IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Shahid Hussain

15-Jan-2024

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection through API and Web Scrapping
  - Data Wrangling
  - EDA with SQL and Data Visualization
  - Building an interactive map with Folium
  - Building a Dashboard with Plotly dash
  - Predictive Analysis
- Summary of all results
  - EDA results
  - Interactive analysis
  - Predictive analysis

Introduction

- Project background and context

  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.

- Problems you want to find answers

  The project task is to predicting if the first stage of the SpaceX Falcon 9 rocket will land successfully

Section 1

Methodology

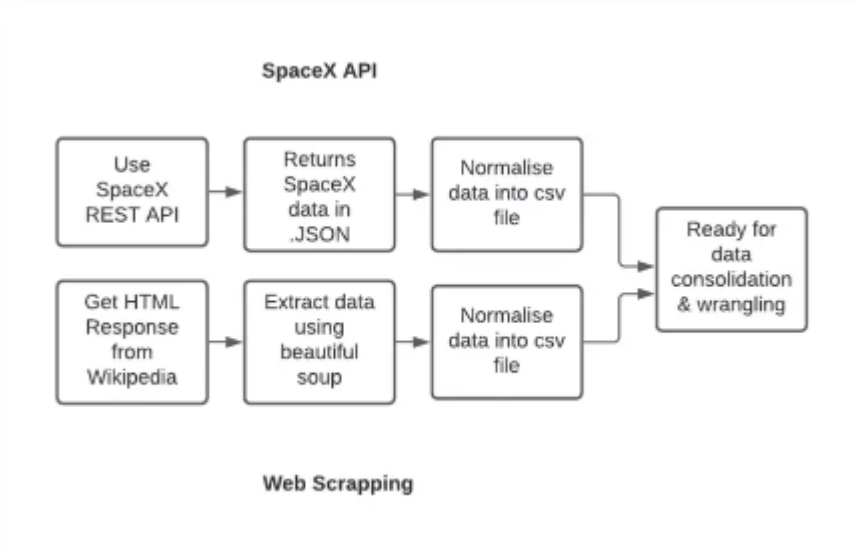
Methodology

Executive Summary

- Data collection methodology:
  - SpaceX Rest API and Web scrapping from Wikipedia
- Perform data wrangling
  - Data was processed using one-hot encoding for catagorical feature and cleaning of null values and irrelevant columns
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - LR, KNN SVM, DT models have been built and evaluated for the best classifier methodology

Data Collection

- Describe how data sets were collected.
  - SpaceX launch data that is gathered from the SpaceX Rest API, gives information about the rocket used, payload delivered, launch specifications, landing specifications and landing outcome
  - Web scrapping was also done to gather the data



Data Collection – SpaceX API

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

- GitHub URL of the completed SpaceX API calls notebook - https://github.com/uniqueshahid/data-science-capstone-project/blob/master/jupyter-labs-spacex-data-collection-api.ipynb

```
spacex_url = "https://api.spacexdata.com/v4/launches/past"
```

```
response = requests.get(spacex_url)
```

```
# Show the head of the dataframe
df.head()
```

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	
0	1	2006-03-24	Falcon 1	20.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	M
1	2	2007-03-21	Falcon 1	NaN	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	M
2	4	2008-09-28	Falcon 1	165.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	M
3	5	2009-07-13	Falcon 1	200.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	M
4	6	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	

Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

- GitHub URL of the completed web scraping notebook - https://github.com/uniqueshahid/data-science-capstone-project/blob/master/jupyter-labs-webscraping.ipynb

```
# use requests.get() method with the provided static_url
# assign the response to a object
data = requests.get(static_url).text
```

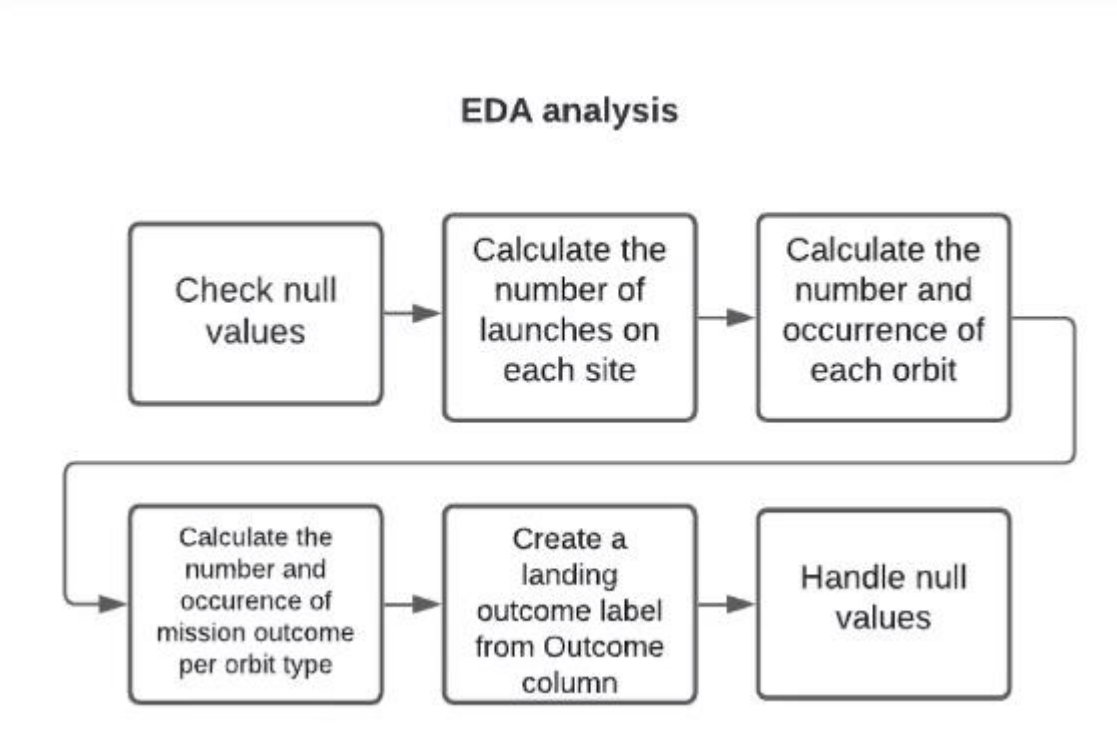
```
# Use BeautifulSoup() to create a BeautifulSoup object from a response text content
soup = BeautifulSoup(data, 'html5lib')
```

```
df= pd.DataFrame({ key:pd.Series(value) for key, value in launch_dict.items() })
df
```

	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n	F9 v1.0B0003.1	Failure	4 June 2010	18:45
1	2	CCAFS	Dragon	0	LEO	NASA	Success	F9 v1.0B0004.1	Failure	8 December 2010	15:43
2	3	CCAFS	Dragon	525 kg	LEO	NASA	Success	F9 v1.0B0005.1	Not attempted\n	22 May 2012	07:44
3	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA	Success\n	F9 v1.0B0006.1	No attempt	8 October 2012	00:35
4	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA	Success\n	F9 v1.0B0007.1	Not attempted\n	1 March 2013	15:10

Data Wrangling

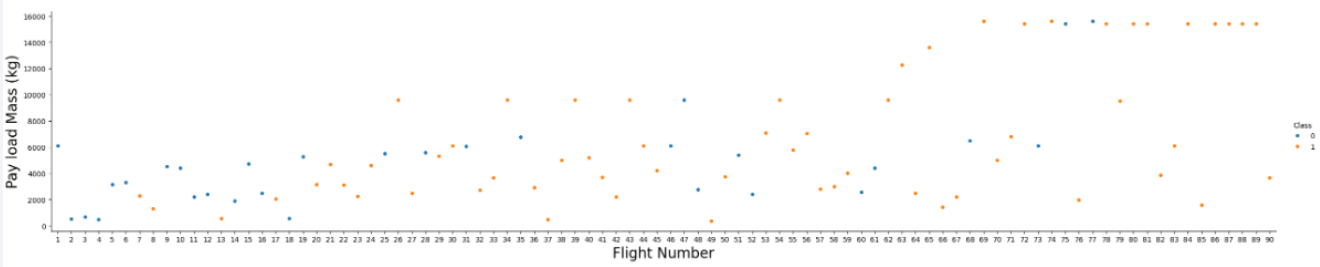
- how data were processed

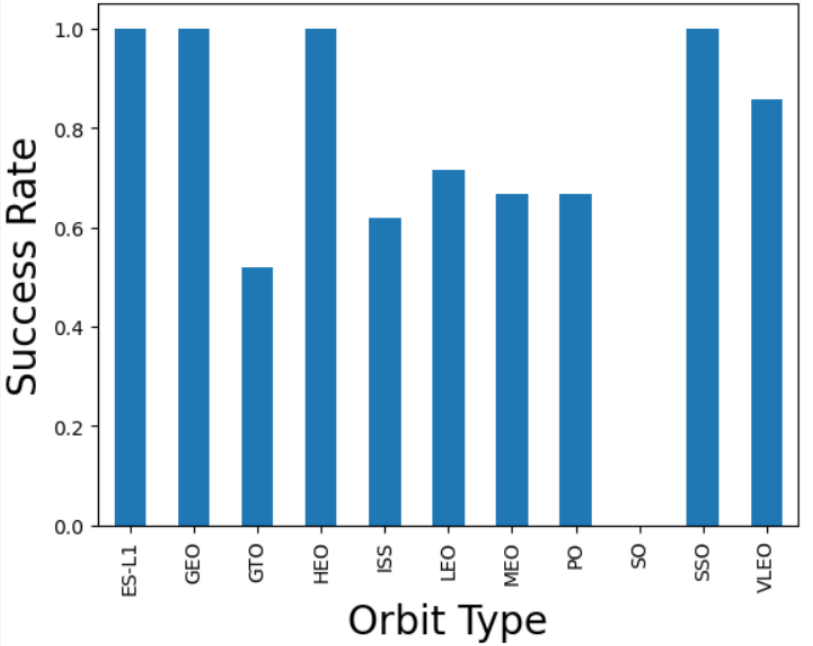


- GitHub URL of your completed data wrangling related notebooks- https://github.com/uniqueshahid/data-science-capstone-project/blob/master/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts





