Winning Space Race with Data Science

Rafael Ramos Wagner
18/07/2022

Outline

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

Executive Summary

- In this report we will use all the data science methodology learned at this point on the course. The main points are:
  - Business Understanding
  - Data Collection
  - Exploratory Data Analysis (EDA) using visualization and SQL
  - Visual Analytics using Folium and Plotly Dash
  - Modeling
  - Evaluation
- All the models evaluated had the same results with accuracy of 83% and present just false positive errors on the confusion matrix.

Introduction

- We will predict if the Falcon 9 first stage will land successfully. 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.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected using Web Scraping and Collection API from SpaceX
- Perform data wrangling
  - Data was processed using Pandas Dataframe to treat and select features for the machine learning modeling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Perform One Hot Encoder at the categorical features, than 4 classifications models were trained and evaluate to determine the best model. The models tested were SVM, KNN, Decision Tree and Logistical Regression.

Data Collection

- The data sets were collected using web scraping from the .
- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts

- Link to GitHub

Place your flowchart of SpaceX API calls here

Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

- Link to GitHub

Place your flowchart of web scraping here

Data Wrangling

- Describe how data were processed
- You need to present your data wrangling process using key phrases and flowcharts
- Link to GitHub

EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
- Link to GitHub

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
- Link to GitHub

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- Explain why you added those objects
- Link to GitHub

Build a Dashboard with Plotly Dash

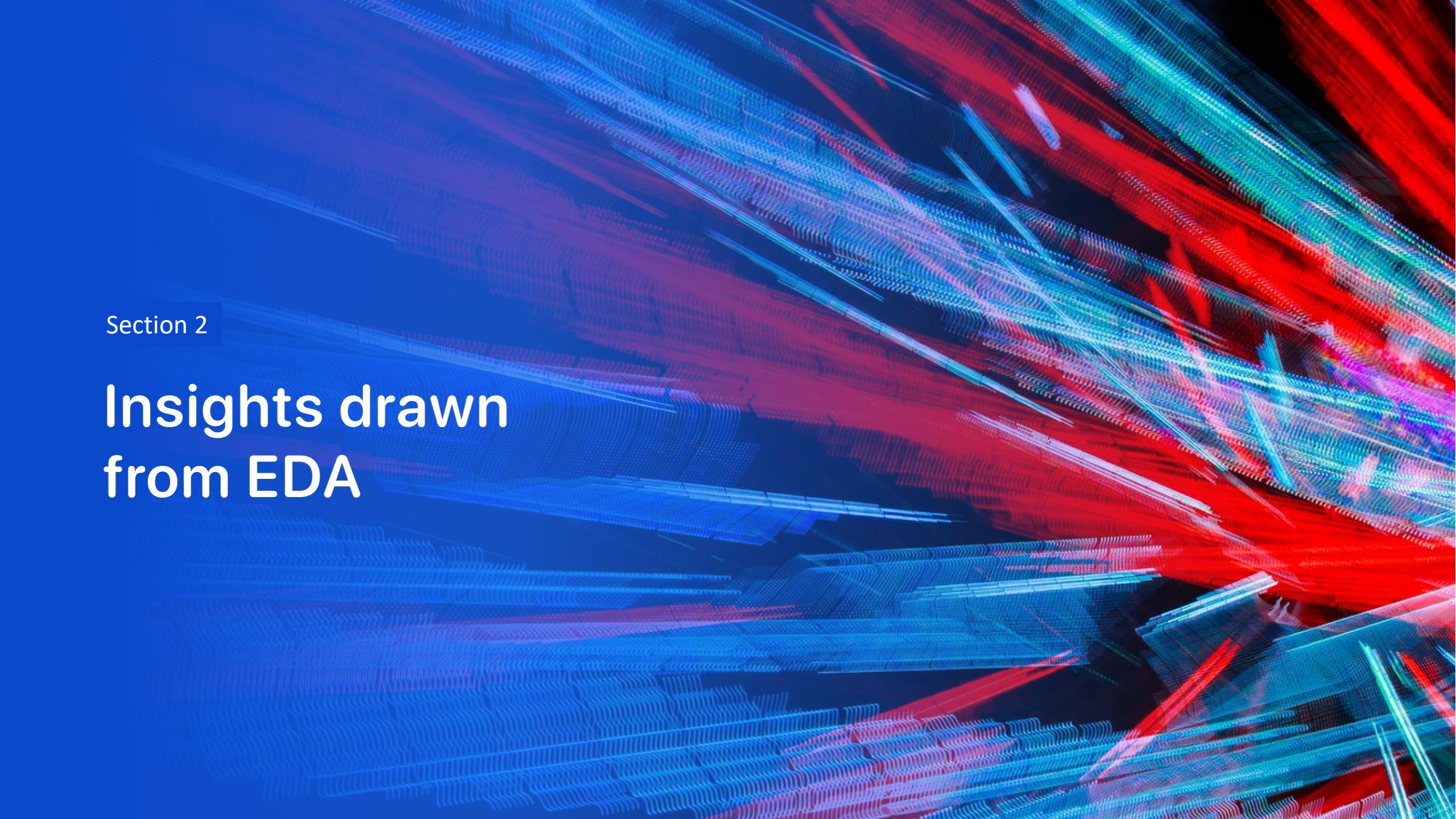
- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- Link to GitHub

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- You need present your model development process using key phrases and flowchart
- Link to GitHub

