IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Sudhesh Solomon
10/30/2023

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection through API
  - Data Collection using Web Scraping
  - Data Wrangling
  - Exploratory Data Analysis using SQL, and Pandas
  - Visual Analytics using Folium and dashboard analytics using Plotly
  - Machine Learning Algorithms
- Summary of all results
  - Exploratory Data Analysis results
  - Machine Learning Algorithms for Space X – a comparative report

Introduction

- Project background and context

Space X 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 Space X 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 space X for a rocket launch. This goal of the project is to create a machine learning pipeline to predict if the first stage will land successfully.

- Problems you want to find answers
  - Factors that determine successful landing
  - Correlation between features that determine if the landing will be successful or not.
  - Operating conditions for successful landing

Section 1

Methodology

Methodology

Executive Summary

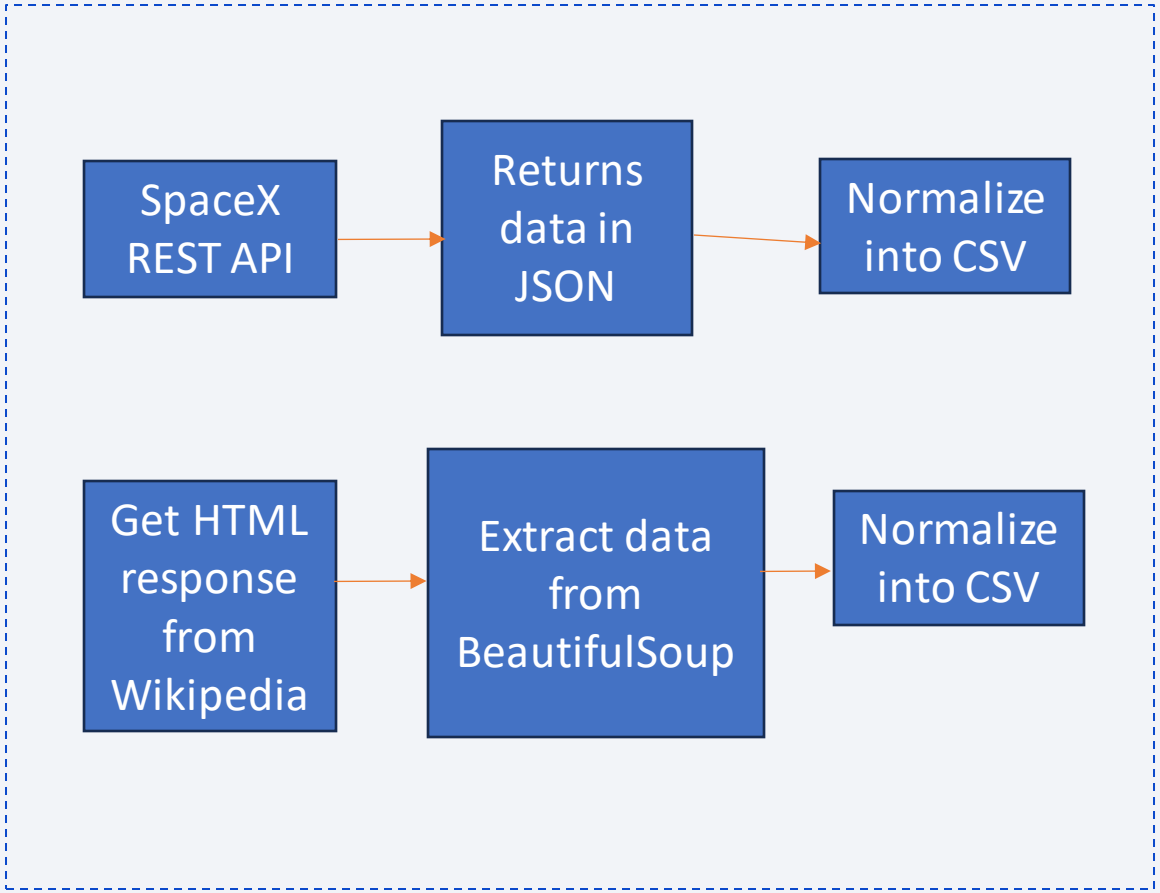
- Data collection methodology:
  - Data was collected using SpaceX API and web scraping from Wikipedia
- Perform data wrangling
  - Applied one-hot encoding to the categorical features
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Split the train and test dataset into 80% for training and 20% for testing.
  - Build the model and test with various combinations of hyperparameters using Grid Search.

Data Collection

- Describe how data sets were collected.
  - The requests module was used to collect data from the SpaceX API.
  - The response was obtained as a JSON and saved into a dataframe using json_normalize method.
  - Data Cleaning was performed to handle missing values.
  - Additionally, BeautifulSoup module was used to scrape data for Falcon 9 launch records

Data Collection – SpaceX API

- Collect SpaceX launch data from SpaceX API
- This dataset gives information about launch, including rocket used, payload delivered, landing outcome, etc.
- The URL of the endpoint is of the format api.spacexdata.com/v4/
- Falcon 9 launch data was obtained by web scraping Wikipedia using BeautifulSoup.



Data Collection - Scraping

- Data Collection using SpaceX REST API
- Link to Github is https://github.com/SudheshSolomon1992/IBM-Data-Science-Professional-Certificate/blob/master/Capstone/jupyter-labs-spacex-data-collection-api.ipynb

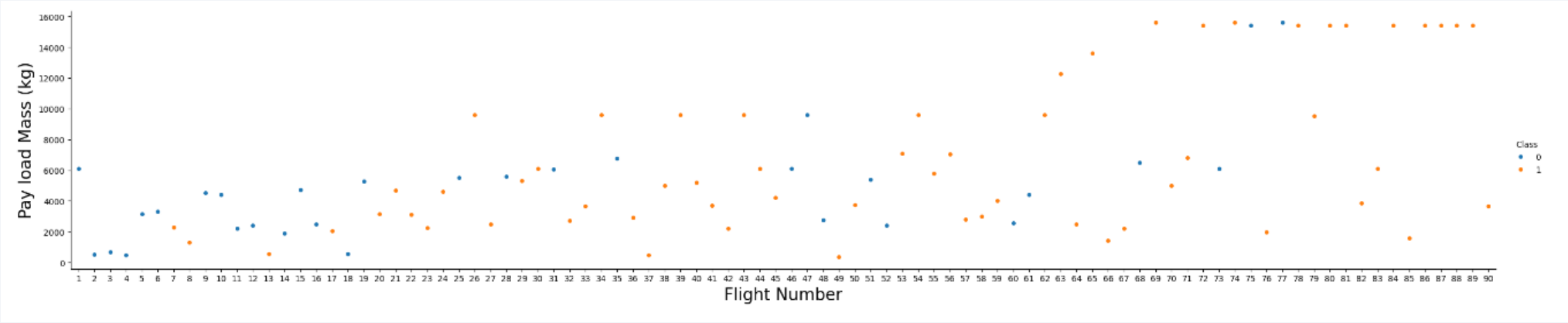
Data Wrangling

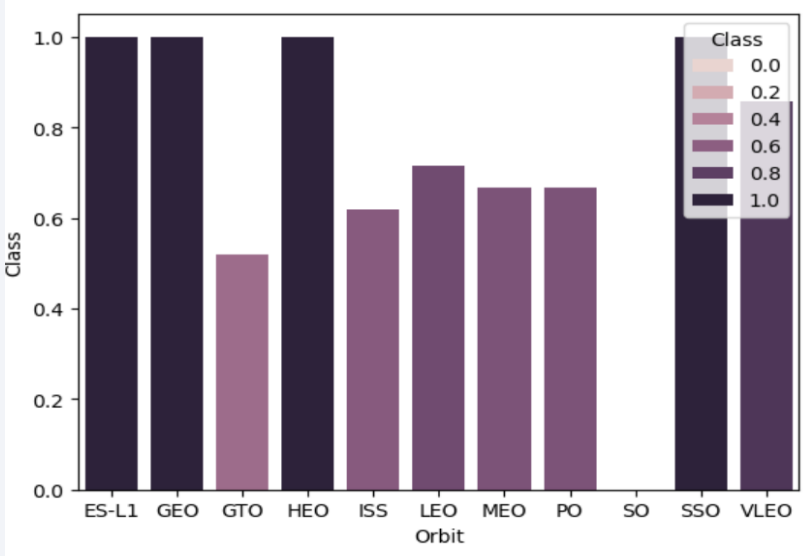
- Data wrangling steps
  - Check and handle null values
  - Calculate the number of launches on each site
  - Calculate the number of occurances of each orbit
  - Calculate the number of occurances of mission outcome per orbit type
  - Create landing outcome label

