IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Coleman Wenzl
07Feb2024

Outline

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

Executive Summary

- The following methodologies were used to analyze data:
  - Data collection was performed using web scraping and Space X API.
  - Data wrangling, Exploratory Data Analysis (EDA), data visualization, and interactive visual analytics.
  - Machine learning for prediction analysis.
- Summary of all results
  - Data was able to be collected via public sources.
  - EDA was successful in identifying which features are best for predicting a successful first stage landing.
  - Machine learning showed the optimal model to predict which characteristics are important in diving this opportunity.

Introduction

- The purpose of this project is to determine the viability of a new company, Space Y, to compete with Space X.
- Desirable Answers:
  - What is the best way to estimate cost of launches by predicting the success rate of the first strange landings, allowing for reuse.
  - Where is the optimal launch site location.

Section 1

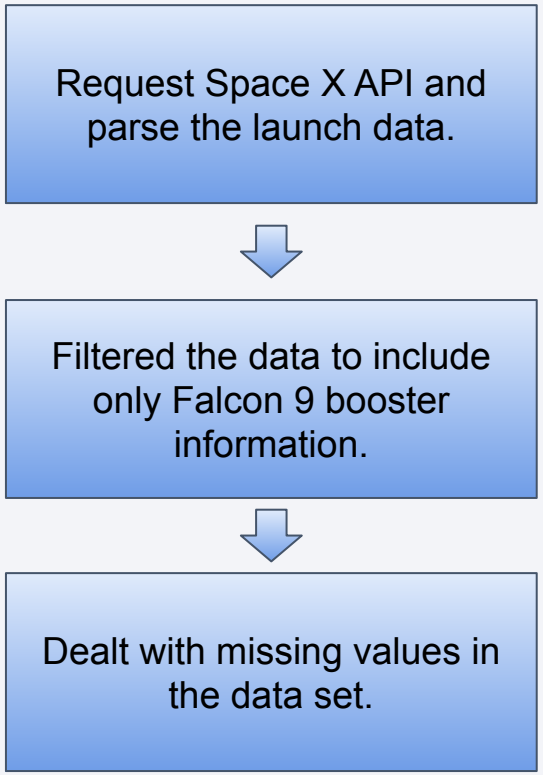
Methodology

Methodology - Executive Summary

- Data collection methodology:
  - Space X data was collected from multiple sources:
    - Space X API : https://api.spacexdata.com/v4
    - Web scraping Space X wikipedia:
      https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922
- Perform data wrangling
  - The collected data was
- 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

- Space X launch data was obtained through the open source API.
- The provided public API was used following the flow chart provided.

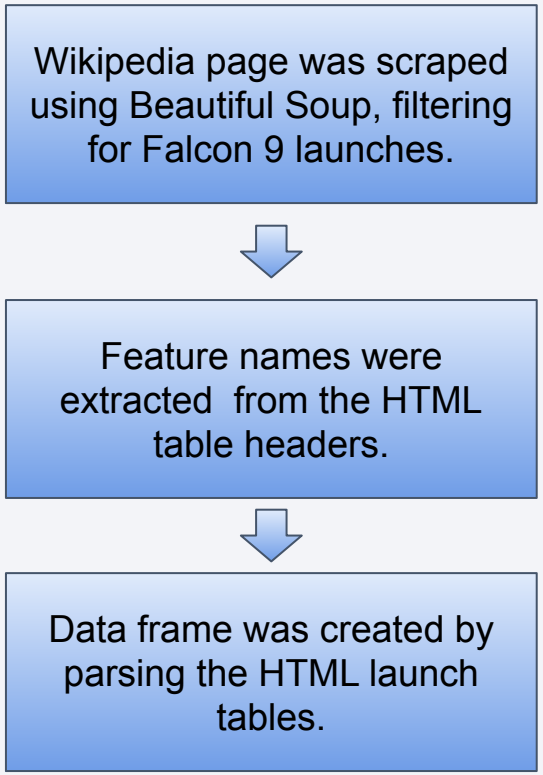


GitHub URL: [API Data Collection]

Data Collection - Scraping

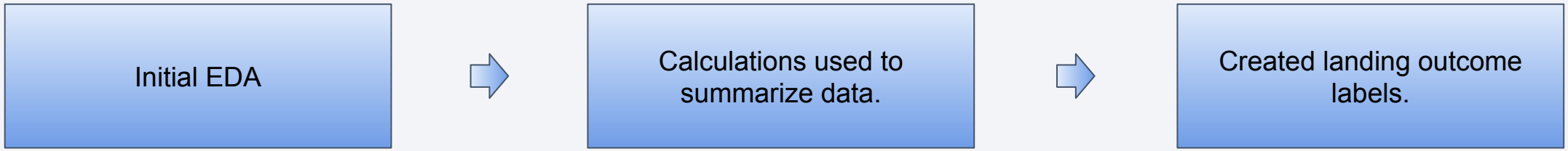
- Space X launch data was obtained form Wikipedia.
- Data was collected and organized following the flow chart provided.

GitHub URL: Web Scraping



Data Wrangling

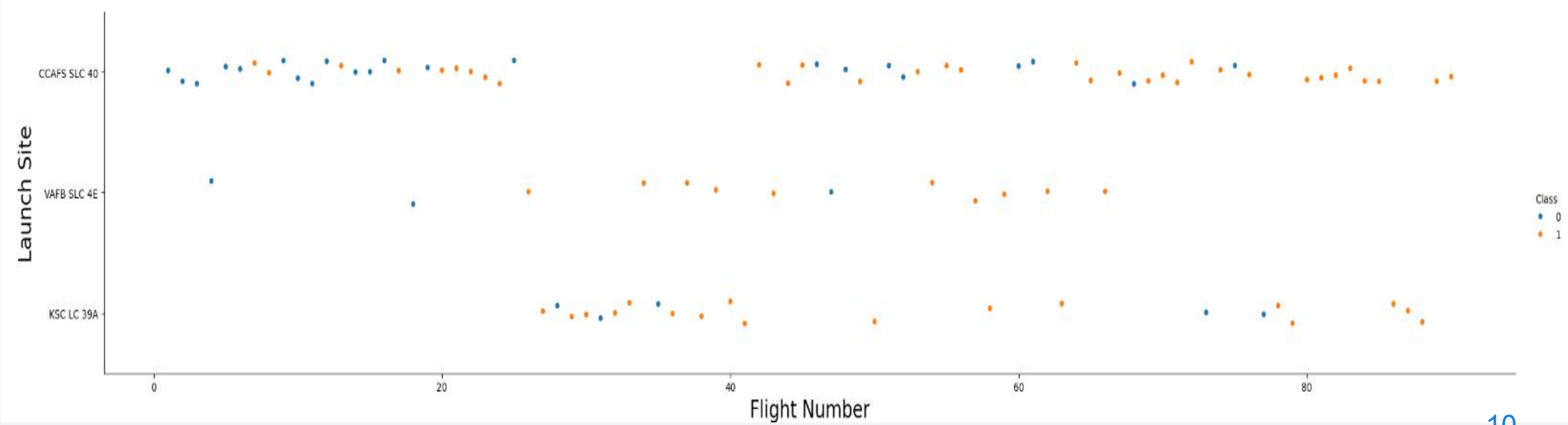
- Initial EDA was performed.
- The number of launches at each site, occurrence at each orbit, mission outcome for each orbit, and landing outcomes were calculated.



GitHub URL: Data Wrangling

EDA with Data Visualization

- Charts, such as scatter, bar, and line, were created to visualize the relationship between different feature pairs.
  - Flight Number x Launch Sites, Payload x Launch Site, Success Rate x Orbit Type, Flightnumber x Orbit Type, Payload x Orbit Type, and Launch Success Yearly Trend.



GitHub: [EDA Data Vizulization]

EDA with SQL

- The following SQL queries were performed:
  - The names of each unique launch site.
  - Show the top 5 entries for ‘CCA’ launch site.
  - Display the the total payload mass by boosters launched by NASA.
  - Average payload mass carried by booster version F9 v1.1.
  - When the first successful landing on a ground pad was achieved.
  - Queried the names of boosters that has successful drone ship landings with payload mass between 4000 and 6000 kg.
  - Counted the total number of successful and failed missions.
  - Listed the names of boosters that have carried the maximum payload mass.
  - Listed the missions that occurred in 2015 by month.
  - Listed the landing outcomes by occurrence between 2010 and 2017.

GitHub URL: EDA with SQL

Build an Interactive Map with Folium

- Each launch site location was marked with a circular maker.
  - The number indicated on the marker is how many launches has been performed.
- By zooming in to a launch site, a marker cluster can be viewed indicating the successful and failed landings of that launch site.
- Lines were added to launch site CCAFS-LC40 to indicate distance from different geographical features.

