IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Mohamed Yasin Mohamed Yahaiya
02-25-2025

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

  -Interactive analytics in screenshots

  -Predictive Analytics result

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 This goal of the project is to create a machine learning pipeline to predict if the first stage will land successfully

- Problems you want to find answers

-What factors determine if the rocket will land successfully?

-The interaction amongst various features that determine the success rate of a successful landing.

-What operating conditions needs to be in place to ensure a successful landing program.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected using the SpaceX API and Web scraping from Wikipedia
- Perform data wrangling
  - 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

Data Collection

- The data was collected using various methods

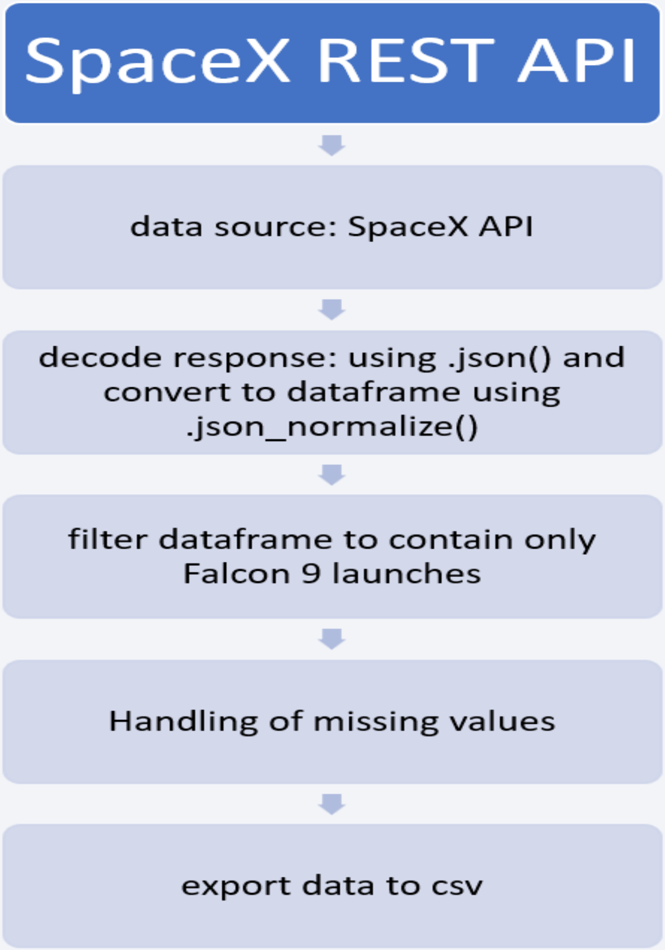
  - Data collection was done using get request to the SpaceX API
  - Next, we decoded the response content as a Json using json() function call and turn it into a pandas dataframe using json_normalize
  - We then cleaned the data, checked for missing values and fill in missing values where necessary
  - In addition, we performed web scraping from Wikipedia for Falcon 9 launch records Beautiful Soup
  - The objective was to extract the launch records as HTML table, parse the table and convert it to a pandas dataframe for future analysis

Data Collection – SpaceX API

- We used the get request to the SpaceX API to collect data, clean the requested data and did some basic data wrangling and formatting.
- Request data from SpaceX API (rocket launch data)
- Decode response using .json() and convert to a dataframe using .json_normalize()
- Request information about the launches from SpaceX API using custom functions
- Create dictionary from the data
- Create dataframe from the dictionary
- Filter dataframe to contain only Falcon 9 launches
- Replace missing values of Payload Mass with calculated .mean()
- Export data to csv file

- The link to the notebook is

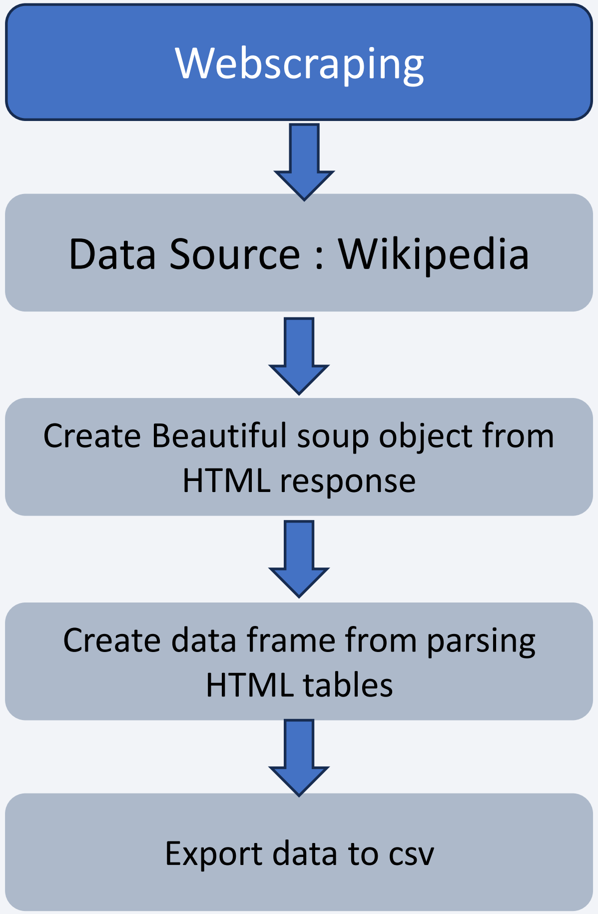
https://github.com/yasinryd/Win-the-Space-race-by-Data-Science/blob/main/jupyter-labs-spacex-data-collection-api-v2_answers.ipynb



