IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

- Summary of methodologies:
  - Data Collection & Data wrangling
  - EDA with data visualization & SQL
  - Building an interactive map with Folium
  - Building a dashboard with Plotly
  - Predictive analysis (Classification)
- Summary of all results
  - Exploratory data analysis results
  - Interactive and predictive analytics results

Introduction

- In this capstone, we will predict if the Falcon 9 first stage will land successfully. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. Therefore if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch. In this lab, you will collect and make sure the data is in the correct format from an API. The following is an example of a successful and launch.Problems you want to find answers

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - SpaceX API and web scraping Wikipedia
- Perform data wrangling
  - One hot encoding
- 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 – SpaceX API

- Get data with get request from SpaceX API, then normalize, clean and wrangle data
- GitHub URL: https://github.com/carical906/data-science/blob/8296521381edffad58080f2ed0b20b283196ff66/Data%20Collection%20API.ipynb

1. Get request for rocket launch data from SpaceX API
2. Normalize data and convert to dataframe
3. Perform data cleaning and data wrangling to fill in missing values and only include Falcon 9 launches in dataframe

Data Collection - Scraping

- Extracted Falcon 9 launch records from Wikipedia and converted it to pandas dataframe

- GitHub URL: https://github.com/carical906/data-science/blob/39c380b7e1a4c003ae23b564711f7d8634374222/Data%20Collection%20Web%20Scraping.ipynb

1. HTTP GET method to request the Falcon9 Launch HTML page, as an HTTP response
2. Create a BeautifulSoup object from the HTML response
3. collect all relevant column names from the HTML table header
4. collect all relevant column names from the HTML table header

Data Wrangling

- Calculated the number of launches on each site
- Calculated the number and occurrence of each orbit
- Calculated the number and occurence of mission outcome per orbit type
- Created a landing outcome label from Outcome column

- GitHub URL: https://github.com/carical906/data-science/blob/f16007a341b71ad4c6fe244f0537a5463f2b73ae/Data%20Wrangling.ipynb

EDA with Data Visualization

- Scatter point chart FlightNumber vs. PayloadMass to see how Payload affects the launch outcome
- Scatter point chart Launch Site vs. Payload to observe if there is any relationship between launch sites and their payload mass
- Bar chart to visualize the relationship between success rate of each orbit type
- Scatter point chart to visualize the relationship between FlightNumber and Orbit type
- GitHub URL: https://github.com/carical906/data-science/blob/226cfa8c8f5d39165657bd054557342ef3632836/jupyter-labs-eda-dataviz%20(1).ipynb

EDA with SQL

- Display the names of the unique launch sites in the space mission
- Display various other records
- List the total number of successful and failure mission outcomes
- 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
- GitHub URL: https://github.com/carical906/data-science/blob/afe6a8e08da2b756ff7e8ef24616a0d3d65e4d3c/EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

- Marked all launch sites, marked the success/failed launches for each site on the map
- Added a MousePosition on the map to get coordinate for a mouse over a point on the map. As such, while you are exploring the map, you can easily find the coordinates of any points of interests (such as railway)
- Explain why you added those objects
- GitHub URL: https://github.com/carical906/data-science/blob/2f31ddf141cf53ac3323efbe2ad0738286f99e93/lab_jupyter_launch_site_location%20(2).ipynb

Build a Dashboard with Plotly Dash

- Built interactive dashboard with Plotly dash
- Plotted pie charts showing the total launches by launch site
- Plotted scatter point graph showing correlation between payload and success for all sites

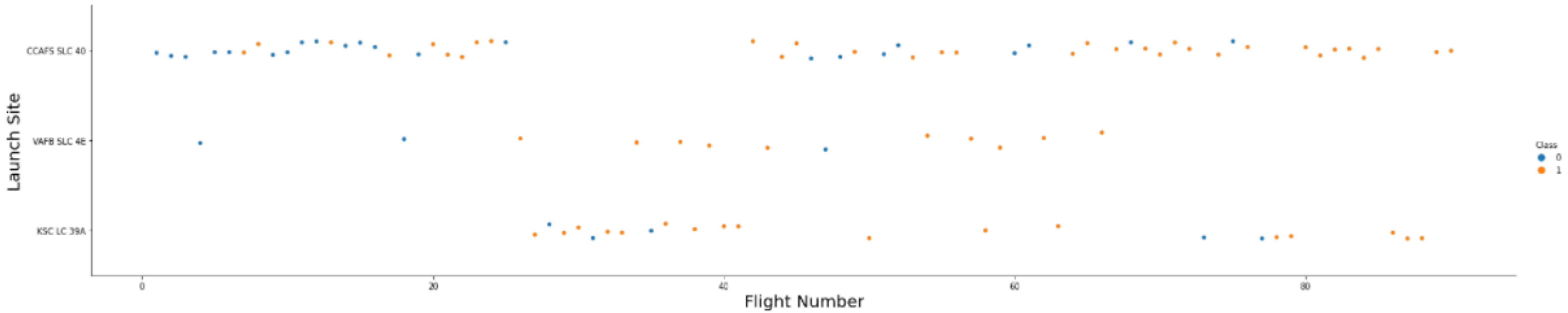
Predictive Analysis (Classification)

