SpaceX Falcon 9 First Stage Landing Prediction

Presented by: Isabell Odambo

Report Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion/Insights
- References

Executive Summary

SpaceX is currently the most cost-effective rocket manufacturing company, retailing their space crafts at approximately $62 million comparative to the $165 million market price with other competitor companies. This is highly attributed to SpaceX's Falcon 9 reusable technologies [1] The goal of this project is to train a machine learning model to predict whether SpaceX will re-use its first stage after launching a rocket.

The following data science techniques were utilized to predict SpaceX Falcon 9 First Stage Landing:

1.Data Collection - Web scraping and APIs

2.Data Wrangling - Pre-processing SpaceX Data

3.Exploratory Data Analysis - SQL and Python Visualization

4. Data Visualization - Folium and Dashboard Applications

5.Predictive analysis - Machine Learning Classification algorithms.

The results analyzed demonstrate 83% accuracy of the model to predict successful landing of Falcon 9 Stage 1

Introduction

- Space X is a company , founded in 2002 that has revolutionized space technology by designing, manufacturing and successfully launching advanced rockets and spacecraft. [2]
- SpaceX has optimized cost of production by re-using it’s first stage. Therefore, predicting whether the stage will land will assist in determining the cost of the first launch [2].
- In this project, I train a machine learning model to predict whether the first stage of a Falcon 9 can be re-used for landing. This information is useful for SpaceY to bid an alternative rocket ship to customers.
- I utilize various Data Science techniques to collect, pre-process, visualize data and identify patterns and trends that will help develop an algorithm to predict re-usability of the first stage of Falcon 9 rocket.

Methodology - Data Collection

1. Use GET request to obtain and parse SpaceX launch data. Look out for a ‘200’ response code to signify successful response.
2. Use *json_normalize* method to convert the json result into a dataframe.
3. Use the APIs to obtain additional information information for our desired columns:
    a. Rocket- Booster name
    b. Payloads - Mass and Orbit
    c. Launchpad - Launch site (Latitude and longitude)
    d. Cores - Serial, re-usability, landing pad used etc
4. Create a pandas dataframe to display and store the information obtained.
5. Filter the data frame to exclusively get data for the Falcon 9 launches.
6. Refer to my **Data Collection** code for reference of this process**.**

Methodology - Data Wrangling

Data wrangling involves transforming raw data into the structured data and the interested format that will be useful for analysis and machine learning [3]. Utilize the following steps to clean/restructure the raw data obtained:

1. From the data frame, remove rows with missing values as observed in the launching pad column data.

2. Verify the data types and convert the class into necessary training labels.

3. Calculate the mean value of the payload mass and replace the np.nan values with its mean value

4. Export it to a .csv file.

5.Refer to my **Data Wrangling** code for reference of this process**.**

Methodology - Exploratory Data Analysis (EDA)- SQL

- Exploratory Data Analysis (EDA) is key step in data processing used to identify trends and patterns in data. It offers a better understanding of the data which can be used for predictive modelling and testing various hypothesis. The figure below shows the various steps utilized for EDA.
- Refer to my **EDA SQL** code for reference of this process**.**

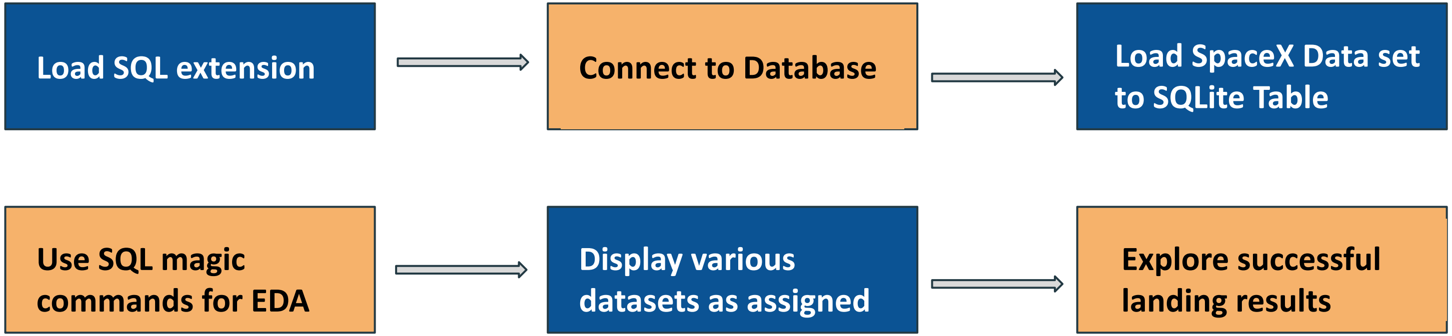


Figure 1: Flowchart representing the sequence of steps used in Exploratory Data Analysis

