IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Joseoh Okelly
July 22, 2024

Outline

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

Executive Summary

- Summary of methodologies
- Jupyter labs spacex data collection api
- Jupyter labs webscraping Falcon 9
- Labs Jupyter spacex data wrangling
- Jupyter labs eda sql coursera sqlite
- Eda dataviz using pandas and matplotlib
- Spacex launch site locations Analysis
- Build Interactive Dashboard – spacex dashboard
- Spacex Mchine Learning Preductions

Summary of results

- EDA
- Visual Analsis
- Predictive Analysis

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

  In this capstone, we will predict if the Falcon 9 first stage will land successfully using data from Falcon 9 rocket launches advertised on its website.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- Describe how data sets were collected.
- Data was collected using Spacex Restful webservices, using Get request Spacecx api.

  The request was later stored for processing using Pandas Dataframe

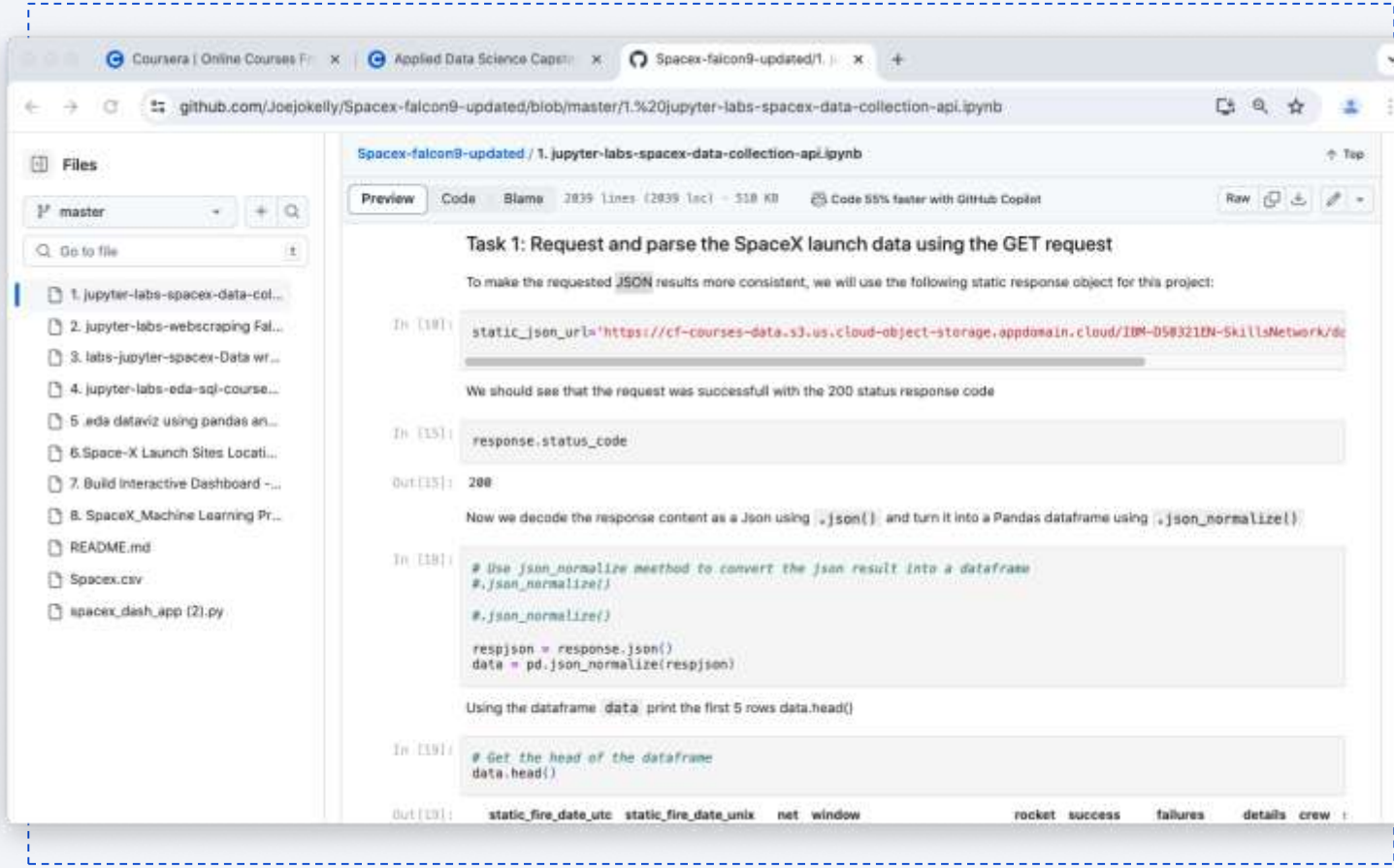
  Data was collected using Spacex Restful webservices, Get request Spacecx api.

- Web scraping performed to Falcon9 historical records.

- You need to present your data collection process use key phrases and flowcharts

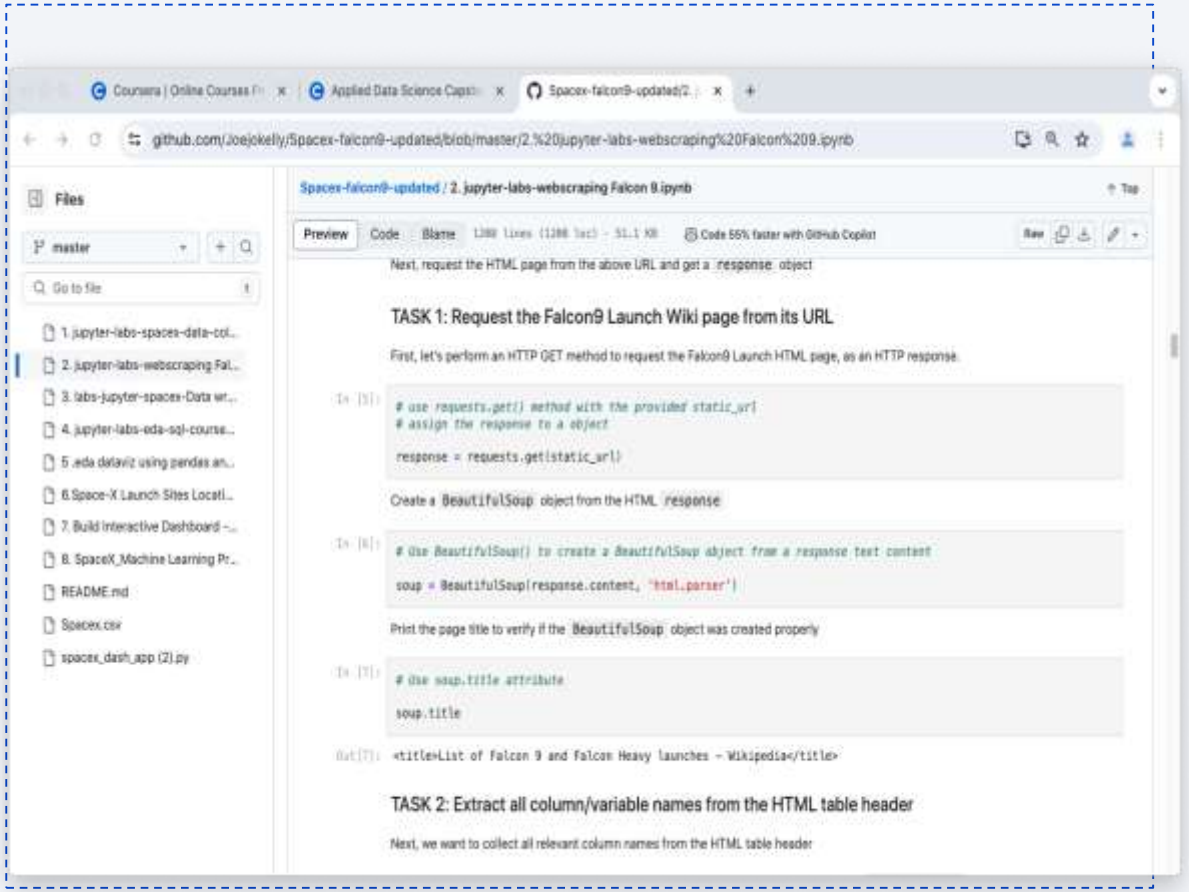
Data Collection – SpaceX API

- Data collected using Spacex Api (Restful call using get parameter)
- Get request is decoded by pandas data frame and it is stored for processing.
- https://github.com/Joejokelly/Spacex-falcon9-updated/blob/master/1.%20jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- Webscraping is performed using historical launch records using Beautiful utilility, next the Data frame is created by passing the Launch HTML file.
- Github url :
- https://github.com/Joejokelly/Spacex-falcon9-updated/blob/master/2.%20jupyter-labs-webscraping%20Falcon%209.ipynb purpose



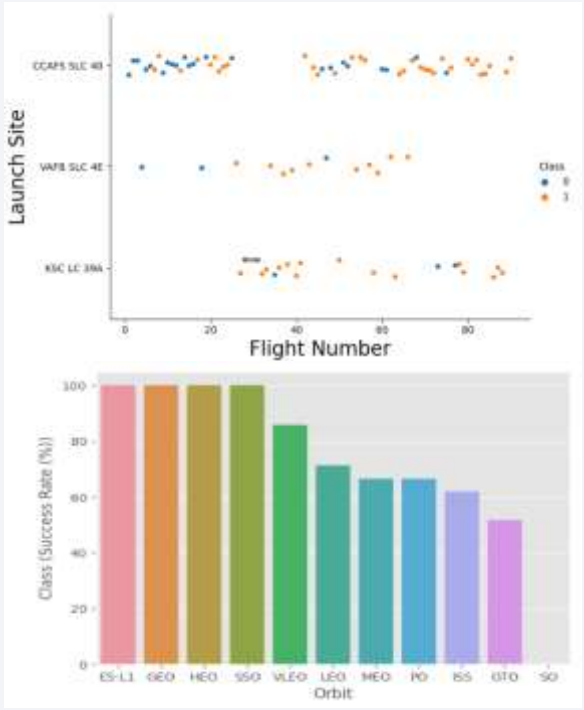
EDA with Data Visualization

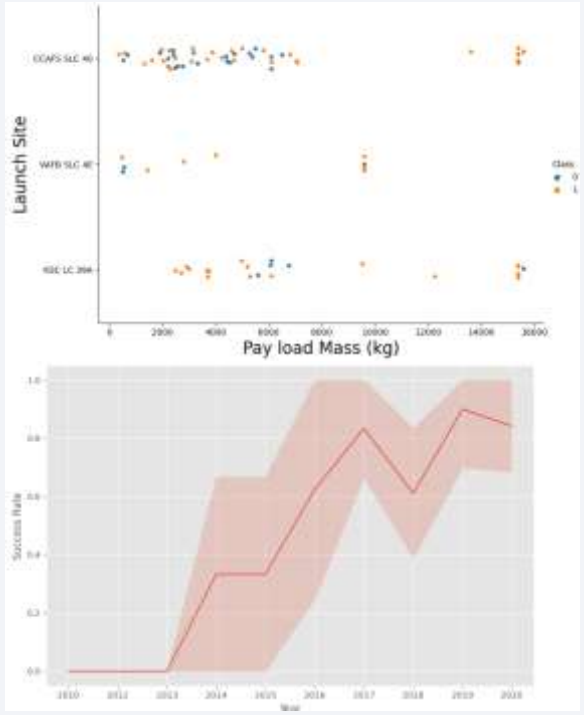
- Data Analysis and Feature Engineering using Pandas and Matplotlib
  - Exploratory Data Analysis
  - Preparing Data Feature Engineering

- https://github.com/Joejokelly/Spacex-falcon9-updated/blob/master/5%20.eda%20dataviz%20using%20pandas%20and%20matplotlib%20.ipynb

EDA with sql (contd)..

EDA with Data Visualization (Plots Cont….)





EDA with SQL

- **Display the names of the unique launch sites in the space mission**
- **%sql** select distinct Launch_site as Launch_site from SPACEXTABLE;

- **Display 5 records where launch sites begin with the string 'CCA'**
- **%sql** select * from 'spacextbl' where Launch_site like 'CCA%' limit 5

- **Display the total payload mass carried by boosters launched by NASA (CRS)**
- *#%sql select sum(payload_mass_kg) as payload_mass from 'SPACEXTBL'* **%sql** SELECT SUM(PAYLOAD_MASS__KG_) as "Total Payload KG)", Customer FROM 'SPACEXTBL' WHERE Customer = 'NASA (CRS)';

Sql contd..