- Created a column for the class
- Standardized the data
- Split into training data and test data
- Found best hyperparameter for SVM, Classification Trees and Logistic Regression
- Found best performing method using test data
- GitHub URL: https://github.com/carical906/data-science/blob/425ad0a2c2a690642ecbcd9ff782b4ae7115fbf4/SpaceX_Machine%20Learning%20Prediction_Part_5%20(2).ipynb

Section 2

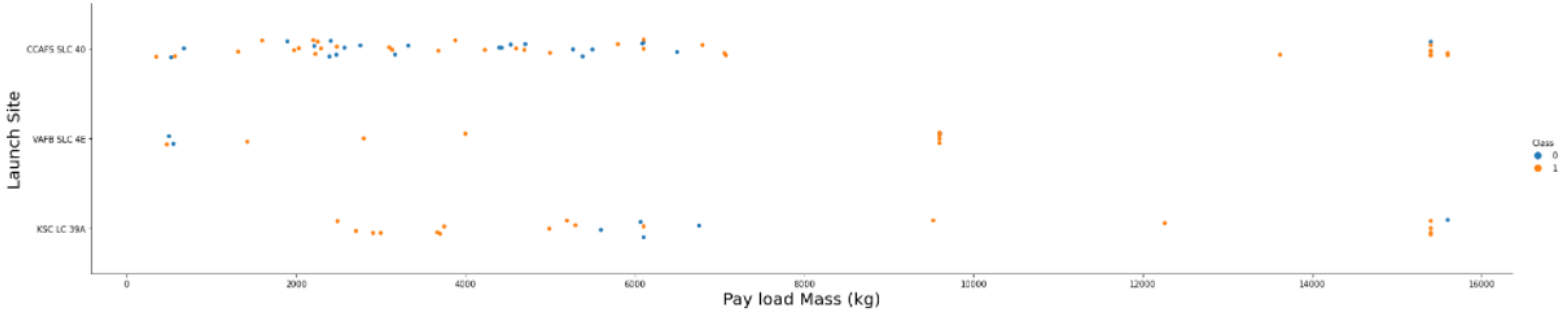
Insights drawn from EDA

Flight Number vs. Launch Site



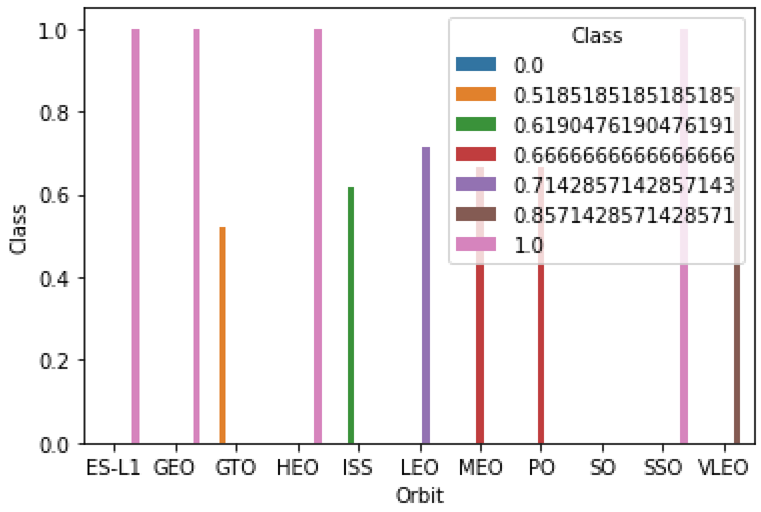

```
[ ]: More flights, more success
```

Payload vs. Launch Site



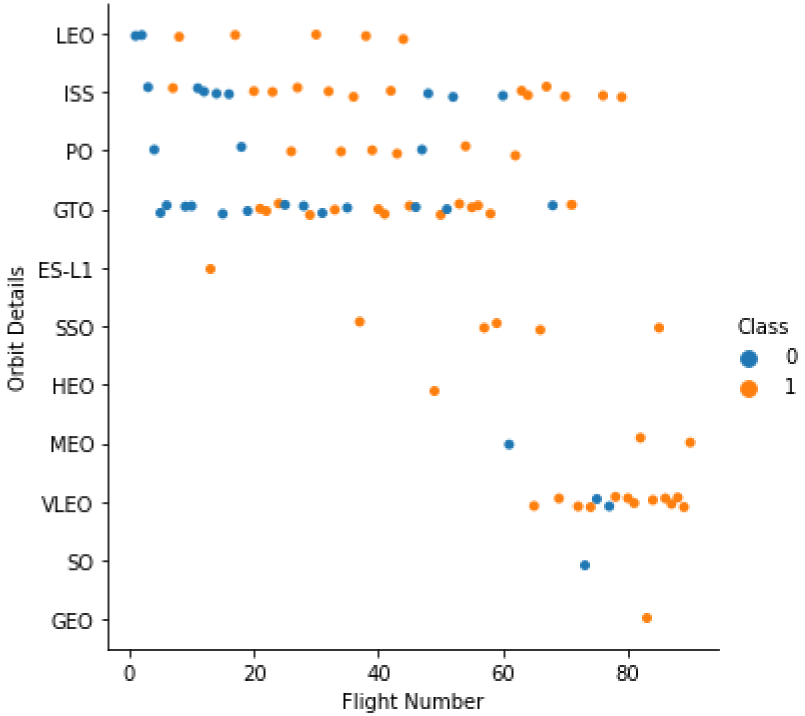
Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

