Outline

- Executive Summary

- Introduction

- Methodology

- Results

- Conclusion

- Appendix

Executive Summary

Problem:

According to CB Insights statistics, 20% of startups fail because they got outcompeted while 8% of startups were sunk because of poor products. Our competitor SpaceX, dominates the market because of the reusability of its rockets. To outcompete them, it is imperative to determine our product performance. In this case, we need a way to predict the landing outcome of our rockets after they are launched.

Solution:

We collected data on SpaceX's Falcon 9 rocket launches and used it to train Machine Learning models that will predict whether the first stage of the rocket will land successfully. This was done through the following ways:

- Data Collection

- Data Analysis

- Predictive Modeling using Machine Learning Algorithms

Note that the data collection stage was carried out using API requests and Web Scraping in order to minimize costs.

Introduction

- The aim of this project is to allow the company to compete with SpaceX. In order to achieve the goal, it necessary to predict whether they will re-land their first stage.

- 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 the cost of a launch. This information can be used if an alternate company wants to bid against space X for a rocket launch.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:

  - Collecting data from SpaceX API.

  - Web scraping Wikipedia for Falcon 9 related data.

- Perform data wrangling

  - Data has been cleaned , explored to find patterns and labeled accordingly for Supervised machine learning models.

- Perform exploratory data analysis (EDA) using visualization and SQL

- Perform interactive visual analytics using Folium and Plotly Dash

- Perform predictive analysis using classification models

  - Create a machine learning pipeline to predict if the first stage will land or not.

  - Find the best performing model to train the data and make Predictions.

Data Collection

The data collection stage is arguably the most crucial stage in the project. This is because we use data to train our machine learning models to make precise predictions.

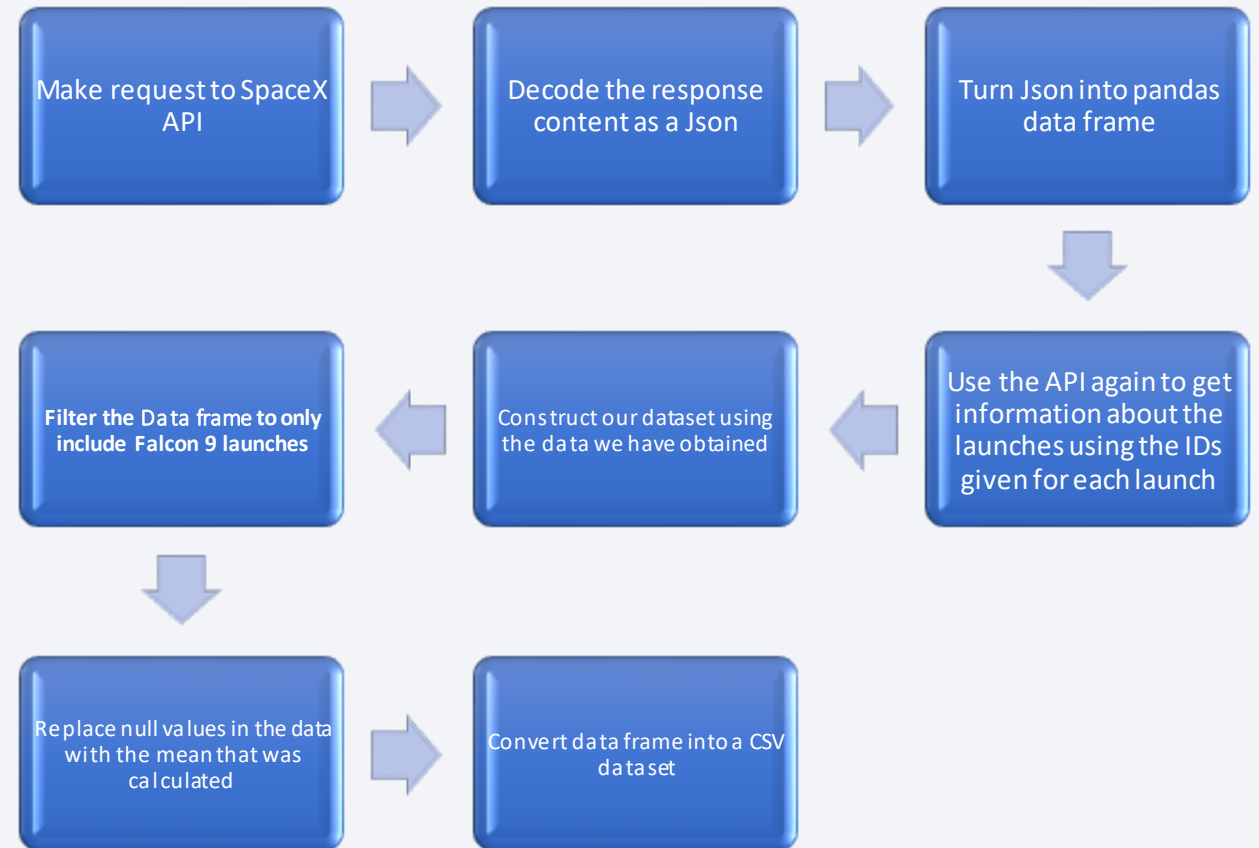There are different ways to go about collecting the data but we used two methods:

- Data collection by SpaceX API request.

- Data collection by Web Scraping

These are all low-cost methods that only require a functioning internet connection.
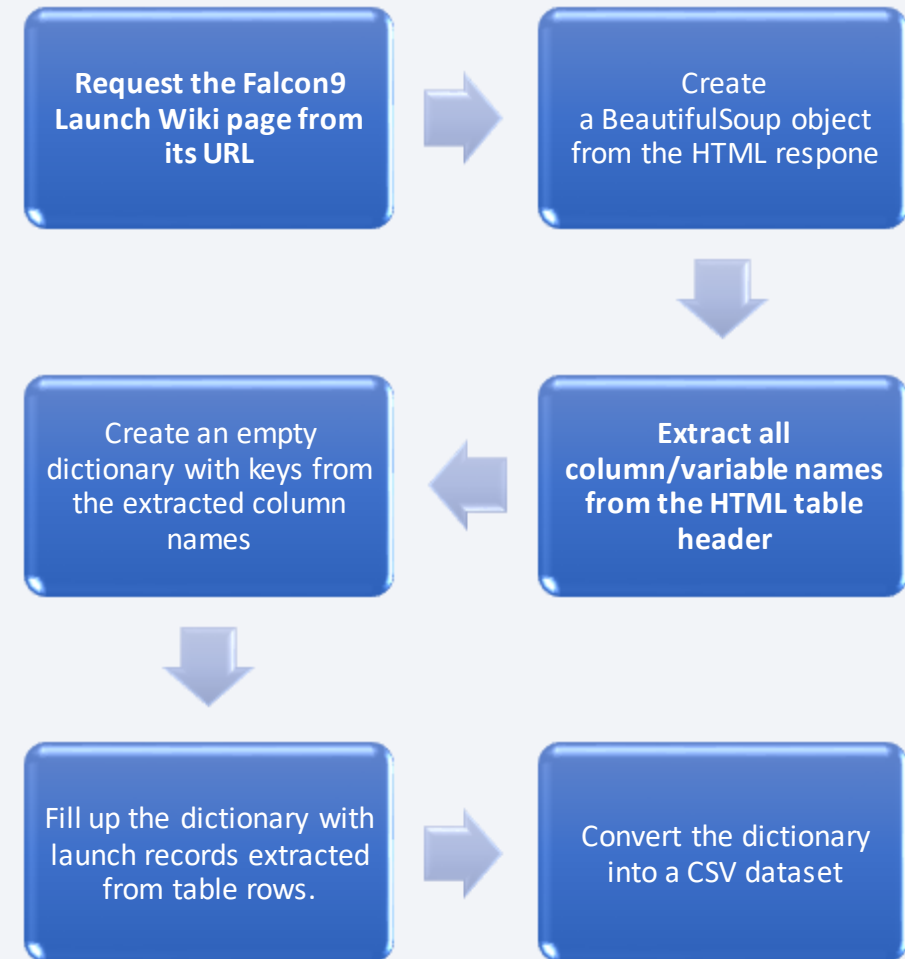
Data Collection – SpaceX API

- Make a request to SpaceX API and make sure the data is in the correct format.

- Perform some basic data wrangling and formatting in order to clean the requested data.

- Convert our data frame into a CSV dataset.

- URL: https://github.com/Kumaresanveerasamy/IBM_Datascience_Professional_Course/blob/master/Applied_Capstone_Project/Week_1/DataScraping.ipynb\

Make request to SpaceX API	→	Decode the response content as a Json	→	Turn Json into pandas data frame

↓

Filter the Data frame to only include Falcon 9 launches	←	Construct our dataset using the data we have obtained	←	Use the API again to get information about the launches using the IDs given for each launch

↓

Replace null values in the data with the mean that was calculated	→	Convert data frame into a CSV dataset

Data Collection - Web Scraping

- Using Beautiful Soup to perform web scraping on the Wikipedia page with title: List of Falcon 9 and Falcon Heavy launches

- Store the launch records in an HTML table.

- Parse the table and convert it into a CSV dataset.