- **Display average payload mass carried by booster version F9 v1.1**
- **%sql** select avg(PAYLOAD_MASS___KG_) as "Average mass kg", customer Booster_Version FROM 'SPACEXTBL' WHERE Booster_Version LIKE 'F9 v1.1%';

- **List the date when the first succesful landing outcome in ground pad was acheived.**
- **%sql** SELECT min(date) FROM 'SPACEXTBL' WHERE "Landing_outcome" = "Success (ground pad)"
- (%sql SELECT DISTINCT Booster_Version, Payload FROM SPACEXTBL WHERE "Landing _Outcome" = "Success (drone ship)" AND PAYLOAD_MASS___KG_ > 4000 AND PAYLOAD_MASS___KG_ < 6000;
- Github link :

- https://github.com/Joejokelly/Spacex-falcon9-updated/blob/master/4.%20jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Created folium map to mark launch sites, and markers, circles, lines to mark failure or success of launches for each site
- Created launch outcomes (failure = 0, success = 1)

- https://github.com/Joejokelly/Spacex-falcon9-updated/blob/master/6.Space-X%20Launch%20Sites%20Locations%20Analysis%20with%20Folium-Interactive%20Visual%20Analytics.ipynb

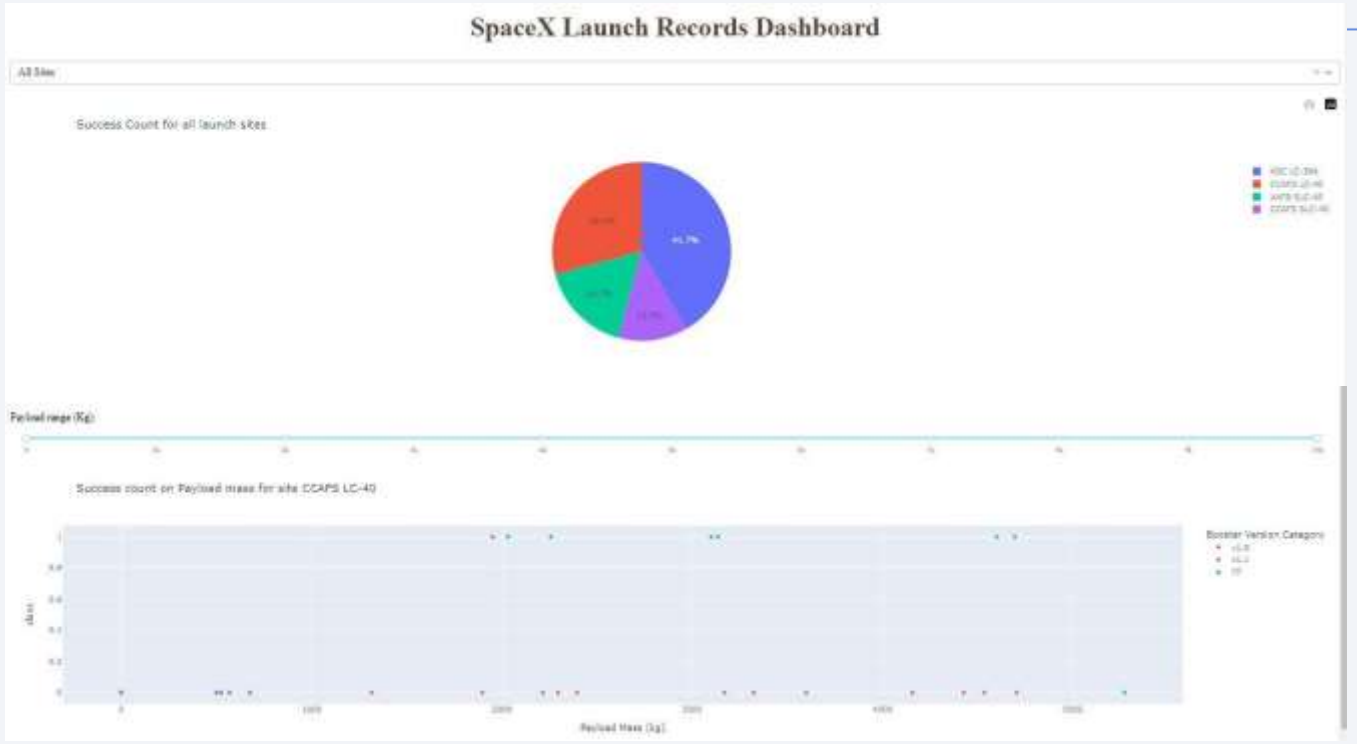
Build a Dashboard with Plotly Dash

- Build an interactive dashboard application with Plotly dash ;
- Adding Launch Site Drop-down Input component
- Adding call-back function success-pie-chart
- Adding range slider to selected payload
- Adding call back function success-payload-scaler –chart

- Github link :
- https://github.com/Joejokelly/Spacex-falcon9-updated/blob/master/7.%20%20Build%20Interactive%20Dashboard%20-%20spacex_dash_app.py

Lauch Record Dashboard

SpaceX Dash App



Predictive Analysis (Classification)

- Summary of how model was built, evaluated and improved :
  - After initial processing, data was loaded to the model using Pandas Data frame, created Numpy Array, then used StandardScaler function, to standardize the data.
  - Split the resulting data into training data and test data, train_test_Split from sklearn.
- In order to find the best Machine Language
  - Created GridsearchCV object, after fitting the training data, into GridsearchCV, displayed the best parameter using data attributes.
  - Used method score to calculate accuracy of each model and plotted confusion matrix for teach test and predicted outcomes.

Github link :

- https://github.com/Joejokelly/Spacex-falcon9-updated/blob/master/6.Space-X%20Launch%20Sites%20Locations%20Analysis%20with%20Folium-Interactive%20Visual%20Analytics.ipynb

Predictive Analysis (Classification, contd)

- The table below shows the test accuracy score :

Out[68]:

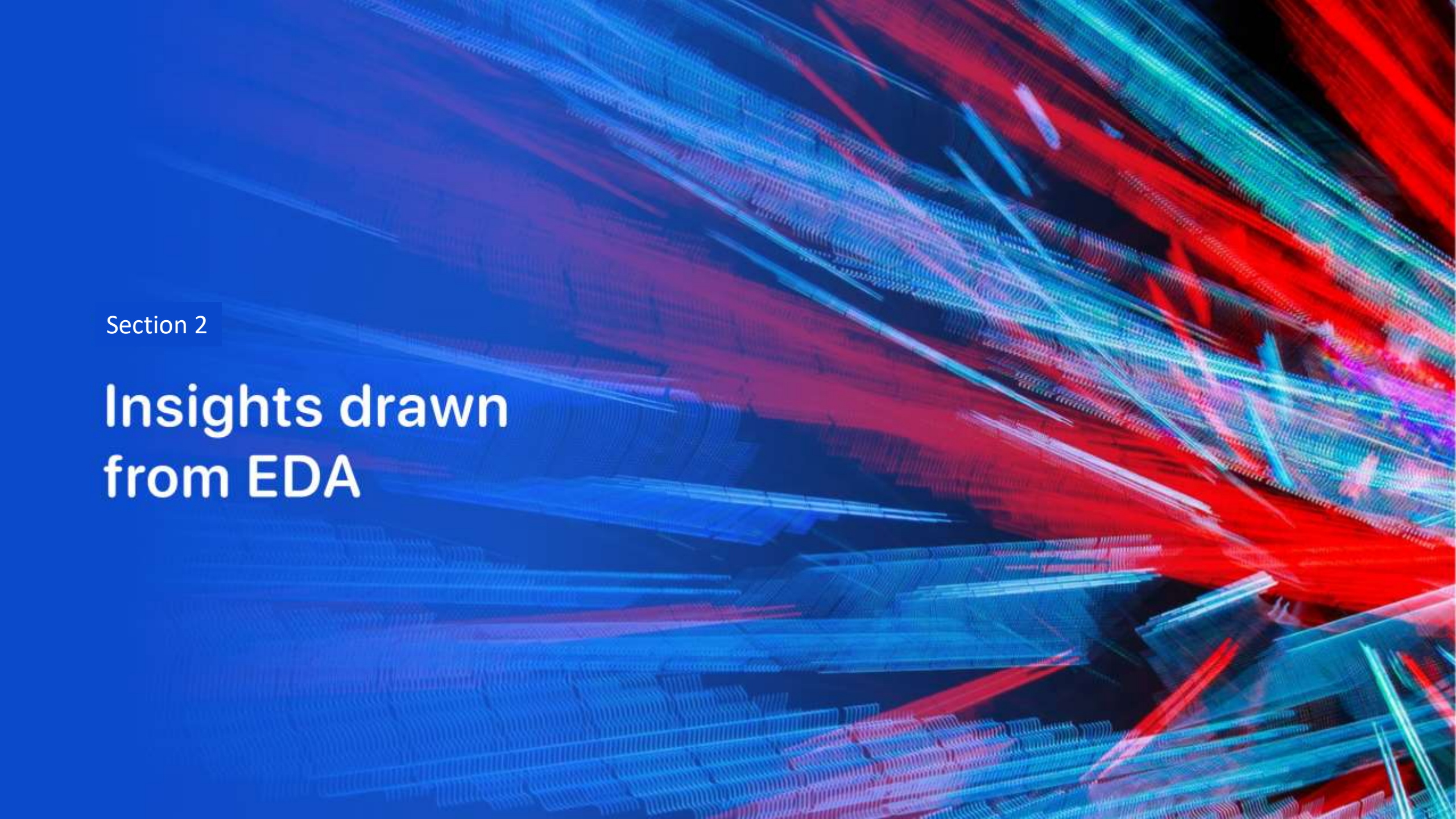
	0
Method	Test Data Accuracy
Logistic_Reg	0.833333
SVM	0.833333
Decision Tree	0.833333
KNN	0.833333

- Github link :

https://github.com/cgatama/SpaceX-Falcon-9-1st-stage-Success-Landing-Prediction/blob/main/8.%20SpaceX%20Machine%20Learning%20Prediction.ipynbhttps://github.com/cgatama/SpaceX-Falcon-9-1st-stage-Success-Landing-Prediction/blob/main/8.%20SpaceX%20Machine%20Learning%20Prediction.ipynb

