Predict First Stage Landing

Ramachandran Chettiar

08-12-2023

IBM Developer

SKILLS NETWORK

OUTLINE

- Executive Summary
- Introduction
- Methodology
  - Data Collection
  - Data Wrangling
  - EDA
  - Interactive Analysis and Visualization
  - Predictive Analysis
- Results
  - EDA Visualization
  - EDA with SQL
  - Dashboard
  - Interactive Visual Analytics
- Discussion
  - Findings & Predictions
- Conclusion
- Appendix

EXECUTIVE SUMMARY

- To predict if the Falcon 9 first stage will land successfully
- Determining success of first stage will help to know the cost of launch
- The bidding company against SpaceX will be able to provide competitive price for rocket launch

INTRODUCTION

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.

METHODOLOGY – Data Collection

- Data is collected using SpaceX provided Rest API
- Takes the dataset and uses the rocket column to call the API and append the data to the list
- Takes the dataset and uses the launchpad column to call the API and append the data to the list
- Takes the dataset and uses the payloads column to call the API and append the data to the lists
- Takes the dataset and uses the cores column to call the API and append the data to the lists
- Below is SpaceX API URL
  - "https://api.spacexdata.com/v4/launches/past"

METHODOLOGY – Data Wrangling

- Performing Data analysis on loaded SpaceX dataset
- Identifying missing values
- Understand the numerical and categorical data
- Determine number launch in different site
- Understanding landing outcome for each orbit
- Create class that informs success rate with ‘1’ as landing successfully and ‘0’ as landing unsuccessfully

METHODOLOGY – EDA

- Performing exploratory analysis on SpaceX data to find variable affecting launch outcome
- Performed feature engineering to identify the variable affect success rate
- Used pandas libraries for data analysis
- Used seaborn and matplotlib to plot various charts

METHODOLOGY – Interactive Visual Analytics

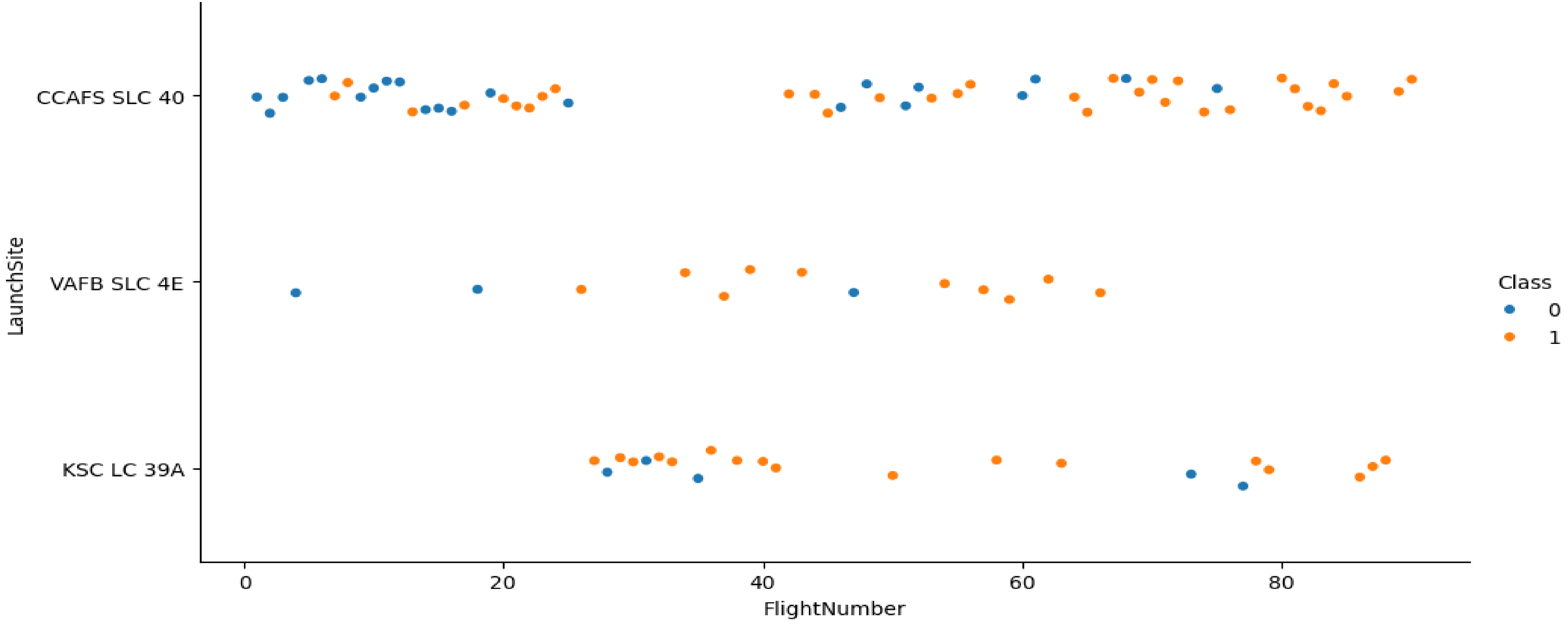
- Added a Launch Site drop down to select “All sites” or the list of launch sites from data frame
- Create call function that build Pie chart based for drop down selection
- Add a slider for payload range
- Create call function that build scatter chart based for drop down selection and given pay load range

METHODOLOGY – Predictive Analysis

- Create a column class to data frame for classifying the launch success
- Standardize the data using GridSearchCV
- Split the data into training data and test data
- Performing below statistical algorithm to determine the better predict score for landing success on test data
  - logistic regression
  - Support vector machine
  - Decision Tree Classifier
  - K nearest Neighbors

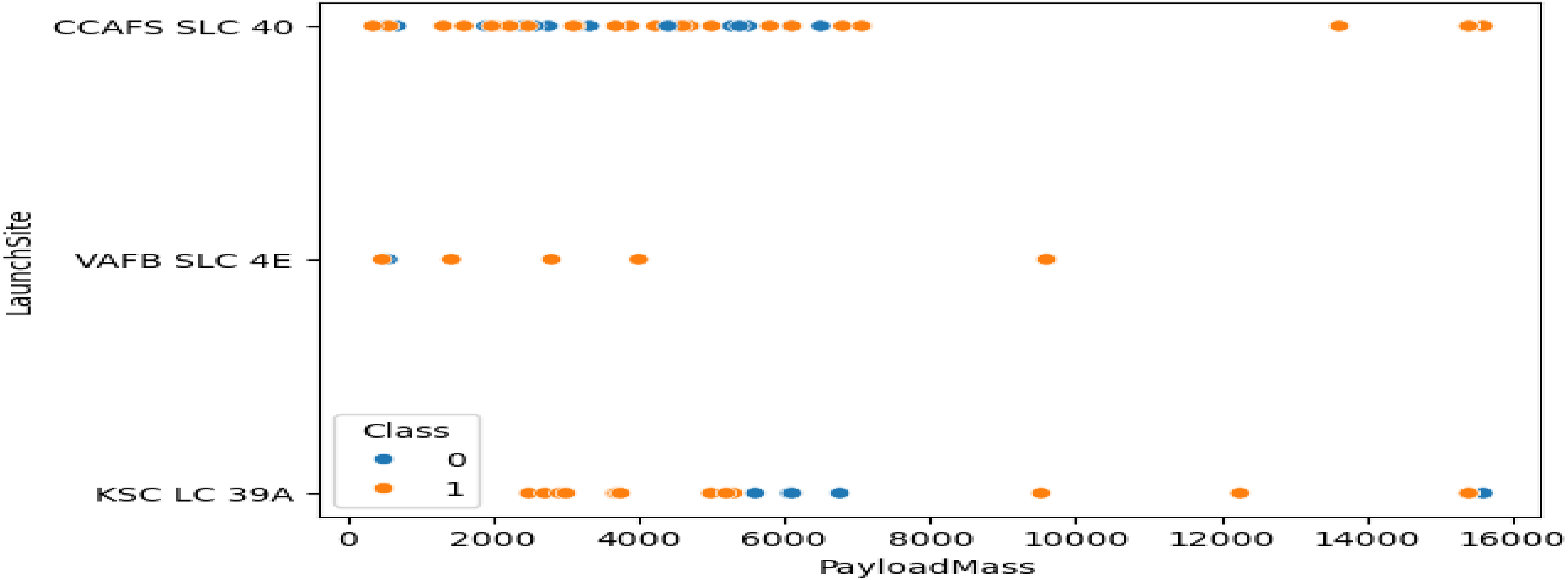
RESULTS - EDA Visualization

CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%.



RESULTS - EDA Visualization

The VAFB-SLC launch site has no rockets launched for heavy payload mass(greater than 10000).

