IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Akeem Davis
Dec. 27, 2023

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection with API and Web Scraping
  - Data Wrangling(Cleaning)
  - Data Analysis with SQL
  - Data Visualization
  - Interactive Maps with Folium
  - Live Dashboards With Dash
  - Model Prediction with Machine Learning
- Summary of all results
  - Exploring Data Analysis Results
  - Interactive Analysis Visualization
  - Predictive Model Results

Introduction

- Project background and context
  - SpaceX's Falcon 9 Rocket launches at a cost of $62M in comparison to other providers which can cost anywhere from $165M upwards. SpaceX is able to accomplish this by reusing the first stage. This means if we can determine if the first stage will land successfully we can also determine the cost of a launch. This data can be helpful if other companies choose to bid against SpaceX for a rocket launch.
- Problems you want to find answers
  - What can we use to determine if a rocket launch was successful?
  - What are the factors that might affect a successful landing?
  - Are there any conditions that must be met for a successful landing?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected using the SpaceX API and Web Scraping.
- Perform data wrangling
  - The data was processed with: One Hot-Encoding was applied to categorical features.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models.

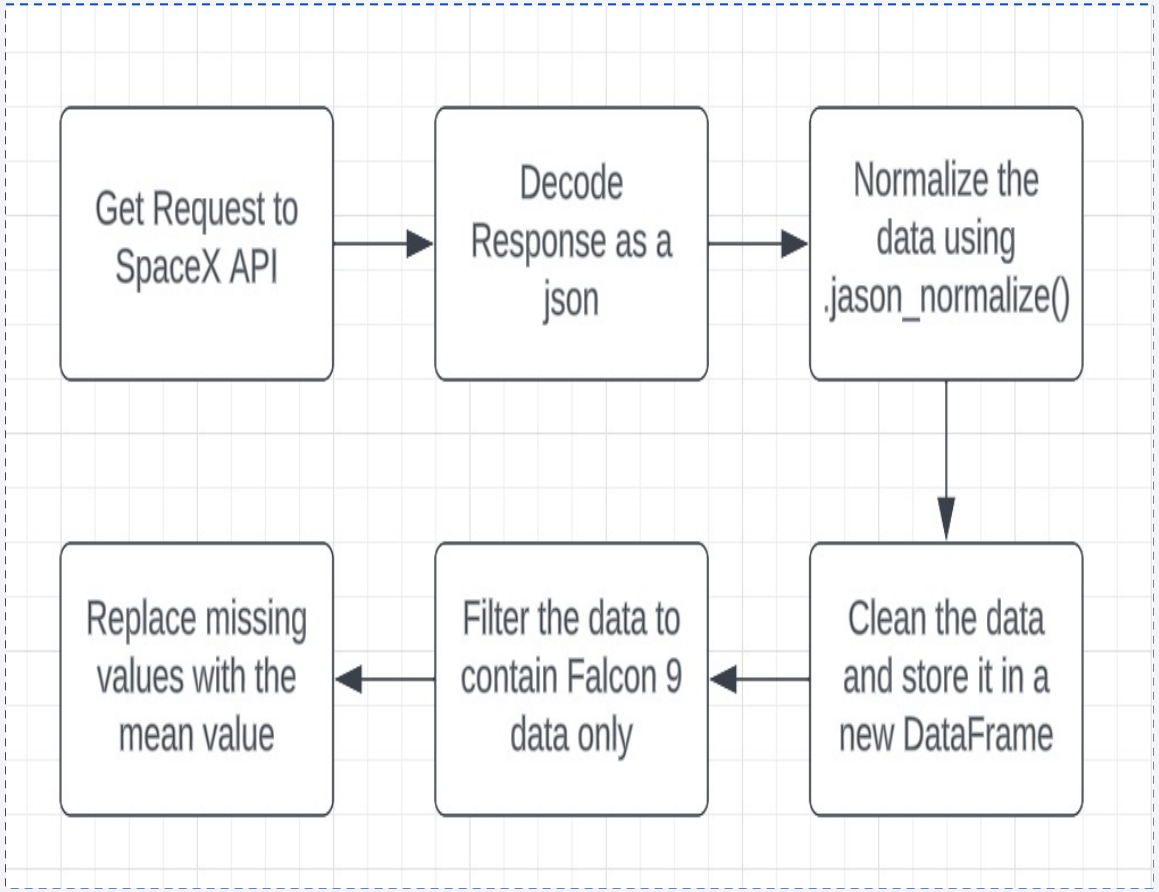
Data Collection

- Describe how data sets were collected.
  - Data collection was due with the SpaceX API. The data was then decoded as a Json with the help of the ".json()" function, the data was normalized using ".json_normalize()". The data was placed into a DataFrame using pandas. The Falcon Launch Data was also collected using Web Scraping through the use of BeautifulSoup. This data was then parsed and placed into a pandas data frame.
  - The dataframe was cleaned – the data was filtered unwanted launch data was removed. The data frame was filter to only include falcon 9 data only. All null values were observed and all missing values were checked and filled where necessary (Null values were replaced).
- You need to present your data collection process use key phrases and flowcharts
  - Please See next Slides

Data Collection – SpaceX API

- A get request was used to collect data using the SpaceX Rest API, the data was returned and stored in a data frame. The data was cleaned and filter to only include Falcon 9 Launch Data, missing values were also replaced.
- Click here to see the Notebook.

