IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

By BULENDA MWIMANIRE Albert

1st October 2021

SUMMARY

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

Executive Summary

SpaceX, the Allon Mask private space company has gained worldwide attention for a series of historic milestones.

It is the only private company ever to return a spacecraft from low-earth orbit, which it first accomplished in December 2010,

On 30 May 2020, SpaceX, successfully launched two NASA astronauts Bob Behnken and Doug Hurley into Orbit on a crew Dragon spacecraft during crew Demo-2. This was an incredible challenge for a private start-up, but they have done that because they were able to use data science methods and apply various Machine learning technics to predict with high confidence the mission outcome.

Introduction

SpaceX accomplishment include sending aircraft to the international Space Station for different purpose include but not limited to satellite internet access and satellite imagery. The SpaceX rocket are likely the most unexpansive. 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.

In this project, SpaceY will try to determine whether the parameter set in place are ideal for the success of the rocket launch mission, as well as the coast of the launch by providing a complete dashboard to the SpaceY Management department.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium
- Perform predictive analysis using classification models

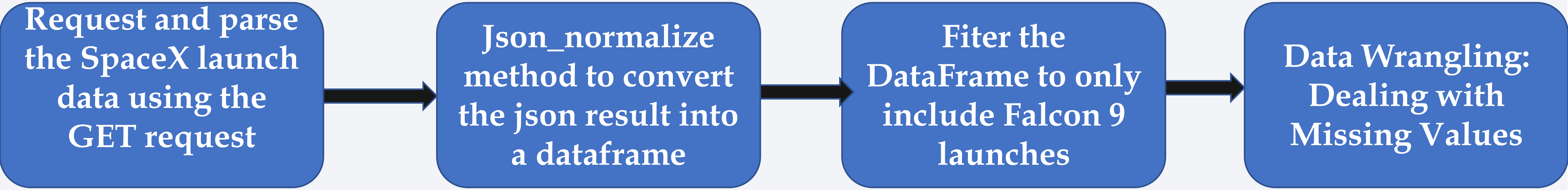
Data Collection

In order to determine the landing outcome of the rocket launch for SpaceY we will use historical dataset from the SpaceX company provided on the website using an API, specifically the SpaceX REST API which give us information about the Delivered Payload Mass, Launch Specifications, Landing Specification and Landing Outcome.

Data Collection - SpaceX API

In this section, we will make a get request to the SpaceX API. We will also do some basic data wrangling and formatting:

- Request to the SpaceX API
- Clean the requested data



Data Collection - Scraping

Objectives:

Web scrap Falcon 9 launch records with BeautifulSoup:

- Extract a Falcon 9 launch records HTML table from Wikipedia
- Parse the table and convert it into a Pandas data frame

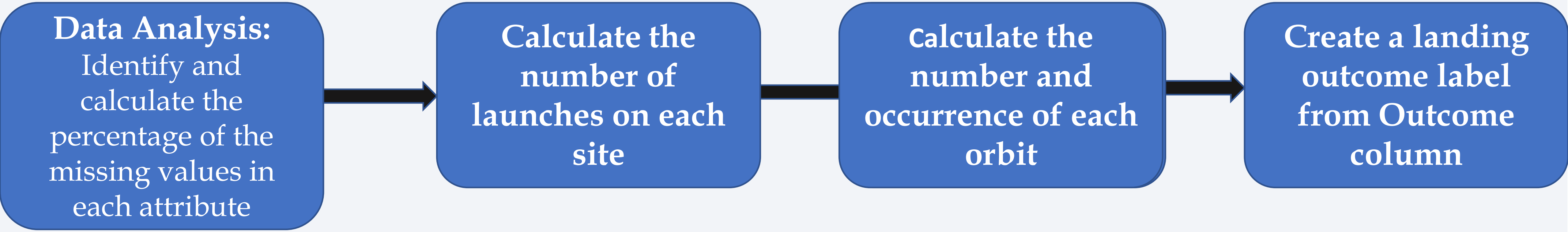
Request the Falcon9 Launch Wiki page from its URL → Create a BeautifulSoup Object from the HTML Response → Extract all column/variable names from the HTML table header → Create a data frame by parsing the launch HTML tables

Data Wrangling

Objectives:

Perform exploratory Data Analysis and determine Training Labels

- Exploratory Data Analysis
- Determine Training Labels



EDA with Data Visualization

The main Objectives of this section is perform Data Analysis and Feature Engineering, The different plot below was been used to explore the data:

- A CatPlot to Visualize the relationship between Flight Number and Launch Site,
- Scatter Plot to Visualize the relationship between Payload and Launch Site
- An ordered Bar Graph to Visualize the relationship between success rate of each orbit type,
- A CatPlot to Visualize the relationship between FlightNumber and Orbit type,
- Scatter plot Visualize the relationship between Payload and Orbit type,
- A Line Plot to Visualize the Yearly Success Trend

EDA with SQL

The main Objectives of this section is to analyse data using SQL Command lines:

- Display the names of the launch sites in the space mission
- Filtering dataset based on Input Parameter using SQL
- List the date when the first successful landing outcome in ground pad was acheived
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Applying SQL function to analyse data manually in database
- Use a subquery

Build an Interactive Map with Folium

The launch success rate may depend on many factors such as payload mass, orbit type, and so on. It may also depend on the location and proximities of a launch site, i.e., the initial position of rocket trajectories. Finding an optimal location for building a launch site certainly involves many factors and hopefully we could discover some of the factors by analysing the existing launch site locations.

- We could use Circle to add highlighted circle area with a text label on specific coordinates,
- We can also use Marker to point different disperced point on the map, to point a location site

Predictive Analysis (Classification)

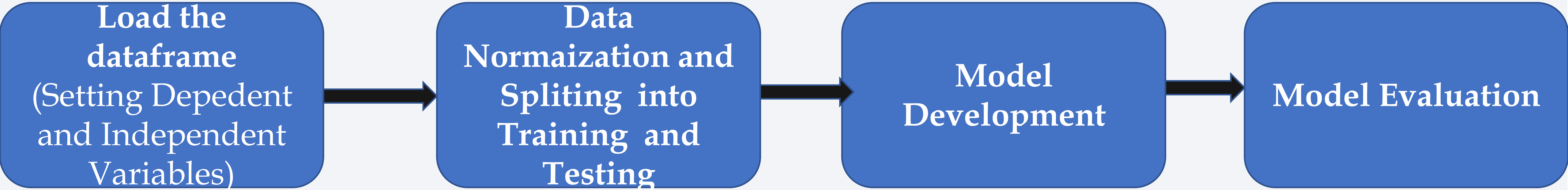
Objectives

Perform exploratory Data Analysis and determine Training Labels

- create a column for the class
- Standardize the data
- Split into training data and test data

