Winning Space Race with Data Science

WY Chan

7 Nov 2023

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection with API and Web Scraping
  - Data Wrangling
  - Exploratory Data Analysis (EDA)
  - Data Visualization
  - Launch Sites Locations Analysis (Folium Map)
  - Interactive Dashboard (Ploty Dash)
  - Machine Learning Prediction
- Summary of all results
  - Insights drawn from EDA
  - Launch sites proximities analysis
  - Plotly Dash Dashboard
  - Predictive Analysis (Classification)

Introduction

SpaceX is a successful space company with relatively inexpensive rocket launches than the other competitors, much of the savings is because the first stage can be reused.

If an alternate company wants to bid against SpaceX for a rocket launch, what information is needed?

- Will the Falcon 9 first stage will land successfully?
- What factors are affecting the success of landing?
- How to increase the success rate of landing?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Space X API, Web Scraping
- Perform data wrangling
  - Convert outcomes into Training Labels
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Logistics Regression, Support Vector Machine (SVM), Decision Tree, K Nearest Neighbors (KNN)

Data Collection

- API - Space X API:

  https://api.spacexdata.com/v4/launchpads/

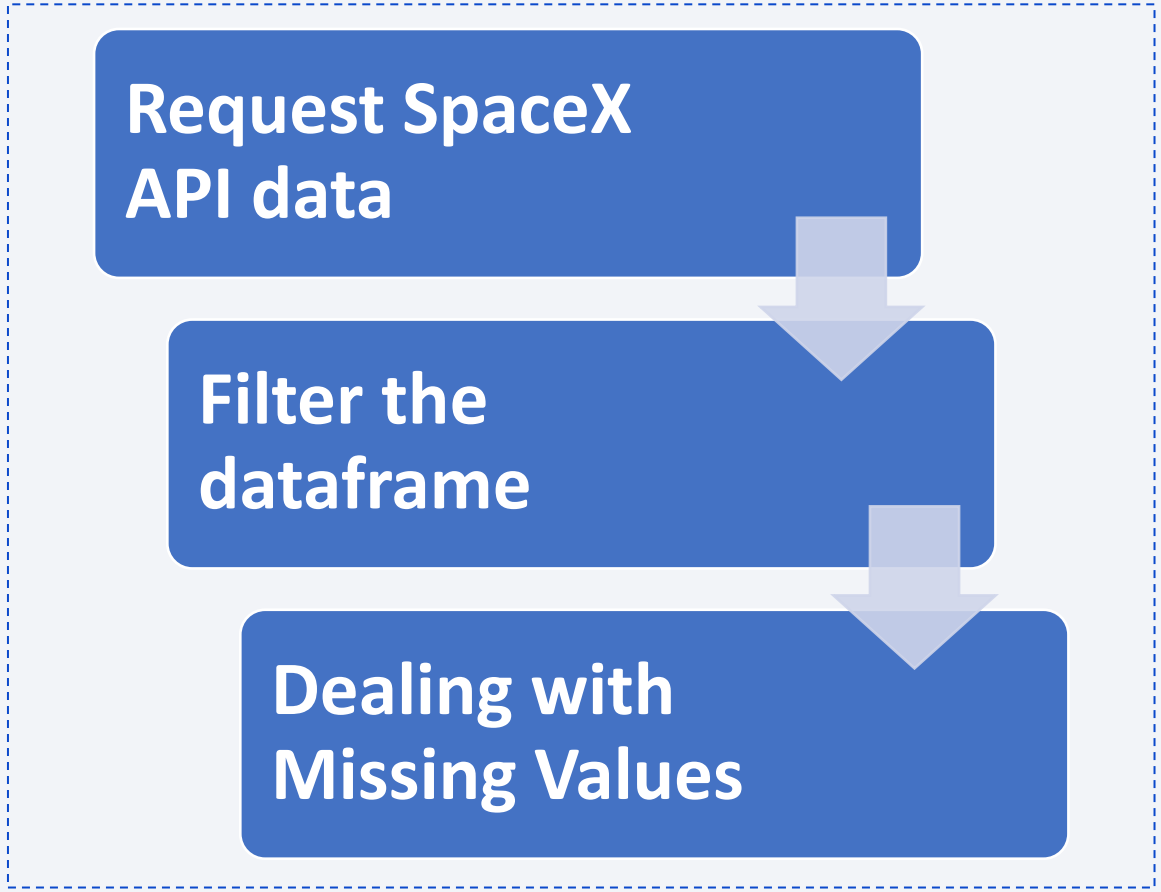
- Web Scraping - Wikipedia:

  https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922

Data Collection – SpaceX API

- Request rocket launch data using the GET request
- Clean the requested data
  - Only include Falcon 9 launches
  - Replace missing value with mean
- GitHub URL - SpaceX API calls notebook:

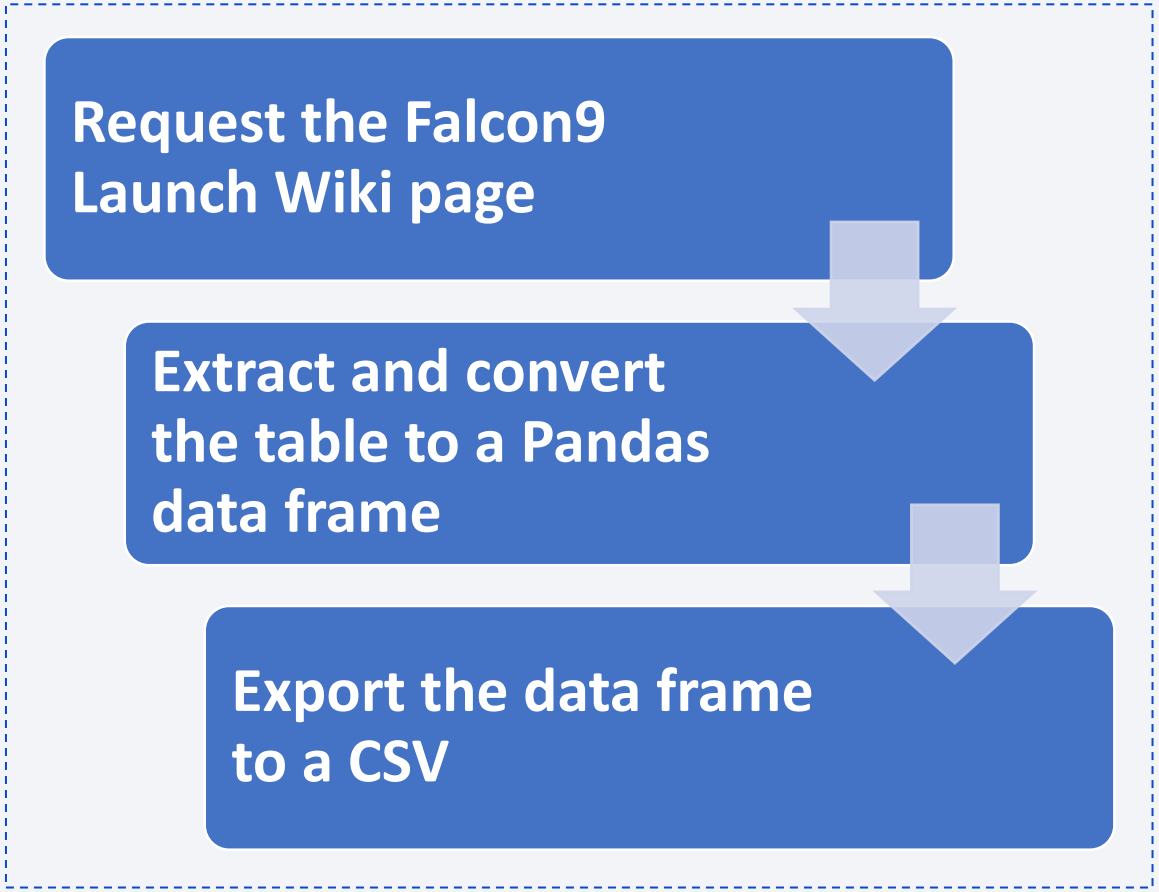
  https://github.com/wy-chan/Applied-Data-Science-Capstone/blob/main/Week1_1_Complete%20the%20Data%20Collection%20API%20Lab.ipynb



Data Collection - Scraping

- Request Falcon 9 launch record using the GET request
- Convert data format:
  - Extract Falcon 9 launch records HTML table
  - Parse the table and convert it into a Pandas data frame
- GitHub URL - Web scraping notebook:

https://github.com/wy-chan/Applied-Data-Science-Capstone/blob/main/Week1_2_Complete%20the%20Data%20Collection%20with%20Web%20Scraping%20lab.ipynb



Data Wrangling

- Perform some Exploratory Data Analysis (EDA) to find some patterns in the data and determine what would be the label for training supervised models.

