IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Name : Heet Shah
Date : 25-0-2023

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

- Summary of methodologies
  - Data collection
  - Data wrangling
  - EDA with data visualization
  - EDA with SQL
  - Building an interactive map with folium
  - Building a Dashboard with Plotly Dash
  - Predictive analysis(Classification)
- Summary of all results
  - EDA results
  - Interactive analytics
  - Predictive analysis

Introduction

- Project background and context
  - SpaceX advertise Falcon 9 rocket launches on its websites, with a cost of 62 million dollars; other providers cost upwards of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.
- Problems you want to find answers
  - The project task is to predicting if the first stage of SpaceX Falcon 9 rocket will land successfully.

Section 1

Methodology

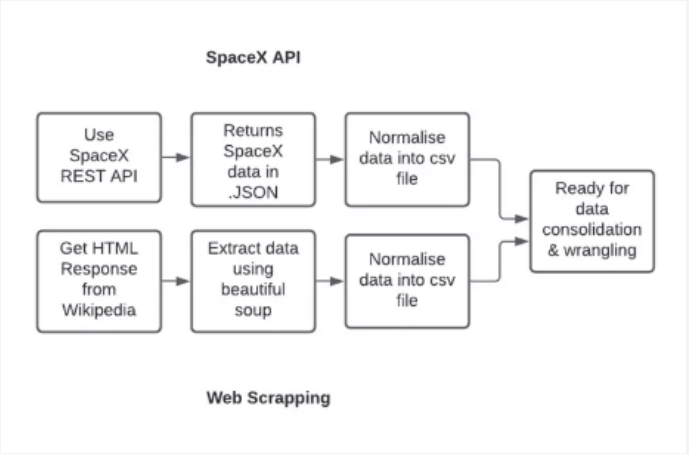
Methodology

Executive Summary

- Data collection methodology:
  - SpaceX Rest API
  - Web Scrapping from Wikipedia
- Perform data wrangling
  - One Hot Encoding data fields for machine learning and data cleaning of null values and irrelevant columns.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - LR, KNN, SVM, DT models have been build and evaluated for the best classifier

Data Collection

- Describe how data sets were collected.
  - SpaceX Launch data is gathered from SpaceX REST API
  - This API will give data about launches, including information about rocket used, payload delivered, launch specifications, landing specifications and landing outcome.
  - The SpaceX REST API endpoints , or URLS starts with api.spacexdata.com/v4/
  - Another popular data source for Falcon 9 Launch data is web scrapping Wikipedia.



Data Collection – SpaceX API

- Data collection with SpaceX REST calls

- Github Link : https://github.com/Heet09/Heet09--IBM-Data-Science-Capstone-SpaceX

Now let's start requesting rocket launch data from SpaceX API with the following URL:

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

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

Now we decode the response content as a Json using `.json()` and turn it into a Pandas dataframe using `.json_normalize()`

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

Finally lets construct our dataset using the data we have obtained. We we combine the columns into a dictionary.

```
launch_dict = {'FlightNumber': list(data['flight_number']),
'Date': list(data['date']),
'BoosterVersion':BoosterVersion,
'PayloadMass':PayloadMass,
'Orbit':Orbit,
'LaunchSite':LaunchSite,
'Outcome':Outcome,
'Flights':Flights,
'GridFins':GridFins,
'Reused':Reused,
'Legs':Legs,
'LandingPad':LandingPad,
'Block':Block,
'ReusedCount':ReusedCount,
'Serial':Serial,
'Longitude': Longitude,
'Latitude': Latitude}
```

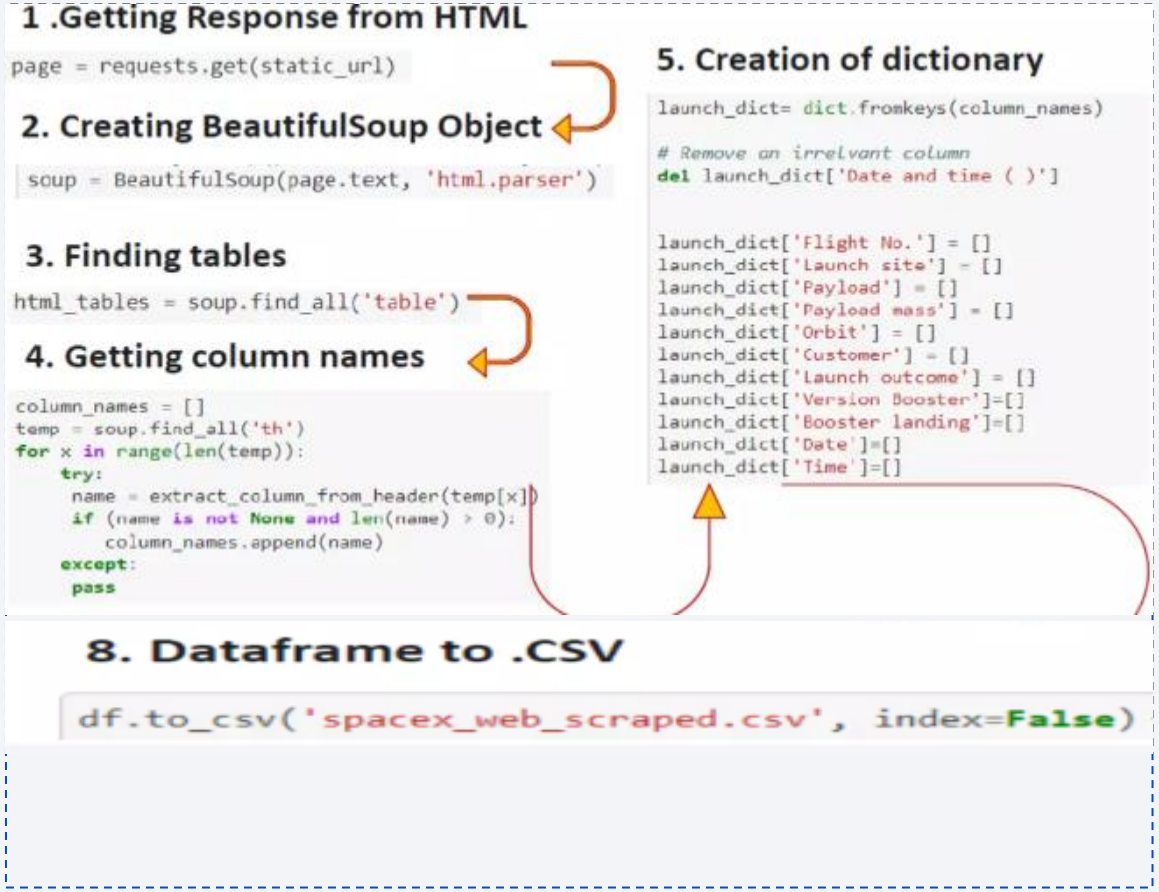
```
data_falcon9.to_csv('dataset_part_1.csv', index=False)
```


```
data_falcon9.to_csv('dataset_part_1.csv', index=False)
```

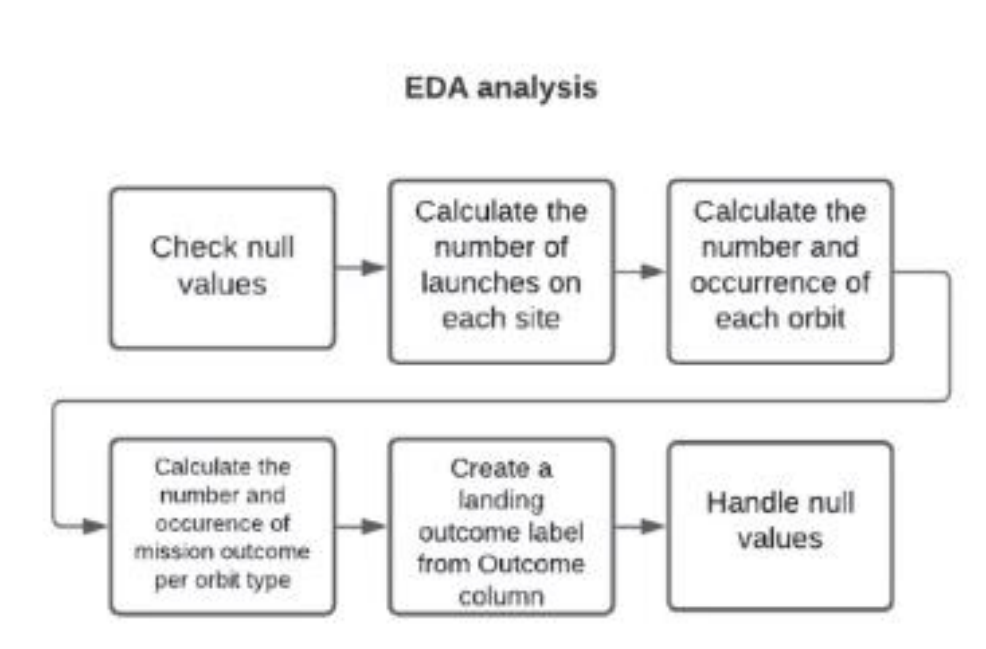
Data Collection - Scraping

- Web Scrapping from Wikipedia.

