IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Facundo Antunez
19-06-2024

Outline

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

Executive Summary

- In order to complete this Data Science project, I needed to implement the hole of the data science proccedure. To do this, I started with Data collection, where I needed to use both web scrapping and APIs. After that, I cleaned the data for it to be ready to be studied. The logic step after this is to do the Exploratory Data Analysis (EDA), using Pandas, SQL and Matplotlib. In addition to that, I created an interactive dashboard and maps with Plotly Dash and Folium, respectively, where I was able to analyze launch records wth more detail.
- The final step was to use Machine Learning algorithms in order to perform the predictive analysis, where I implemented Logistic Regression, SVM, Decision Tree and KNN neighbors. Surprisingly, all of them are equally good to predict results.

Introduction

- Project Background and Context:
  - In the competitive space industry, SpaceX's reusable Falcon 9 rockets have reshaped launch economics. Our task at Space Y is to predict whether SpaceX will successfully recover the first stage post-launch. This prediction is crucial for pricing our launches competitively and planning operations effectively.
- Problems You Want to Find Answers:
  - Our goal is to forecast the reusability of SpaceX's Falcon 9 first stages. This prediction will enable Space Y to optimize launch pricing and operational strategies, positioning us as a contender in the commercial space market.

Section 1

Methodology

Methodology

Executive Summary

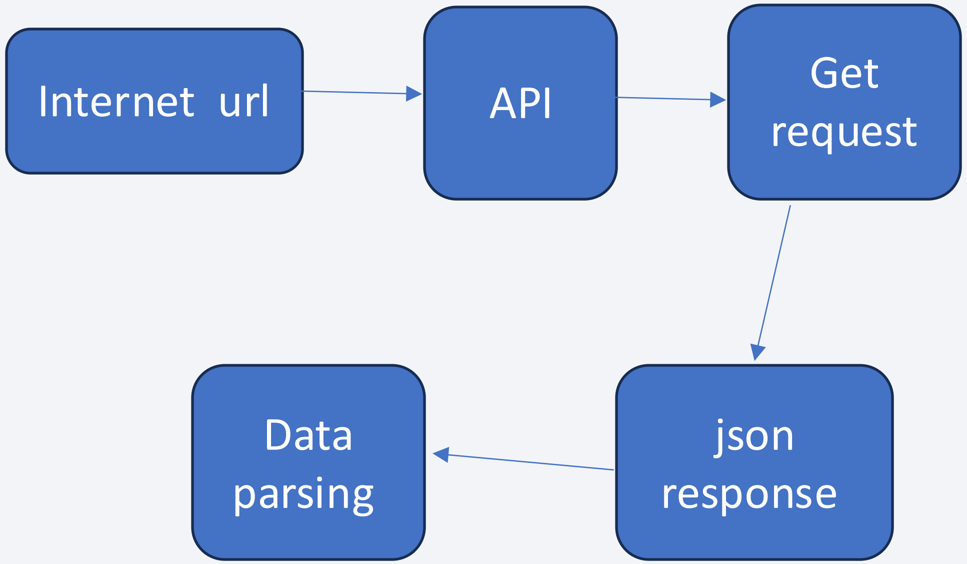
- Data collection methodology:
    - Using APIs and web scrapping, two data sources where used.
- Perform data wrangling
    - With pandas library, I cleaned the data and filled missing values.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - How to build, tune, evaluate classification models

Data Collection

- Two methodoligies applied: APIs and Web Scrapping.
- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

- Here, I used SpaceX API in order to collect the needed data.
- The procces is shown in the following figure:

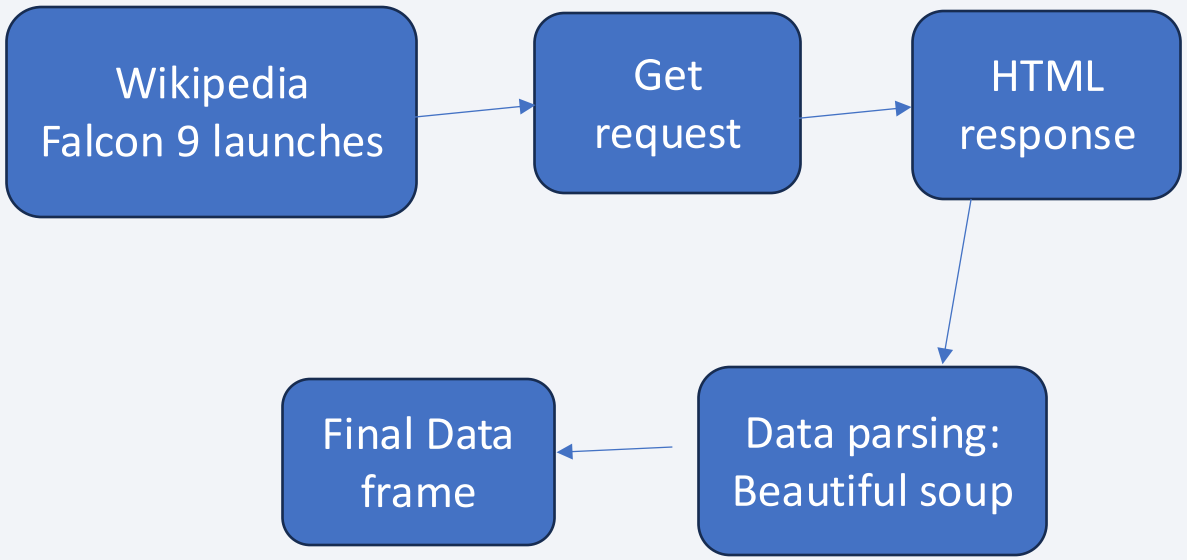


- GitHub URL of SpaceX API calls notebook: capstone_project/jupyter-labs-spacex-data-collection-api.ipynb at main · facundoa98/capstone_project (github.com)

