IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ibe Ezuma
2024-04-28

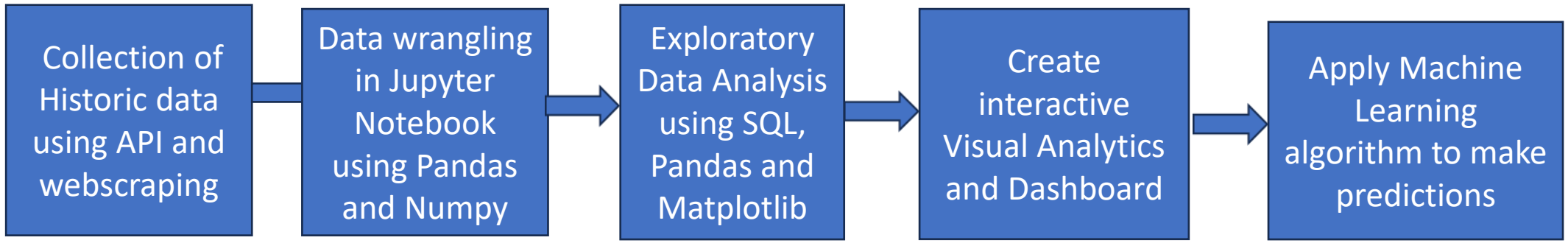
Outline

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

Summary of results

- From the study findings, after performing Optimize Hyperparameter grid search, the Support Vector Machine(SVM) is the model with most accuracy of 0.88, and it made a prediction of 83.33% success rate of Falcon 9 first stage landing.

Summary of Methodology



Introduction

- 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. Problems you want to find answers
- We will predict if the Falcon 9 first stage will land successfully. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

Section 1

Methodology

Methodology

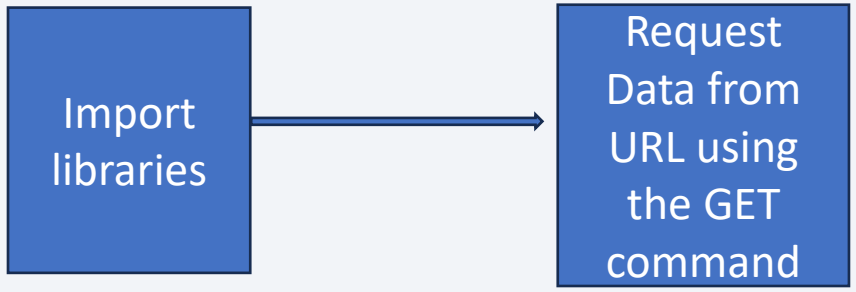
Executive Summary

- 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 – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts

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

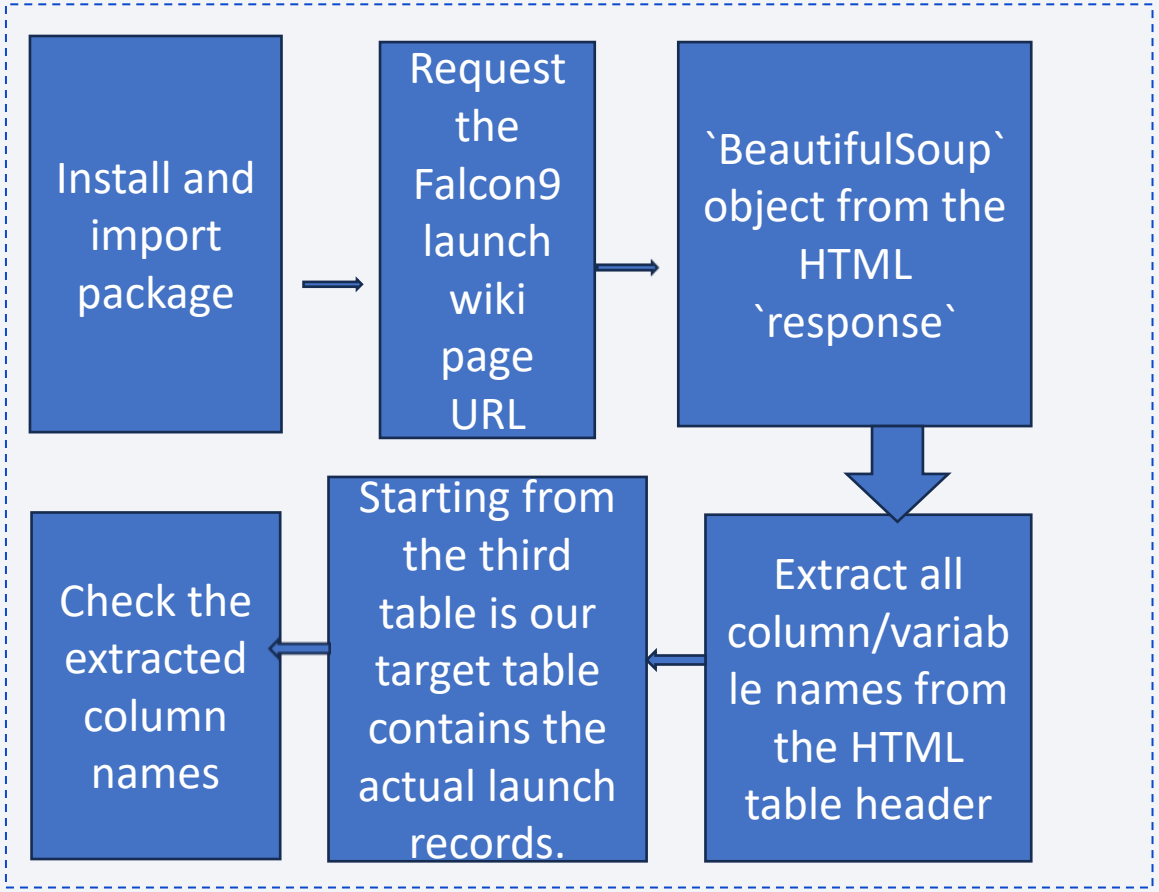


https://github.com/ibeezuma/Data-Science-Project

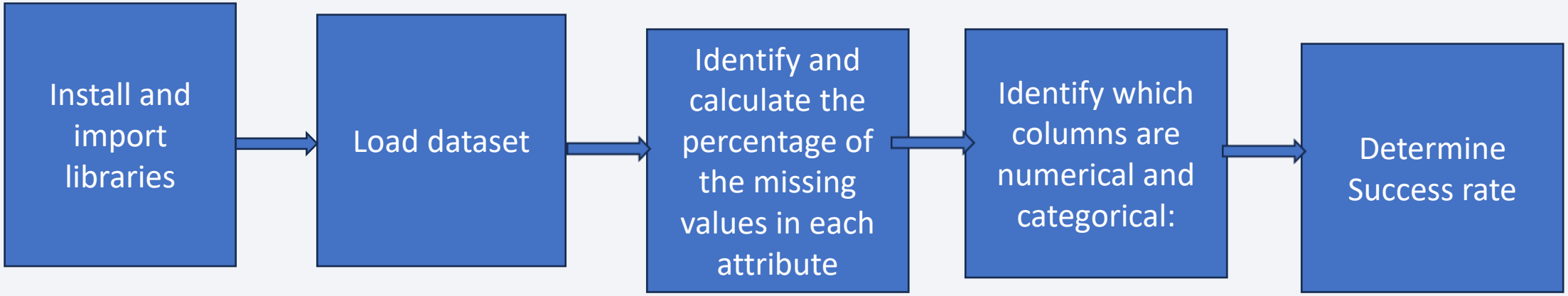
Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