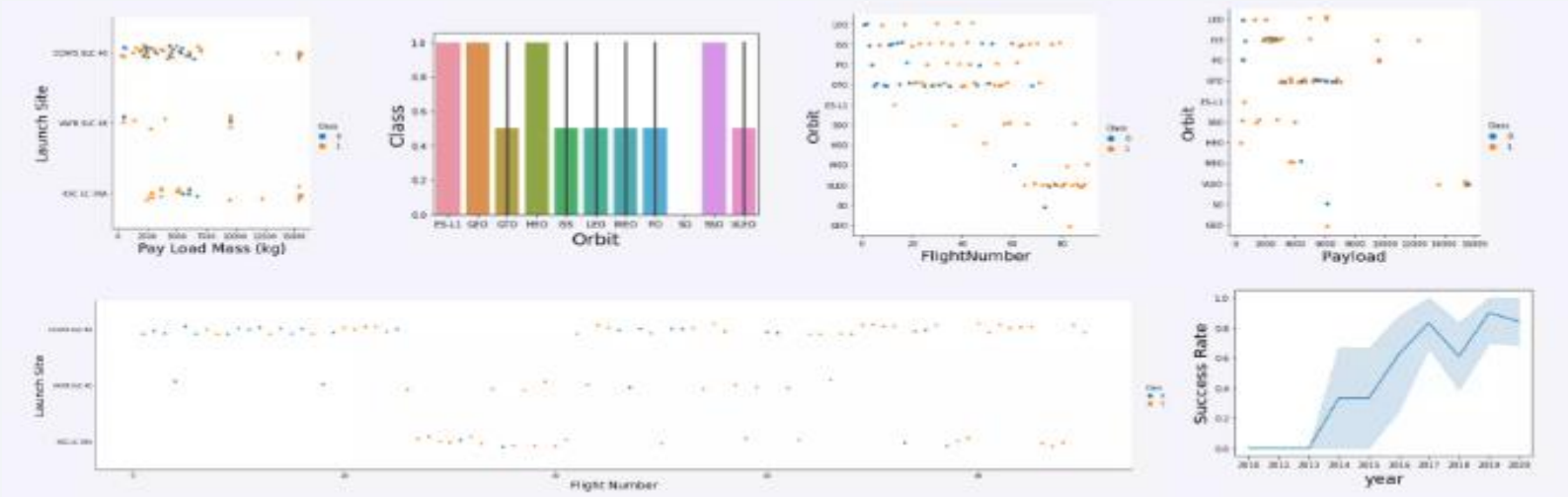
- Github Link : https://github.com/Heet09/Heet09--IBM-Data-Science-Capstone-SpaceX



Data Wrangling



EDA with Data Visualization



- Github Link : https://github.com/Heet09/Heet09--IBM-Data-Science-Capstone-SpaceX

EDA with SQL

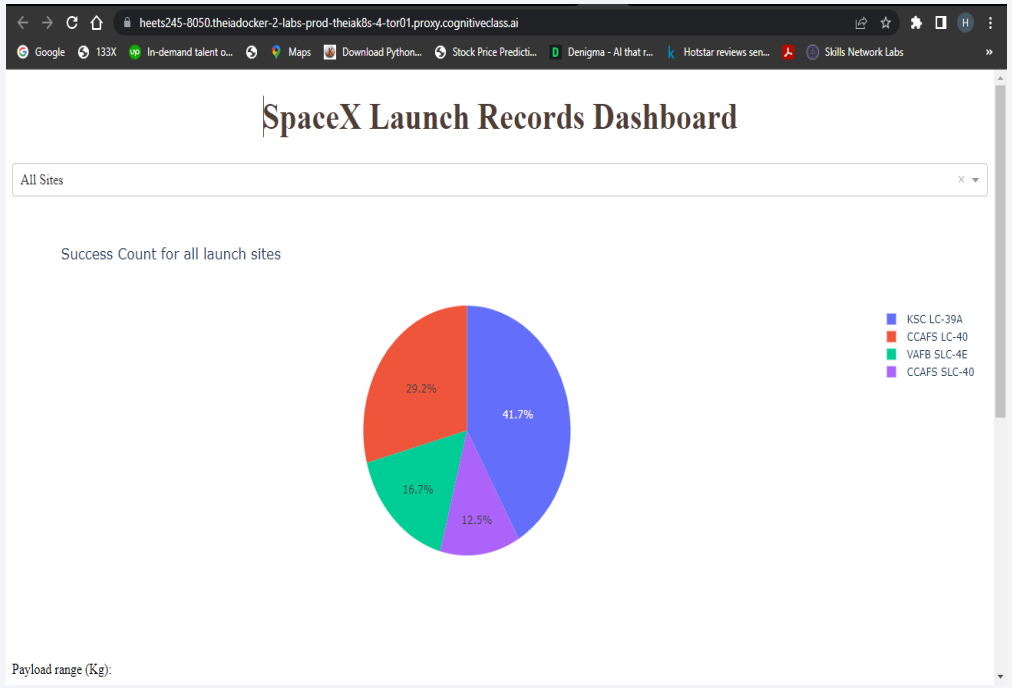
- **SQL Queries performed include:**
  - Displaying the names of unique launches in Space mission
  - Displaying 5 records where launch sites begin with string ‘KSC’
  - Displaying total payload mass carried by boosters launched by NASA (CRS)
  - Displaying average payload mass carried by boosters version F9 v1.1
  - Listing the date when the successful landing outcome in drone ship was achieved.
  - Listing the names of the boosters which have success in ground pad and have payload mass greater than 4000 but less than 6000.
  - Listing the total number of successful and failure mission outcomes
  - Listing the names of the booster_versions which have carried the maximum payload mass. Use a subquery
  - Listing the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015
  - Ranking the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order

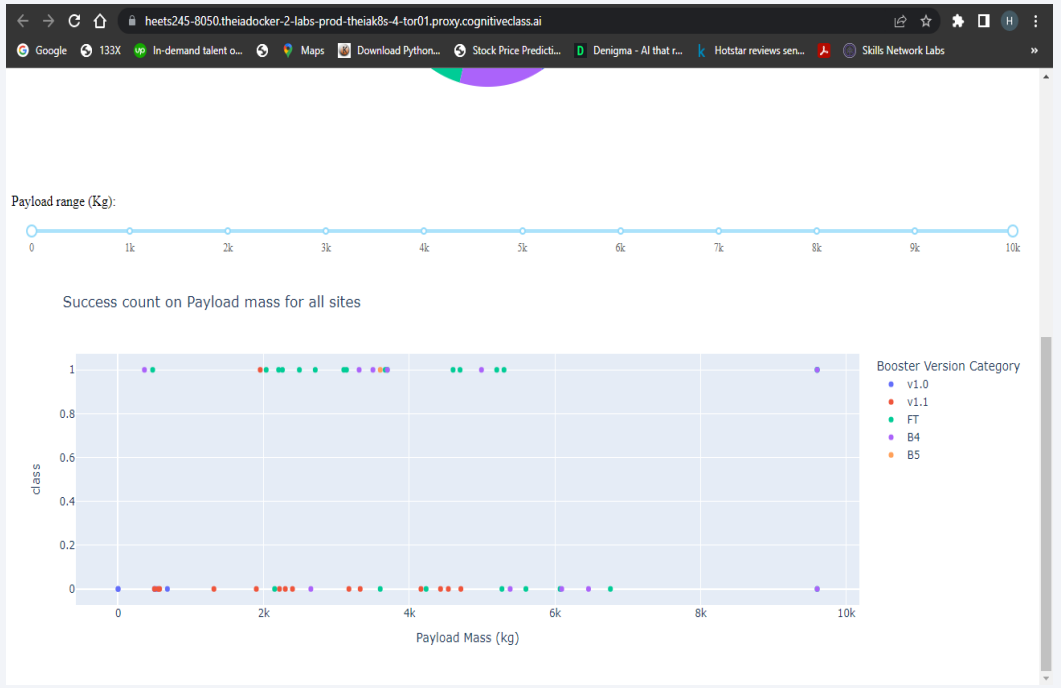
- Github Link : https://github.com/Heet09/Heet09--IBM-Data-Science-Capstone-SpaceX

