IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Saikat Basak
March 10, 2022

Outline

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

Executive Summary

- Methodologies are used to analyze the data of SpaceX
  - Raw data collected from SpaceX Public API and SpaceX Wiki Page
  - Data processed by data collection API and web scraping.
  - Exploratory Data Analysis (EDA) with data wrangling, SQL, data visualization
  - Interactive Visual Analytics with Folium and interactive dashboards with Ploty Dash
  - Logistic Regression, Support Vector Machine, Decision Tree Classifier, and K Nearest Neighbors machine learning models were used to predict the outcome.
- EDA allowed to identify which features are the best to predict success of launchings.
- Among four machine learning prediction model, decision tree model proves more promising, and accuracy about 94%

Introduction

- The objective is to evaluate the viability of the new company SpaceY to compete with SpaceX.
- Problems to find answers
  - Choosing best features to predict success of launchings
  - Find a best machine learning prediction model to predict successful Stage-1 recovery
  - Find the best way to estimate the total cost for launches, by predicting successful landings of the first stage of rockets.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was obtained from 2 sources: SpaceX Public API and SpaceX Wikipedia page
- Perform data wrangling
  - Collected data was processed by creating a landing outcome label based on outcome data after summarizing and analyzing features
- 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 was normalized, tuned models using GridSearchCV to find the best parameters for each model

Data Collection

- Data collection was involved the following processes
  - API requests from Space X Public API
  - Web Scraping data from a table in Space X's Wikipedia Page

Data Collection – SpaceX API

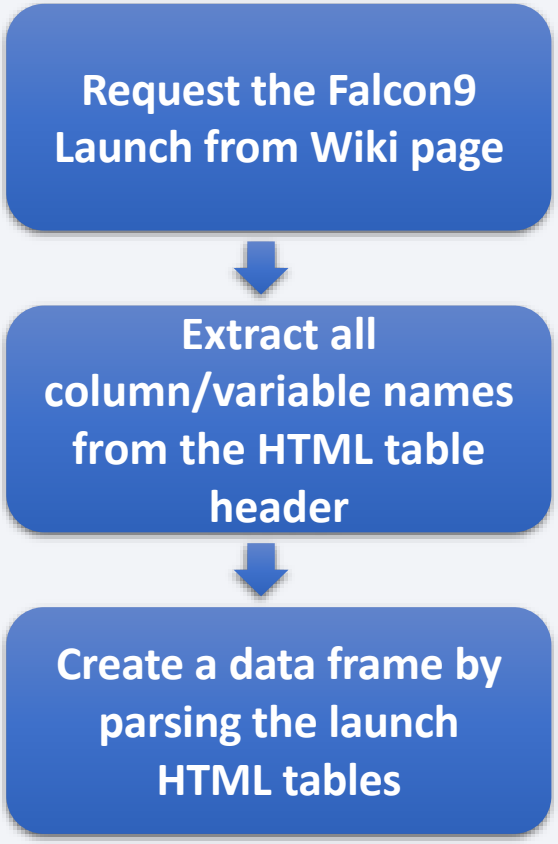
- Data was collected from SpaceX Public API

GitHub URL: Click here

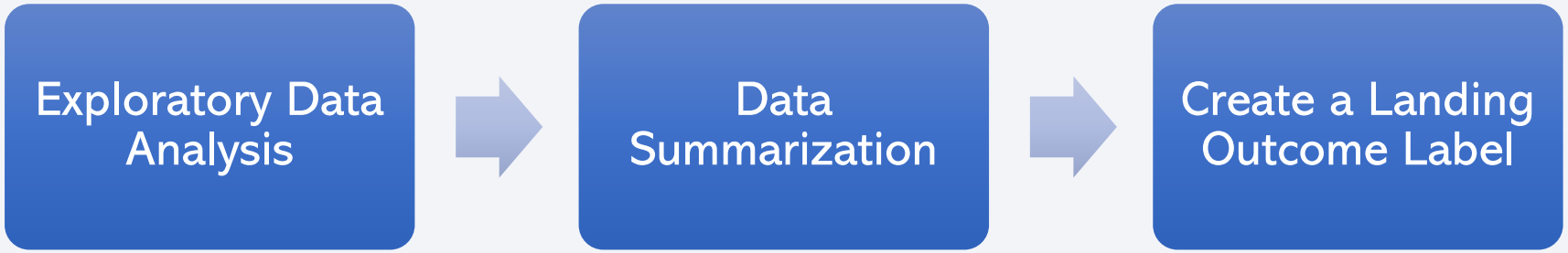
Data Collection – Scraping

- Data was collected from SpaceX's Wikipedia page

GitHub URL: Click here



Data Wrangling



GitHub URL: Click here

EDA with Data Visualization

Scatterplots, barplots and lineplots were used to visualize the relationship between the following features:

- Flight Number and Payload Mass
- Flight Number and Launch Site
- Payload and Launch Site
- Success rate of Each Orbit Type
- Flight Number and Orbit Type
- Payload and Orbit Type
- Launch Success Yearly Trend

GitHub URL: Click here

EDA with SQL

The following SQL queries were performed:

- List the names of the unique launch sites in the space mission
- 5 records where launch sites begin with the string 'CCA'
- Total payload mass carried by boosters launched by NASA (CRS)
- Average payload mass carried by booster version F9 v1.1
- List the date when the first successful landing outcome in ground pad was achieved
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Total number of successful and failure mission outcomes
- List the names of the booster versions which have carried the maximum payload mass
- List the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015
- Rank the count of landing outcomes between the date 2010-06-04 and 2017-03-20

GitHub URL: Click here

Build an Interactive Map with Folium

The following map objects such as markers, circles, lines, marker clusters were added to a folium map:

- Mark all launch sites on a map
- Circles each launch site on the map
- Color-labeled markers in marker clusters to identify which launch sites have relatively high success rates.
- Draw line to display the distances between a launch site to its proximities (such as coastline point, highway)

