IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Amir Mobayen
9/25/2023

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collecting using SpaceX API and Web Scraping
  - Data Wrangling
  - Data Exploration using SQL and Visualization
  - Building dashboards via Plotly and Folium
  - Machine Learning Prediction
- Summary of all results
  - Explratory Data Analysis results
  - Interactive Dashboards
  - ML Prediction Results

Introduction

- Project background and context
  - 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.
- Problems you want to find answers
  - 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.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Space X Json API, Calling in Jupiter environment (1).
  - Web Scraping Wikipedia Page (2)
- Perform data wrangling
  - Filtering Data
  - Filling missed Data with mean values
  - One Hot encoding for classification Problem
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Building Different models, cross validating different parameters

Data Collection

- Data is collected from Public API, then filtered to desired parameters (Falcon 9) and finally, dealing with Null Values and replace them with calculated mean values.

- Sources: Appendix 1, 2

Data Collection – SpaceX API

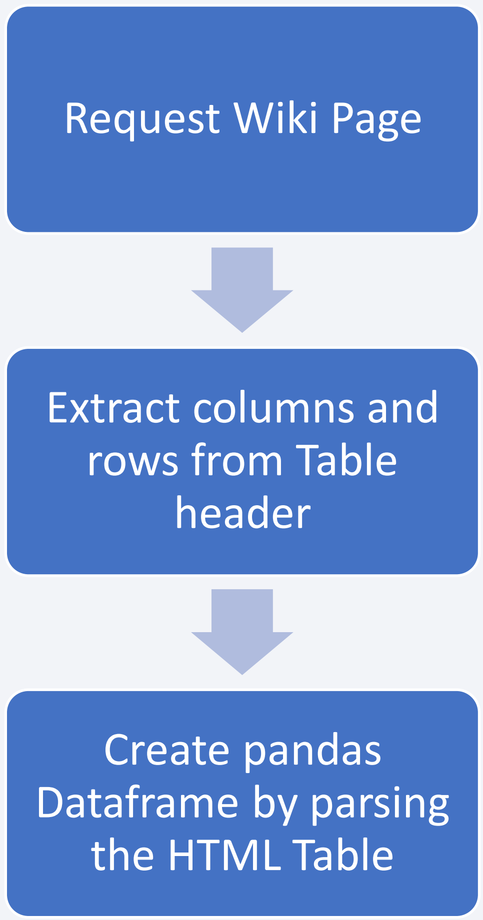
- Data is collected from Public API, then filtered to desired parameters (Falcon 9) and finally, dealing with Null Values and replace them with calculated mean values.

- Github: https://github.com/ahmobayen/Personal_Projects/blob/IBM/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

- Space X Launches Data are collected from the Wikipedia page (appendix2)
- It's Downloaded via the beautifulsoup library
- We parsed the info in jupyter lab.

- **Github:** https://github.com/ahmobayen/Personal_Projects/blob/IBM/jupyter-labs-webscraping.ipynb



Data Wrangling

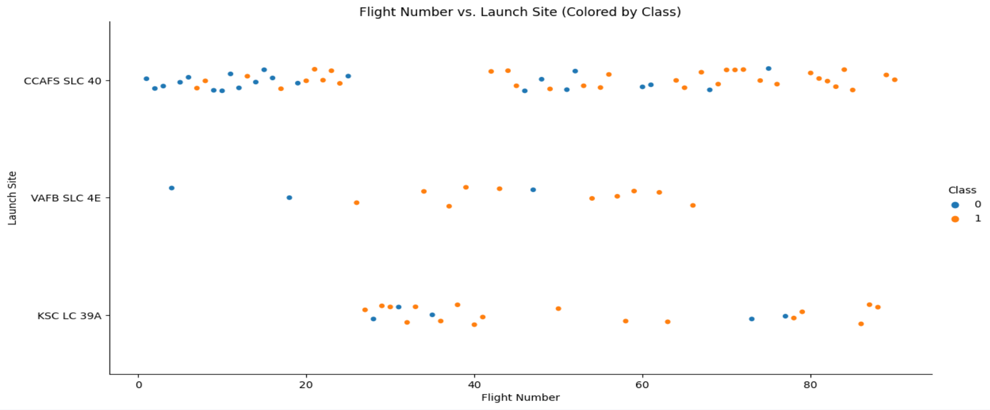
- Perform some Exploratory Data Analysis (EDA) to find some patterns in the data and determine what would be the label for training supervised models.
- Perform exploratory Data Analysis and determine Training Labels
  - **Calculate the number of launches on each site**
  - **Calculate the number and occurrence of each orbit**
  - **Calculate the number and occurence of mission outcome per orbit type**
  - **Create a landing outcome label from Outcome column**

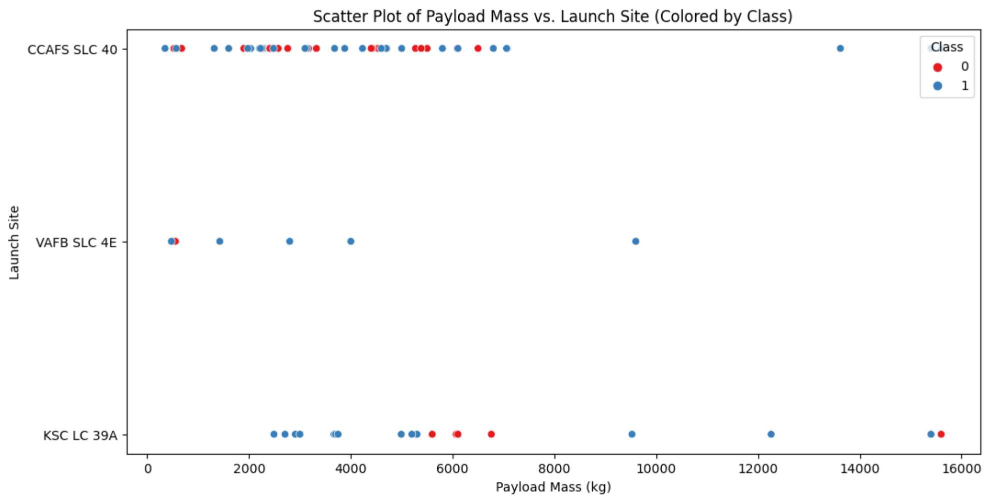
- Github:
  - https://github.com/ahmobayen/Personal_Projects/blob/IBM/labs-jupyter-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb

EDA with Data Visualization

- Try to find the relationships which are the factor to the successful landing, Hence, base on this we try to work on features like, FlighNumer, PayloadMass, Launch site, orbit type

- Github:
  - https://github.com/ahmobayen/Personal_Projects/blob/IBM/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb





EDA with SQL

- Summarize of the performed SQL queries
  - **Find the names of the unique launch sites in the space mission**
  - **Find the total payload mass carried by boosters launched by NASA (CRS)**
  - **Find average payload mass carried by booster version F9 v1.1**
  - **Find the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000**
  - **Find the total number of successful and failure mission outcomes**
  - **Find the names of the booster_versions which have carried the maximum payload mass. Use a subquery**
  - **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,**
- Github:
  - https://github.com/ahmobayen/Personal_Projects/blob/IBM/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Markers, Circles, and clusters were used with Folium maps:
  - Markers indicate launch sites
  - Circles indicate highlighted Space center area
  - Marker cluster indicate the group of events

- GitHub:
  - https://github.com/ahmobayen/Personal_Projects/blob/IBM/lab_jupyter_launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

- The Following interactive charts used for:
  - Percentage of launches by site
  - Payload Range
- This combination allows us to find the relationship between the payload and launch sites, and find the best launch location according to payloads
- GitHub:
  - https://github.com/ahmobayen/Personal_Projects/blob/IBM/Interactive%20Dashboard.py

Predictive Analysis (Classification)

Building Models
- Load dataset
- Normalize and split dataset
- Choosing Classification models

Evaluating Models
- Check Accuracy
- Plot Confusion matrix

Improving Models
- Use gridCV to find best parameters
- Feature engineering

Finding the best Model
- Choose highest accuracy model

- GitHub:
  - https://github.com/ahmobayen/Personal_Projects/blob/IBM/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

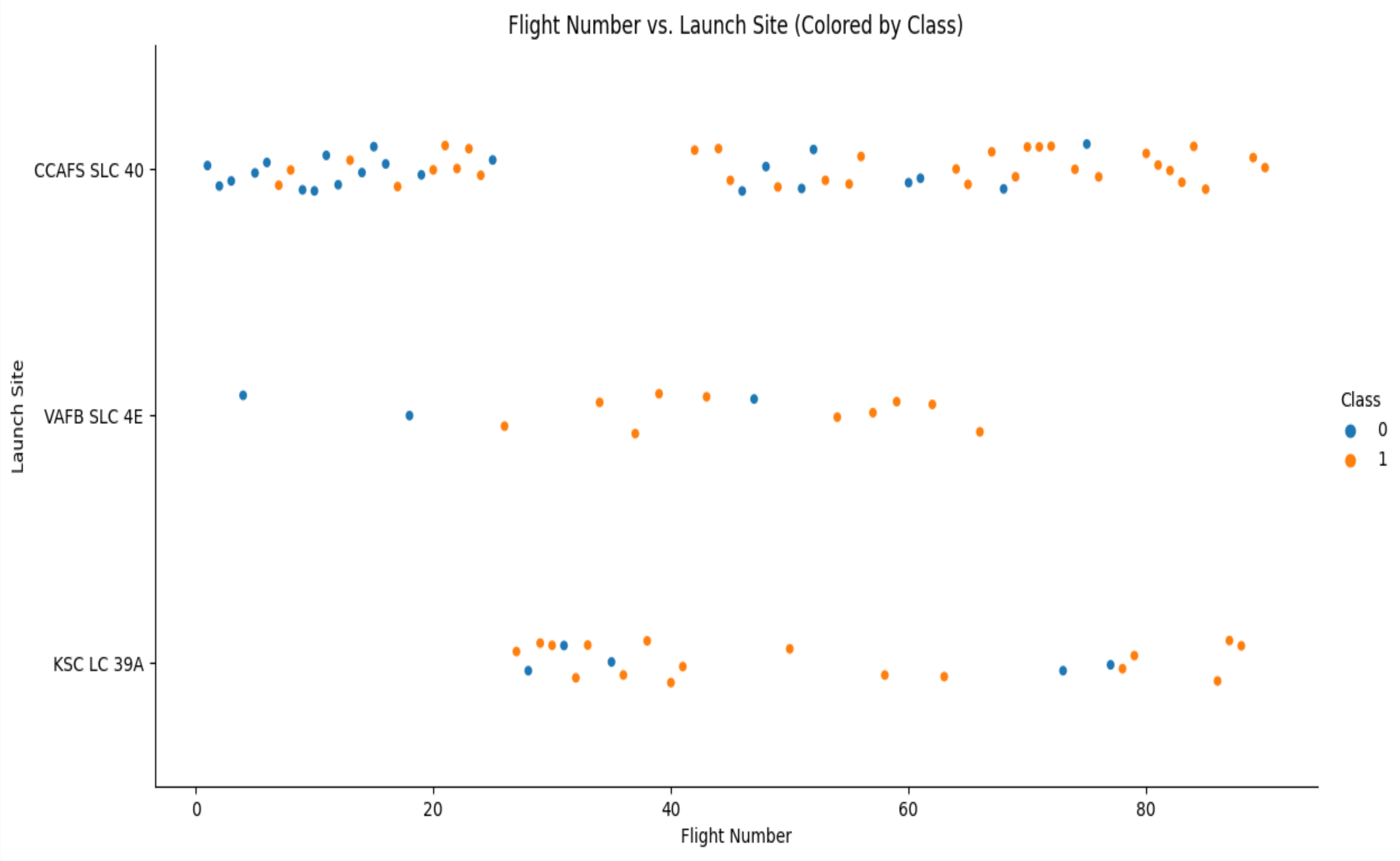
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

