IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Michael Nix
02-12-2025

Outline

SUMMARY

Executive Summary

- Methodology
  - Data comes from [SpaceX DataSet](). It contains launch data for SpaceX's Falcon 9 rocket launches, including launch sites, achieved orbit, mission outcome, and if the booster was able to land successfully.
  - Data cleaning / wrangling / extracting landing outcome data to serve as the dependent variable in the machine learning models.
  - SQL queries and data visualization, including plots, graphs, charts, maps, and an interactive dashboard to gain insights about the data set.
  - Predictive analysis was performed using: Logistic Regression, Support Vector Machine (SVM), Decision Tree, and k-Nearest Neighbors (KNN).
- Results
  - Logistic Regression, SVM, Decision Tree, and KNN all performed equally well on the dataset for predictive purposes.

INTRODUCTION

Introduction

- We want to make predictions about the success of SpaceX's Falcon 9 booster rocket first stage landings.
- Questions to explore:
  - Which machine learning model will best predict the outcome of a Falcon 9 booster rocket safely landing from a future launch?
  - Will a future launch of a Falcon 9 booster rocket successfully land?
- See the appendix for the link to the code repository for this project

Section 1

Methodology

Methodology

Executive Summary

- Data on the SpaceX Falcon 9 first stage landings was provided with the course in a CSV file format and from a public API (https://api.spacexdata.com/); additional data scraped from a Wikipedia article.
  - Included data: Launch Sites, Payload Mass, Orbit, Successful Launch, Successful Landing
- Data was wrangled/cleaned in preparation for visualization, queries, and machine learning models
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Pandas used to build a dataframe; sklearn used for modeling & machine learning

Data Collection

- Data provided by the course and from SpaceX public API

EDA w/ SQL	EDA & Plots	Folium Maps	Dash	Machine Learning
Utilized SQL to perform initial data exploration so to become familiar with the data; i.e. to understand the type of data contained in the file.	Utilized Python Pandas, Numpy, Matplotlib, and Seaborn to place the data in a dataframe for use in plot visualization to quickly see success rates of launches and landings.	Created maps of launch locations and landings for the Falcon 9 rockets. The maps show successful and unsuccessful launches and landings.	Used Plotly Dash to build an interactive dashboard for easy retrieval lookup of graphs and charts for various rocket launch and landing attempts.	Created several machine learning models to determine which launches and landings were successful so to be able to predict future successful launches and landings of the Falcon 9 rocket.

Data Collection – SpaceX API

- SpaceX data publicly available at https://api.spacexdata.com/
- The API is not affiliated with SpaceX
- The completed SpaceX API calls notebook: here on GitHub

1. Sent GET request to API
2. Extract Data
3. Convert data format
4. Used defined functions to generate columns of data
5. Combined columns into a new dataframe
6. Filtered out all rockets other than the Falcon 9
7. Handle missing values

Data Collection - Scraping

- Data was split into training and testing sets.
- Machine learning models employed:
  - Logistic Regression
  - SVM (Support Vector Machine)
  - Decision Tree
  - KNN (k-Nearest Neighbors)
- Hyper-parameters selected
- The completed web scraping notebook: here on GitHub

1. Web Scrape the Falcon 9 Wikipedia Page
2. Created a BeautifulSoup object from the HTML
3. Selected the relevant tables
4. Extracted all column names from the <th></th> tags
5. Made a Pandas Dataframe from the extracted data
6. Saved the output as a csv file

Data Wrangling

- The data contained information about launch sites, achieved orbit, and mission outcomes – both launch and landings whether successful or unsuccessful.
- The mission outcomes were converted to 1 (successful) or 0 (unsuccessful) for the Falcon 9 booster rockets.
- The outcome was added as a column to the dataframe.
- The completed data wrangling notebook: [here]() on GitHub

1. Load the CSV file as input data
2. Determine the number of launches for each site
3. Identify the achieved orbit
4. Determine the outcome for each successful and unsuccessful mission launch
5. Add the outcome to a column in the dataframe for each mission launch
6. Compile all data into a single dataframe

EDA with Data Visualization

- Scatterplots were created to show trends between the launch sites, payload mass, flight number all showing the mission outcome - a blue dot for 0 (unsuccessful) and a yellow dot for 1 (successful)
  - See Appendix for charts figures 1 through 5
- Bar chart shows the relationship between the success rate and the achieved orbit
  - See Appendix figure 6
- Line chart shows the success rate over time for mission outcomes
  - See Appendix figure 7

- The completed EDA with Data Visualization notebook: here on GitHub

EDA with SQL

- SQL queries written to extract information about:
    - Launch sites
    - Landing outcomes
    - Payload mass
    - Booster types
    - Dates
- The completed EDA with SQL notebook: here on GitHub

Interactive Map with Folium

- Map objects were created and added to a Folium map
    - **Circles** indicate the different launch sites
    - **Markers** indicate mission success rate outcomes
        - Red indicate mission failures (figure 8)
        - Green indicate mission success
    - **Lines** drawn from launch point locations to nearby geographic features
- These objects were added to quickly see key indicators important to mission objectives/outcomes
- The completed interactive map with Folium map: [here]() on GitHub

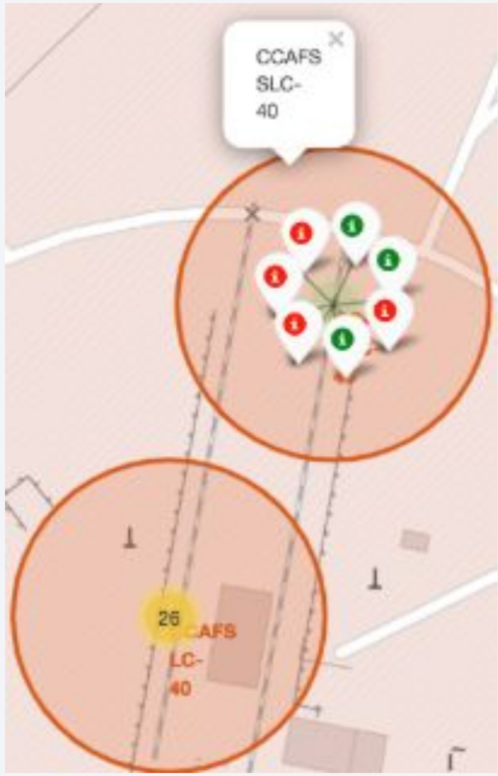


Figure 8

Dashboard with Plotly Dash

- The Plotly Dash dashboard allows for selection for one or all the launch sites to see relevant data for the selection.
- The selected option will show the percentage of successful launches as compared to unsuccessful launches
    - ‘All’ will show comparison of successful outcomes across all sites (figure 9)
    - If you select an individual launch site, it will show the ratio between successful and unsuccessful launches for that site
- The slider will filter the payload mass for the scatterplot
- The completed Plotly Dash lab: [here](#) on GitHub

Figure 9

Predictive Analysis (Classification)

- The dataset was split into training and testing sets
- Trained four machine learning models on the training set:
  - Logistic Regression
  - SVM
  - Decision Tree
  - KNN
- The completed predictive analysis lab: here on GitHub

1. Created Pandas dataframe from cleaned data
2. Split data into training and test sets
3. Trained 4 different models on the training data sets
4. Evaluated the 4 different models using the testing data sets
5. The models were compared on accuracy to determine effectiveness at prediction

RESULTS

Results

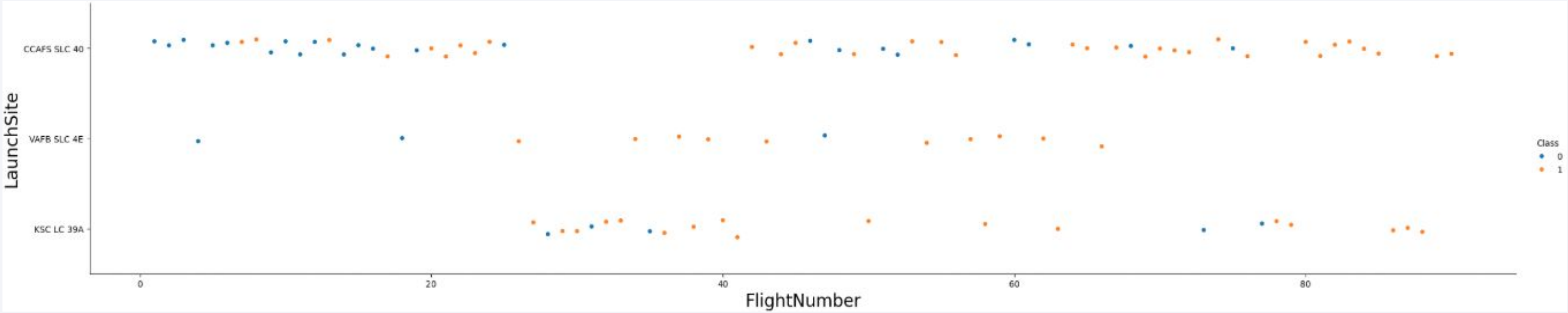
- Insights from Exploratory Data Analysis
    - Data Visualizations
    - SQL Queries
- Launch Sites Proximities Analysis
    - Folium Maps (Screenshots)
- Build a Dashboard with Plotly Dash
    - Plotly Dash (Screenshots)
- Predictive Analysis (Classification)
    - Machine Learning

