SpaceX Falcon 9 Landing Predictive Analysis

(Applied Data Science Capstone Final Presentation)

Massinissa TINOUCHE

04-19-2023

OUTLINE

- Executive Summary
- Introduction
- METHODOLOGY
  - Data Collection and Data Wrangling
  - EDA and Interactive Visual Analytics
  - Predictive Analysis
- RESULTS
  - EDA with Visualization
  - EDA with SQL
  - Interactive Map With Folium
  - Plotly Dash Dashboard
  - Predictive Analysis (Classification)
- Conclusion

EXECUTIVE SUMMARY

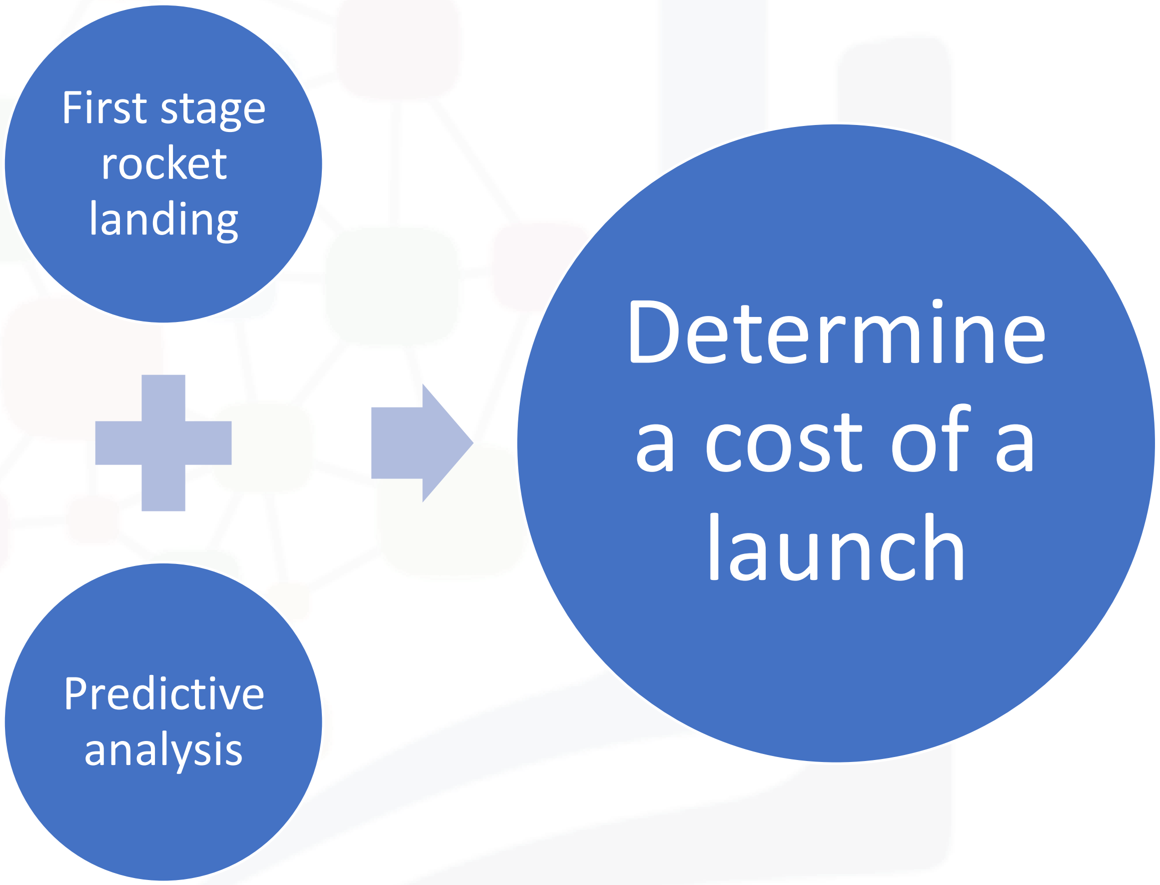
103 million dollars is the saving if the first stage of the rocket will be reusable (launch and success landing), rocket launches (Falcon 9) cost 62 M$ for SpaceX while this amount can cost upward 165 M$ if the first stage will not reusable. A predictive analysis study will help us to determine the cost of a launch.

First of all, we used API's to collect data from "spacexdata.com" website. After exploring and preparing data, we perform some Exploratory Data Analysis (EDA) in order to find relationship between data and determine what would be the label for training supervised models.

We used machine learning models such as logistic regression, support vector machine, decision tree and K nearest neighbors to find the best corresponding model to our study. After splitting data into train and test sets, we perform calculation with each model, we calculate the coefficient of determination R^2 and the accuracy using K Fold. We find that the Logistic Regression was the best model that fit with the best performance.

INTRODUCTION

- **Project subject:** SpaceX Falcon 9 Landing.
- **Study:** Landing Predictive Analysis.
- **Predictive analysis tools:** Machine Learning models.
- **ML models:**
  - Logistic Regression (LR).
  - Support Vector Machine (SVM).
  - Decision Tree (DT).
  - K Nearest Neighbors (KNN).
- **Program Language:** Python



Data Collection and Data Wrangling Methodology

- Data collection

  1) https://api.spacexdata.com/v4/rockets/

  2) https://api.spacexdata.com/v4/launchpads/

  3) https://api.spacexdata.com/v4/payloads/

  4) https://api.spacexdata.com/v4/cores/

  5) https://api.spacexdata.com/v4/launches/past

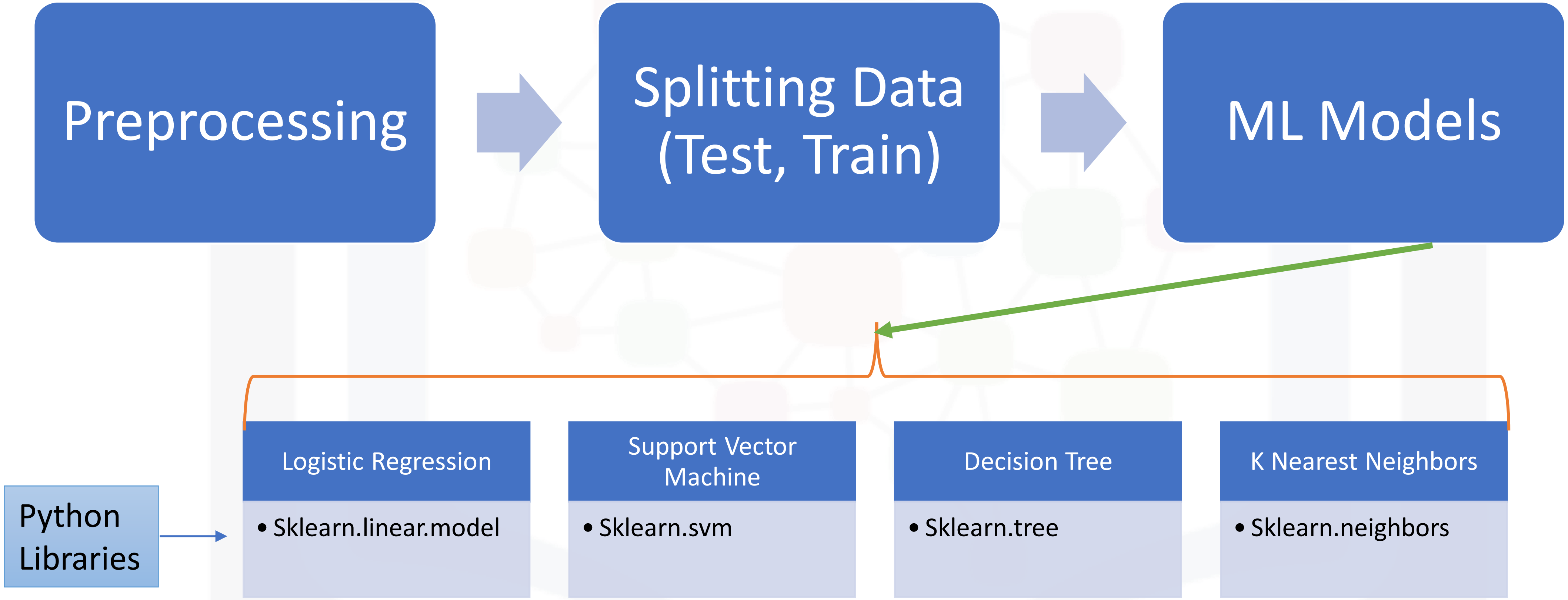
- Data Wrangling

  - Dealing with missing values.
  - Identifying which column are numerical and categorical.
  - Organizing data into datasets.

