IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Katie
3/11/2022

Outline

EXECUTIVE SUMMARY

INTRODUCTION

METHODOLOGY

RESULTS

CONCLUSION

APPENDIX

Executive Summary

- The purpose of this project is to determine if the Falcon 9 first stage will successfully land. This project had key stages to predict if this first stage will land which included:
  - Data Collection
  - Data Wrangling
  - Exploratory Data Analysis (EDA)
  - Interactive Visualizations with Folium and Plotly Dash
  - Predictive Analysis

Introduction

- Space Y is a new rocket company with competitors that include Space X. As a new company in an highly competitive market, it is necessary to find as much information as possible on the competition.
- By determining the cost of the first launch, our company can successfully bid against Space X for a rocket launch.
- This report will seek to answer the following questions:
    - What is the price of each launch for Space X?
    - What other critical information is there on Space X?
    - Will Space X reuse their first stage for Falcon 9?

Section 1

Methodology

Methodology

- Executive Summary
- Data collection methodology:
  - SpaceX launch data was gathered from the SpaceX Rest API
- Perform data wrangling
  - Performed EDA and determined which training labels to use for supervised learning.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Used classification models to predict if the first stage will land.

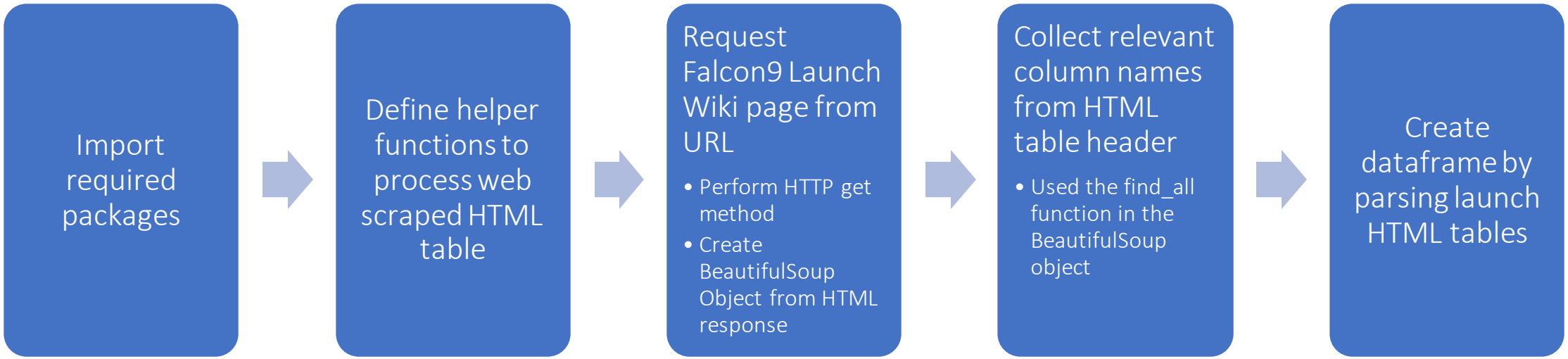
Data Collection

- The SpaceX launch data was gathered using the **SpaceX Rest API.** This API provides information about:
  - Launches and rocket information
  - Payload Delivered
  - Launch Specifications
  - Landing Specifications
  - Landing Outcome
- Data is obtained from this API using a get request
- JSON objects are converted to data frames and normalized into a flat table

Data Collection – SpaceX API

Github Link

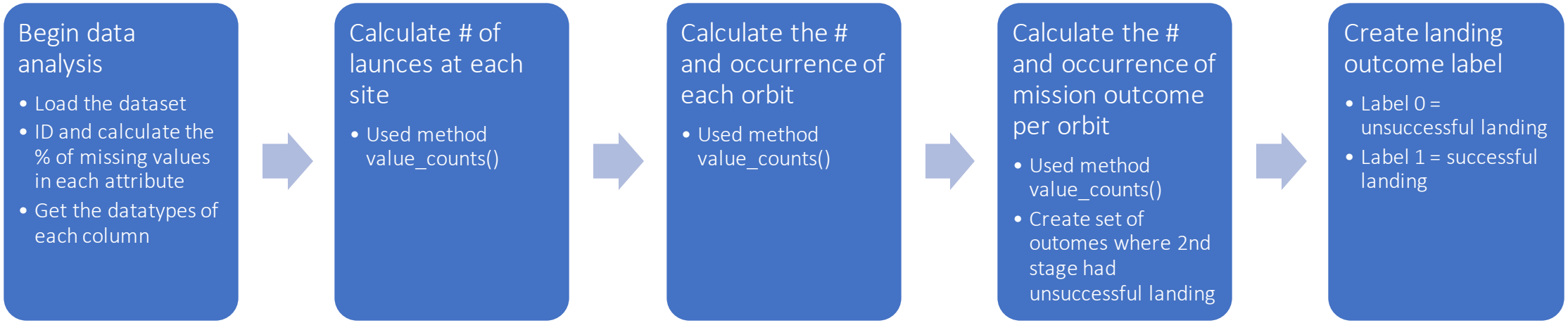
Data Collection – Scraping



Github Link

Data Wrangling

- Exploratory Data Analysis (EDA) was performed to find patterns in the data
- The dataset contained various mission outcomes which were converted into training labels
- Github Link



EDA with Data Visualization

- Summary of charts plotted to better understand the dataset:
  - Scatter point chart to visualize the relationship between flight number and launch site.
  - Scatter point chart to visualize the relationship between payload mass and the launch site.
  - Bar chart to visualize the relationship between the success rate of each orbit type.
  - Scatter point chart to visualize the relationship between flight number and orbit type
  - Scatter point chart to visualize the relationship between payload mass and orbit type.
  - Line chart to visualize the yearly launch success trend
- Github Link

