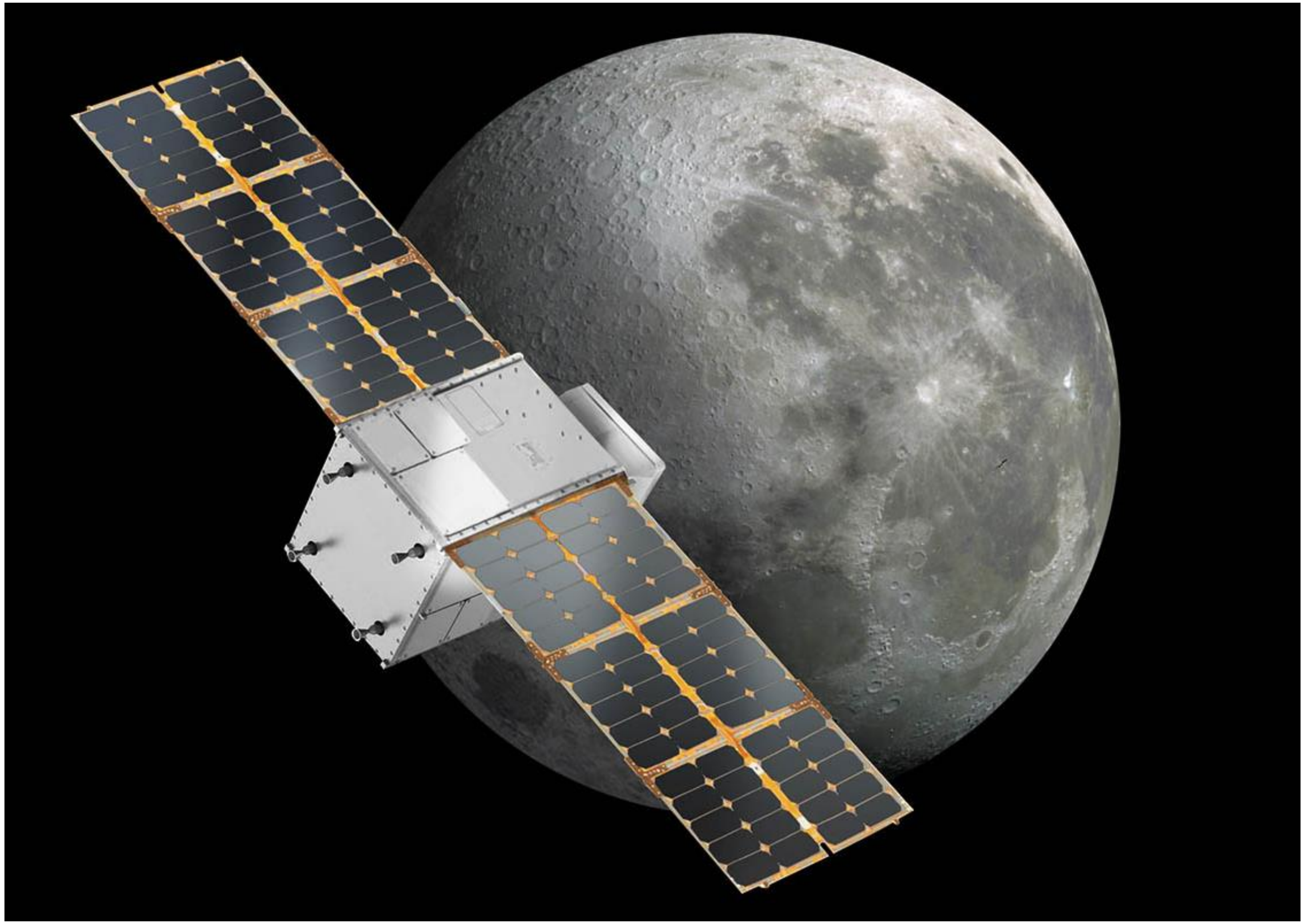
Winning Space Race with Data Science

Sholeh asl Sariraei

11_08_2023

GitHub : sarirsholeh/Capstone

Outline



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 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 the 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

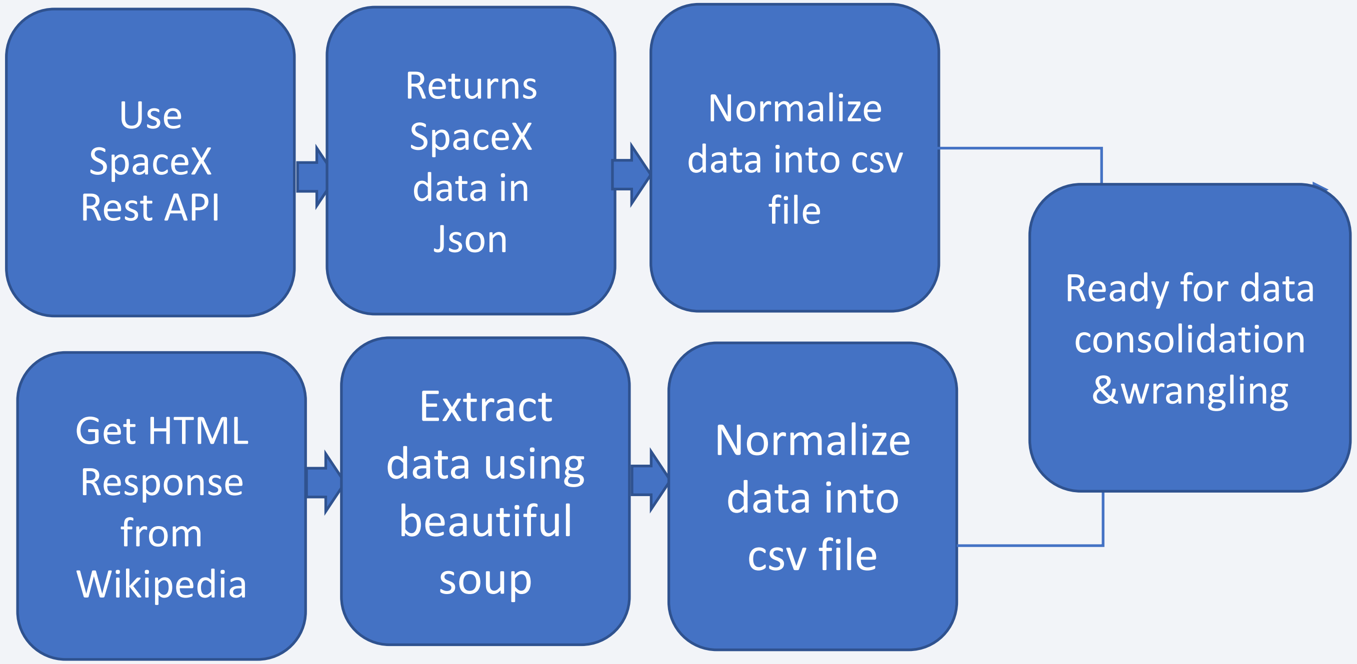
LR, KNN, SVM, DT models have been built and evaluated for the best classifier

Data Collection

SpaceX launch data that is gathered from the SpaceX REST API.

This API will give us data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome.

The SpaceX REST API endpoints, or URL, starts with api.spacexdata.com/v4/.
Another popular data source for obtaining Falcon 9 Launch data is web scraping Wikipedia using Beautiful Soup.



Data Collection – SpaceX API

Data collection with SpaceX REST calls

Task 1: Request and parse the SpaceX launch data using the GET request

To make the requested JSON results more consistent, we will use the following static response object for this project:

```
static_json_url='https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_
```

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
import pandas as pd
import requests

