Winning Space Race with Data Science

Justice Nxumalo
19/01/2024

Outline

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

Executive Summary

- **Summary of methodologies**

The primary objective of this research is to identify the key factors contributing to the successful landing of the first stage in rocket launches, with the ultimate goal of reducing the overall cost of rocket launches at SpaceX. The successful recovery and reuse of the first stage have significant implications for cost efficiency and sustainability in space exploration. The methodologies used are as follows-:

- **Data requirements:** The foundation of the analysis is based on a dataset encompassing records of past rocket launches. Key variables in the dataset include payload mass, launch site, launch outcome, launch dates, and the orbit to which the rocket was launched. By leveraging this diverse set of data, the aim is to discern patterns and correlations that contribute to the successful landing of the first stage.
- **Data collection:** The SpaceX data crucial for the analysis was collected utilizing web scraping techniques on the Wikipedia SpaceX page, and leveraged the SpaceX API to enrich our dataset with real-time and detailed information, ensuring the dataset's relevance and accuracy. The collected data spans a range of parameters, providing a holistic view of SpaceX rocket launches and their outcomes.
- **Data understanding and data Intergration**: The dataset was loaded into corresponding tables within a Db2 database, facilitating efficient management and retrieval of information. SQL queries were executed to assess data quality, ensuring consistency and accuracy for subsequent analyses.

Executive Summary

- **Data Wrangling and Preparation**: A meticulous data wrangling process was implemented to ensure the dataset's cleanliness and usability. Cleaning procedures involved handling missing values, addressing inconsistencies, verifying data integrity, determining feature labels for training models.
- **Graphical Data Exploration and Feature Engineering:** Graphical visualization techniques were applied to explore relationships between features, such as payload vs flight number and launch site vs flight number vs landing outcome, and highlighting launching sites locations on the map. Feature engineering was performed to enhance the predictive power of our models, incorporating domain knowledge and insights gained from the exploratory analysis.
- **Data Standardization and Model Exploration:** To ensure consistency in model training, data was standardized before being split into training and testing sets. Various machine learning models, including logistic regression, decision tree, and k-nearest neighbors (kNN) and support vector machine (SVM), were explored to identify the best-performing model for predicting first-stage landing success. Model evaluation metrics and cross-validation techniques were employed to assess performance, providing a comprehensive understanding of each model's strengths and limitations.

Summary of all results

- **Exploratory Data Analysis:**
  - Among the 100 launches analyzed, CCAFS SLC 40 emerged as the launch site with the highest frequency.

Executive Summary

Summary of all results continuation..

- Over time, a positive trend in success rates was observed, indicating an improvement in the success of first-stage landings.
- VAFB SLC 4E demonstrated the highest landing success rate among the launch sites

- **Orbit vs Success Rate Analysis:**
  - Specific orbits, including ES-LI, GEO, HEO, and SSO, exhibited a 100% success rate, highlighting their reliability.
  - Conversely, the SO orbit showed a 0% success rate.
- **Visualization Analysis:**
  - Geographical visualization revealed that launch sites are strategically located near the equator, leveraging Earth's rotational speed for optimal launches.
  - Coastal launch sites were chosen, providing a safer and less inhabited surrounding in the event of unforeseen complications.
- **Data Model Outcomes:**
  - All models exhibited commendable performance in predicting first-stage landing success.
  - The decision tree model outperformed others

Introduction

Project background and context

Space X 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 Space X can reuse the first stage. Therefore if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against space X for a rocket launch.

Problems you want to find answers

- What types of data are essential for understanding the factors influencing first-stage landings?
- How do launch sites, orbits, payloads mass, flight number influence the success rates of first-stage landings?
- What insights can be gained from analyzing the relationship between different features i.e. orbits and landing outcomes or payload mass and landing outcome?
- What is the best predictive model for a successful landing

NB. All related documents, code, outputs, datasets and this presentation can be found on github-:

https://github.com/justicebhekani88/IBM_Data_Science_Capstone_Project

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Collect data from the SpaceX API and web scrapping the Wikipedia page
- Perform data wrangling
  - Segment dataset and retain only the data we need for analysis, cleaning the data, finding and filling or removing missing values, standardizing out data and performing one hot encoding to have it read for modeling.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Feature engineering, splitting the dataset into training and testing data, selecting the classification models. Training the models and evaluating the results to select the best

Data Collection

Data was collected from the SpaceX API as follows-:

1. Request data from the SpaceX API using the GET request and assign it to a variable response
2. Decode the response using json() converting it to a dataframe using the json_nomalize() function.
3. Perform some exploratory data analysis using head() to view a snapshot of the dataset, shape to view the number of rows and columns
4. Segment the dataset by extracting only the columns needed for data analysis.
5. Perform some data preparation by removing some rows, formatting some columns.
6. Create a dictionary
7. Create a dataframe using the dictionary
8. Filter the dataframe to only include Falcon 9 launches
9. Replace missing values from the payloadMass column using the calculated mean()
10. Export dataset to a .csv file

Data Collection – SpaceX API

```
static_json_url='https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_
response = requests.get(static_json_url)
data = pd.json_normalize(response.json())
print(data.head(5))
```

```
# Create a data from launch_dict and filter dataset to retain only Falcon 9 launches
data2 = pd.DataFrame.from_dict(launch_dict)
data_falcon9 = data2[data2['BoosterVersion'] == 'Falcon 9']
```

```
# Calculate the mean value of PayloadMass column
data_falcon9 = data2[data2['BoosterVersion'] == 'Falcon 9']
payloadMean = data_falcon9['PayloadMass'].mean()

# Replace the np.nan values with its mean value
data_falcon9['PayloadMass'].replace(np.nan, payloadMean)
data_falcon9['PayloadMass'].isnull().sum()
```

```
data_falcon9.to_csv('dataset_part_1.csv', index=False)
```

Data Collection

Data was scrapped from the Falcon 9 launches page on Wikipedia as follows-:

1. Request the html page using the HTTP Get request
2. Create a beautifulsSoup object
3. Extract a Falcon 9 launch records HTML table from Wikipedia
4. Extract columns names from the table header
5. Collect data by parsing the launch HTML tables
6. Create a dictionary
7. Create a dataframe using the dictionary
8. Filter the dataframe to only include Falcon 9 launches
9. Export dataset to a .csv file

Data Collection - Scraping

```
static_url = "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922"
response = requests.get(static_url).text
soup = BeautifulSoup(response, "html.parser")
html_tables = soup.find_all('table')
first_launch_table = html_tables[2]
column_names = []
for th in first_launch_table.find_all('th'):
    colname = extract_column_from_header(th)
    if colname != None and len(colname) > 0:
        column_names.append(colname)
```

```
df.to_csv('spacex_web_scraped.csv', index=False)
```

```
print(column_names)
```

```
['Flight No.', 'Date and time ( )', 'Launch site', 'Payload', 'Payload mass', 'Orbit', 'Customer', 'Launch outcome']
```

```
df= pd.DataFrame({ key:pd.Series(value) for key, value in launch_dict.items() })
```

```
df.to_csv('spacex_web_scraped.csv', index=False)
```

Data Wrangling

The objective of data wrangling is to perform exploratory Data Analysis and determine Training Labels.

- After loading the dataset we use the head() function to get a snapshot of the the dataset
- Check for missing values by calculating the percentage of the missing data in each column and check data types for each column to identify categorical and numerical data

```
df.isnull().sum()/len(df)*100
df.dtypes
```

- Calculate the number of launches per site

```
# Apply value_counts() on column LaunchSite
df['LaunchSite'].value_counts()
```

```
CCAFS SLC 40    55
KSC LC 39A      22
VAFB SLC 4E     13
Name: LaunchSite, dtype: int64
```

Data Wrangling

- Calculate the number and occurrence of each orbit
- Calculate the number and occurrence of mission outcome of the orbits
- Create a landing outcome label from Outcome column

```
# Apply value_counts on Orbit column
df['Orbit'].value_counts()
```

```
GTO      27
ISS      21
VLEO     14
PO        9
LEO       7
SSO       5
MEO       3
ES-L1     1
HEO       1
SO        1
GEO       1
Name: Orbit, dtype: int64
```

```
# landing_outcomes = values on Outcome column
landing_outcomes = df['Outcome'].value_counts()
landing_outcomes
```

```
True ASDS      41
None None      19
True RTLS      14
False ASDS      6
True Ocean      5
False Ocean     2
None ASDS       2
False RTLS      1
Name: Outcome, dtype: int64
```

```
landing_class = list(np.where(df['Outcome'].isin(bad_outcomes), 0, 1))
df['Class']=landing_class
df[['Class']].head(8)
```

	Class
0	0
1	0
2	0
3	0
4	0
5	0
6	1
7	1

```
# landing_class = 0 if bad_outcome
# landing_class = 1 otherwise
landing_class = list(np.where(df['Outcome'].isin(bad_outcomes), 0, 1))
```

EDA with Data Visualization

For data visualization the following charts were used-:

- Pie chart to express the landing success rate proportion for each launching site
- Scatter plots to visualize the relationship between Flight number vs Payload Mass, Launch site vs Payload Mass, Orbit vs Flight Number and Orbit vs Payload Mass.
- Categorical Scatter plot to visualize the relationship between Flight number and Launch site.
- Bar Chart to visualize the relationship between success rate of each orbit type
- Line graph to visualize the success rate yearly trend
- Maps with markers and popups to show the launch sites and success rates from each site

EDA with SQL

Using SQL for exploration data analysis we queried the database to -:

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- Listed the date when the first successful landing outcome in ground pad was acheived.
- Listed the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Listed the total number of successful and failure mission outcomes
- Listed the names of the booster_versions which have carried the maximum payload mass. Use a subquery
- Listed the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
- 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.