Data Collection - Scraping

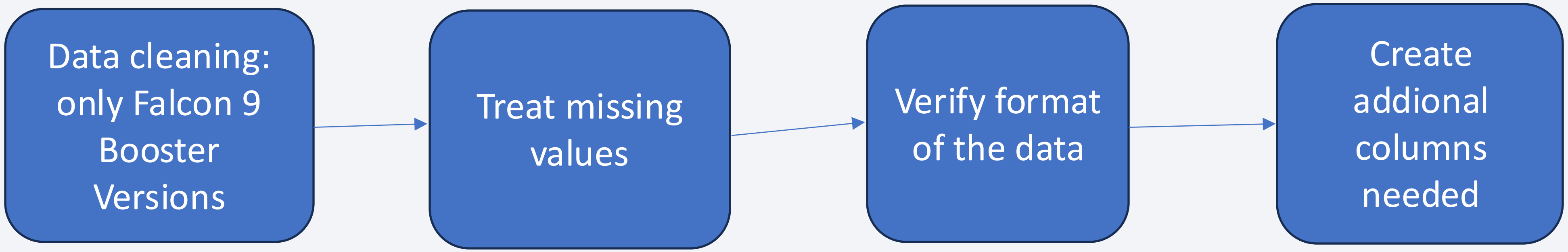
- Web scrapping process in order to obtain data from Wikipedia. Shown in the figure:

- GitHub URL:capstone_project/jupyter-labs-webscraping.ipynb at main · facundoa98/capstone_project (github.com)



Data Wrangling

- Data wrangling proccess shown in the following figure:
- GitHub URL of completed data wrangling: capstone_project/labs-jupyter-spacex-Data wrangling.ipynb at main · facundoa98/capstone_project (github.com)



EDA with Data Visualization

- FlightNumber vs PayloadMass overlaying outcome of the launch. Scatter plot. To identify correlations between these two variables and positive outcomes.
- Flight number-LaunchSite-Outcome. To see if there is any correlation between launch site and final outcome.
- PayloadMass-LaunchSite-Outcome. To check correlations between these variables.
- Success rate per Orbit type. Bar Chart. Determine if there are orbits that are more reliable considering launch outcomes.
- Payload-Orbit type. To reveal any posible relationship between them.
- Launch success trend. Line plot to see trends in succeded launches.
  - GitHub URL: capstone_project/edadataviz.ipynb at main · facundoa98/capstone_project (github.com)

EDA with SQL

SQL queries performed:

- Displaying unique launch sites
- Displaying 5 records where launch sites begin with 'CCA'
- Display 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 records for failure landings in drone ships
- Rank the count of landing outcomes
    - GitHub URL: capstone_project/jupyter-labs-eda-sql-coursera_sqllite.ipynb at main · facundoa98/capstone_project (github.com)

Build an Interactive Map with Folium

- In order to determine how does launch site impacts on the mission being succesful, I created some Folium visualizations, which are an interactive way to understand the available data. Also, I did calculate and indicate distance to some points of interest, such as the closest city, road and sea border.
- Particularly, I created:
    - Circles: to see which are the places where the launches took place.
    - Markers: to visualize success rate per launch site, since the markers where colored depending on outcomes.
    - Lines: to determine the closest point of interest: cities, roads and sea proximities.

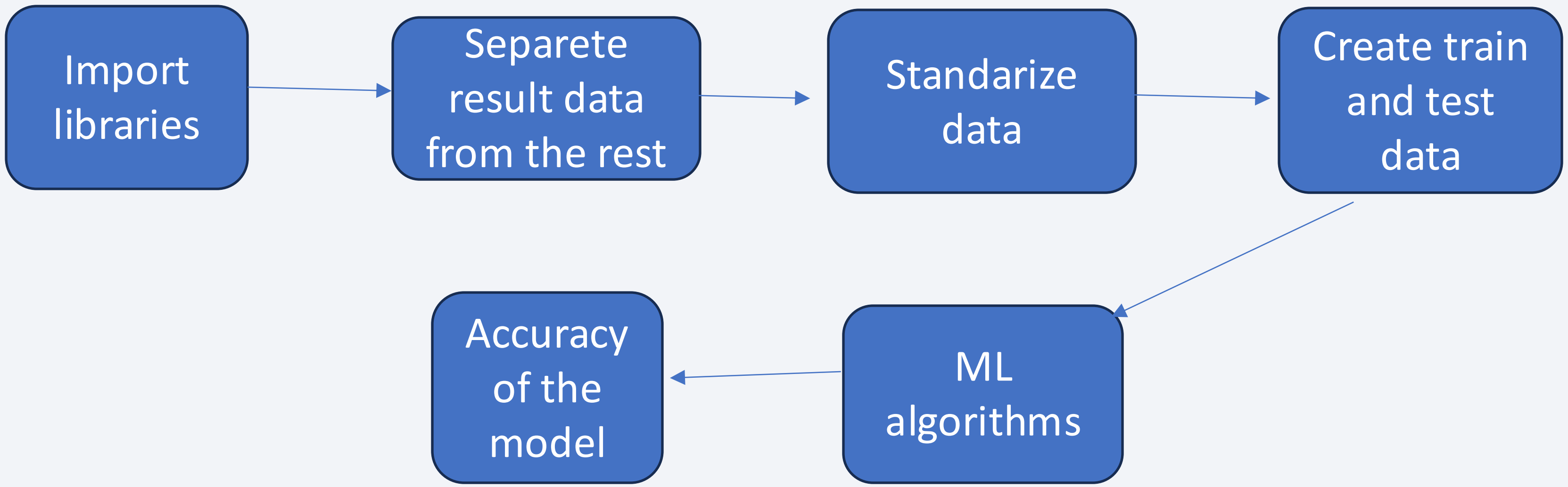
    - GitHub URL: capstone_project/lab_jupyter_launch_site_location.ipynb at main · facundoa98/capstone_project (github.com)

Build a Dashboard with Plotly Dash

- I chose two graphs to add to the dashboard:
  - Pie chart: success rate per launch site.
  - Scatter plot: correlation between Payload and Success.
  - I also added on markdown list and one range selector where you can choose payload mass.
- This plots help to see, dynamically, how different factors affect the success of a launch.

