IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

- Summary of methodologies
- Public data regarding SpaceX's falcon 9 rocket launches was gathered. Subsequently, the data was analyzed using exploratory data analysis, and machine learning techniques were utilized to predict if the first stage will land successfully.
- Summary of all results

Introduction

- Project background and context
- The commercial space age is here and companies are making space travel affordable for everyone. Currently, SpaceX advertises Falcon 9 launches for 62 million dollars, while other providers cost upwards of 165 million dollars. Most of the savings is because SpaceX is able to reuse the first stage. We are a new rocket company that would like to compete with SpaceX.
- Problems you want to find answers
- We would like to know the price of each launch.
- If SpaceX will reuse the first stage, in other words, if it will land succesfully

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was scraped from the SpaceX Wikipedia page and combined with the SpaceX API
- Perform data wrangling
  - A landing outcome label was created from the outcome column. Also missing values were imputed with the mean.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - We used logistic regression, support vector machine, decision tree, and a k-nearest neighbor to make predictions. Utilized a gridsearch method to tune the hyperparameters. A confusion matrix and accuracy score to evaluate the models.

Data Collection

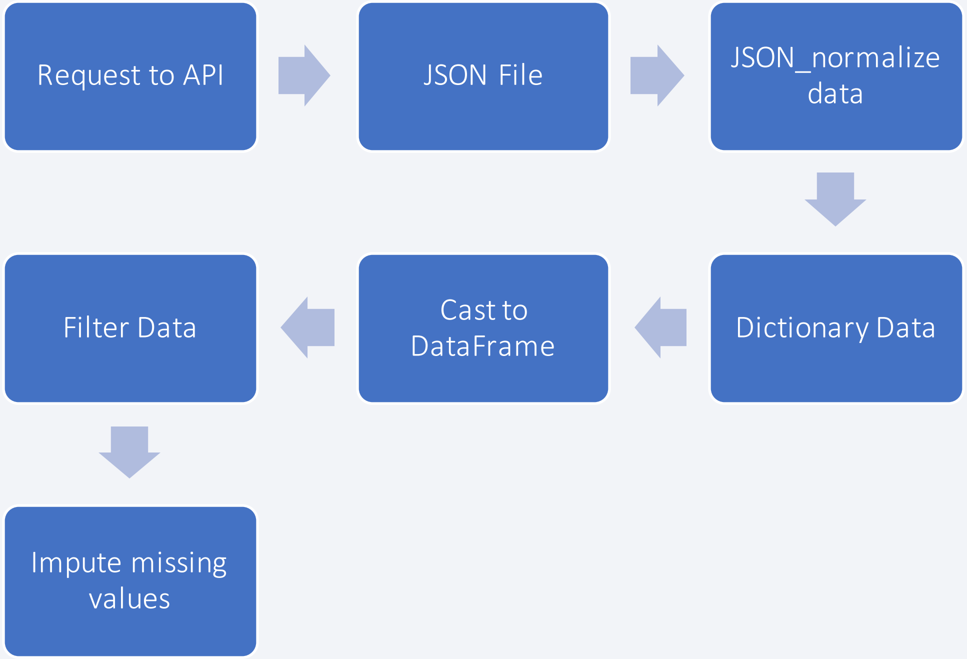
- The data sets were collected from a combination of two sources: a table within the SpaceX Wikipedia page and also responses from the SpaceX API.
- The information contained within the table includes: Launch site, flight number, customer, orbit, payload, payload mass, outcome, version booster, booster landing, date, and time.
- The information contained within the API includes: Booster version, Payload mass, orbit, launch site, date, flight number, gridfins, reused, outcome, latitude, longitude, legs, landing pad, block, reused count, and serial

Data Collection – SpaceX API

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

- 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

- Github URL:
  - https://github.com/IBM-Capstone/blob/main/Data%20Collection%20API%20Lab.ipynb

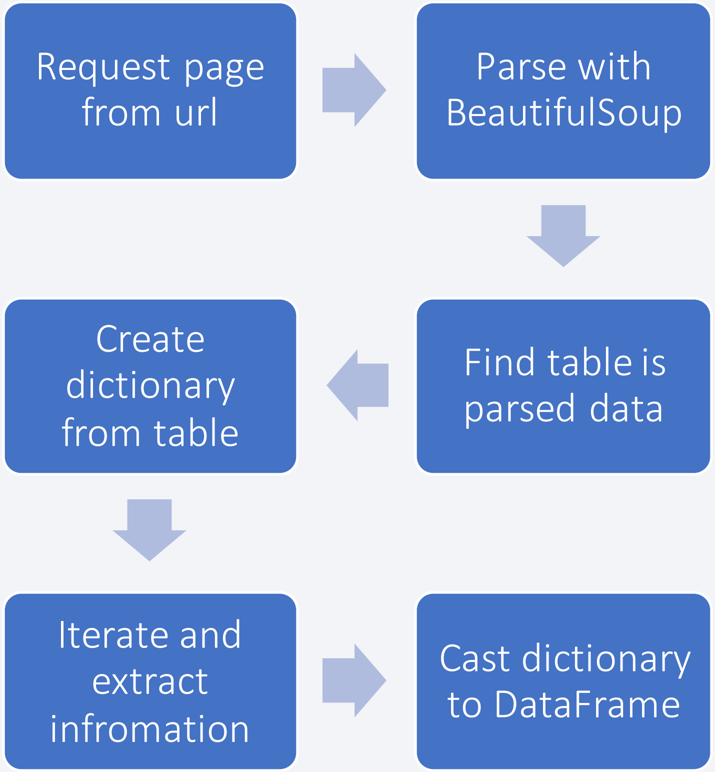


Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose

- Github URL:
  - https://github.com/IBM-Capstone/blob/main/Data%20Collection%20with%20Web%20Scrapping%20Lab.ipynb