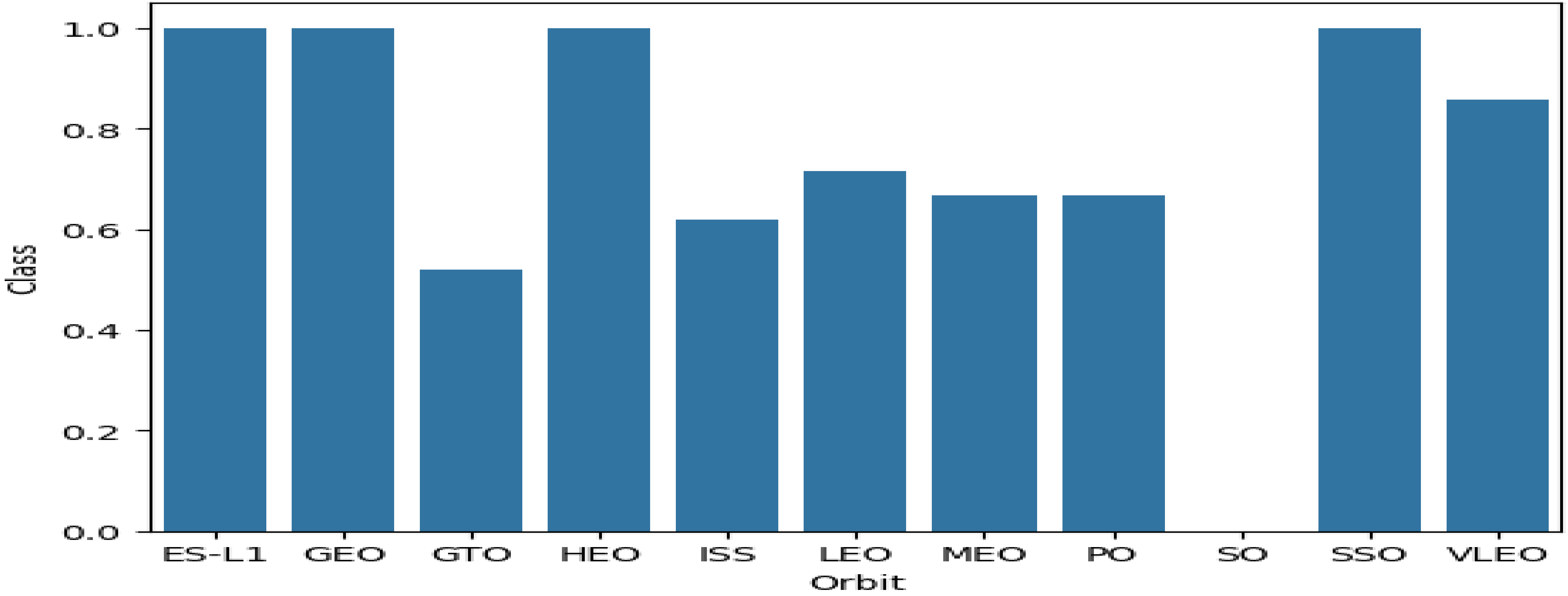


IBM Developer

SKILLS NETWORK

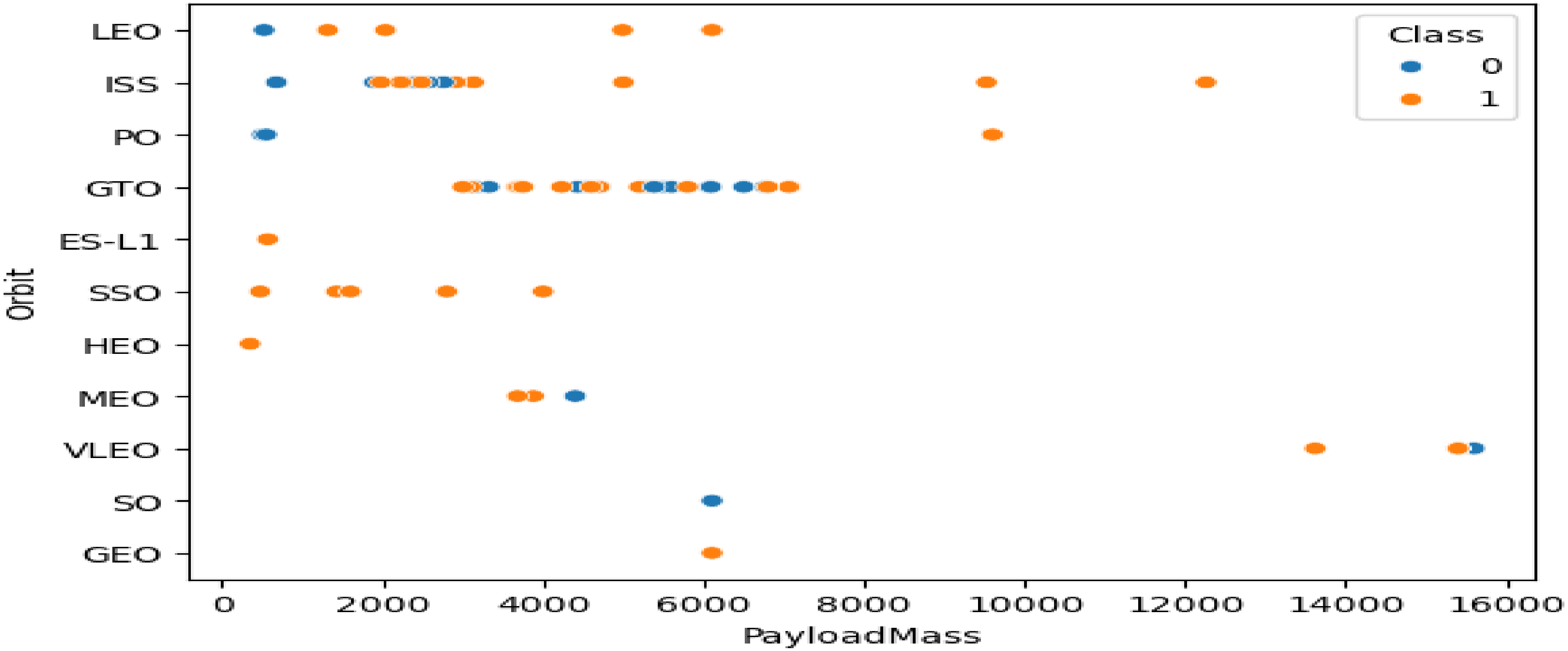
RESULTS - EDA Visualization

The success rate is high in orbit ES-L1, GEO, HEO, SSO



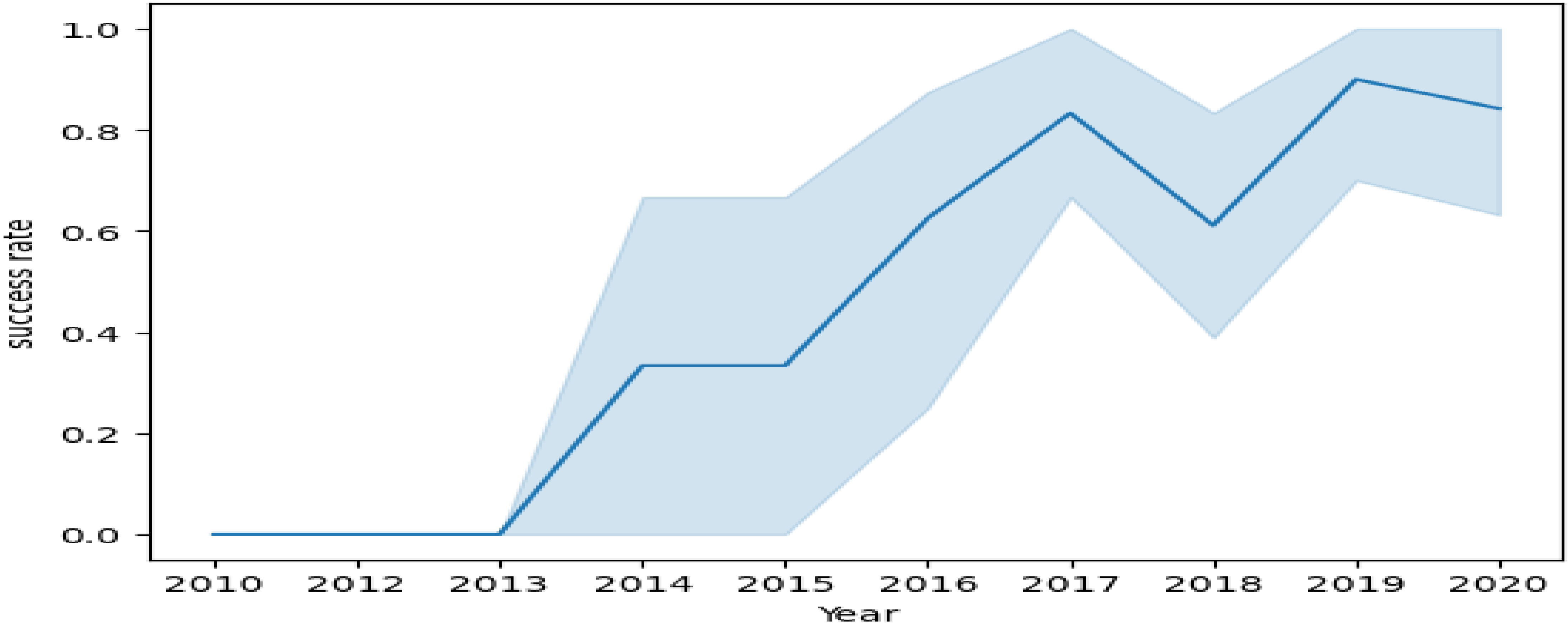
RESULTS - EDA Visualization

The landing rate are more positive for Polar, LEO and ISS orbit type

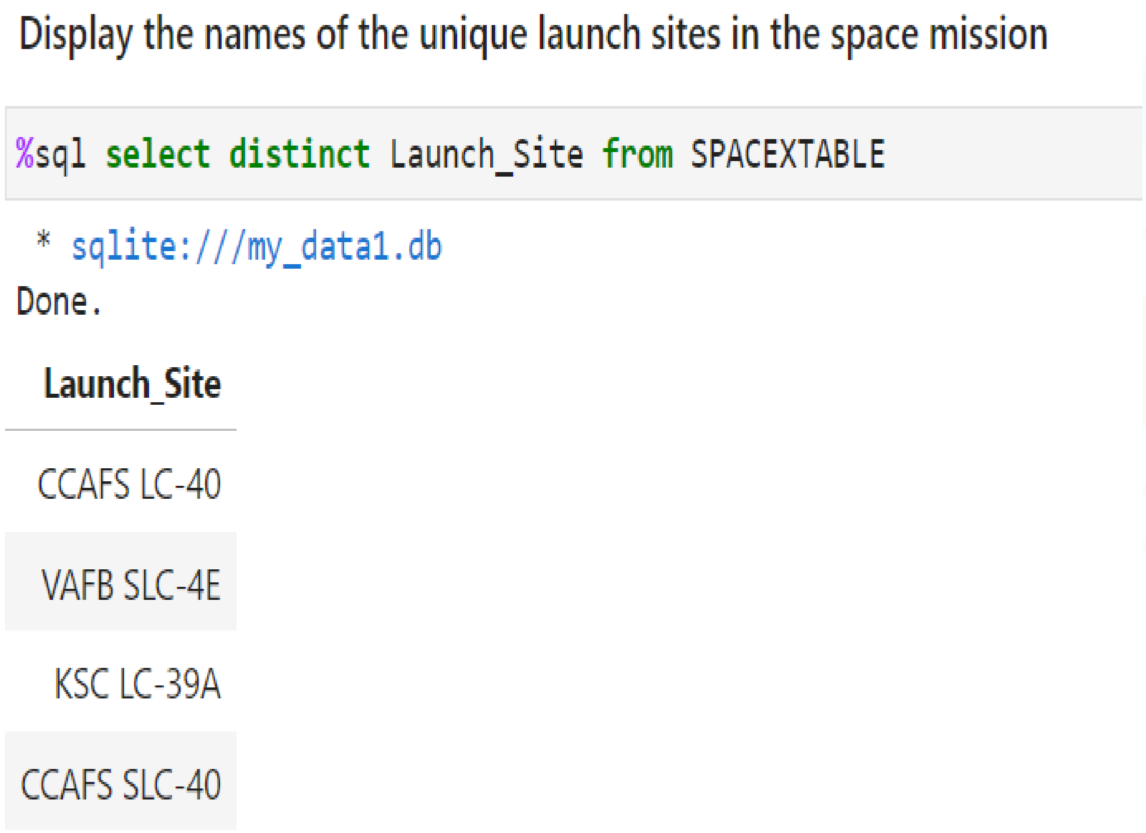


RESULTS - EDA Visualization

The success rate since 2013 kept increasing till 2020



RESULTS - EDA with SQL



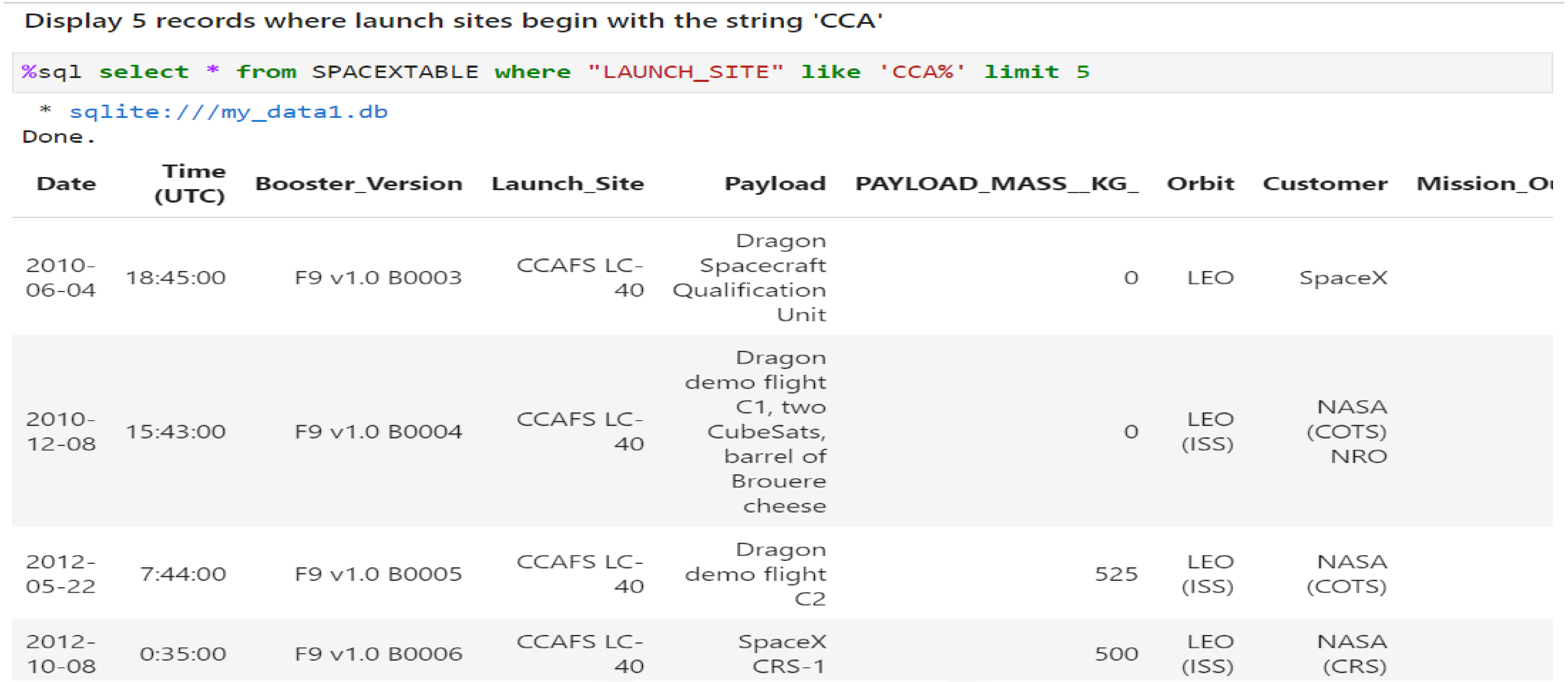
Display the names of the unique launch sites in the space mission

```
%sql select distinct Launch_Site from SPACEXTABLE
```

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

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

RESULTS - EDA with SQL



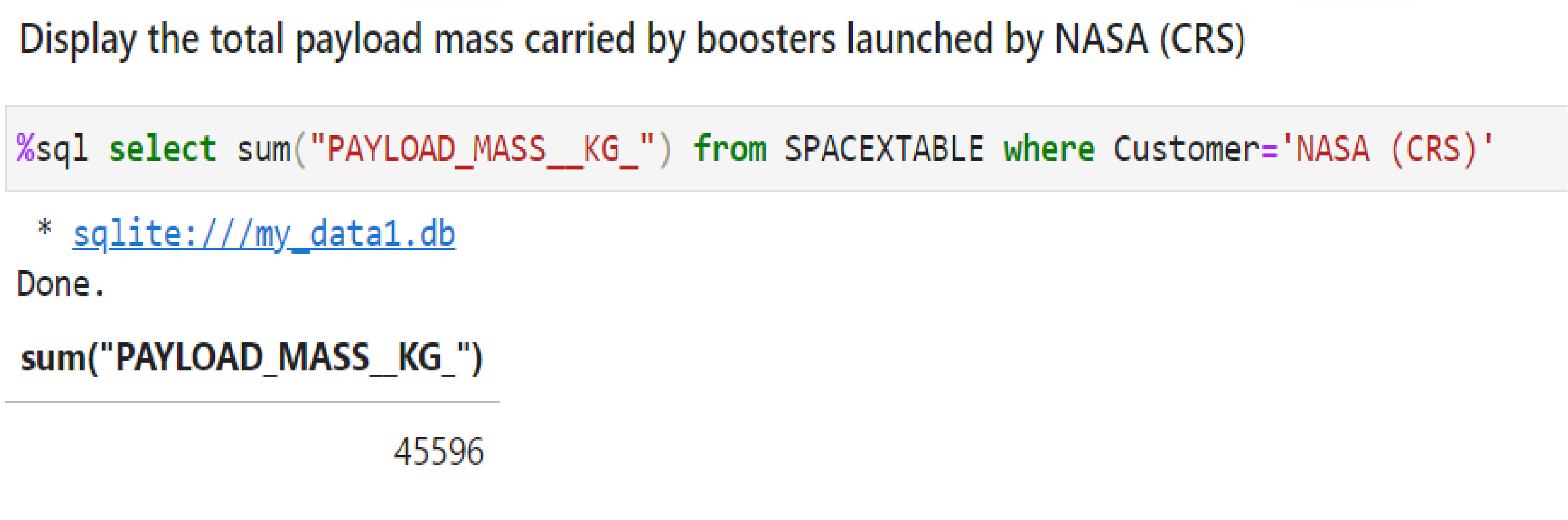
Display 5 records where launch sites begin with the string 'CCA'

```
%sql select * from SPACEXTABLE where "LAUNCH_SITE" like 'CCA%' limit 5
```

