IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Megha SALVE

28th November 2021

Outline

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

Executive Summary

- Data Collection is done using Spacex API and Webscrapping
- Exploratory Data Analysis is done with Visualization and SQL
- Launch sites are located on map and launch outcomes at eact site are marked
- Interactive plots are generated using Plotly dash
- Classifier models, log Regression, SVM, Decision Tree and KNN are built and tested for best prediction accuracy

Introduction

Project background and context

- This Project uses data analysis to determine if the first stage of Falcon-9 launched by SpaceX will safely land. The successful landing of the first stage enables reusing the launcher and can end up cutting launch cost to half or even more.

Problems to answer

- What factors are capable of affecting the landing outcome?
- Does individual site location has any effect on landing outcome?
- What is the best classifier model to predict the landing outcome of Falcon9 first stage?

Section 1

Methodology

Methodology

Data collection

- SpaceX API
- https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

Data Wrangling

- Convert data into meaningful and understanding form
- Taking care of missing data

Methodology

Exploratory data analysis (EDA) using visualization and SQL

Interactive visual analytics using Folium

- Understanding geographical patterns about launch sites.

Interactive visual analytics using Plotly Dash

- Interactive visual analytics on SpaceX launch data in real-time.
- Pie chart and Payload mass Slider to visualize success and failure data of all sites or for individual site

Predictive analysis using classification models

- Logarithmic Regression
- Support Vector Machine
- Decision Tree
- K Nearest Neighbour

Data Collection

- Request to the SpaceX API
- Webscraping the wikipedia page:
    - https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

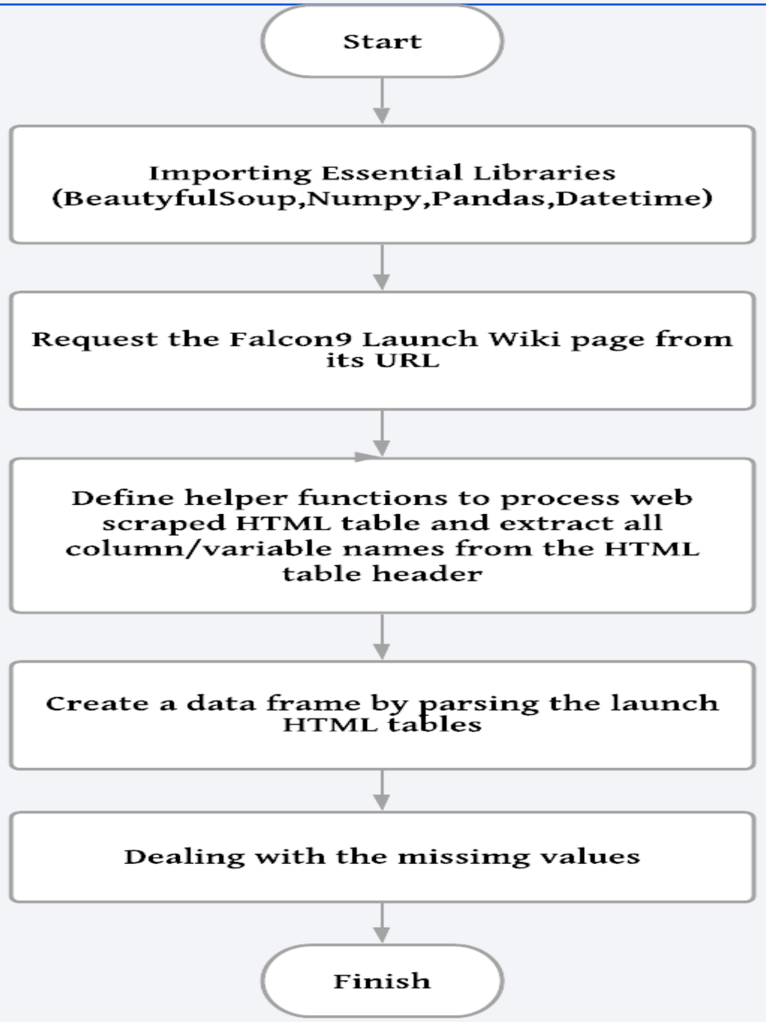
- SpaceX API requests are used to get information on rocket, payloads, launchpad, and cores to make data consistent.
- Collected data is filtered for only Falcon 9 launches
- Missing values of Payload masses are replaced by mean of all payload masses

