Winning Space Race with Data Science

Maroof Chowdhury
30-June-2024

Outline

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

Executive Summary

Summary of methodologies

- API Based Data Collection/ Data Wrangling
- EDA (Data Visualisation) & EDA SQL
- UI Based: Interactive Map with Folium & Dashboard with Plotly Dash
- Classification Based: Predictive Analysis

Summary of all results

- EDA Represented Results
- Interactivity Based Analytics
- Predictive Analysis

Our Mission

How?

- Primarily SpaceX API / Wikipedia SpaceX
- Opportunity To Save Large Cost In Adoption ($62m vs. $165m)

Objective

- Usage of Data Machine Learning Models
- Falcon 9 First Stage Learning – Predictive Analysis To Determine Future Launch
- Let's Use Data To Make It A Success!

Section 1

Methodology

Methodology

How?

- Data Collection:Primarily SpaceX API / Wikipedia SpaceX
- Data Wrangling
- Exploratory Data Analysis(EDA)

With?

- SQL
- Folium
- Plotly Dash
- Classification Models (GridSearch CV/Predictive Analysis)

Data Collection

How?

- Primarily SpaceX API / Wikipedia SpaceX
- Opportunity To Save Large Cost In Adoption ($62m vs. $165m)

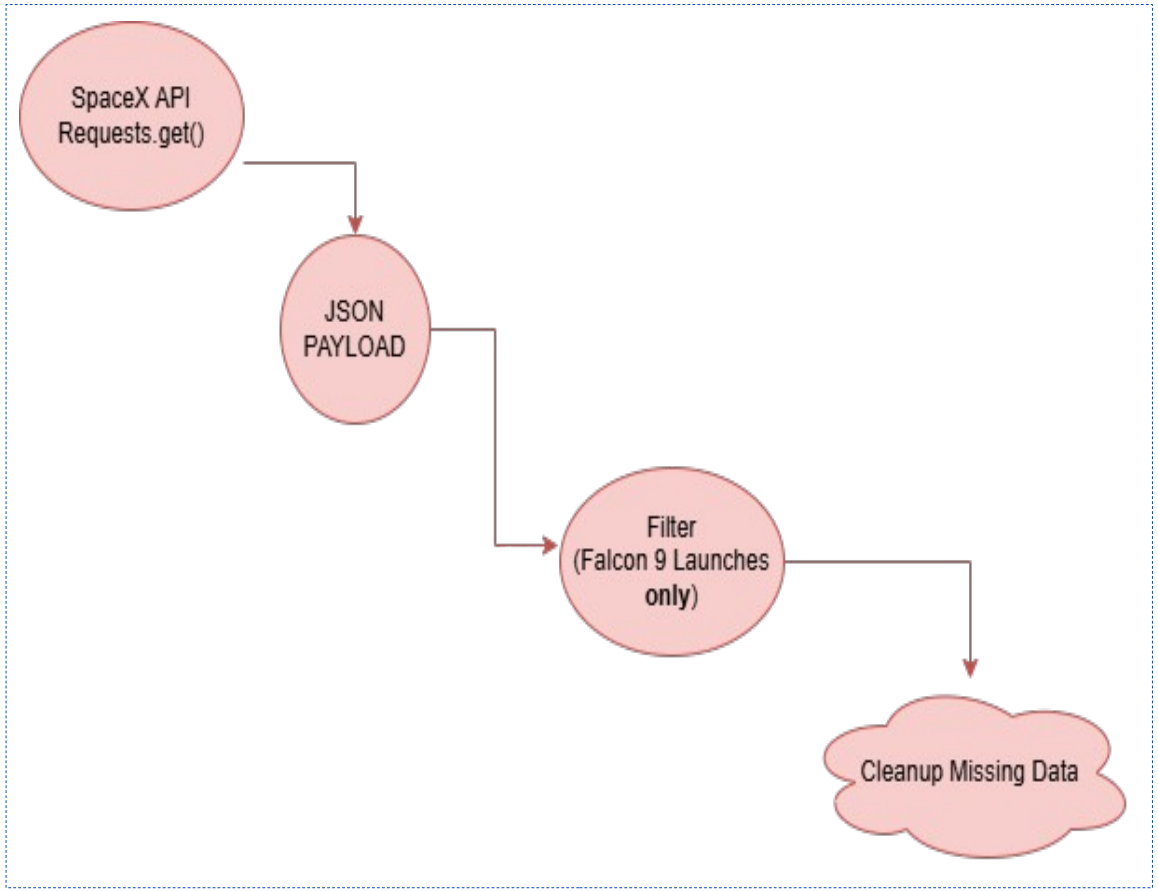
Objective

- Usage of Data Machine Learning Models
- Falcon 9 First Stage Learning – Predictive Analysis To Determine Future Launch
- Let's Use Data To Make It A Success!

SpaceX API

- **Public API:**

  https://api.spacexdata.com/v4
- Payload - JSON Format
- Dictionary Based



https://github.com/maroofc/Applied-Data-Science-Capstone-Project/blob/main/Data%20Collection%20API.ipynb

Scraping Wikipedia

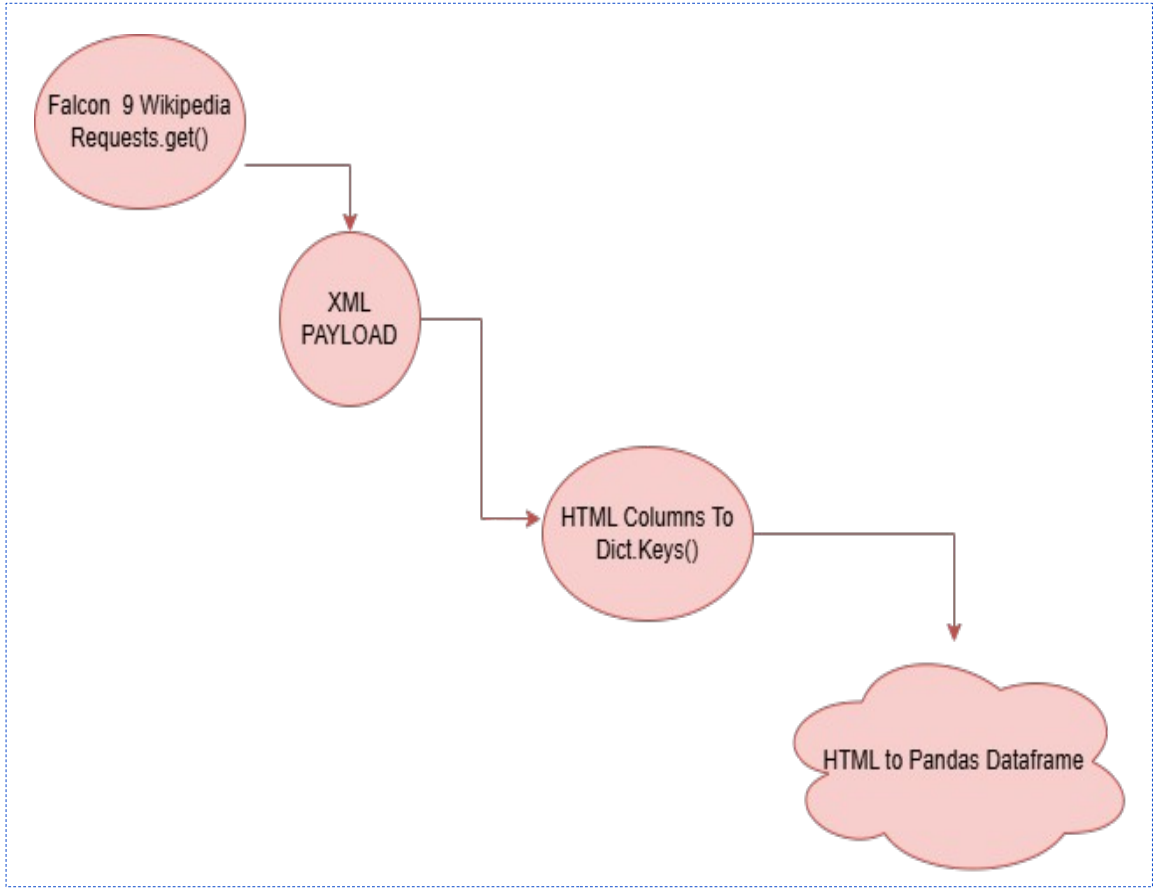
Public URL:

