IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Rakesh Mangore
7th Sep 2024

Outline

Executive Summary

- Summary of methodologies
  - Data collection from SpaceX API and web scrapping
  - Exploratory Data Analysis (EDA) using SQL, Pandas and Matplotlib
  - Interactive Visual Analytics and Dashboard using plotly Dash and folium
  - Predictive Analysis (Classification)
- Summary of all results
  - Machine learning Classification model to predict the best model
  - Exploratory Data Analysis identifies the best location and booster Rocket

Introduction

- To be in a competition of Space Race with SpaceX, a rival company SpaceY wants to make a predictions about the failure/success of SpaceX Rockets first stage landings.
- Which machine learning model will fit the data and predict the future launch.
- Will a future launch landing be successful?

Section 1

Methodology

Methodology

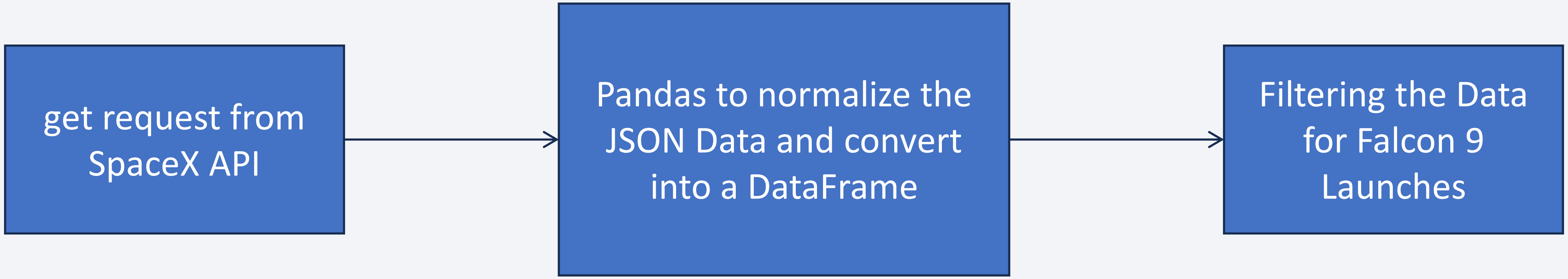
Executive Summary

- Data collection methodology:
  - Data was collected from the SpaceX Api and the web Scrapping of wiki pages.
- Perform data wrangling
  - Data collection is in the form of JSON object with HTML parser, data is converted into the pandas Dataframe for further analysis
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Different Machine learning Classification model performed on the data to predict the launching outcome

Data Collection

- Data is collected from two sources
  - ➢First, from SpacaeX API by get request
  - ➢Second, from web scrapping of SpaceX wiki Pages

Data Collection – SpaceX API

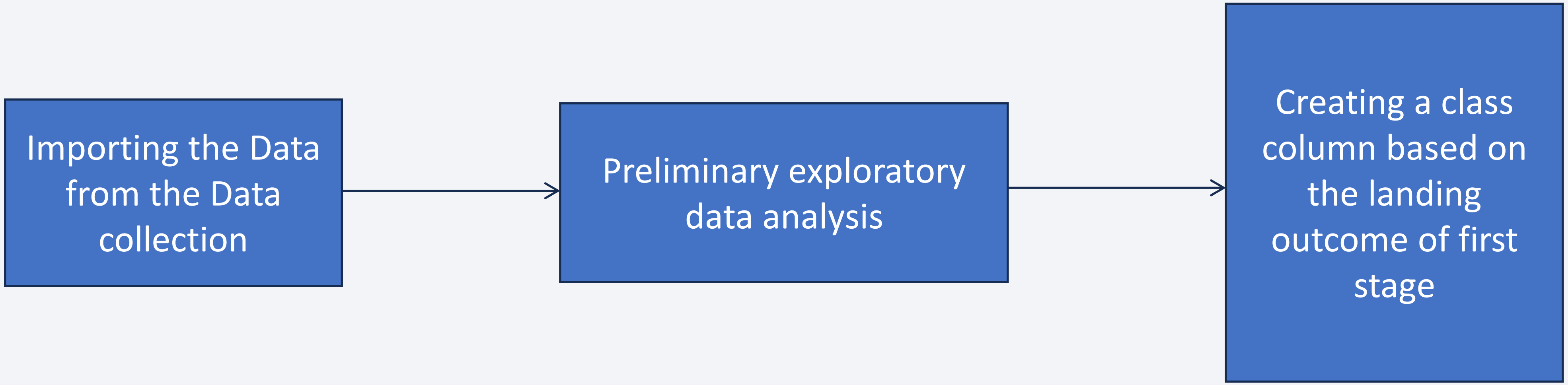


https://github.com/rmangore/Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

https://github.com/rmangore/Capstone/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling



https://github.com/rmangore/Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Visualize the relationship between Flight Number and Launch Site
  - Launch site VAFB-SLC-4E success rate increases with the flight number
- Visualize the relationship between Payload Mass and Launch Site
  - Launch site VAFB-SLC-4E success rate increases with the Payload Mass
- Visualize the relationship between success rate of each orbit type
  - Orbits ES-L1, GEO, HEO, SSO have 100% success rate
- Visualize the relationship between Flight Number and Orbit type
  - Orbits LEO, SSO have increased success rate by flight number
- Visualize the relationship between Payload Mass and Orbit type
  - Orbits ISS, SSO has higher success rate by increased payload mass
- Visualize the launch success yearly trend
  - The success rate since 2013 kept increasing till 2020

https://github.com/rmangore/Capstone/blob/main/edadataviz.ipynb

EDA with SQL

- Displayed the names of the unique launch sites in the space mission
- Displayed the total payload mass carried by boosters launched by NASA (CRS)
- Displayed 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
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List the names of the booster versions which have carried the maximum payload mass

https://github.com/rmangore/Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Marked all launch sites on a map using the folium marker and circle functions
- Marked the success/failed launches for each site on the map by using marker cluster function
- Calculated the distances between a launch site to its proximities are railway, coastline, highway and cities by using the lines to connect the coordinate points

https://github.com/rmangore/Capstone/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Pie chart for all the individual sites can be generated which provides the information of success and failure percentage
- Scatter plot for all the sites or individual site can be generated which provides the information of success rate and by parameters of booster versions varying by rolling parameters of Payload mass

https://github.com/rmangore/Capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- Performing the exploratory Data Analysis
  - create a column for the class using the NumPy array
  - Standardize the data by standardScaler().fit_transform()
  - Split the data into training data and test data
- Finding the best Hyperparameter for SVM, Classification Trees and Logistic Regression
  - Using the GridSearchCV Machine learning model of different estimator to predict the best hyperparameters

https://github.com/rmangore/Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