EDA and Interactive Visual Analytics Methodology

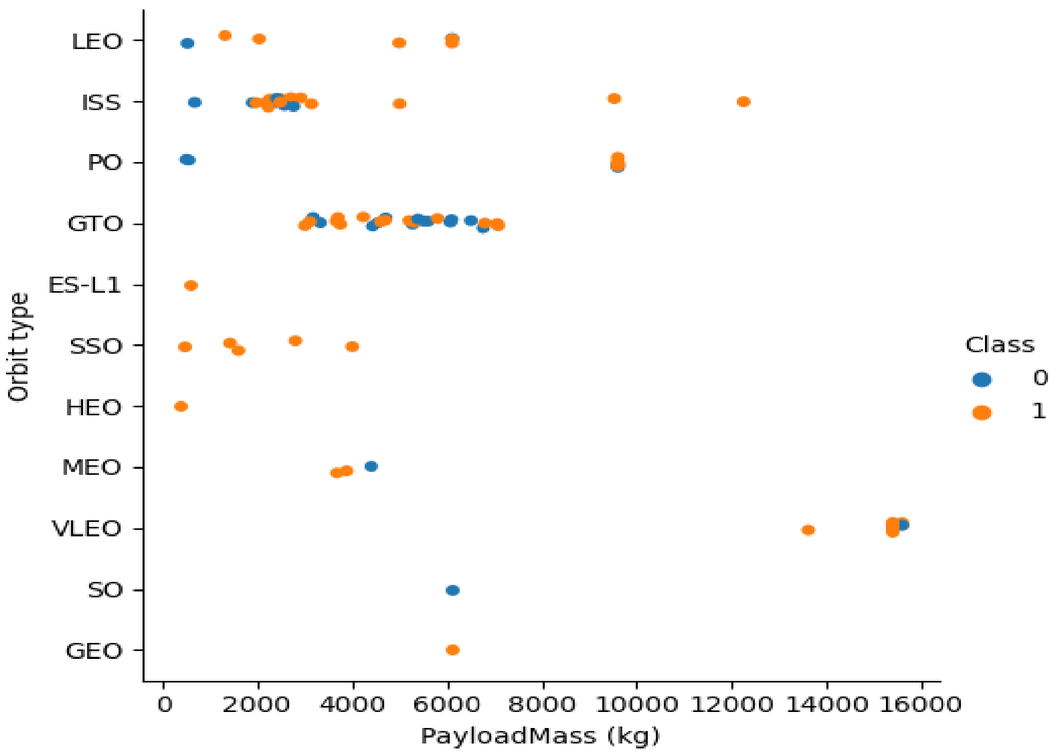
- Python libraries for Exploratory Data Analysis:
  - Pandas
  - Numpy
  - Matplotlib
  - Seaborn
- Plots and Charts (Relationship between):
  - Flight Number & Launch Site
  - Payload & Launch Site
  - Success rate & Orbit type
  - Flight Number & Orbit type
  - Payload & Orbit type
  - Launch Success Yearly Trend

Predictive Analysis Methodology

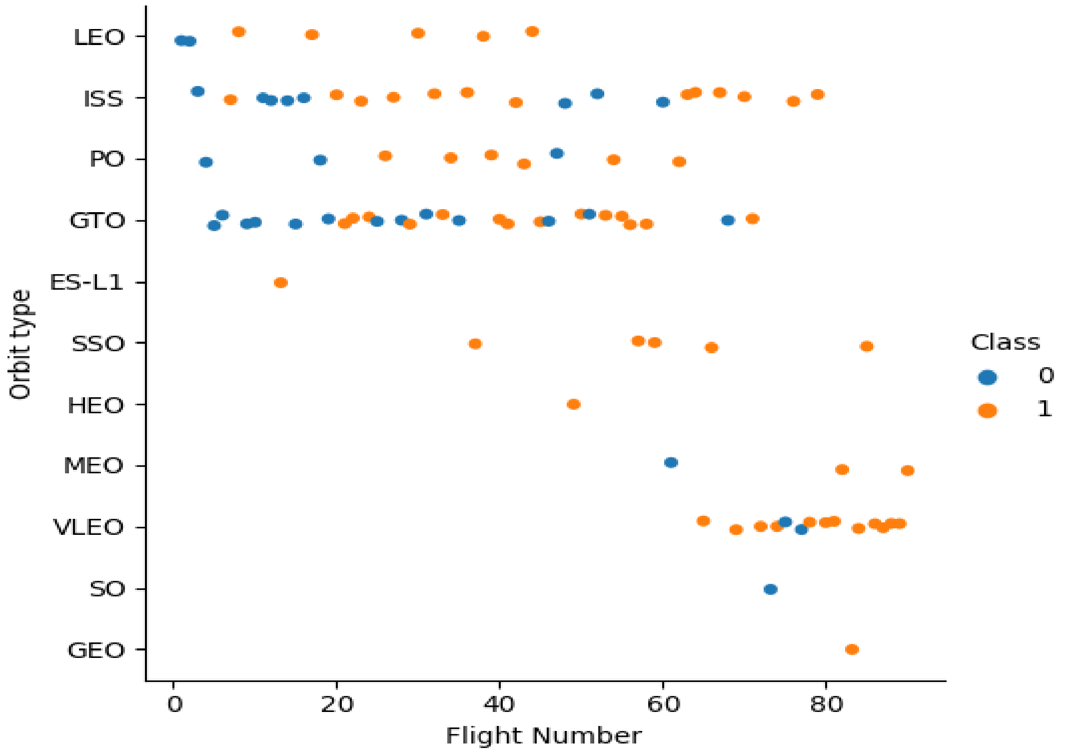


EDA with Visualization Results

Payload Mass Vs. Orbit type



Flight Number Vs. Orbit type

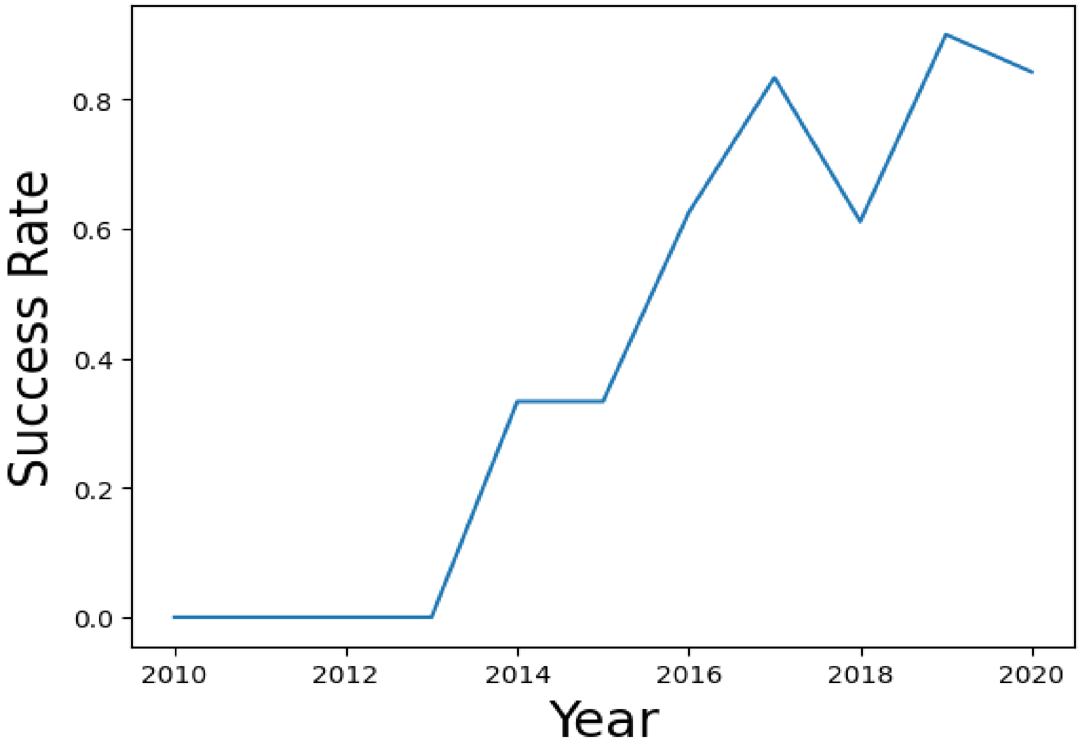


Success rate Vs. Year

According to the figure beside, we can divide the success rate from 2010 to 2020 into three intervals:

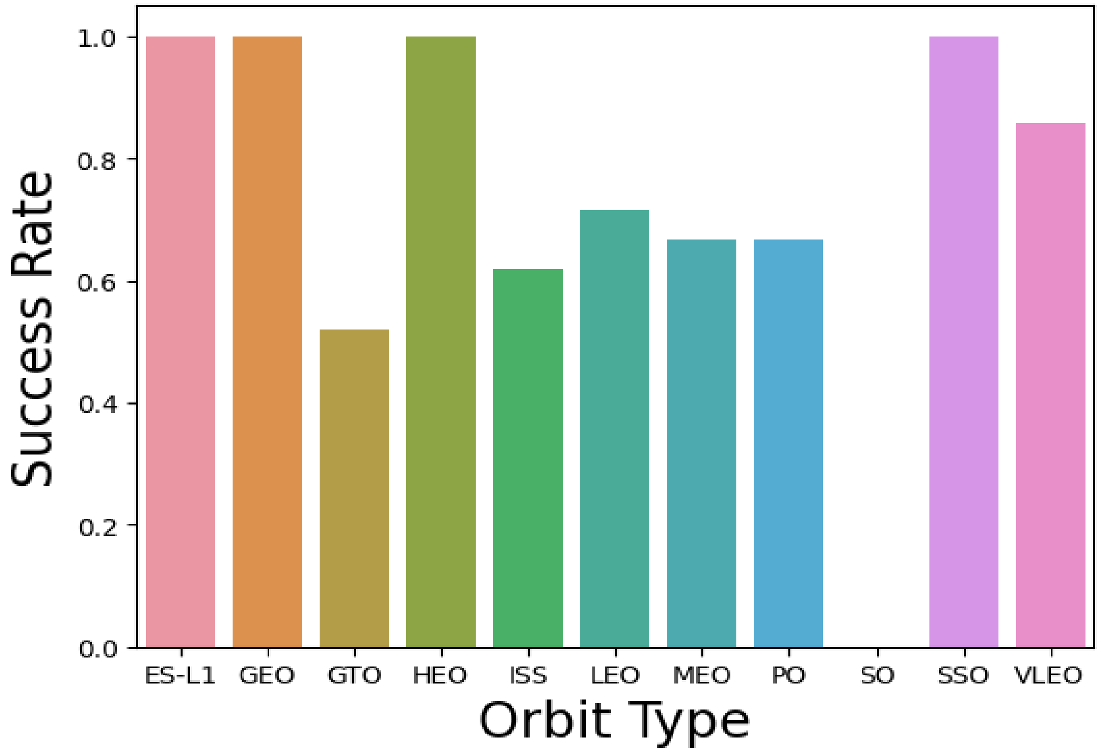
- **[2010, 2013]:** 2010 was the first launch for Falcon 9.
- **[2013, 2017]:** during this period, we notice a significant and rapid increase of the success rate.
- **[2017, 2020]:** the success rate is disturbed but overall it's increases. This perturbation would probably was related to the "Starship" development project.

Overall, the success rate increases significantly.



Success rate Vs. Orbit

It is important to study the success rate for each Orbit type, since the aim of the existence of the rockets is to carry load from the surface of the earth to a given altitude (orbit).



According to the figure beside, the success rate does not depend on the altitude or on the shape of the orbit (elliptic or circular).

ALL Launch Site Names

```
LAUNCH_SITE = pd.read_sql("select Unique(LAUNCH_SITE) from SpaceX", pconn)
LAUNCH_SITE
```

SQL Statement

Launch Sites
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E

Launch Site Names Beginning with “CCA”

```
CCA = pd.read_sql("select * from SpaceX where (LAUNCH_SITE) LIKE 'CCA%' LIMIT 5", pconn)

CCA
```

SQL Statement

DATE	TIME UTC	BOOSTER VERSION	LAUNCH SITE
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40

Payload Mass Carried by Boosters

```
payloadmass = pd.read_sql("select sum(PAYLOAD_MASS_KG_) as payloadmass from SpaceX", pconn)
payloadmass
```

SQL Statement

Total Payload Mass (kg)	Boosters
619967	NASA (CRS)

Average Payload Mass (kg)	Boosters
6138	F9 v1.1

```
payloadmass_avg= "select avg(PAYLOAD_MASS_KG_) as payloadmass from SpaceX"
payloadmass_avg = pd.read_sql(payloadmass_avg, pconn)
print(payloadmass_avg)
```

Boosters which have success in drone ship and have payload mass between 4000 and 6000 kg

```
BOOSTER_VERSION= "select Booster_Version from SpaceX where Landing_Outcome= 'Success (drone ship)' and PAYLOAD_MASS_KG_ BETWEEN 4000 and 6000"
BOOSTER_VERSION = pd.read_sql(BOOSTER_VERSION, pconn)
BOOSTER_VERSION
```

SQL Statement

Booster Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

2010-06-04

Was the date when the first successful landing outcome in ground pad was achieved

List of the failed landing outcome in drone ship

```
Landing_Outcomes_2015= "SELECT MONTH(DATE),Mission_Outcome,Booster_Version,Launch_Site FROM
SpaceX where EXTRACT(YEAR FROM DATE)='2015'"
Landing_Outcomes_2015 = pd.read_sql(Landing_Outcomes_2015, pconn)
Landing_Outcomes_2015
```

SQL Statement

Mission Outcome	Booster Version	Launch Site
Success	F9 v1.1 B1012	CCAFS LC-40
Success	F9 v1.1 B1013	CCAFS LC-40
Success	F9 v1.1 B1014	CCAFS LC-40
Success	F9 v1.1 B1015	CCAFS LC-40
Success	F9 v1.1 B1016	CCAFS LC-40
Failure (in flight)	F9 v1.1 B1018	CCAFS LC-40
Success	F9 FT B1019	CCAFS LC-40