https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

Requests.get() method

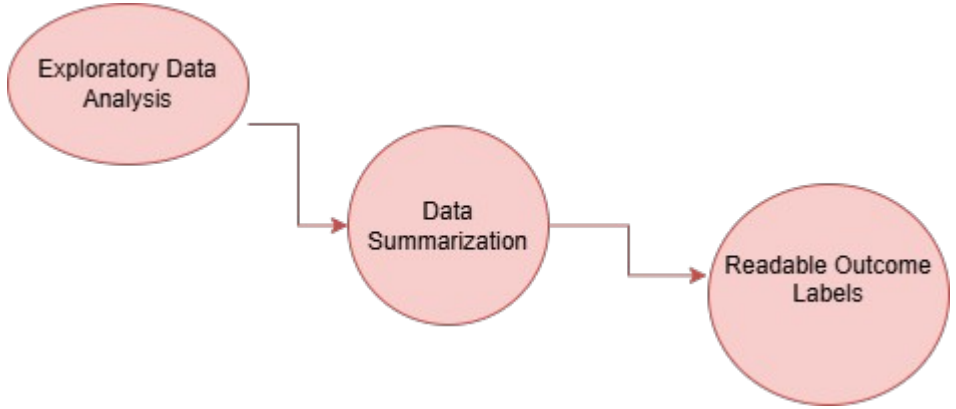
BeautifulSoup To Dict Keys

https://github.com/maroofc/Applied-Data-Science-Capstone-Project/blob/main/Data%20Collection%20with%20Web%20Scraping.ipynb



Data Wrangling

- Exploratory Data Analysis (EDA)
- Ran Against Dataset
- Data Cleanup
- Value Mapping Against True(1)/False(0)



https://github.com/maroofc/Applied-Data-Science-Capstone-Project/blob/main/Data%20Wrangling.ipynb

EDA With SQL

Examples

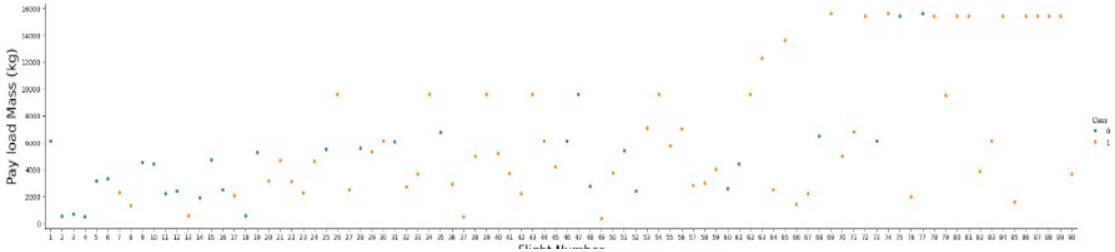
- Distinct Launch Sites
- Top 5 Launch Sites Starting with “CCA”
- Total Payload Mass per Carried Boosters Launched (NASA/CRS)
- Avg. Payload Mass Carried (F9 v.1.1)
- Failed Landing Outcomes with Details

https://github.com/maroofc/Applied-Data-Science-Capstone-Project/blob/main/EDA%20With%20SQL%20Analysis.ipynb

EDA With Data Visualisation

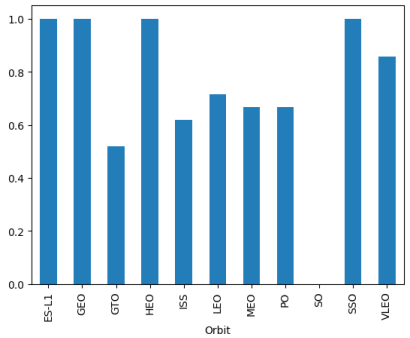
Data Inputs:

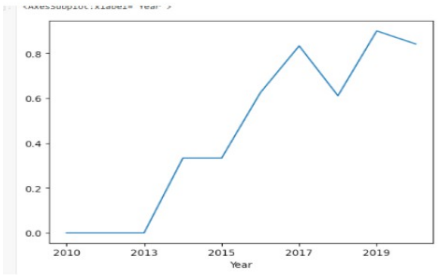
Flight Number vs. Payload Mass, Flight Number vs. Launch Site, Payload Mass vs. Launch Site, Orbit vs. Success Rate, Flight Number vs. Orbit, Payload vs Orbit, and Success Yearly Trend



EDA on Combination:

- Bar Plots
- Scatter Plots
- Line Plots





https://github.com/maroofc/Applied-Data-Science-Capstone-Project/blob/main/EDA%20With%20Data%20Visualisation.ipynb

Interactive Map with Folium

Map Objects Used:

- Circles – Points Of Interest Around Areas
- Markers – Launch Site Locations
- Lines – Distance Measurements Between 2 Coordinates

Correlations:

- Proximity To Key Locations
- Success/Failure Outcomes vs. Distance
- Location of Launch Sites vs. Outcomes

https://github.com/maroofc/Applied-Data-Science-Capstone-Project/blob/main/LaunchSitesWithFolium.ipynb

Dashboard with Plotly Dash

Main Graphs/Plots Used:

- Pie Chart
- Scatter Plot

Correlations:

- Landing Site vs. Launch Site Success Rate
- Success vs. Launchsite vs. Payload Mass vs. Booster Version

https://github.com/maroofc/Applied-Data-Science-Capstone-Project/blob/main/spacex_dash_app.py

Predictive Analysis(Classification)

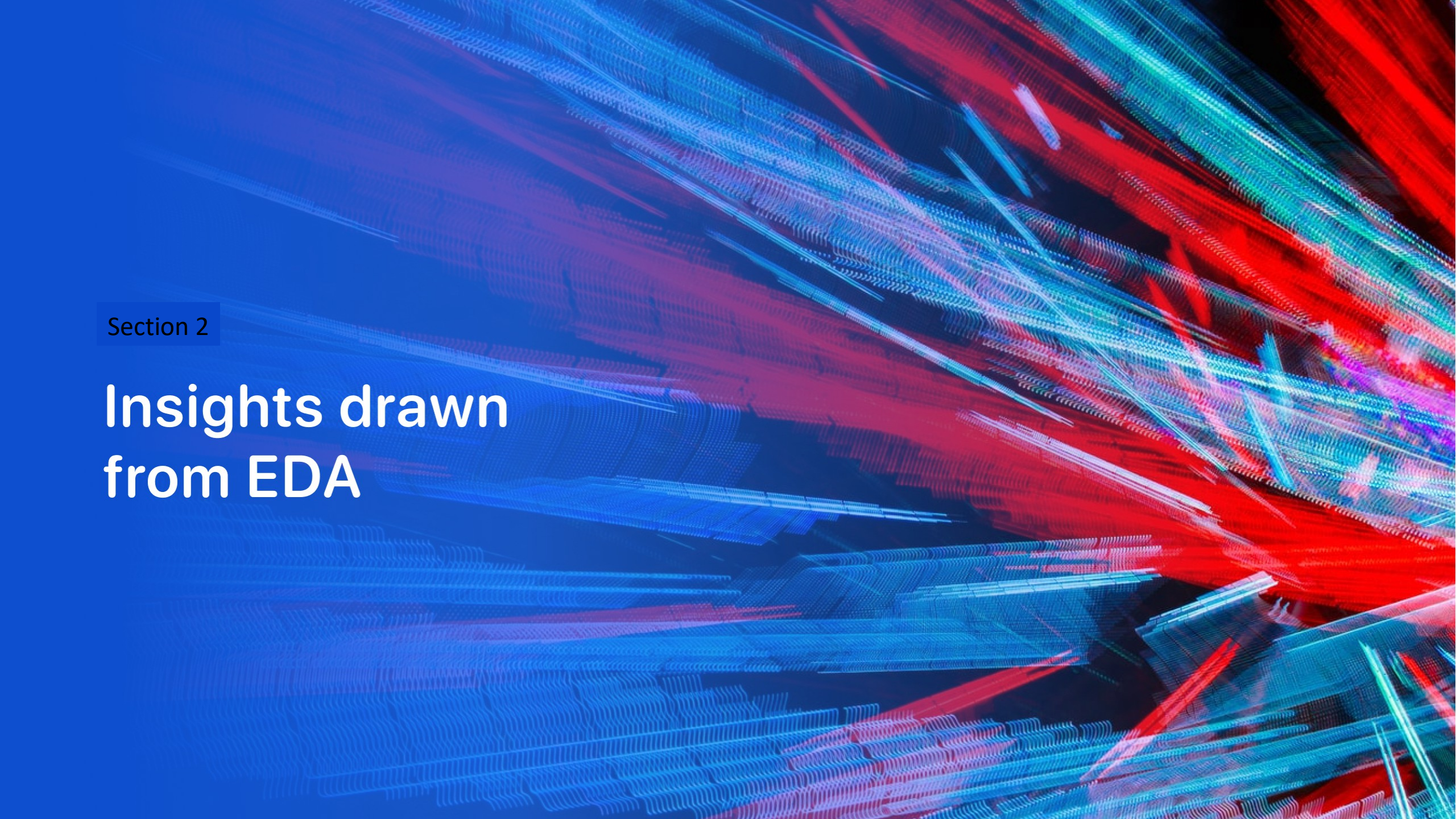
Classification Models:

- Support Vector Machine
- K Nearest Neighbours
- Decision Tree
- Logistic Regression

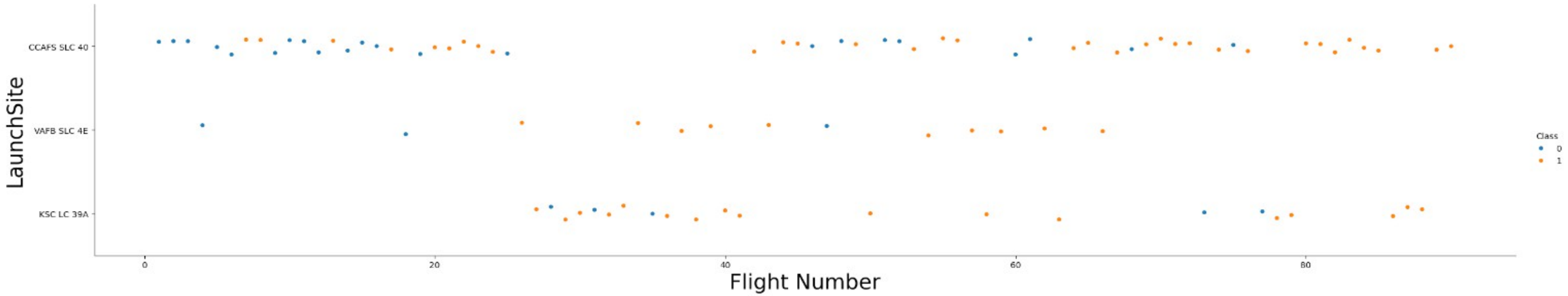
https://github.com/maroofc/Applied-Data-Science-Capstone-Project/blob/main/Predictive%20Analysis.ipynb