- GitHub URL - Data wrangling notebook:
  https://github.com/wy-chan/Applied-Data-Science-Capstone/blob/main/Week1_3_Data%20Wrangling.ipynb

EDA with Data Visualization

- Charts plotted:
  - **Scatter plot** - visualize the frequency and distribution of data
  - **Bar chart** - compare the difference between categorized data
  - **Line chart** – visualize the trend of data

- GitHub URL - completed EDA with data visualization notebook:

  https://github.com/wy-chan/Applied-Data-Science-Capstone/blob/main/Week2_2_EDA%20with%20Visualization%20Lab.jupyterlite.ipynb

EDA with SQL

- SQL queries:
  1. Display the names of the unique launch sites in the space mission
  2. Display 5 records where launch sites begin with the string 'CCA'
  3. Display the total payload mass carried by boosters launched by NASA (CRS)
  4. Display average payload mass carried by booster version F9 v1.1
  5. List the date when the first succesful landing outcome in ground pad was acheived.
  6. List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  7. List the total number of successful and failure mission outcomes
  8. List the names of the booster_versions which have carried the maximum payload mass. Use a subquery
  9. List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
  10. 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.

- GitHub URL - completed EDA with SQL notebook:

  https://github.com/wy-chan/Applied-Data-Science-Capstone/blob/main/Week2_1_Complete%20the%20EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

- Map objects:
  - **Markers** - mark all launch sites on a map with names
  - **Circles** - point out launch sites location and success/fail indication
  - **Lines** - show distance between locations

- Github URL - Completed interactive map with Folium map:

  https://github.com/wy-chan/Applied-Data-Science-Capstone/blob/main/Week3_1_Launch%20Sites%20Locations%20Analysis%20with%20Folium.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

- **Bar Chart**:
  - Visualizing launch success counts
  - Launch site dropdown menu to show success rate of selected launch site

- **Scatter Plot**:
  - Visualizing correlation between payload and success
  - Launch site dropdown menu to show success launches of selected launch site
  - Interact with payload range to show success launches for specific range

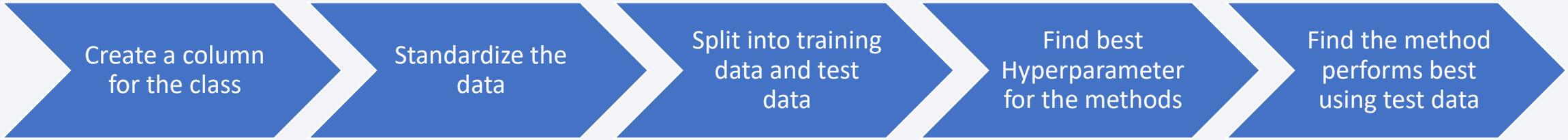
- GitHub URL - completed Plotly Dash lab:

  https://github.com/wy-chan/Applied-Data-Science-Capstone/blob/main/Week3_2_Build%20an%20Interactive%20Dashboard%20with%20Ploty%20Dash/spacex_dash_app.py

Predictive Analysis (Classification)

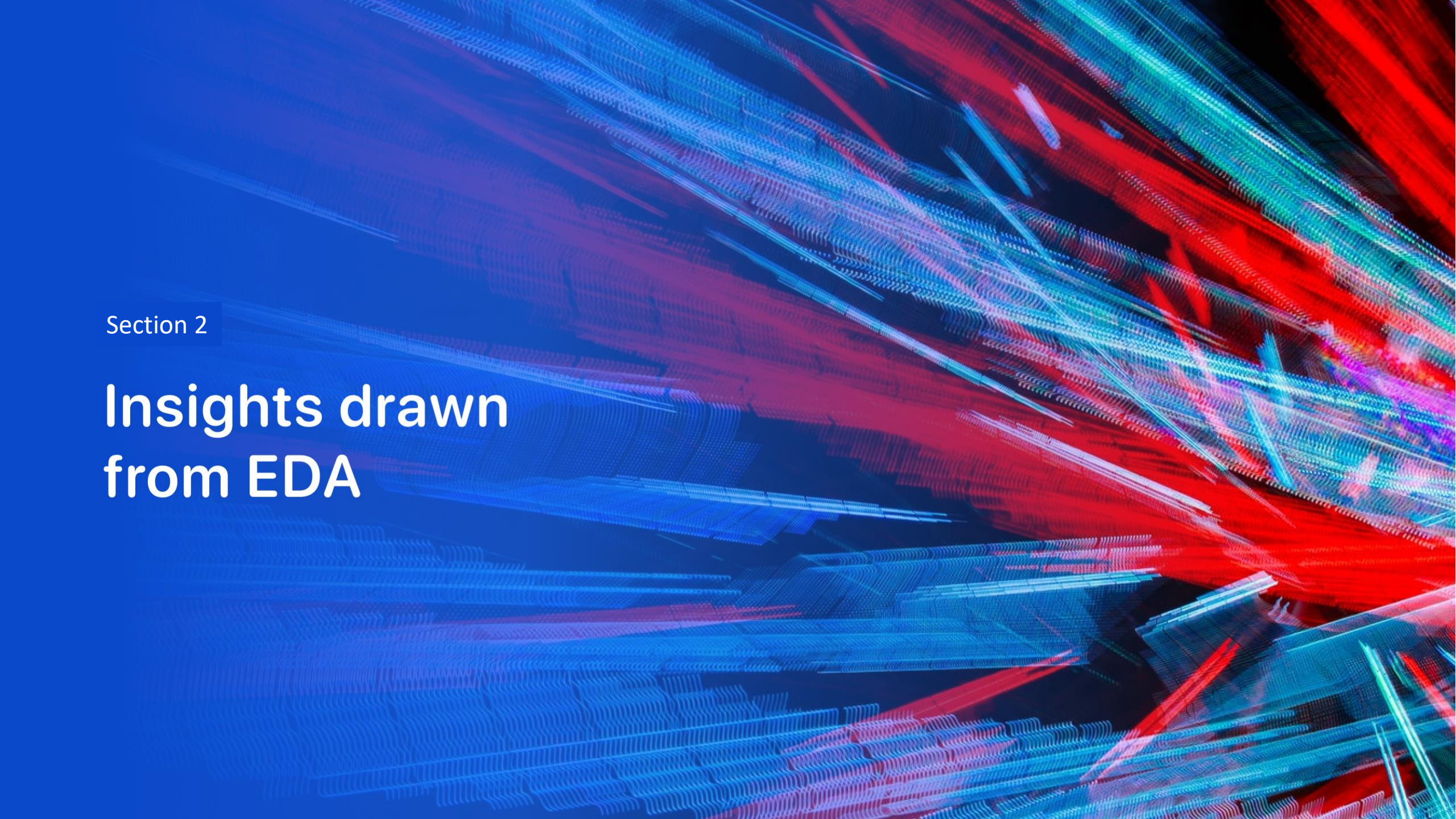
- The accuracy of 4 methods are calculated:
  - Logistic Regression
  - Support Vector Machine (SVM)
  - Classification Trees
  - K Nearsdt Neighbors (KNN)
- GitHub URL - completed predictive analysis lab:

https://github.com/wy-chan/Applied-Data-Science-Capstone/blob/main/Week4_1_Complete%20the%20Machine%20Learning%20Prediction.jupyterlite.ipynb

