IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Mohamed Ayaz Nasar
13/01/2024

Outline

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

Executive Summary

- This project focuses on predicting the success of Falcon 9 first stage landings during SpaceX rocket launches.
- The data, collected through a combination of API calls and web scraping, enables us to leverage machine learning for forecasting.
- This predictive capability not only aids in estimating launch costs but also enhances competitiveness for companies entering the commercial space launch sector.

Introduction

- SpaceX's groundbreaking reusability concept for the Falcon 9 first stage has significantly reduced launch costs.
- 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.
- This project aims to leverage machine learning to predict the success of first stage landings, aiding in estimating launch costs and enhancing competitiveness for companies entering the commercial space launch sector.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - The data was gathered using both an API and web scraping techniques.
- Perform data wrangling
  - Data was processed by using pandas library in python.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - To build, tune, evaluate classification models, python library scikit-learn was used.

Data Collection

In this project, the data was gathered from two primary sources: the SpaceX REST API and relevant Wiki pages.

- SpaceX REST API:

Utilized endpoint api.spacexdata.com/v4/launches/past for detailed SpaceX launch information. → Conducted a GET request using Python's requests library, obtaining JSON-format launch data. → Employed json_normalize to convert the JSON into a flat table suitable for Pandas dataframe.

- Web Scraping:

Leveraged web scraping with Python's BeautifulSoup on Falcon 9-related Wiki pages. → Extracted valuable launch records from HTML tables, converting them into a Pandas dataframe for analysis.

Data Collection – SpaceX API

Data collection with SpaceX REST calls

https://github.com/AyazNazar/Applied_Data_Science_Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Requesting rocket launch data from SpaceX API

```
spacex_url="https://api.spacexdata.com/v4/launches/past"
```

```
response = requests.get(spacex_url)
```

Use json_normalize meethod to convert

the json result into a dataframe

```
# Use json_normalize meethod to convert the json result into a dataframe
response_data = response.json()
data = pd.json_normalize(response_data)
```

Data Collection - Scraping

Web scraping process

https://github.com/AyazNazar/Applied_Data_Science_Capstone/blob/main/jupyter-labs-webscraping.ipynb

Assigning response to an object

```
# use requests.get() method with the provided static_url
# assign the response to a object
response = requests.get(static_url)
```

Using BeautifulSoup() to create a

BeautifulSoup object called as soup

```
# Use BeautifulSoup() to create a BeautifulSoup object from a response text content
soup = BeautifulSoup(response.content,  'html.parser')
```

Assigning all the table data to

an object called html_tables

```
# Use the find_all function in the BeautifulSoup object, with element type `table`
# Assign the result to a list called `html_tables`
html_tables = soup.find_all('table')
```

Data Wrangling

Data Processing Steps

Launch Site Categorization:

- Identified and categorized different launch sites, including Vandenberg AFB, Kennedy Space Center, and others.

Orbit Classification:

- Described various payload orbits, such as Low Earth Orbit (LEO) and Geosynchronous Orbit (GTO).

Outcome Conversion to Classes:

- Addressed the 'Outcome' column, which indicates the success of the first stage landing.
- Created a binary classification variable 'Y' where 0 represents a failed landing, and 1 represents a successful landing.
- For instance, 'True ASDS' signifies a successful landing on a drone ship, while 'False ASDS' indicates an unsuccessful landing.

https://github.com/AyazNazar/Applied_Data_Science_Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

Charts plotted

- Scatter plot between flight number vs. payload mass – To identify patterns between fight number vs. payload mass
- Scatter plot between flight number vs launch site – To identify patterns between flight number vs. launch site
- Bar chart - For visualize the success rate of each orbit
- Scatter plot between flight number vs. Orbit – To identify patterns between success flights in different orbits
- Scatter plot between payload mass vs. Orbit – To identify patterns between success flights in different orbits
- A line chart between year and success rate

https://github.com/AyazNazar/Applied_Data_Science_Capstone/blob/main/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with SQL

- Distinct Launch Sites
- Filter Launch Site with 'CCA%'
- Total Payload Mass for NASA (CRS)
- Average Payload Mass for F9 v1.1 Booster Version
- Earliest Success Landing Date
- Booster versions for Successful Landings on Drone Ship with Payload Mass Between 4000 and 6000
- Count of Each Landing Outcome
- Booster Version with Maximum Payload Mass
- Launch Details for June 2015
- Count of Each Landing Outcome Between 2010-06-04 and 2017-03-20

https://github.com/AyazNazar/Applied_Data_Science_Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Markers used for mark the locations of launch sites.
- Circles are used to mark the proximity of launch sites.
- Lines are used to measure distances between two points.
- Use clusters for display launch sites located nearby as a cluster.

https://github.com/AyazNazar/Applied_Data_Science_Capstone/blob/main/lab_jupyter_launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

- Two graphs, a pie chart and a scatter plot added to the dashboard.
- A dropdown list is addedd to enable launch site selection.
- A slider is added to select payload range.

