IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<DORCELANT Usnaelo>
<July 10th, 2023>

Outline

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

Executive Summary

- Summary of methodologies

-Data collection

-Data wrangling

-Data Visualization

-Data query with SQL

-Interactive Map with Folium

-Dashboard with Plotly Dash

-Classification : Predictive Machine Learning Analysis

- Summary of all results

-Exploratory Data analysis

-Interactive Dashboard analysis

-Machine Learning analysis

Introduction

- Project background and context

- In this project we build models to predict if the falcon 9 first stage will land successfully. The falcon 9 rocket launches are advertising on SpaceX website as one of the most economic-saving technological innovation in space industry. With a cost of 62 millions dollars they cost less than time the traditional cost of other providers which is upward to 165 millions dollars each. Consequently, if we can determine the cost of a launch, we can use this information to bet on SpaceX outcome rocket launch.

- Problems you want to find answers

- What is the nature of the relationship with rocket launches variables and the land outcome?

- What are the factors that have relevant impact on rocket land outcome ?

- What factors can SpaceX improve on to get best results and increase the rocket success landing rate?

Section 1

Methodology

Methodology

- Data collection methodology:

-SpaceX Rest API

-Web Scrapping data from Wikipedia

-Clean requested data

- Perform data wrangling

-Dropping irrelevant data

-Exploratory Data analysis

- Perform exploratory data analysis (EDA) using visualization and SQL

-Using graphics to explore relationship between variables

- Perform interactive visual analytics using Folium and Plotly Dash

- Perform predictive analysis using classification models

-Standardize the data

-split data into training data and test data

- Find best Hyperparameter for SVM, Classification Trees and Logistic Regression

Data Collection

- Describe how data sets were collected.

- First, we worked with SpaceX launch data that was 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/.
- Then we use another popular data source for obtaining Falcon 9 Launch data is web scraping Wikipedia using BeautifulSoup.

GitHub Link

Data Wrangling

- Describe how data were processed

-Calculate the number of the launches at each site

-Calculate the number and occurrence of each site

-Create a landing outcome from outcome column

-Export dataset as .CSV

GitHub Link

EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts

➢ Scatter graphs

-Flight Number VS. Payload Mass
-Flight Number VS. Launch Site
-Payload VS. Launch Site
-Orbit VS. Flight Number
-Payload VS. Orbit Type
-Orbit VS. Payload Mass

Bar Graph
-Mean VS Orbit

Line Graph
Success rate VS Year

- GitHub Link

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed

•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.

- GitHub Link

Build an Interactive Map with Folium

- Summarize :
- Recording of Latitude and Longitude Coordinates at each launch site
- *Establishing Circle Marker around each launch site with a label of the name of the launch site*
- **calculate the distance from the Launch Site to various landmarks to find various trends about what is around the Launch Site**
- Explain why you added those objects
- We ask those objects to discover pattern between those different parameters, we found that as examples that :
- Launch sites are not in close proximity to railroads and highways, there are more in close proximity to coastline and they usually keep a certain distance for cities

- GitHub Link

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard

-We construct the dashboard with flask and dash plotly
-Add a Launch Site Drop-down Input Component
-Add a callback function
-Add a Range Slider to Select Payload
-Add a callback function

- Explain why you added those plots and interactions
- We use those plots and interactions : to render success-pie-chart based on selected site dropdown Select Payload, to render the success-payload-scatter-chart scatter plot

- GitHub Link

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- **BUILDING MODEL**
- •**Load our dataset into NumPy and Pandas**
- •Transform Data
- •Split our data into training and test data sets
- •Check how many test samples we have
- •Decide which type of machine learning algorithms we want to use
- •Set our parameters and algorithms to GridSearchCV
- •Fit our datasets into the GridSearchCVobjects and train our dataset.

- **EVALUATING MODEL**
- •Check accuracy for each model
- •Get tuned hyperparameters for each type of algorithms
- •Plot Confusion Matrix

- **IMPROVING MODEL**
- •Feature Engineering
- •Algorithm Tuning

- **FINDING THE BEST PERFORMING CLASSIFICATION MODEL**
- •The model with the best accuracy score wins the best performing model
- •In the notebook there is a dictionary of algorithms with scores at the bottom of the notebook.