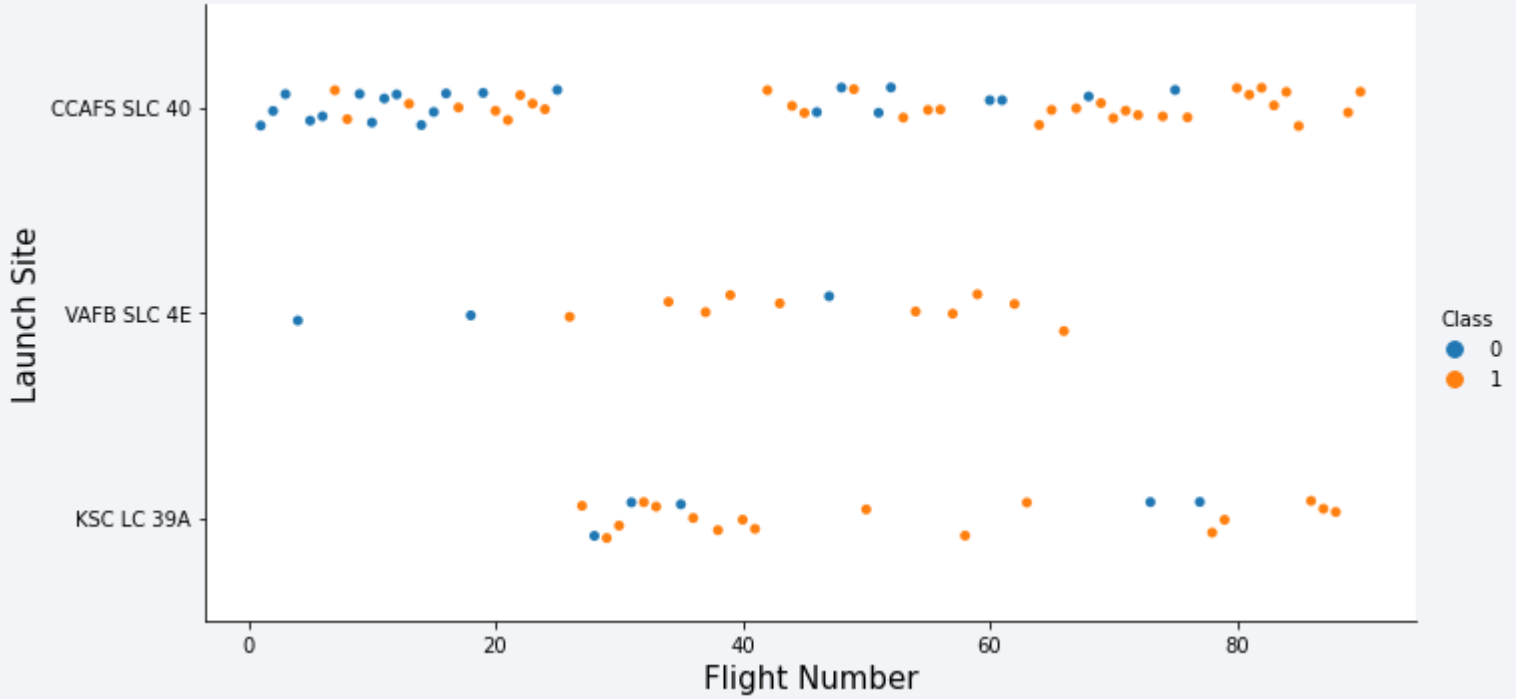
Results

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


Section 2
Insights drawn from EDA

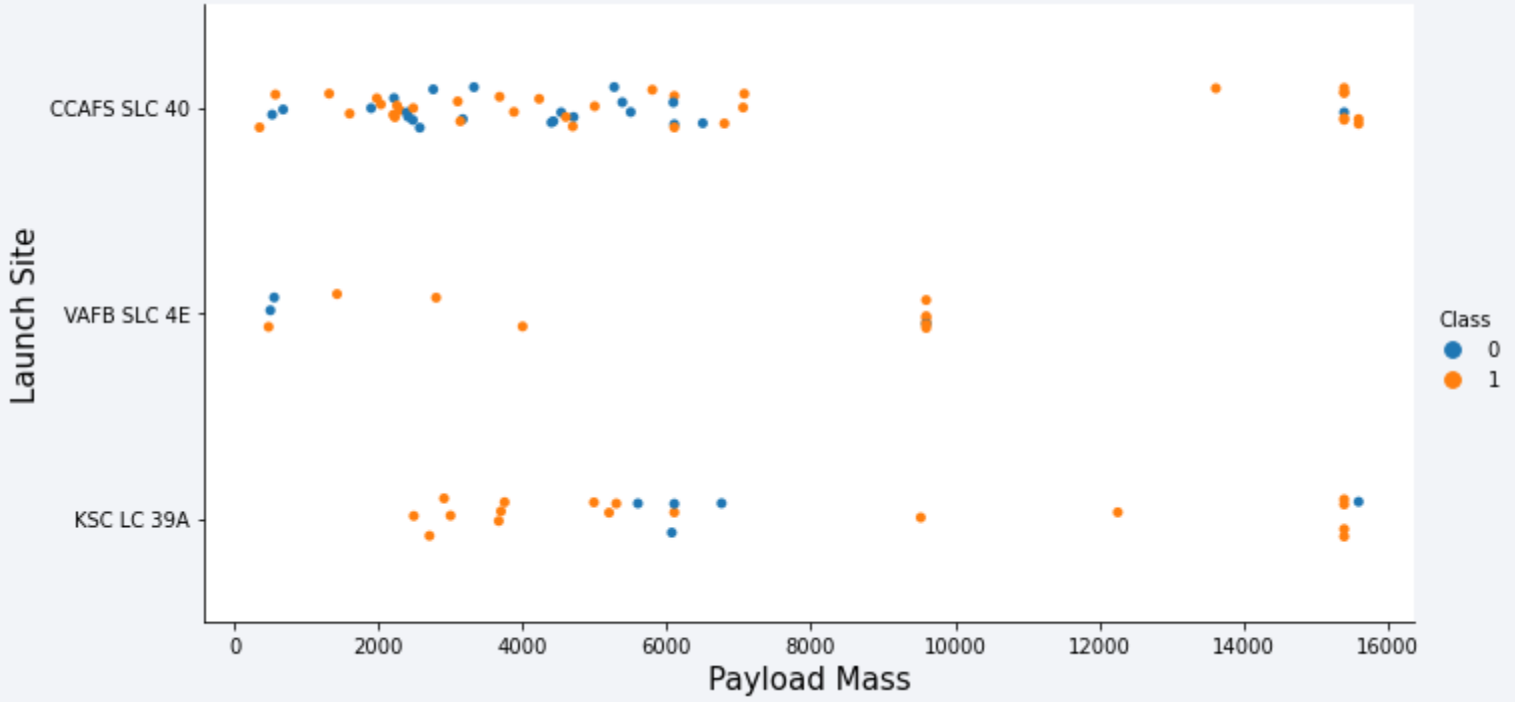
Flight Number vs. Launch Site

- Show a scatter plot of Flight Number vs. Launch Site
- Show the screenshot of the scatter plot with explanations



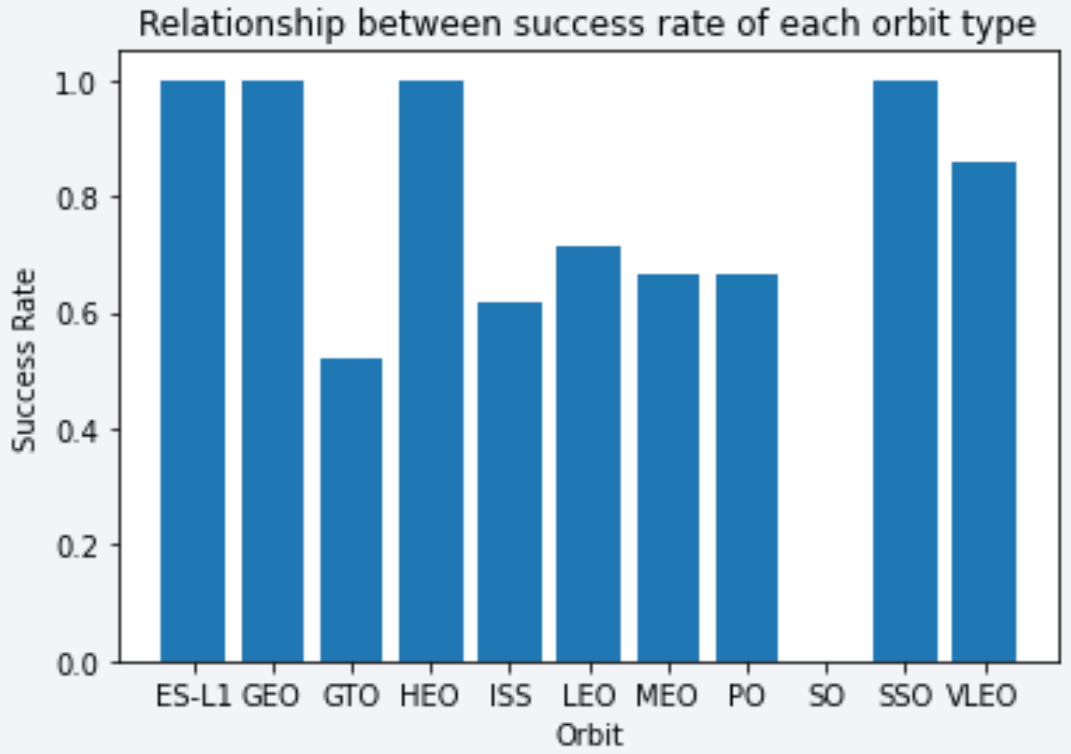
Payload vs. Launch Site

- Show a scatter plot of Payload vs. Launch Site
- Show the screenshot of the scatter plot with explanations



