IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Santiago López Pavón
18/12/2022

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection with Rest Api and Web Scraping
  - Data Wrangling using Numpy and Pandas
  - Exploratory Data Analysis with Pandas and SQL
  - Visual Analytics and Dashboard with Folium and Dash
  - Predictive Analysis with multiple Machine Learning algorithms
- Summary of all results
  - Exploratory Data Analysis result
  - Predictive Analytics result with Machine Learning

Introduction

- Project background and context
  - We live in the space age, there are several companies in the market but the most successful is SpaceX.
  - One of the success factors that SpaceX has had is that rocket launches are cheaper than the competition. Being thanks to Falcon 9 whose launch is 3 times cheaper than
- Problems you want to find answers
  - What factors determine the successful landing of the rocket?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - The data was collected from the Rest Api and use Web Scraping techniques, using Python techniques
- Perform data wrangling
  - Feature engineering techniques are used to normalize the data and improve it so that it can be better processed by machine learning algorithms.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash

Methodology

Executive Summary

- Perform predictive analysis using classification models
  - Several classification algorithms are used, such as logistic regression, support vector machines, decision trees, and the nearest neighbor algorithm.
  - To evaluate the models, the confusion and scoring matrices that measure the precision of the model are used.

Data Collection

- Describe how data sets were collected.
  - Data sets are collected from SpaceX API this is the URL (https://api.spacexdata.com/v4/rockets/)
  - Another data is geted from Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches), using web scraping in Python for get information from tables.

Data Collection – SpaceX API

- For get the information and clean data. With request lib we get the information of Rest Api and with json_normalize method we convert the Json to DataFrame.

- For More Github: https://github.com/santilopezpavon/ibm-network-jupyter/blob/master/Data%20Collection%20API.ipynb

```python
spacex_url="https://api.spacexdata.com/v4/launches/past"
```

```python
response = requests.get(spacex_url)
```

```python
# Use json_normalize meethod to convert the json result into a dataframe
data = pd.json_normalize(response.json())
```

```python
# Lets take a subset of our dataframe keeping only the features we want and the flight number, and date_utc.
data = data[['rocket', 'payloads', 'launchpad', 'cores', 'flight_number', 'date_utc']]

# We will remove rows with multiple cores because those are falcon rockets with 2 extra rocket boosters and rows that ha
data = data[data['cores'].map(len)==1]
data = data[data['payloads'].map(len)==1]

# Since payloads and cores are lists of size 1 we will also extract the single value in the list and replace the feature
data['cores'] = data['cores'].map(lambda x : x[0])
data['payloads'] = data['payloads'].map(lambda x : x[0])

# We also want to convert the date_utc to a datetime datatype and then extracting the date leaving the time
data['date'] = pd.to_datetime(data['date_utc']).dt.date

# Using the date we will restrict the dates of the launches
data = data[data['date'] <= datetime.date(2020, 11, 13)]
```

Data Collection - Scraping

- From the response text of Request we get an BeautifulSoup object.
- With BeautifulSoup object we can access to HTML data and get DOM information.
- For More: https://github.com/santilopezpavon/ibm-network-jupyter/blob/master/Data%20Collection%20with%20Web%20Scraping.ipynb

```python
# use requests.get() method with the provided static_url
# assign the response to a object
response = requests.get(static_url)
```

Create a `BeautifulSoup` object from the HTML `response`

```python
# Use BeautifulSoup() to create a BeautifulSoup object from a response text content
soup = BeautifulSoup(response.text, 'html')
```

Data Wrangling

- The data is processed in these steps:
    1. Calculate the number of launches on each site
    2. Calculate the number and occurrence of each orbit
    3. Calculate the number and occurence of mission outcome per orbit type
    4. Create a landing outcome label from Outcome column
    5. And the last step was generate a new csv.
- More information in Github https://github.com/santilopezpavon/ibm-network-jupyter/blob/master/Data%20Wrangling.ipynb

EDA with Data Visualization

- For Data Visualization we use charts
  - Scatter plot with relation of Flight Number and Launch Site
  - Scatter plot for see the relationship between Payload and Launch Site
  - Bar char for see the relationship between success rate of each orbit type
  - Scatter plot for see the relationship between FlightNumber and Orbit type
  - Scatter plot for see the relationship between Payload and Orbit type
  - Line char for see the yearly trend, relating the Class and years
- https://github.com/santilopezpavon/ibm-network-jupyter/blob/master/EDA%20with%20Visualization.ipynb

