IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Isaac Ramirez

31/03/2024

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection through API
  - Data Wrangling
  - Exploratory Data Analysis with SQL
  - Exploratory Data Analysis with Data Visualization
  - Interactive Visual Analytics with Folium
  - Machine Learning Prediction
- Summary of all results
  - Exploratory Data Analysis result
  - Interactive analytics in screenshots
  - Predictive Analytics result

Introduction

- **Project background and context:** In this project, we will predict if the Falcon 9 first stage will land successfully. 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.

- **Problems you want to find answers.**
  - What factors determine if the rocket will land successfully?
  - What operating conditions needs to be in place to ensure a successful landing

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - https://api.spacexdata.com/v4/
- Perform data wrangling
  - Data from 3 endpoints was grouped, null data was eliminated, and a new attribute was created: 'class'
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - The data was normalized and then several learning models with several parameters each were used to find the best one.

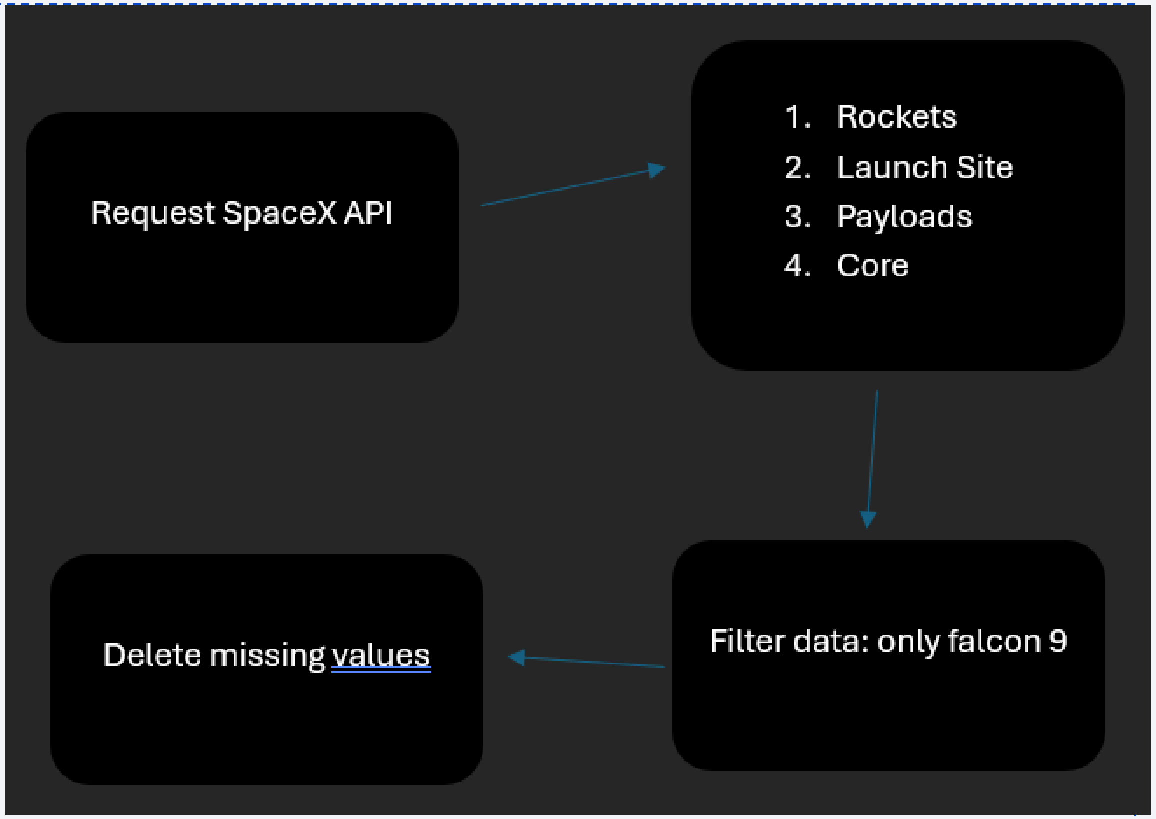
Data Collection

- Describe how data sets were collected.
- Data were collected from SpaceX API https://api.spacexdata.com/v4/

https://api.spacexdata.com/v4/payloads/
https://api.spacexdata.com/v4/payloads/

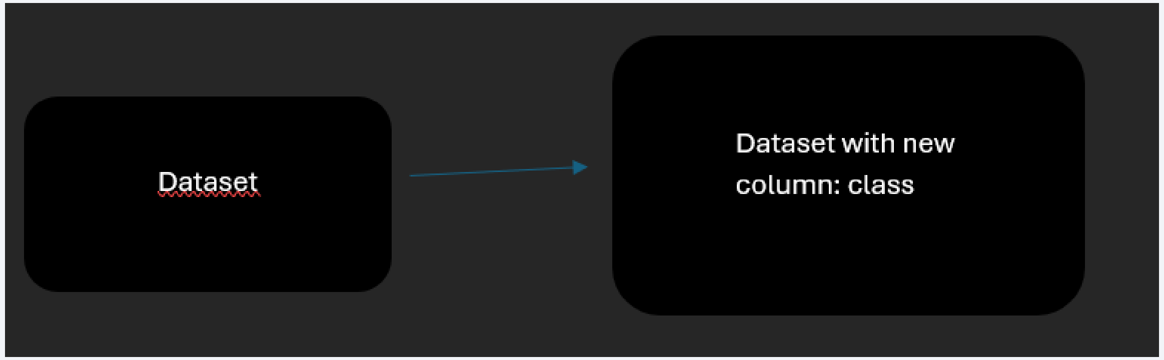
Data Collection – SpaceX API

- https://github.com/ulisac04/final-project/blob/main/jupyter-labs-spacex-data-collection-api.ipynb
    - Rockets
    - Launch Site
    - Payload
    - Core



Data Wrangling

- We were discussing what factors we could use to determine whether a launch was successful or not.
- A new attribute was created: class; 1=successful launch, 0=failed launch
- **GitHub URL:** https://github.com/ulisac04/final-project/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb



EDA with Data Visualization

- Used graphics
  - **Catplot** to understand how the payload mass and the number of launches are distributed, with respect to the place of launch
  - **Countplot** to understand the ratio of successful launches PER rocket orbit.
  - **Scatterplot** to understand how the payload mass and the number of launches are distributed, with respect to the orbit
  - **Lineplot** to observe how the rate of successful launches evolved over time

- GitHub URL https://github.com/ulisac04/final-project/blob/main/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- Names of the launch sites
- All launches that were made from sites starting with the string 'CCA'
- The total payload mass sent from the "NASA (CRS)" station.
- the average payload mass for Booster_Version = 'F9 v1.1'
- date when the first succesful landing outcome in ground pad was acheived.
- Names of the boosters which have succeeded in landing on a drone ship and have a payload mass between 4000 and 6000
- Total count of successful and failure mission outcomes
- Which booster_versions have transported the maximum recorded load?