Success Rate vs. Orbit Type



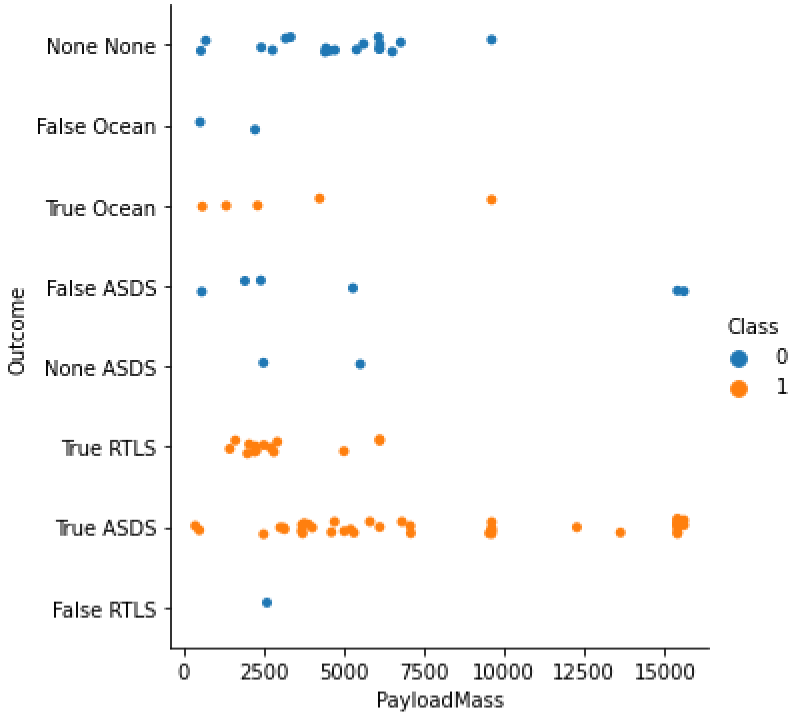

```
[ ]: ES-L1, GEO, HEO, SSO have best success rate
```

Flight Number vs. Orbit Type



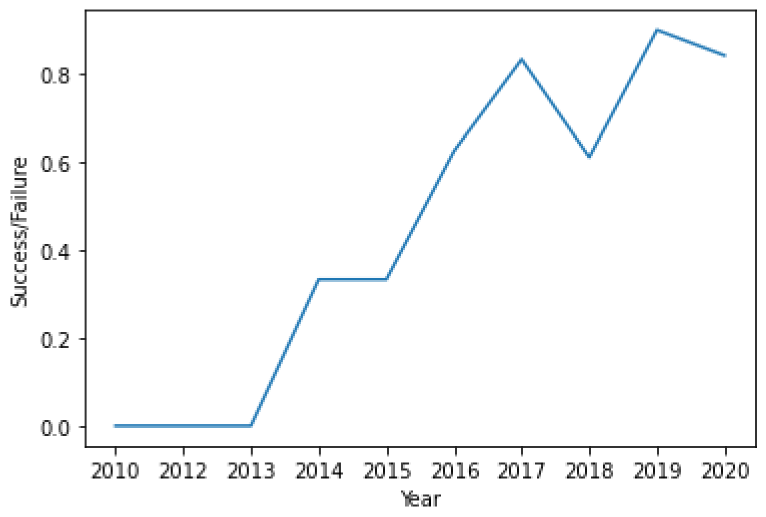
You should see that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type



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

Launch Success Yearly Trend



you can observe that the sucess rate since 2013 kept increasing till 2020

All Launch Site Names

- Select Distinct key word is necessary here

```
In [10]: task_1 = '''
                 SELECT DISTINCT LaunchSite
                 FROM SpaceX
         '''
         create_pandas_df(task_1, database=conn)
```

Out[10]:

	launchsite
0	KSC LC-39A
1	CCAFS LC-40
2	CCAFS SLC-40
3	VAFB SLC-4E

Launch Site Names Begin with 'CCA'

Display 5 records where launch sites begin with the string 'CCA'

In [11]:
```
task_2 = '''
        SELECT *
        FROM SpaceX
        WHERE LaunchSite LIKE 'CCA%'
        LIMIT 5
        '''
create_pandas_df(task_2, database=conn)
```

Out[11]:

	date	time	boosterversion	launchsite	payload	payloadmasskg	orbit	customer	missionoutcome	landingoutcome
0	2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
1	2010-08-12	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of...	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2	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
3	2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
4	2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

Display the total payload mass carried by boosters launched by NASA (CRS)

In [12]:
```
task_3 = '''
        SELECT SUM(PayloadMassKG) AS Total_PayloadMass
        FROM SpaceX
        WHERE Customer LIKE 'NASA (CRS)'
        '''
create_pandas_df(task_3, database=conn)
```

