IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Abraham Jhared Flores Azcona
January 2nd, 2026

Outline

Executive Summary

- Our result findings suggest that *the success of SpaceX as company for commercially reliable rocket launches is to be continued in the foreseeable future.*
- The data analysis pipeline consisted on data extraction from publicly available information, Exploratory Data Analysis using SQL to facilitate the queries, and different Machine Learning models that converge on our interpretation that the success trend will continue.

Introduction

- SpaceX is a rocket launch company
- Many have been quite skeptical about its future since the start of the company
- The historical data analyzed suggests that **success is the trend of this company**
- ***What is the historical launch success rate for the company?***
- ***Which launch sites are there in the US?***
- ***Which boosters offer the best success rate?***
- ***How much cargo weight can we employ to maintain the best-boosters success rate?***

Section 1

Methodology

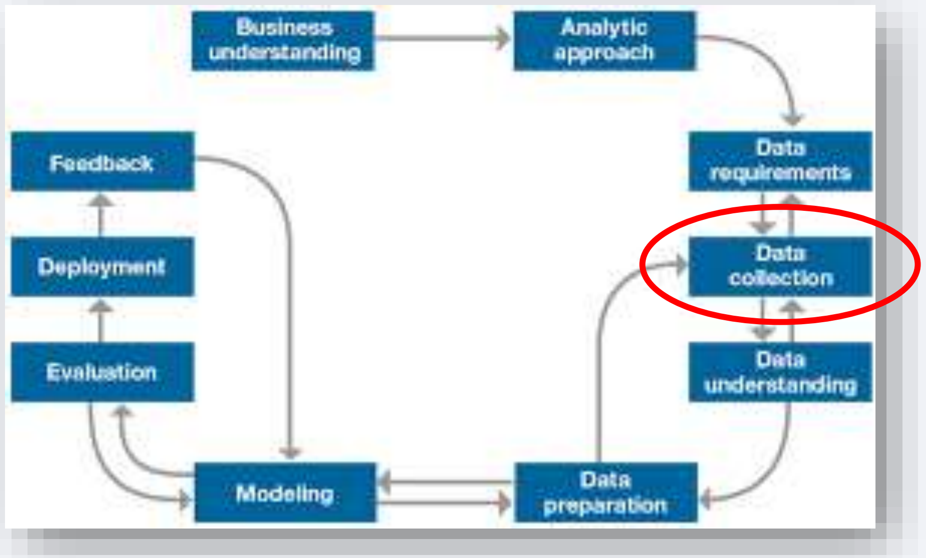
Methodology

Executive Summary

- Data collection methodology:
  - Webscraped data from the Wikipedia article to then port in into an CSV file
- Perform data wrangling
  - Used pandas & SQL for easy manipulation and cleanup based in their APIs
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

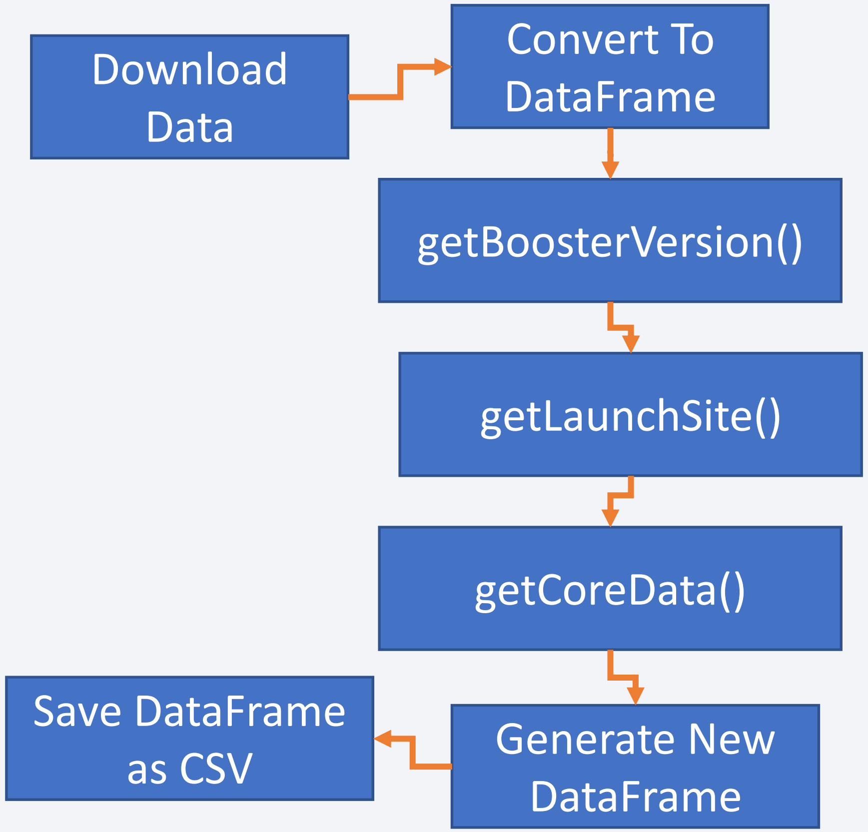
Data Collection

- Webscraped the data from the Wikipedia entry, load the data into a dataframe to clean it
- Afterwards, we made several SQL queries as Exploratory Data Analysis to gain insight on the landing sites
- Using geospatial data, we mark on map the launching sites for reference
- Then, we detect which has the highest success rate, and execute ML models to predict the future success rate



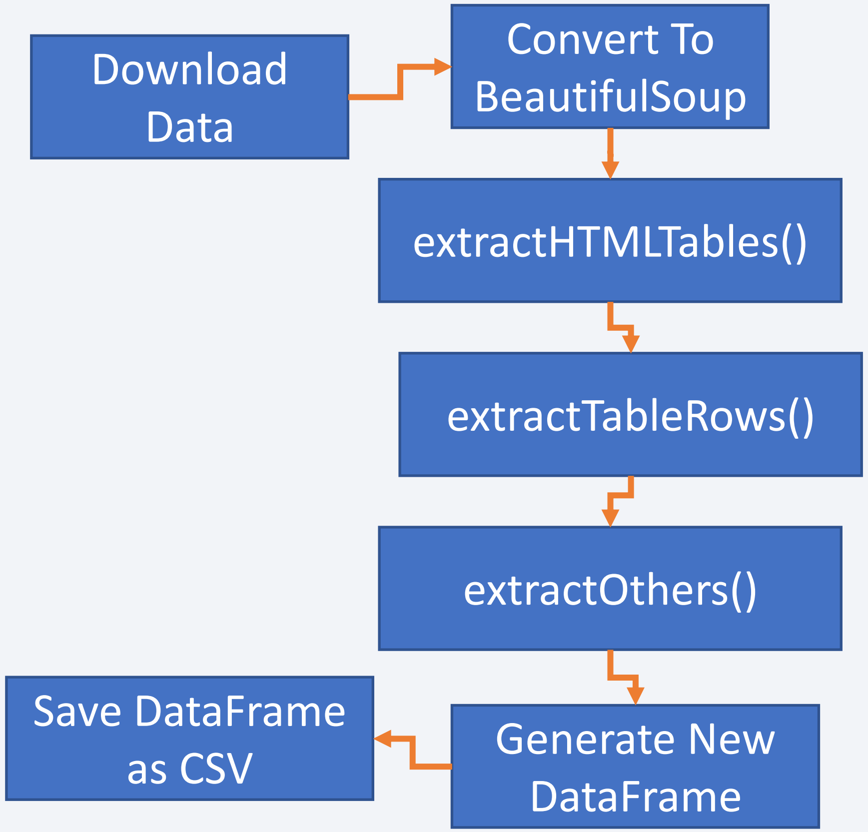
Data Collection – SpaceX API

- Call the SpaceX API and get the required data from the expected endpoint for further analysis
- GitHub URL: https://github.com/NotsoJharedtrollOx17/Module10_AppliedDataScienceCapstone/blob/main/00_DataCollection.ipynb



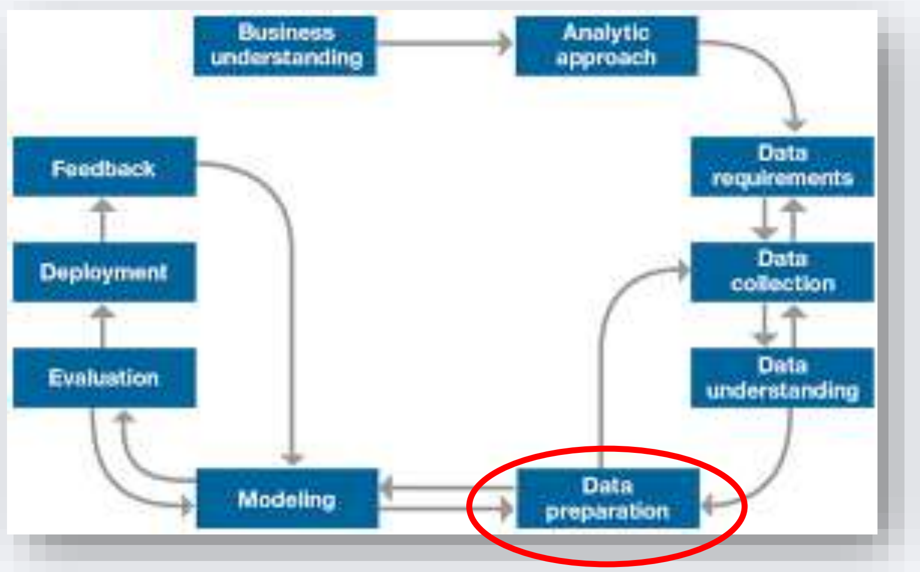
Data Collection - Scraping

- Basically, read the Wikipedia entry of launches and scrape the HTML table found inside.
- GitHub URL: https://github.com/NotsoJharedtrollOx17/Module10_AppliedDataScienceCapstone/blob/main/01_HandsOnLab_Webscraping.ipynb



Data Wrangling

- By using pandas (a Data Science Python library), we where able to clean up and inject missing values per common procedure
- GitHub URL: https://github.com/NotsoJharedtrollOx17/Module10_AppliedDataScienceCapstone/blob/main/02_HandsOnLab_DataWrangling.ipynb



EDA with Data Visualization

- We plotted scatter plots, bar charts & histograms to get a glimpse into the payloads, launch sites & successes to expect a sensible model behavior
- GitHub URL: https://github.com/NotsoJharedtrollOx17/Module10_AppliedDataScienceCapstone/blob/main/04_HandsOnLab_EDAVisualization.ipynb