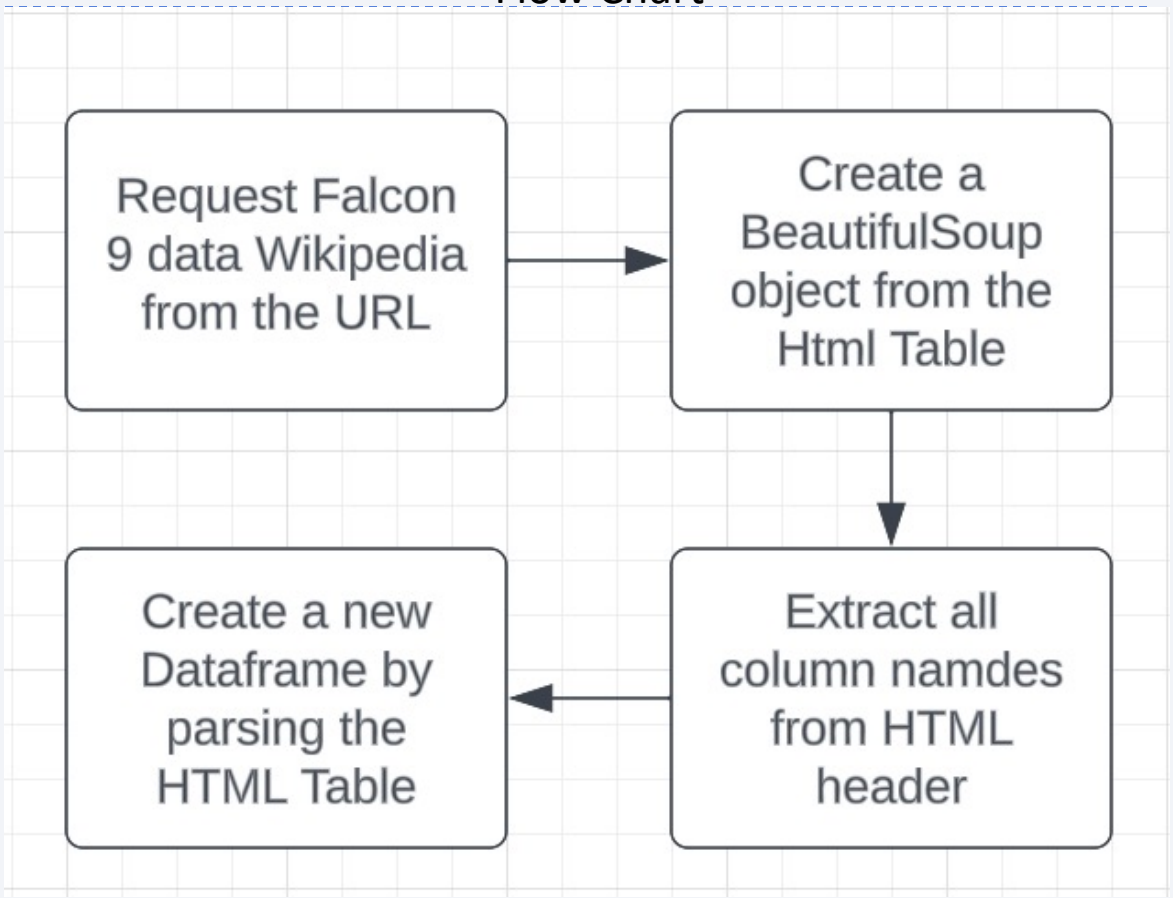
Flow Chart



Data Collection - Scraping

- Web scraping was done on the Falcon 9 Wikipedia page through the use of BeautifulSoup. The data was parsed and converted to a pandas Dataframe
- Click here to view the notebook.

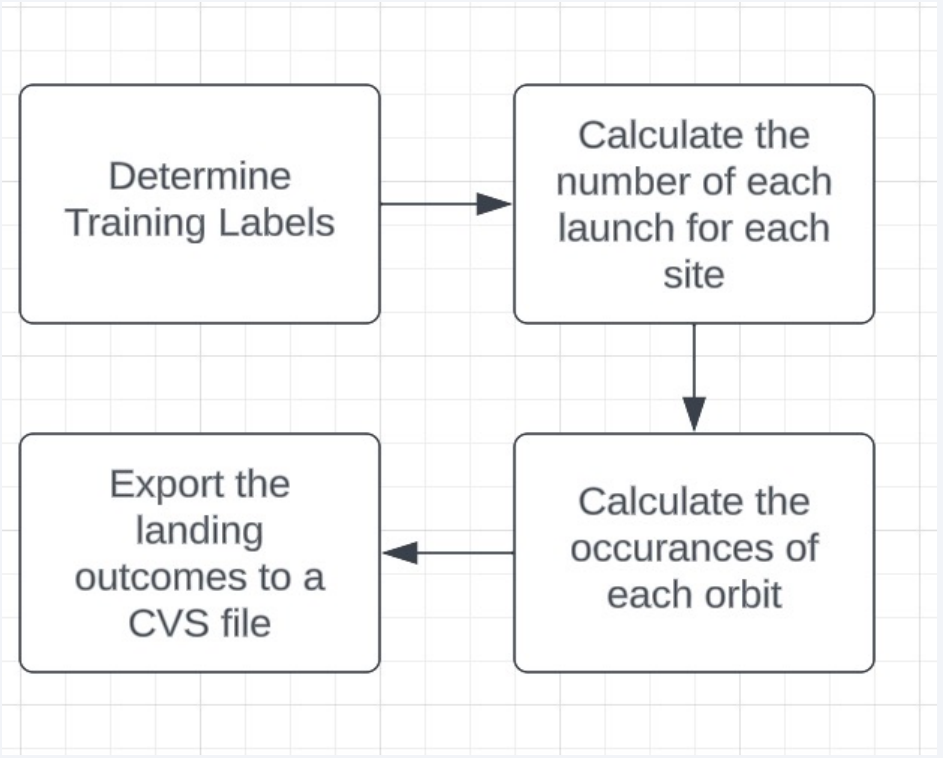
Flow Chart



Data Wrangling

- Exploratory Data Analysis was done to determine our Training labels. The number of launches were also calculated for each launch site along with the occurrences of each orbit. After which landing outcome labels were created and exported as a CSV file.
- Click here to view the Notebook.

Flow Chart



EDA with Data Visualization

- Data Visualizations were done to visualize the relationship between:
  - Flight number and Launch Site
  - Payload and Launch Sites
  - Success Rate of each Orbit type
  - Flight number and Orbit Types
  - Launch Success yearly Trends
- Click here to view Notebook.

EDA with SQL

- The SpaceX data was loaded into a Sql database.
- EDA with Sql was applied to the Dataset and Queries were made to find the
  - Name of launch sites.
  - Total Payload Mass carried by Boosters launch by Nasa(CRS).
  - Average Payload mass carried by Booster Version F9 V1.1.
  - Number of Successful / Failed mission outcomes.
  - Failed landing attempts outcome by drone ship, launch site named and booster version.
- Click here to view the Notebook.

Build an Interactive Map with Folium

- With the help of folium all the launch sites were marked, map objects were also added to depict the success/fail (1/0) of all launches for each site.Success was depicted by 1 and Failure by 0. Colored and cluster labels were also added to the maps.
- Objects were added to make the maps more colorful and easier to understand.
- Click here to view the notebook.

Build a Dashboard with Plotly Dash

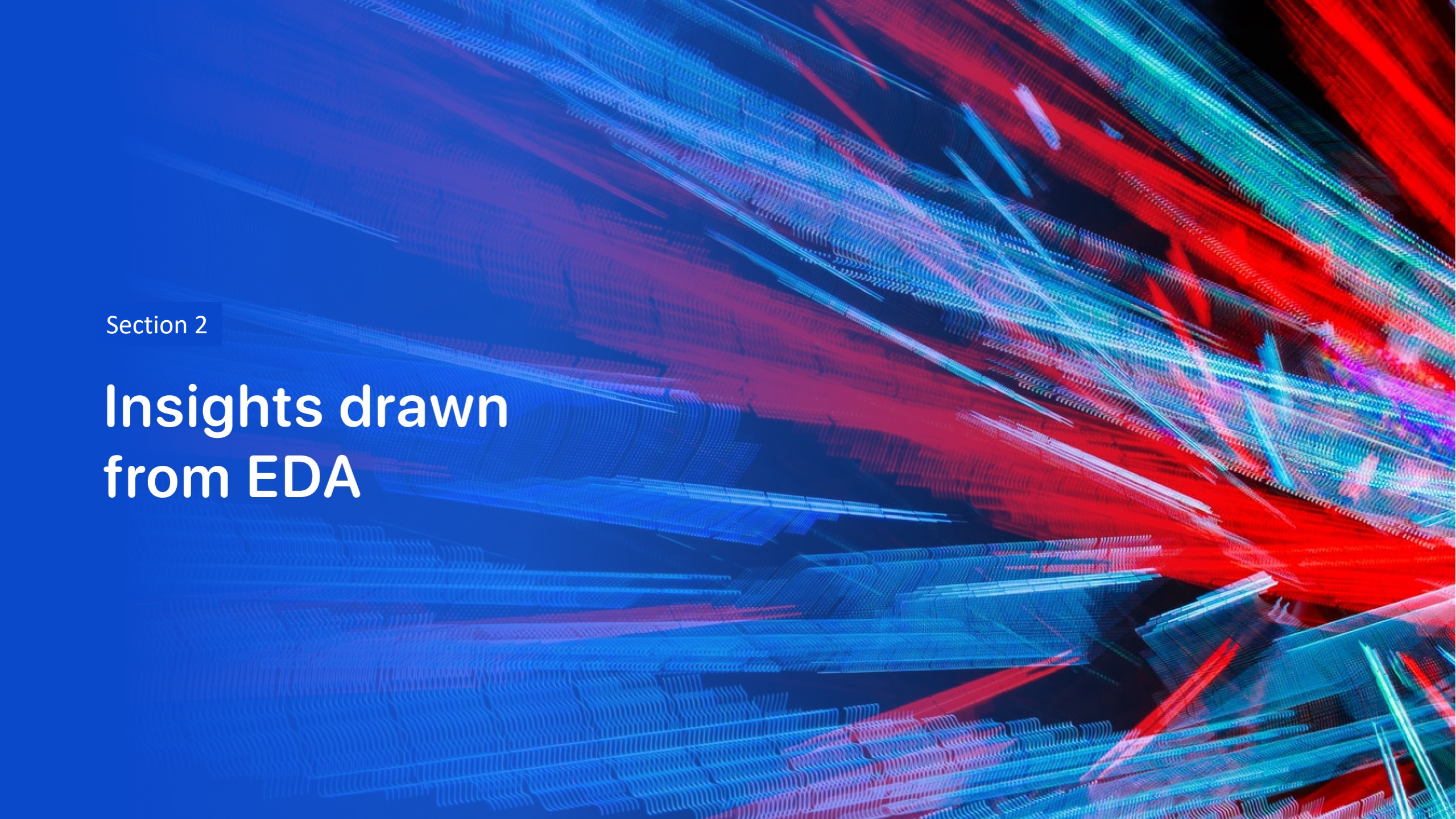
- Interactive Dashboards were created with Plotly Dash.
  - Pie Charts showing the total launches by each site.
  - Scatter Plot displaying the relationship of Outcome and PayLoad Mass in kg for each Booster Version.
- These Plots were selected because it allows us to have a true view of the data.
- Click Here To View The Lab.