Section 2

Insights drawn from EDA

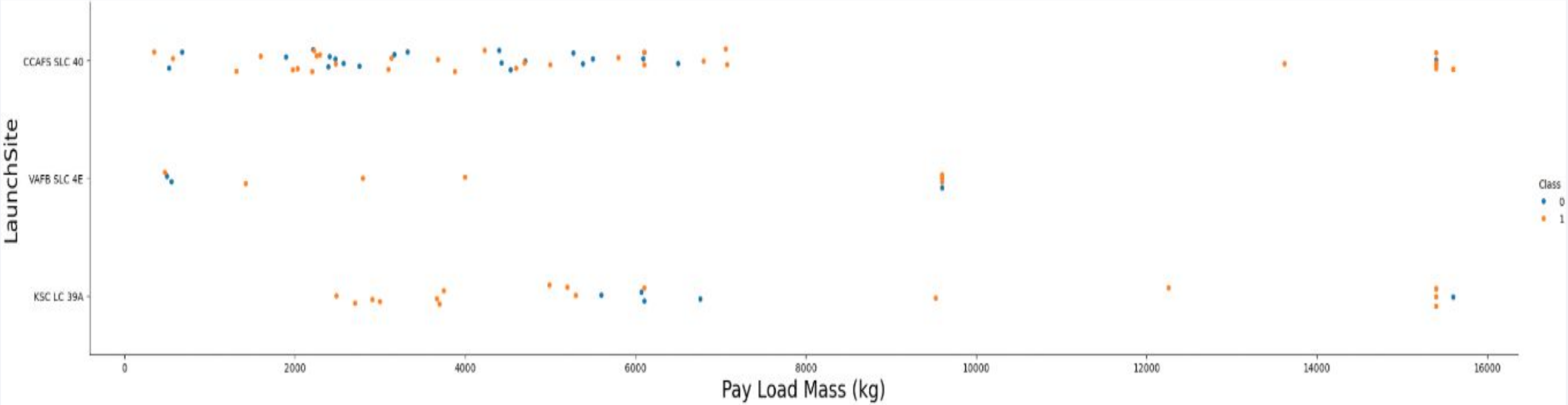
Flight Number vs. Launch Site

- Variation among launch sites in success rates
- Falcon 9 first stage landings increase across all launch sites with more mission launches
    - 0 failed landings (blue dot)
    - 1 successful landings (orange dot)



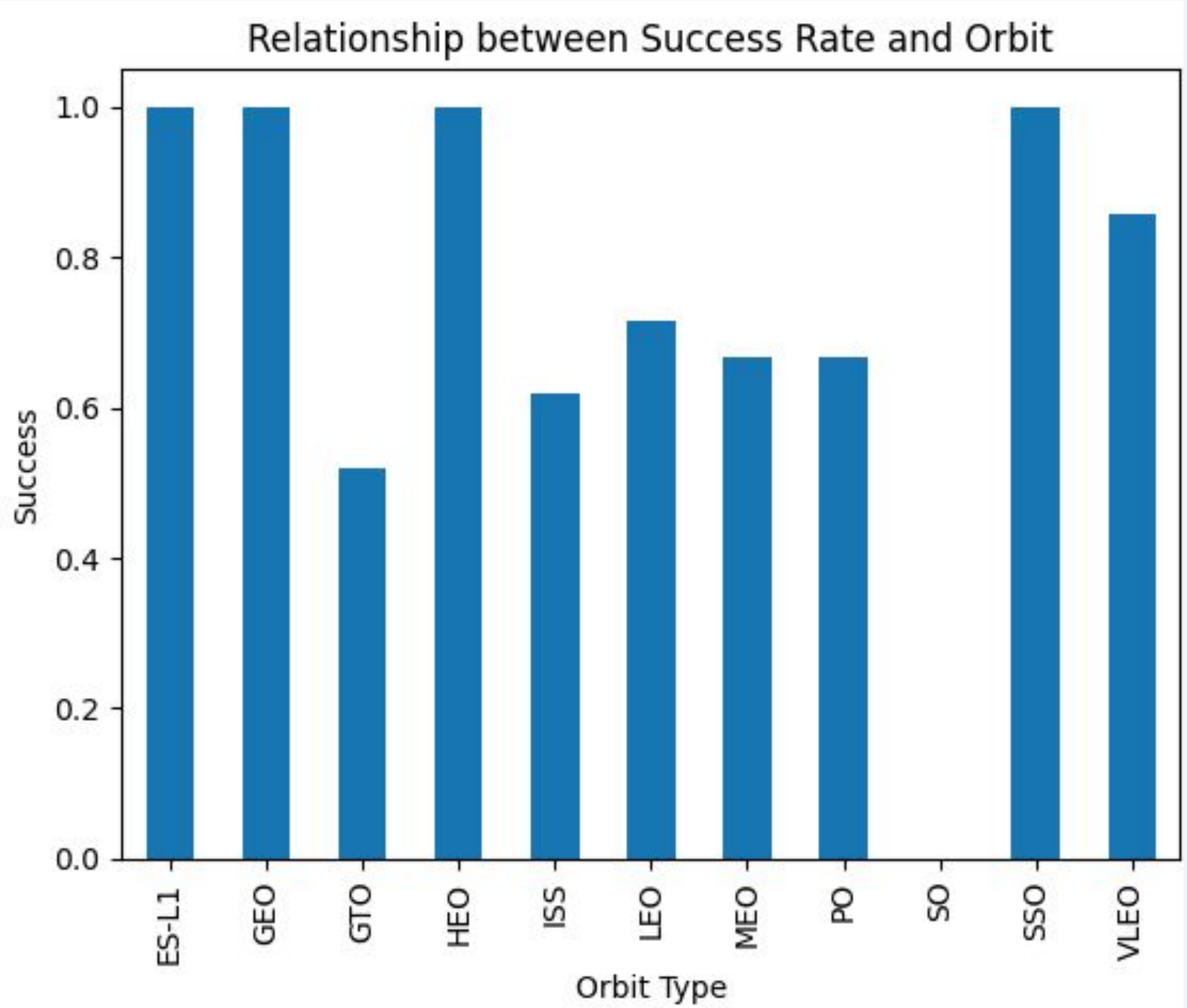
Payload vs. Launch Site

- Multiple launch sites success outcomes compared across payload mass
  - 0 failed landings (blue dot)
  - 1 successful landings (orange dot)



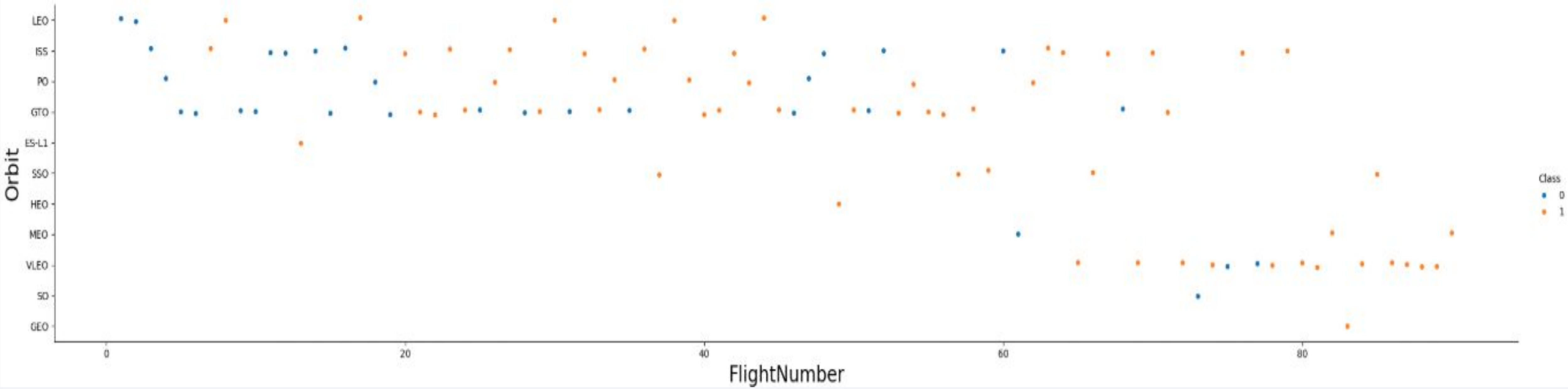
Success Rate vs. Orbit Type

- ES-L1, GEO, HEO, and SSO orbits have no failed first stage landings.
- SO orbits have no successful first stage landings.



