IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Chia Wen Yung
24th December 2024

Outline

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

Executive Summary

This project aims to predict if the first stage of a rocket will land successfully using a SpaceX dataset containing Falcon 9 rocket attributes and launch sites. By leveraging machine learning models, the analysis identifies critical factors influencing landing success rate and provides actionable insights for investors and stakeholders. The project concludes with a regression model capable of accurately predicting landing success rate based on key features such as launch sites, payload mass and booster version.

Introduction

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. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

Key questions addressed in this project:

- Which factors significantly influence the first stage landing success rate?
- How accurately can we predict first stage landing success rate using machine learning models?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- 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

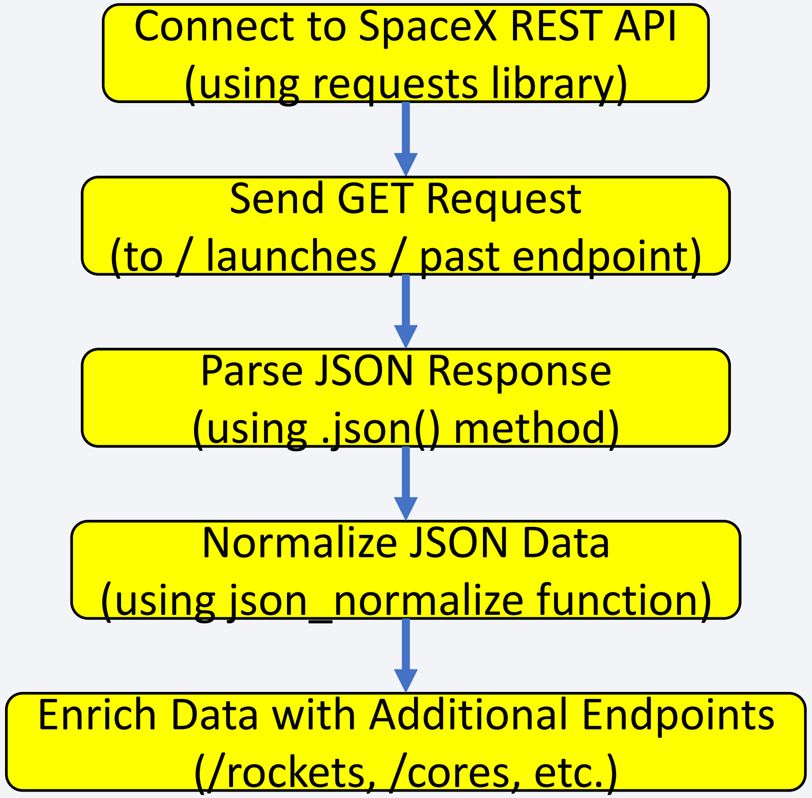
Data Collection

SpaceX launch data is collected from two primary sources:

1. SpaceX REST API
2. Web scraping Wiki pages

Data Collection – SpaceX API

- GitHub URL (https://github.com/wychia5201314/Coursera-IBM-Applied-Data-Science-Capstone/blob/main/Data-Collection/API/jupyter-labs-spacex-data-collection-api.ipynb)



Connect to SpaceX REST API (using requests library)

Send GET Request (to / launches / past endpoint)

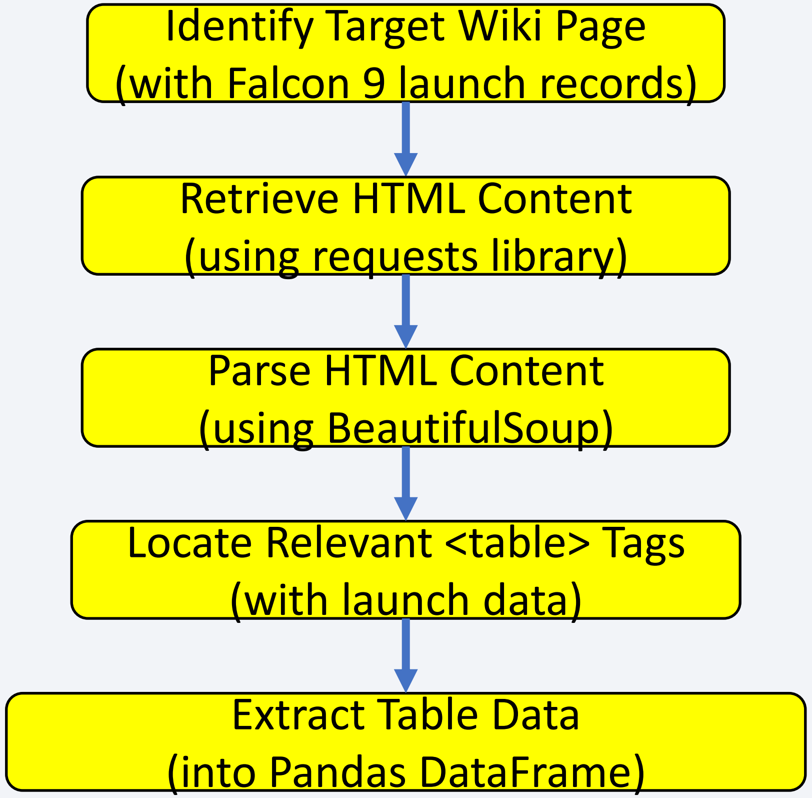
Parse JSON Response (using .json() method)

Normalize JSON Data (using json_normalize function)

Enrich Data with Additional Endpoints (/rockets, /cores, etc.)

Data Collection - Scraping

- GitHub URL (https://github.com/wychia5201314/Coursera-IBM-Applied-Data-Science-Capstone/blob/main/Data-Collection/Web-Scraping/jupyter-labs-webscraping.ipynb)



Identify Target Wiki Page (with Falcon 9 launch records)

Retrieve HTML Content (using requests library)

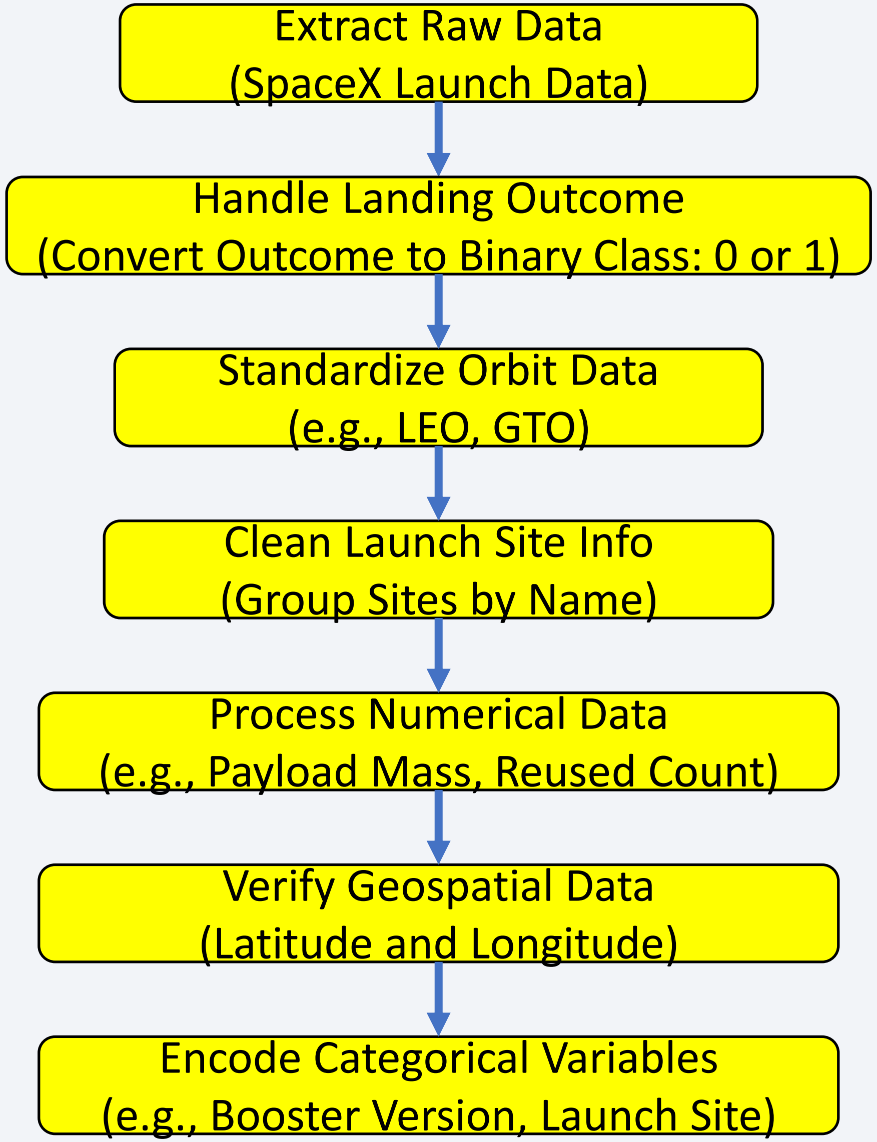
Parse HTML Content (using BeautifulSoup)

Locate Relevant <table> Tags (with launch data)

Extract Table Data (into Pandas DataFrame)

Data Wrangling

- GitHub URL (https://github.com/wychia5201314/Coursera-IBM-Applied-Data-Science-Capstone/blob/main/Data-Wrangling/labs-jupyter-spacex-Data%20wrangling.ipynb)



Extract Raw Data
(SpaceX Launch Data)

Handle Landing Outcome
(Convert Outcome to Binary Class: 0 or 1)

Standardize Orbit Data
(e.g., LEO, GTO)

Clean Launch Site Info
(Group Sites by Name)

Process Numerical Data
(e.g., Payload Mass, Reused Count)

Verify Geospatial Data
(Latitude and Longitude)

Encode Categorical Variables
(e.g., Booster Version, Launch Site)

EDA with Data Visualization

1. <u>**Relationship between Flight Number and Launch Site**</u>

Chart: Scatter plot using catplot.

Purpose: To observe how flight numbers are distributed across different launch sites, with success class (hue) highlighting successful vs. unsuccessful landing. A scatter plot effectively shows the categorical-numerical relationship.

EDA with Data Visualization

2. **Relationship between Payload Mass and Launch Site**

Chart: Scatter plot using catplot.

Purpose: To analyze the relationship between payload mass (numerical) and launch site (categorical). This helps identify if specific payload ranges are associated with certain launch sites, with success class indicated by hue.

EDA with Data Visualization

3. **<u>Relationship between Success Rate of Each Orbit Type</u>**

Chart: Bar chart.

Purpose: To compare success rates across different orbit types. A bar chart is ideal for summarizing the mean success rate for categorical variables (orbits) and provides a clear visual comparison.

EDA with Data Visualization

4. **Relationship between Flight Number and Orbit Type**

Chart: Scatter plot using catplot.

Purpose: To analyze the relationship between flight number (numerical) and orbit type (categorical). This helps identify if specific flight number ranges are associated with certain orbit type, with success class indicated by hue.

EDA with Data Visualization

5. <u>**Relationship between Payload Mass and Orbit Type**</u>

Chart: Scatter plot using catplot.

Purpose: To analyze the relationship between payload mass (numerical) and orbit type (categorical). This helps identify if specific payload ranges are associated with certain orbit type, with success class indicated by hue.

EDA with Data Visualization

6. **Launch Success Yearly Trend**

Chart: Line chart.

Purpose: To observe how the success rate of launches changes over time. A line chart is suitable for showing trends over a continuous variable like time (years).