- These objects were added for quick reference of launch data in relation to geographical area and features.

GitHub: Folium Map

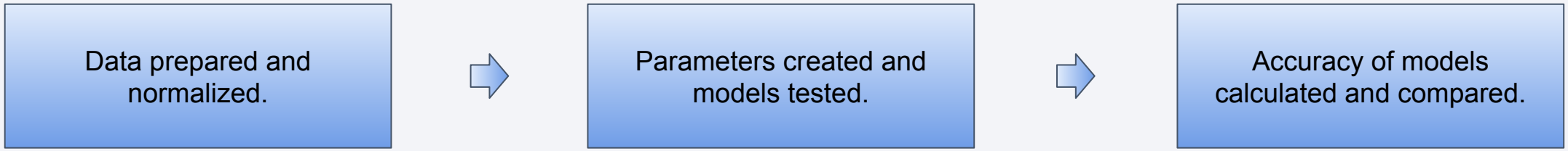
Build a Dashboard with Plotly Dash

- There were two type of graphs used to visualize the data:
  - A pie chart to summarize the success and failures by launch site.
  - A scatter plot to show payload range in relation to launch site and booster type and the effects on landing success.
- The combination of these plots allow for quick analysis of launch sites, payloads, and booster version in relation to success rates.

GitHub URL: Plotly Dash

Predictive Analysis (Classification)

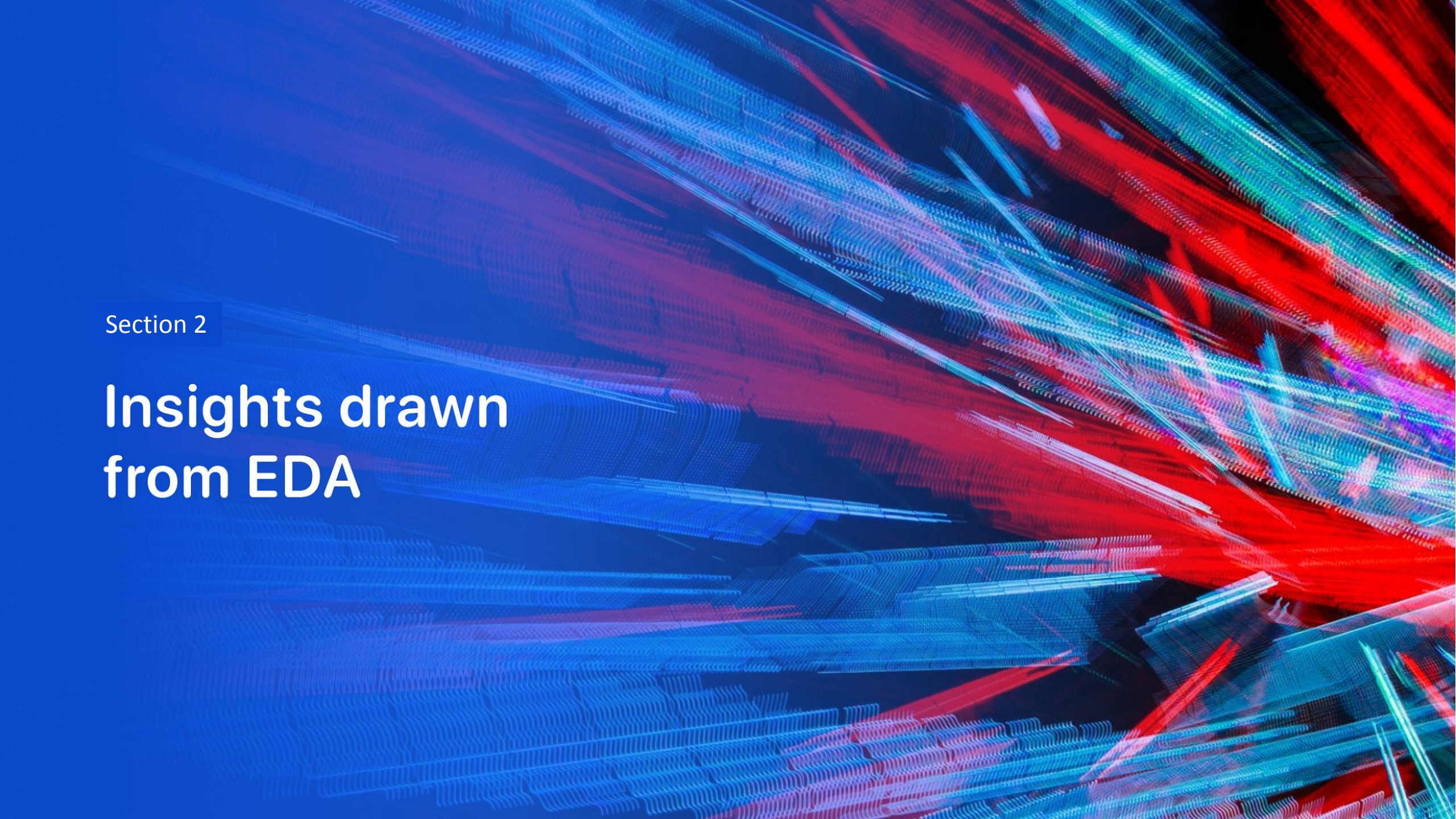
- The data set was split into test and train data
- Multiple models were then compared: logistic regression, support vector machine, decision tree, and k nearest neighbors.



