Winning Space Race with Data Science

Pranav Sonawane
March 21st , 2022

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection & wrangling
  - EDA with data visualization
  - EDA with SQL
  - Building a map with Folium and dashboard with Potly
  - Predictive Analysis (Classification)
- Summary of all results
  - EDA results
  - Interactive analytical demo
  - Classification model

Introduction

- Project background and context
  - The commercial space age is here, companies are making space travel affordable for everyone. the most successful company is SpaceX. SpaceX can do this by reusing the first stage makes rocket launches are relatively inexpensive. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch
- Problems you want to find answers
  - Determine the first stage of Falcon 9 will land successfully ?
  - Determine the price of each launch.
  - Determine the mission parameter like Payload, orbit, etc.

Section 1

Methodology

Methodology

Executive Summary

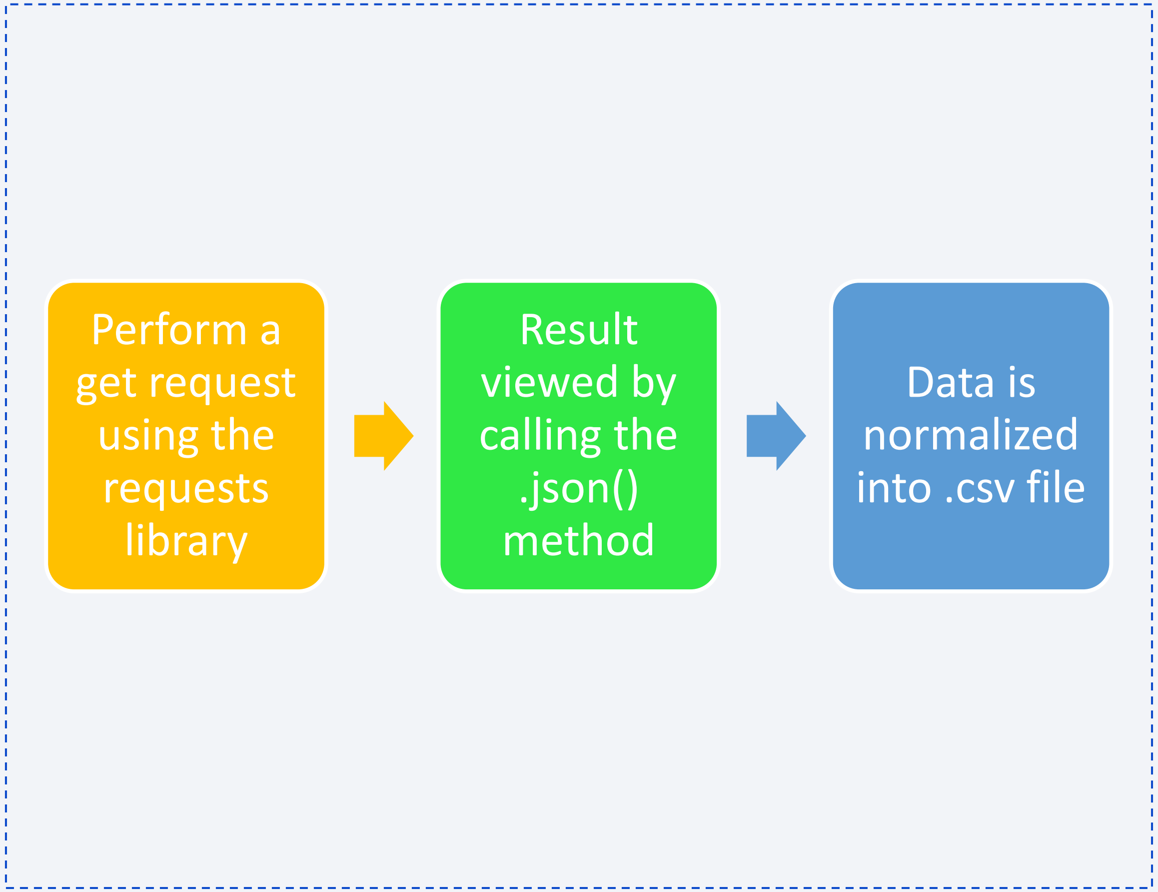
- Data collection methodology:
  - SpaceX launch data is gathered from an SpaceX REST API and Web scraping of wiki page about Falcon 9
- Perform data wrangling
  - Data was filter for Falcon 9 rocket only, dealing with null values.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection

- Data has been collected from two data Source.
    - SpaceX REST API
    - Web Scraping from Wikipedia Page of Falcon 9

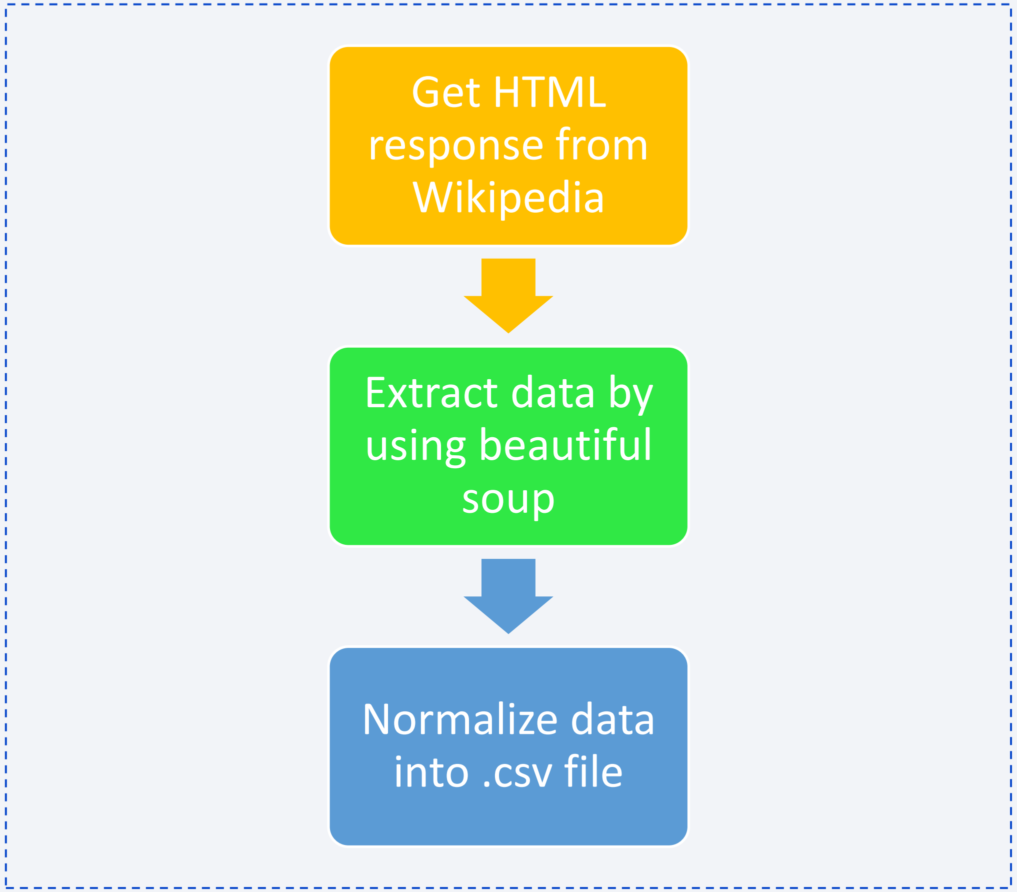
Data Collection – SpaceX API

- API gives data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome.
- Add the GitHub URL of the completed SpaceX API calls notebook (must include completed code cell and outcome cell), as an external reference and peer-review purpose



