IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ibrahim Abdullah
26 - 2 - 2024

Outline

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

Executive Summary

- **Summary of methodologies:**

  - Data Collection via API and Web Scraping

  - Data Wrangling

  - Exploratory Data Analysis with SQL

  - Exploratory Data with Data Visualization

  - Interactive Visual Analytics with Folium

  - Machine Learning Prediction

- **Summary of all results:**

  - Exploratory Data Analysis

  - Interactive Analytics

  - Predictive Analytics from Machine Learning

Introduction

- **Project background and context:**

SpaceX launched a rocket called Falcon9 at an estimated cost of $62 million. On the other hand, the cost of the missile to competing companies is high, reaching $165 million. This is due to the rocket being re-landed for use in the next mission. The goal of this project is to create a machine learning prediction of future first-stage landing outcomes.

- **Problems you want to find answers:**

- Find the features that affect the first stage landing success

- predict if the first stage will land successfully ?

Section 1

Methodology

Methodology

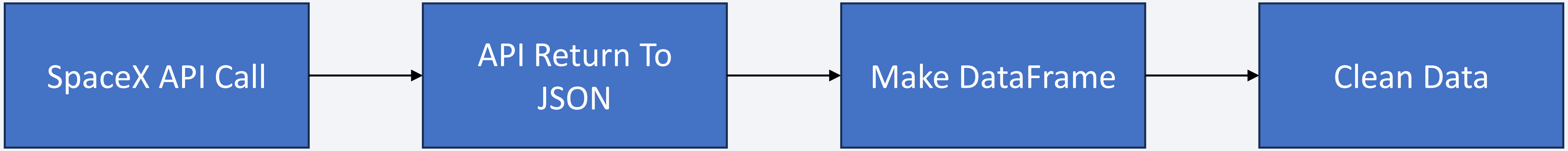
Executive Summary

- Data collection methodology:
  - Data was collected using SpaceX API and web scraping from Wikipedia.
- Perform data wrangling
  - One-hot encoding was applied to categorical features.
- 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

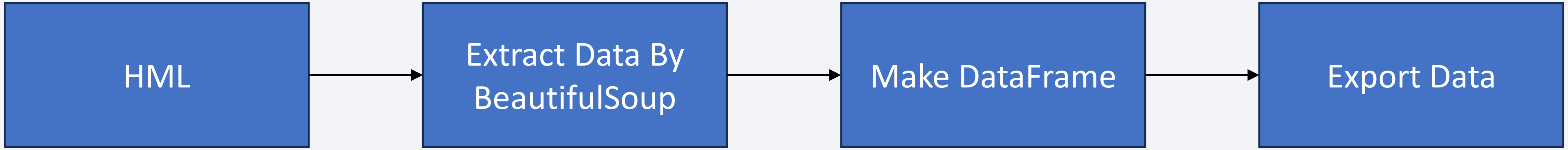
Data Collection

- Datasets are collected from SpaceX API and Web Scraping Wikipedia:

- From SpaceX API :



- From Web Scraping Wikipedia :

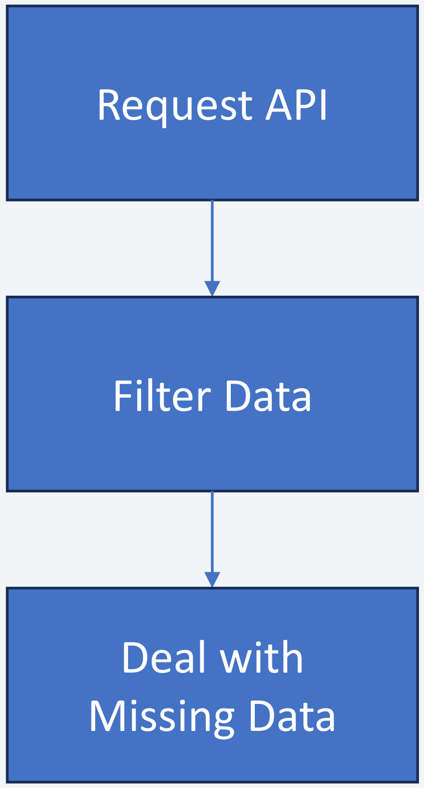


Data Collection – SpaceX API

- Used the get request to the SpaceX API to collect data, clean the requested data and data wrangling and formatting data.

- GitHub URL :

https://github.com/IbrahimAbdullah93/Applied-Data-Science-Capstone/blob/main/Complete%20the%20Data%20Collection%20API%20Lab.ipynb

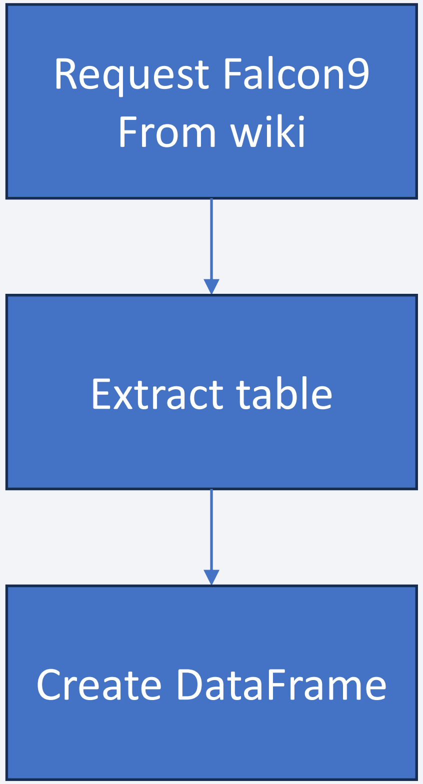


Data Collection - Scraping

- We applied webscraping to Falcon 9 launch with BeautifulSoupm and make it in the DataFrame.

- GitHub URL :

https://github.com/IbrahimAbdullah93/Applied-Data-Science-Capstone/blob/main/Complete%20the%20Data%20Collection%20with%20Web%20Scraping%20lab.ipynb



Data Wrangling

- There are two cases in the dataset:

- True Ocean, This means that the landing mission was successful

- False Ocean, This means that the landing mission was a failure

Here we will convert the categories into numbers, where 1 means a successful landing and 0 means a failed landing.