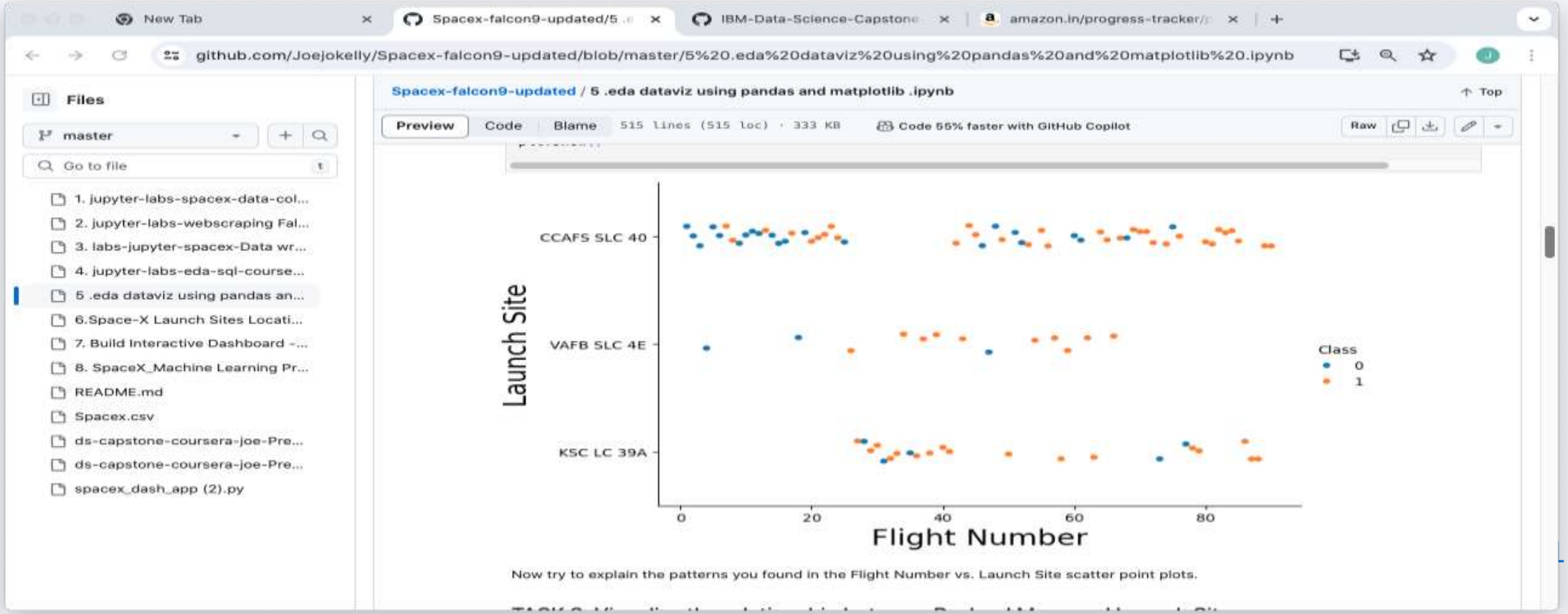
Results

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


Section 2
Insights drawn
from EDA

Flight Number vs. Launch Site

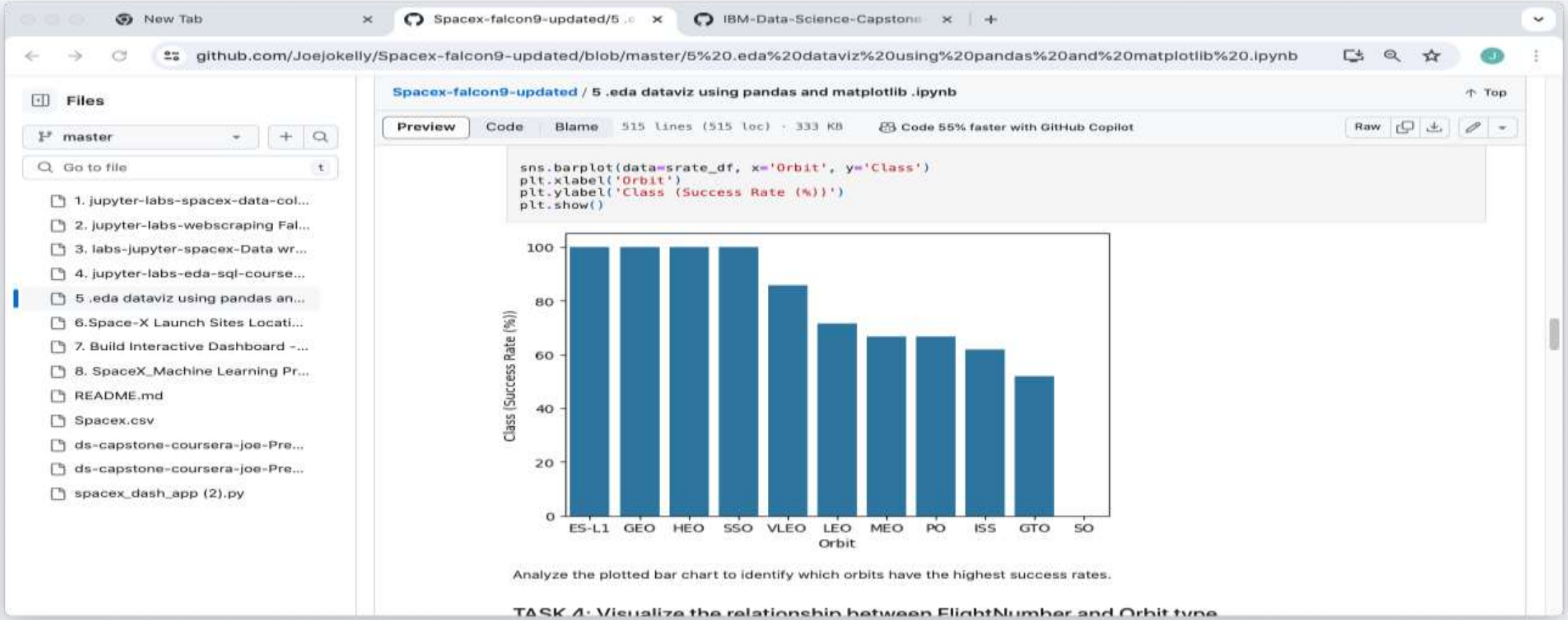
- scatter plot of Flight Number vs. Launch Site



Payload vs. Launch Site

Payload vs Launch site scatter plot, there are no rocket lauches for heavy payload > 10000

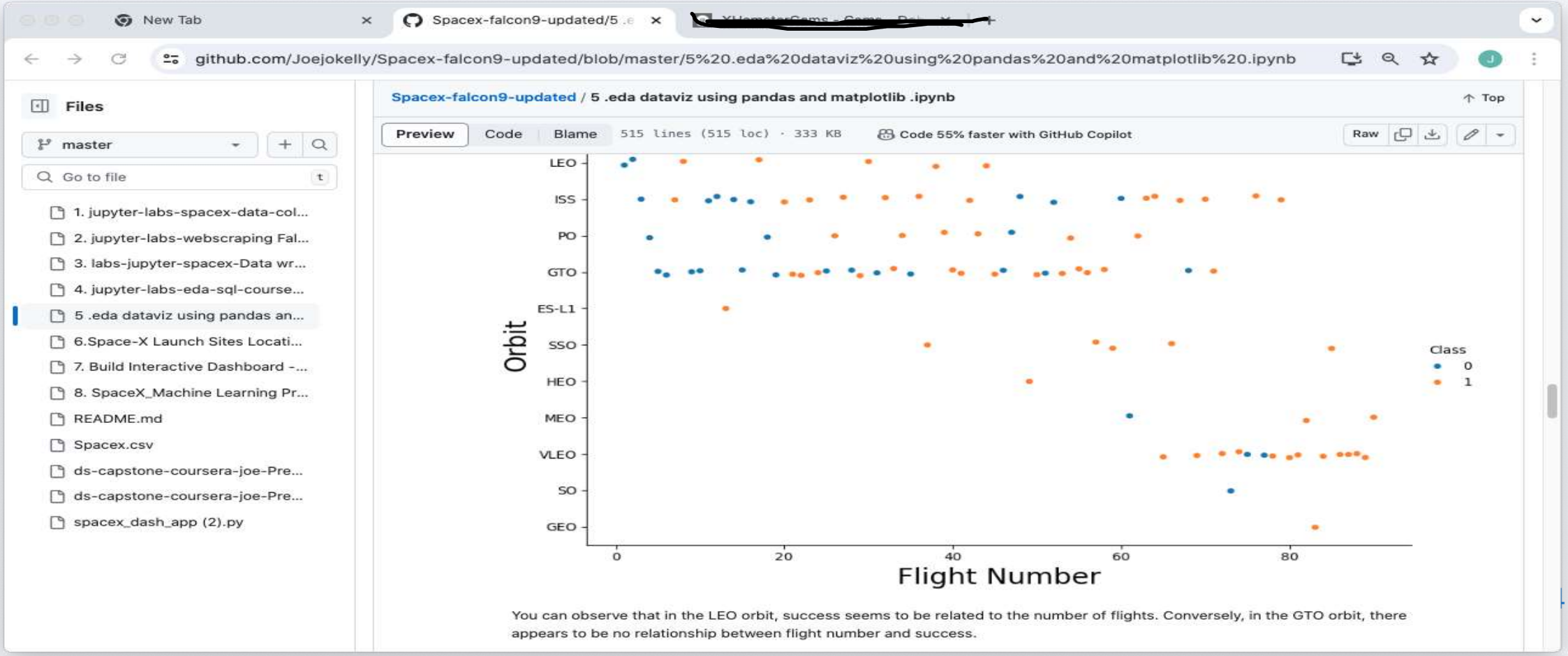
Success Rate vs. Orbit Type



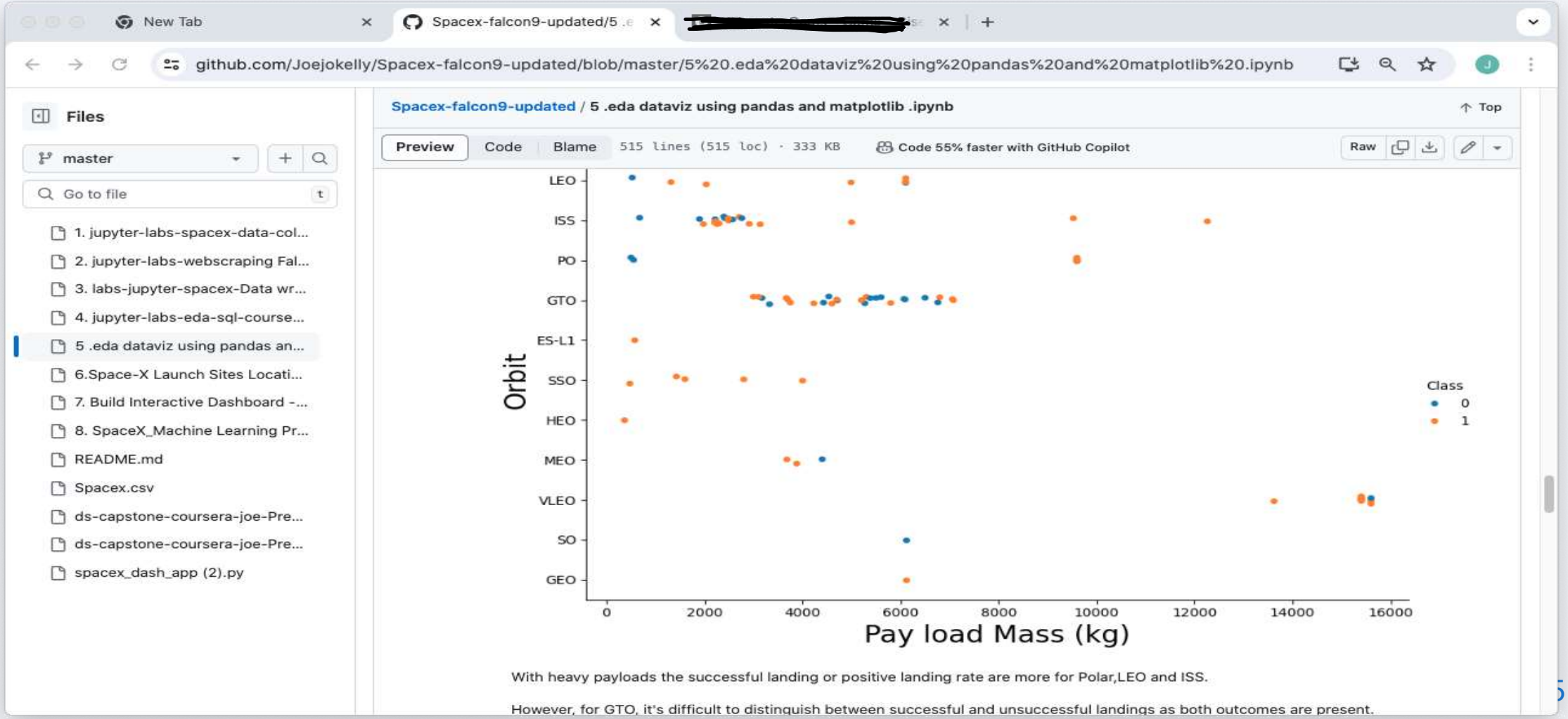
Orbits ES-L1, Geo, Heo, seo, have a higher rate of success, 100%m while GTO has about 50%