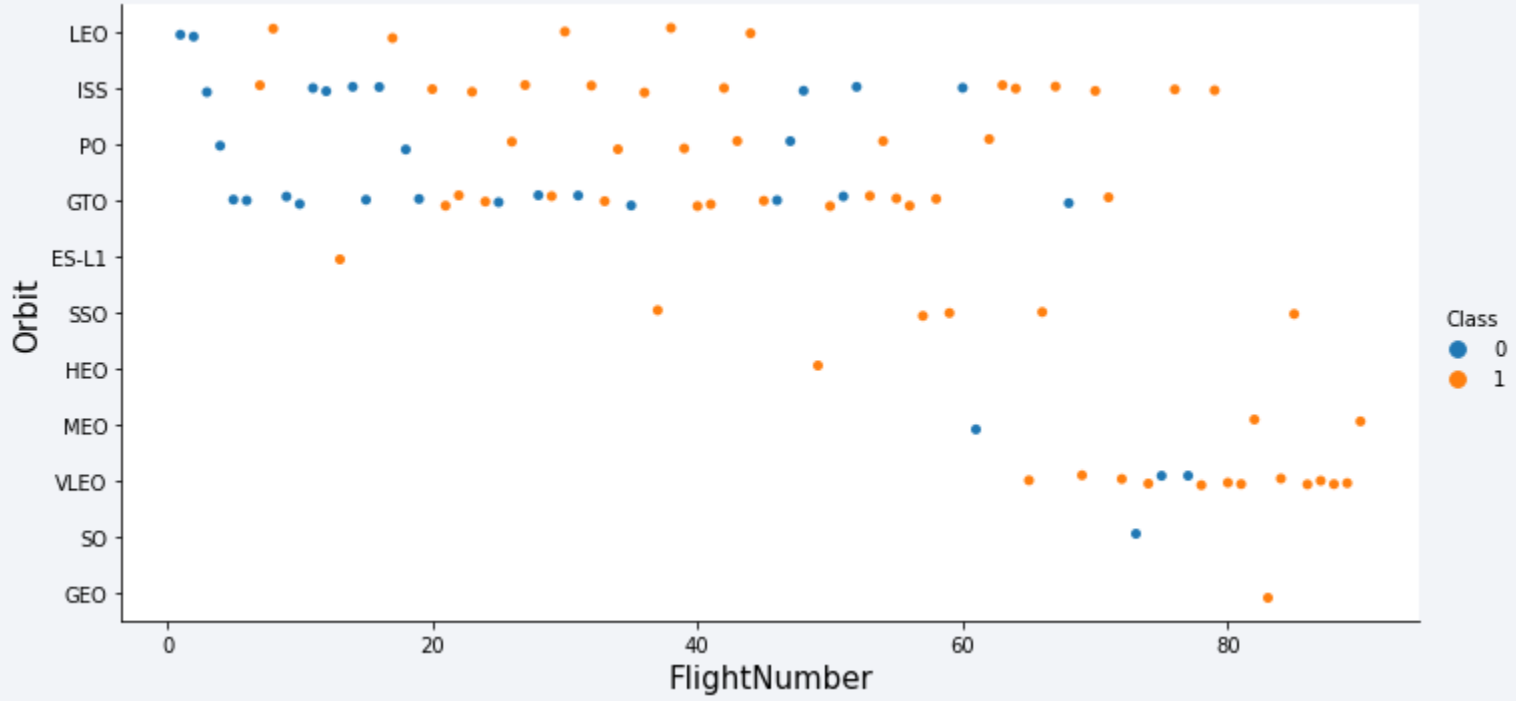
Success Rate vs. Orbit Type

- Show a bar chart for the success rate of each orbit type
- Show the screenshot of the scatter plot with explanations



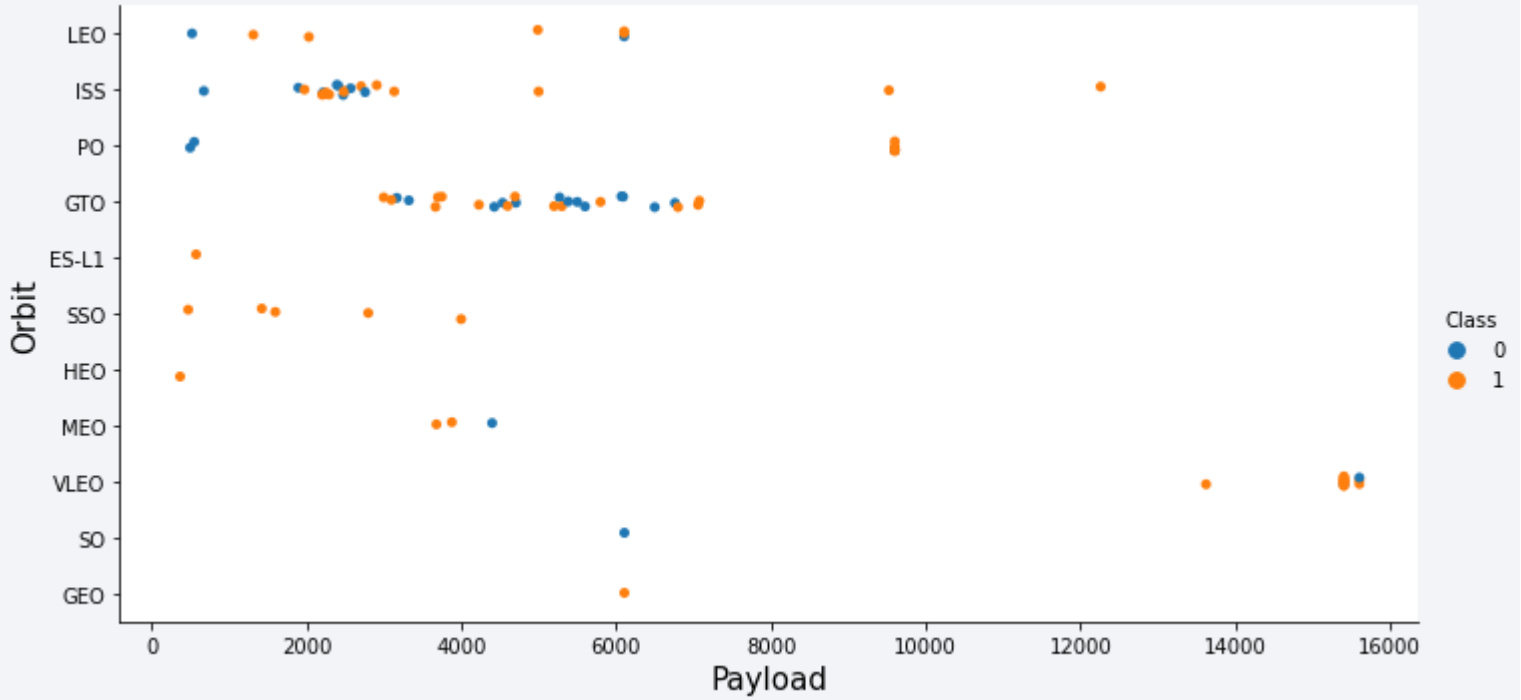
Flight Number vs. Orbit Type

- Show a scatter point of Flight number vs. Orbit type
- Show the screenshot of the scatter plot with explanations



