Accessing the Success of Falcon 9's First-Stage Landing in Space X Rocket launches

JOKOSENUMI SANUSI DAMILARE

18/04/2023

IBM CAPSTONE PROJECT

OUTLINE

1. EXECUTIVE SUMMARY
2. INTRODUCTION
3. METHODOLOGY
4. RESULTS
5. DISCUSSION
6. CONCLUSION

Executive Summary

This report presents the result of a data science project that aimed to gain insights into Space X launches data , to build, evaluate and refine predictive models for discovering more insights.

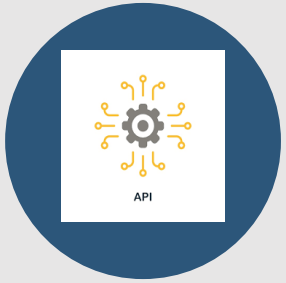
The analysis was conducted on a dataset of Space X launches spanning over the years from 2010 to 2021

INTRODUCTION

SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars, while other providers cost upward of 165 million dollars each, much of its savings is because SpaceX can reuse the first stage. So, if we can determine if the first stage will land, we can determine the cost of a launch.

What we want to know is, what are the factors that affect the most how successful the first stage landing will be and with this have the capability to determine the cost of launch

METHODOLOGY



DATA COLLECTION
SPACEX REST API

Using the public API that SpaceX offers, data was obtained by performing the following steps:

- Perform a request to the API
- Parse the response data to extract relevant information
- Store the data in a dataframe
- Filter the data for the Falcon 9 launches
- Replace null values

Source: API Data Collection

DATA COLLECTION
WEB SCARPING

Wikipedia contains information for SpaceX launches which can be obtained by performing the following :

- Make a request to Wikipedia
- Parse the HTML content using beautiful soup
- Save the extracted data and load into a dataframe

Source: Webscrapping Data Collection

DATA WRANGLING

The following steps were taken to clean, transform and prepare the data for analysis.

- Data Cleaning: The data was cleaned by looking for and handling missing values
- Data Transformation: The data was transformed by converting categorical variable to numerical. Then a landing outcome is created from the outcome column.

Source: Data Wrangling

In this project, we aimed to analyze the factors that affect the landing outcome of the SpaceX falcon 9's first-stage. As part of the methodology, we used data visualization techniques such as **scatter/cat plots**, **line charts** and **bar chart** to explore the data and gained insights into the relationships between:

- Payload mass and flight number
- Launch site and flight number
- Orbit and flight number
- Launch site and Payload mass
- Payload and Orbit

DATA VISUALIZATION

Source: Visualization

EDA: SQL

I also analyzed the dataset using SQL to extract and transform the data, in order to gain insights.

Analysis was performed using SQL queries to get the following information:

- Name of the unique launch sites in the space mission
- Five(5) records where launch sites begin with ‘CCA’
- Total payload mass carried by booster launched by NASA (CRS)
- Average payload mass carried by booster version F9 v1.1
- Dates when the first successful landing outcome in group pad was achieved.

EDA: SQL (cont’d)

- Names of the boosters which have success in drone ship and have payload mass between 4000 and 6000
- Total number of successful and failure mission outcomes
- Names of the booster versions which have carried the maximum payload mass
- Records which will display month names, failure landing outcomes in drone ship, booster versions, launch site for the months in year 2015
- Count of successful landing outcomes between the date 04-06-2010 and 20-03-2017

Source: Analysis with SQL

INTERACTIVE MAP WITH FOLIUM

To visualize the spatial distribution of the data. An interactive map was created using the python library Folium. The steps taken in creating these maps are as follows:

- Import the required libraries including Folium and pandas
- Loading the data into a pandas dataframe
- Creating a map object using folium.map()
- Adding markers or circles to the map
- Adding additional layer.

Source: Folium Map

Dashboard with Plotly Dash

To provide an interactive and customizable interface for exploring the data, a dashboard was created using Plotly Dash library. The steps involved are as follow:

- Importing the necessary libraries including pandas, dash, dash components (html and core, dash dependencies
- Defining the layout and structure of the dashboard such as the dropdown menus, sliders and graphs
- Defining the callbacks for each component of the app
- Finally, the app was ran using app.run_server()

Source: Dashboard

Predictive Analysis: Classification

Predictive analysis was performed on the dataset using machine learning techniques. Four classification models were used; Logistic Regression, Support Vector Machine, Decision Tree and K Nearest Neighbors.

With the following steps, predictive analysis was performed with these models:

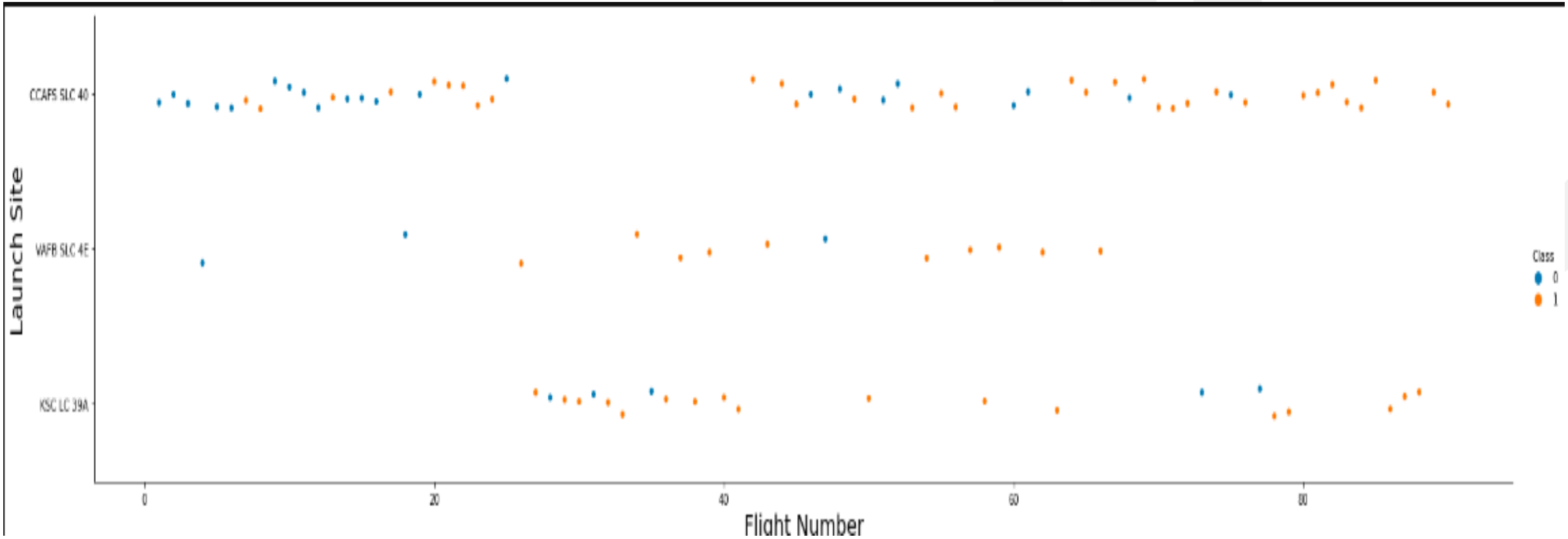
- Data preparation and standardization
- Split data into training and test sets
- Evaluate the models using GridSearch
- Obtain accuracy of the model and compared the result

