Winning Space Race with Data Science

Anna Balmer
06/03/2023

Outline

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

Executive Summary

- Summary of methodologies(See methodology section for more details)
- Collected data utilizing public Space X API and scraping Space X Wikipedia page
- Classified landing outcomes as successful(class=1) and unsuccessful(class=0)
- Build models by standardizing the data and finding the best parameters using GridSearchCV
- Summary of all results(see results section for more details)
- Orbits ES-L1, GEO, HEO, and SSO have the highest success rate.
- KSC LC-39A has the highest success rate.
- All of the models(Logistic Regression, SVM, Decision Tree, and KNN) have a accuracy score of proximately 83.33%.

Introduction

- Project background and context
- During the commercial space age companies are making space travel more affordable.
- Space X conducts the most inexpensive launches ($62 million vs $165 million).
- This is due mostly to Space X being able to reuse the first stage of the rocket.
- Space Y wants to compete with Space X by using data science to determine what factors depend on the outcome.

- Problems you want to find answers
- Whether or not Stage One will land successfully.
- By collecting data and training a machine learning model.

Section 1

Methodology

Methodology

Executive Summary

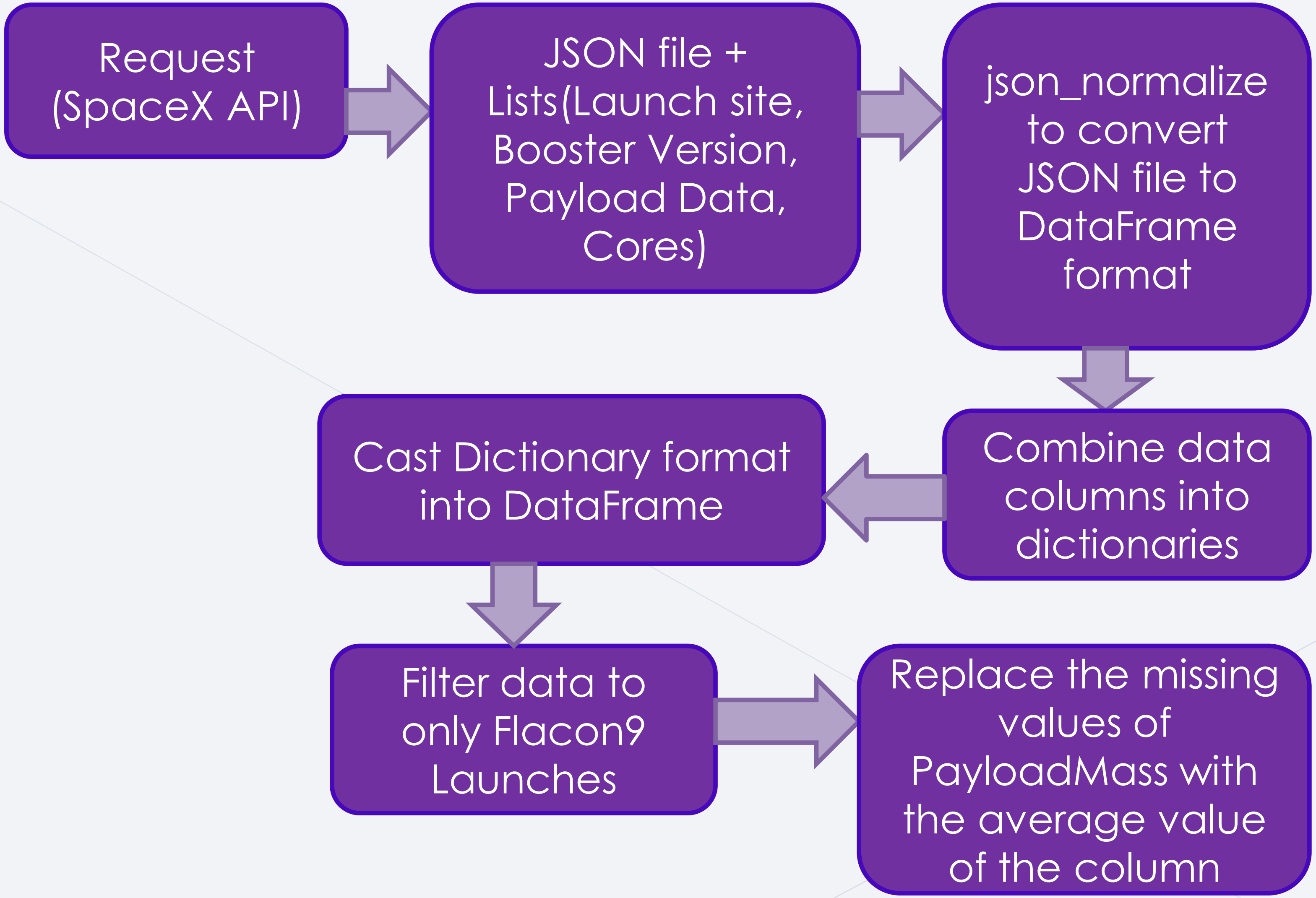
- Data collection methodology:
  - Collected data utilizing public Space X API and scraping Space X Wikipedia page
- Perform data wrangling
  - Classified landing outcomes as successful(class=1) and unsuccessful(class=0)
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Build models by standardizing the data and finding the best parameters using GridSearchCV

Data Collection

- The data collection technique comprised a combination of API queries from Space X's public API and web scraping a table from Space X's Wikipedia entry.
- The next slides describe flowcharts for data gathering from the public Space X API and data collection using web scraping.

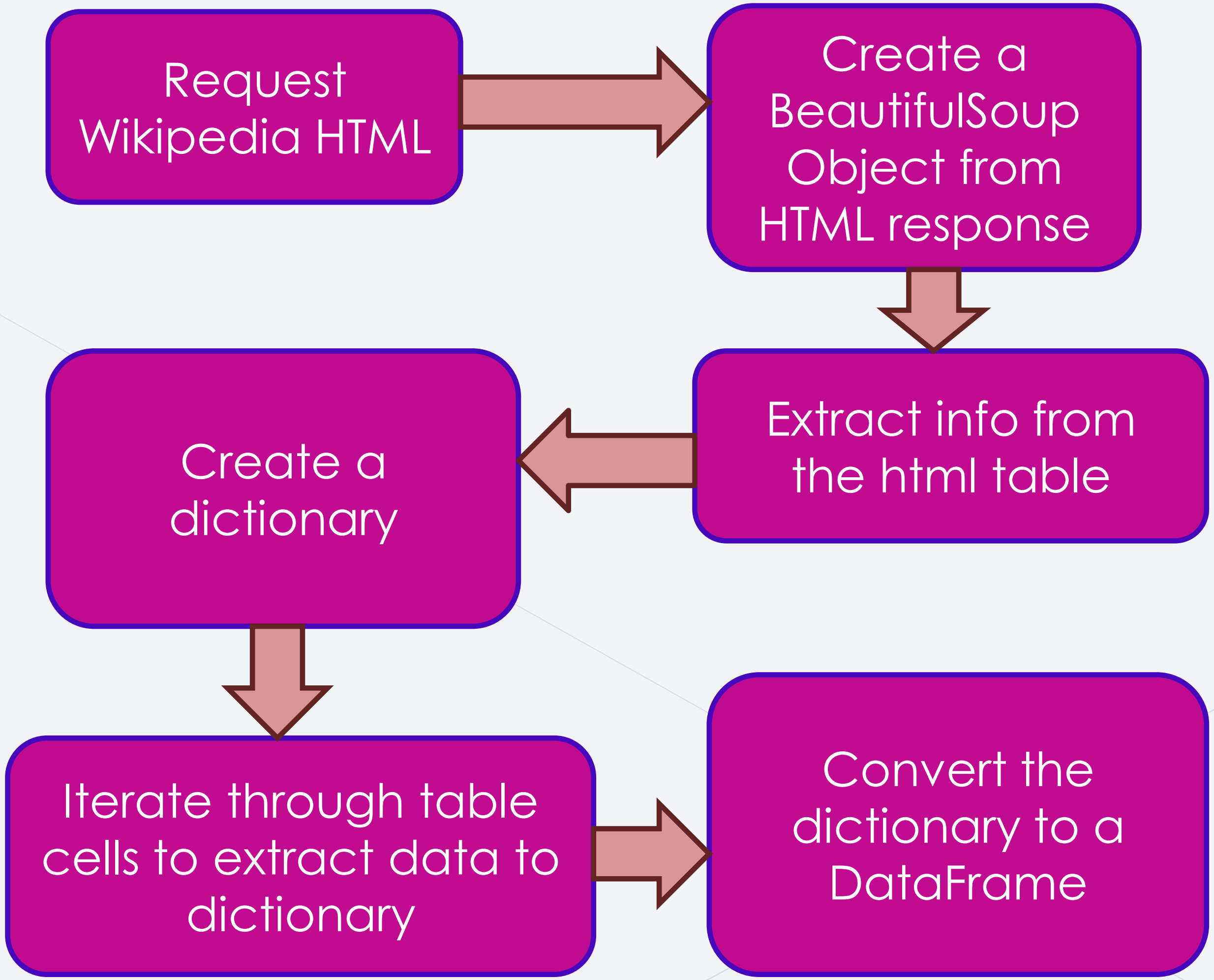
Data Collection – SpaceX API

- GitHub URL
- https://github.com/Ladyamethyst811/Data-Sience-Capstone/blob/main/jupyter-labs-spacex-data-collection-api-bak-2023-05-30-19-14-19Z.ipynb