Data Wrangling

- True Ocean, True RTLS, and True ASDS are considered successes and were converted to a label of 1.
- Likewise, False Ocean, False RTLS, and False ASDS were considered failures and converted to a label of 0.
- Github URL:
  - https://github.com/IBM-Capstone/blob/main/Data%20Wrangling%20Lab.ipynb

EDA with Data Visualization

- Scatter plots, bar plots, and line plots were used in order to help visualize trends in the data. Particularly, relationships between two variables
- Github URL:
  - https://github.com/IBM-Capstone/blob/main/EDA%20with%20Pandas%20and%20Matplotlib.ipynb

EDA with SQL

- Loaded data into Db2 database
- Connected to database with sql DB2 magic
- Queried data in database

- GitHub URL:
  - https://github.com/IBM-Capstone/blob/main/EDA%20with%20SQL%20Lab.ipynb

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- Markers, marker clusters, mouse pointers, circles, poly lines
- To identify launch sites on maps, identify success and failures at each site, identify distance between two points
- https://github.com/IBM-Capstone/blob/main/Interactive%20Visual%20Analytics%20and%20Dashboard%20Lab%20.ipynb

Build a Dashboard with Plotly Dash

- The dashboard includes both a pie chart and a scatter plot
- The pie chart showed a visual of the total successful launches for all sites or the ratio or successful to failed for each site
- The scatter plot identified if the launch was a success or failure with the payload mass for each booster version
- Explain why you added those plots and interactions
- GitHub URL:
  - https://github.com/derrickRmartinez/blob/main/Plotly%20Dashoboard.ipynb

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- Evaluated the models using both a confusion matrix and accuracy score. The models were improved by testing different parameters by utilizing GridSearchCV
- GitHub URL:
  - https://github.com/IBM-Capstone/blob/main/Prediction%20Lab.ipynb

Results

- Exploratory data analysis showed that there are certain trends for the different relationships among the variables
- All models achieved around the same accuracy score of .83333 and same confusion matrix.

Section 2

Insights drawn from EDA

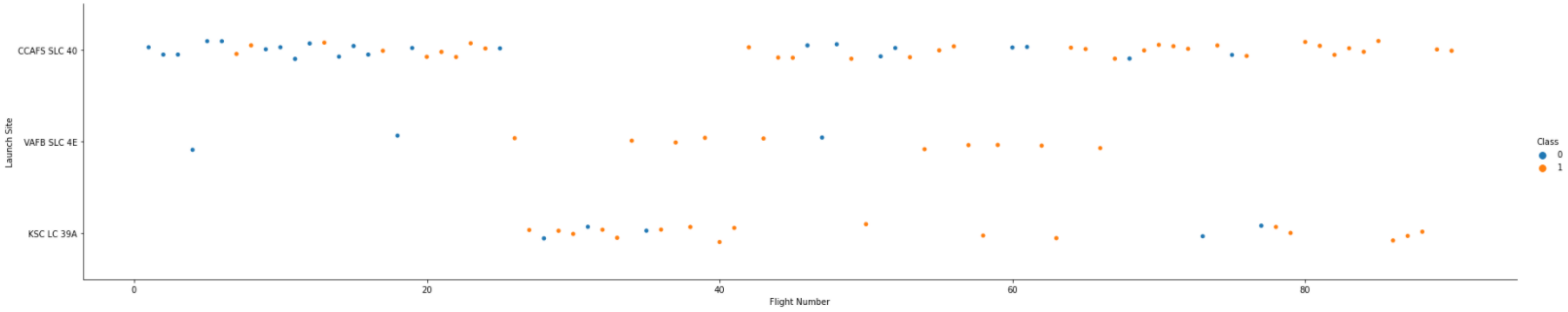
Flight Number vs. Launch Site

TASK 1: Visualize the relationship between Flight Number and Launch Site

Use the function `catplot` to plot `FlightNumber` vs `LaunchSite`, set the parameter `x` parameter to `FlightNumber`,set the `y` to `Launch Site` and set the parameter `hue` to `'class'`

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')
plt.ylabel('Launch Site')
plt.show()
```



- This scatter plot shows the relationship between the flight number, launch site and if these were successes or failures

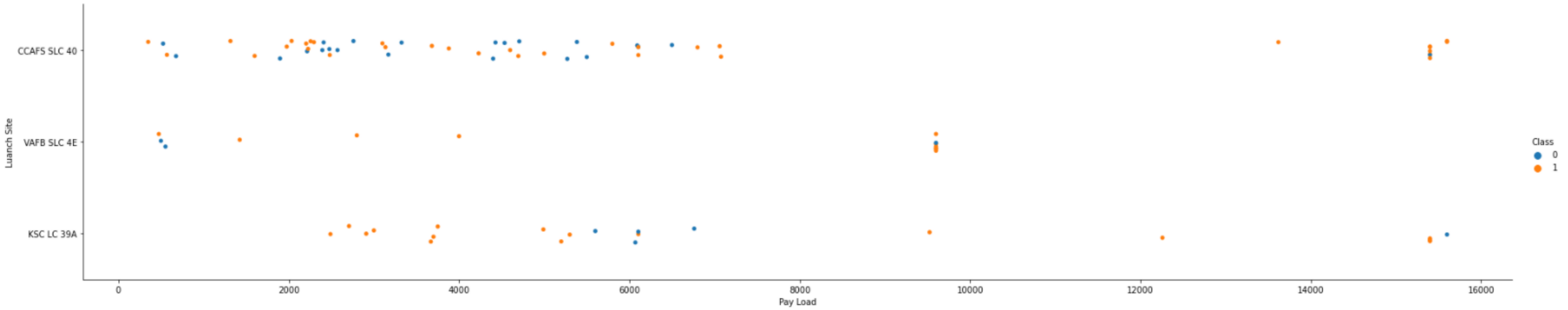
Payload vs. Launch Site

TASK 2: Visualize the relationship between Payload and Launch Site

We also want to observe if there is any relationship between launch sites and their payload mass.

In [6]:

```
# 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 class value
sns.catplot(x = 'PayloadMass', y ='LaunchSite', hue = 'Class', data = df, aspect = 5)
plt.xlabel('Pay Load')
plt.ylabel('Luanch Site')
plt.show()
```



- Shows the relationship between payload and launch site and if these were successful or failed

Success Rate vs. Orbit Type

TASK 3: Visualize the relationship between success rate of each orbit type

Next, we want to visually check if there are any relationship between success rate and orbit type.

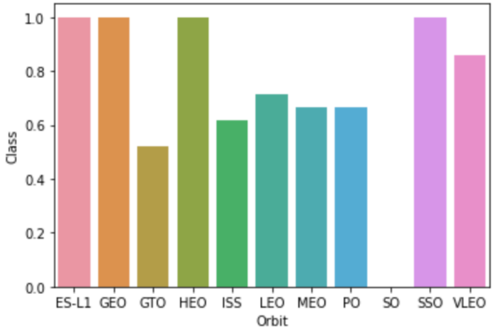
Let's create a `bar chart` for the sucess rate of each orbit

In [20]:

```
# HINT use groupby method on Orbit column and get the mean of Class column
grouped = df[['Orbit', 'Class']].groupby('Orbit').mean()
grouped['Orbit'] = grouped.index
sns.barplot(x = 'Orbit', y = 'Class', data = grouped)
```

Out[20]:

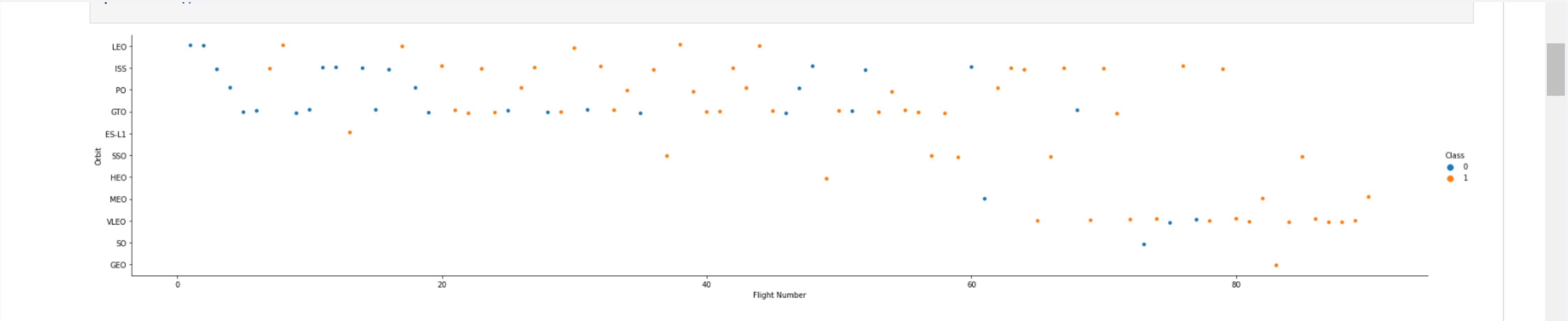
```
<AxesSubplot:xlabel='Orbit', ylabel='Class'>
```



Analyze the ploted bar chart try to find which orbits have high sucess rate.

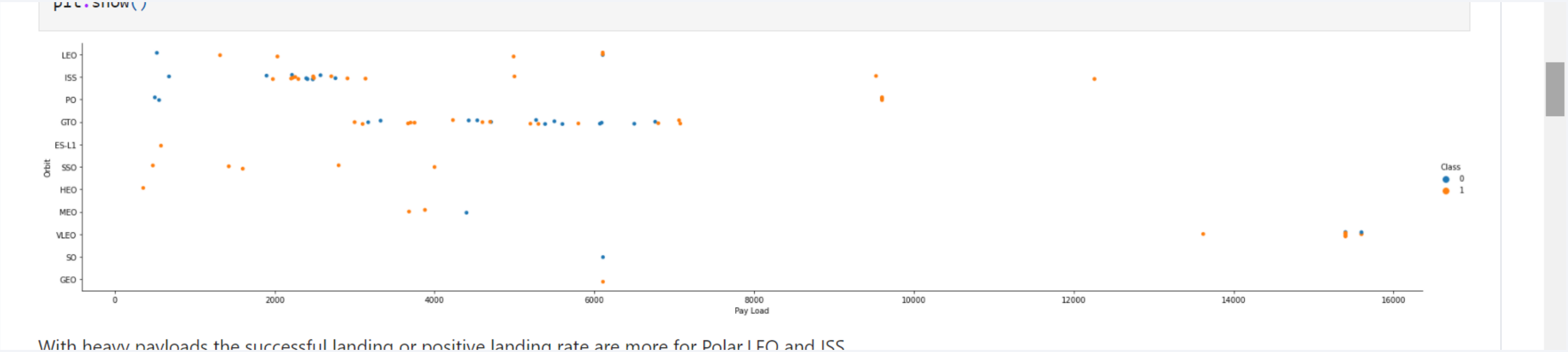
- This graph show the success rate for each orbit type

Flight Number vs. Orbit Type



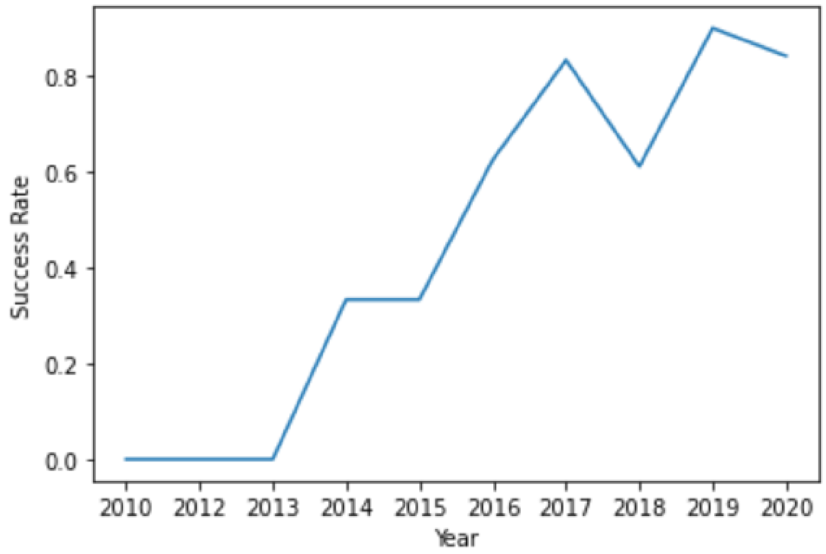
- This scatter plot shows the relationship between flight number and orbit type and if the launch succeeded or failed

Payload vs. Orbit Type



- This scatter plot shows the relationship between payload and orbit type and if each launch succeeded or failed

Launch Success Yearly Trend



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

]:

```
%%sql

select unique(LAUNCH_SITE)
FROM SPACEXTBL;
```

```
 * ibm_db_sa://lkx88098:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

Out[6]:

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

- The table above lists all of the unique launch site names

Launch Site Names Begin with 'CCA'

Task 2

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

In [10]:
```
%%sql

SELECT *
FROM SPACEXTBL
WHERE LAUNCH_SITE LIKE 'CCA%'
LIMIT 5;
```

```
 * ibm_db_sa://lkx88098:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

Out[10]:

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

- These are the results of 5 records with a site name that begins with 'CCA'. The query is limited to only the first 5 records.

Total Payload Mass

Task 3

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

[16]:

```
%%sql

SELECT CUSTOMER, SUM(PAYLOAD_MASS__KG_) AS TOTAL_MASS
FROM SPACEXTBL
WHERE CUSTOMER LIKE '%NASA%(CRS)%'
GROUP BY CUSTOMER;
```

```
 * ibm_db_sa://lkx88098:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

[16]:

customer	total_mass
NASA (CRS)	45596
NASA (CRS), Kacific 1	2617

- The total for only NASA is 45596, and the total for NASA with Kacific 1 is 2617

Average Payload Mass by F9 v1.1

Task 4

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

In [17]:
```
%%sql

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

```
 * ibm_db_sa://lkx88098:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

Out[17]:

1
2928

- The average payload carried by booster version F9 v1.1 is 2928 kg

First Successful Ground Landing Date

Task 5

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

Hint:Use min function

[25]:
```
%%sql

SELECT MIN(DATE) AS First_Success
FROM SPACEXTBL
WHERE LANDING__OUTCOME LIKE 'Success (ground pad)';
```

```
 * ibm_db_sa://lkx88098:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

[25]:

first_success
2015-12-22

- The first successful landing outcome on a ground pad occurred on December 22, 2015

Successful Drone Ship Landing with Payload between 4000 and 6000

Task 6

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

[29]:

```
%%sql

SELECT BOOSTER_VERSION
FROM SPACEXTBL
WHERE LANDING__OUTCOME LIKE 'Success (drone ship)' AND (4000 < PAYLOAD_MASS__KG_ < 6000);
```

```
 * ibm_db_sa://lkx88098:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

[29]:

booster_version
F9 FT B1021.1
F9 FT B1023.1
F9 FT B1029.2
F9 FT B1038.1
F9 B4 B1042.1
F9 B4 B1045.1
F9 B5 B1046.1

- The table above includes the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Total Number of Successful and Failure Mission Outcomes

Task 7

List the total number of successful and failure mission outcomes

]:
```
%%sql

SELECT MISSION_OUTCOME, COUNT(*) as TOTAL
FROM SPACEXTBL
GROUP BY MISSION_OUTCOME;
```

```
 * ibm_db_sa://lkx88098:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

]:

mission_outcome	total
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

- The total number of successful and failure by mission outcomes in listed in the table above

Boosters Carried Maximum Payload

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

```
%%sql

SELECT BOOSTER_VERSION, PAYLOAD_MASS__KG_
FROM SPACEXTBL
WHERE PAYLOAD_MASS__KG_ IN (SELECT MAX(PAYLOAD_MASS__KG_)
                            FROM SPACEXTBL);
```

```
 * ibm_db_sa://lkx88098:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
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

- The names of the boosters which have carried the maximum payload mass are listed in the table above.

