IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Samar Saad
July 2023

Outline

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

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 advertises Falcon 9 rocket launches on its website, with a cost of 62 million dollars other providers cost upward o 165 million dollars each. much ot the savinas Is because Space can reuse the first stage.

- Problems you want to find answers

 The project task is to predicting if the first stage of the Space Falcon 9 rocket will land successtully

Section 1

Methodology

Methodology

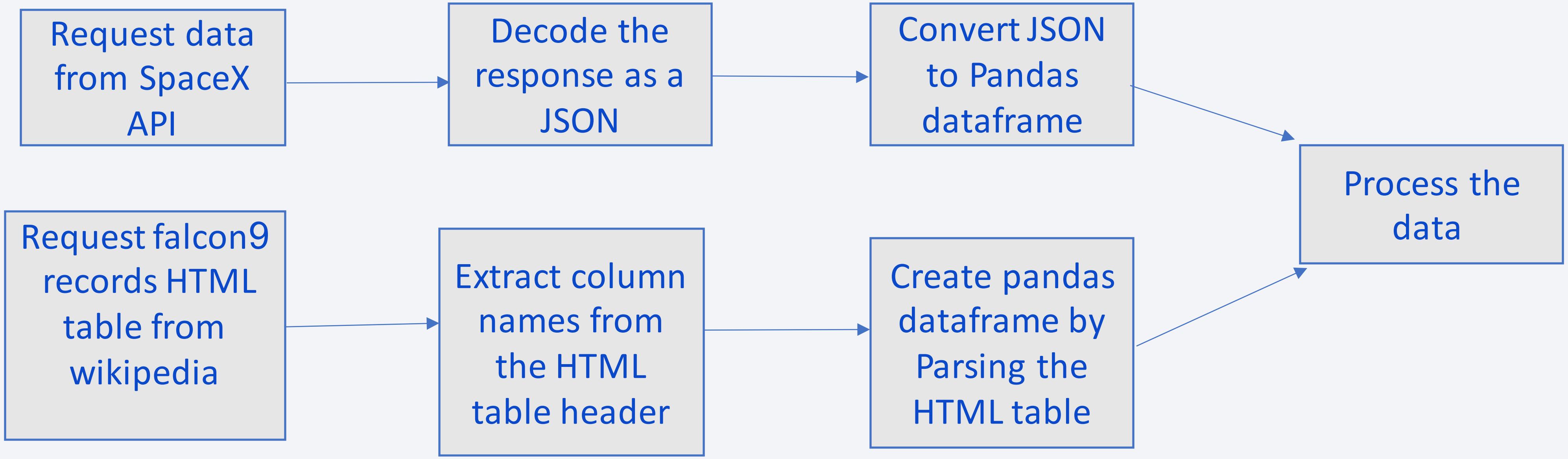
Executive Summary

- Data collection methodology:
  - SpaceX Rest API
  - Web Scraping from Wikipedia
- Perform data wrangling
  - Data cleaning and 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
  - Build and evaluate LR , KNN , SVM and DT

Data Collection

Data sets were collected from :

- SpaceX REST API
- Wikipedia via web scraping using BeautifulSoup



Data Collection – SpaceX API

1. Request data from SpaceX API
2. Decode the response as a JSON
3. Convert JSON to Pandas datafra me

https://github.com/SamarSaad056/Data-Sience-Project-2/blob/main/data-collection-api.ipynb

1

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

2

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

3

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

Data Collection - Scraping

1. Request falcon9 records HTM L table from wikipedia
2. Get the response
3. Create a BeautifulSoup object from a response text content
4. Create pandas dataframe by parsing the HTML table

1

```
static_url = "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922"
```

2

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

3

```
soup = BeautifulSoup(response.text, 'html.parser')
```

4

```
df = pd.DataFrame.from_dict(launch_dict, orient='index')
df = df.transpose()
df.head()
```

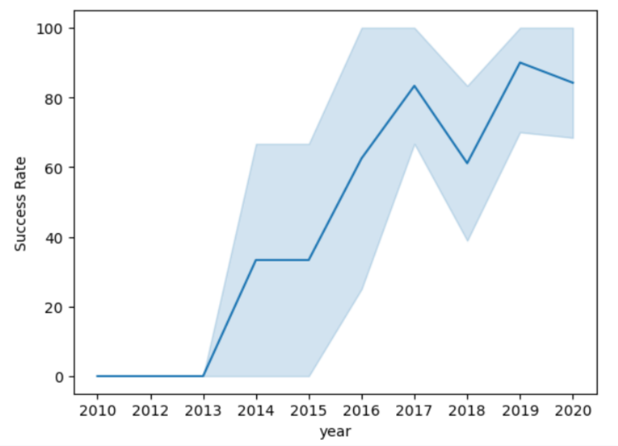
https://github.com/SamarSaad-tcejorP-ecneiS-ataD/056
bnypi.gniparcsbew_noticelloc_atad/niam/bolb/2

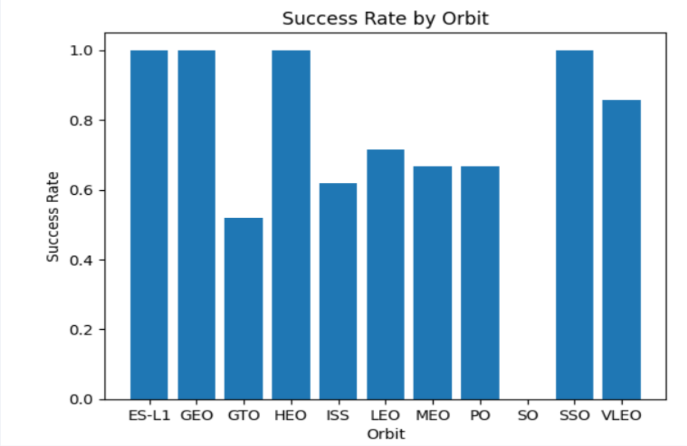
Data Wrangling

- Identify and calculate the percentage of the missing values in each attribute
- Calculate the number of launches on each site
- Calculate the number and occurrence of each orbit
- Calculate the number and occurence of mission outcome per orbit type
- Create a landing outcome label from Outcome column

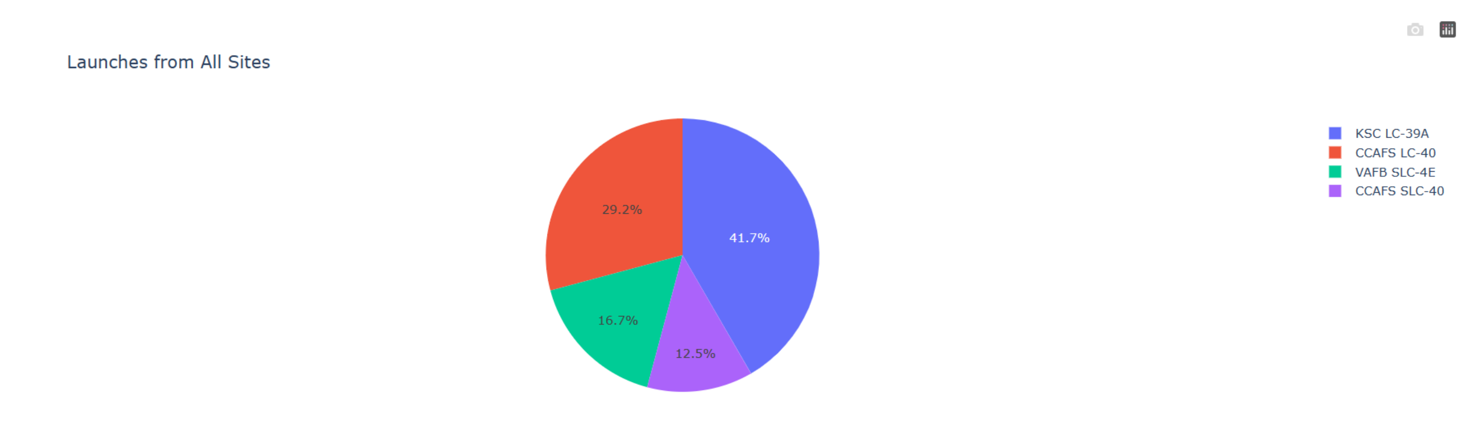
https://github.com/SamarSaad056/Data-Sience-Project-2/blob/main/data_wrangling.ipynb

EDA with Data Visualization





https://github.com/SamarSaad-ecneiS-ataD/056
bnypi.siVataD_ADE/niam/bolb/2-tcejorP

