IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Rodrigue DJOUONTU

04-04-2024

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Project background and context
- Problems you want to find answers

Section 1

Methodology

Methodology

~~Executive Summary~~

- Data collection methodology:
  - Data is collected in the web using webscraping methodology, we also collected data through API REST
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- Describe how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts

- GitHub URL of the completed SpaceX API calls notebook:

https://github.com/djouontu/Spacex-launch-prediction/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Place your flowchart of SpaceX API calls here

Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts
- GITHUB URL:

https://github.com/djouontu/Spacex-launch-prediction/blob/main/jupyter-labs-webscraping.ipynb

Place your flowchart of web scraping here

Data Wrangling

- Describe how data were processed
- You need to present your data wrangling process using key phrases and flowcharts
- GitHub URL for data wrangling related notebooks:

https://github.com/djouontu/Spacex-launch-prediction/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
- Add the GitHub URL of your completed EDA with data visualization notebook:

https://github.com/djouontu/Spacex-launch-prediction/blob/main/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
- Add the GitHub URL of your completed EDA with SQL notebook:

https://github.com/djouontu/Spacex-launch-prediction/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- This step consist to add on the map the location of the launch site. Analys the distance between each site.
- Add the GitHub URL of your completed interactive map with Folium map:

https://github.com/djouontu/Spacex-launch-prediction/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

We have builded a Plotly Dash application for users to perform interactive visual analytics on SpaceX launch data in real-time. this dashboard application contains pie chart and a scatter point chart, so you can see the correlation between payload mass and success of launch. You can also identifie the launch sit with highest rate of success.

Add the GitHub URL of your completed Plotly Dash lab:

https://github.com/djouontu/Spacex-launch-prediction/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

We built various classification models such decision tree, logistique regression, super vector machine and k nearest neighbors. We have improved this model by finding the best parameters, after those steps we have shoose the best model

GitHub URL of your completed predictive analysis lab:

https://github.com/djouontu/Spacex-launch-prediction/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