- GitHub URL of your completed EDA with data visualization notebook - https://github.com/uniqueshahid/data-science-capstone-project/blob/master/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed

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

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

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

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

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

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

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

TOTAL_PAYLOAD_MASS
45596

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

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

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

AVG_PAYLOAD_MASS
2534.6666666666665

List the total number of successful and failure mission outcomes

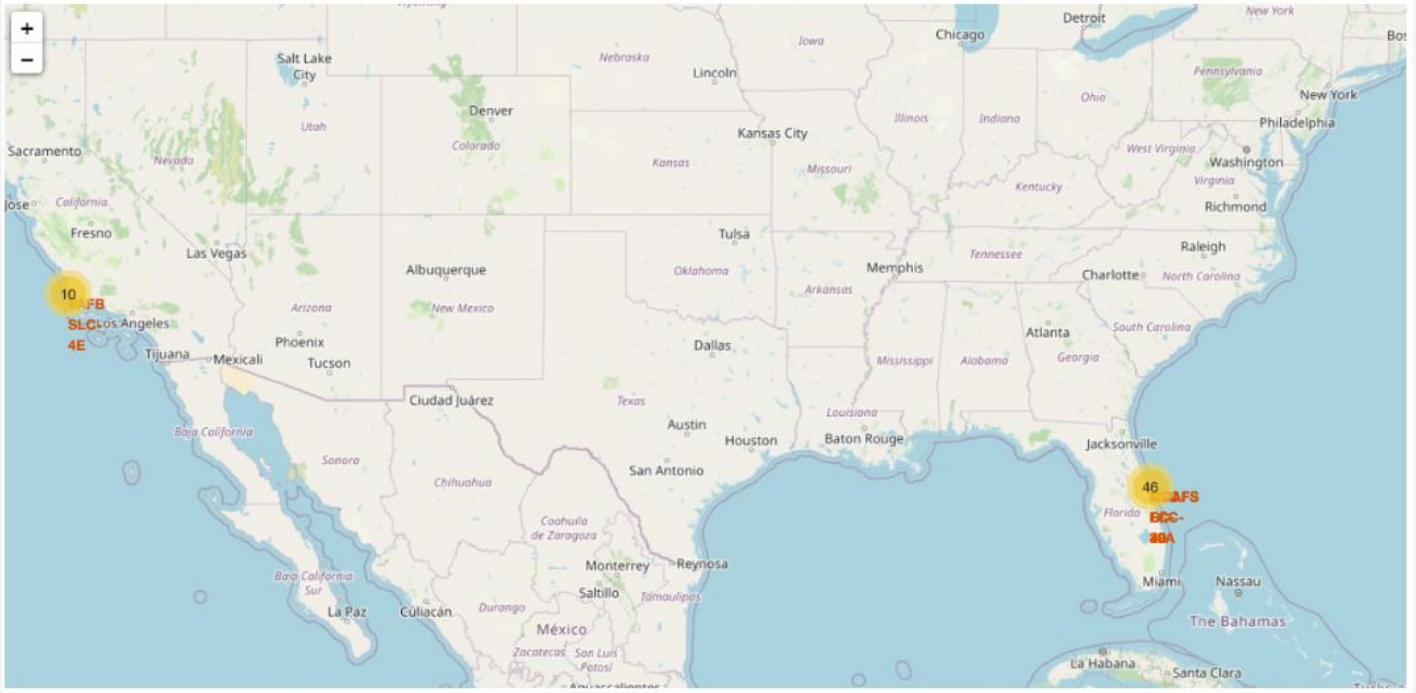
```
%%sql
SELECT MISSION_OUTCOME, COUNT(MISSION_OUTCOME) AS TOTAL_NUMBER
FROM SPACEXTBL
GROUP BY MISSION_OUTCOME;
```

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

Mission_Outcome	TOTAL_NUMBER
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

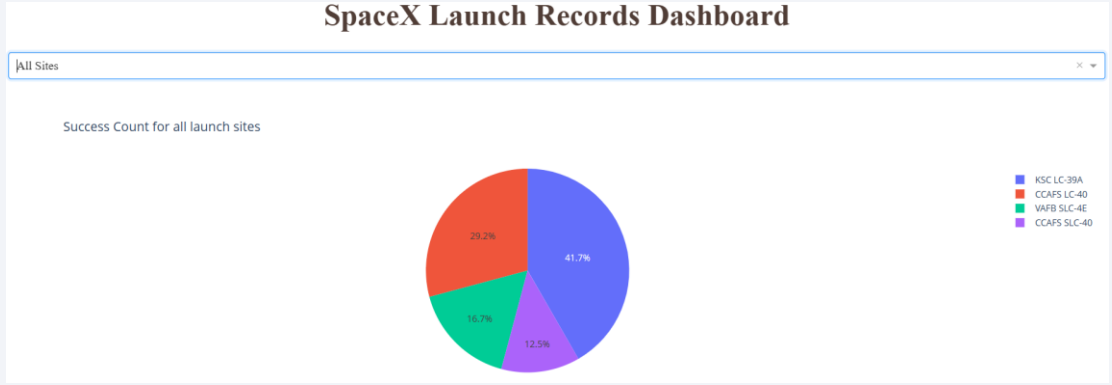
- GitHub URL of your completed EDA with SQL notebook - https://github.com/uniqueshahid/data-science-capstone-project/blob/master/jupyter-labs-eda-sql-coursera_sqllite.ipynb

