IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Coco Kan
2022.02.14

Outline

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

Executive Summary

- Perform data collection and wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Introduction

- 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

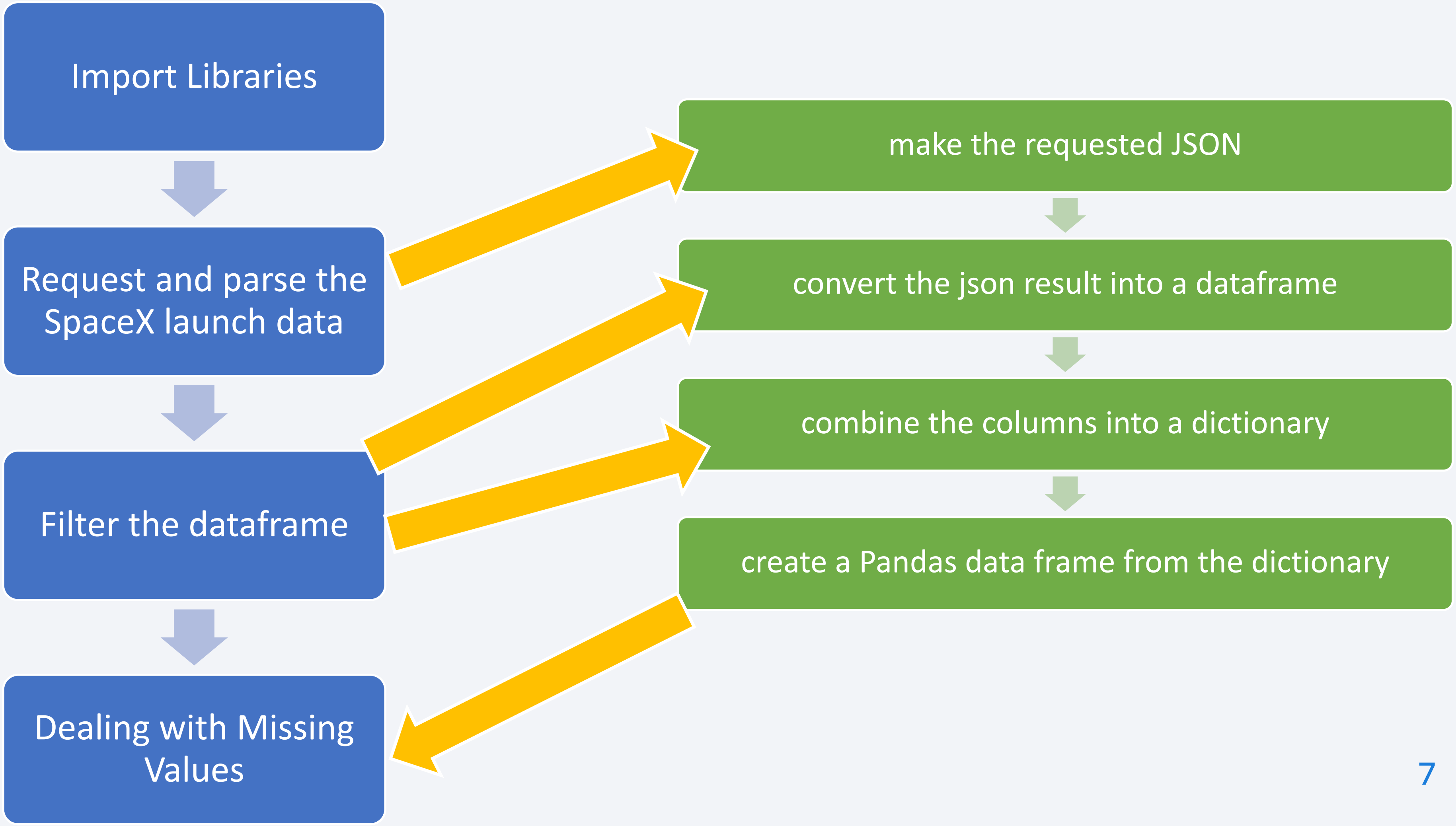
Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data Scraping
- Perform data wrangling
  - Calculate number of outcome and create new dataframe
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Build Logistic Regression, Support Vector Machine, Decision Tree, K-Nearest Neighbor