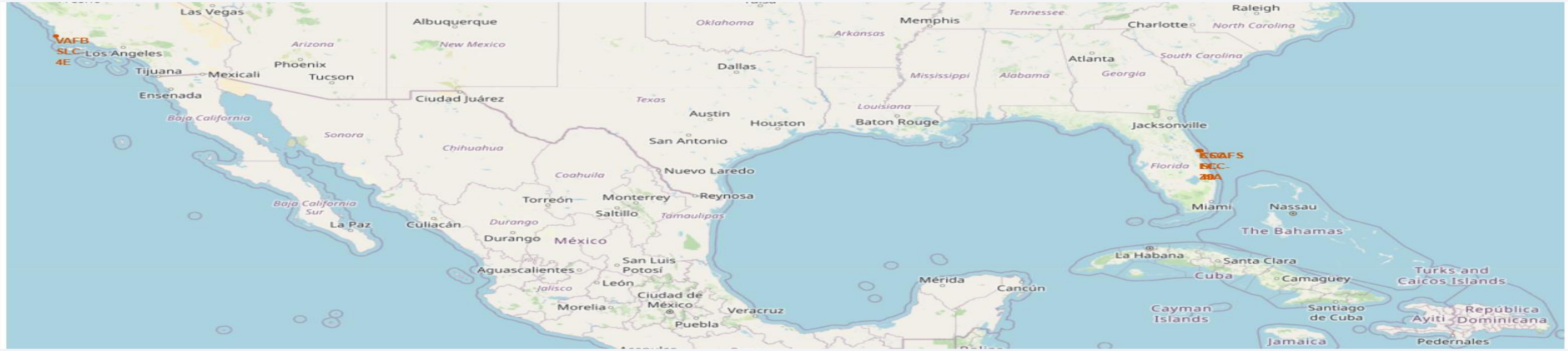


EDA with SQL

- `select distinct "Launch_Site" from SPACEXTBL`
- `select * from SPACEXTBL where Launch_Site like 'CCA%' limit 5`
- `select sum(PAYLOAD_MASS__KG_)as 'Total' from SPACEXTBL where Customer='NASA (CRS)'`
- `select avg(PAYLOAD_MASS__KG_) as 'average' from SPACEXTBL where Booster_Version like 'F9 v1.1%'`
- `select min(date) from SPACEXTBL where mission_outcome like 'Success'`
- `select Booster_Version from SPACEXTBL where landing_outcome = 'Success (drone ship)' and PAYLOAD_MASS__KG_> 4000 and PAYLOAD_MASS__KG_ <6000`
- `select Mission_Outcome , count(*) as total from SPACEXTBL group by Mission_Outcome`

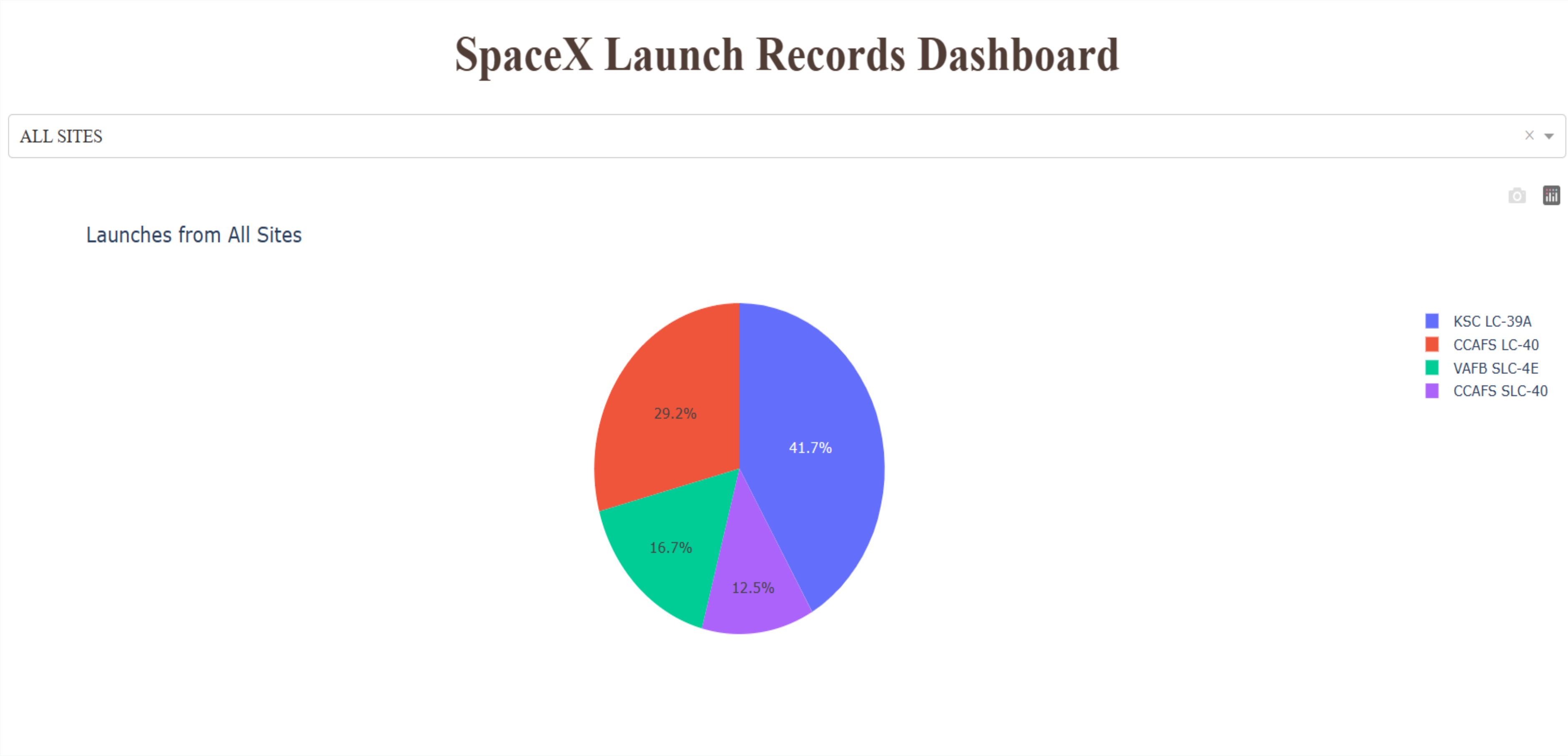
https://github.com/SamarSaad-tcejorP-ecneiS-ataD/056
bnypi.lqS_ADE/niam/bolb/2

Build an Interactive Map with Folium



https://github.com/SamarSaad-tcejorP-ecneiS-ataD/056
bnypi..notiacol_etis_hcnual/niam/bolb/2

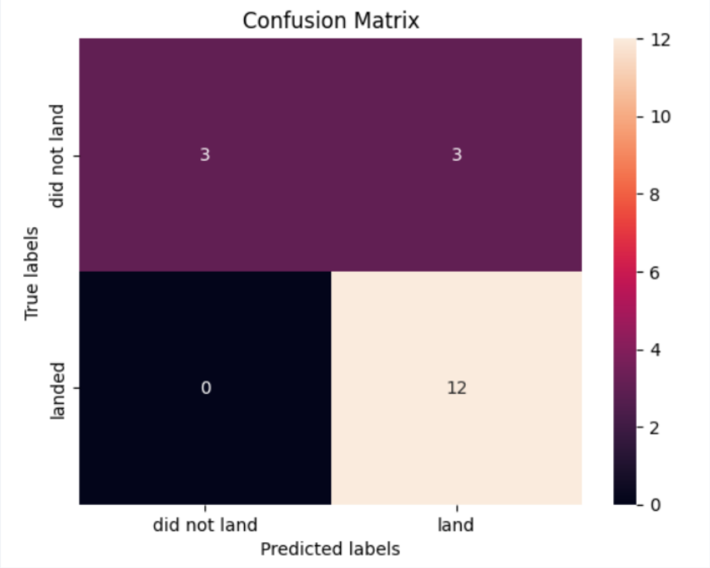
Build a Dashboard with Plotly Dash

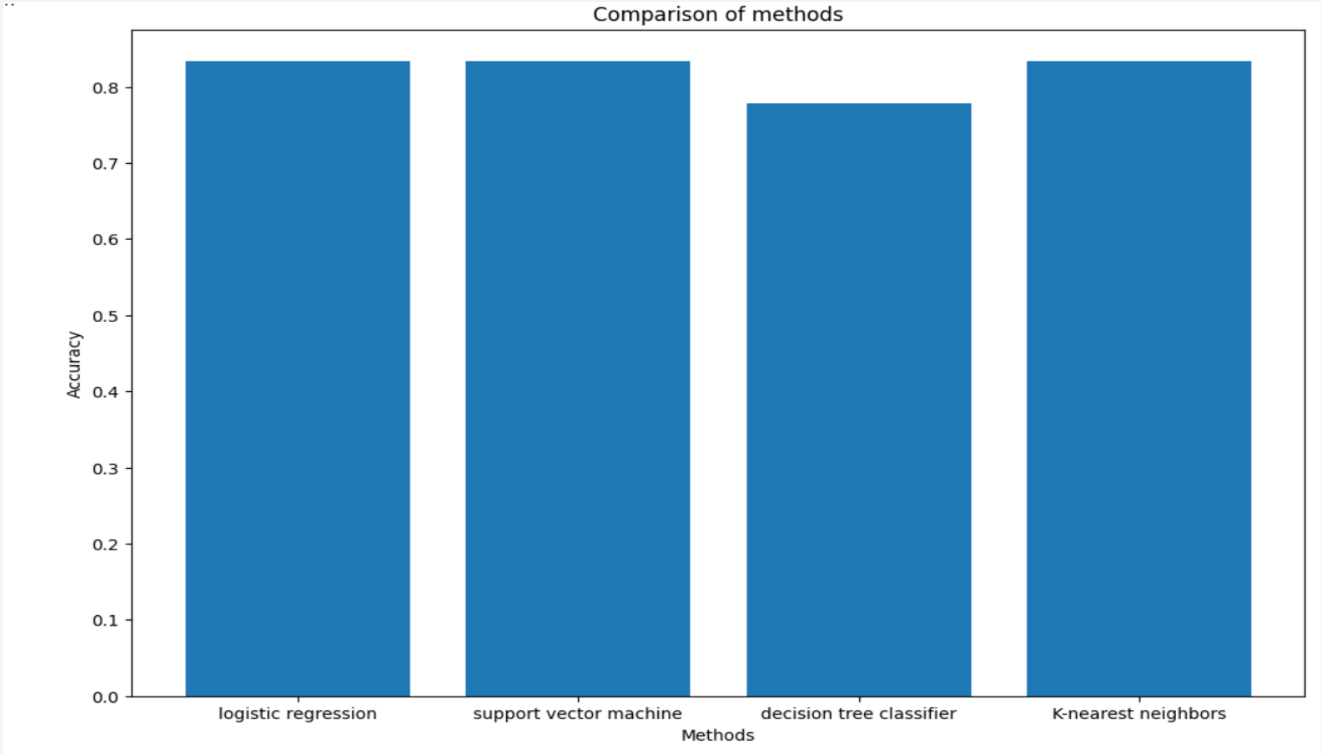


https://github.com/SamarSaadyp.MBIbew/niam/bolb/2-tcejorP-ecneiS-ataD/056

Predictive Analysis (Classification)

