IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Name: Nguyen Kien Quoc

Date: 01/09/2023

Outline

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

Executive Summary

- Methodologies
  - Data Collection using Web Scrapping
  - Data Wrangling
  - Data Preparation using SQL
  - Visualization
  - Modeling and Evaluating
- Results
  - Exploratory data analysis
  - Interactive analytics
  - Predictive analysis

Introduction

- **Project background and context**
  - SpaceX is an American spacecraft manufacturer, launcher, and a satellite communications corporation headquartered in Hawthorne, California which was founded in 2002 by Elon Musk.
  - SpaceX's accomplishments include: Sending spacecraft to the International Space Station. In order to send manned missions to Space. SpaceX uses the rocket launches which are relatively inexpensive.
  - SpaceX advertises Falcon 9 rocket launches on its website with a cost of **62 million dollars**; Other providers cost upwards of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.

- **Problems want to answer**
  - Determine the price of each launch
  - Predict if the first stage will land successfully
  - And determine if SpaceX will reuse the first stage

Section 1

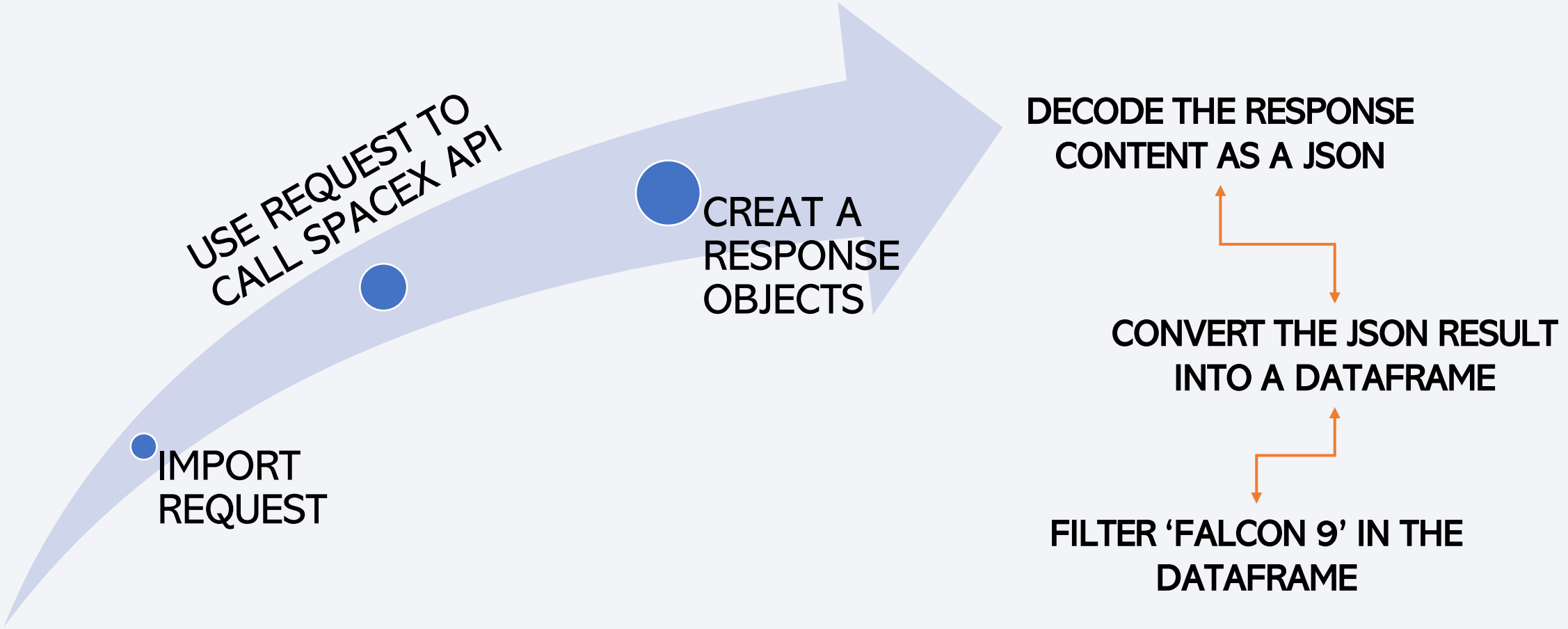
Methodology

Methodology

Executive Summary

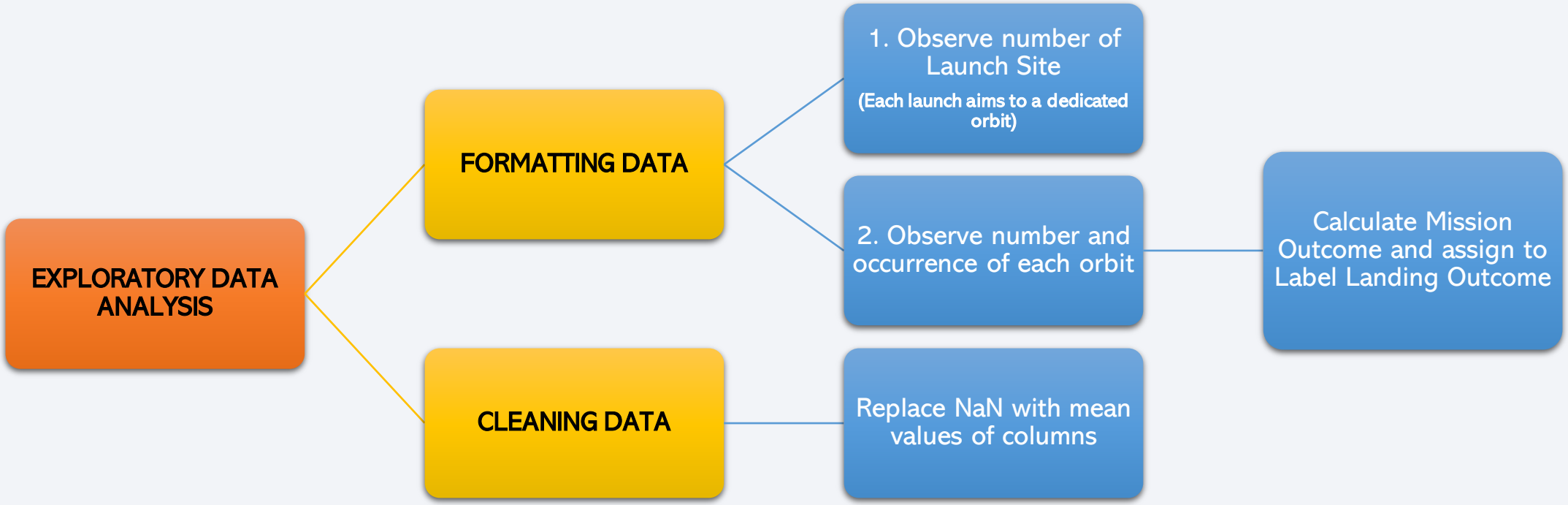
- Data collection methodology:
  - Use Request/Response to the SpaceX API and clean the requested data by doing some basic data wrangling and formatting.
- Perform data wrangling
  - Calculate statistical data to observe and determine Training Labels
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection - SpaceX API



GitHub --- https://github.com/Mikey3300/IBM-Capstone.git

Data Wrangling



GitHub --- https://github.com/Mikey3300/IBM-Capstone.git

EDA with Data Visualization

In this section, the total charts we use are 6 charts including:

- **Scatter Chart** – The chart will help showing the relationship between two variables and the trend of them.
- **Bar Chart** – The chart will help measuring the success rate based on each Orbit, in order to observe which one works the best.
- **Line Chart** – The chart will help identifying the trend of the average success rate through each year.