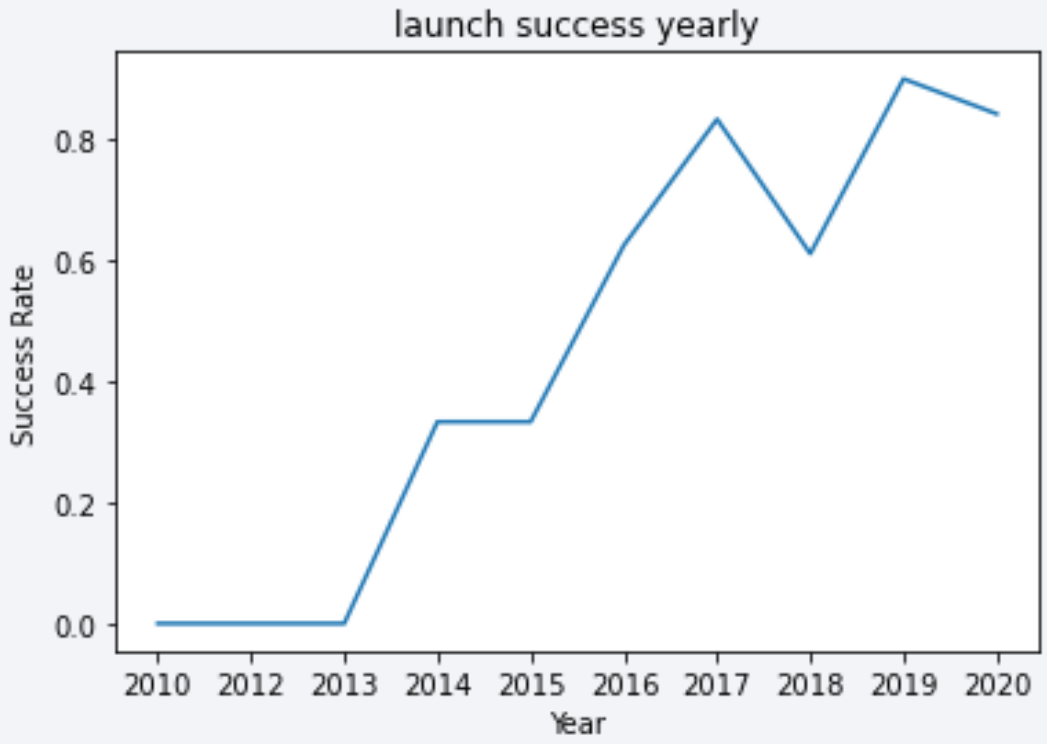
Payload vs. Orbit Type

- Show a scatter point of payload vs. orbit type
- Show the screenshot of the scatter plot with explanations



Launch Success Yearly Trend

- Show a line chart of yearly average success rate
- Show the screenshot of the scatter plot with explanations



All Launch Site Names

- Find the names of the unique launch sites
- Present your query result with a short explanation here

```
In [8]: %%sql

        SELECT Distinct(Launch_Site) FROM SPACEXTBL

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

Out[8]:

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

Launch Site Names Begin with 'CCA'

In [9]:
```
%%sql

SELECT *
FROM SPACEXTBL
WHERE Launch_Site like 'CCA%'
LIMIT 5
```

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

Out[9]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	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
- Present your query result with a short explanation here

```
In [10]: %%sql

         SELECT sum(PAYLOAD_MASS__KG_)
         FROM SPACEXTBL
         WHERE Customer = 'NASA (CRS)'

          * sqlite:///my_data1.db
         Done.

Out[10]:
 sum(PAYLOAD_MASS__KG_)
 ----------------------
                  45596
```

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- Present your query result with a short explanation here

```
In [11]: %%sql

         SELECT avg(PAYLOAD_MASS__KG_)
         FROM SPACEXTBL
         WHERE Booster_Version like 'F9 v1.1%'

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

Out[11]:

avg(PAYLOAD_MASS__KG_)
2534.6666666666665

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- Present your query result with a short explanation here

```
In [12]: %%sql

         SELECT min(Date)
         FROM SPACEXTBL
         WHERE "Landing _Outcome" like 'Success%'

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

Out[12]:

min(Date)
01-05-2017

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

- Present your qu

In [13]:
```
%%sql

SELECT Booster_Version
FROM SPACEXTBL
WHERE "Landing _Outcome" like 'Success (drone ship)'
    AND PAYLOAD_MASS__KG_ between 4000 and 6000
```

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

Out[13]:

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
- Present your query result with a short explanation here

In [14]:
```
%%sql

SELECT Mission_Outcome, count(Mission_Outcome)
FROM SPACEXTBL
GROUP BY Mission_Outcome
```

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

Out[14]:

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

Boosters Carried Maximum Payload

- List the names of the booster which hav
  mass
- Present your query result with a short ex

In [15]:
```
%%sql

SELECT Booster_Version
FROM SPACEXTBL
WHERE PAYLOAD_MASS__KG_ = (SELECT max(PAYLOAD_MASS__KG_)
        FROM SPACEXTBL)
```

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

Out[15]:

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

In [16]:
```
%%sql

SELECT substr(Date, 4, 2) as month, "Landing _Outcome", Booster_Version, Launch_Site
FROM SPACEXTBL
WHERE substr(Date,7,4)='2015'
    AND "Landing _Outcome" like 'Failure (drone ship)%'
```

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

Out[16]:

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

In [17]:
```
%%sql