Predictive Analysis (Classification)

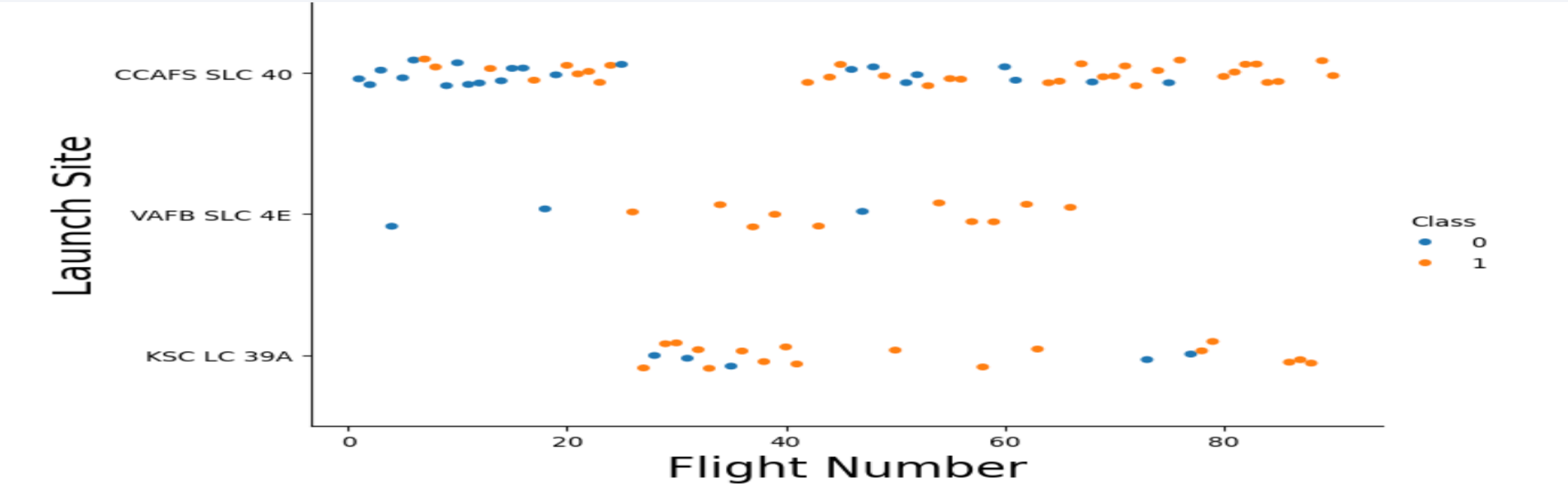
- The data was loaded using Pandas and Numpy, the data was transformed and split into training and testing data
- A few different machine learning models were created, the best model was found to be DecisionTree model with a score of 0.9.
- Click Here to View The Lab.