Out[12]:

	total_payloadmass
0	45596

Average Payload Mass by F9 v1.1

Display average payload mass carried by booster version F9 v1.1

In [13]:
```
task_4 = '''
        SELECT AVG(PayloadMassKG) AS Avg_PayloadMass
        FROM SpaceX
        WHERE BoosterVersion = 'F9 v1.1'
        '''
create_pandas_df(task_4, database=conn)
```

Out[13]:

	avg_payloadmass
0	2928.4

First Successful Ground Landing Date

```
In [14]: task_5 = '''
                 SELECT MIN(Date) AS FirstSuccessfull_landing_date
                 FROM SpaceX
                 WHERE LandingOutcome LIKE 'Success (ground pad)'
                 '''
         create_pandas_df(task_5, database=conn)
```

Out[14]:

	firstsuccessfull_landing_date
0	2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

```
task_6 = '''
        SELECT BoosterVersion
        FROM SpaceX
        WHERE LandingOutcome = 'Success (drone ship)'
            AND PayloadMassKG > 4000
            AND PayloadMassKG < 6000
        '''
create_pandas_df(task_6, database=conn)
```

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

```
task_7a = '''
        SELECT COUNT(MissionOutcome) AS SuccessOutcome
        FROM SpaceX
        WHERE MissionOutcome LIKE 'Success%'
        '''

task_7b = '''
        SELECT COUNT(MissionOutcome) AS FailureOutcome
        FROM SpaceX
        WHERE MissionOutcome LIKE 'Failure%'
        '''
print('The total number of successful mission outcome is:')
display(create_pandas_df(task_7a, database=conn))
print()
print('The total number of failed mission outcome is:')
create_pandas_df(task_7b, database=conn)
```

```
The total number of successful mission outcome is:
```

	successoutcome
0	100

```
The total number of failed mission outcome is:
```

	failureoutcome
0	1

Boosters Carried Maximum Payload

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
task_8 = '''
        SELECT BoosterVersion, PayloadMassKG
        FROM SpaceX
        WHERE PayloadMassKG = (
                                SELECT MAX(PayloadMassKG)
                                FROM SpaceX
                                )
        ORDER BY BoosterVersion
        '''
create_pandas_df(task_8, database=conn)
```

	boosterversion	payloadmasskg
0	F9 B5 B1048.4	15600
1	F9 B5 B1048.5	15600
2	F9 B5 B1049.4	15600
3	F9 B5 B1049.5	15600
4	F9 B5 B1049.7	15600
5	F9 B5 B1051.3	15600
6	F9 B5 B1051.4	15600
7	F9 B5 B1051.6	15600
8	F9 B5 B1056.4	15600
9	F9 B5 B1058.3	15600
10	F9 B5 B1060.2	15600
11	F9 B5 B1060.3	15600

2015 Launch Records

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

```
task_9 = '''
        SELECT BoosterVersion, LaunchSite, LandingOutcome
        FROM SpaceX
        WHERE LandingOutcome LIKE 'Failure (drone ship)'
            AND Date BETWEEN '2015-01-01' AND '2015-12-31'
        '''
create_pandas_df(task_9, database=conn)
```

	boosterversion	launchsite	landingoutcome
0	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
1	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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))