EDA with SQL

- Summary of SQL queries performed to better understand the dataset used.
  - Selected the distinct launch sites from table to display the names of the unique launch sites in the mission.
  - Selected launch site from table where launch sites begins with "CCA". The list of results was limited to 5.
  - Selected the sum of payload mass from the table and grouped by payload
  - Selected average payload mass and booster version from the table and grouped by booster version
  - Selected the minimum date and mission outcome from the table and grouped by mission outcome. This showed the date when the 1st successful landing outcome in ground pad was achieved.
  - Selected booster version from table where mission outcome is "Success (drone ship)" and payload mass is between 4000 and 6000
- Github Link

EDA with SQL Continued

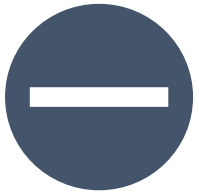
- (Continued) Summary of SQL queries performed to better understand the dataset used.
  - Selected the count of mission outcome and mission outcomes from the table and grouped by mission outcome to get the total number of successful and failure mission outcomes
  - Selected booster version from table where payload mass = the max payload mass from the table. This query uses a subquery.
  - Selected the launch site, booster version, and landing outcome from the table where the landing outcome = "Failure (Drone Ship)" and the Year is 2015
  - Selected the landing outcome from the table where the date is between 2010-06-04 and 2017-03-20. This list was then ordered by a descending date.

Build an Interactive Map with Folium

Marked all launch sites on the map

Added each launch sites location on the map using the sites latitude and longitude coordinates.

Visualized these latitude and longitude coordinates with a circle, to add a highlighted circle area, and with a marker to add a text label on a specific coordinate.



Marked the successful/failed launches for each site

Added markers for all launch records, where a successful launch (class =1) was marked with a green marker and an unsuccessful launch (class=0) was marked with a red marker.

Because many launch records all occur at the same coordinate, used a MarkerCluster()

Calculated the distances between launch sites and its proximities

Added a MousePosition to get the coordinates for a mouse over any point on the map

Added a Marker and drew a PolyLine to show the distance between the closest coastline point and the launch site.

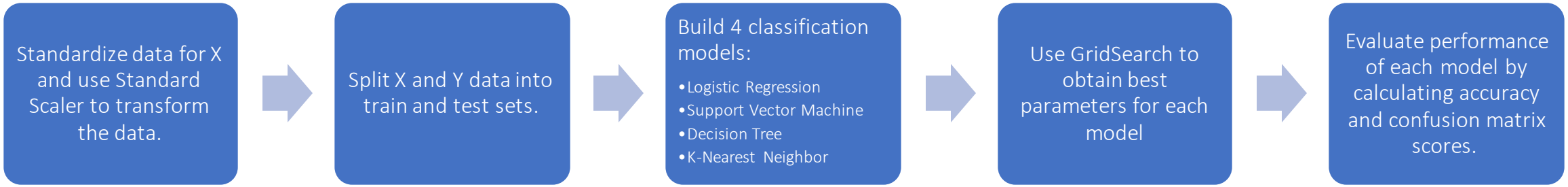
Github Link

Build a Dashboard with Plotly Dash

- The dashboard was built in Plotly Dash and used to find insights from the Space X data.
- Summary of dashboard features:
  - Added a dropdown list for launch site selection. This allows plots to be filtered per site.
  - Added a pie chart to show the total count of successful launches for all sites. When a specific launch site is selected, it will show the successful vs. failed counts for that particular site.
  - Added a slider to select payload range (kg).
  - Added a scatter chart which shows the correlation between payload mass (kg) and launch success.
- Github link

Predictive Analysis (Classification)

- To predict whether the first stage would land or not, the data was standardized and split into training and test sets. Then, 4 different classification models were built and evaluated. The best hyperparameters would found for each model and the accuracy and confusion matrix scores were calculated and compared.
- Github link



