IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Muhammad Imtiaz
09-09-2023

Outline

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

Executive Summary

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.

Value

The predictive models enabled the company to determine the success of a rocket landing. This was especially important, because a failed landing could result in
a loss of tens of millions of dollars. Furthermore, it enabled the company to determine costs and make more attractive offers than SpaceX.

Final Thoughts and Next Steps

Powerful Analytics and Machine Learning tools not only increase our competitiveness – they can also increase customer confidence in our offerings. With the
attention of these offerings, we expect to break into a market projected to reach USD 26.16 billion by 2027. For more information, visit the full project repository:
https://github.com/25911/SpaceX_Landing_Prediction_Project_Final

Introduction

The primary objective of this Data Science project is to allow the company to compete with SpaceX. In order to achieve this goal, it is necessary to predict if the first stage of the SpaceX Falcon 9 rocket will land successfully.

SpaceX 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 SpaceX can reuse the first stage.

Therefore if we can accurately predict the likelihood of the first stage rocket landing successfully, we can determine the cost of a launch. With the help of the Data Science findings and models, the company can make more informed bids against SpaceX for a rocket launch.

Section 1

Methodology

Methodology

Executive Summary

• Make requests to the SpaceX API.

• Perform web scraping to collect Falcon 9 historical launch records on the Wikipedia page titled: List of

Falcon 9 and Falcon Heavy launches

• Perform data wrangling

• Clean the data and explore it to find patterns in the data to determine the labels for training supervised

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 given the data.

• Train the best performing model to make accurate predictions.

Methodology

Data Collection

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

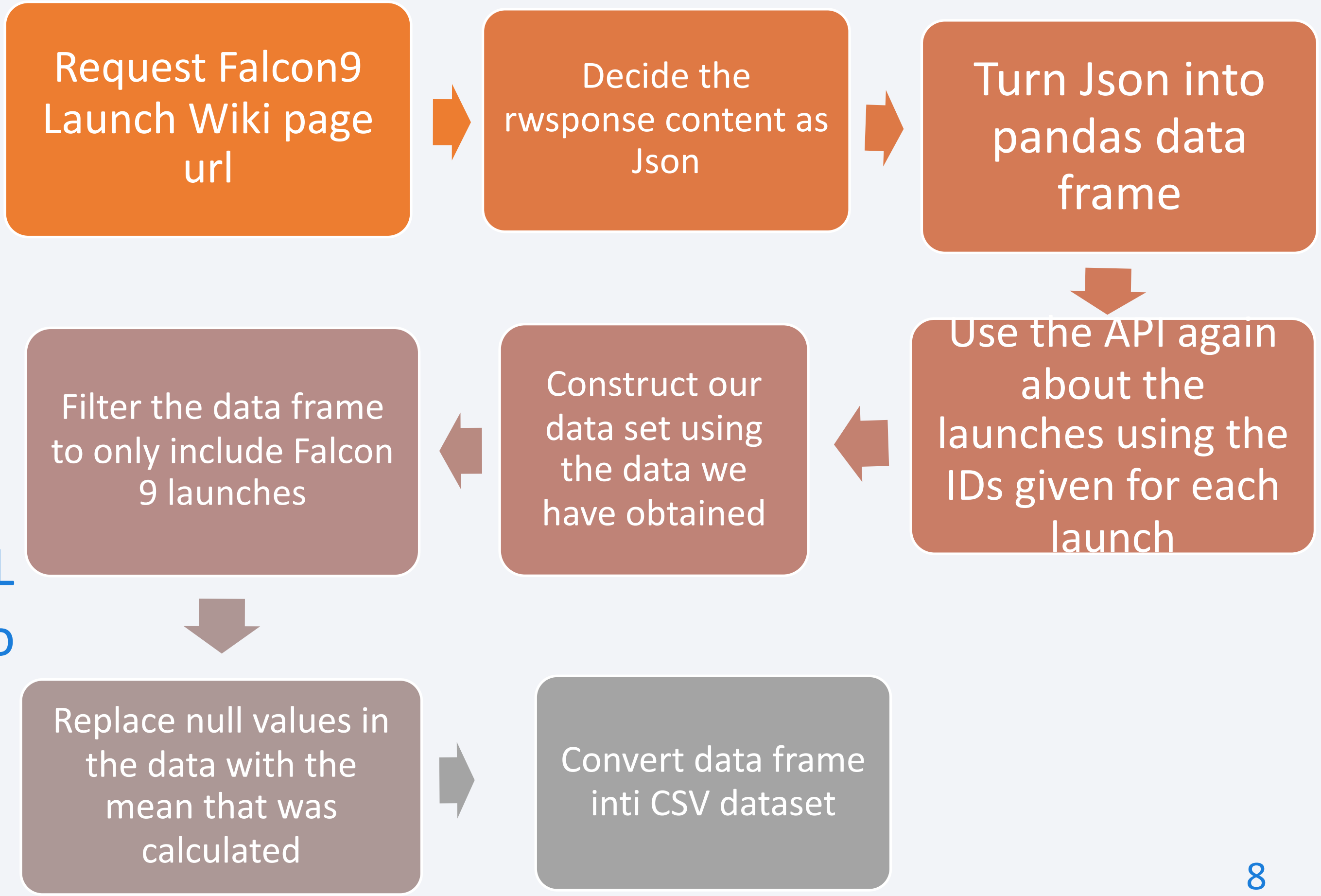
There are different ways to go about collecting the data but we used two methods:

- Data collection by SpaceX API request.
- Data collection by Web Scraping

Data Collection

Data Collection – SpaceX API

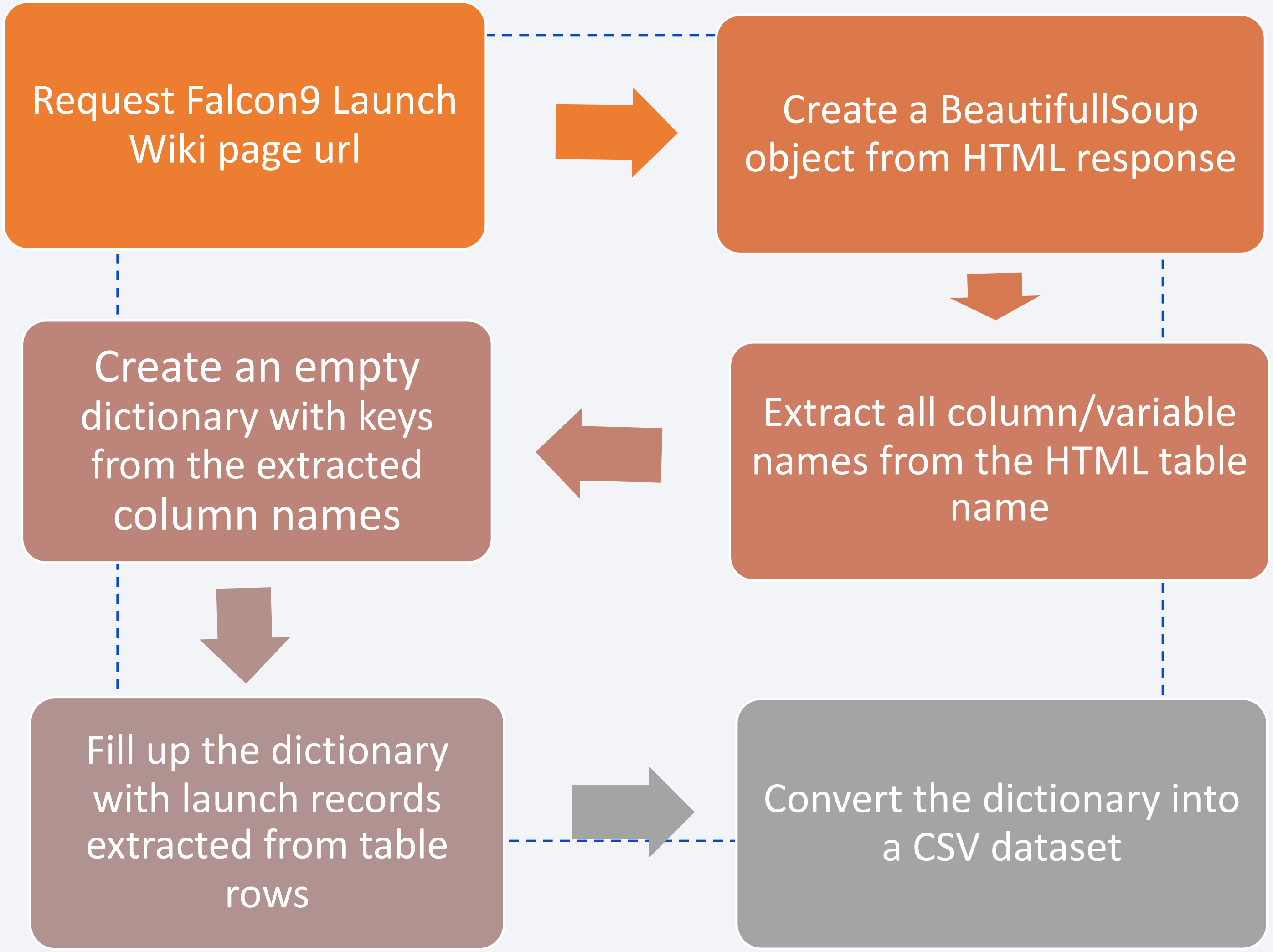
- Present your data collection with SpaceX REST calls using key phrases and flowcharts
- Add the GitHub URL of the completed SpaceX API calls notebook https://github.com/25911/SpaceX_Landing_Prediction_Project_Final/blob/main/Spacex-data-collection-api.ipynb



Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

- GitHub URL
- https://github.com/25911/SpaceX_Landing_Prediction_Project_Final/blob/main/Spacex-webscraping.ipynb



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.

Determine the number of launches on each facility

Calculate the number and occurrences of each orbit

Calculate the number and occurrences 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.

https://github.com/25911/SpaceX_Landing_Prediction_Proj

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.

https://github.com/25911/SpaceX_Landing_Prediction_Project_Final/blob/main/Spacex- Data-Visualization.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

https://github.com/25911/SpaceX_Landing_Prediction_Project_Final/blob/main/SpaceX_EDA_SQL.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.

https://github.com/25911/SpaceX_Landing_Prediction_Project_Final/blob/main/Spacex-interactive-map-with-folium.ipynb

