Winning Space Race with Data Science

Laith Al Habahbeh
07-09-2022

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection
  - Data Wrangling
  - EDA with data visualization
  - EDA with SQL
  - Building an interactive map with Folium
  - Predictive analysis (Classification)
- Summary of all results
  - EDA results
  - Interactive analytics
  - Predictive analysis

Introduction

- Project background and context
  - SpaceX advertising their new Falcon 9 with the feature of “first stage reuse” which is leading to almost 60% less cost comparing to other providers; therefore, we are in this report checking the reliability of the new feature out of the historical data we have in their website and confirming if the advertised cost will be applicable or not, which should be used to bid against SpaceX prices.
- Problems you want to find answers
  - Determine if Flacon 9 first stage will land, so we can determine the cost of a launch.

Section 1

Methodology

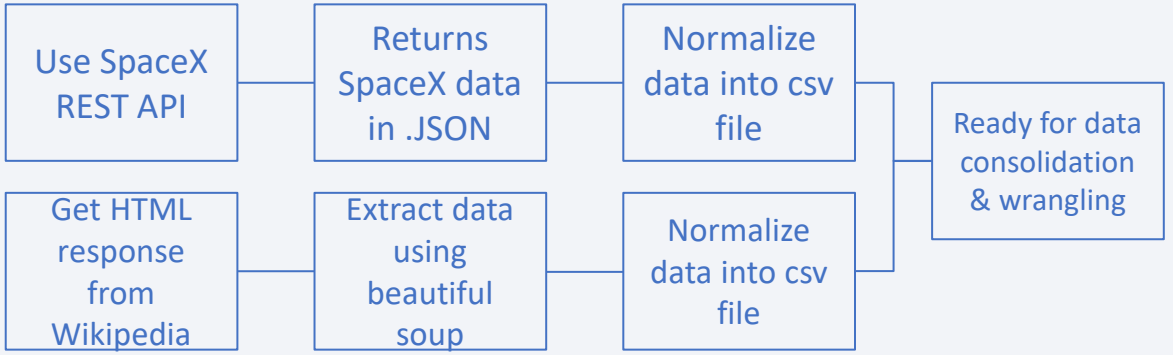
Methodology

Executive Summary

- Data collection methodology:
  - SpaceX launch data has been collected through Request & Parse using GET request.
  - Data frame has been filtered to include Falcon 9 launches only.
- Perform data wrangling
  - Missing values has been checked in the data set
    - PayLoadMss missing data replaced with the mean value.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

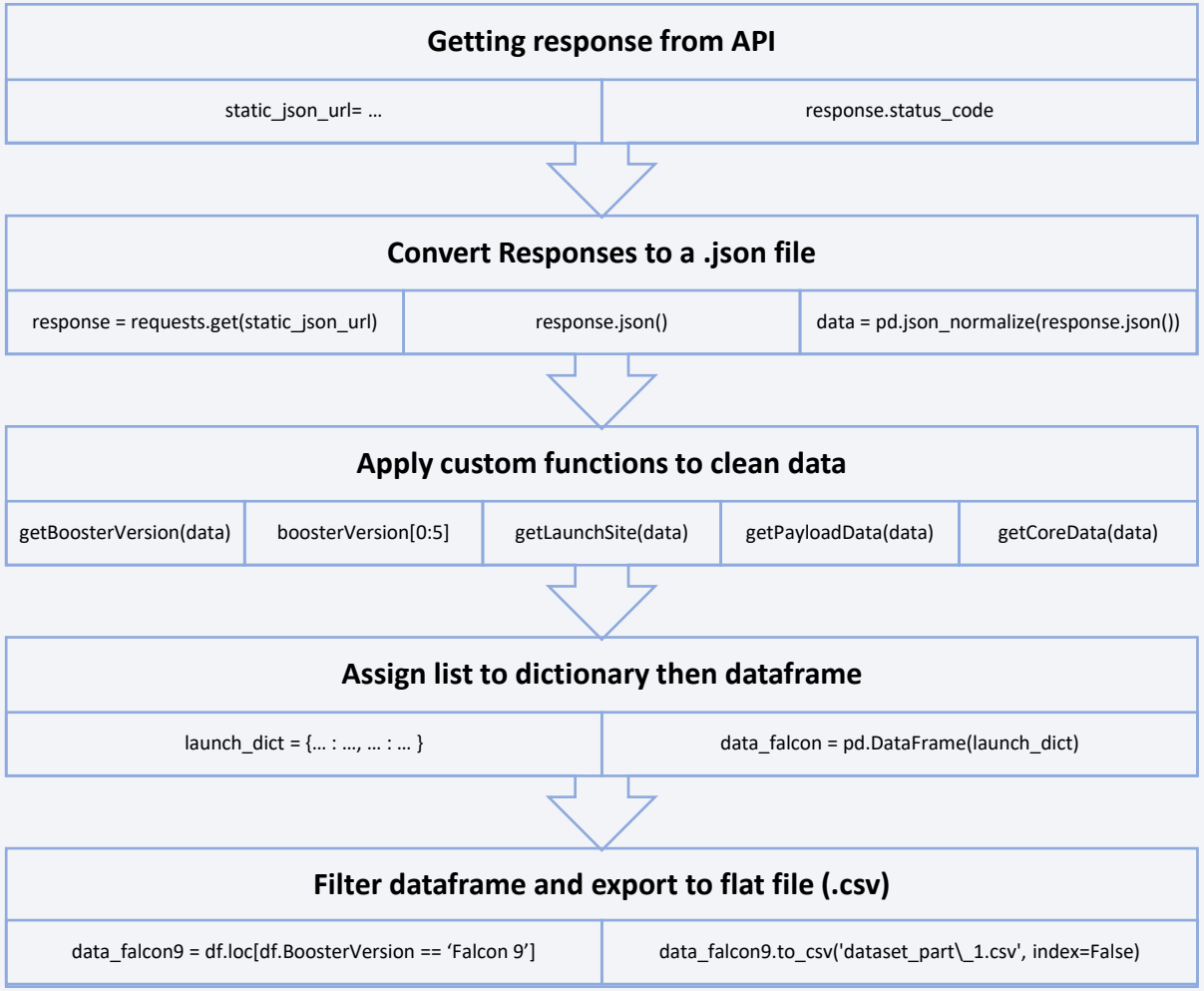
- The following datasets was collected:
  - SpaceX launch data that is gathered from the SpaceX REST API.
  - This API will give us data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome.
  - The Space REST API endpoints, or URL, starts with api.spacexdata.com/v4/.
  - Another popular data source for obtaining Falcon 9 Launch data is web scraping Wikipedia using BeautifulSoup.



Data Collection – SpaceX API

- Data collection with SpaceX REST calls

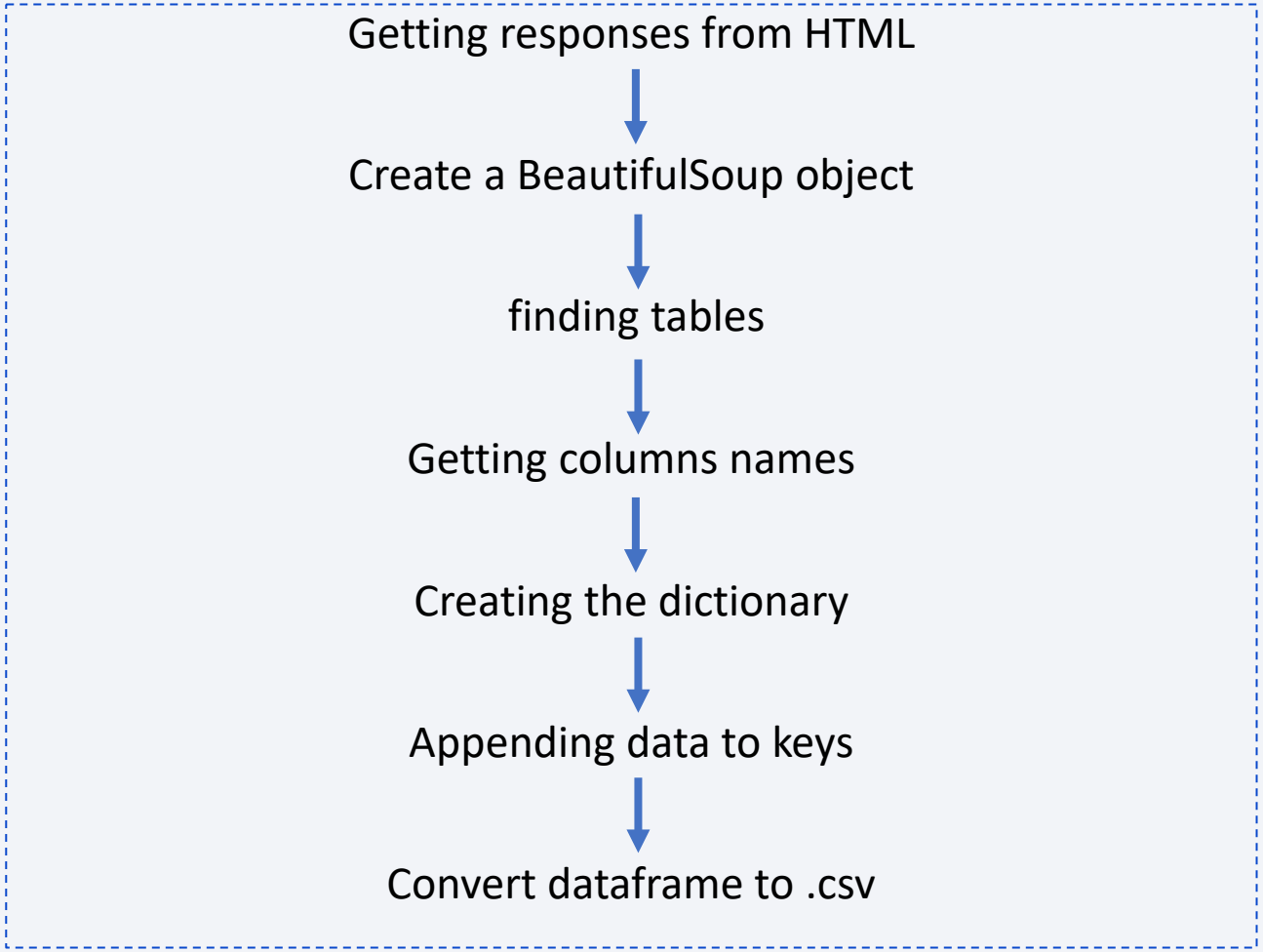
- Notebook Link: https://github.com/Lhabahbeh/DS_Capstone_Project/blob/master/Collecting%20the%20data.ipynb



