IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Christian Prasetya

Outline

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

Executive Summary

- **Summary of methodologies**
  - Data Collection
  - Data Wrangling
  - EDA with Data Visualization
  - EDA with SQL
  - Interactive Map with Folium
  - Dashboard with Ploty Dash
  - Predictive Analysis
- **Summary of all results**
  - EDA results, Interactive analytics demo, and predictive analysis results.

Introduction

- **Project background and context**
  - We predict if the first stage will land successfully. We can also determine the cost of a launch to let us know about the cost savings compare to the other rocket providers. This information can be used if there is any providers want to bid SpaceX for launching rocket.
- **Problems you want to find answers**
  - What kind of parameters that impact or influence the success rate of landing?
  - What kind of condition that SpaceX need to prepare to achieve the best result of success rocket landing?

Section 1

Methodology

Methodology

Executive Summary

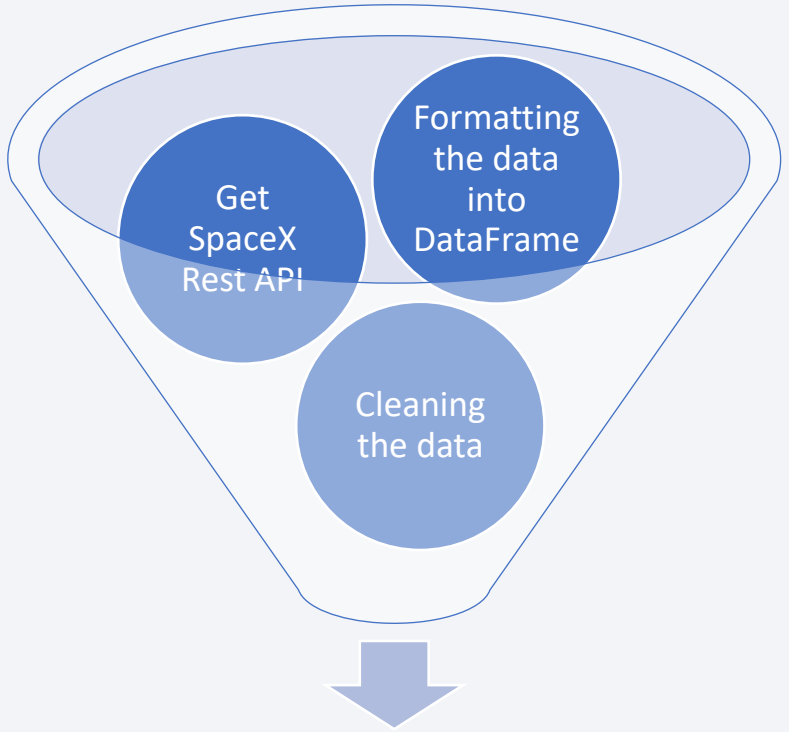
- Data collection methodology:
  - SpaceX Rest API & Web Scrapping from *Wikipedia*.
- Perform data wrangling
  - Removing irrelevant values & One hot encoding data implementation.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium 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

1. Getting Response from API
2. Converting Response to a .json file
3. Apply Custom Function to clean data
4. Assign List to Dictionary to DataFrame
5. A bit data filtering & exporting file.

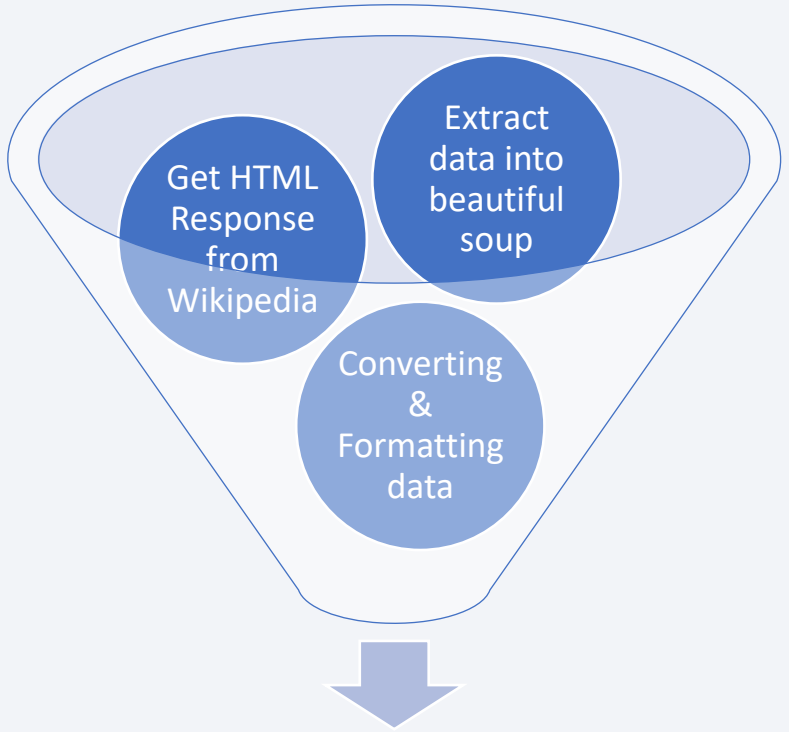


Export result data in csv files

[Github Url]

Data Collection - Scraping

1. Getting Response from HTML
2. Creating BeautifulSoup Object
3. Finding Tables
4. Getting Column names
5. Dictionary & Appending data
6. Converting/Formatting data
7. Dataframe into csv files



Export result data in csv files

[Github Url]

Data Wrangling

EDA on SpaceX Dataset

- Calculate #of launches at each site
- Calculate #of occurrence of each orbit
- Calculate # occurrence of mission outcome / orbit type
- Exporting dataset
- Labeling landing column
- Measuring % success for every landing

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**
  - Mean vs Orbit
- **Line Graph**
  - Success Rate vs Year

Github Url

EDA with SQL

- Displaying the names of the unique launch sites in the space mission
- Displaying 5 records where launch sites begin with the string 'KSC’
- Displaying the total payload mass carried by boosters launched by NASA (CRS)
- Displaying average payload mass carried by booster version F9 v1.1
- Listing the date where the successful landing outcome in drone ship was achieved.
- Listing the names of the boosters which have success in ground pad and have payload mass greater than 4000 but less than 6000
- Listing the total number of successful and failure mission outcomes
- Listing the names of the booster_versions which have carried the maximum payload mass.
- Listing the records which will display the month names, successful landing_outcomes in ground pad ,booster versions, launch_site for the months in year 2017
- Ranking the count of successful landing_outcomes between the date 2010-06-04 and 2017-03-20 in descending order

Github Url

Build an Interactive Map with Folium

Github Url

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 assigned the dataframe launch_outcomes(failures, successes) to classes 0 and 1 with Green and Red markers on the map in a MarkerCluster().

Build a Dashboard with Plotly Dash

The dashboard is built with Flask and Dash web framework.