Build an Interactive Map with Folium

- Github Link : https://github.com/Heet09/Heet09--IBM-Data-Science-Capstone-SpaceX

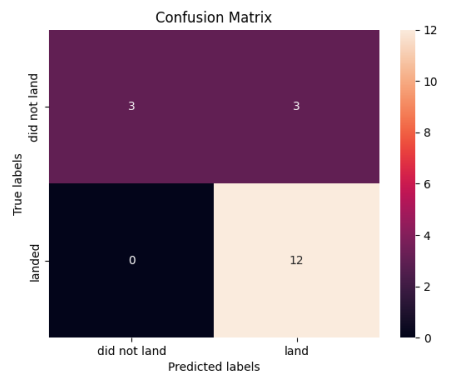
Build a Dashboard with Plotly Dash



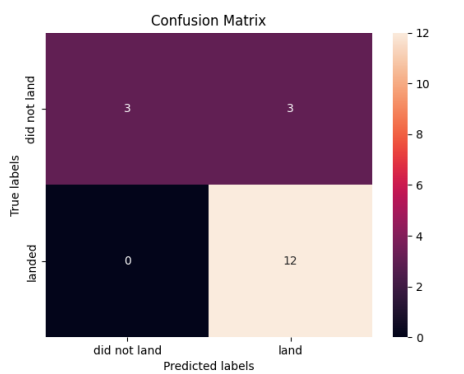


- Github link: https://github.com/Heet09/Heet09--IBM-Data-Science-Capstone-SpaceX

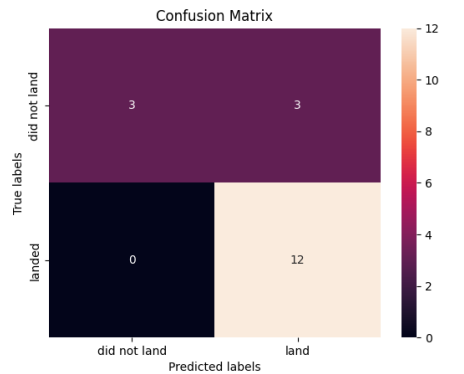
Predictive Analysis (Classification)



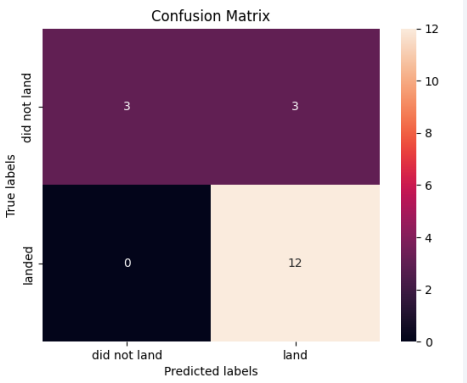
- LR



SVM



Decision Tree

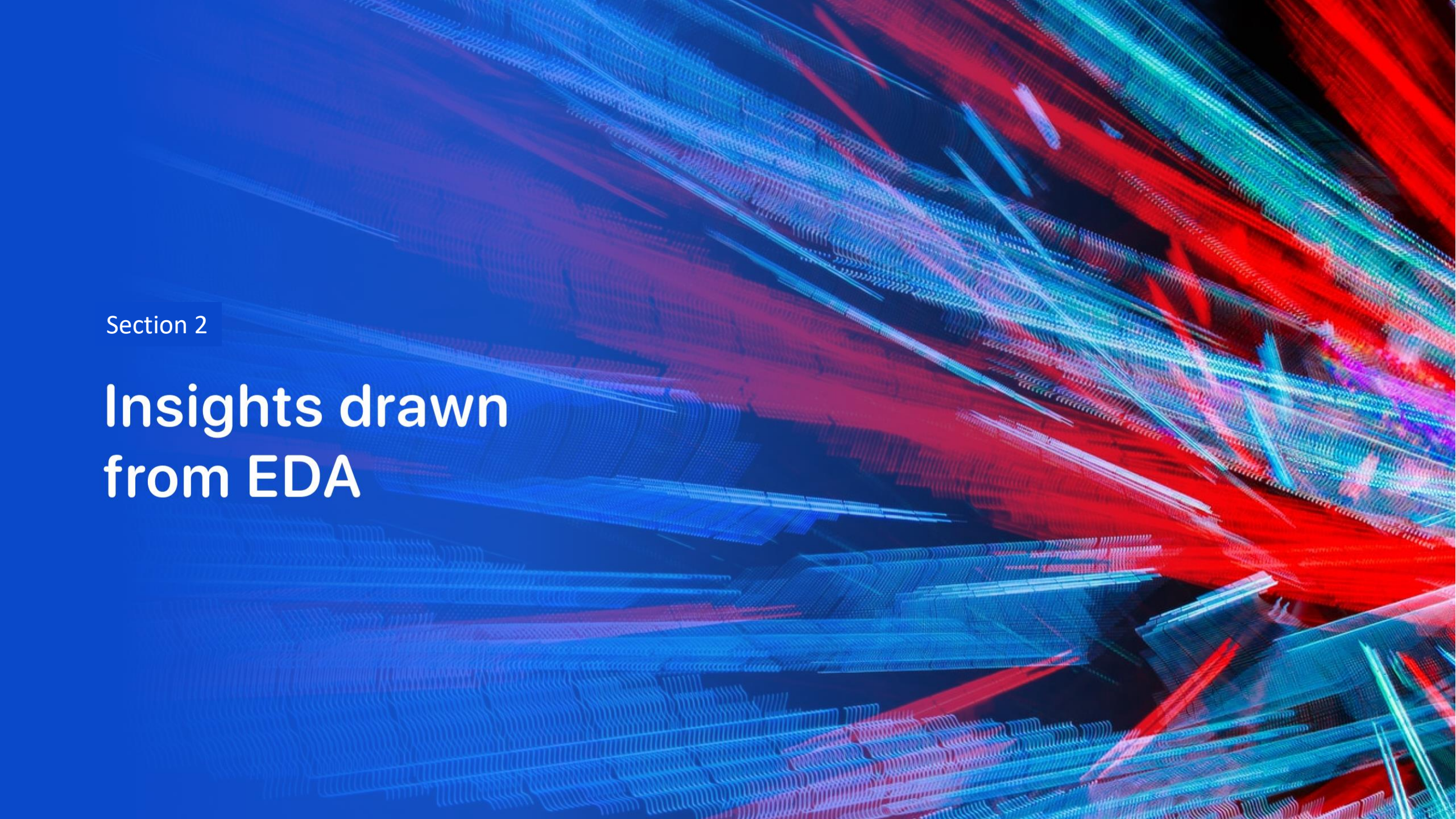


KNN

- The LR, SVM and KNN model achieved the highest accuracy at 83.3% while the SVM performs the best in terms of Area under the curve at 0.958.
- Add the GitHub URL : https://github.com/Heet09/Heet09--IBM-Data-Science-Capstone-SpaceX