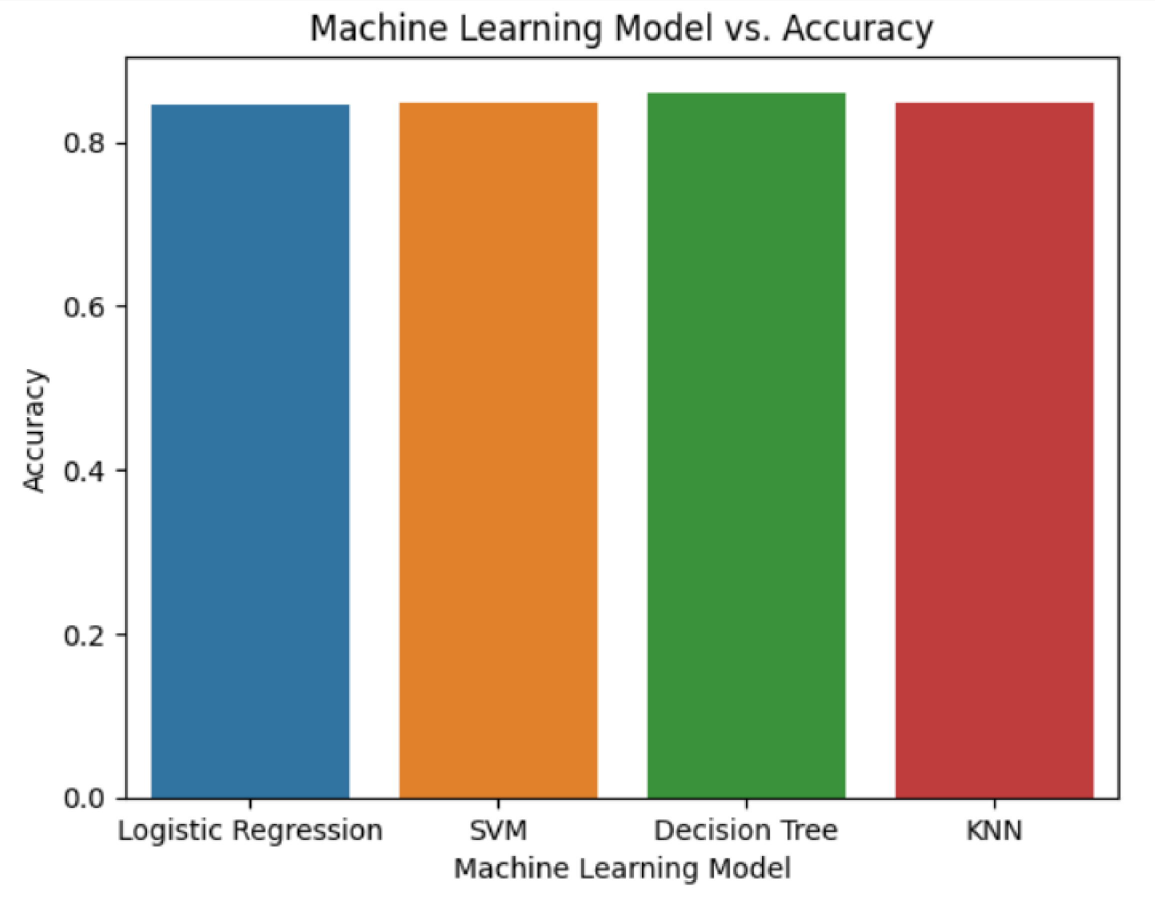
- Exploratory data analysis results
  - SpaceX uses 4 different location to launch the Rockets
  - The first success landing outcome happened in 2015 five year after the first launch
  - Orbits ES-L1, GEO, HEO, SSO have 100% success rate
  - The number of landing outcomes becomes as better as years passed
  - The success rate since 2013 kept increasing till 2020
  - The average payload mass carried by booster version F9 v1.1 is 2534.67 kg

Results

- Interactive analytics demo in screenshots

Results

- Predictive analysis results



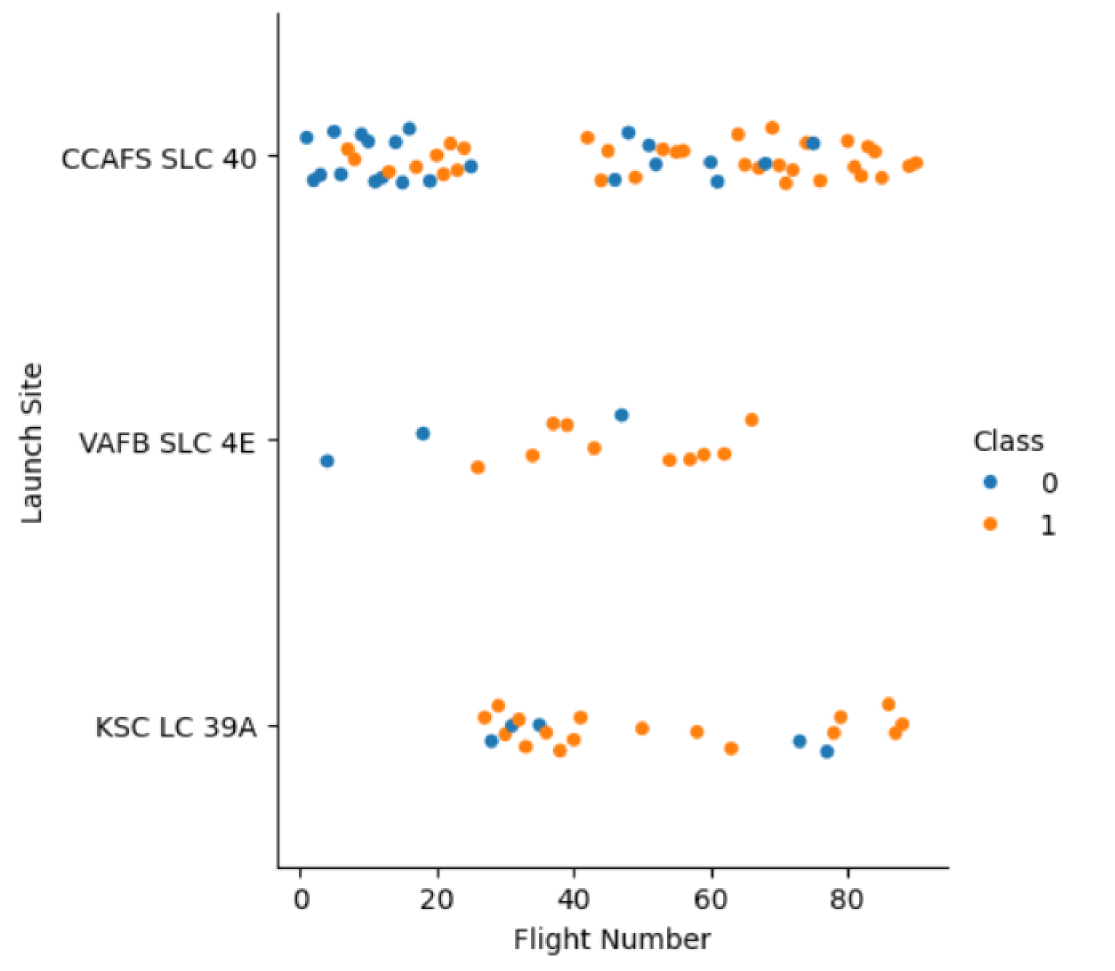
Decision Tree Classifier Machine learning model is perfect choice for predictions