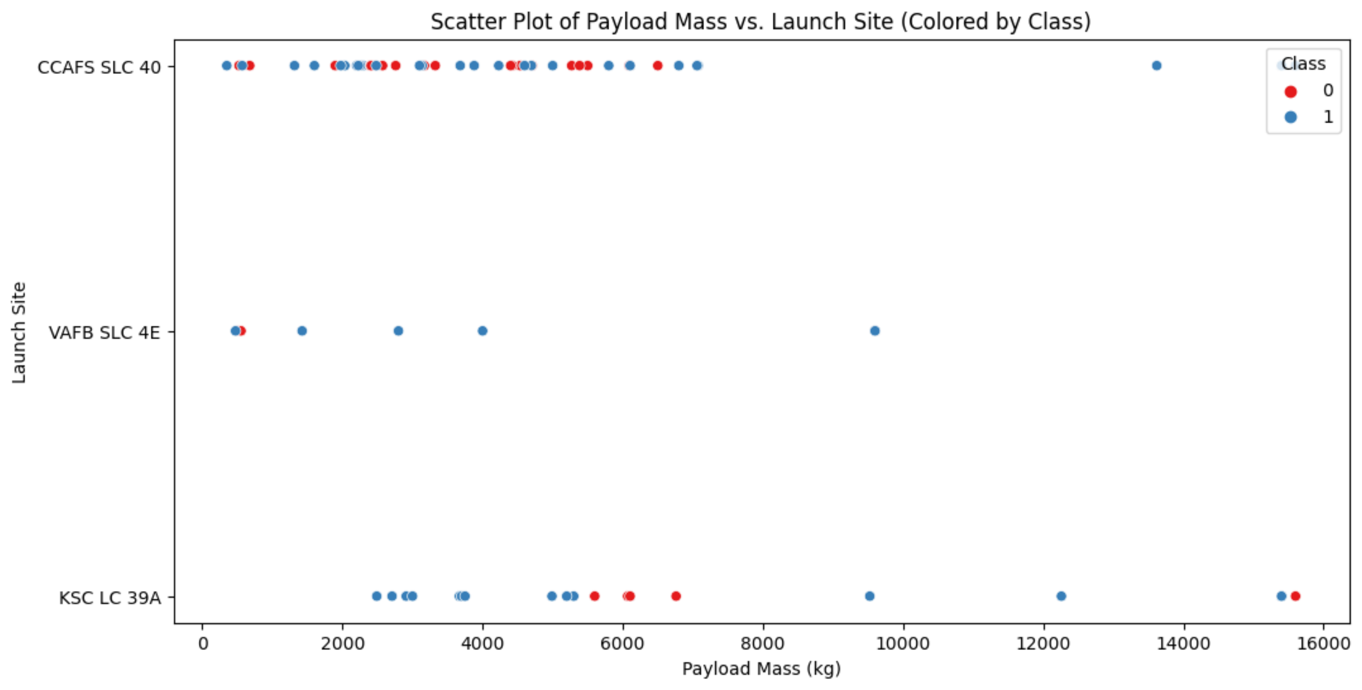
- This Scatter plot shows that the larger flights increases the chance of success
- Also, we can see the launch site also have affect the success rate



Payload vs. Launch Site

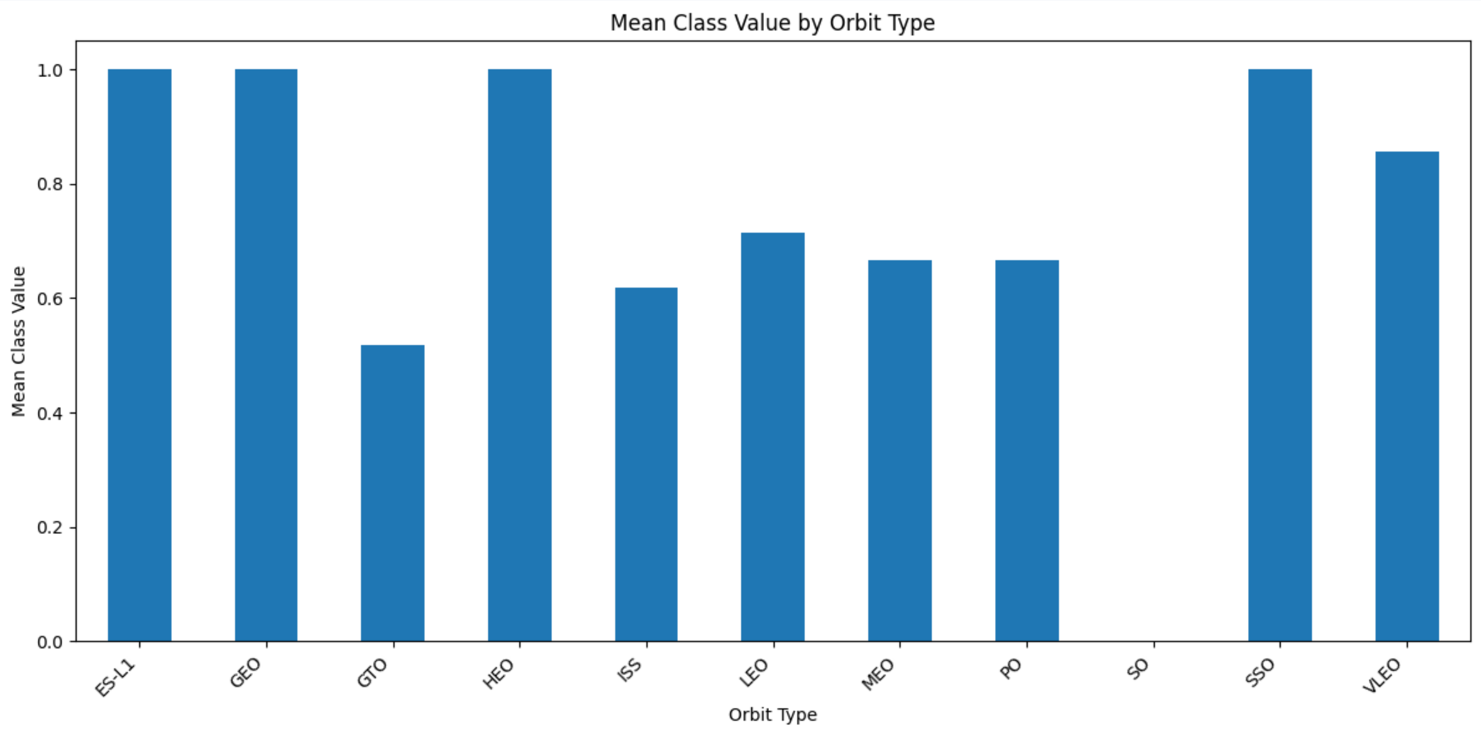
- We find for the VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).