EDA with SQL

- In brief, the main queries made where:
  - ❑ Data counts
  - ❑ Range of values per successes
  - ❑ Locations of successful launches
  - ❑ Payloads per booster & location
  - ❑ Etc.
- GitHub URL: https://github.com/NotsoJharedtrollOx17/Module10_AppliedDataScienceCapstone/blob/main/03_HandsOnLab_EDASQL.ipynb

Build an Interactive Map with Folium

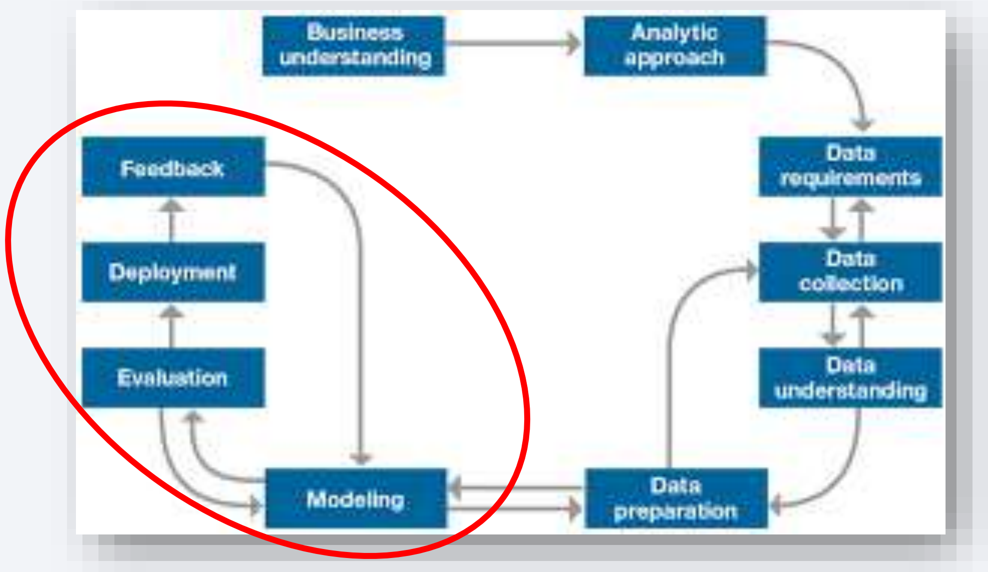
- We added the launch site markers, the labels of the sites, and the over counts as different regions in the map object
- It is imperative to understand *where* the launches have taken place to assess which areas of the world may prevail in extending the successes of the company
- GitHub URL: https://github.com/NotsoJharedtrollOx17/Module10_AppliedDataScienceCapstone/blob/main/05_InteractiveVisual.ipynb

Build a Dashboard with Plotly Dash

- We added a pie chart and a scatterplot of the overall successes within site, and the success/failure launch result within site
- It is necessary to display a brief glimpse of the successes per site, which can help executives to gain insight clearly
- GitHub URL: https://github.com/NotsoJharedtrollOx17/Module10_AppliedDataScienceCapstone/blob/main/06_HandsOnLab_InteractiveDashBoardScript.py.txt

Predictive Analysis (Classification)

- Evaluated 4 models with GridSearchCV: *i) LogisticRegression, ii) SVC, iii) DecisionTreeClassifier & iv) KNeighborsClassifier*
- *With the given data, we preprocess the dataset into train/test splits and then start training each of the variants with GridSearchCV to find the best model overall that estimates the launch success*
- GitHub URL: https://github.com/NotsoJharedtrollOx17/Module10_AppliedDataScienceCapstone/blob/main/07_HandsOnLab_ML.ipynb



Results

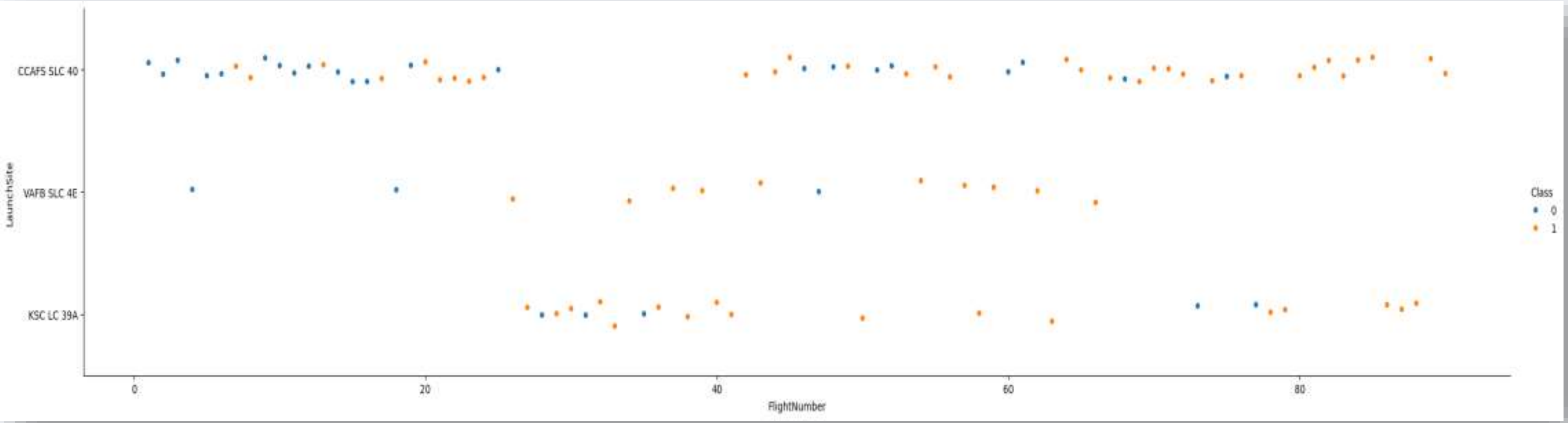
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

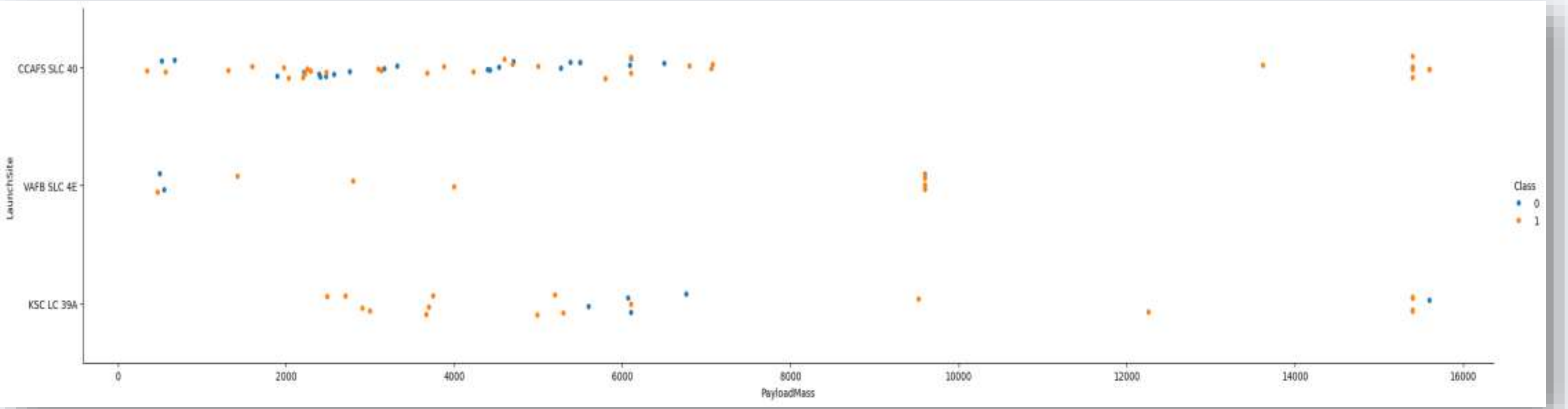
Flight Number vs. Launch Site

- ***CCADFS SLC 40*** *has a greater number of orange dots, indicating a greater success number of flights*



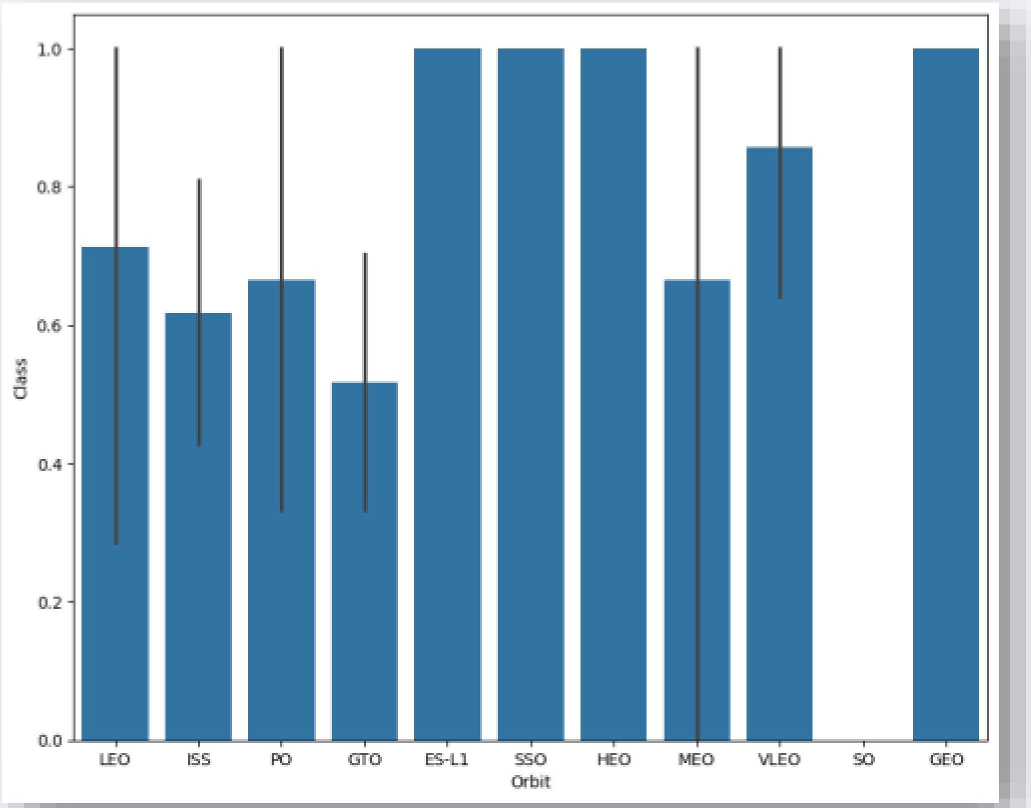
Payload vs. Launch Site

- Overall, from 2000 kg to 6000 kg, the successful launch count is greater.