Results

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

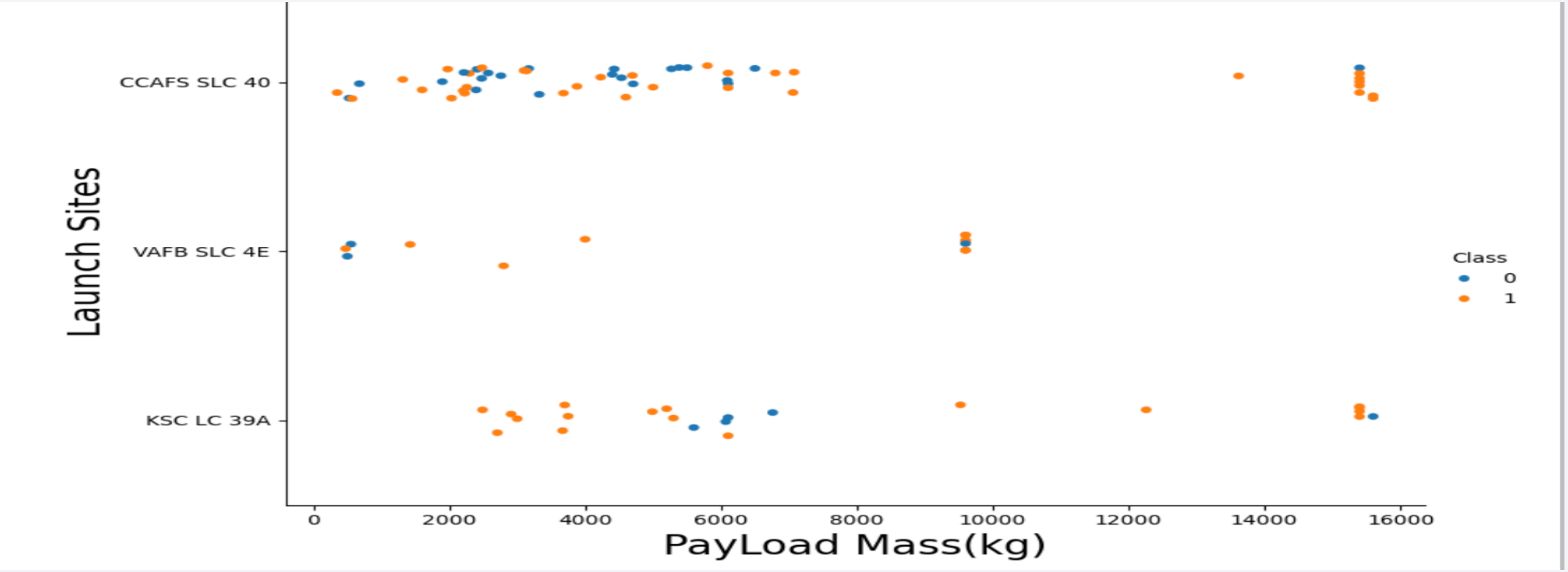

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



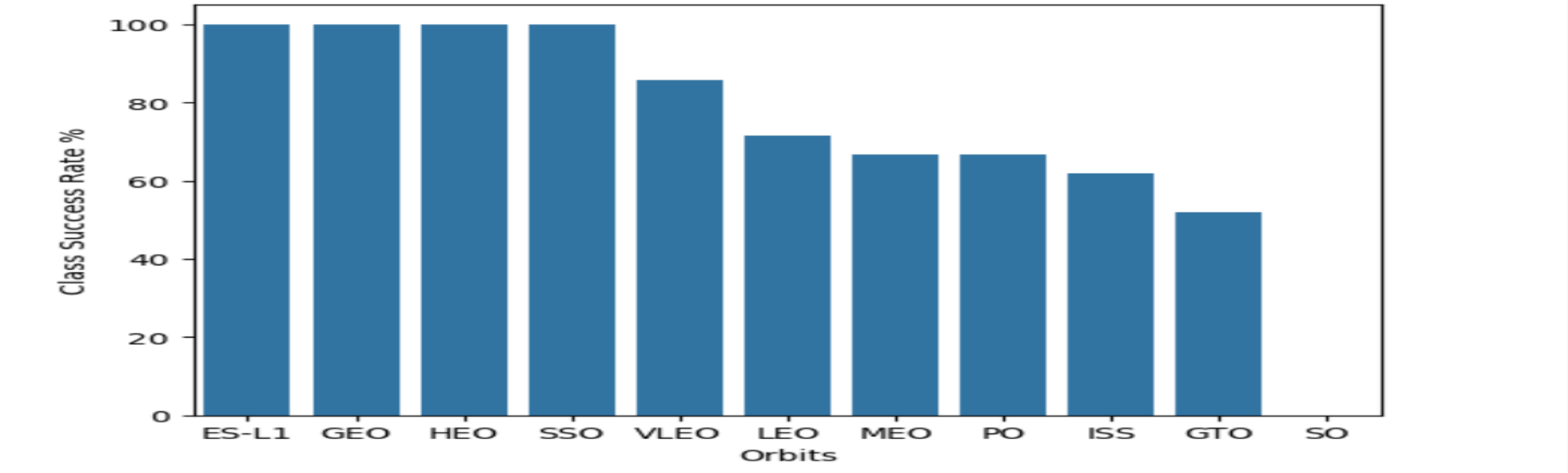
- When we look at this plot we see where as the flight amount increases so does the success rate at each launch site.

Payload vs. Launch Site



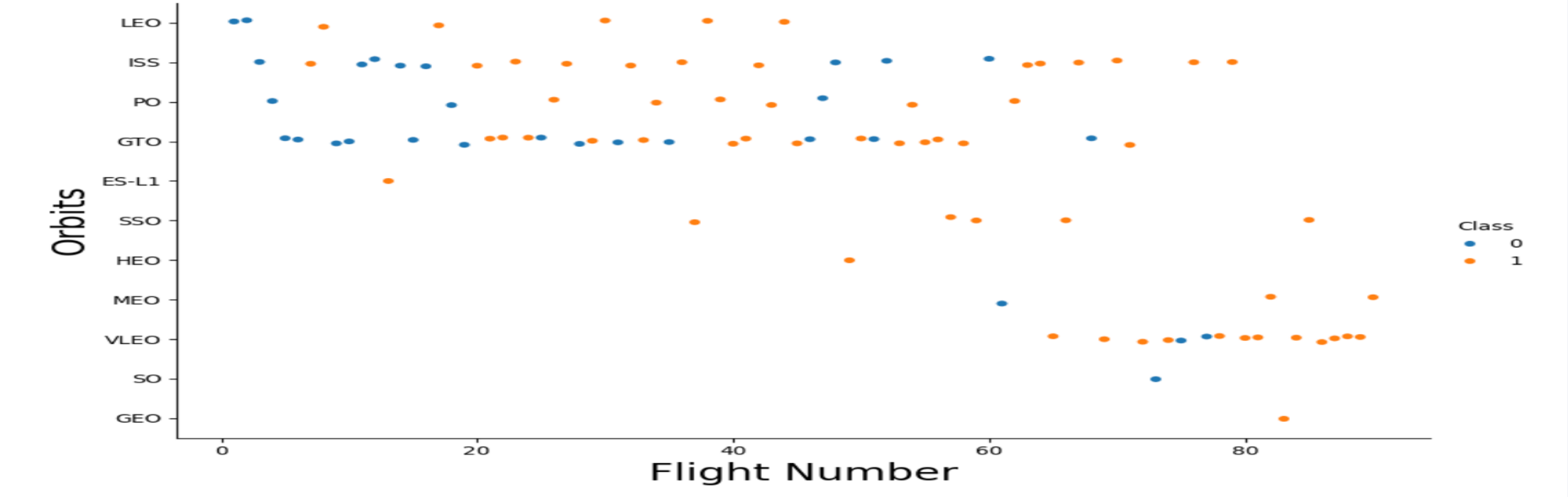
- From this plot we see that the greater the payload mass for CCAFS SLC 40 Site the higher the success rate.

Success Rate vs. Orbit Type



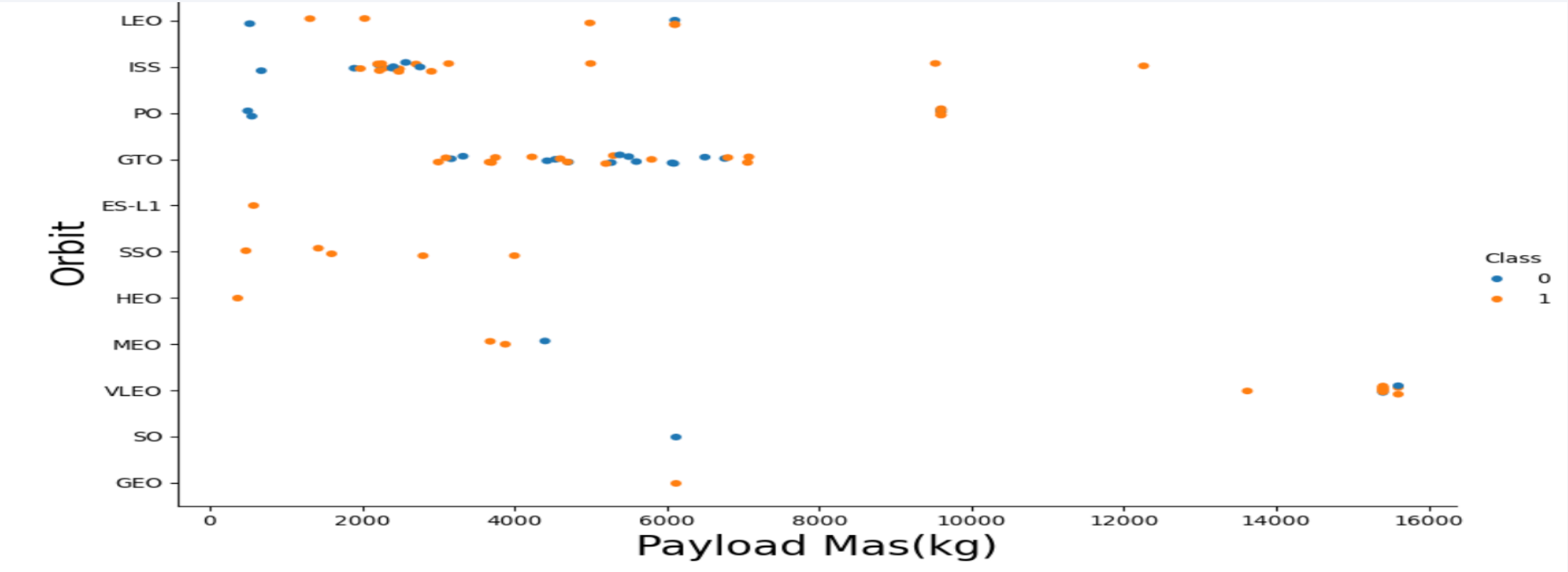
- The Es-L1,Geo, Heo and SSo orbits has the highest success rates of 100%. We also see where SO has the lowest success rate of 0%