2015 Launch Records

Task 9

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

[42]:

```
%%sql

SELECT LANDING__OUTCOME, BOOSTER_VERSION, LAUNCH_SITE, DATE
FROM SPACEXTBL
WHERE LANDING__OUTCOME = 'Failure (drone ship)' and DATE LIKE '2015%';
```

```
 * ibm_db_sa://lkx88098:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

[42]:

landing__outcome	booster_version	launch_site	DATE
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40	2015-01-10
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40	2015-04-14

- The failed landing outcomes for a drone ship, their booster versions, and launch site names for year 2015 is listed in the table above

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(*) AS TOTAL
FROM SPACEXTBL
WHERE DATE BETWEEN to_date('2010-06-04', 'yyyy-mm-dd') AND to_date('2017-03-20', 'yyyy-mm-dd')
GROUP BY LANDING__OUTCOME
ORDER BY TOTAL DESC;
```

```
 * ibm_db_sa://lkx88098:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

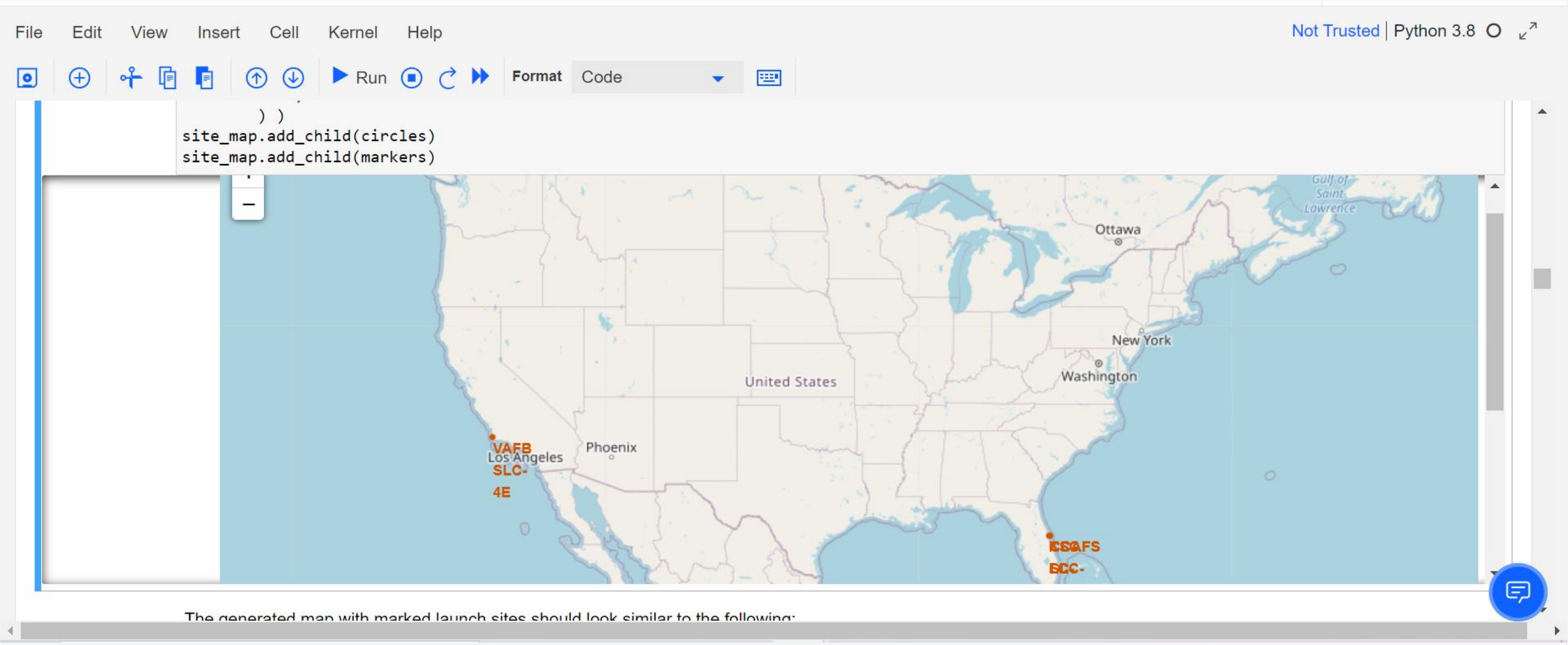
landing__outcome	total
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

- The Rank for total landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, are listed in descending order in the table above

Section 3

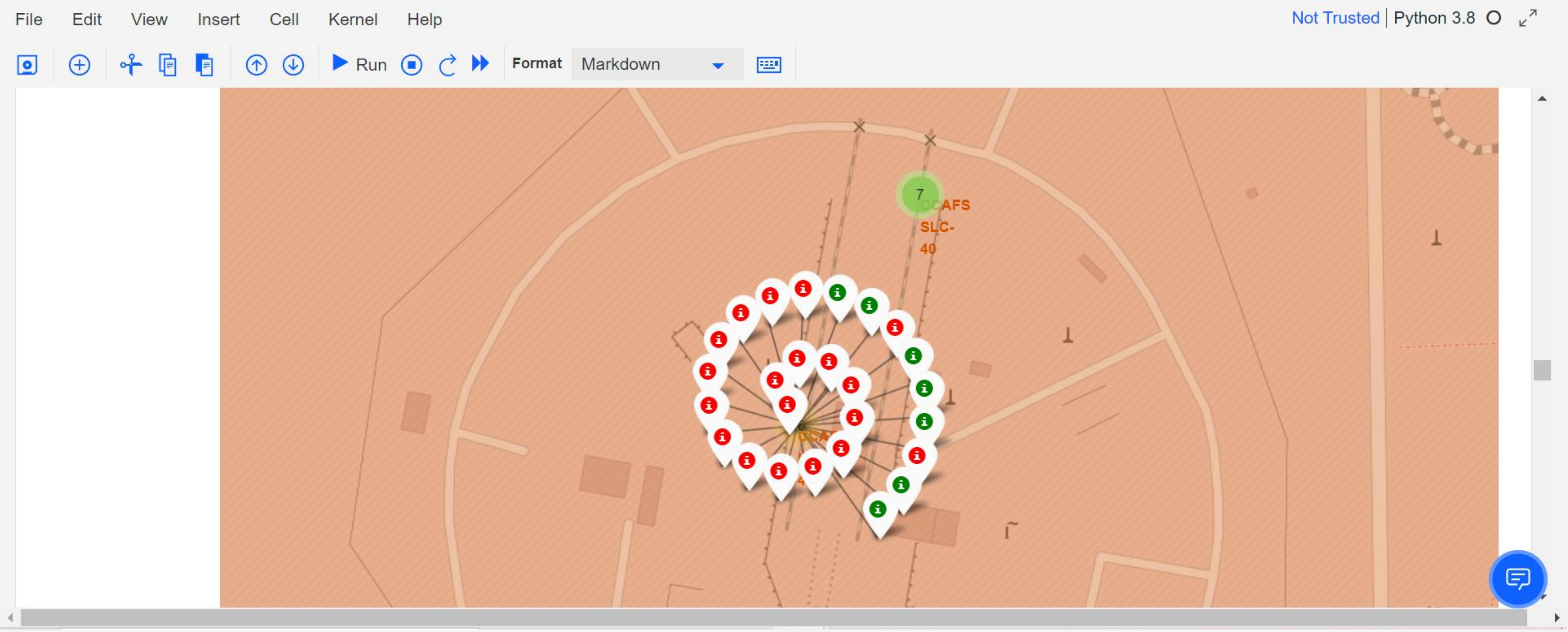
Launch Sites Proximities Analysis

Map of Marked Launch Sites



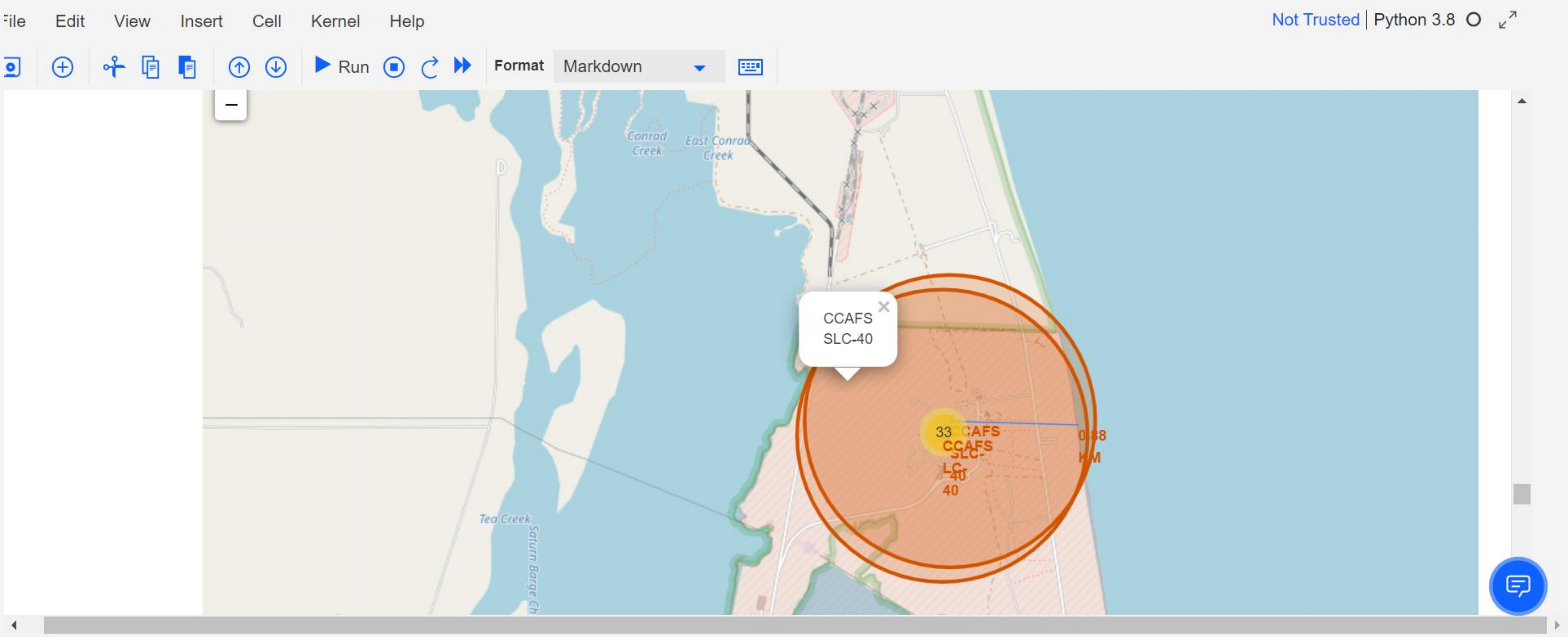
- This map includes the launch sites for the rockets

Success Rate at Launch Sites



- The folium map to identify the success rates at each launch site with color labels

Distance to Coast