- URL: https://github.com/Kumaresanveerasamy/IBM_Datascience_Professional_Course/blob/master/Applied_Capstone_Project/Week_1/DataScraping.ipynb\
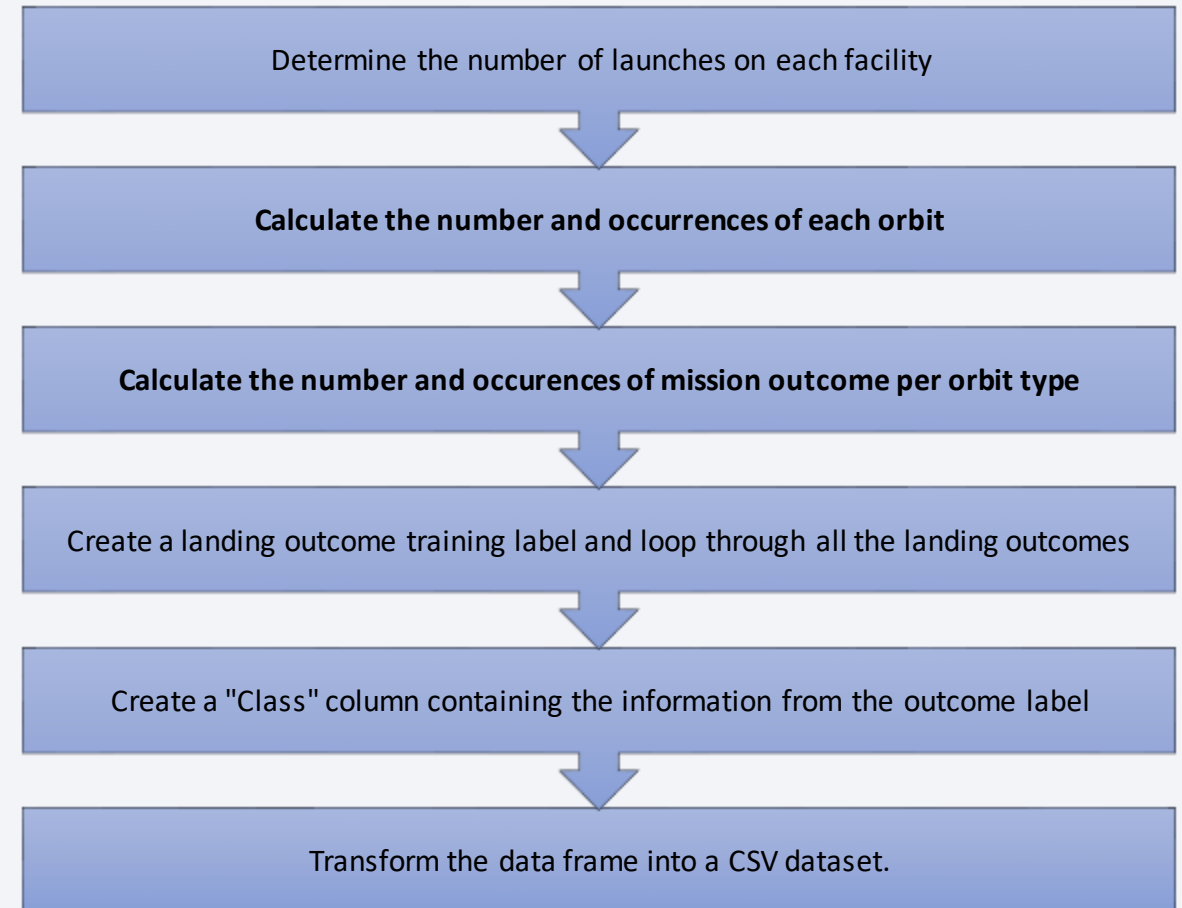
```
Request the Falcon9 Launch Wiki page from its URL
  →
Create a BeautifulSoup object from the HTML respone
  ↓
Create an empty dictionary with keys from the extracted column names
  ←
Extract all column/variable names from the HTML table header
  ↓
Fill up the dictionary with launch records extracted from table rows.
  →
Convert the dictionary into a CSV dataset
```

Data Wrangling

- The goal in this stage is to find patterns in the data and determine the label for training supervised machine learning models.

- In the data set, there are several different cases where the rocket did not land successfully. For example, *True RTLS* means the rocket successfully landed on a ground pad while *False RTLS* means the rocket unsuccessfully landed on a ground pad.

- Those outcomes were converted into Training Labels whereby *1* means the rocket landed successfully while *0* means it was unsuccessful.

URL: https://github.com/Kumaresanveerasamy/IBM_Datascience_Professional_Course/blob/master/Applied_Capstone_Project/Week_1/DataWrangling.ipynb\

Determine the number of launches on each facility

Calculate the number and occurrences of each orbit

Calculate the number and occurences of mission outcome per orbit type

Create a landing outcome training label and loop through all the landing outcomes

Create a "Class" column containing the information from the outcome label

Transform the data frame into a CSV dataset.

EDA with Data Visualization

Data visualization helps us understand data by curating it into a form that's easier to understand, highlighting the trends and outliers. Several types of charts were used in the visualization of the data:

- Cat plots and scatter plots were used to view the relationships of categorical variables like *Launch Site* and *Orbit*.

- A bar chart was used to visualize the success rate of each orbit type.

- A line chart was used to visualize the launch success yearly trend.

URL: https://github.com/Kumaresanveerasamy/IBM_Datascience_Professional_Course/blob/master/Applied_Capstone_Project/Week_2/EDA_pandas%20%26%20Mattplotlib.ipynb\

EDA with SQL

Summary of SQL queries that were used:

- Display the names of the unique launch sites in the space mission

- Compare the payload mass with boosters launched by NASA (CRS)

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

- List the total number of successful and failure mission outcomes

- Determine the dates of different landing outcomes

URL: https://github.com/Kumaresanveerasamy/IBM_Datascience_Professional_Course/blob/master/Applied_Capstone_Project/Week_2/SQLLite.ipynb\

Build an Interactive Map with Folium

- Folium Markers were used to show the SpaceX launch sites and their nearest important landmarks like railways, highways, cities and coastlines.

- Polylines were used to connect the launch sites to their nearest land marks.

- Furthermore, Folium Circles were used to highlight circle area of launch sites.

- In order to mark the success/failed launches for each site, marker clusters were used on the map. Whereby Red represents rocket launch failures while Green represents the successes.

- URL: https://github.com/Kumaresanveerasamy/IBM_Datascience_Professional_Course/blob/master/Applied_Capstone_Project/Week_3/Dashboards_folium.ipynb\
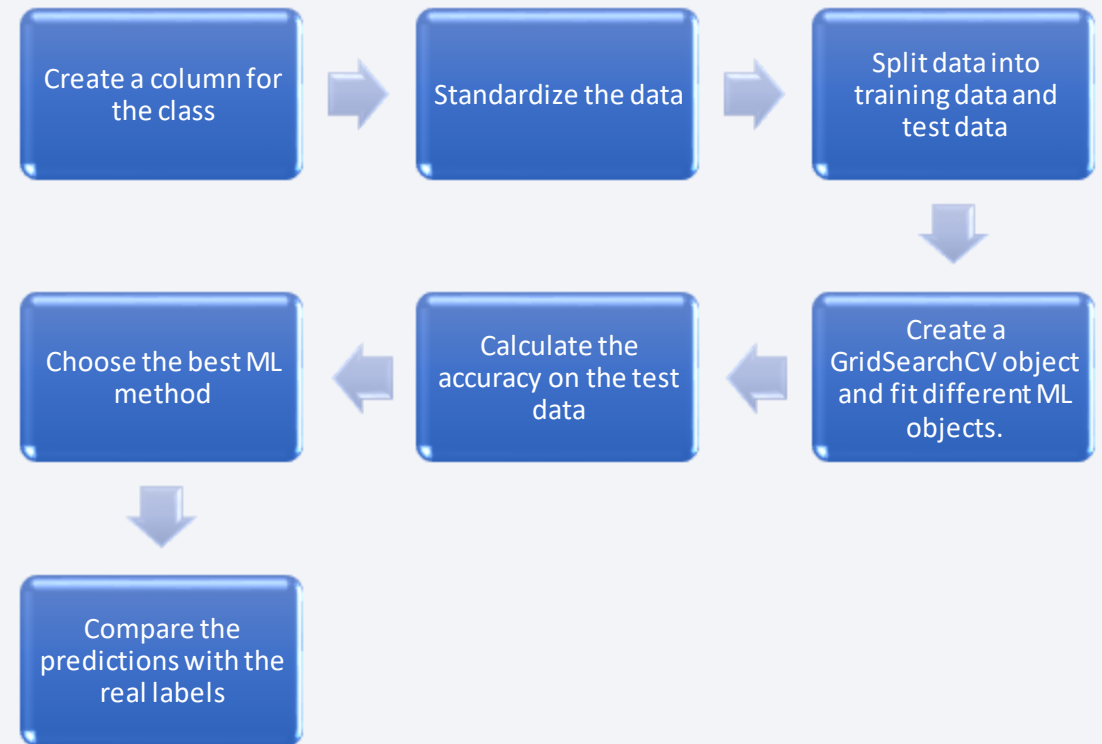
Build a Dashboard with Plotly Dash

- Pie charts and scatter charts were used to visualize the launch records of SpaceX.

- These charts displayed the rocket launch success rate per launch site. We are were able to get an understanding of the factors that may have been influencing the success rate at each site. Such as the payload mass and booster versions.

- Successful launches were represented by 1 while failures were represented by 0.

- URL: https://github.com/Kumaresanveerasamy/IBM_Datascience_Professional_Course/blob/master/Applied_Capstone_Project/Week_3/main.py\

Predictive Analysis (Classification)

Scikit-learn is the primary ML(machine learning) library that was used for predictive analysis. The following took place:

- Created a machine learning pipeline to predict if the first stage will land given the data.

- Using '*GridSearchCV*', found the best ML method for predictions.

- Compared the predictions with the real labels.