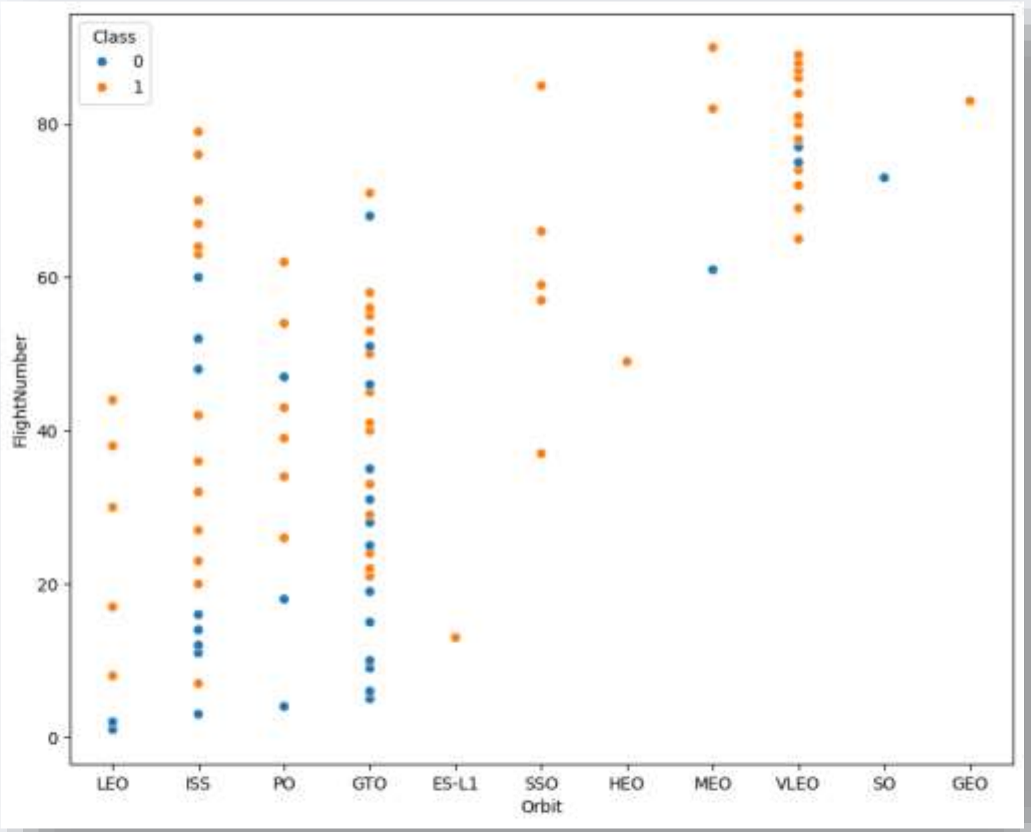
Success Rate vs. Orbit Type

- ES-L1, SSO, HEO & GEO orbits are the ones that show a ***greater success rate bar none.***



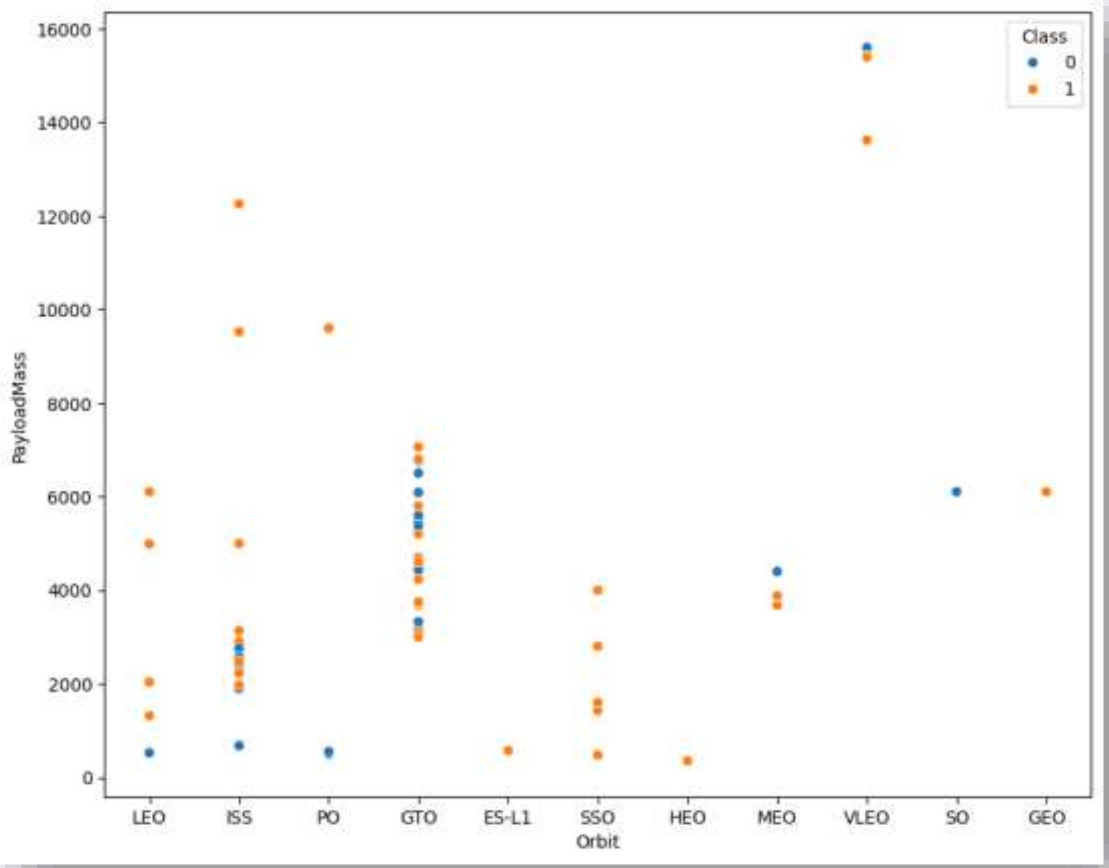
Flight Number vs. Orbit Type

- In contrast to the previous bar chart, this plot indicates that the most truthful successes are present in the LEO, ISS, PO & VLEO
- *The other previously accounted success orbits seems to suffer from a lot of positive confirmation bias*



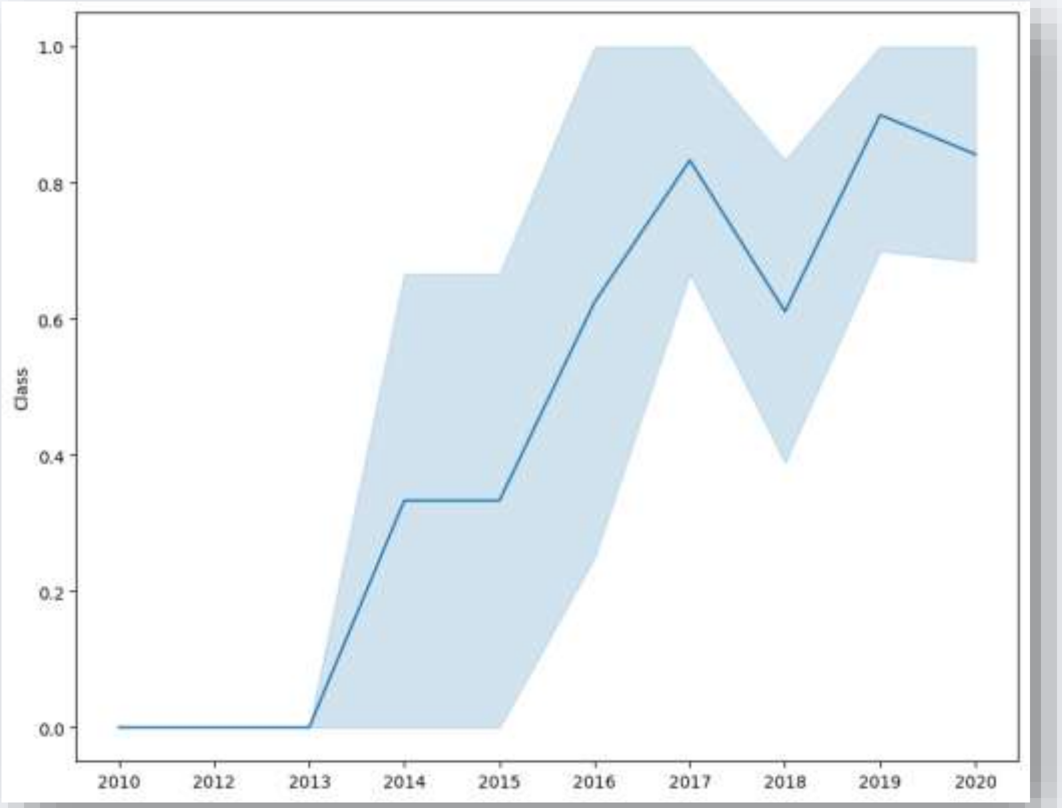
Payload vs. Orbit Type

- In a similar manner to the Payload Mass plot from before, the reliable threshold for the Payload to estimate success is from 2000 kg to 6000 kg
- *Other orbits seem to have greater success under lower payloads*



Launch Success Yearly Trend

- If historical data around the launches is to be trusted, we can expect that the *successful launches will be much more common occurrence than before.*



All Launch Site Names

- Find the names of the unique launch sites
- Current query shows a unique launch site that did not appeared in the plots *(CCAFS SLC-40)*

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

In [10]:
```
%sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE;
```

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

Out[10]:

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`
- We can appreciate that the query obtained successful launches from the get-go

```
In [11]: %sql SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE 'CCA%' LIMIT 5;
 * sqlite:///my_data1.db
Done.
Out[11]:
```

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

- Calculate the total payload carried by boosters from NASA
- One of their customers, NASA, has launched 45596 kg total!

```
In [12]: %sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE Customer = 'NASA (CRS)';

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

Out[12]:

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- Our focus for one of the boosters, has launched ~ 2928 kg total!