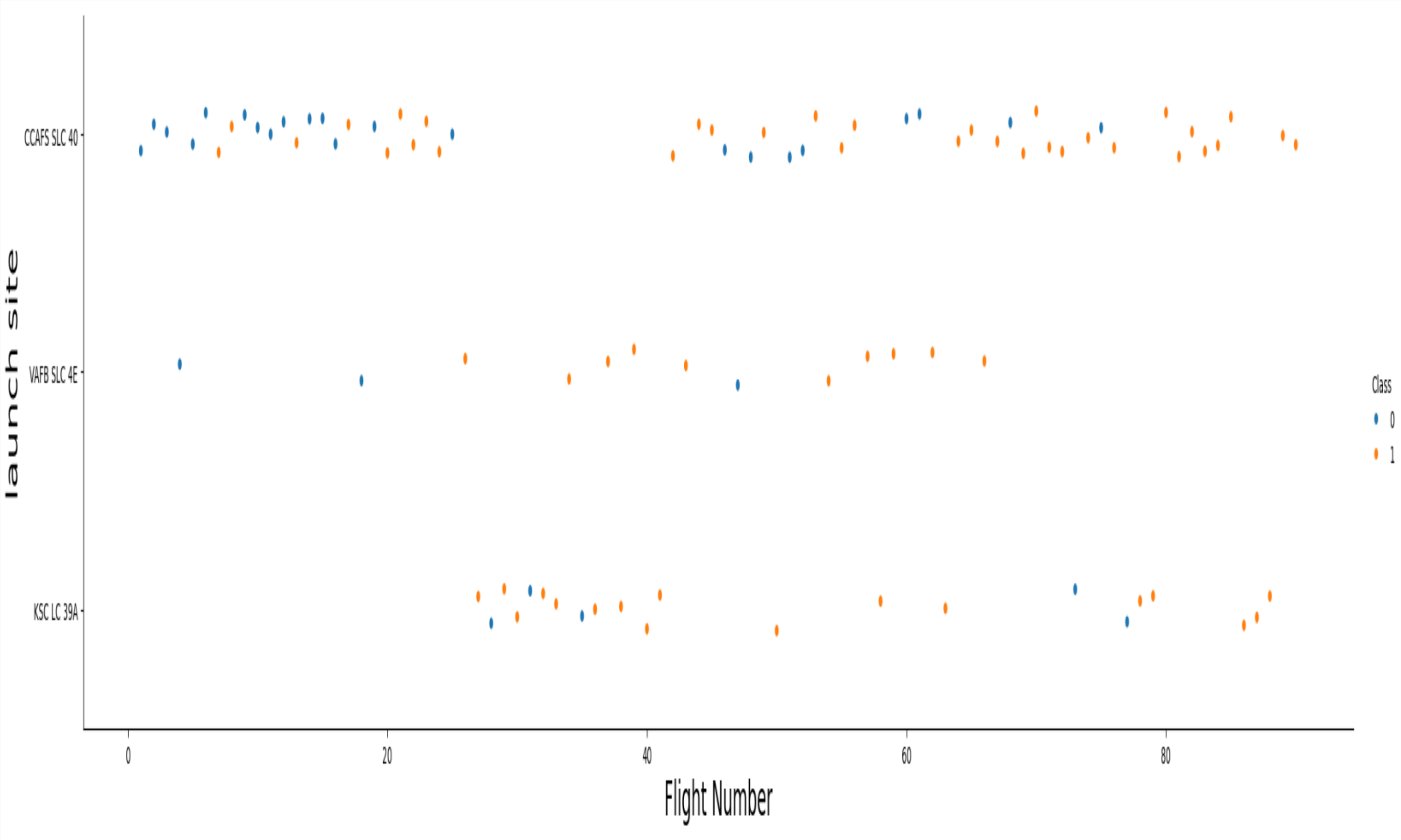
Results

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

Section 2

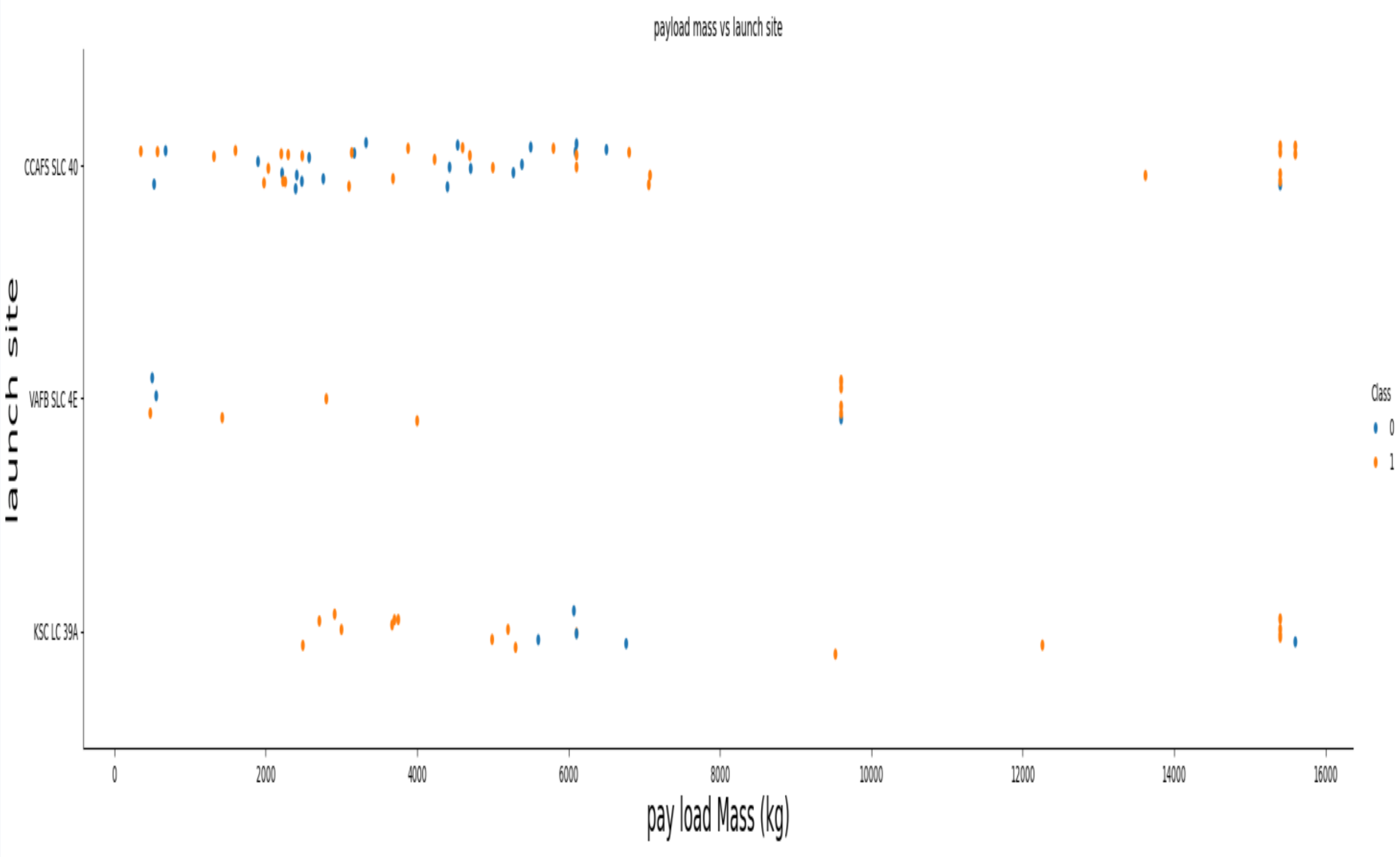
Insights drawn from EDA

Flight Number vs. Launch Site



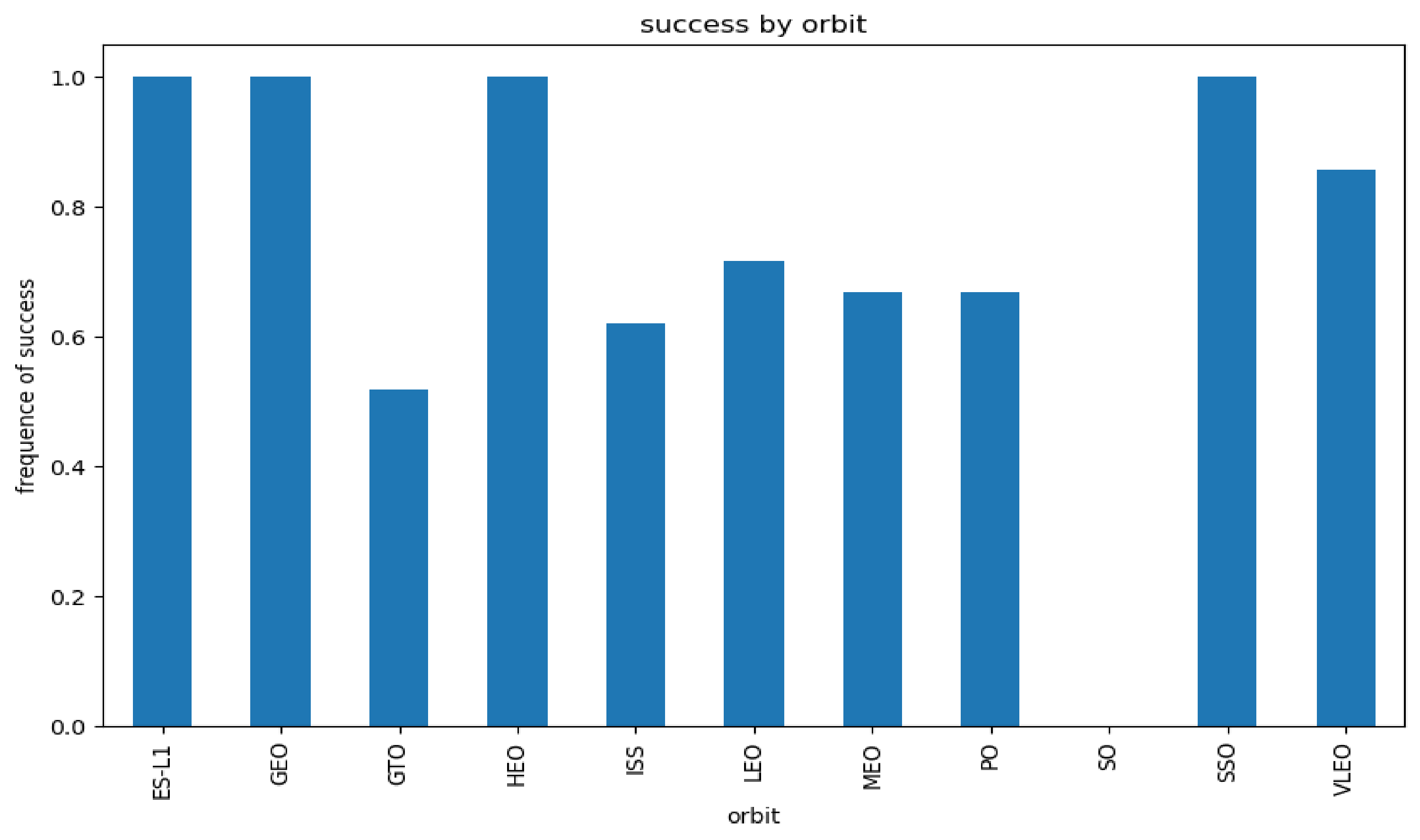
This scatter plot show that most launches are done from the site CCAFS LC-40.

