IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Abdul Qudus
2024-1-11

Outline

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

Executive Summary

- **Summary of methodologies**

❖ Data collection through API and web-scrapping

❖ Data wrangling

❖ EDA with SQL

❖ EDA with data visualization

❖ Interactive Visual Analytics with Folium

❖ Machine Learning Prediction

- **Summary of all results**

❖ EDA Results

❖ Interactive Analytics Result

❖ Predictive Analytics Result

Introduction

SpaceX is a most successful rocket launcher. We will predict if the Falcon 9 first stage will land successfully. 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.

Objective:

- To determine the price of each launch.
- To gather information about SpaceX and creating a dashboard.
- To determine if SpaceX will reuse the first stage.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Request using SpaceX API and web scrapping from Wikipedia
- Perform data wrangling
  - Perform some Exploratory Data Analysis (EDA) to find some patterns in the data and determine what would be the label for training supervised
- 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

1. Import Libraries and Define Auxiliary Functions
2. Request and parse the SpaceX launch data using the GET request
3. Filter the dataframe to only include Falcon 9 launches
4. Dealing with missing value

Data Collection – SpaceX API

Import requests, pandas, numpy, datetime

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

To only include Falcon 9 launches

https://github.com/sorayafavorite/IBM-Applied-Data-Science-Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Import libraries

Request and parse data using GET request
response = requests.get(spacex_url)

Filter dataframe

Dealing with missing value

Data Collection - Scraping

Web scrap Falcon 9 launch records with BeautifulSoup:

- Extract a Falcon 9 launch records HTML table from Wikipedia

- Parse the table and convert it into a Pandas data frame

https://github.com/sorayafavorite/IBM-Applied-Data-Science-Capstone/blob/main/jupyter-labs-webscraping.ipynb

Import libraries

Extract from Wikipedia

Convert into dataframe

Data Wrangling

Analysis on the data using method value_counts() to:

1. Calculate the number of launches on each site
2. Calculate the number and occurrence of each orbit
3. Calculate the number and occurence of mission outcome of the orbits
4. Create a landing outcome label from Outcome column

https://github.com/sorayafavorite/IBM-Applied-Data-Science-Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

Charts were plotted to visualize the relationship between:

1. Flight Number and Launch Site
2. Payload and Launch Site
3. success rate of each orbit type
4. FlightNumber and Orbit type
5. Payload and Orbit type
6. the launch success yearly trend

https://github.com/sorayafavorite/IBM-Applied-Data-Science-Capstone/blob/main/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- 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 succesful landing outcome in ground pad was acheived.
- 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.
- 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.

https://github.com/sorayafavorite/IBM-Applied-Data-Science-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite_v1.ipynb

Build an Interactive Map with Folium

- **Mark all launch sites on a map**
- **Mark the success/failed launches for each site on the map**
- **Calculate the distances between a launch site to its proximities**

https://github.com/sorayafavorite/IBM-Applied-Data-Science-Capstone/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Predictive Analysis (Classification)

Perform exploratory Data Analysis and determine Training Labels

- create a column for the class
- Standardize the data
- Split into training data and test data

To find best Hyperparameter for SVM, Classification Trees and Logistic Regression

- https://github.com/sorayafavorite/IBM-Applied-Data-Science-Capstone/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

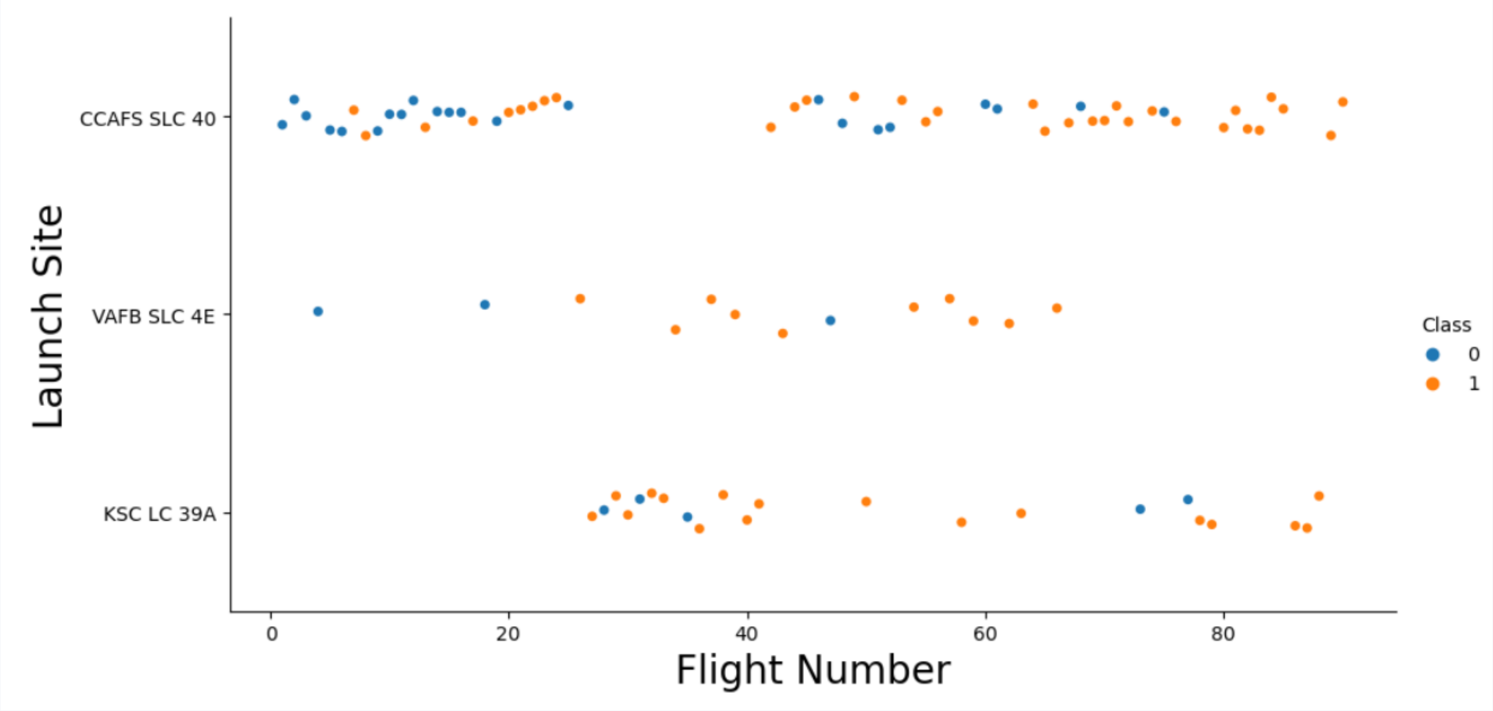
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