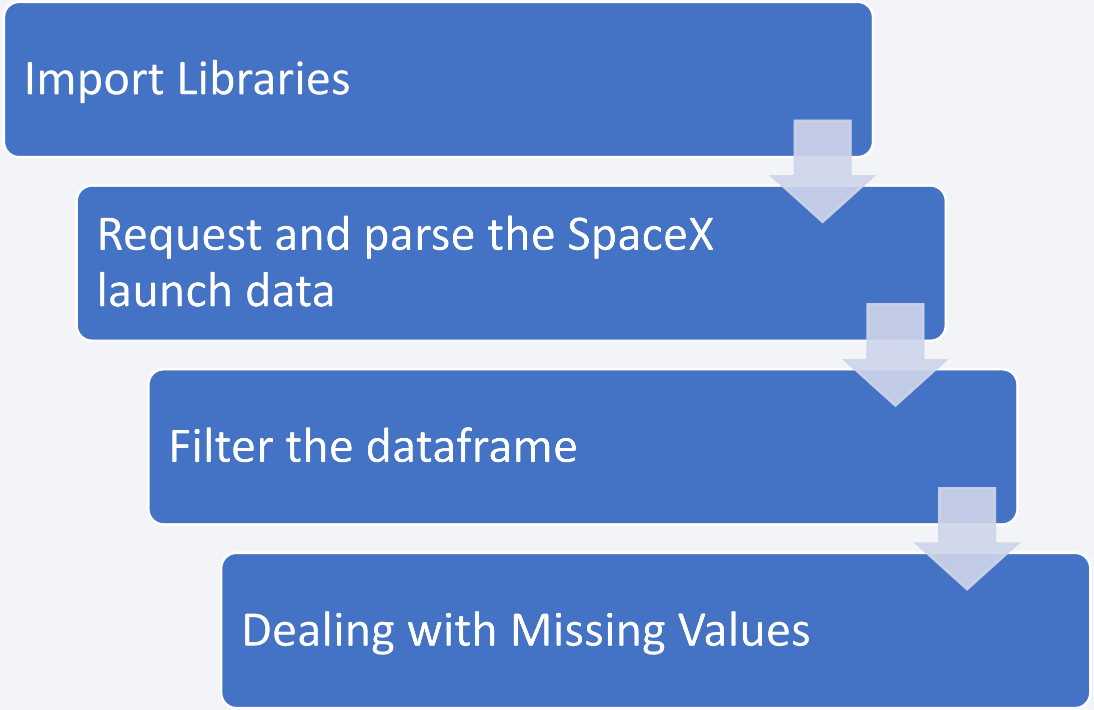
Data Collection



Data Collection – SpaceX API

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

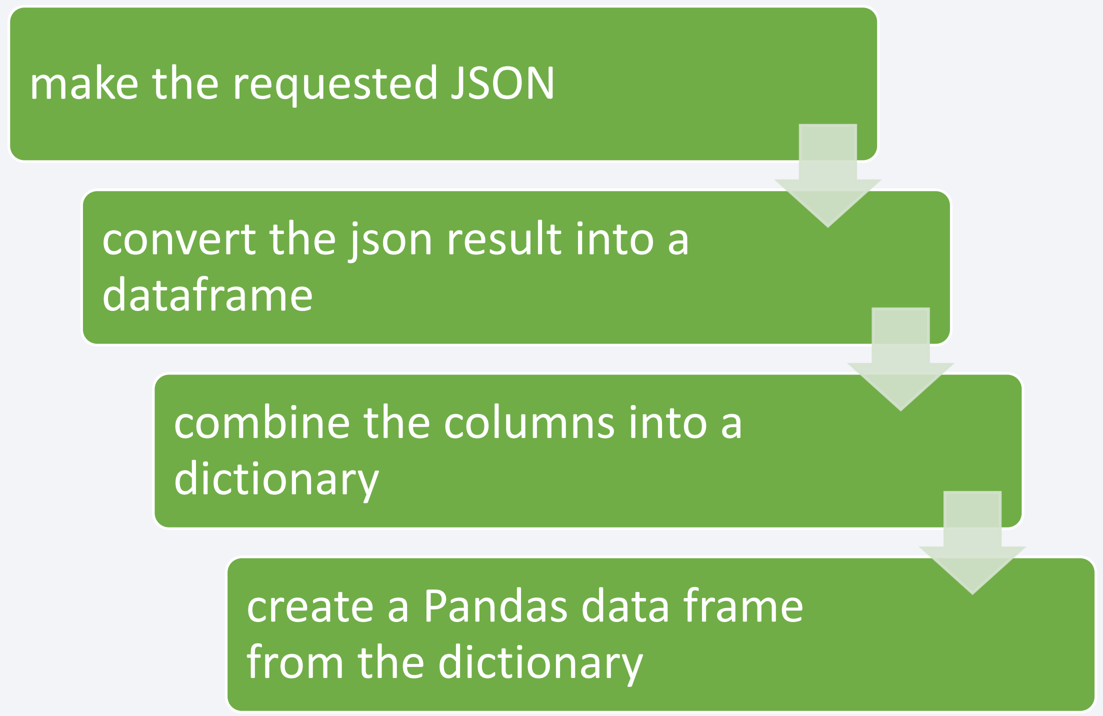
- https://github.com/cocokan05/IBM_DATA_SCIENCE_CERTIFICATE/blob/main/Complete%20the%20Data%20Collection%20API%20Lab.ipynb



Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

- https://github.com/cocokan05/IBM_DATA_SCIENCE_CERTIFICATE/blob/main/Complete%20the%20Data%20Collection%20API%20Lab.ipynb



Data Wrangling

Calculate the number of launches on each site

Calculate the number and occurrence of each orbit

Calculate the number and occurence of mission outcome per orbit type

Create a landing outcome label from Outcome column

- https://github.com/cocokan05/IBM_DATA_SCIENCE_CERTIFICATE/blob/main/Complete%20the%20EDA%20lab%20-Data%20wrangling.ipynb

EDA with Data Visualization

- Scatter plot
- Bar plot
- Line plot
- https://github.com/cocokan05/IBM_DATA_SCIENCE_CERTIFICATE/blob/main/Complete%20the%20EDA%20with%20Visualization%20lab.ipynb

EDA with SQL

- 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)
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first successful 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
- List the total number of successful and failure mission outcomes
- List the names of the booster_versions which have carried the maximum payload mass.
- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- 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
- https://github.com/cocokan05/IBM_DATA_SCIENCE_CERTIFICATE/blob/main/Complete%20the%20EDA%20with%20SQL%20lab.ipynb

Build an Interactive Map with Folium

- Mark all launch sites on a map
- Mark the success/failed launches for each site on the map
- Calculate the distances between a launch site to its proximities
- https://github.com/cocokan05/IBM_DATA_SCIENCE_CERTIFICATE/blob/main/Complete%20the%20Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb

Build a Dashboard with Plotly Dash

- Pie chart
- Scatter plot
- https://github.com/cocokan05/IBM_DATA_SCIENCE_CERTIFICATE/blob/main/Build%20an%20Interactive%20Dashboard%20with%20Ploty%20Dash.ipynb

Predictive Analysis (Classification)

- Build Logistic Regression
- Support Vector Machine
- Decision Tree
- K-Nearest Neighbor
- https://github.com/cocokan05/IBM_DATA_SCIENCE_CERTIFICATE/blob/main/Complete%20the%20Machine%20Learning%20Prediction%20lab.ipynb

Results

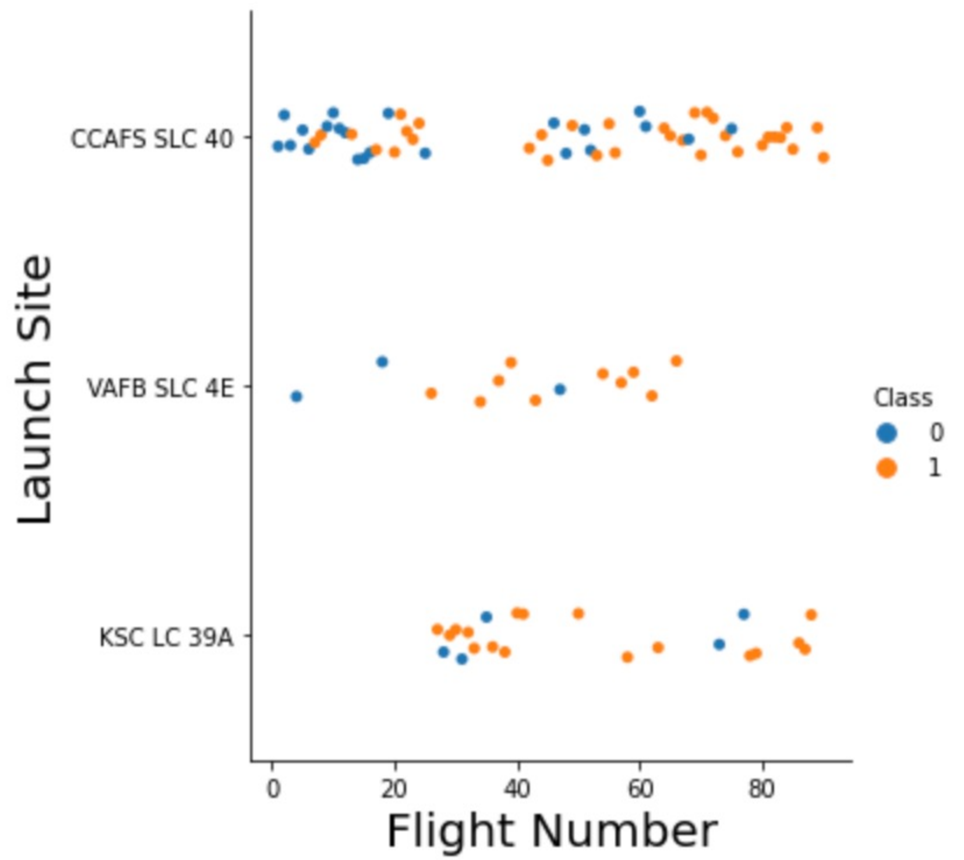
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