Data Collection – SpaceX API

- https://github.com/Megha1210/Data-Science-Capstone-Project/actions/new

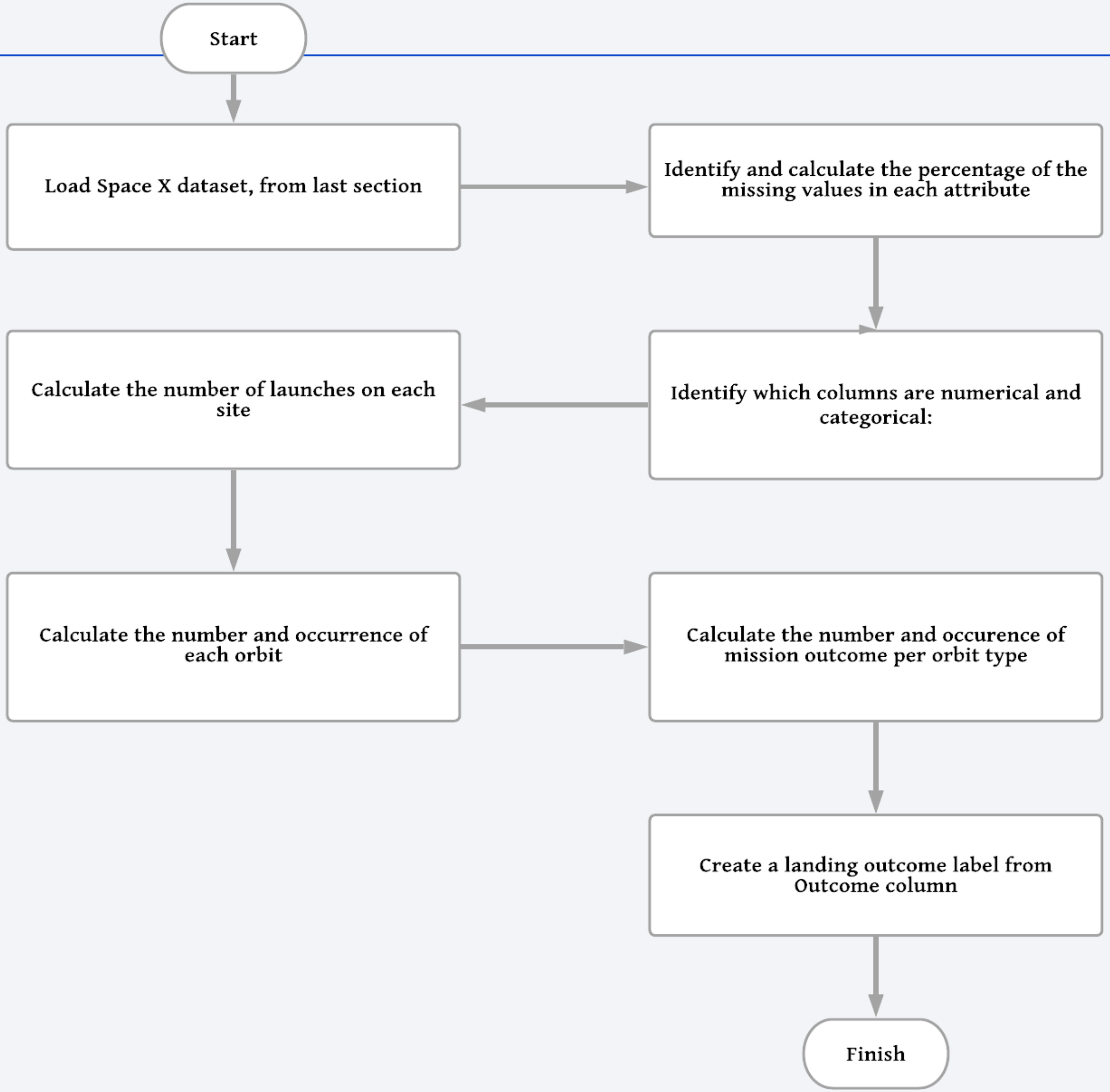
Data Collection - Scraping

- https://github.com/Megha1210/Data-Science-Capstone-Project/blob/main/Capstone%20Project_%20Webscraping.ipynb



Data Wrangling

- https://github.com/Megha1210/Data-Science-Capstone-Project/blob/main/Capstone%20Project_%20EDA.ipynb



EDA with Data Visualization

- https://github.com/Megha1210/Data-Science-Capstone-Project/blob/main/Capstone%20Project_%20EDA%20with%20Data%20Visualization.ipynb
- Relationship between Flight Number and Launch Site
- Relationship between Payload and Launch Site
- Relationship between success rate of each orbit type
- Relationship between Flight Number and Orbit type
- Relationship between Payload and Orbit type
- Launch success yearly trend

EDA with SQL

- https://github.com/Megha1210/Data-Science-Capstone-Project/blob/main/Capstone%20Project_%20EDA%20with%20SQL_SOLVED.ipynb

SQL queries performed

- SELECT UNIQUE(LAUNCH_SITE) FROM SPACEXDATASET
- SELECT DATE, LAUNCH_SITE FROM SPACEXDATASET where LAUNCH_SITE like 'CCA%' LIMIT 5
- SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXDATASET where CUSTOMER='NASA (CRS)'
- SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXDATASET where BOOSTER_VERSION='F9 v1.1'
- SELECT MIN(DATE) FROM SPACEXDATASET where LANDING__OUTCOME='Success (ground pad)'
- SELECT BOOSTER_VERSION,PAYLOAD_MASS__KG_,LANDING__OUTCOME FROM SPACEXDATASET

where LANDING__OUTCOME='Success (drone ship)' AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000

EDA with SQL

- SELECT COUNT(DATE) TOTAL, MISSION_OUTCOME FROM SPACEXDATASET
GROUP BY MISSION_OUTCOME
ORDER BY COUNT(DATE) DESC ;

- SELECT booster_version,PAYLOAD_MASS__KG_ FROM SPACEXDATASET
WHERE PAYLOAD_MASS__KG_= (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXDATASET);

- SELECT BOOSTER_VERSION,launch_site,LANDING__OUTCOME,DATE FROM SPACEXDATASET
WHERELANDING__OUTCOME='Failure (drone ship)' AND YEAR(DATE) = 2015

- SELECT COUNT(DATE) AS FREQUENCY, LANDING__OUTCOME FROM SPACEXDATASET
WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY LANDING__OUTCOME
ORDER BY COUNT(DATE) DESC;

Build an Interactive Map with Folium

- https://github.com/Megha1210/Data-Science-Capstone-Project/blob/main/Capstone%20Project_%20Folium.ipynb

  - The launch success rate may also depend on the location and proximities of a launch site, i.e., the initial position of rocket trajectories
  - Following slides show launch site locations and launch outcomes for each site location.

Build an Interactive Map with Folium

- **map objects to clearly visualize loacations and proximities to Launch sites**
- **Folium object 'Circle' to point the site location** ex.for NASA Centre
  - circle = folium.Circle(nasa_coordinate, radius=1000, color='#d35400', fill=True).add_child(folium.Popup('NASA Johnson Space Center'))

- **Folium object 'Marker' to show name of the site location** ex. NASA Johnson Space Center's coordinate with a icon showing its name
  - marker = folium.map.Marker(nasa_coordinate, icon=DivIcon(icon_size=(20,20), icon_anchor=(0,0), html='<div style="font-size: 12; color:#d35400;">%s</div>' % 'NASA JSC',))

    site_map.add_child(circle)