Build a Dashboard with Plotly Dash

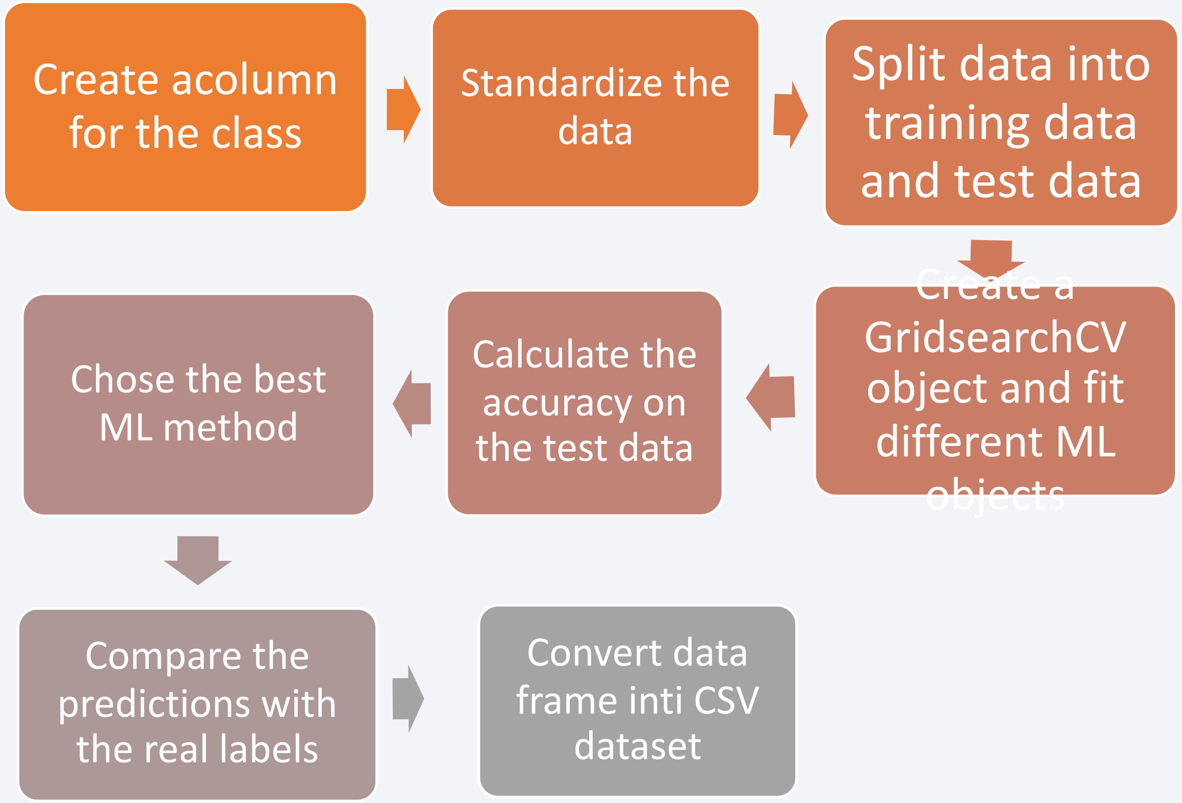
- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

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%

https://github.com/25911/SpaceX_Landing_Prediction_Project_Final/blob/main/Spacex_Machine_Learning_Prediction.ipynb



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.

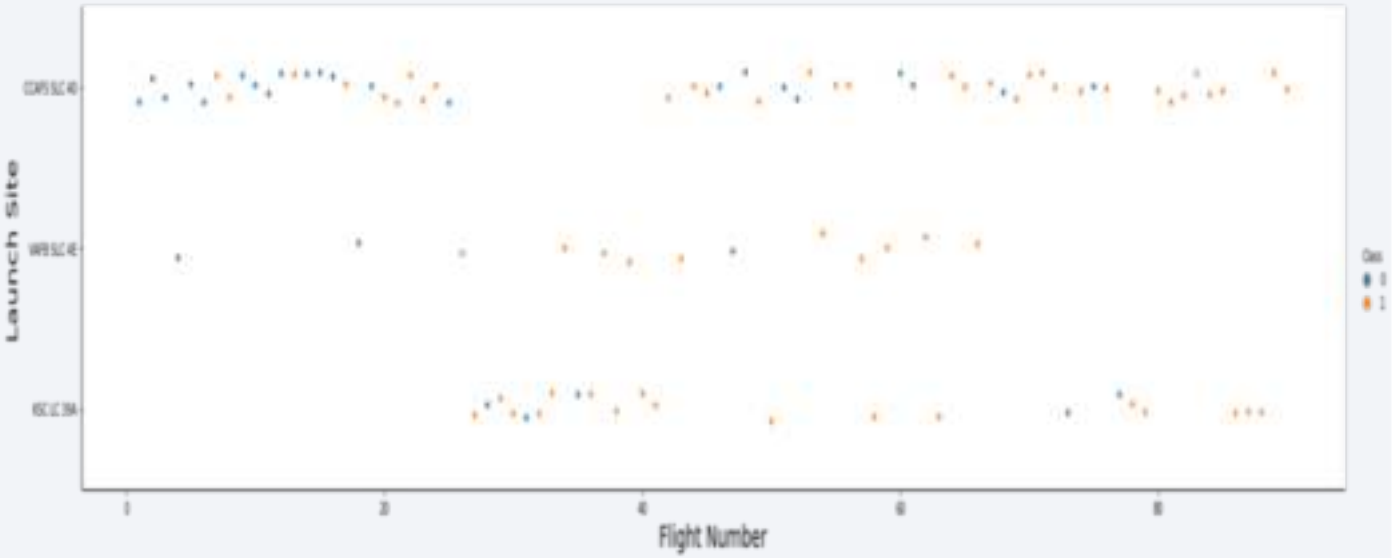
- The machine learning models that were built, were able to predict the landing success of rockets with an accuracy score of 83.33%. This accuracy can be increased in future projects with more data.

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 SLC4E had the least number of attempts