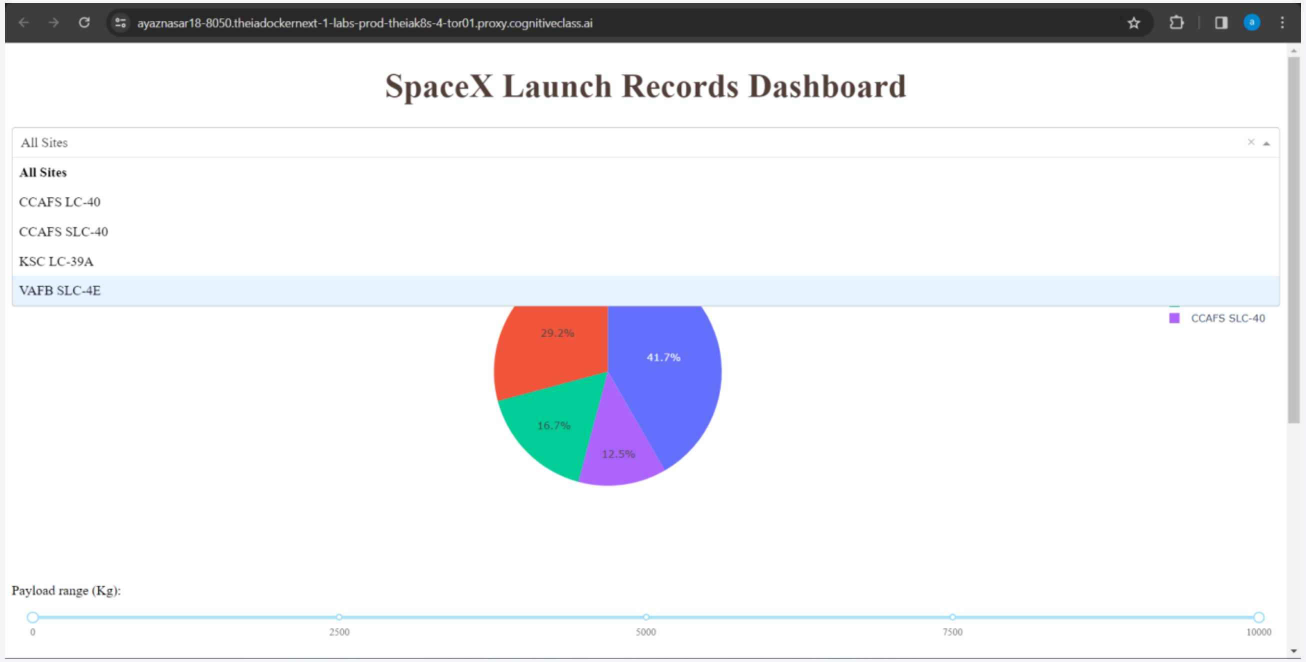
https://github.com/AyazNazar/Applied_Data_Science_Capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- Standardize input features for consistent scales.
- Split the data into training and test Data
- Identify optimal parameters for SVM, Classification Trees, and Logistic Regression.
- Train and evaluate models using test data.
- Choose the best-performing method based on metrics.

https://github.com/AyazNazar/Applied_Data_Science_Capstone/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

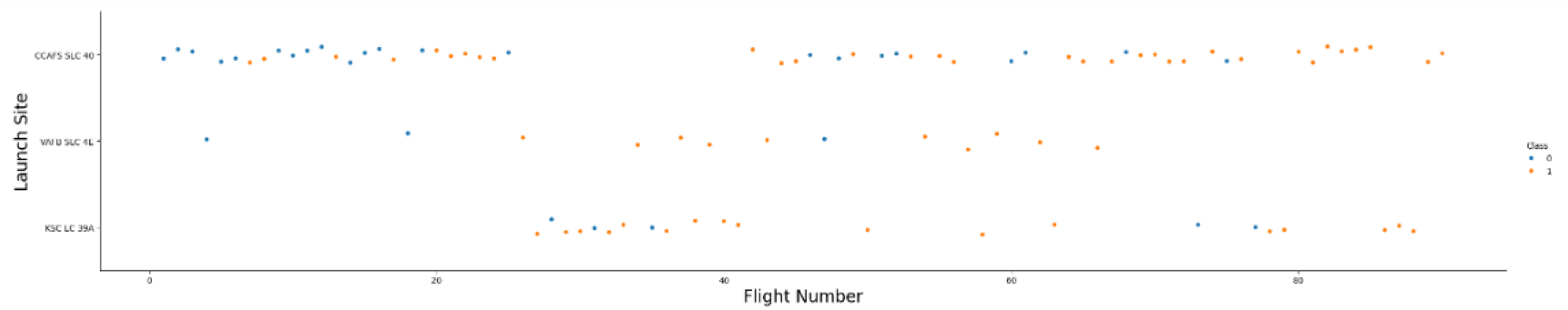



```
Best Model: DecisionTreeClassifier(criterion='entropy', max_depth=2, max_features='sqrt',
                       min_samples_leaf=4)
```