EDA with SQL

- The queries are:
  - The names of the unique launch sites in the space mission
  - 5 records where launch sites begin with the string 'CCA'
  - the total payload mass carried by boosters launched by NASA (CRS)
  - average payload mass carried by booster version F9 v1.1
  - the date when the first successful landing outcome in ground pad was acheived.
  - the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - the total number of successful and failure mission outcomes
  - the names of the booster_versions which have carried the maximum payload mass. Use a subquery
  - the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
  - 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

For more information: https://github.com/santilopezpavon/ibm-network-jupyter/blob/master/EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
  - Markers indicate points like launch sites
  - Circles indicates highlighted areas
  - Marker clusters for group events
  - Lines for indicates distances
- For more information: https://github.com/santilopezpavon/ibm-network-jupyter/blob/master/Visual%20Analytics%20with%20Folium.ipynb

Build a Dashboard with Plotly Dash

- In Plotly Dash there are two plots:
  - Pie graph for percentage of launches by site
  - Interactive line filter for payload range
- For More information https://github.com/santilopezpavon/ibm-network-jupyter/blob/master/spacex_dash_app.py

Predictive Analysis (Classification)

- We test four classification model: logistic regresion, support machine vector, k nearest neighbords and decision tree.
  - Get data, clean, and standardization data
  - Split data in train data and test data
  - Test and train each model for get the best hyperparameters
  - Compare results for each model.
- For more information: https://github.com/santilopezpavon/ibm-network-jupyter/blob/master/Machine%20Learning%20Prediction.ipynb

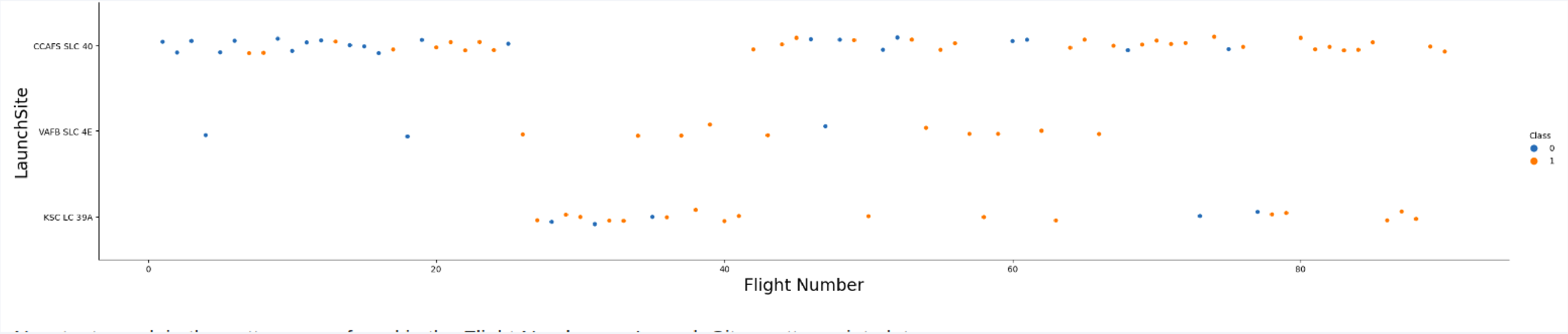
Results

- From of 2014, the trend of improvement in success stories begins.
- Launch location can improve success rate
- The orbit factor can improve to success rate
- A small payload can increase the success stories

Section 2

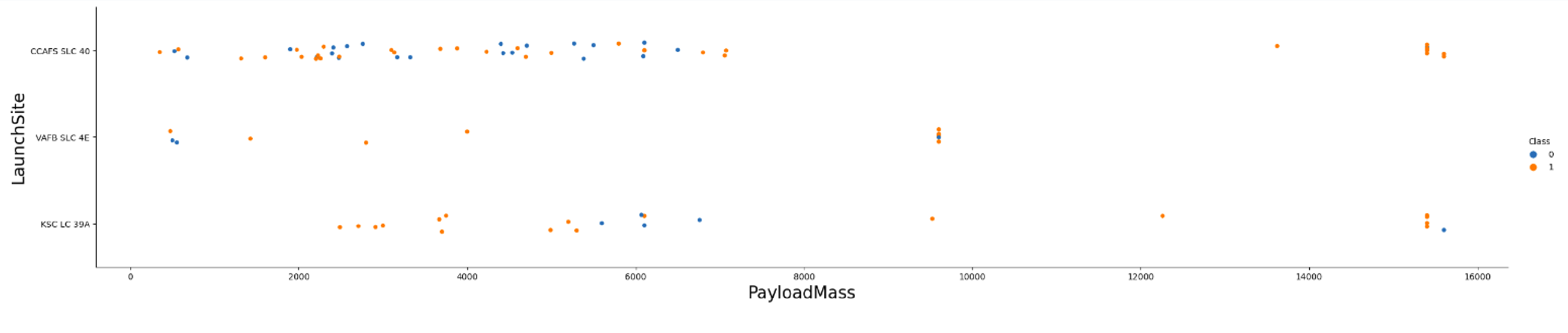
Insights drawn from EDA

Flight Number vs. Launch Site



- It seems that the place of launch can affect the result. The most successful launches are those of CCAF5 SLC 40
- The higher the number of flights at any launch site, the better the success rate.
- It may be that as more flights are made the technique improves and thus the success.