SELECT *
FROM SPACEXTBL
WHERE "Landing _Outcome" like  'Success%'
    AND Date between '05-06-2010' and '20-03-2017'
ORDER BY substr(Date,7,4) desc, substr(Date, 4, 2) desc, substr(Date, 0, 2) desc
```

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

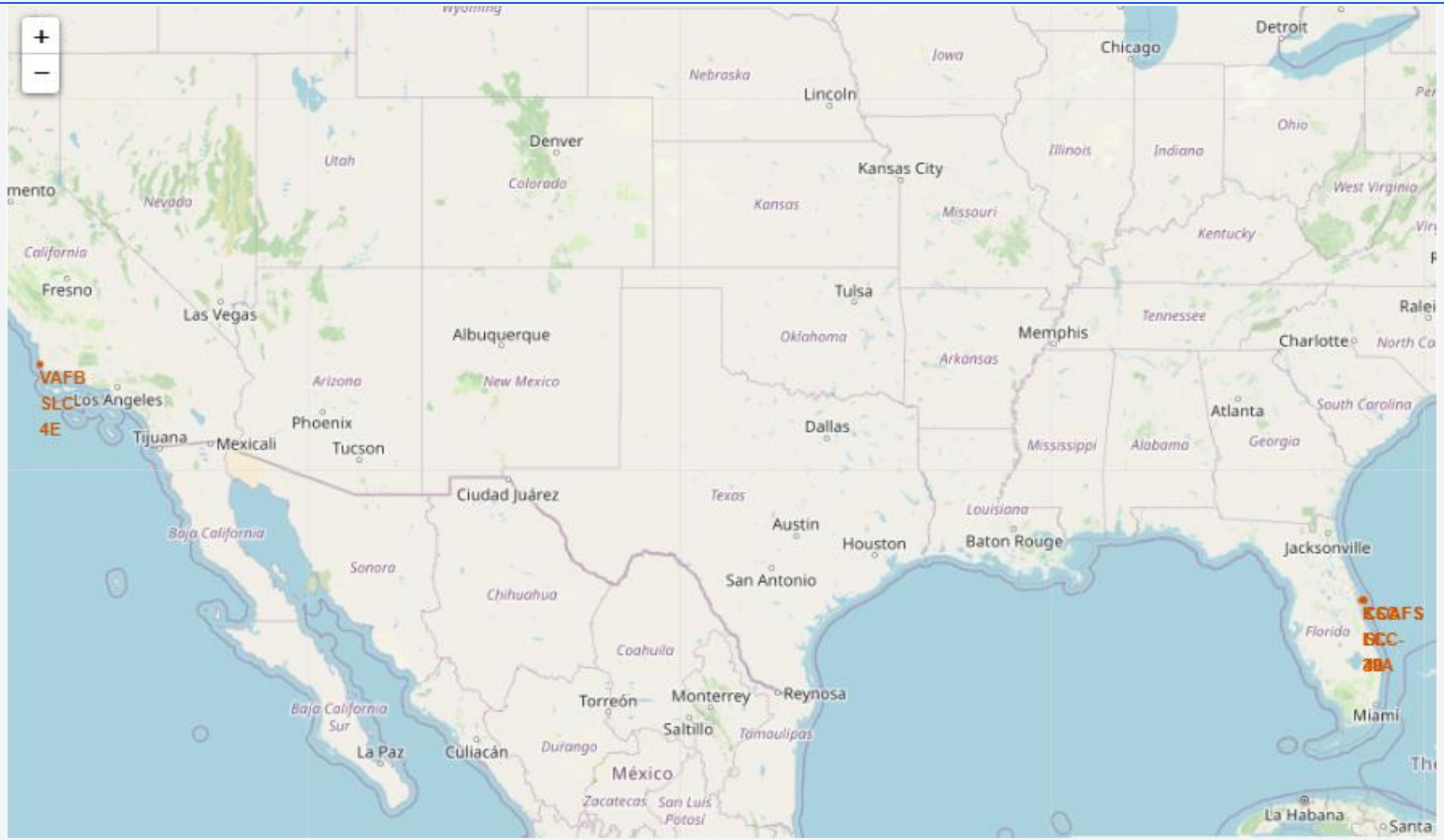
Out[17]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
06-12-2020	16:17:08	F9 B5 B1058.4	KSC LC-39A	SpaceX CRS-21	2972	LEO (ISS)	NASA (CRS)	Success	Success