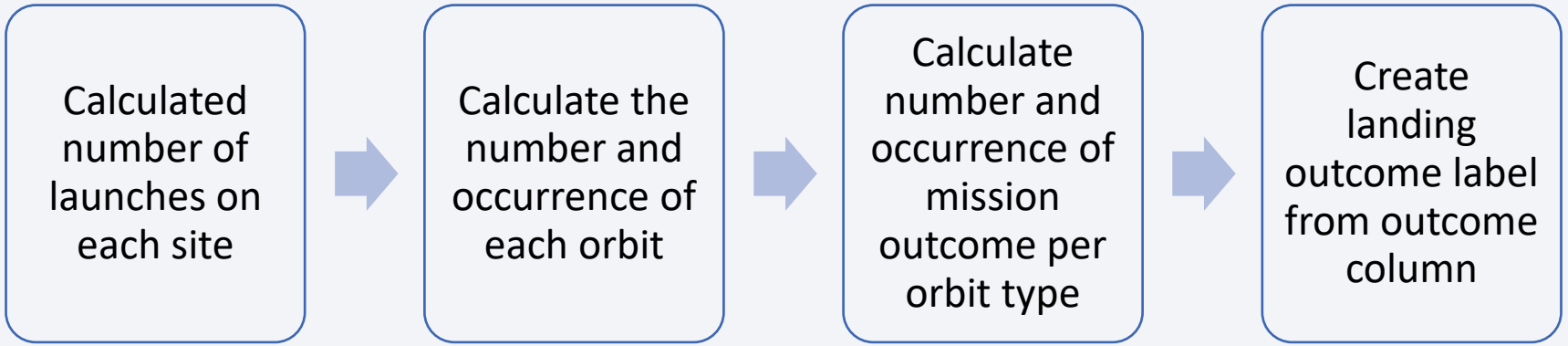
Data Collection - Scraping

- Web scraping from Wikipedia

- Notebook Link: https://github.com/Lhabahbeh/DS_Capstone_Project/blob/master/Data%20Collection%20and%20Web%20Scraping.ipynb

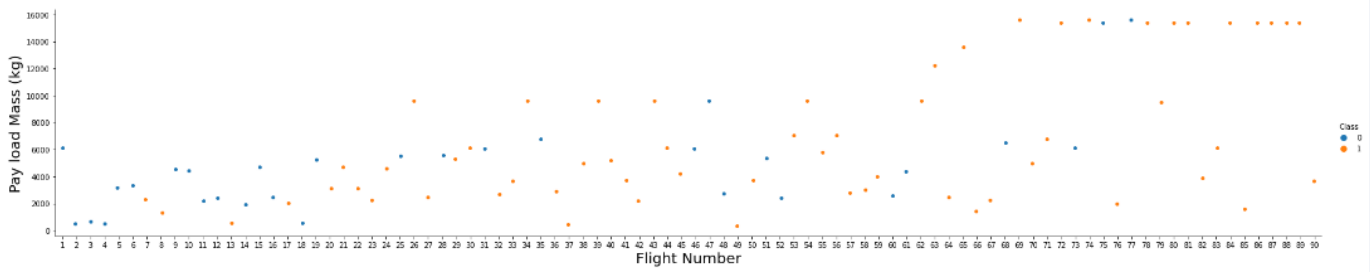


Data Wrangling

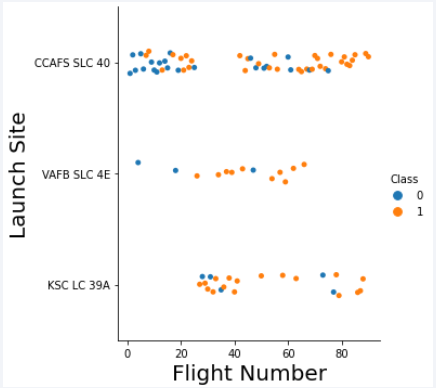


- Notebook Link: https://github.com/Lhabahbeh/DS_Capstone_Project/blob/master/Data%20Wrangling.ipynb

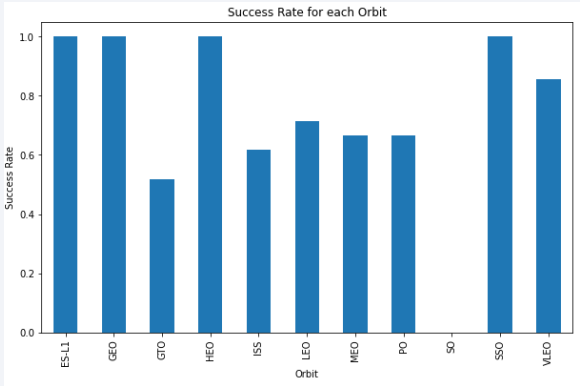
EDA with Data Visualization



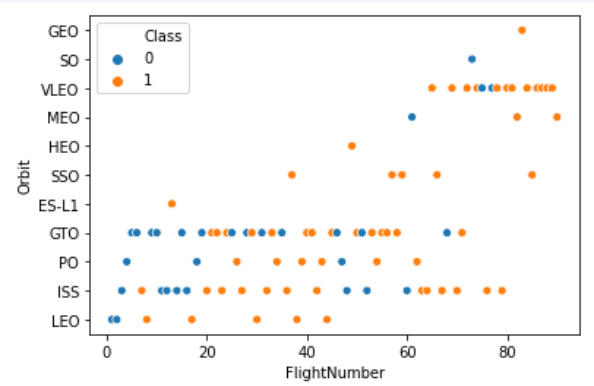
Cat Plot – Relationship of Flight Number and Pay Load Mass with Launch Outcome



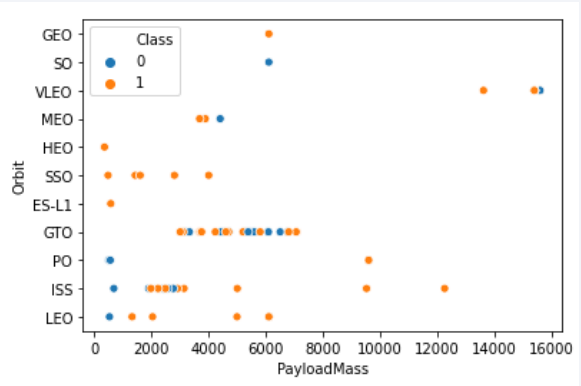
Cat Plot – Relationship Between Flight Number and Launch Site



