IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Rubin Buddhacharya
11/11/2021

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
  - Building a Dashboard with Plotly Dash
  - Predictive Analysis (Classification)
- **Summary of all results**
  - Exploratory data analysis results
  - Interactive analytics demo in screenshots
  - Predictive analysis results

Introduction

- Project background and context

To predict if the Falcon 9 first stage will land successfully. 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. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

- Problems that needs answers

➢ Factors that affects the successful landing of rocket?

➢ Relationship between different variables that impacts the successful landing?

➢ Which methods perform the best using the data?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:

  The data was collected from SpaceX API and also did web scraping with BeautifulSoup

- 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

- Data sets collected.
- The data were collected from the SpaceX REST API which gave me data about launches, landing specifications and landing outcomes.
- The Objective is to predict whether SpaceX will land successfully or not.
- Also used web scraping to collect Falcon 9 launch data using Beautifulsoup

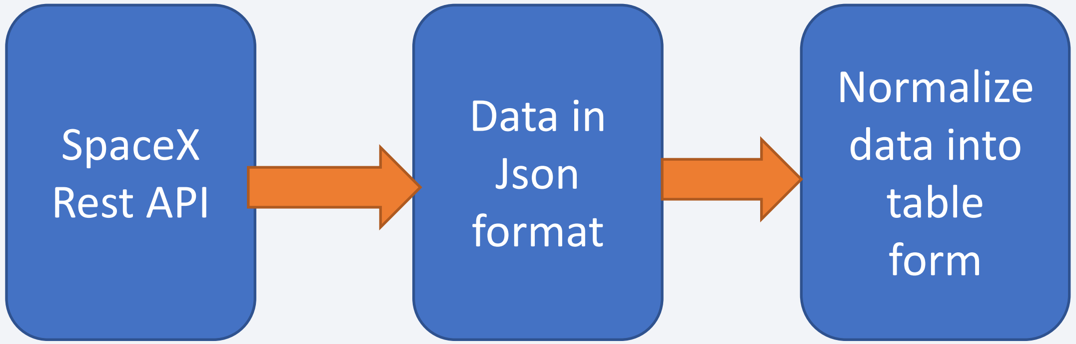
- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

- Data collection with SpaceX REST calls using key phrases and flowcharts

- GitHub URL of the completed SpaceX API calls notebook (must include completed code cell and outcome cell), as an external reference and peer-review purpose

[Github link for Data Collection]

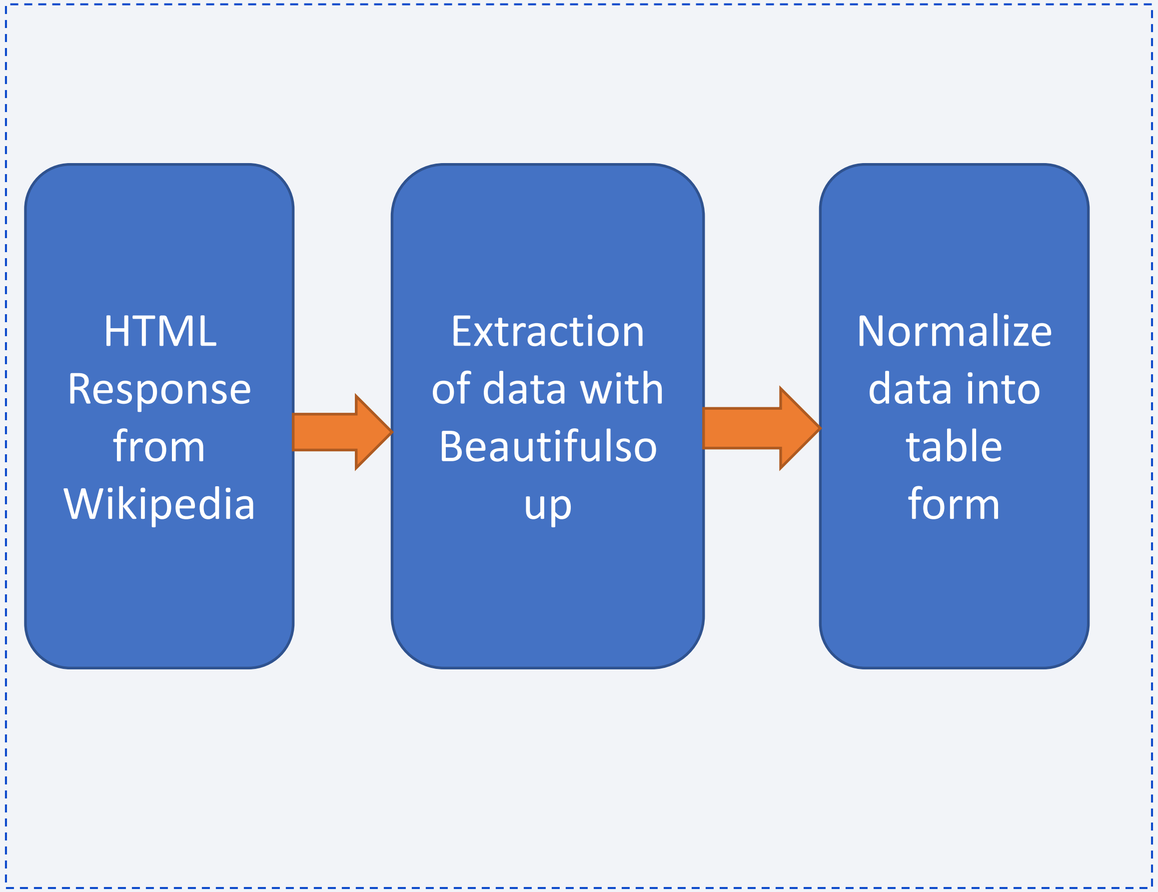


Data Collection - Scraping

- Web scraping process using key phrases and flowcharts

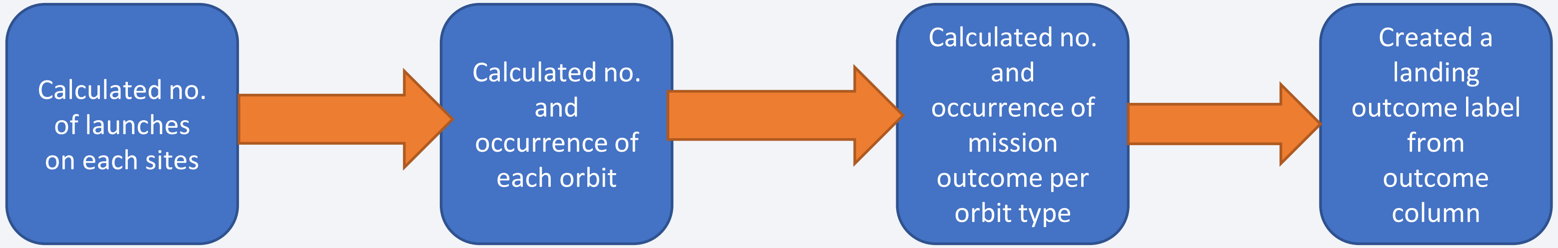
- GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose

- [Github link for Data Collection -Scraping]()



Data Wrangling

- We perform the Exploratory Data Analysis (EDA) to find some patterns in the data. We calculated the number of launches on each sites, calculated the number and occurrence of each orbit, also the mission outcome per orbit type and created a landing outcome label from outcome column.
- Data wrangling process using key phrases and flowcharts



- [Github Link]

EDA with Data Visualization

- Summary of charts and their uses:

  We plotted scatter plot for the following:

➢ FlightNumber vs PayloadMass
➢ FlightNumber vs LaunchSite
➢ PayloadMass vs LaunchSite
➢ Orbit vs FlightNumber
➢ PayloadMass vs Orbit
➢ bar chart for groupby Orbit column
➢ line graph for success rate vs year

