IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Thant Htet Oo @ Feniz
31.12.2022

Outline

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

Executive Summary

- Summary of methodologies
  - Data are collected from SpaceX API and processed
  - EDA with SQL
  - Visualize using Plotly and Folium
  - Train model with data
- Summary of all results
  - All the results are described in the coming slides

Introduction

- Project background and context
  - 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. 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
  - Predict if the Falcon 9 first stage will land successfully

Section 1

Methodology

Methodology

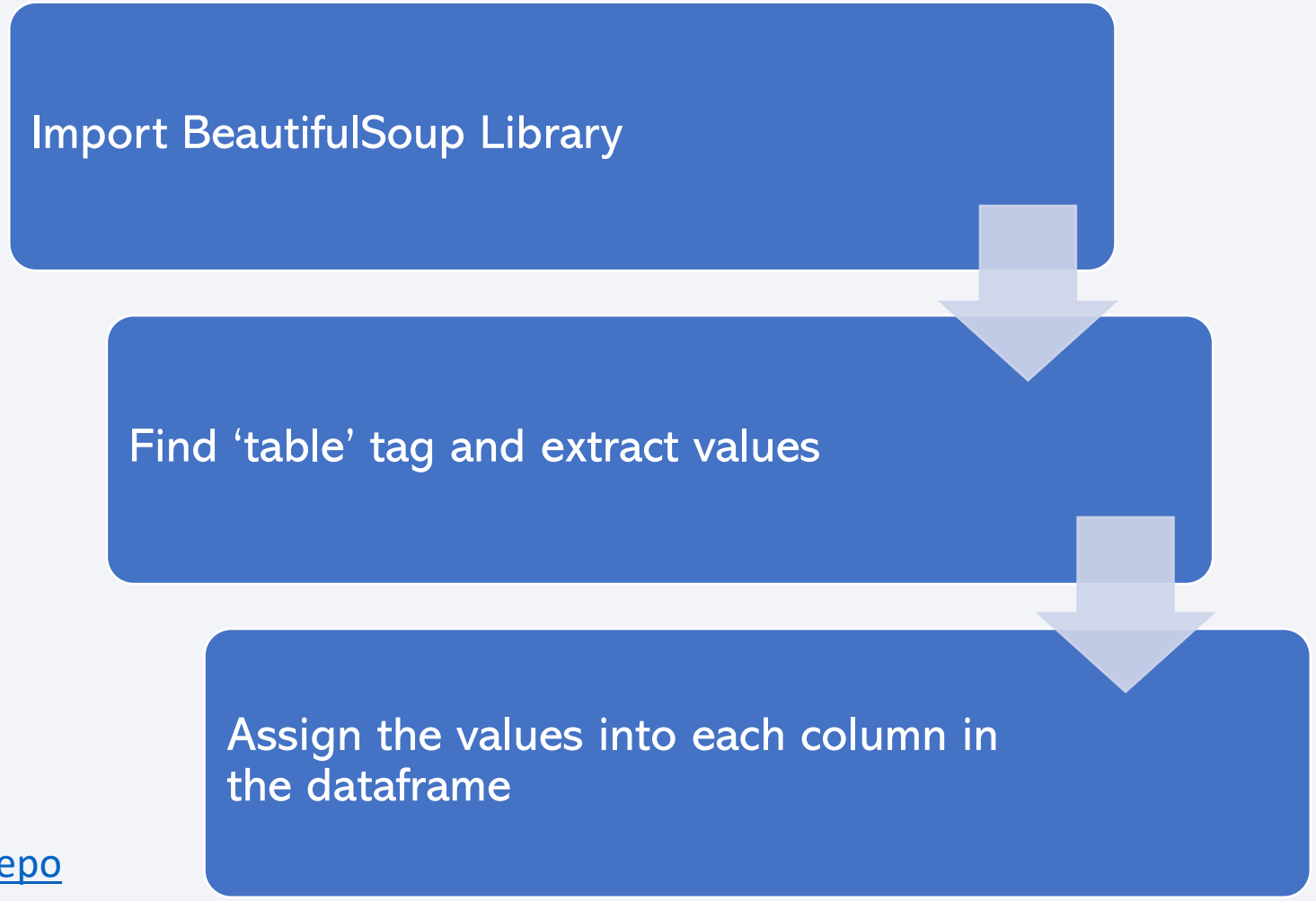
Executive Summary

- Data collection methodology:
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection & Wrangling – SpaceX API

GitHub Repo

Data Collection - Scraping



GitHub Repo

EDA with Data Visualization

- In this project, scatter plot and line plots are visualized

  Scatter Plot

  - To visualize how much extent variables depend each other for the success of the outcome
  - To visualize how payload mass affects the outcome of the launch depending on the site

  Line Plot

  - To determine the trend to look up the factors how and why these results happened

GitHub Repo

EDA with SQL

- SQL queries used in this project are mostly
- ‘SELECT query’
- Using min(), max(), count() function
- Subqueries method to extract only the data wanted

GitHub Repo

Build an Interactive Map with Folium

- Markers and Circles are added in the launch sites in the map using latitude and longitude from the data collected so that the launch sites are more visible and distinct in the map

