IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Mamadou Sagnon
09/19/2022

Outline

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

Executive Summary

- In this capstone project, we will predict whether the SpaceX Falcon 9 first stage will land successfully using several machine learning classification algorithms. The main steps in this project include:
- Data collection, wrangling, and formatting
- Exploratory data analysis
- Interactive data visualization
- Machine learning prediction
- Our graphs show that some features of the rocket launches are correlated to the success or failure of the landing. It is also concluded that decision tree may be the best machine learning algorithm to predict the landing outcome.

Introduction

- In this capstone, we will predict if the Falcon 9 first stage will land successfully. SpaceX is a revolutionary company that has disrupted the space industry by launching its Falcon 9 with a cost of 62 million dollars while other providers cost upward of 165 million dollars each. Most of the savings is because SpaceX can reuse the first stage by landing the rocket for the next mission . Reusing the rocket many time will make the price go down even further. As a data scientist our goal is to build machine learning algorithms to predict the landing outcome of the first stage. This project is crucial in identifying the right price to bid against SpaceX for a rocket launch.
- Most unsuccessful landings are planned. Sometimes, SpaceX performs controlled landings in the ocean. The main question that we are trying to answer is, for a given set of features about a Falcon 9 rocket launch which include its payload mass, orbit type, launch site, and so on, will the first stage of the rocket land successfully?

Section 1

Methodology

Methodology

- Data collection methodology:
  - SpaceX REST API.
  - Web Scraping on Wikipedia.
- Perform data wrangling
  - Data was processed using one-hot encoding for categorical features.
- Perform exploratory data analysis (EDA):
  - SQL
  - Visualization
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models:
  - Logistic regression
  - Support vector machine (SVM)
  - Decision tree
  - K-nearest neighbors (KNN)

Data Collection

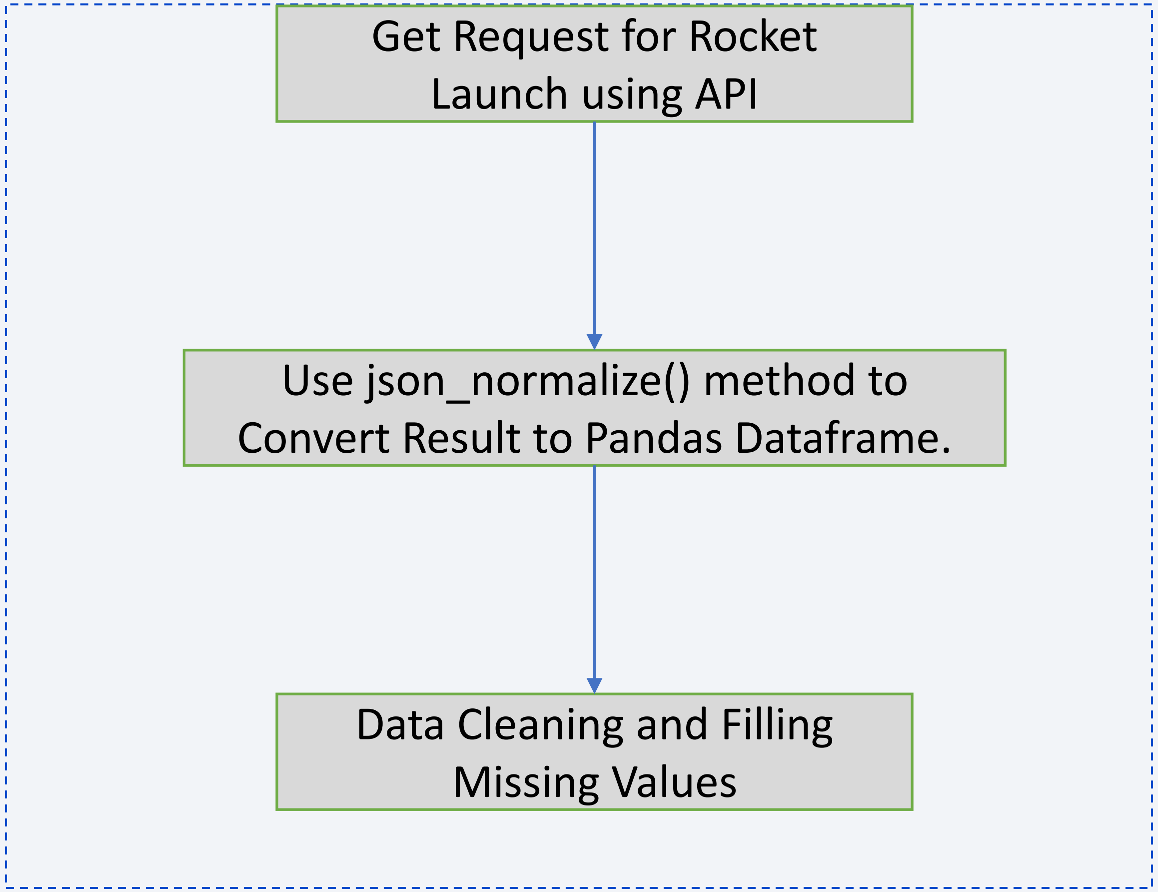
Data was collected using SpaceX REST API and Web Scraping of Wikipedia

- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

- The launch data was collected with SpaceX REST API calls. The data obtained contained information about the rocket used, the payload, launch specifications, landing specification and landing outcome.

- The link to the notebook is: https://github.com/msagnon1988/Applied-Data-Science-Capstone/blob/main/1.%20Collecting%20Using%20SapceX%20API.ipynb



Data Collection - Scraping

- We applied web scraping on Wikipedia to web scrap Falcon 9 launch records using BeautifulSoup.
- Link: https://github.com/msagnon1988/Applied-Data-Science-Capstone/blob/main/2.%20Web%20Scraping%20Falcon%209.ipynb