Source: Predictive Analysis

RESULTS

- Exploratory Data Analysis results
  - Visualization
  - SQL
  - Folium Map
  - Dashboard
- Predictive Analysis results
  - Classification

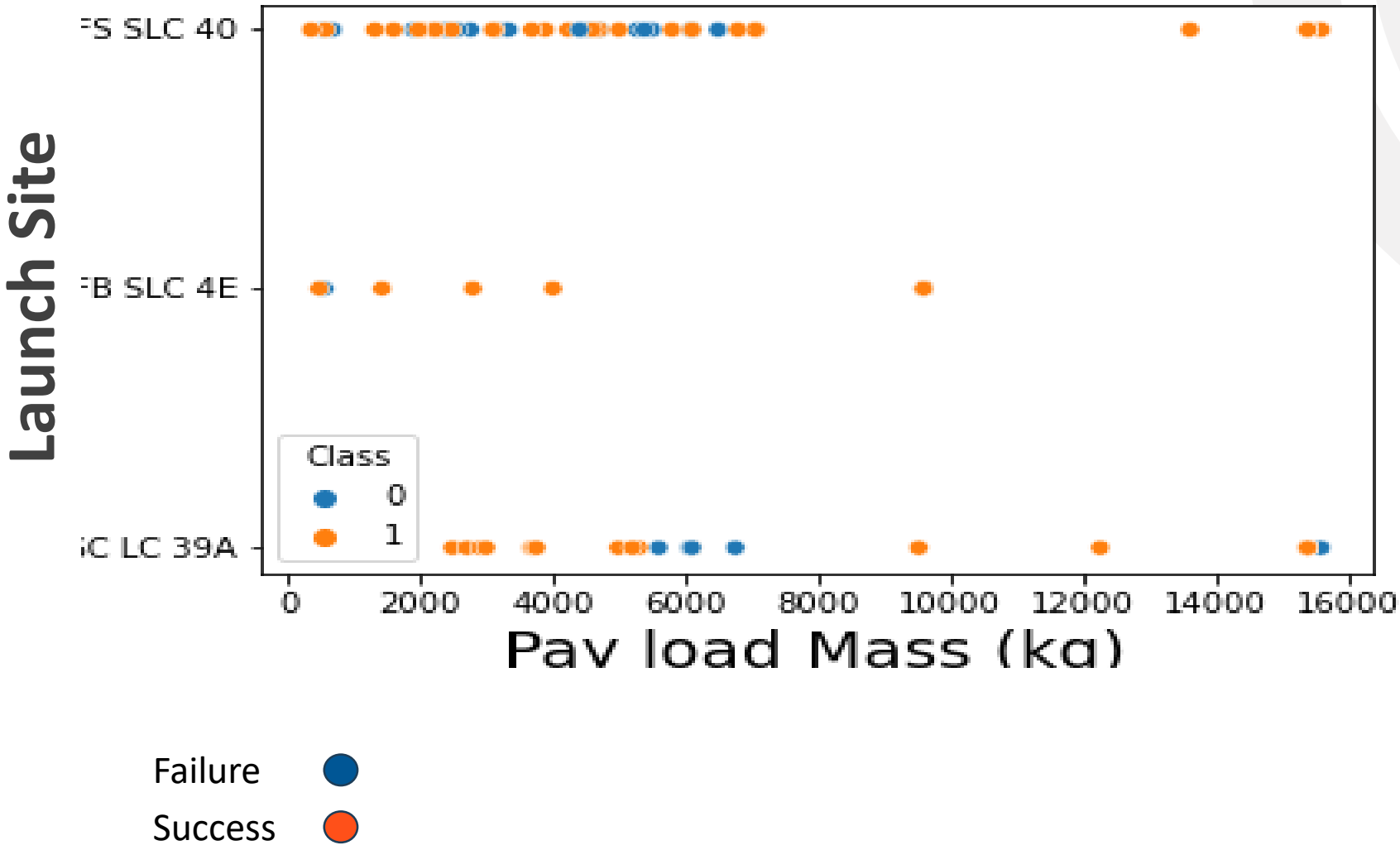
VISUALIZATION



INSIGHTS

- The success rate increased over time in each launch-site
- The launch-site with the highest number of falcon 9 launches is CCAFS LC-40
- The launch site with the highest success rate is KSC LC- 39A