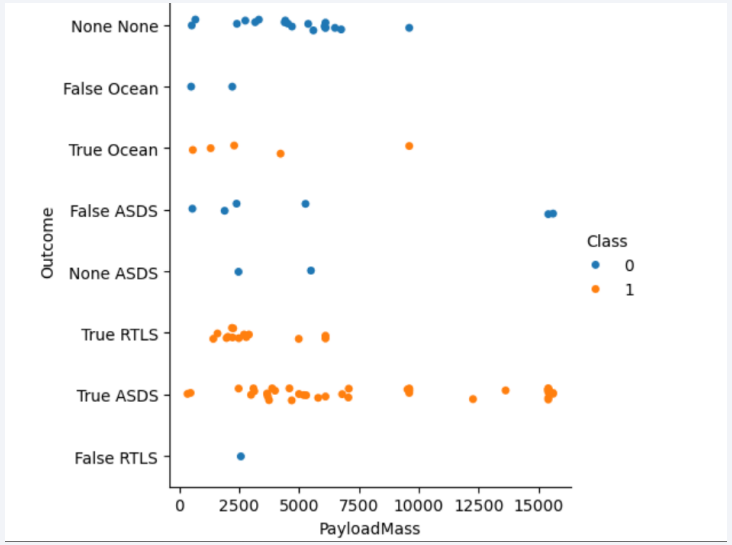
- Github URL is https://github.com/SudheshSolomon1992/IBM-Data-Science-Professional-Certificate/blob/master/Capstone/labs-jupyter-spacex-Data%20wrangling.ipynb

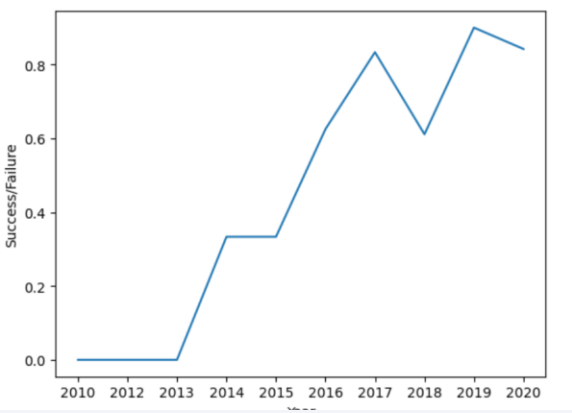
EDA with Data Visualization

Github URL: https://github.com/SudheshSolomon1992/IBM-Data-Science-Professional-Certificate/blob/master/Capstone/IBM-DS0321EN-SkillsNetwork_labs_module_2_jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb









EDA with SQL

- Queries included
  - Finding the unique launch sites in the space mission.
  - Listing top 5 launch sites with name beginning with the string 'CCA'.
  - Calculating total payload_mass carried by boosters launched by NASA (CRS).
  - Calculating average payload_mass carried by booster version F9 v1.1.
  - Finding the date when the first landing outcome in ground pad was achieved.
  - Listing the names of the boosters which had success in drone ship and payload_mass ranged between 4000 kg to 6000 kg.
  - Finding the total number of successful and failure mission outcomes.
  - Finding the names of the booster versions which had carried the maximum payload_mass.
  - Listing the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
  - Ranking 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.
  - Github URL: https://github.com/SudheshSolomon1992/IBM-Data-Science-Professional-Certificate/blob/master/Capstone/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Markers, circles, lines, and marker clusters were used.
- Launch sites have been indicated by **markers**.
- Highlighted areas around specific coordinates like NASA Johnson Space Center have been indicated by **circles**.
- Group of events in each coordinate like launches in the launch site have been indicated using **marker clusters**.
- Distance between two coordinates have been indicated using **lines**.
- Github URL: https://github.com/SudheshSolomon1992/IBM-Data-Science-Professional-Certificate/blob/master/Capstone/IBM-DS0321EN-SkillsNetwork_labs_module_3_lab_jupyter_launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

- Interactive dashboard was built using **Plotly Dash**.
- Pie charts were plotted to show total launches by certain sites.
- Scatter graphs were created to show the relationship with outcome and payload_mass (kg) for different booster version.
- Github URL: https://github.com/SudheshSolomon1992/IBM-Data-Science-Professional-Certificate/blob/master/Capstone/spacex_dash_app.py

Predictive Analysis (Classification)

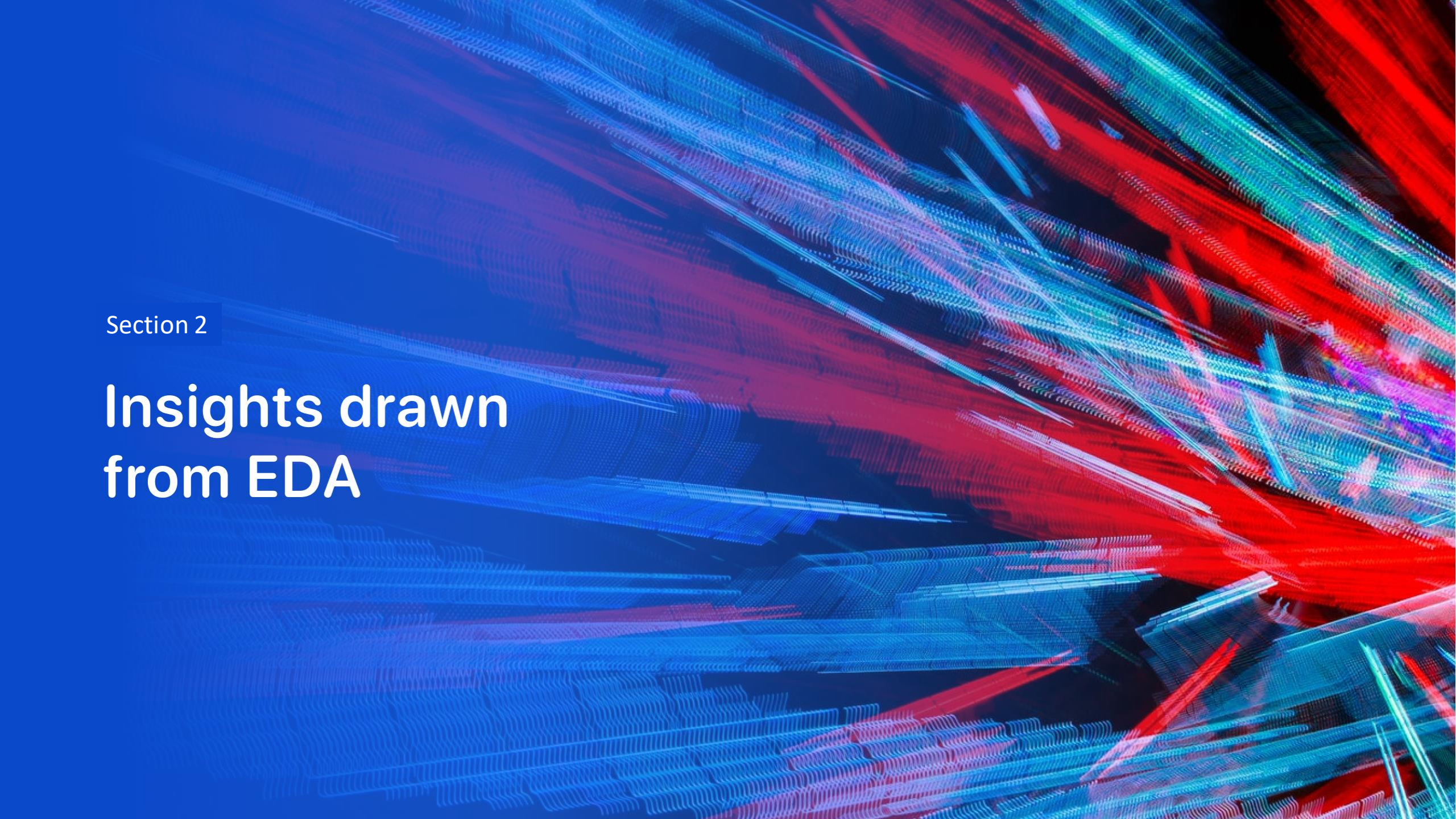
- Data was loaded using numpy, pandas.
- It was then transformed and split into training and test dataset.
- Various machine learning models were developed and hyper-parparameters for those models were tuned using GridSearchCV.
- The best performing model was determined from the results of the prediction.
- Github URL: https://github.com/SudheshSolomon1992/IBM-Data-Science-Professional-Certificate/blob/master/Capstone/IBM-DS0321EN-SkillsNetwork_labs_module_4_SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