16-11-2020	00:27:00	F9 B5B1061.1	KSC LC-39A	Crew-1, Sentinel-6 Michael Freilich	12500	LEO (ISS)	NASA (CCP)	Success	Success
05-11-2020	23:24:23	F9 B5B1062.1	CCAFS SLC-40	GPS III-04 , Crew-1	4311	MEO	USSF	Success	Success
18-10-2020	12:25:57	F9 B5 B1051.6	KSC LC-39A	Starlink 13 v1.0, Starlink 14 v1.0	15600	LEO	SpaceX	Success	Success
06-10-2020	11:29:34	F9 B5 B1058.3	KSC LC-39A	Starlink 12 v1.0, Starlink 13 v1.0	15600	LEO	SpaceX	Success	Success
18-08-2020	14:31:00	F9 B5 B1049.6	CCAFS SLC-40	Starlink 10 v1.0, SkySat-19, -20, -21, SAOCOM 1B	15440	LEO	SpaceX, Planet Labs, PlanetIQ	Success	Success
07				Starlink 9 v1.0			SpaceX, Spaceflight		

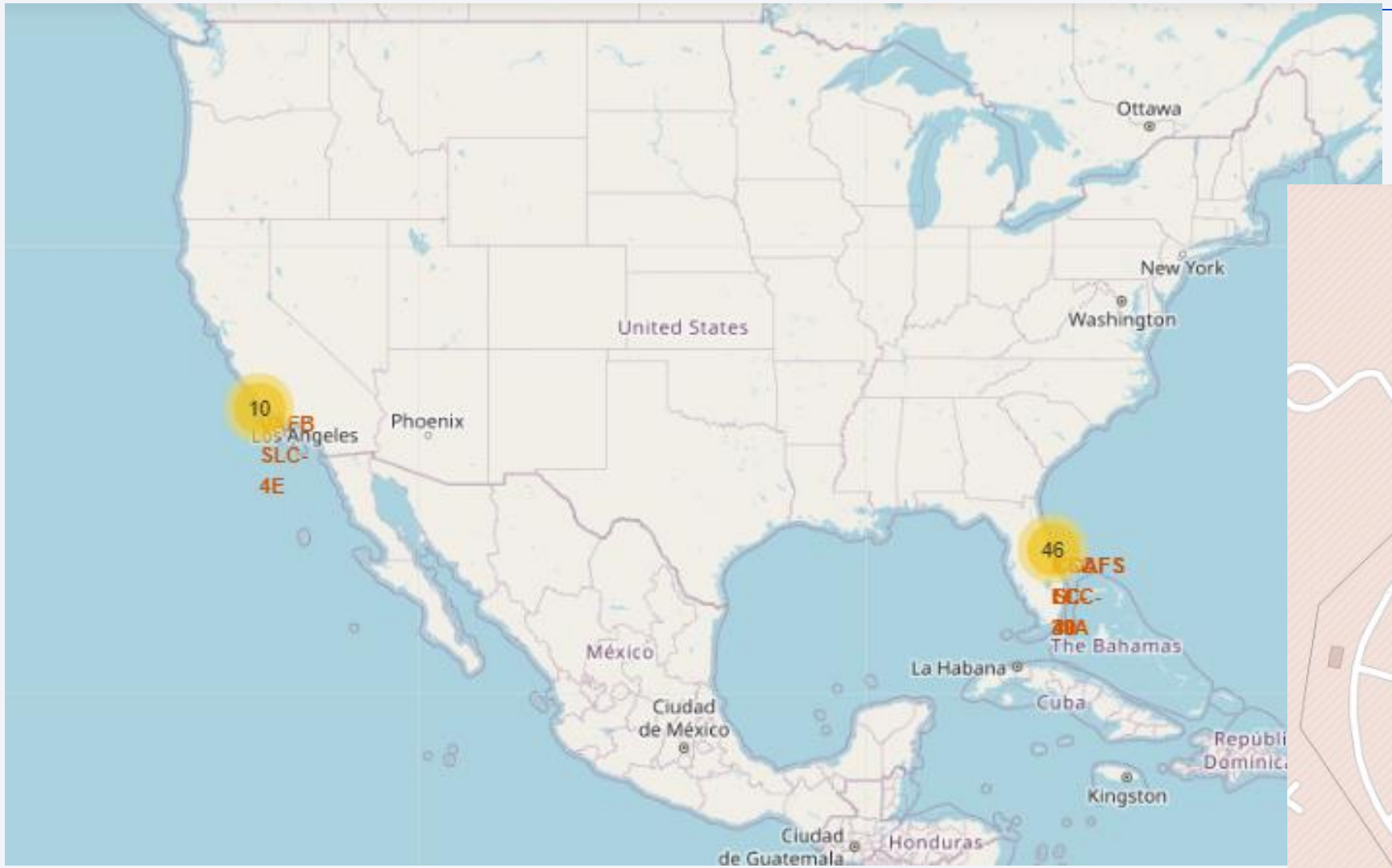

Section 3
Launch Sites
Proximities Analysis

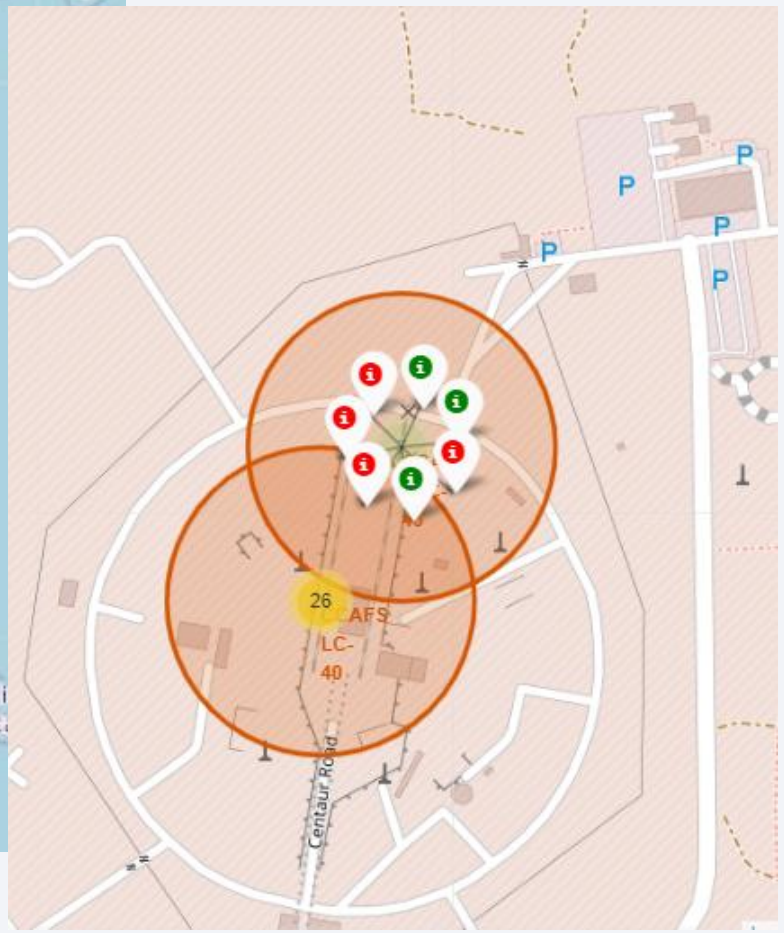
Folium Map – Launch Sites

-
-
-

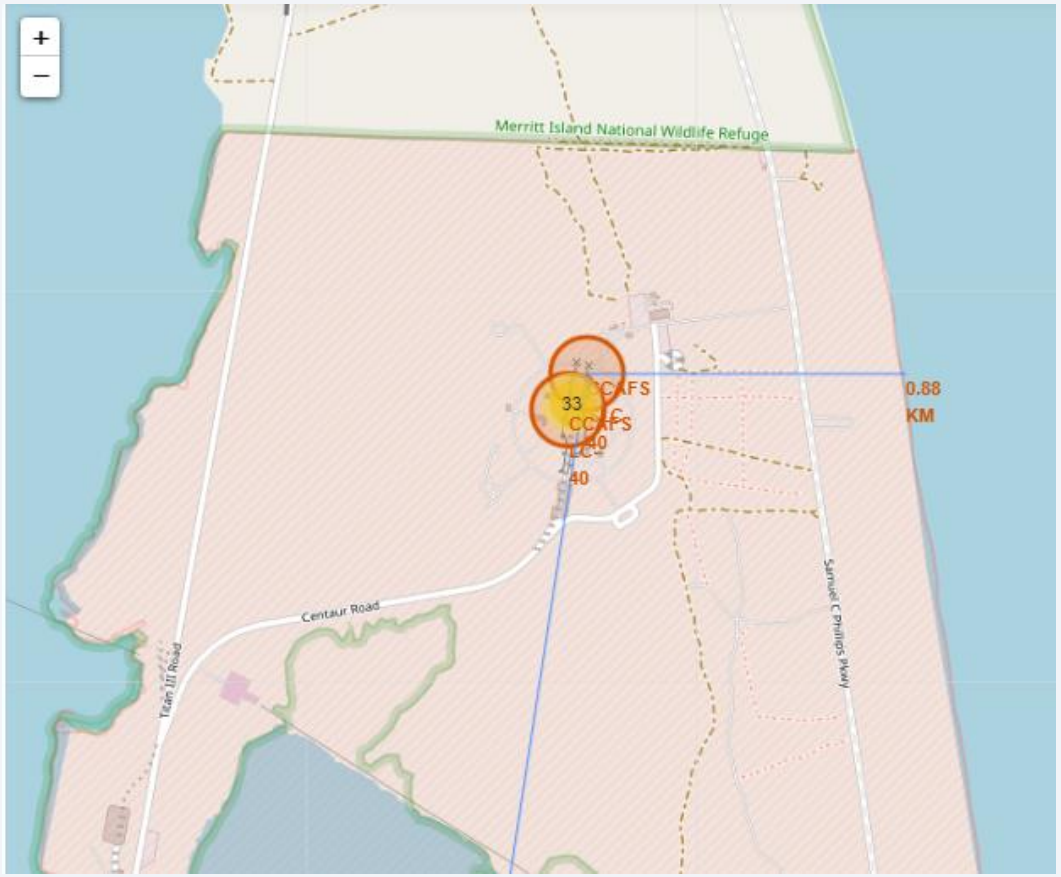


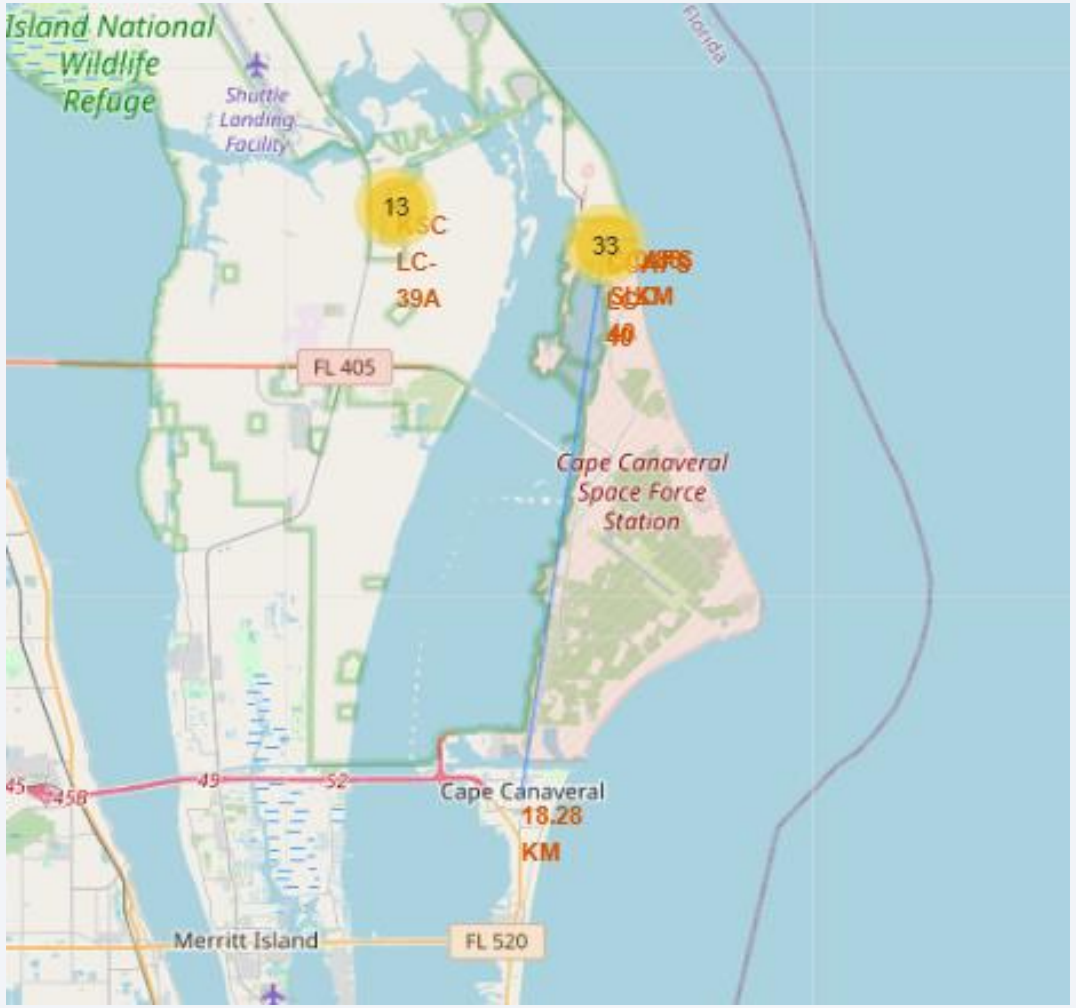
Folium Map – Cluster Success / Fail

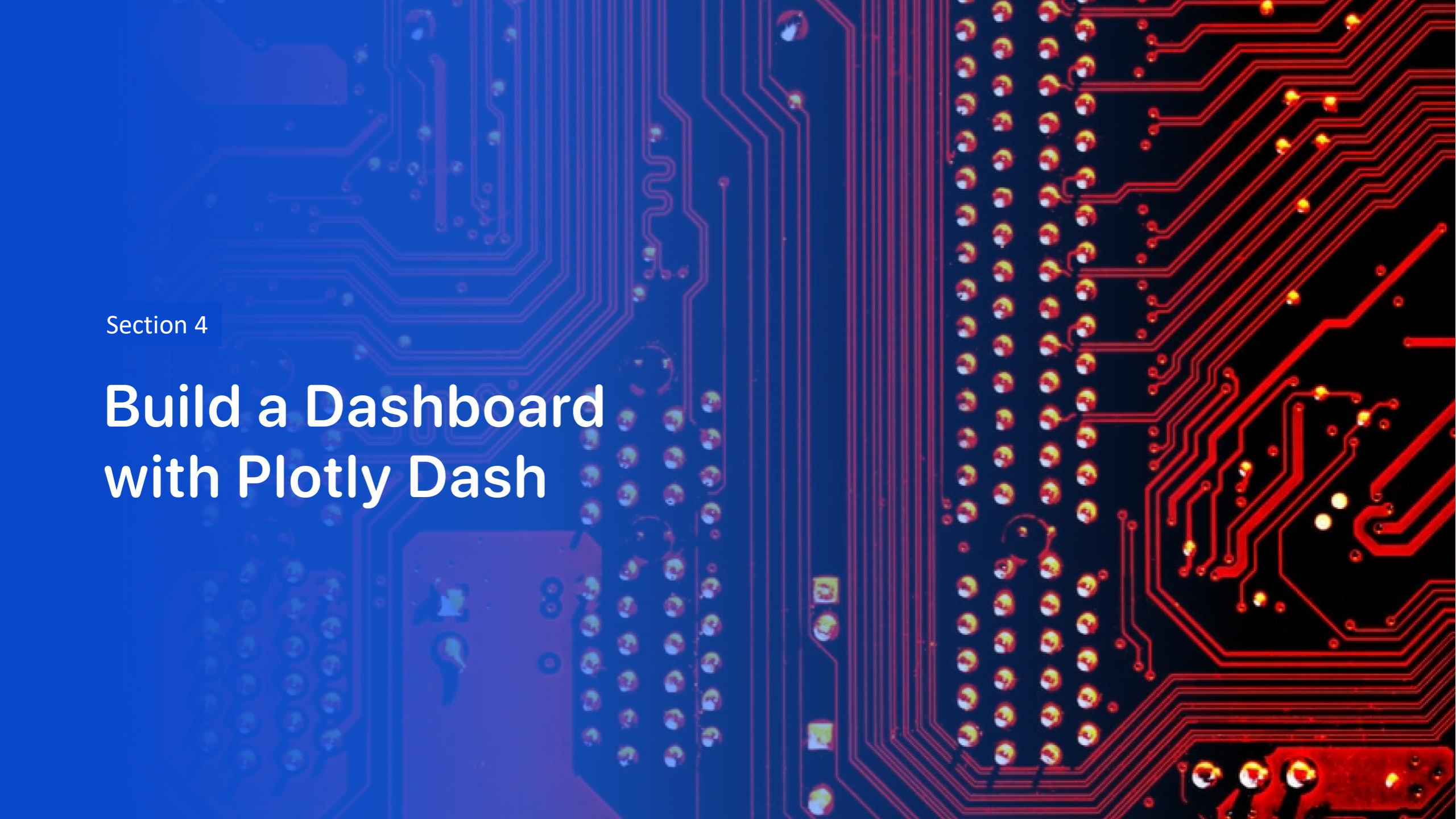




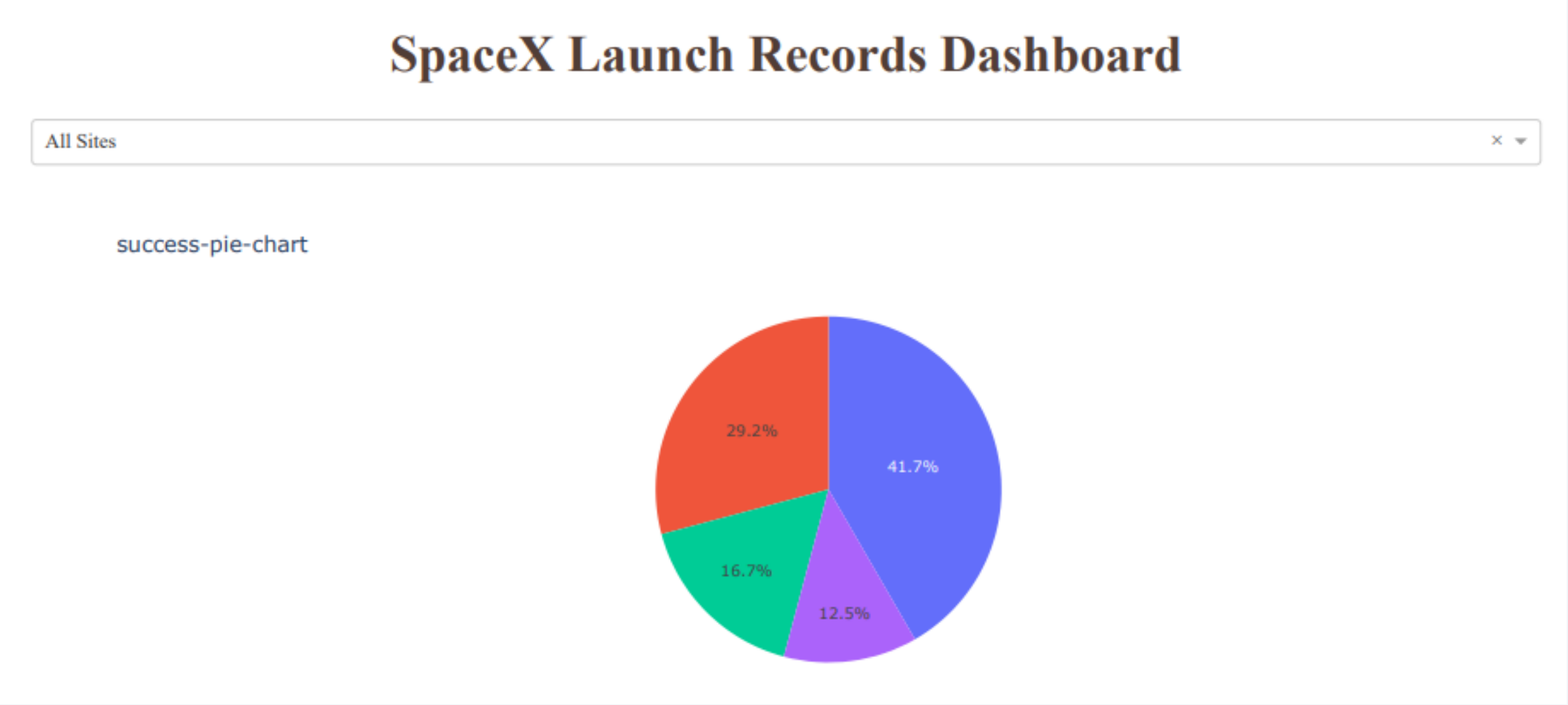