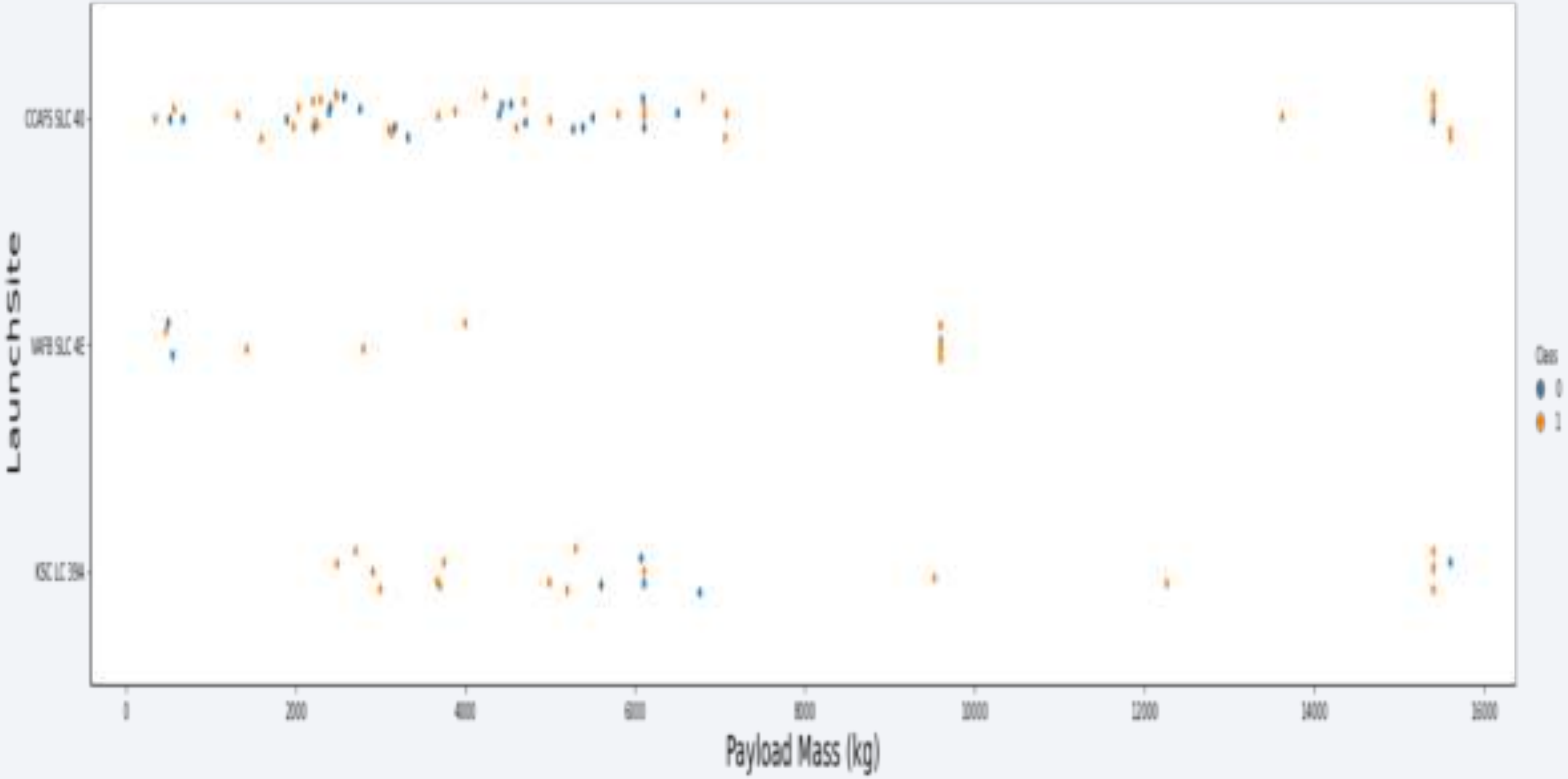


- Looking at the second chart, we can see that there is no Launch Site with a success rate below 60%.

Payload vs. Launch Site

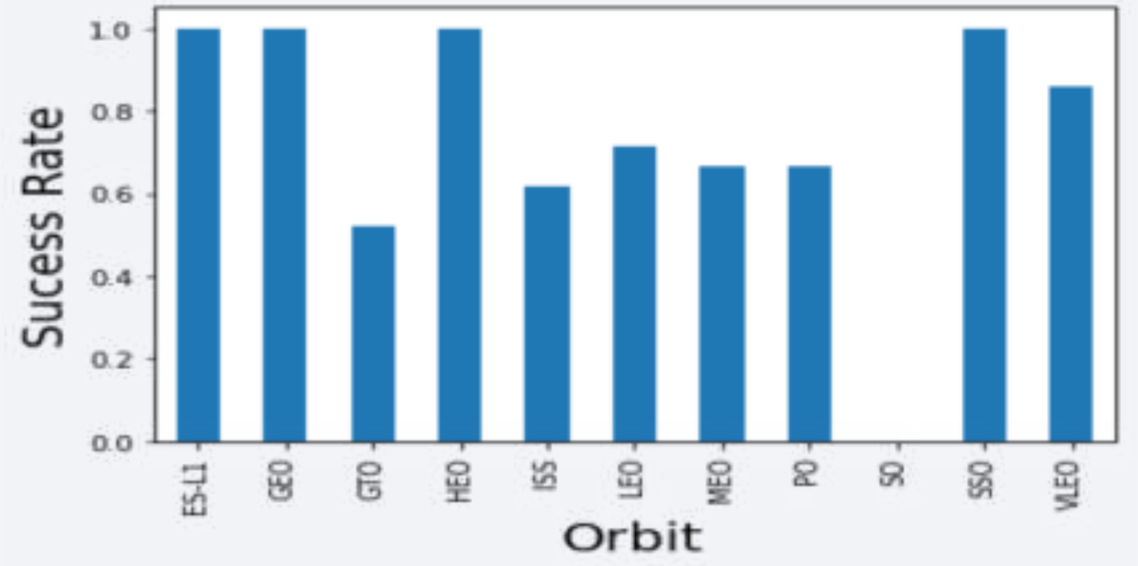
- Show a scatter plot of Payload vs. Launch Site