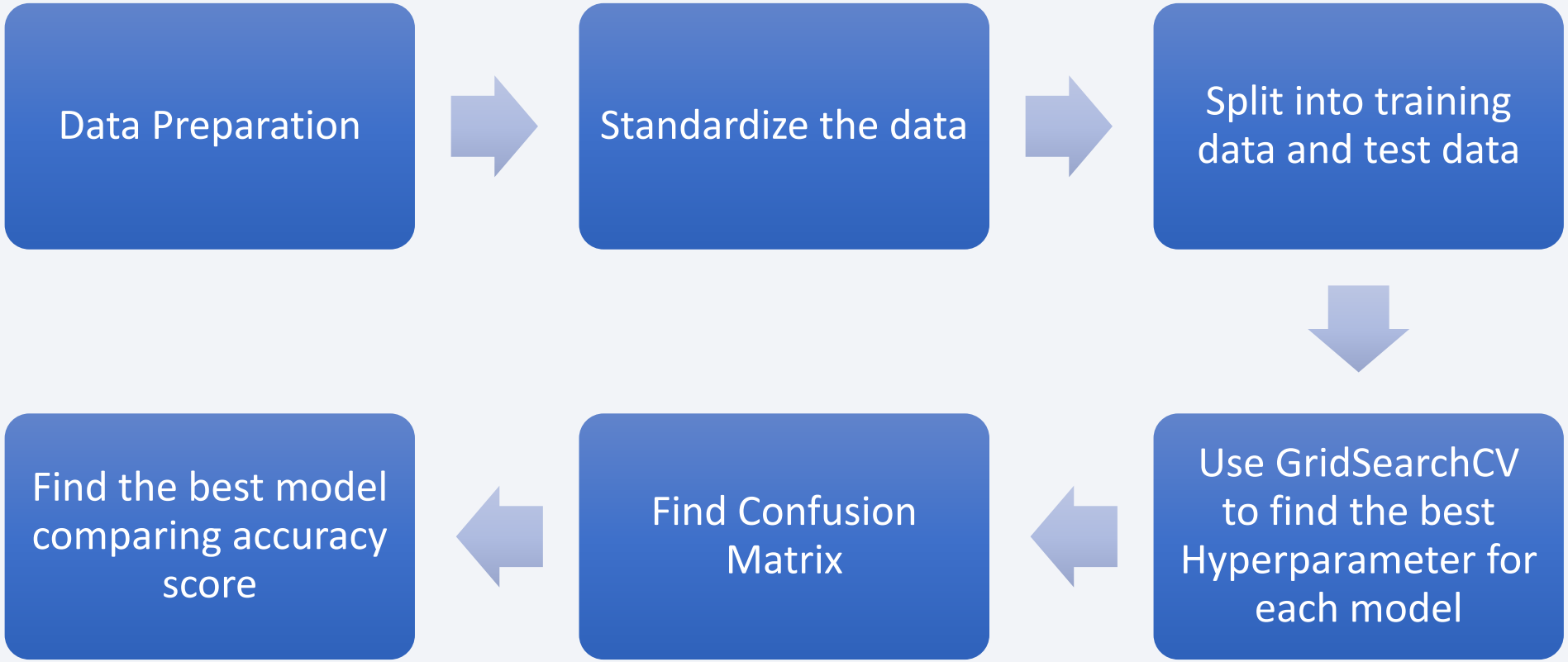
GitHub URL: Click here

Build a Dashboard with Plotly Dash

- Dashboard includes a pie chart and a scatter plot.
  - Pie chart can be selected to show distribution of successful landings across all launch sites and can be selected to show individual launch site success rates.
  - Scatter plot takes two inputs: All sites or individual site and payload mass on a slider between 0 and 10000 kg.
- The pie chart and scatter plot were used to quickly analyze the relation between payloads and launch sites, helping to identify where is best place to launch according to payloads.

Github URL: Click here

Predictive Analysis (Classification)



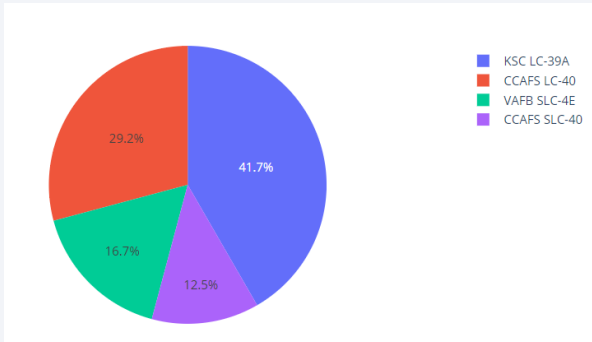
GitHub URL: [Click here]

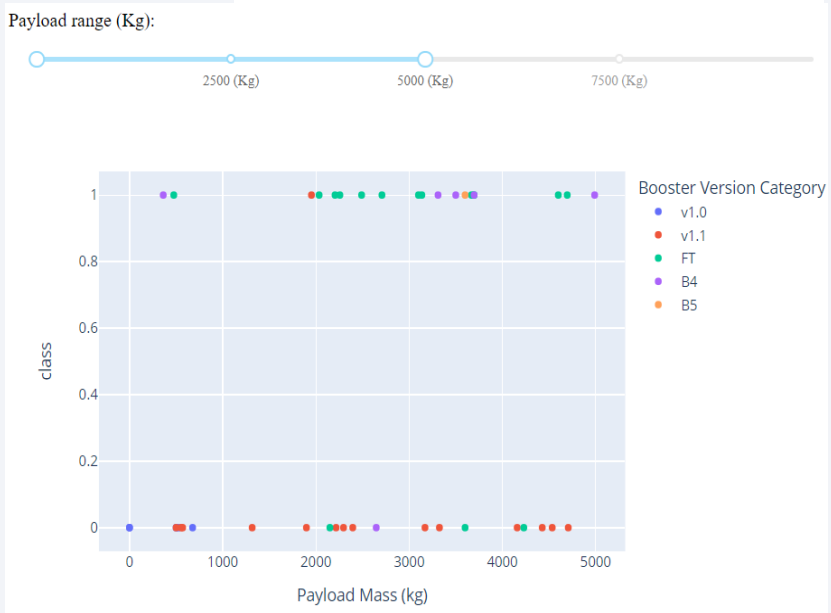
Results

- Exploratory data analysis results
  - SpaceX uses 4 different launch sites.
  - After 5 years of first launch, the first success landing outcome happened in December 2015.
  - The number of landing outcomes became as better as years passed.
  - The average payload of F9 v1.1 booster is 2,534 kg.
  - Many Falcon 9 booster versions were successful at landing in drone ships having payload above the average.
  - Almost 100% of mission outcomes were successful.
  - Two booster versions failed at landing in drone ships in 2015: F9 v1.1 B1012 and F9 v1.1 B1015.

Results

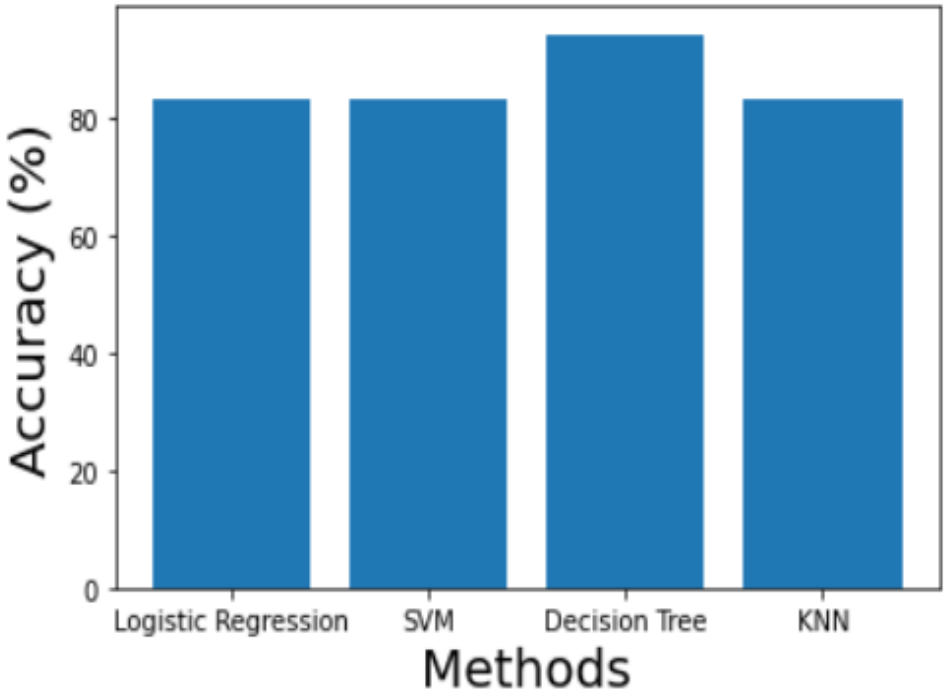
- The locations of all launch sites are near the sea.
- Booster version category “FT” is appeared with highest success outcome.
- KSC LC-39A launch site has maximum success outcome.