Methodology - Exploratory Data Analysis (EDA)- Python Visualization

Visualize the following relationships between the various parameters by creating:

1. Scatter plots for:

- Flight Number and Launch Site
- Payload Mass and Launch Site
- Success rate of each orbit type
- Flight Number and Orbit Type
- Payload Mass and Orbit type

2. Bar Charts for the success rate of each orbit type.

3. Line Plot for the annual trend of the lau launch successes.

Refer to my **EDA Python Visualization** code for reference of this process**.**

Methodology - Folium

- Folium is a Python library tool used to create interactive geographical maps by binding data to a map for choropleth visualizations[4]. Functions such as folium.circle, folium.map, MarkerCluster(), Folium.PolyLine() were used to represent this data.
- Refer to my **Folium and Dashboard** code for reference of this process**.**

Figure 2: Flowchart representing the sequence of steps used in representing data on Folium

Methodology - Plotly Dashboard

- A Dashboard is a Python library tool used to represent data using plots and images using a preferred layout as developed by a user[5].
- Refer to my **Folium and Dashboard** code for reference of this process.

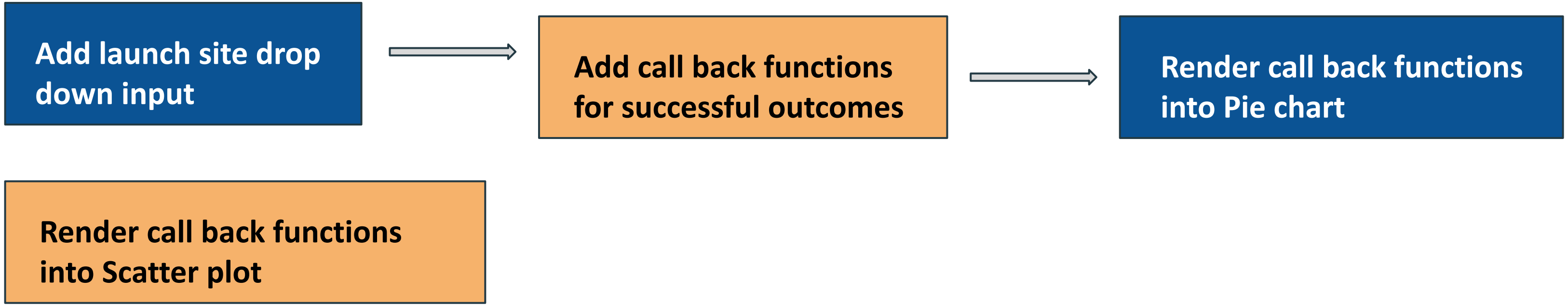


Figure 3: Flowchart representing the sequence of steps used in representing data on a Plotly Dashboard

Methodology - Predictive Analysis

- Build and evaluate the classification models used in this data set by conducting the steps shown below.
- Refer to my **Predictive Analysis** code for reference of this process**.**

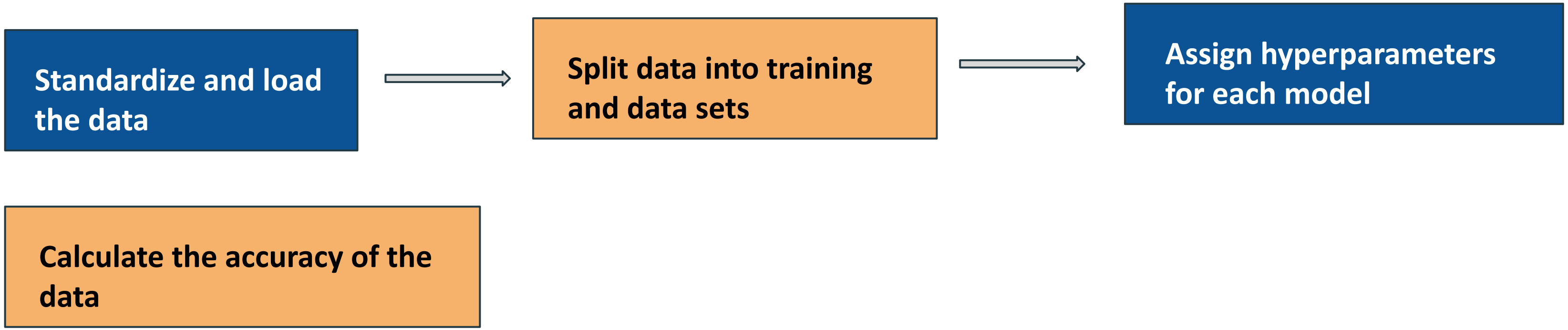


Figure 4: Flowchart representing the sequence of steps used for Predictive Analysis

