IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

By Stavros N. Lamprou

30-11-2021

Executive Summary

Methodology

Data Collection

EDA – Visualization, SQL

Visual Analytics & Dashboards

Predictive Analysis

Summary of Results

- Focus on the orbits have the highest success rates: ES-L1, GEO, HEO & SSO,
- Launch Site of Preference: KSC LC-39A,
- Focus on Mid & Heavy Range Payloads,
- Use FT Booster versions over others,
- Use Decision Tree Classifier for future predictions!

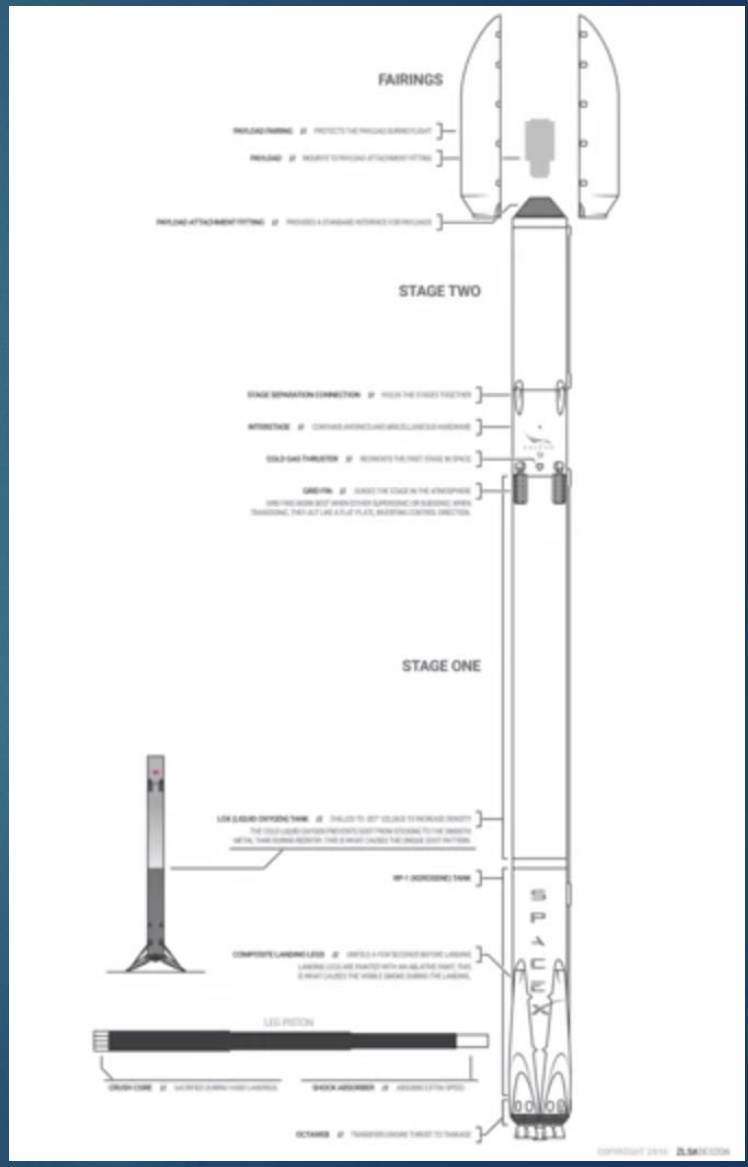
INTRODUCTION


Allon Mask

SpaceY

- True Competitor of SpaceX for the Commercial Space Age Trophy.
- Lower cost of rocket launches as Stage One can be re-used,
- Not always the Stage One would land. Sometimes it would crush or be sacrificed!

- Determine whether the First Stage would Land!
- Determine the price of each Launch.
- Build ML models to predict whether Stage 1 would be reused.



Section 1

Methodology

Methodology

Executive Summary

Data collection methodology:

- Data Collection via APIs
- Web Scrapping

Perform data wrangling

- Use pandas & numpy libraries
- .value_counts()
- Create 0/1 Outcome column
- Use .mean method for success rate

Perform exploratory data analysis (EDA) using matplotlib & seaborn visualization and SQL

Perform interactive visual analytics using Folium and Plotly Dash

Perform predictive analysis using classification models

- Build, tune, evaluate LR, SVM, Decision Tree, KNN classification models

Data Collection

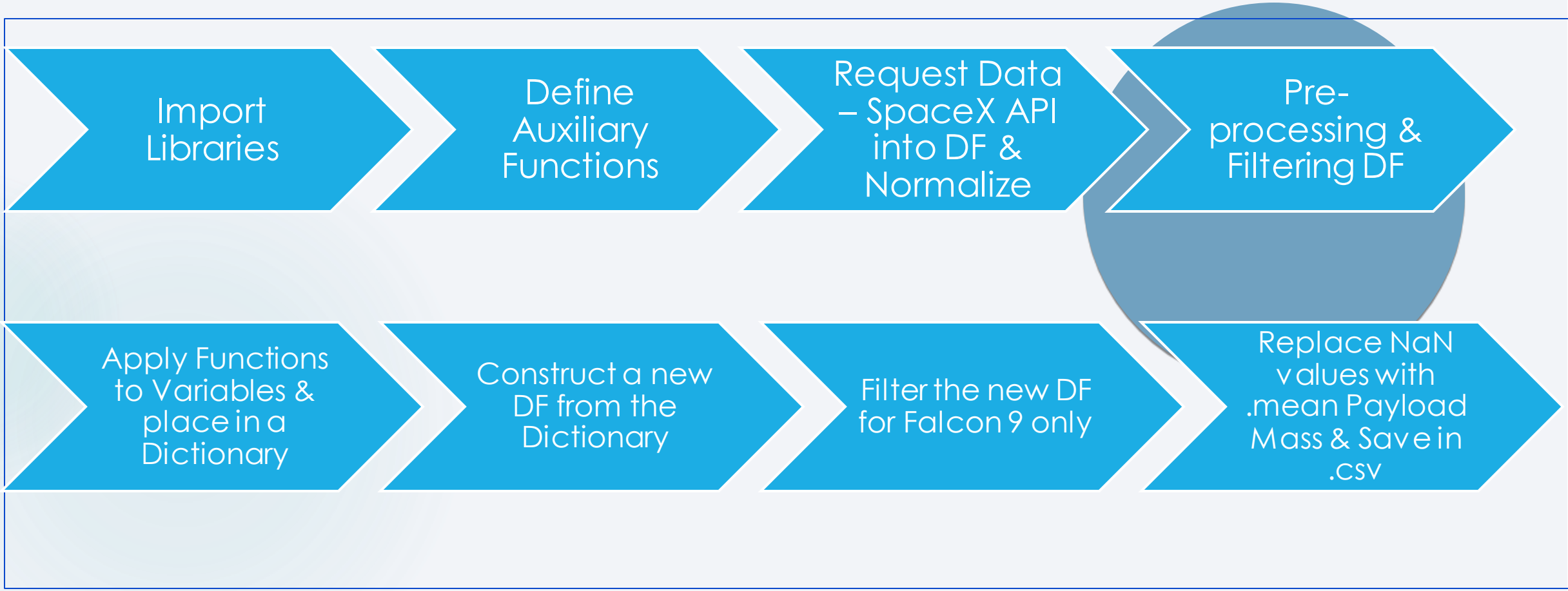
Datasets Collection Sources:

-SpaceX API URL

-Web scrapping Falcon 9 launch records HTML table from Wikipedia

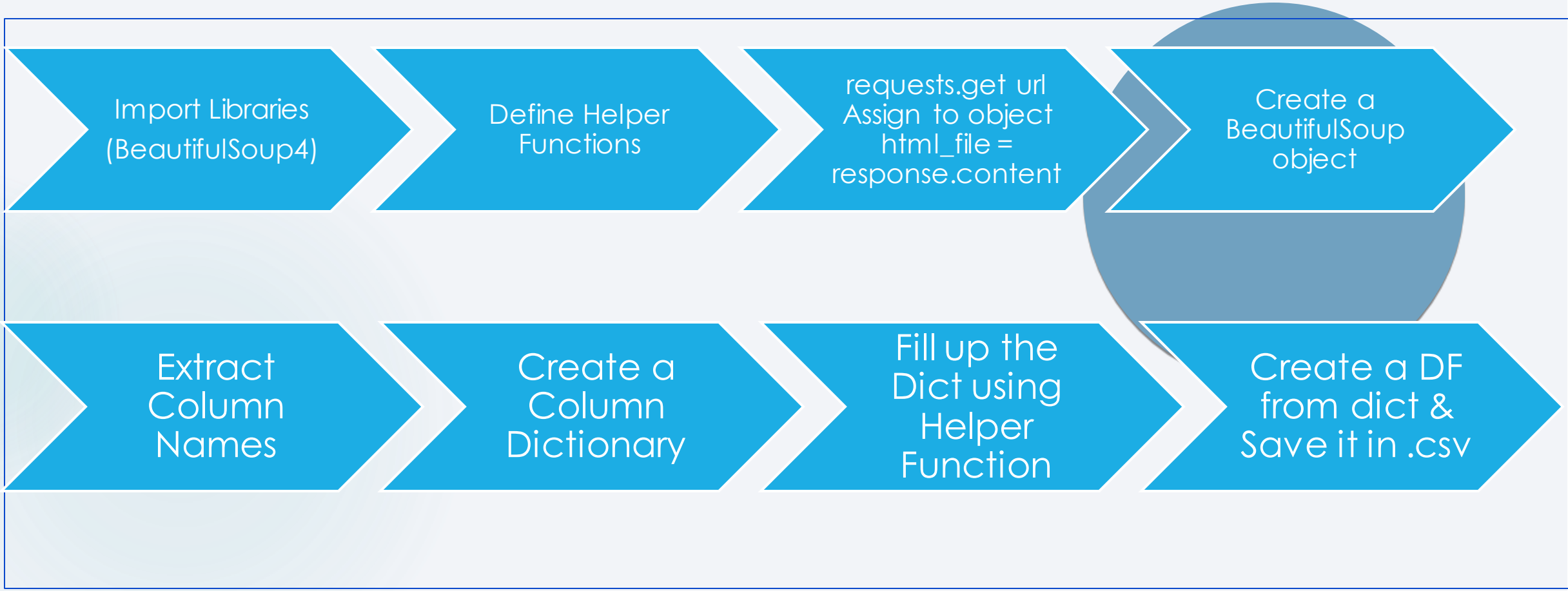
Then, Data Wrangling in Order to Determine the Mean Rate of Success Landing