- The ML model scored an accuracy of **83.33%**

URL: https://github.com/Kumaresanveerasamy/IBM_Datascience_Professional_Course/blob/master/Applied_Capstone_Project/week_4/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb\

```
Create a column for the class  →  Standardize the data  →  Split data into training data and test data
                                                                      ↓
Choose the best ML method  ←  Calculate the accuracy on the test data  ←  Create a GridSearchCV object and fit different ML objects.
        ↓
Compare the predictions with the real labels
```

Results

- The exploratory data analysis has shown us that successful landing outcomes are somewhat correlated with flight number. It was also apparent that successful landing outcomes have had a significant increase since the year 2015.

- All launch sites are located near the coast line. Perhaps, this makes it easier to test rocket landings in the water.

- Furthermore, the sites are also located near highways and railways. This may facilitate transportation of equipment and research material.

- The machine learning models that were built, were able to predict the landing success of rockets with an accuracy score of 83.33%. This accuracy can be increased in future projects with more data.

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

- It appears that there were more successful landings as the flight numbers increased. It also seems that launch site **CCAFS SLC 40** had the most number of landing attempts while the site **VAFB SLC 4E** had the least number of attempts.

- Looking at the second chart, we can see that there is no Launch Site with a success rate below 60%.
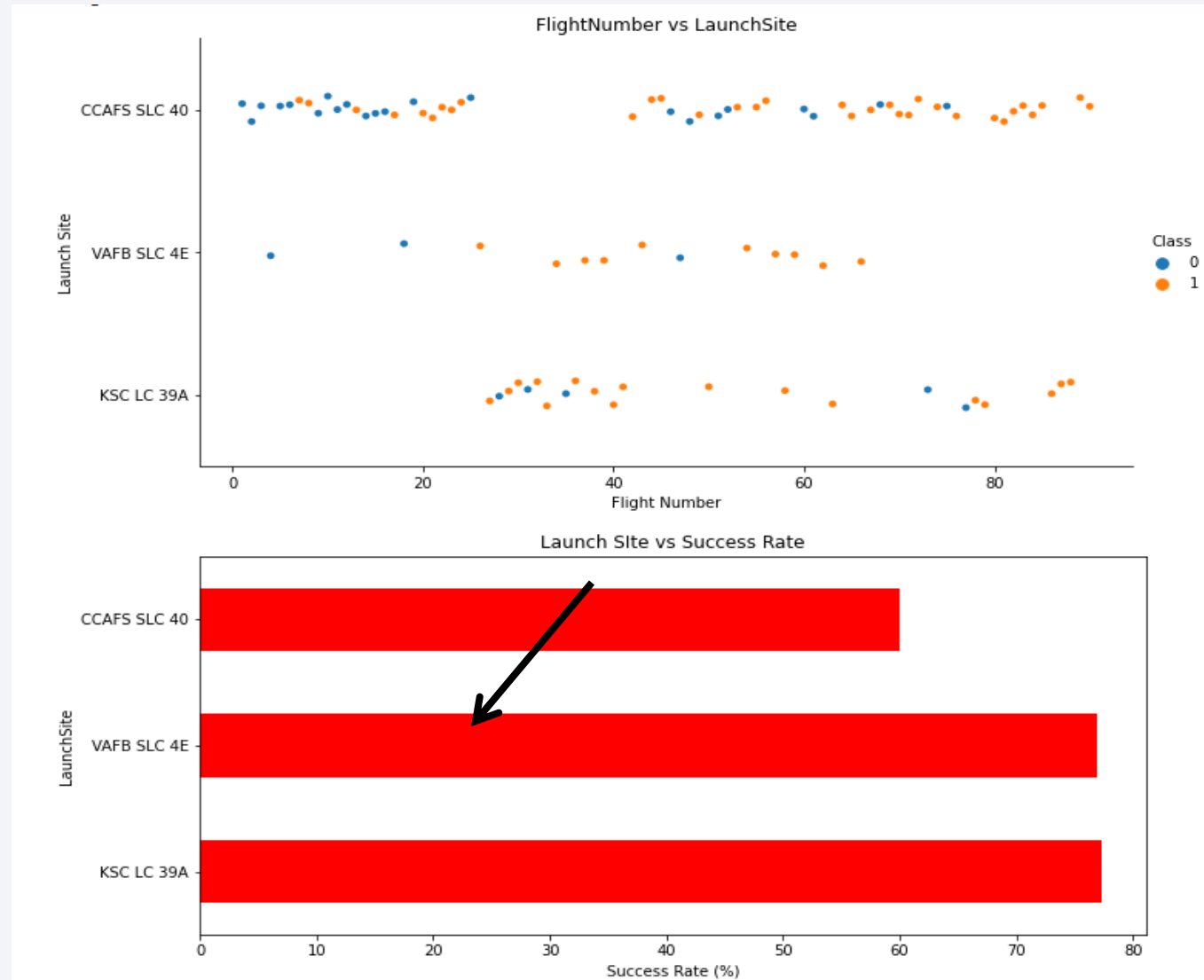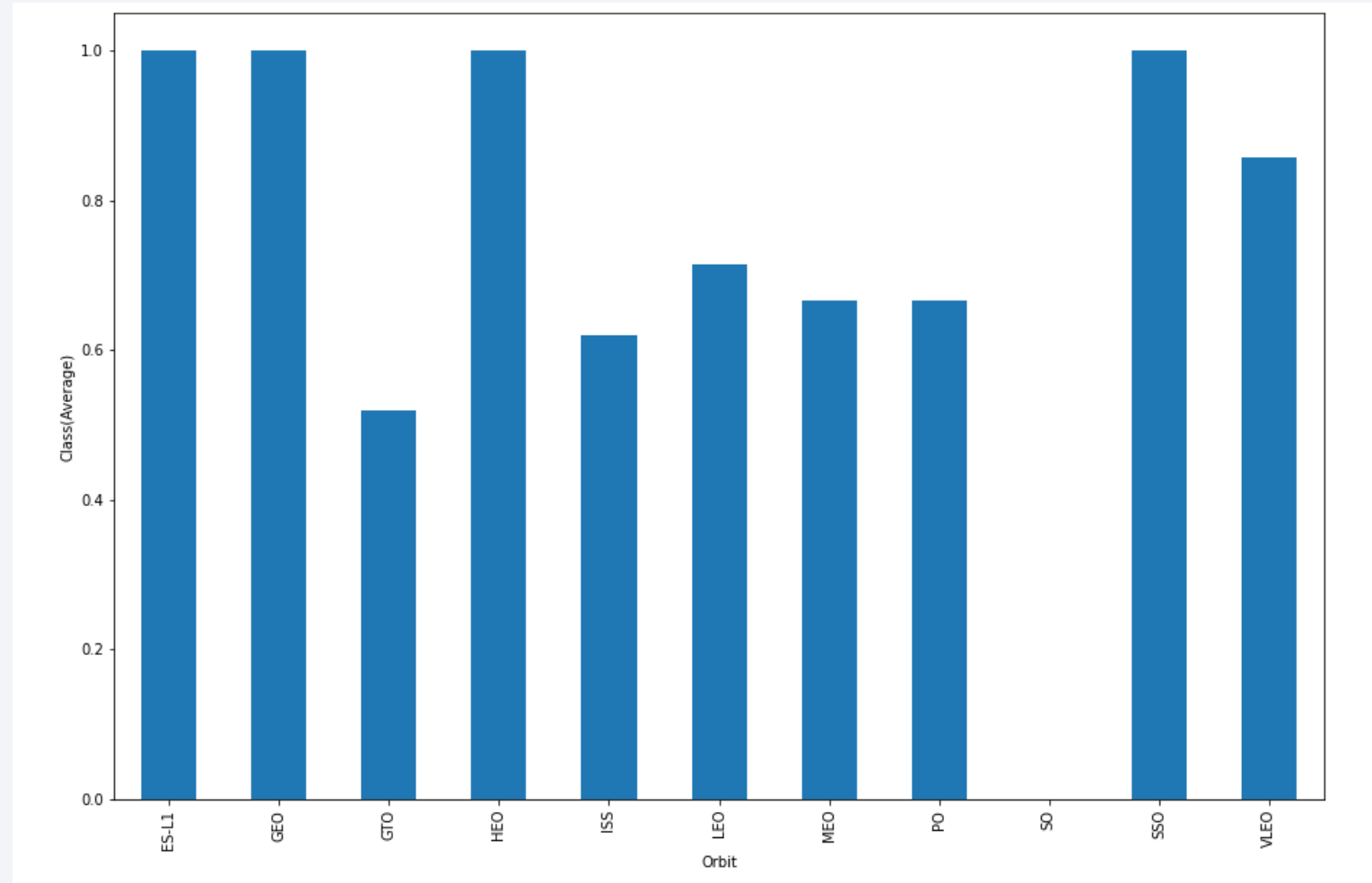
Payload vs. Launch Site