- A line is drawn from the launch site and the nearest coastline

GitHub Repo

Build a Dashboard with Plotly Dash

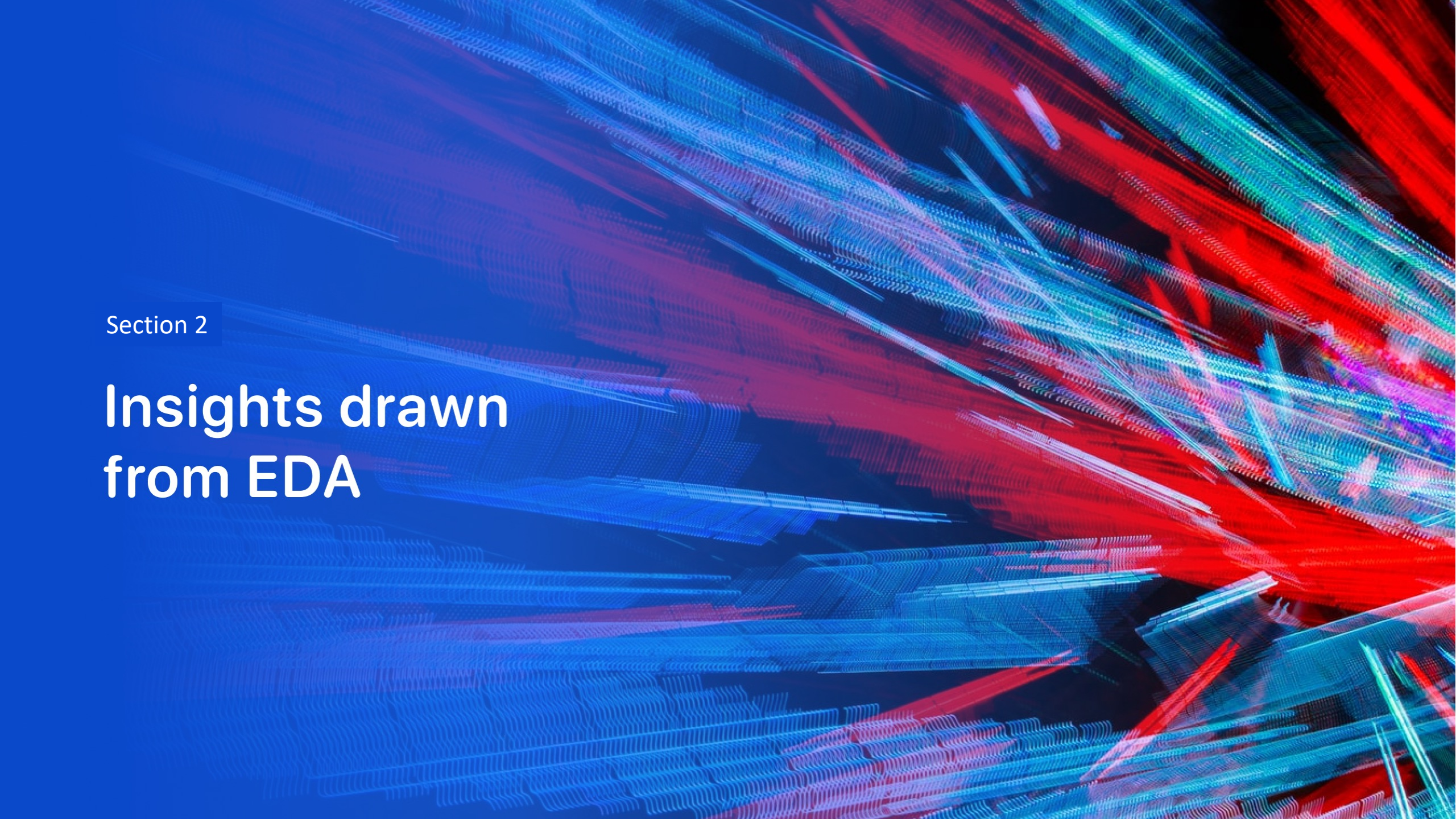
- Pie charts and Scatter plots are added in the dashboard with options of all launch sites or only one site
- Pie charts are effective in showing the proportions as a whole
- Scatter plots are effective in determining the outliers and relationship between the variables

GitHub Repo

Predictive Analysis (Classification)

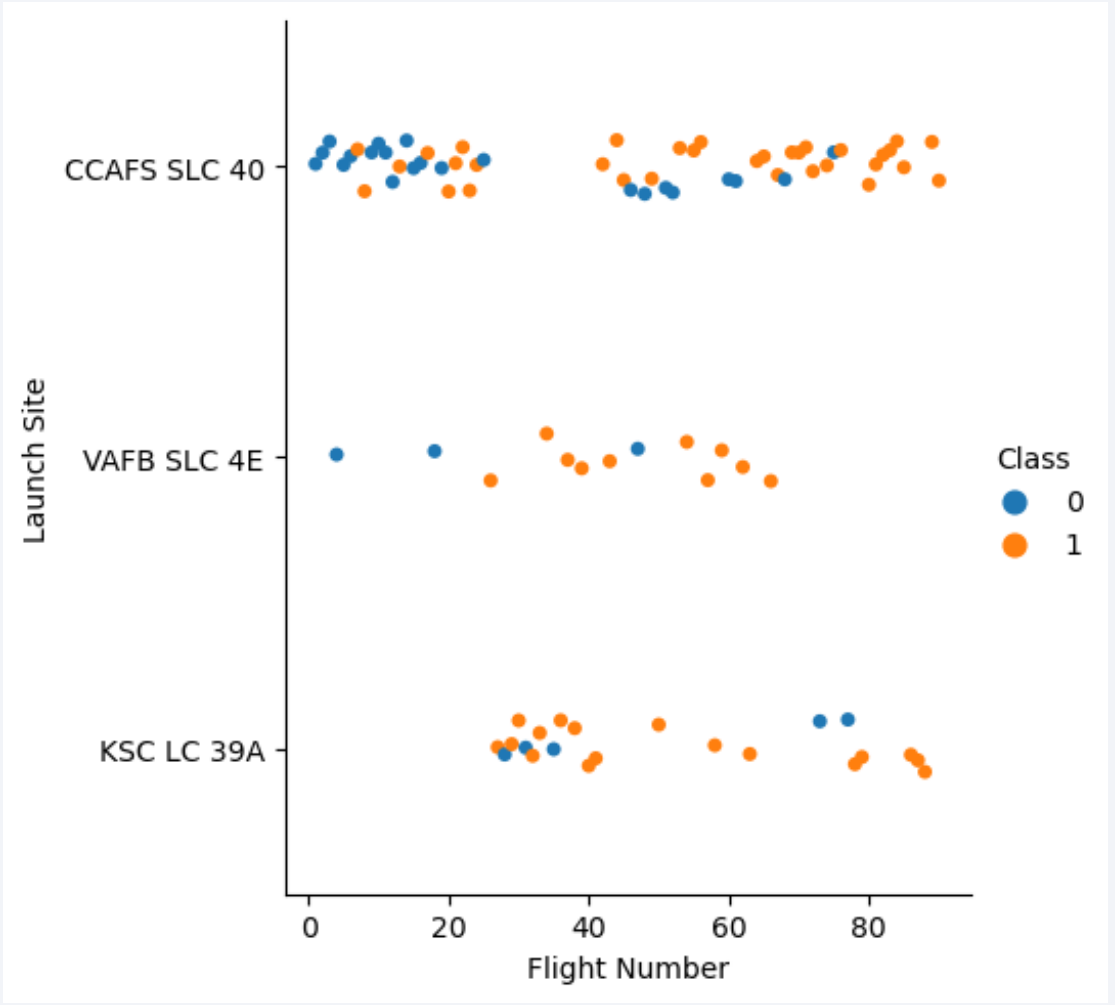
- Some machine learning algorithms, such as Logistic Regression, SVM, Decision Tree, KNN, are trained with the collected data
- Each algorithm is trained with GridSearchCV function to determine the best parameters ,which can maximize the accuracy score
- After that, compare the best accuracy score out of all algorithm and decide one to predict