- https://github.com/ibeezuma/Data-Science-Project



Data Wrangling



- https://github.com/ibeezuma/Data-Science-Project

EDA with Data Visualization

- Flight Number vs. Launch Site: To check for linear relationship
- Payload vs. Launch Site: To check for linear relationships
- Success Rate vs. Orbit Type: To check for orbit types with their success rates
- Flight Number vs. Orbit Type: To check for linear relationship
- Launch Success Yearly Trend: To check for yearly trends in success rates

- https://github.com/ibeezuma/Data-Science-Project

EDA with SQL

- Selecting All Launch Site Names
- Selecting Launch Site Names Begin with 'CCA'
- Calculating Total Payload Mass
- Calculating the Average Payload Mass of booster version F9 v1.1
- First Successful Ground Landing Date
- Successful Drone Ship Landing with Payload between 4000 and 6000
- Total Number of Successful and Failure Mission Outcomes
- 2015 Launch Records
- Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

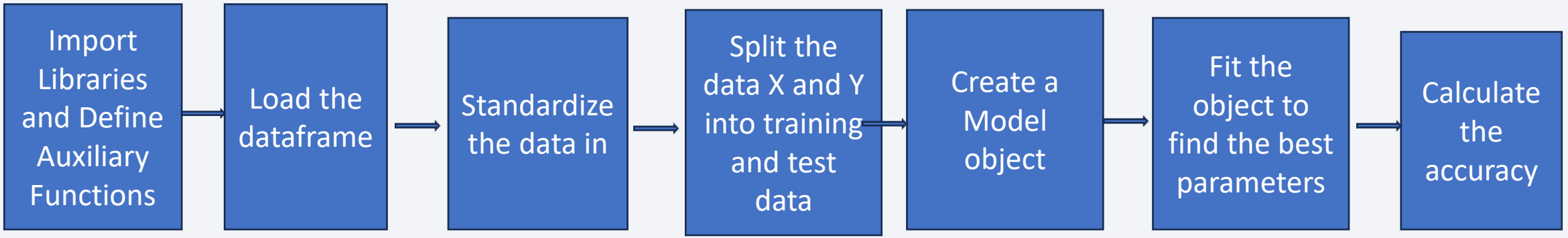
- https://github.com/ibeezuma/Data-Science-Project

Build an Interactive Map with Folium

- The four launch sites circles on the map and a marker
- Success/Failed Launches for each site on the Map
- Launch site VAFBSLC4E is located at 1.8km to the coastline

- https://github.com/ibeezuma/Data-Science-Project

Predictive Analysis (Classification)