Data Collection – SpaceX API



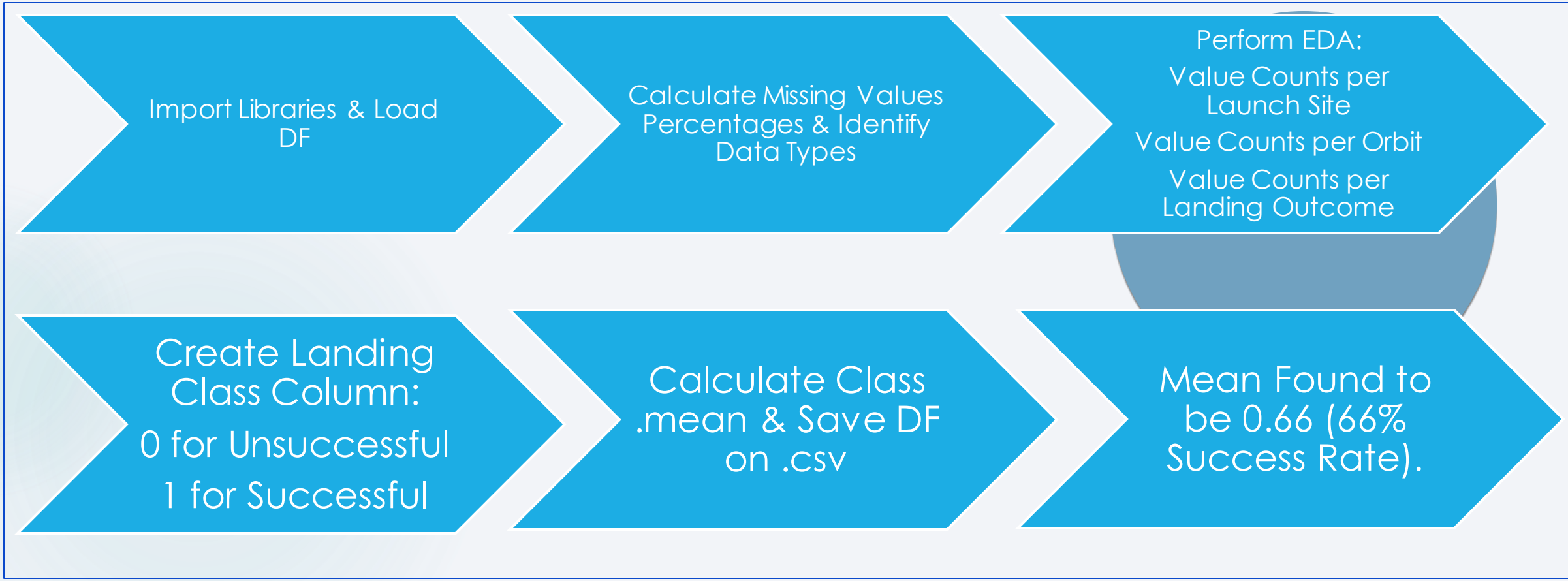
GitHub URL: https://github.com/ageanwave/Applied-Data-Science-Capstone/blob/main/Week%201/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping



GitHub URL: https://github.com/ageanwave/Applied-Data-Science-Capstone/blob/main/Week%201/jupyter-labs-webscraping.ipynb

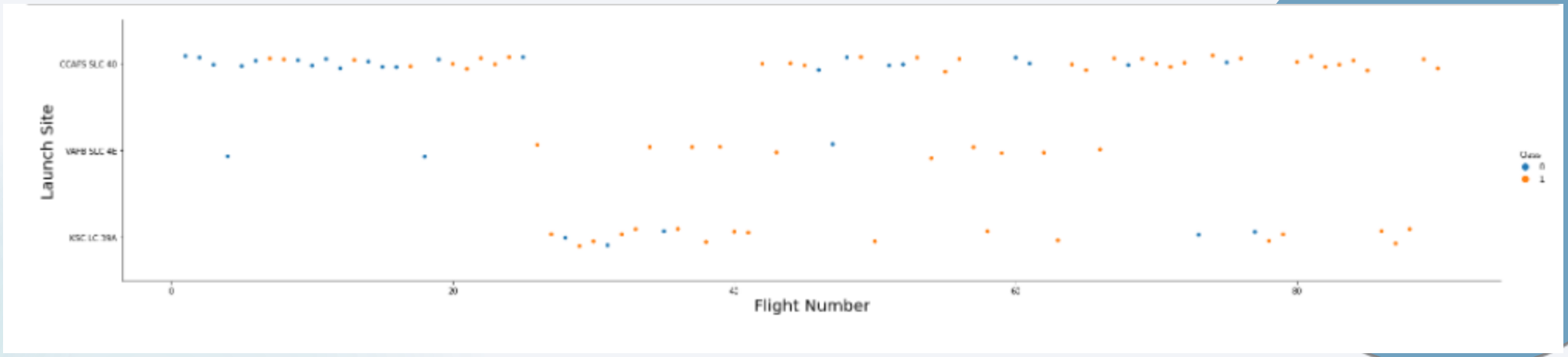
Data Wrangling



GitHub URL: https://github.com/ageanwave/Applied-Data-Science-Capstone/blob/main/Week%201/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

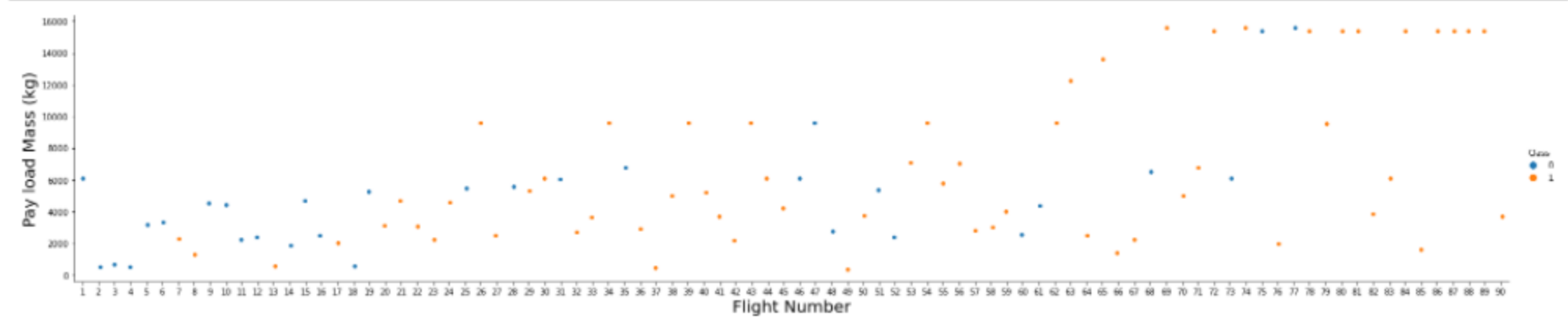
- Libraries used for the plots: numpy, matplotlib, seaborn



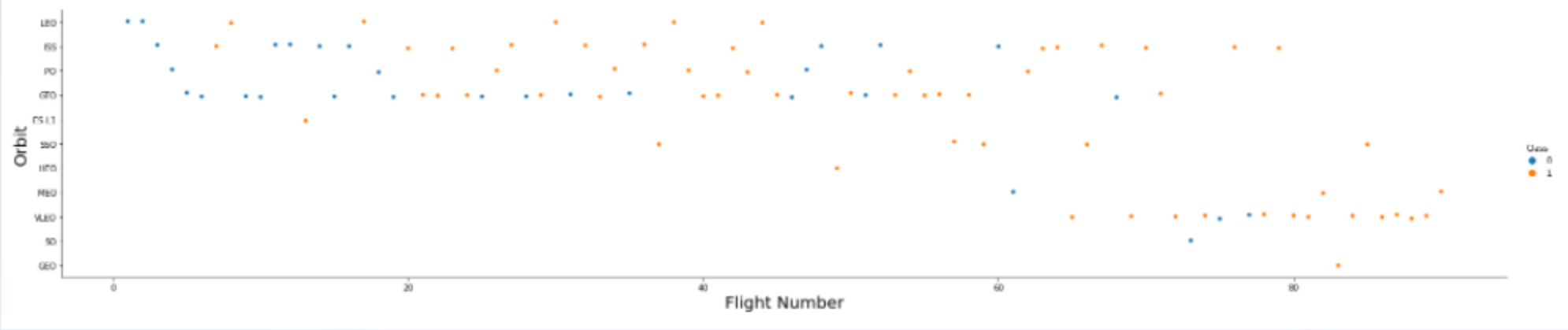
Flight Number vs Launch Site

GitHub URL: https://hub-binder.mybinder.ovh/user/quantstack-xeus-cling-9q2cyus4/notebooks/notebooks/jupyter-labs-eda-dataviz.ipynb