Find best Hyperparameter for SVM, Classification Trees and Logistic Regression

- Find the method performs best using test data



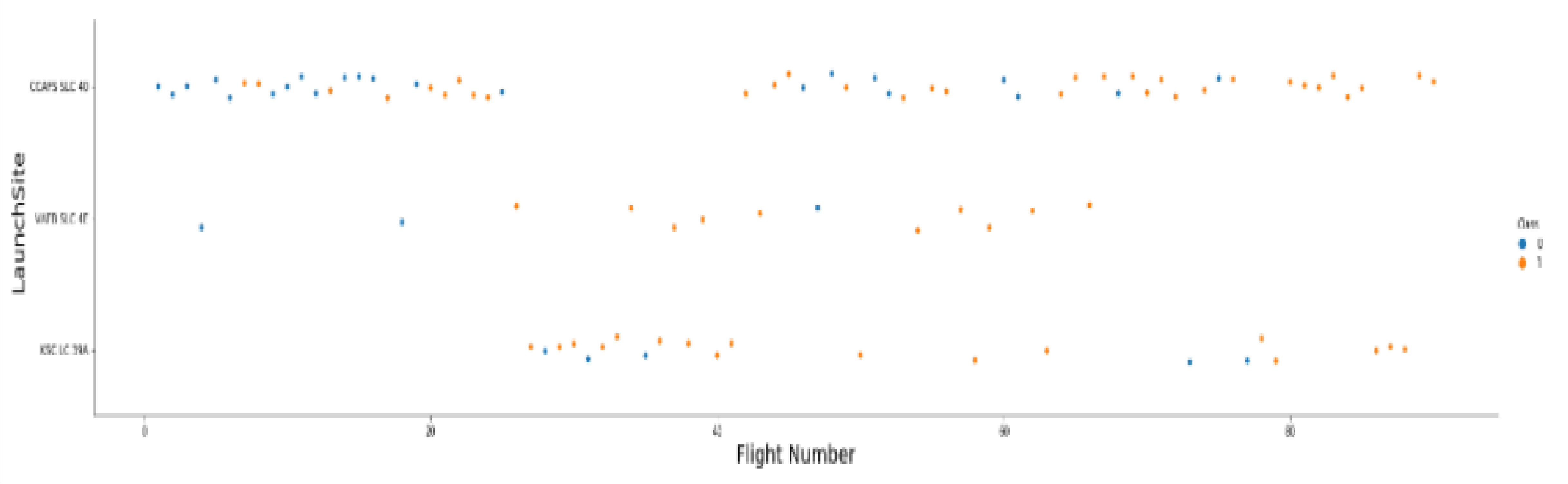
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