- Show the screenshot of the scatter plot with explanations

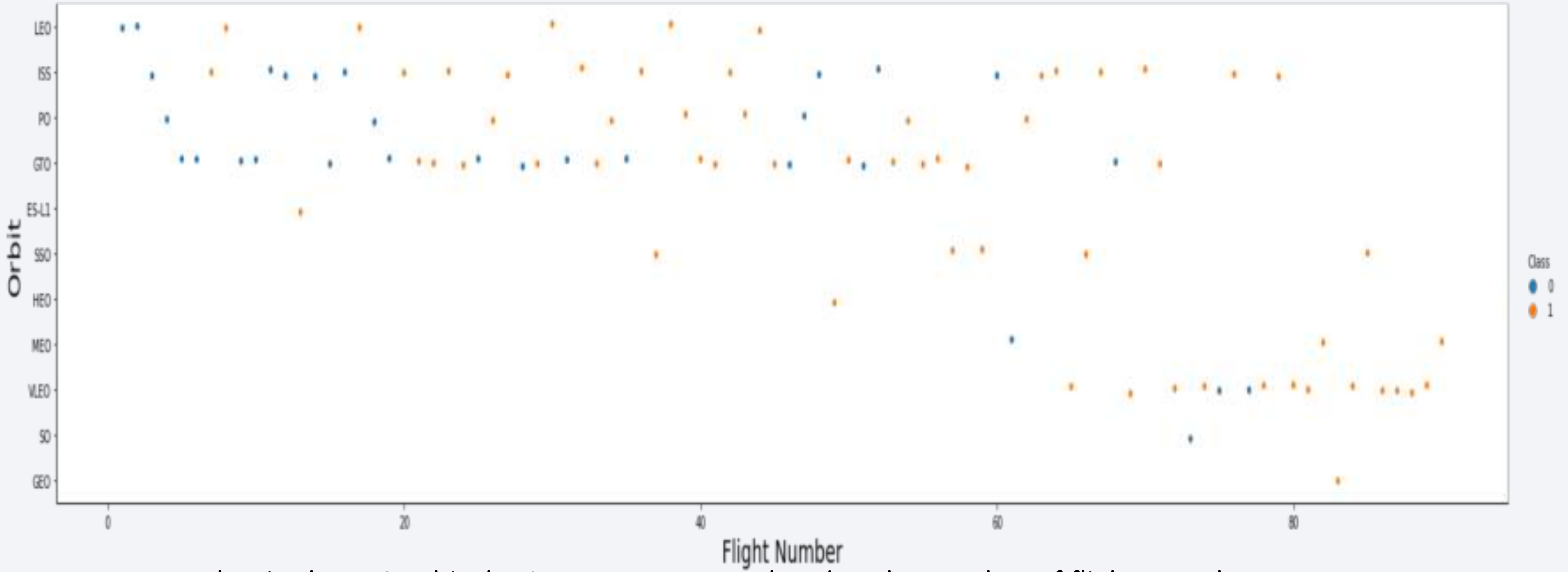


Success Rate vs. Orbit Type

The orbit types SSO, HEO, GEO and ES -L1 had the highest success rate.

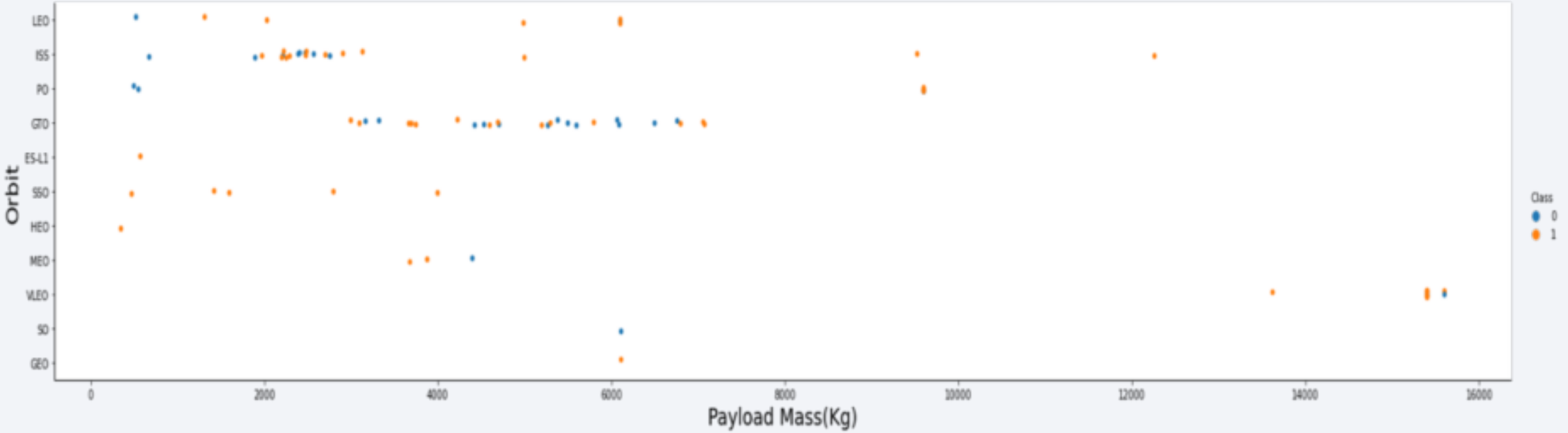


Flight Number vs. Orbit Type



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.