- Github Link for EDA Visualization

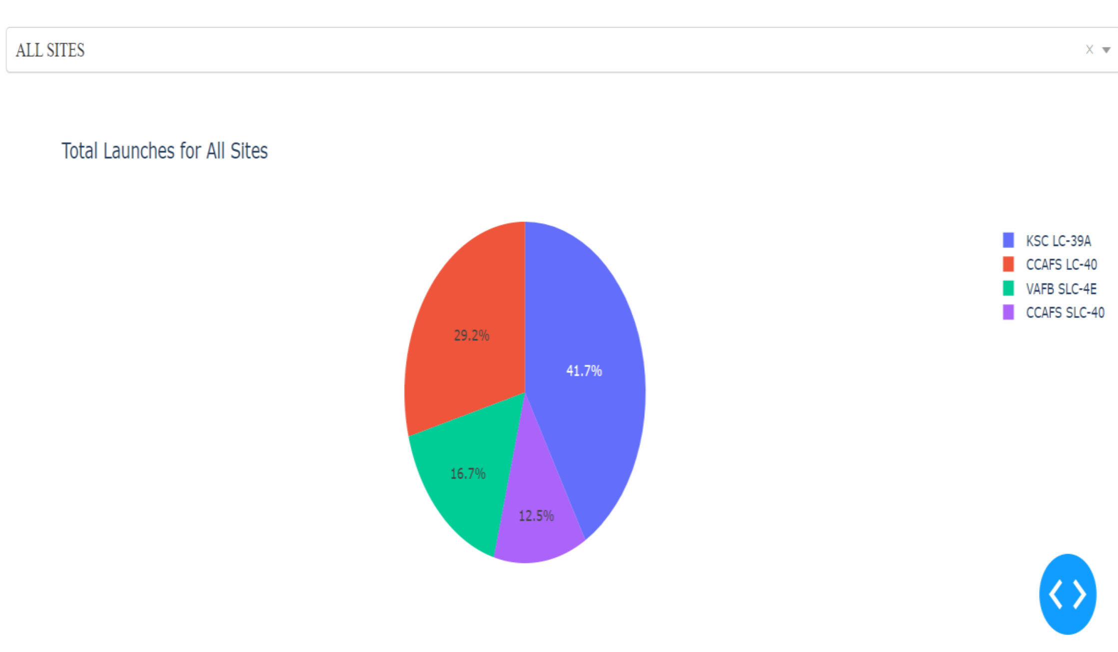
EDA with SQL

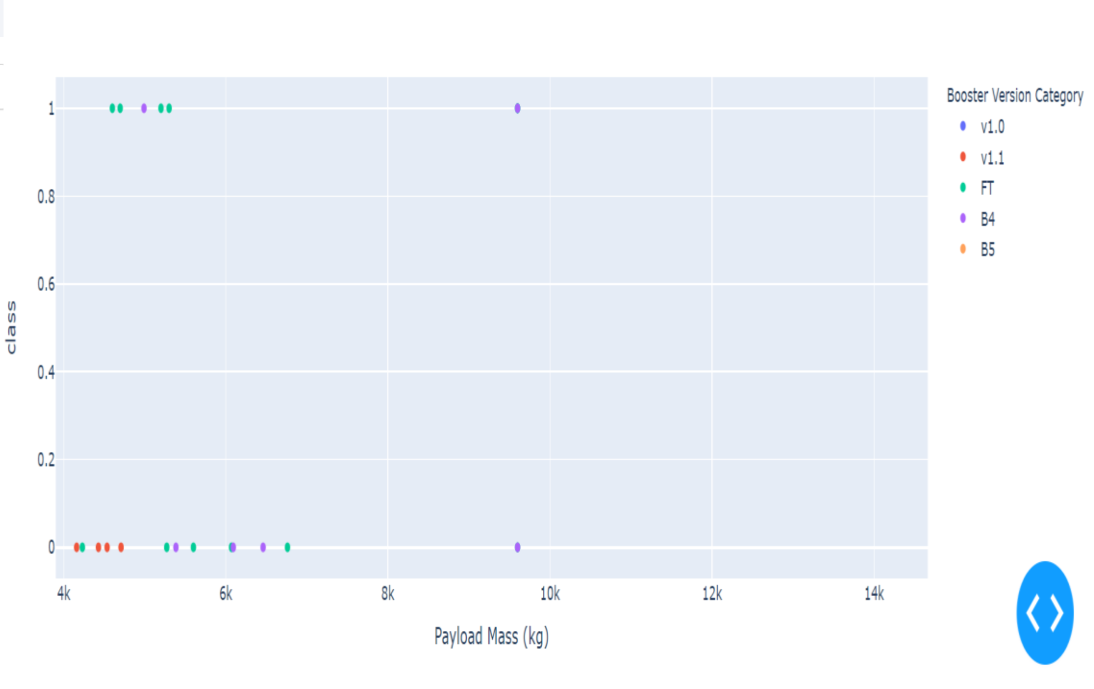
- Using bullet point format, summarize the SQL queries you performed
- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose
- GitHub link to EDA SQL

Build an Interactive Map with Folium

- **We used site's latitude and longitude coordinates to mark all the launch sites on the map, used circle to add a highlighted circle area with the text label on a specific coordinate.**
- **We also marked the success and failed launches for each site on the map based on the class and gave red for failed and green for successful.**
- **We also calculated the distance between the launch sites to other landmarks to see the trend.**
- **Github Link for Folium**

Build a Dashboard with Plotly Dash

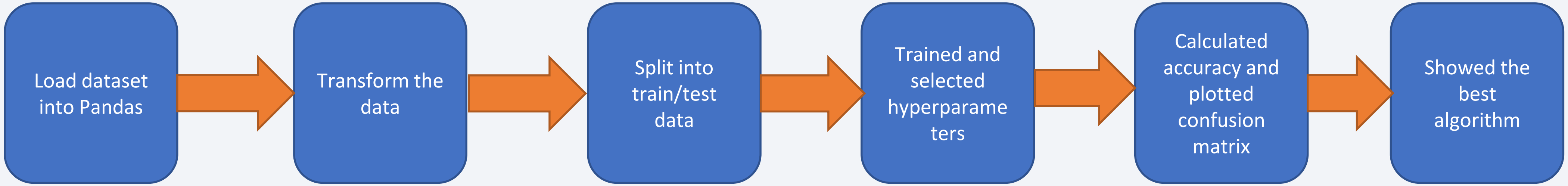




- Added a bar graph for all successful launch sites and a scatter plot for all launch sites showing which sites have higher percentage of success and what payload mass has higher success.
- GitHub URL

Predictive Analysis (Classification)

- Loaded the dataset into pandas and numpy
- Transform the data
- Split the data into training and testing data
- Models are trained and hyperparameters are selected using GridSearchCV function
- To Evaluate the model, we calculated the accuracy of each model
- Plotted the confusion matrix
- Displayed the best algorithms



- [Github link for Predictive Analysis]

