IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Alexander Häfele
December 16th, 2021

Outline

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

Executive Summary

- Target of the project was to predict the costs of a rocket launch from the competitor SpaceX in order to prepare a successful bid of Space Y. The cost could be determined based on the outcome of a Falcon 9 rocket. So will it land successfully or not.
- Summary of methodologies: In order to achieve the respective description, I collected the data via a data get request from a SpaceX API and web scraping from Wikipedia. The merged and cleaned data was used for an Exploratory Data Analysis in order to find patterns which were than used for the creation of an prediction model. This model provides a relatively high accuracy for the prediction if a landing of a SpaceX rocket will be successful for a specific customer scenario with launch site, Orbit and Payload mass.
- Summary of all results

Introduction

- As a data scientists of SpaceY I was asked to provide a model which could predict the outcome of the rocket landing from SpaceX as this determines the costs of the mission. This knowledge would enable SpaceY to create a bid which is competitive with the ones from SpaceX.
- Based on the available data from a SpaceX API and Wikipedia I should find patterns which have a statistical correlation to a successful outcome and transfer those patterns into a predictive model.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Get request from Data API from SpaceX and web scraping with BeautifulSoup from Wikipedia
- Perform data wrangling
  - Data was cleaned, especially missing values were replaced with mean. First we analyzed with counting the launch site and the orbital targets in order to see if there is a correlation of those with the outcome of the rocket landing. The data was than enriched with a class column which indicates the outcome.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