Github Url

Graphs
- Pie Chart showing the total launches by a certain site/all sites.

Scatter Graph showing the relationship with Outcome and Payload Mass (Kg) for the different Booster Versions
- It shows the relationship between two variables.
- It is the best method to show you a non-linear pattern.
- The range of data flow, i.e. maximum and minimum value, can be determined.

Predictive Analysis (Classification)

Building Model

- Load & Transform data
- Split into training & test dataset
- Decide ML Algorithm
- GridSearchCV Configuration
- Implement dataset into GridSearchCV Algorithm

Github Url

Evaluating Model

- Check Accuracy of model
- Confusion matrix

Improving Model

- Feature Engineering
- Algorithm Tuning

Finding Best Performance Classification

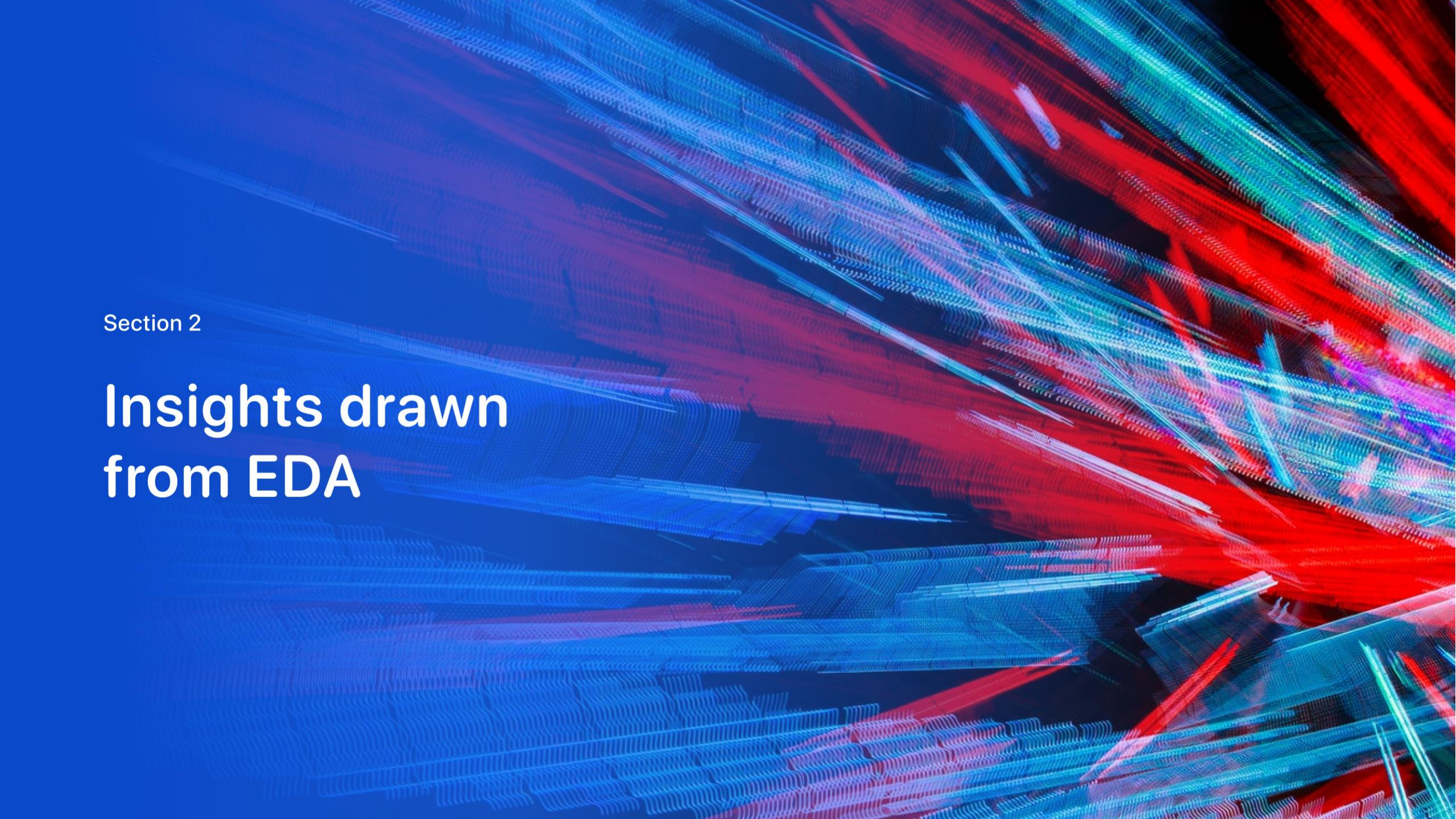
- Based on best accuracy score

Results

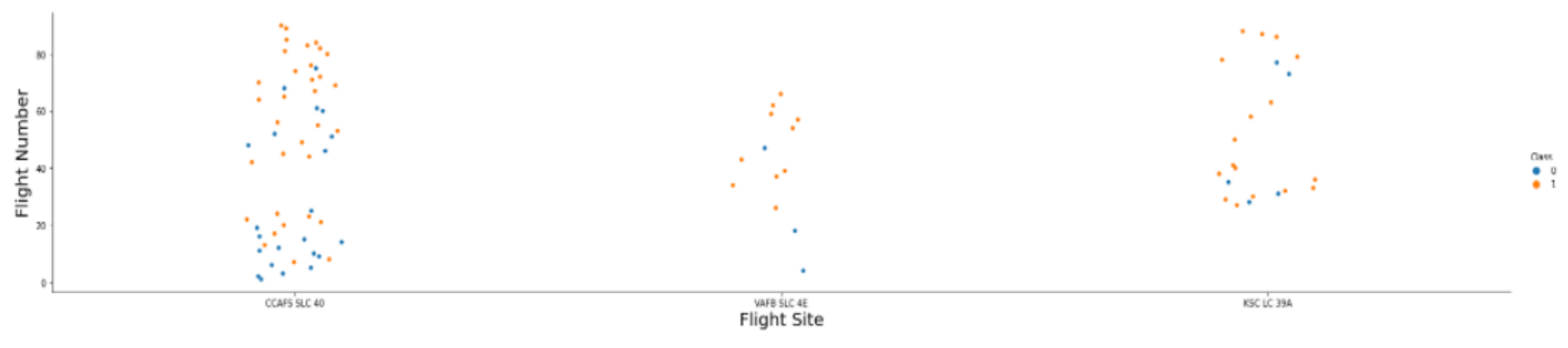
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

