IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Leandro José dos Santos
January 4, 2025

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection through API
  - Data Collection with Web Scraping
  - Data Wrangling
  - Exploratory Data Analysis with SQL
  - Exploratory Data Analysis with Data Visualization
  - Interactive Visual Analytics with Folium
  - Machine Learning Prediction
- Summary of all results
  - Exploratory Data Analysis result
  - Interacive analytics ins screnshots
  - Predictive Analytics result

Introduction

- Project background and context

  We will predict if the Falcon 9 first stage will land successfully. 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. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch.

- Problems you want to find answers

  During the investigation, several obstacles were overcome, the main one being the complex list of variables that make up the situation.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - The data was collected from the SpaceX website with SpaceXAPI

    https://api.spacexdata.com/v4/rockets/
  - WebScraping

    (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)
  - Perform data wrangling
  - The data has been processed, cleaned and normalized.

Methodology

Executive Summary

- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Using Python's seaborn library
  - Data was normalized, divided in training and test data sets and evaluated by four different classification models.

Data Collection

- Describe how data sets were collected.

  Datasets were collected from SpaceX API (https://api.spacexdata.com/v4/rockets/) and from Wikipedia

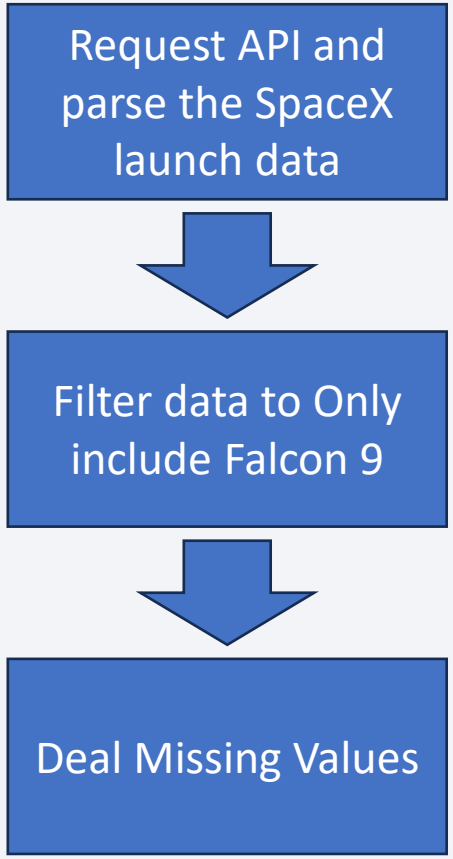
  (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches), using web scraping

  technics

Data Collection – SpaceX API

- Request to the SpaceX API
- Clean the requested data
- Source code:

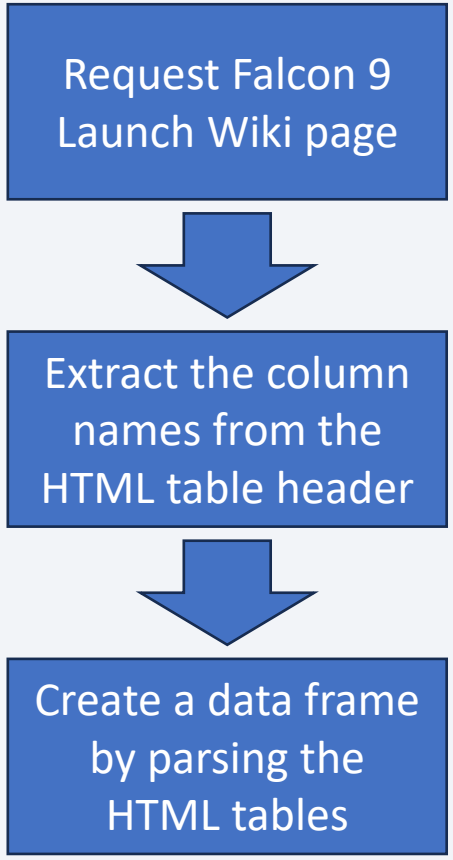
https://github.com/leandrojdsantos/Applied-Data-Science-Capstone/blob/fb1dd9bd98bc2698a539c68cb8239fbbed111f74/jupyter-labs-webscraping.ipynb



Data Collection - Scraping

- Data was obtained from Wikipedia;
- Source code:

https://github.com/brt-h/Applied-Data-Science-Capstone/blob/9751b6a3c1cf4144a8ed9ac884b4281f194bc52a/Hands-on%20Lab_%20Data%20Collection%20with%20Web%20Scraping.ipynb



Data Wrangling

- Some Exploratory Data Analysis (EDA) was performed on the dataset.
- Then the summary launches per site, occurrences of each orbit and occurrences of mission outcome per orbit type were calculated.
- Finally, the landing outcome label was created from Outcome column.
- Source code:

  https://github.com/leandrojdsantos/Applied-Data-Science-Capstone/blob/fb1dd9bd98bc2698a539c68cb8239fbbed111f74/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

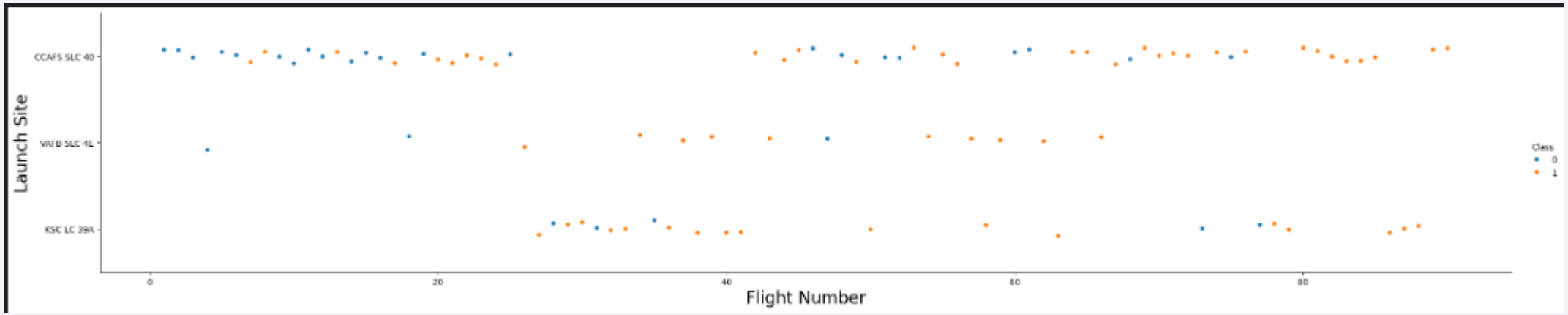
- Summarize what charts were plotted and why you used those charts
- Names of the unique launch sites in the space mission;
- Top 5 launch sites whose name begins with the string 'CCA';
- Total pay load mass carried by boosters launched by NASA (CRS);
- Average payload mass carried by booster version F9 v1.1;
- Date when the first successful landing outcome in ground pad was achieved;
- Names of the boosters which have success in drone ship and have payload mass between 4000 and 6000
- kg;
- Total number of successful and failure mission outcomes;
- Names of the booster versions which have carried the maximum payload mass;
- Failed landing out comes in droneship, their booster versions, and launch site names for in year 2015; and
- Rank of the count of landing outcomes (such as Failure (droneship) or Success (ground pad)) between the
- date 2010-06-04 and 2017-03-20

- Source code: https://github.com/leandrojdsantos/Applied-Data-Science-Capstone/jupyter-labs-eda-sql-coursera_sqllite.ipynb

EDA with SQL

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

  Payload Mass X Flight Number, Launch Site X Flight Number, Launch Site X Payload Mass, Orbit and Flight Number, Payload and Orbit



- Code: https://github.com/leandrojdsantos/edadataviz.ipynb

Build an Interactive Map with Folium

- Markers, circles, lines and marker clusters were used with Folium Maps
- Marker clusters indicates groups of events in each coordinate, like launches in a launch site; and Markers indicate points like launch sites;
- Circles indicate highlighted areas around specific coordinates, like NASA Center;
- Lines are used to indicate distances between two coordinates.Explain why you added those objects
- Code:

https://github.com/leandrojdsantos/lab_jupyter_launch_site_location%20.ipynb

Build a Dashboard with Plotly Dash

- We built an interactive dashboard with Plotly dash
- We plotted pie charts showing the total launches by a certain sites
- We plotted scatter graph showing the relationship with Outcome and Payload
- Mass (Kg) for the different booster version.