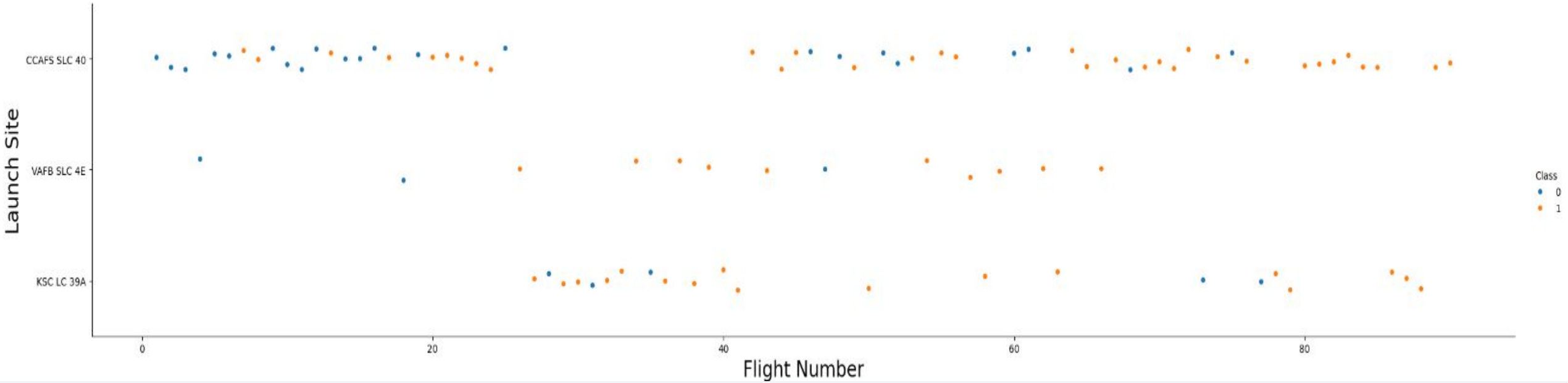
GitHub URL: Predictive Analysis

Section 2

Insights drawn from EDA

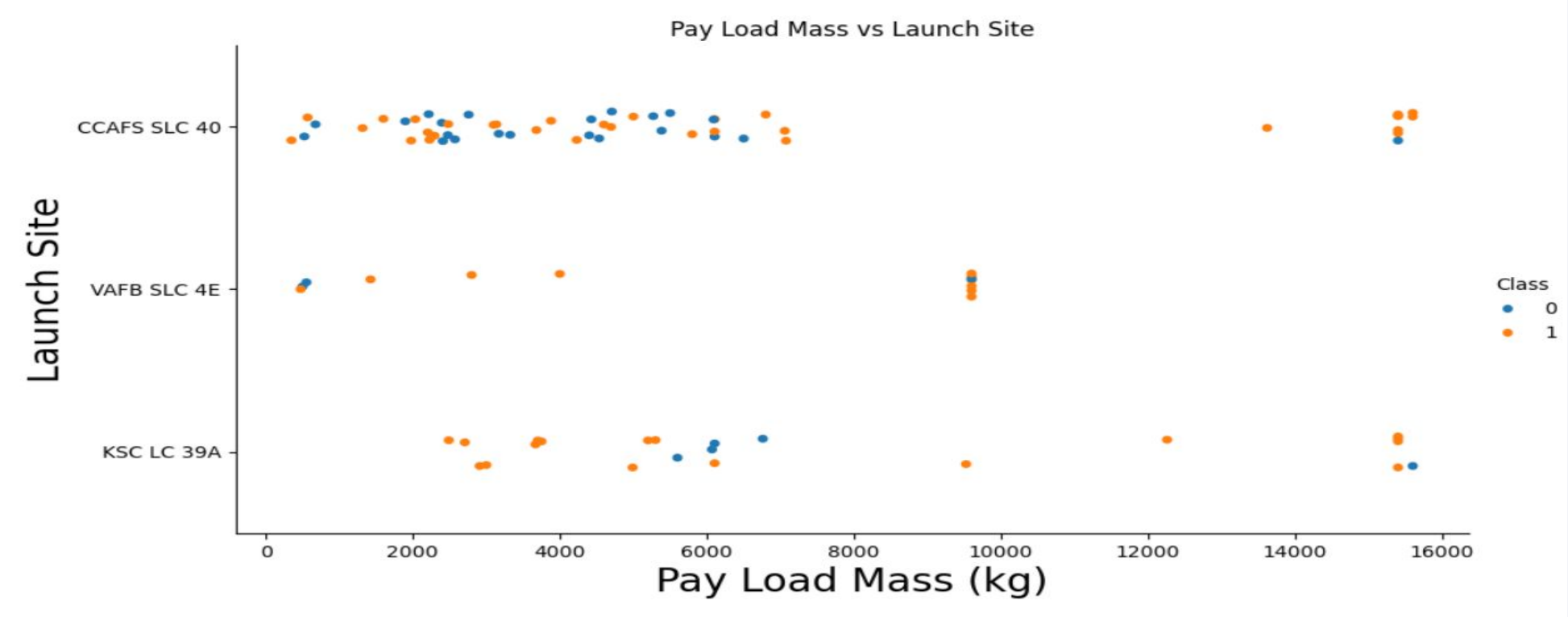


Flight Number vs. Launch Site



- It can be seen that landing success rate has increased over time, with the increasing flight number.
- The CCAFS SLC 40 site has been the most used site over time and could be inferred to be the most reliable site to recreate a successful landing.
- Following the CCAFS site would be the KSC site then the VAFB site.

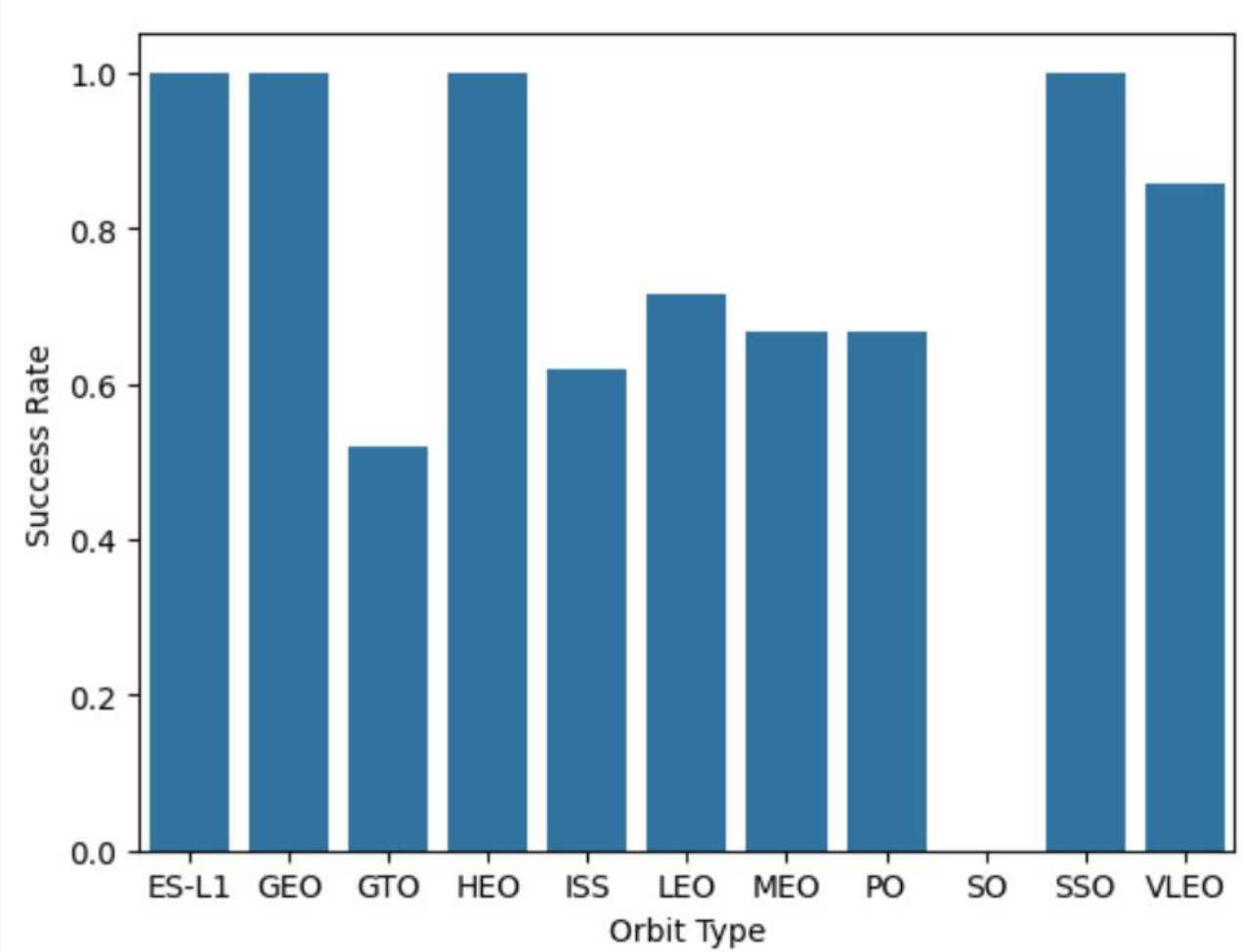
Payload vs. Launch Site



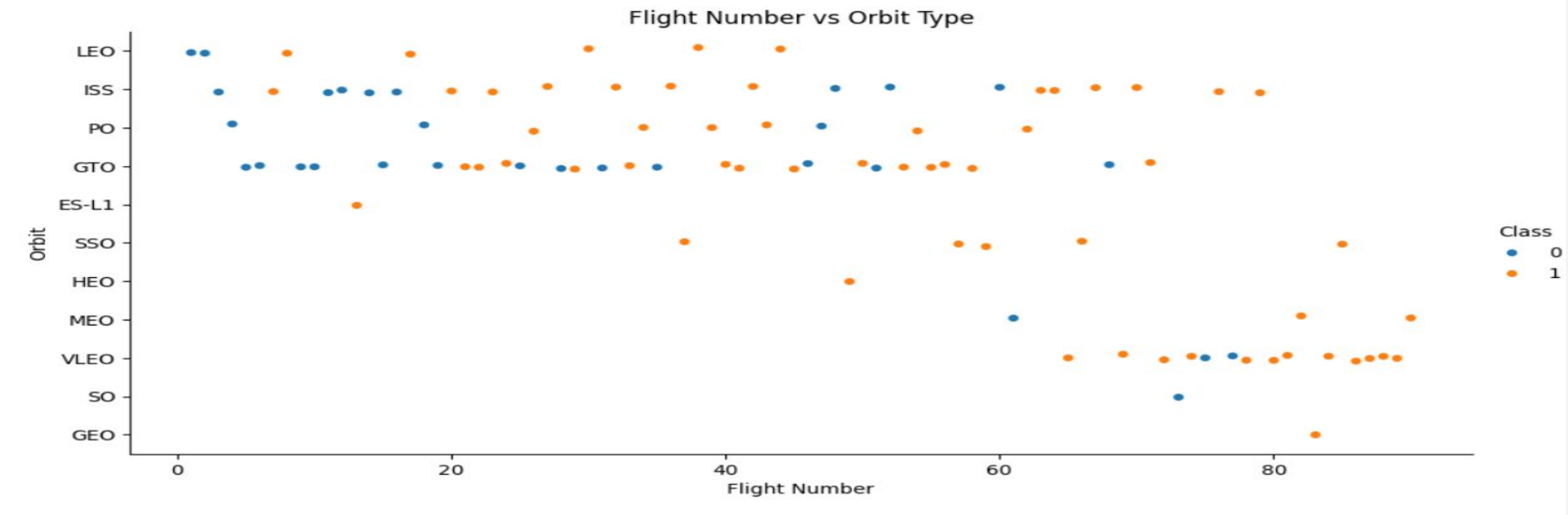
- Higher success rates are seen with payloads over 8,000kg.
- Payloads over 12,000kg are only possible at sites CCAFS and KSC.
- Payloads just under 10,000kg are highly successful at the VAFB site.