Apply HTML get() Method to Request Rocket Launch Page

↓

Create BeautifulSoup Object

↓

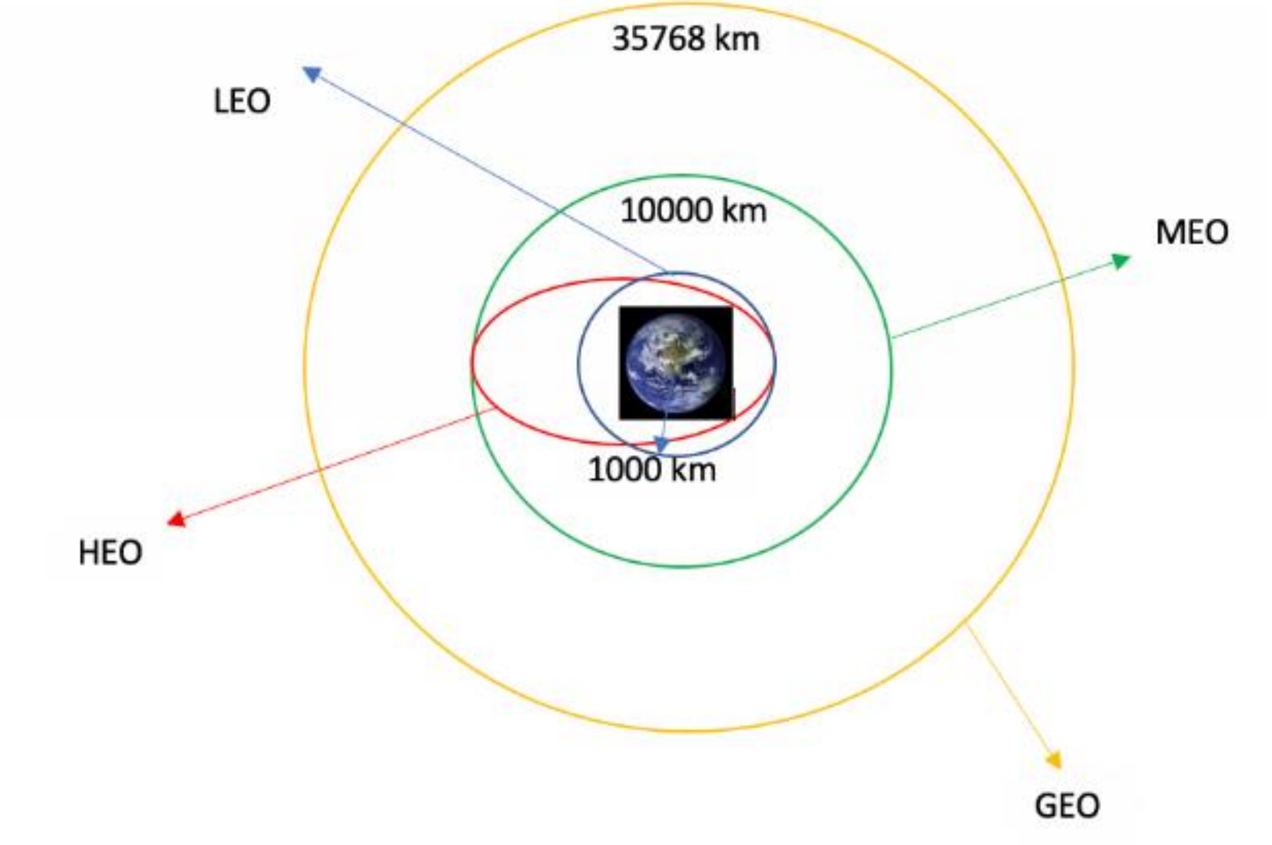
Extract all Columns Name from HTML Table Header

↓

Create a Dataframe

Data Wrangling

- Exploratory Data Analysis
- Features Selection
- Calculation of Number of launches at each site
- Calculation of occurrence of each orbit.
- Creation of landing outcome labels from Outcome column
- GitHub URL: https://github.com/msagnon1988/Applied-Data-Science-Capstone/blob/main/3.%20Data%20Wrangling.ipynb



EDA with Data Visualization

- We explored the data by plotting the relationship between:
  - Flight number and Launch Site
  - Payload and Launch Site
  - Success Rate of each orbit Type
  - Flight number and Orbit Type
  - Payload and Orbit type
  - The launch Success Yearly Trend.
- GitHub URL: https://github.com/msagnon1988/Applied-Data-Science-Capstone/blob/main/5.%20Exploring%20and%20Preparing%20Data.ipynb

EDA with SQL

- We loaded SpaceX dataset into IBM db2 database.
- We applied EDA with SQL to get insights from the dataset. We wrote SQL queries to:
  - Display the names of unique launch sites in the space mission
  - Show find records where launch sites begin with the string ‘CCA’
  - Display the total payload mass carried by boosters launched by NASA (CRS)
  - Display average payload mass carried by booster version F9 v1.1
  - ***List the date when the first successful landing outcome in ground pad was achieved.***
  - ***List the total number of successful and failure mission outcomes***
- GitHub URL: https://github.com/msagnon1988/Applied-Data-Science-Capstone/blob/main/4.%20EDA%20Using%20SQL.ipynb

Build an Interactive Map with Folium

- We marked the launch sites, added markers, circles, lines to mark the failure or success of launches for each site on the folium map
- We assigned 0 to failed launch and 1 to successful launch
- We identified which sites have relatively high success rate using the color-labelled marker cluster.
- The distance between launch sites and their proximities were calculated. We answered the following questions:
    - Are launch sites near to railways, highways, coastlines?
    - Do launch sites keep certain distance away from cities?

Build a Dashboard with Plotly Dash

- We built an interactive dashboard with Plotly Dash
- We plotted pie chart showing the total launches by sites
- We built a scatter plot showing the relationship between launch outcome and payload mass for the different booster version

Predictive Analysis (Classification)

- We loaded the data into Jupiter notebook using pandas and numpy and transformed it.
- We split the data into training set and test set.
- We built different machine learning models and tune the hyperparameters using GridSearchCV.
- We used accuracy as metric to measure the performance of the models
- We selected the best model by comparing the accuracy of all models
- GitHub URL: https://github.com/msagnon1988/Applied-Data-Science-Capstone/blob/main/7.%20Falcon%209%20Machine%20Learning%20Prediction.ipynb

Results

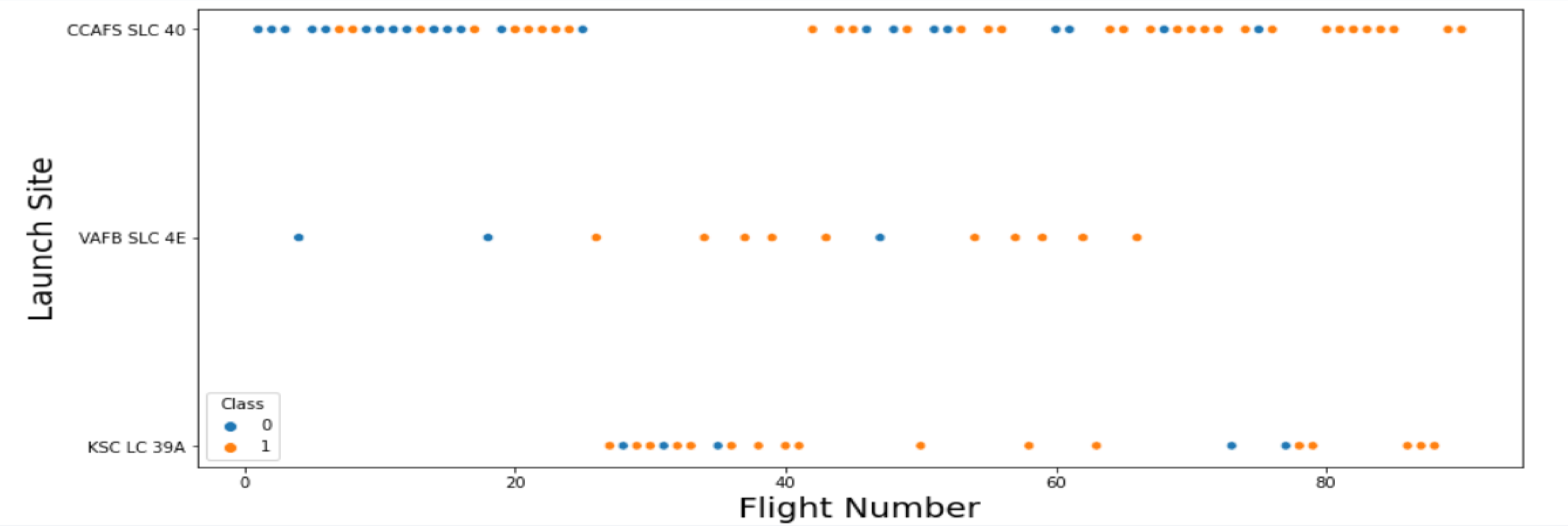
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

