APPLIED DATA SCIENCE CAPSTONE

Jorge Navarrete

2024 09 07

OBJECTIVES

- Excecutive Summary
- Introduction
- Methology
- Results
- Conclusion
- Appendix

EXCECUTIVE SUMMARY

- **Summary of methodologies**
  - **Data Collection through API**
  - **Data Collection with Web scraping**
  - **Sata Wrangling**
  - **Exploratory Data Analysis with SQL**
  - **Exploratory Data Analysis with Data Visualization**
  - **Interactive Visual Analytics with Folium**
  - **Machine Learning Prediction**

- **Summary of results**
  - **Exploratory Data Analysis result**
  - **Interactive analytics in screenshots**
  - **Predictive Analytics result from Machine Learning Lab**

INTRODUCTION

SpaceX: Revolutionizing Space Exploration

Founded in 2002 by entrepreneur Elon Musk, SpaceX (Space Exploration Technologies Corp.) is a private aerospace manufacturer and space transportation company that has revolutionized the space industry. With a bold mission to reduce the cost of space travel and make life multiplanetary, SpaceX has pushed the boundaries of innovation in both space exploration and technology.

The company became known for achieving several historic milestones, including being the first privately-funded company to launch, orbit, and recover a spacecraft (Dragon), and the first to dock a spacecraft with the International Space Station (ISS). In 2020, SpaceX made history again by sending astronauts to the ISS aboard the Crew Dragon, marking the first time a commercial company successfully launched humans into space.

SpaceX is also renowned for its development of reusable rocket technology with the Falcon 9 and Falcon Heavy rockets, significantly reducing the cost of space missions by reusing key components. In addition to launching satellites and cargo to space, SpaceX has ambitious plans to colonize Mars through its development of the Starship spacecraft, a fully reusable next-generation rocket designed for deep space missions.

METHODOLOGY

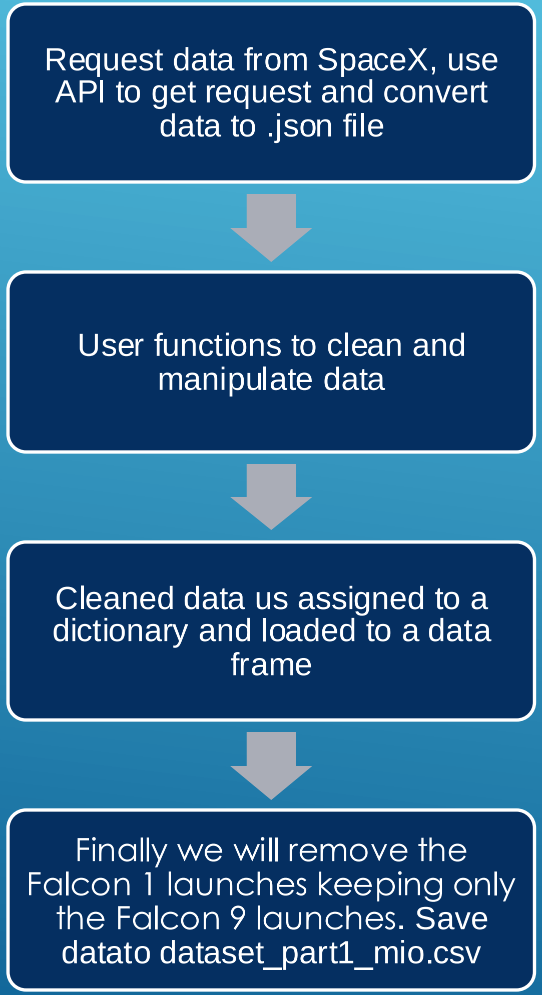
- Data Collection Methodology
  - Data was collected using SpaceX REST API and web scrapping from Wikipedia
- Perform data wrangling
  - Data was processed using one-hot encoding for categorical features
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

DATA COLLECTION

- Data collection is a process of gathering and measuring information on identified variables in a system or collected data. This allow us to asnwer relevant question and analyze outome data.

- REST API: By using get request , the resonse content as JSON can be decoded and loaded into a pandas dataframe. Then the data can be evaluated, cleaned and be used.
- Web SCRAPPING: "BeatifullSoup" will be used to extract launch records as HTML Table, then parse de table to load the data into a dataframe in order to be used and analized
- Data Wrangling: To perform some Exploratory Data Analysis (EDA) to find some patterns in the data and determine what would be the label for training supervised models.

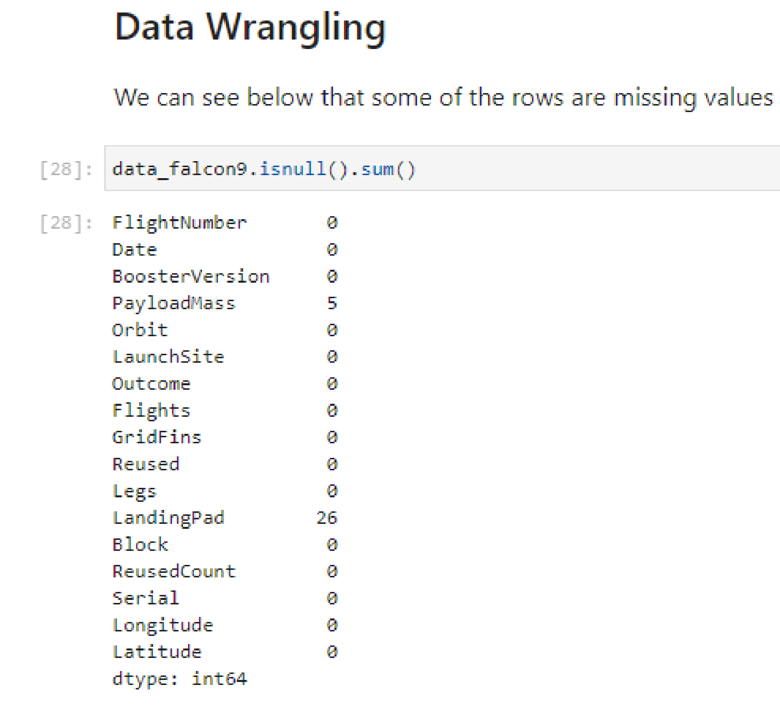
REST API


Request data from SpaceX, use API to get request and convert data to .json file

User functions to clean and manipulate data

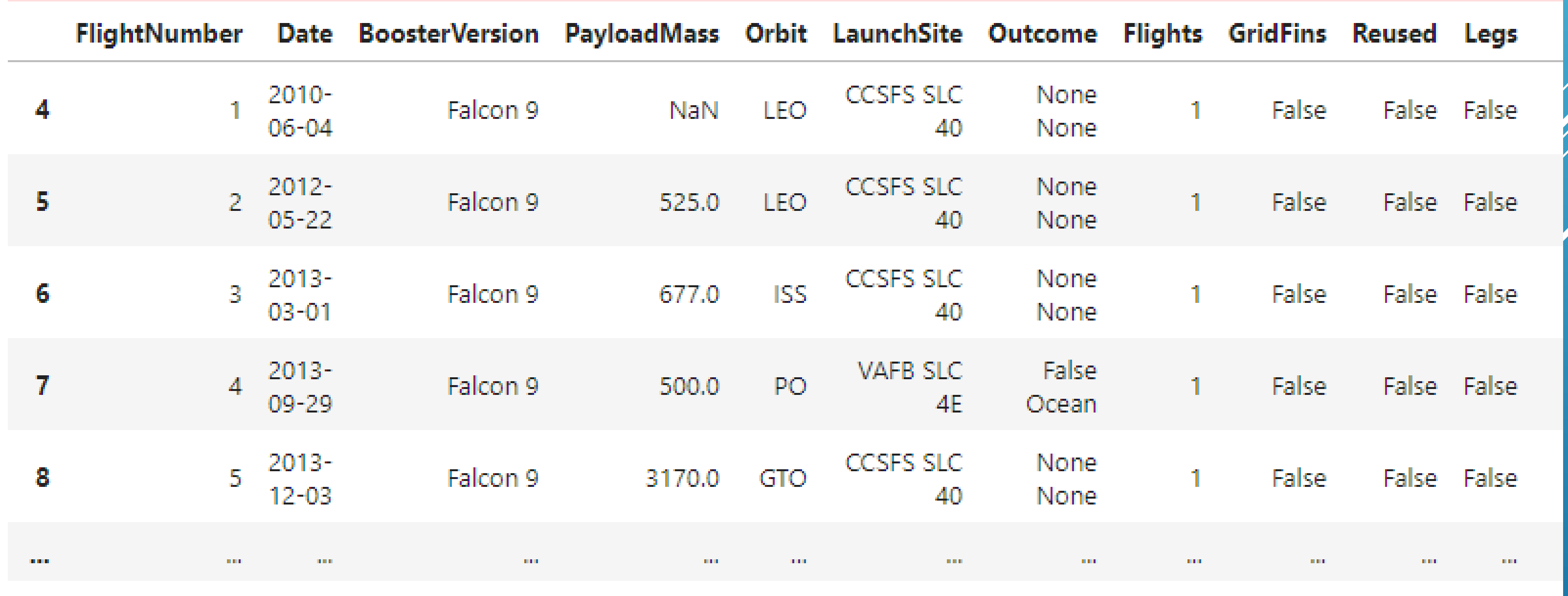
Cleaned data us assigned to a dictionary and loaded to a data frame

Finally we will remove the Falcon 1 launches keeping only the Falcon 9 launches. Save datato dataset_part1_mio.csv


Data Wrangling

We can see below that some of the rows are missing values

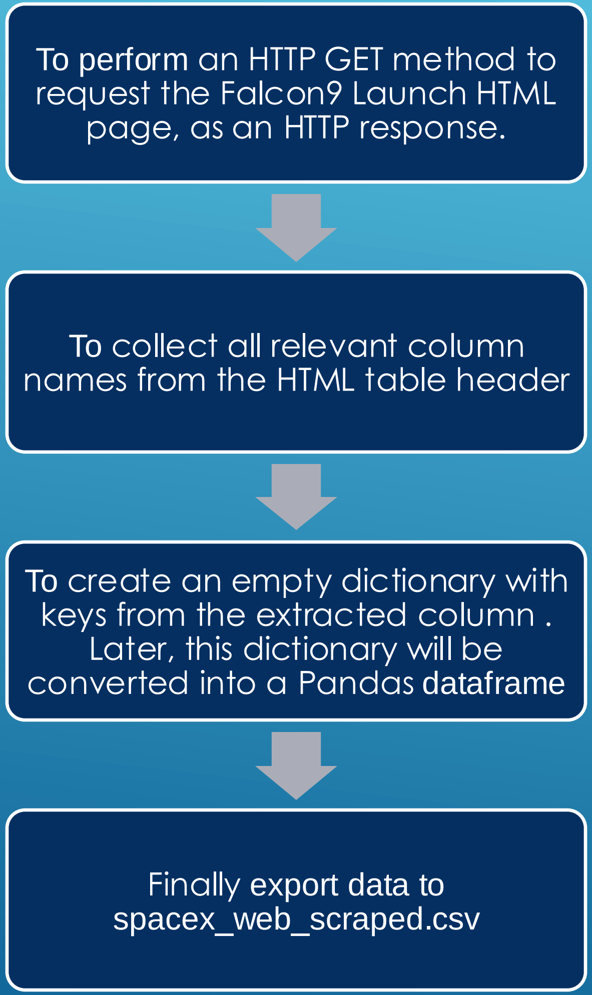
```
[28]: data_falcon9.isnull().sum()

[28]: FlightNumber       0
      Date               0
      BoosterVersion     0
      PayloadMass        5
      Orbit              0
      LaunchSite         0
      Outcome            0
      Flights            0
      GridFins           0
      Reused             0
      Legs               0
      LandingPad        26
      Block              0
      ReusedCount        0
      Serial             0
      Longitude          0
      Latitude           0
      dtype: int64
```



	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False
...	...	...	...	...	...	...	...	...	...	...	...

Github link: https://github.com/gioboto/Applied_Data_Science_Capstone/tree/a84e9f2b6b98450ed8ee383aedaf84c063e04c6b/module01/05-Hands-on%20Lab%20Complete%20the%20Data%20Collection%20API%20Lab

WEB SCRAPPING

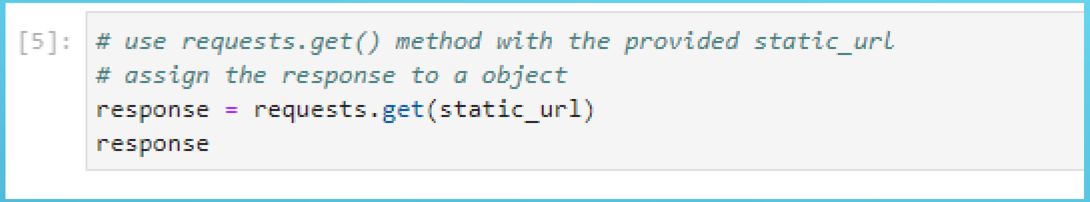


To perform an HTTP GET method to request the Falcon9 Launch HTML page, as an HTTP response.