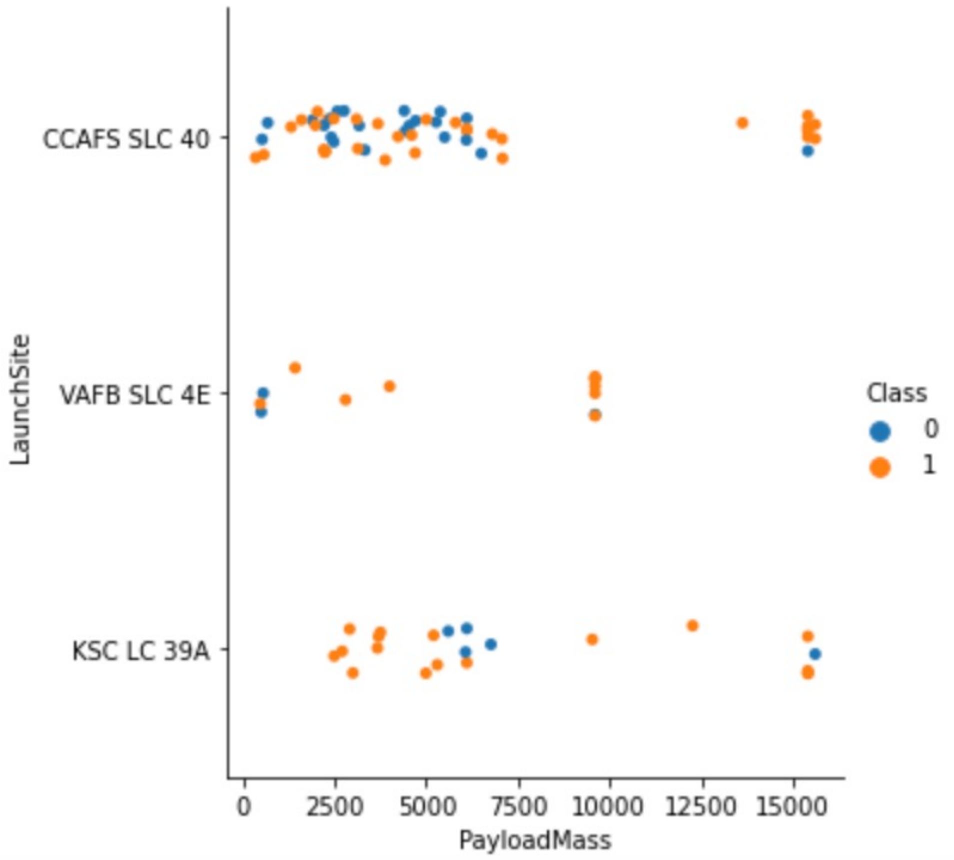
```
sns.catplot(y="LaunchSite", x="FlightNumber", hue="Class", data=df)
plt.xlabel("Flight Number",fontsize=20)
plt.ylabel("Launch Site",fontsize=20)
plt.show()
```



Payload vs. Launch Site

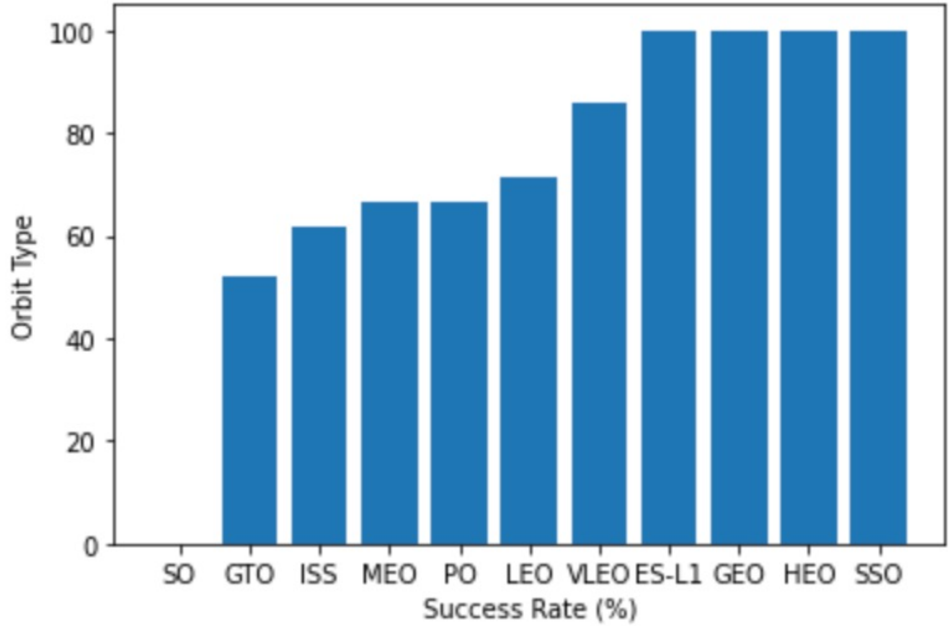
```
sns.catplot(y="LaunchSite", x="PayloadMass", hue="Class", data=df)
```

```
<seaborn.axisgrid.FacetGrid at 0x7f84560824f0>
```