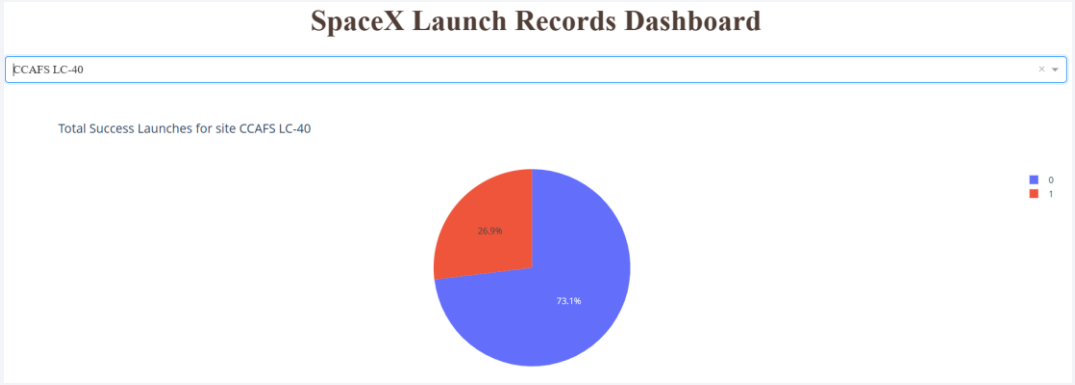
Build an Interactive Map with Folium

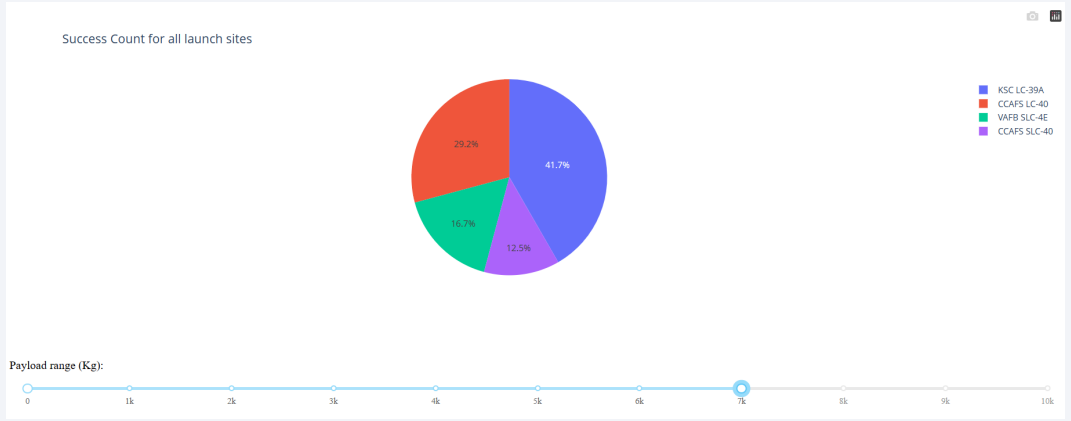


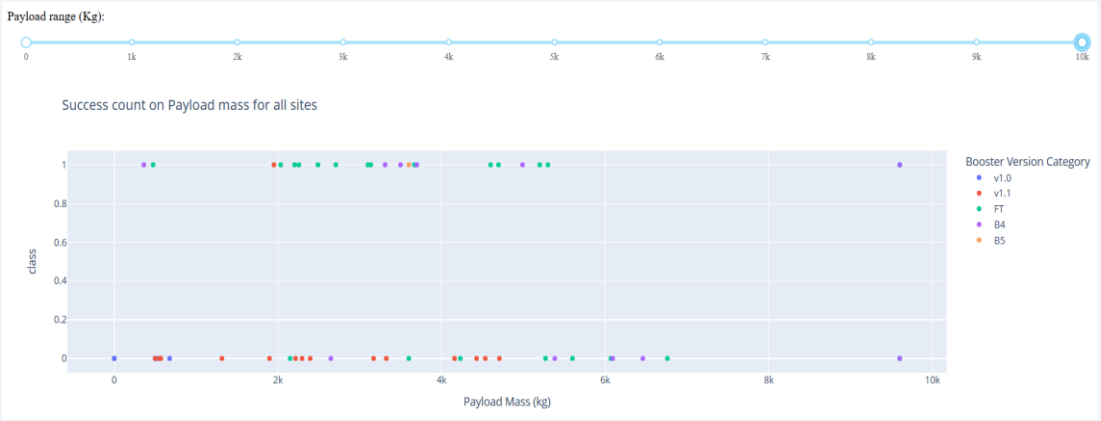
- Map markers has been added to the map with aim to finding an optimal location for building a launch site
- GitHub URL of your completed interactive map with Folium map - https://github.com/uniqueshahid/data-science-capstone-project/blob/master/lab_jupyter_launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash





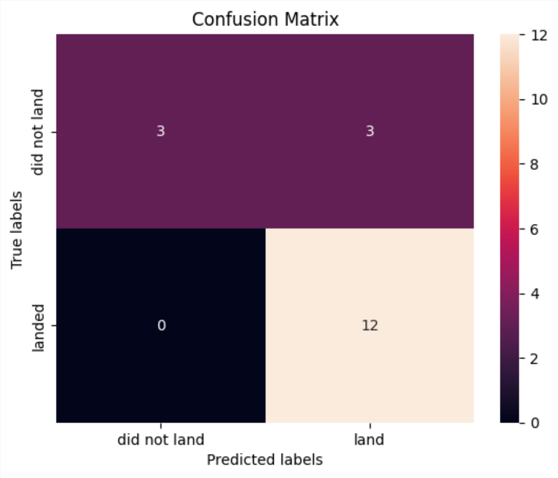


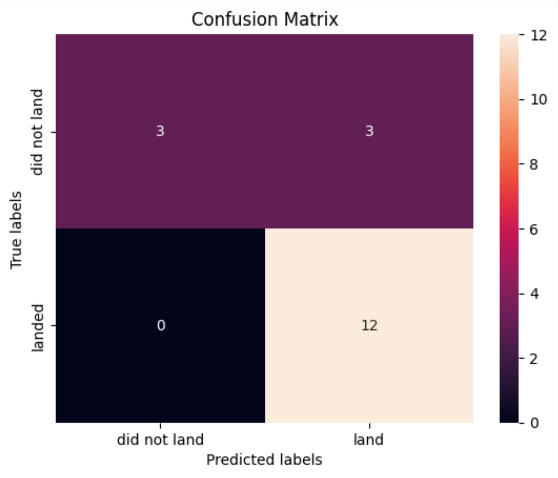


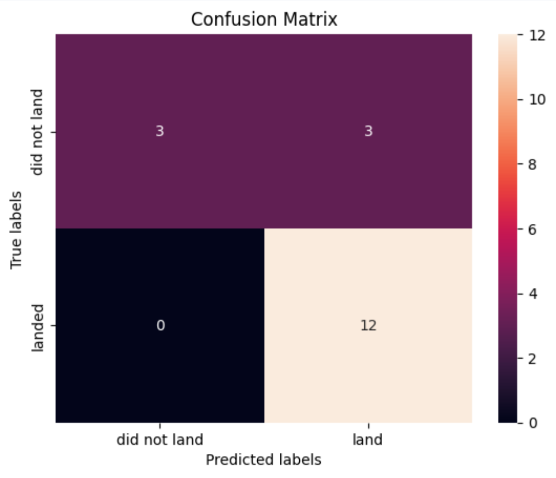
- GitHub URL of your completed Plotly Dash lab - https://github.com/uniqueshahid/data-science-capstone-project/blob/master/spacex_dash_app.py

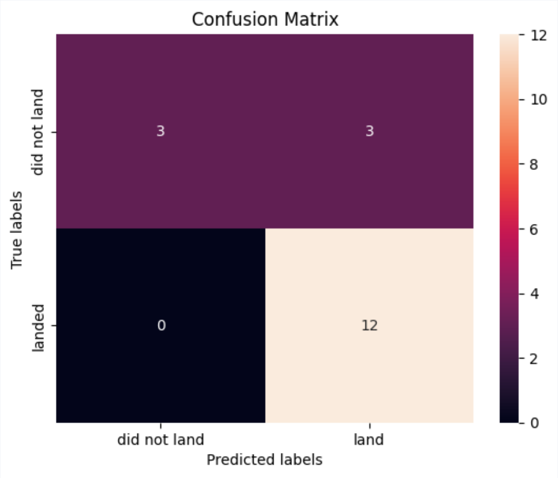
Predictive Analysis (Classification)

- The SVM, KNN and Logistic Regression model achieved the highest accuracy of 83.3%, while the SVM performs the best in terms of Area under Curve at 0.958









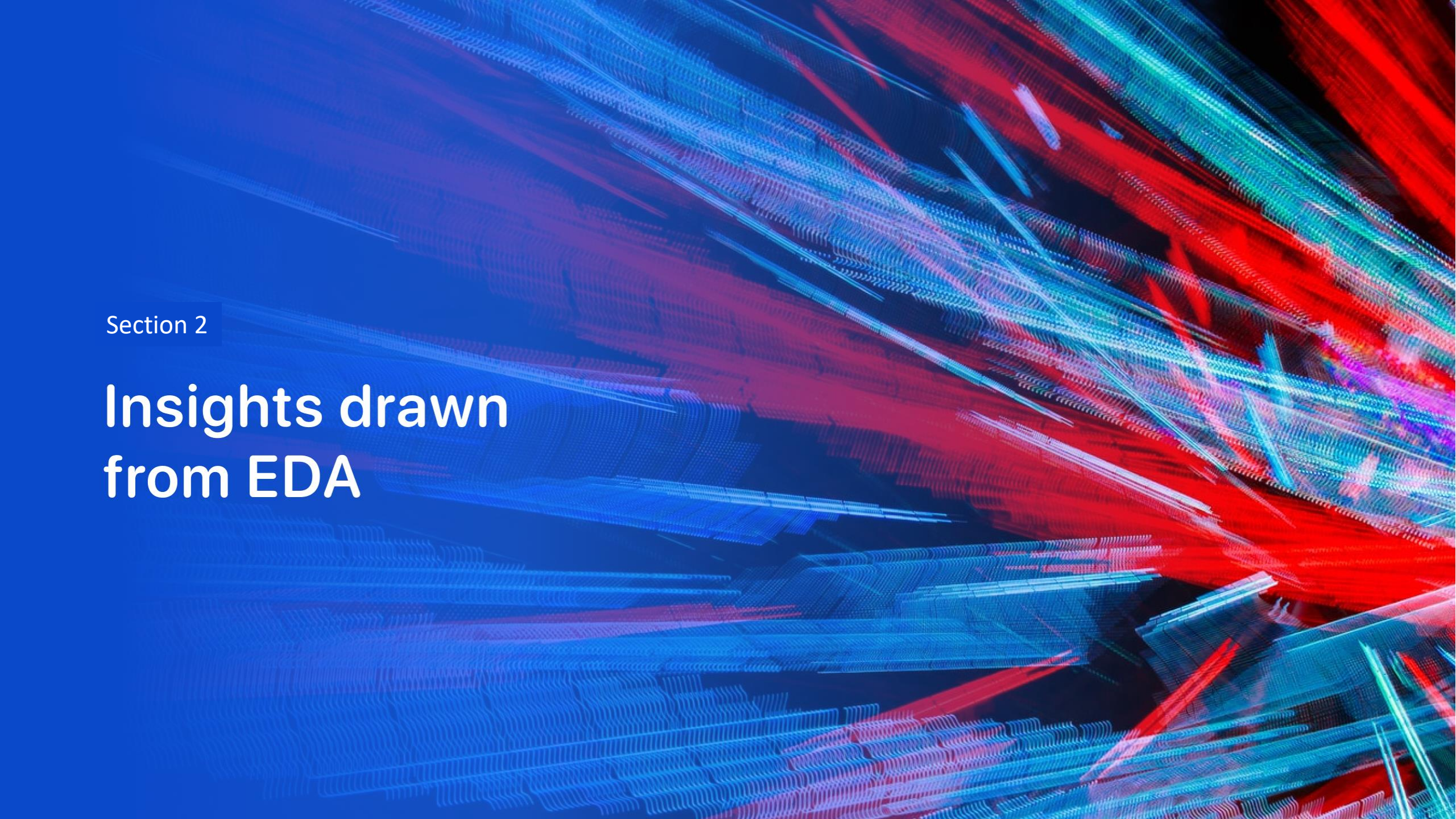
- GitHub URL of your completed predictive analysis lab - https://github.com/uniqueshahid/data-science-capstone-project/blob/master/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