Pay Load Mass vs LaunchSite
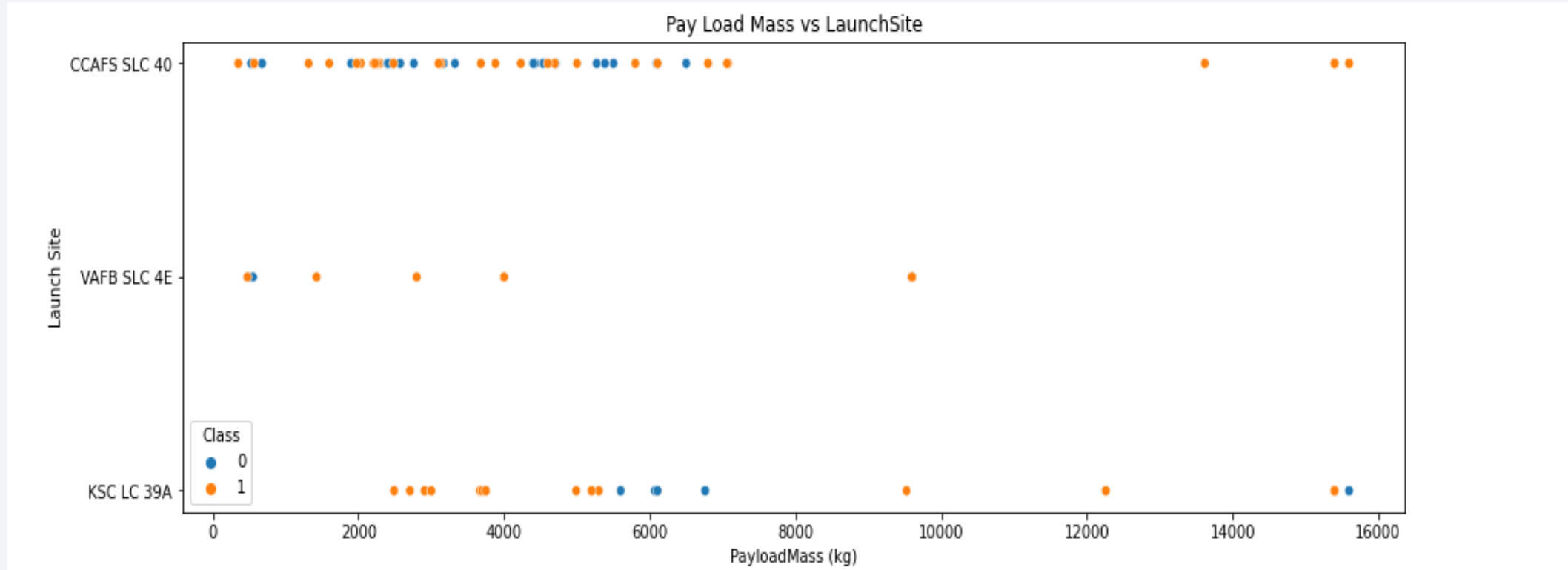
- Now if you observe the scatter point chart, you will find for the VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).
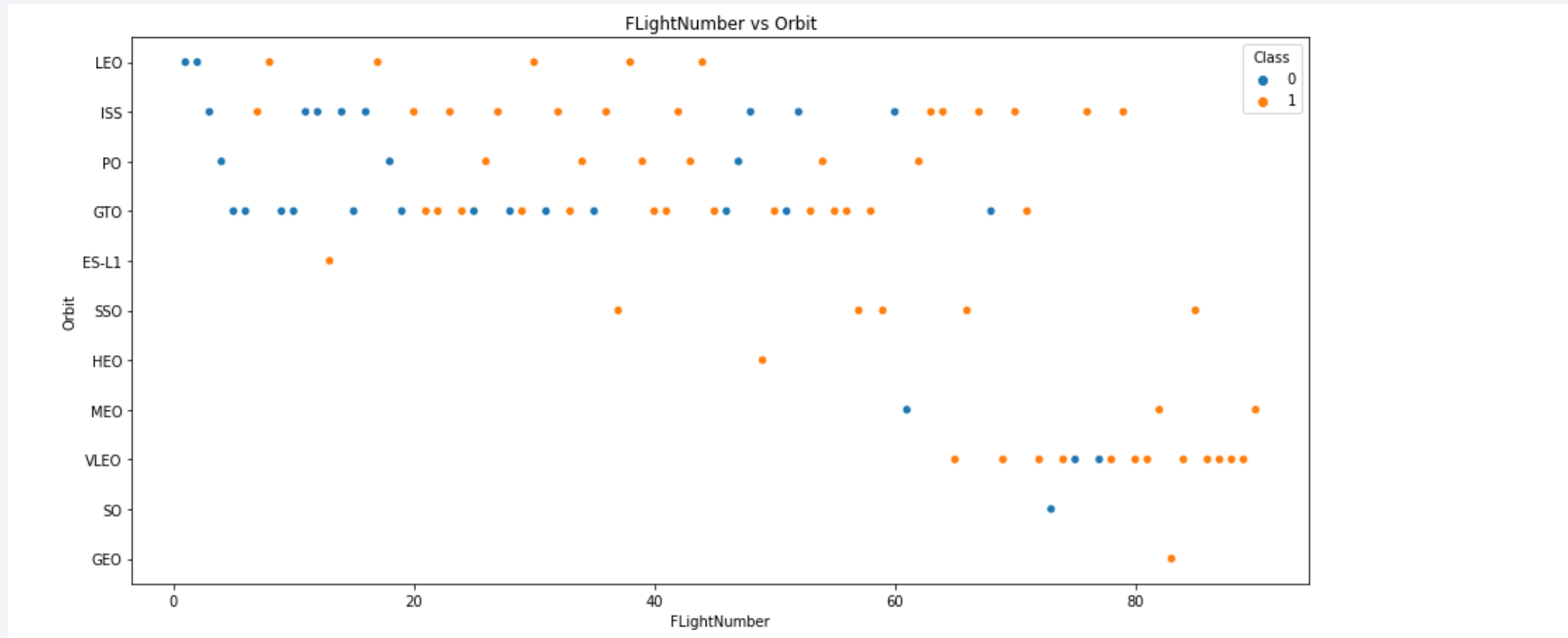
Success Rate vs. Orbit Type

The orbit types **SSO**, **HEO**, **GEO** and **ES-L1** had the highest success rate.
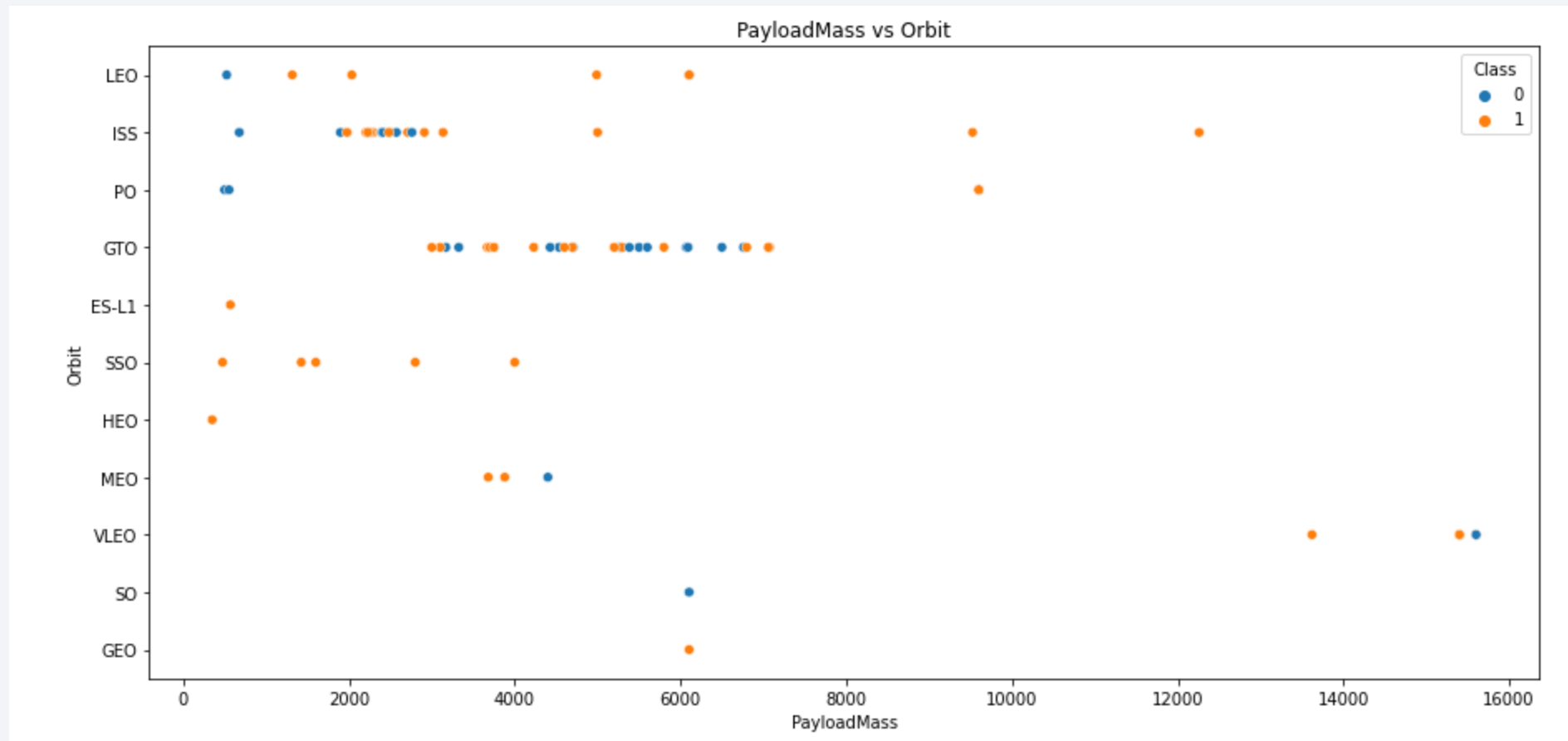
Flight Number vs. Orbit Type