Section 2

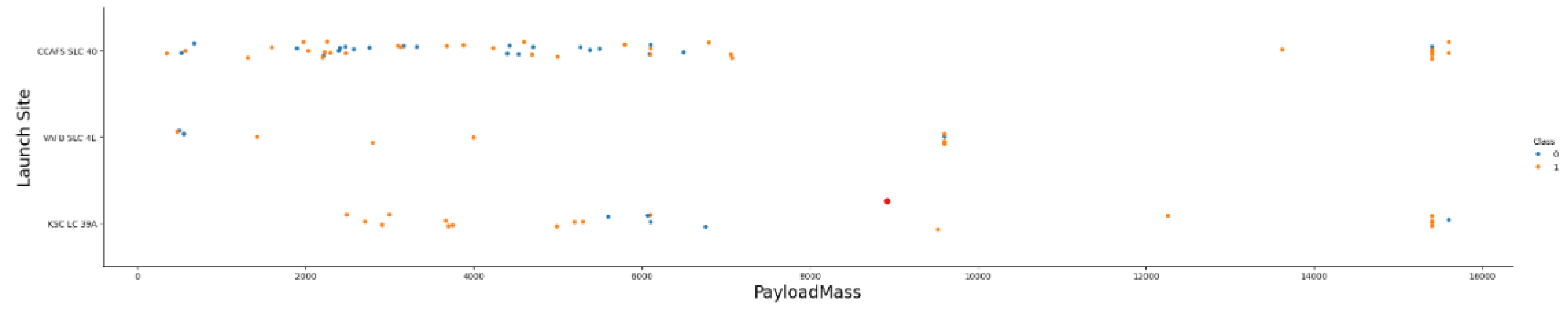
Insights drawn from EDA

Flight Number vs. Launch Site



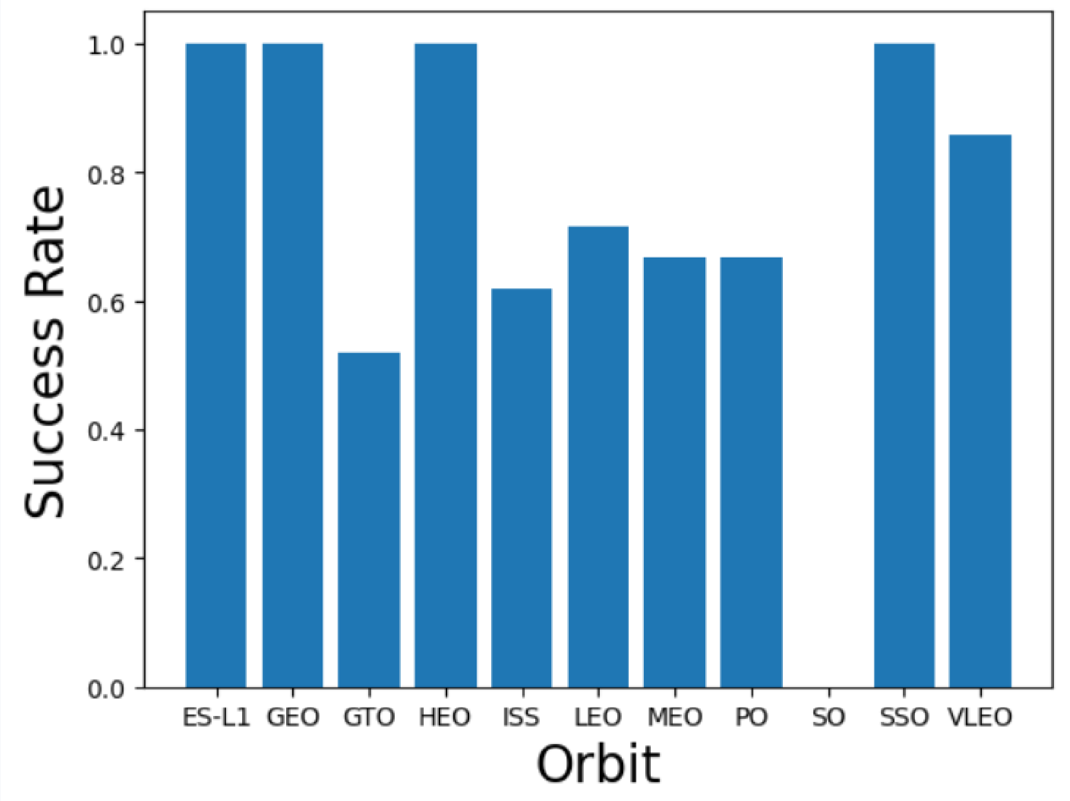
For the recent launches the launch site VAFB SLC 4E is not used.

Payload vs. Launch Site



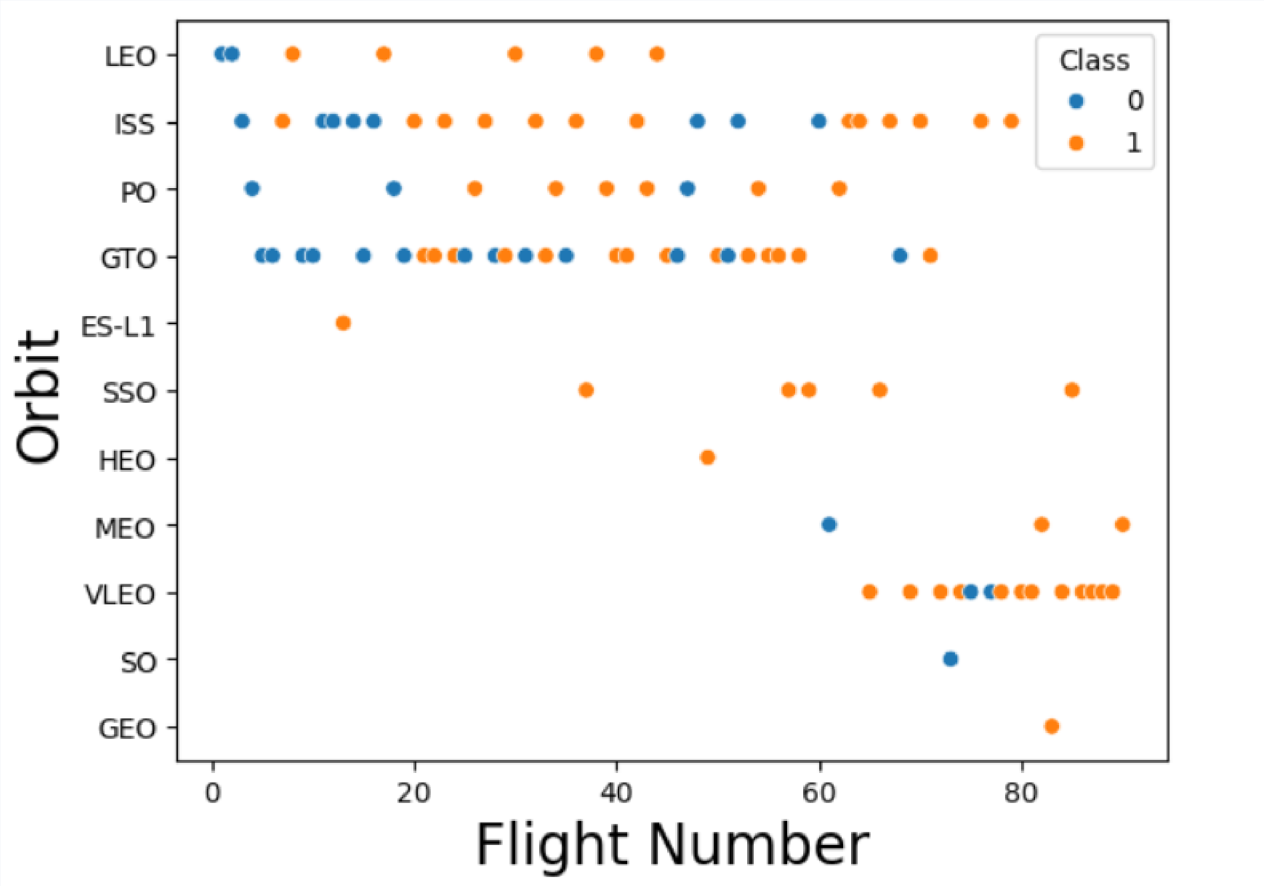
The VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