Payload vs. Launch Site



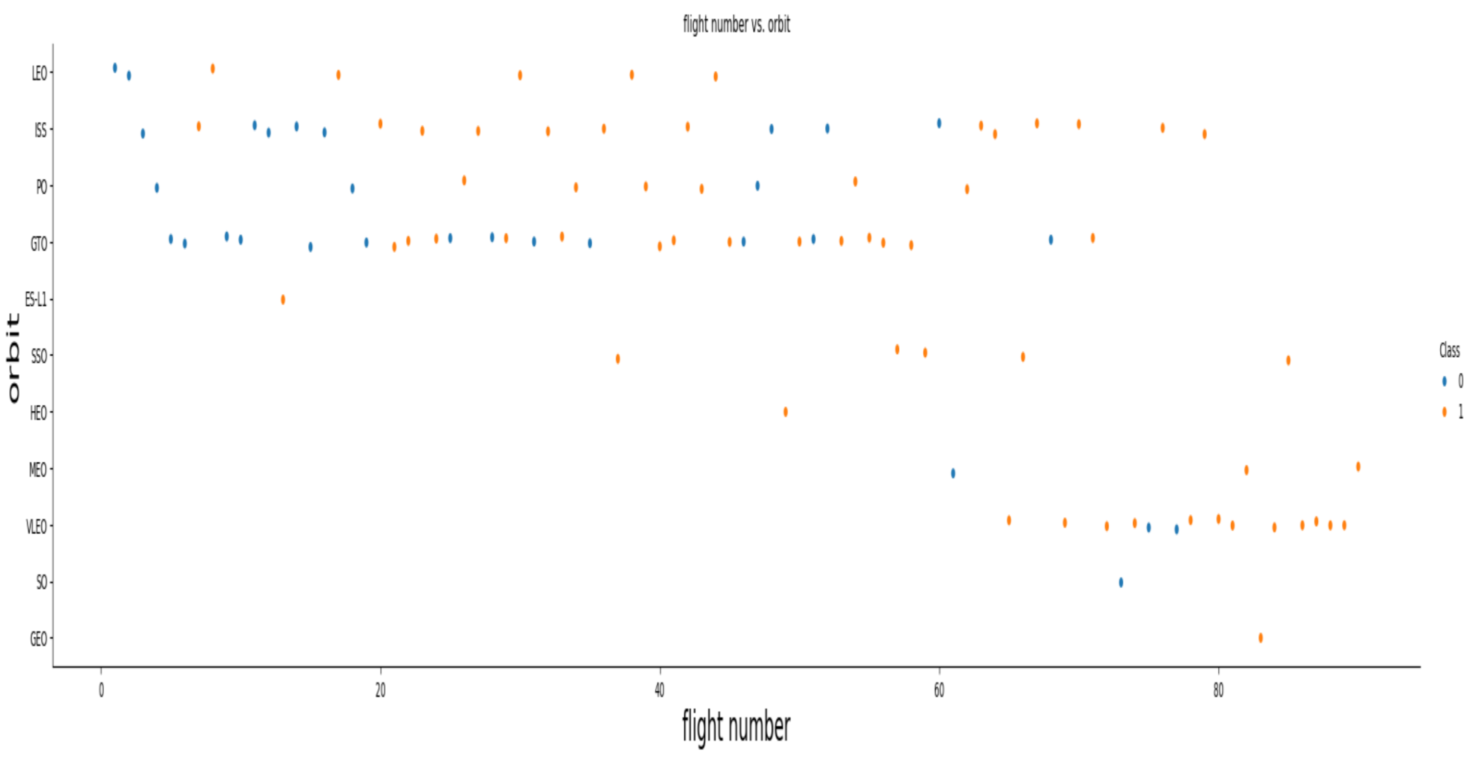
Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