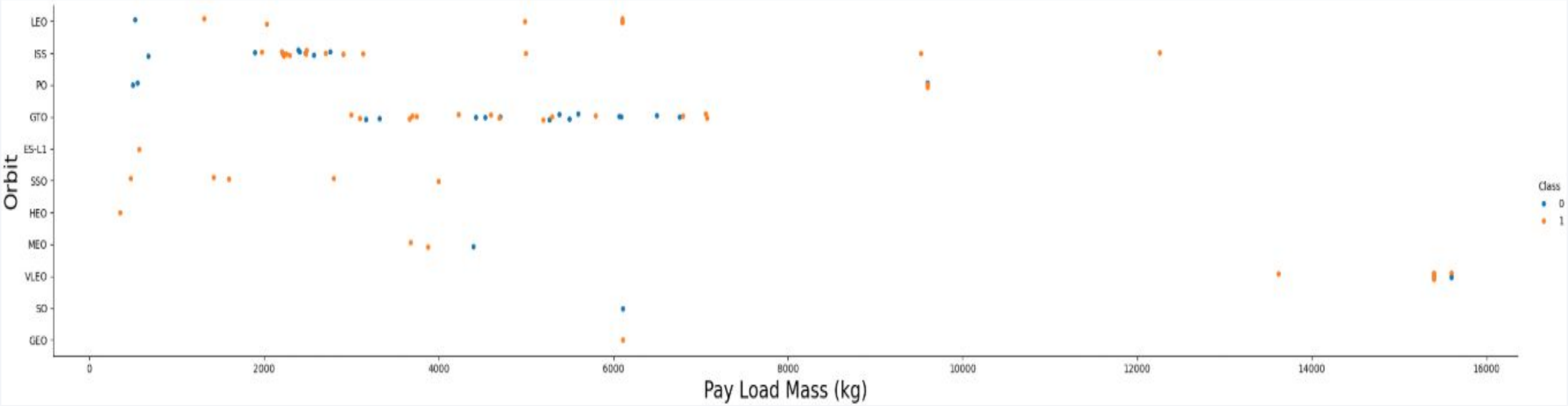
Flight Number vs. Orbit Type

- There is a general trend of greater success the more attempts that are made
  - 0 failed landings (blue dot)
  - 1 successful landings (orange dot)



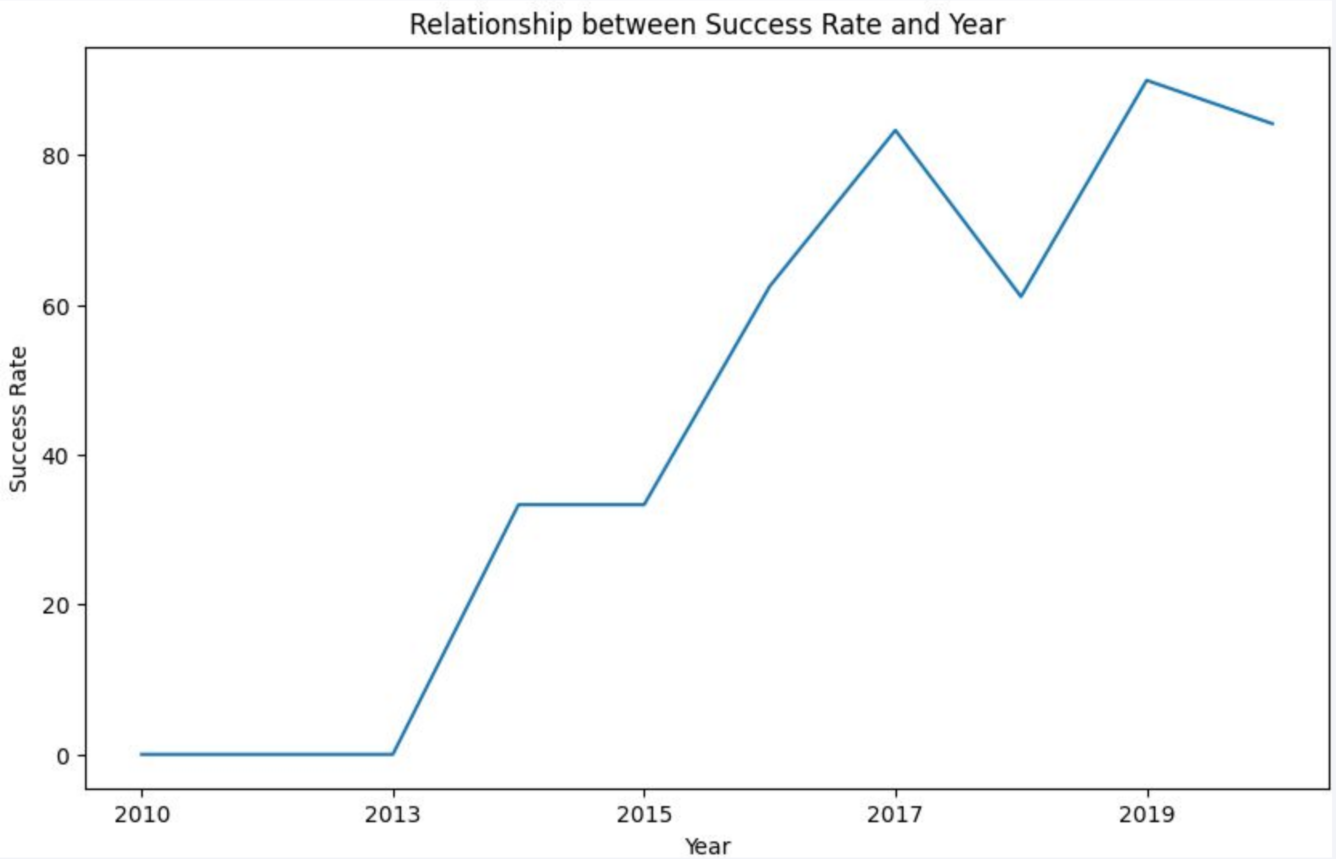
Payload vs. Orbit Type

- There does not appear to be a correlation between success rate between achieved orbit and payload mass
  - 0 failed landings (blue dot)
  - 1 successful landings (orange dot)



Launch Success Yearly Trend

- The rate of success increased each year from 2013 until 2018 when there was a dip and then increased again afterwards.



All Launch Site Names

- **Task:** Find all the launch site names.
- **Query:** **SELECT DISTINCT** Launch_Site **FROM** SPACEXTABLE;
- **Explanation:** There are four unique launch sites.

Query Returned:

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

Launch Site Names Begin with 'CCA'

- **Task:** Retrieve 5 records of launch sites beginning with ‘CCA’
- **Query:** **SELECT** * **FROM** SPACEXTABLE **WHERE** Launch_Site **LIKE**"CCA%" **LIMIT** 5;
- **Explanation:** Returns a table providing insights to the type of data contained in the database.

