IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ruth Okolo

August 8Th 2024

Outline

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

Executive Summary

The methods listed below were applied to the data analysis:

- Data collection is accomplished using SpaceX API and Webscraping.
- Exploratory Data Analysis, data wrangling and data visualization using folium.
- Machine Learning Prediction.

A summary of all the outcomes shows that:

- It was feasible to gather useful information from open sources;
- EDA made it possible to determine which characteristics are most predictive of launch success;
- Machine Learning Prediction demonstrated the most effective model for predicting the qualities that are crucial to drive this opportunity in the most efficient manner, utilizing all the data gathered.

Introduction

Objective:

- The purpose of this project is to predict if the Falcon first stage will land successfully.

Issues for which solutions are required:

- The most accurate method of estimating launch of costs overall is to forecast the first stage rockets' successful landings.
- The method that performs best for machine learning prediction.

Section 1

Methodology

Methodology

Executive Summary

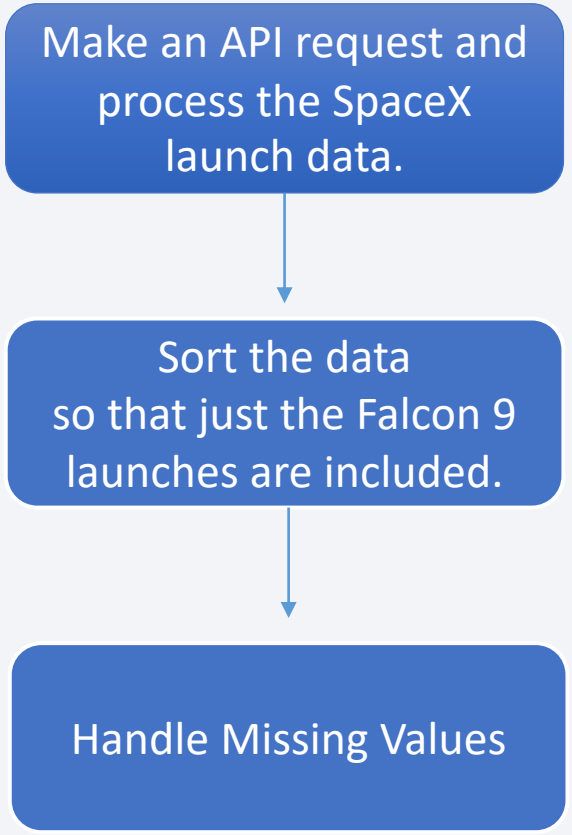
- Perform Data collection methodology:
- Data from Space X was obtained from 2 sources:
- Space X API (https://api.spacexdata.com/v4/rockets/)
- WebScraping (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)
- Perform data wrangling
- Following feature analysis and feature summarization, collected data was enhanced by the creation of a landing result label based on outcome data.
- 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

- Data Collection was performed by collecting data sets from Space X API (https://api.spacexdata.com/v4/rockets/) and Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches), using Webscraping.

Data Collection - SpaceX API

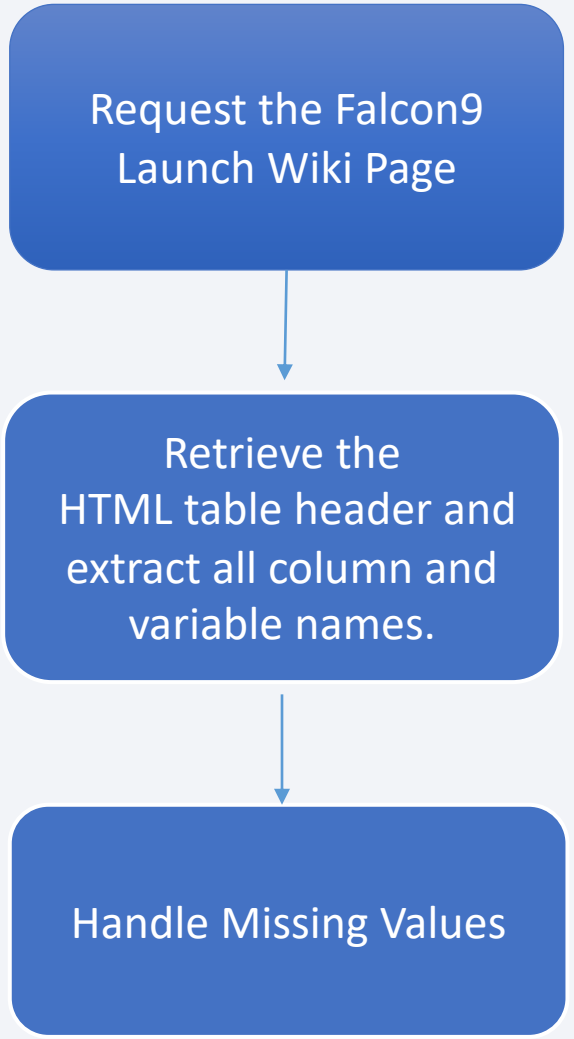
- SpaceX provides a public API through which information may be acquired and utilized.