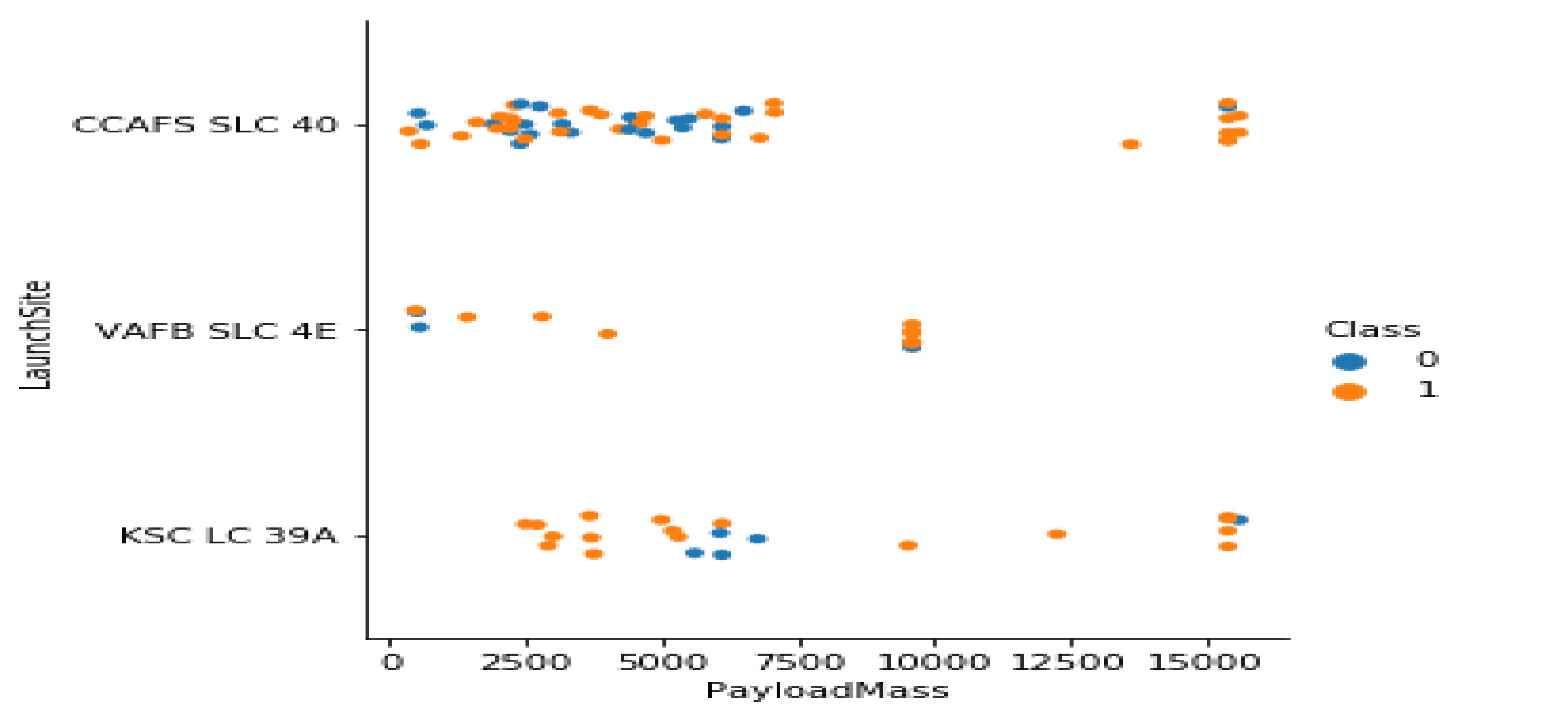
Insights drawn from EDA

Flight Number vs. Launch Site



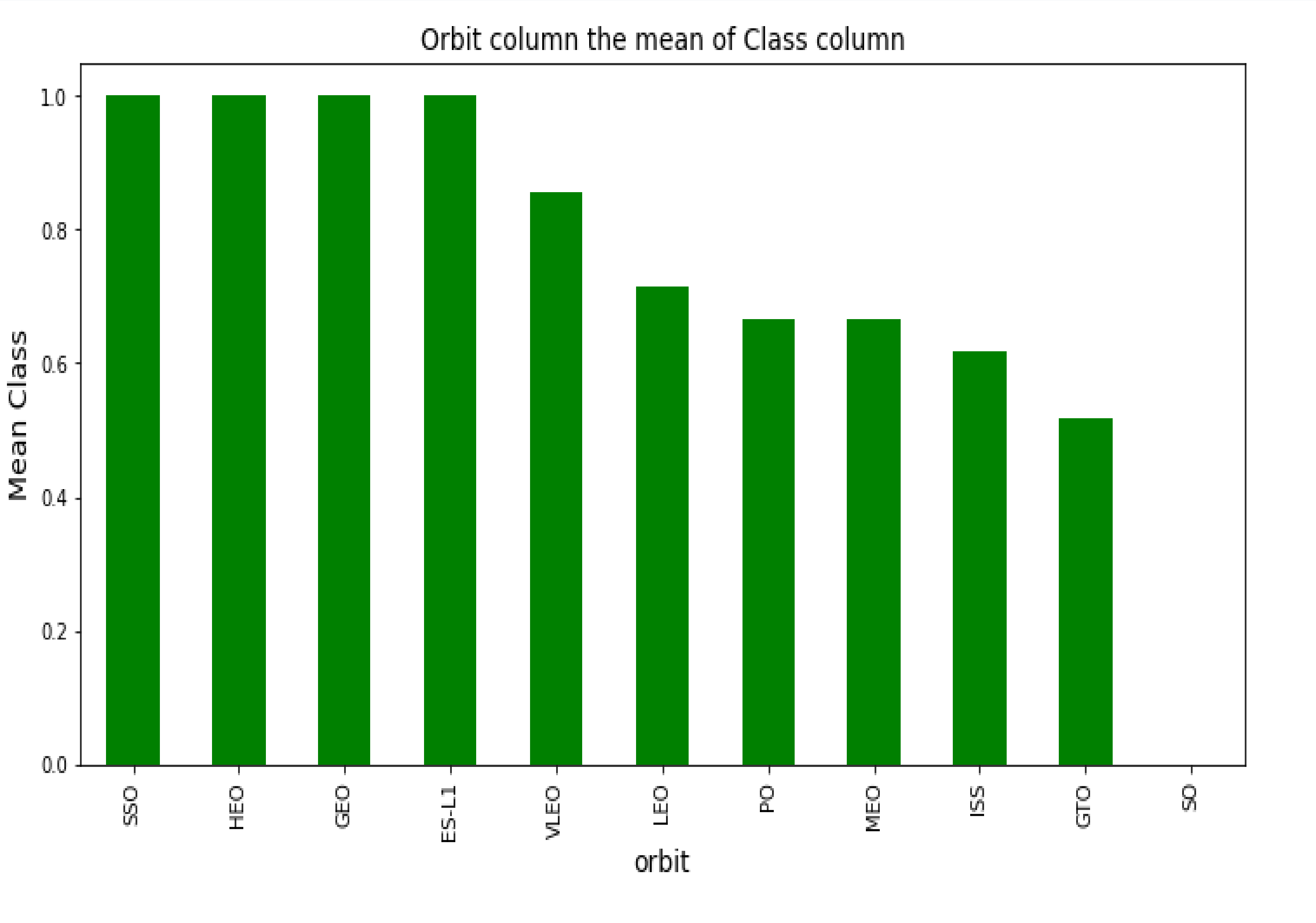
The above plot shows the correlation between flight number and the launch site, the succeeded launhs colored in Yellow while the dailure coloredin blue, we can see there is relationship between the two variables because when the Flight Number increase, the launch to succeed at every Launch site,

Payload vs. Launch Site



The above plot shows the correlation between Payload Mass and the launch site, we can see there is relationship between the two variables because when the Payload Mass increases, the launch to succeed at every Launch site,

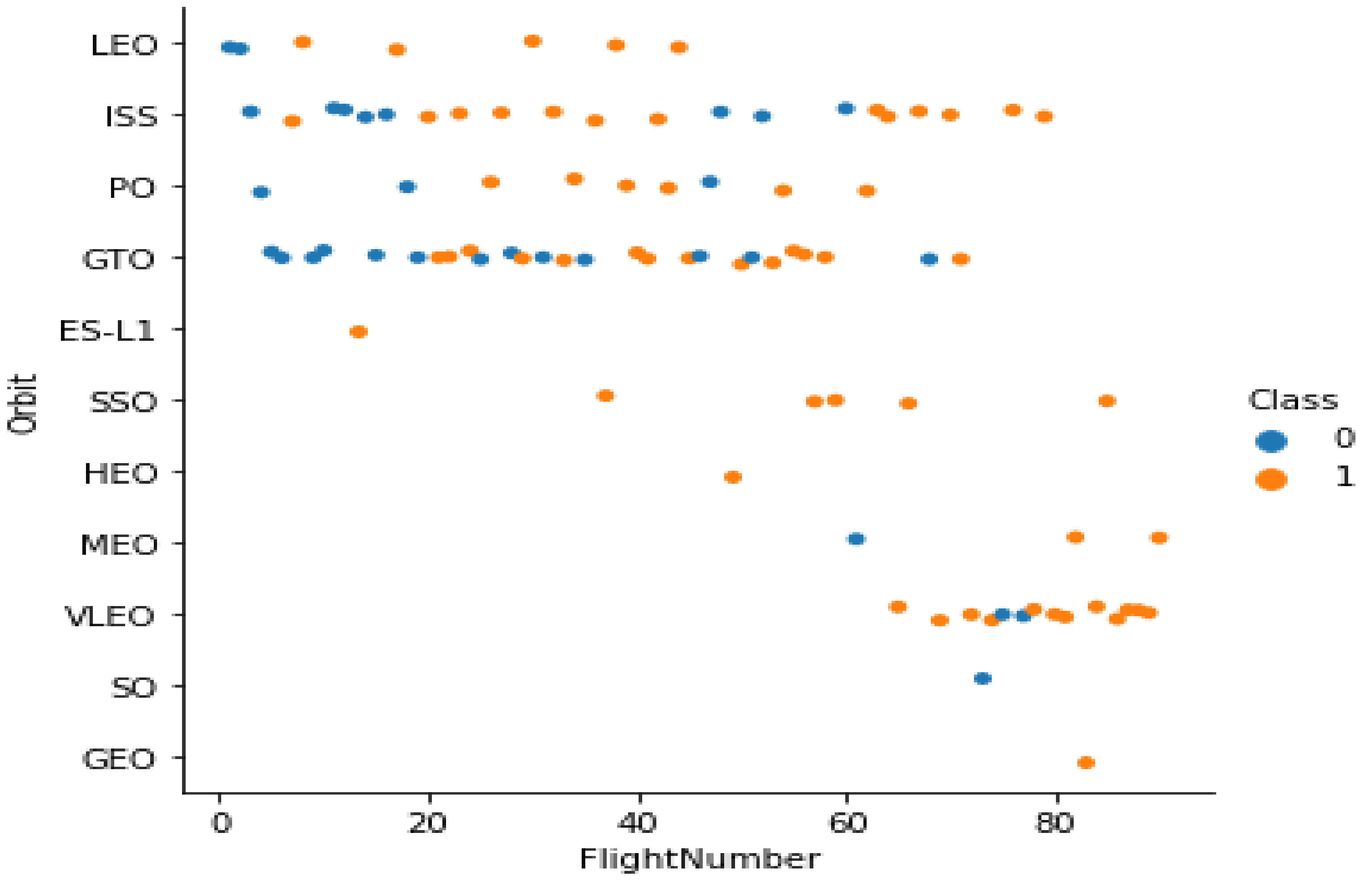
Success Rate vs. Orbit Type



✓ **Figure Interpretation**

The figure two show the mean results of each launch by the orbit criteria. You will notice, the diagram classed in descending order that the four first Orbit in occurrence SSO, HEO, GEO, ES-L1 have launched their rocket successfully without failure. While the misson carried out at the SO orbit only resulted in total failure.