Build an Interactive Map with Folium

To highlight key points in the map we use-:

- Circles to highlight all the launch site locations on the map.
- Popup to show the site name when launch highlight (circle) is clicked
- Markers to show launch site point locations
- Marker clusters to show the success or failure of the launches on each site
- Lines to show the distances from different proximities such as roads, rails, and coastlines.
- Explain why you added those objects

Build a Dashboard with Plotly Dash

To interactively visualize our dataset with Plotly Dash we used-:

- A dropdown list to choose a launch site to visualize
- A pie chart to visualize the selected launch site first stage landing success and failure rates.
- A range slider to highlight the points of interest in our dataset using payload mass
- A scatter plot to show the correlation between payload and launch site success

Predictive Analysis (Classification)

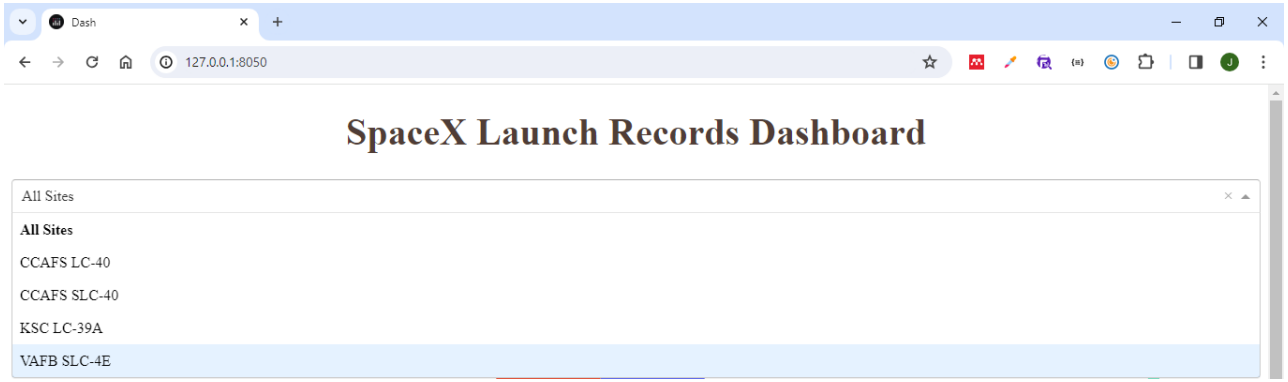
- Summarize how you built, evaluated, improved, and found the best performing classification model
- You need present your model development process using key phrases and flowchart
- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose

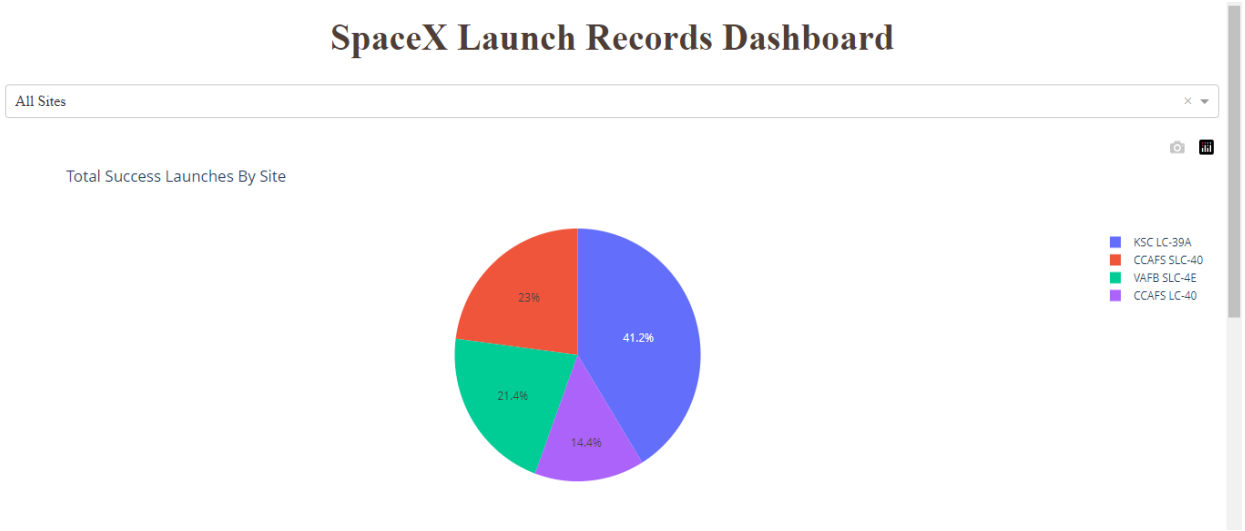
Results

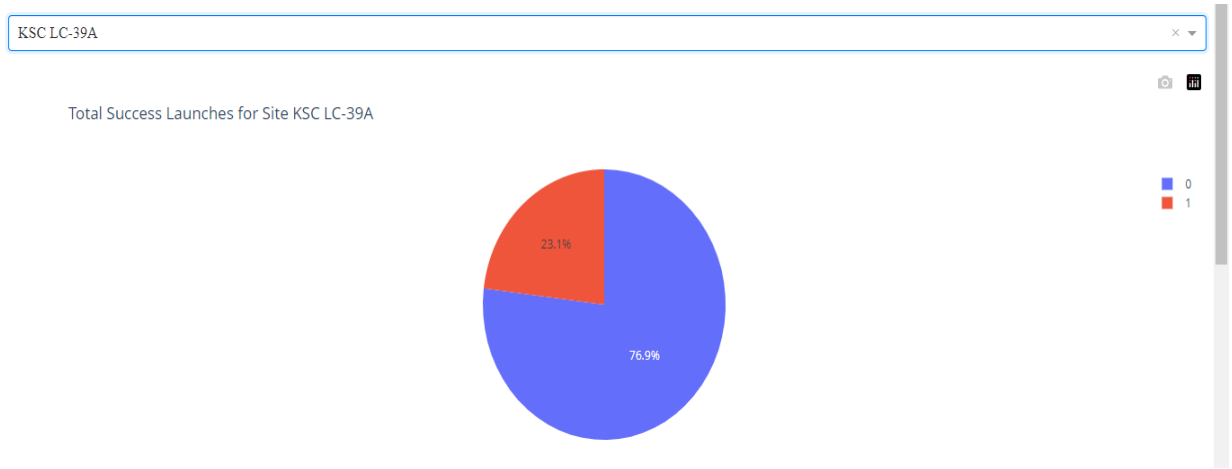
- Exploratory data analysis results
  - Among the 100 launches analyzed, CCAFS SLC 40 emerged as the launch site with the highest frequency.
  - Over time, a positive trend in success rates was observed, indicating an improvement in the success of first-stage landings.
  - VAFB SLC 4E demonstrated the highest landing success rate among the launch sites
  - Specific orbits, including ES-LI, GEO, HEO, and SSO, exhibited a 100% success rate, highlighting their reliability.
  - Conversely, the SO orbit showed a 0% success rate.
  - Geographical visualization revealed that launch sites are strategically located near the equator, leveraging Earth's rotational speed for optimal launches.
  - Coastal launch sites were chosen, providing a safer and less inhabited surrounding in the event of unforeseen complications.