- GitHub URL :

https://github.com/IbrahimAbdullah93/Applied-Data-Science-Capstone/blob/main/Data%20Wrangling.ipynb

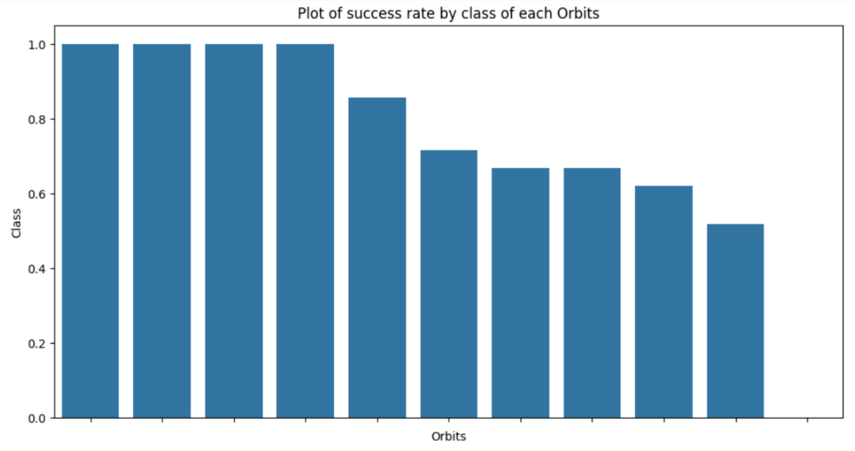
EDA with Data Visualization

- **Scatter Graphs :**

- Flight Number vs. Payload Mass
- Flight Number vs. Launch Site
- Payload vs. Launch Site
- Orbit vs. Flight Number
- Payload vs. Orbit Type
- Orbit vs. Payload Mass

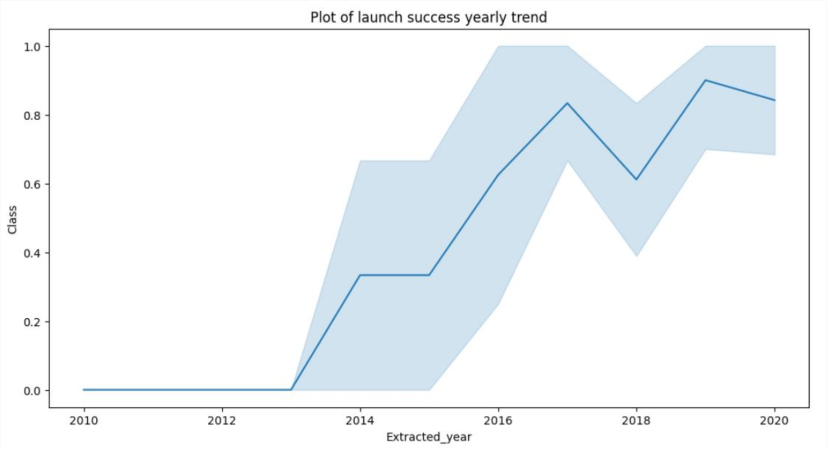
- **Bar Graph :**

- Success rate vs. Orbit



- **Line Graph :**

- Success rate vs. Year



- GitHub URL:

https://github.com/IbrahimAbdullah93/Applied-Data-Science-Capstone/blob/main/EDA%20with%20Visualization%20Lab.ipynb

EDA with SQL

- Loaded SpaceX dataset into the SQLITE.
- We applied EDA with SQL to get insight from the data.

    - The names of unique launch sites in the space mission.

    - The total payload mass carried by boosters launched by NASA (CRS)

    - The average payload mass carried by booster version F9 v1.1

    - The total number of successful and failure mission outcome

    - The failed landing outcomes in drone ship, their booster version and launch site names.

- GitHub URL :

https://github.com/IbrahimAbdullah93/Applied-Data-Science-Capstone/blob/main/The%20EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

- To visualize the launch data into an interactive map. We took the latitude and longitude coordinates at each launch site and added a circle marker around each launch site with a label of the name of the launch site.
- We then assigned the dataframe launch_outcomes(failure,success) to classes 0 and 1 with Red and Green markers on the map in MarkerCluster().
- We then used the Haversine's formula to calculated the distance of the launch sites to various landmark to find answer to the questions of:

  - How close the launch sites with railways, highways and coastlines?

  - How close the launch sites with nearby cities?

- GitHub URL :

https://github.com/IbrahimAbdullah93/Applied-Data-Science-Capstone/blob/main/Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb

Build a Dashboard with Plotly Dash

- We built an interactive dashboard with Plotly dash which allowing the user to play around with the data as they need.

- We plotted pie charts showing the total launches by a certain sites.

- We then plotted scatter graph showing the relationship with Outcome and Payload Mass (Kg) for the different booster version.

Predictive Analysis (Classification)

Building the Model :

- Load the dataset into NumPy and Pandas
- Transform the data and then split into training and test datasets
- Decide which type of ML to use

Evaluating the Model :

- Check the accuracy for each model
- Get tuned hyperparameters for each type of algorithms.
- Plot the confusion matrix

Improving the Model :

- Use Feature Engineering and Algorithm Tuning

Find the Best Model :

- The model with the best accuracy score will be the best performing model

- GitHub URL :

https://github.com/IbrahimAbdullah93/Applied-Data-Science-Capstone/blob/main/Machine%20Learning%20Prediction.ipynb

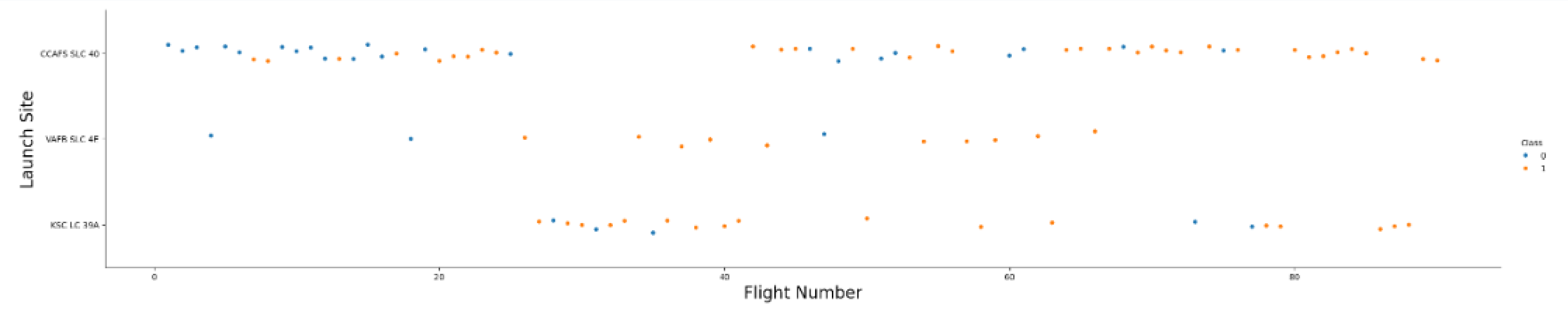
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