Results

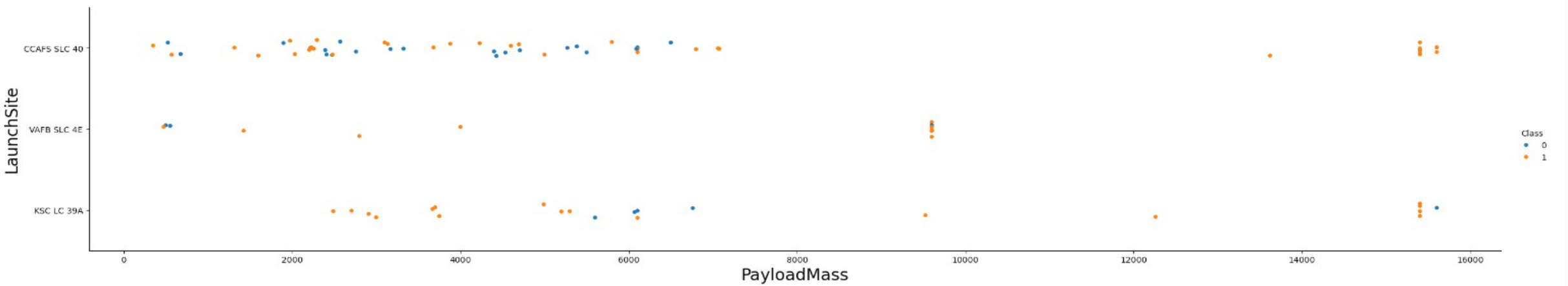
- 83% Accuracy
- Landing Outcomes Improved Over Time
- 1st Successful Year: 2015
- SpaceX Utilized 4 Launch Sites


Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

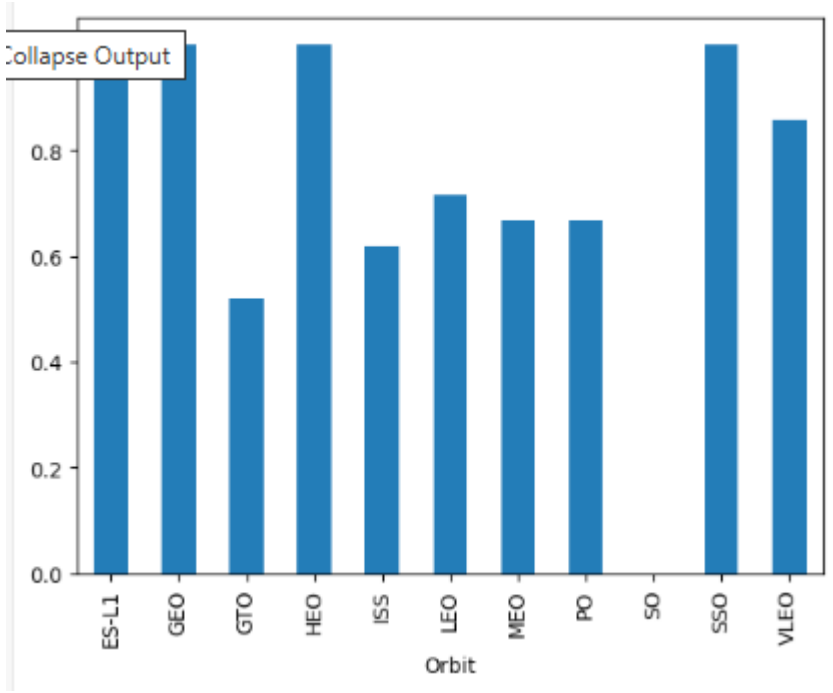


Payload vs. Launch Site

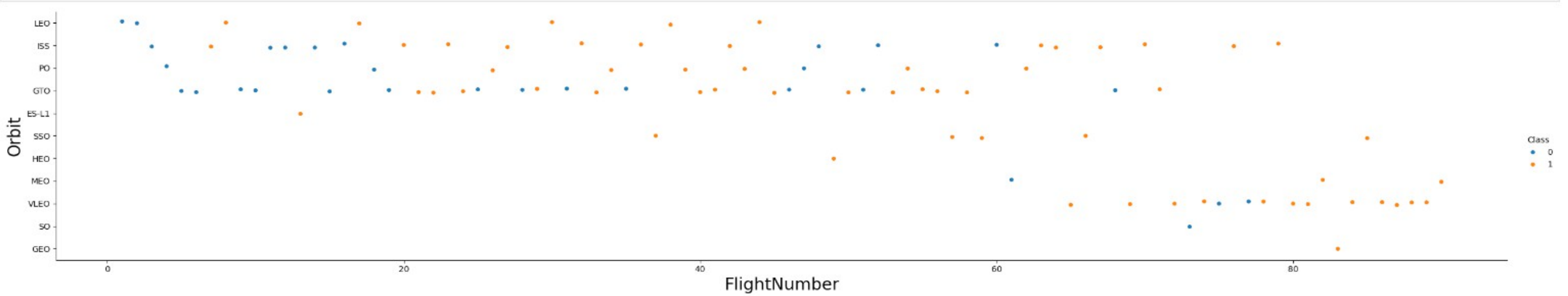


We also want to observe if there is any relationship between launch sites and their payload mass