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_O
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	
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	
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	

RESULTS - EDA with SQL



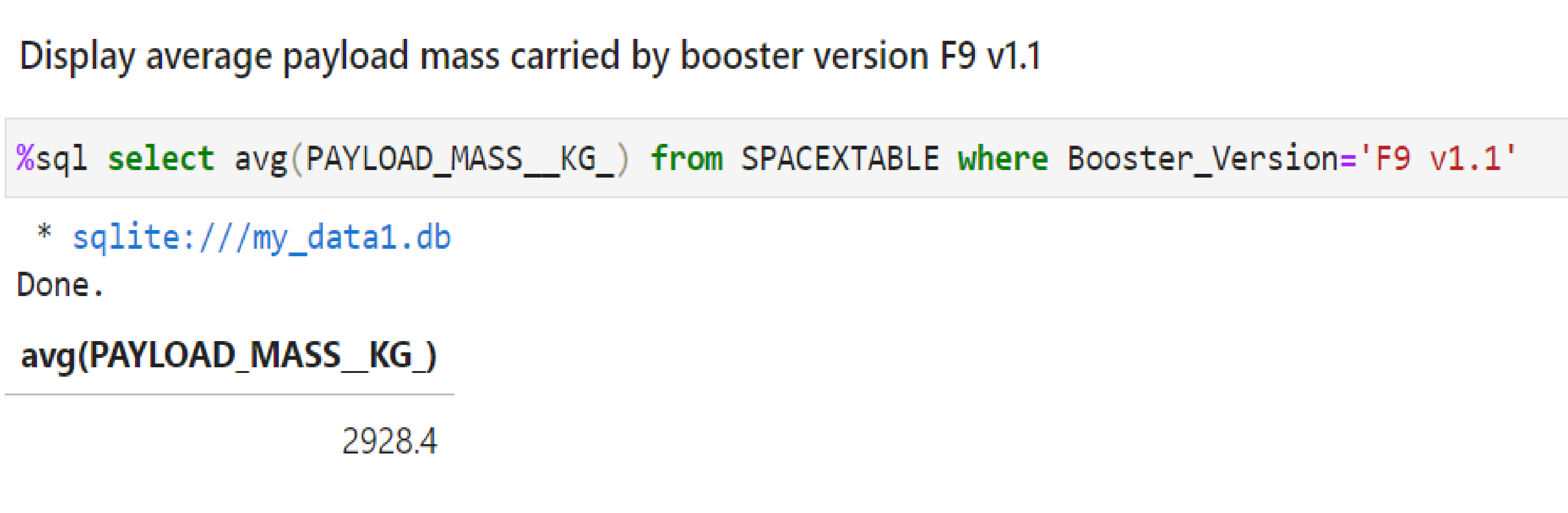
Display the total payload mass carried by boosters launched by NASA (CRS)

```
%sql select sum("PAYLOAD_MASS__KG_") from SPACEXTABLE where Customer='NASA (CRS)'
```

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

sum("PAYLOAD_MASS__KG_")
45596

RESULTS - EDA with SQL



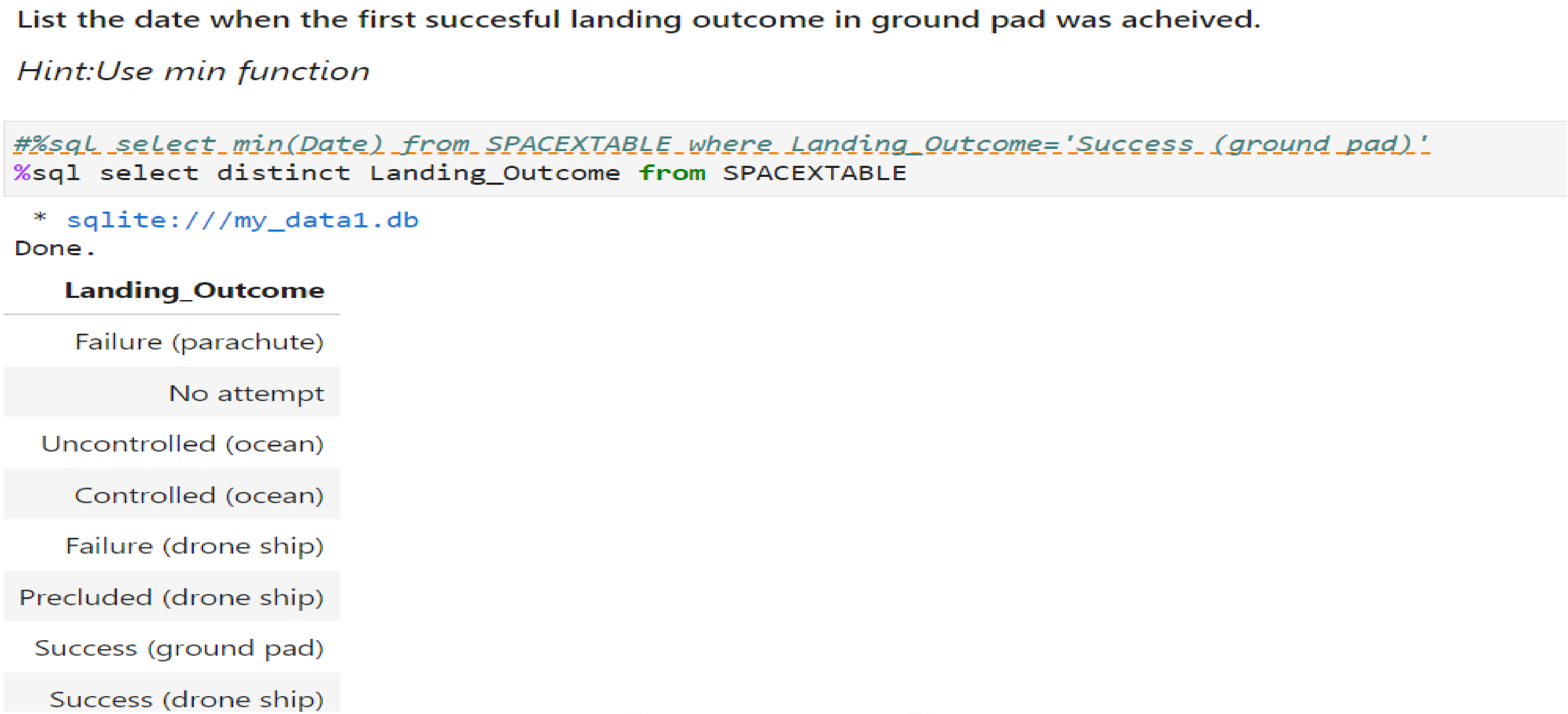
Display average payload mass carried by booster version F9 v1.1

```
%sql select avg(PAYLOAD_MASS__KG_) from SPACEXTABLE where Booster_Version='F9 v1.1'
```

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

avg(PAYLOAD_MASS__KG_)
2928.4

RESULTS - EDA with SQL



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

Hint:Use min function

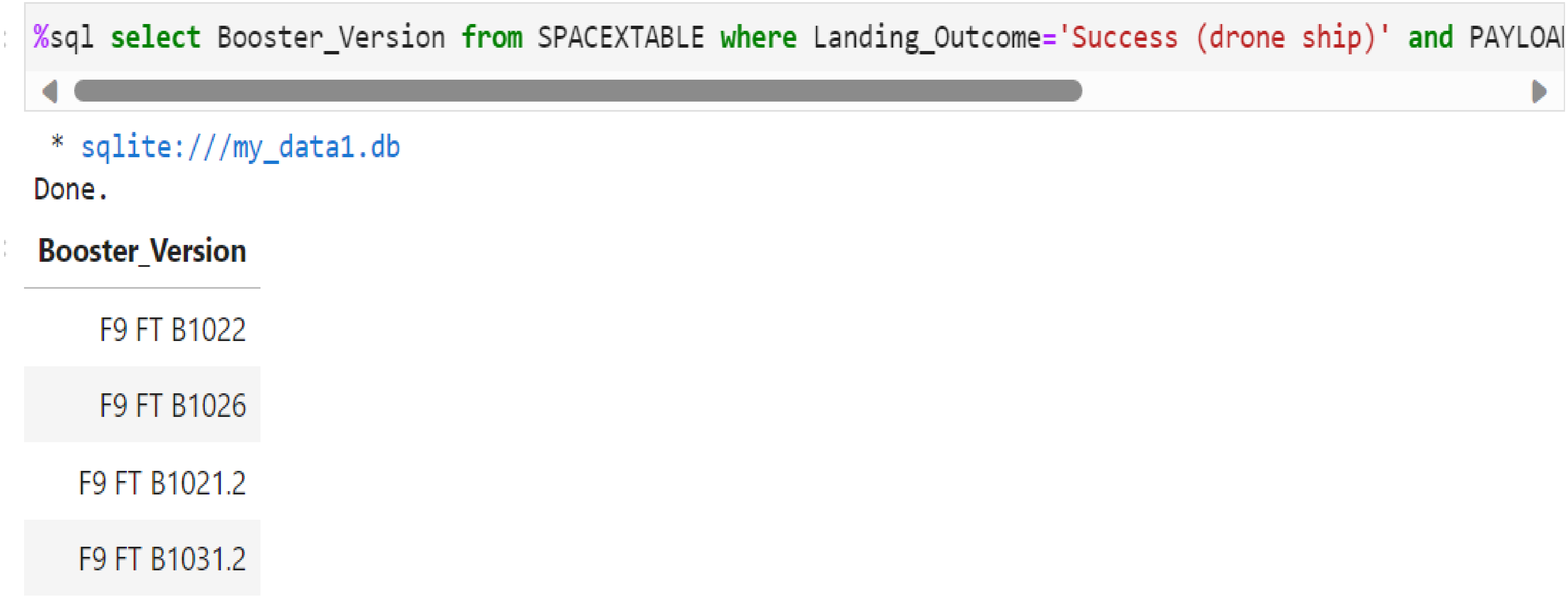
```
#%sql select min(Date) from SPACEXTABLE where Landing_Outcome='Success (ground pad)'
%sql select distinct Landing_Outcome from SPACEXTABLE
```

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

Landing_Outcome
Failure (parachute)
No attempt
Uncontrolled (ocean)
Controlled (ocean)
Failure (drone ship)
Precluded (drone ship)
Success (ground pad)
Success (drone ship)

RESULTS - EDA with SQL

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

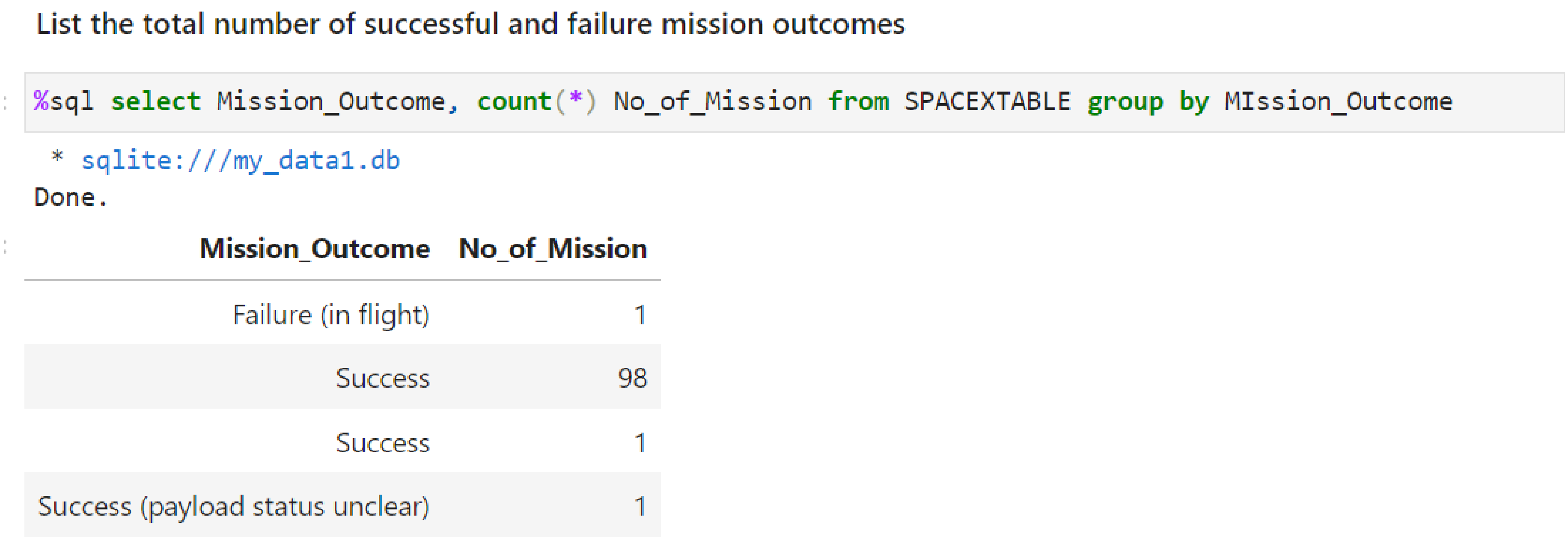


```
%sql select Booster_Version from SPACEXTABLE where Landing_Outcome='Success (drone ship)' and PAYLOA
```