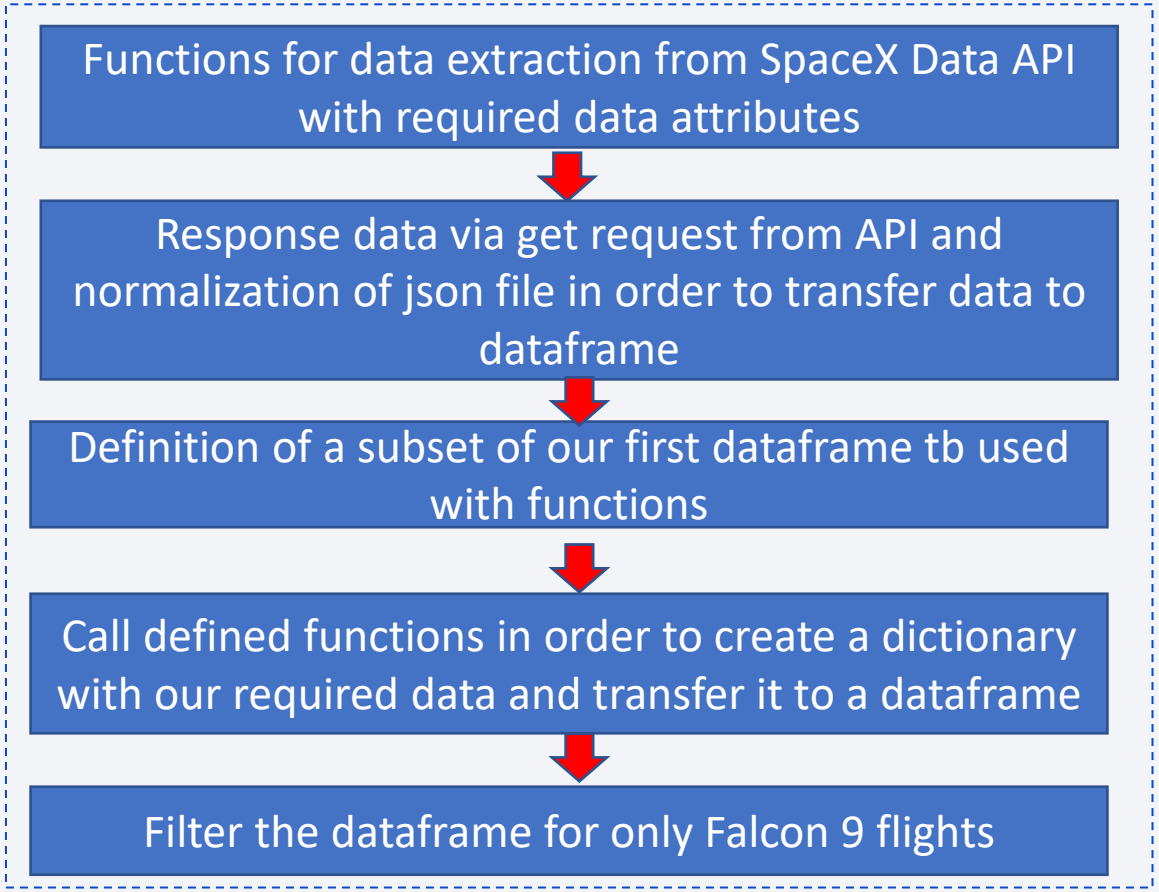
Data Collection

- Describe how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts

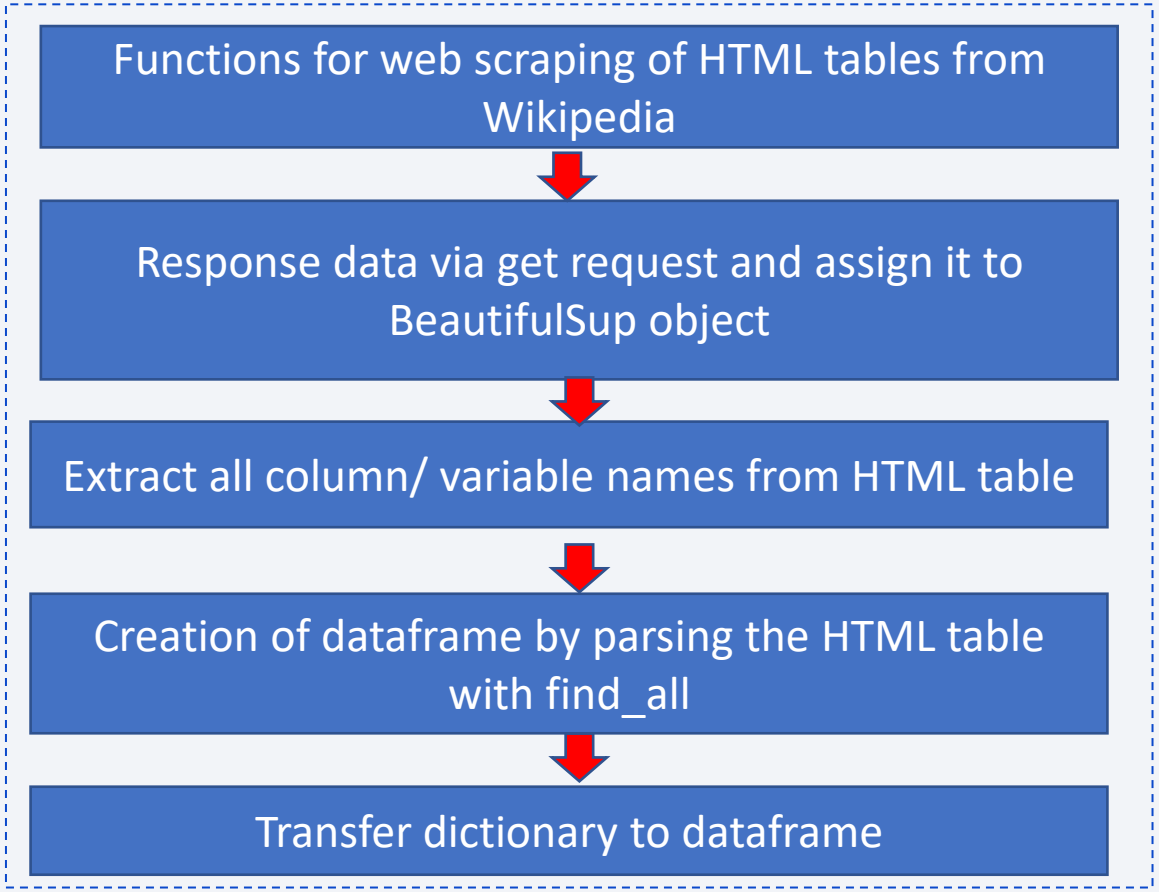
- Github - API notebook



Data Collection - Scraping

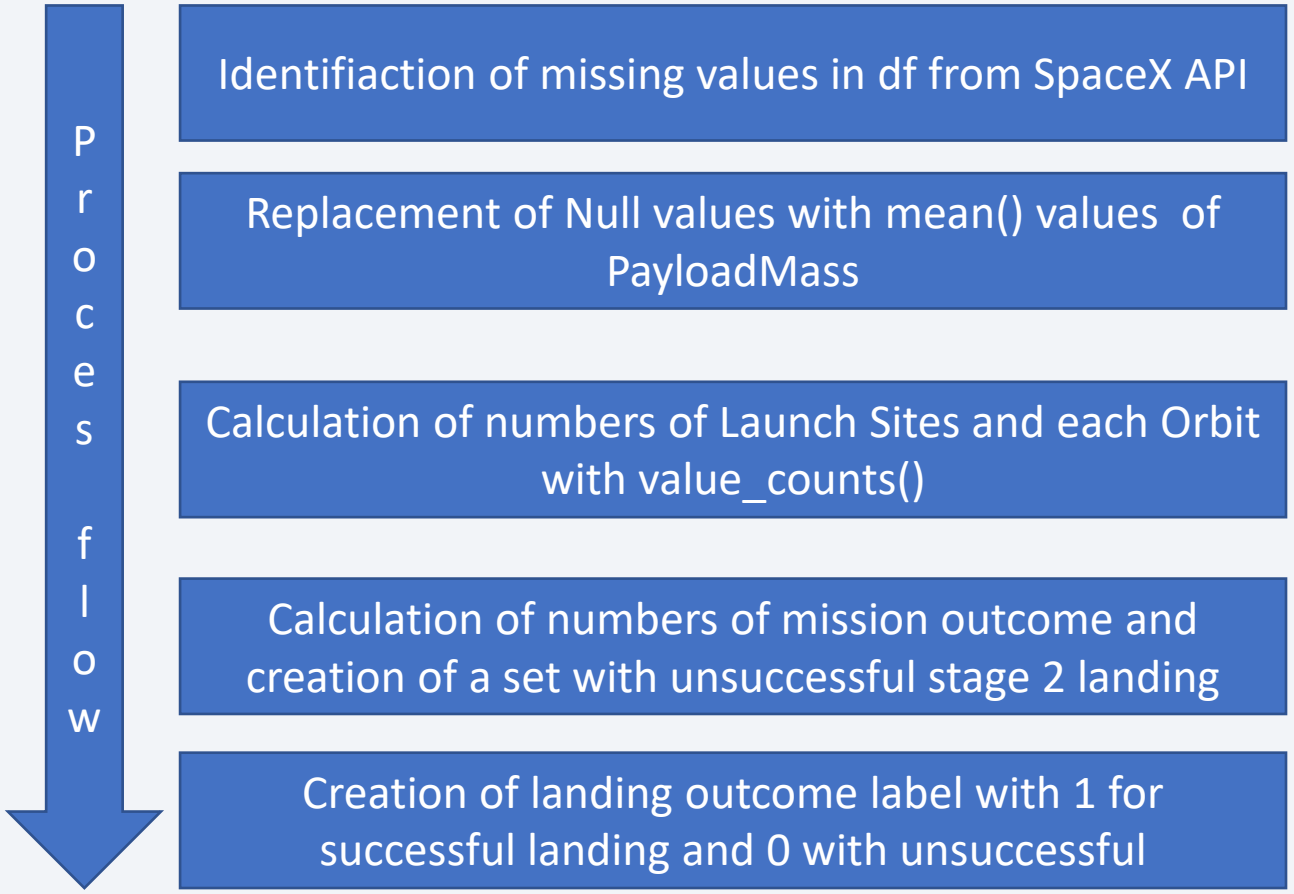
- Present your web scraping process using key phrases and flowcharts

- Github - Web Scraping Notebook



Data Wrangling

- Github - Data Wrangling notebook



EDA with Data Visualization

- Scatterplots in order to visualize the relationship of Payload Mass and Flight Number, Launch Site and Flight Number, Launch Site and Payload Mass, Orbit and Flight Number and Orbit and Payload Mass. The color coding showed if the landing was successful.
- Bar chart was created to show the success rate of each orbit.
- Line chart was created to show the yearly increasing success rate trend.
- Github - EDA with visualization notebook

EDA with SQL

- Select Distinct
- Group BY
- Where x Like AND between
- Limit
- sum, avg, min, count
- Order by Desc
- Subquerry

- Github - EDA with SQL notebook

Build an Interactive Map with Folium

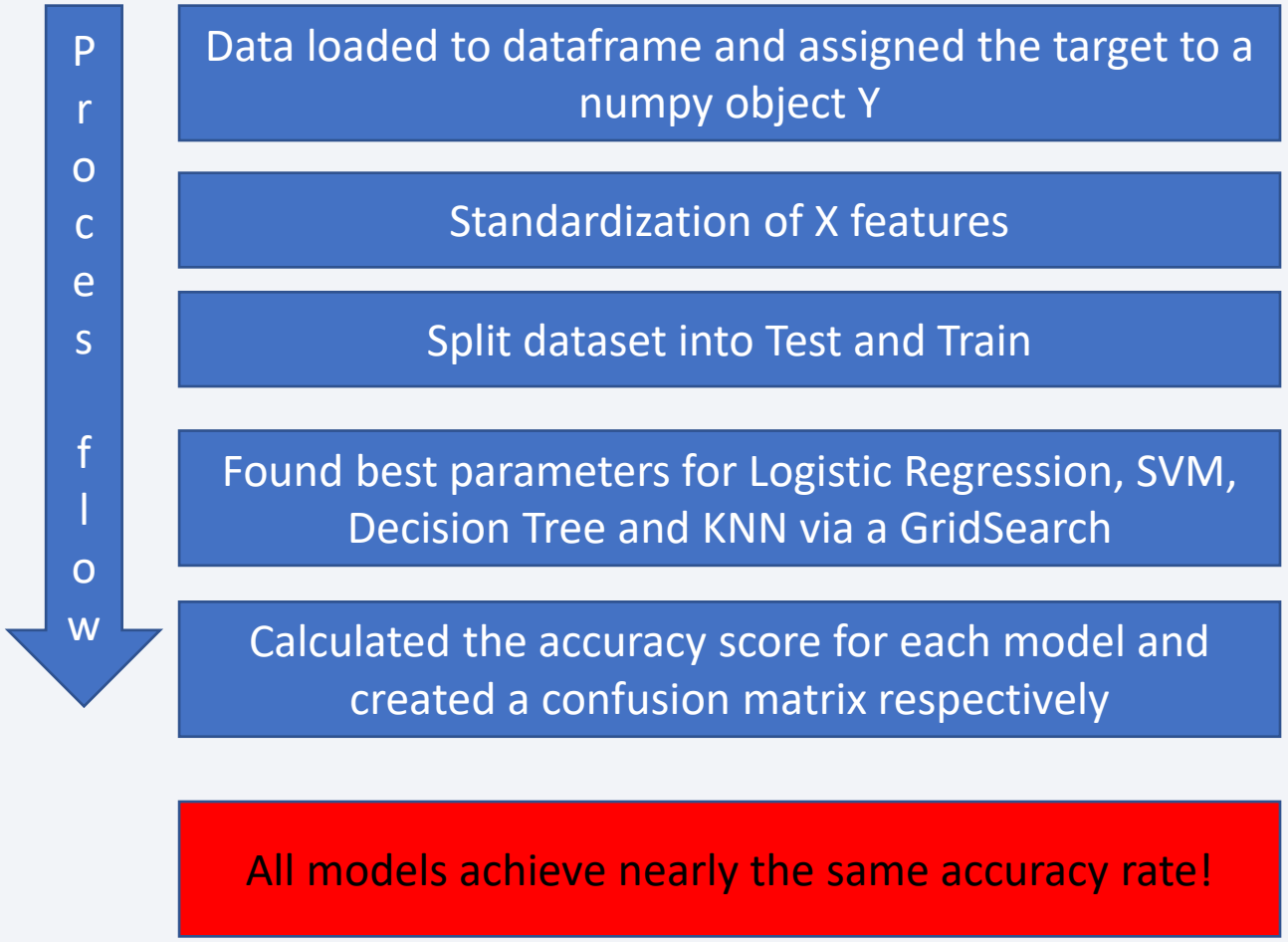
- The following objects were added to the map:
  - Circles to show the location of the launch sites,
  - Marker_cluster with colored markers to show the outcome of the landings and to easily identify which site has a relatively high success rates
  - Coordinates of the mouse position
  - Polylines in order to show the distance btw several defined points with markers which state the distance
- Github - Interactive Folium map notebook