VISUALIZATION



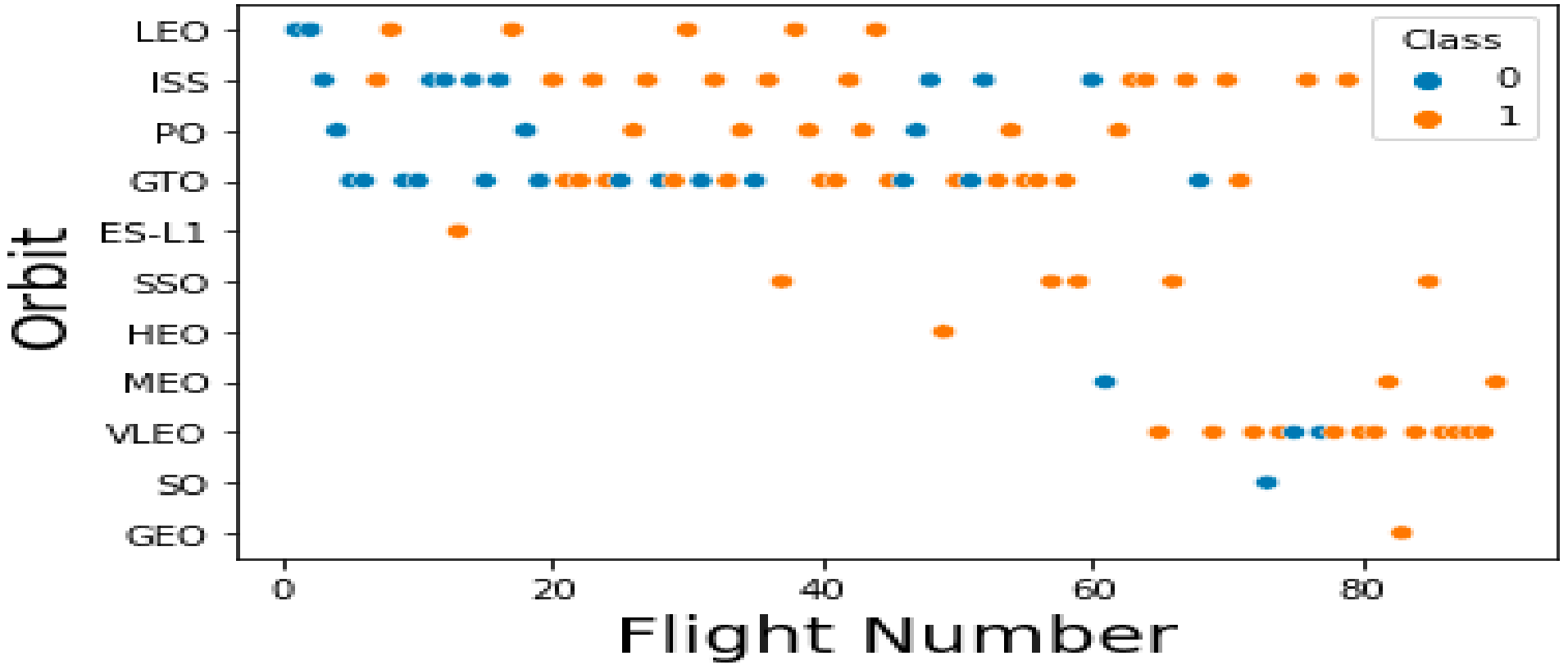
Failure

Success

Insights

- Payloads of over 8000 kg have a high success rate
- VAFB-SLC launch-site there are no rockets launched for heavy payload mass(greater than 10000).
- KSC LC-39A has a high success rate for payloads under 6000 kg

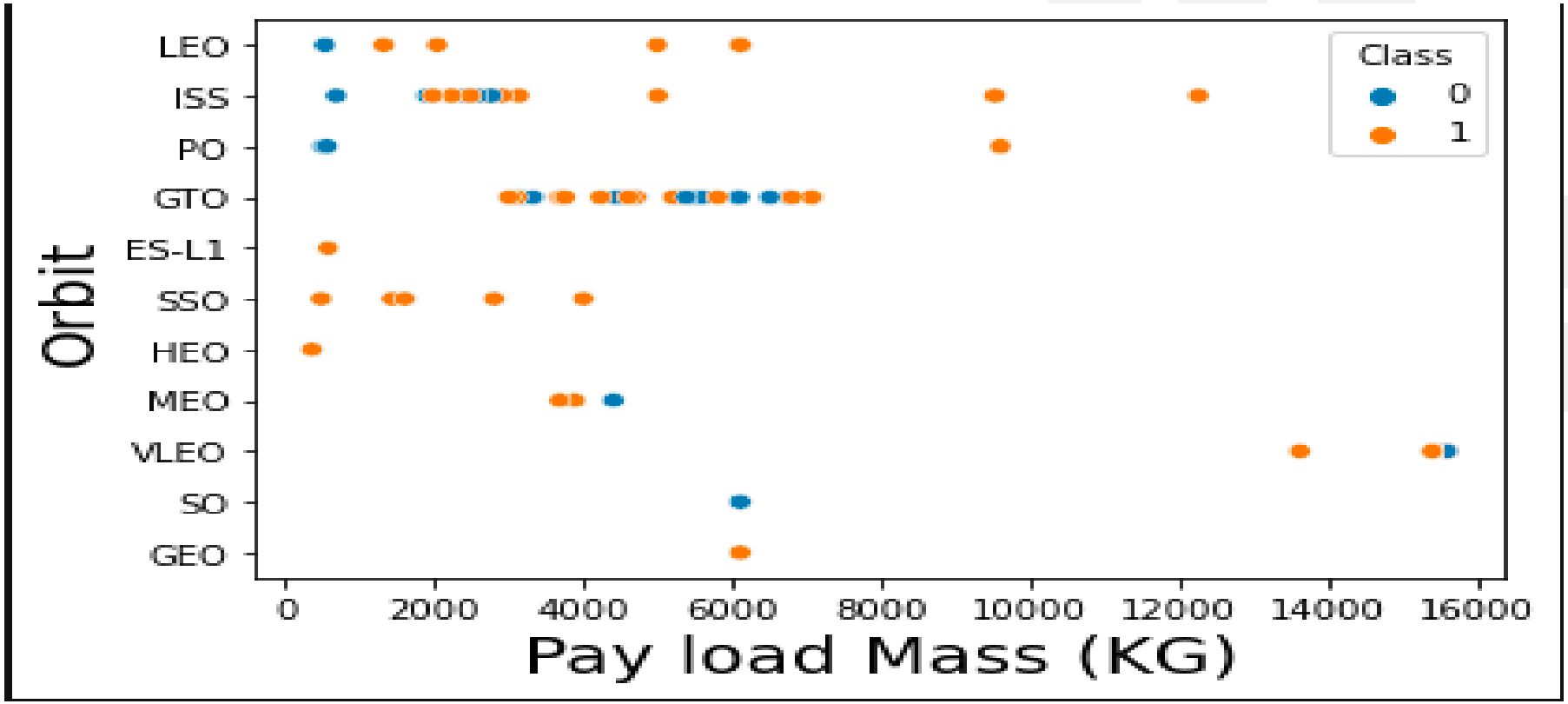
VISUALIZATION



INSIGHTS

- In the LEO orbit the Success appears related to the number of flights
- There seems to be no relationship between flight number when in GTO orbit.