# Decode the response content as JSON
json_data = response.json()

# Convert JSON data to a pandas DataFrame using json_normalize
data= pd.json_normalize(json_data)

# Display the resulting DataFrame
data
```

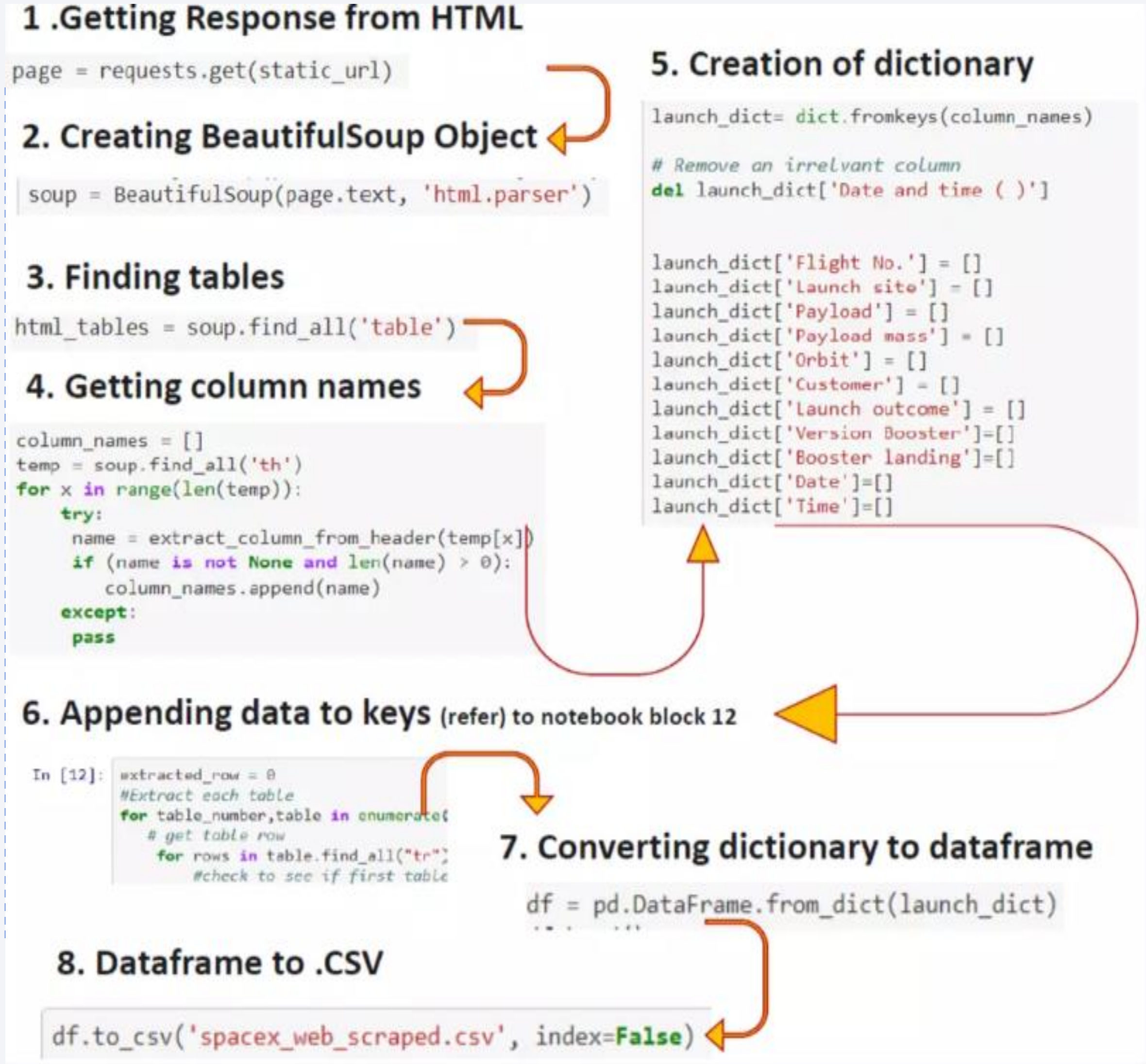
Now let's start requesting rocket launch data from SpaceX API with the following URL:

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

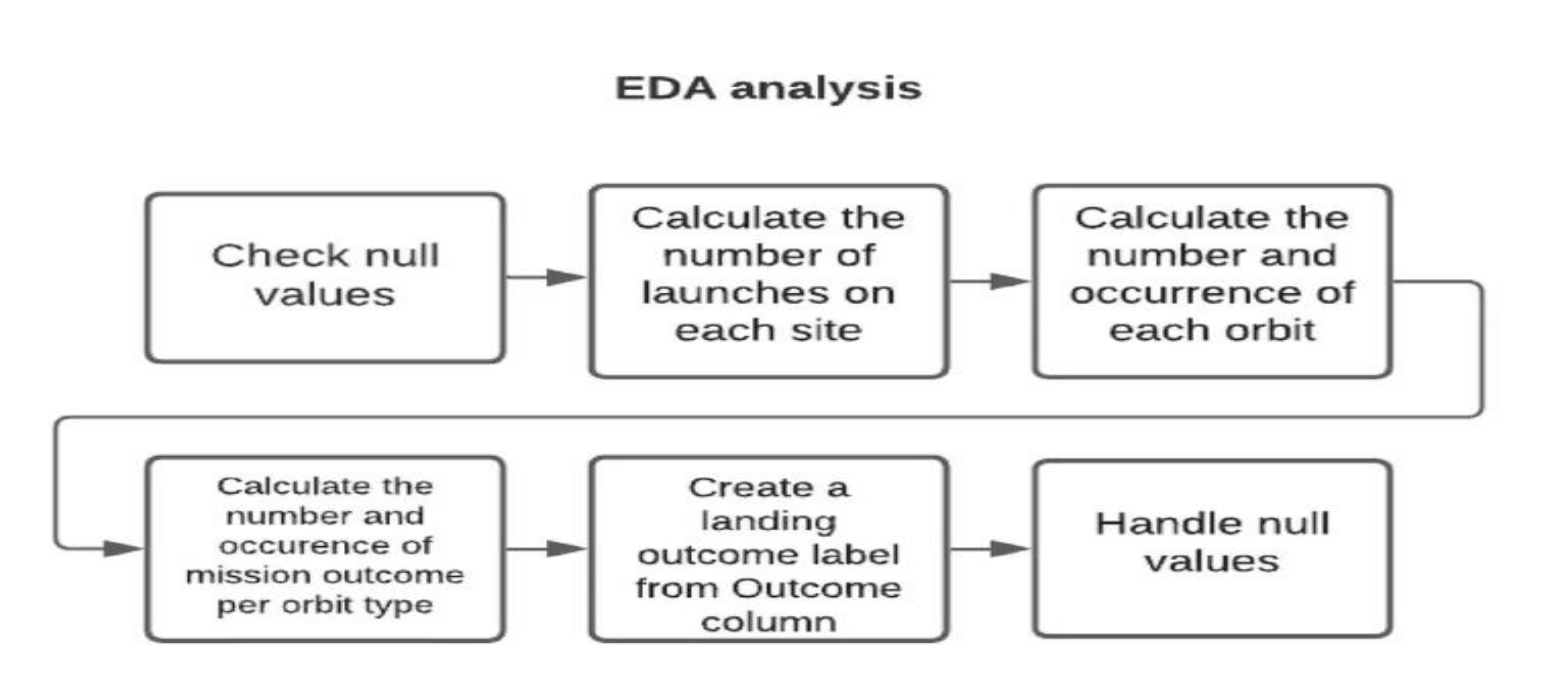
```
response = requests.get(spacex_url)
```

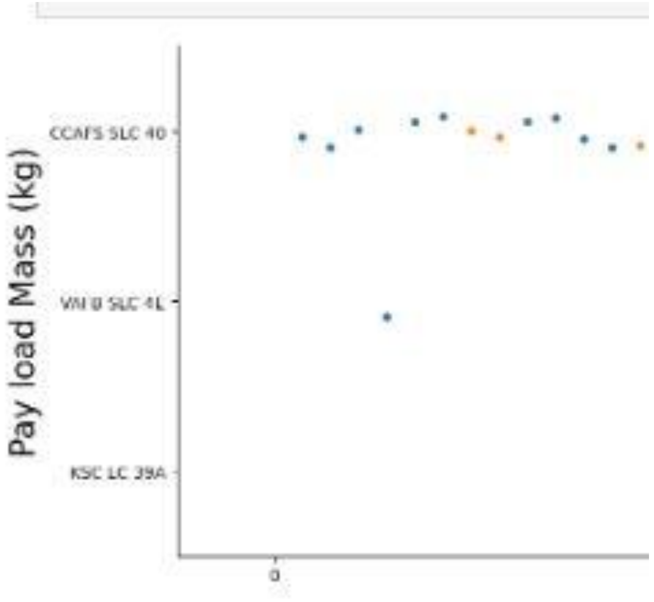
Data Collection - Scraping

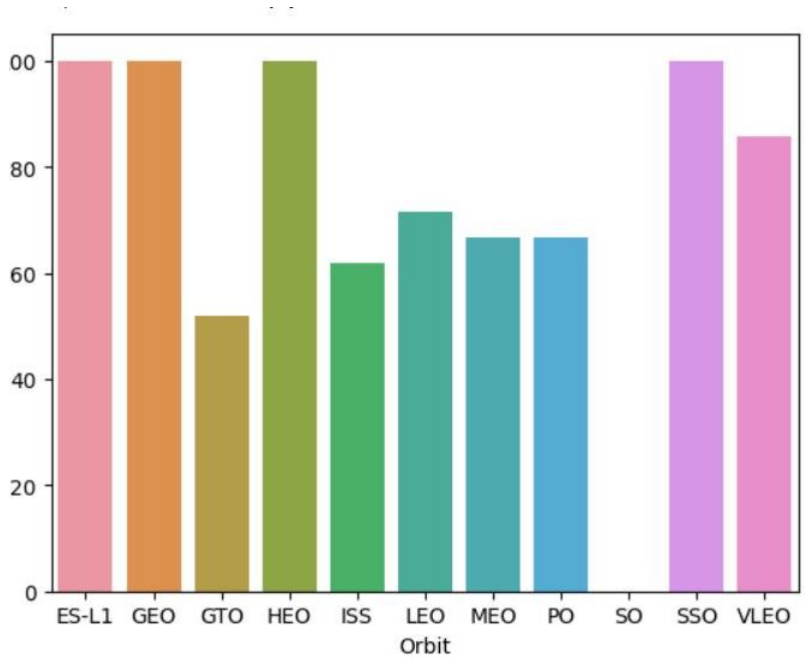
Web Scraping from Wikipedia

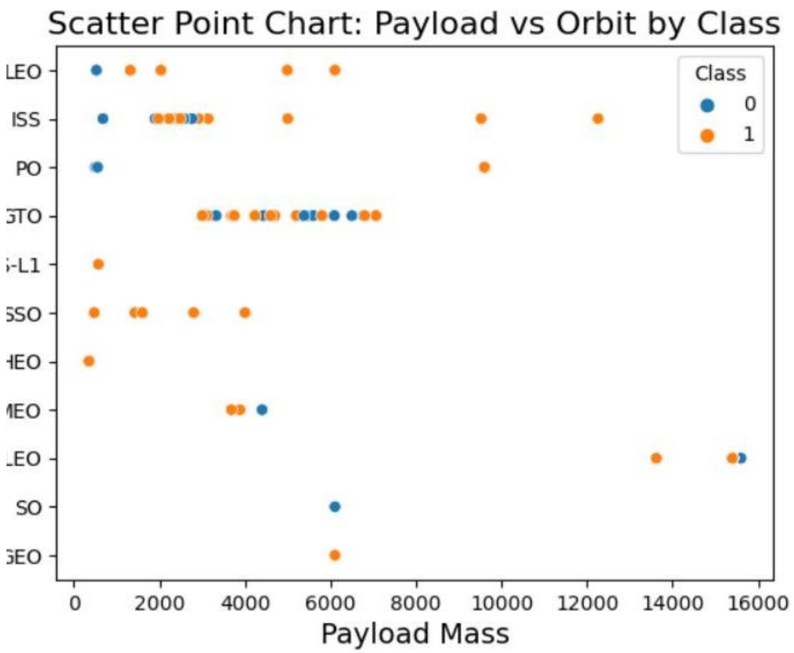


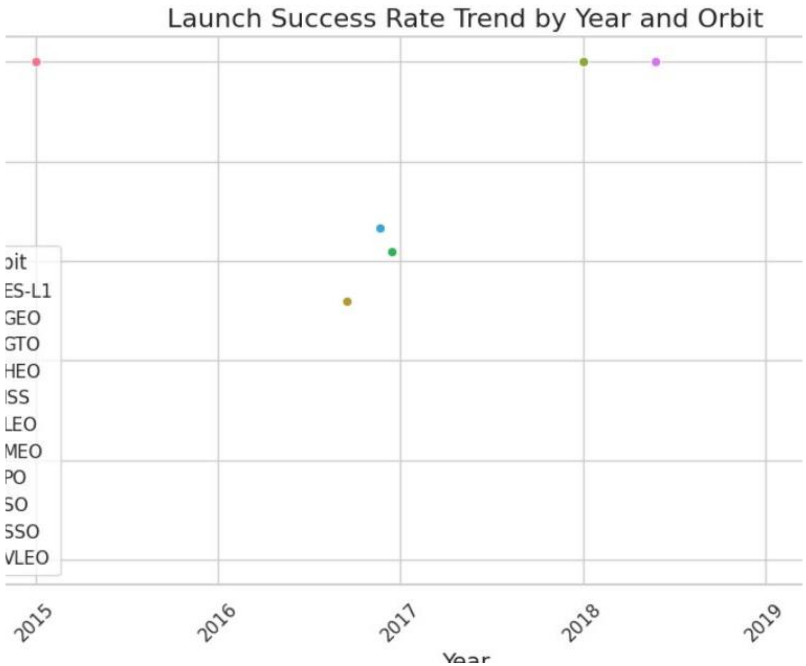
Data Wrangling

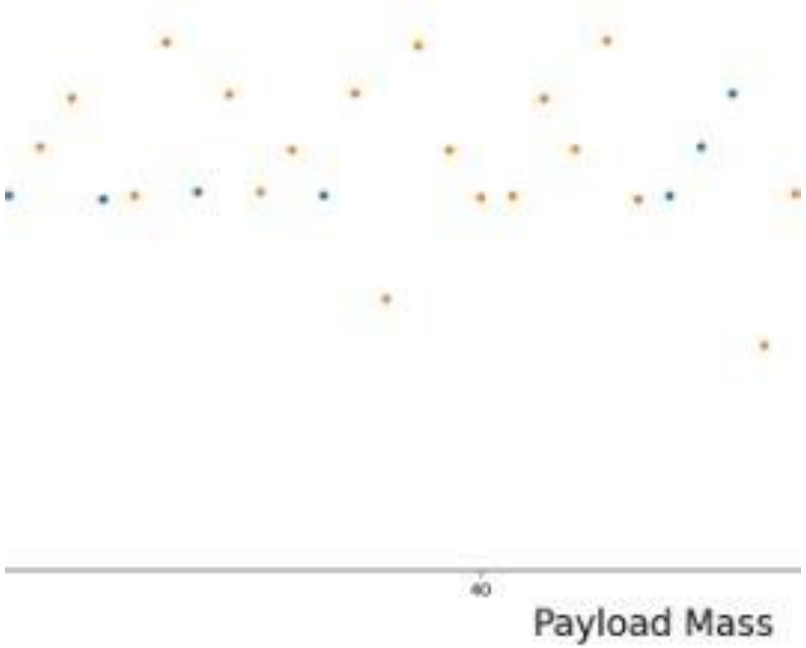



Pay load Mass (kg)
CCAFS SLC 40
VAFB SLC 4E
KSC LC 39A
0


00
80
60
40
20
0
ES-L1
GEO
GTO
HEO
ISS
LEO
MEO
PO
SO
SSO
VLEO
Orbit


Scatter Point Chart: Payload vs Orbit by Class
Class
0