Results

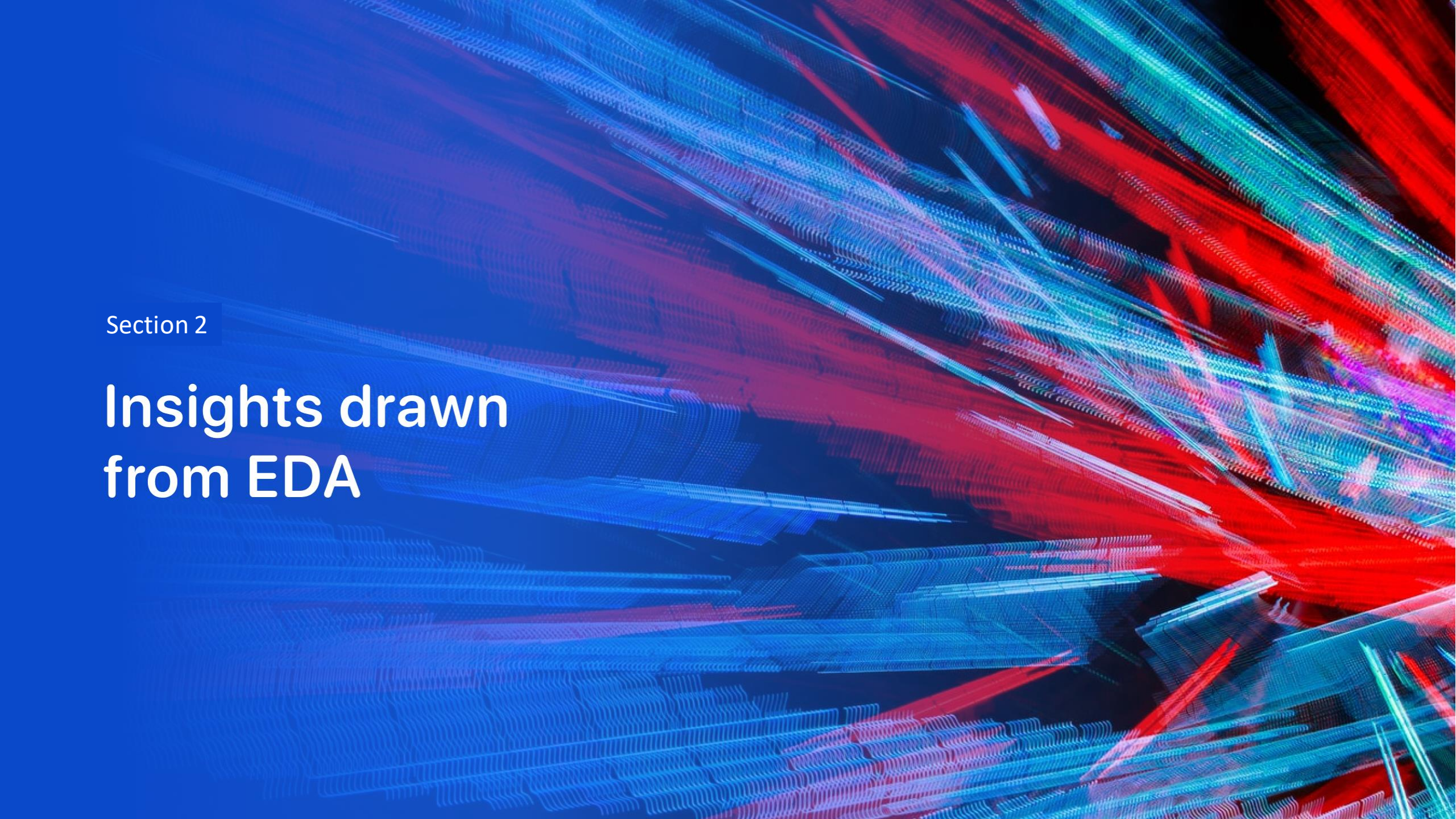
EXPLORATORY DATA ANALYSIS RESULTS

INTERACTIVE ANALYTICS DEMO IN SCREENSHOTS

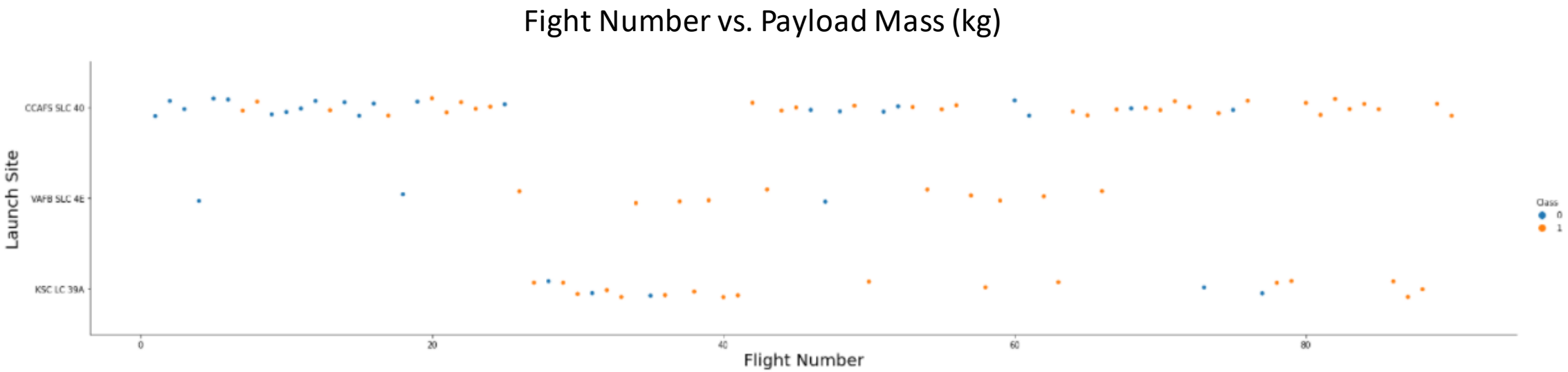
PREDICTIVE ANALYSIS RESULTS

Section 2

Insights drawn from EDA

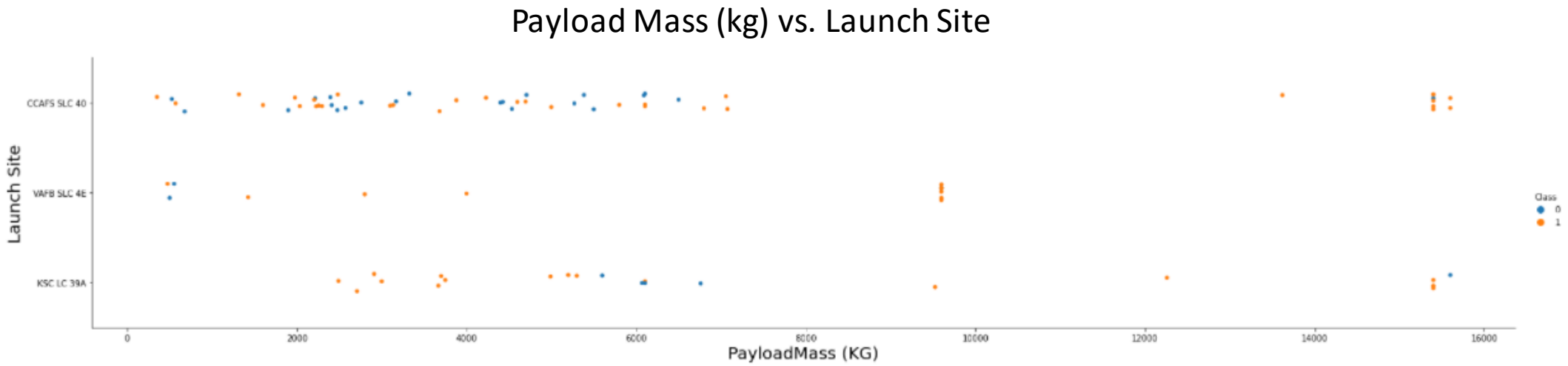


Flight Number vs. Launch Site



- This plot shows the flight number vs launch site.
- For site CCAFS SLC 40, observe no flights between approximately flight number 20 and flight number 40
- For site VAFB SLC 4E, all class 1 flights are between approximately flight number 20 and 70
- For site KSC LC 39A, no flights before approximately flight number 30