https://github.com/RuthOkolo/ProfessionalCertifications/tree/IBM-DATA-SCIENCE

Data Collection – Webscraping

- Wikepedia provides data from SpaceX launches.



https://github.com/RuthOkolo/ProfessionalCertifications/tree/IBM-DATA-SCIENCE

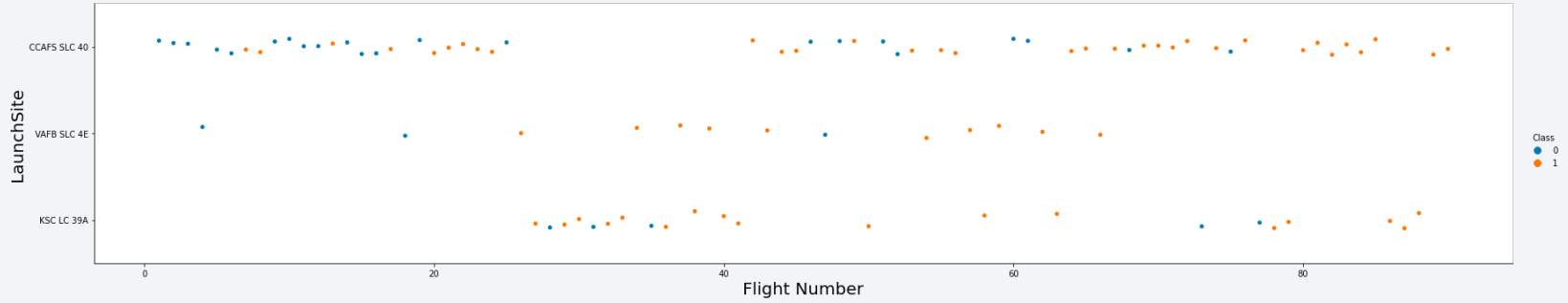
Data Wrangling

- Exploratory Data Analysis (EDA) was performed on the dataset first.
- Then the summaries launches per site, occurrences of each orbit and occurrences of mission outcome per orbit type were evaluated.
- Finally, the landing outcome label was created from Outcome column.

- https://github.com/RuthOkolo/ProfessionalCertifications/tree/IBM-DATA-SCIENCE

EDA with Data Visualization

- To explore data, scatterplots and barplots were used to visualize the relationship between pair of features:Payload Mass X Flight Number, Launch Site X Flight Number, Launch Site X Payload Mass, Orbit and Flight Number, Payload and Orbit



https://github.com/RuthOkolo/ProfessionalCertifications/tree/IBM-DATA-SCIENCE

EDA with SQL

The following SQL queries were performed:

1. Names of the unique launch sites in the space mission;
2. Top 5 launch sites whose name begin with the string 'CCA';
3. Total payload mass carried by boosters launched by NASA (CRS);
4. Average payload mass carried by booster version F9 v1.1;
5. Date when the first successful landing outcome in ground pad was achieved;
6. Names of the boosters which have success in drone ship and have payload mass between 4000 and 6000 kg;
7. Total number of successful and failure mission outcomes;
8. Names of the booster versions which have carried the maximum payload mass;
9. Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015.
10. Rank of the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20.

Build an Interactive Map with Folium

Markers, circles, lines and marker clusters were used with Folium Maps

- Markers indicate points like launch sites;
- Circles indicate highlighted areas around specific coordinates, like NASA Johnson Space Center;
- Marker clusters indicates groups of events in each coordinate, like launches in a launch site;
- Lines are used to indicate distances between two coordinates.

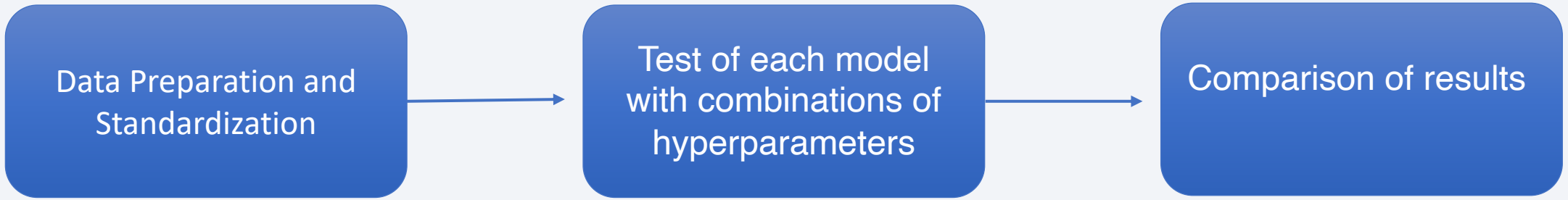
https://github.com/RuthOkolo/ProfessionalCertifications/tree/IBM-DATA-SCIENCE

Build a Dashboard with Plotly Dash

The following graphs and plots were used to visualize data:

- Percentage of launches by site
- Payload range

- This combination allowed to quickly analyze the relation between payloads and launch sites, helping to identify where is best place to launch according to payloads.



https://github.com/RuthOkolo/ProfessionalCertifications/tree/IBM-DATA-SCIENCE

Predictive Analysis (Classification)

Four classification models were compared:

- ✓ logistic regression
- ✓ support vector machine
- ✓ decision tree
- ✓ k nearest neighbors

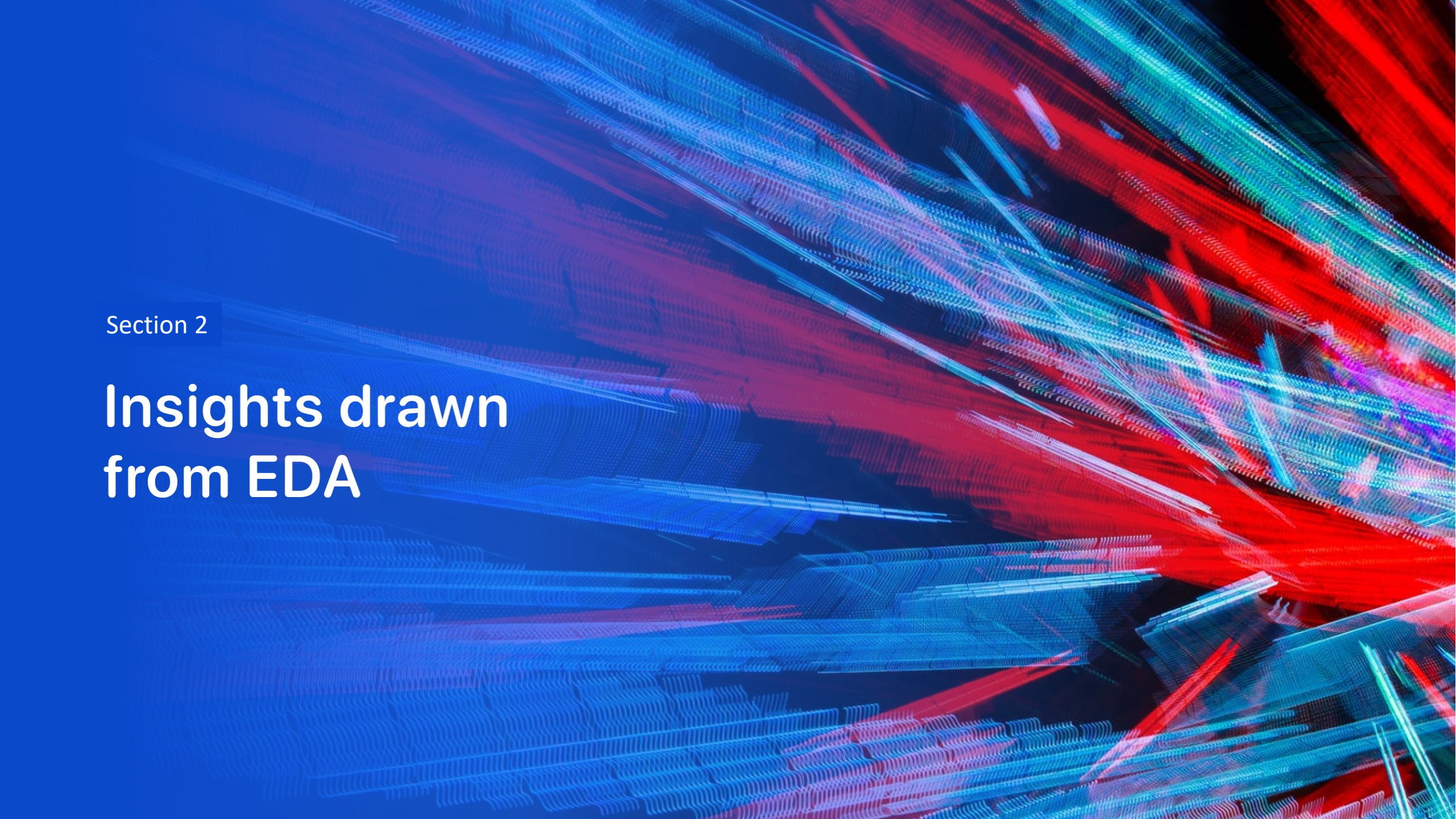
https://github.com/RuthOkolo/ProfessionalCertifications/tree/IBM-DATA-SCIENCE

Results

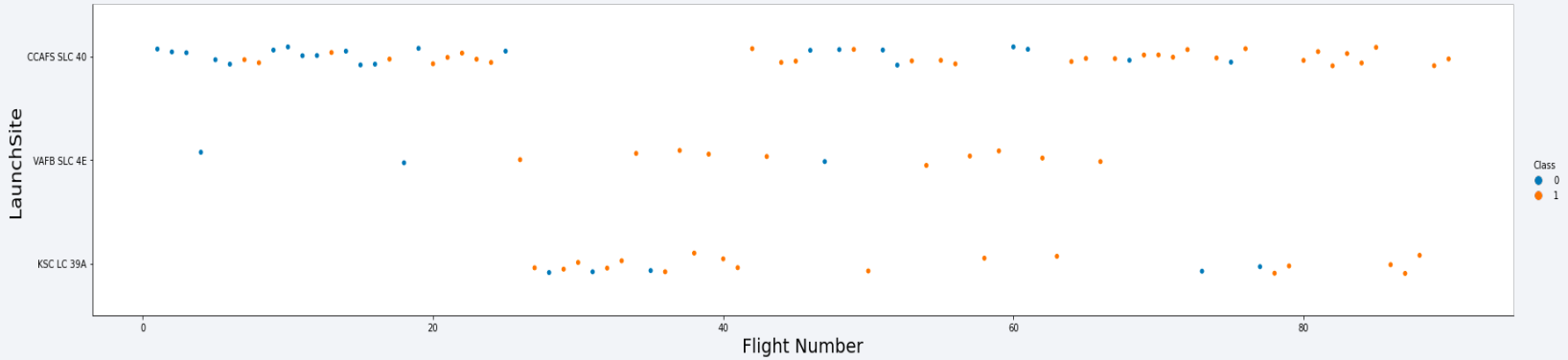
Exploratory data analysis results:

- Space X uses 4 different launch sites;
- The first launches were done to Space X itself and NASA;
- The average payload of F9 v1.1 booster is 2,928 kg;
- The first success landing outcome happened in 2015 fiver year after the first launch;
- Many Falcon 9 booster versions were successful at landing in drone ships having payload above the average;

https://github.com/RuthOkolo/ProfessionalCertifications/tree/IBM-DATA-SCIENCE


Section 2
Insights drawn from EDA

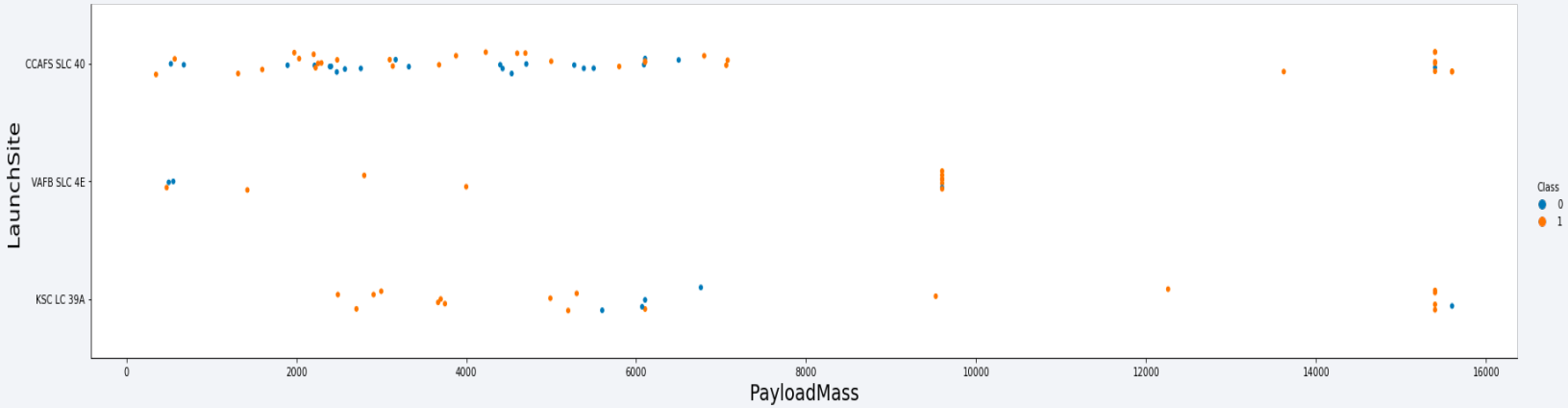
Flight Number vs. Launch Site



-
- According to the plot above, it's possible to verify that the best launch site
- nowadays is CCAF5 SLC 40, where most of recent launches were successful;
- In second place VAFB SLC 4E and third place KSC LC 39A;
- It's also possible to see that the general success rate improved over time

https://github.com/RuthOkolo/ProfessionalCertifications/tree/IBM-DATA-SCIENCE

Payload vs. Launch Site



- Payloads over 9,000kg (about the weight of a school bus) have excellent success rate;
- Payloads over 12,000kg seems to be possible only on CCAFS SLC 40 and KSC LC 39A launch sites.

