IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Polad Valiyev
29 April 2024

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Project background and context
- Problems you want to find answers

Section 1

Methodology

Methodology

Executive Summary

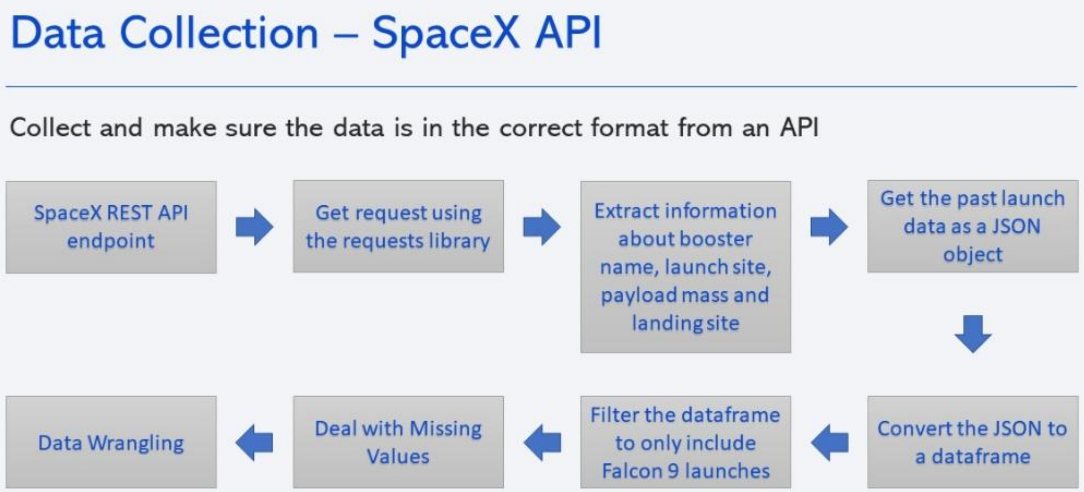
- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- Describe how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

- SpaceX API calls notebook

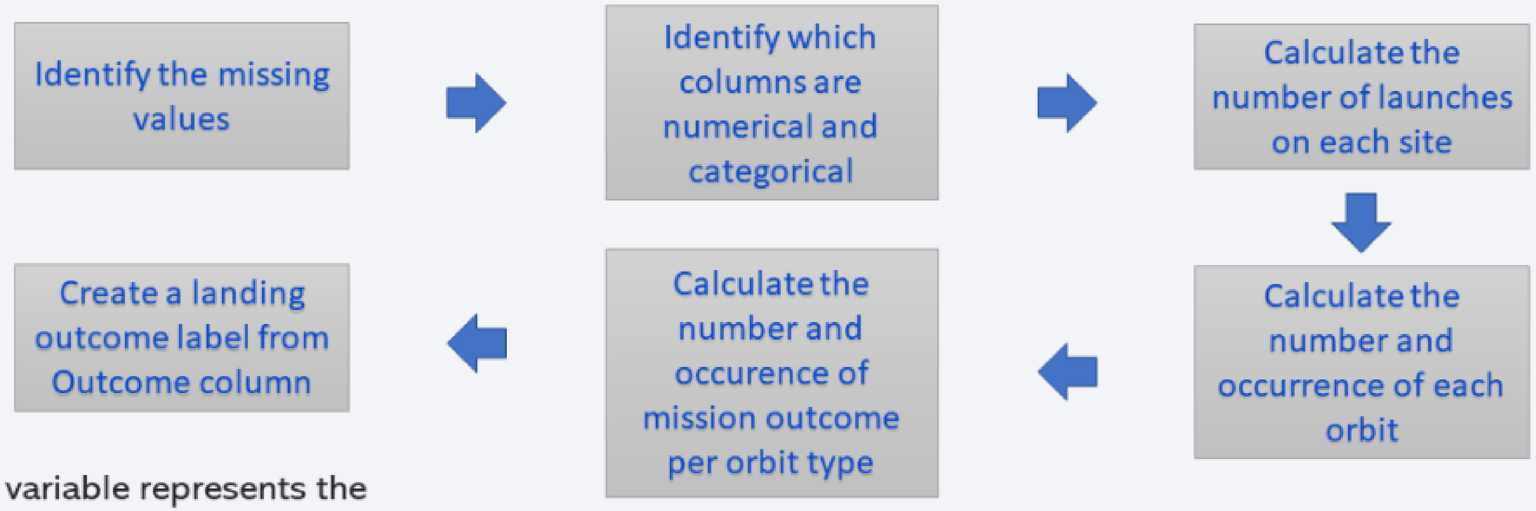


Data Collection - Scraping

- web scraping notebook

Data Wrangling

Perform Exploratory Data Analysis (EDA) to find patterns in the data and determine what would be the label for train supervised models



The variable represents the classification outcome of each launch. Zero means, the first stage did not land successfully; one means the first stage landed successfully.

[Data wrangling notebook]()

EDA with Data Visualization

Summary of charts that were plotted:

- Catplot to visualize the relationship between Flight Number and Payload.
- Catplot to visualize the relationship between Flight Number and Launch Site.
- Catplot to visualize the relationship between Payload and Launch Site.
- Bar chart to visualize the relationship between success rate of each Orbit type.
- Catplot to visualize the relationship between Flight Number and Orbit type.
- Catplot to visualize the relationship between Payload and Orbit type.
- Line chart to visualize the launch success yearly trend.

[EDA with Data Visualization]

EDA with SQL

SQL queries performed:

- Display the names of the unique launch sites in the space mission:
  SELECT DISTINCT(launch_site) FROM SPACEXTBL;

- Display 5 records where launch sites begin with the string 'CCA':
  *SELECT * FROM SPACEXTBL WHERE launch_site LIKE 'CCA%' LIMIT 5;*

- Display the total payload mass carried by boosters launched by NASA (CRS):
  SELECT SUM(payload_mass__kg_) AS TOTAL_PAYLOAD_MASS FROM SPACEXTBL WHERE customer='NASA (CRS)';

- Display average payload mass carried by booster version F9 v1.1:
  SELECT AVG(payload_mass__kg_) AS AVG_PAYLOAD_MASS FROM SPACEXTBL WHERE booster_version='F9 v1.1';

- List the date when the first successful landing outcome in ground pad was achieved:
  SELECT MIN(DATE) AS first_successful_landing FROM SPACEXTBL WHERE (landing_outcome)='Success (ground pad)';

EDA with SQL notebook

Build an Interactive Map with Folium

Summary of map objects that were created and added to the Folium map

- *folium.Circle* and *folium.Marker* to add a highlighted circle area with a text label on a specific coordinate for each launch site on the site map.
- *MarkerCluster* object for simplify a map containing many markers having the same coordinate.
- *MousePosition* on the map to get coordinate for a mouse over a point on the map.
- *folium.PolyLine* object to draw a line between a launch site to its closest city, railway and highway.