- GitHub URL (https://github.com/wychia5201314/Coursera-IBM-Applied-Data-Science-Capstone/blob/main/Exploratory-Data-Analysis/Data-Visualization/edadataviz.ipynb)

EDA with SQL

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

```
SELECT DISTINCT Launch_Site FROM SPACEXTABLE;
```

- Display 5 records where launch sites begin with the string ‘CCA’

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

EDA with SQL

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

```
SELECT SUM(PAYLOAD_MASS__KG_) AS TOTAL_PAYLOAD_MASS FROM
SPACEXTABLE WHERE Customer='NASA (CRS)'
```

- Display average payload mass carried by booster version F9 v1.1

```
SELECT AVG(PAYLOAD_MASS__KG_) AS AVERAGE_PAYLOAD_MASS FROM
SPACEXTABLE WHERE Booster_Version='F9 v1.1'
```

EDA with SQL

- List the date when the first successful landing outcome in ground pad was acheived.

```
SELECT MIN(Date) AS FIRST_DATE FROM SPACEXTABLE WHERE
Landing_Outcome='Success (ground pad)‘
```

- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
SELECT Booster_Version FROM SPACEXTABLE WHERE
Landing_Outcome='Success (drone ship)' AND PAYLOAD_MASS__KG_
BETWEEN 4001 AND 5999
```

EDA with SQL

- List the total number of successful and failure mission outcomes

```
SELECT COUNT(Mission_Outcome) AS SUCCESS_MISSION_OUTCOME FROM
SPACEXTABLE WHERE Mission_Outcome LIKE '%Success%‘


SELECT COUNT(Mission_Outcome) AS FAILURE_MISSION_OUTCOME FROM
SPACEXTABLE WHERE Mission_Outcome LIKE '%Fail%'
```

EDA with SQL

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

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

EDA with SQL

- List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

```
SELECT SUBSTR(Date, 6, 2) AS Month_Name, Landing_Outcome,
Booster_Version, Launch_Site FROM SPACEXTABLE WHERE
Landing_Outcome='Failure (drone ship)' AND SUBSTR(Date, 0, 5)='2015'
```

EDA with SQL

- 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.