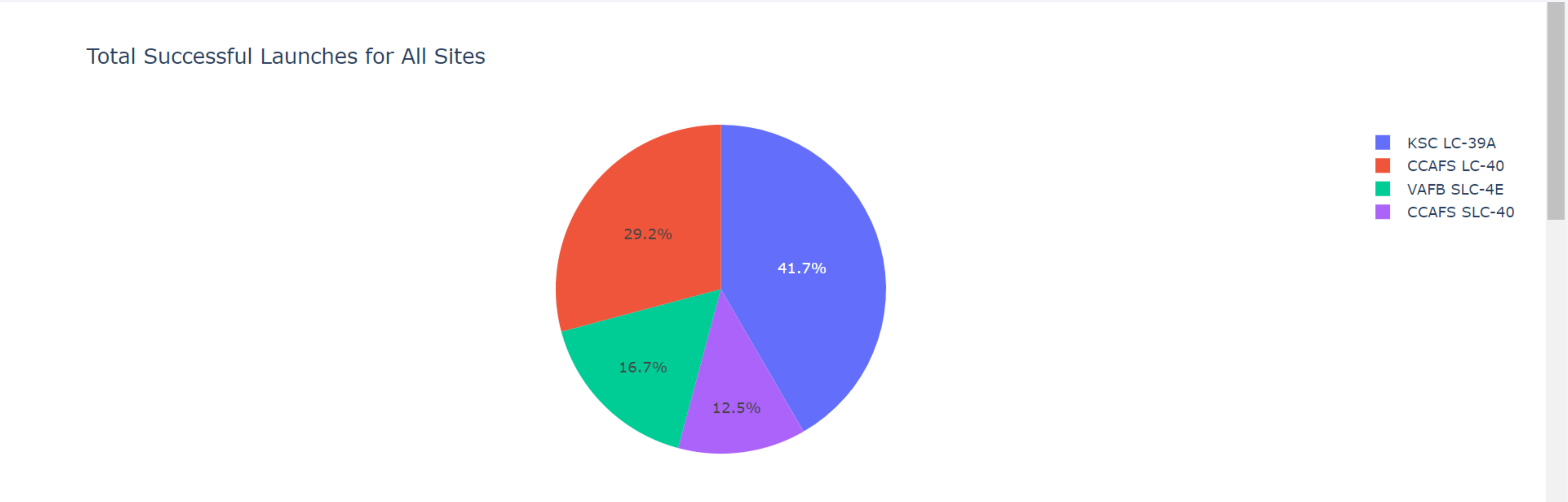


- This folium map shows the distance from a launch site to the nearest coastal point

Section 4

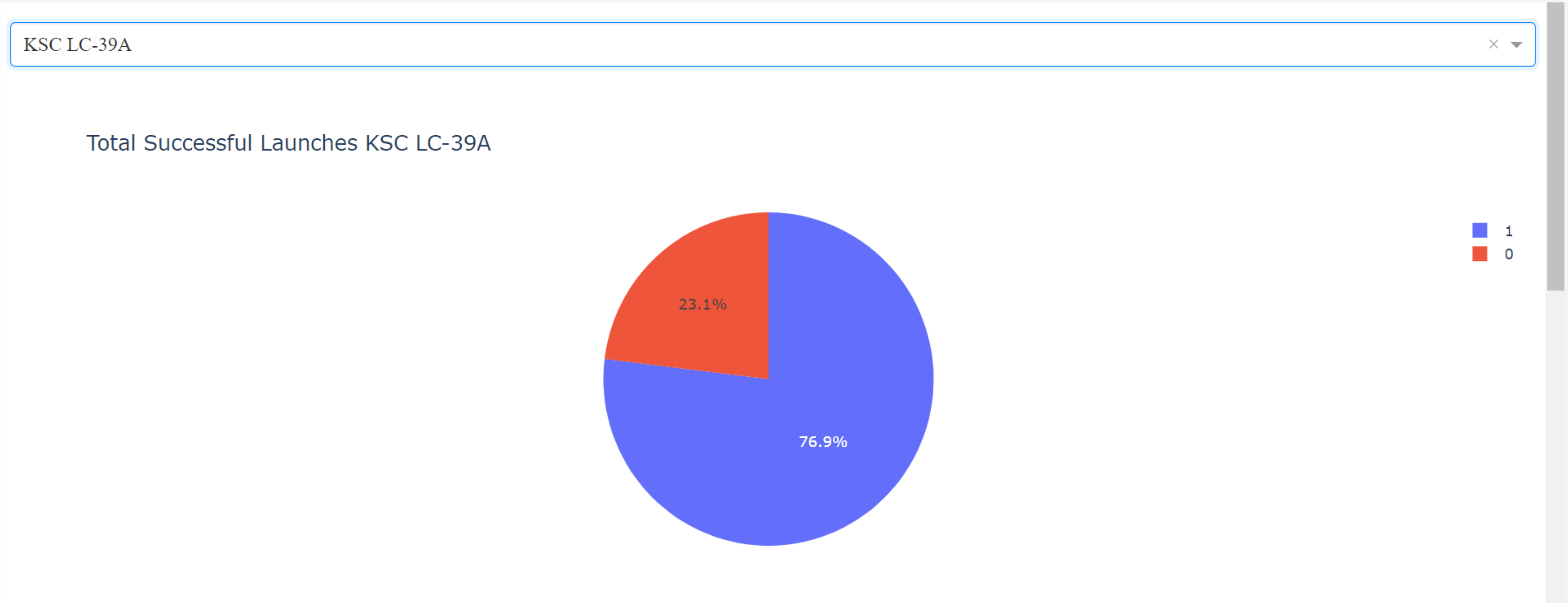
Build a Dashboard with Plotly Dash

Successful Launches Across all Sites



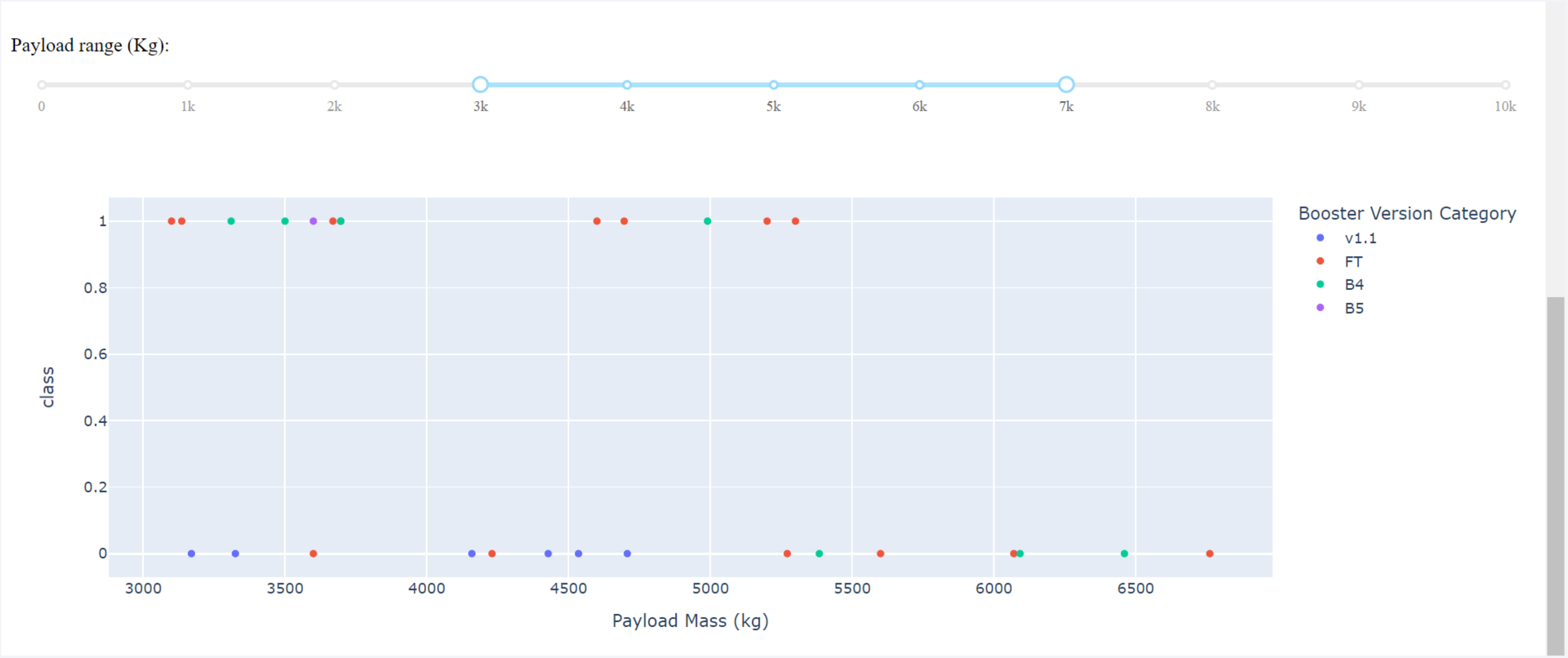
- This is the pie chart for the percent of successful launches as a ratio of all successful launches for each site

Site with Highest Launch Success



- The launch site with highest launch success ratio is site KSC LC-39A

Payload vs Launch Outcome



- Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider. Currently set to the range of 3k to 7k

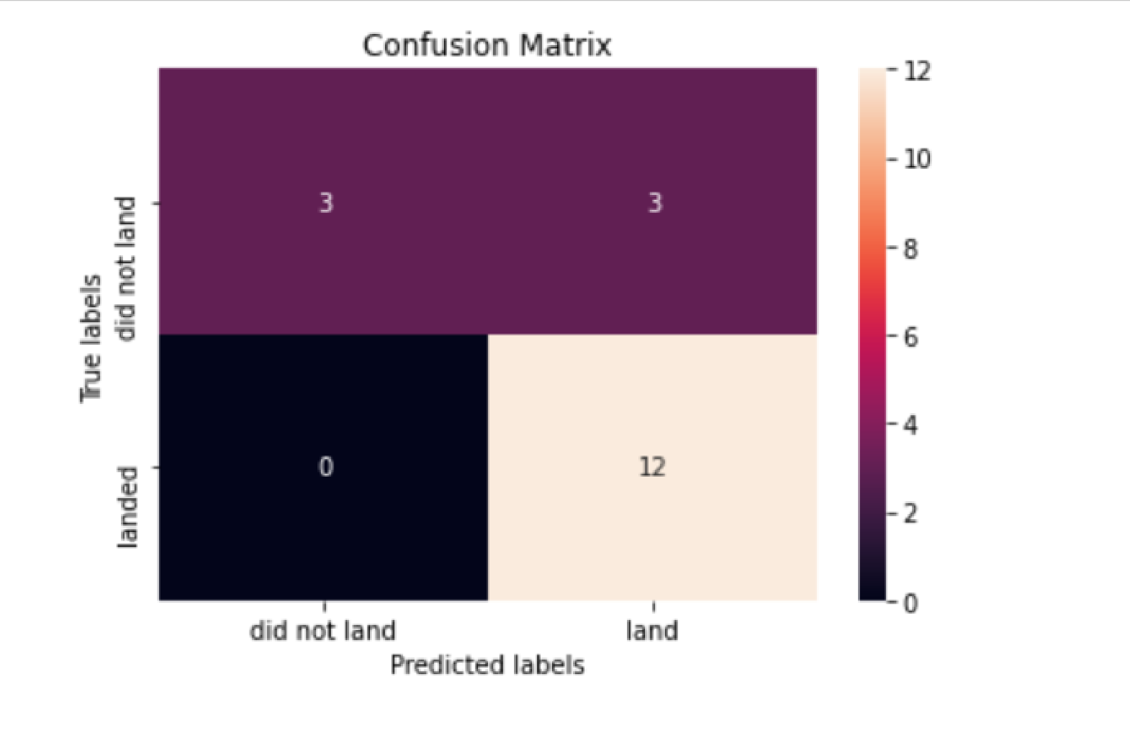
Section 5

Predictive Analysis (Classification)

Classification Accuracy

- All of the models have the same classification accuracy of approximately 0.8333

Confusion Matrix



- All of the models have the same confusion matrix

Conclusions

- We have collected data and visualized trends that will aid us in predicting if a launch will be successful
- Our predictions will enable us to determine if we will be able to compete with SpaceX
- All of our models have a prediction with an accuracy rate of approximately .8333. Though this is decent, if we have more data we would likely be able to both improve the accuracy as well as determine which model is best for the predictions
- Overall, the public information through Wikipedia and the SpaceX API have provided invaluable insights into the success rate for the launches

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!