- Success rate for mass above 7KT is almost 100%



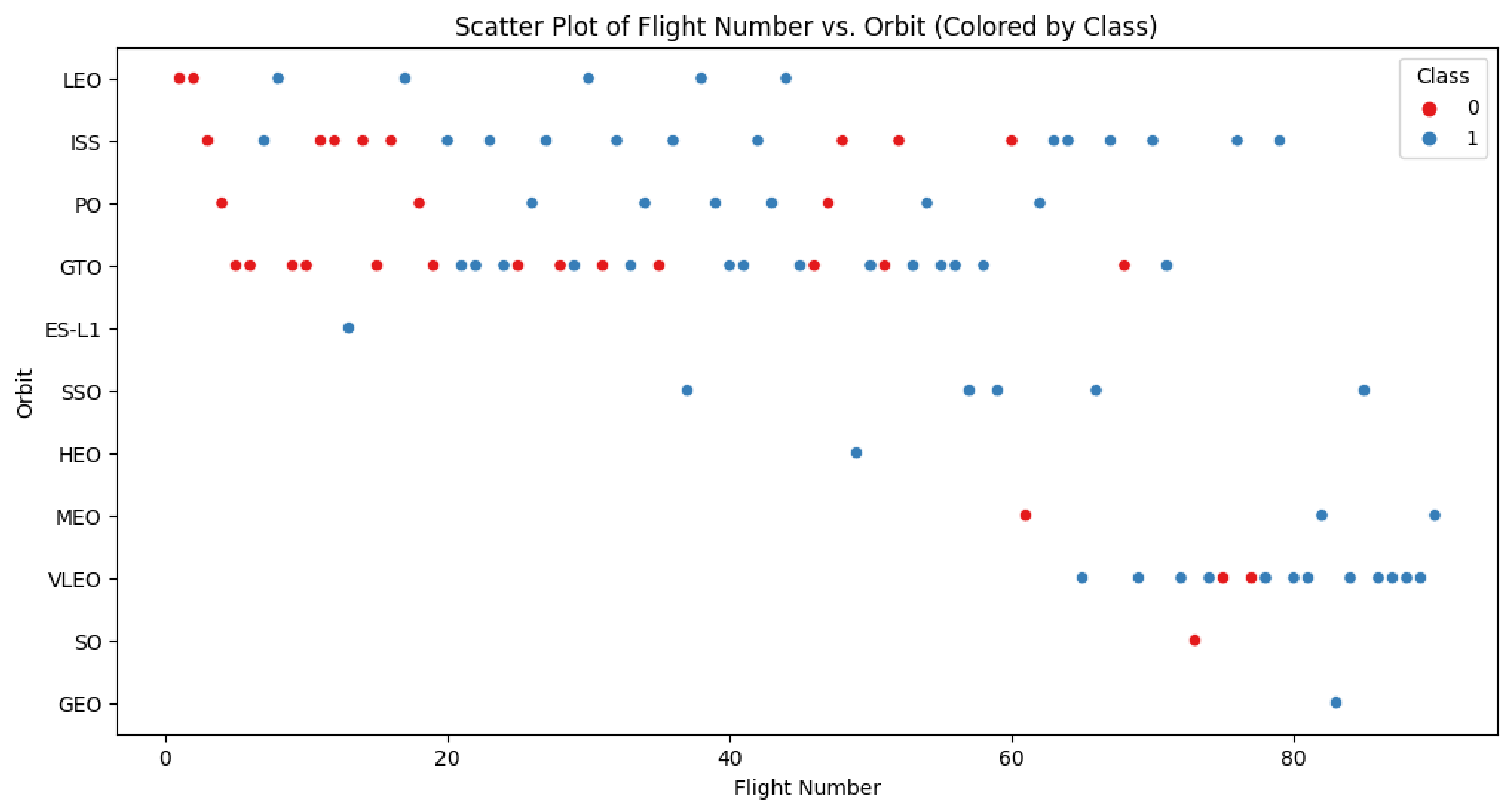
Success Rate vs. Orbit Type

- SO orbit has 0% success rate
- SSO, HEO, GEO and ES-L1 have 100% success rate



Flight Number vs. Orbit Type

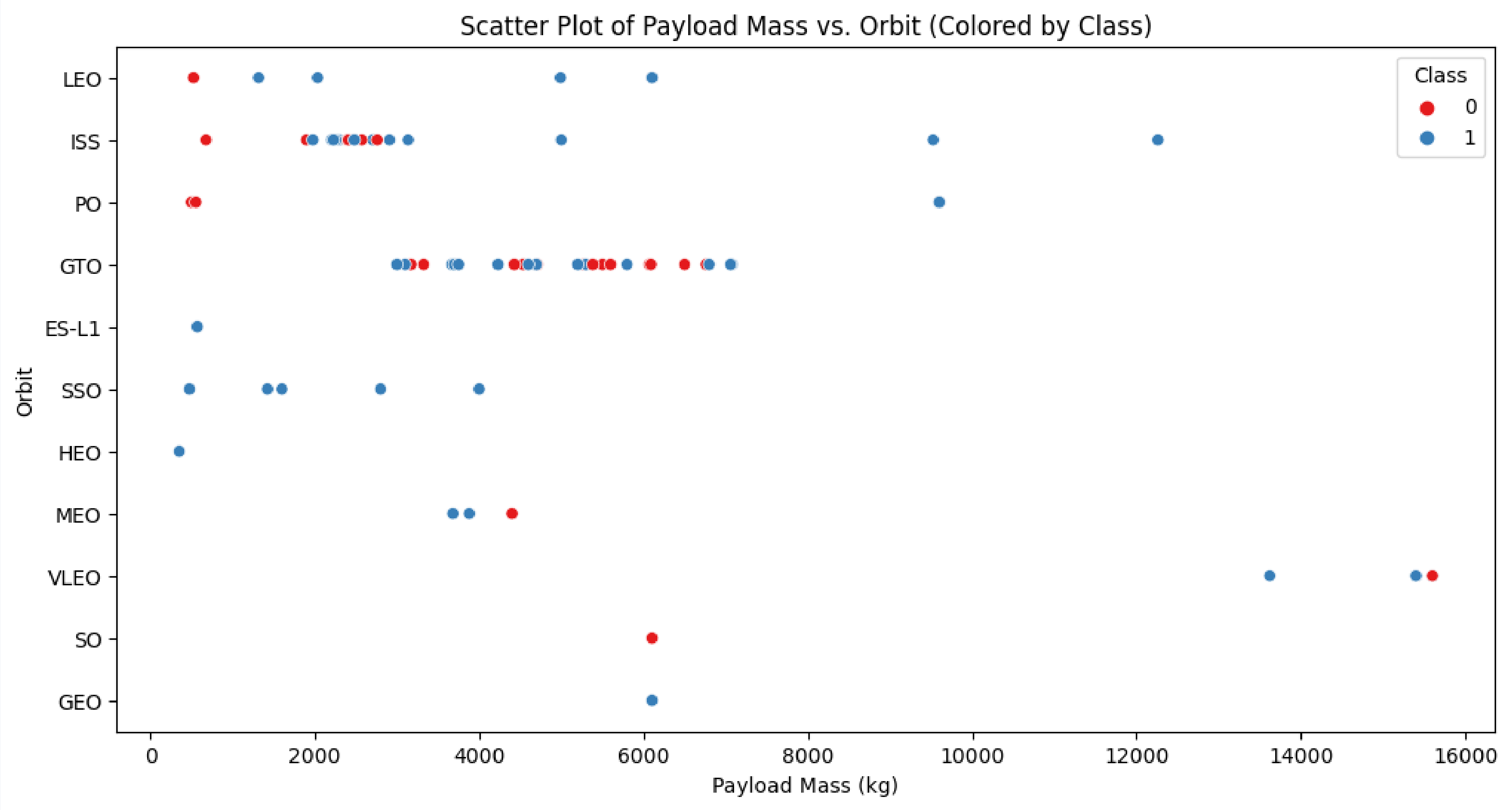
- In the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit



Payload vs. Orbit Type

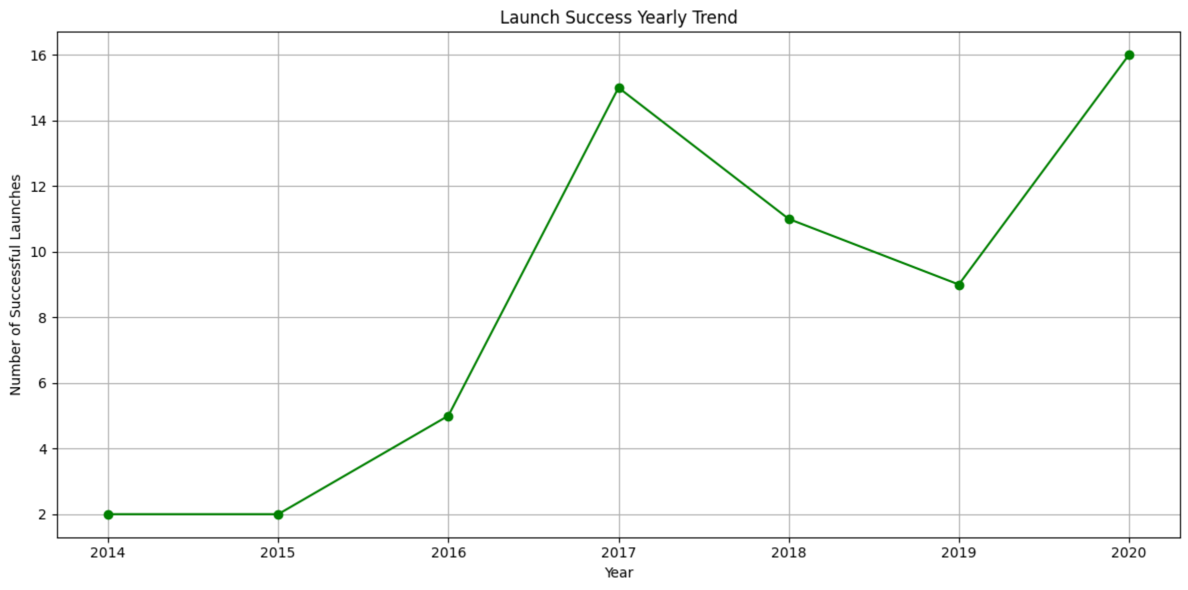
- With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

- 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

- We can observe that the sucess rate since 2013 kept increasing till 2020



All Launch Site Names

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

- Use Query:
  - select distinct(Launch_Site) from SPACEXTABLE

Launch Site Names Begin with 'CCA'

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

Use Query: select * from SPACEXTABLE where Launch_Site like 'CCA%' limit 5

Total Payload Mass

- Calculate the total payload carried by boosters from NASA:
  - 99980 KG
- Query:
  - select sum(PAYLOAD_MASS__KG_) as 'total payload Mass' from SPACEXTABLE where Customer like 'NASA%'

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1:
    - 2534.6666666666665
- Query
    - select avg(PAYLOAD_MASS__KG_) as 'average' from SPACEXTABLE where Booster_Version like 'F9 v1.1%'

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad:
  - 2015-12-22
- Query:
  - SELECT MIN(Date) AS 'First successful date' FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (ground pad)'

SELECT Booster_Version **FROM** SPACEXTABLE **WHERE** Landing_Outcome = 'Success (drone ship)' **and** PAYLOAD_MASS__KG_ **between** 4000 **and** 6000

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
- Query:
  - SELECT Booster_Version FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (drone ship)' and PAYLOAD_MASS__KG_ between 4000 and 6000

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

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

- Query:
  - SELECT Mission_Outcome, COUNT(*) AS Count
  - FROM SPACEXTABLE
  - GROUP BY Mission_Outcome;

Boosters Carried Maximum Payload

- List the names of the 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

Query:

```
SELECT Booster_Version
FROM SPACEXTABLE
WHERE PAYLOAD_MASS__KG_ = (
    SELECT MAX(PAYLOAD_MASS__KG_)
    FROM SPACEXTABLE
);
```

2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015:
    - The query returns nothing.

- Query:
    - This SQL query transforms numerical month values into corresponding names and retrieves data on failed drone ship landings in 2015. It presents the month, landing outcome, booster version, and launch site for the specified events.

```
SELECT
   CASE
      WHEN substr(Date, 4, 2) = '01' THEN 'January'
      WHEN substr(Date, 4, 2) = '02' THEN 'February'
      WHEN substr(Date, 4, 2) = '03' THEN 'March'
      WHEN substr(Date, 4, 2) = '04' THEN 'April'
      WHEN substr(Date, 4, 2) = '05' THEN 'May'
      WHEN substr(Date, 4, 2) = '06' THEN 'June'
      WHEN substr(Date, 4, 2) = '07' THEN 'July'
      WHEN substr(Date, 4, 2) = '08' THEN 'August'
      WHEN substr(Date, 4, 2) = '09' THEN 'September'
      WHEN substr(Date, 4, 2) = '10' THEN 'October'
      WHEN substr(Date, 4, 2) = '11' THEN 'November'
      WHEN substr(Date, 4, 2) = '12' THEN 'December'
   END AS Month,
   Landing_Outcome AS Landing Outcome,
   Booster_Version AS Booster Version,
   Launch_Site AS Launch Site
FROM
   SPACEXTABLE
WHERE
   Landing_Outcome = 'Failure (drone ship)'
   AND substr(Date, 7, 4) = '2015';
```

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

- Query:
  - SELECT Landing_Outcome, COUNT(*) AS Outcome_Count
  - FROM SPACEXTABLE
  - WHERE Date >= '2010-06-04' AND Date <= '2017-03-20'
  - GROUP BY Landing_Outcome
  - ORDER BY Outcome_Count DESC;

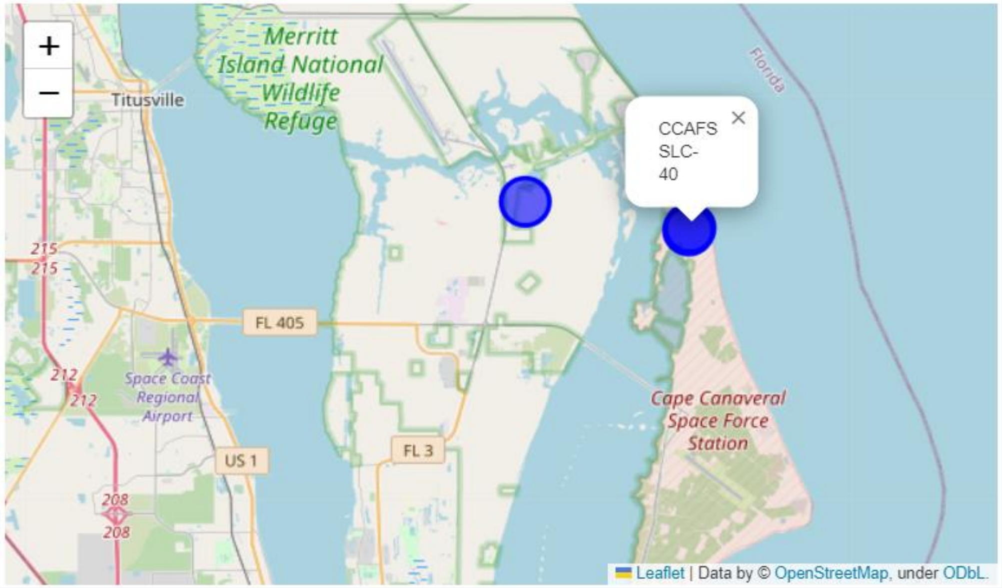
Landing_Outcome	Outcome_Count
No attempt	10
Success (ground pad)	5
Success (drone ship)	5
Failure (drone ship)	5
Controlled (ocean)	3
Uncontrolled (ocean)	2
Precluded (drone ship)	1
Failure (parachute)	1