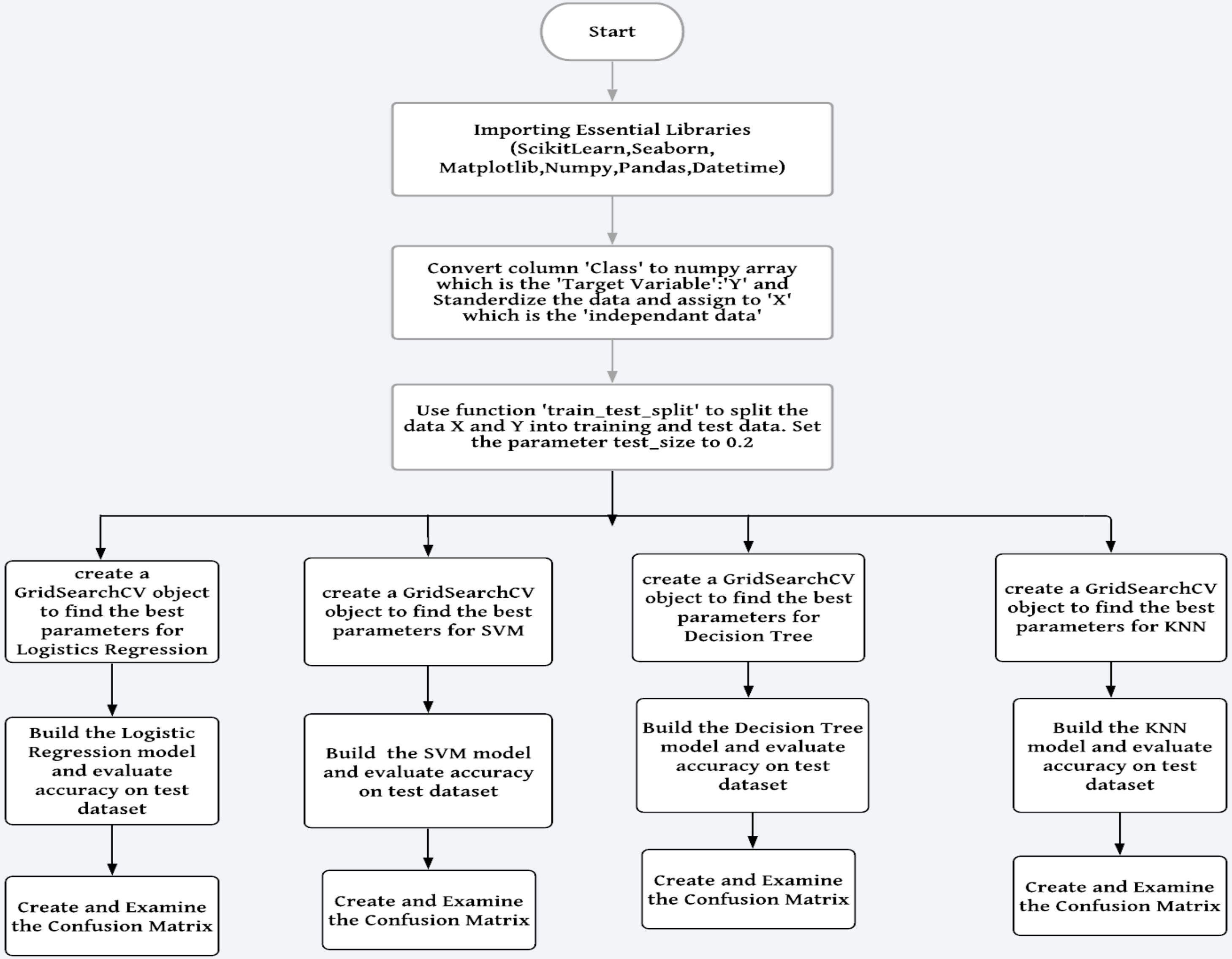
    marker.add_to(site_map)

Build a Dashboard with Plotly Dash

- https://github.com/Megha1210/Data-Science-Capstone-Project/blob/main/CapstoneProject_PlotlyDash.ipynb

Predictive Analysis (Classification)

https://github.com/Megha1210/Data-Science-Capstone-Project/blob/main/Capstone%20Project_%20Machine%20Learning.ipynb