Success Rate vs. Orbit Type

- The highest landing success rates are at orbit types:
  - ES-L1
  - GEO
  - HEO
  - SSO
- The orbit type with the lowest landing rate was SO.

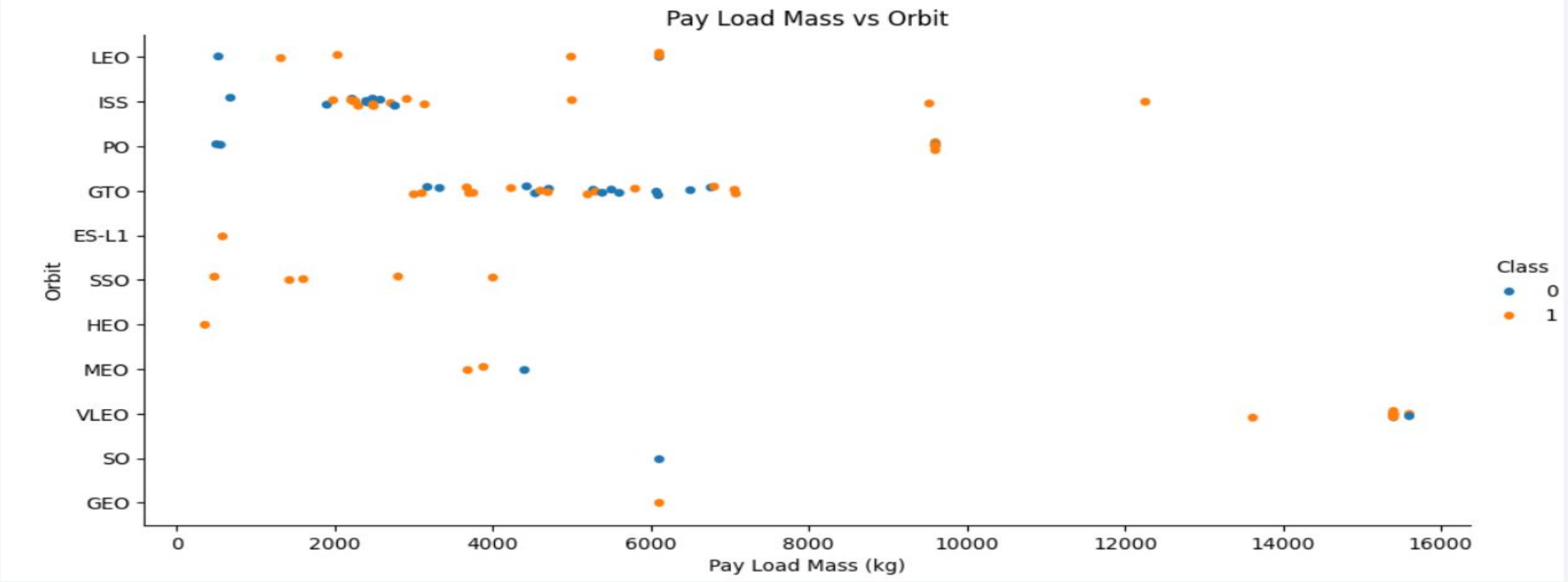


Flight Number vs. Orbit Type



- Success rate of all orbits increased over time.
- Missions to the VLEO orbit are quite frequent in later launches with high success rates.

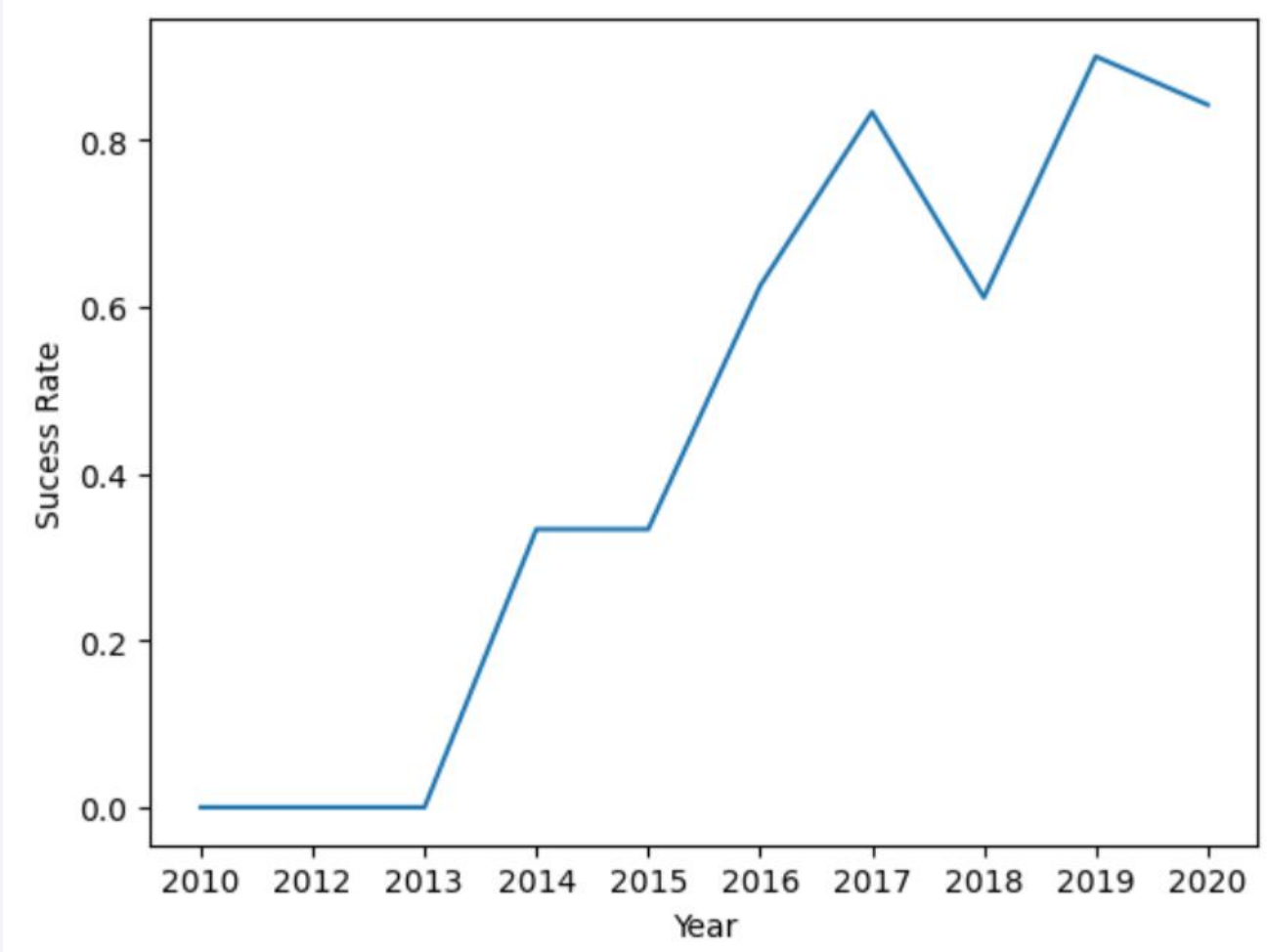
Payload vs. Orbit Type



- VLEO orbit missions require the highest payloads.
- There looks to be little correlation between payload and success rate in GTO orbit.
- ISS orbit missions have the highest variance of payloads with a good success rate.

Launch Success Yearly Trend

- The average success rate of landings increased over time.
- There was a large drop in success rate in 2018.
-
- Show the screenshot of the scatter plot with explanations



All Launch Site Names

- Using a SQL query, a list of all launch site names were found and listed.

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

Launch Site Names Begin with 'CCA'

- The top 5 records for launch sites with CCA are shown to the right.
- It can be seen that there is some missing data that would need to be corrected.