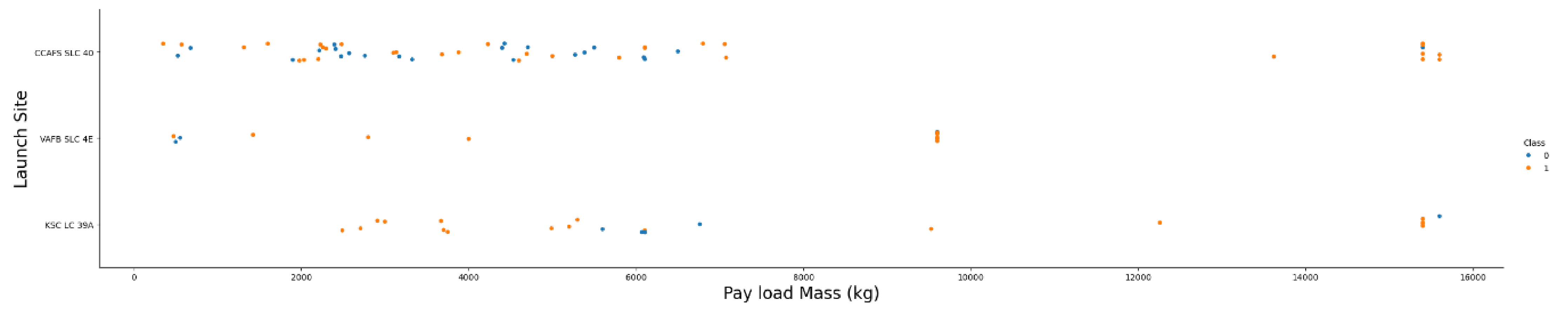
Insights drawn from EDA

Flight Number vs. Launch Site

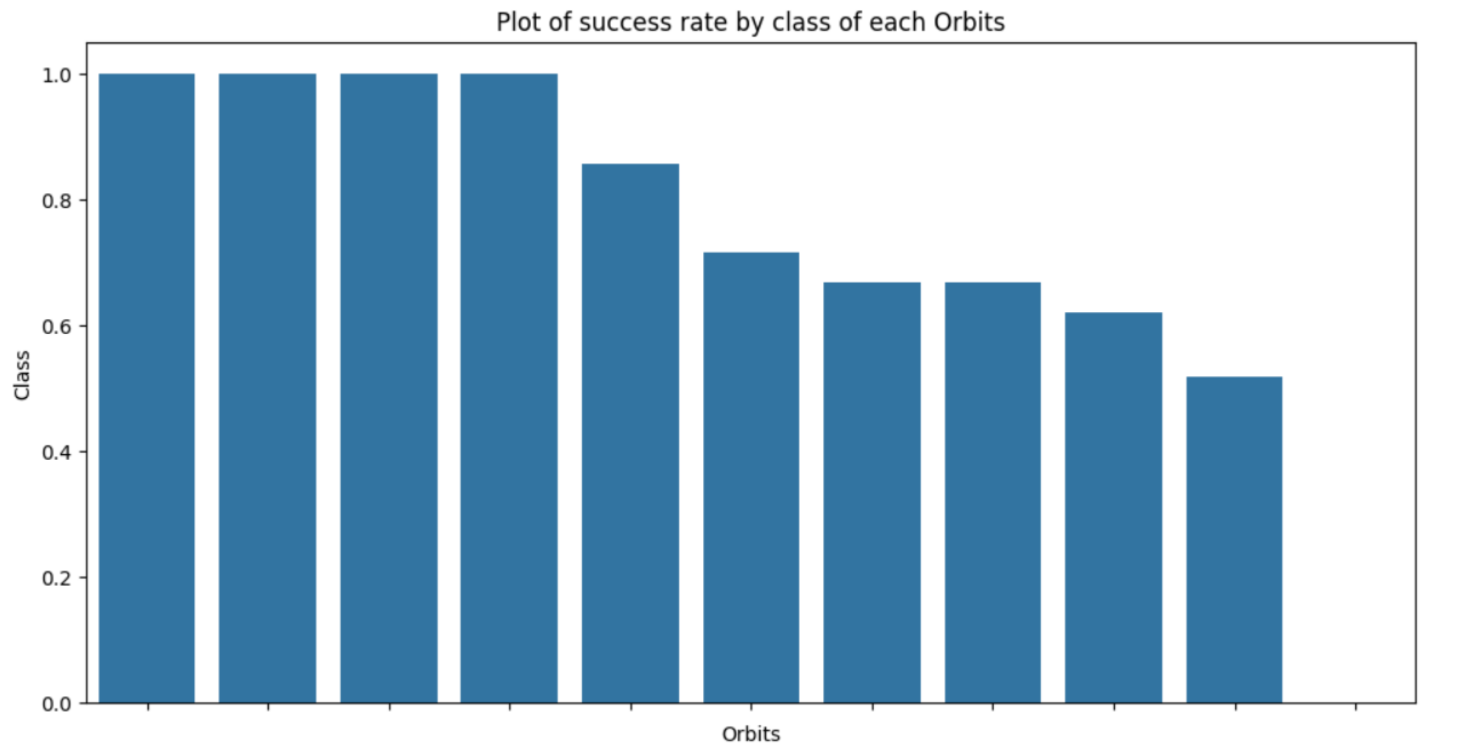


- The success rate is increasing

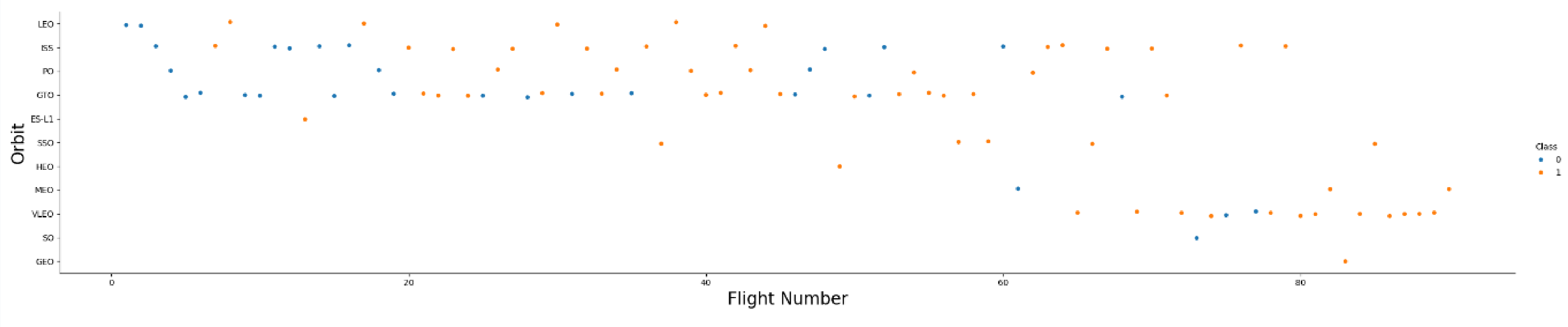
Payload vs. Launch Site



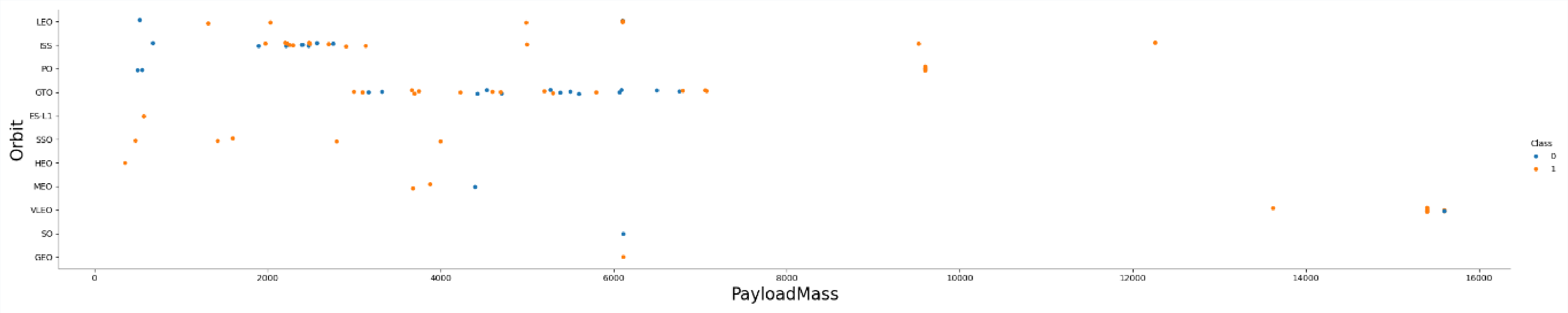
Success Rate vs. Orbit Type



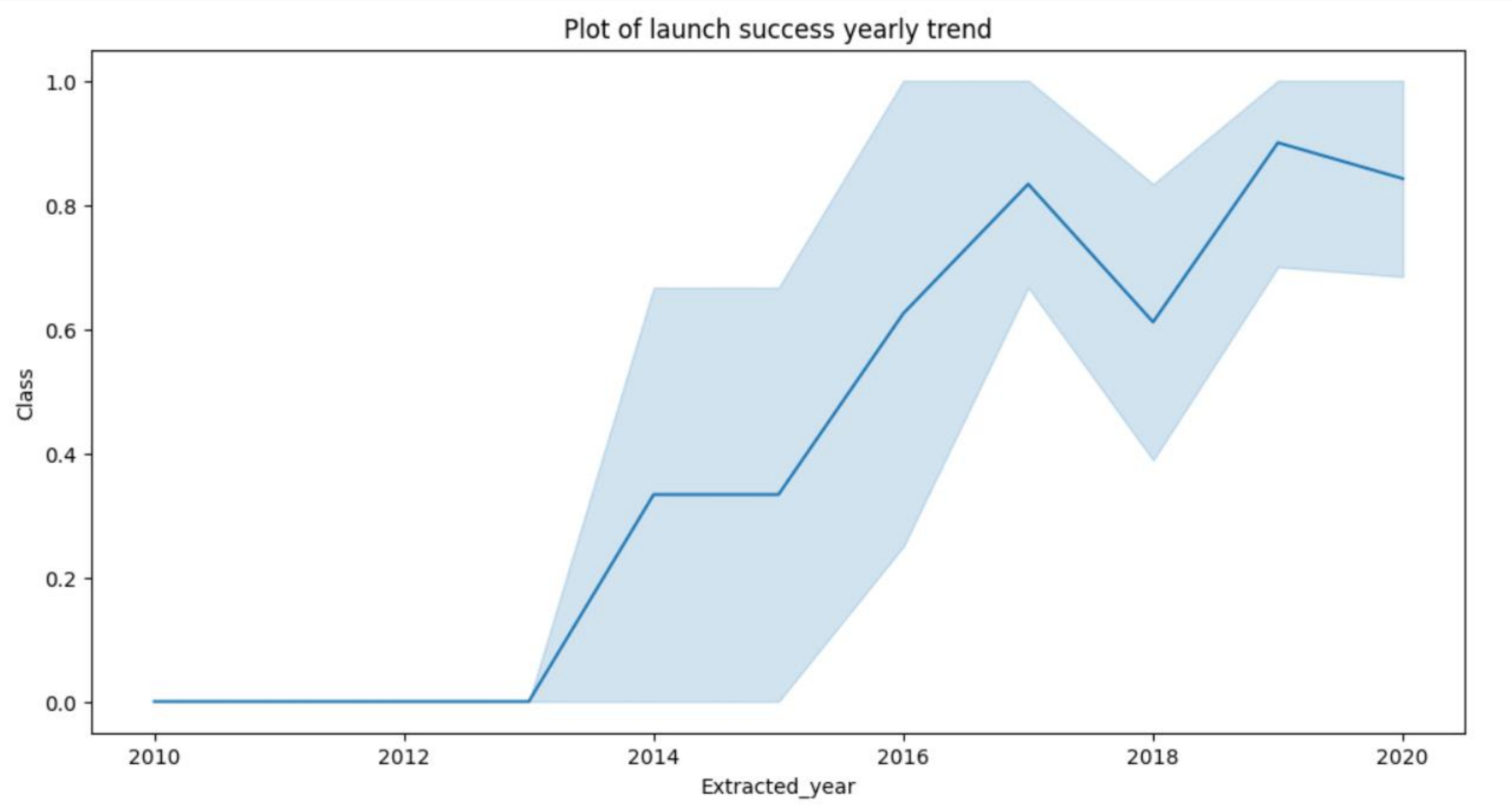
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

```
%sql SELECT DISTINCT LAUNCH_SITE as "Launch_Sites" FROM SPACEXTBL;
```

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

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

We used the key word **DISTINC**

To show only unique launch sites from the SpaceX data

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

The **WHERE** clause followed by **LIKE** clause filters launch sites that contain the substring CCA.

LIMIT 5 shows 5 records fromfiltering.

Total Payload Mass

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) as "Total Payload Mass(Kgs)", Customer FROM SPACEXTBL WHERE Customer = 'NASA (CRS)';
```

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

Total Payload Mass(Kgs)	Customer
45596	NASA (CRS)

Average Payload Mass by F9 v1.1

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) as "Payload Mass Kgs", Customer, Booster_Version FROM SPACEXTBL WHERE Booster_Version LIKE 'F9 v1.1%';
```

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

Payload Mass Kgs	Customer	Booster_Version
2534.6666666666665	MDA	F9 v1.1 B1003

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

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%%sql SELECT DISTINCT Booster_Version, Payload FROM SPACEXTBL WHERE Landing_Outcome = "Success (drone ship)"
AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000;
```

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

Booster_Version	Payload
F9 FT B1022	JCSAT-14
F9 FT B1026	JCSAT-16
F9 FT B1021.2	SES-10
F9 FT B1031.2	SES-11 / EchoStar 105

Total Number of Successful and Failure Mission Outcomes

```
%sql SELECT Mission_Outcome, COUNT("Mission_Outcome") as Total FROM SPACEXTBL GROUP BY "Mission_Outcome";
```

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

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

Boosters Carried Maximum Payload