Success Rate vs. Orbit Type

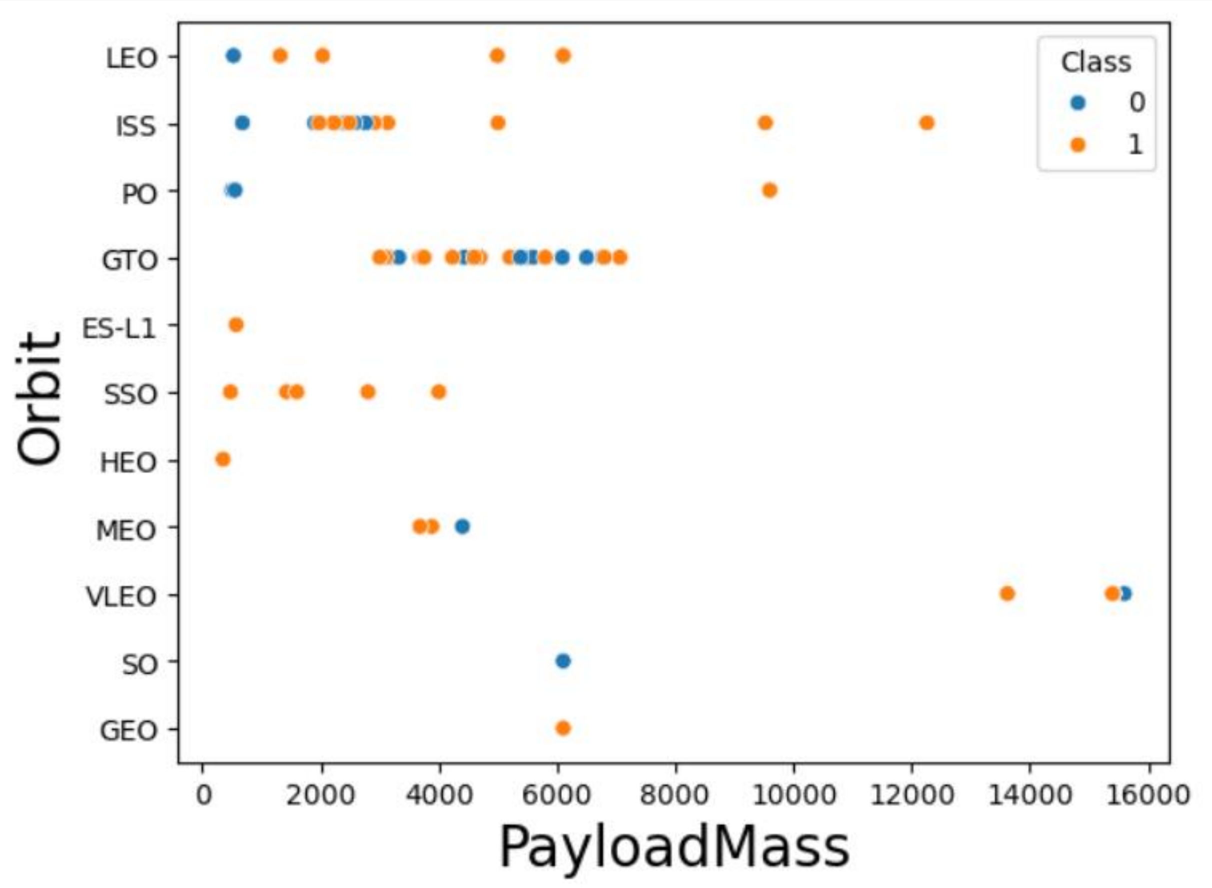


The launches to orbits ES-L1, GEO, HEO and SSO are all success and launches to orbit SO are all failed.

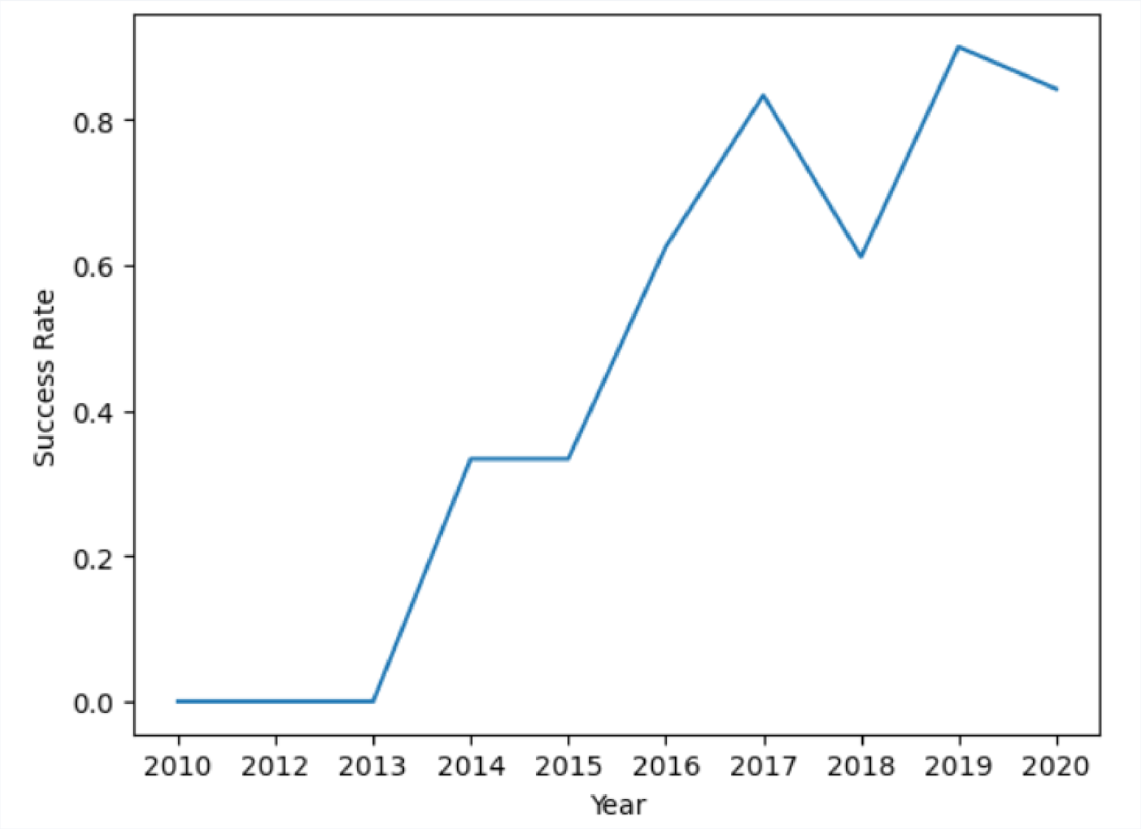
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