- SpaceX used 4 different launch sites.
- The average payload mass for F9 v1.1 was 2,928 kg.
- The first successful landing outcome happened in 2015 which was 5 years after the first launch.
- Many Falcon9 booster versions were successful at landing drone ships having payload_mass above the average value.
- Two booster versions failed at landing drone ships in 2015.
- The landing outcomes became better as the years passed.

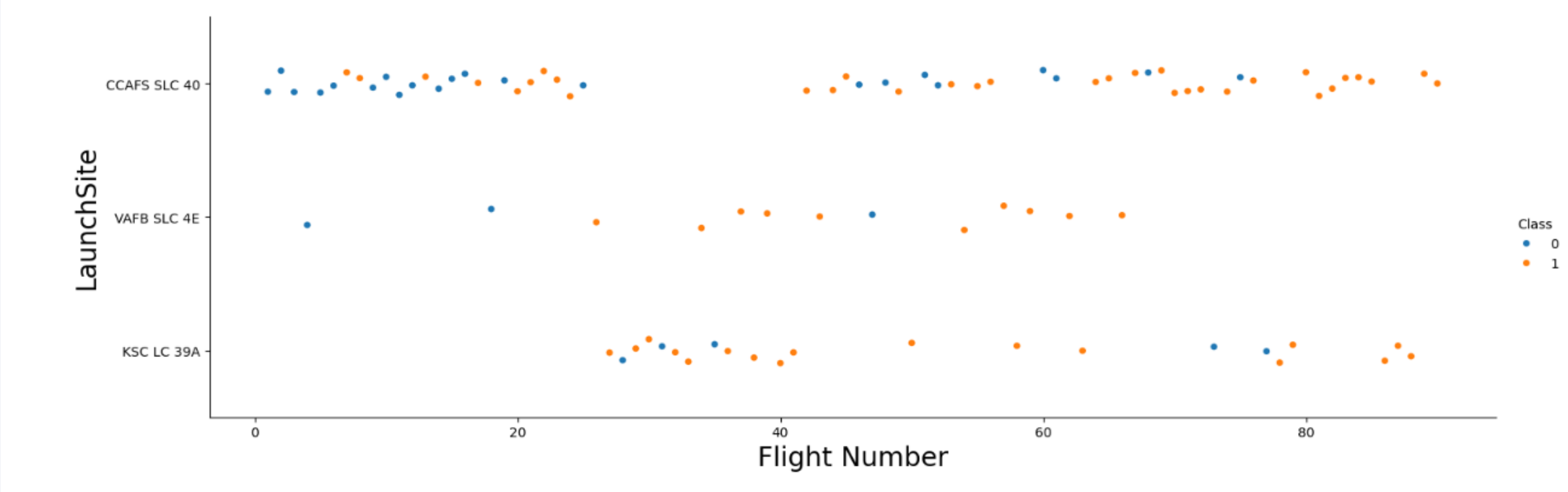
Section 2

Insights drawn from EDA



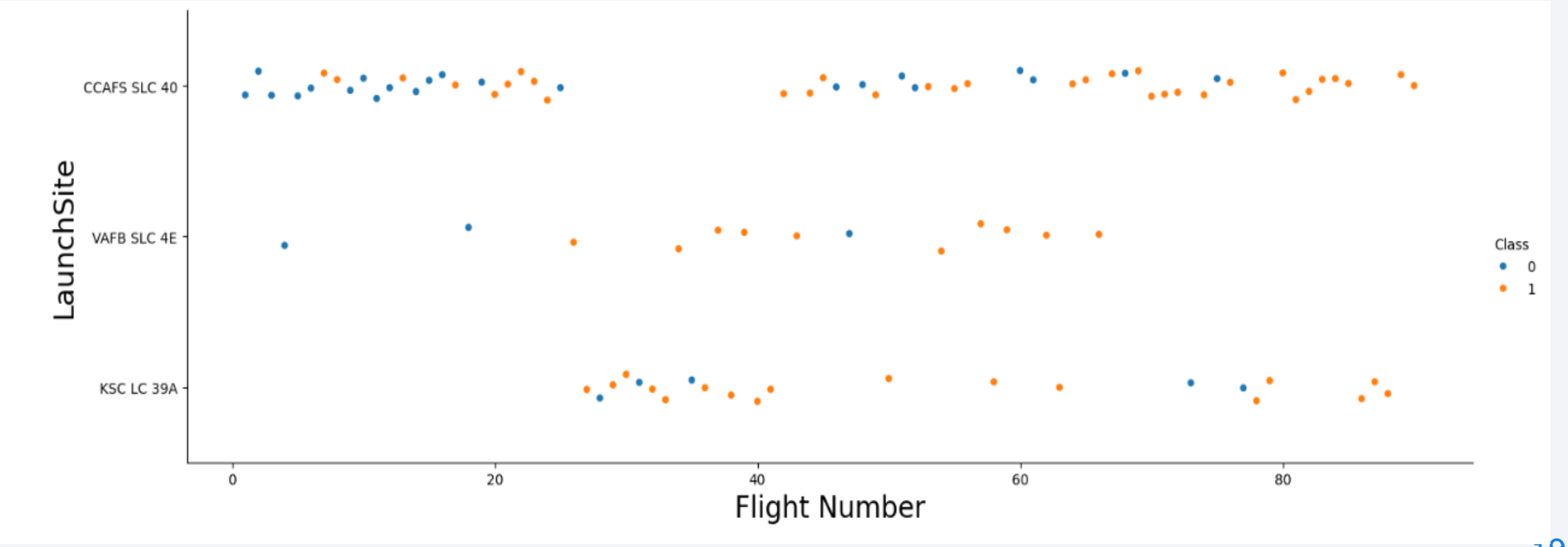
Flight Number vs. Launch Site

- **Inference**: Larger the flight amount at the launch site, greater is the success rate.



