SpaceX

First Stage Reuse

Aditya Akuskar

22 /11/2024

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion
  - Findings & Implications
- Conclusion

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

As SpaceX has achieved groundbreaking advancements in reusable rocket technology and space exploration, making it a leader in the global space industry.

This project showcases the practical applications of data science in the aerospace industry. It demonstrates how data-driven insights can be used to enhance operational efficiency, reduce costs, and support strategic decision-making in groundbreaking companies like SpaceX.

Section 1

METHODOLOGY

METHODOLOGY

Executive Summary

- Data collection Method:
  - Data was collected using 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
  - How to build, tune, evaluate classification models

RESULTS

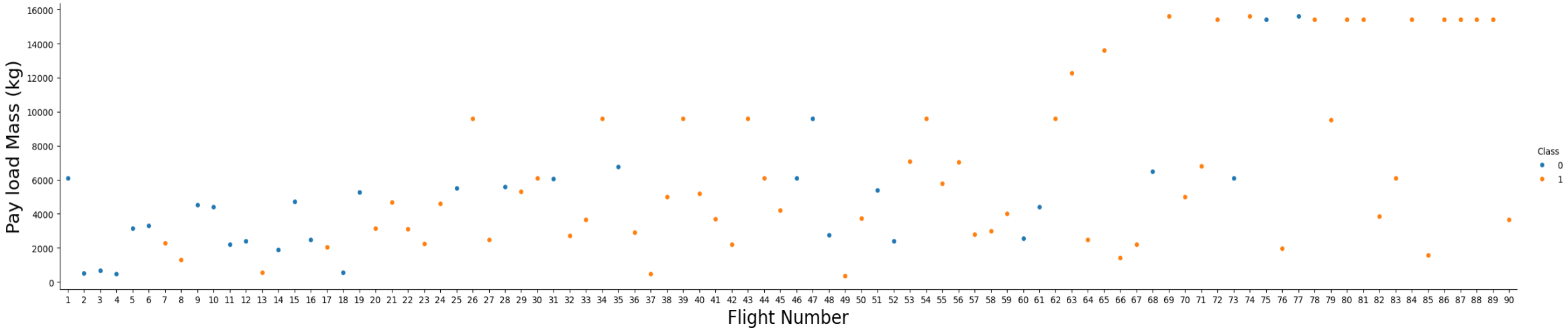
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn through EDA

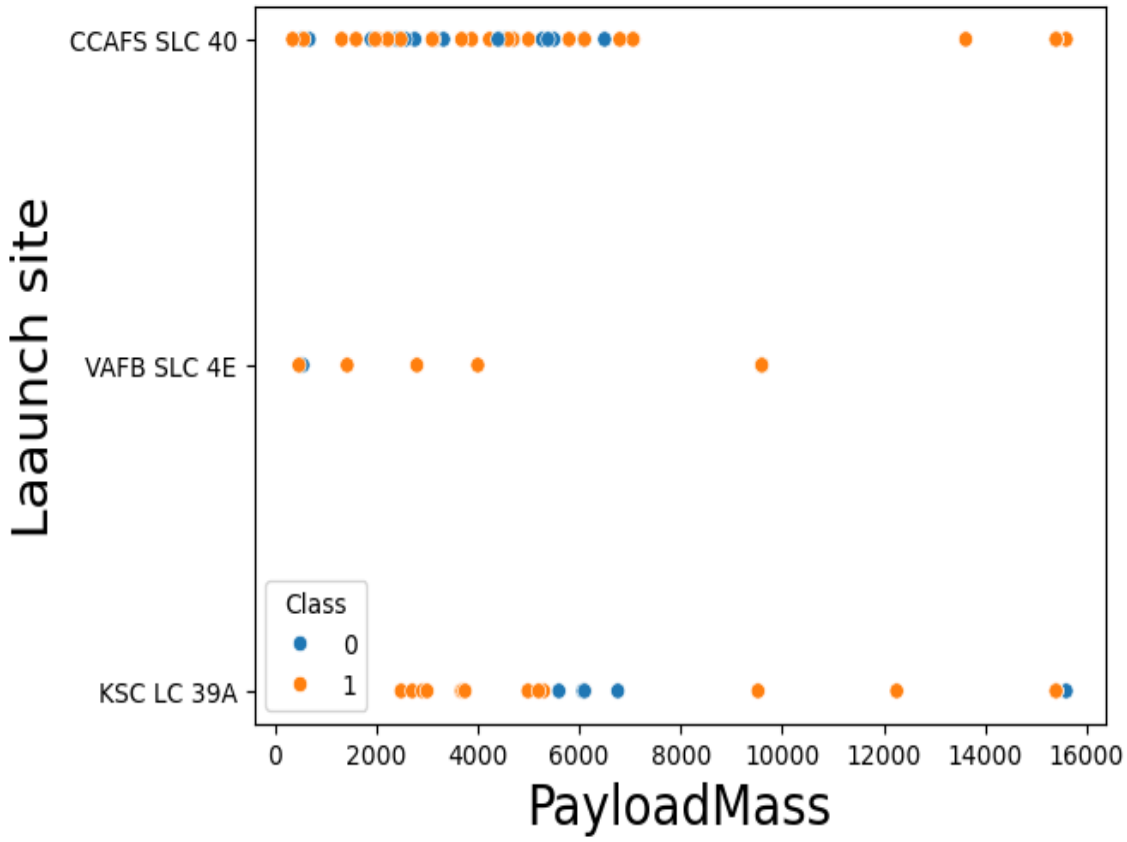
Flight Number vs. Pay load Mass(kg)

- From the plot, we see that as the flight number increases, the first stage is more likely to land successfully.
- The payload mass also appears to be a factor; even with more massive payloads, the first stage often returns successfully.