Success Rate vs. Orbit Type



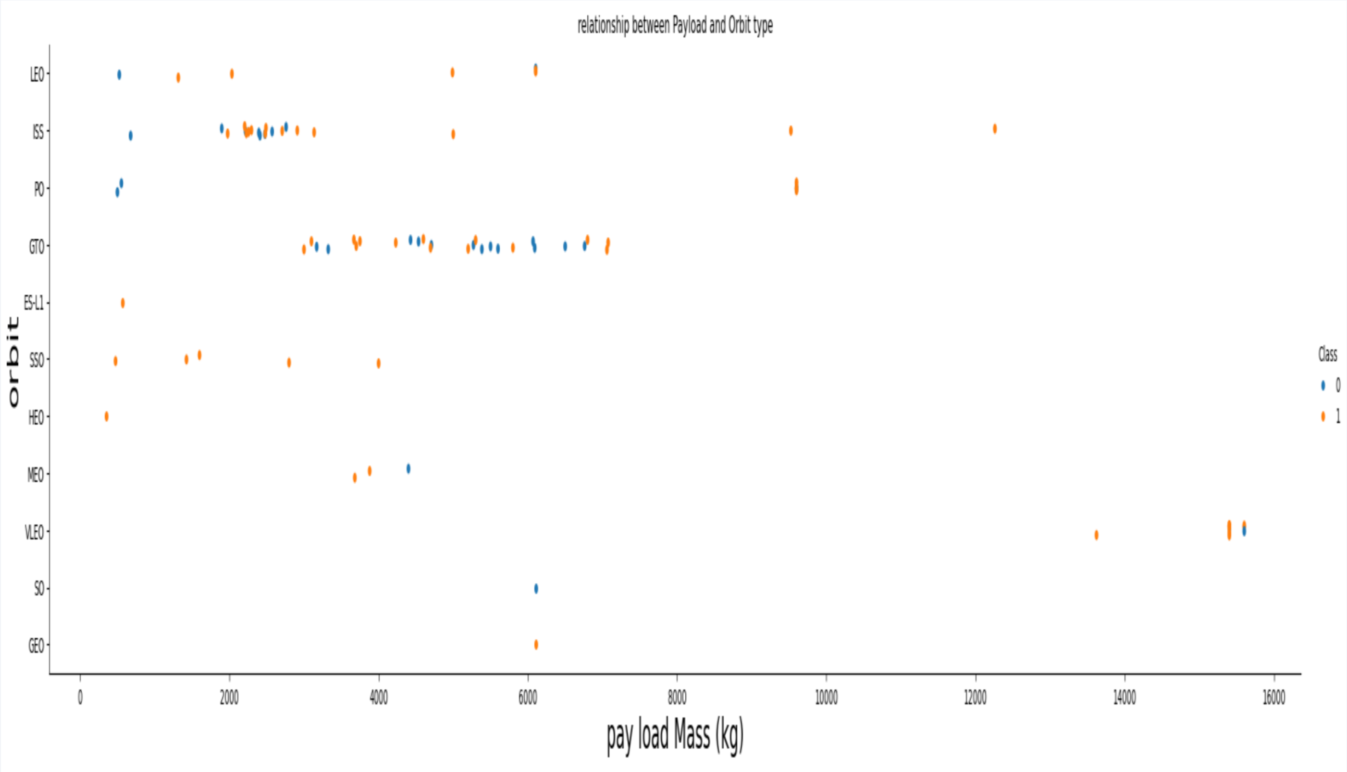
Orbit ES-L1, GEO, HEO and SSO have highest rates of success.

Flight Number vs. Orbit Type



You should see that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

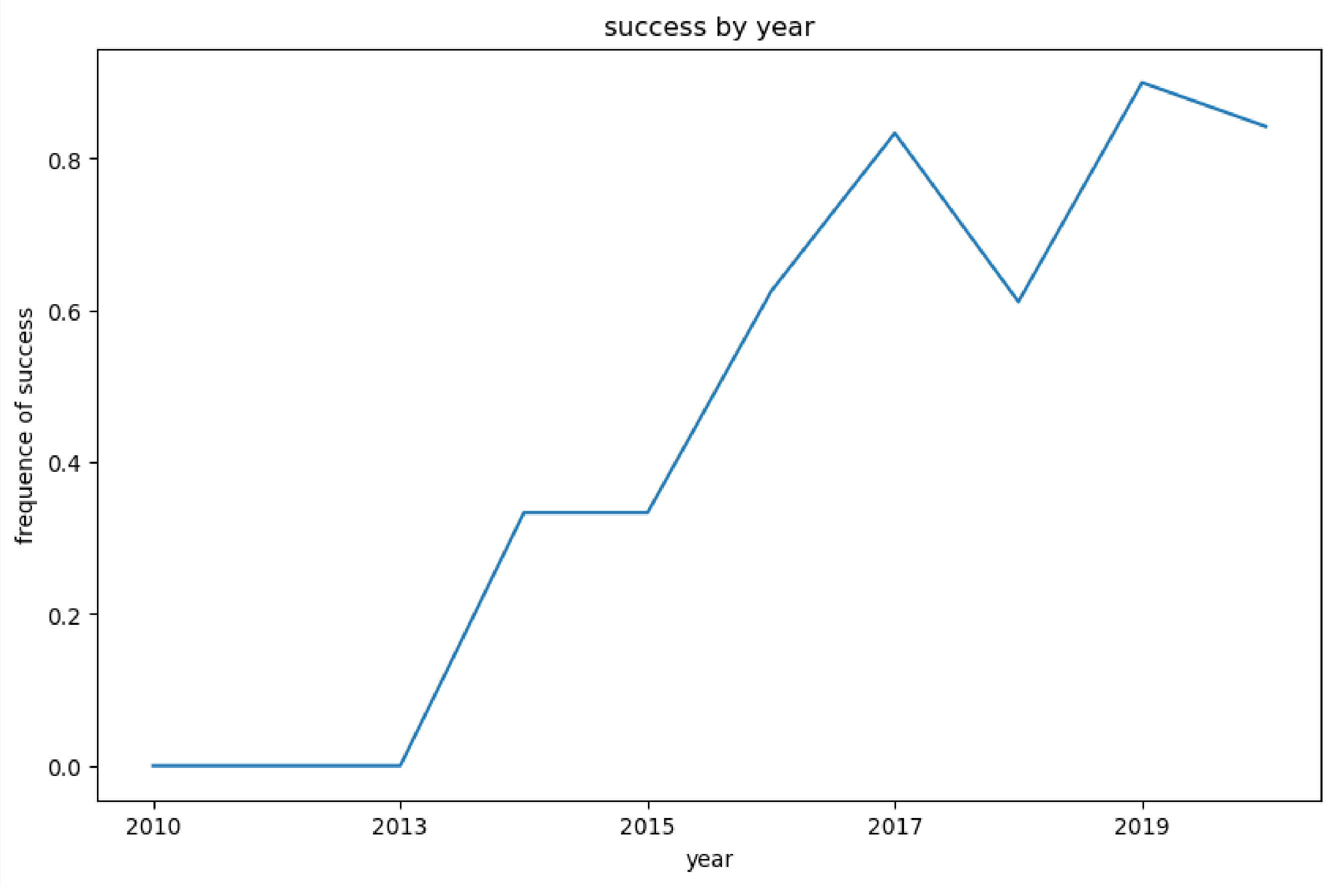
Payload vs. Orbit Type



With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.

However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.

Launch Success Yearly Trend



You can observe that the success rate since 2013 kept increasing till 2017 (stable in 2014) and after 2015 it started increasing.