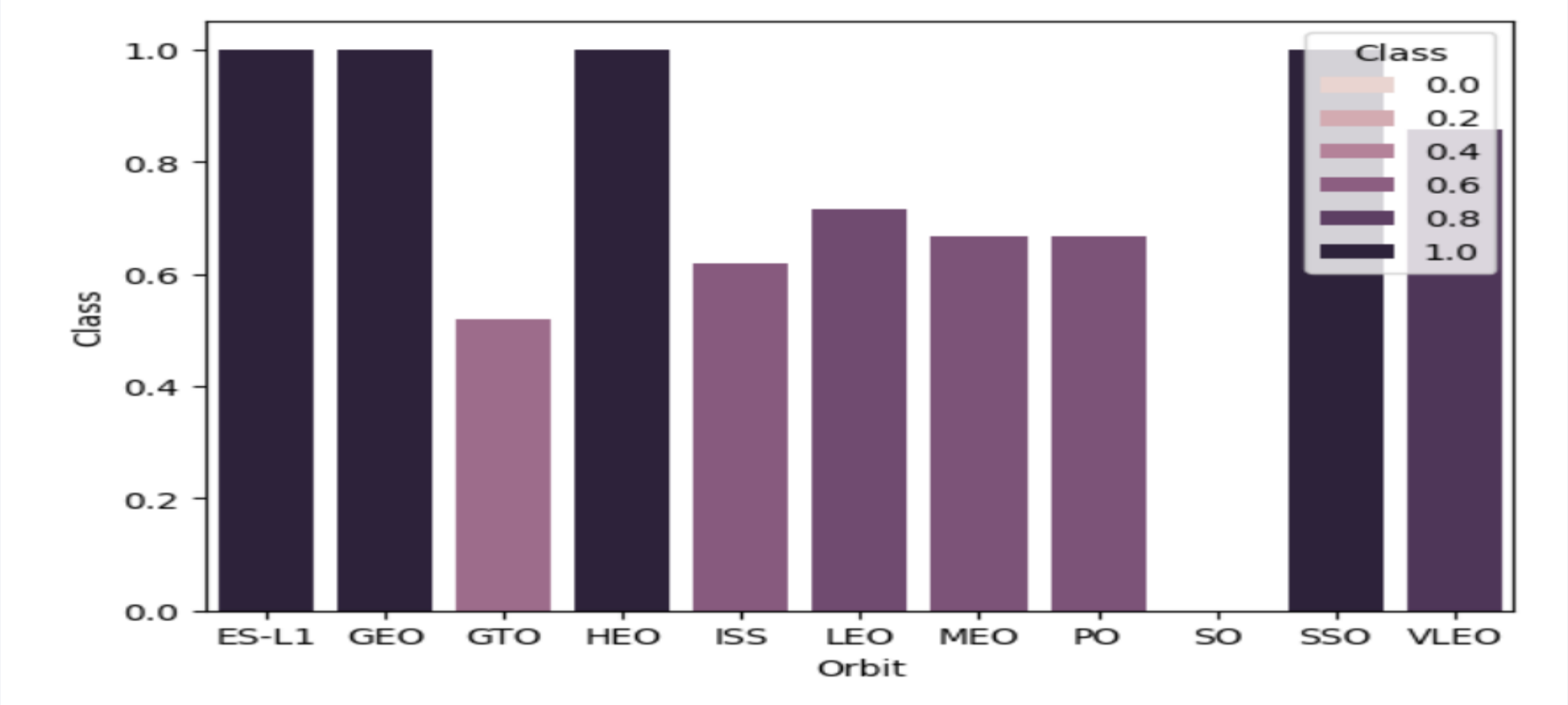
Payload vs. Launch Site

- **Inference**: Greater the payload_mass for CCAFS SLC40 the higher the success rate.



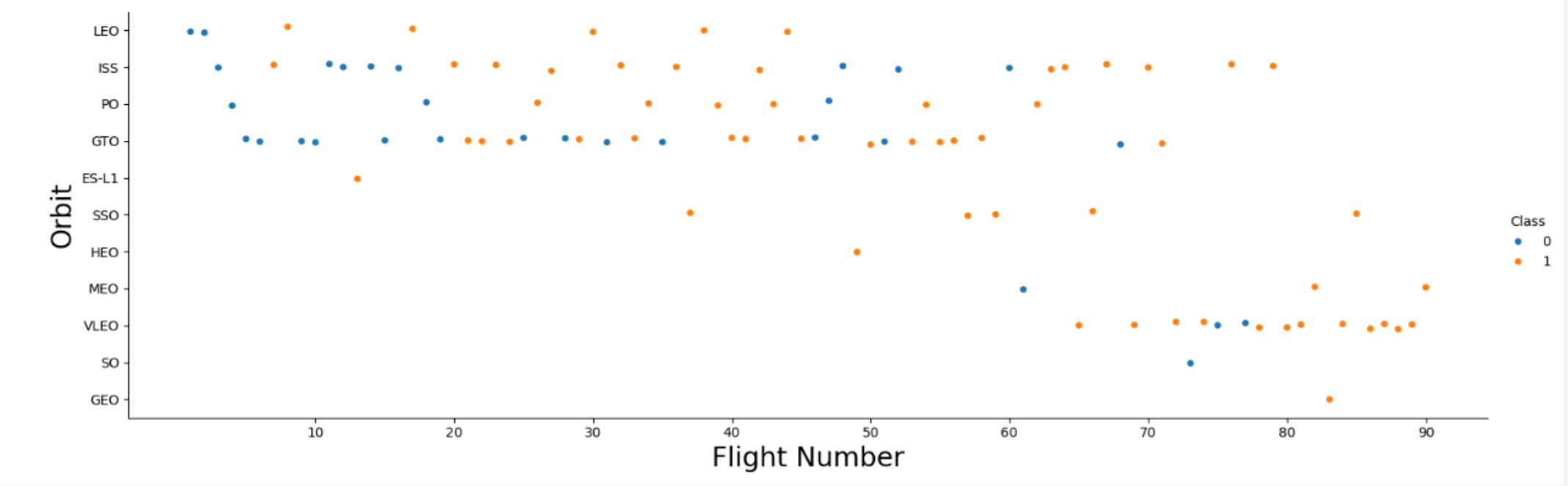
Success Rate vs. Orbit Type

- The highest success were achieved by ES-L1, GEO, HEO, SSO, and VLEO



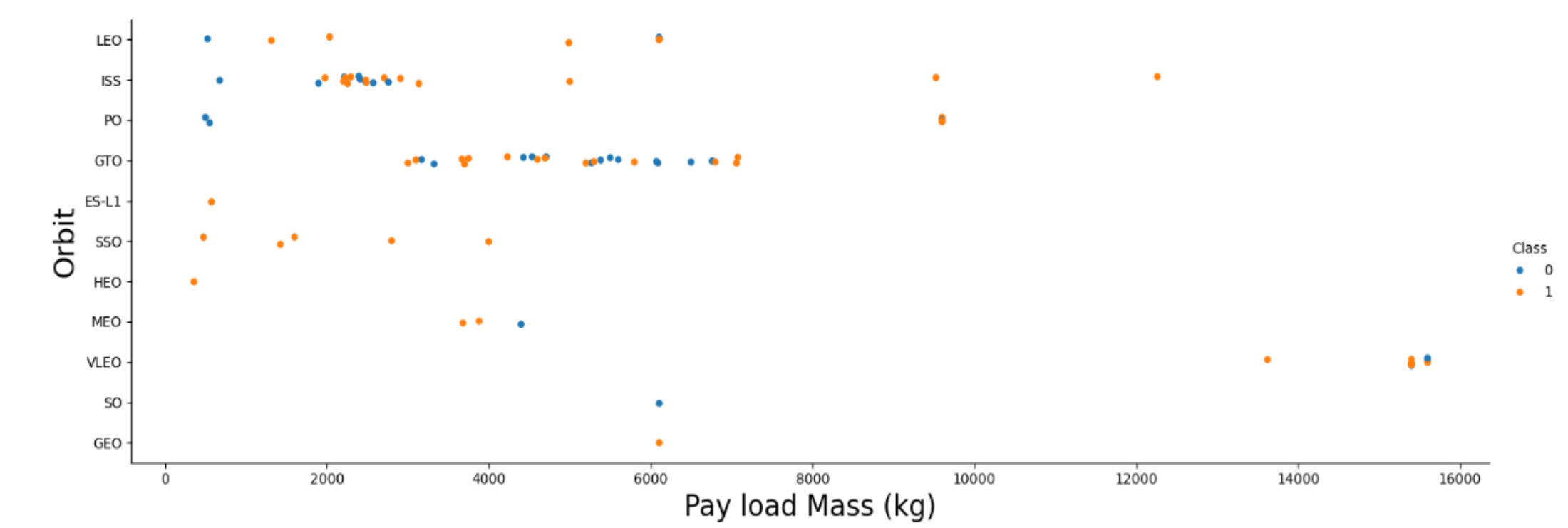
Flight Number vs. Orbit Type

- **Inference**: Higher success is found in LEO and VLEO with increase in the number of flights.