- GitHub URL: https://github.com/leandrojdsantosspacex_dash_app.py

Predictive Analysis (Classification)

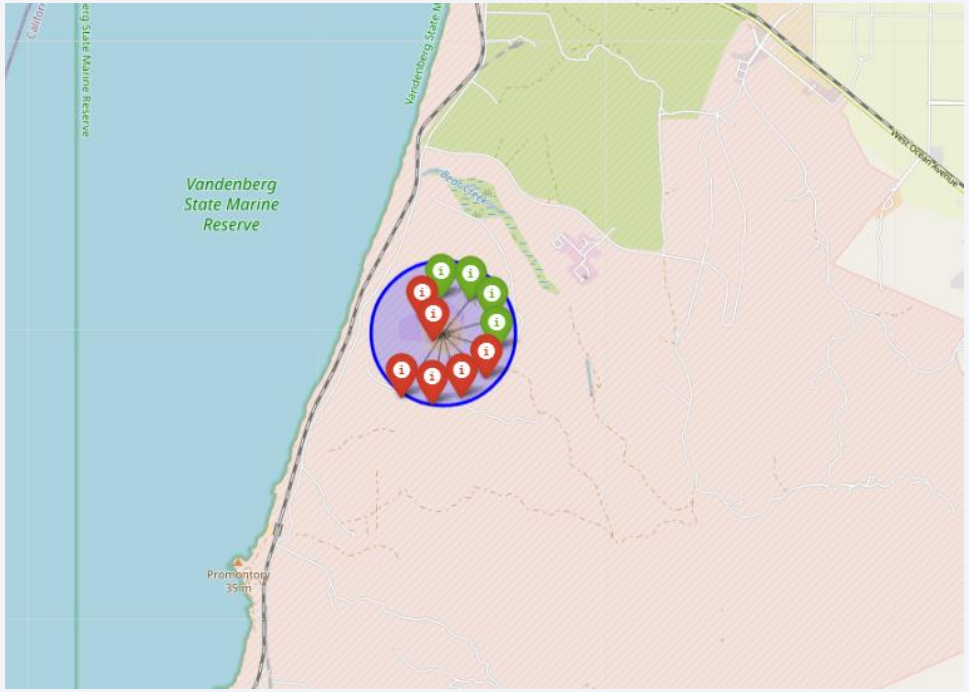
- We loaded the data using numpy and pandas, transformed the data, split our data into training and testing.
- We built different machine learning models and tune different hyperparameters using GridSearchCV.
- We used accuracy as the metric for our model, improved the model using feature engineering and algorithm tuning.
- We found the best performing classification model.
- GitHub: https://github.com/leandrojdsantos/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

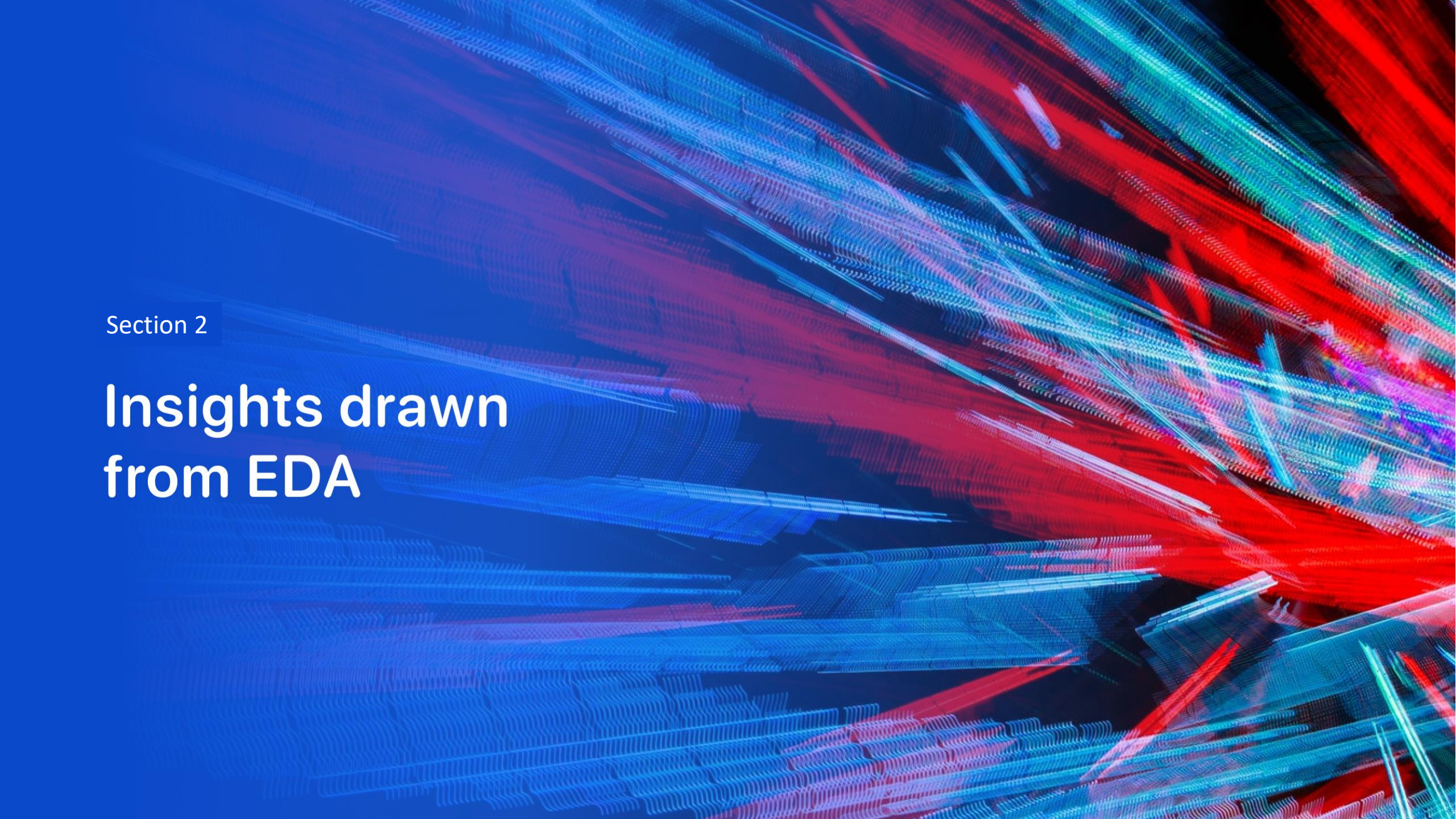
Results

- The first success landing outcome happened in 2015 fiver year after the first launch;
- Two booster versions failed at landing in drone ships in 2015: F9 v1.1 B1012 and F9 v1.1 B1015;
- Space X uses 4 different launch sites;
- The first launches were done to Space X itself and NASA;
- The number of landing outcomes became as better as years passed
- Many Falcon 9 booster versions were successful at landing in drone ships having payload above the average;
- Almost 100% of mission outcomes were successful;
- The average payload of F9 v1.1 booster is 2,928 kg;