- GitHub URL https://github.com/ulisac04/final-project/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Map objects used: Markers, circles, lines
- Markers: to indicate points such as launch sites
- Circles: to indicate areas around specific coordinates.
- Lines are used to indicate distances between two points

- GitHub URL https://github.com/ulisac04/final-project/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Use a standing chart to see the ratios of successful launch vs. failed launch. Explain why you added those plots and interactions
- Use a plotted scatter graph to explain the relationship between uutcome and ayload mass peer booster version.

- GitHub URL https://github.com/ulisac04/final-project/blob/main/dash_interactivity.py

Predictive Analysis (Classification)

- Load the data and create a dataset.
- Divided the training and test data
- Tested the data with several classification methods and several hyperparameters.
- GitHub URL https://github.com/ulisac04/final-project/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

- Exploratory data analysis results
  - The average payload of F9 v1.1 booster almost 2.900kg.
  - Space X has 4 launch sites
  - Almost 100% of the missions were a success.

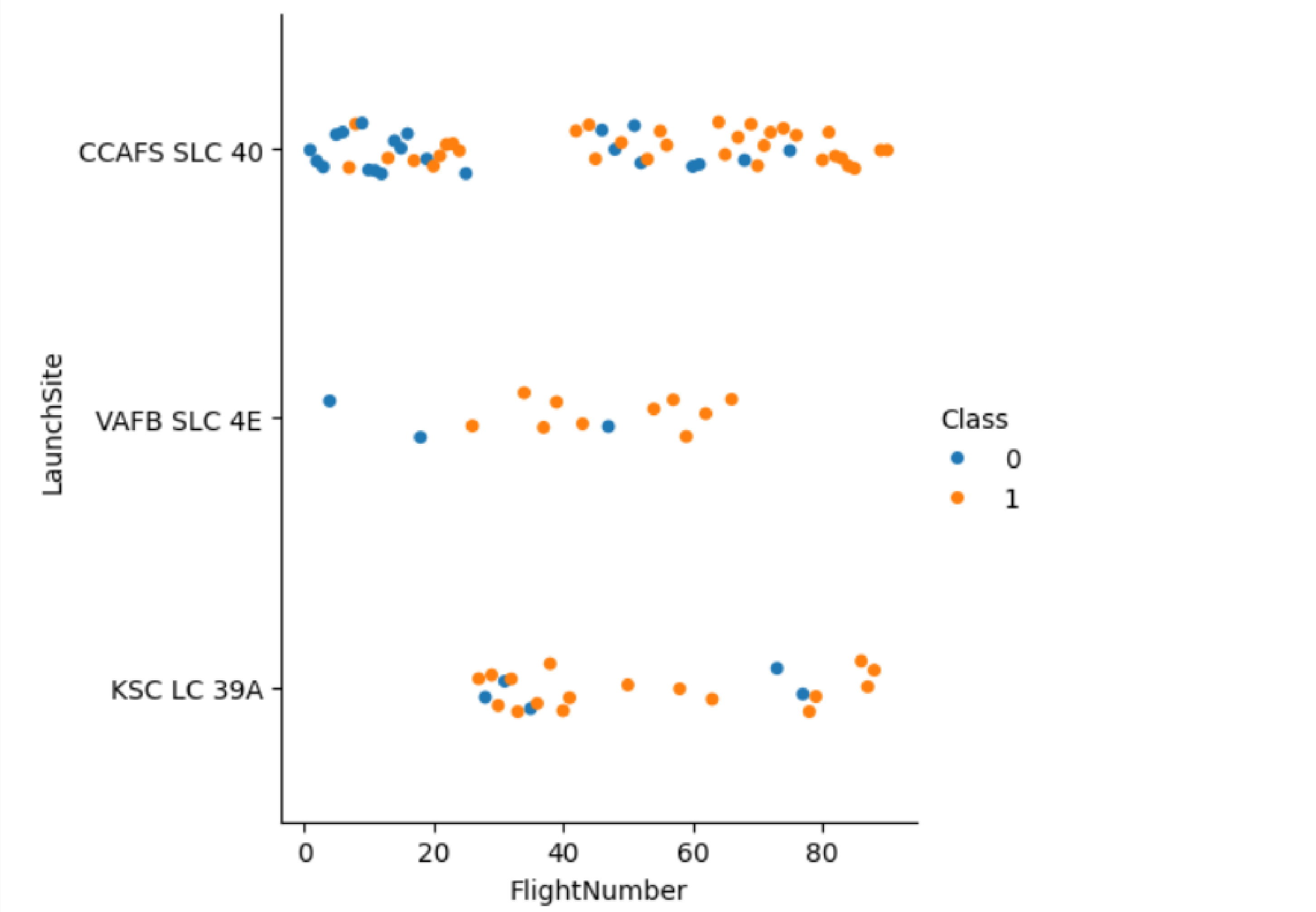
- Interactive analytics demo in screenshots
  - launches are always carried out near a coast

