IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Rahul K S
August 15, 2024

Outline

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

Executive Summary

Summary of methodologies

- Data Collection using SpaceX REST API, Web Scraping
- Data Wrangling
- Exploratory Data Analysis (EDA) with SQL
- EDA with Data Visualization
- Interactive Maps using Folium
- Interactive Dashboards with Plotly Dash
- Predictive Analysis

Summary of all results

- Exploratory Data Analysis results
- Interactive maps and dashboards
- Predictive analysis results

Introduction

Project background and context

SpaceX is a private company owned and founded by Elon Musk. The company provides space transport services to government and private companies and develops rockets and spacecraft.

Falcon 9 is a partially reusable, human-rated, two-stage-to-orbit, medium-lift launch vehicle designed and manufactured in the United States by SpaceX.

The aim of the project is to predict if the first stage of the Falcon 9 LV will successfully land. The cost to launch the LV is claimed to be 62 million dollars. Other providers cost upwards of 165 million dollars each. This price difference is due to the fact that SpaceX can reuse the first stage.

Therefore, if we can determine if the first stage will land, we can determine the cost of the launch.

Introduction

Questions to be answered

- What are the main characteristics of a successful landing?
- How do variable such as payload mass, launch sites, number of launches and orbit affect the success rate of landings?
- Does rate of successful landings increase over the years?
- What are the conditions that will allow SpaceX to achieve the best landing success rate?

Section 1

Methodology

Methodology

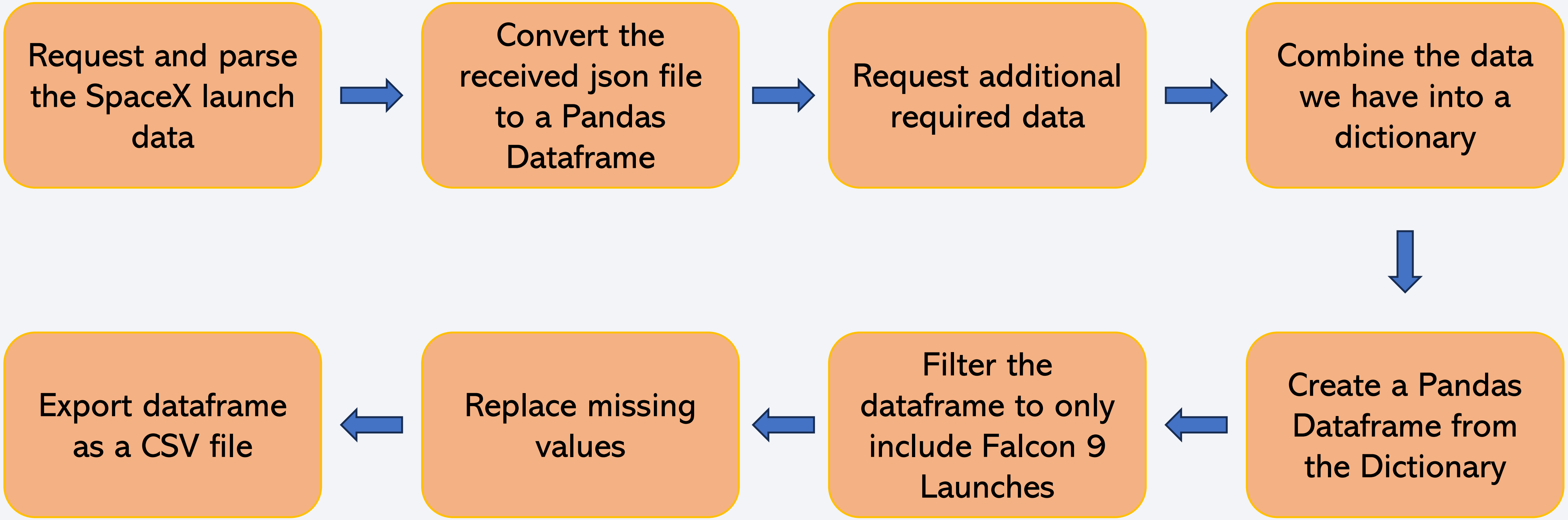
Executive Summary

- Data collection methodology:
  - SpaceX REST API
  - Web scraping from Wikipedia
- Perform data wrangling
  - Dealing with missing data
  - Using One Hot Encoding to prepare data for classification
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Comparing different models and utilizing Hyperparameter tuning methods to ensure best model performance