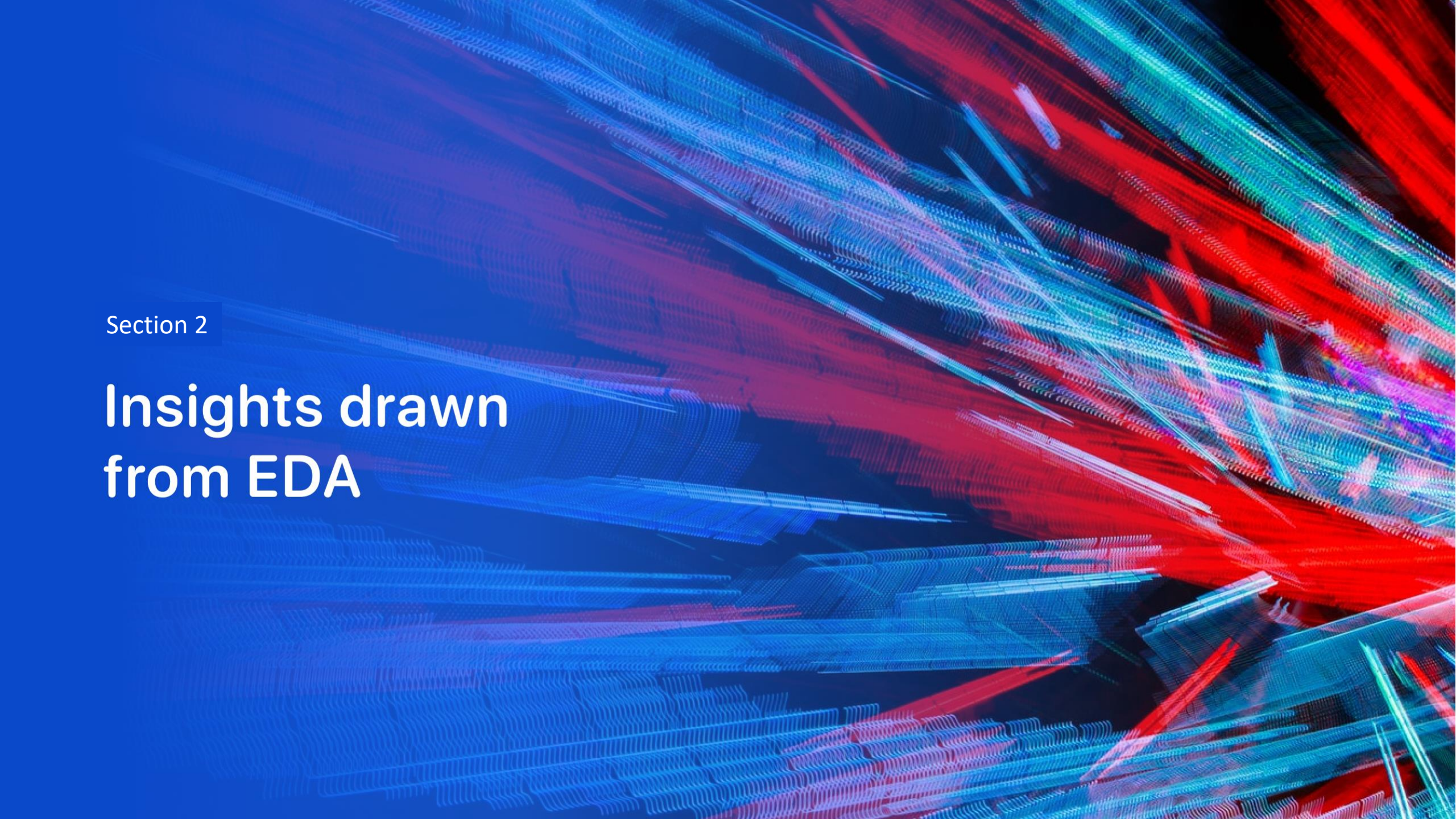




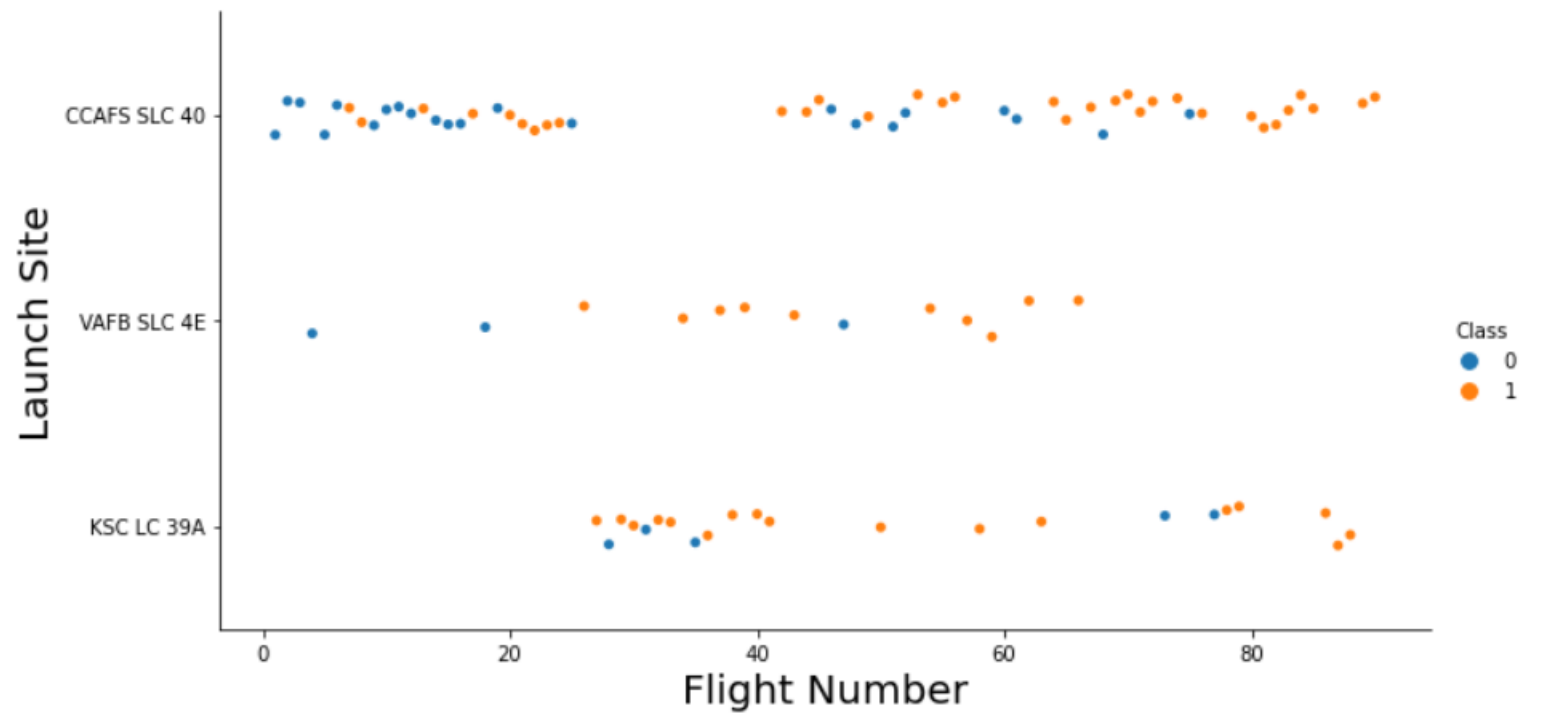
Results

- Predictive analysis results shows “Decision Tree” model has the highest accuracy about 94.44% compared to other three models: Logistic Regression, Support Vector Machine, and K Nearest Neighbors



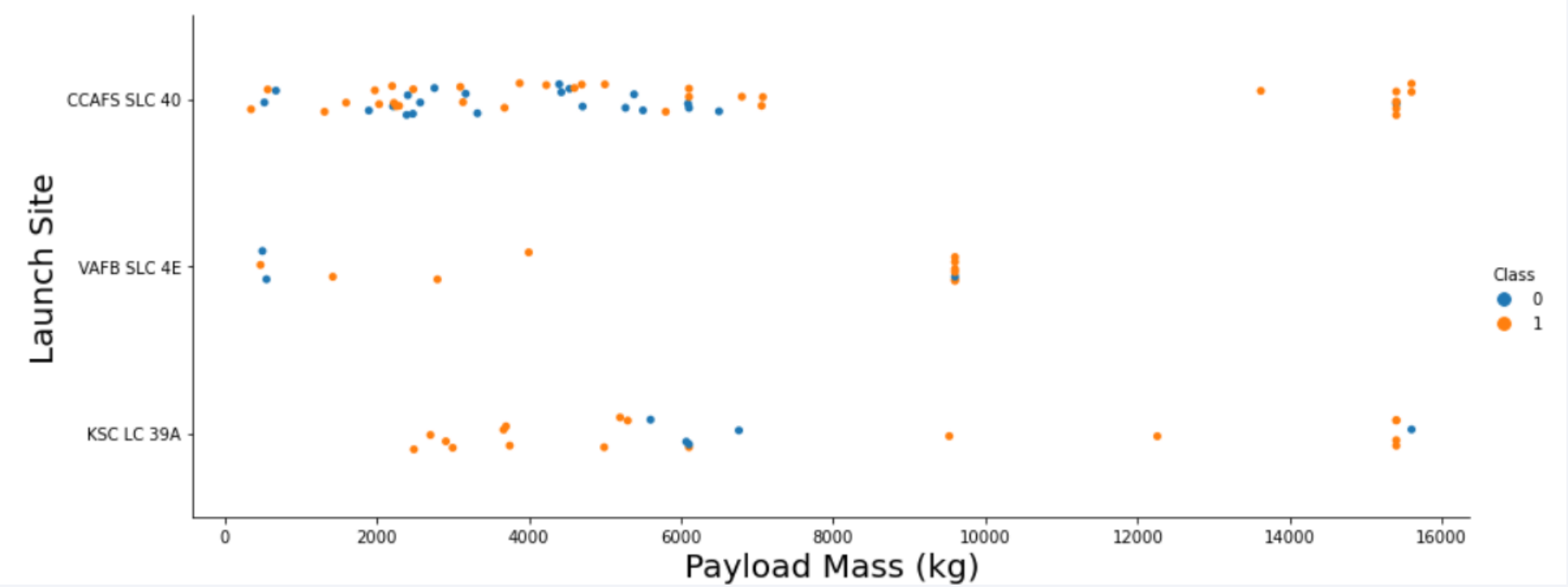

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



- Figure shows an increase in success rate over time (indicated in Flight Number).
- CCAFS SLC 40 appears to be the main launch site as it has the most volume.
- After 20 flights, success rate was increased significantly for all launch sites.