Data Collection - Scraping

- Request data (Falcon 9 launch data) from Wikipedia
- Create BeautifulSoup object from HTML response
- Extract column names from HTML table header
- Collect data from parsing HTML tables
- Create dictionary from the data
- Create dataframe from the dictionary
- Export data to csv file

- URL : https://github.com/yasinryd/Win-the-Space-race-by-Data-Science/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- Exploratory data analysis:

Check null values and data types

Calculate number of launches on each site and occurrence of associated orbits

Calculate mission outcome dependent on orbit type

- Create binary training labels as dependent variable:

Create label from outcome column

1 = good outcome

0 = bad outcome

Url : https://github.com/yasinryd/Win-the-Space-race-by-Data-Science/blob/main/labs-jupyter-spacex-Data%20wrangling-v2.ipynb

EDA with Data Visualization

- Charts:

Flight Number vs Payload (scatter category plot)

Flight Number vs Launch Site (box plot)

Payload Mass (kg) vs Launch Site (box plot)

Payload Mass (kg) vs Orbit (scatter category plot)

- Rationale:

Scatter plots to investigate relationship between two numerical variables

Boxplots to investigate trends of discrete categories like launch site

GITHUB URL : https://github.com/yasinryd/Win-the-Space-race-by-Data-Science/blob/main/jupyter-labs-eda-dataviz-v2.ipynb

EDA with SQL

Summary of SQL Queries:

- Display unique launch sites
- Display total payload carried by rockets dependent on launch site
- List date when first successful landing was achieved on ground pad
- List the total number of successful and failure mission outcomes
- Performed queries and subqueries in sqllite with sqlalchemy

- Url : https://github.com/yasinryd/Win-the-Space-race-by-Data-Science/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

Markers:

- Used for characterizing launch sites:
- Circle to denote NASA Johnson Space Center's coordinate as well as other launch sites
- Colored markers to indicate successful (green) and unsuccessful (red) launches at each launch site to show which launch sites have the highest success rate
- Lines: to show the distance between several points of interest e.g. distance to highways, railways, coasts and city

Github : https://github.com/yasinryd/Win-the-Space-race-by-Data-Science/blob/main/lab-jupyter-launch-site-location-v2.ipynb

Build a Dashboard with Plotly Dash

Dropdown List with launch sites

Allow user to select all launch sites or a distinct launch sites

Pie Chart showing successful launches

Allow users to see proportional rate of successful launches

Slider of payload mass range

Allow user to select payload mass range to study the influence of this factor on success rate

Scatter plot showing payload mass vs success rate by booster version

Github : https://github.com/yasinryd/Win-the-Space-race-by-Data-Science/blob/main/ploty%20dash.ipynb

Predictive Analysis (Classification)

Step process

1. Create NumPy array from class column = dependent variable

2. Standardize and transform the data

3. Split the data into test and training set using scikit lean's train test split function

4. Apply Grid Search CV on different machine learning algorithms

    - Logistic regression
    - Support vector machine
    - Decision tree
    - K nearest neighbour

5. Calculate accuracy

Github: https://github.com/yasinryd/Win-the-Space-race-by-Data-Science/blob/main/SpaceX-Machine-Learning-Prediction-Part-5-v1.ipynb

Results

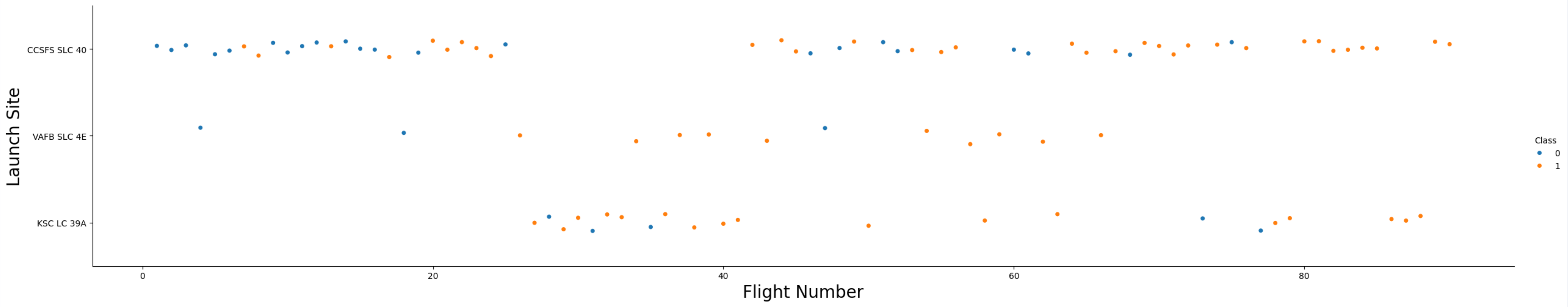
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