Section 2

Insights drawn from EDA

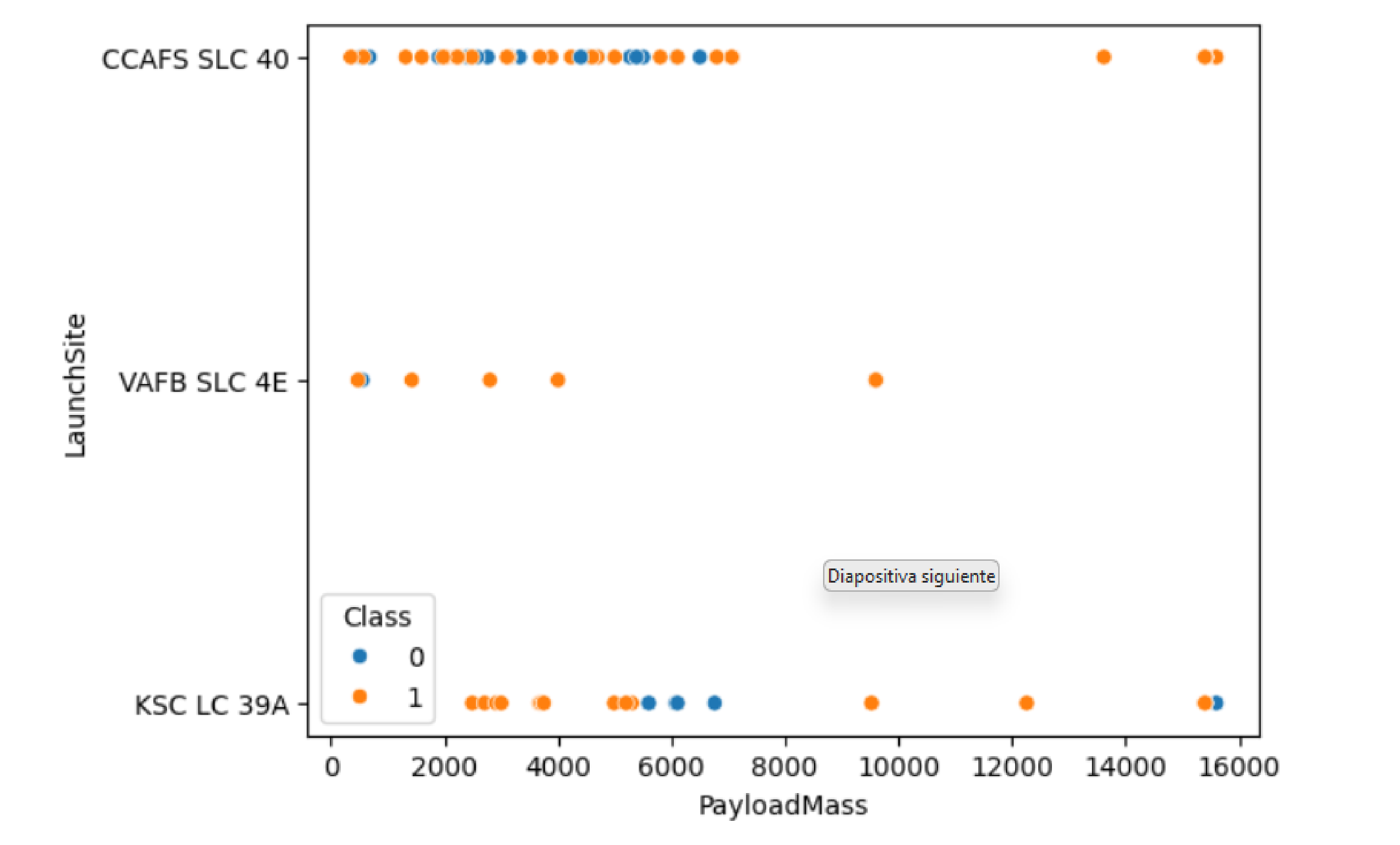
Flight Number vs. Launch Site

- The more launches, the higher the success rate



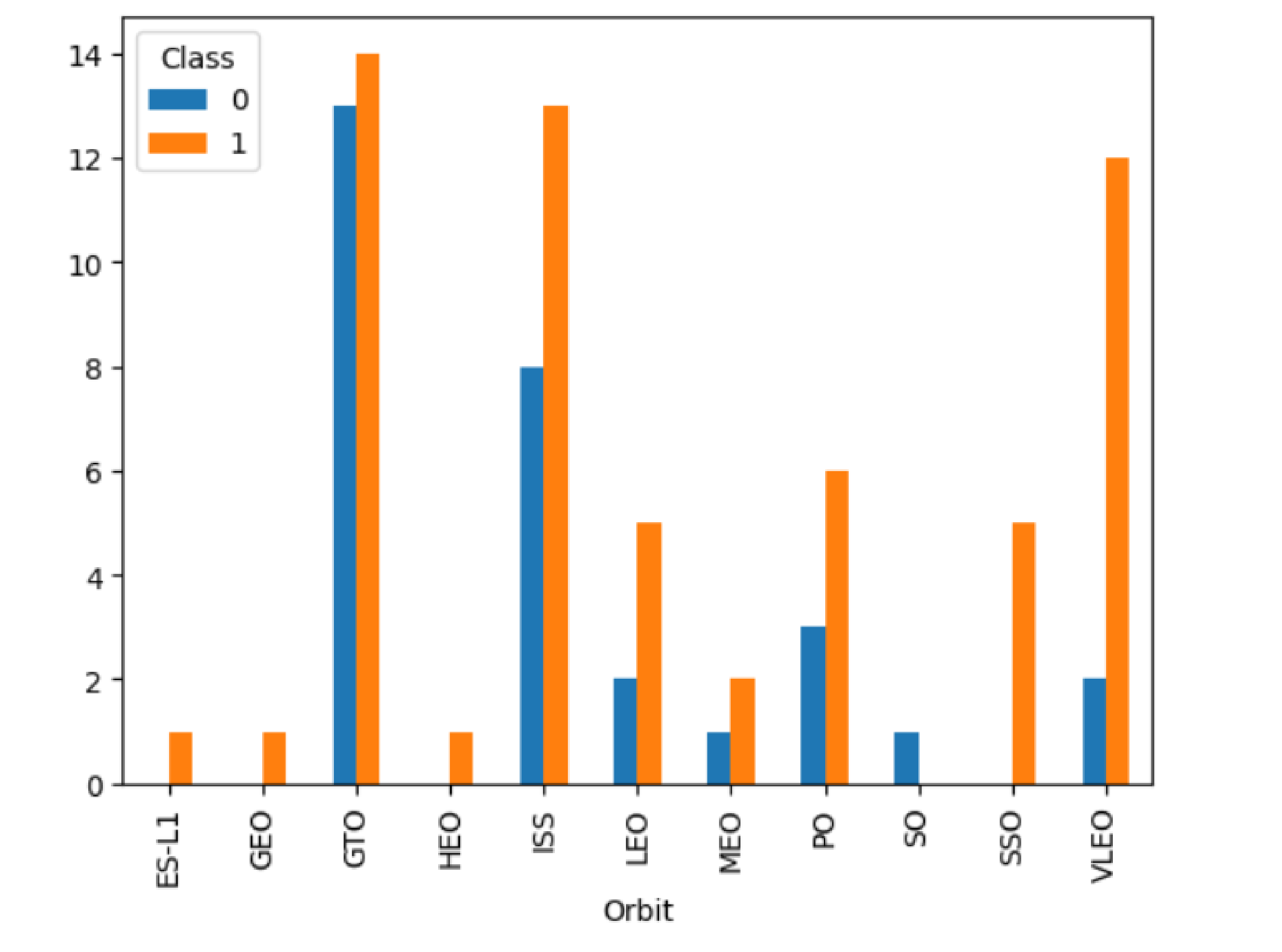
Payload vs. Launch Site