GitHub Repo


Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

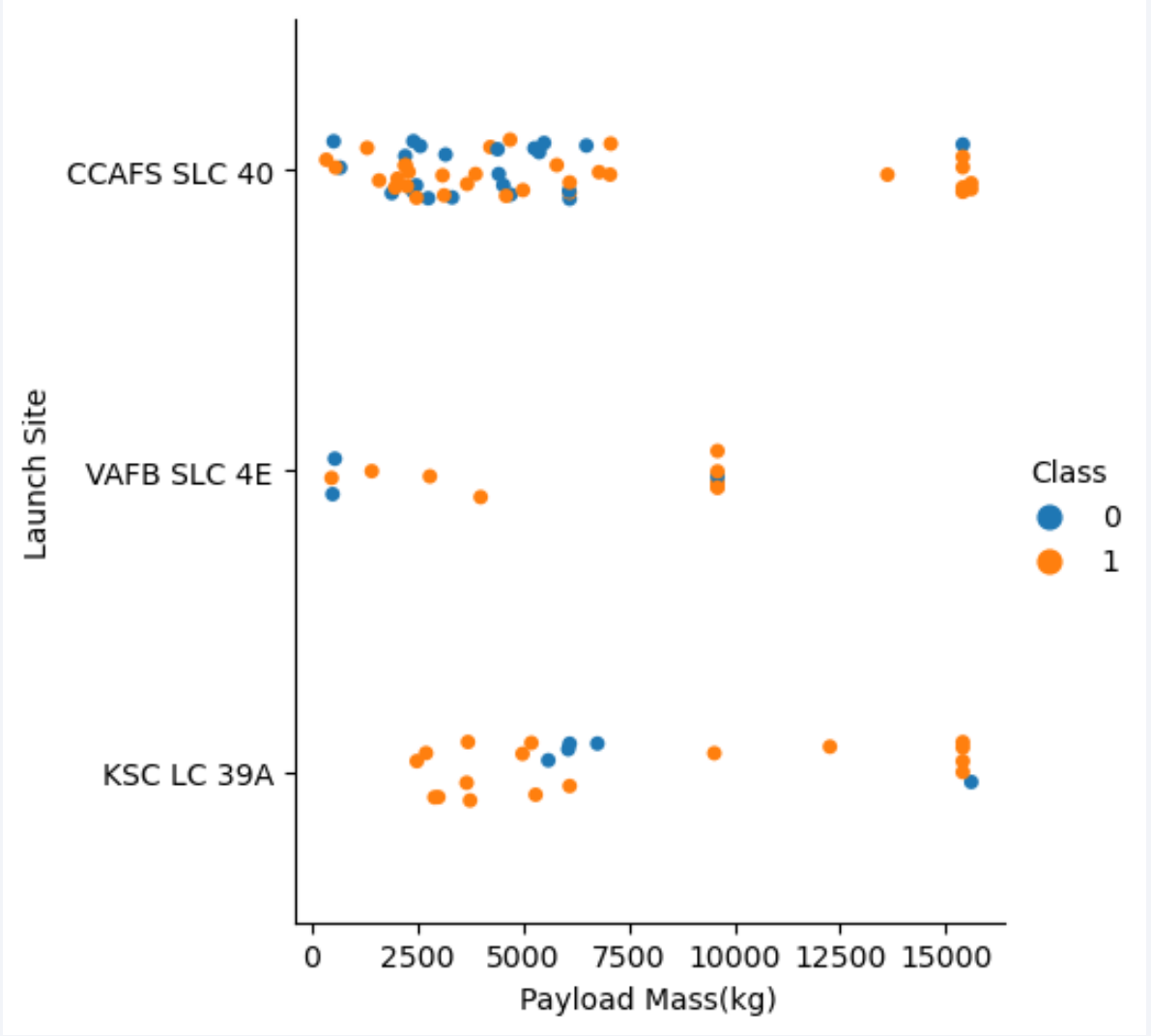
Overall, the more the flights, the higher the success rate of launch in all sites



GitHub Repo

Payload vs. Launch Site

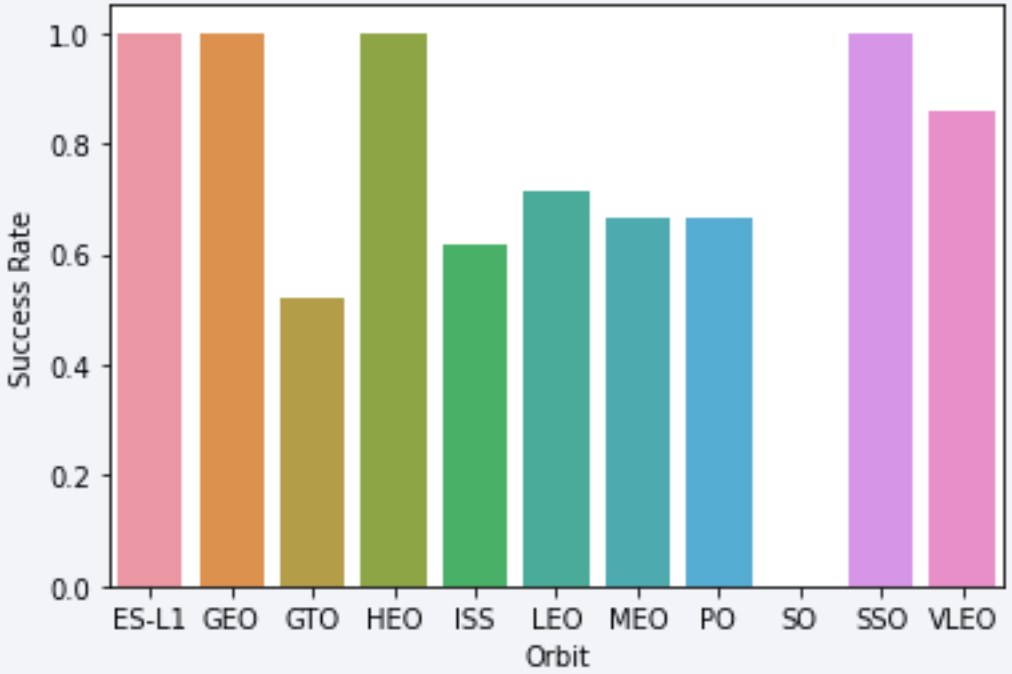
For the VAFB-SLC launch site, there are no rockets launched for heavy payload mass(> 10000).



