

The background of the slide is a deep space image featuring a dense field of stars of various colors (white, yellow, blue) against a dark blue and black sky. A prominent nebula with blue and purple hues is visible in the upper left quadrant. Faint, white, wavy lines resembling gravitational field lines or orbital paths are overlaid on the star field.

SpaceX Falcon 9 First Stage Analysis

By: Vikrant Singh

Executive Summary

- + 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.

Introduction

- + Data gathering or collection using web scrapping techniques and the data is cleaned, pre-processed and transformed into a well-structured data set which can be used to figure out some useful insights from it.
- + Also, finding out the parameters that are responsible for the success or failure of SpaceX's Falcon 9 first stage landing.
- + From these Parameters, multiple classification predictive model will be built to predict the future outcome for the Falcon 9's landing and all are compared to figure out the most accurate model to predict these outcomes. It will help the bidders to bid accordingly on the projects which has higher success possibilities.

Data Collection Methodology

+ Data is collected from the Wikipedia's and SpaceXdata's websites and stored as CSV files:

- + Requests Library

To Get the website's content by sending out a request through a URL.

- + BeautifulSoup Object

To parse this response file in Html format, and using built-in tag functions to get the table required for the analysis of SpaceX falcon 9's first stage landing.

- + Pandas Library

To create a Data Frame for manipulating or updating of this tabular dataset we extracted.

Data Wrangling Methodology

- + Data Wrangling or cleaning is required to remove the redundant, missing or false data in the dataset.
- + Removing Parameters not important for the analysis.
- + Handling missing values by imputing mean.
- + Normalizing Data
- + Data formatting by converting Object datatype to Int/Float datatype.

EDA and Visualization Methodology

EDA helps understand data better by summarizing main characteristics of the data and uncovering the relationships between variables.

- Descriptive Statistics
 - `Describe()`, `value_counts()`, `pivot()`, `groupby()`, `dtype()`, `isnull()` etc.
 - `Mean()`, `median()`, `mode()`, `var()`, `std()`, etc.
- Visualization using `matplotlib`, `seaborn`, `plotly`.
 - Box Plot, Scatter Plot, Heat Map, Histogram, etc.

EDA and Visualization Methodology

- EDA with SQL
 - Using SQLite and SQL magic library of Python, data can be visualized and analyzed by running queries to get the desired result.
- EDA with Folium Maps
 - Folium is used to create interactive maps to analyze the geographical parameters in the data set. Markers and Polygons can be created.
- EDA with DASH by Plotly
 - Dash is used to create web based interactive Dashboards for data visualization using multiple charts and graphs.

Predictive Analysis Methodology

- Discovering Target Variable
 - If Data is labeled or has Past outcomes for target variable
 - If Continuous, Regression models can be used.
 - If Categorical, Classification models can be used.
 - If Data is not labeled,
 - Clustering Models can be used.

Scikit-Learn Library of Python is used to Preprocess Data, Create and analyze Predictive Models, dividing data for training and testing and many other functionalities for machine learning.

EDA and Visualization

Target variable

+ Class: 0 for Failure, 1 for Success

Average success rate of all landings outcomes for Falcone 9' First stage is 67%

Number of launches on each Launch sites,

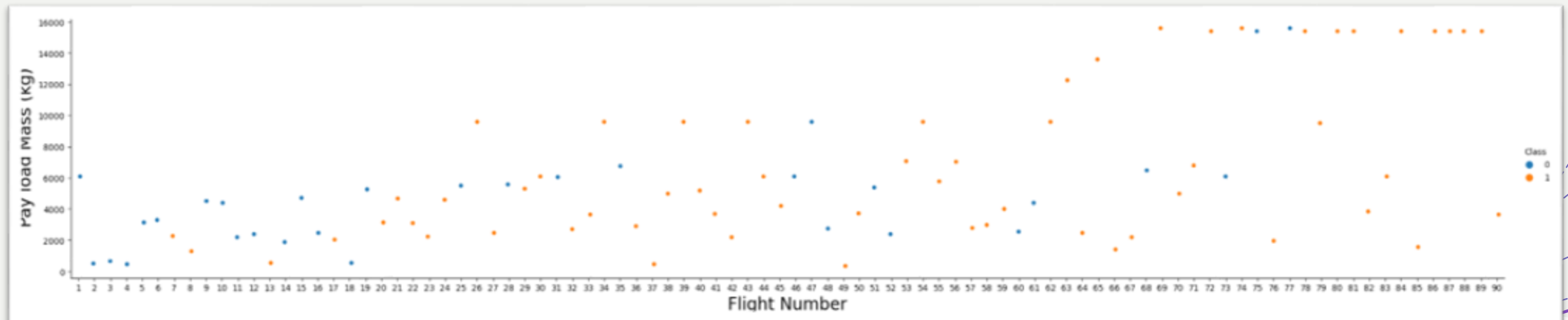
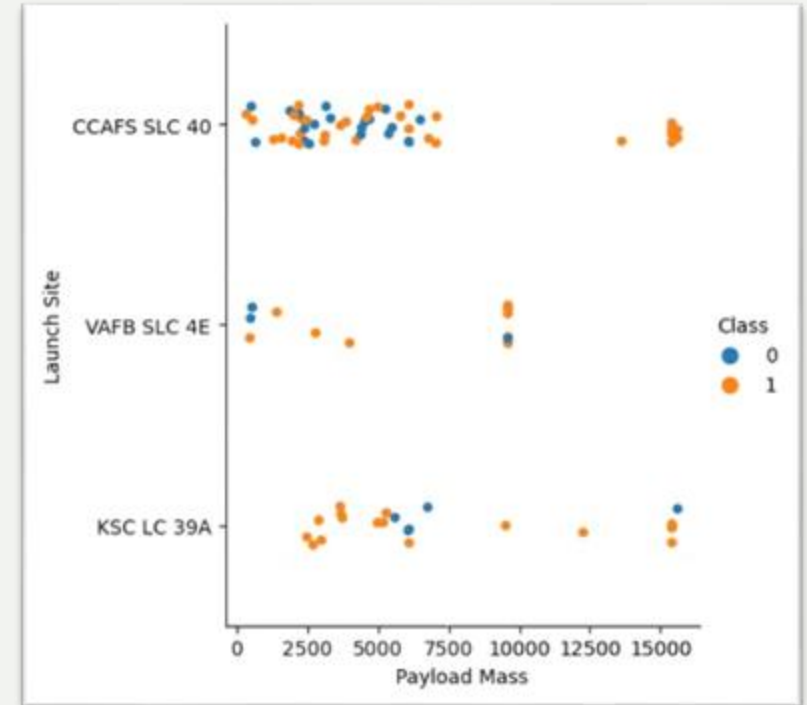
- CCAFS SLC 40 ---- 55
- KSC LC 39A ---- 22
- VAFB SLC 4E ---- 13

In Launch Site vs Payload Mass scatter plot,

- As the Payload Mass increases, Success rate of landing increases.

