IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Divyanshu Raj
5 July 2023

Outline

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

Executive Summary

- The taking after techniques were utilized to analyze data::
  - Data Collection utilizing web scratching and SpaceX API,
  - Exploratory Data Analysis (EDA), including data wrangling, data visualization and interactive visual analytics;
  - Machine Learning Prediction.
- Summary of all results was conceivable to collected profitable information from open sources;
  - EDA permitted to distinguish which highlights are the most excellent to predict success of launchings
  - Machine Learning Forecast appeared the leading demonstrate to anticipate which characteristics are important to drive this opportunity by perfect way">the most perfect, way utilizing all collected information.

Introduction

- The **goal** is to evaluate the viability of the new Space Y **company** to compete with Space X.
- Desirable answers:
  - Best way to predict successful landings to estimate total launch cost The first stage of the rocket;
  - Where is the best place to **fire?**

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data from Space X was obtained from 2 sources:
    - Space X API (https://api.spacexdata.com/v4/rockets/)
    - WebScraping (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)
- Perform data wrangling
  - The Collected data was enriched by creating a landing outcome label based on outcome data after summarizing and analyzing features
- Perform exploratory data analysis (EDA) using visualization and SQL

Methodology

Executive Summary

- Perform interactive visual **analysis** using Folium and Plotly **Dash.**
- Perform predictive **analytics** using classification models
  - The data collected up to this stage were normalized, divided into training and test data sets, and evaluated with four classification models, the accuracy of each model was evaluated using different parameter combinations.

Data Collection

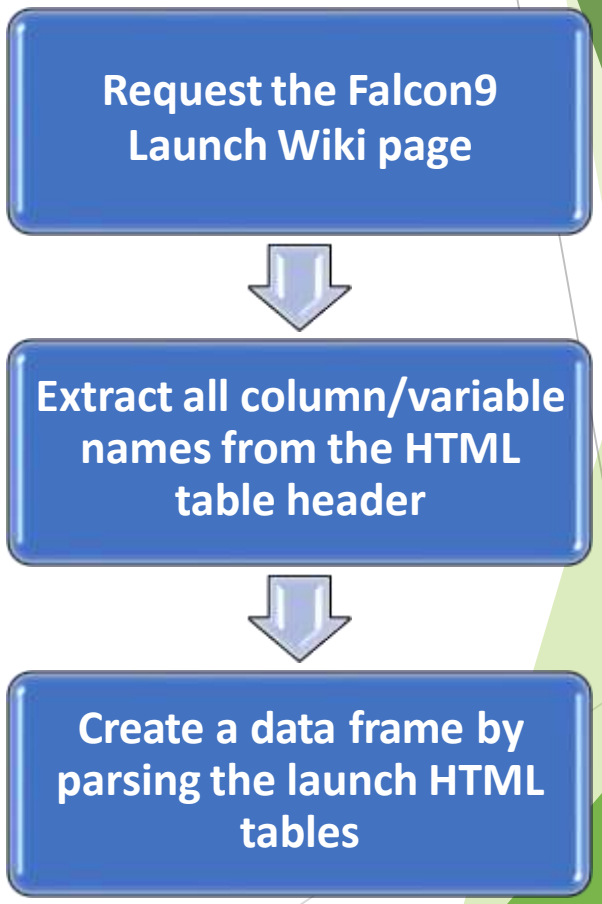
- Data sets were collected from Space X API (https://api.spacexdata.com/v4/rockets/) and from Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches), using web scraping technics.

Data Collection – SpaceX API

- SpaceX **provides** a public API **for retrieving and usin g data.**
- I used this API and saved data according to the flow chart below.

Data Collection - Scraping

- SpaceX **launch data is** als o **available on Wikipedia**
- **Follow the flowchart to lo ad data** from Wikipedia and then **save it.**.

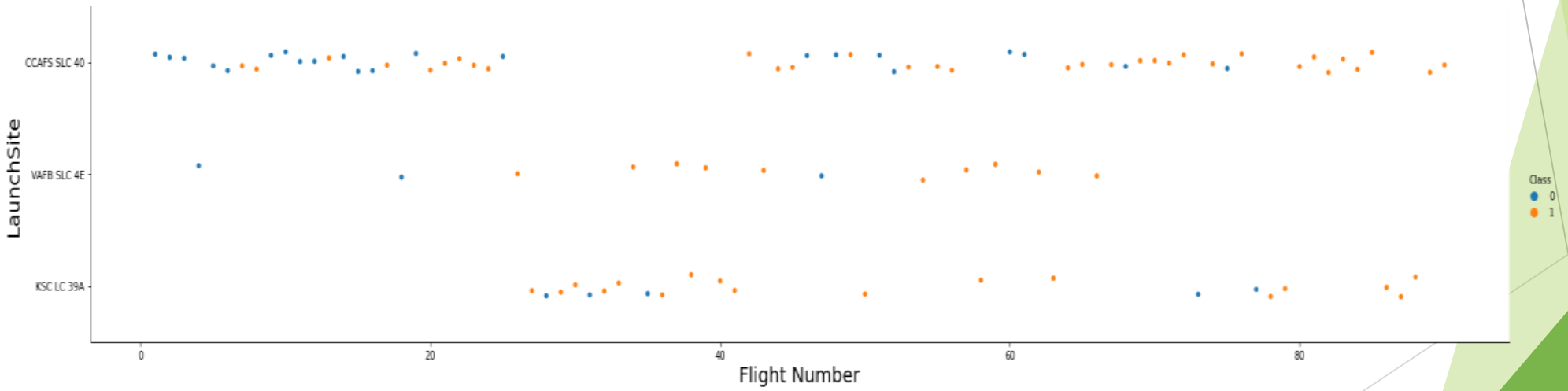


Data Wrangling

- **Initially,** some **exploratory data analysis** (EDA) was performed on the **data set.**
- **Then, a site-by-site summary** of **launches, what** each **trajectory looks like,** and **Mission** outcome **cases were counted by** orbit **type.**
- Finally, **a Landing Result** label was created **in the Result column.**

EDA with Data Visualization

- **Scatterplots** and **histograms** were used **to explore the data** to visualize the relationship between **a** pair of **features**:
  - Payload Mass X Flight Number, Launch Site X Flight Number, Launch Site X Payload Mass, **Track** and Flight Number, Payload and **Track**



EDA with SQL

- The following SQL **query was executed**:
  - **The name** of the unique **space mission launch pad,**
  - Top 5 launch sites whose **names start** with **"CCA“,**
  - Total payload mass carried by **a NASA-launched projectile (CRS)**;
  - Average payload mass **of the launch vehicle** version F9 v1.1;
  - Date when the first successful landing outcome in ground pad was achieved;
  - The name of a projectile successfully used for unmanned vehicles and with a payload mass of 4000-6000 kg;
  - The total number of success and failure outcomes of the mission ;
  - Names of the booster versions which have carried the maximum payload mass;
  - Landing failure result of drone ship, booster version and launcher name 2015 and,
  - Ranking of the number of landing results, e.g. "failure" (drone) or "success" (land pad)) between the dates 2010-06-04 and 2017-03-20.