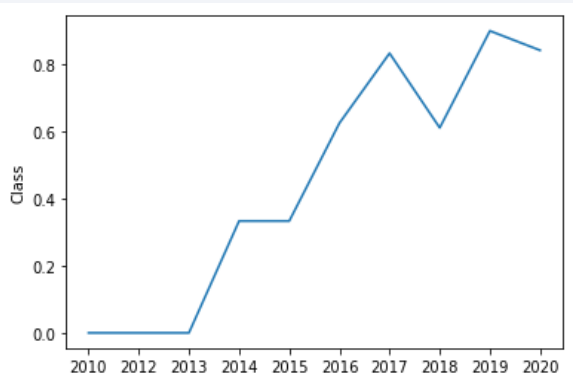
Bar Chart – Success Rate for Each Orbit



Scatter Plot – Relationship between Flight Number and Orbit Type



Scatter Plot – Relationship between Pay Load and Orbit Type



Line Plot – Launch Success Yearly Trend

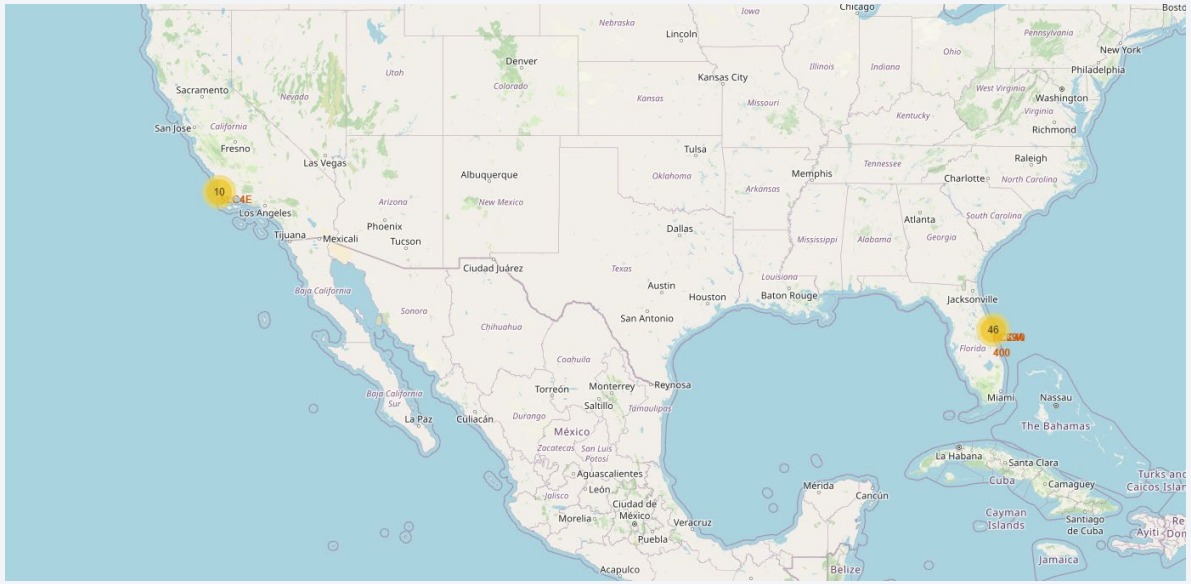
- Notebook Link: https://github.com/Lhabahbeh/DS_Capstone_Project/blob/master/EDA%20with%20Date%20Visualization.ipynb

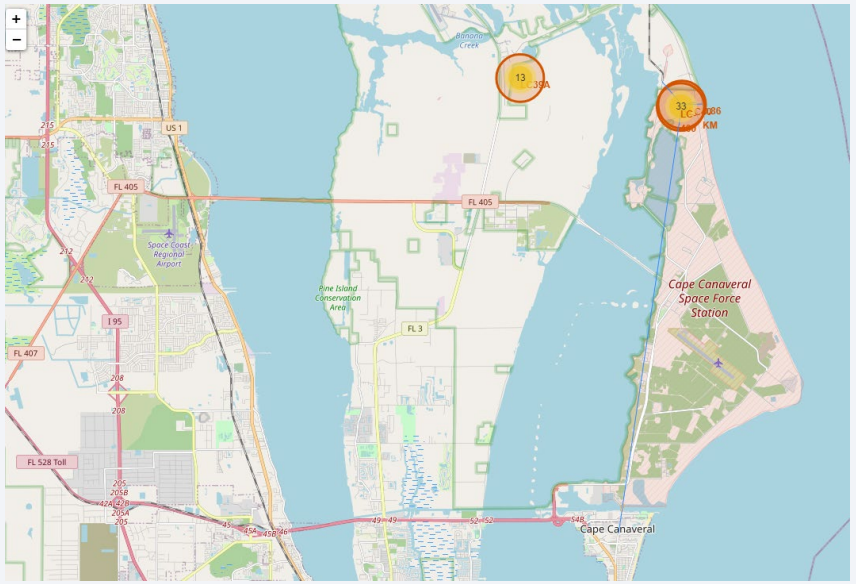
EDA with SQL

1. Displayed the names of the unique launch sites in the space mission
2. Displayed 5 records where launch sites begin with the string 'CCA’
3. Displayed the total payload mass carried by boosters launched by NASA (CRS)
4. Displayed average payload mass carried by booster version F9 v1.1
5. Listed the date when the first successful landing outcome in ground pad was achieved.
6. Listed the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
7. Listed the total number of successful and failure mission outcomes
8. Listed the names of the booster_versions which have carried the maximum payload mass. Use a subquery
9. Listed the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
10. Ranked the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

- Notebook Link: https://github.com/Lhabahbeh/DS_Capstone_Project/blob/master/EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

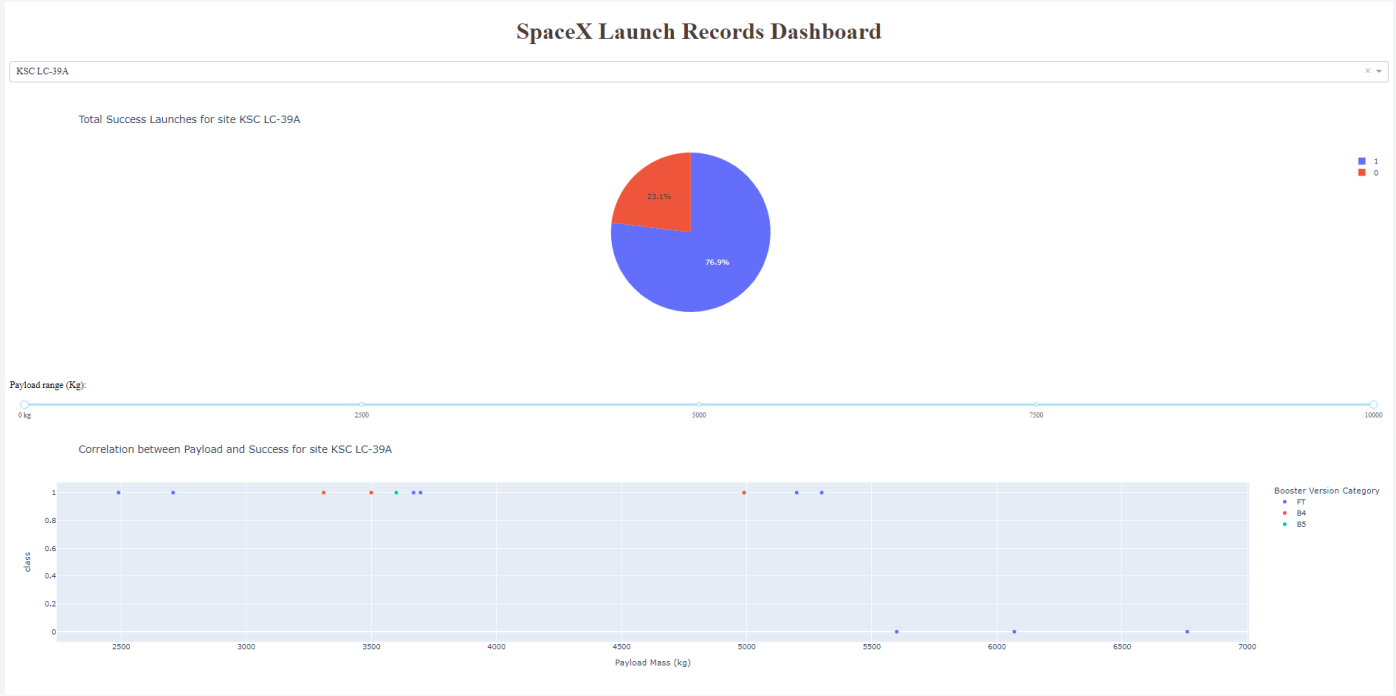




- Markers and lines has been added to an interactive map to find the optimal location for building a launch site.

- Notebook Link: https://github.com/Lhabahbeh/DS_Capstone_Project/blob/master/Launch%20Sites%20Locations%20Analysis%20with%20Folium.ipynb

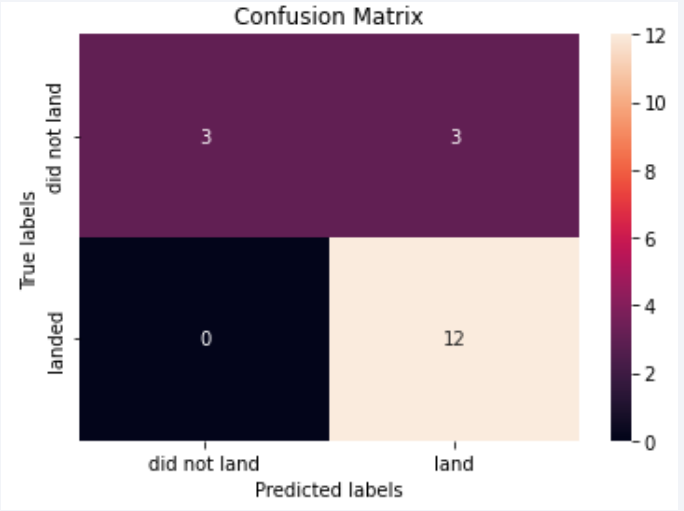
Build a Dashboard with Plotly Dash



- Dashboard has been built to show the highest and largest successful launches, rates and payloads
- Notebook Link: https://github.com/Lhabahbeh/DS_Capstone_Project/blob/master/spacex_dash_app.py

Predictive Analysis (Classification)

Convert Class Column to Pandas Series → Standardize the data → Split data to train/test data → Create models with train data → Test models with test data using score method =



- KNN, Decision tree, SVM & Logistics Regression models had all achieved the highest accuracy at 83.3%.