```
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;
```

- GitHub URL (https://github.com/wychia5201314/Coursera-IBM-Applied-Data-Science-Capstone/blob/main/Exploratory-Data-Analysis/SQL/jupyter-labs-eda-sql-coursera_sqllite.ipynb)

Build an Interactive Map with Folium

- **Map objects**

  1. **Markers**

  - Represent specific points of interest on the map, such as launch sites, cities, or landmarks.

  2. **Circles**

  - Highlight areas around a point to represent proximity or zones of influence.

  3. **Lines**

  - Show connections or distances between locations.

- GitHub URL (https://github.com/wychia5201314/Coursera-IBM-Applied-Data-Science-Capstone/blob/main/Folium/lab_jupyter_launch_site_location.ipynb)

Build a Dashboard with Plotly Dash

This dashboard provides interactive visualizations of SpaceX's launch records, allowing us to explore success rates and payload characteristics based on launch sites and payload mass.

1. **Dropdown menu for launch site selection**
   - Users can select a specific launch site or view data for all sites.
   - To allow focused analysis on specific sites or a broader comparison across all sites.
   - Updates the success pie chart and scatter plot dynamically.

Build a Dashboard with Plotly Dash

2. Success pie chart

- Displays for all sites (total successful launches per site) and for a specific site (breakdown of successful and failed launches).
- To provide an overview of success rates and distribution for all sites or detailed outcomes for individual sites.
- Helps identify sites with higher launch success rates, providing insights into operational performance.

Build a Dashboard with Plotly Dash

3. Payload mass range slider

- Users can adjust the range of payload mass using a slider, with values ranging from 0 kg to 10,000 kg in increments of 1,000 kg.
- To filter launches based on payload mass and analyze its impact on success rates.
- Payload mass is a critical factor in launch outcomes, and this interaction allows us to explore its correlation with success rates.

Build a Dashboard with Plotly Dash

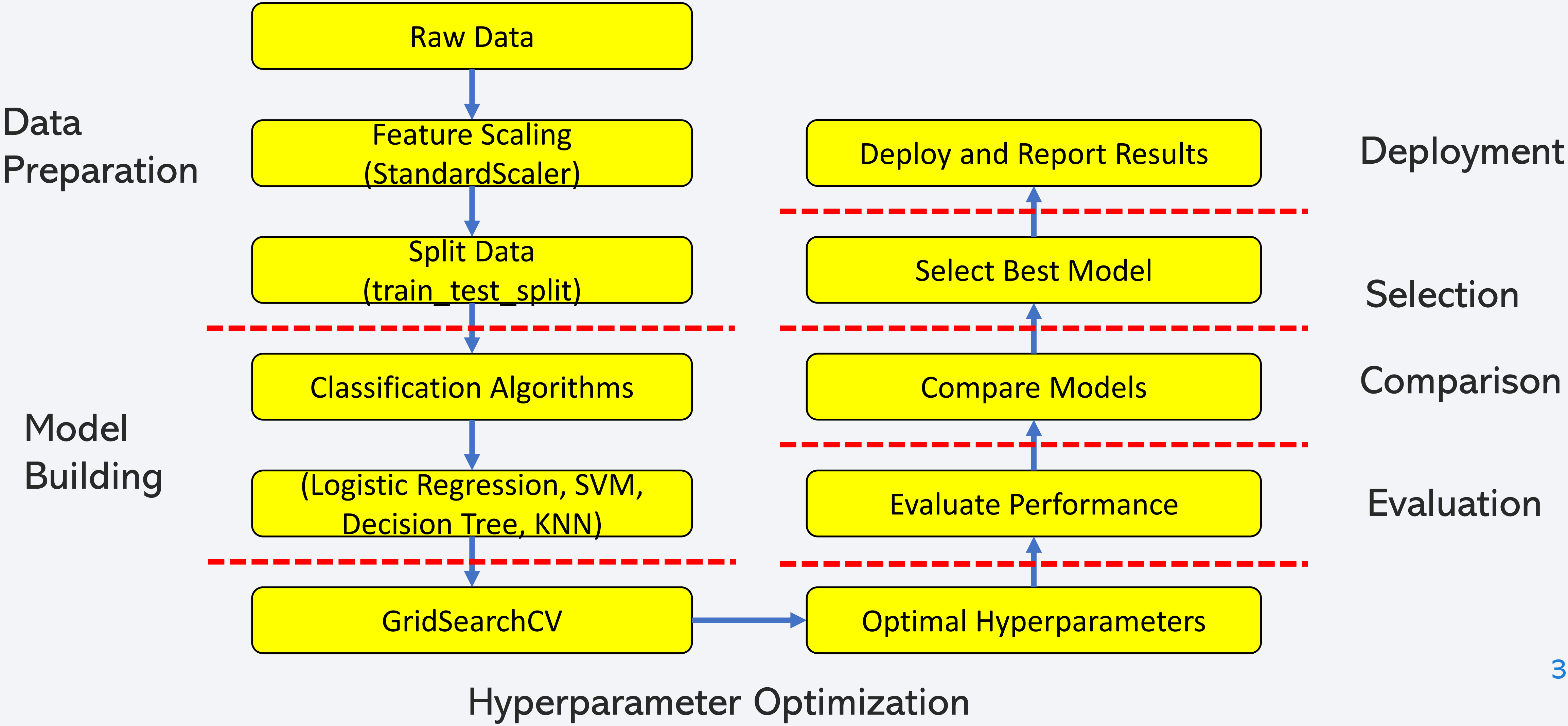
4. Success-payload scatter plot

- Displays payload mass on the x-axis and launch outcome on the y-axis.
- Differentiates data point by the booster version category using color.
- Dynamically updates based on selected launch site and payload mass range set by the slider.
- To explore the relationship between payload mass and launch success for all sites or a specific site.
- To understand whether certain payload masses or booster versions correlate with better outcomes.
- Provides actionable insights into technical performance and booster capabilities.

Build a Dashboard with Plotly Dash

- GitHub URL (https://github.com/wychia5201314/Coursera-IBM-Applied-Data-Science-Capstone/tree/main/Plotly-Dash)

Predictive Analysis (Classification)



Raw Data
Data Preparation
Feature Scaling (StandardScaler)
Split Data (train_test_split)
Classification Algorithms
Model Building
(Logistic Regression, SVM, Decision Tree, KNN)
GridSearchCV
Optimal Hyperparameters
Hyperparameter Optimization
Evaluate Performance
Evaluation
Compare Models
Comparison
Select Best Model
Selection
Deploy and Report Results
Deployment

Predictive Analysis (Classification)

1. **Data Preparation**
   - Standardized the features using *StandardScalder* to ensure all variables are on the same scale.
2. **Model Building**
   - Created classification models using Logistic Regression, Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors (KNN).
3. **Hyperparameter Optimization**
   - Used *GridSearchCV* to perform cross-validated hyperparameter tuning with *cv=10*.
4. **Evaluation**
   - Evaluated models using *score* method to compute train and test accuracies.
5. **Comparison**
   - Compared test accuracies across models and ensured no overfitting by checking train-test accuracy gaps.
6. **Selection of Best Model**
   - Selected the best-performing model based on test accuracy and balanced performance between train and test accuracies.

Predictive Analysis (Classification)

- GitHub URL (https://github.com/wychia5201314/Coursera-IBM-Applied-Data-Science-Capstone/blob/main/Predictive-Analysis/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb)

Results

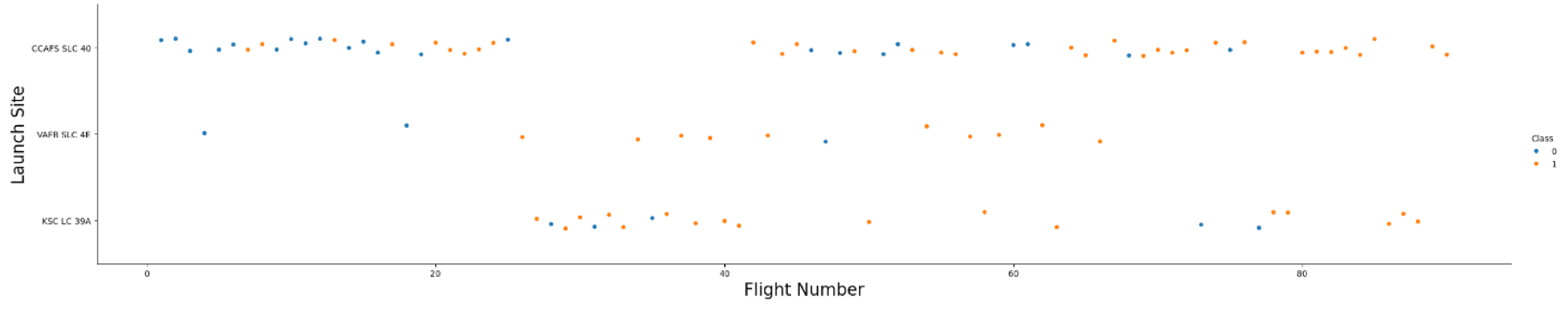
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

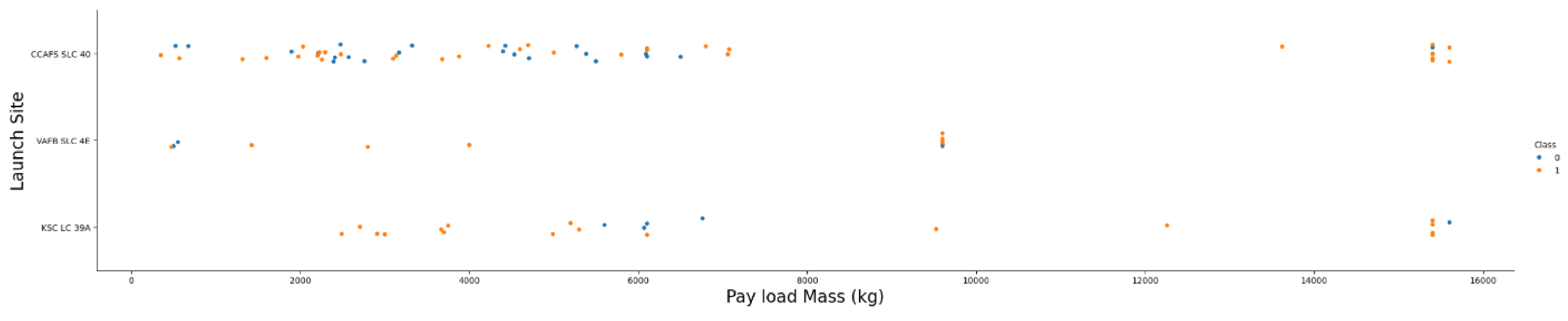
- Scatter plot of Flight Number vs. Launch Site



- For CCAFS SLC-40 launch site, it has the highest flight number while VAFB SLC-4E launch site has the least flight number.
- The success rate is getting higher when flight number is increasing, especially flight number above 80 is 100% success observed from scatter plot.

Payload vs. Launch Site

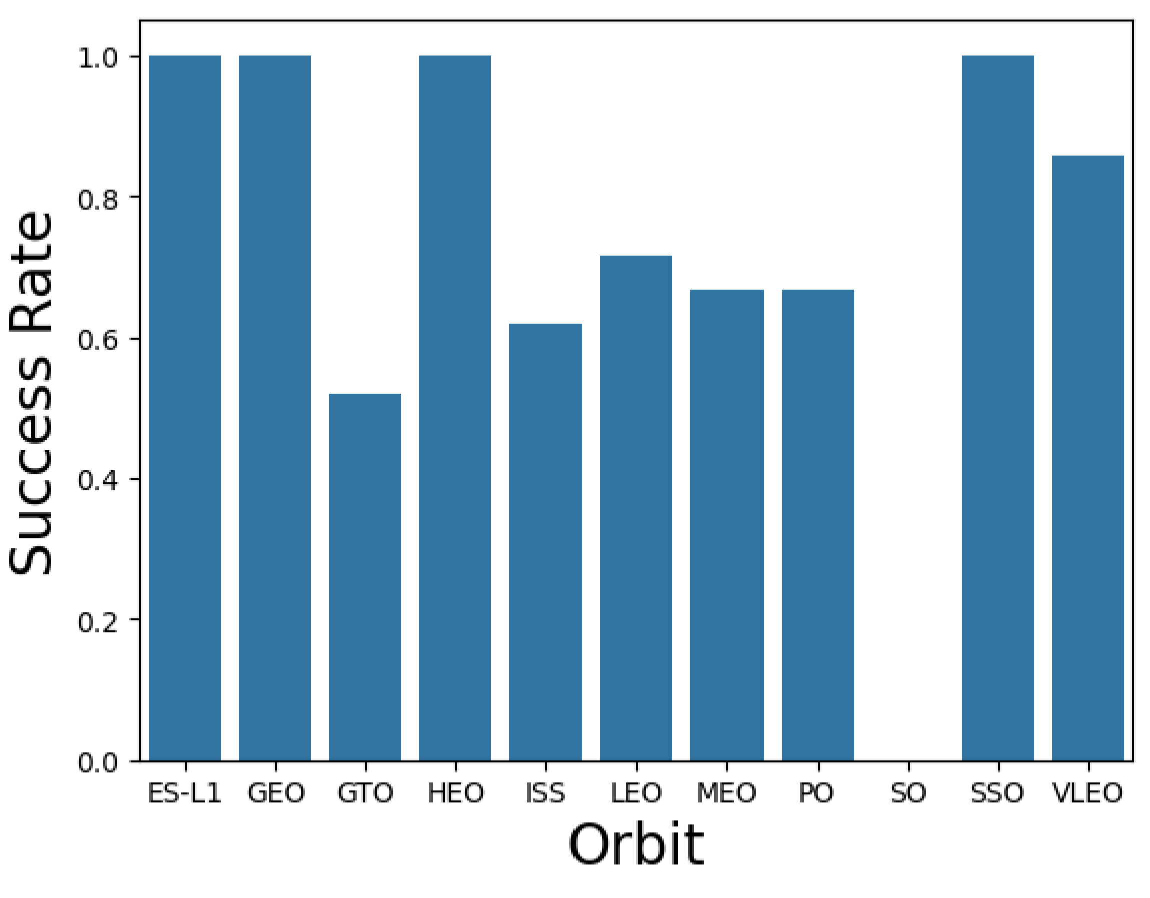
- Scatter plot of Payload vs. Launch Site



- For the VAFB SLC-4E launch site there are no rockets launched for heavy payload mass (greater than 10,000kg).

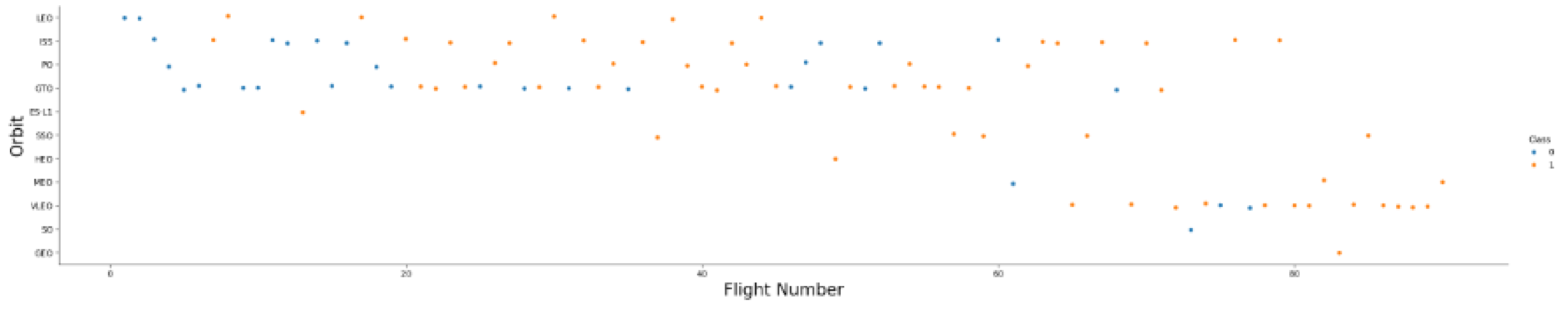
Success Rate vs. Orbit Type

- The orbits have the highest success rate are ES-L1, GEO, HEO and SSO.
- SO orbit has no success rate achieved so far.



Flight Number vs. Orbit Type

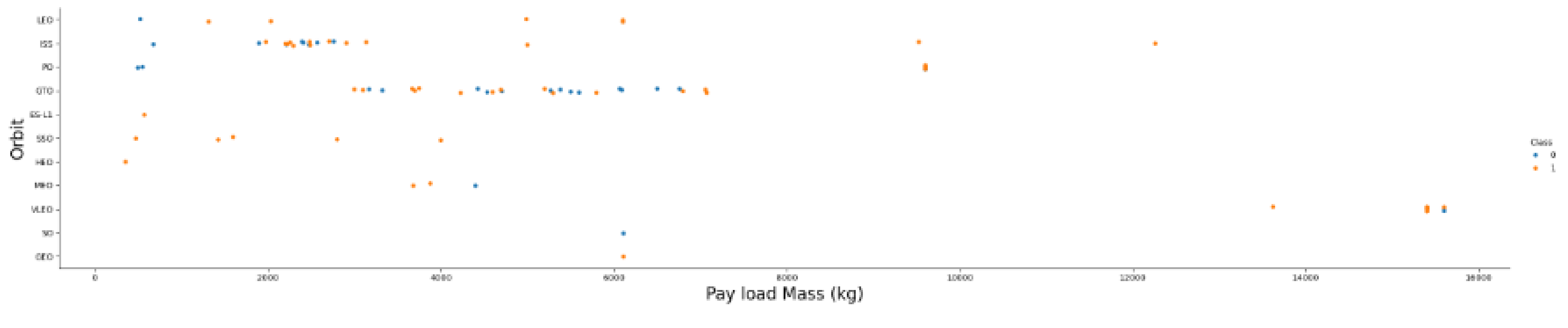
- Scatter plot of Flight Number vs. Orbit Type



- The LEO orbit, success seems to be related to the number of flights. Conversely, in the GTO orbit, there appears to be no relationship between flight number and success.

Payload vs. Orbit Type

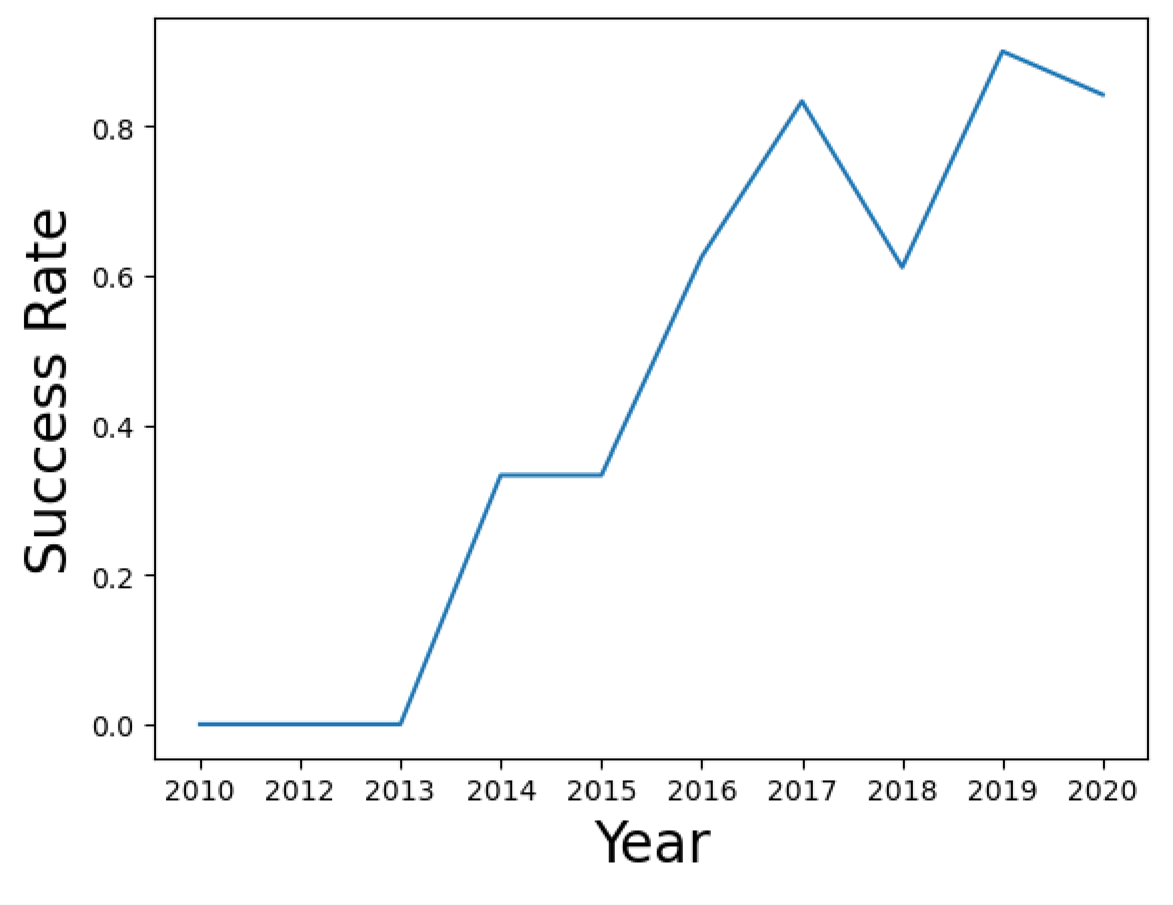
- Scatter plot of Payload vs. Orbit Type



- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.
- However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.

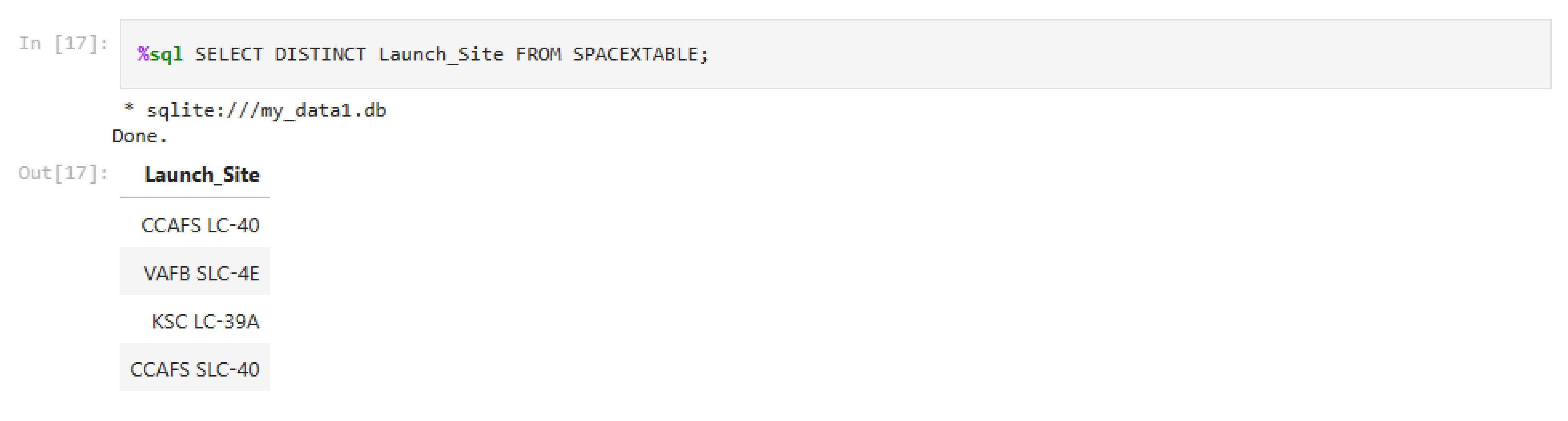
Launch Success Yearly Trend

- The success rate since year 2013 kept increasing until year 2020.



All Launch Site Names

- The names of the unique launch sites as shown below:



```
In [17]: %sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE;
```

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

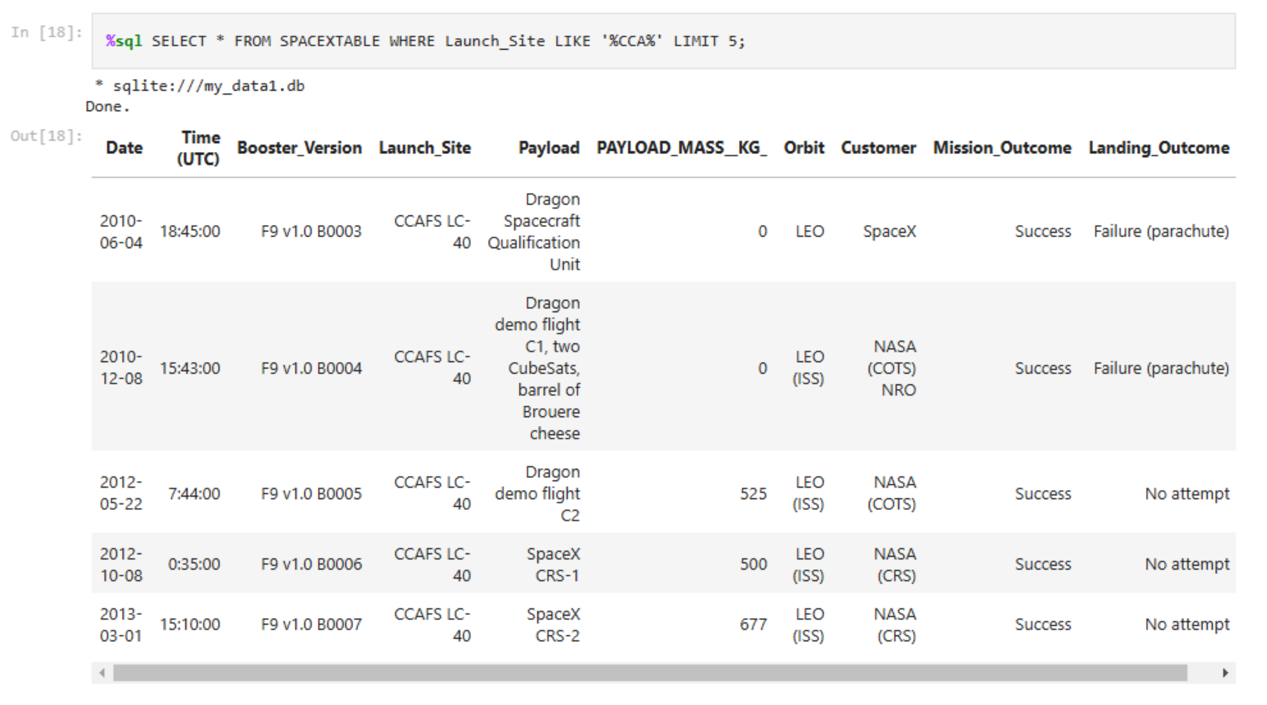
Out[17]:

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

- There are total 4 launch sites.

Launch Site Names Begin with 'CCA'

- The 5 records where launch sites begin with `CCA` as shown below:



```
In [18]: %sql SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE '%CCA%' LIMIT 5;

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

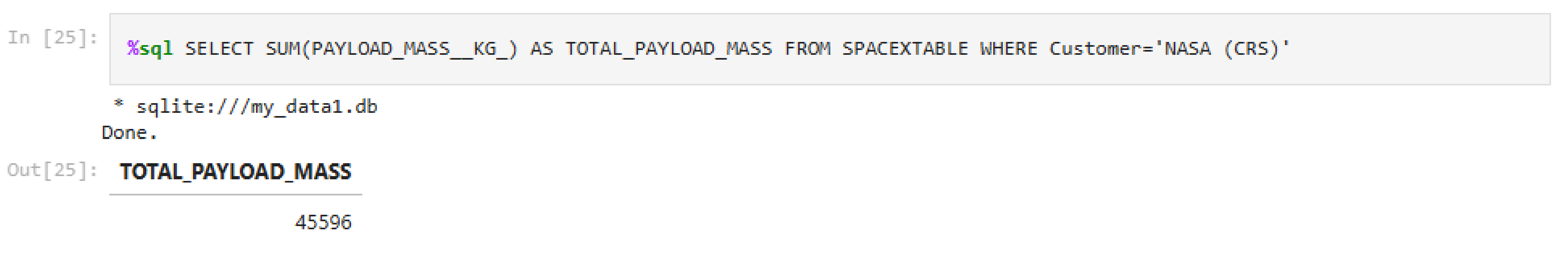
Out[18]:

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

- The records above shown is from CCAFS LC-40 launch site.

Total Payload Mass

- The total payload carried by boosters from NASA as shown below:

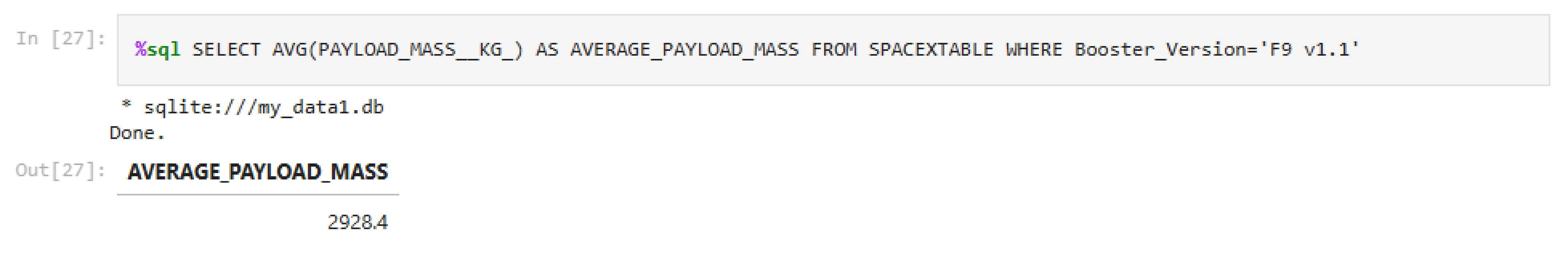


```
In [25]:  %sql SELECT SUM(PAYLOAD_MASS__KG_) AS TOTAL_PAYLOAD_MASS FROM SPACEXTABLE WHERE Customer='NASA (CRS)'

           * sqlite:///my_data1.db
          Done.

Out[25]:  TOTAL_PAYLOAD_MASS

                       45596
```

- The total payload mass carried is 45,596kg.

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 as shown below:

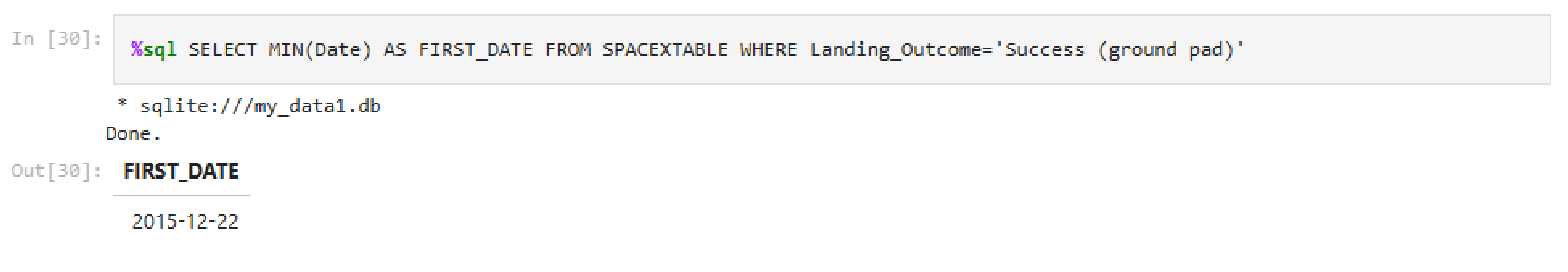


```
In [27]:  %sql SELECT AVG(PAYLOAD_MASS__KG_) AS AVERAGE_PAYLOAD_MASS FROM SPACEXTABLE WHERE Booster_Version='F9 v1.1'

 * sqlite:///my_data1.db
Done.

Out[27]:  AVERAGE_PAYLOAD_MASS
                        2928.4
```

- The average payload mass carried is 2,928.40kg.