```
task_10 = '''
        SELECT LandingOutcome, COUNT(LandingOutcome)
        FROM SpaceX
        WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20'
        GROUP BY LandingOutcome
        ORDER BY COUNT(LandingOutcome) DESC
        '''
create_pandas_df(task_10, database=conn)
```

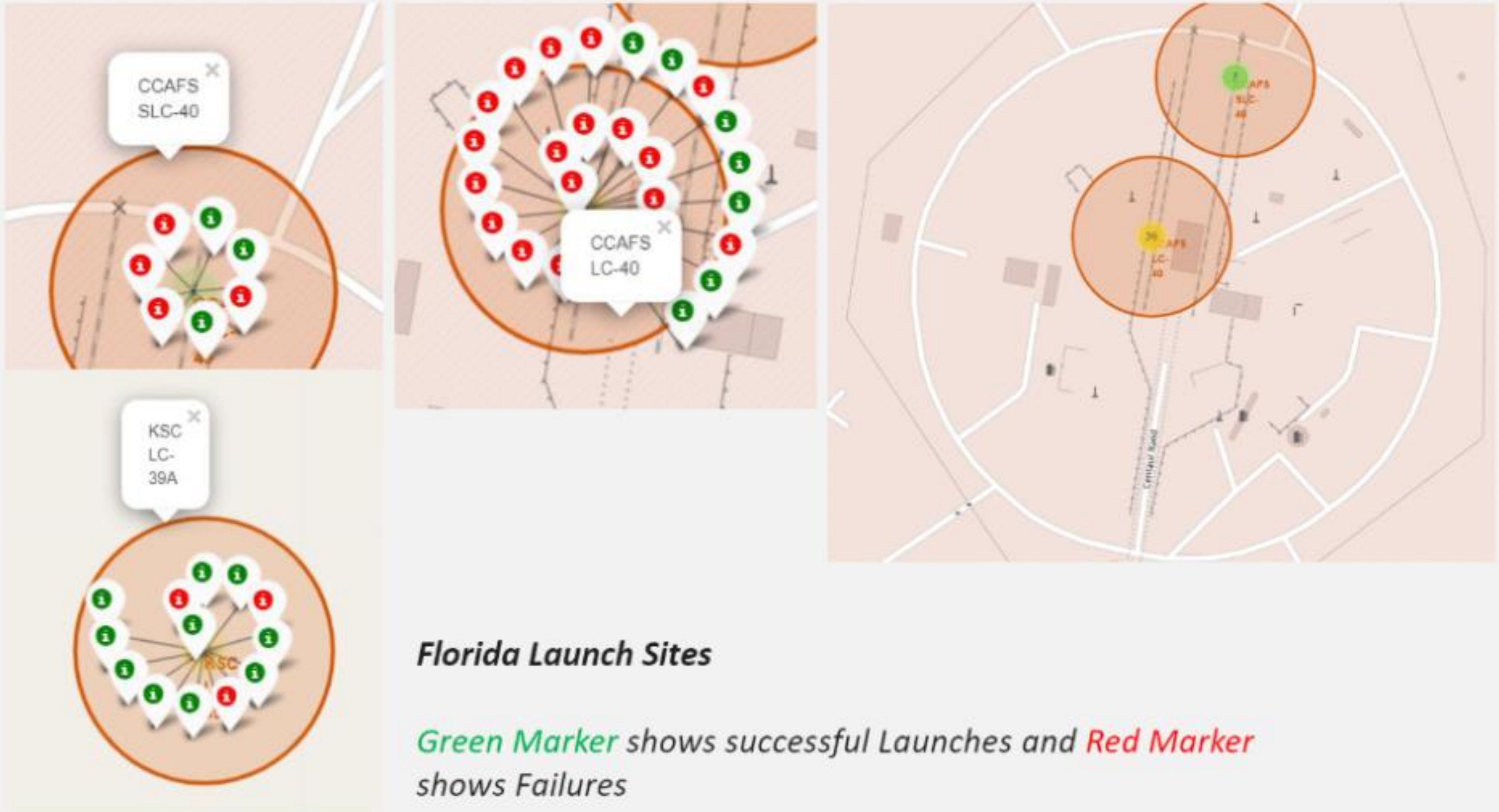
	landingoutcome	count
0	No attempt	10
1	Success (drone ship)	6
2	Failure (drone ship)	5
3	Success (ground pad)	5
4	Controlled (ocean)	3
5	Uncontrolled (ocean)	2