Logistic Regression , SVM , and KNN achieved the highest accuracy at 83.3%





https://github.com/SamarSaad-ecneiS-ataD/056
-tcejorP
bnypi.noticiderP_gninraeL_enihcaM/niam/bolb/2

Results

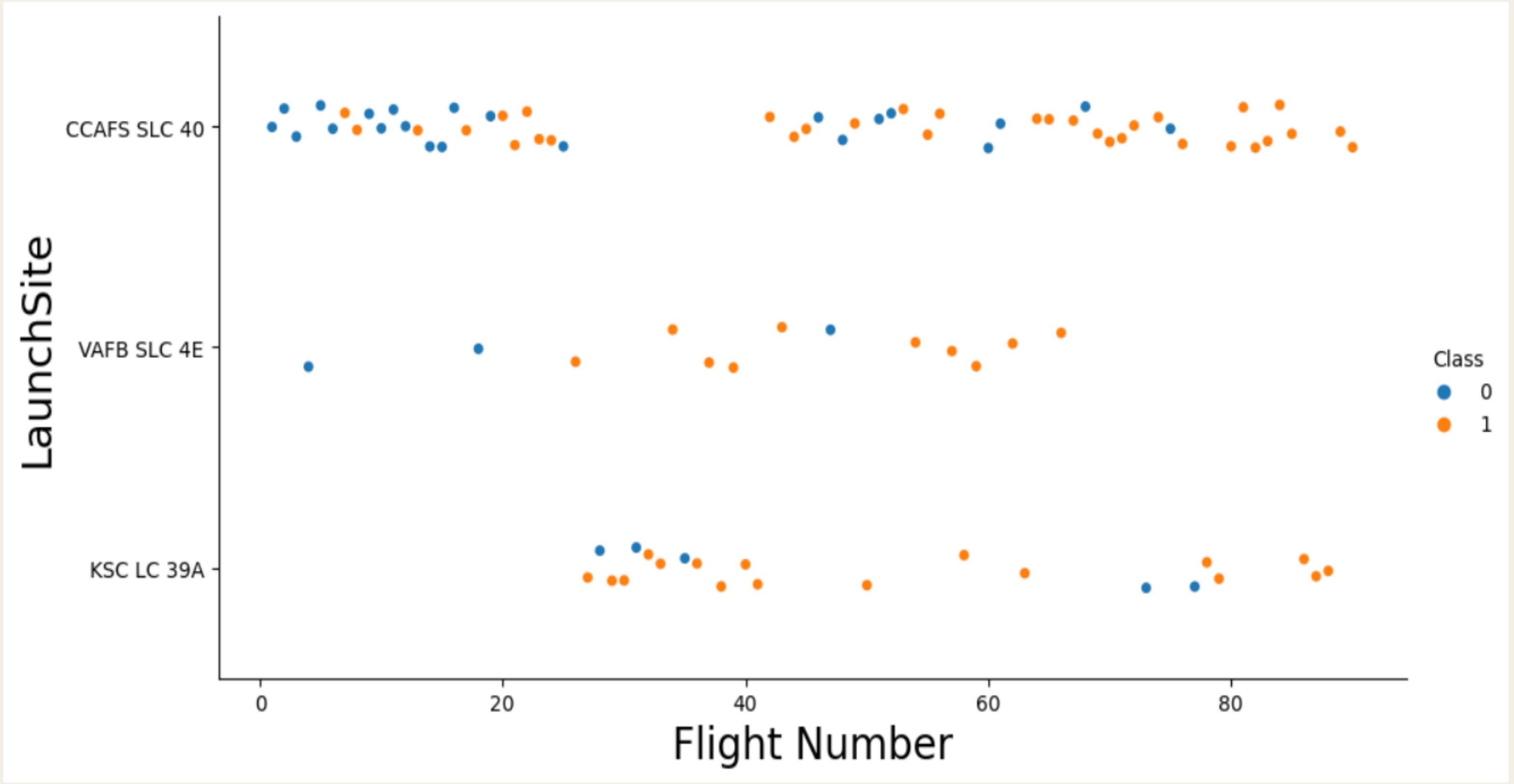
- The SVM, KNN, and Logistic Regression models are the best in terms of prediction accuracy for this dataset.
- Low weighted payloads perform better than the heavier payloads.
- The success rates for Space launches is directly proportional time in years they will eventually perfect the launches.
- KSC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