- VAFB SLC 4E: has a 100% success rate



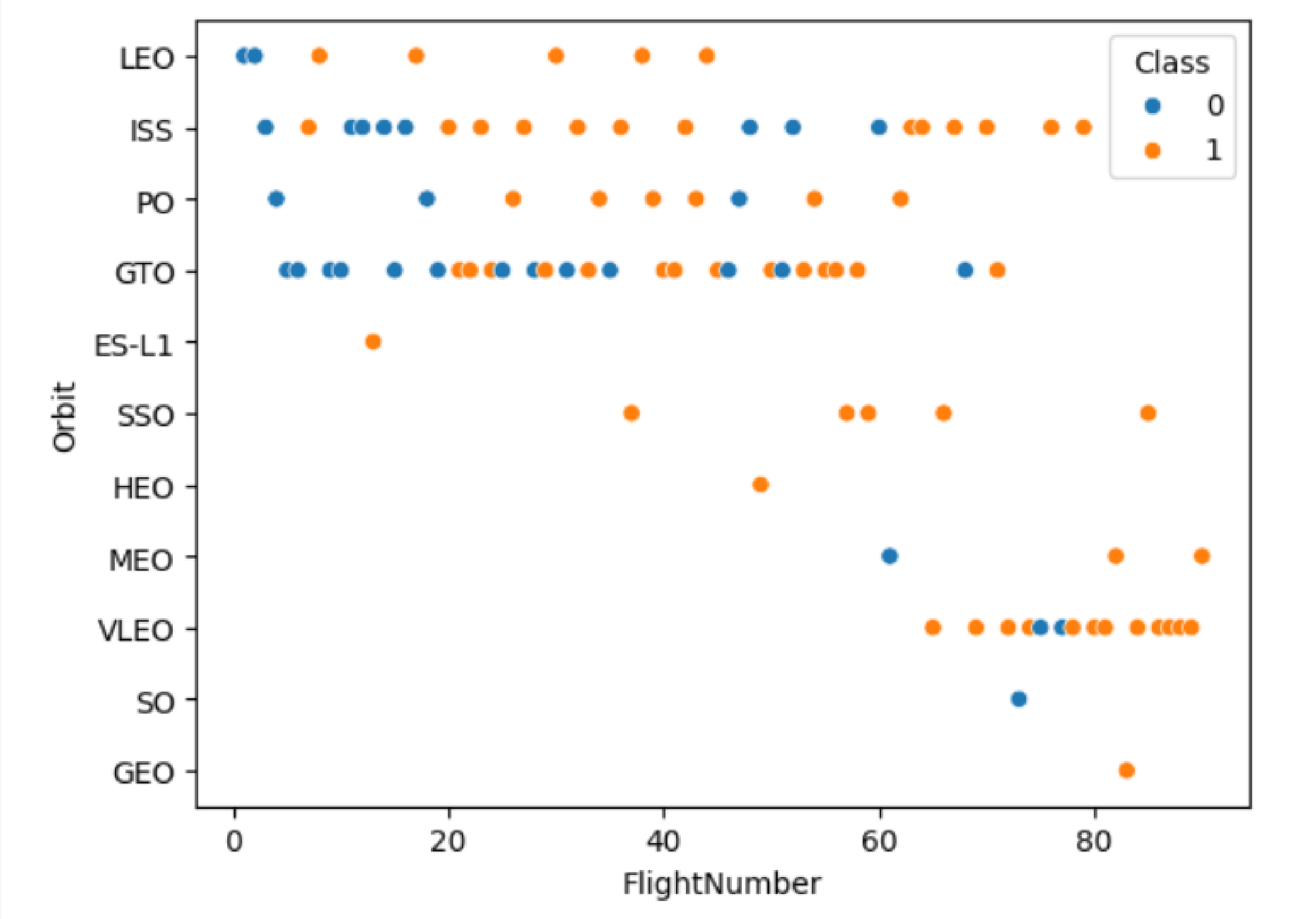
Success Rate vs. Orbit Type

- SSO is 100% successful
- GTO is the most deficient

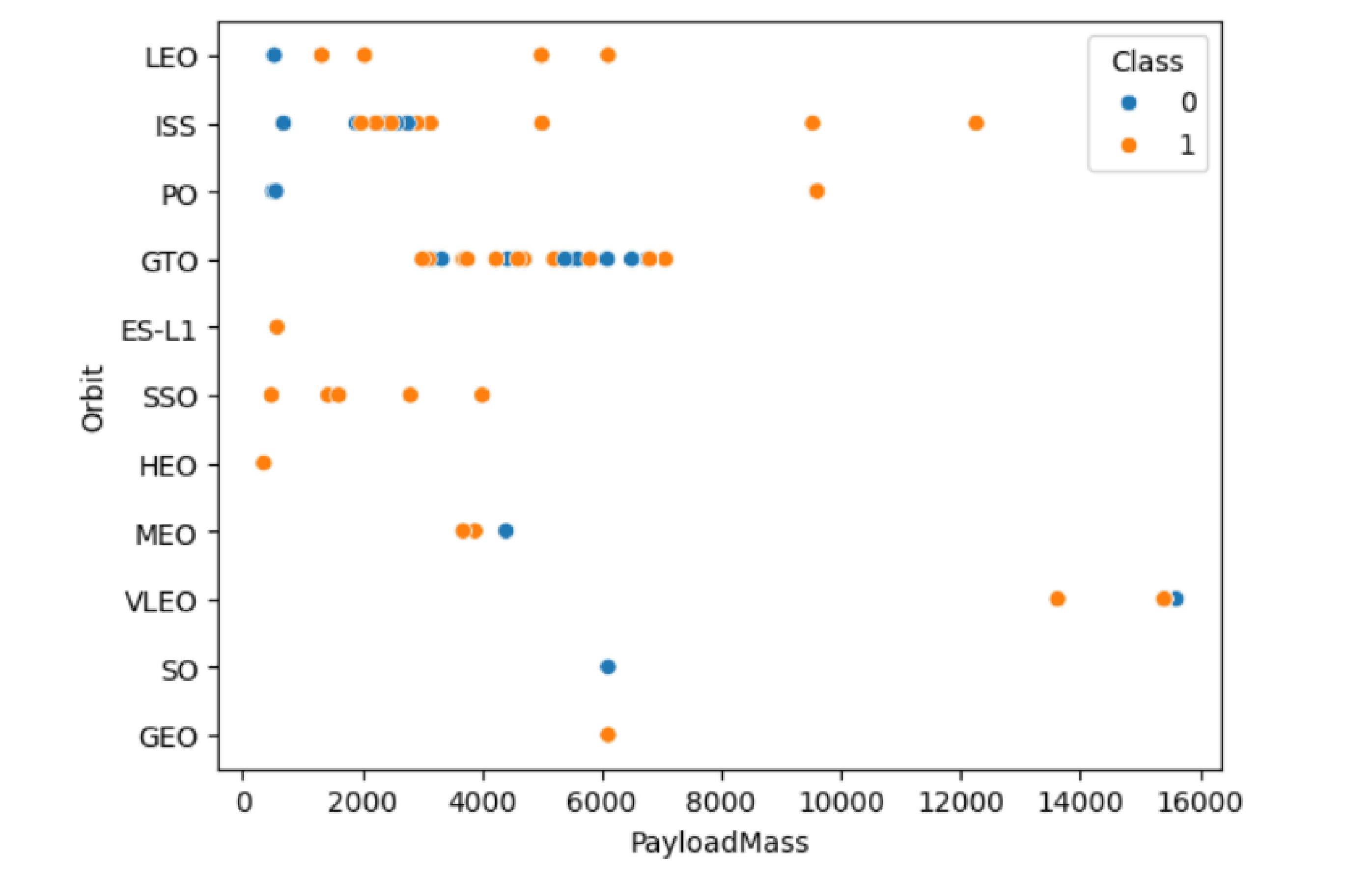


Flight Number vs. Orbit Type