Results

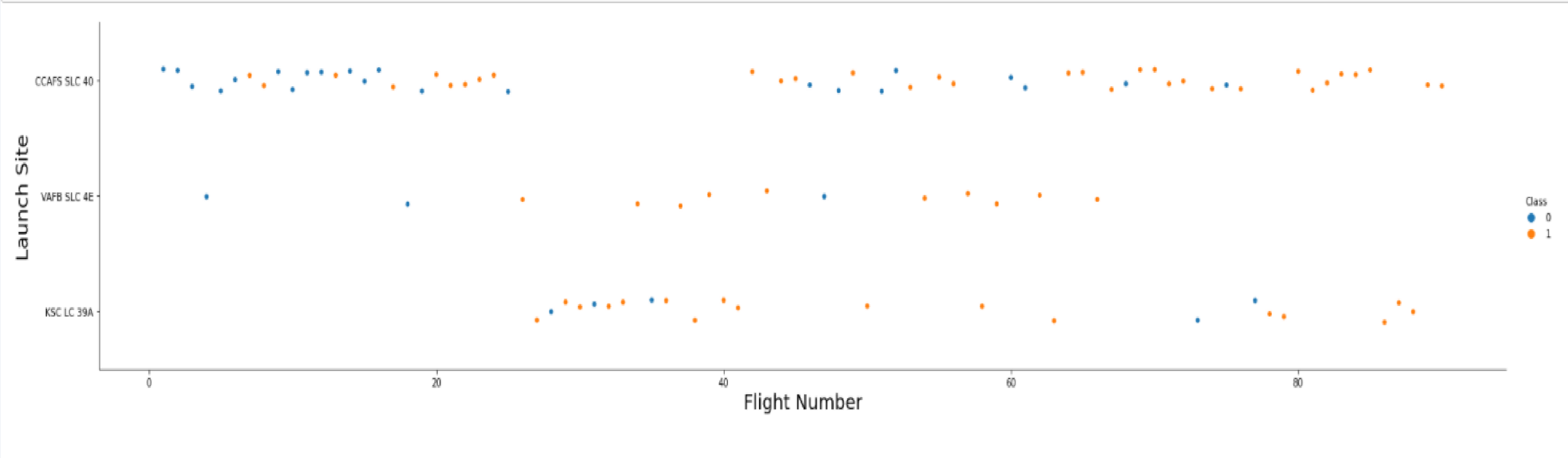
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

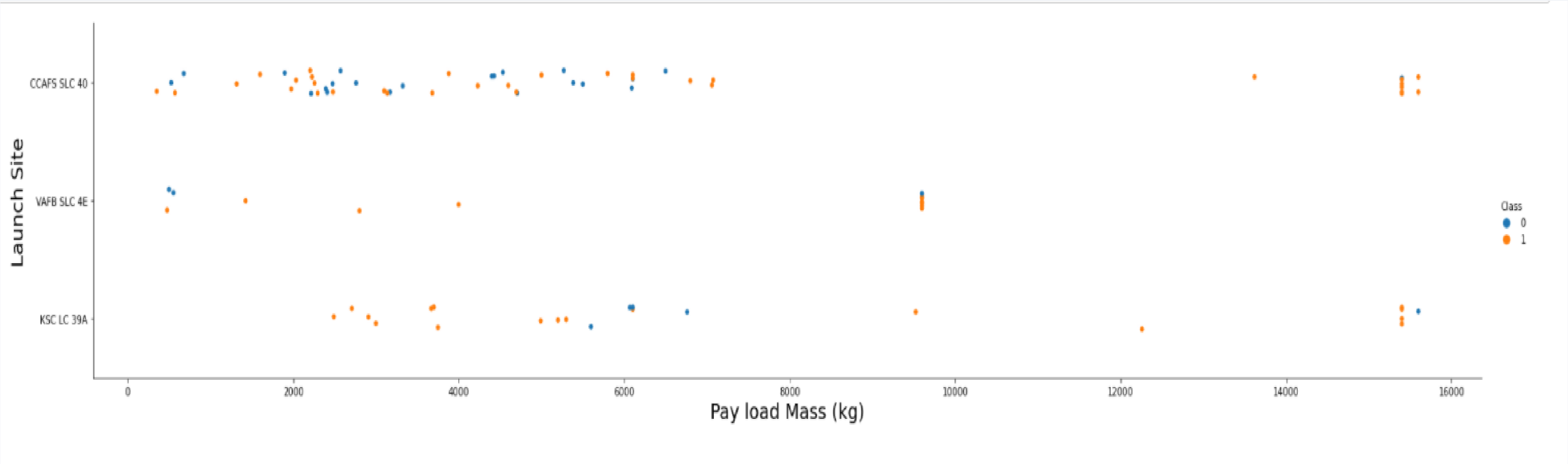
- Scatter plot of Flight Number vs. Launch Site



- The more the number of flights at launch site, greater the success rate.

Payload vs. Launch Site

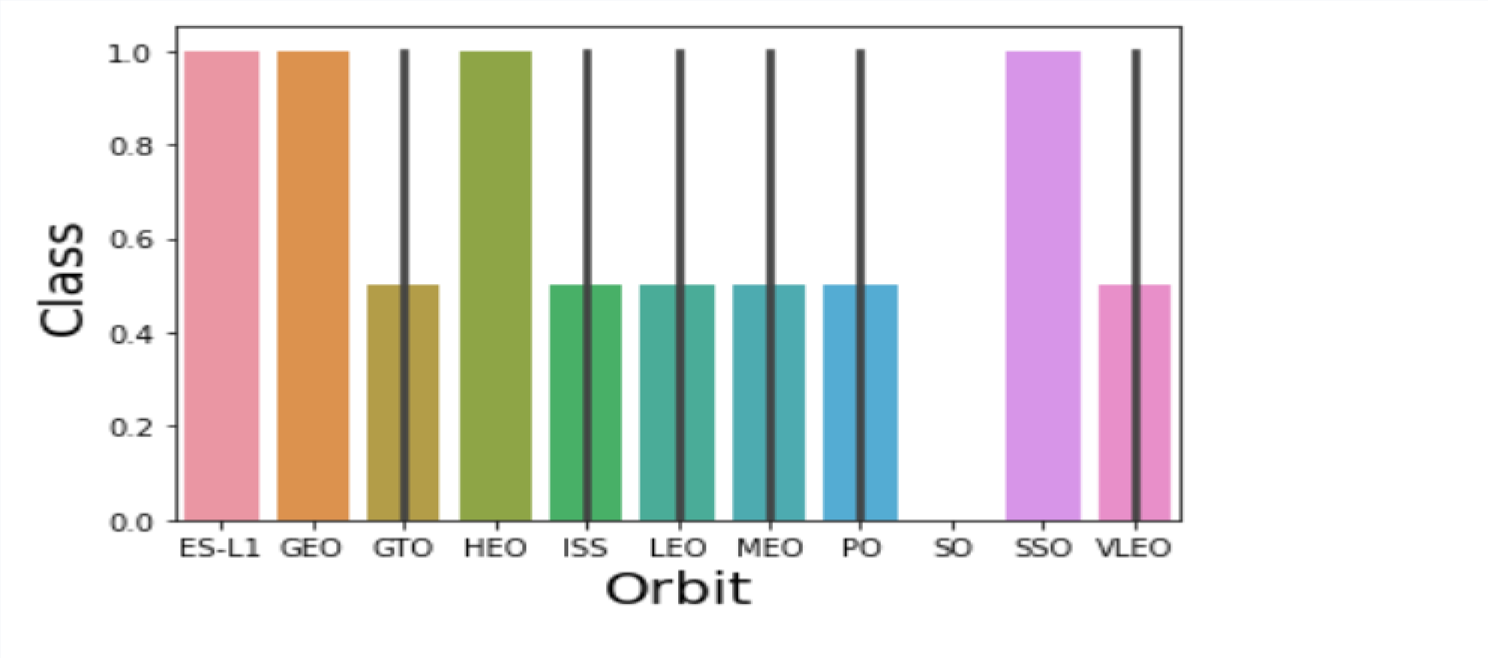
- Scatter plot of Payload vs. Launch Site



- The more the payload mass, the more the success rate for launch sites.