First Successful Ground Landing Date

- The dates of the first successful landing outcome on ground pad as shown below:



```
In [30]: %sql SELECT MIN(Date) AS FIRST_DATE FROM SPACEXTABLE WHERE Landing_Outcome='Success (ground pad)'

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

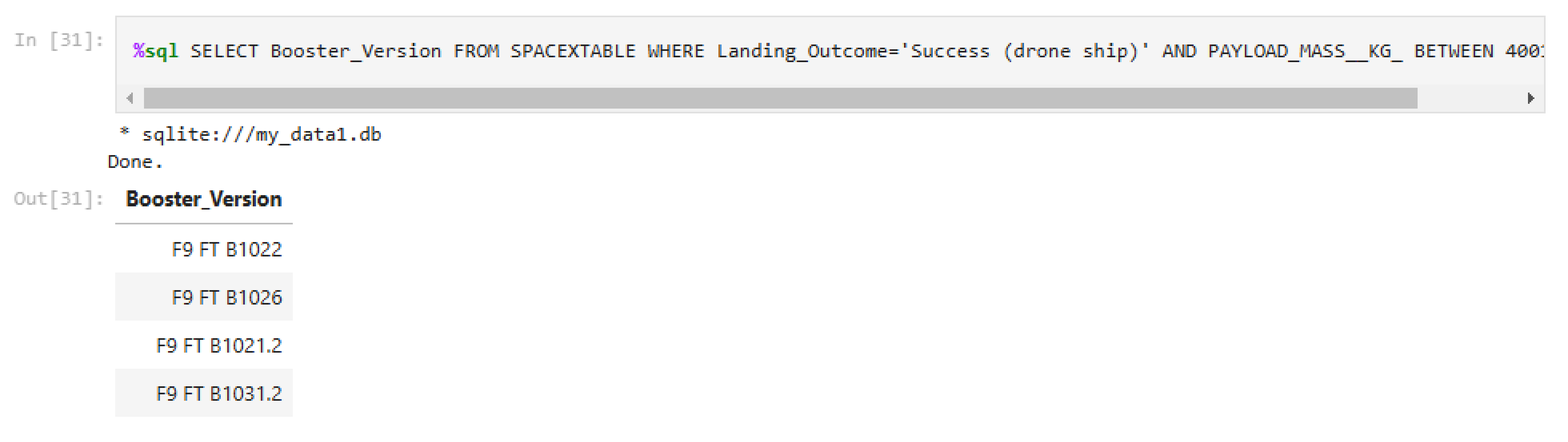
Out[30]:

FIRST_DATE
2015-12-22

- The first successful landing date was fall on 22nd December 2015, just few days before celebrating Christmas day.

Successful Drone Ship Landing with Payload between 4000 and 6000

- The names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 as shown below:



```
In [31]: %sql SELECT Booster_Version FROM SPACEXTABLE WHERE Landing_Outcome='Success (drone ship)' AND PAYLOAD_MASS__KG_ BETWEEN 400

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

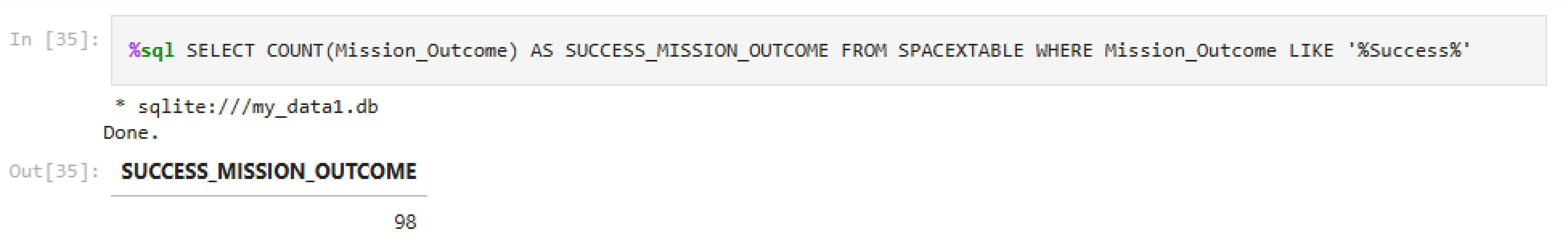
Out[31]:

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

- There are total 4 boosters.

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes as shown below:

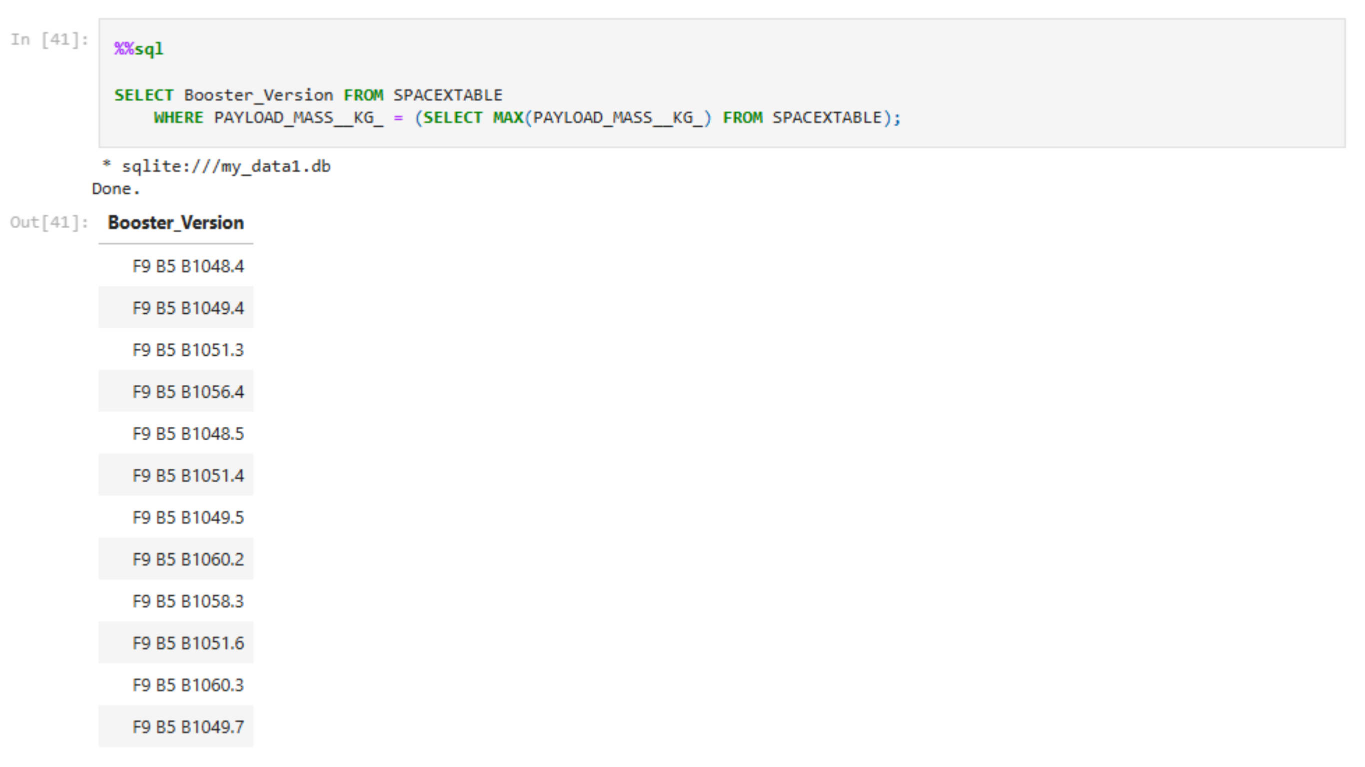


```
In [35]:  %sql SELECT COUNT(Mission_Outcome) AS SUCCESS_MISSION_OUTCOME FROM SPACEXTABLE WHERE Mission_Outcome LIKE '%Success%'

           * sqlite:///my_data1.db
          Done.

Out[35]:  SUCCESS_MISSION_OUTCOME
                                98
```

- There are total 98 missions carried out.

Boosters Carried Maximum Payload

- The names of the booster which have carried the maximum payload mass as shown below:



```
In [41]: %%sql

         SELECT Booster_Version FROM SPACEXTABLE
             WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE);

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

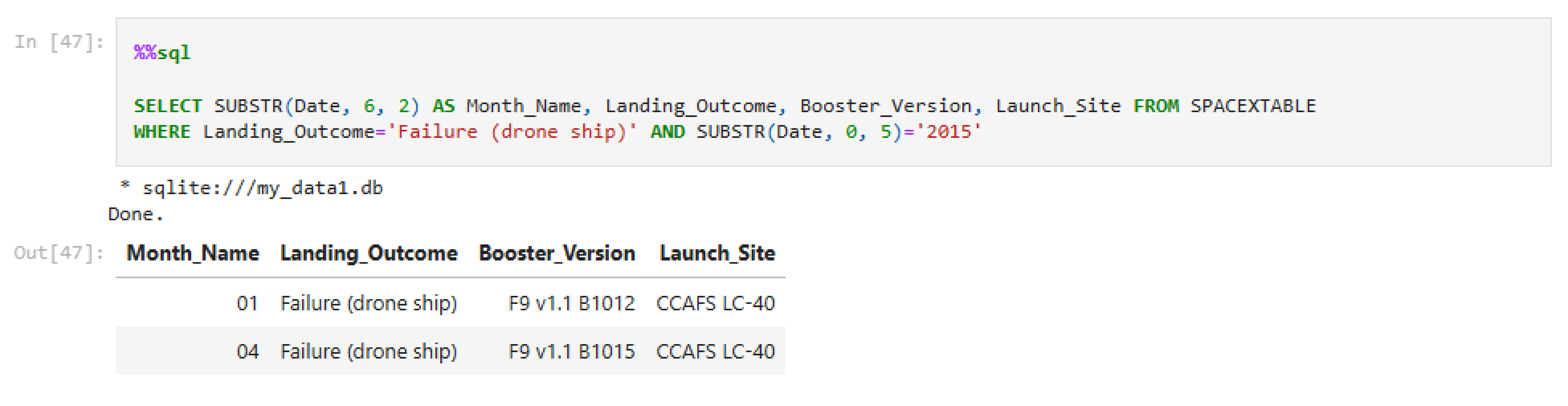
Out[41]:

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

- There are total 12 boosters which have carried the maximum payload mass.

2015 Launch Records

- The failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015 as shown below:



```
%%sql

SELECT SUBSTR(Date, 6, 2) AS Month_Name, Landing_Outcome, Booster_Version, Launch_Site FROM SPACEXTABLE
WHERE Landing_Outcome='Failure (drone ship)' AND SUBSTR(Date, 0, 5)='2015'
```

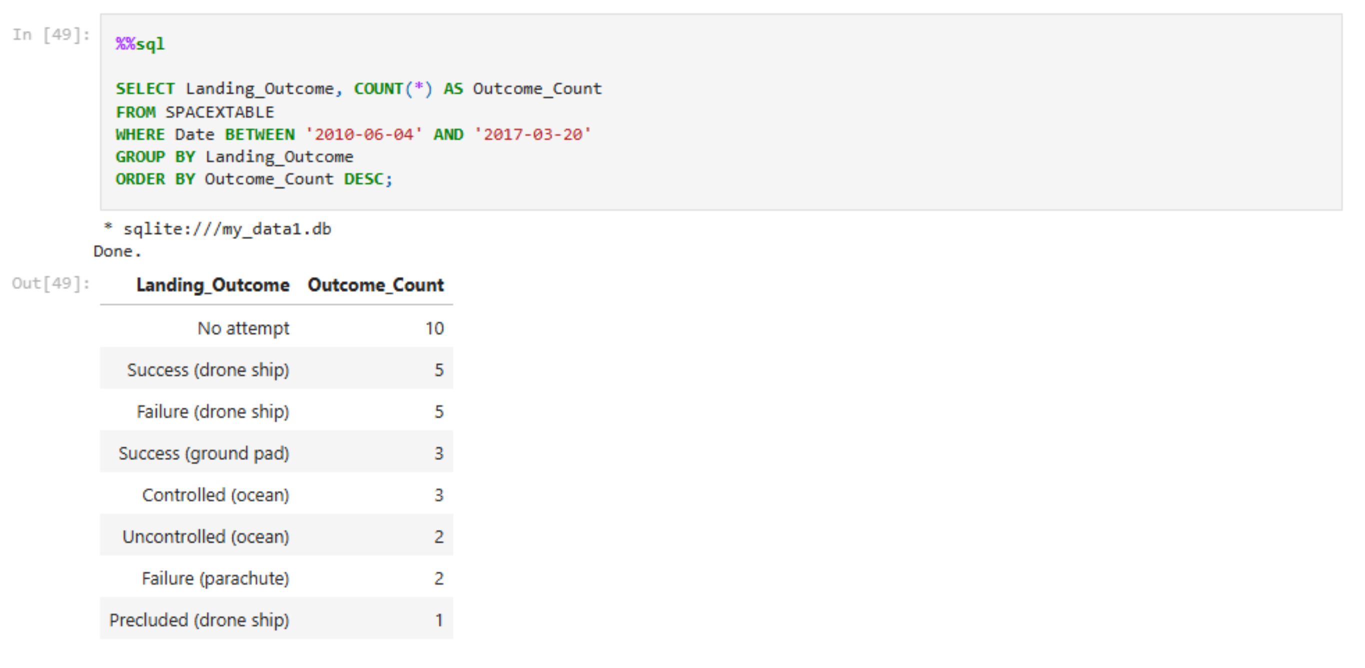
```
 * sqlite:///my_data1.db
Done.
```

Month_Name	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

- On January and April of 2015, there was 1 failure incident happened on each respective month which occurred at CCAFS LC-40 launch site.

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

- 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 as shown below:



```
%%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.
```

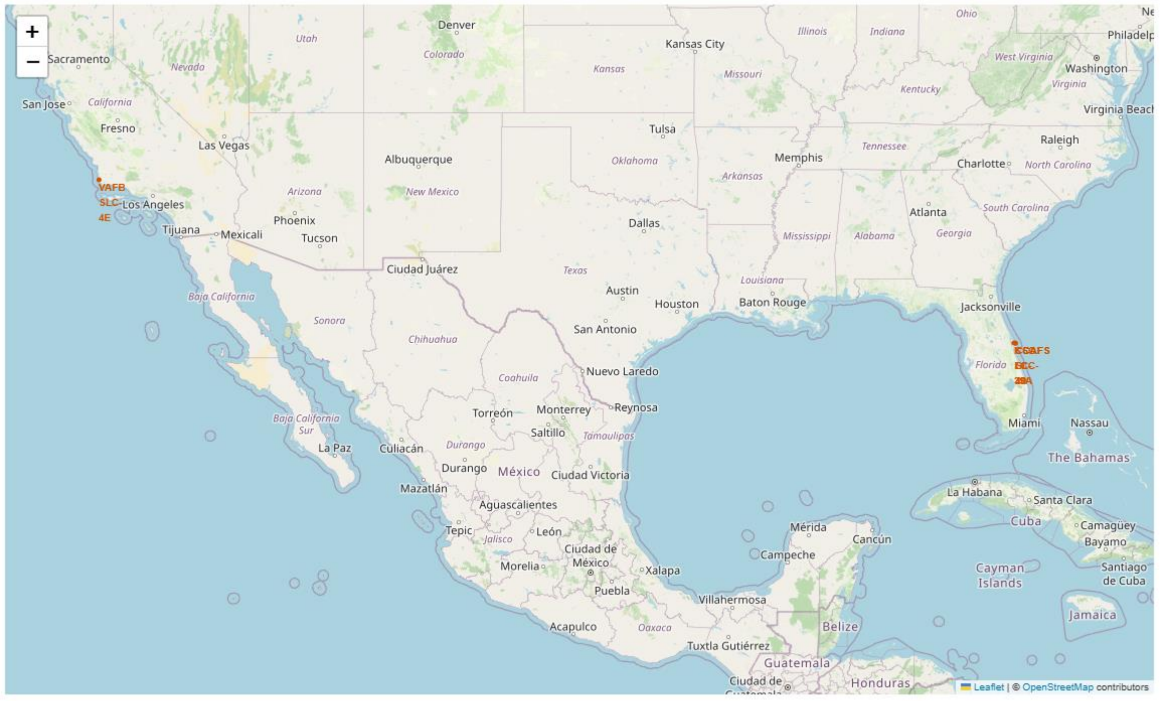
Landing_Outcome	Outcome_Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

- From the observation above, landing on ground pad having higher success rate while using parachute to land is not ideal.

Section 3

Launch Sites Proximities Analysis

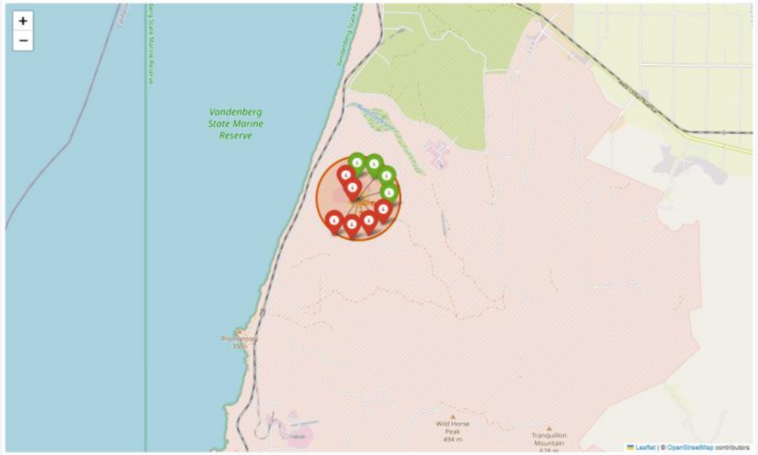
Launch Sites Location



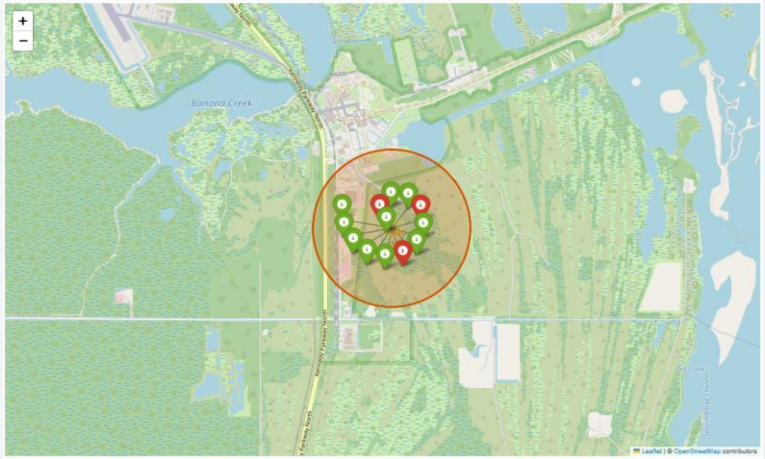
- All launch sites is located in close proximity to coastline.

Launch Outcomes for All Sites

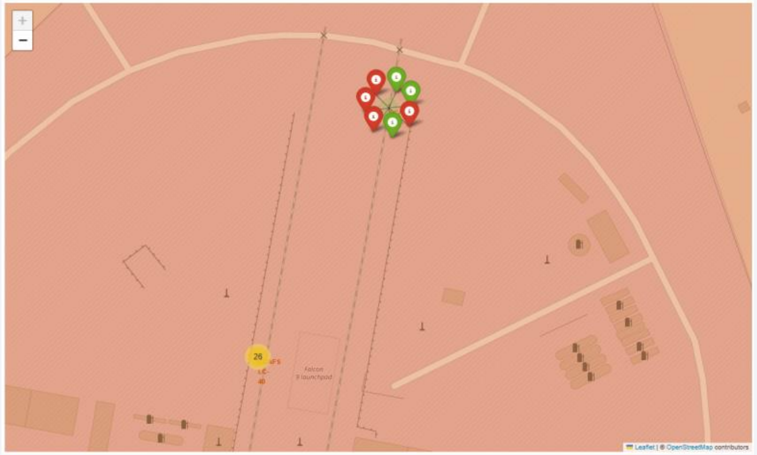
VAFB SLC-4E (40%)


Vandenberg State Marine Reserve
Leaflet | © OpenStreetMap contributors

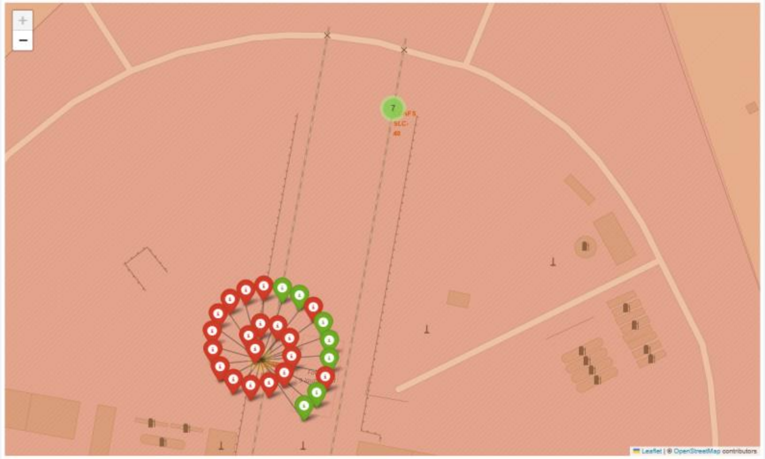
KSC LC-39A (77%)



CCAFS LC-40 (43%)



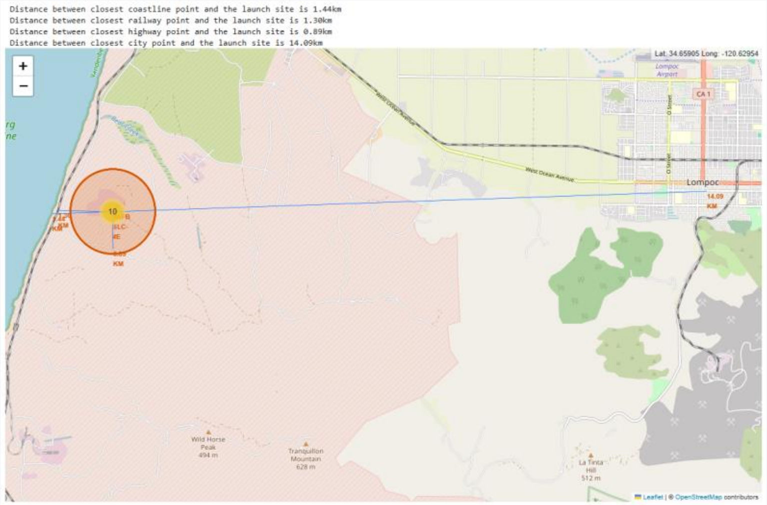
CCAFS SLC-40 (27%)



- KSL LC-39A launch site has the highest success rate among all the launch sites.
- CCAFS SLC-40 launch site has the lowest success rate among all the launch sites.

Launch Sites to Proximities

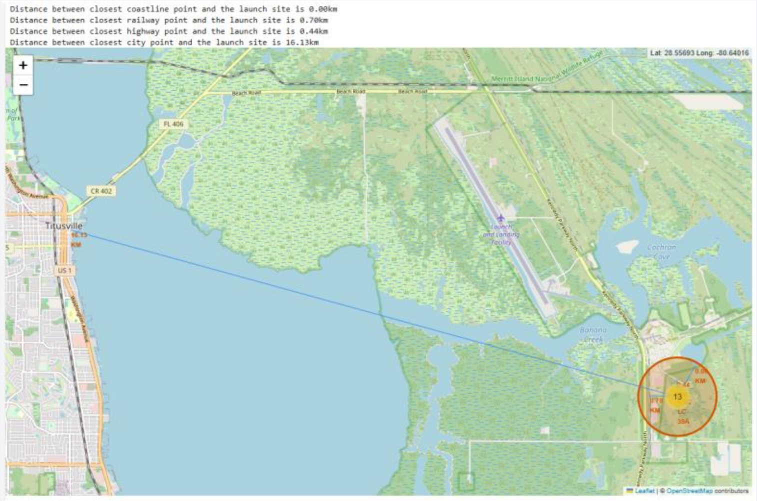
VAFB SLC-4E