You can see that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.
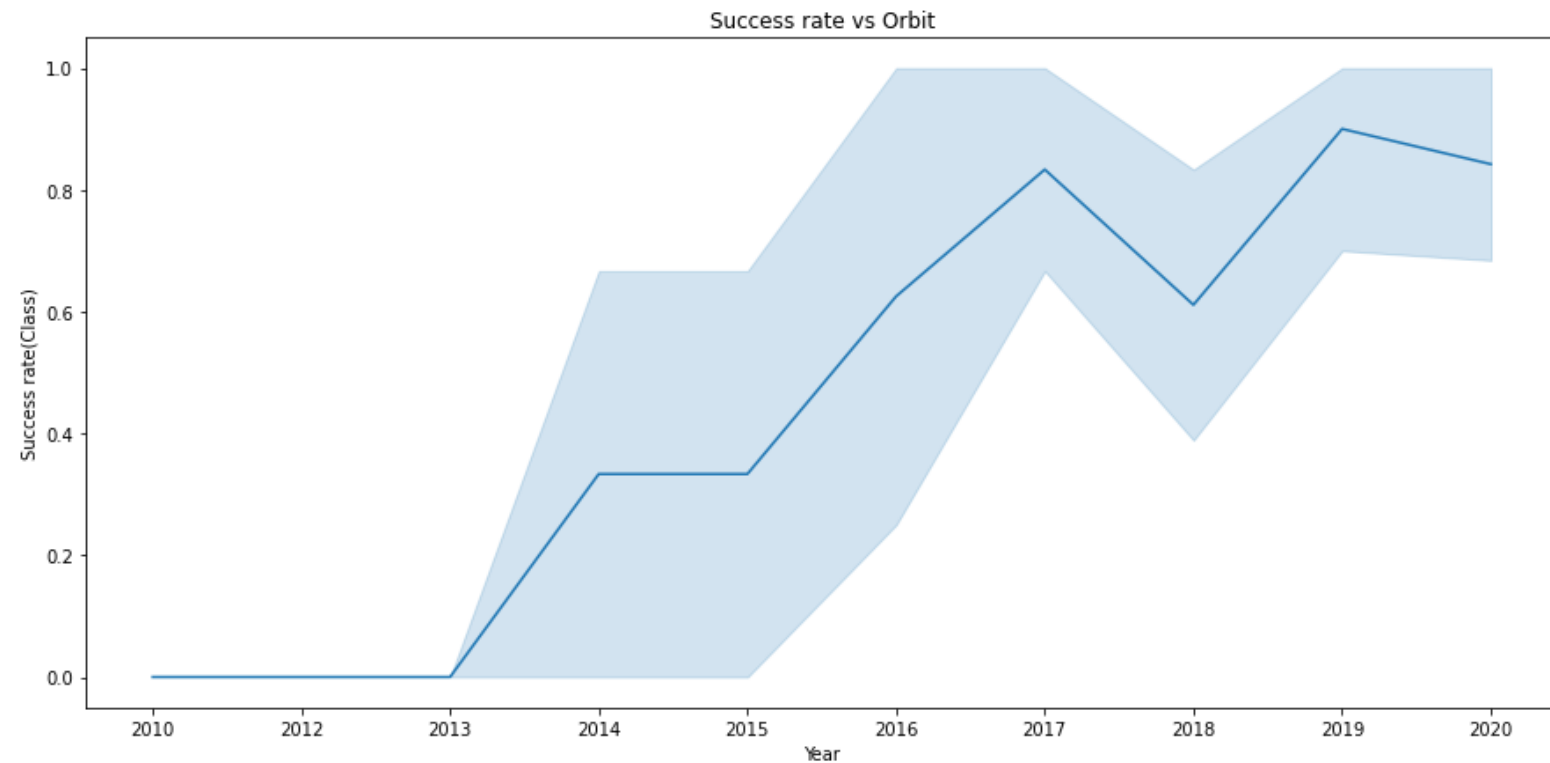
Payload vs. Orbit Type

PayloadMass vs Orbit

- 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.

Launch Success Yearly Trend

It is apparent that the success rate has significantly increased from 2013 to 2020.

All Launch Site Names

Given the data, these are the names of the launch sites where different rocket landings where attempted:

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

Launch Site Names Beginning with 'CCA'

Date	Launch_Site	Orbit	Customer	Mission_Outcome	Landing_Outcome
04-06-2010	CCAFS LC-40	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	CCAFS LC-40	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
22-05-2012	CCAFS LC-40	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	CCAFS LC-40	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	CCAFS LC-40	LEO (ISS)	NASA (CRS)	Success	No attempt

These are 5 records where launch sites begin with the letters 'CCA'. As we can see, there are other organizations besides SpaceX that were testing their rockets.

Total Payload Mass

- The information in the table displays the total payload mass carried by boosters launched by NASA .

- It seems that *NASA (CRS)* had a significantly higher total payload mass compared to the rest.

Customer	Total_Payload_Mass
NASA (CRS)	45596
NASA (CCDev)	12530
NASA (CCP)	12500
NASA (CCD)	12055
NASA (CTS)	12050
NASA (CRS), Kacific 1	2617
NASA / NOAA / ESA / EUMETSAT	1192
NASA (LSP) NOAA CNES	553
NASA (COTS)	525
NASA (LSP)	362
NASA (COTS) NRO	0

Average Payload Mass by F9 v1.1

Average_Payload_Mass (kg)	Booster_Version
2928.4	F9 v1.1

- The average payload mass carried by F9 v1.1 was 2928.4 kg.

First Successful Ground Landing Date

Date	Landing_Outcome
22-12-2015	Success (ground pad)

- The first successful ground pad landing took place in December 2015. This was a historic reusable-rocket milestone for both SpaceX and the world.

- Prior to this, no one had ever brought an orbital class booster back intact.

Successful Drone Ship Landing with Payload between 4000 and 6000

Booster_Version	PAYLOAD_MASS__KG_	Landing_Outcome
F9 FT B1022	4696	Success (drone ship)
F9 FT B1026	4600	Success (drone ship)
F9 FT B1021.2	5300	Success (drone ship)
F9 FT B1031.2	5200	Success (drone ship)

- It appears that there only 4 Boosters with a payload mass between 4000 and 6000.

- It is interesting to see that they all had successful landing outcomes.

Total Number of Successful and Failure Mission Outcomes

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

- It appears that missions generally tend to be successful with the exception of one failure.

Boosters That Carried the Maximum Payload Mass

- 12 boosters have carried the maximum payload mass of 15600 kg.

- Since the version names are similar, they might be from the same manufactures.

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records - Failed Landing Outcomes

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

- It appears that 2 boosters failed to land at the beginnnig of the year..

- The first successful landing took place later that year in December as we saw earlier.

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

- If we observe the table, it is apparent that the number of successful landings have increased since 2015.

- Before 2013, it seems that there were no attempts to land the boosters.

date	Landing_Outcome	Outcomes
2016-04-08	Success (drone ship)	14
2015-12-22	Success (ground pad)	9
2015-06-28	Precluded (drone ship)	1
2015-01-10	Failure (drone ship)	5
2014-04-18	Controlled (ocean)	5
2013-09-29	Uncontrolled (ocean)	2
2012-05-22	No attempt	22

Section 3

Launch Sites Proximities Analysis

Launch Site Locations

- We can see that all launch sites are in very close proximity to the coast and they are also a couple thousand kilometers away from the equator line.

- It is interesting to see that most launch sites are concentrated near Miami.
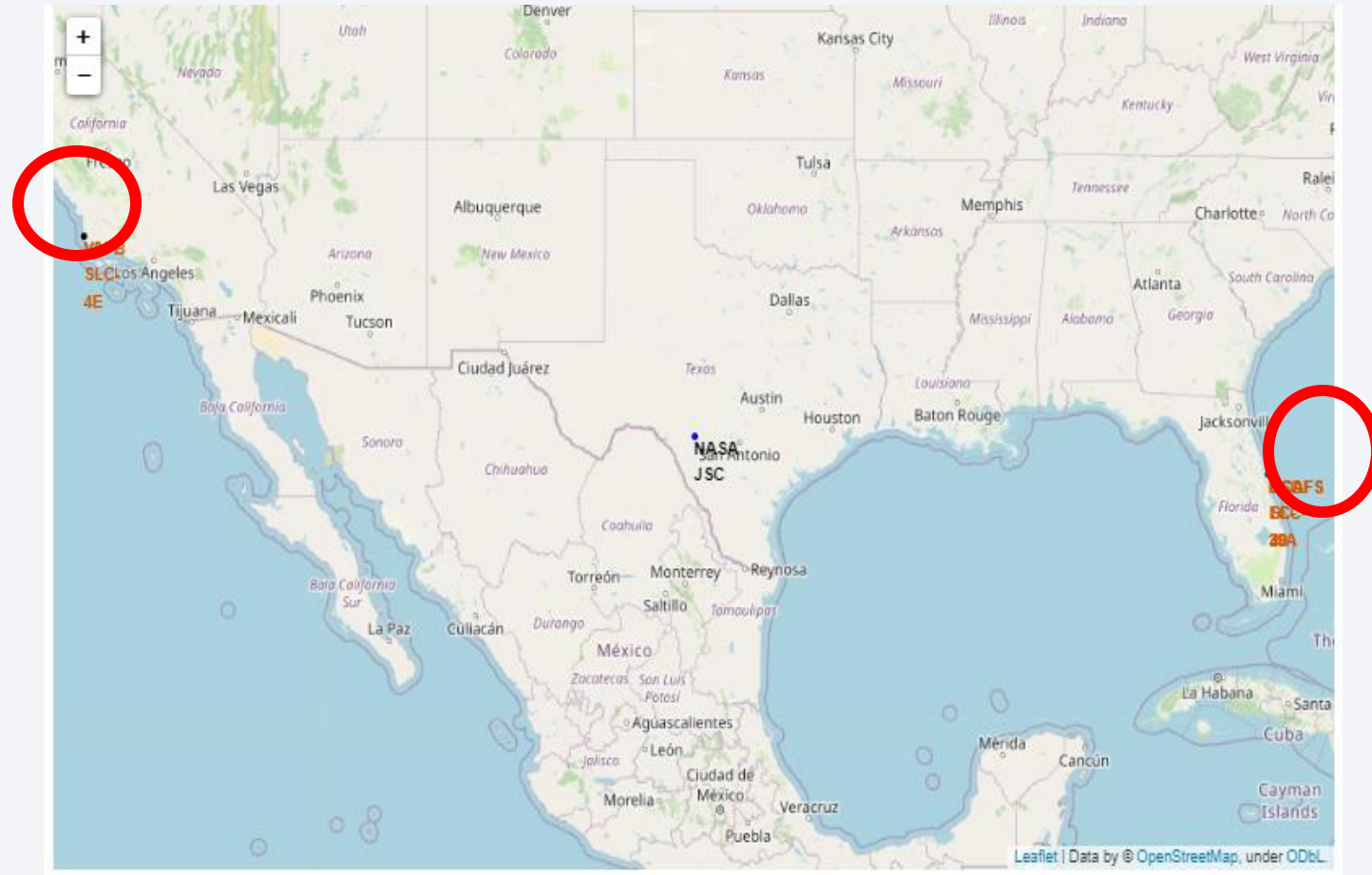
# Success Rate of Rocket Launches