- SSO and HEO are the most successful
- LEO is successful when there are more than 5 launches

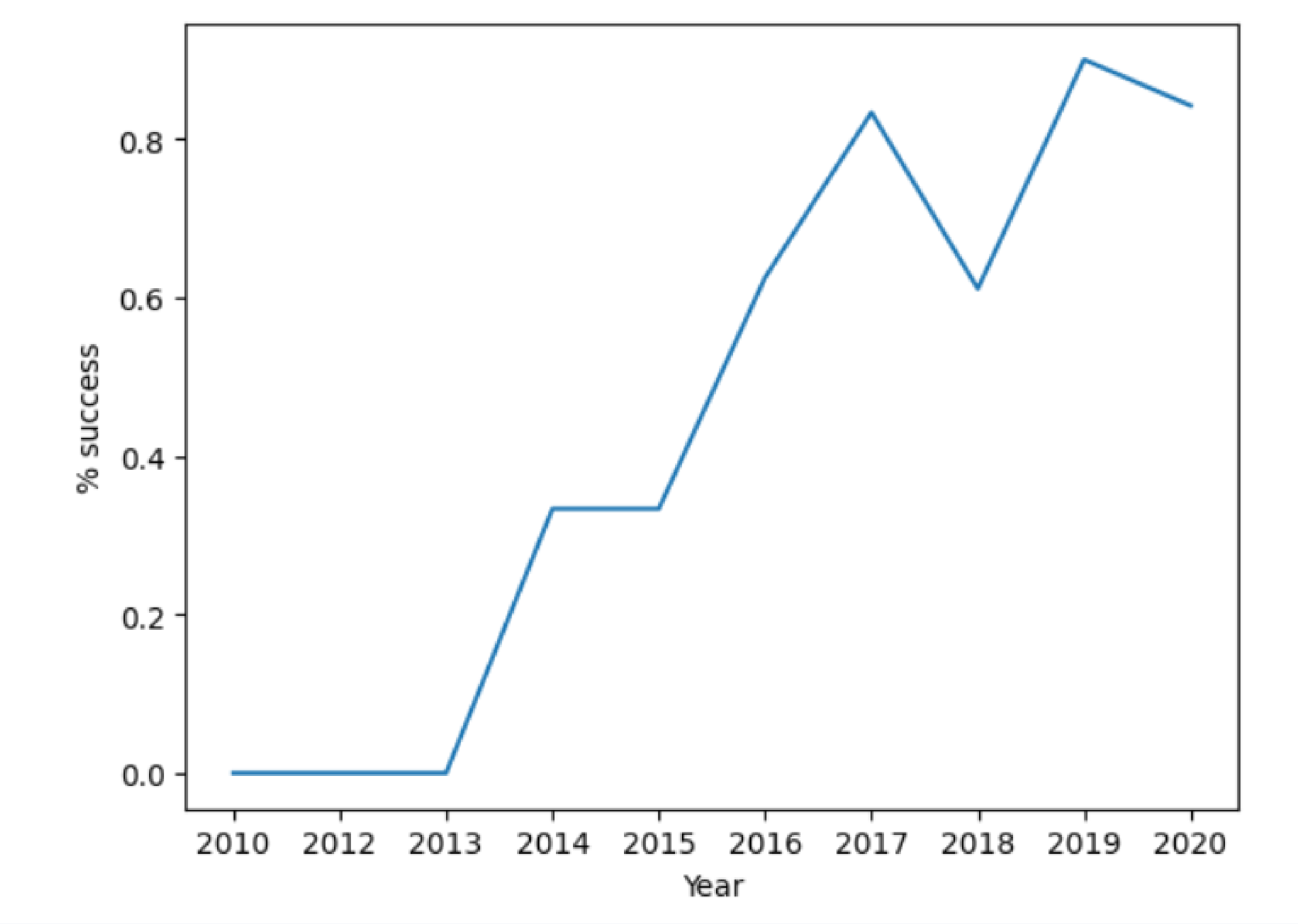


Payload vs. Orbit Type



Launch Success Yearly Trend

- The success rate has improved over the years.