Folium Map – Distances from de Sites



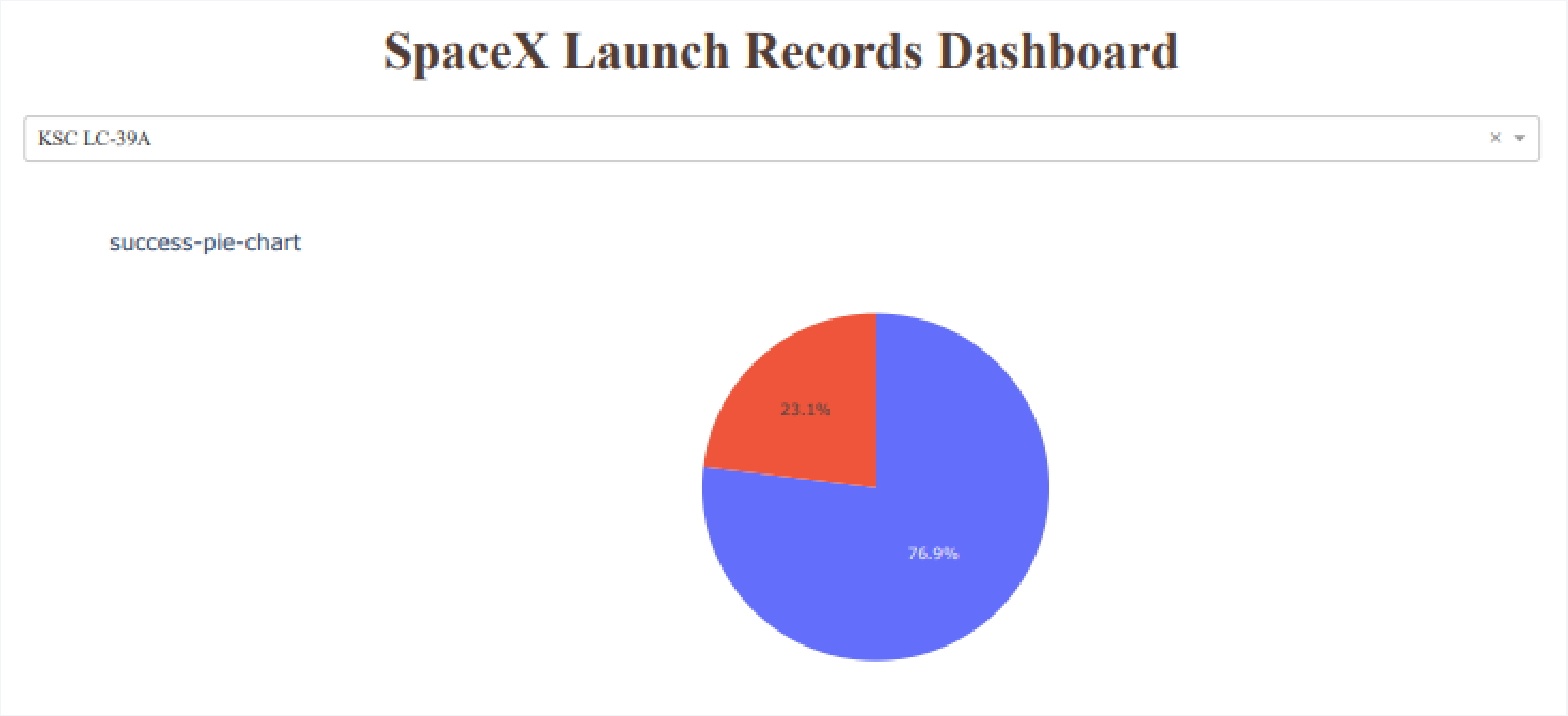



Section 4
Build a Dashboard with Plotly Dash

Dashboard – Success Pie - All



Dashboard – Success Pie - KSC LC-39A



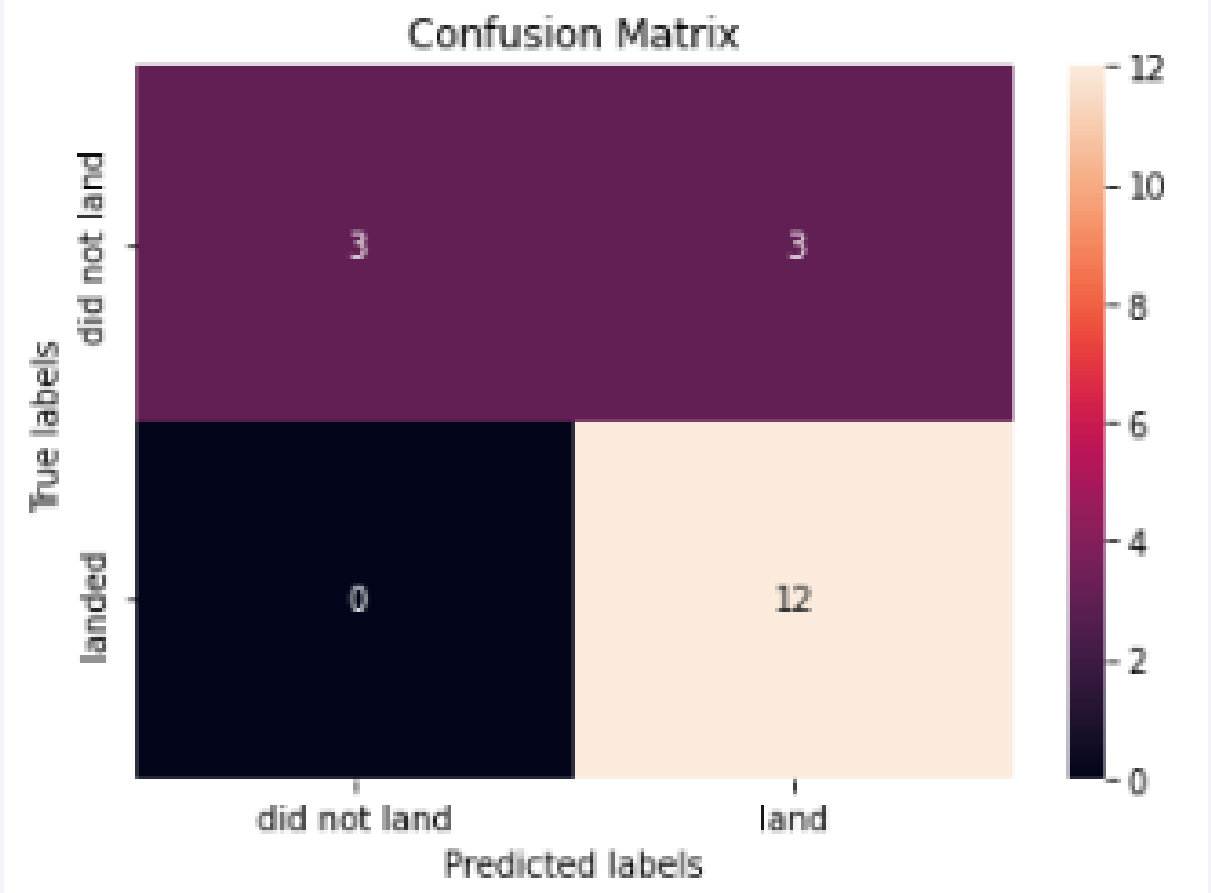
Dashboard – Payload - All

Section 5

Predictive Analysis (Classification)

Confusion Matrix

- Show the confusion matrix of the best performing model with an explanation



Conclusions

- All the models builted have the same accuracy of 83,3%

Thank you!