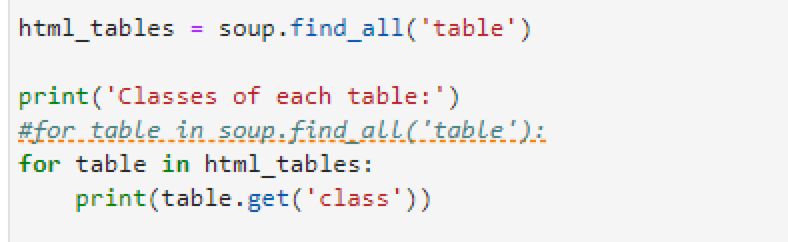
To collect all relevant column names from the HTML table header

To create an empty dictionary with keys from the extracted column . Later, this dictionary will be converted into a Pandas dataframe

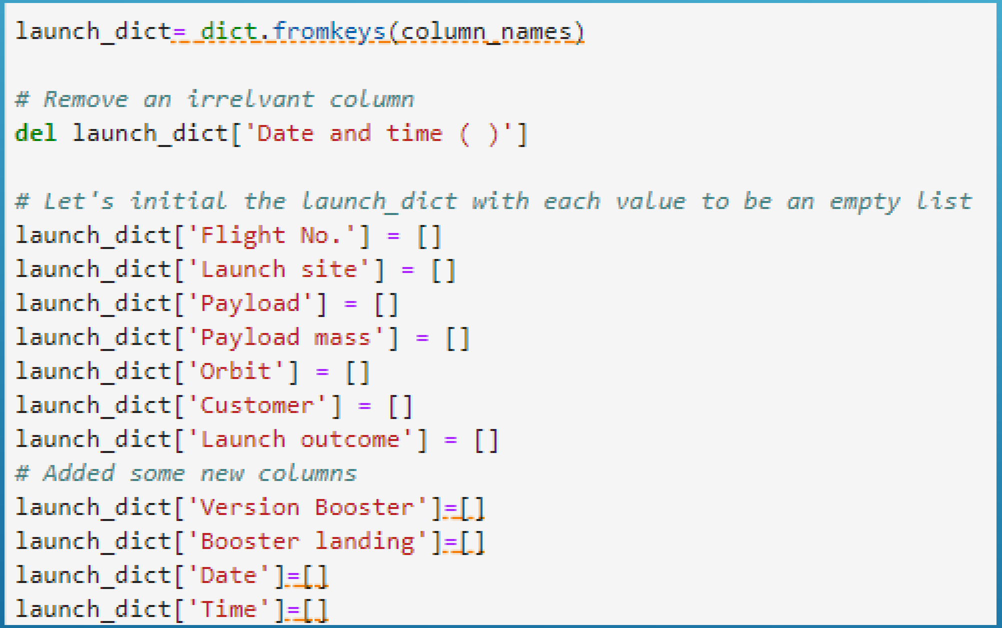
Finally export data to spacex_web_scraped.csv



```
[5]: # use requests.get() method with the provided static_url
     # assign the response to a object
     response = requests.get(static_url)
     response
```



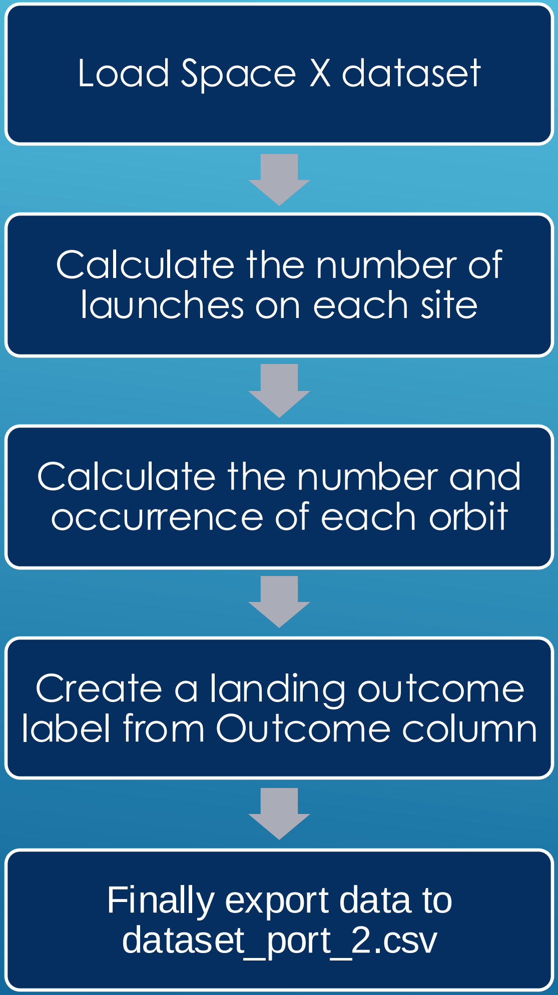
```
html_tables = soup.find_all('table')

print('Classes of each table:')
#for table in soup.find_all('table'):
for table in html_tables:
    print(table.get('class'))
```



```
launch_dict= dict.fromkeys(column_names)

# Remove an irrelvant column
del launch_dict['Date and time ( )']

# Let's initial the launch_dict with each value to be an empty list
launch_dict['Flight No.'] = []
launch_dict['Launch site'] = []
launch_dict['Payload'] = []
launch_dict['Payload mass'] = []
launch_dict['Orbit'] = []
launch_dict['Customer'] = []
launch_dict['Launch outcome'] = []
# Added some new columns
launch_dict['Version Booster']=[]
launch_dict['Booster landing']=[]
launch_dict['Date']=[]
launch_dict['Time']=[]
```

Github link: https://github.com/gioboto/Applied_Data_Science_Capstone/tree/a84e9f2b6b98450ed8ee383aedaf84c063e04c6b/module01/06-Hands-on%20Lab%20Complete%20the%20Data%20Collection%20with%20Web%20Scraping%20lab

DATA WRANGLING

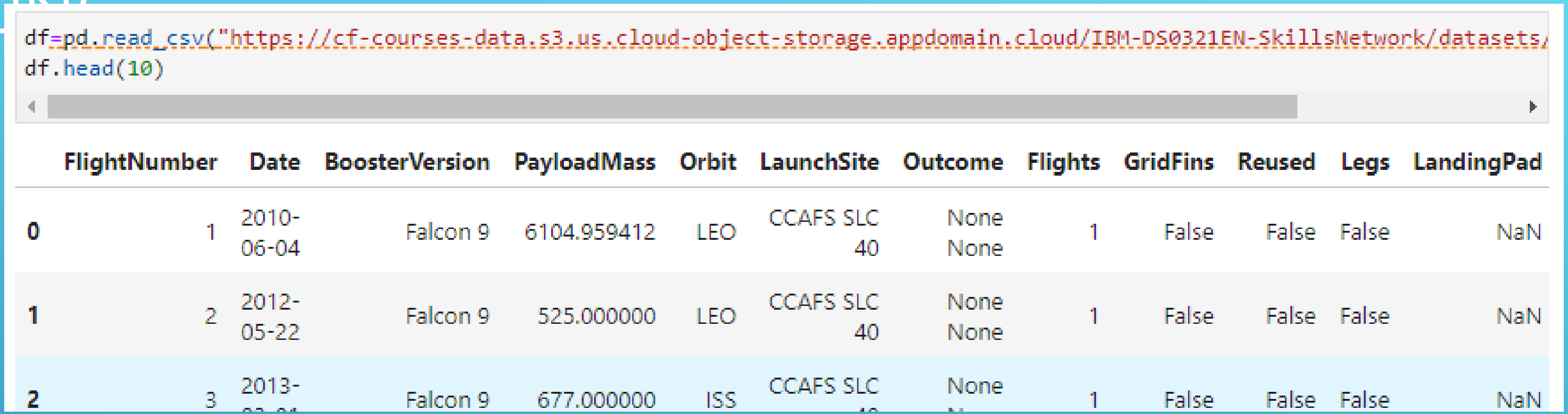

Load Space X dataset

Calculate the number of launches on each site

Calculate the number and occurrence of each orbit

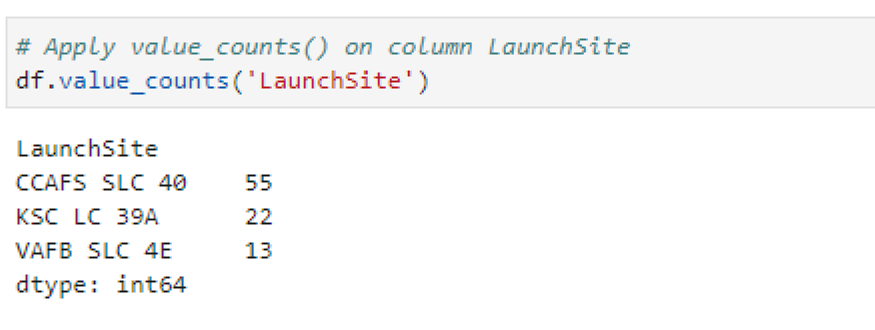
Create a landing outcome label from Outcome column

Finally export data to dataset_port_2.csv



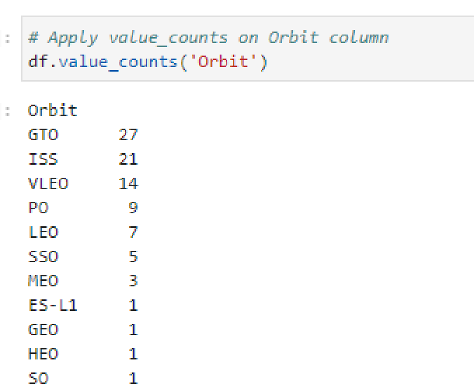
```
df=pd.read_csv("https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/
df.head(10)
```

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN
2	3	2013-	Falcon 9	677.000000	ISS	CCAFS SLC	None	1	False	False	False	NaN



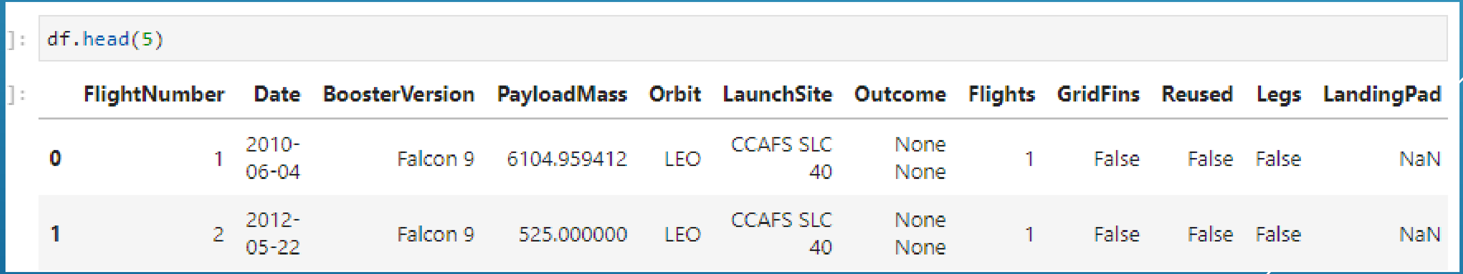
```
# Apply value_counts() on column LaunchSite
df.value_counts('LaunchSite')
```

```
LaunchSite
CCAFS SLC 40    55
KSC LC 39A      22
VAFB SLC 4E     13
dtype: int64
```



```
# Apply value_counts on Orbit column
df.value_counts('Orbit')
```

```
Orbit
GTO      27
ISS      21
VLEO     14
PO        9
LEO       7
SSO       5
MEO       3
ES-L1     1
GEO       1
HEO       1
SO        1
```