Flight Number vs. Orbit Type

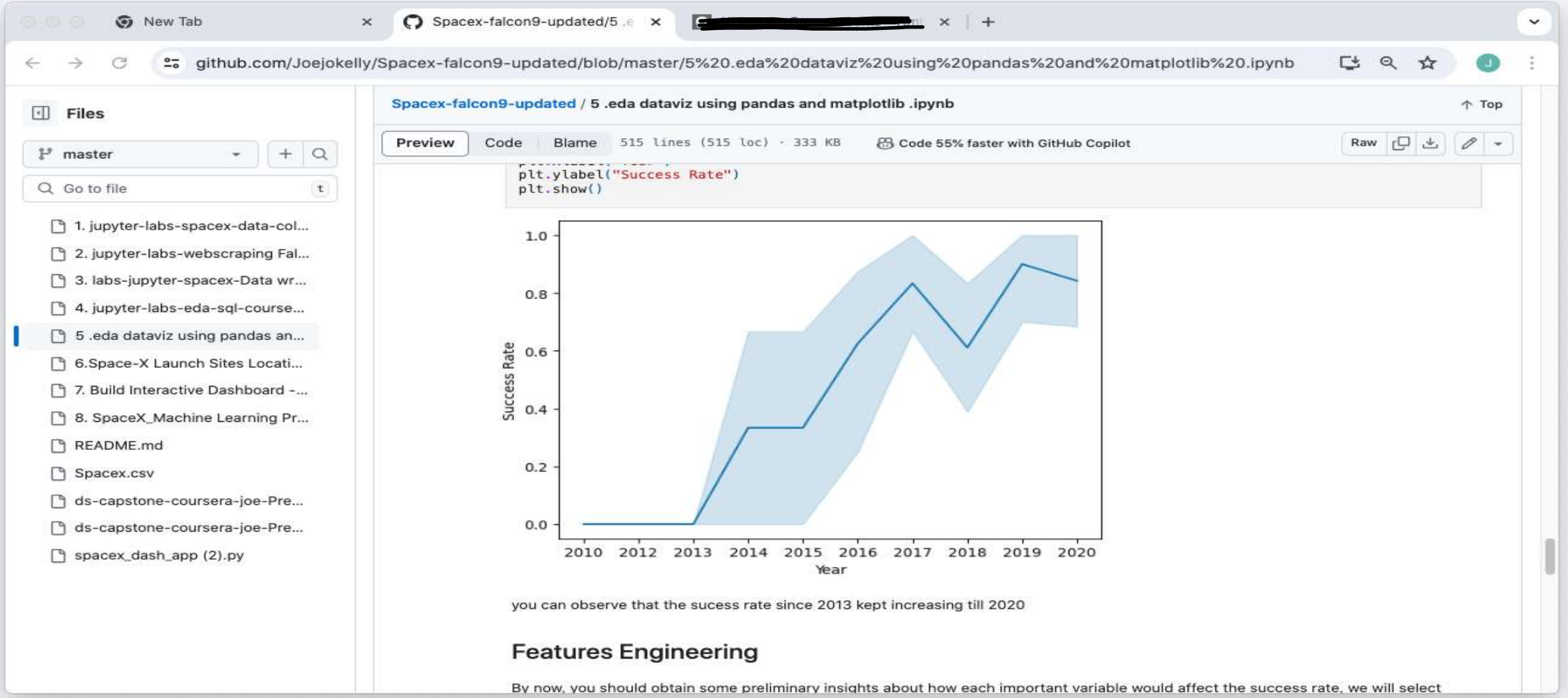
- Show a scatter point o Flight number vs. Orbit type



Payload vs. Orbit Type

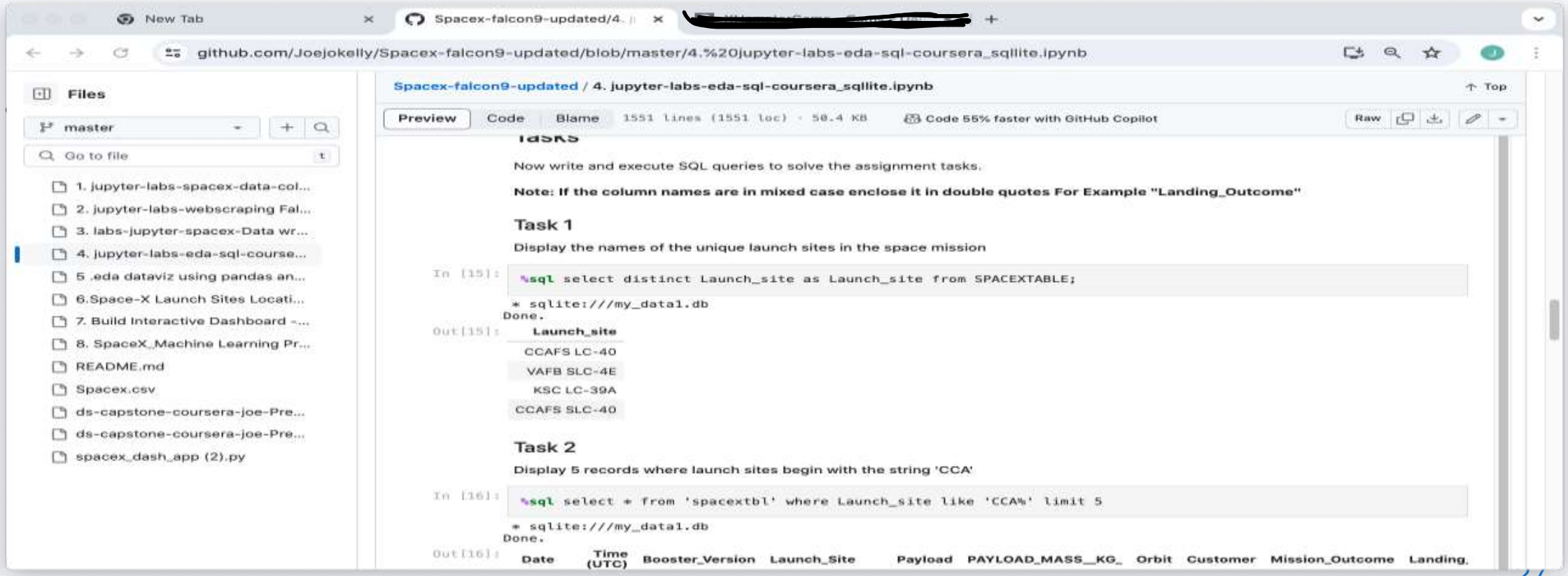


Launch Success Yearly Trend



All Launch Site Names

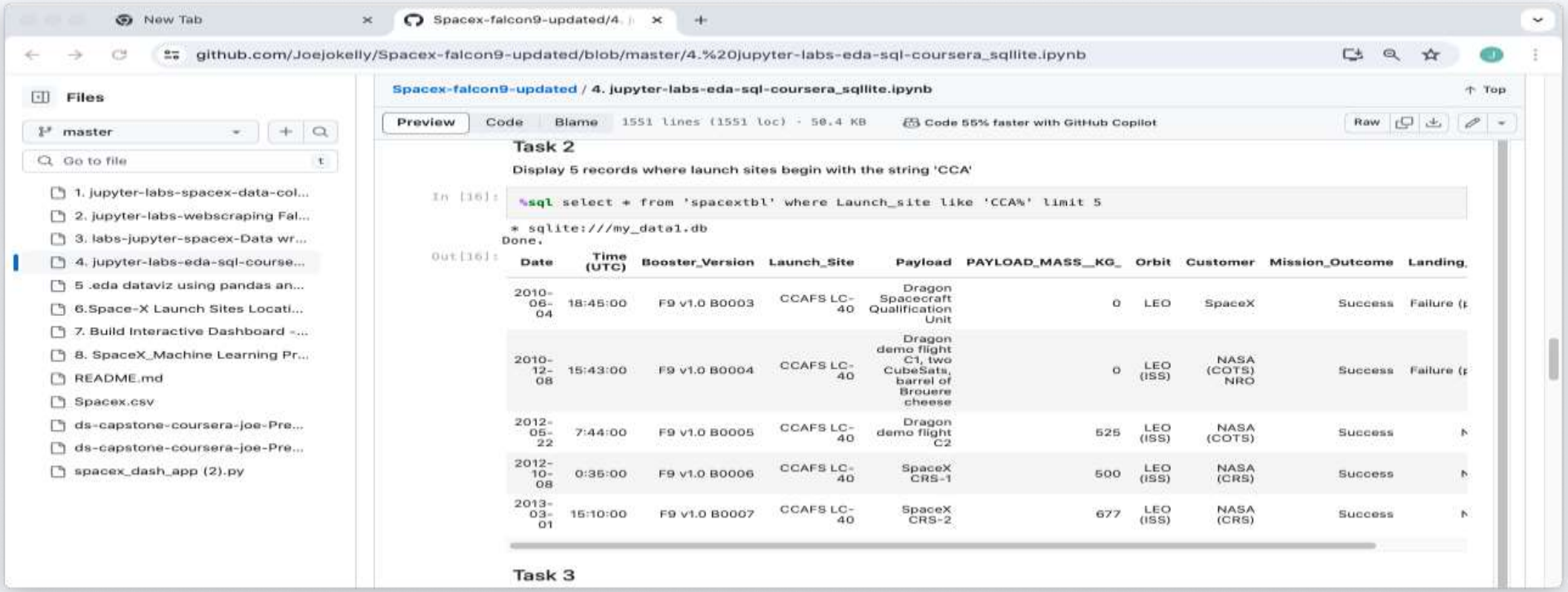
- Find the names of the unique launch sites



%sql select distinct Launch_site as Launch_site from SPACEXTABLE;

Launch Site Names Begin with 'CCA'

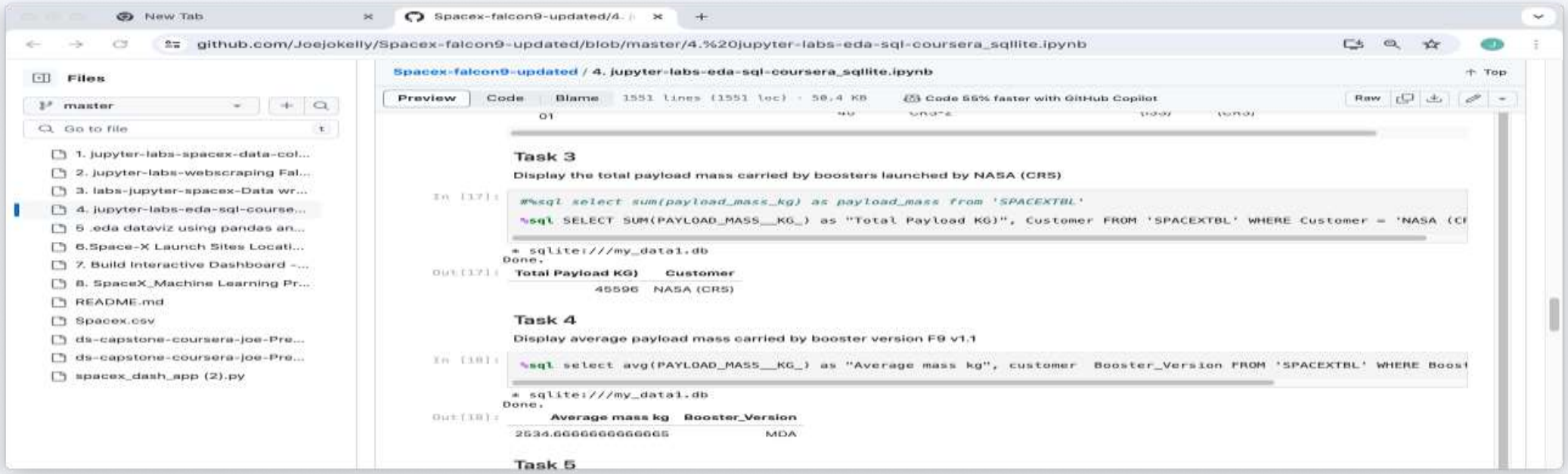
- Find 5 records where launch sites begin with `CCA`



Used wild cars in the above query to display records with site begin with ‘CCA’