- The SVN, KNN and Logistic regression models are the best in terms of prediction accuracy for this dataset.
- Low weighted payloads perform better than heavier payloads
- The success rate for SpaceX launches is directly proportional time in years they will eventually perfect the launches
- KSC LC 39A had the most successful launches from all the sites.
- Orbit GEO, HEO, SSO, ES L1 has the best Success Rate

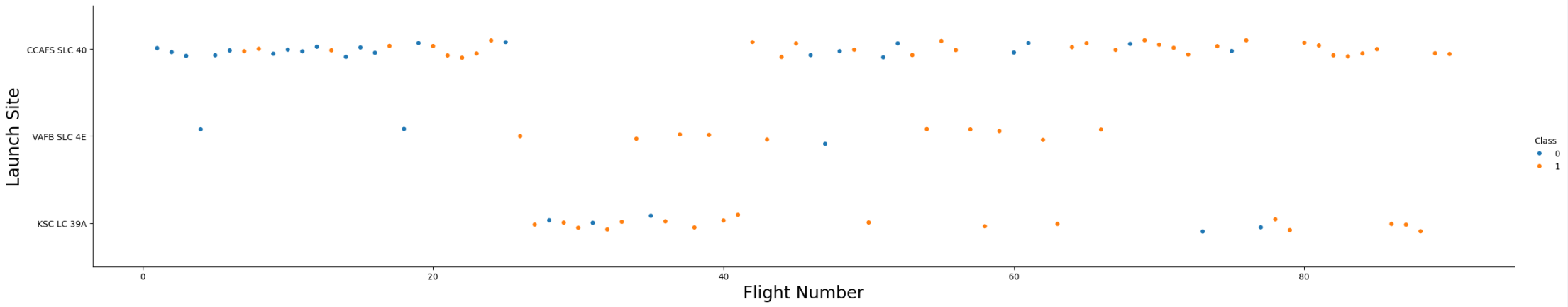
Section 2

Insights drawn from EDA



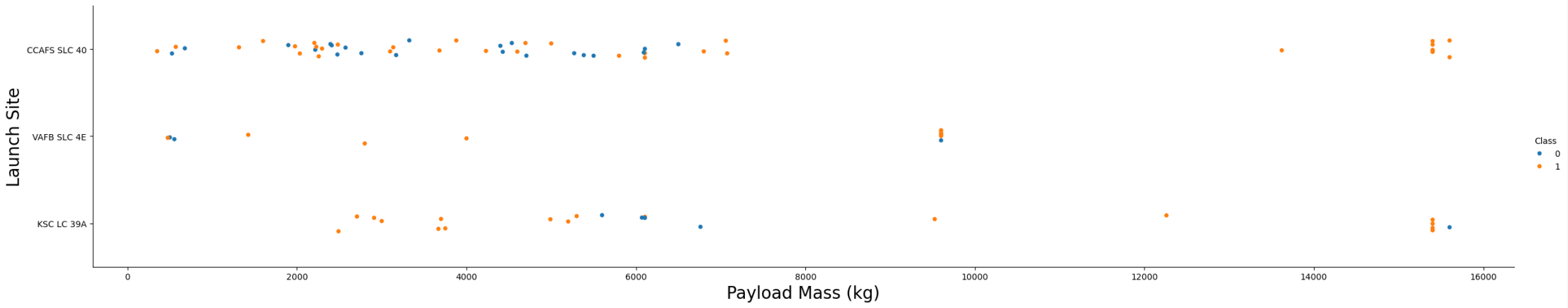
Flight Number vs. Launch Site

- Launches from CCAFS SLC 40 are significantly higher than launches from other sites.



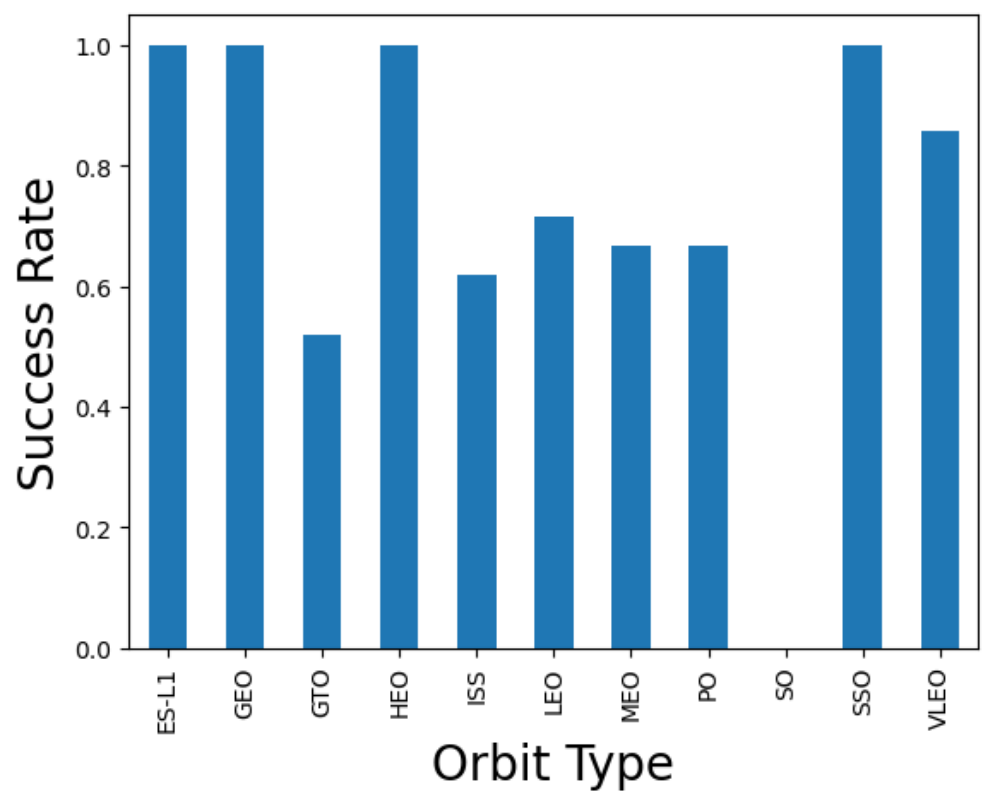
Payload vs. Launch Site

- If we observe Payload Vs. Launch Site scatter point chart, we find that for the VAFB-SLC launchsite there are no rockets launched for heavy payload mass (greater than 10000)