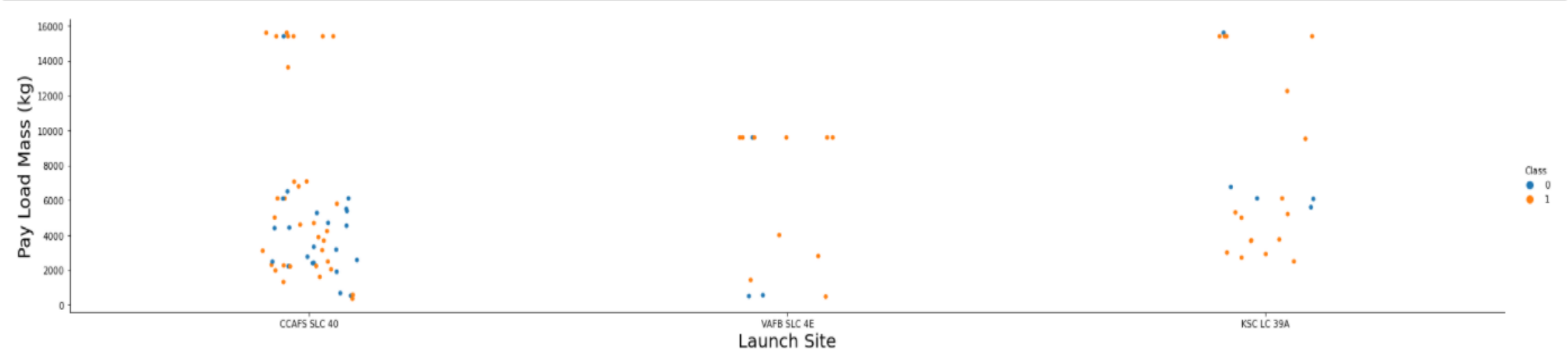


Flight Number vs. Launch Site



- **More amount of # flights at a launch site have inclination of greater the success rate at a launch site.**

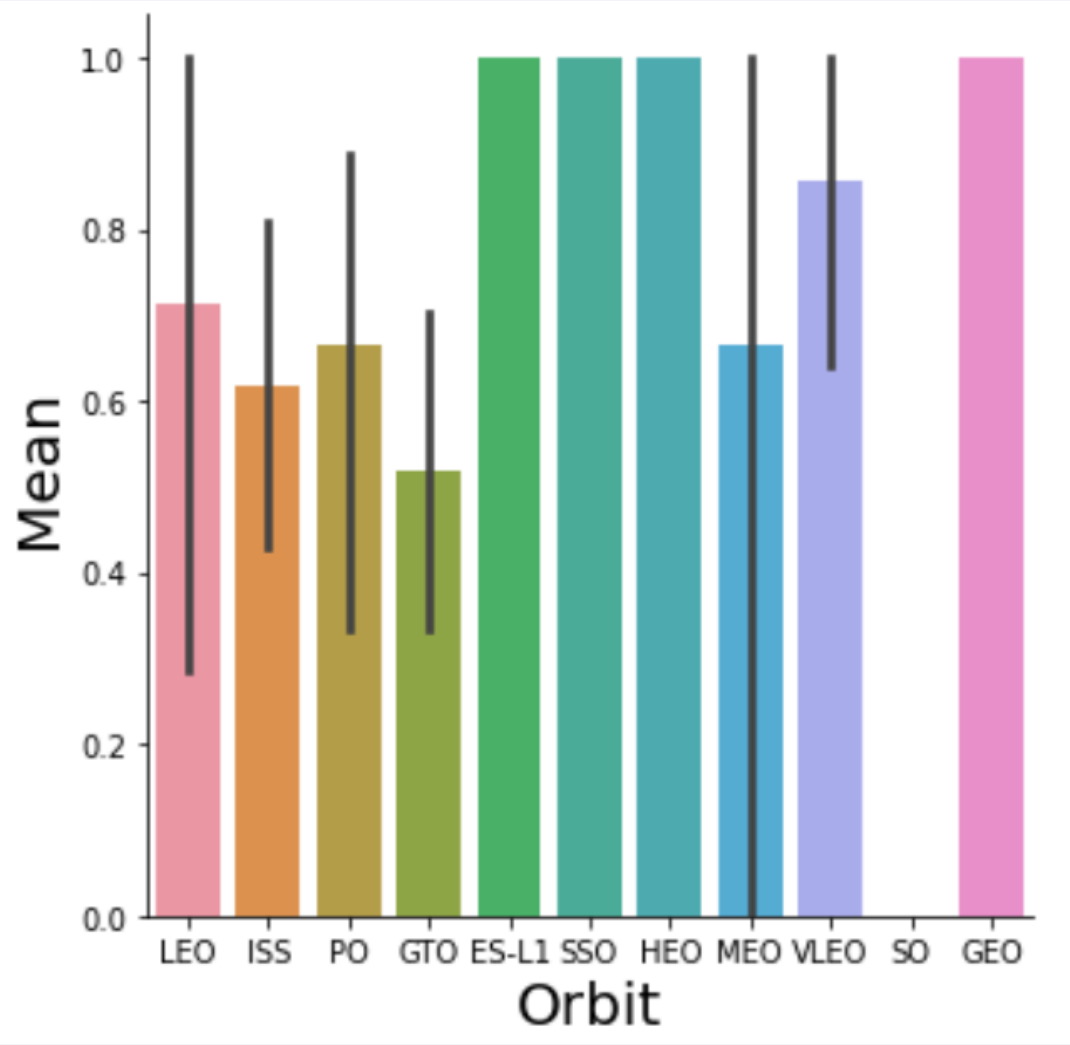
Payload vs. Launch Site



- **The greater the payload mass for CCAFS SLC 40, the higher the success rate for the Rocket.**
- **For the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10,000)**