Section 3

Launch Sites Proximities Analysis

all launch sites on a map

- We can see that All of the SpaceX launch sites are located inside the United States near the west and east coastlines.



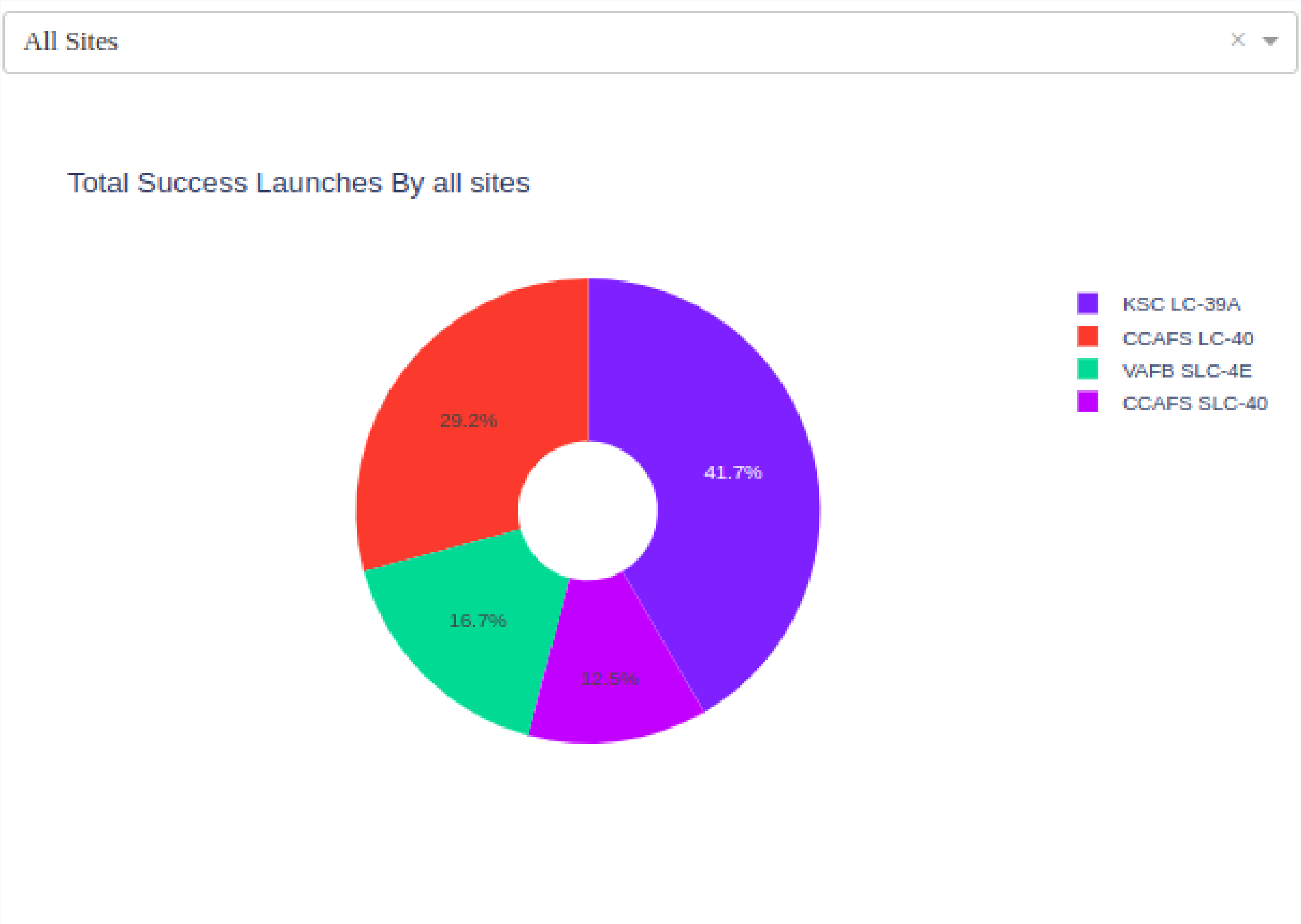
Mark the success/failed launches for each site on the map

- We can see total number of launches. If we zoom to the location , we can see the number of success or failed launches.
- The colour Green indicates success and The colour Red indicates failure.