Success Rate vs. Orbit Type

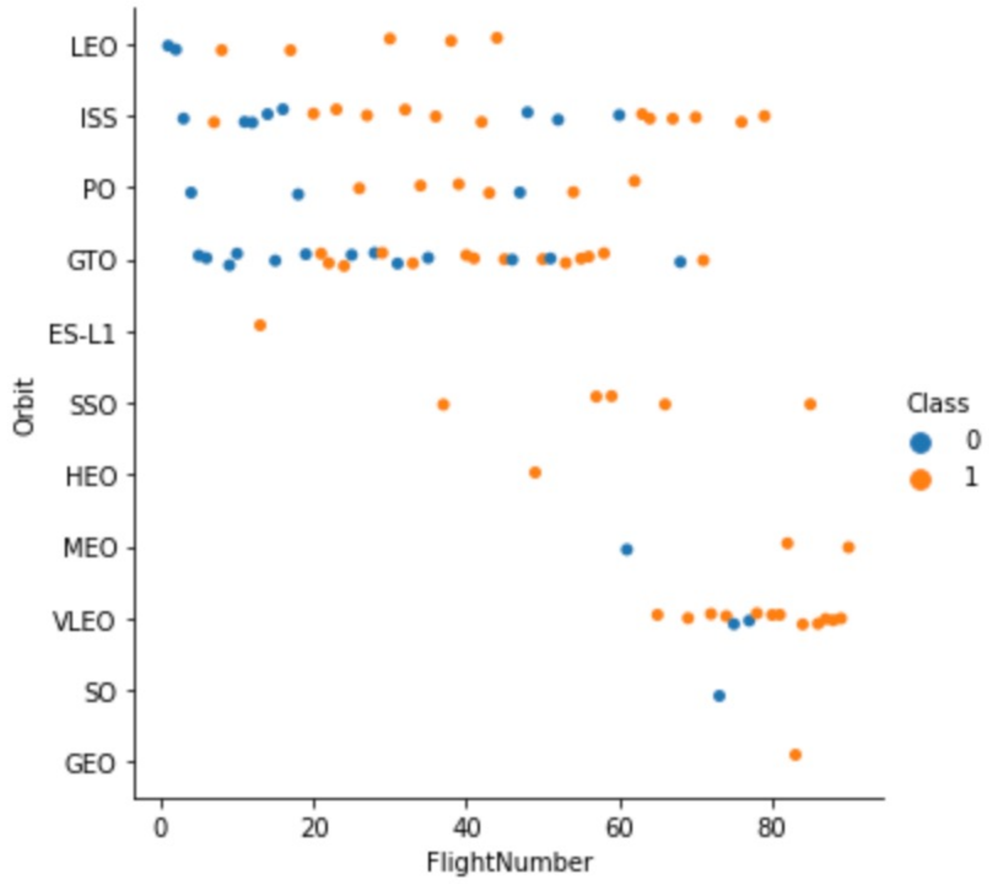
```
df_sorted = df.groupby('Orbit').mean()['Class'].reset_index().sort_values(['Class'], ascending=True)

fig, ax = plt.subplots()

ax.bar(df_sorted.Orbit, df_sorted.Class * 100)

plt.xlabel('Success Rate (%)')
plt.ylabel('Orbit Type')
plt.show()
```



Flight Number vs. Orbit Type

```
sns.catplot(y="Orbit", x="FlightNumber", hue="Class", data=df)
```

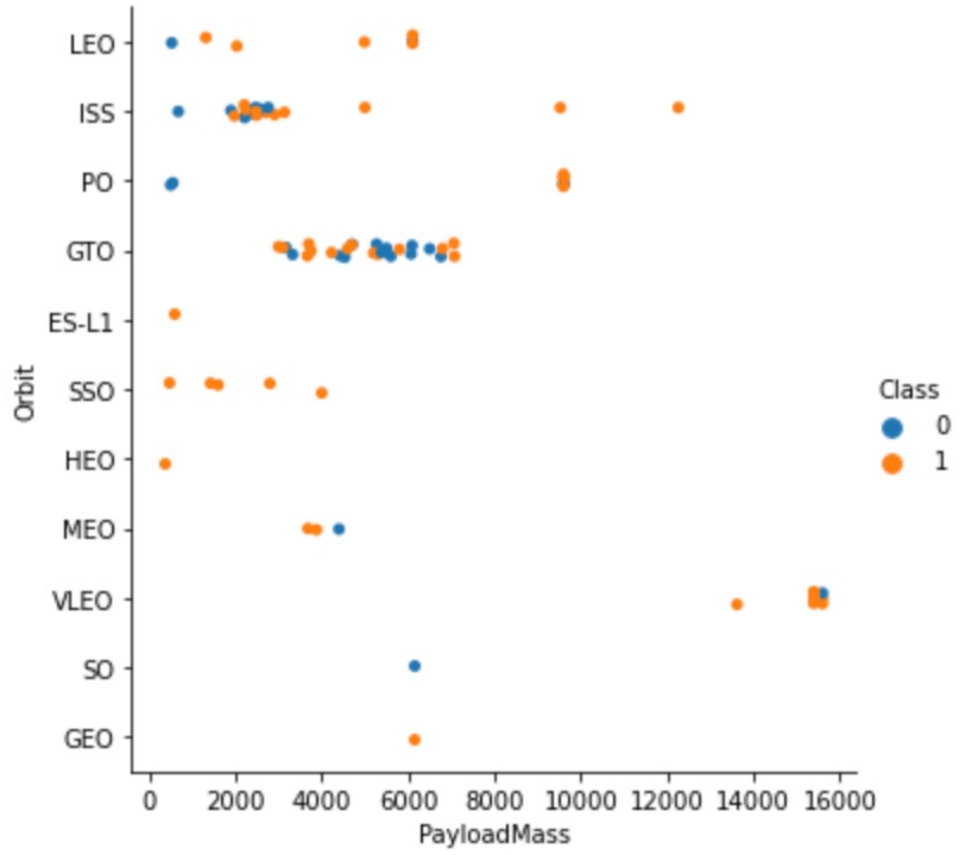
```
<seaborn.axisgrid.FacetGrid at 0x7f84562dde80>
```



Payload vs. Orbit Type

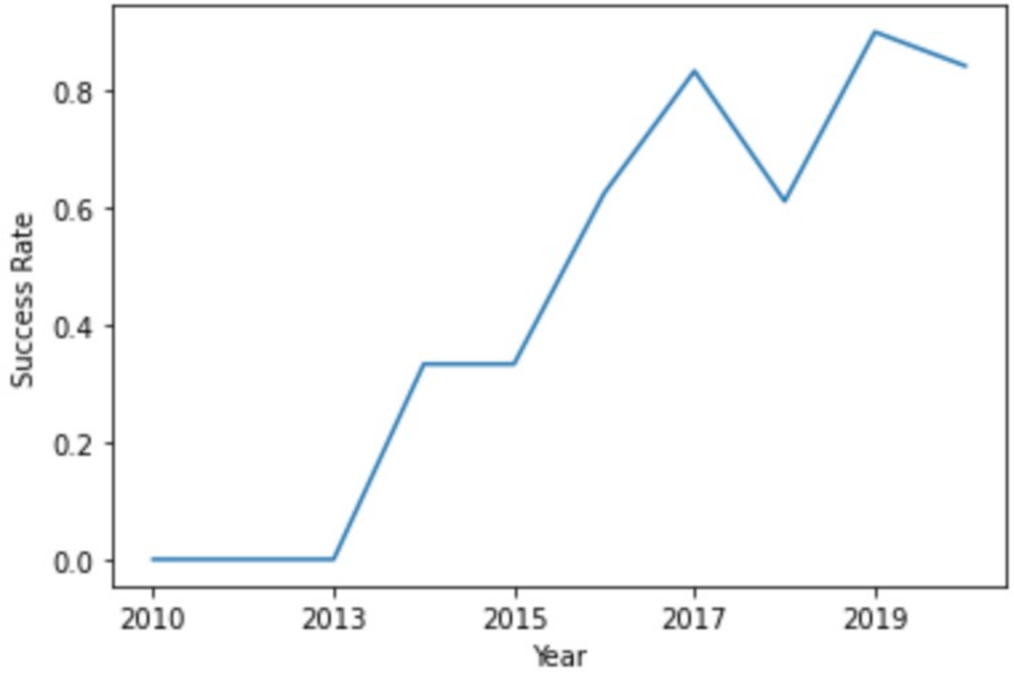
```
sns.catplot(y="Orbit", x="PayloadMass", hue="Class", data=df)
```

```
<seaborn.axisgrid.FacetGrid at 0x7f84562da070>
```