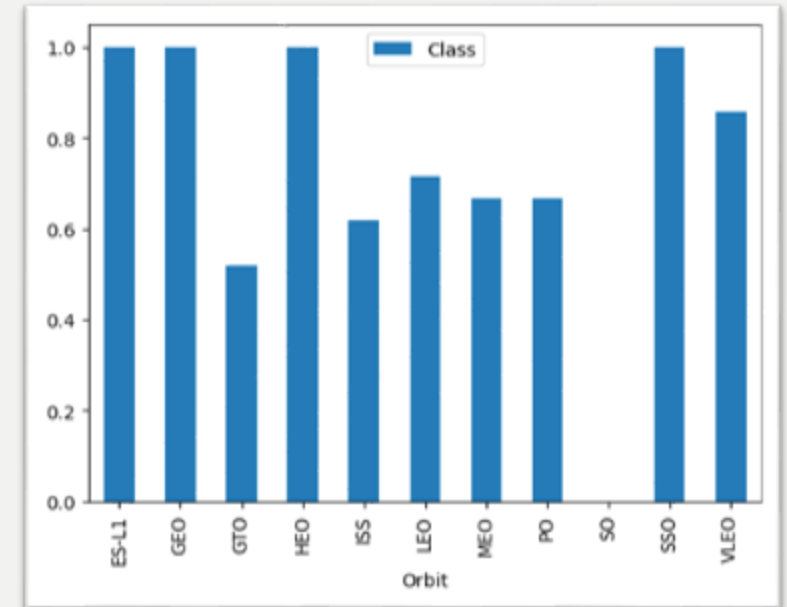
In Payload mass vs Flight number scatter plot,

- Launch Site VAFB SLC 4E does not launch higher than 10,000 kg of Payload.
- Success rate of landing is 100% at CCAFS SLC 40 with Payload higher than 12,500.



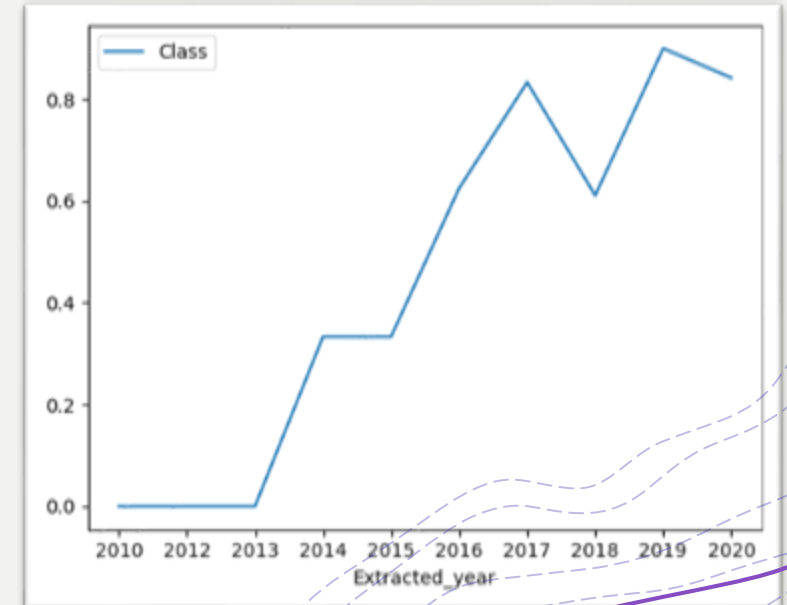
In Orbit vs Class Bar Graph,

- ES-L1, GEO, HEO, SSO has 100% landing success rate, that shows that all the first stage of Falcon 9' are successfully landed on either land, water or drone ship.
- SO has no successful landing. It has only one occurrence and that is a failure.



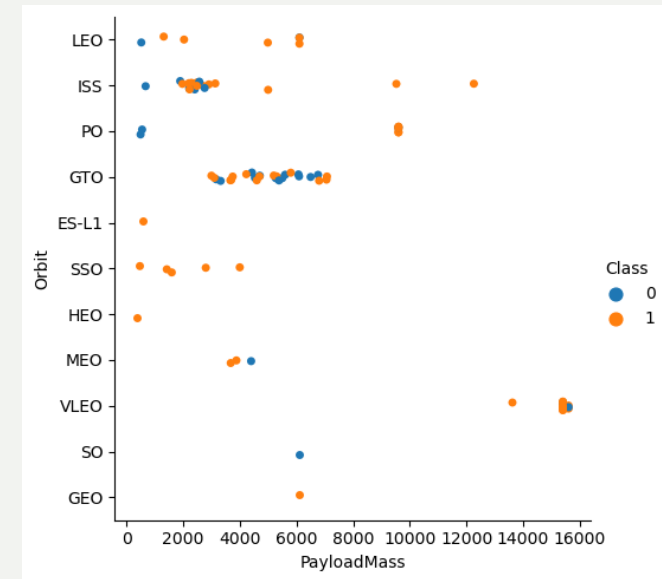
In Years vs Class Line plot,

- We can see that as the years pass by, the landing success rate increases shows the improvement done in the past years
- In the first 4 years, 2010-2013, There was no successful landing of the Falcon 9' First Stage was recorded.