Results - EDA with Visualization

- Launch site CCAFS SLC 40 has launched the most rockets.
- Higher successful landing rates are shown for VAFB SLC 4E and KSC LC 39A launch sites.

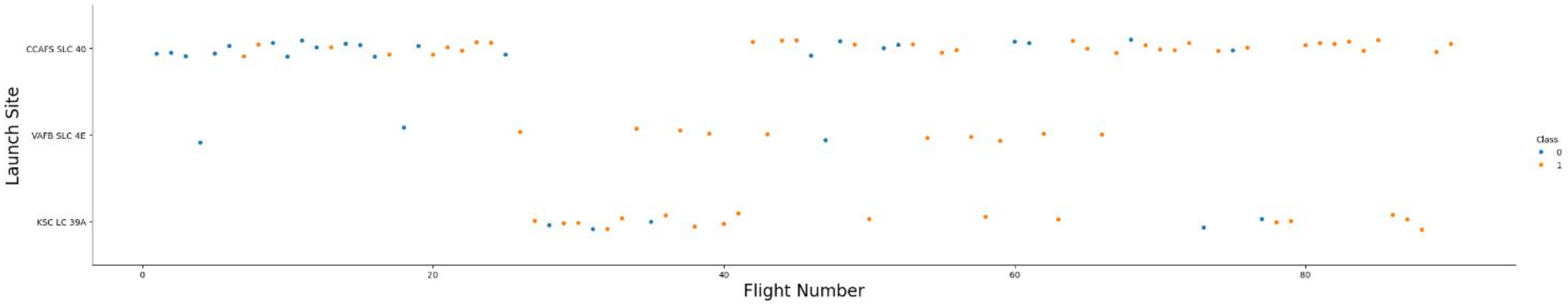


Figure 5: Scatter Plot of Flight Number vs Launch Site

Results - EDA with Visualization

- Most of the rockets launched have a payload mass below 9000Kg.
- CCAFS SLC 40 has the highest landing success rate for rockets weighing between 14000Kg and 16000Kg.

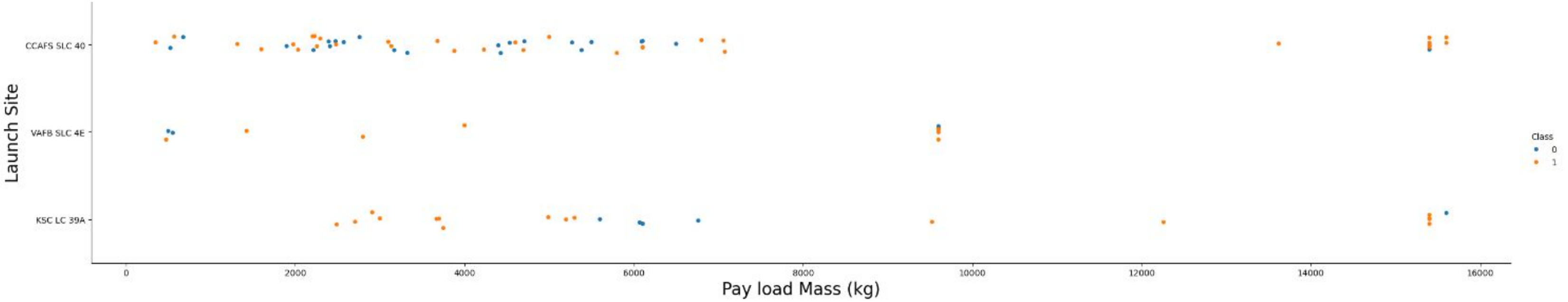


Figure 6: Scatter Plot of Pay load Mass vs Launch Site

Results - EDA with Visualization

- Orbis SO has the least success rate whereas orbits SSO, HEO, GEO and ES-L1 have the highest success rates.