GitHub Repo

Success Rate vs. Orbit Type

- Orbits
- ES-L1, GEO, HEO and SSO have a success rate of 1.0



GitHub Repo

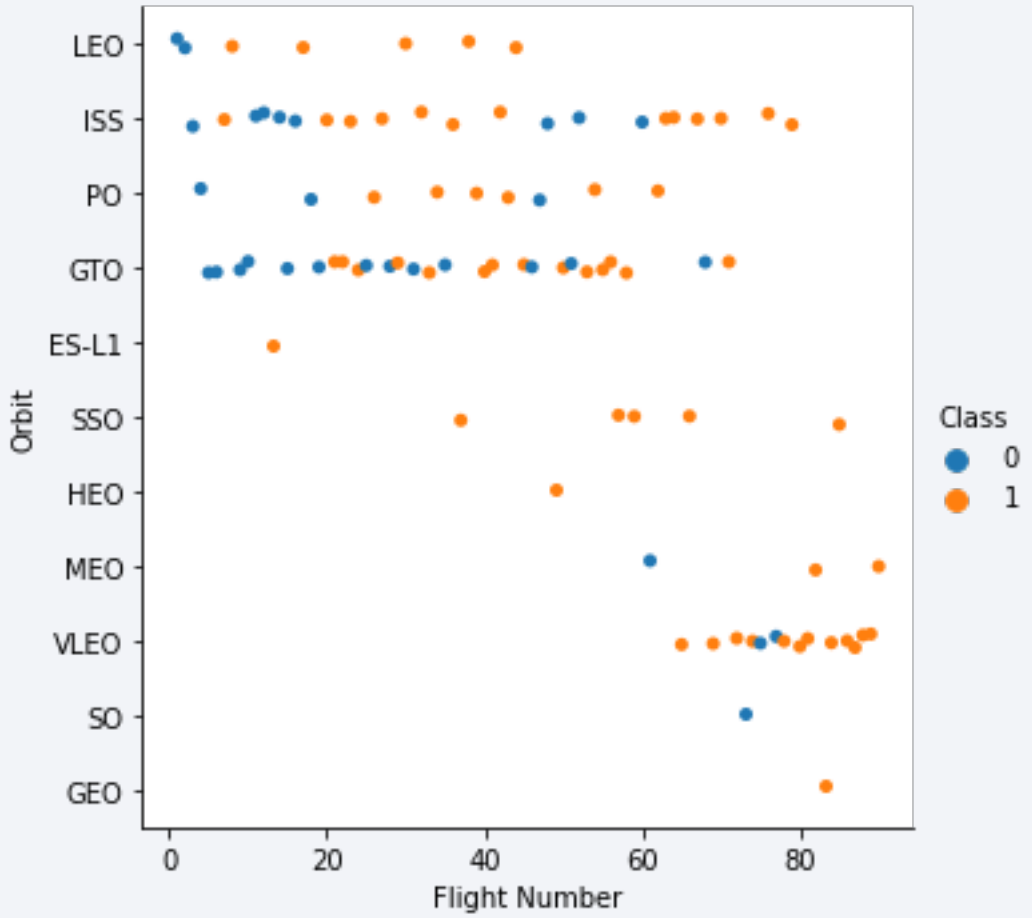
Flight Number vs. Orbit Type

In the LEO orbit,

the Success appears related to the number of flights;

On the other hand,

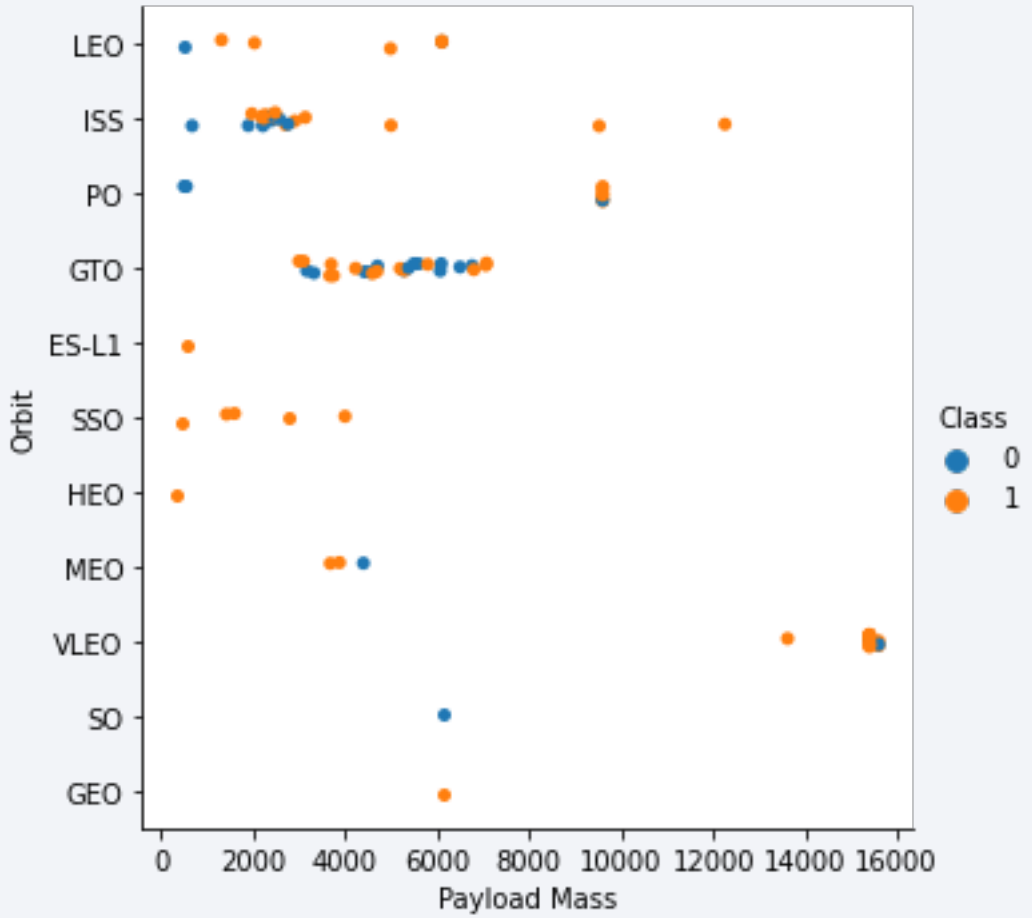
there seems to be no relationship between flight number when in GTO orbit