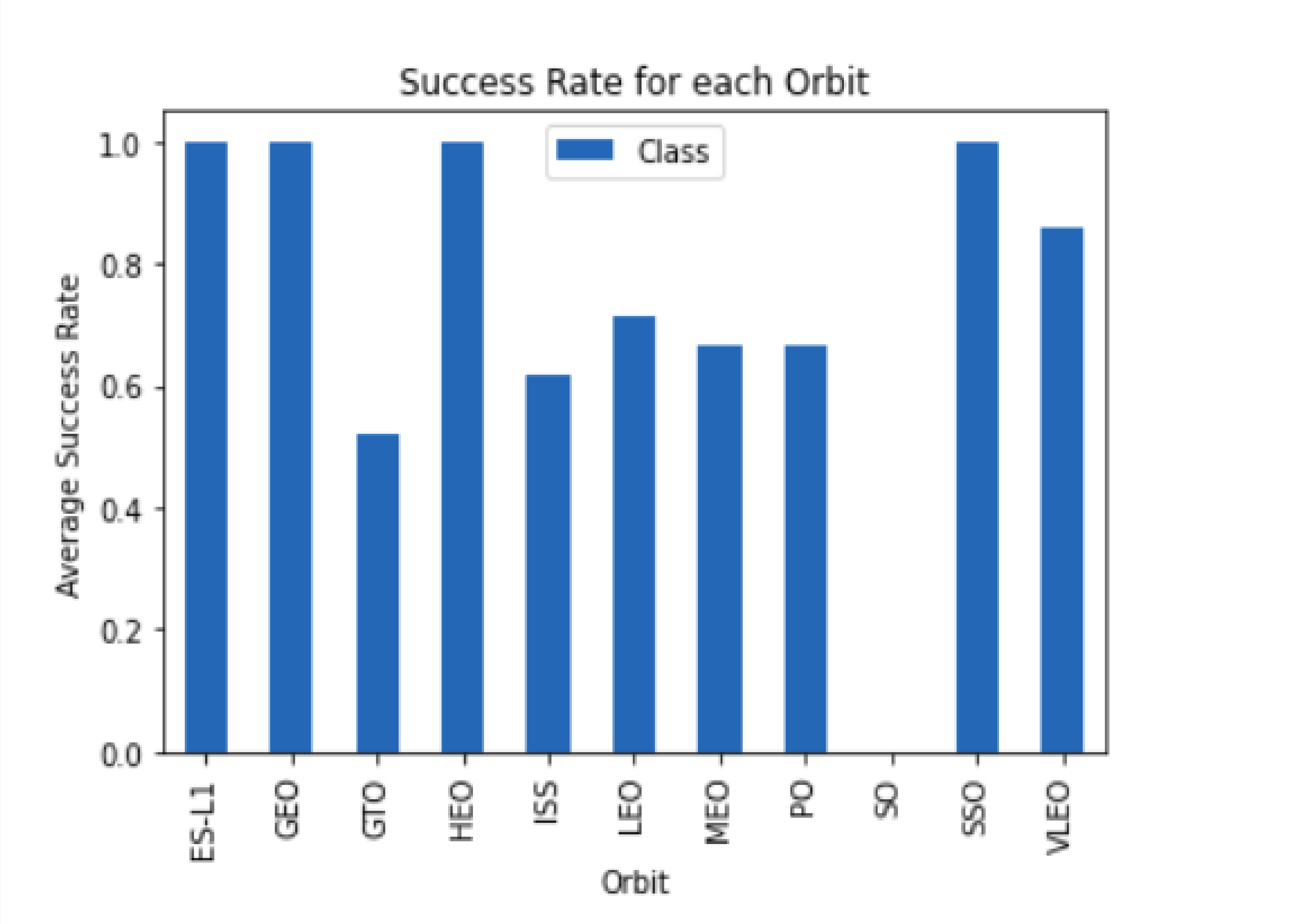


Section 2

Insights drawn from EDA

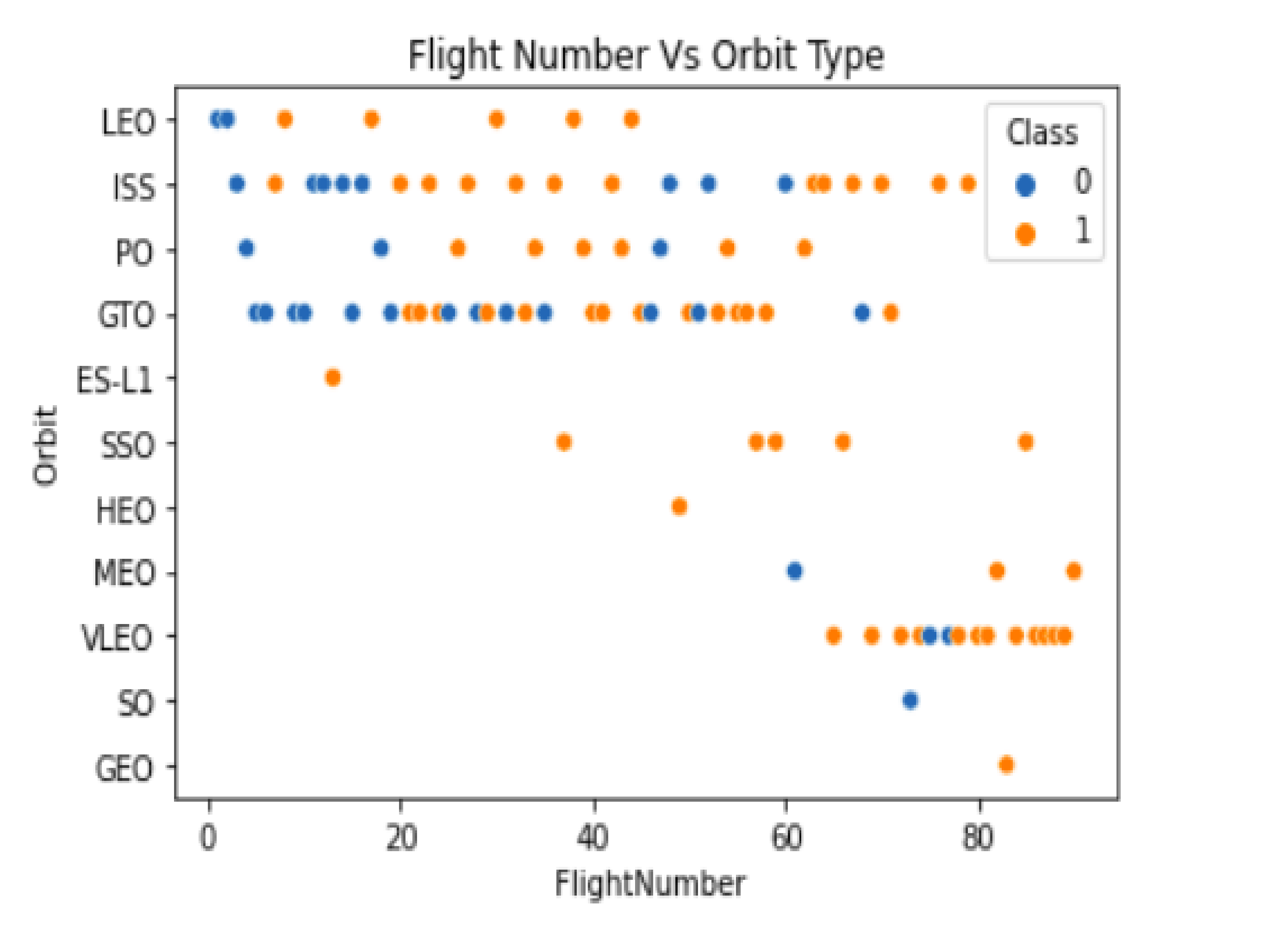
Success Rate vs. Orbit Type

- ES-L1, HEO, LEO, SSO orbits have highest success rate

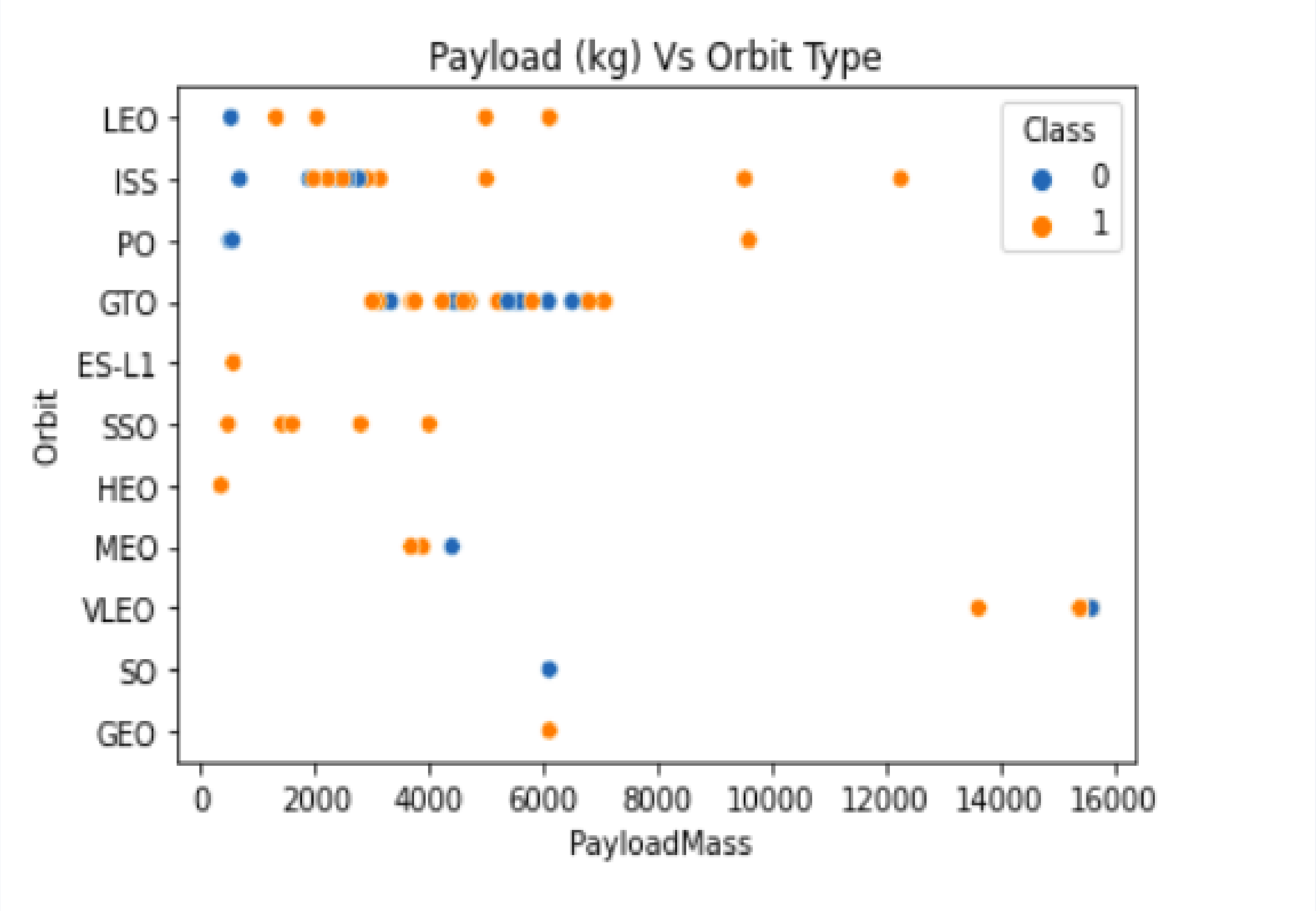


Flight Number vs. Orbit Type

- No clear relationship is established between Flight number and orbit types



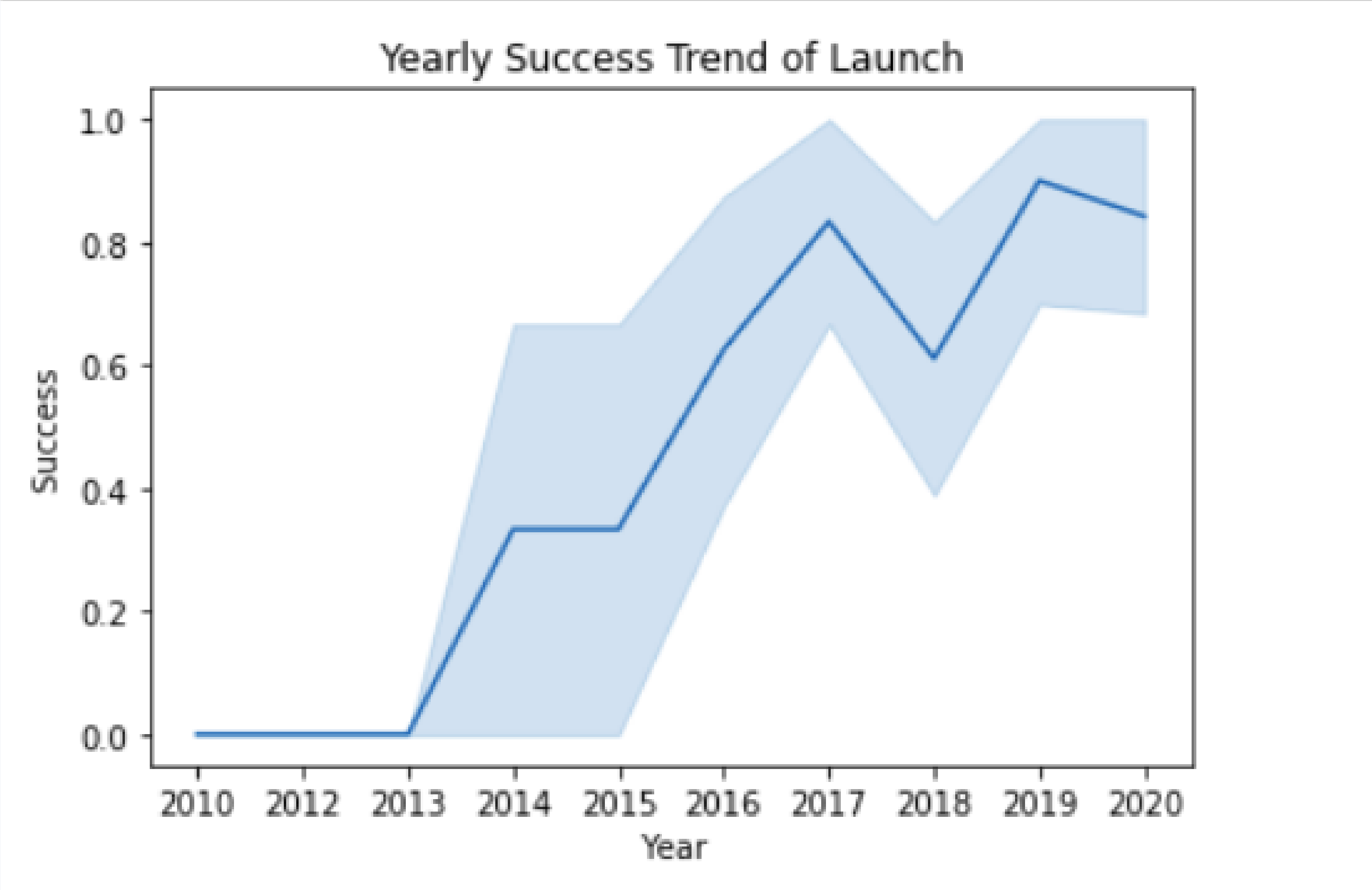
Payload vs. Orbit Type



- Heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS however, GTO doesn’t show any realtion

Launch Success Yearly Trend

- Success rate from year 2013 has kept increasing



All Launch Site Names

```
In [11]: %sql SELECT UNIQUE(LAUNCH_SITE) FROM SPACEXDATASET

 * ibm_db_sa://scp19627:***@ea286ace-86c7-4d5b-8580-3fbfa46b1c66.bs2io90l08kqb1od8lcg.datab
ases.appdomain.cloud:31505/BLUDB
Done.
```

Out[11]:

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

Launch Site Names Begin with 'CCA'

```
In [12]: %sql SELECT DATE, LAUNCH_SITE FROM SPACEXDATASET where LAUNCH_SITE like 'CCA%' LIMIT 5

 * ibm_db_sa://scp19627:***@ea286ace-86c7-4d5b-8580-3fbfa46b1c66.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31505/BLUDB
Done.

Out[12]:
```

DATE	launch_site
2010-06-04	CCAFS LC-40
2010-12-08	CCAFS LC-40
2012-05-22	CCAFS LC-40
2012-10-08	CCAFS LC-40
2013-03-01	CCAFS LC-40

Total Payload Mass

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

In [13]:
```
%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXDATASET where CUSTOMER='NASA (CRS)'
```

```
 * ibm_db_sa://scp19627:***@ea286ace-86c7-4d5b-8580-3fbfa46b1c66.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31505/BLUDB
Done.
```

Out[13]:

1
45596

Average Payload Mass by F9 v1.1

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

```
In [14]: %sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXDATASET where BOOSTER_VERSION='F9 v1.1'

 * ibm_db_sa://scp19627:***@ea286ace-86c7-4d5b-8580-3fbfa46b1c66.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31505/BLUDB
Done.
```