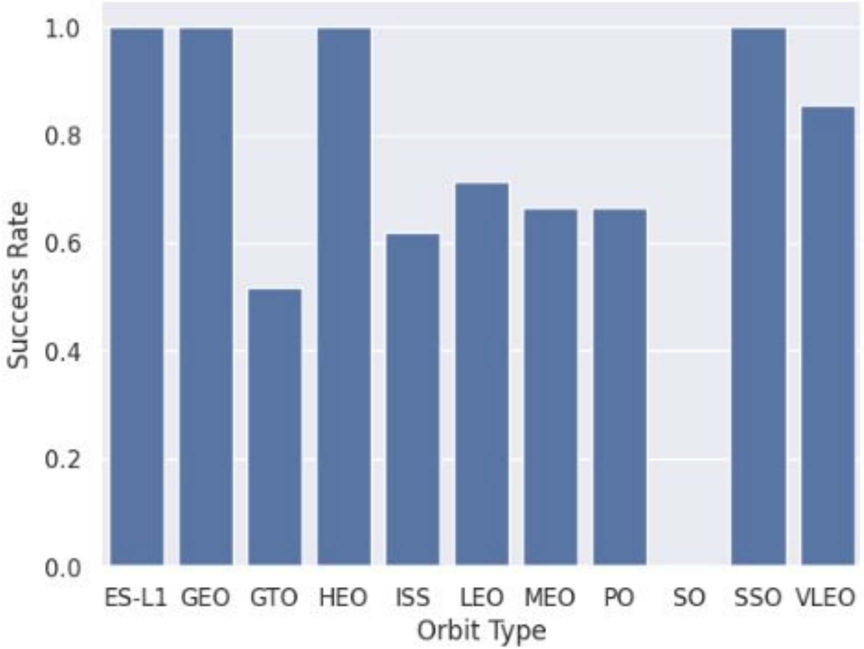


Figure 7: Bar Chart of Orbit Type vs Success Rate

Results - EDA with Visualization

- VLEO orbit displays successful landing rates for flight numbers greater than 60

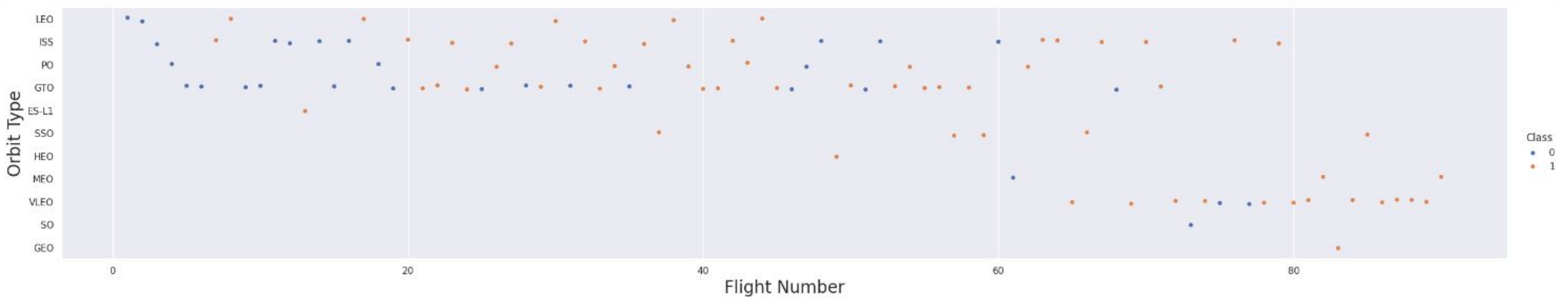


Figure 8: Scatter Plot of Flight number vs Orbit Type

Results - EDA with Visualization

- Orbits LEO, ISS and PO have more successful landing rates exhibiting a positive correlation with increasing payload mass.