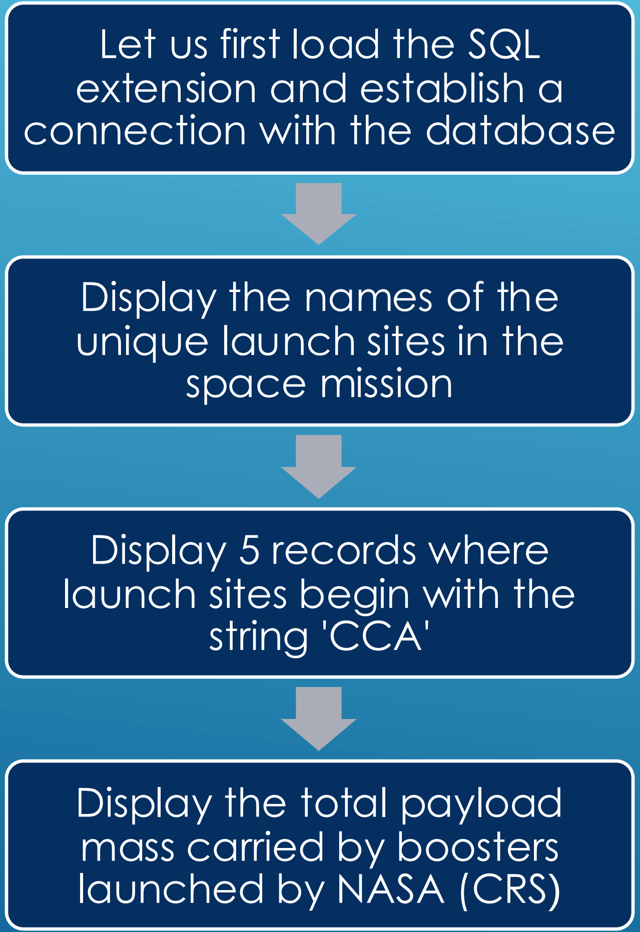
```
df.head(5)
```

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN

Github link: https://github.com/gioboto/Applied_Data_Science_Capstone/tree/a84e9f2b6b98450ed8ee383aedaf84c063e04c6b/module01/09-Data%20Wrangling%20Overview

EDA WITH VISUALIZATION LAB

Understand the Spacex DataSet

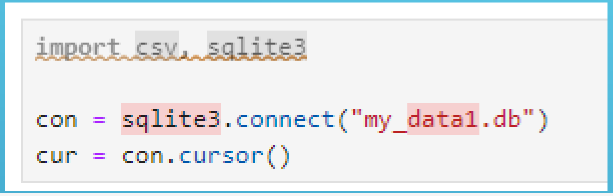


Let us first load the SQL extension and establish a connection with the database

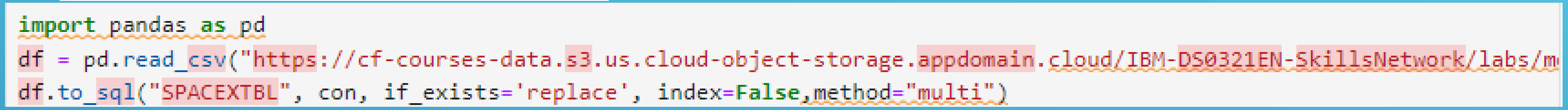
Display the names of the unique launch sites in the space mission

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

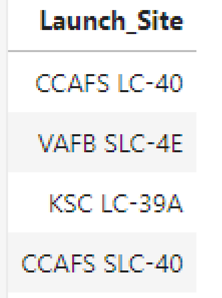
Display the total payload mass carried by boosters launched by NASA (CRS)



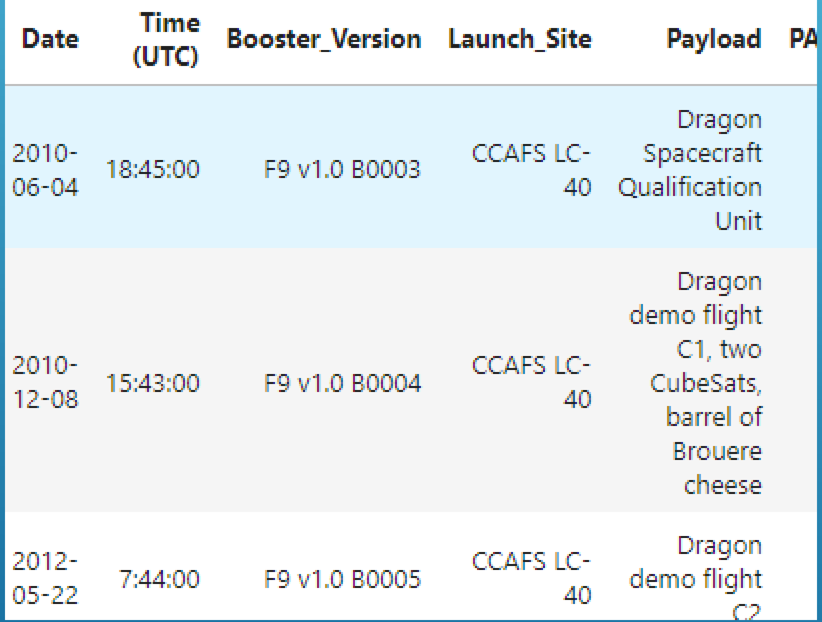
```
import csv, sqlite3

con = sqlite3.connect("my_data1.db")
cur = con.cursor()
```



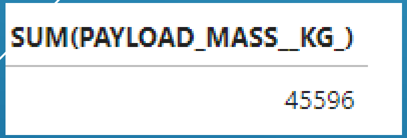
```
import pandas as pd
df = pd.read_csv("https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/labs/m
df.to_sql("SPACEXTBL", con, if_exists='replace', index=False,method="multi")
```



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



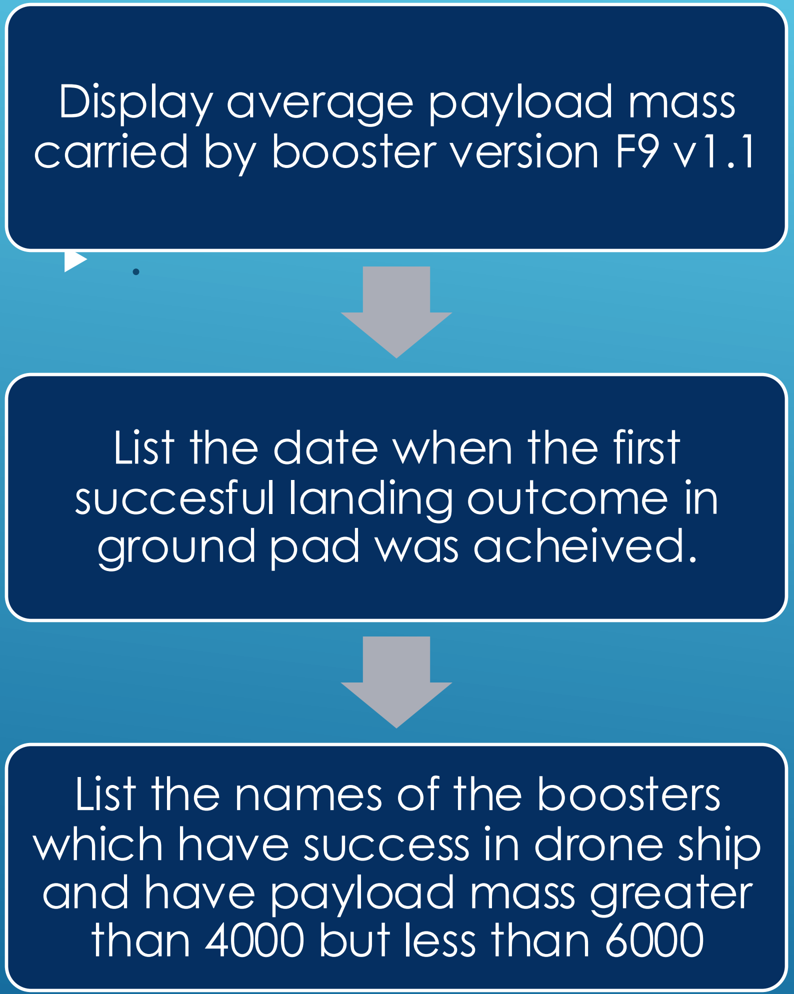
Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PA
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	



SUM(PAYLOAD_MASS__KG_)
45596

Github link: https://github.com/gioboto/Applied_Data_Science_Capstone/tree/a84e9f2b6b98450ed8ee383aedaf84c063e04c6b/module02/02-Hands-on%20Lab%20Complete%20the%20EDA%20with%20SQL

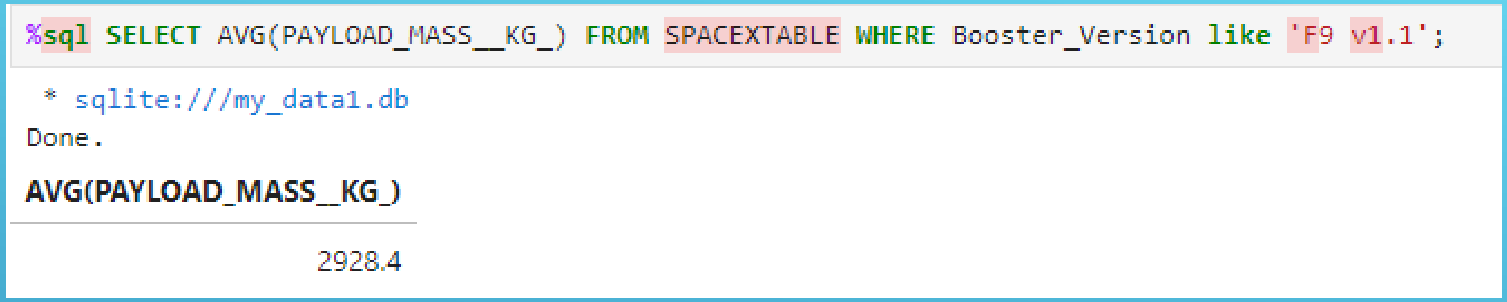
EDA WITH VISUALIZATION LAB



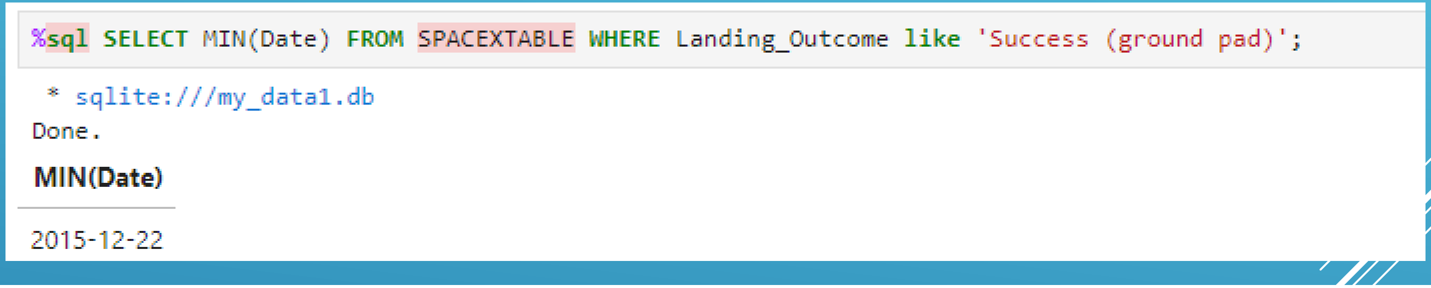
Display average payload mass carried by booster version F9 v1.1

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

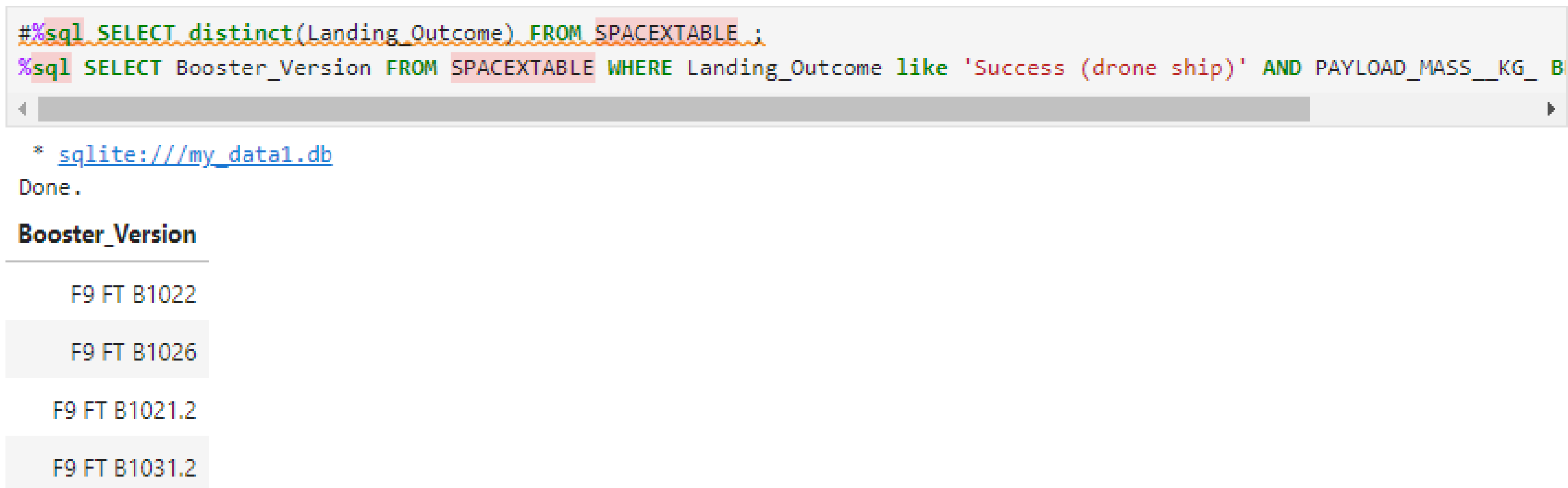
List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000



```
%sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE Booster_Version like 'F9 v1.1';

 * sqlite:///my_data1.db
Done.
AVG(PAYLOAD_MASS__KG_)
                2928.4
```