```
In [13]: %sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE Booster_Version = 'F9 v1.1';

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

Out[13]:

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- The first successful capture of any booster happened in late 2015!

```
In [14]: %sql SELECT MIN(Date) FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (ground pad)';

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

Out[14]:

MIN(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
- A notable observation is that F9-variant boosters can carry those weight ranges!

In [15]:

```
%sql SELECT Booster_Version FROM SPACEXTABLE WHERE
```

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

Out[15]:

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
- Curiously enough, most launches have been considered successes!

In [16]:
```
%sql SELECT Mission_Outcome, COUNT(*) FROM SPACEXTABLE GROUP BY Mission_Outcome;
```

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

Out[16]:

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

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- F9-variant boosters also successfully launch their max payloads!

In [17]:

```
%sql SELECT Booster_Version FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE);
```

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

Out[17]:

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
- January and April 2025 suffered the most defining failures of launch

```
In [21]: %sql SELECT SUBSTR(Date, 6, 2) AS Month, Landing_Outcome, Booster_Version, Launch_Site FROM SPACEXTABLE WHERE SUBSTR(Date, 6

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

Out[21]:

Month	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	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
- Most of the dataset consists of launches that were not attempted!

```
In [20]: %sql SELECT Landing_Outcome, COUNT(*) AS Count FROM SPACEXTABLE WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY La

 * sqlite:///my_data1.db
Done.
Out[20]:
```

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

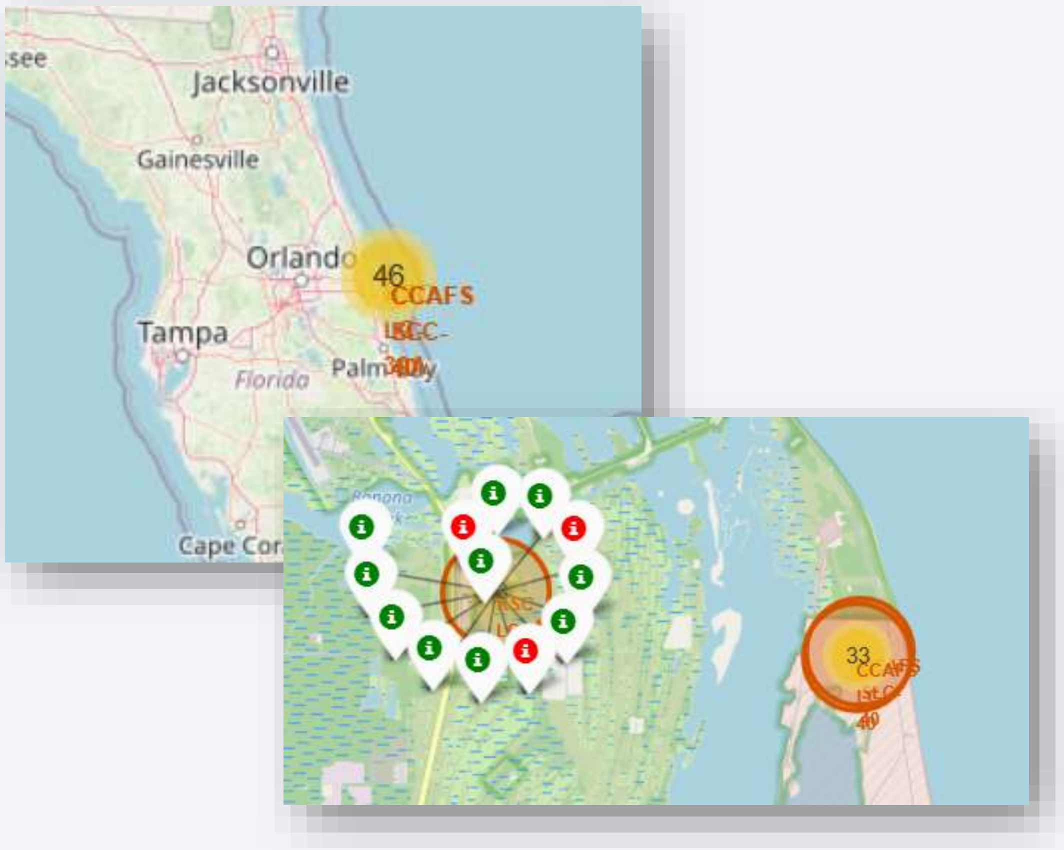
Launch Sites Proximities Analysis

Area of Operations (AO)

- As we tend to expect, all launch sites analyzed belong to the United States of America
- Launch sites are geographically divided into the *East Coast* & *West Coast* respectively

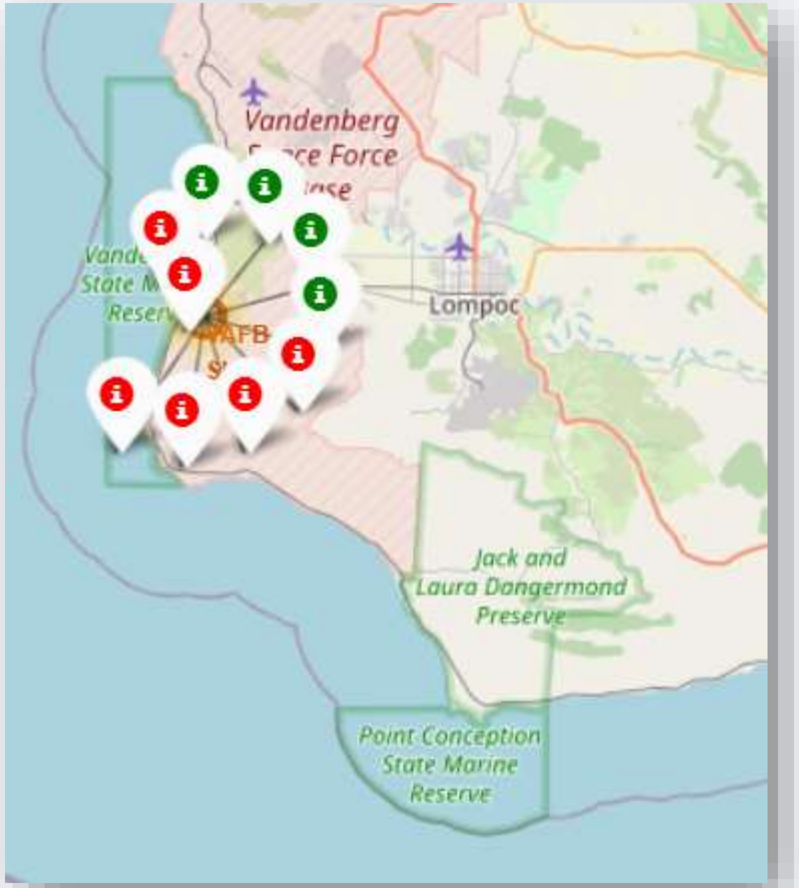
East Coast’s AO

- Most Launches Happen in Cabo Cañaveral
- They’re further separated into inland Cabo (where most green markers appear), and the actual NASA property (the 33 count)
- We may expect that most successful launches happen there



West Coast's AO

- Most of the launches are near a military base.
- Interestingly, the launch site is around a very isolated area, which may influence the risk assessment of the launches.