EDA with Data Visualization

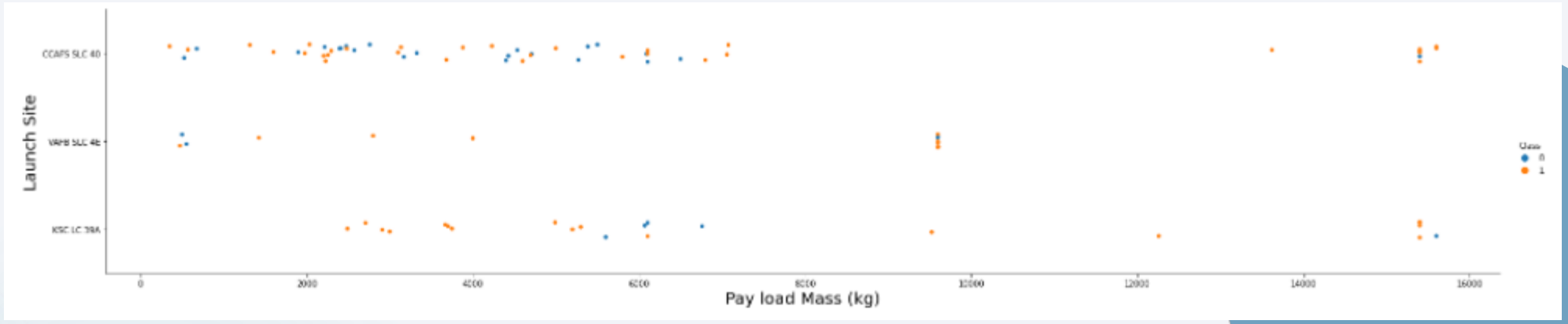


Flight Number vs Payload Mass Plot

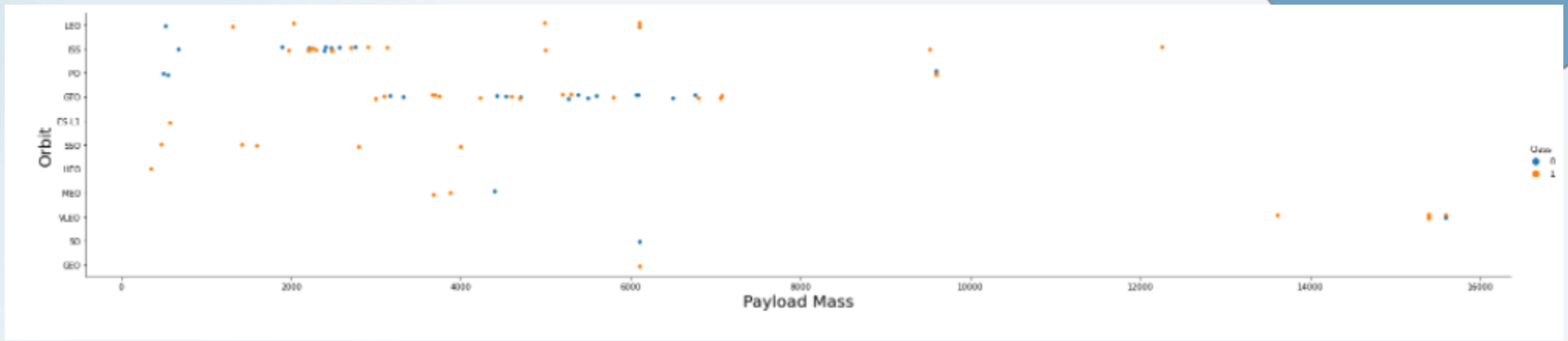


Flight Number vs Orbit

EDA with Data Visualization



Payload vs Launch Site

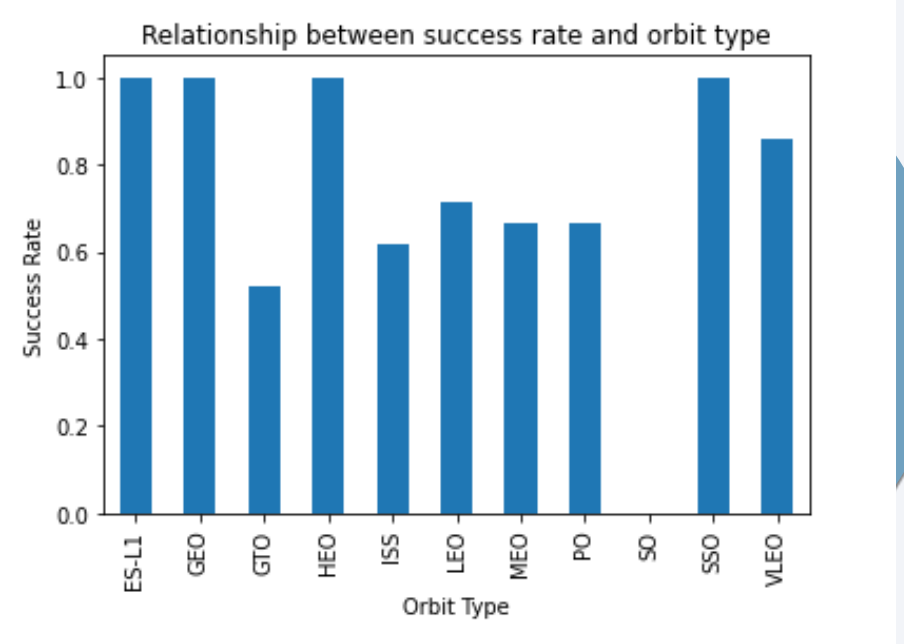


Payload vs Orbit

EDA with Data Visualization

Some Useful Conclusions:

- Different launch sites have different success rates,
- Most Launches had under 8k ton PL. However, 15.5-ton launches were mostly successful,
- CCAFS SLC40 had the largest number of flight while VAFB SLC 4E had the lowest
- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS. With GTO, we cannot distinguish,
- ES-L1, GEO, HEO & SSO orbits have the highest success rates.



GitHub URL: https://hub-binder.mybinder.ovh/user/quantstack-xeus-cling-9q2cyus4/notebooks/notebooks/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

SQL queries performed:

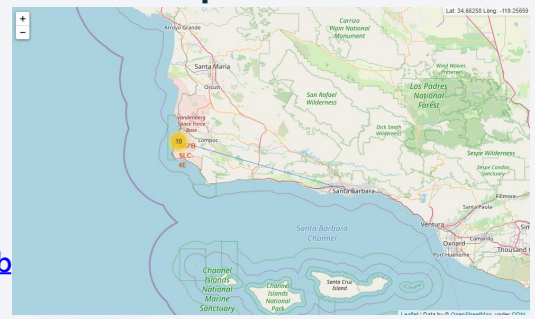
- Names of the unique launch sites
- Five records where launch sites begin with the string 'CCA'
- Total payload mass carried by boosters launched by NASA (CRS)
- Average payload mass carried by booster version F9 v1.1
- Date when the first successful landing outcome in ground pad was achieved
- Names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Total number of successful and failure mission outcomes
- Names of the booster versions which have carried the maximum payload mass
- Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015
- Count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, descending.

GitHub URL: https://github.com/ageanwave/Applied-Data-Science-Capstone/blob/main/Week%202/jupyter-labs-eda-sql-coursera.ipynb

Build an Interactive Map with Folium

- Aim is to create an interactive Folium Map that includes Site Launch locations, Successful & Failed Launches and Proximity to key geographic components such as Coastlines, Cities, Hightways, Railways etc.
- Object used on the Analysis:

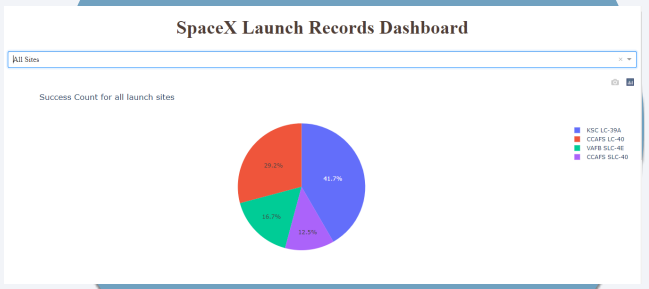
1. Circles to mark all Launch Sites with Names popups
2. Markers to Group Launch attempts and drill down on successful or failed attempts per site
3. MousePositions, to get the coordinates of specific points on the map by the mouse pointer
4. Lines, used to calculate the distances of Launch Sites to its proximities

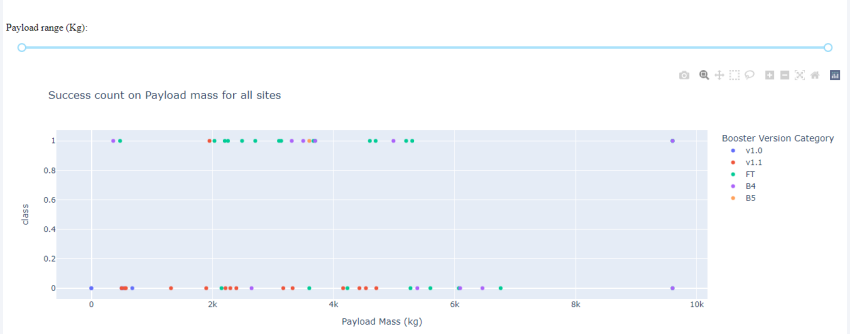


- GitHub URL: https://github.com/ageanwave/Applied-Data-Science-Capstone/blob/main/Week%203/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Drop down menu includes All_Sites and each individual Launch Site,
- Pie Charts of Success Launches % for All Sites & Success/Failure rates per individual Site,
- Slider for selecting Payload ranges,
- Create various Success count graphs for various payload ranges, marked by Booster Version Category,
- Draw insights & conclusions on the optimal payload ranges & the best booster versions.



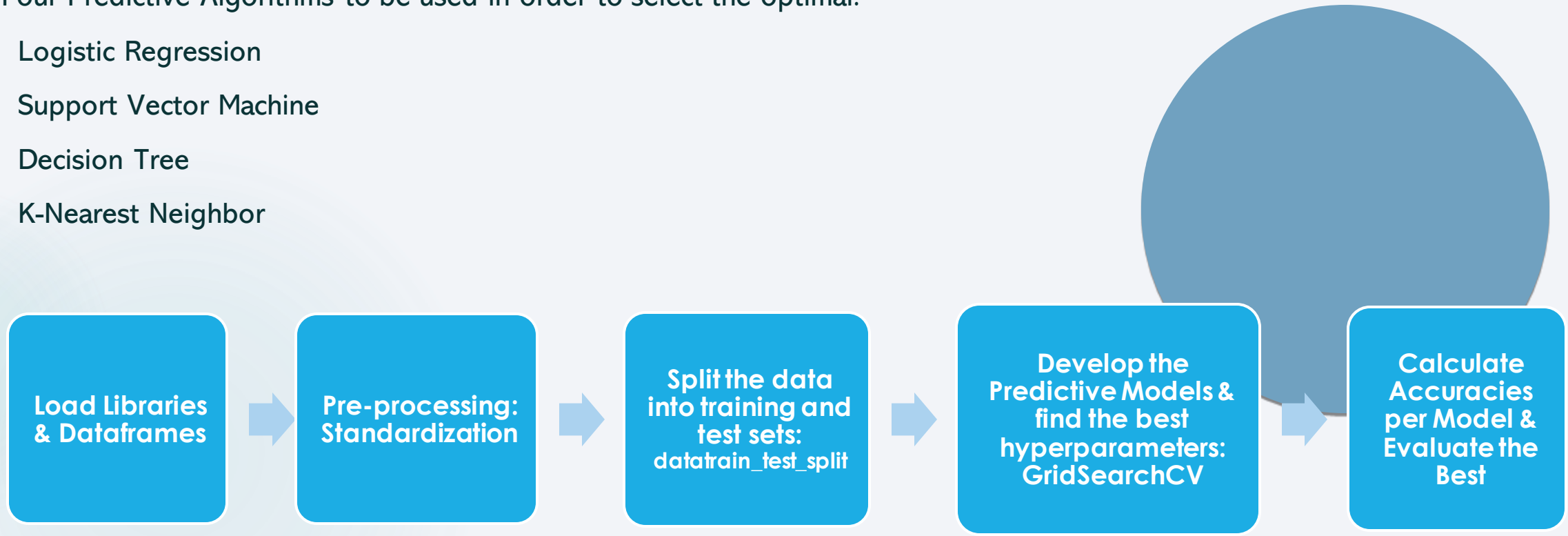


- GitHub URL: https://github.com/ageanwave/Applied-Data-Science-Capstone/blob/main/Week%203/spacex_dash_app.py