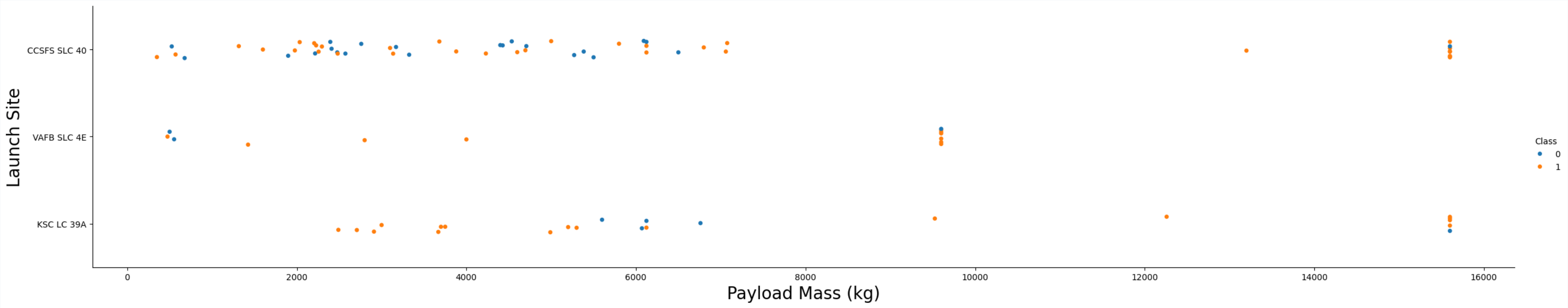
Flight Number vs. Launch Site

- Below is the scatter plot of Flight Number vs. Launch Site



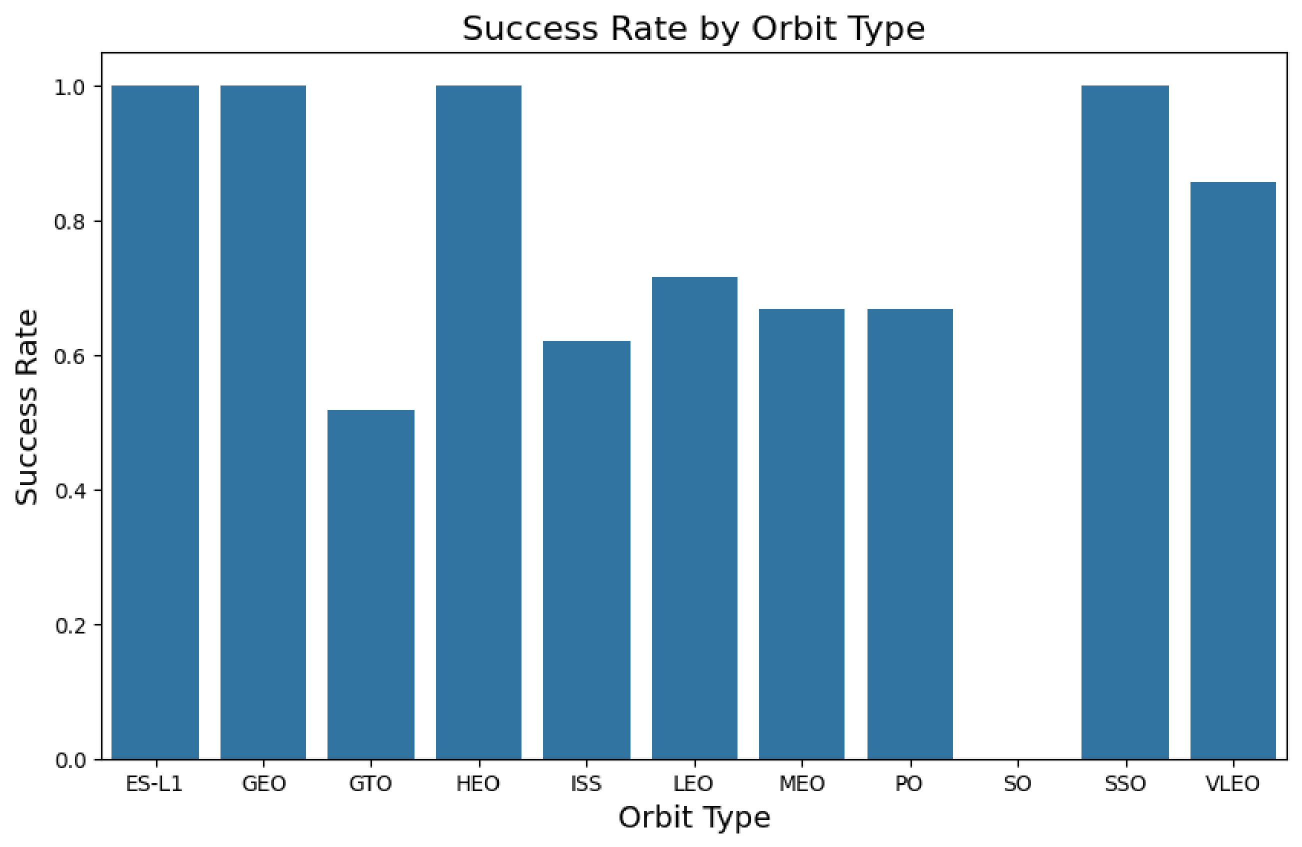
Payload vs. Launch Site

Scatter plot of Payload vs. Launch Site



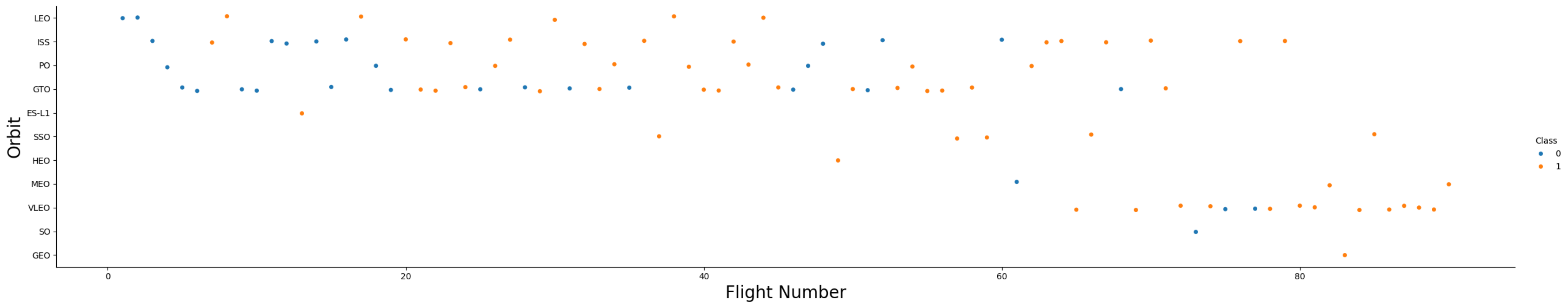
Success Rate vs. Orbit Type

Bar chart for the success rate of each orbit type



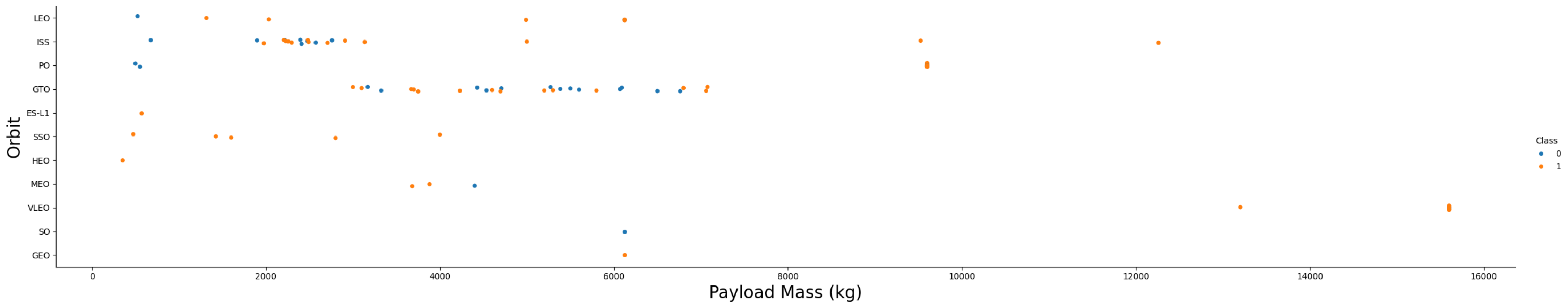
Flight Number vs. Orbit Type

Scatter point of Flight number vs. Orbit type



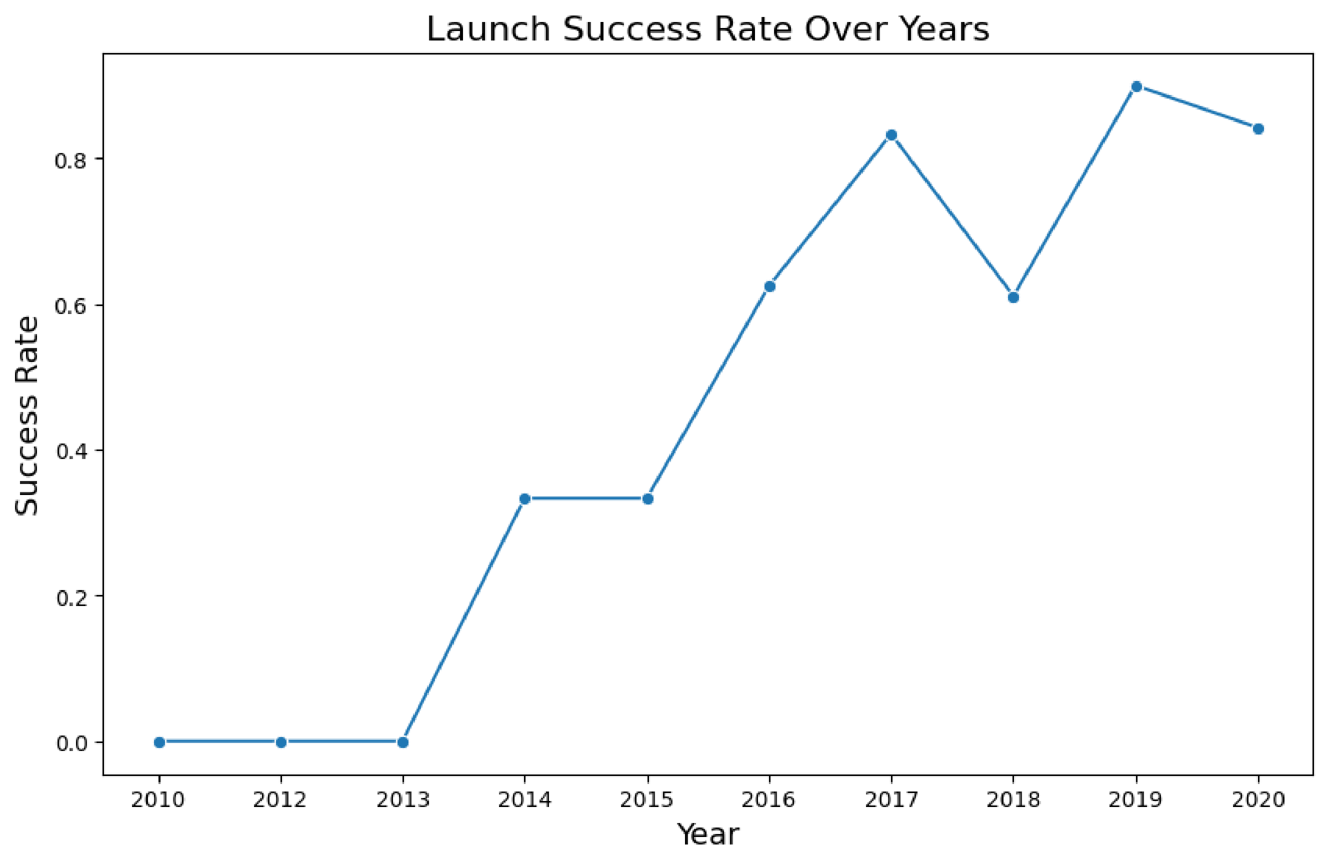
Payload vs. Orbit Type

Scatter point of payload vs. orbit type



Launch Success Yearly Trend

- Show a line chart of yearly average success rate



All Launch Site Names

- Names of the unique launch sites
- SQL query was used to list all unique launch sites
- 4 launch sites were used for Falcon 9 rocket launches as listed on the right

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

Launch Site Names Begin with 'CCA'

5 records where launch sites begin with `CCA`

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