All Launch Site Names

Names of the unique launch sites

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

Query result

%sql select distinct Launch_Site from SPACEXTABLE

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`

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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt

- Query result with a short explanation here

%sql select * from SPACEXTABLE where Launch_Site like 'CCA%' limit 5

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

Total
45596

- Query result with a short explanation here

```
%sql select SUM(PAYLOAD_MASS__KG_) as Total from SPACEXTABLE where Customer='NASA (CRS)'
```

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1

mean
2928.4

- Query result

%sql select AVG(PAYLOAD_MASS__KG_) as mean from SPACEXTABLE where Booster_Version='F9 v1.1'

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

Date: 2018-07-22

- Query result

%sql select min(Date) as date from (select * from SPACEXTABLE where Landing_Outcome= 'Success')

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Booster_Version	PAYLOAD_MASS__KG_
F9 v1.1	4535
F9 v1.1 B1011	4428
F9 v1.1 B1014	4159
F9 v1.1 B1016	4707
F9 FT B1020	5271
F9 FT B1022	4696
F9 FT B1026	4600
F9 FT B1030	5600
F9 FT B1021.2	5300
F9 FT B1032.1	5300
F9 B4 B1040.1	4990
F9 FT B1031.2	5200
F9 B4 B1043.1	5000
F9 FT B1032.2	4230
F9 B4 B1040.2	5384
F9 B5 B1046.2	5800

Query result

%sql select Booster_Version, PAYLOAD_MASS__KG_ from SPACEXTABLE where PAYLOAD_MASS__KG_> 4000 and PAYLOAD_MASS__KG_< 6000

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

Total number: 101

- Query result

```
%sql select count(Mission_Outcome) from SPACEXTABLE
```

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

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

- Query result

```
%%sql select Booster_Version, PAYLOAD_MASS__KG_ from SPACEXTABLE where
PAYLOAD_MASS__KG_ = (select max( PAYLOAD_MASS__KG_) from SPACEXTABLE)
```

2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

month	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

- Present your query result with a short explanation here

%%sql select substr(Date, 6,2) as month, Landing_Outcome, Booster_Version,Launch_Site from SPACEXTABLE where (Landing_Outcome like 'Failure (drone ship)' and Date like '2015%')

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

- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

- Present your query result with a short explanation here

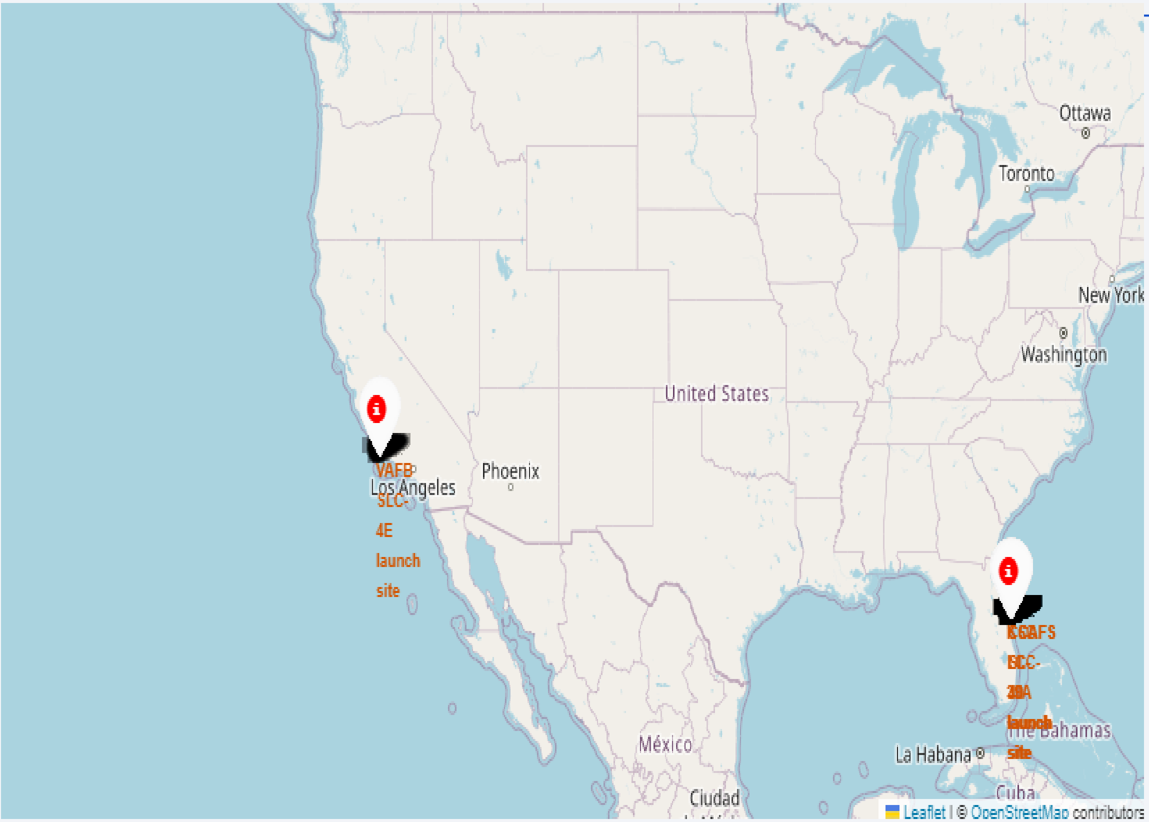
Section 3

Launch Sites Proximities Analysis

Geographics locations of launch sites

We see that most launch site are located at west cote of USA when only one is at East coast