- Notebook Link: https://github.com/Lhabahbeh/DS_Capstone_Project/blob/master/Machine%20Learning%20Prediction%20lab.ipynb

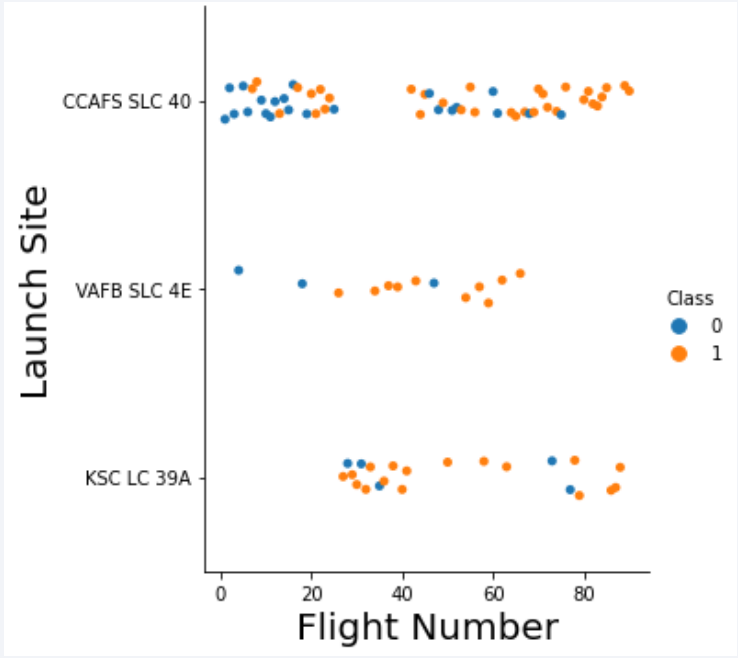
Results

- The SVM, KNN, and Logistic Regression models are the best in term of prediction accuracy for this dataset
- Low weighted payloads perform better then 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 rates

Section 2

Insights drawn from EDA

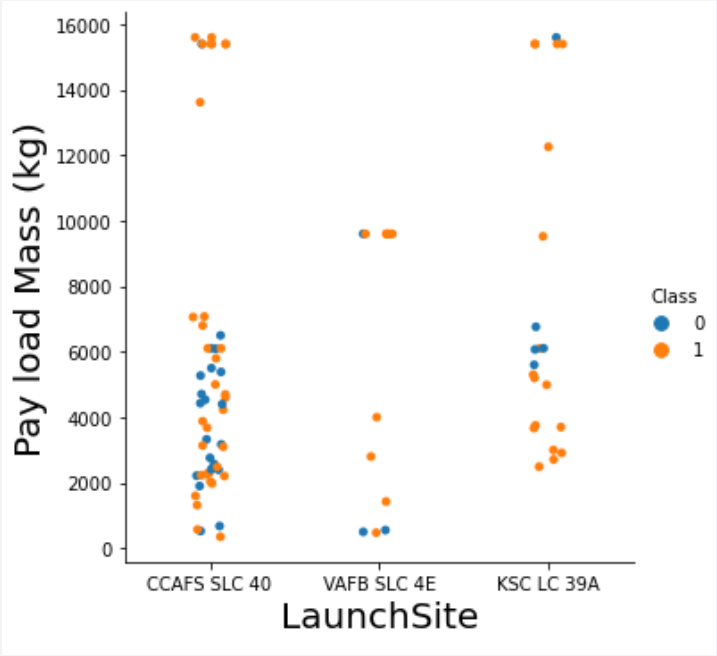
Flight Number vs. Launch Site



Observations:

- CCAFS SLC 40 launch site has higher number of launches than other locations.

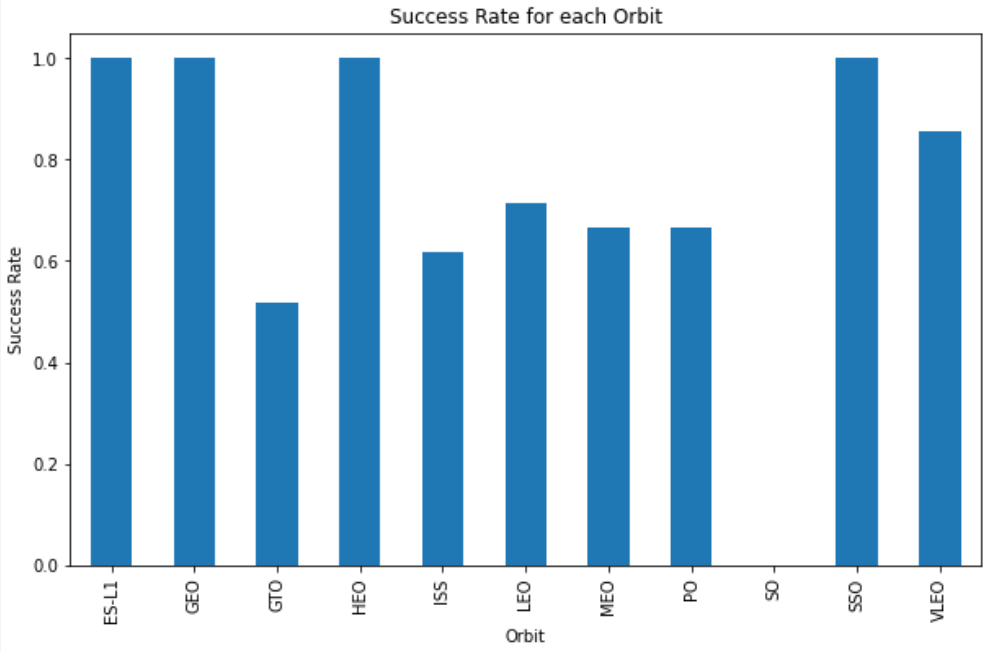
Payload vs. Launch Site



Observations:

- No heavy load launches (greater than 10000) happened in VAFB SLC 4E site

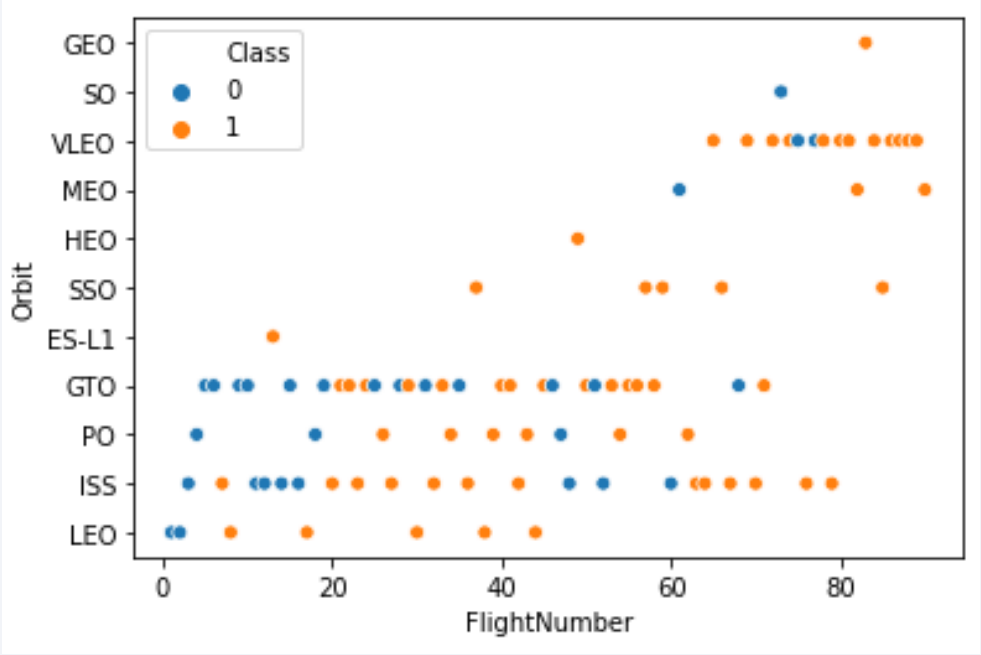
Success Rate vs. Orbit Type



Observations:

- Orbits ES-L1, GEO, HEO, SSP have the highest success rates

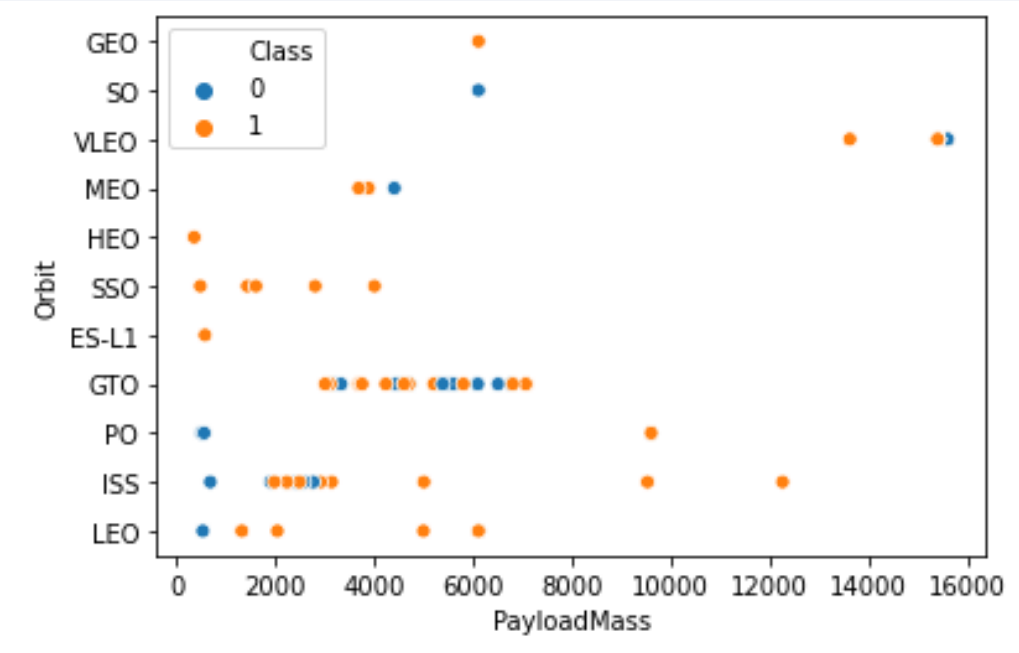
Flight Number vs. Orbit Type



Observations:

- LEO orbit has better success rate related to flight number

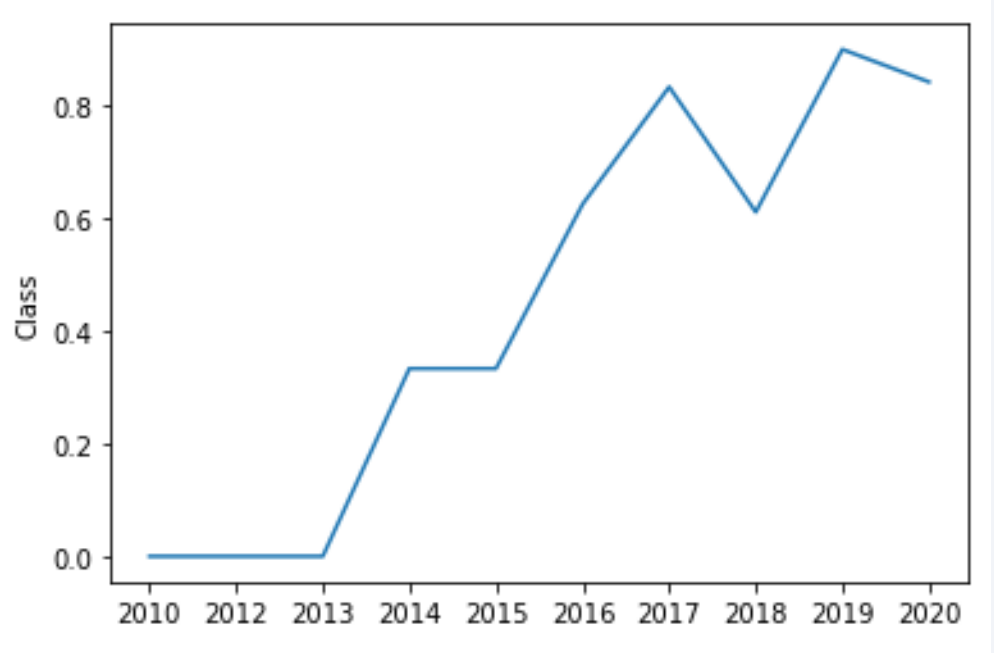
Payload vs. Orbit Type



Observations:

- Successful landing are more for Polar, LEO, and ISS with heavy payloads.

Launch Success Yearly Trend



Observations:

- Success rate kept increasing from 2013 till 2020

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
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	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 Customer like 'NASA (CRS)';
```

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

Total_Payload
45596

Average Payload Mass by F9 v1.1

```
%%sql

SELECT avg(PAYLOAD_MASS__KG_) as Average_Payload
FROM SPACEXTBL
WHERE Booster_Version like 'F9 v1.1%';
```

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

Average_Payload
2534.6666666666665

First Successful Ground Landing Date

```
%%sql

SELECT min(Date)
FROM SPACEXTBL
WHERE Mission_Outcome like 'Success';
```

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

min(Date)
01-03-2013

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%%sql

SELECT *
FROM SPACEXTBL
WHERE PAYLOAD_MASS__KG_ > 4000
AND PAYLOAD_MASS__KG_ < 6000
AND "Landing _Outcome" like "Success (drone ship)";
```

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
06-05-2016	05:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
14-08-2016	05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
30-03-2017	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
11-10-2017	22:53:00	F9 FT B1031.2	KSC LC-39A	SES-11 / EchoStar 105	5200	GTO	SES EchoStar	Success	Success (drone ship)

Total Number of Successful and Failure Mission Outcomes

```
%%sql

SELECT count(Mission_Outcome)
FROM SPACEXTBL;
```

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

count(Mission_Outcome)
101

Boosters Carried Maximum Payload

```
%%sql

SELECT 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, 4, 2) as month
FROM SPACEXTBL
WHERE substr(Date, 7, 4) = '2015';
```

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome	month
10-01-2015	09:47:00	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2395	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)	01
11-02-2015	23:03:00	F9 v1.1 B1013	CCAFS LC-40	DSCOVR	570	HEO	U.S. Air Force NASA NOAA	Success	Controlled (ocean)	02
02-03-2015	03:50:00	F9 v1.1 B1014	CCAFS LC-40	ABS-3A Eutelsat 115 West B	4159	GTO	ABS Eutelsat	Success	No attempt	03
14-04-2015	20:10:00	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6	1898	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)	04
27-04-2015	23:03:00	F9 v1.1 B1016	CCAFS LC-40	Turkmen 52 / MonacoSAT	4707	GTO	Turkmenistan National Space Agency	Success	No attempt	04
28-06-2015	14:21:00	F9 v1.1 B1018	CCAFS LC-40	SpaceX CRS-7	1952	LEO (ISS)	NASA (CRS)	Failure (in flight)	Precluded (drone ship)	06
22-12-2015	01:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)	12

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

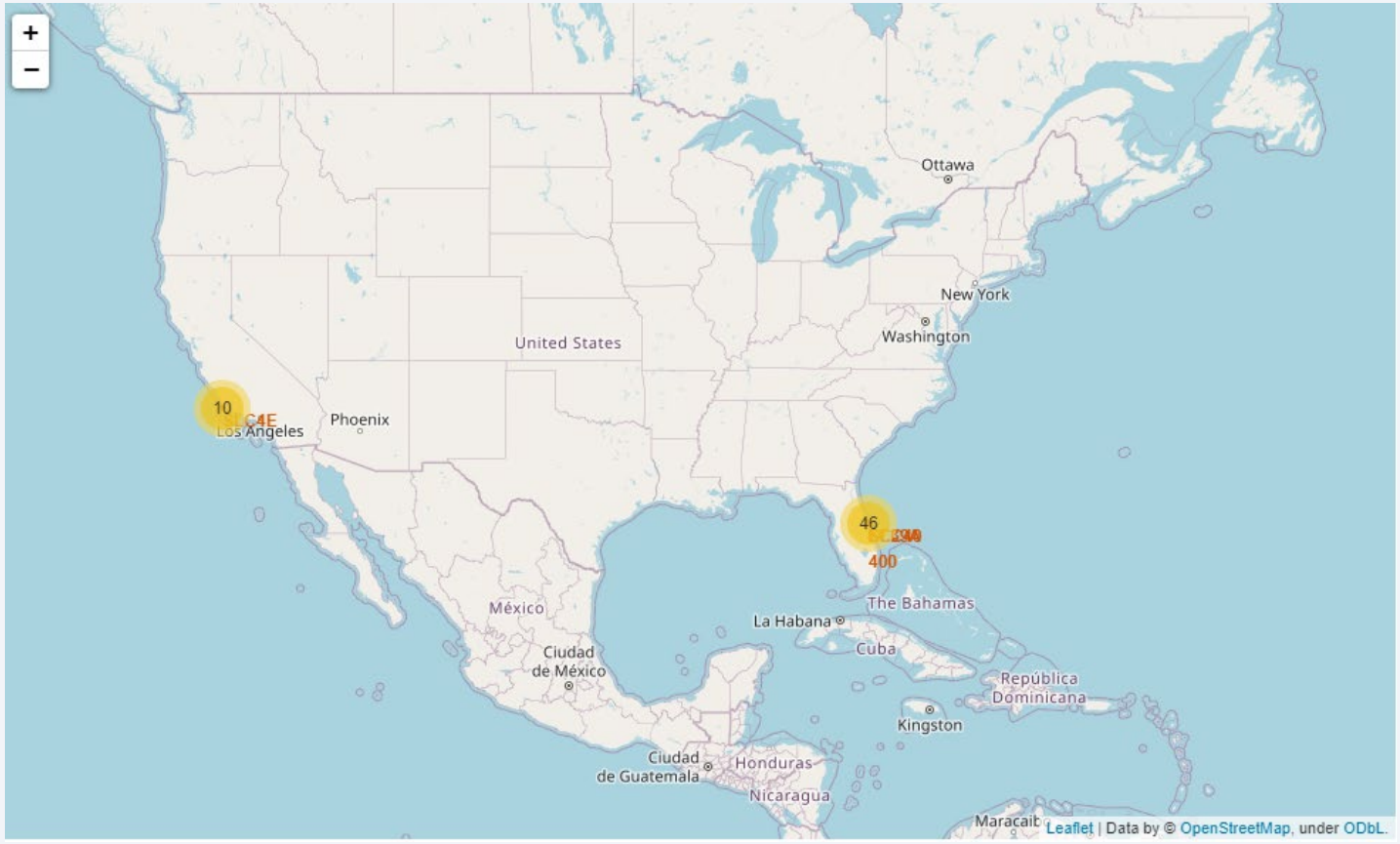
```
%%sql

SELECT count('Landing _Outcome')
FROM SPACEXTBL
WHERE "Landing _Outcome" like "Success%" and DATE between '04-06-2010' and '20-03-2017'
group by ("Landing _Outcome")
order by count('Landing _Outcome') desc;
```