  - GitHub URL: [capstone_project/dash app at main · facundoa98/capstone_project (github.com)]()

Predictive Analysis (Classification)



- GitHub URL: capstone_project/SpaceX_Machine Learning Prediction_Part_5.ipynb at main · facundoa98/capstone_project (github.com)

Results

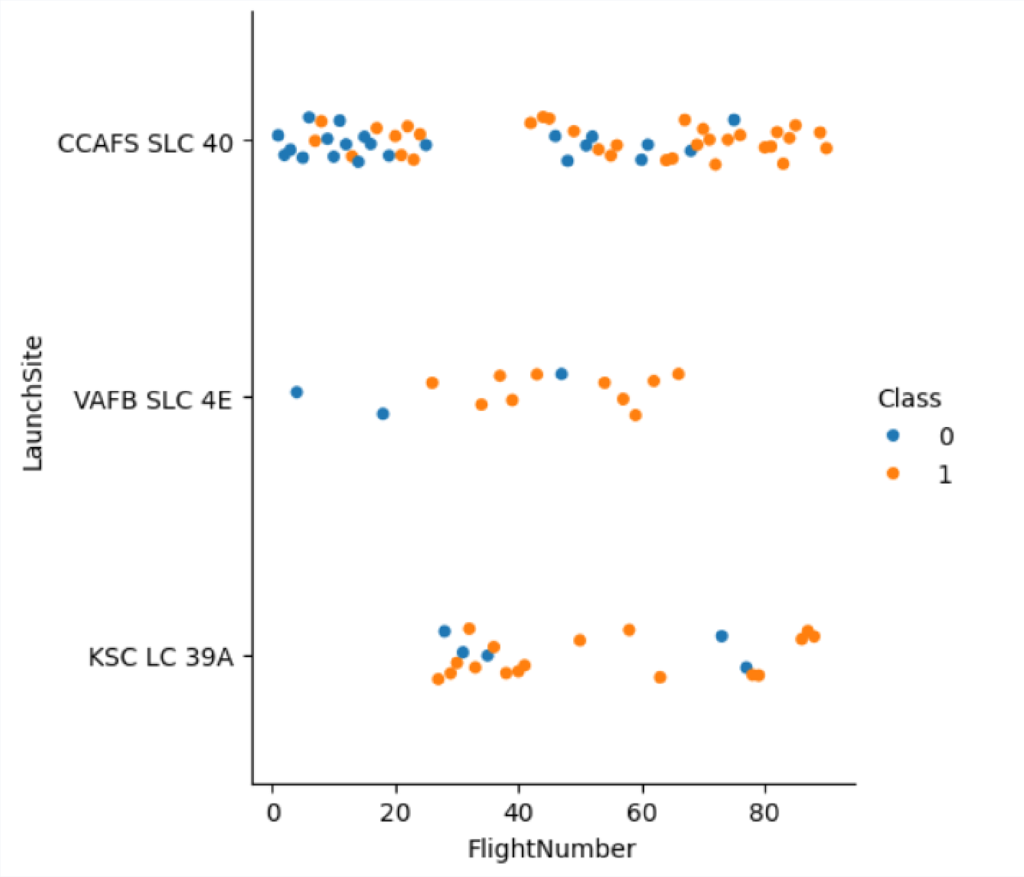
- Exploratory data analysis results
- High flight number --> succesful landing
- Orbits ES-L1, GEO, HEO, SSO --> 100% success rate
- With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS
- Success rate is increasing every year
- Predictive analysis results: all of the ML algorithms created are equally good to predict results with test data.

Section 2

Insights drawn from EDA

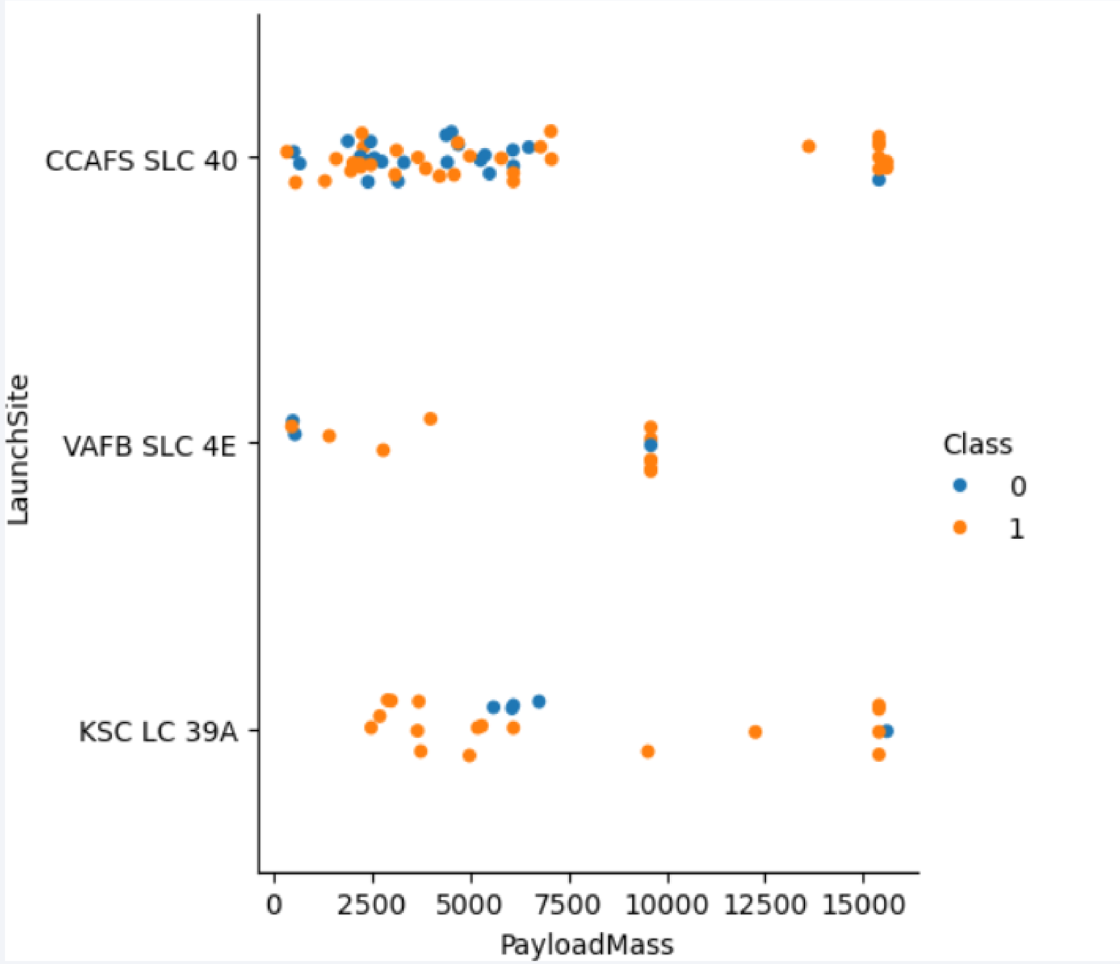
Flight Number vs. Launch Site

- With different rates, we can see that for the three sites missions are more succesfull when the flight number is higher.