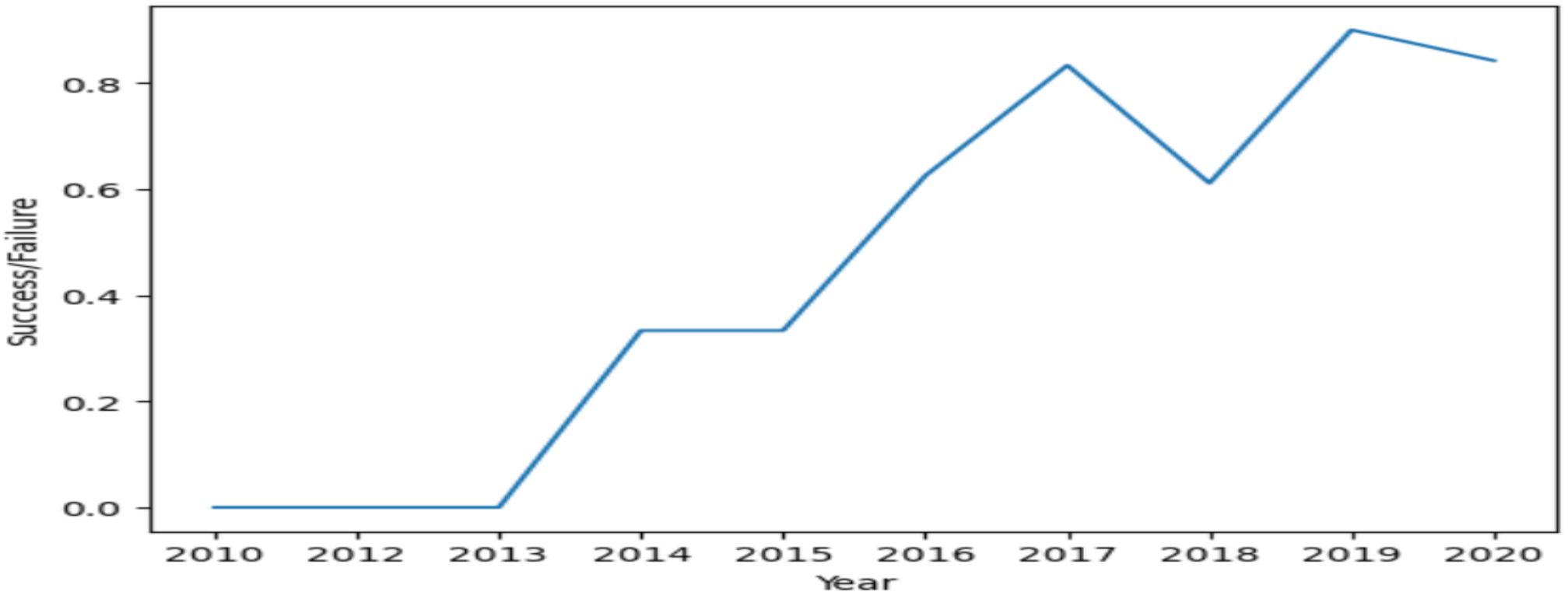
Payload vs. Orbit Type

- **Inference**: Higher payload result in successful landing for PO, LEO, and ISS orbits



Launch Success Yearly Trend

- **Inference**: Success rate is increasing but there was a sharp decline in 2018 after which I again started increasing



All Launch Site Names

- **Query**: **%sql** select Unique(LAUNCH_SITE) from SPACEXTBL;

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

Launch Site Names Begin with 'CCA'

- **Query**: %sql select * from SPACEXTBL where lower(launch_site) like 'cca%' limit 5;

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing_outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- **Query**: **%sql** SELECT SUM(PAYLOAD_MASS__KG_) \
- FROM SPACEXTBL \
- WHERE CUSTOMER = 'NASA (CRS)';

1
45596

Average Payload Mass by F9 v1.1

- **Query**: **%sql** SELECT AVG(PAYLOAD_MASS__KG_) \
- FROM SPACEXTBL \
- WHERE BOOSTER_VERSION = 'F9 v1.1'

1
2928

First Successful Ground Landing Date

- **Query**: `%sql SELECT MIN(DATE) \`
- `FROM SPACEXTBL \`
- `WHERE LANDING_OUTCOME = 'Success (ground pad)'`

1
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Query: `%sql SELECT PAYLOAD \`
- `FROM SPACEXTBL \`
- `WHERE LANDING_OUTCOME = 'Success (drone ship)' \`
- `AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000;`

payload
JCSAT-14
JCSAT-16
SES-10
SES-11 / EchoStar 105

Total Number of Successful and Failure Mission Outcomes

- **Query**: `%sql SELECT MISSION_OUTCOME, COUNT(*) \`
- `FROM SPACEXTBL \`
- `GROUP BY MISSION_OUTCOME`

mission_outcome	2
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- **Query**: `%sql SELECT BOOSTER_VERSION \`
- `FROM SPACEXTBL \`
- `WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);`

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

2015 Launch Records

- **Query**: `%sql SELECT substr(Date,4,2) as month, DATE,BOOSTER_VERSION, LAUNCH_SITE, LANDING_OUTCOME \`
- `FROM SPACEXTBL \`
- `where LANDING_OUTCOME = 'Failure (drone ship)' and substr(Date,7,4)='2015';`

MONTH	DATE	booster_version	launch_site	landing_outcome

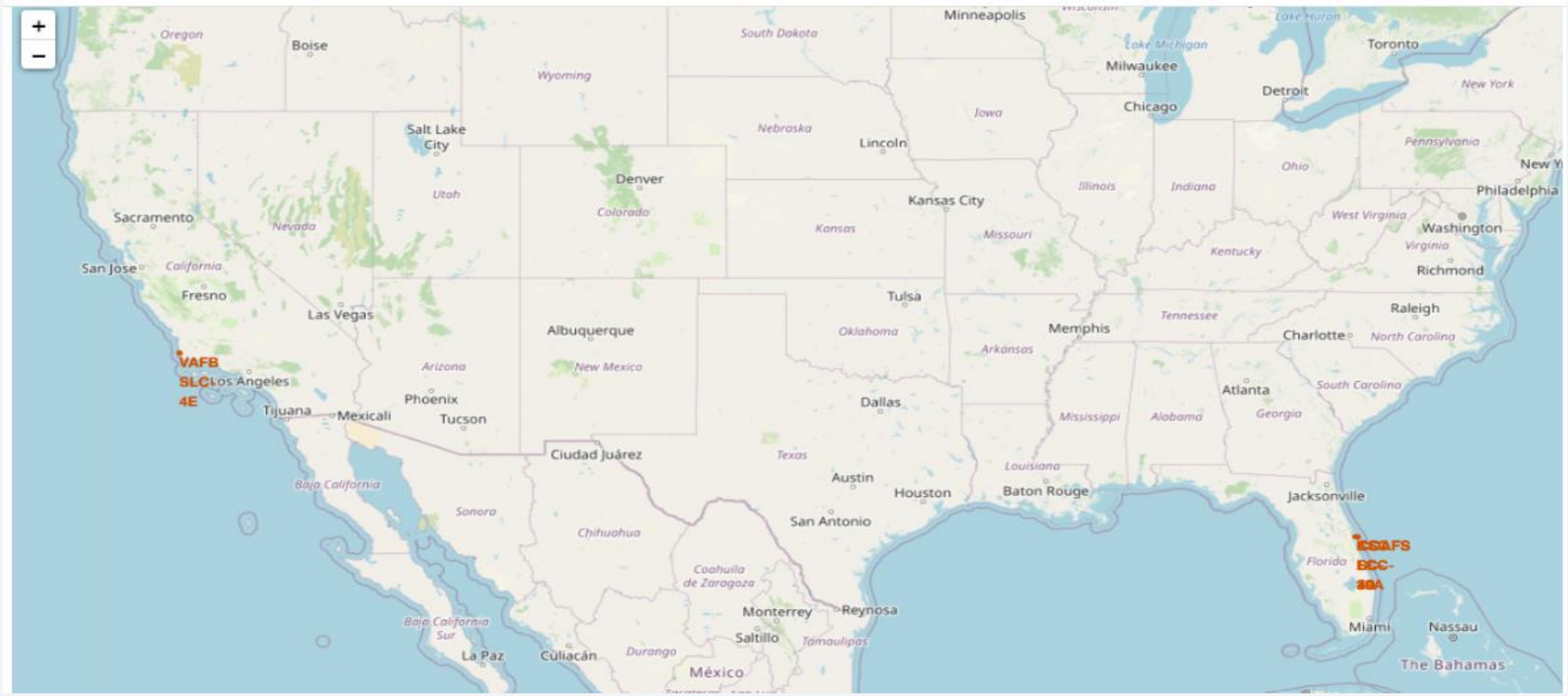
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- Query: `%sql SELECT LANDING_OUTCOME, count(*) as count_outcomes \`
- `FROM SPACEXTBL \`
- `WHERE DATE between DATE '2010-06-04' and DATE '2017-03-20' \`
- `group by LANDING_OUTCOME \`
- `order by count_outcomes DESC;`

landing_outcome	count_outcomes
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Success (ground pad)	5
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	1
Precluded (drone ship)	1