Build a Dashboard with Plotly Dash

- Dropdown menu to change the launch site which impacts the data used for a pie chart which shows either the success rate of the all launch sites or the outcome of the selected launch site.
- A range slider in order to chose the payload mass kg which is than reflected in a plotter chart. This chart shows the correlation between payload mass and landing outcome for each Booster Version via color coding.
- Github - Plotly Dash lab code

Predictive Analysis (Classification)

- Github - Predictive analysis notebook

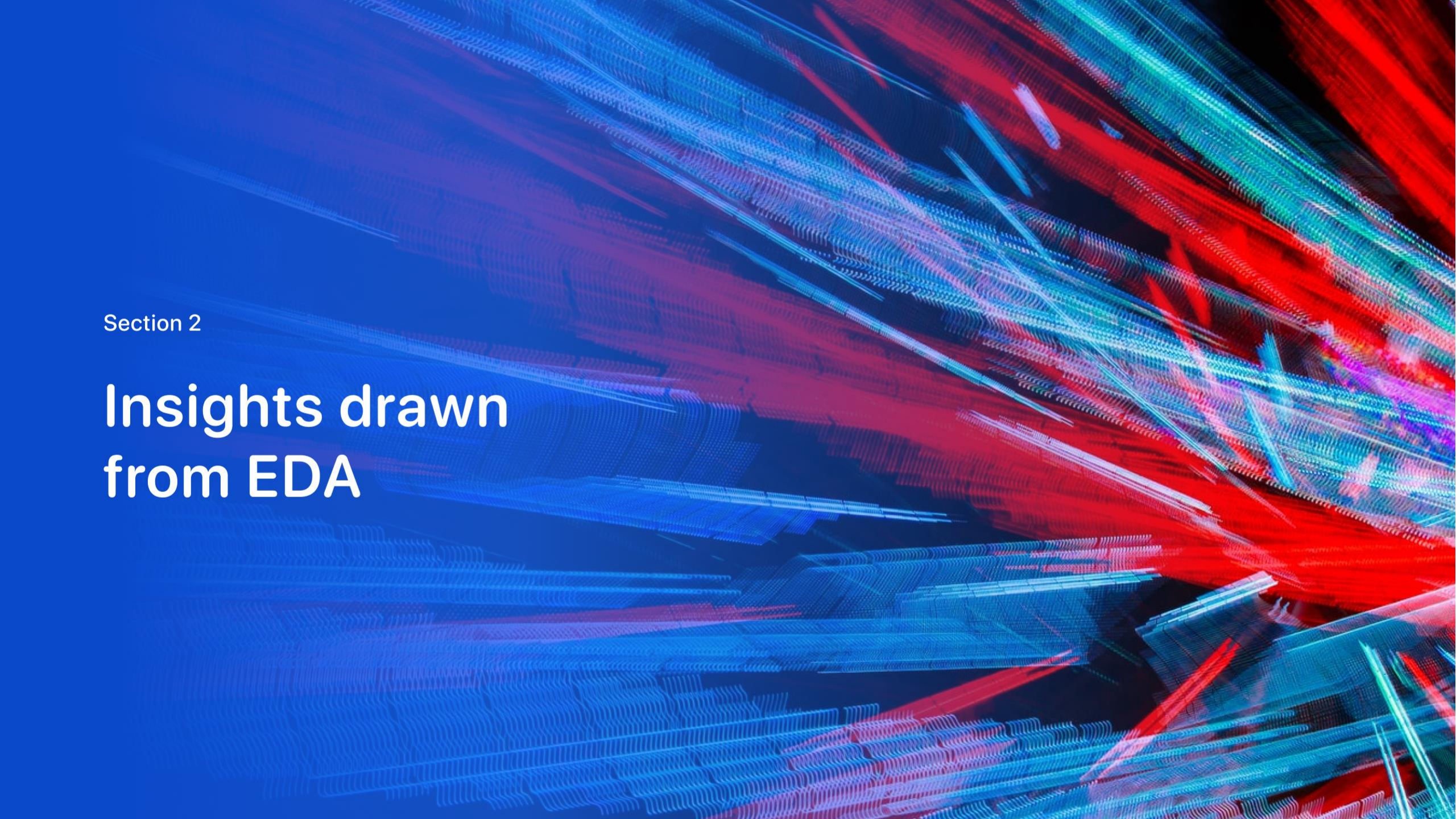


Results

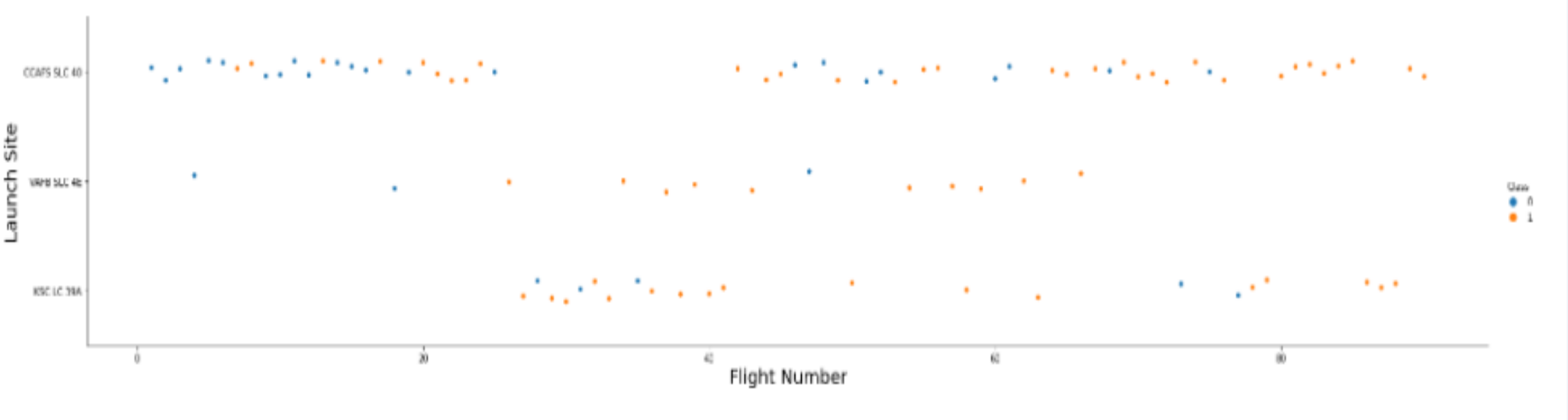
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