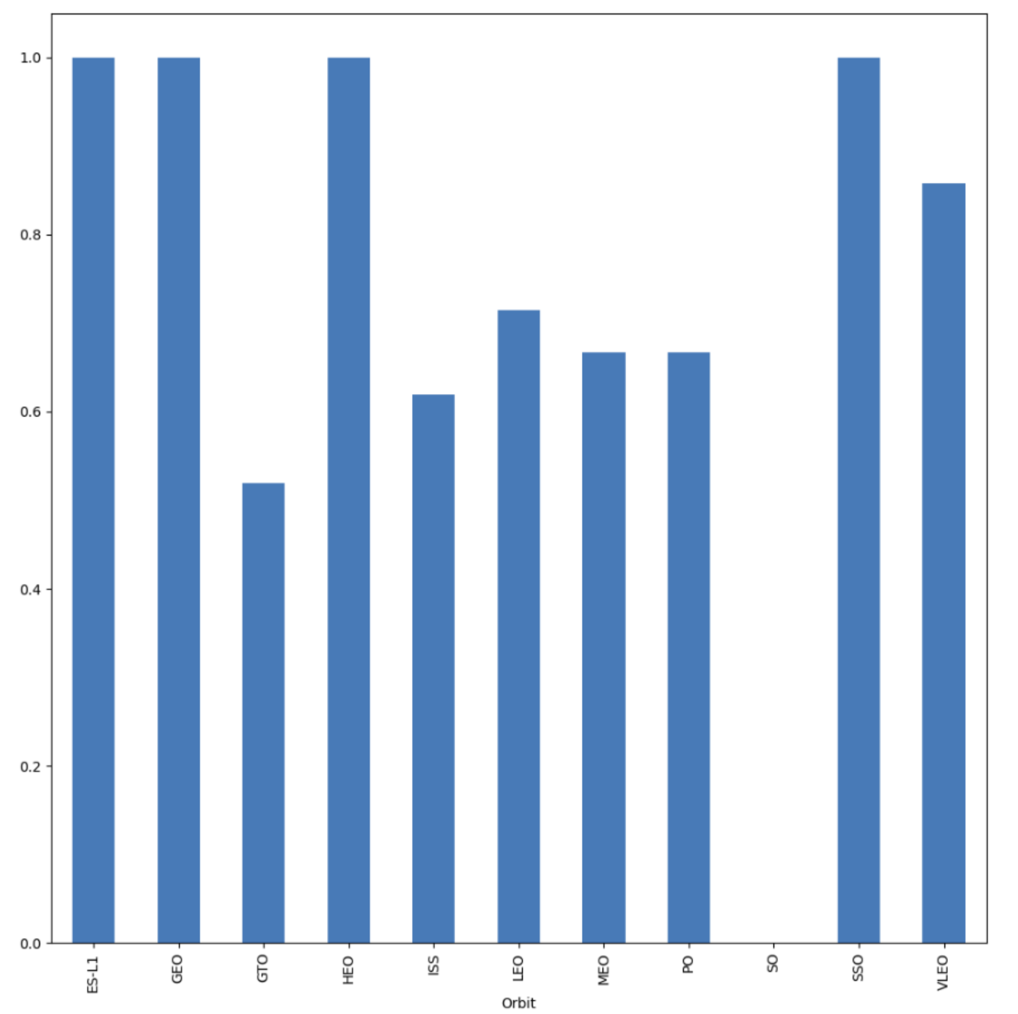
Payload vs. Launch Site



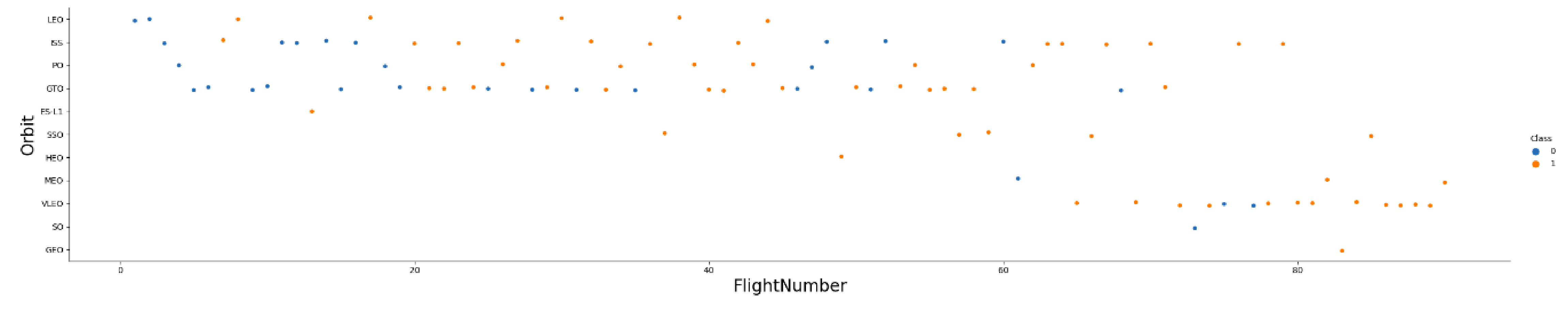
- It seems that a lower payload contributes to more success, especially if it is less than 7,000
- A higher payload at the CAF5 SLC 40 location does not appear to reduce success stories, although there is little data to assess this.

Success Rate vs. Orbit Type

- The type of orbit can affect success stories, with 4 types of orbit standing out.