Out[14]:

1
2928

First Successful Ground Landing Date

```
In [20]: %sql SELECT MIN(DATE) FROM SPACEXDATASET where LANDING__OUTCOME='Success (ground pad)'

 * ibm_db_sa://scp19627:***@ea286ace-86c7-4d5b-8580-3fbfa46b1c66.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31505/BLUDB
Done.

Out[20]:
```

1
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

In [22]:
```
%%sql
SELECT BOOSTER_VERSION,PAYLOAD_MASS__KG_,LANDING__OUTCOME FROM SPACEXDATASET
where LANDING__OUTCOME='Success (drone ship)' AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000
```

```
 * ibm_db_sa://scp19627:***@ea286ace-86c7-4d5b-8580-3fbfa46b1c66.bs2io90l08kqb1od8lcg.datab
ases.appdomain.cloud:31505/BLUDB
Done.
```

Out[22]:

booster_version	payload_mass__kg_	landing__outcome
F9 FT B1022	4696	Success (drone ship)
F9 FT B1026	4600	Success (drone ship)
F9 FT B1021.2	5300	Success (drone ship)
F9 FT B1031.2	5200	Success (drone ship)

Total Number of Successful and Failure Mission Outcomes

In [60]:

```
%%sql
SELECT COUNT(DATE) TOTAL, MISSION_OUTCOME
FROM SPACEXDATASET
GROUP BY MISSION_OUTCOME
ORDER BY COUNT(DATE) DESC ;
```

```
 * ibm_db_sa://scp19627:***@ea286ace-86c7-4d5b-8580-3fbfa46b1c66.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31505/BLUDB
Done.
```

Out[60]:

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

Boosters Carried Maximum Payload

In [37]:
```
%%sql
SELECT booster_version,PAYLOAD_MASS__KG_ FROM SPACEXDATASET
WHERE PAYLOAD_MASS__KG_= (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXDATASET);
```

```
 * ibm_db_sa://scp19627:***@ea286ace-86c7-4d5b-8580-3fbfa46b1c66.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31505/BLUDB
Done.
```

Out[37]:

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

2015 Launch Records

```
In [40]: %%sql
SELECT BOOSTER_VERSION,launch_site,LANDING__OUTCOME,DATE FROM SPACEXDATASET
where LANDING__OUTCOME='Failure (drone ship)' AND YEAR(DATE) = 2015

 * ibm_db_sa://scp19627:***@ea286ace-86c7-4d5b-8580-3fbfa46b1c66.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31505/BLUDB
Done.
```

Out[40]:

booster_version	launch_site	landing__outcome	DATE
F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)	2015-01-10
F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)	2015-04-14

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

In [57]:

```
%%sql
SELECT COUNT(DATE) AS FREQUENCY, LANDING__OUTCOME
FROM SPACEXDATASET
WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY LANDING__OUTCOME
ORDER BY COUNT(DATE) DESC;
```

```
 * ibm_db_sa://scp19627:***@ea286ace-86c7-4d5b-8580-3fbfa46b1c66.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31505/BLUDB
Done.
```

Out[57]:

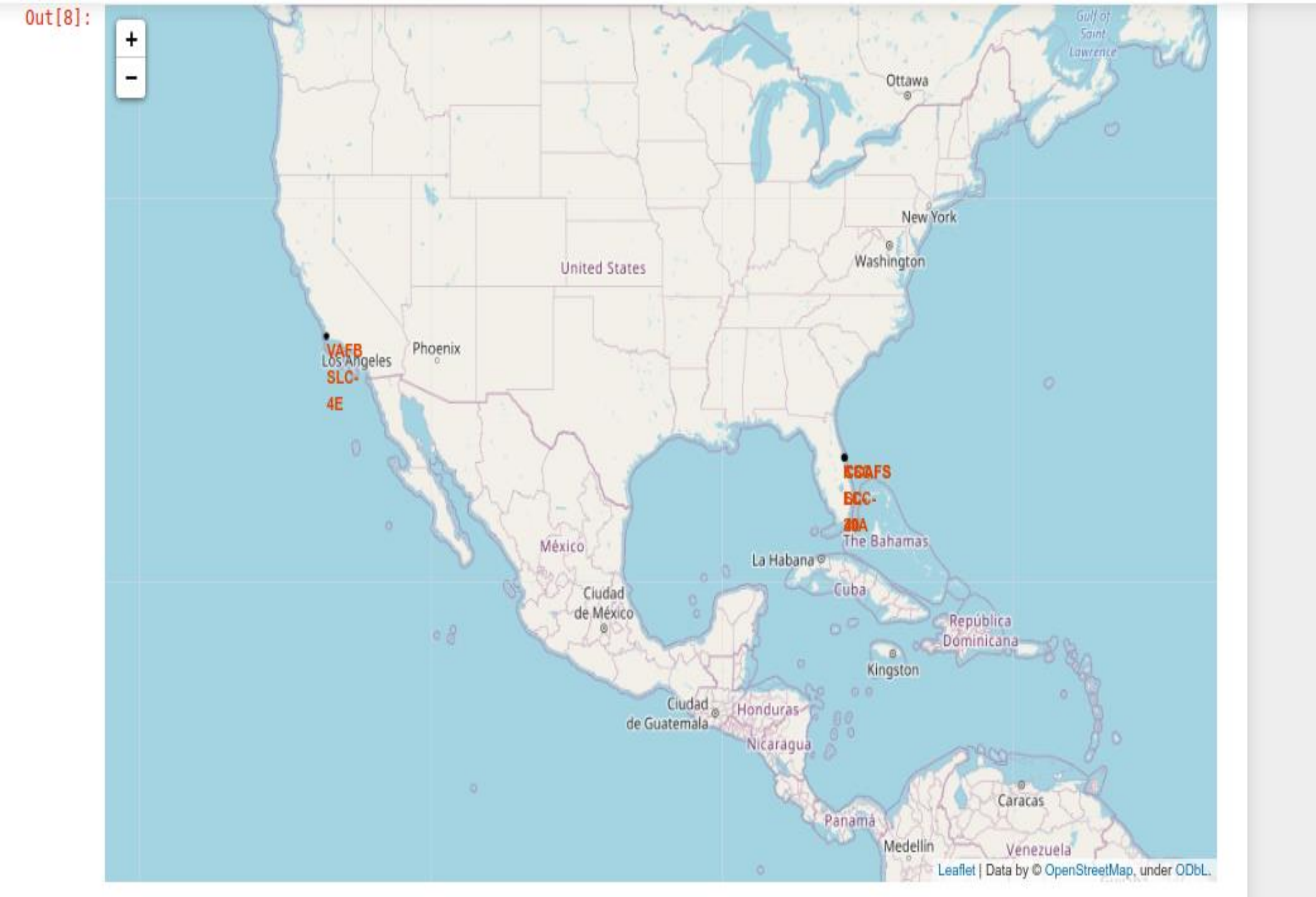
frequency	landing__outcome
10	No attempt
5	Failure (drone ship)
5	Success (drone ship)
3	Controlled (ocean)
3	Success (ground pad)
2	Failure (parachute)
2	Uncontrolled (ocean)
1	Precluded (drone ship)