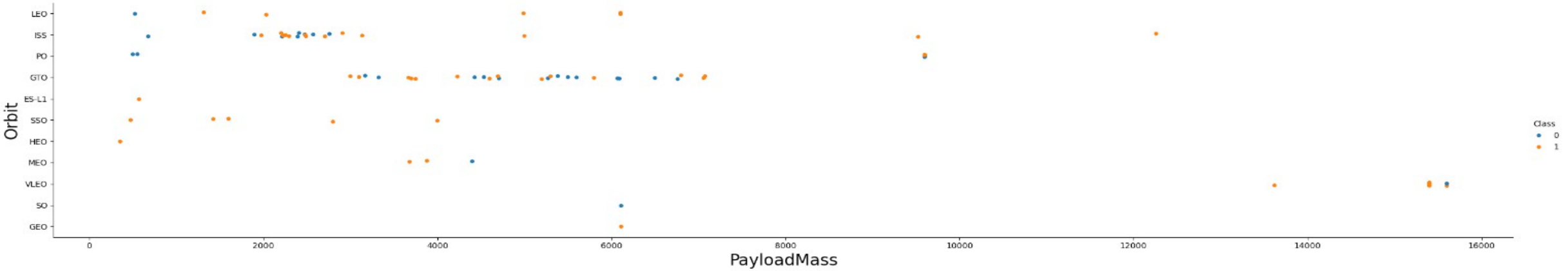
Success Rate vs. Orbit Type



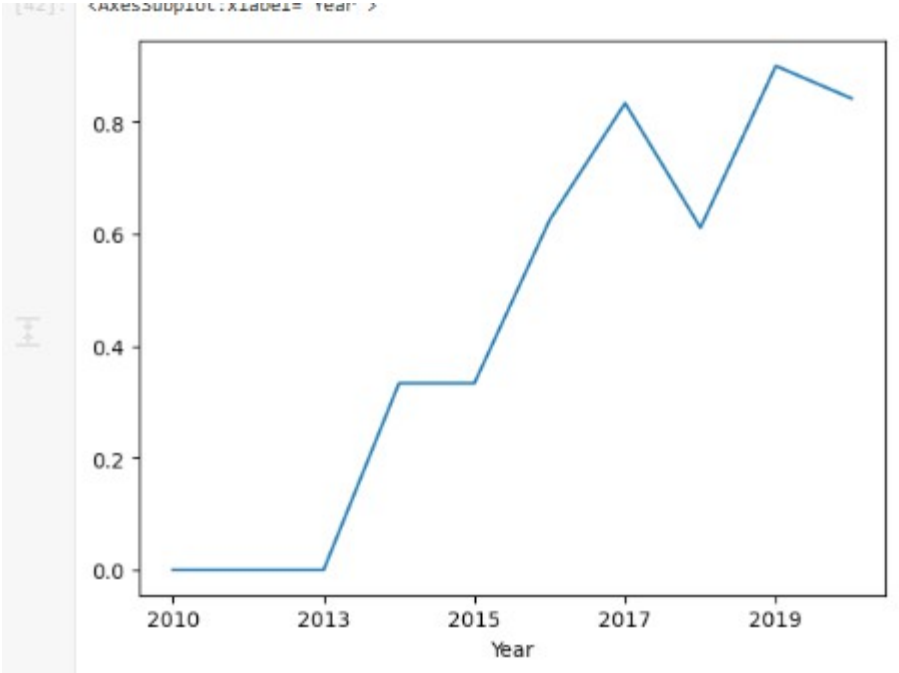
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



Launch Site Names

In [9]:
```
sql SELECT DISTINCT LAUNCH_SITE FROM SPACEXTBL ORDER BY 1;
```

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

Out[9]:

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

Launch Site Names with 'CCA'

```
10]: sql SELECT * FROM SPACEXTBL WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5;
 * sqlite:///my_data1.db
Done.
```

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

sql SELECT SUM(PAYLOAD_MASS__KG_) AS TOTAL_PAYLOAD FROM SPACEXTBL WHERE PAYLOAD LIKE '%CRS%';

Out[10]:

total_payload
111268

Average Payload Mass for Booster Version

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

Out[11]:

avg_payload
2928

First Successful Grounding Date

```
sql SELECT MIN(DATE) AS FIRST_SUCCESS_GP FROM SPACEXTBL WHERE LANDING__OUTCOME = 'Success (ground pad)';
```

Out[13]:

first_success_gp
2015-12-22

Successful Drone Ship Landing Stats

sql SELECT DISTINCT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000 AND LANDING__OUTCOME = 'Success (drone ship)';

Out[14]:

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

Total No. Success vs. Failure Outcomes

sql SELECT MISSION_OUTCOME, COUNT(*) AS QTY FROM SPACEXTBL GROUP BY MISSION_OUTCOME ORDER BY MISSION_OUTCOME;

Out[15]:

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

Boosters Carried vs. Max Load

sql SELECT DISTINCT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL) ORDER BY BOOSTER_VERSION;

booster_version
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

2015 Launch Records

sql SELECT BOOSTER_VERSION, LAUNCH_SITE FROM SPACEXTBL WHERE LANDING__OUTCOME = 'Failure (drone ship)' AND DATE_PART('YEAR', DATE) = 2015;

Out[24]:

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

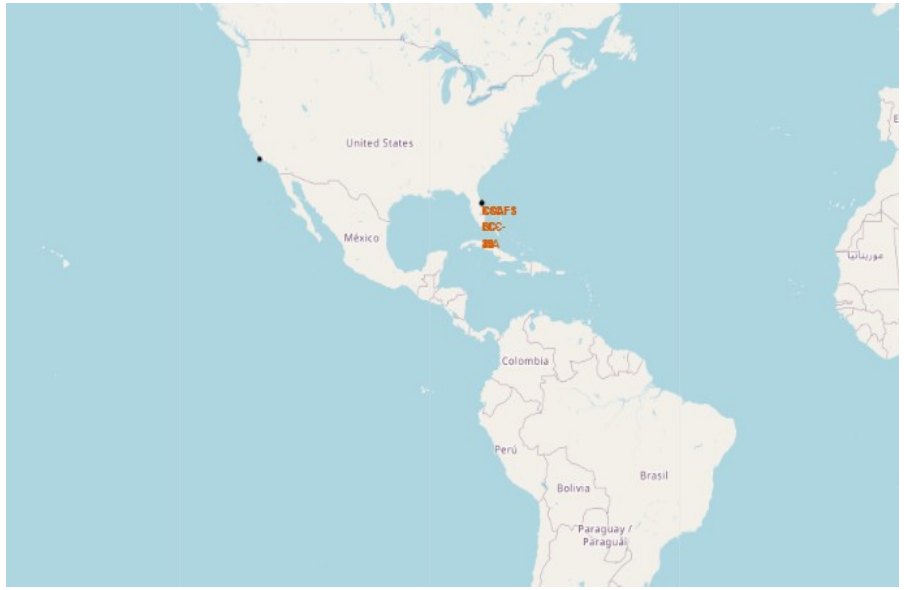
2010-2017 Landing Outcomes

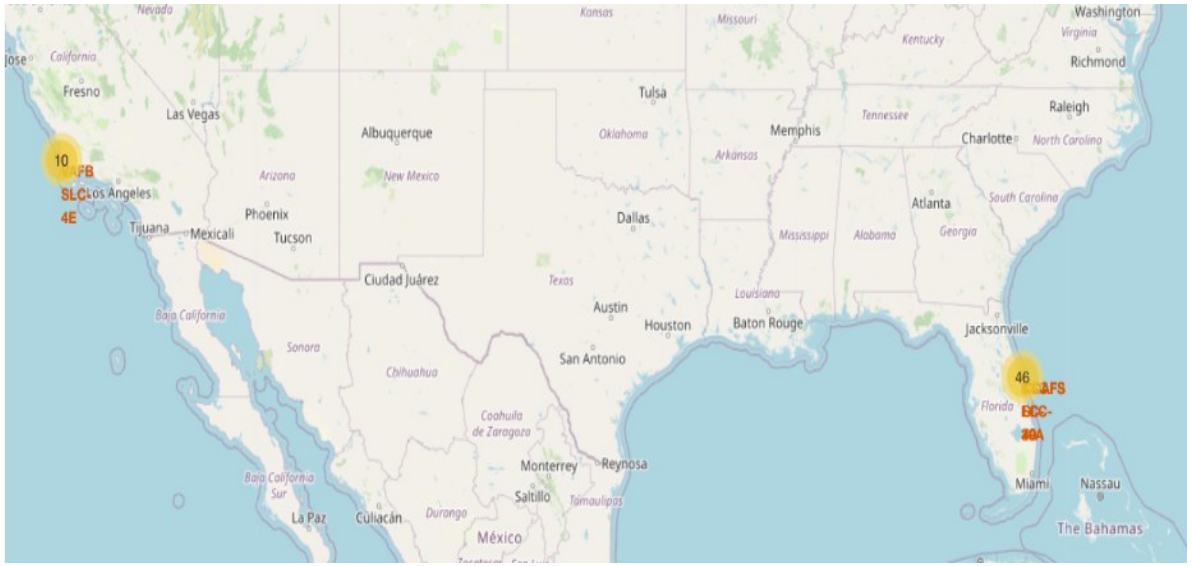
sql SELECT LANDING__OUTCOME, COUNT(*) AS QTY FROM SPACEXTBL WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY LANDING__OUTCOME ORDER BY QTY DESC;