Data Collection - Scraping

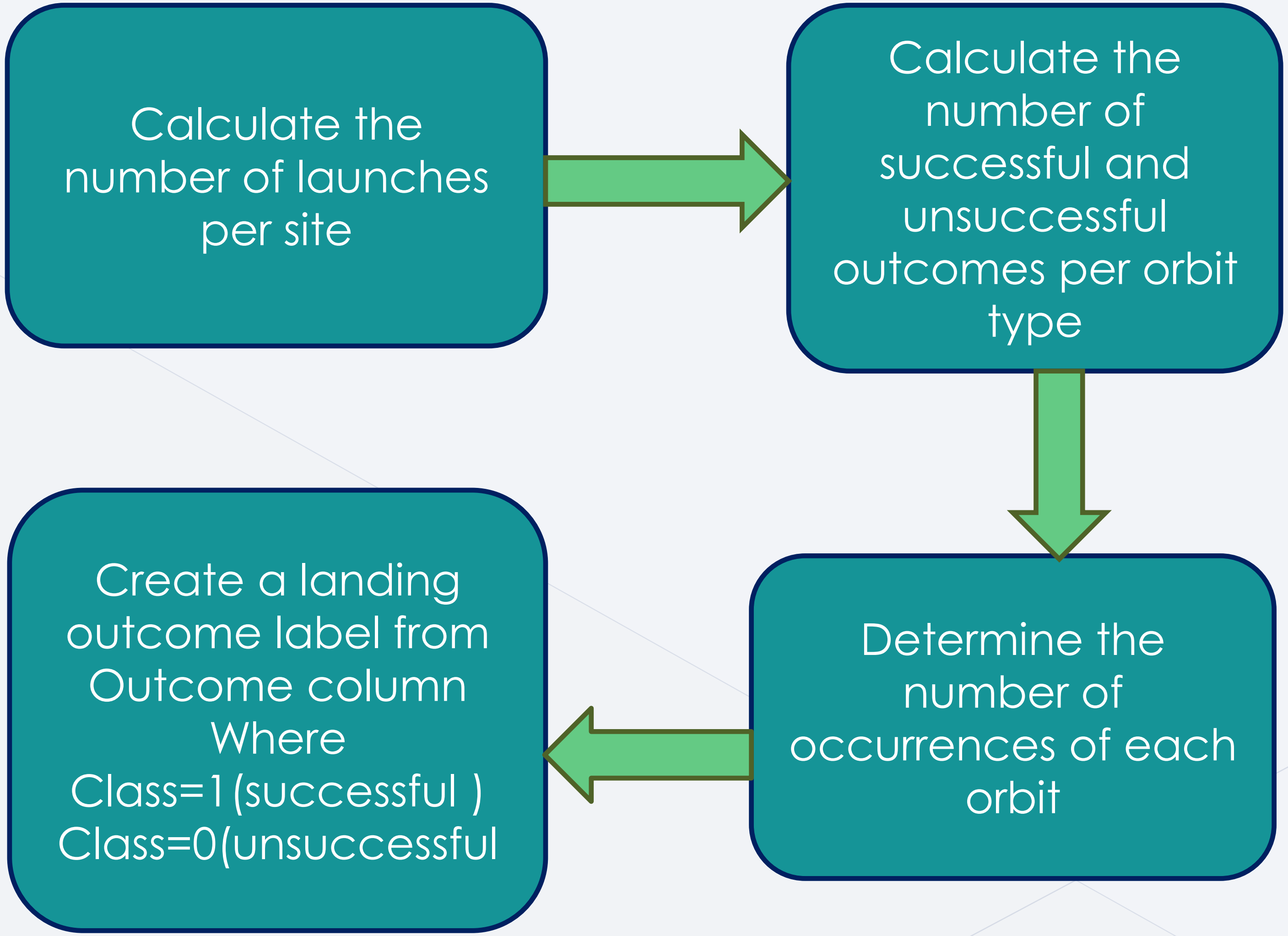
- GitHub URL
- https://github.com/Ladyamethyst811/Data-Science-Capstone/blob/main/jupyter-labs-webscraping(1).ipynb



Data Wrangling

GitHub URL

https://github.com/Ladyamethyst811/Data-Science-Capstone/blob/main/IBM-DS0321EN-SkillsNetwork_labs_module_1_L3_labs-jupyter-spacex-data_wrangling_jupyterlite.jupyterlite(1).ipynb



EDA with Data Visualization

Plotted Charts:

- **Scatter Point Plot of FlightNumber vs PayloadMass:**
- **To determine if the number of attempts and amount of weight would affect the launch outcome.**
- **Scatter Point Plot of FlightNumber vs LaunchSite:**
- **To determine if the number of attempts and the location of the launch has any bearing on the outcome.**
- **Scatter Point Plot of Payload vs LaunchSite:**
- **To observe if there is a relationship between launch sites and payload mass.**
- **Bar chart of Orbit vs Success Rate:**
- **To determine which orbit types have the highest success rate.**
- **Scatter Point Plot of FlightNumber vs Orbit Type:**
- **To determine if there is a relationship between the number of flights and the type of orbit.**
- **Scatter Point Plot of Payload vs Orbit Type:**
- **To determine if payload mass has any bearing on the success or failure of the orbit type.**
- **Line chart of Average Success Rate vs Year:**
- **To observe the average launch success trend.**

GitHub URL

https://github.com/Ladyamethyst811/Data-Science-Capstone/blob/main/IBM-DS0321EN-SkillsNetwork_labs_module_2_jupyter-labs-eda-dataviz.ipynb.jupyterlite(1).ipynb

EDA with SQL

- SQL Queries:

- **Display the names of the unique launch sites.**
- **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 successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.**

GitHub URL

https://github.com/Ladyamethyst811/Data-Science-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite(1).ipynb

Build an Interactive Map with Folium

Objects on Folium Maps such as markers, circles, and lines helps us locate the
launch sites highlight successful and unsuccessful launches and the proximity to key locations such as: Cities, Railways, Coastlines, and Highways.

Adding objects to maps helps us visualize which launch sites have the most successful launch outcomes and key locations.

GitHub URL

https://github.com/Ladyamethyst811/Data-Science-Capstone/blob/main/IBM-DS0321EN-SkillsNetwork_labs_module_3_lab_jupyter_launch_site_location.jupyterlite(1).ipynb

Build a Dashboard with Plotly Dash

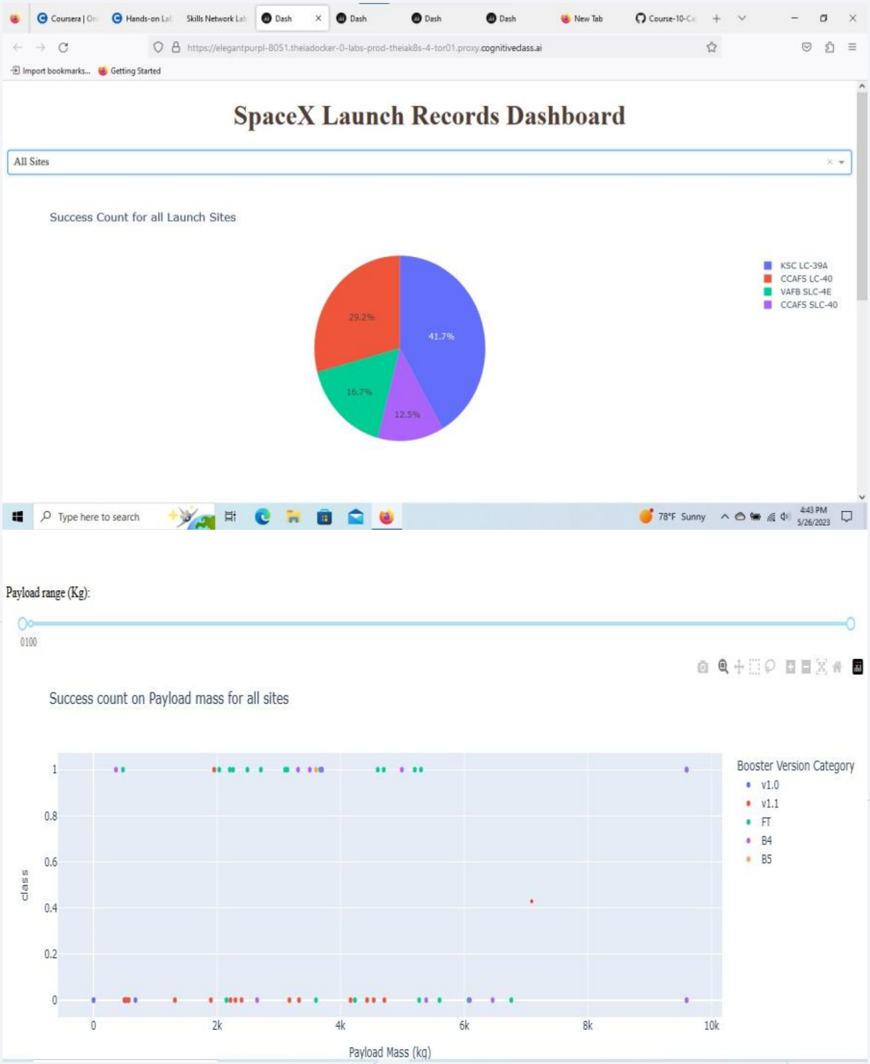
- Dashboard contains Drop-Down Input, Pie Chart, Range Slider, and Scatter Plot.

- Drop-Down input contains the option to select all launch sites or an individual site.

- Pie Chart is a representation of the success and failure of the launch sites.

- Range Slider lets you see the difference payload mass has on the success rate by moving it from 0 – 10,000 kg.

- Scatter Plot shows how the success rate depends on launch site, payload mass, and booster version.