    - ES-L1
    - GEO
    - HEO
    - SSO

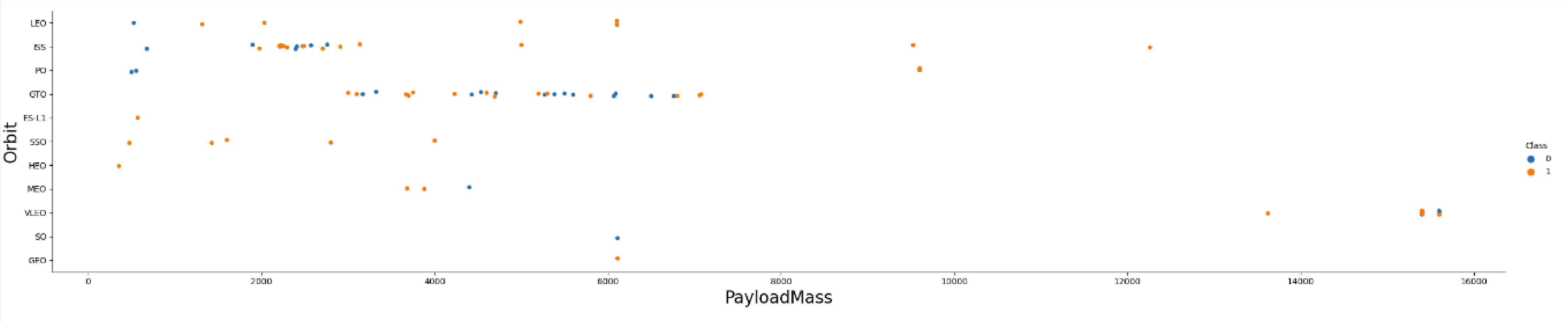


Flight Number vs. Orbit Type



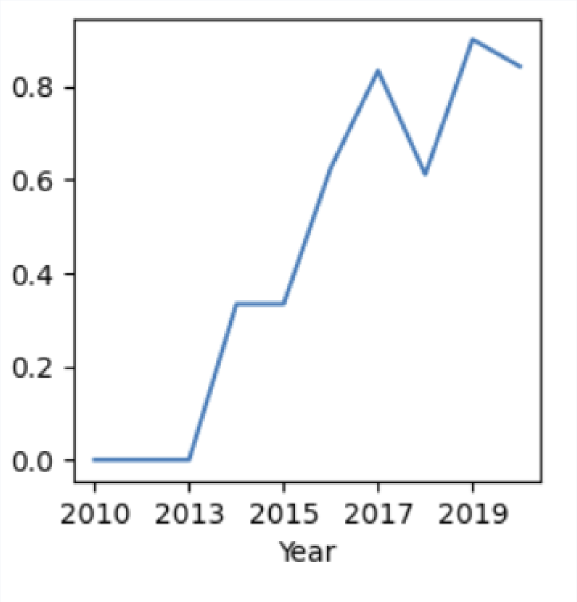
- As we have previously verified: It depends on the type of orbit we have a higher success rate, and as flights increase too.

Payload vs. Orbit Type



- Heavy payloads, have best results for PO, LEO and ISS orbits.

