IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Saad Almalki
28 Jun 2025

Outline

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

Executive Summary

- Summary of methodologies

1- Data Collection : collects a data using SpaceX launch data API and web scraping it .
2- EDA processing (includes a data wrangling - visualization and Interactive visual analytics)
3- Using ML Techniques

- Summary of all results

1- Understanding all of falcon9 stages and visualizating the maps and geographical informations .
2- SVM and K-Nearest have a 83.33% accuracy , but Decision Tree have 94.44% accuracy .

Introduction

- Project background and context

The objective of this project is focusing in predict the successful landing of the Falcon 9 first stage .

- Problems you want to find answers

Good questions to be asked in the pre-project process stage :

- What's the best place to launch a rocket ?
- What's the best ML model to predict the landing successful ?

Section 1

Methodology

Executive Summary

- Data collection methodology:

We collected data from 2 sources :

- SpaceX API
- Web Scraping via Wikipedia

- Perform data wrangling

We wrangling and cleaning a data to involves handling the errors and missing values .

Methodology

- Perform interactive visual analytics using Folium and Plotly Dash
  - Used Folium to create interactive visual analytics and plotly dash app to analysis the visual analytics .
- Perform predictive analysis using classification models

Data Collection

- Describe how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts

We collected data from 2 sources :

- SpaceX API
- Web Scraping via Wikipedia

Data Collection - API

Requesting API

Filtering a data to receives only falcon9 results

Working and Fixing the problems

Data Collection - Web Scraping

Requesting The wiki page → Extracting the content from HTML → Creating dataframe by HTML Parsing

Data Wrangling

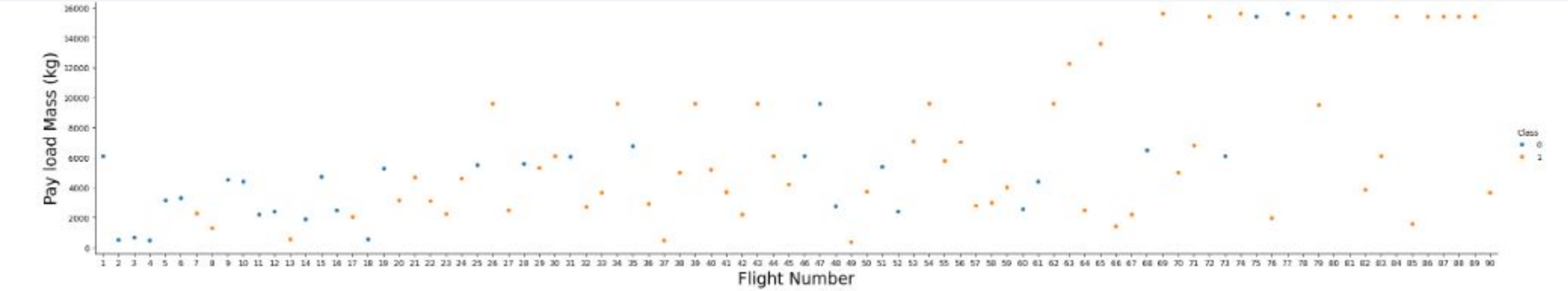
EDA processes → Summarizing → Working on landing outcomes per orbit type

To see github repository :
https://github.com/Saad711T/IBM-Applied-Data-Science-Capstone

EDA with Data Visualization

- I’m using scatterplots to visualize the relations between many of features in the analytics .

    To see github repository :
    https://github.com/Saad711T/IBM-Applied-Data-Science-Capstone/blob/main/Labs/jupyter-labs-eda-dataviz-v2.ipynb



EDA with SQL

Using SQL queries with SQLite to analyzing and summarizing some points like :

1. Perfect and failure launches
2. Filtering a data to focus on specific launches
3. Sorting a data to identifying what the useful data and not
4. Nested and Subqueries useful in data analysis using SQL

To see github repository :
https://github.com/Saad711T/IBM-Applied-Data-Science-Capstone/blob/main/Labs/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

Making an interactive map visualization with folium library using Markers, lines ,circles, and a clusters.

- Markers to identify a launch site
- Circles to identify an areas
- Marker clusters to identify any group of events happened in launch site or something like this
- Lines used to know a distance and space between areas.

To see github repository :
https://github.com/Saad711T/IBM-Applied-Data-Science-Capstone/blob/main/Labs/lab-jupyter-launch-site-location-v2.ipynb

Build a Dashboard with Plotly Dash

Using plotly library to make a dashboard hosting a visualizations about :

- Percentage of launches by site
- Range of a payload

Plotly is allowed to interactive and communicate between user and visualizations with a web app dashboard

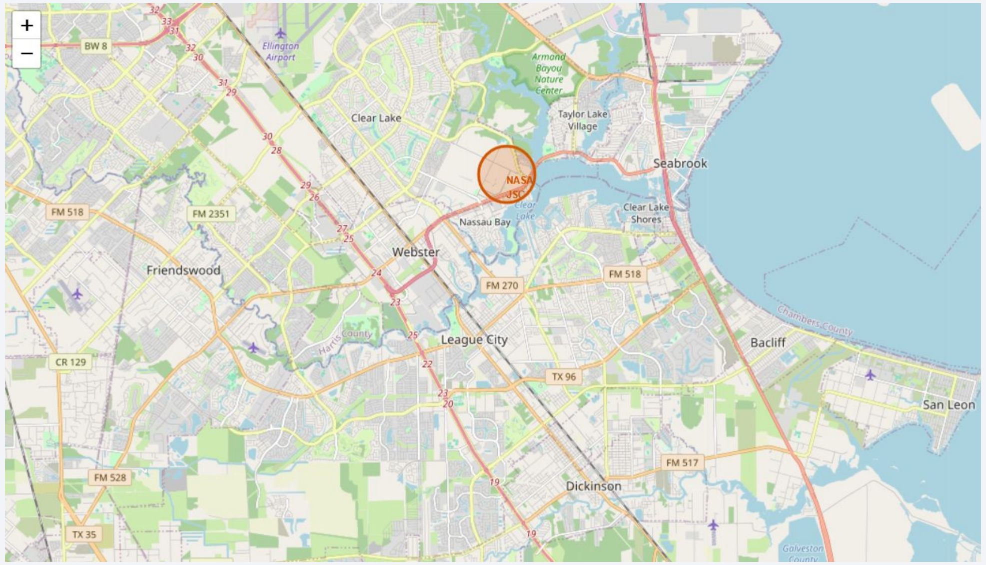
To see github repository :
https://github.com/Saad711T/IBM-Applied-Data-Science-Capstone/blob/main/Labs/spacex_dash_app.py