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

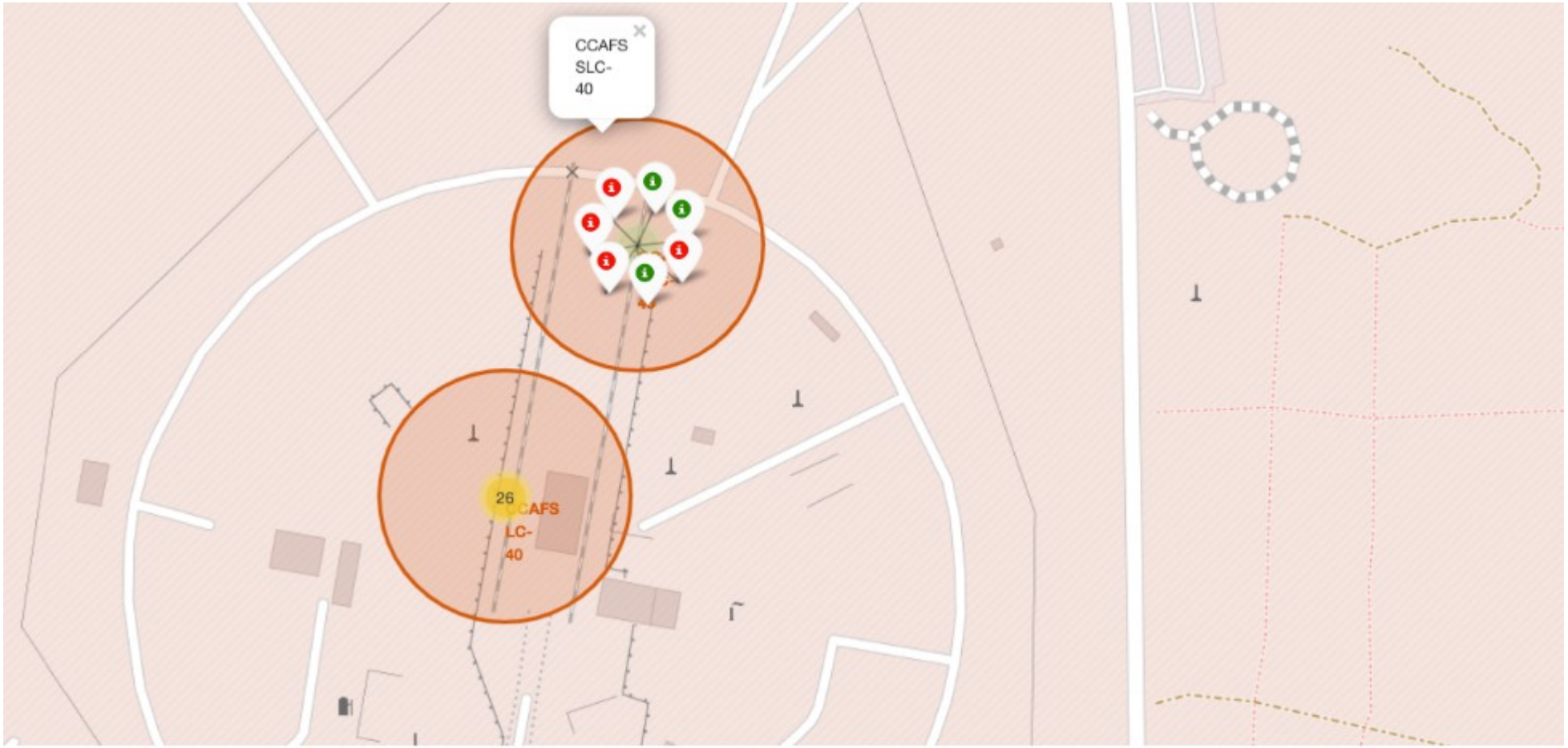

Section 3
Launch Sites
Proximities Analysis

Launch Sites

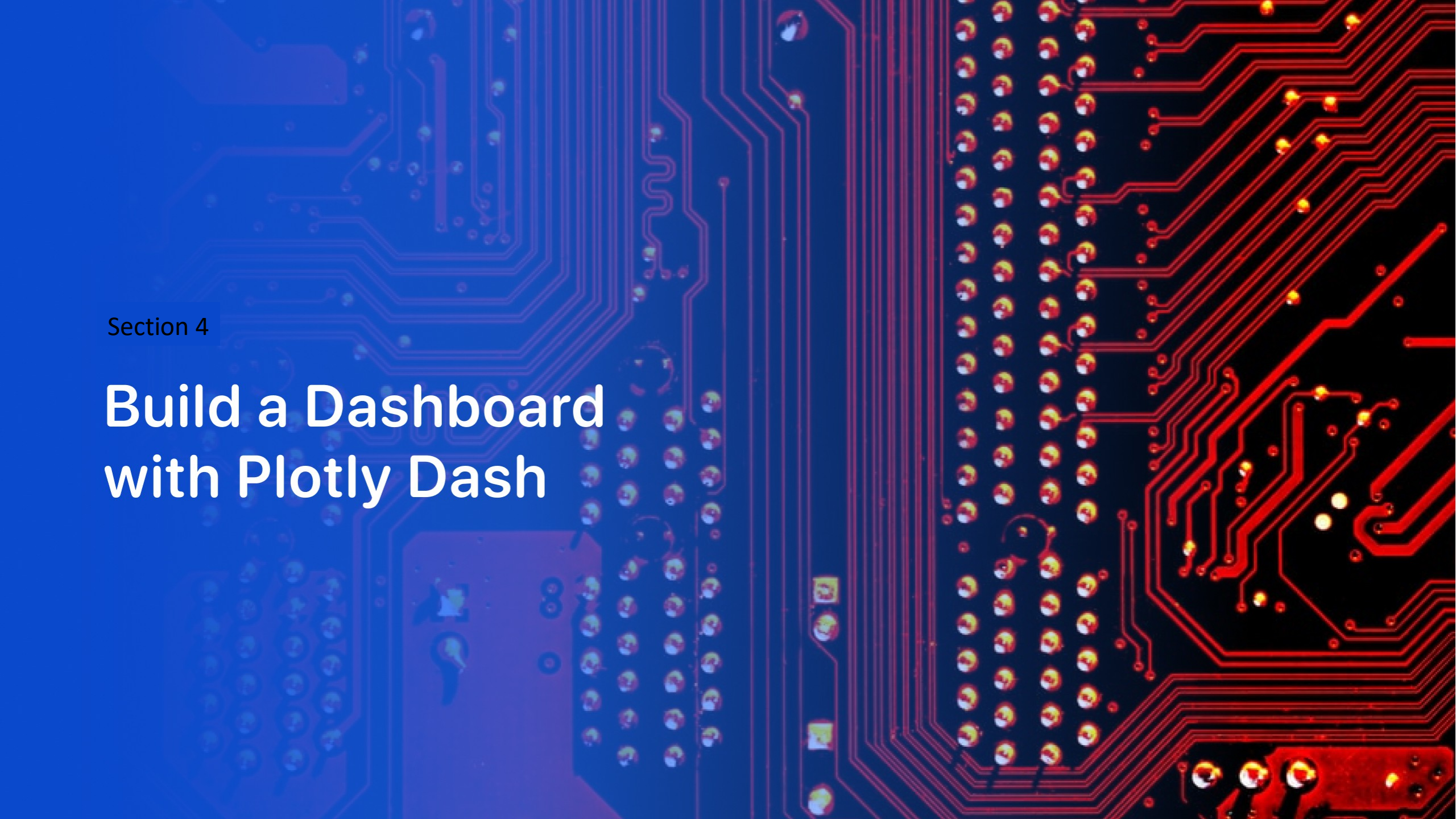




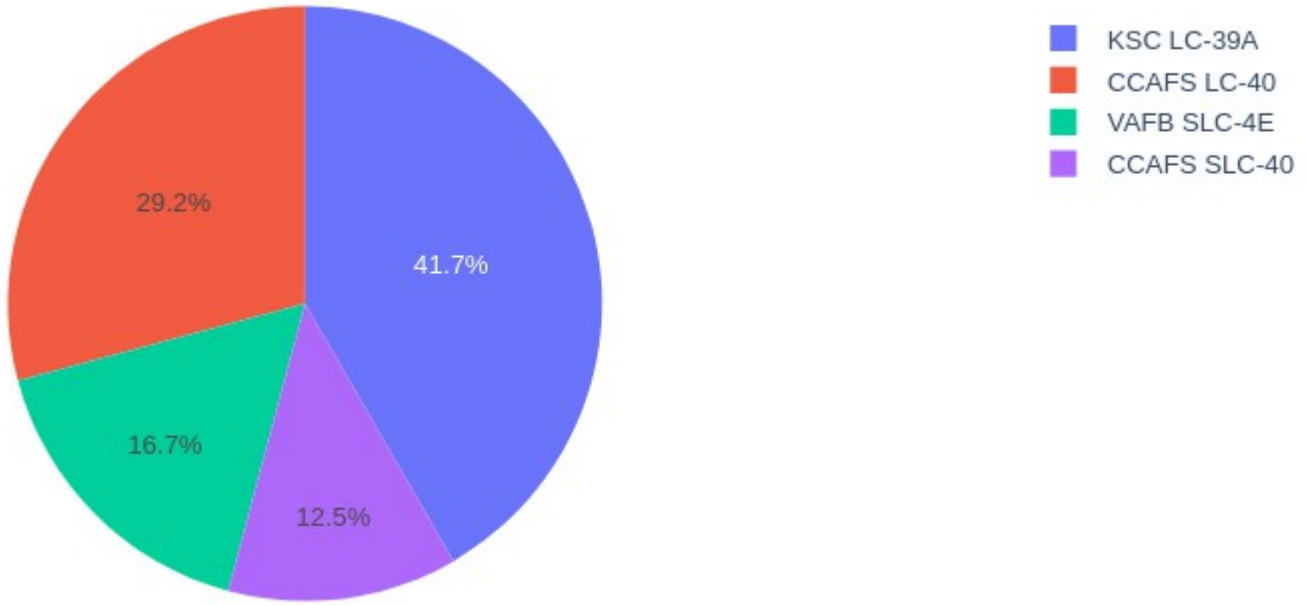