- The successful launches are represented by a green marker while the red marker represents failed rocket launches.

- It appears that **KSC LC-39A** had the highest success rate of rocket launches compared to other launch sites.
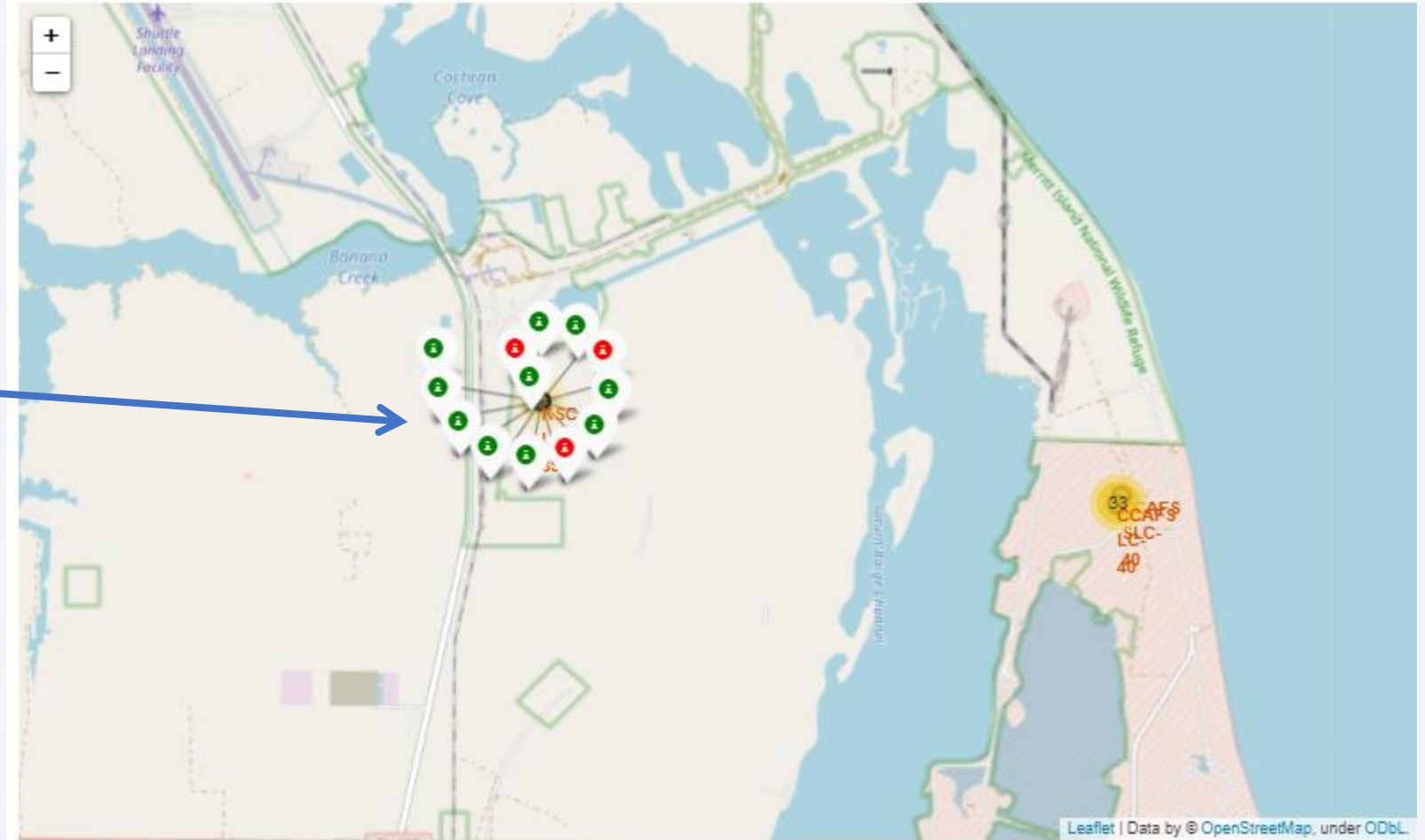
Surrounding Landmarks

- It appears that launch sites are usually set up at least 18 km away from cities. This may be because of the desire to prevent any crashes near populated areas.

- It is also apparent that launch sites are in very close proximity to railways and highways. Perhaps, due to the necessary transportation requirements for rocket parts.

- The sites are close the coast line. This is evident with the many rocket landing tests on water bodies like the ocean.
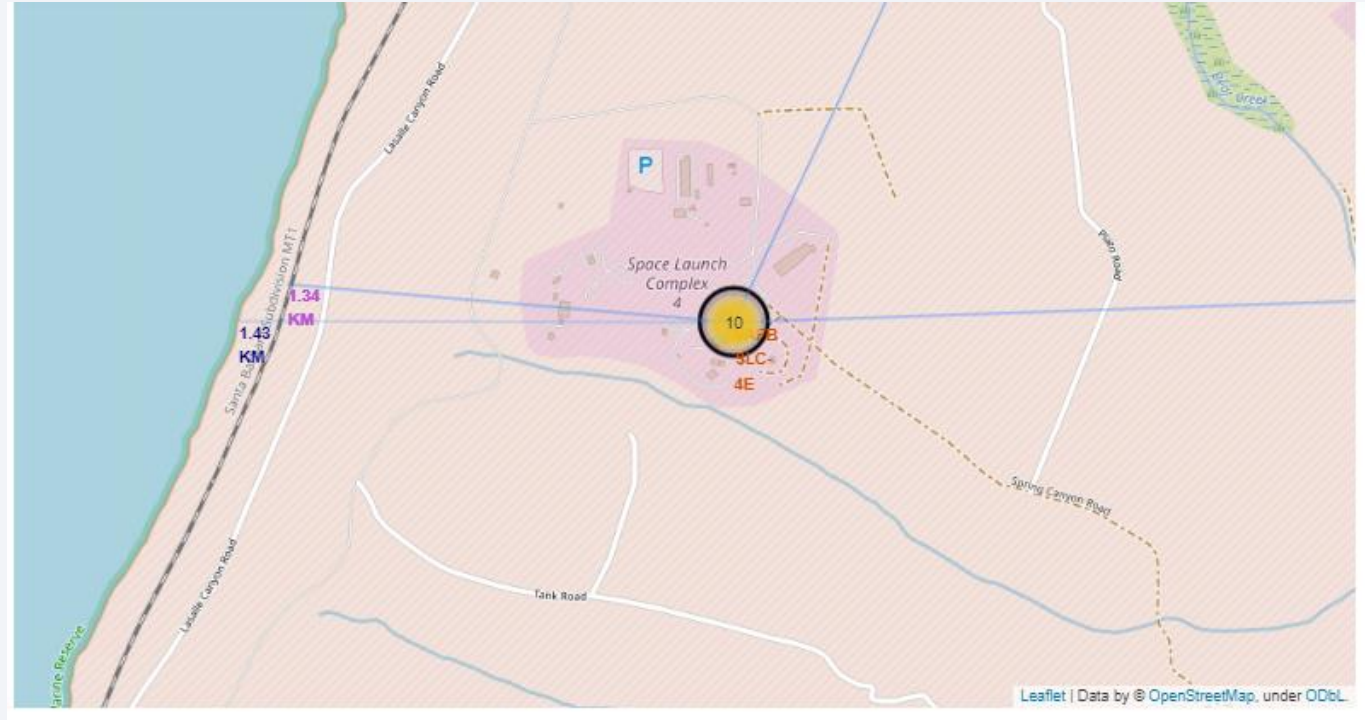
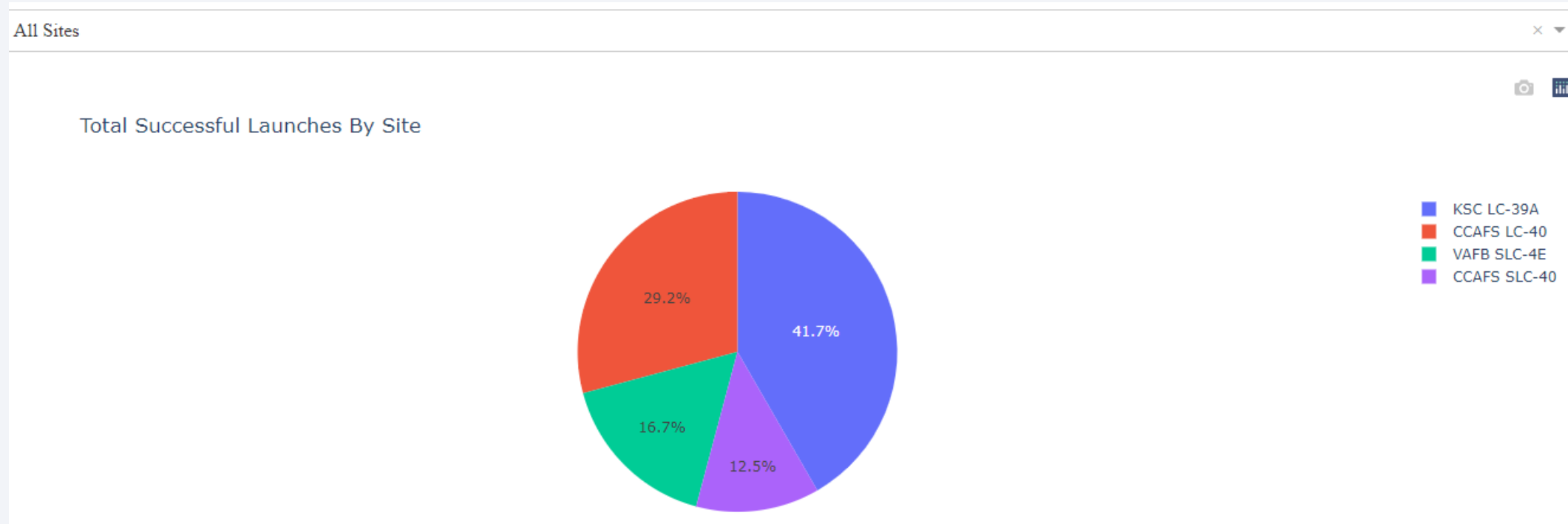| Map Object	Colour
Nearest Highway	Green
Nearest Railway	Purple
Nearest City	Crimson
Nearest Coastline	Dark Blue