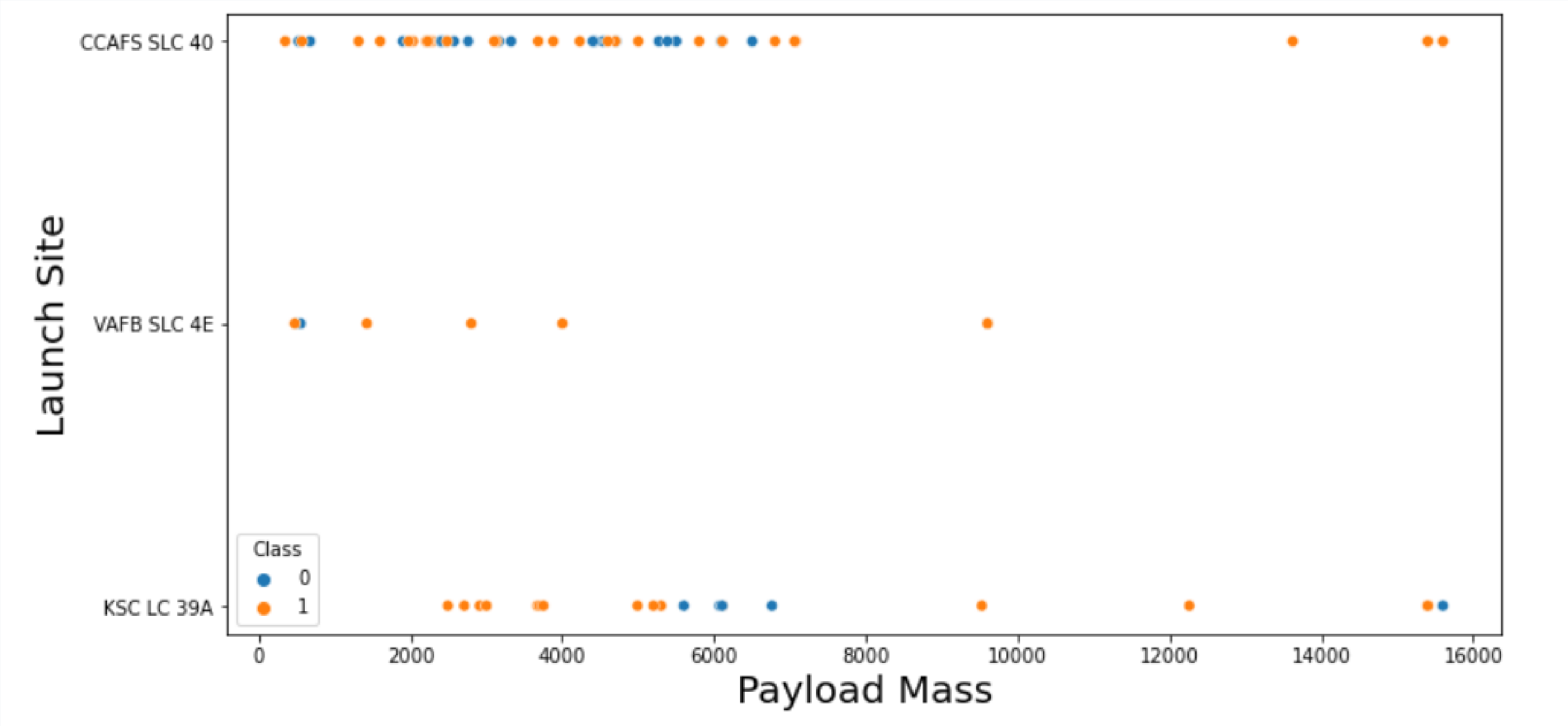
Flight Number vs. Launch Site

The larger the flight number the greater the success rate.



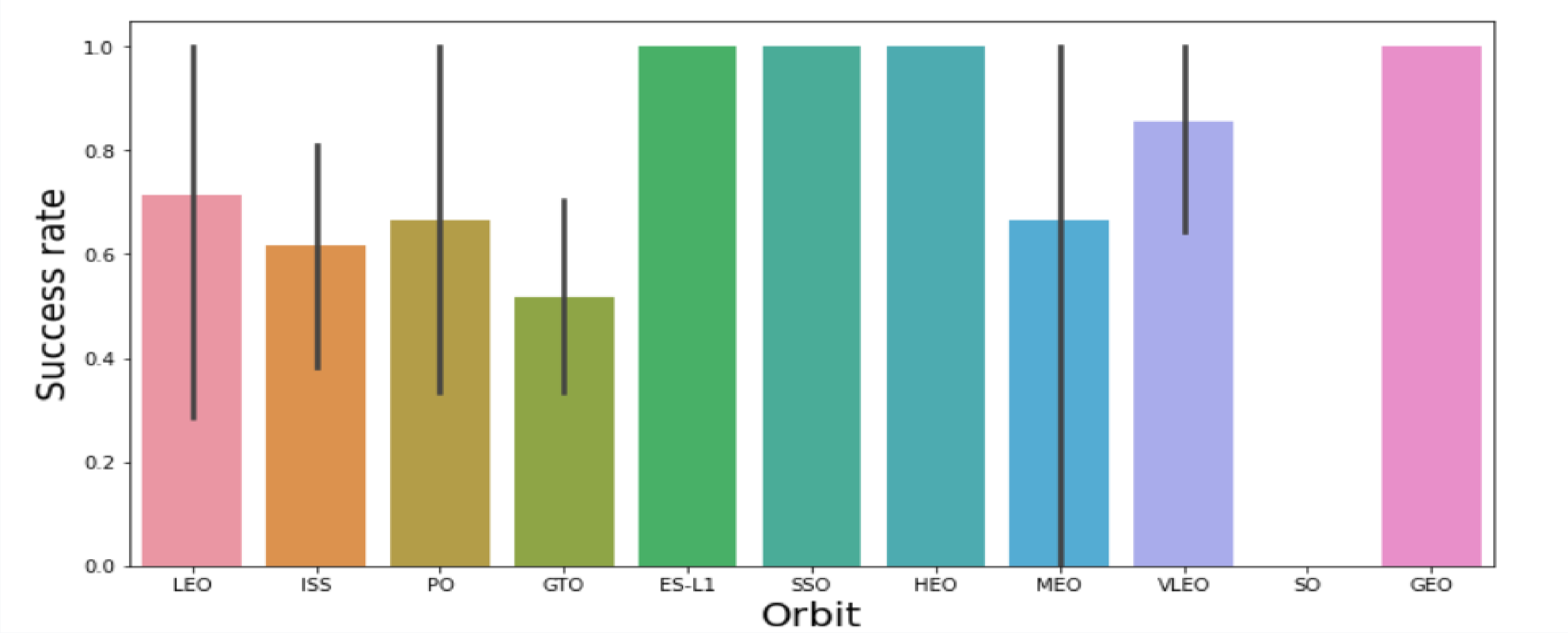
Payload vs. Launch Site

For launch site CCAFS SLC, the bigger the payload mass, the higher the success rate.