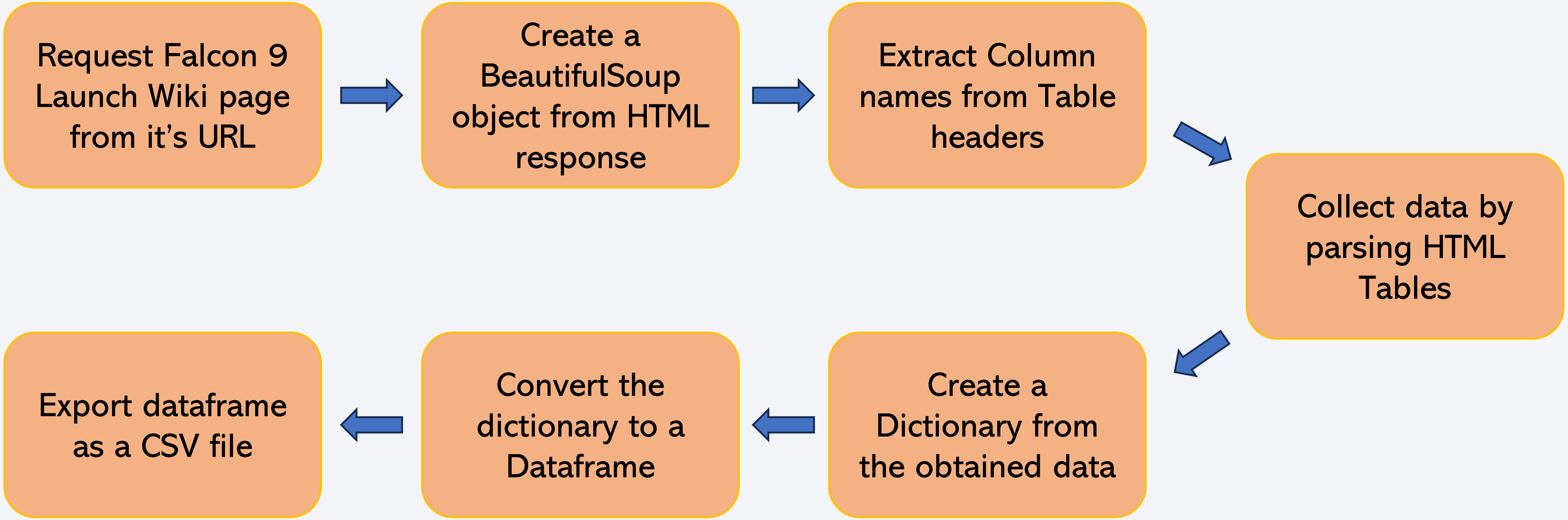
Data Collection

- There were two ways the data was sourced. One was to use the SpaceX REST API and the other was to Web Scrape data from SpaceX's Wikipedia page.
- Data obtained by using SpaceX REST API include FlightNumber, Date, BoosterVersion, PayloadMass, Orbit, LaunchSites, Outcomes.
- Data obtained by using Wikipedia Web Scraping include FlightNo, LaunchSite, Customer, LaunchOutcome, VersionBooster, Booster Landing, Date, Time.

Data Collection – SpaceX API



Data Collection - Scraping



Data Wrangling

Calculate the number of launches on each site → Calculate the number and occurrence of each orbit → Calculate the number and occurrence of mission outcome of each orbit → Create a Landing outcome label from Outcome column

We Translate the results into Training Labels, where ‘1’ indicates a Successful Landing and ‘0’ indicates a Failure in Landing.

EDA with Data Visualization

Scatter Plots:

Scatter Plots depict the relation between variables

- Flight Number vs Payload Mass
- Flight Number vs Launch Site
- Payload vs Launch Site
- Payload vs Orbit
- Orbit vs Flight Number
- Orbit vs Payload Mass

Bar Graphs:

Bar Graphs display comparison between distinct categories

- Success Rate vs Launch Site

Line Charts:

Line charts depict pattern in data across time

- Success Rate vs Year

EDA with SQL

The following SQL Queries were performed:

- 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
- List the date when the first successful landing outcome in ground pad was achieved
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List the names of the booster_versions which have carried the maximum payload mass. Use a subquery
- 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
- 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

- Markers have been added to all Launch Sites on the map including a Circle, Popup Label and a Text Label.
- Successful/Failed Launches have been marked for each Launch Site. Green indicates a successful launch whereas Red indicates a Failed launch.
- Distances have been calculated between each Launch Site and key locations such as railway, highway, coastlines and cities.

Build a Dashboard with Plotly Dash

- Dropdown that allows a user to choose a launch site or all launch sites.
- A Pie chart shows the total success and the total failure for the launch site chosen from the dropdown.
- Rangeslider allows the user to select Payload mass in a fixed range.
- Scatter plot shows the relation between Success and the Payload mass selected by the user.

Predictive Analysis (Classification)

- Data Preparation
  - Load Dataset
  - Normalize Data
  - Train Test Split
- Model Preparation
  - Selection of ML Algorithm
  - Determining the range of Parameters
  - Utilizing GridSearchCV to tune the Hyperparameters
- Model Evaluation
  - Compute accuracy score using the test dataset
  - Plot Confusion matrix
- Model Comparison
  - Compare different model performance based on their accuracy score
  - The model with the best performance is chosen