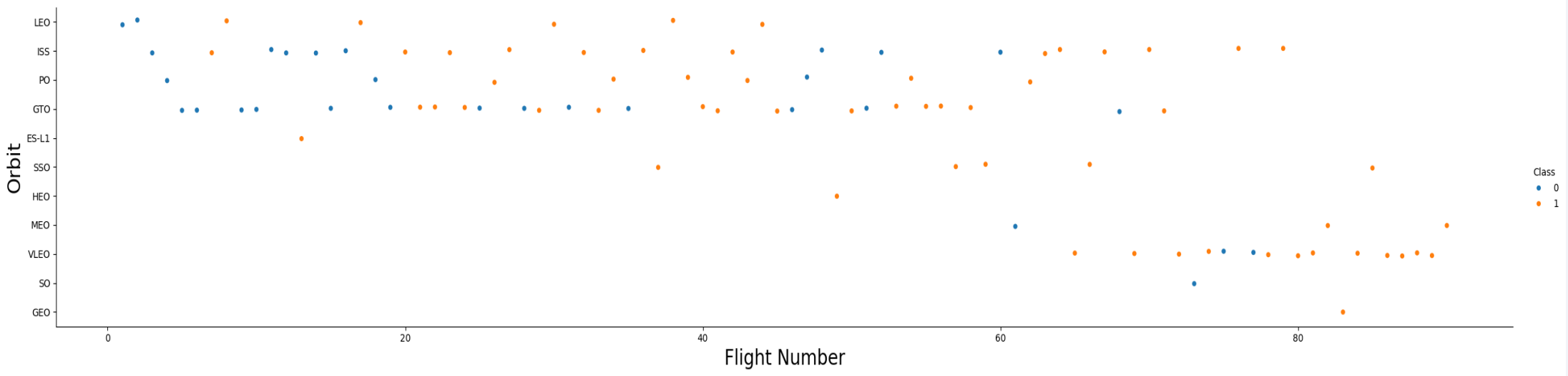
Success Rate vs. Orbit Type

- The Orbit type of ES-L1, GEO, HEO, SSO are among the highest success rate



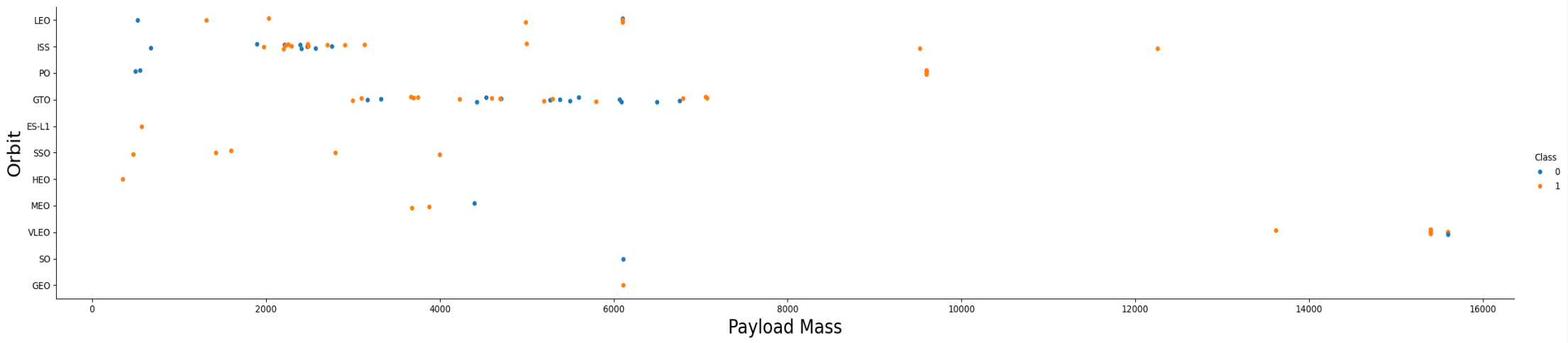
Flight Number vs. Orbit Type

- From Flight number vs Orbit type scatter plot, a trend can be observed of shifting to VLEO launches in recent years



Payload vs. Orbit Type

- There are strong correlation between ISS and Payload at the range around 2000, as well as between GTO and the range of 4000-8000.



All Launch Site Names

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

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

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

Launch Site Names Begin with 'CCA'

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

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

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_MASS FROM SPACEXTBL WHERE Customer = 'NASA (CRS)';
```

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

TOTAL_PAYLOAD_MASS
45596

Average Payload Mass by F9 v1.1

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

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

AVG_PAYLOAD_MASS
2534.6666666666665

First Successful Ground Landing Date

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

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

MIN(Date)
2015-12-22

Succesful Drone Ship Landing with Payload between 4000 and 6000

```
%%sql
SELECT BOOSTER_VERSION
FROM SPACEXTBL
WHERE LANDING_OUTCOME = 'Success (drone ship)'
    AND 4000 < PAYLOAD_MASS__KG_ < 6000;
```

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

Booster_Version
F9 FT B1021.1
F9 FT B1022
F9 FT B1023.1
F9 FT B1026
F9 FT B1029.1
F9 FT B1021.2
F9 FT B1029.2
F9 FT B1036.1
F9 FT B1038.1
F9 B4 B1041.1
F9 FT B1031.2
F9 B4 B1042.1
F9 B4 B1045.1
F9 B5 B1046.1

Total Number of Successful and Failure Mission Outcomes

```
%%sql
SELECT MISSION_OUTCOME, COUNT(MISSION_OUTCOME) AS TOTAL_NUMBER
FROM SPACEXTBL
GROUP BY MISSION_OUTCOME;
```

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

Mission_Outcome	TOTAL_NUMBER
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

```
%%sql
SELECT DISTINCT BOOSTER_VERSION
FROM SPACEXTBL
WHERE PAYLOAD_MASS__KG_ = (
    SELECT MAX(PAYLOAD_MASS__KG_)
    FROM SPACEXTBL);
```

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

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