1
LEO
ISS
PO
GTO
SSO
HEO
MEO
SO
GEO
0
2000
4000
6000
8000
10000
12000
14000
16000
Payload Mass


Launch Success Rate Trend by Year and Orbit
ES-L1
GEO
GTO
HEO
ISS
LEO
MEO
PO
SO
SSO
VLEO
2015
2016
2017
2018
2019

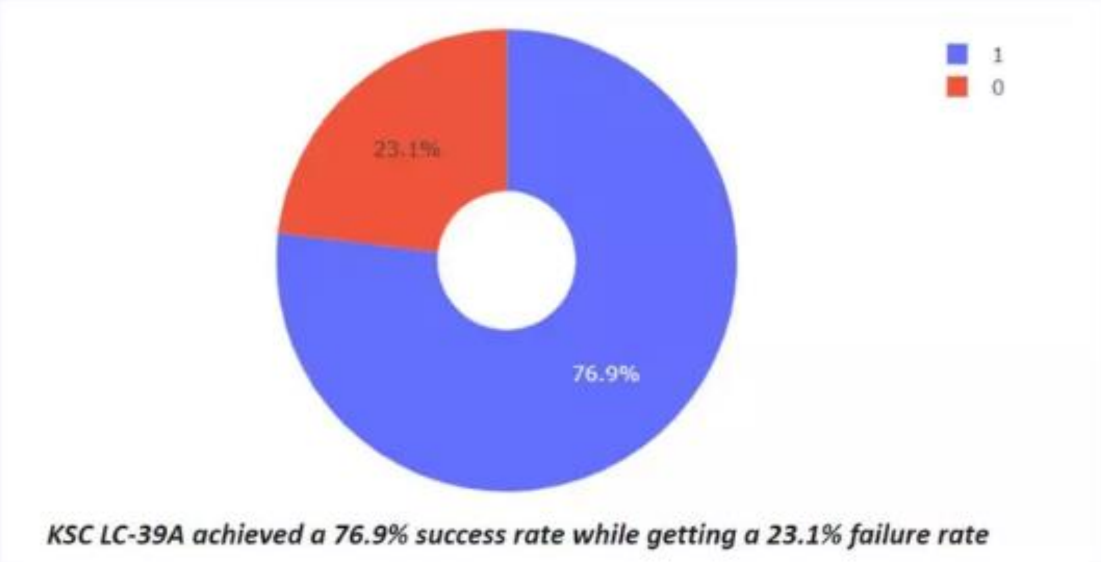

40
Payload Mass


Launch Site

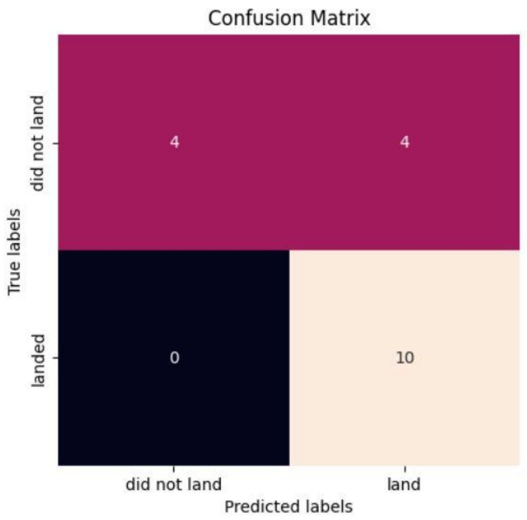
EDA with SQL

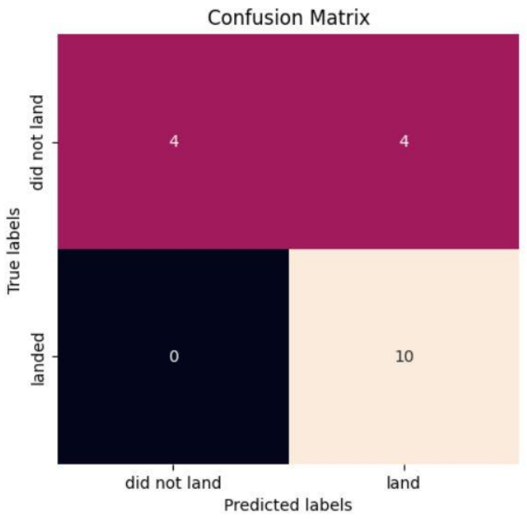
- Displaying the names of the unique launch sites in the space mission
- Displaying 5 records where launch sites begin with the string 'KSC
- Displaying the total payload mass carried by boosters launched by NASA (CRS)
- Displaying average payload mass carried by booster version F9 v1.1
- Listing the date where 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
- Listing the records which will display the month names, successful landing outcomes in ground pad booster versions, launch site for the months in year 2017