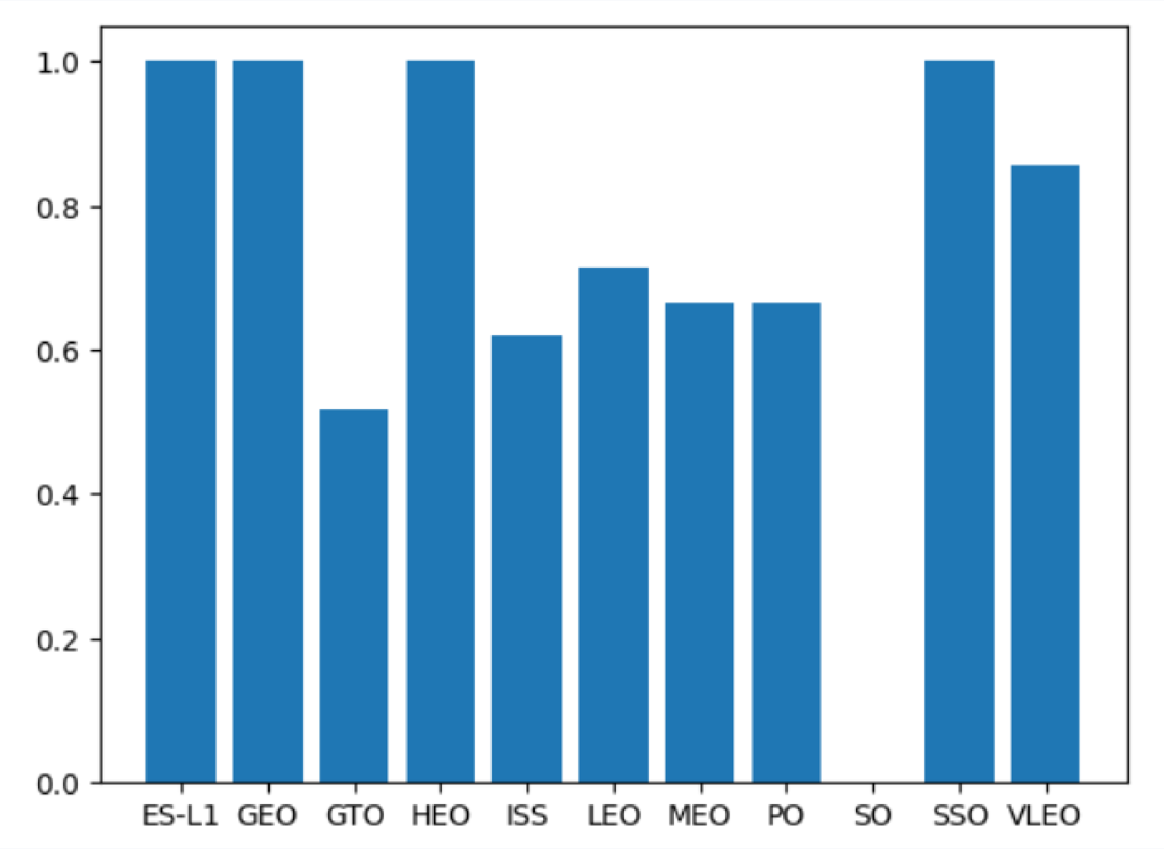
Payload vs. Launch Site

- There are some conclusion to the sight here:
    - CCAFS SLC 40: more success with higher payload mass.
    - VAFB SLC 4E: max payload mass of 10.000 Kg.
    - KSC LC 39A: succesfull with light payload mass, in comparison with CCAFS SLC 40.