6	Precluded (drone ship)	1
7	Failure (parachute)	1

Section 3

Launch Sites Proximities Analysis

Launch Site Locations

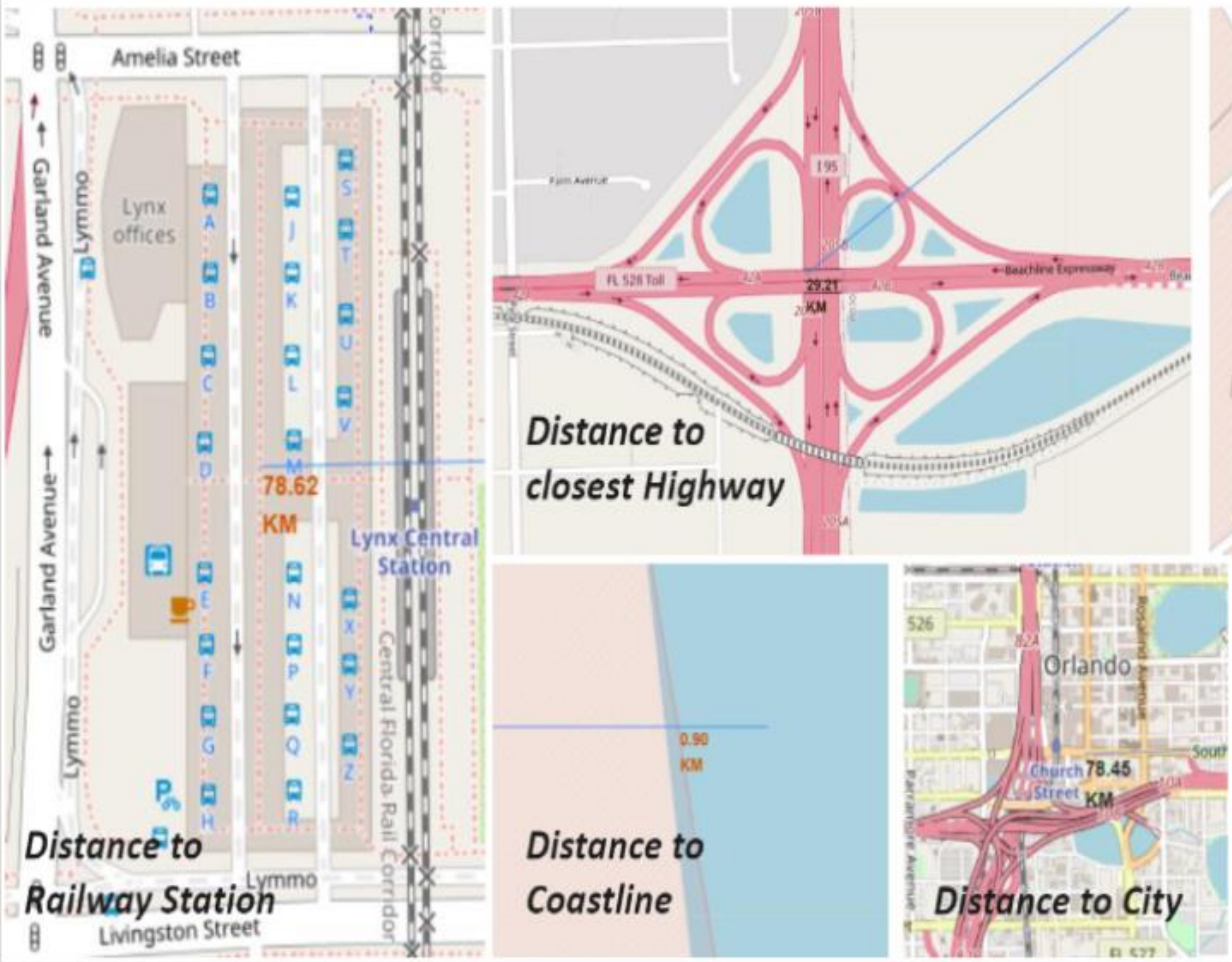
Florida Launch Sites



Florida Launch Sites

Green Marker shows successful Launches and Red Marker shows Failures

Launch Site distance to landmarks

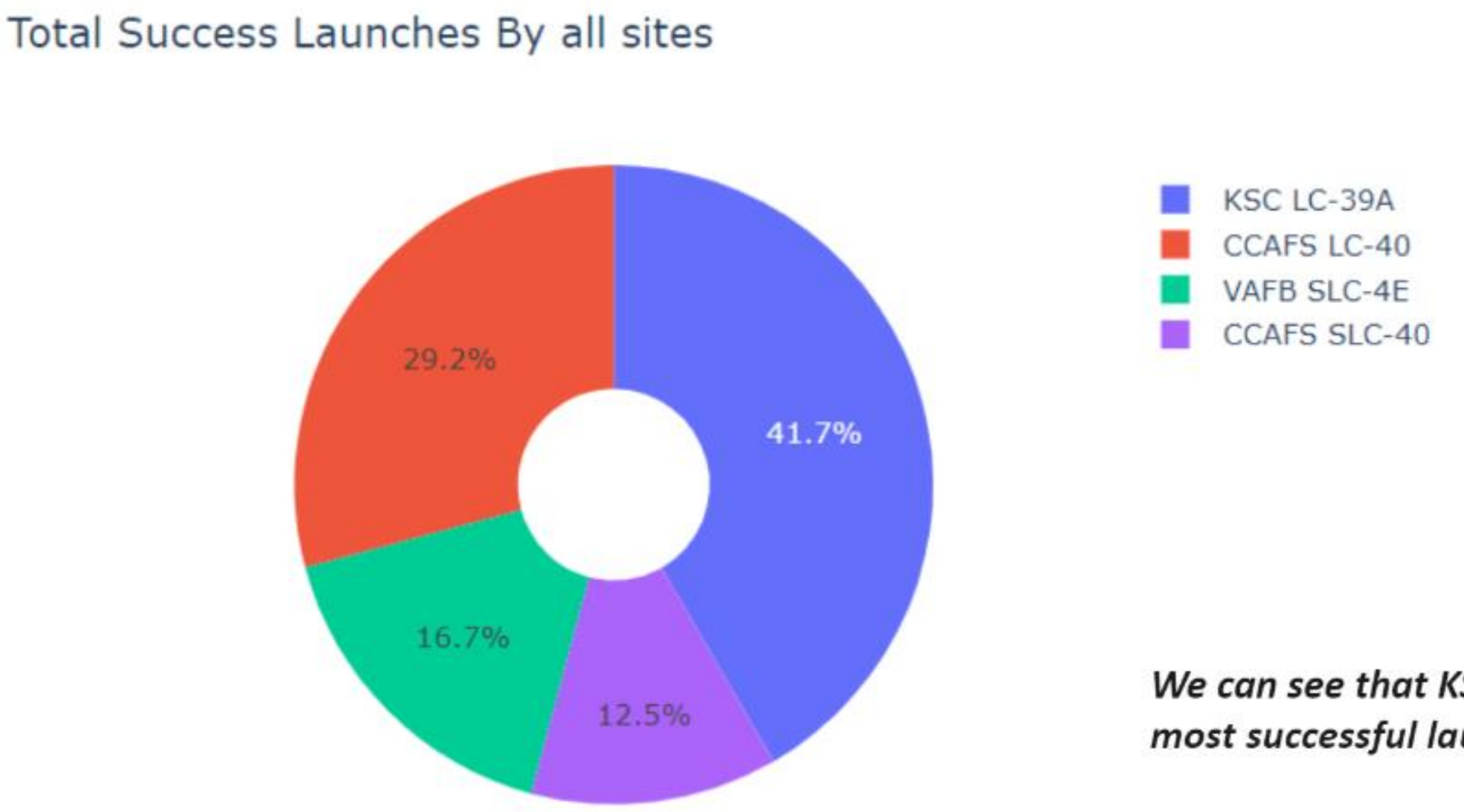


- Are launch sites in close proximity to railways? No
- Are launch sites in close proximity to highways? No