Launch Success Yearly Trend



- From 2014 success stories increase significantly

All Launch Site Names

- The launch sites are:

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

Launch Site Names Begin with 'CCA'

- These are the 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	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

- Calculate the total payload carried by boosters from NASA

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

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is 2928

avg_payload
2928

First Successful Ground Landing Date

- The first successful landing outcome on ground pad is 2015-12-22

first_success_gp
2015-12-22

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

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes

mission_outcome	qty
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

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

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

2015 Launch Records

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

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

- 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

landing__outcome	qty
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

- The Space X launch sites are in United States of America, in Florida and California

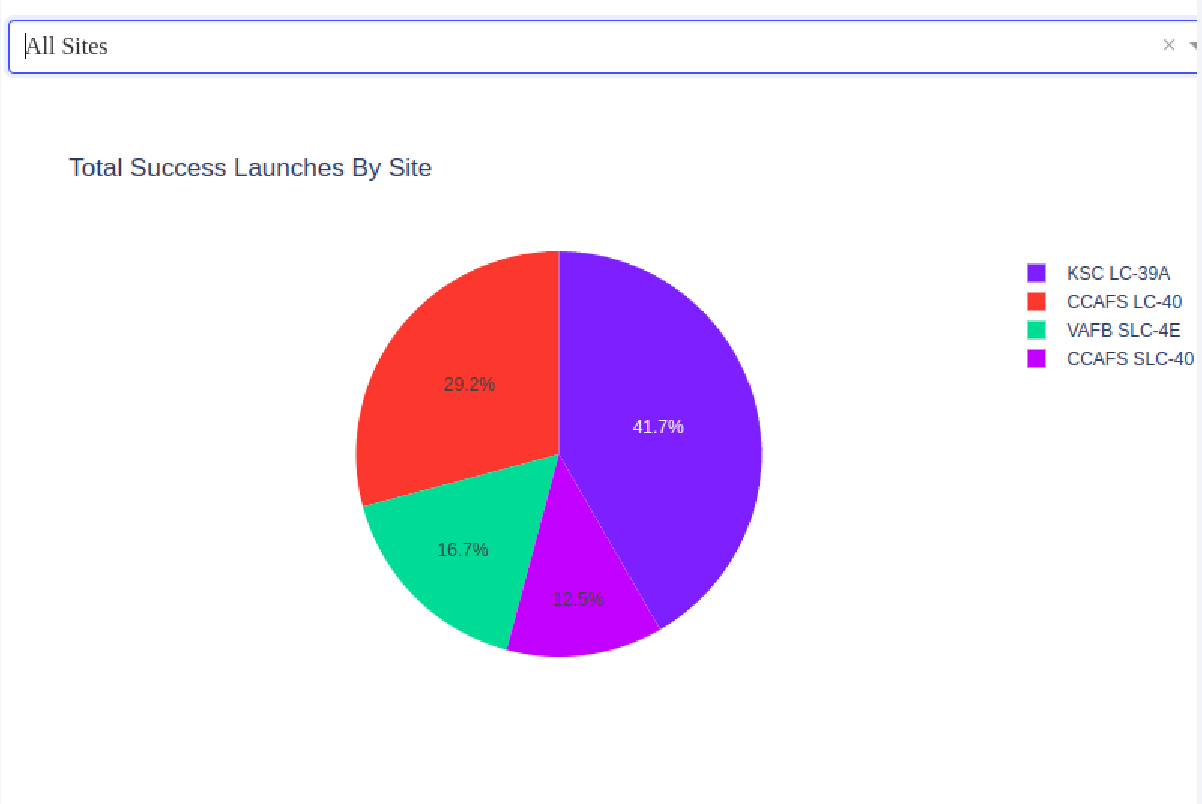
Launch Sites in the coast

- The Green Markers are for successful launches, and Red Markers are for failures launches