Rank the count of landing outcomes between 2010-06-04 and 2017-03-20

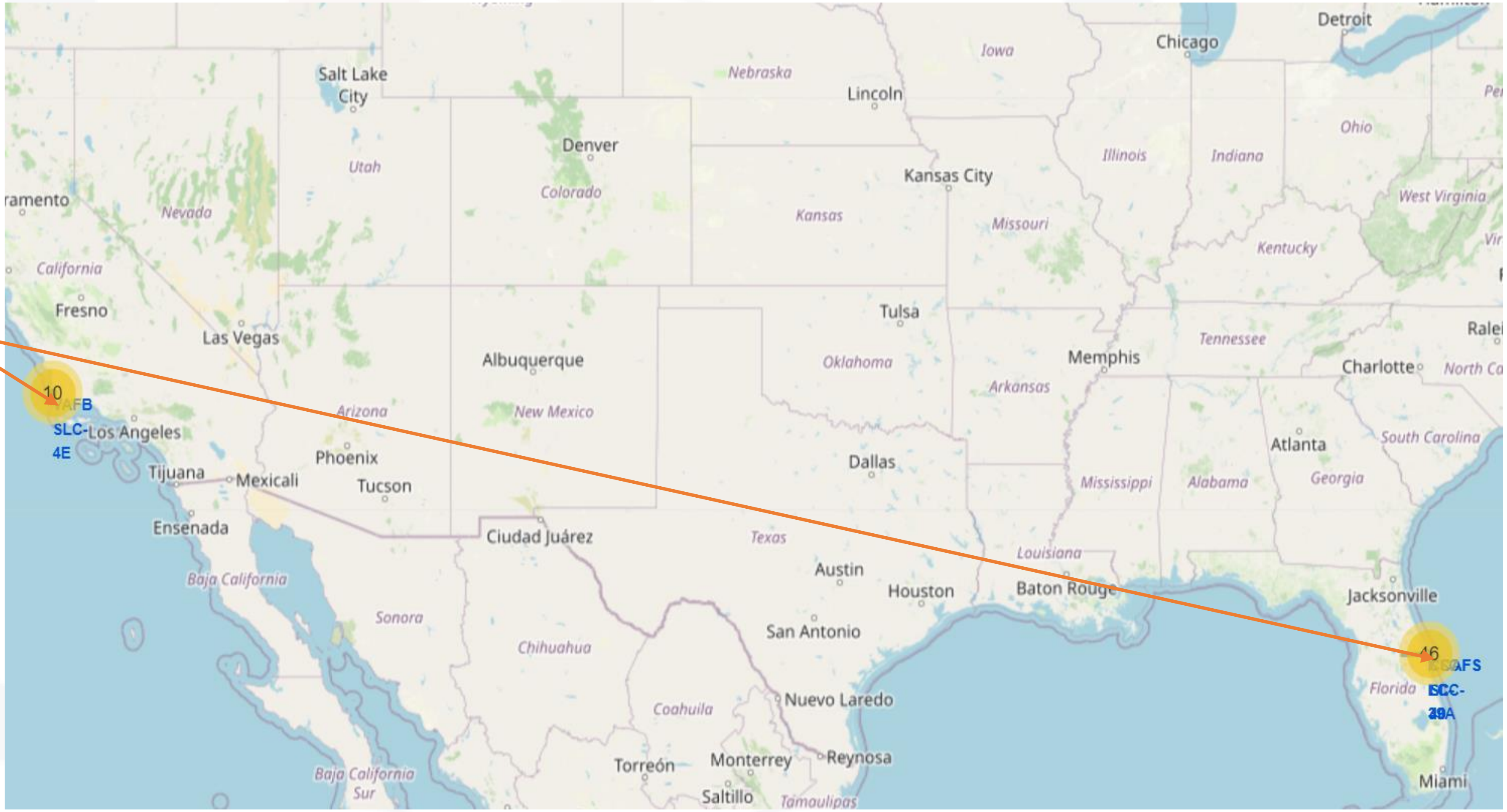
```
Rank_Landing_Outcomes= "select LANDING_OUTCOME, count(*) from SpaceX where Date between
'2011-06-04' and '2017-03-20' group by LANDING_OUTCOME order by 2 desc"
Rank_Landing_Outcomes = pd.read_sql(Rank_Landing_Outcomes, pconn)
Rank_Landing_Outcomes
```

SQL Statement

Landing Outcome	Count
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Uncontrolled (ocean)	2
Precluded (drone ship)	1

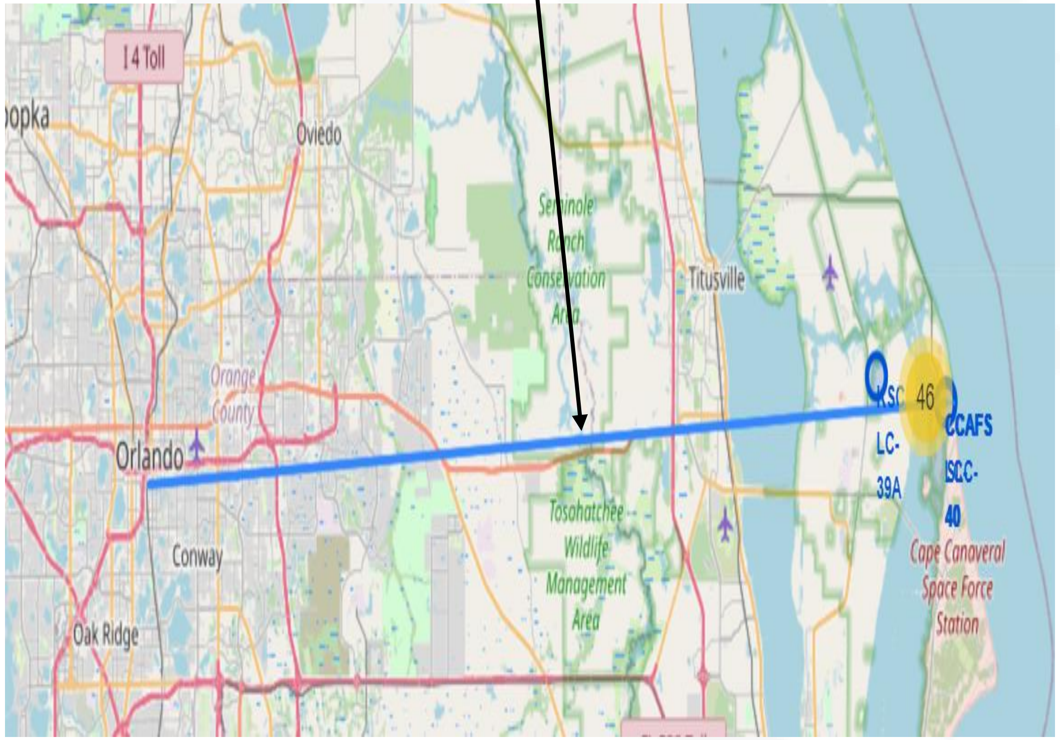
Interactive Map With Folium Results

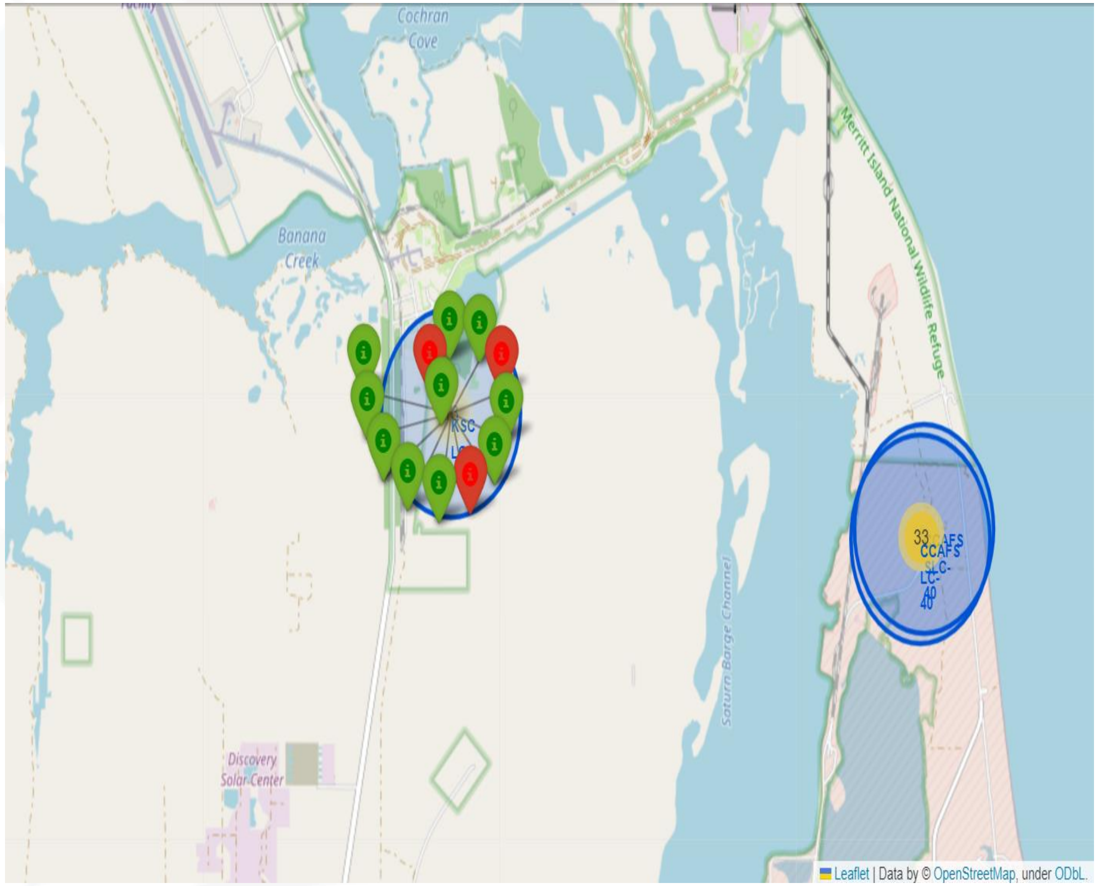
Displaying different Launch Sites using Folium.