Predictive Analysis (Classification)

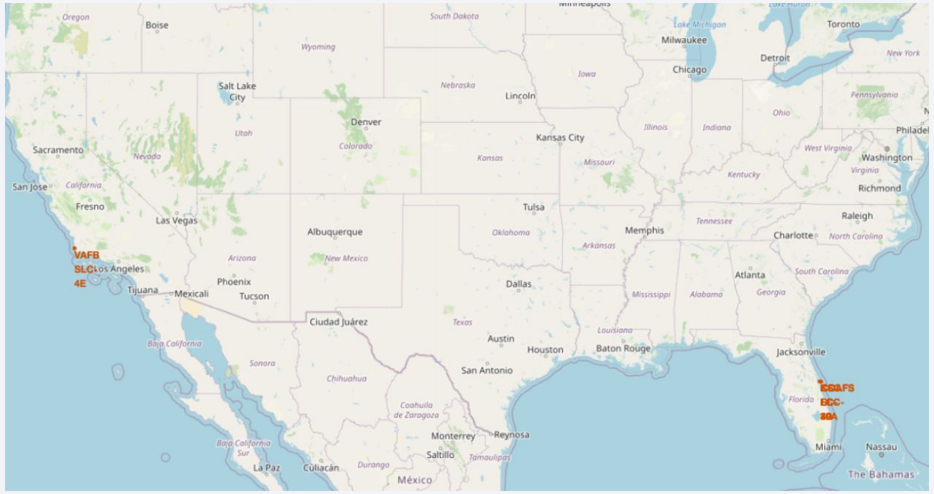
Data preparation → Test of each ML models → Comparison the results

To see github repository :
https://github.com/Saad711T/IBM-Applied-Data-Science-Capstone/blob/main/Labs/SpaceX-Machine-Learning-Prediction-Part-5-v1.ipynb

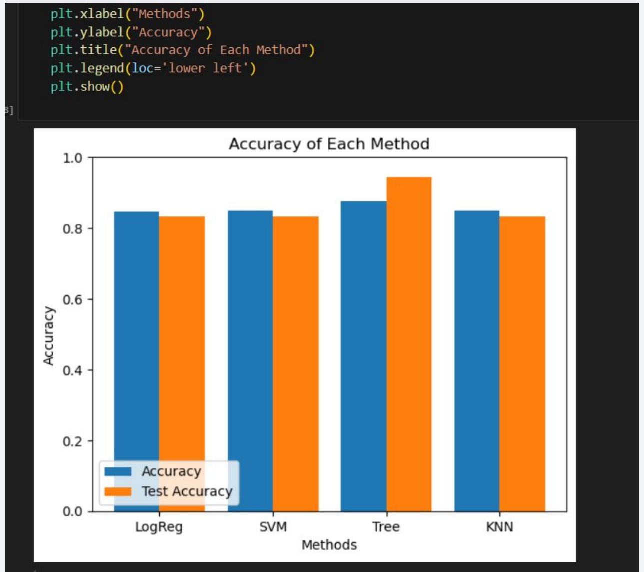
Results



Clear Lake
Taylor Lake Village
Seabrook
NASA JSC
Nassau Bay
Clear Lake Shores
Webster
Friendswood
League City
Bacliff
San Leon
Dickinson
Ellington Airport
Armand Bayou Nature Center
FM 518
FM 2351
FM 270
CR 129
TX 96
FM 517
FM 528
TX 35
BW 8



VAFB
SLC
4E
CCAFS
SLC-
40A
Boise
Salt Lake City
Denver
Sacramento
San Jose
Fresno
Las Vegas
Los Angeles
Tijuana
Mexicali
Phoenix
Tucson
Albuquerque
Ciudad Juárez
Lincoln
Kansas City
Tulsa
Dallas
Austin
Houston
San Antonio
Memphis
Chicago
Milwaukee
Detroit
Toronto
Atlanta
Charlotte
Raleigh
Richmond
Washington
Jacksonville
Baton Rouge
Miami
Nassau
The Bahamas
Monterrey
Reynosa
Saltillo
México
La Paz
Culiacán



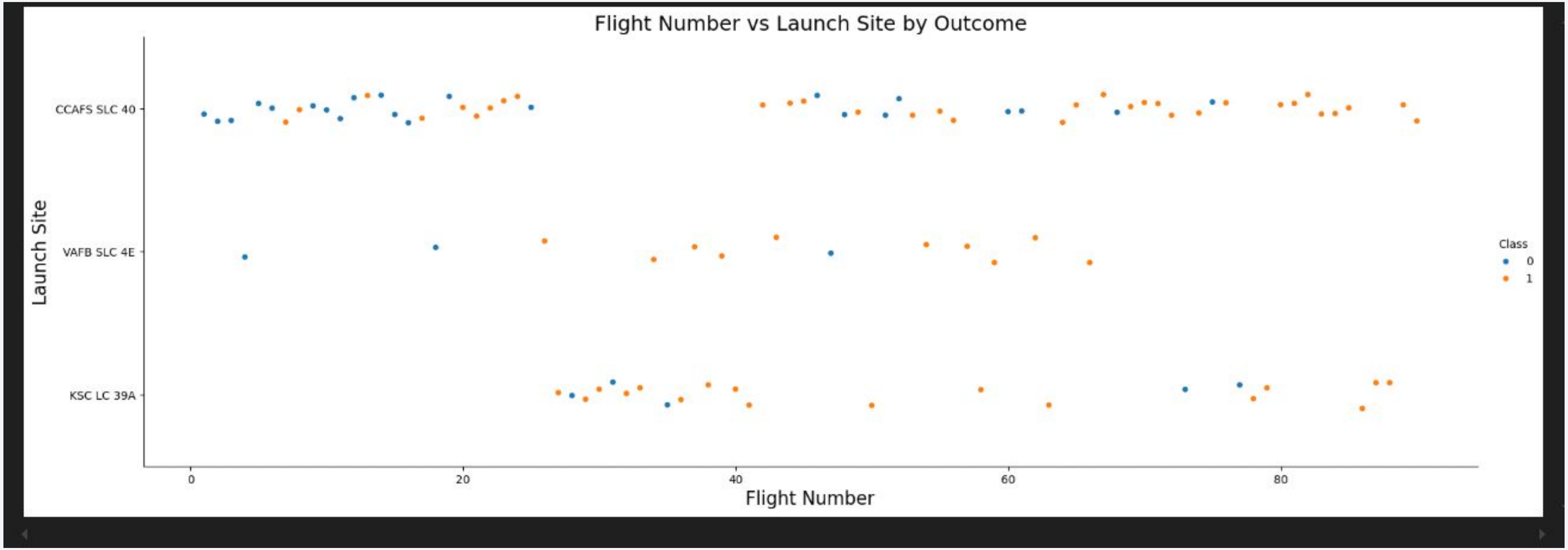
plt.xlabel("Methods")
plt.ylabel("Accuracy")
plt.title("Accuracy of Each Method")
plt.legend(loc='lower left')
plt.show()
Accuracy of Each Method
Accuracy
Test Accuracy
LogReg
SVM
Tree
KNN
Methods
Accuracy
0.0
0.2
0.4
0.6
0.8
1.0