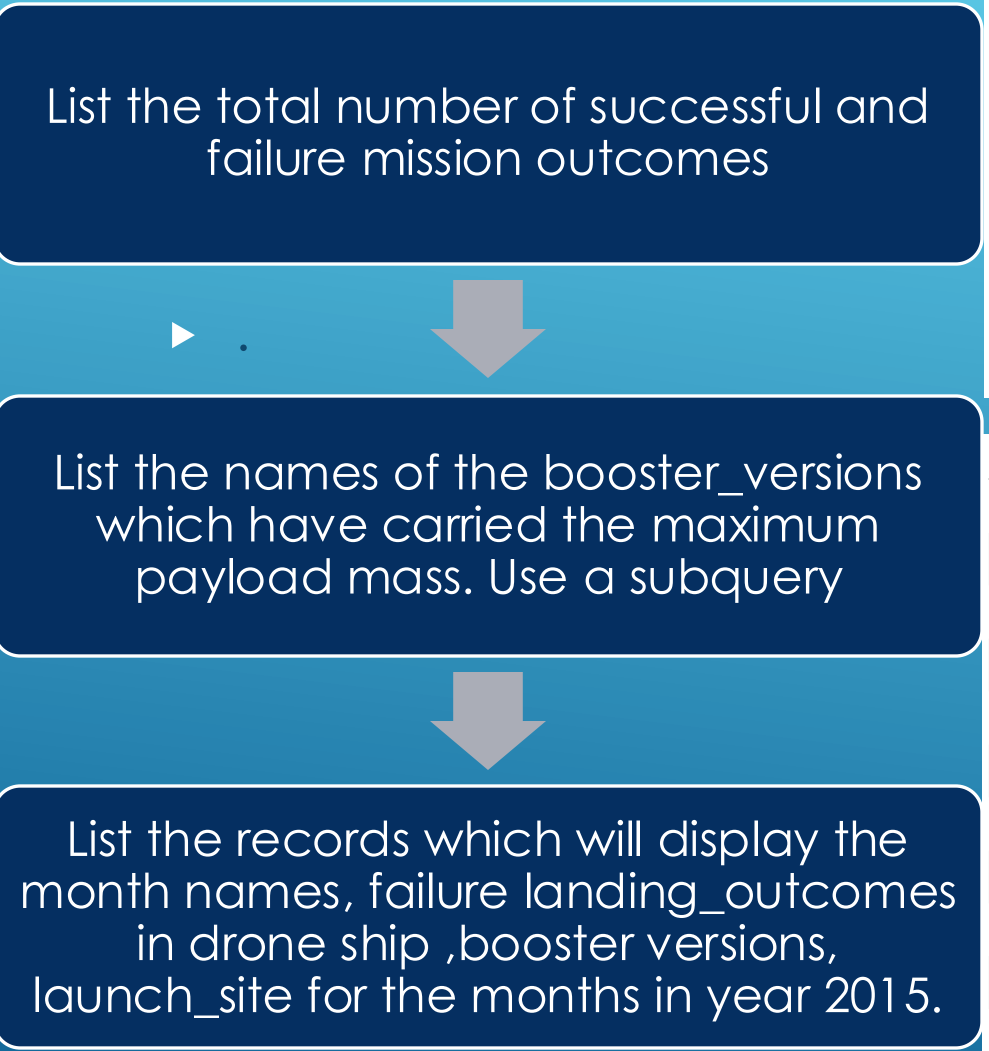
```
%sql SELECT MIN(Date) FROM SPACEXTABLE WHERE Landing_Outcome like 'Success (ground pad)';

 * sqlite:///my_data1.db
Done.
MIN(Date)
2015-12-22
```



```
#%sql SELECT distinct(Landing_Outcome) FROM SPACEXTABLE ;
%sql SELECT Booster_Version FROM SPACEXTABLE WHERE Landing_Outcome like 'Success (drone ship)' AND PAYLOAD_MASS__KG_ B

 * sqlite:///my_data1.db
Done.
Booster_Version
   F9 FT B1022
   F9 FT B1026
 F9 FT B1021.2
 F9 FT B1031.2
```

Github link: https://github.com/gioboto/Applied_Data_Science_Capstone/tree/a84e9f2b6b98450ed8ee383aedaf84c063e04c6b/module02/02-Hands-on%20Lab%20Complete%20the%20EDA%20with%20SQL

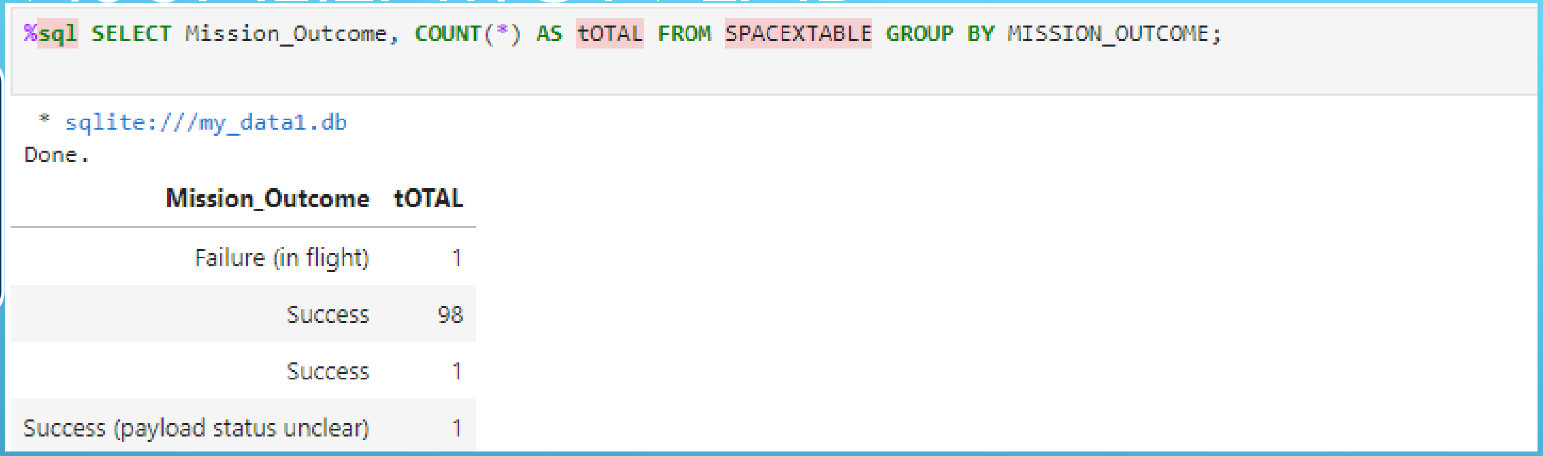
EDA WITH VISUALIZATION LAB



List the total number of successful and failure mission outcomes

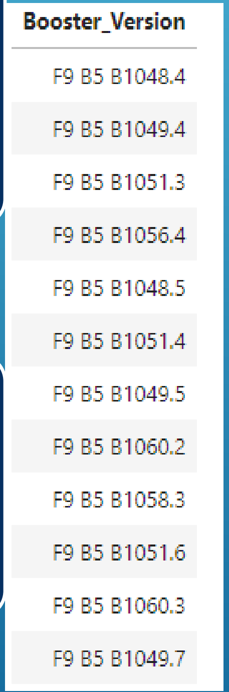
List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

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.

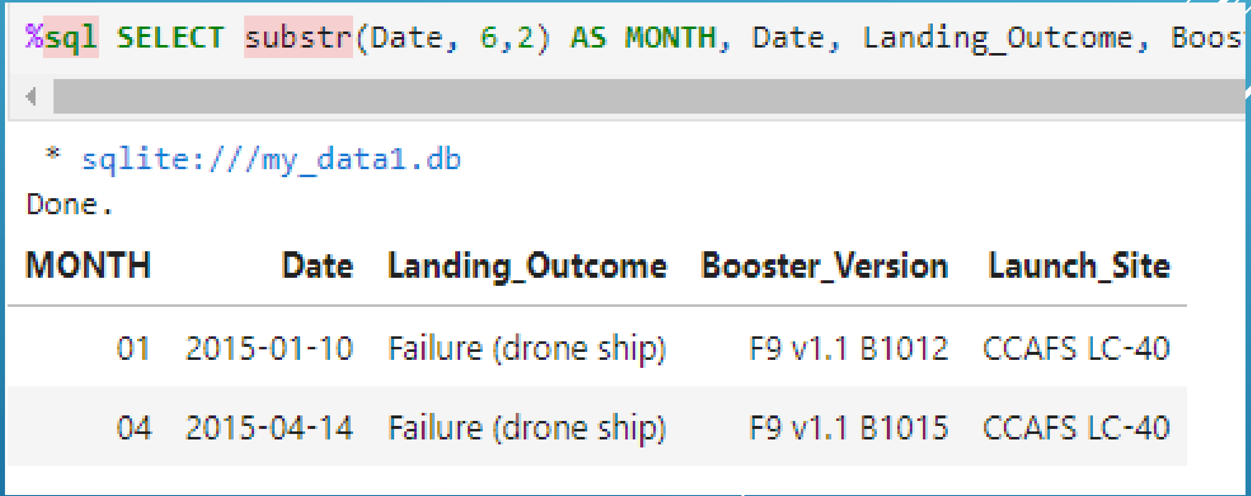


```
%sql SELECT Mission_Outcome, COUNT(*) AS tOTAL FROM SPACEXTABLE GROUP BY MISSION_OUTCOME;

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

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



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

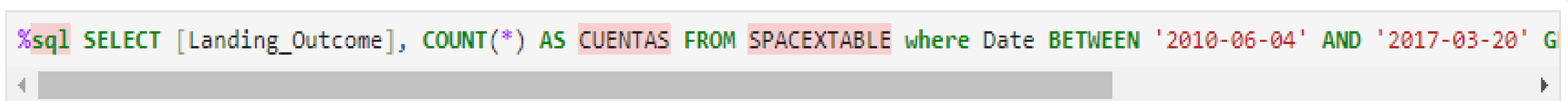


```
%sql SELECT substr(Date, 6,2) AS MONTH, Date, Landing_Outcome, Boos

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

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

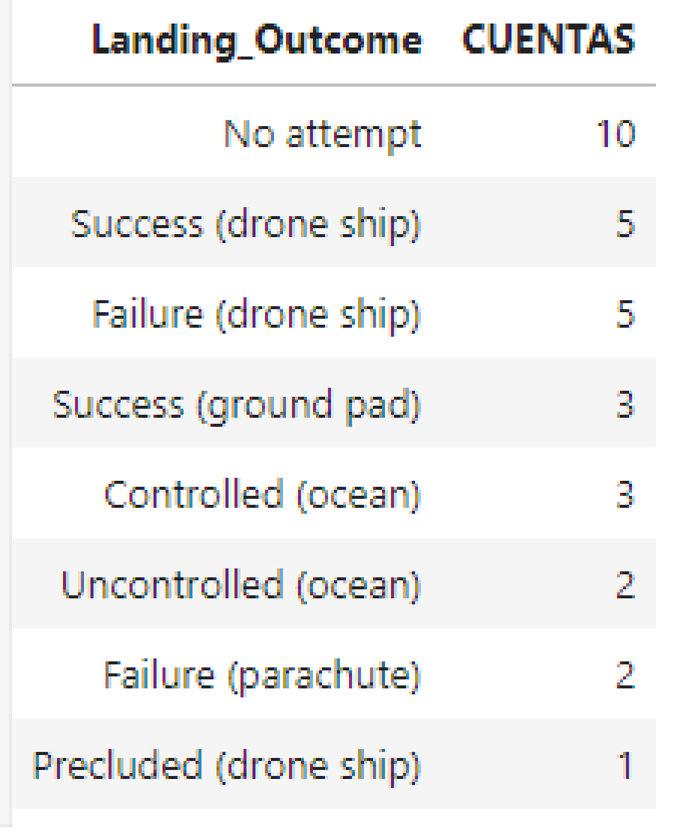
Github link: https://github.com/gioboto/Applied_Data_Science_Capstone/tree/a84e9f2b6b98450ed8ee383aedaf84c063e04c6b/module02/02-Hands-on%20Lab%20Complete%20the%20EDA%20with%20SQL

EDA WITH VISUALIZATION LAB



```
%sql SELECT [Landing_Outcome], COUNT(*) AS CUENTAS FROM SPACEXTABLE where Date BETWEEN '2010-06-04' AND '2017-03-20' G
```

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.



