Landing with Payload between 4000 and 6000
  - Average Payload Mass by F9 v1.1
- IBM-Captone-Project/EDA_with_SQL_lab.ipynb at main · taylaivory/IBM-Captone-Project (github.com)

Build an Interactive Map with Folium

- A Folium map was created and included markers, circles, lines, and marker clusters
- Markers were used to show the launch sites
- Circles were used to highlight areas around specific coordinates
- Marker clusters were used to group the launches from each launch site and were colour coded dependent on success
- Lines were added to show distances between two coordinates
- IBM-Captone-Project/Interactive_Visual_Analytics_Folium.ipynb at main · taylaivory/IBM-Captone-Project (github.com)

Build a Dashboard with Plotly Dash

- Dash was used to create an interactive dashboard
- The dashboard included a dropdown menu, graphs and plots and a slider
- The graphs and plots allowed for visualization of
    - Percentage of launches by site
    - Payload range
- IBM-Captone-Project/Interactive_Dashboard_DASH_Lab.py at main · taylaivory/IBM-Captone-Project (github.com)

Predictive Analysis (Classification)

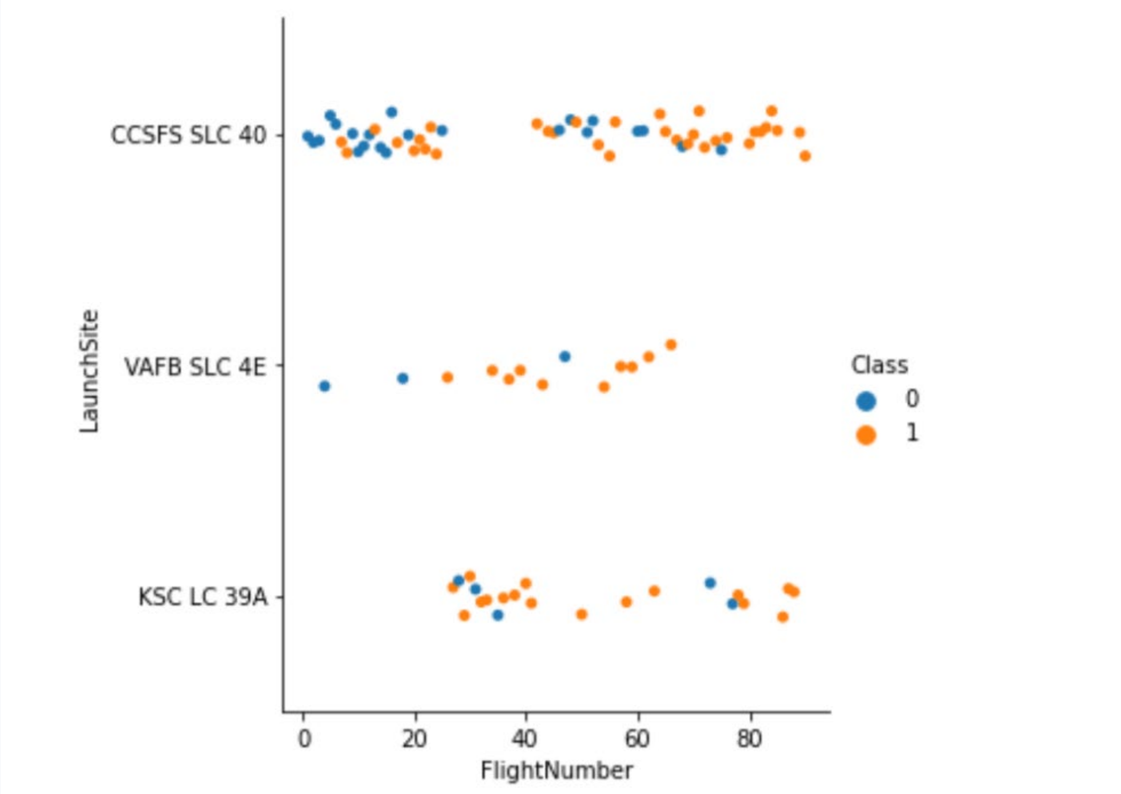
- Four Machine Learning models were tested: Logistic Regression, Decision Tree, Support Vector Machine and K-Nearest Neighbors
- The method to evaluate the models was as followed
  - Normalize the data
  - Split the data into training and testing data sets
  - Create the machine learning models
  - Fit the models on the training data set
  - Find the best hyperparameters for each model
  - Evaluate the models based on the accuracy and the test accuracy and compare the confusion matrixes
- IBM-Captone-Project/Machine_Learning_Lab.ipynb at main · taylaivory/IBM-Captone-Project (github.com)