Results

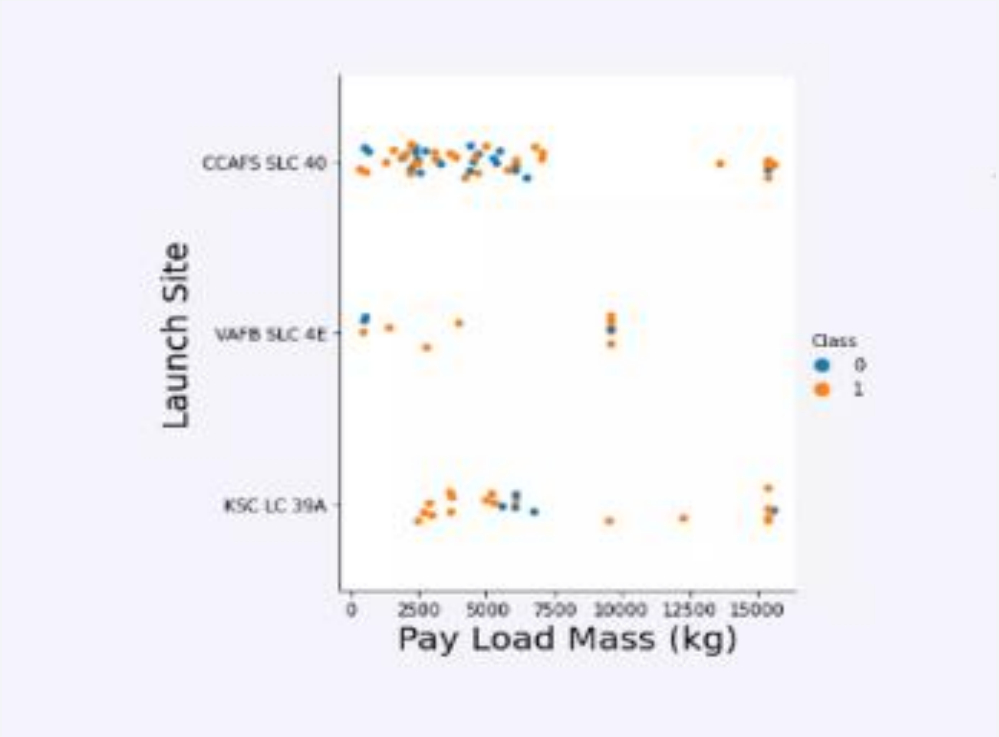
- The SVM, KNN and LR models are the best in terms of prediction accuracy of dataset.
- Low Weighted payloads perform better than heavy ones.
- The success rate of SpaceX launches is directly proportional time in years they will eventually perfect the launches.
- KSC LC 39A has most successful launches from all sites.


Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

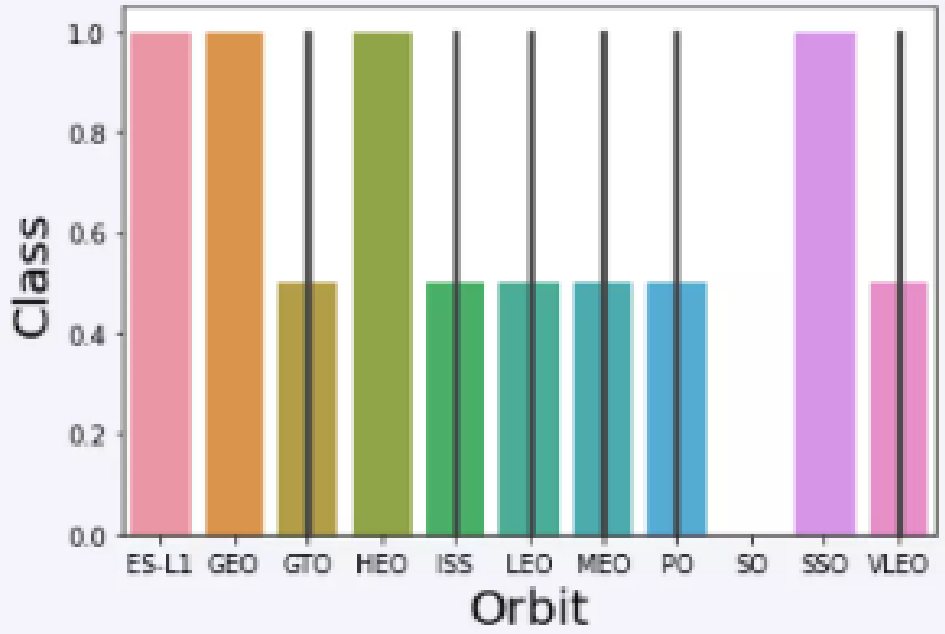
Launches from the site KSC LC-39A are significantly lower than the other launches

Payload vs. Launch Site



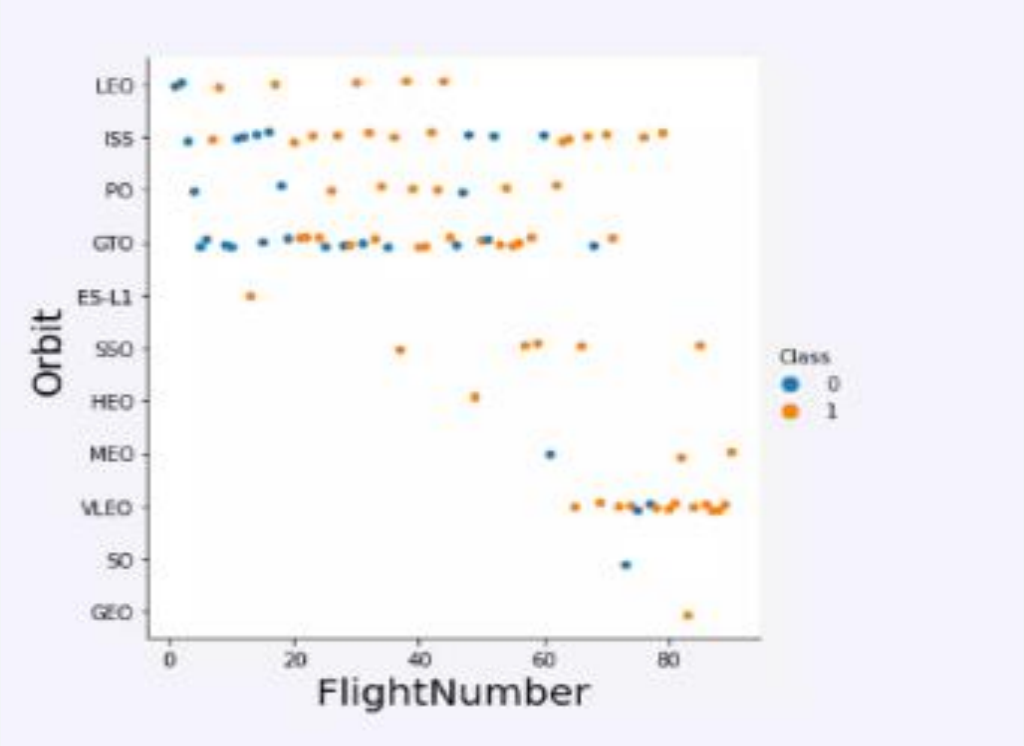
The average IPay loads with lower mass have been launched from VAFB SLC 4E

Success Rate vs. Orbit Type



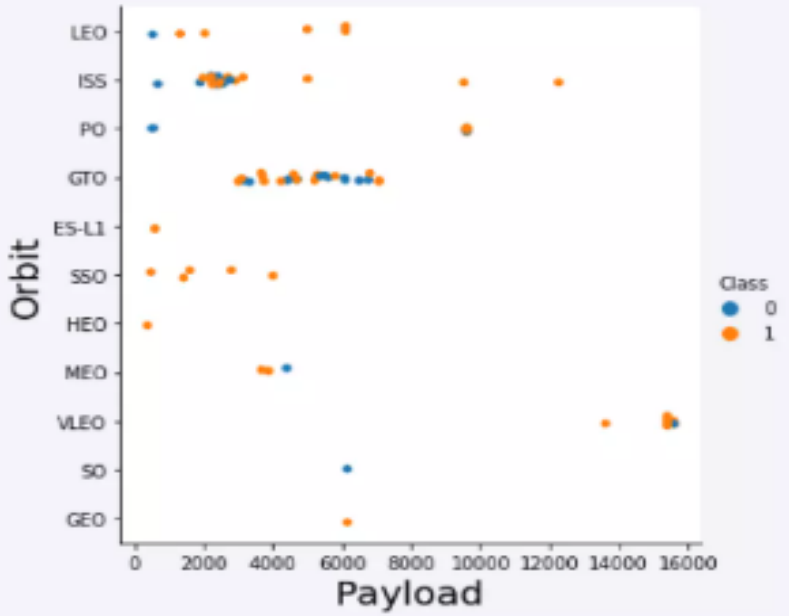
- The orbit types of ES-L1, GEO, HEO, SSO are among the highest success rate.