Build an Interactive Map with Folium

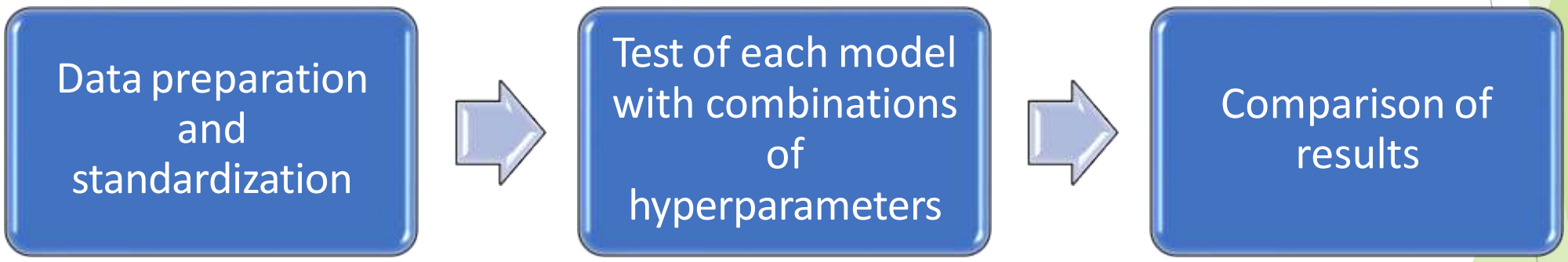
- **Folium maps used markers,** circles, **lines,** and marker **clusters.**
  - Markers indicate points **such as** launch **pads.**
  - Circles **represent** highlighted areas around specific coordinates, **such as** NASA Johnson **Space center.**
  - **A cluster of markers represents a group** of events **at** each coordinate, **such as** a launch **pad.**
  - **A line is** used to indicate **the distance** between two coordinates.

Build a Dashboard with Plotly Dash

- **I used the** following **charts and** graphs to visualize **the data**
  - Site Launch rate
  - Payload range
- This combination allowed **us** to quickly analyze the **relationship** between **the payload** and **the** launch **site,** helpin g to **determine the** best place to launch **based on the payload.**

Predictive Analysis (Classification)

- Four classification models **(logistic** regression, support vector machine, decision **tree,** and k nearest **neighbours) were compared.**



Results

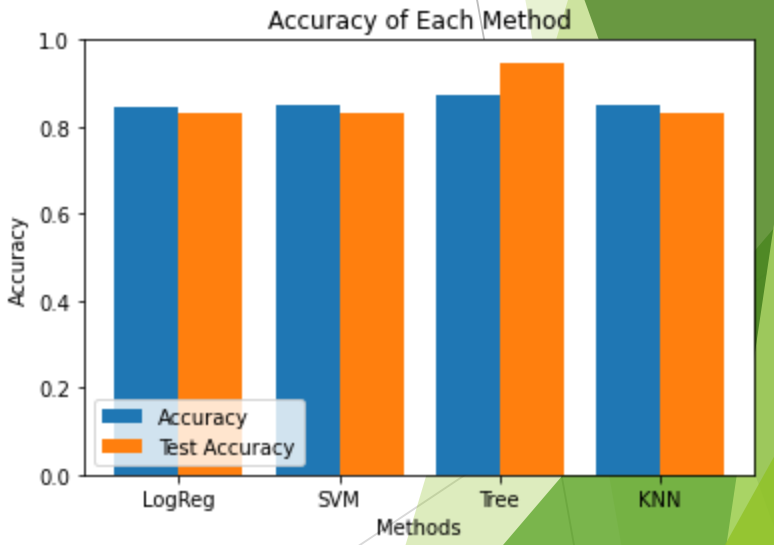
- Exploratory data analysis results:
  - Space X uses four different launch pads;
  - The first launches were made for Space X itself and NASA;
  - The average payload of F9 v1.1 projectile is 2,928 kg;
  - The first successful landing took place in 2015, five years after the first launch;
  - Many versions of Falcon9 launch vehicles have successfully landed drones with above the average;
  - Almost 100% of mission results were successful;
  - **In 2015, two launch vehicle versions,** F9 v1.1 B1012 and F9 v1.1 **B1015, failed to land on drone ships**;
  - The number of **planting results has gotten** better **over the years.**

Results

- Using interactive analysis, we were able to determine which launch pads are generally where. A safe place, for example the sea**,** and good logistical infrastructure around it.
- Most **of the** launches **will be from the East Coast** launch **pad.**