Section 2

Insights drawn from EDA

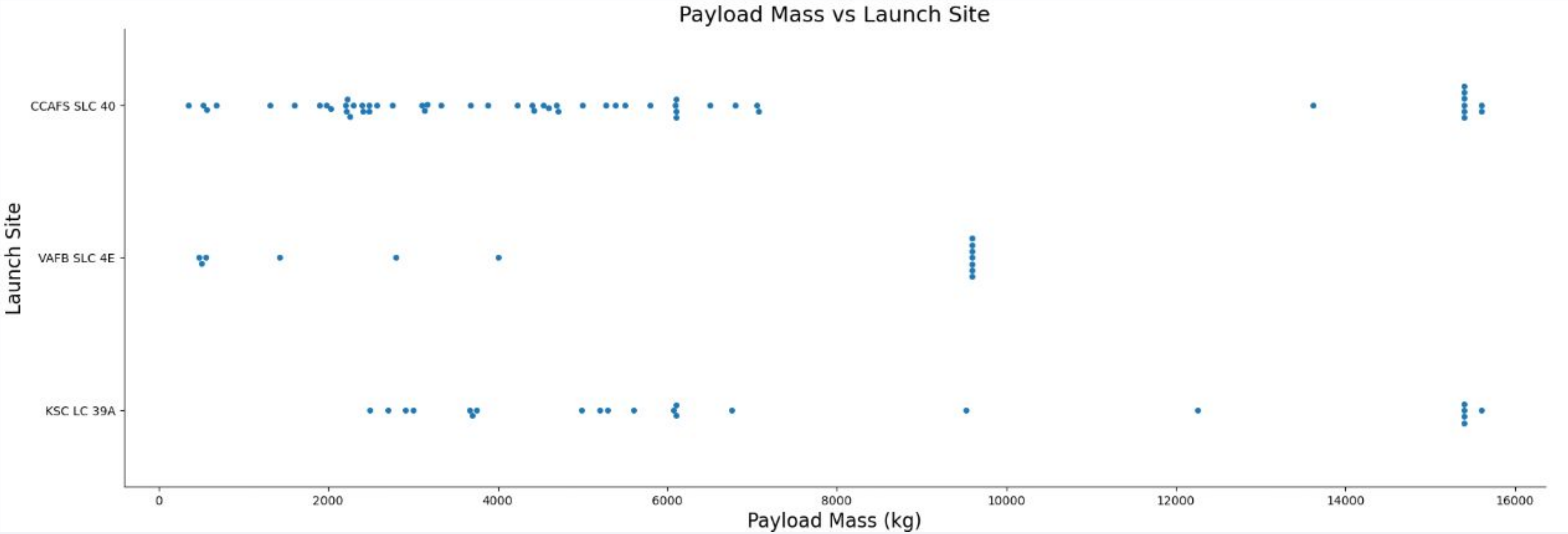
Flight Number vs. Launch Site

- Visualization :



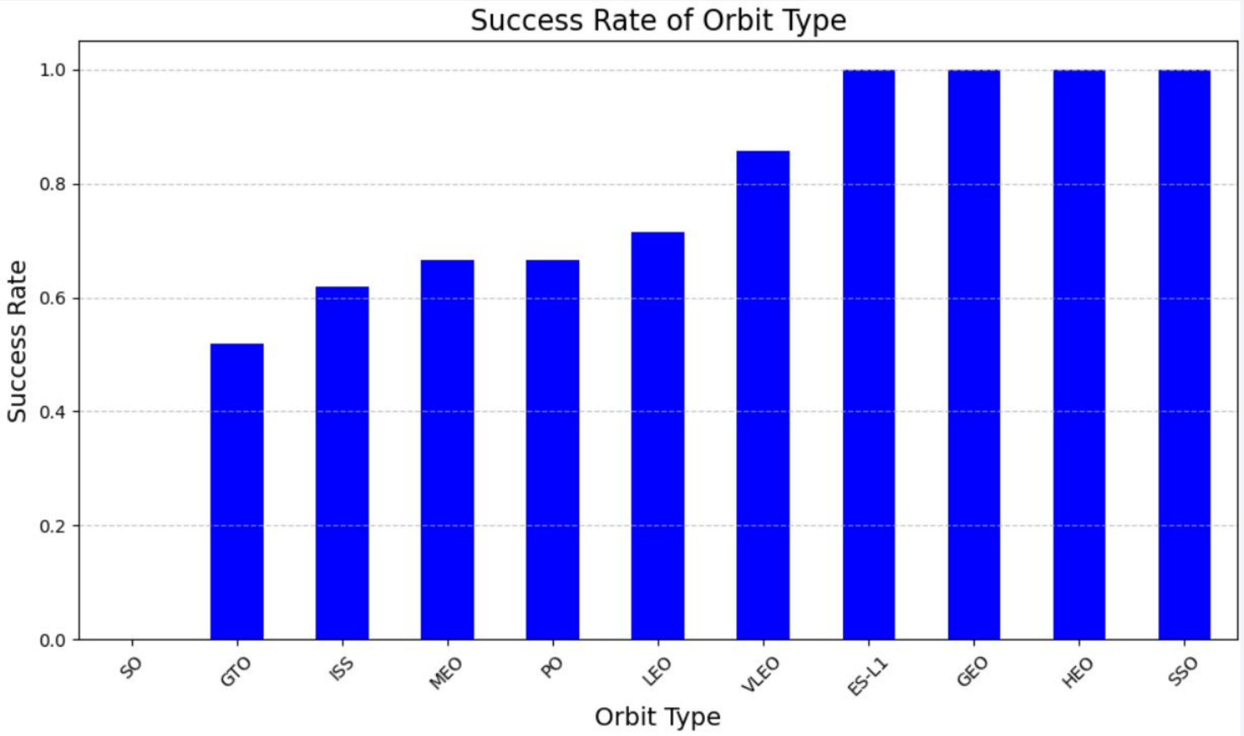
Payload vs. Launch Site

- Visualization :