Results

Interactive analytics demo in screenshots

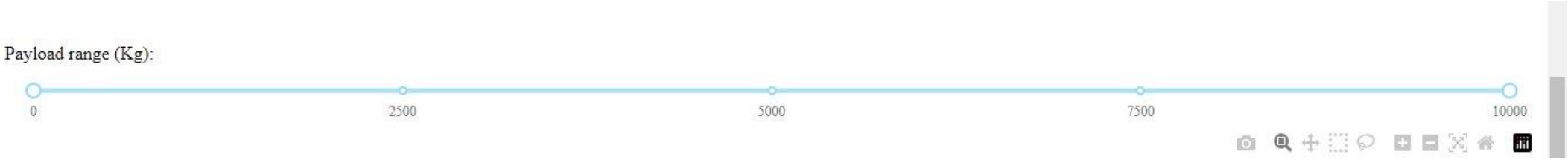


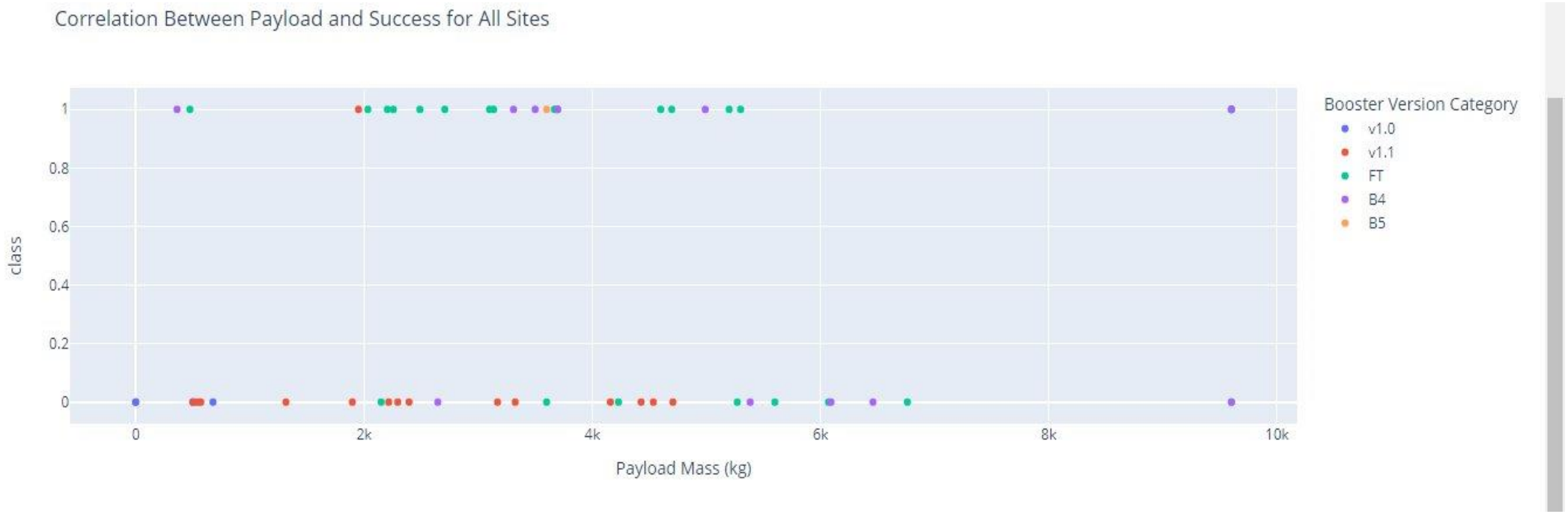




Results

Interactive analytics demo in screenshots



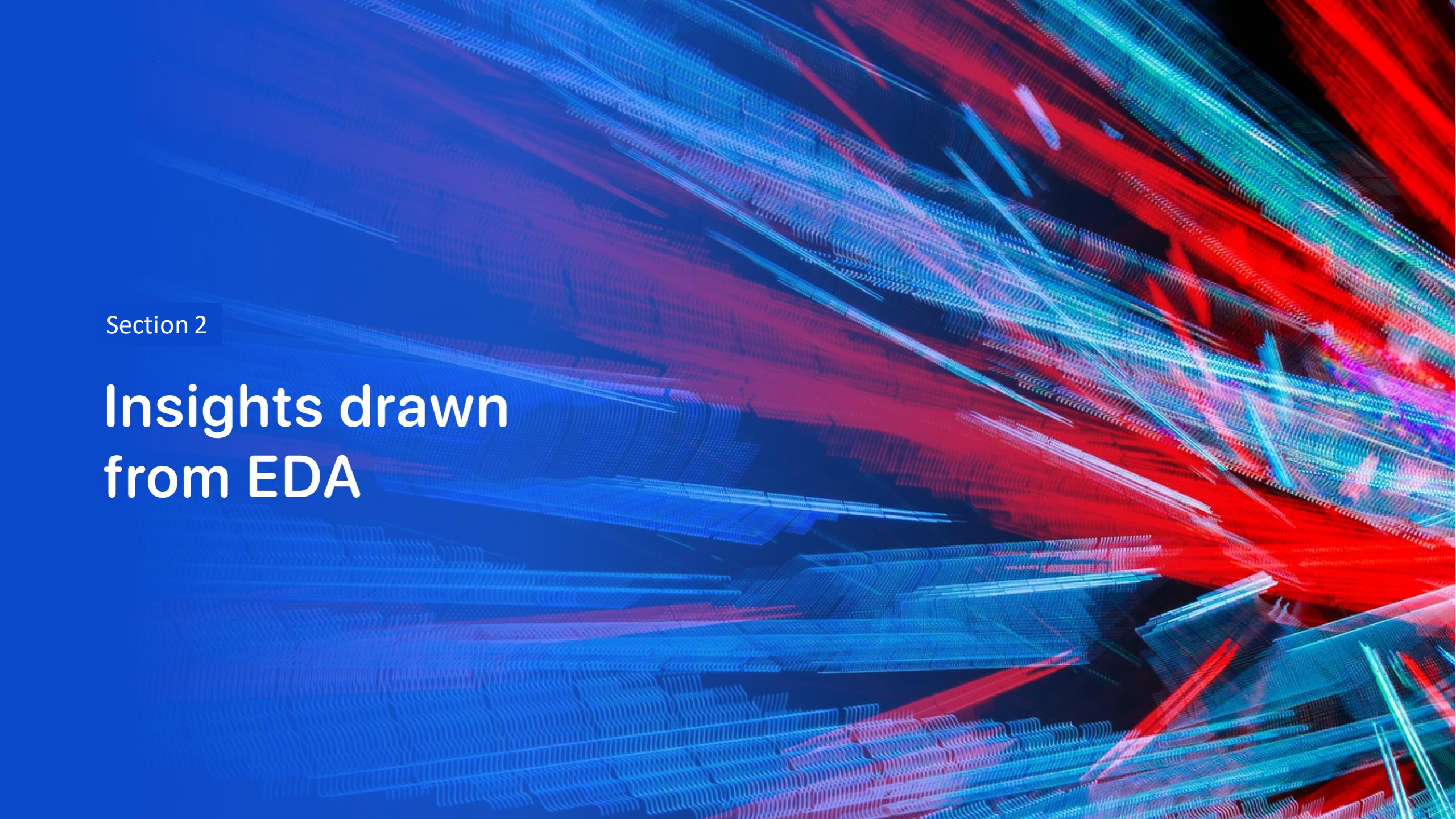


Results

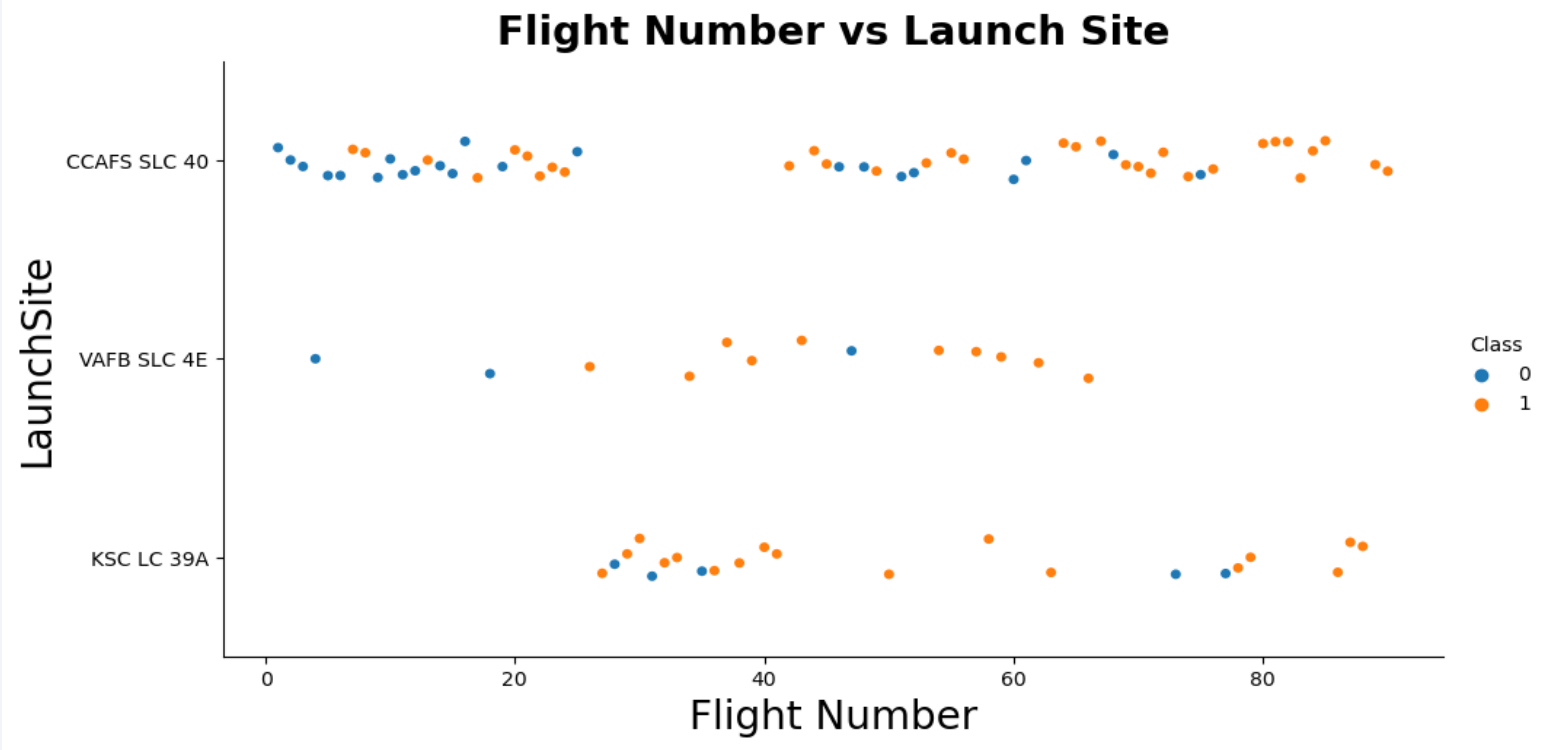
- **Predictive analysis results**
  - Through rigorous testing, all models exhibited commendable performance in predicting first-stage landing success.
  - The decision tree model outperformed others, showcasing superior accuracy and robustness in its predictions.