Launch Markers



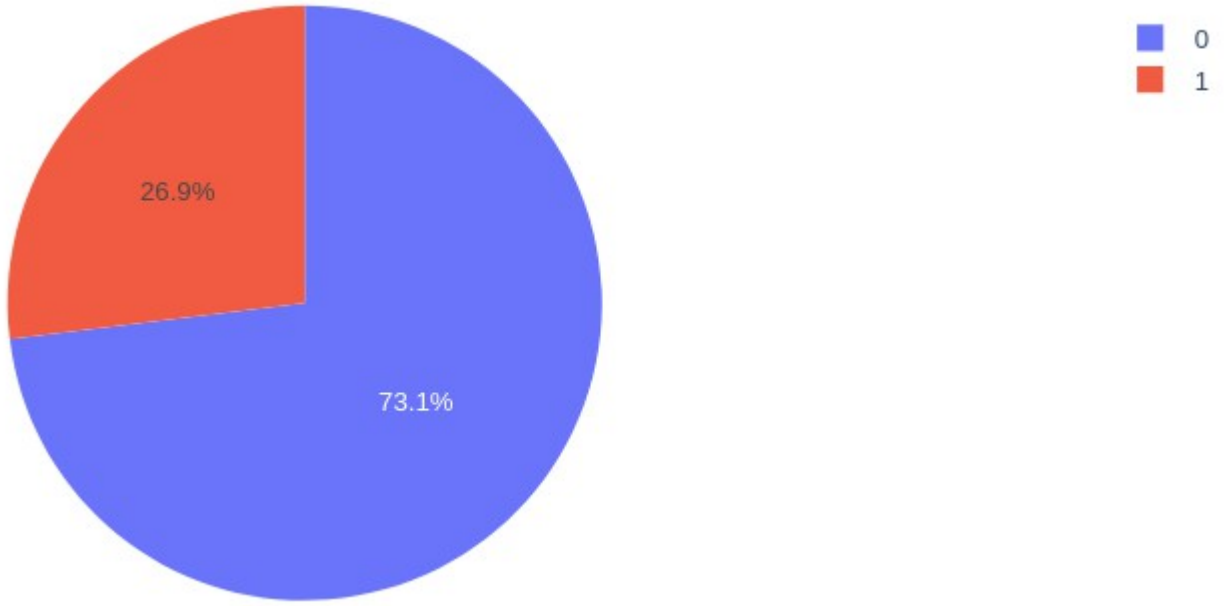
Proximity Based Line Marker


Section 4
Build a Dashboard with Plotly Dash

Visual Success vs. Launch Site



Visual Success(0) vs. Failure(1) Rate



Section 5

Predictive Analysis (Classification)