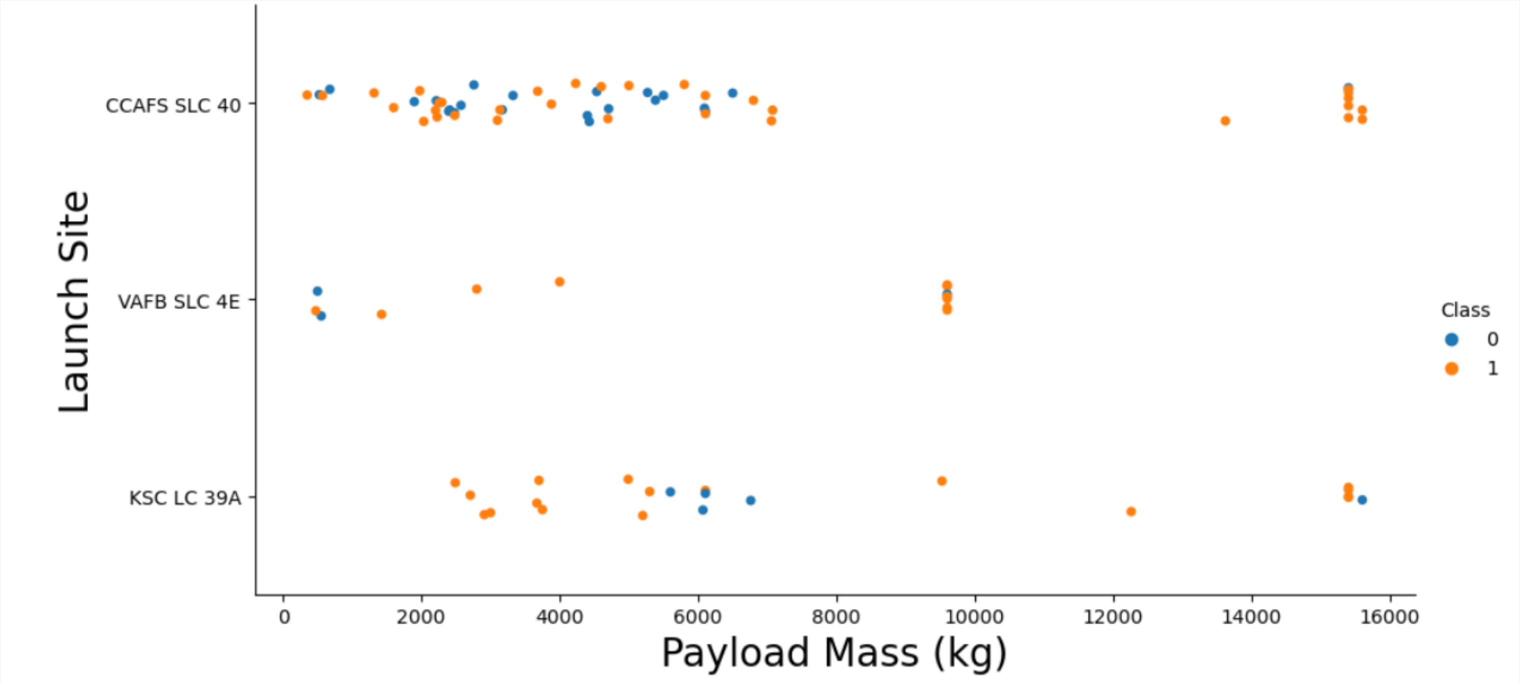
Flight Number vs. Launch Site

- As the flight number increases, the first stage is more likely to land successfully
- Different launch sites have different success rates