Section 2

Insights drawn from EDA

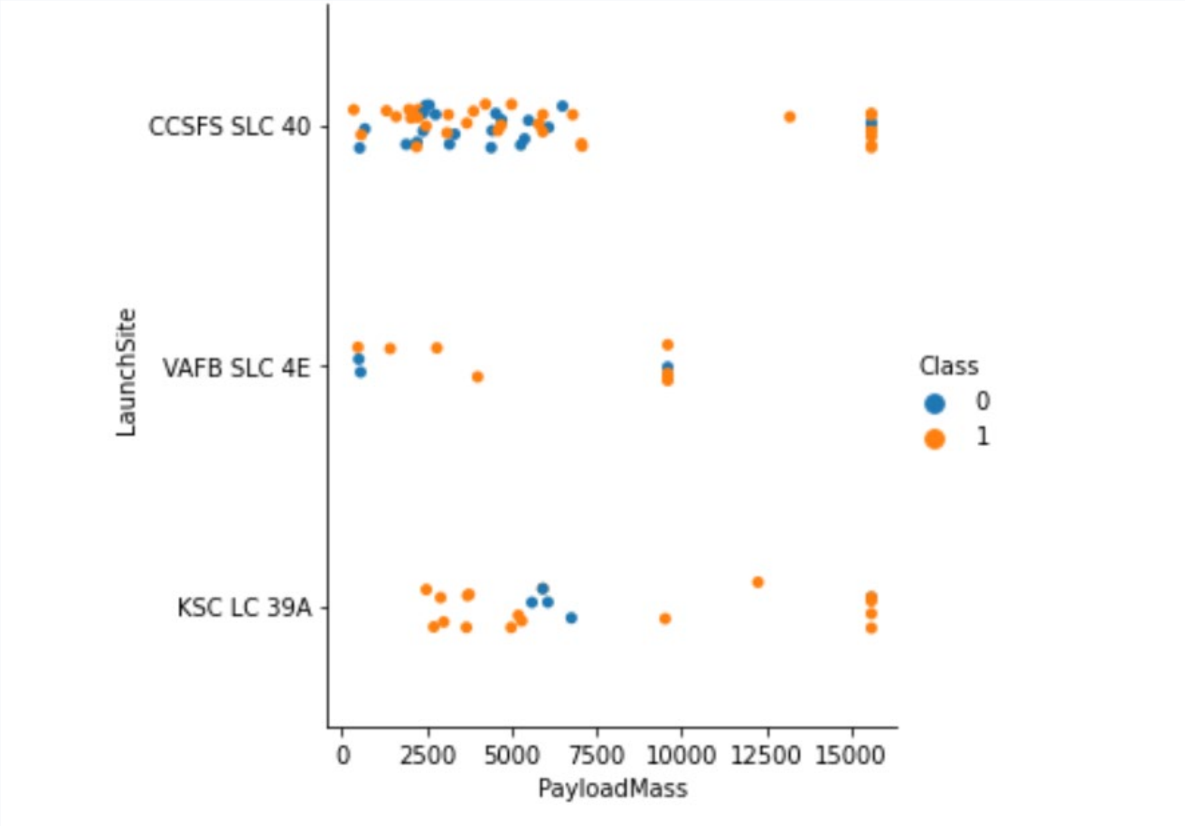
Flight Number vs. Launch Site

- From the plot we can determine the most successful launch site
- Class 1 represents successful flights therefore it is clear that CCAF5 SLC 40 has been the most successful launch site
- Can also see that over time the general success rate has improved at every site



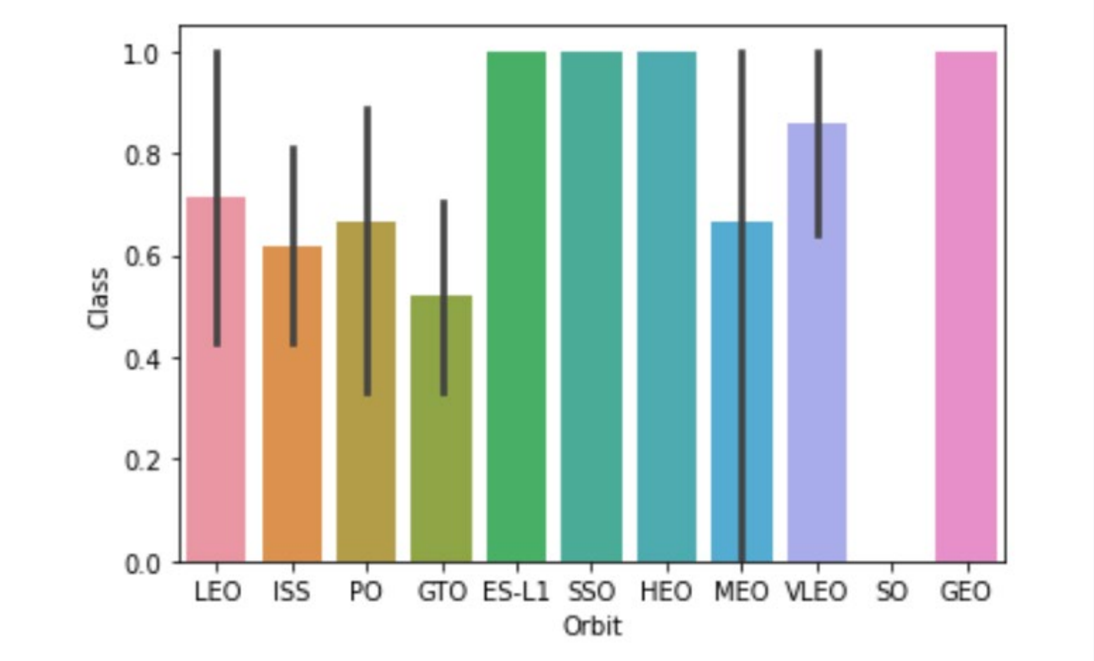
Payload vs. Launch Site

- Payloads over 9,000 kg have an excellent success rate
- Payloads over 12,000 kg have only been launched from CCAF5 SLC 40 and KSC LC 39A launch sites