GitHub Repo

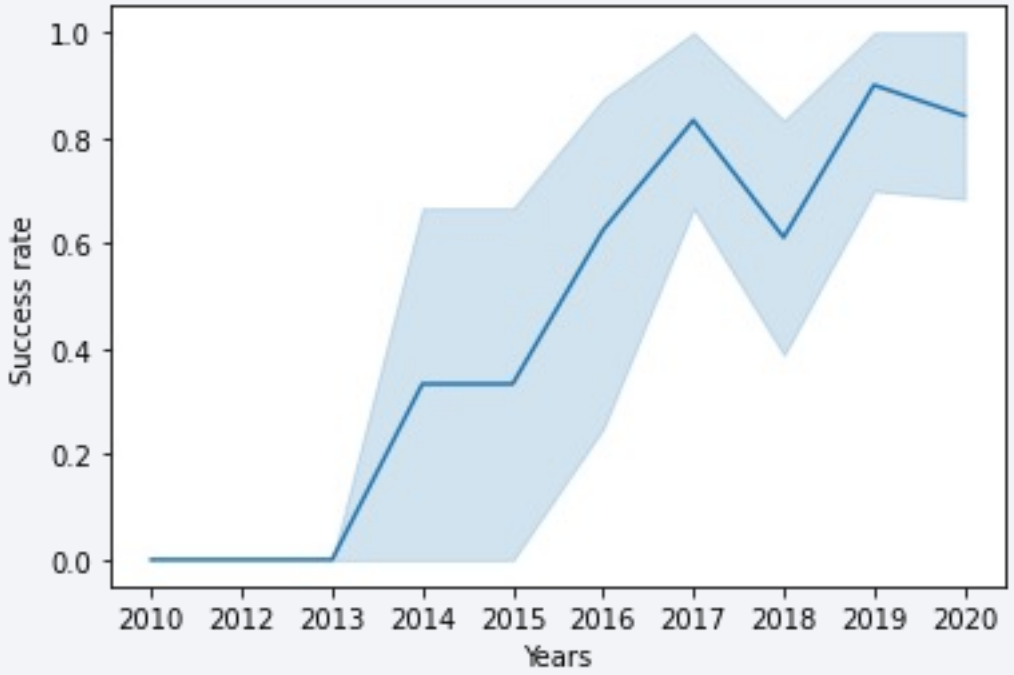
Payload vs. Orbit Type

- 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



GitHub Repo

Launch Success Yearly Trend



The success rate since 2013 kept increasing till 2020

GitHub Repo

All Launch Site Names

```
(%sql SELECT DISTINCT(launch_site) FROM SPACEX)
```

- CCAFS
- LC-40CCAFS
- SLC-40KSC
- LC-39AVAFB SLC-4E

GitHub Repo

Launch Site Names Begin with 'CCA'

- `%sql SELECT * FROM SPACEX WHERE launch_site LIKE 'CCA%' LIMIT 5`

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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

GitHub Repo

Total Payload Mass

- `%sql SELECT SUM(payload_mass__kg_) FROM SPACEX WHERE customer='NASA (CRS)'`

45596

GitHub Repo

Average Payload Mass by F9 v1.1

- `%sql SELECT AVG(payload_mass__kg_) FROM SPACEX WHERE booster_version = 'F9 v1.1'`

2928

GitHub Repo

First Successful Ground Landing Date

- `%sql SELECT min(DATE) FROM SPACEX WHERE landing__outcome = 'Success (ground pad)'`

2015-12-22

GitHub Repo

Successful Drone Ship Landing with Payload between 4000 and 6000

- `%sql SELECT payload FROM SPACEX WHERE landing__outcome = 'Success (drone ship)' and payload_mass__kg_ > 4000 and payload_mass__kg_<6000`

payload
JCSAT-14
JCSAT-16
SES-10
SES-11 / EchoStar 105

GitHub Repo

Total Number of Successful and Failure Mission Outcomes

- `%sql SELECT COUNT(mission_outcome) FROM SPACEX`

101

GitHub Repo

Boosters Carried Maximum Payload

- `%sql SELECT 'booster_version' FROM SPACEX WHERE payload_mass__kg_ = (SELECT max(payload_mass__kg_) FROM SPACEX)`

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

GitHub Repo

2015 Launch Records

- `%sql SELECT landing__outcome, booster_version, launch_site FROM SPACEX WHERE DATE LIKE '2015%'`

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Controlled (ocean)	F9 v1.1 B1013	CCAFS LC-40
No attempt	F9 v1.1 B1014	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40
No attempt	F9 v1.1 B1016	CCAFS LC-40
Precluded (drone ship)	F9 v1.1 B1018	CCAFS LC-40
Success (ground pad)	F9 FT B1019	CCAFS LC-40

GitHub Repo

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

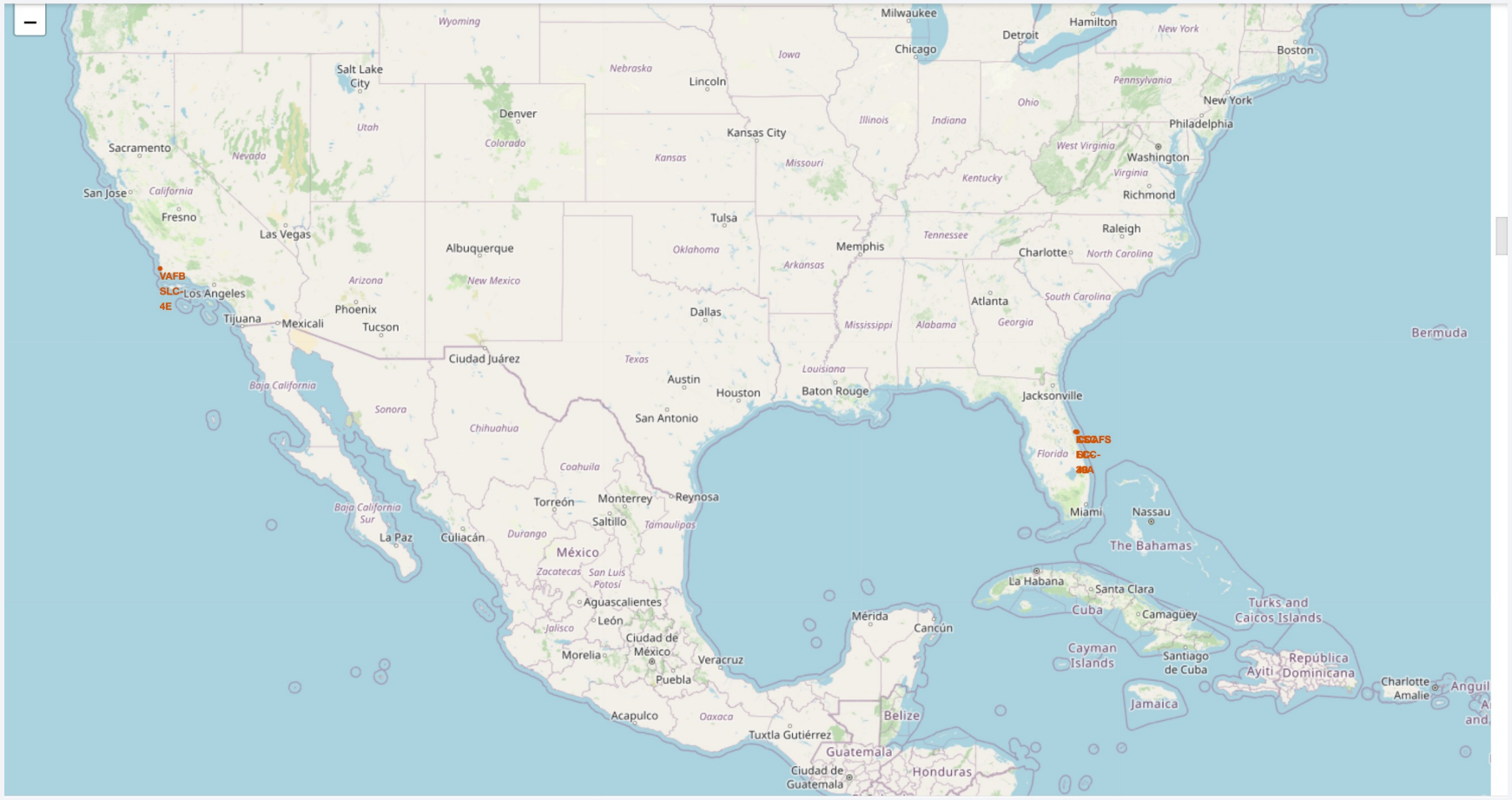
- `%sql SELECT landing__outcome, COUNT(*) as Count FROM (SELECT * FROM SPACEX WHERE DATE > '2010-06-04' and DATE > '2017-03-20') GROUP BY landing__outcome ORDER BY Count DESC`