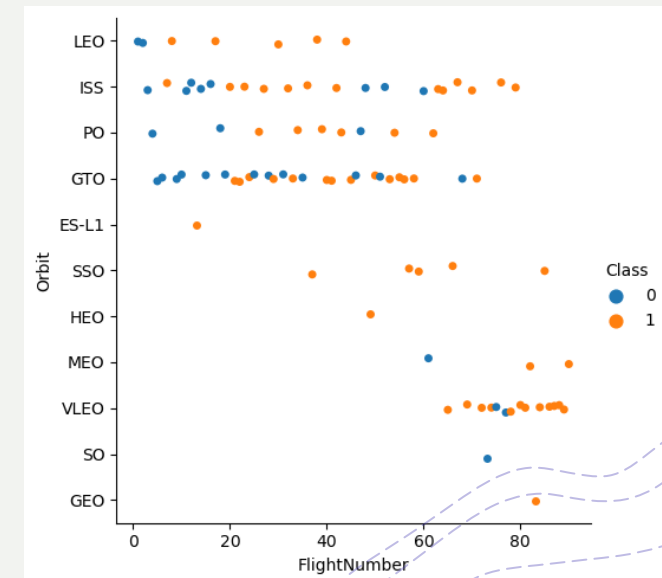
In Orbit vs Payload mass Scatter Plot,

- LEO, ISS, PO has 100% landing success rate for High Payload mass value, that shows that all the first stage of Falcon 9' are successfully landed on either land, water or drone ship.
- VLEO has only High payload mass landings and 0 low Payload mass landing



In Orbit vs Flight Number Scatter Plot,

- Leo has success appeared for different different Flight numbers, but on the other hand GTO has no relation between Flight numbers and successful landing possibility.



EDA with SQL

Task 1: Getting names of all the Launch Sites.

Task 2: Display 5 records where launch sites begin with the string 'CCA'.

Launch_Site

CCAFS LC-40

VAFB SLC-4E

KSC LC-39A

CCAFS SLC-40

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MAS S_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

Task 3: Display the total payload mass carried by boosters launched by NASA (CRS)

avg_payload_mass
2928.4

Task 4: Display average payload mass carried by booster version F9 v1.1

total_payload_mass
45596

Task 5: List the date when the first successful landing outcome in ground pad was achieved.

min(Date)
01-05-2017

Task 6: 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
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Task 7: List the total number of successful and failure mission outcomes.

Task 8: List the total number of successful and failure mission outcomes.

Mission_Outcome	count(Date)
Success	101

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

Task 9: List the records which will display the month names, failure landing outcomes in drone ship ,booster versions, launch site for the months in year 2015.

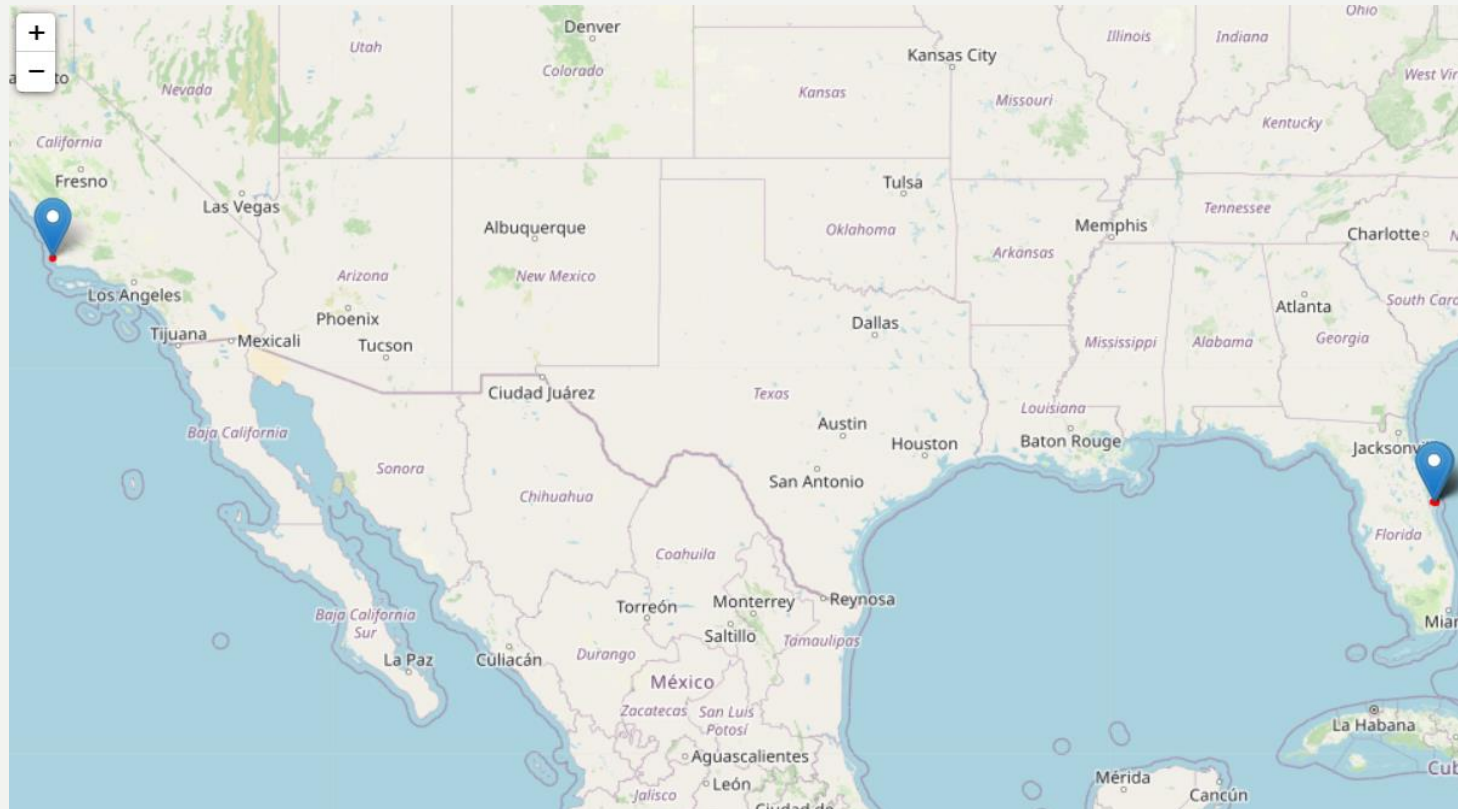
month s	Landing _Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFLC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFLC-40

Task 10: Rank the count of successful landing outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