- Ranking the count of successful landing outcomes between the date 2010 06 04 and 2017 03 20 in descending order.

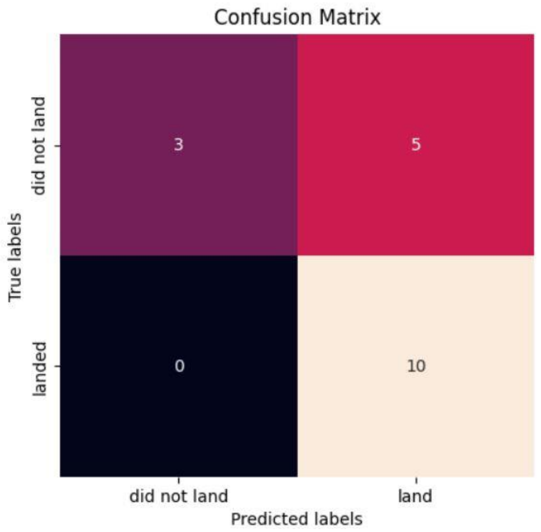
Build an Interactive Map with Folium

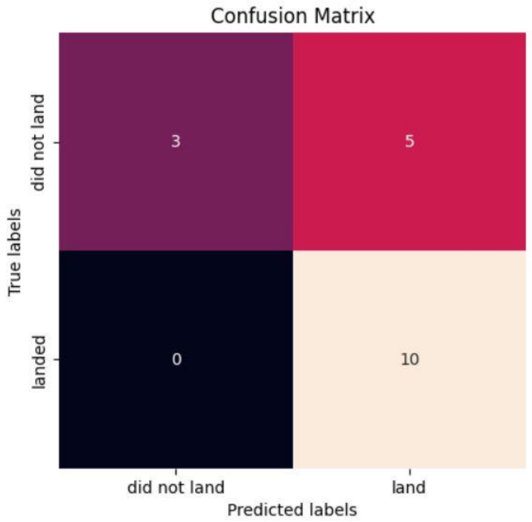


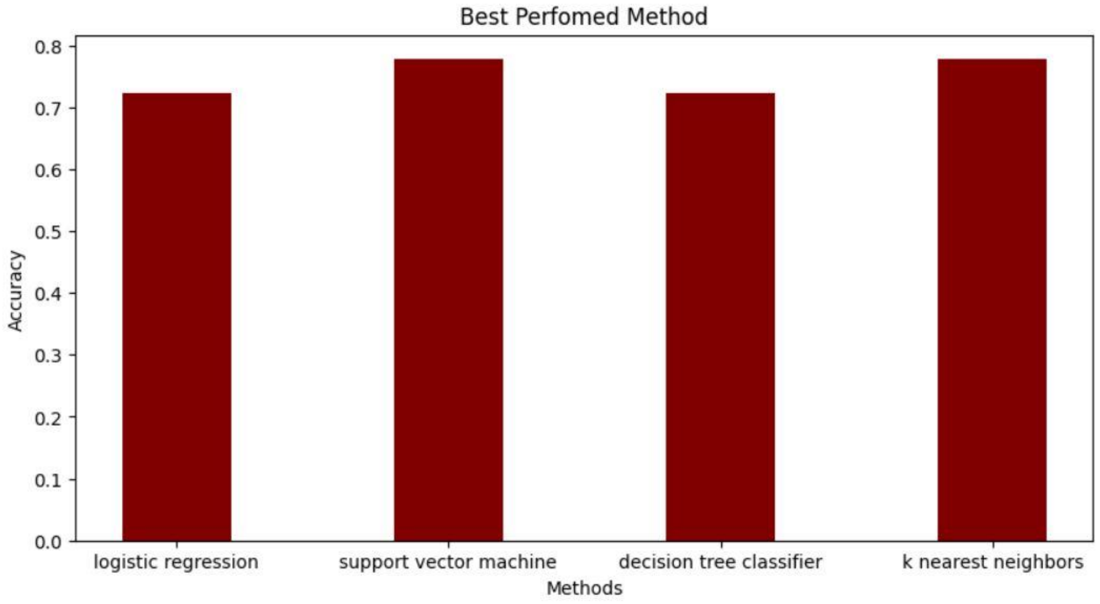
Dashboard with Plotly Dash











Predictive Analysis (Classification)

- The SVM, KNN, and Logistic Regression model achieved the highest accuracy at 83.3%, while the SVM performs the best in terms of Area Under the Curve at 0.958.

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 SpaceX launches is directly proportional time in years they will eventually perfect the launches.

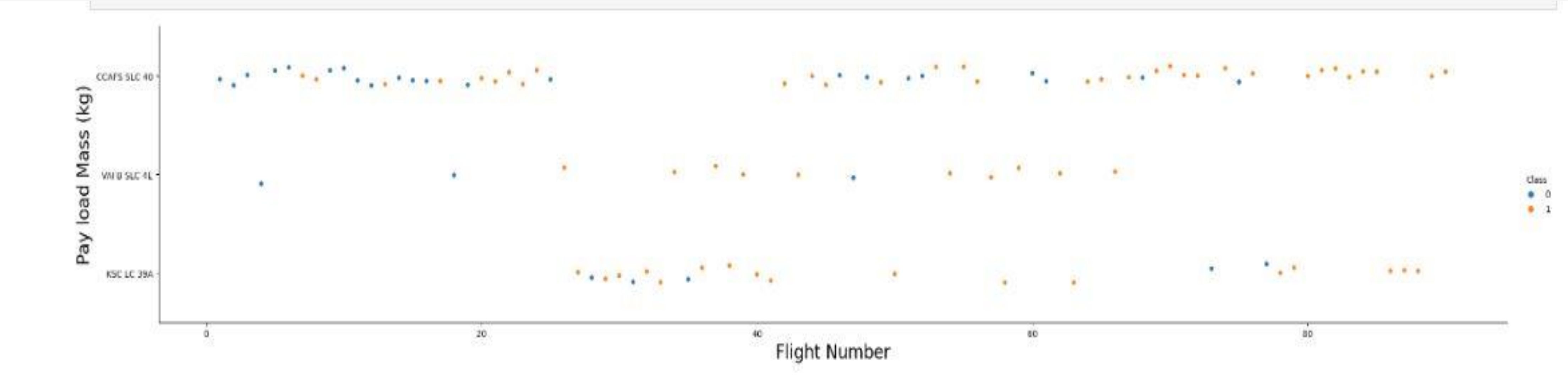
KSC LC 39A had the most successful launches from all the sites.

• Orbit GEO,HEO,SSO,ES L1 has the best Success Rate.

Section 2

Insights drawn from EDA

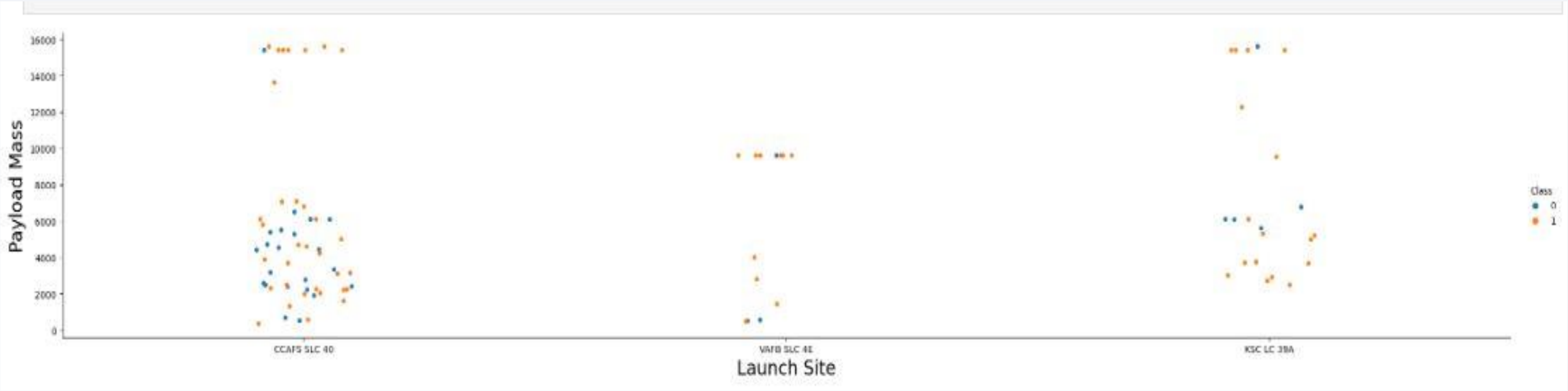
Flight Number vs. Launch Site



Launches from the site of CCAFS SLC 40 are launches form other sites.