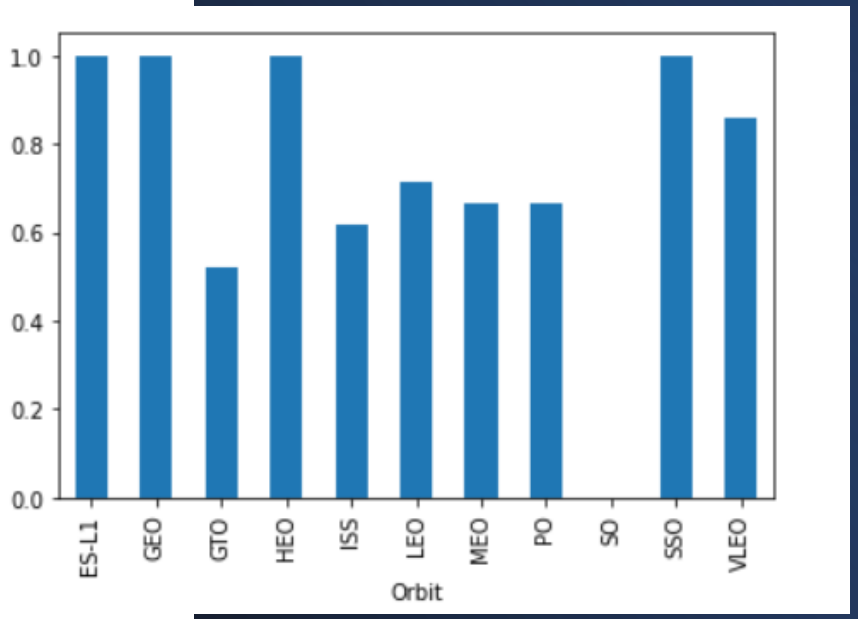
Payload vs. Launch Site



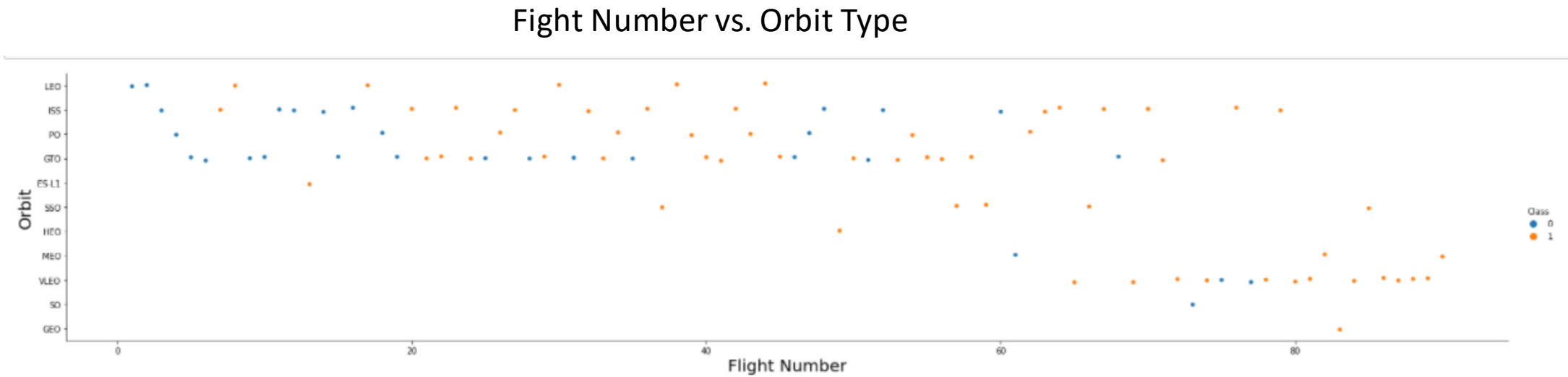
- This plot shows the payload mass (kg) vs launch site.
- Results show most payload mass for all 3 launch sites fell between 0 and 8000 kg.
- For site VAFB LC, there are no rockets launched with a mass greater than 10,000 kg.

Success Rate vs. Orbit Type

- This plot visualizes the relationship between success rate of each orbit type.
- Orbits with high success rates include: ES_L1, GEO, HEO, and SSO
- SO has a success rate of 0

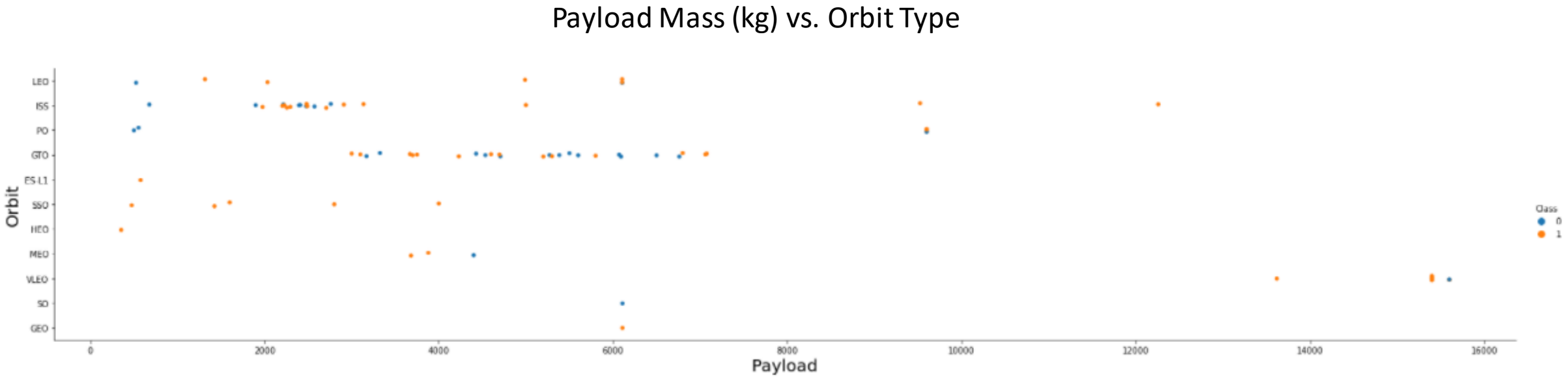


Flight Number vs. Orbit Type



- This plot shows the flight number vs. Orbit type.
- For LEO orbit, the success appears to be related to the number of flights.
- For GTO orbit, there appears to be no relationship