GitHub --- https://github.com/Mikey3300/IBM-Capstone.git

EDA with SQL

- In this section, the total SQL queries were used to perform including:
  - SELECT
  - FROM
  - WHERE (LIKE)
  - MAX/MIN
  - BETWEEN … AND …
  - LIMIT
  - ORDER BY
  - GROUP BY
  - SUBQUERY

GitHub --- https://github.com/Mikey3300/IBM-Capstone.git

Build a Dashboard with Plotly Dash

Interaction added including:

- Dropdown Menu
- A Range Slider

Plots/Graphs added including:

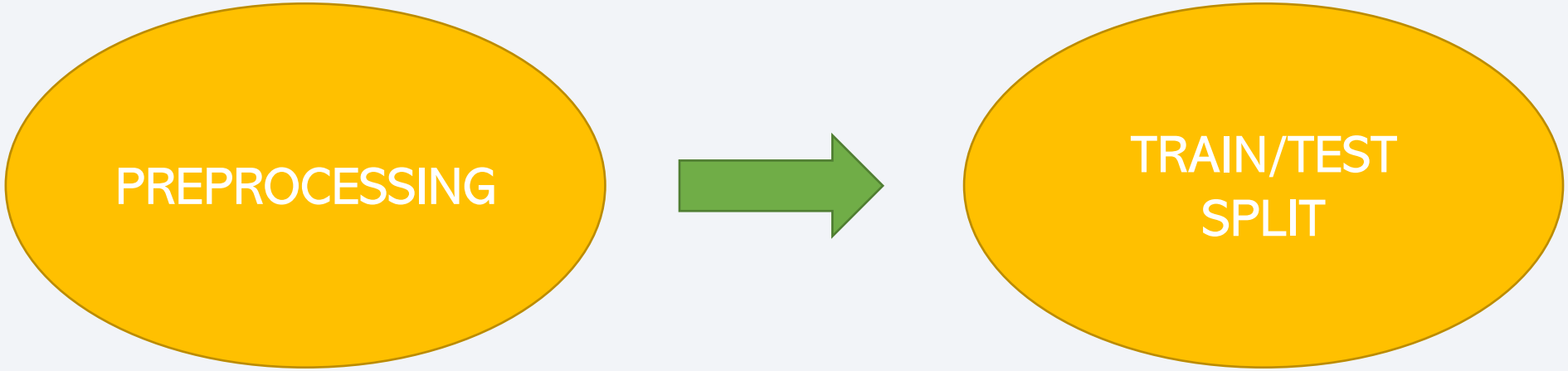
- **A Pie Chart** – To observe when the input is All Launch Sites which one has the most successful landings and when the input is Each Launch, we will observe the percentage of Failure and Success Landing Rate
- **A Scatter Chart** – To observe the strength of the relationship between Payload and Classification of Landing when the input is All Launch Sites or Each Launch.

GitHub --- https://github.com/Mikey3300/IBM-Capstone.git

Predictive Analysis (Classification)

MODEL DEPLOYMENT PROCESS

1. STANDARDIZE AND SPLIT DATA

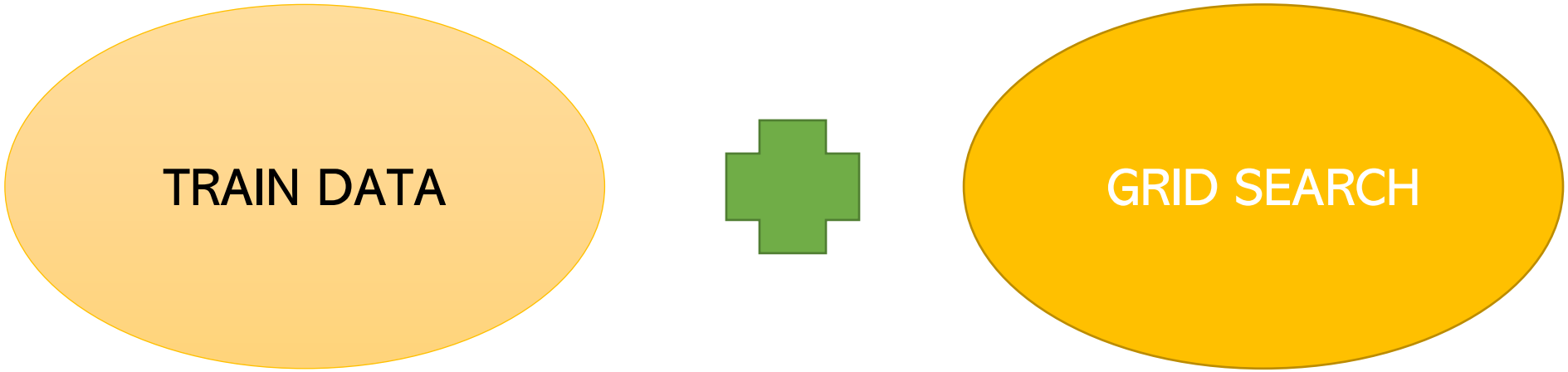


GitHub --- https://github.com/Mikey3300/IBM-Capstone.git

Predictive Analysis (Classification)

MODEL DEPLOYMENT PROCESS

2. Find the hyperparameters that allow a given algorithm to perform best by perform **GRIDSEARCH** with **TRAIN DATA**

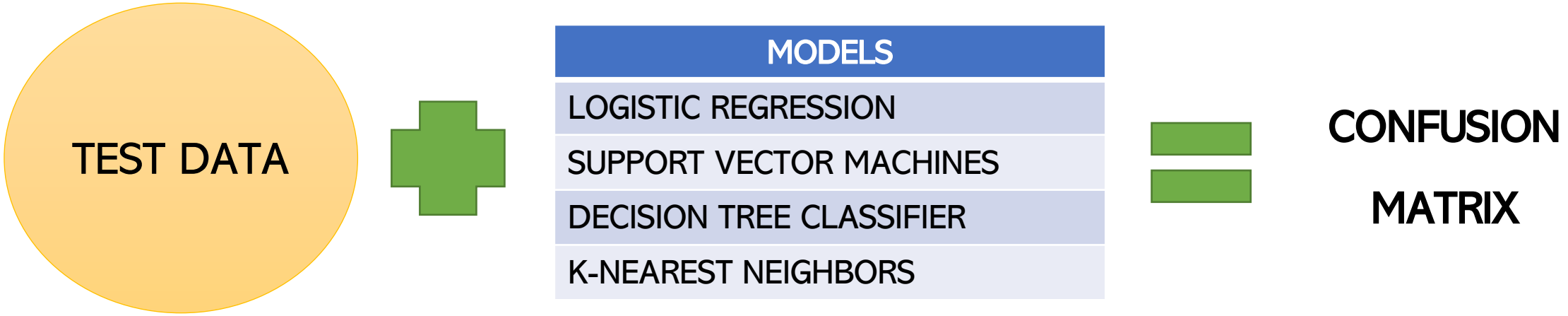


GitHub --- https://github.com/Mikey3300/IBM-Capstone.git

Predictive Analysis (Classification)

MODEL DEPLOYMENT PROCESS

3. Improve the model by testing data with several models and output the Confusion Matrix, then looking for the best model by its accuracy and output.



GitHub --- https://github.com/Mikey3300/IBM-Capstone.git

Results

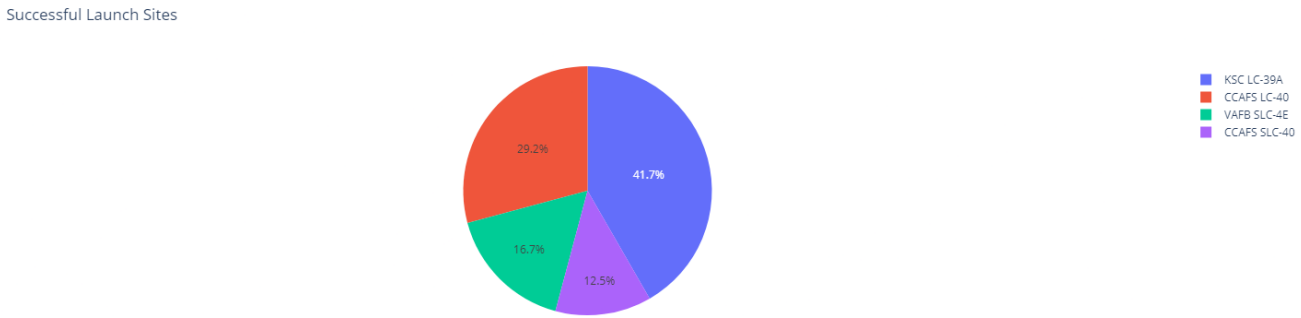
- Exploratory data analysis results
  - Determine Class as Success Rate which shows that if Class = 0 means Fail, otherwise Class = 1 means Success

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Class
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	0
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	0
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	0
3	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	0
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	0
...	...	...	...	...	...	...	...
85	86	2020-09-03	Falcon 9	15400.000000	VLEO	KSC LC 39A	1
86	87	2020-10-06	Falcon 9	15400.000000	VLEO	KSC LC 39A	1
87	88	2020-10-18	Falcon 9	15400.000000	VLEO	KSC LC 39A	1
88	89	2020-10-24	Falcon 9	15400.000000	VLEO	CCAFS SLC 40	1
89	90	2020-11-05	Falcon 9	3681.000000	MEO	CCAFS SLC 40	1