Predictive Analysis (Classification)

- Four Predictive Algorithms to be used in order to select the optimal:

1. Logistic Regression
2. Support Vector Machine
3. Decision Tree
4. K-Nearest Neighbor



GitHub URL: https://github.com/ageanwave/Applied-Data-Science-Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

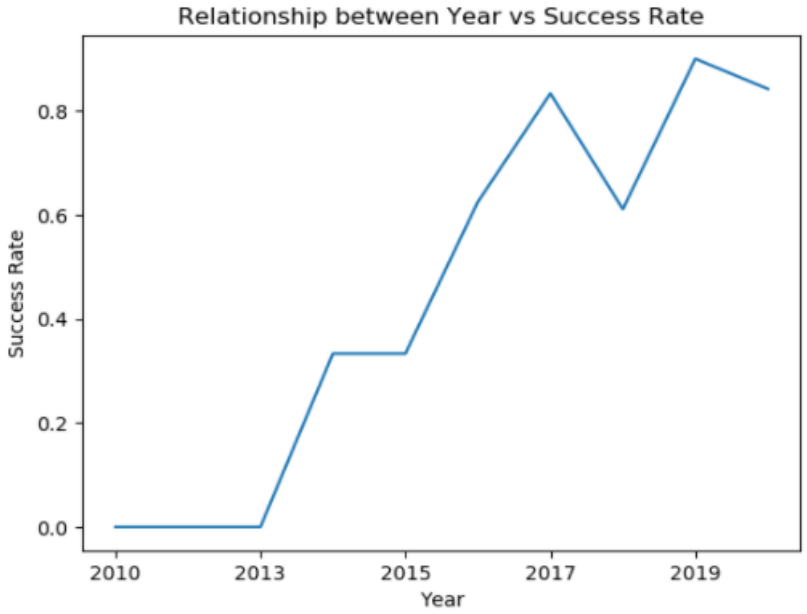
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Evaluate & find the Best Predictive Model by Accuracy Rate

Section 2

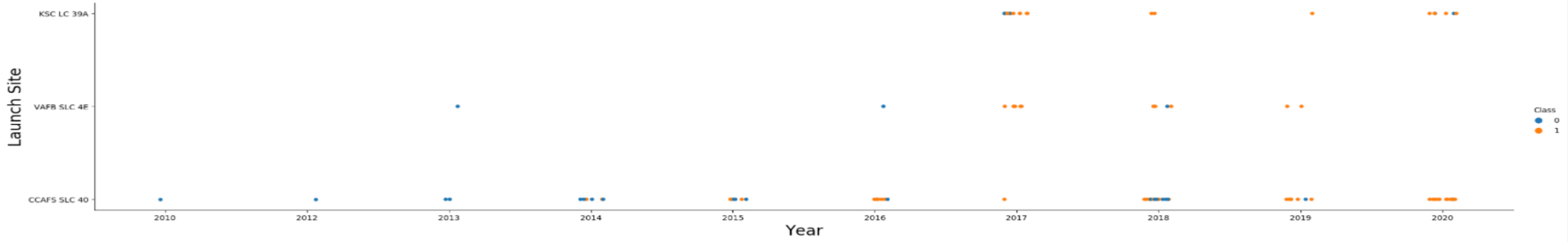
Insights drawn from EDA

Launch Success Yearly Trend

- A Line Chart of Success Rate vs Year is given on the right. We can easily identify the upward trend
- A Scatter Plot of Launch Sites per Year, hue Class. Again, visibly we can verify the increase in successful launches from 2016 onwards.




```
[52]: # Plot a scatter plot of Launch Site vs Year, hue Class
sns.catplot(y="LaunchSite", x="Year", hue="Class", data=df, aspect = 5)
plt.xlabel("Year",fontsize=20)
plt.ylabel("Launch Site",fontsize=20)
plt.show()
```



All Launch Site Names

- Names of the unique launch sites:
- CCAFS LC-40,
- CCAFS SLC-40,
- KSC LC-39A,
- VAFB SLC-4E

```
In [5]: %sql SELECT DISTINCT(LAUNCH_SITE) from SPACEXDATASET
 * ibm_db_sa://jnp40121:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb
Done.
```

Out[5]:

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

Launch Site Names Begin with 'CCA'

- Listed 5 records where launch sites begin with `CCA`:

```
In [9]: %sql SELECT * from SPACEXDATASET WHERE (LAUNCH_SITE like 'CCA%%') LIMIT 5
 * ibm_db_sa://jnp40121:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb
Done.
```

Out[9]:

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

- Total payload carried by boosters from NASA (CRS) given:

```
In [46]: %sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXDATASET WHERE CUSTOMER='NASA (CRS)'
          * ibm_db_sa://jnp40121:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb
         Done.
Out[46]:
    1
45596
```

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1:

```
In [13]: %sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXDATASET WHERE Booster_Version LIKE 'F9 v1.1%'
          * ibm_db_sa://jnp40121:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb
         Done.
Out[13]:    1
            2534
```

First Successful Ground Landing Date

- Date of the first successful landing outcome on ground pad:

```
In [51]: %sql SELECT MIN(DATE) FROM SPACEXDATASET WHERE LANDING__OUTCOME = 'Success (ground pad)'
 * ibm_db_sa://jnp40121:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb
Done.
Out[51]:
1
2015-12-22
```

Successful Drone Ship Landing with Payload between 4000 and 6000

- Names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000:

```
In [93]: %sql SELECT booster_version, PAYLOAD_MASS__KG_ FROM SPACEXDATASET WHERE LANDING__OUTCOME = 'Success (drone ship)' AND PAYLOAD_MAS
 * ibm_db_sa://jnp40121:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb
Done.
```

Out[93]:

booster_version	payload_mass__kg_
F9 FT B1022	4696
F9 FT B1026	4600
F9 FT B1021.2	5300
F9 FT B1031.2	5200

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes:

```
In [92]: %sql SELECT Mission_Outcome, COUNT Mission_Outcome FROM SPACEXDATASET GROUP BY Mission_Outcome
 * ibm_db_sa://jnp40121:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb
Done.
```

Out[92]:

mission_outcome	mission_outcome_1
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- Names of the booster which have carried the maximum payload mass:

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

In [83]:
```
%sql SELECT BOOSTER_VERSION, PAYLOAD_MASS__KG_ FROM SPACEXDATASET WHERE PAYLOAD_MASS__KG_=(SELECT MAX(PAYLOAD_MASS__KG_) FROM SPA
```

```
 * ibm_db_sa://jnp40121:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb
Done.
```

Out[83]:

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

- Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015:

```
In [74]: %sql SELECT DATE, LANDING__OUTCOME, BOOSTER_VERSION, LAUNCH_SITE FROM SPACEXDATASET WHERE LANDING__OUTCOME='Failure (drone ship)'
 * ibm_db_sa://jnp40121:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb
Done.
```

Out[74]:

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

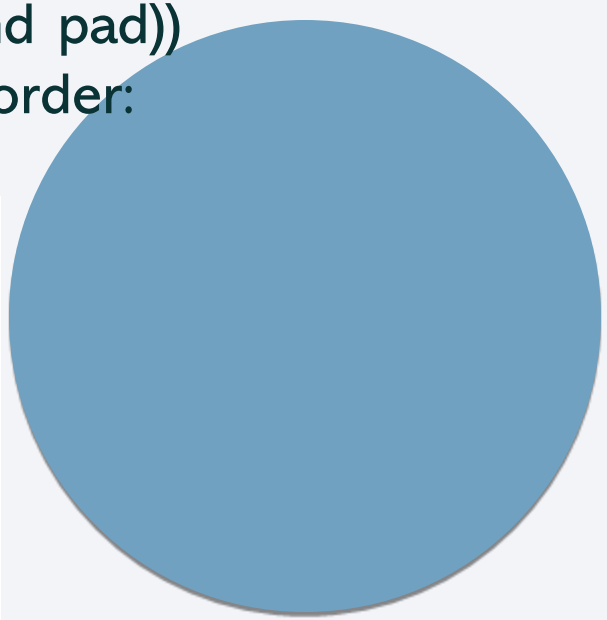
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- 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) FROM SPACEXDATASET WHERE DATE BETWEEN '2010-06-04' AND
```