The success rate since 2013 kept increasing till 2020

All Launch Site Names

Name of the SpaceX launch sites

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`

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

Payload masses carried by boosters from NASA

sum("PAYLOAD_MASS__KG_")
45596

Average Payload Mass by F9 v1.1

Average payload mass carried by booster version F9 v1.1

avg("PAYLOAD_MASS__KG_")
2534.6666666666665

First Successful Ground Landing Date

Dates of the first successful landing outcome on ground pad

min("Date")
2018-07-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

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

Total Number of Successful and Failure Mission Outcomes

The total number of successful and failure mission outcomes

Landing_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

Names of the booster which have carried the maximum payload mass

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

Failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

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

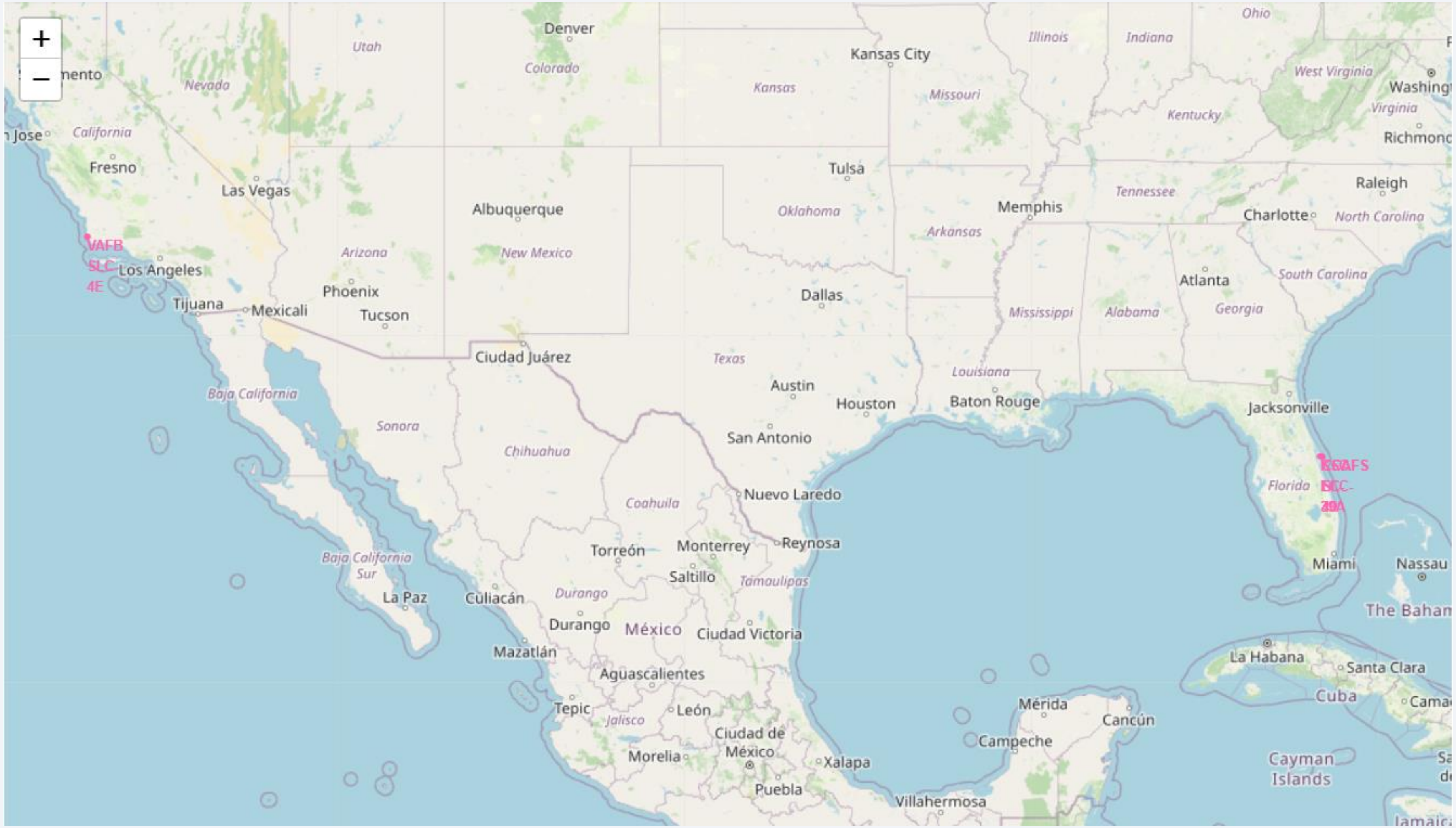
Count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