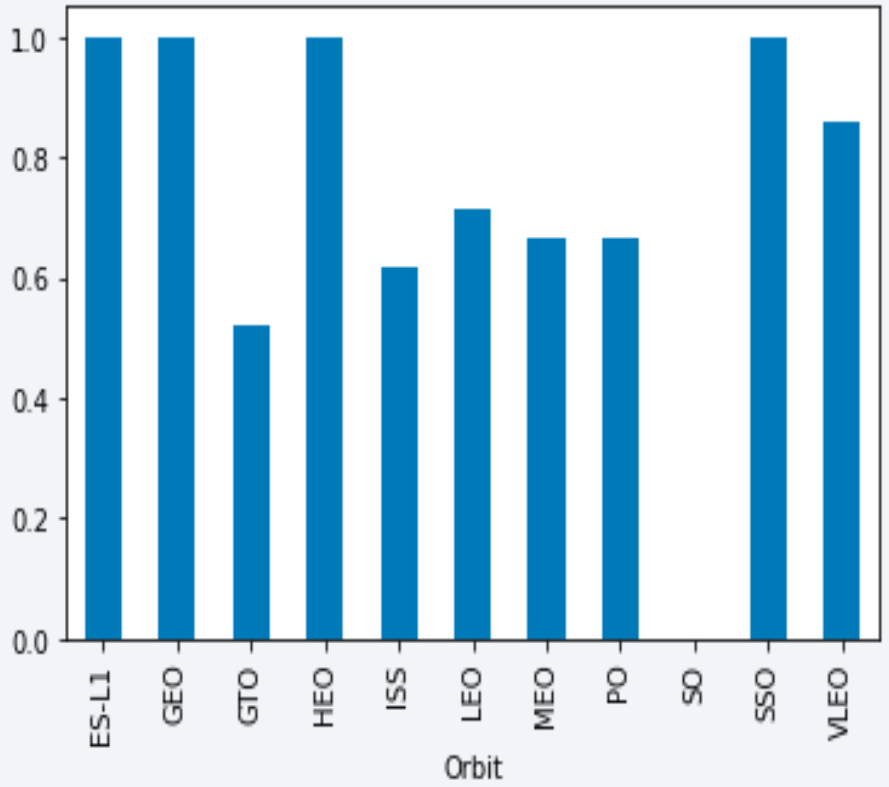
https://github.com/RuthOkolo/ProfessionalCertifications/tree/IBM-DATA-SCIENCE

Success Rate vs. Orbit Type

The biggest success rates happens to orbits:

- ES-L1;
- GEO
- HEO
- SSO.

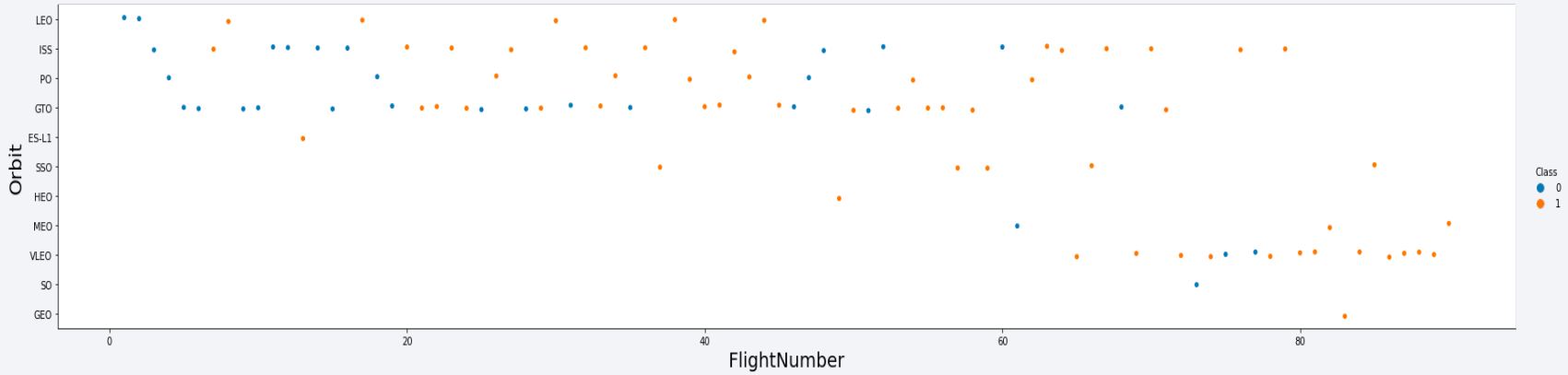
- Followed by:
- VLEO (above 80%); and
- LFO (above 70%).



https://github.com/RuthOkolo/ProfessionalCertifications/tree/IBM-DATA-SCIENCE

Flight Number vs. Orbit Type

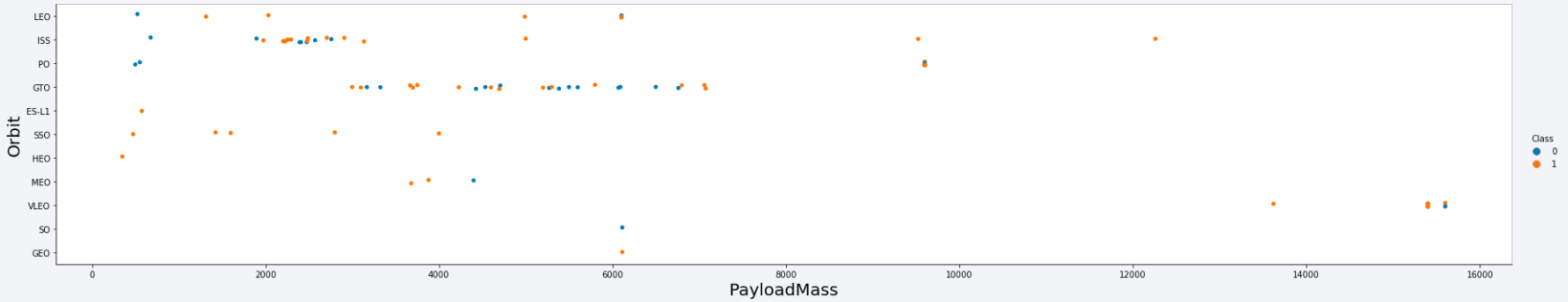
- The success rate greatly improved over time to all orbits;
- VLEO orbit seems a new business opportunity, due to recent increase of its
- frequency.