```
 * ibm_db_sa://jnp40121:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb
Done.
```

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



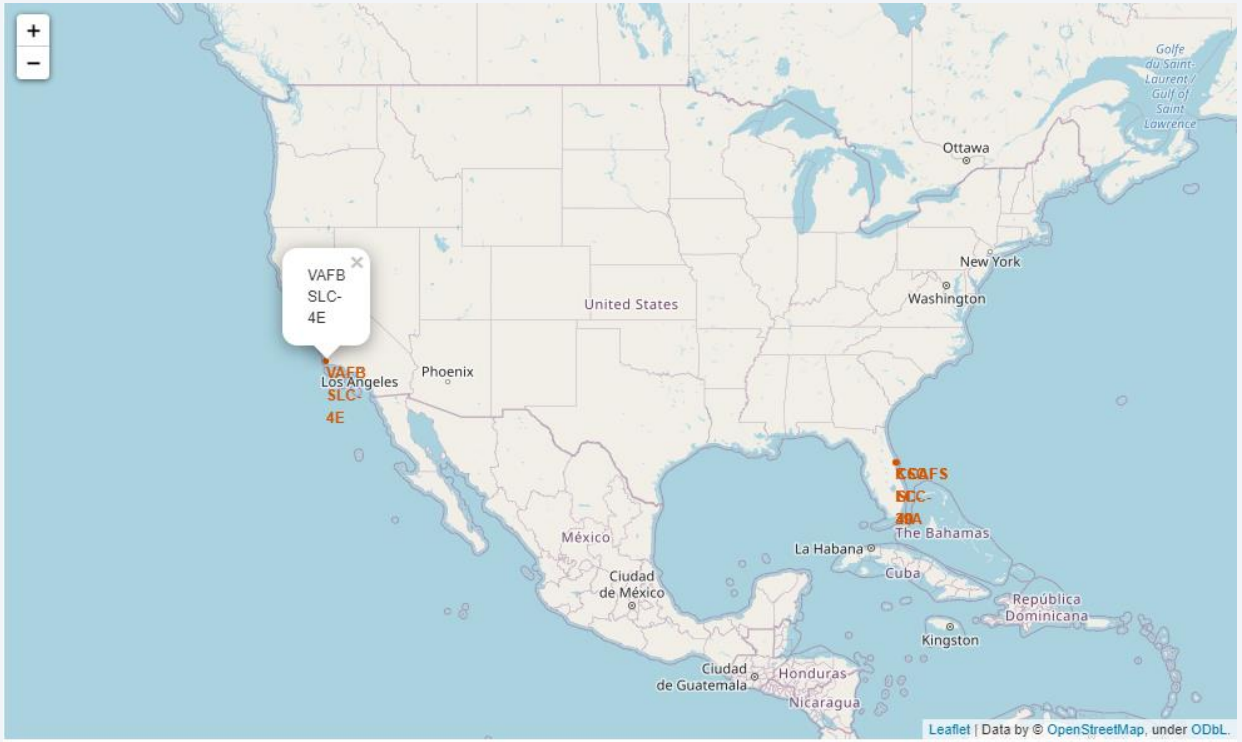
Section 4

Launch Sites Proximities Analysis

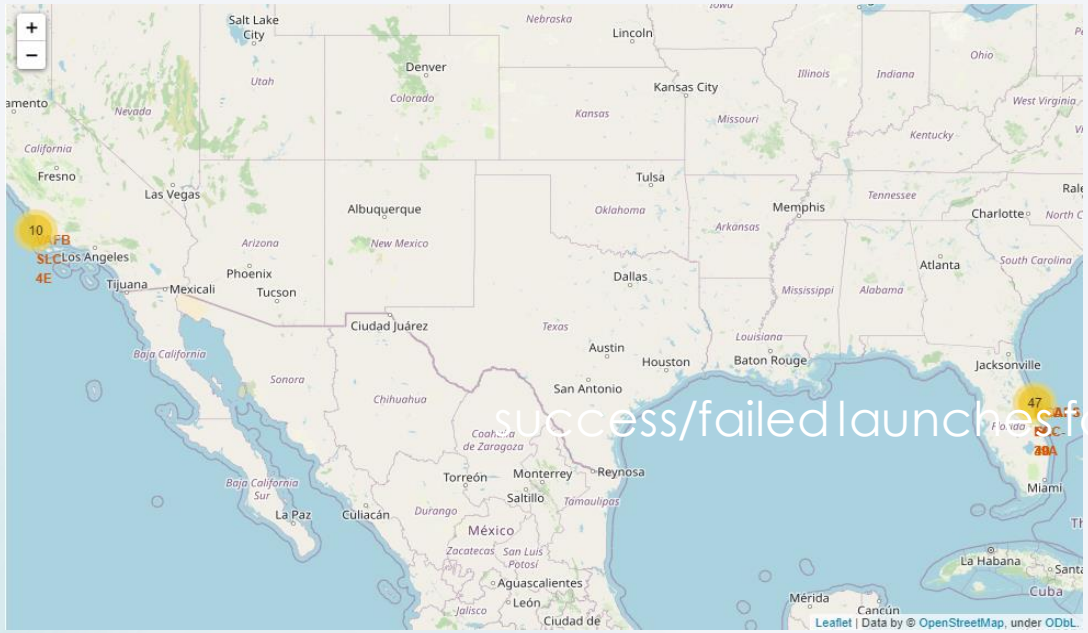
Launch Sites Used

	Launch Site	Lat	Long
0	CCAFS LC-40	28.562302	-80.577356
1	CCAFS SLC-40	28.563197	-80.576820
2	KSC LC-39A	28.573255	-80.646895
3	VAFB SLC-4E	34.632834	-120.610746

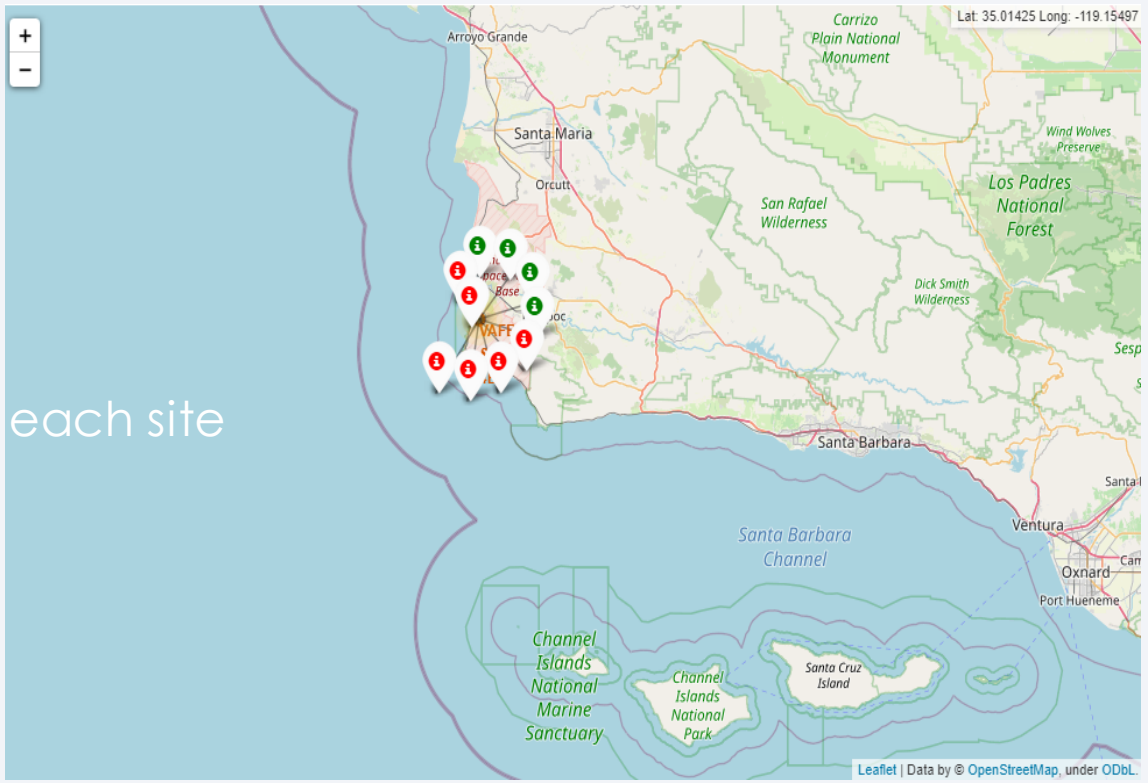
- Four Launch Site Locations across the US.
- A Circle to Mark each Site
- Markers Added to Name each Site



Showing Success & Failed launches for each site



success/failed launches for each site



- Marker Clusters added for each site
- Green for Successful Launches – Red for Failed

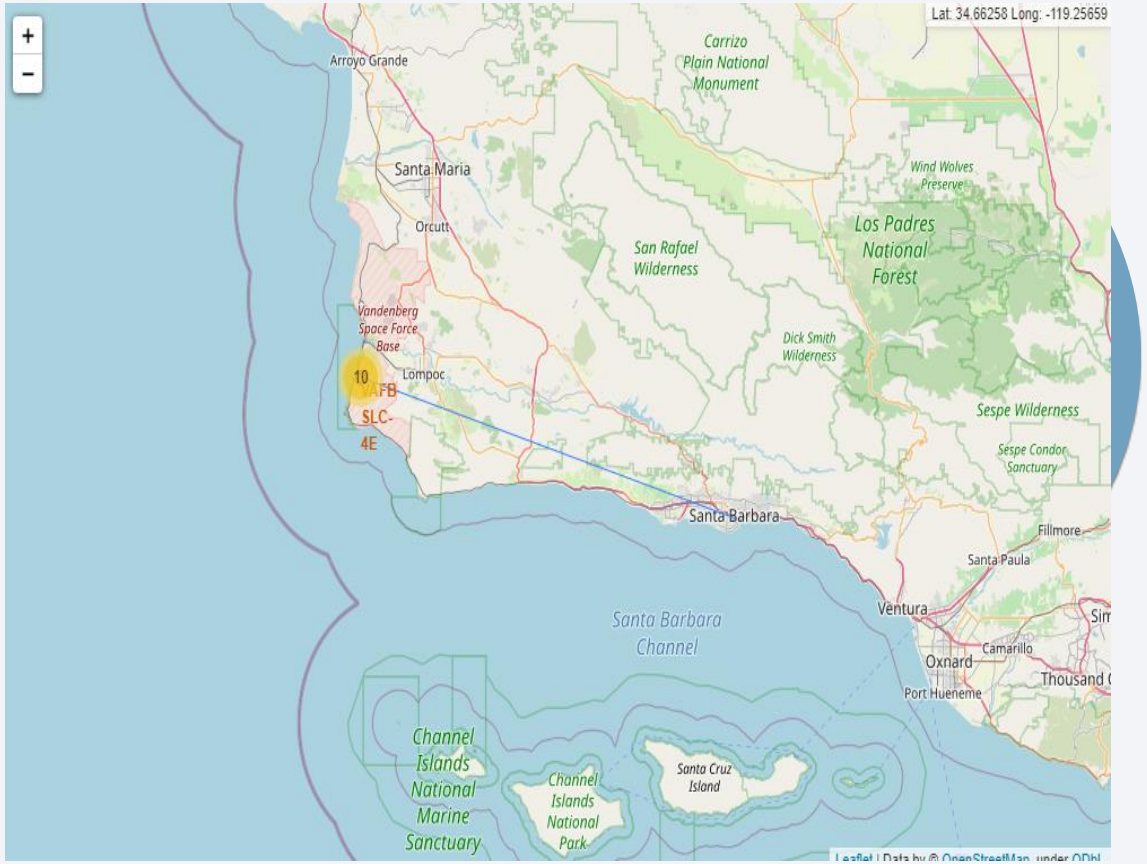
Proximity Analysis

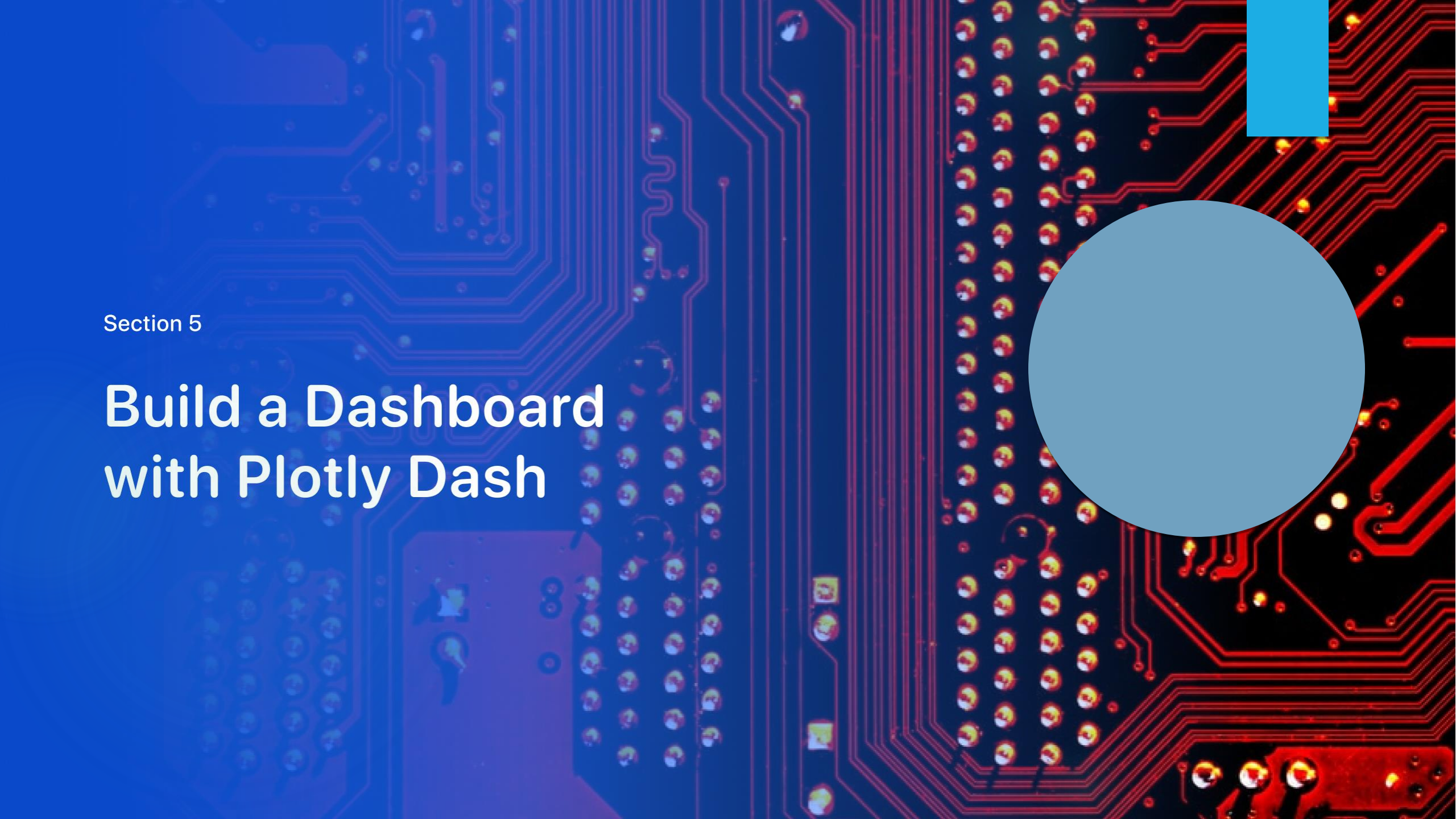