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

- The calculated payload mass carried by NASA boosters is listed below:

total_sum
48213

Average Payload Mass by F9 v1.1

- The calculated average payload mass of booster version F9 v1.1 is:

total_sum
2534.6666666666665

First Successful Ground Landing Date

- The first successful ground landing date was:

MIN(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- The names of boosters which have successfully landed on drone ship and had payload mass between 4000kg and 6000kg are:

Booster_Version		
F9 v1.1	F9 FT B1032.1	F9 B5 B1048.3
F9 v1.1 B1011	F9 B4 B1040.1	F9 B5 B1051.2
F9 v1.1 B1014	F9 FT B1031.2	F9 B5B1060.1
F9 v1.1 B1016	F9 B4 B1043.1	F9 B5 B1058.2
F9 FT B1020	F9 FT B1032.2	F9 B5B1062.1
F9 FT B1022	F9 B4 B1040.2	
F9 FT B1026	F9 B5 B1046.2	
F9 FT B1030	F9 B5 B1047.2	
F9 FT B1021.2	F9 B5B1054	

Total Number of Successful and Failure Mission Outcomes

- The calculated total number of successful and failed mission outcomes are:

Mission_Success	Mission_Failure
100	1

- This calculation is the overall mission and not specifically the landing outcome, which has been used in previous data analysis.
- There seems to be only on mission failure.

Boosters Carried Maximum Payload

- A list of the boosters that have carried the maximum was compiled.
- There are a total of 12 boosters that have carried the maximum payload, shown in a list below:

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

2015 Launch Records

- Below is a list of failed landings on drone ships in the year 2015, listed by month in numerical format:

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

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

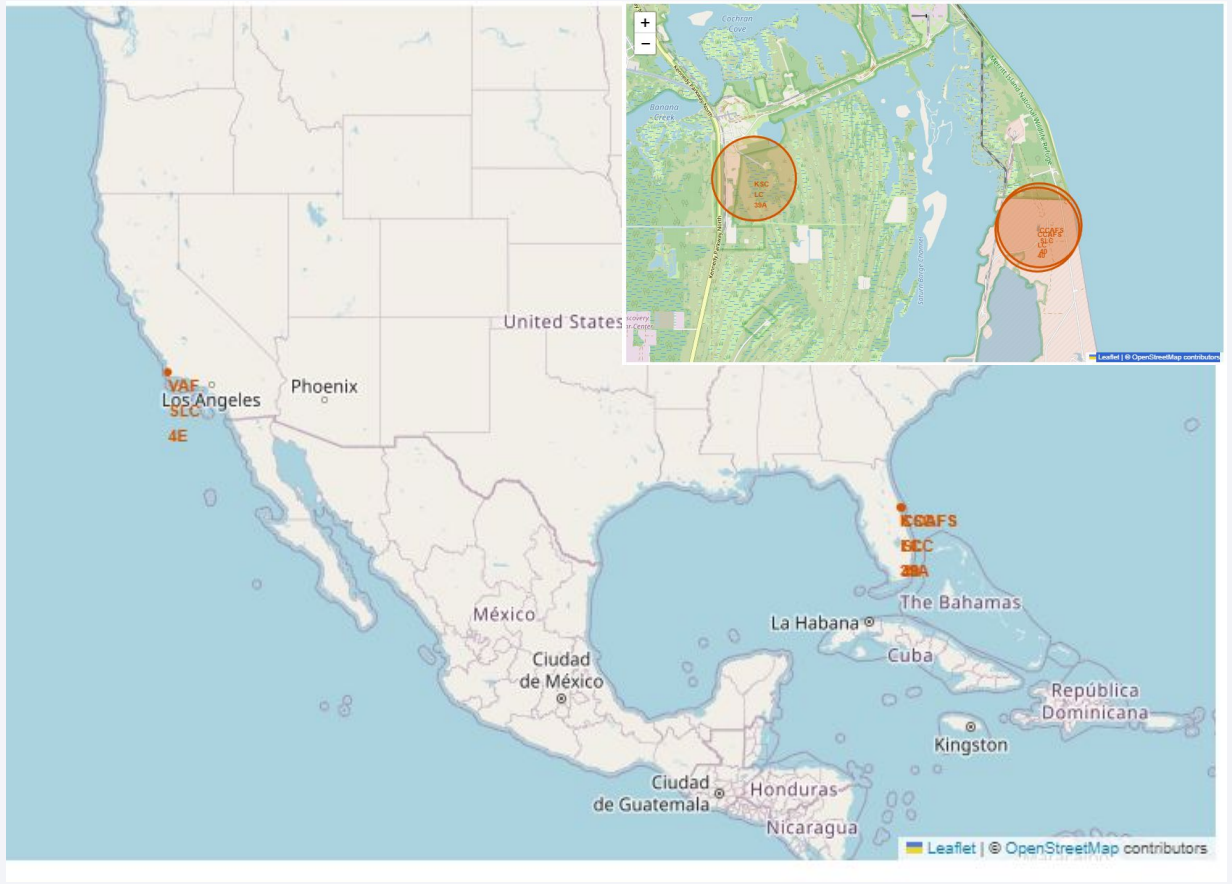
- The count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, are ranked in descending order below:

Landing_Outcome	Landing_Outcomes
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

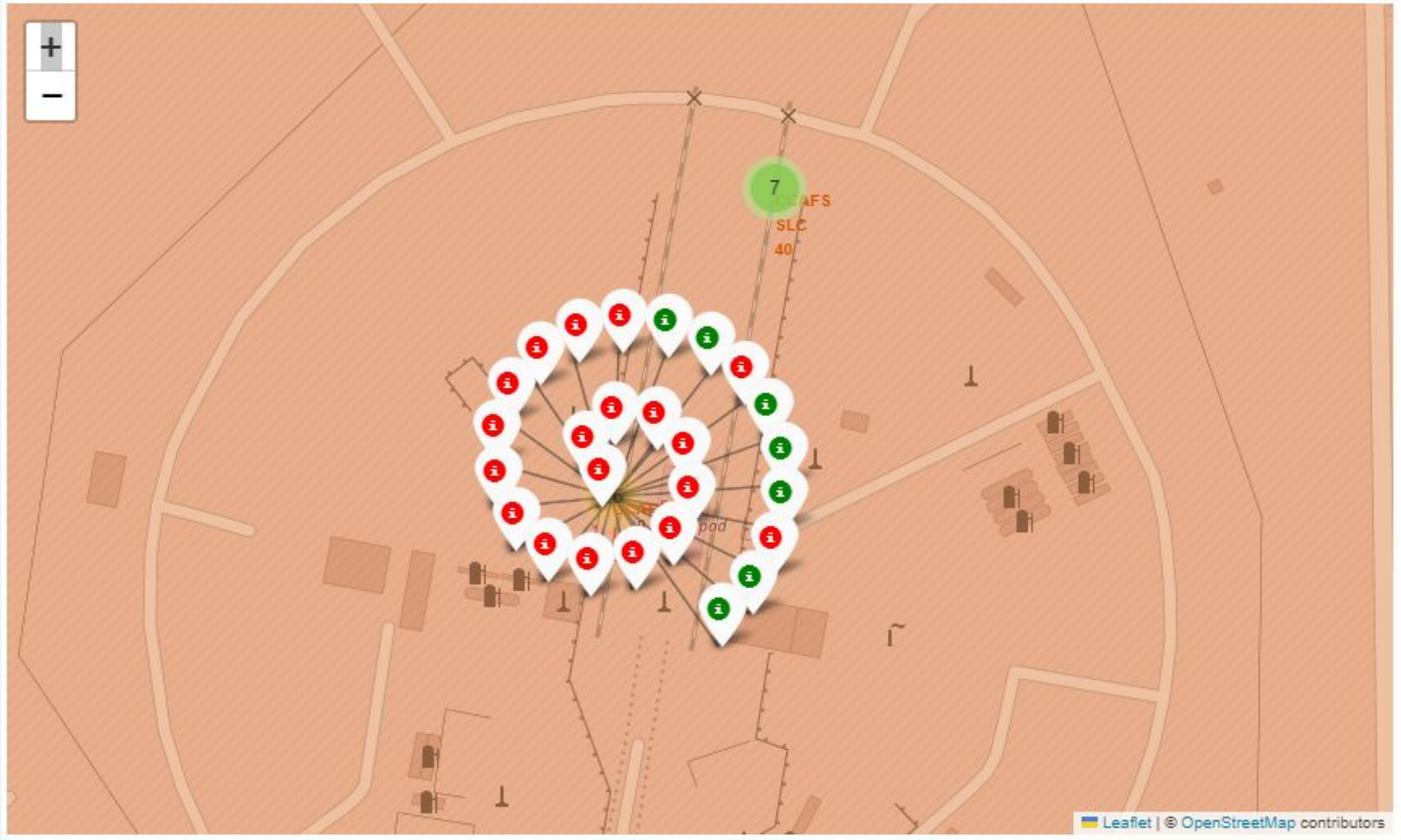
Map of all Launch Sites

- On the map to the right, there are two red spots representing the launch sites.
- The dot in Florida contains 3 launch sites which can be distinguished by zooming in closer, shown in the top right.
    - Note that two launch site circles overlap slightly.