Success Rate vs. Orbit Type

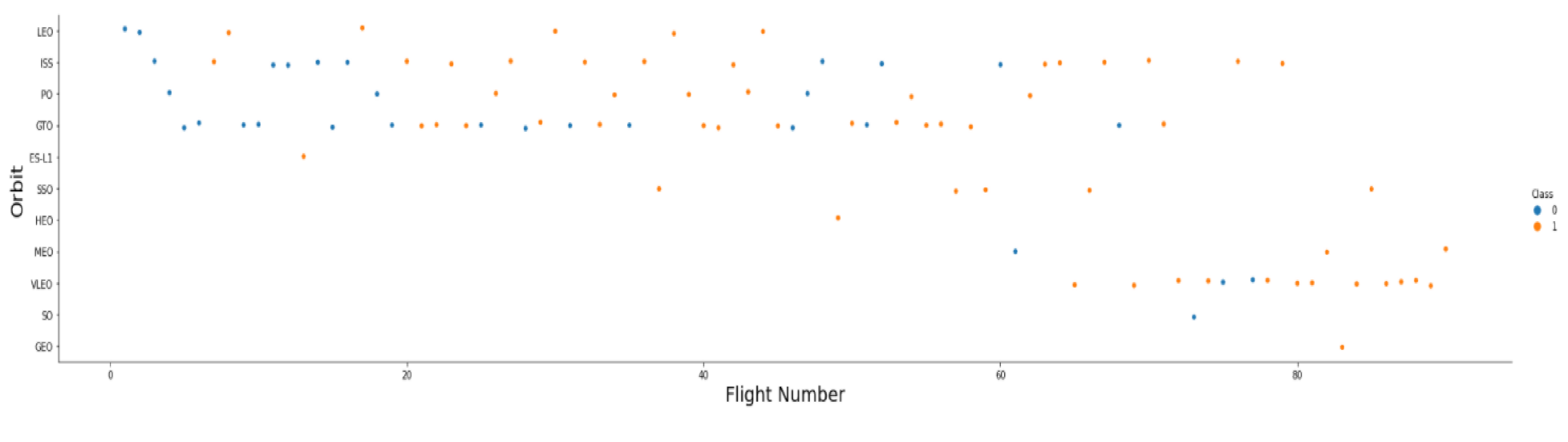
- A bar chart for the success rate of each orbit type



- The orbits ES-L1, GEO,HEO,SSO have higher success rate.

Flight Number vs. Orbit Type

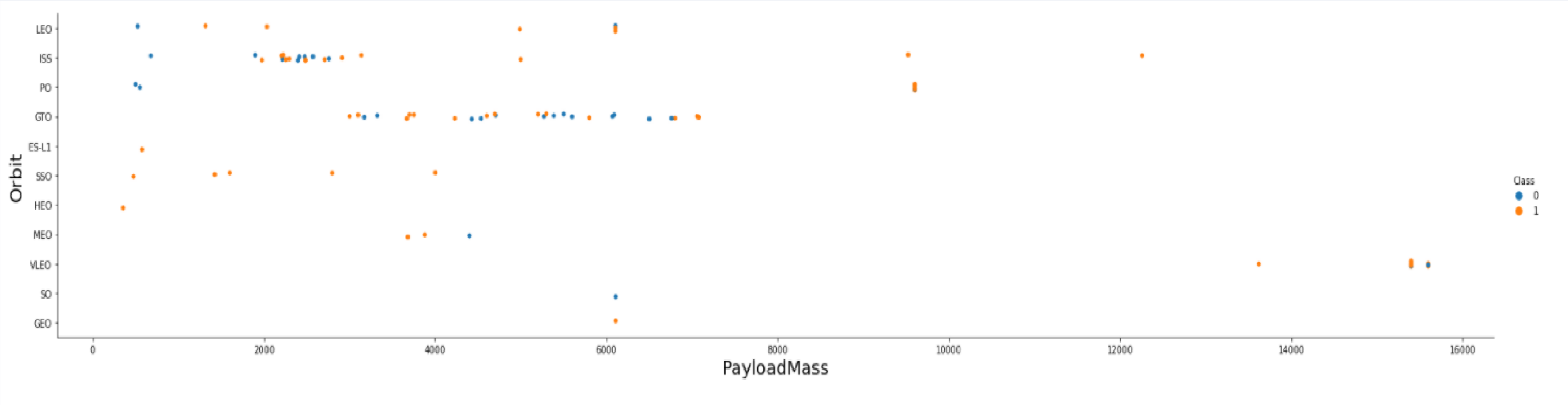
- Scatter point of Flight number vs. Orbit type



- **The LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.**

Payload vs. Orbit Type

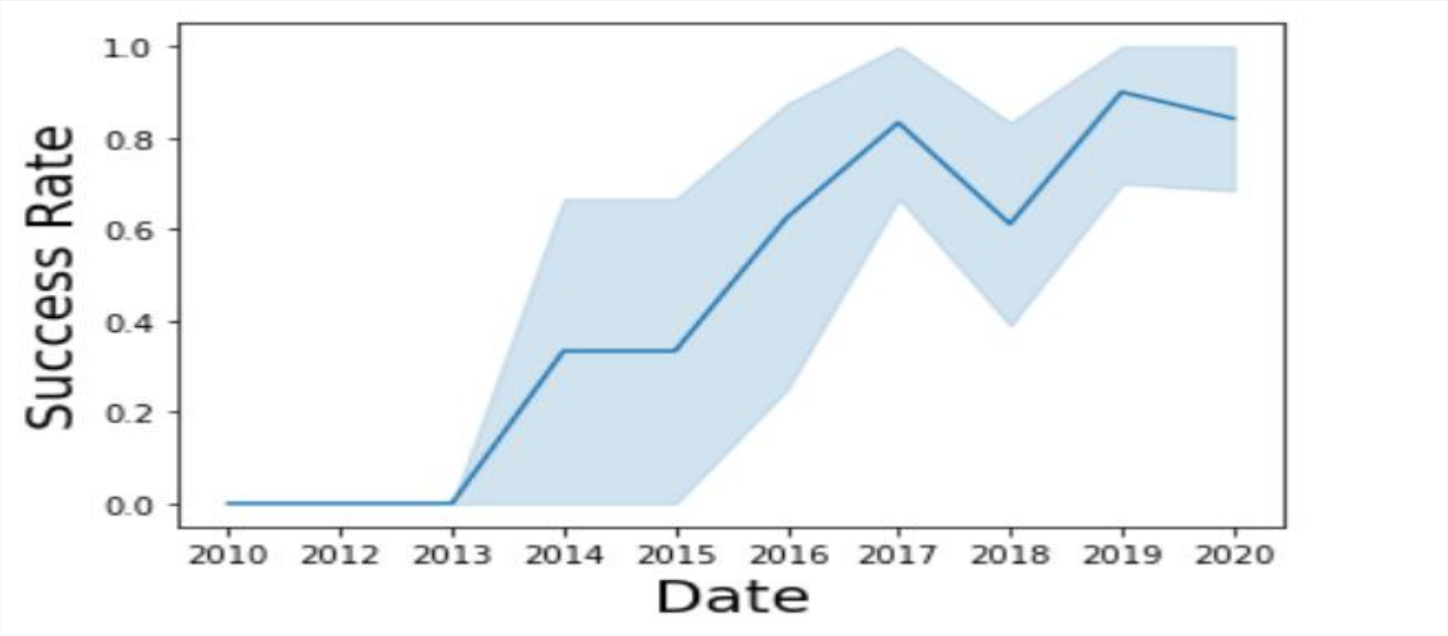
- Scatter point of payload vs. orbit type



- With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

Launch Success Yearly Trend

- A line chart of yearly average success rate



- The success rate kept on increasing from 2013 until 2020.

All Launch Site Names

- Name of the Unique Launch Sites
  - CCAFS LC-40
  - CCAFS SLC-40
  - KSC LC-39A
  - VAFB SLC-4E

- The query distinct will only show the unique values in the launch site.

```
%sql select distinct(LAUNCH_SITE) from SPACEXTBL

 * ibm_db_sa://ksy22772:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/bludb
Done.
```

Out[7]:

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

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with `CCA`

```
%sql select * from SPACEXTBL where launch_site like 'CCA%' limit 5
```

```
 * ibm_db_sa://ksy22772:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/bludb
Done.
```

Out[8]:

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

- The query with like and % after the word 'CCA' will give any value that start with 'CCA' and limit 5 give 5 records.