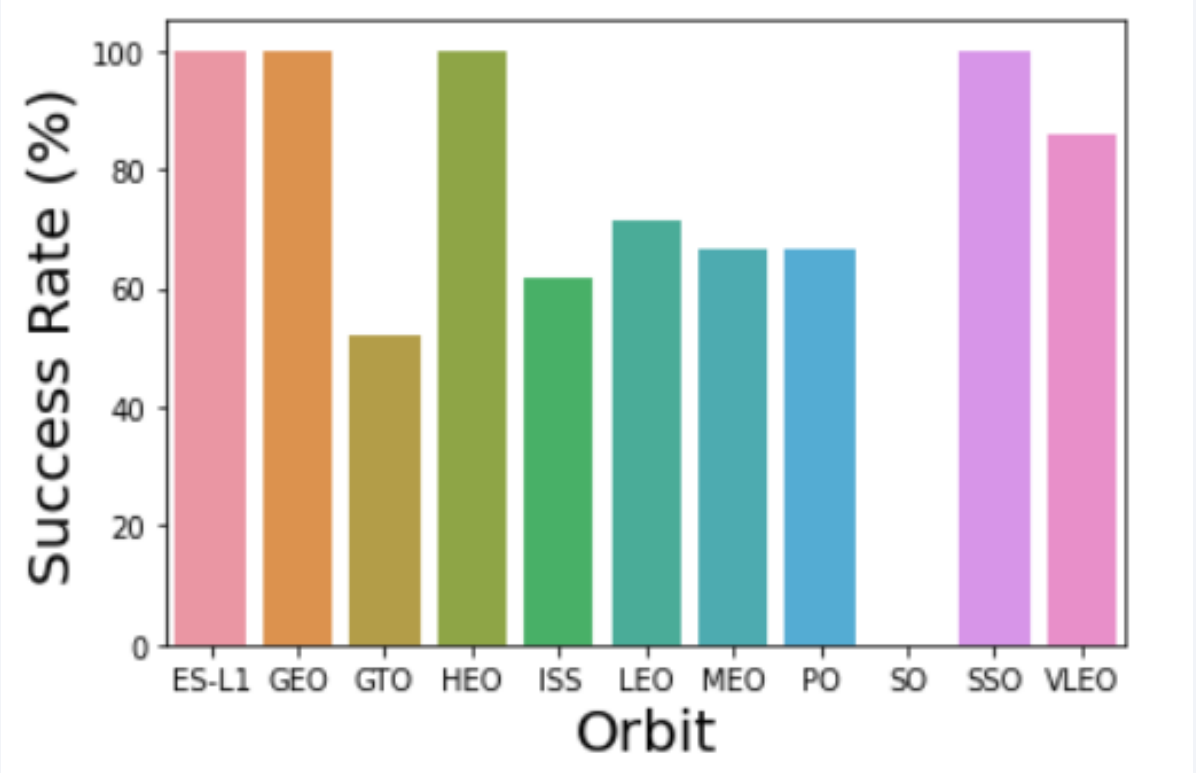
Payload vs. Launch Site



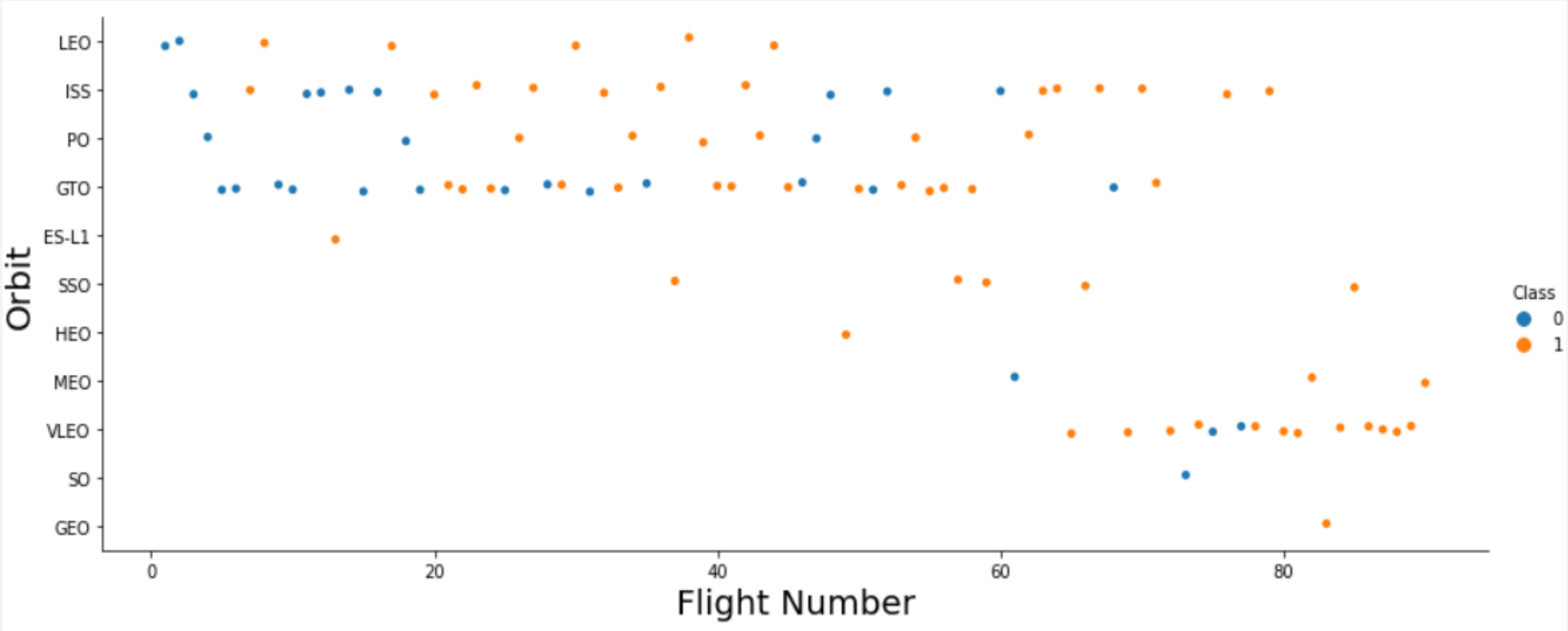
- The success rate is significantly high over payload mass approx. 9,000 kg.
- Most payload mass appears below 8,000 kg.
- Payloads over 12,000 kg seems to be possible only on CCAFS SLC 40 and KSC LC 39A launch sites.

Success Rate vs. Orbit Type

- 100% success rate for the following orbits:
  - ES-L1
  - GEO
  - HEO
  - SSO
- “SO” orbit has no success.