significantly higher than lunches from other sites

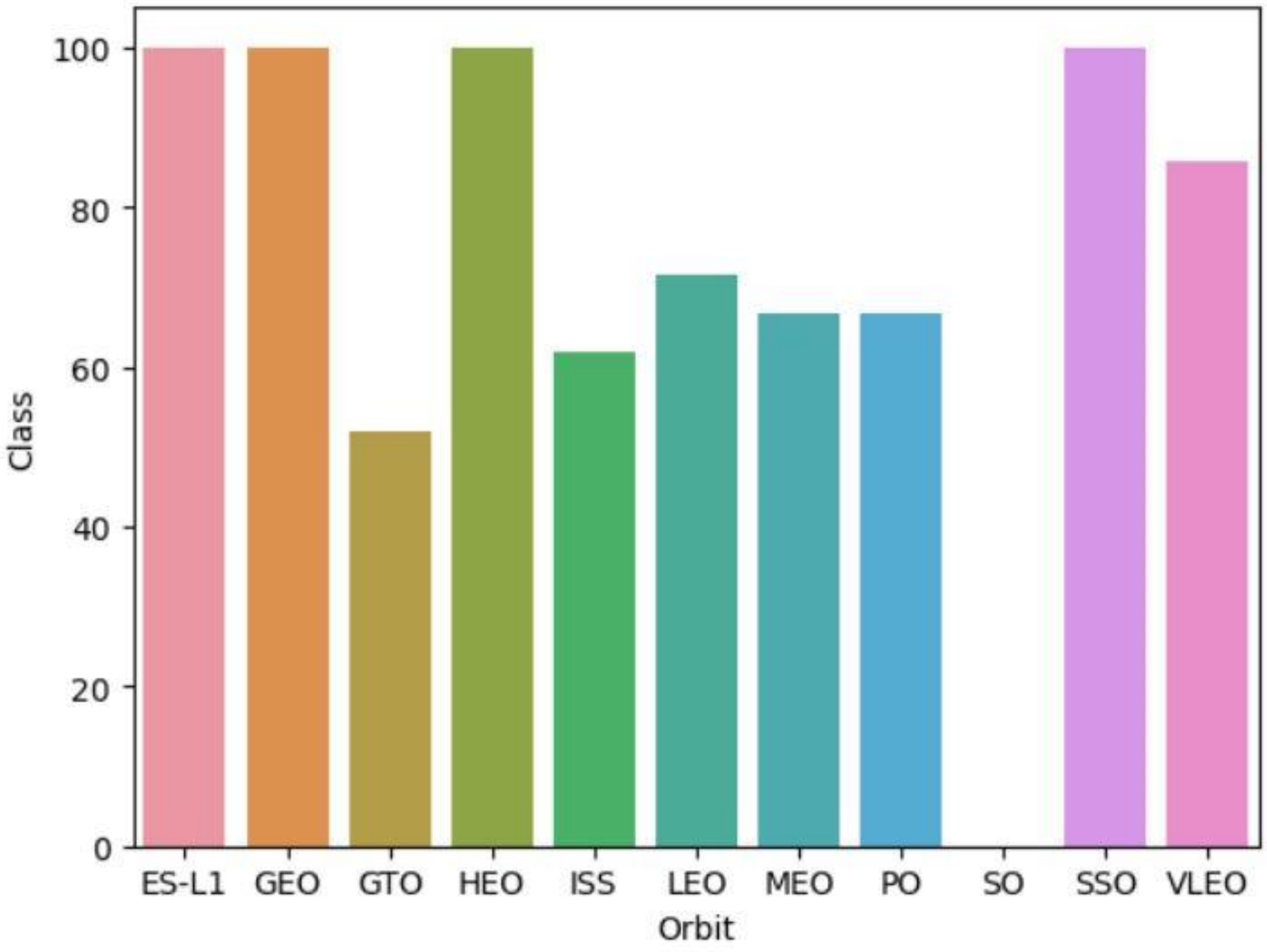
Payload vs. Launch Site

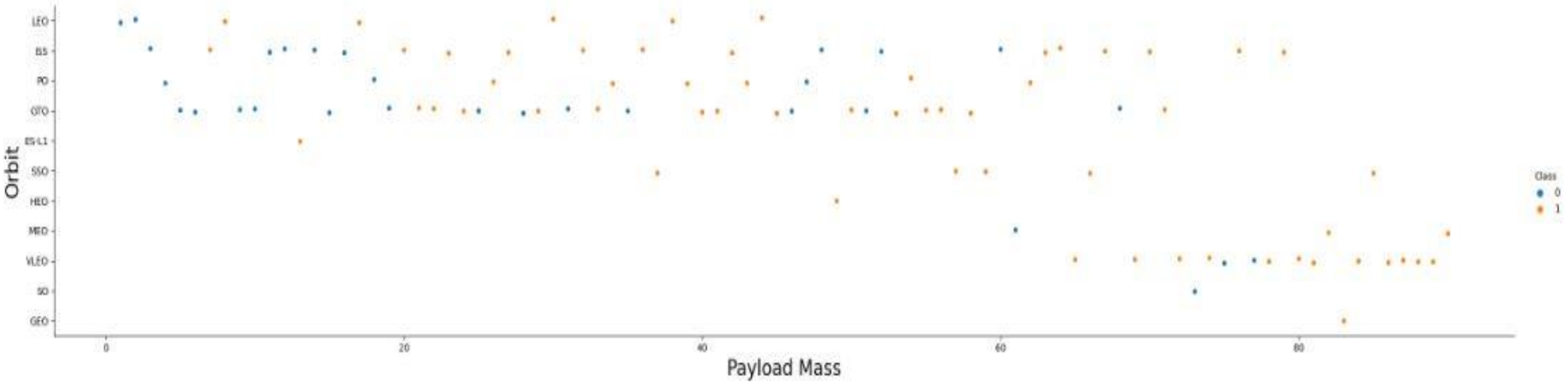


- The majority of IPay Loads with lower Mass have been launched from CCAFS SLC 40.

Success Rate vs. Orbit Type

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



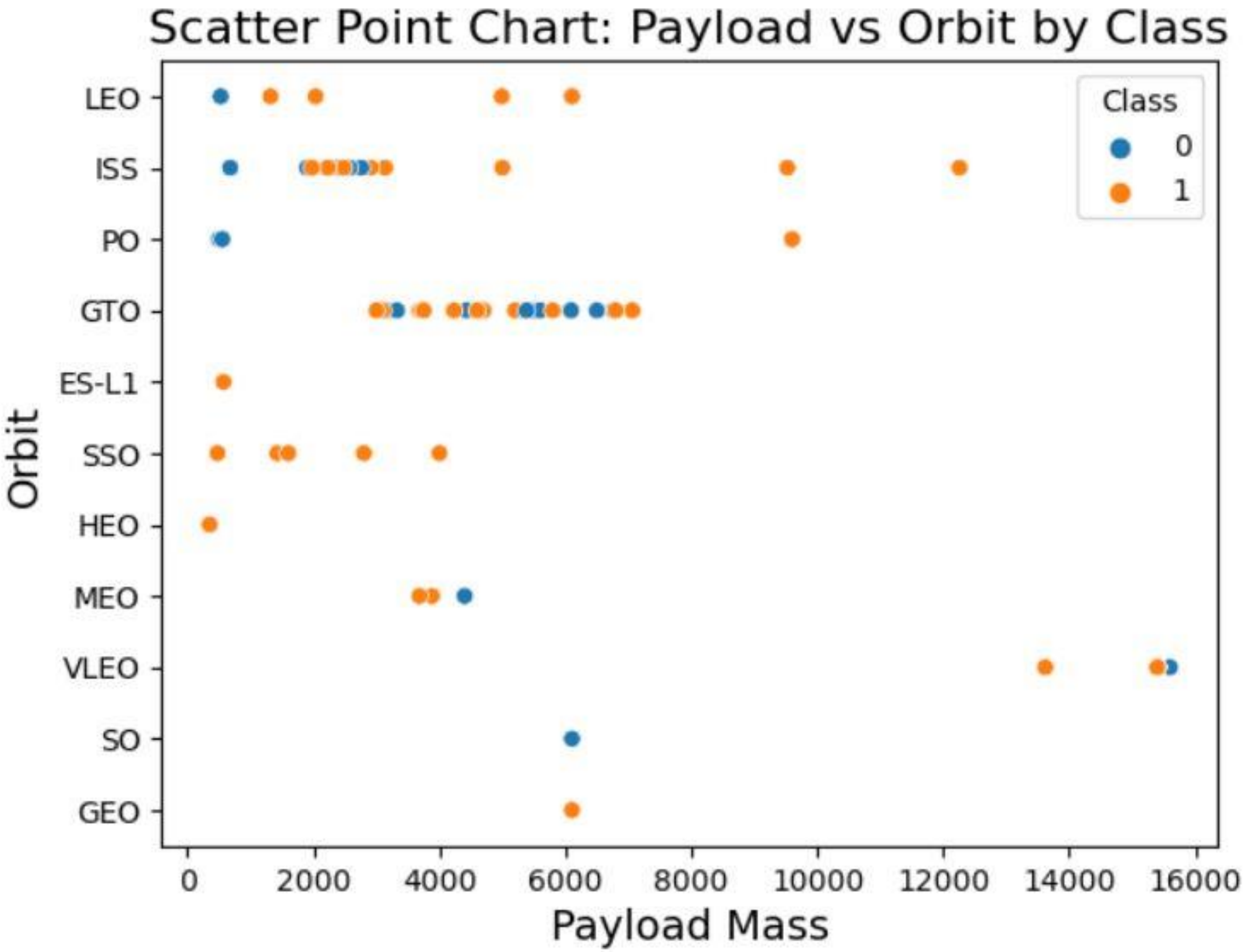


Flight Number vs. Orbit Type

- A trend can be observed of shifting to VLEO launches in recent years.

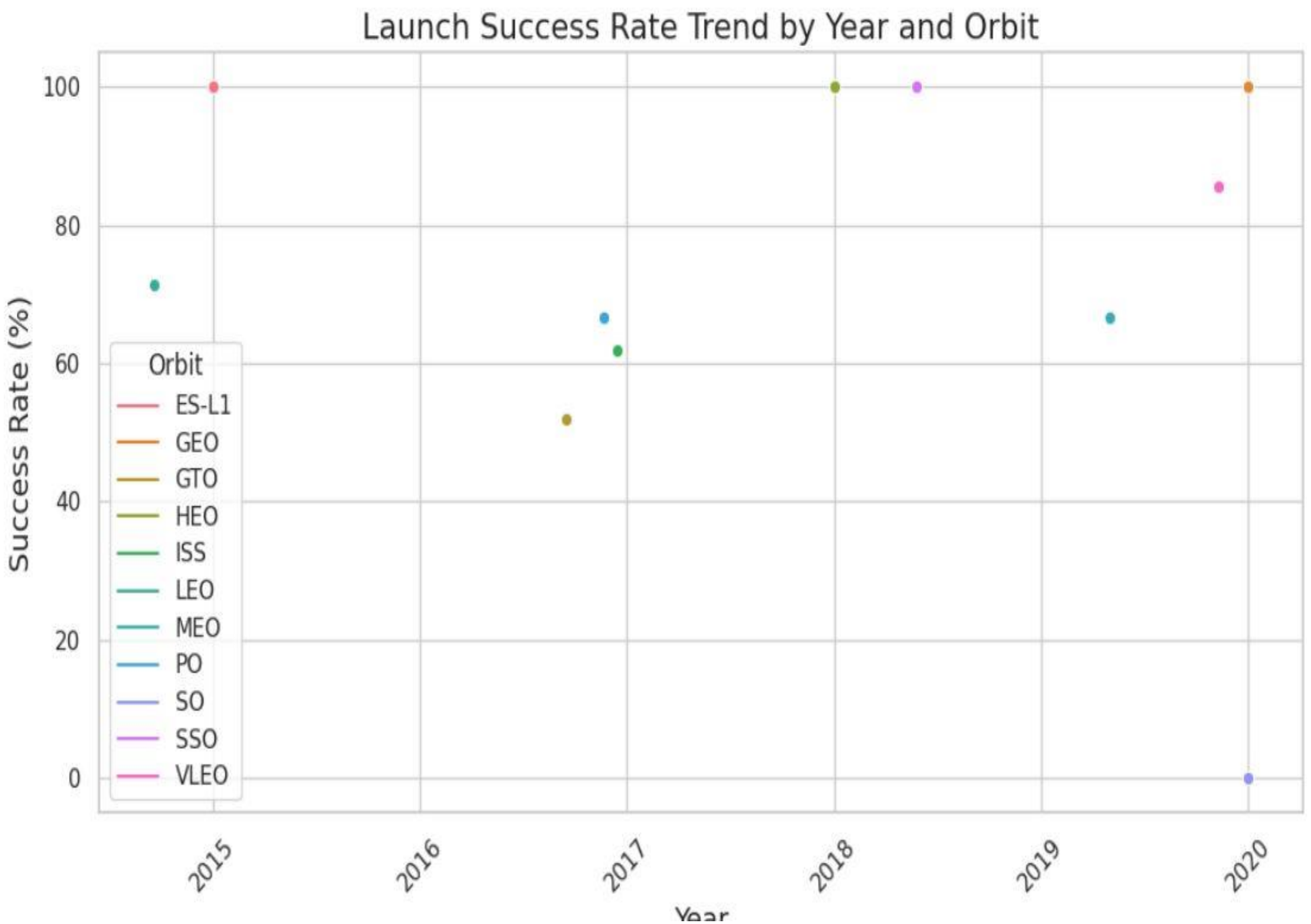
Payload vs. Orbit Type

There are strong correlation between ISS and Payload at the range around 2000, as well as between GTO and the range of 4000-8000.



Launch Success Yearly Trend

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



All Launch Site Names

- %sql select distinct(LAUNCH_SITE) from SPACEXTBL

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

Launch Site Names Begin with 'CCA'