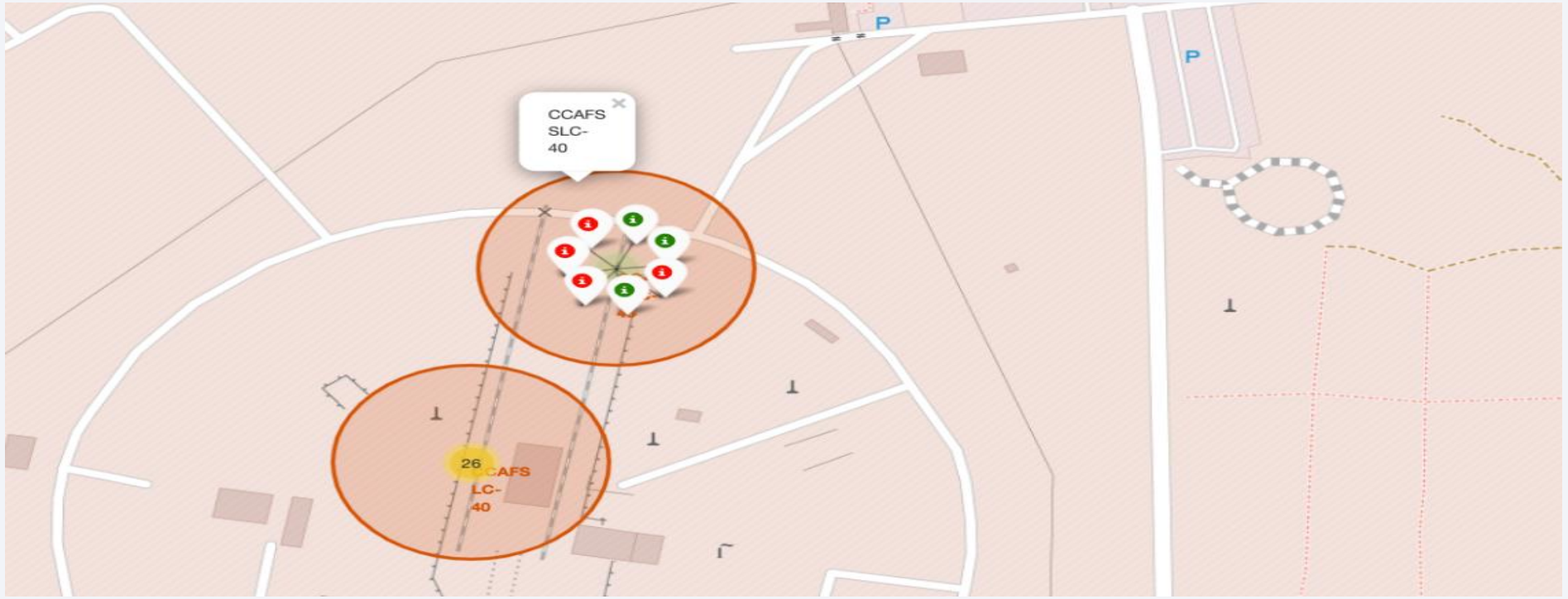

Section 3
Launch Sites
Proximities Analysis

All Launch Sites Global Markers



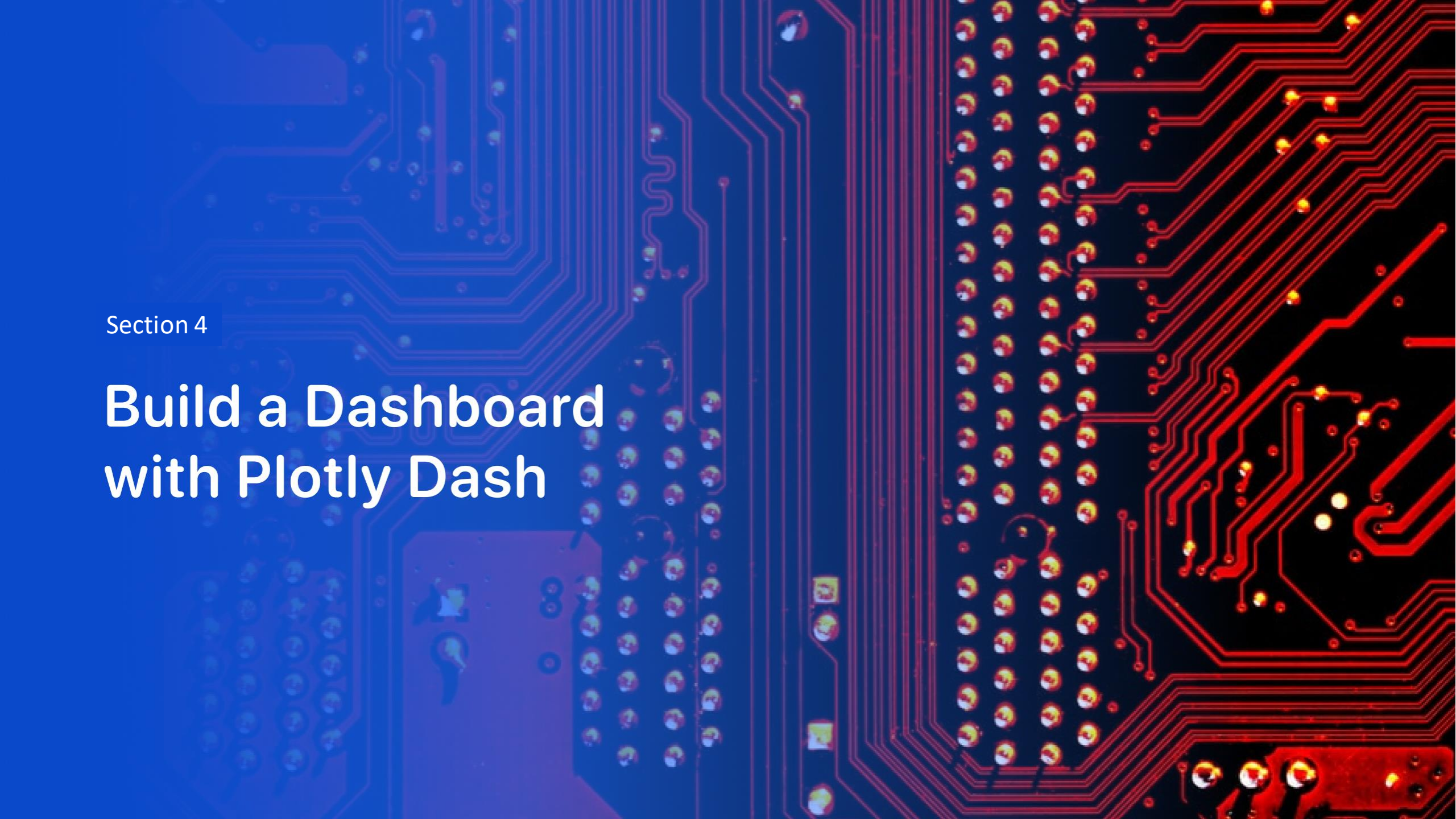
Markers showing outcomes with color labels

- Green markers depict successful launches while red markers denote failed launches.