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

Section 3

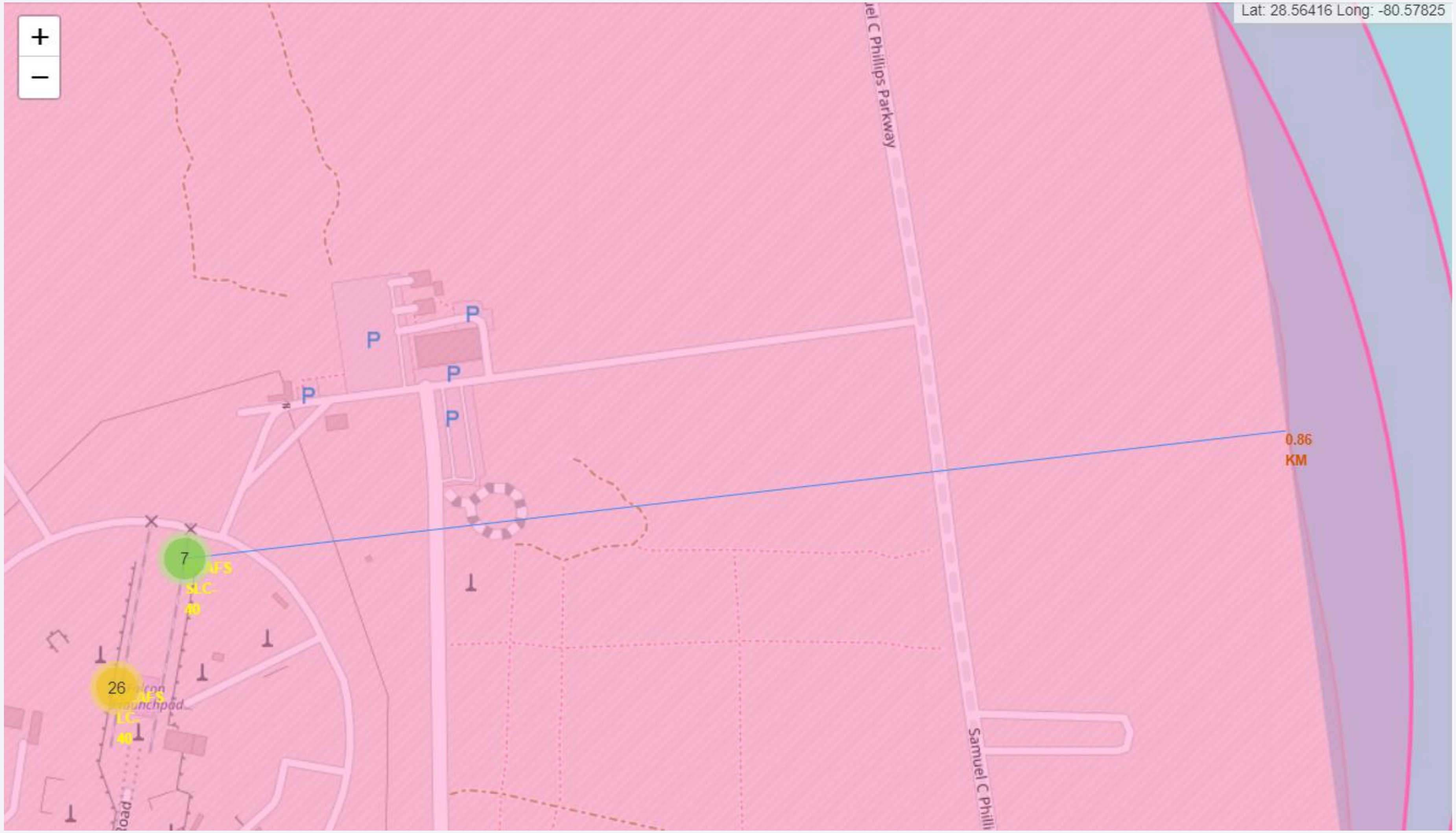
Launch Sites Proximities Analysis

Launch sites' location markers on a global map



Marker cluster to current site map

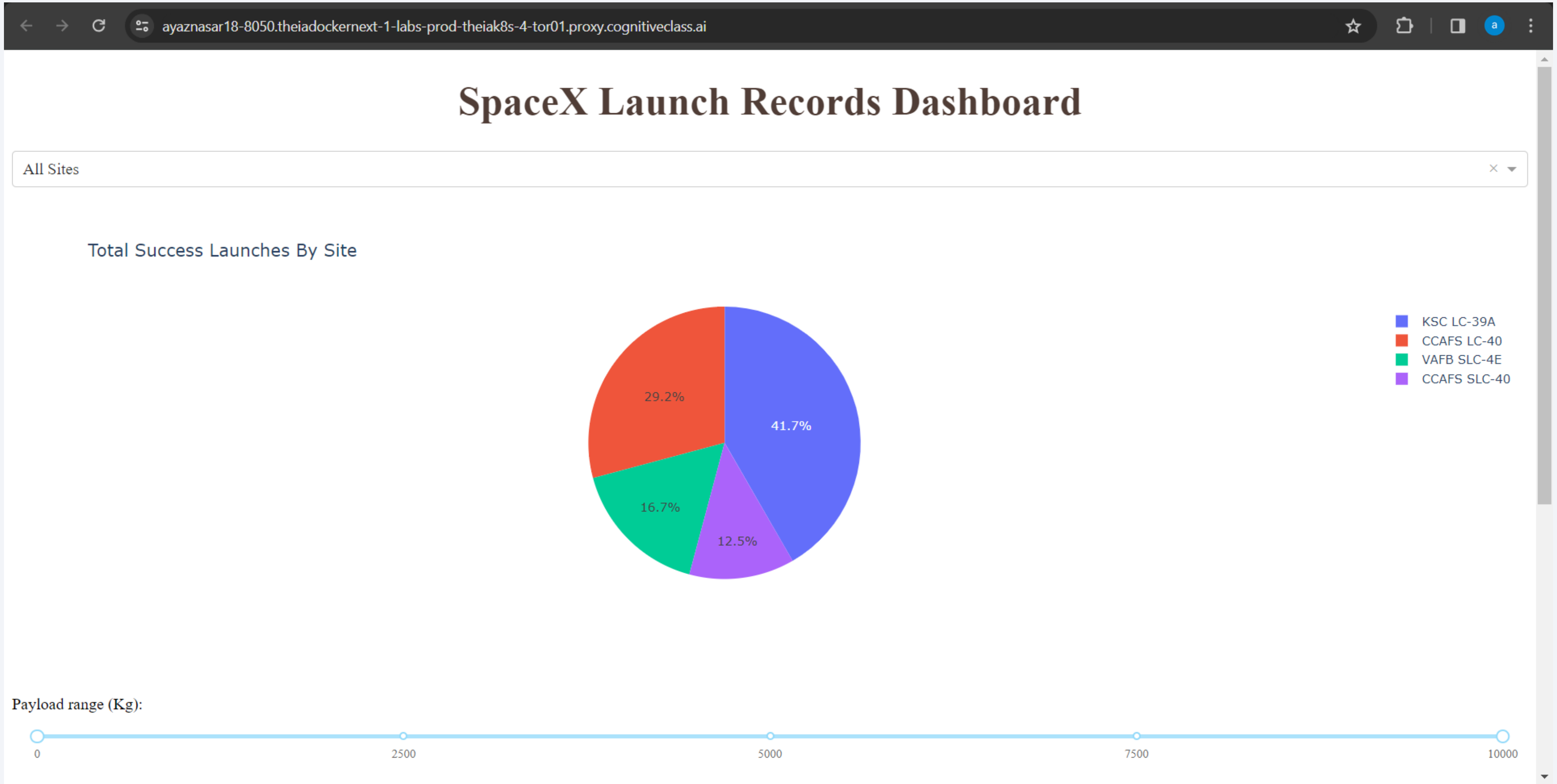
Closest coastline point on the map



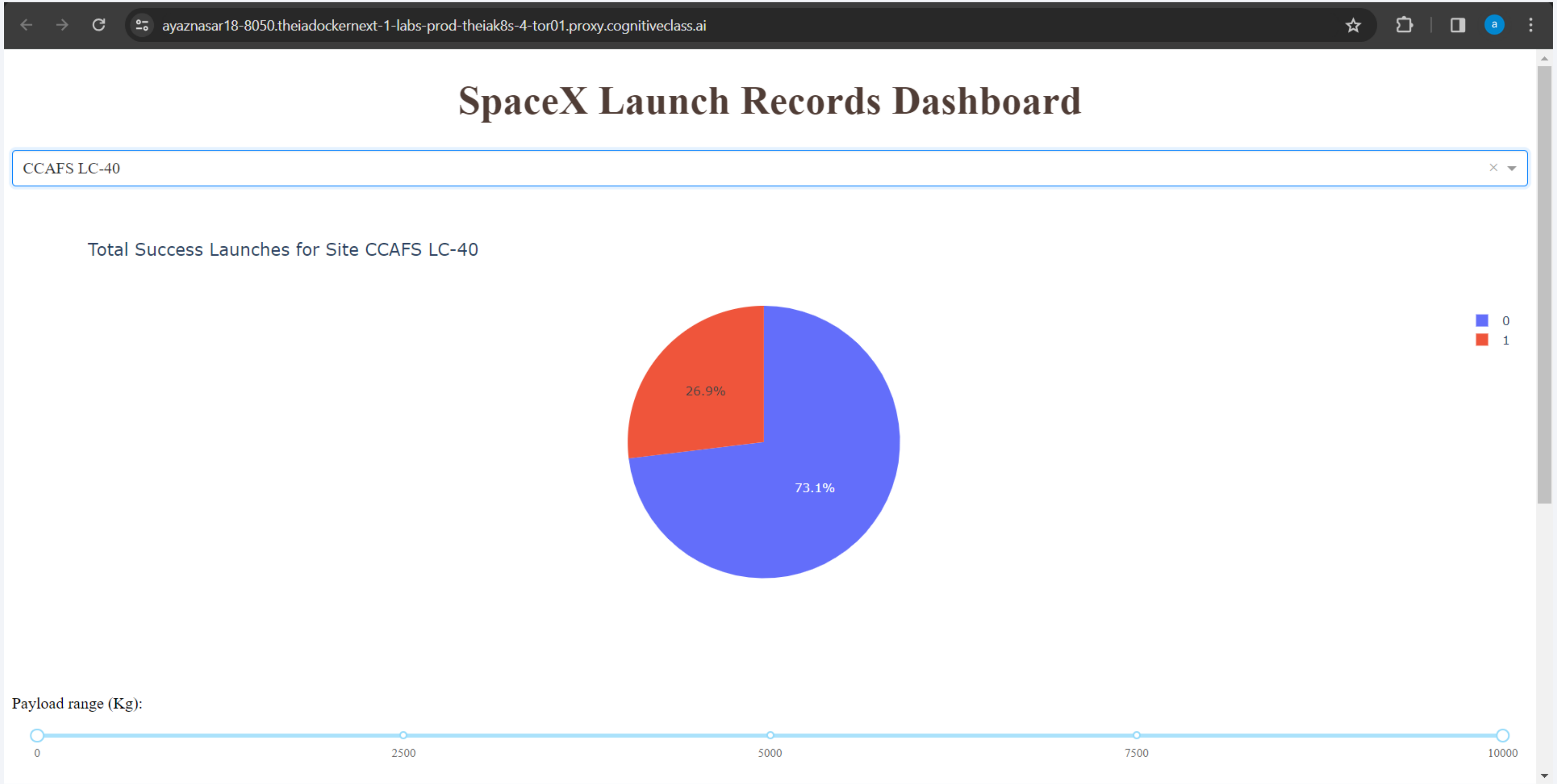
Section 4

Build a Dashboard with Plotly Dash