Query Returned:

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

Total Payload Mass

- **Task:** Calculate the total payload carried by boosters from NASA
- **Query:** SELECT SUM(PAYLOAD_MASS__KG_) AS 'Total Payload Mass (kg)' FROM SPACEXTABLE WHERE Customer LIKE '%NASA (CRS)%';
- **Explanation:** The total payload carried by all lossters launched by Nasa (CRS)

Query Returned:

Total Payload Mass (kg)
48213

Average Payload Mass by F9 v1.1

- **Task:** Calculate the average payload mass carried by booster version F9 v1.1
- **Query:** **SELECT AVG(**PAYLOAD_MASS__KG_**) FROM** SPACEXTABLE **WHERE** Booster_Version **LIKE** "F9 v1.1";
- **Explanation:** The average mass carried by booster version F9 v1.1

Query Returned:

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

- **Task:** Find the dates of the first successful landing outcome on ground pad
- **Query:** SELECT MIN(Date) FROM SPACEXTABLE WHERE Landing_Outcome LIKE 'Success (ground pad)';
- **Explanation:** First date of the a successful landing outcome on ground pad

Query Returned:

min(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- **Task:** List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000
- **Query:** **SELECT** Booster_Version **FROM** SPACEXTABLE **WHERE** Landing_Outcome **LIKE** 'Success (drone ship)' **AND** PAYLOAD_MASS__KG_ **BETWEEN** 4000 **AND** 6000;
- **Explanation:** All the names of the boosters to successfully carry by drone ship payloads between 4000 kg and 6000 kg.

Query Returned:

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

Total Number of Successful and Failure Mission Outcomes

- **Task:** Calculate the total number of successful and failure mission outcomes
- **Query:** **SELECT(SELECT COUNT(*) FROM** SPACEXTABLE **WHERE** Landing_Outcome **LIKE** '%success%') AS 'Success', **COUNT(*) AS** 'Failure' **FROM** SPACEXTABLE **WHERE** Landing_Outcome **NOT LIKE** '%success%';
- **Explanation:** The total number of successful (61) and failed (40) missions

Query Returned

Success	Failure
61	40

Boosters Carried Maximum Payload

- **Task:** List the names of the booster which have carried the maximum payload mass
- **Query:** **SELECT** Booster_Version, PAYLOAD_MASS__KG_ **FROM** SPACEXTABLE **WHERE** PAYLOAD_MASS__KG_ = **(SELECT MAX(**PAYLOAD_MASS__KG_**) FROM** SPACEXTABLE**)**;
- **Explanation:** Names of the boosters that carried the most weight, which is 15600 kg

Query Returned (see image to the right)

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

- **Task:** List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- **Query: SELECT** substr(Date, 6, 2) AS 'Month', Landing_Outcome, Booster_Version, Launch_Site **FROM** SPACEXTABLE **WHERE** Landing_Outcome='Failure (drone ship)' **AND** substr(Date, 0, 5)='2015';
- **Explanation:** There are two instances of failure in drone ship landings in 2015, both at the CCAFS LC-40 launch site.

Query Returned:

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

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

- **Task:** 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
- **Query:** **SELECT** 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(**Landing_Outcome**)** **DESC**;
- **Explanation:** The most common outcome was “No attempt”

Query Returned: see image to the right

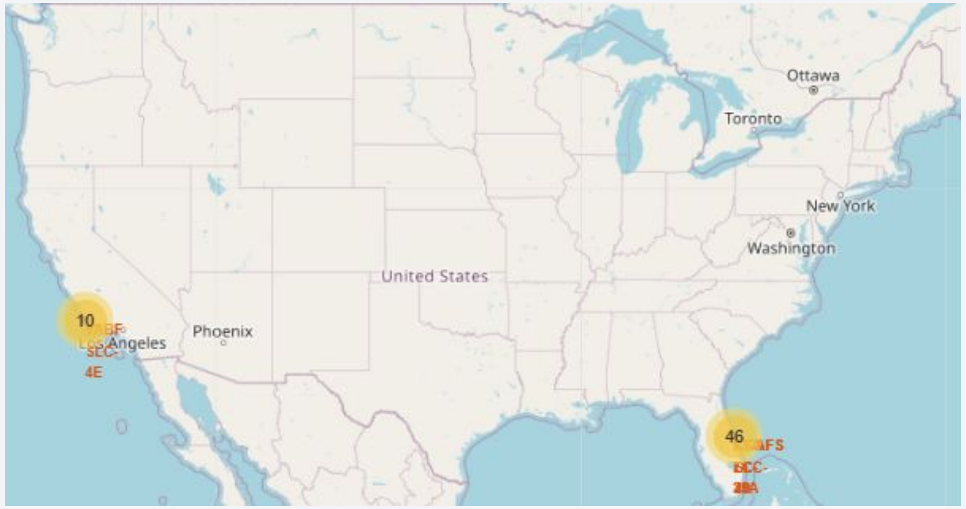
Landing_Outcome	Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

Falcon 9 Launch Site Locations

- California, USA
  - VAFB SLC-4E | Vandenberg Space Launch Complex 4
- Florida, USA
  - KSC LC-39A | Kennedy Space Center Launch Complex 39A
  - CCAFS SLC-40 | Cape Canaveral Space Launch Complex 40
  - CCAFS LC-40 | Cape Canaveral Space Launch Complex 40



- CCAFS SLC-40 and CCAFS LC-40 are the same location

Map Markers for Successful/Failed Landings

- Each marker on the map corresponds with a launch site's outcome, either successful or failure for the Falcon 9 first stage landings. Each marker cluster is located on the map where the launch site is located.

VAFB SLC-4E

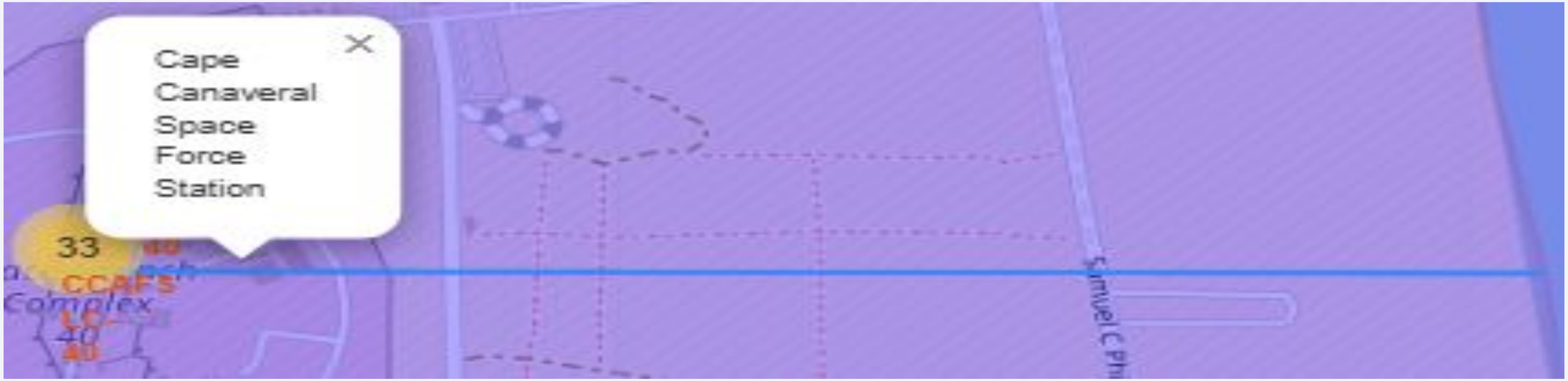
KSC LC-39A

CCAFS LC-40

CCAFS SLC-40

Distance from CCAFS LC-40 to Coastline

- Explore the generated folium map and show the screenshot of a selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed
- The pop-up indicates that the launch site location is Cape Canaveral Space Force Station and the blue line perpendicular to the location is the marker distance to the coastline.