Launch Success Yearly Trend

```
df.groupby(Extract_year(df['Date'])).mean()['Class'].plot(kind='line')
plt.xlabel('Year')
plt.ylabel('Success Rate')
plt.show()
```



All Launch Site Names

- Find the names of the unique launch sites

```
%%sql
SELECT DISTINCT Launch_Site
FROM SPACEXTBL;
```

```
 * sqlite://
Done.
```

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`

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

```
 * sqlite://
Done.
```

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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	0:35:00	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

```
%%sql
SELECT SUM(PAYLOAD_MASS__KG_)
FROM SPACEXTBL
WHERE Customer='NASA (CRS)';
```

```
 * sqlite://
Done.
```

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

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

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

```
 * sqlite://
Done.
```

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

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

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

```
 * sqlite://
Done.
```

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

```
%%sql
SELECT Booster_Version
FROM SPACEXTBL
WHERE (Landing_Outcome='Success (drone ship)') AND (PAYLOAD_MASS__KG_ BETWEEN 4000 and 6000);
```

```
 * sqlite://
Done.
```

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

```
%%sql
SELECT Mission_Outcome, COUNT(Mission_Outcome) AS Number
FROM SPACEXTBL
GROUP BY Mission_Outcome
```

```
 * sqlite://
Done.
```

Mission_Outcome	Number
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

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

```
%%sql
SELECT Booster_Version, PAYLOAD_MASS__KG_
FROM SPACEXTBL
WHERE PAYLOAD_MASS__KG_= (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL)
```

```
 * sqlite://
Done.
```

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

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

```
%%sql
SELECT Booster_Version, Launch_Site
FROM SPACEXTBL
WHERE (Landing_Outcome='Failure (drone ship)') AND (YEAR(Date)='2015');
```

```
 * sqlite://
```

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
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

```
%%sql
SELECT Landing_Outcome, COUNT(Landing_Outcome) AS Number
FROM SPACEXTBL
WHERE Date BETWEEN '2010-06-04' and '2017-03-20'
GROUP BY Landing_Outcome
ORDER BY COUNT(Landing_Outcome) DESC
```

```
 * sqlite://
Done.
```

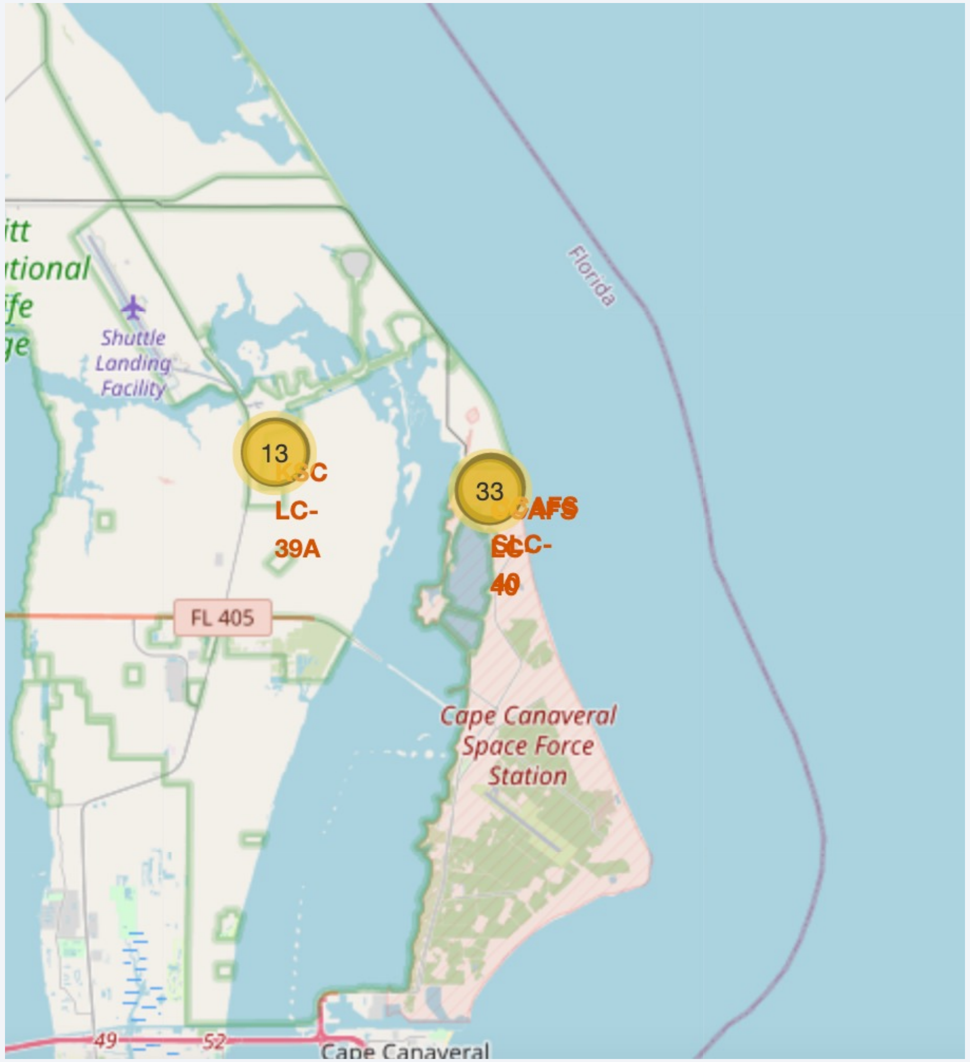
landing__outcome	total_number
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