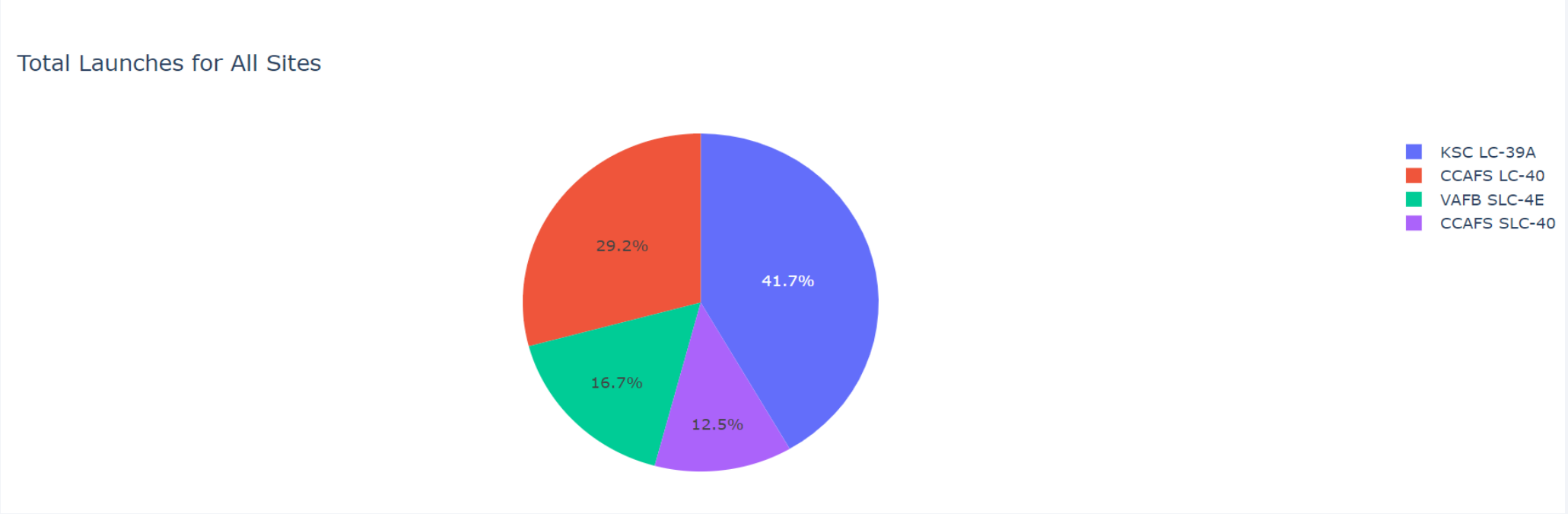
Launch Site Distance to Coast

- The figure below depicts the distance to coast from the launch site.


Section 4
Build a Dashboard with Plotly Dash

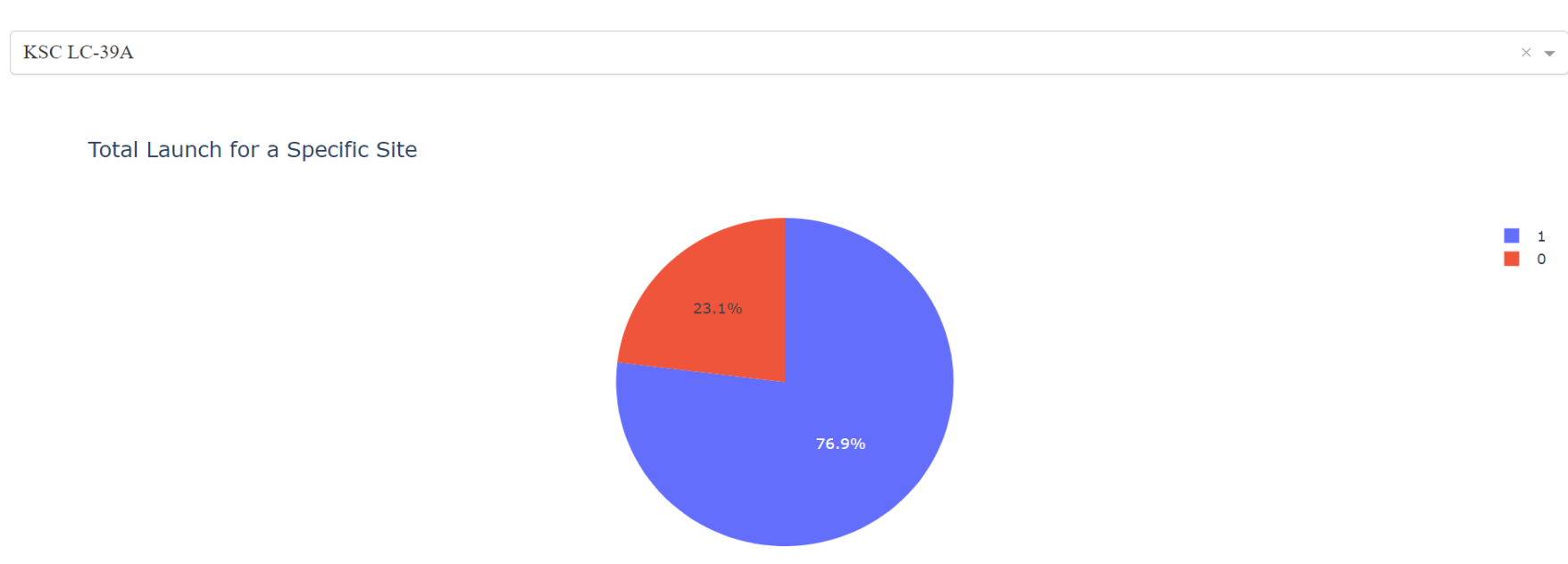
Success Percentage per Launch Site

- We can see that KSC LC-39A had the most successful launches among all the sites.



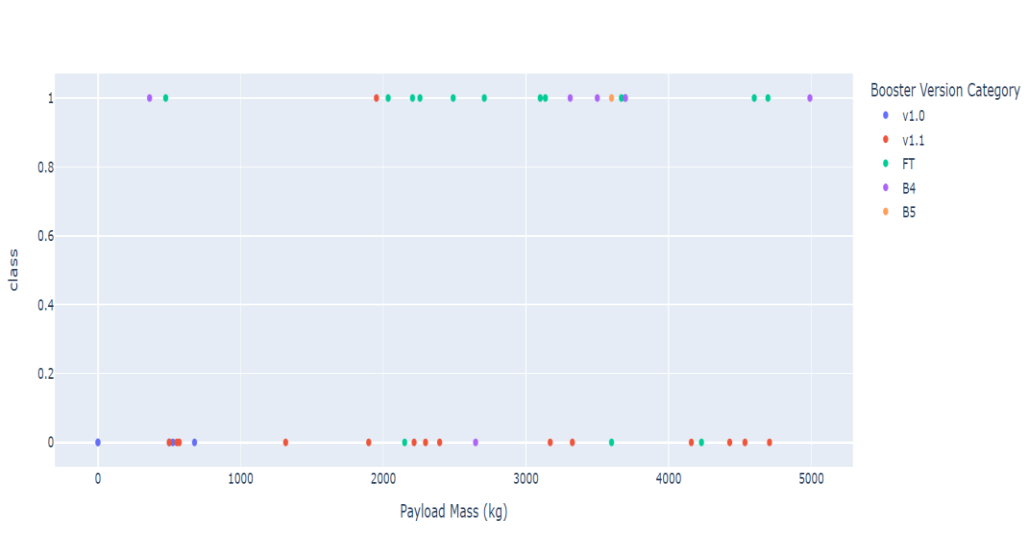
<Dashboard Screenshot 2>

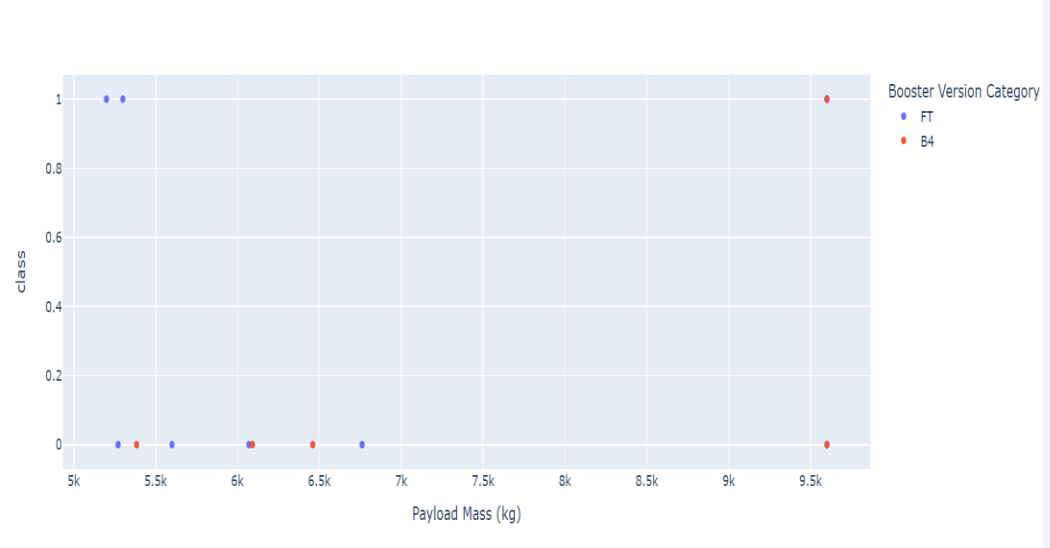
- KSC LC-39A achieved a 76.9% success and a 23.1% failure rate