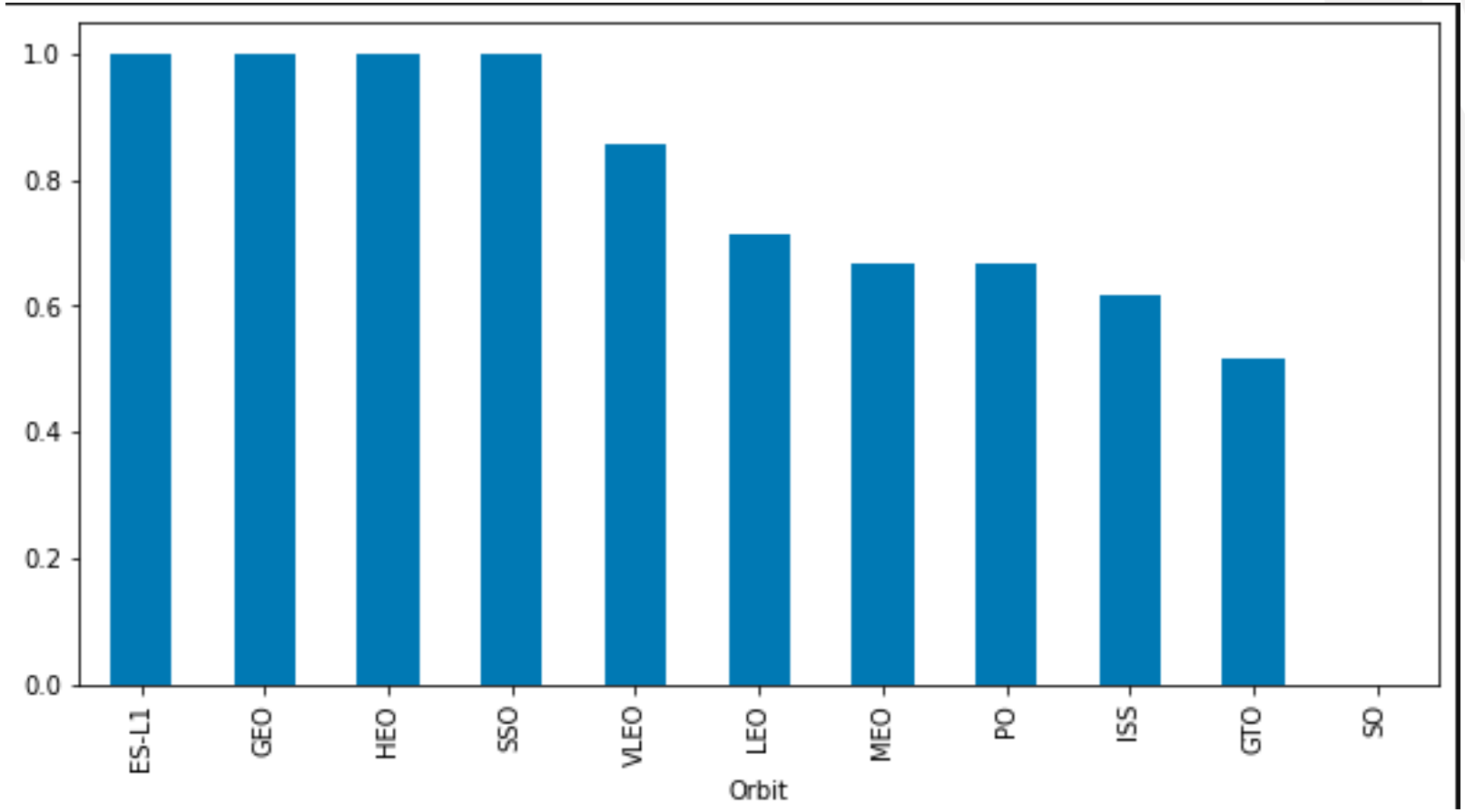
VISUALIZATION



INSIGHTS

- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS
- In GTO we cannot distinguish well between the successful landing and unsuccessful ones.