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

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

RESULTS - EDA with SQL



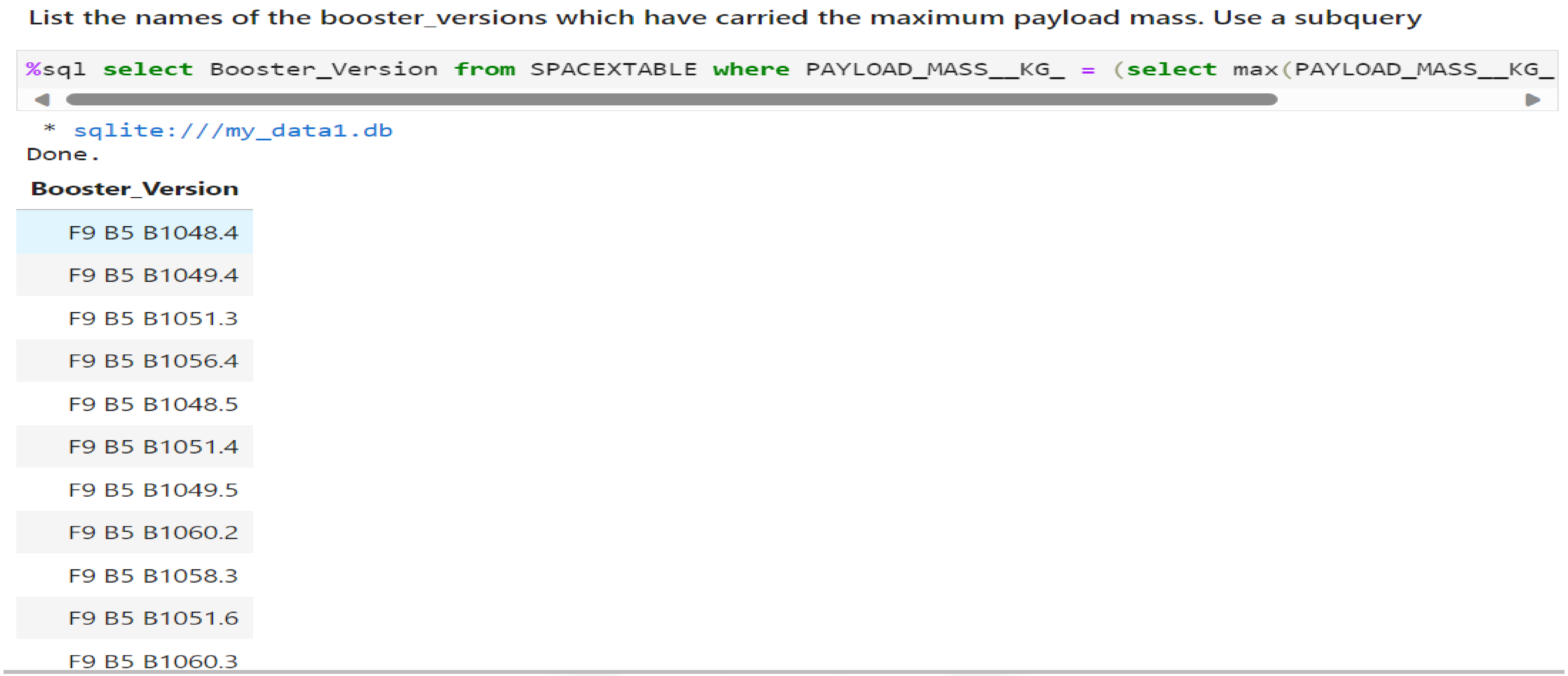
List the total number of successful and failure mission outcomes

```
%sql select Mission_Outcome, count(*) No_of_Mission from SPACEXTABLE group by MIssion_Outcome
```

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

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

RESULTS - EDA with SQL



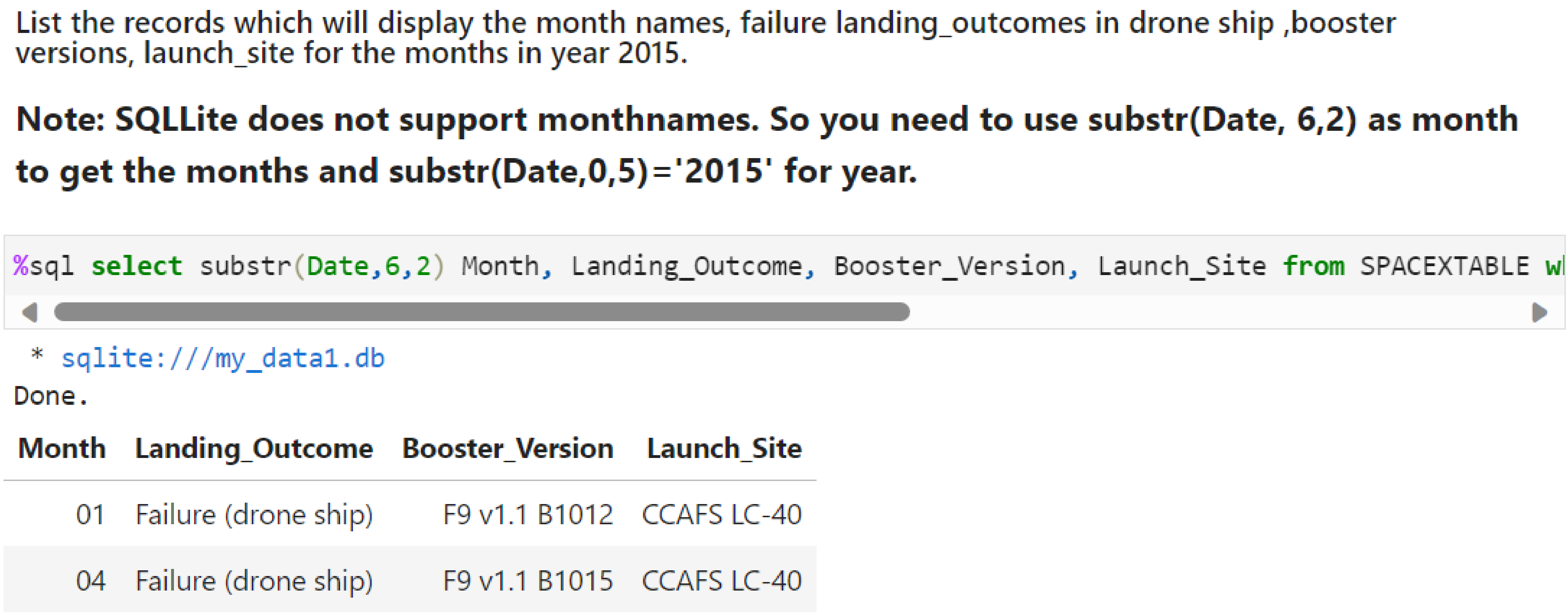
List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
%sql select Booster_Version from SPACEXTABLE where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_
```

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

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

RESULTS - EDA with SQL



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.

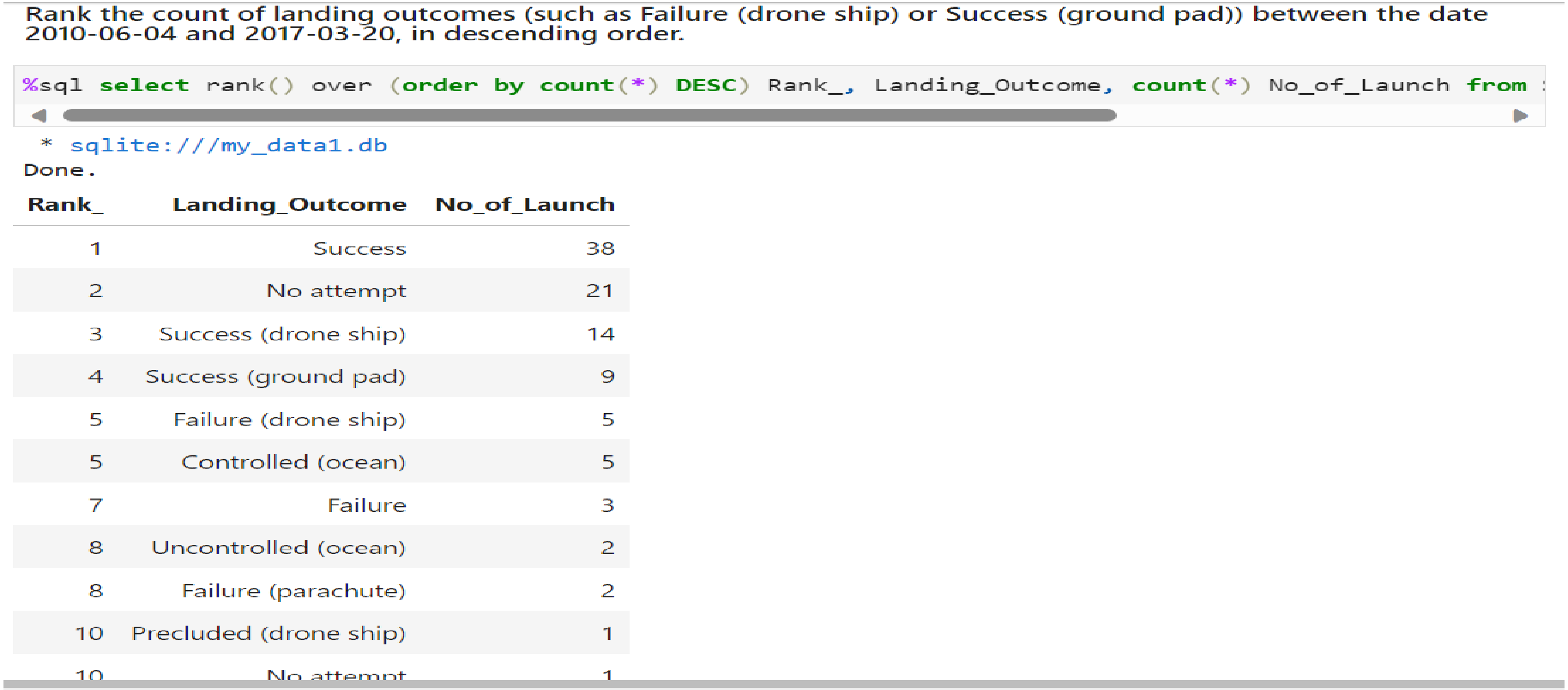
Note: SQLLite does not support monthnames. So you need to use substr(Date, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.