Section 4

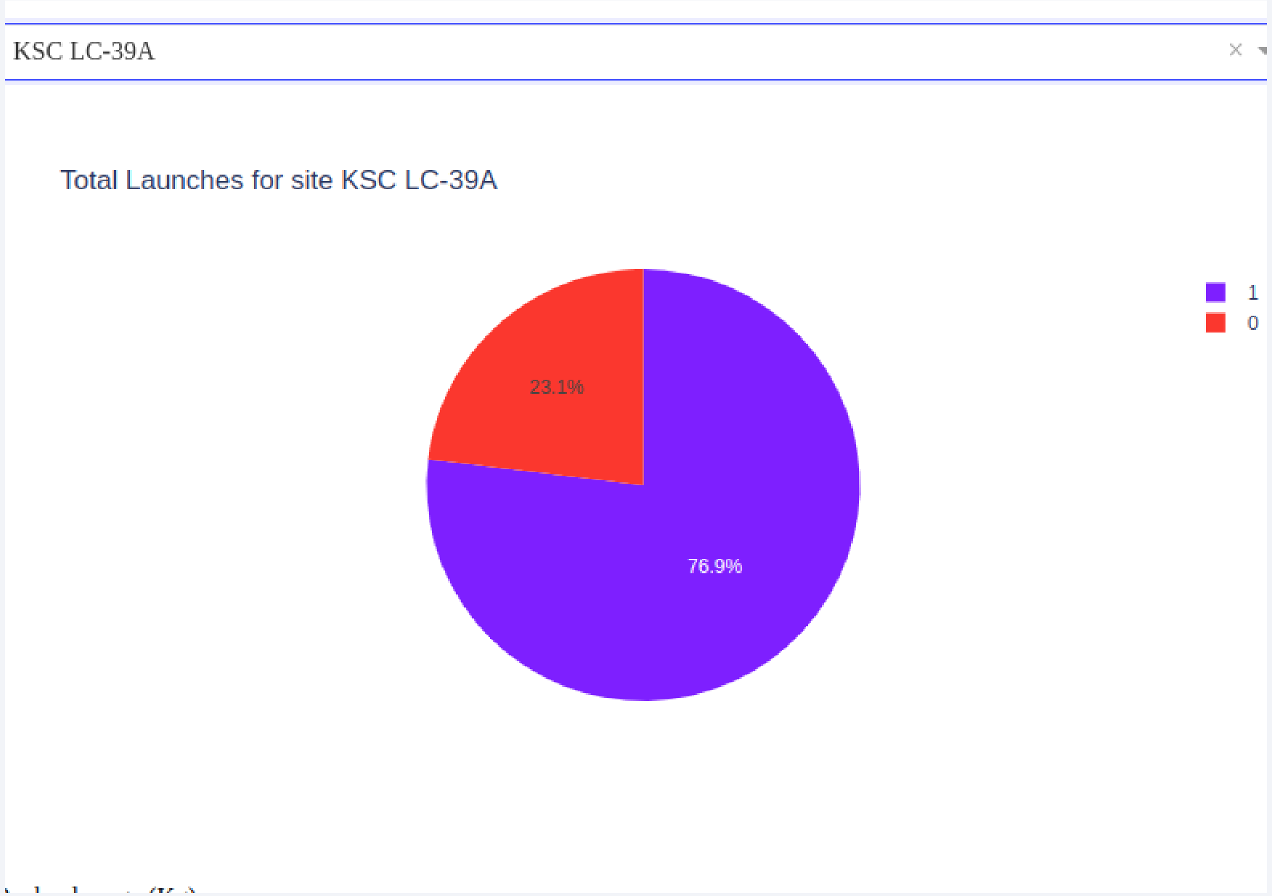
Build a Dashboard with Plotly Dash

Launch Success all sites



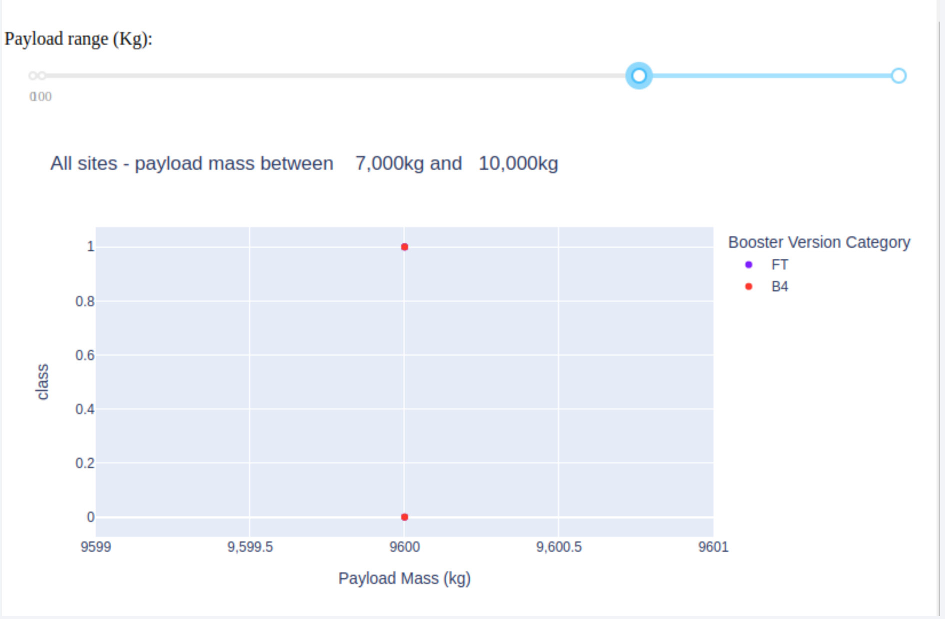
- KSC LC 39A has more success cases for All sites.