```
%%sql SELECT Booster_Version, Payload, PAYLOAD_MASS__KG_ FROM SPACEXTBL
WHERE PAYLOAD_MASS__KG_ = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTBL);
```

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

Booster_Version	Payload	PAYLOAD_MASS__KG_
F9 B5 B1048.4	Starlink 1 v1.0, SpaceX CRS-19	15600
F9 B5 B1049.4	Starlink 2 v1.0, Crew Dragon in-flight abort test	15600
F9 B5 B1051.3	Starlink 3 v1.0, Starlink 4 v1.0	15600
F9 B5 B1056.4	Starlink 4 v1.0, SpaceX CRS-20	15600
F9 B5 B1048.5	Starlink 5 v1.0, Starlink 6 v1.0	15600
F9 B5 B1051.4	Starlink 6 v1.0, Crew Dragon Demo-2	15600
F9 B5 B1049.5	Starlink 7 v1.0, Starlink 8 v1.0	15600
F9 B5 B1060.2	Starlink 11 v1.0, Starlink 12 v1.0	15600
F9 B5 B1058.3	Starlink 12 v1.0, Starlink 13 v1.0	15600
F9 B5 B1051.6	Starlink 13 v1.0, Starlink 14 v1.0	15600
F9 B5 B1060.3	Starlink 14 v1.0, GPS III-04	15600
F9 B5 B1049.7	Starlink 15 v1.0, SpaceX CRS-21	15600

2015 Launch Records

```
%%sql SELECT substr(Date,6,2), substr(Date,0,5), Booster_Version, Launch_Site, Payload, PAYLOAD_MASS__KG_, Mission_Outcome, Landing_Outcome
FROM SPACEXTBL WHERE substr(Date,0,5)='2015' AND Landing_Outcome = 'Failure (drone ship)';
```

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

substr(Date,6,2)	substr(Date,0,5)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Mission_Outcome	Landing_Outcome
01	2015	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2395	Success	Failure (drone ship)
04	2015	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6	1898	Success	Failure (drone ship)

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