Section 2

Insights drawn from EDA

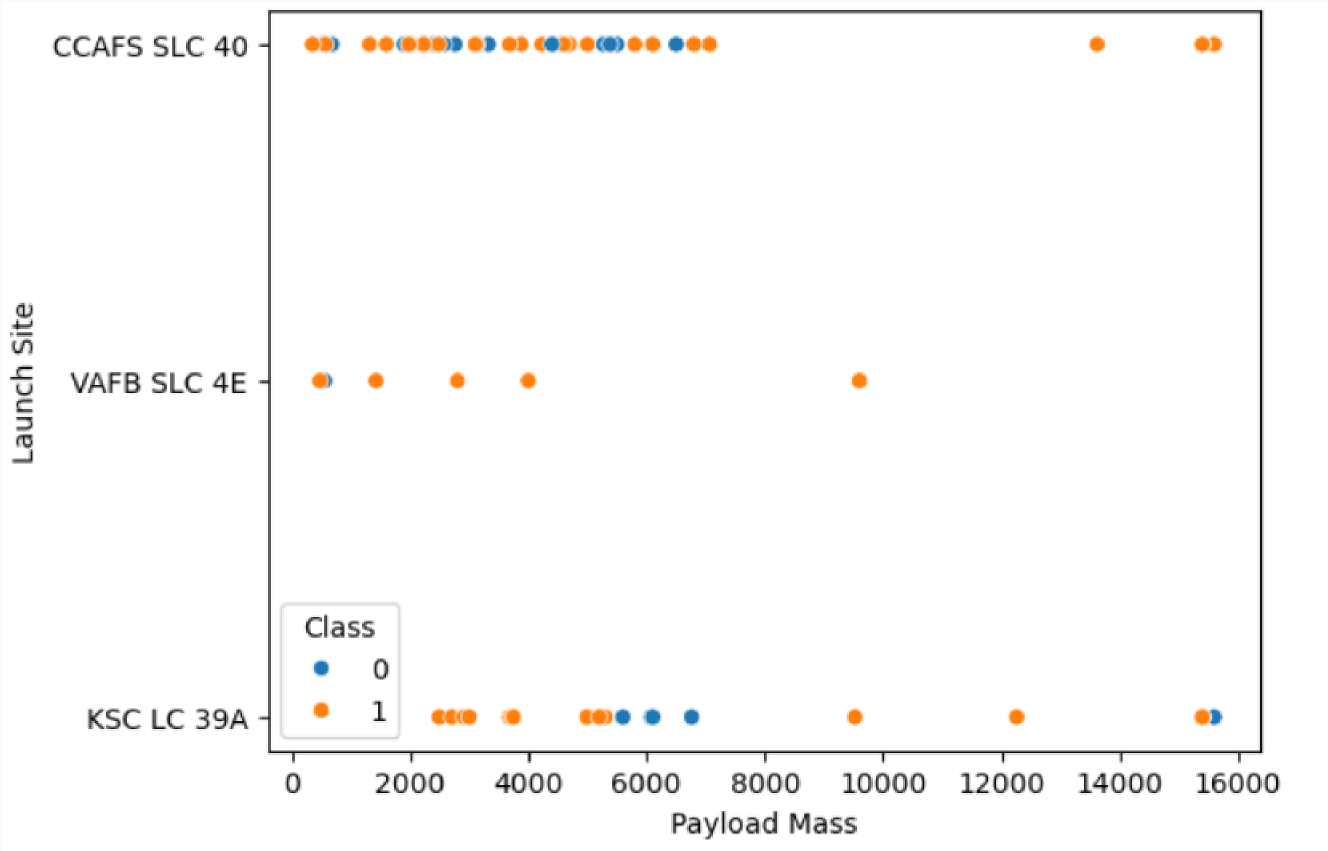
Flight Number vs. Launch Site

- VAFB-SLC-4E success rate increases with the flight number
- Other sites doesn’t have any correlations to get derived any conclusion



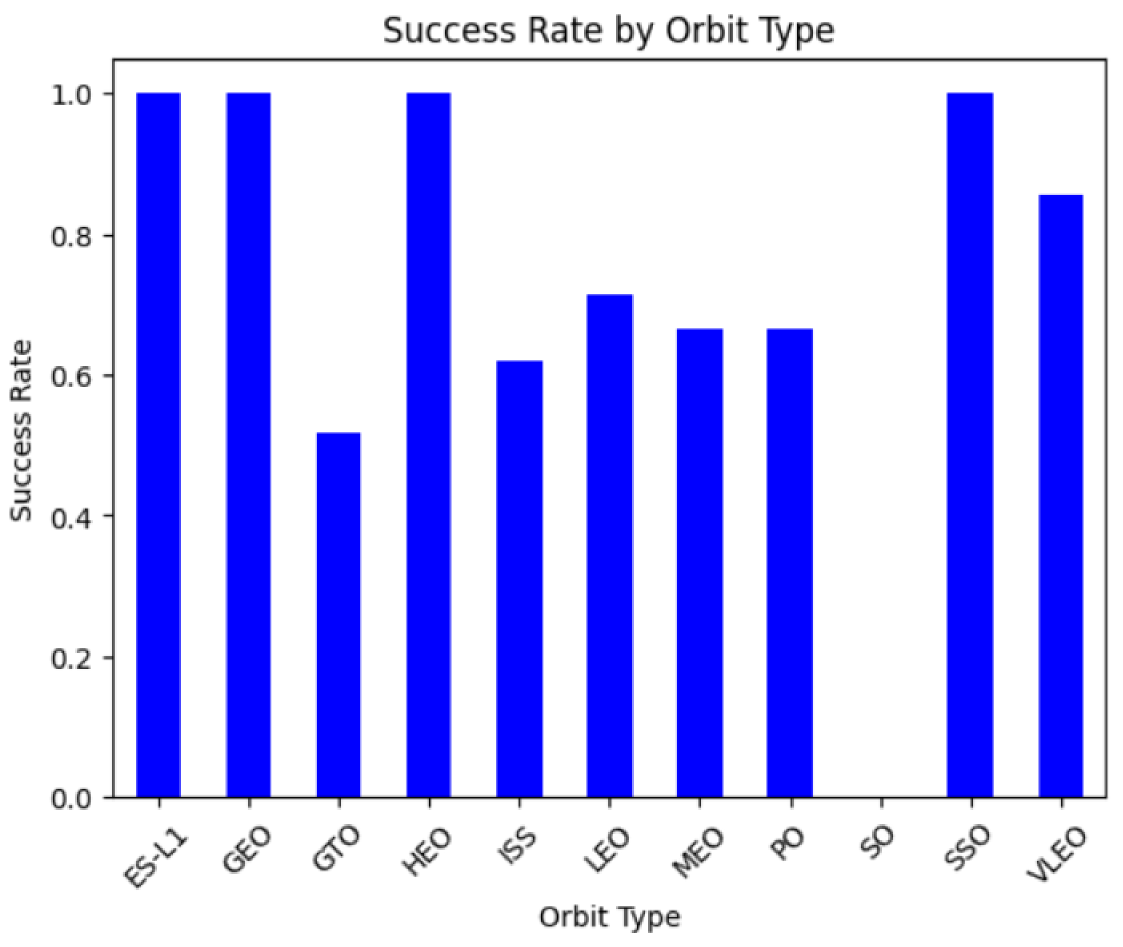
Payload vs. Launch Site

- VAFB-SLC-4E success rate increases with the flight number
- Other sites doesn't have any correlations to get derived any conclusion