```
%%sql
SELECT substr(DATE,6,2) as Month, LANDING_OUTCOME, BOOSTER_VERSION, LAUNCH_SITE
FROM SPACEXTBL
WHERE Landing_Outcome = 'Failure (drone ship)' AND substr(DATE,0,5) = '2015';
```

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

Month	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

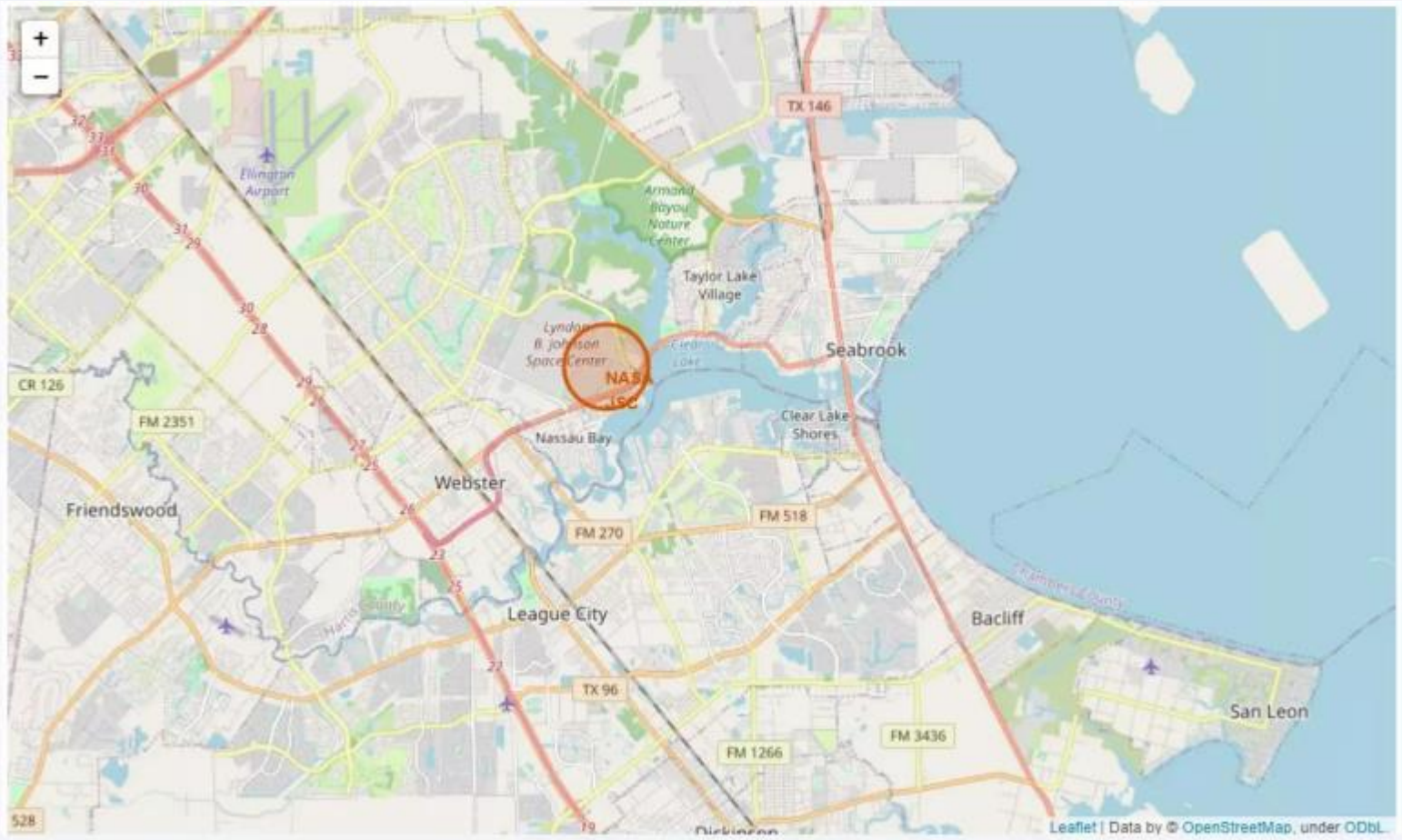
```
%%sql
SELECT LANDING_OUTCOME, COUNT(LANDING_OUTCOME) AS TOTAL_NUMBER
FROM SPACEXTBL
WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY LANDING_OUTCOME
ORDER BY TOTAL_NUMBER DESC
```

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

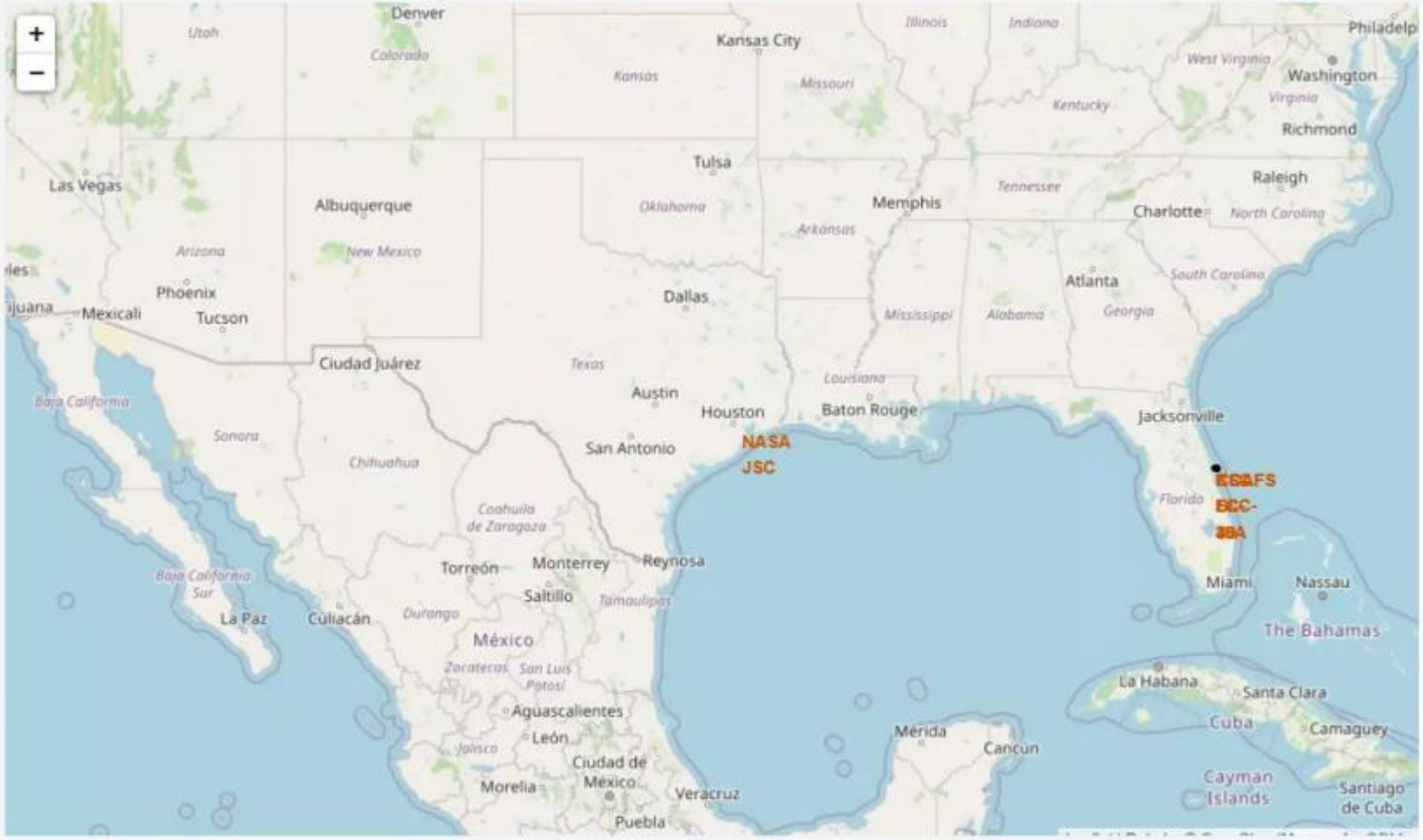
Landing_Outcome	TOTAL_NUMBER