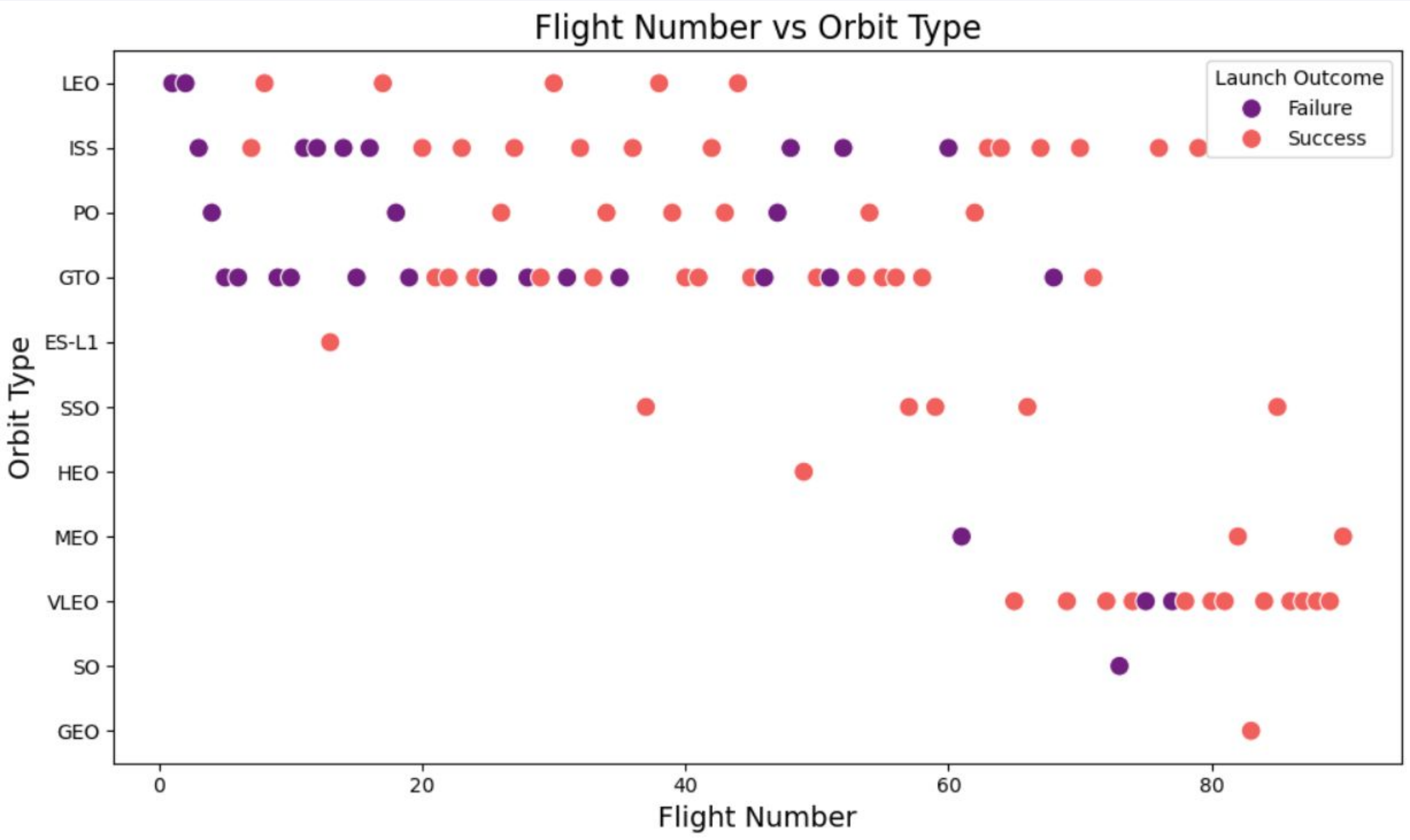
Success Rate vs. Orbit Type

Visualization :