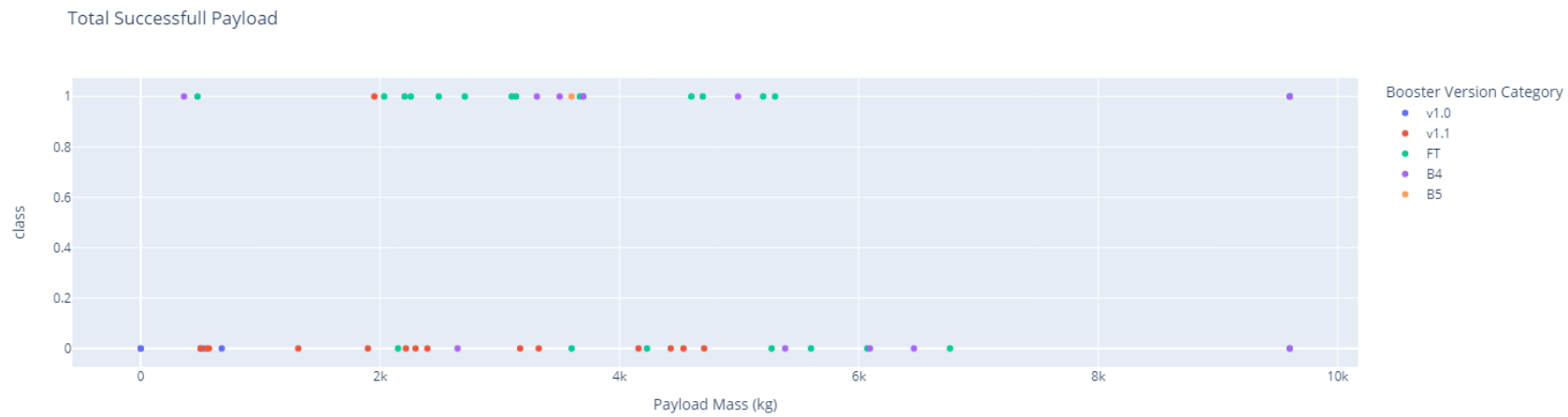
90 rows × 7 columns

GitHub --- https://github.com/Mikey3300/IBM-Capstone.git

Results



INTERACTIVE ANALYTICS DEMO IN SCREENSHOTS



https://kienquoc3300-8050.theiadocker-3-labs-prod-theiak8s-4-tor01.proxy.cognitiveclass.ai/

Results

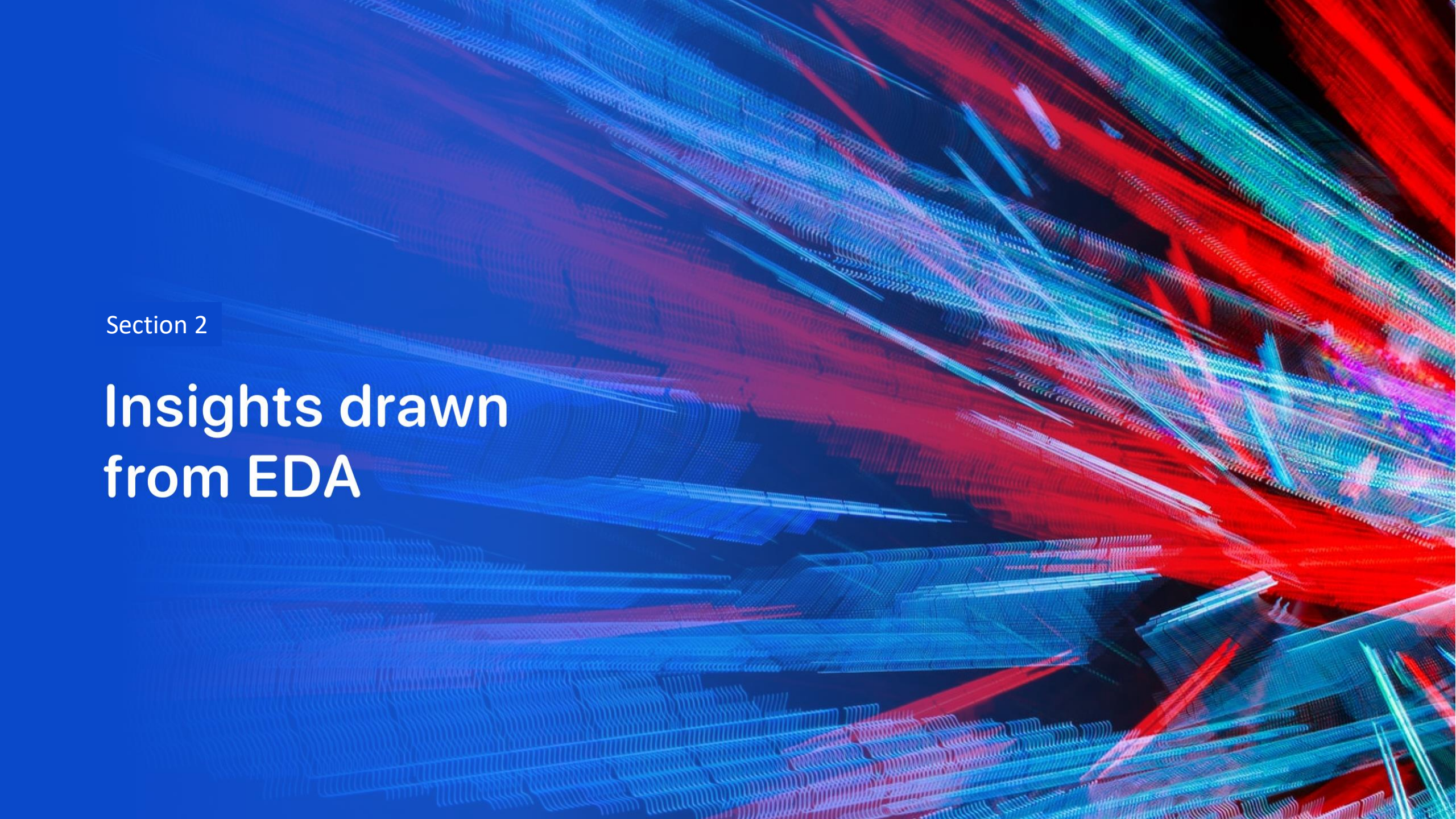
PREDICTIVE ANALYSIS RESULTS

```
print("Tree accuracy :",tree_cv.best_score_)
print("KNN accuracy :",knn_cv.best_score_)
print("Logression accuracy :",logreg_cv.best_score_)
print("SVM accuracy :",svm_cv.best_score_)
```

```
Tree accuracy : 0.8892857142857145
KNN accuracy : 0.8482142857142858
Logression accuracy : 0.8464285714285713
SVM accuracy : 0.8482142857142856
```

The best performing methods are LogisticRegression, SVM,

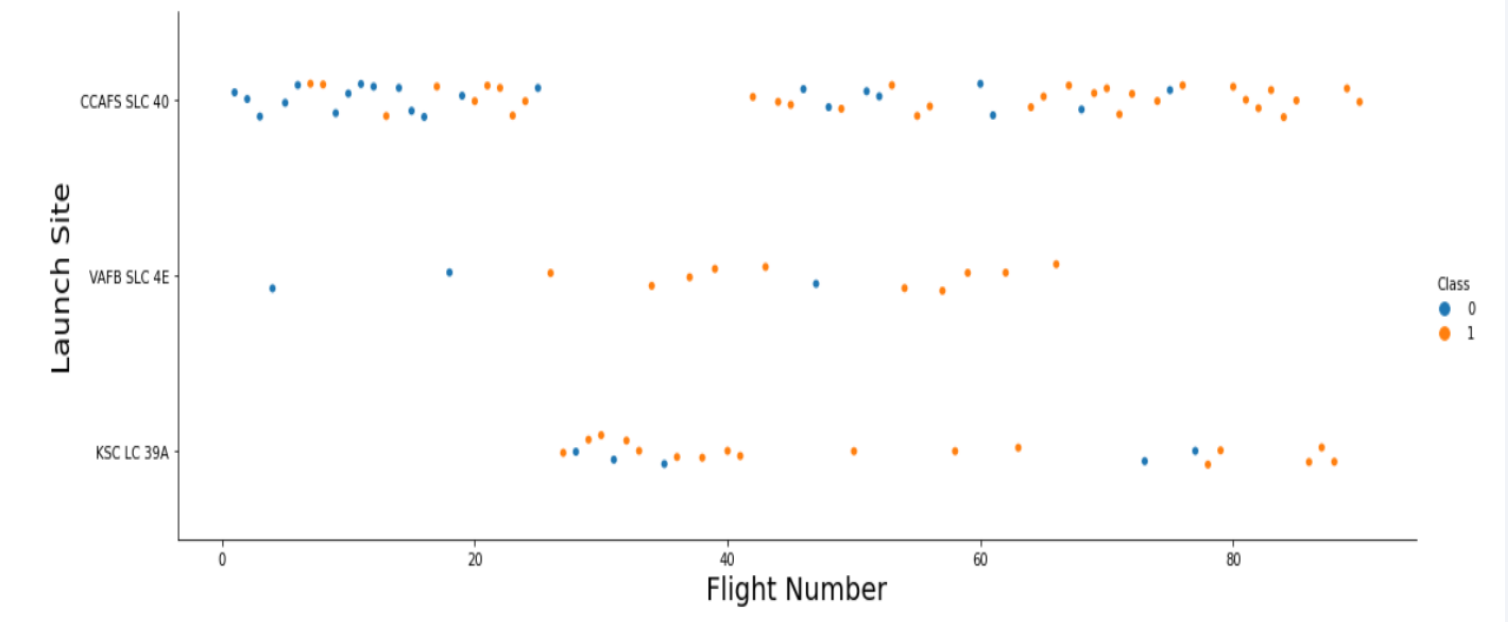
GitHub --- https://github.com/Mikey3300/IBM-Capstone.git


Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

Flight Number vs Launch Site

- Each Launch Site has a significantly different amount of Flights
  - **CCAFS LC-40** has the highest number of flights and it lies down respectively from number 0 to 90
  - **VAFB SLC 4E** can be said as a Launch Site that has a least amount of flights but the success rate is still in a good condition.
  - While **KSC LC 39A**, as we can see, it does not have any flight from number 0 to 23.

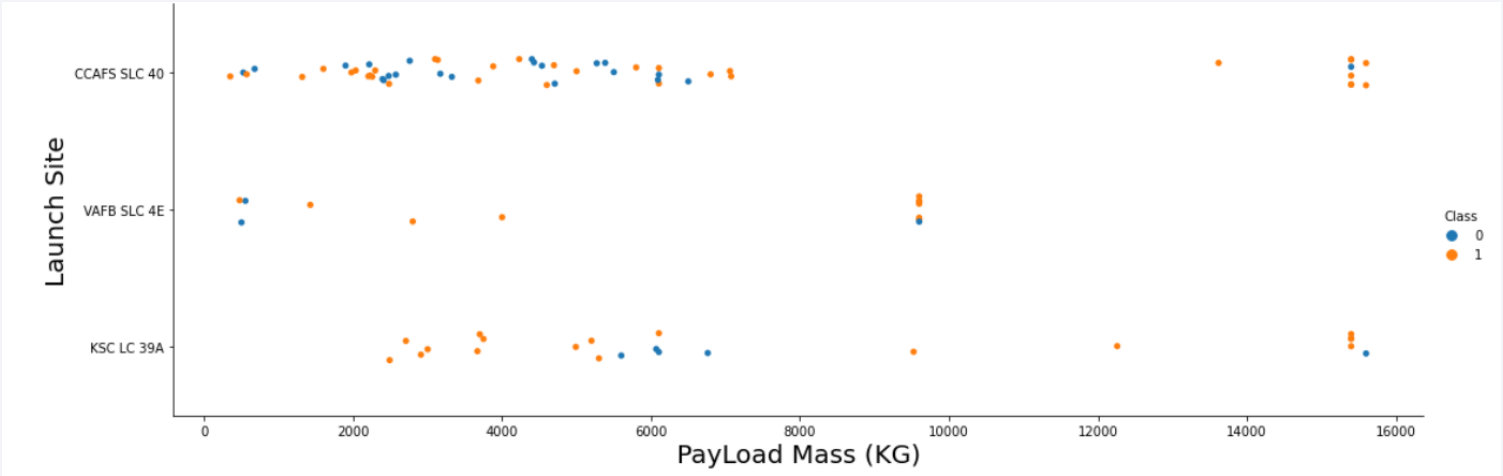


GitHub --- https://github.com/Mikey3300/IBM-Capstone.git

Payload vs. Launch Site

Payload vs Launch Site

- Notice from the chart, the **CCAFS SLC 40** can be able to store strongly payloads **from 100 to approximately 7500 KG**, while the strength in that range decreases in other launches.
- As we can see, the **VAFB SLC 4E** Launch Site **does not have any rocket store more than 10000 KG**, while two other launch are available for heavy payload at some rockets.

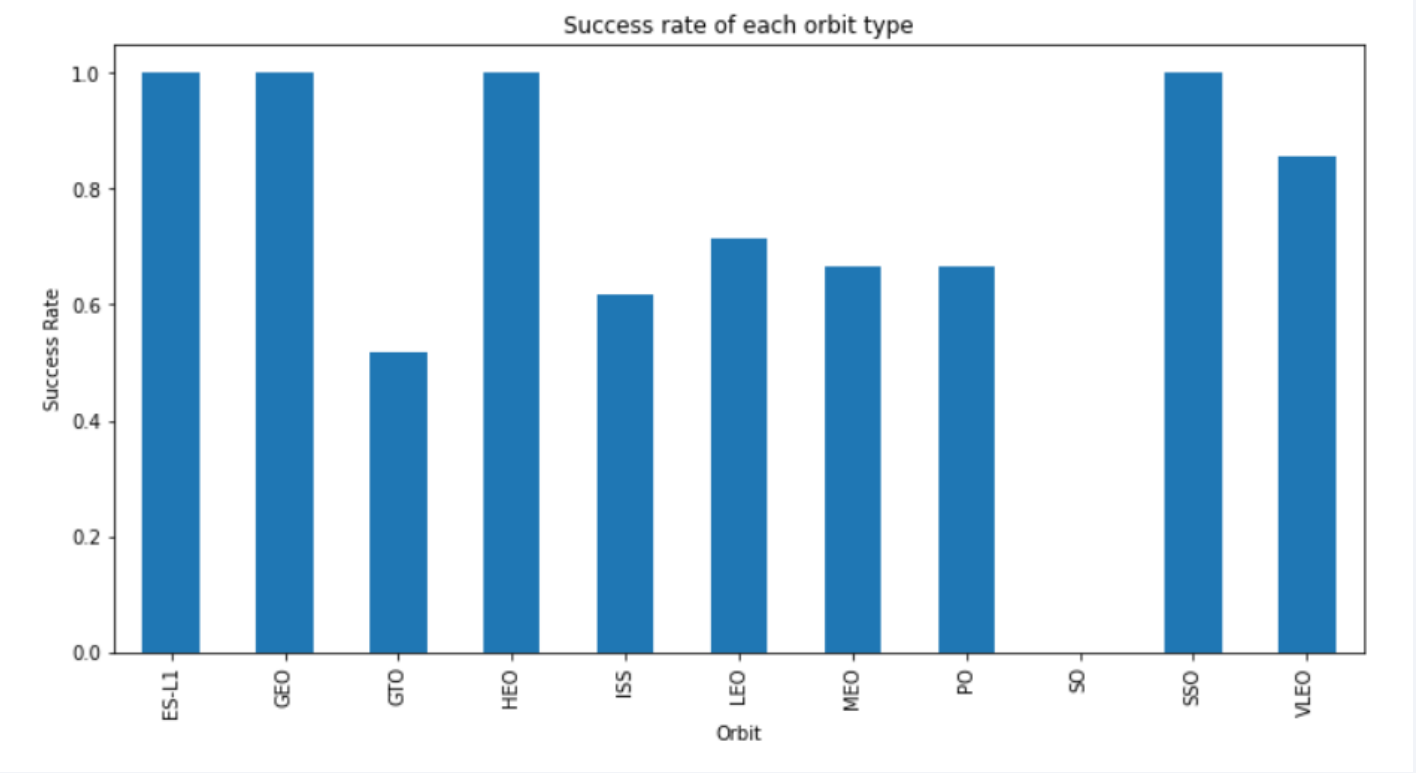


GitHub --- https://github.com/Mikey3300/IBM-Capstone.git

Success Rate vs. Orbit Type

Success rate of each orbit type

- There are some noticeable observations in the chart.
  - **Orbit type – SO** does not contain any success rate which means it is failed to land and might continue if we use this type.
  - **ES-L1, GEO, HEO and SSO** have the highest landing success rate.
  - **GTO** has the lowest success rate after excluding the zero-rating **SO**.