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1


Section 3
Launch Sites
Proximities Analysis

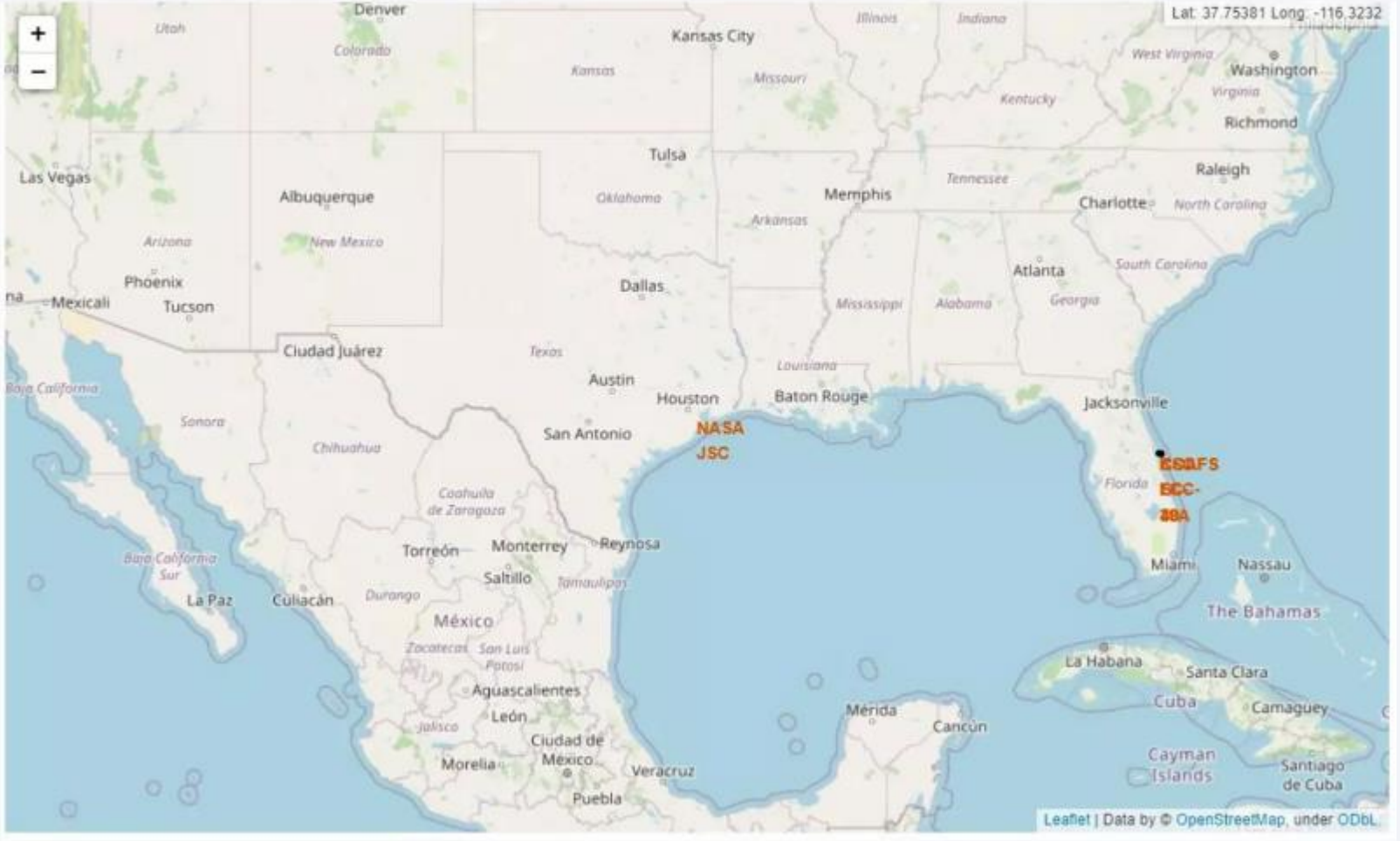
Launch Sites Marked on a Map

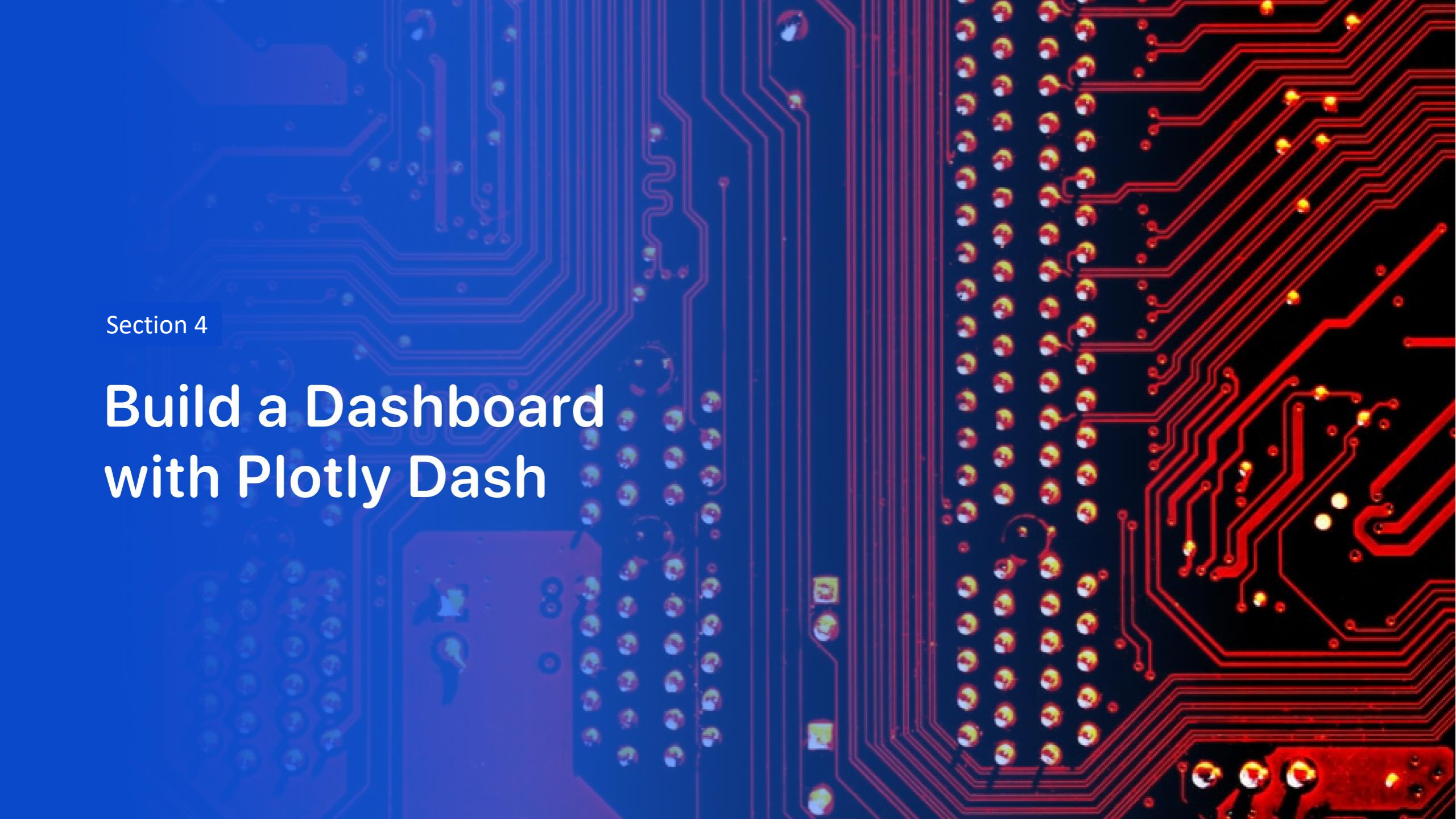


Success / Failed launches marked on Map

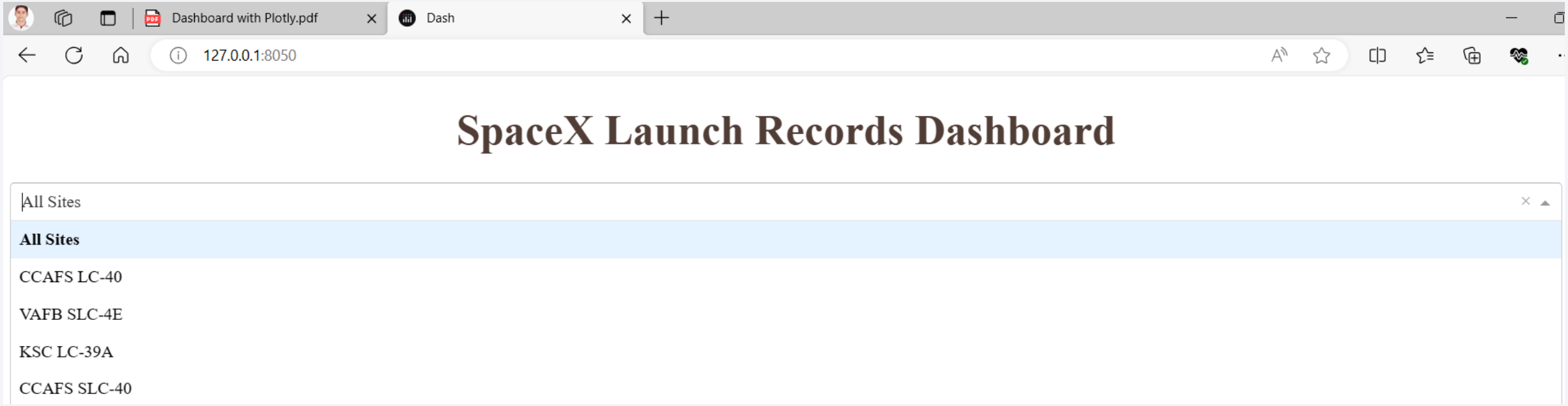


Distances between a launch site to its proximities

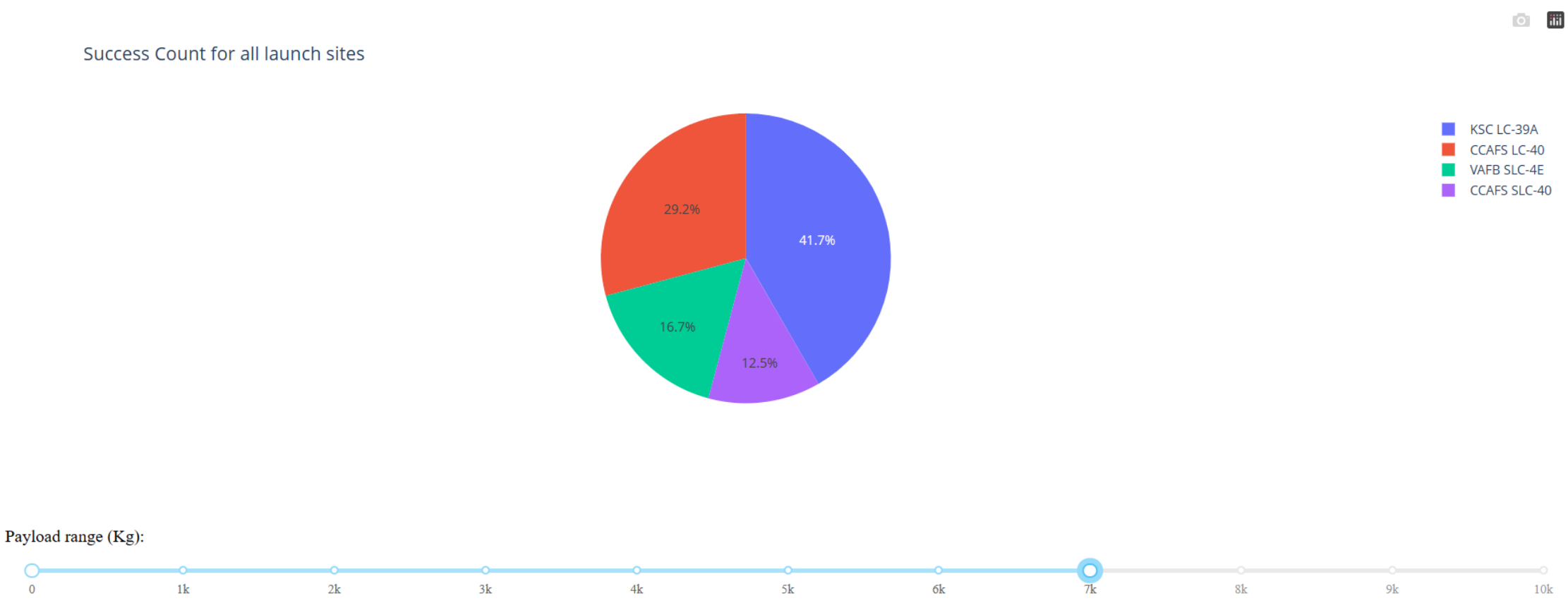



Section 4
Build a Dashboard with Plotly Dash

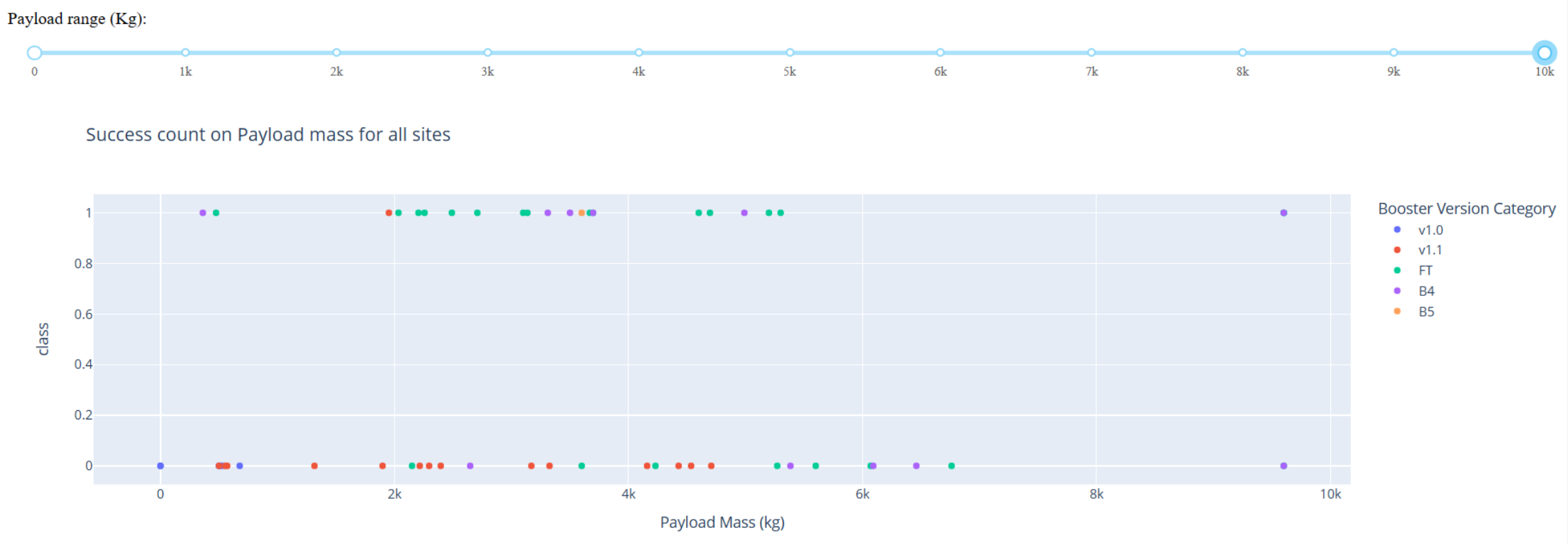
SpaceX Launch Records Dashboard



Total Success Launches for all sites



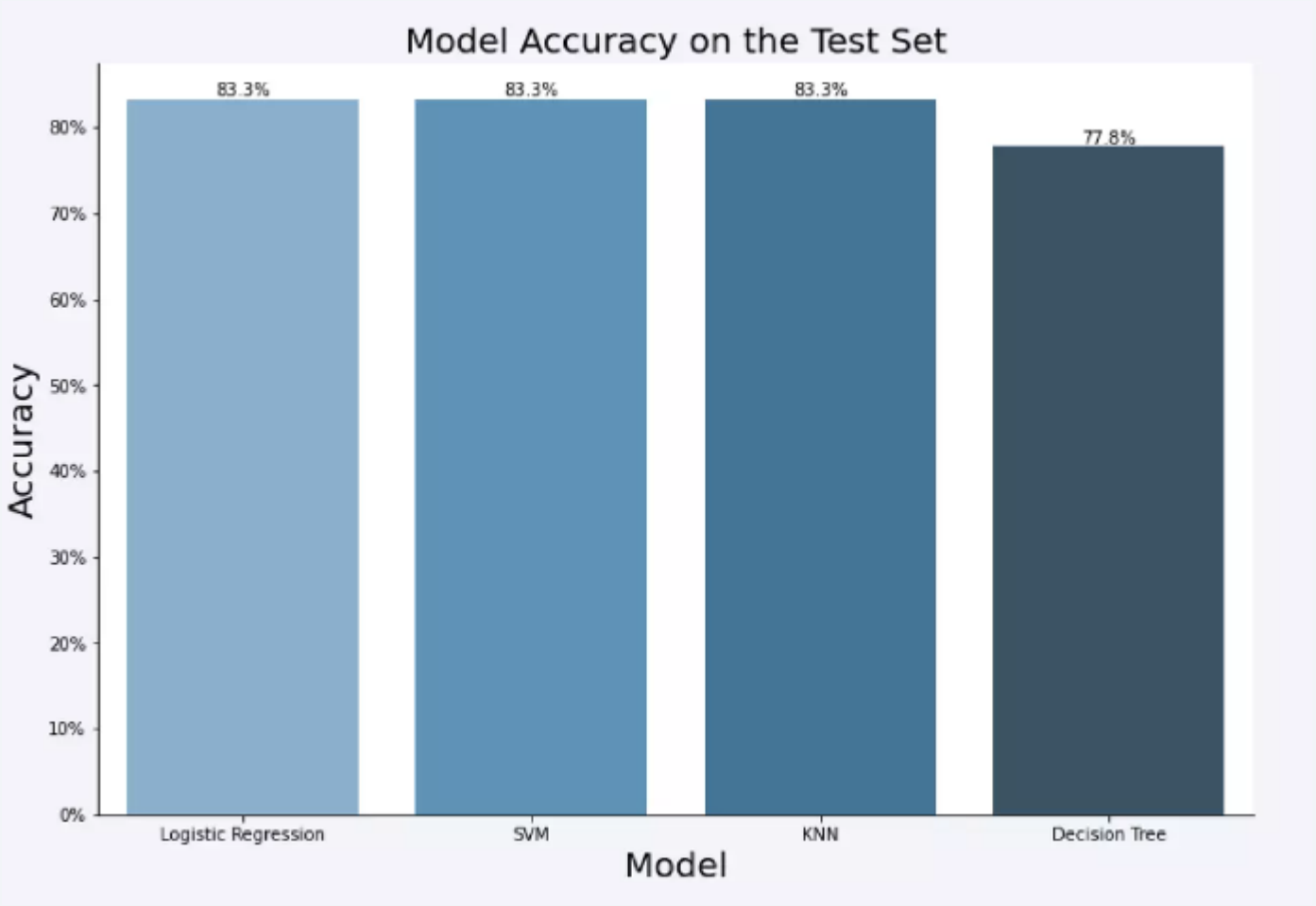
Payload vs Launch Outcome



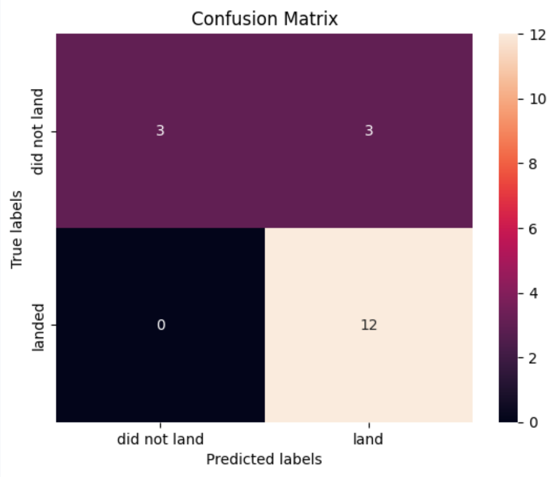