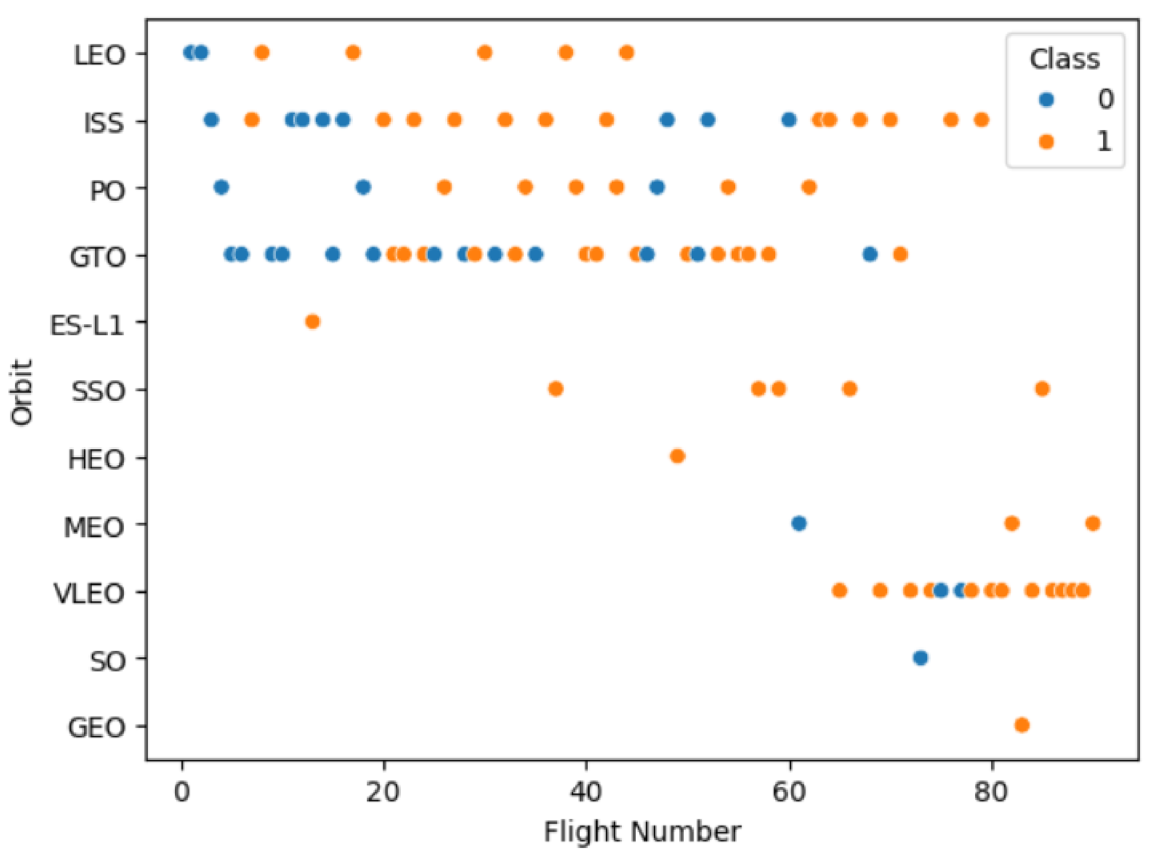
Success Rate vs. Orbit Type

- ES-L1, GEO, HEO and SSO orbits has 100% success rate
- Most of the orbit has more than 50% success rate



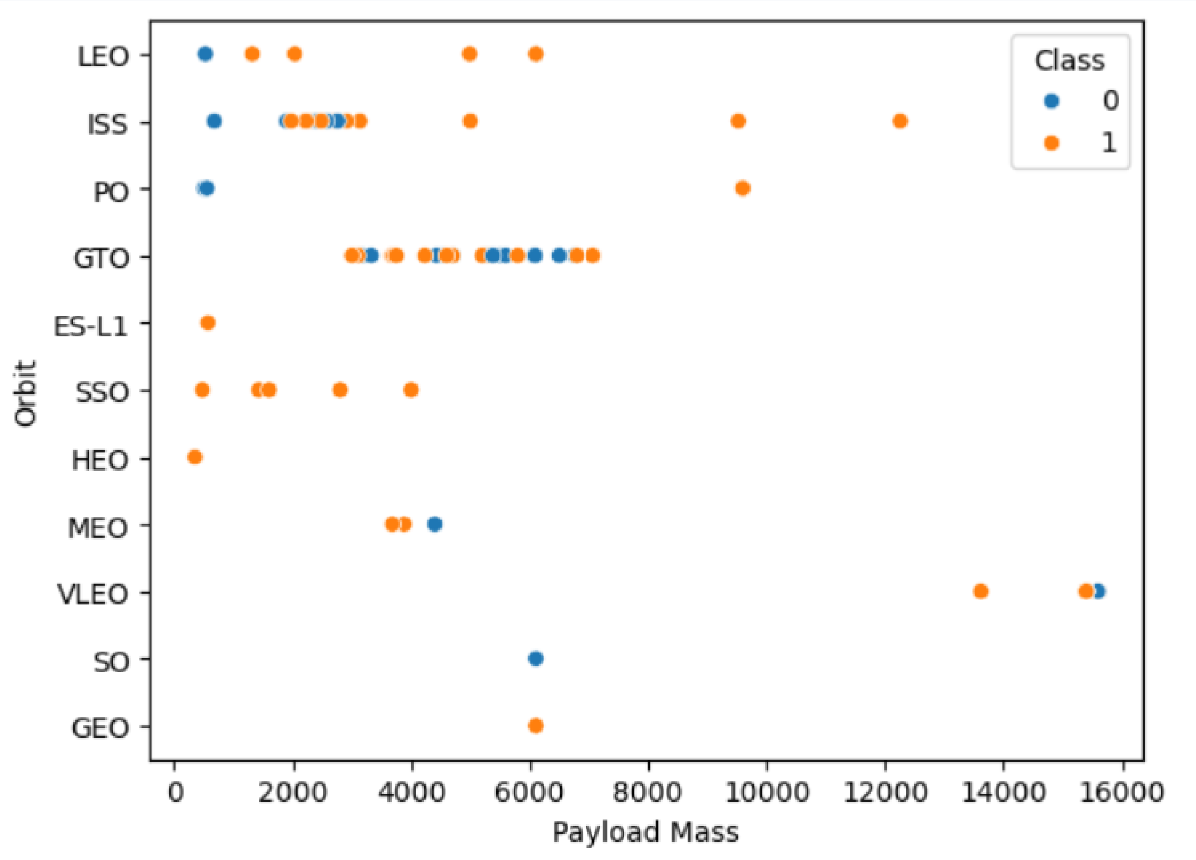
Flight Number vs. Orbit Type

- LEO and SSO orbits success rate increases with flight number
- Other orbits have no correlation between the flight number to derive any conclusion