Flight Number vs. Orbit Type



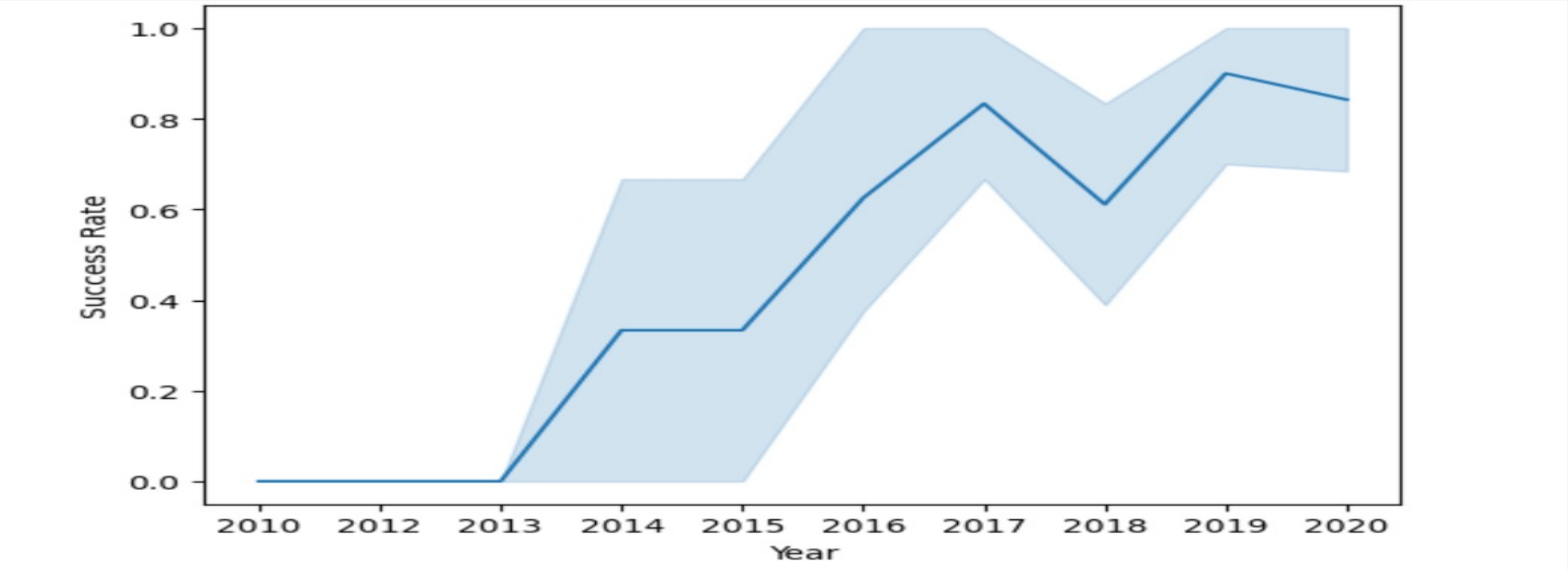
- Here you can see that in the LEO orbit the success is related to # of flights. On the other hand the GEO orbit there is no clear relation between orbit and # of flights.

Payload vs. Orbit Type



- We see here where heavey Payload results in more success for PO, ISS and LEO orbit.

Launch Success Yearly Trend



- The line plot depicts that there has been a drastically increase win success rate from 2013 to 2020.(This is good)

All Launch Site Names

```
[8]: %sql select DISTINCT launch_site from SPACEXTABLE
 * sqlite:///my_data1.db
Done.
```

[8]:

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

- Here we use the Distinct statement to return a list of unique launch sites from the database.

Launch Site Names Begin with 'CCA'

```
[9]: %sql select * from SPACEXTABLE where launch_site like 'CCA%' limit 5
 * sqlite:///my_data1.db
Done.
```

[9]:

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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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

- Here we display 5 records from the CCA launch site.

Total Payload Mass

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

```
[13]: %sql select sum(PAYLOAD_MASS__KG_) from SPACEXTABLE where customer = 'NASA (CRS)'
```

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

[13]:

sum(PAYLOAD_MASS__KG_)
45596

- The Total payload mass from NASA CRS is calculate to be 45596 KG.

Average Payload Mass by F9 v1.1

```
[14]: %sql select AVG(PAYLOAD_MASS__KG_) from SPACEXTABLE where Booster_Version = 'F9 v1.1'

 * sqlite:///my_data1.db
Done.

[14]: AVG(PAYLOAD_MASS__KG_)
      2928.4
```

- The Average Payload Mass carried by Booster Version F9 v1.1 is calculated and displayed here to be 2928.4KG.

First Successful Ground Landing Date

```
[15]: %sql select min(DATE) from SPACEXTABLE where Landing_Outcome = 'Success (ground pad)'

 * sqlite:///my_data1.db
Done.

[15]: min(DATE)
      2015-12-22
```

- Here the first successful ground landing date was calculated to be 2015-12-22 with the use of the min() function.

Successful Drone Ship Landing with Payload between 4000 and 6000

```
[17]: %%sql select DISTINCT Booster_Version, Payload_Mass__KG_ from SPACEXTABLE where PAYLOAD_MASS__KG_
      between 4000 and 6000 and Landing_Outcome = 'Success (drone ship)'
```

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

[17]:

Booster_Version	PAYLOAD_MASS__KG_
F9 FT B1022	4696
F9 FT B1026	4600
F9 FT B1021.2	5300
F9 FT B1031.2	5200

- The Where Clause is used to filter Boosters with successful landing on drone ships, then the and condition was used apply additional filter to get the payload mass greater than 4000 and less than 6000

Total Number of Successful and Failure Mission Outcomes

```
[20]: %sql select Mission_Outcome, count(*) from SPACEXTABLE group by Mission_Outcome
 * sqlite:///my_data1.db
Done.
```

[20]:

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

- Displayed here is the number of successful and mission failure outcomes.

Boosters Carried Maximum Payload

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

```
[16]: %%sql select DISTINCT Booster_Version from SPACEXTABLE where PAYLOAD_MASS__KG_
= (select Max(PAYLOAD_MASS__KG_) from SPACEXTABLE)
```

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

[16]:

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

- Through the help of the Where clause and the max function is displayed the boosters that carried the max payload.

2015 Launch Records

```
[20]: %%sql select substr(Date,6,2) as month, Date,Booster_Version,Launch_Site,
Landing_Outcome from SPACEXTABLE where Landing_Outcome = 'Failure (drone ship)'
and substr(Date,0,5) = '2015'
```

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

[20]:

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

- Here a query is made to reveal launch records for 2015 through the use of Where, Like , And , and Between.