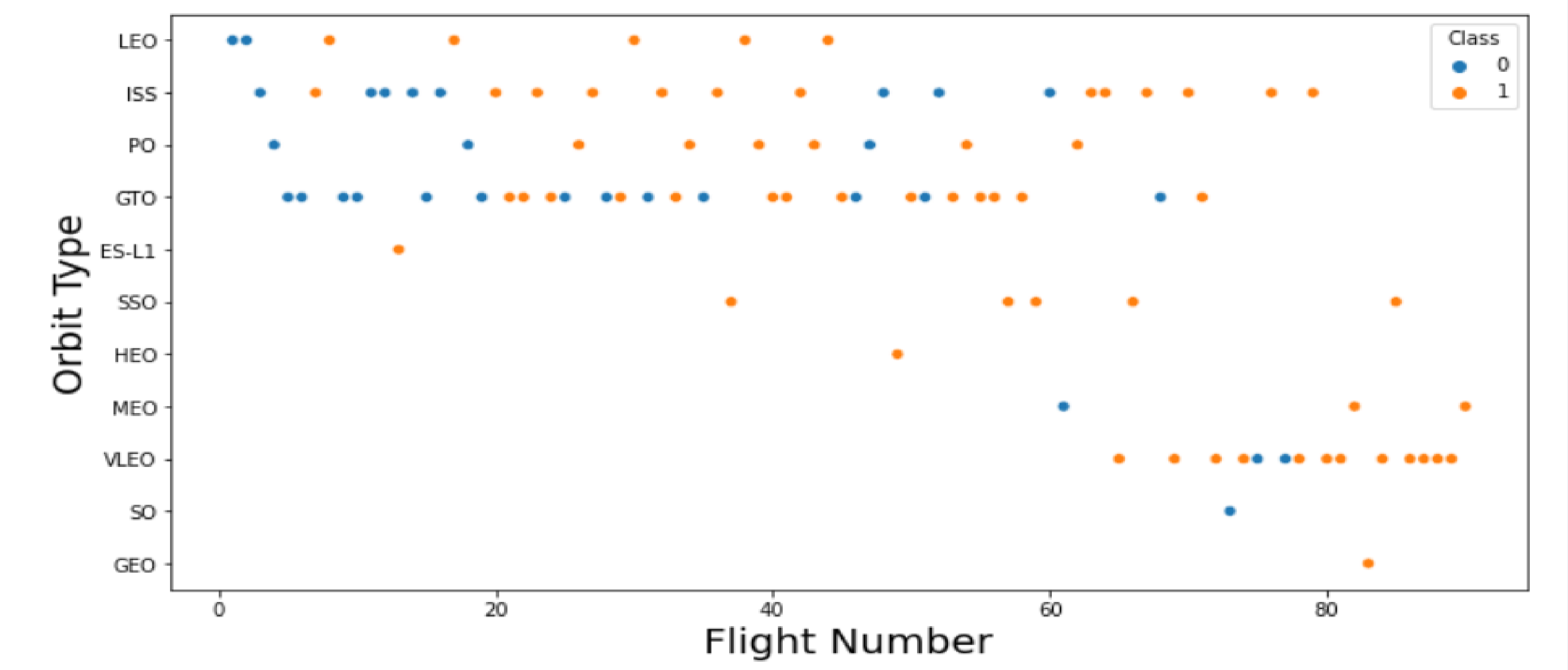
Success Rate vs. Orbit Type

- ES-L1, SSO, HEO, and GEO have the highest success rate



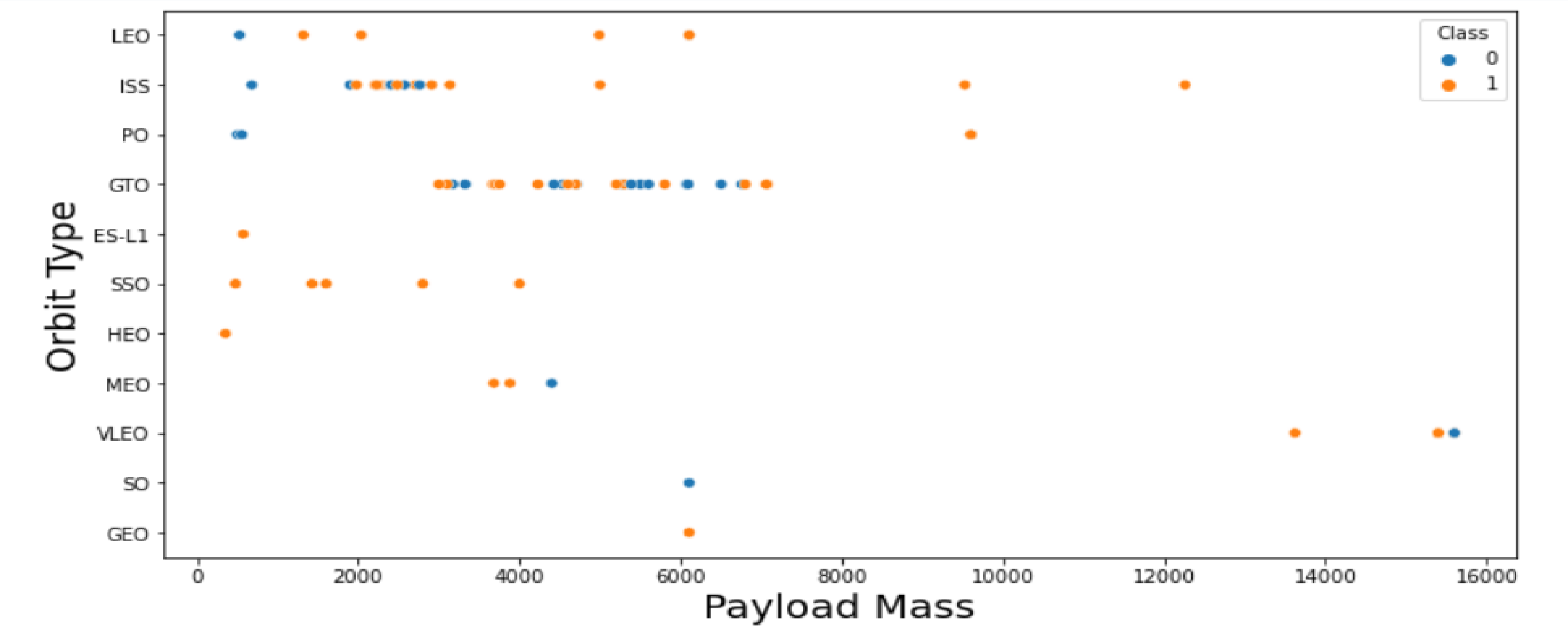
Flight Number vs. Orbit Type

We observed that in the LEO Orbit success is related to the number of flights whereas there is no relationship between flight number and orbit type in the GTO.