- %sql select * from SPACEXTBL where LAUNCH_SITE like 'CCA%' limit 5

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

- %sql select sum(PAYLOAD_MASS_KG_) from SPACEXTBL where CUSTOMER = 'NASA (CRS)’

45596

Average Payload Mass by F9 v1.1

- %sql select avg(PAYLOAD_MASS_KG_) from SPACEXTBL where BOOSTER VERSION = 'F9 v1.1’

✓2928400000

First Successful Ground Landing Date

- %sql select min(DATE) from SPACEXTBL where Landing_Outcome = 'Success (ground pad)’

2015-12-22

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

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 count(MISSION_OUTCOME) from SPACEXTBL where
- MISSION OUTCOME = 'Success' or MISSION_OUTCOME = 'Failure (in flight)'

100

Boosters Carried Maximum Payload

- %sql select BOOSTER_VERSION from SPACEXTBL where
- PAYLOAD_MASS_KG_ = (select max(PAYLOAD_MASS_KG_) from SPACEXTBL)

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 * from SPACEXTBL where Landing Outcome like 'Success%' and (DATE between '2015-01-01' and '2015-12-31') order by date desc

time_utc_	booster_version	launch_site	payload	payload_mass_kg_	orbit	customer	mission_outcome	landing_outcome
14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600	Polar LEO	Iridium Communications	Success	Success (drone ship)
05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
04:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success	Success (drone ship)

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

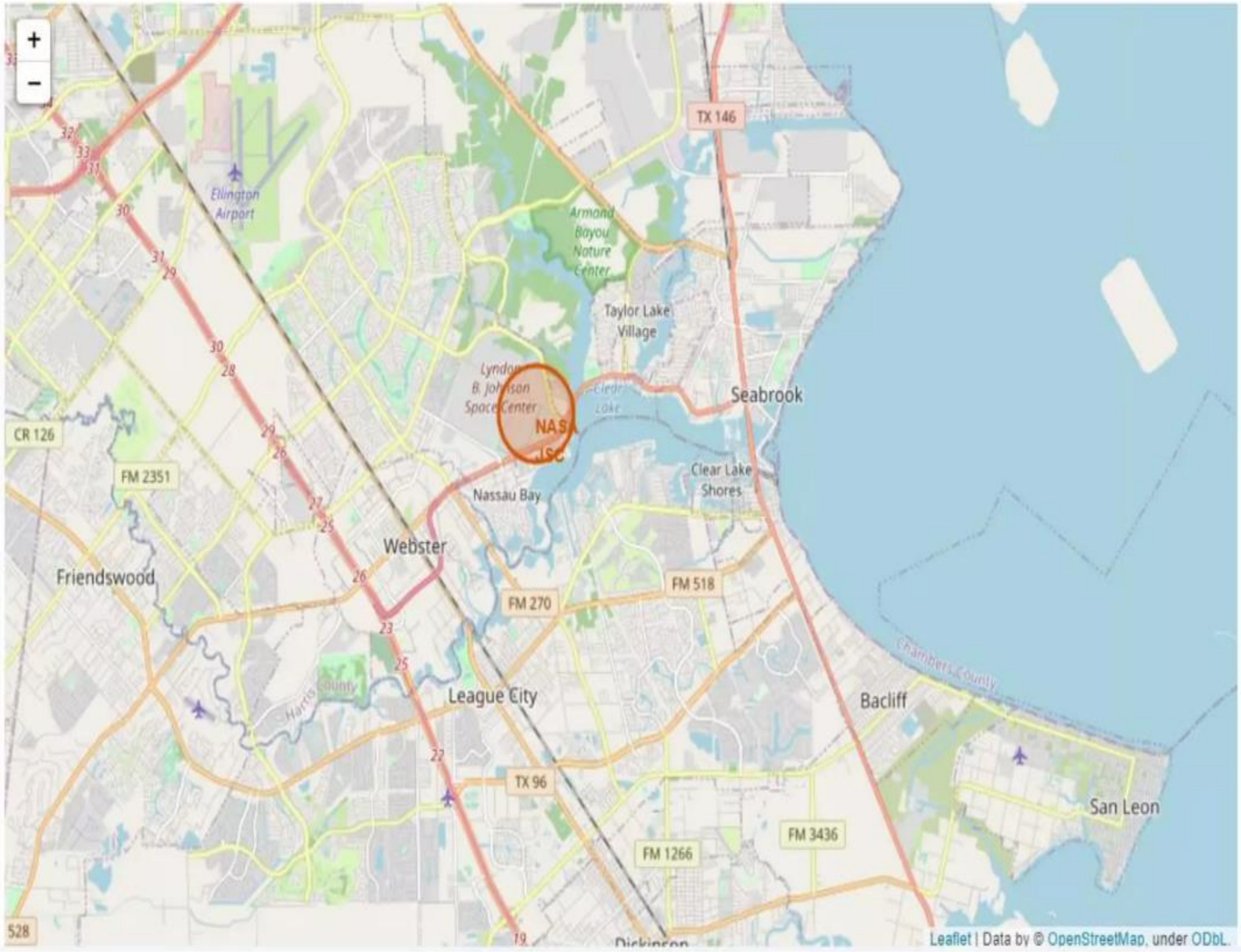
- %sql select * from SPACEXTBL where Landing_Outcome like 'Success%' and (DATE between '2010-06-04' and '2017-03-20') order by date desc

2016-05-27	21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success	Success (drone ship)
2016-05-06	05:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-04-08	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136	LEO (ISS)	NASA (CRS)	Success	Success (drone ship)