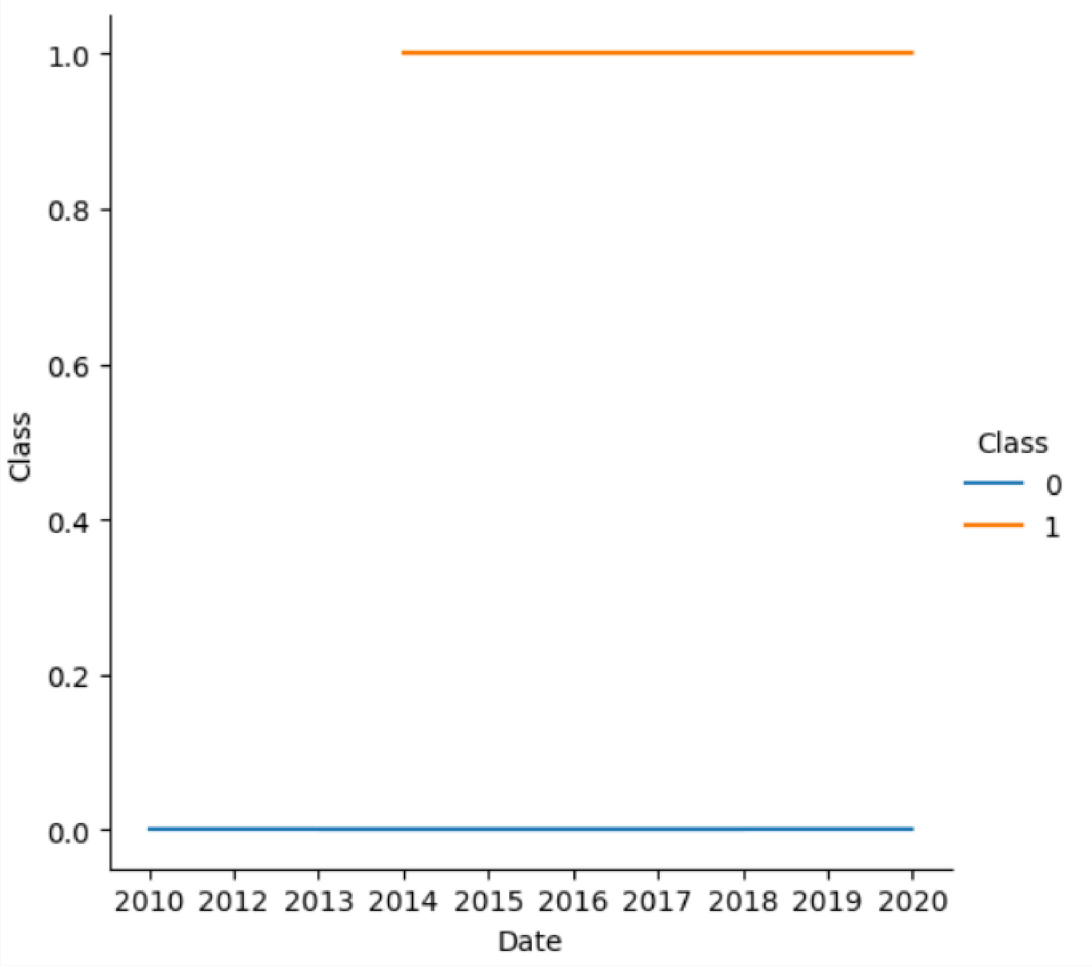
Payload vs. Orbit Type

- LEO and SSO orbits success rate increases with payload mass
- Other orbits have no correlation between the payload mass to derive any conclusion



Launch Success Yearly Trend

- the success rate since 2013 kept increasing till 2020



All Launch Site Names

- Names of the unique launch sites

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

- SQL query

```
%sql select DISTINCT("Launch_Site") from SPACEXTABLE;
```

- 4 Launch Sites are extracted from the datasets

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with `CCA`

Launch_Site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

- SQL Query

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

- 5 samples extracted from the dataset using the SQL query using the where and like clause

Total Payload Mass

- The total payload mass carried by boosters launched by NASA (CRS)

Total_payload_mass	Customer
45596	NASA (CRS)

- SQL query

```
%sql select sum(PAYLOAD_MASS__KG_) as Total_payload_mass, Customer from SPACEXTABLE where Customer="NASA (CRS)"
```

- Total payload mass carried by the booster rocket launched by NASA is 45596 kg calculated by SQL query using where clause

Average Payload Mass by F9 v1.1

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

AVG(PAYLOAD_MASS__KG_)
2534.6666666666665

- SQL query

```
%sql select AVG(PAYLOAD_MASS__KG_) from SPACEXTABLE where Booster_Version like "F9 v1.1%"
```

- The average payload mass carried by booster version F9 v1.1 is 2534.66 kg

First Successful Ground Landing Date

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

min(Date)
2010-06-04

- SQL query

```
%sql select min(Date) from SPACEXTABLE where Mission_Outcome="Success"
```

- First successful first stage landing of Falcon9 occurred on June 2010

Successful Drone Ship Landing with Payload between 4000 and 6000

- List of the names of boosters which have successfully landed on drone ship and had 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

- SQL Query

```
%%sql select Booster_Version from SPACEXTABLE
    where Landing_Outcome="Success (drone ship)"
    and PAYLOAD_MASS__KG_ between 4000 and 6000
```

- Total 4 booster rocket is used for the payload range between 4000 and 6000

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes

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

- SQL query

```
%sql select Mission_Outcome,count(Mission_Outcome) from SPACEXTABLE group by Mission_Outcome
```

- Total 100 missions are successful

Boosters Carried Maximum Payload

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

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

```
 * sqlite:///my_data1.db
Done.
```

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

Total 12 booster Rocket carried the maximum payload mass

2015 Launch Records

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

```
%%sql select substr(Date, 6,2) as Month, Landing_Outcome, Booster_Version, Launch_Site from SPACEXTABLE
    where substr(Date,0,5)='2015' and Landing_Outcome="Failure (drone ship)"
```

```
 * sqlite:///my_data1.db
Done.
```

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

- At the month of Jan and April two booster rocket failed to land by drone ship which is launch on the same loaction

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

```
%%sql select Landing_Outcome, COUNT(*) AS outcome_count from SPACEXTABLE
    where Date between "2010-06-04" and "2017-03-20" and
    (Landing_Outcome="Failure (drone ship)" or Landing_Outcome="Success (ground pad)")
    GROUP BY "Landing_Outcome"
    ORDER BY outcome_count DESC
```

```
 * sqlite:///my_data1.db
Done.
```