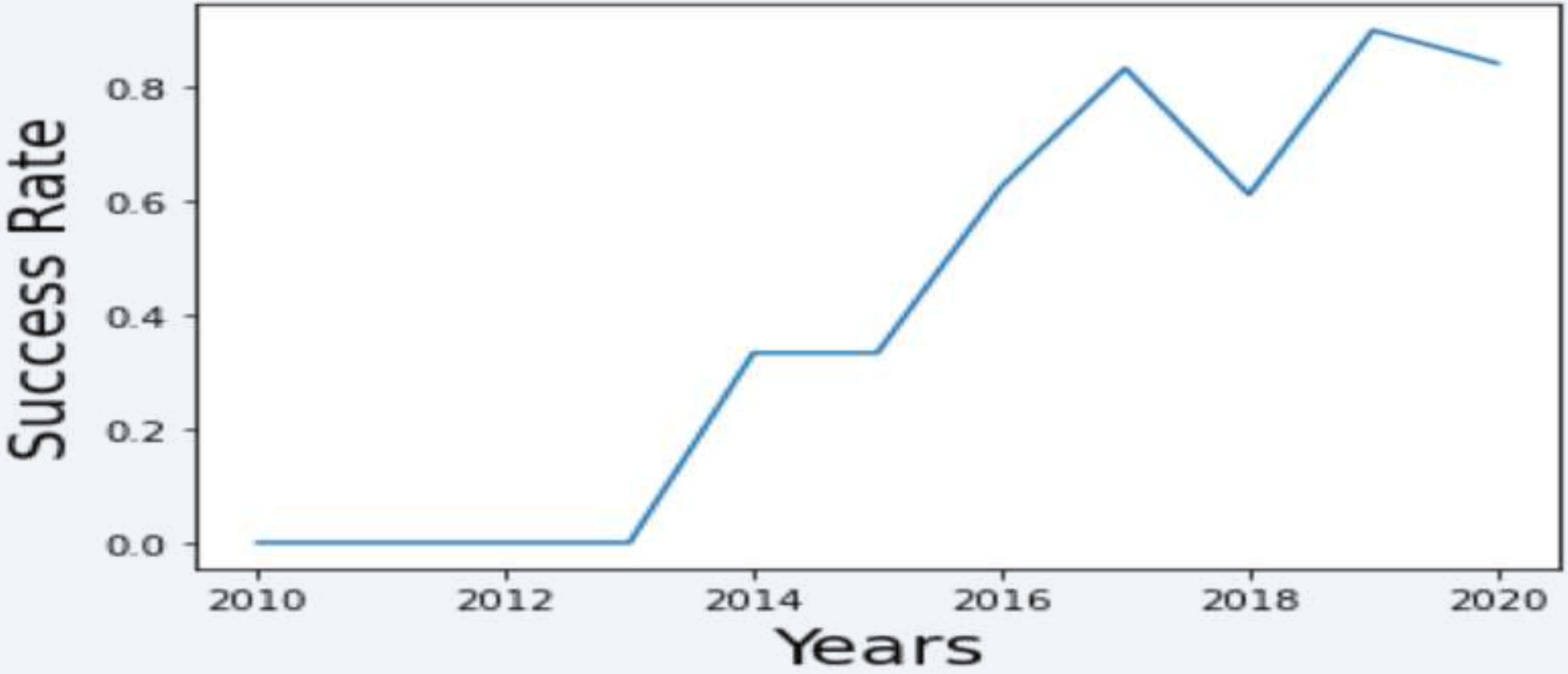
Payload vs. Orbit Type



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 Begin with 'KSC'

Date	Time (UTC)	BoosterVersion	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2017-03-16	06:00:00	F9 FT B1030	KSC LC-39A	EchoStar 23	5600	GTO	EchoStar	Success	No attempt
2017-03-30	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
2017-01-05	11:15:00	F9 FT B1032.1	KSC LC-39A	NROL-76	5300	LEO	NRO	Success	Success (ground pad)
2017-05-15	23:21:00	F9 FT B1034	KSC LC-39A	Inmarsat-5 F4	6070	GTO	Inmarsat	Success	No attempt

Total Payload Mass

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	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

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

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 Carried Maximum Payload

- 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

- List the records which will display the month names, succesful landing_outcomes in ground pad ,booster versions, launch_site for the months in year 2017