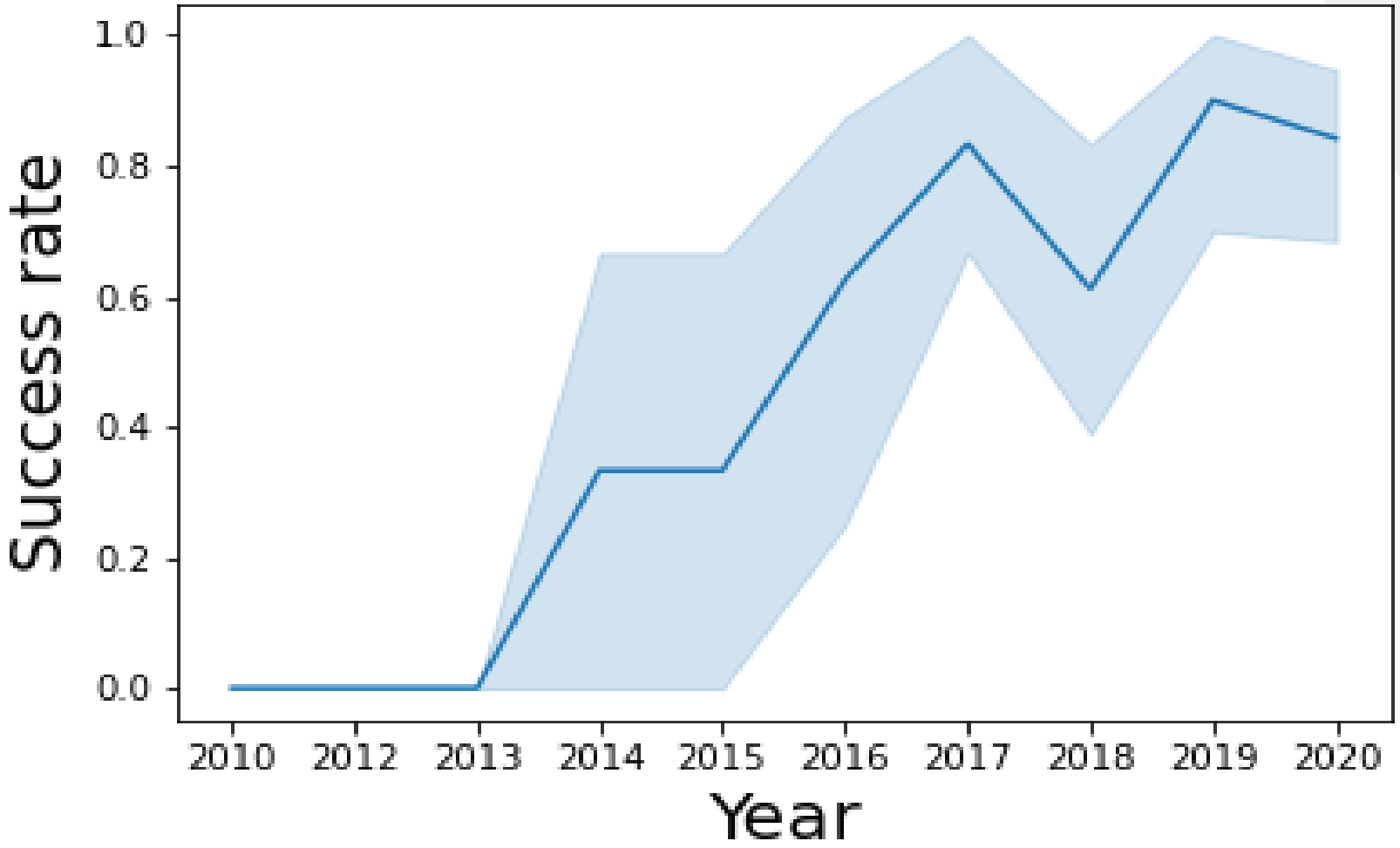
Visualization



INSIGHTS

- ES-L1, GEO, HEO and SSO have a 100% success rate
- SO has 0% success rate

VISUALIZATION



Insights

- Success rate kept on increasing since 2013 to 2020
- There was a dip in the success rate in the year 2018

SQL

5 Records where Launch Sites begin with the string 'CCA'

```
%sql SELECT * FROM SPACEXTBL WHERE launch_site LIKE 'CCA%' LIMIT 5
```

 * sqlite:///my_data1.db
Done.

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS_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

```
%sql SELECT "Landing _Outcome", COUNT(*) AS Total FROM SPACEXTBL WHERE DATE BETWEEN '04-06-2010' AND '20-03-2017'  GROUP BY "Landing _Outcome" HAVING "Landing _
```

 * sqlite:///my_data1.db
Done.

Landing _Outcome	Total
Success	20
Success (drone ship)	8
Success (ground pad)	6

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

SQL

Names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
%sql SELECT DISTINCT(booster_version) FROM SPACEXTBL WHERE "Landing _Outcome" = "Success (drone ship)" AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000
 * sqlite:///my_data1.db
Done.
```

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