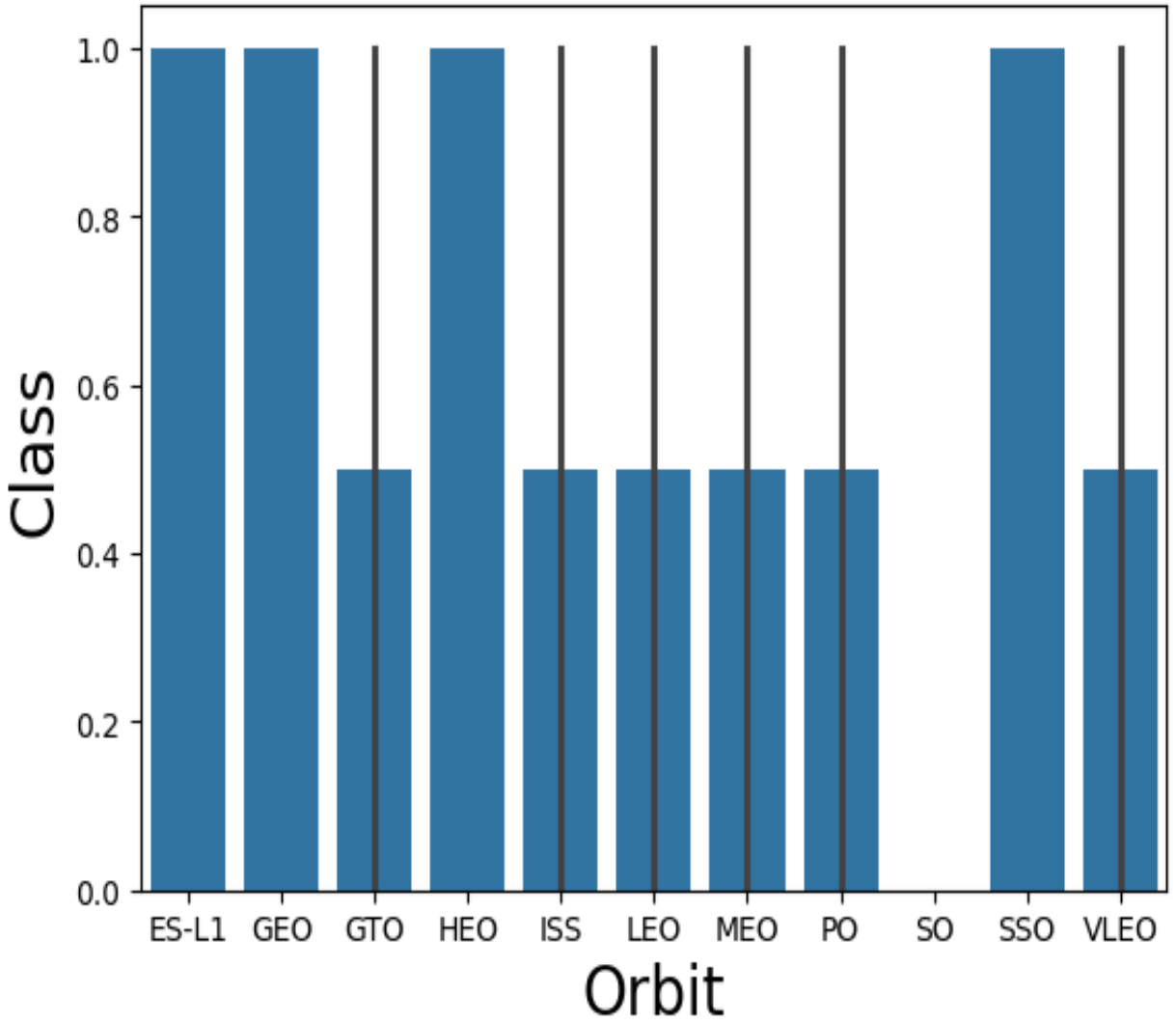
Payload vs. Launch Site

- As you can observe Payload Mass Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).