Section 2

Insights drawn from EDA

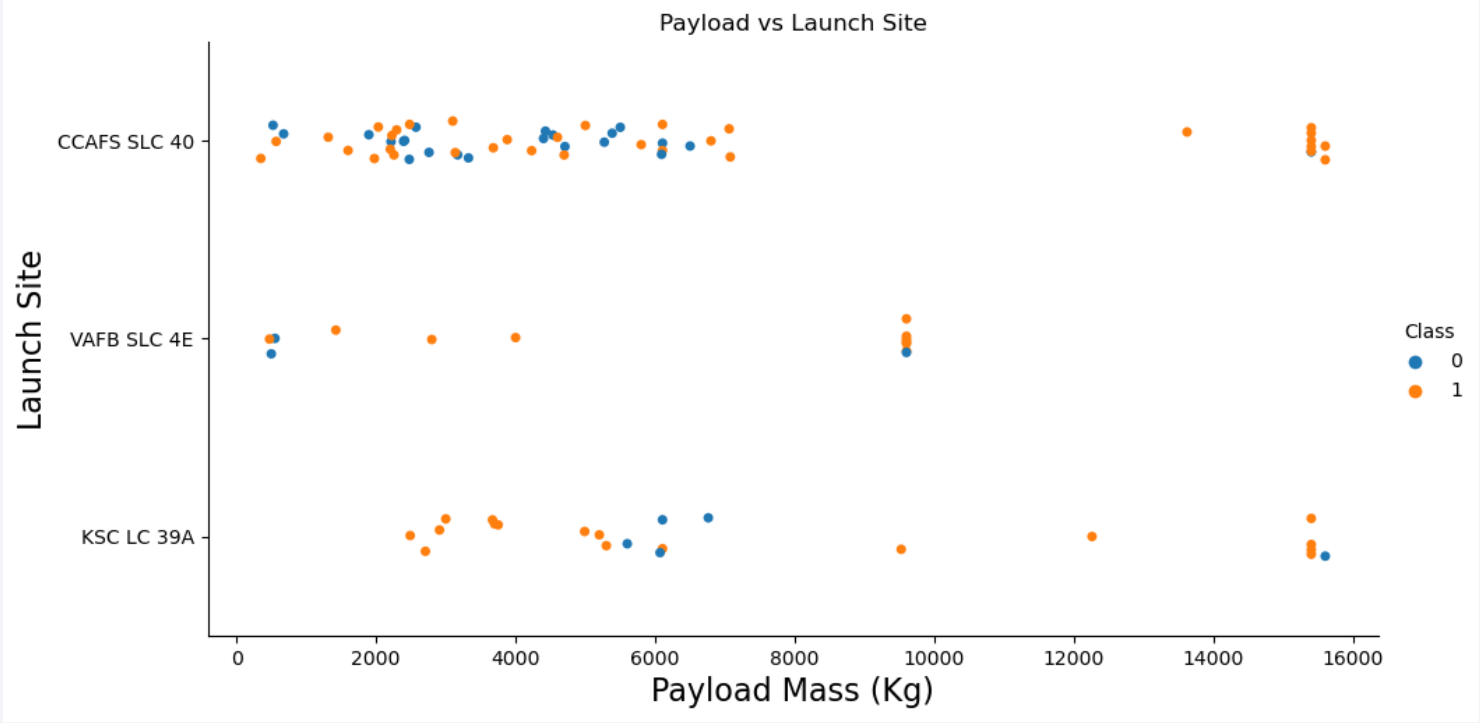


Flight Number vs. Launch Site



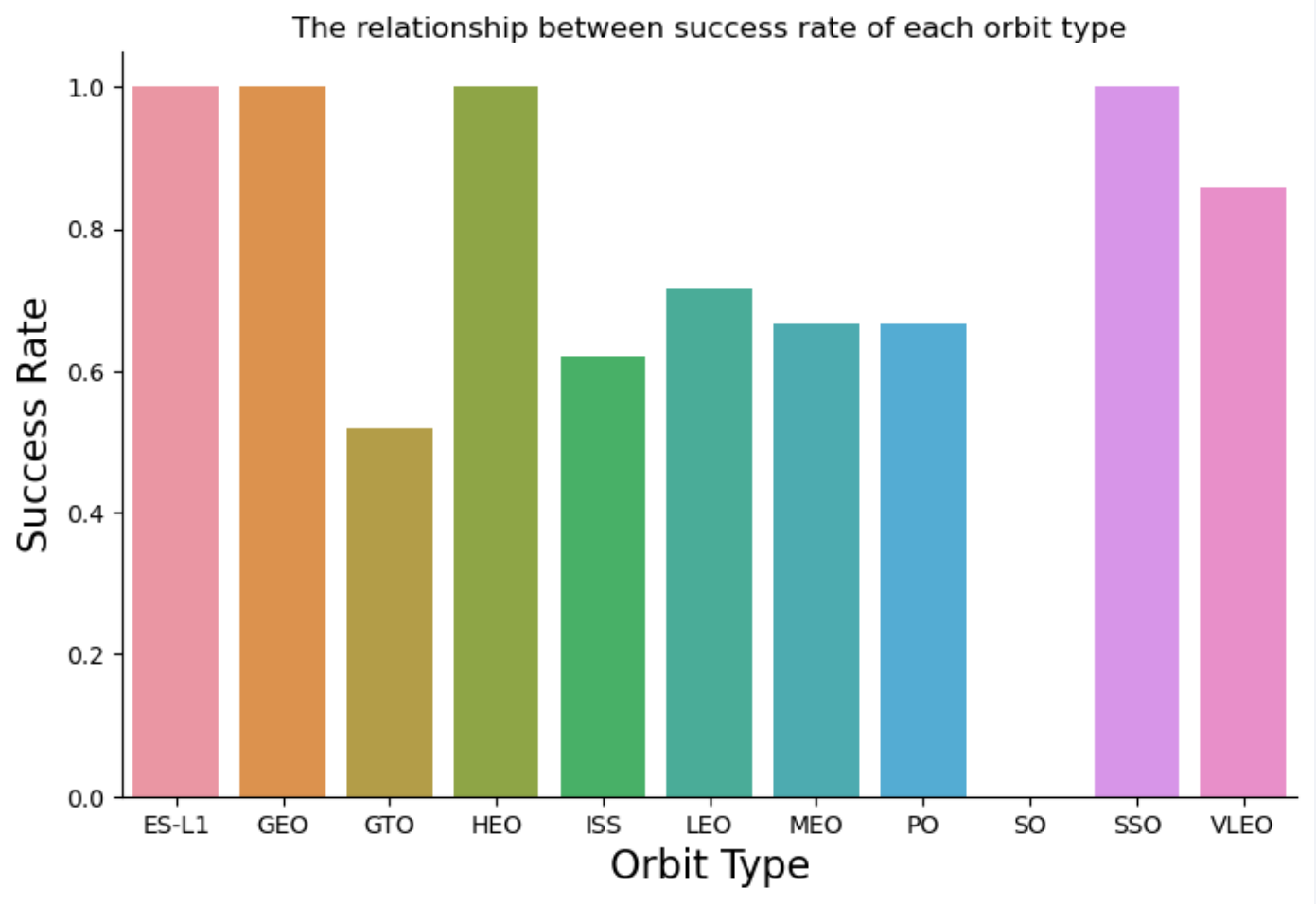
- CCAF LC-40 has launched over 50% of the total launches
- CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%.
- There are more failures than success in the earlier launches and and this is reverse in the later launches.
- Overall the success rate has increased with more launches.

Payload vs. Launch Site



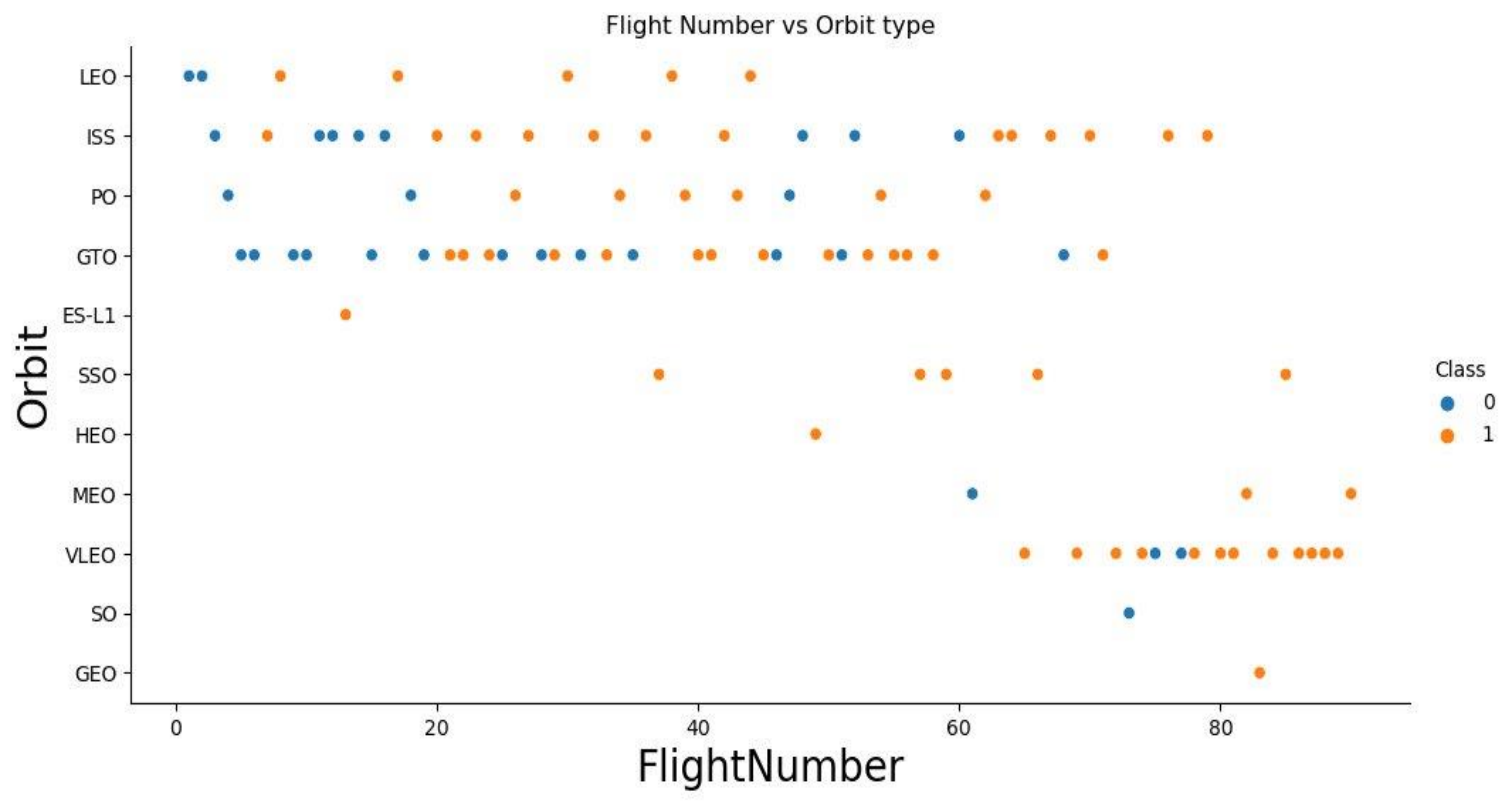
- There is an upward trend, the heavier the payload the higher the success rate as shown by the success rate of launches with payloads heavier than 7500kg
- VAFB-SLC launch site has no rockets launched for heavy payload mass(greater than 10000)
- KSC LC-39A has a 100% success smaller payloads of less than ~5500kg