Total Payload Mass

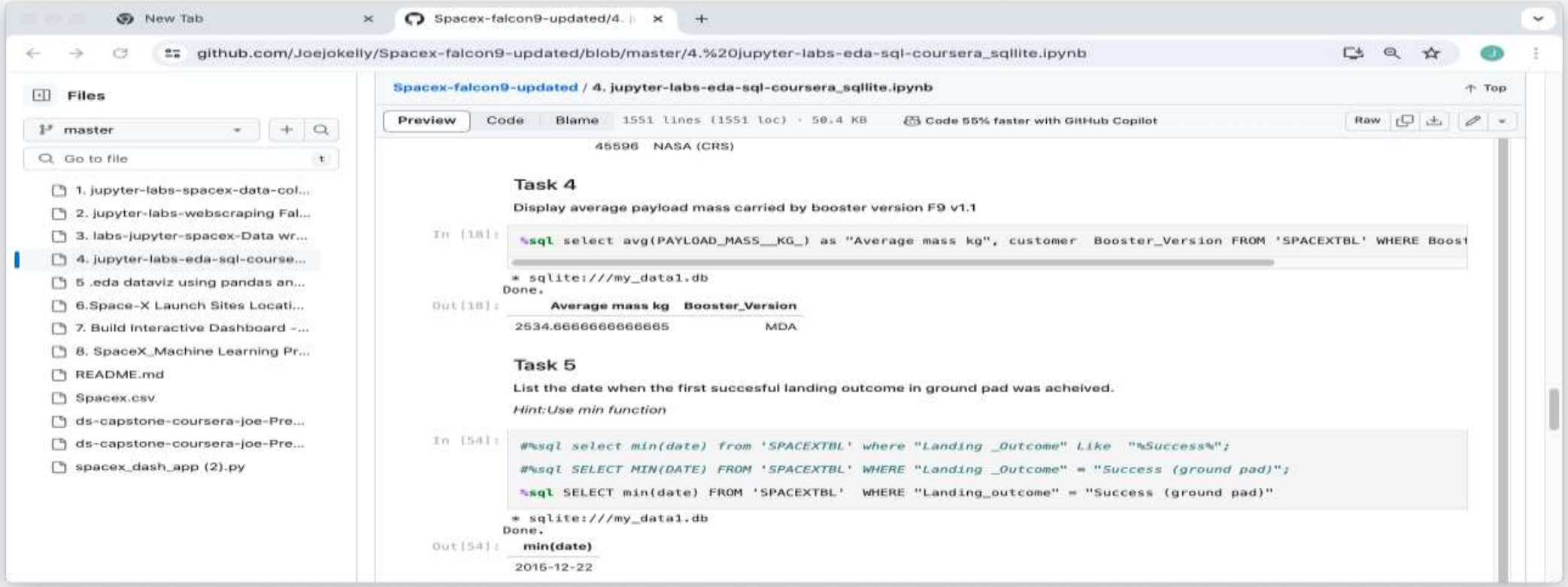
- Calculate the total payload carried by boosters from NASA



Used the sum function to display the total payload (kg)

Average Payload Mass by F9 v1.1

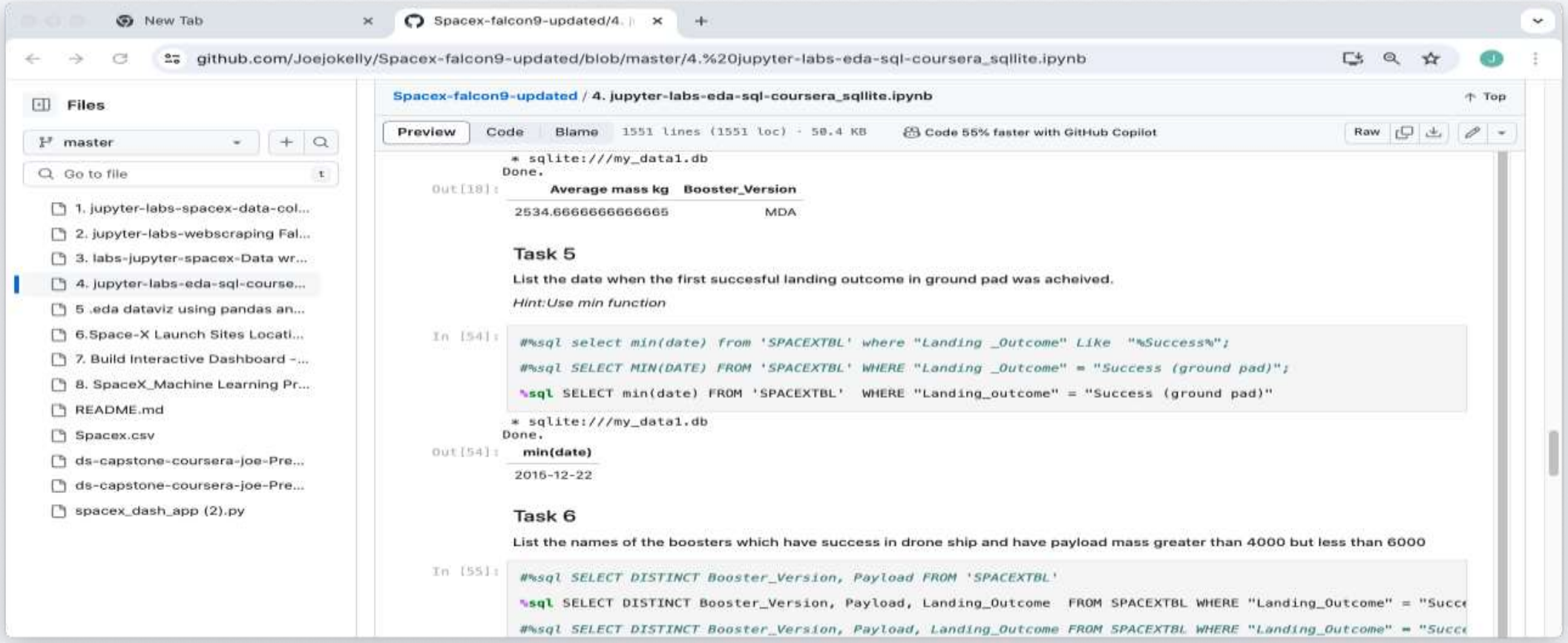
- Calculate the average payload mass carried by booster version F9 v1.1



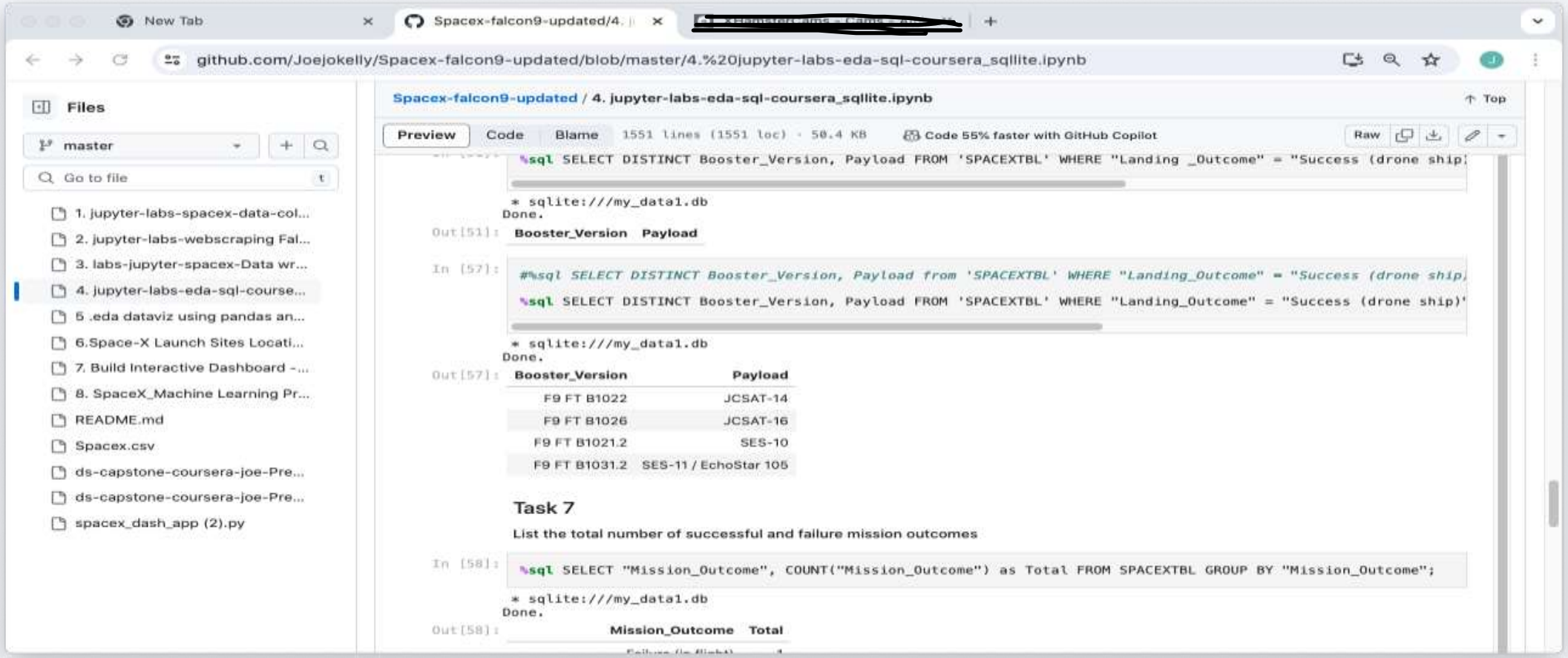
Used the avg() function to display average payload booster version