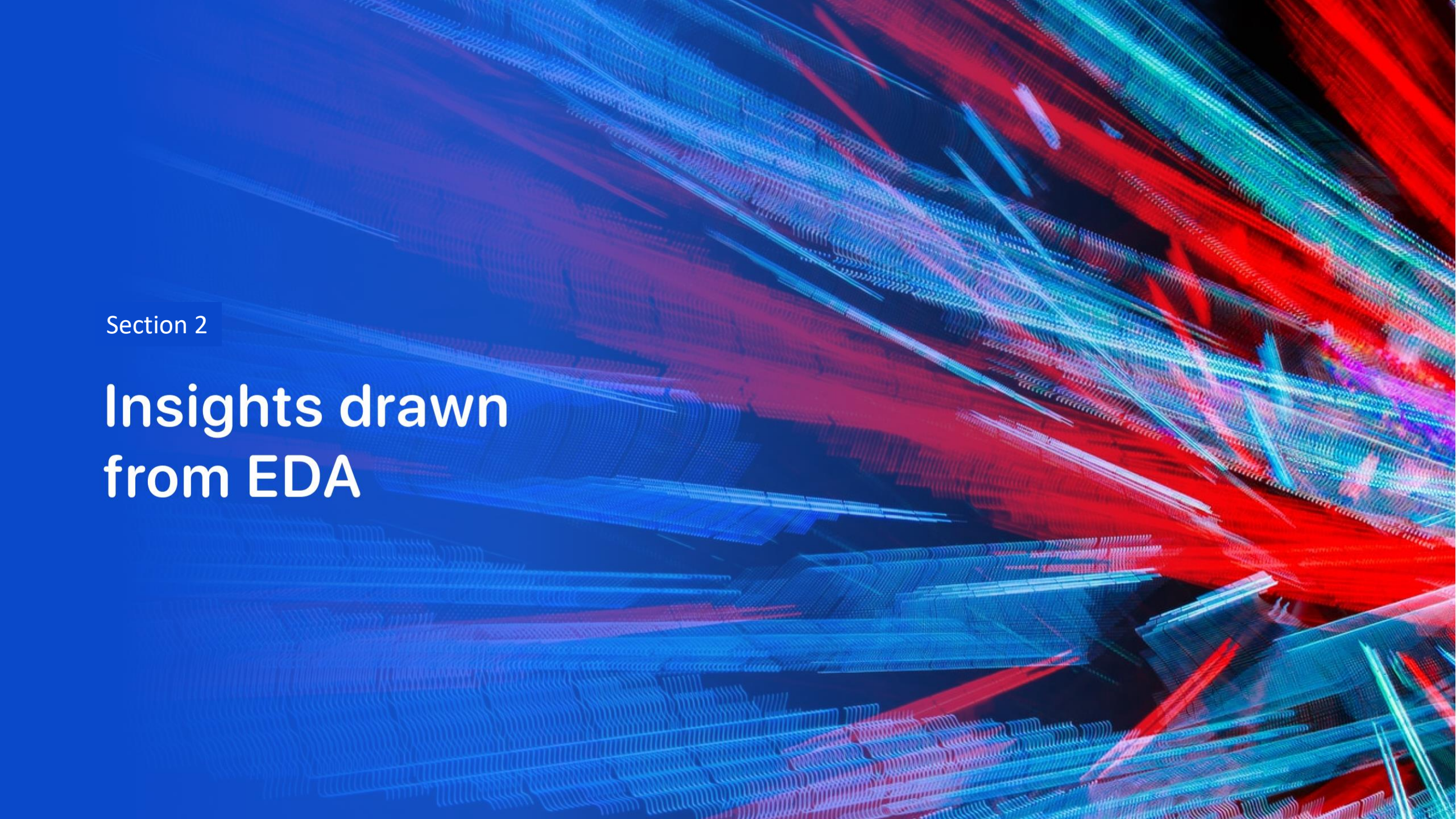


- https://github.com/ibeezuma/Data-Science-Project

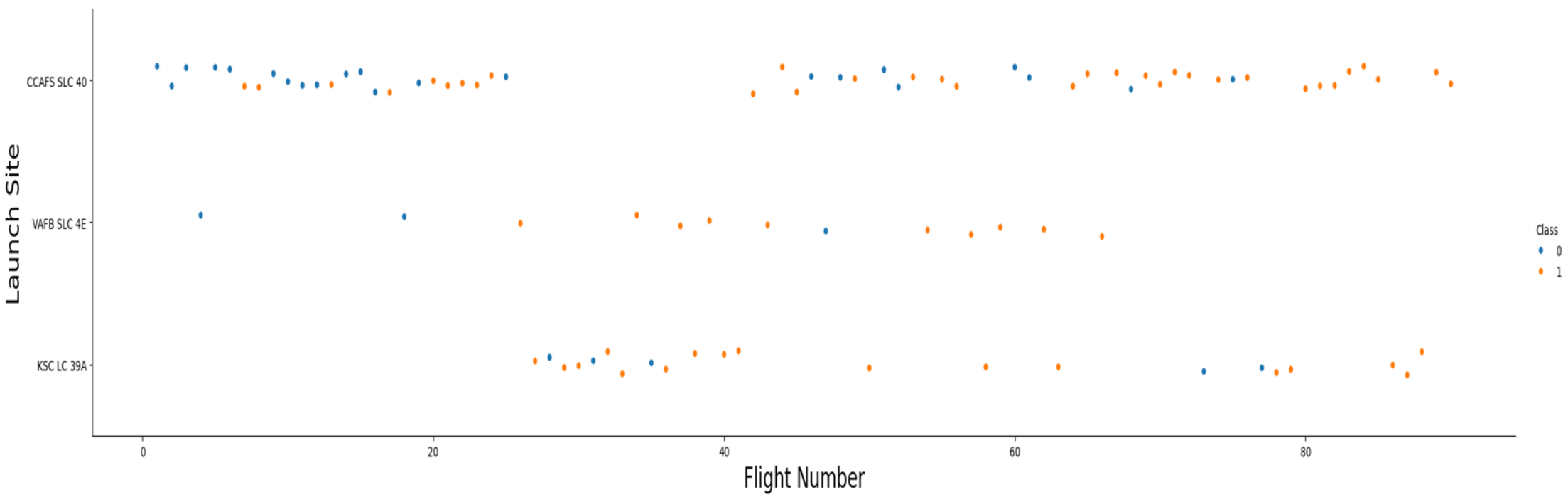
Results

Predictive analysis results

Model	Accuracy score
Support Vector Machine	0.88
Logistic Regression	0.875
Decision Tree	0.76
K- Nearest Neighbour	0.86