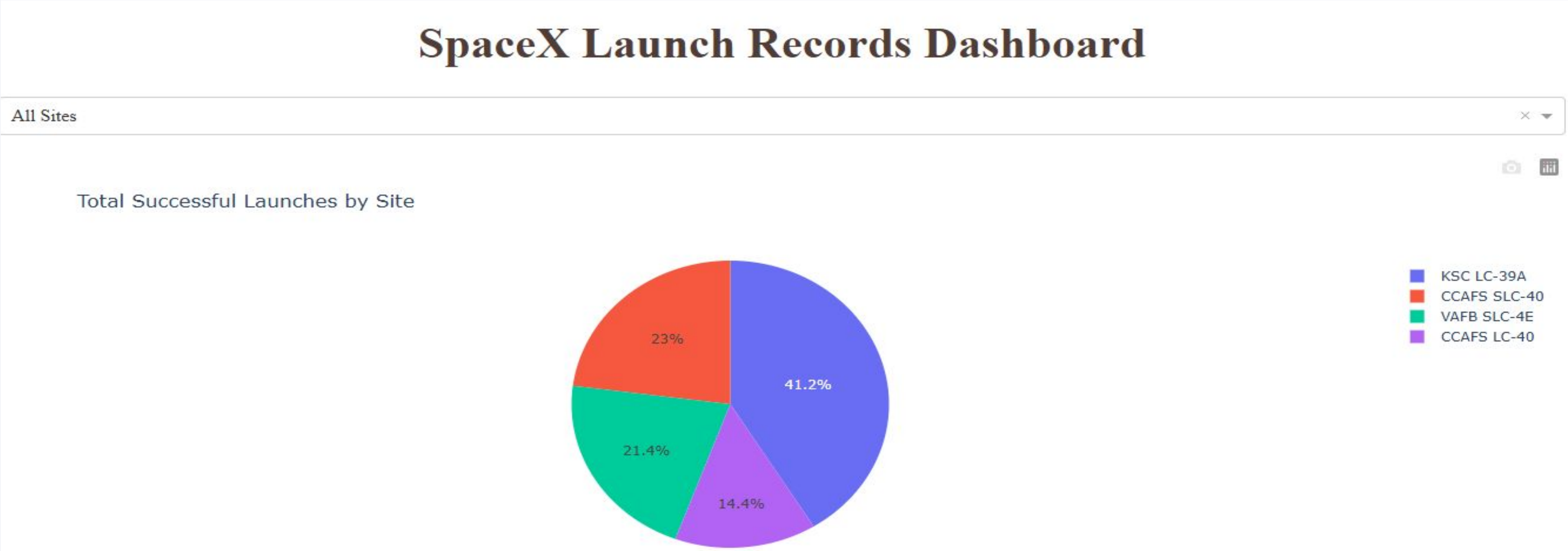


Section 4

Build a Dashboard with Plotly Dash

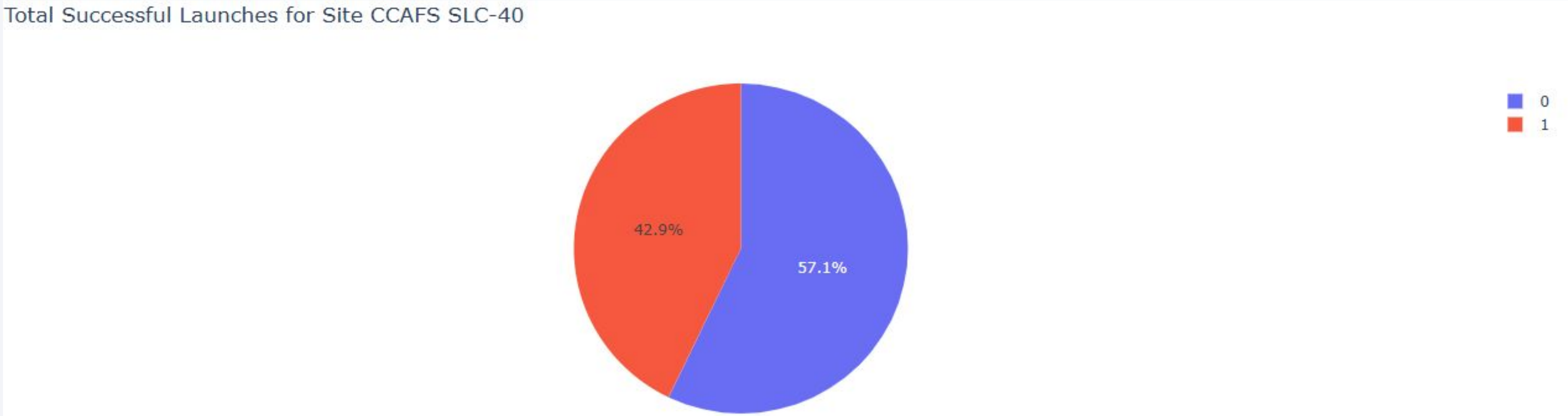
Launch Success Count for All Sites

- Dropdown menu allows for selecting an individual site or all sites
- Selecting “All Sites” displays the success rate for the Falcon 9 first stage landing for each site location.
- KSC LC-39A has the highest success rate of all sites at 49.2%



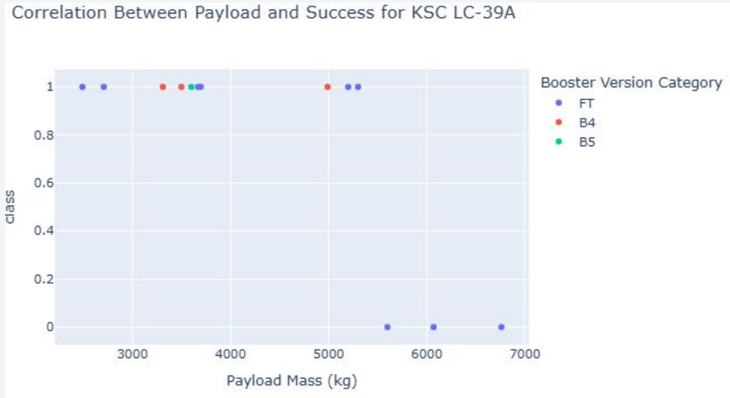
Highest Success Launch Ratio Across all Launch Sites

- Falcon 9 first stage landing success rate is indicated with 1 (red slice) and failure is indicated with 0 (blue slice)
- CCAFS SLC-40 has the highest ratio of success compared to failures across all launch sites with a first stage landing success rate of 42.9%

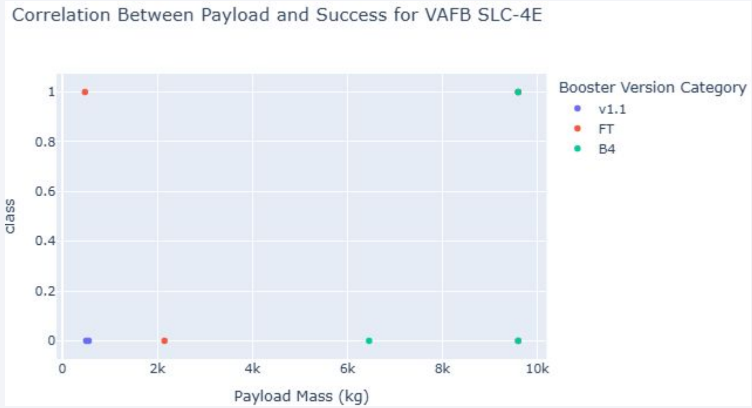


Payload vs Launch Outcomes

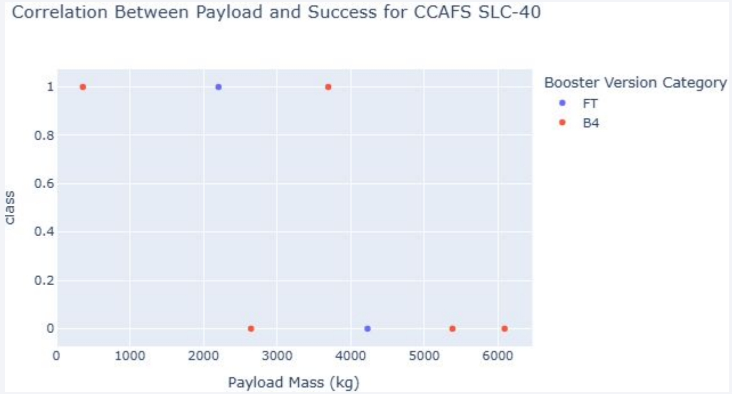
- These screenshots are of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider
- The payload range with the most success is between 2,000 kg and 5,000 kg.
- The ‘FT’ booster has the largest success rate of all the booster types.