Landing_Outcome	outcome_count
Failure (drone ship)	5
Success (ground pad)	3

- Failure by drone ship is highest

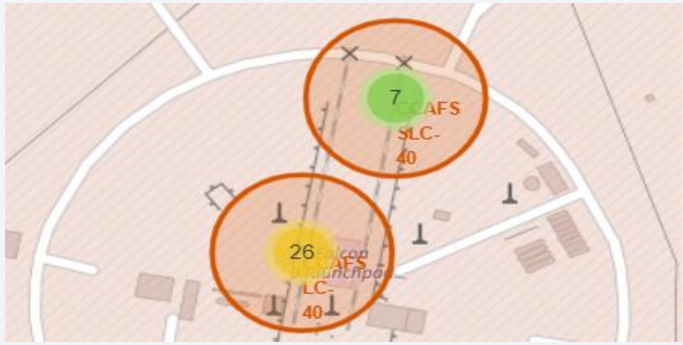
Section 3

Launch Sites Proximities Analysis

All launch sites on a map

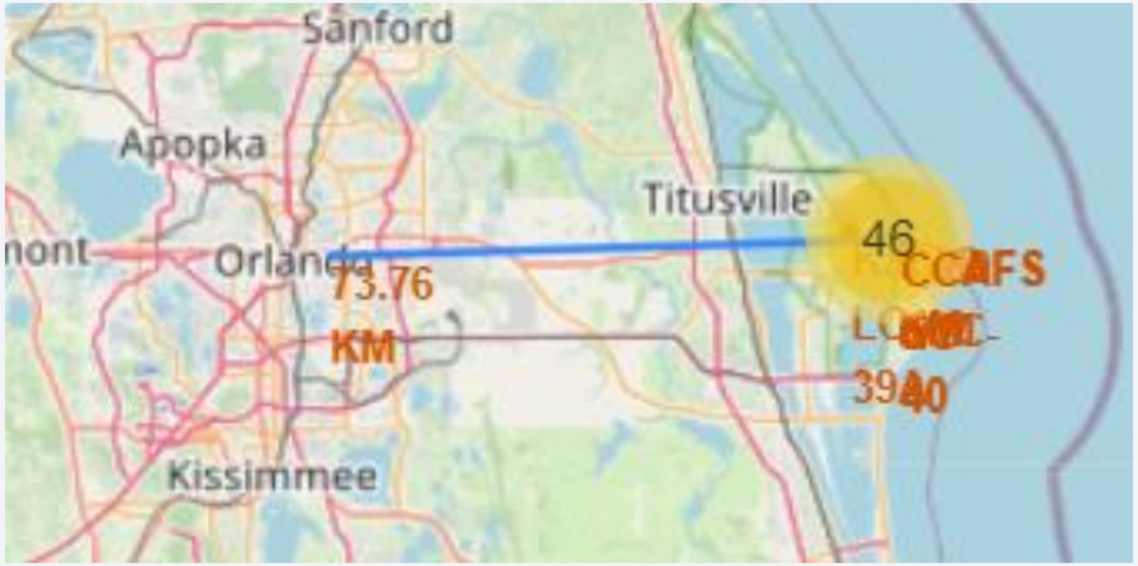
- All the launch sites are in proximity to the Equator line.
- All the launch sites are in very close proximity to the coastlines.

The success/failed launches for each site on the map



- CCAFS-SLC-40 and KSC-LC-39A launch site has higher success rate

The distances between a launch site to its proximities

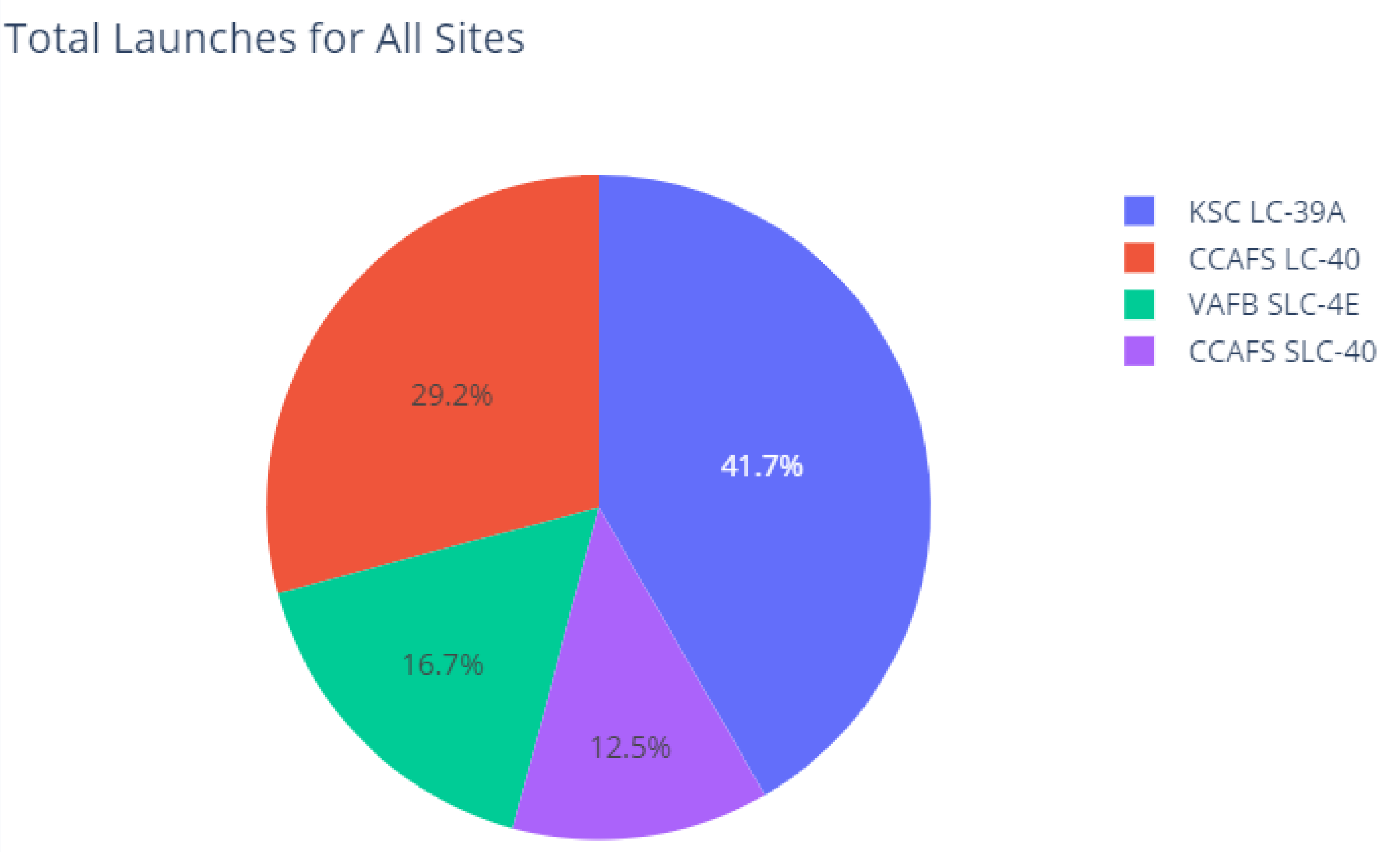


- launch sites are in close proximity to railways by 1.22 km from one of the launch site.
- launch sites are in close proximity to highways by 1.00 km from one of the launch site
- launch sites are in close proximity to coastline by 0.89 km from one of the launch site
- launch sites keep certain distance away from cities by 73.76 km(Orlando) from one of the launch site