Section 4

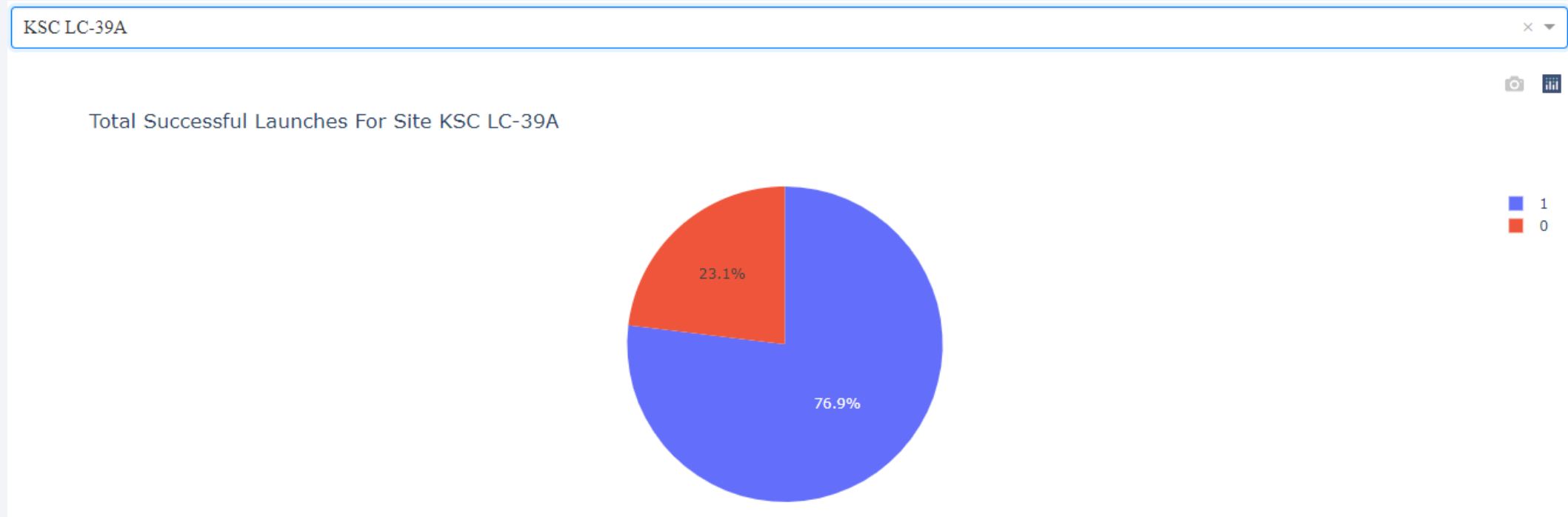# Build a Dashboard with Plotly Dash

Successful Launches by Site

- Site KSC LC-39A has the largest successful launches as well the highest launch success rate.

- More investigation may be needed to determine why KSC LC-39A is the preferred launch site.
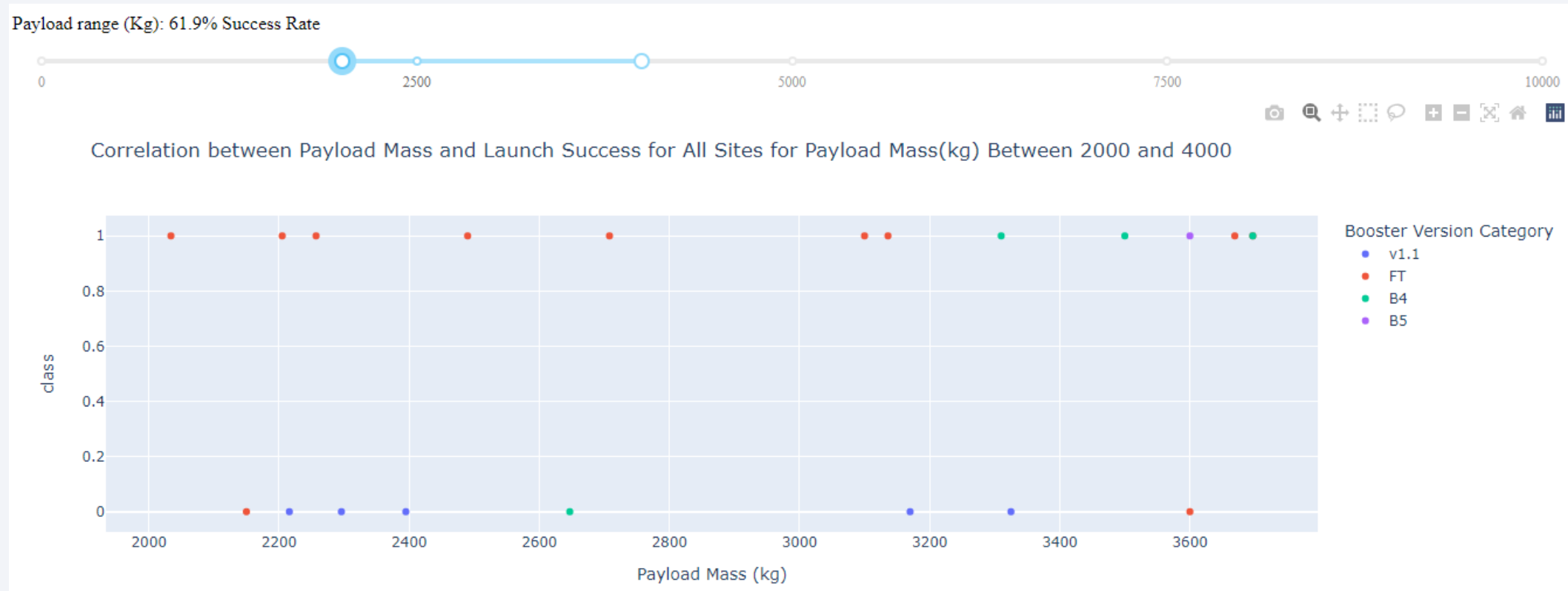
Total Successful Launches for Site KSC LC-39A

- As we can see, 76.9% of the total launches at site KSC LC-39A were successful. This is a the highest success rate of all the different launch sites.

- However, this success rate was only around 3% higher than the runner up; site CCAFS LC-40.

Payload Mass vs. Launch Success for All Sites

- It appears that the payload range between 2000 kg and 4000 kg has the highest success rate.

- The launch success rate was also dramatically low between the payload range of 0kg and 2500kg. Perhaps very low masses decrease launch success.