```
%sql SELECT MISSION_OUTCOME, COUNT(*) AS Total_number FROM SPACEXTBL GROUP BY Mission_outcome
 * sqlite:///my_data1.db
Done.
```

Mission_Outcome	Total_number
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Total number of successful and failure mission outcomes

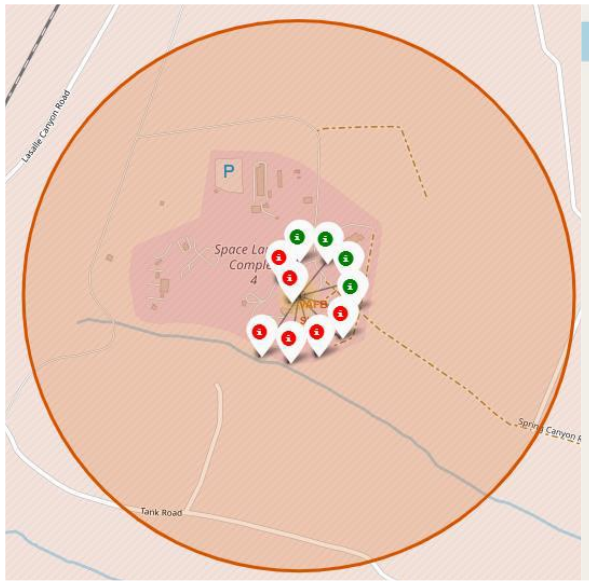
FOLIUM MAP

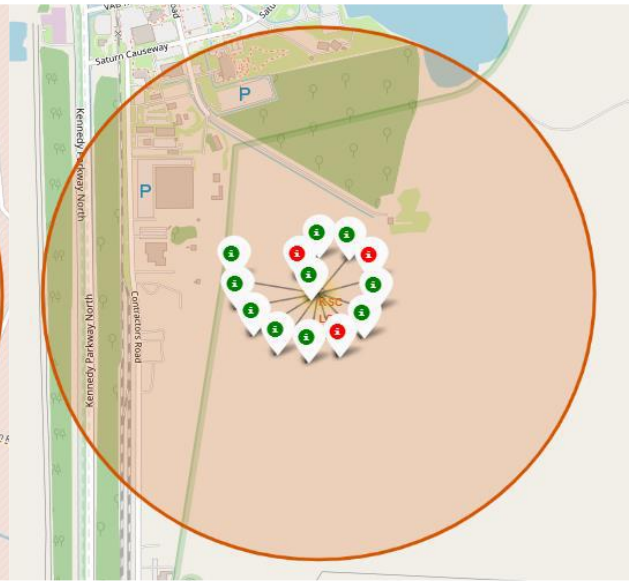
All Launch Sites

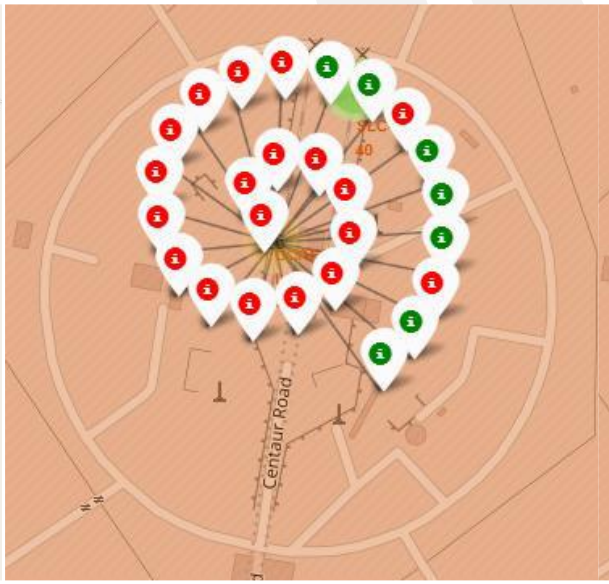
Insights

The map shows the location of each launch site. It can be seen from the map that the launch sites are located near the sea to avoid populated areas

Launch Outcome in each Launch Site





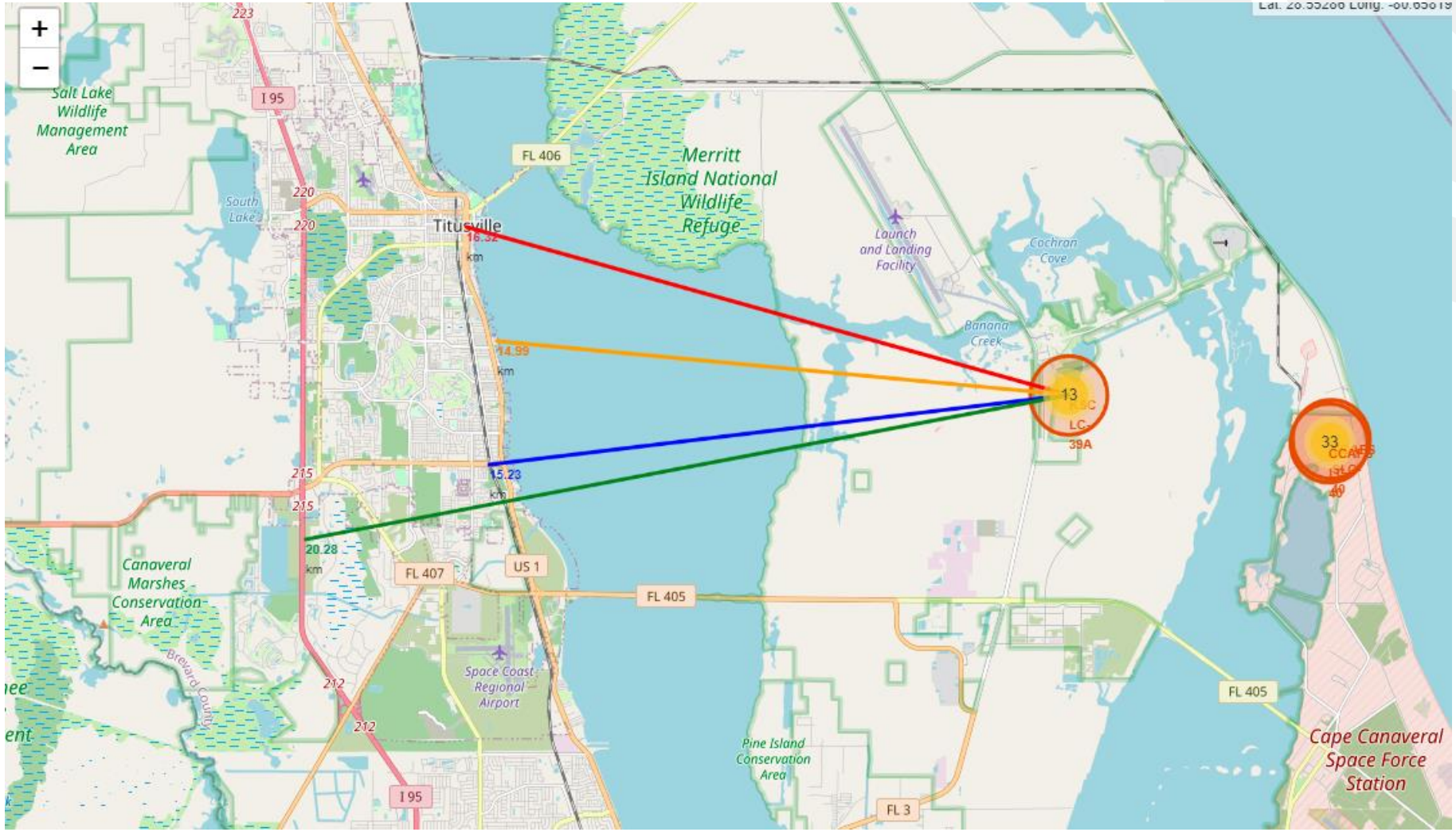


VAFB SLC-4E **KSC LC-39** **CCAFS SLC-40** **CCAFS LC-40**

Failed Launches

Successful Launches

Distance Between KSC LC-39 To The Nearest Highway, Railway, City And Coastline