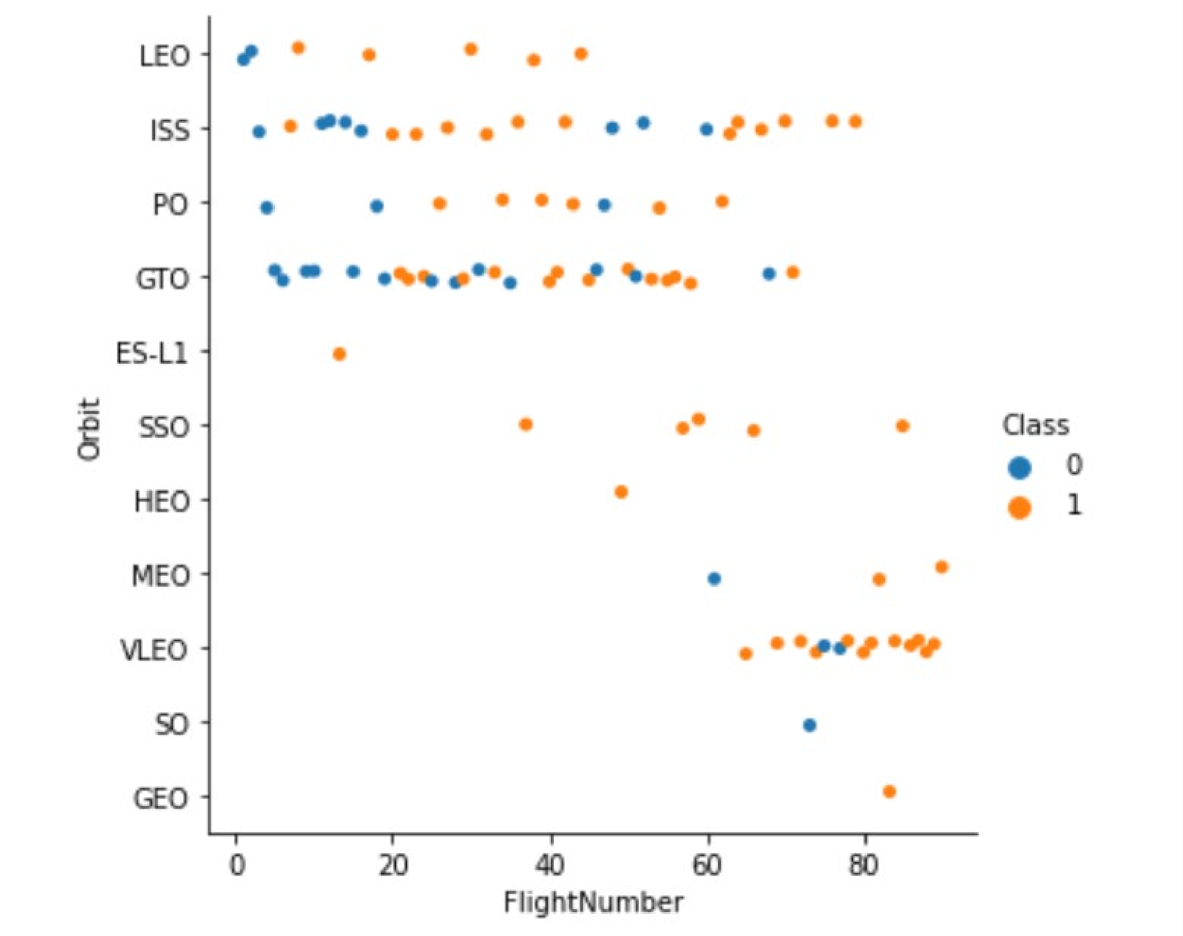
Success Rate vs. Orbit Type

- The greatest success rates are in the following orbits:
  - ES-L1
  - SSO
  - HEO
  - GEO