2015-12-22	01:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

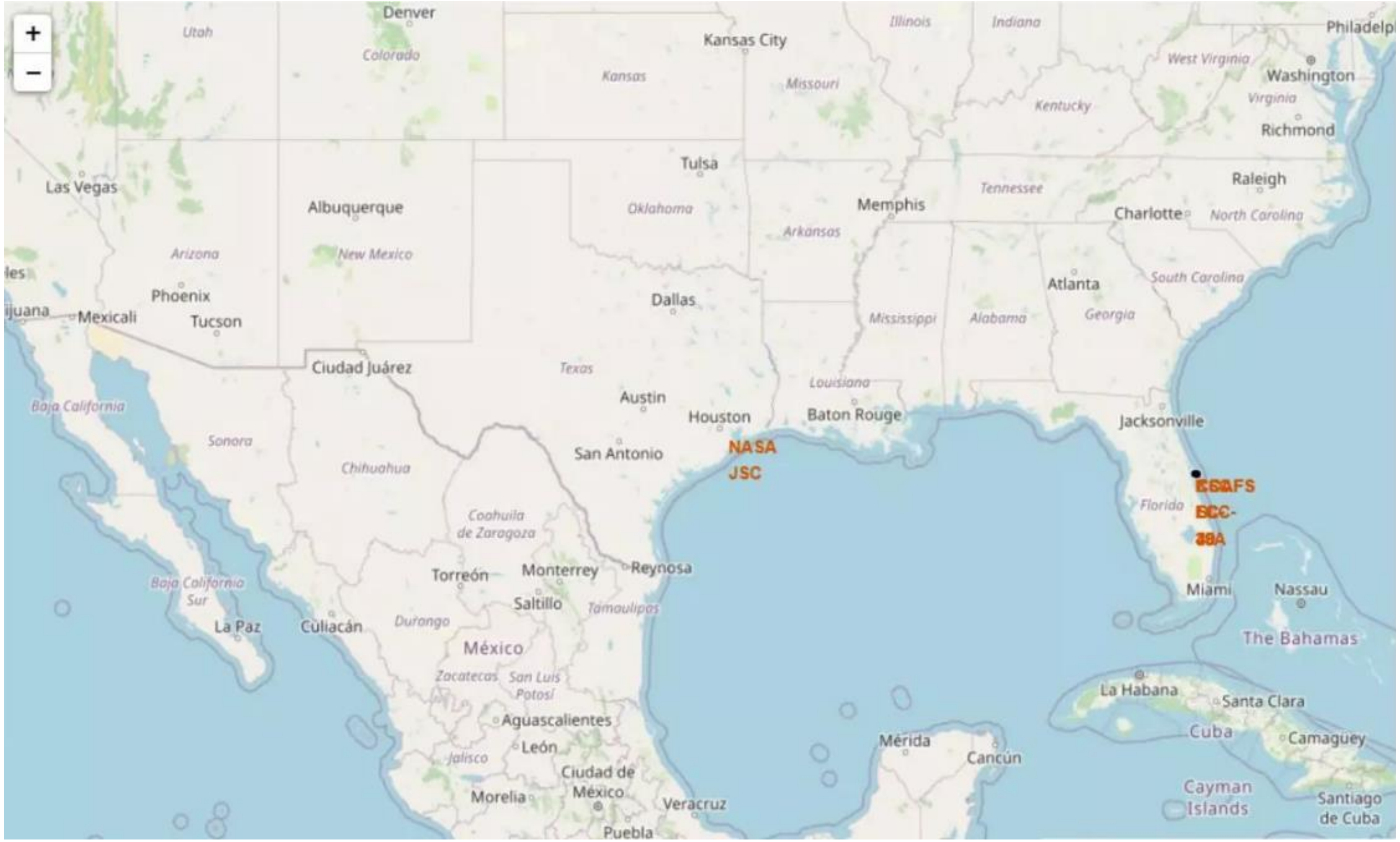
Section 3

Launch Sites Proximities Analysis

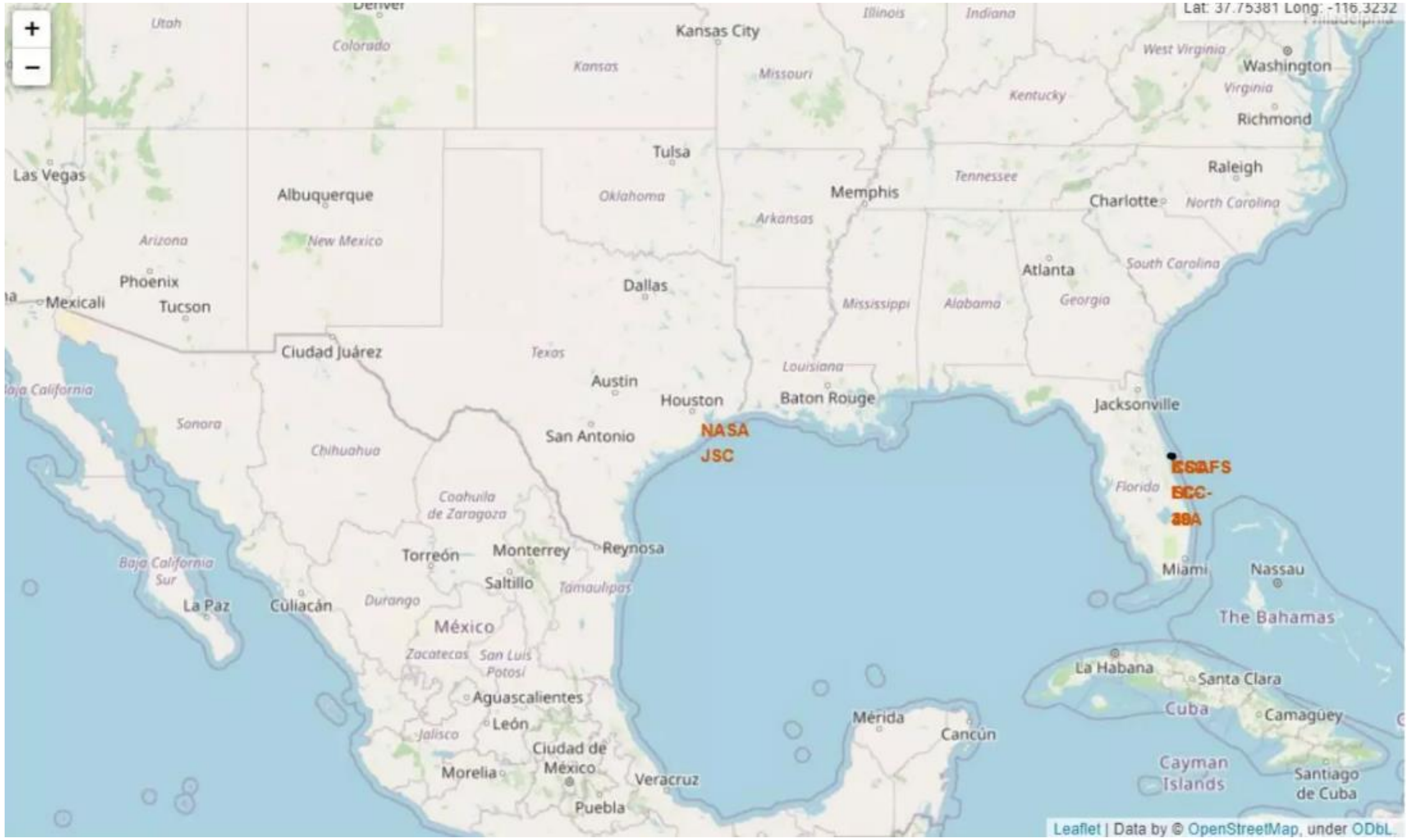
Folium Map Screenshot 1



Folium Map Screenshot 2



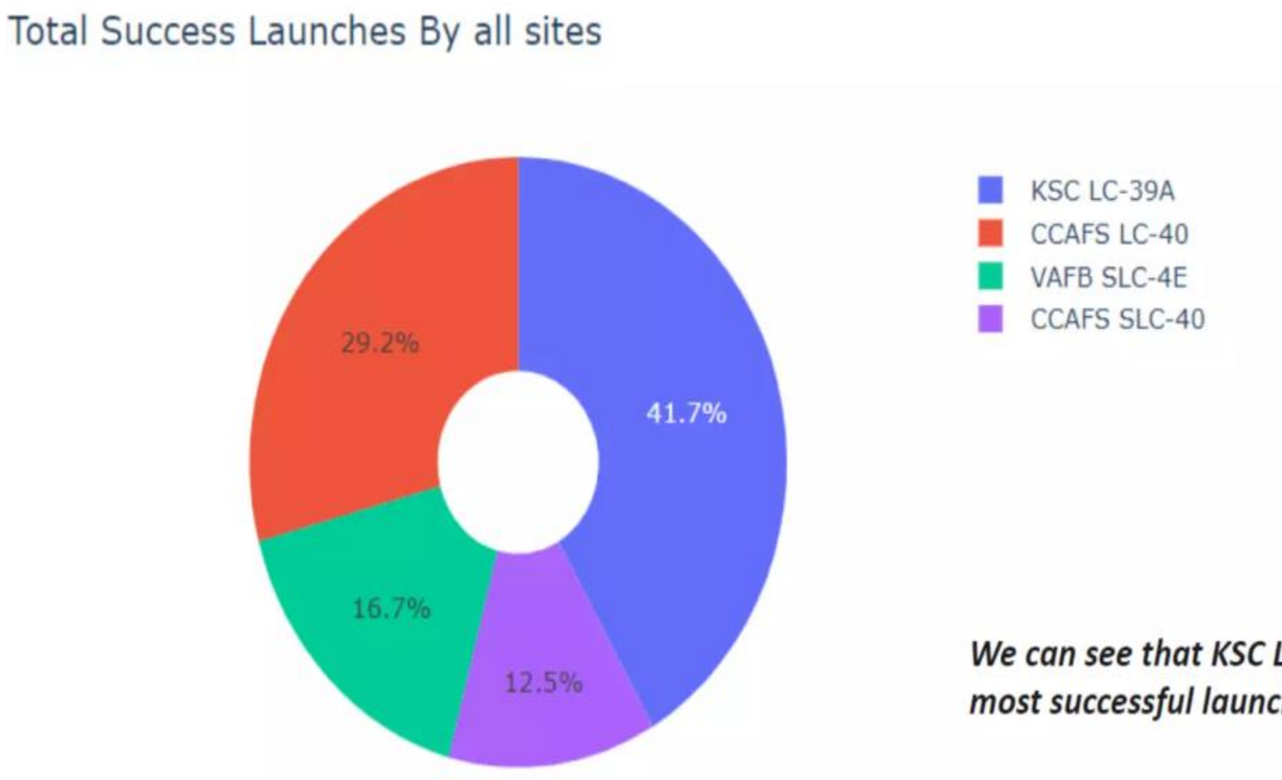
Folium Map Screenshot 3



Section 4

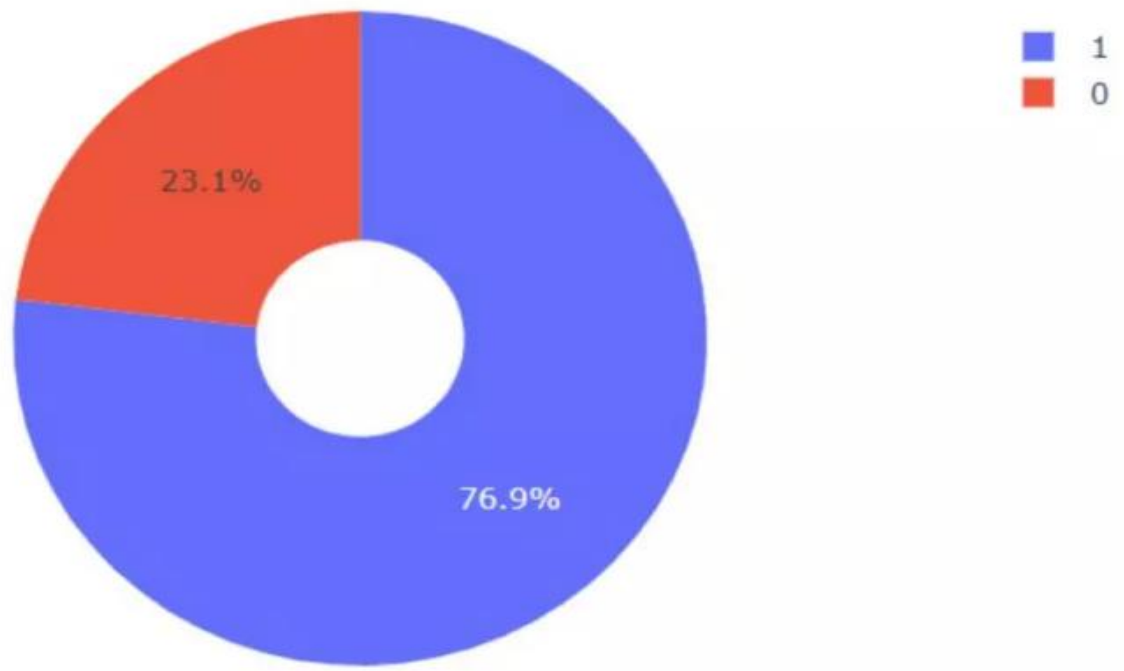
Build a Dashboard with Plotly Dash

Dashboard Screenshot 1



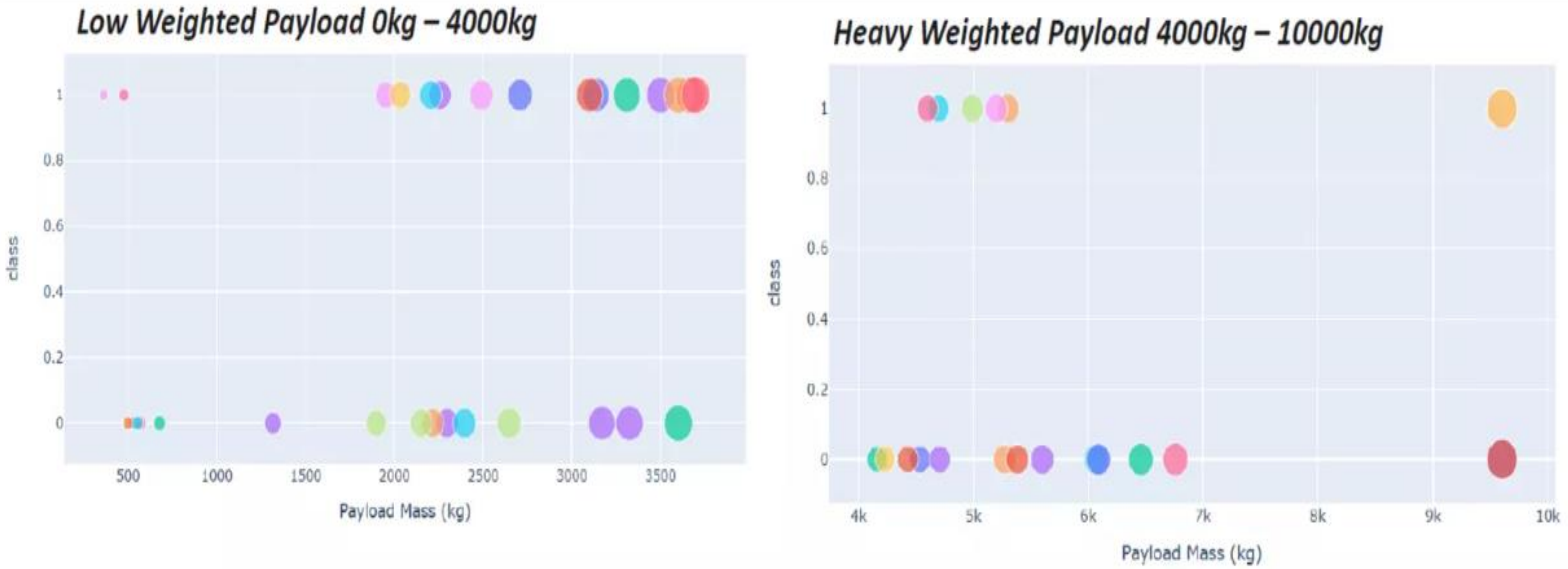
We can see that KSC LC-39A had the most successful launches from all the sites

Dashboard Screenshot 2



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

Dashboard Screenshot 3

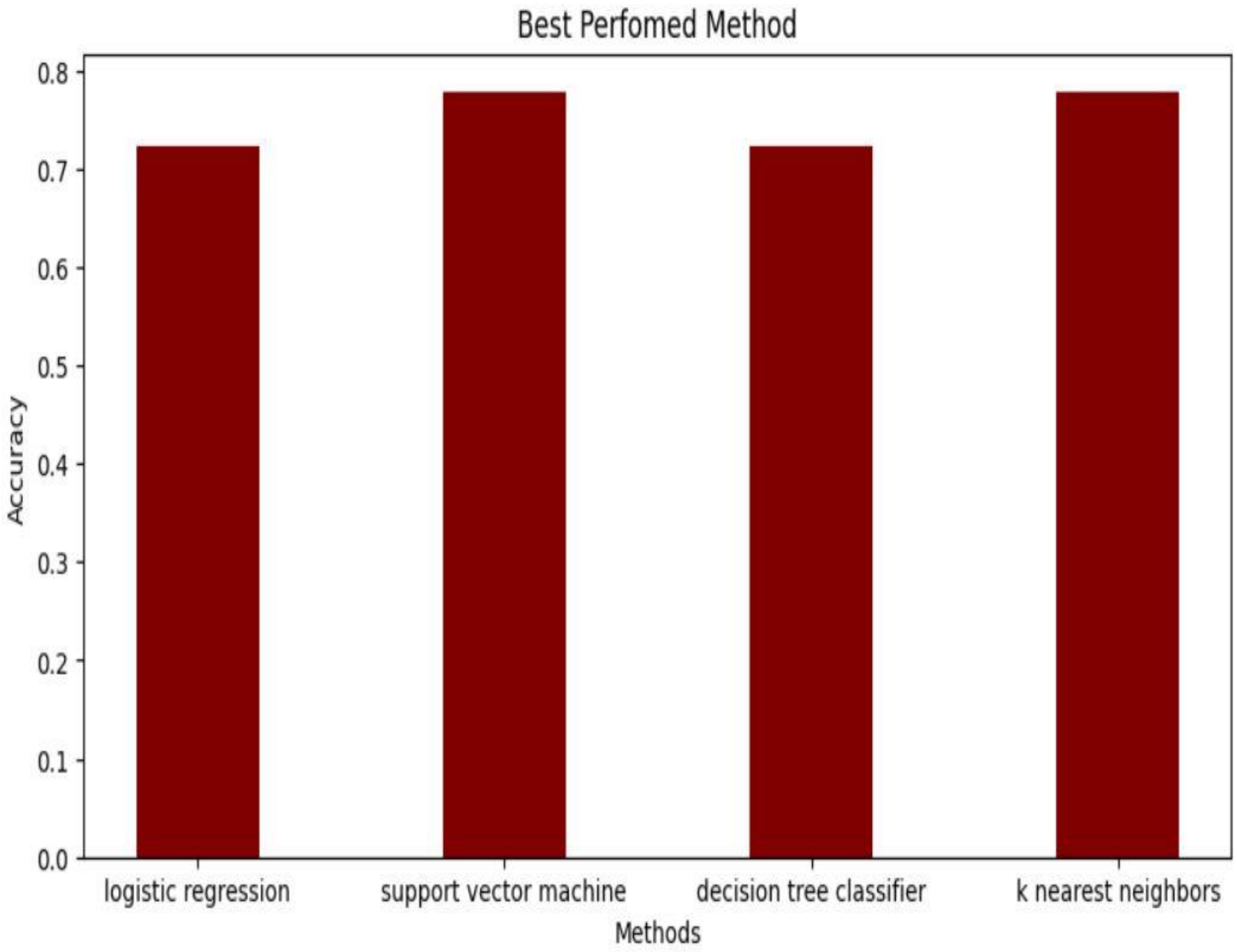


We can see the success rates for low weighted payloads is higher than the heavy weighted payloads

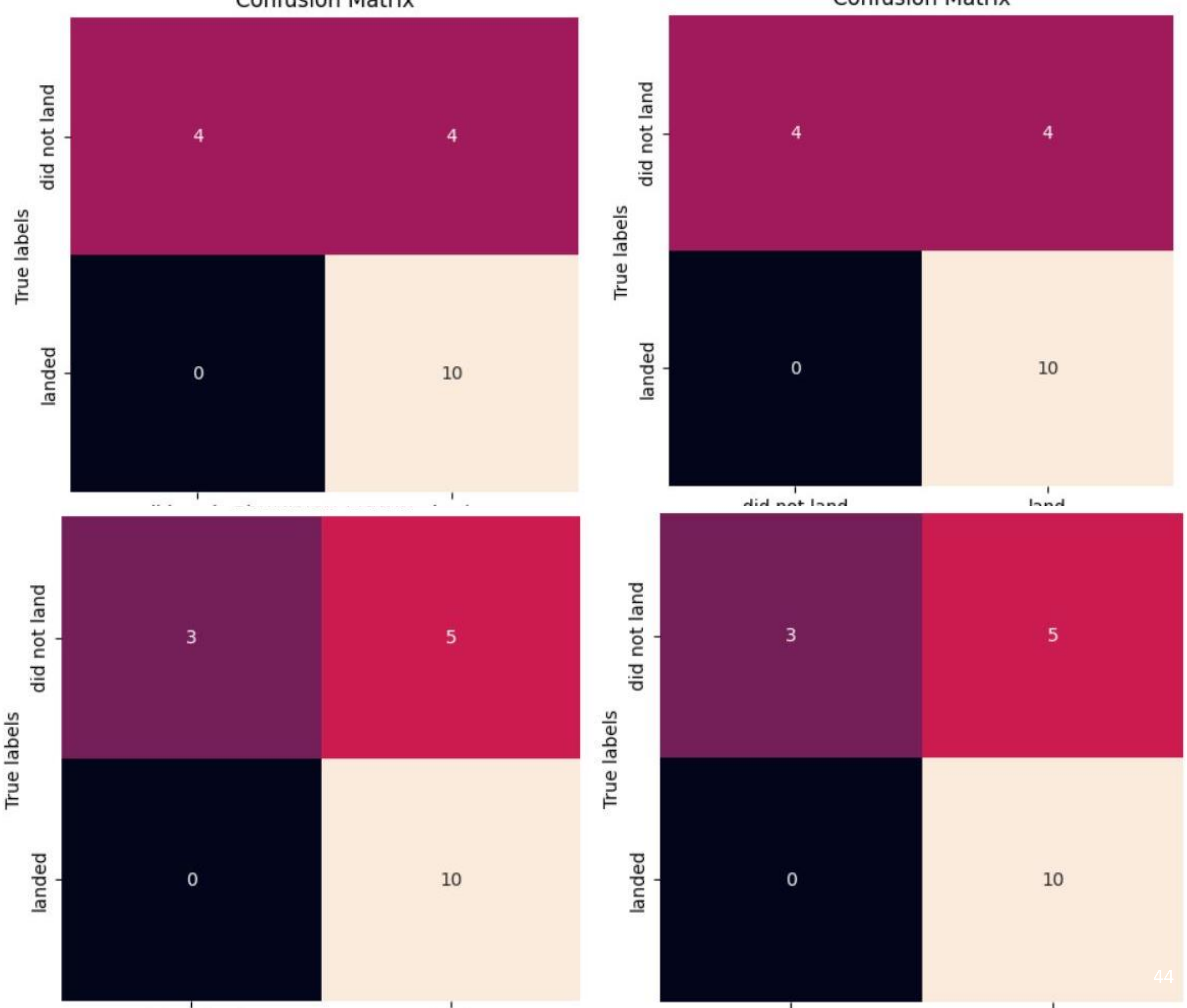
Section 5

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!