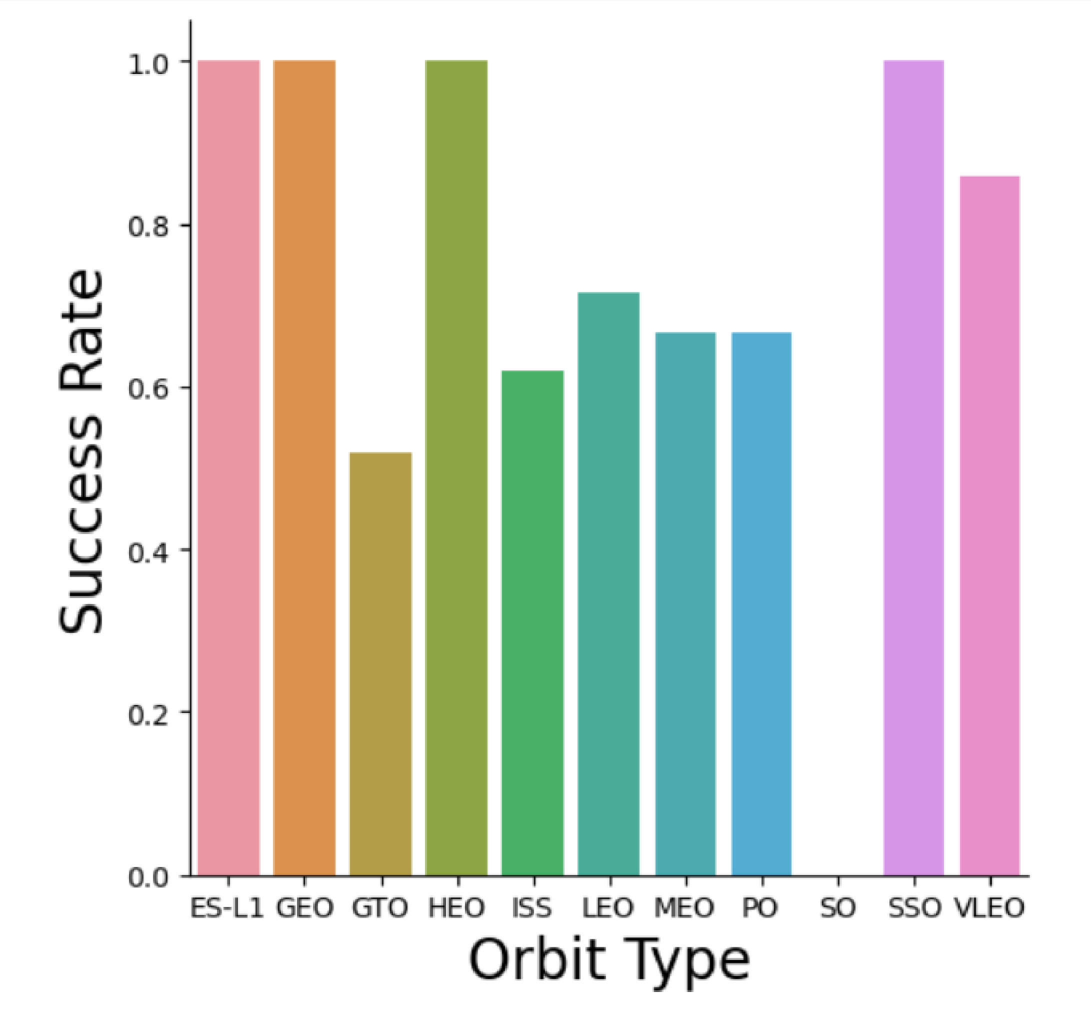
Payload vs. Launch Site

- It seems the more massive the payload, the less likely the first stage will return
- No rocket launch for heavy payload at VAFB SLC 4E



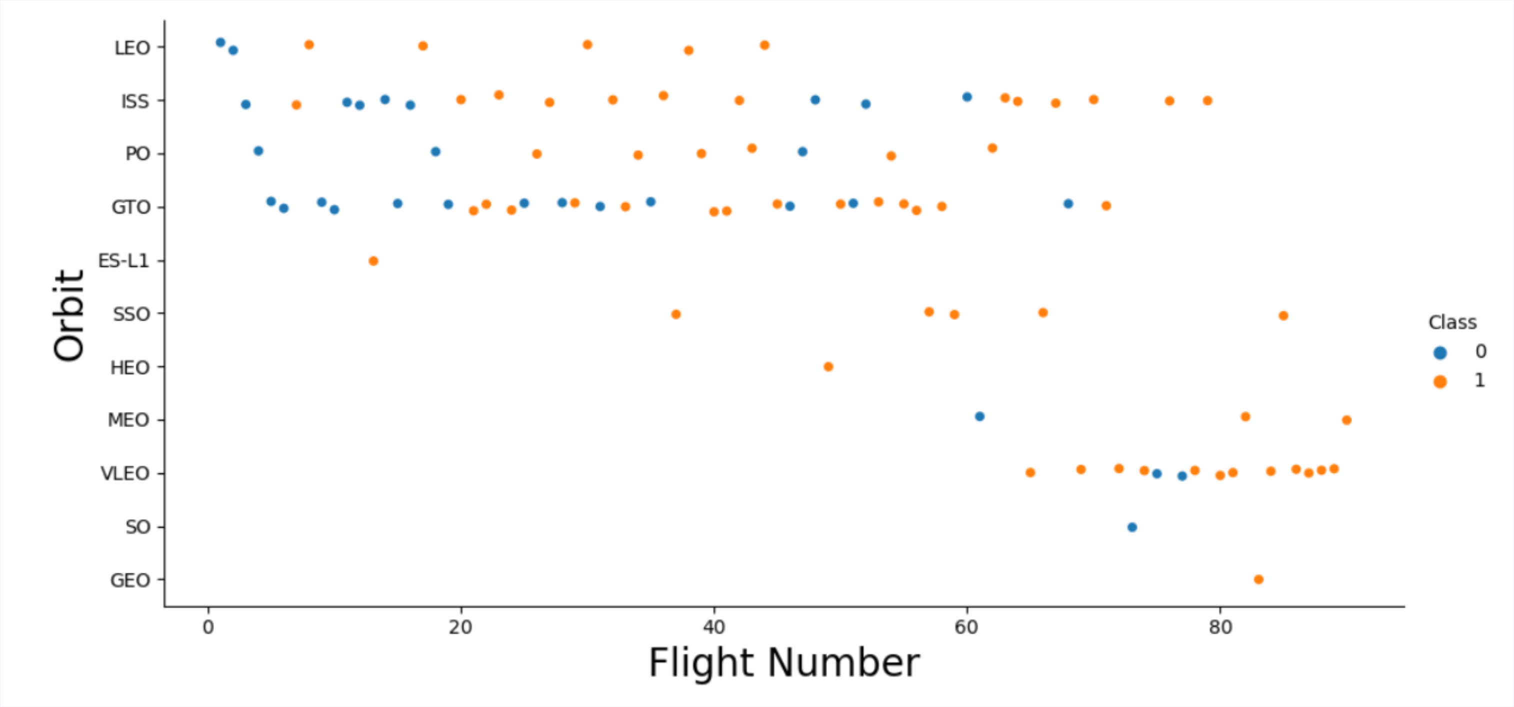
Success Rate vs. Orbit Type

- ES-L1, GEO, HEO and SSO have high successful rate for rocket launch.