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

```
[21]: %%sql select Landing_Outcome, count(*) as QTY from SPACEXTABLE where date between
      '2010-06-04' and '2017-03-20' group by Landing_Outcome order by QTY DESC
 * sqlite:///my_data1.db
Done.
```

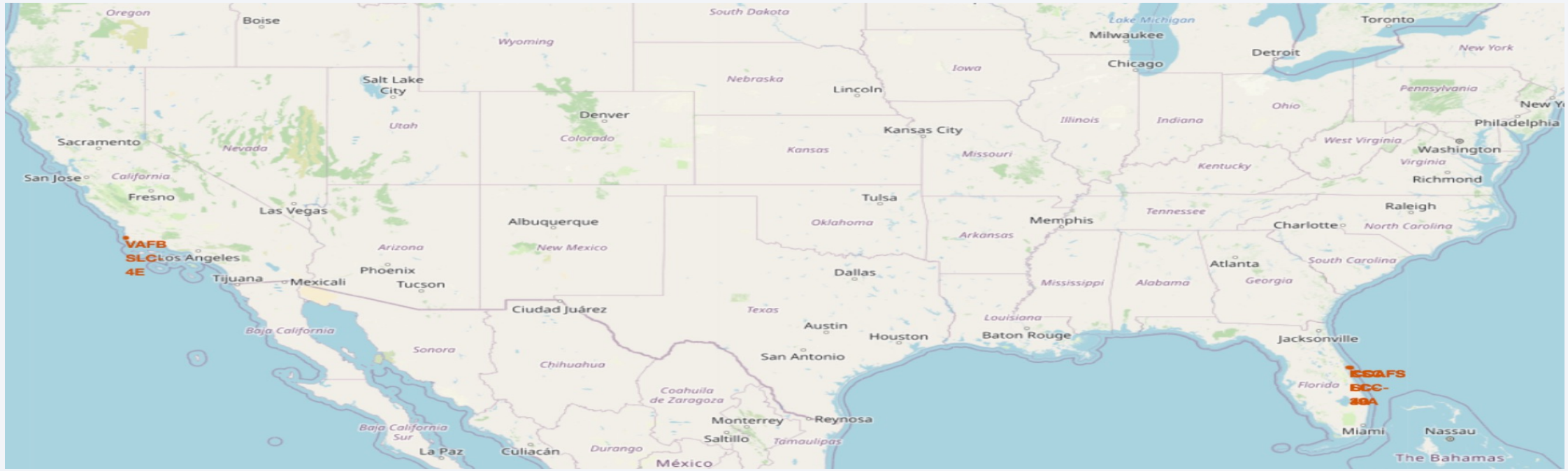
[21]:

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

- Here the landing records are extracted between 2010 and 2017, the data is then grouped by Landing Outcomes in Descending order.

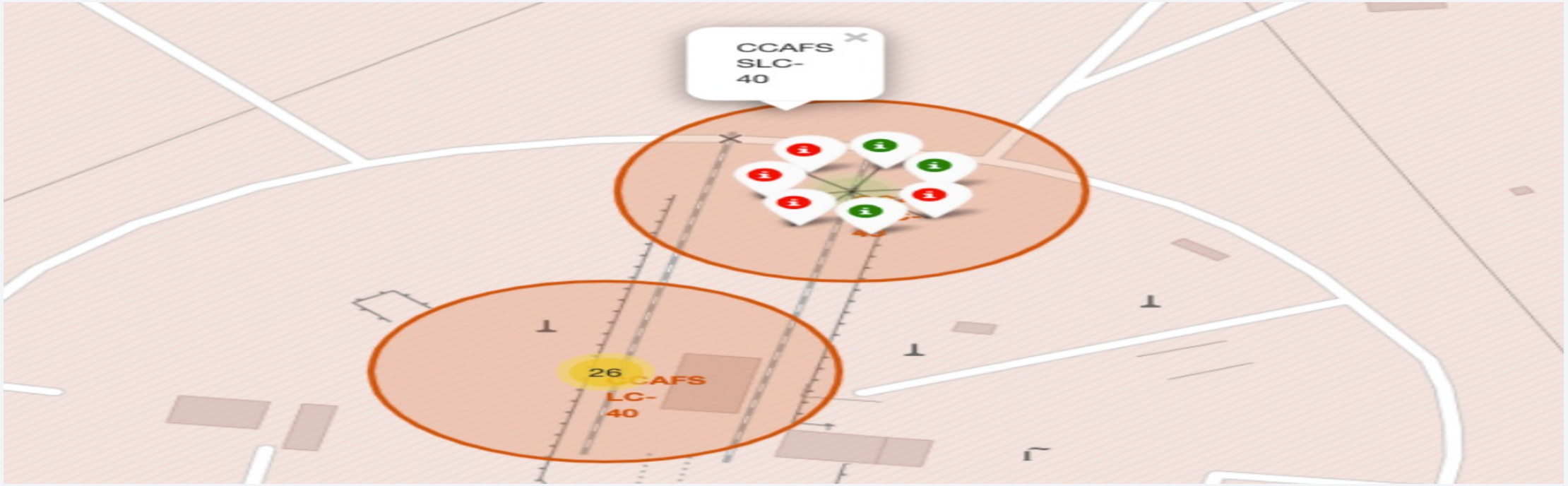

Section 3
Launch Sites
Proximities Analysis

All Launch Sites



- Displayed here is a map with all the launch sites, you can see that all the launch sites are very very close to the coast line, they are also located close to the equator.