landing__outcome	COUNT
Success	38
No attempt	12
Success (drone ship)	9
Success (ground pad)	6
Failure	3
Controlled (ocean)	2

GitHub Repo

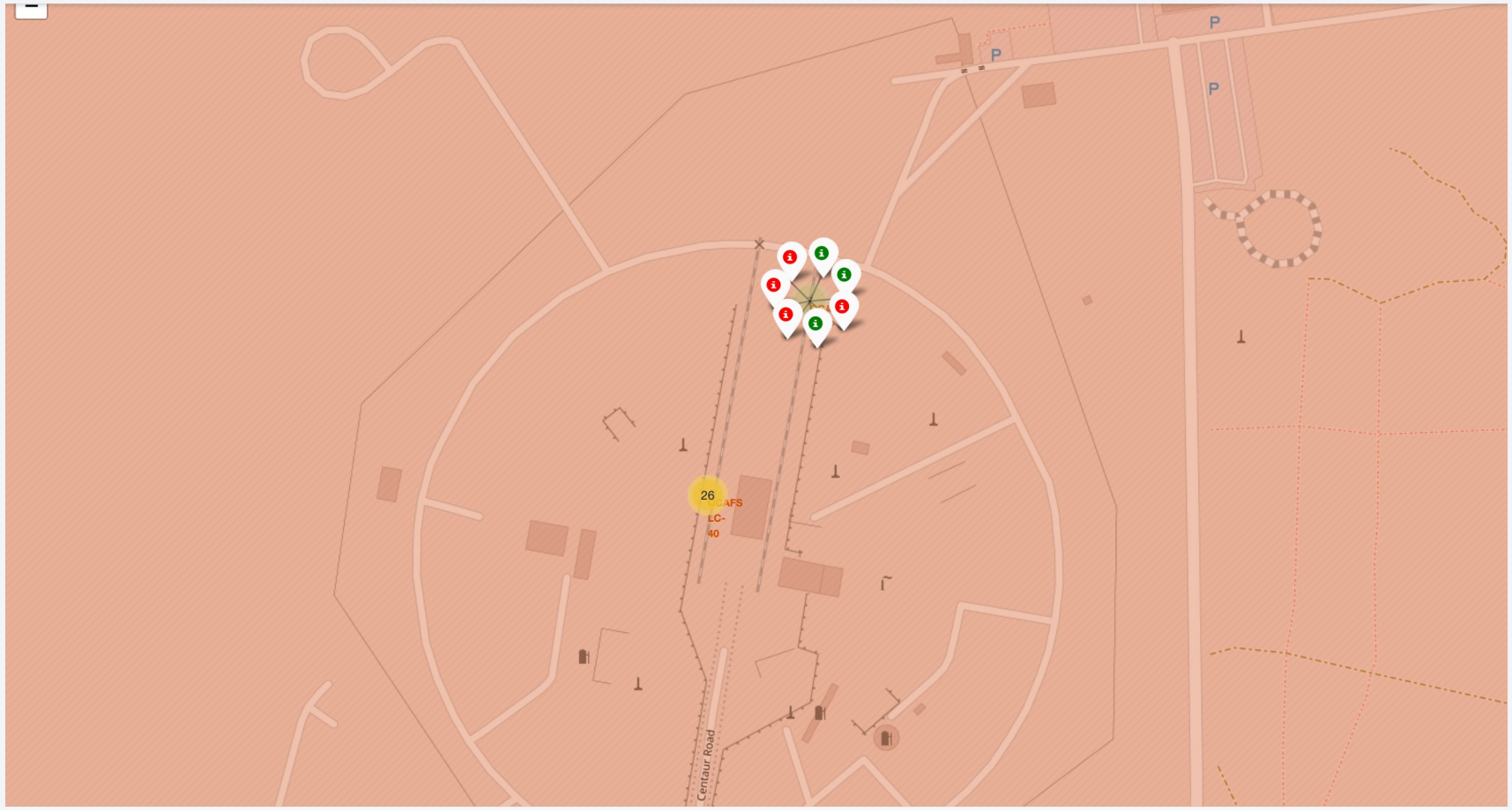

Section 3