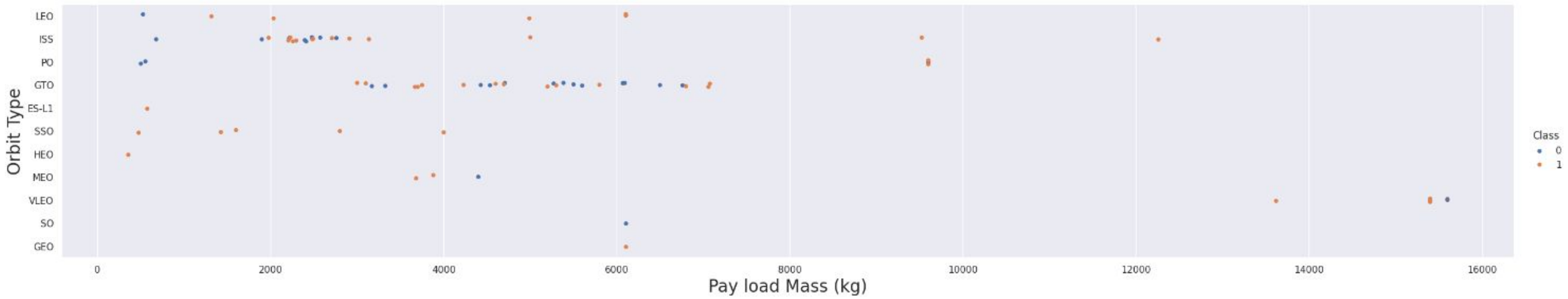


Figure 9: Scatter Plot of Pay load Mass vs Orbit Type

Results - EDA with Visualization

- Success rate of Falcon launching has increased annually, since 2013. There was a drop in success rate in 2018, however, the rate increased in the following year.

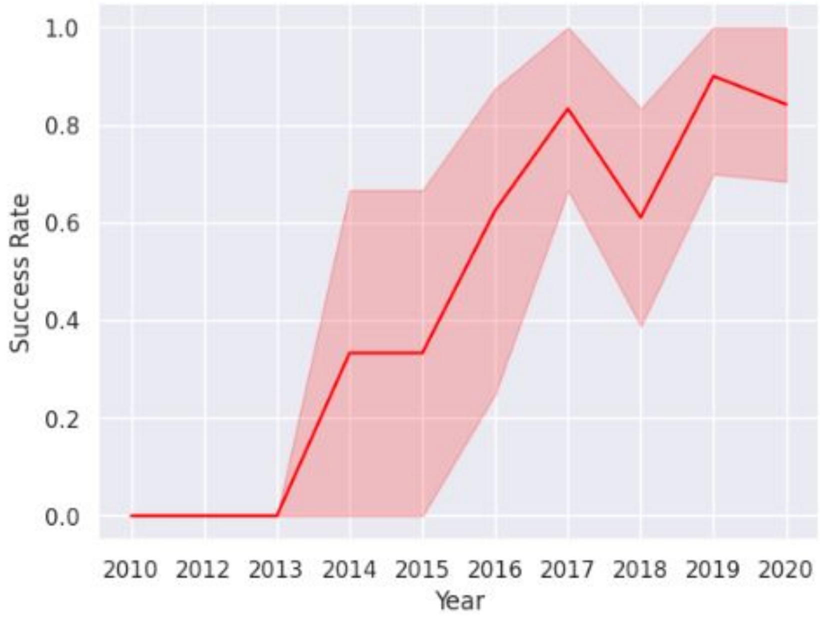


Figure 10: Line Plot of Year vs Success rate

Results - EDA with SQL

The following shows the Unique launch site in the space mission

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

Out[12]:

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

The following shows 5 record where launch sites begin with string ‘CCA

Out[13]:

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

Results - EDA with SQL

The following shows the Total Payload Mass carried by Boosters launched by NASA (CRS)

Out[16]:

Customer	Total_NASA_CRS_mass
NASA (CRS)	45596

The following shows the average Payload mass carried by booster version F9 v1.1

Out[20]:

Booster_Version	avg_Booster_versionF9_v1_1
F9 v1.1	2928.4

Results - EDA with SQL

The following shows the date when the first successful landing outcome in ground pad was achieved.

Out[23]:

Mission_Outcome	Date_First_Succ_Land
Success	2015-12-22

The following shows the boosters which have success in drone ship with payload mass greater between 4000kg and 6000kg

Out[24]:

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

Results - EDA with SQL

The following shows the total number of successful and failure mission outcomes

Out[27]:

Mission_Outcome	Total (Success or failure)
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

The following shows the booster versions which have carried maximum payload mass

Out[28]:

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

Results - EDA with SQL

The following shows months relative to landing outcomes.

Done.

Out[29]:

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)

The following ranks the landing outcomes at a specific period

Out[30]:

Landing_Outcome	Total Count
Success (ground pad)	5
Failure (drone ship)	5

Results - Interactive Map with Folium

- SpaceX Launch sites are typically along the Florida and California Coast lines in the United States.