Total Payload Mass

- The total payload carried by boosters from NASA is 45596

```
%sql select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where CUSTOMER = 'NASA (CRS)'

 * ibm_db_sa://ksy22772:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/bludb
Done.

ut[23]:     1
        45596
```

- The query sum function returns the sum of the total value and where clause is used to filter the customer named ‘NASA (CRS)’

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is 2928

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

```
%sql select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where booster_version = 'F9 v1.1'
```

```
 * ibm_db_sa://ksy22772:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/bludb
Done.
```

Out[25]:

1
2928

- The query average function returns the average value of payload mass kg for booster version F9 v1.1

First Successful Ground Landing Date

- The dates of the first successful landing outcome on ground pad is December 22nd, 2015.
- The query min returns the smallest value of the selected column.

List the date when the first successful landing outcome in ground pad was acheived.

Hint:Use min function

```
]: %sql select min(Date) from SPACEXTBL where landing__outcome = 'Success (ground pad)'
```

```
 * ibm_db_sa://ksy22772:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/bludb
Done.
```

Out[27]:

1
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- The names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 are:
  F9 FT B1022
  F9 FT B1026
  F9 FT B1021.2
  F9 FT B1031.2
- ***List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000***

```
]: %sql select booster_version from SPACEXTBL where landing__outcome = 'Success (drone ship)' and payload_mass__kg_ > 4000 and payload_mass__kg_
```

```
 * ibm_db_sa://ksy22772:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/bludb
Done.
```

Out[28]:

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

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes
- The query ‘count’ counts the mission outcome for success or failure.

List the total number of successful and failure mission outcomes

```
%sql select count(mission_outcome) from SPACEXTBL where mission_outcome = 'Success' or mission_outcome = 'Failure (in flight)'
```

```
 * ibm_db_sa://ksy22772:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/bludb
Done.
```

ut[29]:

1
100

Boosters Carried Maximum Payload

- Below are the names of the booster which have carried the maximum payload mass
- Used select inside a select query to get the list of names of the booster.

```
]: %sql select booster_version from SPACEXTBL where payload_mass__kg_ = (select max(payload_mass__kg_) from SPACEXTBL)
 * ibm_db_sa://ksy22772:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/bludb
Done.
```

Out[30]:

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

- The failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

booster version	launch site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

```
]: %sql select landing__outcome, booster_version, launch_site from SPACEXTBL where Date like '2015%' AND landing__outcome = 'Failure (drone shi
```

```
 * ibm_db_sa://ksy22772:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/bludb
Done.
```

Out[42]:

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

```
]: %sql select * from SPACEXTBL where landing__outcome like 'Success%' and (DATE between '2010-06-04' and '2017-03-20') order by date desc

 * ibm_db_sa://ksy22772:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/bludb
Done.
```

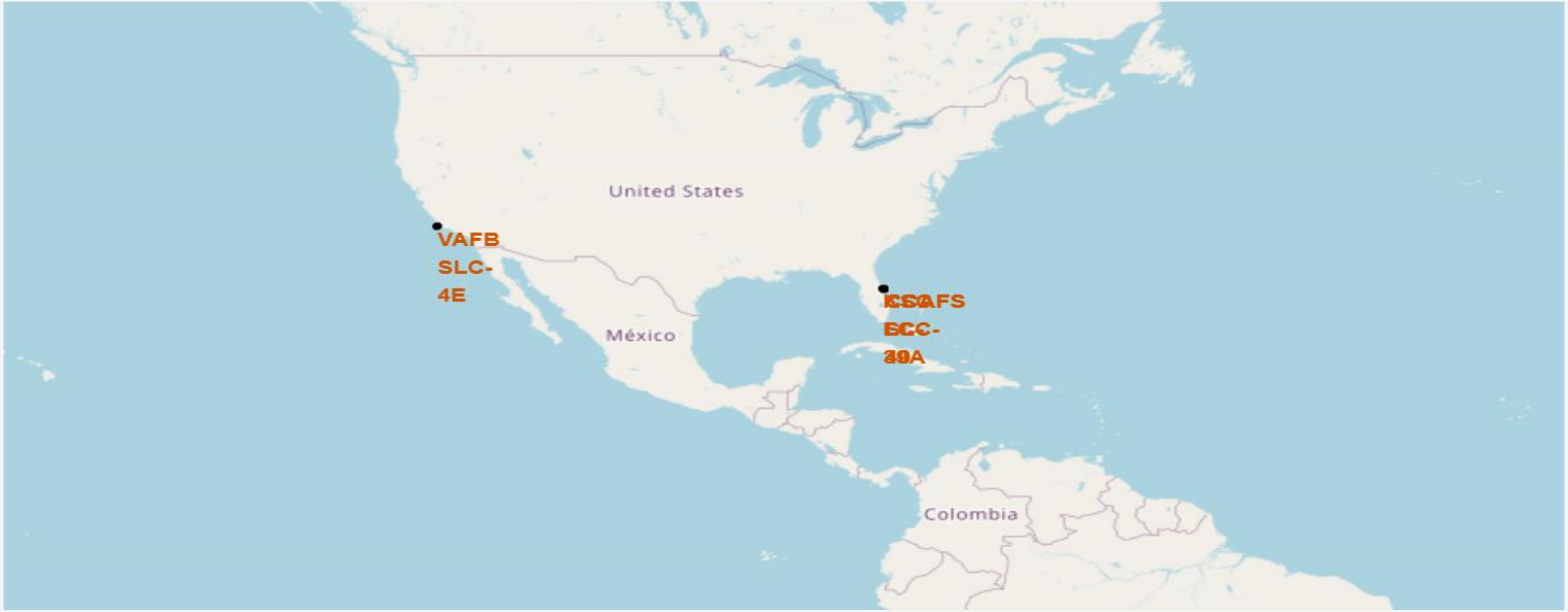
Out[37]:

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	05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-07-18	04:45:00	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	05:21:00	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	01:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

Section 4

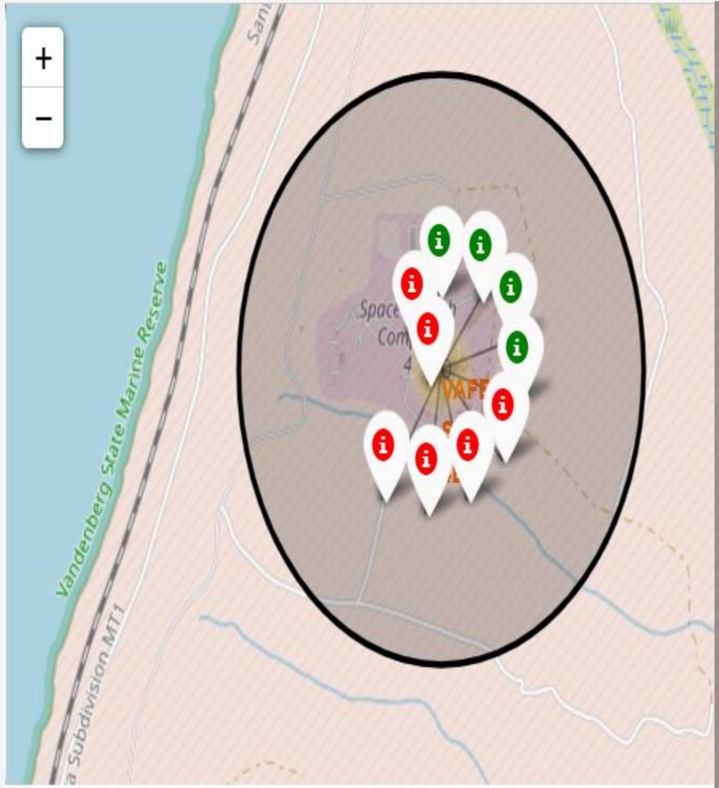
Launch Sites Proximities Analysis

All Launch Site on Global Map



The SpaceX launch sites are all in Unites States of American with sites being in California and Florida.