Data Collection - Scraping

- Web scraping gives data about flight, date, booster version used, orbit, outcomes, grid fines, serial and launching data
- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose



Data Wrangling

- Data was filter for Falcon 9 rocket only, data further explore to see some pattern and determine what would be the label for training supervised models.
- Also, outcomes converted into Training Labels with 1 means the booster successfully landed 0 means it was unsuccessful.
- This replicates Falcon 9 first stage will land successfully
- Add the GitHub URL of your completed data wrangling related notebooks, as an external reference and peer-review purpose

EDA with Data Visualization

- Scatter plot, Bar plot and Line chart has been used
- Scatter plot is used to determine the relationship between FlightNumber, PayloadMass, Launch Site, Orbit.
- Bar plot is used to determine most success rate orbit type
- Line chart is used to identify trend of success of launch
- Add the GitHub URL of your completed EDA with data visualization notebook, as an external reference and peer-review purpose

EDA with SQL

- SQL queries perform on cloud dataset
  - 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 achieved
  - 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. Use a subquery
  - 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
- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose

Build an Interactive Map with Folium

- Determine the pattern of success/failure of launchsite interactive map objects such as markers, circles, lines were created and added to a folium map
- To check the pattern of success at launch site marker were added with success in green and failure as red
- Circles are used to locate launchsite with longitude and latitude data
- Lines are used to plot distance between launch site and highway, city, coastline, railway
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

Build a Dashboard with Plotly Dash

- To show the success or failure outcome for payload mass and launchsite is displayed with pie chart and scatter plot
- Pie chart gives the success & failure measure for launch sites
- Scatter plot display the distribution of payload mass with outcome
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Predictive Analysis (Classification)

Preprocessing Model

- Load the clean dataset.
- Find the feature set and target set from dataset
- Convert them into numpy array
- Use train test split evaluation approach

Building Model

- Decide which type of ml algorithm is best fit for problem statement
- set parameter and algorithm to GridSearchCV
- Fit the our dataset and train the model

Evaluating Model

- Check accuracy for each model and plot confusion matrix
- Model with best accuracy is best fit for dataset

- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose

Results

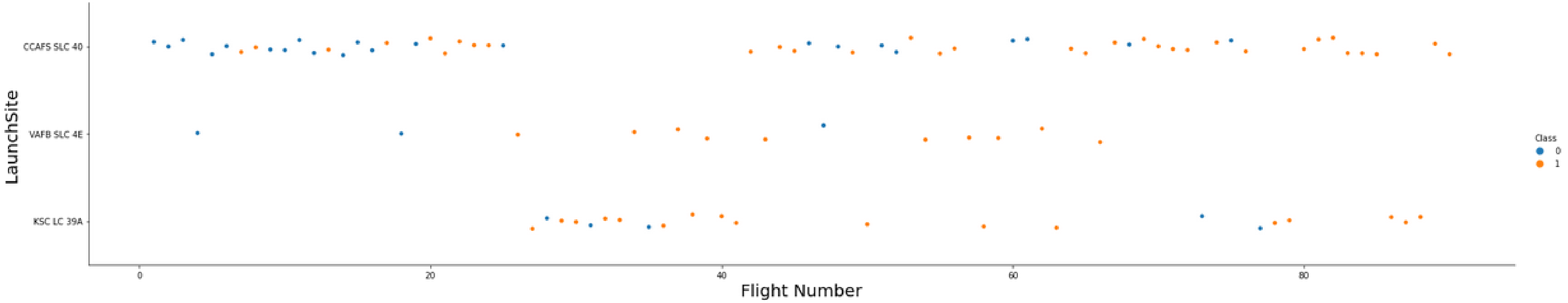
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

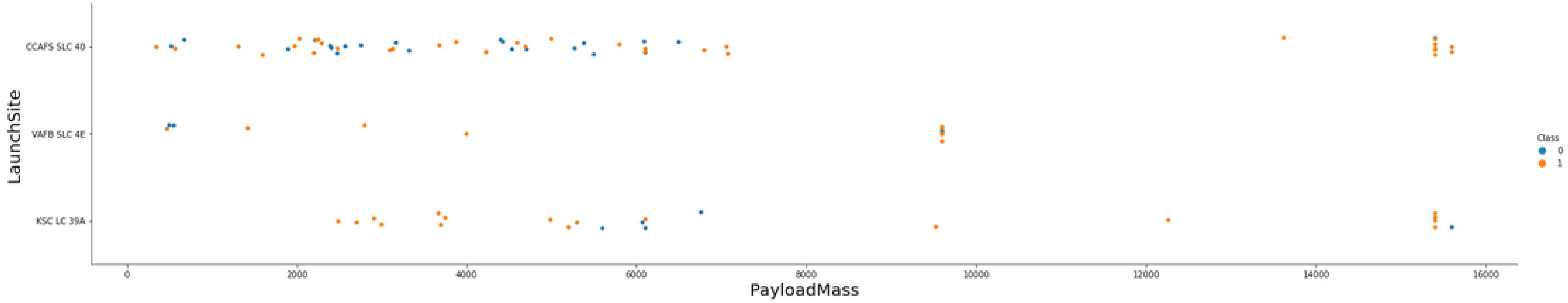
In [4]:
```
# Plot a scatter point chart with x axis to be Flight Number and y axis to be the launch site, and hue to be the class value
sns.catplot(y="LaunchSite", x="FlightNumber", hue="Class", data=df, aspect = 5)
plt.xlabel("Flight Number",fontsize=20)
plt.ylabel("LaunchSite",fontsize=20)
plt.show()
```



KSC LC – 39A Launch site has most successful launches

Payload vs. Launch Site

In [5]:
```
# Plot a scatter point chart with x axis to be Pay Load Mass (kg) and y axis to be the launch site, and hue to be the c
lass value
sns.catplot(y="LaunchSite", x="PayloadMass", hue="Class", data=df, aspect = 5)
plt.xlabel("PayloadMass",fontsize=20)
plt.ylabel("LaunchSite",fontsize=20)
plt.show()
```