```
%sql SELECT * FROM SPACEXTBL WHERE Landing_Outcome LIKE 'Success%' AND (Date BETWEEN '2010-06-04' AND '2017-03-20') ORDER BY Date DESC;
```

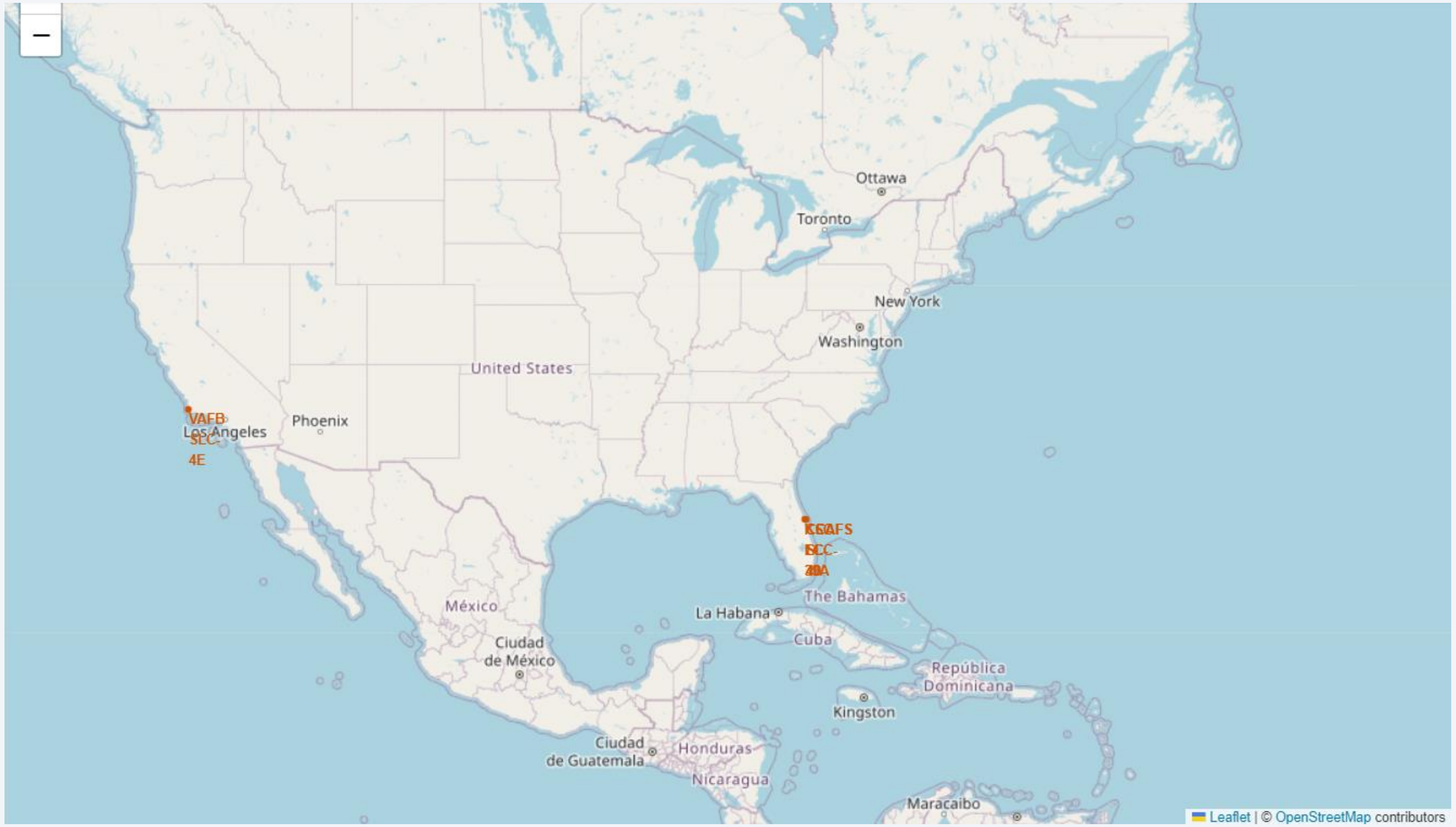
```
 * sqlite:///my_data1.db
Done.
```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2017-01-14	17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600	Polar LEO	Iridium Communications	Success	Success (drone ship)
2016-08-14	5:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-07-18	4:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2016-05-27	21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success	Success (drone ship)
2016-05-06	5:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-04-08	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136	LEO (ISS)	NASA (CRS)	Success	Success (drone ship)