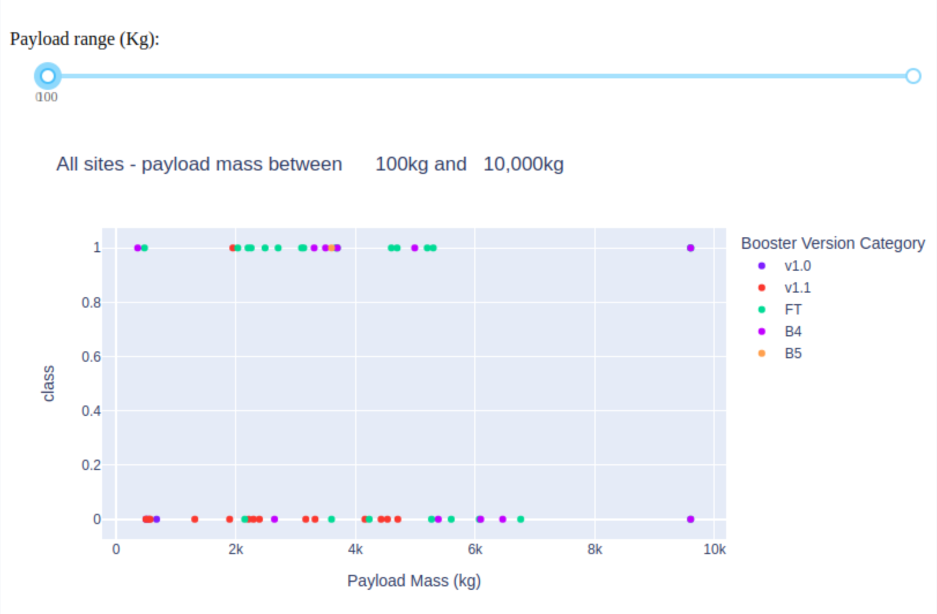
Highest launch success site



- KSC LC-39 A has more success case rate, 76.9% of success case.

Payload vs. Launch Outcome scatter





- For payload range from 7.000 to 10.000 there are low success cases.

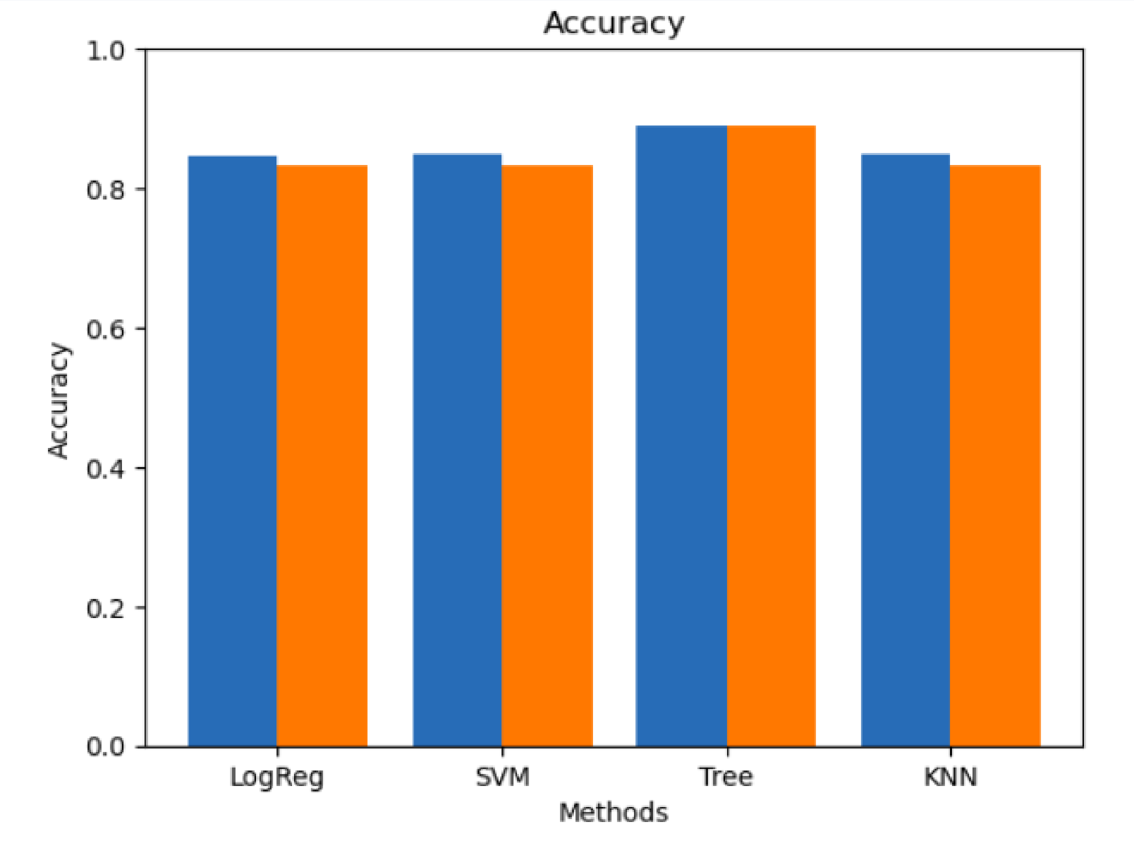
- For payload minors, can increase success stories.