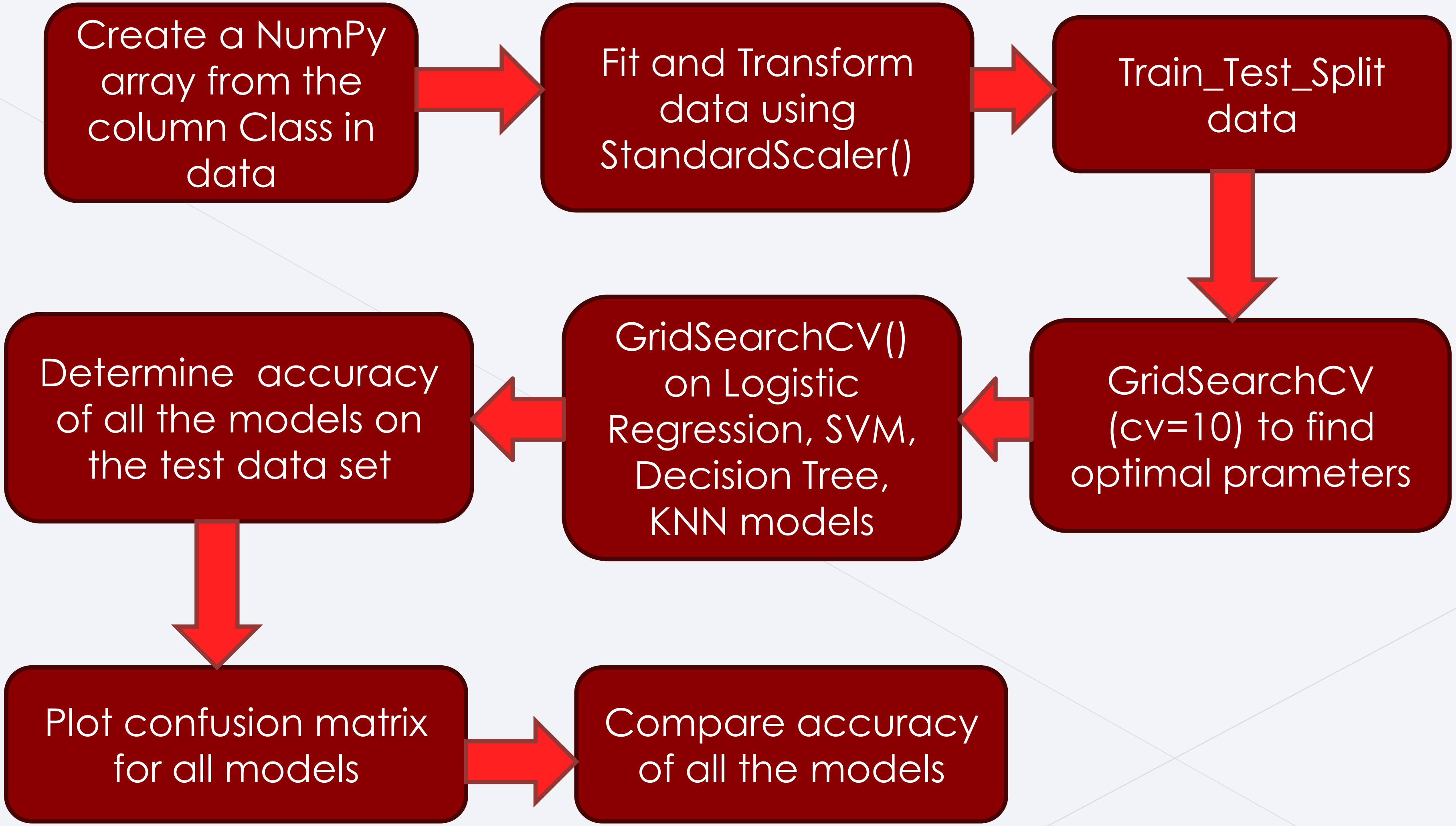
GitHub URL

https://github.com/Ladyamethyst811/Data-Science-Capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

GitHub URL

https://github.com/Ladyamethyst811/Data-Science-Capstone/blob/main/IBM-DS0321EN-SkillsNetwork_labs_module_4_SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite(2).ipynb



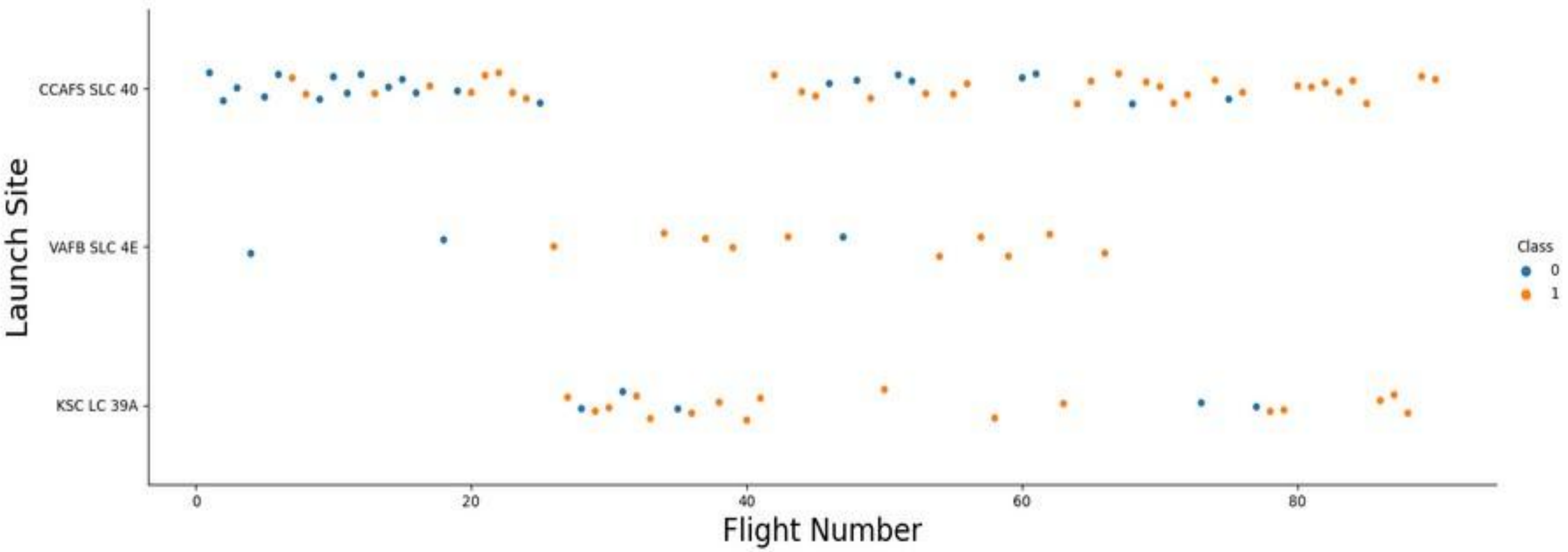
Results

- Exploratory data analysis results
- Orbits ES-L1, GEO, HEO, and SSO have the highest success rate.
- All of the landing outcomes on a drone ship with payload mass between 4,000 and 6,000 are successful.
- Interactive analytics demo in screenshots
- CCAFS SCL-40 has the highest launch success ratio 42.9%.
- KSC LC-39A has the highest success rate.
- Predictive analysis results
- All of the models(Logistic Regression, SVM, Decision Tree, and KNN) have a accuracy score of approximately 83.33%.

Section 2

Insights drawn from EDA

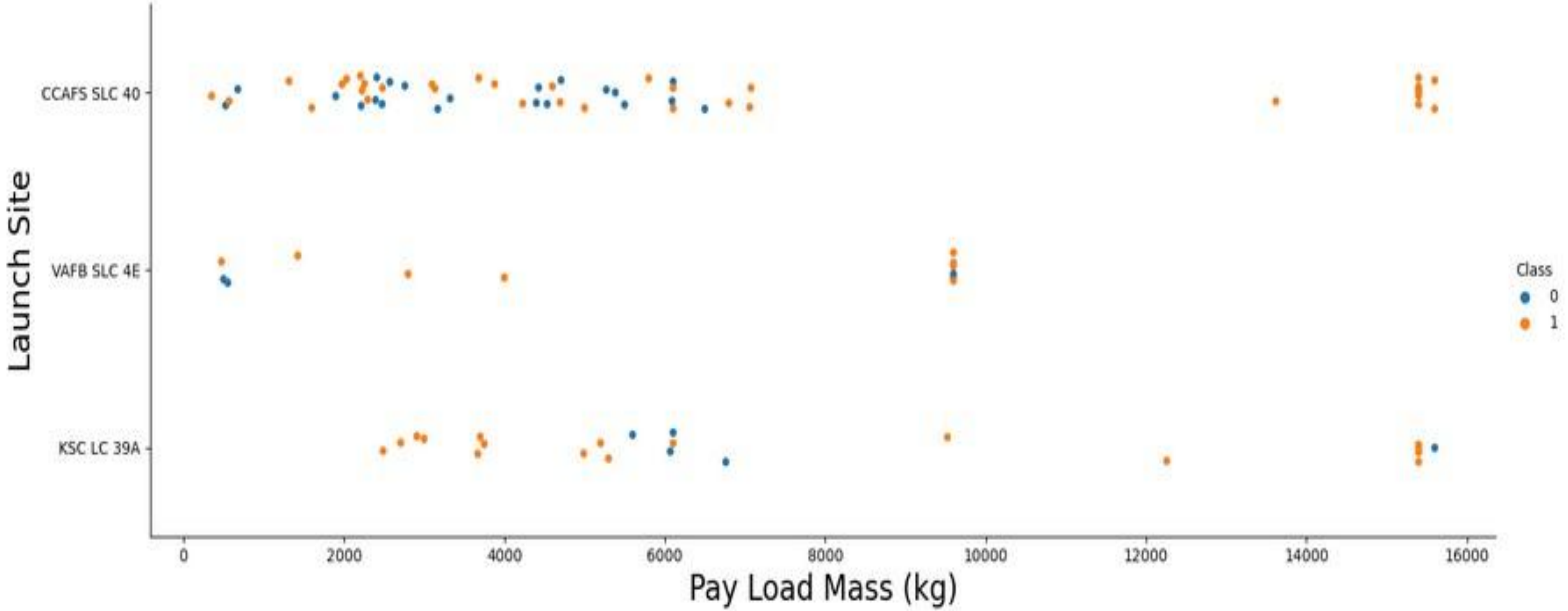
Flight Number vs. Launch Site



Now try to explain the patterns you found in the Flight Number vs. Launch Site scatter point plots.

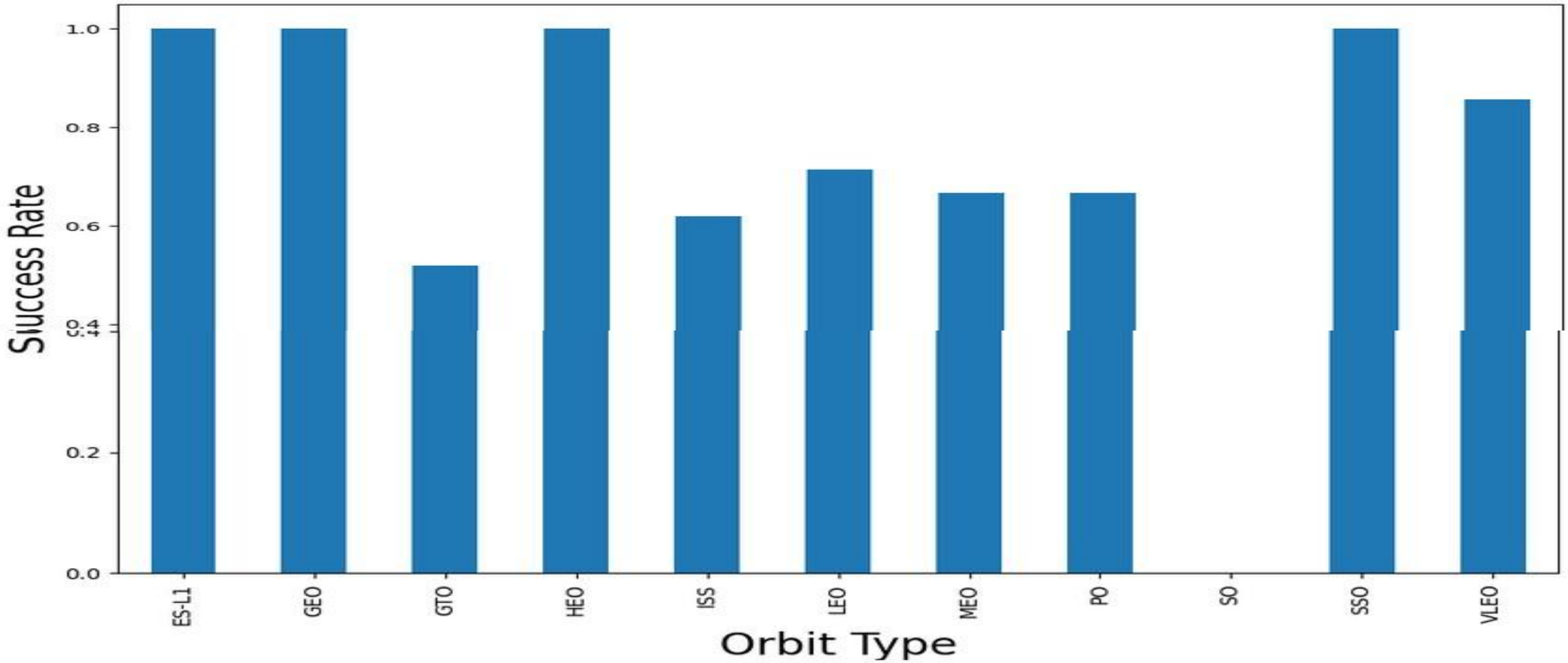
It appears CCAFS SLC 40 has the largest amount of launches. Also the higher the Fight Number the better the success rate.