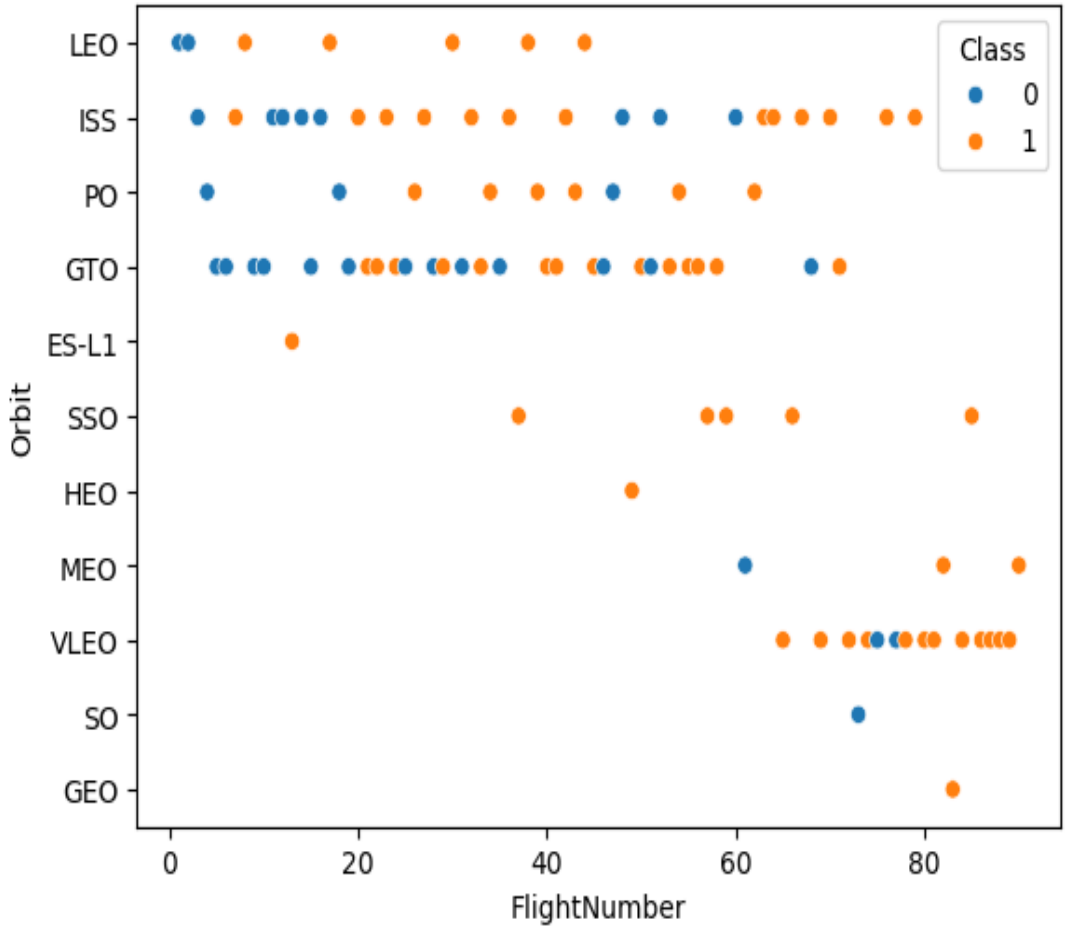
Success Rate vs. Orbit Type

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



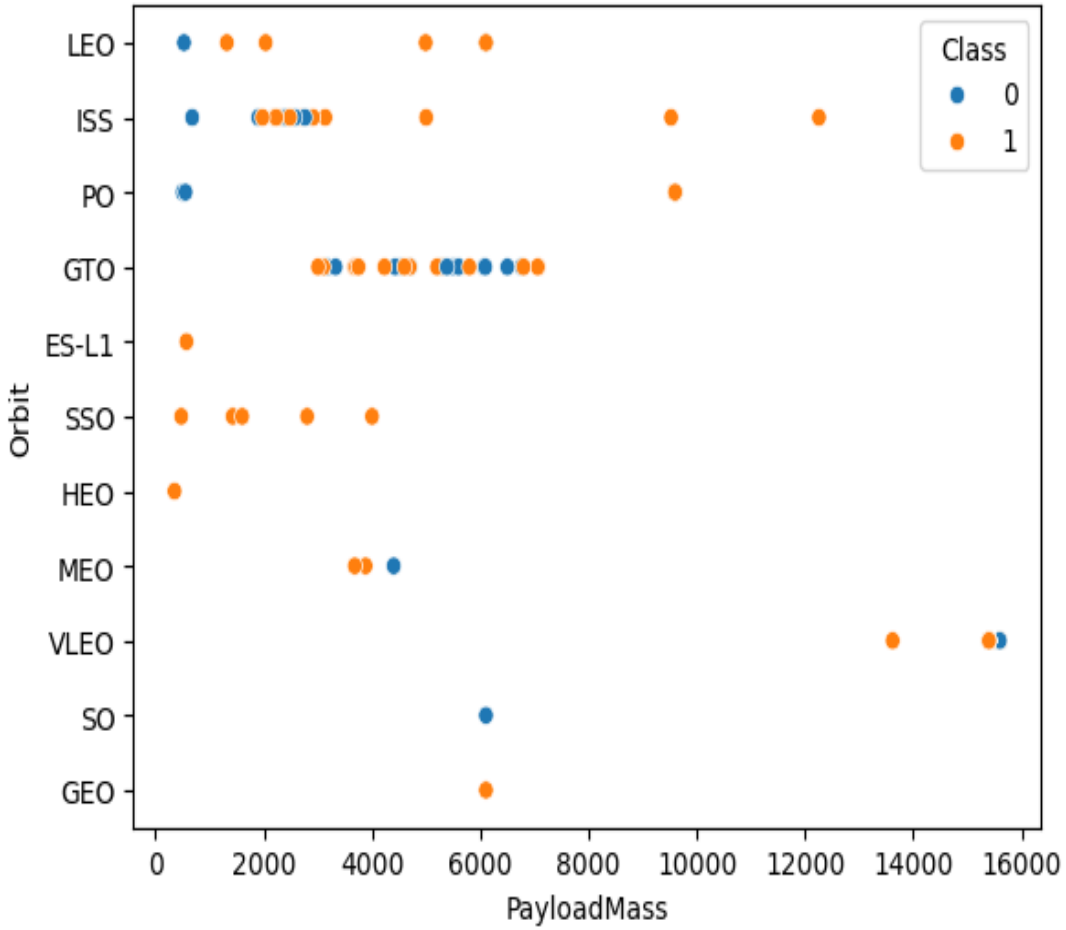
Flight Number vs. Orbit Type

- 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.