Railway 15.23km

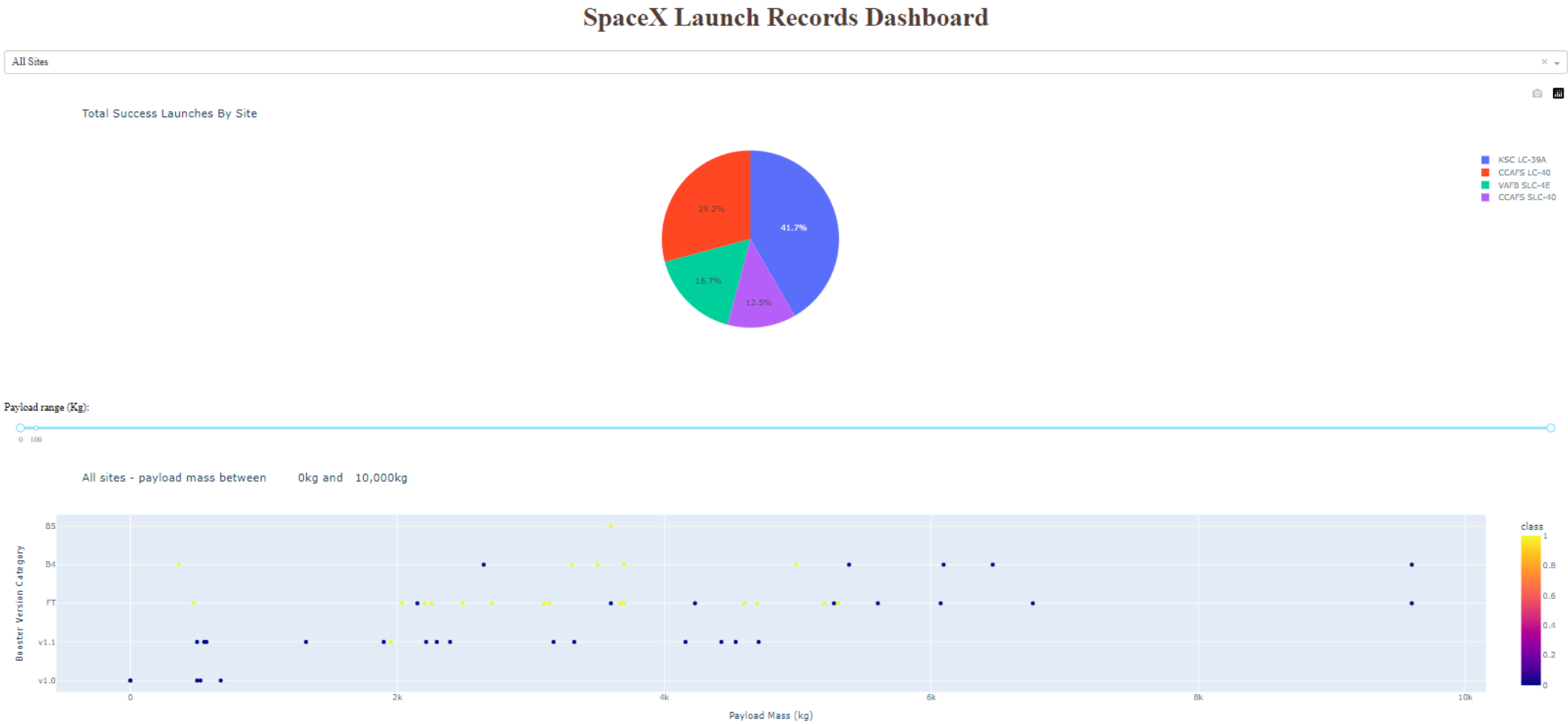
City 16.32km

Highway 20.28km

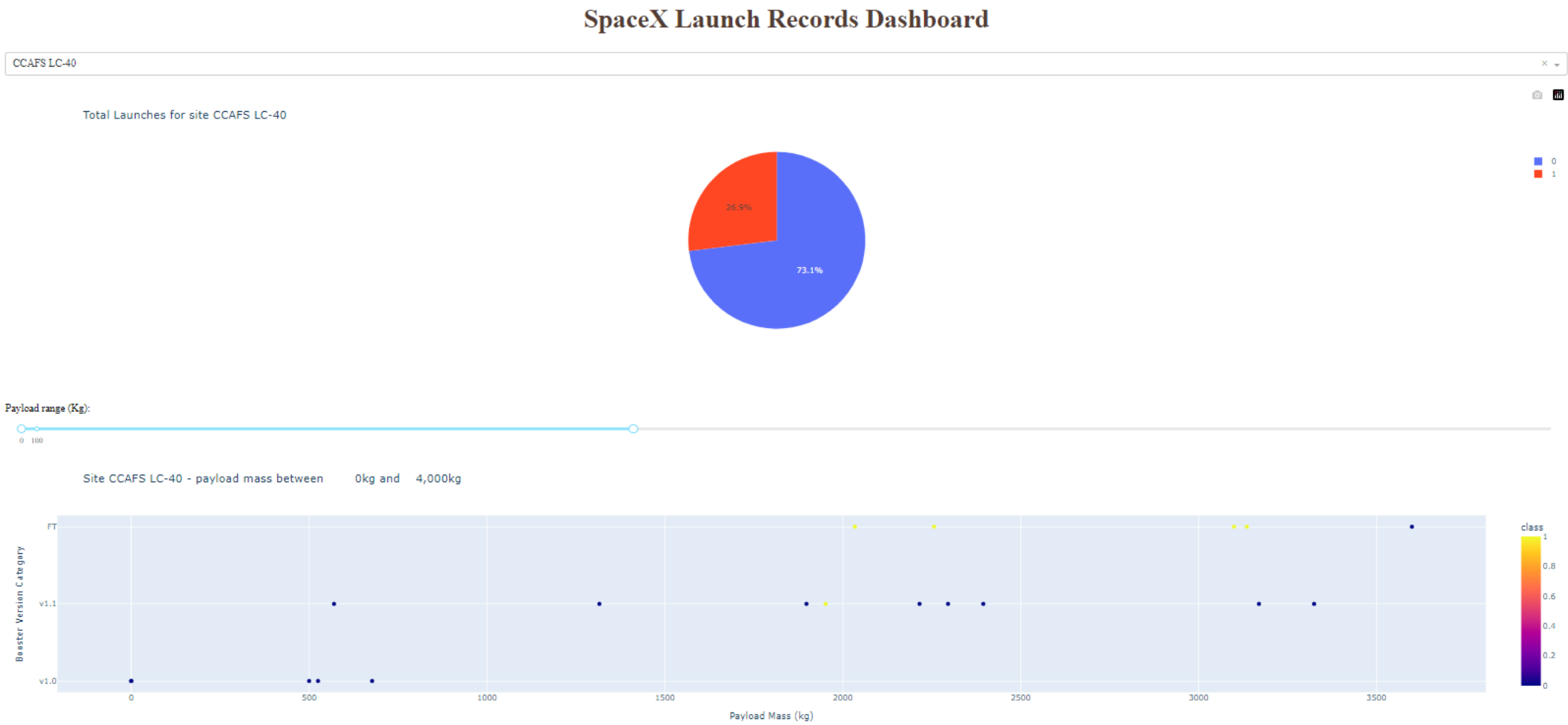
Coastline 14.99km

DASHBOARD

Successful Launches by Sites with Payload Mass Between 0-10000kg



Successful Launches in CCAFS LC-40 with Payload Mass between 0-4000 kg



PREDICTIVE ANALYSIS

CLASSIFICATION

Comparing the different ML models with evaluation metrics

Whole Dataset

[33]:

	LogReg	SVM	Tree	KNN
Jaccard_Score	0.833333	0.845070	0.841270	0.819444
F1_Score	0.909091	0.916031	0.913793	0.900763
Accuracy	0.866667	0.877778	0.888889	0.855556

Test Data

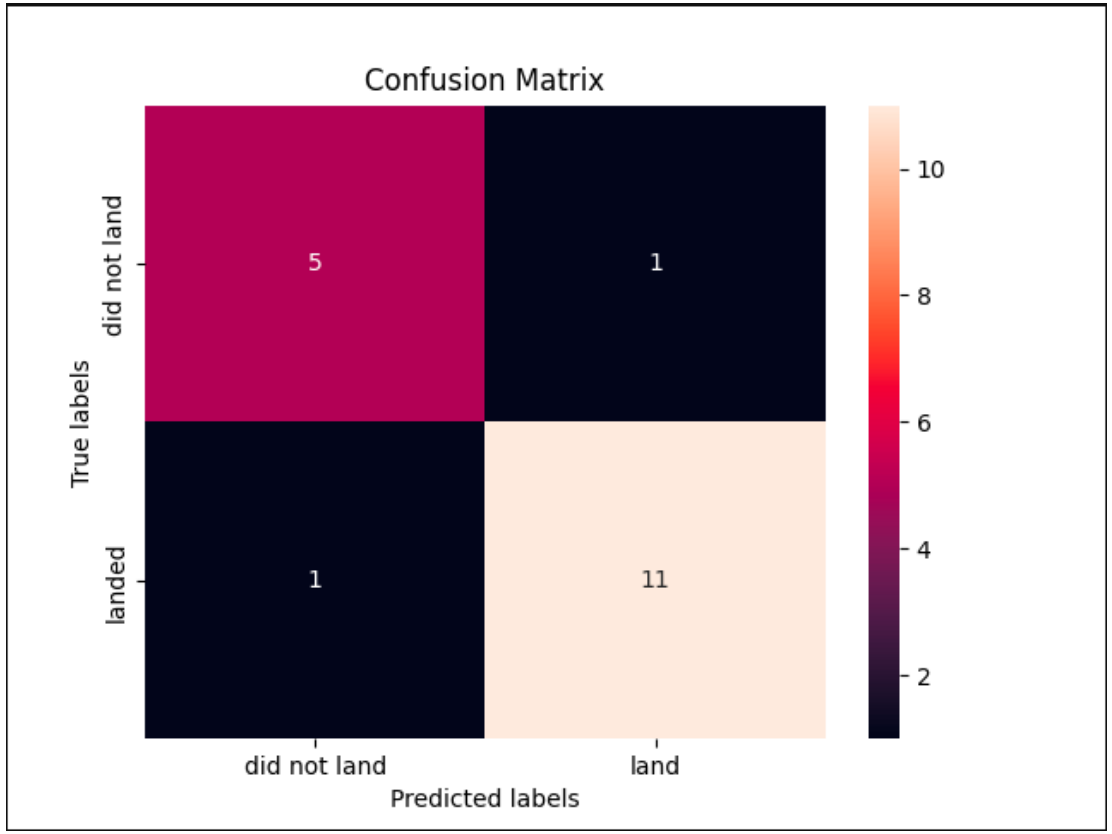
[32]:

	LogReg	SVM	Tree	KNN
Jaccard_Score	0.800000	0.800000	0.846154	0.800000
F1_Score	0.888889	0.888889	0.916667	0.888889
Accuracy	0.833333	0.833333	0.888889	0.833333