[Folium map]

Build a Dashboard with Plotly Dash

Summary of plots/graphs and interactions that were added to the dashboard to perform interactive visual analytics on SpaceX launch data in real-time.

This dashboard application contains input components such as a dropdown list and a range slider to interact with a pie chart and a scatter point chart.

- A launch Site Drop-down Input Component.
  There are four different launch sites and a dropdown menu let us select different launch sites.
- A callback function to render *success-pie-chart* based on selected site dropdown.
  The general idea of this callback function is to get the selected launch site from site-dropdown and render a pie chart visualizing launch success counts.
- A range Slider to Select Payload.
  The Slider is to be able to easily select different payload range and see if we can identify some visual patterns.
- A callback function to render the *success-payload-scatter-chart* scatter plot.
  To visually observe how payload may be correlated with mission outcomes for selected site(s).

Predictive Analysis (Classification)

Summary of the model development process used to predict if the first stage will land given the data from the preceding labs.

- Creation of a NumPy array from the column Class in data.
- Data standardization.
- Use of the function **train_test_split** to split the data X and Y into training and test data.
- Searching for the best Hyperparameters for Logistic Regression, SVM, Decision Tree and KNN classifiers.
- Searching for the method that performs best using test data.

Predictive analysis lab

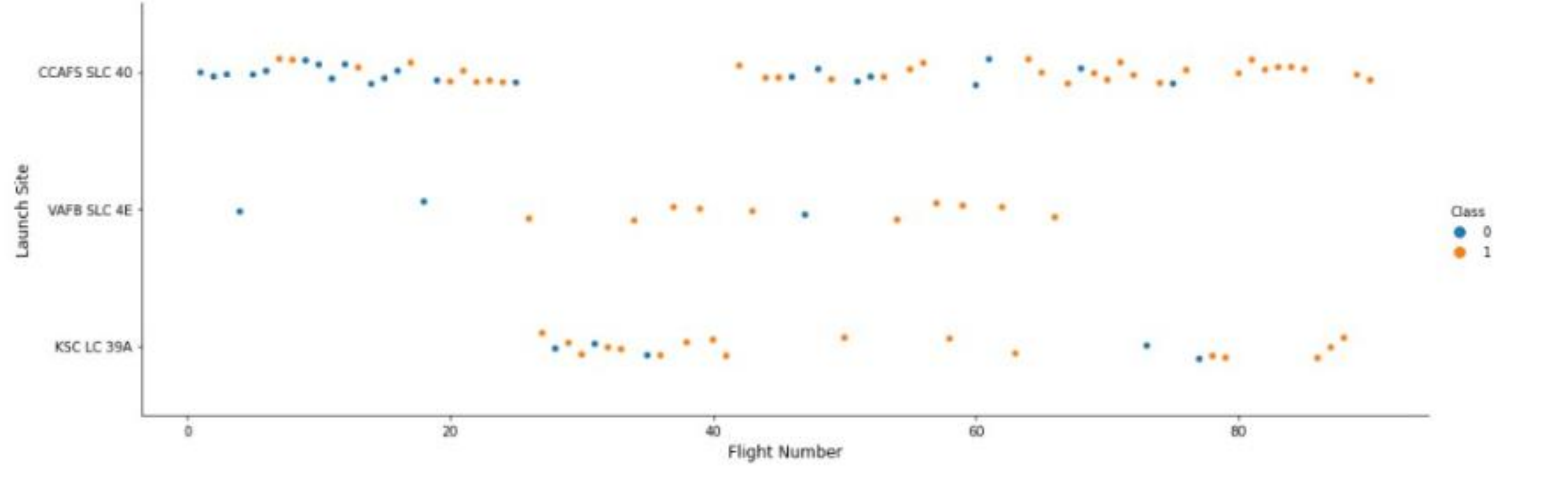
Results

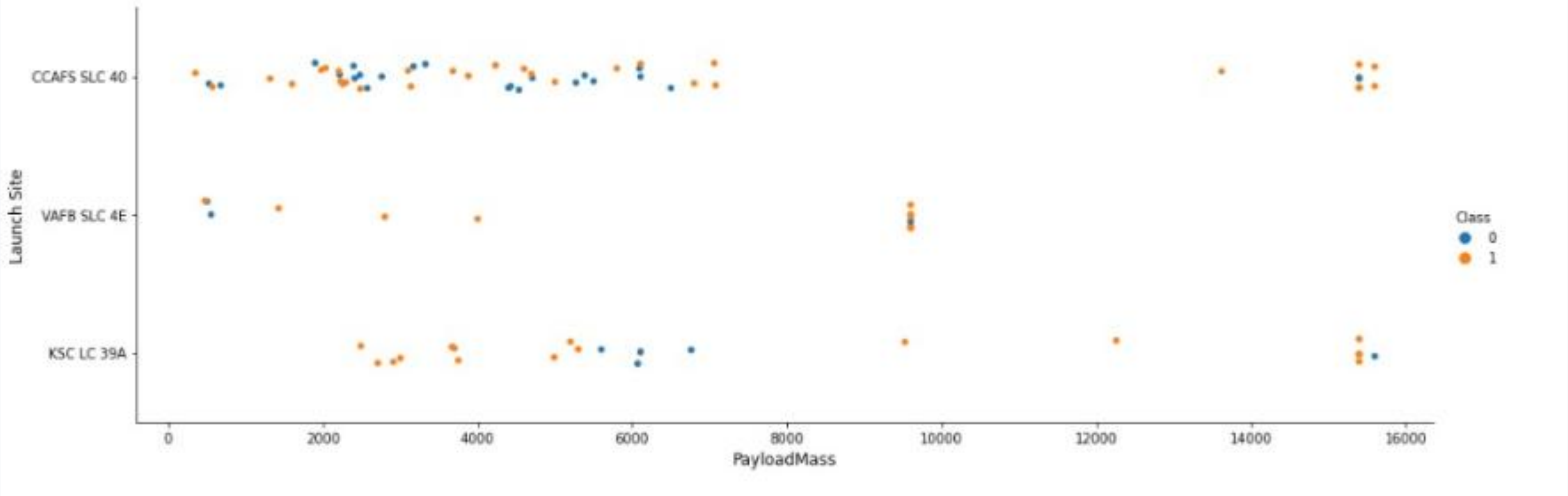
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