2015-12-22	1:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

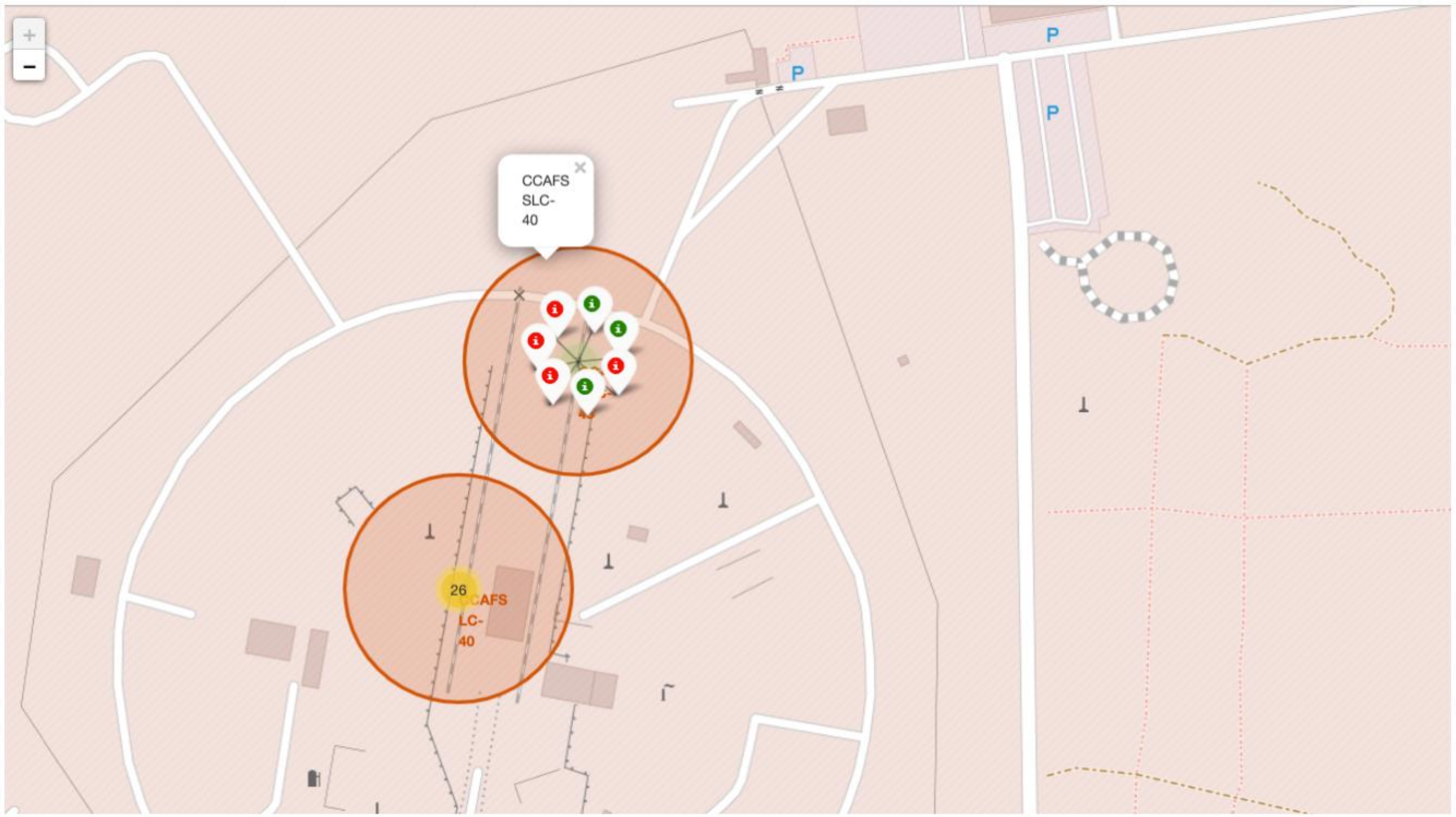
Section 3

Launch Sites Proximities Analysis

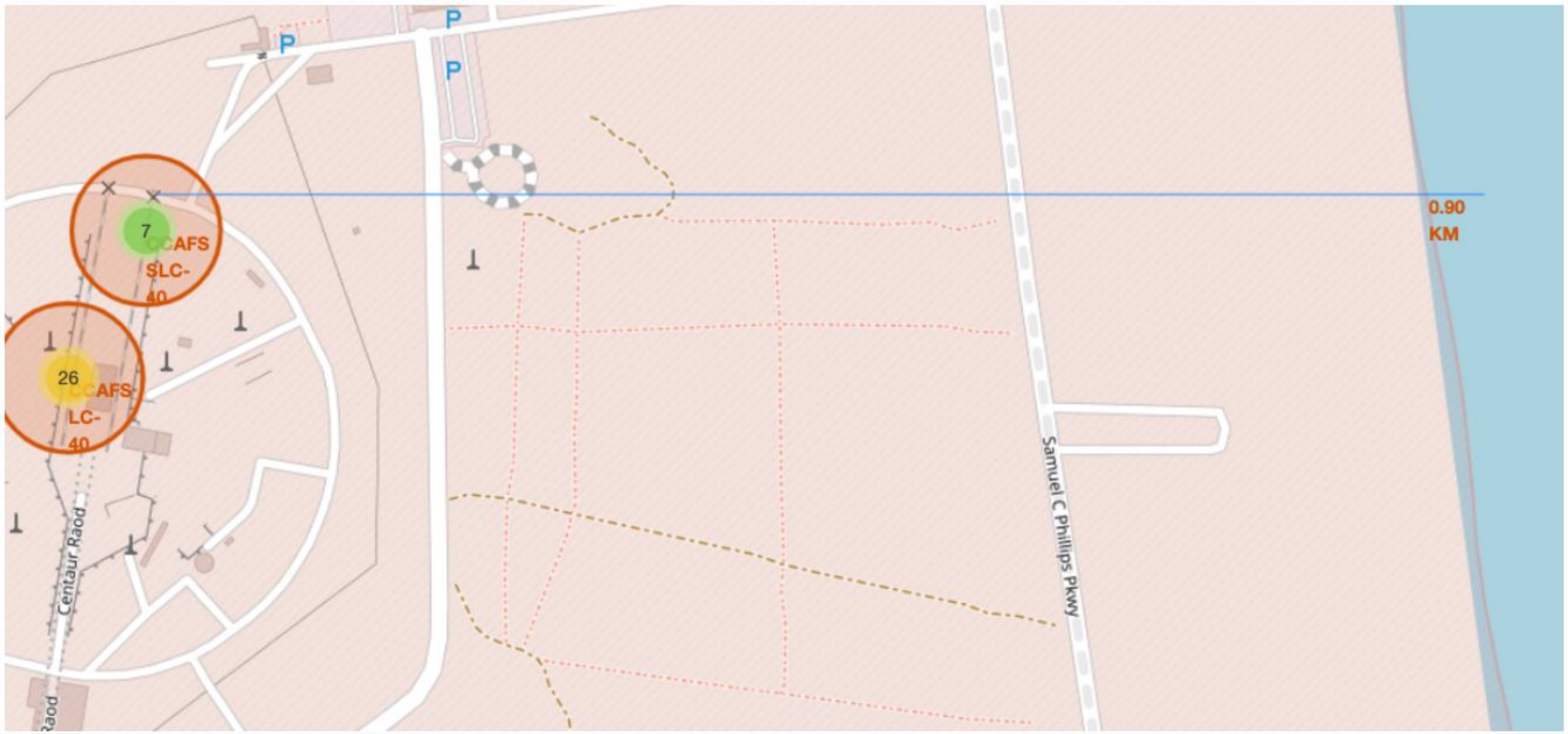
All launch sites global map markers



Markers showing launch sites with color labels



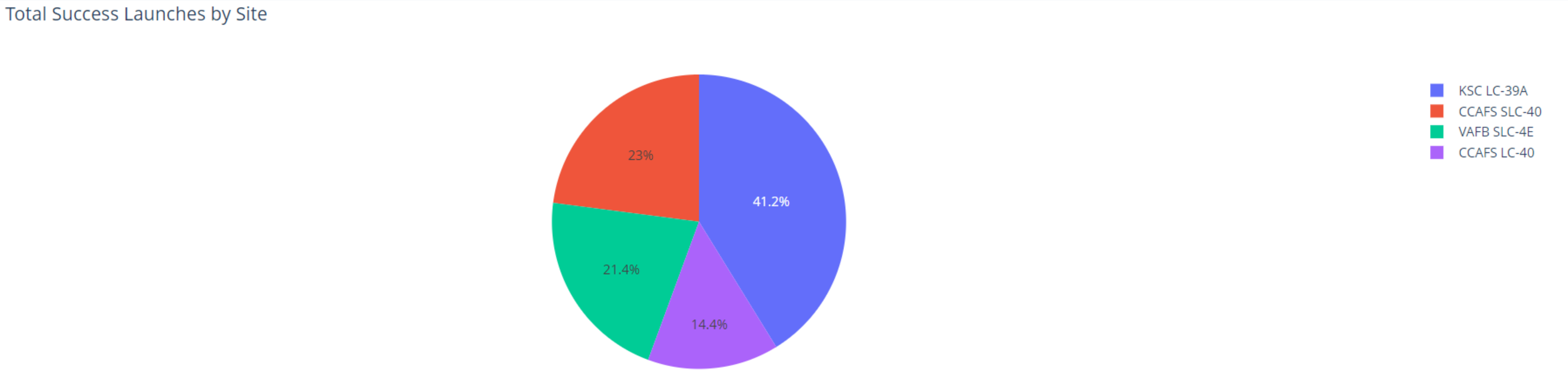
Launch Site distance to landmarks



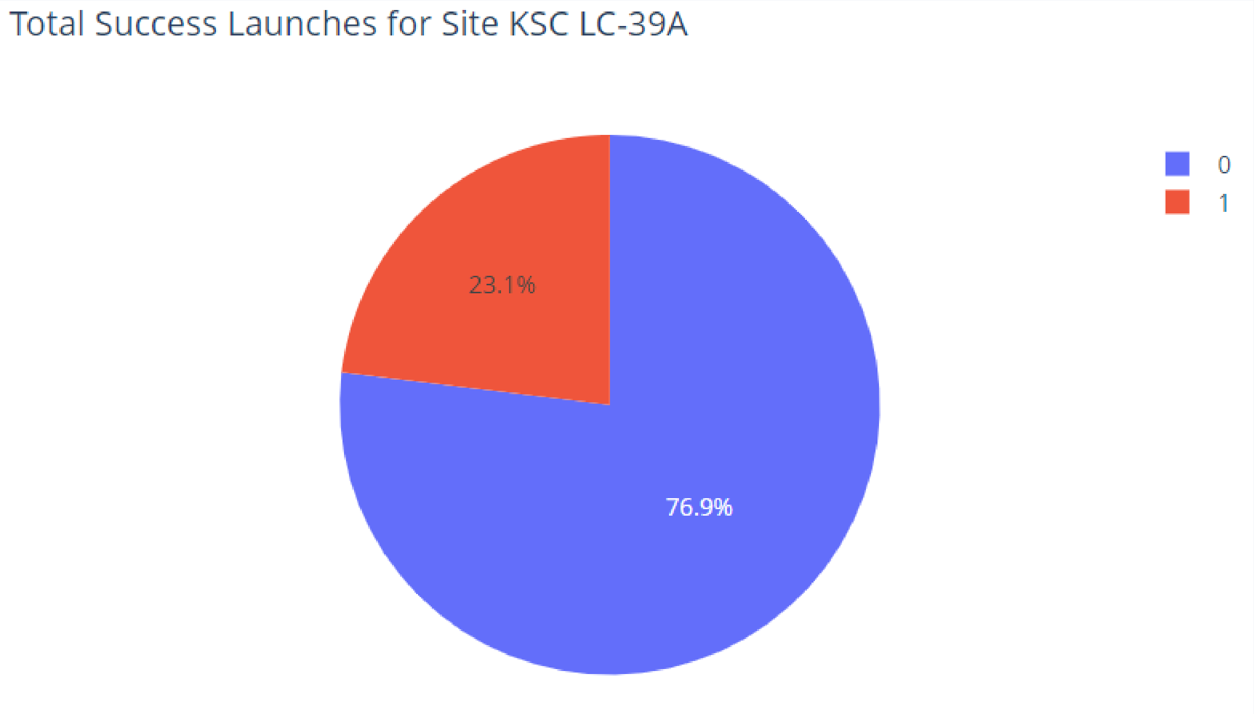
Section 4

Build a Dashboard with Plotly Dash

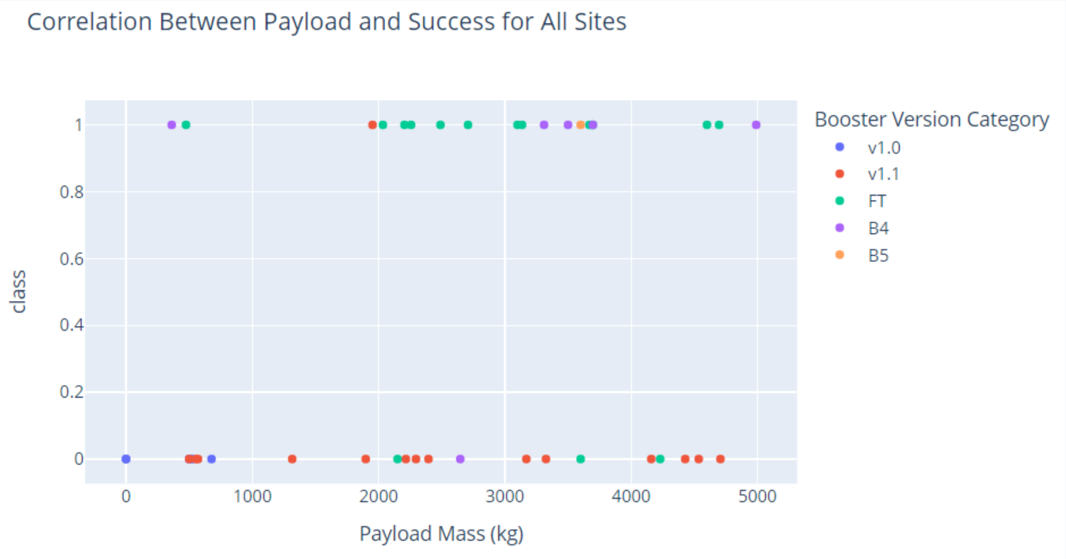
The success percentage by each sites

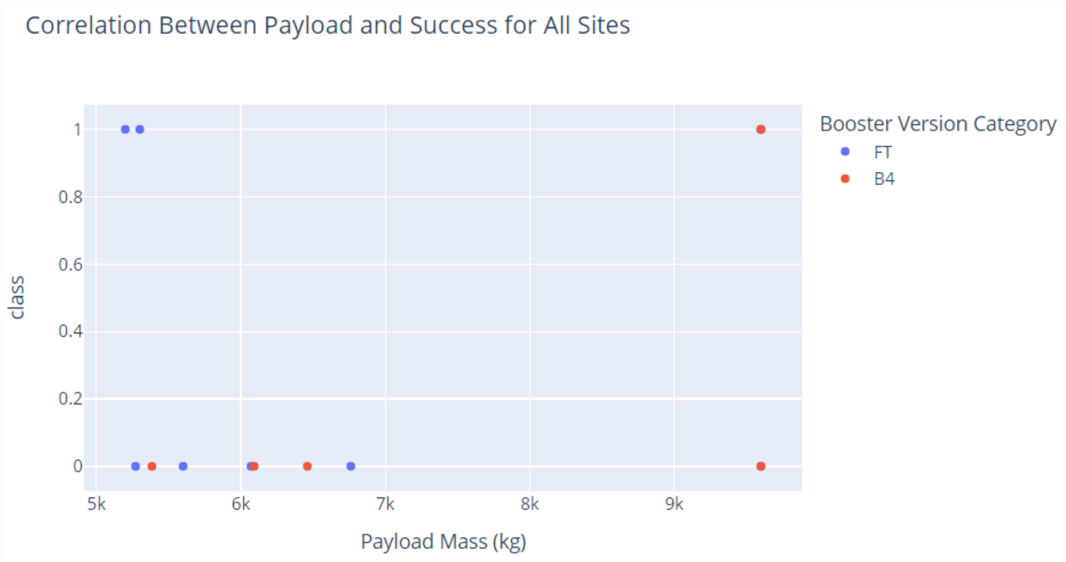


The highest launch-success ratio: KSC LC-39A



Payload vs Launch Outcome Scatter Plot





Low Weighted Payload
(0 – 5000 kg)