Results

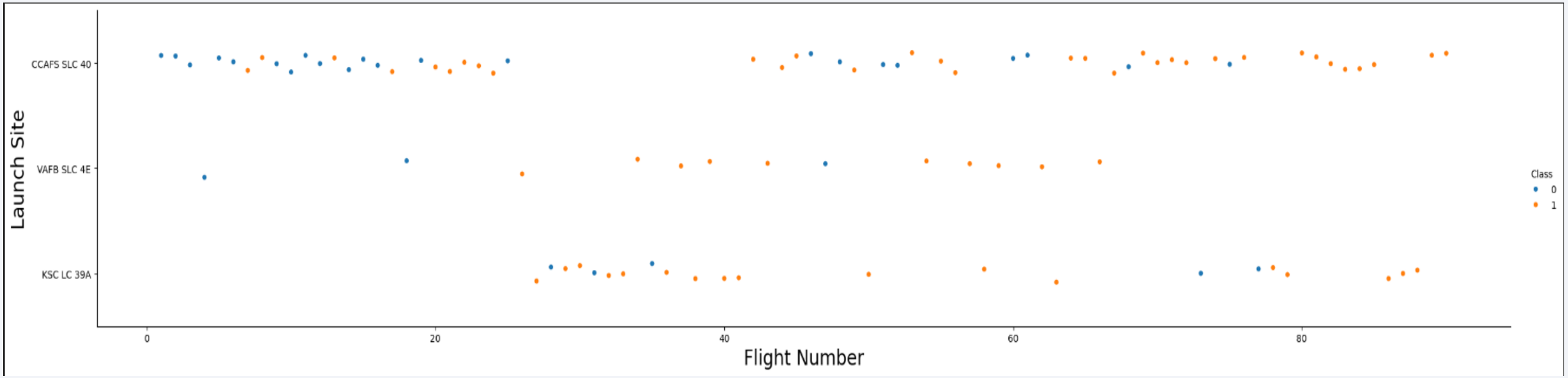
- Using interactive analytics was possible to identify that launch sites use to be in safety places, near sea, for example and have a good logistic infrastructure around.
- Most launches happens at east cost launch sites.




Section 2
Insights drawn from EDA

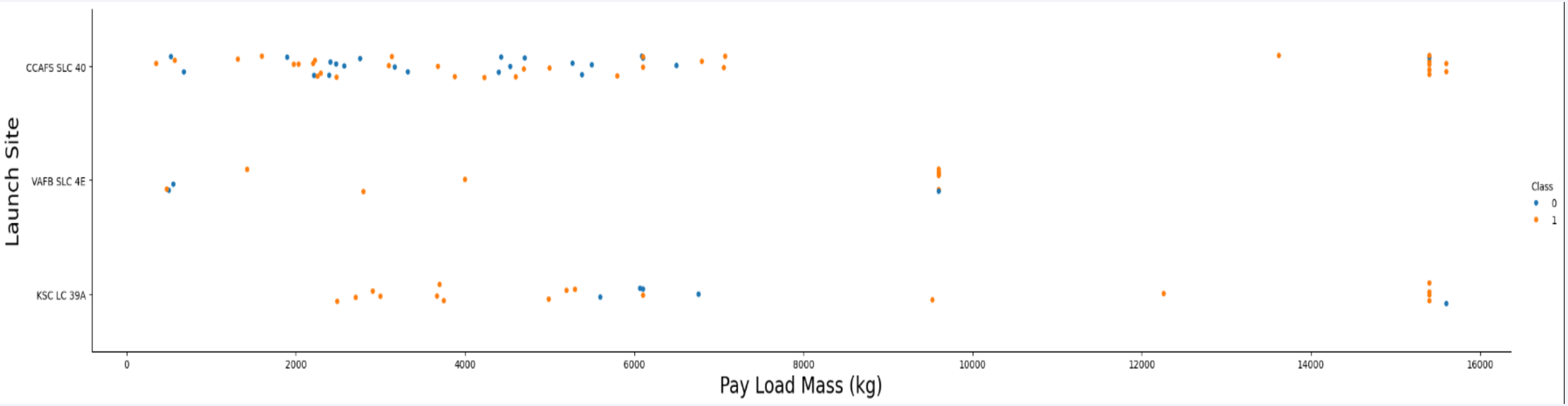
Flight Number vs. Launch Site

- From the plot, we found that the larger the flight amount at a lunch site, the greater the success rate at a launch site.



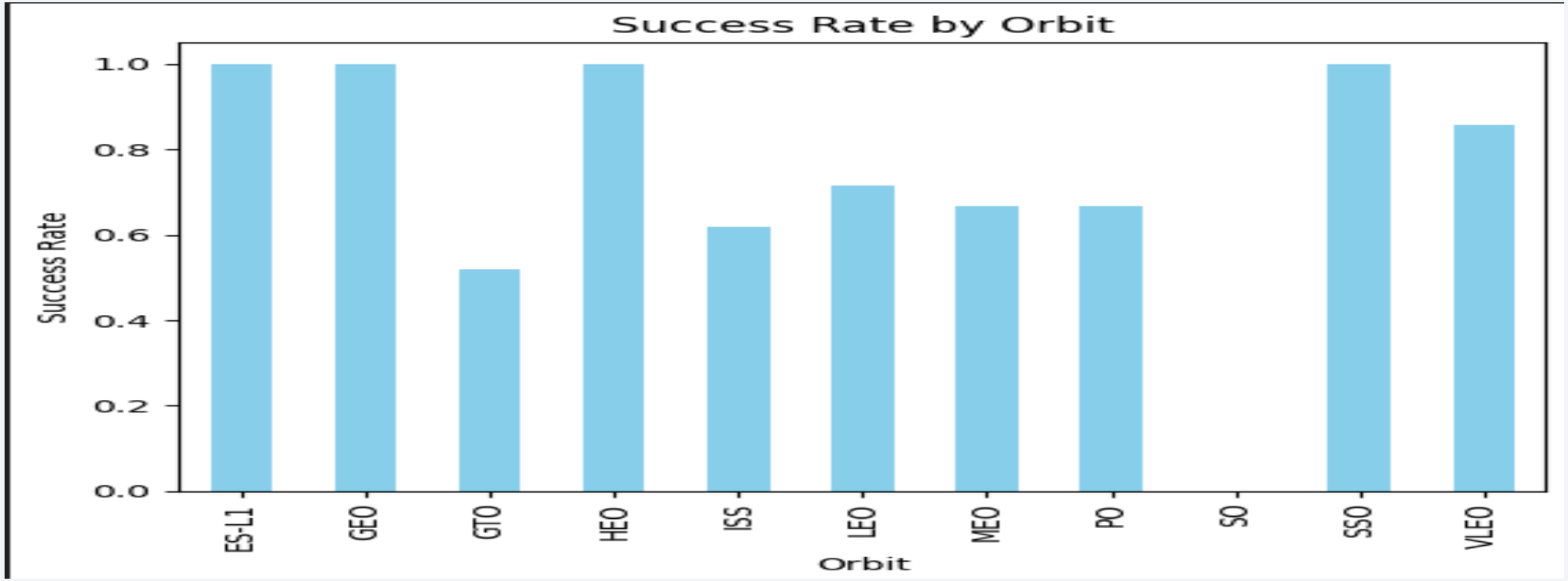
Payload vs. Launch Site

- The greater the payload mass for launch site CCAFS SLC 40 the higher the success rate for the rocket.