- Present your query result with a short explanation here

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

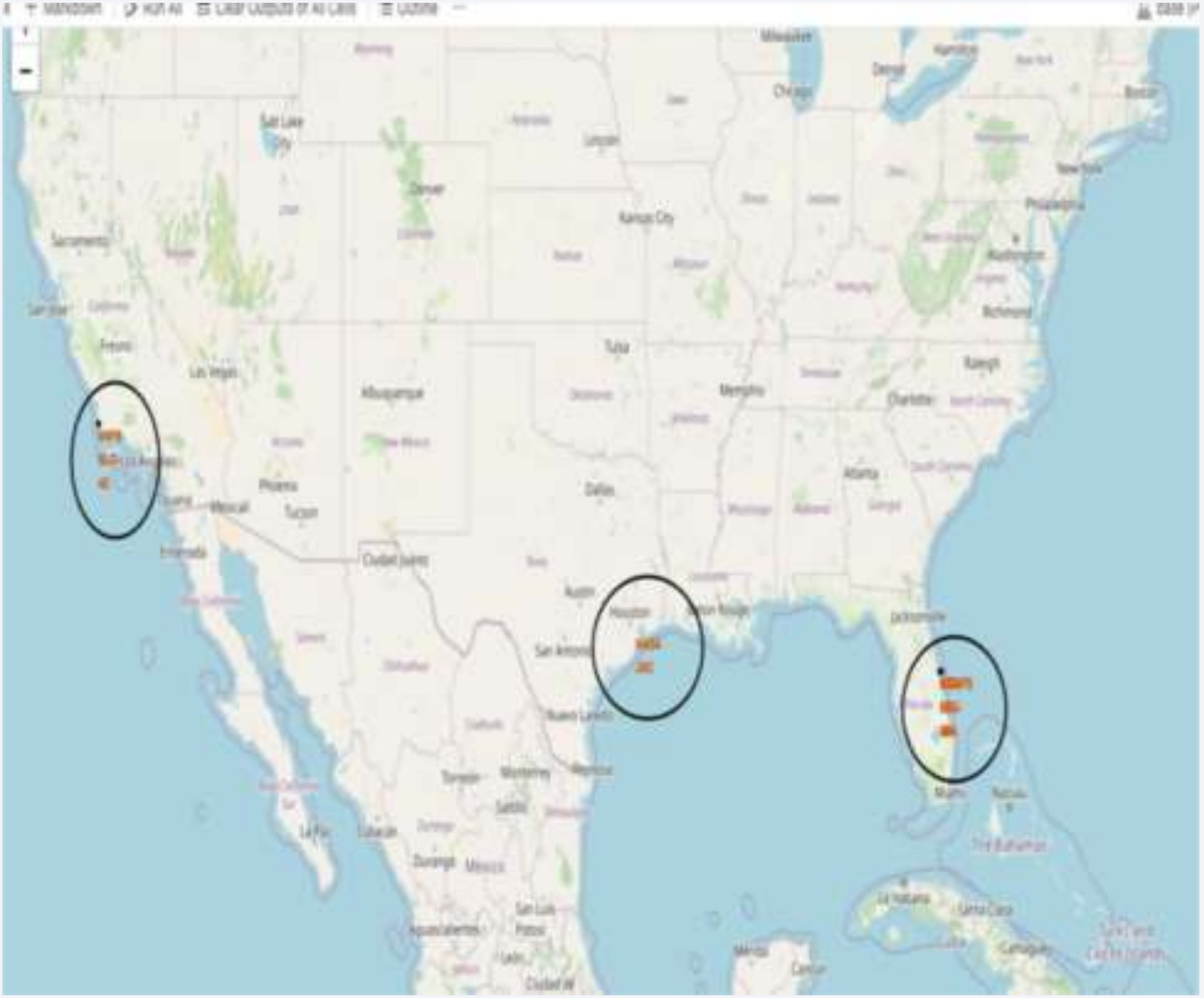
Section 3

Launch Sites Proximities Analysis

Launch Site Location

- 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.



Surrounding Landmarks

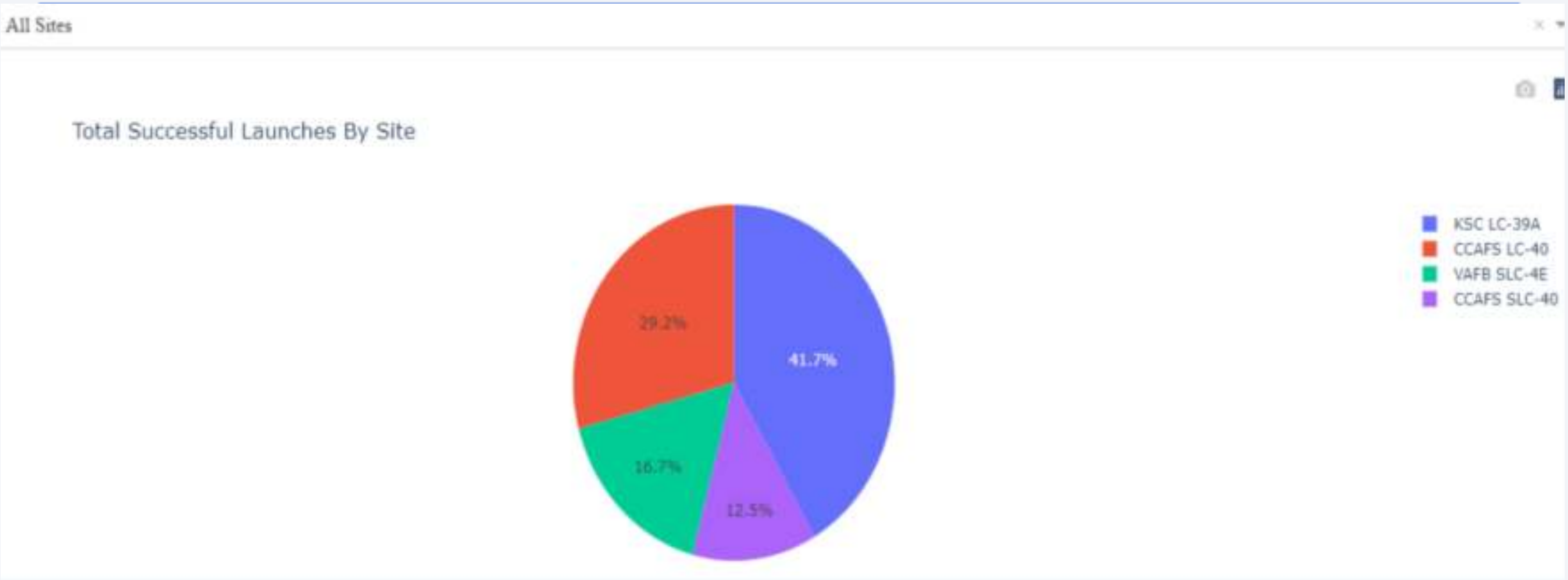
It appears that launch sites are usually set up at least 26 km away from cities. This may be because of the desire to prevent any crashes near populated areas.