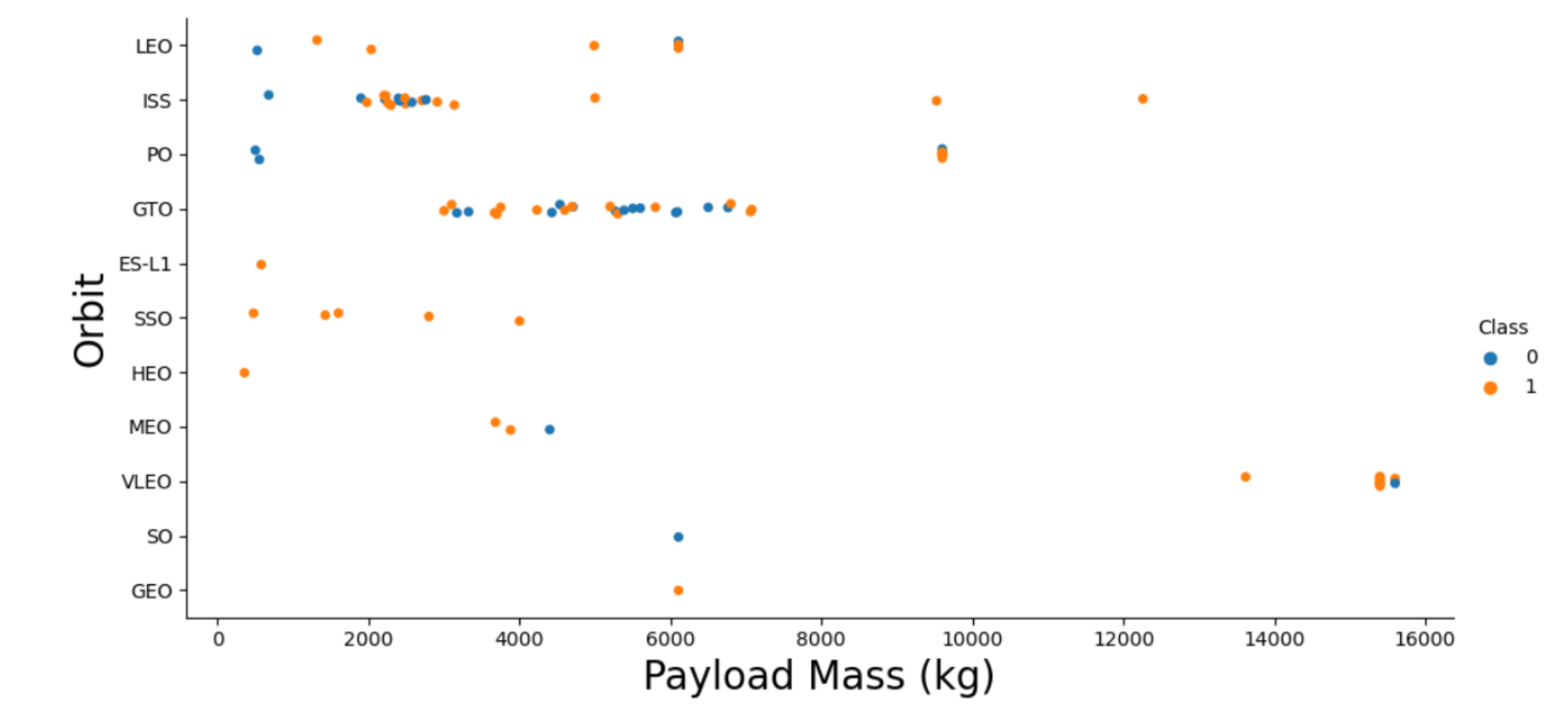
Flight Number vs. Orbit Type

- LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.