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

- The total carried payload mass is 45,596 kg

```
%%sql
SELECT SUM("Payload_Mass__kg_") AS Total_Payload_Mass
FROM SPACEXTABLE
WHERE "Customer" = 'NASA (CRS)';
```

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

Total_Payload_Mass
45596

Average Payload Mass by F9 v1.1

The average payload mass carried by booster version F9 v1.1 is 2928.4 KG

```
%%sql
SELECT AVG("Payload_Mass__kg_") AS Average_Payload_Mass
FROM SPACEXTABLE
WHERE "Booster_Version" = 'F9 v1.1';
```

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

Average_Payload_Mass
2928.4

First Successful Ground Landing Date

The first successful landing outcome on ground pad was on 2015-12-22

```
%%sql
SELECT MIN("Date") AS First_Successful_Landing_Date
FROM SPACEXTABLE
WHERE "Landing_Outcome" = 'Success (ground pad)';
```

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

First_Successful_Landing_Date
2015-12-22

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 are below

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

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

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

```
%%sql
SELECT "Landing_Outcome", COUNT(*) AS Outcome_Count
FROM SPACEXTABLE
GROUP BY "Landing_Outcome";
```

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

Landing_Outcome	Outcome_Count
Controlled (ocean)	5
Failure	3
Failure (drone ship)	5
Failure (parachute)	2
No attempt	21
No attempt	1
Precluded (drone ship)	1
Success	38
Success (drone ship)	14
Success (ground pad)	9
Uncontrolled (ocean)	2

Boosters Carried Maximum Payload

The names of the booster which have carried the maximum payload mas

```
%%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

2015 Launch Records

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

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

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

Month	Booster_Version	Launch_Site
01	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40

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