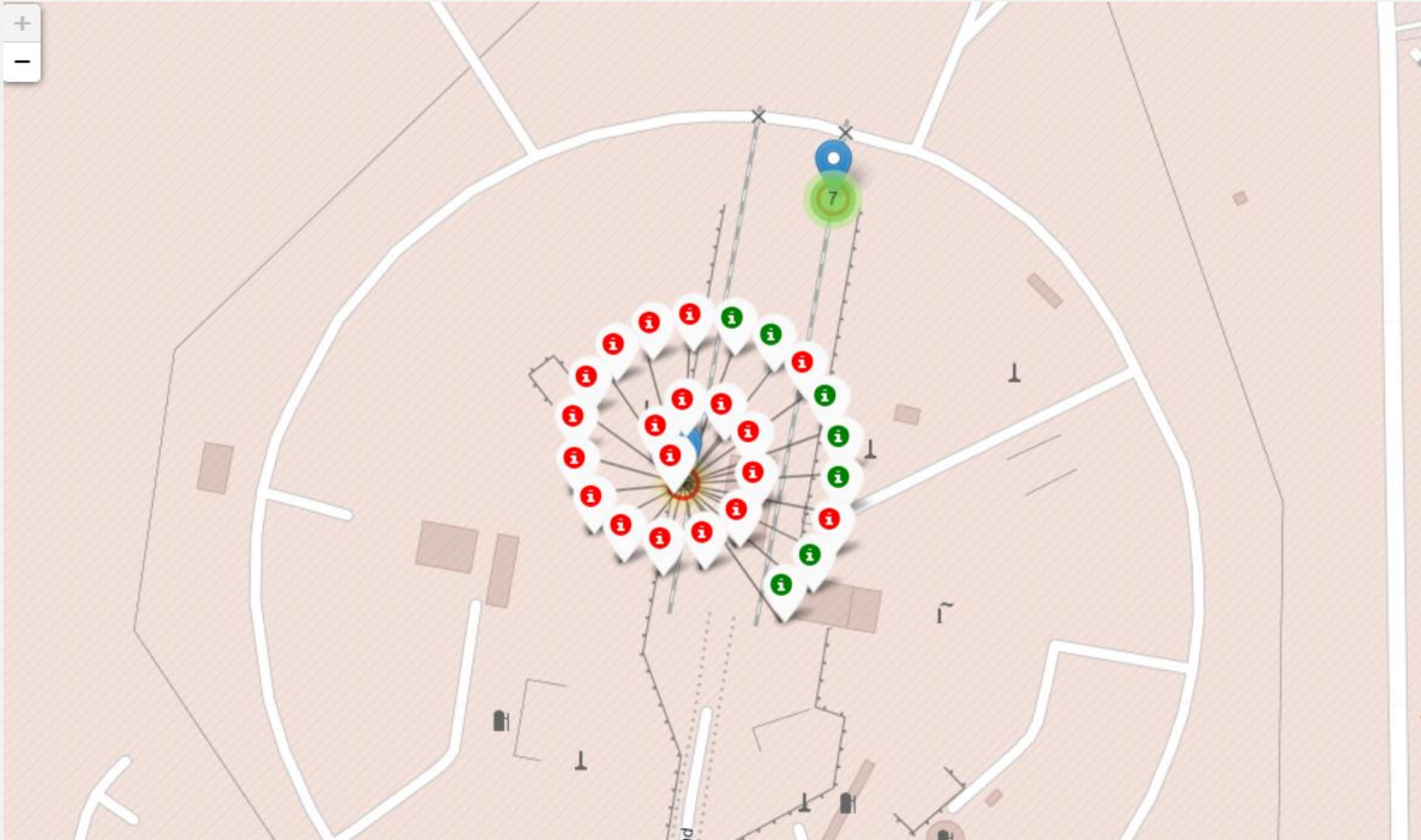
Landing_Outcome	count_total
Success	20
Success (drone ship)	8
Success (ground pad)	6

EDA with Folium

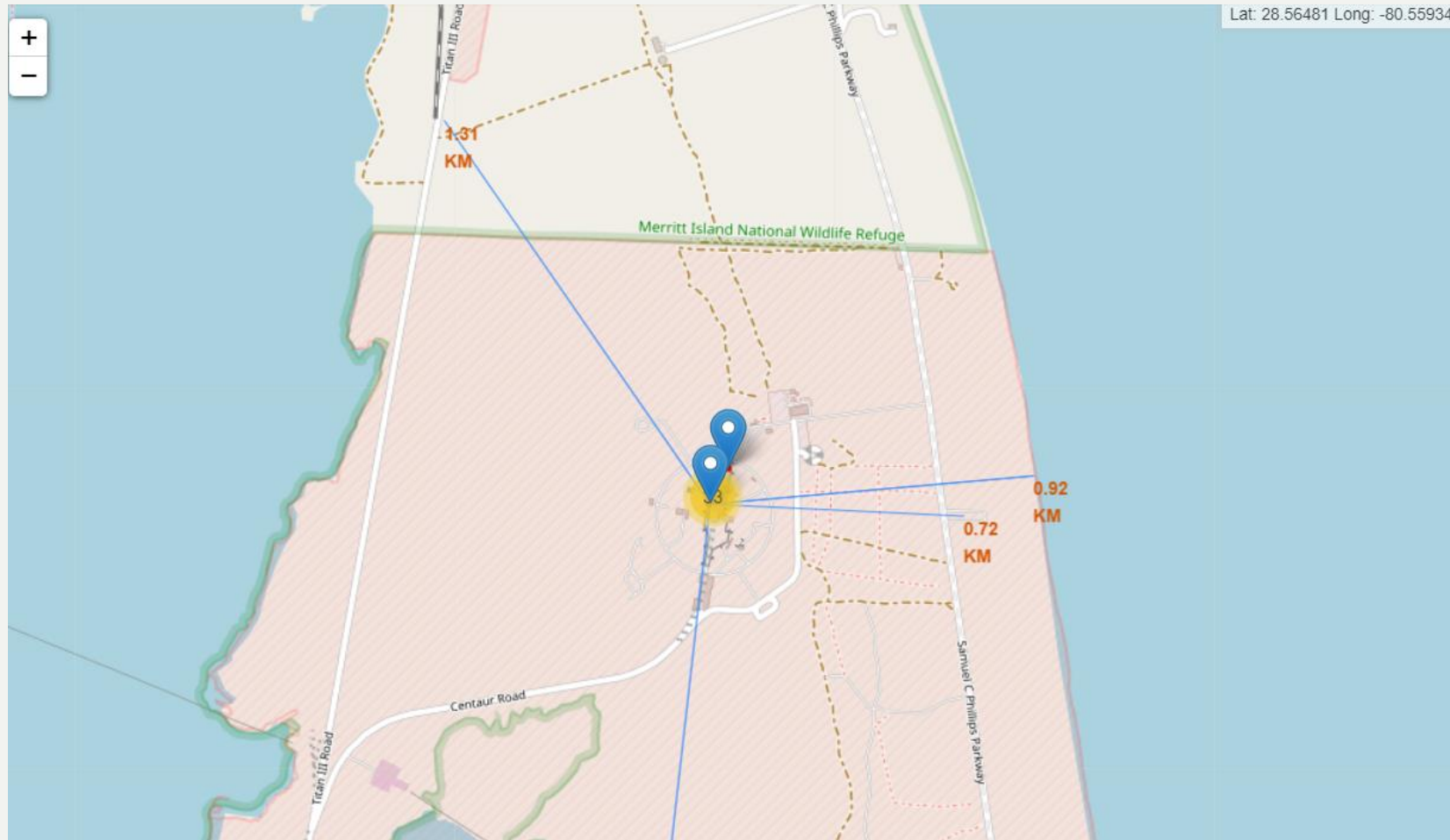
Task1: Creating Markers for each launch sites.