Launch Sites
Proximities Analysis

SpaceX Launch Sites

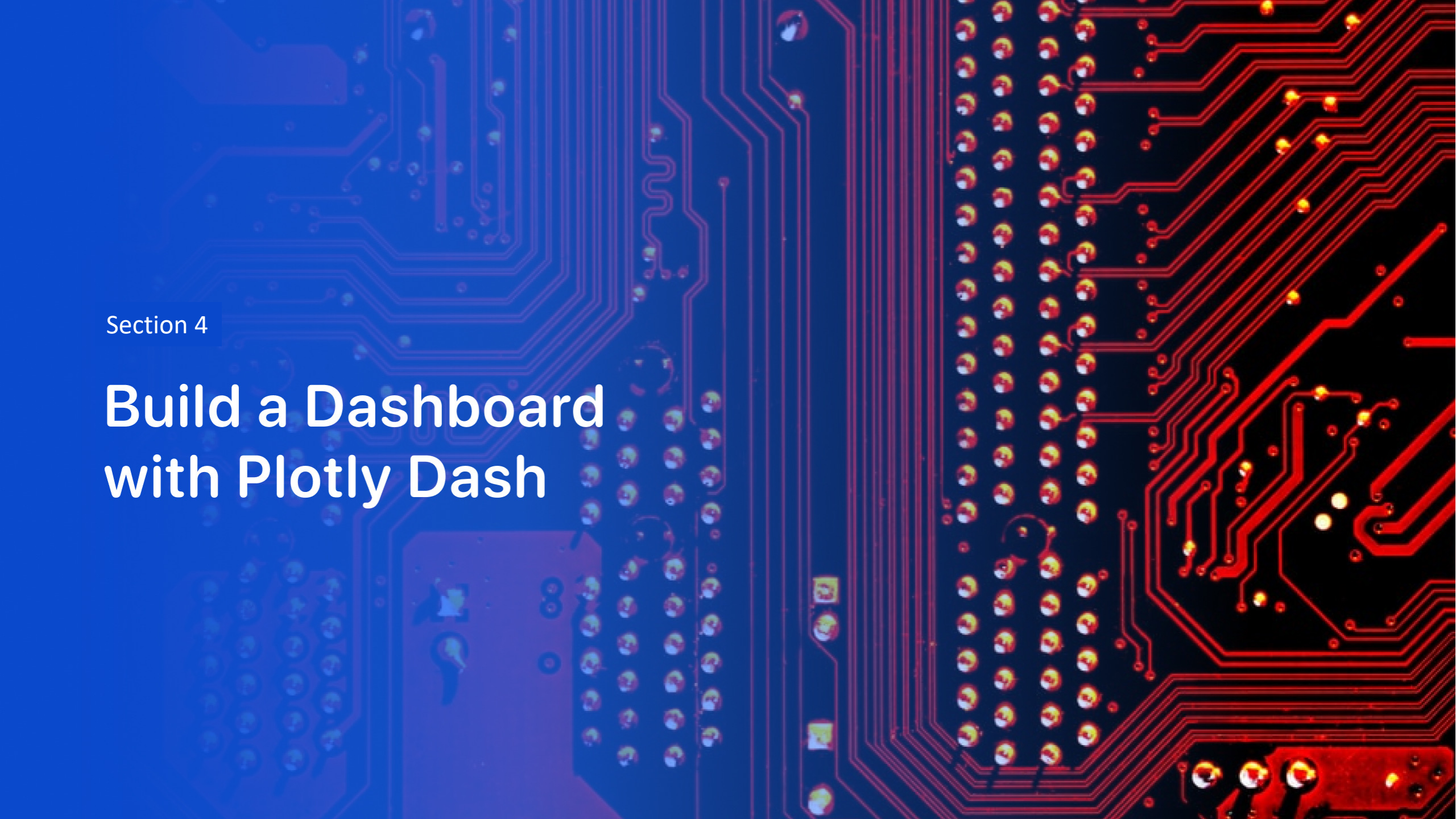


[GitHub Repo]

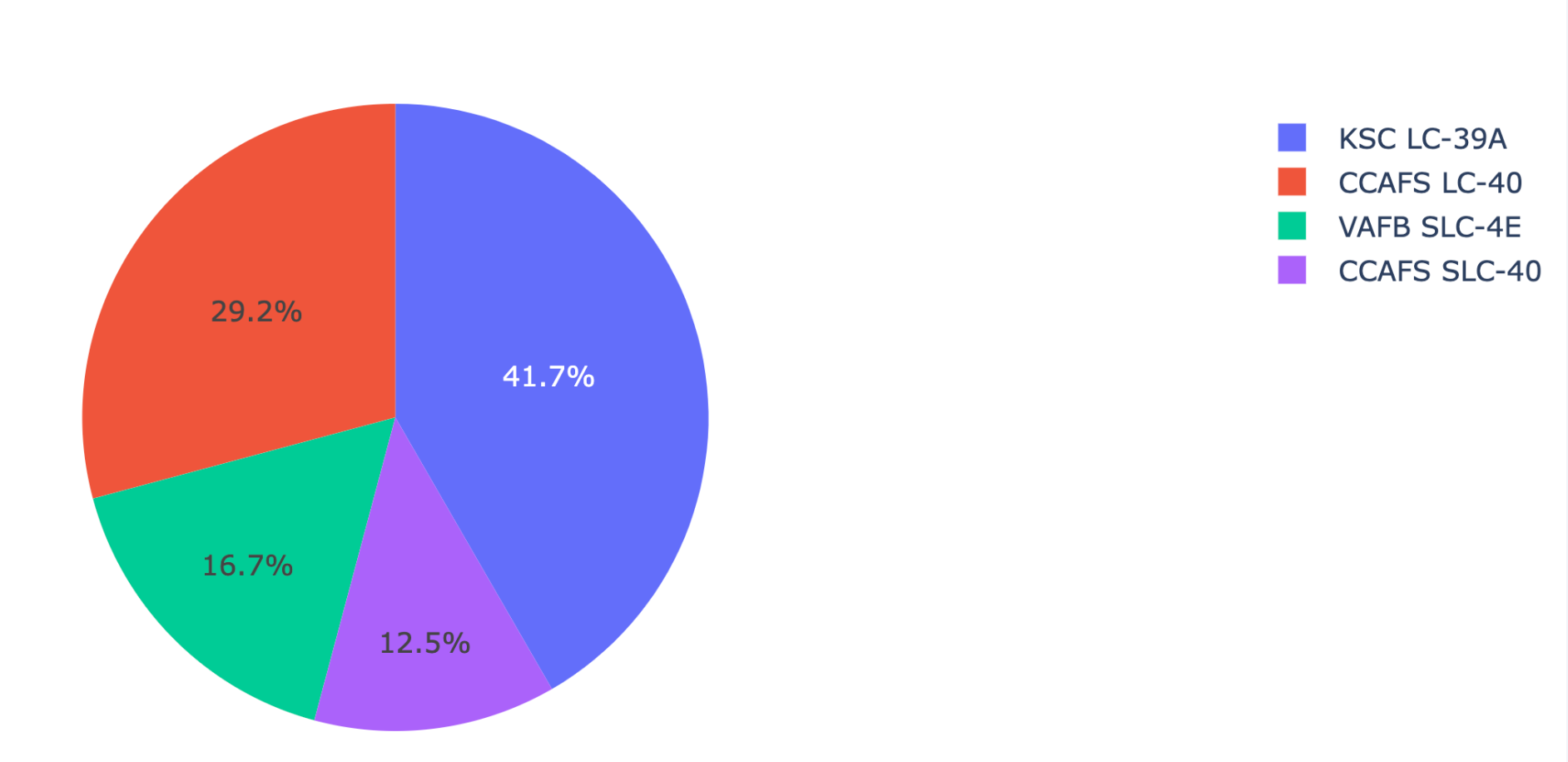
Colored-Labeled Launch Outcomes in each site