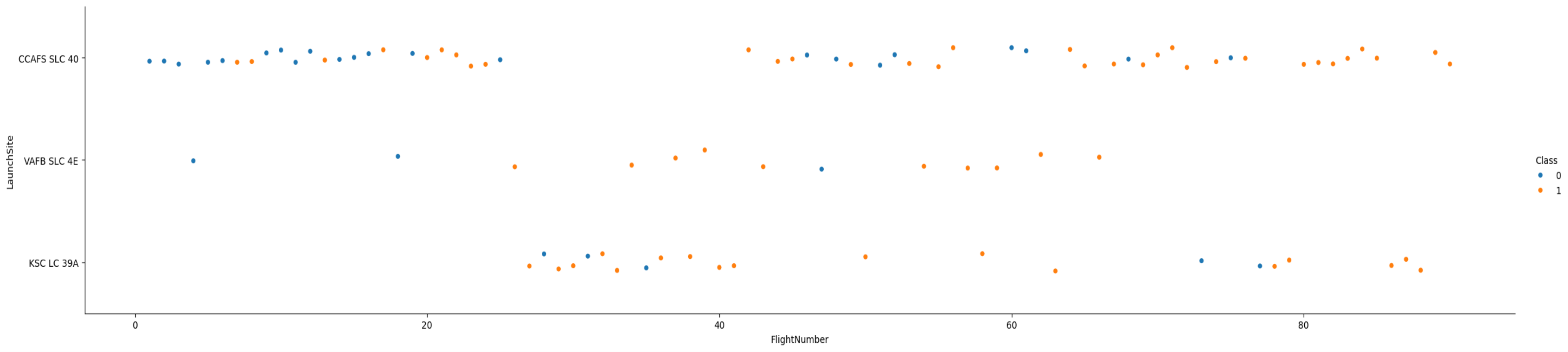
Results

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

Section 2

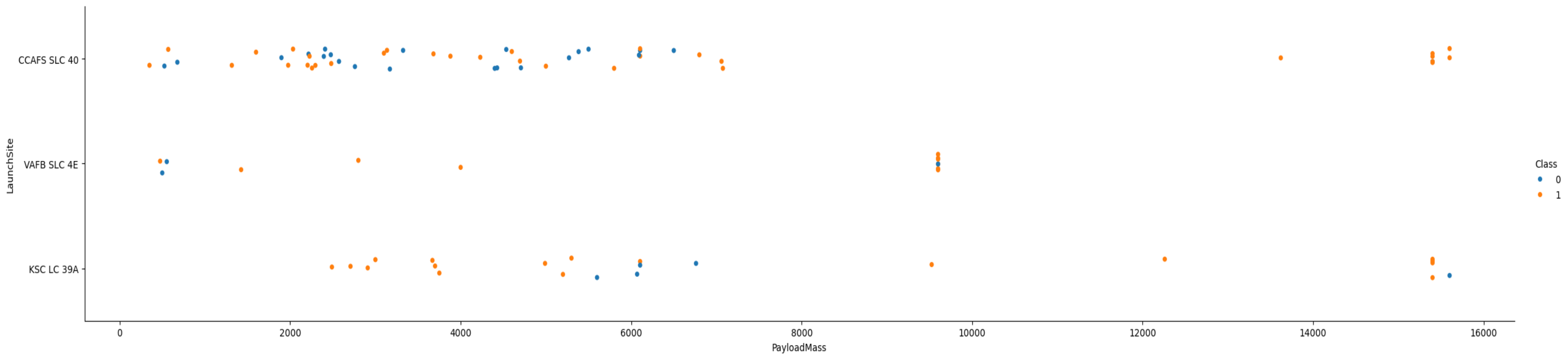
Insights drawn from EDA

Flight Number vs. Launch Site



- The Success rating is increasing for each site.
- Launch Site CCAFS SLC 40 accounts to more number of launches relative to the other two launch sites.

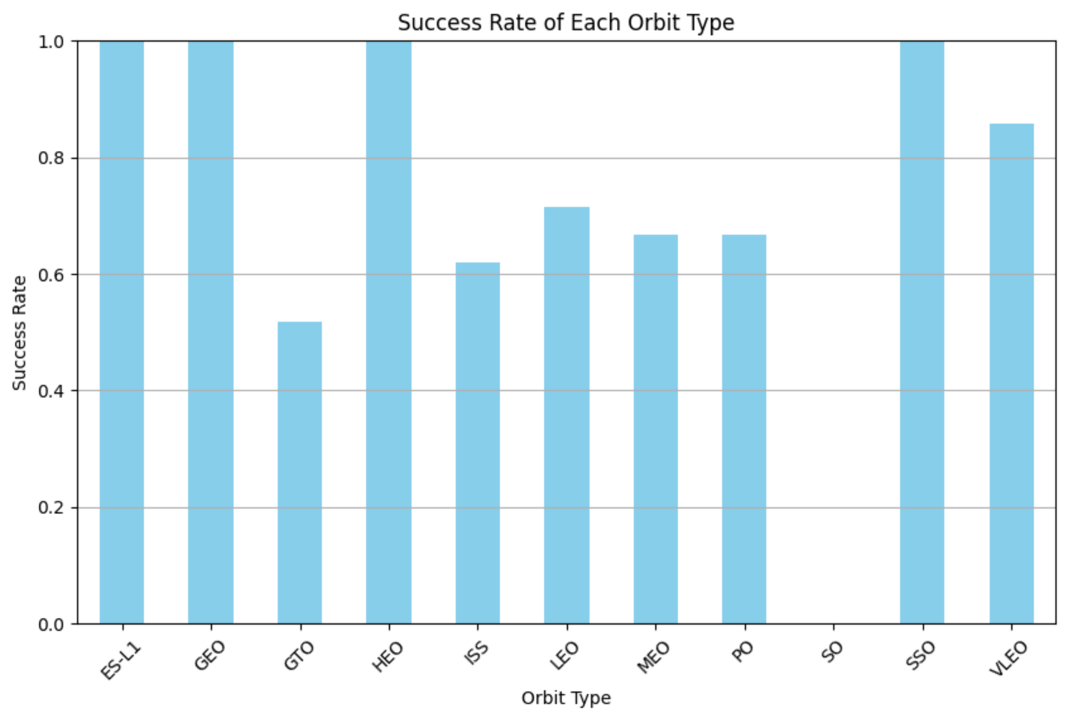
Payload vs. Launch Site



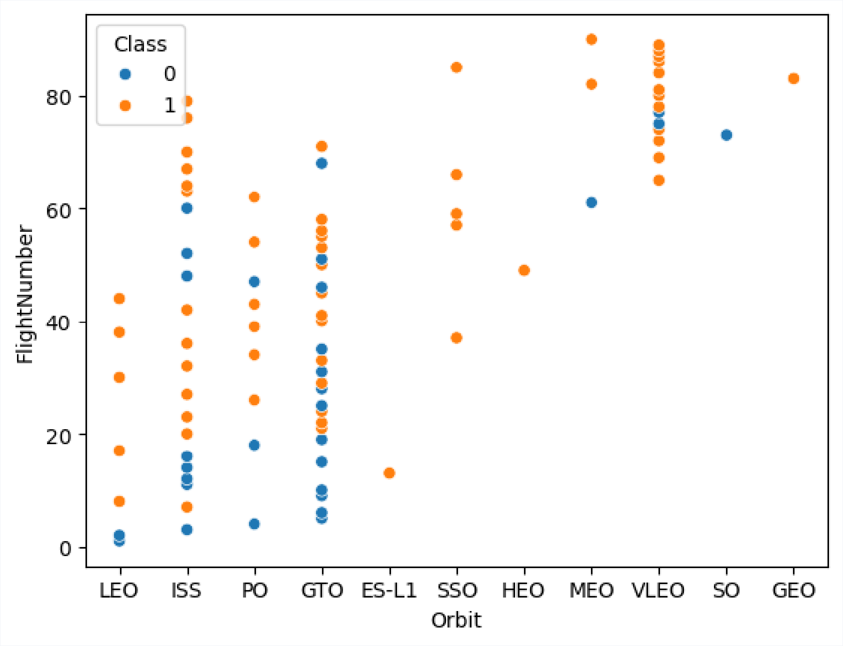
- The majority of the missions weighing above 7000 kg were successful.
- KSC LC 39A has an 100% success rate for payloads below 5500 kg.