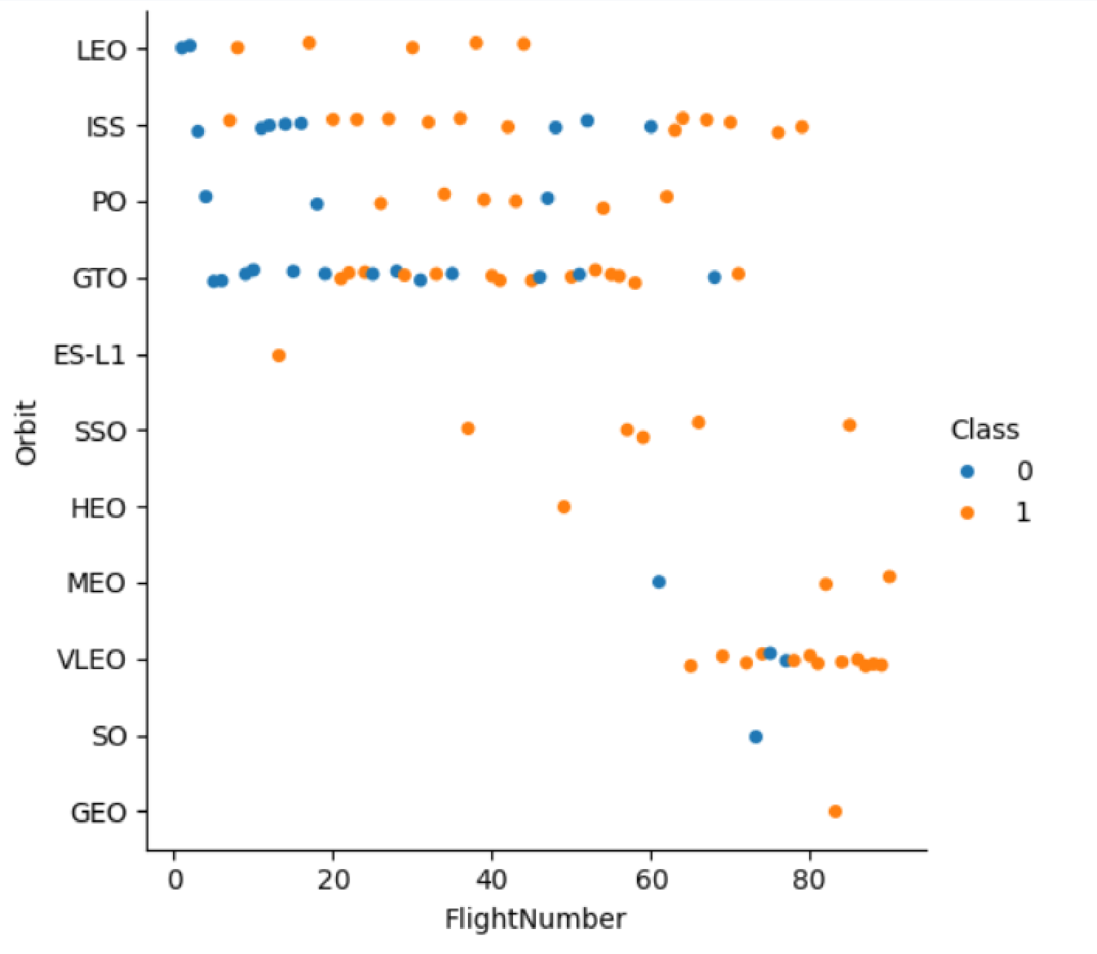
Success Rate vs. Orbit Type

- 100% success rate in four orbits: ES-L1, GEO, HEO and SSO.
- Still not launches to the SO orbit.



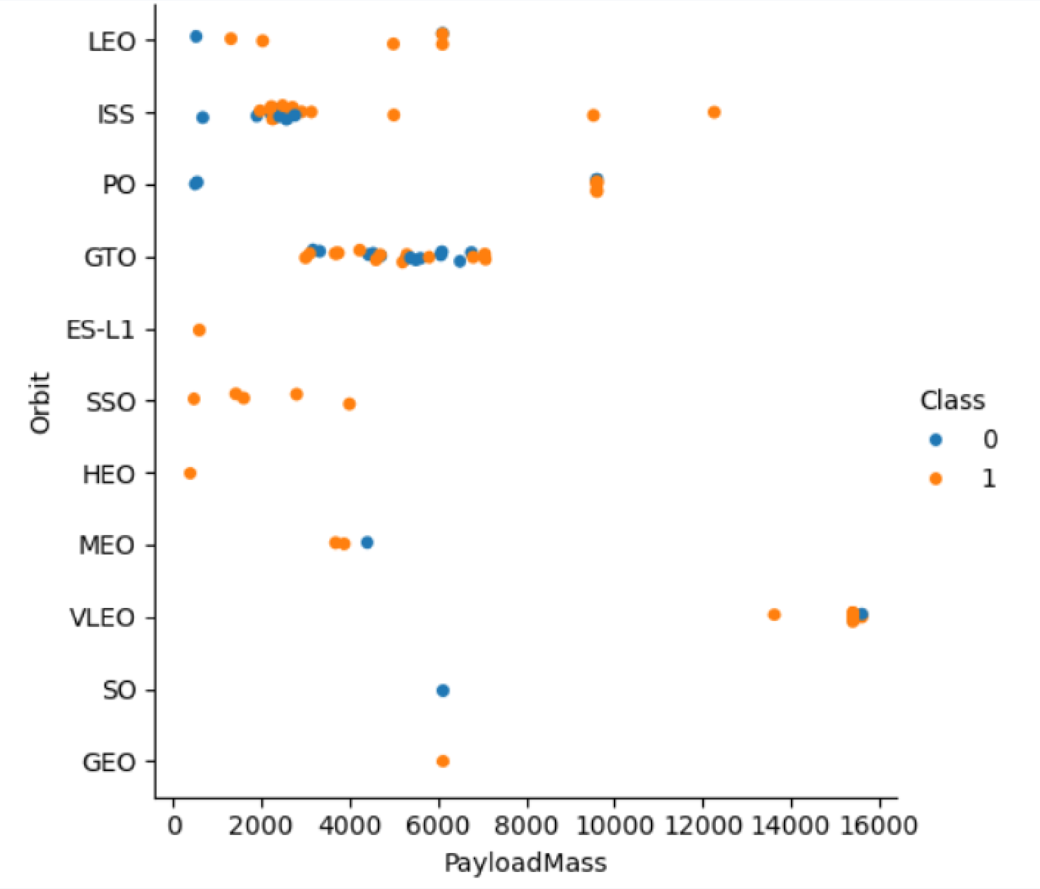
Flight Number vs. Orbit Type

- The orbits that have a 100% success rate, are the ones with less launches.
- LEO launches success increase when the flight number is higher.