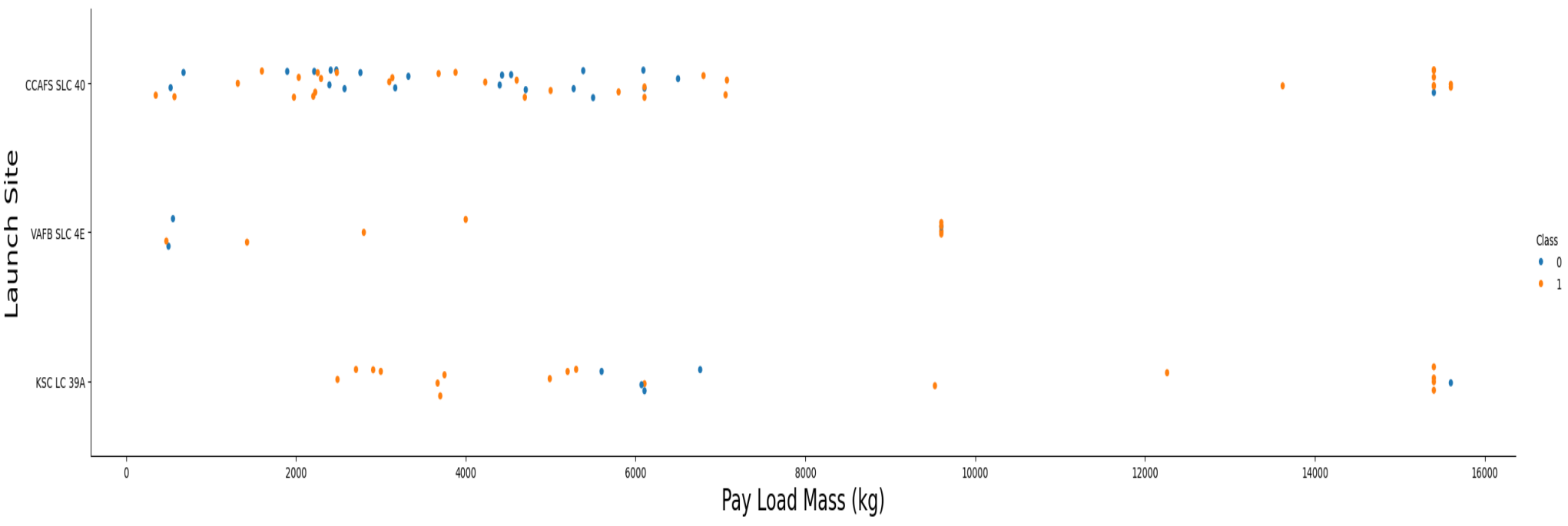

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



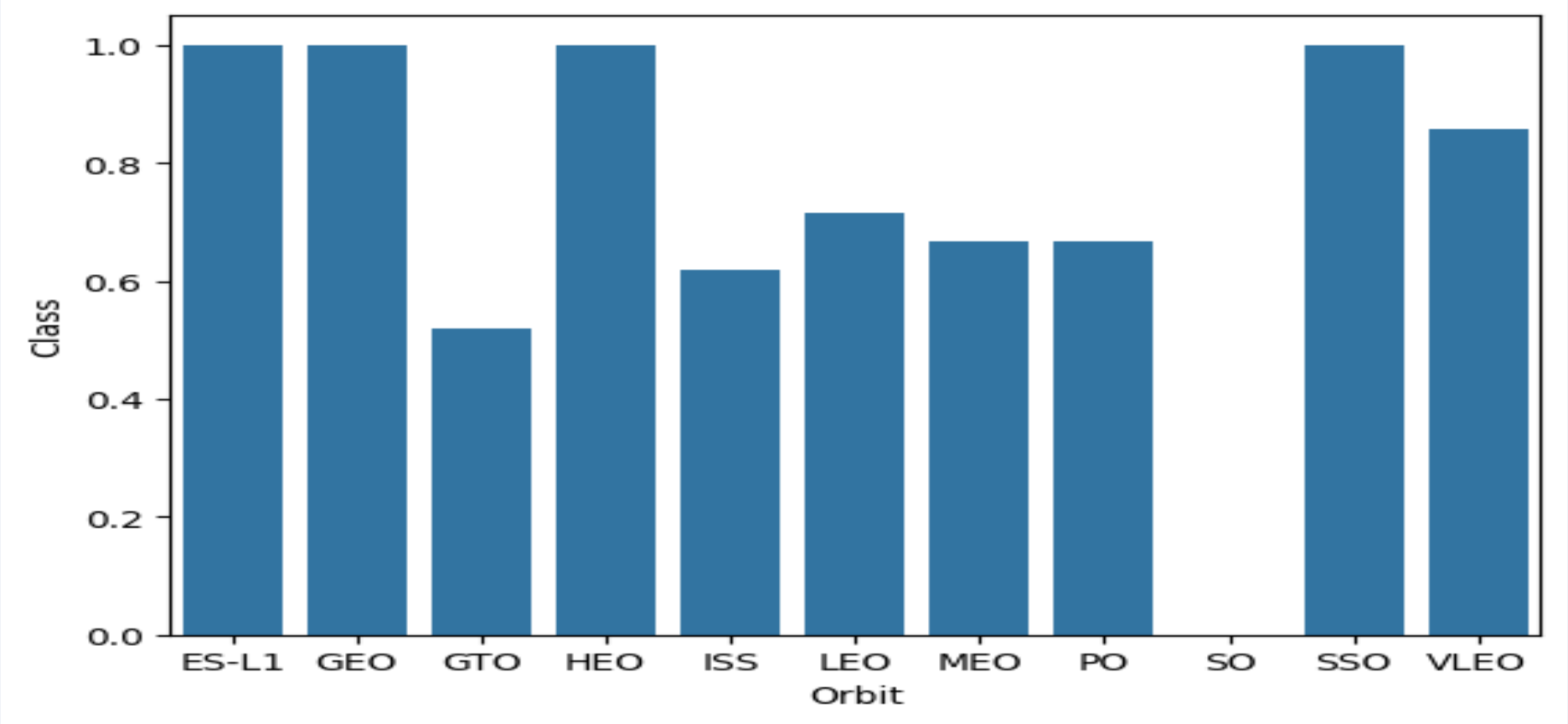
From the Scatterplot, there appears to be no correlation between Flight number and Launch site.

Payload vs. Launch Site



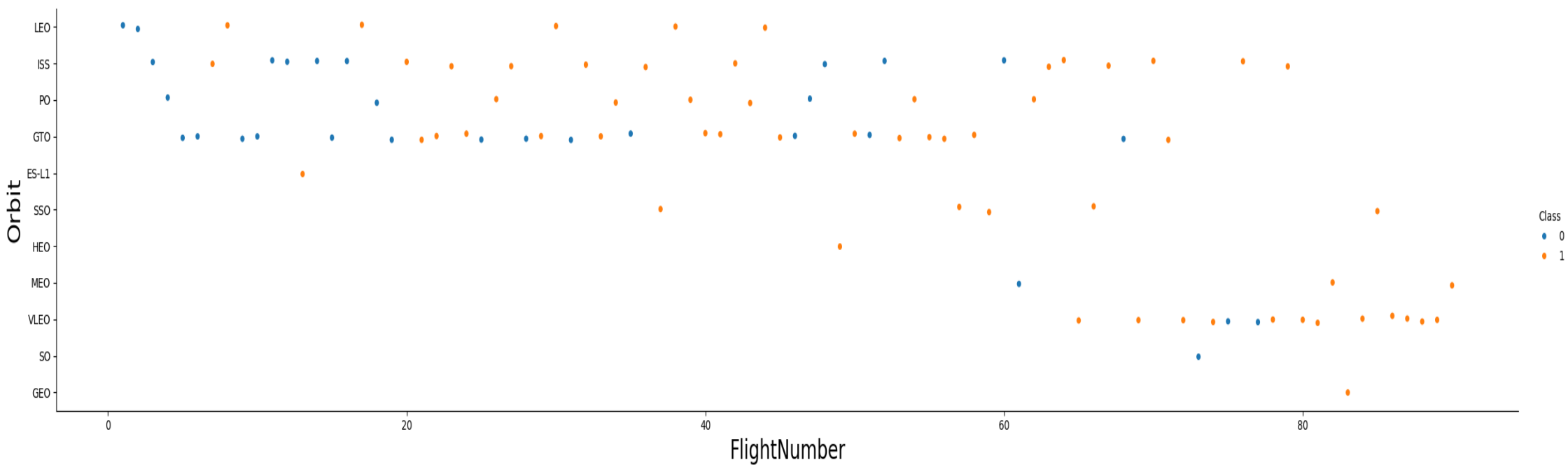
- From the Scatterplot, there appears to be no correlation between Playload and Launch site

Success Rate vs. Orbit Type



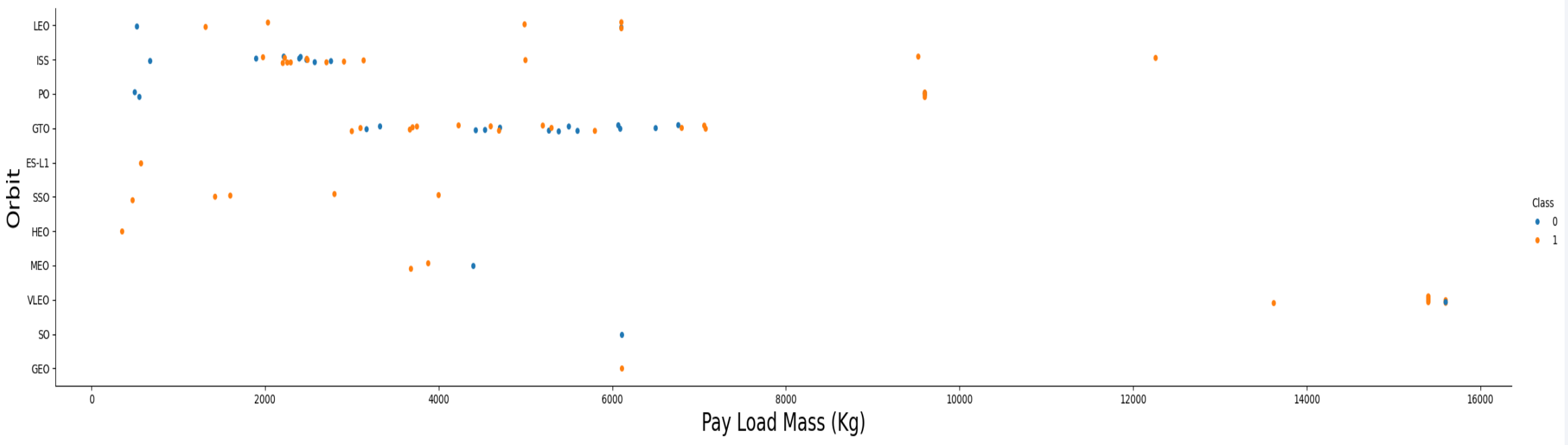
- From the bar chart ES-L1, GEO, HEO and SSO has the highest success rate, while so has no success