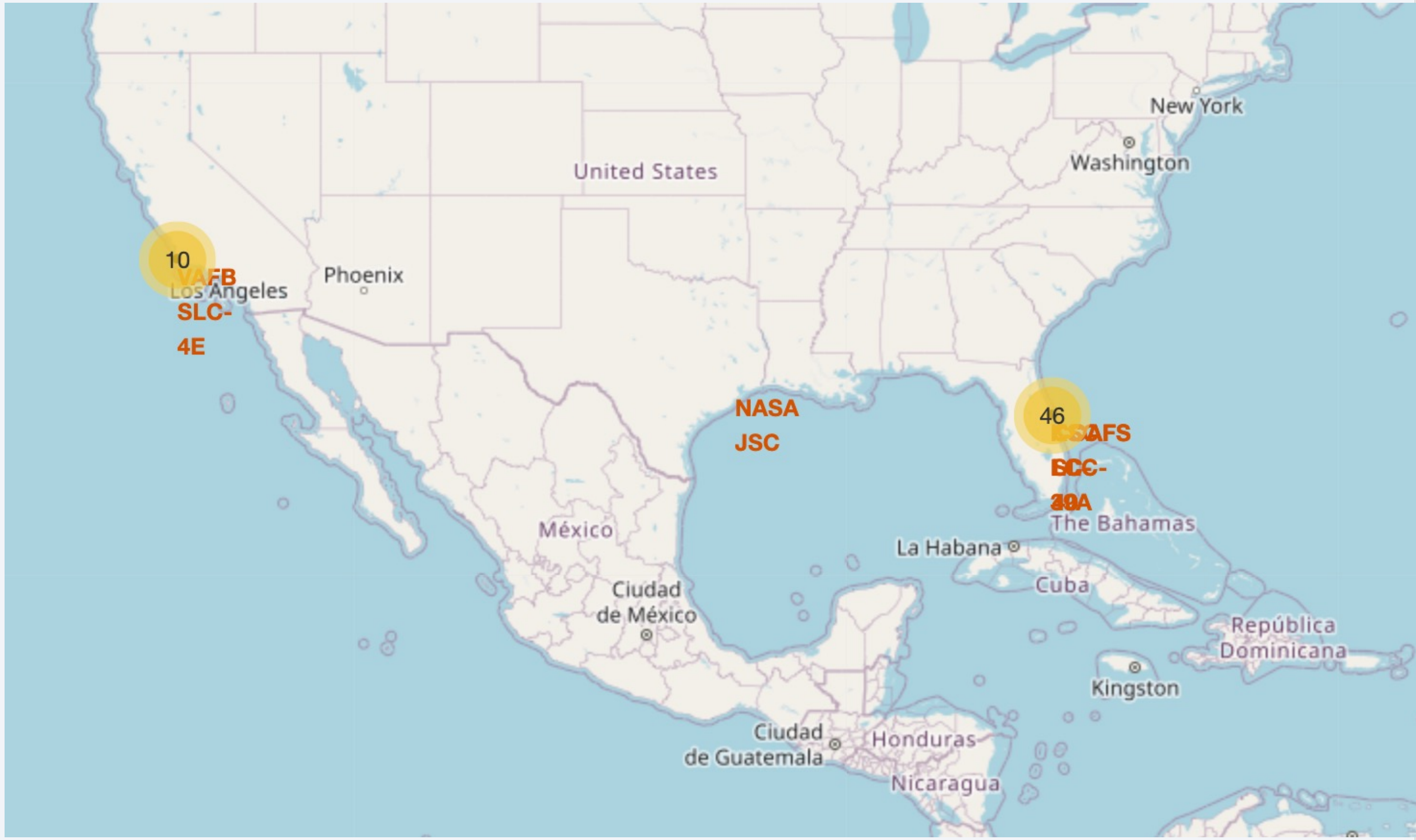
Section 3

Launch Sites Proximities Analysis

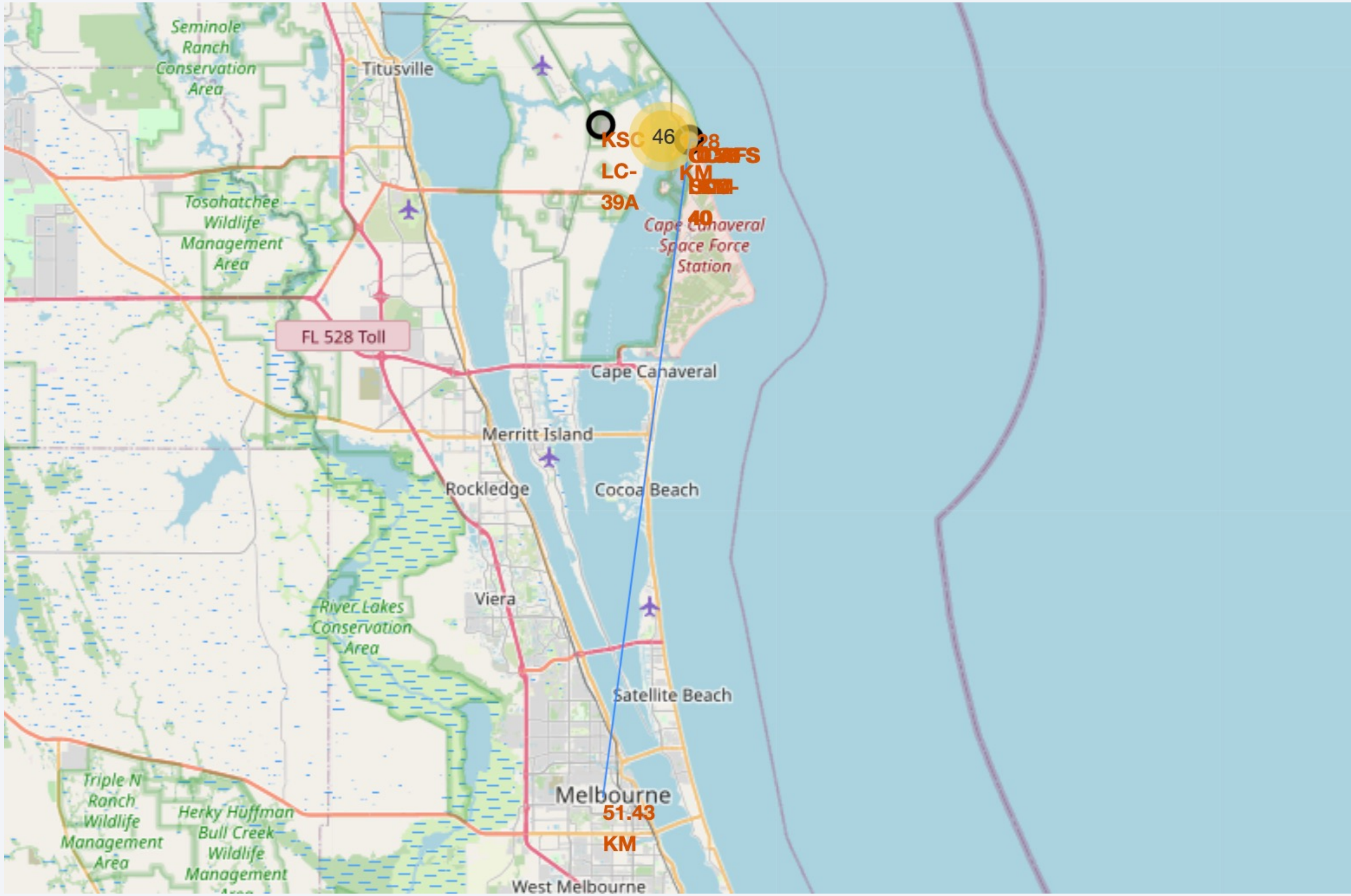
All Launch Site on Map