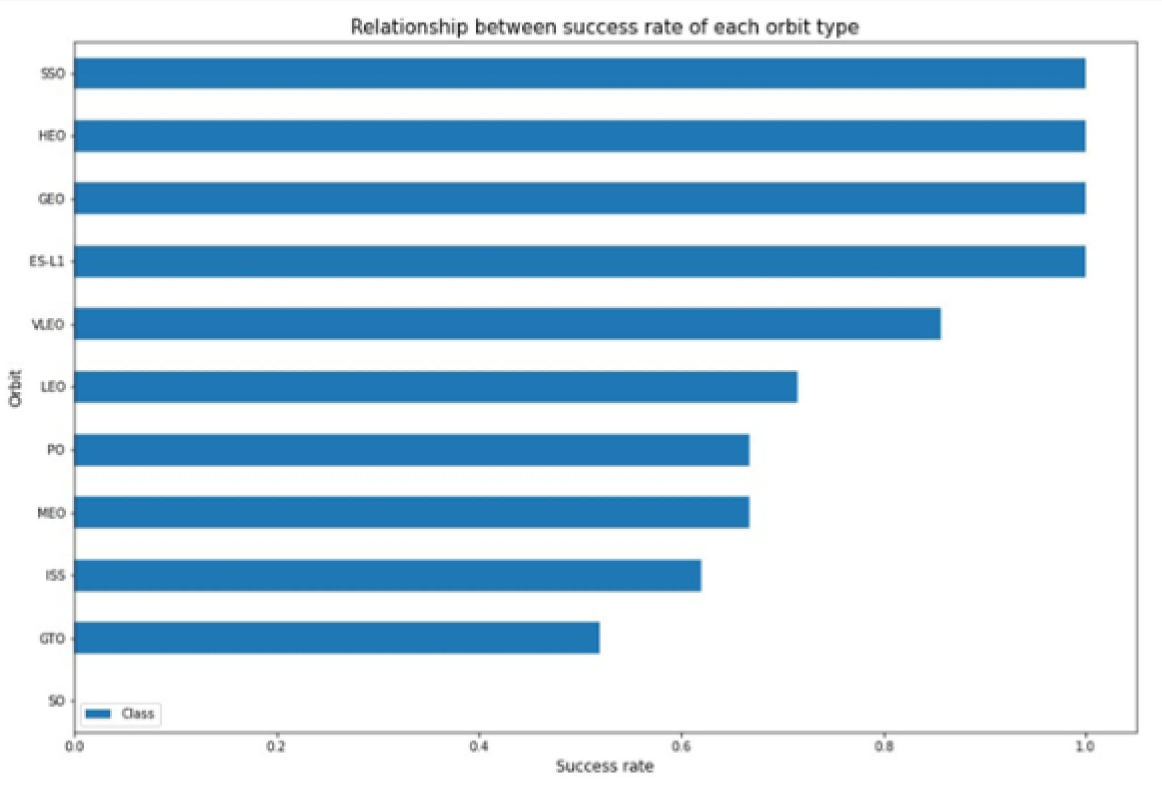
Insights drawn from EDA

Flight Number vs. Launch Site

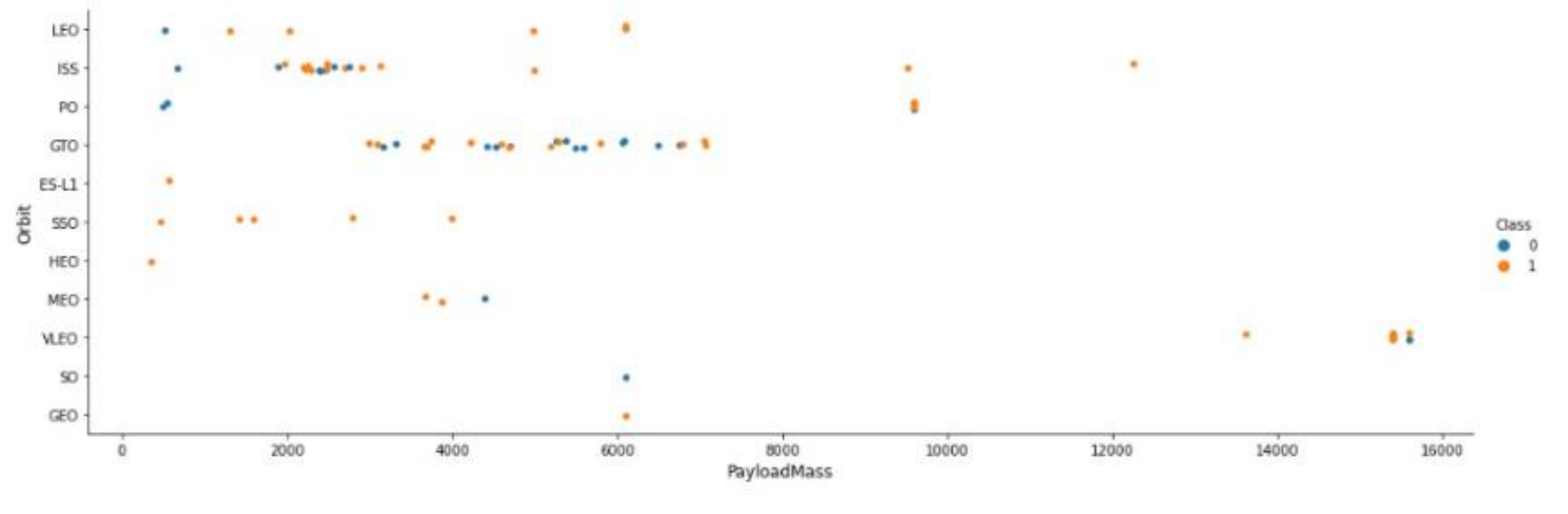




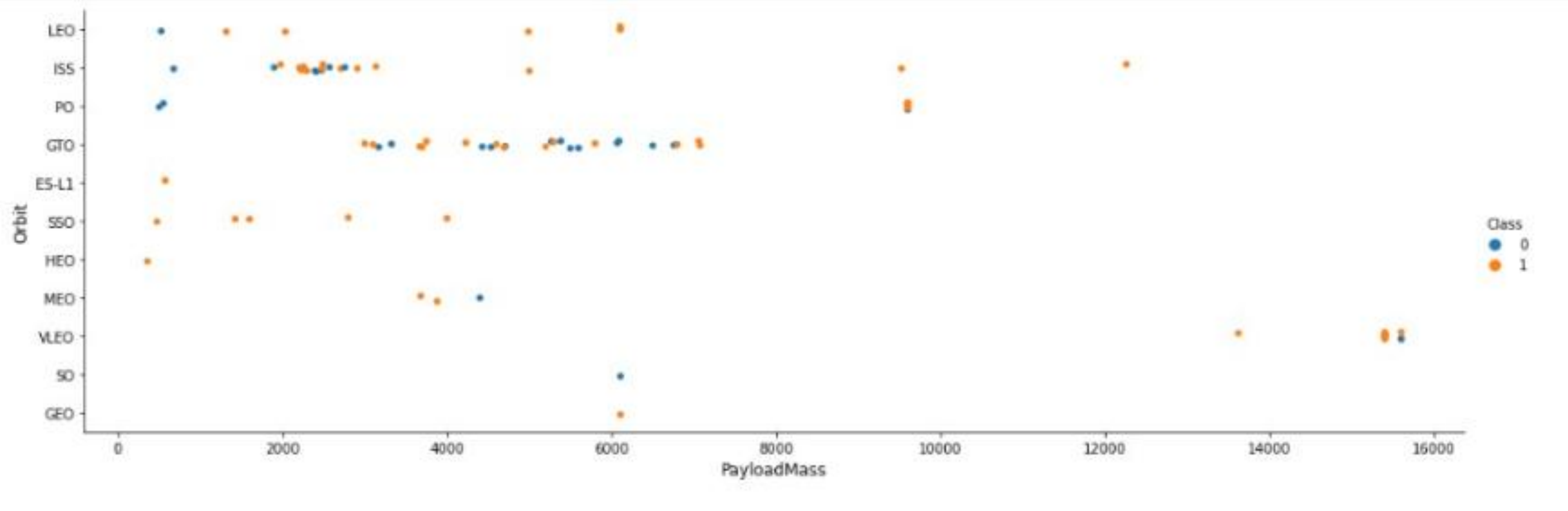
Payload vs. Launch Site



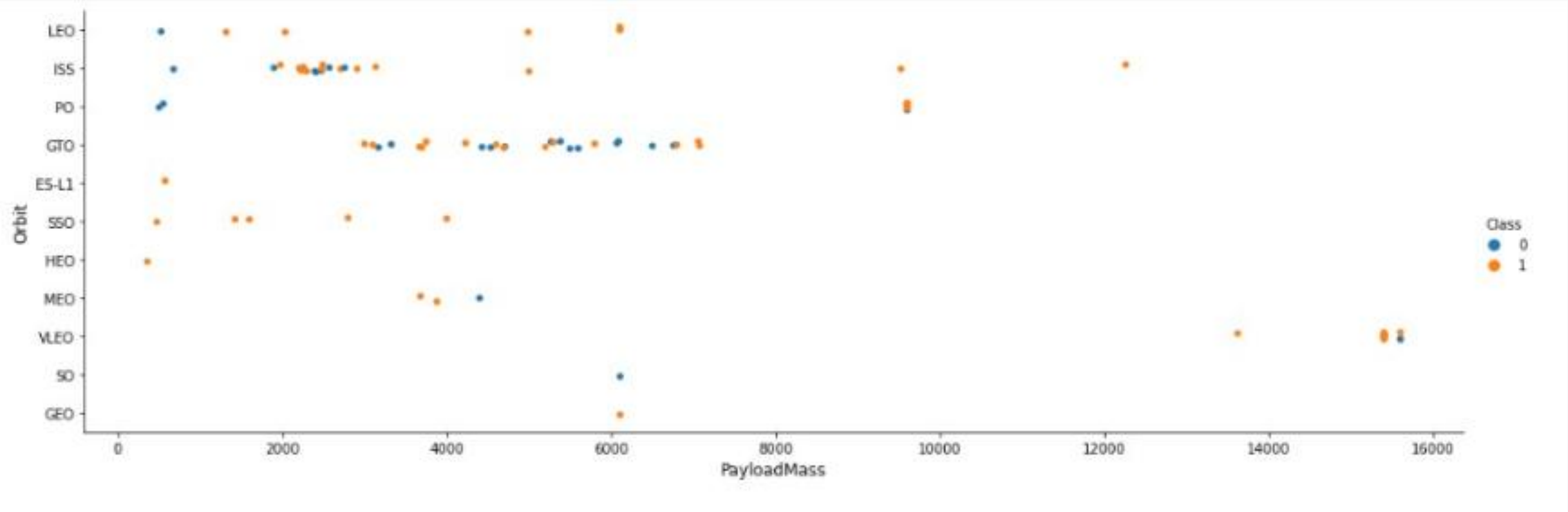
Success Rate vs. Orbit Type



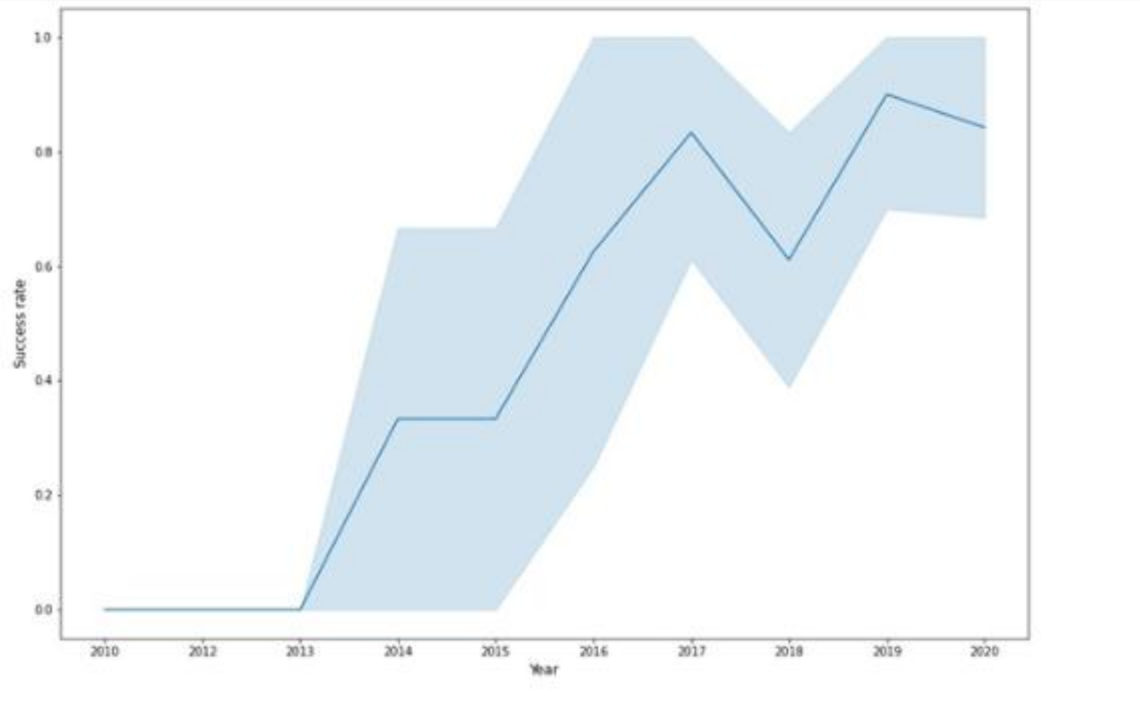
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

```
%sql SELECT DISTINCT(launch_site) FROM SPACEXTBL;
```

```
 * ibm_db_sa://ycy00214:***@3883e7e4-18f5-4afe-be8c-fa3
Done.
```

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

Launch Site Names Begin with 'CCA'

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

```
 * ibm_db_sa://ycy00214:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
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	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

```
%sql SELECT SUM(payload_mass__kg_) AS TOTAL_PAYLOAD_MASS FROM SPACEXTBL WHERE customer='NASA (CRS)';
```

```
 * ibm_db_sa://ycy00214:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdom
Done.
```

total_payload_mass
45596

Average Payload Mass by F9 v1.1