Section 4

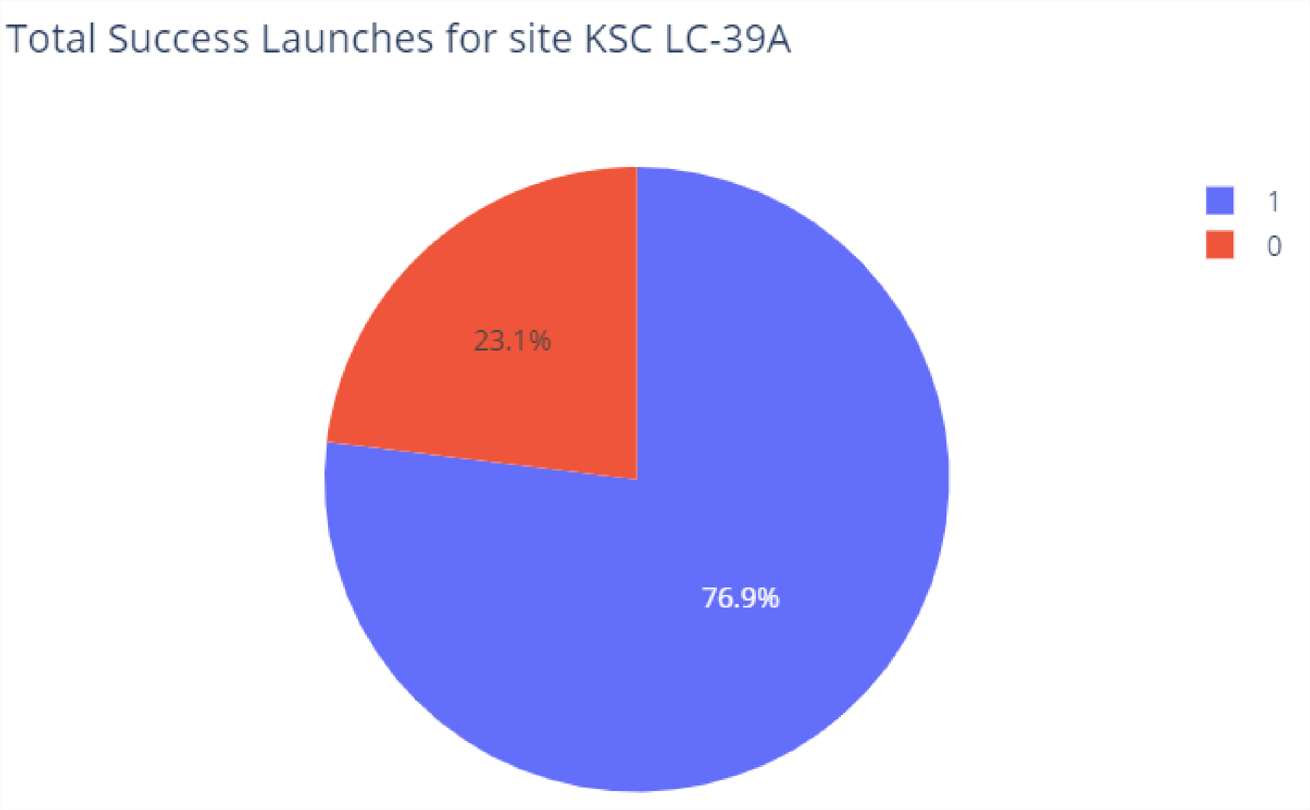
Build a Dashboard with Plotly Dash

Success rate of all the Launch sites



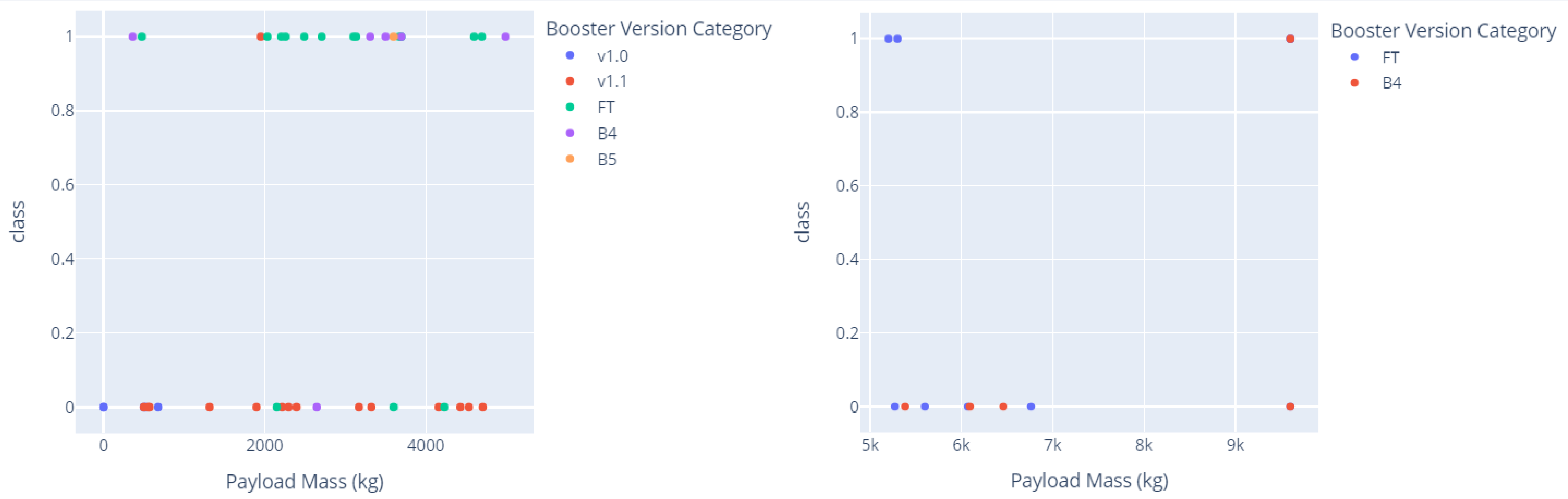
KSC LC-39A had the most successful launch from all the launch sites