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

CCAFS SLC 40 has higher launches than other sites

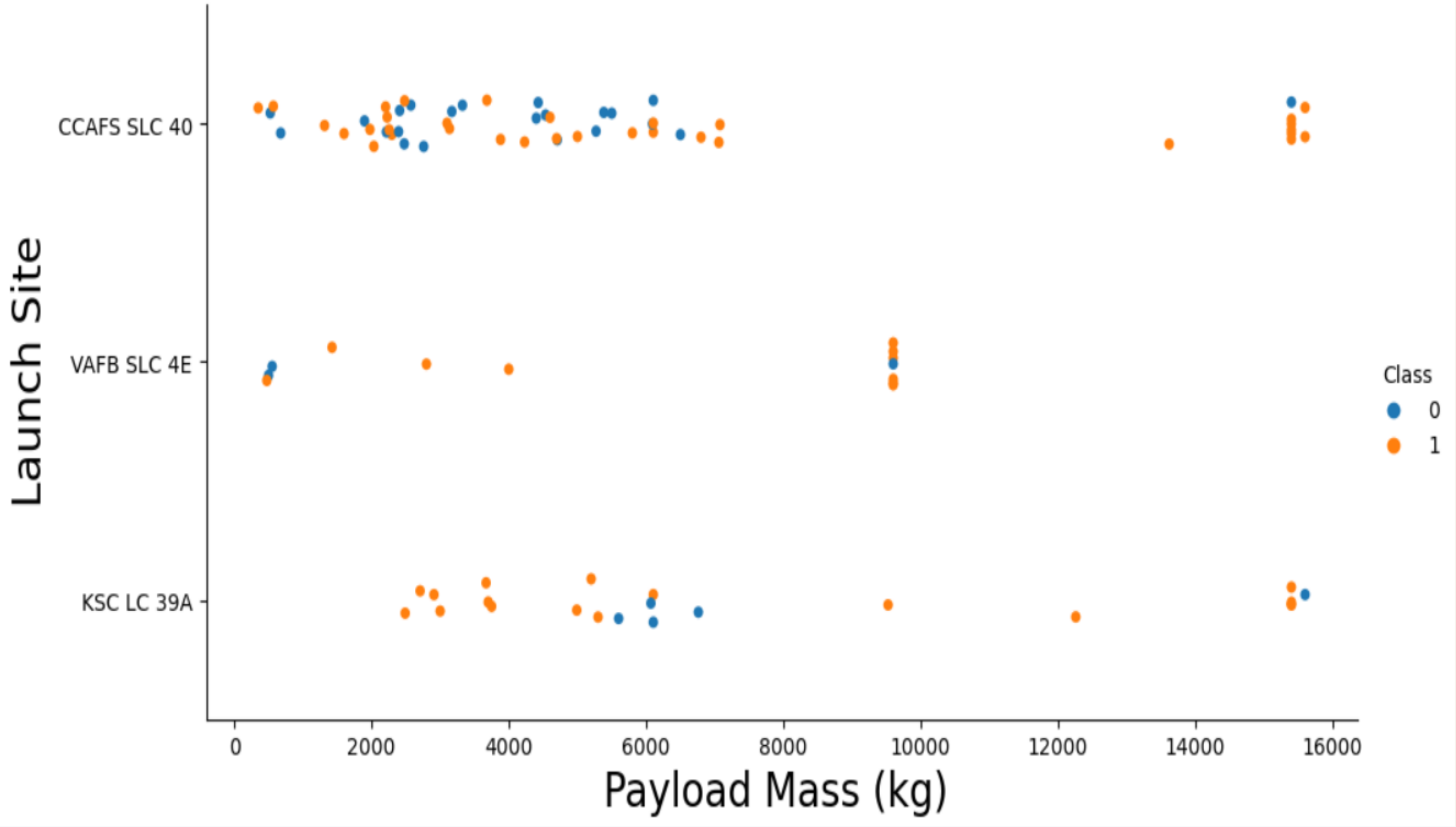


Payload vs. Launch Site

Most of the payload with

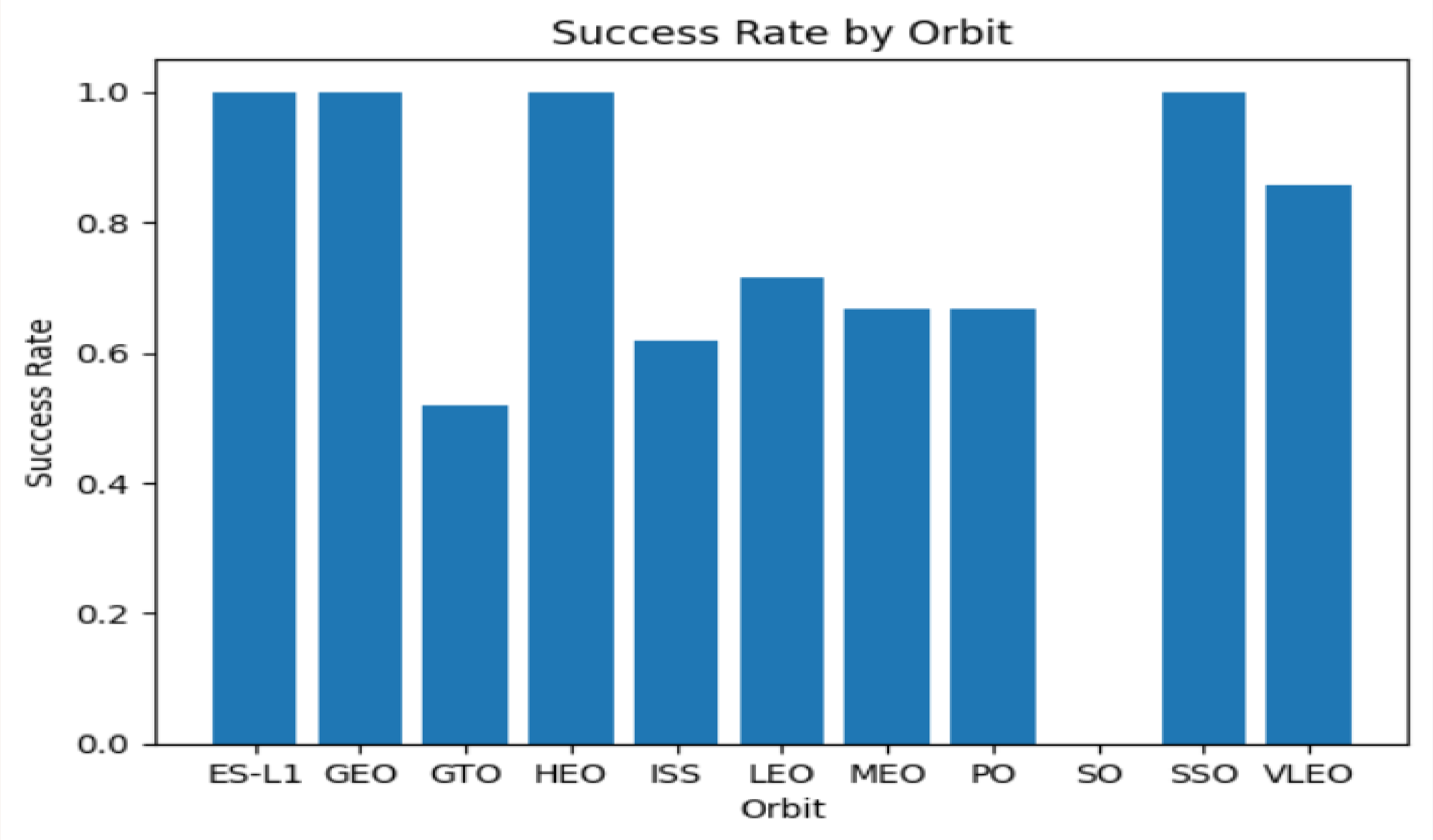
low mass have been launched

From CCAFS SLC 40 site .



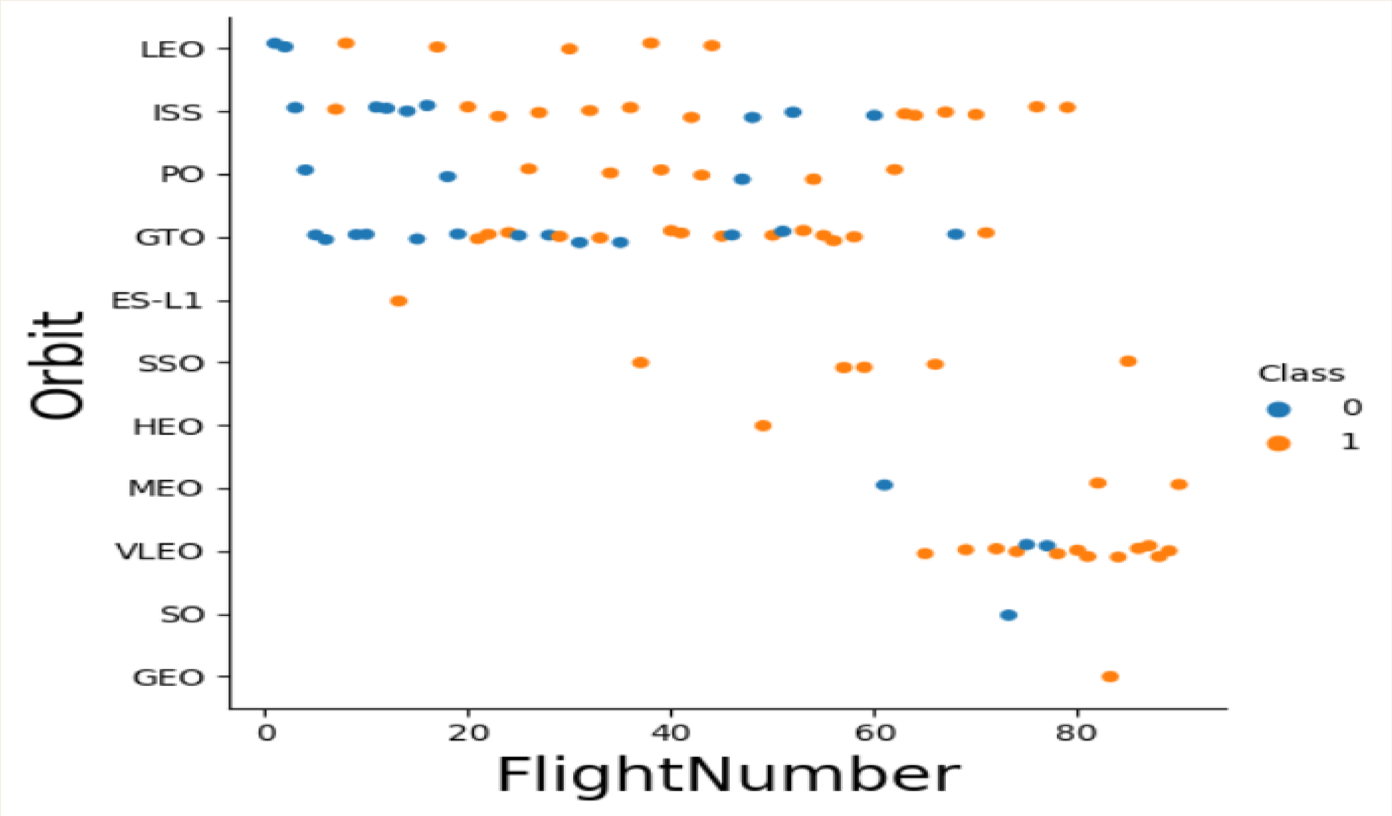
Success Rate vs. Orbit Type

ES-L1 , GEO , HEO and SSO orbits have the higher success rate .