```
%sql select avg(PAYLOAD_MASS__KG_) as payloadmass from SPACEXTBL where Booster_Version = 'F9 v1.1';
```

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

payloadmass
2928.4

First Successful Ground Landing Date

```
%sql SELECT MIN(DATE) AS first_successful_landing FROM SPACEXTBL WHERE (landing_outcome)='Success (ground pad)';
```

```
 * ibm_db_sa://ycy00214:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:3
Done.
```

first_successful_landing
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql SELECT booster_version, payload_mass__kg_, landing_outcome FROM SPACEXTBL \
        WHERE landing_outcome='Success (drone ship)' AND (payload_mass__kg_ BETWEEN 4000 AND 6000) ;
```

```
 * ibm_db_sa://ycy00214:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdc
Done.
```

booster_version	payload_mass__kg_	landing_outcome
F9 FT B1022	4696	Success (drone ship)
F9 FT B1026	4600	Success (drone ship)
F9 FT B1021.2	5300	Success (drone ship)
F9 FT B1031.2	5200	Success (drone ship)

Total Number of Successful and Failure Mission Outcomes

```
%sql SELECT mission_outcome, COUNT(mission_outcome) AS TOTAL FROM SPACEXTBL GROUP BY mission_outcome;
```

```
 * ibm_db_sa://ycy00214:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdom
Done.
```

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

Boosters Carried Maximum Payload

```
%sql SELECT DISTINCT(booster_version), (SELECT MAX(payload_mass__kg_) AS "maximum_payload_mass" FROM SPACEXTBL) FROM SPACEXTBL
```

```
 * ibm_db_sa://ycy00214:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

booster_version	maximum_payload_mass
F9 B4 B1039.2	15600
F9 B4 B1040.2	15600
F9 B4 B1041.2	15600
F9 B4 B1043.2	15600
F9 B4 B1039.1	15600

2015 Launch Records

```
%sql SELECT landing_outcome, booster_version, launch_site, DATE FROM SPACEXTBL WHERE landing_outcome LIKE '%Failure (drone ship)%'
```

```
 * ibm_db_sa://ycy00214:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

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

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

```
%sql SELECT landing_outcome, COUNT(landing_outcome) AS "total" FROM SPACEXTBL WHERE (DATE BETWEEN '2010-06-04' AND '2017-03-20')
```