Interactive Map With Folium Results

Distance between a launch site to its proximities





IBM Developer

SKILLS NETWORK

Plotly Dash Dashboard Results

Interactive plotly dashboard

Runing the python script of dash app

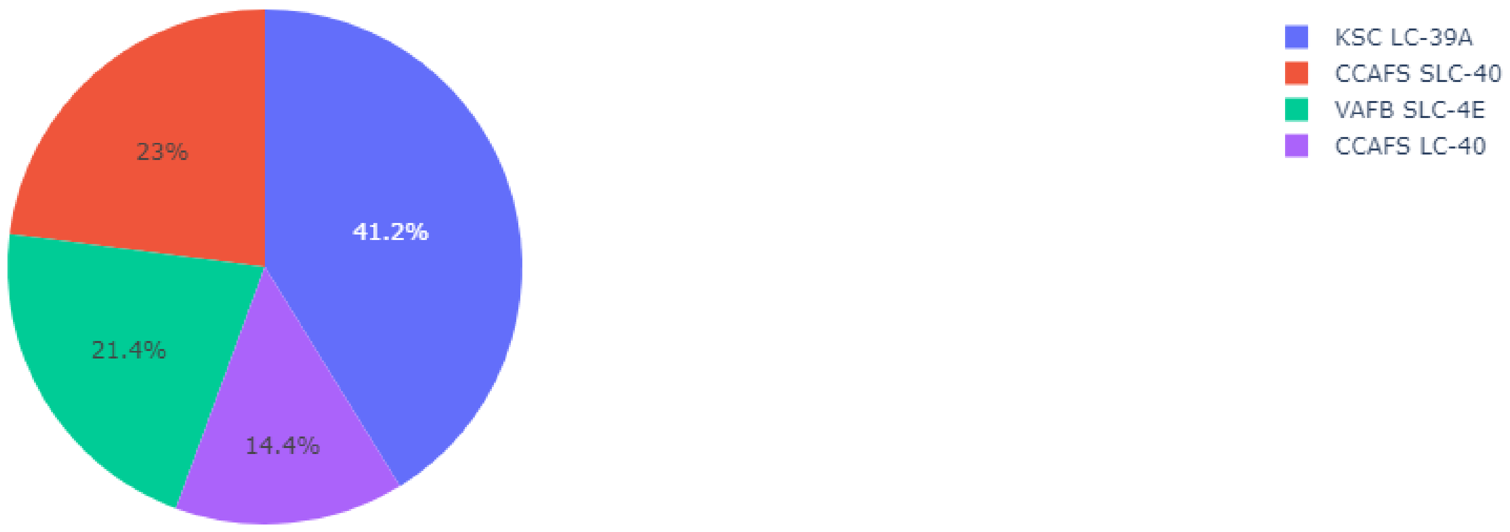
http://127.0.0.1:8050

Plotly Dash Dashboard of Launch Success Rate for all Sites

Launch Success Rate For All Sites

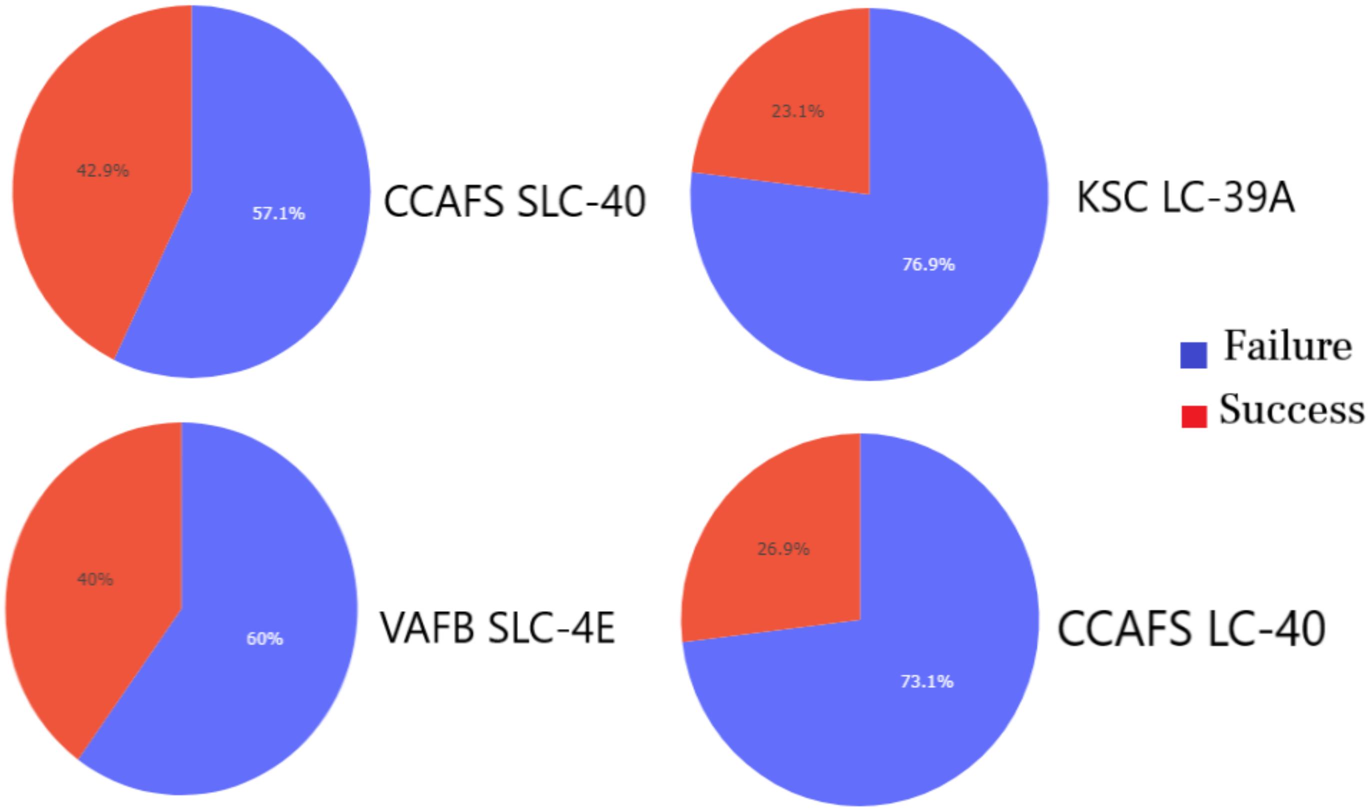
Florida: 78.6 %

California: 21.4 %

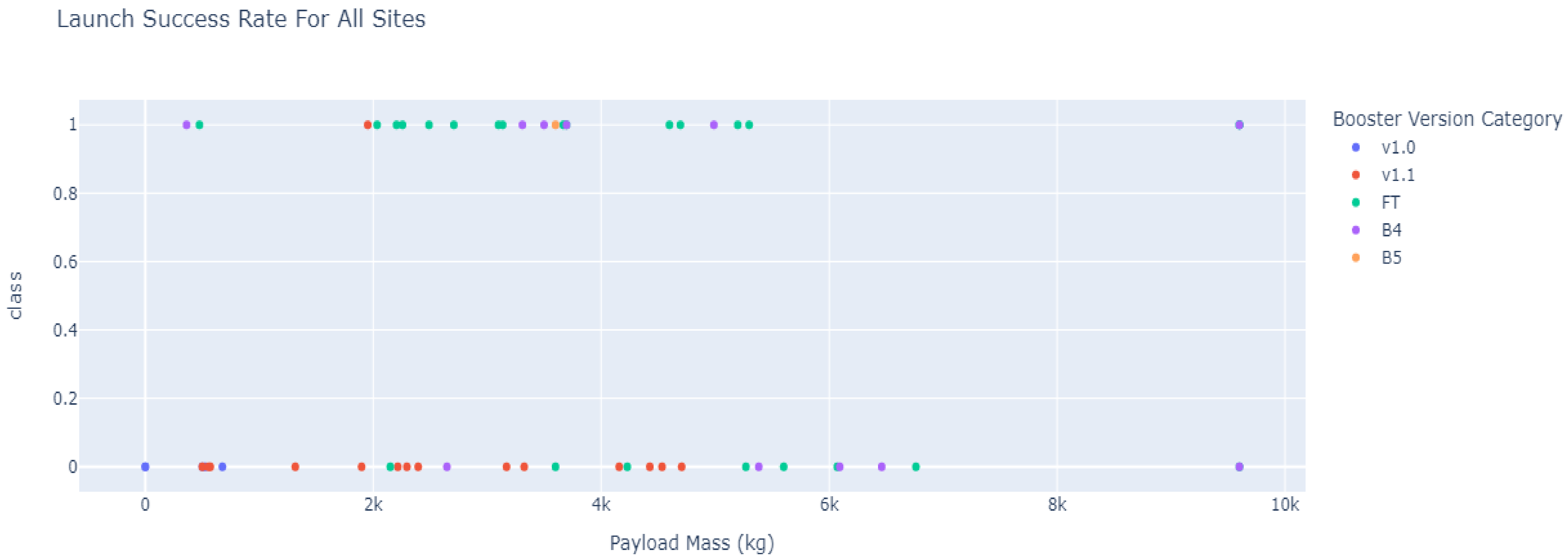


Plotly Dash Dashboard of Launch Success for each site

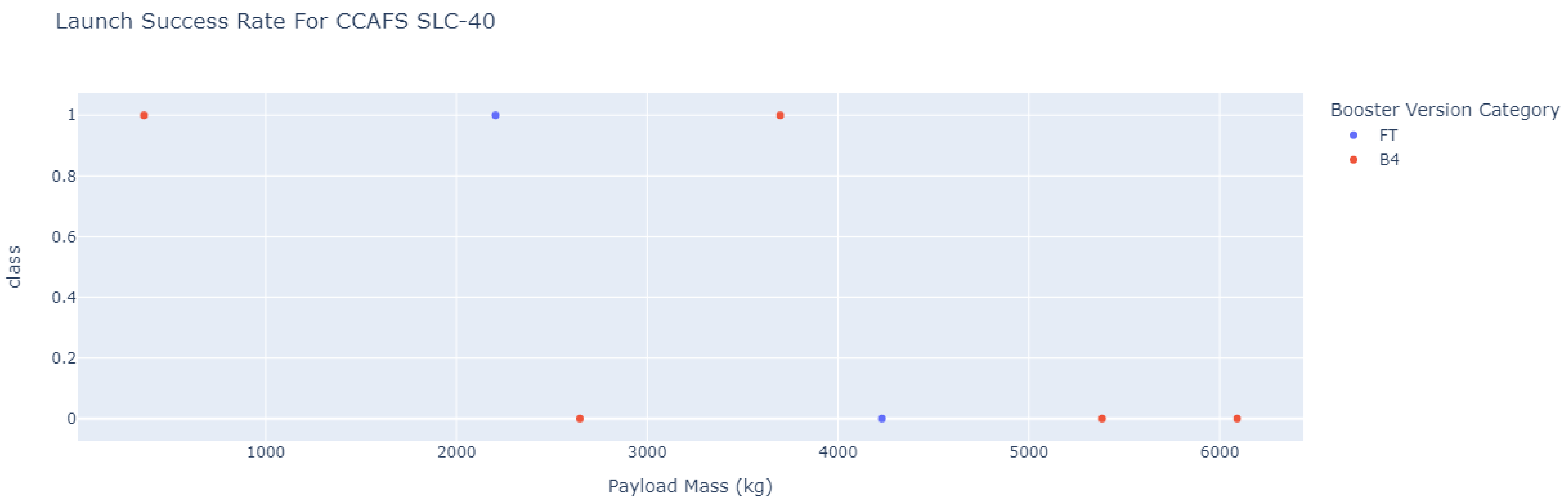