- 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

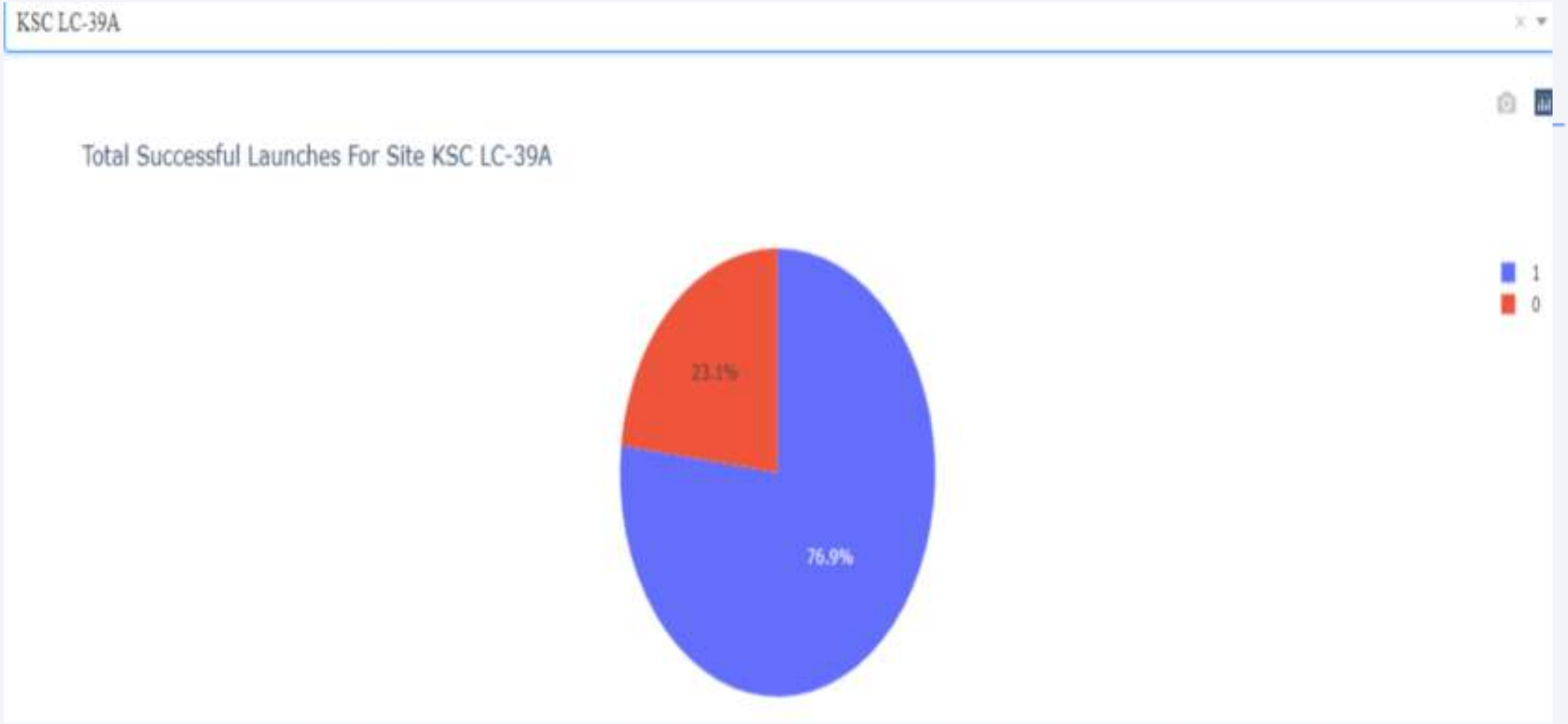
.

Section 4

Build a Dashboard with Plotly Dash



- 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.



- 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.

Section 5

Predictive Analysis (Classification)

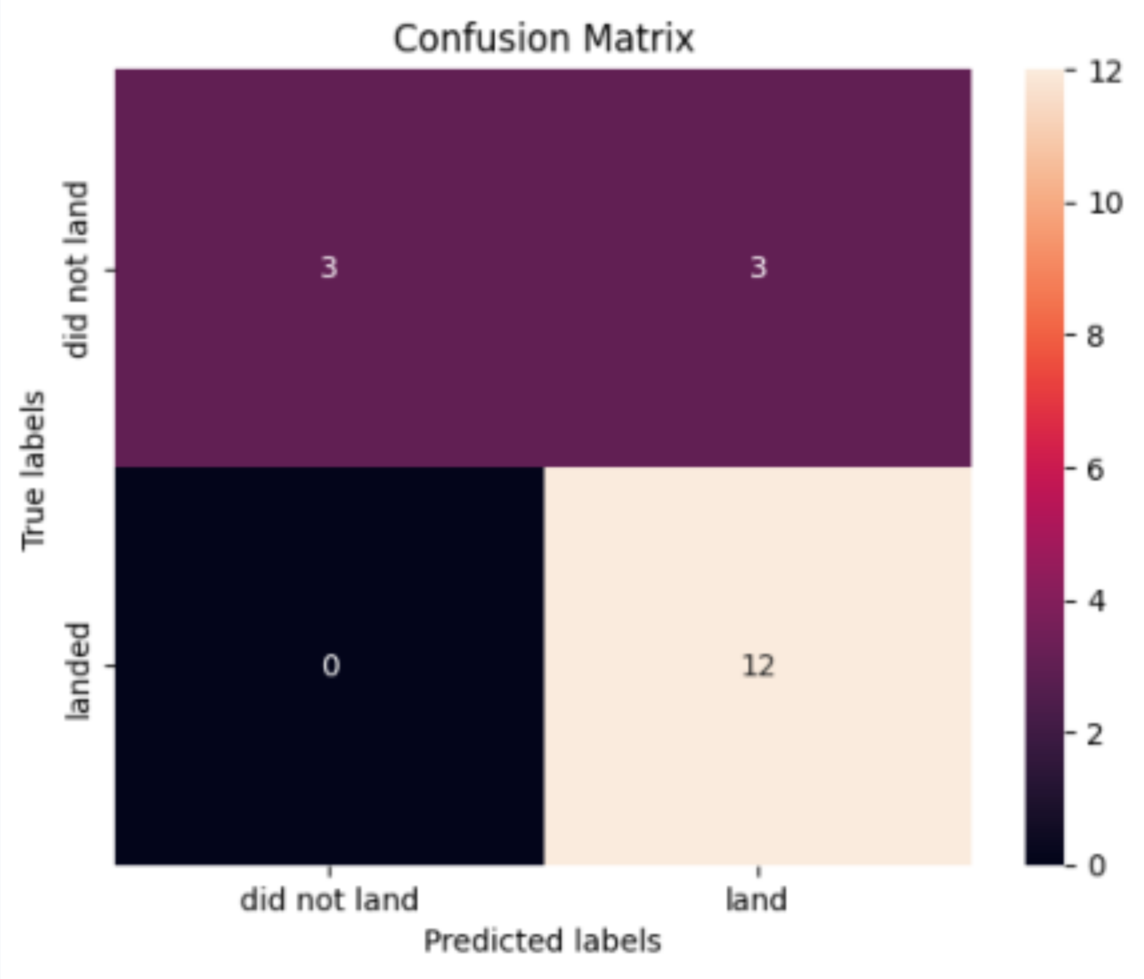
Classification Accuracy

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

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.



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?

Thank you!