```
#Distance to Santa Barbara
launch_site_lat = 34.63283
launch_site_lon = -120.610746
sb_lat = 34.41584
sb_lon = -119.69549

distance_to_santabarbara = calculate_distance(launch_site_lat, launch_site_lon, sb_lat, sb_lon)
print('VAFB SLC-4E Site Distance to Santa Barbara: ', distance_to_santabarbara, 'km')
```

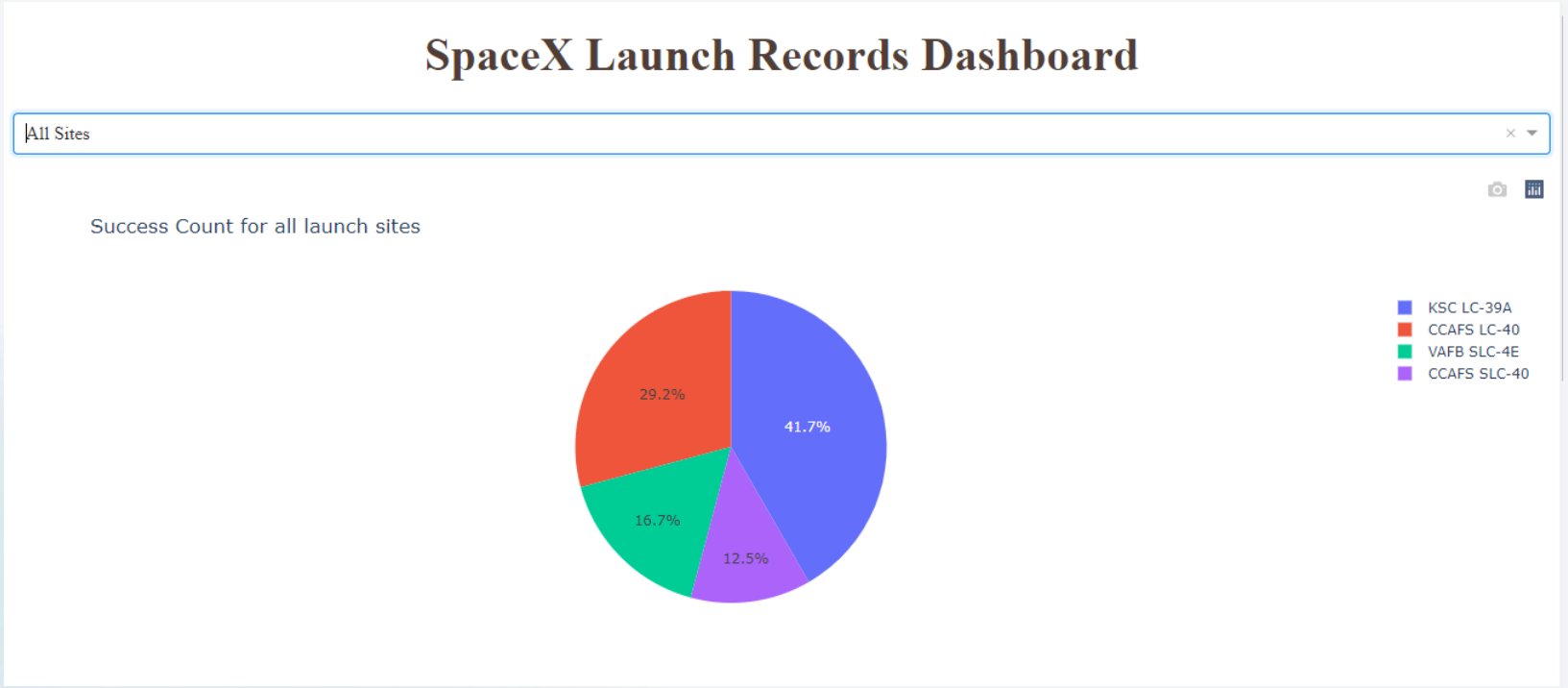
```
VAFB SLC-4E Site Distance to Santa Barbara:  87.27785859247399 km
```

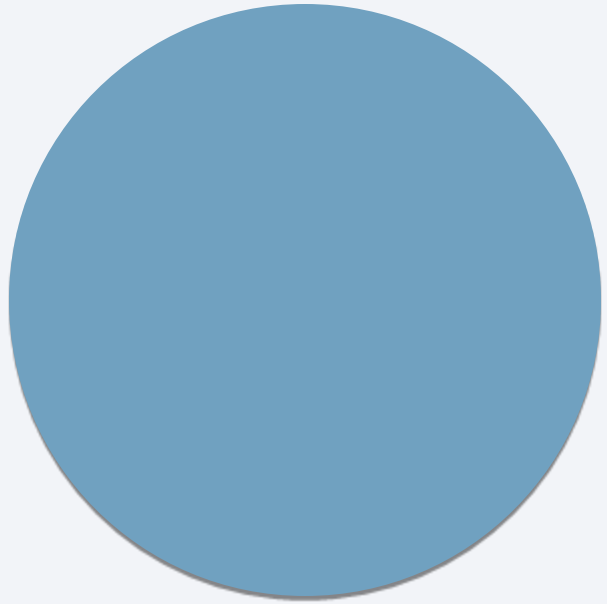
- We've performed an analysis on the VAFB SLC-4E launch site to determine the procimity to nearby cities.
- Distance to Santa Barbara found to be over 87km, long enough not to disturb residents




Section 5
Build a Dashboard with Plotly Dash

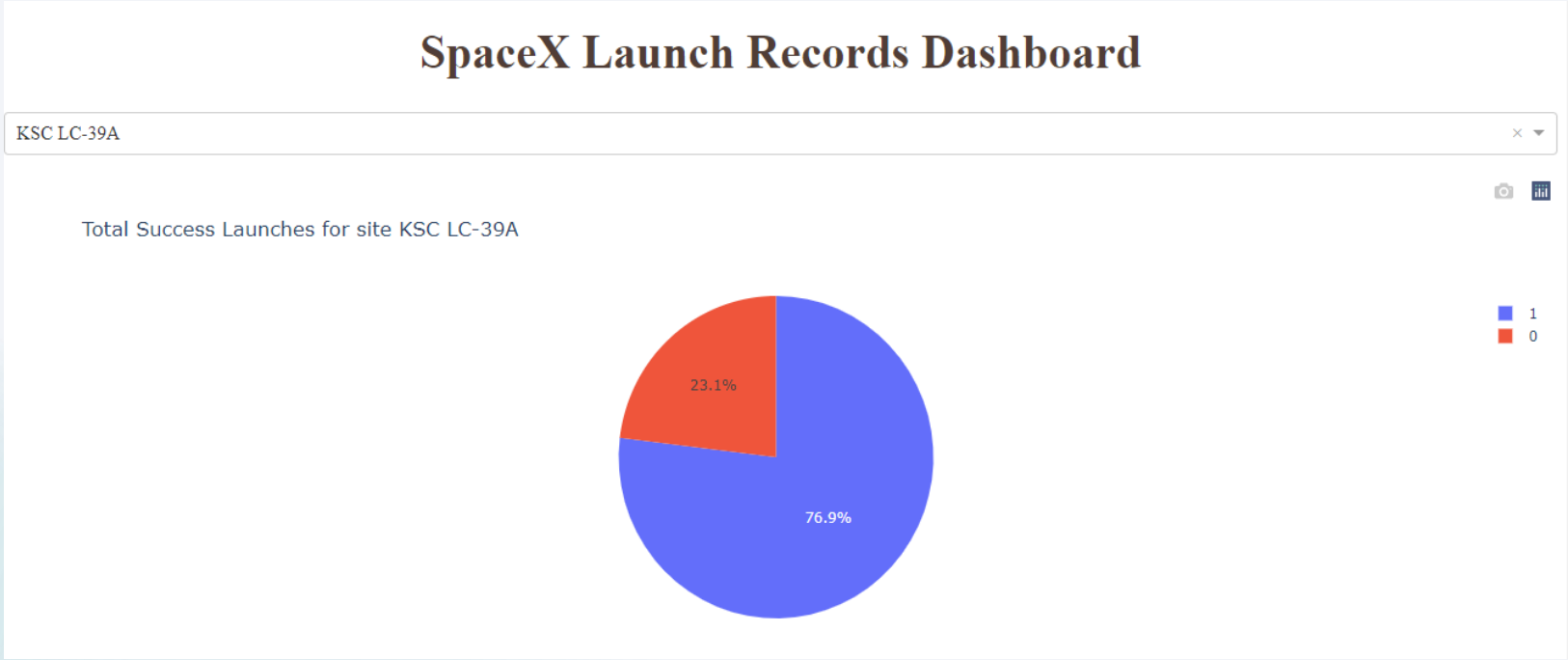
Dash Pie - Success Count for all Sites

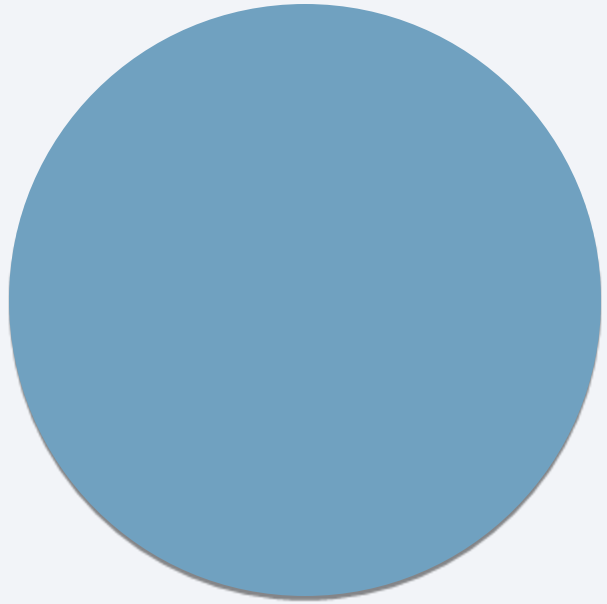




- Successful Launches per site given on the above pie chart, as a percentage of the total successful launches,
- Most successful launches from KSC LC-39A, 41.7% of the total.

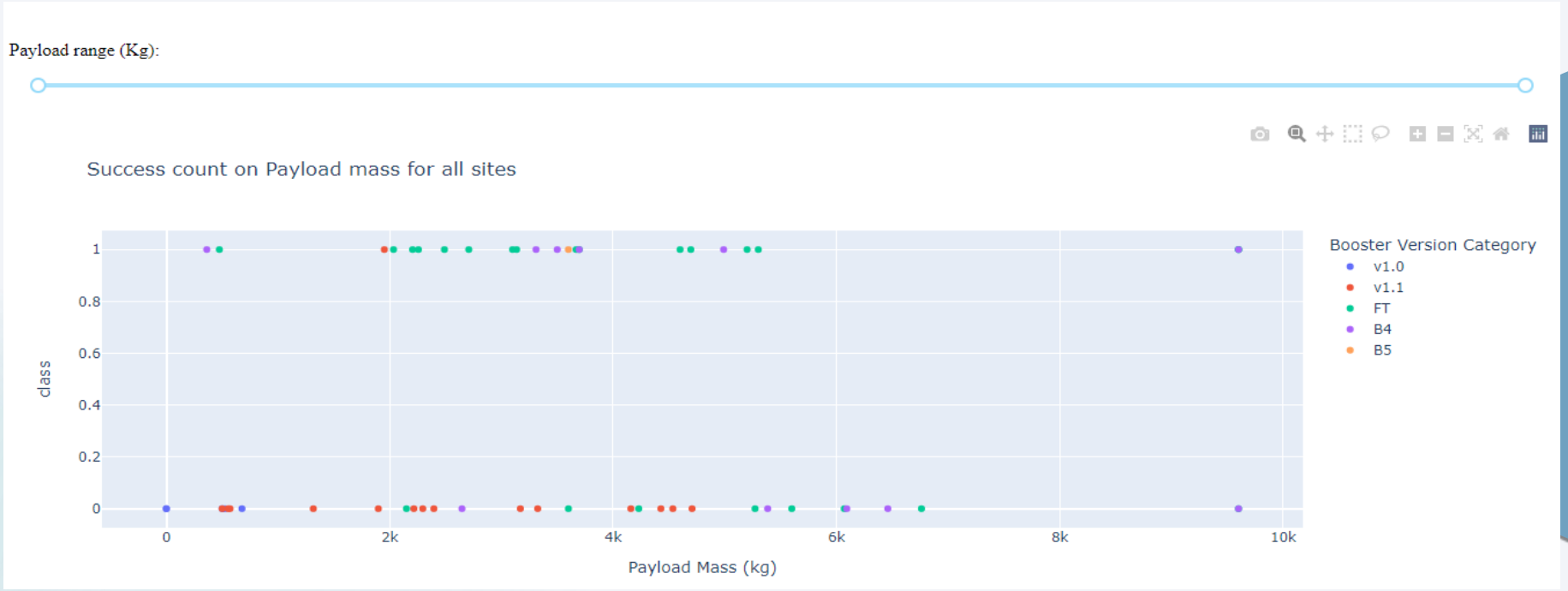