Payload vs. Launch Site



Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

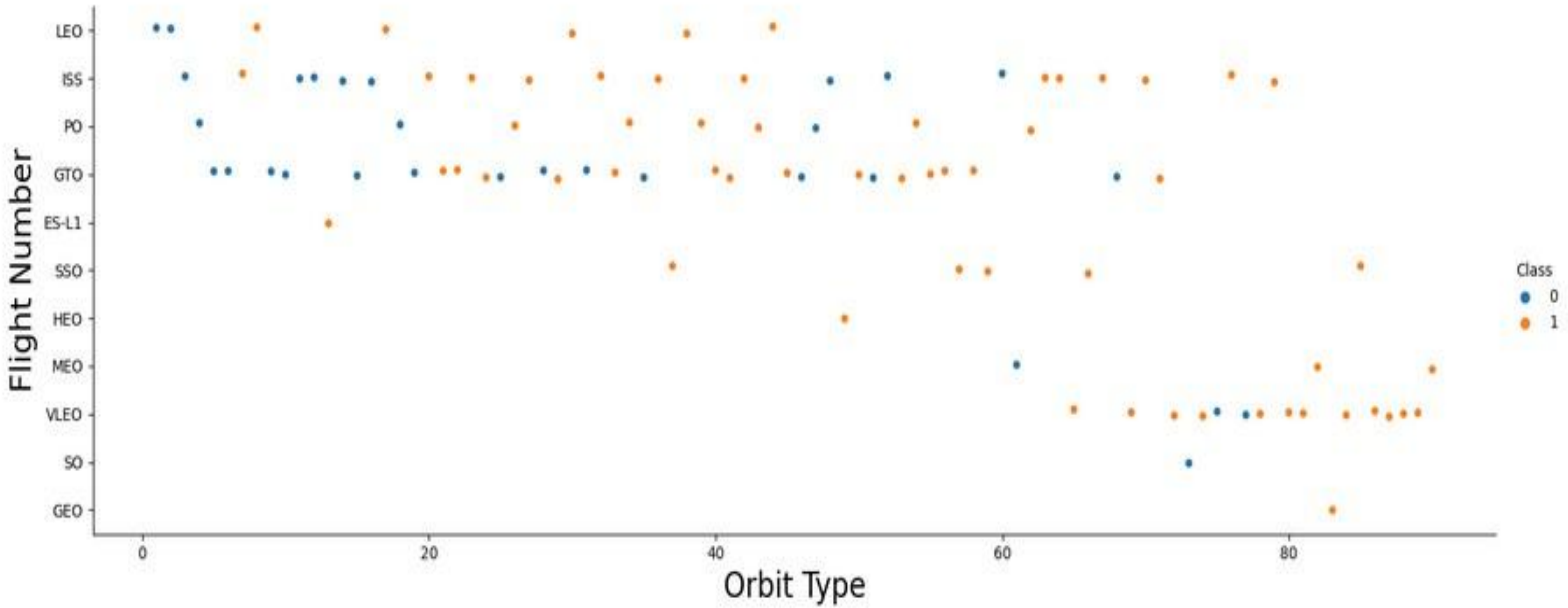
Success Rate vs. Orbit Type



Analyze the ploted bar chart try to find which orbits have high sucess rate.

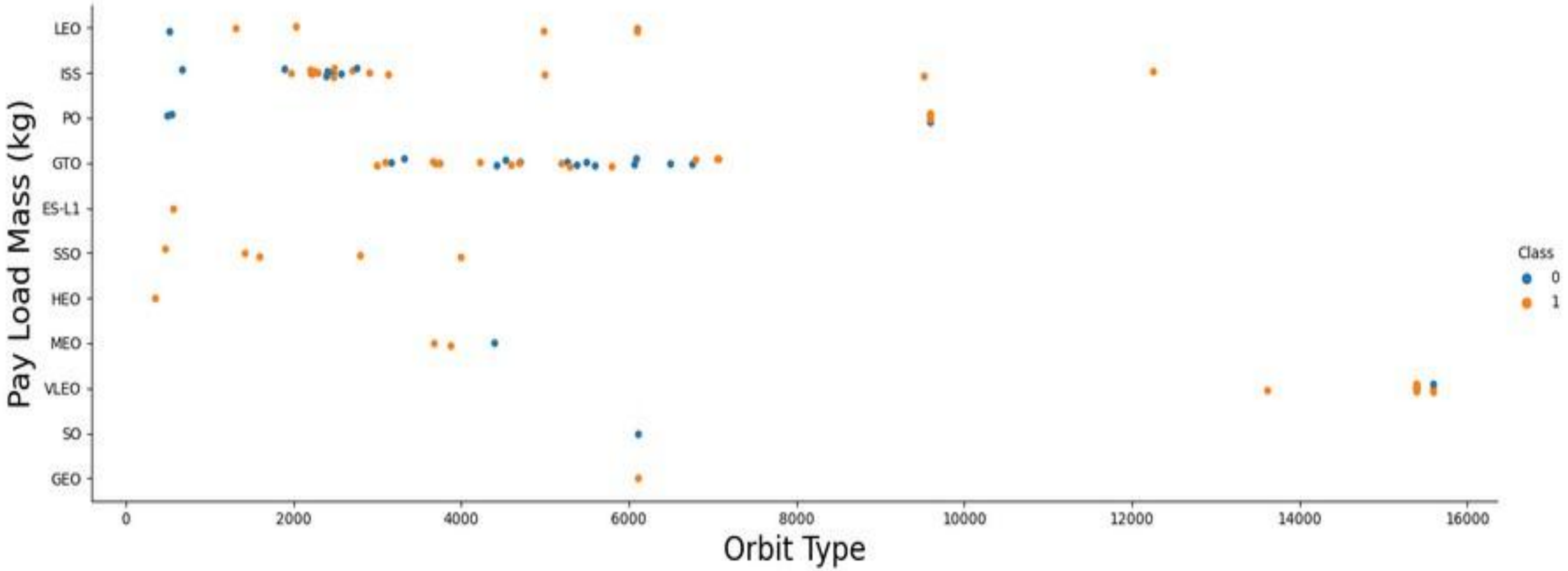
Orbit Types ES-L1, GEO, HEO, and SSO seem to have the highest success rate.

Flight Number vs. Orbit Type



You should see that 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.

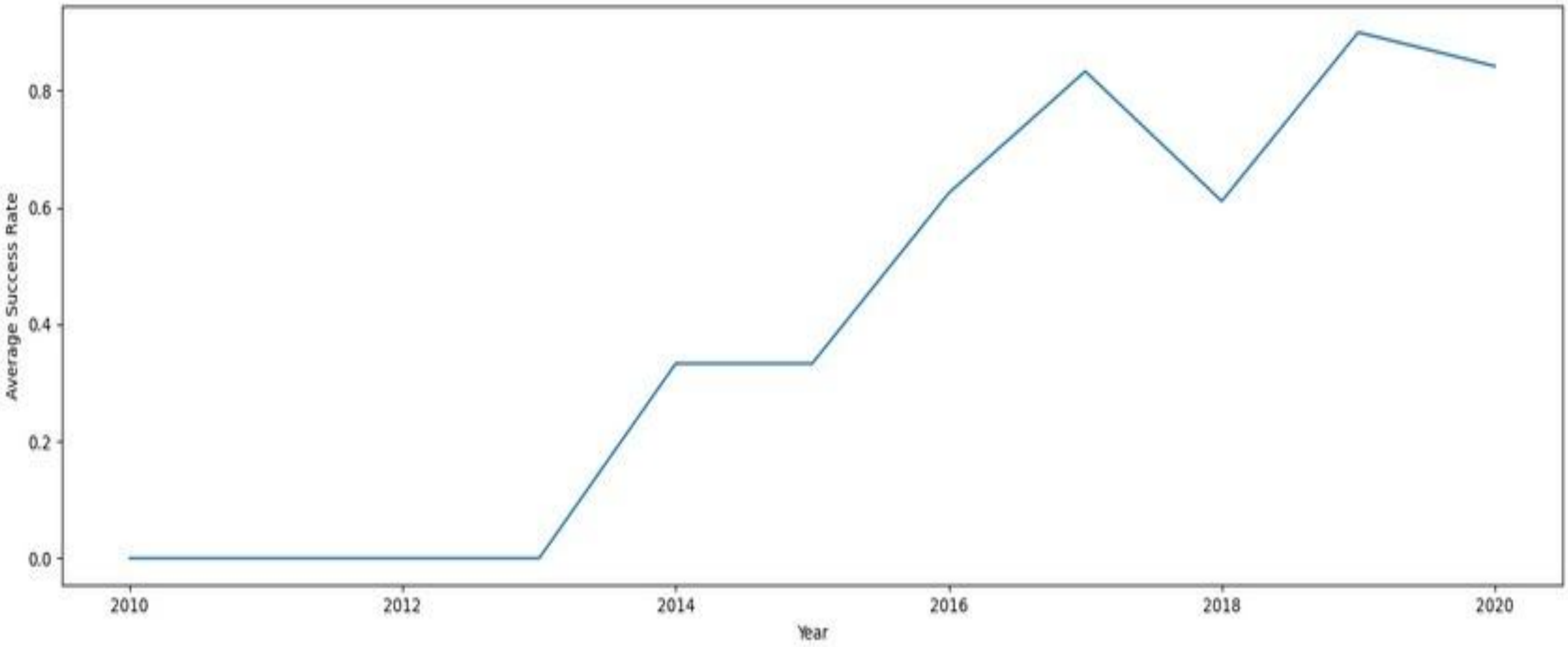
Payload vs. Orbit Type



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



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

All Launch Site Names

Display the names of the unique launch sites in the space mission

```
%sql SELECT DISTINCT "LAUNCH_SITE" FROM SPACEXTBL;
```

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

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40
None

As you can see there are four Launch Sites:
CCAFS LC-40,
VAFB SLC-4E, KSC LC-39A,
and CCAFS SLC-40.