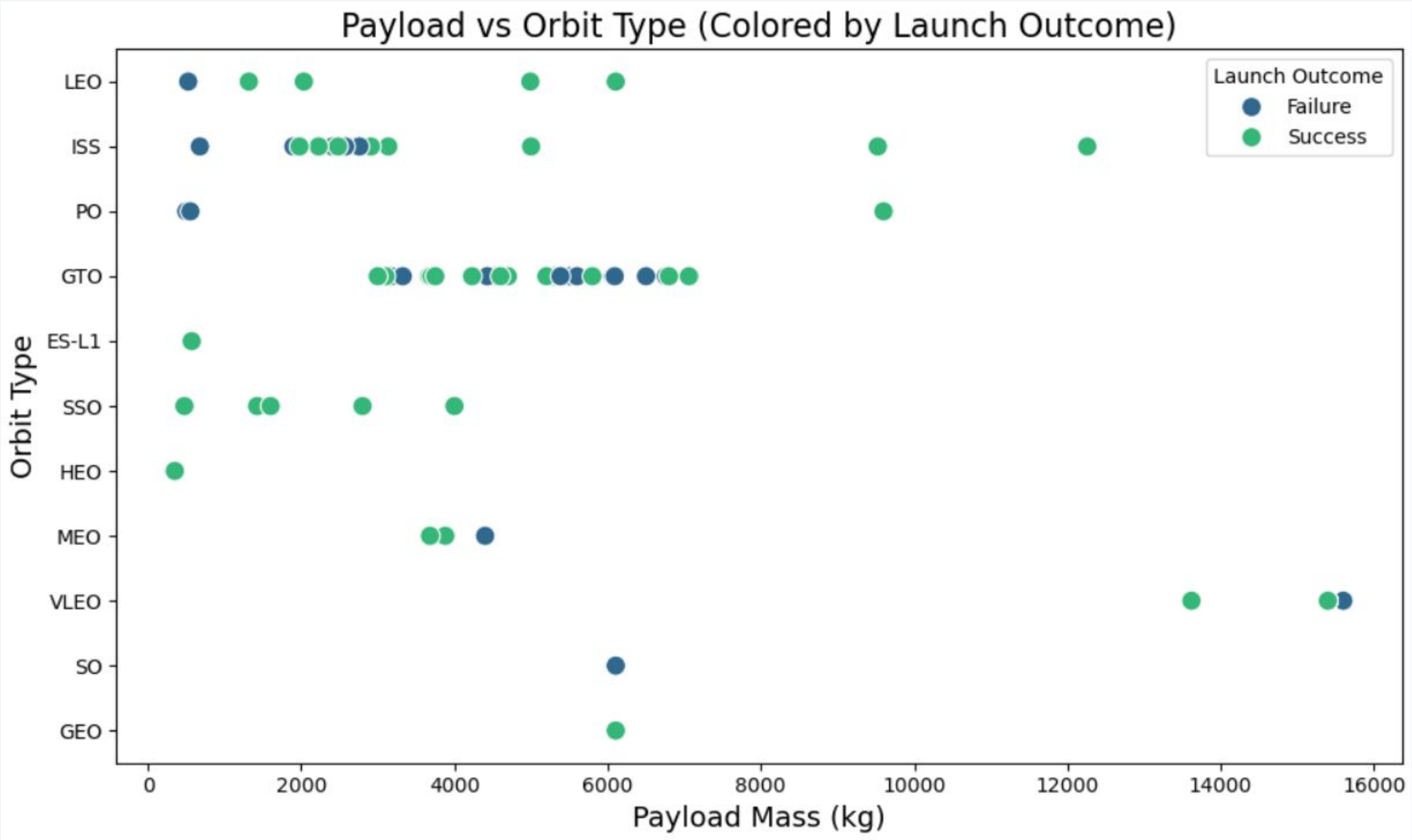
Flight Number vs. Orbit Type

- Visualization :



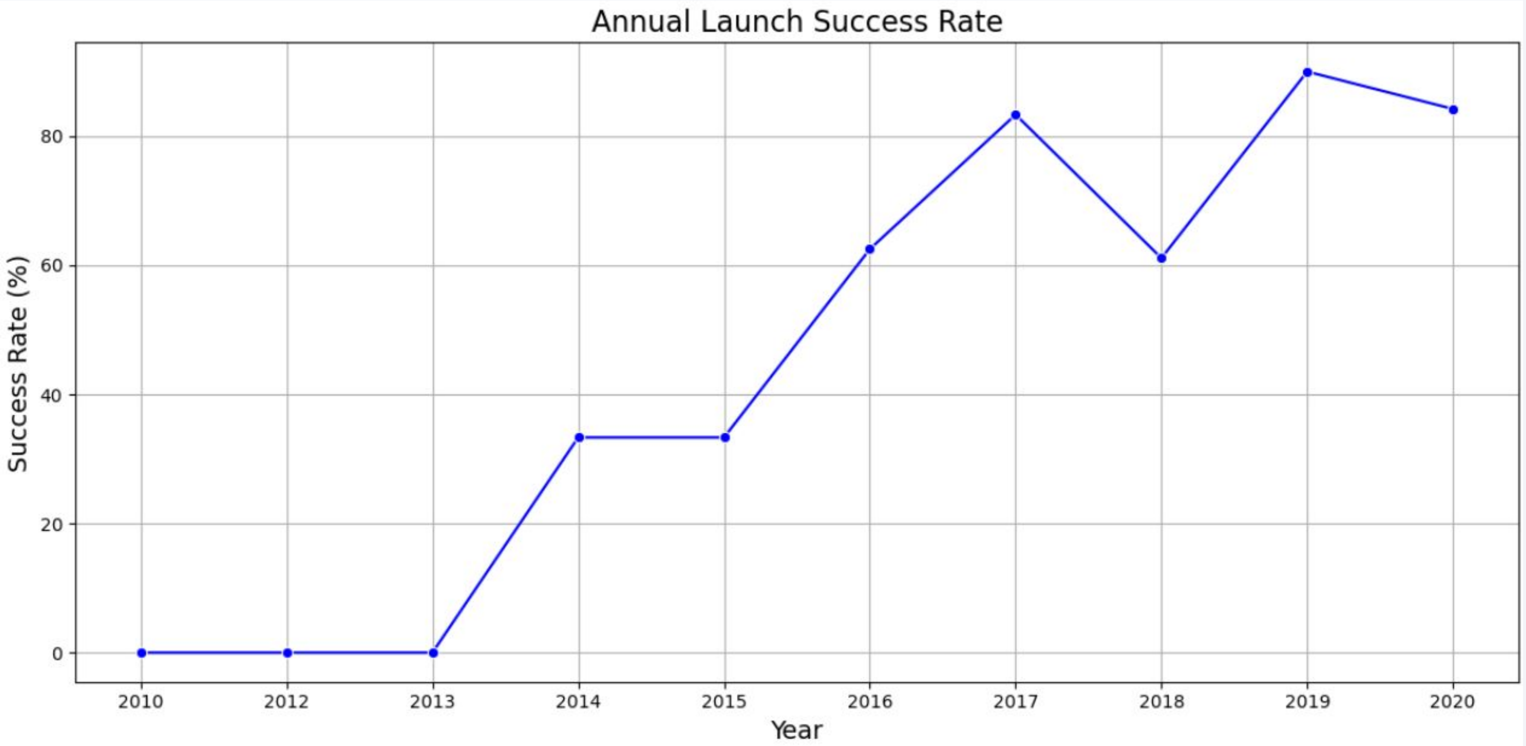
Payload vs. Orbit Type

- Visualization :



Launch Success Yearly Trend

- Visualization :



All Launch Site Names

- Using SQL query :

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

Launch Site Names Begin with 'CCA'

- Using SQL Query :

Done.