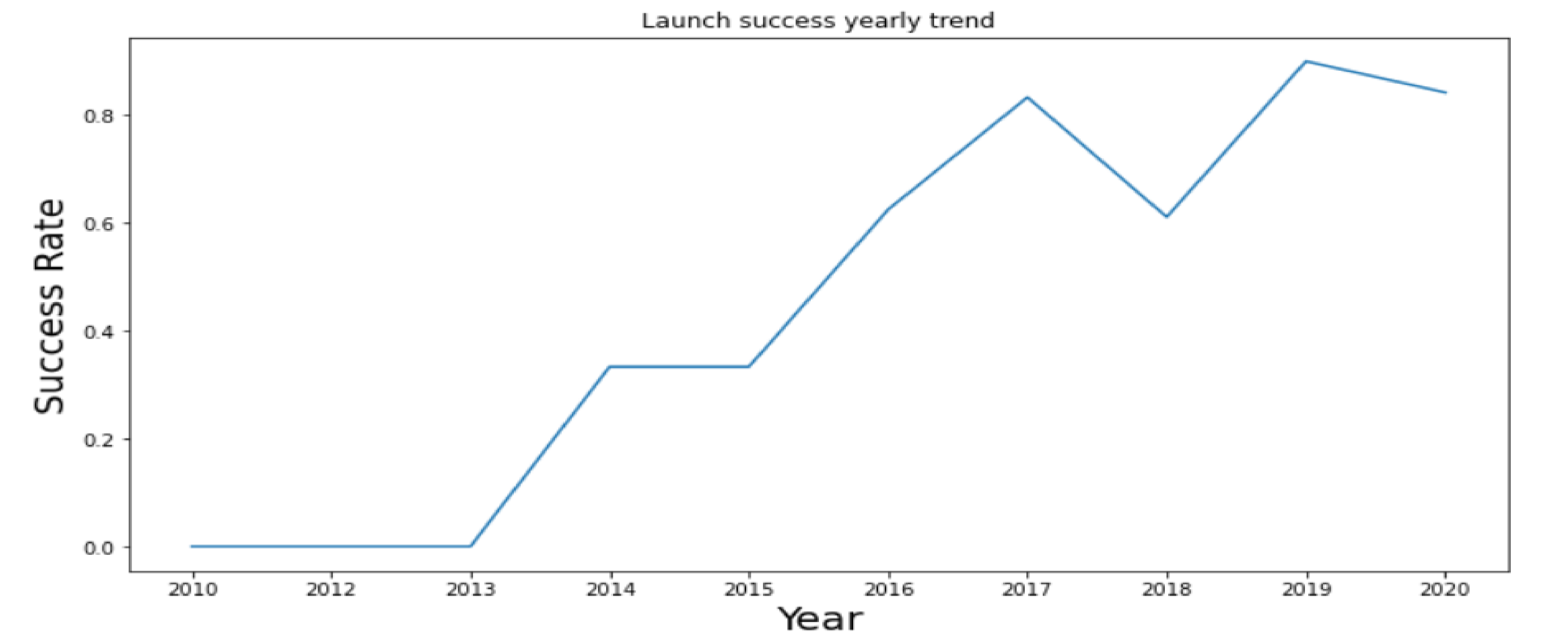
Payload vs. Orbit Type

- For heavy payload the positive landing rate are more for Polar, and ISS
- For GTO we cannot distinguish this well as both positive landing rate and negative landing rate are both there.



Launch Success Yearly Trend

The success rate since 2013 kept increasing till 2020



All Launch Site Names

We used DISTINCT to display the unique site names

```
%%sql select distinct(LAUNCH_SITE)
from SPACEXTBL
```

* ibm_db_sa://[redacted]

Done.

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

Launch Site Names Begin with 'CCA'

The first 5 records with site names beginning with CCA

```
%%sql select *
from SPACEXTBL
where LAUNCH_SITE like 'CCA%'
limit 5
```

 * [redacted]
Done.

[7]:

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	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
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

The total payload mass carried by boosters launched by NASA (CRS) is 45,596 KG

```
%%sql select sum(PAYLOAD_MASS__KG_)
from SPACEXTBL
where CUSTOMER = 'NASA (CRS)'

 * ibm_db_sa://csv49024:***@54a2f15b-
Done.

[8]:      1
      45596
```

Average Payload Mass by F9 v1.1

The average payload mass carried by booster version F9 v1.1 is **2,928** KG.

```
%%sql select avg(PAYLOAD_MASS__KG_)
from SPACEXTBL
where BOOSTER_VERSION = 'F9 v1.1'
```

```
 * ibm_db_sa://csv49024:***@54a
Done.
```

5]:

1
2928

First Successful Ground Landing Date

The date of the first successful landing outcome on ground pad is December 22nd, 2015.

```
%%sql select min(DATE)
from SPACEXTBL
where LANDING__OUTCOME = 'Success (ground pad)'
```

```
 * ibm_db_sa://csv49024:***@54a2f15b-5c0f-46
Done.
```

7]:

1
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

F9 FT B1022, F9 FT B1026, F9 FT B1021.2, and F9 FT B1031.2 have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

```
%%sql select BOOSTER_VERSION
from SPACEXTBL
where LANDING__OUTCOME = 'Success (drone ship)'
and PAYLOAD_MASS__KG_ between 4000 and 6000
```

```
 * ibm_db_sa://csv49024:***@54a2f15b-5c0f-4
Done.
```

21]:

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

Total Number of Successful and Failure Mission Outcomes

- The total number of successful mission outcomes is 100
- The total number of failed mission outcomes is 1

```
%%sql select count(MISSION_OUTCOME) as Total_Success
from SPACEXTBL
where MISSION_OUTCOME like 'Success%'
```

```
 * ibm_db_sa://csv49024:***@54a2f15b-5c0f-46df-8
Done.
```

6]:

total_success
100

```
%%sql select count(MISSION_OUTCOME) as Total_Failure
from SPACEXTBL
where MISSION_OUTCOME like 'Failure%'
```

```
 * ibm_db_sa://csv49024:***@54a2f15b-5c0f-46df-8
Done.
```

7]:

total_failure
1

Boosters Carried Maximum Payload

```
%%sql select BOOSTER_VERSION
from SPACEXTBL
where PAYLOAD_MASS__KG_ =
            (select max(PAYLOAD_MASS__KG_)
            from SPACEXTBL)
```

```
 * ibm_db_sa://csv49024:***@54a2f15b-5
Done.
```

•

1]:

booster_version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

2015 Launch Records

•

```
%%sql select BOOSTER_VERSION, LAUNCH_SITE
from SPACEXTBL
where year(DATE) = 2015
and LANDING__OUTCOME = 'Failure (drone ship)'
```

```
 * ibm_db_sa://csv49024:***@54a2f15b-5c0f
Done.
```

46]:

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

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

```
%%sql select LANDING__OUTCOME, count(LANDING__OUTCOME) as count
from SPACEXTBL
where DATE >= '2010-06-04'
and DATE <= '2017-03-20'
group by LANDING__OUTCOME
order by count(LANDING__OUTCOME)
desc
```

```
 * ibm_db_sa://[redacted]:***@[redacted]
Done.
```

9]:

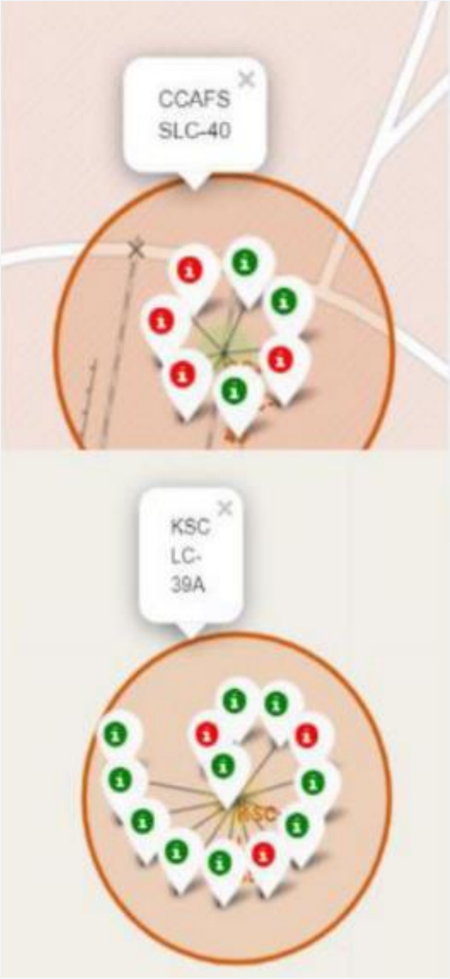
landing__outcome	COUNT
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

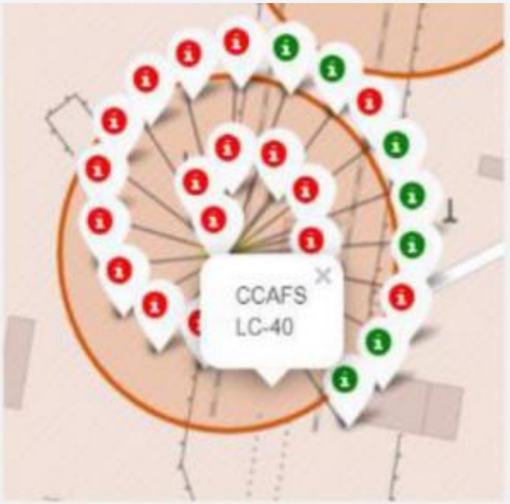
Section 3

Launch Sites Proximities Analysis

All Launch Sites Global Map Markers

Markers Showing Launch Sites with Color Labels





Florida Launch Sites

Green Marker shows successful Launches and Red Marker shows Failures

California Launch Site

Launch Sites Distance to Lanmarks

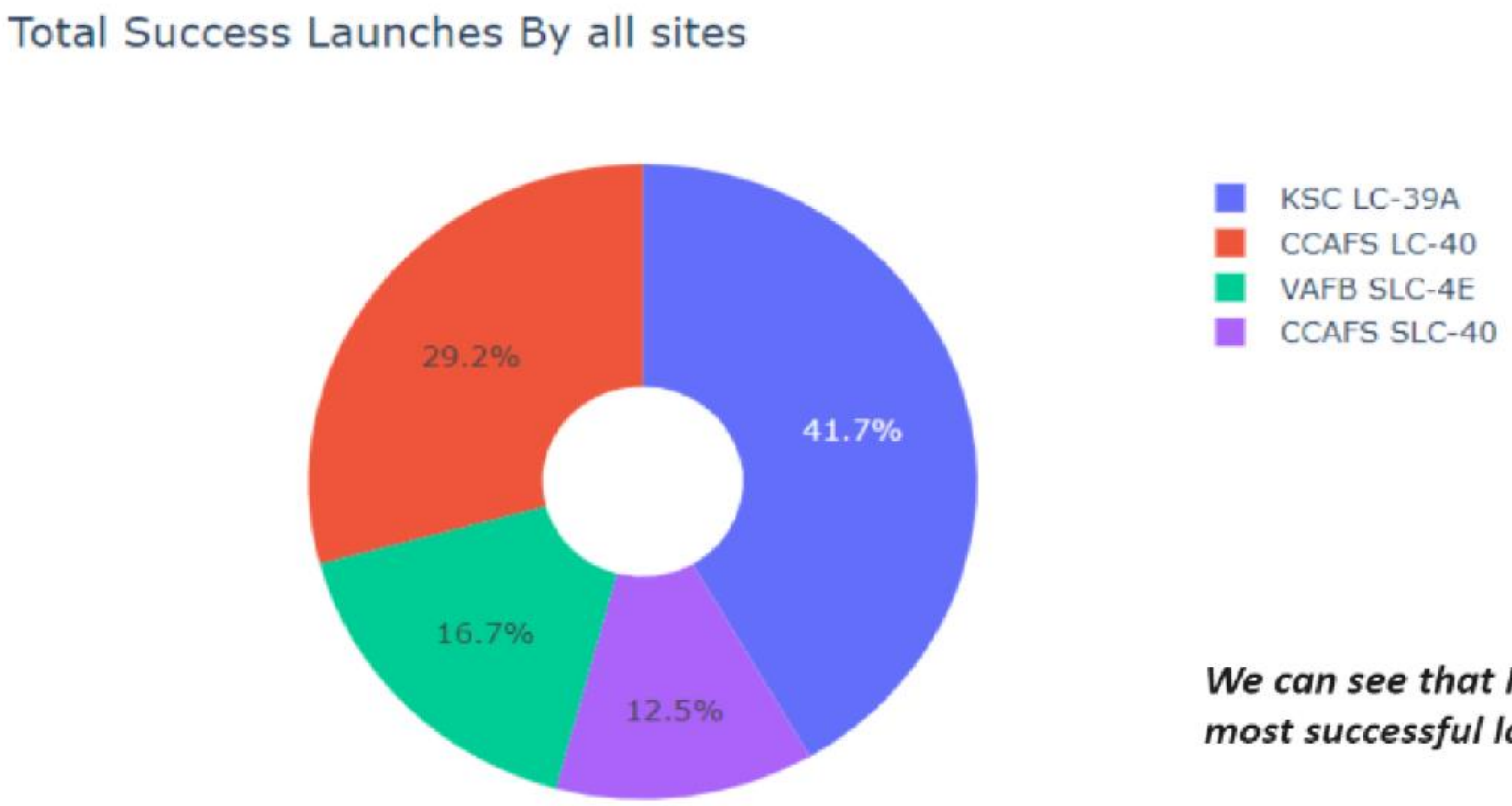
•

- Are launch sites in close proximity to railways? No
- Are launch sites in close proximity to highways? No
- Are launch sites in close proximity to coastline? Yes
- Do launch sites keep certain distance away from cities? Yes