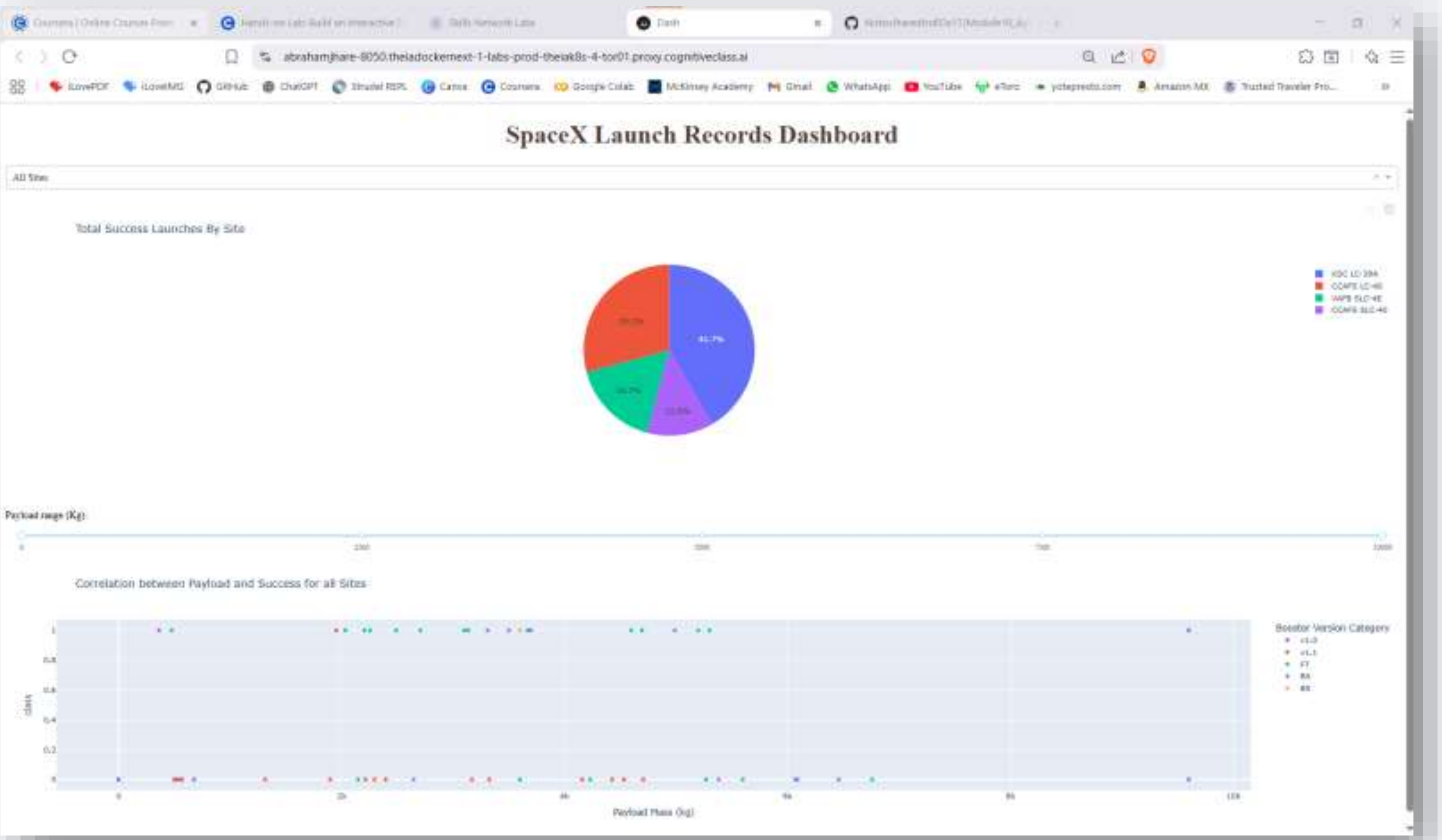


Section 4

Build a Dashboard with Plotly Dash

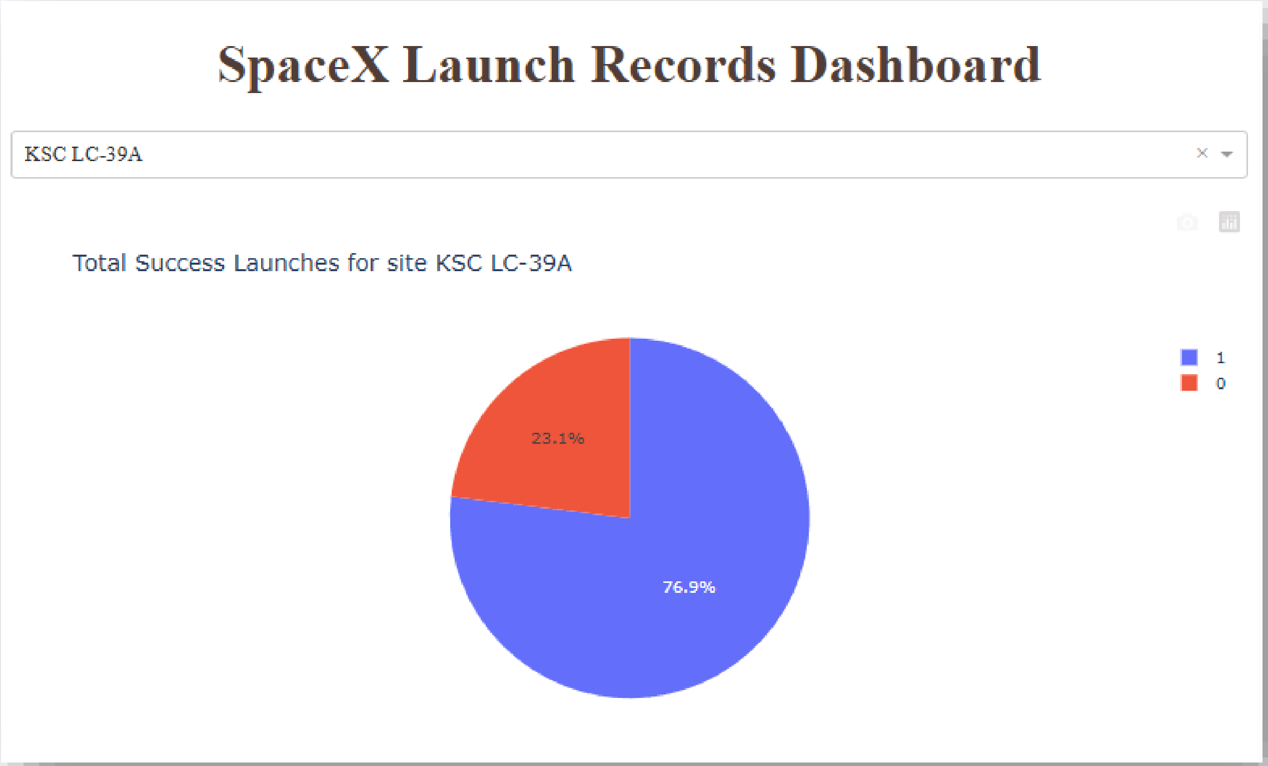
Launch Success for All Sites

- Overall successes happen with a payload of 2000 kg to 6000kg



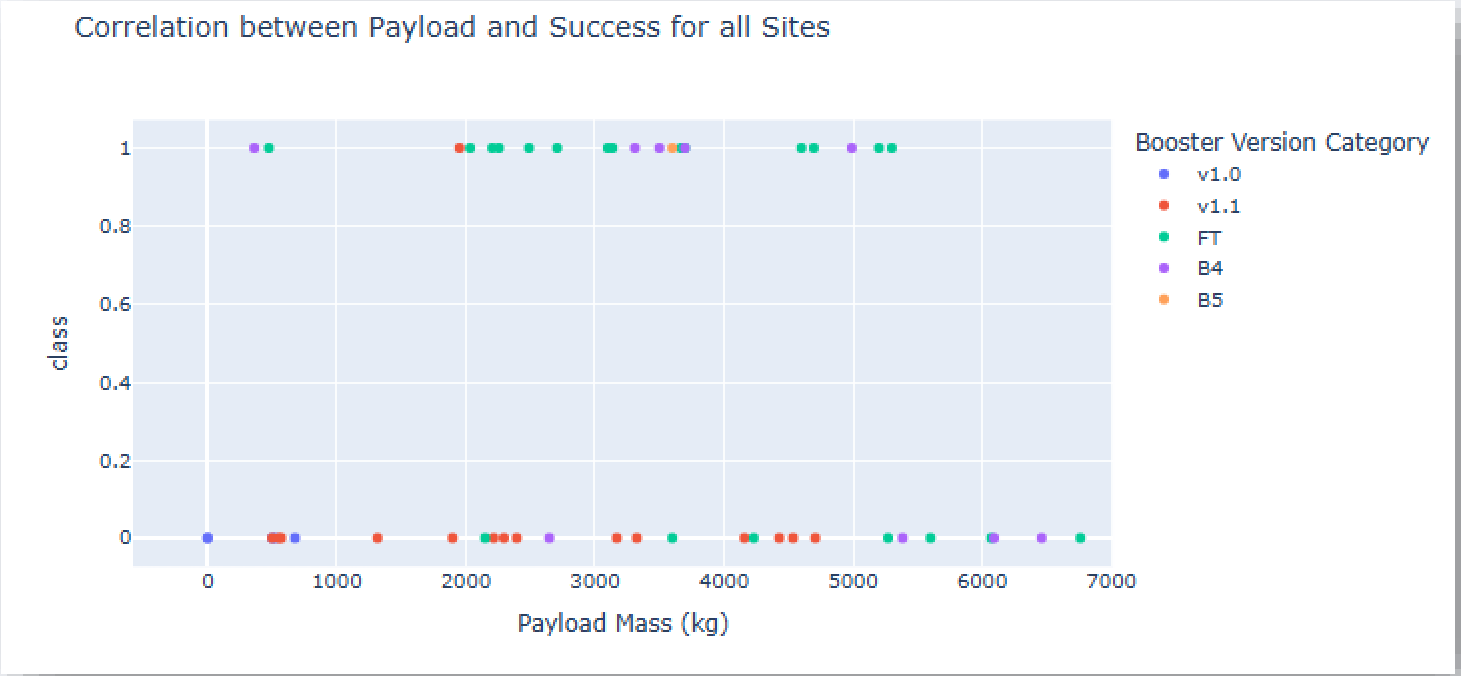
Launch Success for the Best Site

- We can see that *the best site has a success rate of ~ 77% !*



Payload and Success

- For our successful range (2000 kg to 6000 kg) we can appreciate that *the most performant booster category is the FT category,* where our F9-variant belongs!

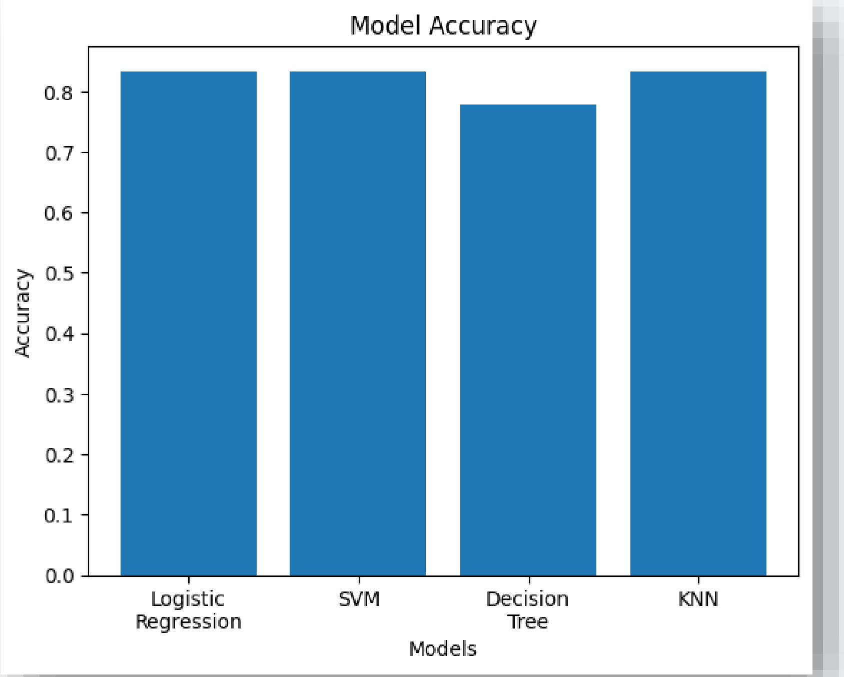


Section 5

Predictive Analysis (Classification)

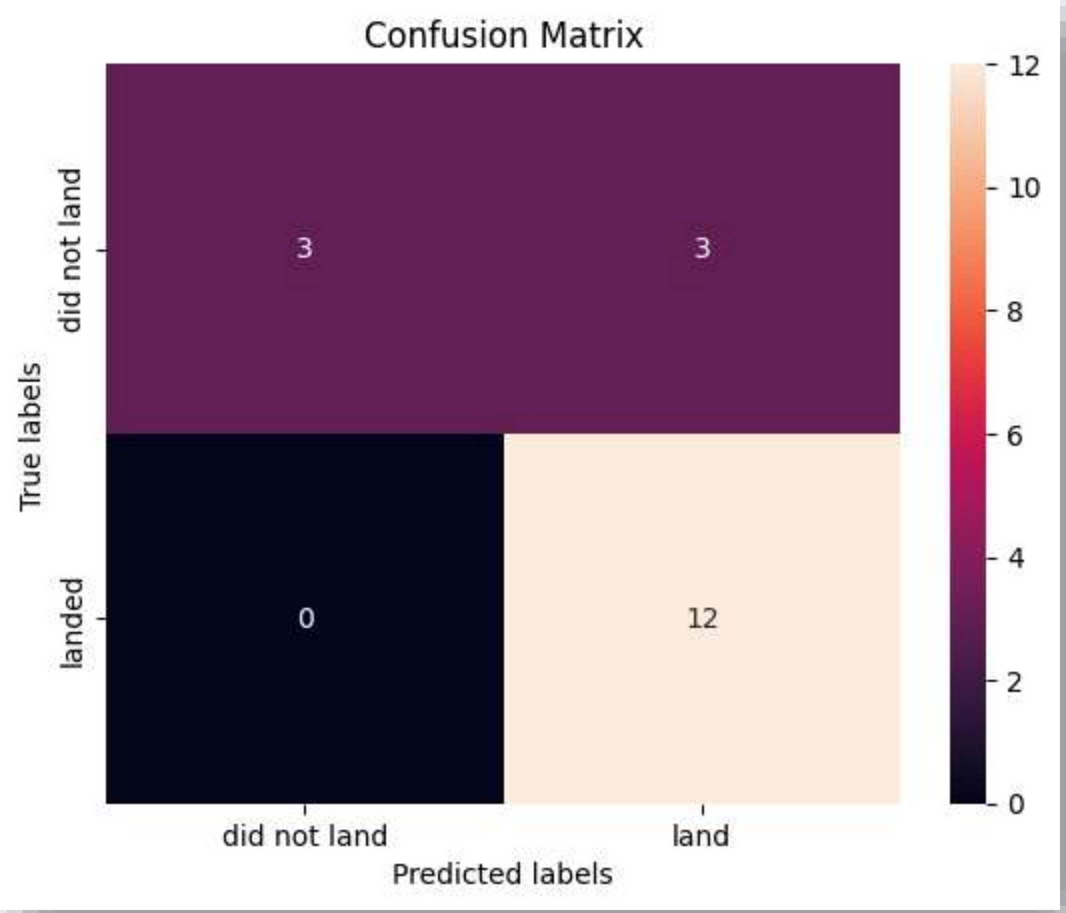
Classification Accuracy

- After ML modelling, we found that, with the current data available, all models except the decision tree obtained a better classification record overall.
- The exception, the Decision Tree, seems to have an extra account of False Negatives that suggests this model may tamper with estimations under a larger dataset.