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

- Using SQL Query :

Task 3

Display the total payload mass carried by boosters launched by NASA (CRS)

[12]:
```
sql SELECT SUM(PAYLOAD_MASS__KG_) AS TOTAL_PAYLOAD FROM SPACEXTBL WHERE PAYLOAD  LIKE '%CRS%';
```

```
 * sqlite:///my_data1.db
Done.
```

[12]:

TOTAL_PAYLOAD
111268

Average Payload Mass by F9 v1.1

- Using SQL Query :

Task 4

Display average payload mass carried by booster version F9 v1.1

```
[13]: sql SELECT AVG(PAYLOAD_MASS__KG_) AS AVG_PAYLOAD FROM SPACEXTBL WHERE BOOSTER_VERSION = 'F9 v1.1';
```

```
 * sqlite:///my_data1.db
Done.
```

[13]:

AVG_PAYLOAD
2928.4

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- Present your query result with a short explanation here

Task 5

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

```
[14]: sql SELECT MIN(DATE) AS FIRST_SUCCESS_GP FROM SPACEXTBL WHERE LANDING__OUTCOME = 'Success (ground pad)';
```

```
 * sqlite:///my_data1.db
(sqlite3.OperationalError) no such column: LANDING__OUTCOME
[SQL: SELECT MIN(DATE) AS FIRST_SUCCESS_GP FROM SPACEXTBL WHERE LANDING__OUTCOME = 'Success (ground pad)' ;]
(Background on this error at: https://sqlalche.me/e/20/e3q8)
```

Successful Drone Ship Landing with Payload between 4000 and 6000

- Present your query result with a short explanation here

Task 6

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

[15]:
```
sql SELECT DISTINCT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000 AND LANDING__OUTCOME = 'Success (drone ship)'
```

```
 * sqlite:///my_data1.db
(sqlite3.OperationalError) no such column: LANDING__OUTCOME
[SQL: SELECT DISTINCT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000 AND LANDING__OUTCOME = 'Success (drone shi
p)' ;]
(Background on this error at: https://sqlalche.me/e/20/e3q8)
```

Total Number of Successful and Failure Mission Outcomes

- Present your query result with a short explanation here

Task 7

List the total number of successful and failure mission outcomes

[16]:
```
sql SELECT MISSION_OUTCOME, COUNT(*) AS QTY FROM SPACEXTBL GROUP BY MISSION_OUTCOME ORDER BY MISSION_OUTCOME;
```

```
 * sqlite:///my_data1.db
Done.
```

[16]:

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

Boosters Carried Maximum Payload

- Present your query result with a short explanation here

[17]:
```
sql SELECT DISTINCT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL) ORDER BY BOOSTER_V
```

```
 * sqlite:///my_data1.db
Done.
```

[17]:

Booster_Version
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

2015 Launch Records

- Present your query result with a short explanation here

Month_Name	Mission_Outcome	Booster_Version	Launch_Site
January	Success	F9 v1.1 B1012	CCAFS LC-40
February	Success	F9 v1.1 B1013	CCAFS LC-40
March	Success	F9 v1.1 B1014	CCAFS LC-40
April	Success	F9 v1.1 B1015	CCAFS LC-40
April	Success	F9 v1.1 B1016	CCAFS LC-40
June	Failure (in flight)	F9 v1.1 B1018	CCAFS LC-40
December	Success	F9 FT B1019	CCAFS LC-40

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

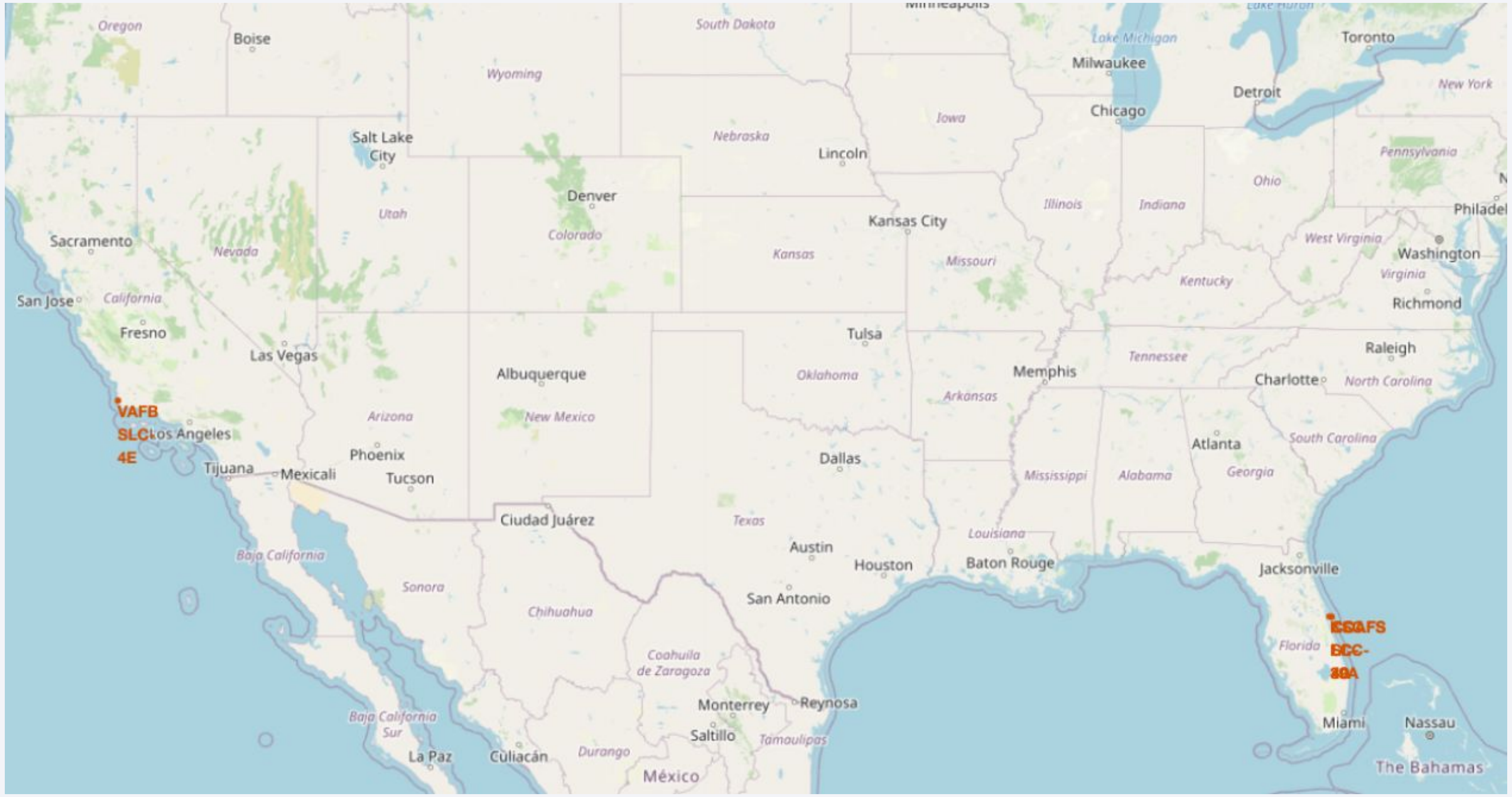
- Present your query result with a short explanation here