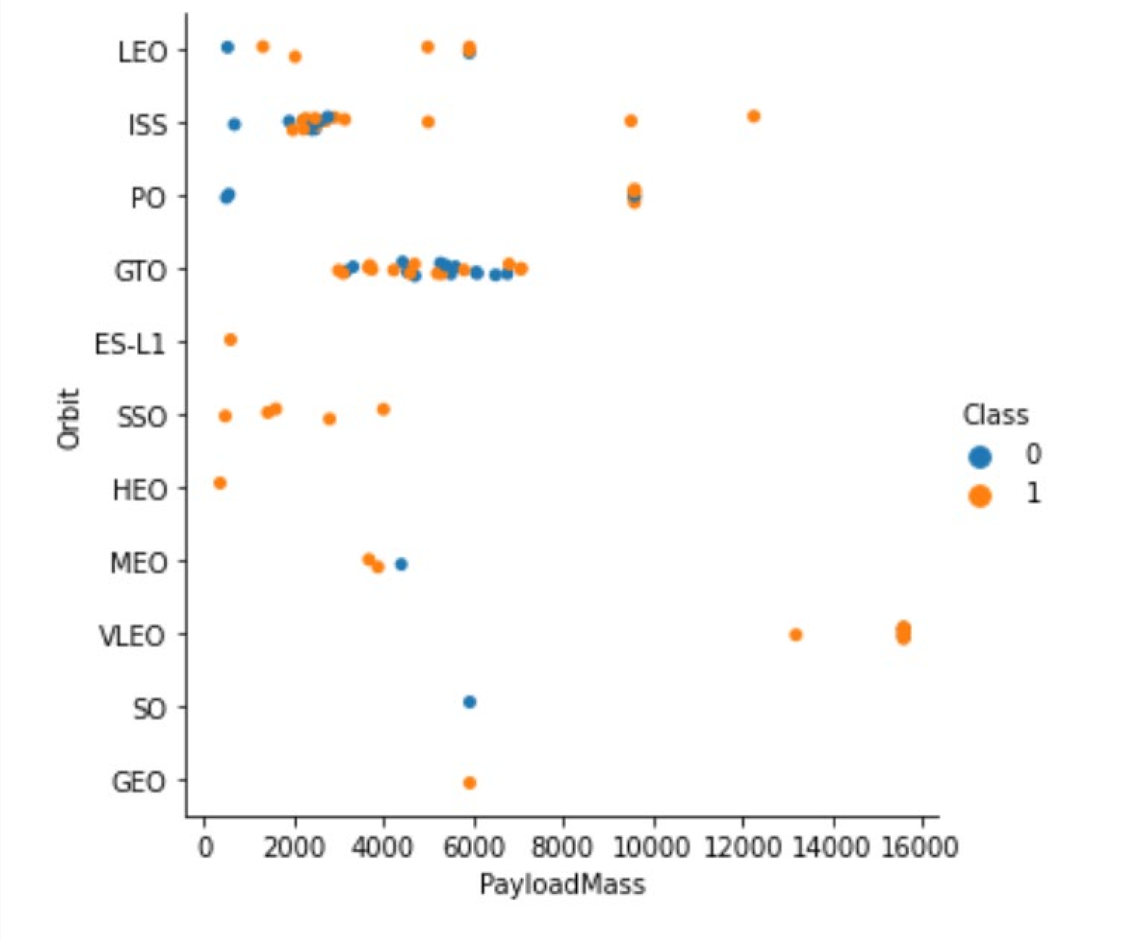
Flight Number vs. Orbit Type

- Success rate has improved over time for all orbit types
- VLEO orbit has recently become more frequent
- Some orbits have only 1 flight at a success, using the previous figure would change our perspective of the most successful orbit types



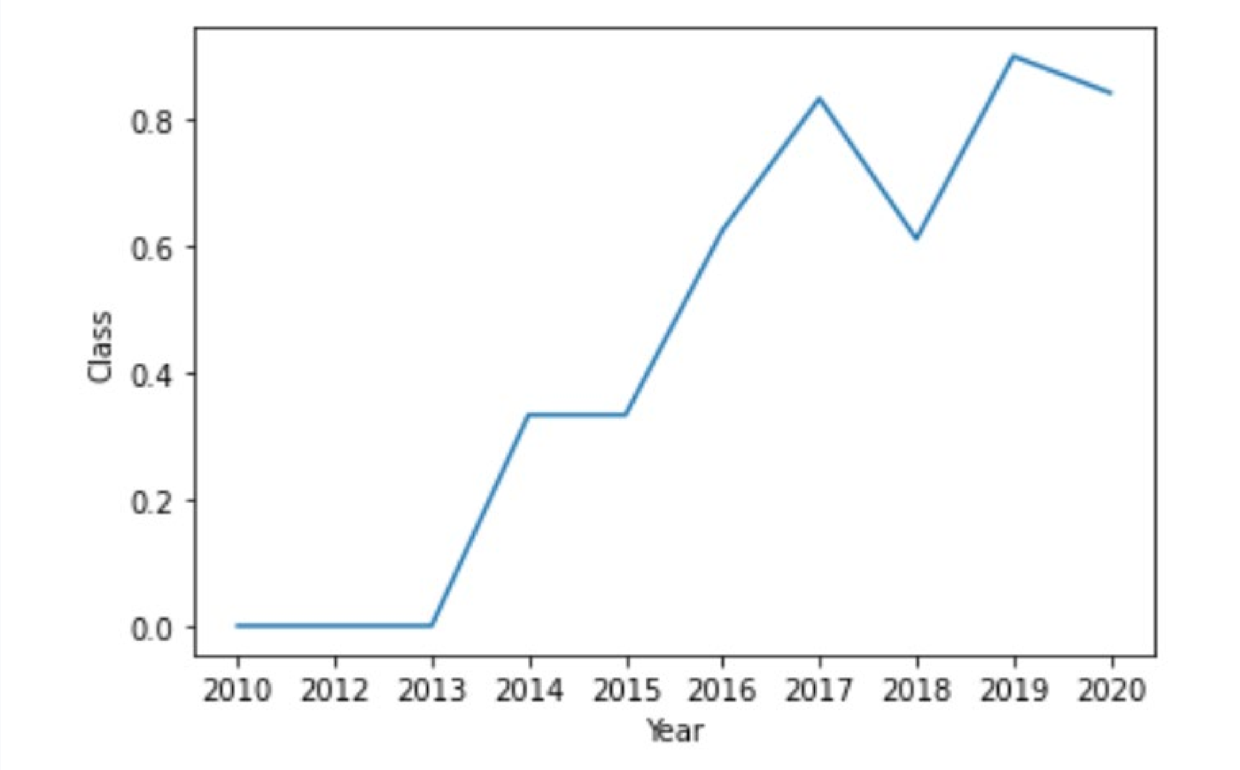
Payload vs. Orbit Type

- ISS orbit has the greatest range of tested payloads, higher payloads seem to have a better success rate
- GTO orbit shows no correlation to the payload mass and the success rate