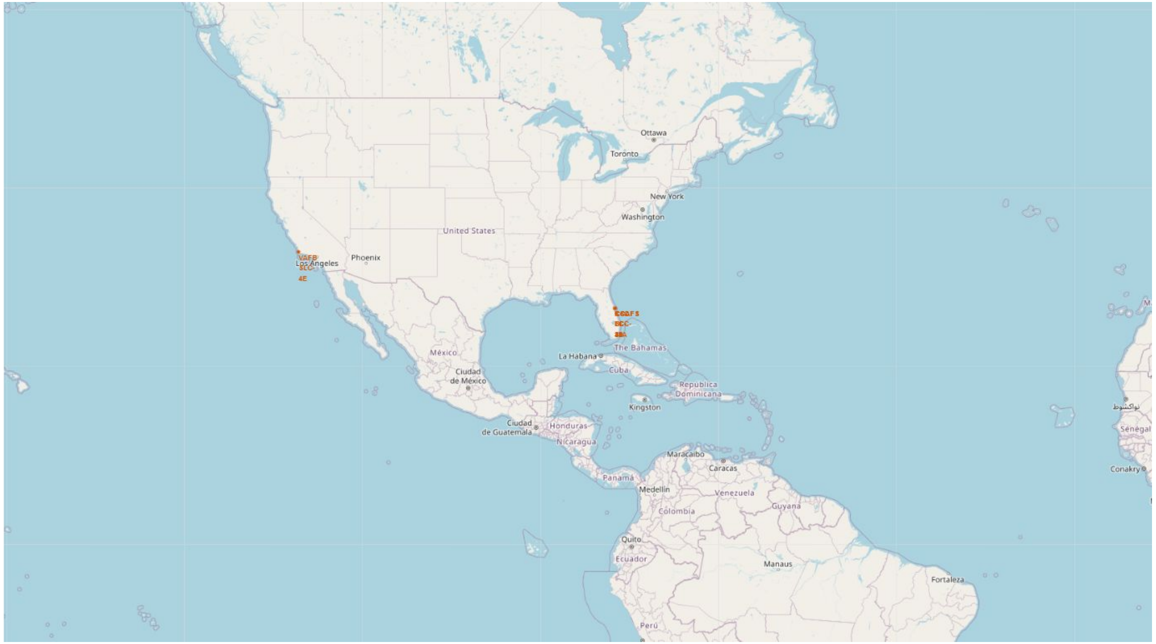


Figure 11: Launch Sites of SpaceX rockets

Results - Interactive Map with Folium

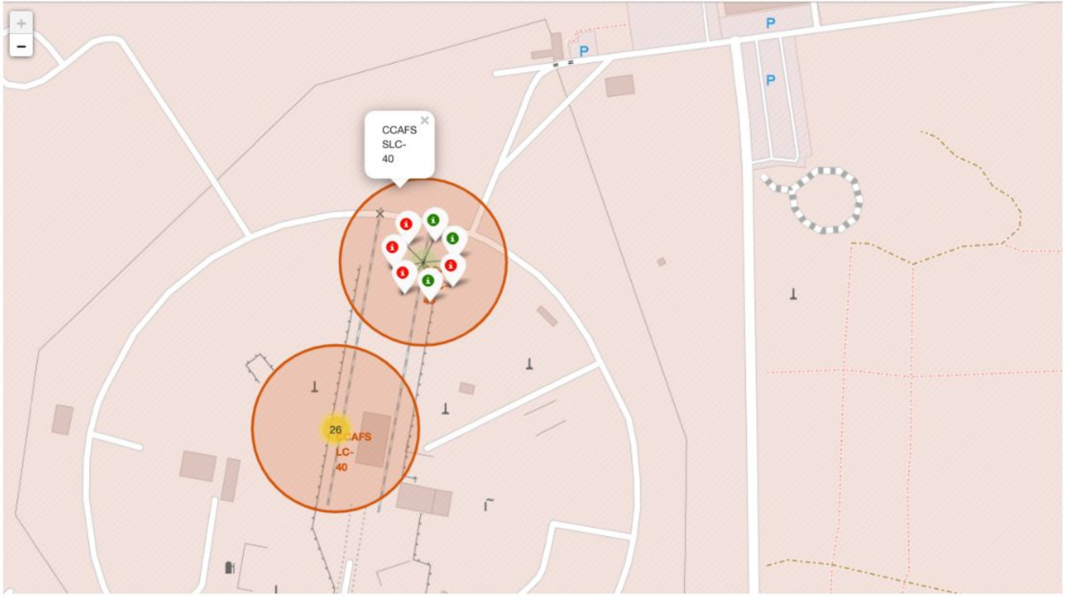
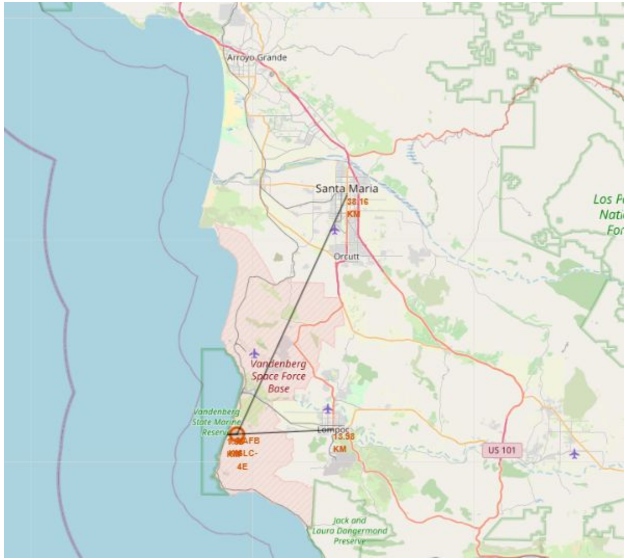