Section 4

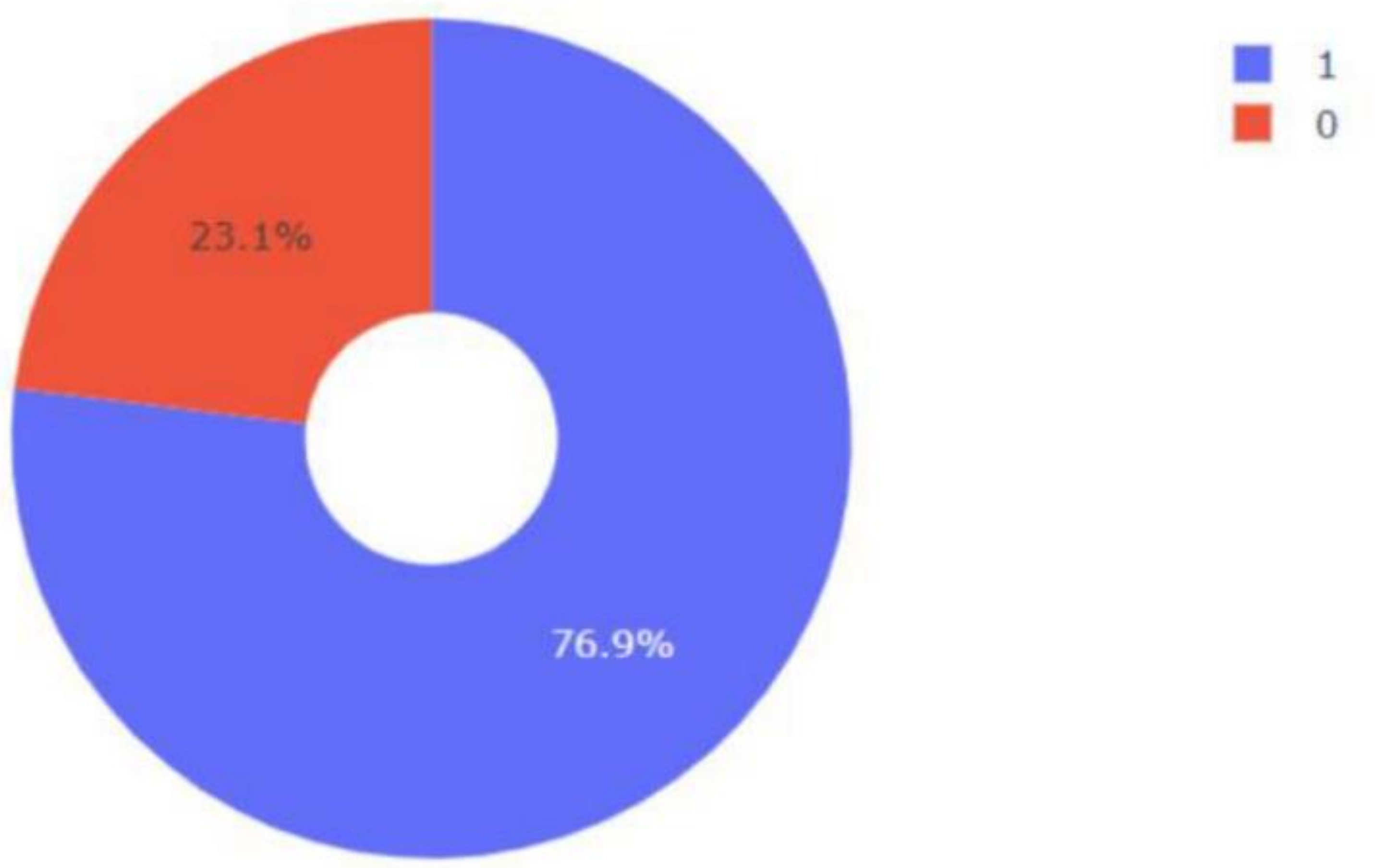
Build a Dashboard with Plotly Dash

Pie Chart of Success rate by Site



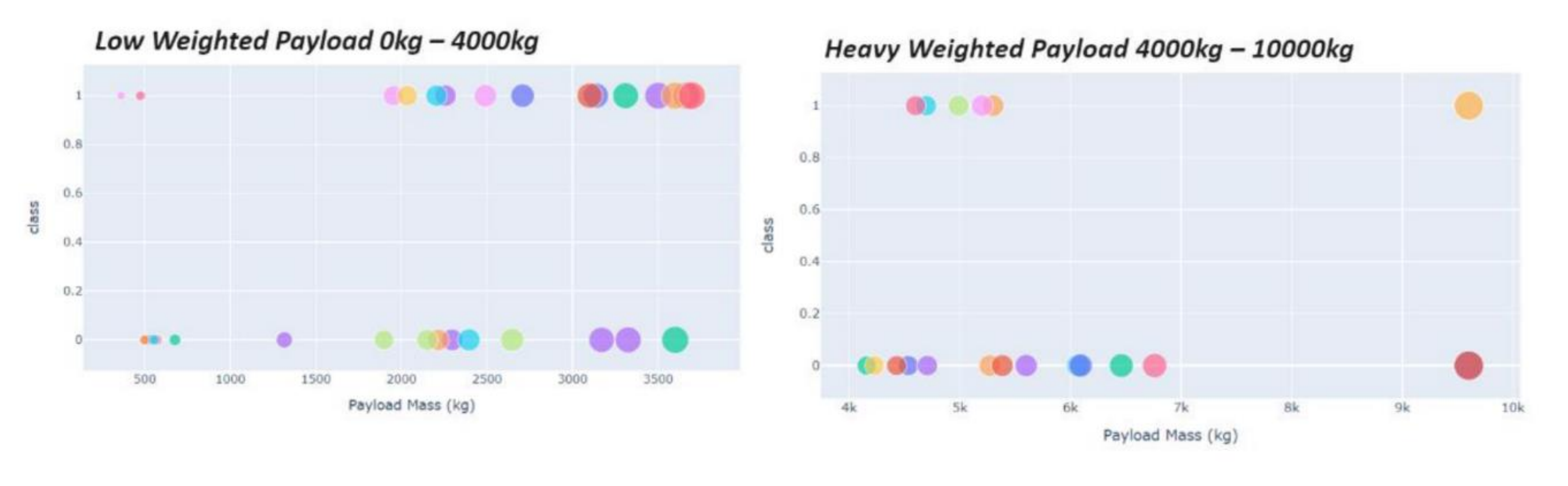
We can see that KSC LC-39A had the most successful launches from all the sites