Flight Number vs. Orbit Type



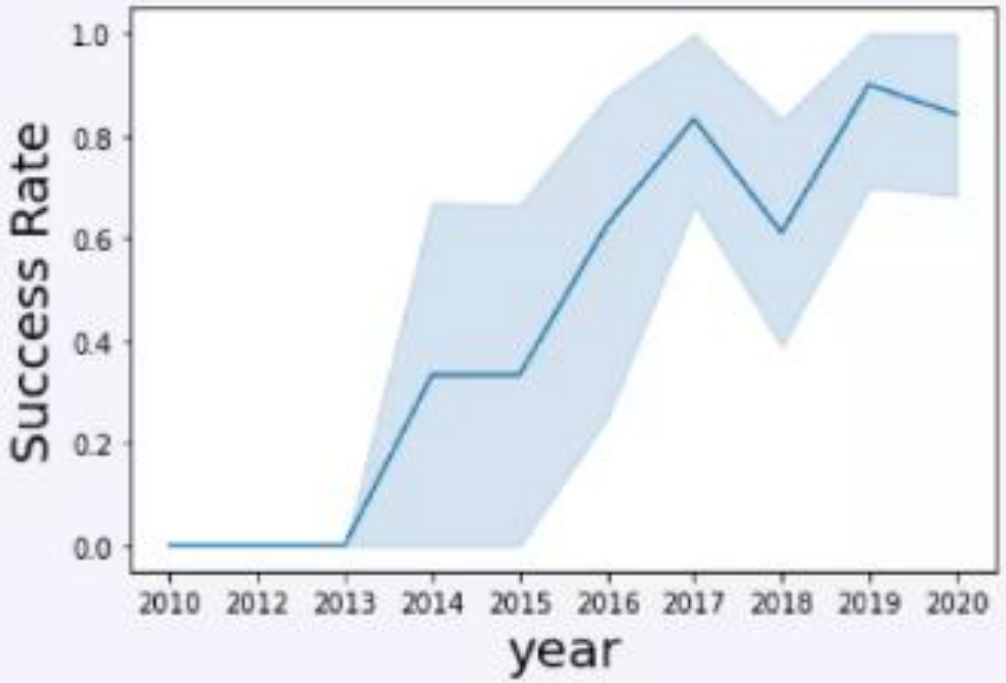
- Earlier LEO, ISS and GTO launches had max flight numbers but in recent years trend shifted to VLEO launches.

Payload vs. Orbit Type



- There are strong correlation between ISS and Payload at the range around 2000, as well as GTO has in range between 2000 to 8000.

Launch Success Yearly Trend



- Launch success rate has increased significantly since 2013 and has stablised since 2019, potentially due to advance in technology and lessons learned.

All Unique Launch Site Names

Display the names of the unique launch sites in the space mission

```
query = "SELECT DISTINCT Launch_Site FROM SPACEXTBL "
cur.execute(query)
for names in cur:
    print(names)
```

```
('CCAFS LC-40',)
('VAFB SLC-4E',)
('KSC LC-39A',)
('CCAFS SLC-40',)
```

Launch Site Names Begin with 'CCA'

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

```
%sql SELECT Launch_Site from SPACEXTBL where (LAUNCH_SITE) LIKE 'CCA%' LIMIT 5;
```

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

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

Total Payload Mass

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

```
%sql SELECT sum(payload_mass__kg_) as sum_payload from SPACEXTBL where (customer) = 'NASA (CRS)'
```

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

sum_payload
45596

Average Payload Mass by F9 v1.1

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

```
%sql SELECT avg(payload_mass__kg_) as average_payload from SPACEXTBL where (booster_version) = 'F9 v1.1'
```

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

average_payload
2928.4

First Successful Ground Landing Date

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

```
%sql SELECT min(date) from SPACEXTBL where LandingOutcome = 'Success (ground pad)'
```

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

min(date)
01-05-2017

Successful Drone Ship Landing with Payload between 4000 and 6000

%sql select BOOSTER_VERSION from SPACEXTBL where LandingOutcome='Success (drone ship)' and PAYLOAD_MASS__KG_ BETWEEN 4001 and 5999

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
%sql select BOOSTER_VERSION from SPACEXTBL where LandingOutcome='Success (drone ship)' and PAYLOAD_MASS__KG_ BETWEEN 40
```

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

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

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

```
%sql SELECT MISSION_OUTCOME, COUNT(MISSION_OUTCOME) AS OUTCOME FROM SPACEXTBL GROUP BY MISSION_OUTCOME
```

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

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

Boosters Carried Maximum Payload

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

```
%sql SELECT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL)
```

```
 * 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

2015 Launch Records

%sql SELECT DATE, BOOSTER_VERSION, LAUNCH_SITE, LandingOutcome FROM SPACEXTBL WHERE LandingOutcome = 'Failure (drone ship)' and substr(Date,7,4)='2015'

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 4, 2) as month to get the months and substr(Date,7,4)='2015' for year.

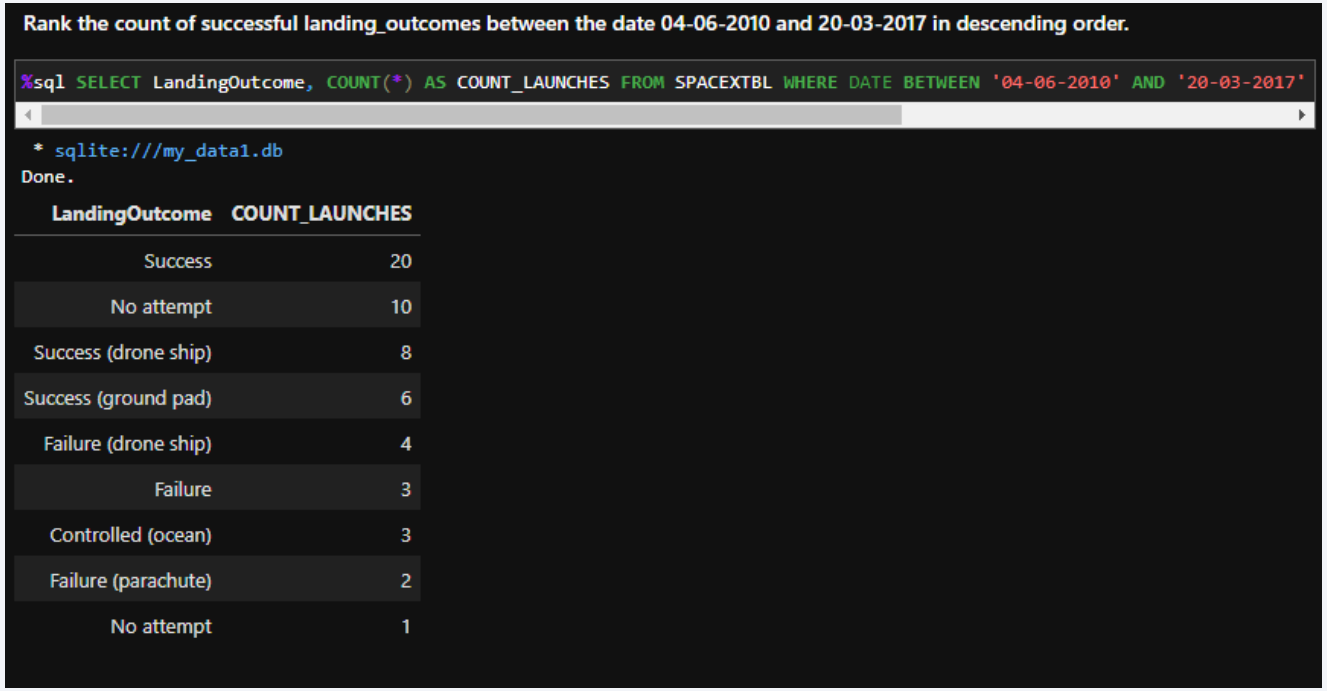
```
%sql SELECT DATE, BOOSTER_VERSION, LAUNCH_SITE, LandingOutcome FROM SPACEXTBL WHERE LandingOutcome = 'Failure (drone sh
```

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

Date	Booster_Version	Launch_Site	LandingOutcome
10-01-2015	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
14-04-2015	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

```
%sql SELECT LandingOutcome, COUNT(*) AS
COUNT_LAUNCHES FROM SPACEXTBL WHERE DATE BETWEEN
'04-06-2010' AND '20-03-2017' GROUP BY LandingOutcome
ORDER BY COUNT_LAUNCHES DESC;
```



Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

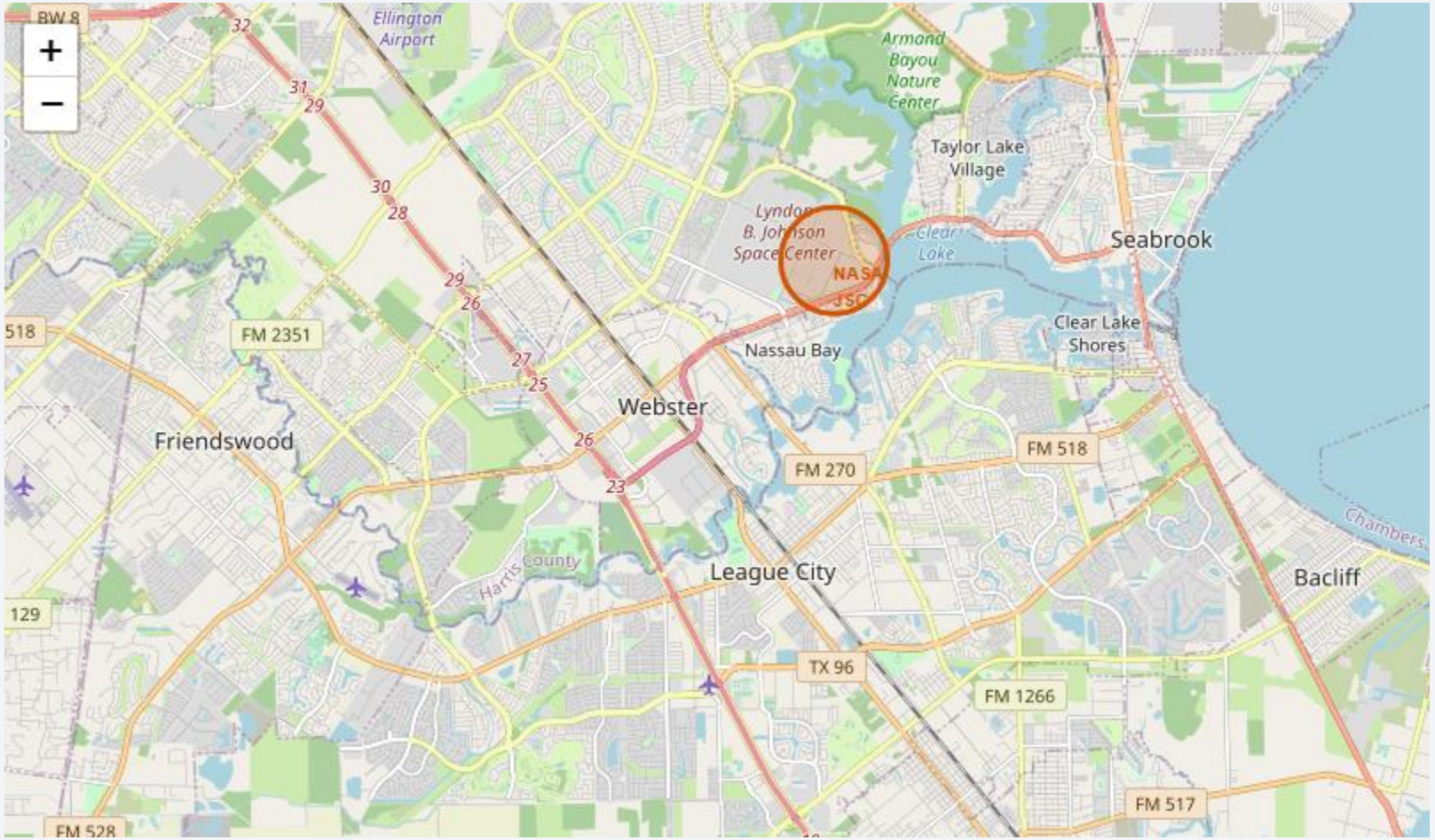
```
%sql SELECT LandingOutcome, COUNT(*) AS COUNT_LAUNCHES FROM SPACEXTBL WHERE DATE BETWEEN '04-06-2010' AND '20-03-2017'
 * sqlite:///my_data1.db
Done.
```

LandingOutcome	COUNT_LAUNCHES
Success	20
No attempt	10
Success (drone ship)	8
Success (ground pad)	6
Failure (drone ship)	4
Failure	3
Controlled (ocean)	3
Failure (parachute)	2
No attempt	1