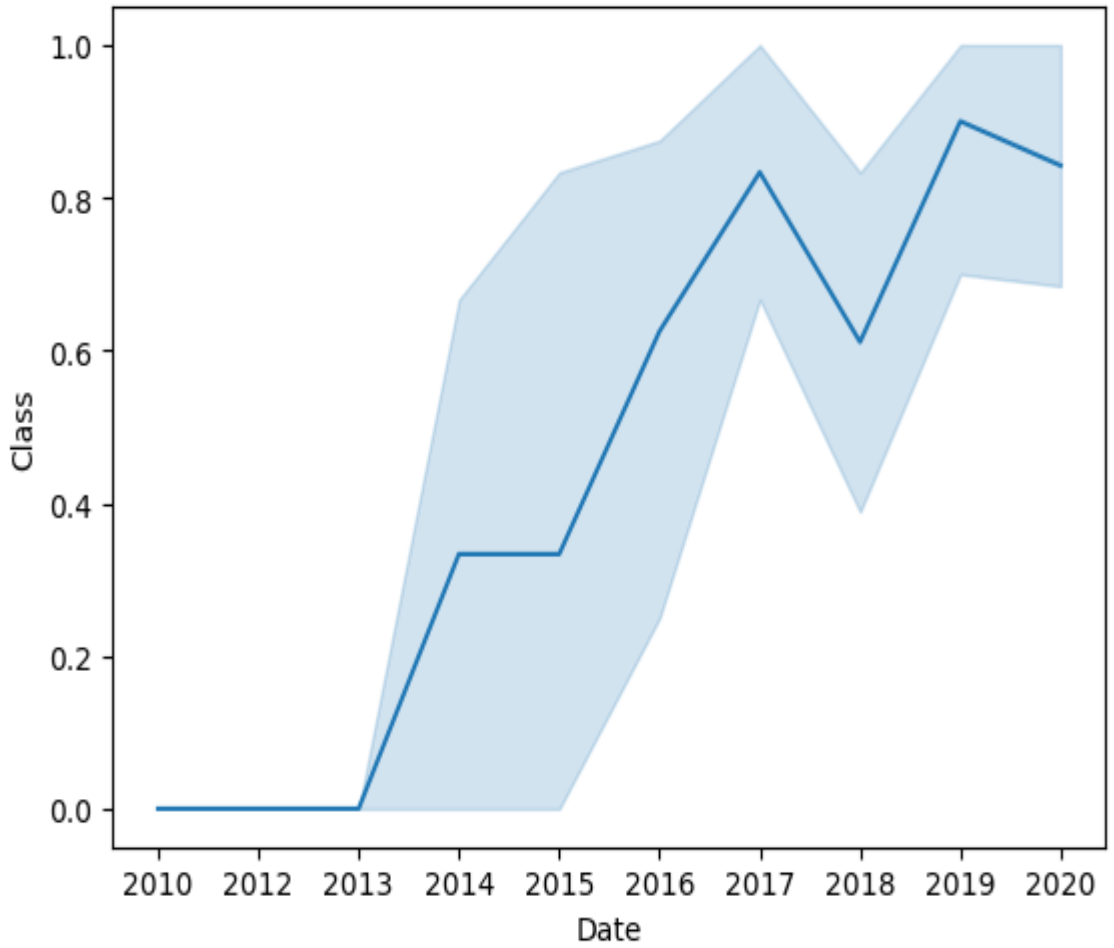
Payload vs. Orbit Type

- 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

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



Launch Sites Information

- Using Distinct keyword we shown all the Launch sites from where Falcon rockets where launched.

```
query = """ SELECT distinct(Launch_Site) FROM df"""
result = pd.read_sql_query(query, conn)
print(result)
```

```
   Launch_Site
0   CCAFS LC-40
1   VAFB SLC-4E
2    KSC LC-39A
3  CCAFS SLC-40
```

- Total payload mass carried by boosters launched by NASA (CRS). Average payload mass carried by booster version F9 v1.1.

```
query = """ SELECT avg(PAYLOAD_MASS__KG_) as Avg_payload_mass
            FROM df
            WHERE Booster_Version ='F9 v1.1'"""
result = pd.read_sql_query(query, conn)
print(result)
```

```
   Avg_payload_mass
0            2928.4
```

- Average payload mass carried by booster version F9 v1.1.

```
query = """ SELECT sum(PAYLOAD_MASS__KG_) as Total_payload_mass
            FROM df
            WHERE Customer = 'NASA (CRS)'"""
result = pd.read_sql_query(query, conn)
print(result)
```

```
   Total_payload_mass
0               45596
```