Flight Number vs. Orbit Type



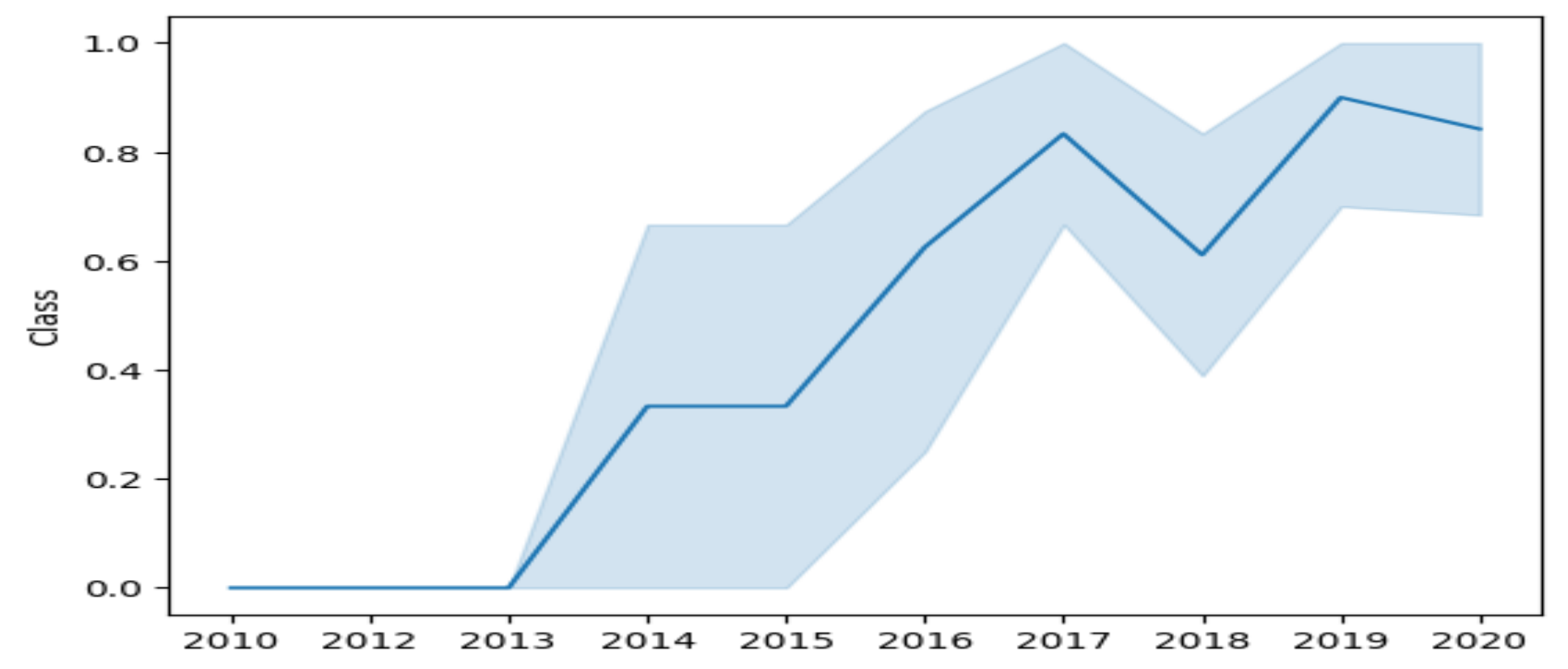
- No Correlation between Flight number and orbit Type

Payload vs. Orbit Type



- No Correlation between Payload and orbit Type

Launch Success Yearly Trend



- You can observe that the success rate since 2013 kept increasing till 2017 (stable in 2014) and after 2015 it started increasing.

All Launch Site Names

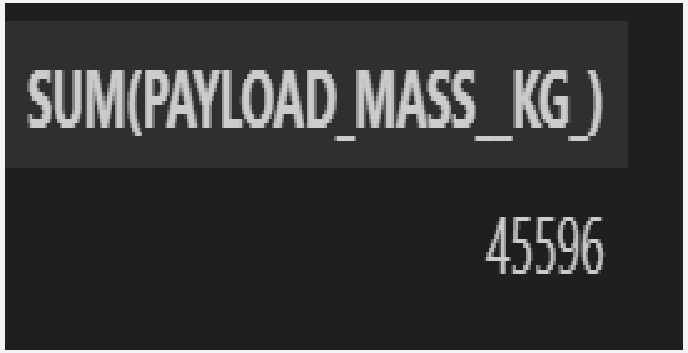
- %sql SELECT DISTINCT launch_site FROM SPACEXTABLE ;

Launch Site Names Begin with 'CCA'

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

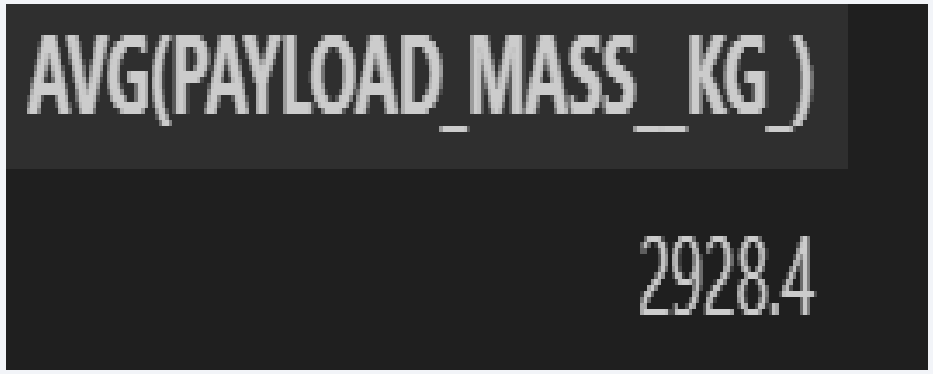
- %sql SELECT * FROM SPACEXTABLE where launch_site LIKE 'CCA%' ;