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

count('Landing _Outcome')
20
8
6

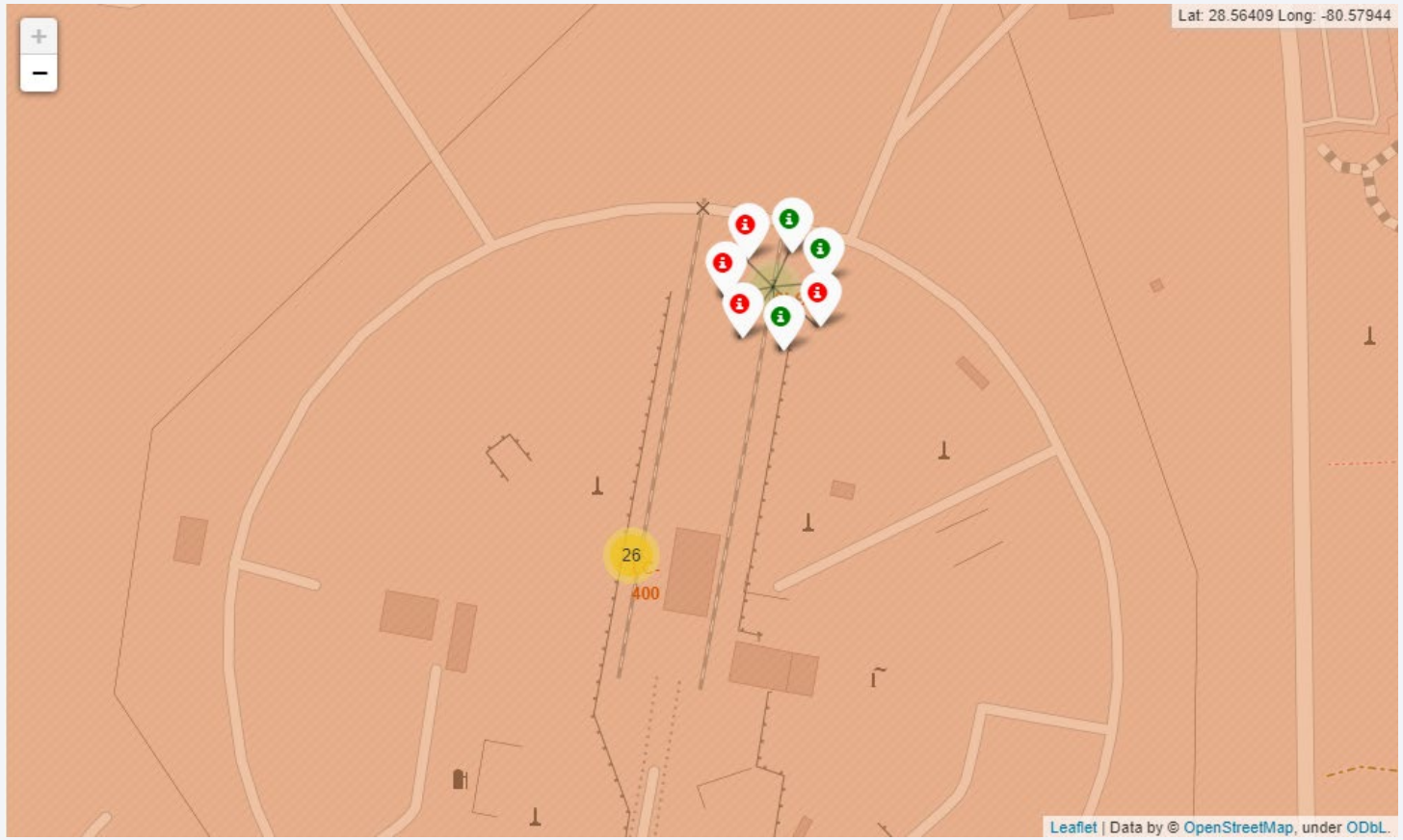

Section 3
Launch Sites
Proximities Analysis

Launch Sites Locations



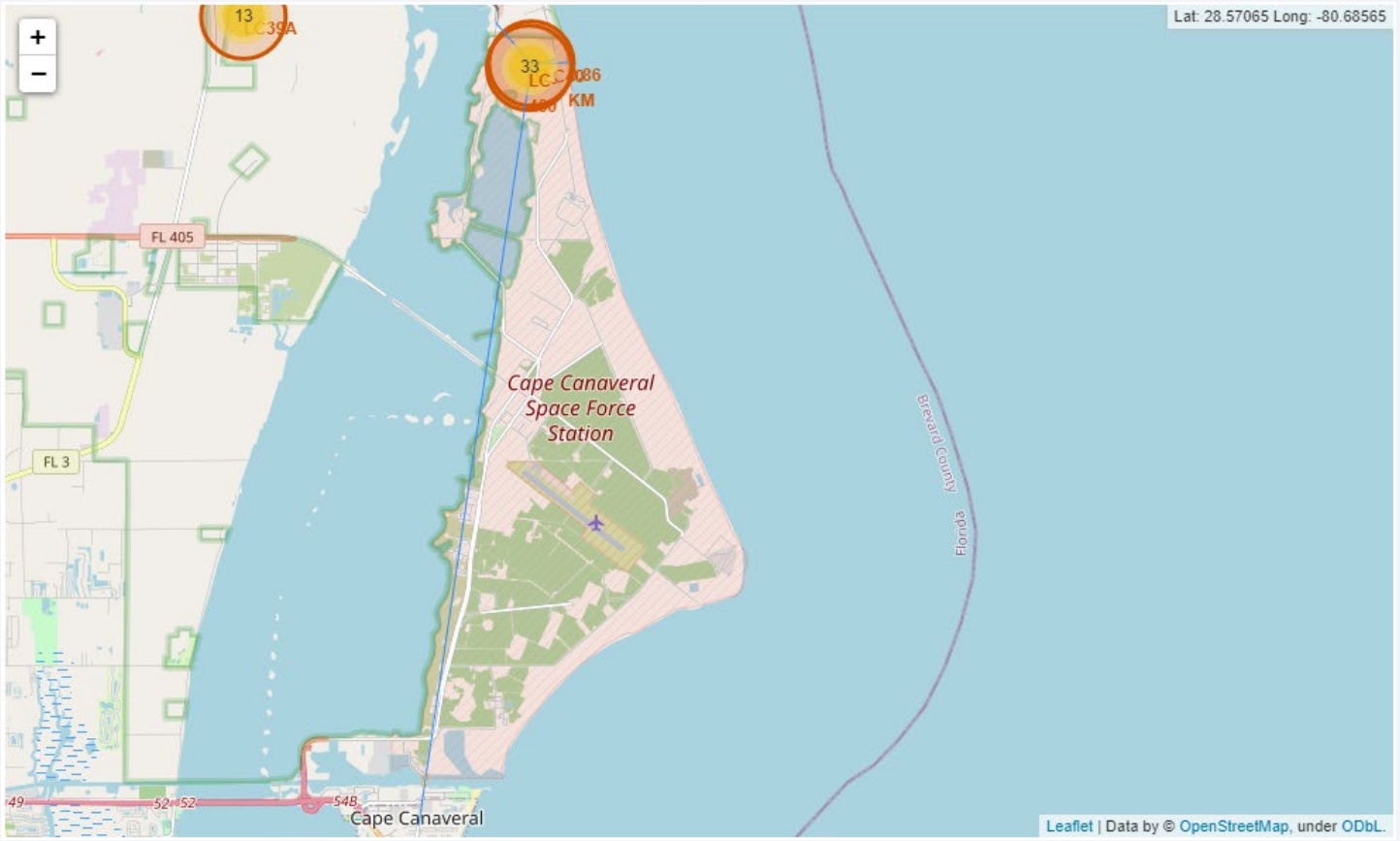
Locations and Number of flights from each location