Confusion Matrix

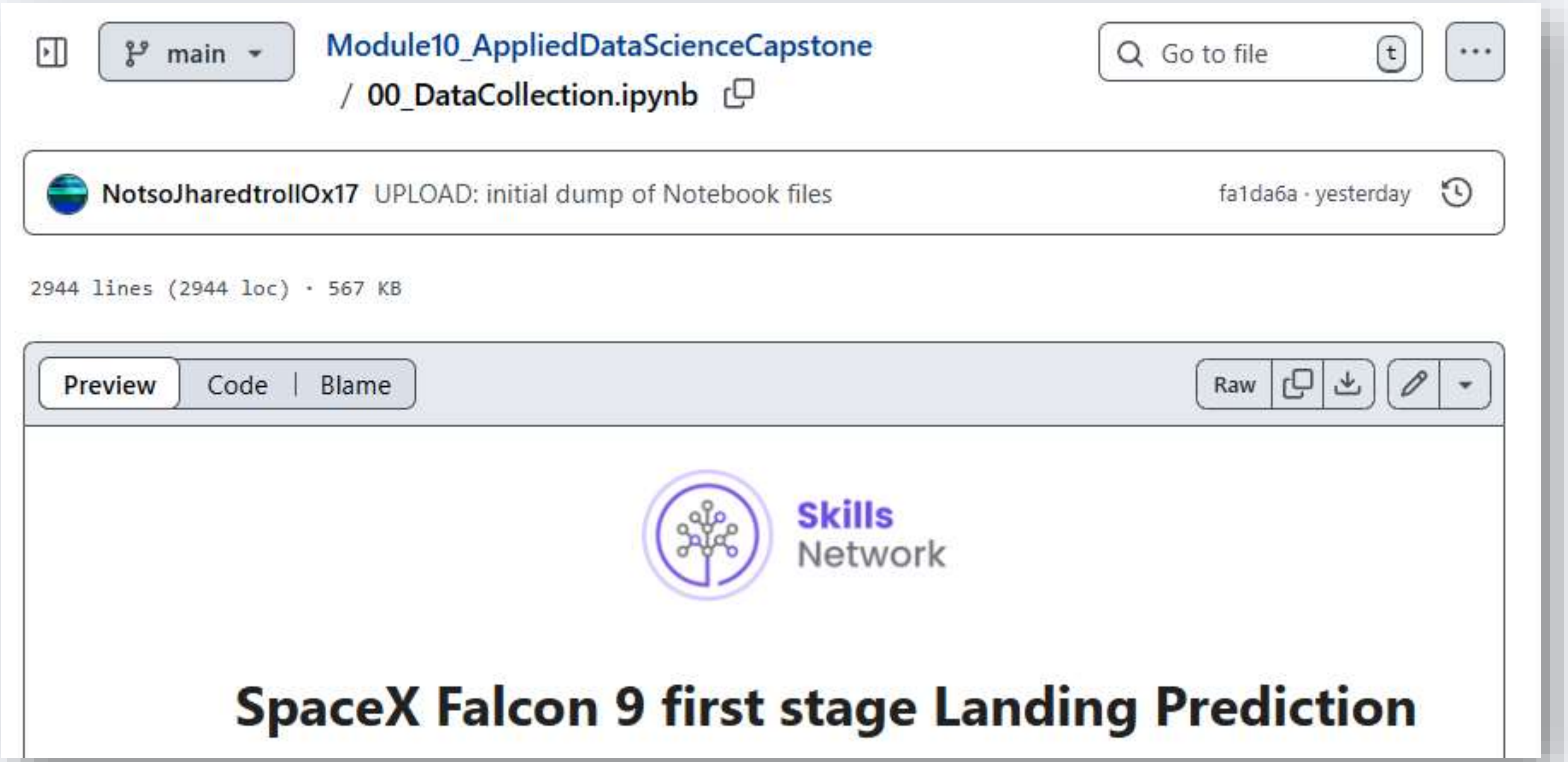
- As stated previously, all the other models seem to possess a better confusion matrix with lesser False Negatives.
- For our criteria, better prediction around all models indicate sufficient evidence for trusting the models.



Conclusions

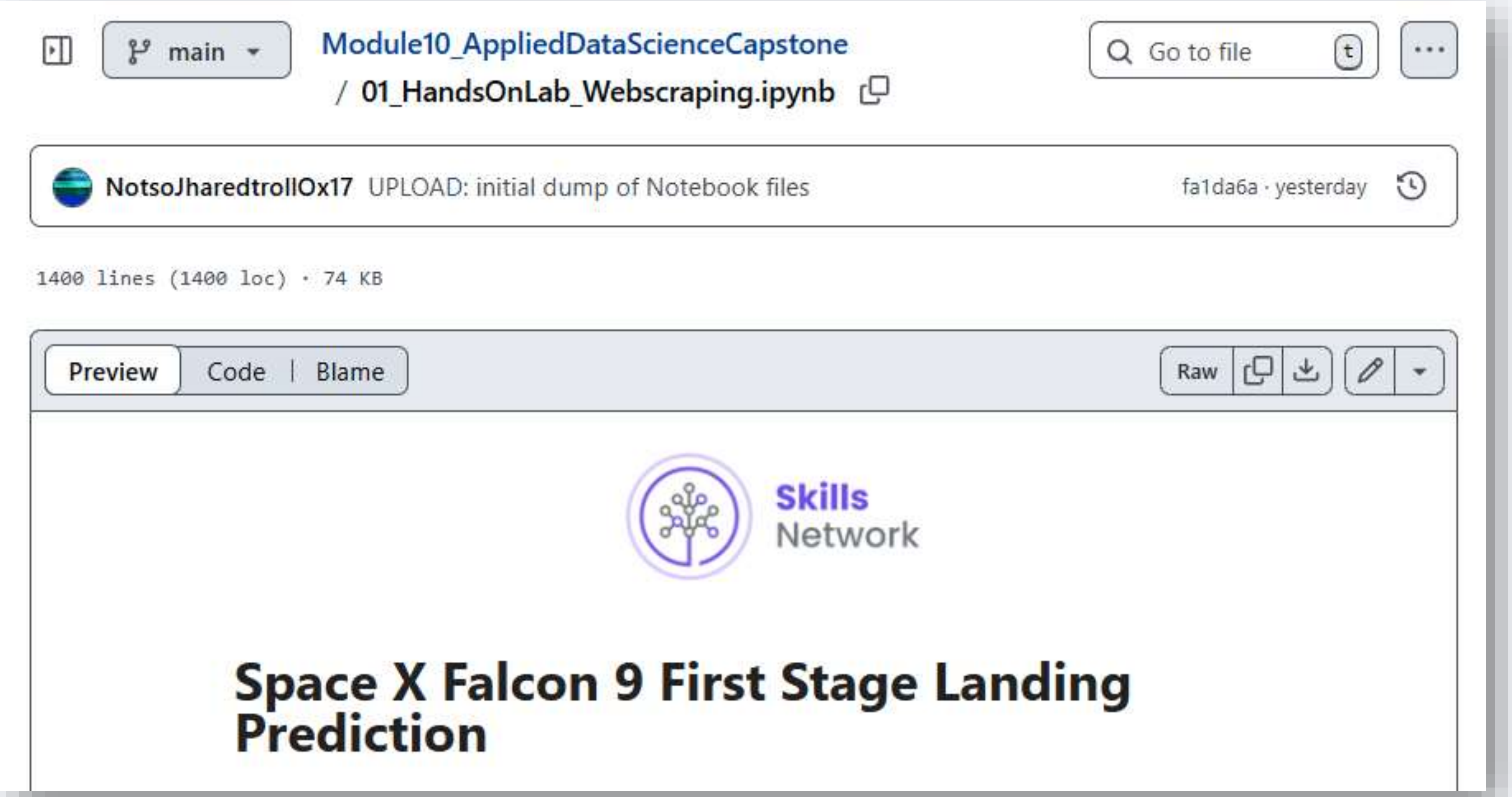
- *SpaceX launch success can be expect for the long term*
- The East Coast overall seems to have more successful launches due to its close proximity to NASA's compounds
- The F9-variant boosters offer the most reliable launch success rate
- Around 2000 kg to 6000 kg can be launched to expect a successful launch