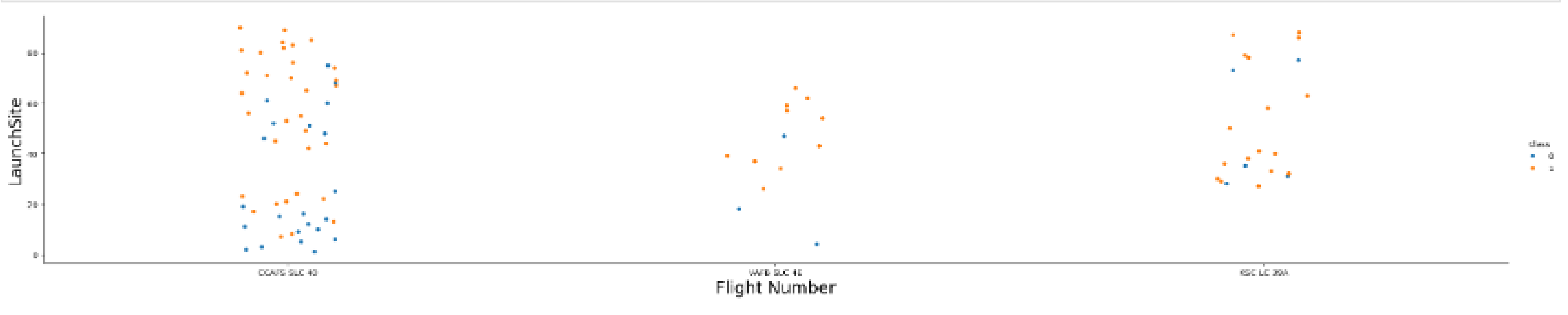
Landing_Outcome	CUENTAS
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Github link: https://github.com/gioboto/Applied_Data_Science_Capstone/tree/a84e9f2b6b98450ed8ee383aedaf84c063e04c6b/module02/02-Hands-on%20Lab%20Complete%20the%20EDA%20with%20SQL

EDA WITH VISUALIZATION LAB

1

- **Read** the SpaceX dataset into a Pandas dataframe and print its summary
- Visualize the relationship between Flight Number and Launch Site

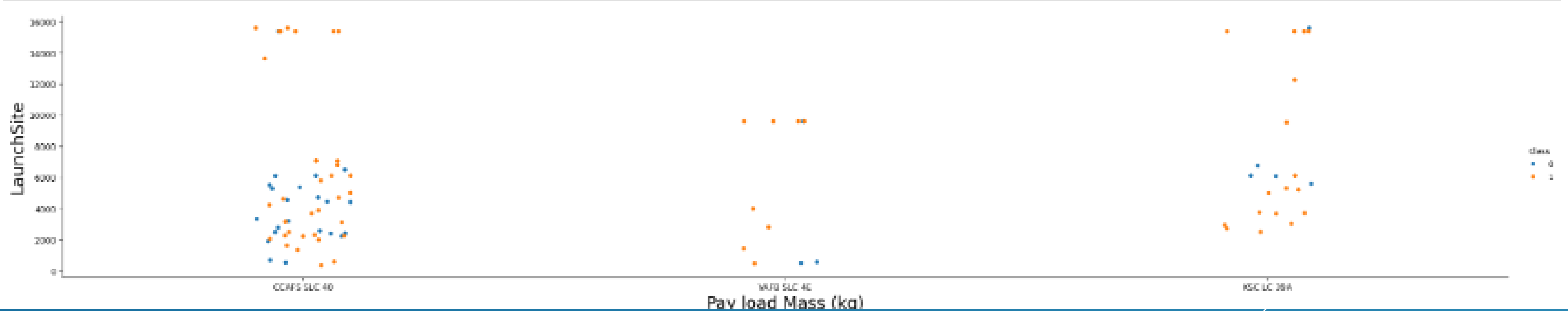


Github link: https://github.com/gioboto/Applied_Data_Science_Capstone/tree/a84e9f2b6b98450ed8ee383aedaf84c063e04c6b/module02/05-EDA%20with%20Visualization%20Lab

EDA WITH VISUALIZATION LAB

2

- Visualize the relationship between Payload Mass and Launch Site

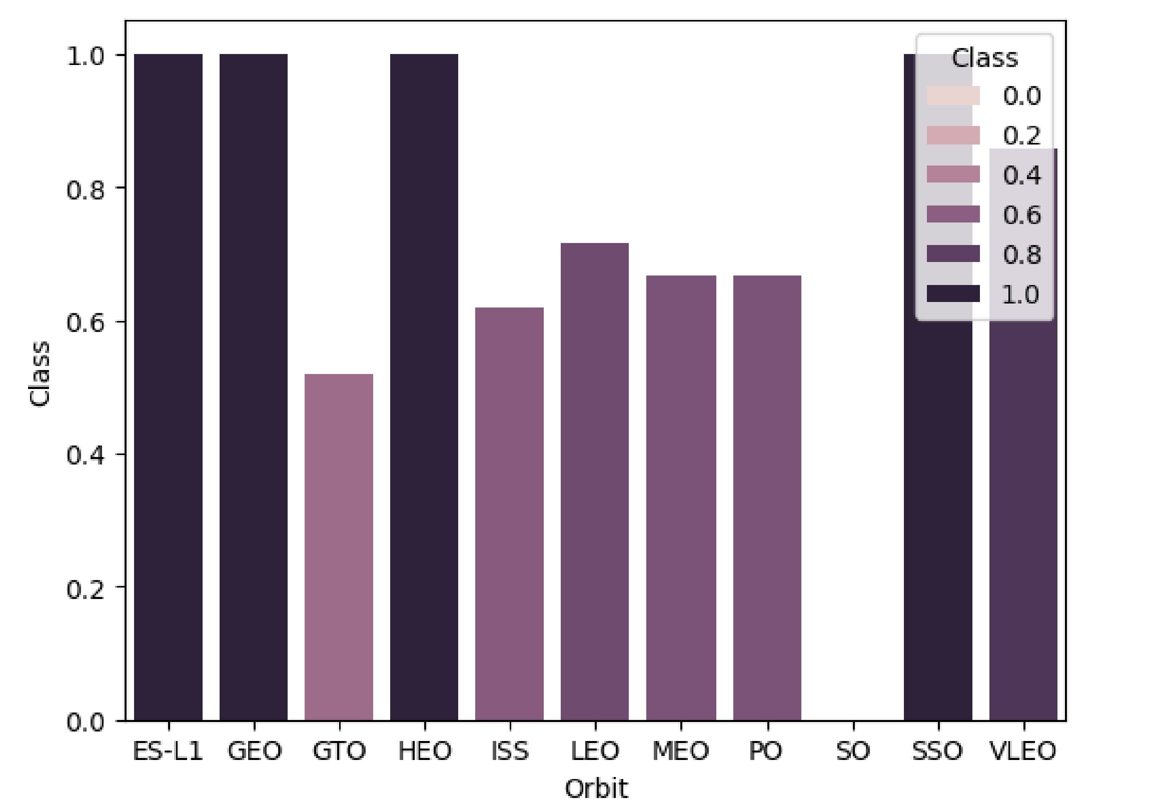


Github link: https://github.com/gioboto/Applied_Data_Science_Capstone/tree/a84e9f2b6b98450ed8ee383aedaf84c063e04c6b/module02/05-EDA%20with%20Visualization%20Lab

EDA WITH VISUALIZATION LAB

3

- Visualize the relationship between success rate of each orbit type

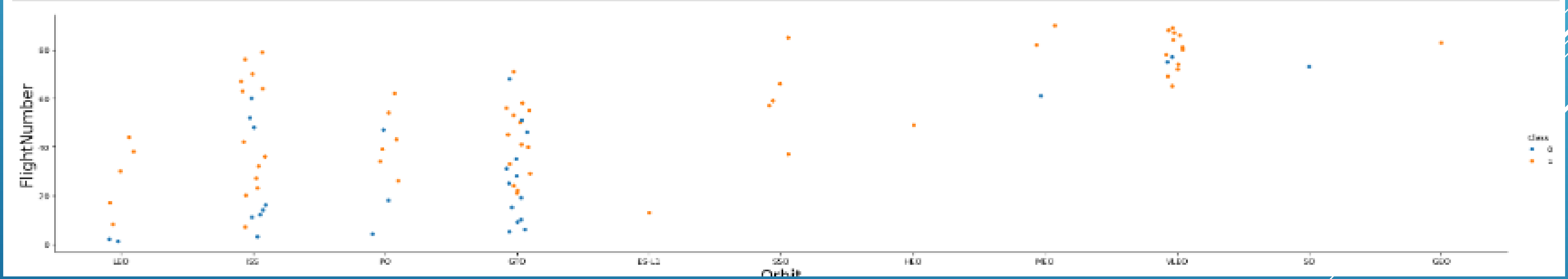


Github link: https://github.com/gioboto/Applied_Data_Science_Capstone/tree/a84e9f2b6b98450ed8ee383aedaf84c063e04c6b/module02/05-EDA%20with%20Visualization%20Lab

EDA WITH VISUALIZATION LAB

4

- Visualize the relationship between FlightNumber and Orbit type

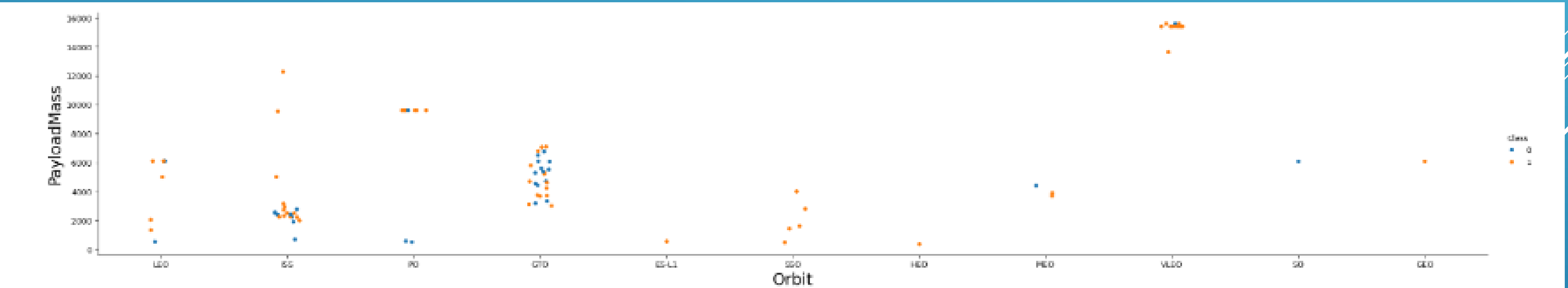


Github link: https://github.com/gioboto/Applied_Data_Science_Capstone/tree/a84e9f2b6b98450ed8ee383aedaf84c063e04c6b/module02/05-EDA%20with%20Visualization%20Lab

EDA WITH VISUALIZATION LAB

5

- Visualize the relationship between Payload Mass and Orbit type

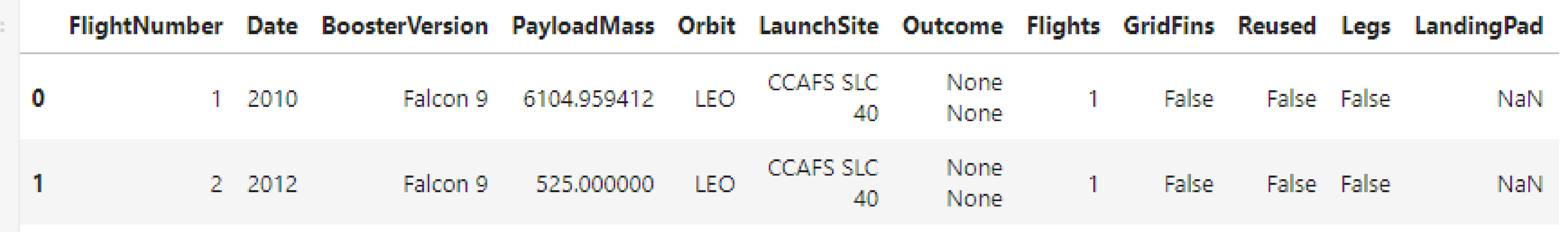


Github link: https://github.com/gioboto/Applied_Data_Science_Capstone/tree/a84e9f2b6b98450ed8ee383aedaf84c063e04c6b/module02/05-EDA%20with%20Visualization%20Lab