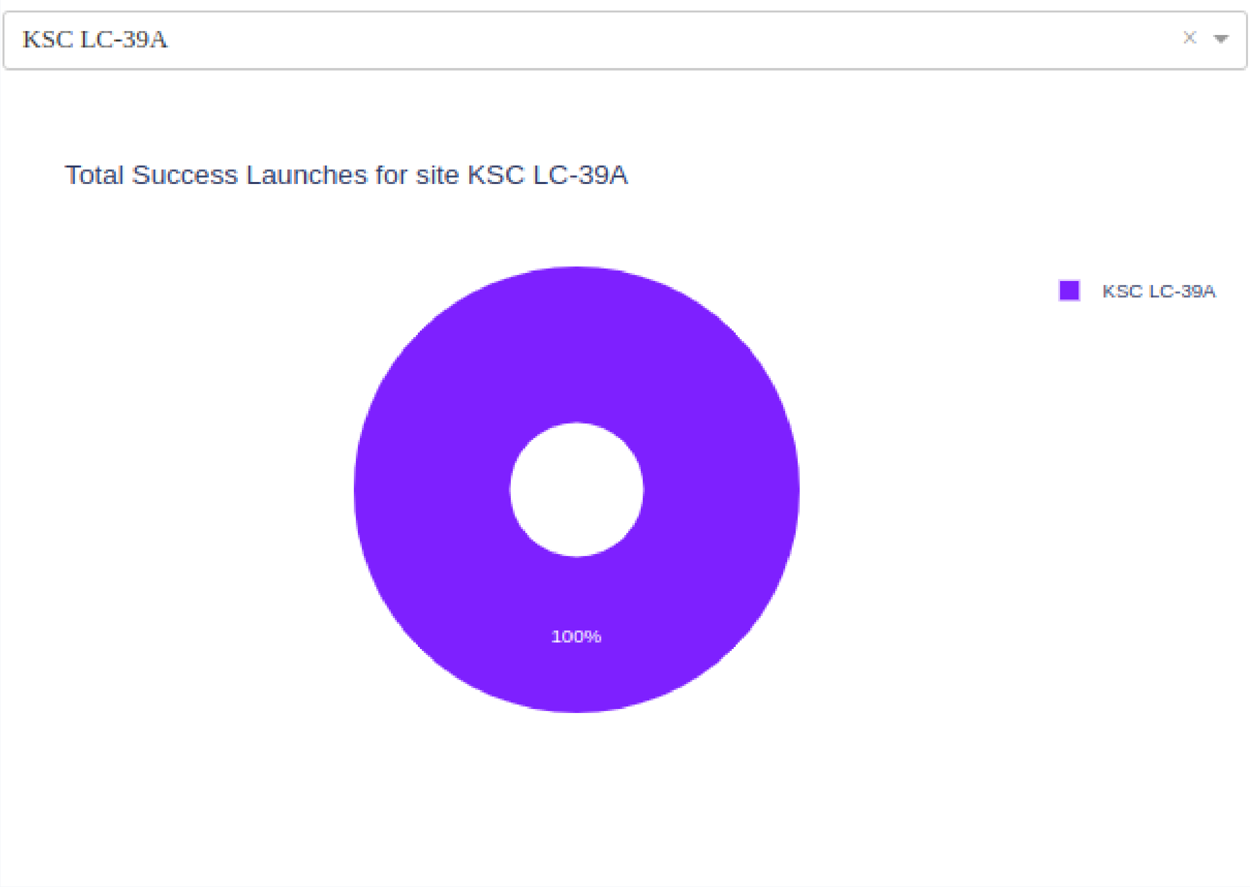
Section 4

Build a Dashboard with Plotly Dash

Total Success launches by All sites

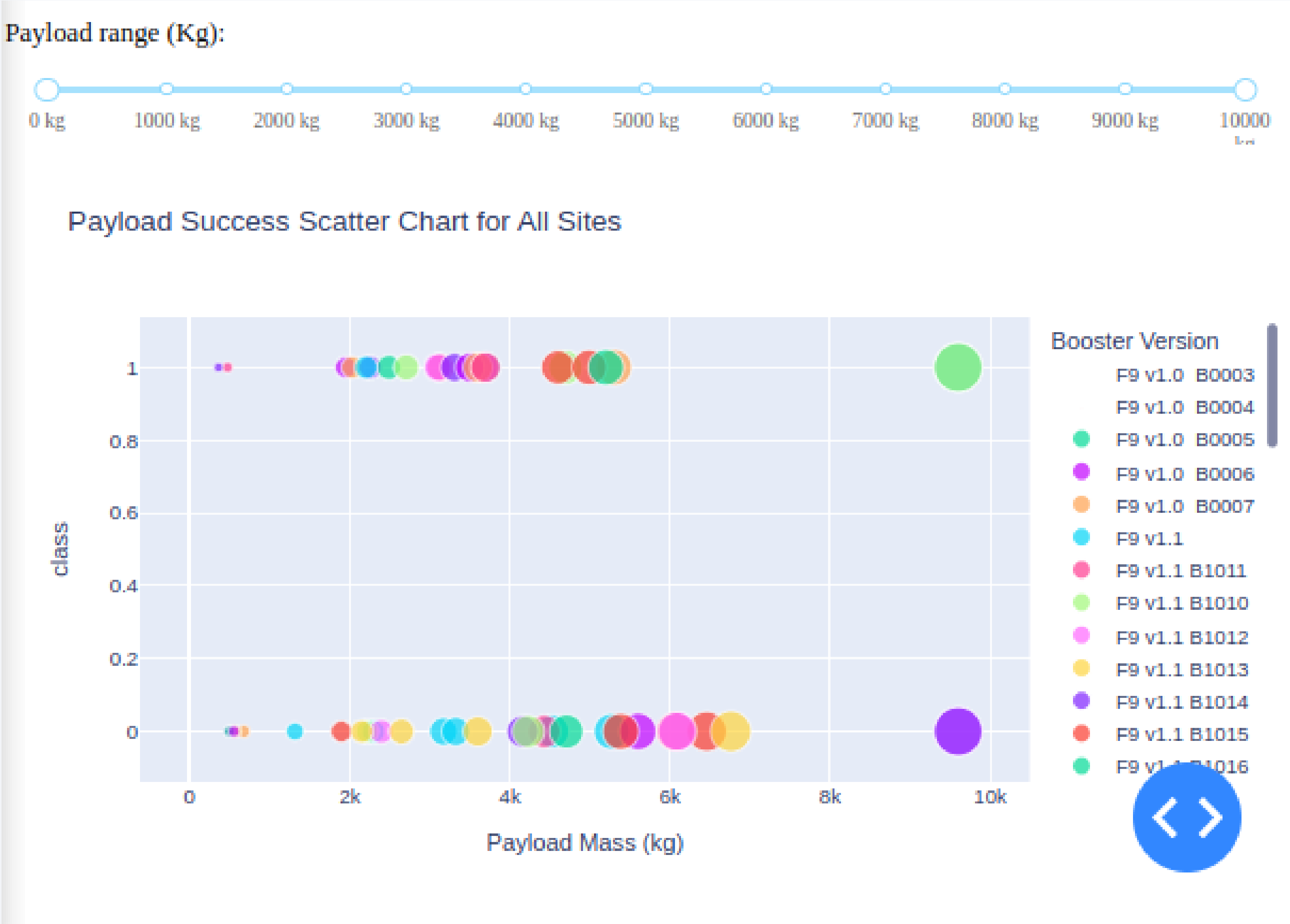


Launch site with highest launch success ratio



Launch Outcome scatter plot for all sites

- The line controller allow us to filter based on the desire weight and create a dynamic chart for us.