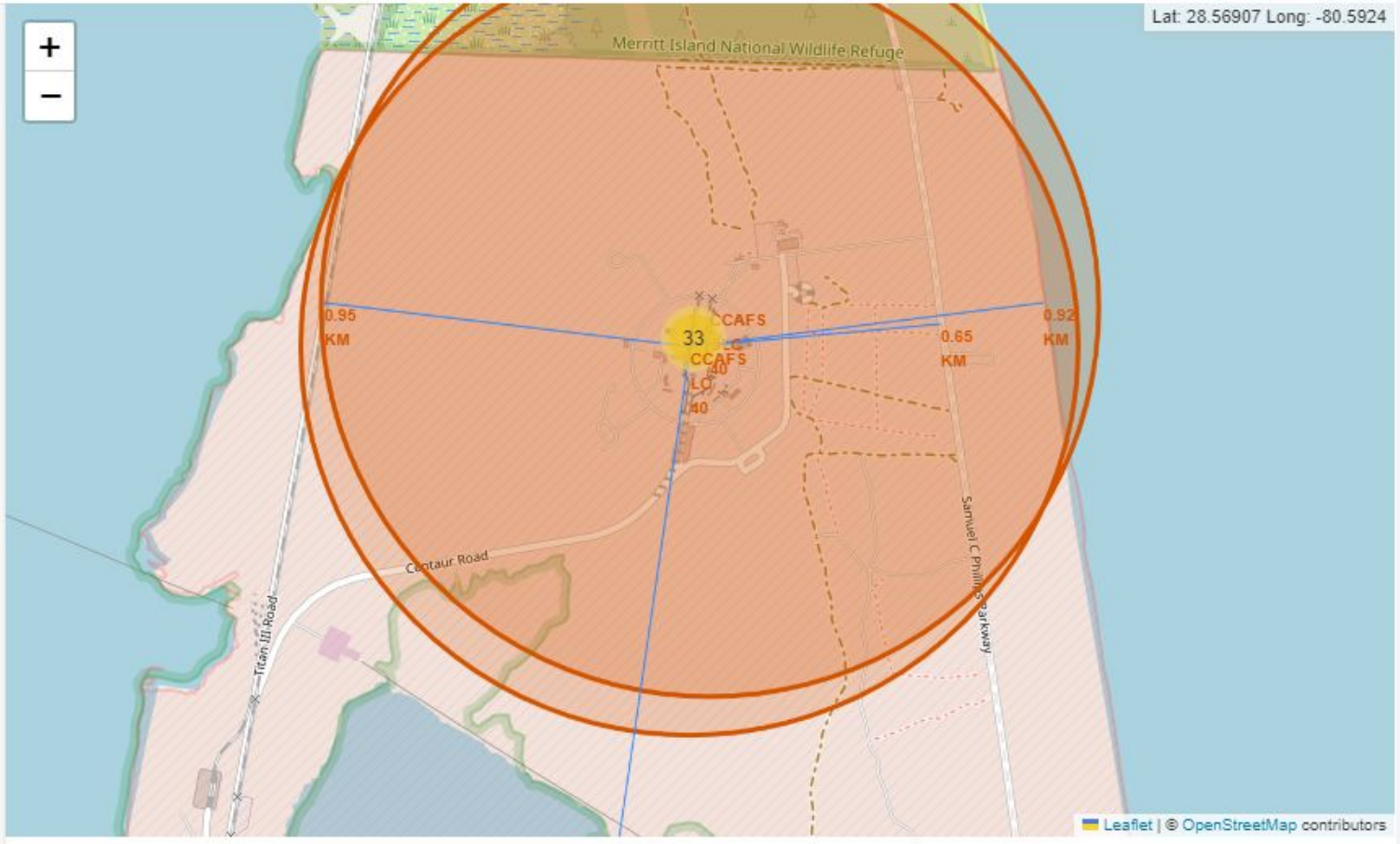
Marker Cluster of Launch Site Outcomes

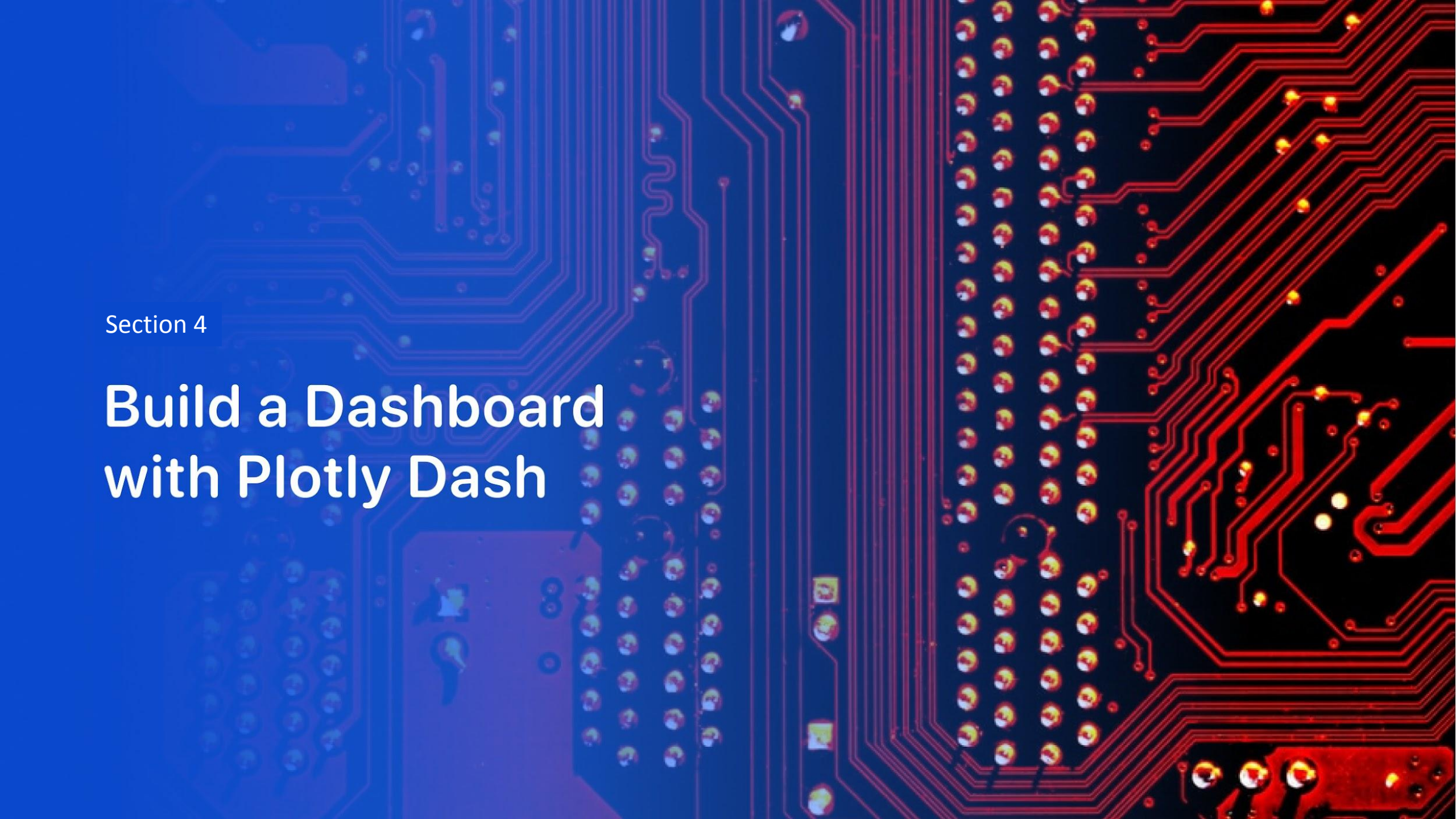
- Each launch site contains a cluster of markers representing failed and successful landings.
  - Green for successful.
  - Red for failed.
- Shown slightly above the cluster, in the map to the right, is a marker for another launch site with a maker indicating the number of missions performed at site.