EDA WITH VISUALIZATION LAB

6

- Visualize the launch success yearly trend



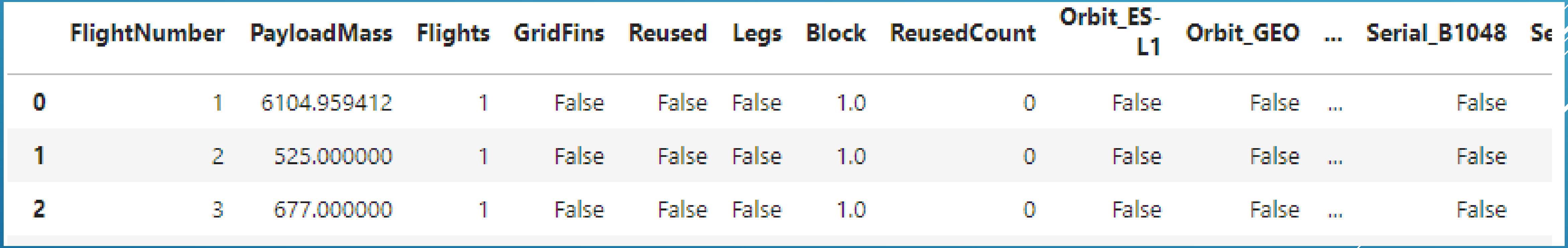
	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad
0	1	2010	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN
1	2	2012	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN

Github link: https://github.com/gioboto/Applied_Data_Science_Capstone/tree/a84e9f2b6b98450ed8ee383aedaf84c063e04c6b/module02/05-EDA%20with%20Visualization%20Lab

EDA WITH VISUALIZATION LAB

7

- Create dummy variables to categorical columns

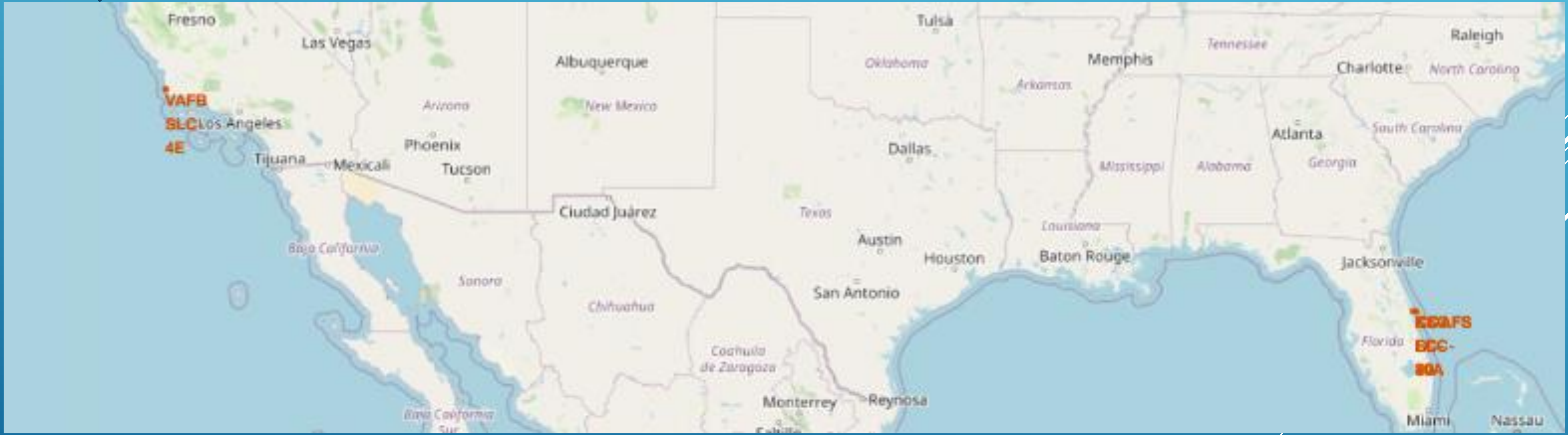


	FlightNumber	PayloadMass	Flights	GridFins	Reused	Legs	Block	ReusedCount	Orbit_ES-L1	Orbit_GEO	...	Serial_B1048	Se
0	1	6104.959412	1	False	False	False	1.0	0	False	False	...	False	
1	2	525.000000	1	False	False	False	1.0	0	False	False	...	False	
2	3	677.000000	1	False	False	False	1.0	0	False	False	...	False	

Github link: https://github.com/gioboto/Applied_Data_Science_Capstone/tree/a84e9f2b6b98450ed8ee383aedaf84c063e04c6b/module02/05-EDA%20with%20Visualization%20Lab

WITH FOLIUM

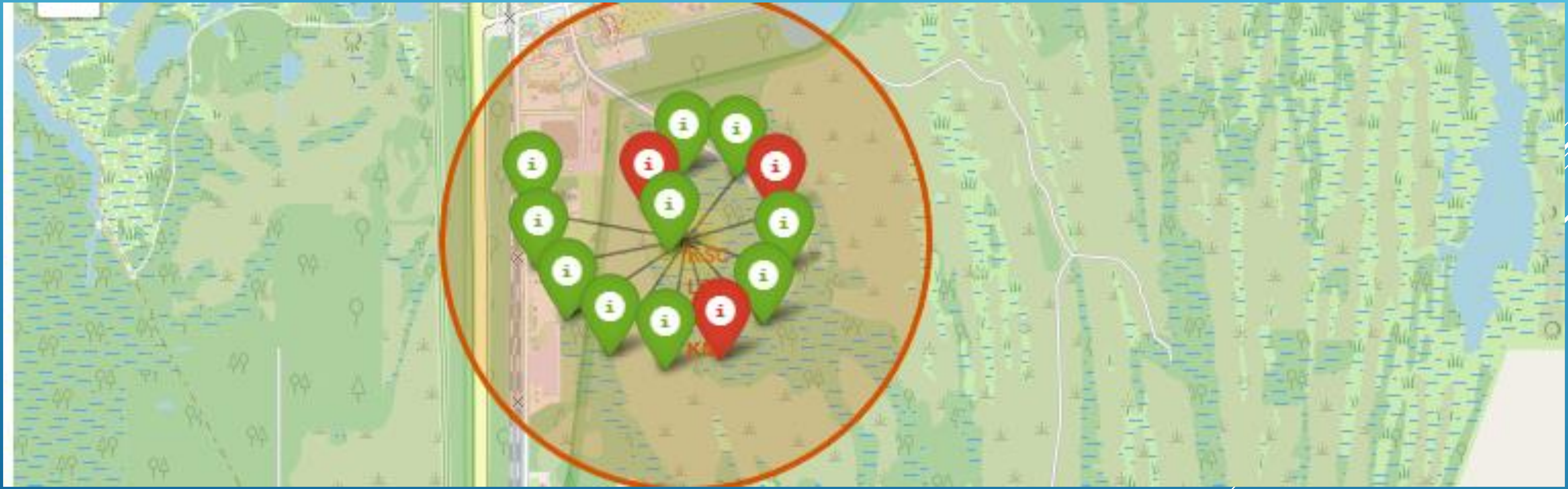
Mark all launch sites on a map



Github link: https://github.com/gioboto/Applied_Data_Science_Capstone/tree/a84e9f2b6b98450ed8ee383aedaf84c063e04c6b/module03/02-Hands-on%20Lab%20Interactive%20Visual%20Analytics%20with%20Folium%20lab

WITH FOLIUM

Mark the success/failed launches for each site on the map

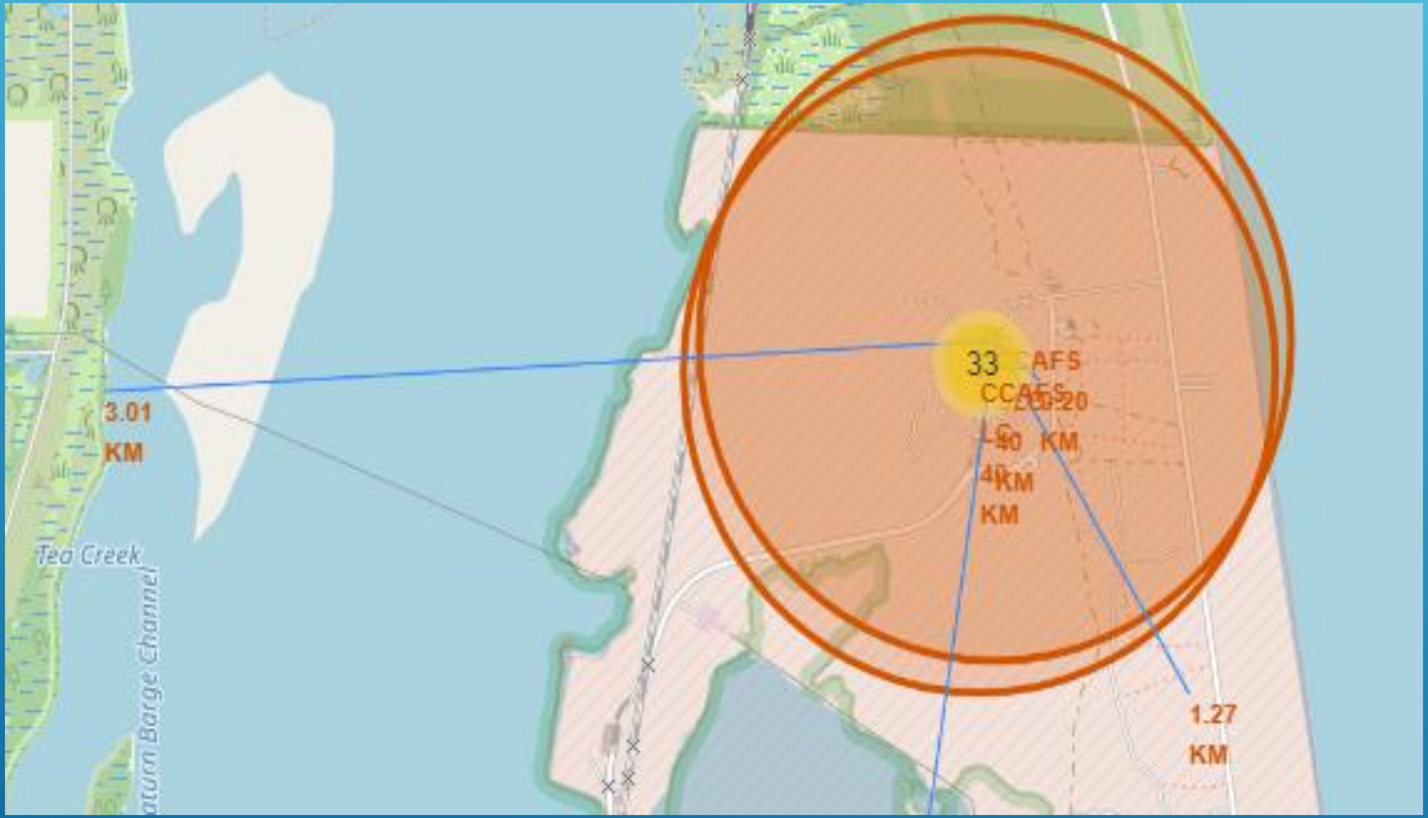


Github link: https://github.com/gioboto/Applied_Data_Science_Capstone/tree/a84e9f2b6b98450ed8ee383aedaf84c063e04c6b/module03/02-Hands-on%20Lab%20Interactive%20Visual%20Analytics%20with%20Folium%20lab

INTERACTIVE VISUAL ANALYTICS WITH FOLIUM

Calculate the distances between a launch site to its proximities

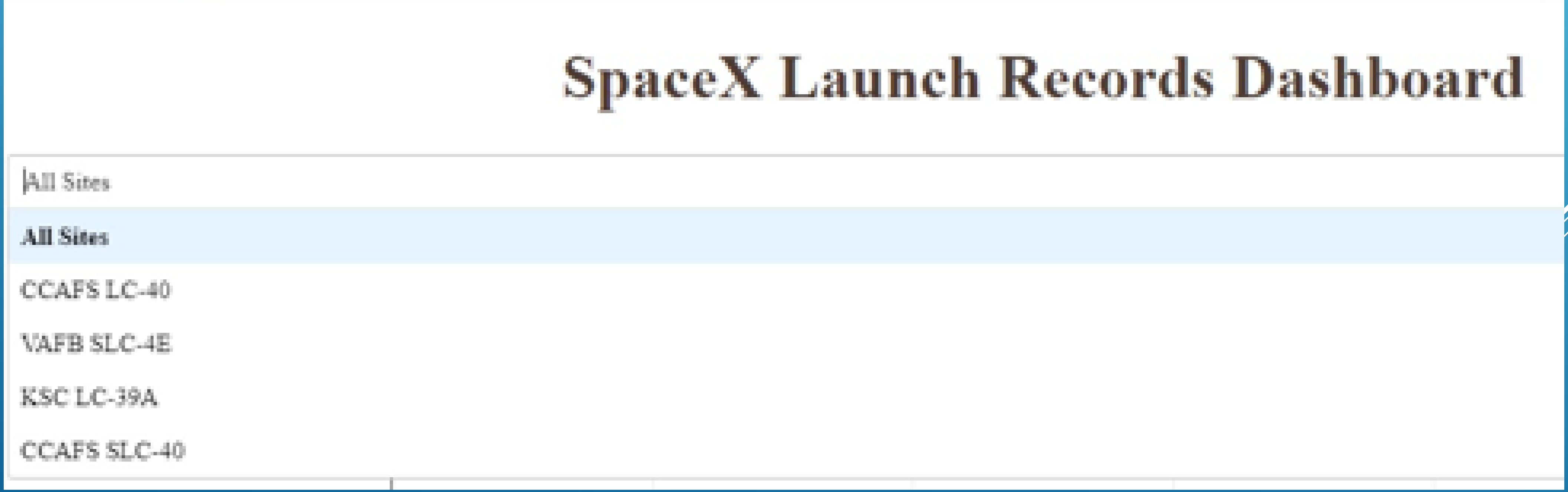
- .