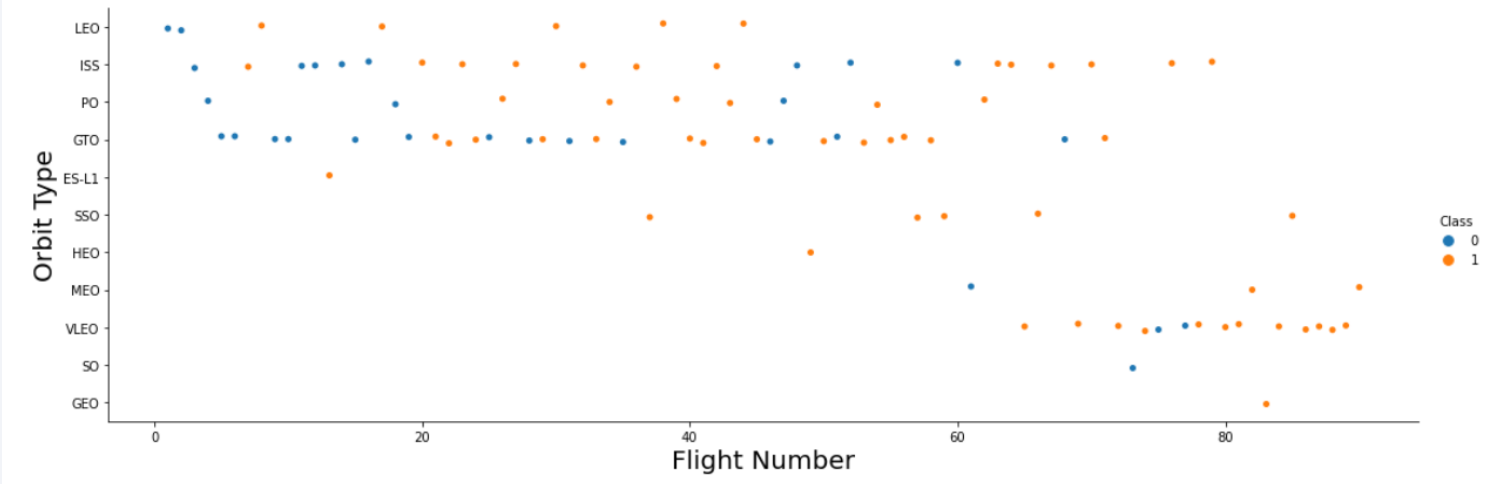


GitHub --- https://github.com/Mikey3300/IBM-Capstone.git

Flight Number vs. Orbit Type

Flight number vs Orbit type

- The Success of EO orbit appears related to the number of flights. On the other hand, it does not seem to be related to the flight number on the GTO orbit.
- Other orbit below after excluding LEO, ISS, PO, GTO can be seen as they do not have any flight from number 0 to 50.
- ES-L1 has only one flight number and previously, we can see it has 100% success rate. Therefore, this orbit still misses information to conclude is it good Orbit.

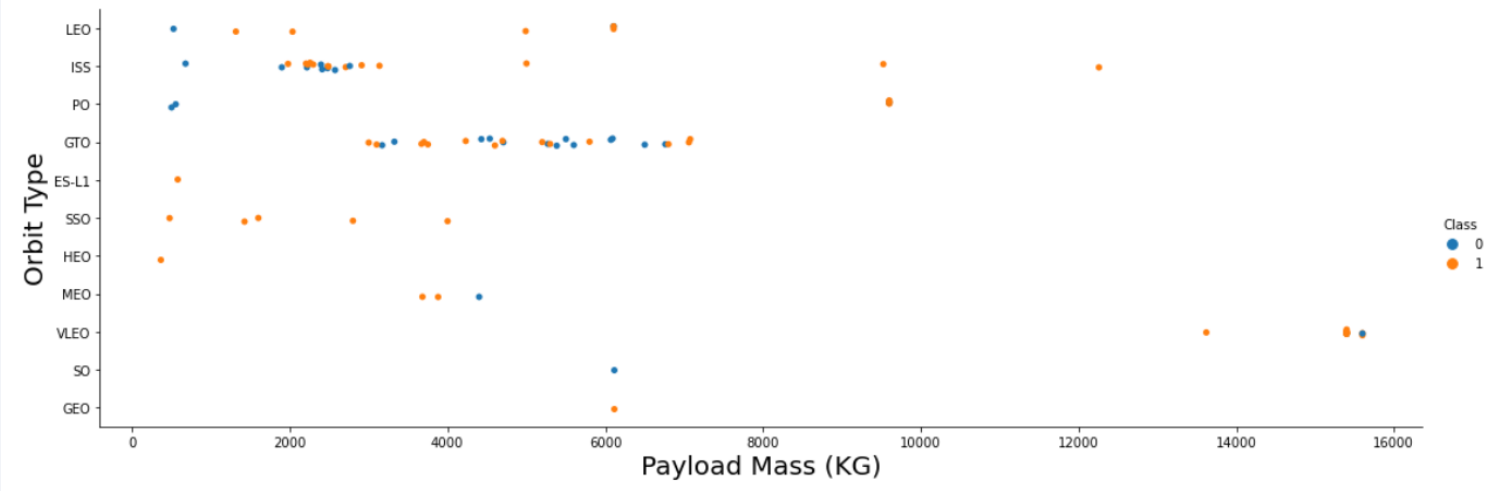


GitHub --- https://github.com/Mikey3300/IBM-Capstone.git

Payload vs. Orbit Type

Payload vs orbit type

- For heavy payloads, Polar, LEO, and ISS have higher success or positive landing rates.
- However, GTO has both positive and negative landing rates (failed missions), so it's hard to distinguish.

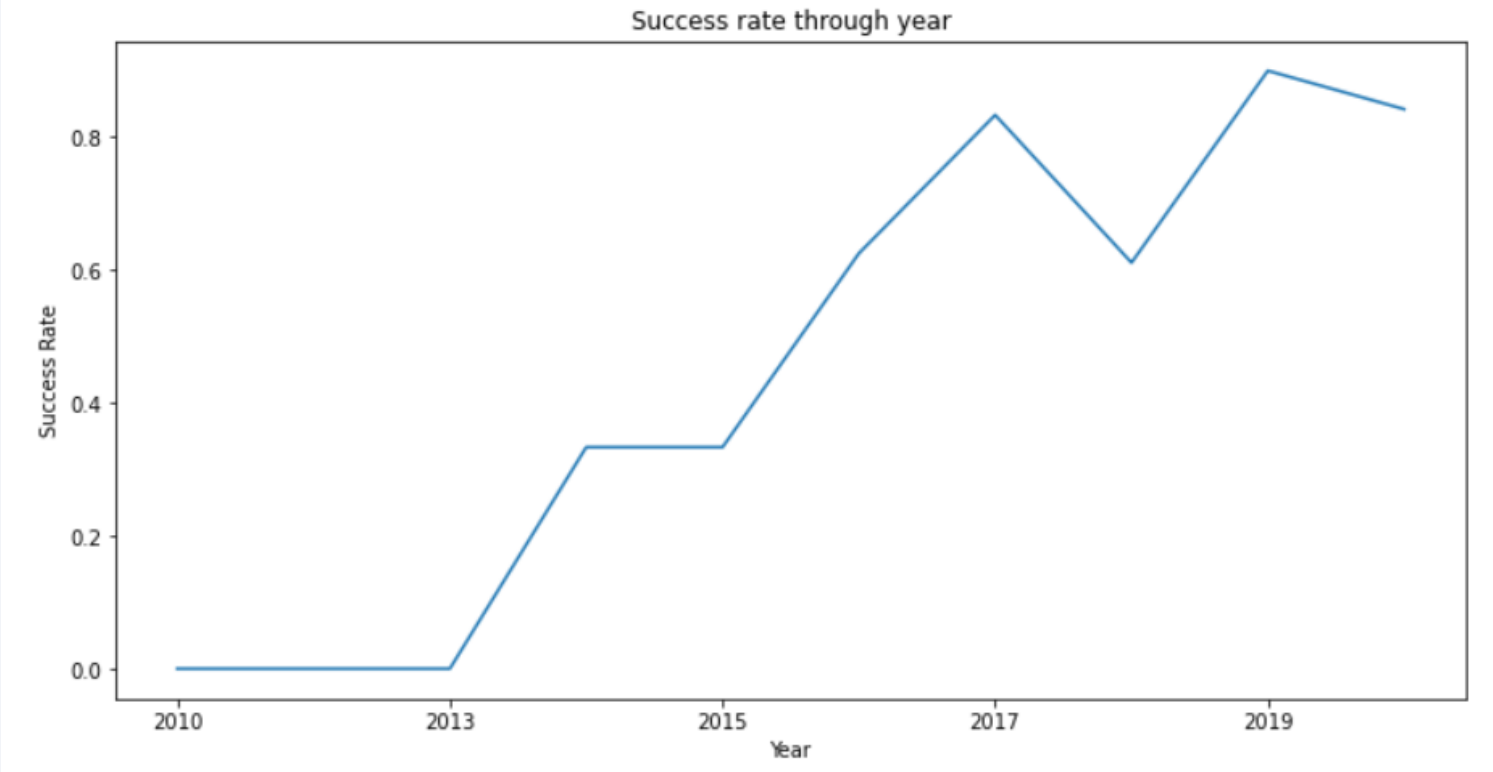


GitHub --- https://github.com/Mikey3300/IBM-Capstone.git

Launch Success Yearly Trend

Yearly average success rate

- As we can see, the trend has increased respectively since 2013 and reach the peak at 2019 before decreasing at the moment.