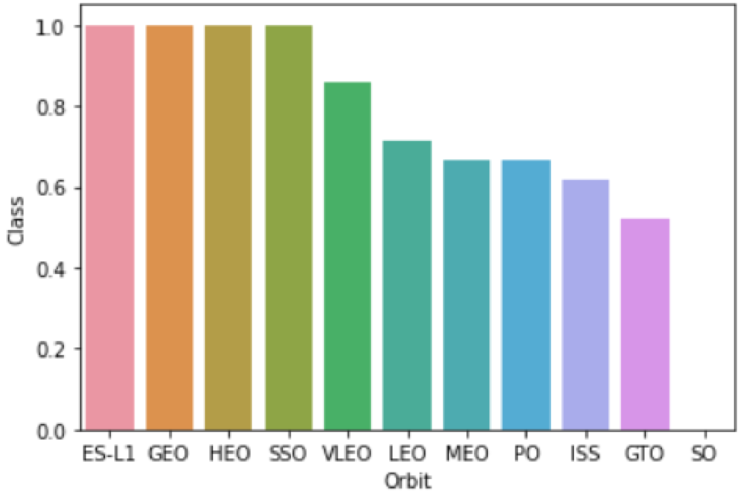


Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

Success Rate vs. Orbit Type

In [6]:
```
# HINT use groupby method on Orbit column and get the mean of Class column
orbit_ = df.groupby(['Orbit'])['Class'].mean().sort_values(ascending=False).reset_index()
sns.barplot(y="Class", x="Orbit", data=orbit_)
```

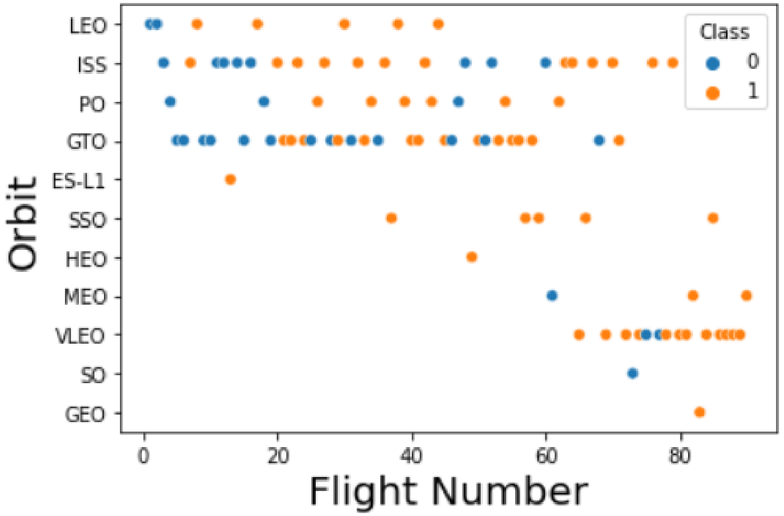
Out[6]: <AxesSubplot:xlabel='Orbit', ylabel='Class'>



ES-L1, GEO, HEO, SSO have highest success rate

Flight Number vs. Orbit Type

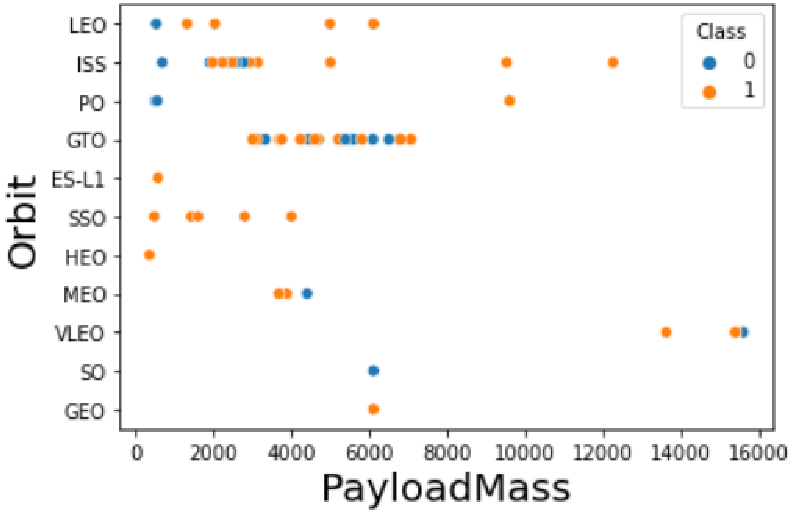
In [9]:
```
# Plot a scatter point chart with x axis to be FlightNumber and y axis to be the Orbit, and hue to be the class value
sns.scatterplot(y="Orbit", x="FlightNumber", hue="Class", data=df)
plt.xlabel("Flight Number",fontsize=20)
plt.ylabel("Orbit",fontsize=20)
plt.show()
```



You should see that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type

In [10]:
```
# Plot a scatter point chart with x axis to be Payload and y axis to be the Orbit, and hue to be the class value
sns.scatterplot(y="Orbit", x="PayloadMass", hue="Class", data=df)
plt.xlabel("PayloadMass",fontsize=20)
plt.ylabel("Orbit",fontsize=20)
plt.show()
```

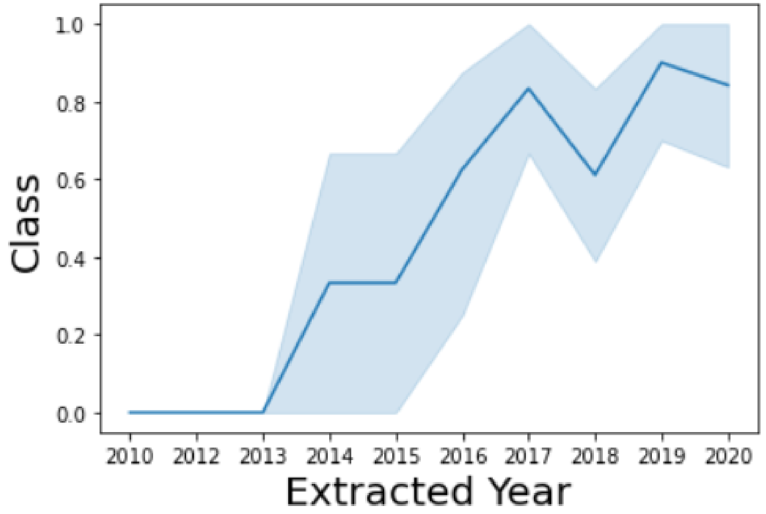


With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.