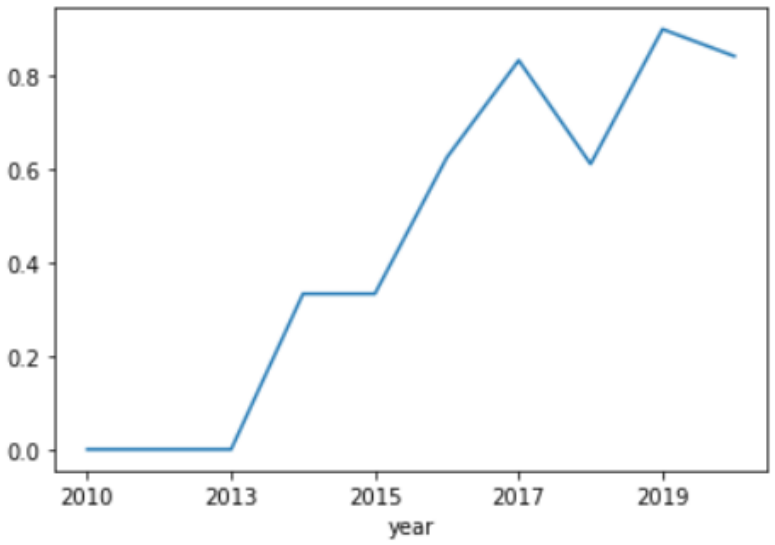
Payload vs. Orbit Type



- This plot shows payload mass (kg) vs. Orbit type.
- Observe that there are more successful landings for Polar, LEO, and ISS with heavy payloads.
- GTO does not show an obvious relationship.

Launch Success Yearly Trend

- This plot visualizes the launch success yearly trend.
- The success rate since 2013 is consistently increasing until 2020.



All Launch Site Names

- This query displays the names of the unique launch sites in the space mission.
- The output shows 4 unique sites including: CCAFS LC-40, CCAFS SLC-40, KSC LC-39A, AND VAFB SLC-4E

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

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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

Launch Site Names Begin with 'CCA'

- This query displays 5 records where launch site begins with the string 'CCA'.
- The output above shows all columns for each of the 5 records.

Total Payload Mass

- This query shows the total payload mass (kg) carried by boosters launched by NASA (CRS).
- Customer and sum of payload mass (kg) were selected from the table. The name of the sum of payload mass was changed to paysum, as shown below.
- The query looked for where customer was like "NASA (CRS)" and grouped by customer.

customer	paysum
NASA (CRS)	45596
NASA (CRS), Kacific 1	2617

Average Payload Mass by F9 v1.1

- This query shows the average payload mass (kg) carried by booster version F9v1.1
- Selected average payload mass (kg) and booster version from the table.
- The query looked for where booster version was like "F9 v1.1" and grouped by booster version.

1	booster_version
2928	F9 v1.1
500	F9 v1.1 B1003
2216	F9 v1.1 B1010
4428	F9 v1.1 B1011
2395	F9 v1.1 B1012
570	F9 v1.1 B1013
4159	F9 v1.1 B1014
1898	F9 v1.1 B1015
4707	F9 v1.1 B1016
553	F9 v1.1 B1017
1952	F9 v1.1 B1018

First Successful Ground Landing Date

- This query lists the date when the first successful landing outcome in ground pad was achieved.
- From the table, the minimum date and mission outcome were selected and grouped by mission outcome.

1	mission_outcome
2015-06-28	Failure (in flight)
2010-06-04	Success
2018-01-08	Success (payload status unclear)

Successful Drone Ship Landing with Payload between 4000 and 6000

- This query lists the names of the boosters which have success in drone ship and have payload mass (kg) between 4000 and 6000.
- Selected booster version from the table and looked for where landing outcome was "Success (drone ship)" and payload mass (kg) was between 4000 and 6000.

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

Total Number of Successful and Failure Mission Outcomes

- This query lists the total number of success and failure mission outcomes.
- Selected count of mission outcome, which is displayed as count in the chart below, and mission outcome from the table. It was then grouped by mission outcome.

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

Boosters Carried Maximum Payload

- This query lists the names of the booster versions which have carried the maximum payload mass.
- Selected booster version from the table and looked for where payload mass was equal to the maximum payload mass from the table.
    - This query used a subquery to find the maximum payload mass.

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

2015 Launch Records

- This query lists the failed landing outcomes in drone ship for the 2015.
- From the table, selected launch site, booster version and landing outcome.
- Looked for where the landing outcome was "Failure (drone ship)" and the year was equal to 2015.

launch_site	booster_version	landing__outcome
CCAFS LC-40	F9 v1.1 B1012	Failure (drone ship)
CCAFS LC-40	F9 v1.1 B1015	Failure (drone ship)

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

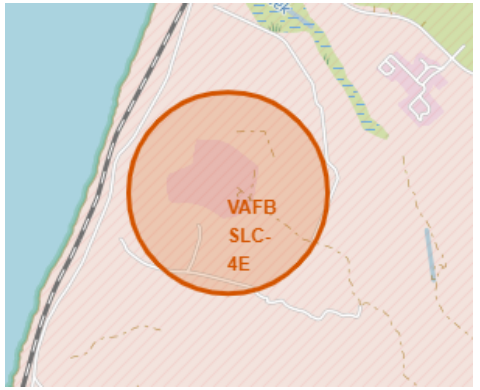
- This query ranks the count of landing outcomes between the dates 2010-06-04 and 2017-03-20.
- Selected landing outcome and count of landing outcome from the table.
- Looked for where the date was between 2010-06-04 and 2017-03-20. Then, grouped by landing outcome and ordered by descending count of landing outcome.