GitHub --- https://github.com/Mikey3300/IBM-Capstone.git

All Launch Site Names

- There are 4 unique Launch Site in SPACEX corporation
  - CCAFS LC-40 is short for Cape Canaveral Launch Complex 40
  - VAFB SLC 4E is short for Vandenberg Space Launch Complex 4
  - KSC LC 39A is short for Kennedy Space Center Launch Complex 39A
  - CCAFS SLC 40 is short for Cape Canaveral Space Launch Complex 40

Click in the underline text for additional details about each Launch Site

GitHub --- https://github.com/Mikey3300/IBM-Capstone.git

Launch Site Names Begin with 'CCA'

- Launch Site Names with CCA including LC (Launch Complex) and SLC (Space Launch Complex), but in the information below, we only take 5 records, therefore, it only contain details about Launch Site CCAFS LC-40.

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

GitHub --- https://github.com/Mikey3300/IBM-Capstone.git

Total Payload Mass

- Total Payload Mass when Customer is **NASA Commercial Resupply Services (CRS)** is **45,596 KG**

GitHub --- https://github.com/Mikey3300/IBM-Capstone.git

Average Payload Mass by F9 v1.1

- **The average payload mass** carried by booster version F9 v1.1 is **2,534.67 KG**

GitHub --- https://github.com/Mikey3300/IBM-Capstone.git

First Successful Ground Landing Date

- The first successful landing was coming from Ground Pad on 01-05-2017

GitHub --- https://github.com/Mikey3300/IBM-Capstone.git

Successful Drone Ship Landing with Payload between 4000 and 6000

- There are 4 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

GitHub --- https://github.com/Mikey3300/IBM-Capstone.git

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes are 101
  - There is 1 Failure in Flight
  - There are 100 Successful outcomes.
    - In General, we have 99 successes and 1 success in specific which comes from payload status unclear

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

GitHub --- https://github.com/Mikey3300/IBM-Capstone.git

Boosters Carried Maximum Payload

- The Booster Versions which have carried the maximum payload mass, are calculated as **12 Booster Versions** which comes from 2 specific Launch Sites such as **CCAFS SLC-40 and KSC LC-39A.**
- The maximum payload mass is **15,600 KG**

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

GitHub --- https://github.com/Mikey3300/IBM-Capstone.git

2015 Launch Records

- Months including Failure Landing Outcome from Drone Ship are 1 and 4. The Booster Version they used are F9 v1.1 B1012 and F9 v1.1 B1012, all of these rockets are coming from CCAFS LC-40 Launch Site

Month	Landing _Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

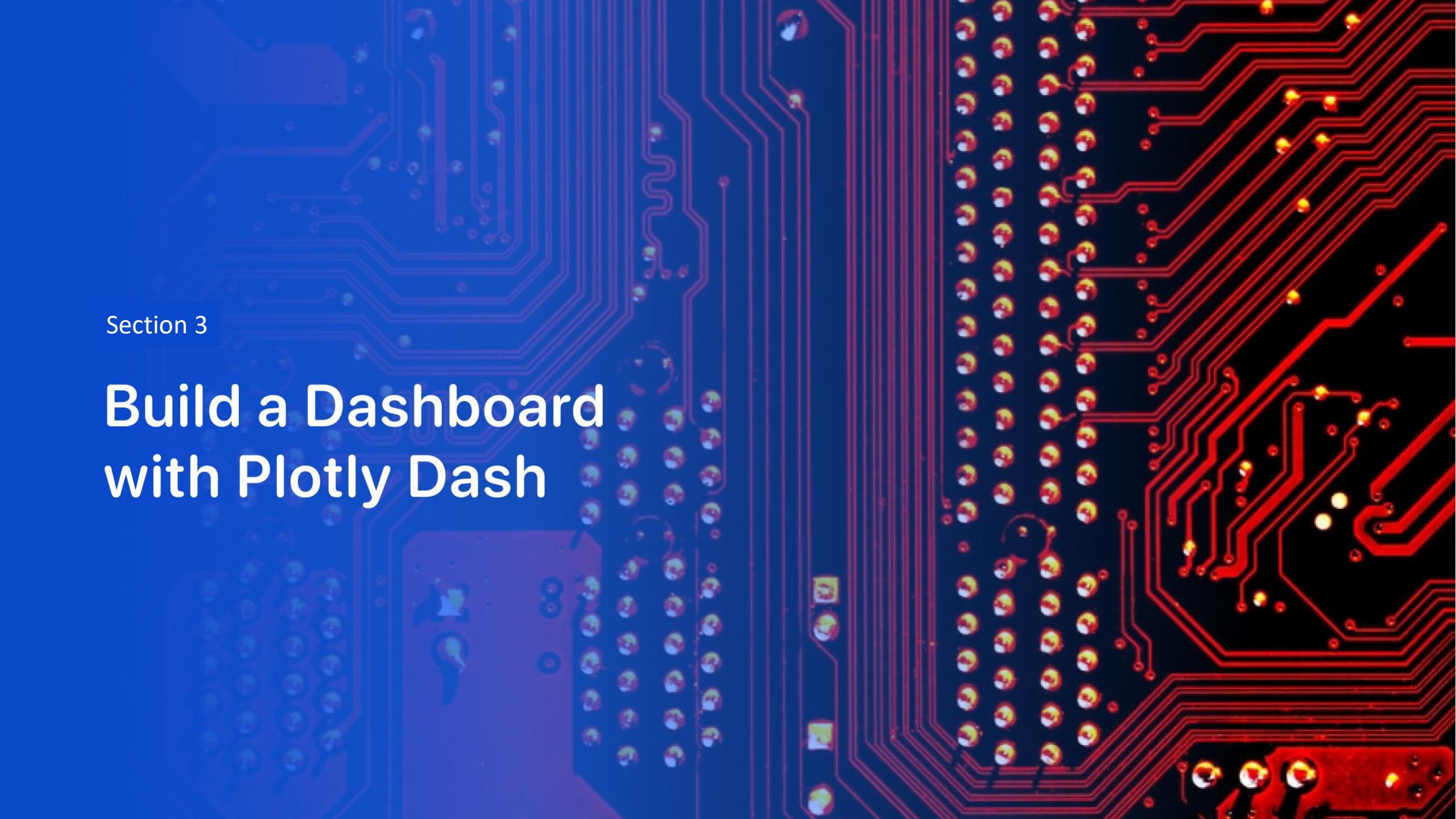
GitHub --- https://github.com/Mikey3300/IBM-Capstone.git

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

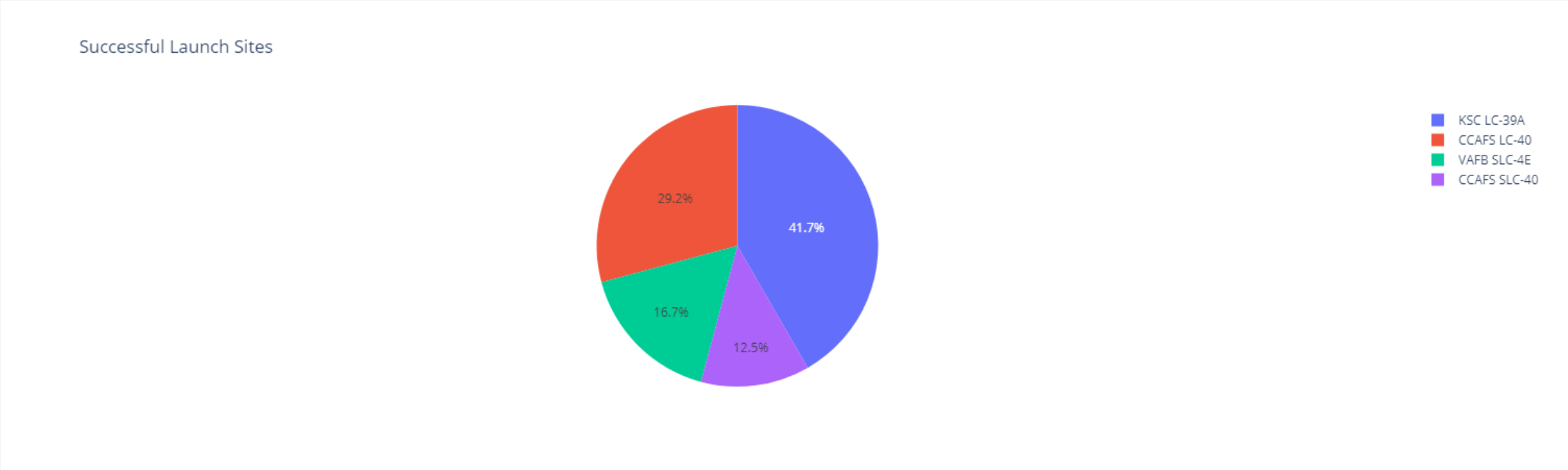
- The count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20 are ranking as:
  - Success in General have the highest rank with 20 times.
  - After that, we have Successful Outcome in Drone Ship at the second with 8 times
  - Finally, 6 times successful outcomes at Ground Pad between the date 2010-06-04 and 2017-03-20