Section 3

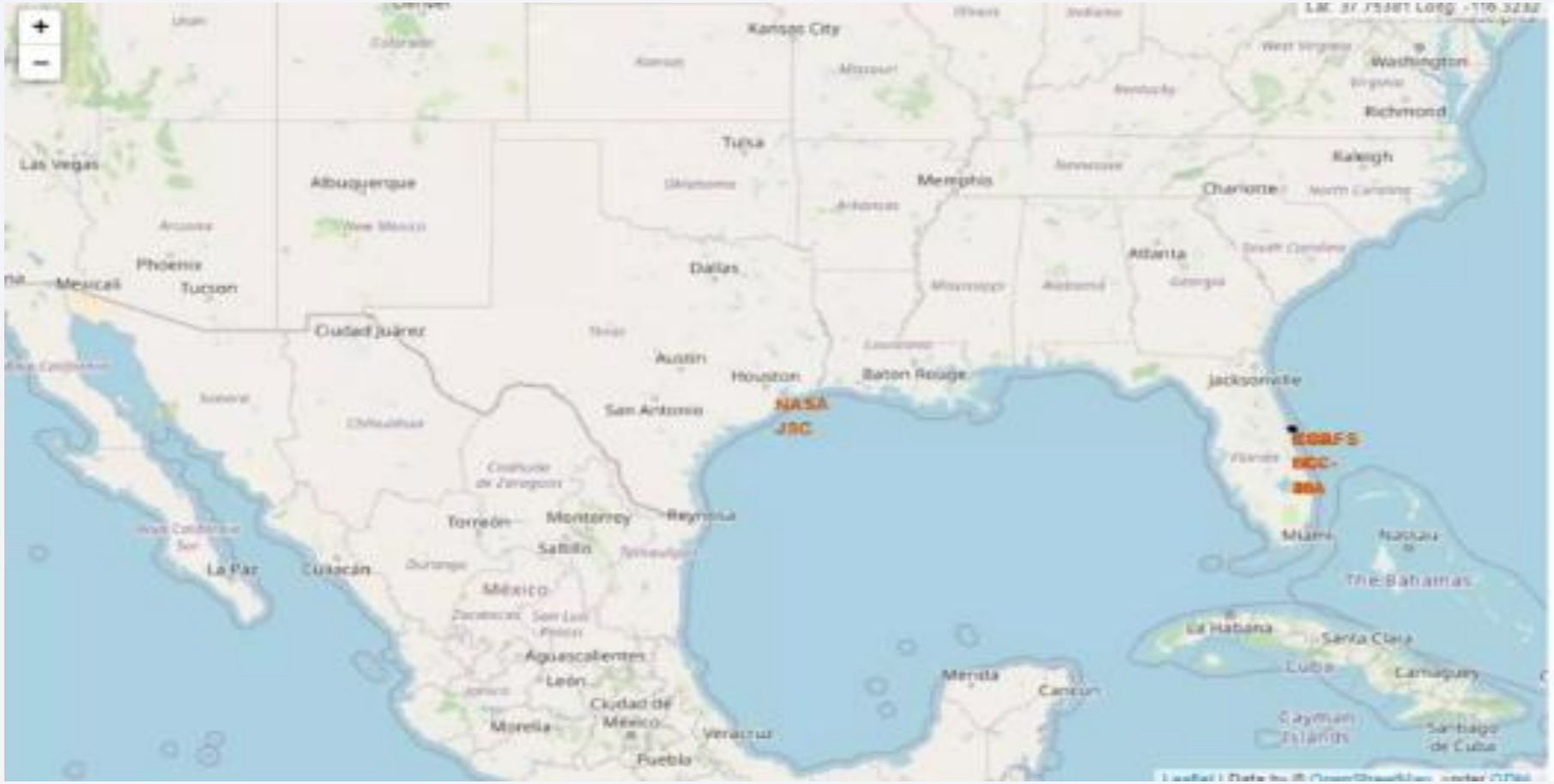
Launch Sites Proximities Analysis

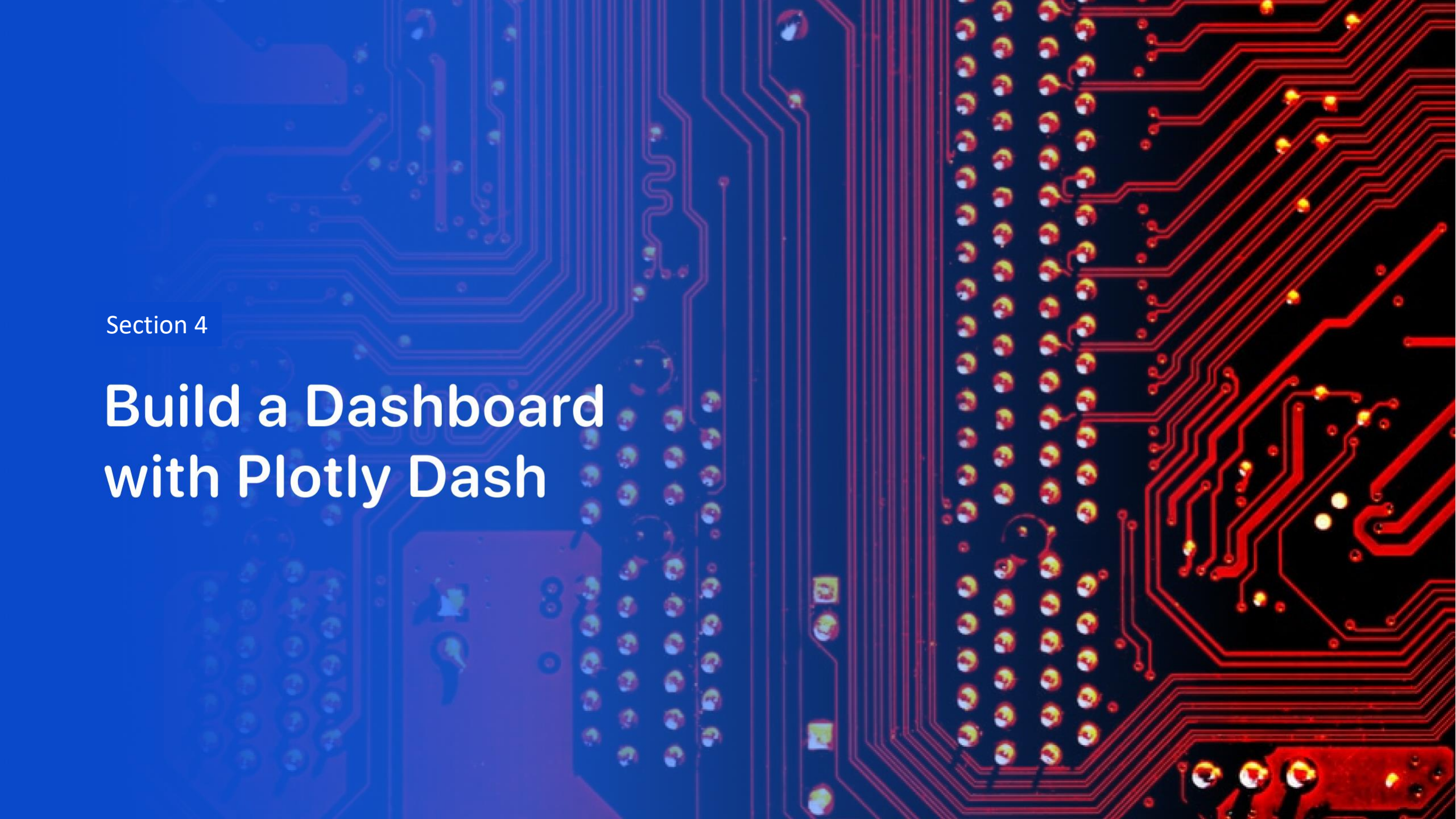
All launch sites marked on map



Success/failed launches marked on map

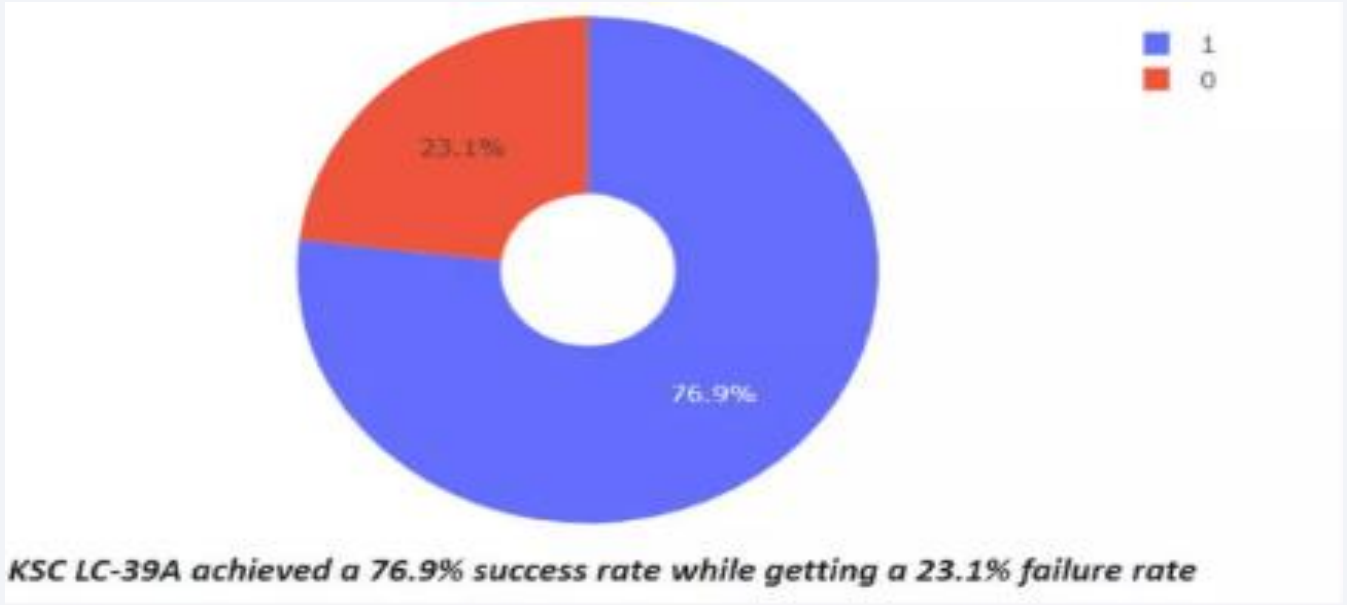
Distance between launch sites and its proximities