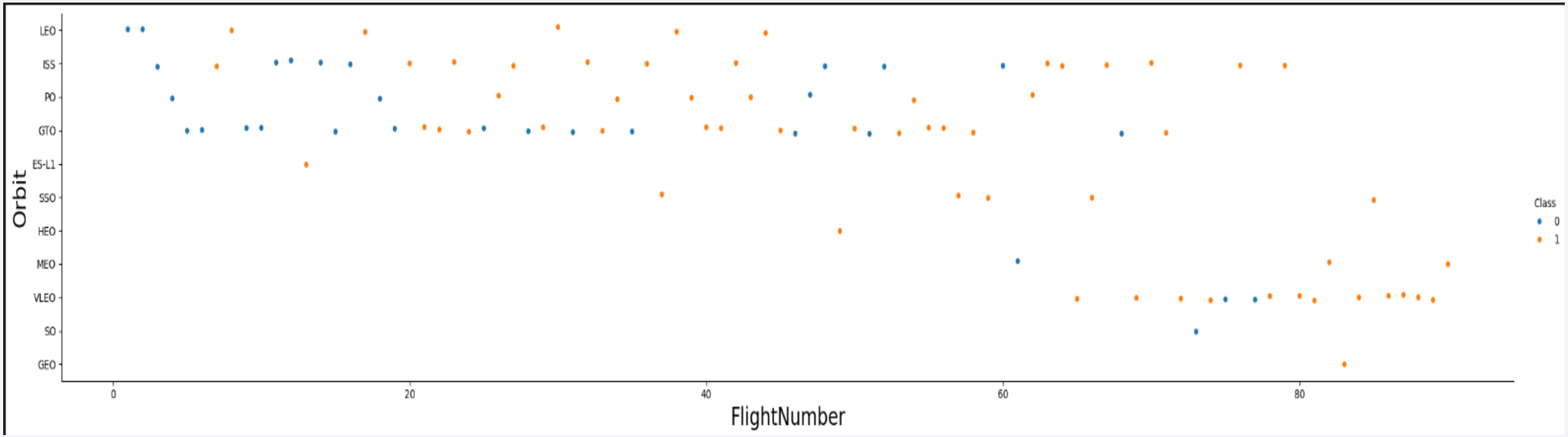
Success Rate vs. Orbit Type

- From the plot, we can see that ES-L1, GEO, HEO, SSO, VLEO had the most success rate.



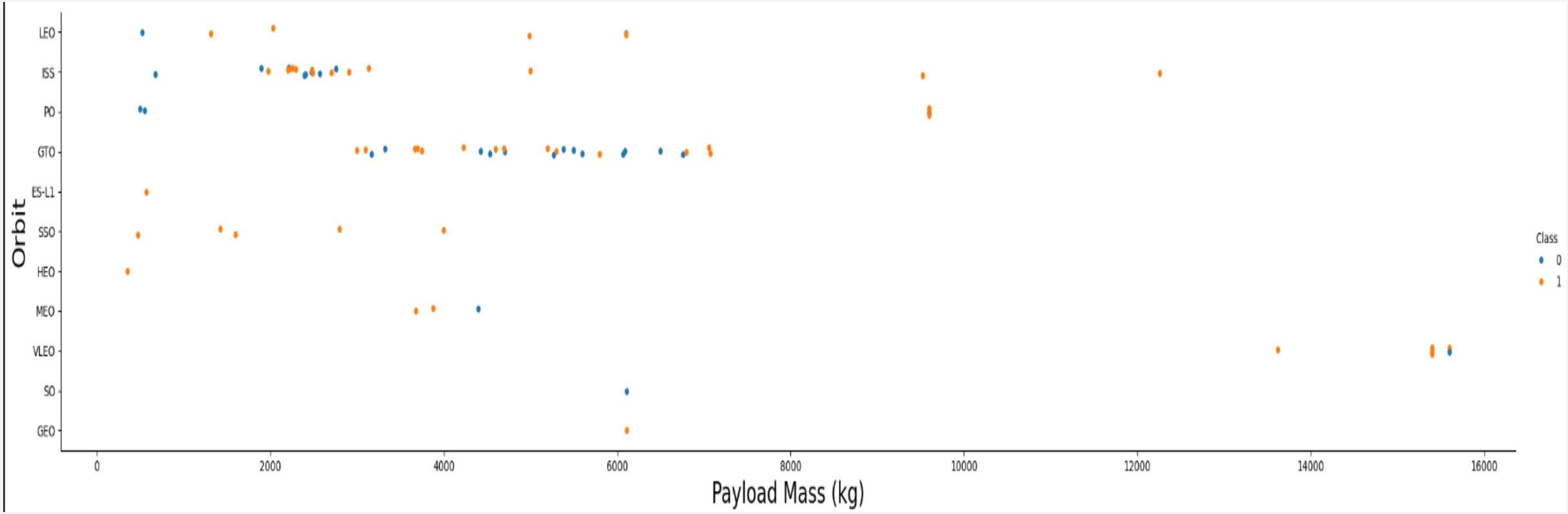
Flight Number vs. Orbit Type

- The plot below shows the Flight Number vs. Orbit type. We observe that in the LEO orbit, success is related to the number of flights whereas in the GTO orbit, there is no relationship between flight number and the orbit.