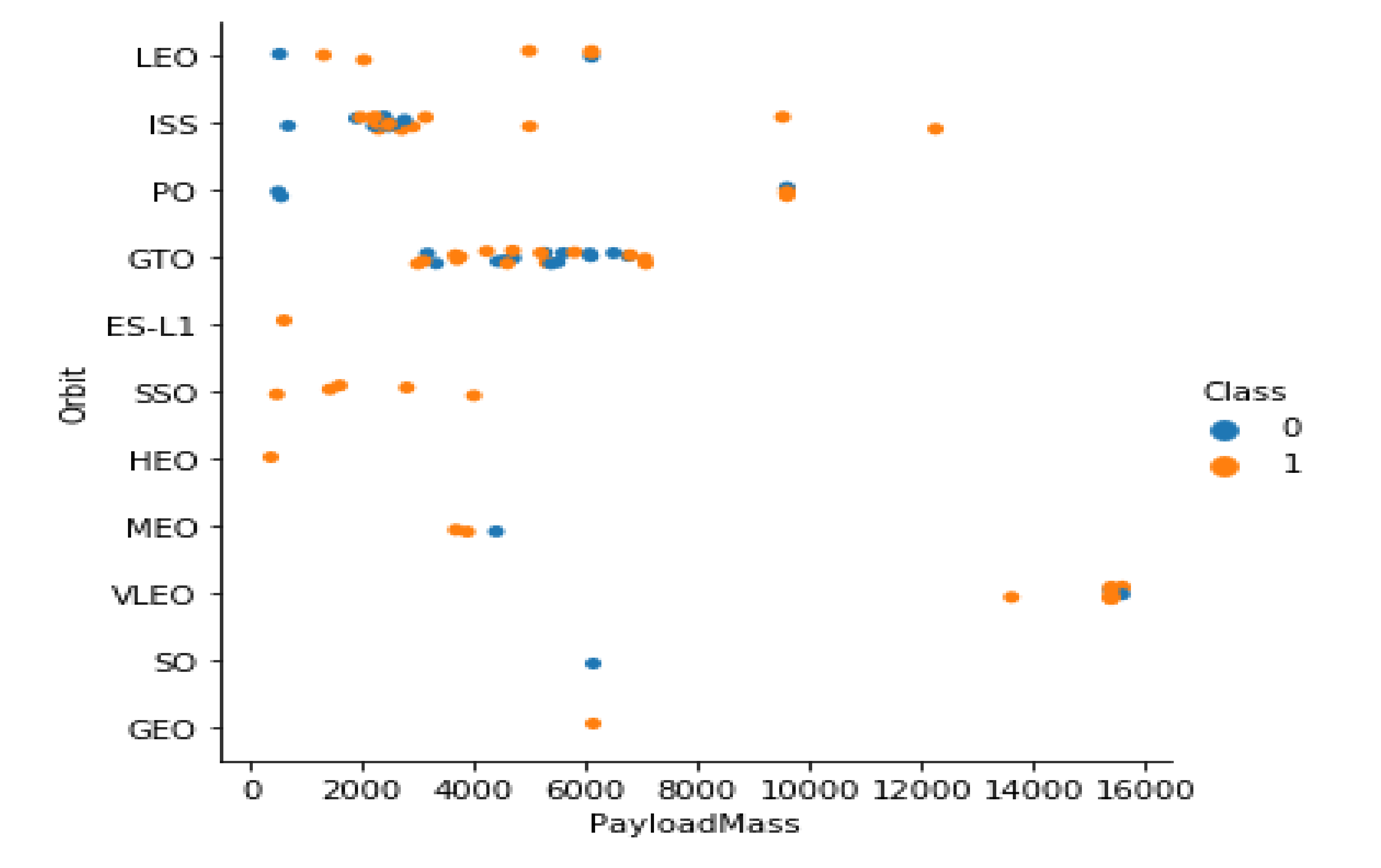
Flight Number vs. Orbit Type



✓ **Interpretation:**

The above plot shows the correlation between the Flight Number and the Orbit, we can see there is a positive relationship between the two variables because when the Flight Number increases, the launch to succeed at every Orbit,