All Launch Site Names

- Find the names of the unique launch sites
- Present your query result with a short explanation here

```
select DISTINCT(Launch_Site) from SPACEXTABLE
```

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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- Present your query result with a short explanation here

```
select * from SPACEXTABLE WHERE Launch_Site like 'CCA%' LIMIT 5
```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcon
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachut
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 (parachut
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attem
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attem
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attem

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
- Present your query result with a short explanation here

```
select SUM(PAYLOAD_MASS__KG_) from SPACEXTABLE WHERE Customer = 'NASA (CRS)' GROUP BY Customer
```

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- Present your query result with a short explanation here

```
select  AVG(PAYLOAD_MASS__KG_) from SPACEXTABLE where Booster_Version = 'F9 v1.1'
```

AVG(PAYLOAD_MASS__KG_)
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

```
select  Date from SPACEXTABLE where Landing_Outcome = 'Success (ground pad)' LIMIT 1
```

Date
2015-12-22

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
- Present your query result with a short explanation here

```
select DISTINCT(Booster_Version) from SPACEXTABLE  where Landing_Outcome = 'Success (drone ship)' and PAYLOAD_MASS__KG_ >= 4000 and PAYLOAD_MASS__KG_ <=6000
```

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
- Present your query result with a short explanation here

```
%sql select COUNT(*) from SPACEXTABLE WHERE Mission_Outcome LIKE '%Success%';
```

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

COUNT(*)
100

```
%sql select COUNT(*) from SPACEXTABLE WHERE Mission_Outcome LIKE '%Failure%';
```

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

COUNT(*)
1

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- Present your query result with a short explanation here

```
select Booster_Version from SPACEXTABLE where PAYLOAD_MASS__KG_ = (select MAX(PAYLOAD_MASS__KG_) from SPACEXTABLE order by PAYLOAD_MASS__KG_ desc)
```

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

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- Present your query result with a short explanation here

```
SELECT substr(Date,6,2), Date, Booster_Version, Landing_Outcome, Launch_Site FROM SPACEXTABLE WHERE Date  LIKE '2015%'  and Landing_Outcome = 'Failure (drone ship)'
```

substr(Date,6,2)	Date	Booster_Version	Landing_Outcome	Launch_Site
01	2015-01-10	F9 v1.1 B1012	Failure (drone ship)	CCAFS LC-40
04	2015-04-14	F9 v1.1 B1015	Failure (drone ship)	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

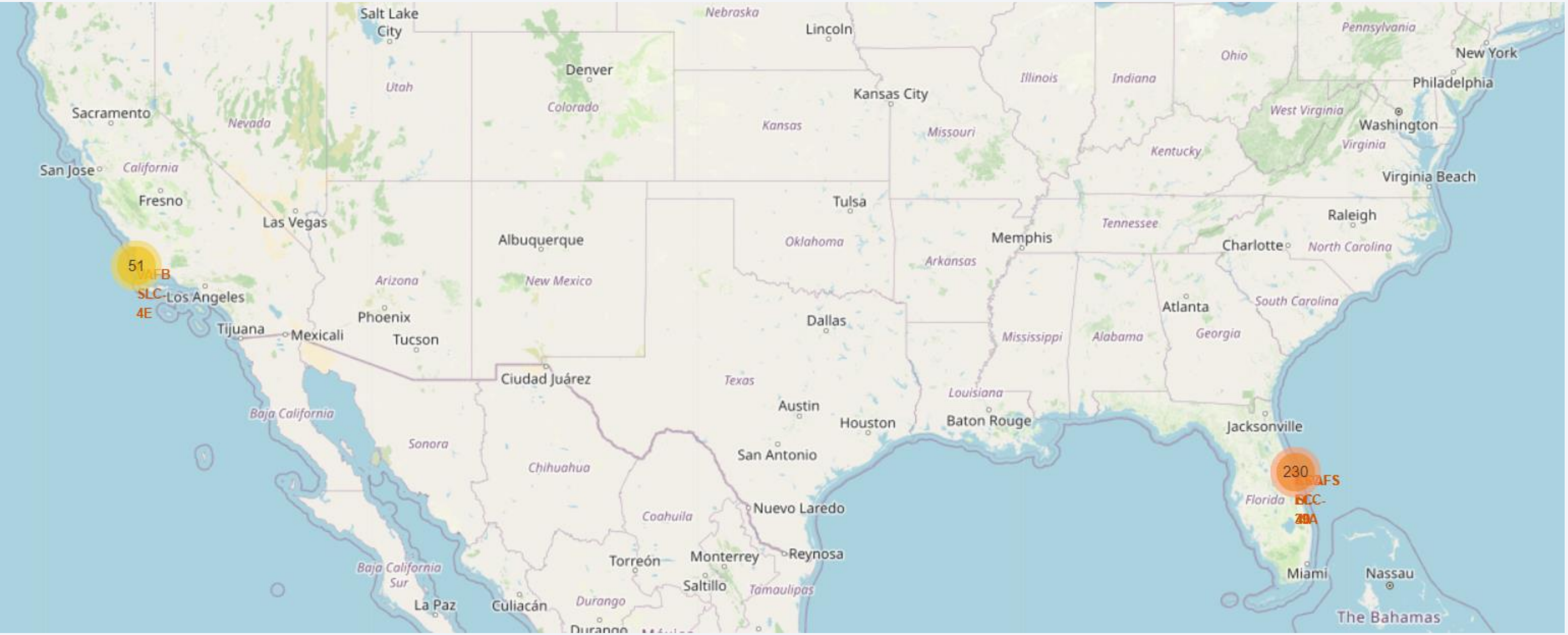
- Present your query result with a short explanation here