Github link: https://github.com/gioboto/Applied_Data_Science_Capstone/tree/a84e9f2b6b98450ed8ee383aedaf84c063e04c6b/module03/02-Hands-on%20Lab%20Interactive%20Visual%20Analytics%20with%20Folium%20lab

BUILD AN INTERACTIVE DASHBOARD WITH PLOTY DASH

Add a Launch Site Drop-down Input Component



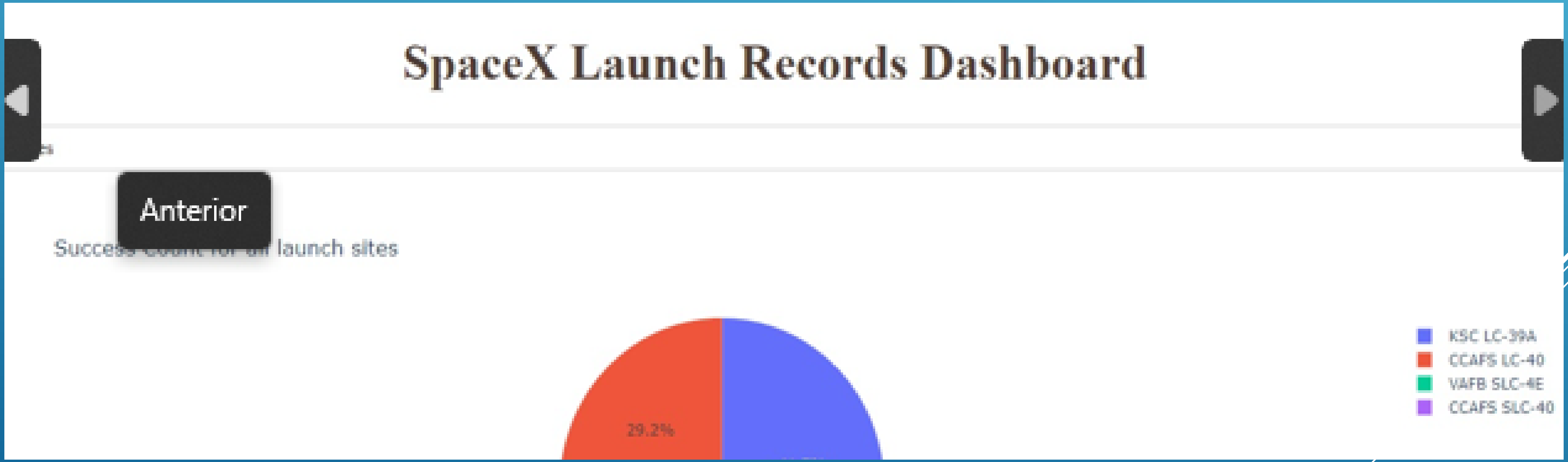
SpaceX Launch Records Dashboard

Github link: https://github.com/gioboto/Applied_Data_Science_Capstone/tree/a84e9f2b6b98450ed8ee383aedaf84c063e04c6b/module03/03-Hands-on%20Lab%20Build%20an%20Interactive%20Dashboard%20with%20Ploty%20Dash

BUILD AN INTERACTIVE DASHBOARD WITH PLOTY DASH

Add a callback function to render `success-pie-chart` based on selected site dropdown

- .



SpaceX Launch Records Dashboard

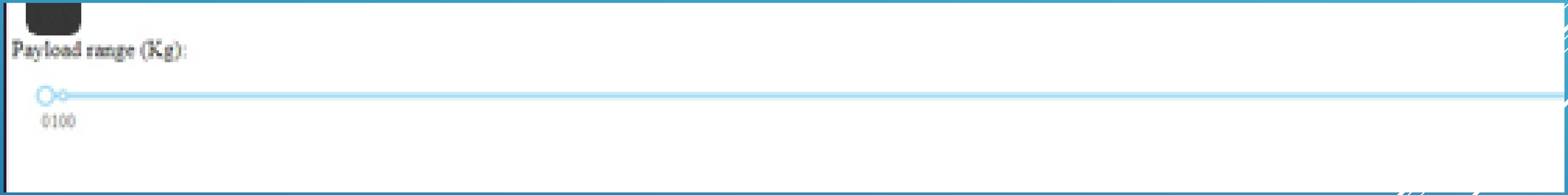
Success Count for all launch sites

Github link: https://github.com/gioboto/Applied_Data_Science_Capstone/tree/a84e9f2b6b98450ed8ee383aedaf84c063e04c6b/module03/03-Hands-on%20Lab%20Build%20an%20Interactive%20Dashboard%20with%20Ploty%20Dash

BUILD AN INTERACTIVE DASHBOARD WITH PLOTY DASH

Add a Range Slider to Select Payload

- .



Payload range (Kg):

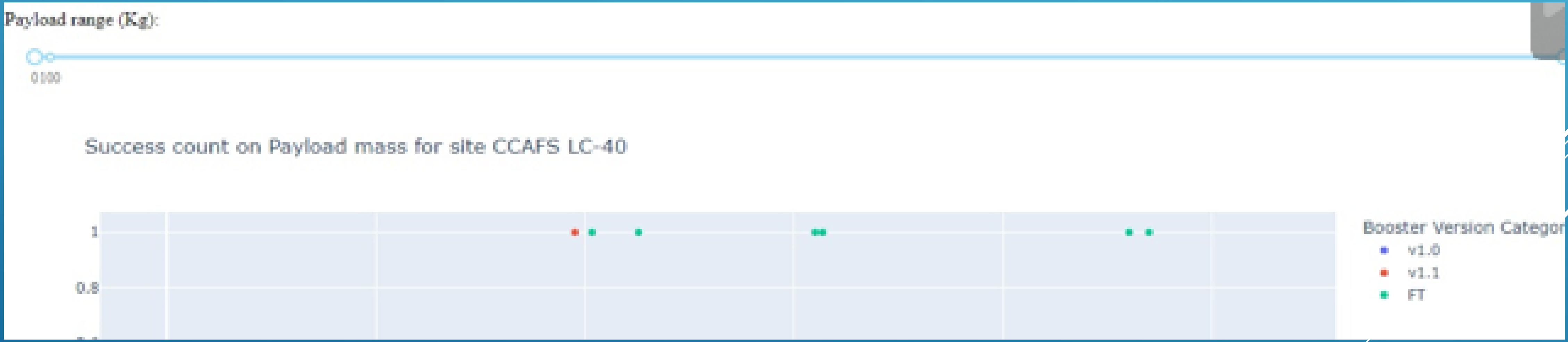
0100

Github link: https://github.com/gioboto/Applied_Data_Science_Capstone/tree/a84e9f2b6b98450ed8ee383aedaf84c063e04c6b/module03/03-Hands-on%20Lab%20Build%20an%20Interactive%20Dashboard%20with%20Ploty%20Dash

BUILD AN INTERACTIVE DASHBOARD WITH PLOTY DASH

- .

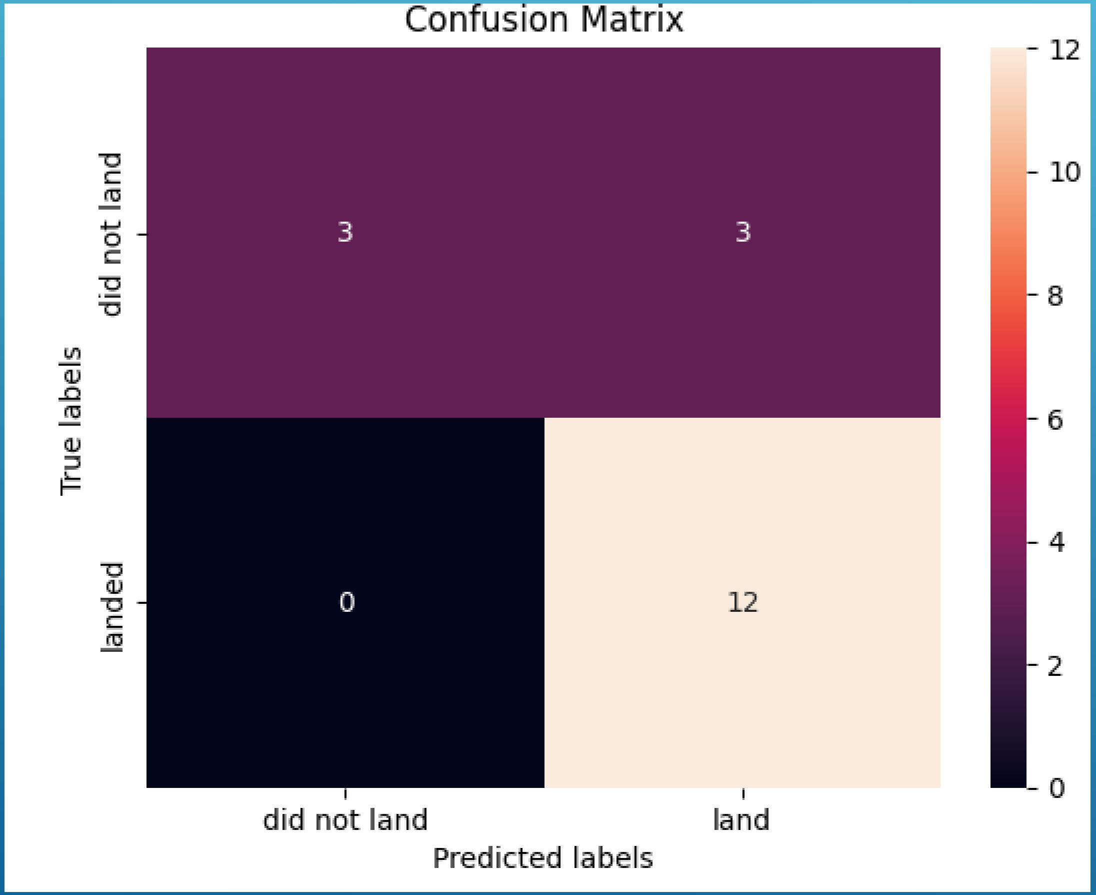
Add a callback function to render the `success-payload-scatter-chart` scatter plot



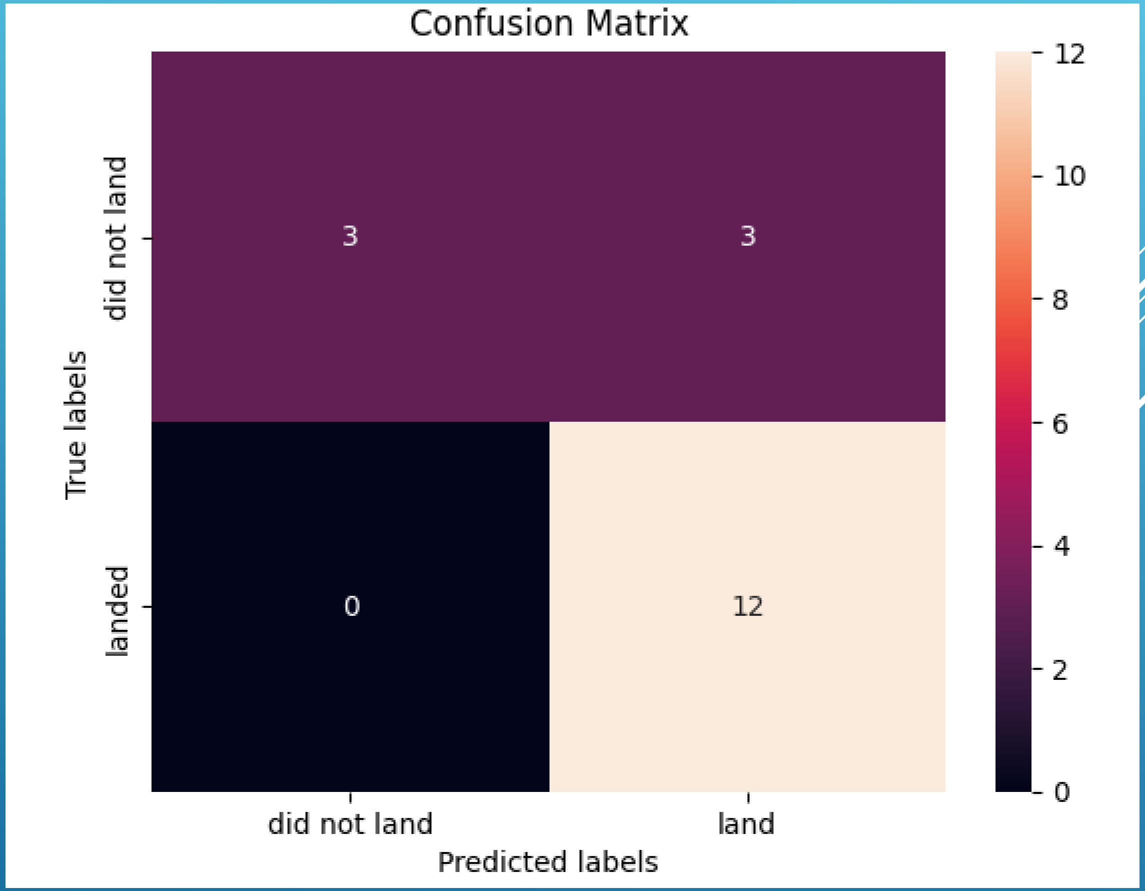
Github link: https://github.com/gioboto/Applied_Data_Science_Capstone/tree/a84e9f2b6b98450ed8ee383aedaf84c063e04c6b/module03/03-Hands-on%20Lab%20Build%20an%20Interactive%20Dashboard%20with%20Ploty%20Dash