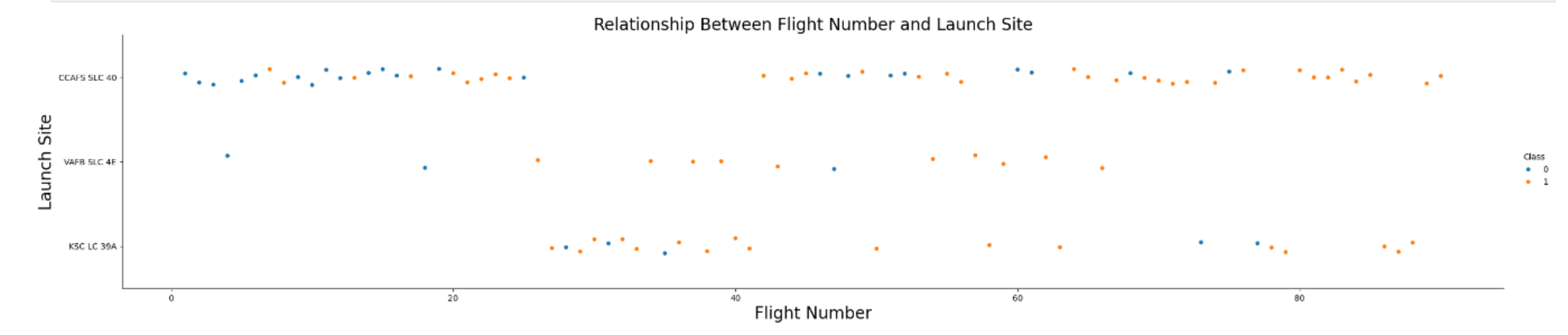


Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

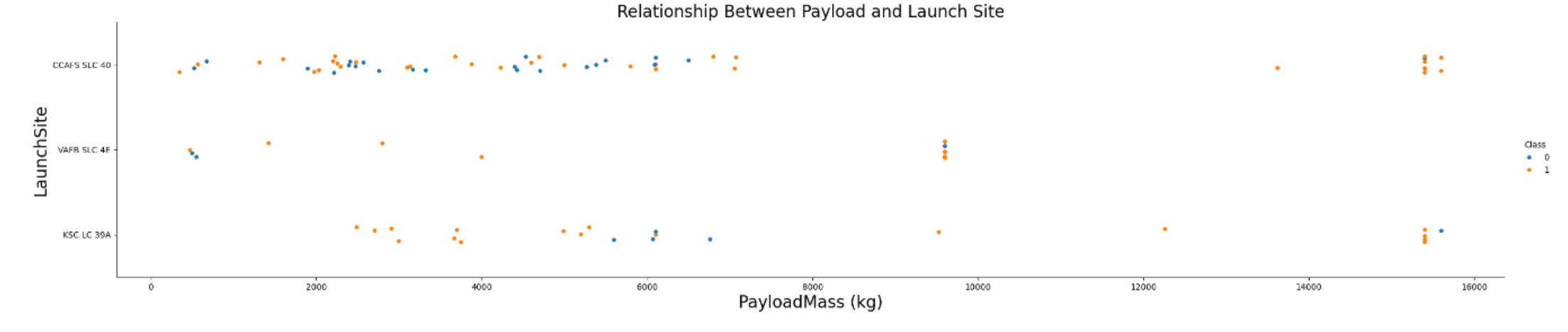

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



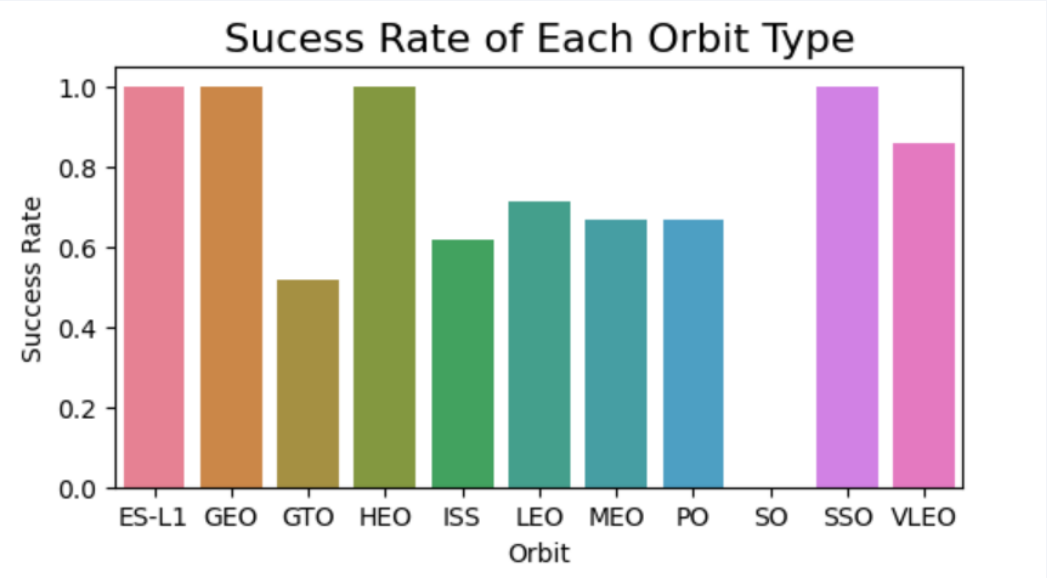
- There is a positive relationship between flight number and success rate
- According to the data, CCAFS LC-40, has a success rate of 60 %, whie KSC LC-39A and VAFB SLC 4E has a success rate of 77%
- CCAFS LC-40 has a lower success rate because there are more data with small flight numbers(continuous launch attempts)

Payload vs. Launch Site



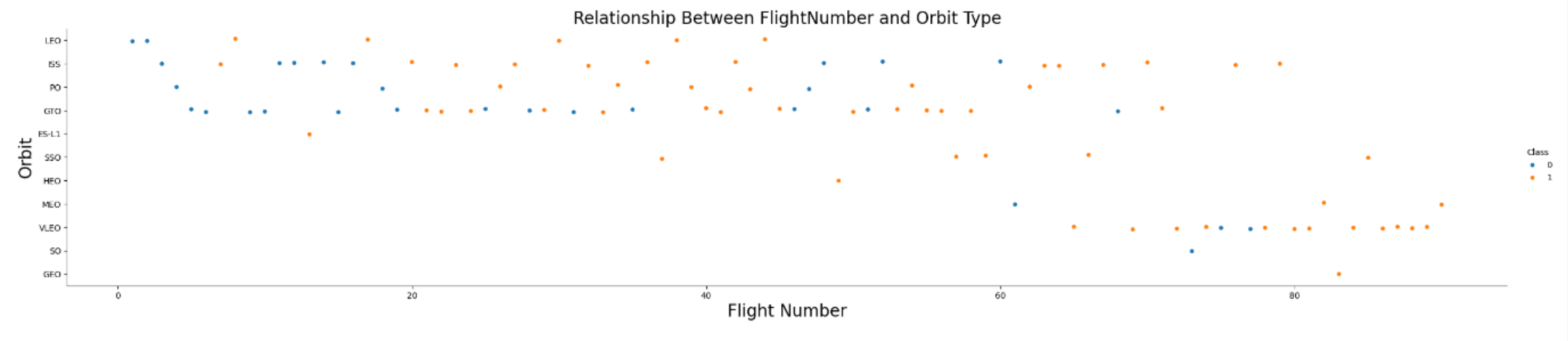
- VAFB-SLC launchsite has no rockets launched for heavy payload mass(greater than 10000).
- Most launches of CCAFS LC-40 and KSC LC-39A are either below 8000kg or between 14000-16000kg
- The success rate is high for payload mass over 9000kg

Success Rate vs. Orbit Type



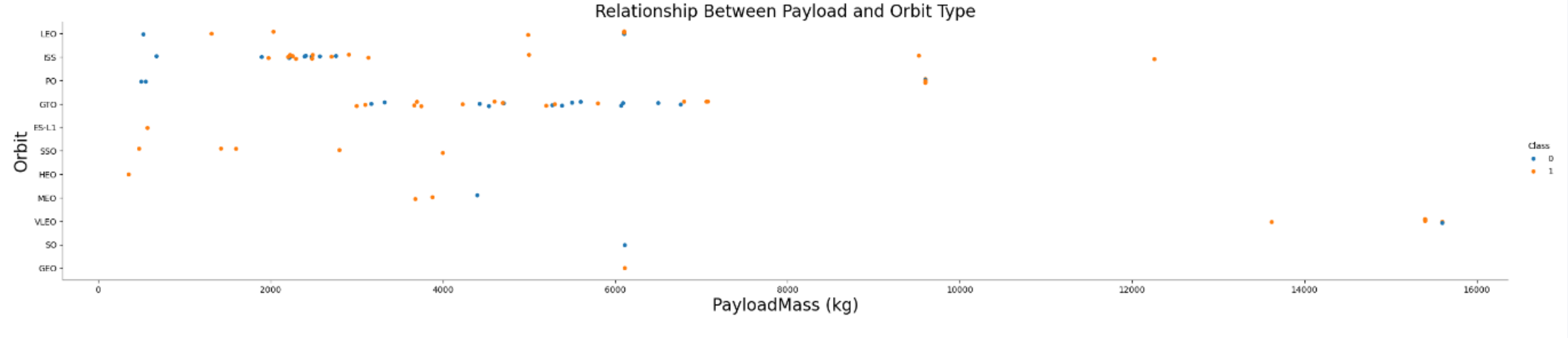
- There is a relationship between success rate and orbit type
- Orbit types EL-L1, GEO, HEO and SSO have high success rate

Flight Number vs. Orbit Type



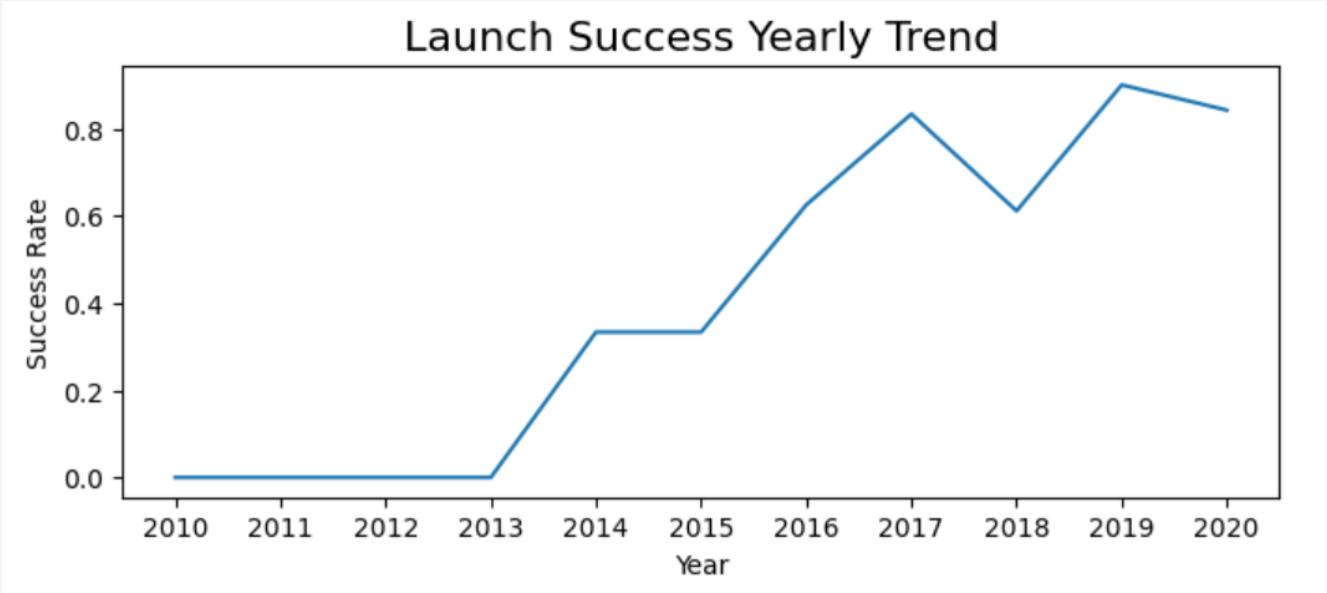
- LEO orbit's success appears related to the number of flights
- Other orbit's success rate has no clear relationship with flight number