Total Payload Mass



- %sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE Customer ='NASA (CRS)' ;

Average Payload Mass by F9 v1.1



- %sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE Booster_Version ='F9 v1.1' ;

First Successful Ground Landing Date

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS_KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
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)

- %sql select * from SPACEXTABLE where Landing_Outcome== 'Success (ground pad)' ORDER BY Date ASC limit 5;

Successful Drone Ship Landing with Payload between 4000 and 6000

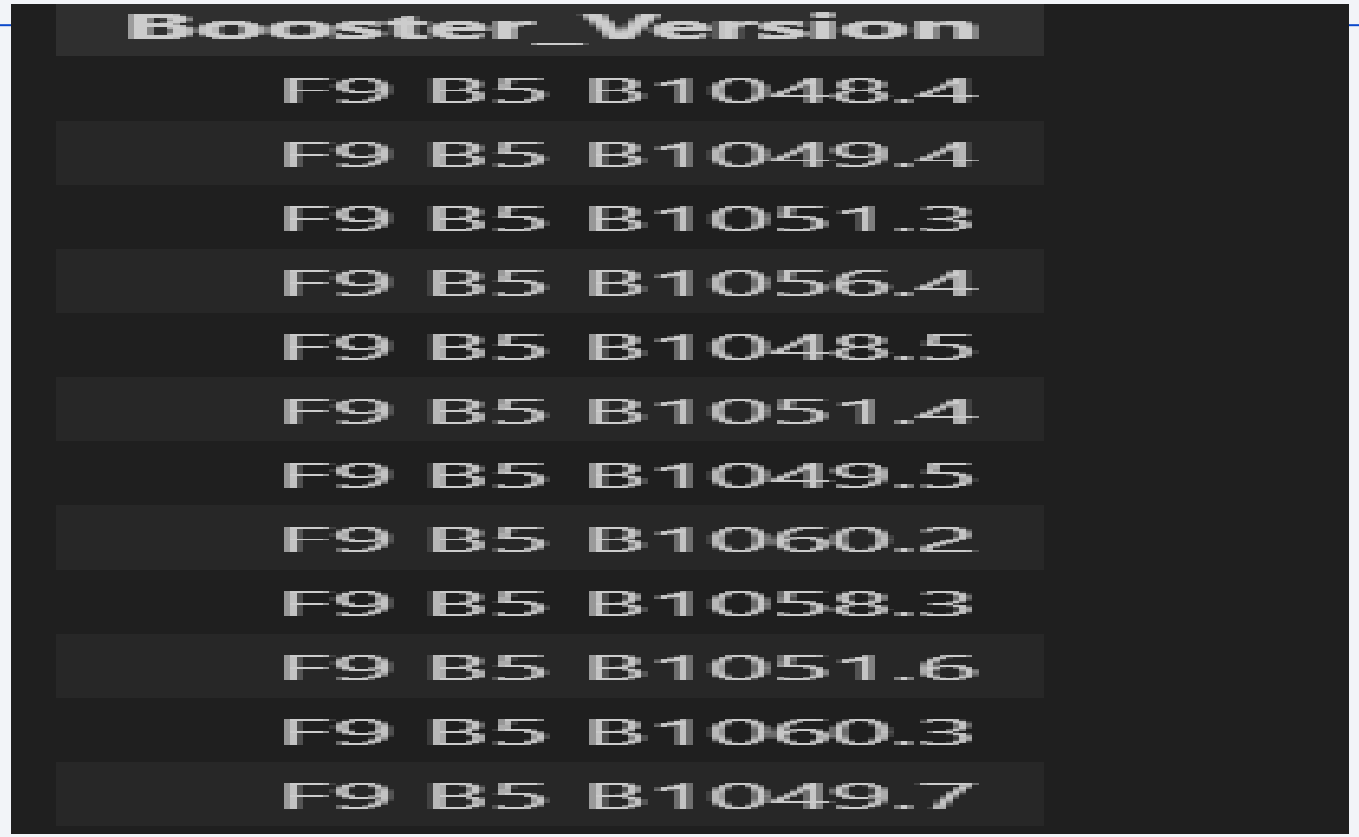
Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

- %sql SELECT Booster_Version FROM SPACEXTABLE WHERE Landing_Outcome=='Success (drone ship)'and PAYLOAD_MASS__KG_ between 4000 and 6000 ;

Total Number of Successful and Failure Mission Outcomes

- %sql SELECT Count(Mission_Outcome) from SPACEXTABLE where Mission_Outcome=='Success'or 'Failure'

Boosters Carried Maximum Payload



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

- %sql SELECT Booster_Version from SPACEXTABLE where PAYLOAD_MASS__KG_==(SELECT MAX(PAYLOAD_MASS__KG_) from SPACEXTABLE)

2015 Launch Records

Month	Booster_Version	Launch_Site
01	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40

- %sql SELECT substr(Date, 6,2) as Month, Booster_Version, Launch_Site from SPACEXTABLE where Landing_Outcome== 'Failure (drone ship)' and substr(Date, 0,5)= '2015';

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

Date	Landing_Outcome	Count
2012-05-22	No attempt	10
2016-04-08	Success (drone ship)	5
2015-01-10	Failure (drone ship)	5
2015-12-22	Success (ground pad)	3
2014-04-18	Controlled (ocean)	3
2013-09-29	Uncontrolled (ocean)	2
2010-06-04	Failure (parachute)	2
2015-06-28	Precluded (drone ship)	1

- %sql SELECT Date, Landing_Outcome, count(Landing_Outcome) as Count from SPACEXTABLE where Date between '2010-06-04' and '2017-03-20' Group by Landing_Outcome order by count Desc ;