- Date when the first succesful landing outcome in ground pad was acheived.

```
query =  """SELECT min(DATE)
            FROM df
            WHERE Landing_Outcome = 'Success (ground pad)'"""
result = pd.read_sql_query(query, conn)
print(result)
```

```
    min(DATE)
0  2015-12-22
```

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

In [63]:

```
query = """SELECT Booster_Version
           FROM df
           WHERE PAYLOAD_MASS__KG_ > 4000
           AND PAYLOAD_MASS__KG_ < 6000
           AND Landing_Outcome = 'Success (ground pad)' """
result = pd.read_sql_query(query, conn)
print(result)
```

```
  Booster_Version
0   F9 FT B1032.1
1   F9 B4 B1040.1
2   F9 B4 B1043.1
```

- Total number of successful and failure mission outcomes

In [69]:

```
query = """ SELECT COUNT(Mission_Outcome) AS SuccessOutcome
            FROM df
            WHERE Mission_Outcome LIKE 'Success%'"""
result = pd.read_sql_query(query, conn)
print(result)

query1 = """
        SELECT COUNT(Mission_Outcome) AS FailureOutcome
        FROM df
        WHERE Mission_Outcome LIKE 'Failure%'"""
result = pd.read_sql_query(query1, conn)
print(result)
```

```
   SuccessOutcome
0             100
   FailureOutcome
0               1
```

- Booster_versions which have carried the maximum payload mass

In [70]:

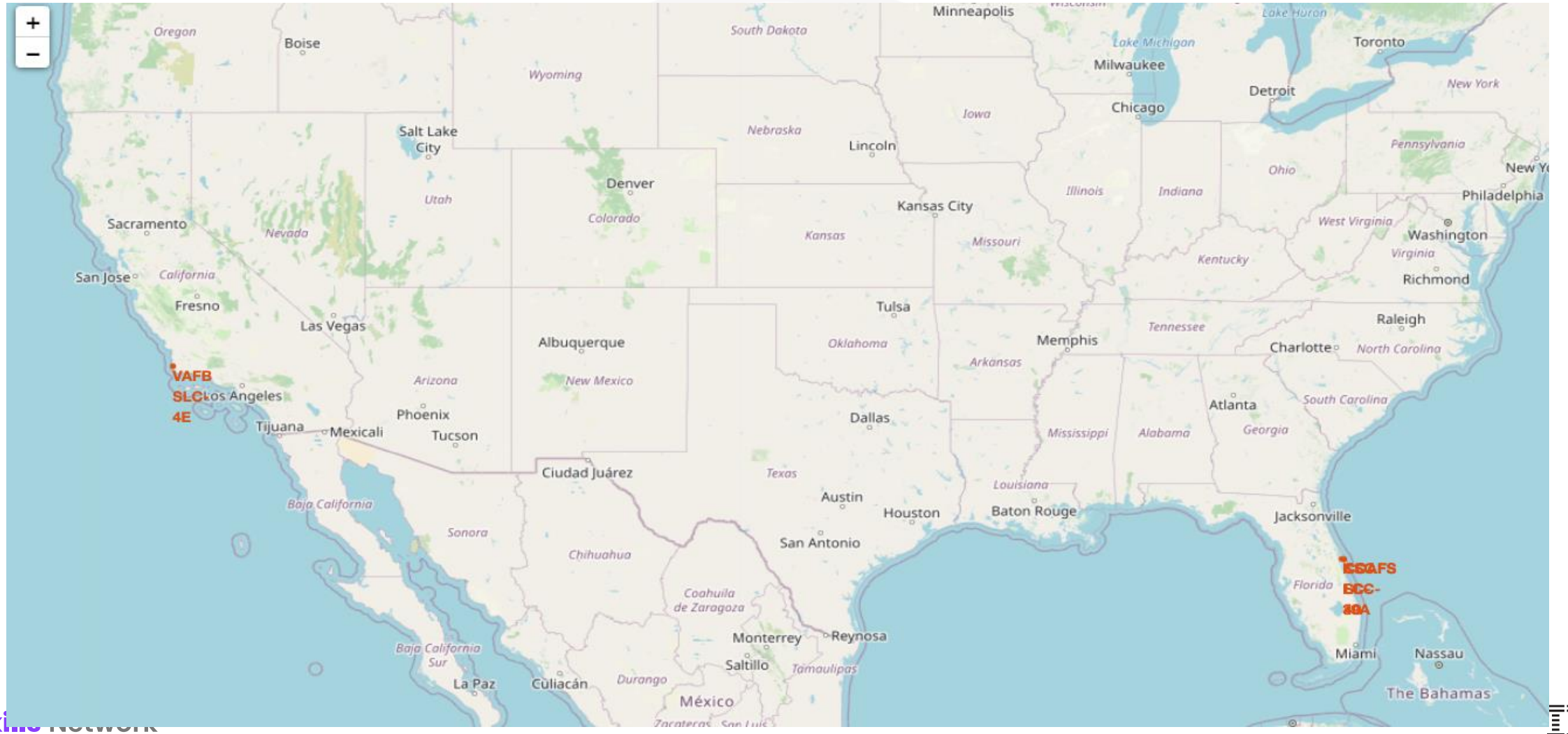
```
query = """SELECT Booster_version
           FROM df
           WHERE PAYLOAD_MASS__KG_ = (select max( PAYLOAD_MASS__KG_) from df)"""
result = pd.read_sql_query(query, conn)
print(result)
```

```
   Booster_Version
0     F9 B5 B1048.4
1     F9 B5 B1049.4
2     F9 B5 B1051.3
3     F9 B5 B1056.4
4     F9 B5 B1048.5
5     F9 B5 B1051.4
6     F9 B5 B1049.5
7    F9 B5 B1060.2
8    F9 B5 B1058.3
9     F9 B5 B1051.6
10    F9 B5 B1060.3
11   F9 B5 B1049.7
```