- The booster version **FT**, seems to have a higher success rate than other booster versions
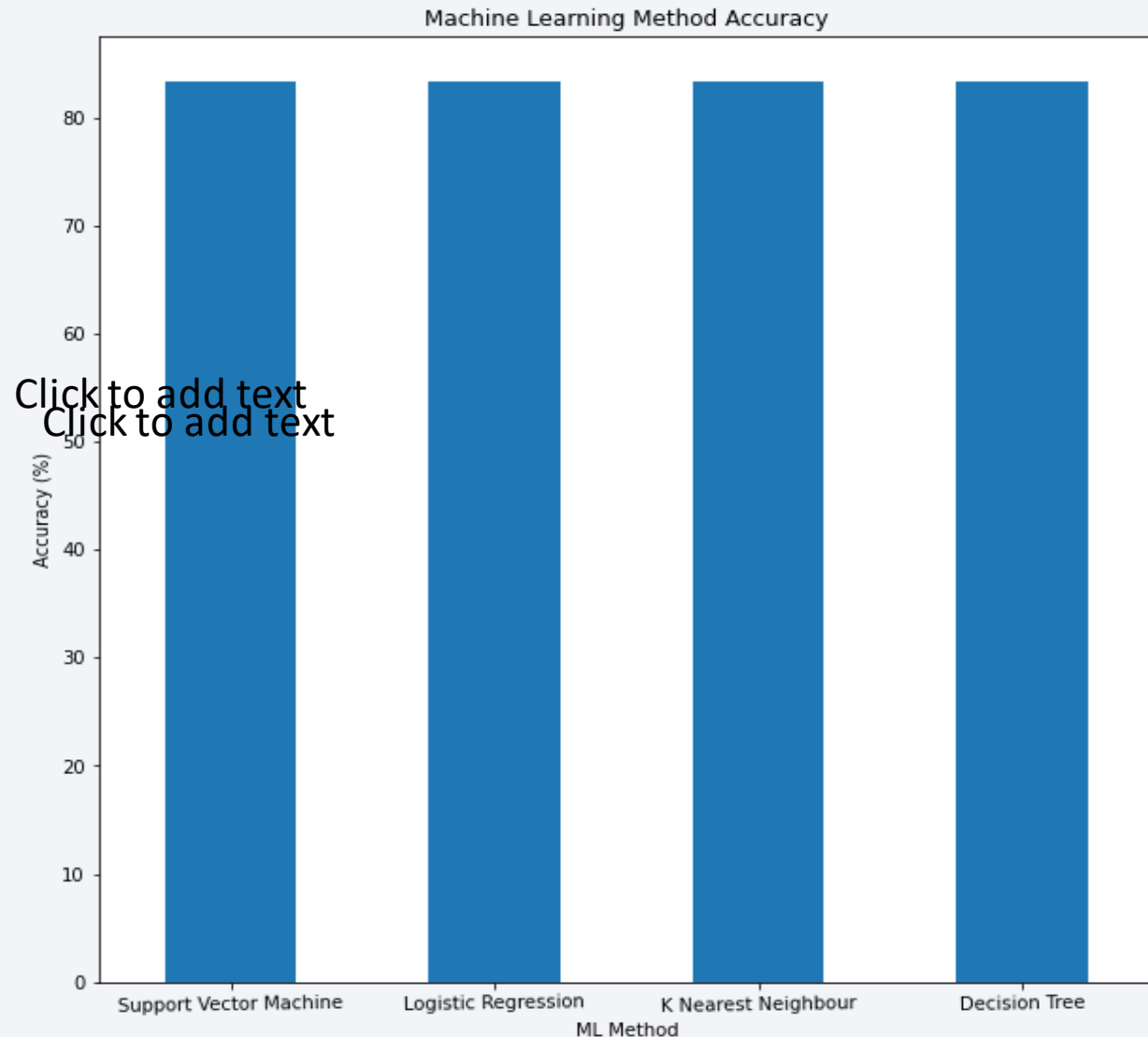
Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Since all the methods have an identical accuracy score of 83.33%, we decided to use Logistic Regression for the classification.

Machine Learning Method Accuracy

Click to add text
Click to add text

Confusion Matrix

- The chart shows the confusion matrix of the Logistic Regression model that was chosen.

- The model only failed to accurately predict 3 labels.

- Model shows the problem with false positive of prediction.
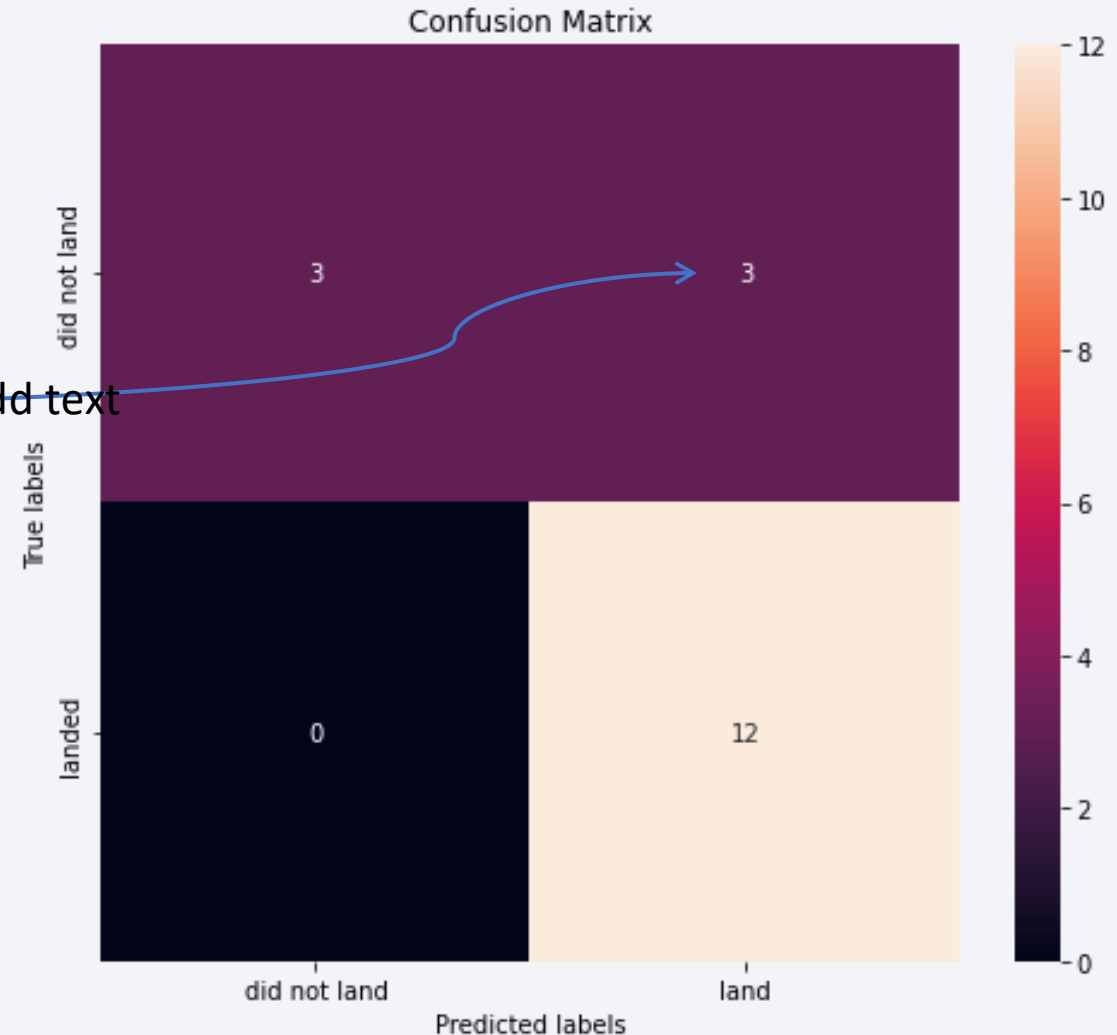
Conclusions

In order to compete with SpaceX, it was crucial to analyze their data. Through this process, a general picture of their success methods was produced.

- All their launch sites are located near the coast, away from nearby cities. This enabled to them to test their rocket landings without much interference.

- Site KSC LC-39A had the highest launch success rate out of all the launch sites.

- From 2015 onwards, the success rate of rocket landings significantly increased. It was also apparent that landing success increased with flight number

All this data was used to train a machine learning model that is able to predict the landing outcome of rocket launches with 83.33% accuracy.

This will allow our company to make more attractive offers than SpaceX and increase the confidence of our investors and customers. Can anyone say "No" to a company that can predict the success of their product?

References

- Fortune Business Insights (2020). *Space launch services market.* https://www.fortunebusinessinsights.com/industry-reports/space-launch-services-market-101931

- CB Insights. *The Top 12 Reasons Startups Fail.* *https://www.cbinsights.com/research/startup-failure-reasons-top/*

- IBM. *Data Science Professional Certificate.* *https://www.coursera.org/professional-certificates/ibm-data-science*

- *Space.com. SpaceX Lands Orbital Rocket Successfully in Historic First.* *https://www.space.com/31420-spacex-rocket-landing-success.html*

Appendix

SQLite Data Set

- The table structure belongs to the SQLite data set used for SQL queries.

Tables (1)

Name	Type	Schema
Spacex		CREATE TABLE "Spacex" ("Date" TEXT, "Time(UTC)" TEXT, "Booster_Version" TEXT, "Launch_Site" TEXT, "Payload" TEXT, "PAYLOAD_MASS__KG_" INTEGER, "Orbit" TEXT, "Customer" TEXT, "Mission_Outcome" TEXT, "Landing_Outcome" TEXT)
Date	TEXT	"Date" TEXT
Time(UTC)	TEXT	"Time(UTC)" TEXT
Booster_Version	TEXT	"Booster_Version" TEXT
Launch_Site	TEXT	"Launch_Site" TEXT
Payload	TEXT	"Payload" TEXT
PAYLOAD_MASS__KG_	INTEGER	"PAYLOAD_MASS__KG_" INTEGER
Orbit	TEXT	"Orbit" TEXT
Customer	TEXT	"Customer" TEXT
Mission_Outcome	TEXT	"Mission_Outcome" TEXT
Landing_Outcome	TEXT	"Landing_Outcome" TEXT

Thank you!