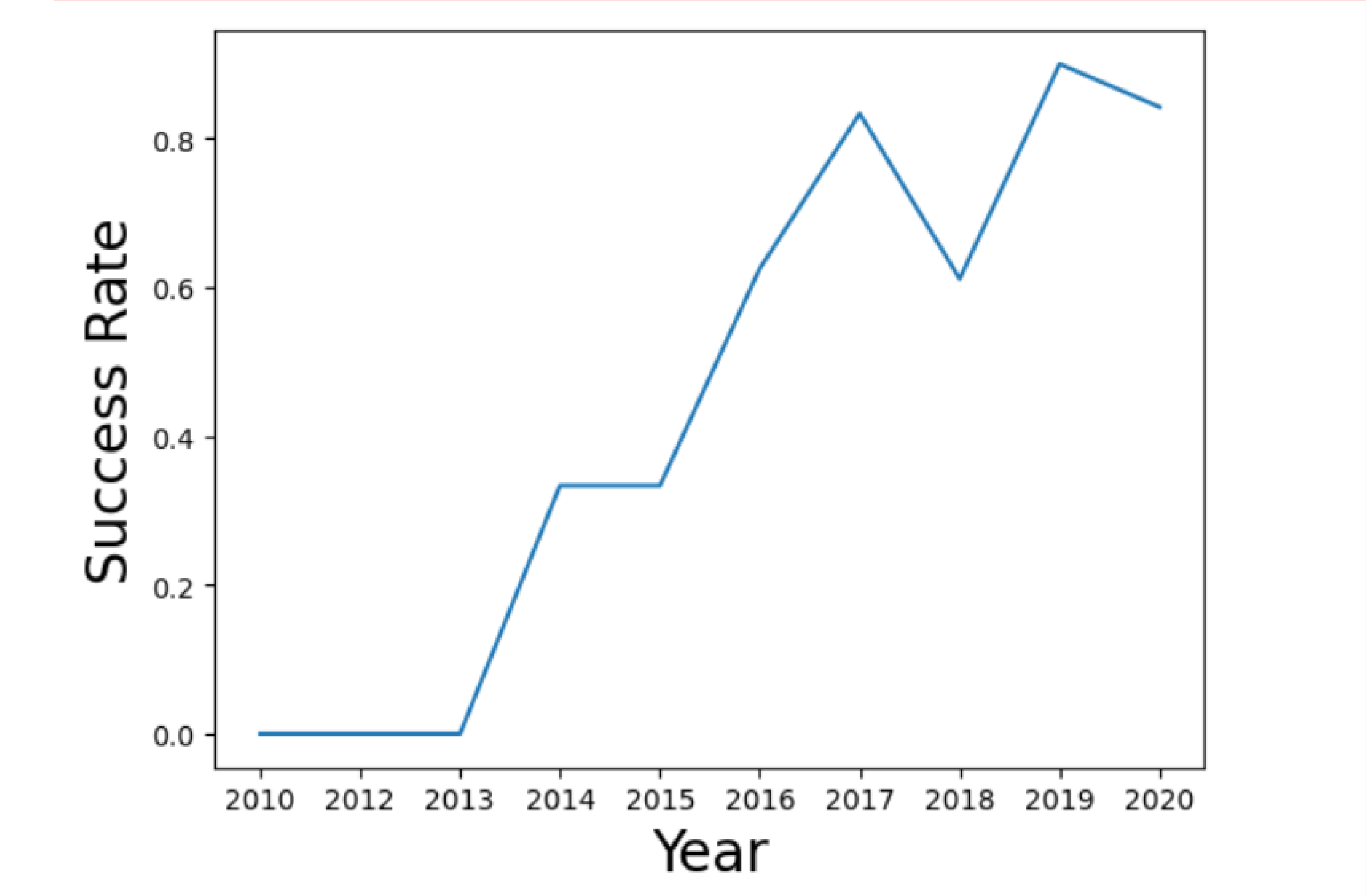
Payload vs. Orbit Type

- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.
- However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.



Launch Success Yearly Trend

- The success rate since 2013 kept increasing till 2020



All Launch Site Names

- Launch site query from the table

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`

Out[13]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	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 payload carried by boosters from NASA

Out[14]:

total_payload_mass
45596

Average Payload Mass by F9 v1.1

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

Out[15]:

average_payload_mass
2534.6666666666665

First Successful Ground Landing Date

- The dates of the first successful landing outcome on ground pad

Out[17]:

first_successful_landing
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

Out[21]:

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

Out[22]:

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

Boosters Carried Maximum Payload

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

Out[26]:

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

Out[70]:

Month	Date	Booster_Version	Launch_Site	Landing_Outcome
10	2015-10-01	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

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

Out[71]:

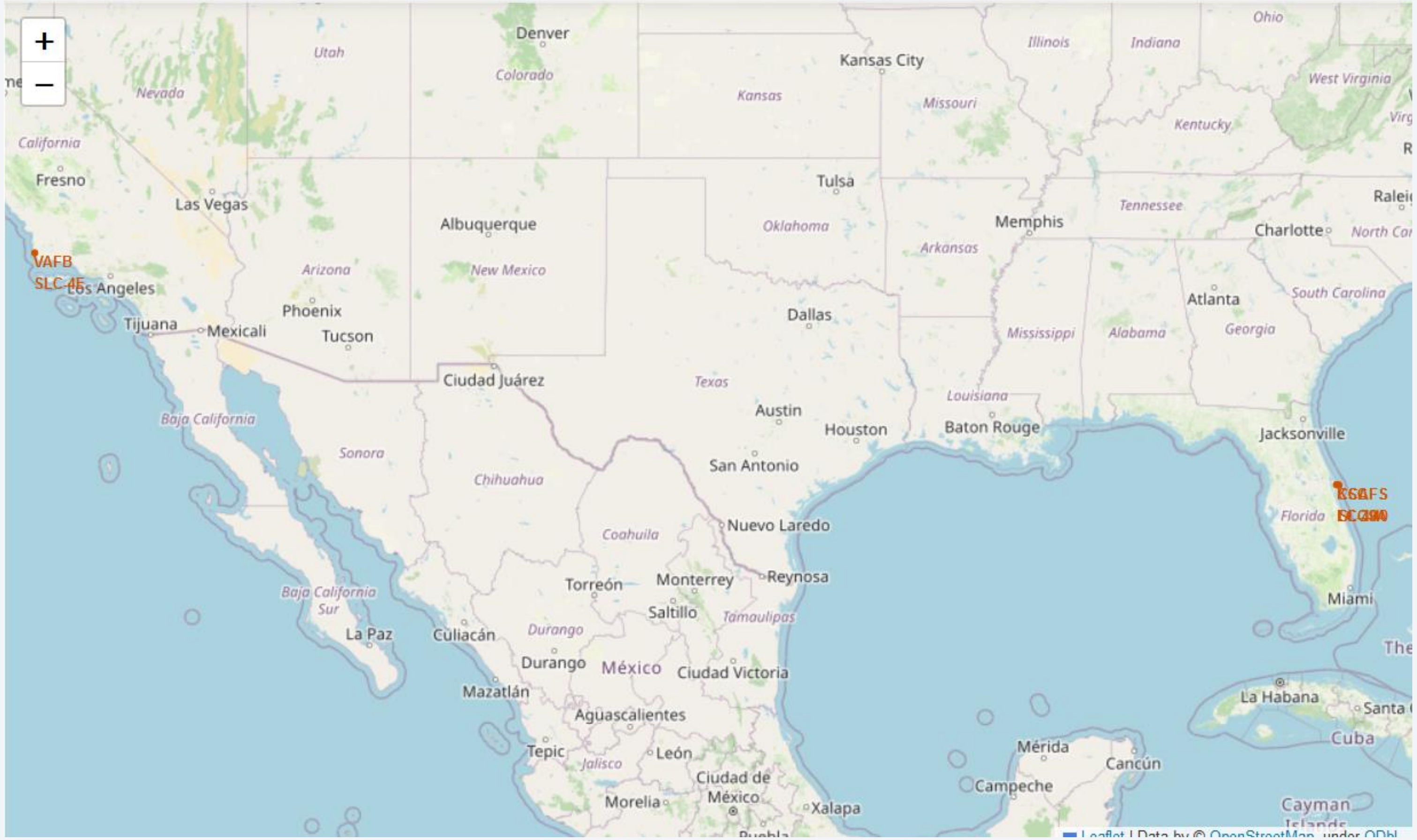
Landing_Outcome	count_outcomes
No attempt	10
Success (ground pad)	5
Success (drone ship)	5
Failure (drone ship)	5
Controlled (ocean)	3
Uncontrolled (ocean)	2
Precluded (drone ship)	1
Failure (parachute)	1