KSC LC-39A Success/Failure percentage



KSC LC-39A launch site achieved a highest 76.9% success rate

Payload vs. Launch Outcome scatter plot for all sites

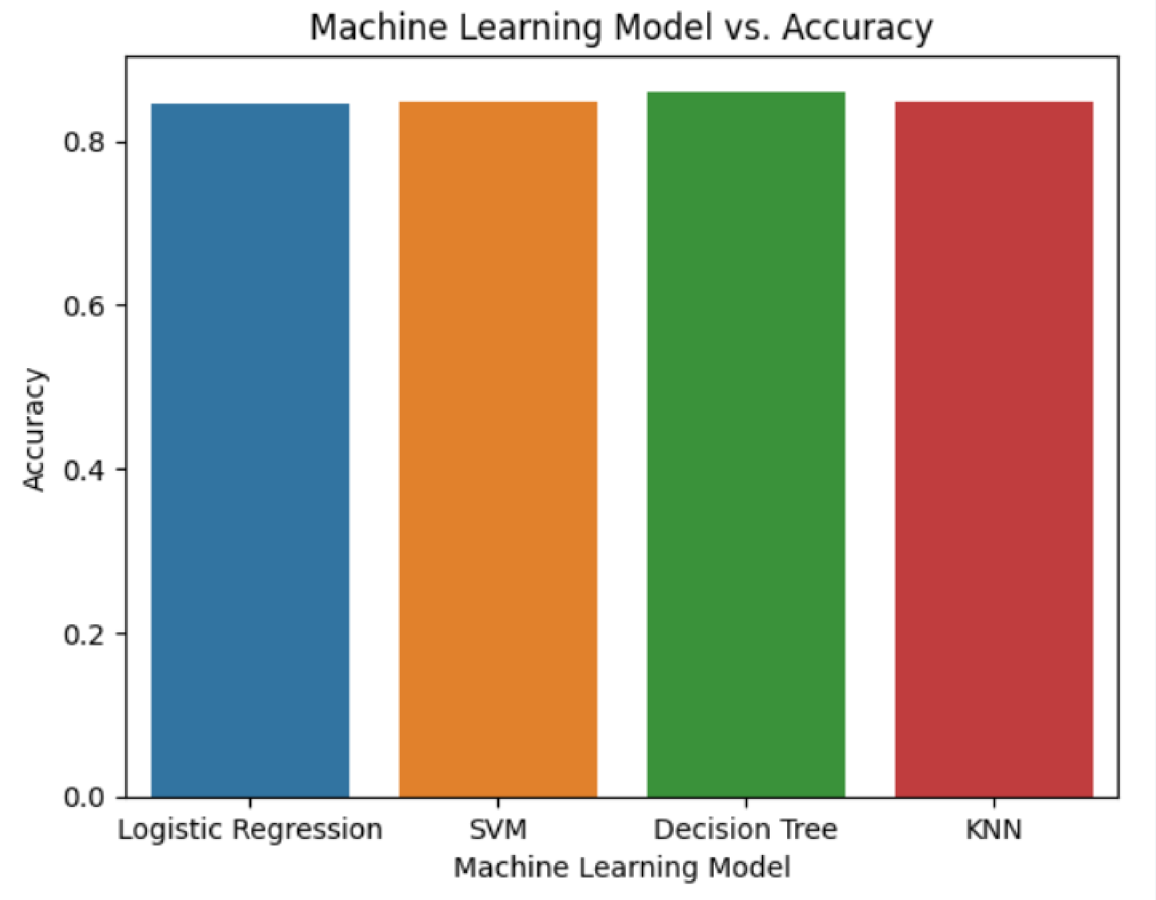


- Success rate of the FT booster is higher for low payload mass
- Failure rate of the B4,B5 and v1.1 are lower for low payload mass

Section 5

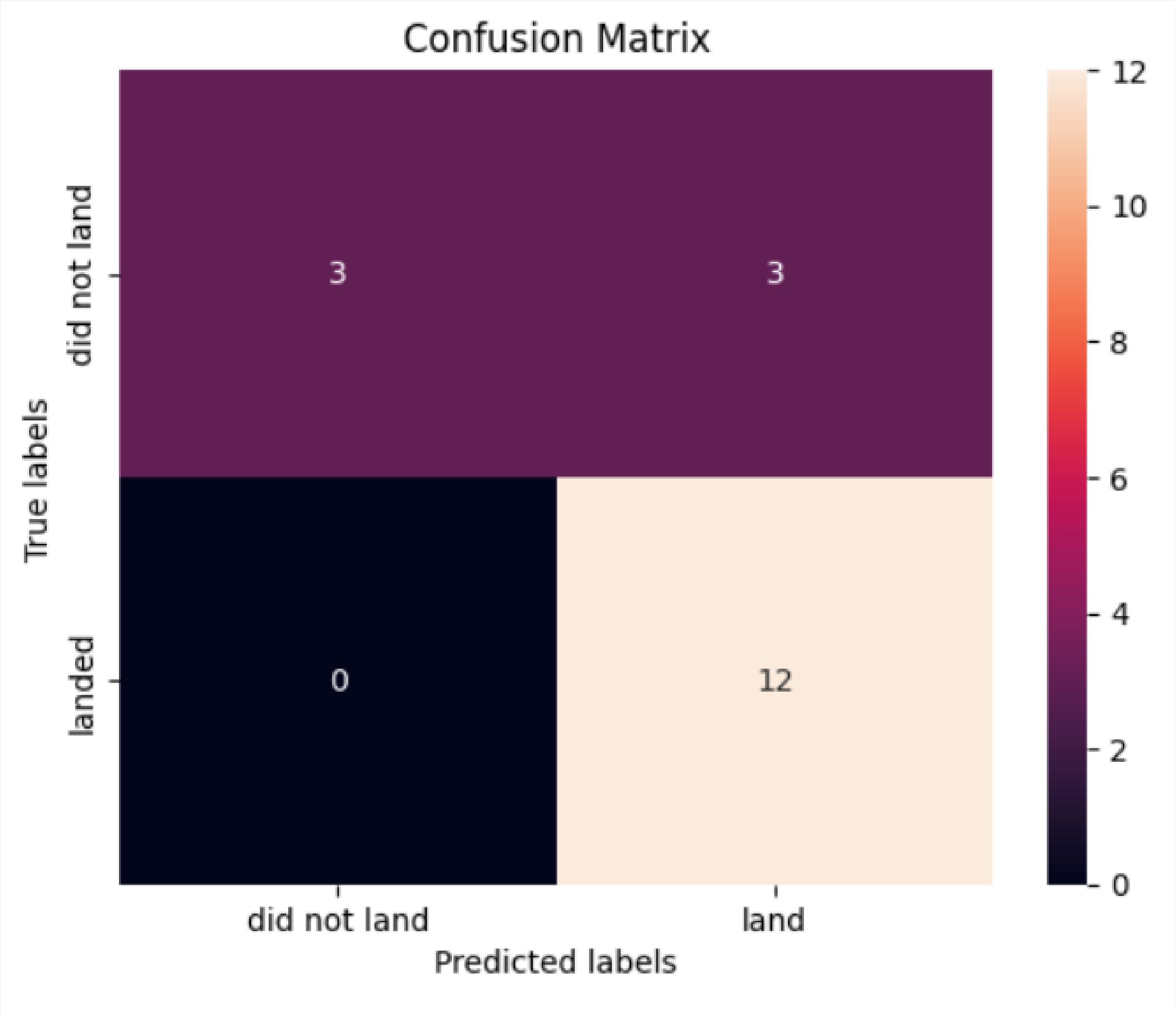
Predictive Analysis (Classification)

Classification Accuracy



Decision Tree has highest score of 86.07% among all the machine learning models

Confusion Matrix of the best performing model



Decision Tree confusion matrix predicted only 3 False positive outcome out of total 18 outcomes

Conclusions

We can conclude that:

- The larger the flight amount at a launch site, the greater the success rate at a launch site.
- Launch success rate started to increase in 2013 till 2020.
- Orbits ES-L1, GEO, HEO, SSO, VLEO had the most success rate.
- KSC LC-39A had the most successful launches of any sites.
- The Decision tree classifier is the best machine learning Model for prediction of success rate.

Appendix

- https://github.com/rmangore/Capstone.git

Thank you!