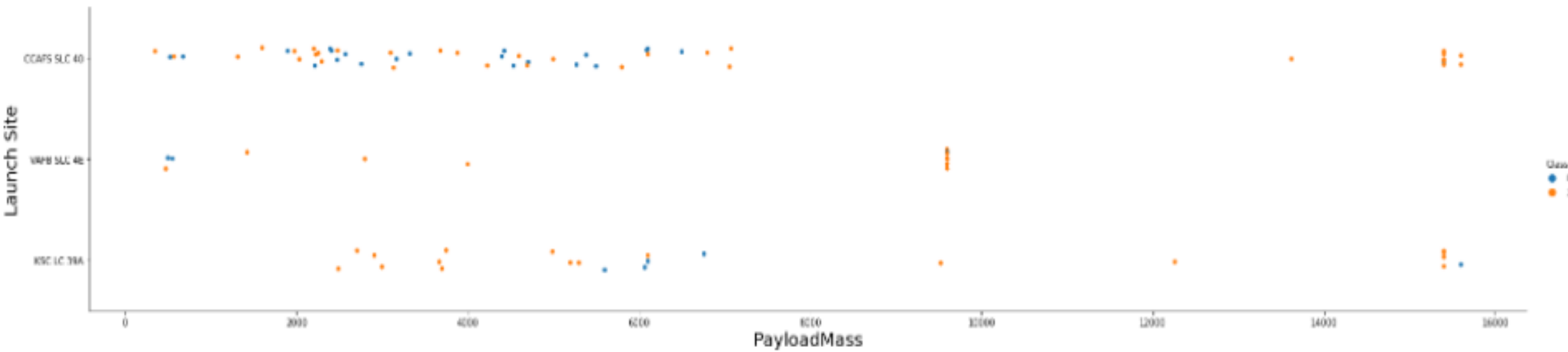


Flight Number vs. Launch Site



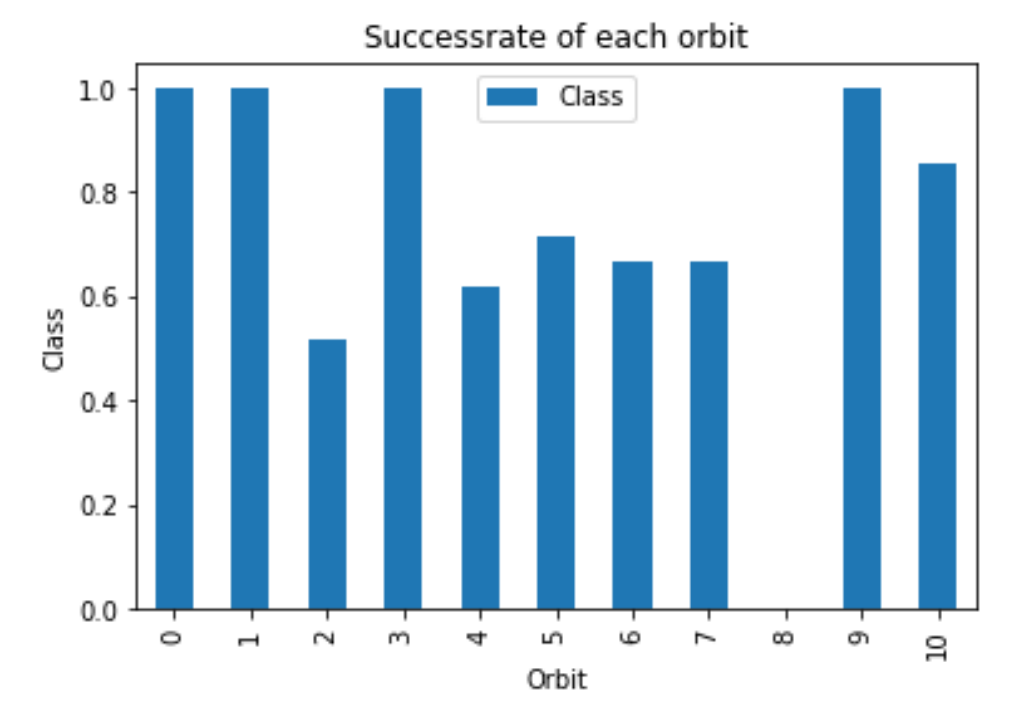
- Launch site KSC LC 39A has the highest success rate (class 1 in orange color)
- CCAFS SLC 40 has the lowest succes rate (class 0 in blue color)

Payload vs. Launch Site



- Higher Payload Mass in kg showed high success rate over 7500 kg independently from the launch site

Success Rate vs. Orbit Type

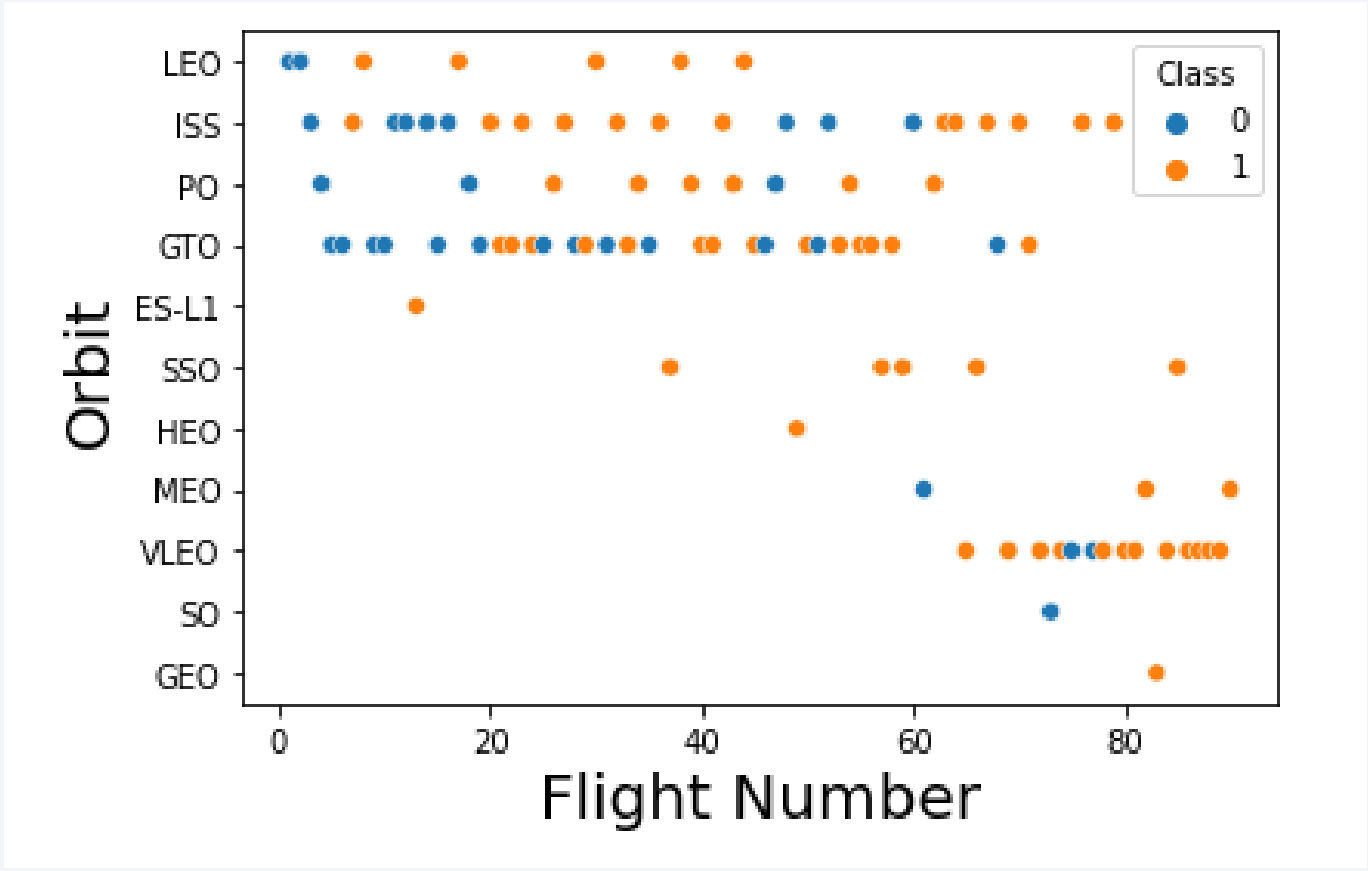


	Orbit	Class
0	ES-L1	1.000000
1	GEO	1.000000
2	GTO	0.518519
3	HEO	1.000000
4	ISS	0.619048
5	LEO	0.714286
6	MEO	0.666667
7	PO	0.666667
8	SO	0.000000
9	SSO	1.000000
10	VLEO	0.857143

- Orbit 0, 1, 3 and 9 have the highest success rate.