Section 3

Launch Sites Proximities Analysis

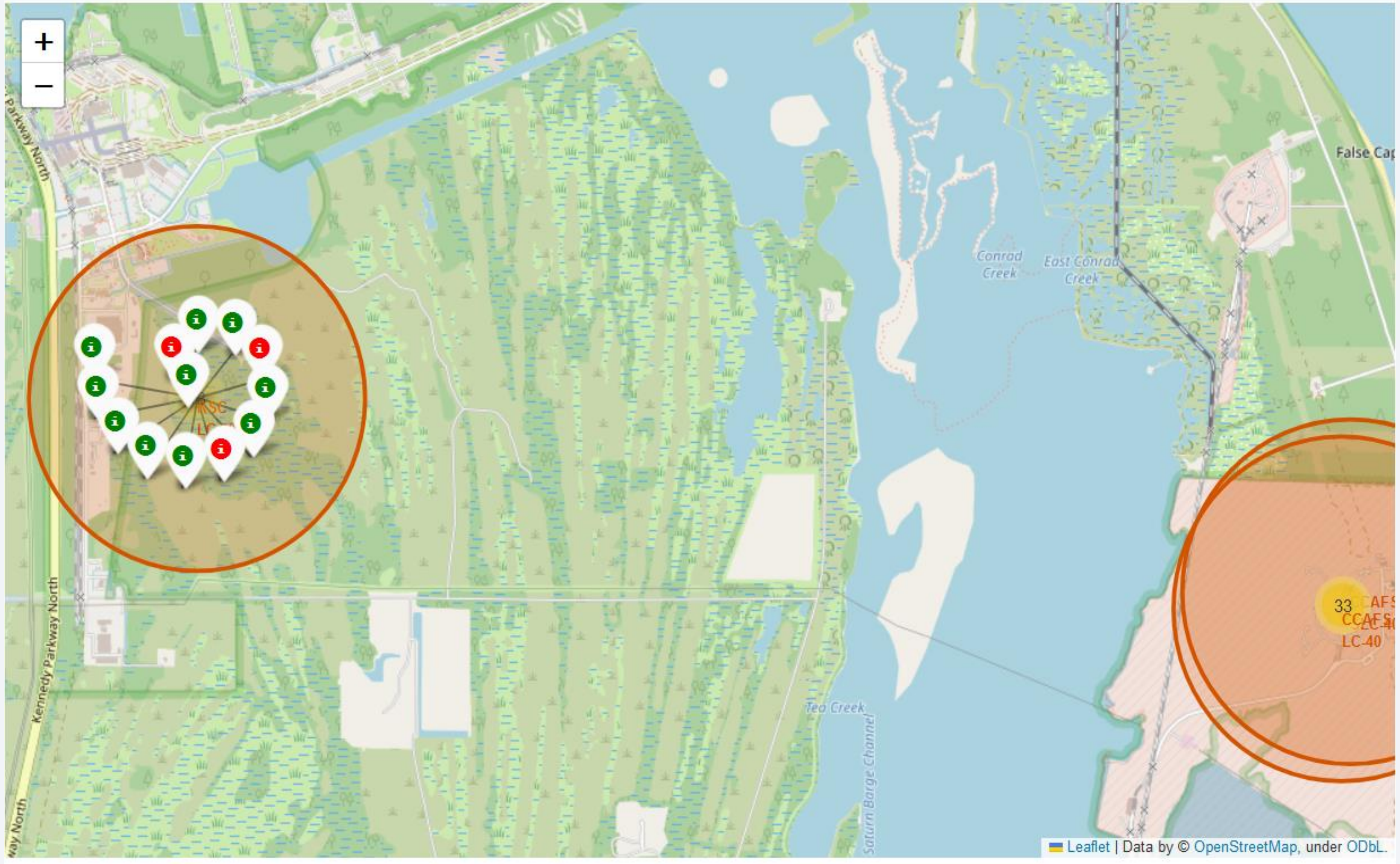
Location of the Launch Sites

- All launch sites in proximity in Equatorial line
- All launch sites in very close proximity to the coast