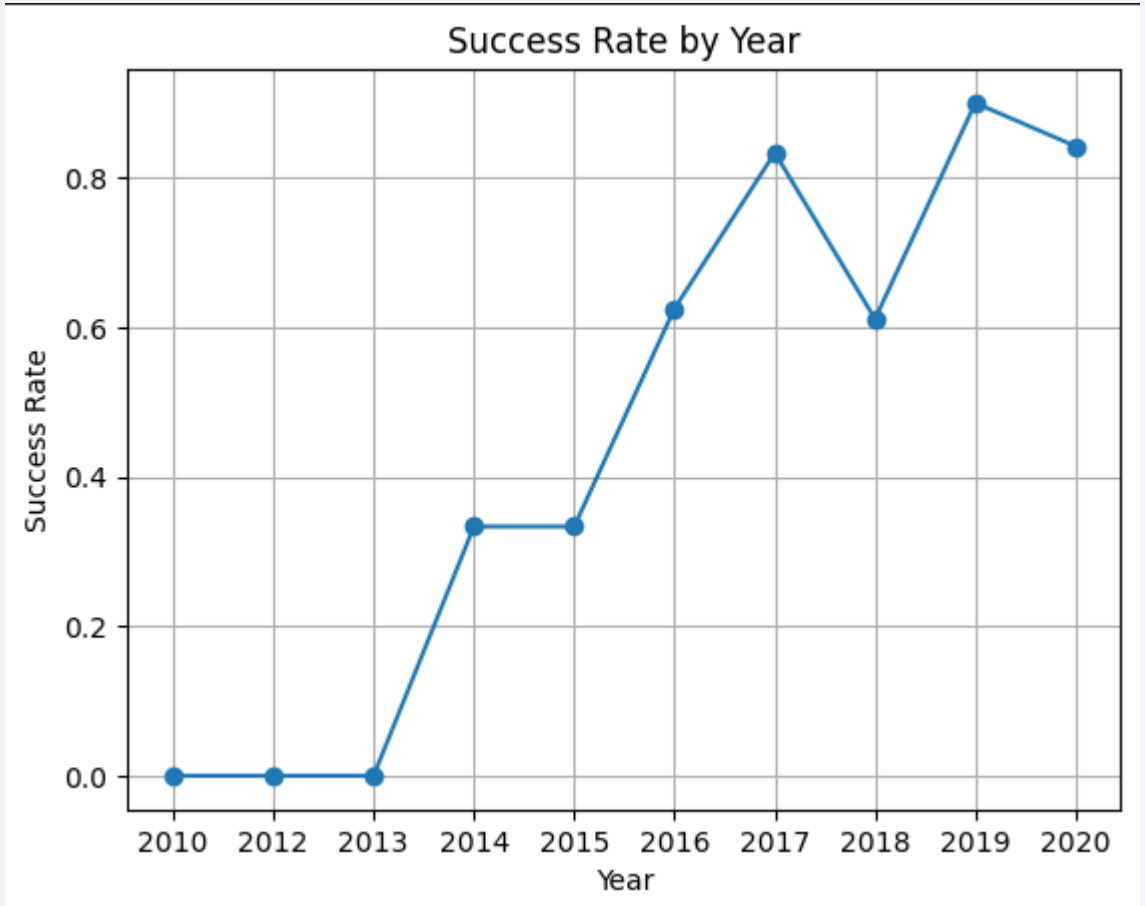
Payload vs. Orbit Type

- We can observe that with heavy payloads, the successful landing are more for PO, LEO and ISS orbits.



Launch Success Yearly Trend

- From the plot, we can observe that success rate since 2013 kept on increasing till 2020.



All Launch Site Names

- We used the key word **DISTINCT** to show only unique launch sites from the SpaceX data.

[16]:
```
# df.columns
# %sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE;
%sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE;
```

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

[16]:

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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`

[26]:
```
%sql SELECT * FROM SPACEXTABLE WHERE "Launch_Site" LIKE  "CCA%" LIMIT 5;
```

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

[26]:

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

Total Payload Mass

- We calculated the total payload carried by boosters from NASA as 45596 using the query below

```
[21]: %sql select sum("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "Customer" LIKE  "NASA (CRS)";
```

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

[21]:

sum("PAYLOAD_MASS__KG_")
45596

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- We calculated the average payload mass carried by booster

```
[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
```

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

  We observed that the dates of the first successful landing outcome on ground pad was 22nd December 2015

```
[15]: %sql select min("Date") FROM SPACEXTABLE WHERE "Landing_Outcome" LIKE  "Success (ground pad)";
 * sqlite:///my_data1.db
Done.
[15]: min("Date")
      2015-12-22
```

Successful Drone Ship Landing with Payload between 4000 and 6000

- We used the WHERE clause to filter for boosters which have successfully landed on drone ship and applied the AND condition to determine successful landing with payload mass greater than 4000 but less than 6000

[16]:
```
%sql select Booster_Version FROM SPACEXTABLE \
WHERE "Landing_Outcome" LIKE  "Success (drone ship)" \
AND "PAYLOAD_MASS__KG_" > 4000 \
AND "PAYLOAD_MASS__KG_" < 6000;
```

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

[16]:

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

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
- We used wildcard like '%' to filter for WHERE MissionOutcome was a success or a failure.

```
[22]: %sql select COUNT(*) FROM SPACEXTABLE WHERE "Mission_Outcome" LIKE  "%Success%";
      #%sql select COUNT(*) FROM SPACEXTABLE WHERE "Mission_Outcome" LIKE  "%Failure%";
       * sqlite:///my_data1.db
      Done.
[22]: COUNT(*)
      ---------
            100
```

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- Whe determined the bosster that carried the maximum with this query

```
[18]: %sql select Booster_Version, PAYLOAD_MASS__KG_ FROM SPACEXTABLE \
          ORDER BY PAYLOAD_MASS__KG_ DESC limit 10;
```

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

[18]:

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600

2015 Launch Records

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

  We used a combinations of the WHERE clause, LIKE, and AND conditions to filter for failed landing outcomes in drone ship, their booster versions, and launch site names for year 2015

[19]:
```
%sql SELECT substr("Date", 6, 2) as month, "Landing_Outcome", "Booster_Version", "Launch_Site" \
FROM SPACEXTABLE \
WHERE substr("Date", 0, 5) = '2015' AND "Landing_Outcome" LIKE '%failure%drone ship%';
```

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

[19]:

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

We selected Landing outcomes and the COUNT of landing outcomes from the data and used the WHERE clause to filter for landing outcomes BETWEEN 2010-06-04 to 2010-03-20. We applied the GROUP BY clause to group the landing outcomes and the ORDER BY clause to order the grouped landing outcome in descending order

```
[20]: %sql SELECT "Landing_Outcome", COUNT(*) as outcome_count \
      FROM SPACEXTABLE \
      WHERE "Date" BETWEEN '2010-06-04' AND '2017-03-20' \
      GROUP BY "Landing_Outcome" \
      ORDER BY outcome_count DESC;
```

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

[20]:

Landing_Outcome	outcome_count
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