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


Section 3
Launch Sites
Proximities Analysis

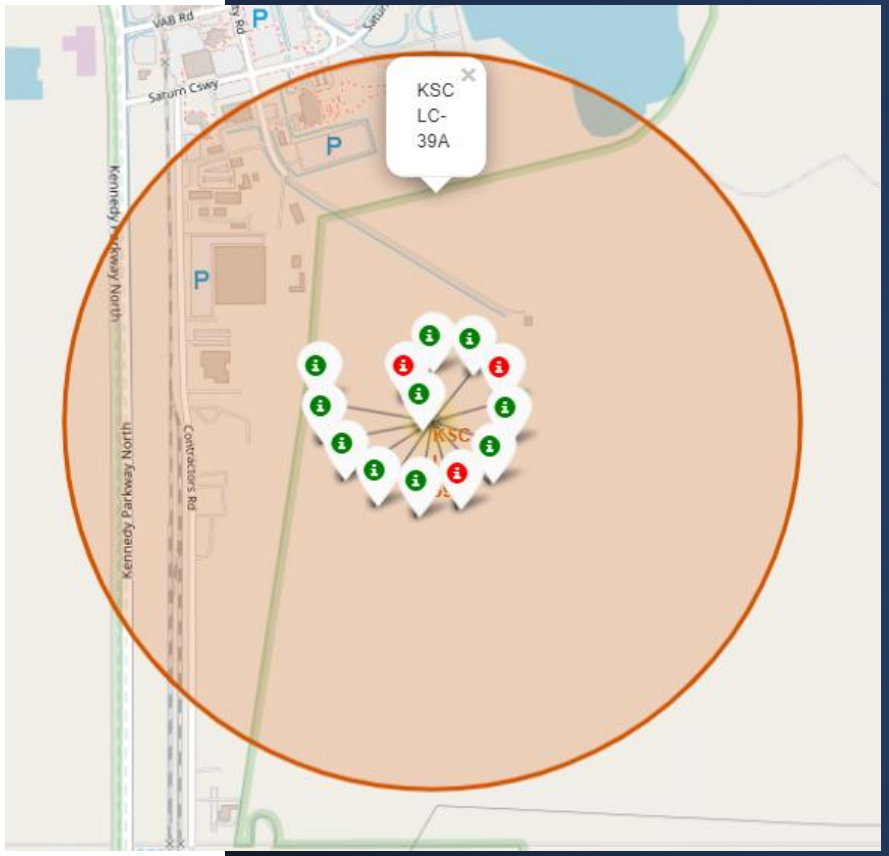
Site Locations on Global Map

- The goal of this task was to map all site locations on the map using the sites latitude and longitude coordinates.
- The bottom left image shows the map will all site locations.
- The smaller, bottom right image shows an example of the circle and marker that was created for each site.