Maximum success rate (42.9 %) was recorded from CCAFS SLC-40 with 55 launches/90.



Plotly Dash Dashboard of Payload Mass Vs. Class and Booster Version Category



Plotly Dash Dashboard Results



Predictive Analysis (Classification) Results

Logistic Regression (LR)

Support Vector Machine (SVM)

Decision Tree

K Nearest Neighbors (KNN)

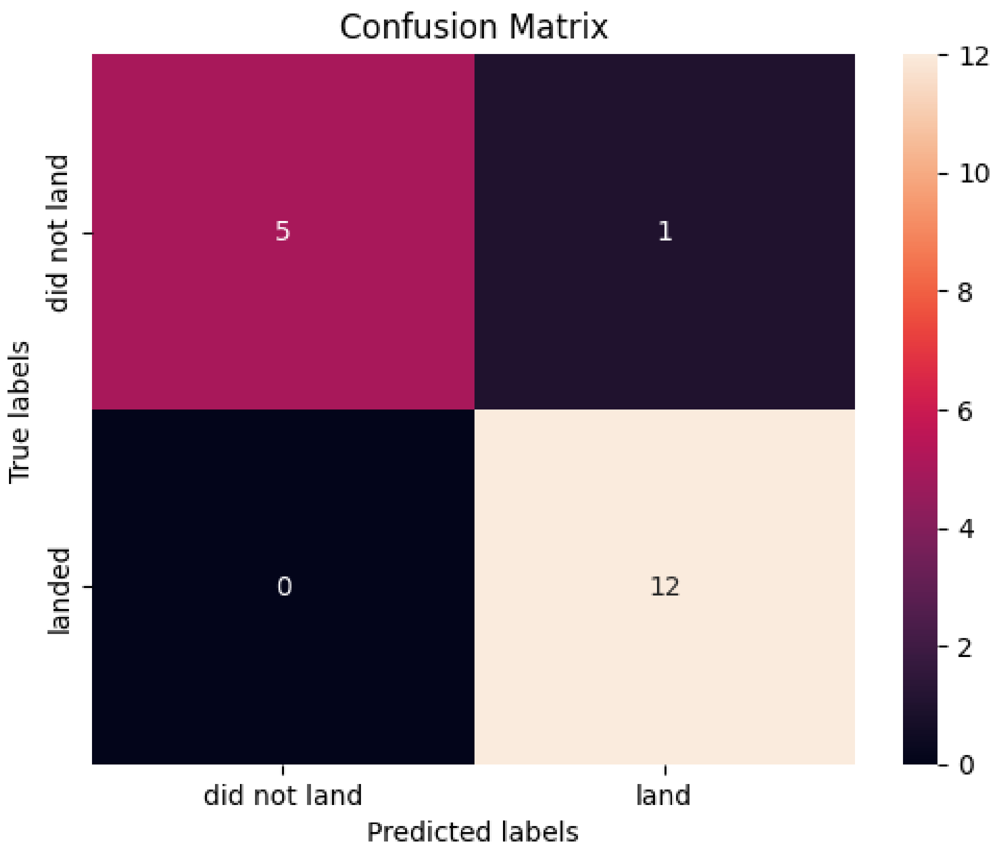
Predictive Analysis (Classification) Results

Logistic Regression (LR)

$$\text{Accuracy} = 0.822 \longrightarrow 82.2\ \%$$

$$\begin{cases} R^2 = 0.875 \longrightarrow \text{Train Data} \\ R^2 = 0.944 \longrightarrow \text{Test Data} \end{cases}$$