Success/Failed Colored Marker Launch Sites

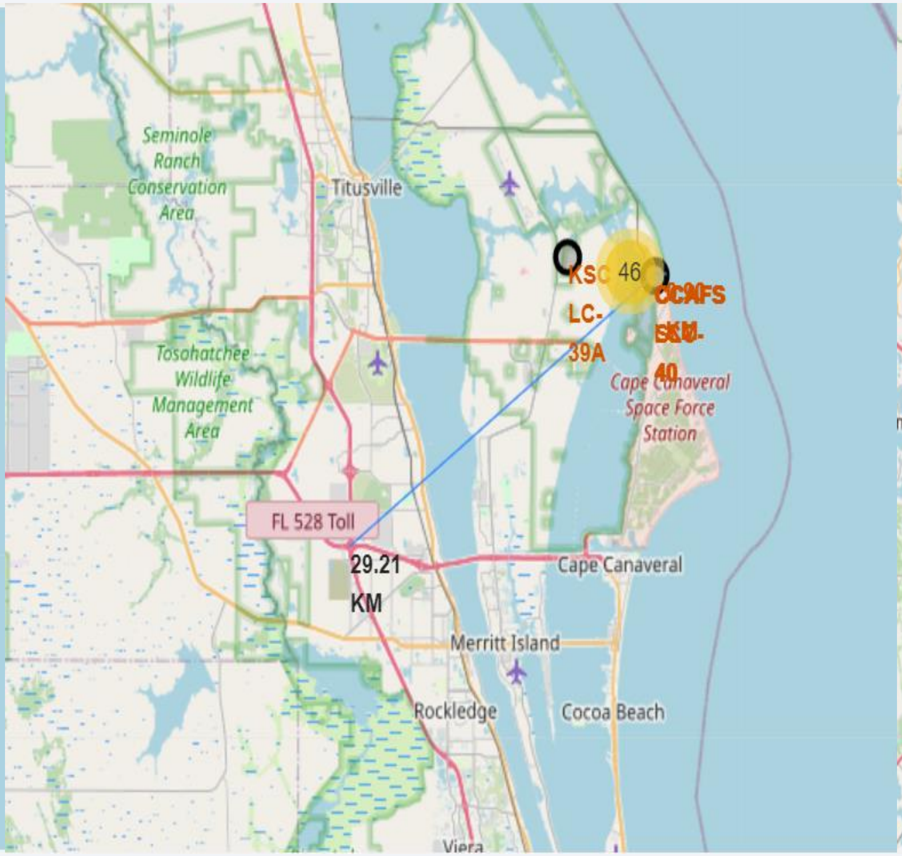


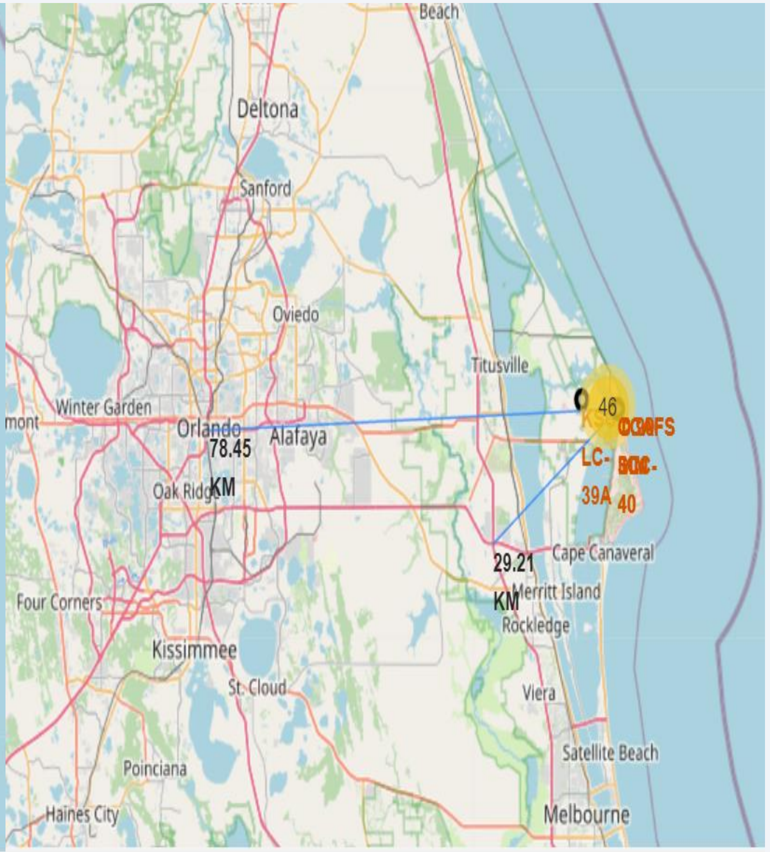
California Launch Site

Florida Launch Sites

The Green Marker are successful launch site while red markers are the failed launch sites.