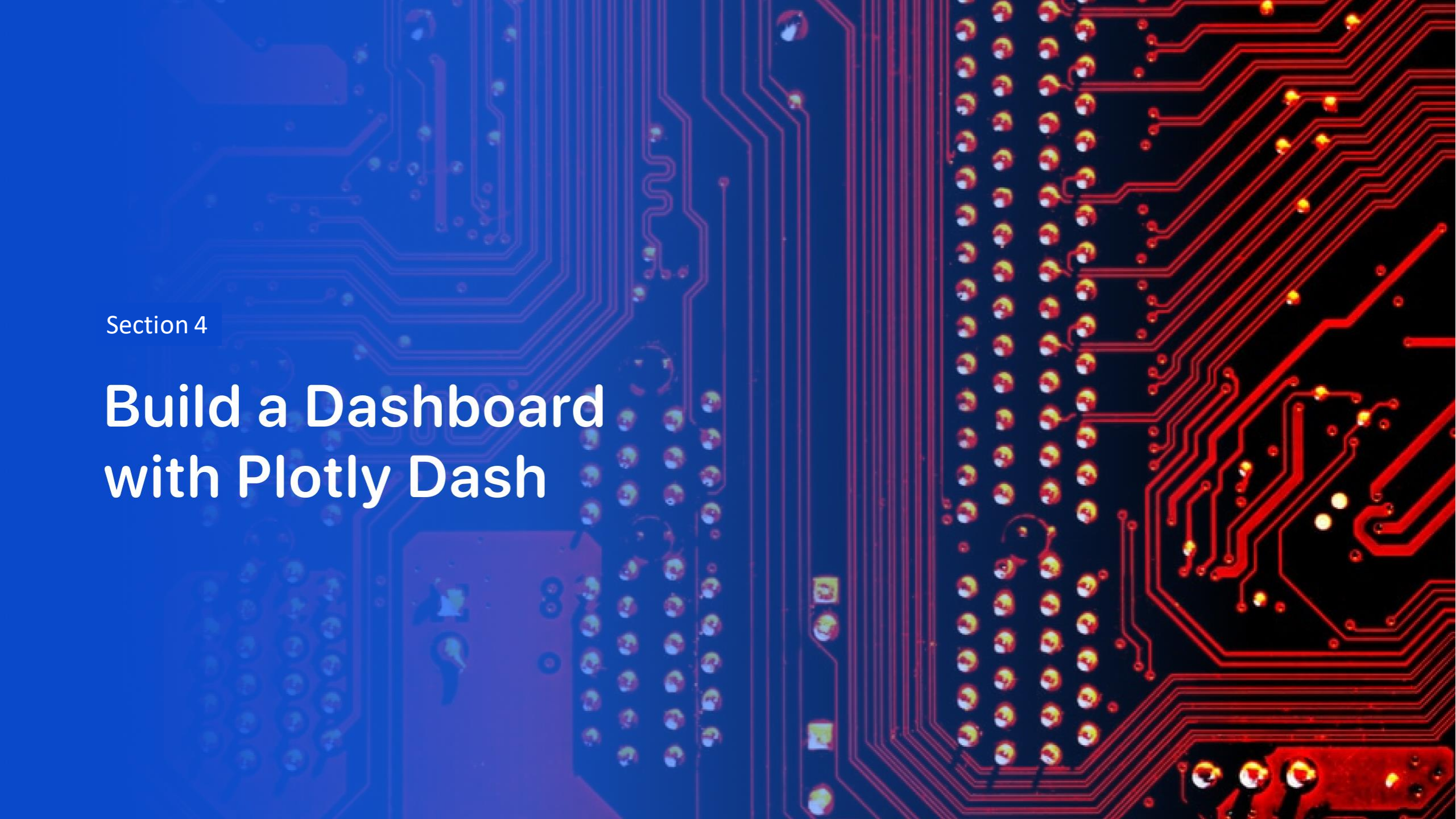
Successful/Failed Launches per Site

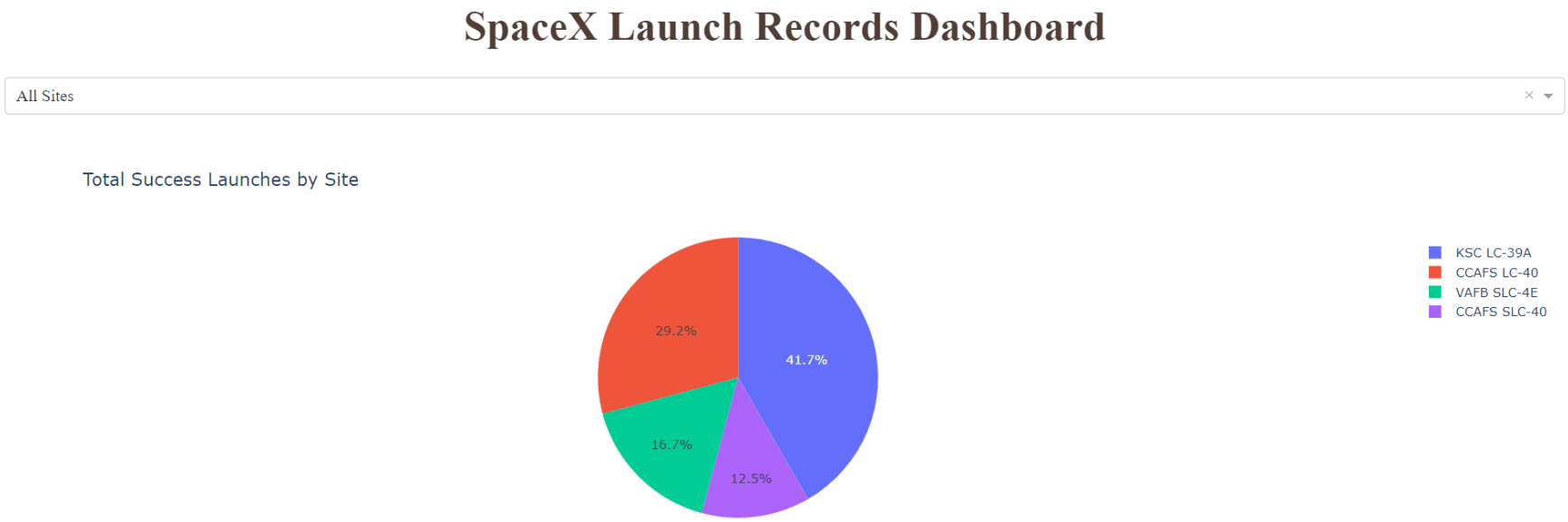
- This task marks successful and failed launches per site. Successful launches are shown with a green marker and unsuccessful launches are shown with a red marker.
- Because several launches happen at the same location, a MarkerCluster object was created to complete this task.
- The image shows an example of what was added for each site location.



- This task shows lines between launch sites and its proximities, such as railways, highways, and coastlines.
- The example shows the distance between a site and a railway.