```
%sql select substr(Date,6,2) Month, Landing_Outcome, Booster_Version, Launch_Site from SPACEXTABLE w
```

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

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

RESULTS - EDA with SQL



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.

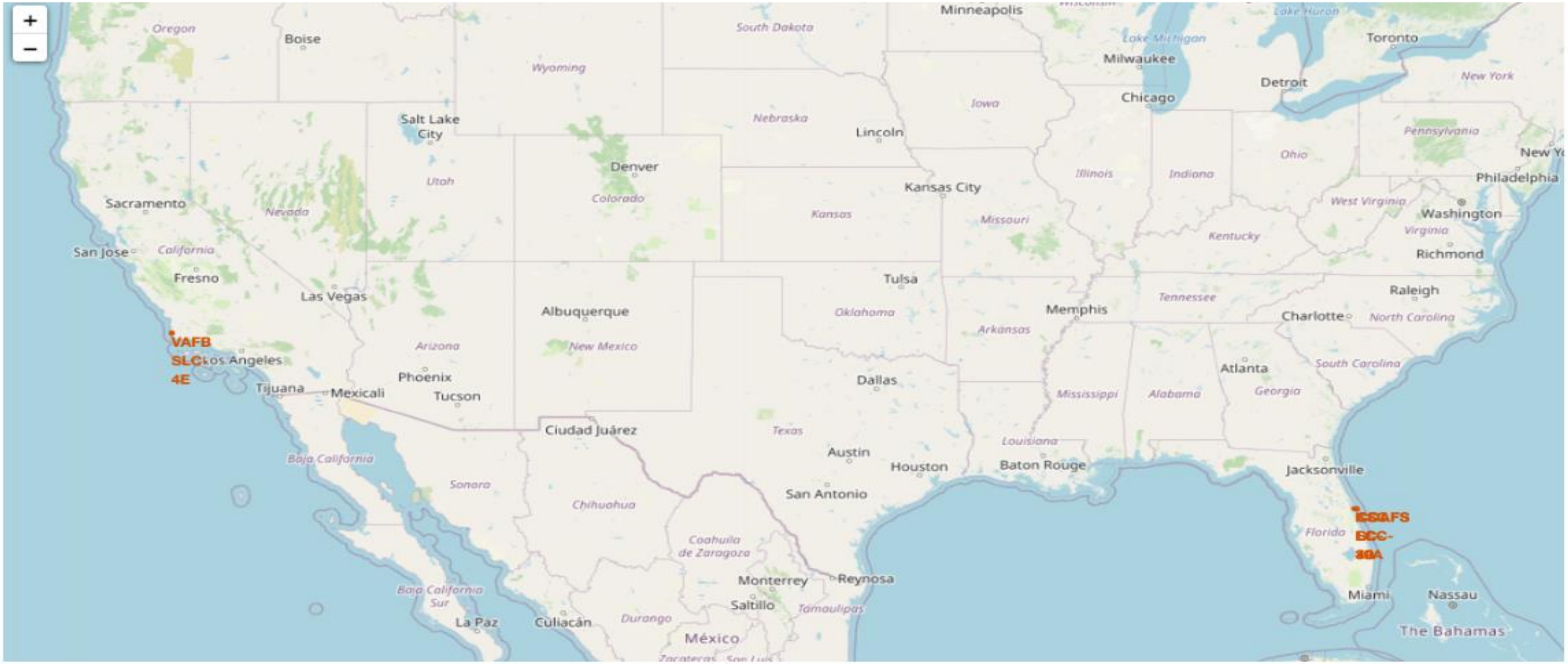
```
%sql select rank() over (order by count(*) DESC) Rank_, Landing_Outcome, count(*) No_of_Launch from :
```

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

Rank_	Landing_Outcome	No_of_Launch
1	Success	38
2	No attempt	21
3	Success (drone ship)	14
4	Success (ground pad)	9
5	Failure (drone ship)	5
5	Controlled (ocean)	5
7	Failure	3
8	Uncontrolled (ocean)	2
8	Failure (parachute)	2
10	Precluded (drone ship)	1
10	No attempt	1

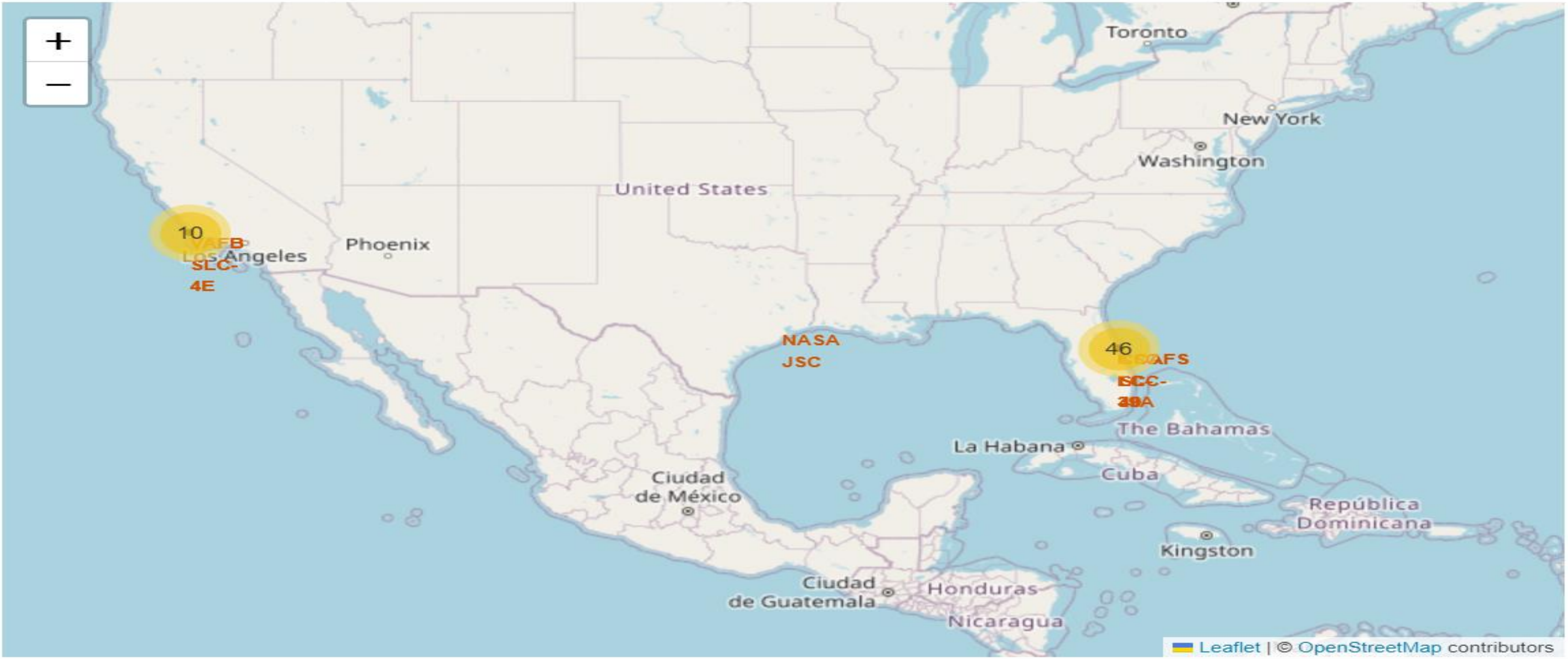
RESULTS – Interactive map with Folium

Mark all launch sites on a map



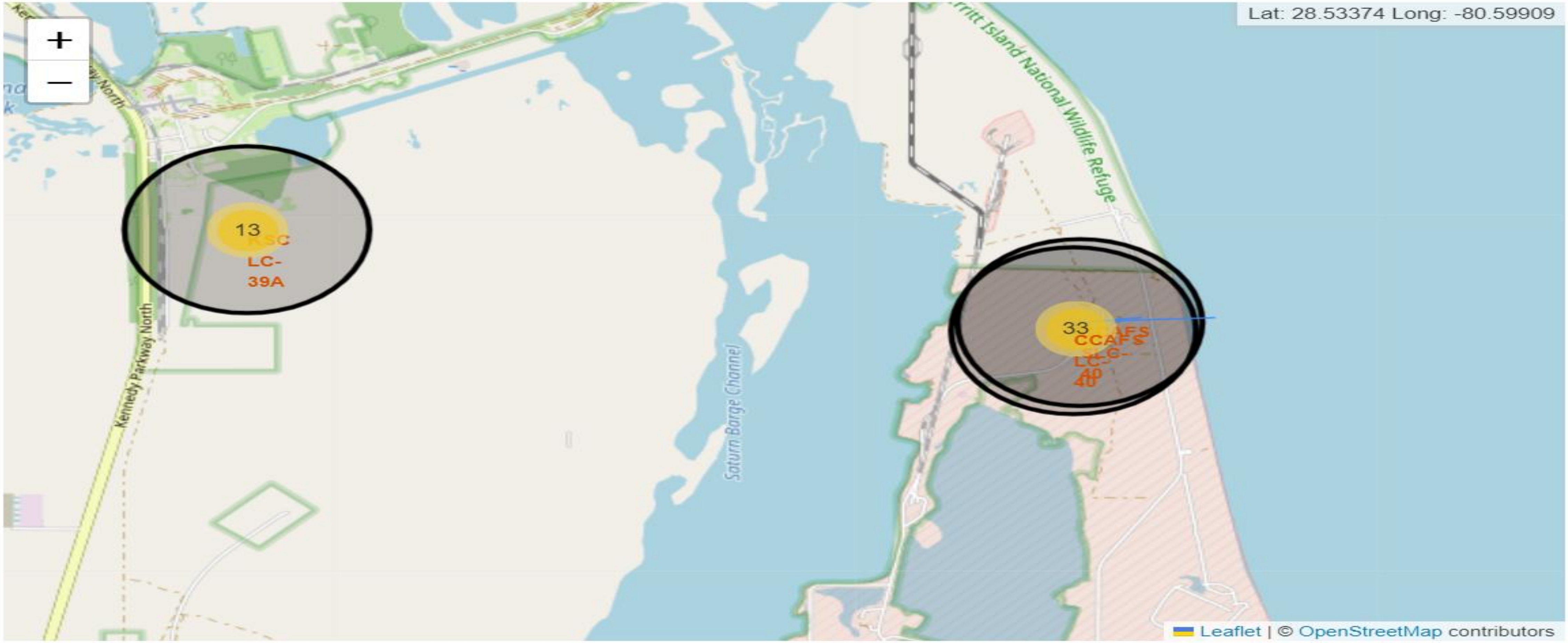
RESULTS – Interactive map with Folium

# Task 2: Mark the success/failed launches for each site on the map



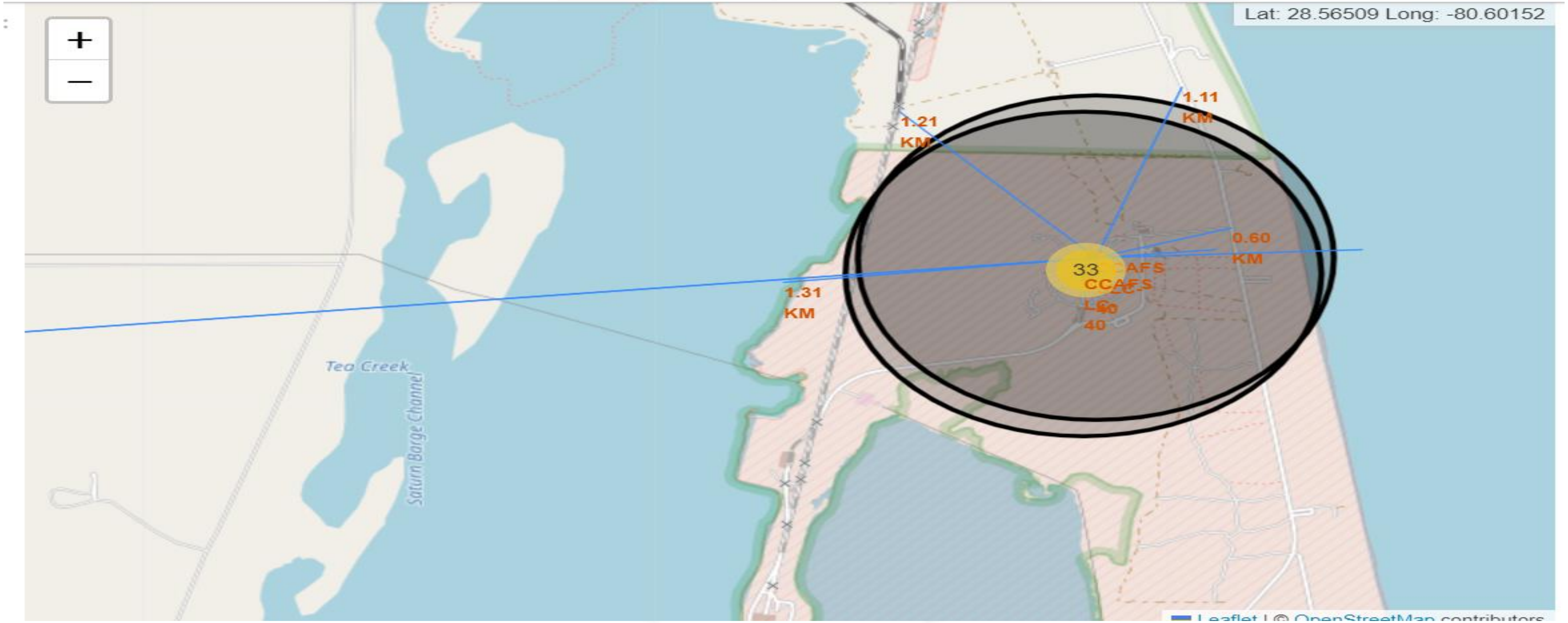
RESULTS – Interactive map with Folium

# TASK 3: Calculate the distances between a launch site to its nearest cost line


Lat: 28.53374 Long: -80.59909
+
—
rritt Island National Wildlife Refuge
Kennedy Parkway North
13
KSC
LC-
39A
Saturn Barge Channel
33
CCAFS
SLC-
40
CCAFS
LC-
40
Leaflet | © OpenStreetMap contributors