Launch Site Names Begin with 'CCA'

Display 5 records where launch sites begin with the string 'CCA'

[29]:
```
%%sql
SELECT * FROM SPACEXTBL WHERE "LAUNCH_SITE" LIKE 'CCA%' LIMIT 5;
```

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

[29]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
06/04/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0.0	LEO	SpaceX	Success	Failure (parachute)
12/08/2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0.0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
22/05/2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525.0	LEO (ISS)	NASA (COTS)	Success	No attempt
10/08/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500.0	LEO (ISS)	NASA (CRS)	Success	No attempt
03/01/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677.0	LEO (ISS)	NASA (CRS)	Success	No attempt

Task 3

… 5 records where launch site name begins with “CCA”.

Total Payload Mass

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

```
%%sql
SELECT sum(PAYLOAD_MASS__KG_) as Payload_Mass, "CUSTOMER" FROM SPACEXTBL WHERE CUSTOMER = 'NASA (CRS)';
```

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

Payload_Mass	Customer
45596.0	NASA (CRS)

The payload mass for NASA(CRS) is 45596.0.

Average Payload Mass by F9 v1.1

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

```
%%sql
SELECT AVG(PAYLOAD_MASS__KG_) from SPACEXTBL where Booster_Version LIKE 'F9 v1.1%';
```

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

AVG(PAYLOAD_MASS__KG_)
2534.6666666666665

Average payload mass for booster F9 v1.1 is approximately 2534.65.

First Successful Ground Landing Date

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

Hint:Use min function

```
%sql select * from spacextbl limit 1 ;
```

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
06/04/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0.0	LEO	SpaceX	Success	Failure (parachute)

```
%%sql
select MIN(Date) from SPACEXTBL where "LANDING_OUTCOME" = 'Success (ground pad)';
```

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

MIN(Date)
01/08/2018

First successful ground landing date is January 8 2018.

Successful Drone Ship Landing with Payload between 4000 and 6000

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

```
%%sql

SELECT booster_version,PAYLOAD_MASS__KG_ ,"Landing_Outcome" from SPACEXTBL where  "Landing_Outcome"

='Success (drone ship)' and PAYLOAD_MASS__KG_ >4000 and PAYLOAD_MASS__KG_ < 6000
```

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

Booster_Version	PAYLOAD_MASS__KG_	Landing_Outcome
F9 FT B1022	4696.0	Success (drone ship)
F9 FT B1026	4600.0	Success (drone ship)
F9 FT B1021.2	5300.0	Success (drone ship)
F9 FT B1031.2	5200.0	Success (drone ship)

All of the landing outcomes on a drone ship with payload mass between 4,000 and 6,000 are successful.

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

```
%%sql
select "Mission_Outcome", count ("Mission_Outcome") from SPACEXTBL GROUP BY Mission_Outcome;
```

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

Mission_Outcome	count ("Mission_Outcome")
None	0
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

There are 100 successful and 1 failure landing outcome.

Boosters Carried Maximum Payload

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
%%sql
select "BOOSTER_VERSION", "Payload_Mass__kg_"
from SPACEXTBL where 'Payload_Mass__kg_' = (select max('Payload_Mass__kg_') from SPACEXTBL)
ORDER BY PAYLOAD_MASS__KG_ DESC LIMIT 5;
```

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

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600.0
F9 B5 B1049.4	15600.0
F9 B5 B1051.3	15600.0
F9 B5 B1056.4	15600.0
F9 B5 B1048.5	15600.0

Maximum payload mass is 15,600.

2015 Launch Records

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.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 4, 2) as month to get the months and substr(Date,7,4)='2015' for year.

```
%%sql
select substr(Date,4,2), "Booster_version", "Landing_Outcome", "Launch_Site" from SPACEXTBL
where substr(Date,7,4) = '2015' and "Landing_Outcome" = "Failure (drone ship)";
```

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

substr(Date,4,2)	Booster_Version	Landing_Outcome	Launch_Site
10	F9 v1.1 B1012	Failure (drone ship)	CCAFS LC-40
04	F9 v1.1 B1015	Failure (drone ship)	CCAFS LC-40

CCAFS LC-40 appears to be the only launch site to suffer failure in 2015.

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

Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