Section 5

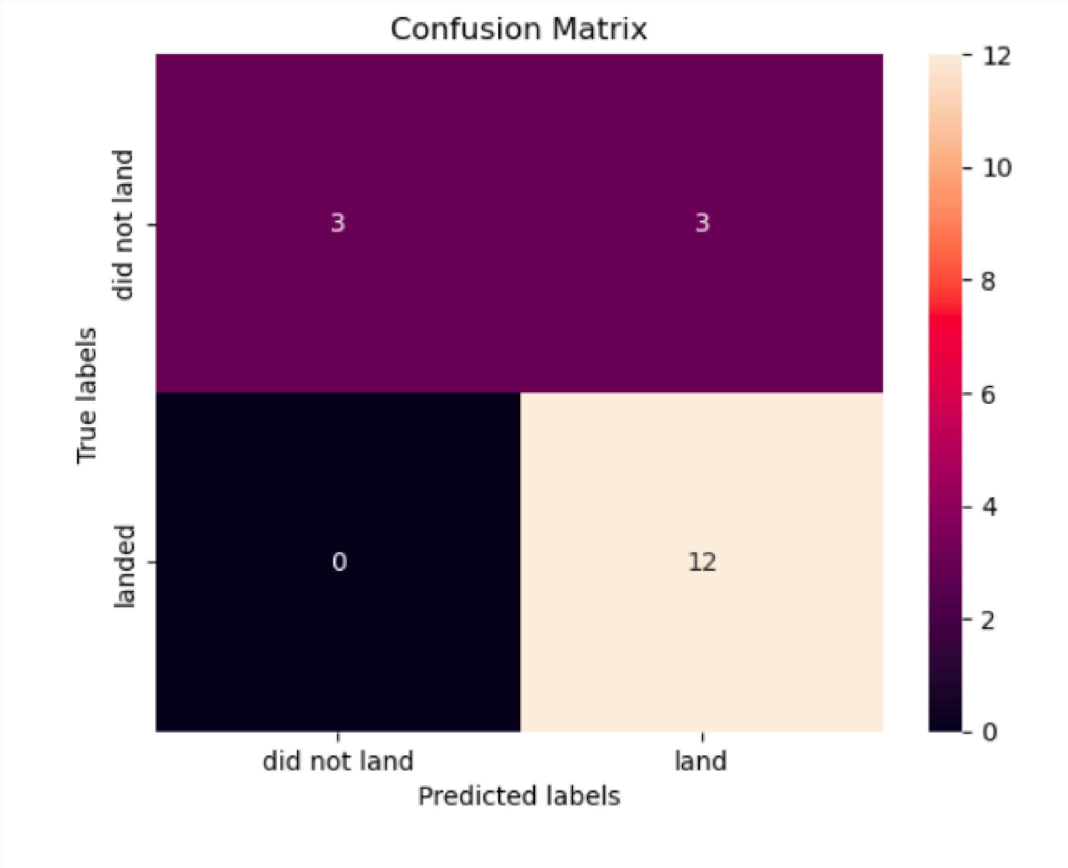
Predictive Analysis (Classification)

Classification Accuracy

- The model with the best results is Tree decision classifier.



Confusion Matrix



- As we can view for this Confusion Matrix of Tree decision algorithm, the major problem is the false positives

Conclusions

- The larger flight number and major date, increase the success case rate. It is surely due to the experience acquired and the improved technique.

- There are orbits with best success rate: VLEO, SSO, GEO, ES-L1 and HEO

- The launch site KSC LC-39A had most successful launches.

- The Decision Tree algorithm is the best machine learning algorithm for this question.

Thank you!