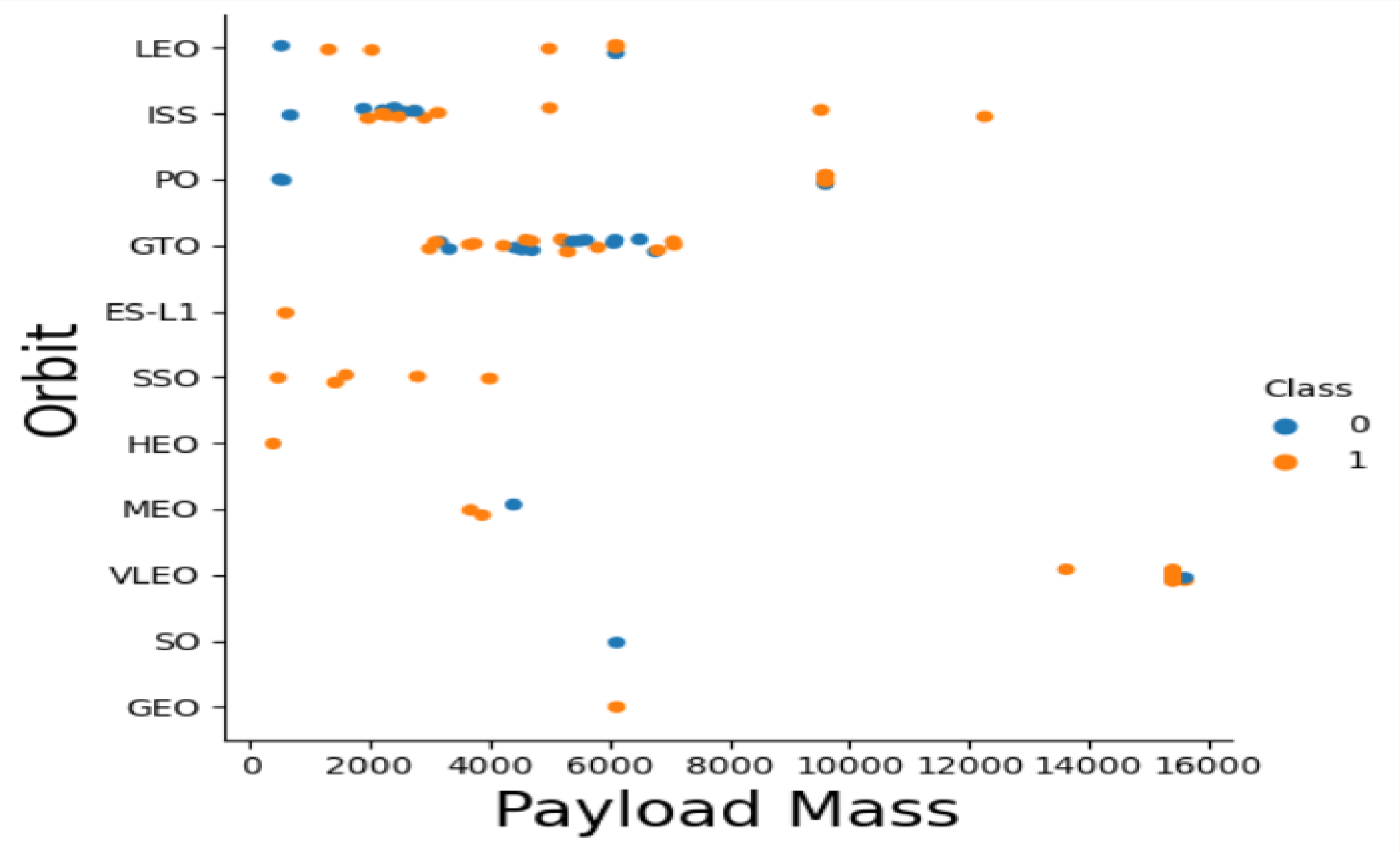
Flight Number vs. Orbit Type

in the LEO orbit the Success appears related to the number of flights



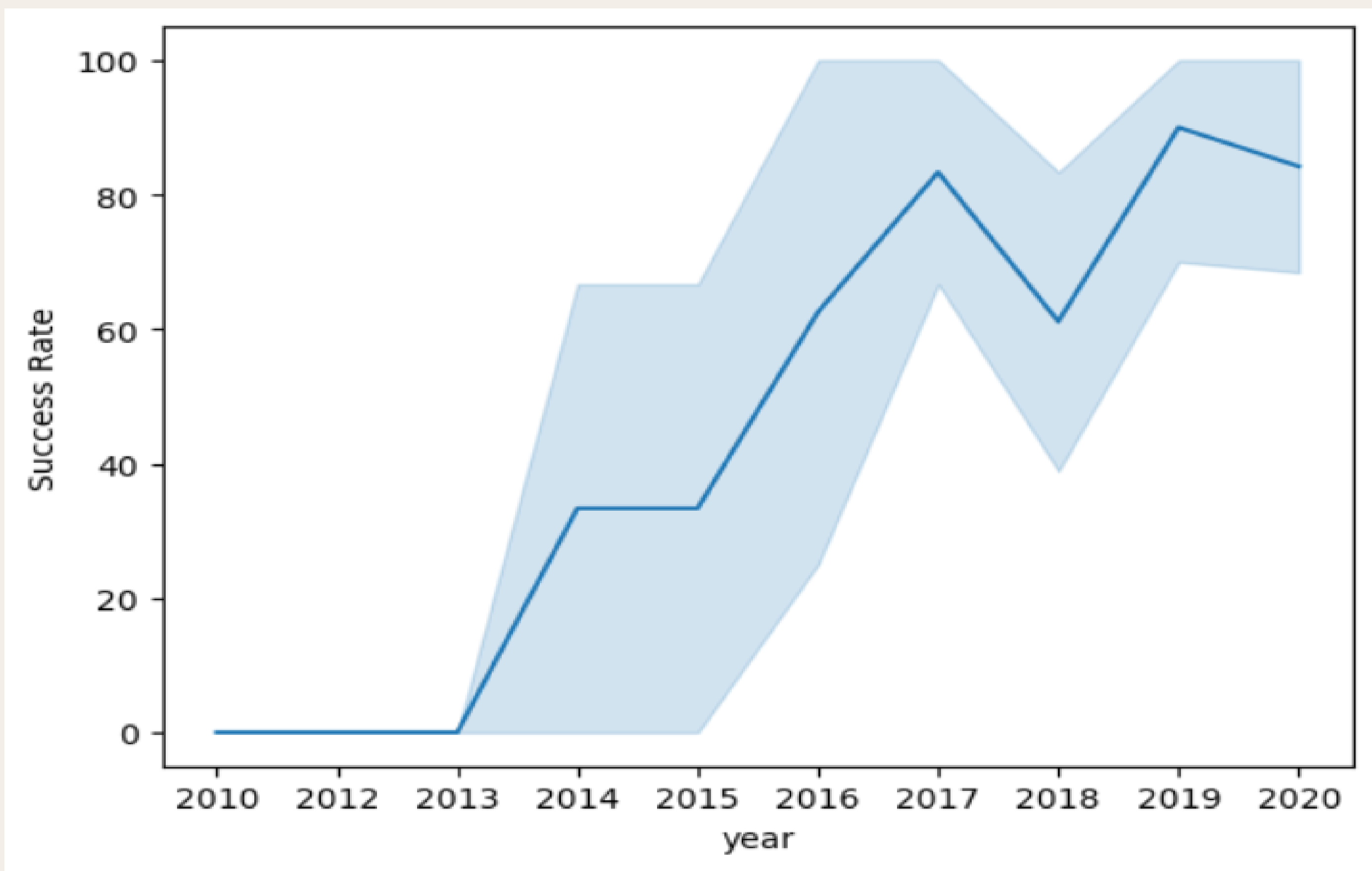
Payload vs. Orbit Type

With heavy payloads the successful landing rate are more for Po ,LEO and ISS.



Launch Success Yearly Trend

Success rate since 2013 kept increasing till 2020



All Launch Site Names

```
%sql select distinct "Launch_Site" from SPACEXTBL
```

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

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40
None

Launch Site Names Begin with 'CCA'

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

Python

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
06/04/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0.0	LEO	SpaceX	Success	Failure (parachute)
12/08/2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0.0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
22/05/2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525.0	LEO (ISS)	NASA (COTS)	Success	No attempt
10/08/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500.0	LEO (ISS)	NASA (CRS)	Success	No attempt
03/01/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677.0	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

```
%sql select sum(PAYLOAD_MASS__KG_)as 'Total' from SPACEXTBL where Customer='NASA (CRS)'
```

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

Total
45596.0

Average Payload Mass by F9 v1.1

```
%sql select avg(PAYLOAD_MASS__KG_) as 'average' from SPACEXTBL where Booster_Version like 'F9 v1.1%'
```

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

average
2534.6666666666665

First Successful Ground Landing Date

```
%sql select min(date) from SPACEXTBL where mission_outcome like 'Success'
```

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

min(date)
01/06/2014

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql select Booster_Version from SPACEXTBL where landing_outcome = 'Success (drone ship)' and PAYLOAD_MASS__KG_> 4000 and PAYLOAD_MASS__KG_ <6000
```

```
 * 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

```
%sql select Mission_Outcome , count(*) as total from SPACEXTBL group by Mission_Outcome
```