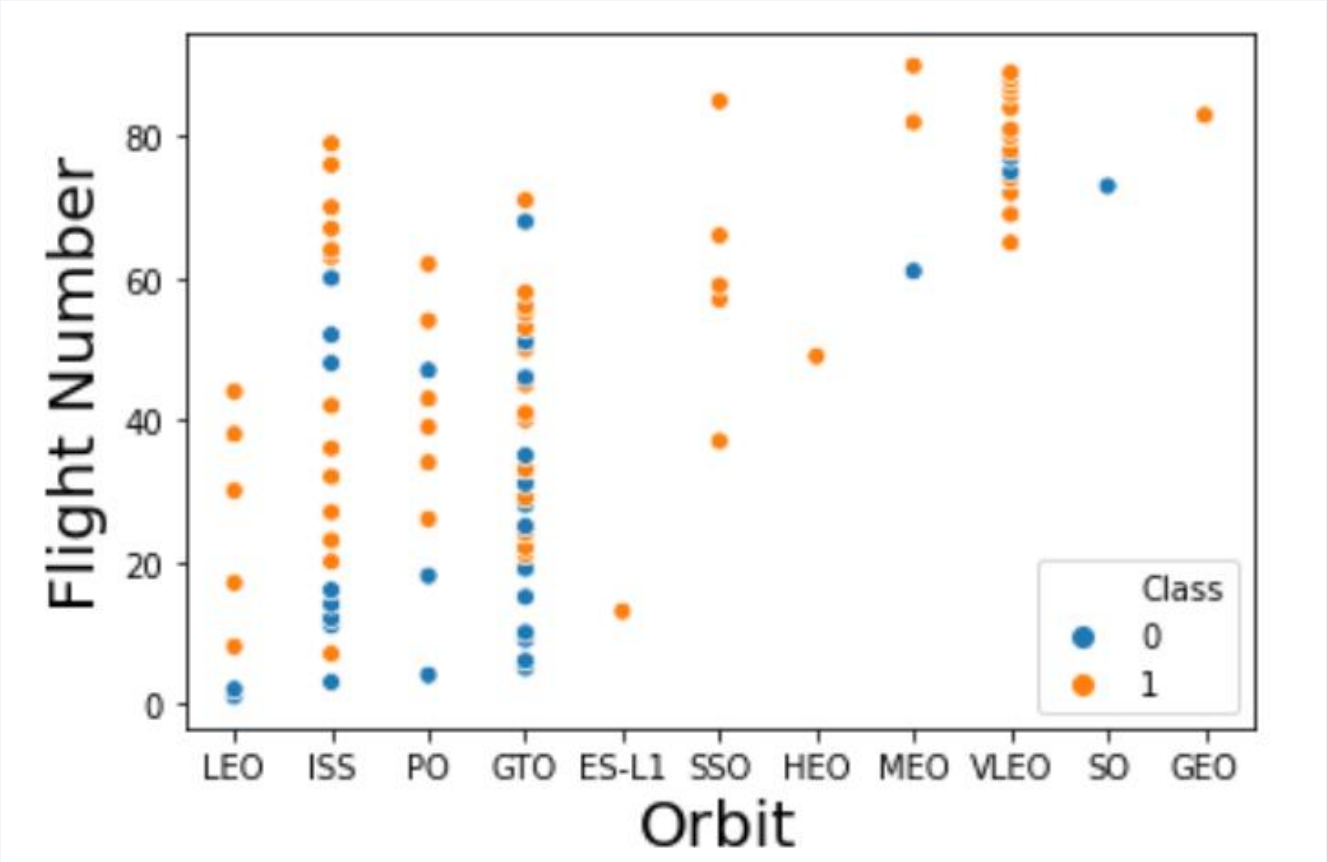
Success Rate vs. Orbit Type

- The best success rates happened on
  - GEO
  - HEO
  - SSO
  - ES-L1



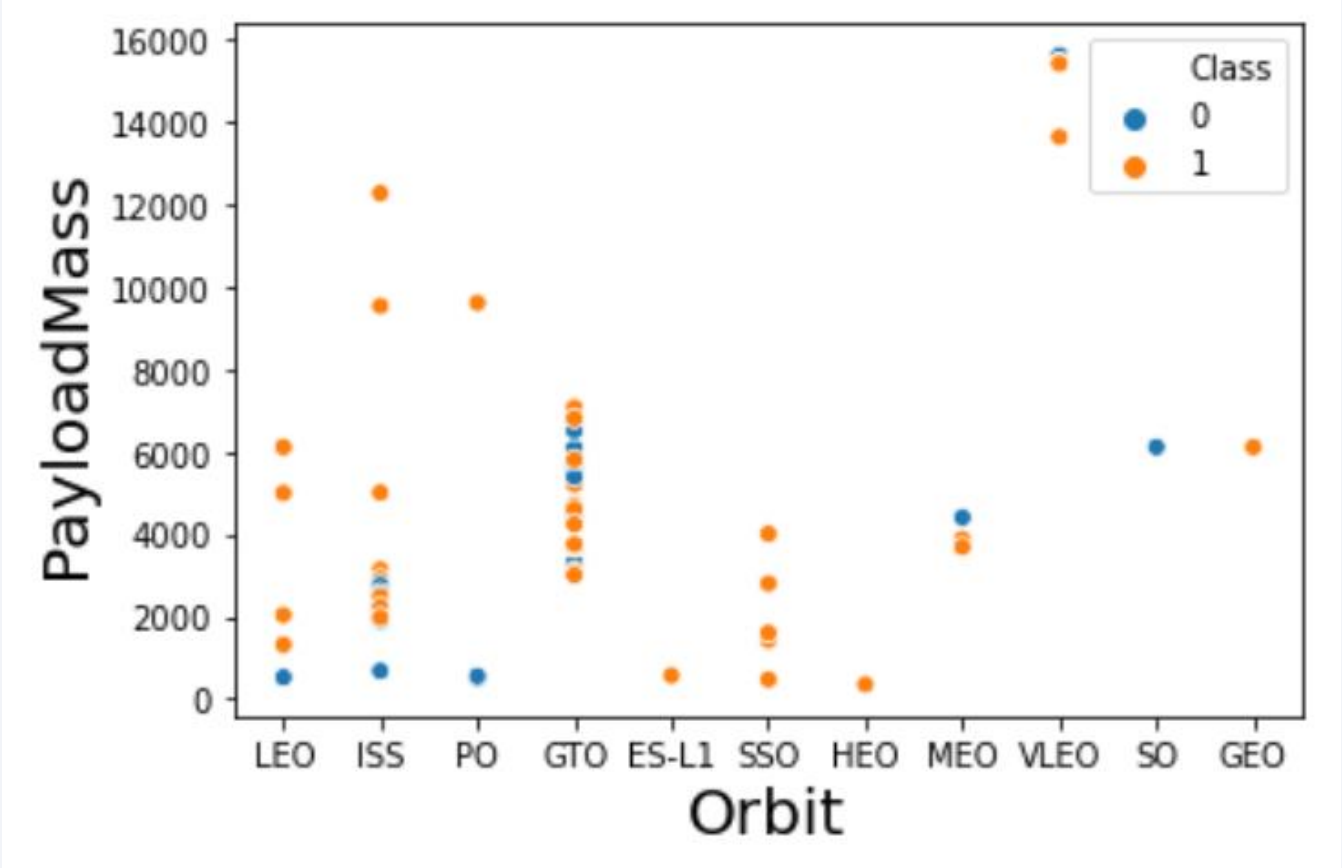
Flight Number vs. Orbit Type

- On LEO Orbit, there are relations between # success appears and number of flights.
- On the other hand, there is no relation in the other orbit.