Github link

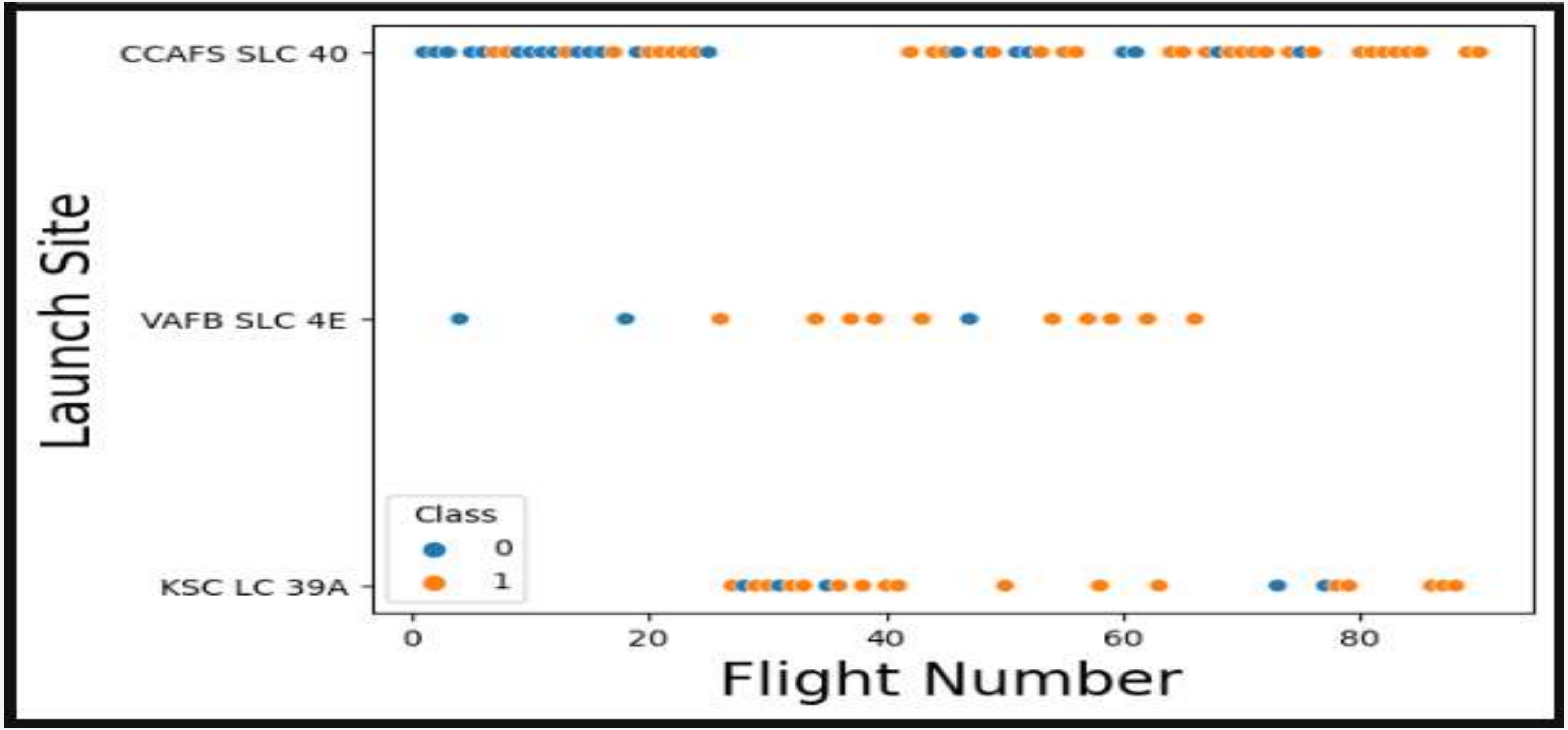
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