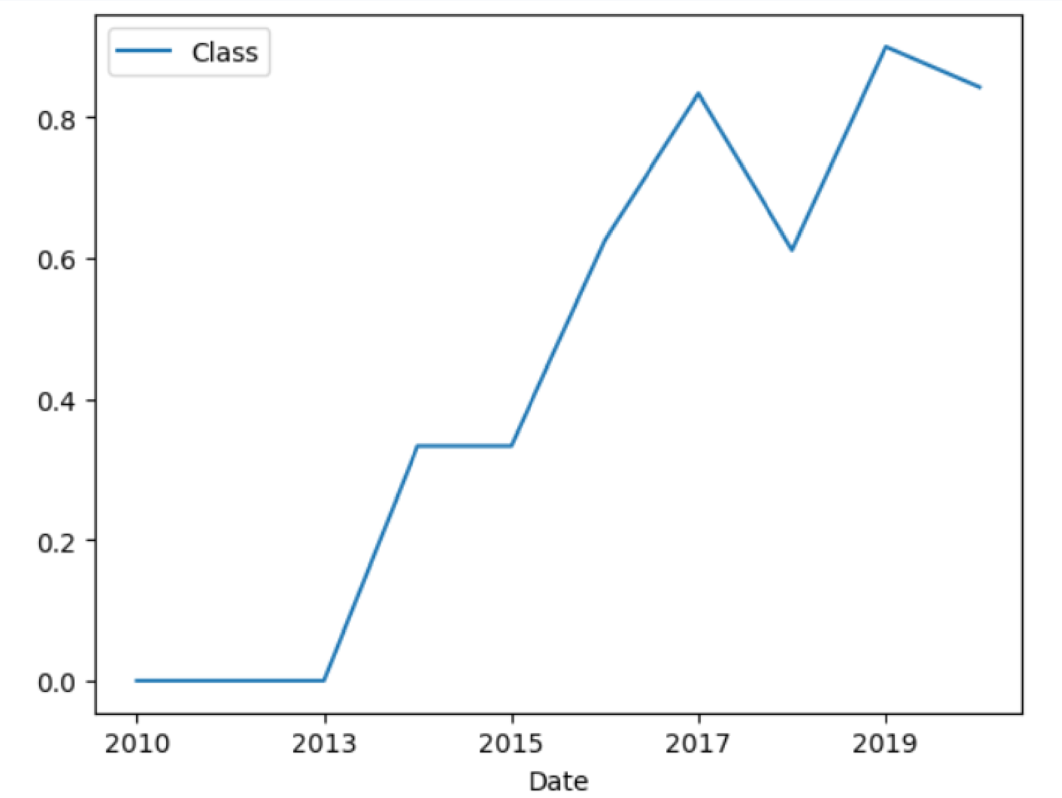
Payload vs. Orbit Type

- Heaviest launches directed to VLEO orbit.
- Most of the launches are lighter than 6000 kg.



Launch Success Yearly Trend

- Clear increase on success rate with an increase on the year, especially from 2013.



All Launch Site Names

- Unique launch sites:
  - `CCAFS LC-40`
  - `VAFB SLC-4E`
  - `KSC LC-39A`
  - `CCAFS SLC-40`

Launch Site Names Begin with 'CCA'

- Query result:
  - ('CCAFS LC-40',)
  - ('CCAFS LC-40',)
  - ('CCAFS LC-40',)
  - ('CCAFS LC-40',)
  - ('CCAFS LC-40',)

Total Payload Mass

- Calculate the total payload carried by boosters from NASA (CRS)
    - Query result: 45596.

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1:
  - 2928.4 Kg

First Successful Ground Landing Date

- Date of the first successful landing outcome on ground pad:
    - 2010-06-04

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000
  - [('F9 FT B1022',),
  - ('F9 FT B1026',),
  - ('F9 FT B1021.2',),
  - ('F9 FT B1031.2',)]

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes

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

Boosters Carried Maximum Payload

Booster which have carried the maximum payload mass:

[('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

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015:
  - ('01', 'Failure (drone ship)', 'F9 v1.1 B1012', 'CCAFS LC-40')
  - ('04', 'Failure (drone ship)', 'F9 v1.1 B1015', 'CCAFS LC-40')

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

- 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:
  - 619967

Section 3

Launch Sites Proximities Analysis

<Folium Map Screenshot 1>

- Launch sites across United States:

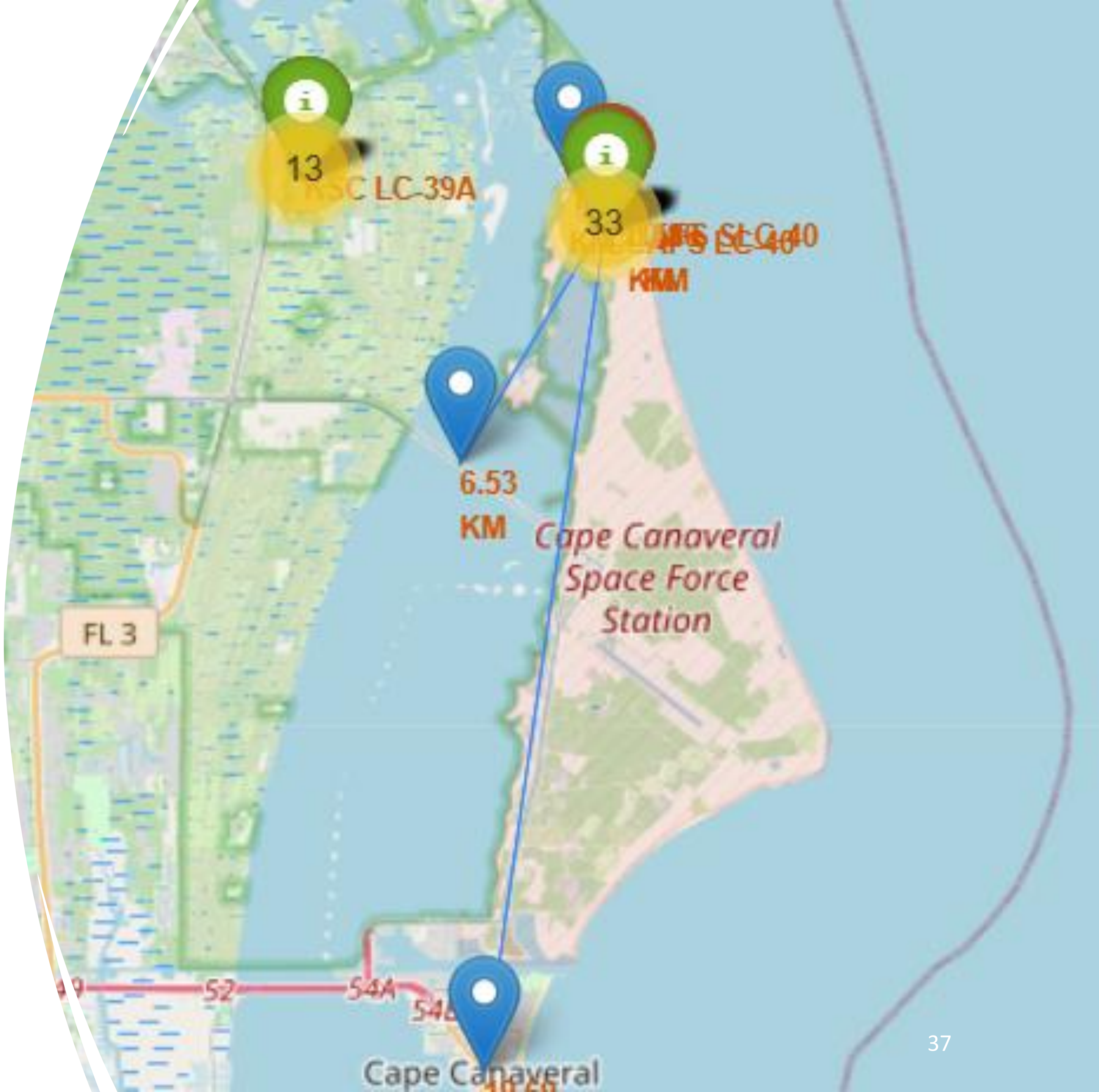
- You can see here that launch sites are in proximity to the coast and to the equator line.