Launch Site with the Best Success Rate



KSC LC-39A achieved a 76.9% success rate while getting a 23.1% failure rate

Payload vs Launch Outcome

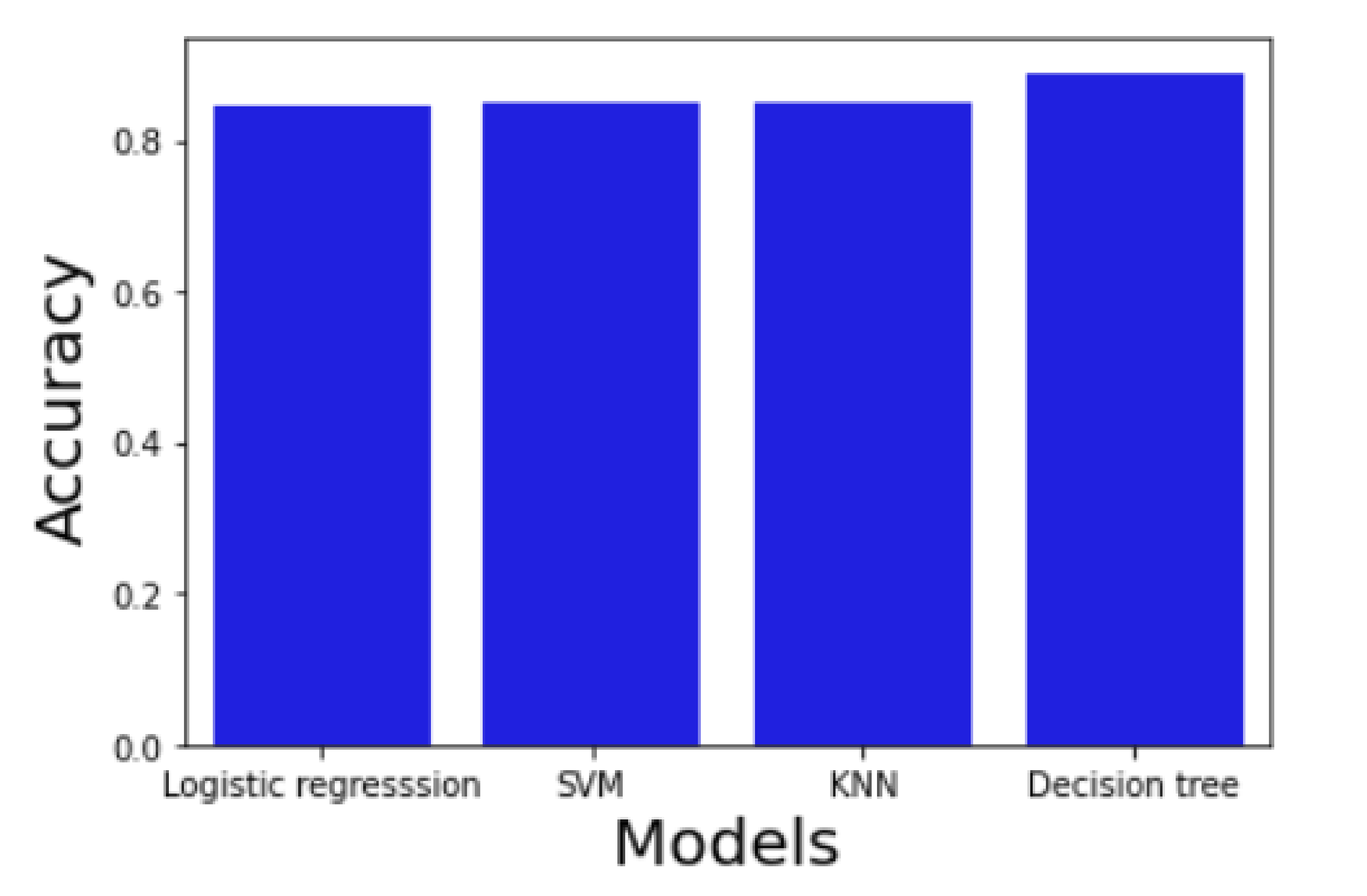


Section 5

Predictive Analysis (Classification)

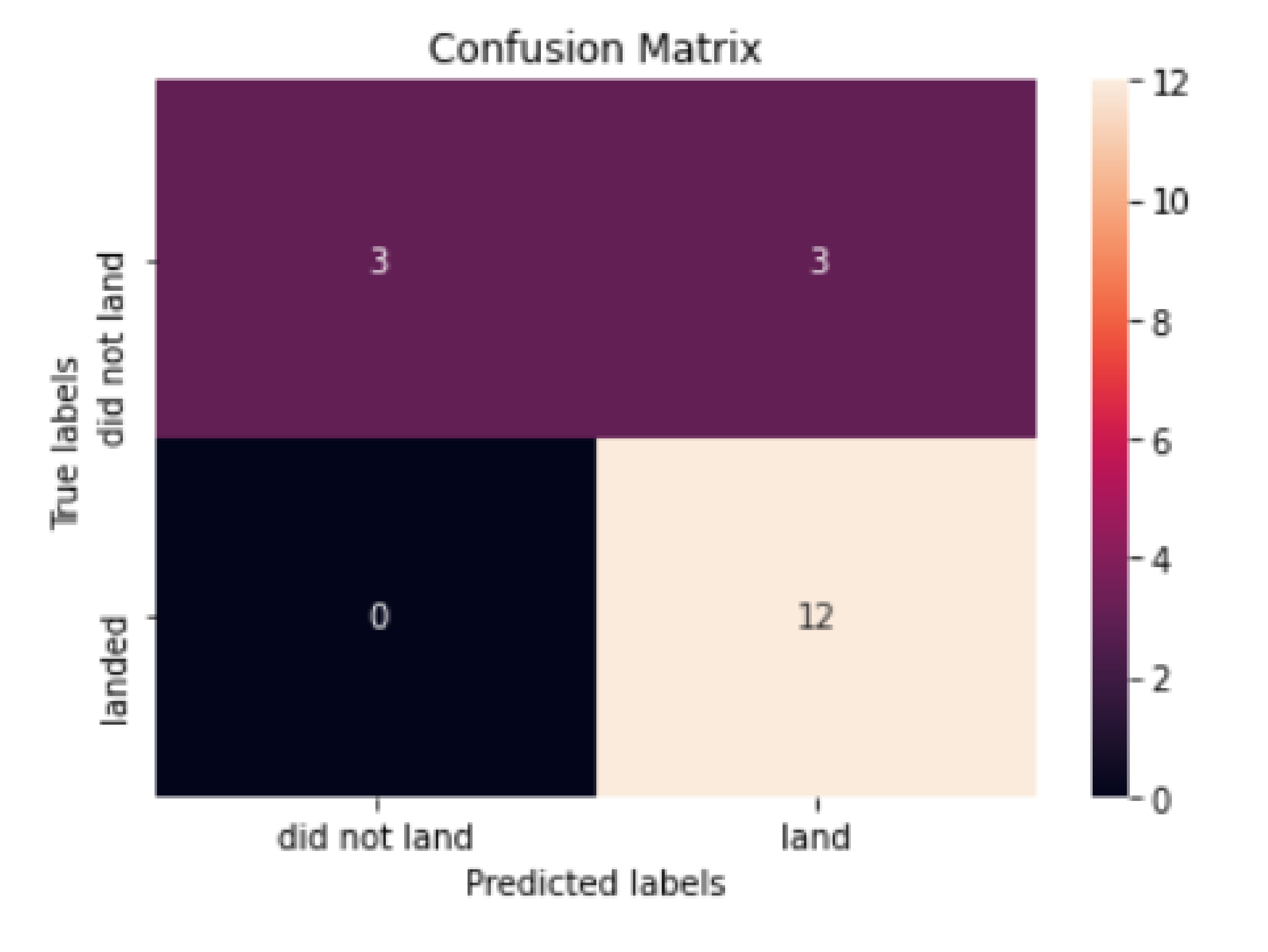
Classification Accuracy

Decision tree has the highest accuracy



Confusion Matrix

The decision tree algorithm is the best predictor of the landing outcome. Its confusion matrix shows that it is able to predict very well the outcome. However, it has false positives



Conclusions

- The larger the flight amount at a Site, the greater the success rate at that site.
- Launch success rate started to increase from 2013.
- Orbits ES-L1, GEO, HEO, VLEO, SSO have the most success rate
- Site KSC-LC 39A has the most successful launches
- Decision tree classifier is the best algorithm in predicting the lauding outcome.

Thank you!