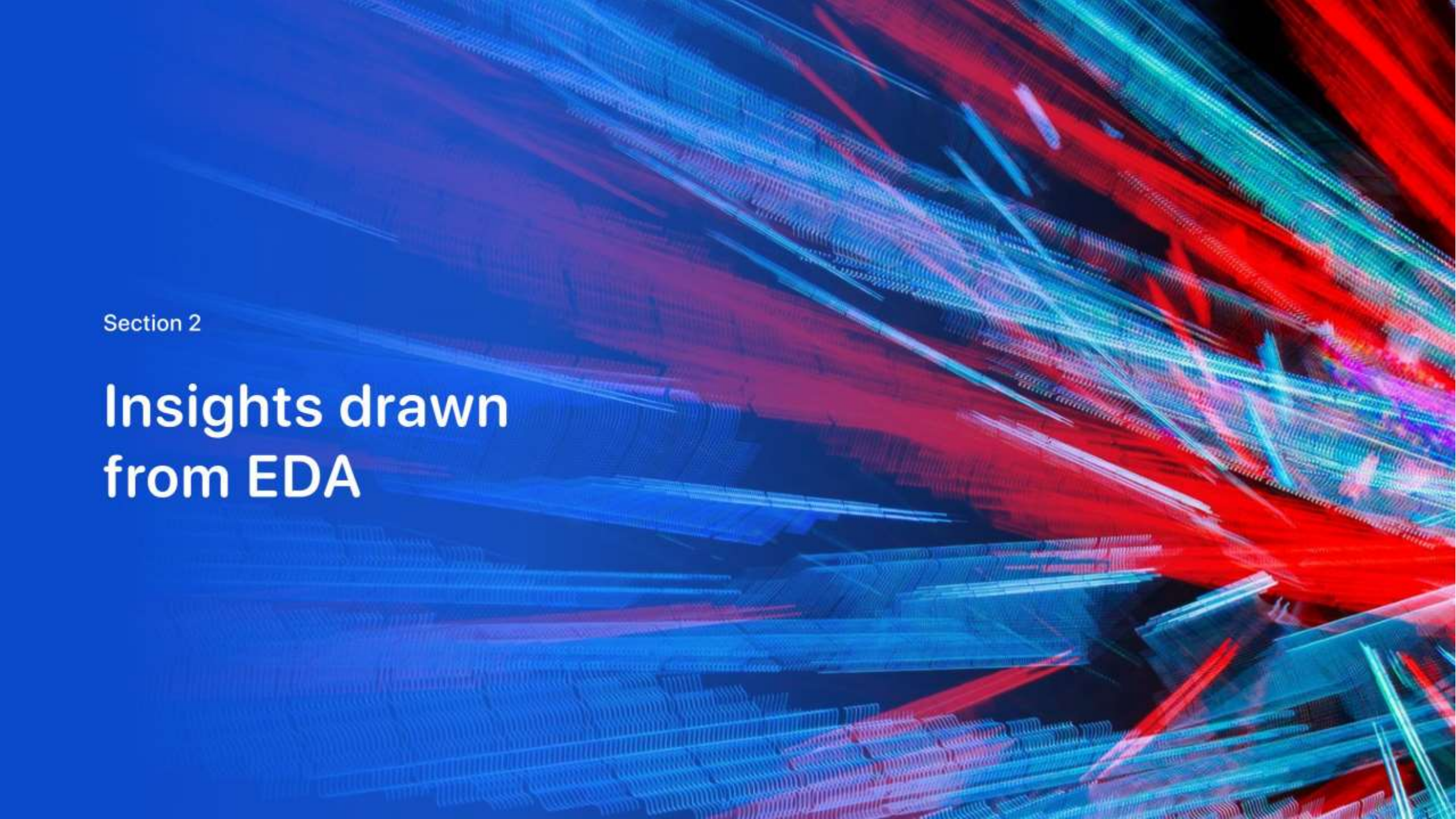
Results

- Predictive **analysis results** showed that **the decision tree classifier** is the best model **for** successful **landing prediction with an** accuracy **of** over 87% and **an** accuracy **of over 94% on the** test **data.**

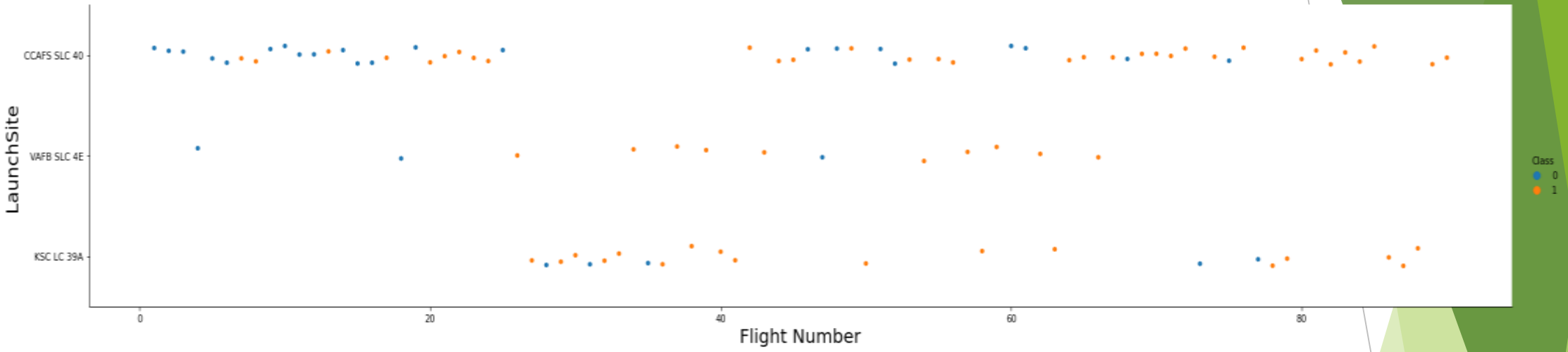


Section 2

Insights drawn from EDA

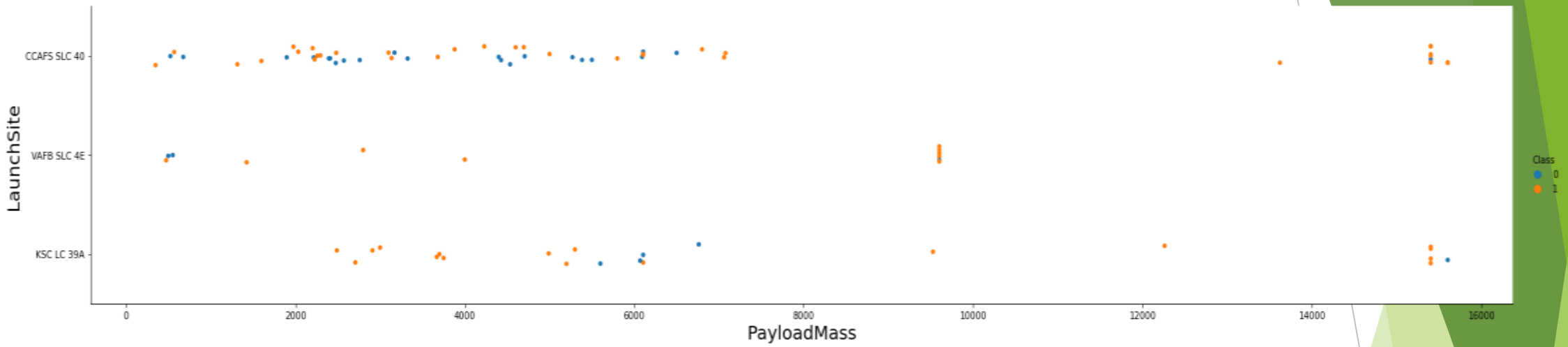


Flight Number vs. Launch Site



- **The best launchers in** the **graph above Currently** the **most recent su ccessful** launch is **the** CCAF5 SLC **40,**
- n second place **is the** VAFB SLC 4E and **in** third place **is the** KSC LC **39A.** ;
- Y**ou will** also **notice** that **your overall** success rate **has** improved over **time.**