Launch Site distance to Different Landmarks





Distance to Coastline

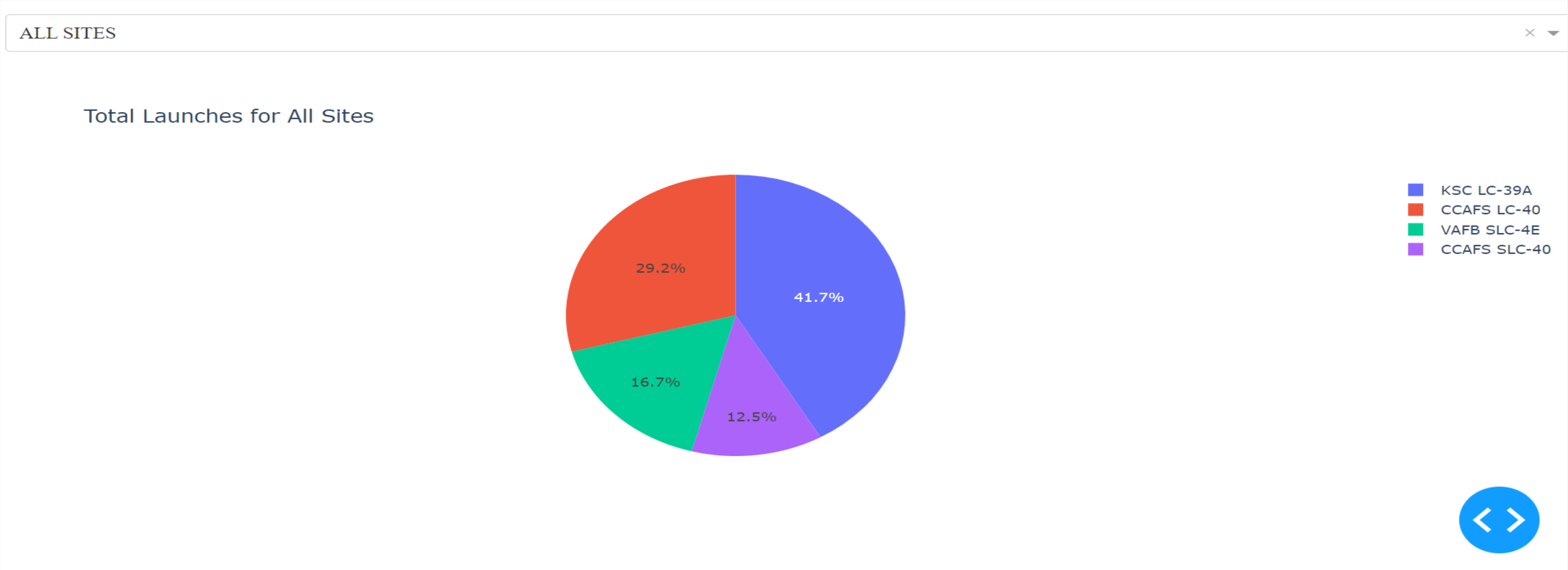
Distance to Highway

Distance to City

Section 5

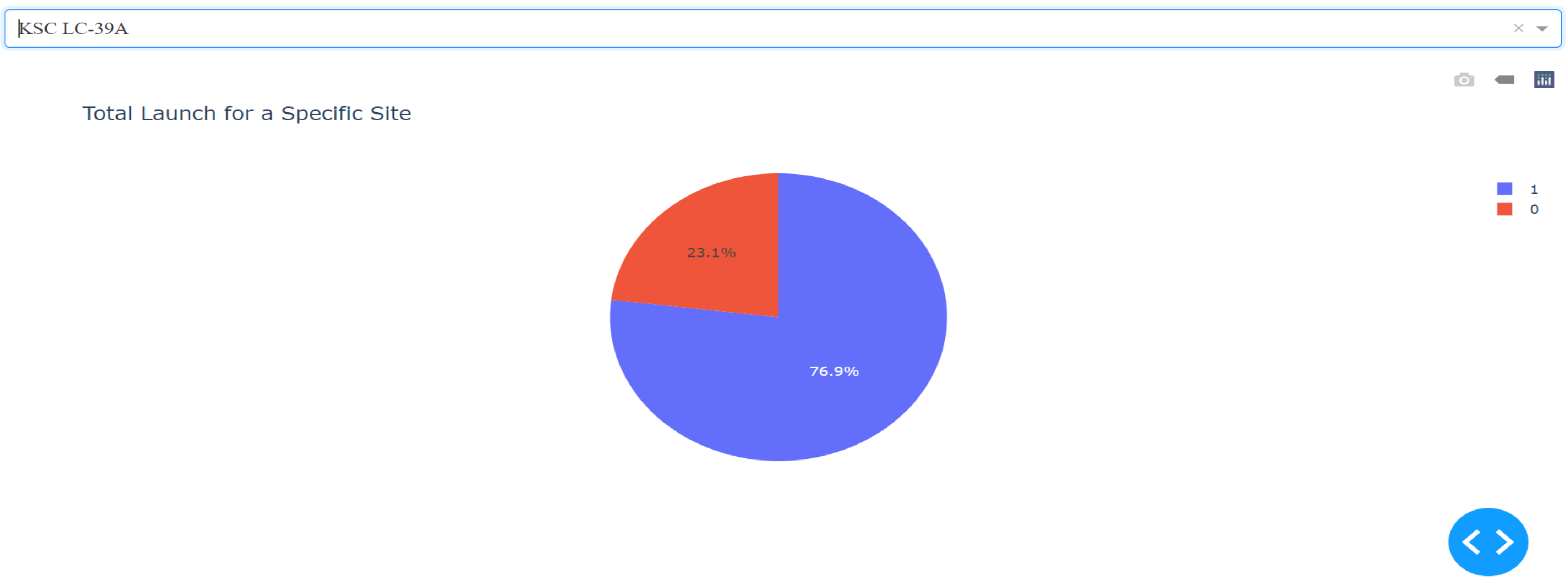
Build a Dashboard with Plotly Dash

Dashboard- Pie chart with Successful Launch Sites



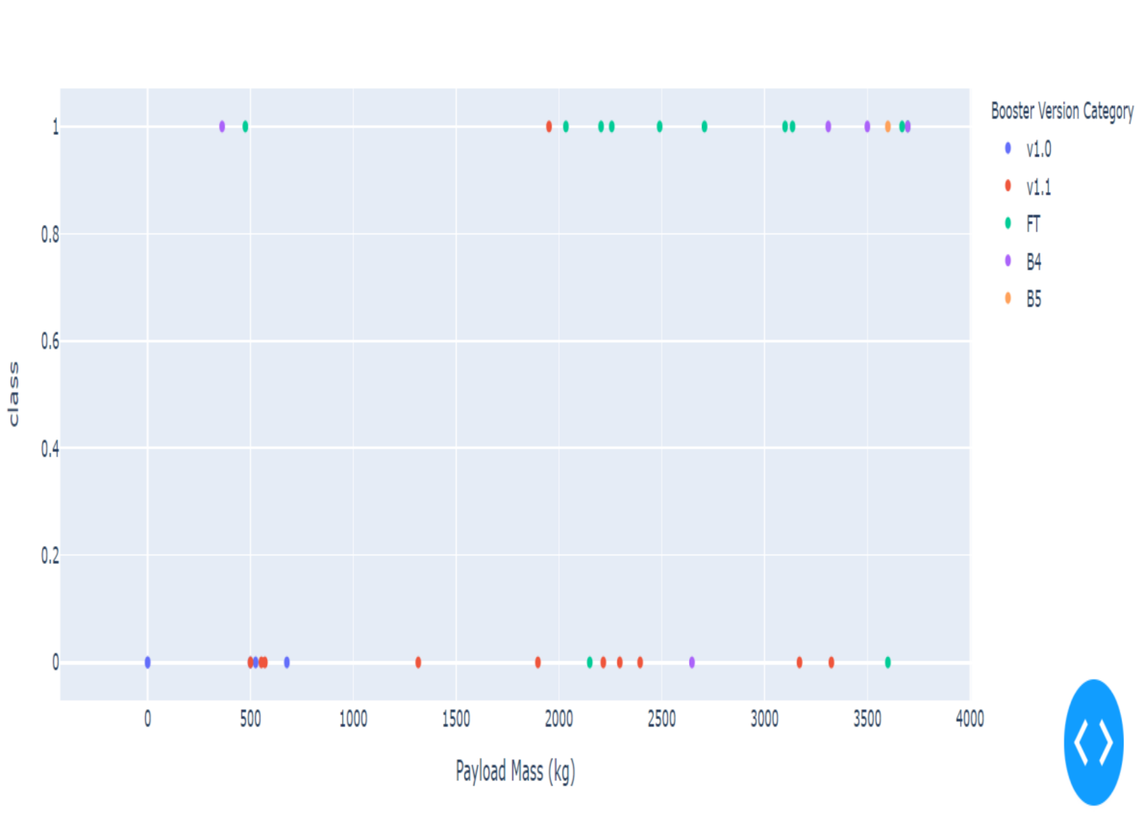
Out of all the launch sites, KSC LC-39A has the highest successful launch percentage with 41.7%

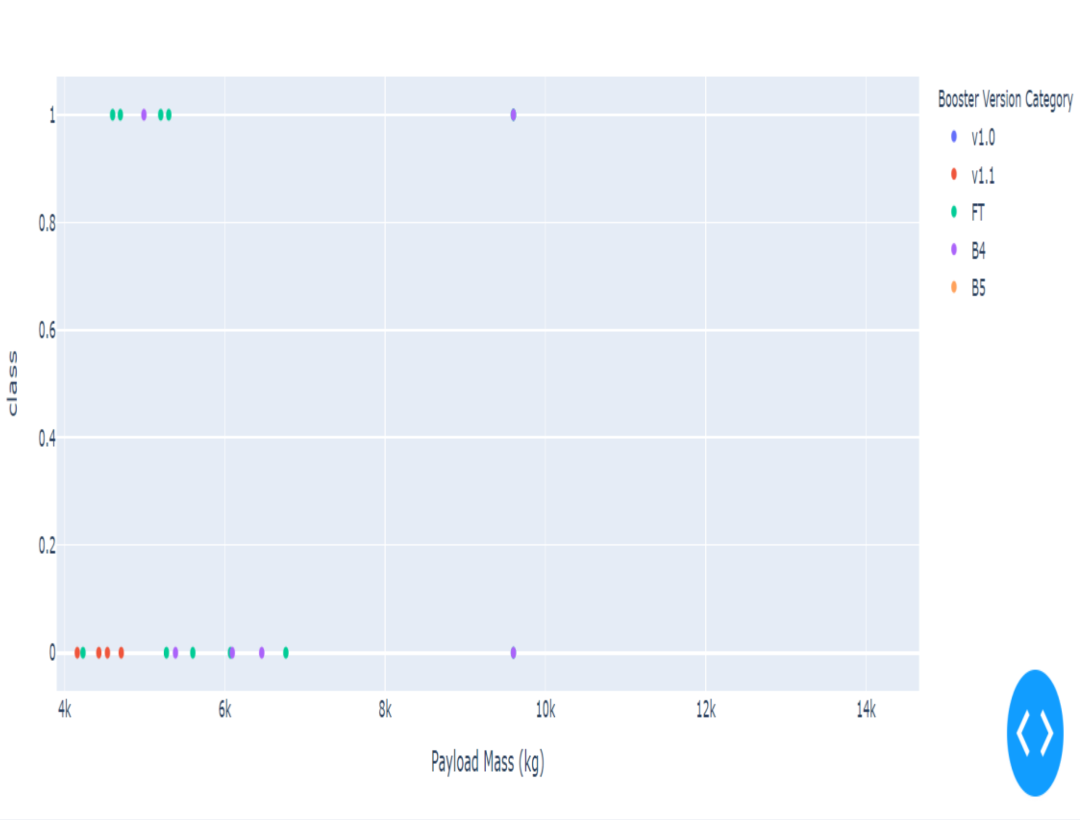
Dashboard - Pie chart with Highest Success Launch Ratio