Launch Site Map With Nearby Features

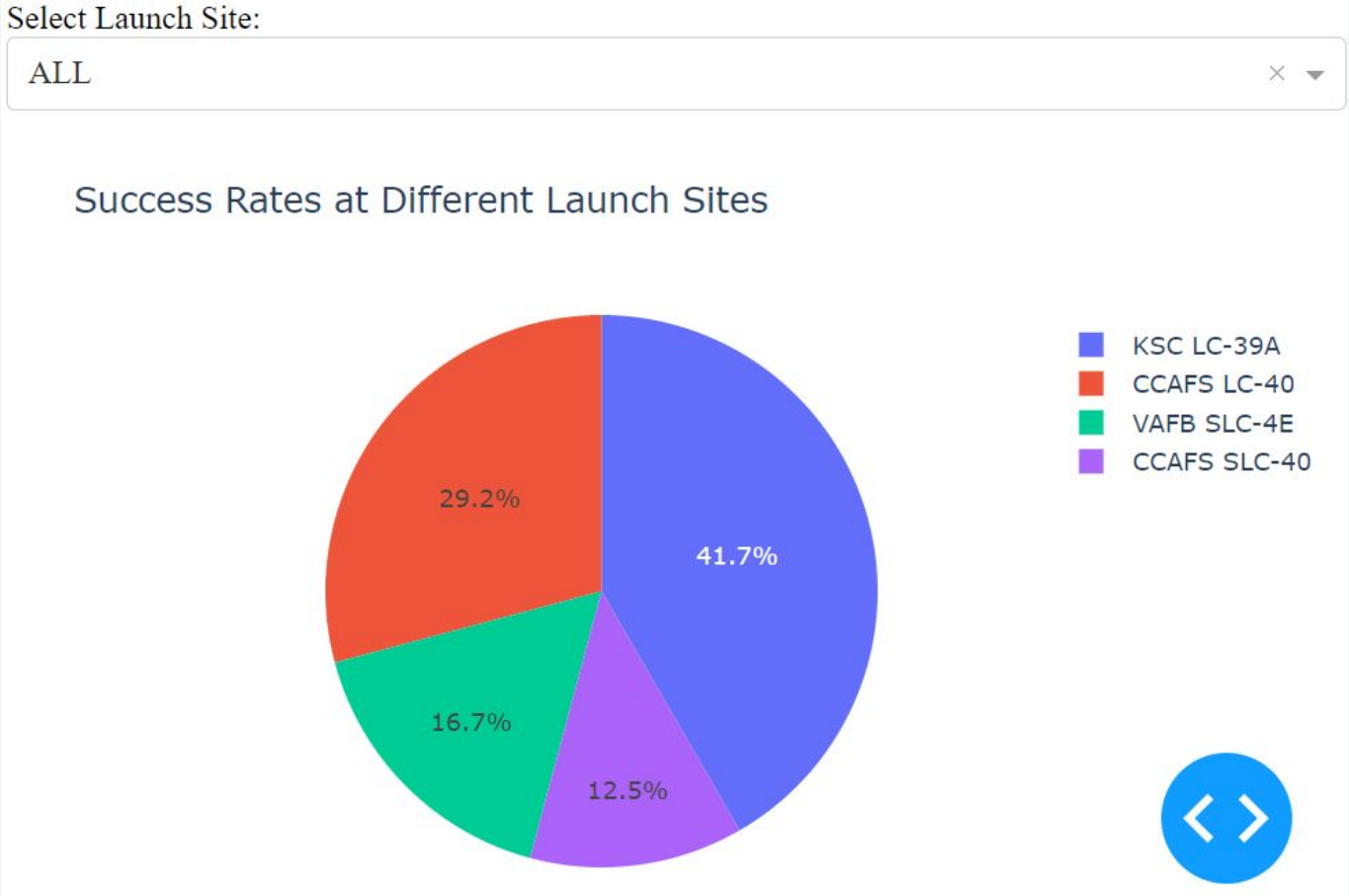
- The launch site CCAFS is close to numerous features.
  - Such as railways, highways, coastline, and city.
- The blue lines show direction of the features, marked with the distance from each feature to the launch site.
- Each Feature in detail
  - Highway - 0.65km
  - Coast line - 0.92km
  - Railway - 0.95km
  - City - 18.66km
- As the markers distance is show by the feature the city distance is too far to be captured at this zoom.




Section 4
Build a Dashboard with Plotly Dash

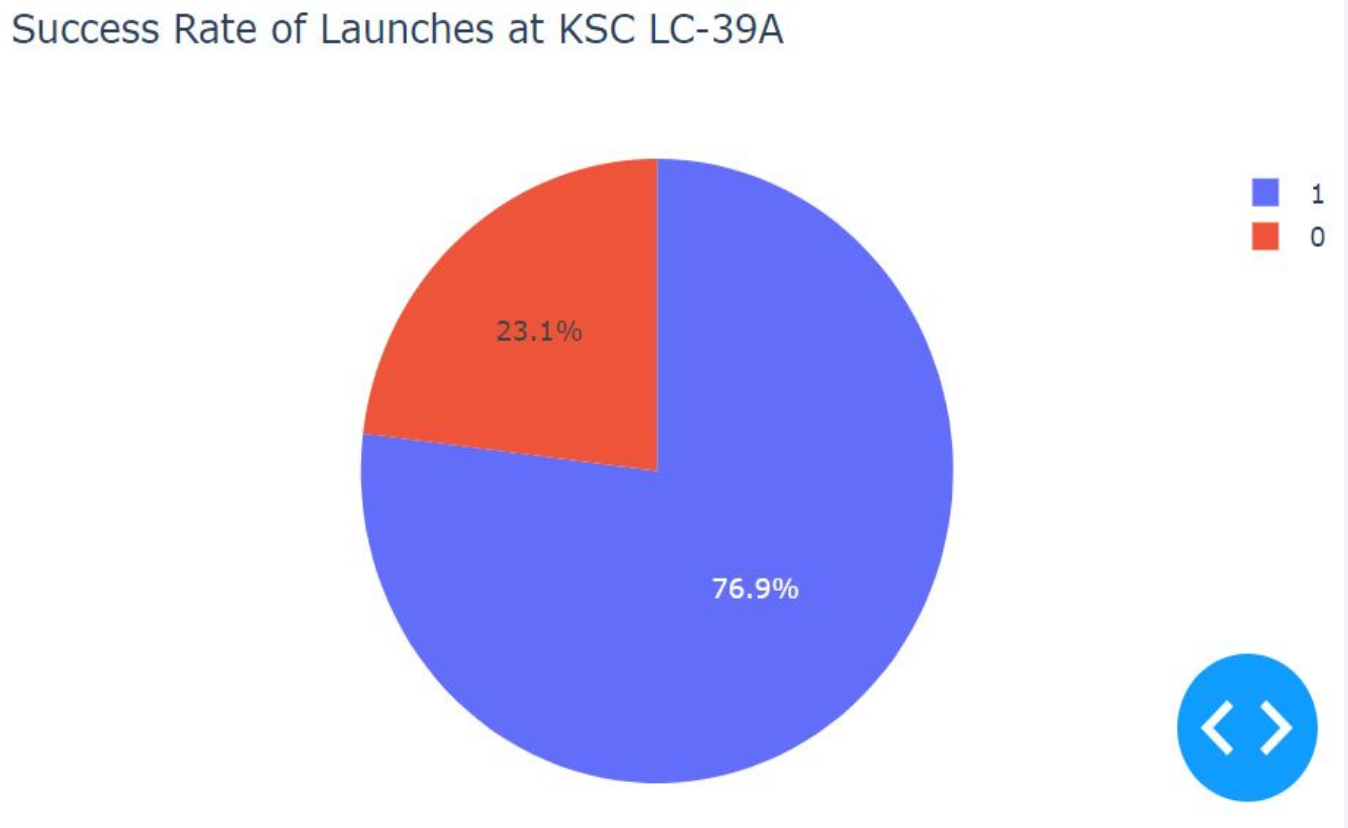
Pie Chart of the Success Rates at Different Launch Sites

- The KSC LC-39A launch site has had the most successful landings at 41.7%.
- The CCAFS SLC-40 has the lowest amount of successes, at only 12.5% of total successful landings.