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

Mission_Outcome	total
None	898
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

```
%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

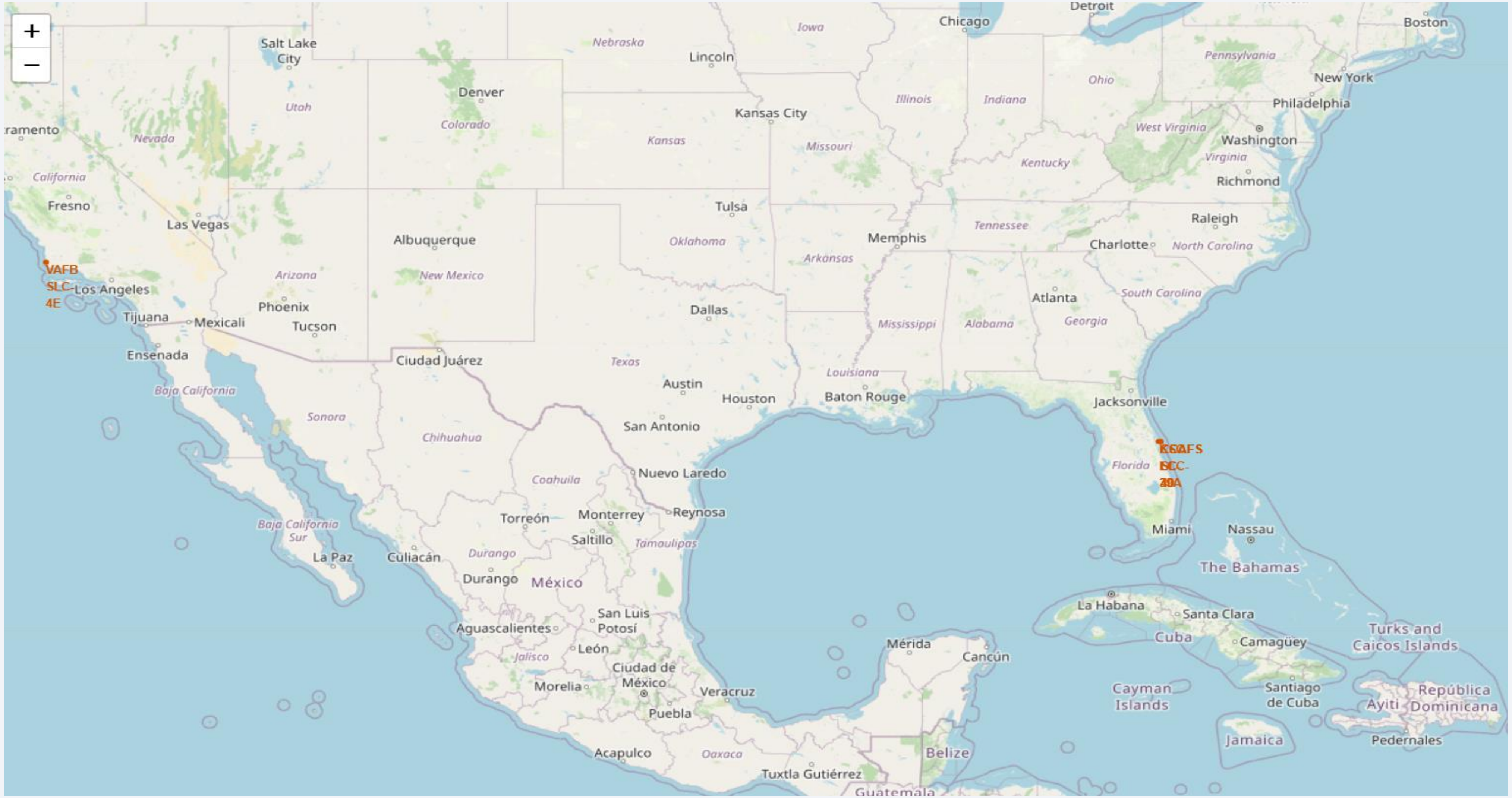
```
%sql select strftime('%m', Date) AS month, landing_outcome ,booster_version, launch_site
from SPACEXTBL where substr(Date, 7, 4) = '2015' AND landing_outcome = 'Failure (drone ship)'
```

month	Landing_Outcome	Booster_Version	Launch_Site
None	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
None	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

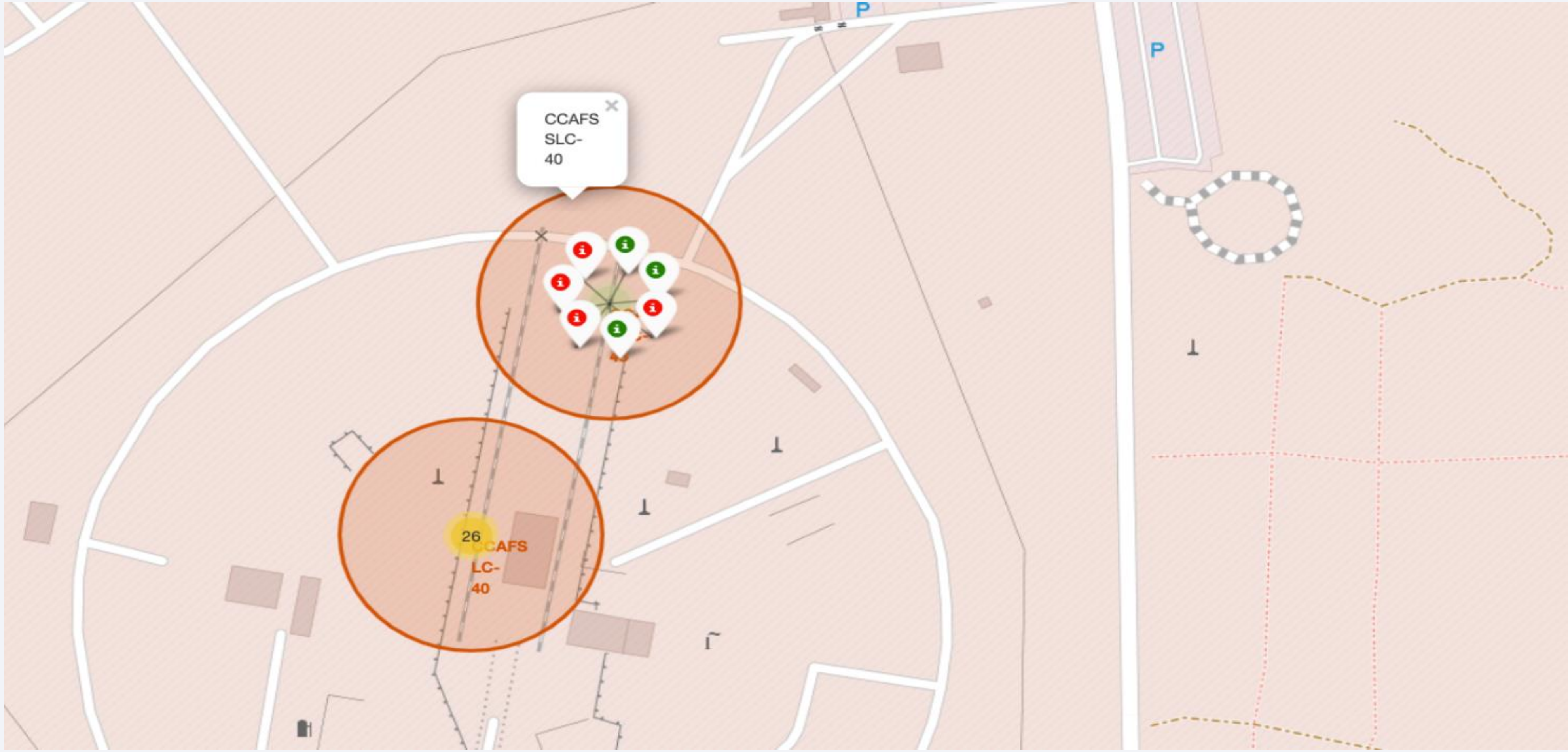
Section 3