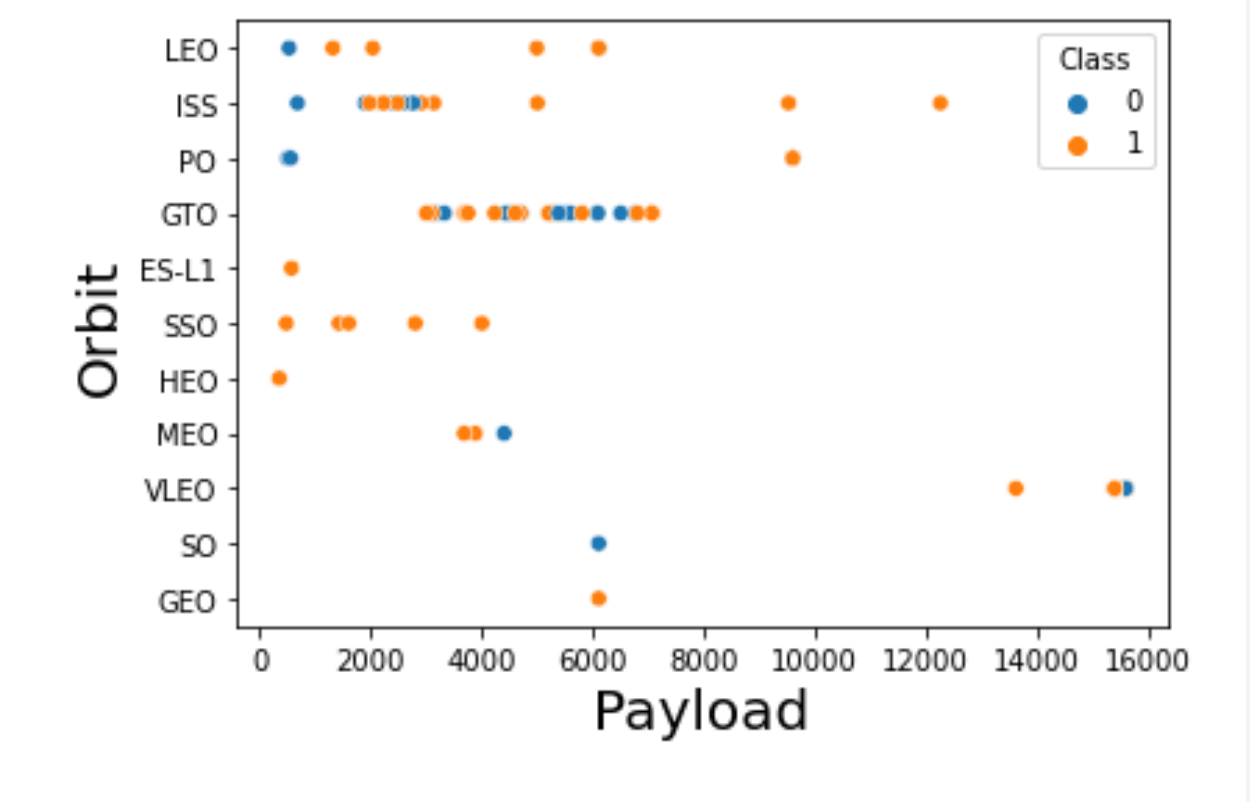
Flight Number vs. Orbit Type

- Orbit VLEO and SSO are showing a high sucess rate



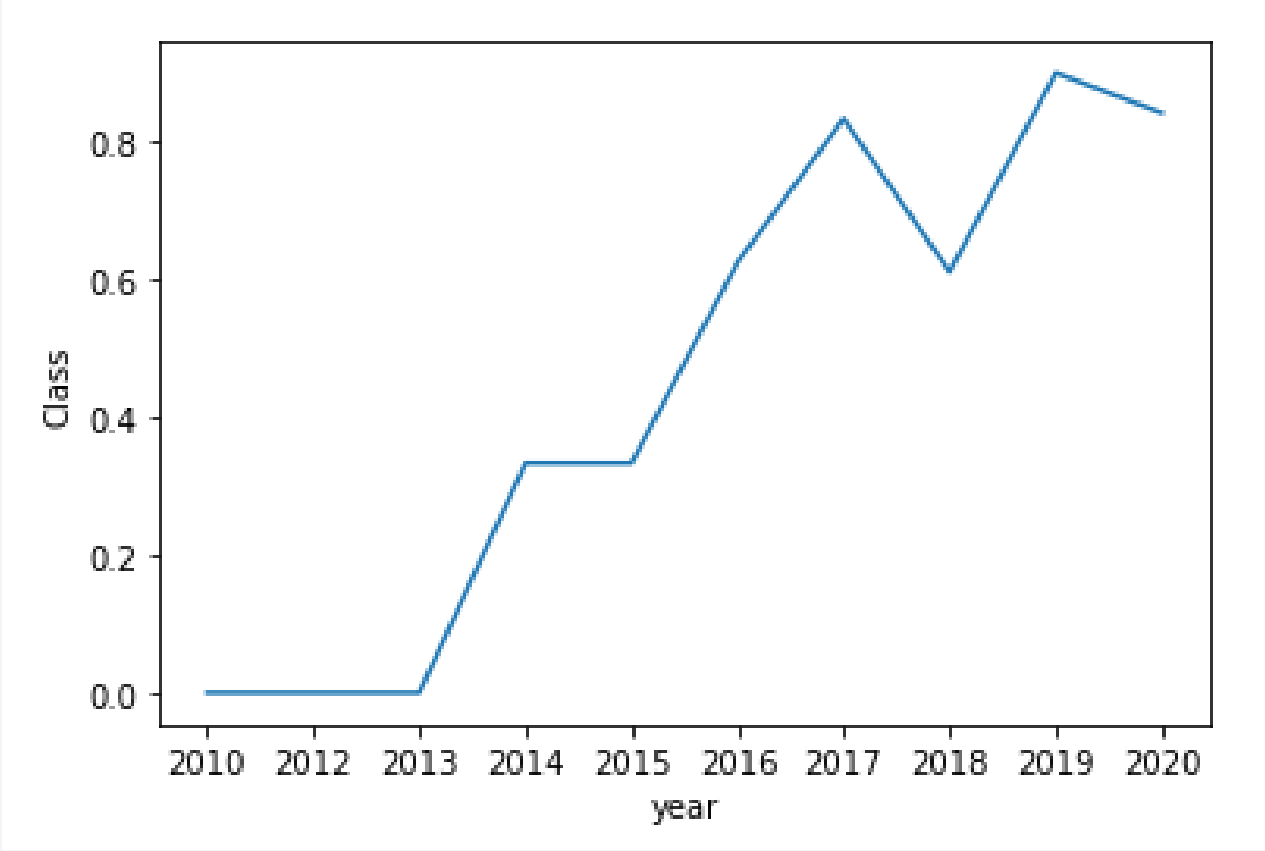
Payload vs. Orbit Type

- SSO most succesful with Payload <6000 kg



Launch Success Yearly Trend

- The success rate is clearly increasing throughout the years.