$$\text{Accuracy (KFold)} = 0.818 \longrightarrow 81.8\ \%$$



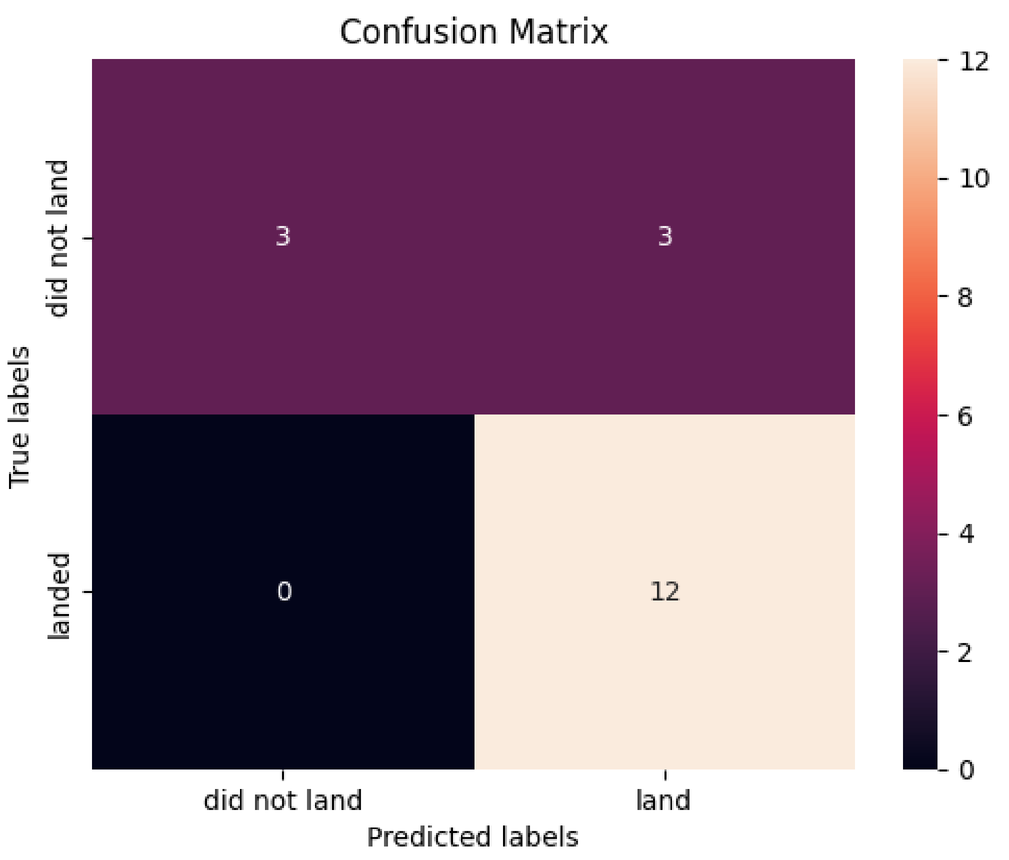
Predictive Analysis (Classification) Results

Support Vector Machine (SVM)

$$\text{Accuracy} = 0.848 \longrightarrow 84.8\ \%$$

$$\begin{cases} R^2 = 0.888 \longrightarrow \text{Train Data} \\ R^2 = 0.833 \longrightarrow \text{Test Data} \end{cases}$$

$$\text{Accuracy (KFold)} = 0.761 \longrightarrow 76.1\ \%$$



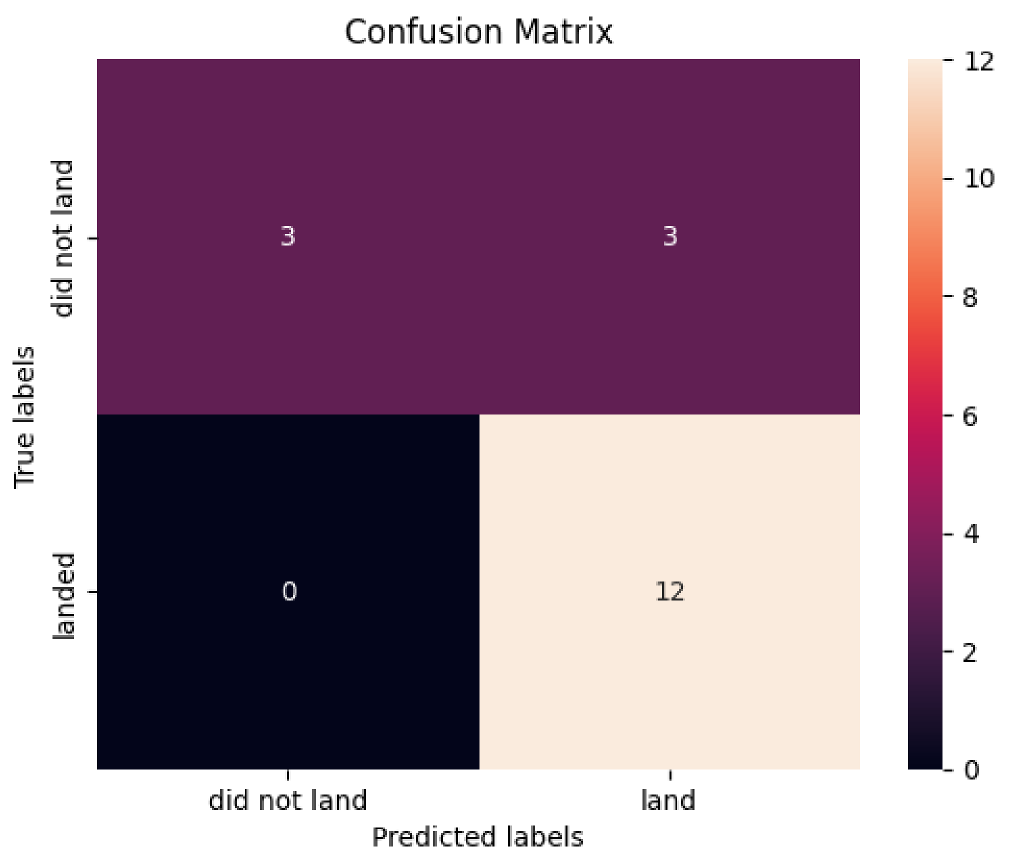
Predictive Analysis (Classification) Results

Decision Tree

$$\text{Accuracy} = 0.875 \longrightarrow 87.5\,\%$$

$$\begin{cases} R^2 = 0.903 \longrightarrow \text{Train Data} \\ R^2 = 0.833 \longrightarrow \text{Test Data} \end{cases}$$

$$\text{Accuracy (KFold)} = 0.764 \longrightarrow 76.4\,\%$$



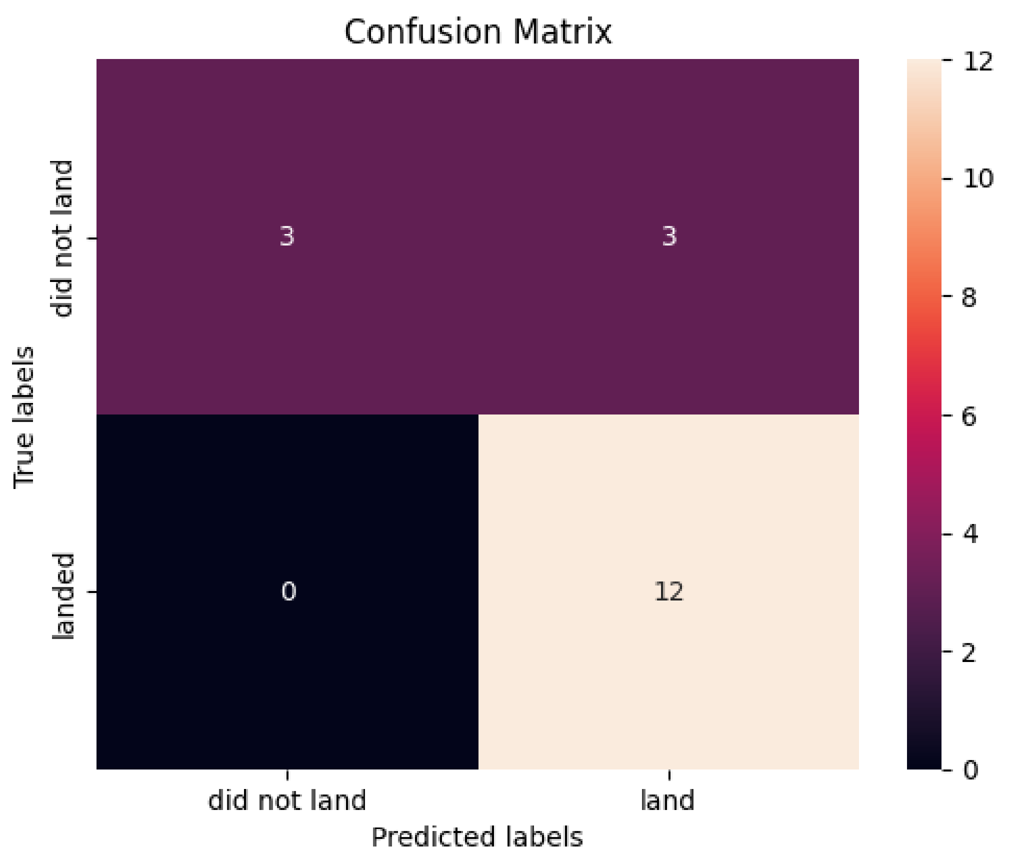
Predictive Analysis (Classification) Results