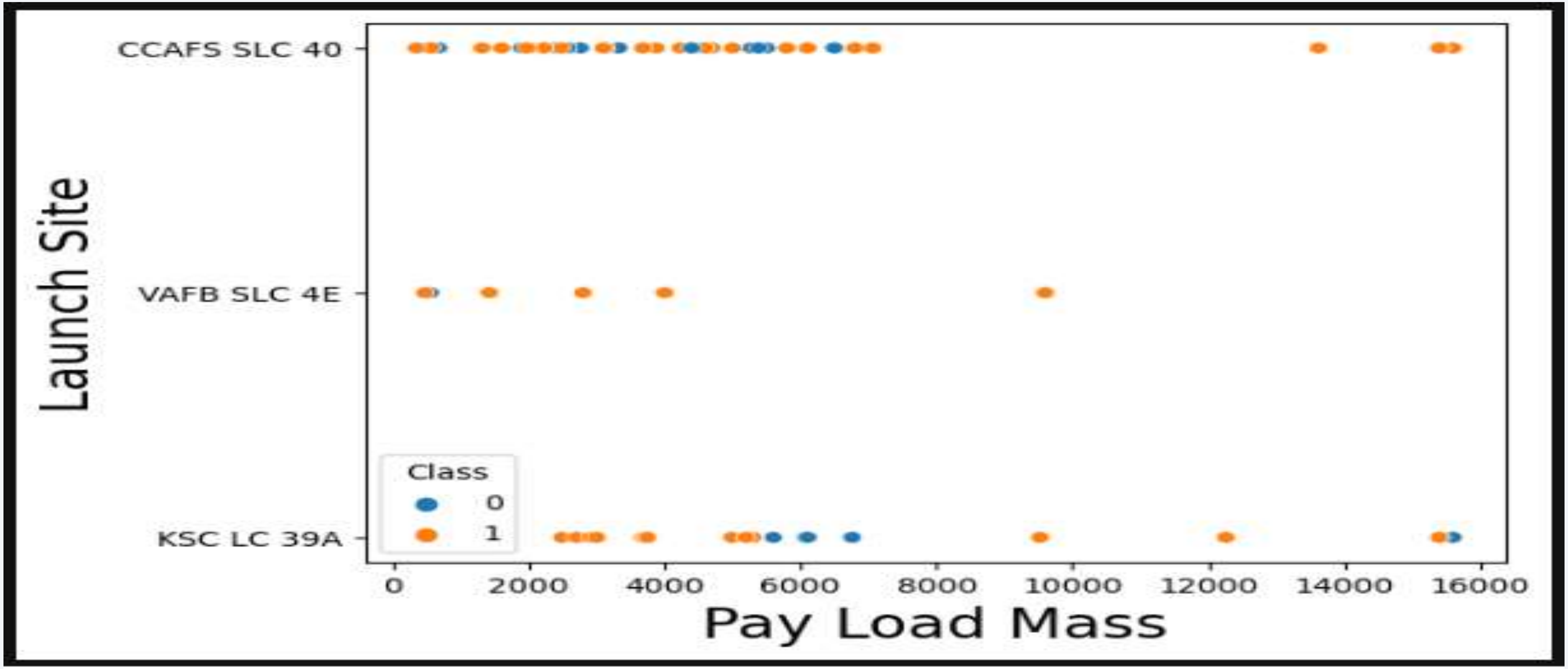
Section 2

Insights drawn from EDA

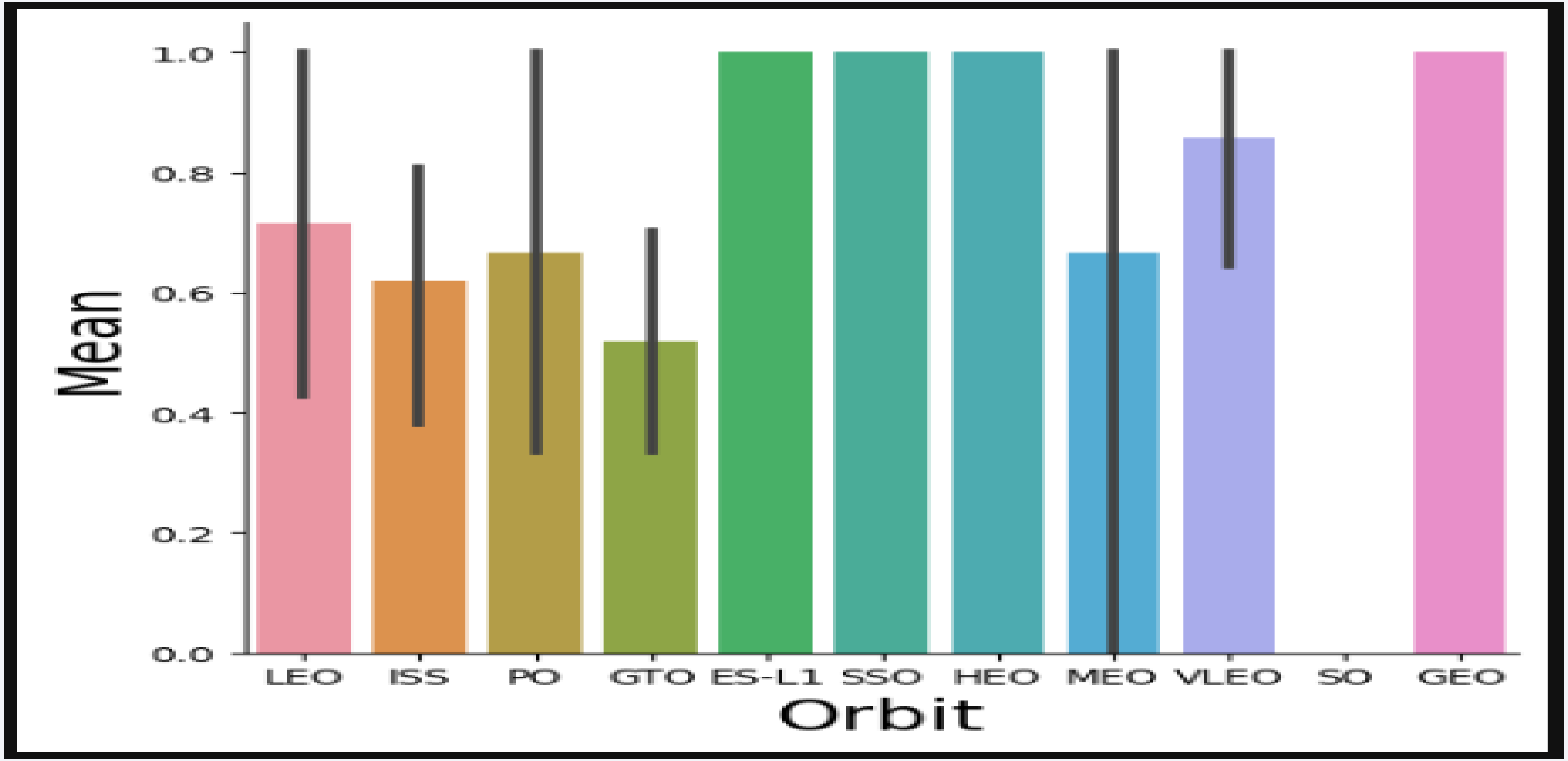
Flight Number vs. Launch Site



Payload vs. Launch Site

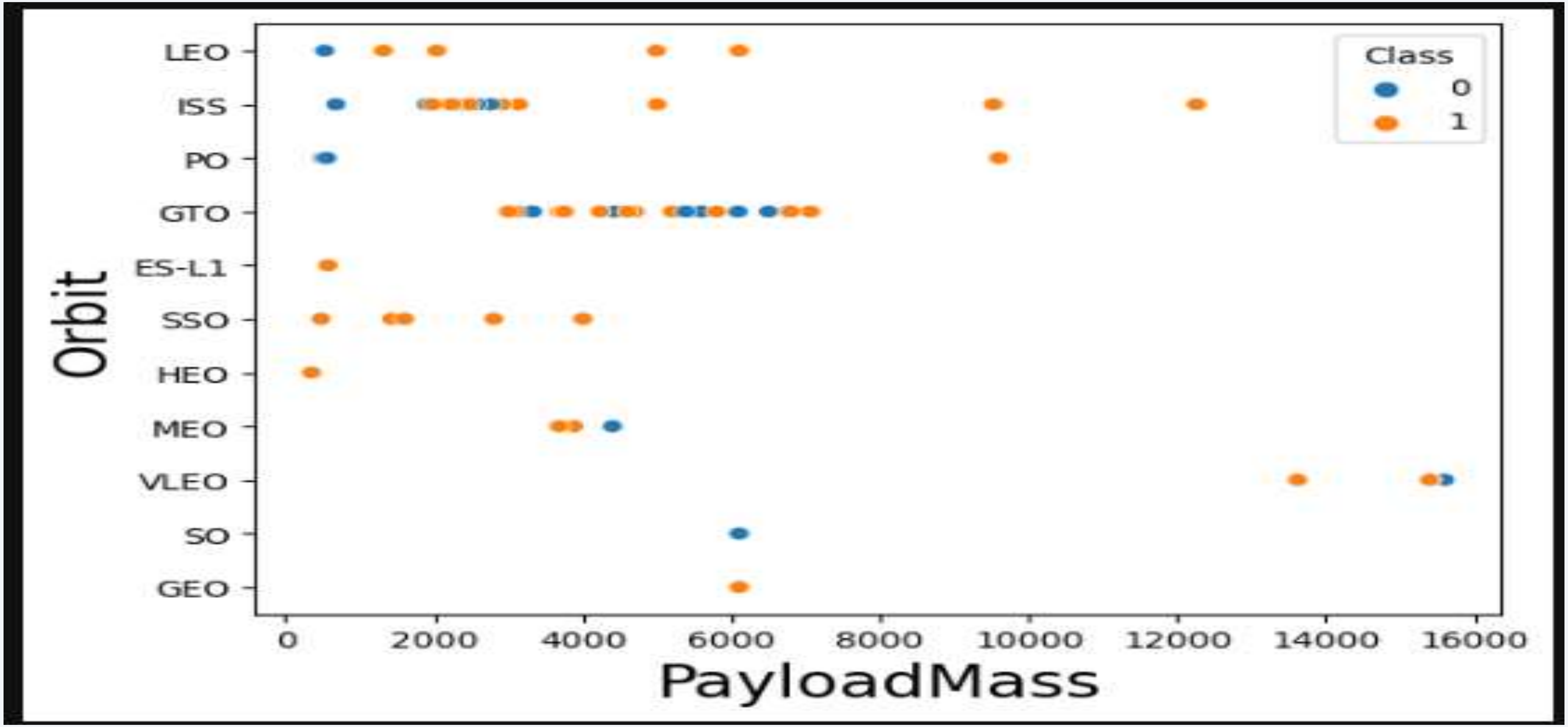


Success Rate vs. Orbit Type

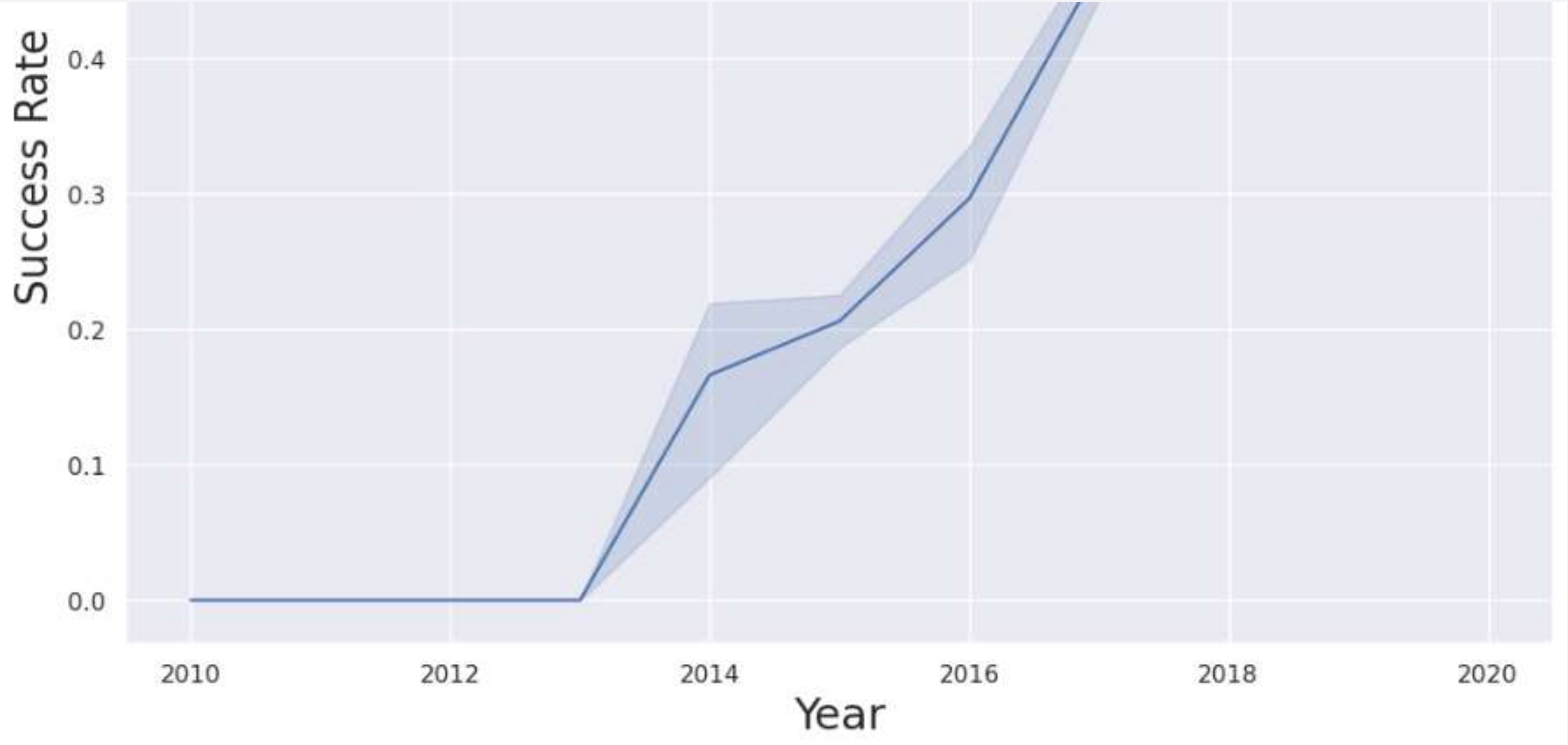


Flight Number vs. Orbit Type

Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

The names of the unique launch sites

```
%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'

5 records where launch sites begin with `CCA`

```
%sql SELECT * FROM SPACEXTBL WHERE Launch_Site LIKE 'KSC%' LIMIT 5
```

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
19/02/2017	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490.0	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
16/03/2017	6:00:00	F9 FT B1030	KSC LC-39A	EchoStar 23	5600.0	GTO	EchoStar	Success	No attempt
30/03/2017	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300.0	GTO	SES	Success	Success (drone ship)
05/01/2017	11:15:00	F9 FT B1032.1	KSC LC-39A	NROL-76	5300.0	LEO	NRO	Success	Success (ground pad)
15/05/2017	23:21:00	F9 FT B1034	KSC LC-39A	Inmarsat-5 F4	6070.0	GTO	Inmarsat	Success	No attempt

Total Payload Mass

Total payload carried by boosters from NASA

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE Customer = 'NASA (CRS)'
```

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

SUM(PAYLOAD_MASS__KG_)
45596.0

Average Payload Mass by F9 v1.1

Average payload mass carried by booster version F9 v1.1

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE Booster_Version='F9 v1.1'
```

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

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

Dates of the first successful landing outcome on ground pad