```
%%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;
```

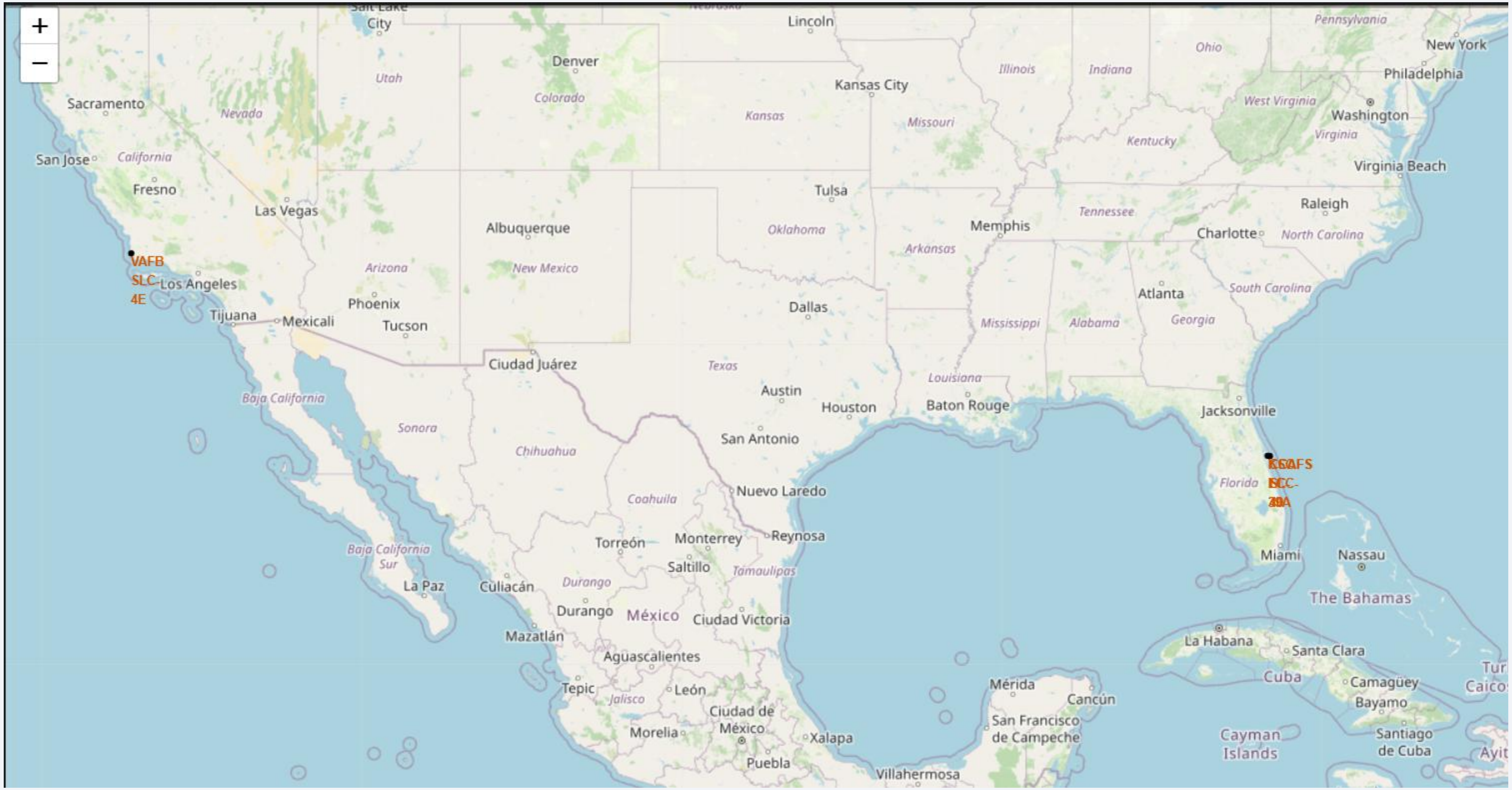
```
 * sqlite:///my_data1.db
Done.
```

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

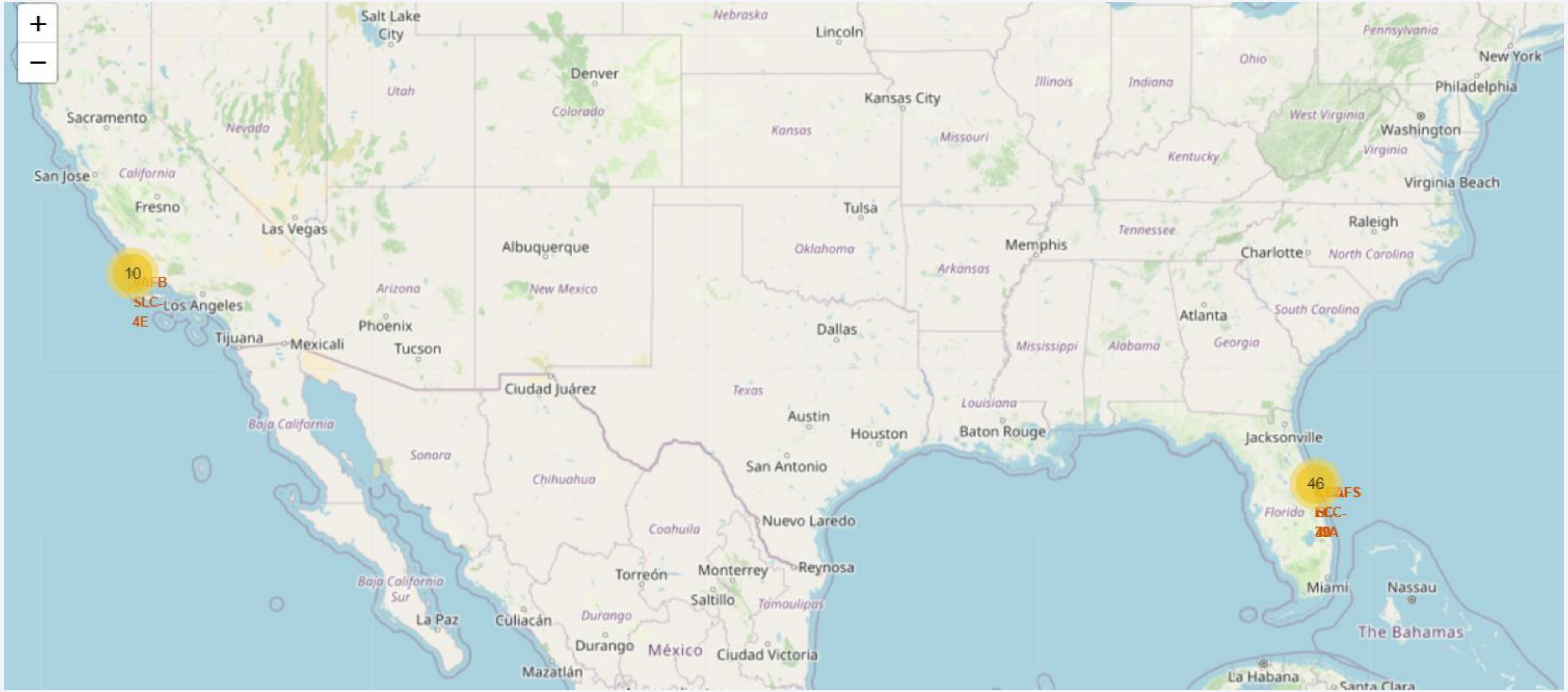
Section 3

Launch Sites Proximities Analysis

Distribution of Launch sites in USA



Number of Launches at each site



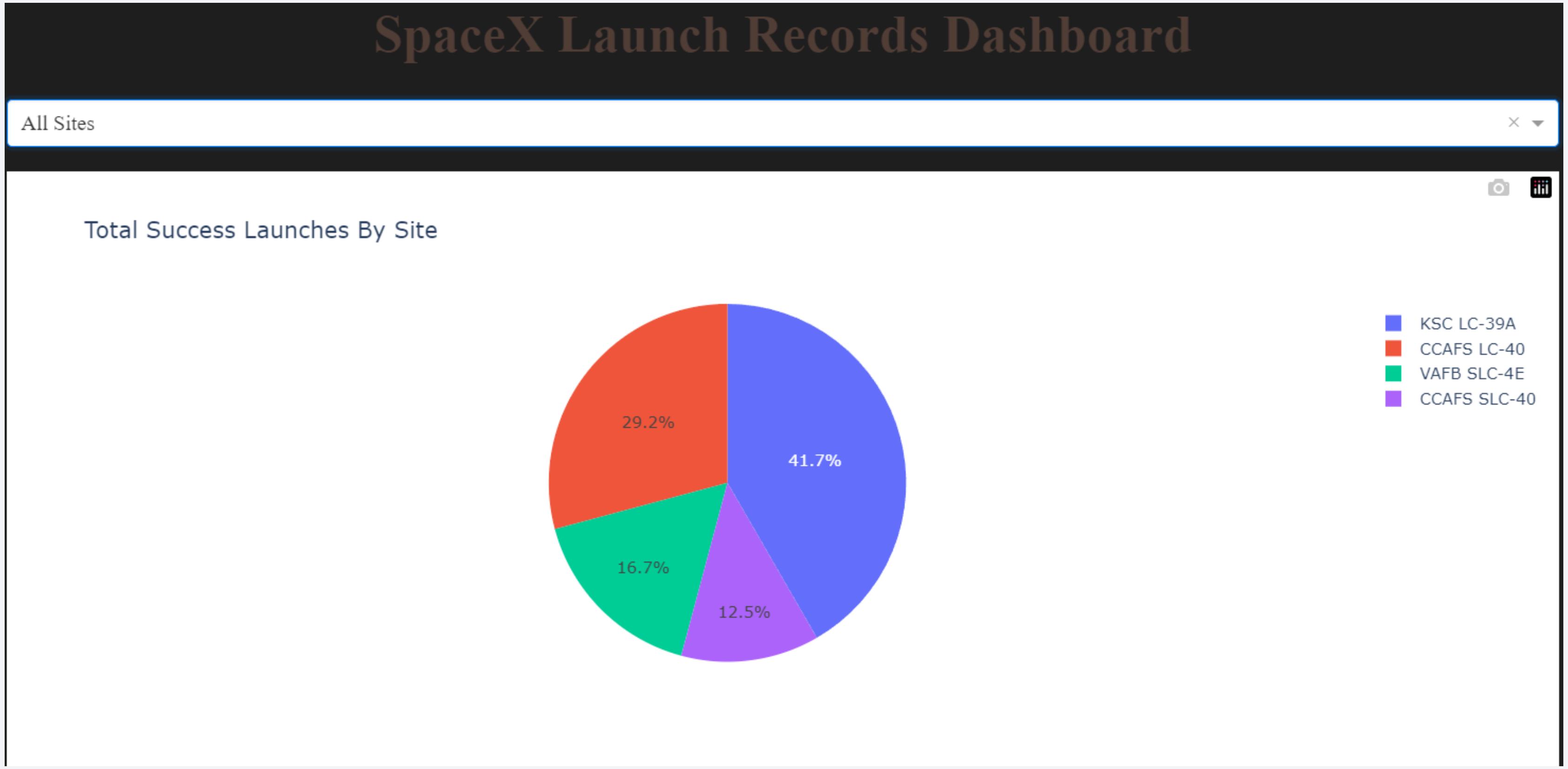
Distance of launch Site to the coast

Section 4

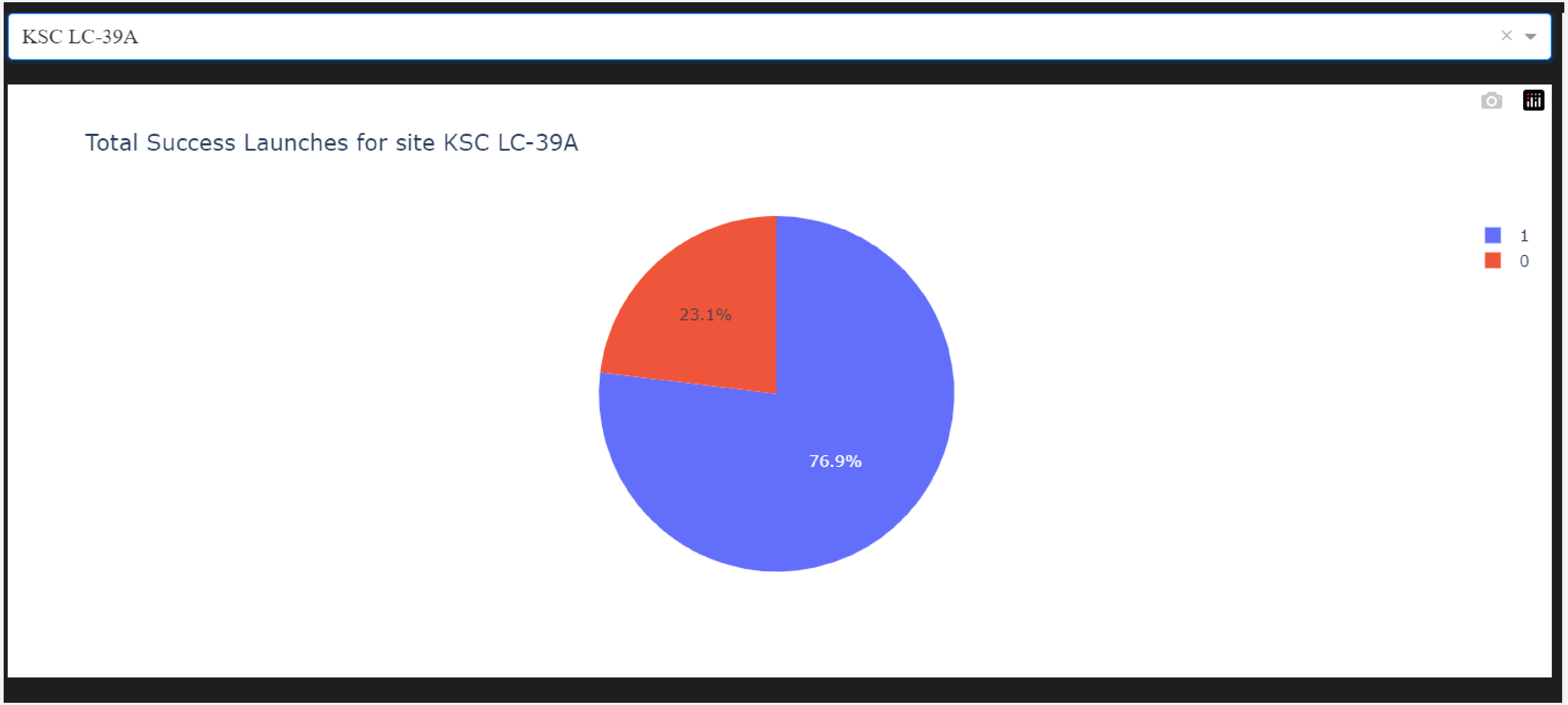
Build a Dashboard with Plotly Dash

Launch Evaluation by sites

- **KSC LC 39 A shows the highest success rate of launches**