All Launch Site Names

- Find the names of the unique launch sites
- Present your query result with a short explanation here

launch_site
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`
- Present your query result with a short explanation here

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

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
- Present your query result with a short explanation here

payloadmass_kg_by_nasa
45596

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- Present your query result with a short explanation here

avg_payload_with_f9
2534

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- Present your query result with a short explanation here

first_landing
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
- Present your query result with a short explanation here

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
- Present your query result with a short explanation here

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

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- Present your query result with a short explanation here

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

2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- Present your query result with a short explanation here

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

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
- Present your query result with a short explanation here

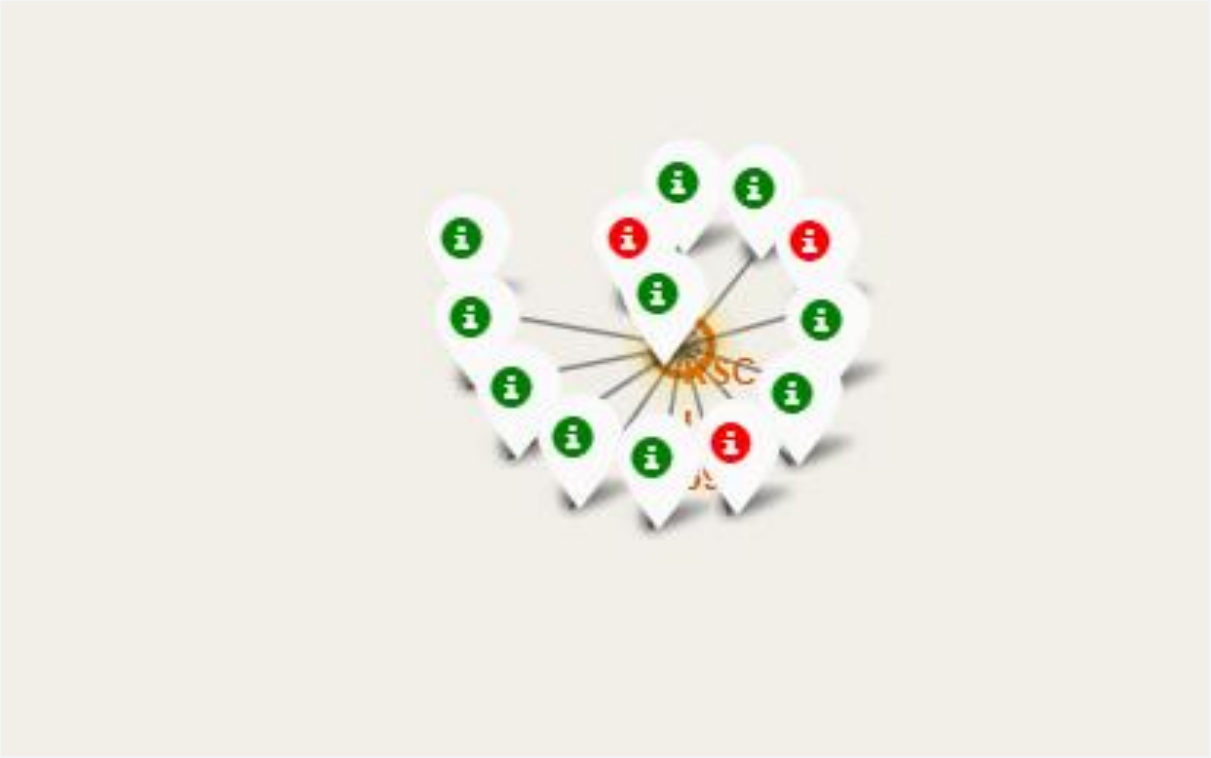
landing__outcome	2
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 4