Launch Success Yearly Trend

In [35]:
```
sns.lineplot(y="Class", x="Date", data=df_copy)
plt.xlabel("Extracted Year",fontsize=20)
plt.ylabel("Class",fontsize=20)
plt.show()
```



you can observe that the sucess rate since 2013 kept increasing till 2020

All Launch Site Names

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

In [12]:
```
%sql select DISTINCT(launch_site) from SPACEXTBL;
```

```
 * ibm_db_sa://wkx47869:***@824dfd4d-99de-440d-9991-629c01b3832d.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:30119
/bludb
Done.
```

Out[12]:

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

Launch Site Names Begin with 'CCA'

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

In [13]:
```
%sql select * from SPACEXTBL where launch_site Like '%CCA%' limit 5;
```

```
 * ibm_db_sa://wkx47869:***@824dfd4d-99de-440d-9991-629c01b3832d.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:30119
/bludb
Done.
```

Out[13]:

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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

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

In [14]:
```
%sql select SUM(payload_mass__kg_) as  total_payload_mass_NASA_CRS from SPACEXTBL where customer = 'NASA (CRS)';
```

```
 * ibm_db_sa://wkx47869:***@824dfd4d-99de-440d-9991-629c01b3832d.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:30119
/bludb
Done.
```

Out[14]:

total_payload_mass_nasa_crs
45596

Average Payload Mass by F9 v1.1

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

In [15]:

```
%sql select AVG(payload_mass__kg_) as average_payload_mass_F9v1_1 from SPACEXTBL where booster_version like '%F9 v1.
1%';
```

```
 * ibm_db_sa://wkx47869:***@824dfd4d-99de-440d-9991-629c01b3832d.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:30119
/bludb
Done.
```

Out[15]:

average_payload_mass_f9v1_1
2534

First Successful Ground Landing Date

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

Hint:Use min function

In [16]:
```
%sql select MIN(DATE) from SPACEXTBL where landing__outcome = 'Success (ground pad)';
```

```
 * ibm_db_sa://wkx47869:***@824dfd4d-99de-440d-9991-629c01b3832d.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:30119
/bludb
Done.
```

Out[16]:

1
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

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

In [19]:
```
%sql select booster_version from SPACEXTBL where landing__outcome = 'Success (drone ship)' AND payload_mass__kg_ > 4000
AND payload_mass__kg_ < 6000;
```

```
 * ibm_db_sa://wkx47869:***@824dfd4d-99de-440d-9991-629c01b3832d.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:30119
/bludb
Done.
```

Out[19]:

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

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

In [28]:
```
%sql select COUNT(*) as mission_success from SPACEXTBL where mission_outcome LIKE '%Success%';
```

```
 * ibm_db_sa://wkx47869:***@824dfd4d-99de-440d-9991-629c01b3832d.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:30119
/bludb
Done.
```

Out[28]:

mission_success
100

In [29]:
```
%sql select COUNT(*) as mission_failure from SPACEXTBL where mission_outcome LIKE '%Failure%';
```

```
 * ibm_db_sa://wkx47869:***@824dfd4d-99de-440d-9991-629c01b3832d.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:30119
/bludb
Done.
```

Out[29]:

mission_failure
1

Boosters Carried Maximum Payload

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

In [32]:
```
%sql select booster_version from SPACEXTBL where payload_mass__kg_ = (select MAX(payload_mass__kg_) from SPACEXTBL);
```

```
 * ibm_db_sa://wkx47869:***@824dfd4d-99de-440d-9991-629c01b3832d.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:30119
/bludb
Done.
```

Out[32]:

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 [33]:
```
%sql select landing__outcome, booster_version, launch_site from SPACEXTBL where mission_outcome LIKE '%Failure%' AND YE
AR(DATE) = 2015;
```

```
 * ibm_db_sa://wkx47869:***@824dfd4d-99de-440d-9991-629c01b3832d.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:30119
/bludb
Done.
```

Out[33]:

landing__outcome	booster_version	launch_site
Precluded (drone ship)	F9 v1.1 B1018	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

In [40]:
```
%sql select landing__outcome, COUNT(landing__outcome) from SPACEXTBL where DATE BETWEEN '2010-06-04' AND '2017-03-20' g
roup by landing__outcome order by COUNT(landing__outcome) DESC;
```

```
 * ibm_db_sa://wkx47869:***@824dfd4d-99de-440d-9991-629c01b3832d.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:30119
/bludb
Done.
```

Out[40]:

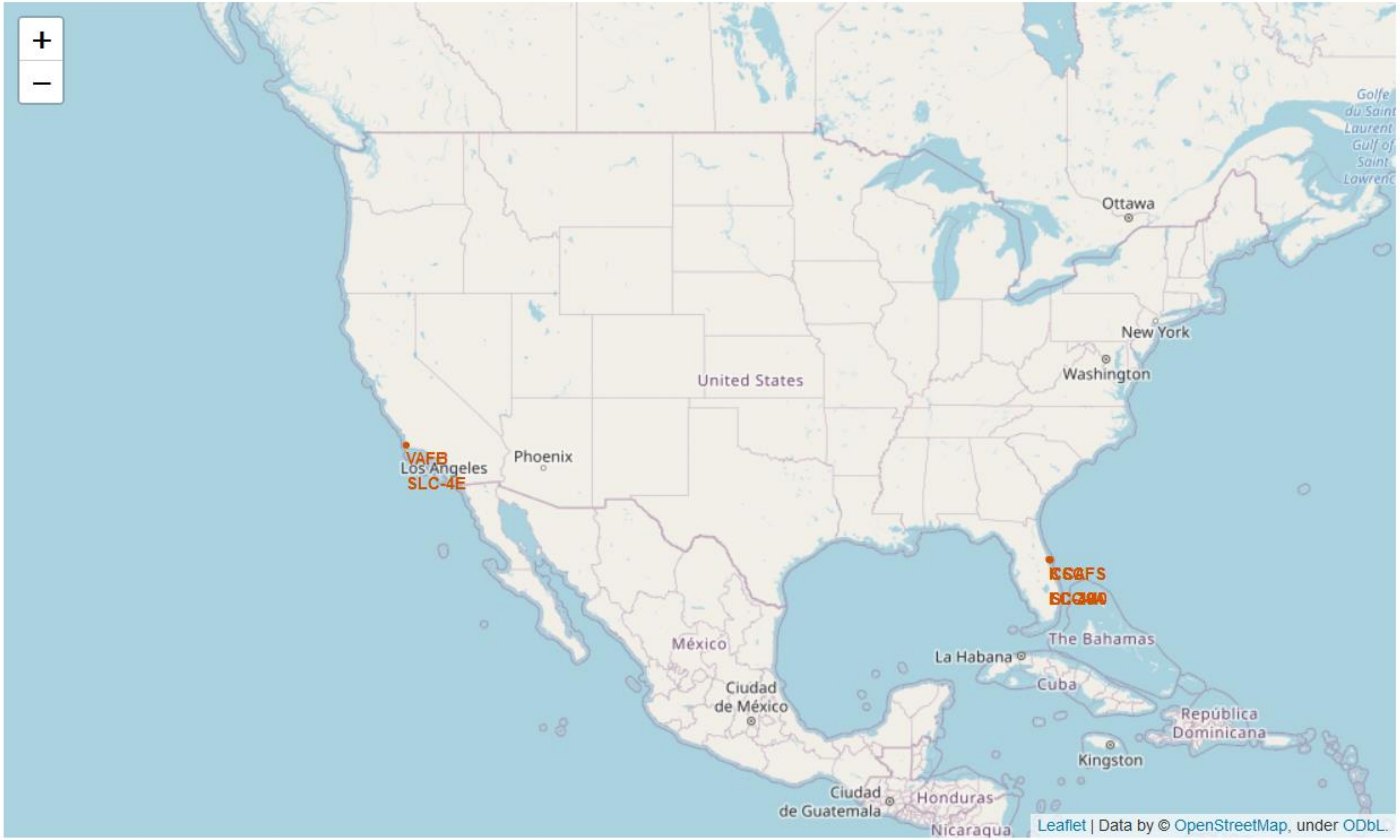
landing__outcome	2
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 3