Task2: Mark the Success/Failure on each launch and creating clusters for each Launch site.

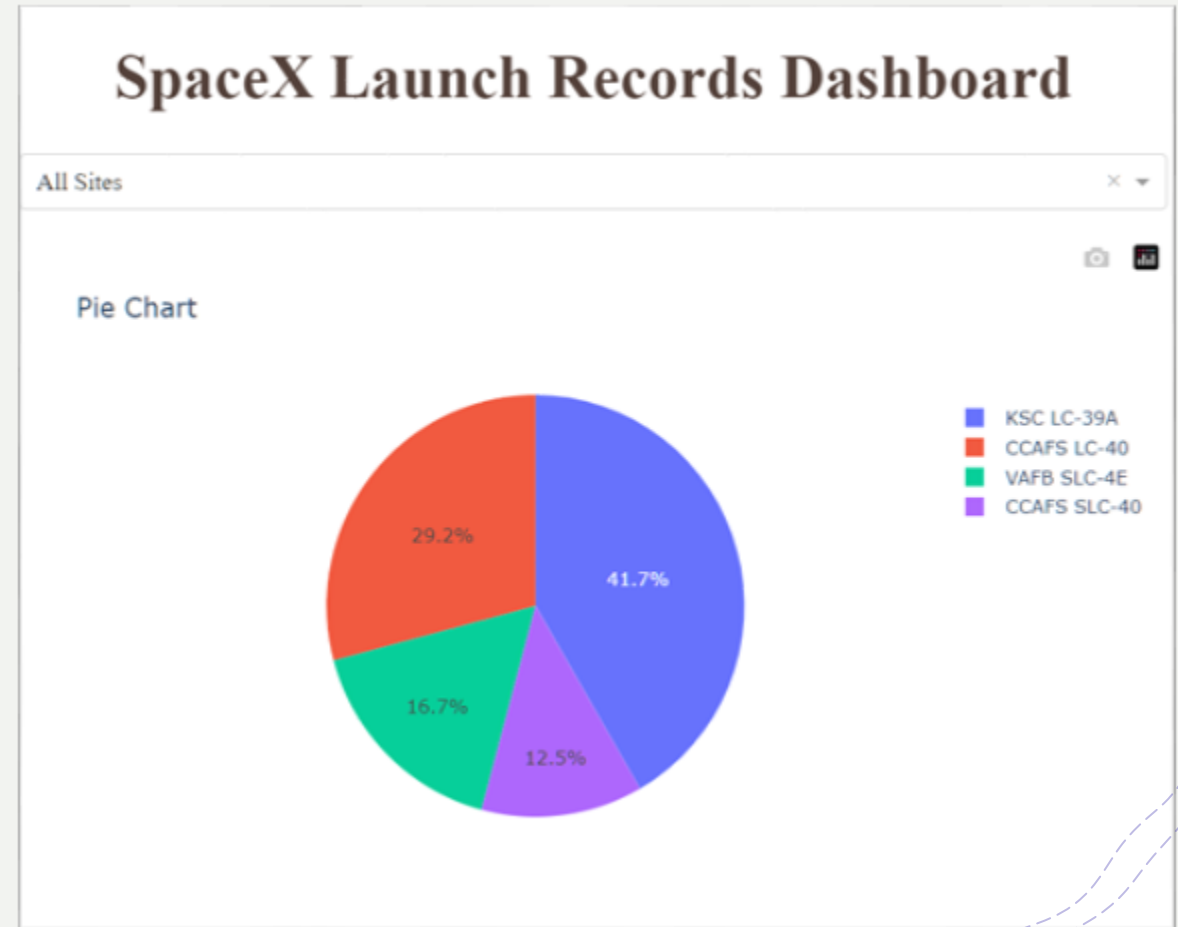


Task3: Calculating Distance from nearest Highway, Railway, City, Coastline.



Dash Application

- Launch site that has the largest successful launches is CCAFS LC-40.
- Launch site that has the highest launch success rate is CCAFS SLC-40 with 42.9%.