KSC LC-39A has highest successful launch with 76.9% and 23.1% failure rate.

Dashboard – Payload vs Launch Outcome





- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

Section 6

Predictive Analysis (Classification)

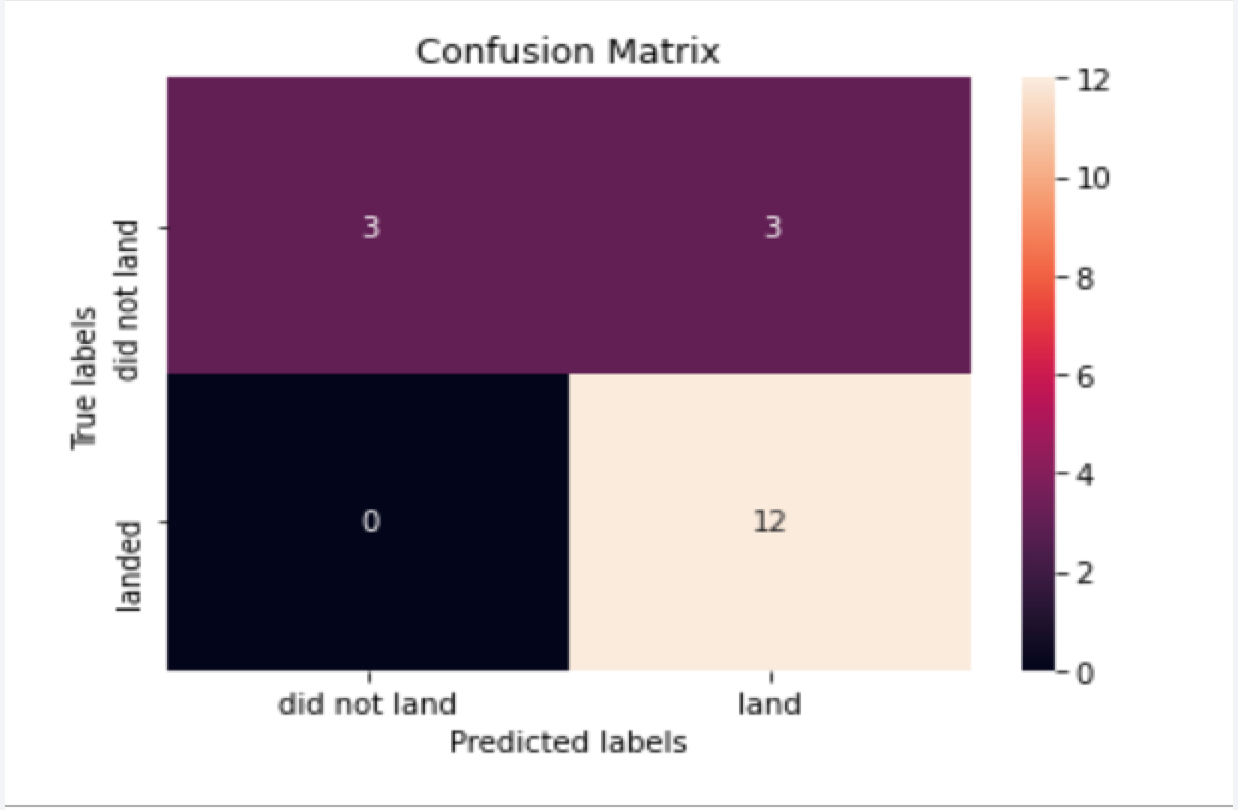
Classification Accuracy

- The DecisionTreeClassifier has the highest accuracy.
- Other model have these accuracy:
  - Logistic Regression: 84.64%
  - SVM : 84.82%
  - KNN: 84.82%

```
Best Method is Tree with a score of 0.875
Best Params is : {'criterion': 'entropy', 'max_depth': 2, 'max_features': 'auto', 'min_samples_leaf': 1, 'min_samples_split': 2, 'split
ter': 'best'}
```

Confusion Matrix

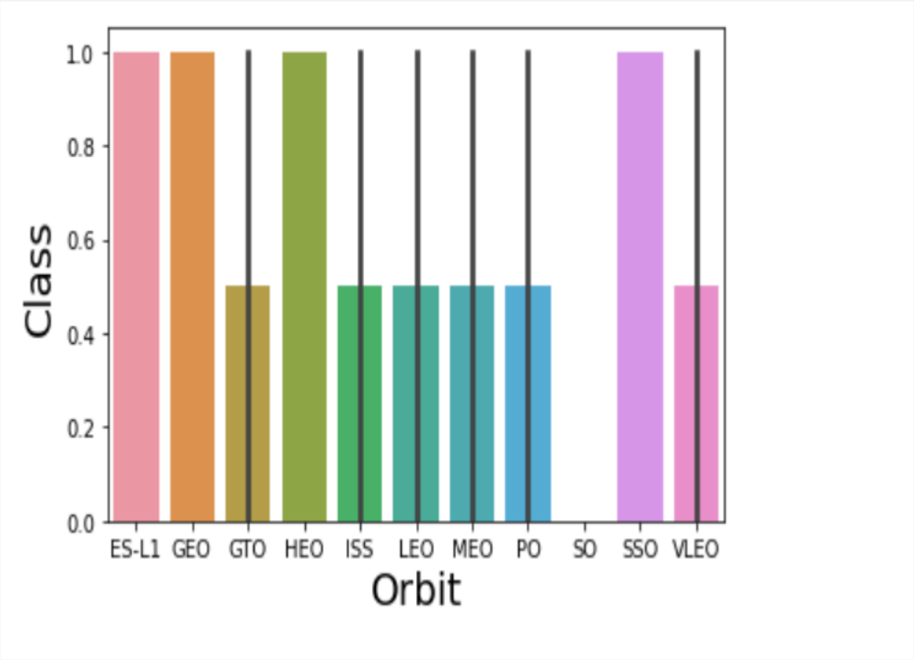
- Confusion matrix of the DecisionTreeClassifier Model. It can differentiate between different class.



Conclusions

- Decision Tree Classifier is the best machine learning algorithm out of all other algorithm for predictive analysis.
- The more the number of flights at launch site, greater the success rate.
- The more the payload mass, the more the success rate for launch sites.
- KSC LC-39A has the most successful launch out of all sites.
- The LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

Appendix



List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
]: %sql select booster_version from SPACEXTBL where landing__outcome = 'Success (drone ship)' and payload_mass__kg_ > 4000 and payload_mass__kg
```

```
 * ibm_db_sa://ksy22772:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/bludb
Done.
```

Out[28]:

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

A bar chart for the success rate of each orbit type

Thank you!