- Are launch sites in close proximity to coastline? Yes
- Do launch sites keep certain distance away from cities? Yes

Section 4

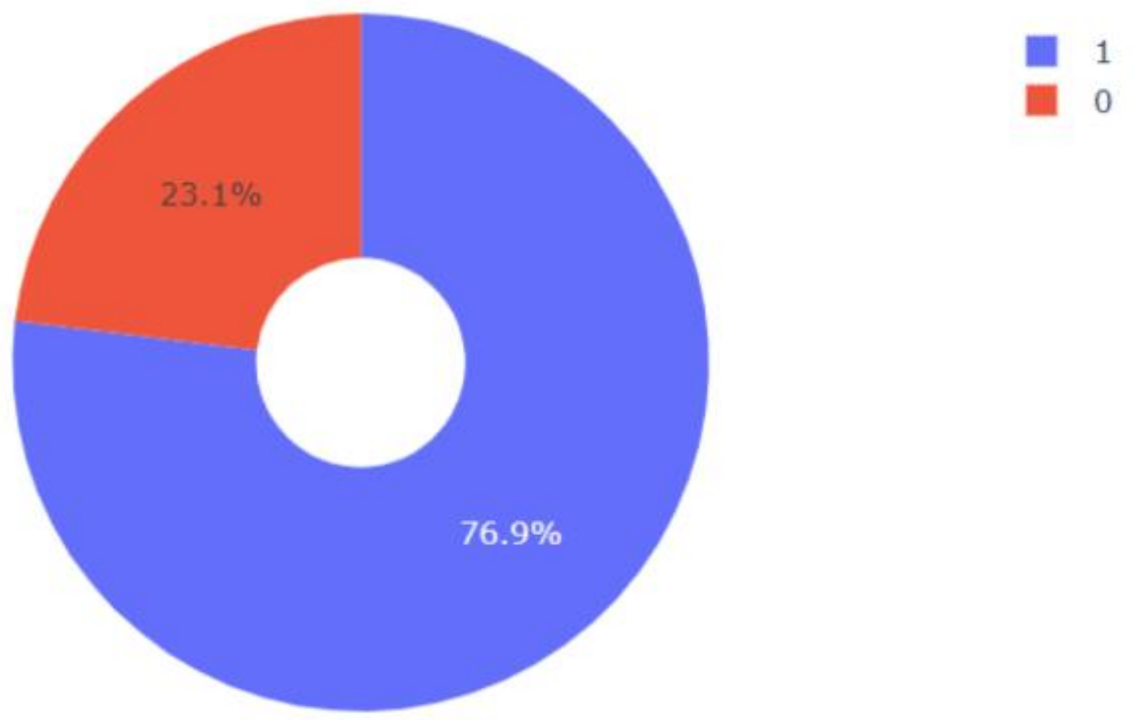
Build a Dashboard with Plotly Dash

Success by launch site



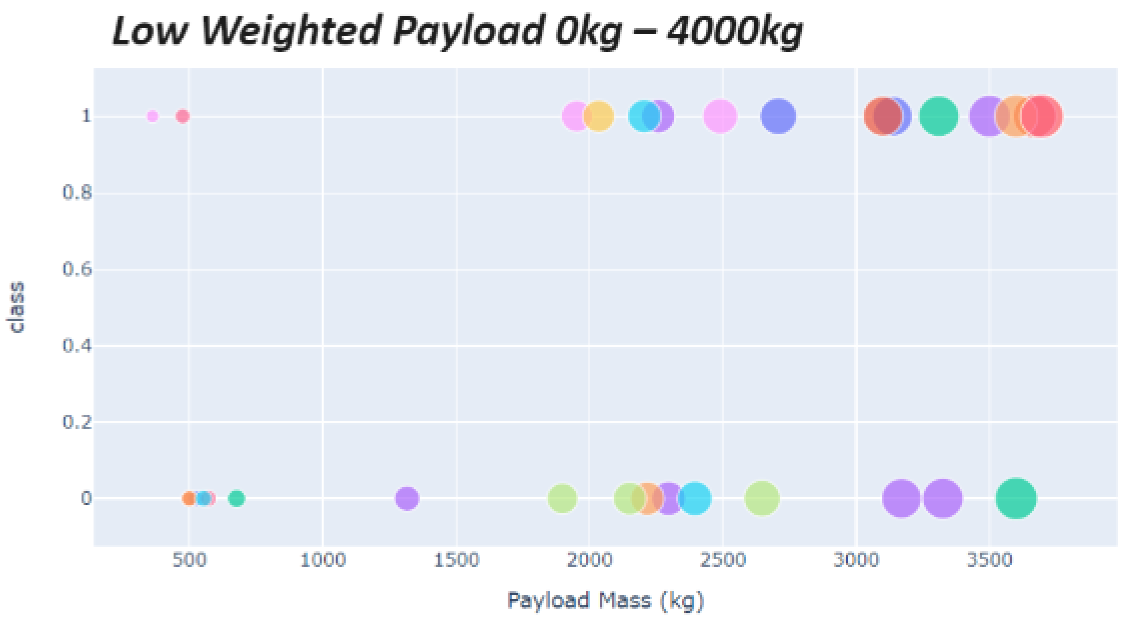
We can see that KSC LC-39A had the most successful launches from all the sites

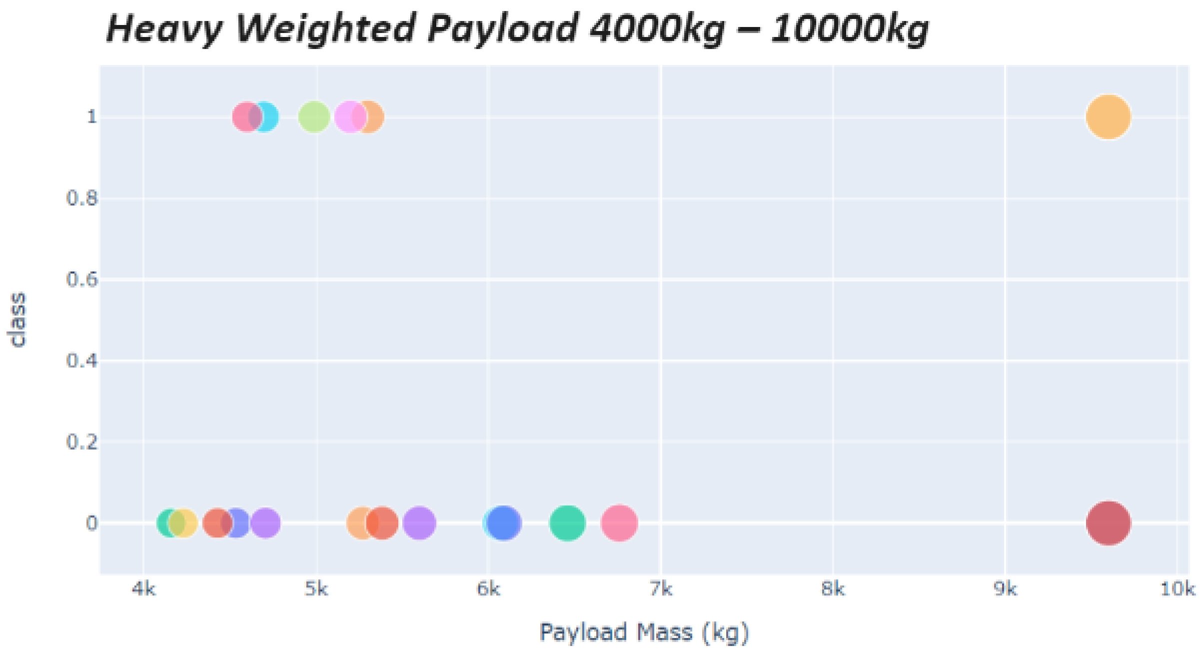
Success rate KSC LC-39A



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

Payload vs. Launch Outcome