Payload vs. Orbit Type



- Polar, LEO and ISS have higher successful landing with heavy payloads
- GTO’s success rate has no clear relationship with payload

Launch Success Yearly Trend



- the success rate since 2013 kept increasing till 2020
- There was a decrease in success rate in year 2018, but it raise back in 2019

All Launch Site Names

- Find the names of the unique launch sites:

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

There are 4 launch sites: CCAFS LC-40, CCAFS SLC-40, KSC LC-39A, VAFB SLC-4E

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`:

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

- Calculate the total payload carried by boosters from NASA

SUM (PAYLOAD_MASS__KG_)
45596

  - Launches with NASA as customer are calculated with the SUM() function
  - The sum of payload mass is 44596 Kg

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1

AVG(PAYLOAD_MASS__KG_)
2928.4

  - Launches with F9 v1.1 as the booster version are calculated with the AVG() function
  - The average payload mass is 2928.4 Kg

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

MIN(DATE)
2015-12-22

  - The date is calculated with the MIN() function
  - The first successful landing date is 2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

  - Data is selected by the criteria:
    - Landing outcome: Success (drone ship)
    - Payload mass: Between 4000 and 6000
  - 4 booster versions are found:
    - F9 FT B1022, F9 FT B1026, F9 FT B1021, F9 FT B1031.2

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

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

Mission_Outcome	count(*)
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

- There is 1 fail outcome, 98 success outcomes and 1 success with unclear payload status

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- Maximum payload mass is 2928.4 Kg
- 12 booster versions are found:

  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

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

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

Month	Landing_Outcome	Booster_Version	Launch_Site
10	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
4	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

- There are 2 launches found in year 2015
- F9 v1.1 B1012 was launched in April and F9 v1.1 B1015 was launched in October

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

- 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

  - During the specified period, 10 of the launches have no attempt on landing
  - Success (ground pad), success (drone ship) and failure (drone ship) have 5 counts each

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

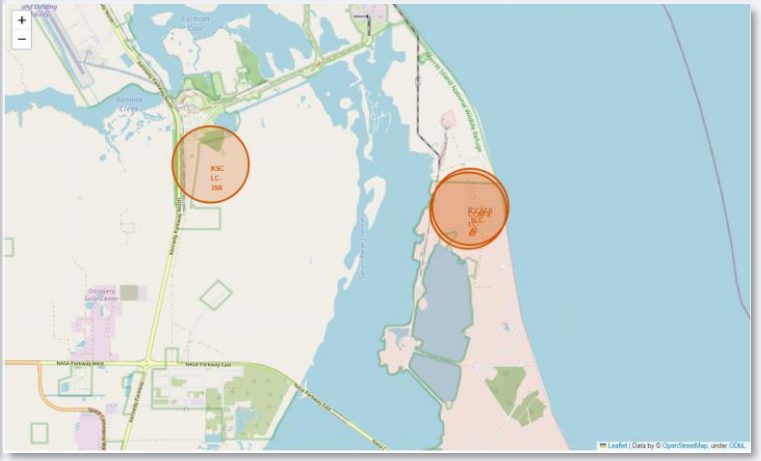
Section 3

Launch Sites Proximities Analysis

Launch Site Locations

There are 4 launch sites on the map:

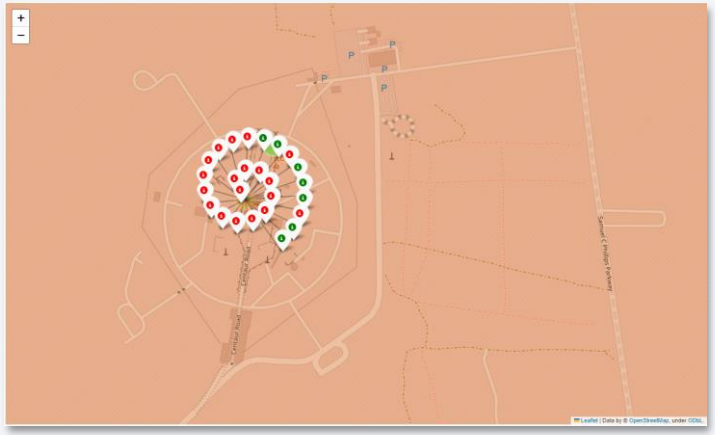
Launch Site	Lat	Long
CCAFS LC-40	28.562302	-80.577356
CCAFS SLC-40	28.563197	-80.576820
KSC LC-39A	28.573255	-80.646895
VAFB SLC-4E	34.632834	-120.610745