Success Rate vs. Orbit Type

- Orbits with 100% success rate: ES-L1, GEO, HEO, SSO
- Orbits with 0% success rate: SO



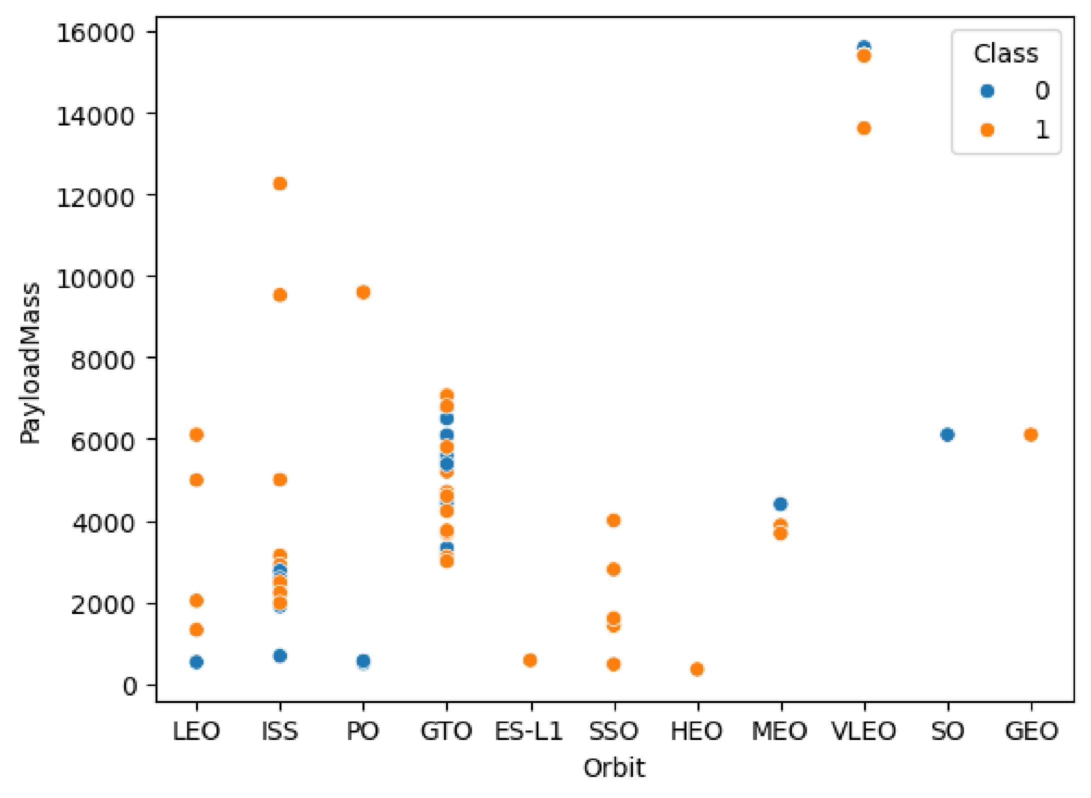
Flight Number vs. Orbit Type



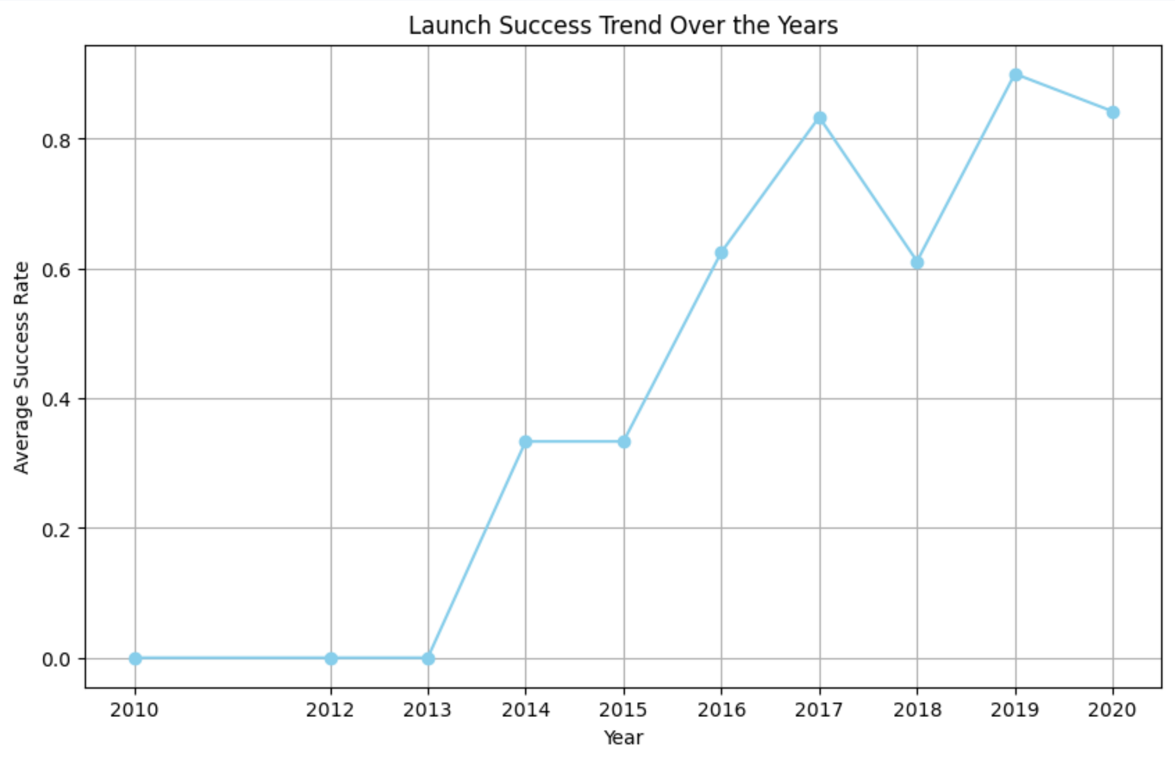
- You can observe that in the LEO orbit, success seems to be related to the number of flights. Conversely, in the GTO orbit, there appears to be no relationship between flight number and success

Payload vs. Orbit Type

- 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 2013 kept increasing till 2020

All Launch Site Names

```
%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

- Displaying the names of distinct launch sites.

Launch Site Names Begin with '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_Outc
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (paracl
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 (paracl
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No atte
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No atte
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No atte

- Displaying 5 entries with Launch Site Name that begins with ‘CCA’.

Total Payload Mass

```
%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

- Displaying Total Payload mass where Customer is NASA.