Success/Failed Launches for each site

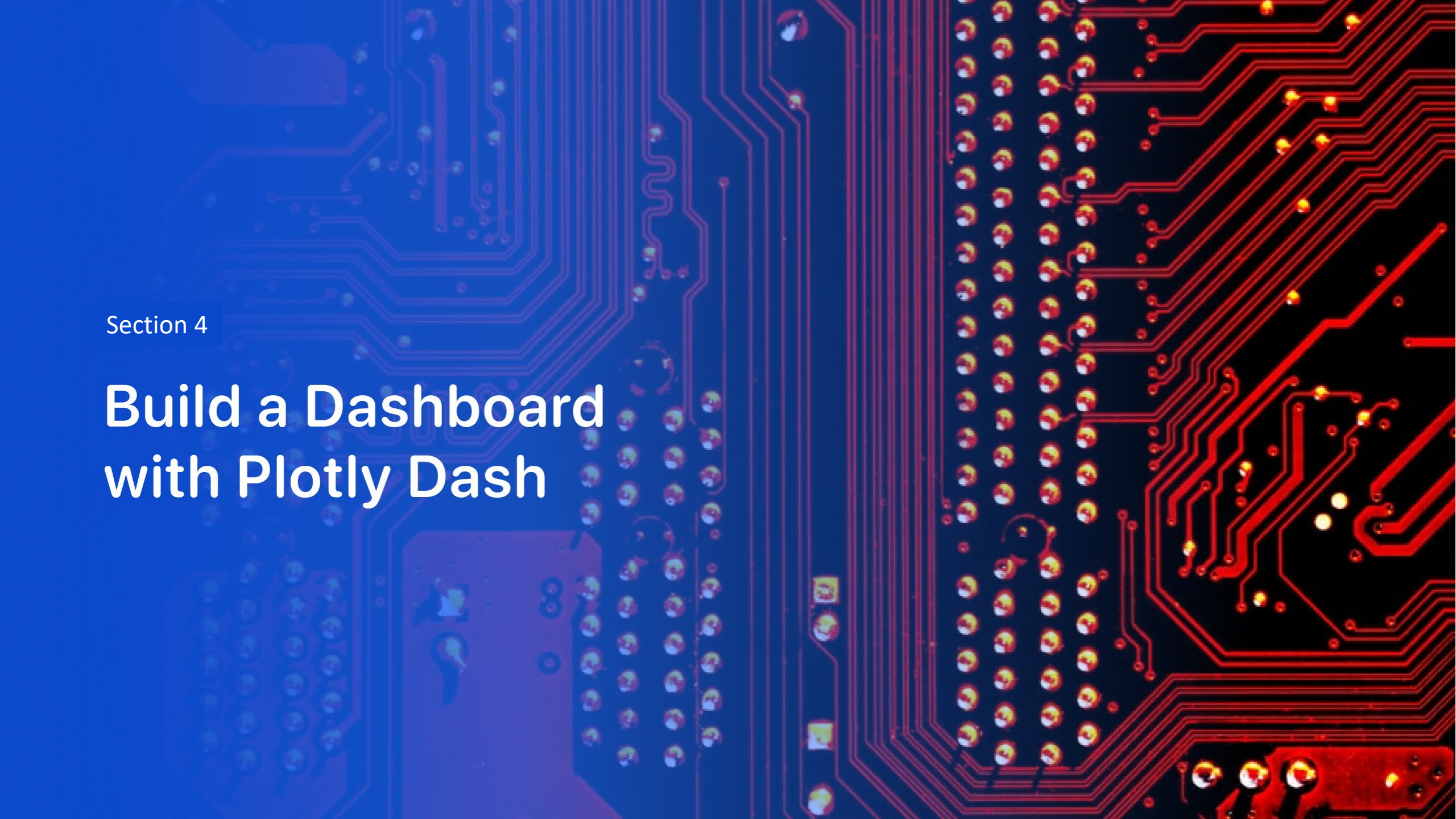


Success/Failed Launches for each site

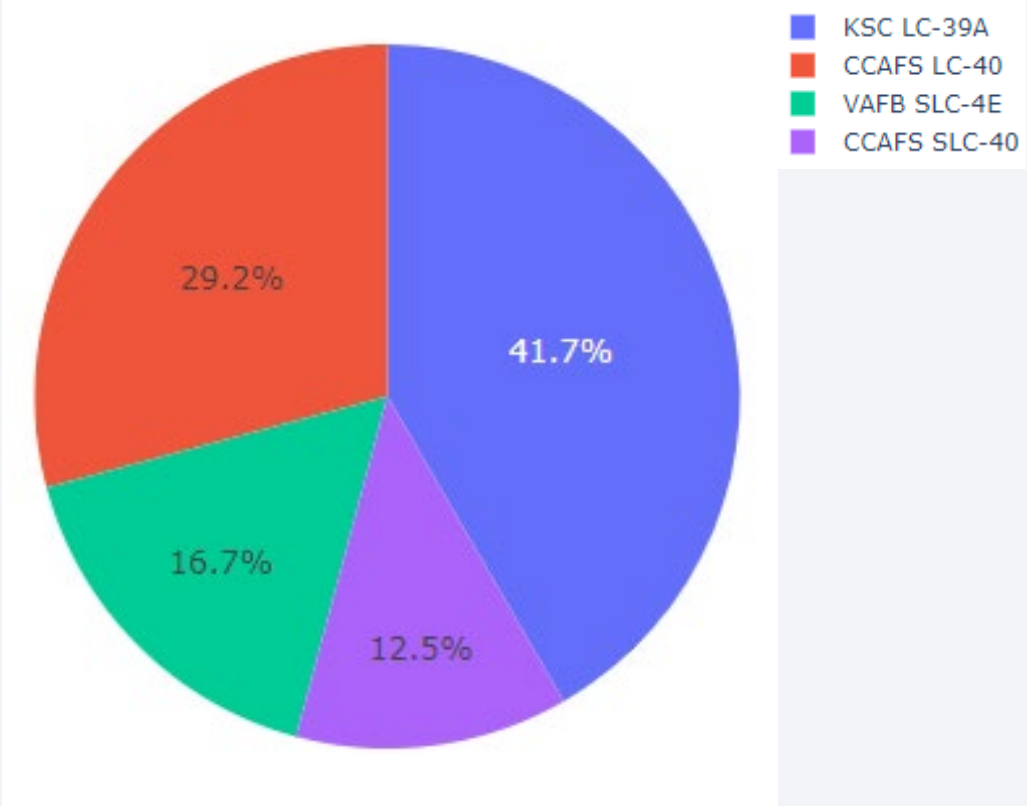
Launch locations with proximities distances