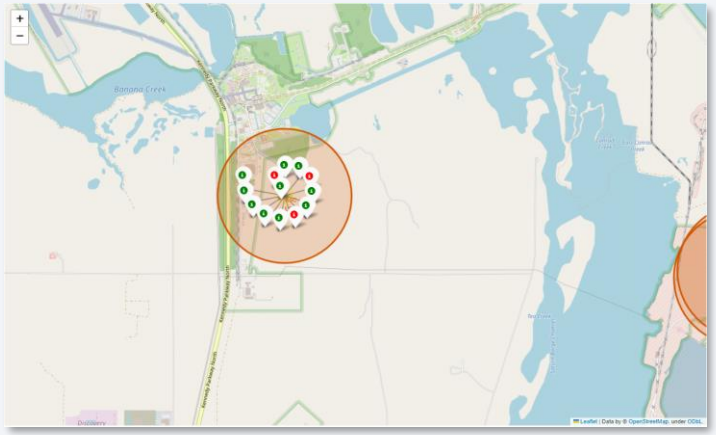
Launch Outcomes of Launch Sites

- Color-labeled launch outcomes are shown on the map

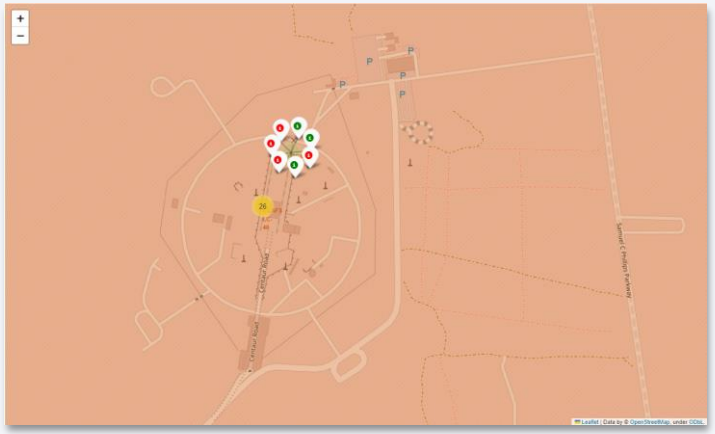
CCAFS LC-40



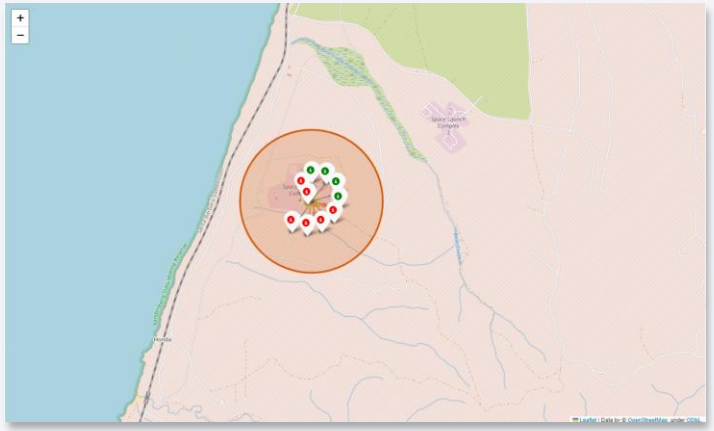
KSC LC-39A



CCAFS SLC-40



VAFB SLC-4E



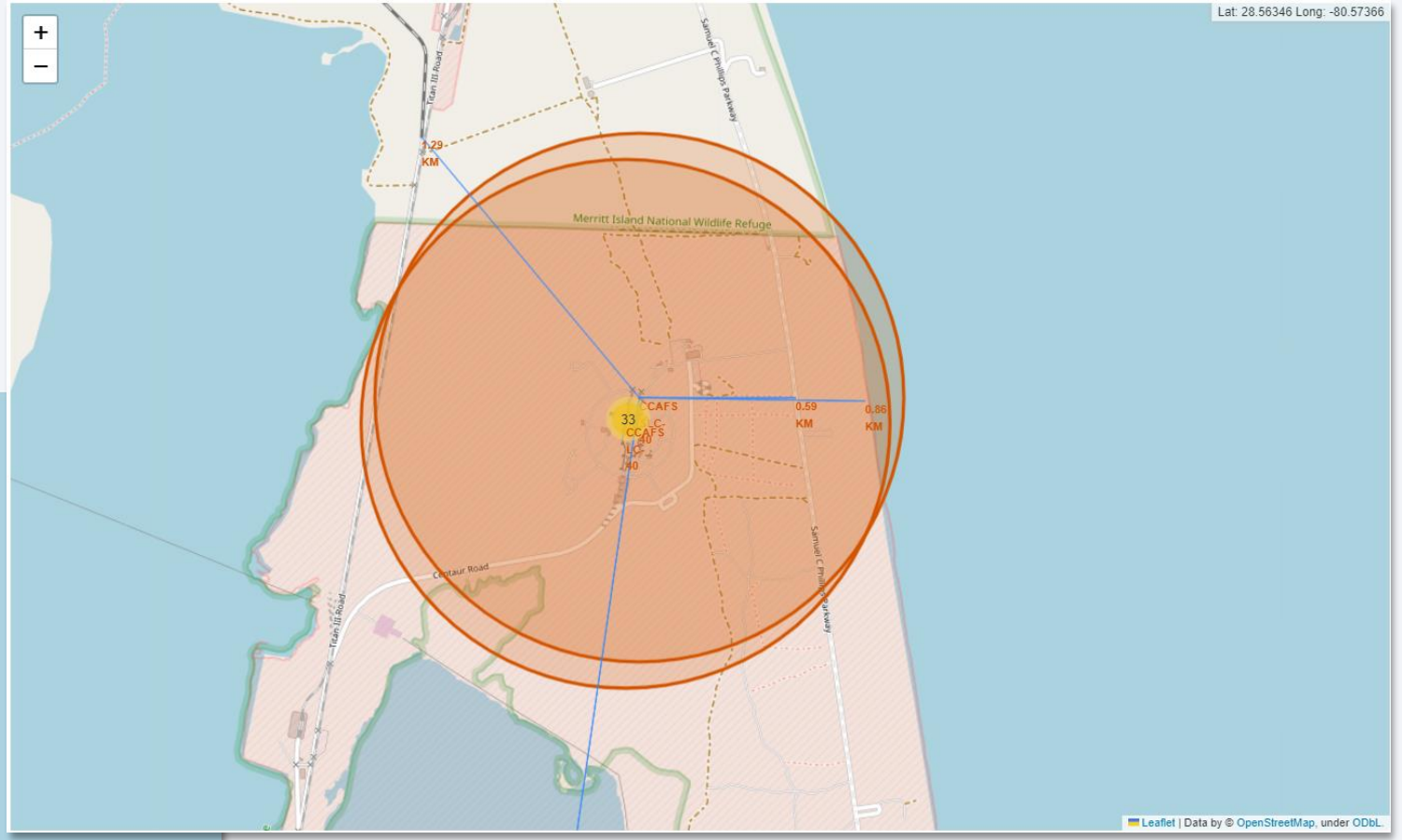
Circle Labels:

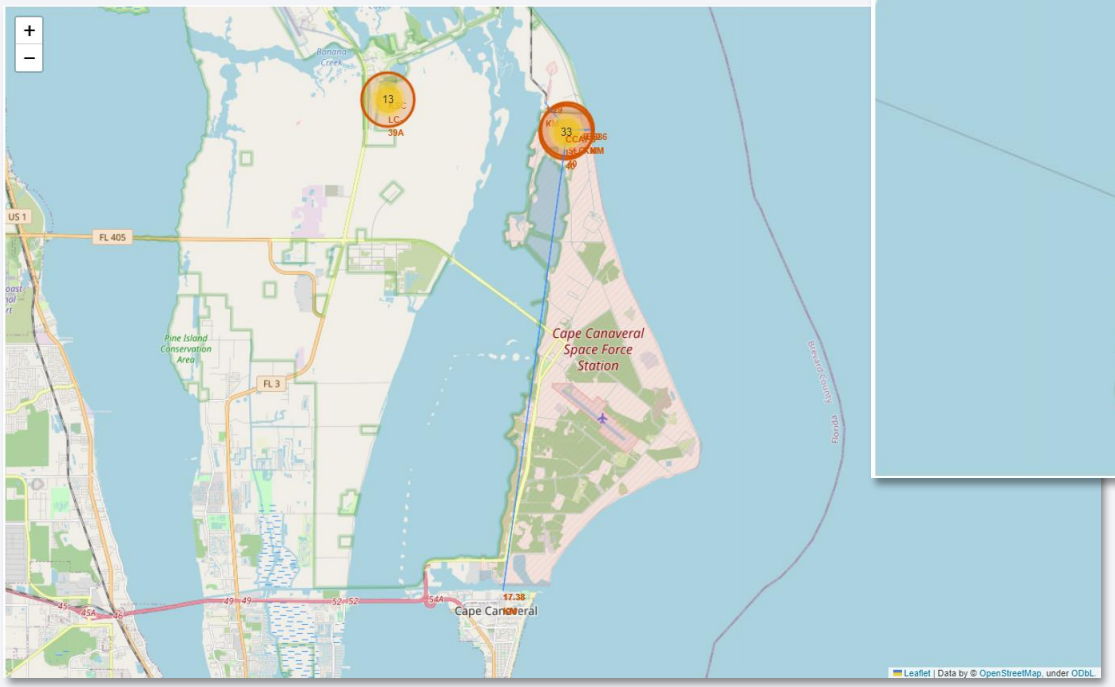
- **Green - Success**
- **Red - Fail**

- CCAFS LC_40 had the most launches but also the most failure
- KSC LC-39A had the highest number of success landing

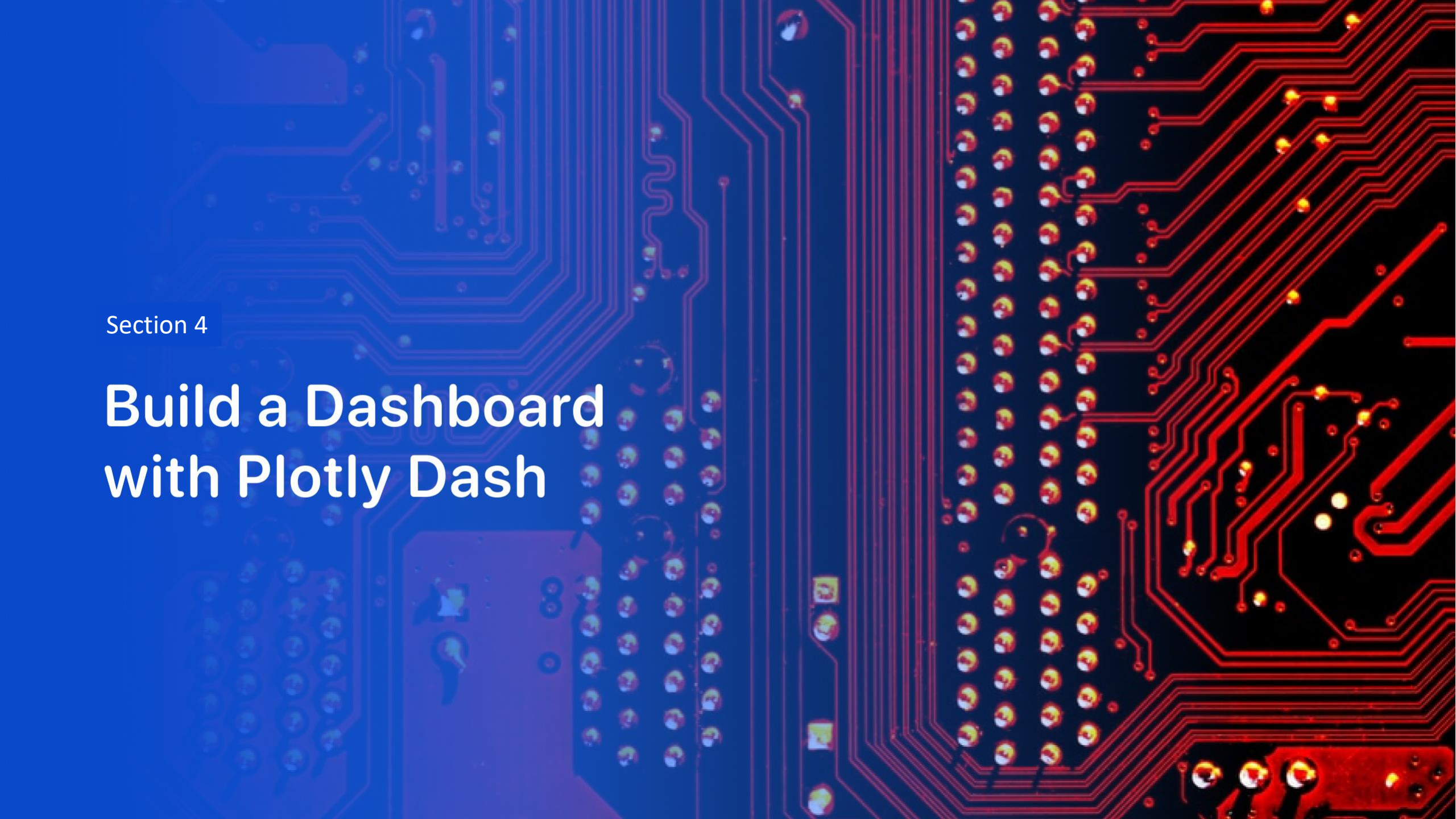
Distance of launch site CCAFS SLC-40 to its proximities