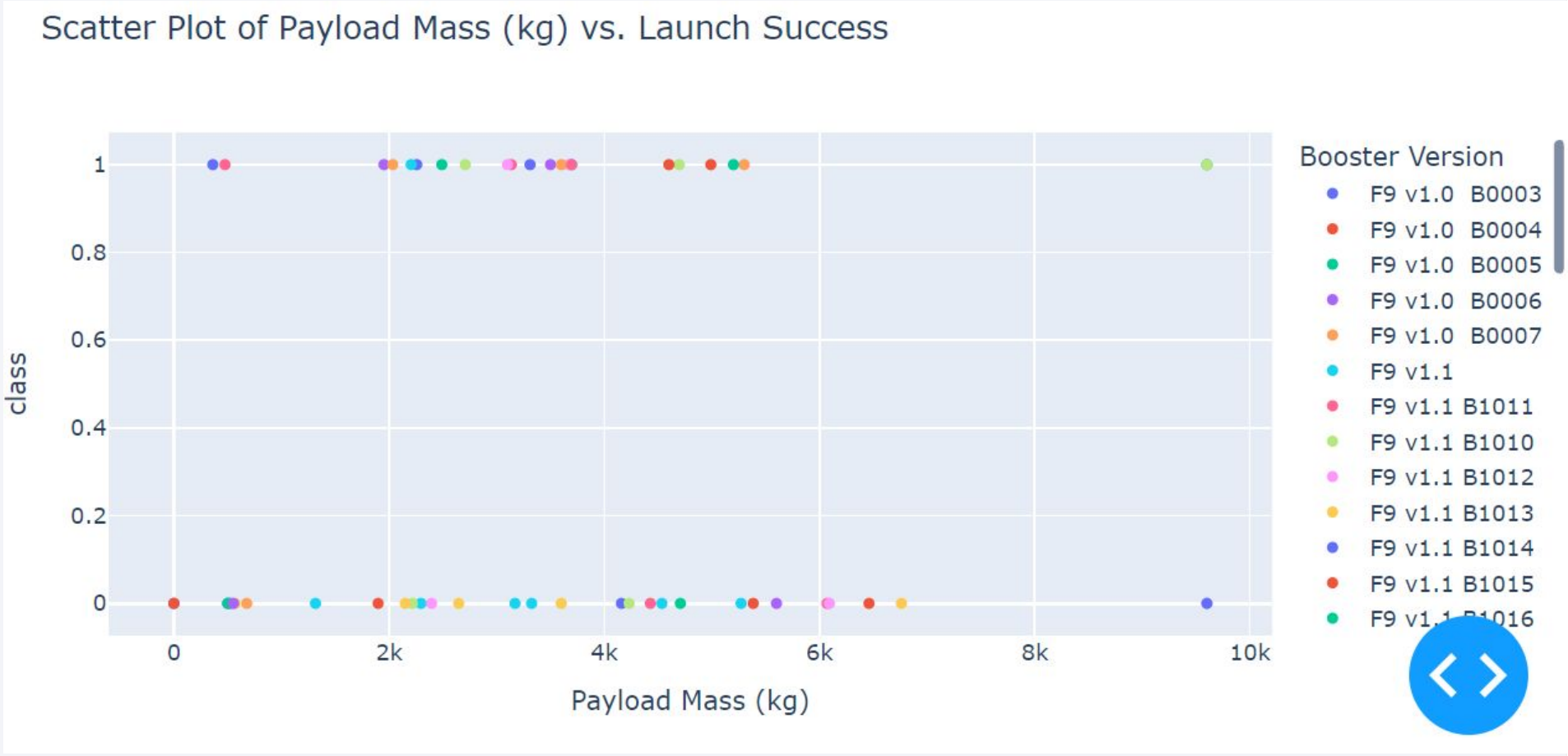
The Launch Site with the Highest Sucess Ratio

- The chart shows the launch site with the highest success ratio at a 76.9% success rate.
- This high success ratios corresponds to the previous slide in which this site has the most overall successes.



Payload vs. Launch Outcome from 500kg to 9000kg Payloads

- The payload mass is listed on the x-axis against success of launch landing from 0 (failure) to success (1) on the y-axis.
- Each booster version is color coded for identification.
- The highest success rate was between 2000kg to 4000kg
- Success seems to depend more on payload than booster version.



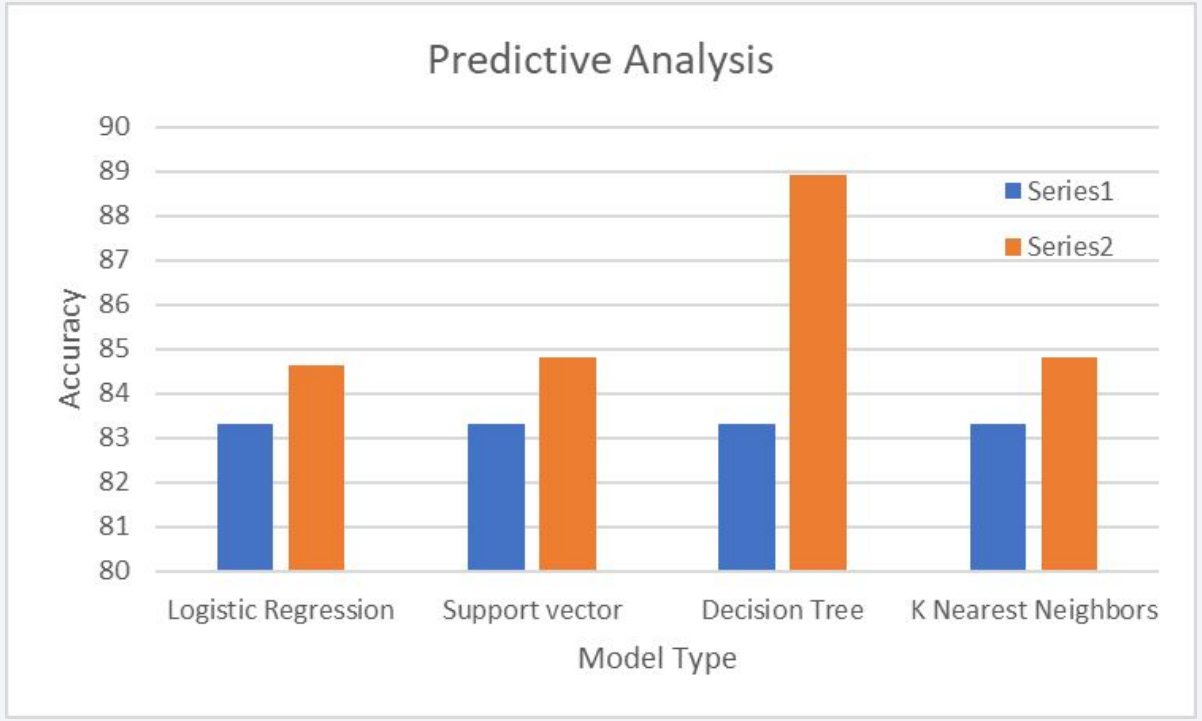
Section 5

Predictive Analysis (Classification)

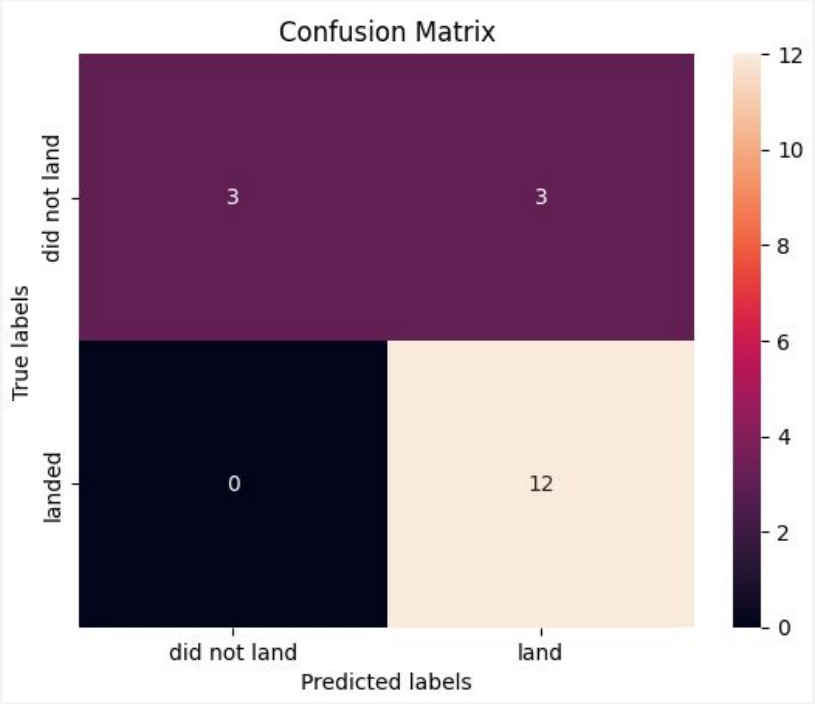
Classification Accuracy

- Four classification models were tested, the accuracies were plotted next to each other.

- The decision tree model has the highest accuracy.



Confusion Matrix Of the Decision Tree



- To the right is the confusion matrix for the Decision Tree Classifier.

Conclusions

- The optimal launch site is KSC LC-39A.
- Launches with payloads over 8,000kg have a higher success rate.
- Generally, mission outcomes are a success. However, this does not influence landing success rates.
- Using a decision tree classifier is the best way to predict landing success conditions.

Appendix

- GitHub Repository: https://github.com/WenzI-C/Data-Science-Capstone

Thank you!