Appendix

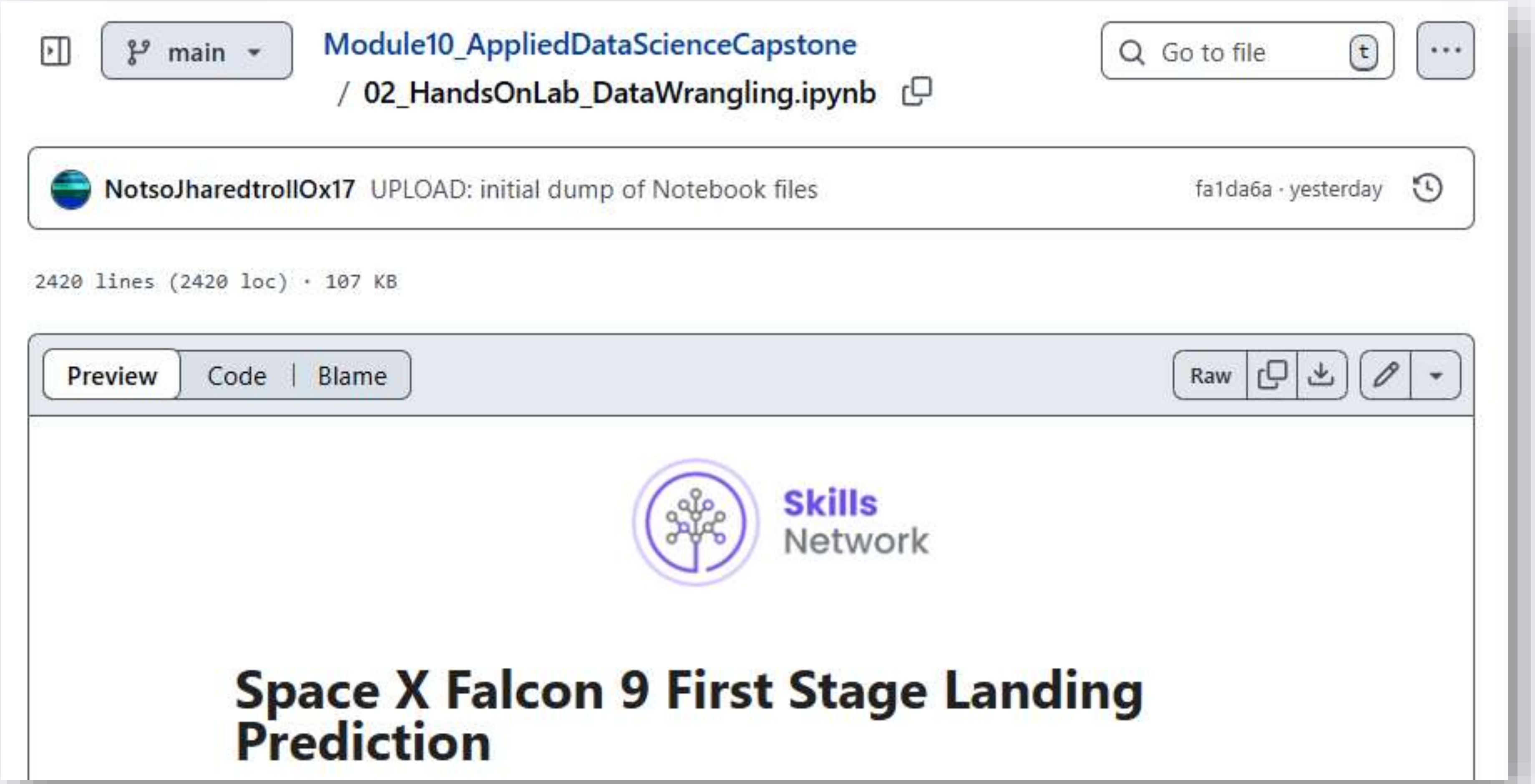


main
Module10_AppliedDataScienceCapstone
/ 00_DataCollection.ipynb
Go to file
t
NotsoJharedtrollOx17 UPLOAD: initial dump of Notebook files
fa1da6a · yesterday
2944 lines (2944 loc) · 567 KB
Preview
Code
Blame
Raw
Skills
Network
SpaceX Falcon 9 first stage Landing Prediction

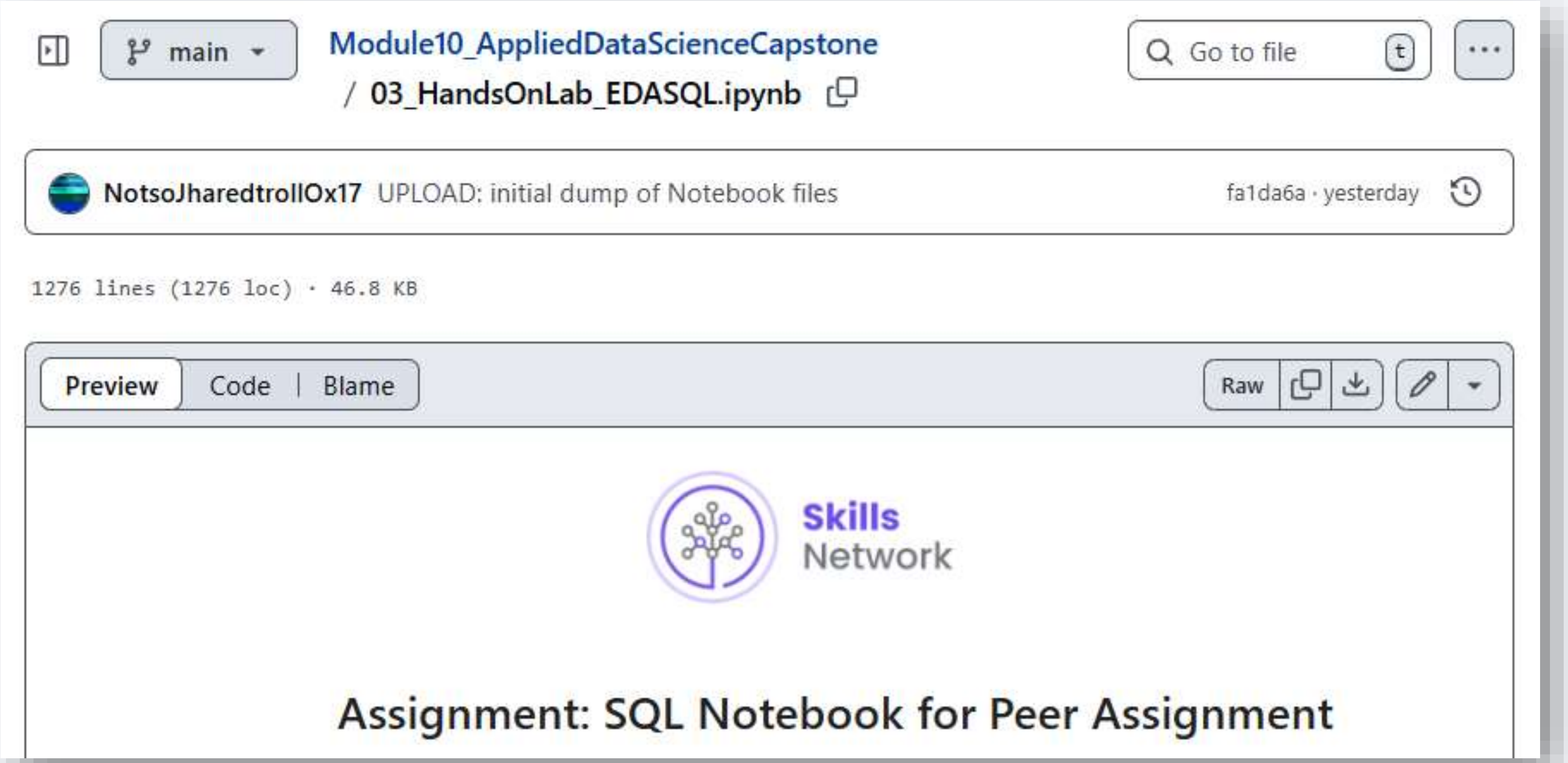
Appendix


main
Module10_AppliedDataScienceCapstone
/ 01_HandsOnLab_Webscraping.ipynb
Go to file
NotsoJharedtrollOx17 UPLOAD: initial dump of Notebook files
fa1da6a · yesterday
1400 lines (1400 loc) · 74 KB
Preview
Code
Blame
Raw
Skills
Network
Space X Falcon 9 First Stage Landing Prediction

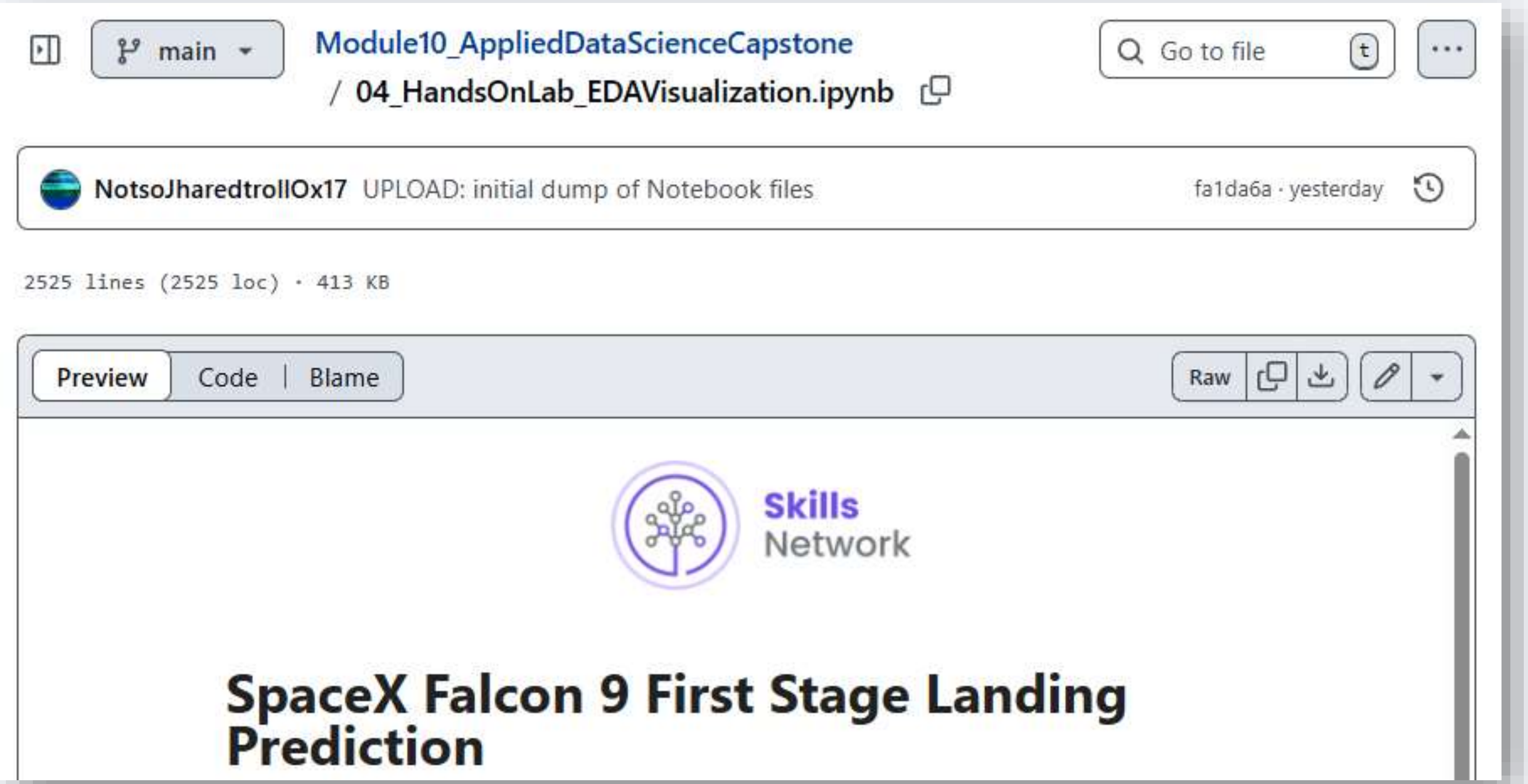
Appendix


main
Module10_AppliedDataScienceCapstone
/ 02_HandsOnLab_DataWrangling.ipynb
Go to file
t
NotsoJharedtrollOx17 UPLOAD: initial dump of Notebook files
fa1da6a · yesterday
2420 lines (2420 loc) · 107 KB
Preview
Code
Blame
Raw
Skills
Network
Space X Falcon 9 First Stage Landing Prediction