Dash Pie - Total Launches for KSC LC-39A

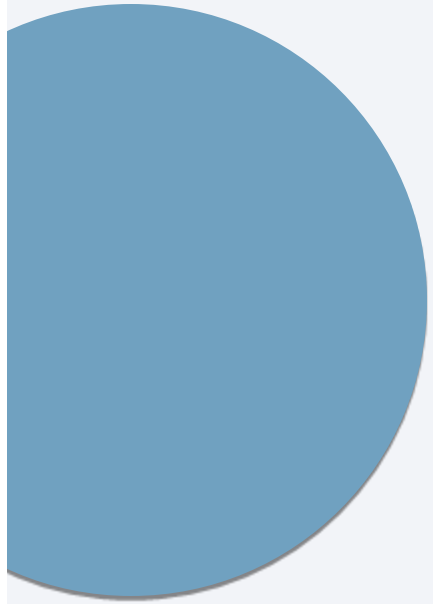




- On KSC LC-39A:
  - 76.9% of Launches were successful,
  - 23.1% were Failures.

Dash Graph - Success Count on All Payloads

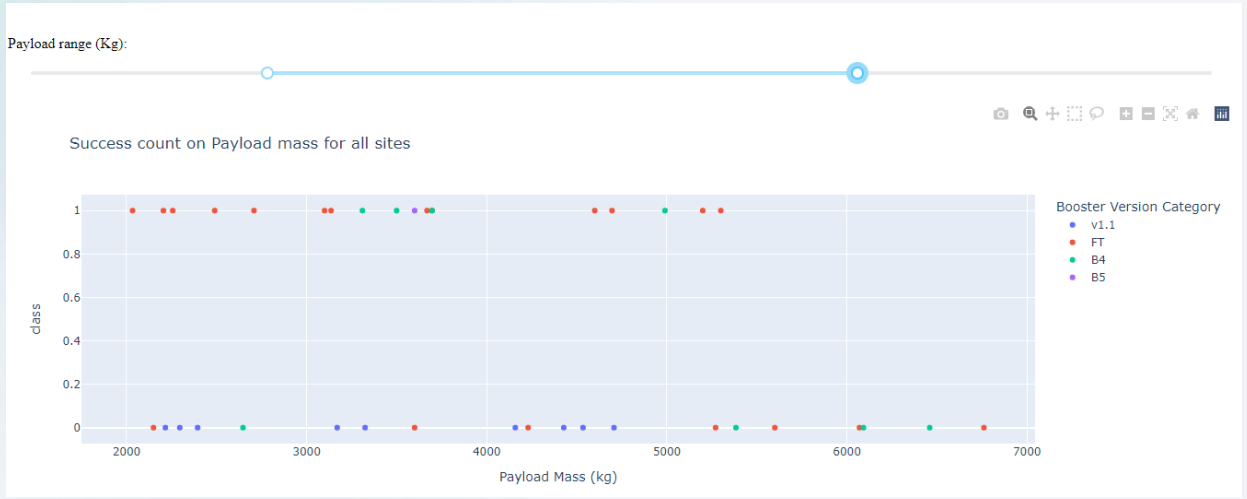




- Success count for all payloads given, colored by Booster Version Category.
- FT category seems to have the most successes overall
- v1.1 mostly Failures.

Dash Graph - Success Count on Various Payloads

- Success Count for 0-2k Payloads,
- Less Successes in this Range (B4 & FT,
- v1.1 most Failures.

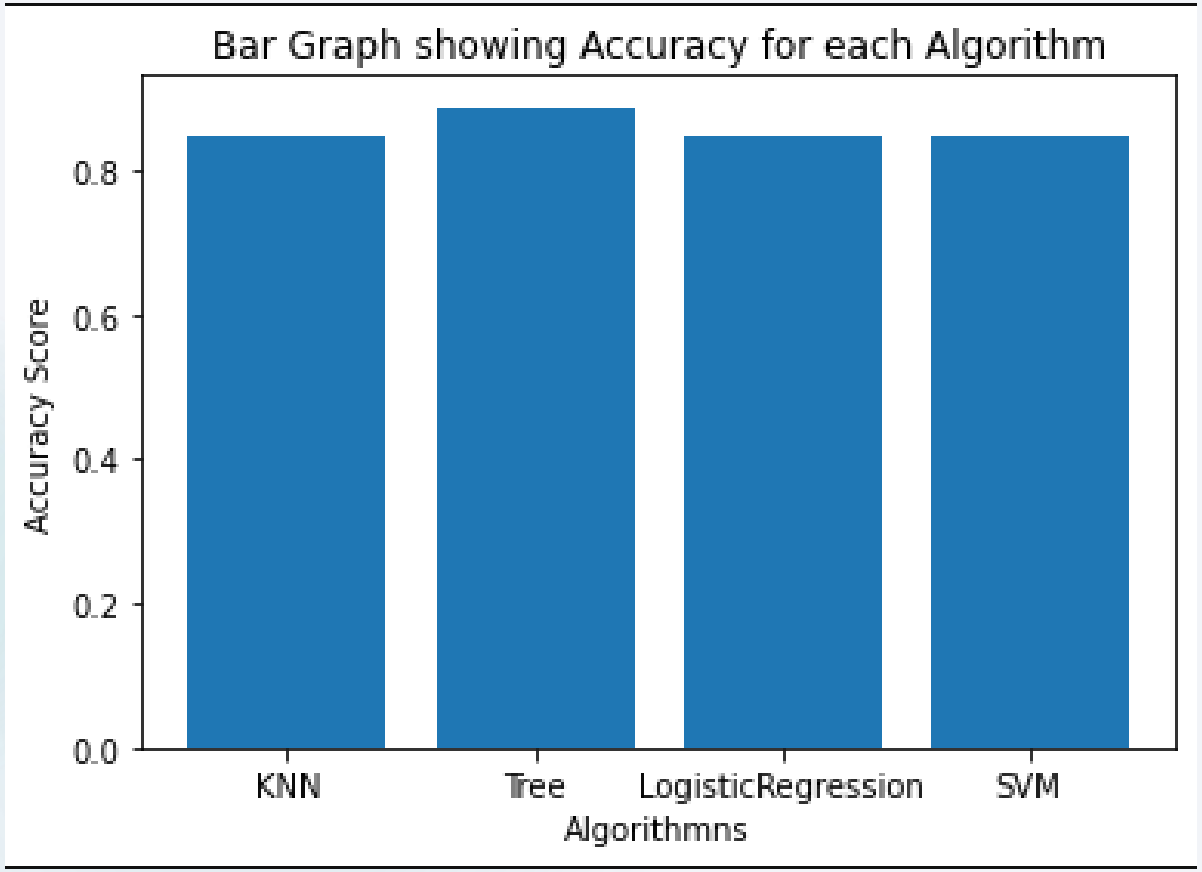


- Success Count for 2-7k Payloads,
- Most Successes in this Range (B4 & FT,
- Again, v1.1 most Failures.

Section 6

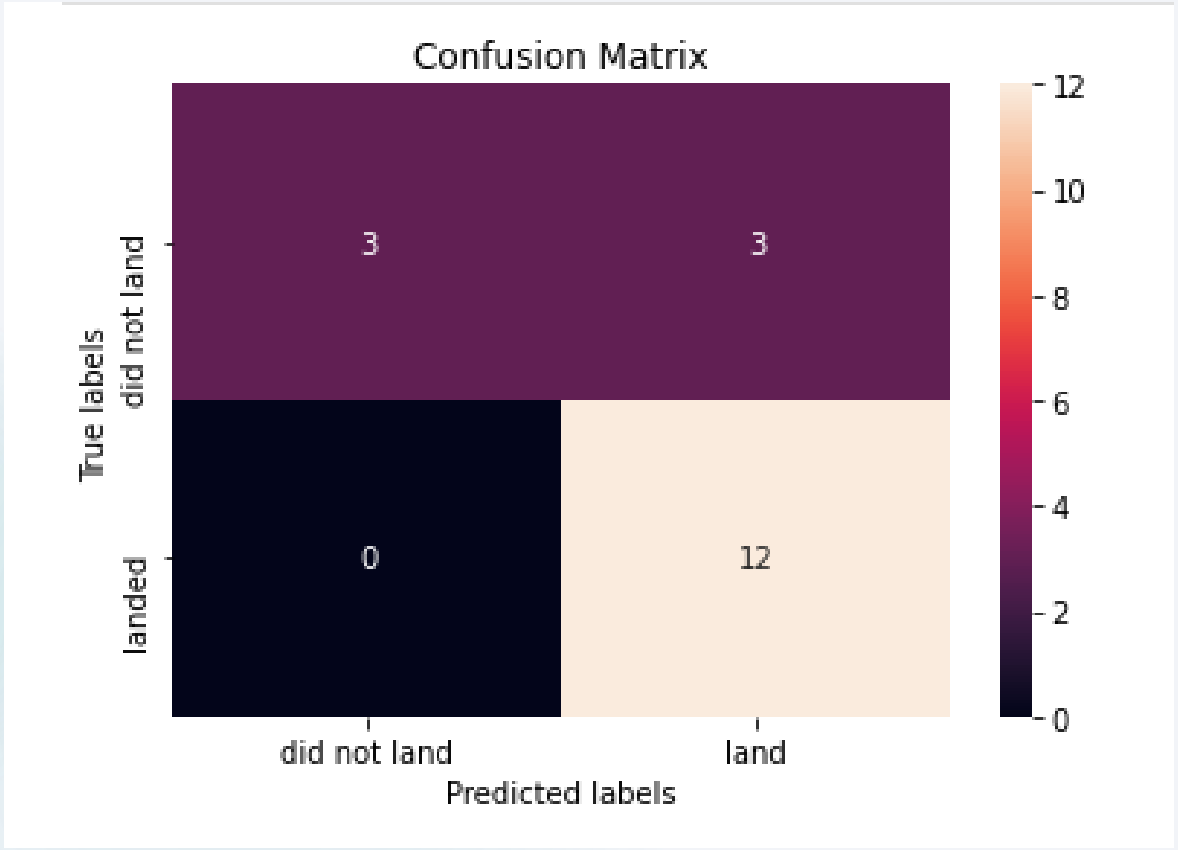
Predictive Analysis (Classification)

Classification Accuracy



- Accuracy per Model:
  1. KNN : 0.848
  2. DT : 0.887
  3. LogReg : 0.846
  4. SVM : 0.848
- Best Performing Algorithm: Decision Tree Classifier
- Best Accuracy: 0.887

Decision Tree Confusion Matrix