```
%sql SELECT Date, Landing_Outcome FROM SPACEXTBL WHERE Landing_Outcome="Success (ground pad)"
```

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

Date	Landing_Outcome
22/12/2015	Success (ground pad)
18/07/2016	Success (ground pad)
19/02/2017	Success (ground pad)
05/01/2017	Success (ground pad)
06/03/2017	Success (ground pad)
14/08/2017	Success (ground pad)
09/07/2017	Success (ground pad)
15/12/2017	Success (ground pad)
01/08/2018	Success (ground pad)

Successful Drone Ship Landing with Payload between 4000 and 6000

- Boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

```
%%sql SELECT Booster_Version, Landing_Outcome, PAYLOAD_MASS__KG_ FROM SPACEXTBL
WHERE Landing_Outcome="Success (drone ship)" AND (PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000)
```

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

Booster_Version	Landing_Outcome	PAYLOAD_MASS__KG_
F9 FT B1022	Success (drone ship)	4696.0
F9 FT B1026	Success (drone ship)	4600.0
F9 FT B1021.2	Success (drone ship)	5300.0
F9 FT B1031.2	Success (drone ship)	5200.0

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

```
%%sql SELECT COUNT(*), Mission_Outcome FROM SPACEXTBL
GROUP BY "Mission_Outcome"
```

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

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

Boosters Carried Maximum Payload

- Booster which have carried the maximum payload mass

```
%%sql
SELECT DISTINCT Booster_Version, PAYLOAD_MASS__KG_ FROM SPACEXTBL WHERE (SELECT MAX(PAYLOAD_MASS__KG_ ) FROM SPACEXTBL)
```

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

Booster_Version	PAYLOAD_MASS__KG_
F9 v1.0 B0003	0.0
F9 v1.0 B0004	0.0
F9 v1.0 B0005	525.0
F9 v1.0 B0006	500.0
F9 v1.0 B0007	677.0
F9 v1.1 B1003	500.0
F9 v1.1	3170.0
F9 v1.1	3325.0
F9 v1.1	2296.0
F9 v1.1	1316.0
F9 v1.1	4535.0
F9 v1.1 B1011	4428.0

2015 Launch Records

- landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

```
%%sql
SELECT (Landing_Outcome), Booster_version, Launch_Site FROM SPACEXTBL
WHERE substr("Date",7,4)='2015' AND Landing_Outcome = "Failure (drone ship)"
ORDER BY Landing_Outcome DESC
```

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

Landing_Outcome	Booster_Version	Launch_Site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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)) between the date 2010-06-04 and 2017-03-20, in descending order