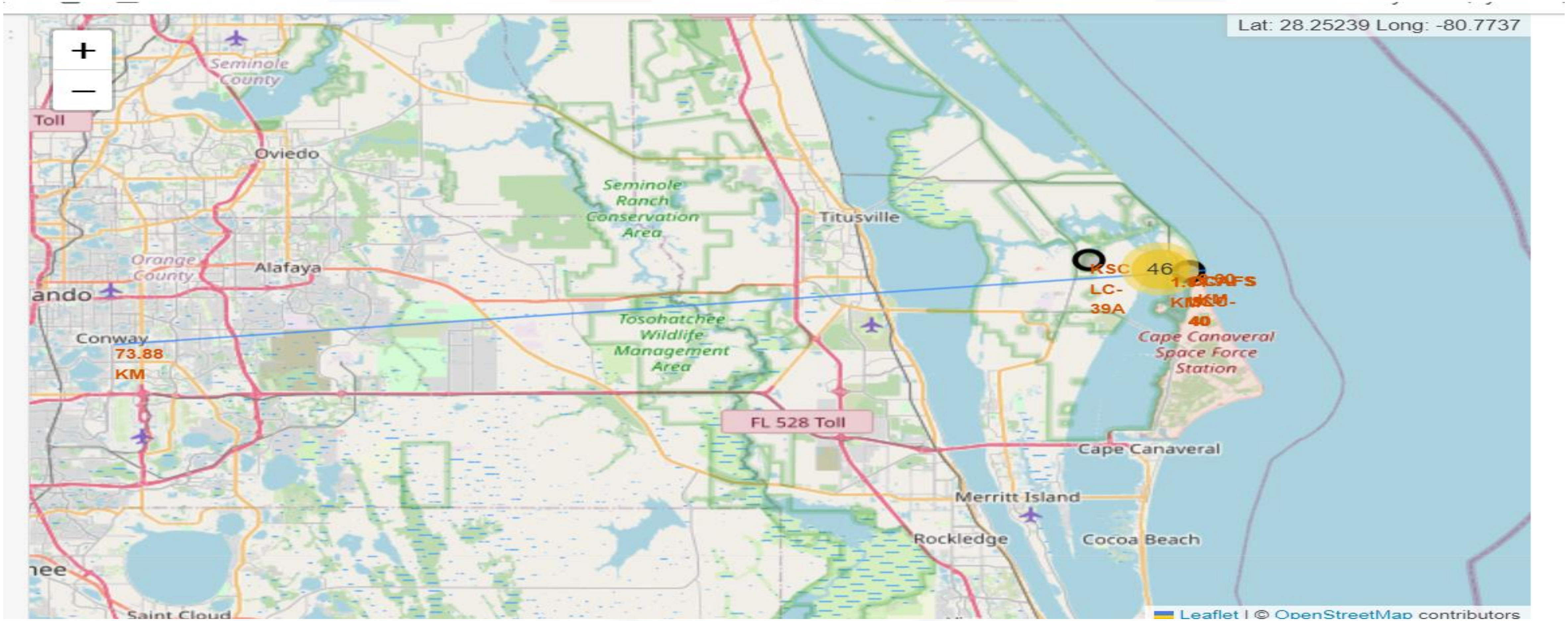
RESULTS – Interactive map with Folium

# TASK 3: Calculate the distances between a launch site to its nearest railroad and highway



RESULTS – Interactive map with Folium

# TASK 3: Calculate the distances between a launch site to its nearest city

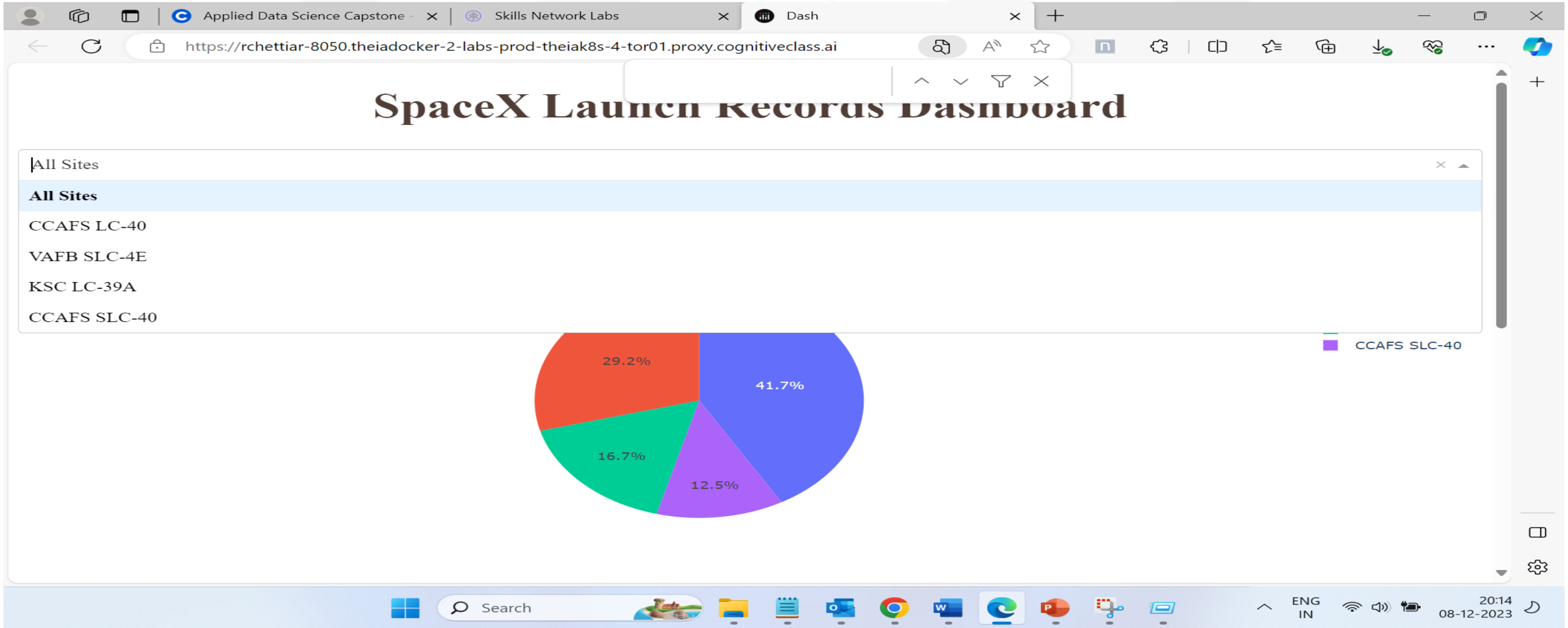


DASHBOARD

https://github.com/r-chettiar/coursera1/blob/main/SpaceX%20Dashboard%20All%20Sites.pdf

DASHBOARD using Plotly

Add a Launch Site Drop-down Input Component



https://rchettiar-8050.theiadocker-2-labs-prod-theiak8s-4-tor01.proxy.cognitiveclass.ai

SpaceX Launch Records Dashboard

All Sites

All Sites

CCAFS LC-40

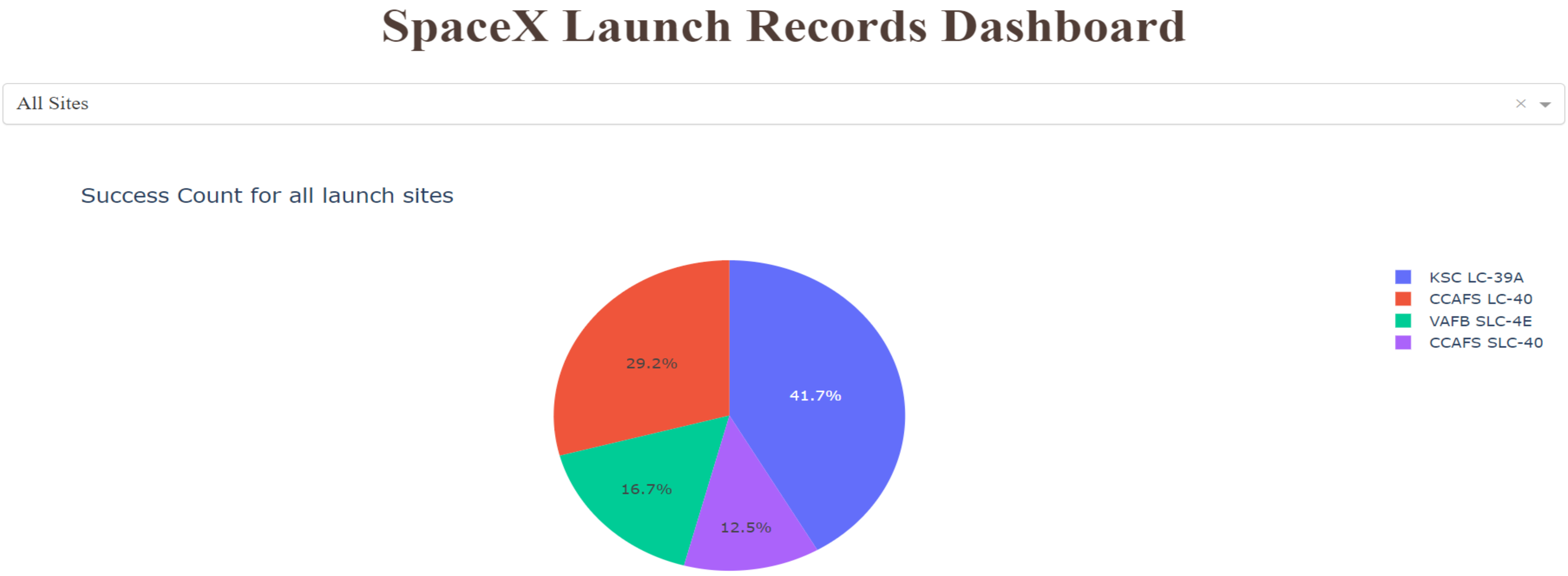
VAFB SLC-4E

KSC LC-39A

CCAFS SLC-40

DASHBOARD using Plotly

Add a callback function for `site-dropdown` as input, `success-pie-chart` as output



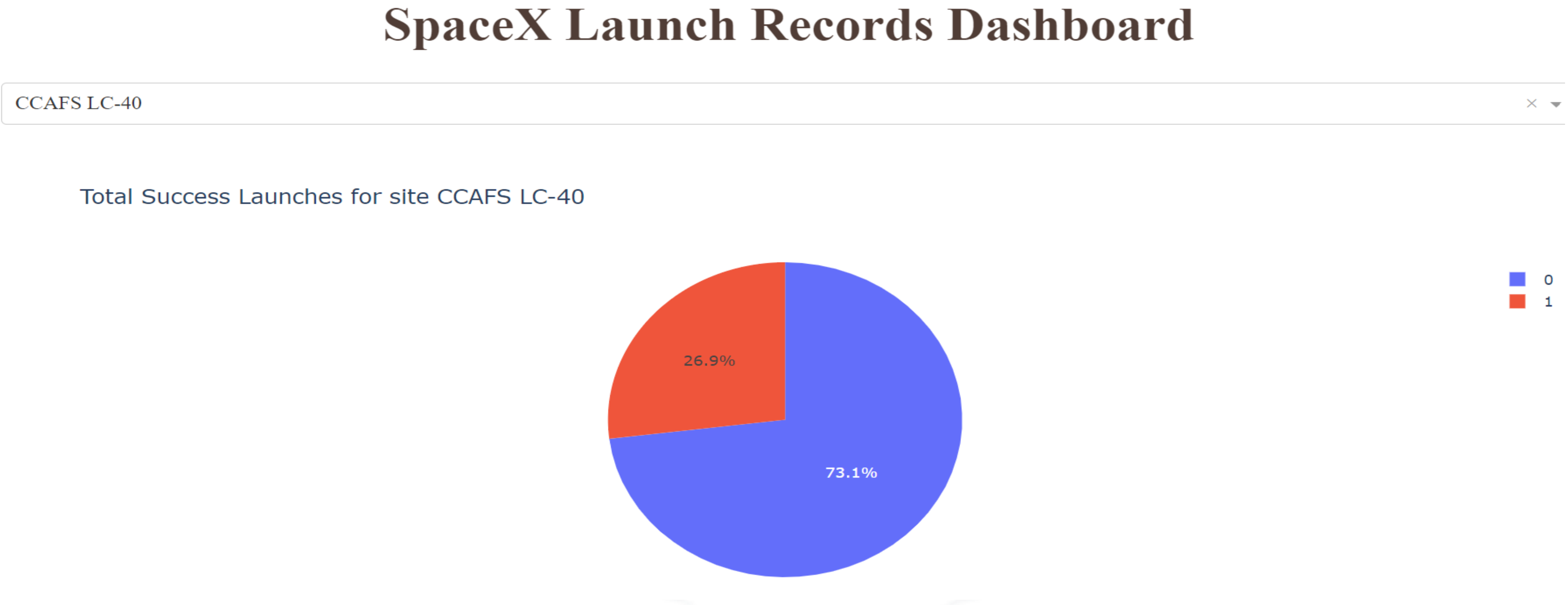
SpaceX Launch Records Dashboard

All Sites

Success Count for all launch sites

DASHBOARD using Plotly

Add a callback function for `site-dropdown` as input, `success-pie-chart` as output



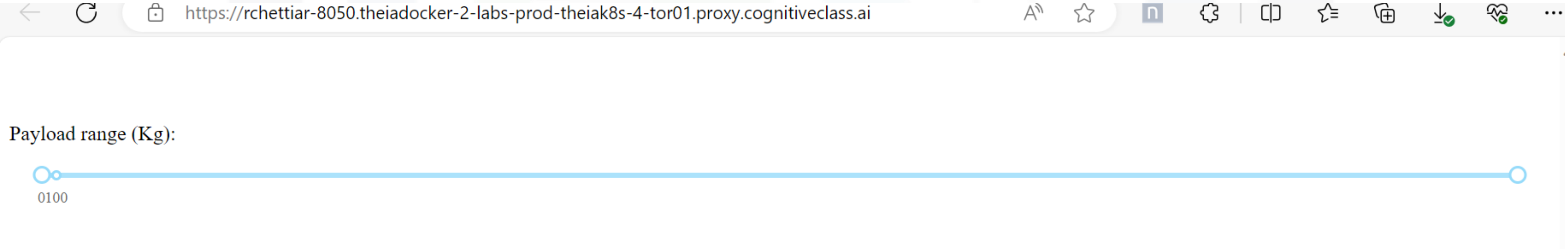
SpaceX Launch Records Dashboard

CCAFS LC-40

Total Success Launches for site CCAFS LC-40

DASHBOARD using Plotly

Add a slider to select payload range

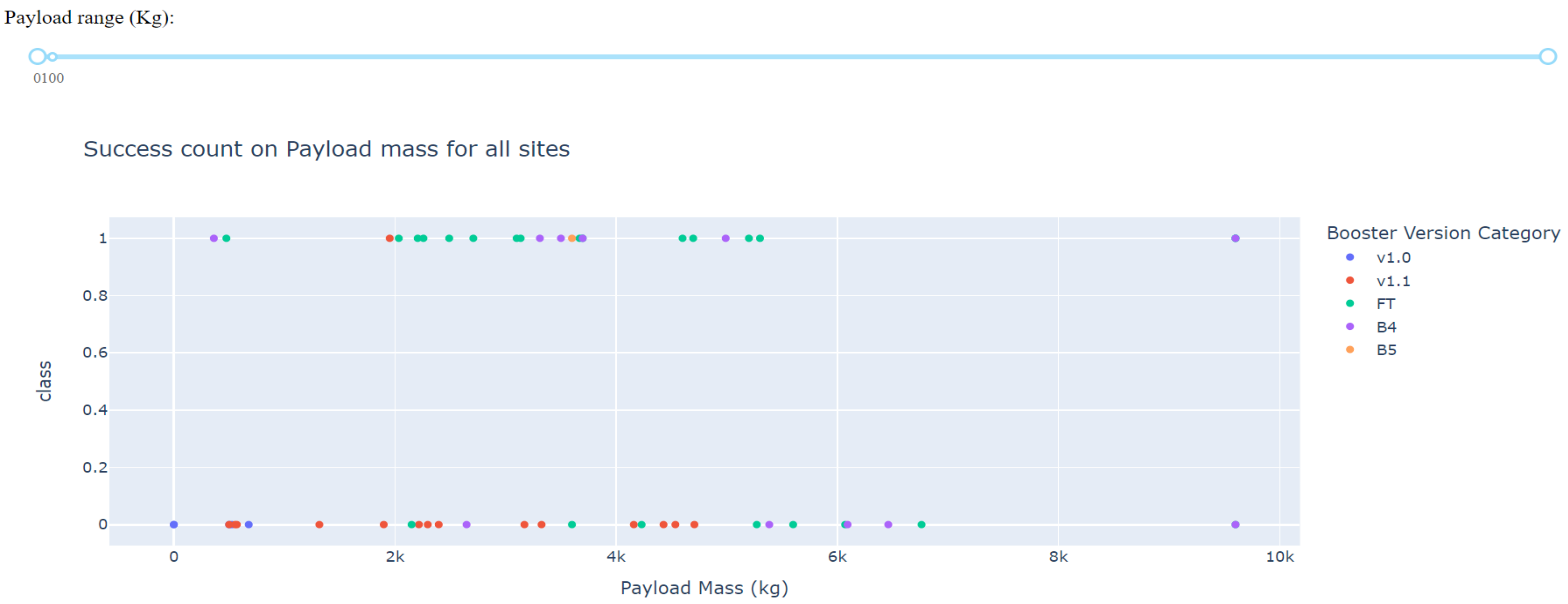


https://rchettiar-8050.theiadocker-2-labs-prod-theiak8s-4-tor01.proxy.cognitiveclass.ai