Launch Success Yearly Trend

- The first 3 years had no success
- Sharp increase from 2013 to 2014
- Period of steady success between 2014 to 2015
- Continuous sharp increase in success between 2015 and 2017
- Small decrease between 2017 and 2018
- Reached the peak of success in 2019



All Launch Site Names

```
SELECT DISTINCT LAUNCH_SITE FROM SPACEXTBL
```

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

Launch Site Names Begin with 'CCA'

SELECT * FROM SPACEXTBL WHERE LAUNCH_SITE LIKE ‘CCA%’ LIMIT 5

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

SELECT SUM(PAYLOAD_MASS__KG_ FROM SPACEXTBL WHERE CUSTOMER == ‘NASA (CRS)’

- 45596 KG

Average Payload Mass by F9 v1.1

SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE BOOSTER_VERSION == ‘F9 v1.1’

- 2928.4 KG

First Successful Ground Landing Date

SELECT MIN(DATE) FROM SPACEXTBL WHERE LANDING_OUTCOME == ‘Success (ground pad)’

- 2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

SELECT BOOSTER_VERSION FROM SPACEXTBL WHERE LANDING_OUTCOME == 'Success (drone sip)' AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 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

SELECT MISSION_OUTCOME, COUNT(*) FROM SPACEXTBL FROM SPACEXTBL GROUP BY MISSION_OUTCOMES

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

Boosters Carried Maximum Payload

SELECT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_=(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

SELECT DATE, BOOSTER_VERSION, LAUNCH_SITE FROM SPACEXTBL WHERE YEAR(DATE) == ‘2015’ AND LANDING_OUTCOME == ‘Failure (drone ship)’

DATE	booster_version	launch_site
2015-01-10	F9 v1.1 B1012	CCAFS LC-40
2015-04-14	F9 v1.1 B1015	CCAFS LC-40

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

SELECT LANDING_OUTCOME, COUNT(LANDING_OUTCOME) FROM SPACEXTBL WHERE DATE BETWEEN ‘2010-06-04’ AND ‘2017-03-20’ GROUP BY LANDING_OUTCOME ORDER BY DATE DESC

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

Section 3

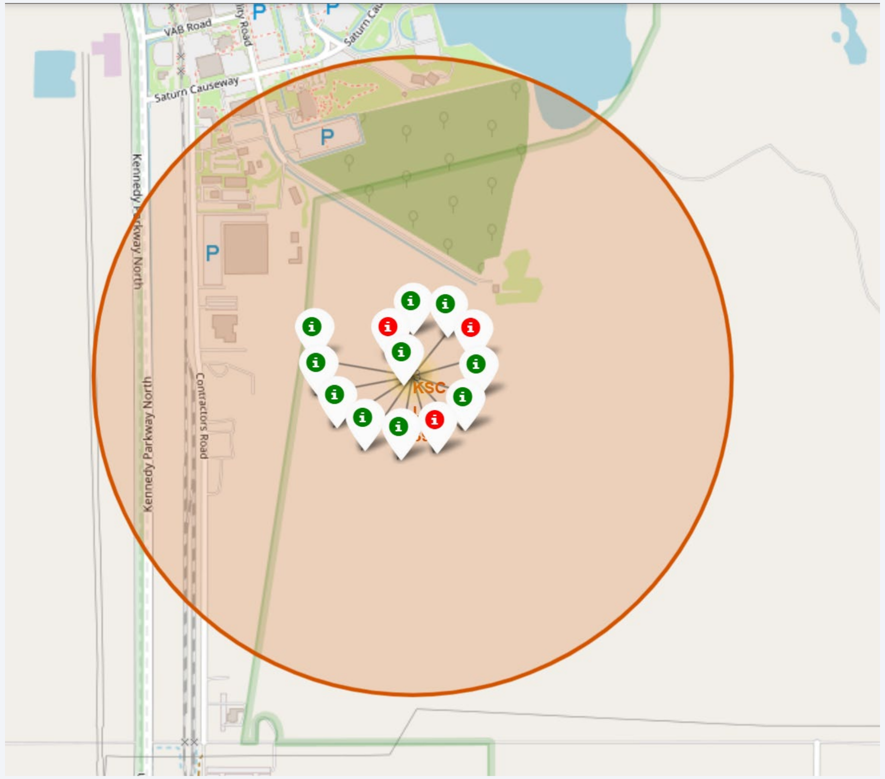
Launch Sites Proximities Analysis

All launch sites

- All the Launch sites can be seen on the interactive folium map,

Launch Outcomes

- This screenshot shows the 13 launch outcomes at KSC LC-39A and they are colour coded depending on their success: red for failure, green for success