```
%sql SELECT "DATE","Landing_Outcome",count("Landing_Outcome") as LANDING_OUTCOME_COUNT,DATE from SPACEXTBL where substr(Date,7,4)
```

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

Date	Landing_Outcome	LANDING_OUTCOME_COUNT	Date_1
22/05/2012	No attempt	10	22/05/2012
22/12/2015	Success (ground pad)	5	22/12/2015
04/08/2016	Success (drone ship)	5	04/08/2016
01/10/2015	Failure (drone ship)	5	01/10/2015
18/04/2014	Controlled (ocean)	3	18/04/2014
29/09/2013	Uncontrolled (ocean)	2	29/09/2013
28/06/2015	Precluded (drone ship)	1	28/06/2015
12/08/2010	Failure (parachute)	1	12/08/2010

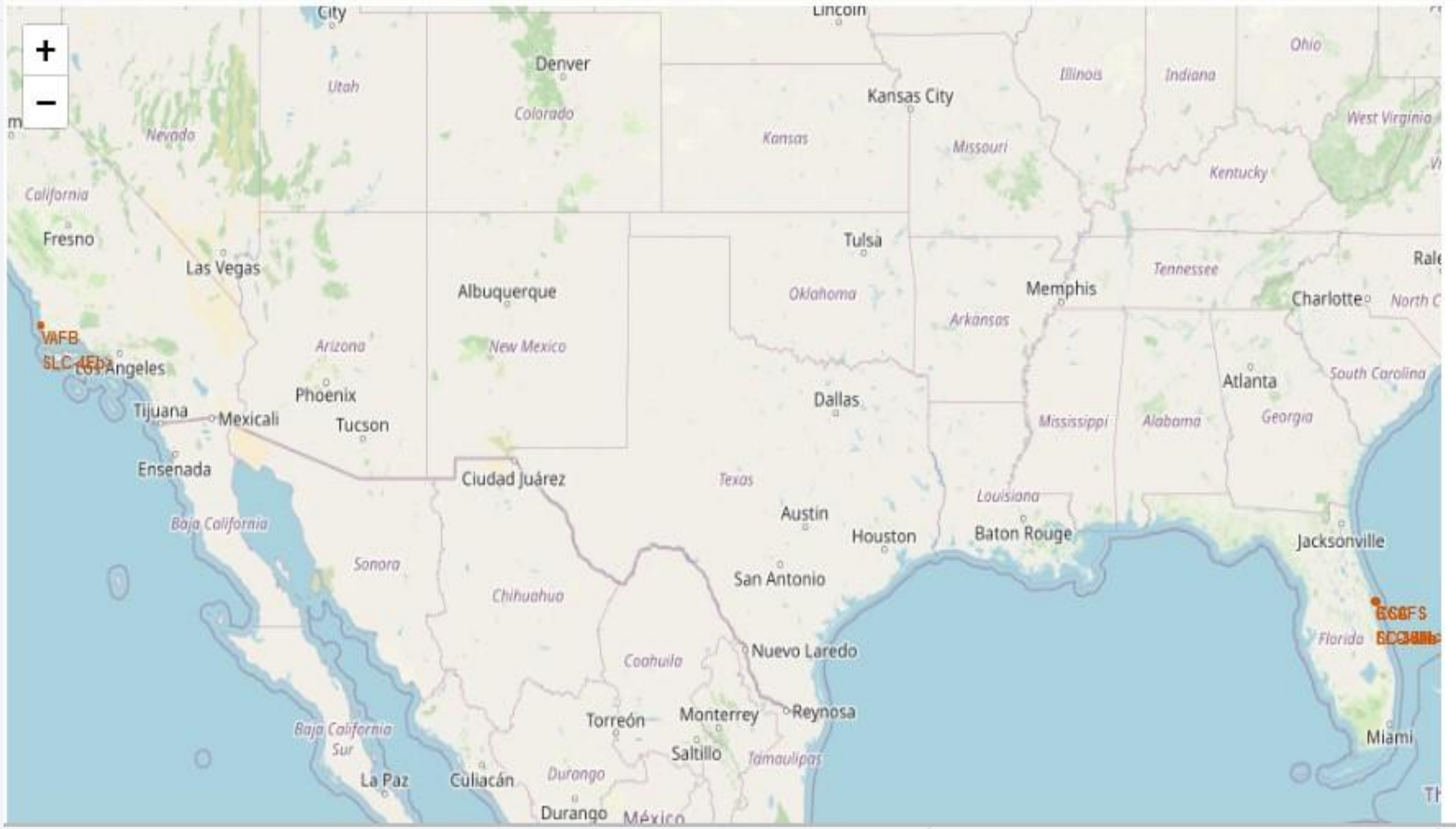
2015 and 2016 had the best success rates

Section 3

Launch Sites Proximities Analysis

Launch Site Locations

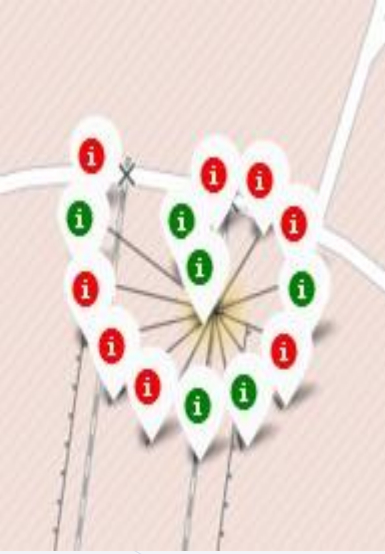
All launch sites are in close proximity to the Equator and Coastline.



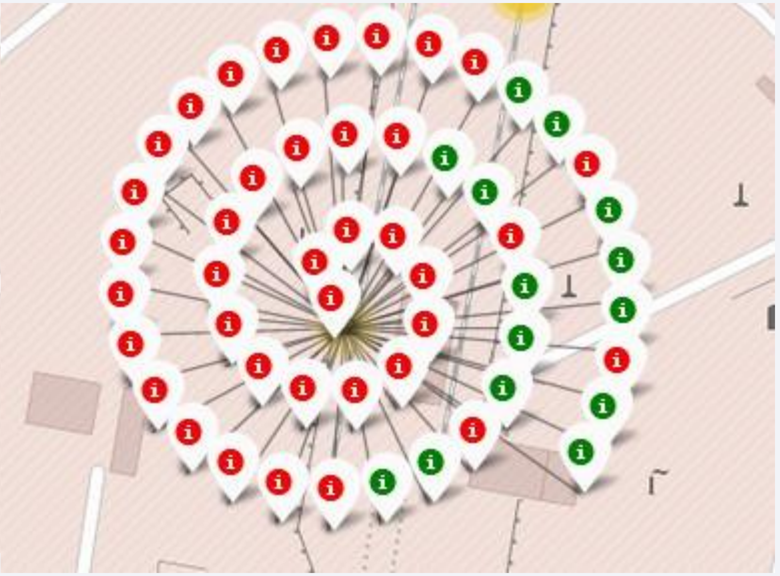
Launch Site Success Rate

KSC LC-39A

VAFB SLC-4E



CCAFS SLC-40



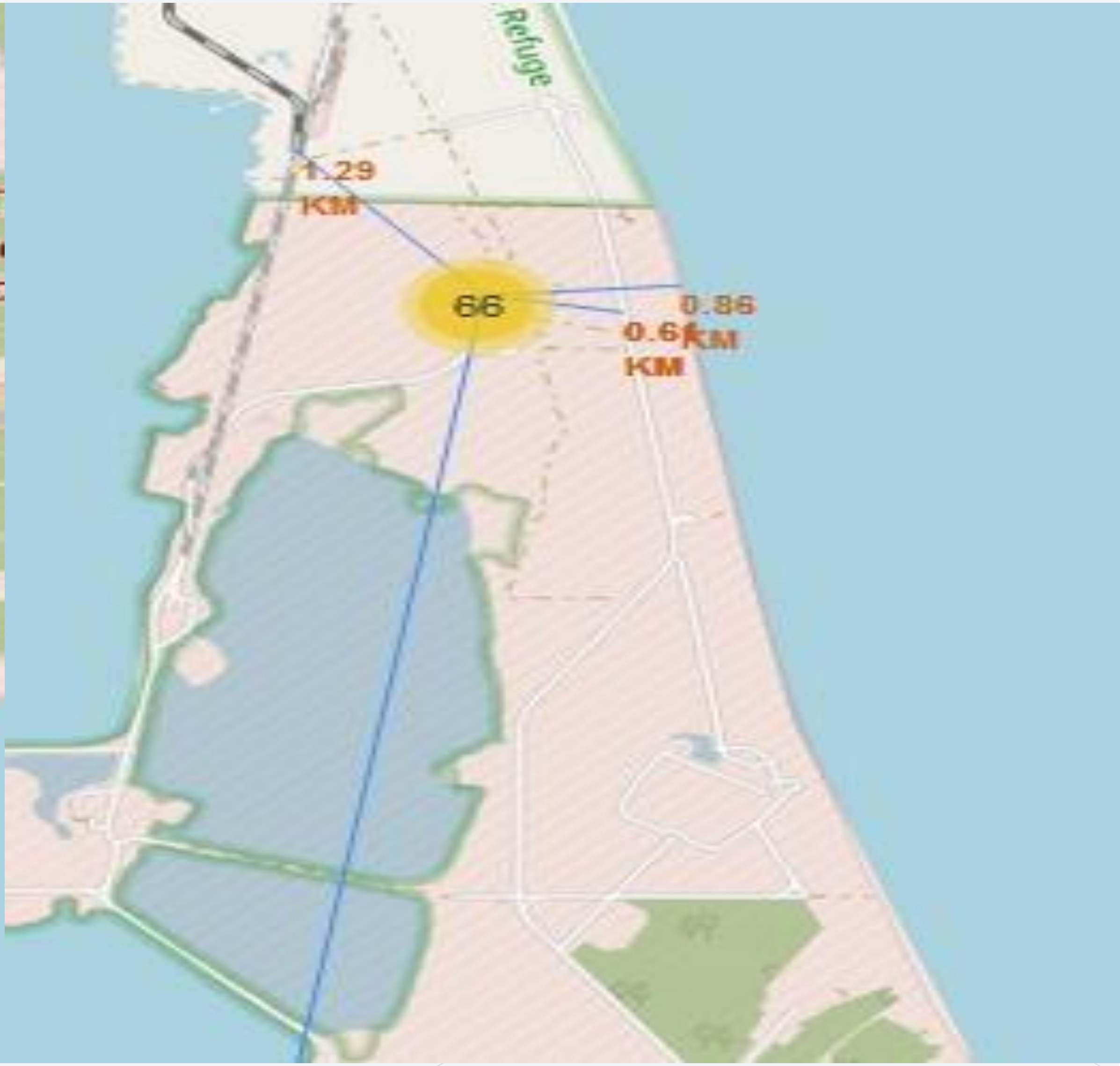
CCAFS LC-40

KSC LC-39A has the highest success rate and CCAFS LC-40 has the lowest.

Launch Site Markers

Launch sites are:

- close to railways to transport heavy cargo with ease.