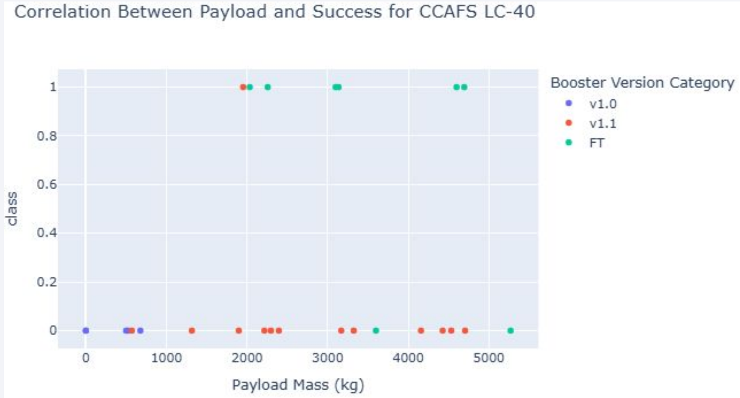
KSC LC-39A



VAFB SLC-4E



CCAFS SLC-40



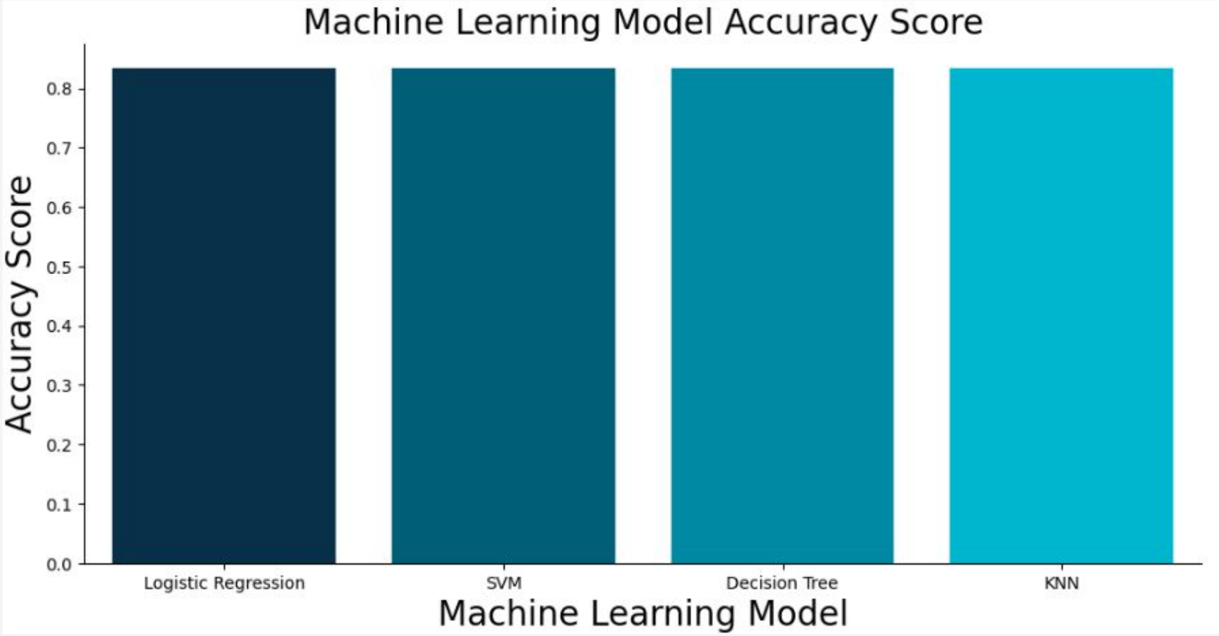
CCAFS LC-40

Section 5

Predictive Analysis (Classification)

Classification Accuracy

- All the models performed equally well.
    - Logistic Regression
      Score: .833333
    - SVM
      Score: .833333
    - Decision Tree
      Score: .833333
    - KNN
      Score: .833333

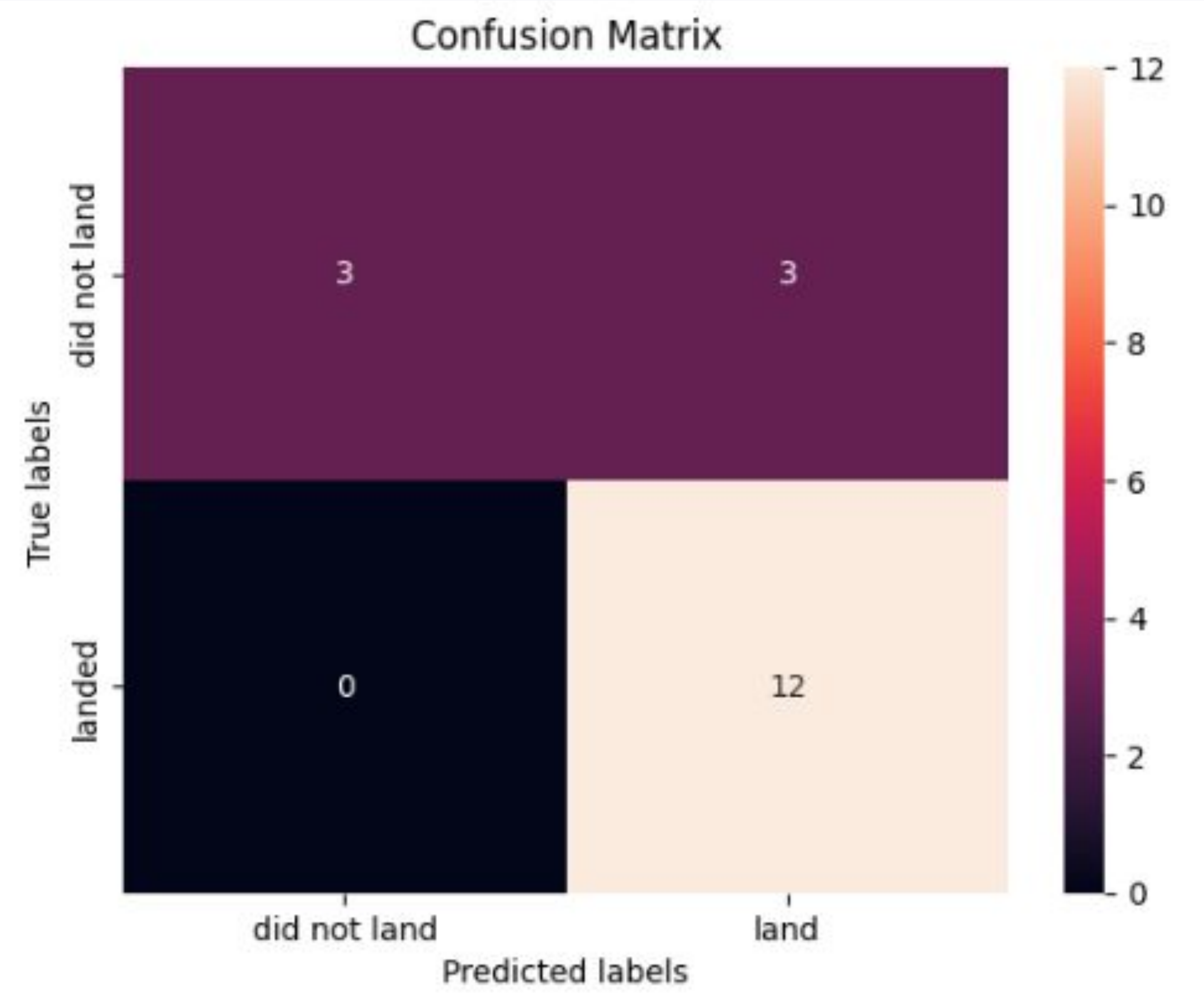


Confusion Matrix

- Shown here is the Logistic Regression confusion matrix (representative of all models)

- The predictions are as follows:
  - True Negative 3
  - True Positive 12
  - False Negative 0
  - False Positive 3

True Negative	False Positive
False Negative	True Positive



CONCLUSION

Conclusions

- In general, the trend has been more successful first stage landing outcomes for the Falcon 9 over time.
- The trend has improved with more launch attempts.
- The machine learning models are capable of accurately predicting future SpaceX Falcon 9 first stage landing outcomes.