Landing _Outcome	Count
Success	20
Success (drone ship)	8
Success (ground pad)	6

GitHub --- https://github.com/Mikey3300/IBM-Capstone.git


Section 3
Build a Dashboard with Plotly Dash

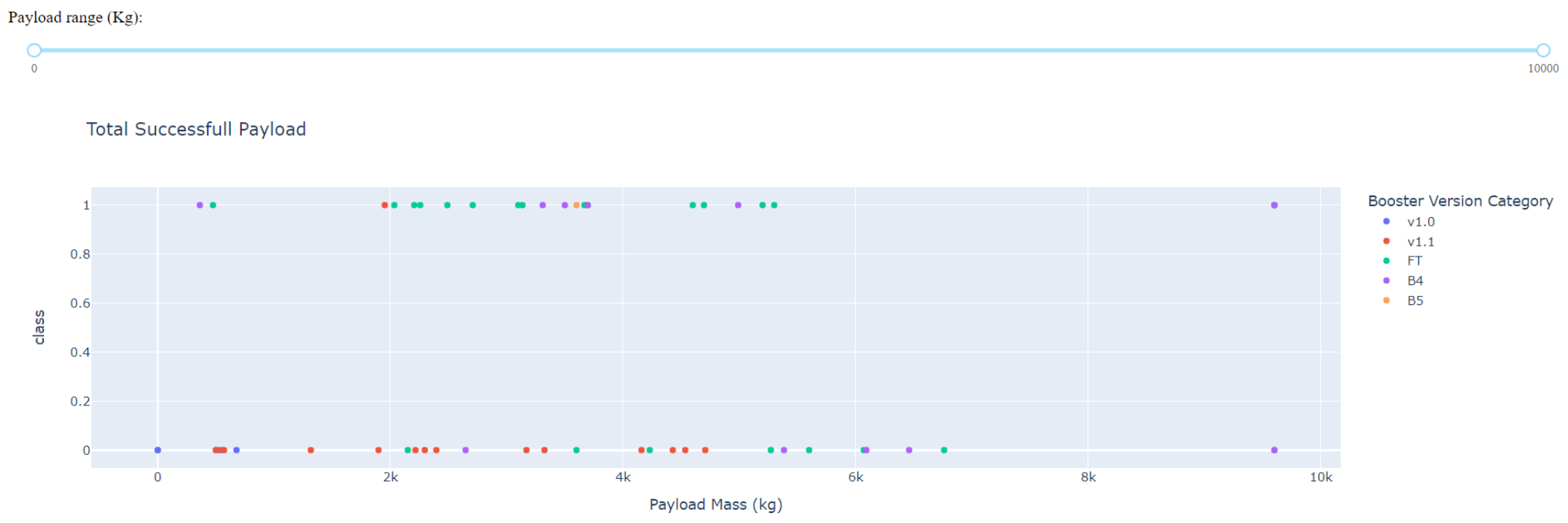
Dashboard for All Successful Launch Sites



The success rate from KSC LC-39A is demonstrated as a highest rate with 41.7%. The other positions from the Pie Chart are respectively from CCAFS LC-40, VAFB SLC-4E and CCAFS SLC-40 with the rating statistics 29.2%, 16.7% and 12.5%.

GitHub --- https://github.com/Mikey3300/IBM-Capstone.git

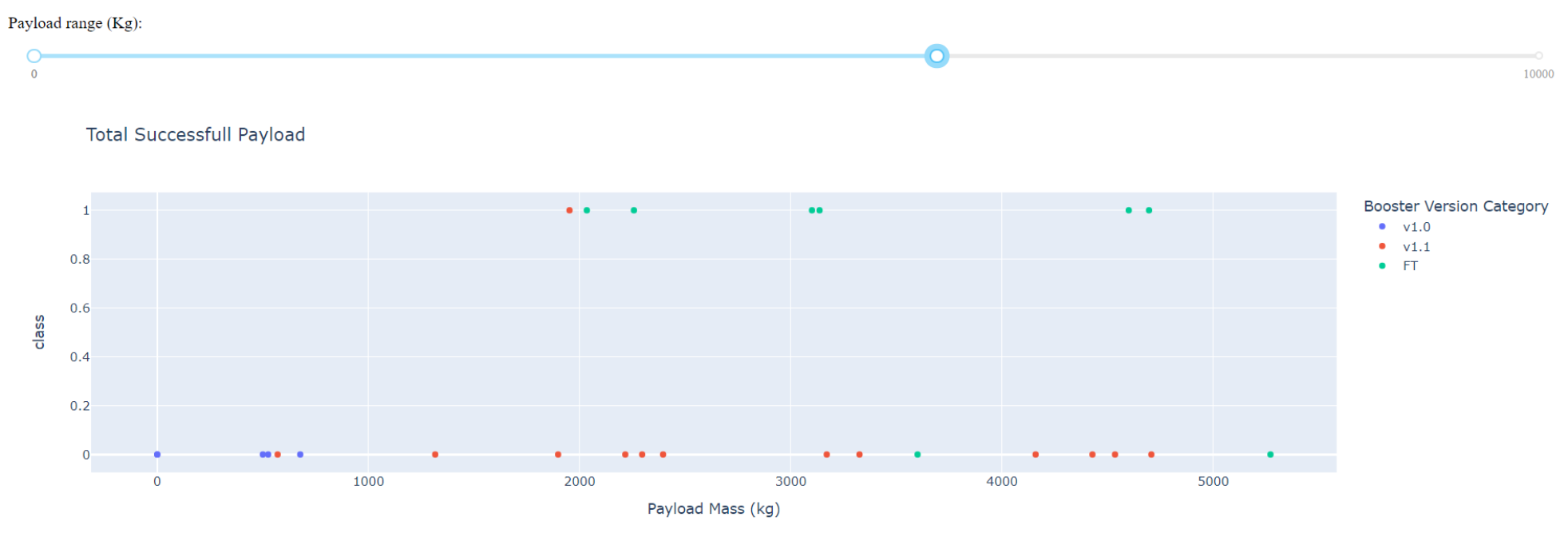
Dashboard for All Sites in Payload vs Launch Outcome



With all sites are selected and the range of Payload is from 0 to 10,000 KG, we can see the success landing happening mostly in F9 FT and F9 V1.0, while Failure appears more dots than Success and F9 v1.1 has the largest unsuccessful rate than others .

GitHub --- https://github.com/Mikey3300/IBM-Capstone.git

Dashboard for Specific Site in Payload vs Launch Outcome



With CCAFS LC-40 site is selected and the range of Payload is from 0 to 6000 KG, F9 FT has the largest success rate, while Failure appears mostly on F9 v1.1. The F9 v1.0 does not have any success rate in term of CCAFS LC-40 site.