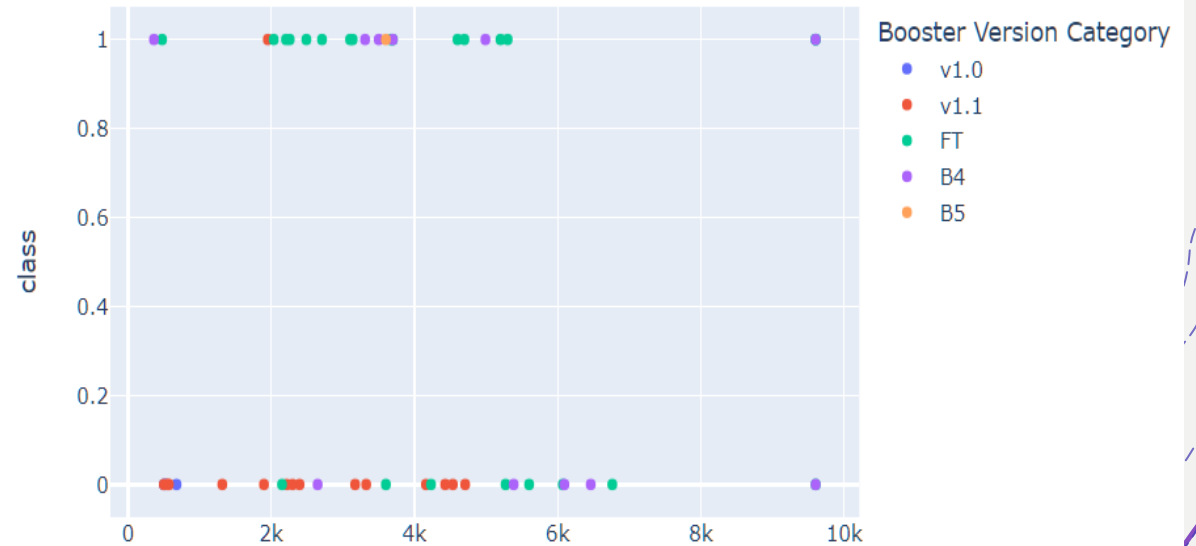
Dash Application

- Payload range(s) that has the highest launch success rate is 3000Kg-4000Kg.
- Payload range(s) that has the lowest launch success rate is 6000Kg-7000Kg.
- F9 Booster version (v1.0, v1.1, FT, B4, B5, etc.) that has the highest launch success rate is FT Booster Version.

Payload range (Kg):



Correlation between Payload Mass and Success for All sites



Predictive Analysis

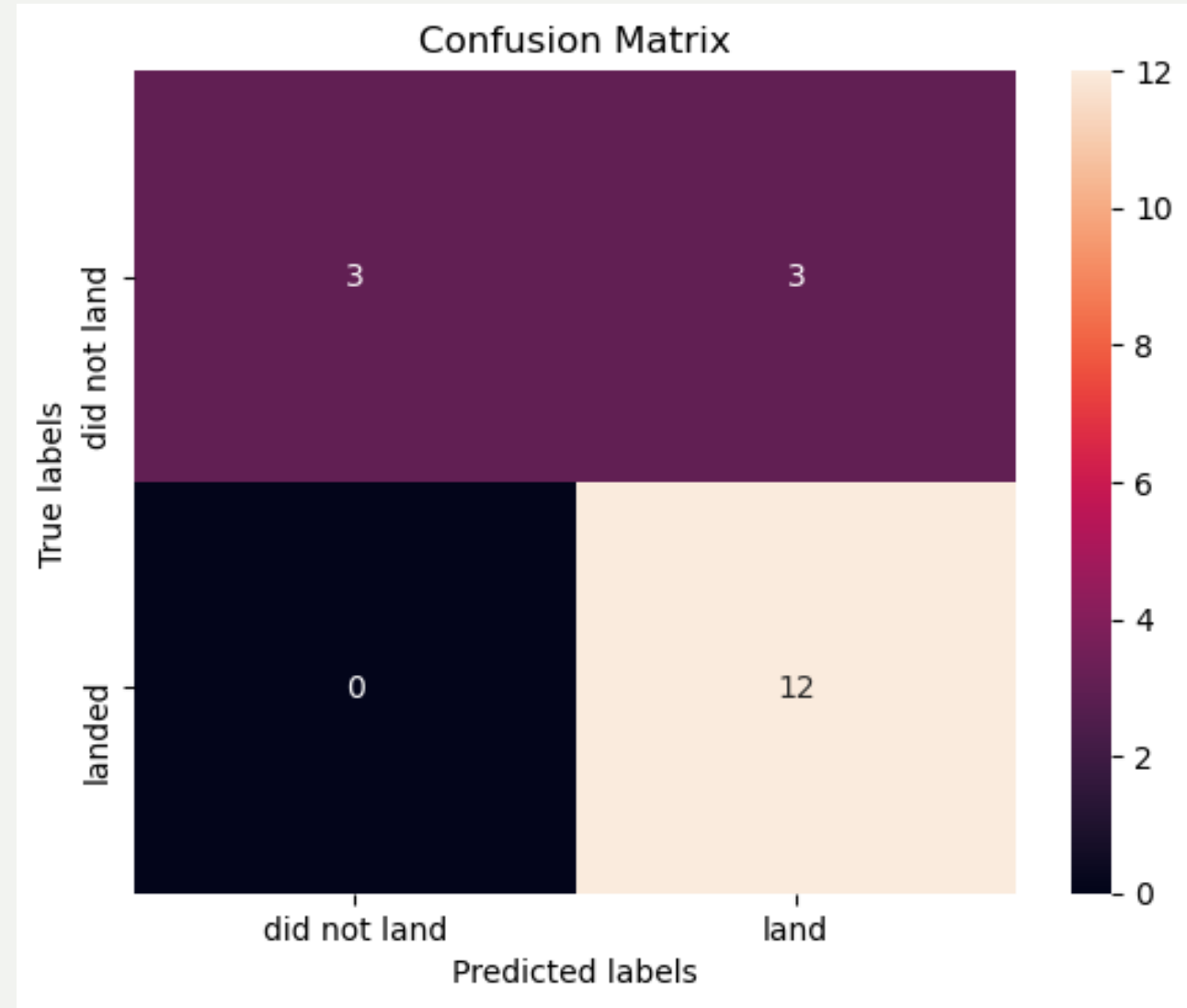
As Target Variable "Class" is a binary categorical variable, therefore Classification Model will be suitable for Predictions.