Success Rate of Launch Site KSC LC 39A



Influence of Payload on Success Rate

Section 5

Predictive Analysis (Classification)

Classification Accuracy

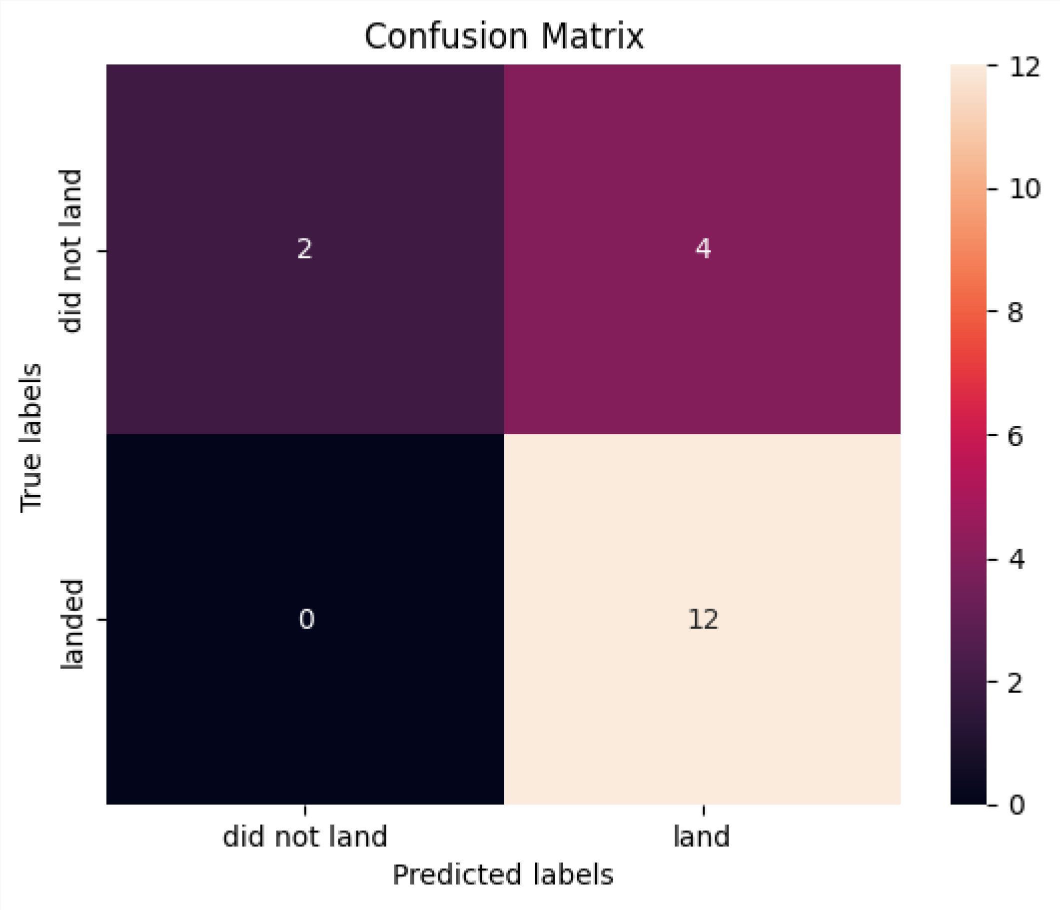
- All evaluated models showed similar accuracy thus no statement can be made which method is best suitable for classification

	ML Method	Accuracy Score (%)
0	Support Vector Machine	83.333333
1	Logistic Regression	83.333333
2	K Nearest Neighbour	83.333333
3	Decision Tree	83.333333

Confusion Matrix

The confusion matrix is the derived from KNN classification

We can see that this model has high accuracy with detecting true positives but performs less well in detection of true negatives since 50 % fall into false negatives for no landing



Conclusions

- To predict the cost of rocket launches , this project aimed to evaluate the success rate of first stage landing

- Conclusions

Rocket launch success rate increased over time

Orbits ES L1, GEO, HEO and SSO have a 100 % success rate in contrast to rest of studied orbits

Most launches are located near the coast

All evaluated classification algorithms showed an accuracy score of 83 %

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!