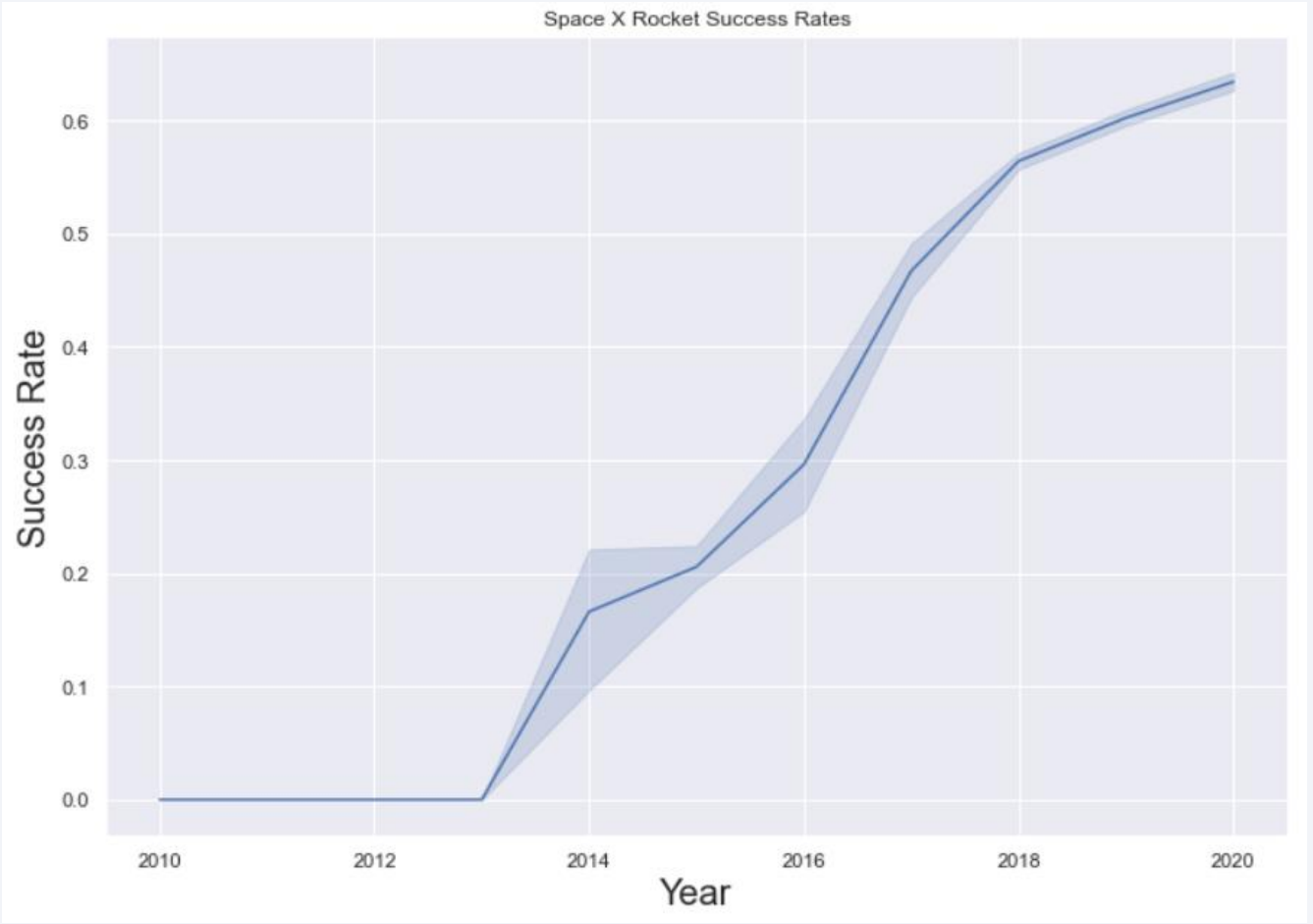
Payload vs. Orbit Type

- **With heavy payloads, the successful landing landing rate happened in LEO and ISS orbit.**
- **However for GTO, we cannot distinguish this well as both positive landing and negative landing.**



Launch Success Yearly Trend

- **Since 2013, the success rates start constantly increasing until 2000.**



All Launch Site Names

select DISTINCT Launch_Site from tblSpaceX

```
['CCAFS LC-40', 'CCAFS SLC-40', 'CCAFSSLC-40', 'KSC LC-39A', 'VAFB SLC-4E']
```

Launch Site Names Begin with KSC'

select TOP 5 * from tblSpaceX WHERE Launch_Site LIKE 'KSC%'

	Date	Time_UTC	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS_KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
0	19-02-2017	2021-07-02 14:39:00.0000000	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
1	16-03-2017	2021-07-02 06:00:00.0000000	F9 FT B1030	KSC LC-39A	EchoStar 23	5600	GTO	EchoStar	Success	No attempt
2	30-03-2017	2021-07-02 22:27:00.0000000	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
3	01-05-2017	2021-07-02 11:15:00.0000000	F9 FT B1032.1	KSC LC-39A	NROL-76	5300	LEO	NRO	Success	Success (ground pad)
4	15-05-2017	2021-07-02 23:21:00.0000000	F9 FT B1034	KSC LC-39A	Inmarsat-5 F4	6070	GTO	Inmarsat	Success	No attempt

Total Payload Mass

select SUM(PAYLOAD_MASS_KG_) TotalPayloadMass from tblSpaceX where Customer = 'NASA (CRS)'

	Total Payload Mass
0	45596

Average Payload Mass by F9 v1.1

select AVG(PAYLOAD_MASS_KG_) AveragePayloadMass from tblSpaceX where Booster_Version = 'F9 v1.1'

	Average Payload Mass
0	2928

First Successful Ground Landing Date

select MIN(Date) SLO from tblSpaceX where Landing_Outcome = 'Success (drone ship)'

0	06-05-2016

Successful Drone Ship Landing with Payload between 4000 and 6000

select Booster_Version from tblSpaceX where Landing_Outcome = 'Success (ground pad)' AND Payload_MASS_KG_ > 4000 AND Payload_MASS_KG_ < 6000

0	F9 FT B1032.1
1	F9 B4 B1040.1
2	F9 B4 B1043.1

Total Number of Successful and Failure Mission Outcomes

SELECT(SELECT Count(Mission_Outcome) from tblSpaceX where Mission_Outcome LIKE '%Success%') as Successful_Mission_Outcomes,(SELECT Count(Mission_Outcome) from tblSpaceX where Mission_Outcome LIKE '%Failure%') as Failure_Mission_Outcomes

	Successful_Mission_Outcomes	Failure_Mission_Outcomes
0	100	1

Boosters Carried Maximum Payload

SELECT DISTINCT Booster_Version, MAX(PAYLOAD_MASS_KG_) AS [Maximum Payload Mass] FROM tblSpaceX GROUP BY Booster_Version ORDER BY [Maximum Payload Mass] DESC

	Booster_Version	Maximum Payload Mass
0	F9 B5 B1048.4	15600
1	F9 B5 B1048.5	15600
2	F9 B5 B1049.4	15600
3	F9 B5 B1049.5	15600
4	F9 B5 B1049.7	15600
...	...	...