Heavy Weighted Payload
(5000 – 10000 kg)

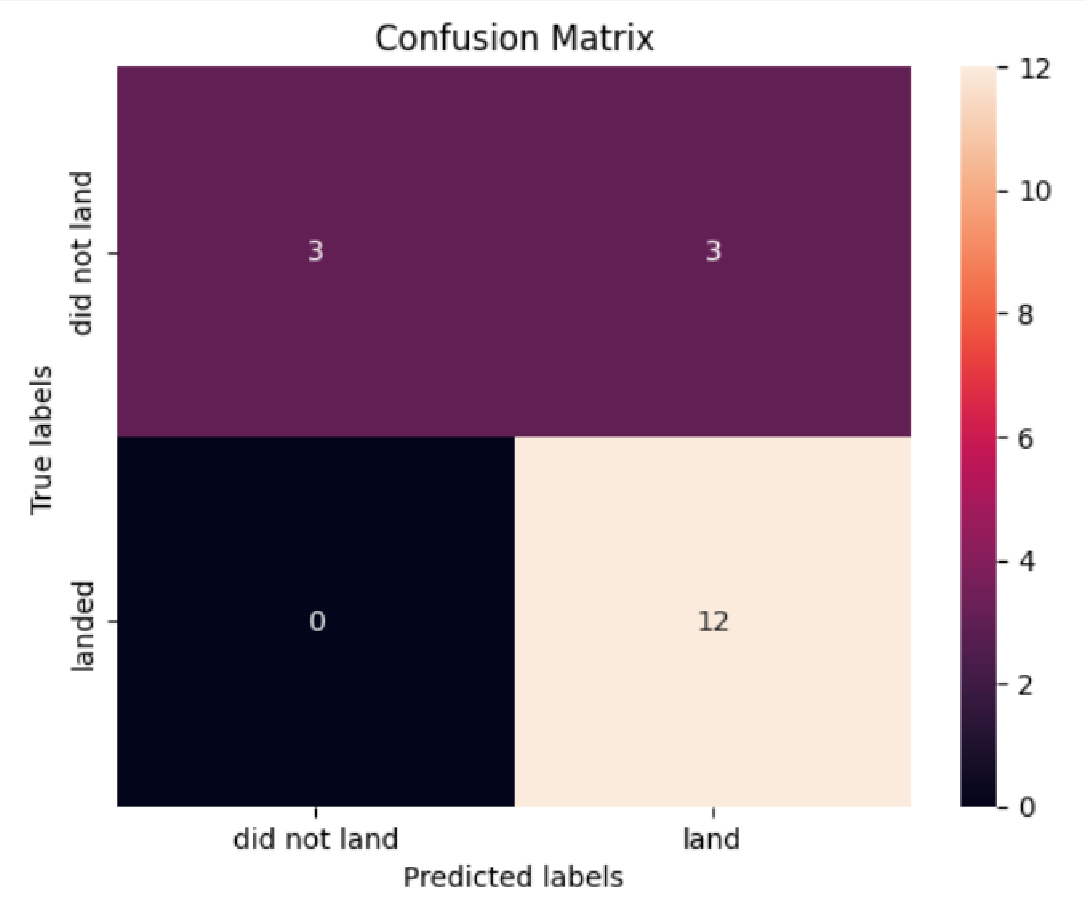
Section 5

Predictive Analysis (Classification)

Classification Accuracy

```
tuned hpyerparameters :(best parameters)  {'criterion': 'gini', 'max_depth': 4, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 5,
'splitter': 'random'}
accuracy : 0.8892857142857142
```

Confusion Matrix



Conclusions

- The larger the flight amount at a launch site, the greater the success rate at a launch site.
- Launch success rate started to increase in 2013 till 2020.
- Orbits ES-L1, GEO, HEO, SSO, VLEO had the most success rate.
- KSC LC-39A had the most successful launches of any sites.
- The KNN classifier is the best machine learning algorithm for this task.

Thank you!