- close proximity to highways to make transportation of people and supplies easier.
- close to coastline for safety reasons
- are not close to cities to minimize injuries.

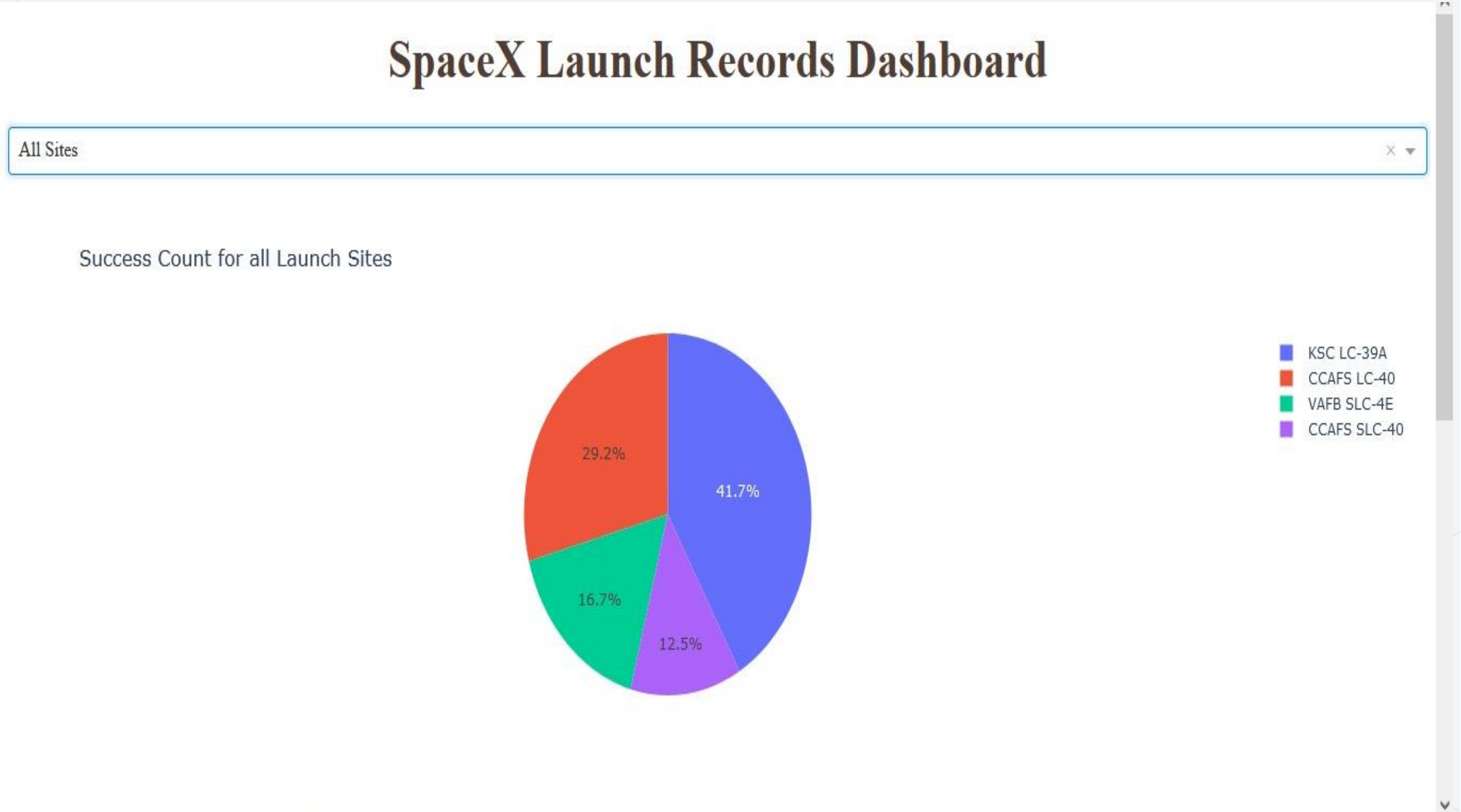


Section 4

Build a Dashboard with Plotly Dash

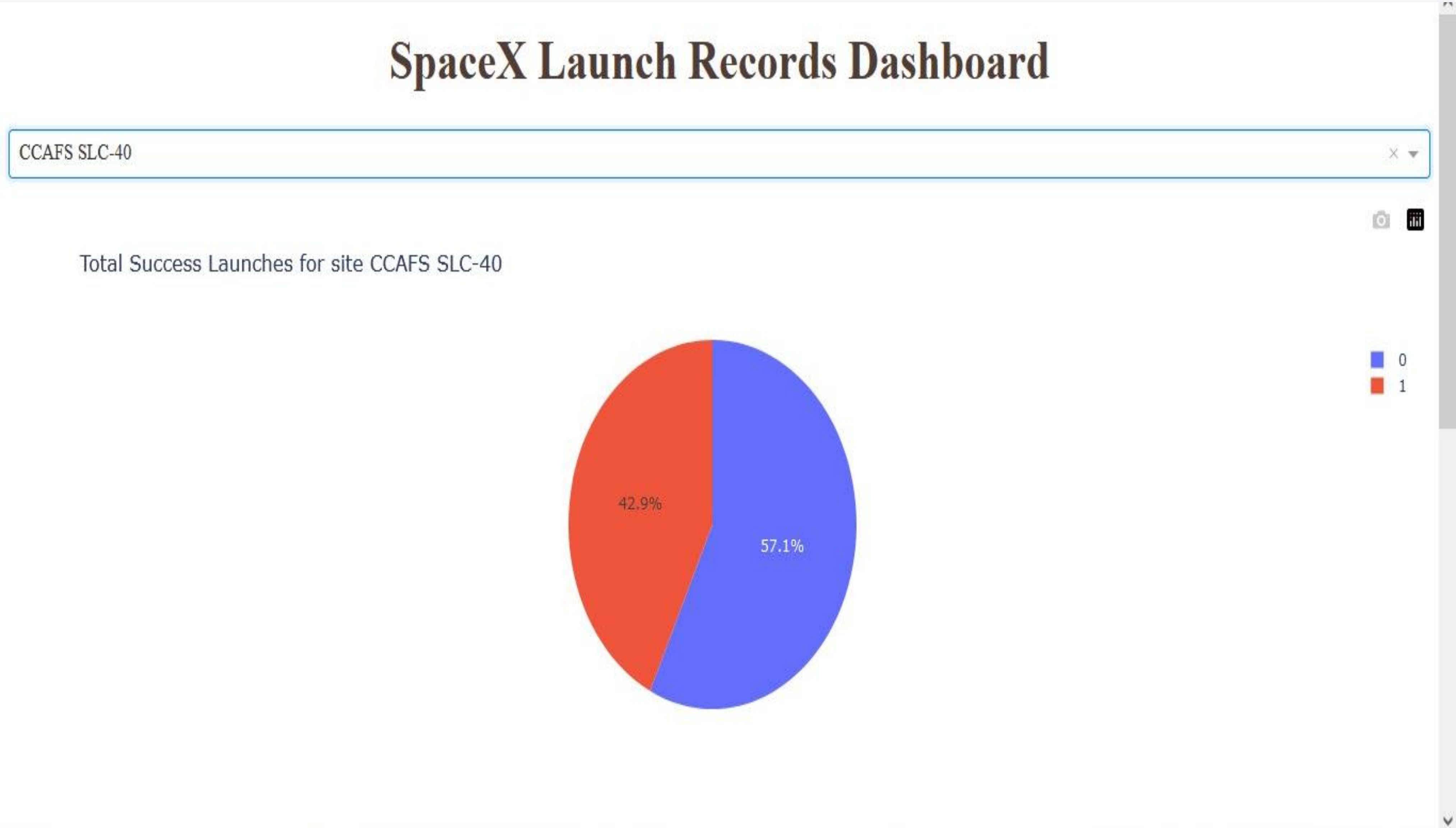
Success Count for all Launch Sites

- KSC LC-39A has the highest success rate.
- CCAFS SLC-40 has the lowest success rate.

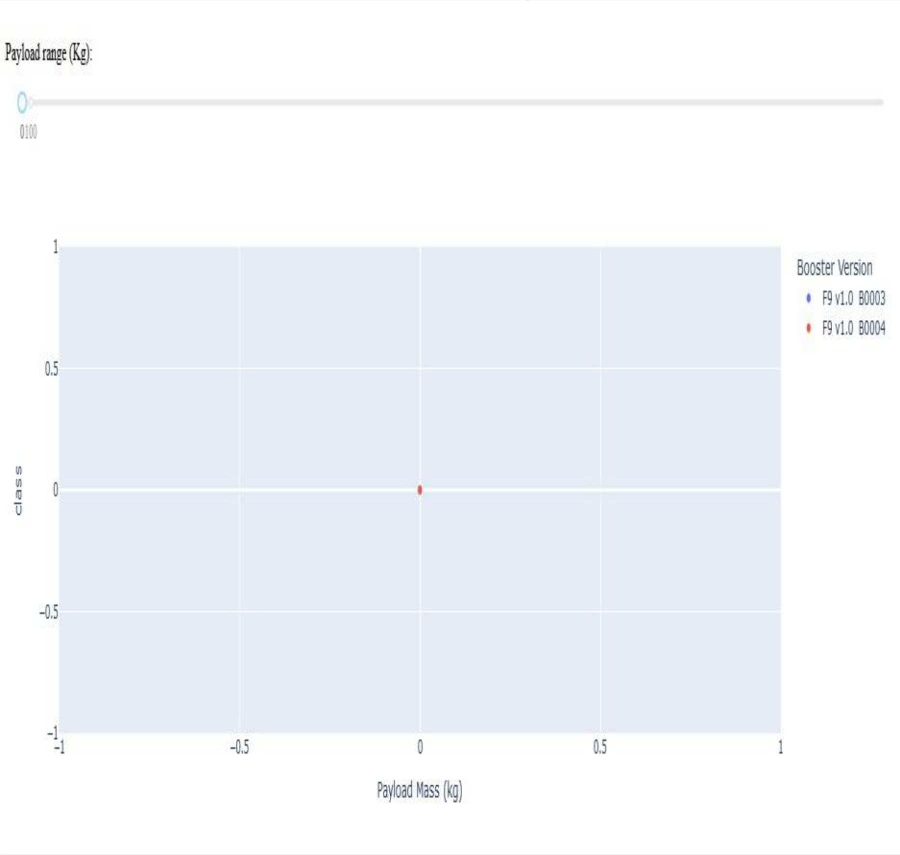


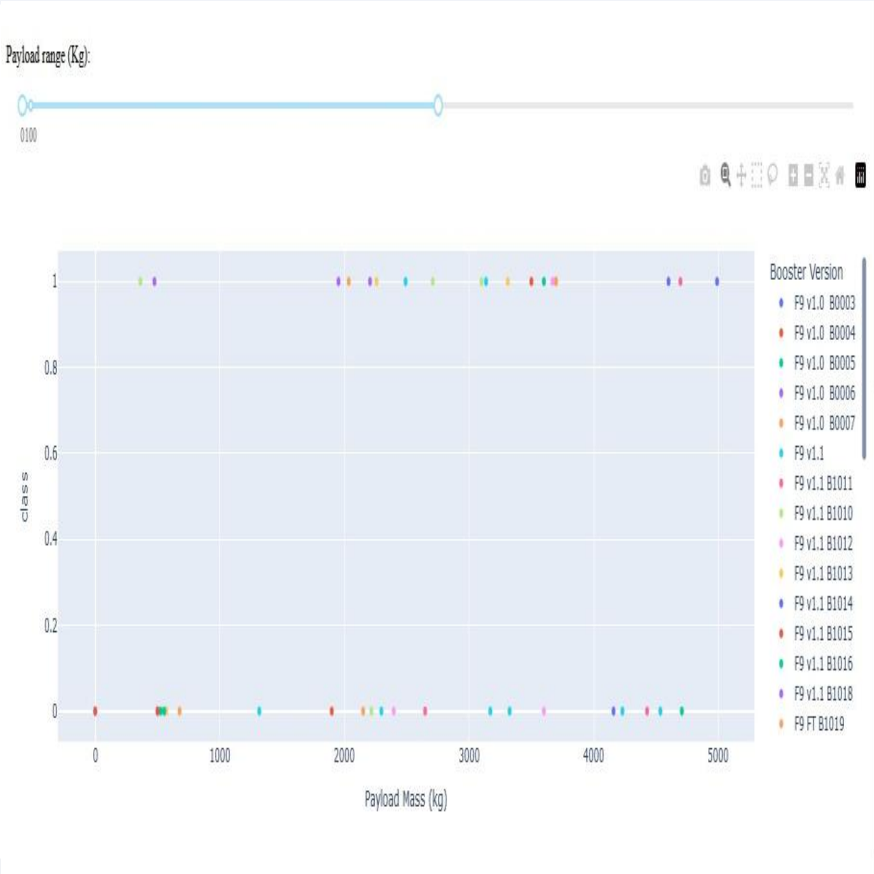
Highest Launch Success Ratio

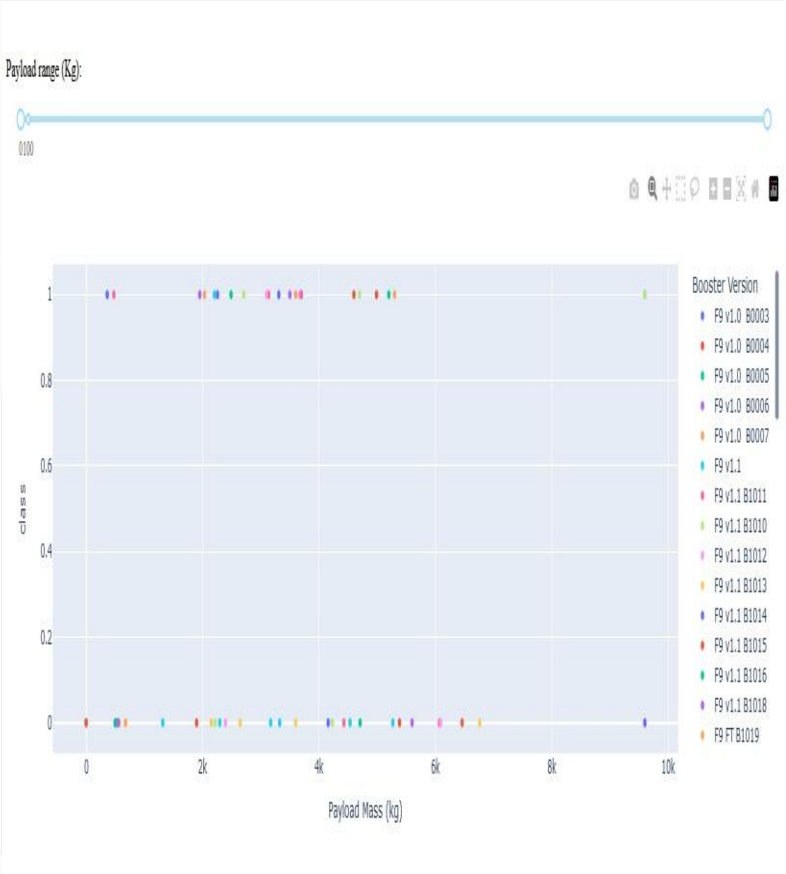
CCAFS SLC-40 has the highest launch success ratio. With success=42.9% ,and failure=57.1% .



Payload vs Launch Outcome







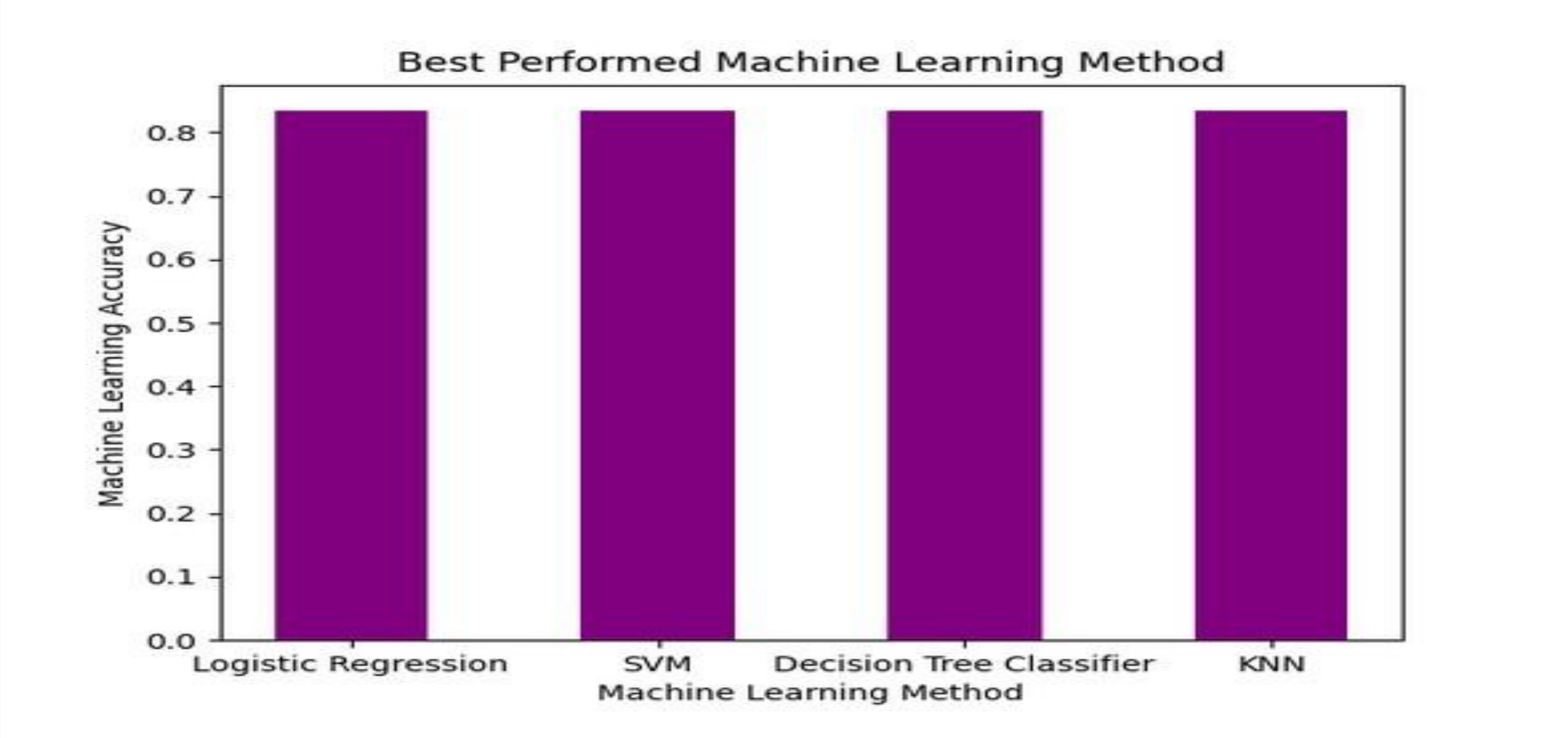
2,000kg-4,000kg has the highest launch success rate and 500-677, 2216-2315, and 4156-4707 have the lowest success rates.

F9 FT Booster has the highest success rate and F9 v1.0 Booster has the lowest success rate.

Section 5

Predictive Analysis (Classification)

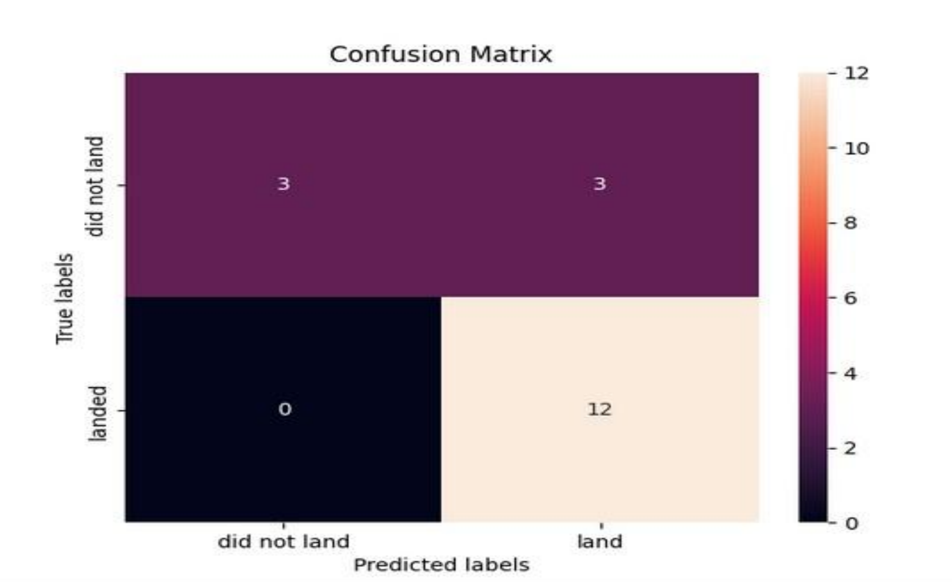
Classification Accuracy

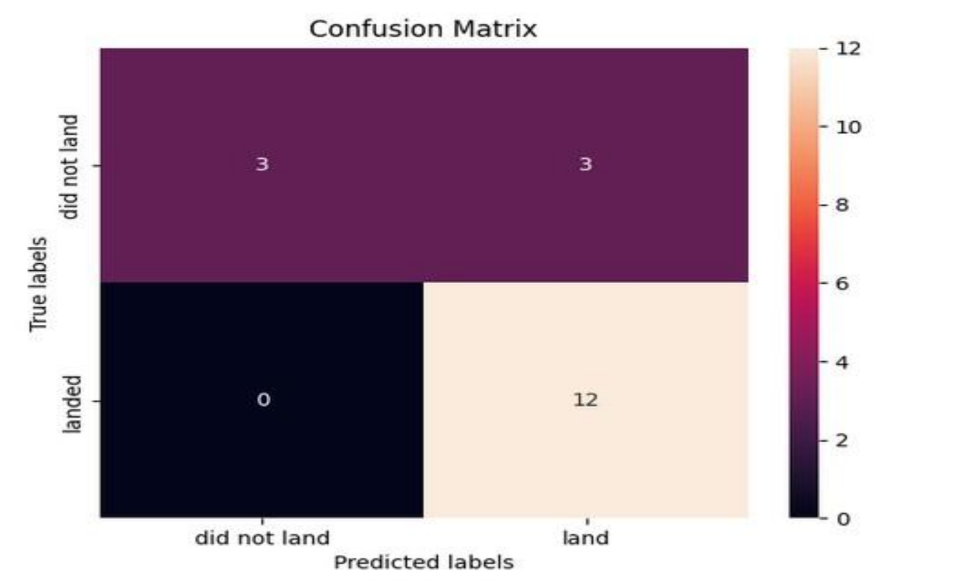


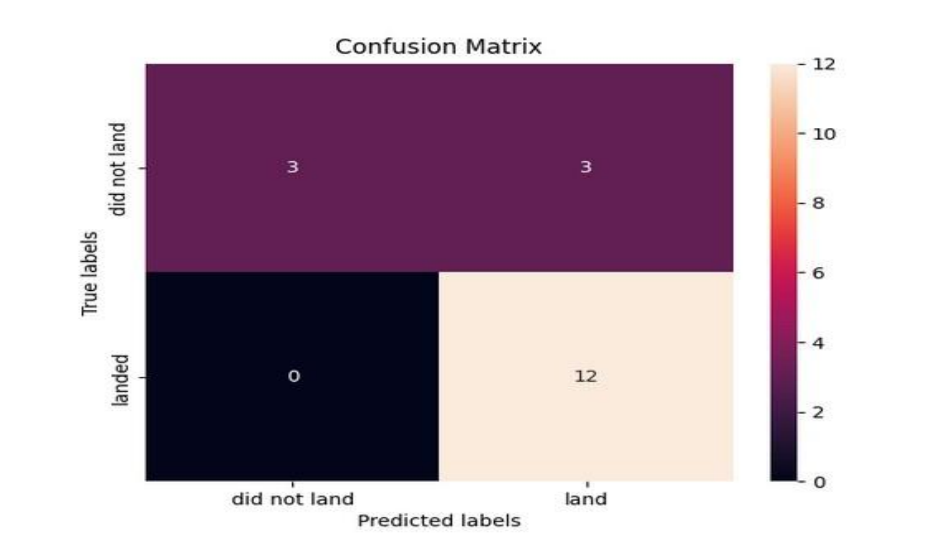