Success Rate vs. Orbit Type



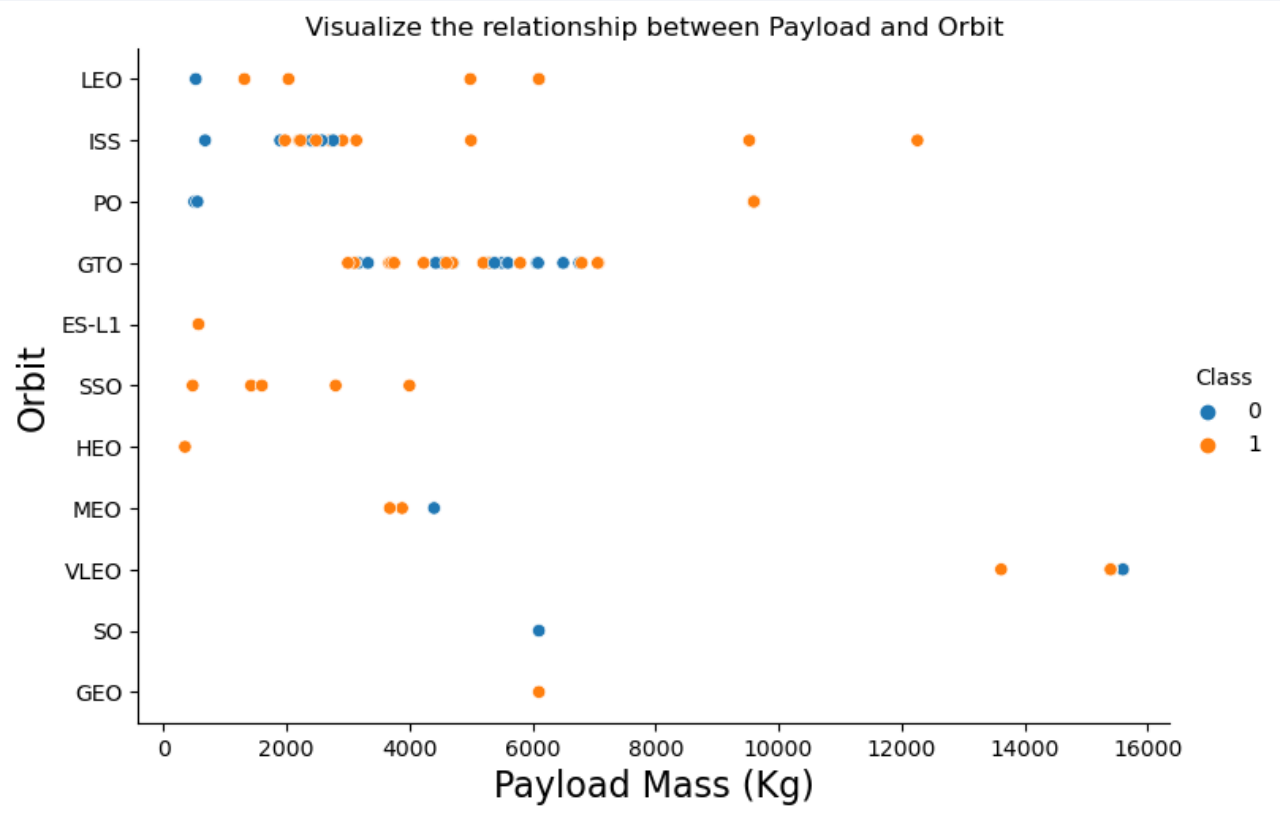
- ES-LI, GEO, HEO and SSO have 100% success rates
- SO has a success rate of zero

Flight Number vs. Orbit Type



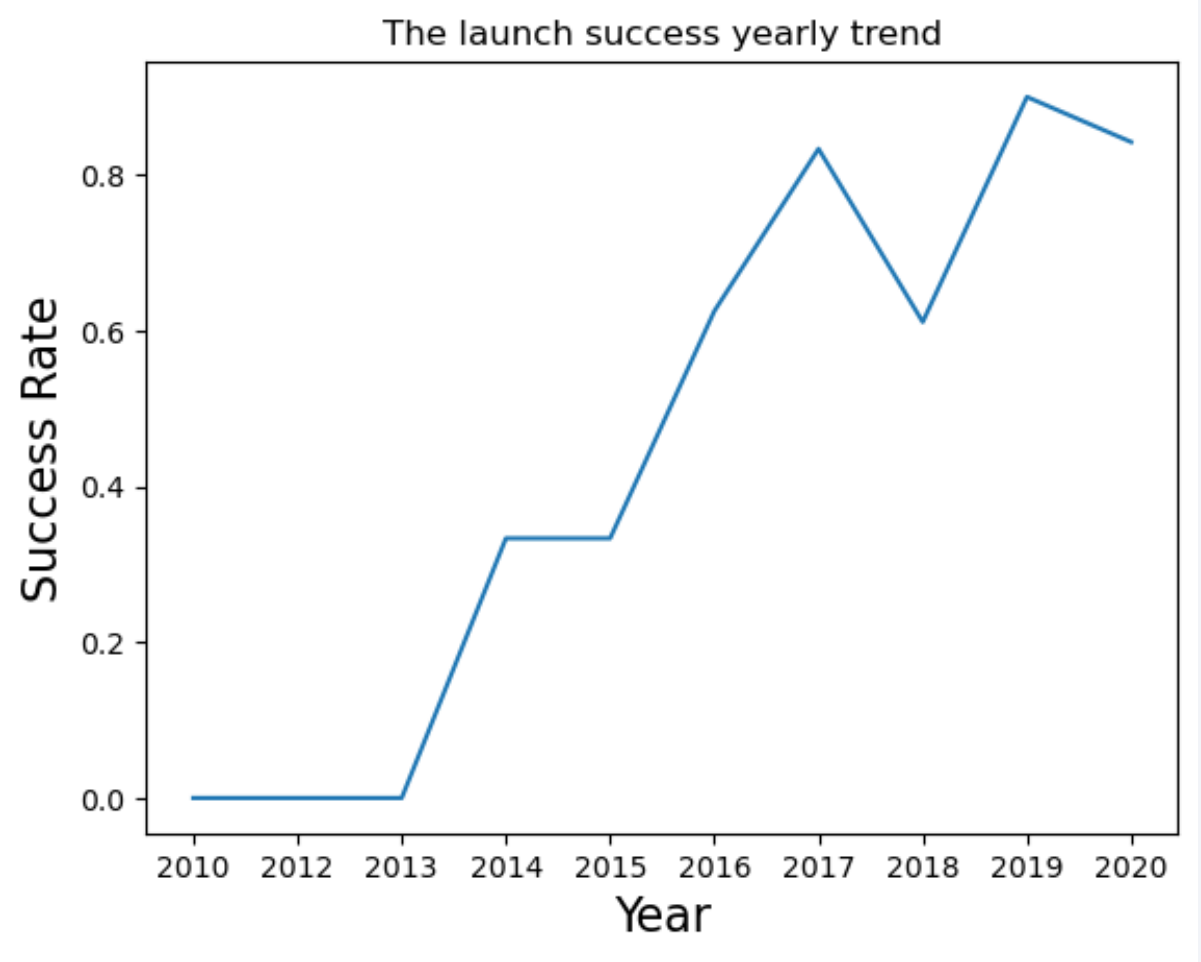
- LEO orbit the Success appears related to the number of flights
- ES-L1, SSO, HEO all have 100% success rates
- With successive launches the success rate improves for all orbits but SO with a single launch

Payload vs. Orbit Type



- With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.
- However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.

Launch Success Yearly Trend



- The success rate since 2013 kept increasing till 2020

All Launch Site Names

Display the names of the unique launch sites in the space mission

```
%sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE;
```

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

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

- Using the select query with distinct or unique launchsite on the spacextable returns a table with only unique values

Launch Site Names Begin with 'CCA'

Display 5 records where launch sites begin with the string 'CCA'

```
%%sql SELECT * FROM SPACEXTABLE
WHERE Launch_Site like 'CCA%'
LIMIT 5;
```

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

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

- Using the select query, select all columns where launch site is 'CCA' and limit output to 5 rows

Total Payload Mass

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

```
%%sql SELECT SUM(PAYLOAD_MASS__KG_) AS Total_Payload_Mass_KG FROM SPACEXTABLE
WHERE Customer = 'NASA (CRS)';
```

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

Total_Payload_Mass_KG
45596

- Return total payload mass using the select query and the sum() function on the spacextable where customer is 'NASA (CRS)'

Average Payload Mass by F9 v1.1

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

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

AVG(PAYLOAD_MASS__KG_)
2928.4

- Query the spacextable to return the average payload mass where booster_version is 'F9 v1.1'

First Successful Ground Landing Date

```
%%sql SELECT MIN("DATE") FROM SPACEXTABLE
WHERE Landing_Outcome = "Success (ground pad)";
```

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

MIN("DATE")
2015-12-22

- Query the spacextable to return the earliest date of successful landing

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%%sql SELECT Booster_Version FROM SPACEXTABLE
WHERE PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000 AND Landing_Outcome = "Success (drone ship)";
```

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

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

- Query the spacextable to return the booster version that has a successful landing with a payload mass that is greater than 4000kg and less than 6000kg

Total Number of Successful and Failure Mission Outcomes

```
%%sql SELECT Mission_Outcome, COUNT(*)
FROM SPACEXTABLE
GROUP BY Mission_Outcome;
```

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

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

- Query the spacextable to return the totals of mission outcomes

Boosters Carried Maximum Payload

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

```
%%sql SELECT Booster_Version FROM SPACEXTABLE
WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE);
```

```
 * 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

- Query the spacextable to return the booster versions with the highest payload mass

2015 Launch Records

```
%%sql SELECT substr(Date, 6, 2) AS Months, Date, Landing_Outcome, Booster_Version, Launch_Site
FROM SPACEXTABLE
WHERE substr(Date, 0, 5) = '2015' and Landing_Outcome = 'Failure (drone ship)';
```

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

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

- Query the spacextable to return months, date, landing outcome, booster version and launch site records for 2015 where landing outcome is 'Failure (drone ship)'

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

```
%%sql SELECT Landing_Outcome, COUNT(*) AS "Count"
FROM SPACEXTABLE
WHERE "Date" BETWEEN "2010-06-04" and "2017-03-20"
GROUP BY Landing_Outcome
ORDER BY Count DESC;
```