Landing Outcomes

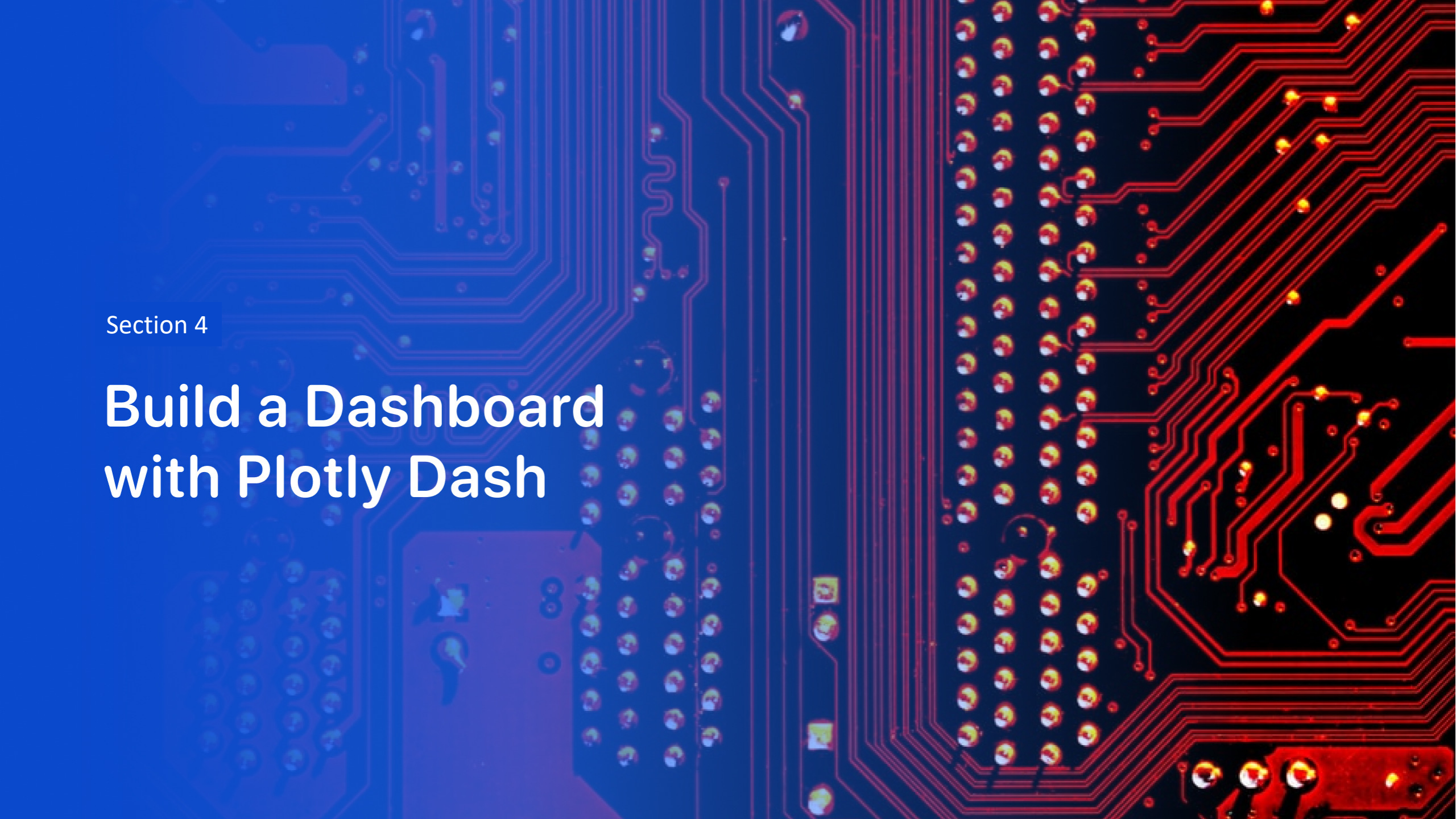


- Here we see where the outcomes are neck to neck on the CCAFS SLC-40 Launch site.

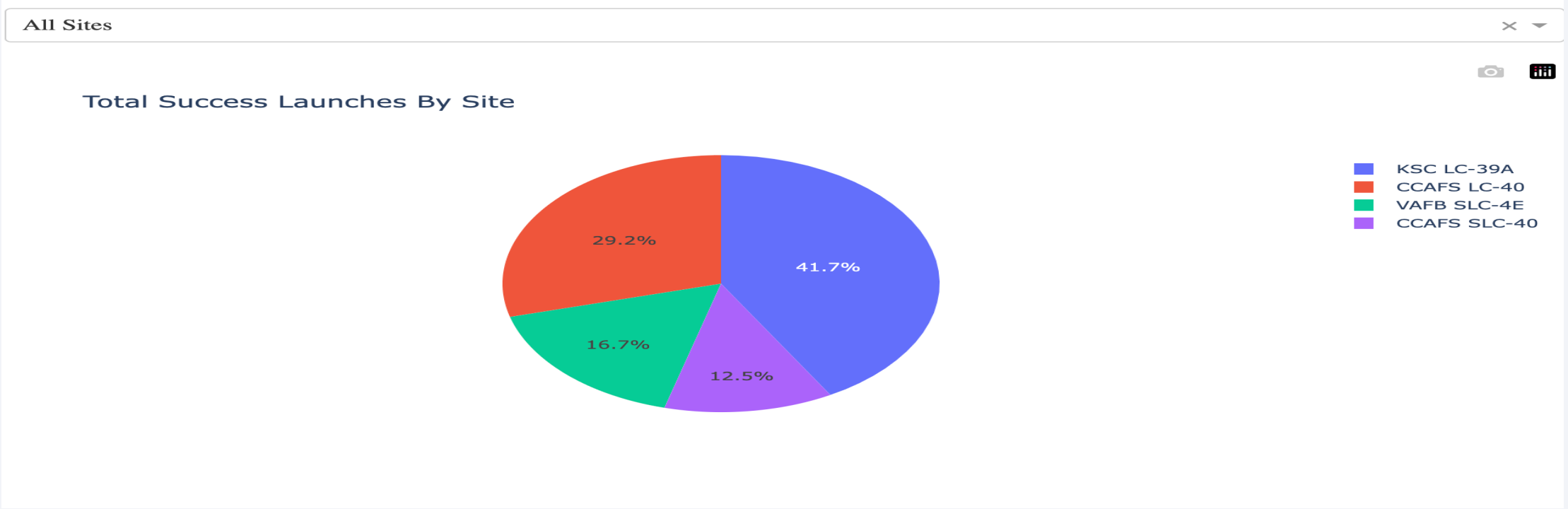
Landing Outcomes Florida

- 

le
- Landing Outcomes for florida launch site, it also shows how close the landing site is to the coast line.