```
%%sql
SELECT COUNT(Landing_Outcome), Landing_Outcome FROM SPACEXTBL
WHERE(("Date">"2010/06/04") AND ("Date">"2017/03/20")) AND Landing_Outcome LIKE "Succes%" OR "Fail%"
ORDER BY Landing_Outcome DESC
```

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

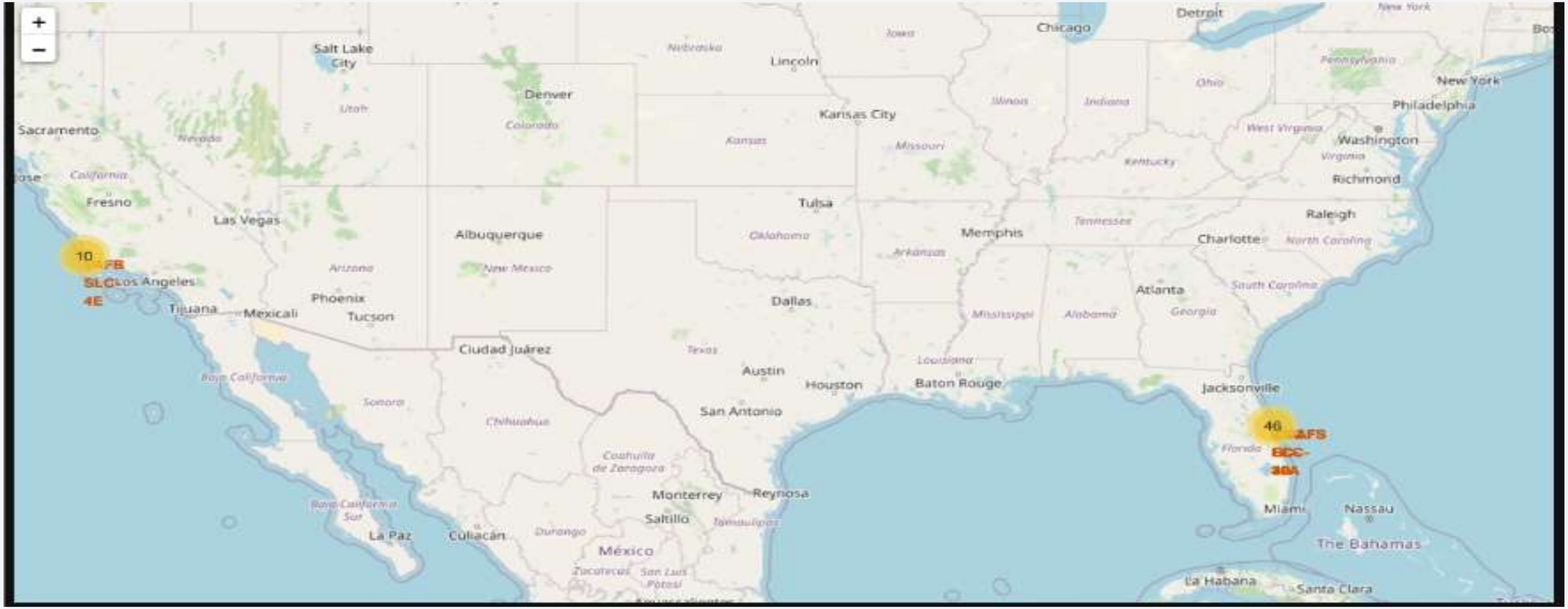
COUNT(Landing_Outcome)	Landing_Outcome
20	Success (ground pad)

Section 3

Launch Sites Proximities Analysis

Launch Site Location

- We can observe that there's more site on the east coast then on the west coast.



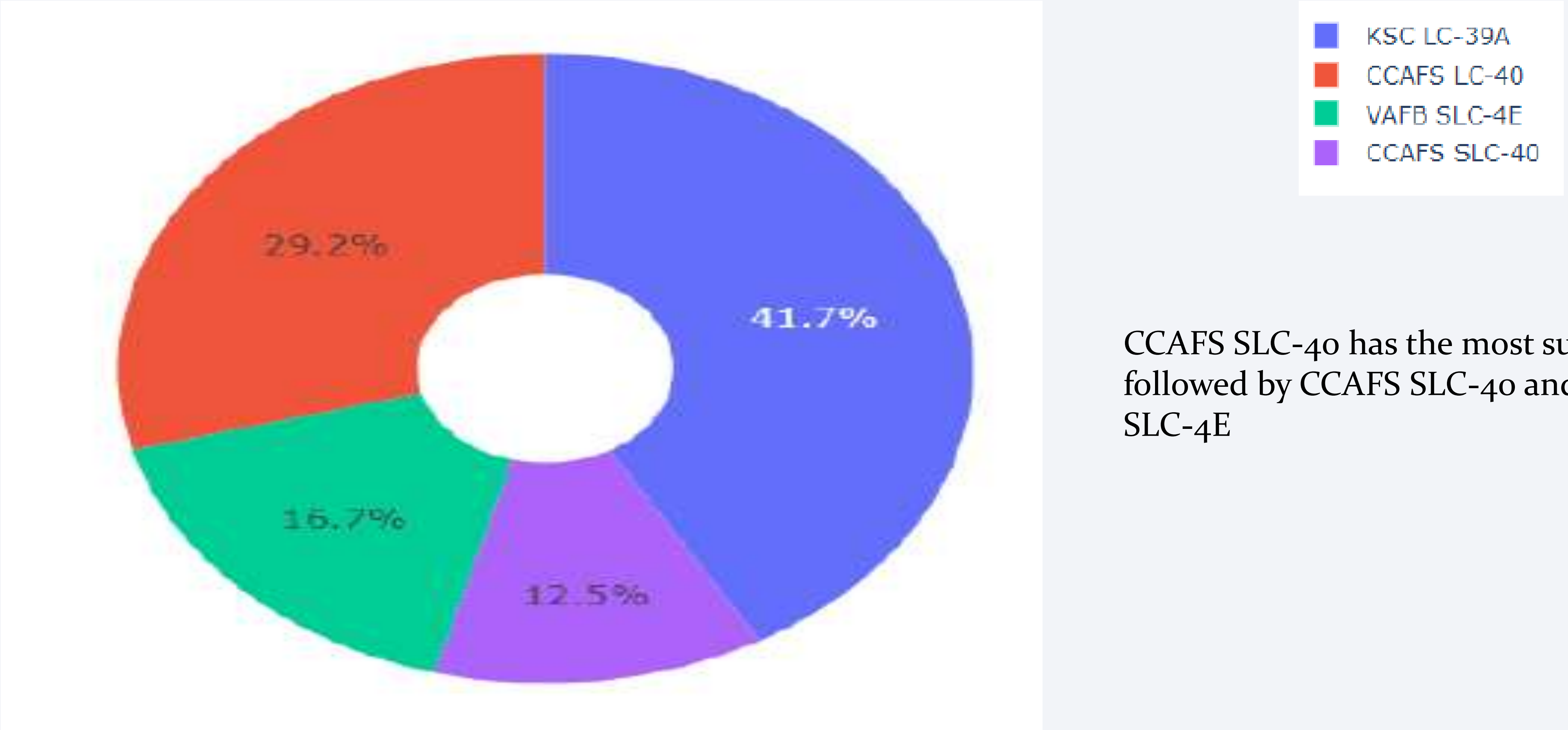
Launch Outcome Mapped

Launch Site Location and Proximities

Section 4

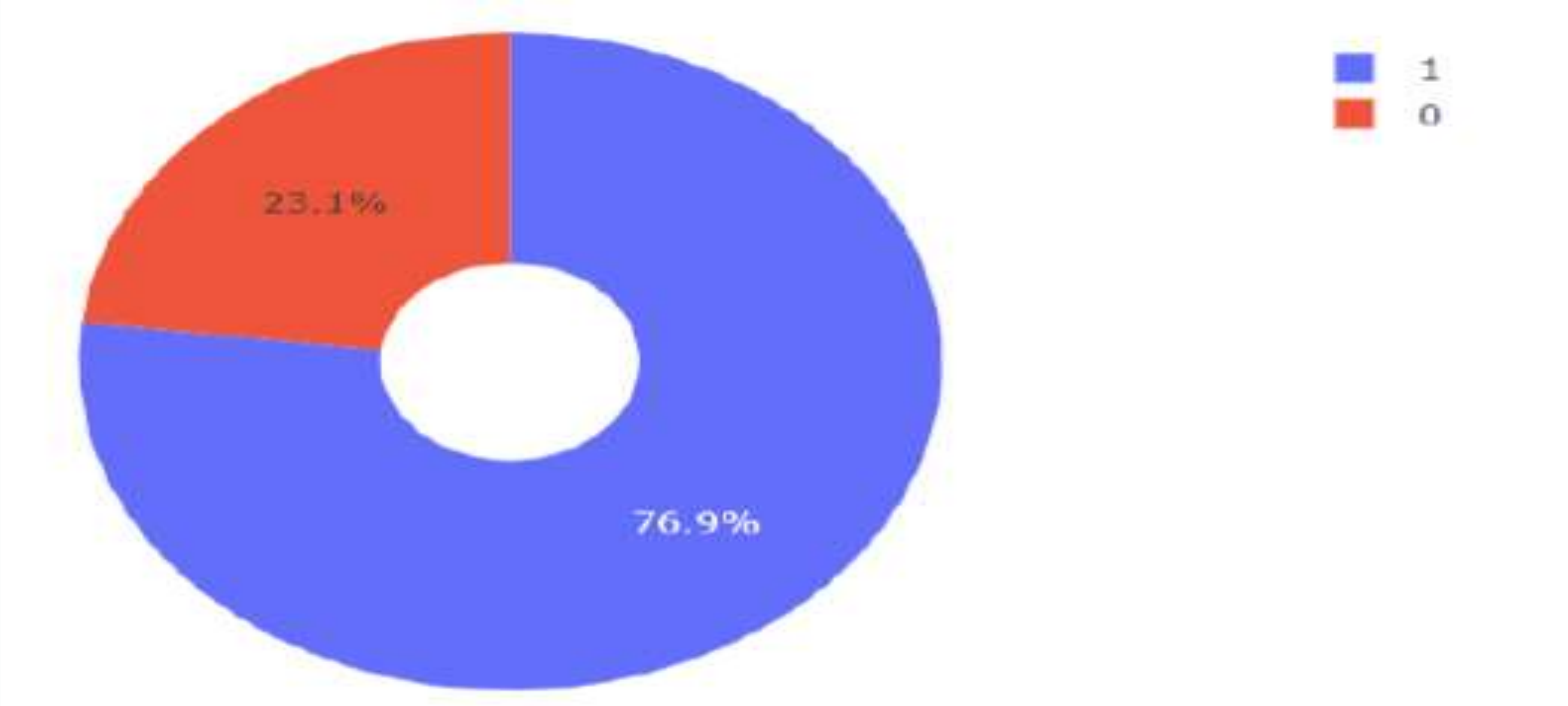