Section 4
Build a Dashboard with Plotly Dash

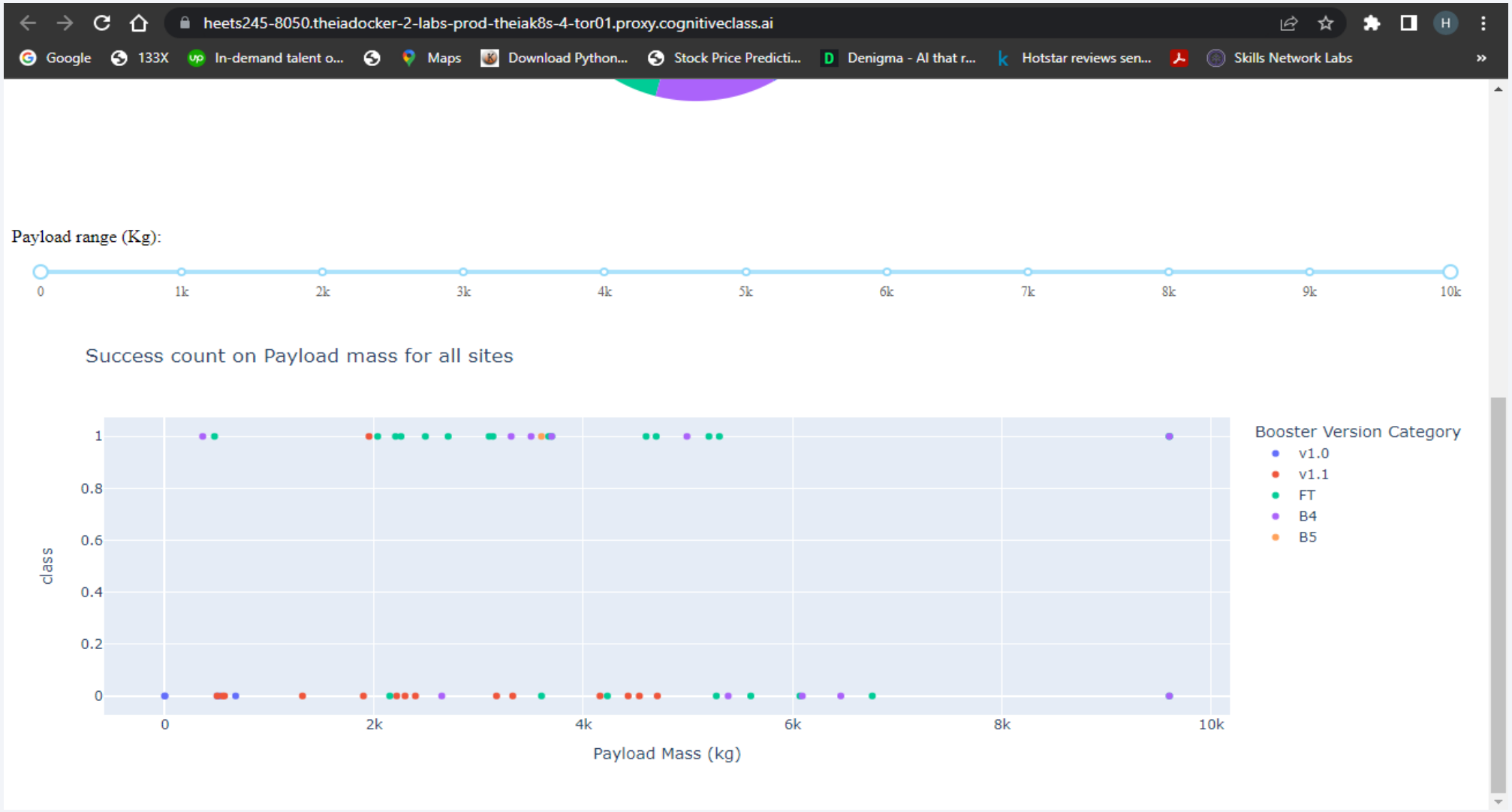
Total success launches by all sites

Success rate by sites



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

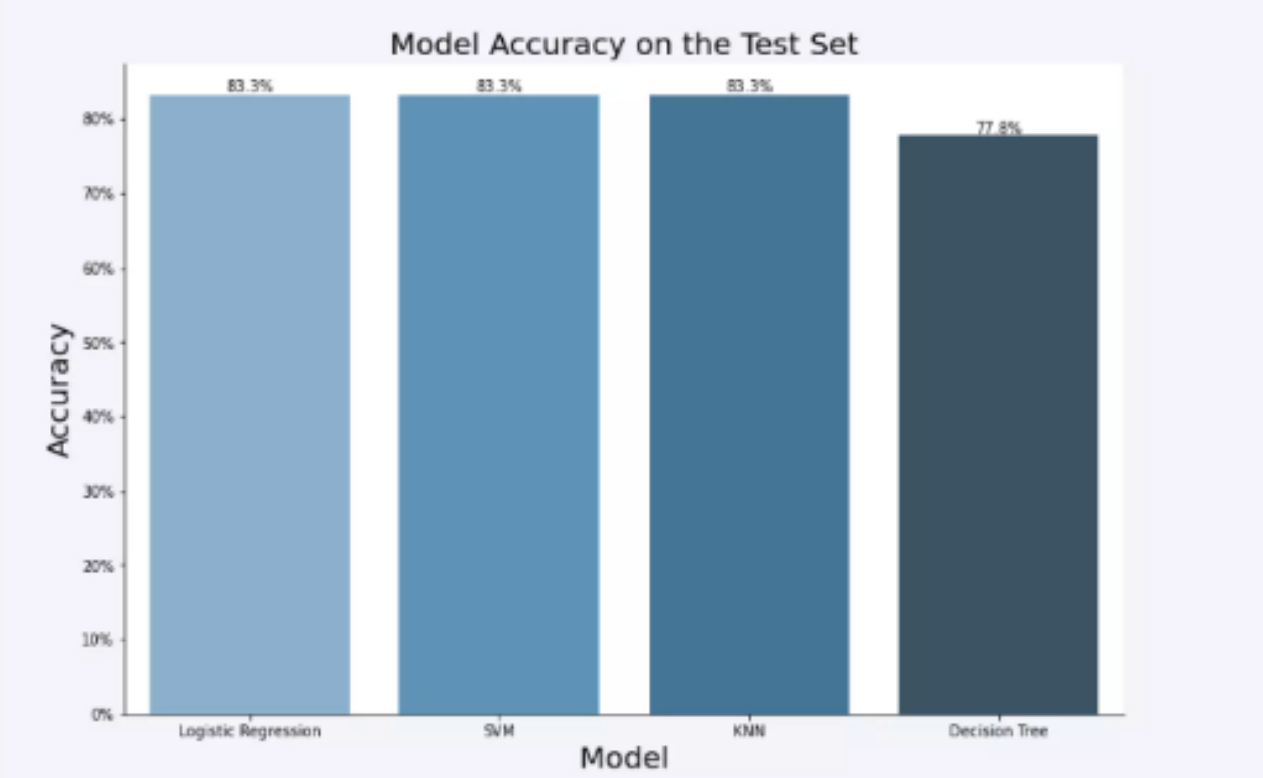
Success count on Payload mass of all sites



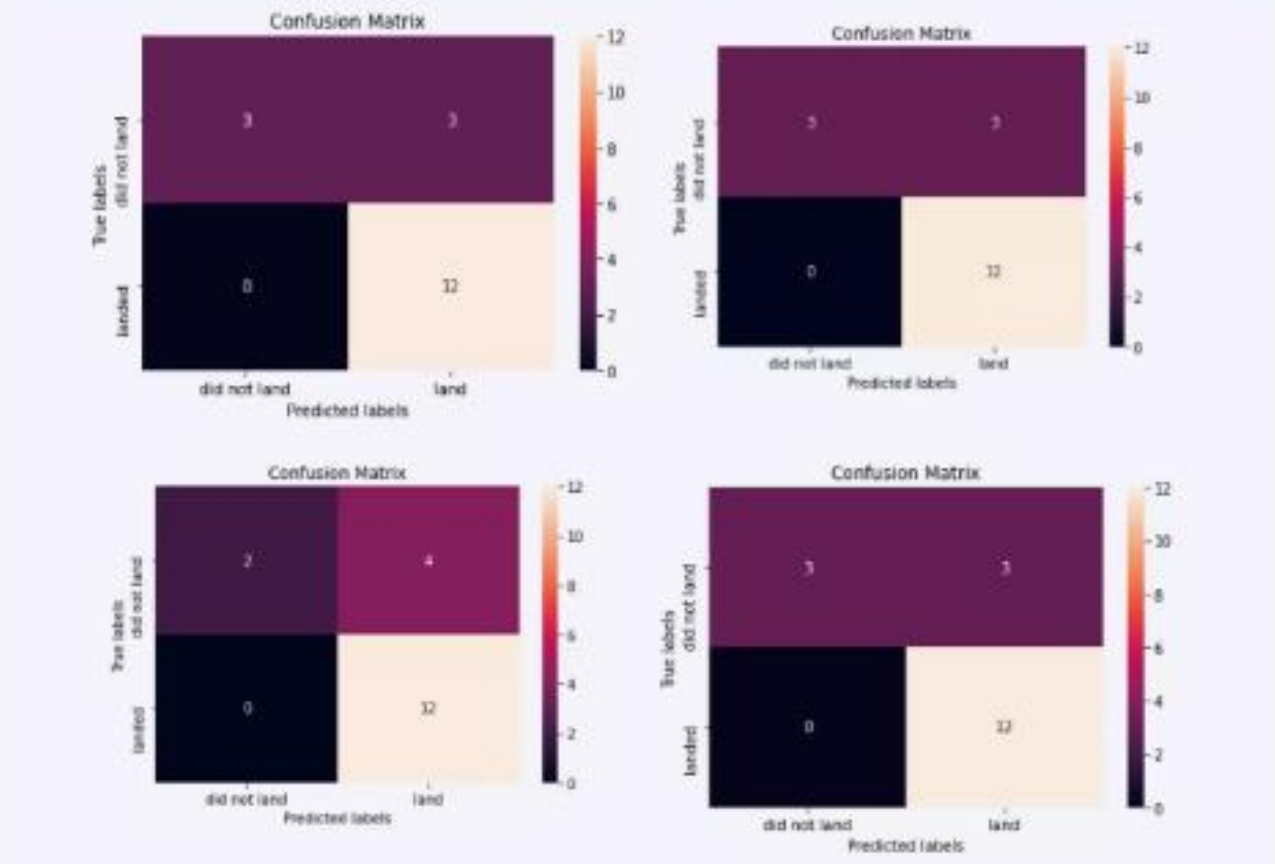
Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix



Conclusions

- The SVM, KNN, and Logistic Regression models are the best in terms of prediction accuracy for this dataset.
- Low weighted payloads perform better than the heavier payloads.
- The success rates for SpaceX launches is directly proportional time in years they will eventually perfect the launches.
- KSC LC 39A had the most successful launches from all the sites.
- Orbit GEO,HEO,SSO,ES L1 has the best Success Rate.

Thank you!