Average Payload Mass by F9 v1.1

```
%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

- Displaying the average Payload mass carried by F9 v1.1

First Successful Ground Landing Date

```
%sql select min(Date) from SPACEXTABLE where Landing_Outcome = 'Success'
```

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

min(Date)
2018-07-22

- Displaying the first successful Ground landing date

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql SELECT booster_version FROM SPACEXTBL WHERE "Landing_Outcome" = 'Success (drone ship)' and payload_mass__kg_ BETWEEN 4(
```

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

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

- Displaying the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Total Number of Successful and Failure Mission Outcomes

```
%sql select Mission_Outcome, count(Mission_Outcome) from SPACEXTABLE group by Mission_Outcome
```

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

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

- Displaying the total number of successful and failure mission outcomes

Boosters Carried Maximum Payload

```
%sql select Booster_Version from SPACEXTABLE where PAYLOAD_MASS__KG_ in (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

- Displaying the names of the booster which have carried the maximum payload mass

2015 Launch Records

```
%%sql select monthname(date) as month, date, booster_version, launch_site, landing__outcome from SPACEXDATASET
      where landing__outcome = 'Failure (drone ship)' and year(date)=2015;
```

```
 * ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90108kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

MONTH	DATE	booster_version	launch_site	landing__outcome
January	2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
April	2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