```
Distance between closest coastline point and the launch site is 1.44km
Distance between closest railway point and the launch site is 1.30km
Distance between closest highway point and the launch site is 0.89km
Distance between closest city point and the launch site is 14.09km
```

- Coastline = 1.44 km
- Railway = 1.30 km
- Highway = 0.89 km
- City = 14.09 km

KSC LC-39A

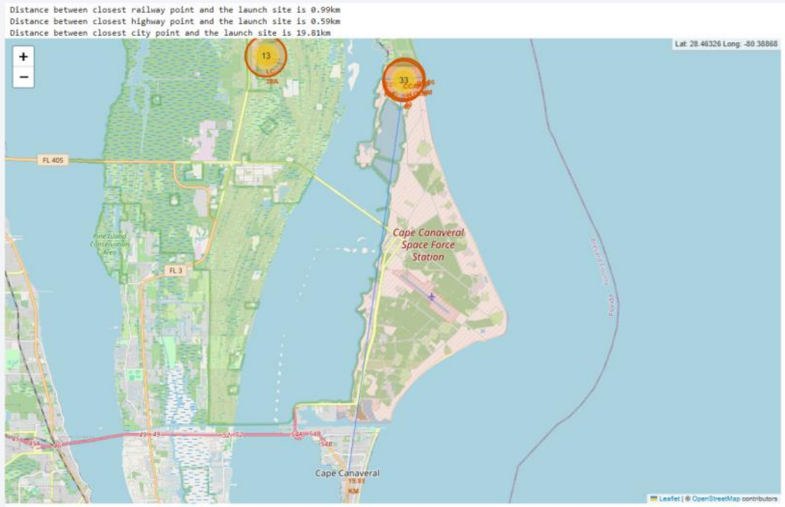


```
Distance between closest coastline point and the launch site is 0.00km
Distance between closest railway point and the launch site is 0.70km
Distance between closest highway point and the launch site is 0.44km
Distance between closest city point and the launch site is 16.13km
```

- Coastline = 0.00 km
- Railway = 0.70 km
- Highway = 0.44 km
- City = 16.31 km

CCAFS SLC-40

CCAFS LC-40