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

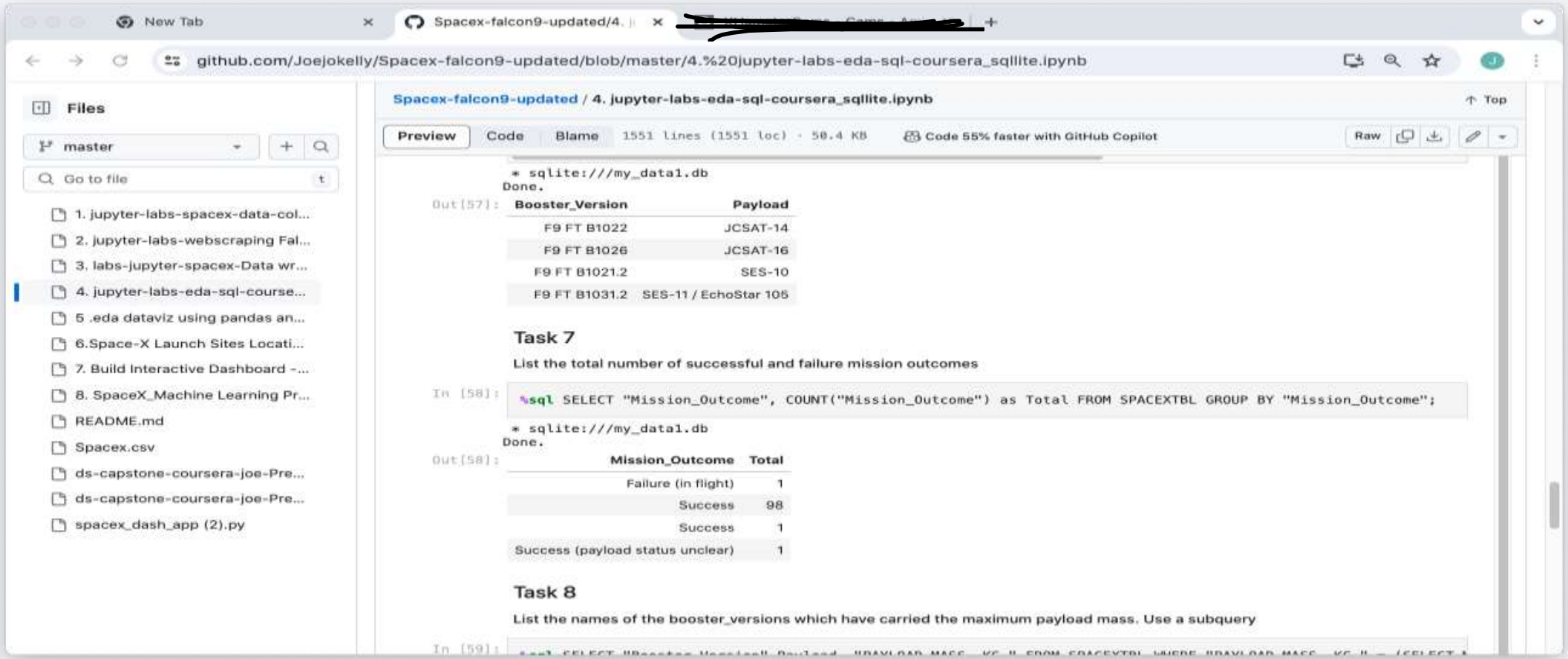


Successful Drone Ship Landing with Payload between 4000 and 6000

- 

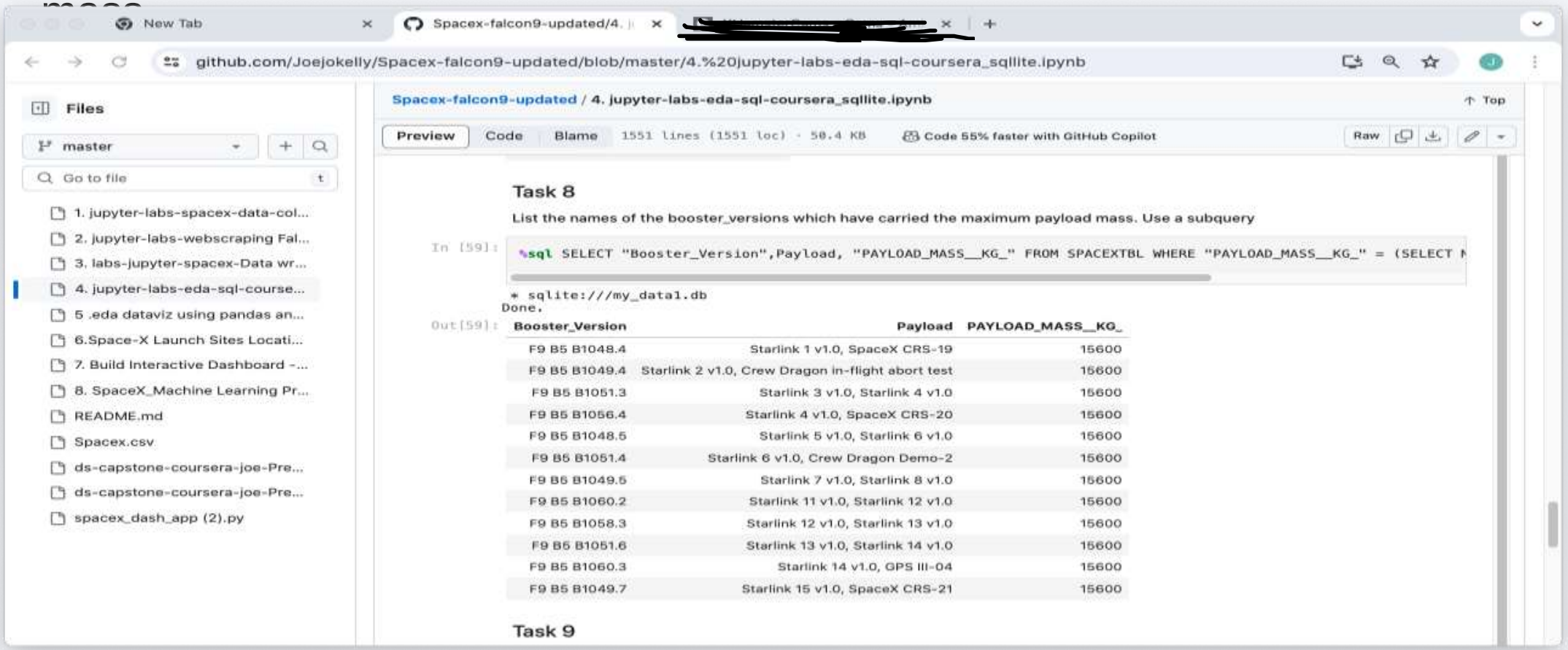
Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes



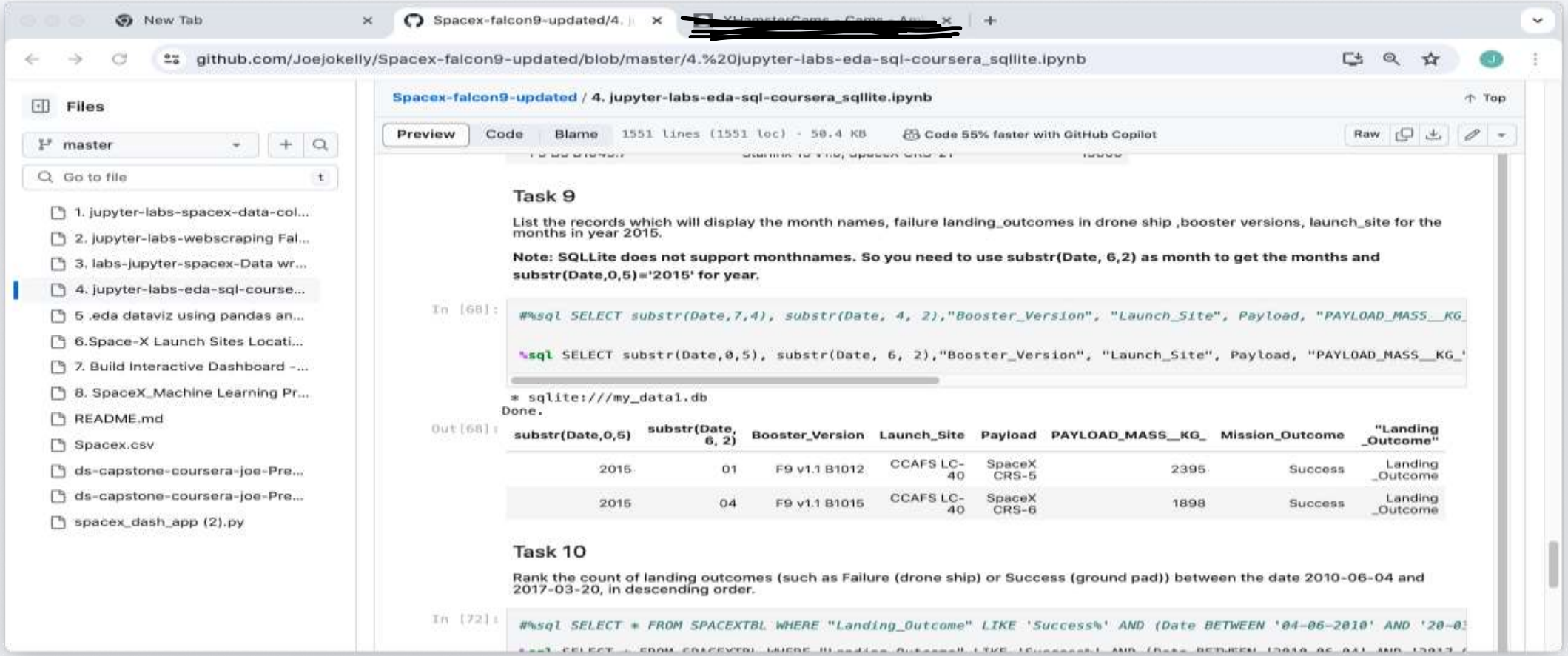
Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass



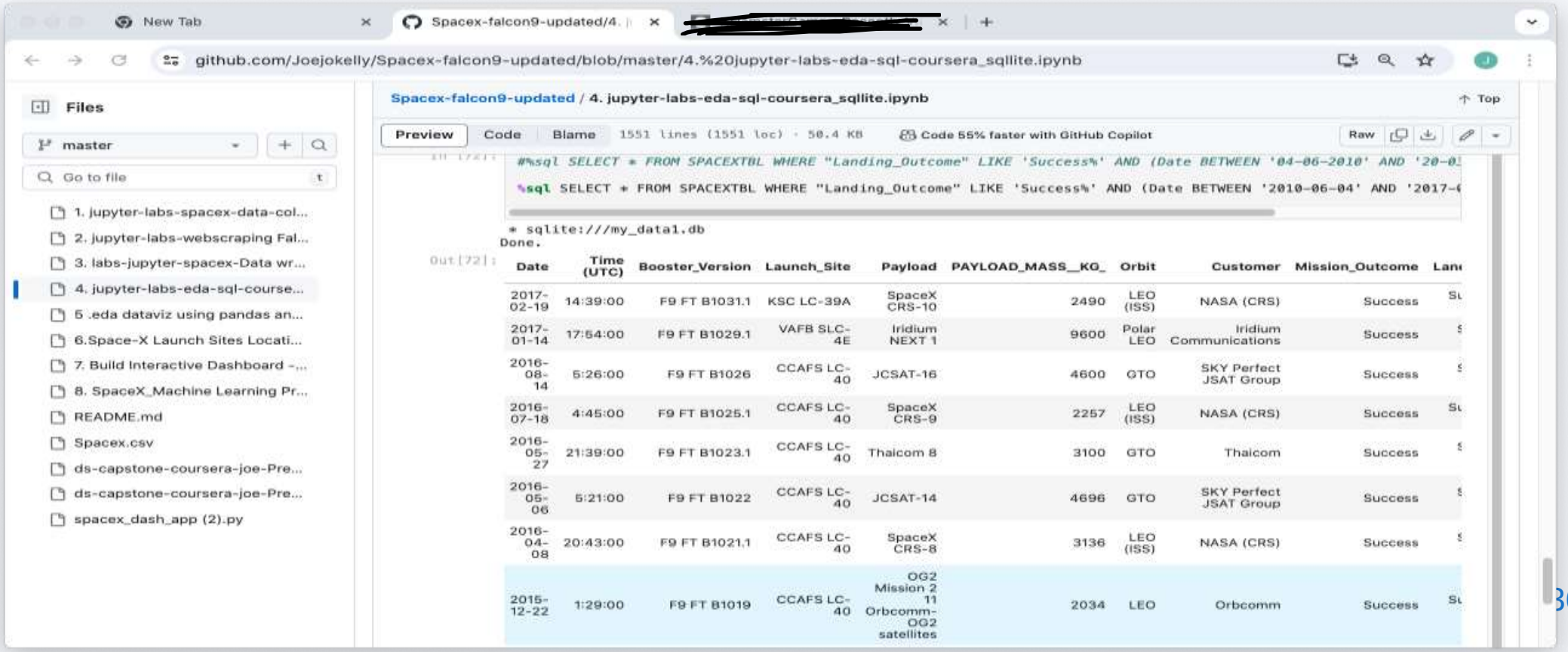
2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015



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

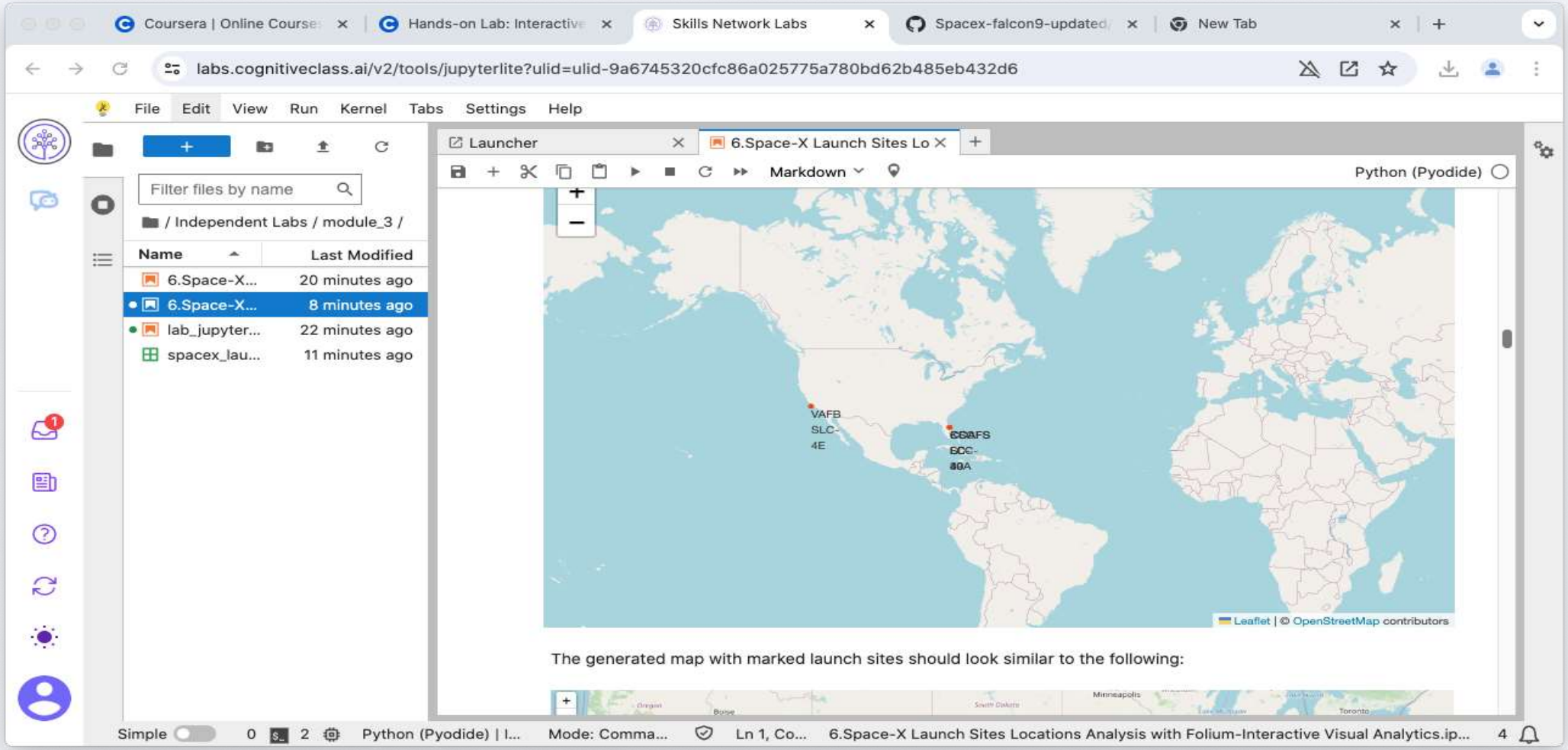
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order