K Nearest Neighbors (KNN)

$$\text{Accuracy} = 0.848 \longrightarrow 84.8\ \%$$

$$\begin{cases} R^2 = 0.861 \longrightarrow \text{Train Data} \\ R^2 = 0.833 \longrightarrow \text{Test Data} \end{cases}$$

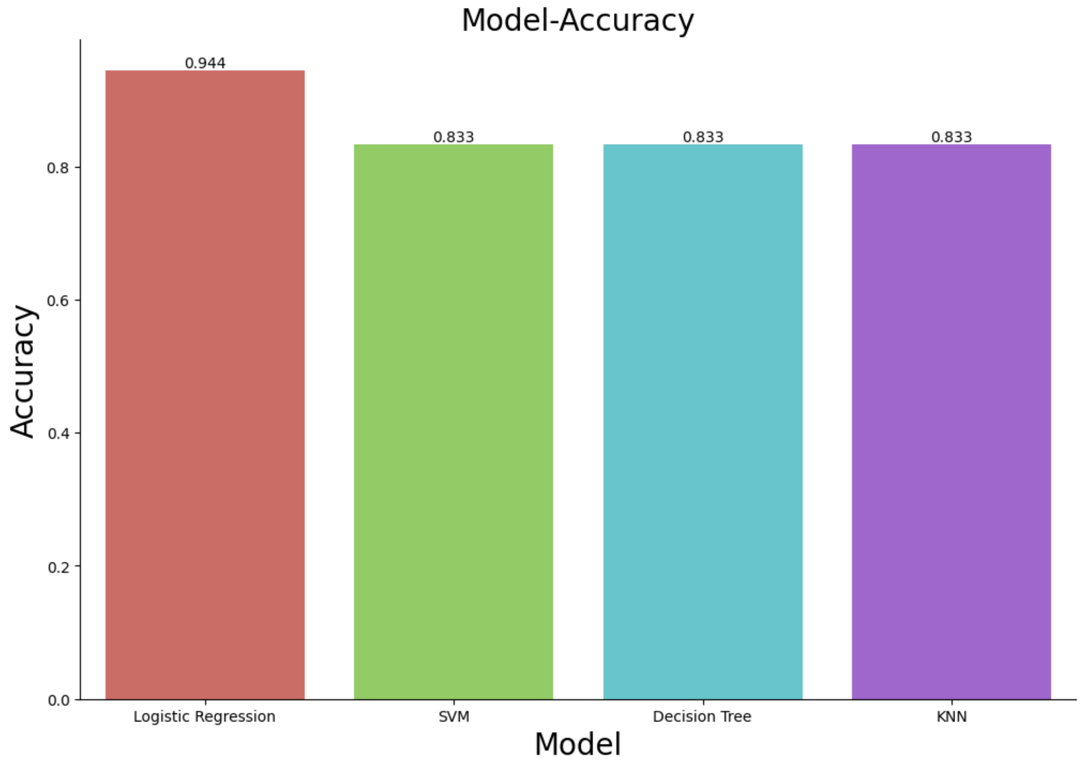
$$\text{Accuracy (KFold)} = 0.834 \longrightarrow 83.4\ \%$$

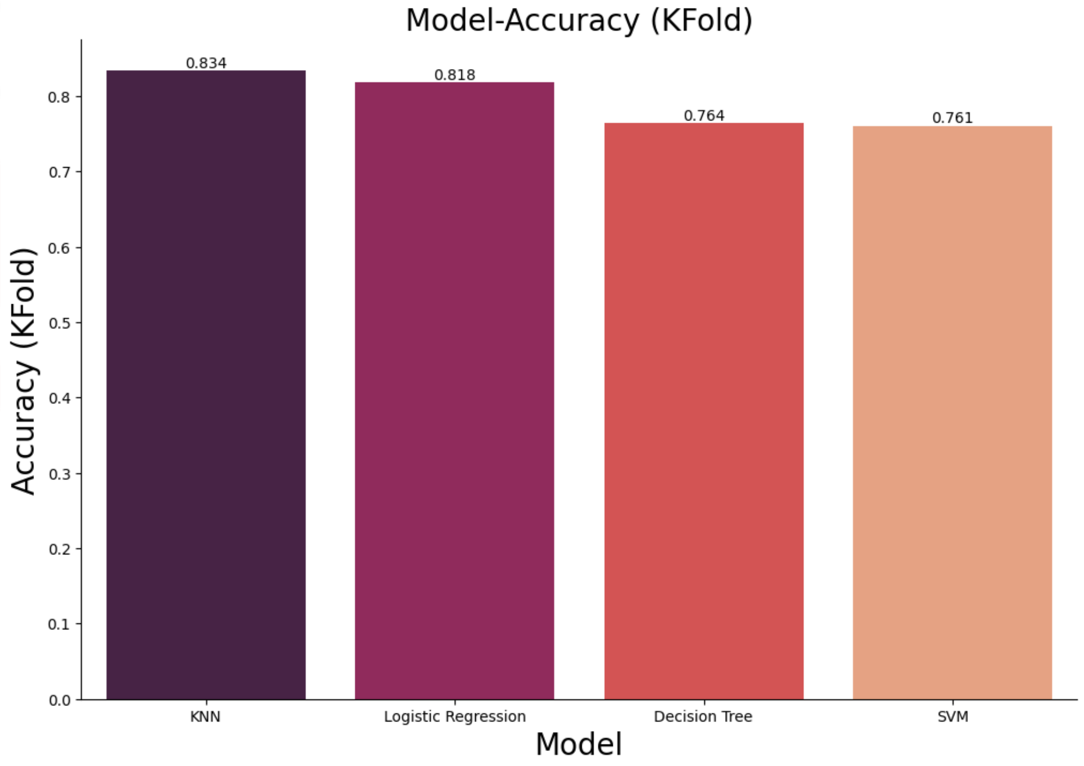


Accuracy and performance of the ML models

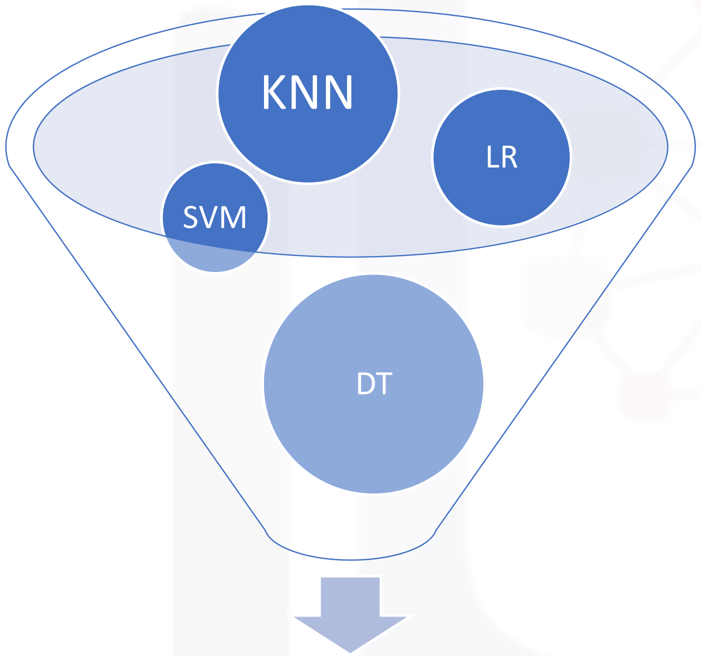
Logistic Regression has the greatest R^2

KNN has the greatest Accuracy using KFold





Best ML models for this study



LR (Logistic regression)

Logistic regression has the best R^2 =0.944 which means that the model has an excellent fit while KNN has R^2 =0.833. On the other hand, KNN has an accuracy (Kfold) of 0.834 and 0.818 for the Logistic regression.

After considering all these results, we can conclude that the best applicable model for this study is the Logistic Regression model.

CONCLUSION

Study the prediction of a success launching and landing of the rockets will let us to estimate the amount of each Launch/Landing.

Our methodology consist on collecting and preparing the data of our project, next step was exploring these data in order to familiarize and understand the relationship between them.

We used four machine learning models (Logistic Regression, Support Virtual Machine, Decision Tree, K nearest neighbors) in order to calculate the accuracy and performance of each model, the calculations show that best model with the best fitting/Accuracy for our project study is the Logistic Regression.