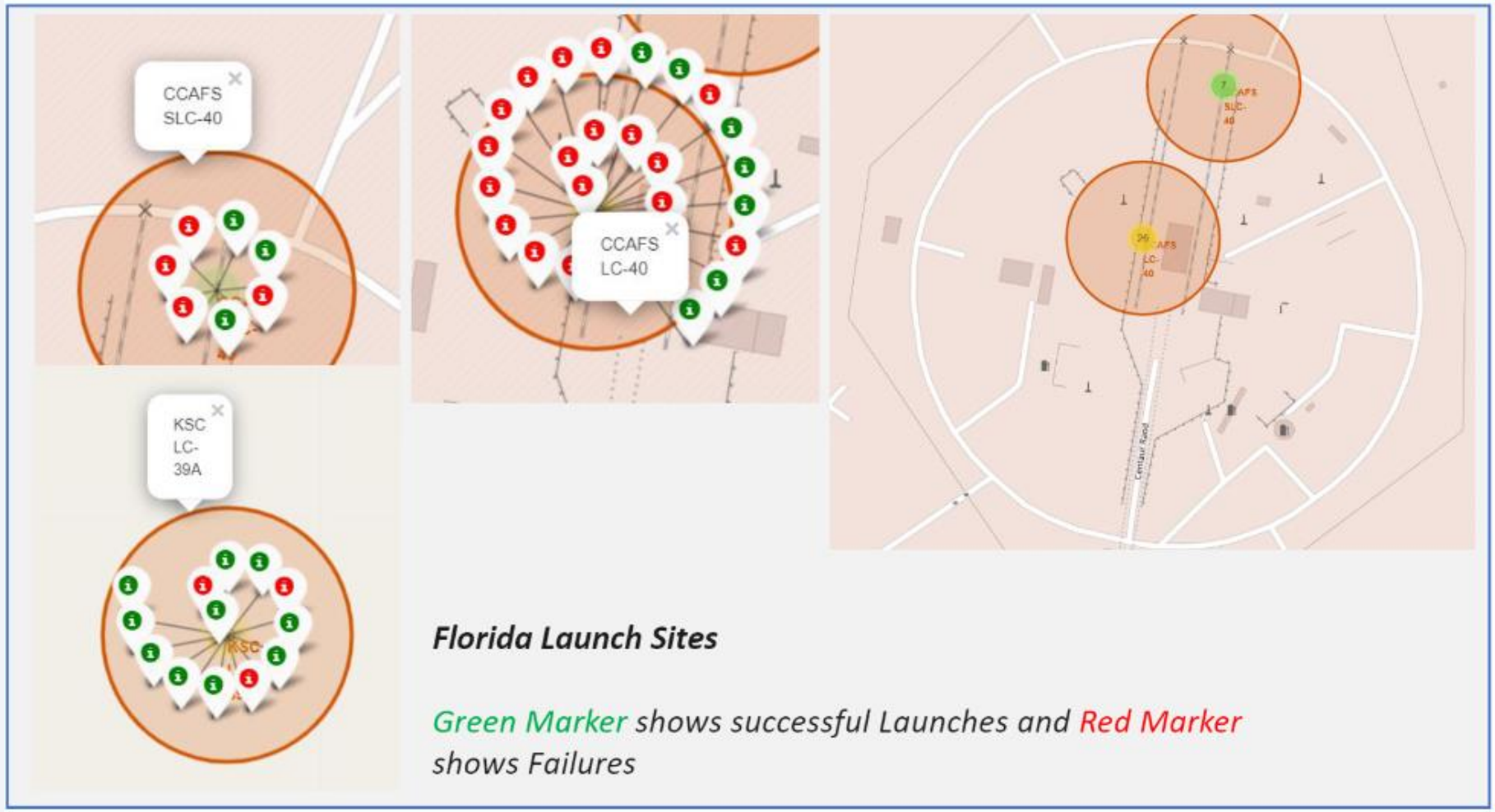
Section 3

Launch Sites Proximities Analysis

All launch sites global map markers

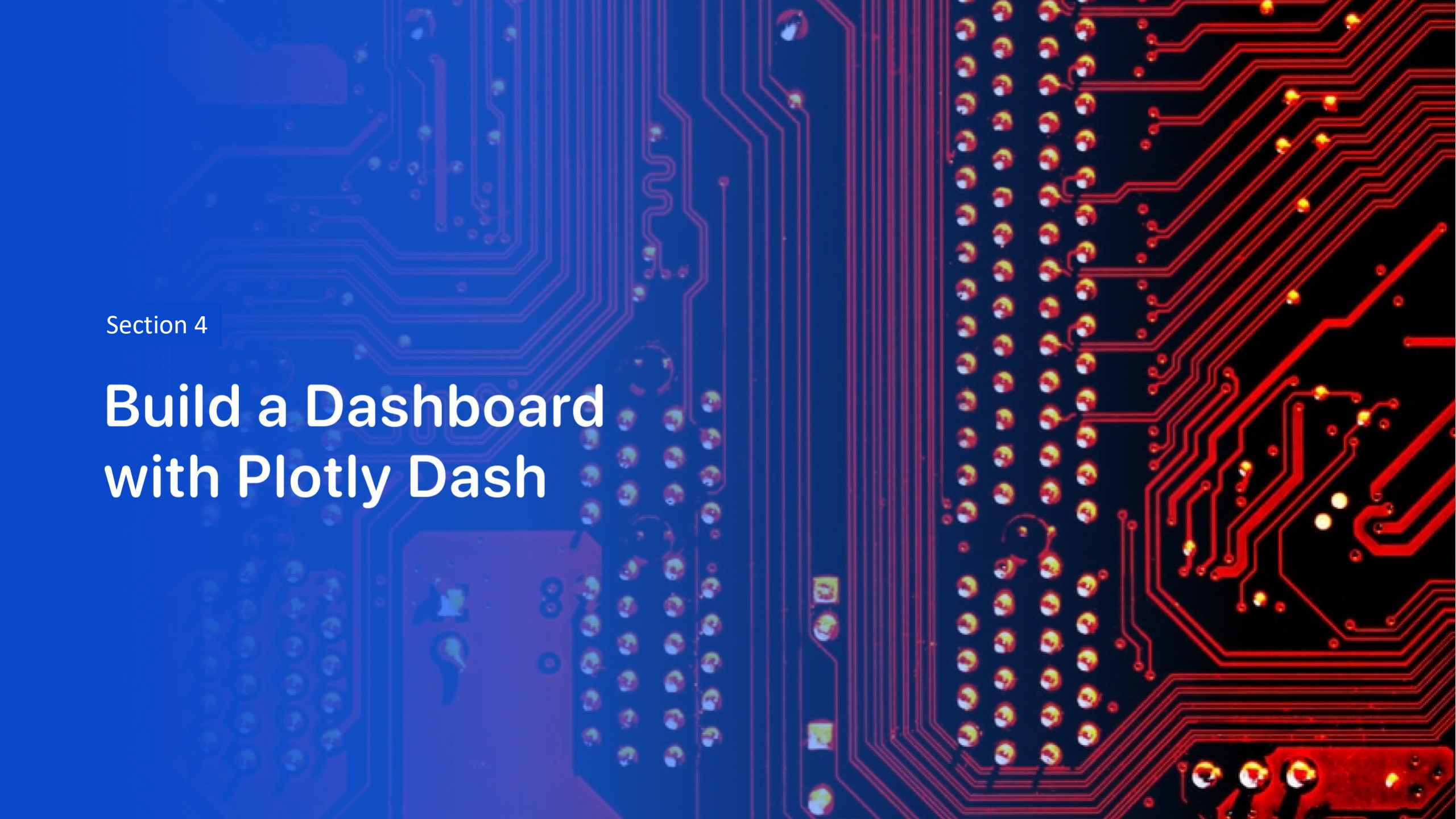
- We can see that the SpaceX launch siter are in the United States of America coasts. Florida and California.