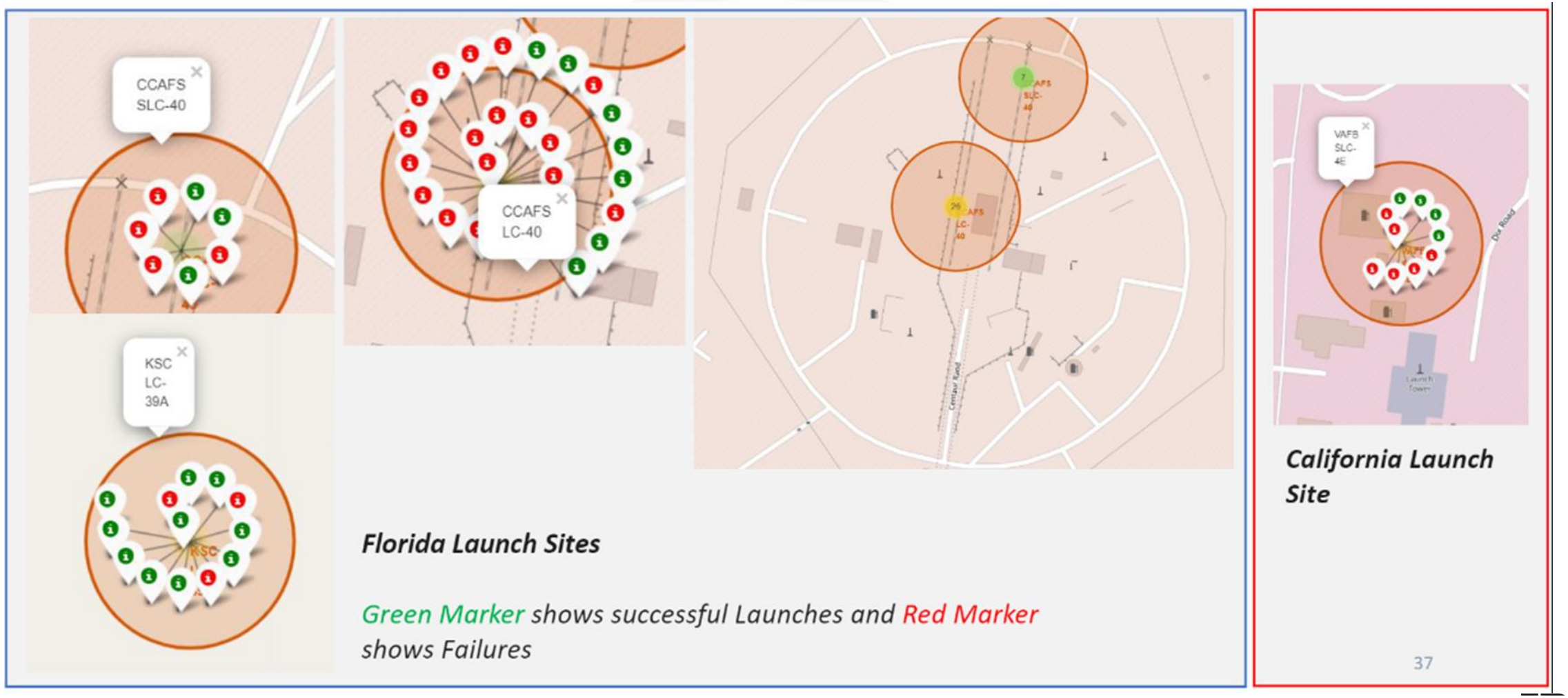
Section 3

Launch Site Analysis

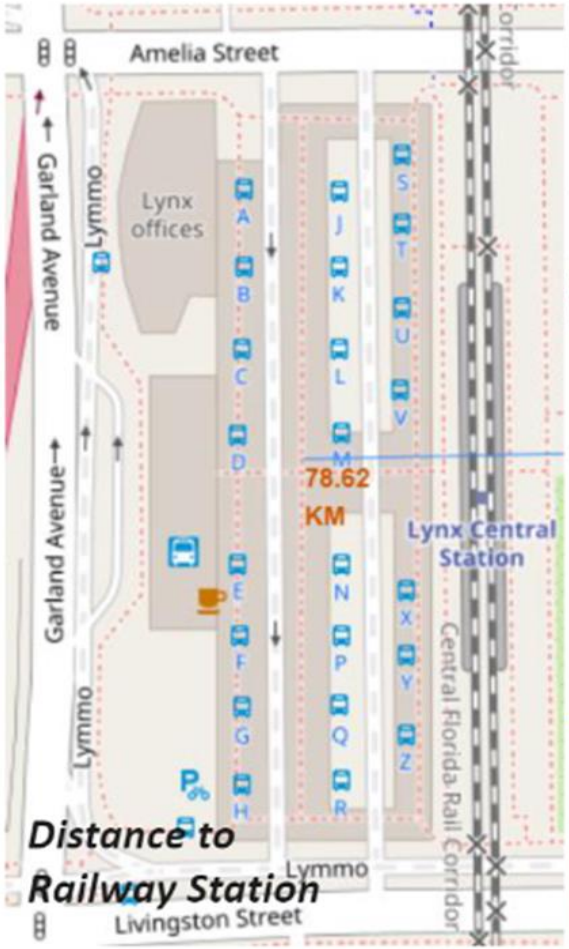
All launch sites global map markers

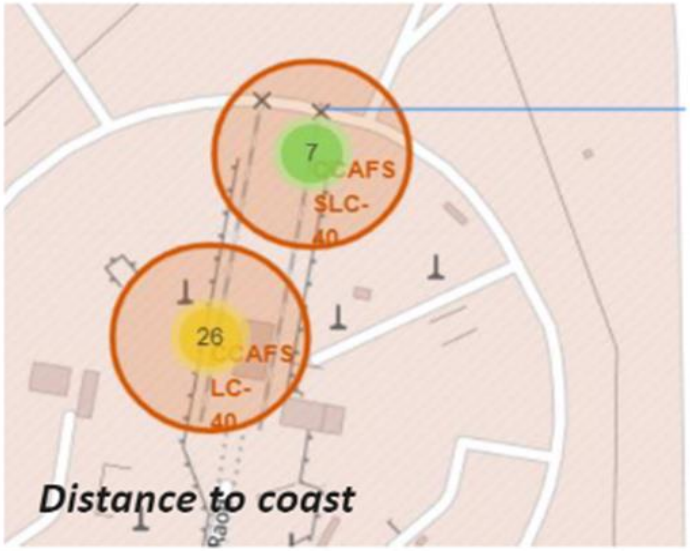


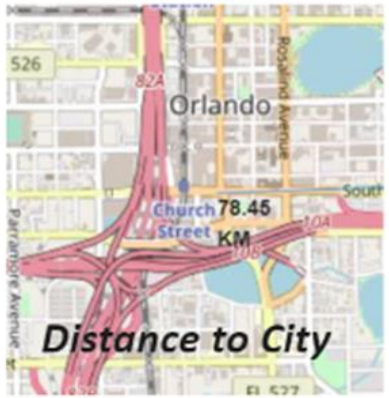
Markers showing launch sites with color labels