Section 3

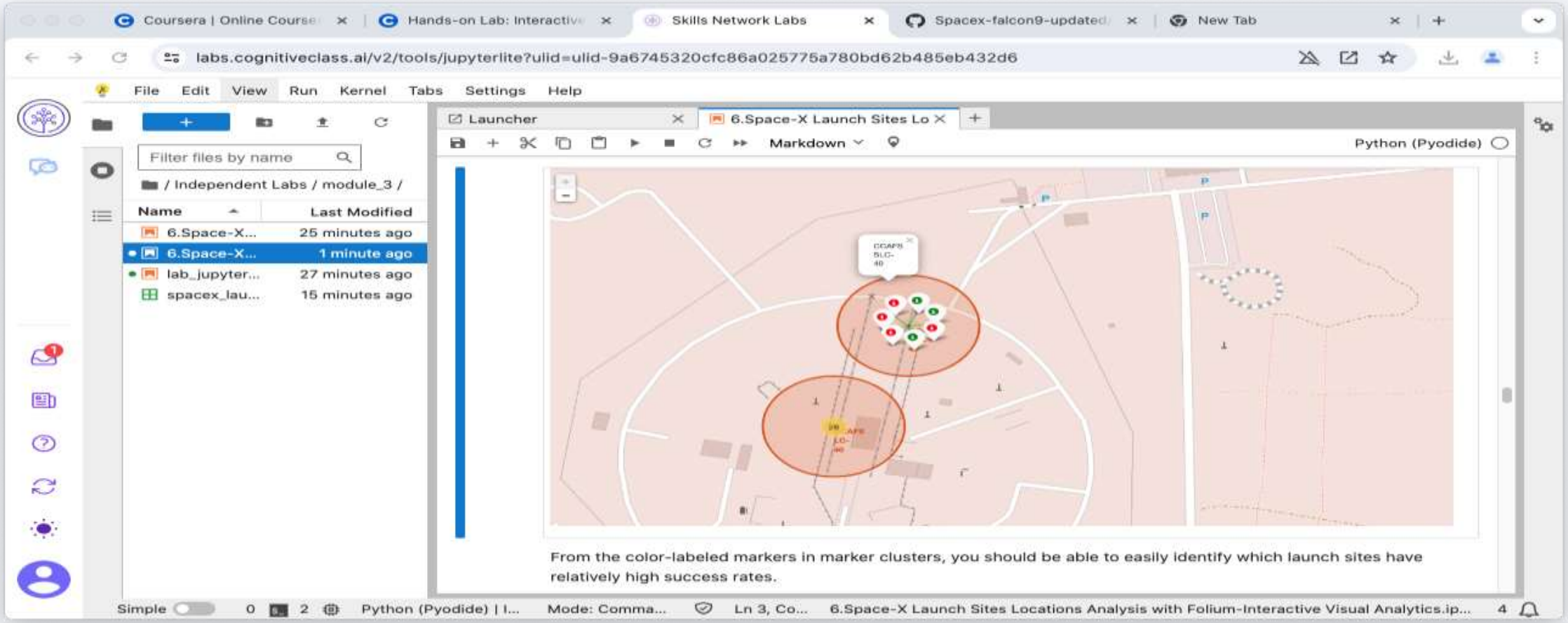
Launch Sites Proximities Analysis

launch site on the global map



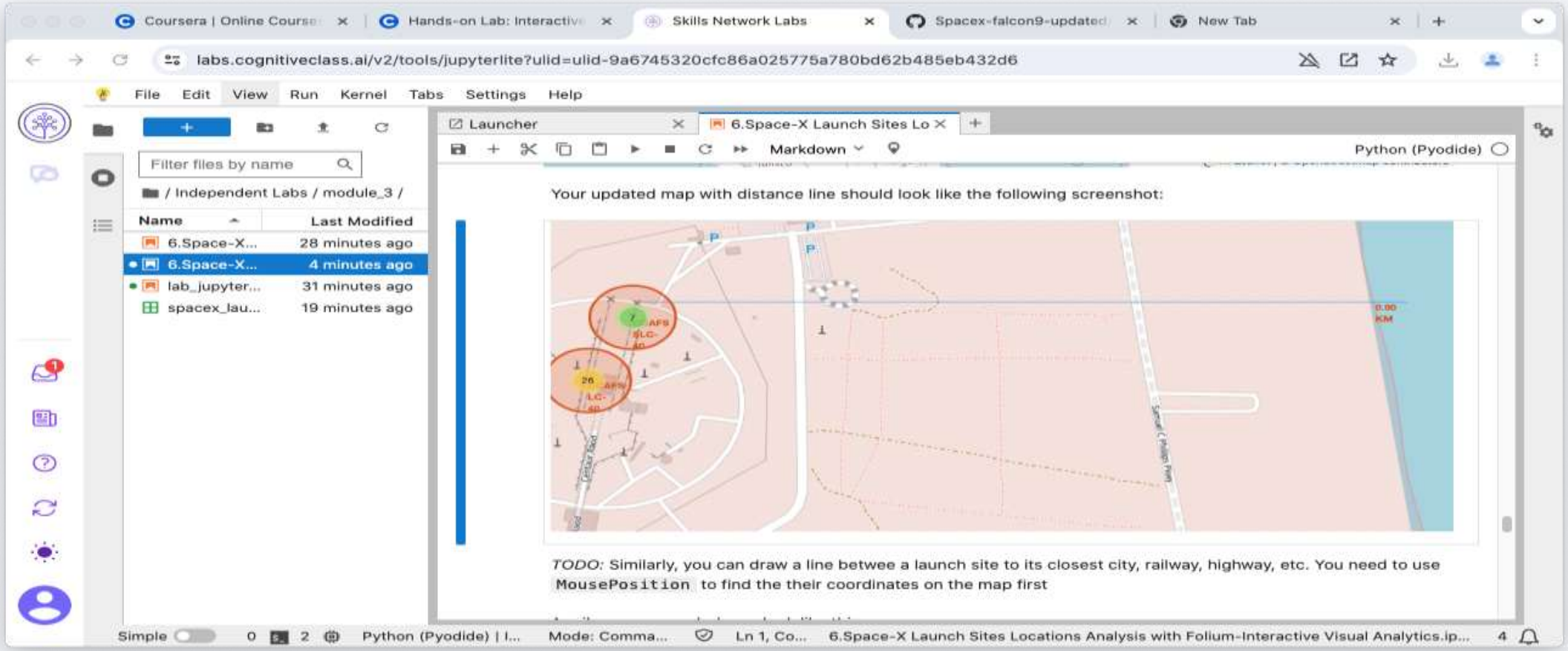
color-labeled launch outcome

- Explore the folium map and make a proper screenshot to show the color-labeled launch outcomes on the map