Markers showing launch sites with color labels

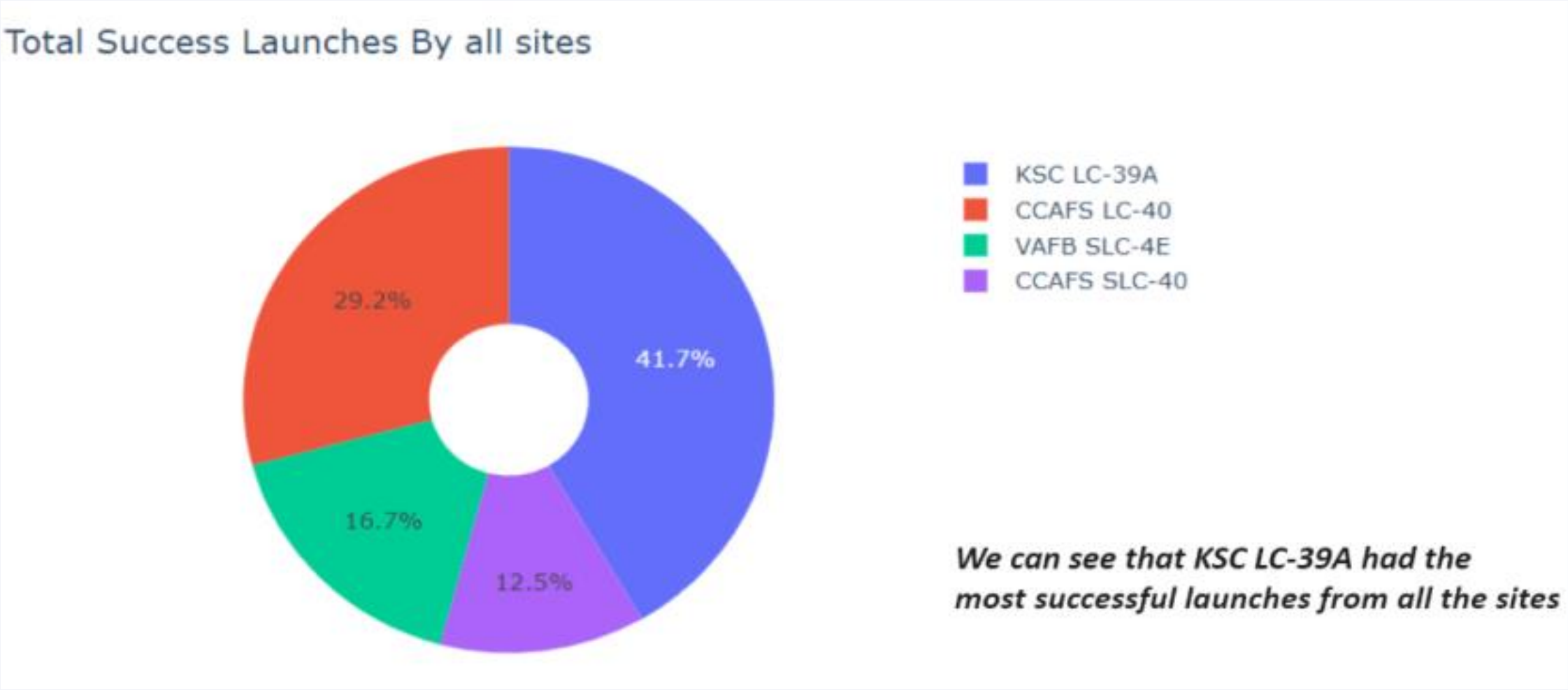


Launch Site distance to landmarks

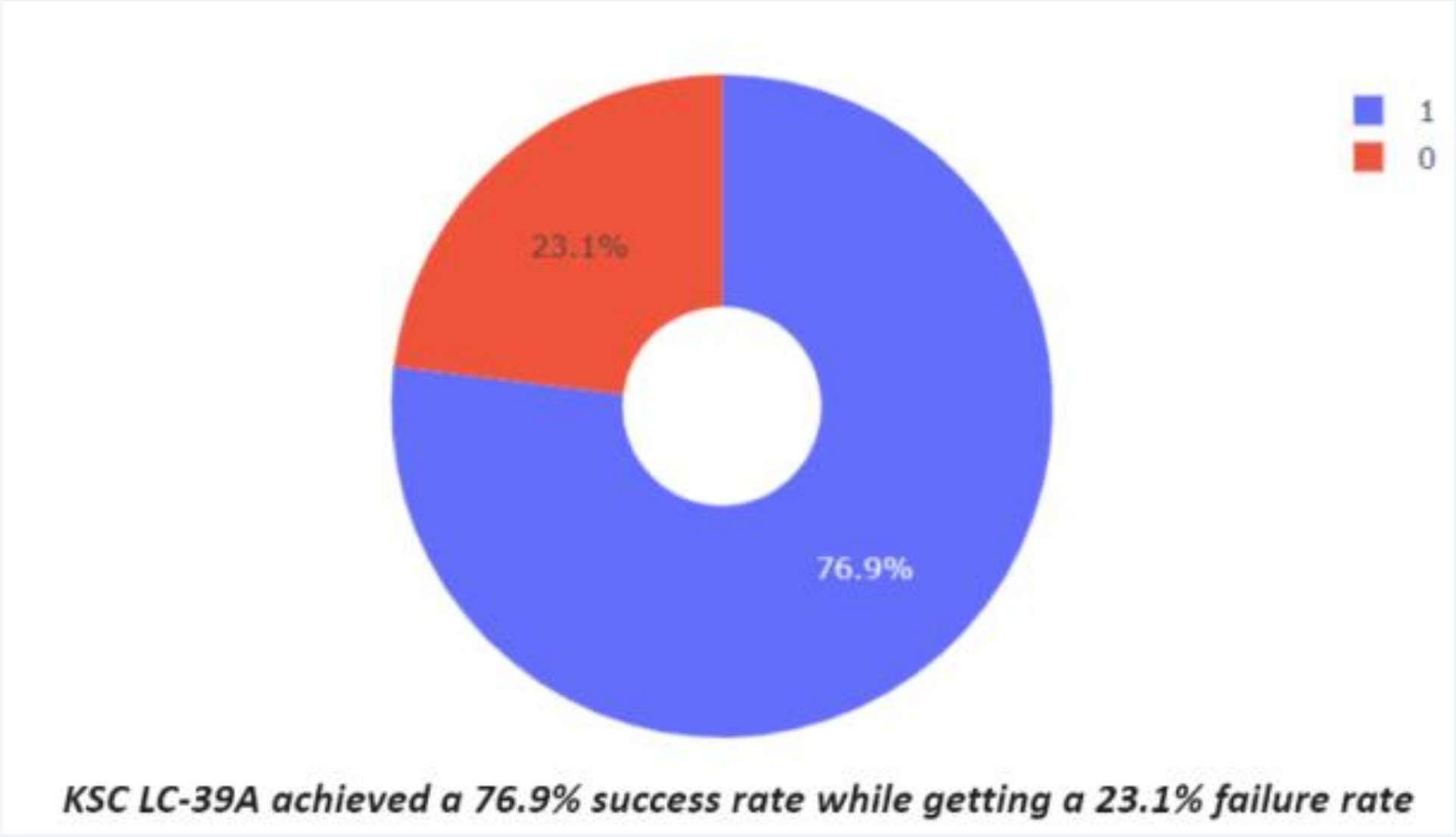
- Are launch sites in close proximity to railways? No
- Are launch sites in close proximity to highways? No
- Are launch sites in close proximity to coastline? Yes
- Do launch sites keep certain distance away from cities? Yes