- True Positive : 12
- True Negative : 3
- False Positive : 3
- False Negative : 0

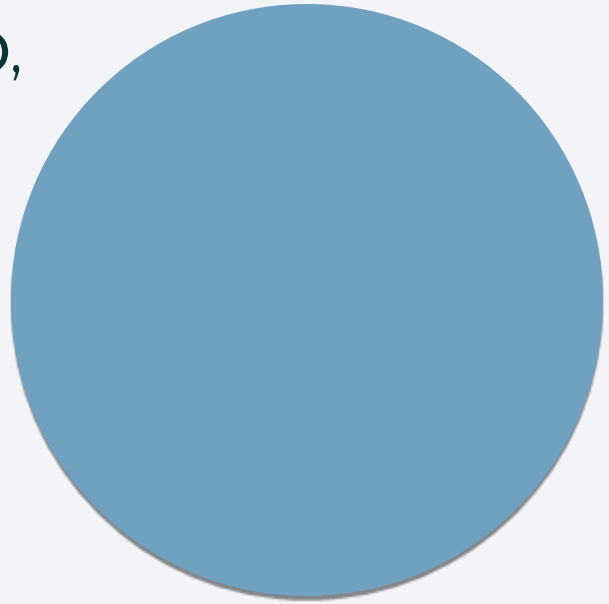
Looks Satisfactory!

Conclusions

- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.
- ES-L1, GEO, HEO & SSO orbits have the highest success rates.
- Mid & Heavy Range Payloads have better Success Rates.
- Proximity Analysis shows Launch Sites far enough from residential areas.
- KSC LC-39A Launch Site has most successful launches,
- FT Booster version category most Successful, v1.1 most Failures,
- Best Performing Predictive Algorithm: Decision Tree Classifier (Accuracy: 0.887) .

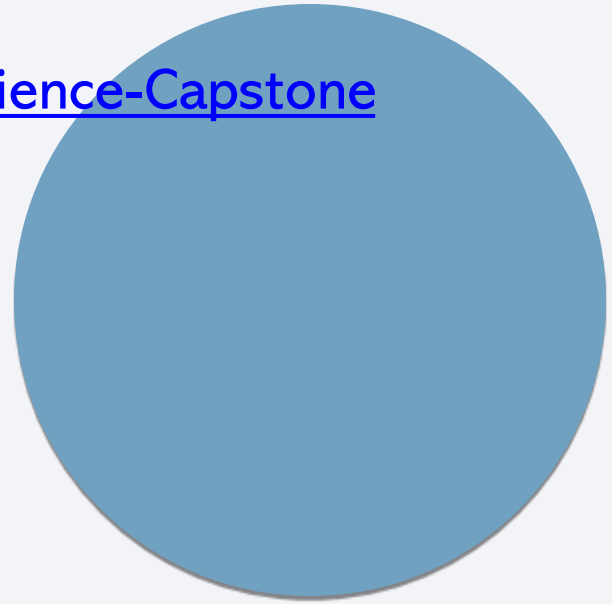
Suggestions

- Focus on the orbits have the highest success rates: ES-L1, GEO, HEO & SSO,
- Launch Site of Preference: KSC LC-39A,
- Focus on Mid & Heavy Range Payloads,
- Use FT Booster versions over others,
- Use Decision Tree Classifier for future predictions!



Appendix

- Top github repo url: https://github.com/ageanwave/Applied-Data-Science-Capstone



Thank you!