```
Distance between closest railway point and the launch site is 0.99km
Distance between closest highway point and the launch site is 0.59km
Distance between closest city point and the launch site is 19.81km
```

Lat: 28.46326 Long: -80.38868
13
33
Cape Canaveral Space Force Station
Cape Canaveral

- Coastline = 0.86 km
- Railway = 0.99 km
- Highway = 0.59 km
- City = 19.81 km

Launch Sites Location

1. Proximity to Railways and Highways

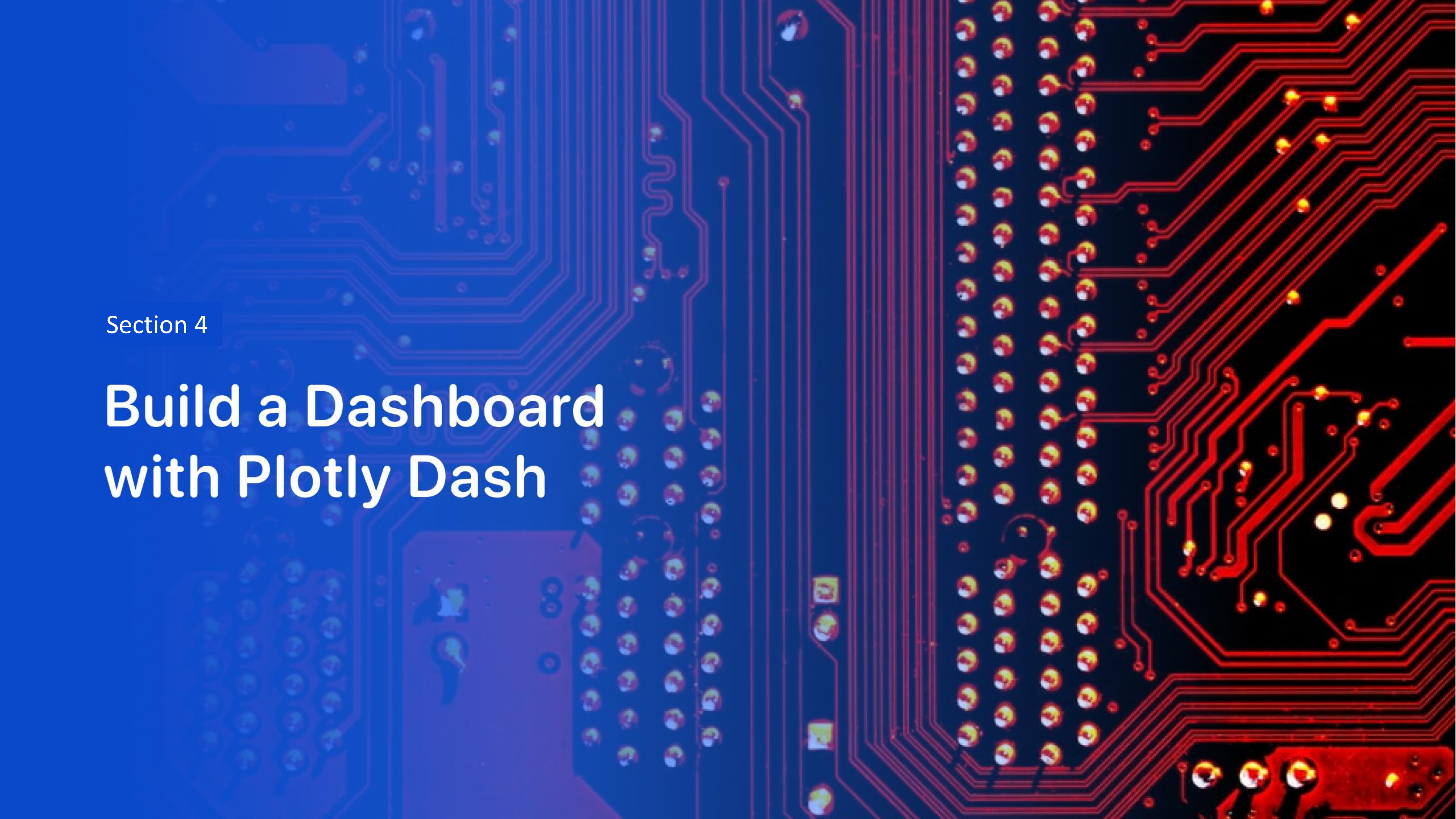
   - Launch sites leverage infrastructure for transportation needs. Proximity to these facilities reduces logistical costs and improves operational efficiency.

2. Proximity to Coastline

   - Safety concerns and environmental considerations influence this. Being near water ensures that failed launches or falling debris pose minimal risk to human settlements.

3. Distance from Cities

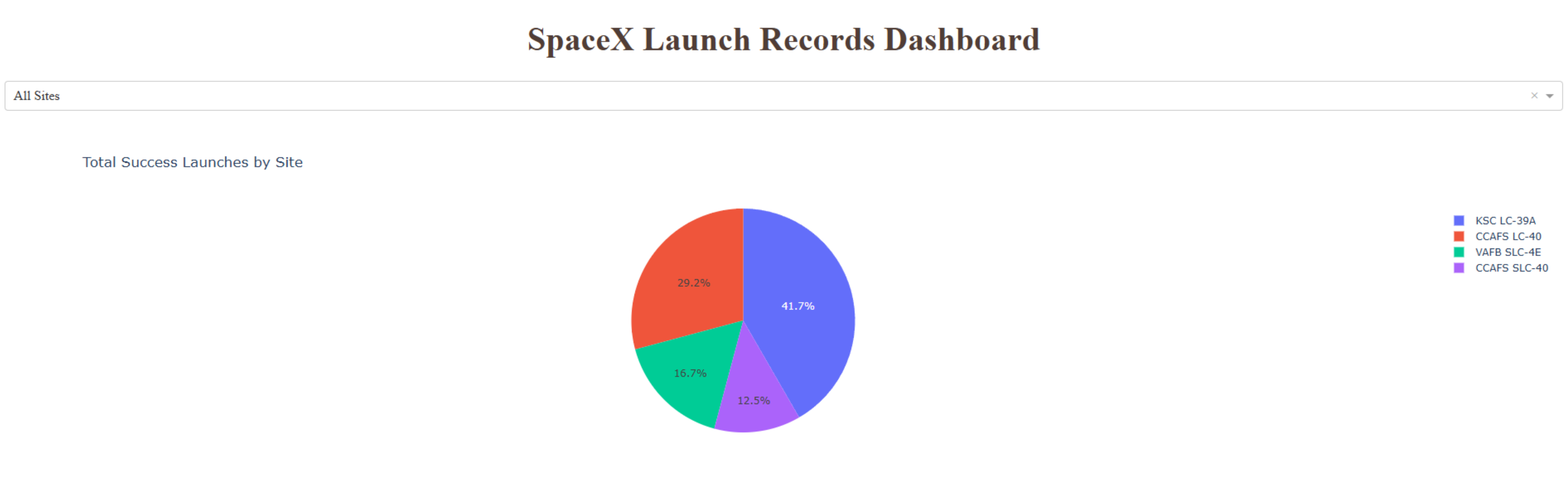
   - Risk mitigation drives the placement of launch sites. Distances are often determined by national safety regulations and international guidelines.



Section 4

Build a Dashboard with Plotly Dash

Total Success Launches by Site Dashboard



SpaceX Launch Records Dashboard

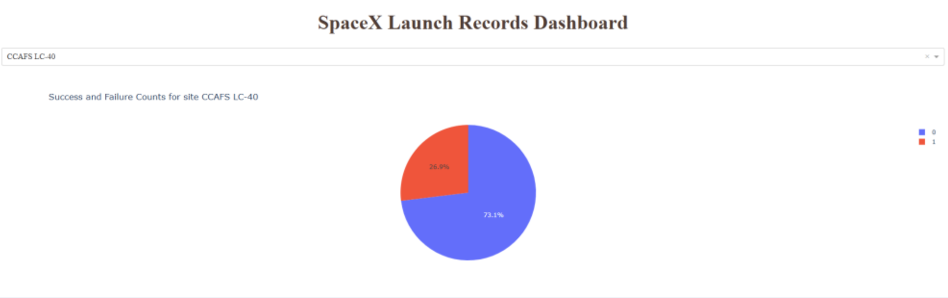
All Sites

Total Success Launches by Site

- From the above pie chart observation, it shows what KSC LC-39A has the largest successful launches (41.7%).

Launch Site Success Rate Dashboard

CCAFS LC-40 (73.1%)

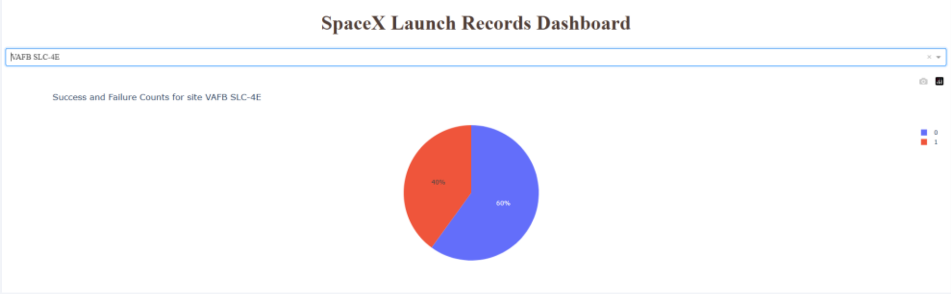


SpaceX Launch Records Dashboard

CCAFS LC-40

Success and Failure Counts for site CCAFS LC-40

VAFB SLC-4E (60%)

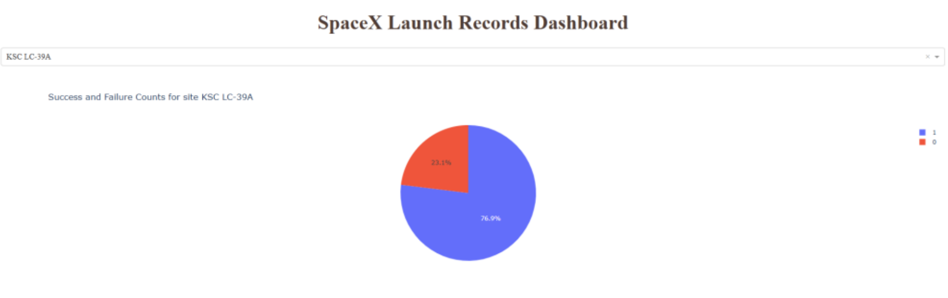


SpaceX Launch Records Dashboard

VAFB SLC-4E

Success and Failure Counts for site VAFB SLC-4E

KSC LC-39A (76.9%)

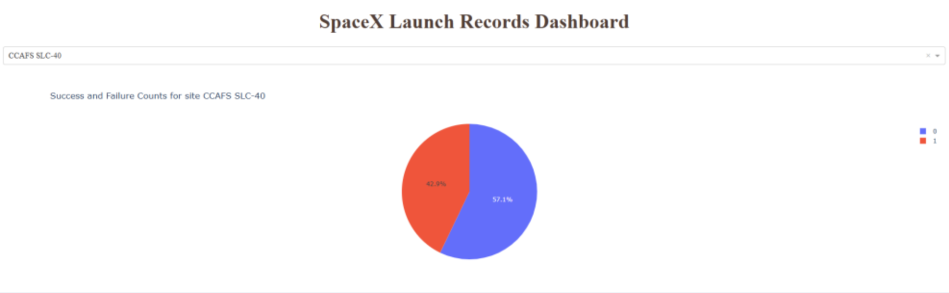


SpaceX Launch Records Dashboard

KSC LC-39A

Success and Failure Counts for site KSC LC-39A

CCAFS SLC-40 (76.9%)



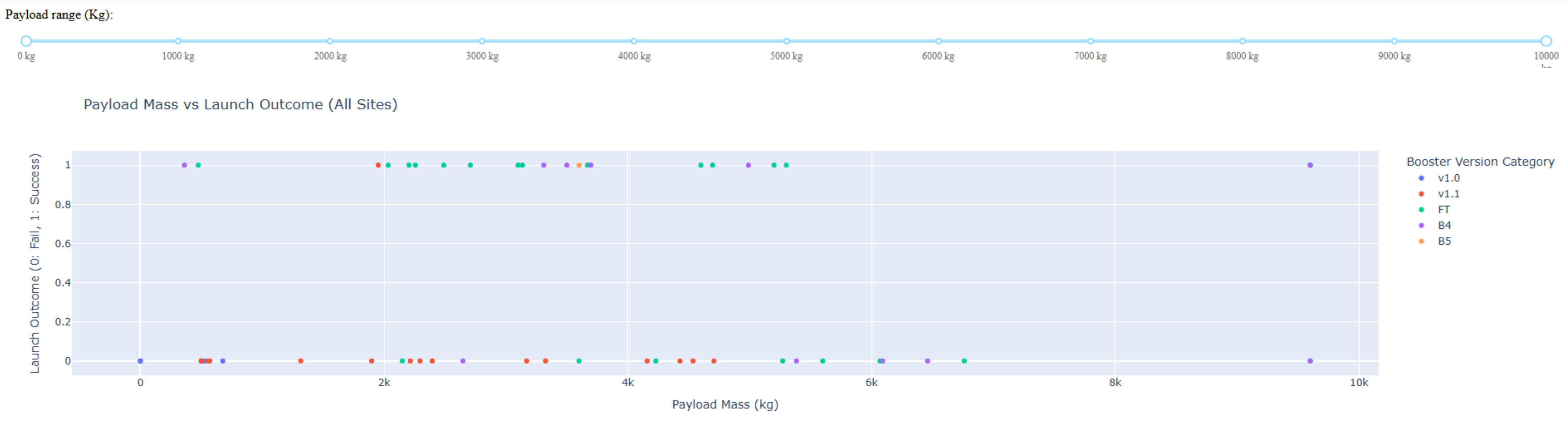
SpaceX Launch Records Dashboard

CCAFS SLC-40

Success and Failure Counts for site CCAFS SLC-40

- From the above pie chart observation, KSC LC-39A has highest launch success rate (76.9%) compared to other launch sites.

Payload vs Launch Outcome Scatter Plot (All Sites)

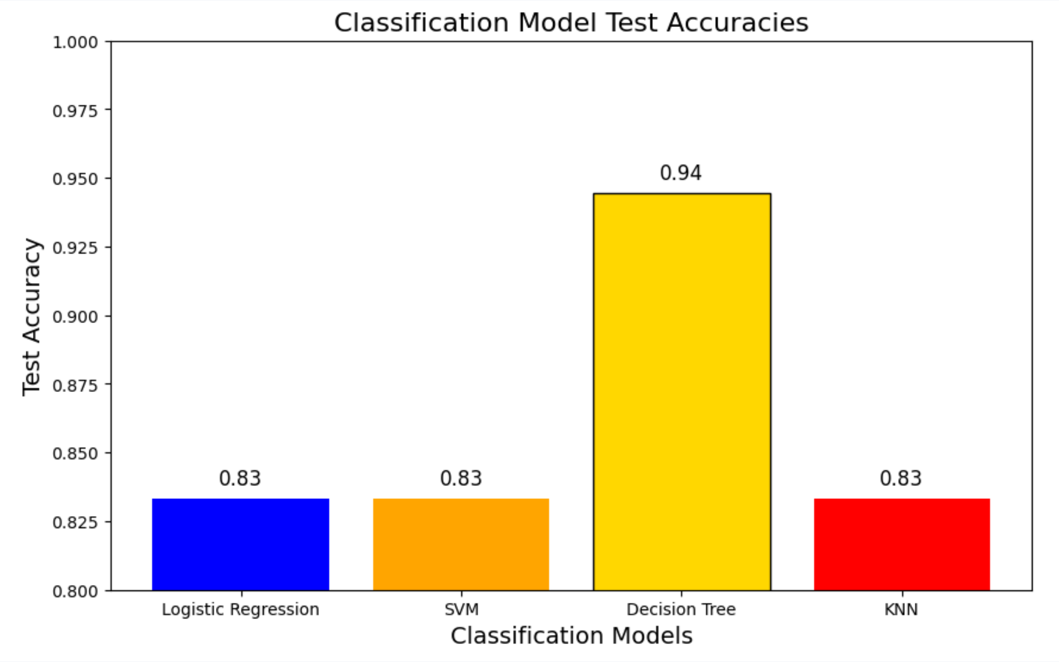


- Payload range fall between 3,000kg and 4,000kg has the highest launch success rate.
- Booster version FT has the highest launch success rate.

Section 5

Predictive Analysis (Classification)

Classification Accuracy



- The best model is Decision Tree with an accuracy of 0.94.

Confusion Matrix of Decision Tree Model

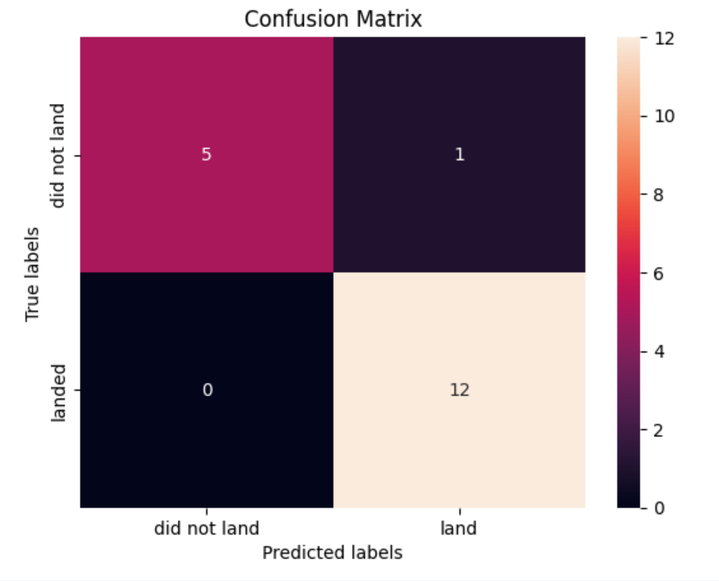
- From the confusion matrix observation, it shows the true labels vs predicted labels as per below:

  1. True positive = 12
  2. True negative = 5
  3. False positive = 1
  4. False negative = 0

Accuracy = 94.44%

Precision = 92.31%

Recall = 100%



- High recall for the “landed” class means the model successfully identifies most of the “landed” cases.
- Minimal false negatives indicate the model is effective in capturing actual “landed” instances.
- There is one false positive where the model incorrectly classified “did not land” as “landed”. This could potentially mislead decision-making.

Conclusions

1. The orbits have the highest success rate are ES-L1, GEO, HEO and SSO.
2. With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.
3. The names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 are F9 FT B1022, F9 FT B1026, F9 FT B1021.2 and F9 FT B1031.2.
4. Landing on ground pad having higher success rate while using parachute to land is not ideal.

Conclusions

5. Launch sites leverage infrastructure for transportation needs. Proximity to railways and highways reduces logistical costs and improves operational efficiency.
6. Safety concerns and environmental considerations influence this. Being near water ensures that failed launches or falling debris pose minimal risk to human settlements.
7. Risk mitigation drives the placement of launch sites. Distances are often determined by national safety regulations and international guidelines.

Conclusions

8. Payload range fall between 3,000kg and 4,000kg has the highest launch success rate.
9. Booster version FT has the highest launch success rate.
10. The decision tree model performs exceptionally well with high accuracy (94.44%) and recall (100%). However, the single false positive suggests a slight room for improvement in precision, potentially by fine-tuning the model further or adjusting the decision threshold.

Thank you!