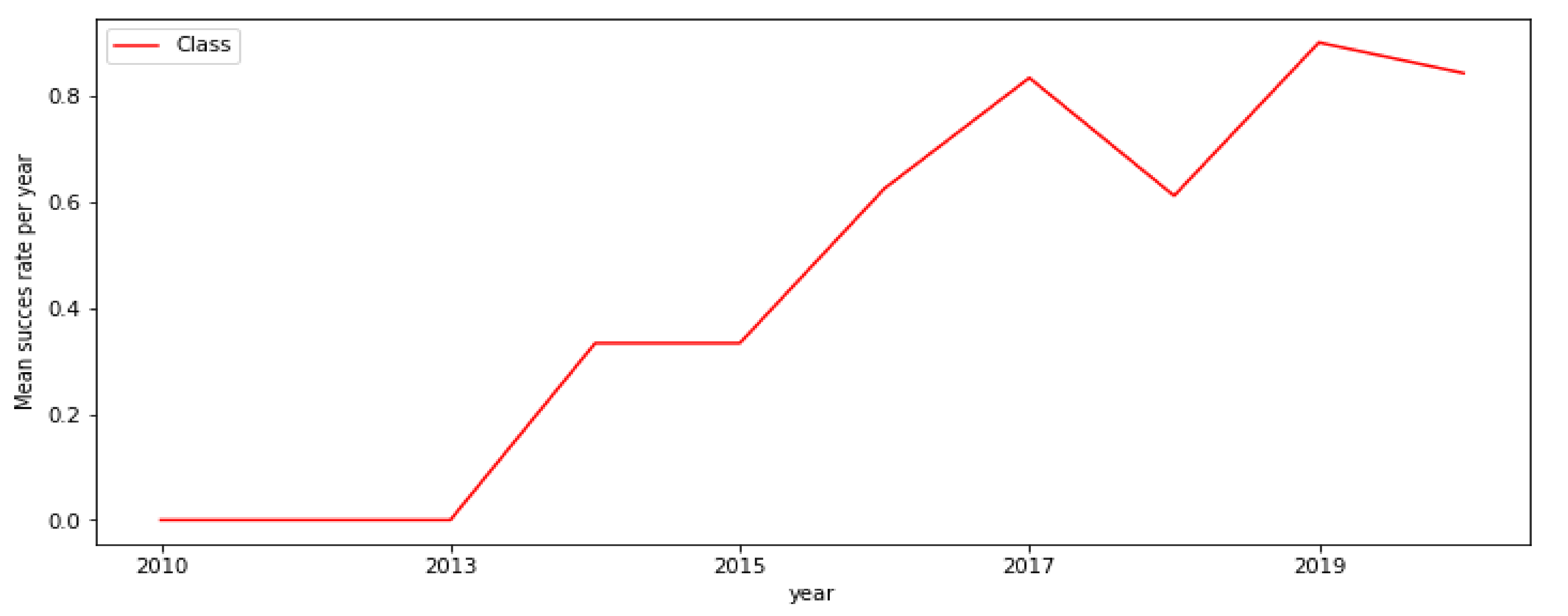
Payload vs. Orbit Type



✓ **Interpretation:**

The above plot shows the correlation between the Payload Mass and the Orbit, we can see there is a positive relationship between the two variables because when the Payload Mass increases, the launch to succeed at every Orbit,

Launch Success Yearly Trend



This diagram shows the Launch Success Yearly Trend which tend to increases the time goes on, this is may due to the evolution in design and Technology during the process

All Launch Site Names

All the Launch Site names can be displayed by requesting the Unique Site in IBM2 database created from the collected data:

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

Launch Site Names Begin with 'CCA'

We can also access more Information about the dataset by displaying the Launch Site Names Beginning with ‘CCA’:

- CCAFS LC-40
- CCAFS LC-40
- CCAFS LC-40
- CCAFS LC-40
- CCAFS LC-40

Total Payload Mass

F9 B4 B1039.1	3310
F9 FT B1021.1	3136
F9 B5 B1058.4	2972
F9 FT B1035.1	2708
F9 B4 B1045.2	2697
F9 B4 B1039.2	2647
F9 B5B1050	2500
F9 B5B1056.1	2495
F9 FT B1031.1	2490
F9 v1.1 B1012	2395
F9 v1.1	2296
F9 B5 B1056.2	2268
F9 FT B1025.1	2257
F9 v1.1 B1010	2216
F9 FT B1035.2	2205

Displayed total payload mass carried by boosters launched by NASA (CRS)

Average Payload Mass by F9 v1.1

F9 v1.1 B1016	4707
F9 v1.1 B1011	4428
F9 v1.1 B1014	4159
F9 v1.1	2928
F9 v1.1 B1012	2395
F9 v1.1 B1010	2216
F9 v1.1 B1018	1952
F9 v1.1 B1015	1898
F9 v1.0 B0007	677
F9 v1.1 B1013	570
F9 v1.1 B1017	553
F9 v1.0 B0005	525
F9 v1.0 B0006	500
F9 v1.1 B1003	500
F9 v1.0 B0003	0
F9 v1.0 B0004	0

Displayed average payload mass carried by booster version F9 v1.1

First Successful Ground Landing Date

The Output below show the date when the first successful landing outcome in ground pad was achieved:

DATE	landing__outcome
0001-01-22	Success (ground pad)

Successful Drone Ship Landing with Payload between 4000 and 6000

Let now list the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

booster_version	landing__outcome	payload_mass__kg_
F9 FT B1022	Success (drone ship)	4696
F9 FT B1026	Success (drone ship)	4600
F9 FT B1021.2	Success (drone ship)	5300
F9 FT B1031.2	Success (drone ship)	5200

Total Number of Successful and Failure Mission Outcomes

List of the 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

F9 B5 B1048.4	15600
F9 B5 B1048.5	15600
F9 B5 B1049.4	15600
F9 B5 B1049.5	15600
F9 B5 B1049.7	15600
F9 B5 B1051.3	15600
F9 B5 B1051.4	15600
F9 B5 B1051.6	15600
F9 B5 B1056.4	15600
F9 B5 B1058.3	15600
,,,,,,,,,	,,,,,,,,,,,,,
,,,,,,,,,	,,,,,,,,,,,,,
F9 B5 B1049.6	15440
F9 B5 B1059.3	15410
F9 B5 B1051.5	14932
F9 B5 B1049.3	13620

Using a subquery, we can also List the names of the booster_versions which have carried the maximum payload mass.

2015 Launch Records

The failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015:

booster_version	landing__outcome	launch_site
15	Failure (drone ship)	CCAFS LC-40

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

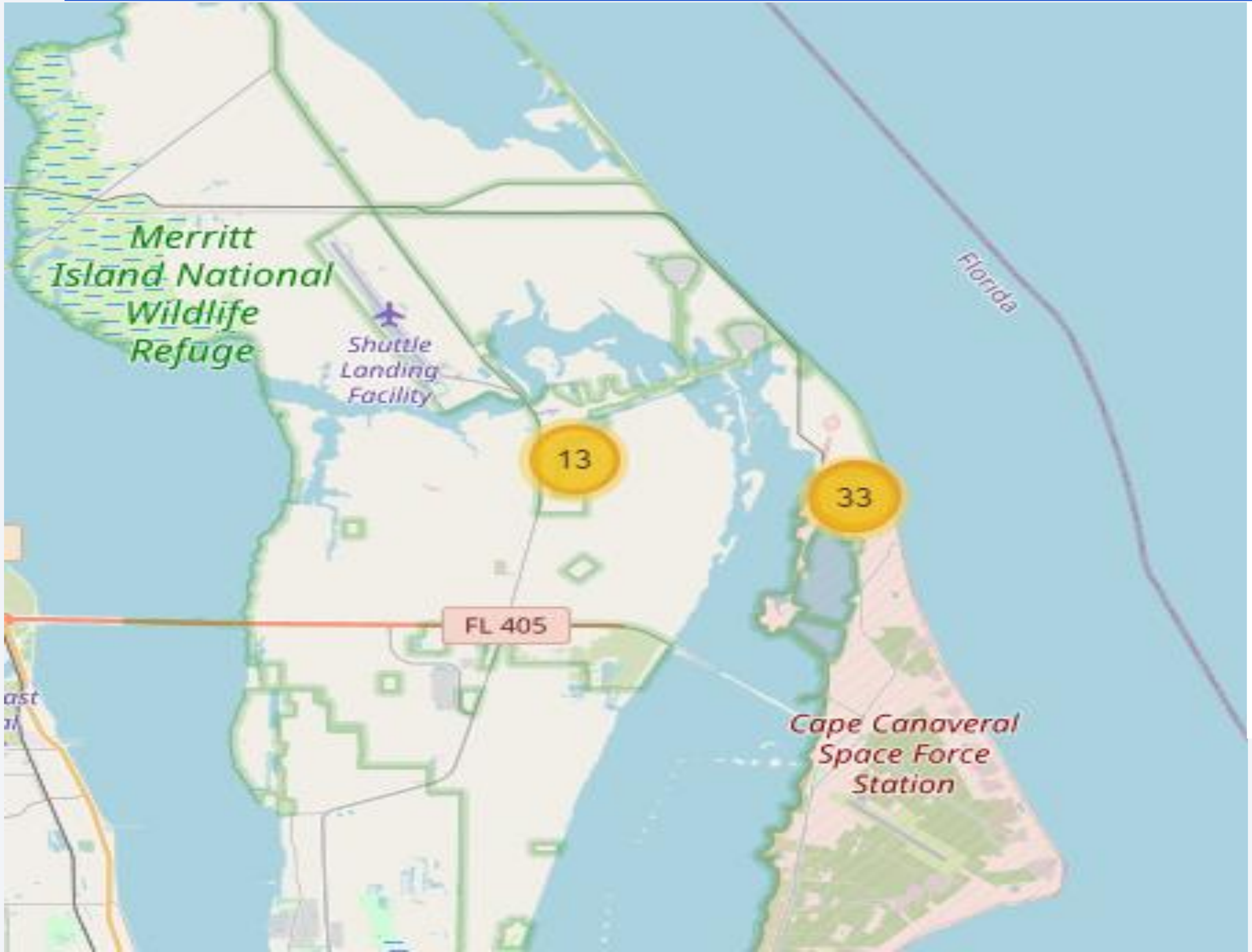
Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order:

landing__outcome	**COUNT**
Success	6
No attempt	3
Success (ground pad)	3
Failure	2
Failure (drone ship)	2
Controlled (ocean)	1
Success (drone ship)	1

Section 4

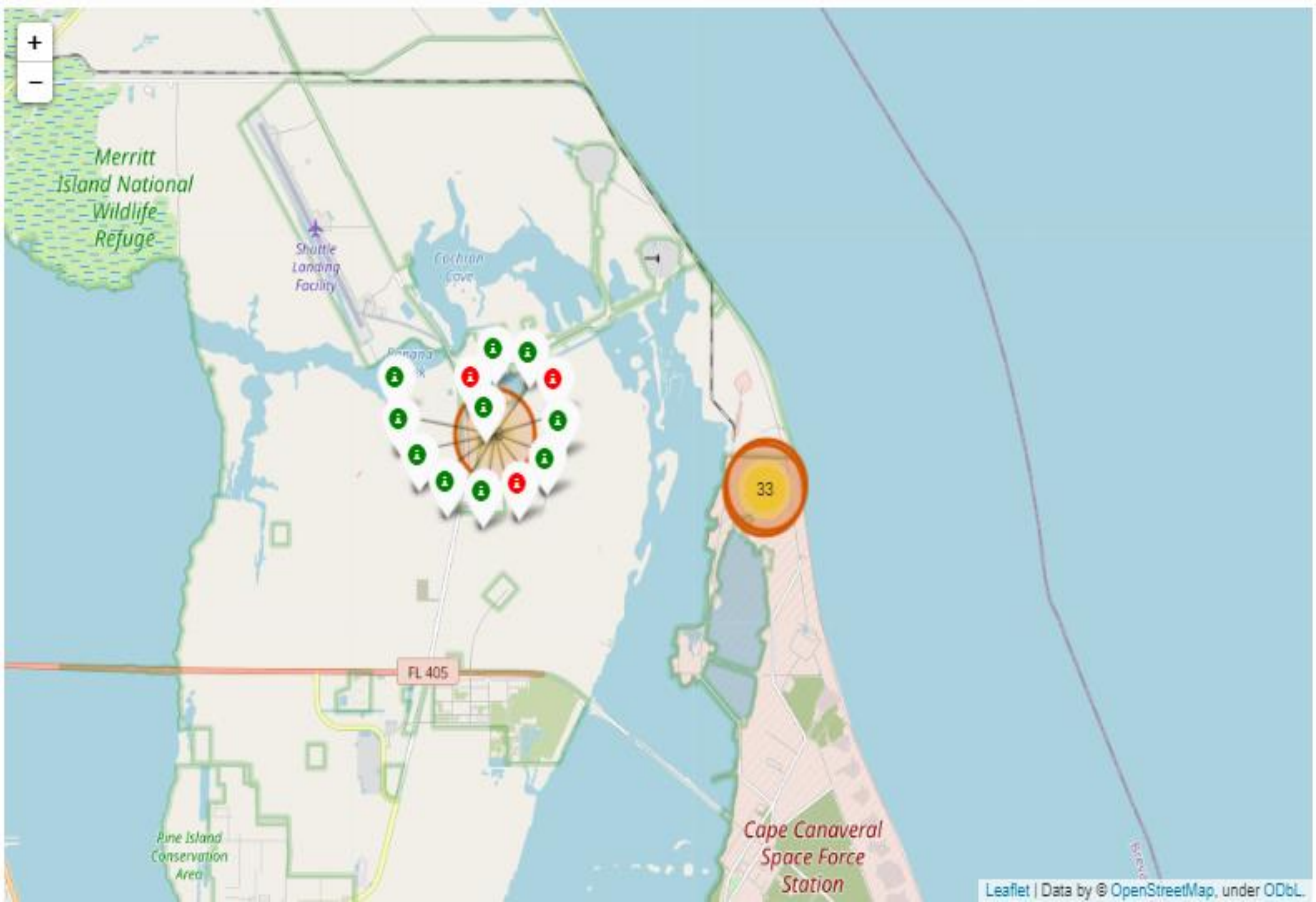
Launch Sites Proximities Analysis

<Folium Map Screenshot 1>



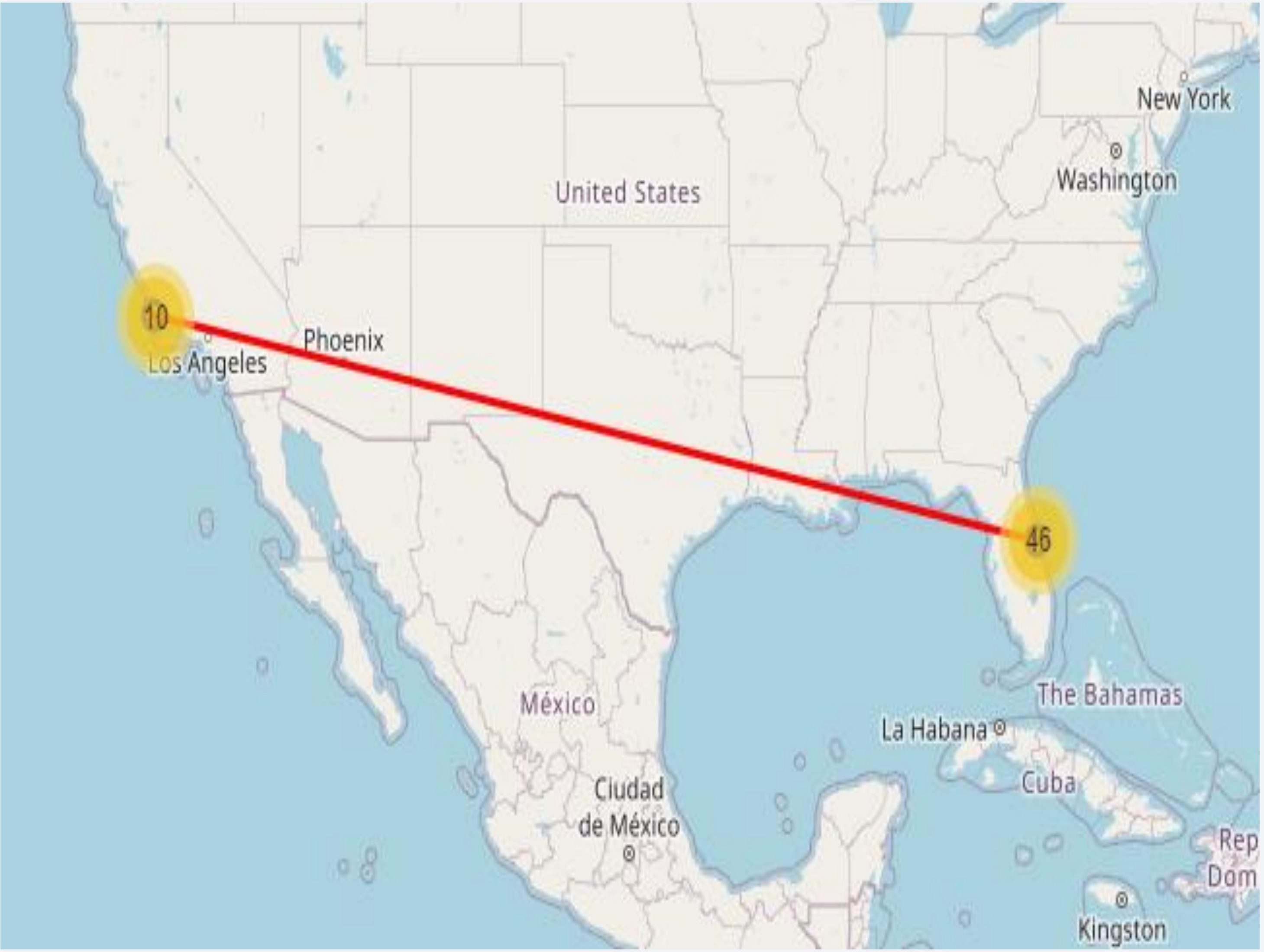
The figure shows the distribution of the site location on a global map,,, The nearest points fo launch are grouped within a market cluster for better visualization,
This is just a section of an entire map that contain 56 launches,

<Folium Map Screenshot 2>