Section 3

Launch Sites Proximities Analysis

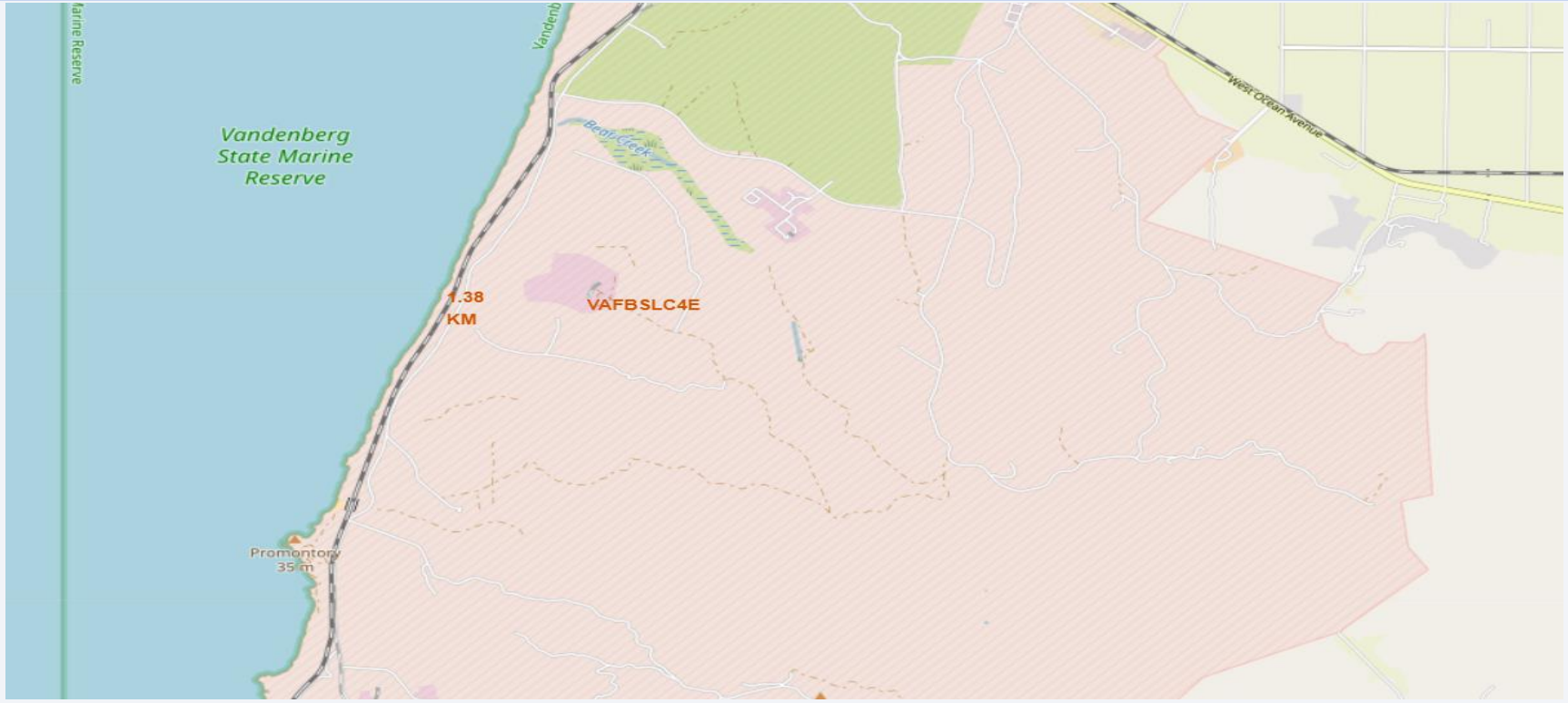
Map showing Launch Sites

- The four launch sites circles on the map and a marker

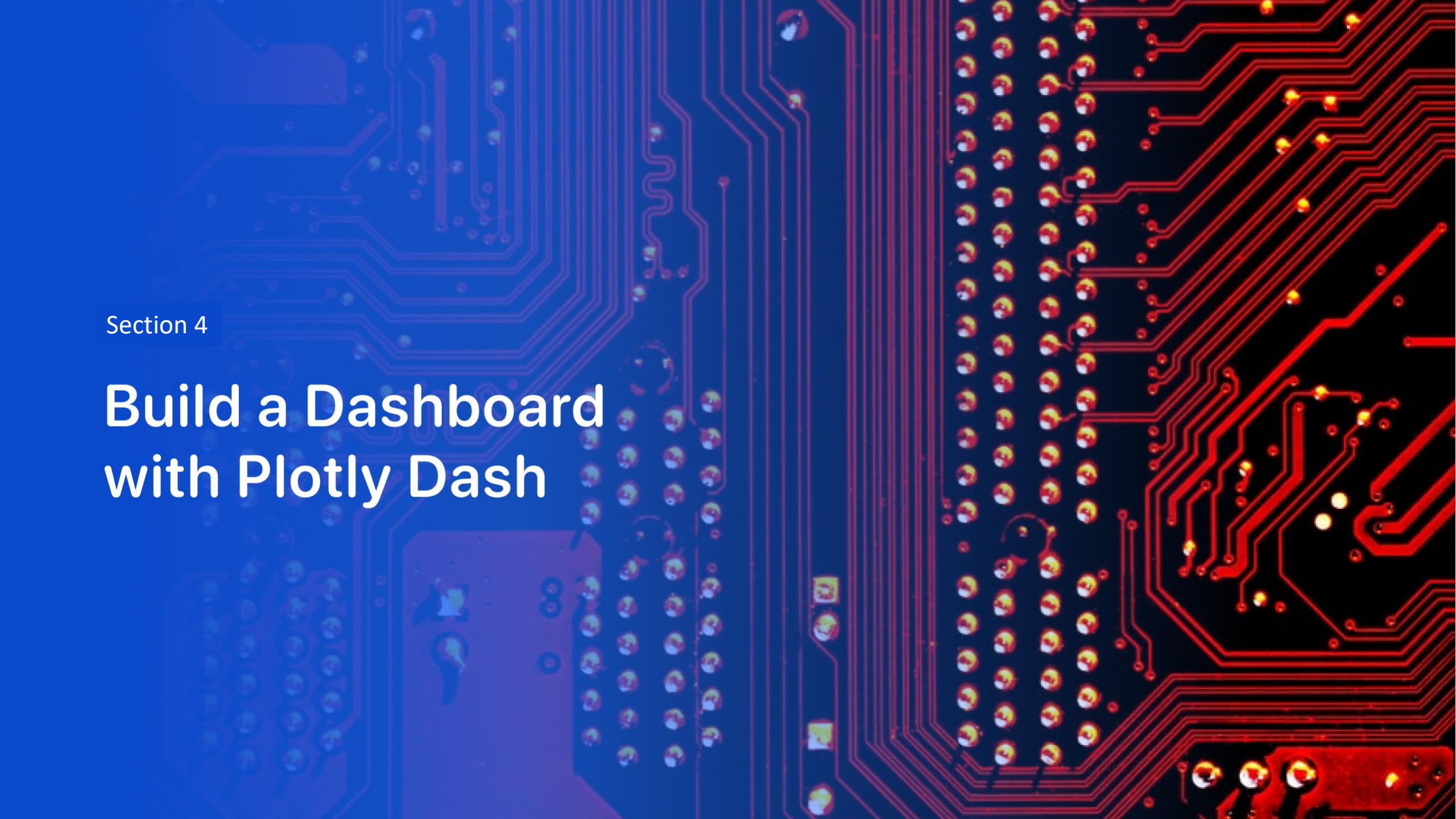
Map showing Success/Failed Launches for each site