Classification Accuracy

[12]:
```
parameters ={'C':[0.01,0.1,1],
             'penalty':['l2'],
             'solver':['lbfgs']}
```

[13]:
```
parameters ={"C":[0.01,0.1,1],'penalty':['l2'], 'solver':['lbfgs']}# l1 lasso l2 ridge
lr=LogisticRegression()
logreg_cv = GridSearchCV(estimator=lr, cv=10, param_grid=parameters)
logreg_cv.fit(X_train, Y_train)
```

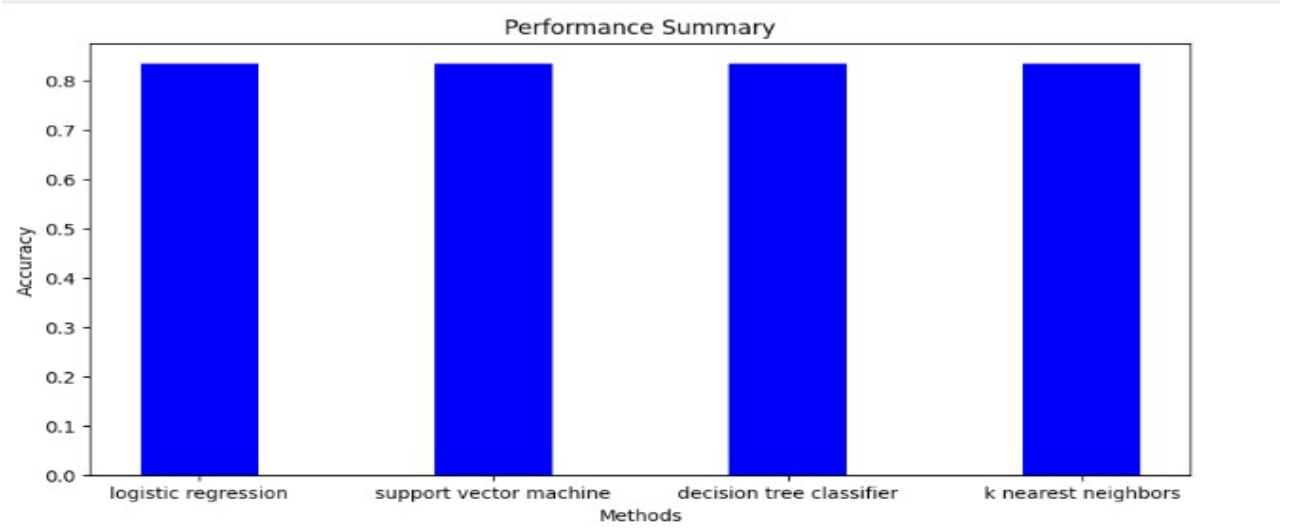
[13]:

GridSearchCV
estimator: LogisticRegression
LogisticRegression

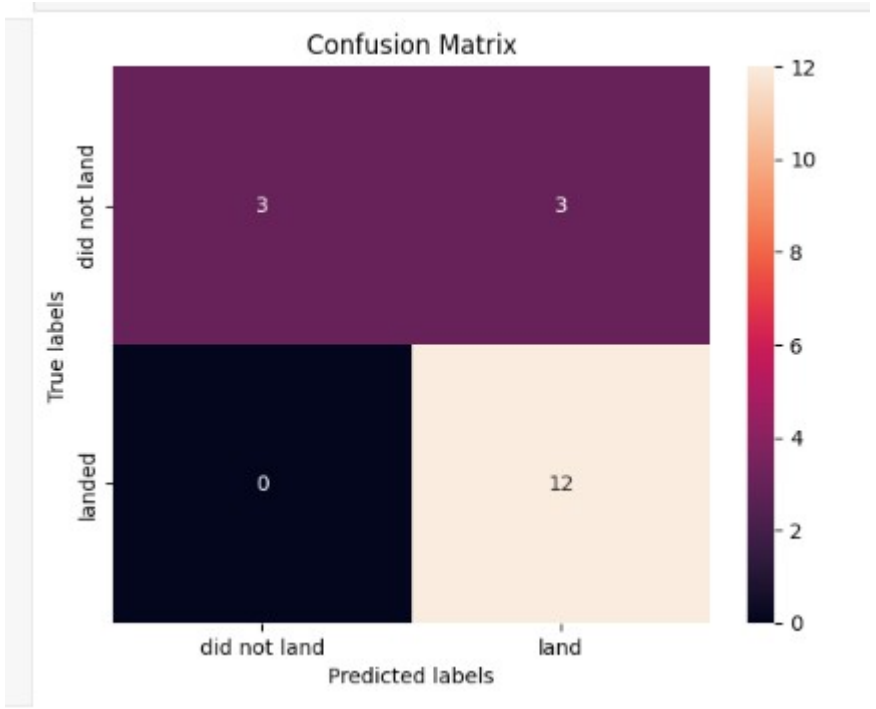
We output the `GridSearchCV` object for logistic regression. We display the best parameters using the data attribute `best_params_` and the accuracy on the validation data using the data attribute `best_score_`.

[14]:
```
print("tuned hpyerparameters :(best parameters) ",logreg_cv.best_params_)
print("accuracy :",logreg_cv.best_score_)
```

```
tuned hpyerparameters :(best parameters)  {'C': 0.01, 'penalty': 'l2', 'solver': 'lbfgs'}
accuracy : 0.8464285714285713
```



Confusion Matrix



Mission Go/No Go?

- SpaceY Has Potential
- Future Landing Success Determination Bases On Further Analysis
- Data CAN be used
- More Data Needed For Greater Accuracy
- More Launches =Greater Success(and Data!)
- Machine Learning Models CAN be used

Appendix

- https://github.com/maroofc/Applied-Data-Science-Capstone-Project
- https://github.com/maroofc/Applied-Data-Science-Capstone-Project/blob/main/spacex_dash_app.py
- https://github.com/maroofc/Applied-Data-Science-Capstone-Project/blob/main/Data%20Collection%20API.ipynb
- https://github.com/maroofc/Applied-Data-Science-Capstone-Project/blob/main/Data%20Wrangling.ipynb
- https://github.com/maroofc/Applied-Data-Science-Capstone-Project/blob/main/EDA%20With%20Data%20Visualisation.ipynb
- https://github.com/maroofc/Applied-Data-Science-Capstone-Project/blob/main/EDA%20With%20SQL%20Analysis.ipynb
- https://github.com/maroofc/Applied-Data-Science-Capstone-Project/blob/main/LaunchSitesWithFolium.ipynb
- https://github.com/maroofc/Applied-Data-Science-Capstone-Project/blob/main/Predictive%20Analysis.ipynb
- https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
- https://api.spacexdata.com/v4

Thank you!