Mark the success/failed launches for each site on the map



<Folium Map Screenshot 3>

- Replace <Folium map screenshot 3> title with an appropriate title

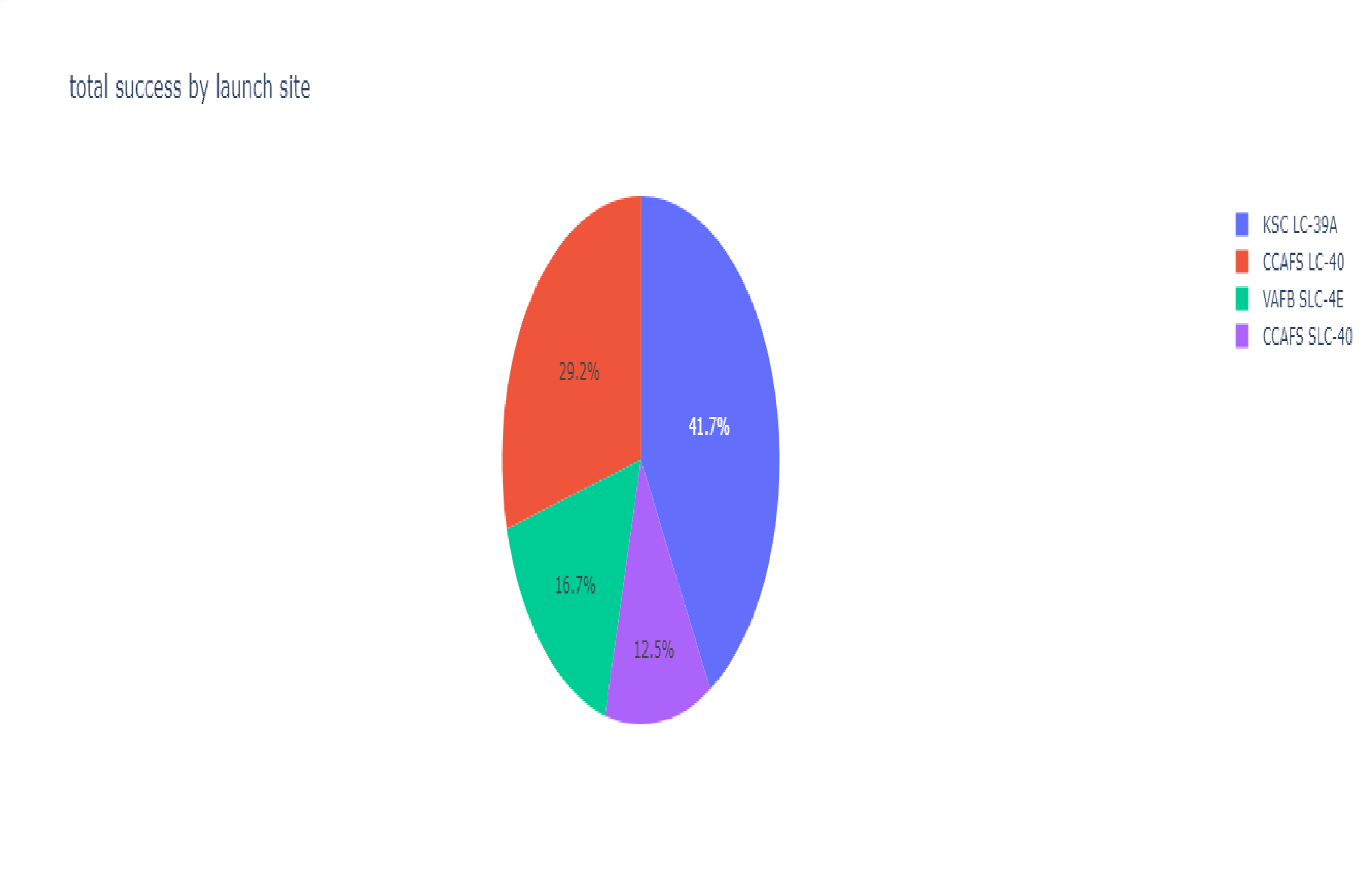
- Explore the generated folium map and show the screenshot of a selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed

- Explain the important elements and findings on the screenshot

Section 4

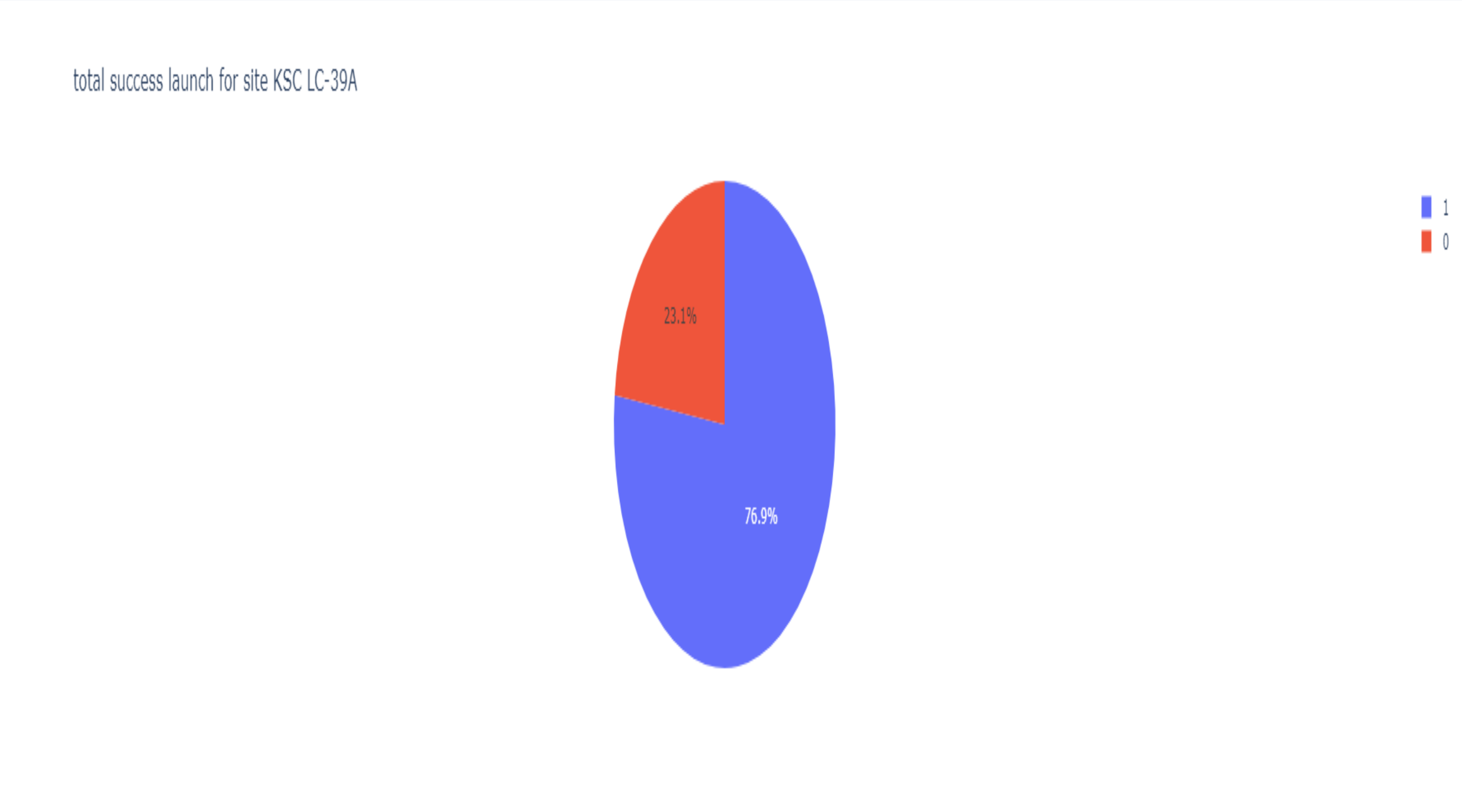
Build a Dashboard with Plotly Dash

Total success launch by site



This piechart represent the distribution od success launch by site. In regard of this piture, the launch site KSC LC-39A has the highest rate of success.

Success ration of Launch site KSC LC-39A

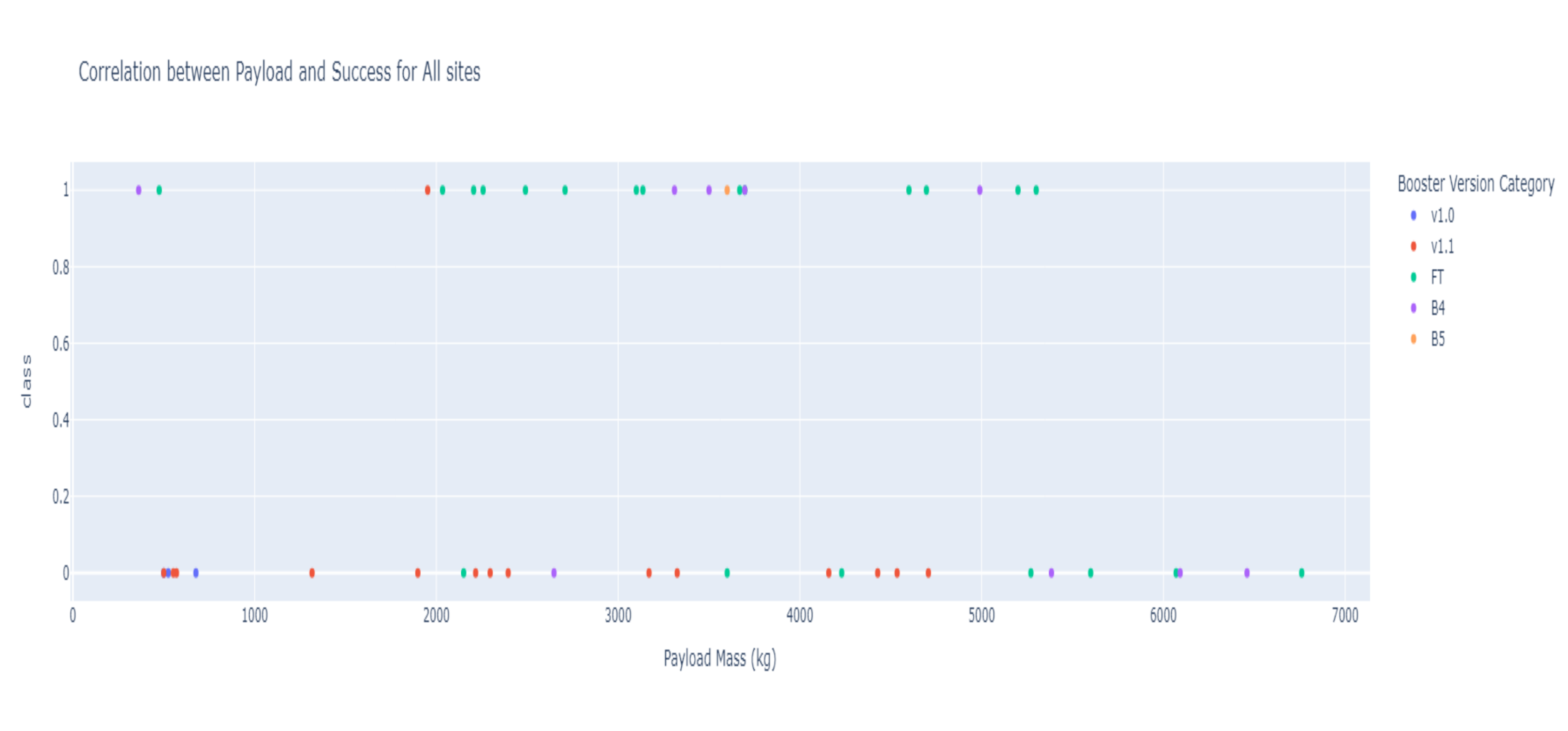


KSC-39A has the best outcome belong all others sites.

This picture shoes the ration of success on this site.

“1” represent de launch success outcoume and “0” represent de loss ie the rocket that has not land.