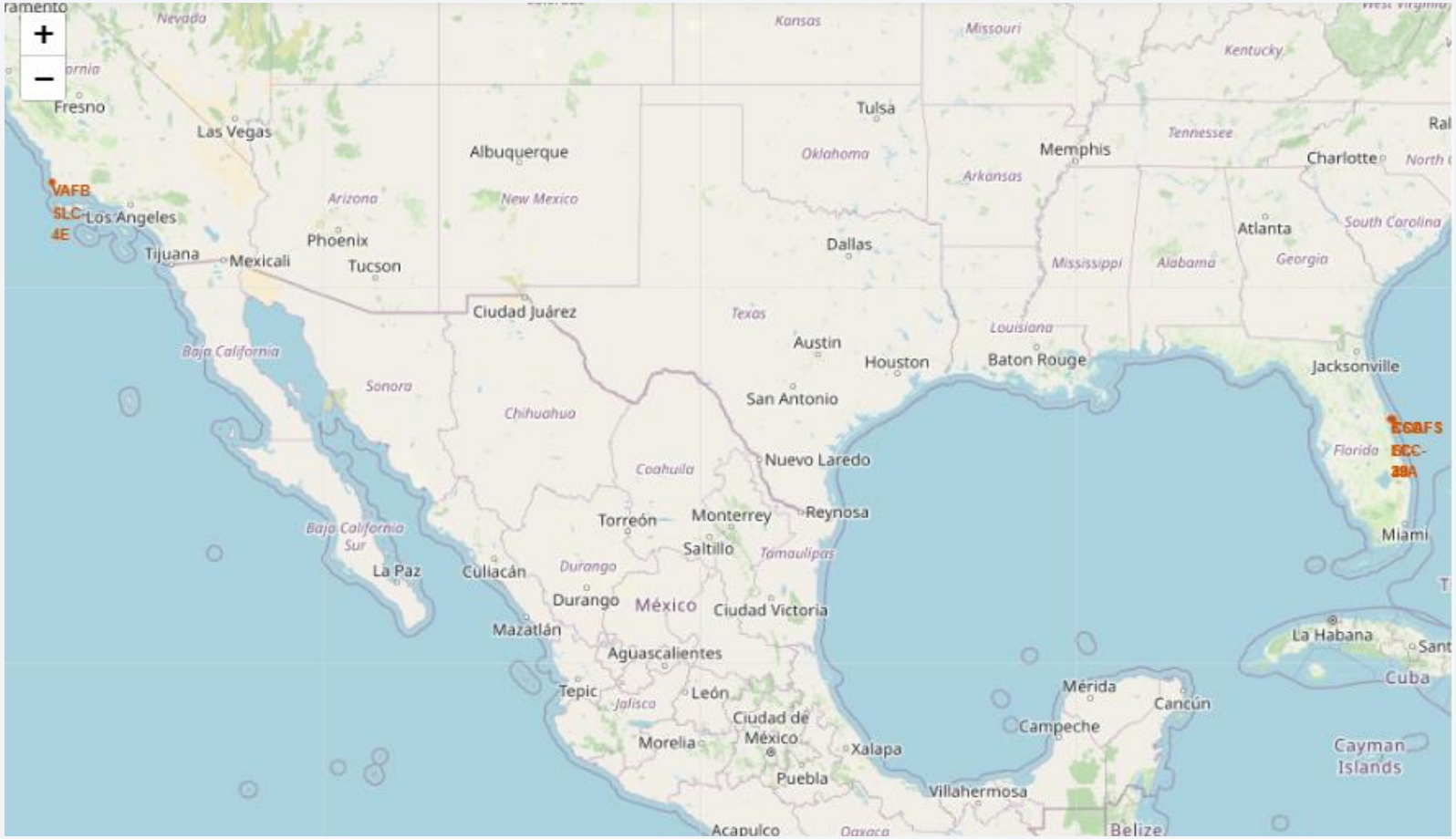
```
 * sqlite:///my_data1.db
Done.
```

Landing_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

- Query the spacextable to return landing outcomes and their totals between 2010-06-04 and 2017-03-20 in descending order

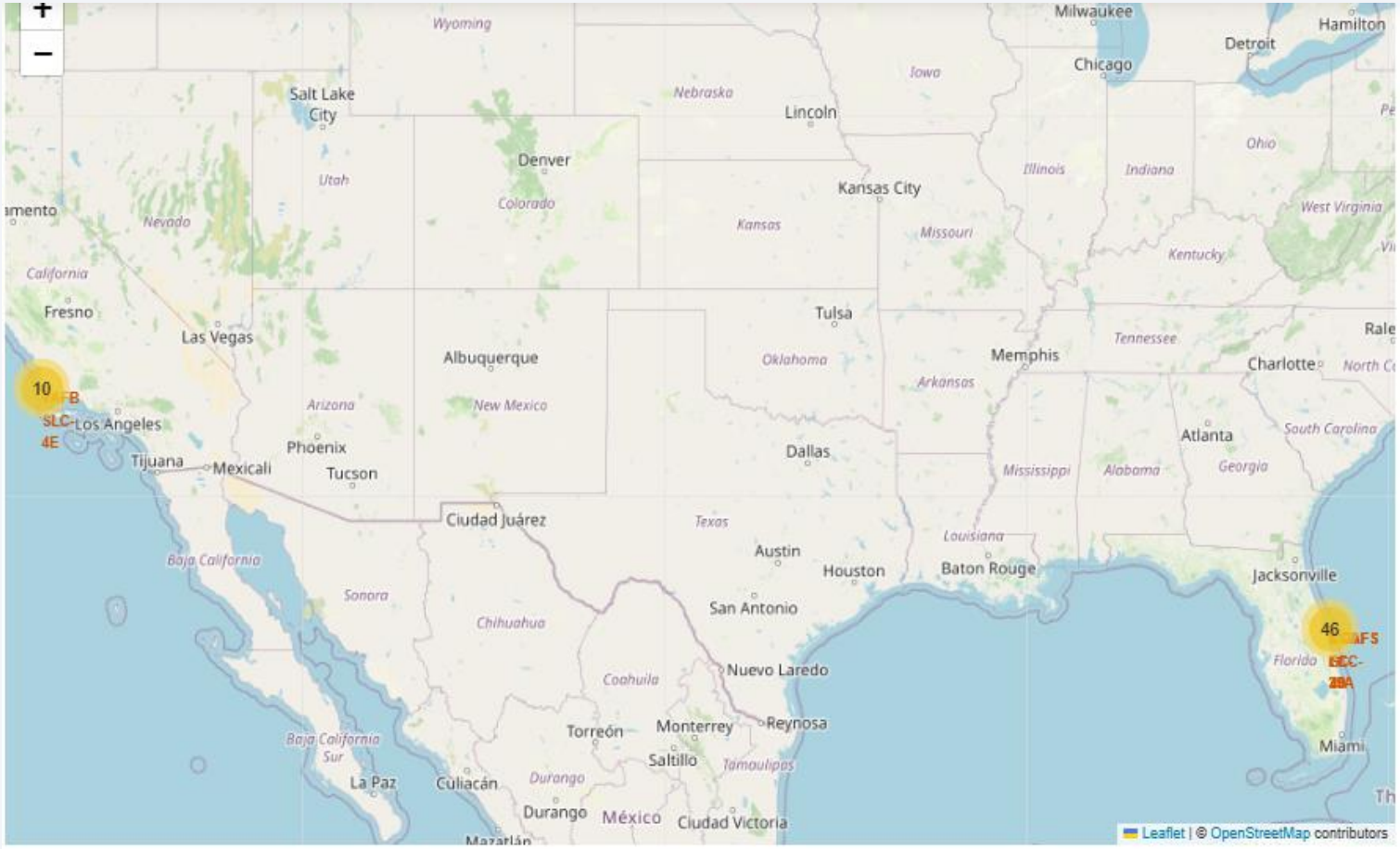

Section 3
Launch Sites
Proximities Analysis

Launch Sites Map



- Map with markers highlighting launch locations

Launch Sites Map 2



- Launch sites with each site's total number of launches highlighted.