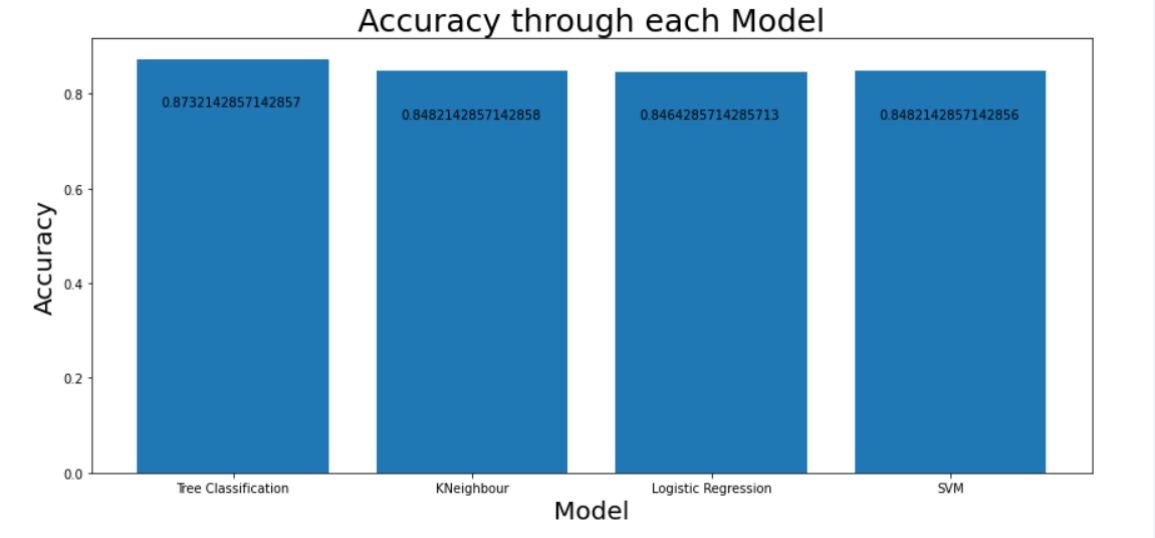
GitHub --- https://github.com/Mikey3300/IBM-Capstone.git

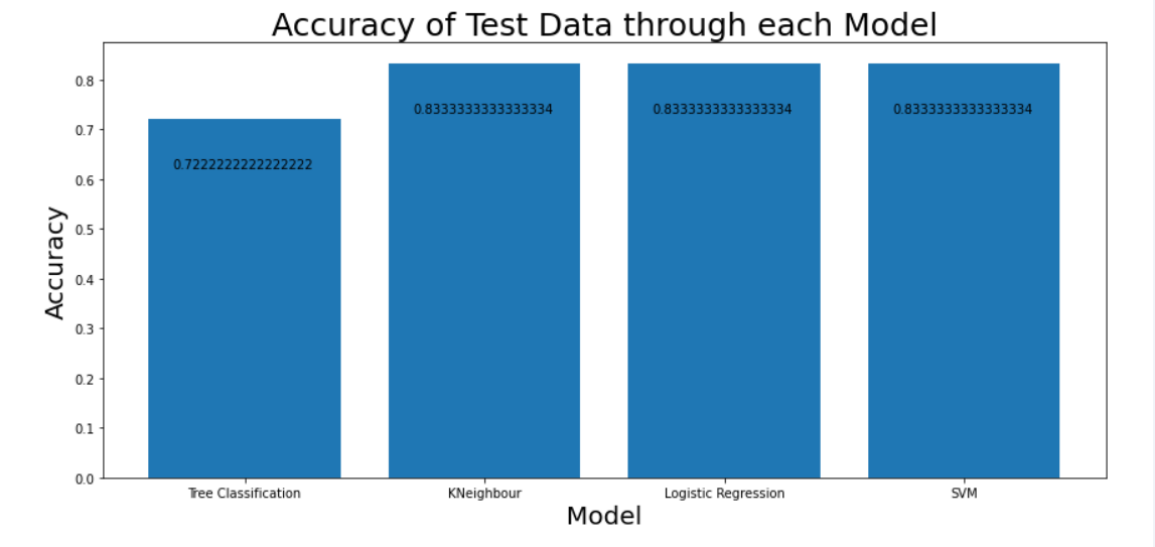
Section 4

Predictive Analysis (Classification)

Classification Accuracy

- The highest accuracy using training Data is belonged to Tree Classification. However, when use testing data to find accuracy, Tree model is seen as the lowest rate in accuracy.
- Therefore, by combining two training and testing data, we select the best models are KNN and SVM



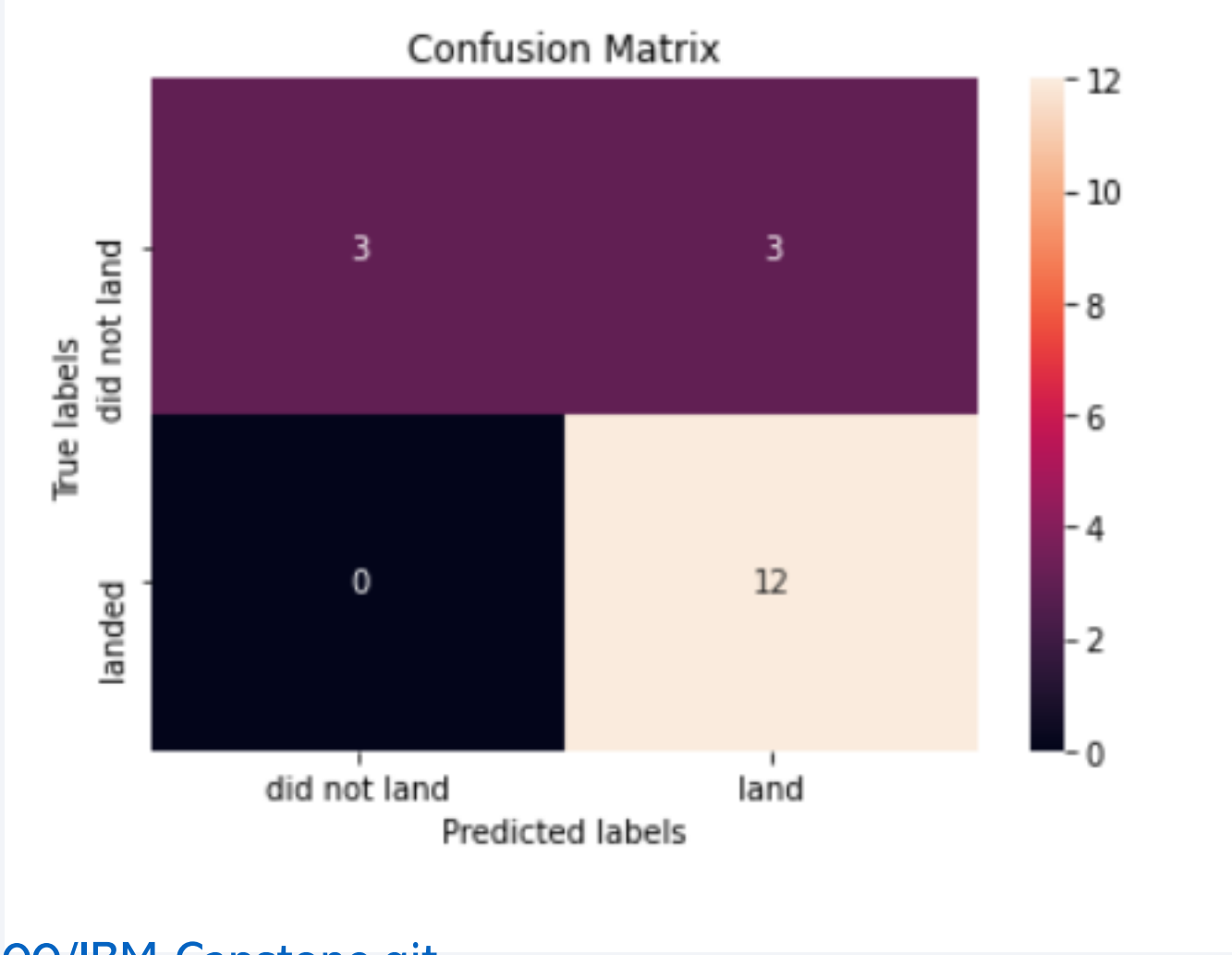


GitHub --- https://github.com/Mikey3300/IBM-Capstone.git

Confusion Matrix

KNN and SVM's

Confusion Matrix



GitHub --- https://github.com/Mikey3300/IBM-Capstone.git

Conclusions

- For Launch Site, we rank from the best performance to the least are respectively **KSC LC-39A, CCAFS SLC-40** and **VAFB SLC-4E**, due to the amounts of rocket, the ability of carrying heavy payload, success rate.
- For Orbit Type, we see **LEO Orbit** is a good implementation because it has a lot of flights with great success rate, also considering in Payload, it has excellent success rate as well.
- For Booster Version, we observe **the F9 FT** performs excellent in carrying (heavy) payload with success rate at a high level. In drone and ground pad, it also has high rate in landing perfectly without any damage.
- For futuristic prediction of successful landing, we will use **KNN and SVM** models because they have been observed as models with highest accuracy in both Test and Train Data.

GitHub --- https://github.com/Mikey3300/IBM-Capstone.git

Appendix

- Python code snippets, SQL queries, charts, Notebook outputs, and data sets are all included in my repository at GitHub in the link below.

→https://github.com/Mikey3300/IBM-Capstone.git

Thank you!