Landmarks to LaunchSite Distance



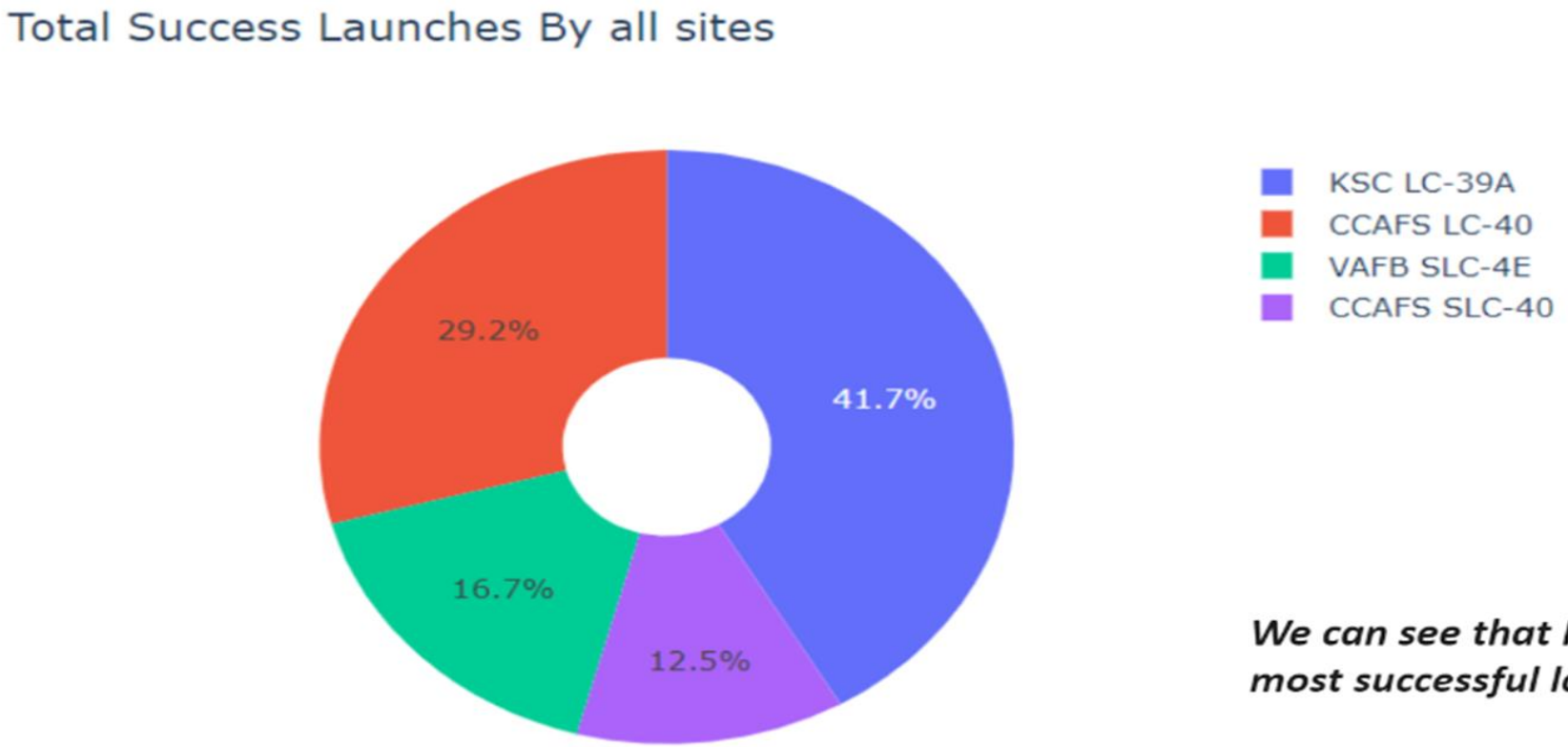




- Are launch sites in close proximity to railways? No
- Are launch sites in close proximity to highways? No
- Are launch sites in close proximity to coastline? Yes
- Do launch sites keep certain distance away from cities? Yes

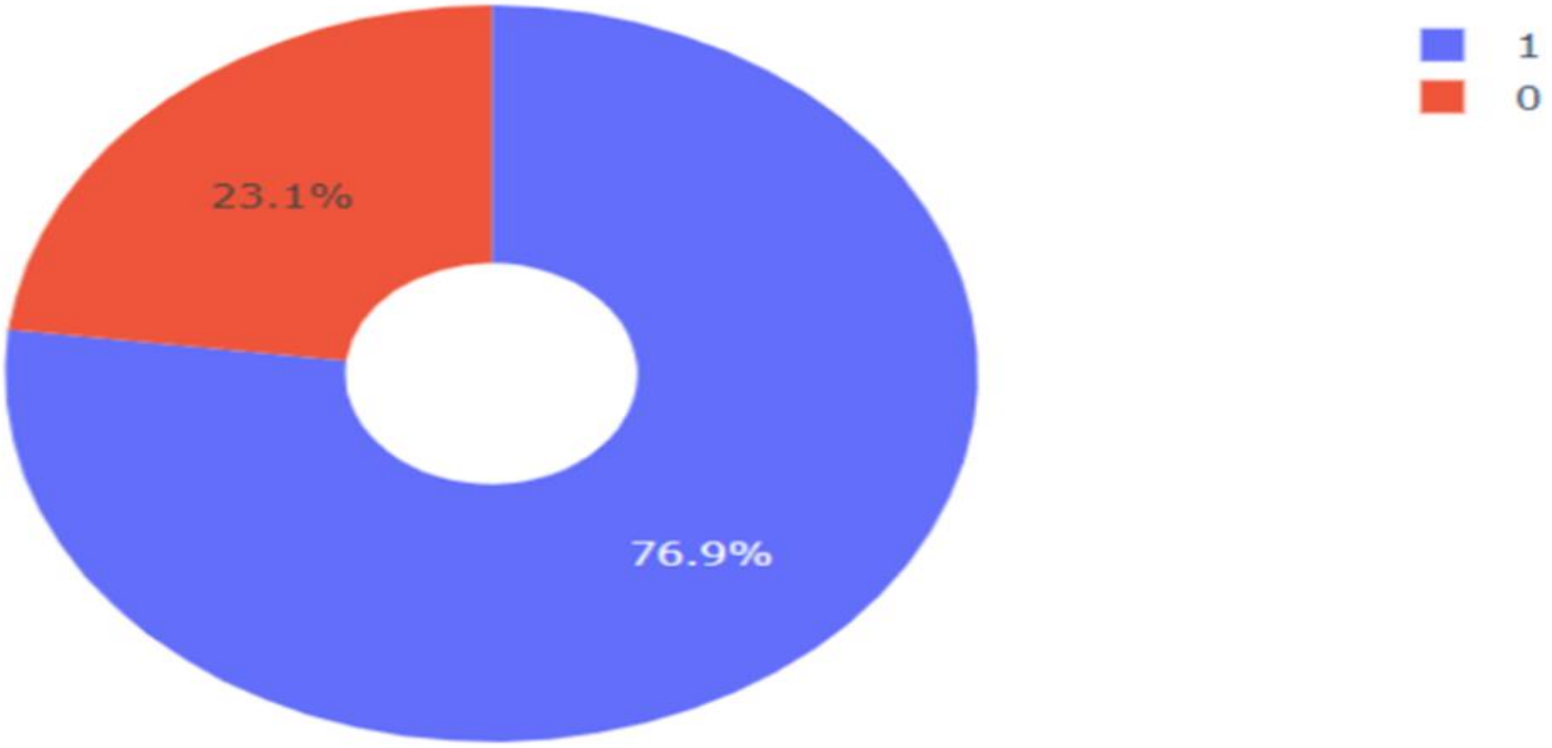

Section 4
DashBoard with Plotly Dash

Pie chart showing the success percentage achieved by each launch site



We can see that KSC LC-39A had the most successful launches from all the sites

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



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

Conclusions

We can conclude that:

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