- Railway - 1.29 KM
- Highway - 0.59KM
- Coastline - 0.86 KM
- City - 17.38 KM



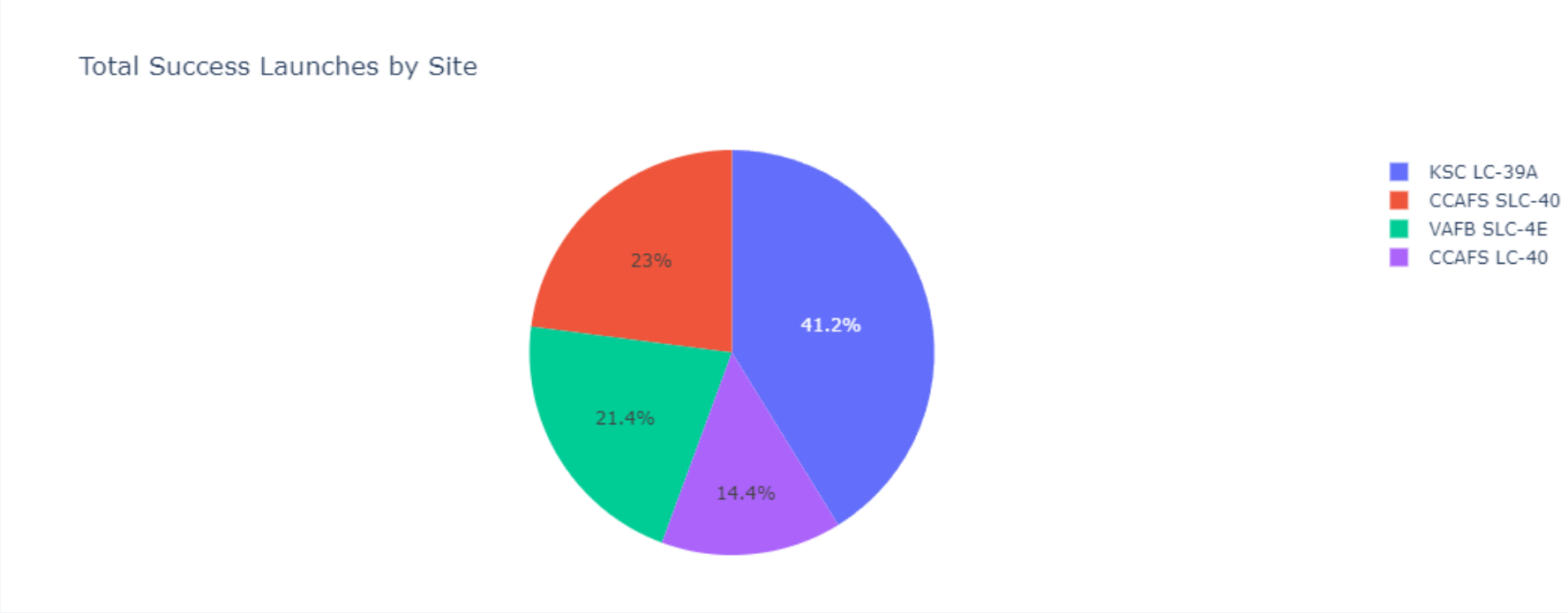


- The launch site is close to railway, highway and coastline, but far from city


Section 4
Build a Dashboard with Plotly Dash

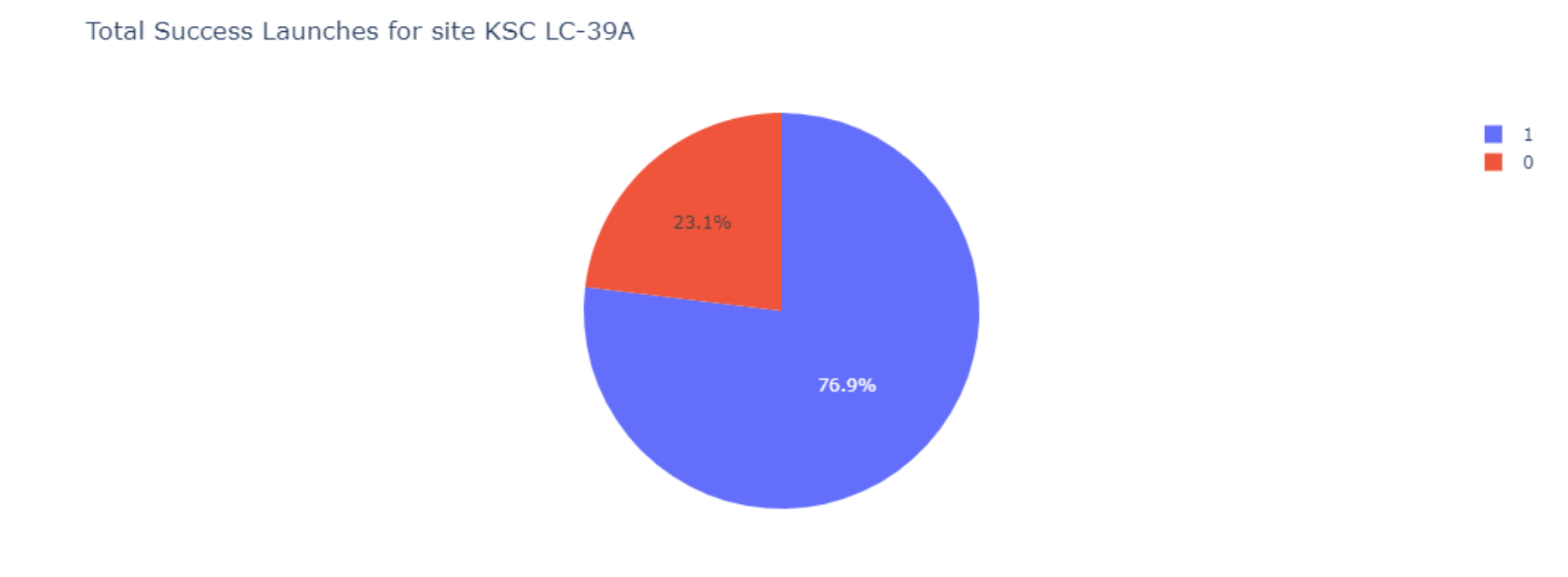
Launch Success Count For All Sites

- The pie chart shows total success launches by all sites
- Out of all launch sites, KSC LC-39A has the largest successful launches