Payload range (Kg):

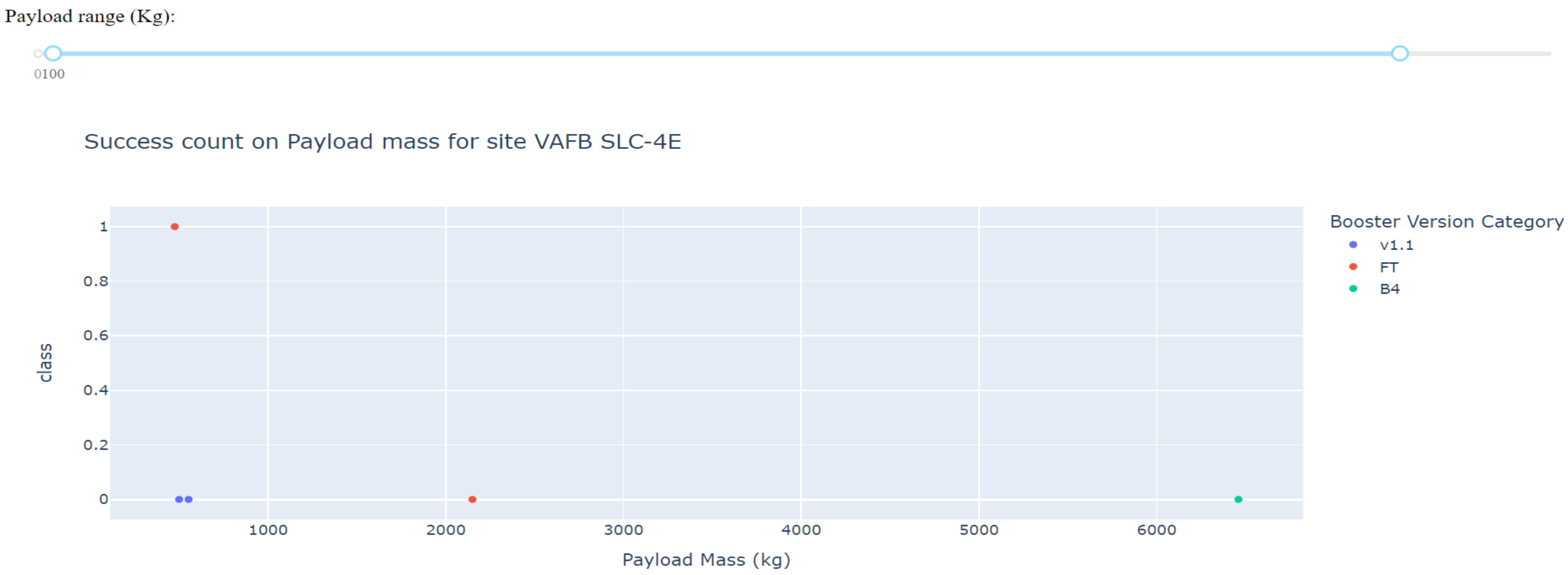
DASHBOARD using Plotly

Add a callback function for `site-dropdown` and `payload-slider` as inputs, `success-payload-scatter-chart` as output


Payload range (Kg):

DASHBOARD using Plotly

Add a callback function for `site-dropdown` and `payload-slider` as inputs, `success-payload-scatter-chart` as output



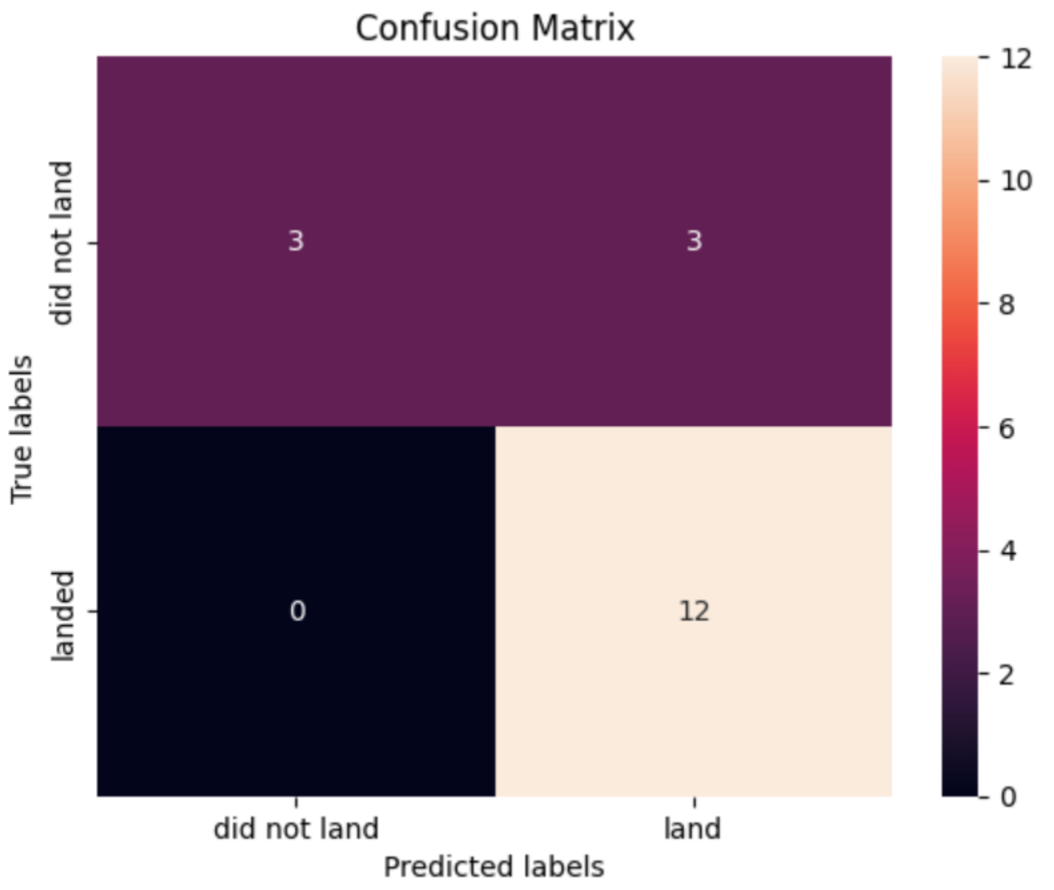
DISCUSSION

Predictive analysis – Logistics Regression

tuned hpyerparameters :(best parameters) {'C': 0.01, 'penalty': 'l2', 'solver': 'lbfgs'}

accuracy : 0.8464285714285713

Accuracy for Logistics Regression method on test data: 0.8333333333333334

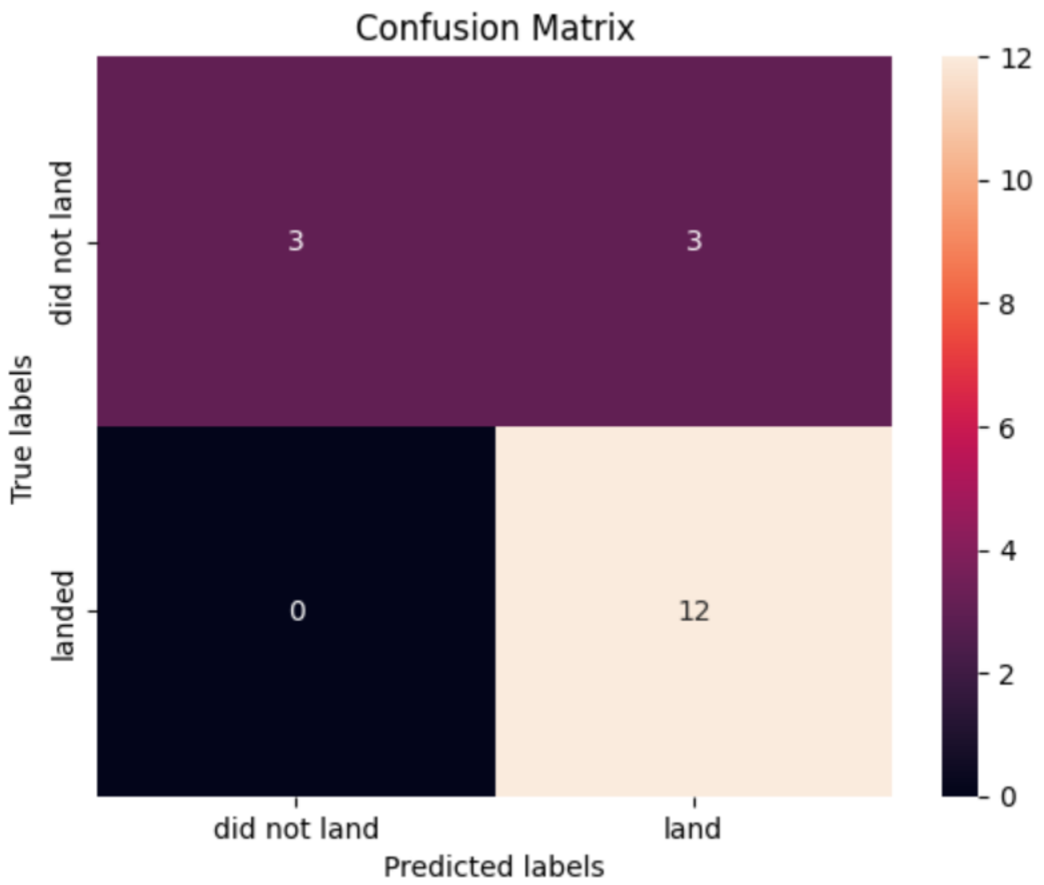


Predictive analysis – Support Vector Machine

tuned hpyerparameters :(best parameters) {'C': 0.01, 'penalty': 'l2', 'solver': 'lbfgs'}

accuracy : 0.8464285714285713

Accuracy for Support Vector Machine method on test data: 0.8333333333333334

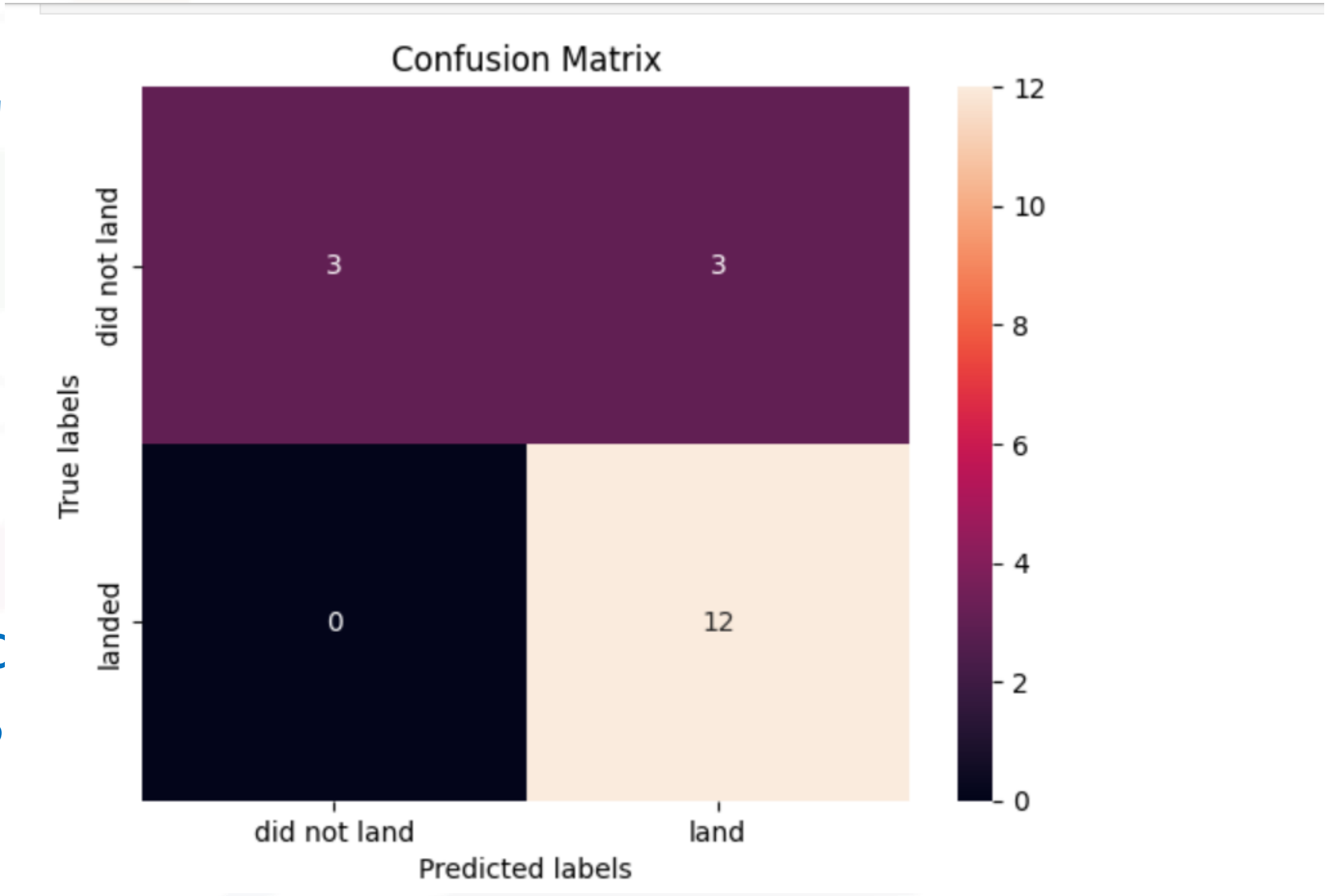


Predictive analysis – Decision Tree

tuned hpyerparameters :(best parameters) {'criterion': 'entropy' 'max_depth': 6, 'max_features': 'sqrt', 'min_samples_leaf': 4, 'min_samples_split': 2, 'splitter': 'random'}

accuracy : 0.875

Accuracy for decision tree methoc on test data: 0.833333333333333

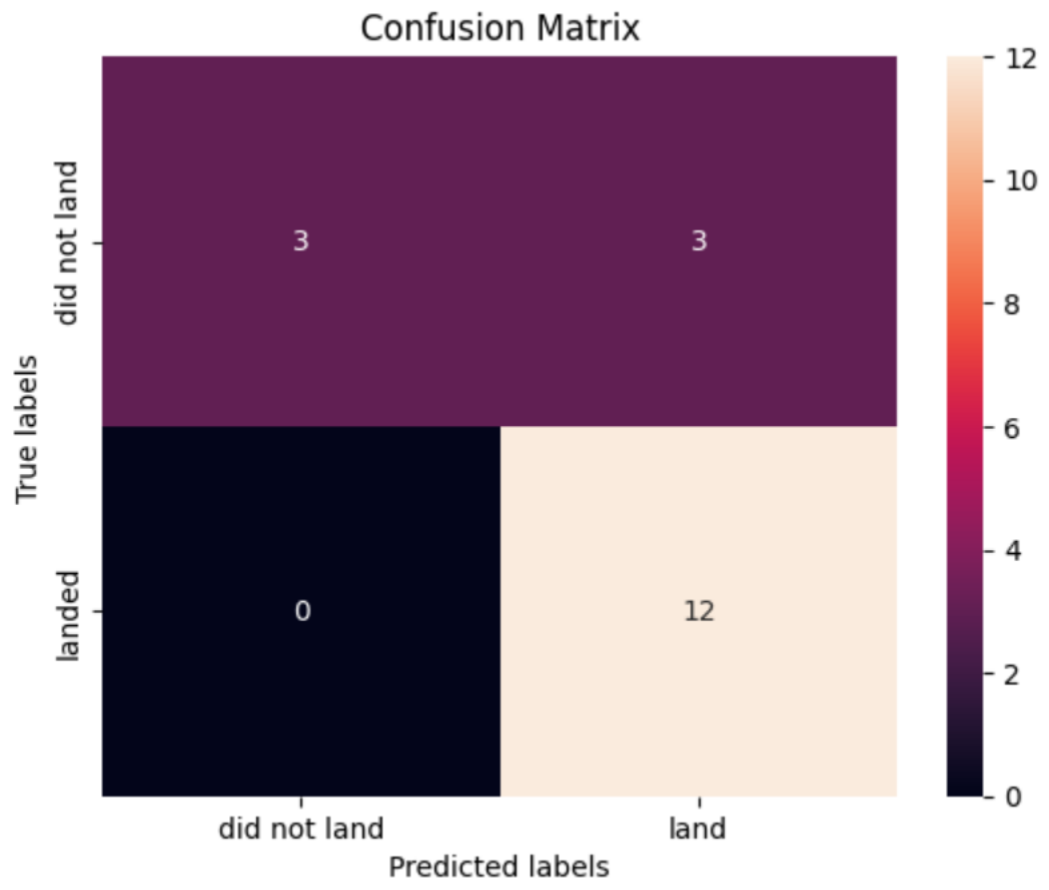