Distances between a launch site to its proximities

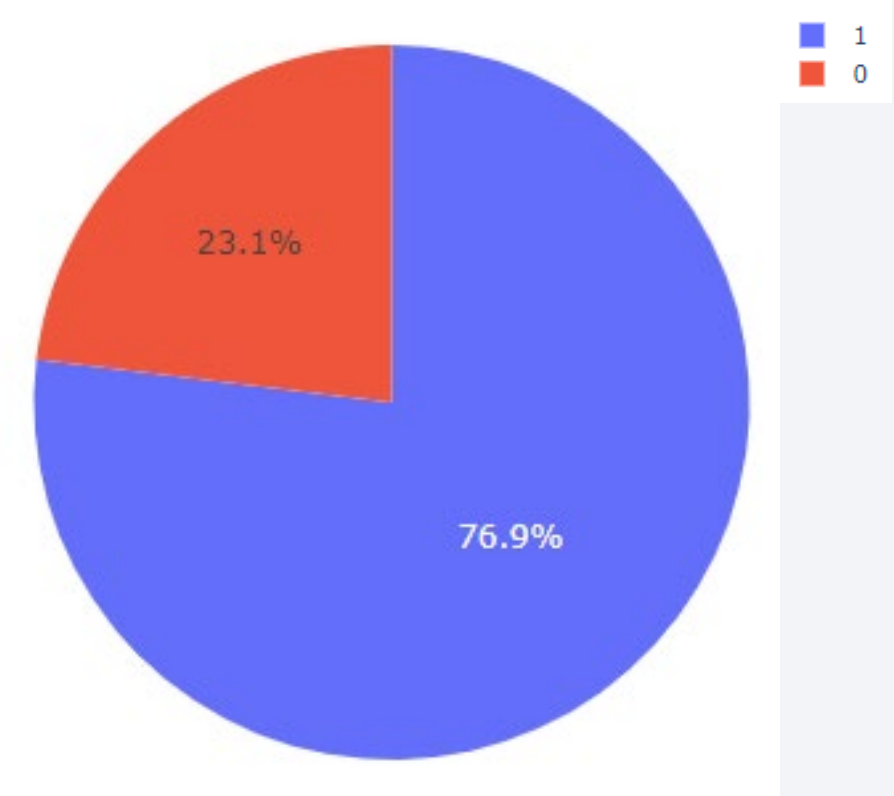

Section 4
Build a Dashboard with Plotly Dash

Launch Success Rate – All Sites



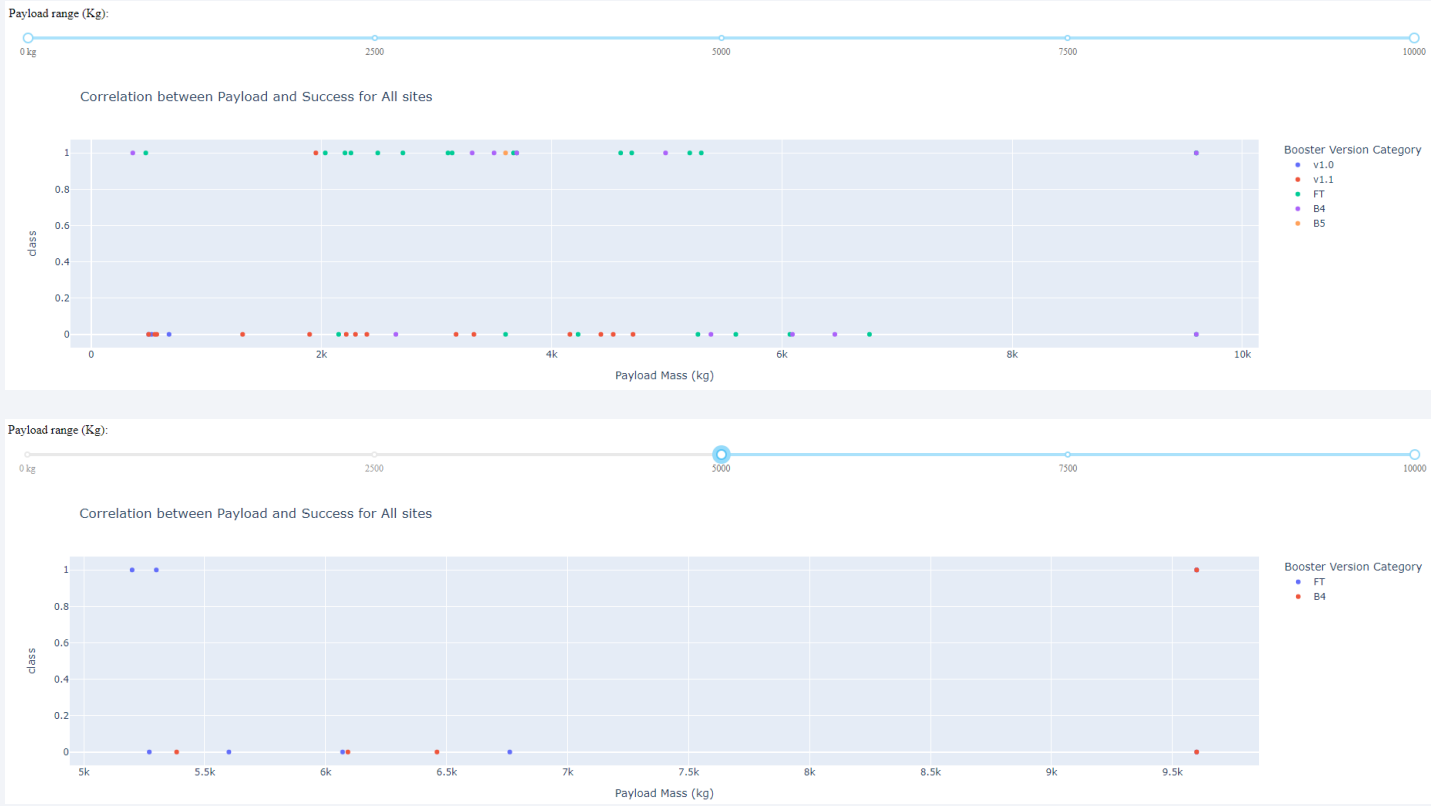
KSC LC-39A has the highest launch success rate

Launch Success Rate – Highest Site



Highest success rate site (KSC LC-39A) has 79.9% successful launches

<Dashboard Screenshot 3>

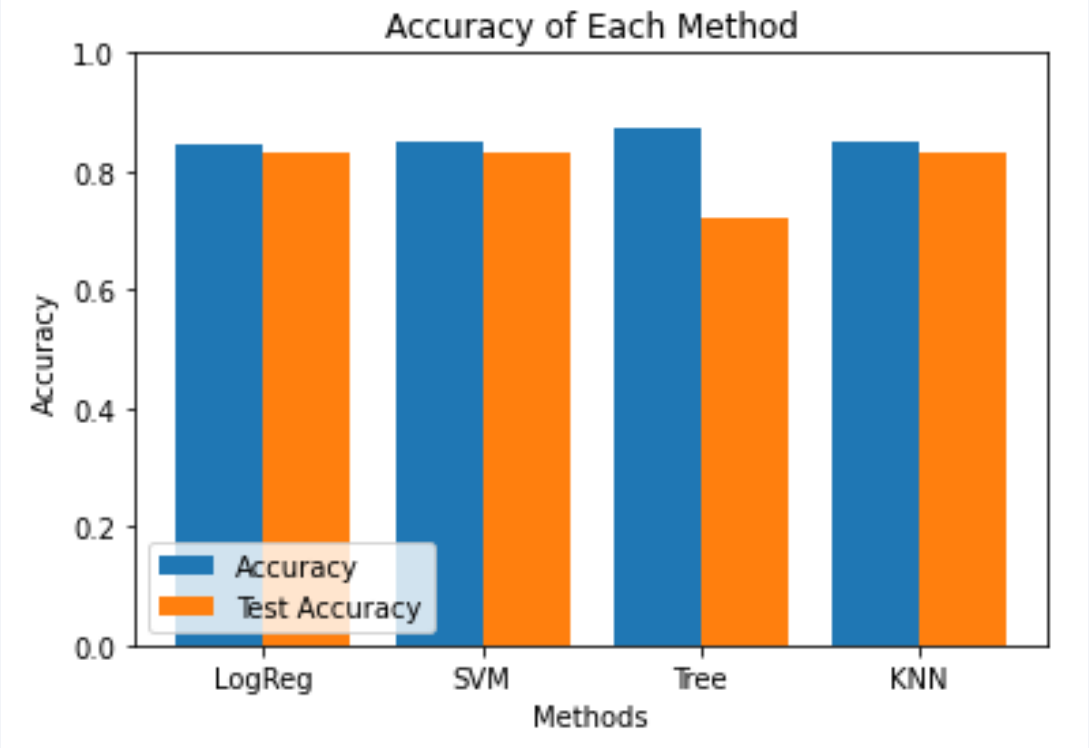


Only B4 version has success launch in heavy payloads greater than 5.5k

Section 5

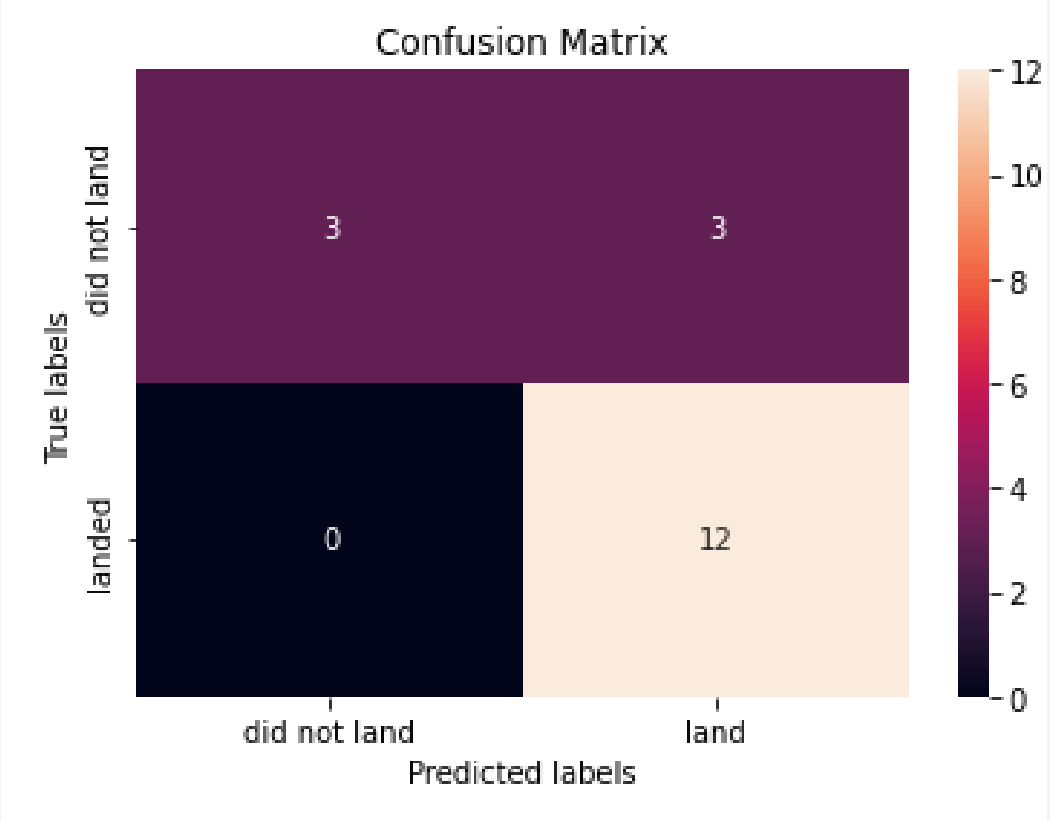
Predictive Analysis (Classification)

Classification Accuracy



Decision Tree model has the highest accuracy

Confusion Matrix



Confusion matrix showing 12 correct predictions and the major issue is the false positives

Conclusions

- The SVM, KNN, and Logistic Regression models are the best in term of prediction accuracy for this dataset
- Low weighted payloads perform better then 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 rates

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!