Some of the Classification Models that can be created and compared are:

- + Logistic Regression Model
- + SVM
- + Decision Tree
- + KNN

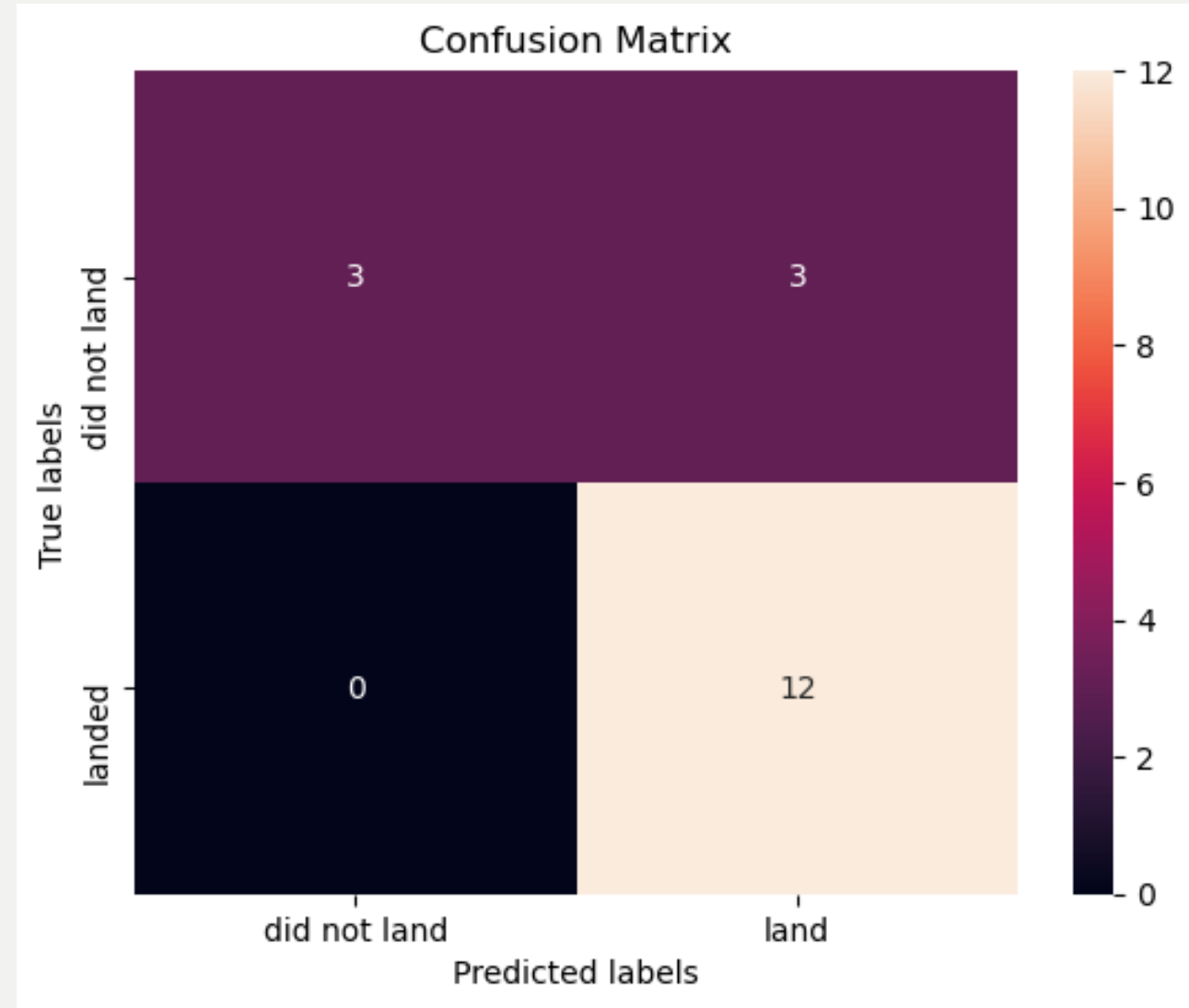
Logistic Regression Model

- + Best Accuracy = 0.847
- + Score on Test Data = 0.833
- + Confusion Matrix:



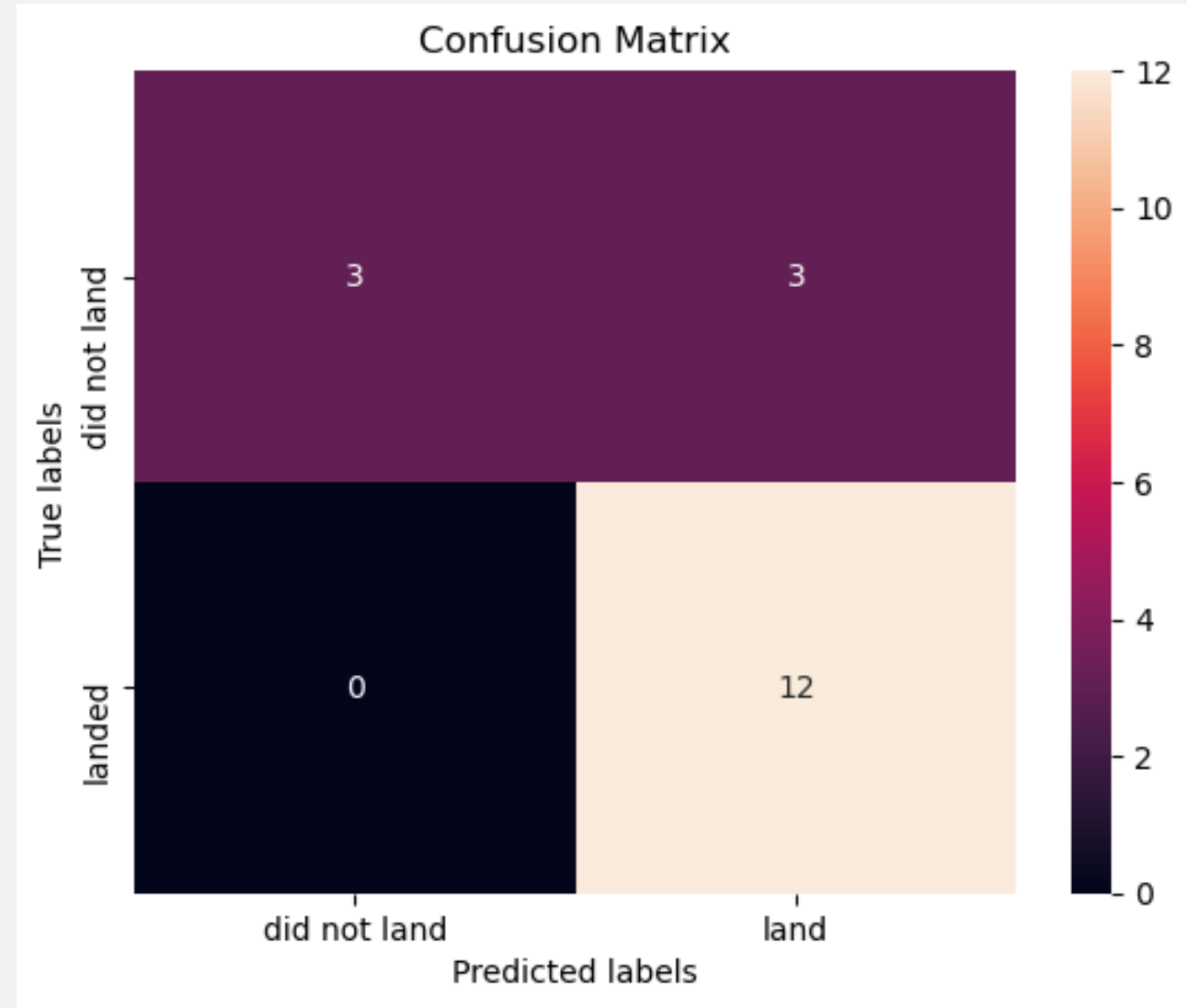
SVM Model

- + Best Accuracy = 0.847
- + Score on Test Data = 0.833
- + Confusion Matrix:



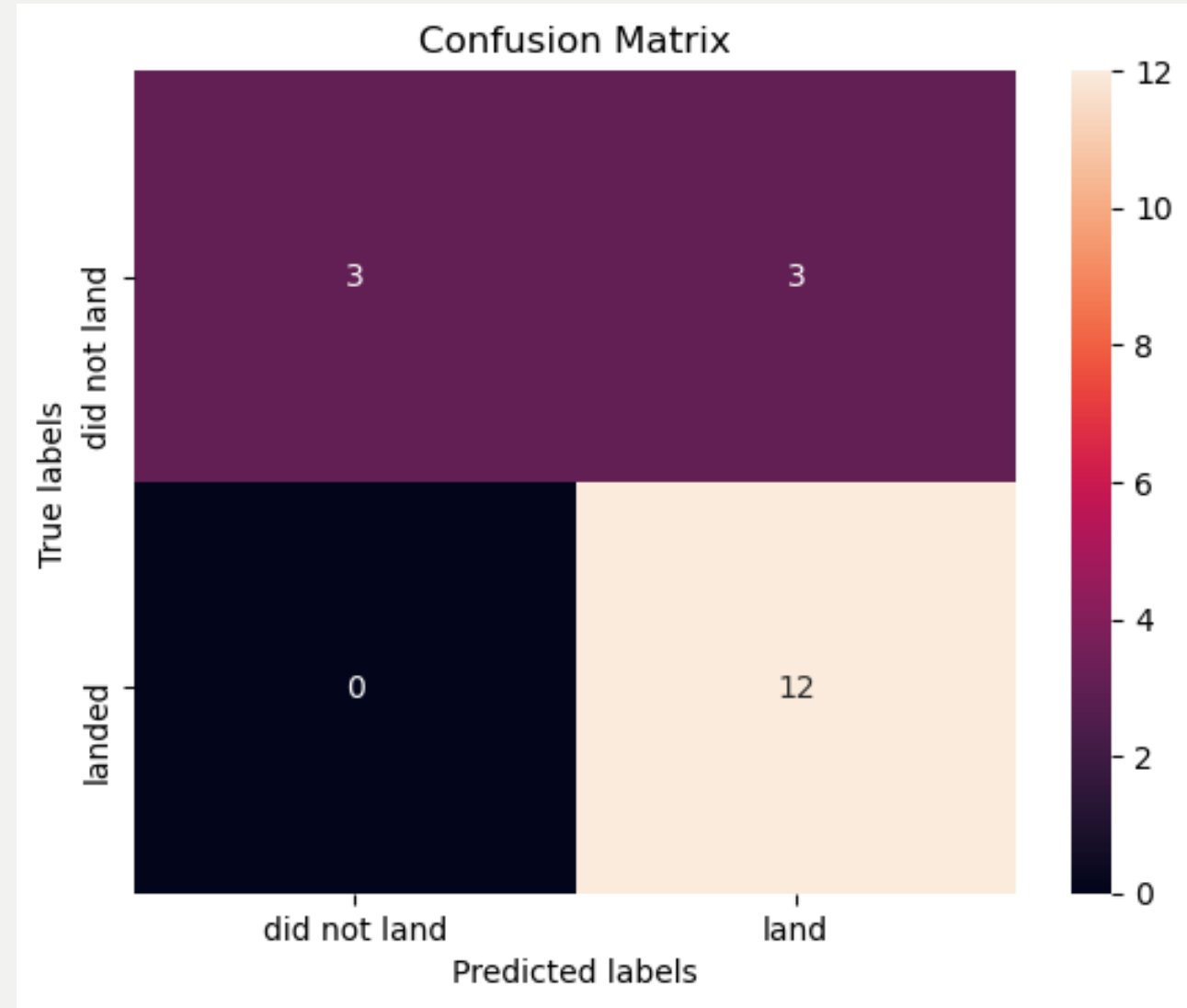
Decision Tree Model

- + Best Accuracy = 0.875
- + Score on Test Data = 0.833
- + Confusion Matrix:



KNN Model

- + Best Accuracy = 0.847
- + Score on Test Data = 0.833
- + Confusion Matrix:



Best Model

	Logistic Regression	SVM	Decision Tree	KNN
Best Accuracy	0.847	0.847	0.916	0.847
Score On Testing data	0.833	0.833	0.778	0.833

Decision Tree has the best Accuracy of 0.916, therefore can be selected as the best model for predicting Falcon 9's Successful landing chances more accurately than other models.

Conclusion

Decision Tree Model can be used to predict the future possibilities of successful landing of Falcon 9's First stage on either ocean, drone ship or ground pad.

There will be 91.6% chances of the prediction to be correct.

Additional Analysis

Launch Site VAFB SLC- 4E has more than 60% Failure rate and is in close proximity of the Lompoc Airport and Railway lines, if a failure with a significant measure happens, there are chances of big loss for the state properties.