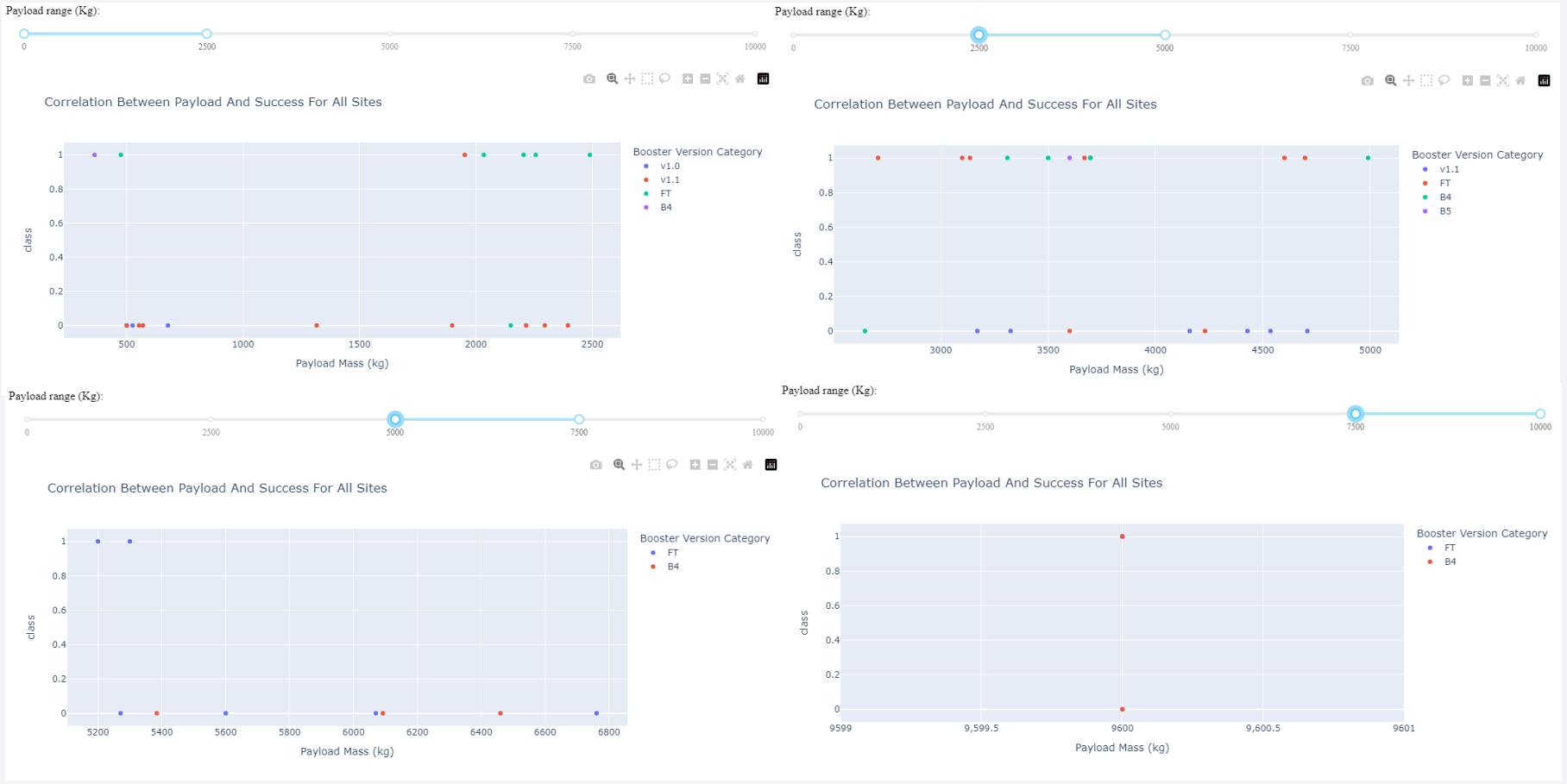
Launch Success Ratio Of Site KSC LC-39A

- Launch site KSC LC-39A also has the highest launch success ratio
- The Launch Success Rate is 76.9%



Correlation Between Payload And Success

- Payload range with highest success rate: 2500 – 5000 kg
- Payload range with lowest success rate: 7500 – 10000 kg
- Booster version with highest success rate: FT

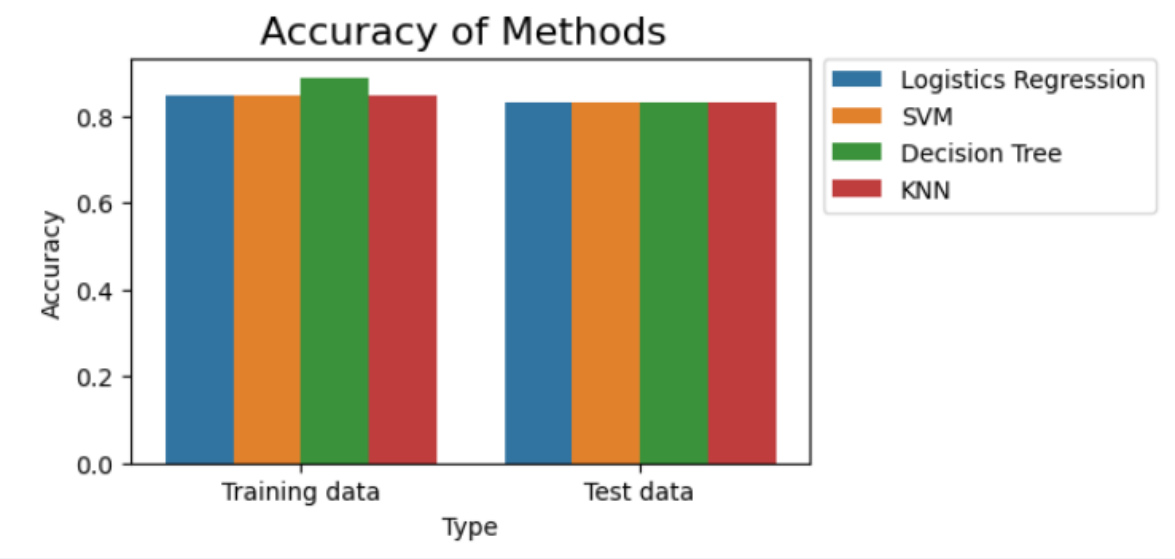


Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Built model accuracy for all built classification models is compared in the bar chart
- Classification accuracy is the same in all models for tested data

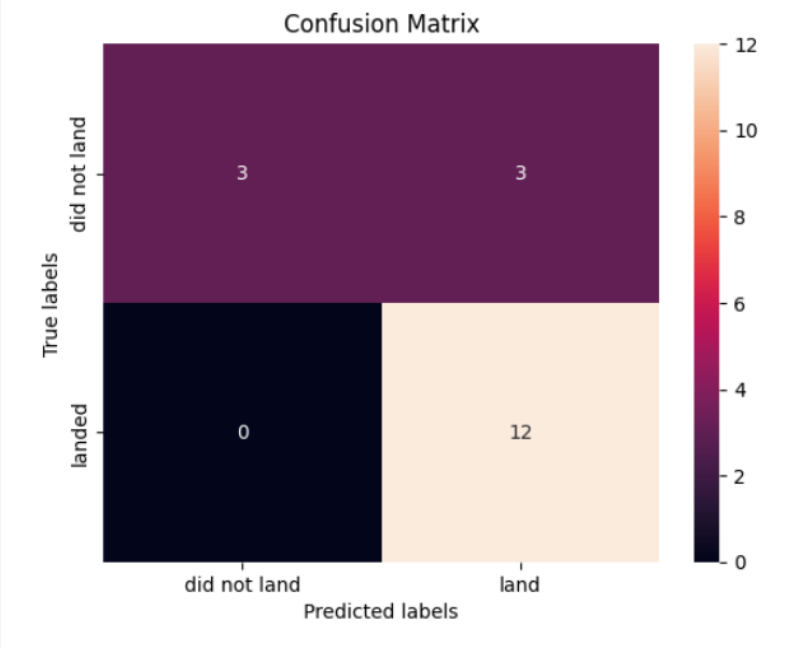


Confusion Matrix

- Confusion matrix of decision tree:

True Positive	12
False Positive	3
True Negative	3
False Negative	0

- Precision: 12/(12+3) = 0.8
- Recall: 12/(12+0) = 1.0
- F1-Score = 2*(1*0.8)/(1+0.8) = 0.88
- F1-Score closer to 1 means higher accuracy