GitHub Repo

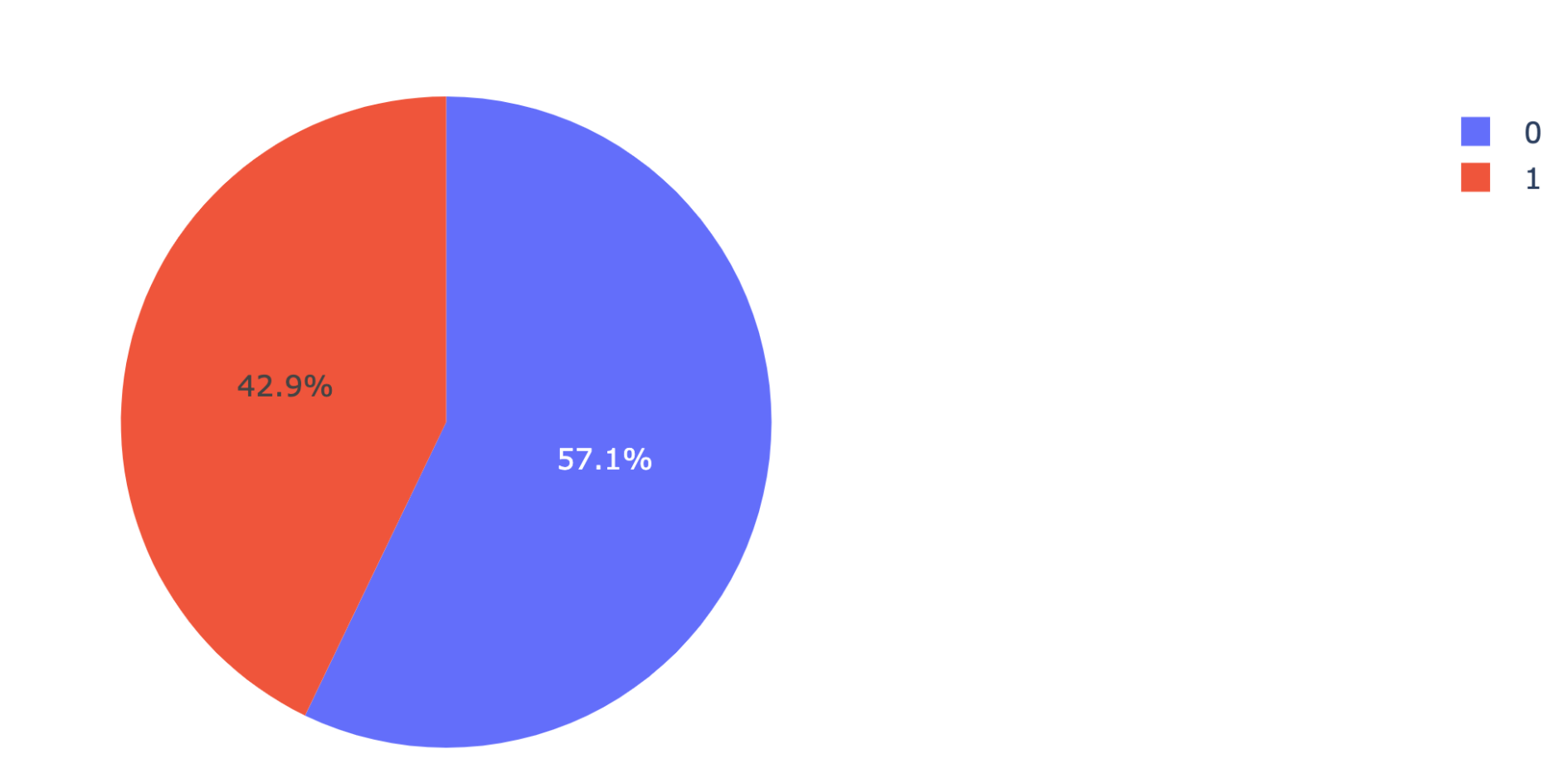

Section 4
Build a Dashboard with Plotly Dash

Total Launches for All Sites



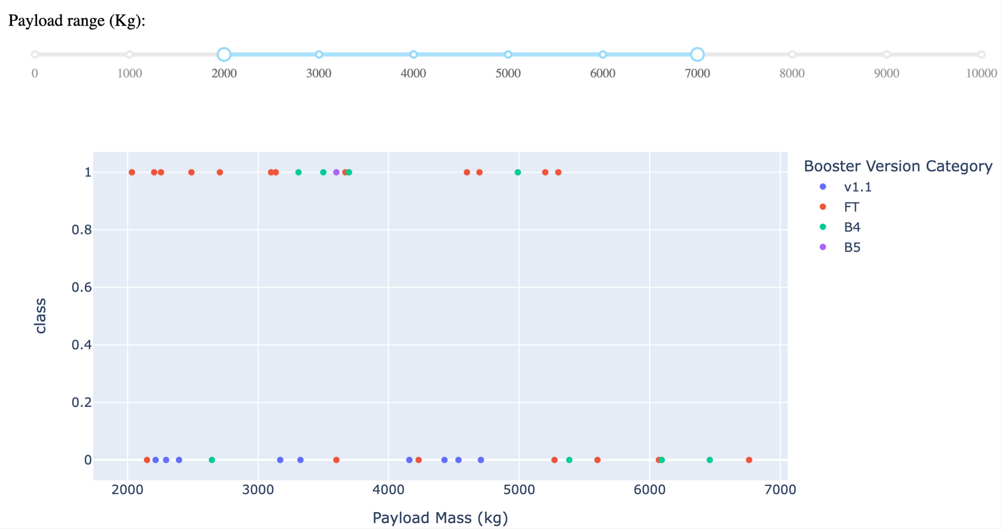
GitHub Repo

CCAFS SLC-40 Launch Success Rate



GitHub Repo

Payload vs Launch Outcome



[GitHub Repo]

Section 5

Predictive Analysis (Classification)

Confusion Matrix of Logistic Regression

GitHub Repo

Thank you!