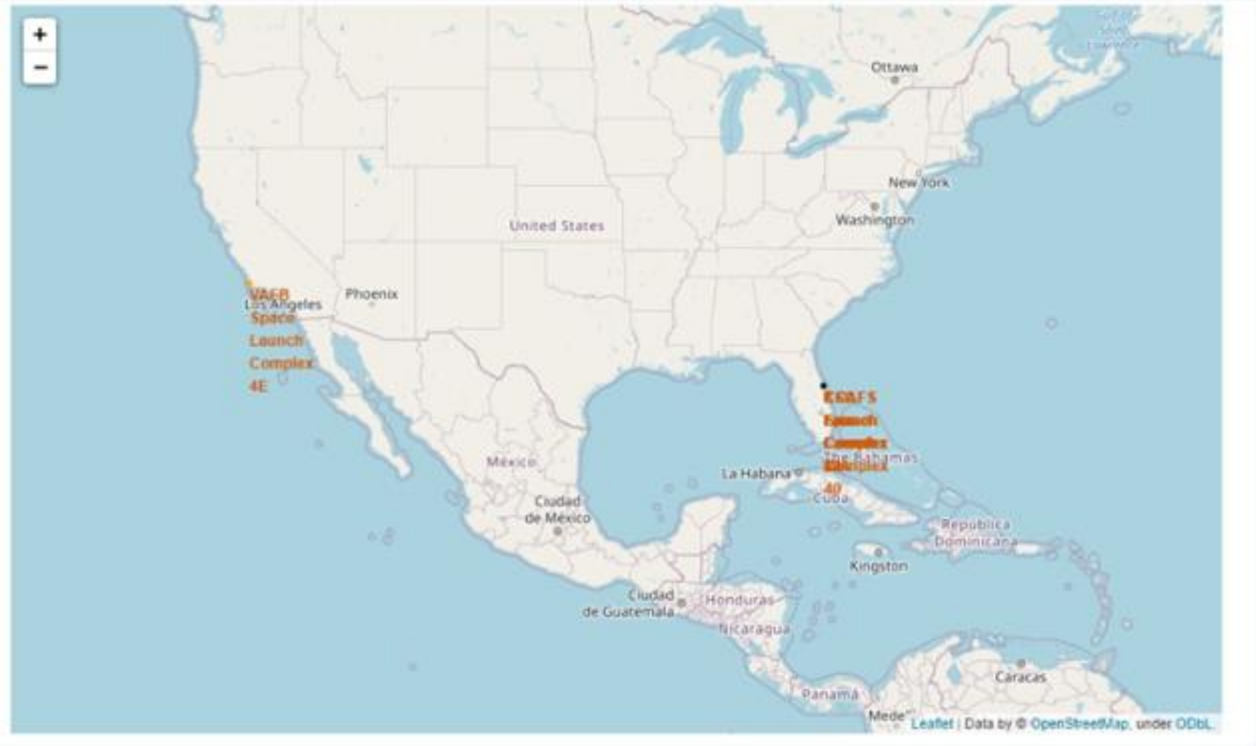
```
 * ibm_db_sa://ycy00214:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

landing_outcome	total
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

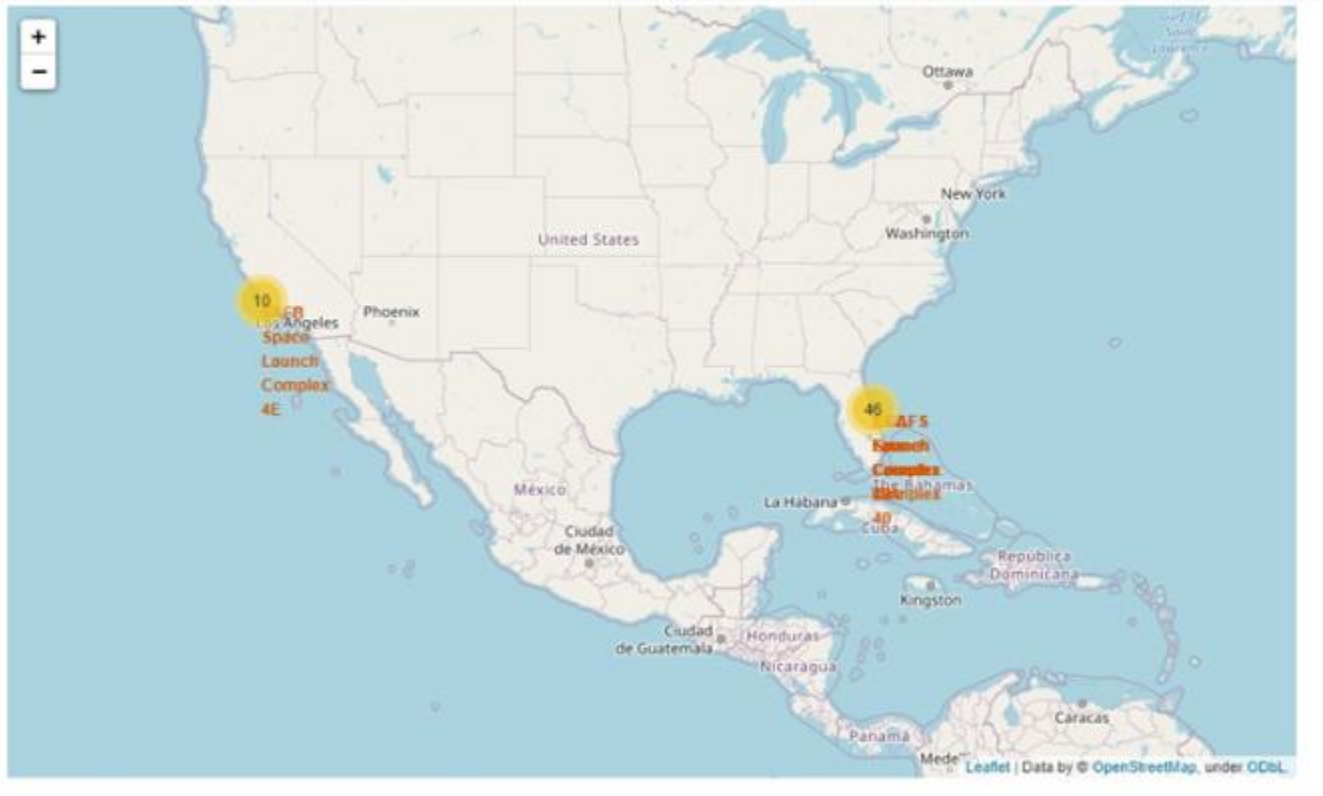
Section 3

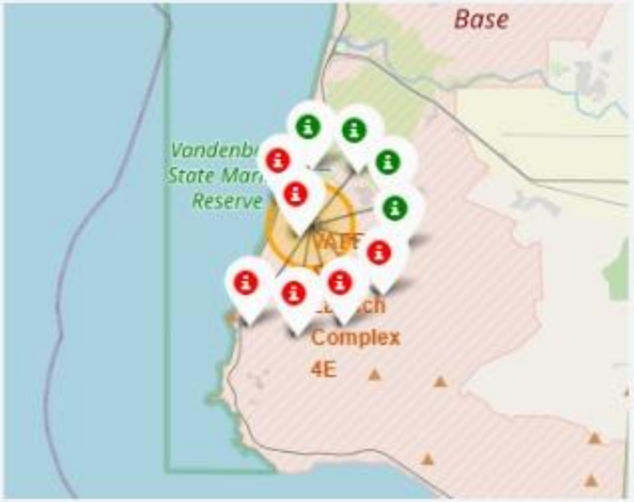
Launch Sites Proximities Analysis

<Folium Map Screenshot 1>