- Displaying the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

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

```
%%sql select landing__outcome, count(*) as count_outcomes from SPACEXDATASET
      where date between '2010-06-04' and '2017-03-20'
      group by landing__outcome
      order by count_outcomes desc;
```

```
 * ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

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

- Ranking 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

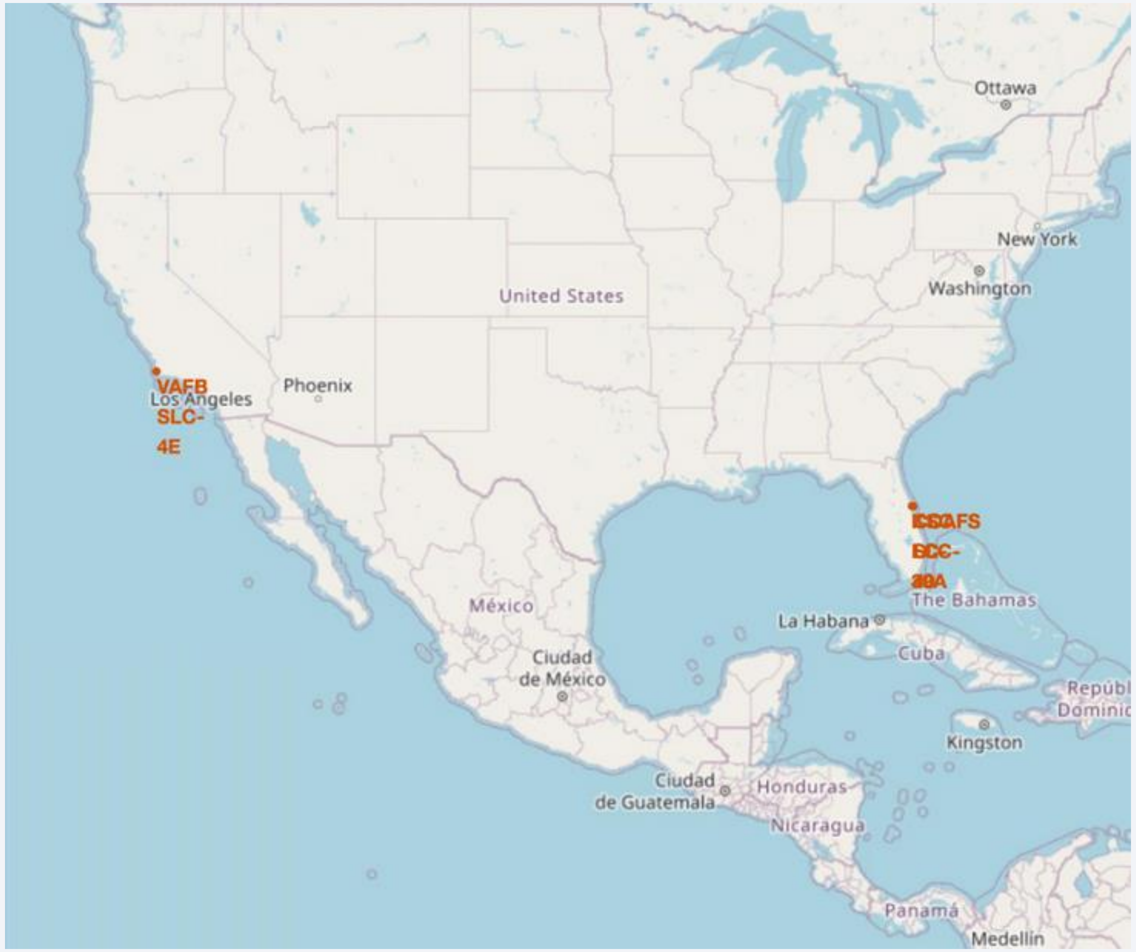
Section 3

Launch Sites Proximities Analysis

SpaceX Launch Sites

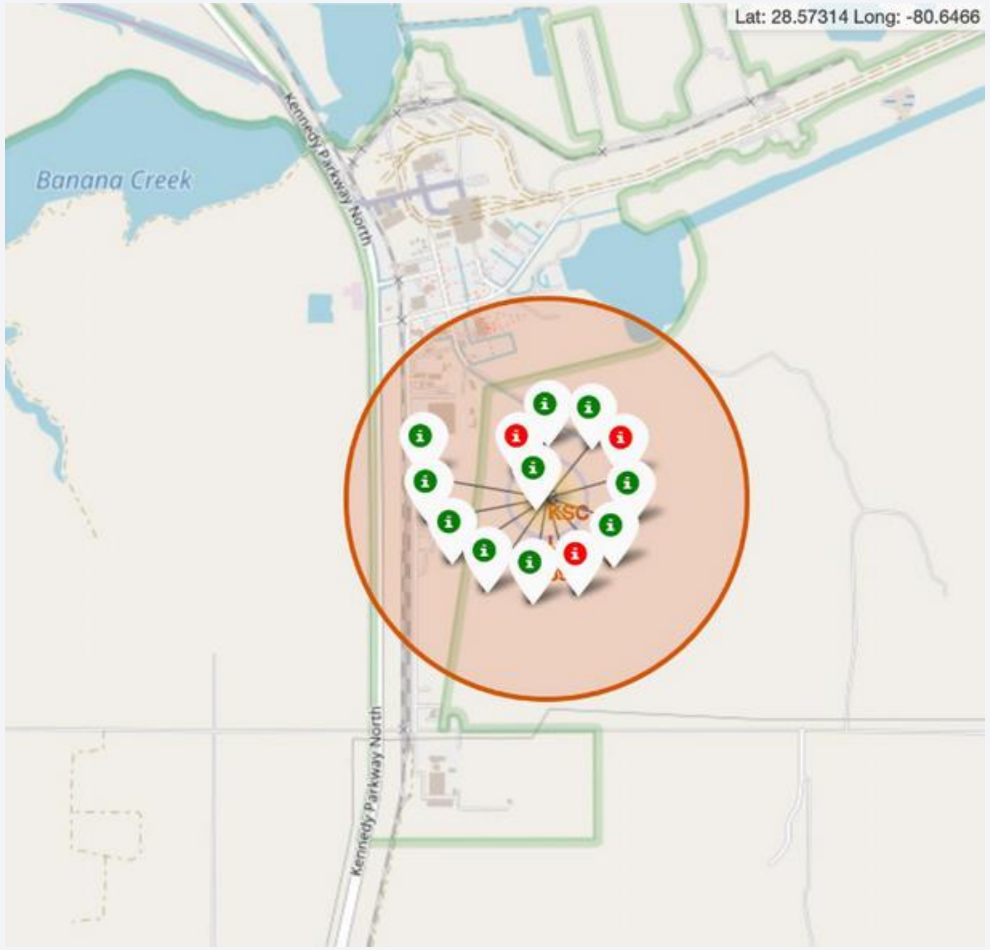
- The Launch Sites are in close proximity to the Equator.

- These Launch sites are also located close to the shore to minimize the damage caused due to any debris from the launch.



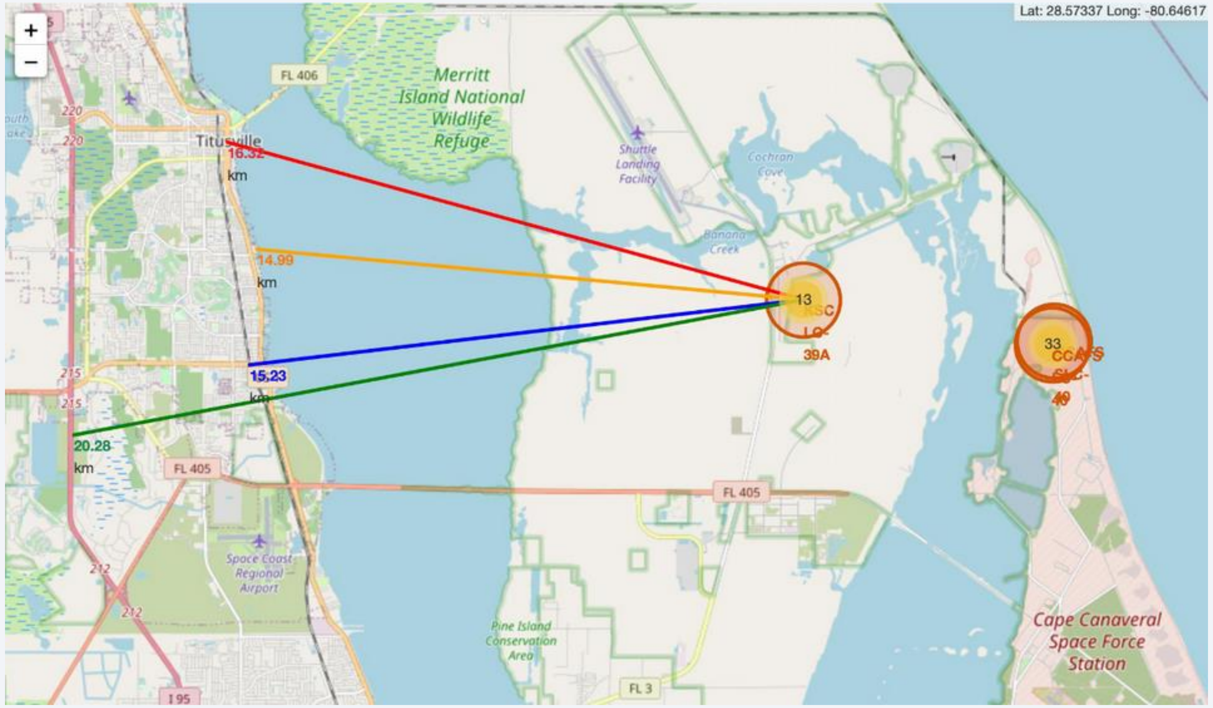
Success/Failed Launches for each Site

- Green marker indicates a Successful Launch