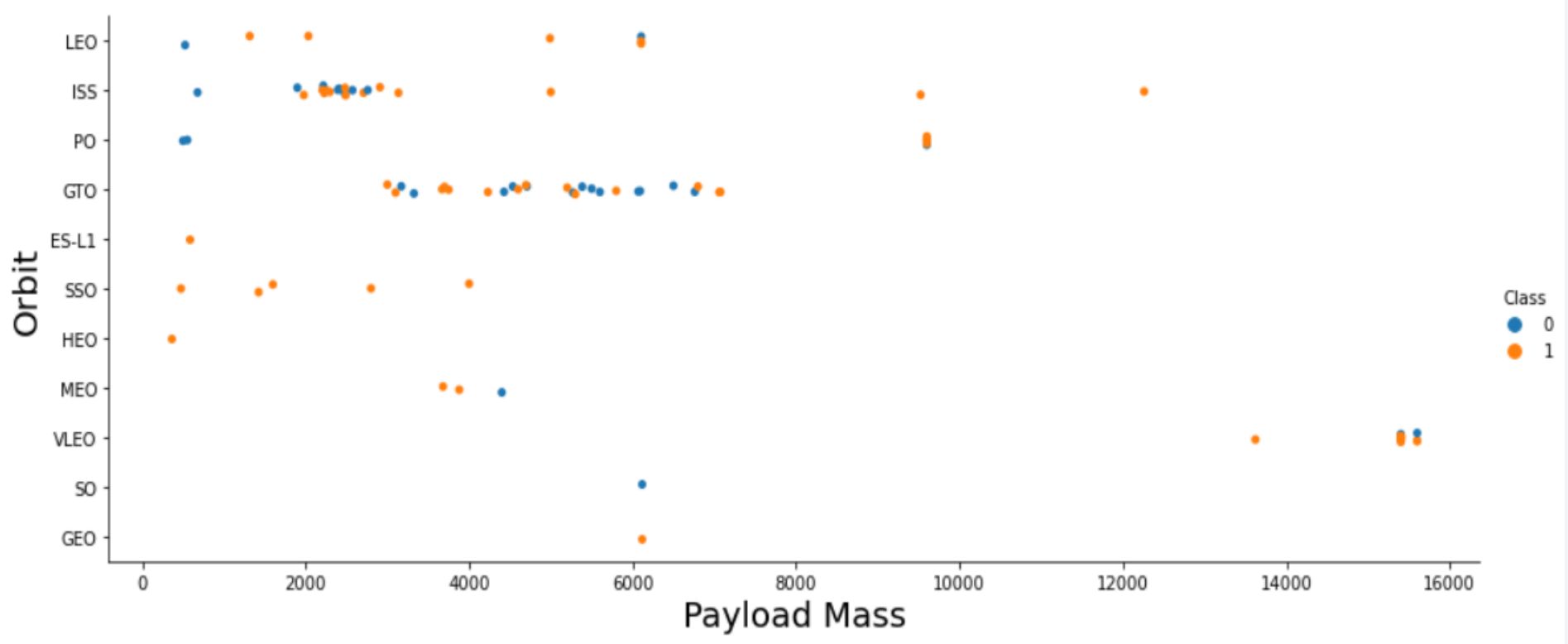


Flight Number vs. Orbit Type



- Orbit preferences was changed over flight number.
- Success rate was improved over time to all orbits, especially flight numbers over 60.

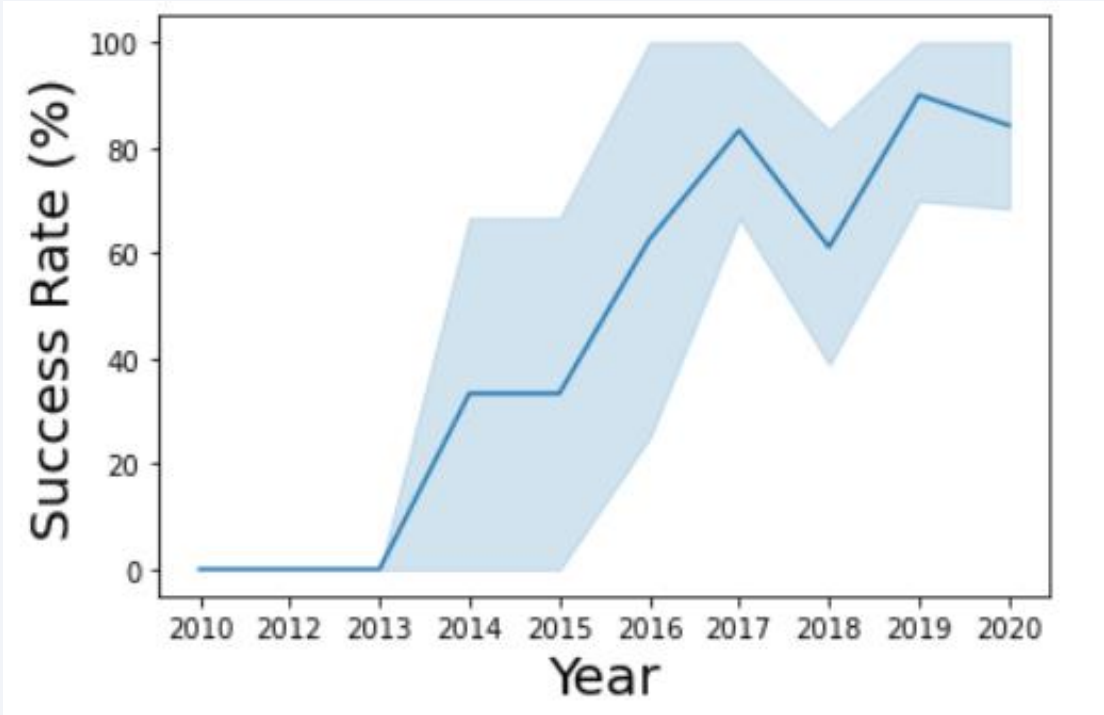
Payload vs. Orbit Type



- GTO, ISS orbits have most of the launches compared to the other orbits
- SO and GEO orbits have fewer launches than other orbits.
- Most of launches are below 8,000 kg payload mass.

Launch Success Yearly Trend

- A trend of increasing success rate is observed between 2013 and 2017.
- No success in 2010, 2011, 2012



All Launch Site Names

- There are four launch sites.

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

- Only unique names are listed for launch sites.

Launch Site Names Begin with 'CCA'

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
2010-06-04	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

- Total payload carried by boosters from NASA

total_payload_mass
45596

- By summing all payloads whose codes contain ‘CRS’, which corresponds to NASA

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1

average_payload_mass
2534

- The average payload mass for booster version F9 v1.1 is found 2,534 kg by filtering data by the booster version

First Successful Ground Landing Date

- Date of the first successful landing outcome on ground pad

DATE
2015-12-22

- The date is found by using “MIN” query and by filtering the data using successful landing outcome on ground pad.

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 greater than 4000 but less than 6000

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

- To find this data, filters are used for payload mass and successfully landed on drone ship.

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes

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

- Total successful mission outcome is 100, and failure is only 1

Boosters Carried Maximum Payload

- Names of the booster which have carried the maximum payload mass

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

- Subquery is used to find the names of the booster versions.

2015 Launch Records

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

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

- In 2015, failed landing outcomes in drone ship appears only in launch site “CCAFS LC-40”.

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

- Count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order

landing__outcome	count_of_landing_outcomes
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