2015 Launch Records

SELECT DateName(month , DateAdd(month , MONTH(CONVERT(date,Date, 105)) , 0) - 1) as Month, Booster_Version, Launch_Site, Landing_Outcome FROM tblSpaceX WHERE (Landing_Outcome LIKE N'%Success%') AND YEAR(CONVERT(date,Date, 105)) = '2015'

	Month	Booster_Version	Launch_Site	Landing_Outcome
0	January	F9 FT B1029.1	VAFB SLC-4E	Success (drone ship)
1	February	F9 FT B1031.1	KSC LC-39A	Success (ground pad)
2	March	F9 FT B1021.2	KSC LC-39A	Success (drone ship)
3	May	F9 FT B1032.1	KSC LC-39A	Success (ground pad)
4	June	F9 FT B1035.1	KSC LC-39A	Success (ground pad)
5	June	F9 FT B1029.2	KSC LC-39A	Success (drone ship)
6	June	F9 FT B1036.1	VAFB SLC-4E	Success (drone ship)
7	August	F9 B4 B1039.1	KSC LC-39A	Success (ground pad)
8	August	F9 FT B1038.1	VAFB SLC-4E	Success (drone ship)
9	September	F9 B4 B1040.1	KSC LC-39A	Success (ground pad)
10	October	F9 B4 B1041.1	VAFB SLC-4E	Success (drone ship)
11	October	F9 FT B1031.2	KSC LC-39A	Success (drone ship)
12	October	F9 B4 B1042.1	KSC LC-39A	Success (drone ship)
13	December	F9 FT B1035.2	CCAFS SLC-40	Success (ground pad)

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

SELECT COUNT(Landing_Outcome) AS sl FROM dbo.tblSpaceX WHERE (Landing_Outcome LIKE '%Success%') AND (Date >'04-06-2010') AND (Date < '20-03-2017')

0	34

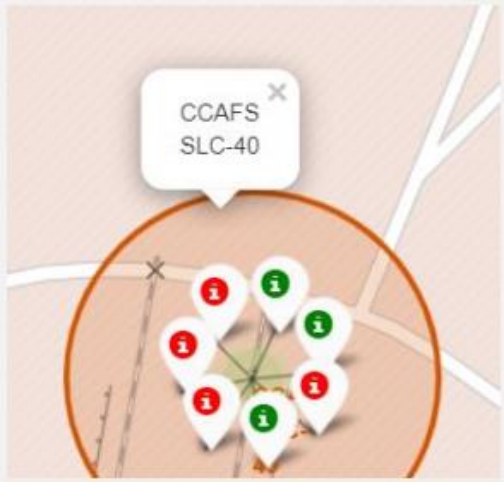
Section 4

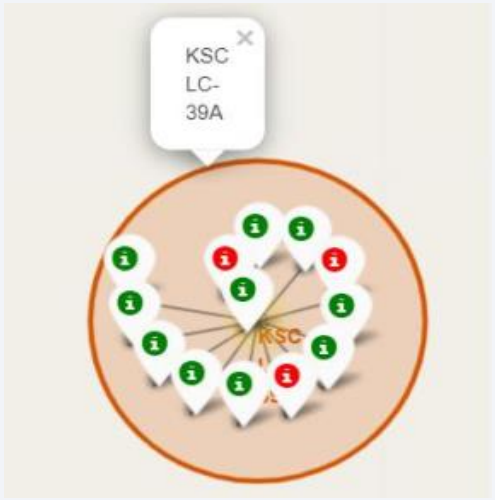
Launch Sites Proximities Analysis

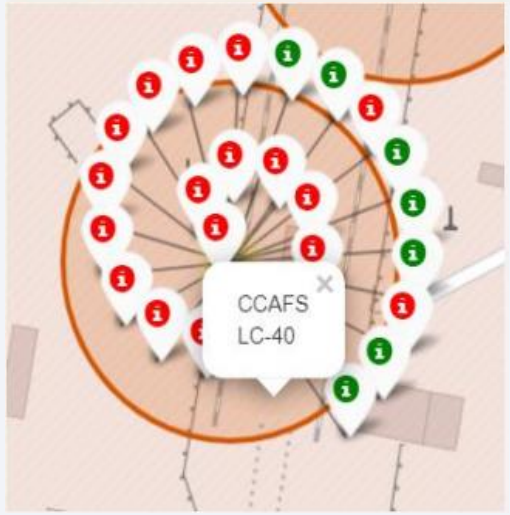
Launch Sites in Global Maps Perspective

Labeling Markers

Florida Launch Sites



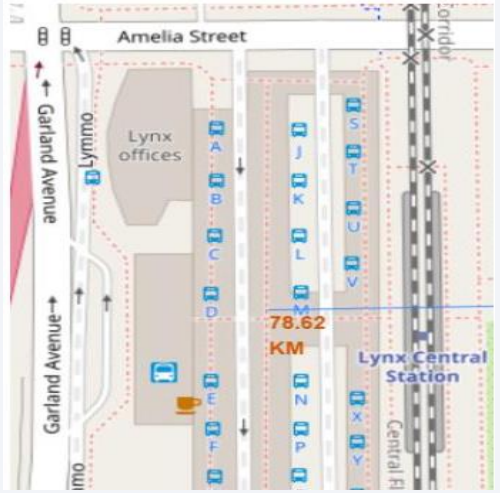




California Launch Sites

Launch sites close proximity to city ~ but far away from railways, highways, and city