<Folium Map Screenshot 2>

- Markers determing amount of launches per site.
- Also, if you zoomed in you could see how many success/failure landing attemts there are.

<Folium Map Screenshot 3>

- Launch sites and close interest points:
  - Railways
  - Coastline
  - Cities

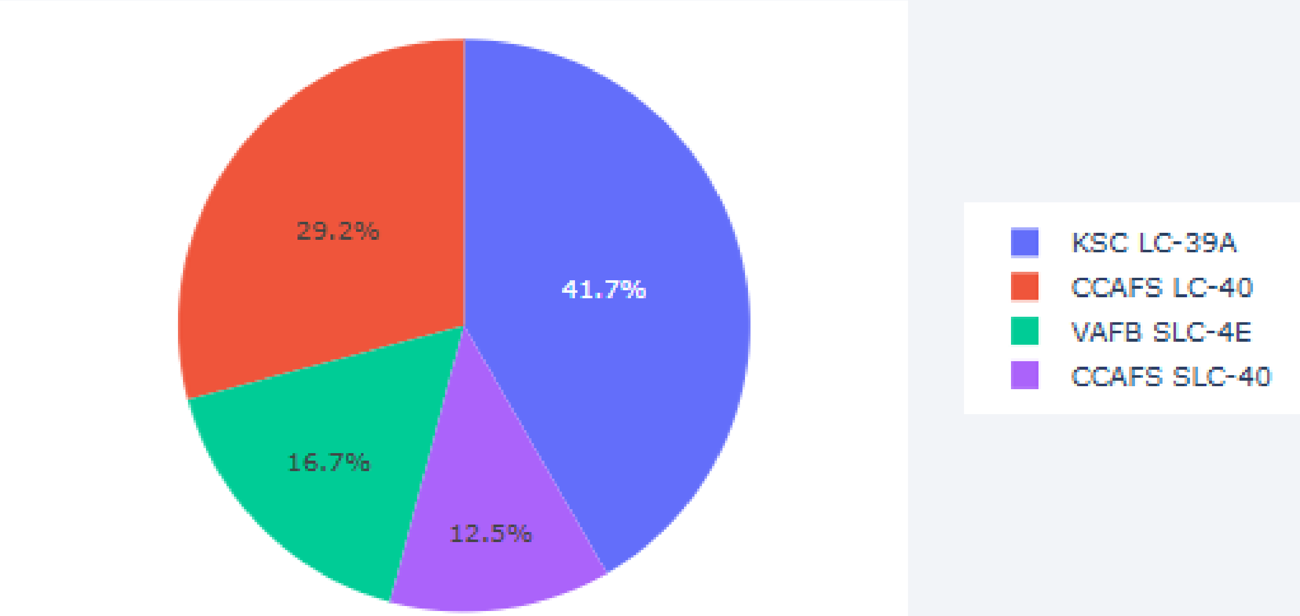


Section 4

Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>

- Success pie chart:



- You can clearly see tat KSC LC-39A has the largest succesfull landings.

<Dashboard Screenshot 2>

- If you chose every site and checked how succesfull each are, you see that:
  - KSC LC-39A has a 76.9% success rate.