The success/failed launches for each site on the map





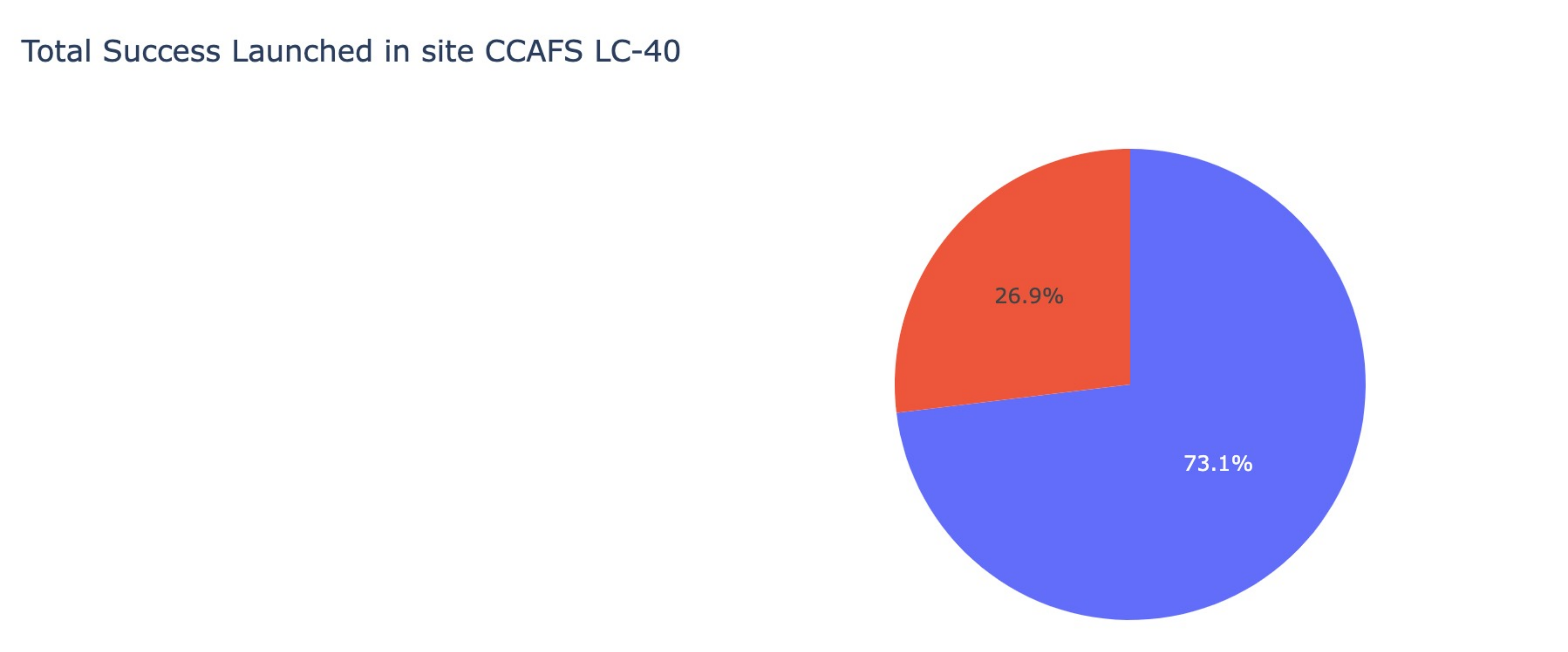
The distance between a launch site to its closest city