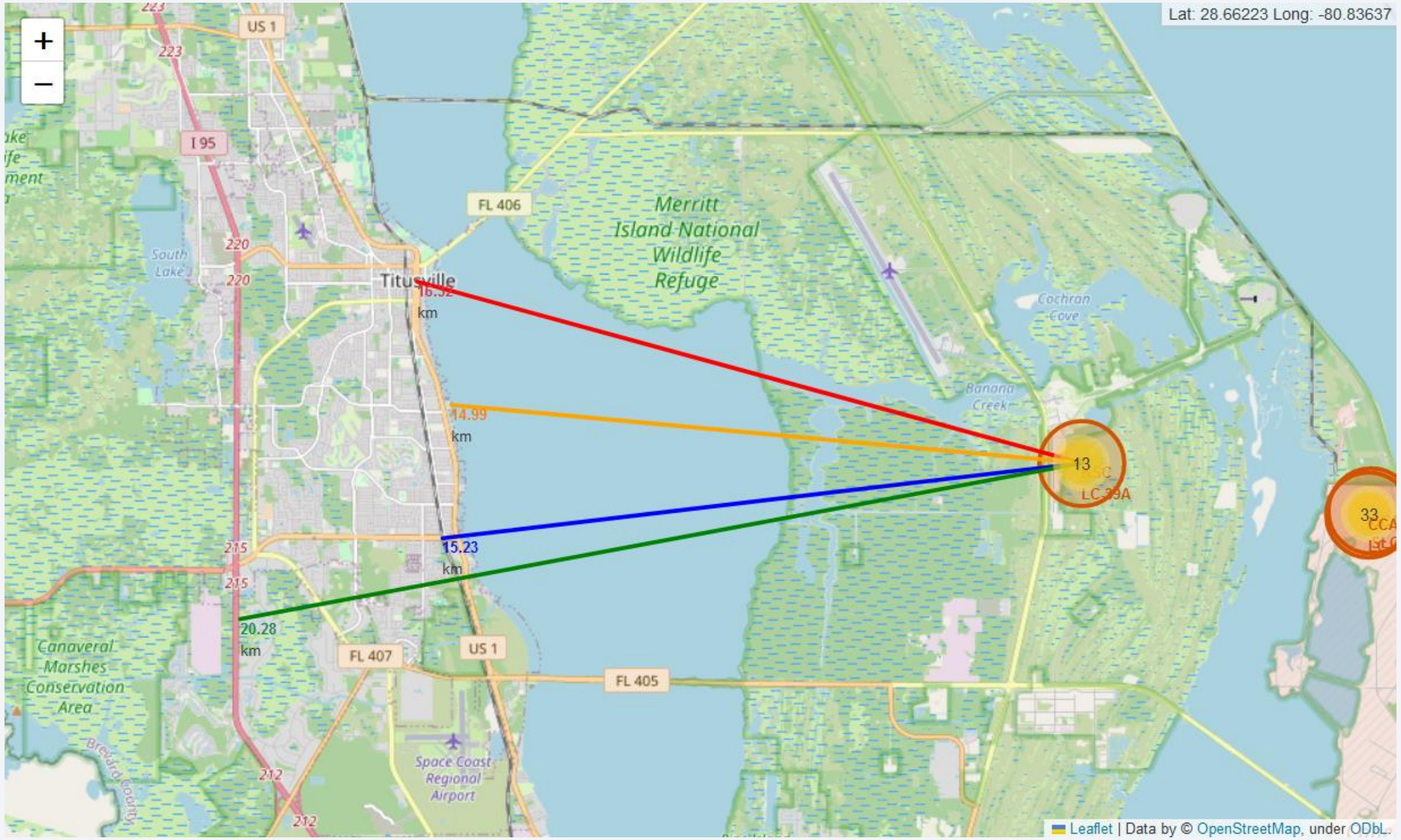
Indicators of Launch Sites

- Green marker indicates successful launch and if else is red
- The red circle shows the launch site location



Proximity in Launch Sites

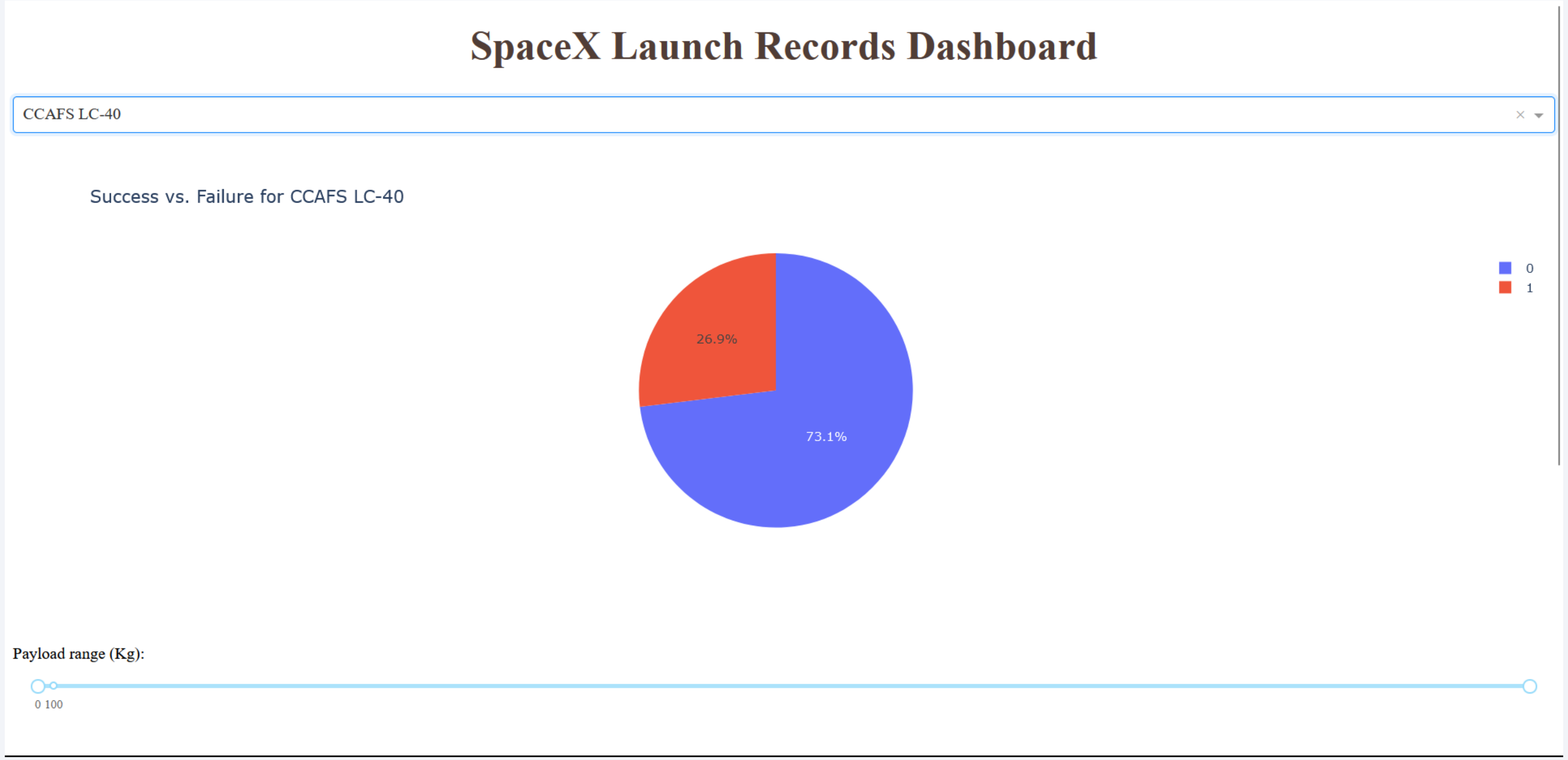
- Nearest city is 16.32 km from KSC LC-39A launch site.
- Nearest coastline is 14.99 km from KSC LC-39A launch site.
- Nearest highway is 20.28 km from KSC LC-39A launch site.
- Nearest railway is 15.23 km from KSC LC-39A launch site.