In this map we explored the Launch Site, where the surface of launch is materialised with a red circle and the point of launch are materialized in white Markers. These markers are grouped in closest area where successful launch are set in green colour inside the whites marker and failed launch are set in red colour inside the white marker

<Folium Map Screenshot 3>

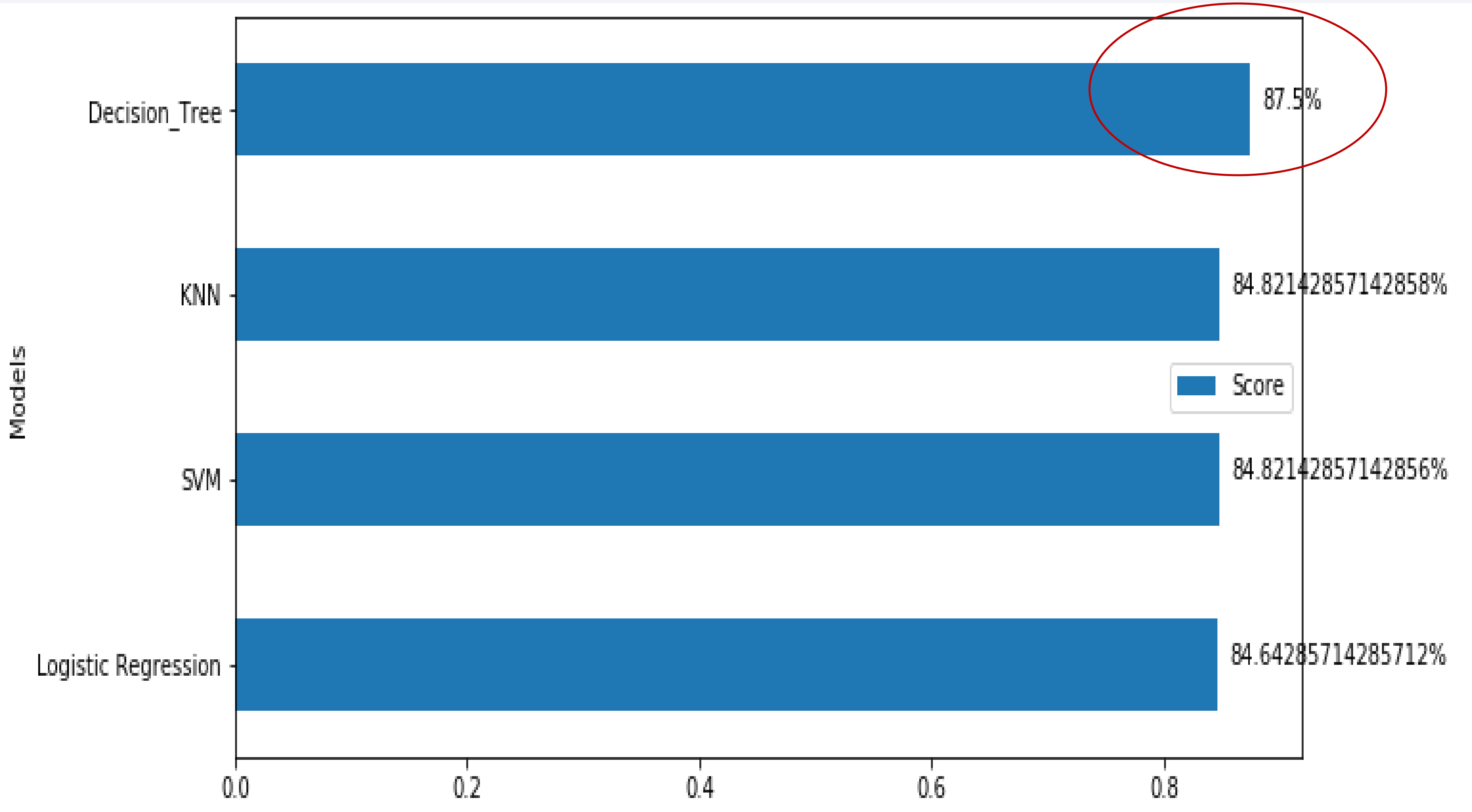


Mark down of a point on the closest launch site and calculate the distance between the railway point to the launch site.
The distance between the 2 Point is about 823.5603820604156 km,