https://github.com/RuthOkolo/ProfessionalCertifications/tree/IBM-DATA-SCIENCE

Payload vs. Orbit Type

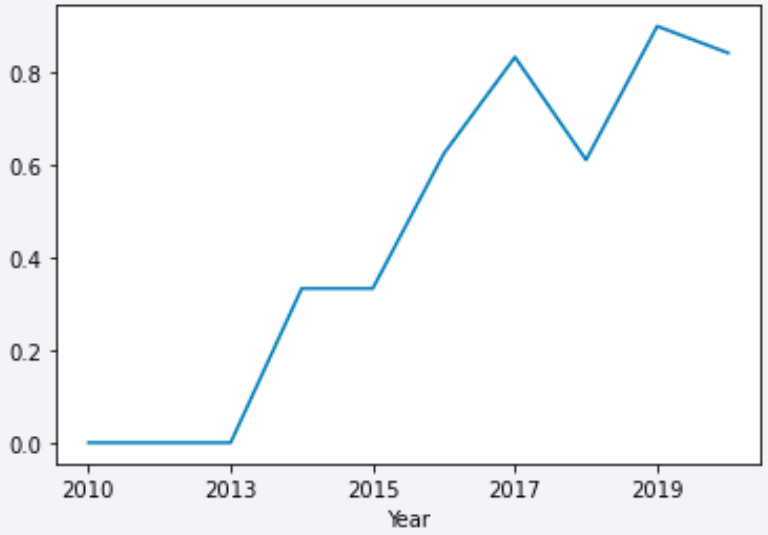
- There is no relation between payload and success rate to orbit GTO;
- ISS orbit has the widest range of payload and a good rate of success;
- There are few launches to the orbits SO and GEO.



https://github.com/RuthOkolo/ProfessionalCertifications/tree/IBM-DATA-SCIENCE

Launch Success Yearly Trend

- Success rate started increasing in 2013 and
- kept until 2020;
- It seems that the first three years were a period of adjusts and improvement of technology.



https://github.com/RuthOkolo/ProfessionalCertifications/tree/IBM-DATA-SCIENCE

All Launch Site Names

- The four launch sites obtained by selecting unique occurrences of “launch site” values from the dataset are:

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

https://github.com/RuthOkolo/ProfessionalCertifications/tree/IBM-DATA-SCIENCE

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 attemp

- Here we can see five samples of Cape Canaveral launches.

https://github.com/RuthOkolo/ProfessionalCertifications/tree/IBM-DATA-SCIENCE

Total Payload Mass

Total Payload (kg)
111.268

- Total payload calculated above, by summing all payloads whose codes
- contain ‘CRS’, which corresponds to NASA.

https://github.com/RuthOkolo/ProfessionalCertifications/tree/IBM-DATA-SCIENCE

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1:
- Filtering data by the booster version above and calculating the average payload mass we obtained the value of 2,928 kg.

Avg Payload (kg)
2.928

https://github.com/RuthOkolo/ProfessionalCertifications/tree/IBM-DATA-SCIENCE

First Successful Ground Landing Date

- By filtering data by successful landing outcome on ground pad and getting the minimum value for date it's possible to identify the first occurrence, that happened on 12/22/2015.

Min Date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000
- Selecting distinct booster versions according to the filters above, these 4 are the result.

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

https://github.com/RuthOkolo/ProfessionalCertifications/tree/IBM-DATA-SCIENCE

Total Number of Successful and Failure Mission Outcomes

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

Boosters Carried Maximum Payload

Booster Version (...)	Booster Version
F9 B5 B1048.4	F9 B5 B1051.4
F9 B5 B1048.5	F9 B5 B1051.6
F9 B5 B1049.4	F9 B5 B1056.4
F9 B5 B1049.5	F9 B5 B1058.3
F9 B5 B1049.7	F9 B5 B1060.2
F9 B5 B1051.3	F9 B5 B1060.3