Launch Sites Proximities Analysis

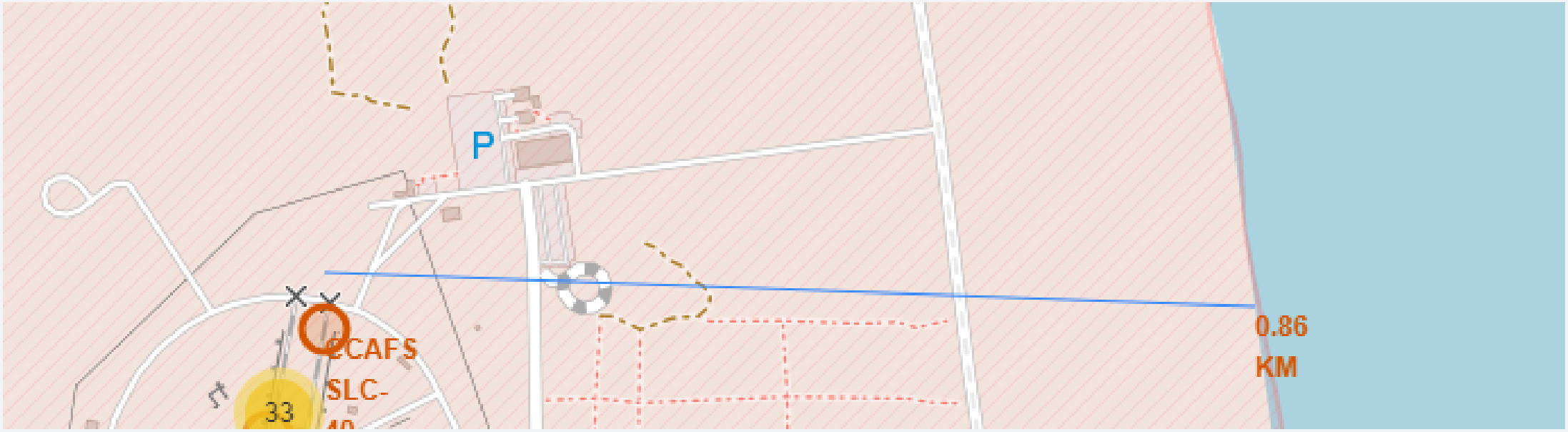
Location of launch sites

Launch outcomes for launch site KSC LCA -39A



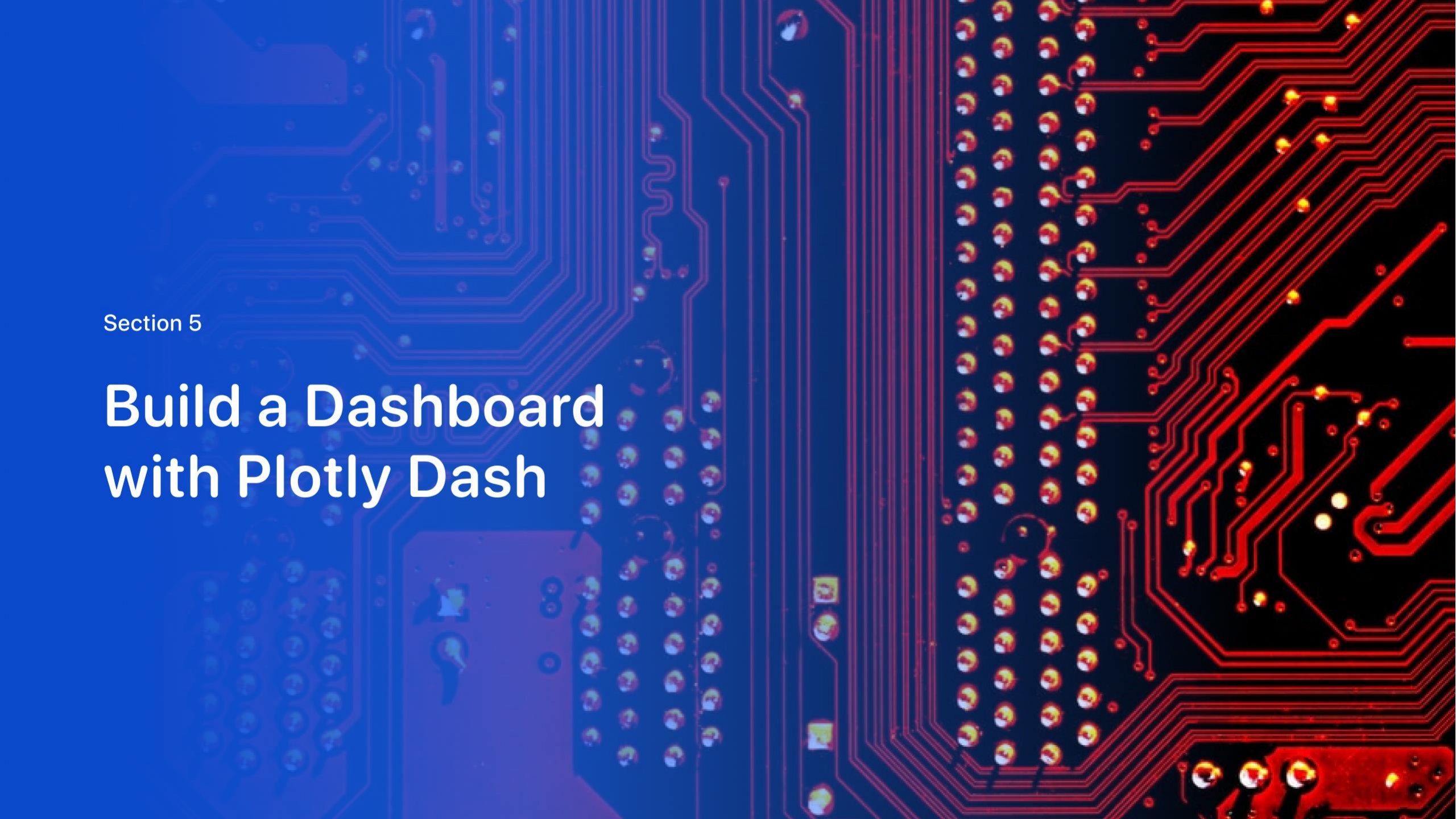
- Green color marks successful launches and red the failed ones