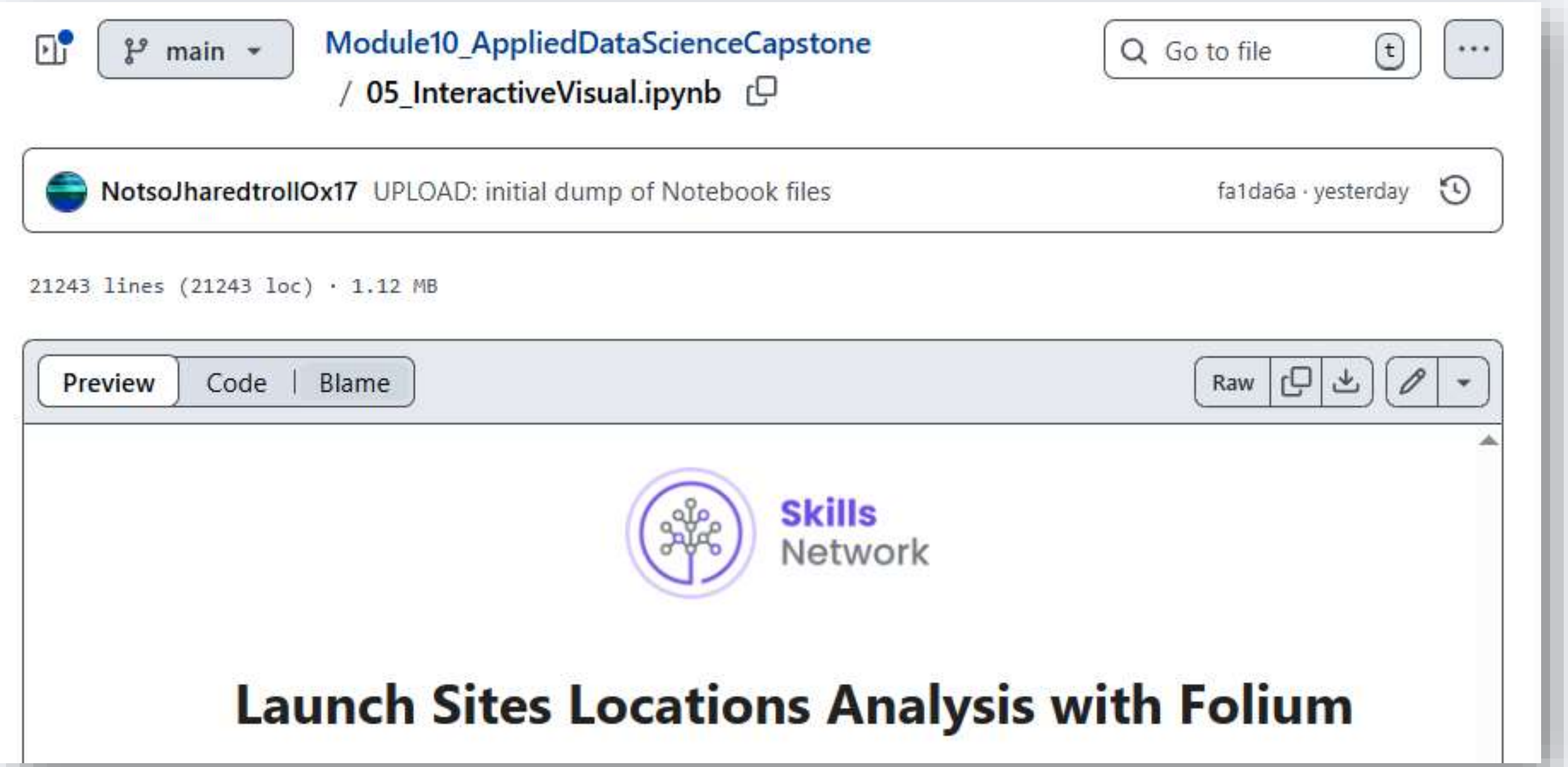
Appendix


main
Module10_AppliedDataScienceCapstone
/ 03_HandsOnLab_EDASQL.ipynb
Go to file
t
NotsoJharedtrollOx17 UPLOAD: initial dump of Notebook files
fa1da6a · yesterday
1276 lines (1276 loc) · 46.8 KB
Preview
Code
Blame
Raw
Skills
Network
Assignment: SQL Notebook for Peer Assignment

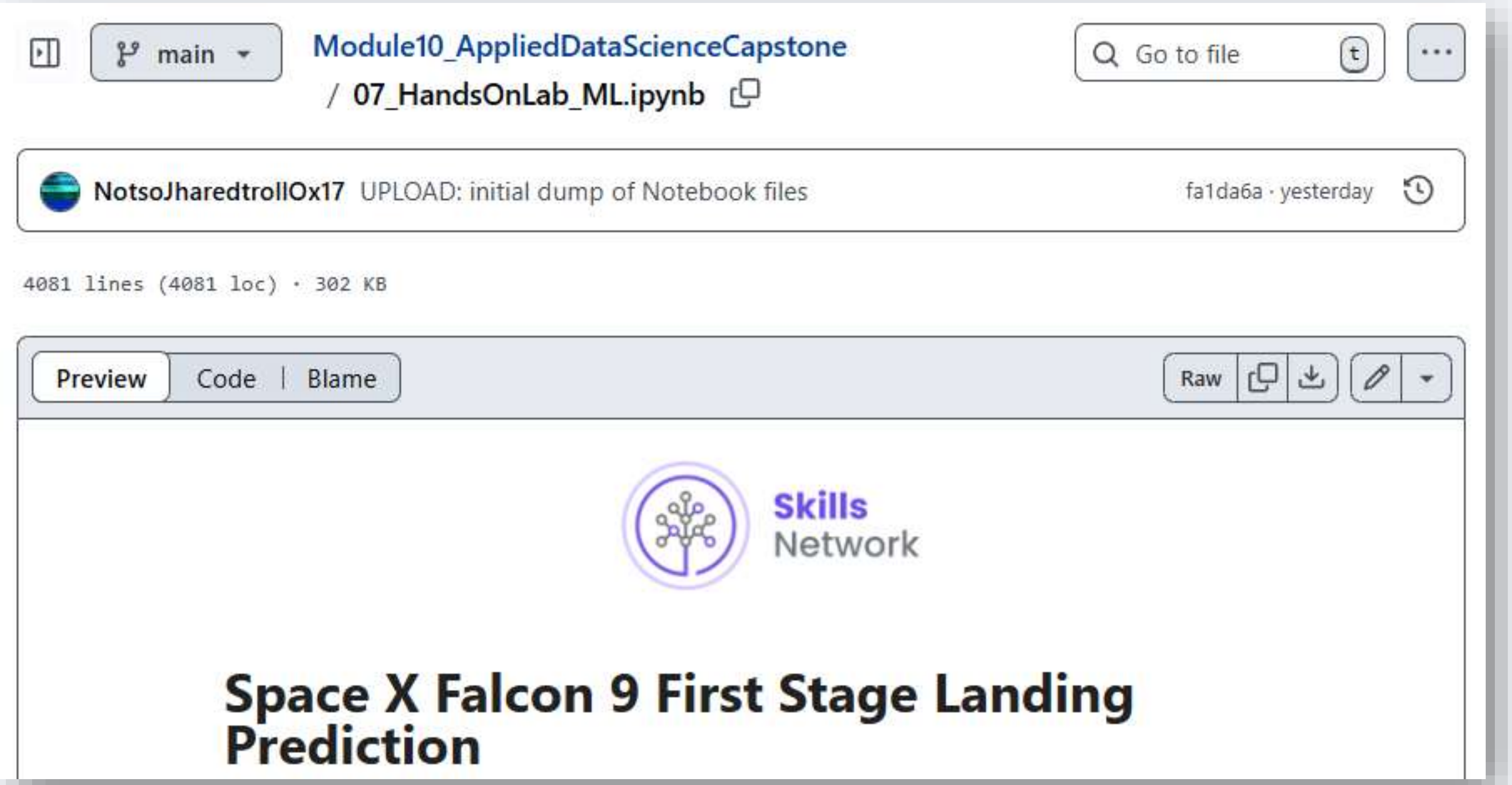
Appendix


main
Module10_AppliedDataScienceCapstone
/ 04_HandsOnLab_EDAVisualization.ipynb
Go to file
t
NotsoJharedtrollOx17 UPLOAD: initial dump of Notebook files
fa1da6a · yesterday
2525 lines (2525 loc) · 413 KB
Preview
Code
Blame
Raw
Skills
Network
SpaceX Falcon 9 First Stage Landing Prediction

Appendix


main
Module10_AppliedDataScienceCapstone
/ 05_InteractiveVisual.ipynb
Go to file
NotsoJharedtrollOx17 UPLOAD: initial dump of Notebook files
fa1da6a · yesterday
21243 lines (21243 loc) · 1.12 MB
Preview
Code
Blame
Raw
Skills
Network
Launch Sites Locations Analysis with Folium

Appendix


main
Module10_AppliedDataScienceCapstone
/ 07_HandsOnLab_ML.ipynb
Go to file
t
NotsoJharedtrollOx17 UPLOAD: initial dump of Notebook files
fa1da6a · yesterday
4081 lines (4081 loc) · 302 KB
Preview
Code
Blame
Raw
Skills
Network
Space X Falcon 9 First Stage Landing Prediction

Thank you!