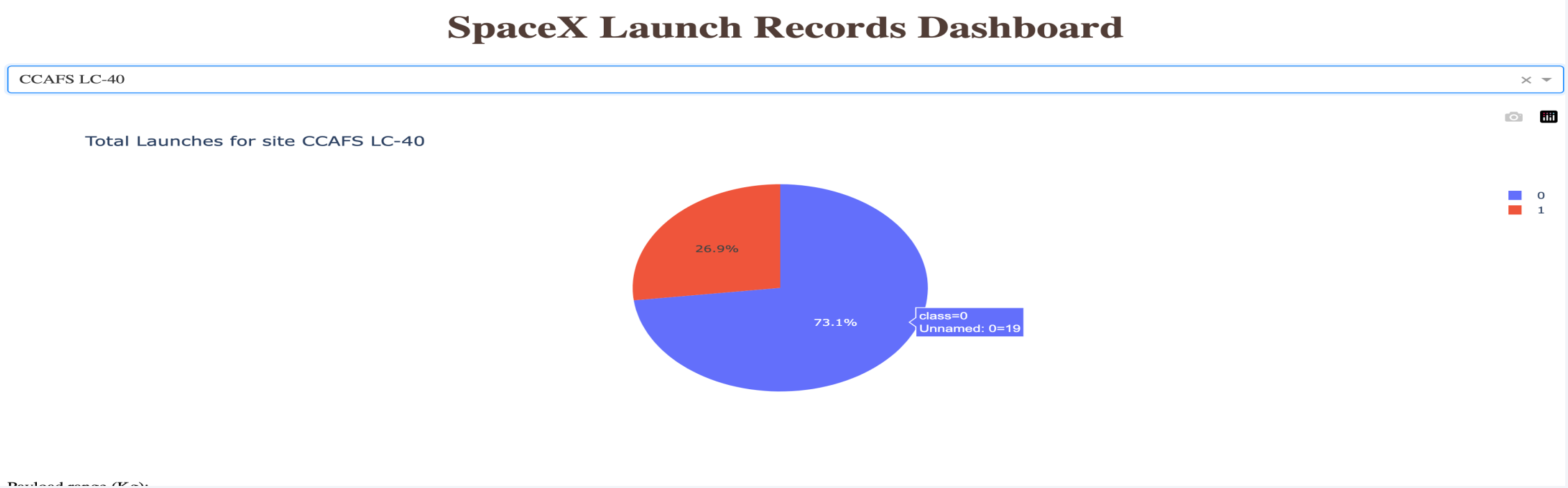

Section 4
Build a Dashboard with Plotly Dash

Pie Chart 1 – showing the landing sites and Launch Success Ratio



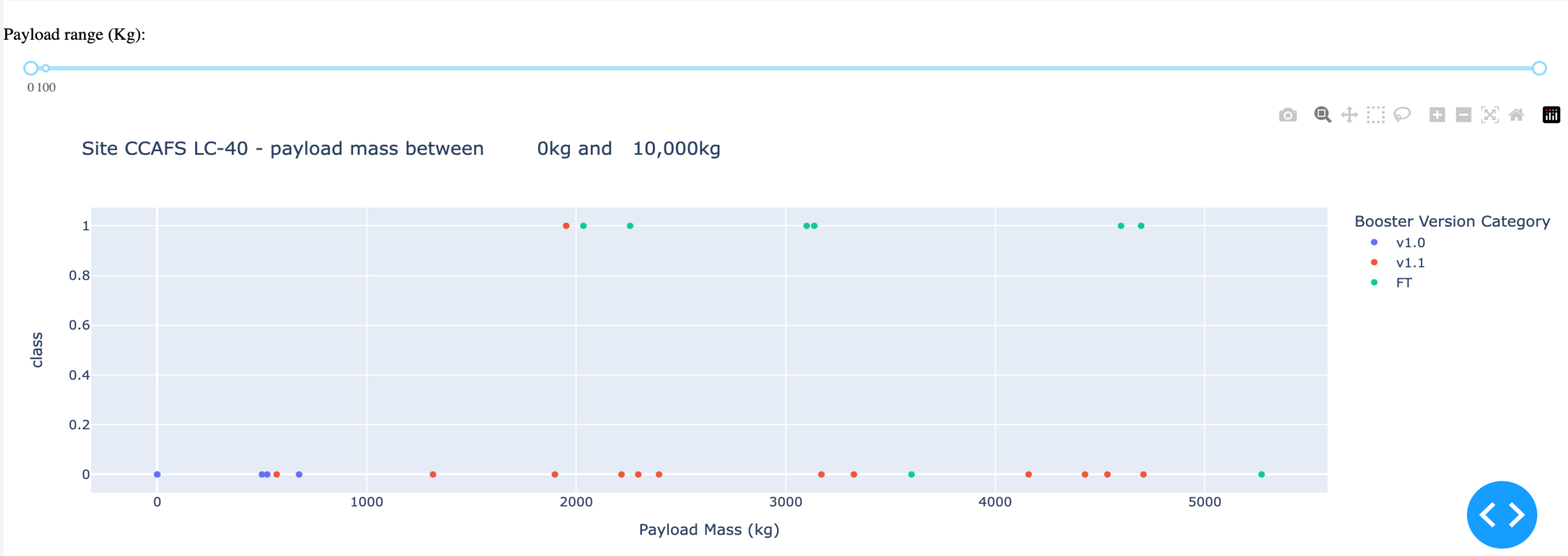
- Here we see where the dashboard displays the Total success launches by each launch site.
- Highest being KSC LC -39A with 41.7%, then CCAFS LC-40 with 29.2%, VAFB SLC-4E with 16.7% and CCAFS SLC-40 with 12.5%.

Pie Chart 2 – CCAFSLC-40 Launch Site



- Here we see Launch site CCAFS LC-40 with a success of 73.1% and a failure value of 26.9%.

Dashboard – Scatter Plot



- Here we see for the CCAFS launch site the booster version has the highest success rate for payload greater than 2000KG.

Section 5

Predictive Analysis (Classification)

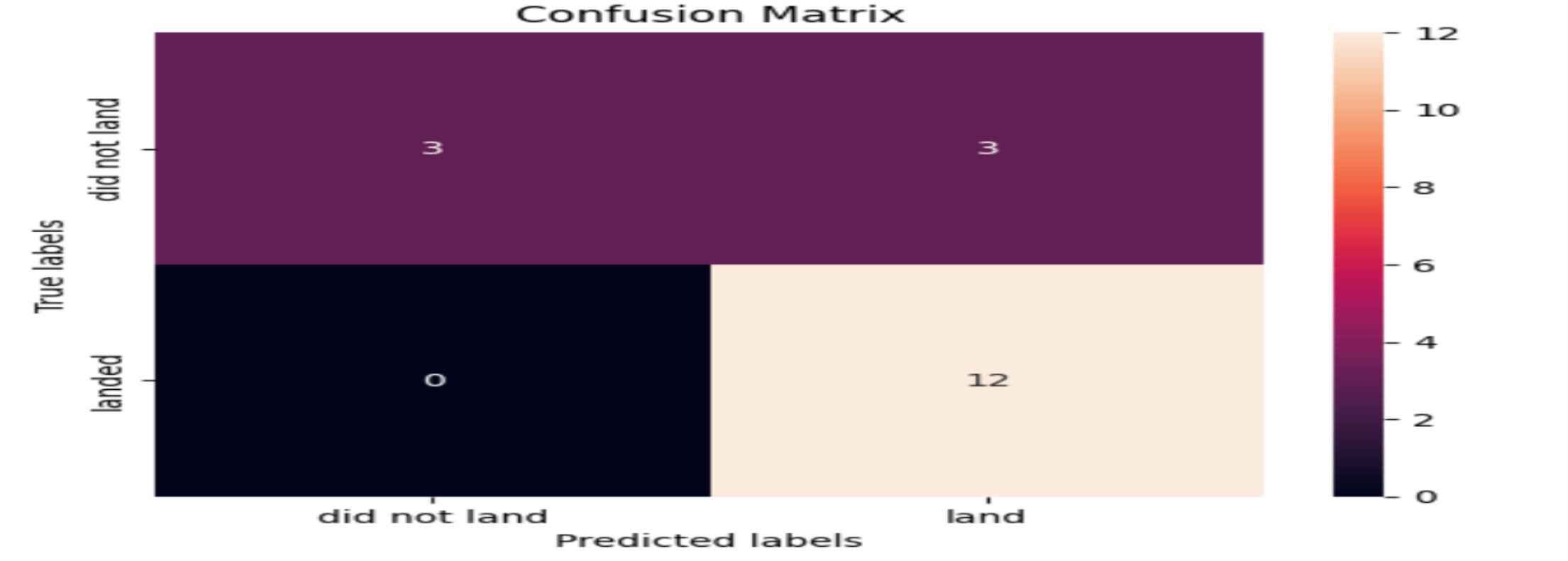
Classification Accuracy

```
[33]: m = {'LogisticRegression':logreg_cv.best_score_,'SupportVectorMachine': svm_cv.best_score_,'DecisionTree':tree_cv.
bpm = max(m, key=m.get)
print('The Method that Performs Best is the', bpm,' model with a score of', m[bpm])
if bpm == 'LogisticRegression':
    print('With Parameters : ', logreg_cv.best_params_)
if bpm == 'SupportVectorMachine':
    print('With Parameters : ', svm_cv.best_params_)
if bpm == 'DecisionTree':
    print('With Parameters : ', tree_cv.best_params_)
if bpm == 'KNeighbours':
    print('With Parameters : ', knn_cv.best_params_)
```

```
The Method that Performs Best is the DecisionTree  model with a score of 0.8875
With Parameters :  {'criterion': 'entropy', 'max_depth': 4, 'max_features': 'sqrt', 'min_samples_leaf': 2, 'min_s
amples_split': 2, 'splitter': 'best'}
```

- The DecisionTree classifier is the best model, it has the highest classification accuracy.

Confusion Matrix



- Here we can see the confusion matrix, the confusion matrix for each model is the same, the major issues see to to be the occurrence of false positive.

Conclusions

- In Conclusion we can can say that , there is an increase in success rate for an increase in flight rate.
- The KSC-LC-39A Launch site has the most successful launches.
- The ES-L1, GEO,SSO, HEO and LEO were the orbits with the most successful rate.
- There is a large success rate increase in 2013 to 2020.
- The Decision tree model is the best machine learning algo in this project.

Appendix

- Please Click here to be taking to the repository.
- Here

Thank you!