- **From the above image, Decision Tree shows to be the best classification model to use while Scoring on the Test dataset and the whole dataset**

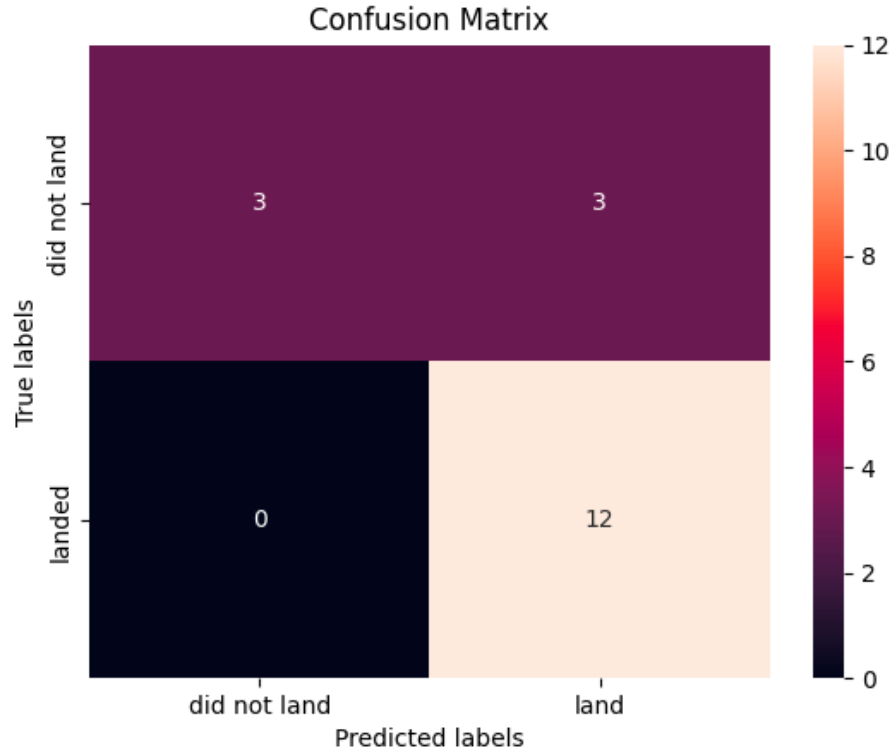
Confusion Matrix

Decision Tree Confusion Matrix



This model correctly predicted 91.67% of the true positive cases and 83.3% of the true negative cases

Logistic Regression Confusion Matrix



This model correctly predicted 100% of the true positive cases and 50% of the true negative cases

CONCLUSION

- The success rate increased over time in each launch-site
- The launch site with the highest success rate is KSC LC- 39A
- ES-L1, GEO, HEO and SSO have a 100% success rate
- Decision tree is the most optimal classification model in this case.
- the launch sites are located near the sea to avoid populated areas

THANK YOU!

Darex0132@gmail.Com

https://github.com/darexboy/