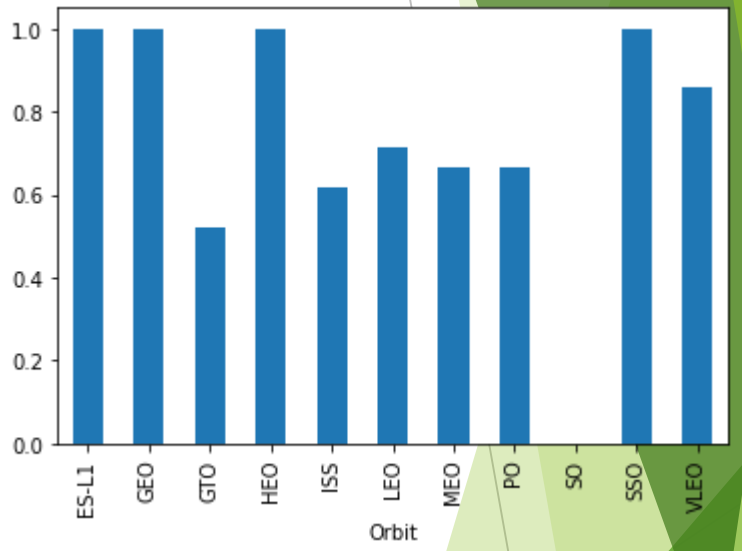
Payload vs. Launch Site



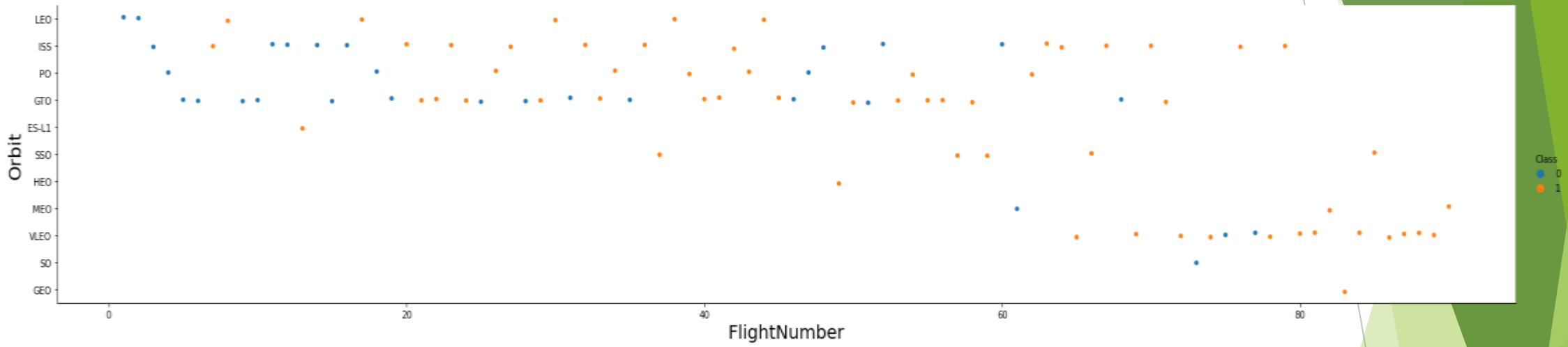
- **A payload of** over **9,000 kg (approximately** the weight of a school bus) **is a huge success.**
- Payloads over **12,000 kg are only** possible **with the** CCAFS SLC 40 and KSC **LC. Launcher 39A.**

Success Rate vs. Orbit Type

- The greatest success rate occurs in orb it:
  - ES-L1;
  - GEO;
  - HEO; and
  - SSO.
- Followed by:
  - VLEO (above 80%); and
  - LFO (above 70%).

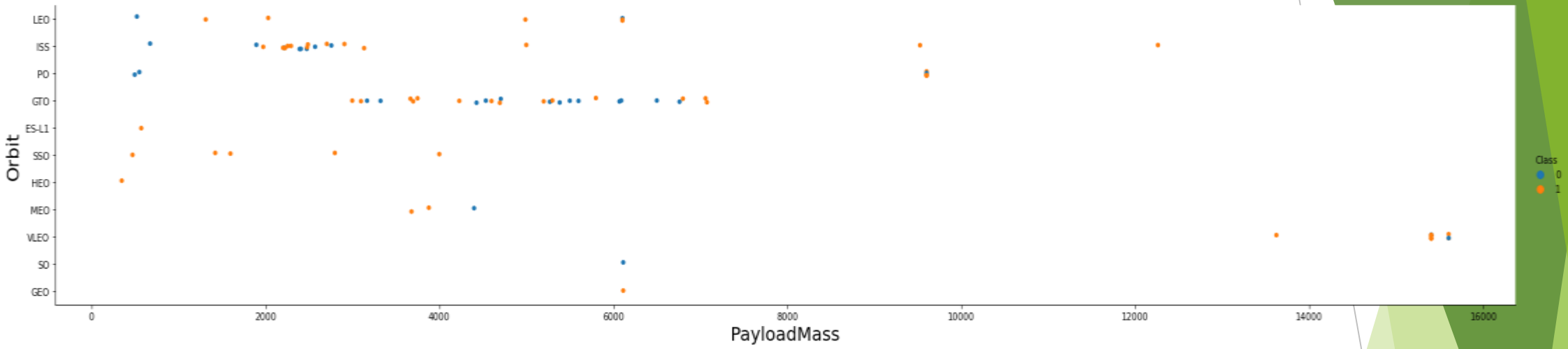


Flight Number vs. Orbit Type



- **The** success rate **appears to have** improved over time **on** all **trajectories,**
- VLEO **orbits appear to be** a new business **opportunity** due to **their** recent **increase.** frequency.

Payload vs. Orbit Type



- **There does not appear to be any relationship** between payload and **GTO orbital** success **rate,**
- **The** ISS orbit has the widest **payload** range and **the highest success rate,**
- **Launches into** SO and **GEO orbits are rare.**

Launch Success Yearly Trend

- Since 2013, the success rate has start ed to increase. Store until 2020,
- The first three years seem to have been a period of adaptation and improvement of the technology.

All Launch Site Names

- According to the data, there are 4 launch pads :

Launch Site
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E

- Obtained by selecting the unique for the “launch_site” value in the data set.