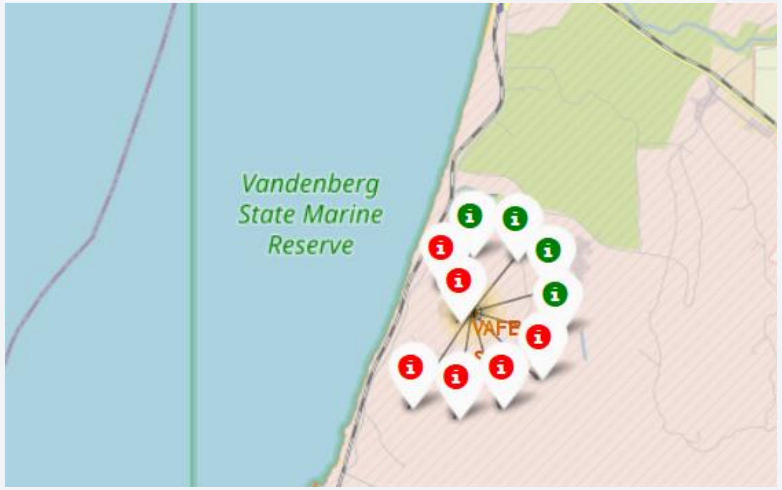
Launch Sites Proximities Analysis

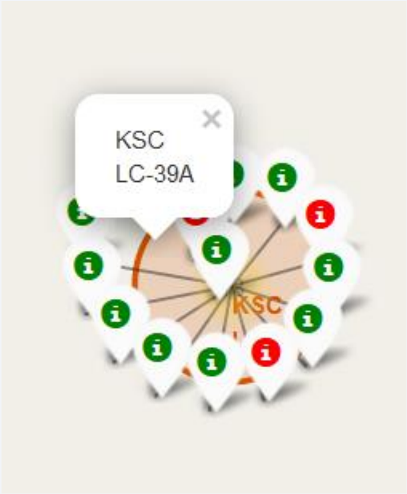
Geographical location of launch sites

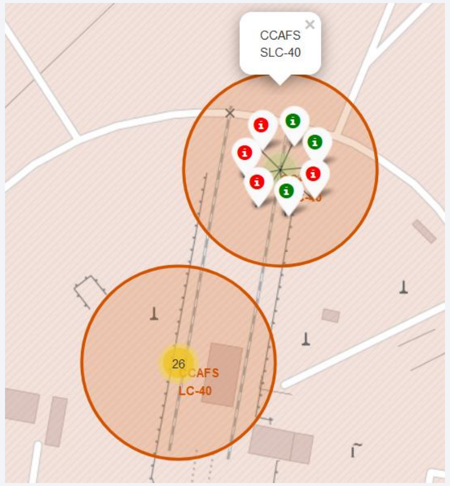
Out[8]:

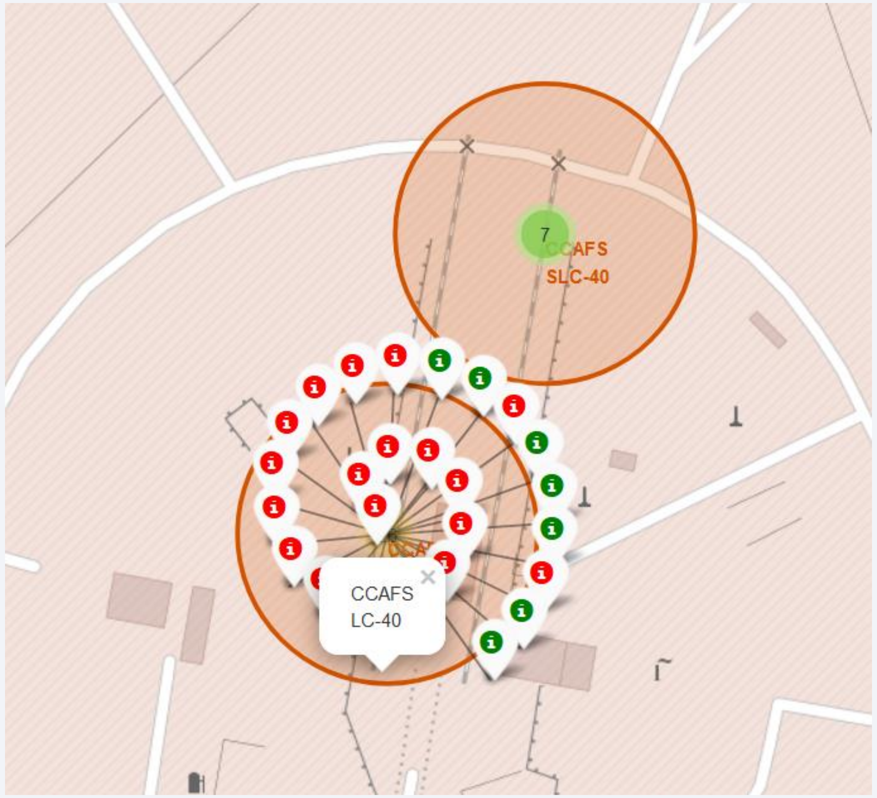


Launch outcomes of Launch sites

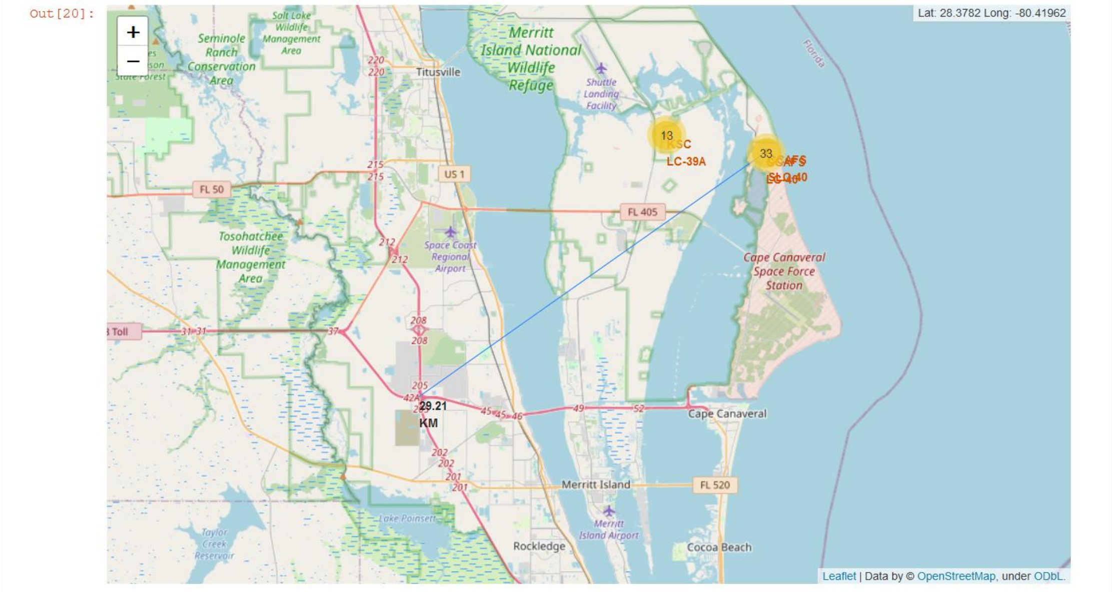


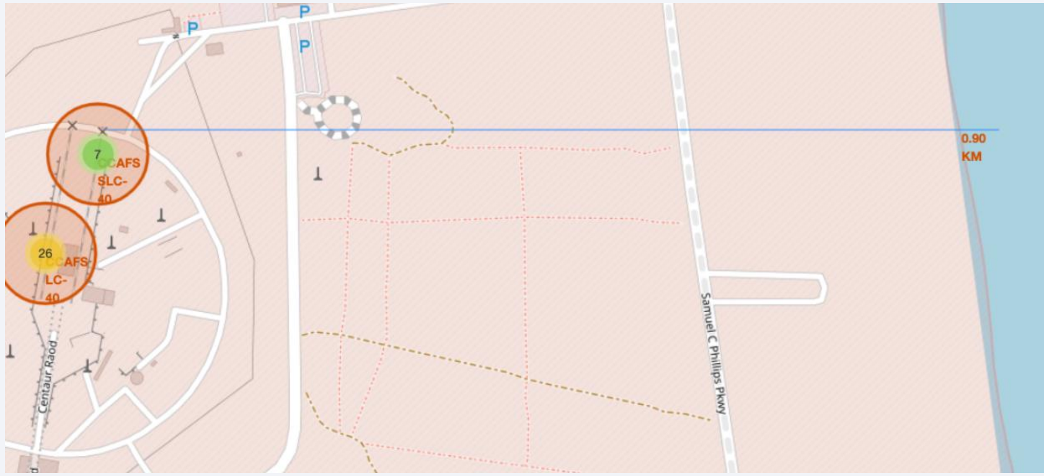






Distance plot to the proximities

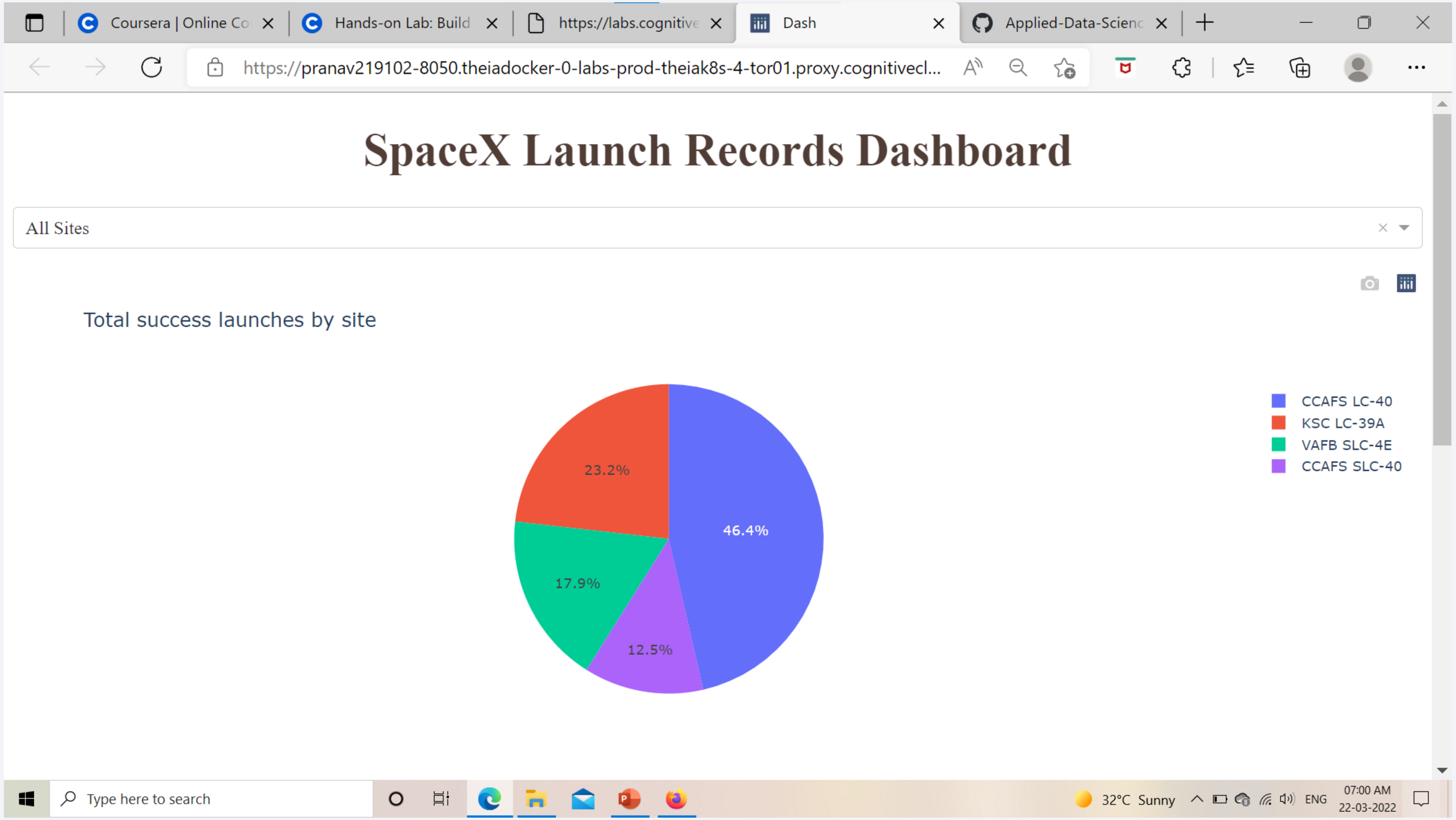




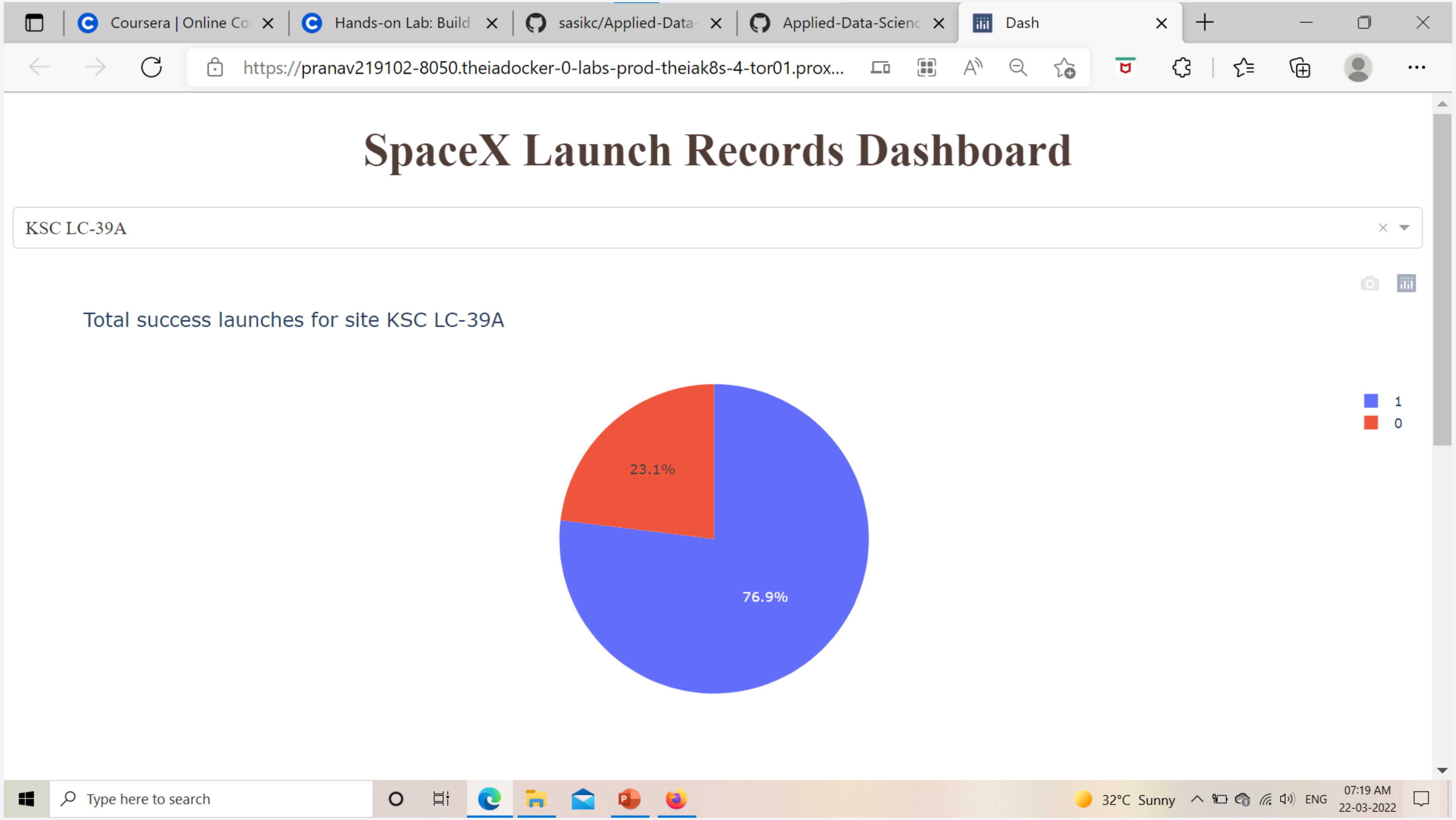
Section 4

Build a Dashboard with Plotly Dash

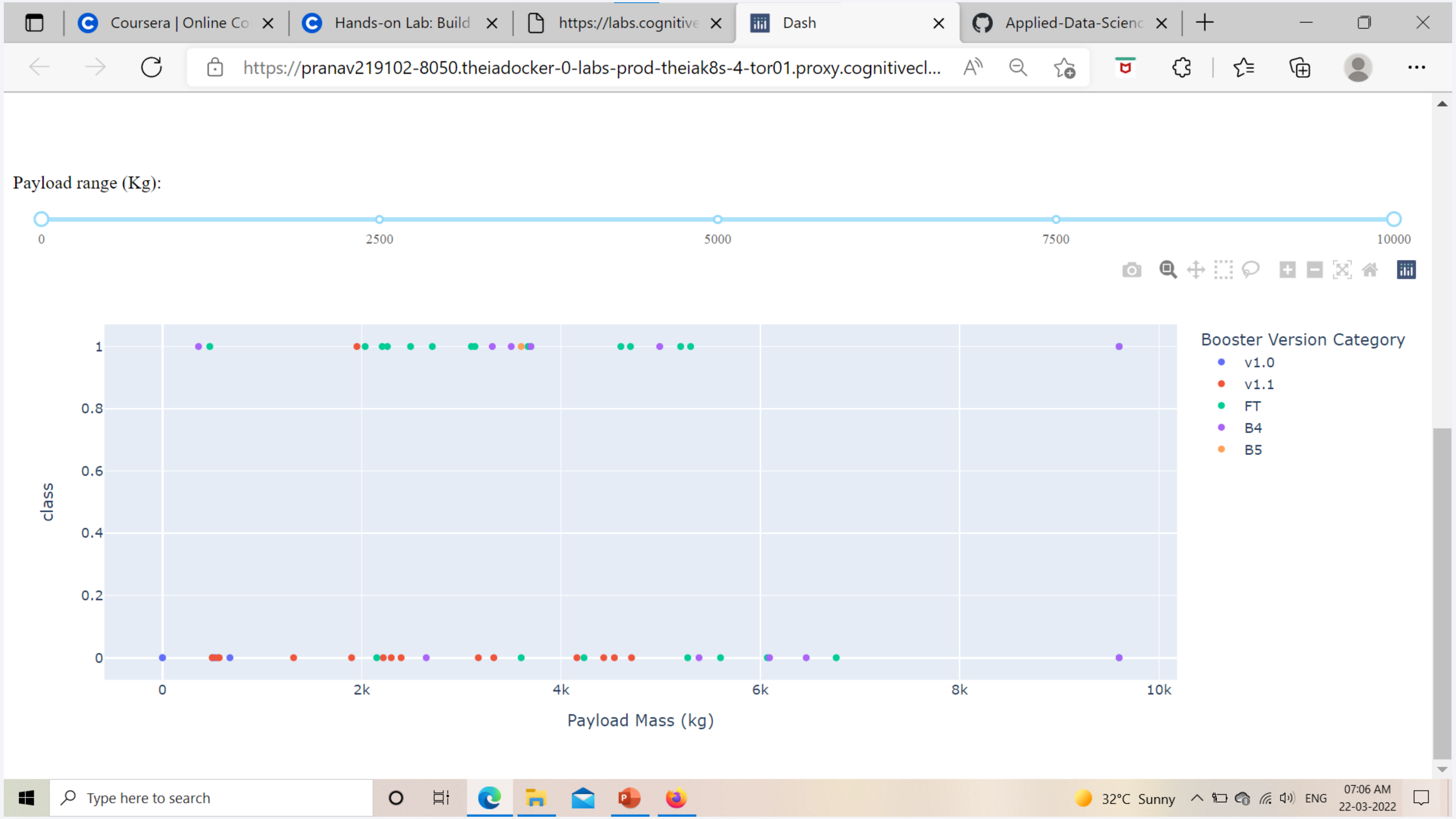
Success rate of launch site



Most Successful Launch Site



Outcome for payload mass



Section 5

Predictive Analysis (Classification)

Classification Accuracy

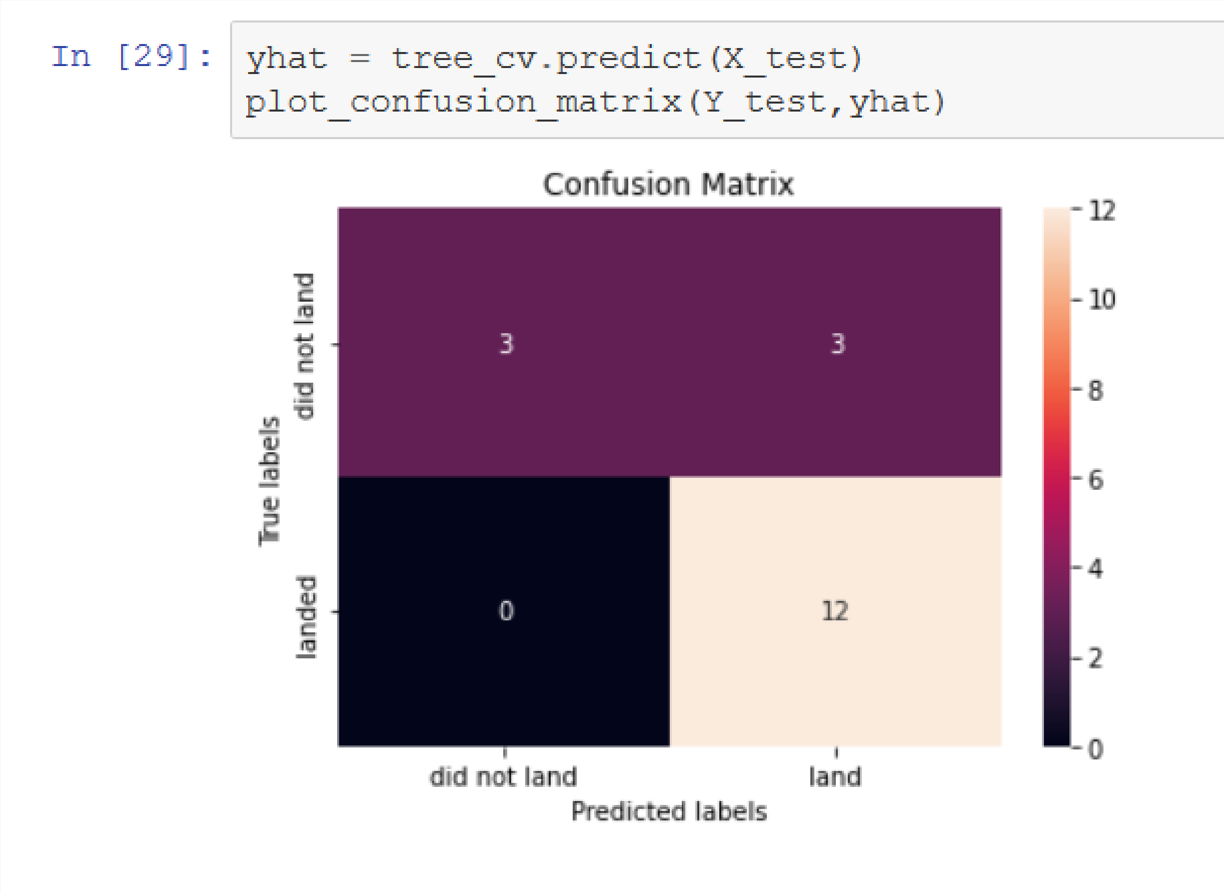
In [35]:
```
models = {'KNeighbors':knn_cv.best_score_,
          'DecisionTree':tree_cv.best_score_,