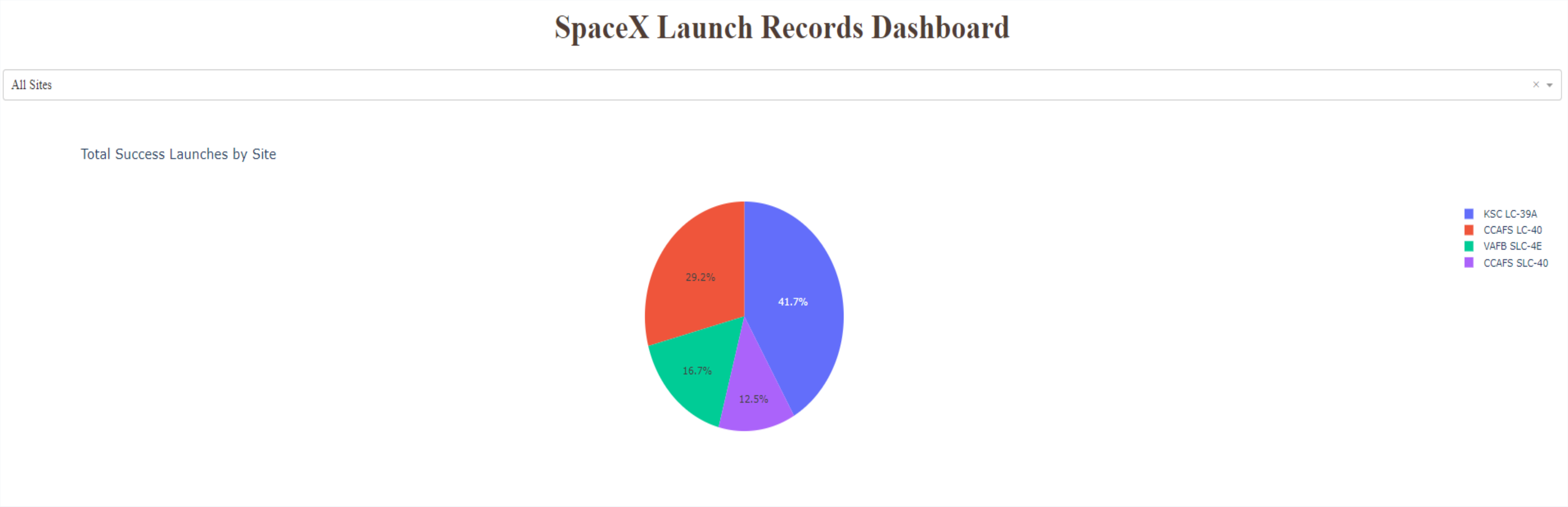


Section 4

Build a Dashboard with Plotly Dash

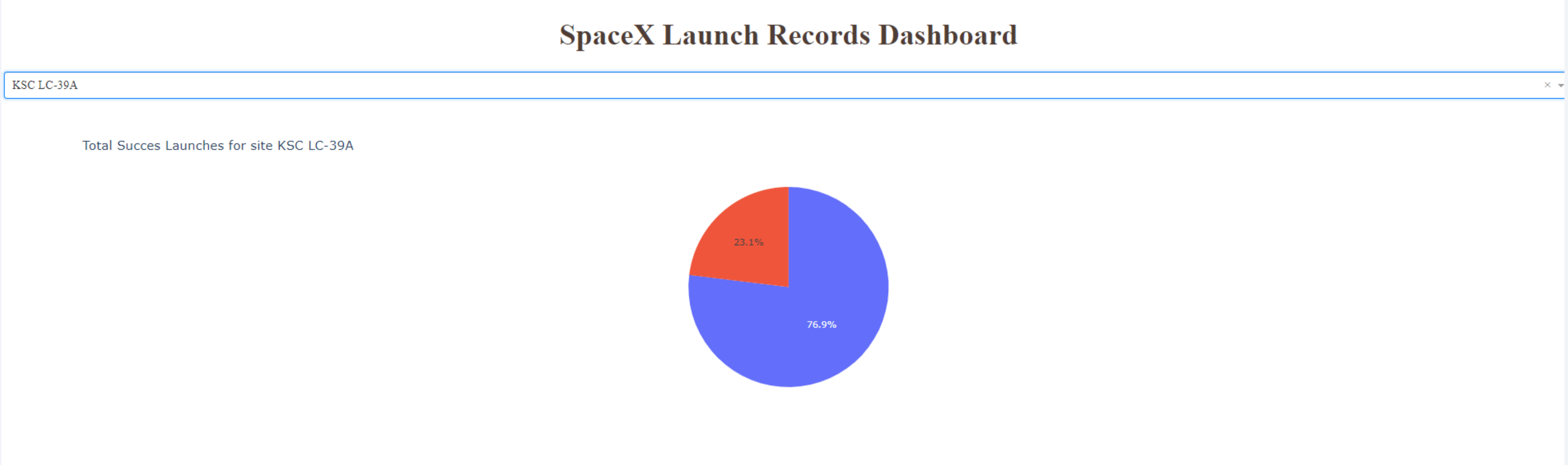
Successful Launches by Site

- The screenshot of the DASH interactive app highlights how the launch sites play an important factor in the success of the launches