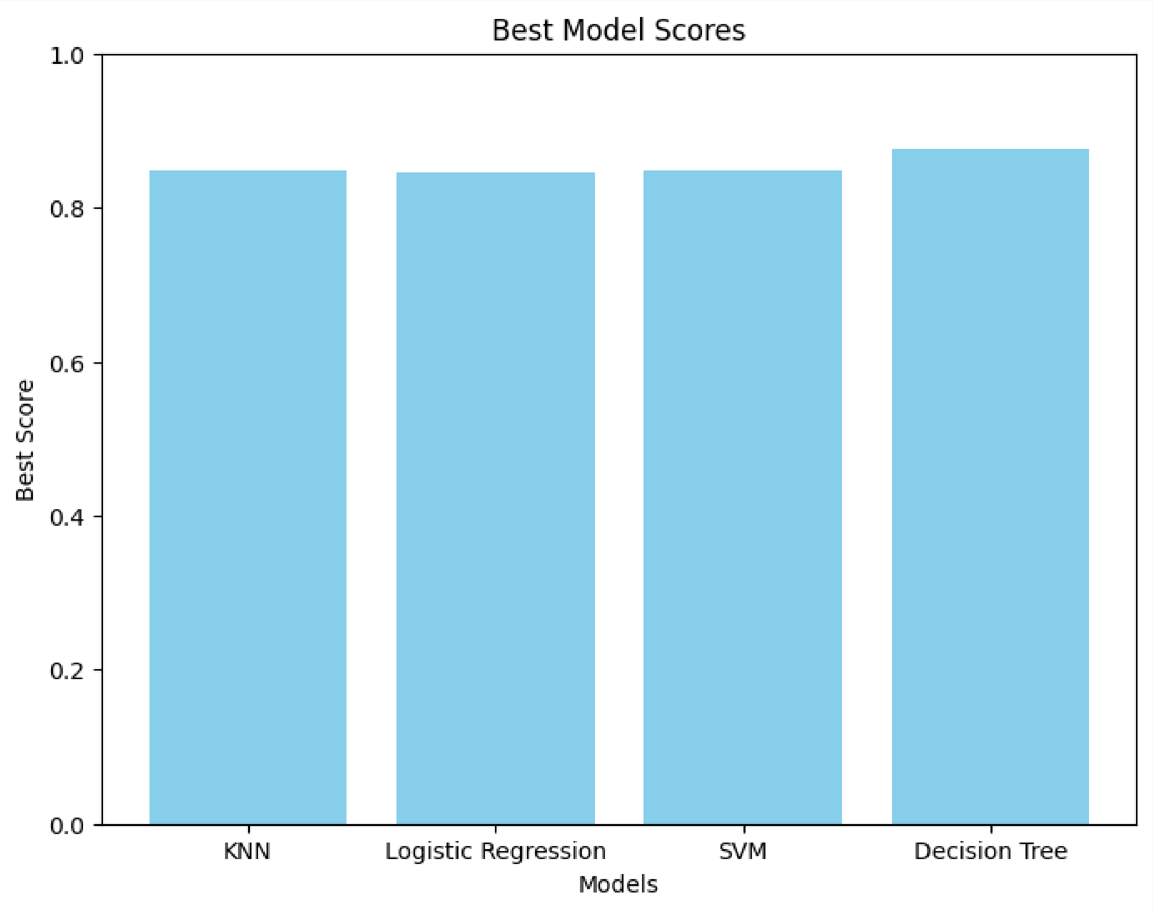


Section 5

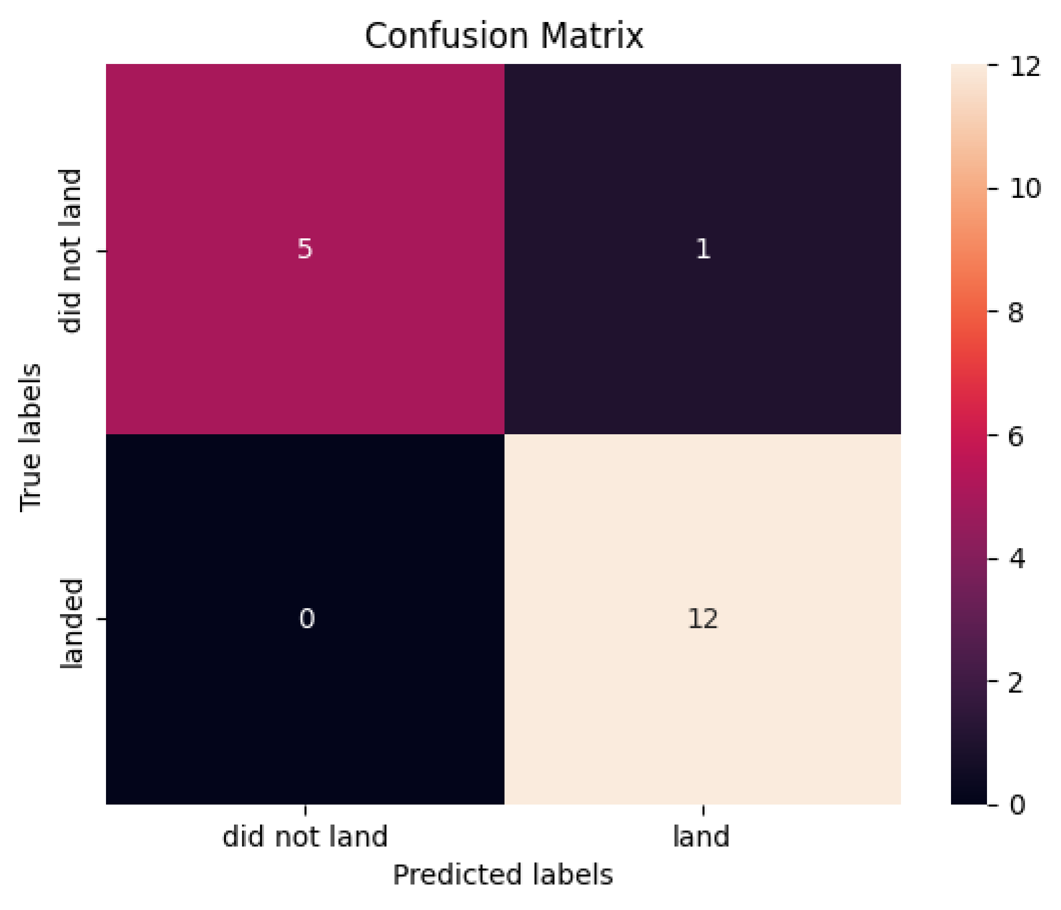
Predictive Analysis (Classification)

Classification Accuracy

- The best model is Decision Tree with higher Accuracy:
  - KNN: 0.848
  - Log Reg: 0.846
  - SVM: 0.848
  - Decision Tree: 0.876



Confusion Matrix



Conclusions

- We can conclude that:
  - The Tree Classifier Algorithm is the best Machine Learning approach for this dataset.
  - Starting from the year 2013, the success rate for SpaceX launches is increased, directly proportional time in years to 2020, which it will eventually perfect the launches in the future.
  - We understand that the larger flights increases the chance of success
  - SO orbit has 0% success rate while SSO, HEO, GEO and ES-L1 have 100% success rate

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
- References:

1. https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json
2. https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922

Thank you!