Launch Site Names Begin with 'CCA'

- 5 things that start sites start 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	**CCA**FS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	**CCA**FS 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	**CCA**FS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	F9 v1.0 B0006	**CCA**FS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	**CCA**FS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attemp

- Here we see examples of five ships from Cape Canaveral.

Total Payload Mass

- Total payload carried by NASA launch vehicles::

Total Payload (kg)
111.268

- Total payload is calculated above by summing all payloads with code. Contains "CRS" for NASA.

Average Payload Mass by F9 v1.1

- Average payload mass of projectile version F9 v1.1:

Avg Payload (kg)
2.928

- Filter the data by the booster version above and calculate the average we got a payload mass of 2.929 kg

First Successful Ground Landing Date

- First successful landing on a dirt platform

Min Date
2015-12-22

- If you filter the data for successful off-road landing results and get the minimum value for the date, you'll see the first event occurred on December 22,2015.

Successful Drone Ship Landing with Payload between 4000 and 6000

- A booster that lands successfully on a drone ship and has a payload mass greater than 4,000 but less than 6,000

Booster Version
F9 FT B1021.2
F9 FT B1031.2
F9 FT B1022
F9 FT B1026

- By choosing different versions of the amplifier according to the filter above, you get 4 options.

Total Number of Successful and Failure Mission Outcomes

- Number of successful and **failed** mission **results**

Mission Outcome	Occurrences
Success	99
Success (payload status unclear)	1
Failure (in flight)	1

- Grouping the mission results and
  counting each group's records yields the above summary.

Boosters Carried Maximum Payload

- Boosters carrying 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

Booster Version
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

- These are the projectiles carrying the largest payload masses recorded in the data set.

2015 Launch Records

- Failed landing results for drone spacecraft, booster versions and launch site names in 2015.

Booster Version	Launch Site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

- There are only two items in the list above.

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

- All landing result ratings between June 4, 2010 and 2017 03-20:

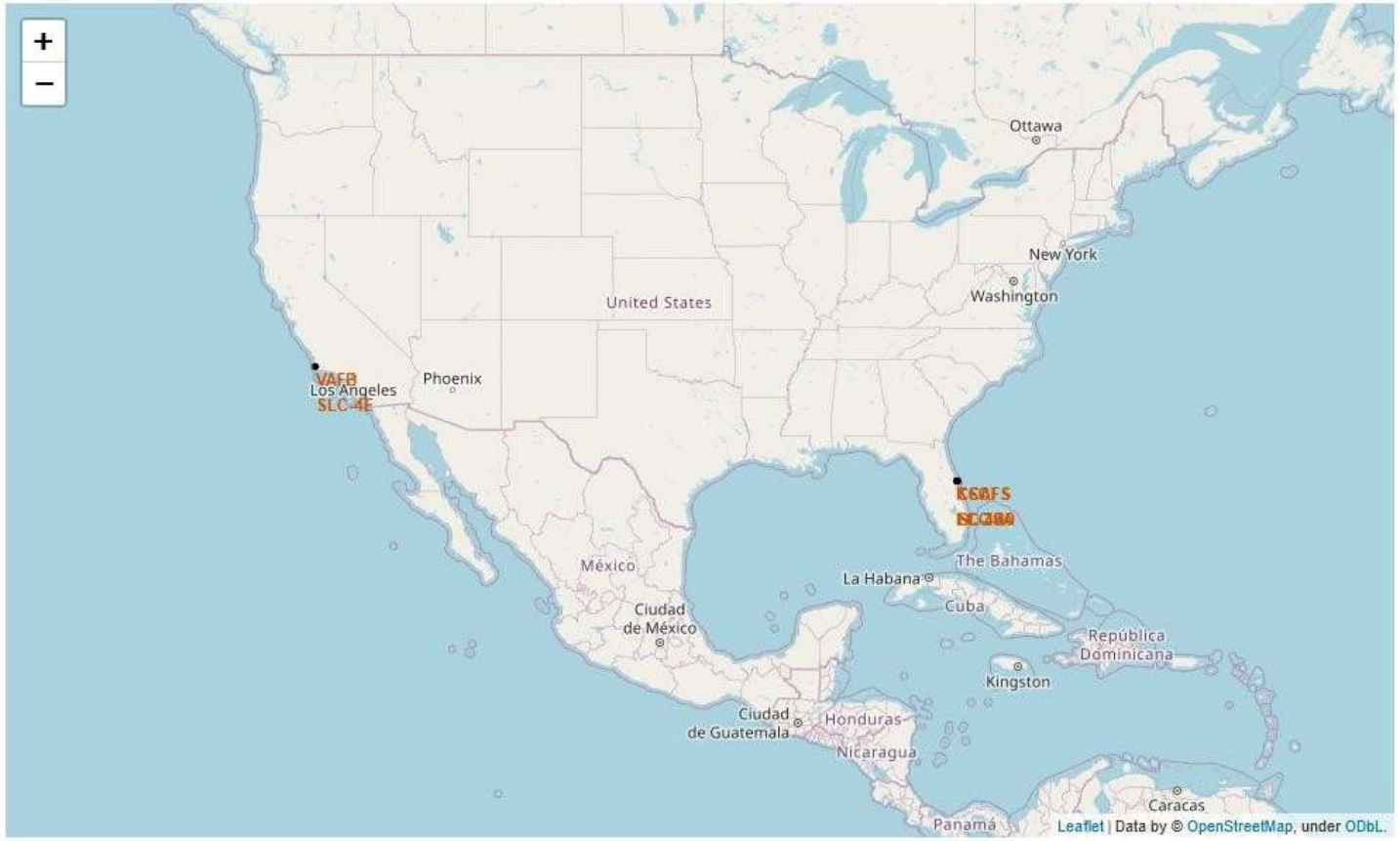
Landing Outcome	Occurrences
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