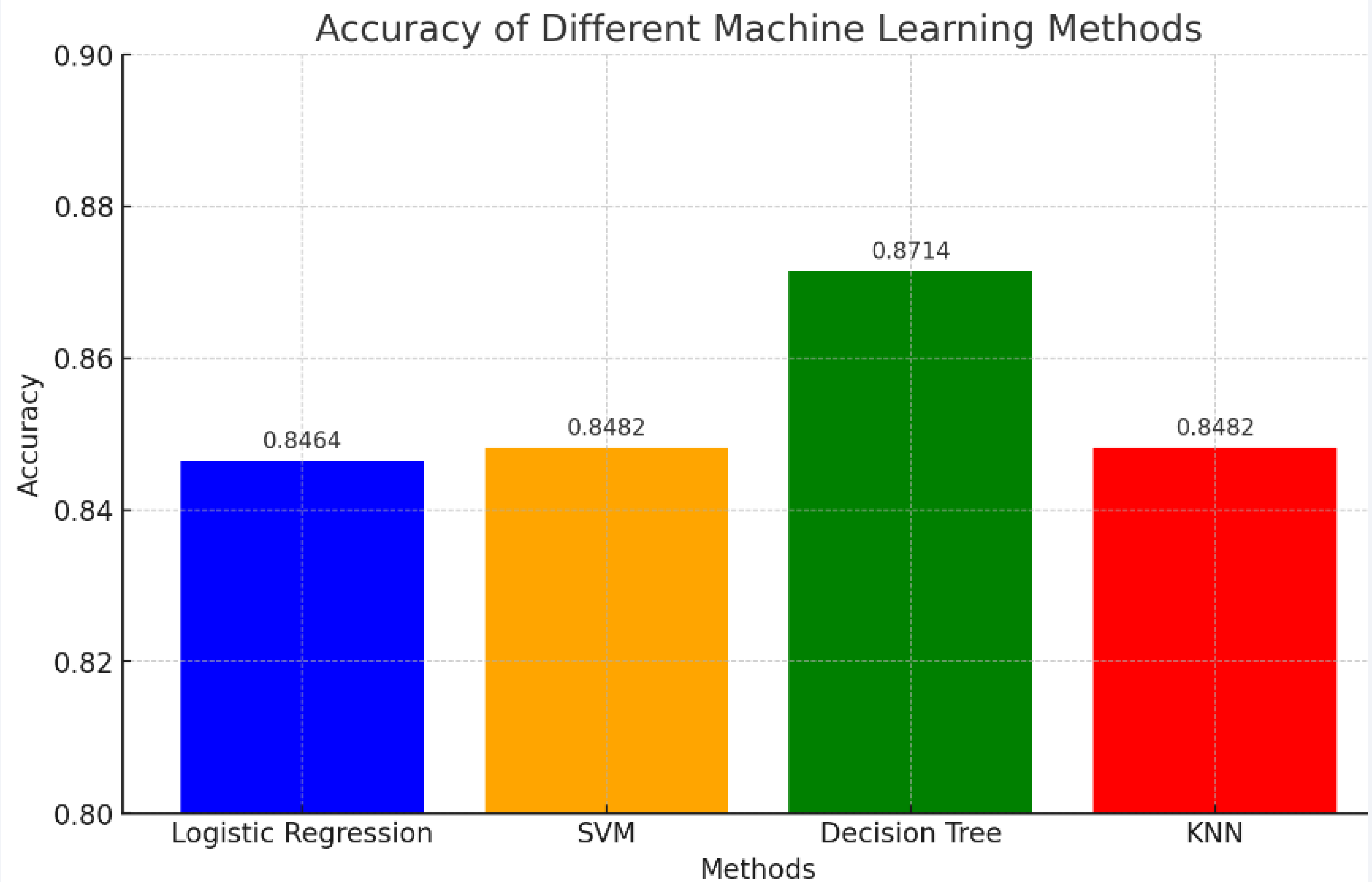
<Dashboard Screenshot 3>

- Payload vs. Launch Outcome scatter plot for all sites

Section 5

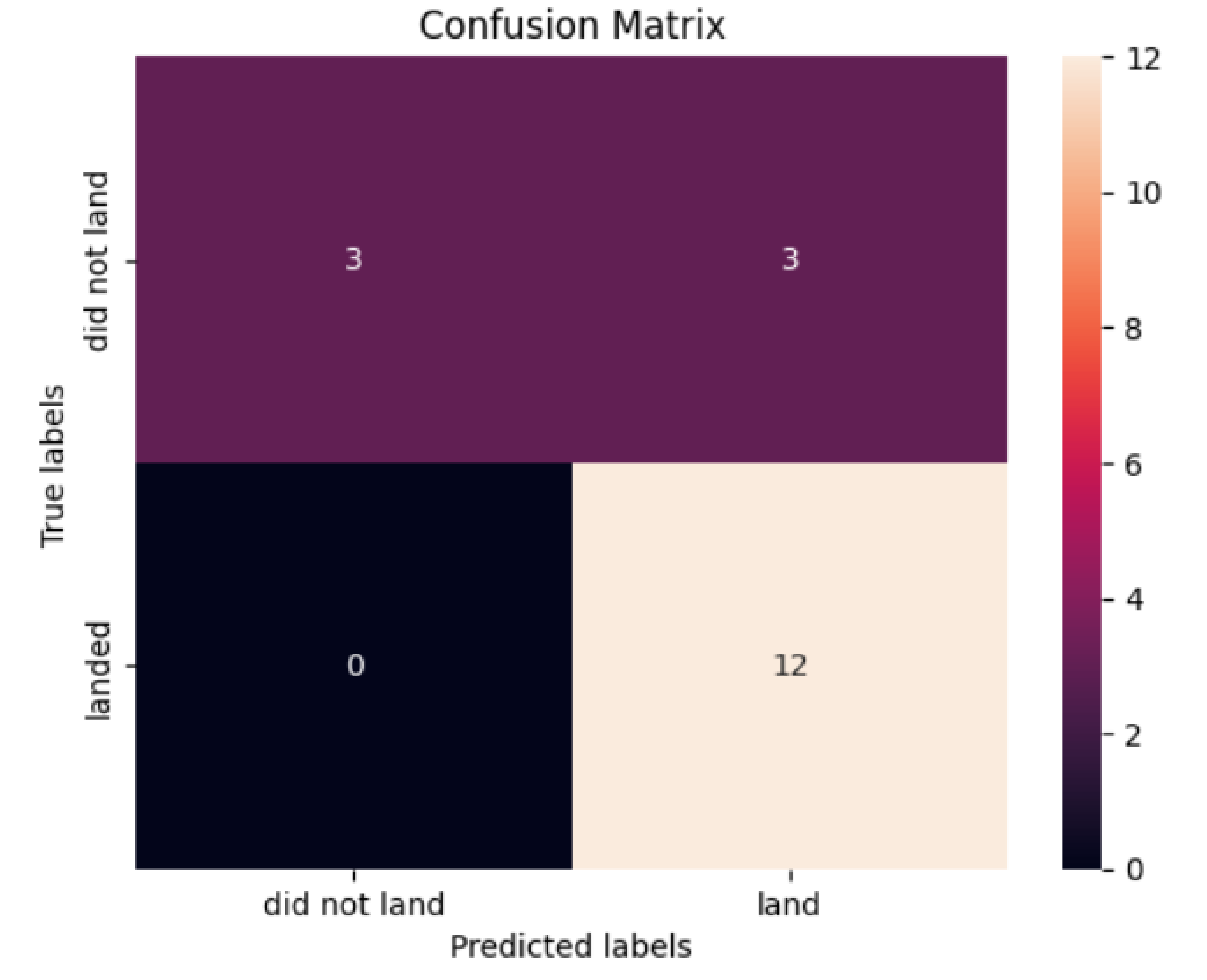
Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix

- Confusion matrix for the Decision tree ML model:



Conclusions

There are many conclusion to take here:

- Best performing ML algorithm is the Decision Tree.
- Higher Flight number is related to more probalities on succesfull landing.
- Every year the success rate is bigger.
- With 76.9%, KSC LC-39A is the best place to perform launches.

Thank you!