          'LogisticRegression':logreg_cv.best_score_,
          'SupportVector': svm_cv.best_score_}

bestalgorithm = max(models, key=models.get)
print('Best model is', bestalgorithm,'with a score of', models[bestalgorithm])
if bestalgorithm == 'DecisionTree':
    print('Best params is :', tree_cv.best_params_)
if bestalgorithm == 'KNeighbors':
    print('Best params is :', knn_cv.best_params_)
if bestalgorithm == 'LogisticRegression':
    print('Best params is :', logreg_cv.best_params_)
if bestalgorithm == 'SupportVector':
    print('Best params is :', svm_cv.best_params_)
```

```
Best model is DecisionTree with a score of 0.8767857142857143
Best params is : {'criterion': 'gini', 'max_depth': 14, 'max_features': 'sqrt', 'min_samples_leaf': 4, 'min_samples_s
plit': 2, 'splitter': 'random'}
```

Confusion Matrix

- Examining the confusion matrix, we see that decision tree can distinguish between the different classes
- It has best score i.e accuracy is 0.8482



Conclusions

- The decision tree classifier is best fit machine learning algorithm for future prediction
- KSC LC – 39A Launch site has most successful launches
- Orbit type GEO, HEO, SSO, ES-L1 have most success rate
- Success is high with less than 5000 kg payload mass compared to high payload mass

Appendix

- https://github.com/IamPranav28/SpaceX-Falcon-9-first-stage-Landing-Prediction-Guided-Project

Thank you!