<Folium Map Screenshot 2>



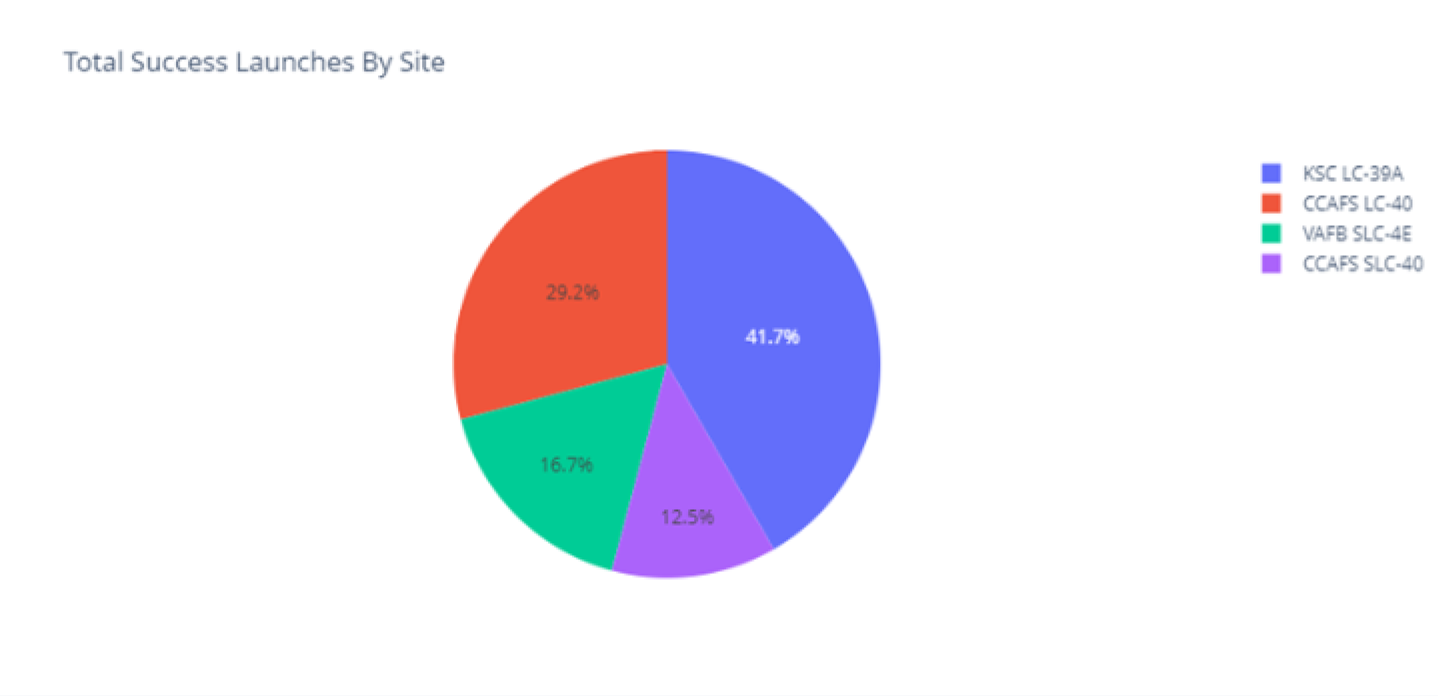


<Folium Map Screenshot 3>

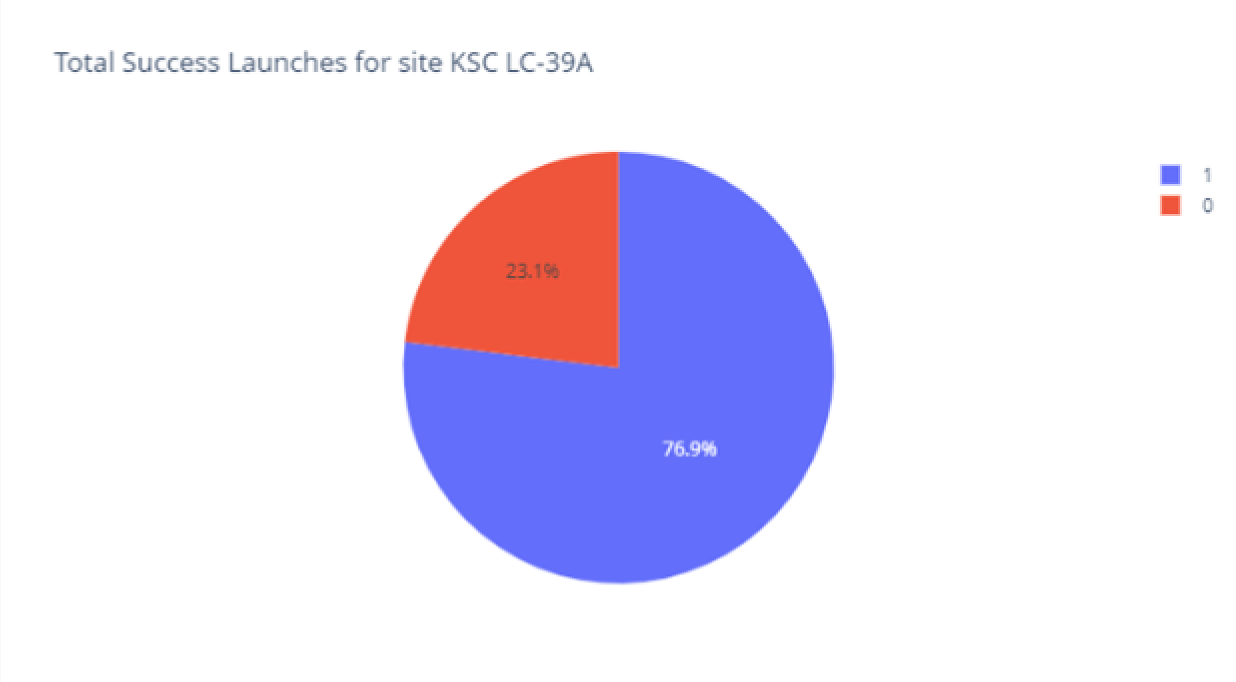
Section 4

Build a Dashboard with Plotly Dash

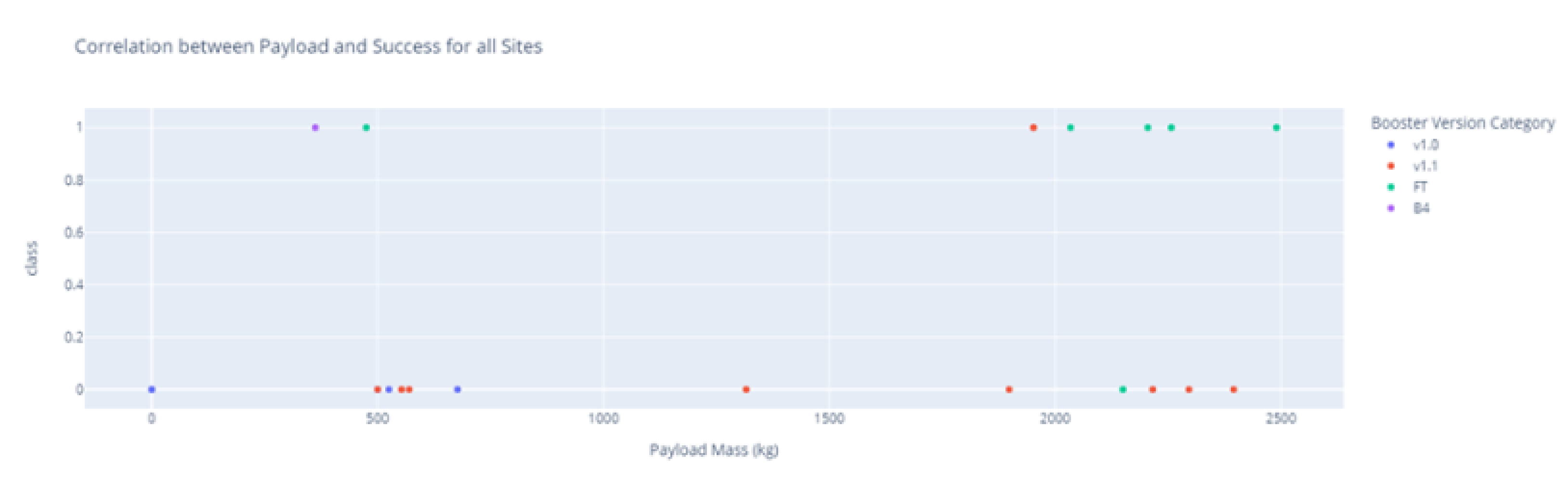
<Dashboard Screenshot 1>



<Dashboard Screenshot 2>



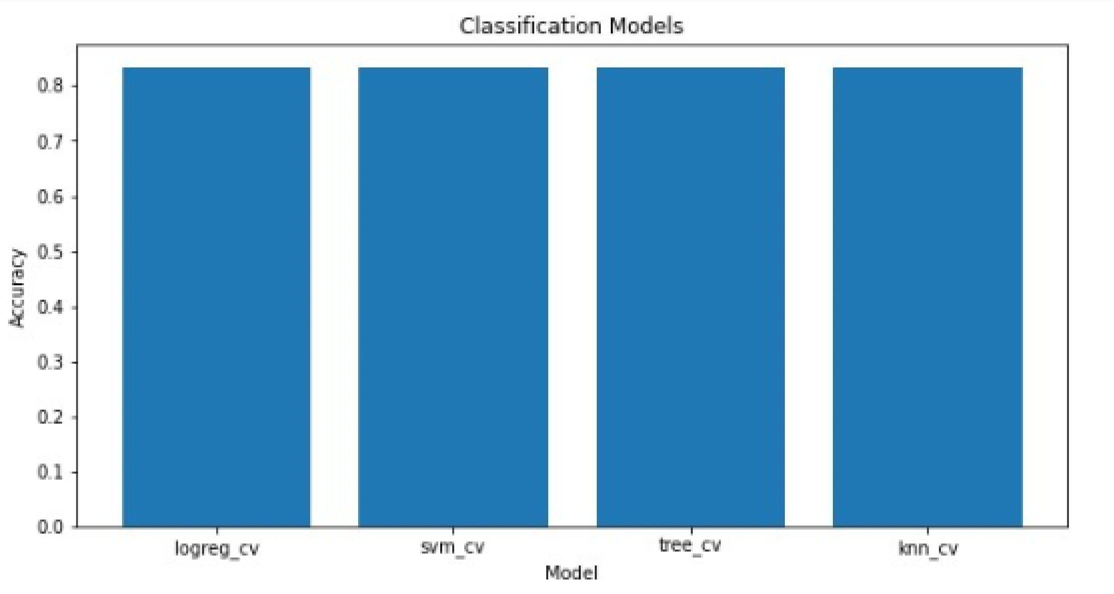
<Dashboard Screenshot 3>



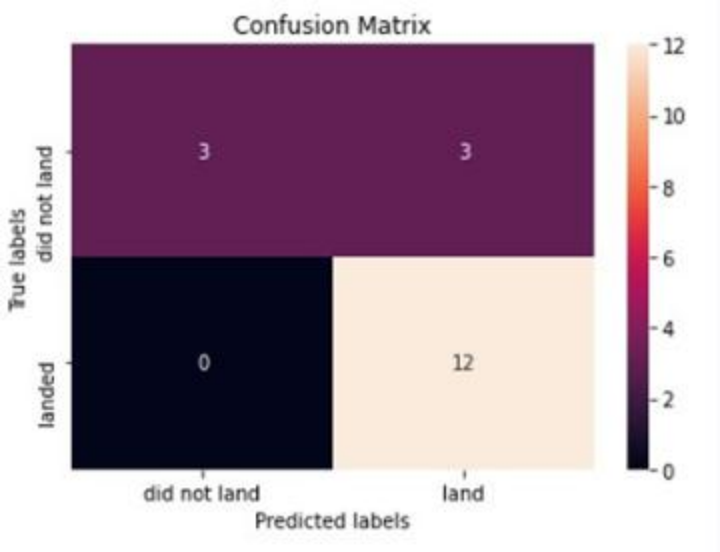
Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix



Conclusions

- As all the algorithms are giving the same accuracy, they all perform practically the same.
- By using our machine learning model, we can predict if the first stage of our competitor will land and determine the cost of a launch.

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!