APPENDIX

Appendix

- This appendix includes figures that were too large to fit onto the related slides, all figures are on slides 52-56.
- Project Github code repository found here

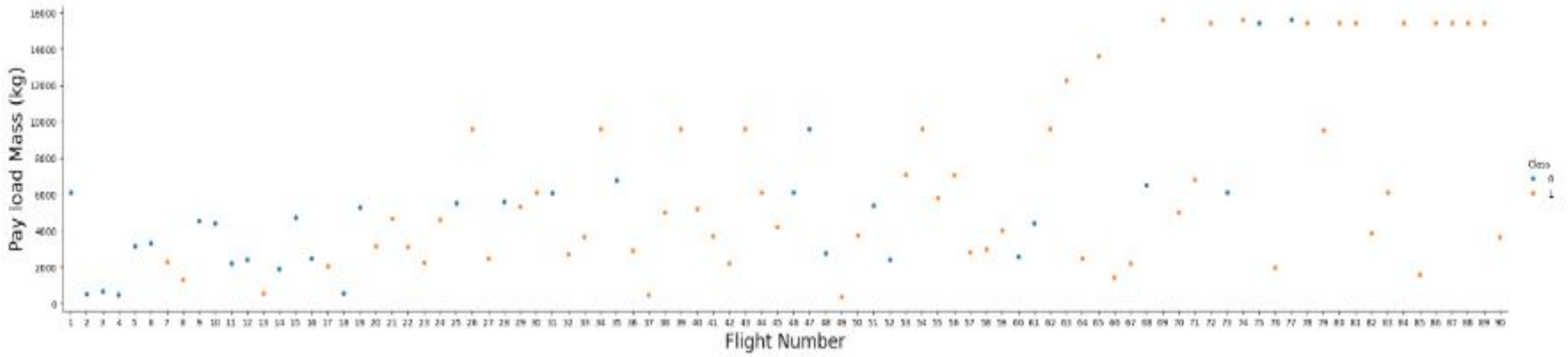


Figure 1

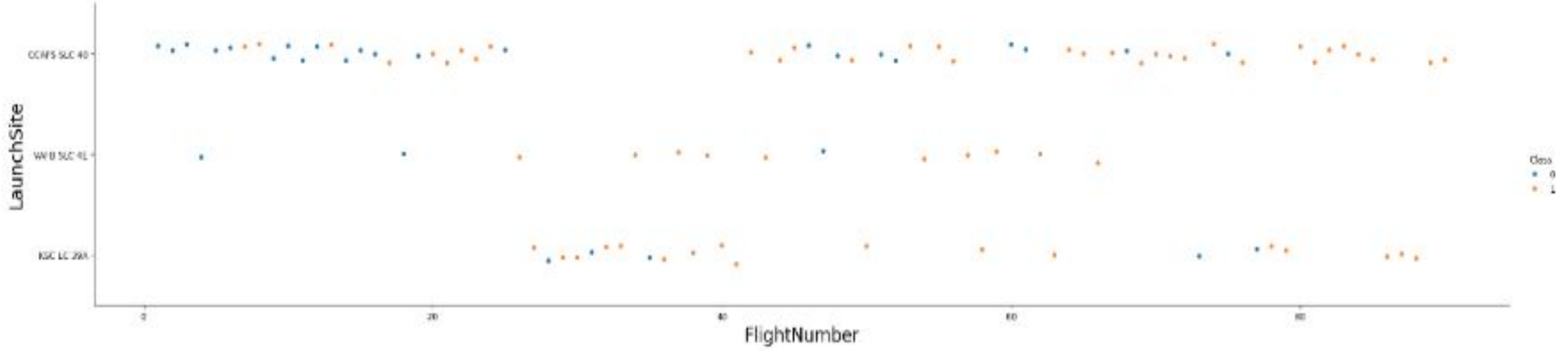


Figure 2

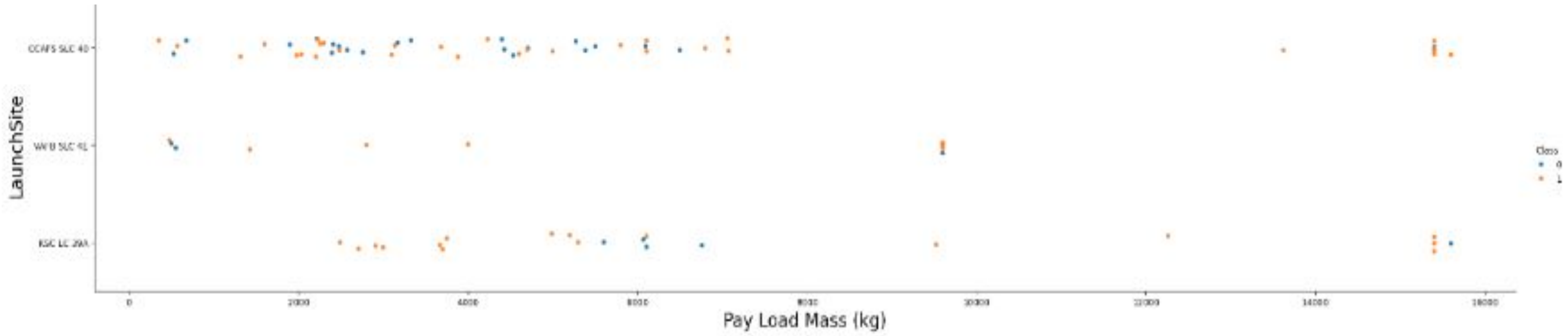


Figure 3

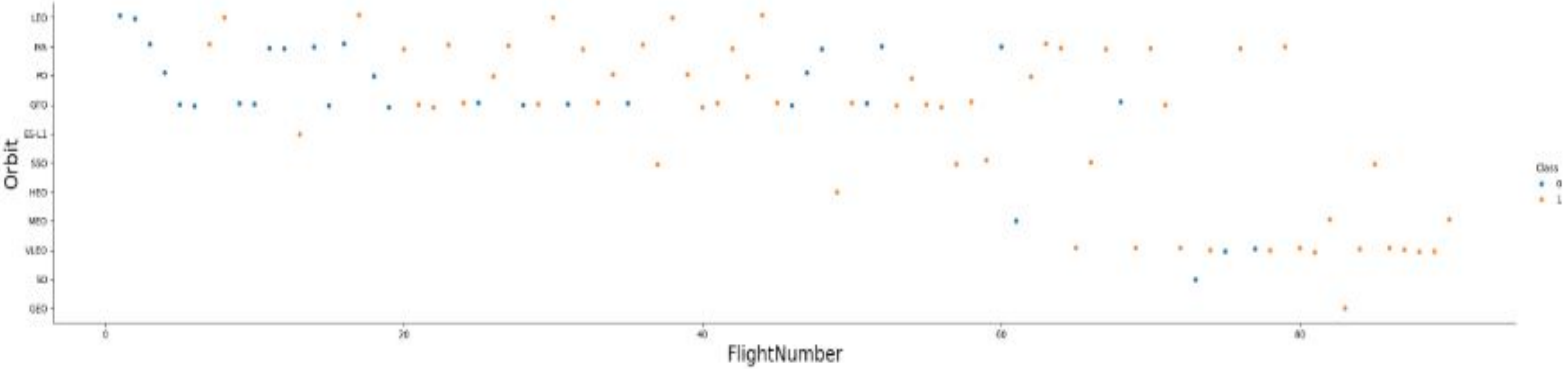
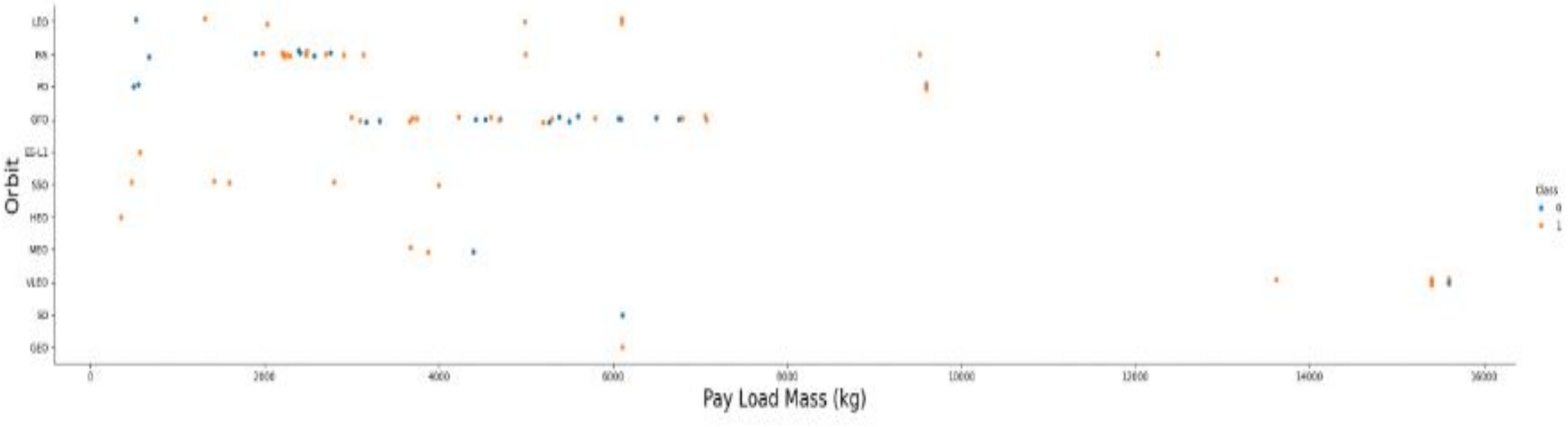