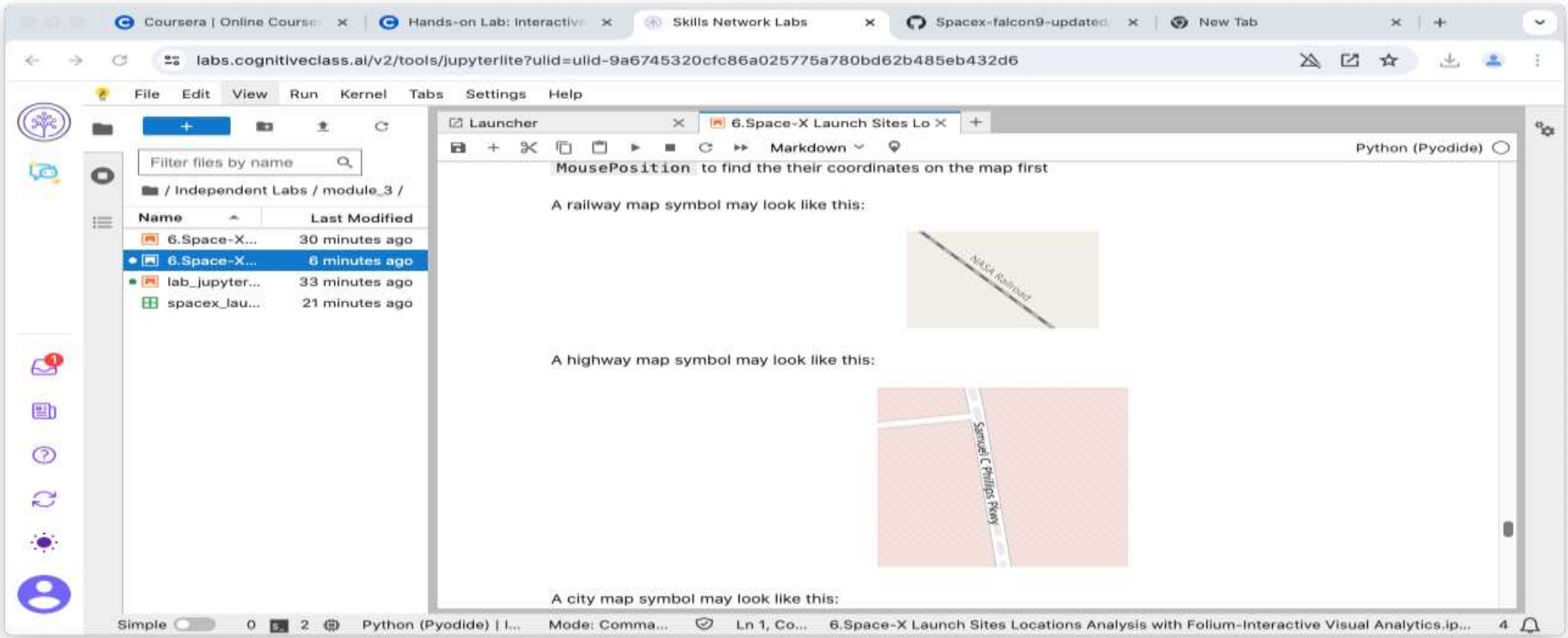


Proximity of Launch site to railway, highway

- Replace <Folium map screenshot 3> title with an appropriate title

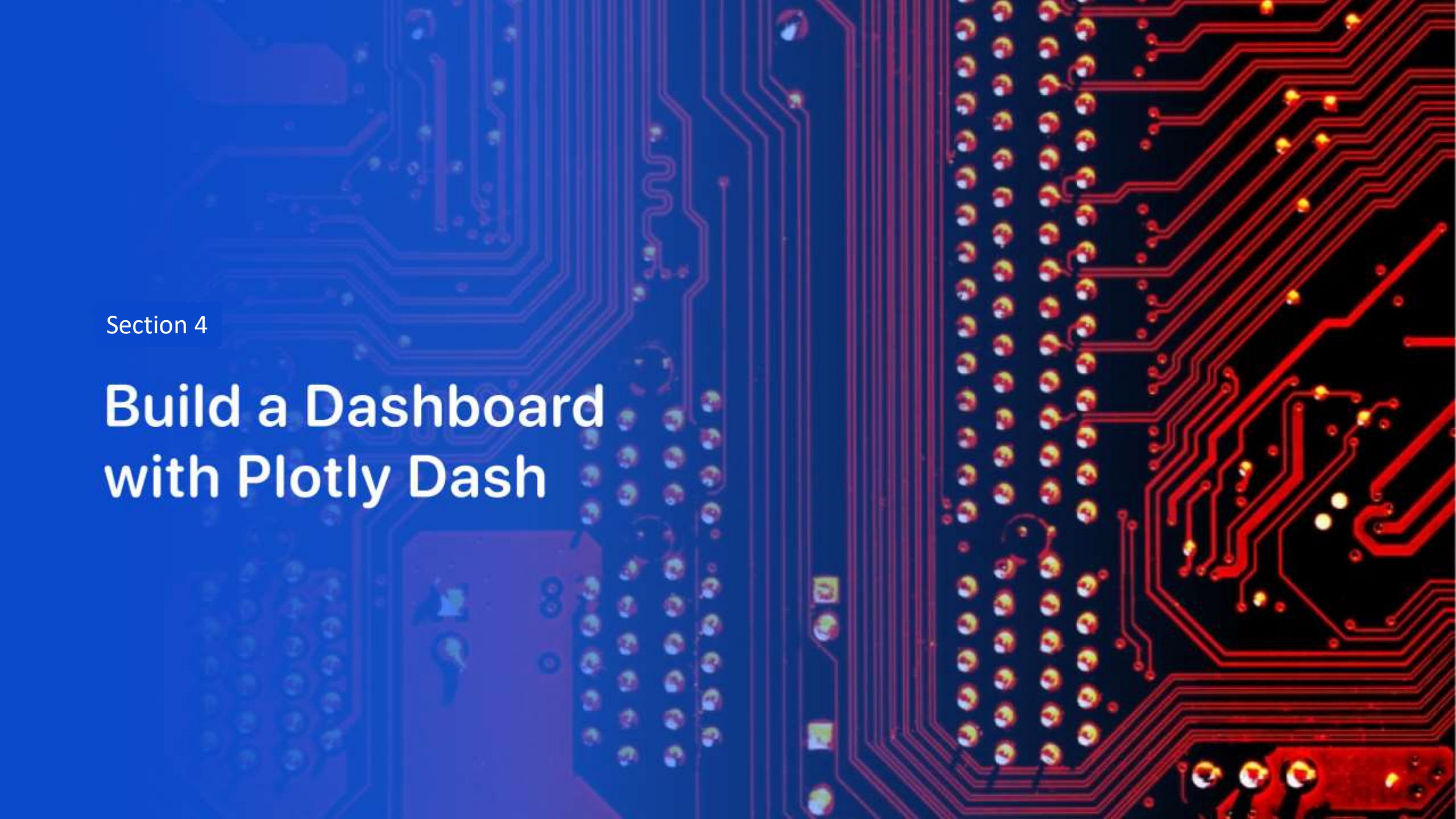


Proximity contd…



Section 4

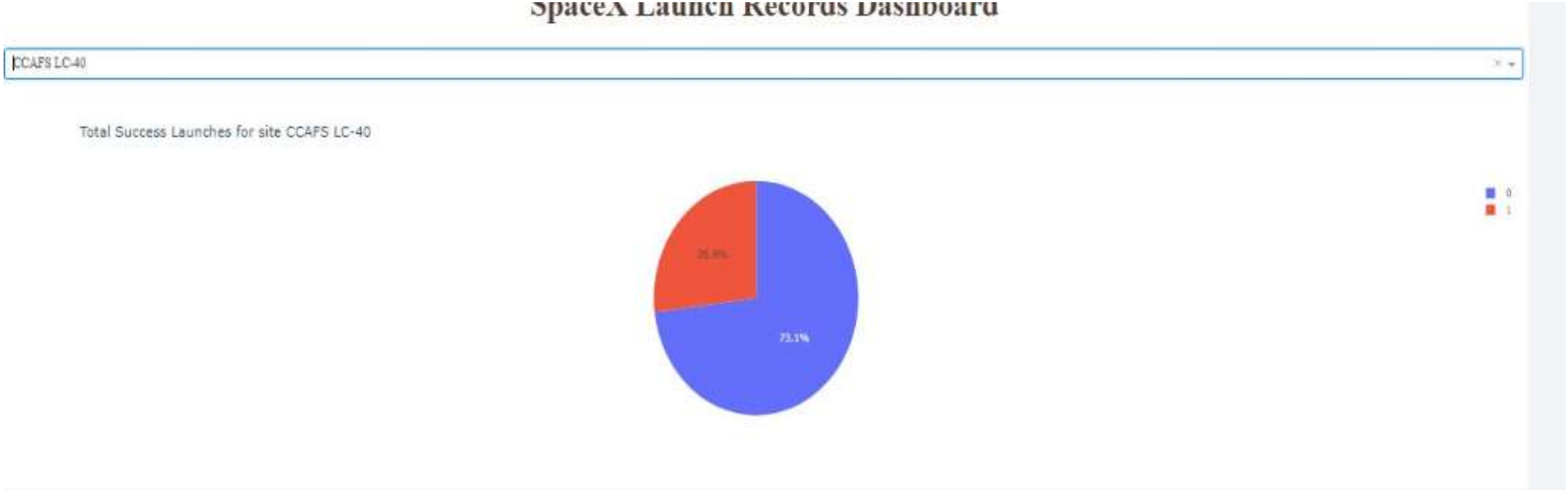
Build a Dashboard with Plotly Dash



Pie chart for launch Success

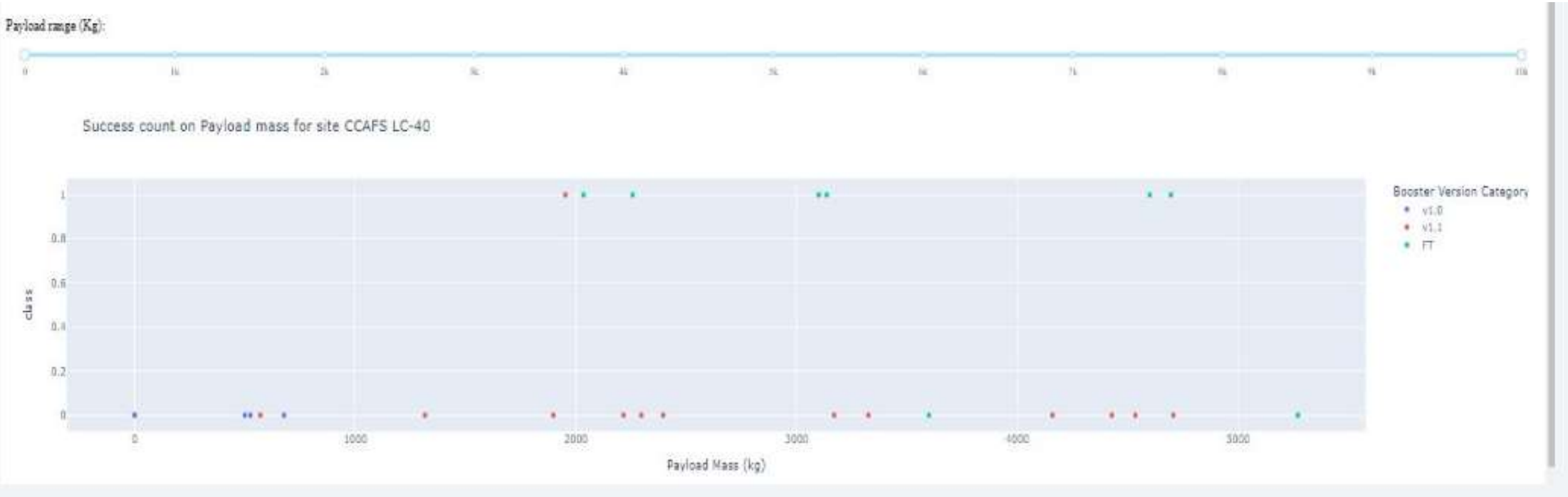
Launch site KSC LC-39A highest success rate 42%
CCAFS LC-40 at 29%, CCAFS LC-40 at 29%

Lauch site success rate



- Launch site CCAFS LC-40 had the 2nd highest success ratio of 73% success against 27% failed launches

Payload vs launch site



For launch ccfa-lc 40, the booster version has the largest success rage payload mass > 2000

Section 5

Predictive Analysis (Classification)

Classification Accuracy

- All models perform the same, accuracy of 0.83333 on test Data

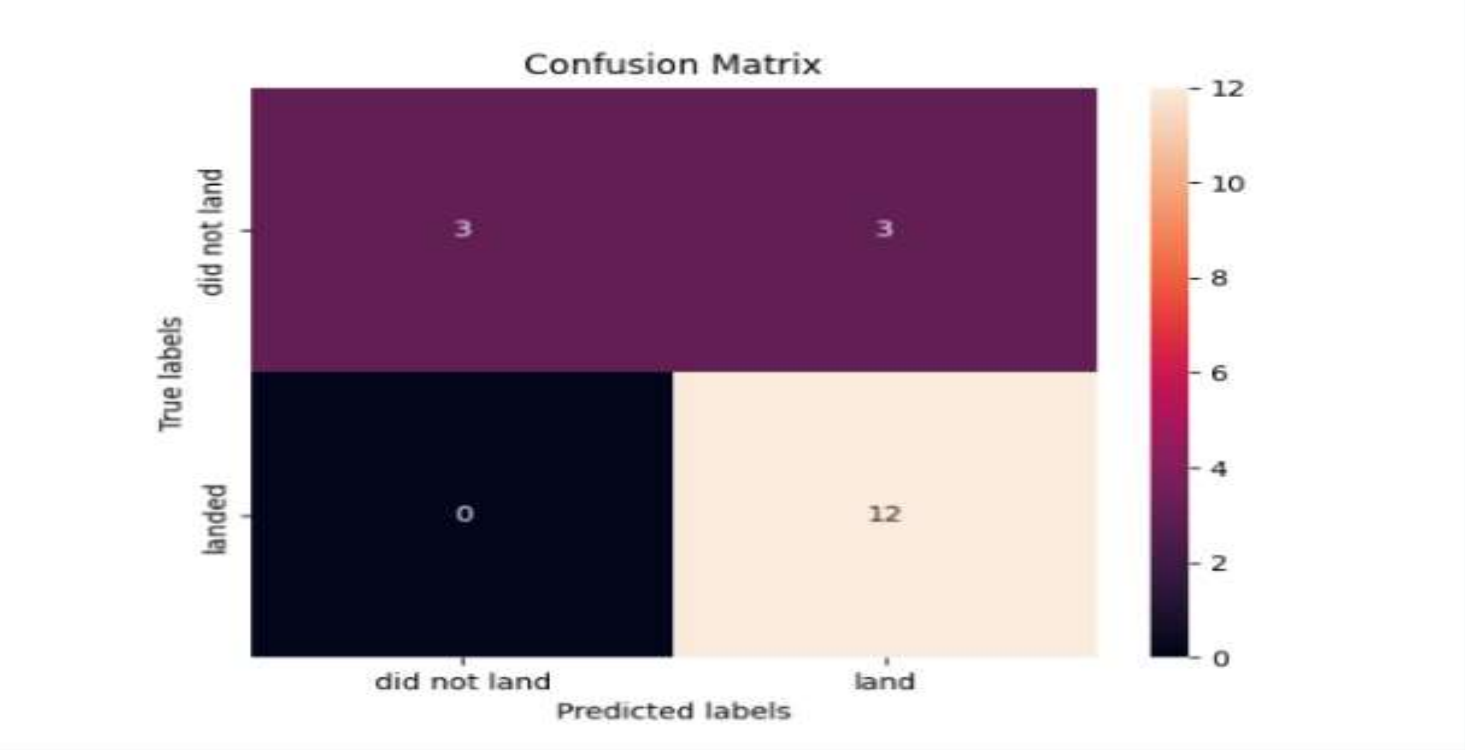
Out[68]:

	0
Method	Test Data Accuracy
Logistic_Reg	0.833333
SVM	0.833333
Decision Tree	0.833333
KNN	0.833333

All the methods perform equally on the test data: i.e. They all have the same accuracy of 0.833333 on the test Data

Confusion Matrix

- All models have the same confusion matrix



Conclusions

- Success rates are different for different launch sites.

  CCAFS LC-40 has success rate of 60%

  KSC LC-39A and VAFB SLC 4E has a success rate of 77%

- The flight number increases so does the success rate,
- Payload vs Lauch site Scatter plot, we find there are no rockets lauched s
- Orbits ES-L1, GEO, HEO & SSO have the highest success rates at 100%, with SO orbit having the lowest success rate at ~50%. Orbit SO has 0% success rate.
- 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

Conclusion contd..

s

- Heavy payloads the successful landing is higher for Polar, LEO and ISS. GTO has both positive and negative landings
- Finally the success rate since 2013 kept increasing till 2020

Thank you!