Figure 12:Color labelled Markers. Green Markers display successful launches whereas Red Markers display failed launches.

Results - Interactive Map with Folium



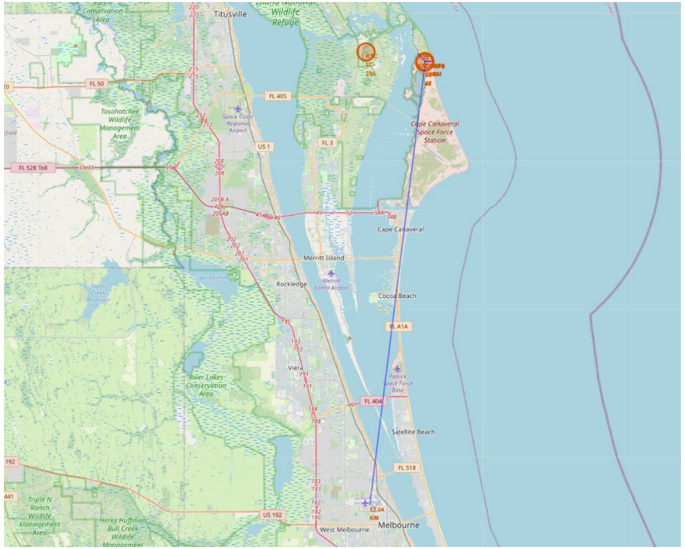


Figure 13:Launch sites distances relative to landmarks such as coast lines, , highways, railways, cities.

Results - Plotly Dash Dashboard

Results - Predictive Analysis (Classification)

- An 83.33% accuracy on test data is achieved when using a decision tree classifier.
- While the remaining two (2) classification methods still have a high accuracy (KNN - 0.6536, Logistic Regression - 0.6679), the decision tree displays the most desirable results at 0.6680.

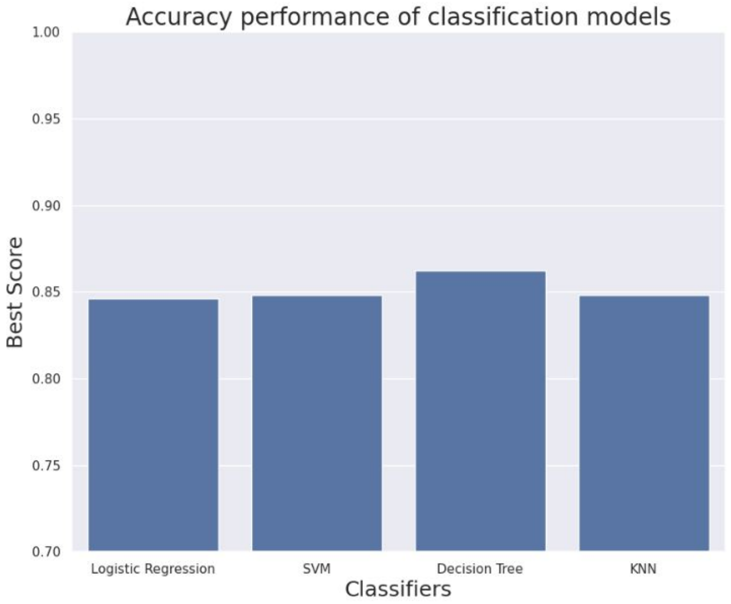
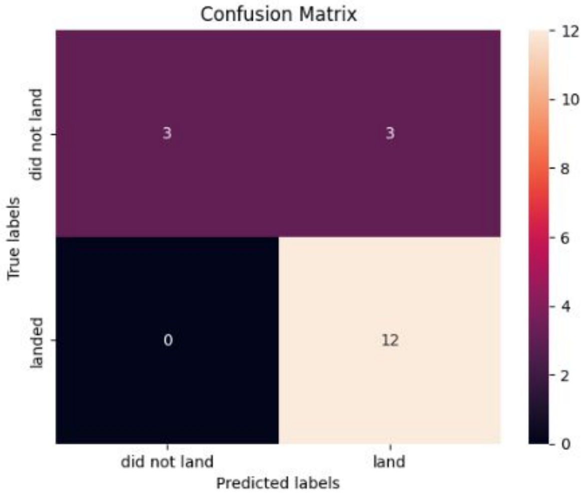