- Red marker indicates a Failed Launch
- Launch site KSC LC-39A has a high success rate



Distances between KSC LC-39A and its proximities

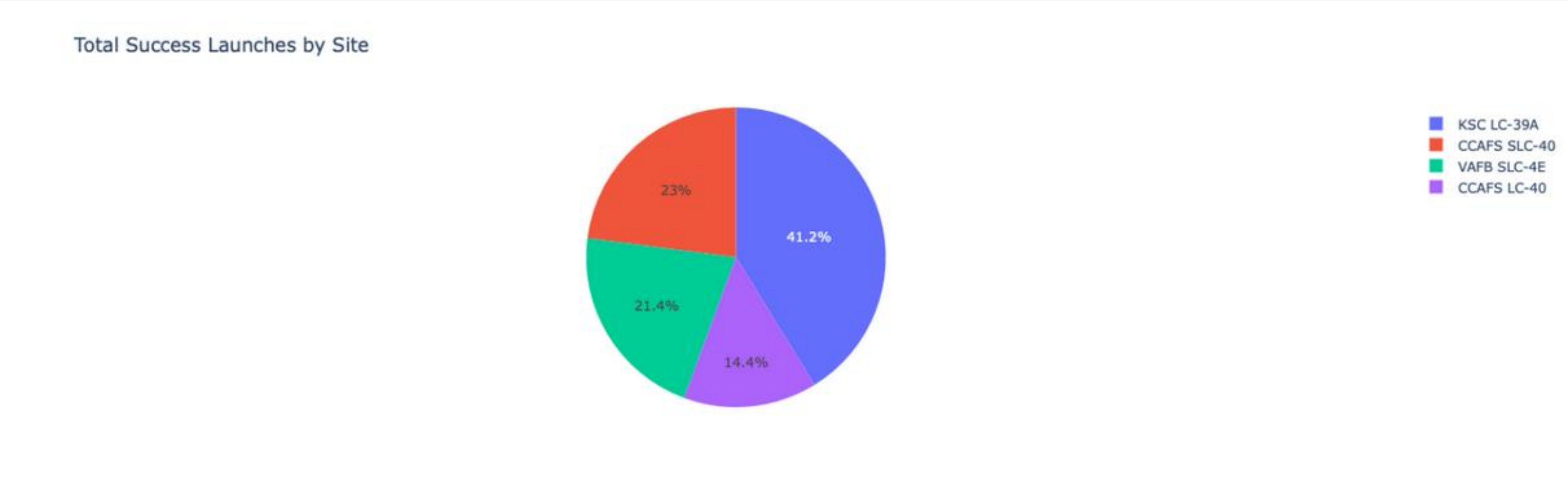
- Launch site KSC LC-39A is close in proximity to a Railway.
- It is close in proximity to a Highway
- It is close in proximity to a Coastline
- It is relatively close to a city with may be dangerous



Section 4

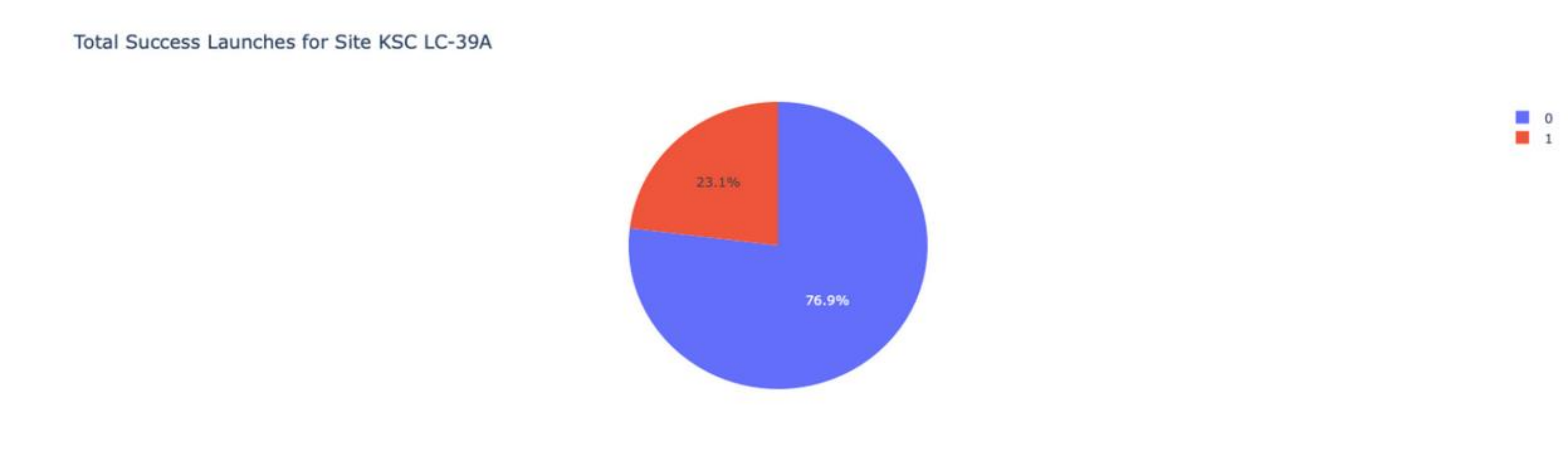
Build a Dashboard with Plotly Dash

Total Successful launches by Site



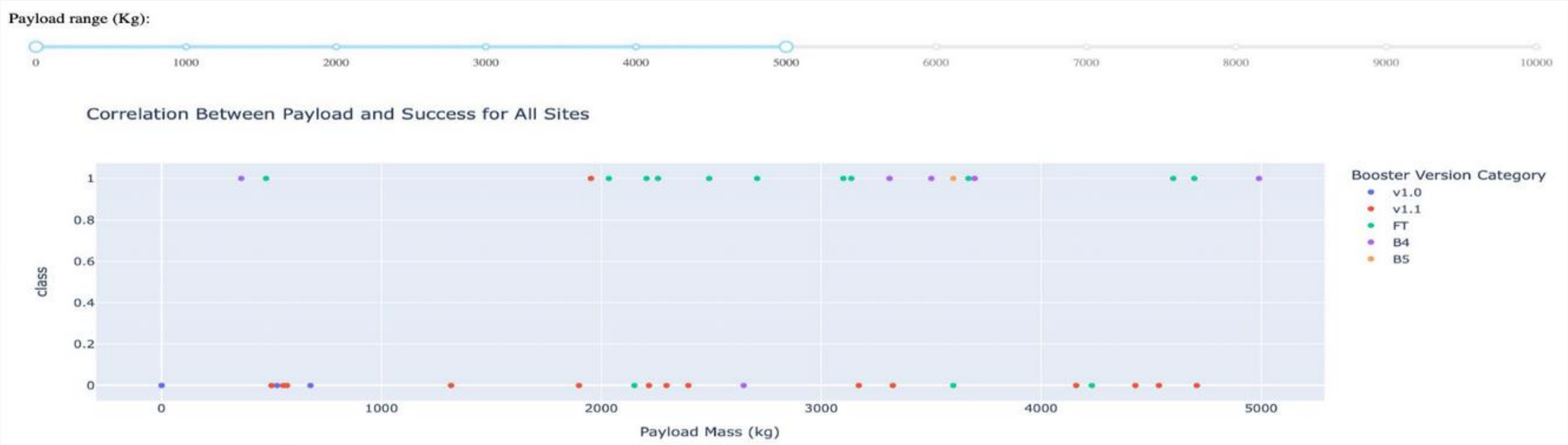
- KSC LC-39A is the most Successful launch site.

Total Successful launches for site KSC LC-39A



- Launch Site KSC LC-39A has a success rate of 76.9% and a failure rate of 23.1%.

<Dashboard Screenshot 3>



- Payloads ranging from 1900 kg to 5000 kg have the highest success rate.

Section 5

Predictive Analysis (Classification)

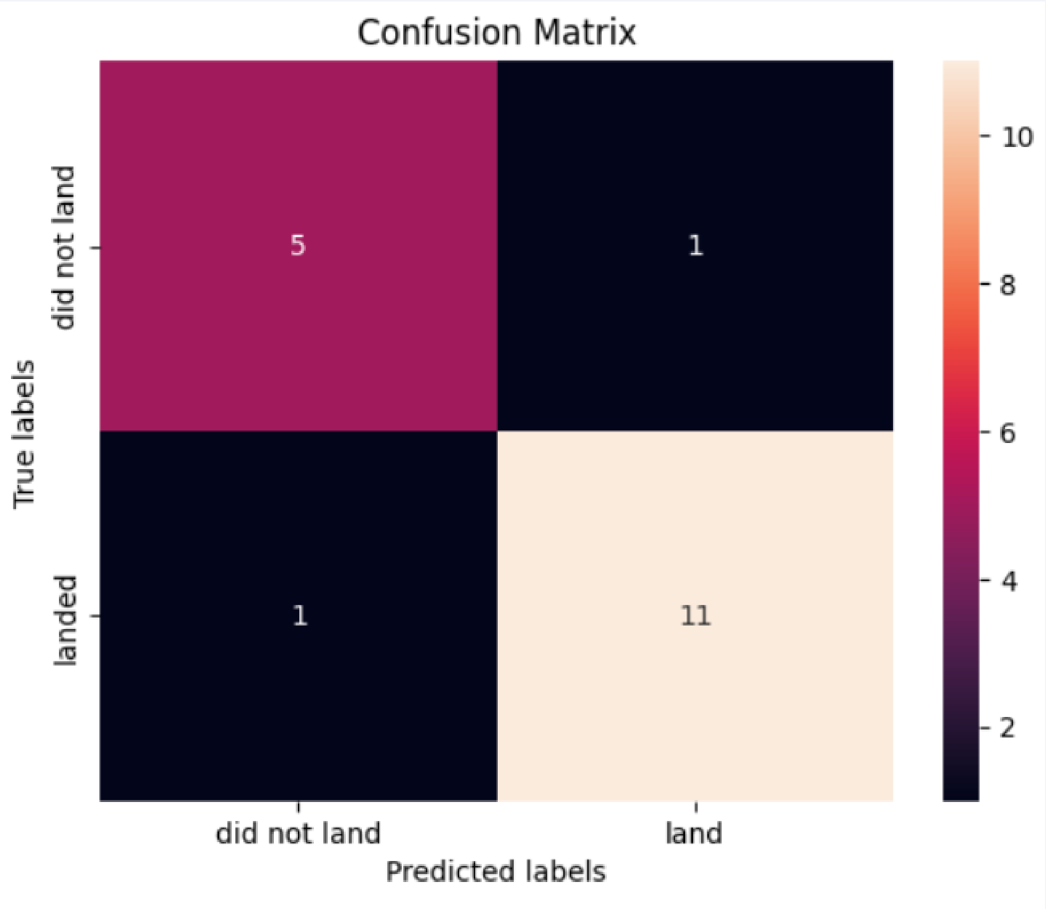
Classification Accuracy

```
Best model is DecisionTree with a score of 0.875
Best params is : {'criterion': 'entropy', 'max_depth': 14, 'max_features': 'sqrt', 'min_samples_leaf': 2, 'min_samples_split'
: 10, 'splitter': 'best'}
```

- The Decision Tree Classifier is the model with the highest accuracy and its best Hyperparameters are shown above.

Confusion Matrix

- The confusion matrix helps assess the Decision Tree classifier's performance by comparing predicted values against actual values for a dataset.
- This model is accurate with only one False Positive and one False Negative.



Conclusions

- A Launch Site's Success rates increased with increase in number of launches.
- The Launch success rate increased from the year 2013 up to 2020.
- Orbits ES-L1, GEO, HEO, SSO and VLEO have high success rates.
- KSC LC-39A is the most successful launch site.
- The Decision Tree Classifier is the model with the highest accuracy score for this problem.

Thank you!