Relation between payload and launch outcome

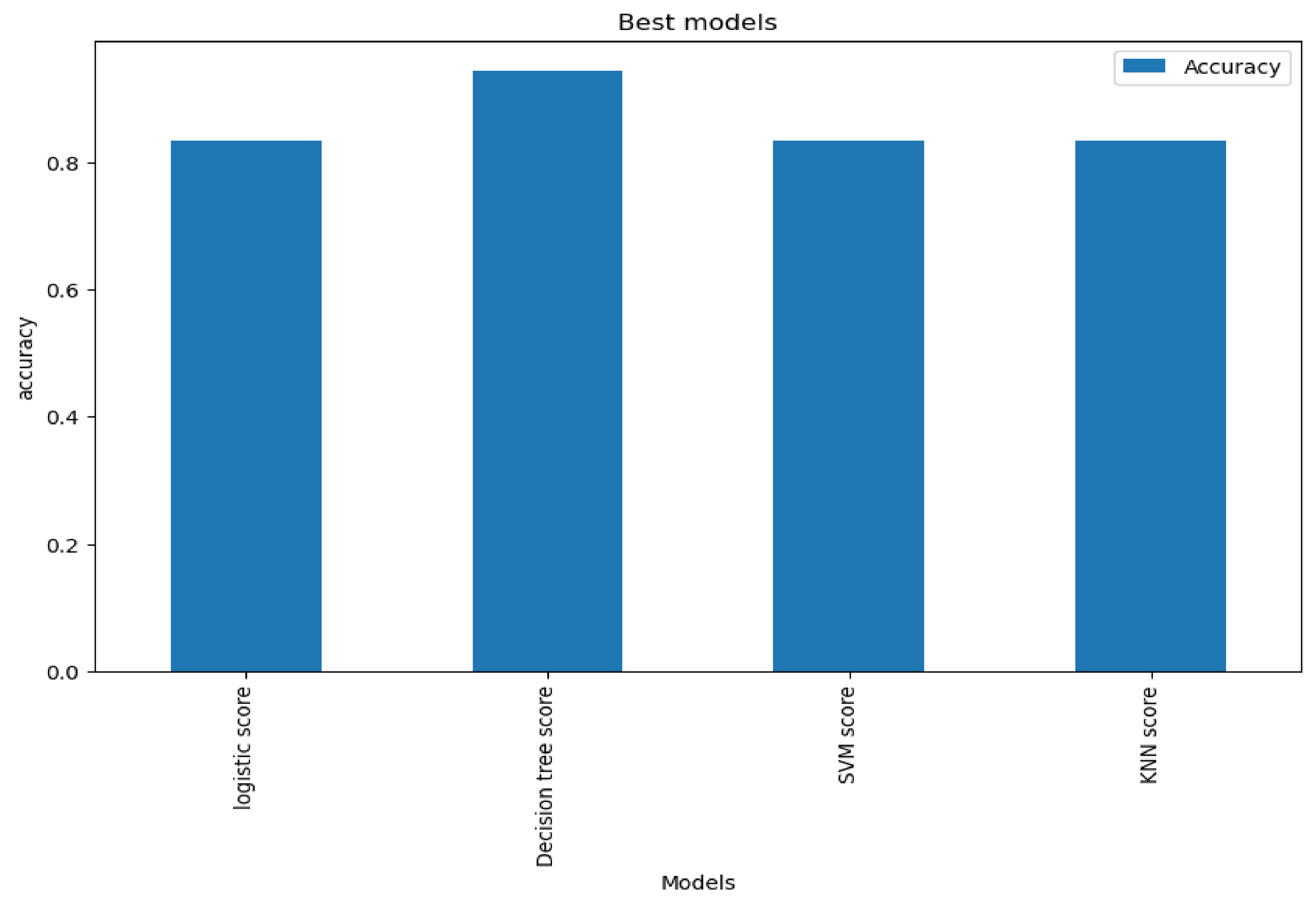


Regarding this image the most success has the payload between 2000 kg and 4000 kg.

Section 5

Predictive Analysis (Classification)

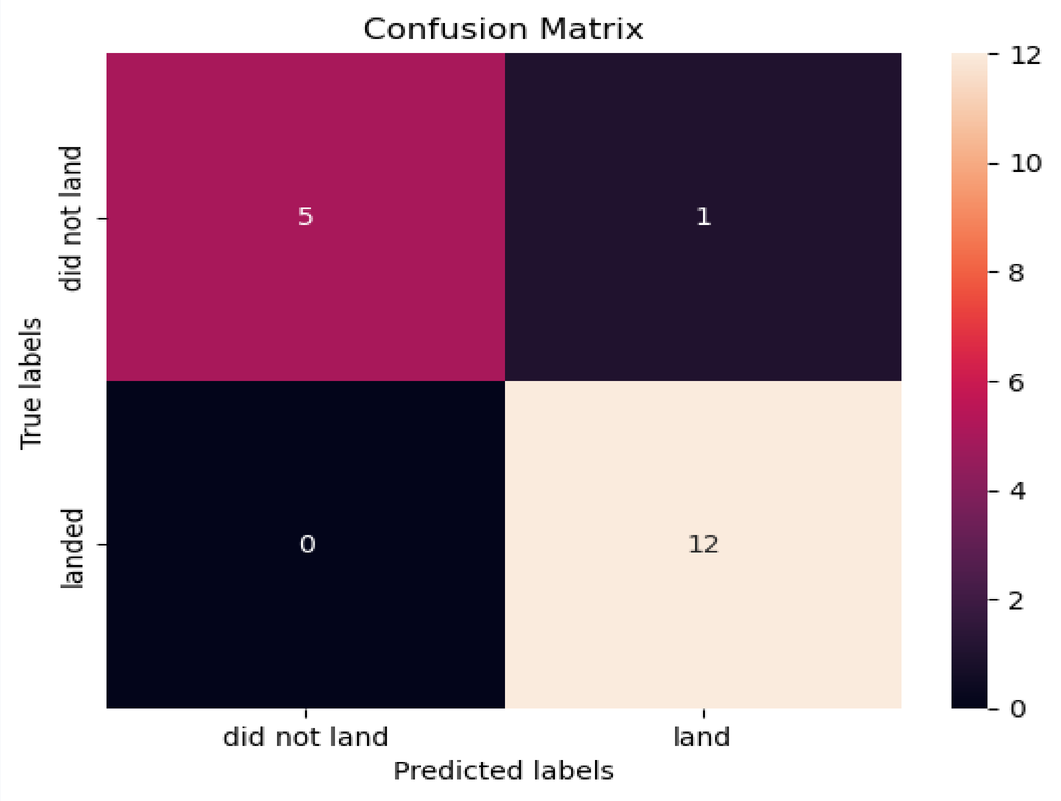
Classification Accuracy



We can see that Decision tree model has the highest classification accuracy.

Models	Accuracy
logistic score	83,33%
Decision tree score	94,44%
SVM score	83,33%
KNN score	83,33%

Confusion Matrix



Confusion matrix of the best model ie decision tree. It shows that the model predict

Conclusions

After completion this project, we see that the best model to predict if the Falcon 9 first stage will land successfully is the decision tree model.

Also the successful land depend of the orbit, the payload mass and the locations

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!