- Success/Failed Launches for each site on the Map

Map showing launch site proximities to the Coast



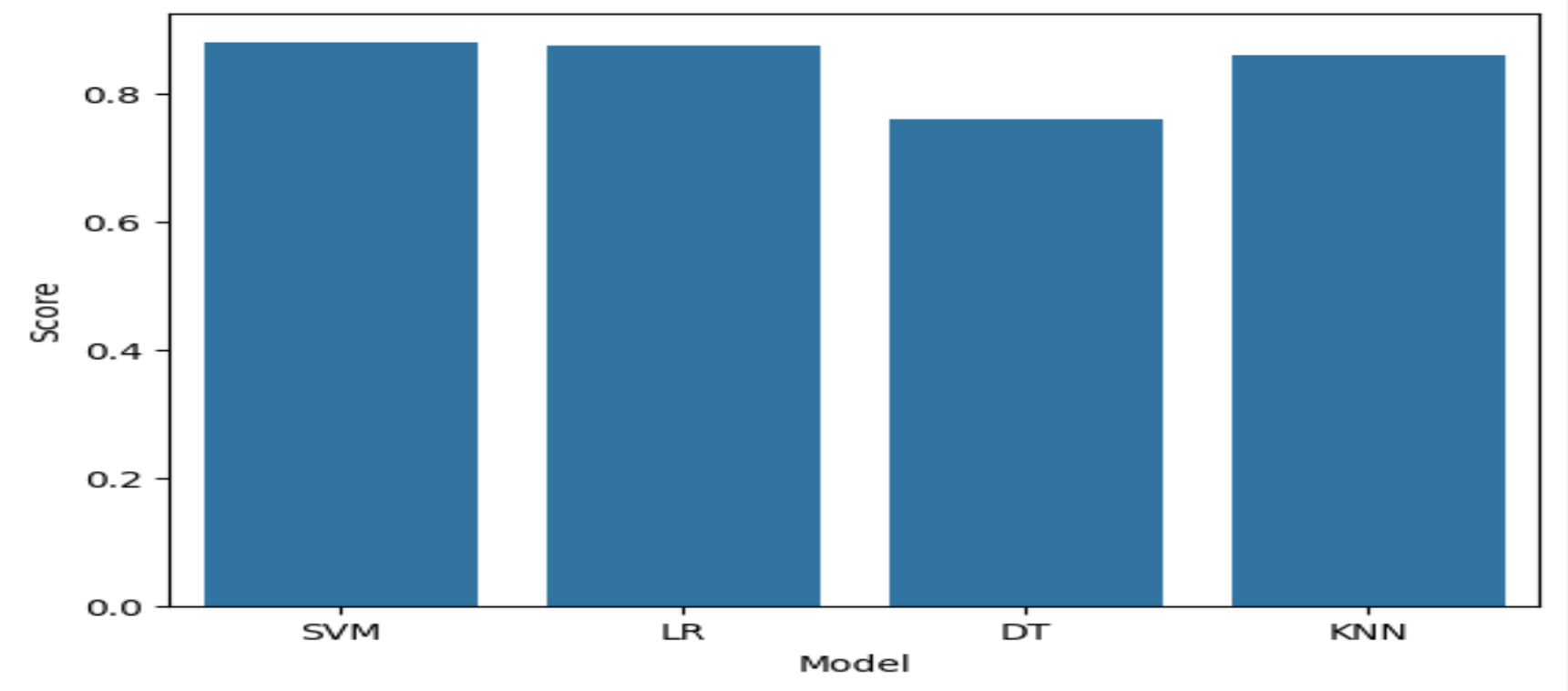
- Launch site VAFBSLC4E is located at 1.8km to the coastline


Section 4
Build a Dashboard with Plotly Dash

Section 5

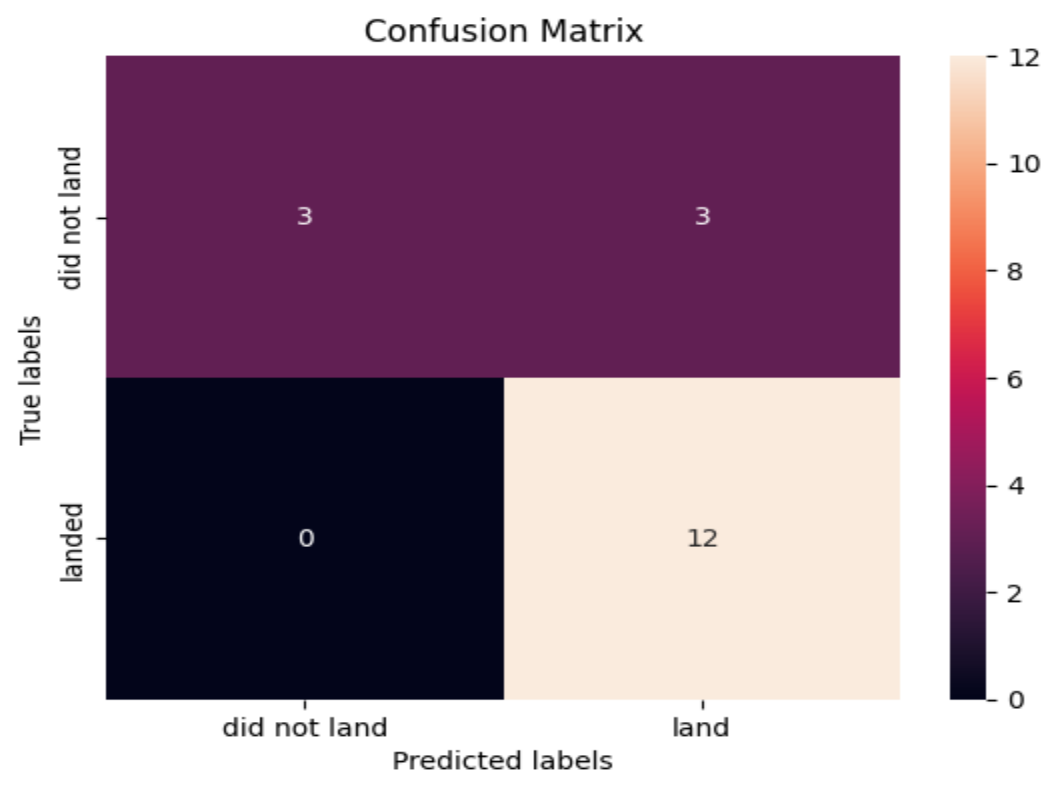
Predictive Analysis (Classification)

Classification Accuracy



- Support Vector Machine model has the highest classification accuracy of 0.88

Confusion Matrix



- The best performing model was the SVM, the model was perfect in predicting the successfully landed but was 50/50 in predicting the unsuccessful landing

Conclusions

- Point 1: Data Science can help business win .
- Point 2 Support Vector Machine model is the most accurate predictive model for this analysis
- Point 3 There is a 83.33% success rate of Falcon 9 first stage landing.

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!