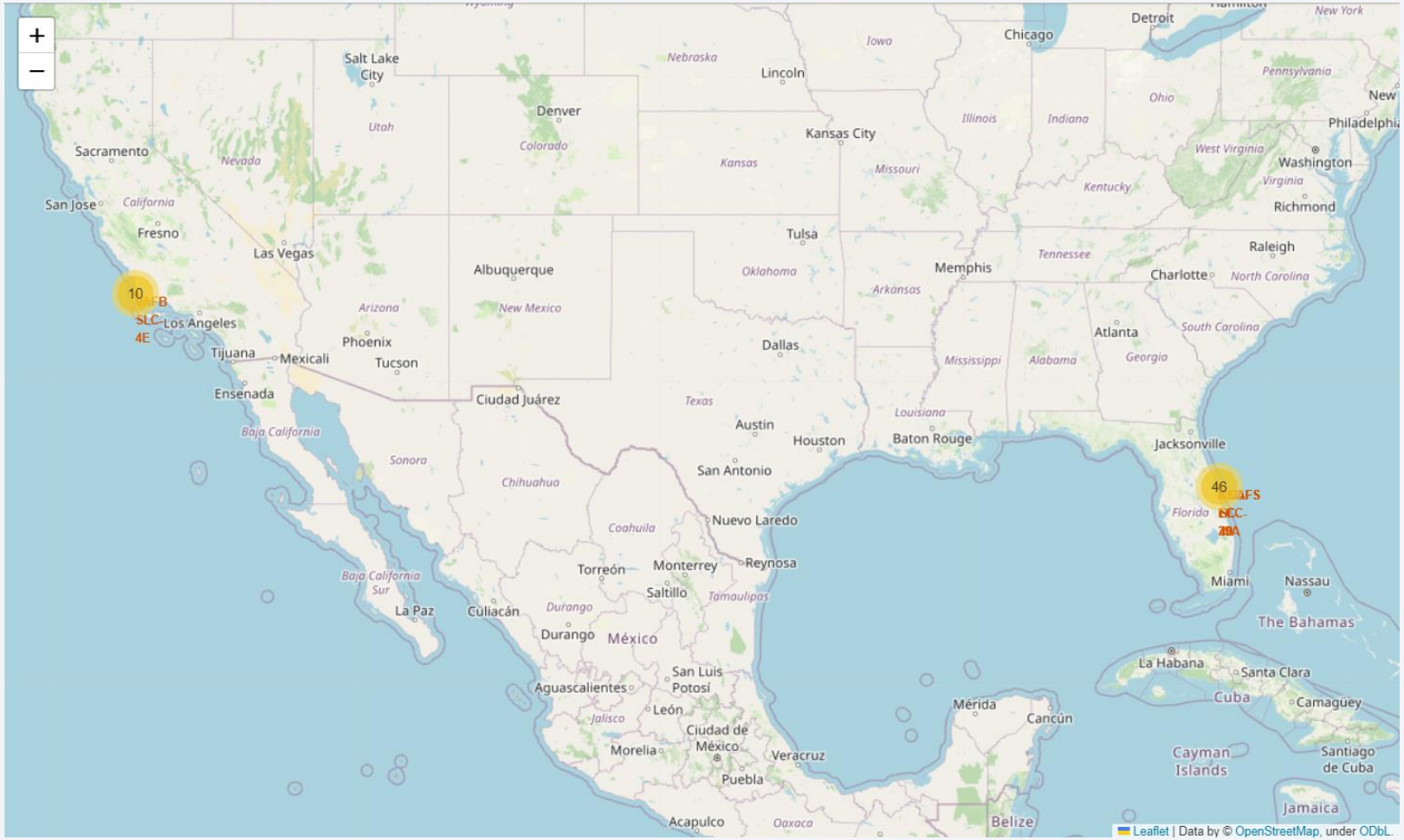
Conclusions

- Factors that increase landing success rate:
  - **Launch site:** KSC LC-39A had the highest success rate
  - **Payload mass:** 2500 – 5000 kg
  - **Orbit types:** EL-L1, GEO, HEO and SSO
  - **Yearly trend**: the success rate is increasing

- Machine learning model:
  - Accuracy is the same among Logistic Regression, SVM, Decision Tree, and KNN

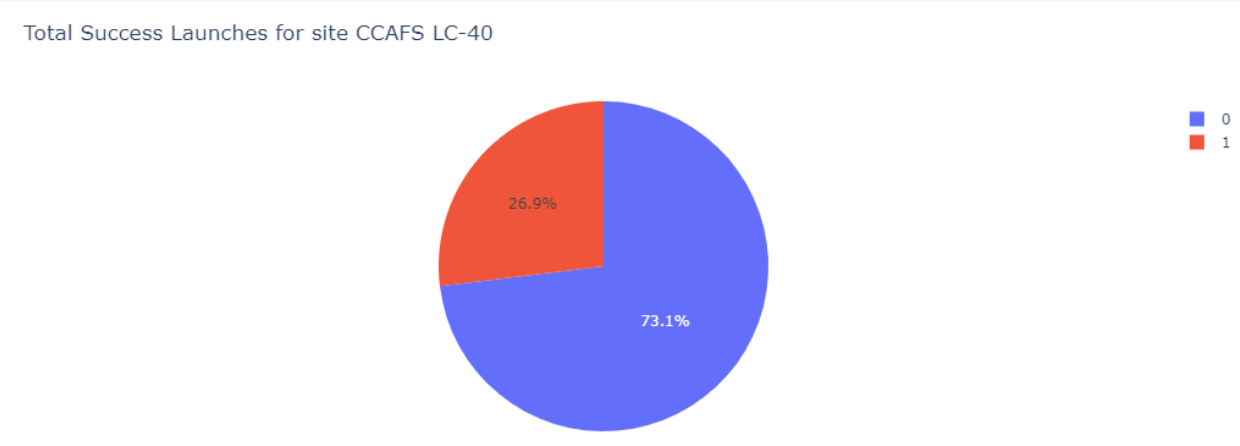
Appendix

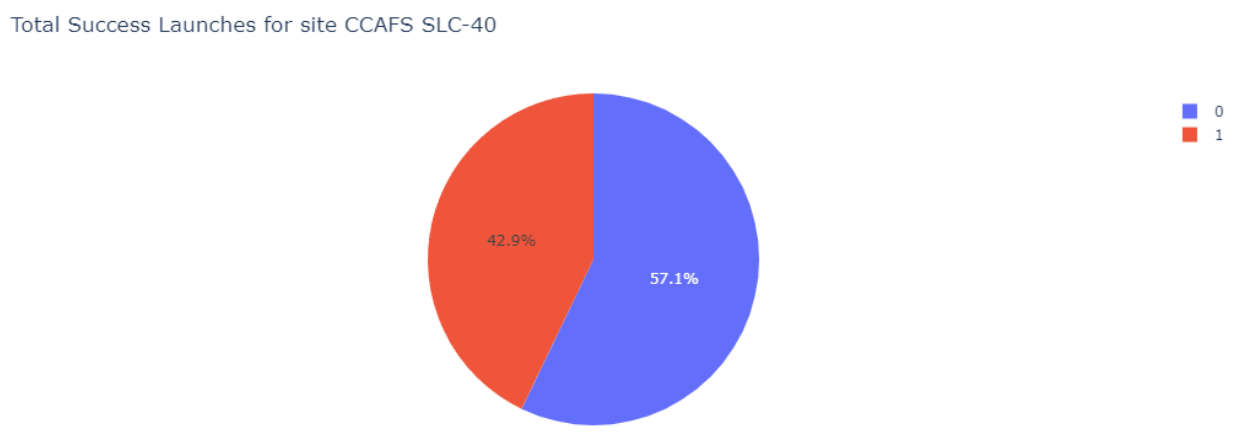
Launch sites and number of launches:

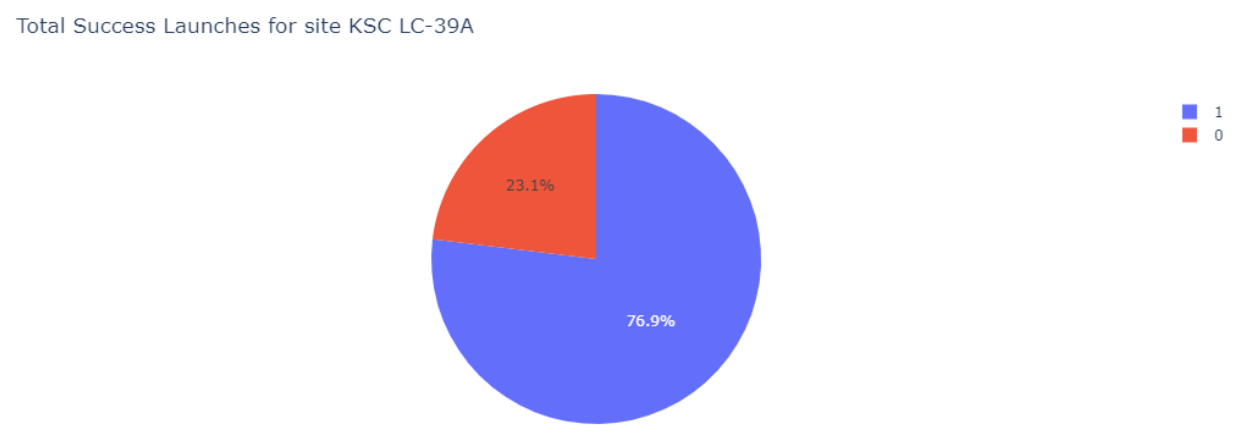


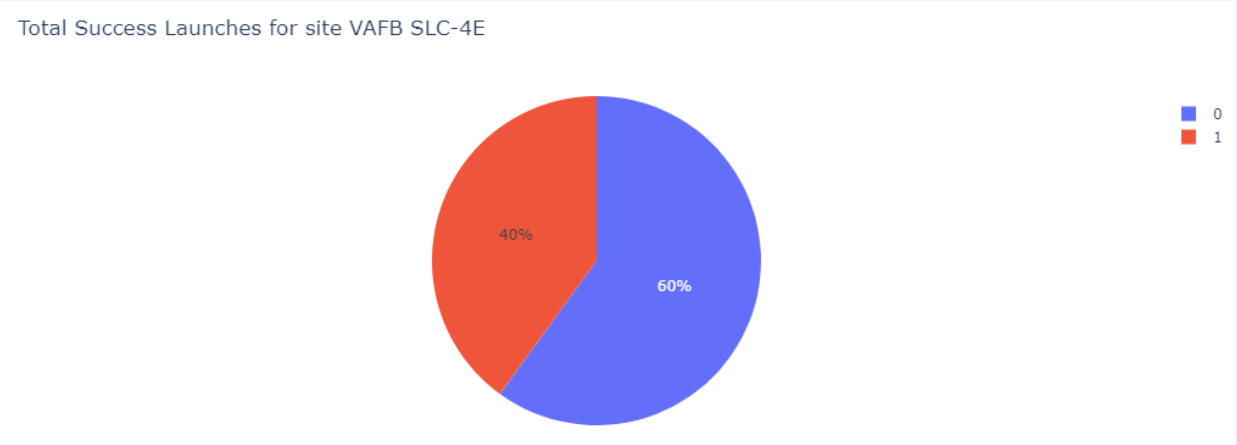
Appendix

Launch success ratio by sites:









Thank you!