Launch Sites Proximities Analysis

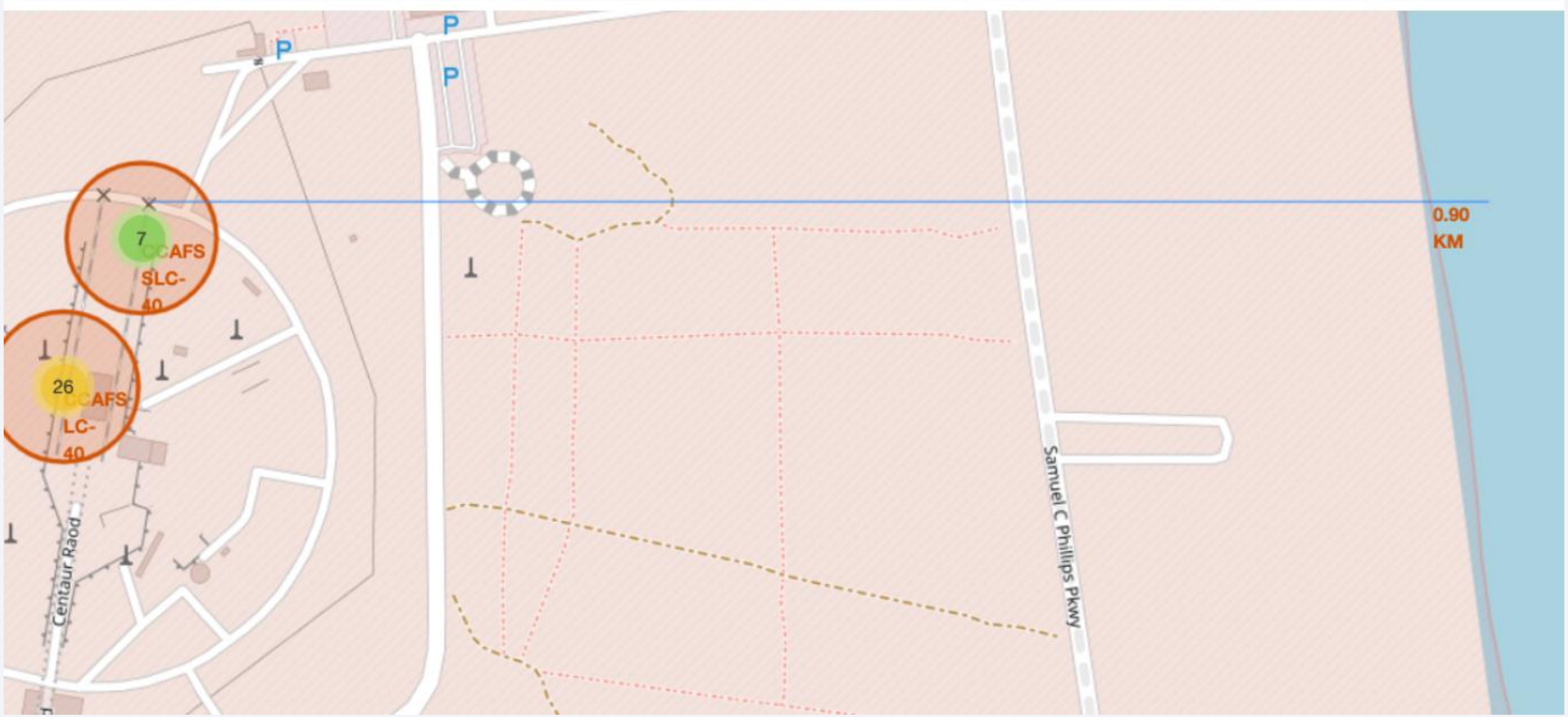
All launch sites on a map



Success/failed launches for each site on the map



Distances between a launch site to its proximities

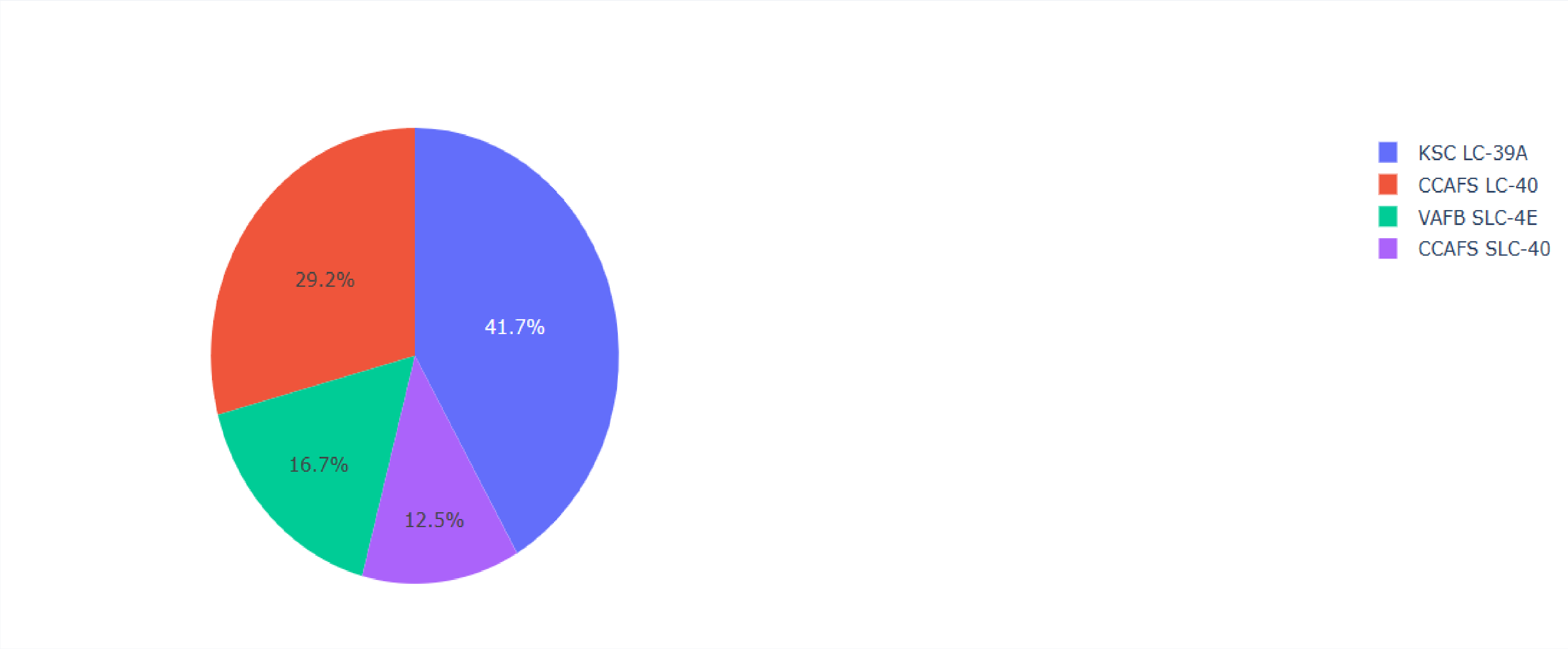


Section 4

Build a Dashboard with Plotly Dash

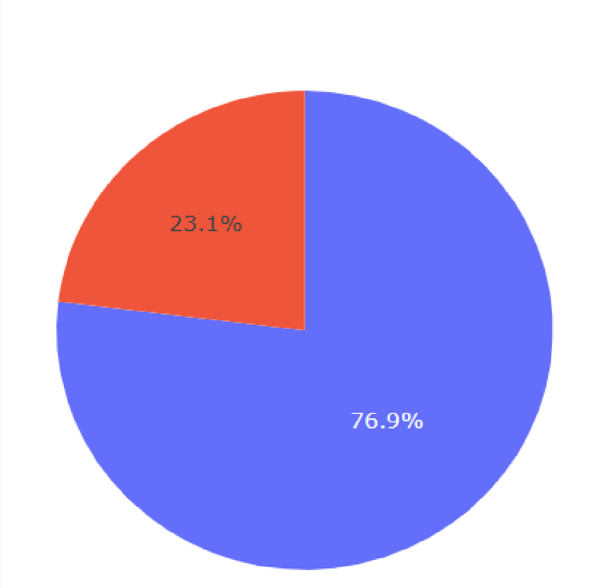
Launches from all sites

KSC LC-39A has the highest launches number

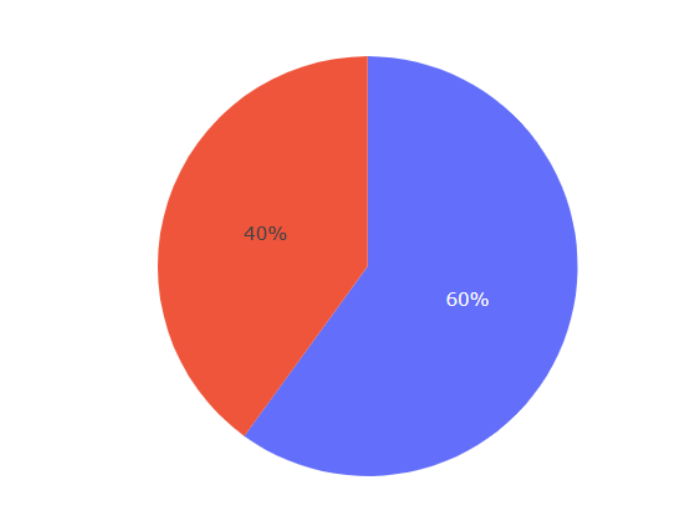


Success Rate by Site

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

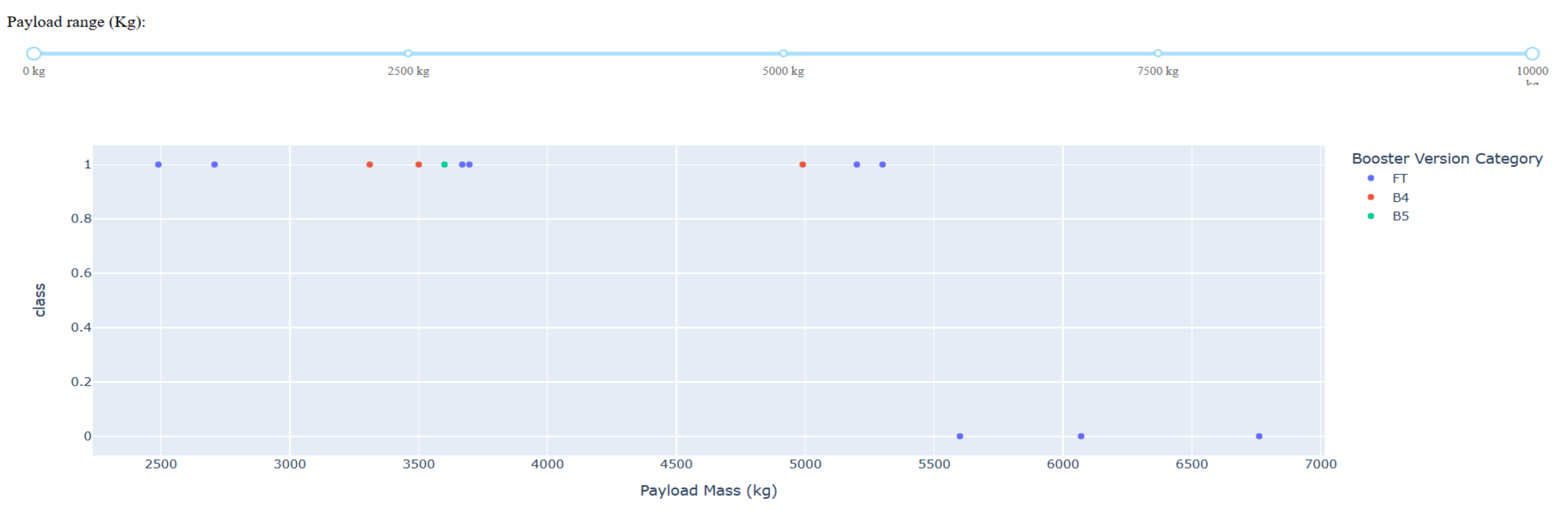


VAFB SLC-4E achieved dna etar sseccus 60%
%40failure rate



Payload vs Launch outcome