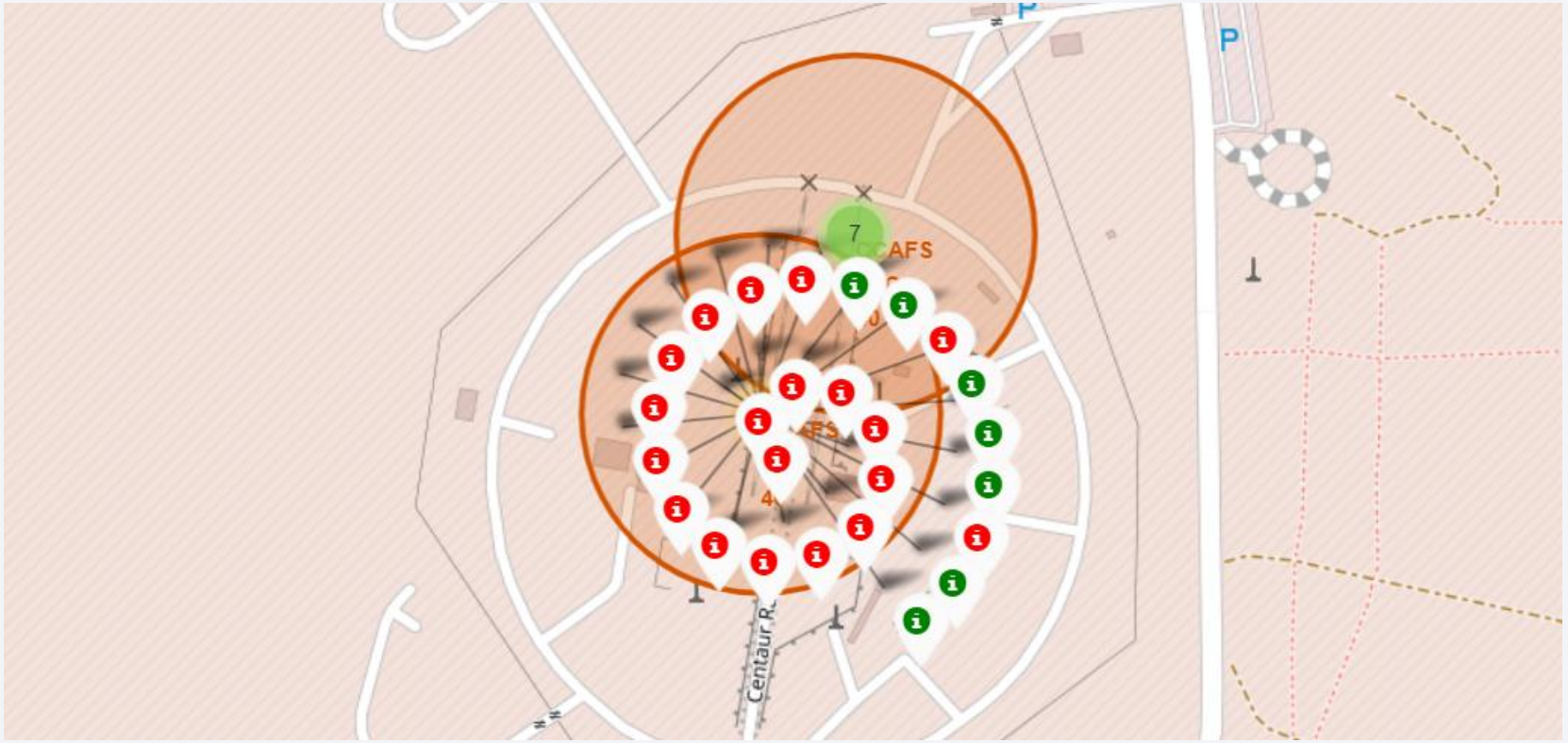
- Landing outcomes are maximum for “No attempt” and minimum for “Precluded (drone ship).


Section 3
Launch Sites
Proximities Analysis

Launch Site Locations

- The locations of all launch sites are near the coast.