- This data **representation warns you to consider "don't try".**

Section 4

Launch Sites Proximities Analysis

All launch sites



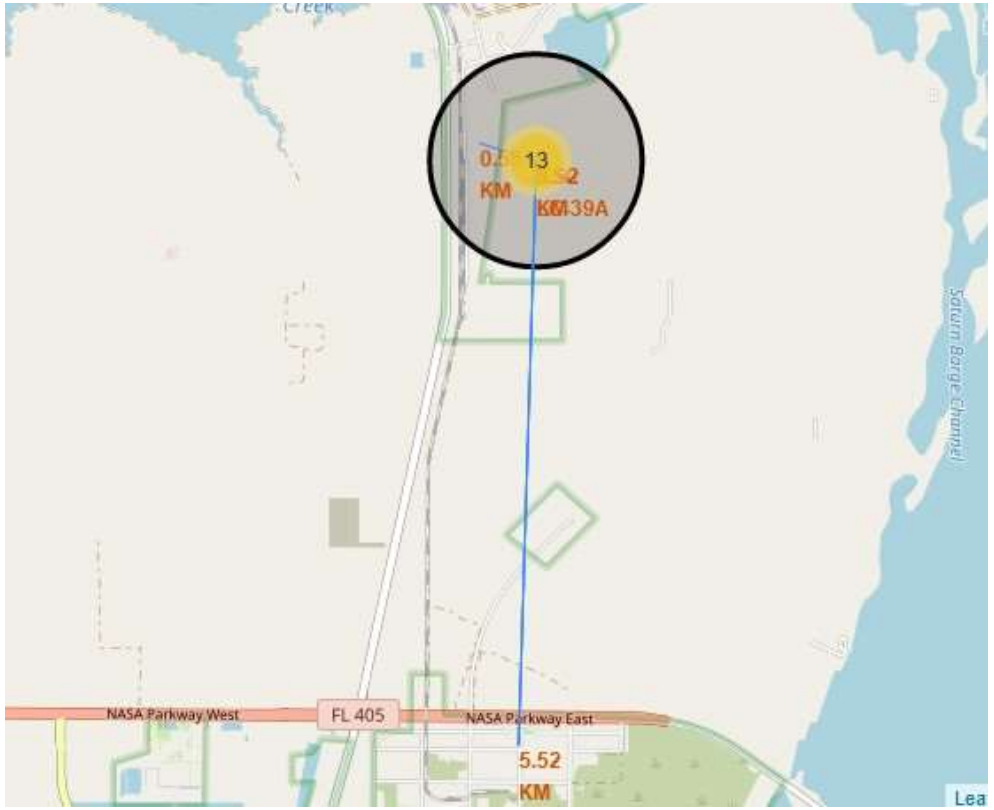
- The launch pad is probably close to the sea for safety reasons, but the road and railroad.

Launch Outcomes by Site

- An example of the results of a KSC LC-39A launch from the space station

A green marker indicates success and a red marker indicates failure.