https://github.com/RuthOkolo/ProfessionalCertifications/tree/IBM-DATA-SCIENCE

2015 Launch Records

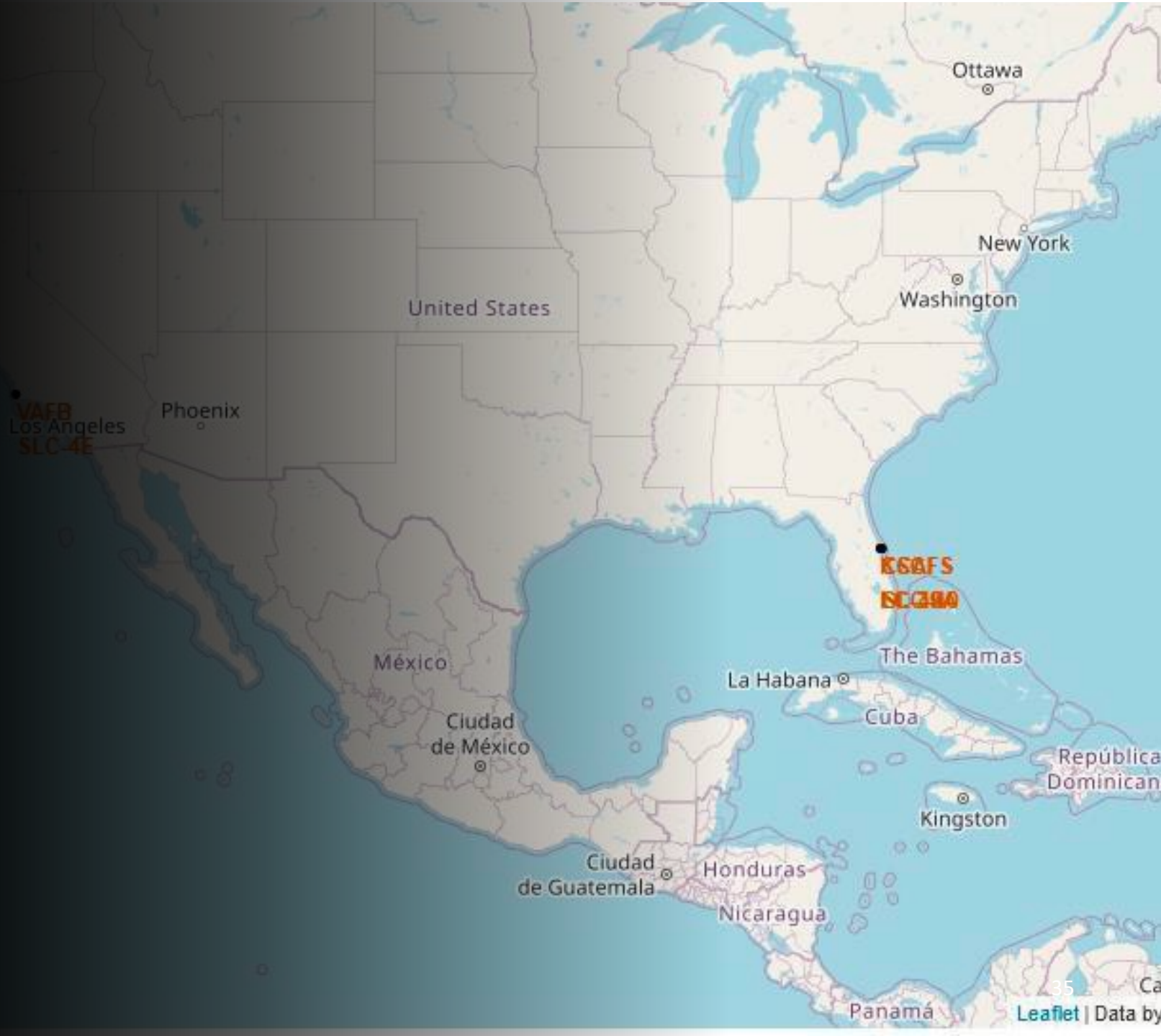
Booster Version	Launch Site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

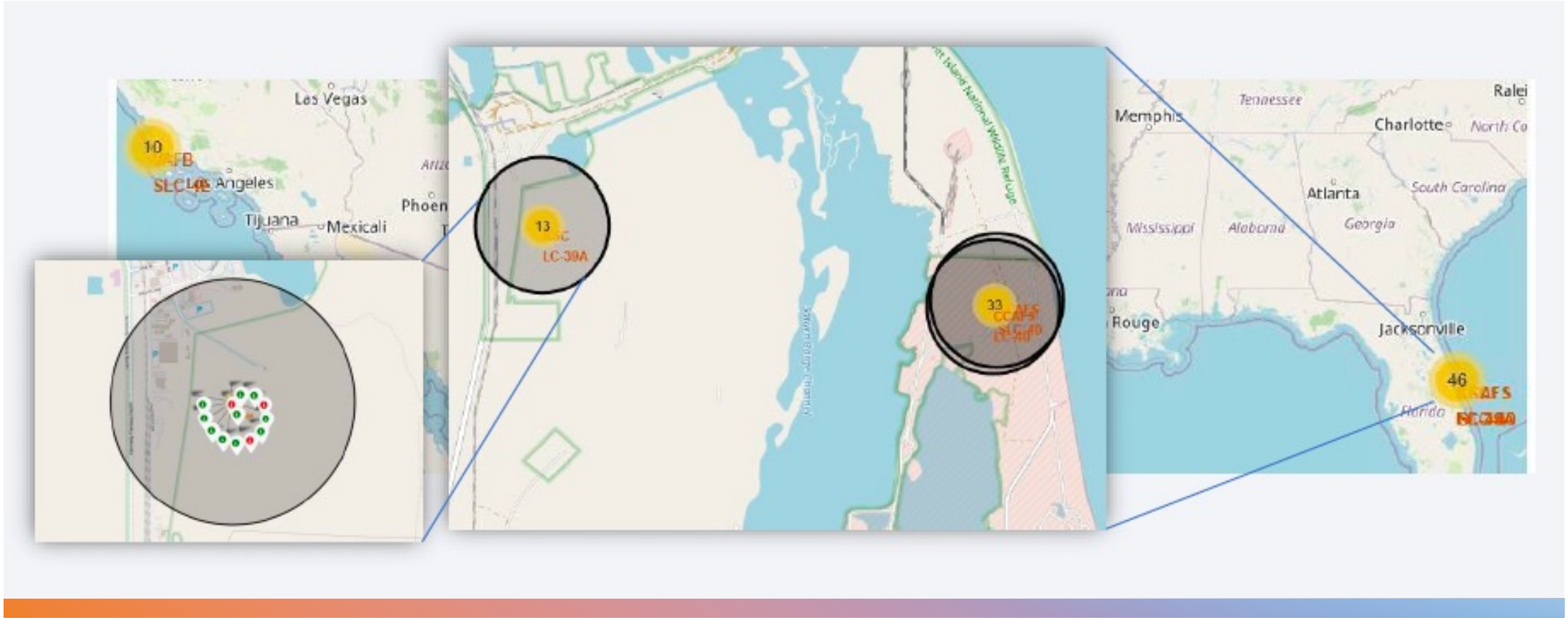
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