The Success rate of low weighted payloads is higher than the heavy weighted payloads

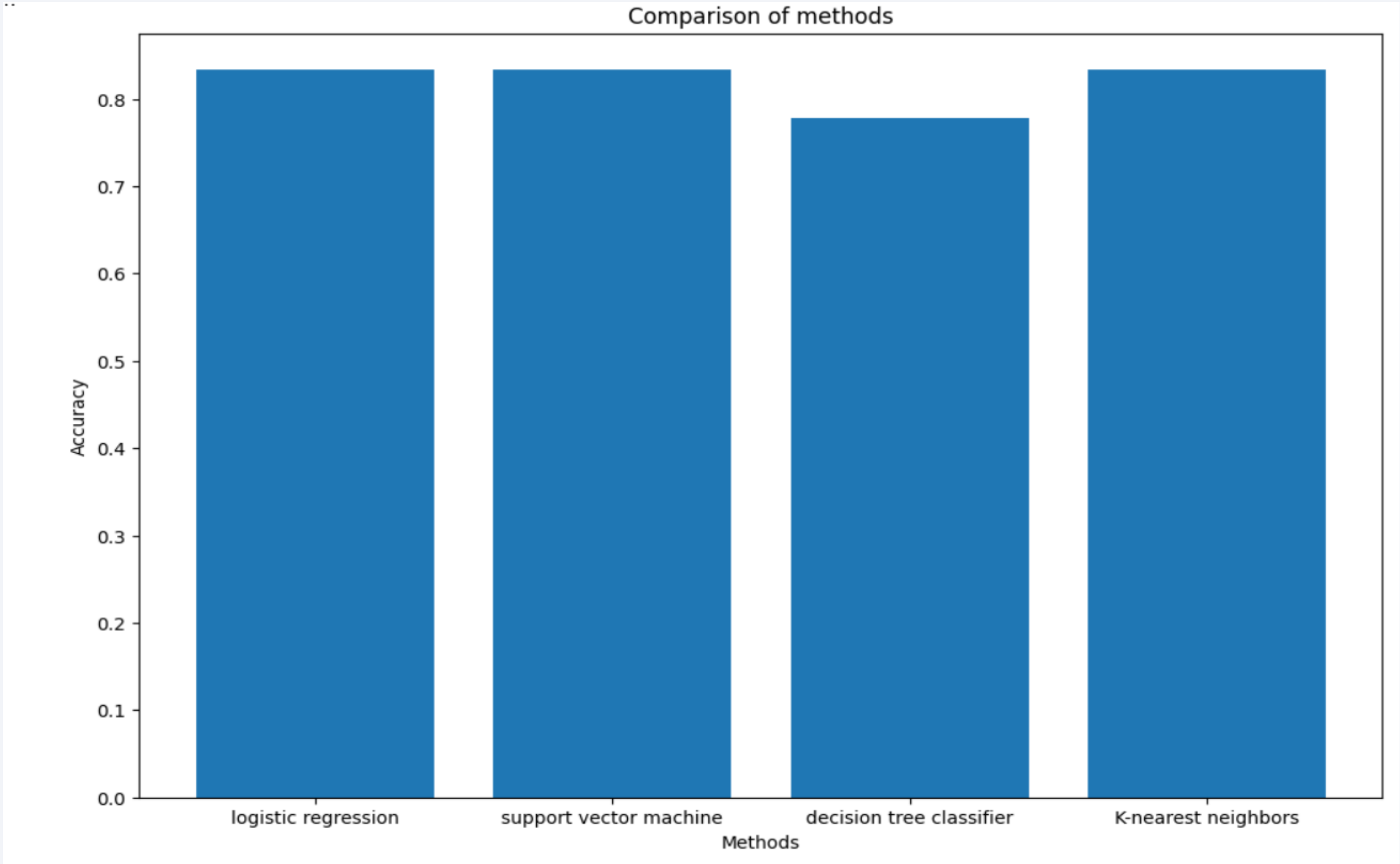


Section 5

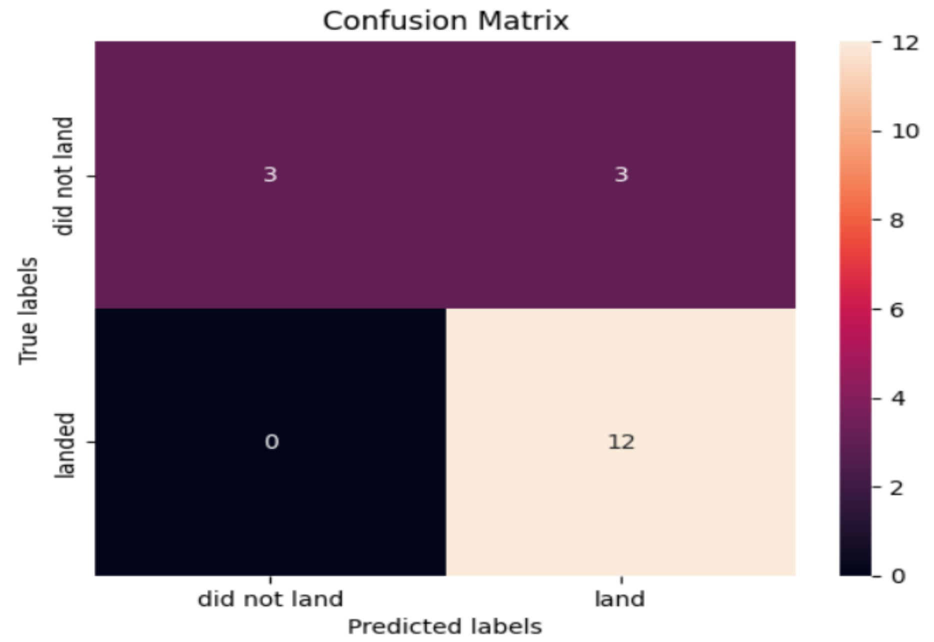
Predictive Analysis (Classification)

Classification Accuracy

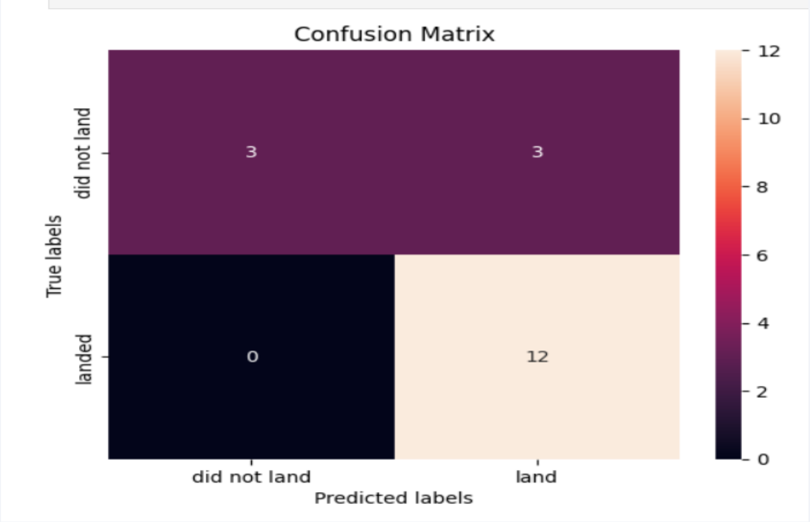
Logistic Regression , SVM and KNN achieved the highest accuracy at 83.3%

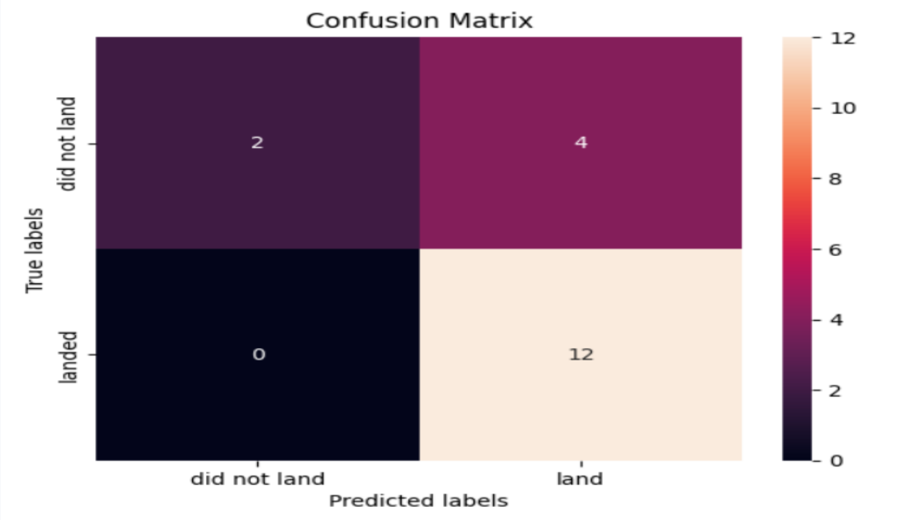


Confusion Matrix



Examining the confusion matrix, we see that logistic regression can distinguish between the different classes. We see that the major problem is false positives.





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 Space 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!