Total success launches by sites



Total success launches for site CCAFS LC-40

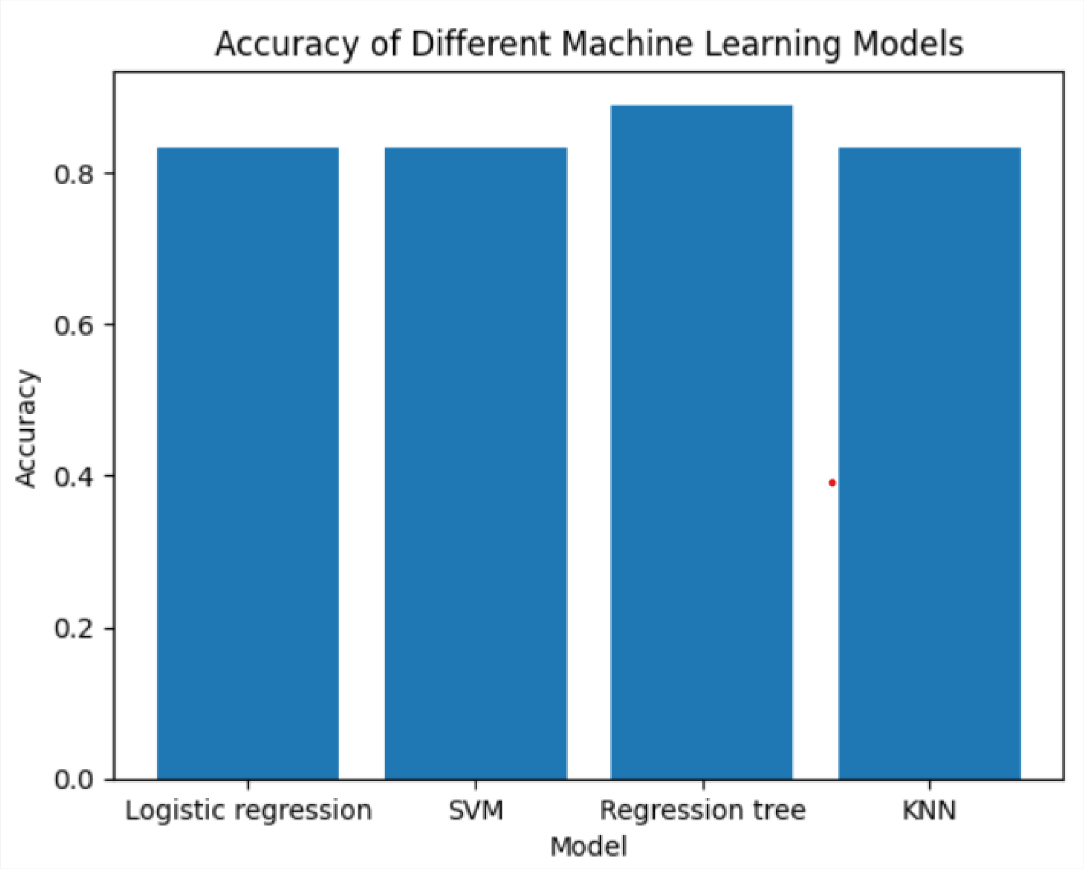


Correlation between payload and success in site CCAFS LC-40

Section 5

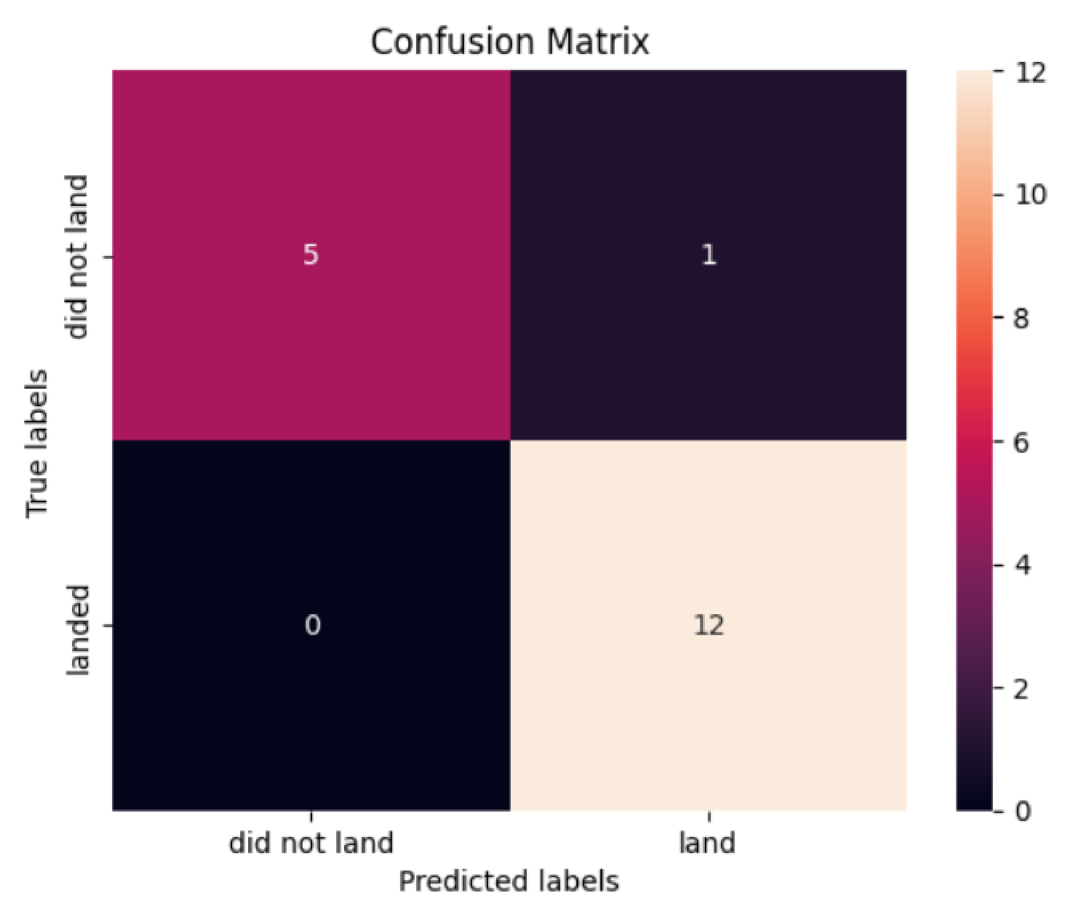
Predictive Analysis (Classification)

Classification Accuracy



Regression tree model has the highest accuracy of 0.94

Confusion Matrix



Regression tree model only has one false positive.

Conclusions

- 'FlightNumber', 'PayloadMass', 'Orbit', 'LaunchSite', 'Flights', 'GridFins', 'Reused', 'Legs', 'LandingPad, ‘Block’, ‘ReusedCount’ and ‘Seriel’ are the important independent variables for predicting the dependent variable ‘Class’(success of first stage landings).
- Regression tree is the best method for predict success of first stage landings.

Thank you!