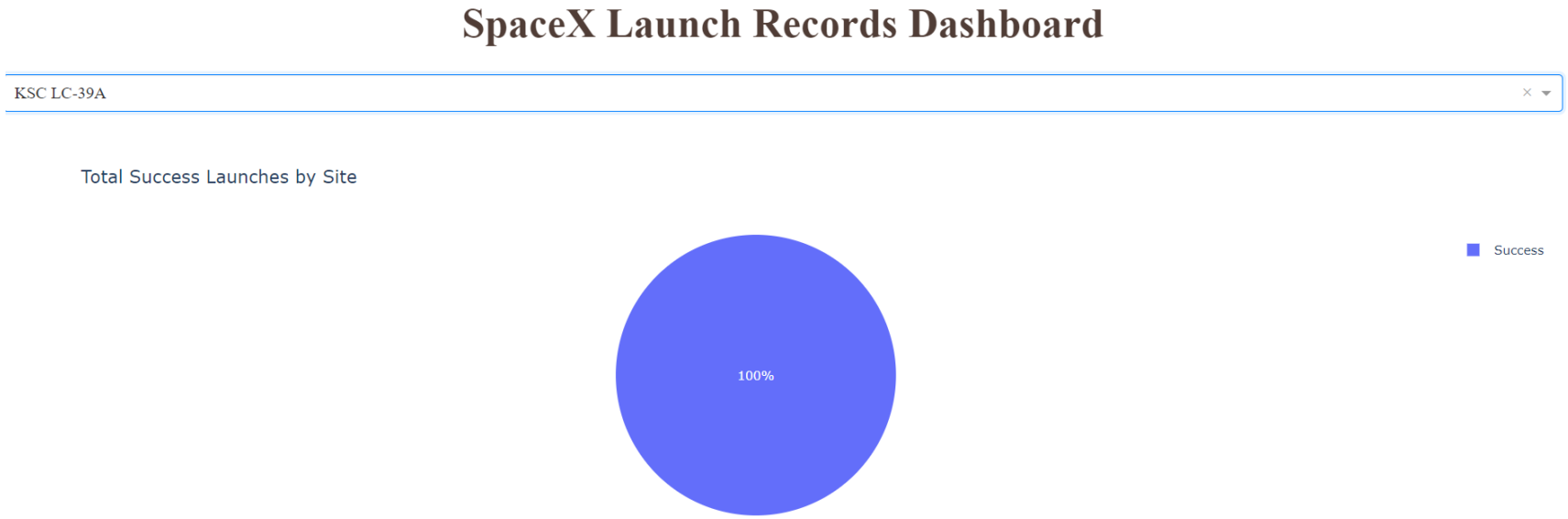
Launch Sites Proximities


Section 4
Build a Dashboard with Plotly Dash



- This screenshot shows the total successful launches for all sites. The legend for the site colors is shown on the right. The dropdown is at the top, with the option to select a specific site.

Launch Success Count for all Sites



- This screenshot shows the site with the highest launch success ratio. According to this pie chart, site KSC LC-39A had 100% success rate.

Launch Site with Highest Launch Success

- This screenshot shows the scatter plot of payload mass (kg) vs. Launch outcome for all sites. The slider range was adjusted to show payload mass between 3k and 7k.

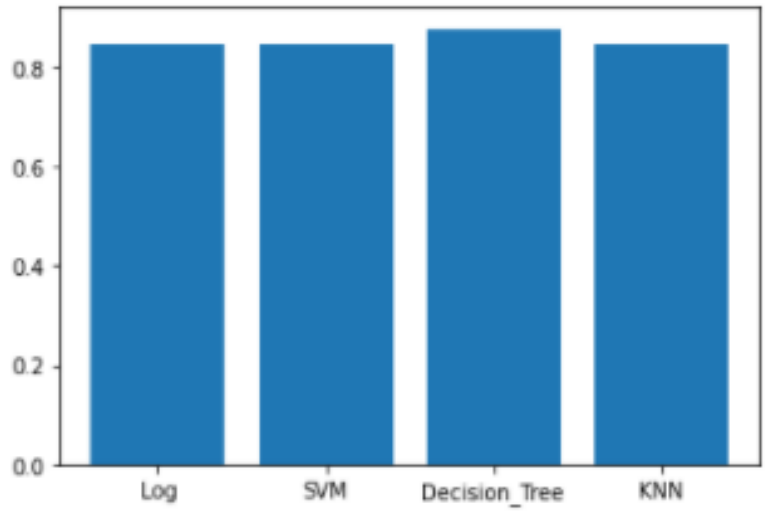
Payload vs. Launch Outcome

Section 5

Predictive Analysis (Classification)

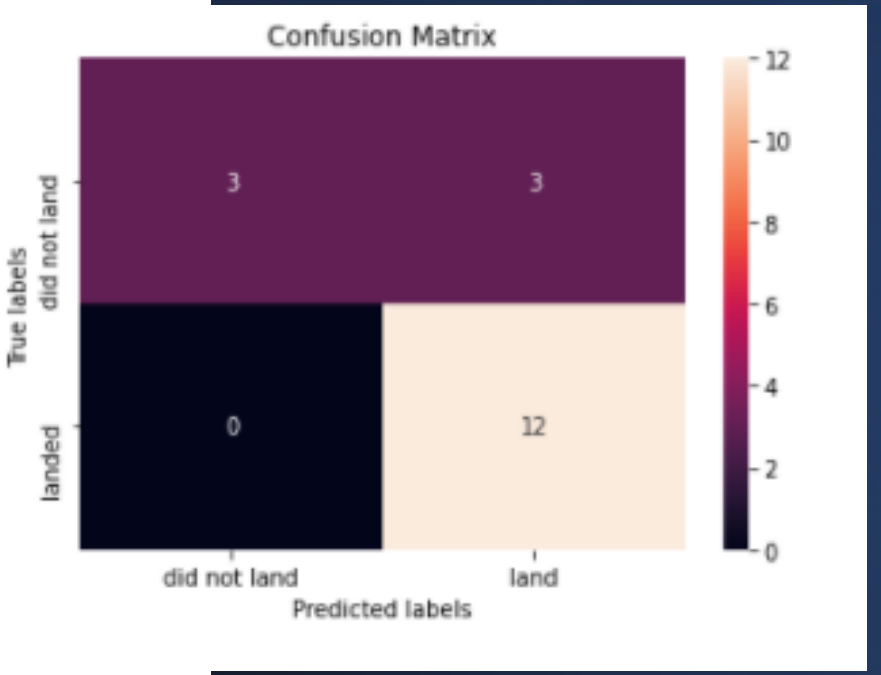
Classification Accuracy

- For all 4 models, the accuracy and confusion matrices were calculated.
- The highest accuracy was 89% and was the best model using the decision tree classifier. When calculating accuracy using the method score, all models achieved an accuracy of 83%.



Confusion Matrix

- Depicted is the confusion matrix for the decision tree model.
- A confusion matric shows true positives, true negatives, false positives, and false negatives.
- Given the results on the right, we see:
  - True Positive: 3 Values
  - False Positive: 3 Values
  - False Negative: 0 Values
  - True Negative: 12 Values



Conclusions

- After evaluating the 4 classifier models, the best accuracy results were obtained from the Decision Tree model.
- According to the Plotly Dashboard, the site with the highest success rate was KSC LC-39A.
- Orbits with the highest success rates include: ES_L1, GEO, HEO, and SSO
- The yearly trend line chart shows the success rate has been increasing from 2013 to 2020.

Thank you!