Distance to the closest coastline for CCAFS – SLC 40

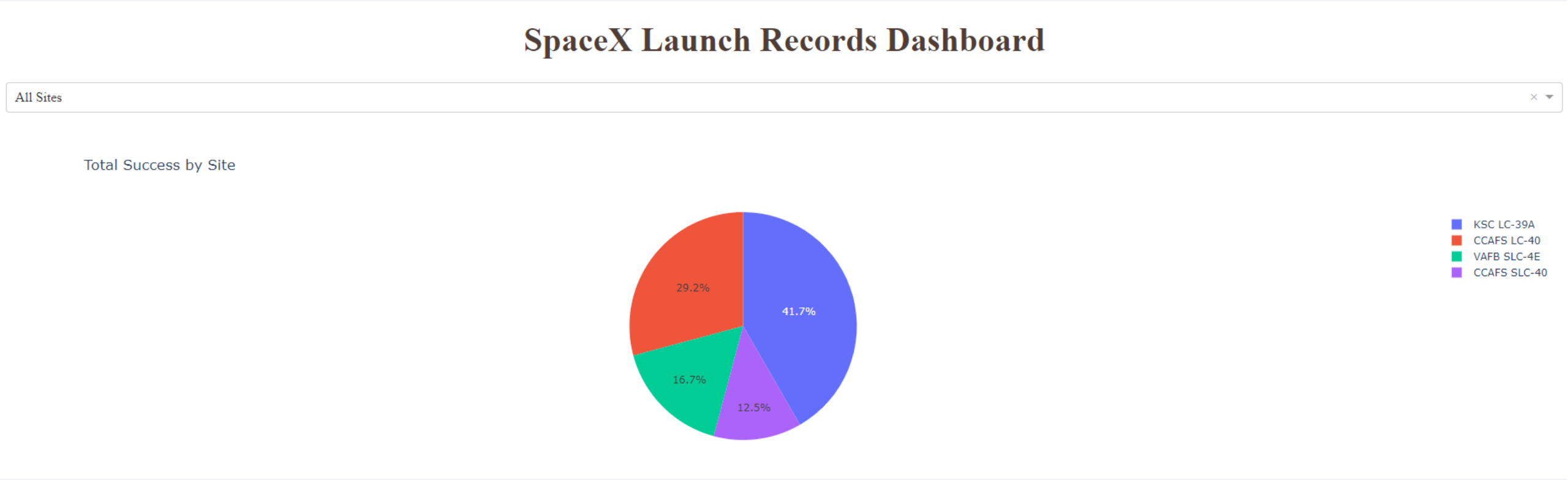


Section 5

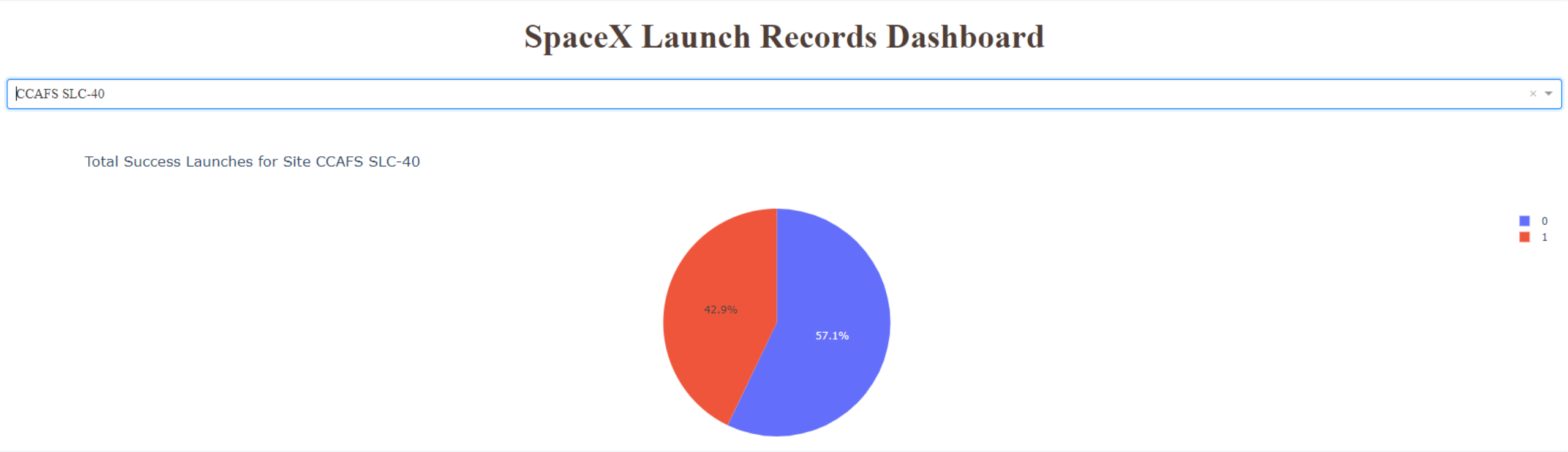
Build a Dashboard with Plotly Dash



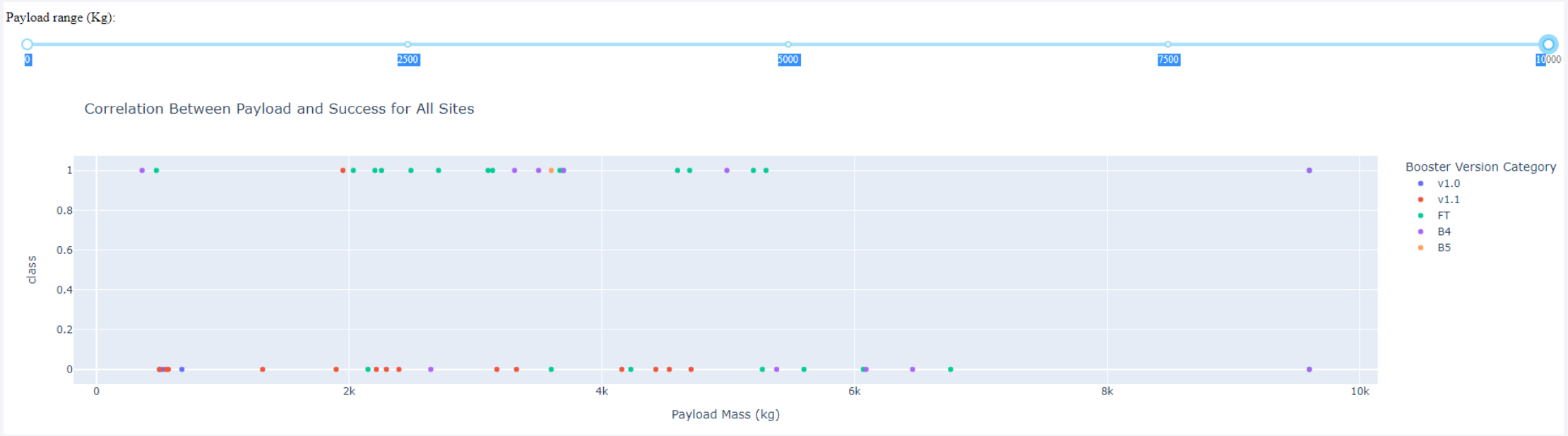
SpaceX success rate for all launch sites

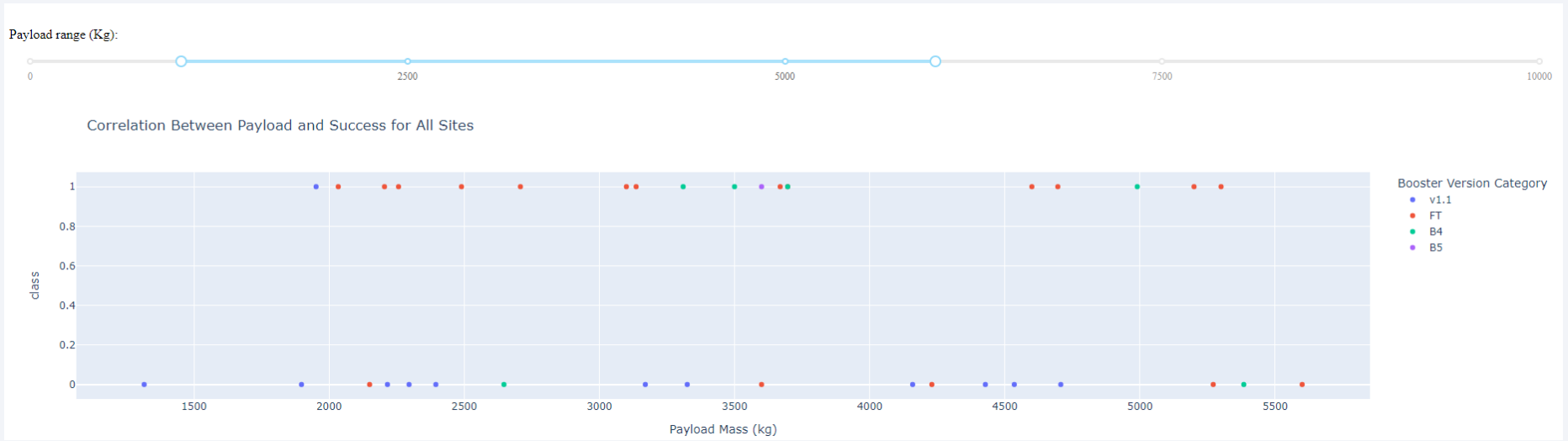


Launch outcome for launch site CCAFS SLC-40



Correlation between Payload and Success for all sites

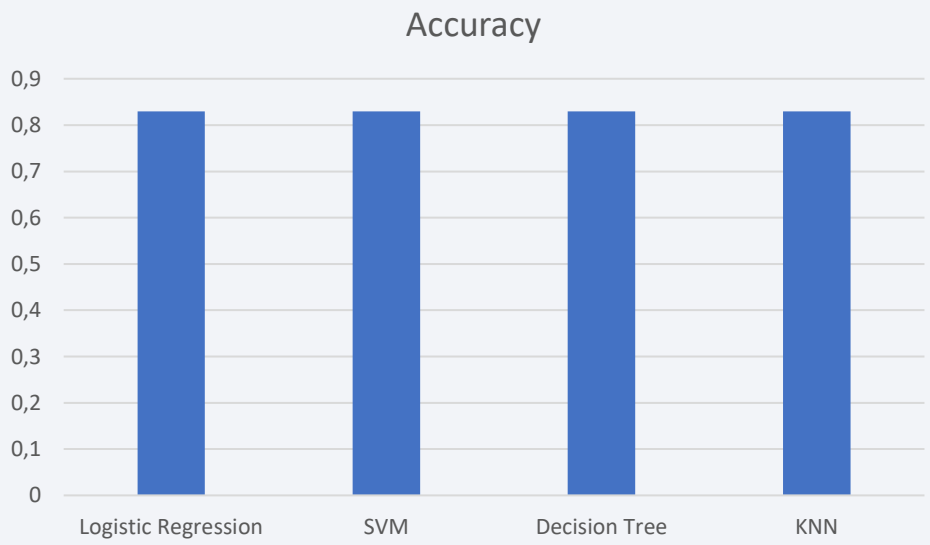




Section 6

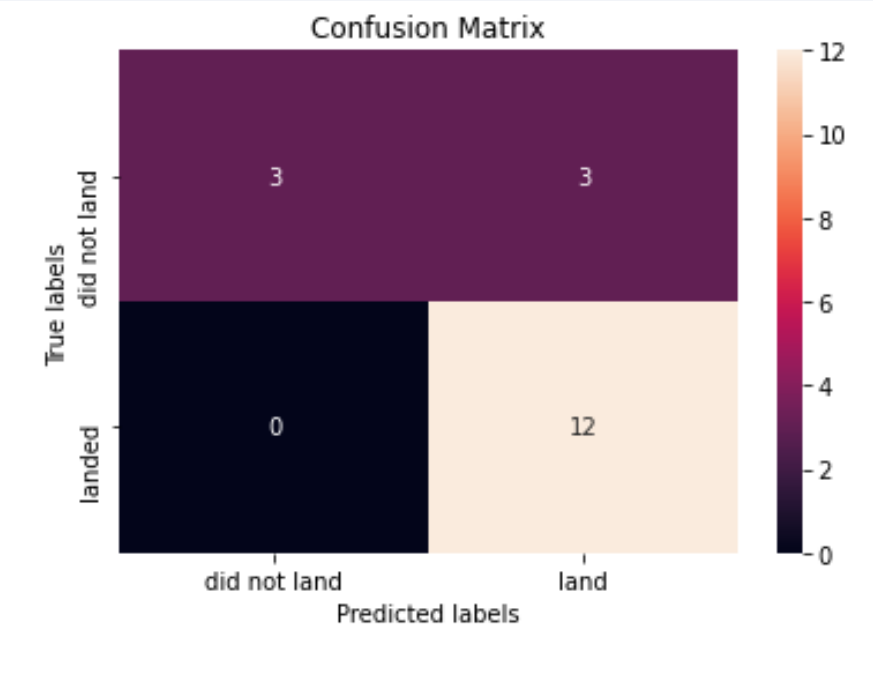
Predictive Analysis (Classification)

Classification Accuracy



- All models achieved the same Accuracy with the test data.

Confusion Matrix for KNN



- Shows 12 True positive and 3 True negative
- 3 false positive and 0 false negative

Conclusions

- I were able to create 4 predictive models with a high accuracy which enable us to predict the landing outcome of a mission depending on the launch site, orbit, payload mass and used booster.
- …

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!