Figure 14: Bar Chart for Classifiers vs Best Score.

Results - Predictive Analysis (Classification)

- Decision tree can distinguish between various classes.



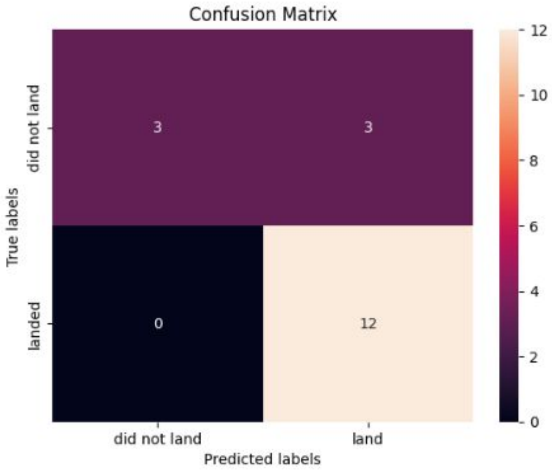
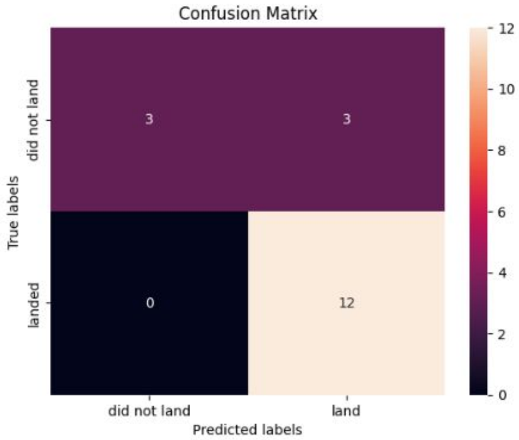


Figure 15: Confusion Matrix charts Part 1

Results - Predictive Analysis (Classification)

- The main challenge is obtaining false negatives.



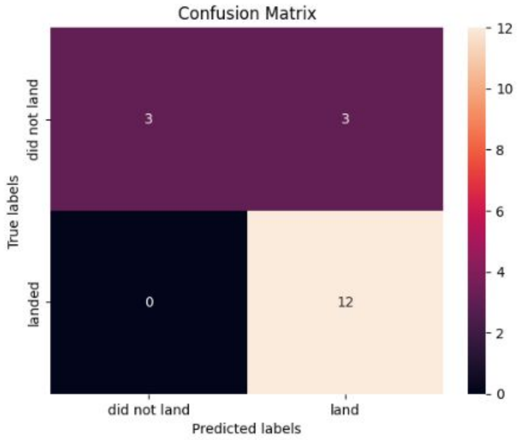


Figure 16: Confusion Matrix charts Part 2

Conclusion/Insights

- The lower the Payload the higher success landing rates are exhibited.
- KSC LC-39A launch site has the most success rates compared to the other launch sites.
- Orbits GEO, SSO, ES-L1 and HEO have the best success rates.
- The Tree Classifier algorithm for machine learning is the best model for this dataset.

References

[1] https://www.coursera.org/learn/applied-data-science-capstone/home/module/1

[2] https://www.spacex.com/

[3]https://www.alteryx.com/glossary/data-wrangling#:~:text=Data%20wrangling%20is%20the%20process,also%20sometimes%20called%20data%20munging.

[4] https://realpython.com/python-folium-web-maps-from-data/

[5]https://plotly.com/python/v3/create-online-dashboard-legacy/#:~:text=A%20dashboard%20is%20a%20collection,plots%2C%20text%20and%20webpage%20images.