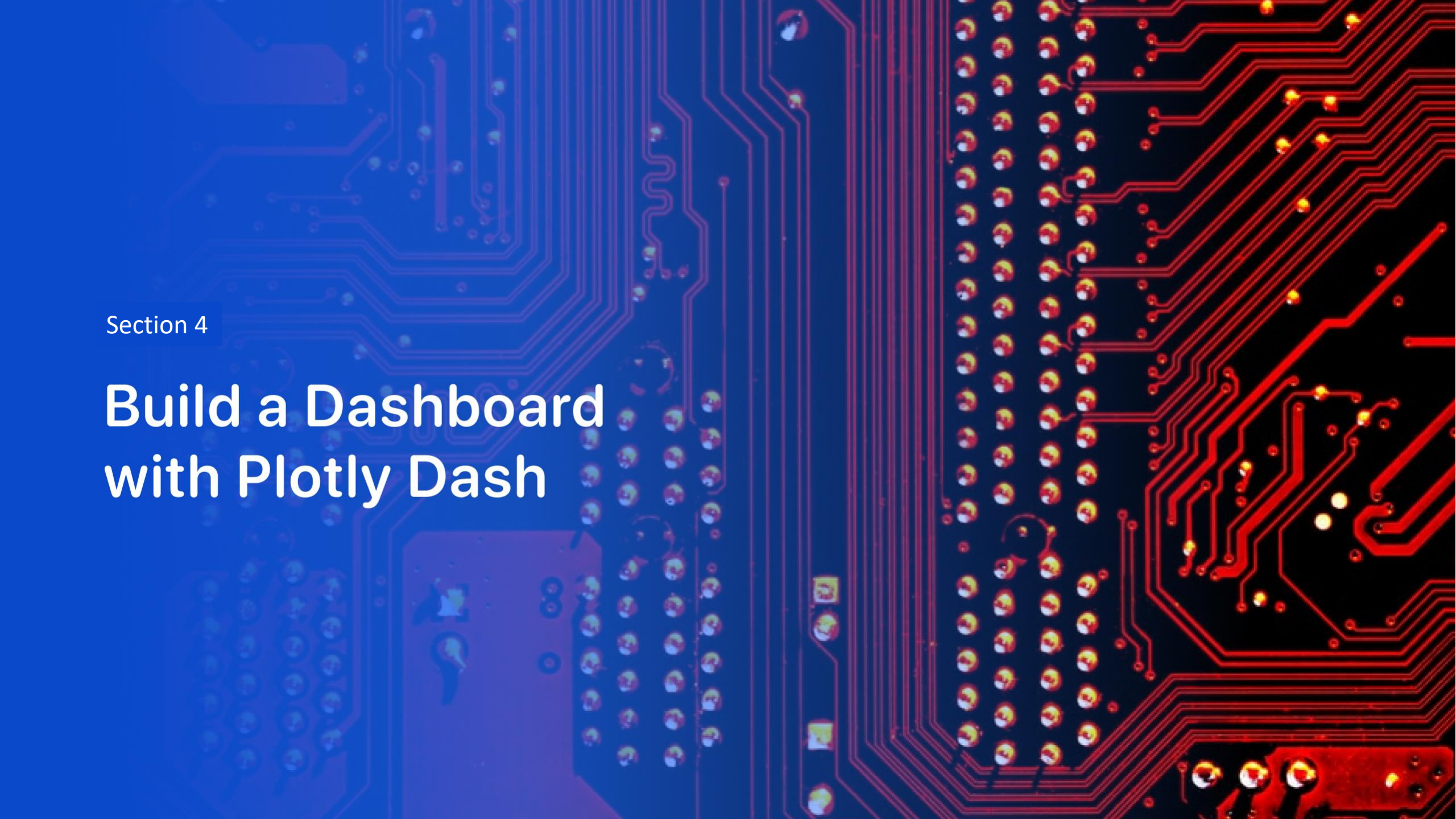
Color-Labeled Launch Outcomes



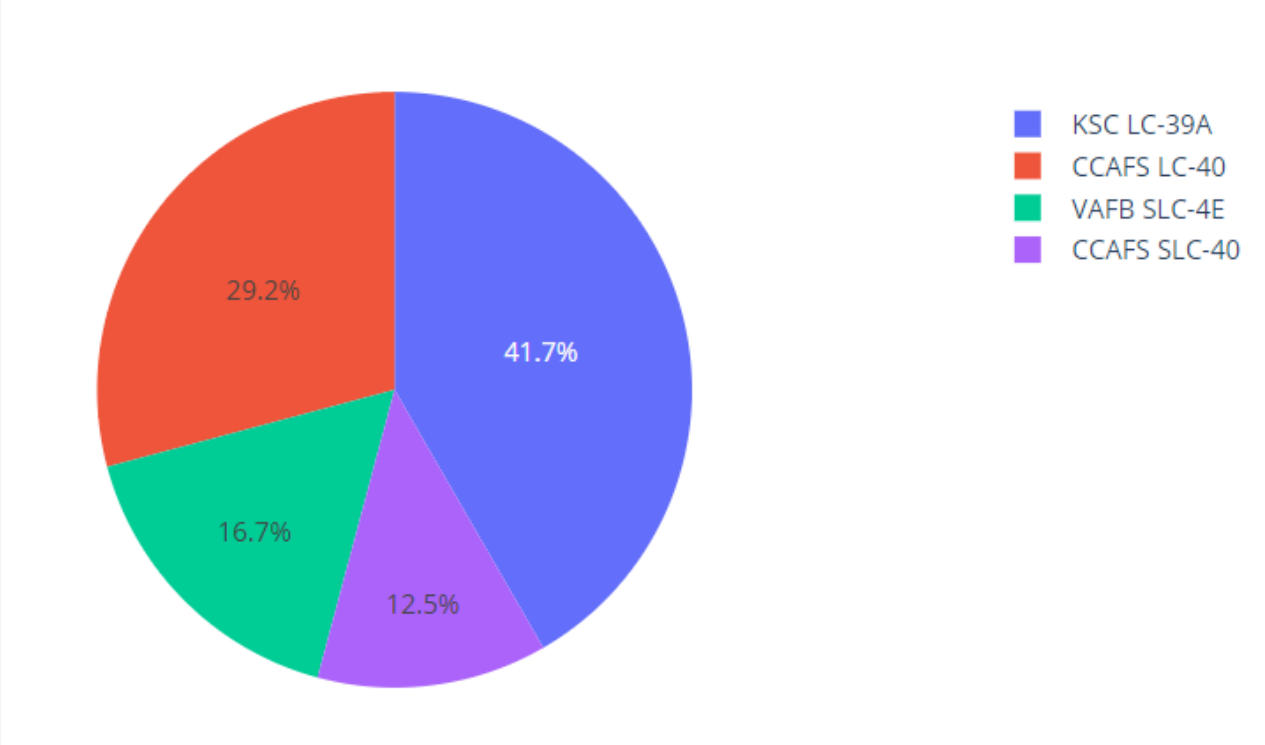
- Green and red maker represents successful and failure outcomes for a launch site respectively.

Distance Measurement from Sites to its Proximities

- In figure, 0.87 KM distance is measured from a launch site to coastline, also 0.59 KM distance from that site to highway.

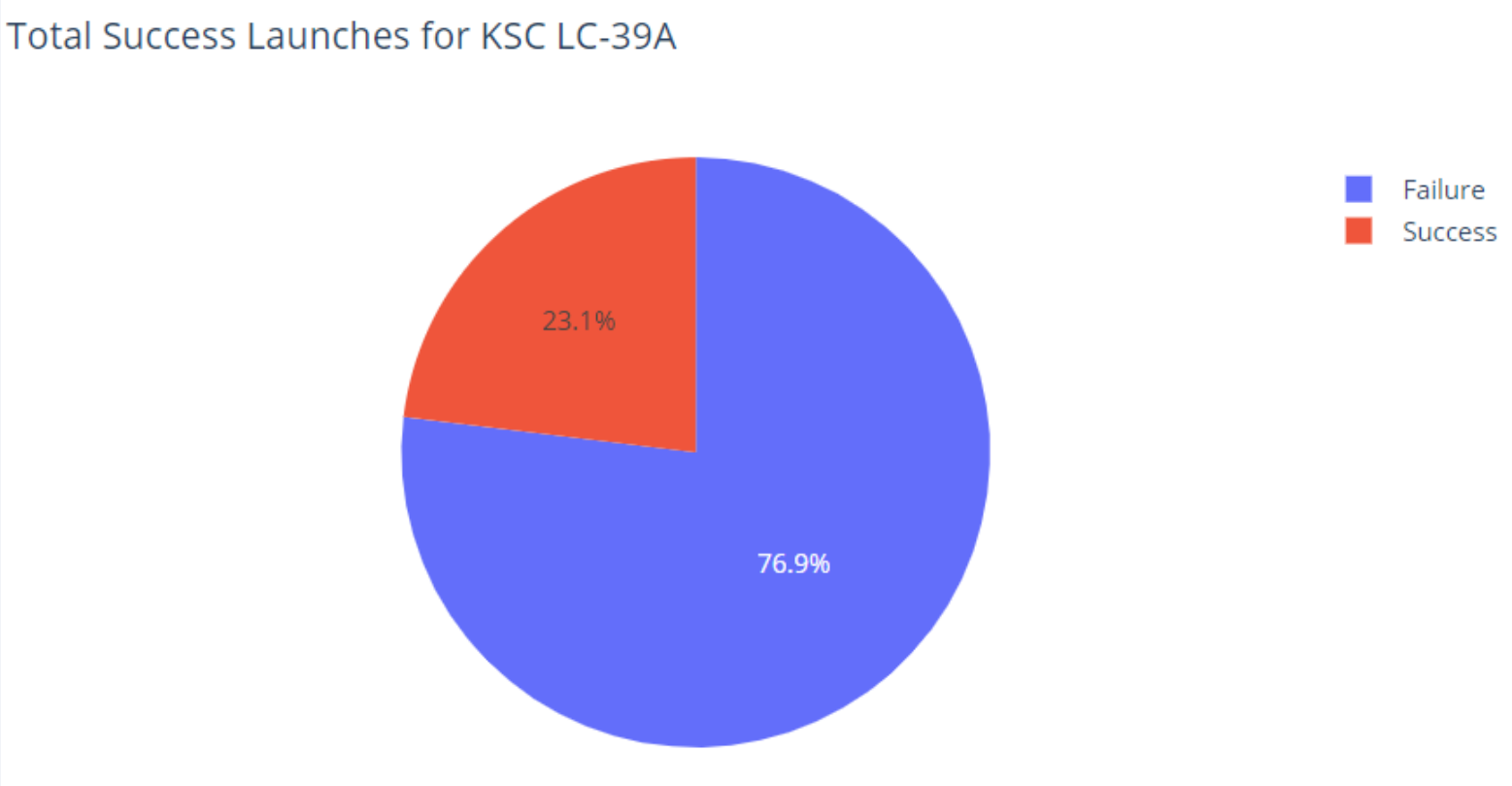

Section 4
Build a Dashboard with Plotly Dash

Launch Success Rate for All Sites



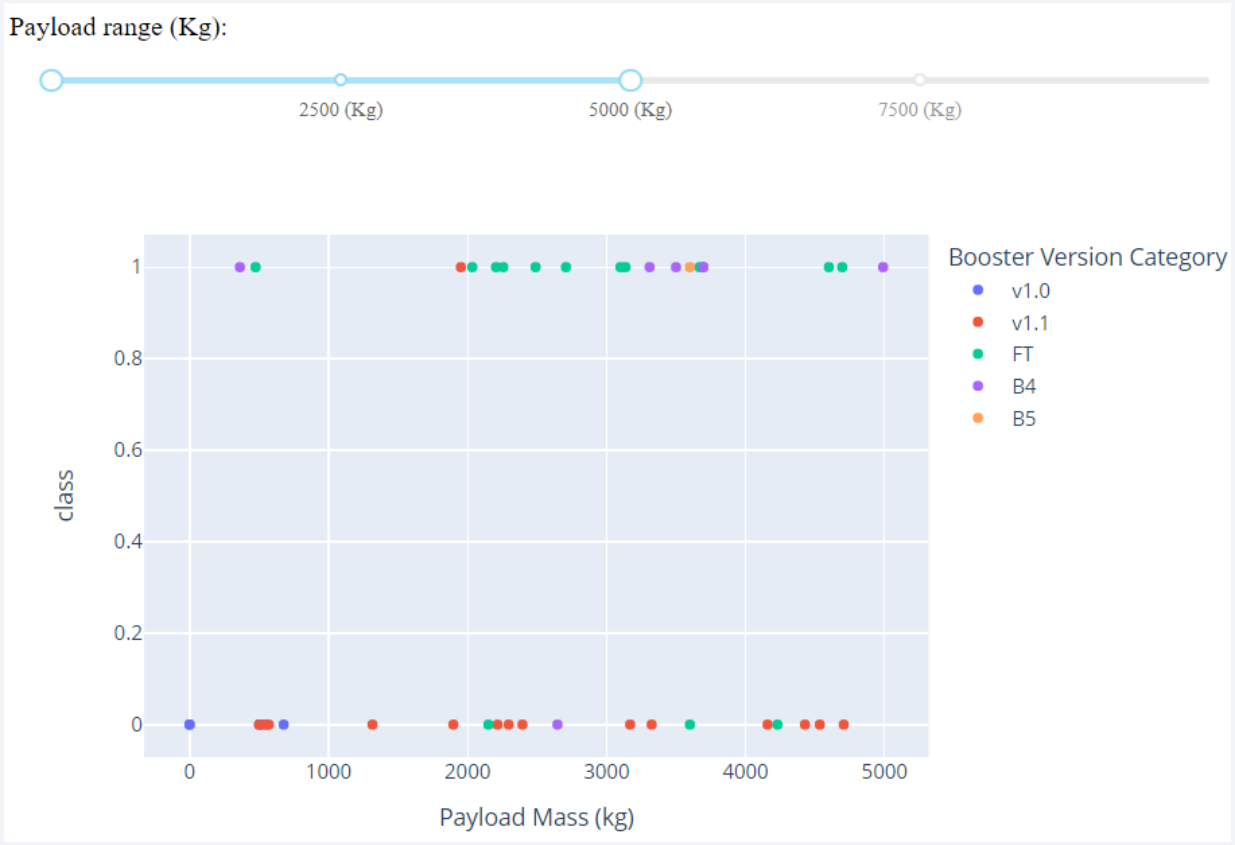
- The launch successful rate is maximum for KSC LC-39A launch site, and minimum for CCAFS SLC-40