MACHINE LEARNING PREDICTION

Confusion Matrix of LogisticRegression



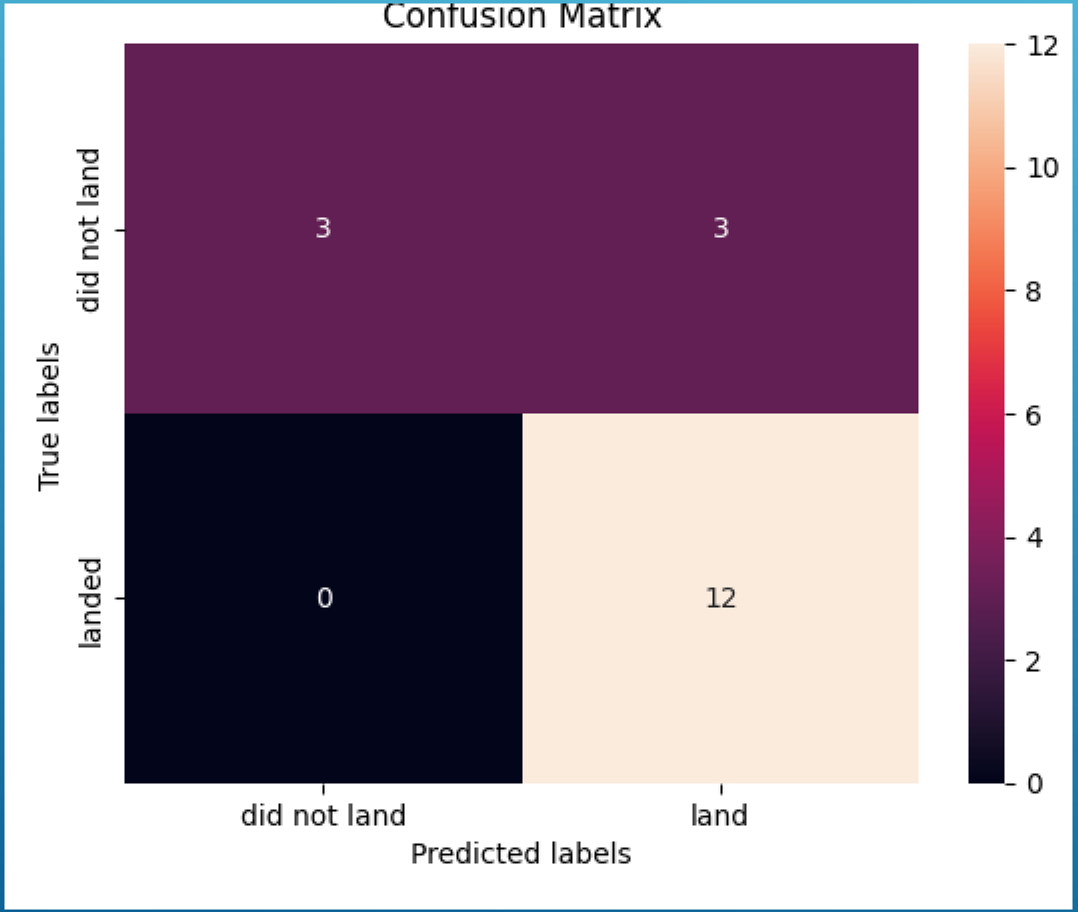
Confusion Matrix of SVC



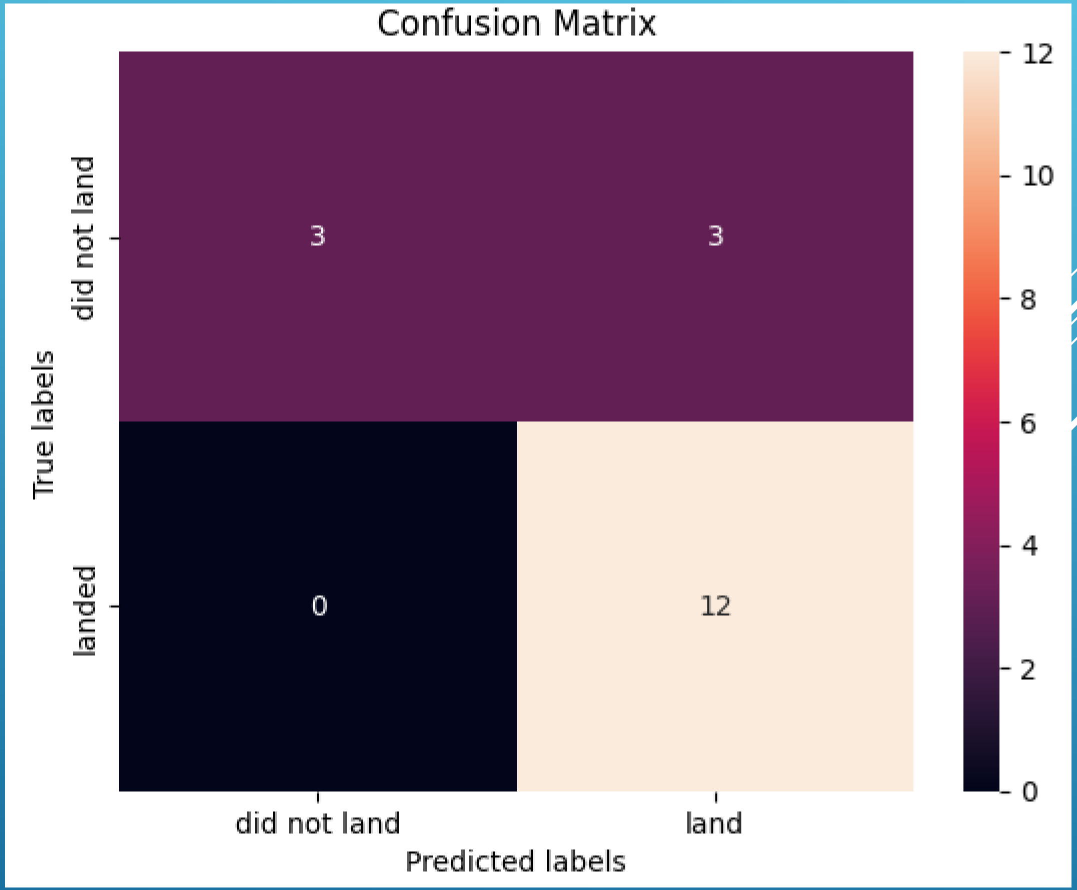
Github link: https://github.com/gioboto/Applied_Data_Science_Capstone/tree/a84e9f2b6b98450ed8ee383aedaf84c063e04c6b/module04/02-Hands-on%20Lab%20Complete%20the%20Machine%20Learning%20Prediction%20lab

MACHINE LEARNING PREDICTION

Confusion Matrix of DecisionTreeClassifier

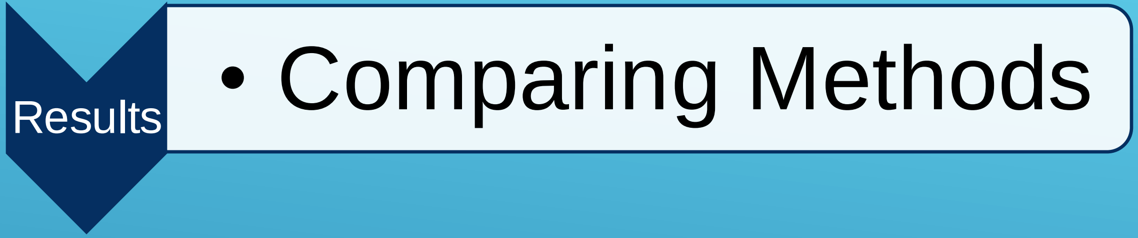


Confusion Matrix of KNeighborsClassier



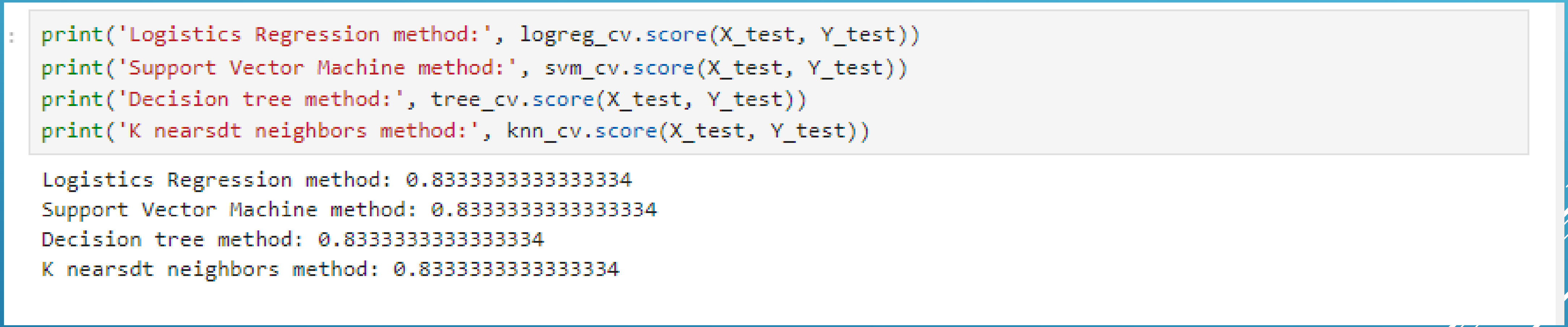
Github link: https://github.com/gioboto/Applied_Data_Science_Capstone/tree/a84e9f2b6b98450ed8ee383aedaf84c063e04c6b/module04/02-Hands-on%20Lab%20Complete%20the%20Machine%20Learning%20Prediction%20lab

MACHINE LEARNING PREDICTION



Results

- Comparing Methods



```
print('Logistics Regression method:', logreg_cv.score(X_test, Y_test))
print('Support Vector Machine method:', svm_cv.score(X_test, Y_test))
print('Decision tree method:', tree_cv.score(X_test, Y_test))
print('K nearsdt neighbors method:', knn_cv.score(X_test, Y_test))
```

```
Logistics Regression method: 0.8333333333333334
Support Vector Machine method: 0.8333333333333334
Decision tree method: 0.8333333333333334
K nearsdt neighbors method: 0.8333333333333334
```

Github link: https://github.com/gioboto/Applied_Data_Science_Capstone/tree/a84e9f2b6b98450ed8ee383aedaf84c063e04c6b/module04/02-Hands-on%20Lab%20Complete%20the%20Machine%20Learning%20Prediction%20lab

CONCLUSIONS

- Woking with data implies to export to formats that allow us wrangling data, to clean it and understand it
- The site with code CCAFS SLC 40 has the bigger number of launches vs KSC LC 39A and VAFB SLC 4E
- According the information extracted of the data there is a success rate oof missions equal to 60%
- Missions with lighter payloads have a higher performance compared to mission with heavier ones.
- Ploting data shows that ES-L1, GEO, HEO, SSO orbit types have the highest rates of successful launches.
- Ploting data in a chart shows increases in success rate since 2013 to 2020
- About the four models for forecasting, (Logistics Regression, Support Vector Machine, Decision tree, K nearest neighbors), present the same accuracy value (0.8333333333333334)