Figure 4



Orbit
LEO
ISS
PO
GTO
ES-L1
SSO
HEO
MEO
VLEO
SO
GEO
Pay Load Mass (kg)
Class
0
1

Figure 5

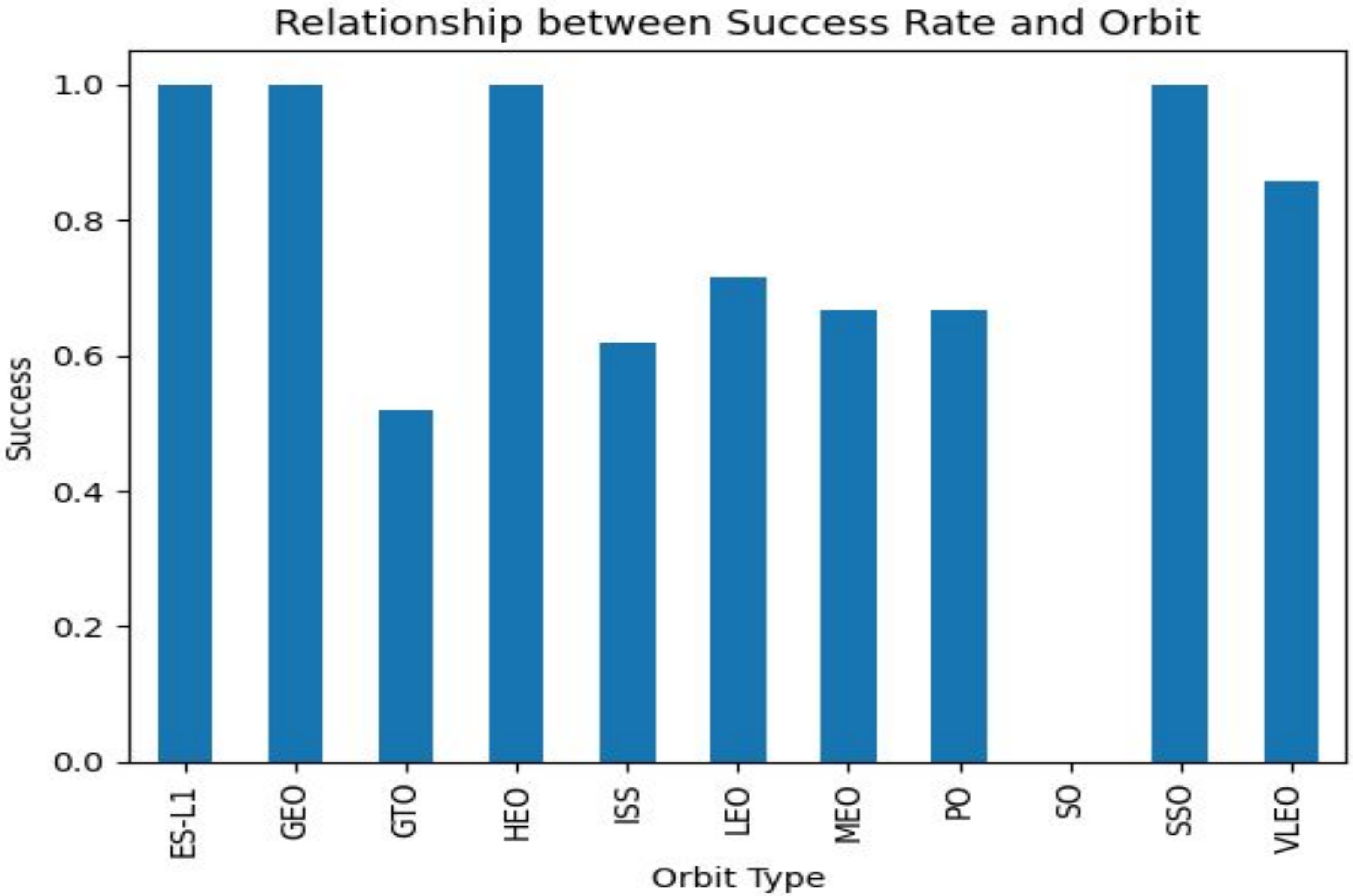

Relationship between Success Rate and Orbit
1.0
0.8
0.6
0.4
0.2
0.0
Success
ES-L1
GEO
GTO
HEO
ISS
LEO
MEO
PO
SO
SSO
VLEO
Orbit Type

Figure 6

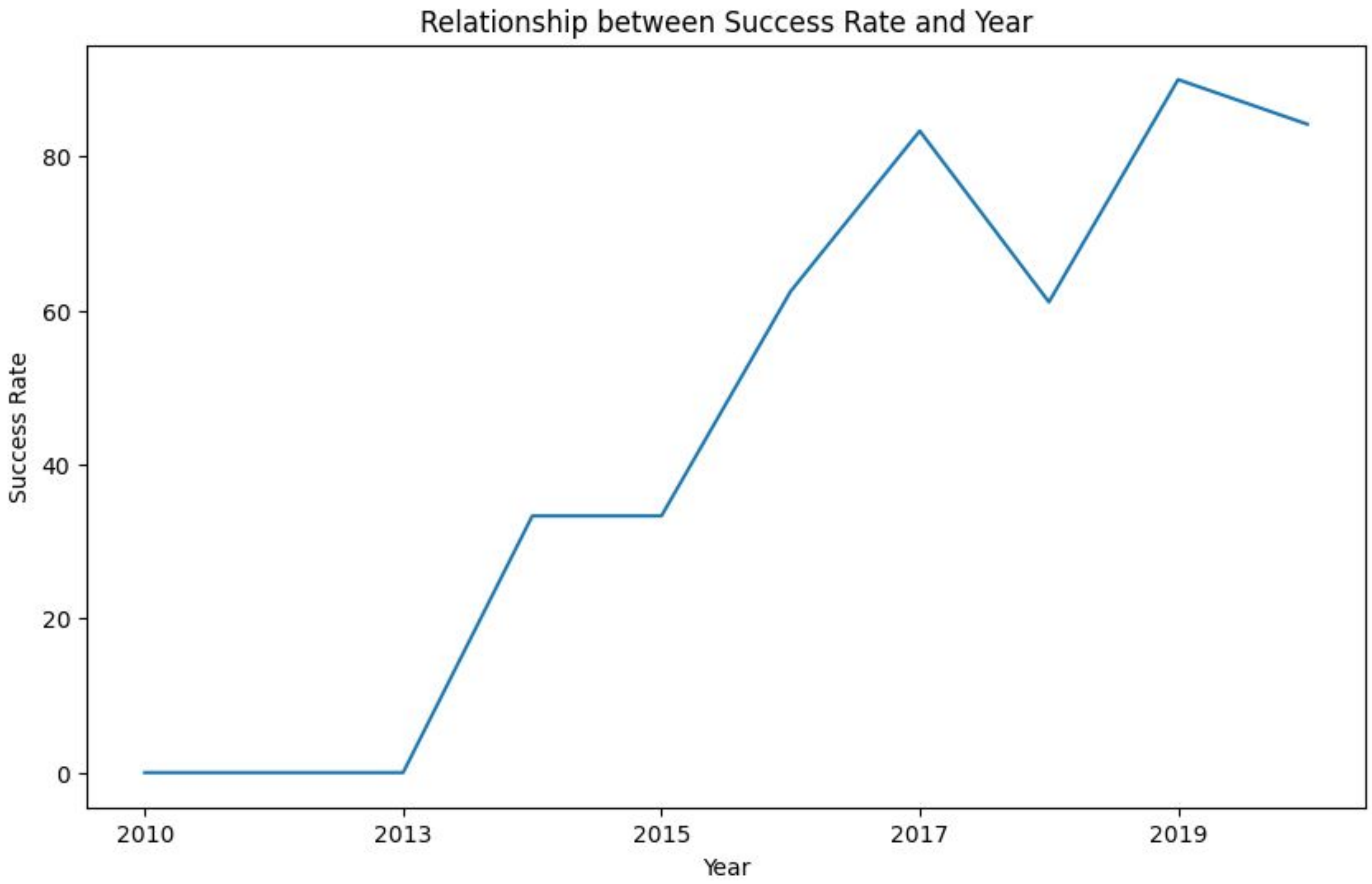

Relationship between Success Rate and Year
Success Rate
0
20
40
60
80
2010
2013
2015
2017
2019
Year

Figure 7

Thank you!