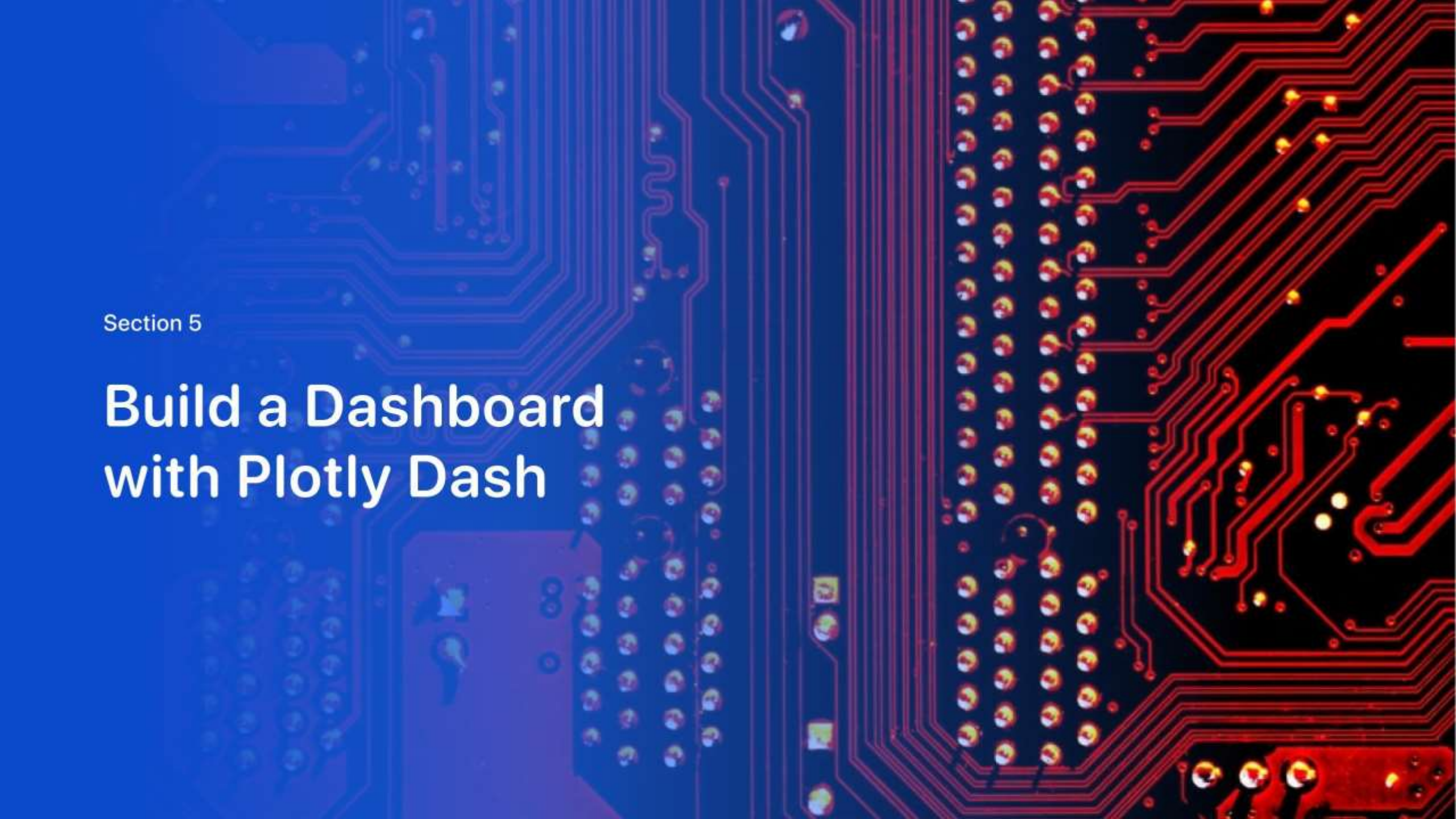
Logistics and Safety



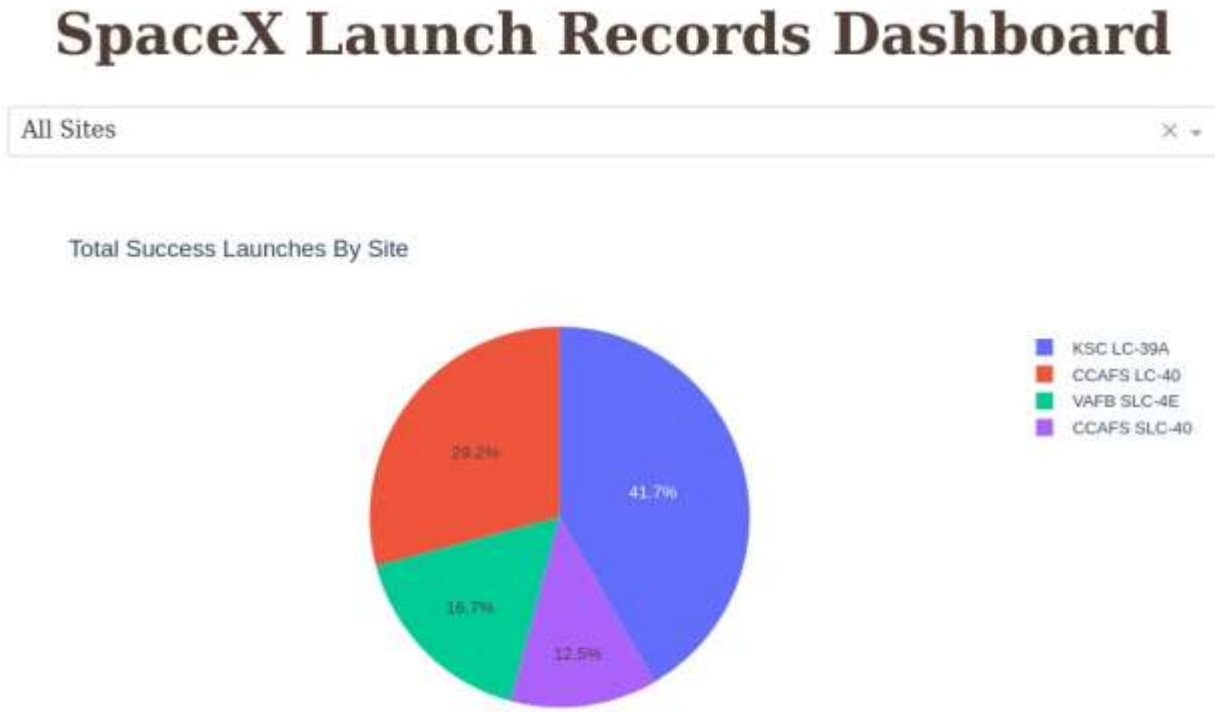
- The KSC LC-39A launch pad is close to railways and roads and relatively far from densely populated areas, so the logistics aspect is excellent

Section 5

Build a Dashboard with Plotly Dash

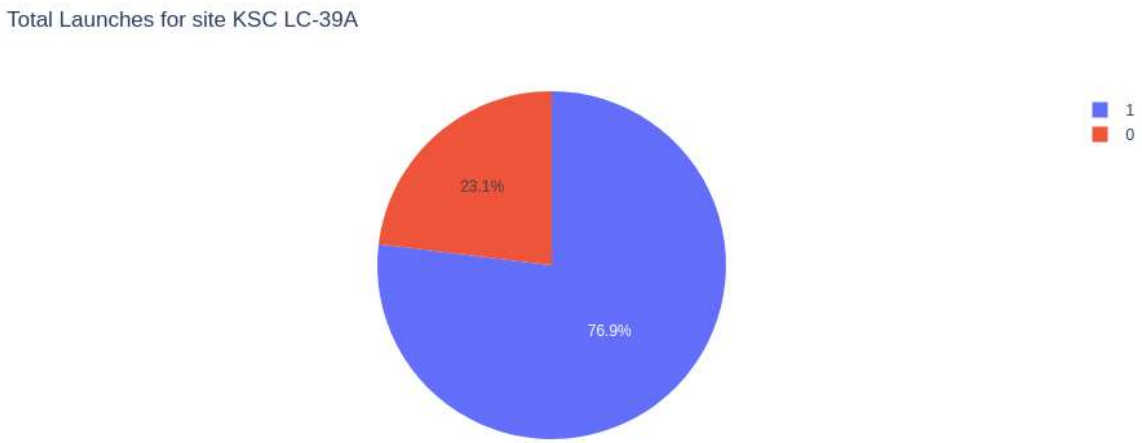


Successful Launches by Site



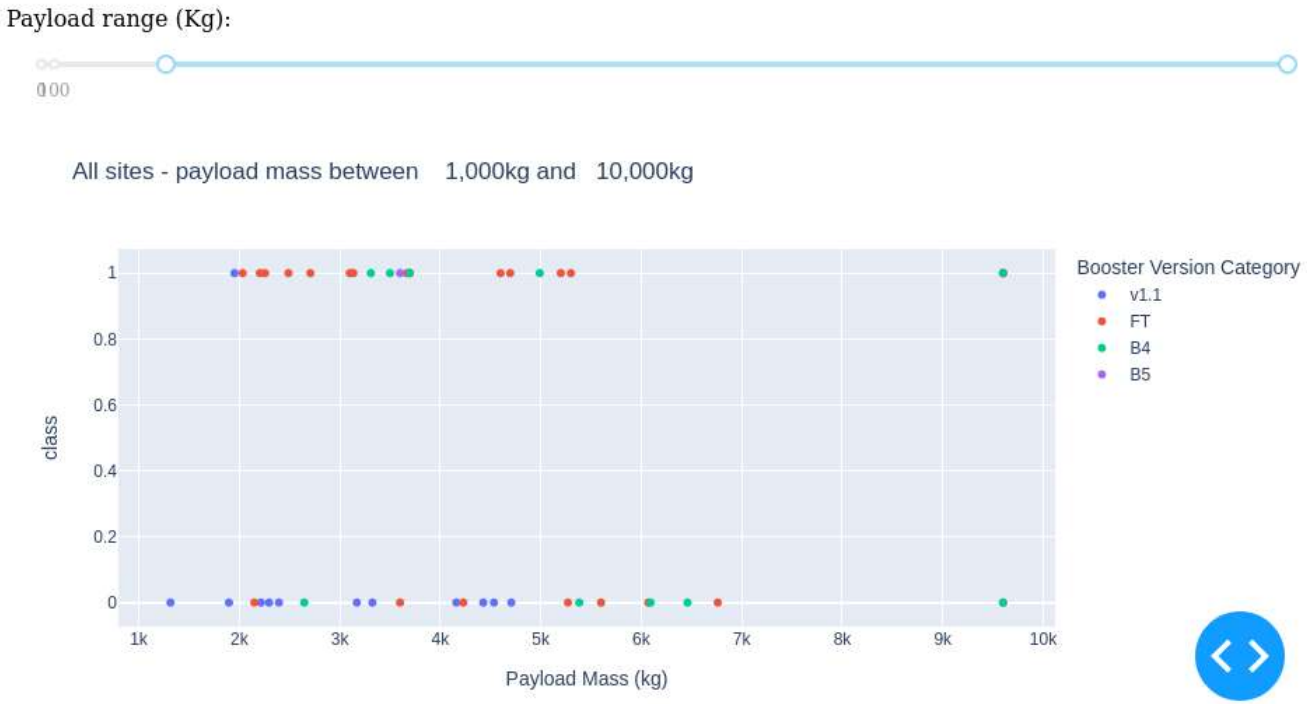
- The location of the launch seems to be a very important factor.
- mission success.