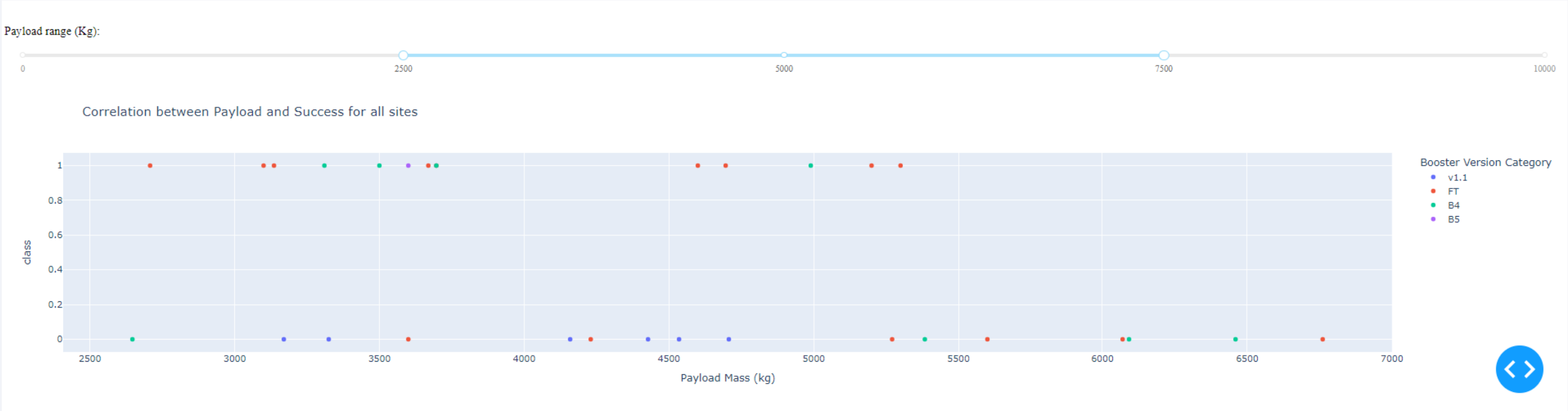
Most successful launch site

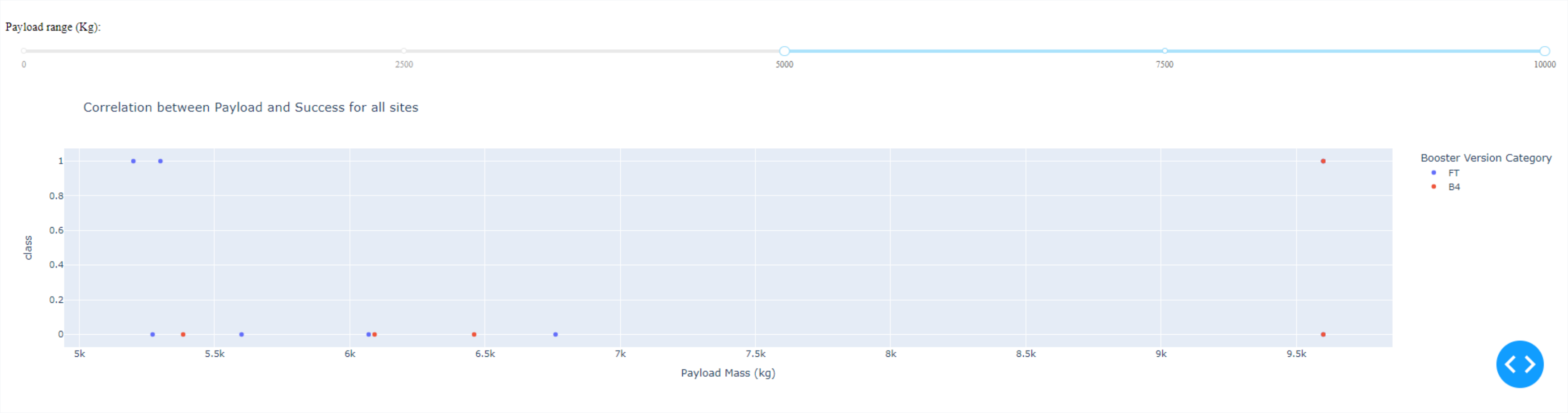
- KSC LC-39A has the highest success ratio with 76.9 % of launches being successful



Payload vs. Launch Outcome for all sites

- The below screenshots are of Payload vs. Launch Outcome of all sites, with different payload selected in the range slider, 2500 to 7500 and 5000 to 10000