Section 6

Predictive Analysis (Classification)

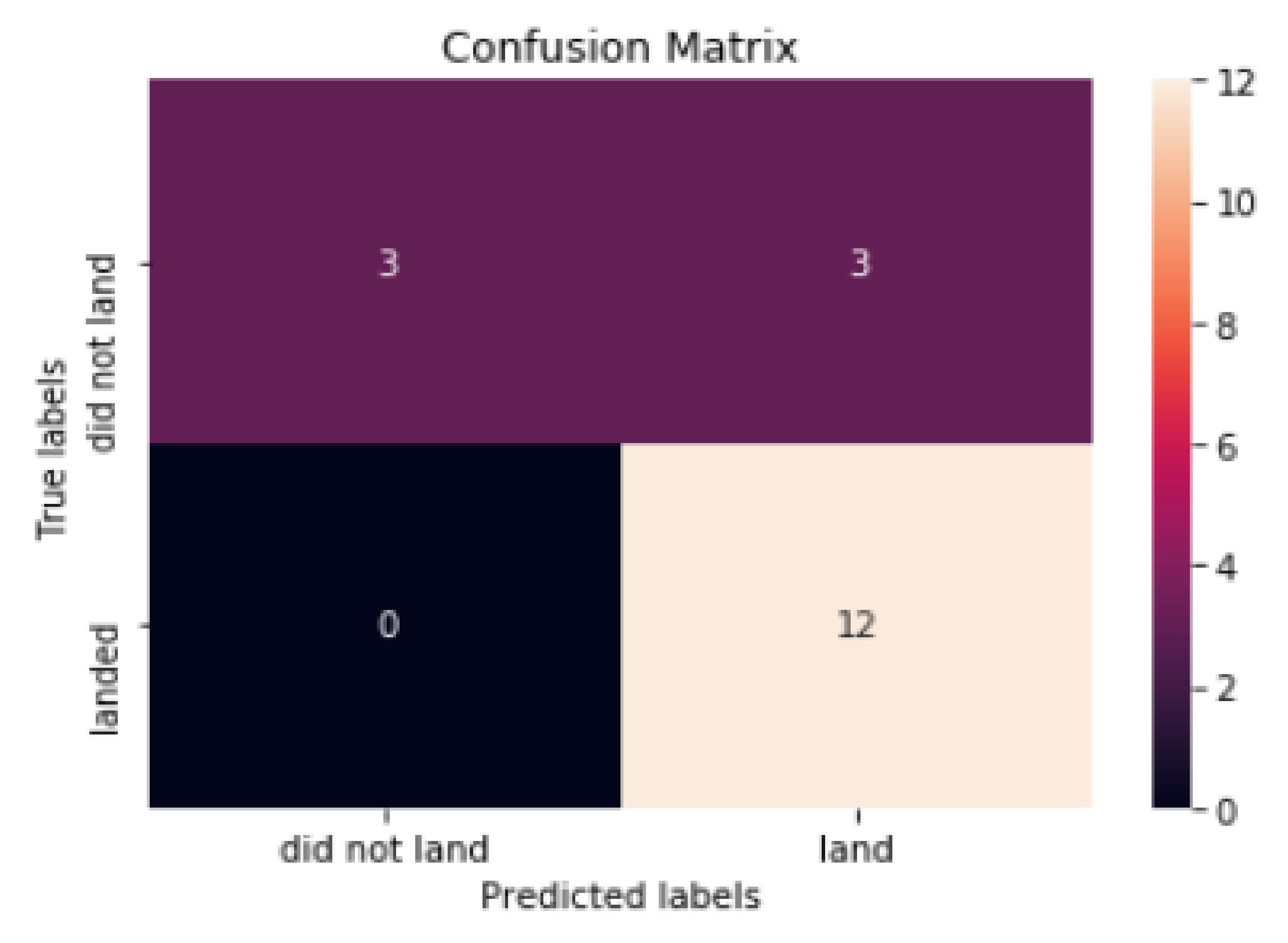
Classification Accuracy



This Horizontal Bar shows the Prediction Accuracy performed by each Model, including the Logistic Regression, the Decision Tree, the Key Nearest Neighbour and Support Vector Machine,,, Note how we selected the Decision Tree as the Best Model to predict the Landing Outcome of the next rocket with 87,5% on accuracy,

Confusion Matrix

- **Find the method performs best using test data**



As we can see on the diagram, the model was tested on 18 launches to check its efficiency… Therefore, the model predicted correctly all the stages that could not launch and little confused on land mission by predicting 3 of them as failed mission.

This model will require more data from the SpaceX source or will need more optimisation process

Conclusions

Machine learning software tools are widely employed in modern industries. In fact, the global market size for machine learning is expected to reach the value of 117.19 billion size by 2027.

During this project SpaceY predict the outcome of the first stage land using data science methods to collect, analyse and Visualize to really understand the challenge of the launch. During the same process, SpaceY tested varieties of machine learning technics including Logistic Regression, K Nearest Neighbour, S.V.M(Support Vector Machine) and Decision Tree to predict the mission Outcome.

By all counts, and with proven results, its no wonder that Decision Tree was considered as the best model to fit the data provided by SpaceX with the Accuracy of **87.68** % and **83.33%** score on the testing data.

Appendix

This Section contain information about data collection, Analysis and Visualiazation:

- https://github.com/albertbulenda/IBM-Data-Science-Professional-Certificate/blob/main/jupyter-labs-webscraping.ipynb
- https://github.com/albertbulenda/IBM-Data-Science-Professional-Certificate/blob/main/lab_jupyter_launch_site_location.ipynb
- https://github.com/albertbulenda/IBM-Data-Science-Professional-Certificate/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

Appendix

This Section contain information about data collection, Analysis and Visualiazation:

- https://github.com/albertbulenda/IBM-Data-Science-Professional-Certificate/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipyn
- https://github.com/albertbulenda/IBM-Data-Science-Professional-Certificate/blob/main/jupyter-labs-eda-dataviz.ipynb
- https://github.com/albertbulenda/IBM-Data-Science-Professional-Certificate/blob/main/jupyter-labs-eda-sql-coursera.ipynb
- https://github.com/albertbulenda/IBM-Data-Science-Professional-Certificate/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Thank you!