Section 3

Launch Sites Proximities Analysis

Launch sites

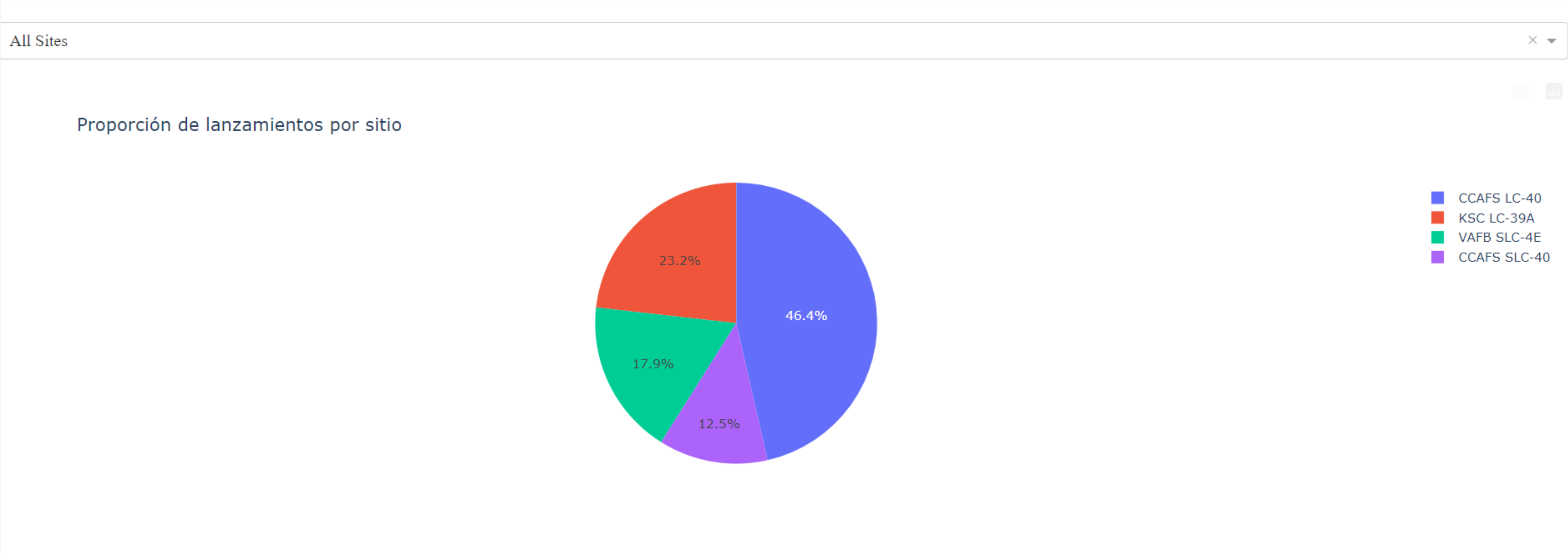
Number of launches, per site



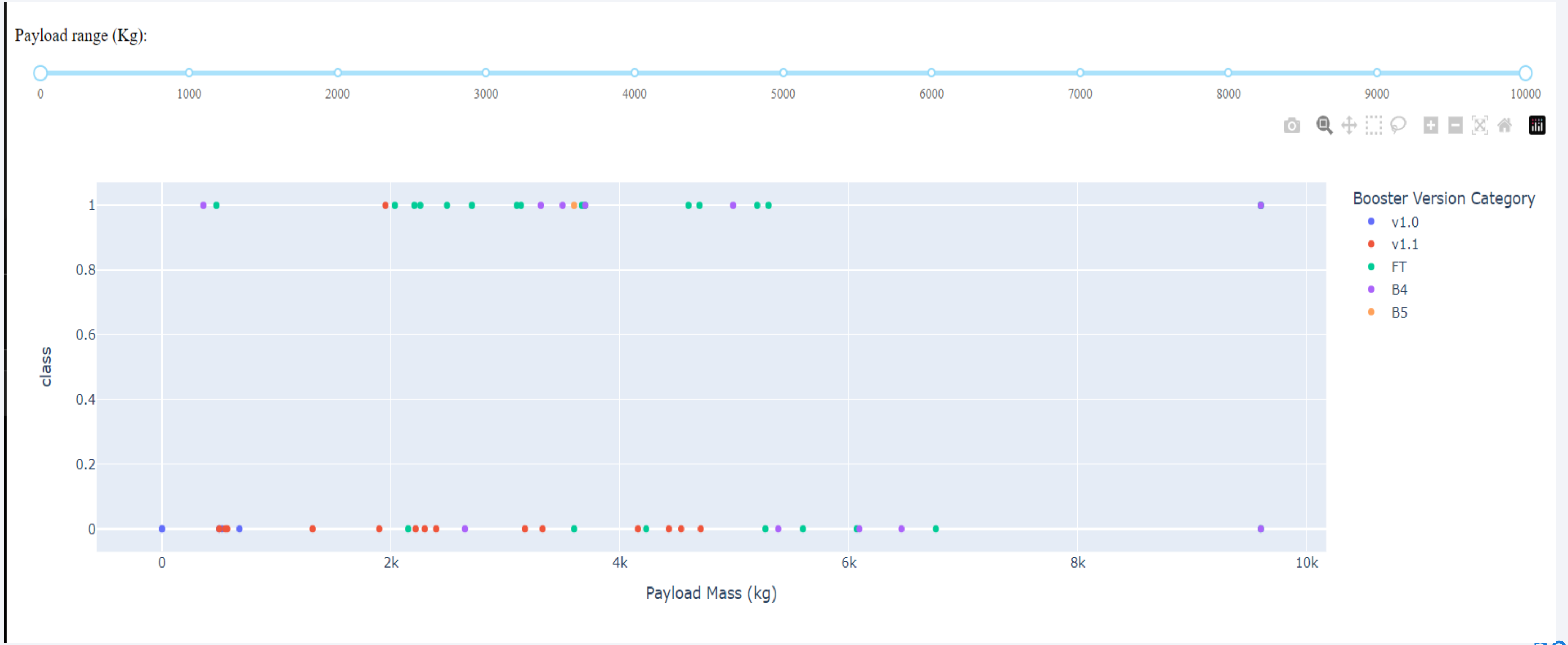
Section 4

Build a Dashboard with Plotly Dash

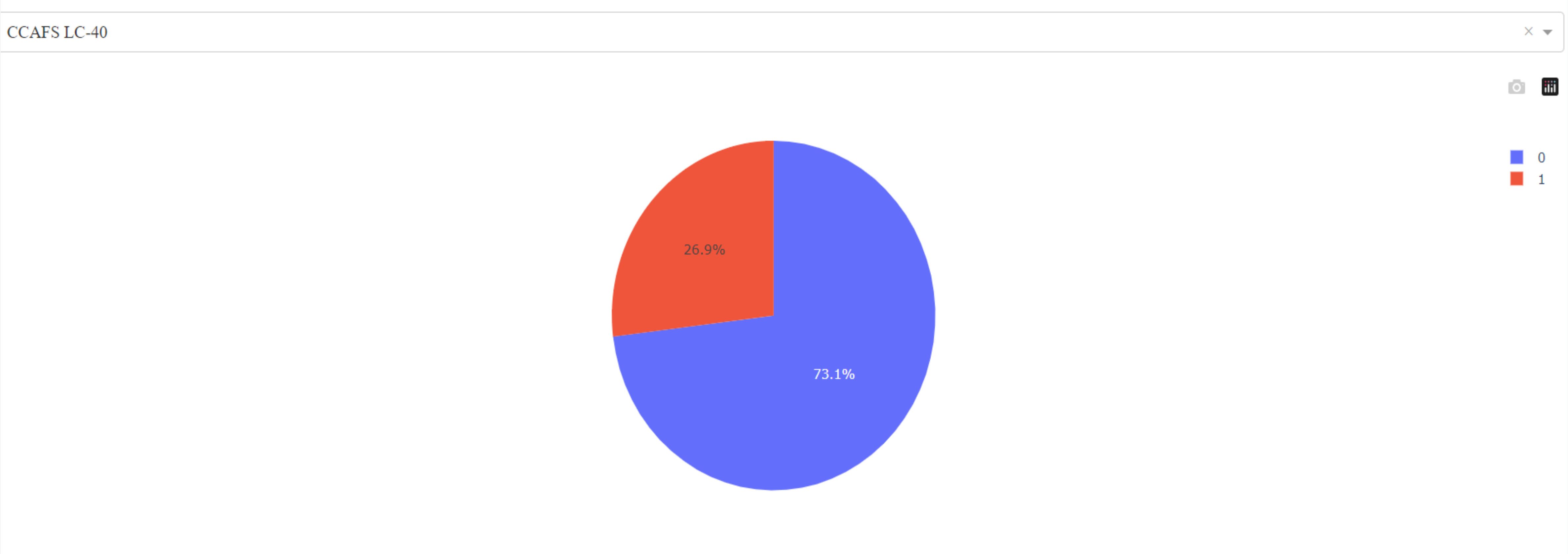
Successful launch rate at all sites



Relationship between payload mass and launch success segmented by booster type



Successful launch rate in CCSFS LC-40

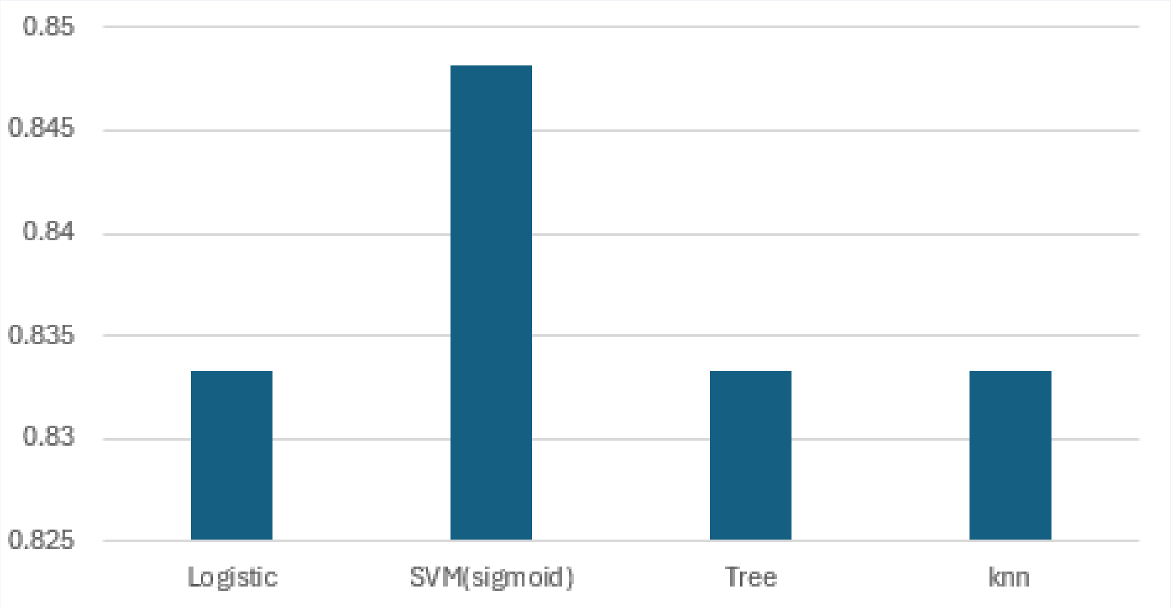


Section 5

Predictive Analysis (Classification)