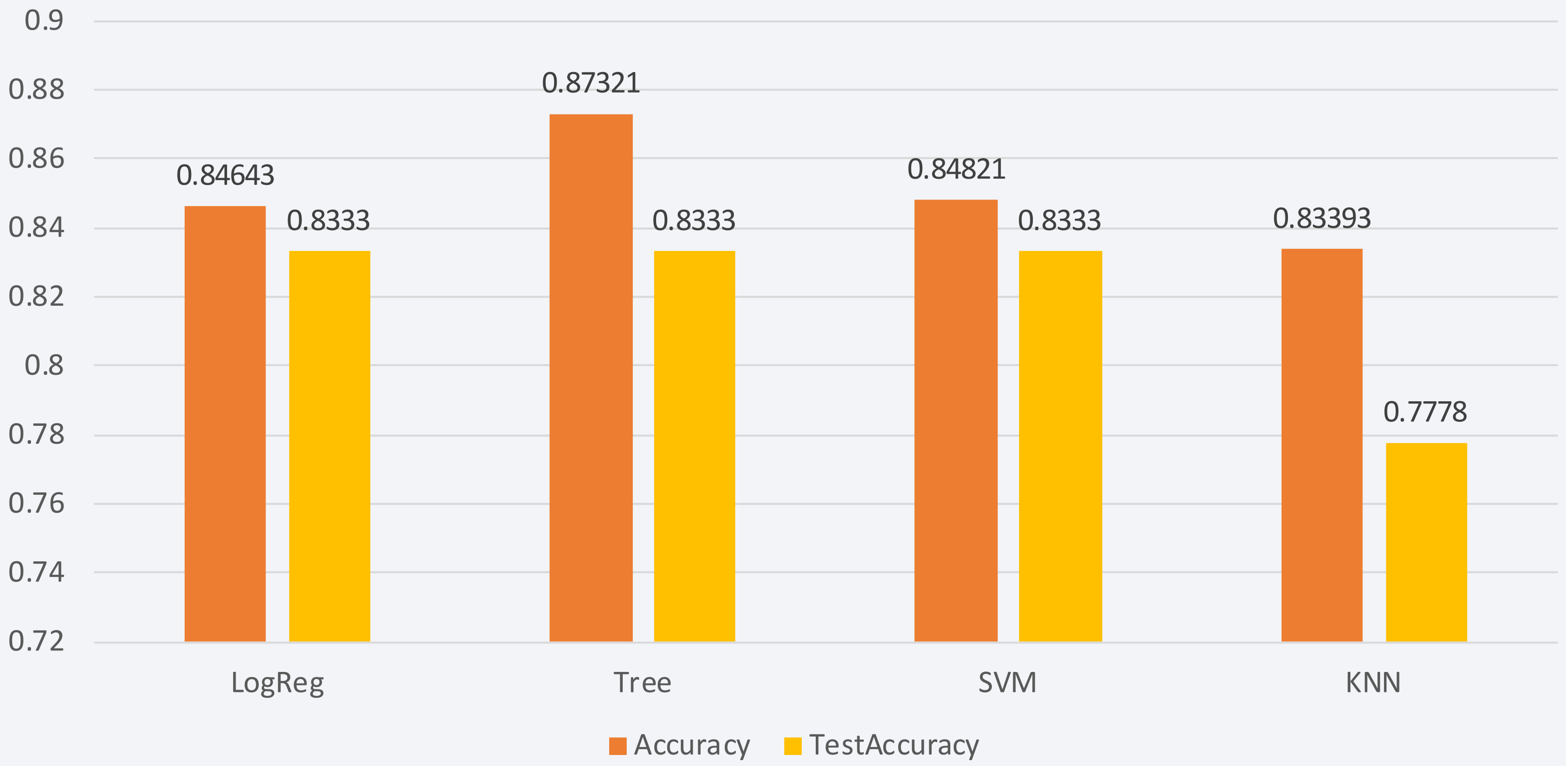


Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Four Machine Learning models were tested: Logistic Regression, Decision Tree, Support Vector Machine and K-Nearest Neighbors

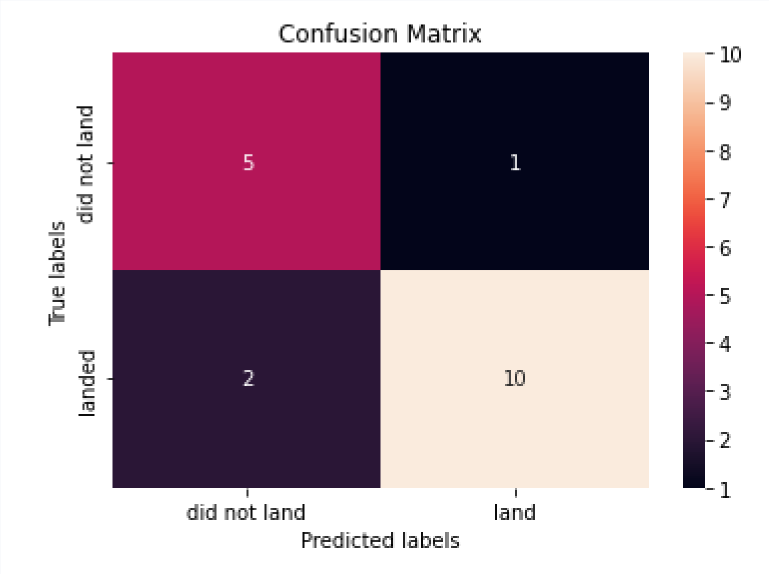


Confusion Matrix

The best Machine Learning model tested was the Decision Tree

It had the highest number of True Positives and True Negatives

Had an accuracy score of 0.833



Conclusions

- Using the different methodologies to analyze the data collected the following conclusions can be drawn:
- The most successful launch site is KSC LC-39A with a success rate of 76.9 %
- Successful landings have become more frequent over time, at a peak in 2019, this is likely due to the evolution of technology as well as learning through experience
- The Decision Tree model was the most accurate at predicting the success of launches and therefore could be used to help predict the cost of launches

Thank you!