Distance to Railway Station



Distance to Highway

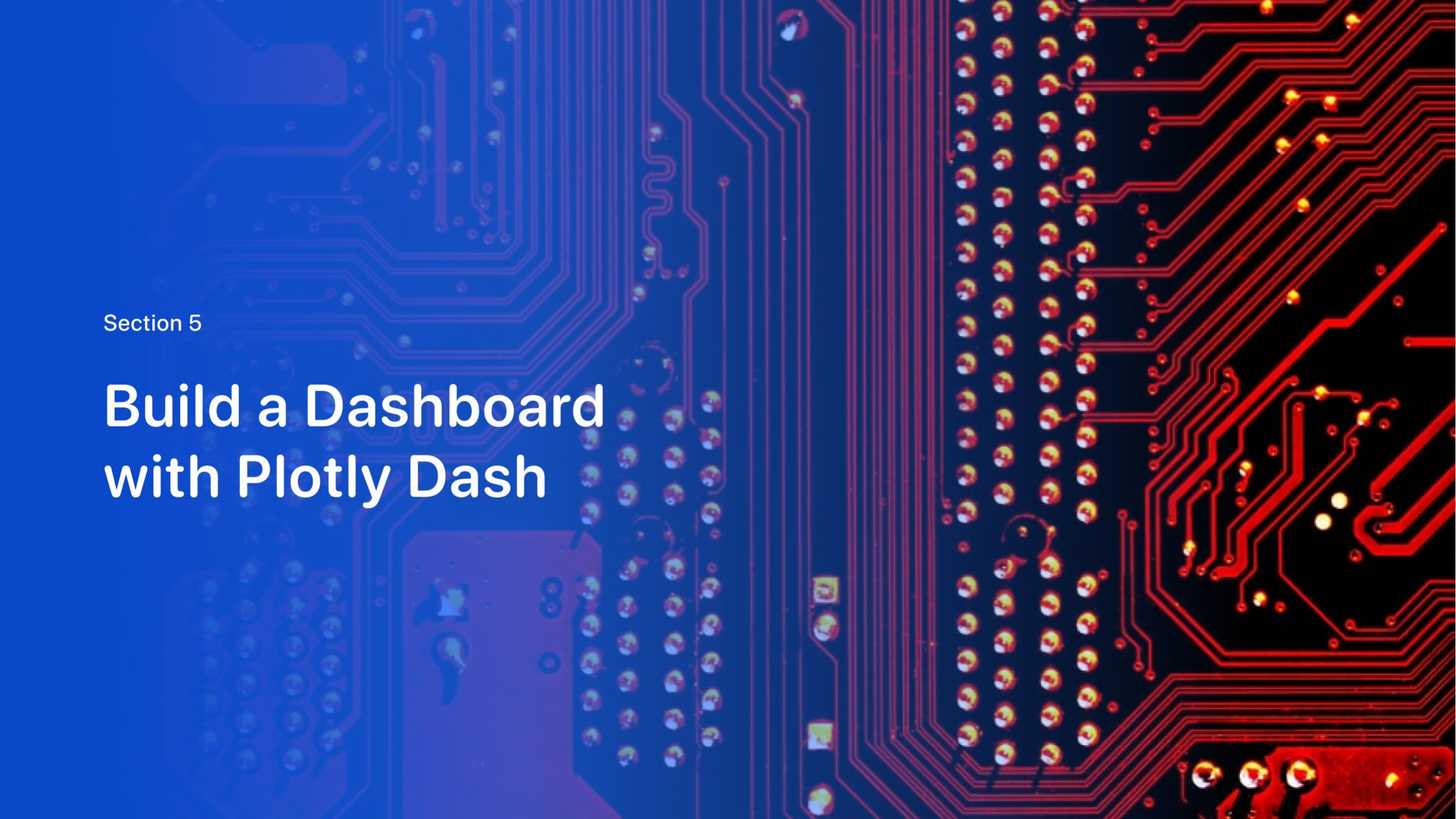
Distance to Coast

Distance to Coastline

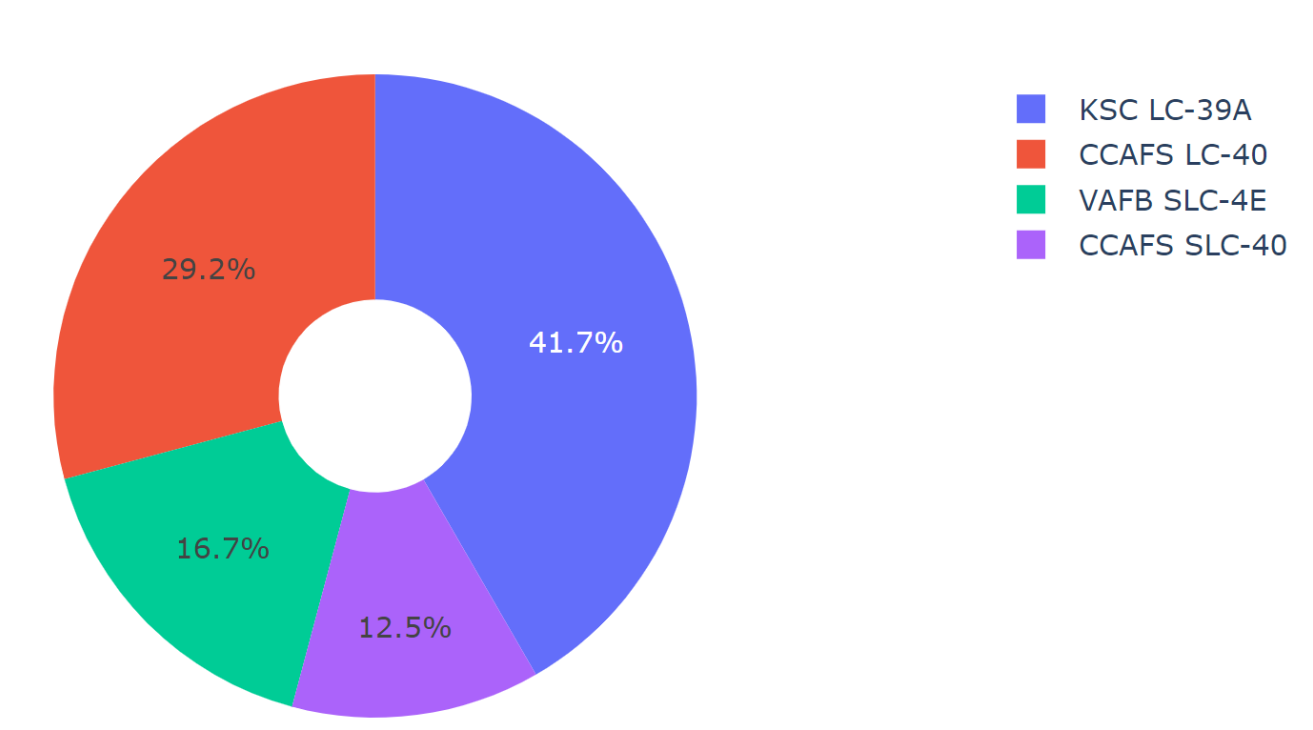
Distance to City

Section 5

Build a Dashboard with Plotly Dash

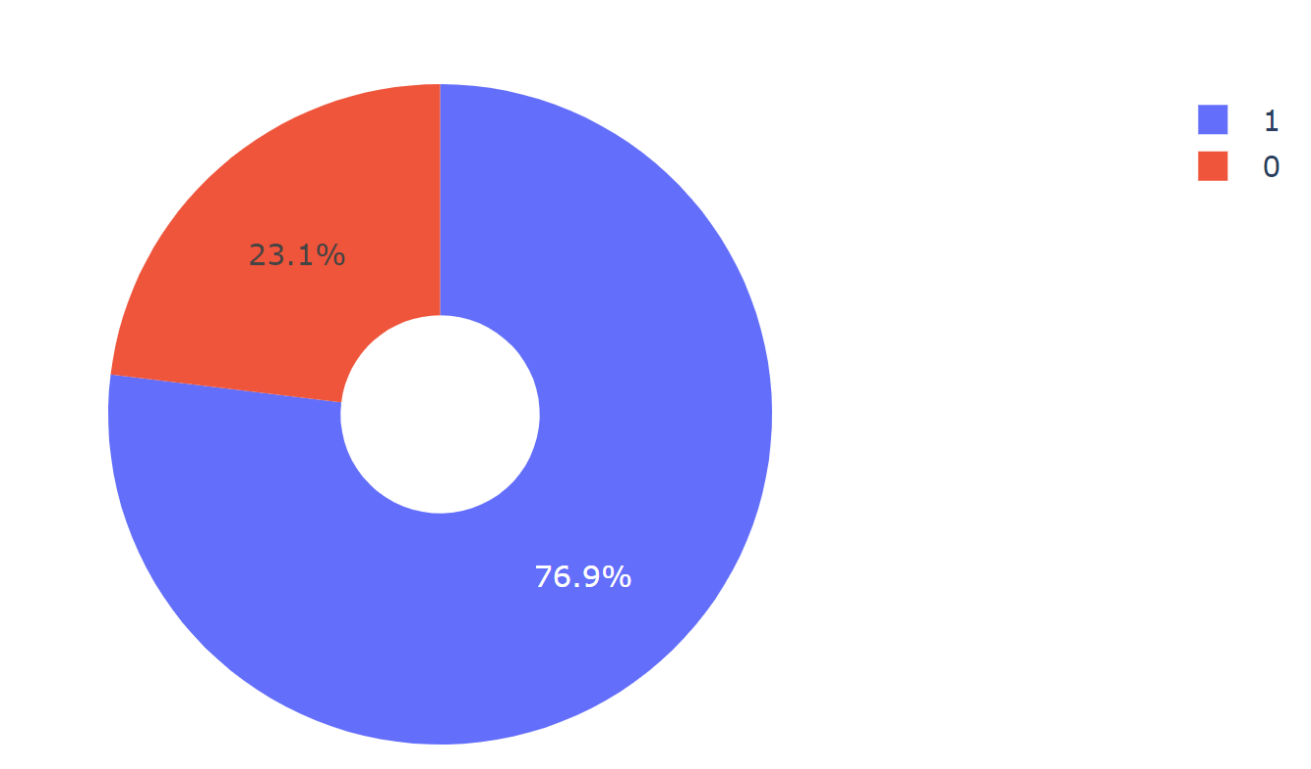


Total Success Launches by All Sites



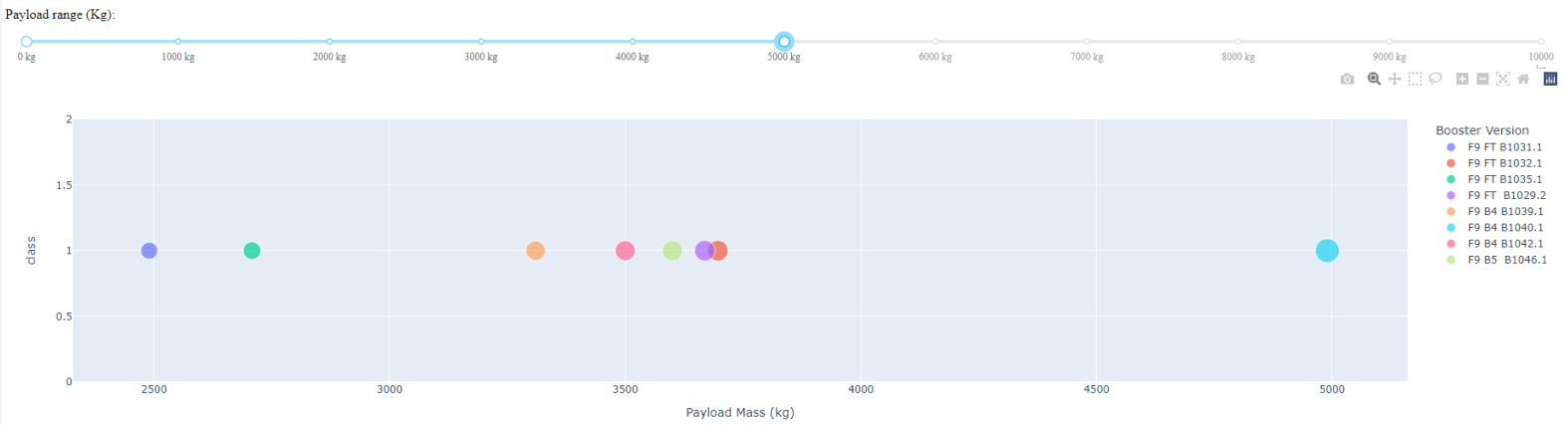
KSC LC-39A have the most successful launches among all the sites

Highest Success Ratio of Launch Sites



KSC LC-39A
76.9% success rate
23.1% failure rate

Payload vs Launch Outcome Plot with Weighted Category Slider



% Success of low weighted payloads (0-5.000 kg) is greater than % success of heavy weighted payloads (5.000-10000 kg)

Section 6

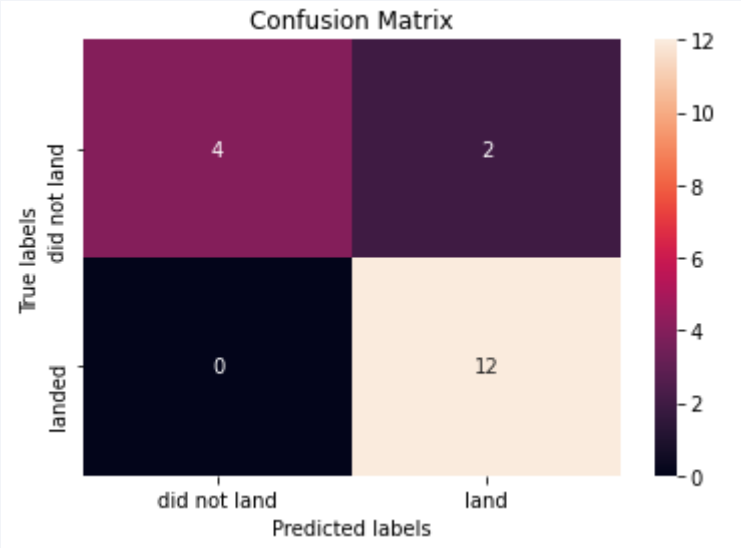
Predictive Analysis (Classification)

Classification Accuracy


```
Best Algorithm is Tree with a score of 0.9035714285714287
Best Params is : {'criterion': 'entropy', 'max_depth': 18, 'max_features': 'auto', 'min_samples_leaf': 1, 'min_samples_split':
2, 'splitter': 'best'}
```

Tree Algorithm have the highest accuracy among other algorithm, with 90% accuracy.

Confusion Matrix



Correct Prediction for all labels, except 2 datas for false positive case.

Conclusions

- Tree algorithm is the best machine learning options to be implemented.
- Low weighted payloads perform better than heavy payloads.
- KSC LC-39A is the most successful launches over all the sites.
- GEO, HEO, ES-L1, and SSO are the best orbit's success rate.

Appendix

- Sql Server Module (ADGSQLSERVER)

Thank you!