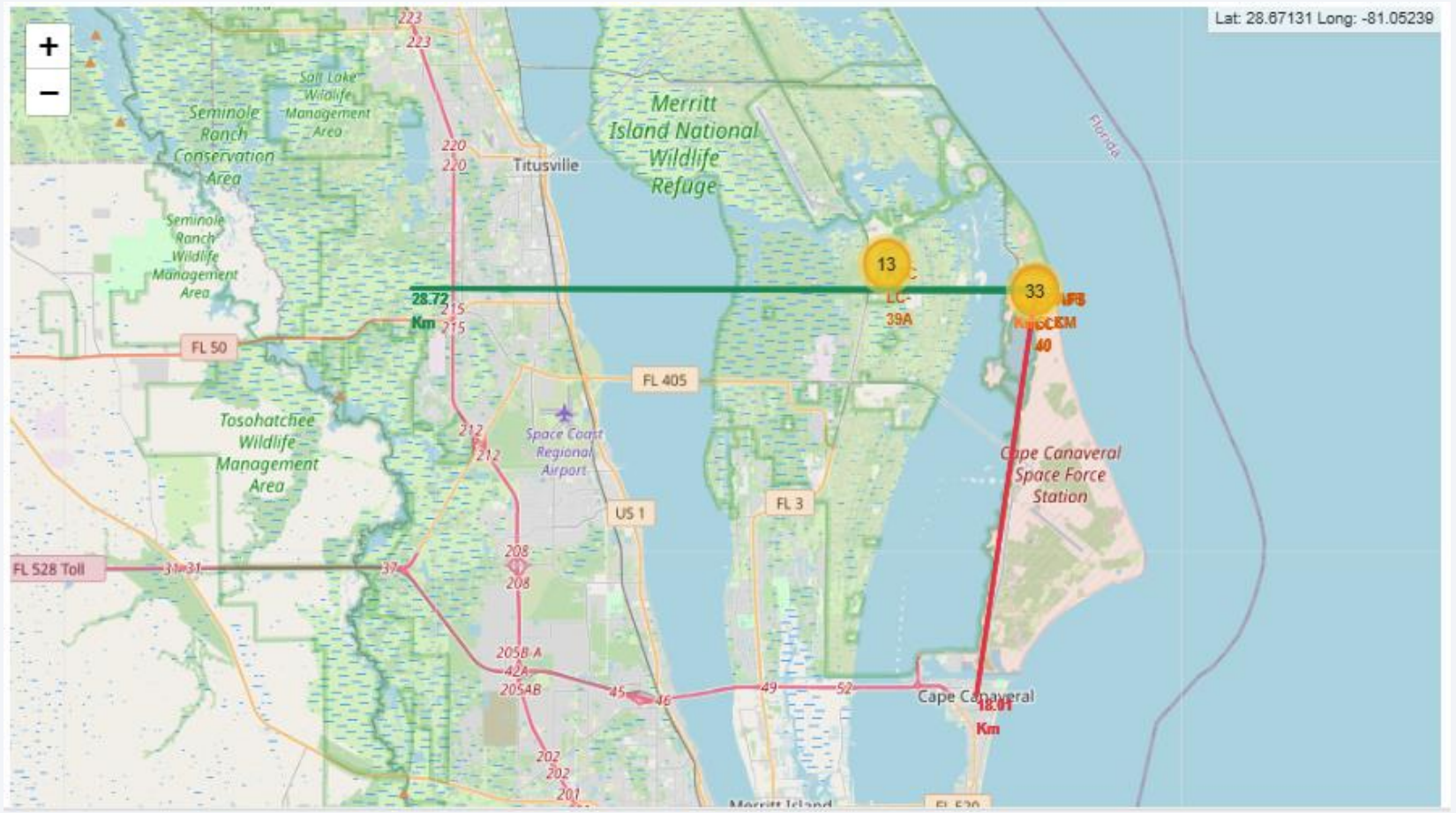
Launch Sites Map 2

- A close up image clearly showing launch sites with their total launch counters

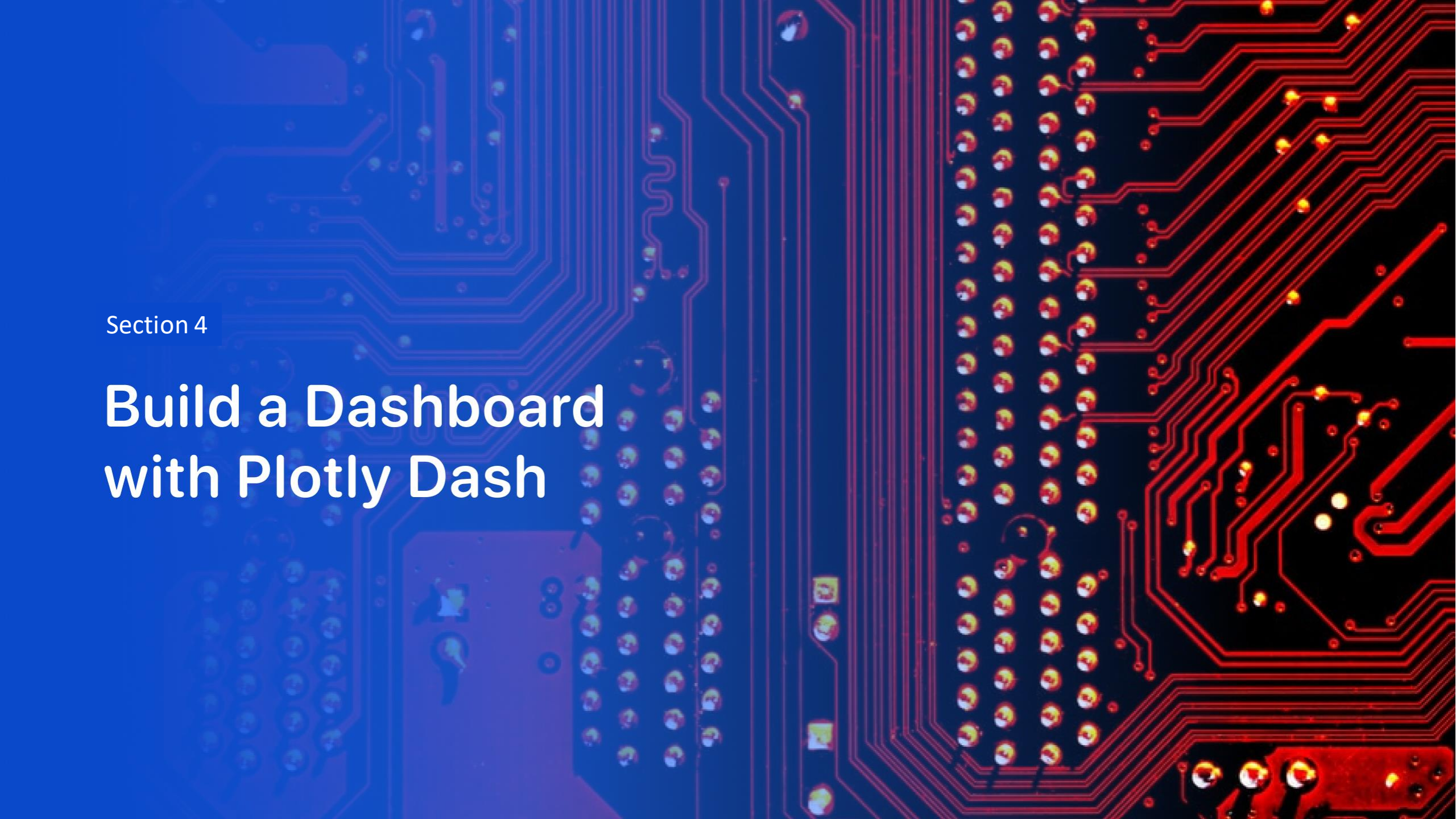
Launch Sites Map 2

An even closer view of the launch site showing the total launches and the success and failures

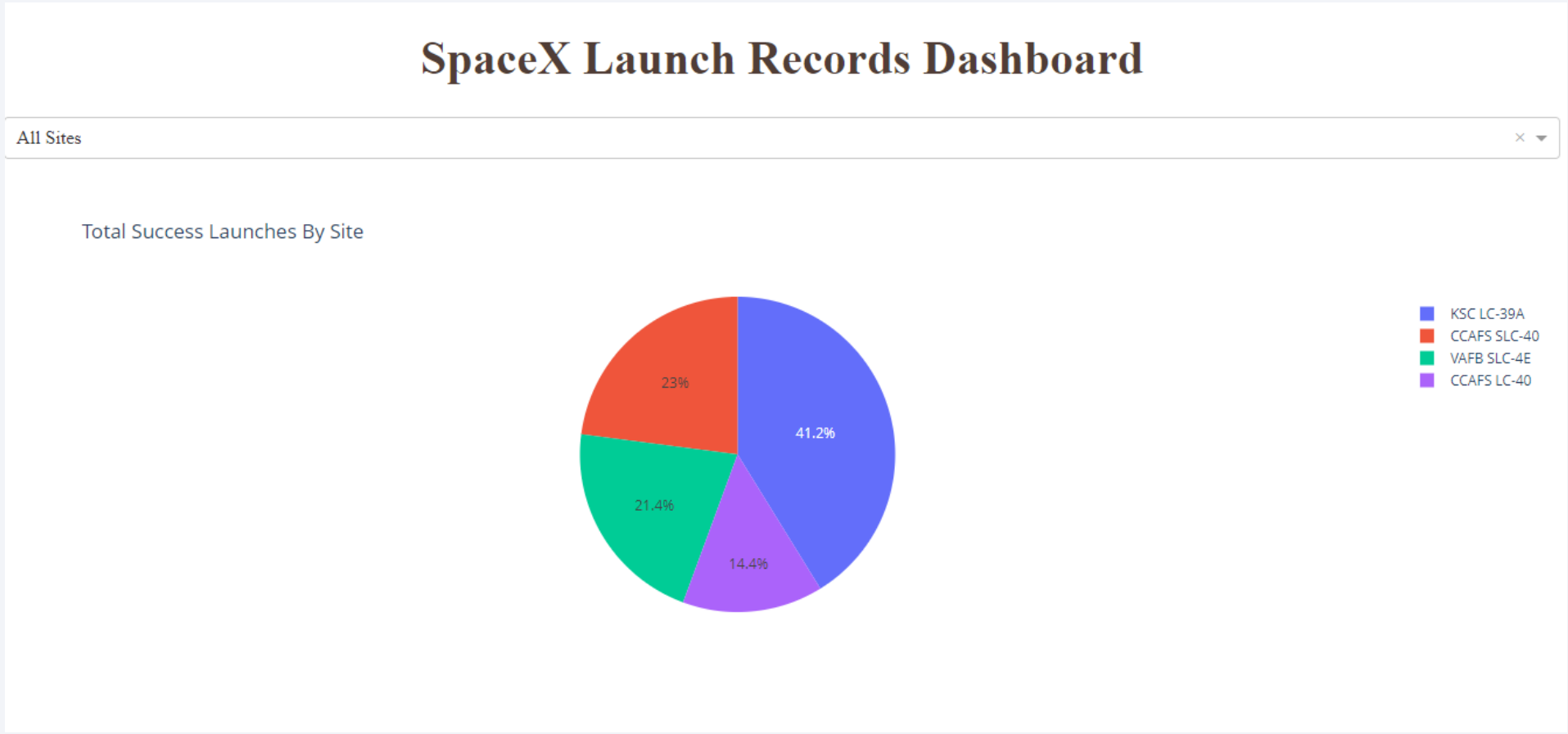
Launch Site Map in relation to its surroundings



- Map showing the launch site in relation to its closest roads, rails, coastline and town.

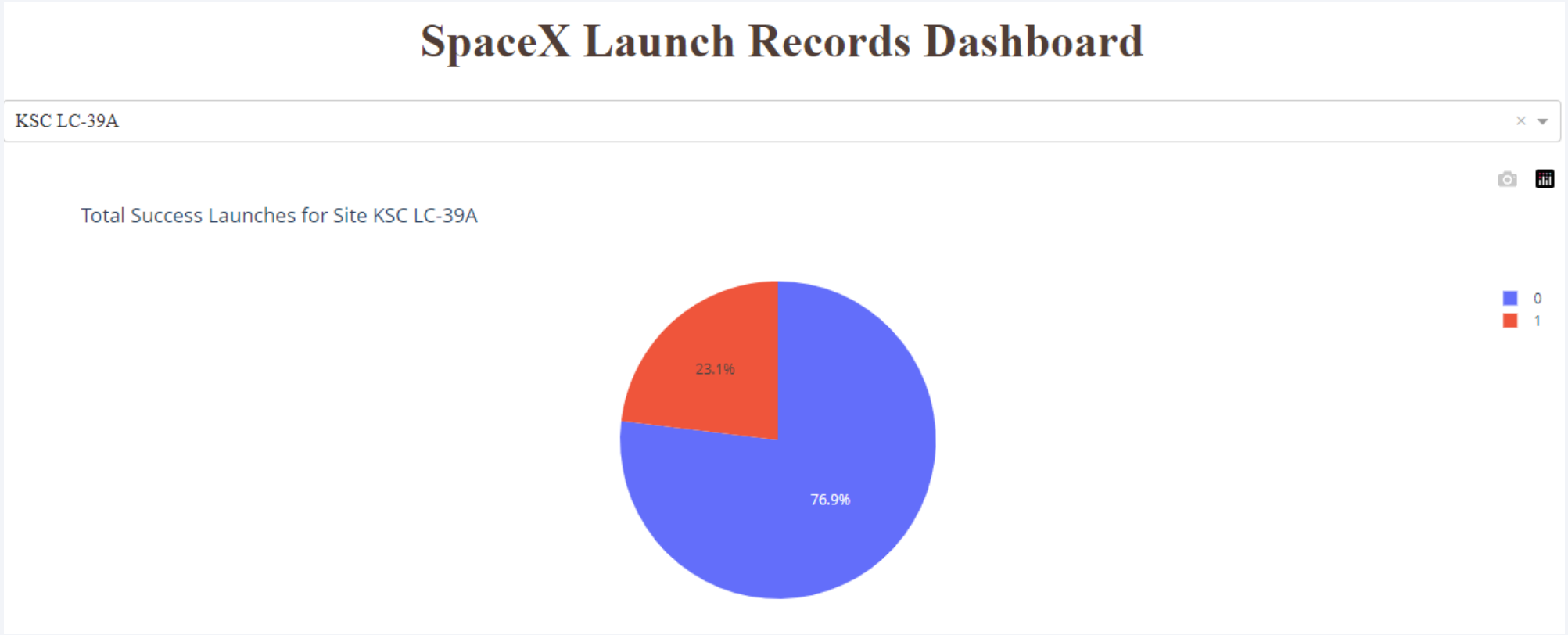

Section 4
Build a Dashboard with Plotly Dash

Success Rate Pie Chart (All Launch Sites)