Section 4
Build a Dashboard with Plotly Dash

Pier chart showing the success percentage achieved by each launch site

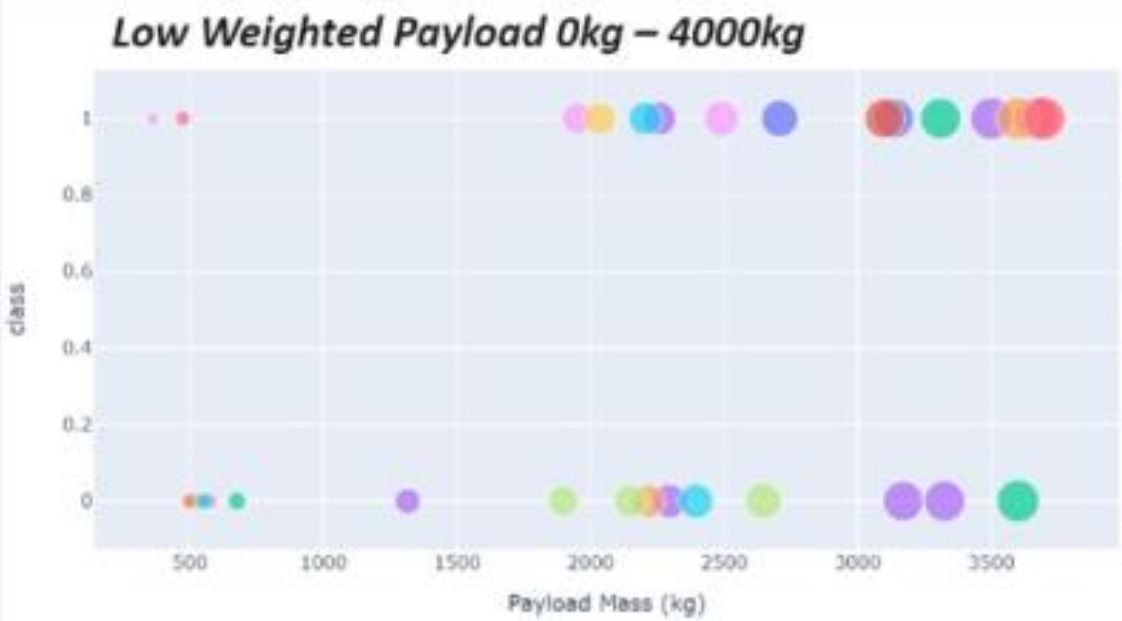


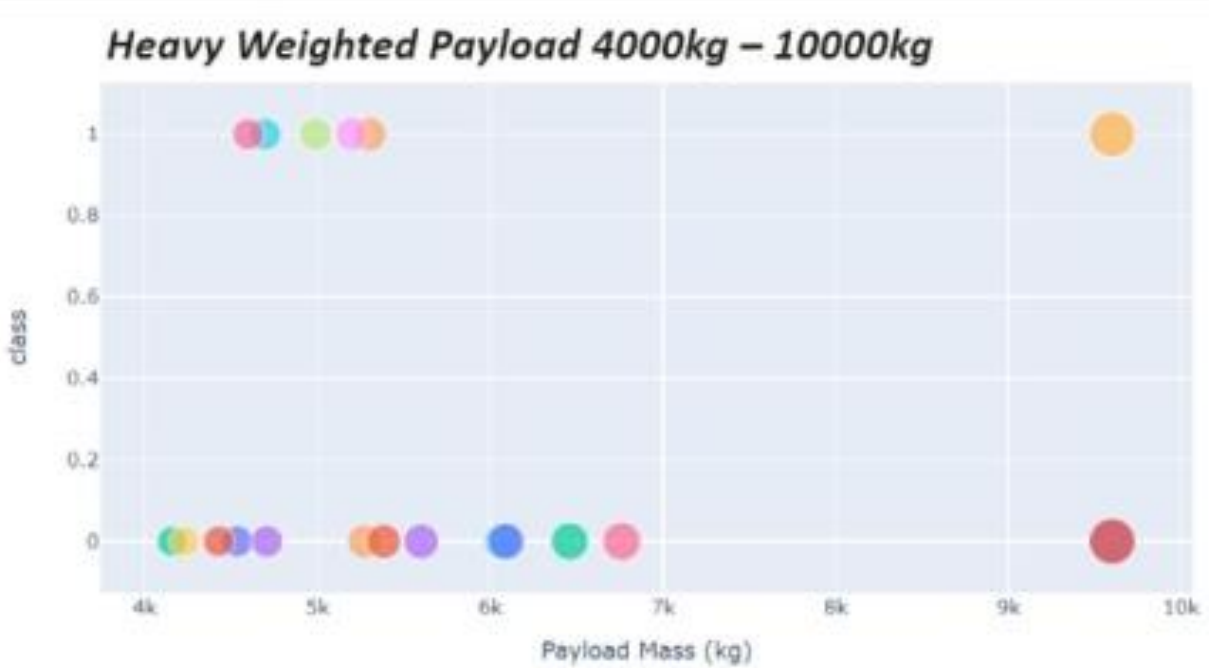
Pie chart showing the Launch site with the highest launch success ratio



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

Scatter plot of Payload vs Launch Outcome for all sites, with different payload selected in the range slider





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

Section 5

Predictive Analysis (Classification)

Classification Accuracy

- The decision tree classifier is the model with the highest classification accuracy

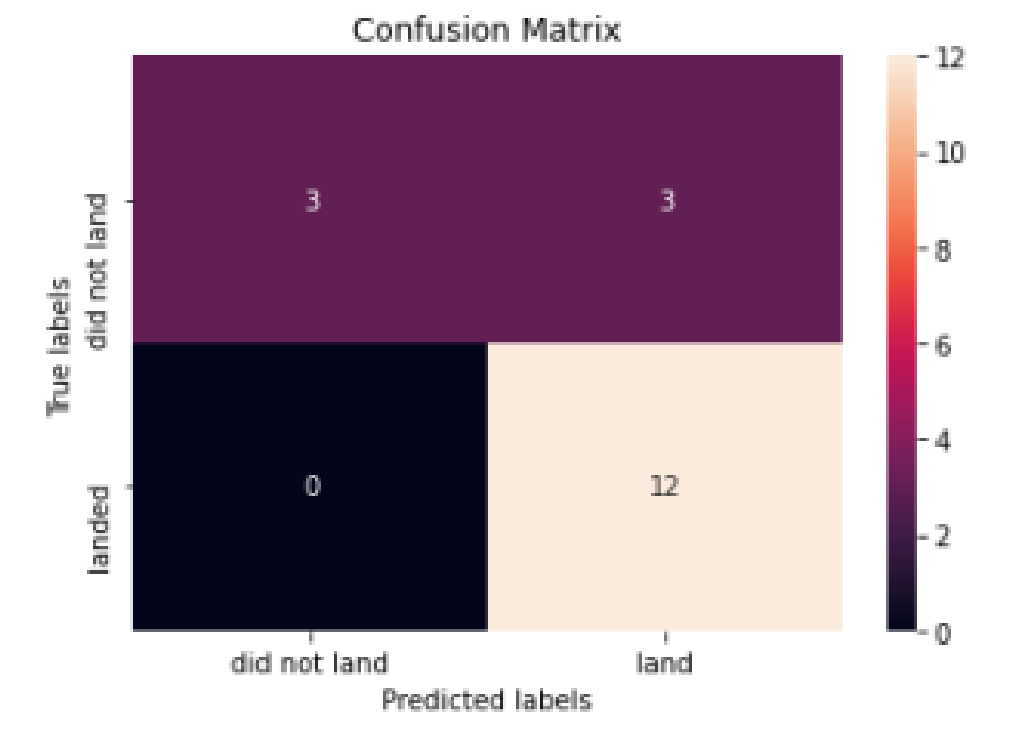
```
models = {'KNeighbors':knn_cv.best_score_,
              'DecisionTree':tree_cv.best_score_,
              'LogisticRegression':logreg_cv.best_score_,
              'SupportVector': svm_cv.best_score_}

bestalgorithm = max(models, key=models.get)
print('Best model is', bestalgorithm,'with a score of', models[bestalgorithm])
if bestalgorithm == 'DecisionTree':
    print('Best params is :', tree_cv.best_params_)
if bestalgorithm == 'KNeighbors':
    print('Best params is :', knn_cv.best_params_)
if bestalgorithm == 'LogisticRegression':
    print('Best params is :', logreg_cv.best_params_)
if bestalgorithm == 'SupportVector':
    print('Best params is :', svm_cv.best_params_)
```

```
Best model is DecisionTree with a score of 0.8732142857142856
Best params is : {'criterion': 'gini', 'max_depth': 6, 'max_features': 'auto', 'min_samples_leaf': 2, 'min_samples_split': 5, 'splitter': 'random'}
```

Confusion Matrix

- The confusion matrix for the decision tree classifier shows that the classifier can distinguish between the different classes. The major problem is the false positives .i.e., unsuccessful landing marked as successful landing by the classifier.



Conclusions

We can conclude that:

- Orbits ES-L1, GEO, HEO, SSO, VLEO had the most success rate.
- The larger the flight amount at a launch site, the greater the success rate at a launch site.
- The Decision tree classifier is the best machine learning algorithm for this task.
- Launch success rate started to increase in 2013 till 2020.
- KSC LC-39A had the most successful launches of any sites.

Thank you!