Section 4

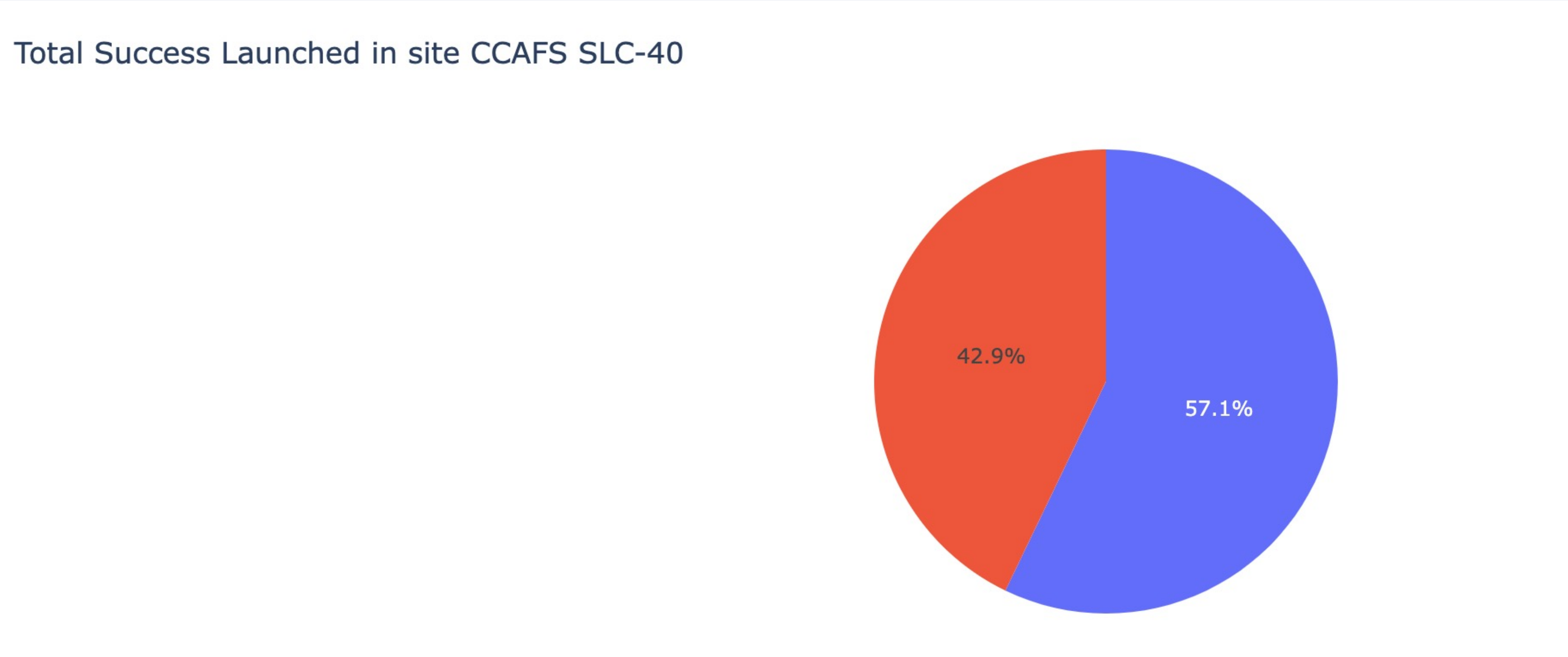
Build a Dashboard with Plotly Dash

Total Success Launched in site CCAFS LC-40 in Pie chart



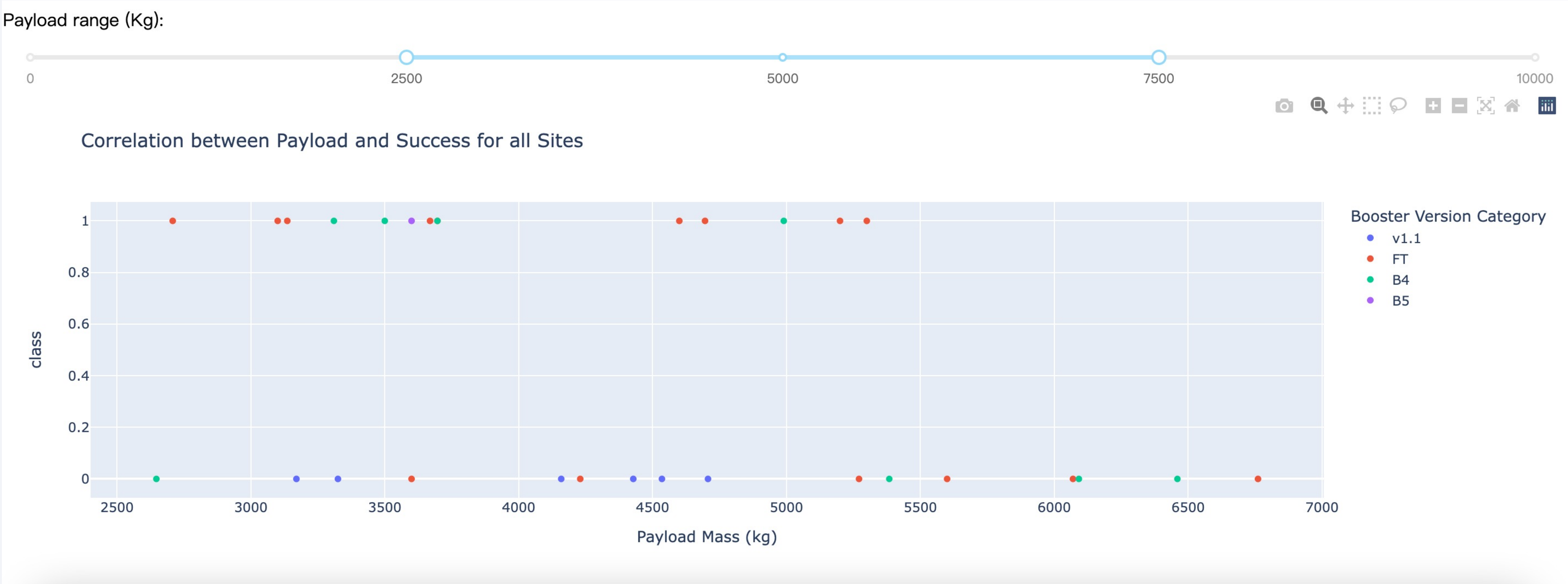
The launch site with highest launch success ratio

- The highest launch success ratio 42.9% in site CCAFS SLC-40



Payload vs. Launch Outcome in Payload between 5000-7500

- Where Class 1=Success, Class 0= Failure

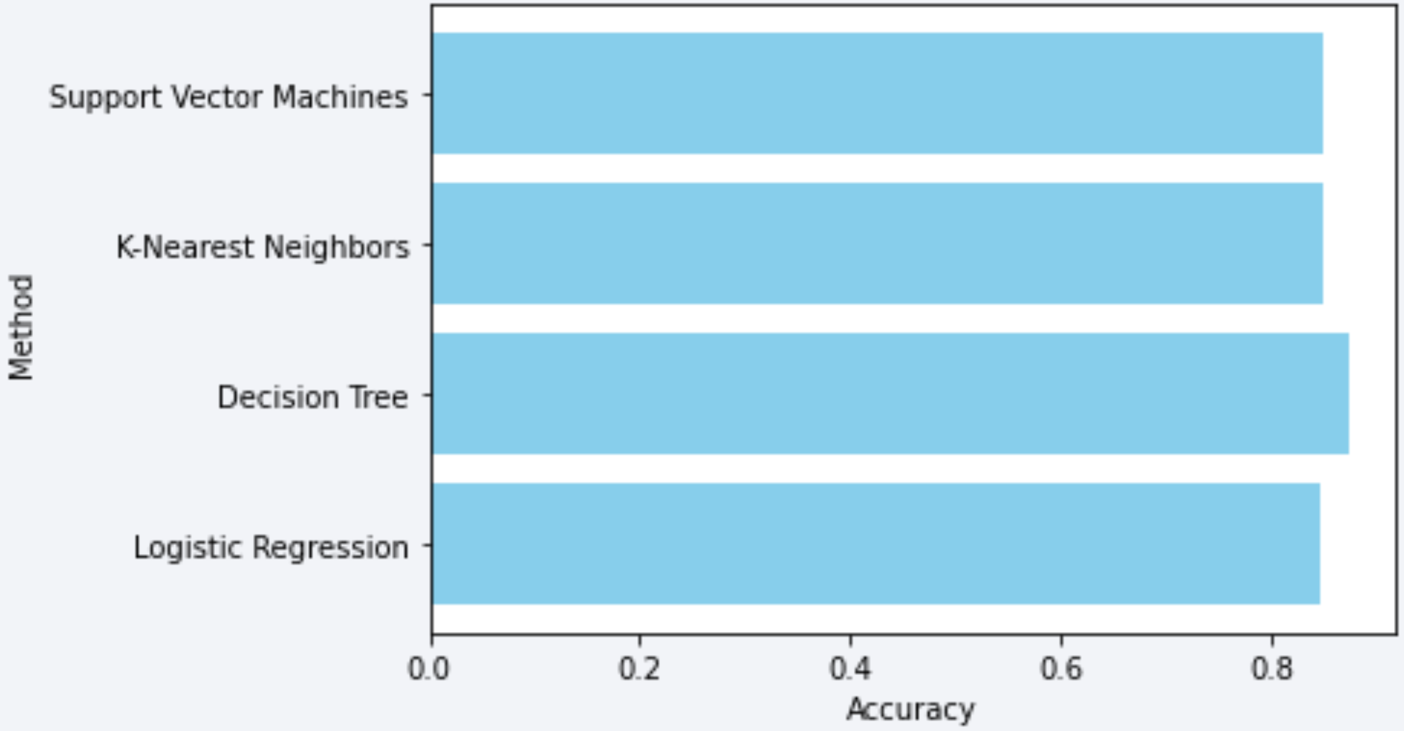


Section 5

Predictive Analysis (Classification)

Classification Accuracy

- The decision tree method has the highest accuracy= 0.875



Confusion Matrix

- The confusion matrix of decision tree

Conclusions

- Launch success ratio rise as year increase
- Site CCAFS SLC-40 has the highest launch success ratio
- The decision tree method has the highest accuracy

Appendix

- File for all notebooks:

  https://github.com/cocokan05/IBM_DATA_SCIENCE_CERTIFICATE

Thank you!