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

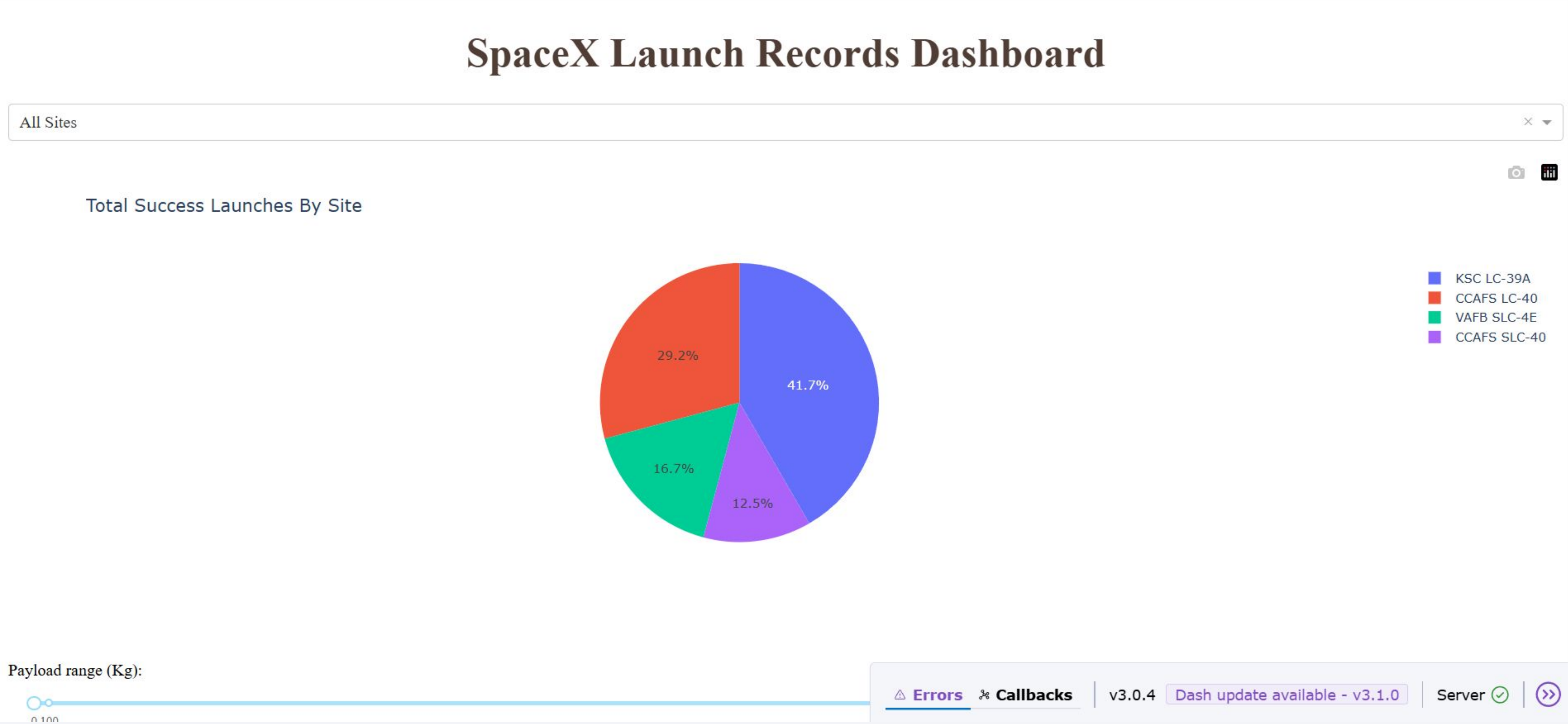
Folium map - Locations



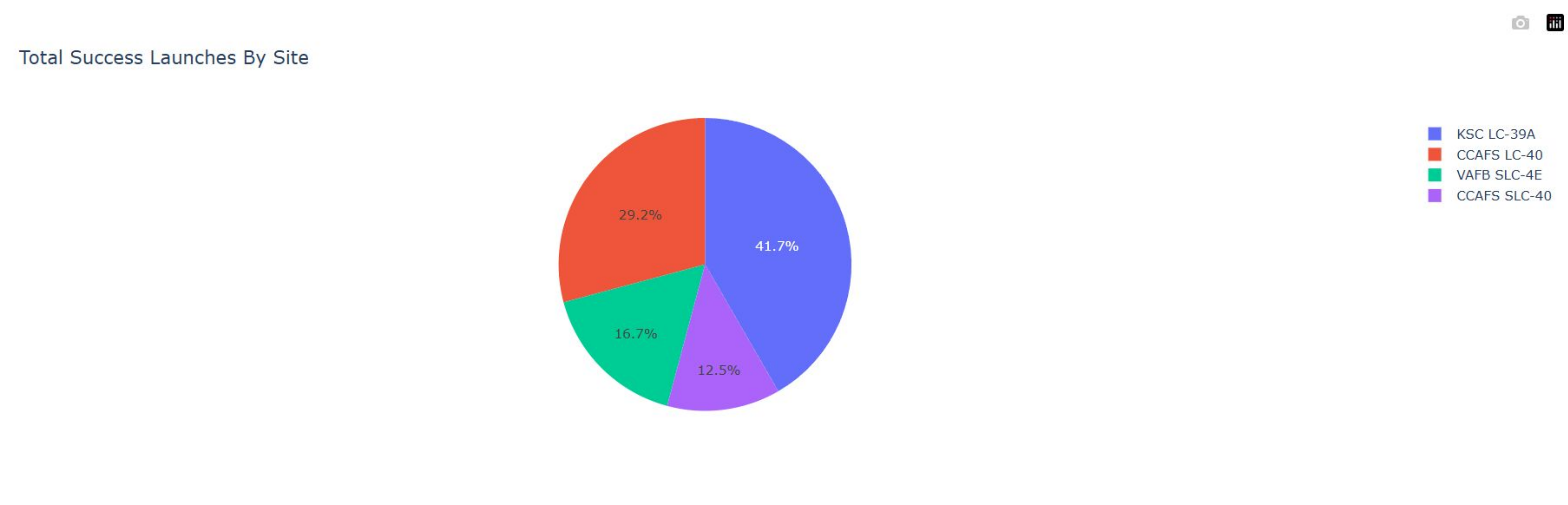
Section 4

Build a Dashboard with Plotly Dash

Screenshot of dashboard



Pie chart - Total Success launches by site



Scatter - payload mass between for all sites

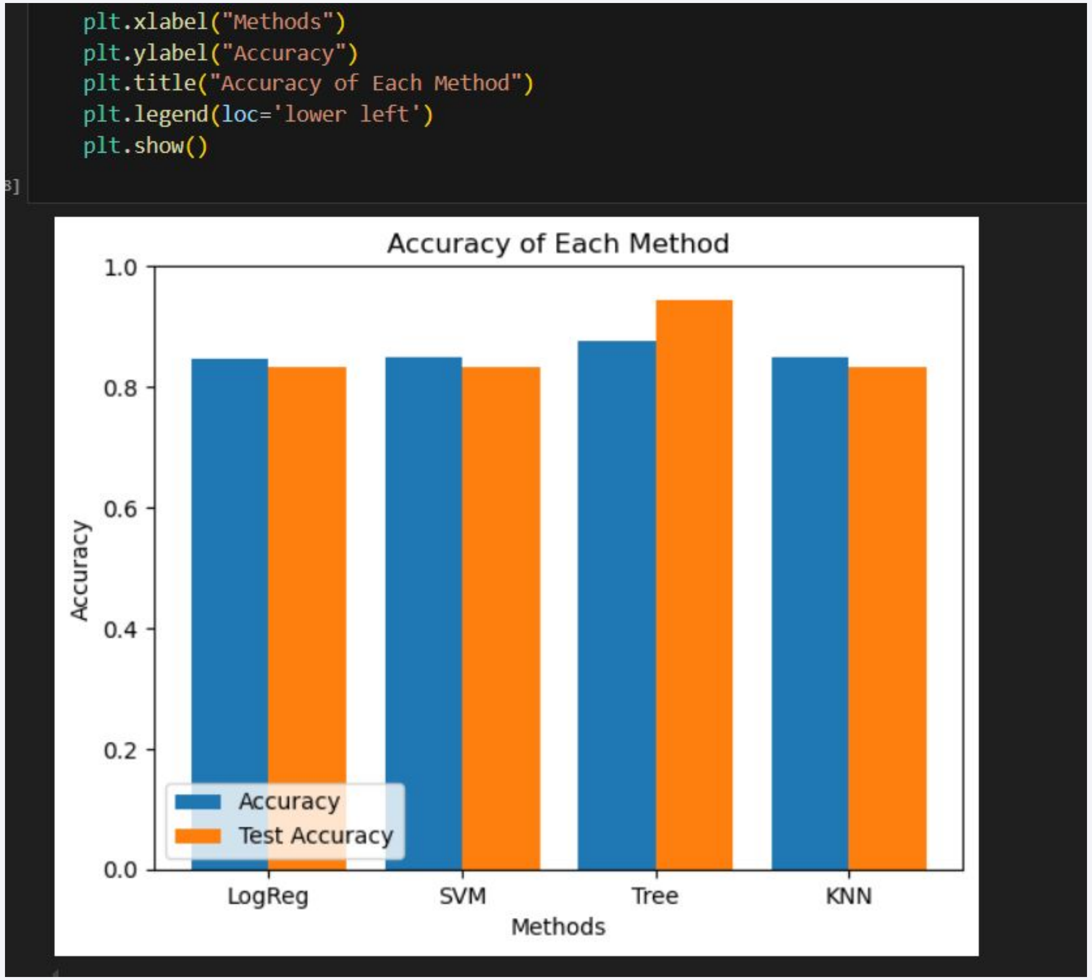
Section 5

Predictive Analysis (Classification)

Classification Accuracy