Payload vs Launch Outcomes for all sites

- Success rates of low weighted payloads (0 – 5000 kg) vs high weighted payloads (5000kg – 10000 kg)



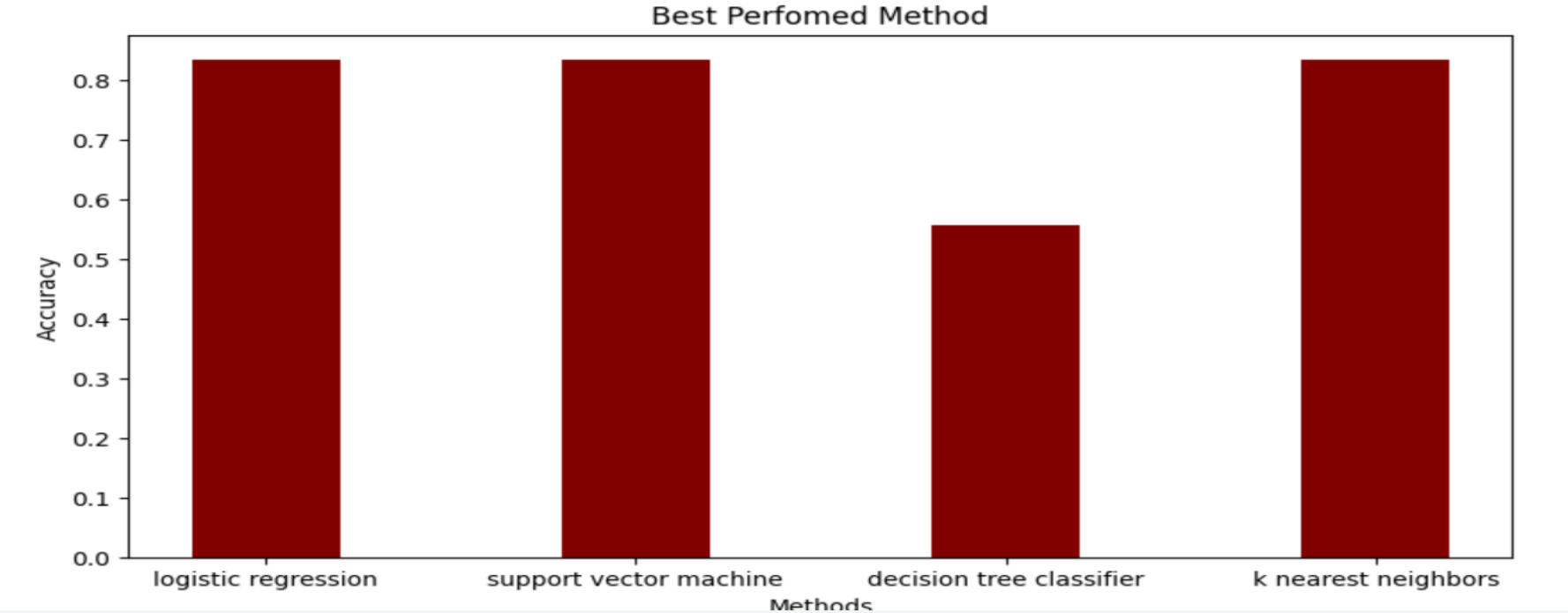


Section 5

Predictive Analysis (Classification)

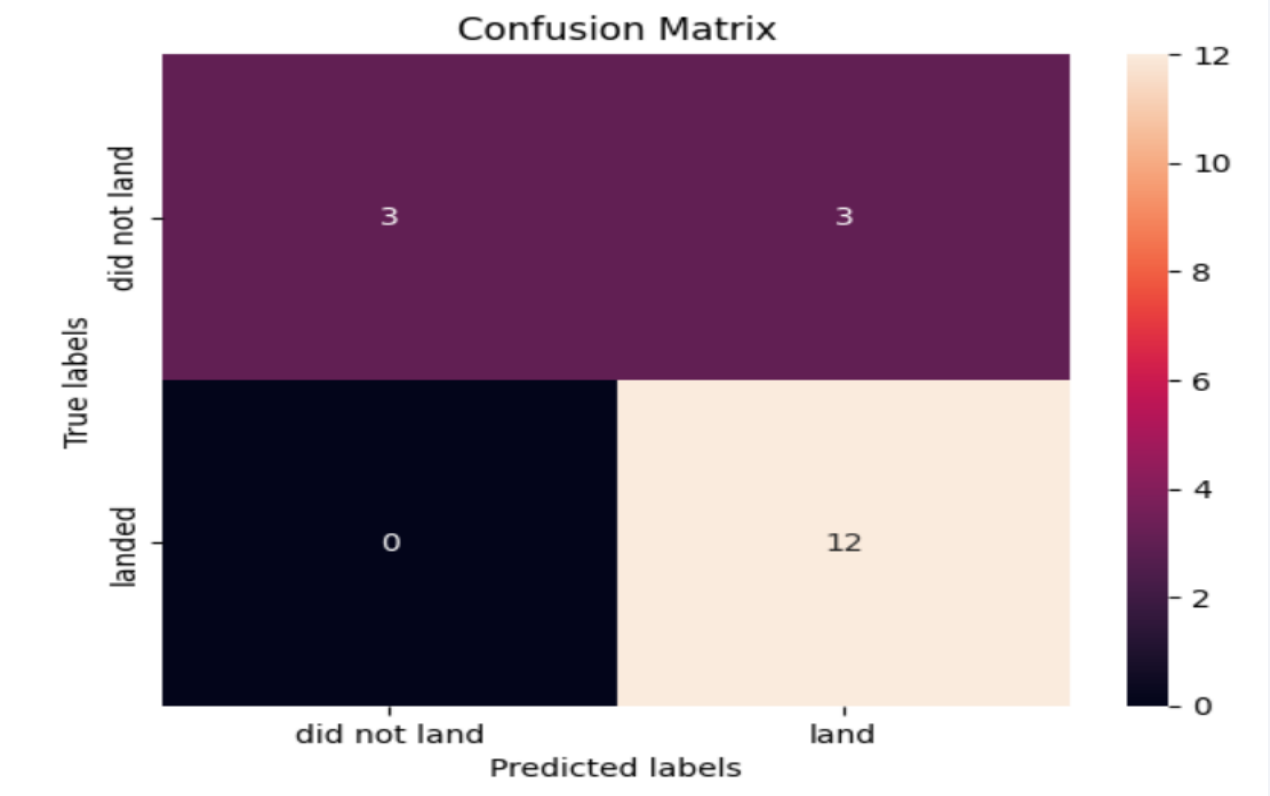
Classification Accuracy

- SVM, Logistic Regression, and KNN achieved almost nearly the same accuracy of 83%.



Confusion Matrix

- Shown here is the confusion matrix of KNN.



Conclusions

- Greater the number of flights at the launch site, greater the success rate.
- The launch success rate started to gradually increase from 2013.
- Orbits ES-L1, GEO, HEO, SSO, VLEO had the most success.
- KSC LC-39A was the most successful launch site in terms of successful launches.
- The KNN was the best performing algorithm achieving nearly 84% accuracy.

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!