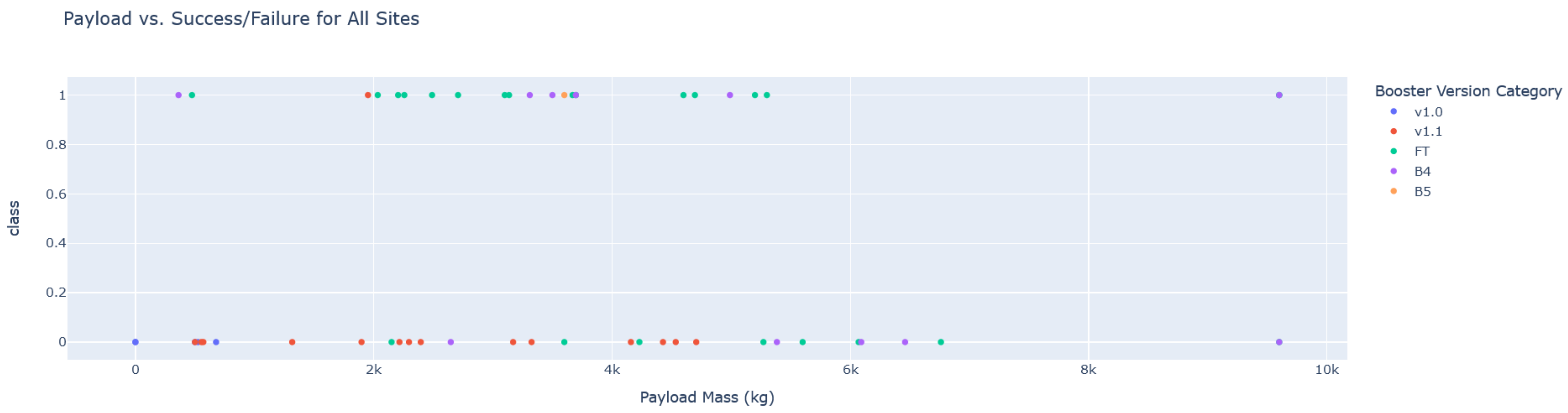
Section 4

Build a Dashboard with Plotly Dash

Successful Rate of CCAFS LC40



Payload Mass by Booster Version Category



Section 5

Predictive Analysis (Classification)

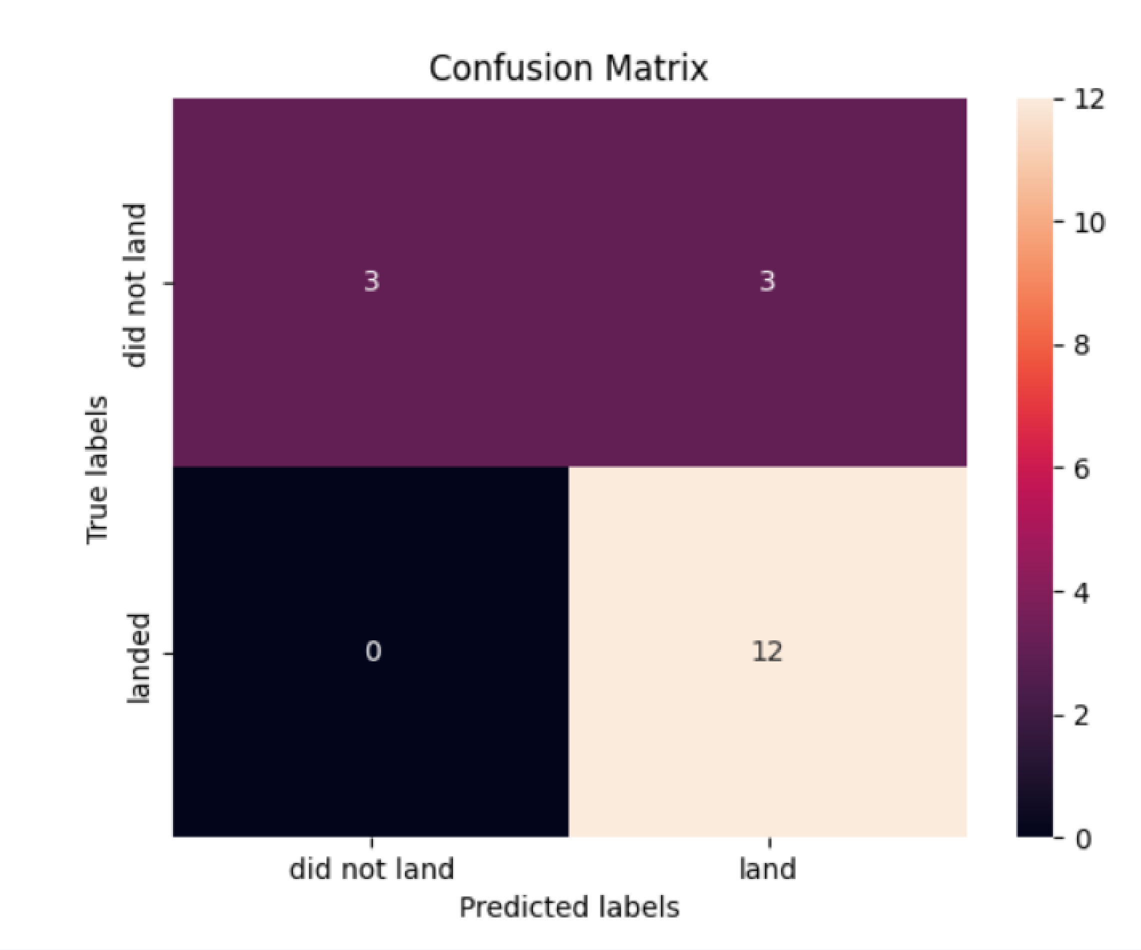
Classification Accuracy

- All classification model accuracy is 0.833333.

	Method	Logistic_Reg	SVM	Decision Tree	KNN
0	Test Data Accuracy	0.833333	0.833333	0.833333	0.833333

Confusion Matrix

- The decision tree model is selected.
- Based on the confusion matrix, the model predicted 2 successful and 3 unsuccessful landed
- False positive predicted 3 unsuccessful landed.



Conclusions

- All classification model accuracy is 0.833333.
- As the flight number increases, the first stage is more likely to land successfully. The success rate since 2013 kept increasing till 2020.
- The more massive the payload, the less likely the first stage will return. With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.
- ES-L1, GEO, HEO and SSO have high successful rate for rocket launch.

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!