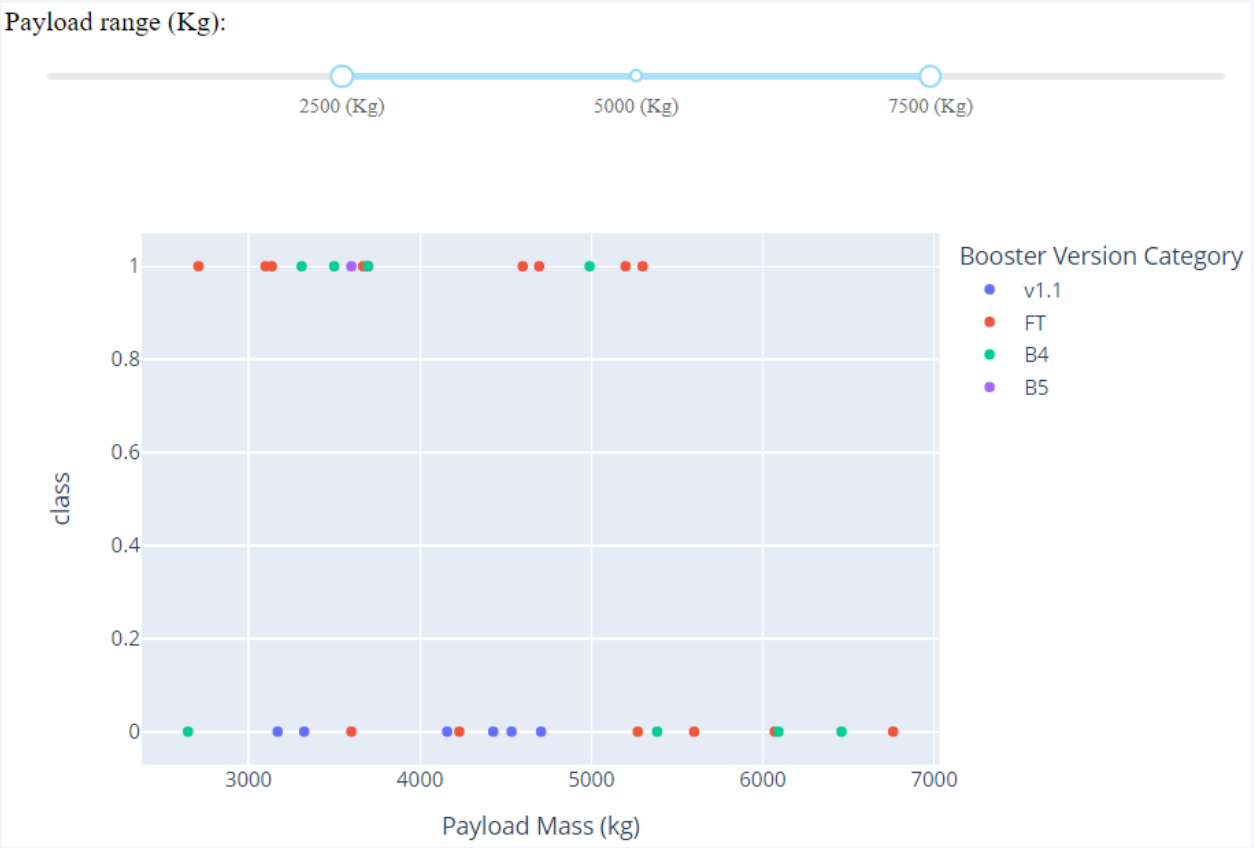
Highest Launch Success Site



- KSC LC-39A launch site has the highest launch success rate.
- For this site, the success rate is 76.9%

Payload vs. Launch Outcome





- Above 2 screenshots have different payload ranges. Left figure has payload range from 0 kg to 2,500 kg, and right one from 2,500 kg to 7,500 kg.
- In both cases, booster version category “FT” is appeared with highest success outcome.

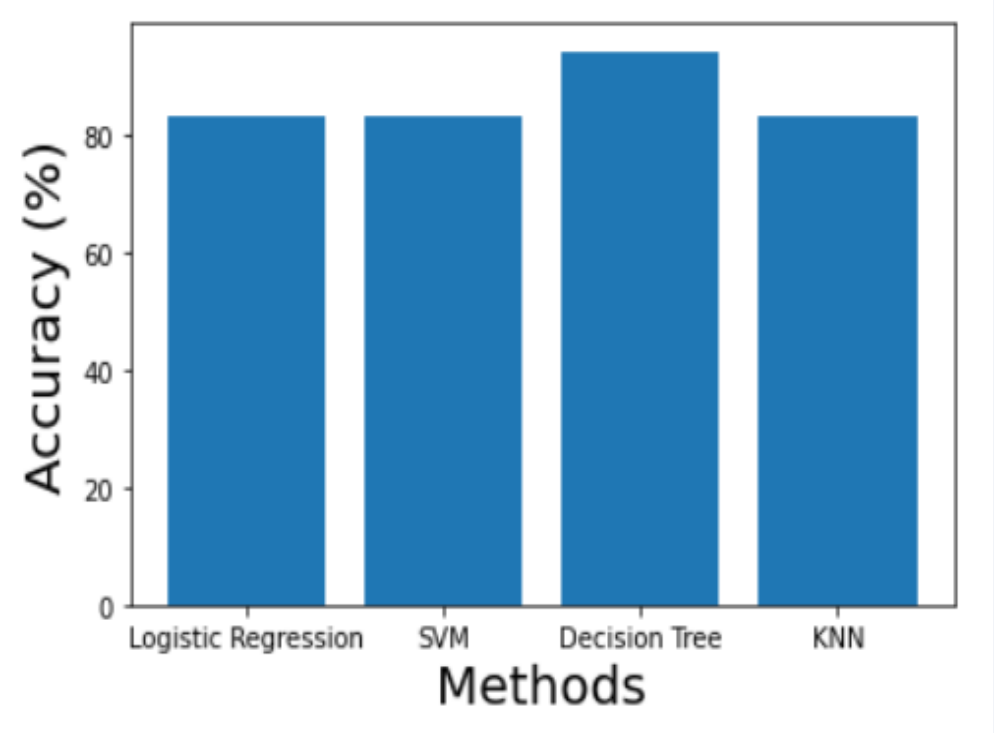
Section 5

Predictive Analysis (Classification)

Classification Accuracy

- The accuracy for four different classification models are plotted in a barplot.

- “Decision Tree” model has the highest classification accuracy about 94.44%.



Confusion Matrix

- The confusion matrix for highest classification accuracy “Decision Tree” model is shown in the figure.
- The accuracy for this model is about 94.44%.
- There is only 1 type-II error, and no type-I error.

Conclusions

- All launch sites are near the sea coast.
- The most successful launch site is KSC LC-39A.
- Booster version category “FT” has highest success outcome.
- Successful landing outcomes seem to improve over time.
- “Decision Tree Classifier” model can be used to predict successful outcomes.

Appendix

- Relevant assets like Python code, SQL queries, charts, notebook outputs that was created and analyzed during this project can be found in Github.

  Github Repository URL: Click here

Thank you!