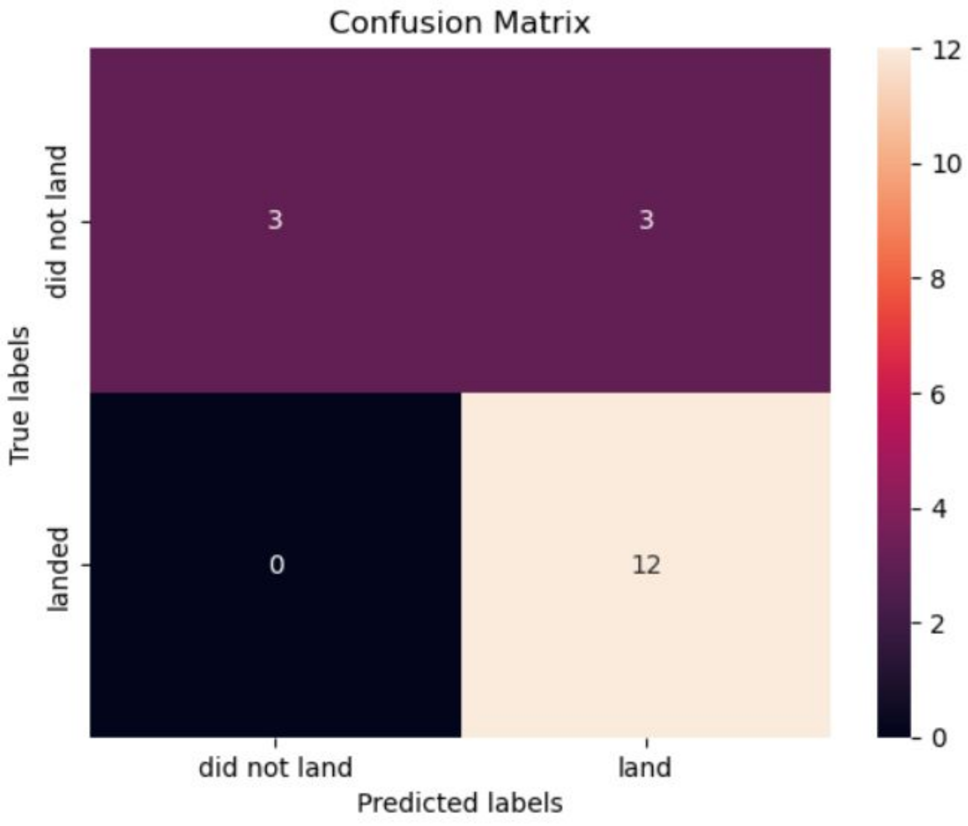
- Visualization using Pandas-Matplotlib for ML models :

- We notice the ML model have a higher classification accuracy is Tree.



Confusion Matrix

•



Conclusions

- Launches have more 7000KG are less riskly
- The tree is higher classification more than other ML models.
- The average payload mass equals “2928.4”
- The total of payload mass equals “111268”
- Mission outcome of June 2015 have a failure.

Thank you!