The results of all the Models are practicly the same.

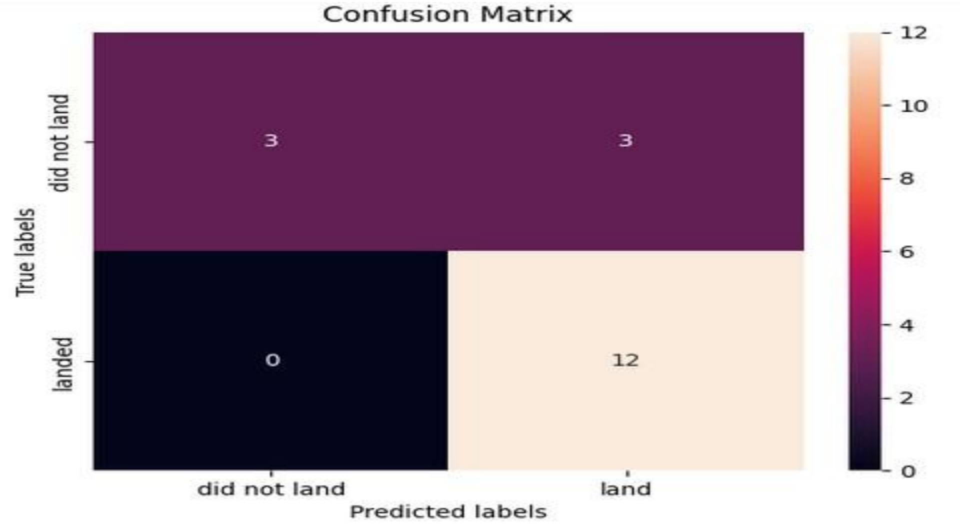
Confusion Matrix

As you can see the results of the confusion matrixes are practically the same.









Conclusions

- Task: Create a machine learning model for Space Y, which wants to compete with Space X.
- A machine learning model with an accuracy of 83.33% was produced.
- The model's purpose is to anticipate when Stage One would successfully arrive in order to save $100 million USD.
- Space Y can use this model to predict if a launch will have a successful Stage One landing before deciding whether or not to launch.
- Suggestion: If possible, collect more data to better establish the appropriate machine learning model and increase model accuracy.

Appendix

API's used:

Space X API

- https://api.spacexdata.com/v4/rockets/
- https://api.spacexdata.com/v4/launchpads/
- https://api.spacexdata.com/v4/payloads/
- https://api.spacexdata.com/v4/cores/
- https://api.spacexdata.com/v4/launches/past
- Space X Wikipedia
- https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922

GitHub URL

https://github.com/Ladyamethyst811/Data-Science-Capstone/tree/main

Thank you!