Build a Dashboard with Plotly Dash

Success Part for each Launch Site



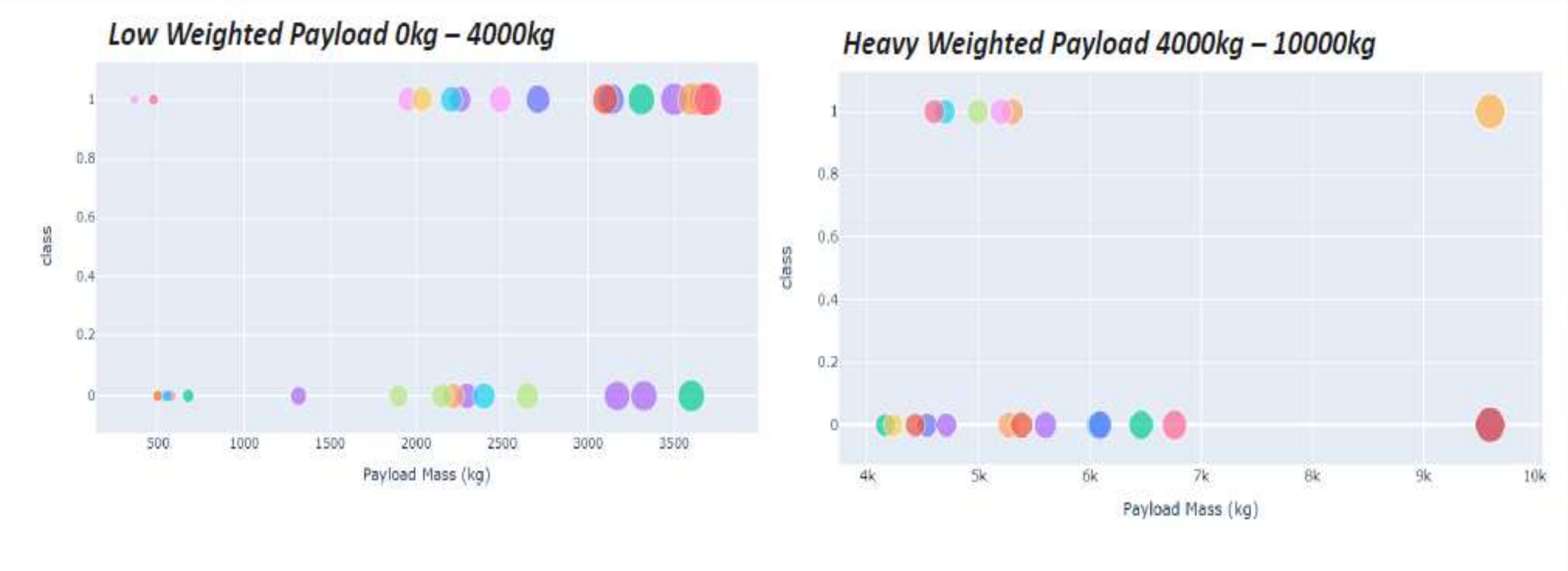
CCAFS SLC-40 has the most success followed by CCAFS SLC-40 and VAFB SLC-4E

Launch Site access with the max ratio



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

Payload Vs Launch Outcome Dash

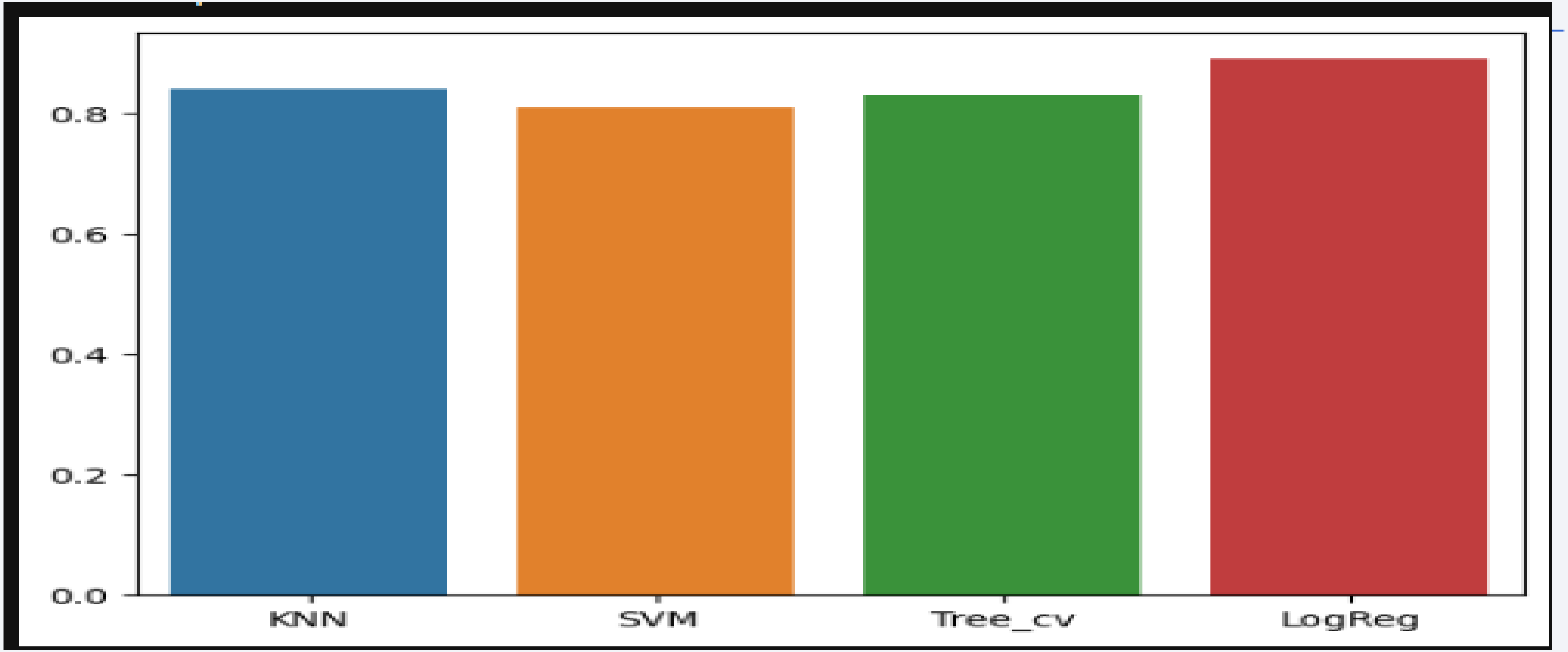


The success rates for low weighted payloads is lower than the heavy weighted payloads

Section 5

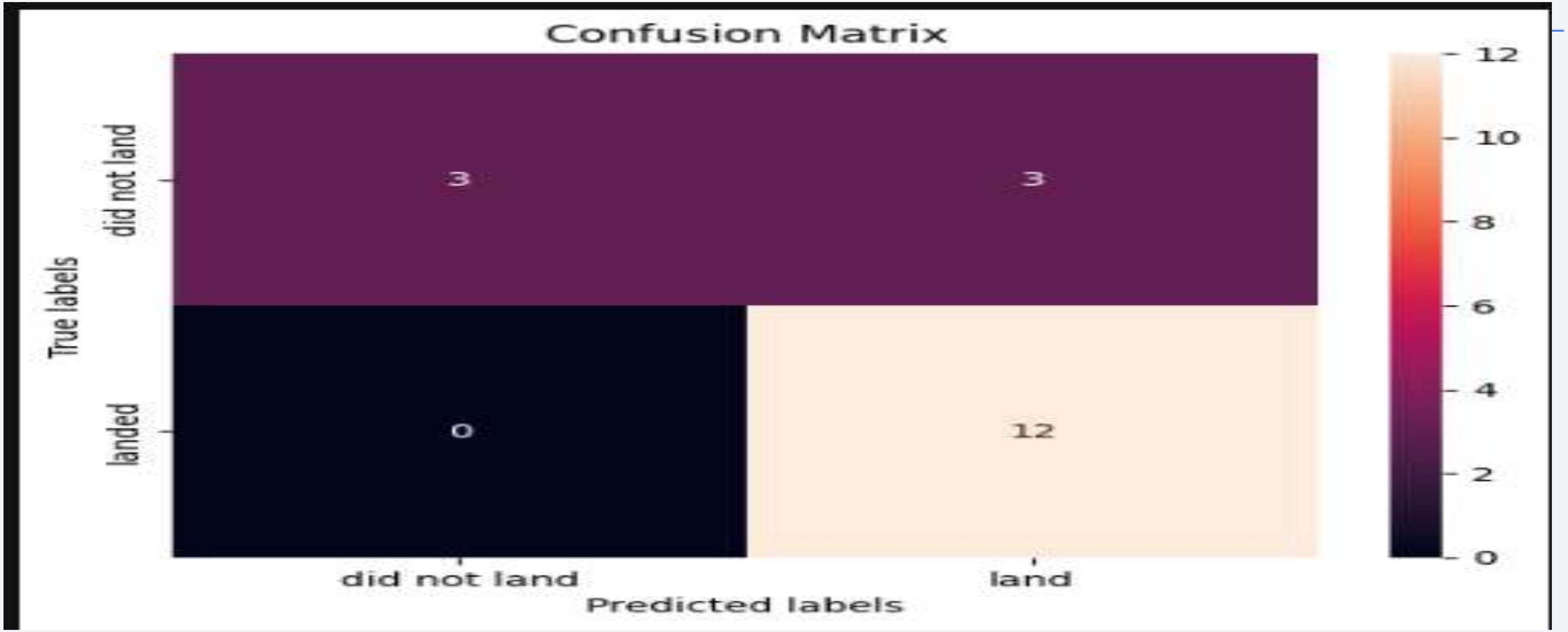
Predictive Analysis (Classification)

Classification Accuracy



- Best Algorithm is Tree with a score of 0.8892857142857142 Best Params is : {'criterion': 'gini', 'max_depth': 14, 'max_features': 'sqrt', 'min_samples_leaf': 4, 'min_samples_split': 2, 'splitter': 'random'}

Confusion Matrix



Conclusions

- Tree Classifier Algorithm is the best for Machine Learning for this dataset
- Low weighted payloads perform better than the heavier payloads
- Orbit GEO,HEO,SSO,ES-L1 has the best Success Rate
- 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
- CCAFS SLC-40 has the most success followed by CCAFS SLC-40 and VAFB SLC-4E
- ***success rates for low weighted payloads is lower than the heavy weighted payloads***

Appendix

- Pandas Documentation
- Seaborn Documentation
- SQLite
- SQLserver
- Dash Library
- Folium Library

Thank you!