Section 4

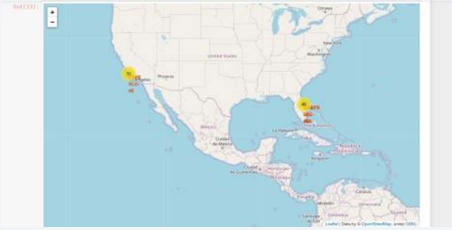
Launch Sites Proximities Analysis

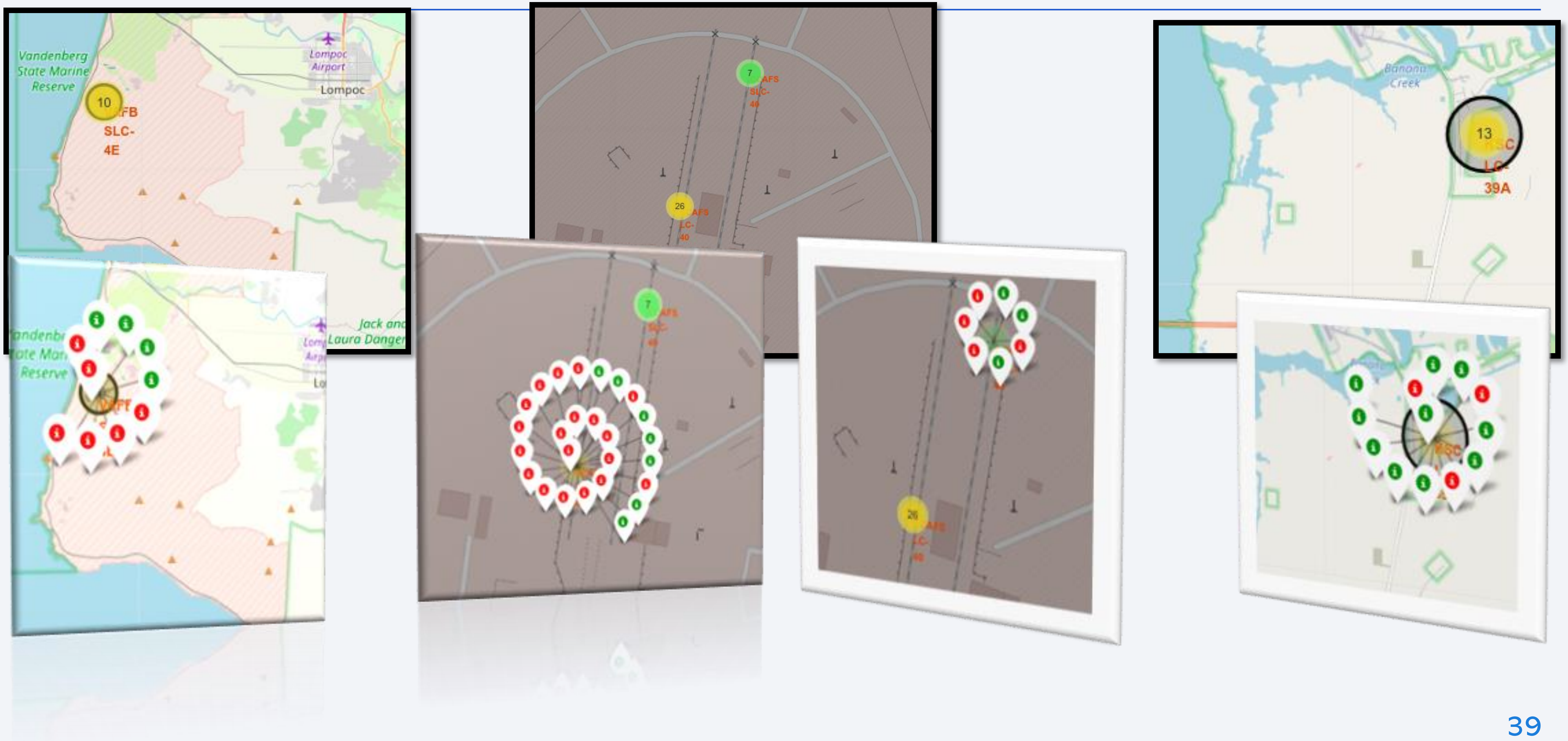
Launch Site Locations

- All Launch sites are near to the coastline



Launch Outcomes of each Site





Section 5

Build a Dashboard with Plotly Dash

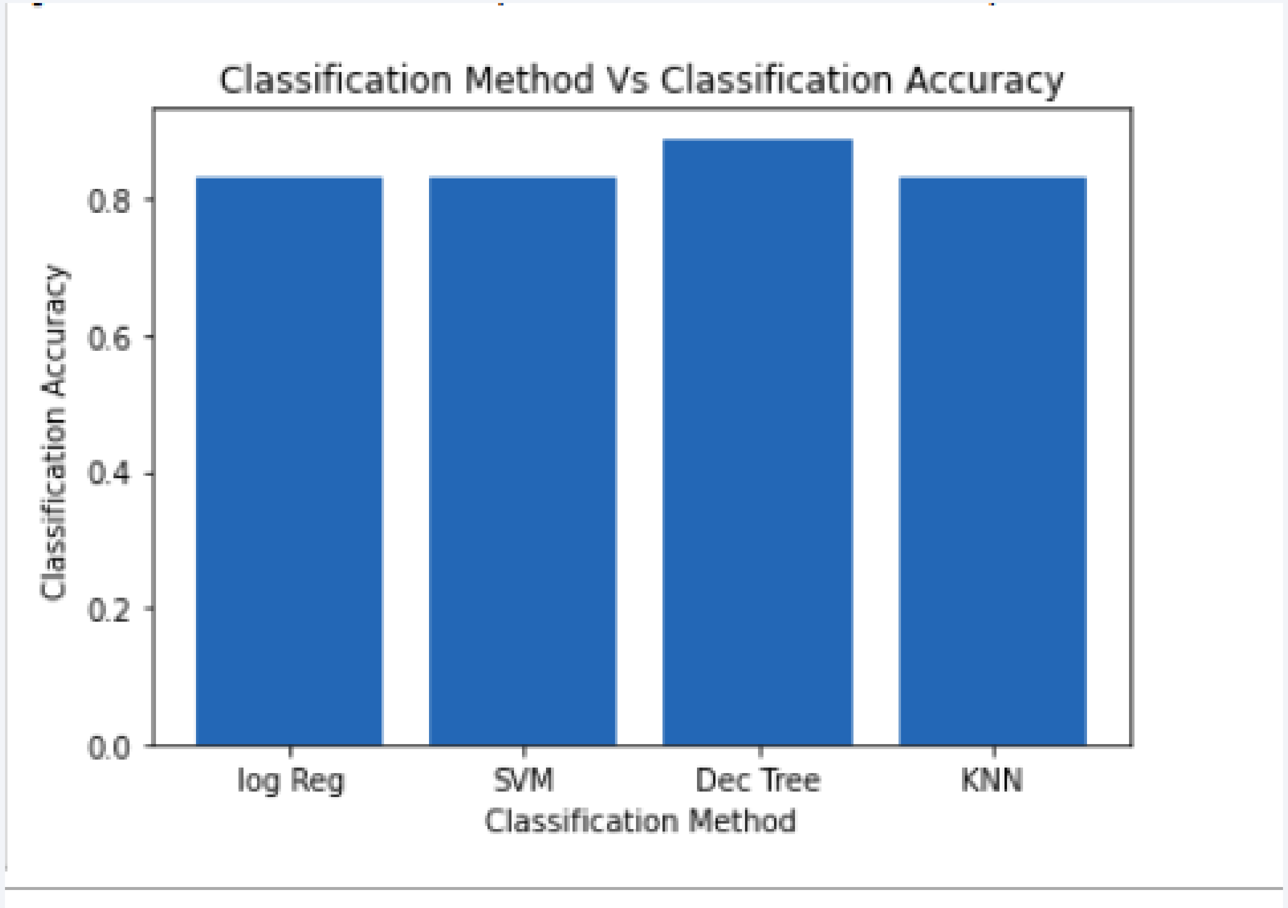
Section 6

Predictive Analysis (Classification)

Classification Accuracy

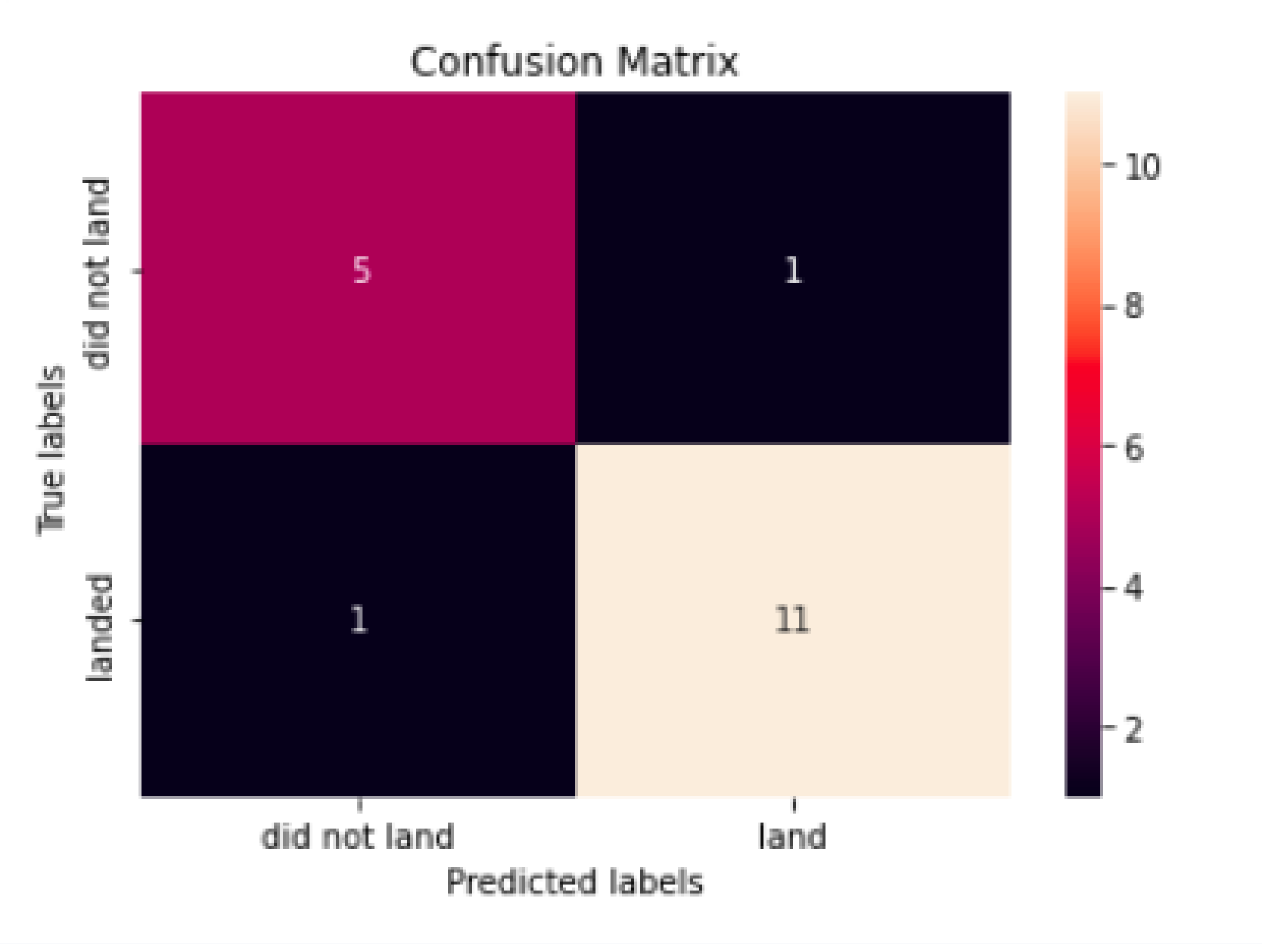
- The Decision Tree Classifier

Model is found to have highest accuracy



Confusion Matrix

- The confusion matrix of Decision tree model shows that the model has accurately predicted 11 out of 12 successful landings and 5 out of 6 failed landings based on test data set of 18 landing outcomes



Conclusions

- **The following factors have positive effect on successful landing**
  - Less Payload mass
  - Launches to ES-L1, HEO, LEO, SSO orbits
  - For heavy payloads ,Polar, LEO and ISS orbits are better choices
- **Obesrvations about landing sites**
  - SSC-LC-39A launch site have heighest success rate of 77%
  - CCAFS-LC-40 launch site have lowest success rate of 27%
- **Predictive Analysis**
  - To predict the successful landing outcome, the decision tree classifier model has highest accuracy of 83.3%

Thank you!