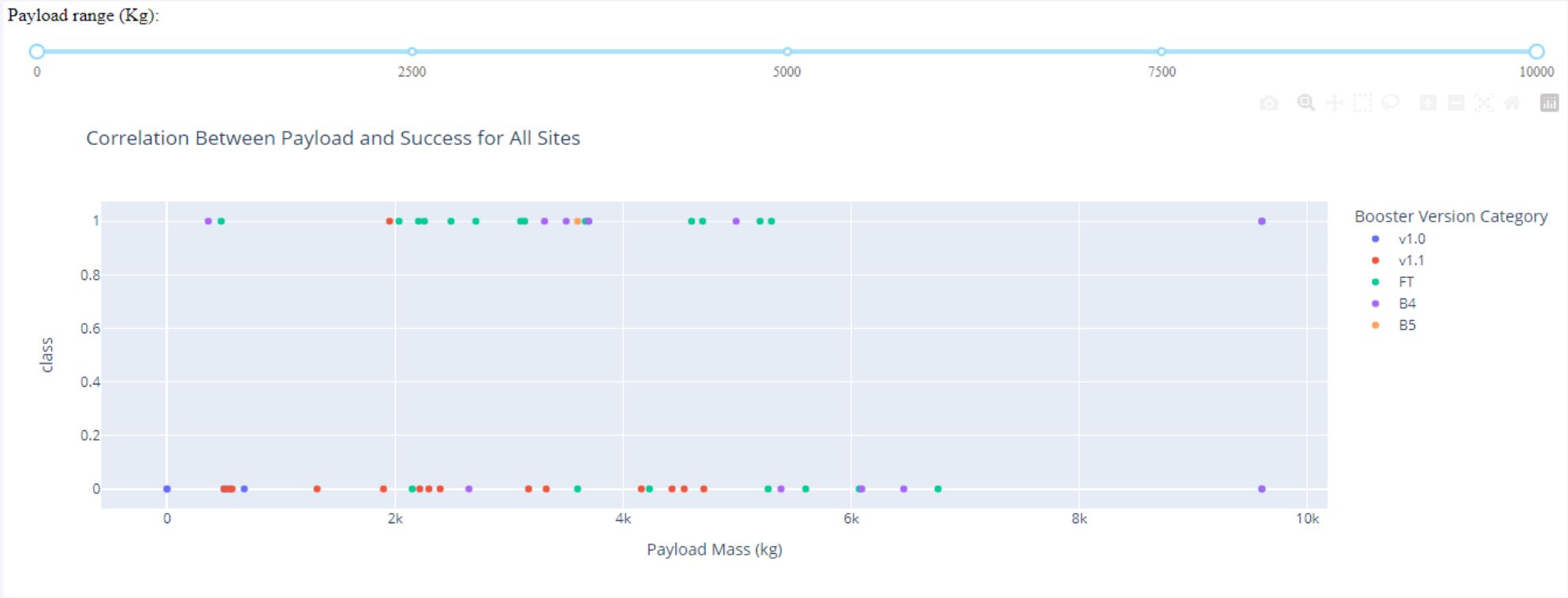
- Pie Chart showing success rate total shares for all launch sites

KSC LC-39A Pie Chart



- Pie Chart showing the success and failure ratio for launch site KSC LC-39A which has the highest success rate.

Payload vs Launch Outcome Scatter Plot On Dash



- 1 shows success and 0 failure
- Payloads between 2000 and 5000kg have the highest success rate

Section 5

Predictive Analysis (Classification)

Classification Accuracy

All models performed well with the decision tree a little better than others.

All this may be a result of a small dataset we used and it remains to be seen if the provided with a lot more data how the models would perform.

	LogReg	SVM	Tree	KNN
Jaccard_Score	0.800000	0.800000	0.800000	0.800000
F1_Score	0.888889	0.888889	0.888889	0.888889
Accuracy	0.833333	0.833333	0.833333	0.833333

```
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 0.9017857142857144
Best params is : {'criterion': 'gini', 'max_depth': 18, 'max_features': 'sqrt', 'min_samples_leaf': 2, 'min_samples_spli
t': 2, 'splitter': 'best'}
```

Confusion Matrix

All the confusion matrices were identical, and the fact that we had false positive is a not good.

Precision = TP / (TP +FP) -> 12 / 15 = .80

Recall = TP / (TP +FN) -> 12 / 12 = 1

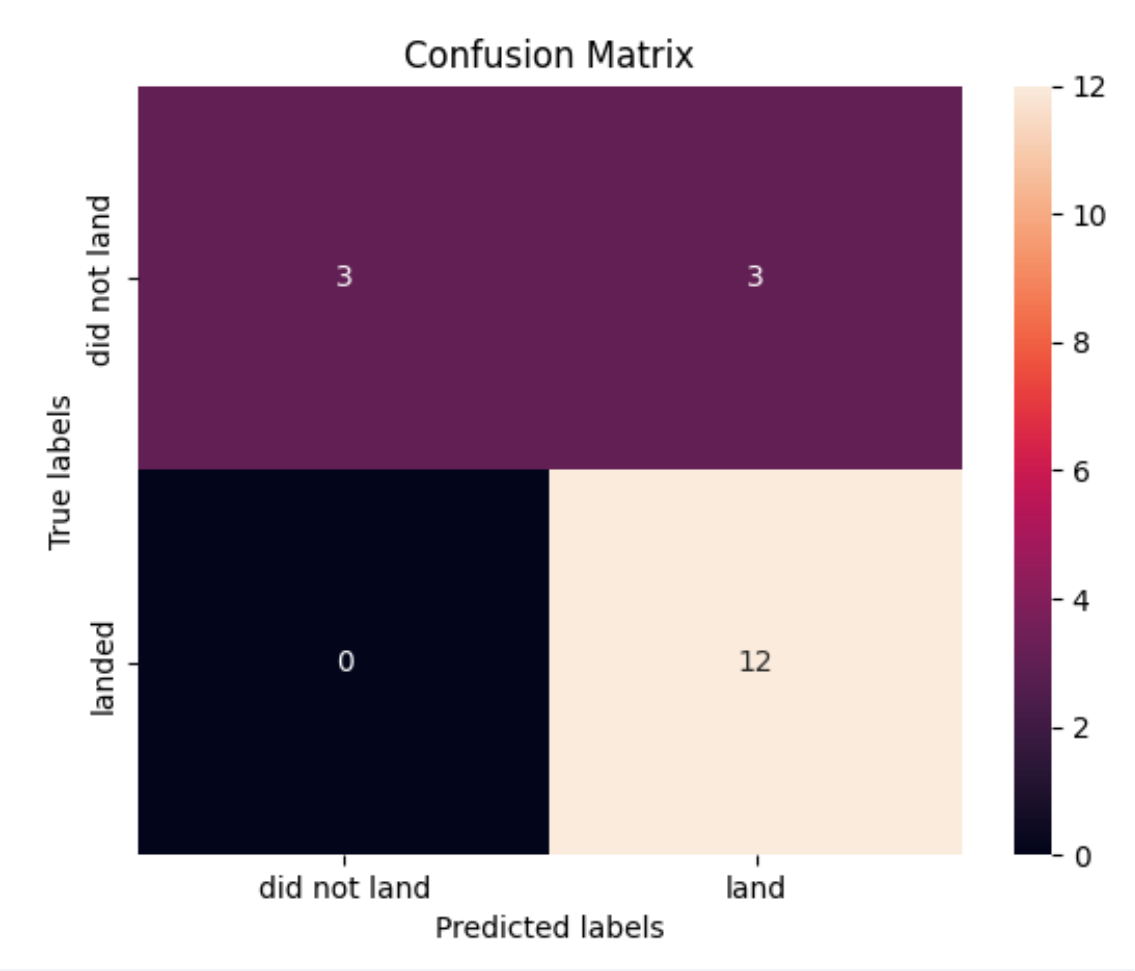
F1 Score = 2 * (Precision * Recall) / (Precision + Recall)

2 * (.8 * 1) / (.8 + 1) = .89

Accuracy = (TP + TN) / (TP + TN + FP + FN) = .833

12 True positive, 3 True negative, 3 False positive

0 False Negative



Conclusions

- **Exploratory Data Analysis:**
  - Among the 100 launches analyzed, CCAFS SLC 40 emerged as the launch site with the highest frequency.
  - Over time, a positive trend in success rates was observed, indicating an improvement in the success of first-stage landings.
  - VAFB SLC 4E demonstrated the highest landing success rate among the launch sites
- **Orbit vs Success Rate Analysis:**
  - Specific orbits, including ES-LI, GEO, HEO, and SSO, exhibited a 100% success rate, highlighting their reliability.
  - Conversely, the SO orbit showed a 0% success rate.
- **Visualization Analysis:**
  - Geographical visualization revealed that launch sites are strategically located near the equator, leveraging Earth's rotational speed for optimal launches.
  - Coastal launch sites were chosen, providing a safer and less inhabited surrounding in the event of unforeseen complications.
- **Data Model Outcomes:**
  - All models exhibited commendable performance in predicting first-stage landing success.
  - The decision tree model outperformed others

Appendix

All related codes, outputs, datasets and this presentation can be found on Github-;

https://github.com/justicebhekani88/IBM_Data_Science_Capstone_Project

Thank you!