Launch Success Ratio for KSC LC-39A



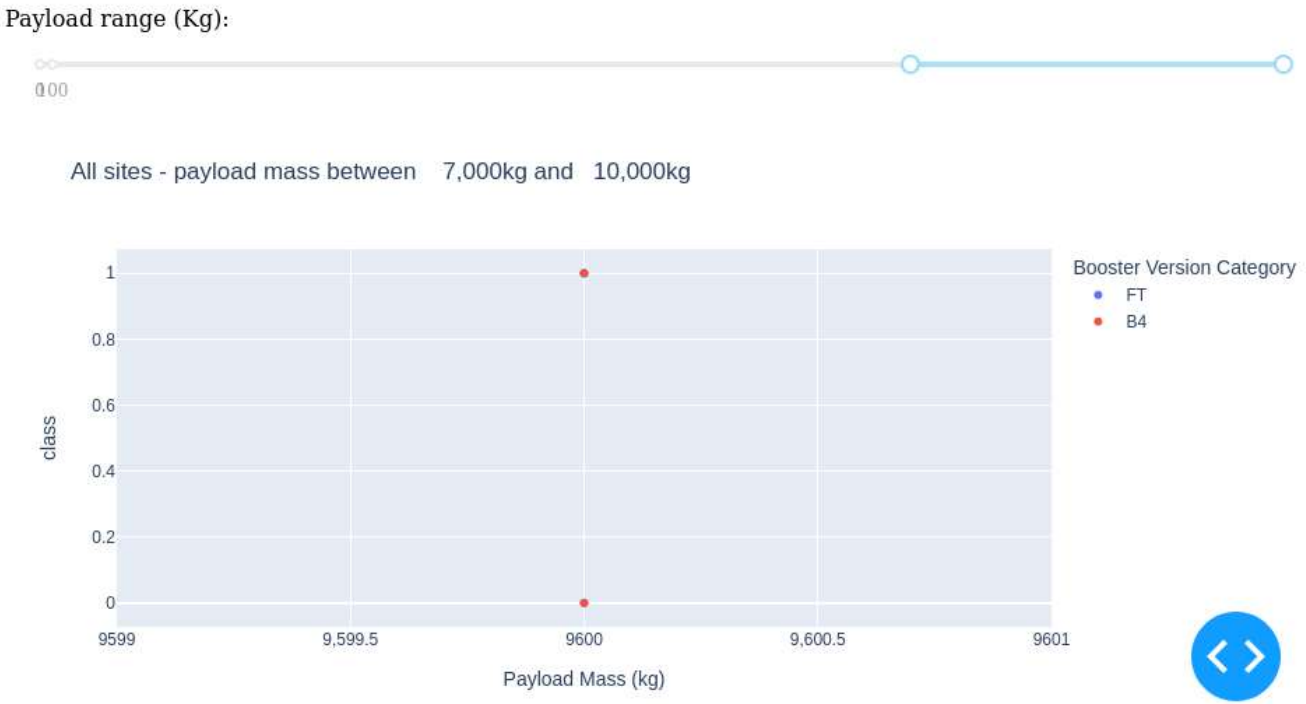
- 76.9% of **the site's** launches **were successful**.

Payload vs. Launch Outcome



- A payload of less than 6000 kg and an FT booster is the most successful combination.

Payload vs. Launch Outcome



- There is insufficient data to assess the risk of launching more than 7000 kg.

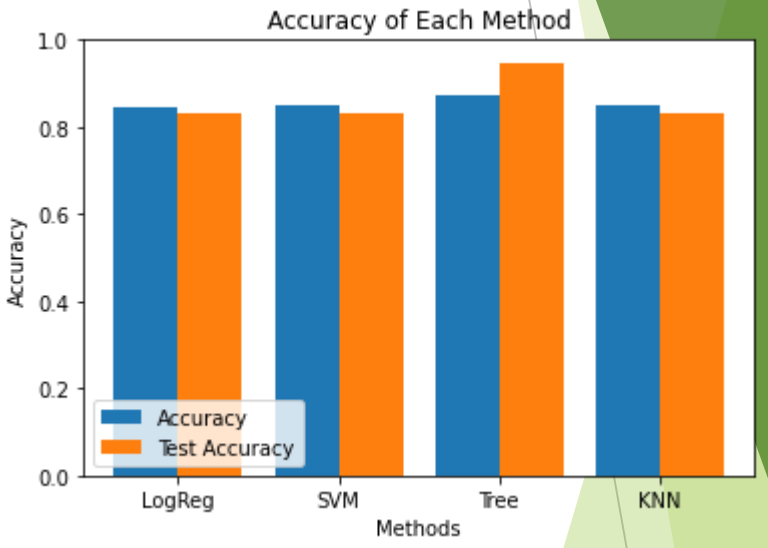
Section 6

Predictive Analysis (Classification)

Classification Accuracy

- Four classification models were tested and their accuracies are displayed next to them,

- The model with the highest classification accuracy is the Decision Tree Classifier with an accuracy of over 87%.



Confusion Matrix of Decision Tree Classifier

- The decision tree classifier confusion matrix proves correctness by showing large. Compares the number of false positives and true negatives.

Conclusions

- By analyzing various data sources, we came to the following conclu sions:-
- The best launcher is the KSC LC-39A.;
- Launches over 7000 kg are less risky;
- Most mission outcomes are successful, but successful landings seem to improve over time as processes and missiles evolve. ;
- A decision tree classifier can be used to predict successful landings and Increase your profits.

Thank you!