We can see the success rates for low weighted payloads is higher than the heavy weighted payloads

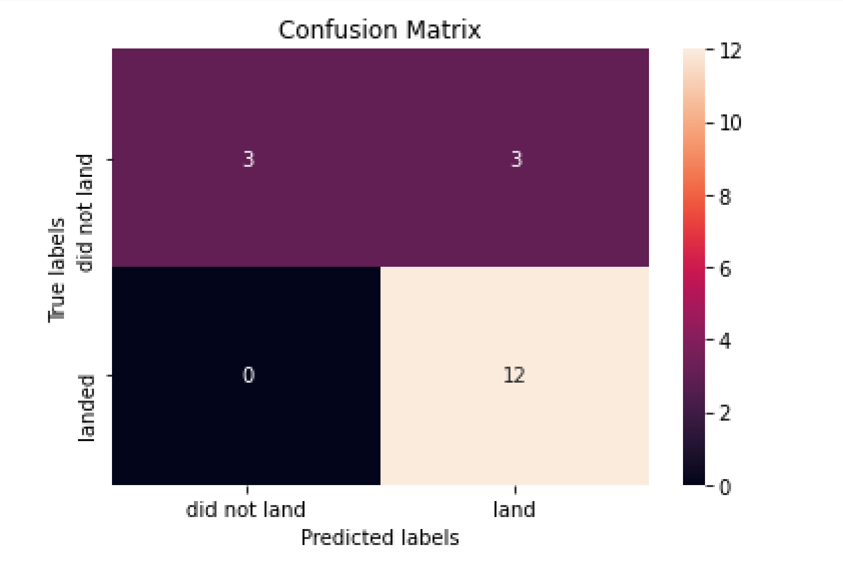
Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Tree algorithm has best accuracy

Confusion Matrix



Conclusions

- Tree Classifier Algorith is best for ML for this data
- Lower payload is more successful than heavier payload
- KSC LC-39A has most successful launches
- ES-L1, GEO, HEO and SSO have best success rate

Thank you!