Section 5

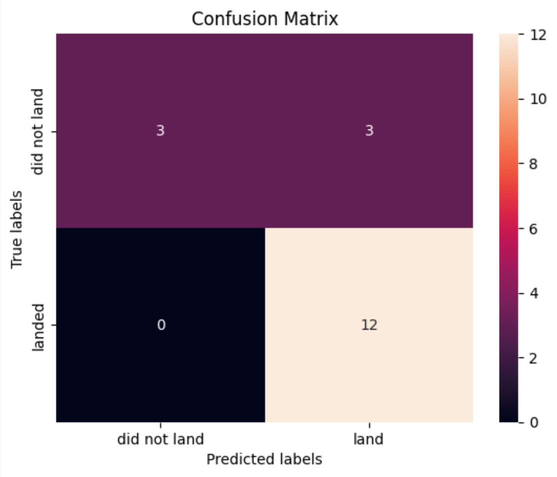
Predictive Analysis (Classification)

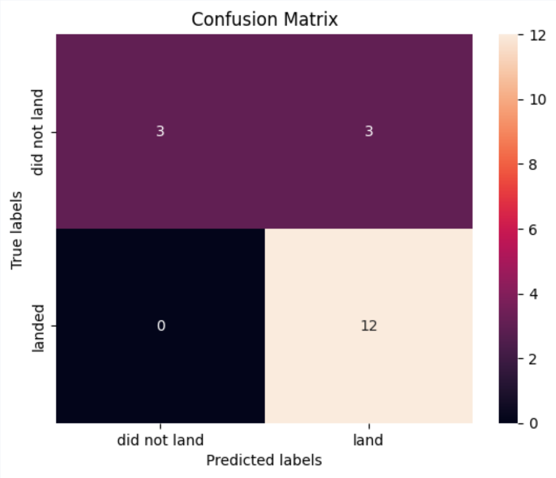
Classification Accuracy

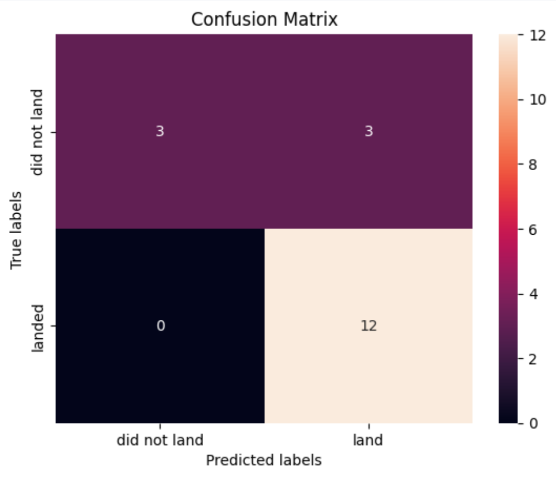


Confusion Matrix









Conclusions

- The SVM, KNN, and Logistic Regression models are the best in terms of prediction accuracy for this dataset.
- Low weighted payloads perform better than the heavier payloads.
- The success rates for SpaceX launches is directly proportional time in years they will eventually perfect the launches.
- KSC LC 39A had the most successful launches from all the sites.
- Orbit GEO,HEO,SSO,ES L1 has the best Success Rate

Thank you!