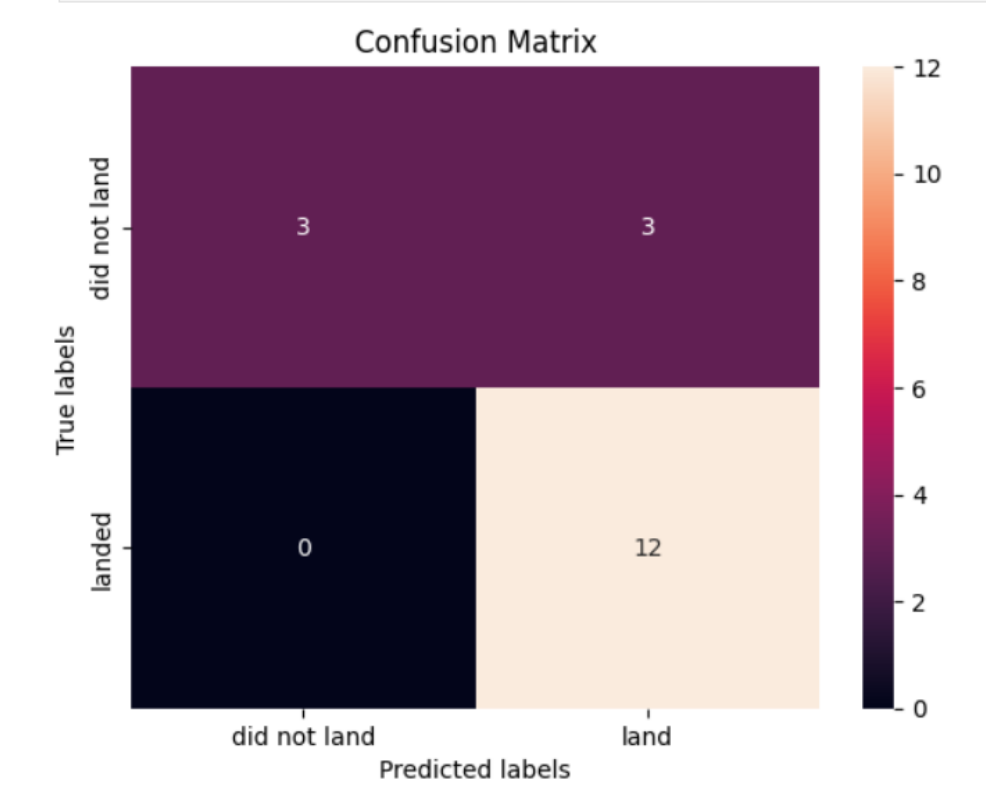
Classification Accuracy

- All models have the same Accuracy. (SVM slightly a little more)



Confusion Matrix

- Show the confusion matrix of the best performing model with an explanation



Conclusions

- The best success rate in launches is positively related to the number of launches that occur on that site Point 2
- Orbits ES-L1, GEO, HEO, SSO, VLEO had the most success rate.
- The SVM model is the best machine learning algorithm for this task.

Thank you!