Predictive analysis – k nearest neighbors

tuned hpyerparameters :(best parameters) {'algorithm': 'auto', 'n_neighbors': 10, 'p': 1}

accuracy : 0.8482142857142858

Accuracy for k nearest neighbors method on test data: 0.8333333333333334



Predictive analysis (classification)

Accuracy for Support Vector Machine method: 0.8333333333333334

Accuracy for Decision tree method: 0.8333333333333334

Accuracy for K nearest neighbors' method: 0.8333333333333334

CONCLUSION

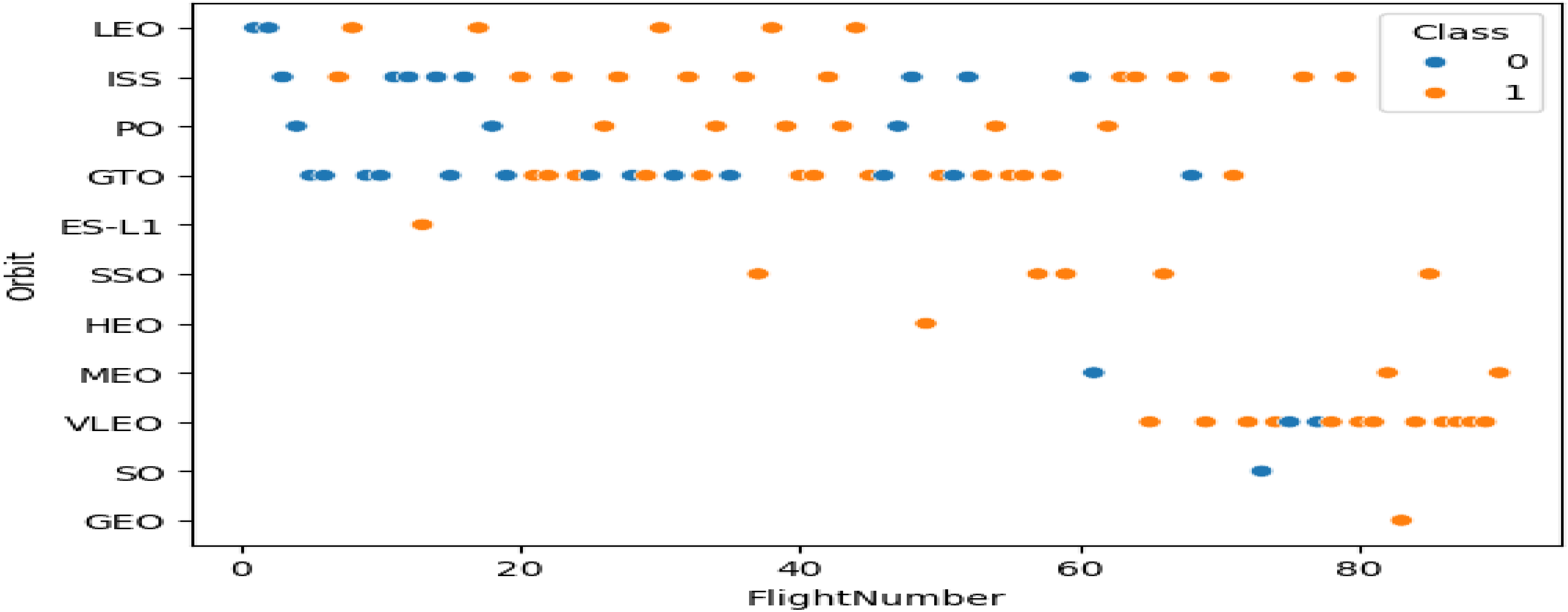
- In all the method, accuracy to predict successfully landing on the test sample was attained 83.33%
- The model has issue mainly in false positives

APPENDIX

- Include any relevant additional charts, or tables that you may have created during the analysis phase.

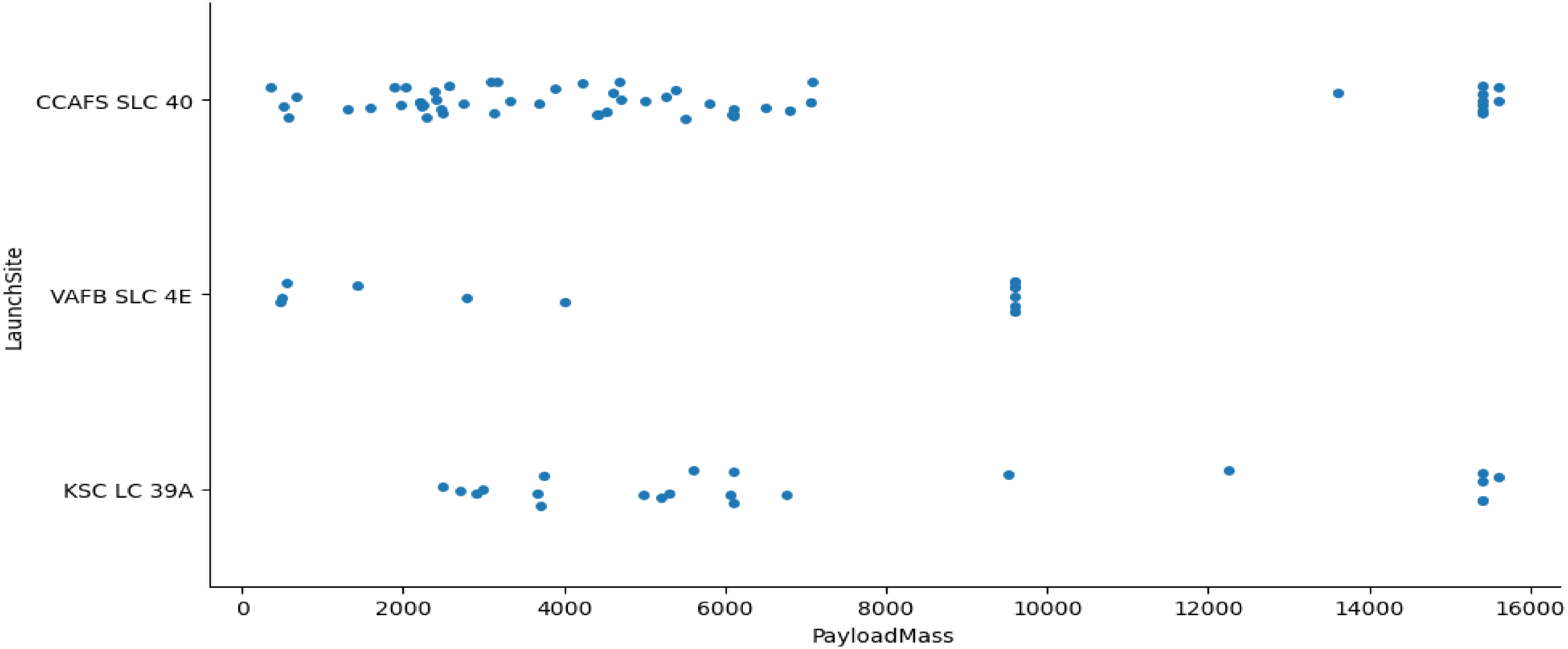
EDA Visualization

There is no relationship between flight number when in GTO orbit.

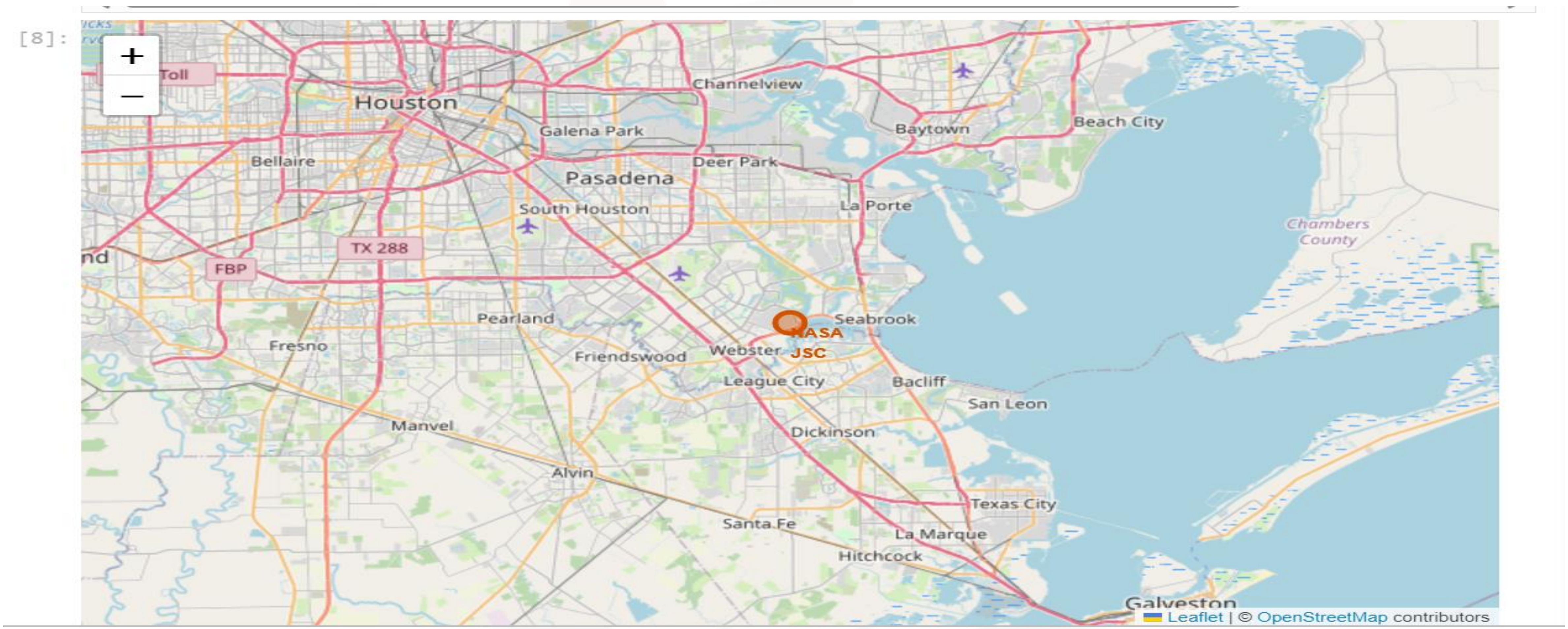


EDA Visualization

There is a relationship between launch site and pay load mass



Marked Nasa in map


[8]:
+
−
Toll
Houston
Channelview
Galena Park
Baytown
Beach City
Bellaire
Deer Park
Pasadena
South Houston
La Porte
Chambers County
TX 288
FBP
nd
Pearland
Seabrook
NASA
JSC
Fresno
Webster
Friendswood
League City
Bacliff
San Leon
Manvel
Dickinson
Alvin
Texas City
Santa Fe
La Marque
Hitchcock
Galveston
Leaflet | © OpenStreetMap contributors