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


Section 3
Launch Sites
Proximities Analysis

Launch sites on a global map

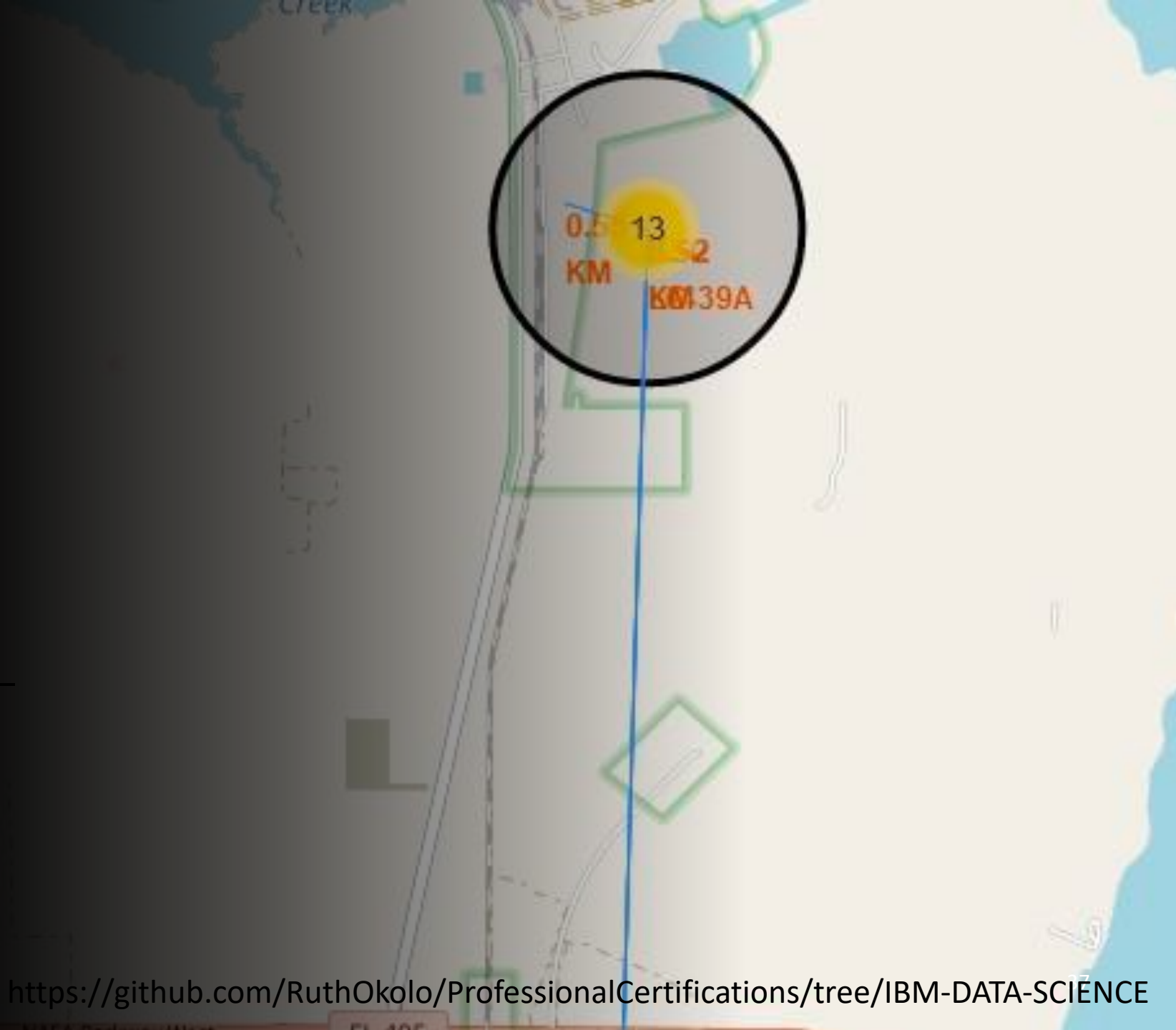




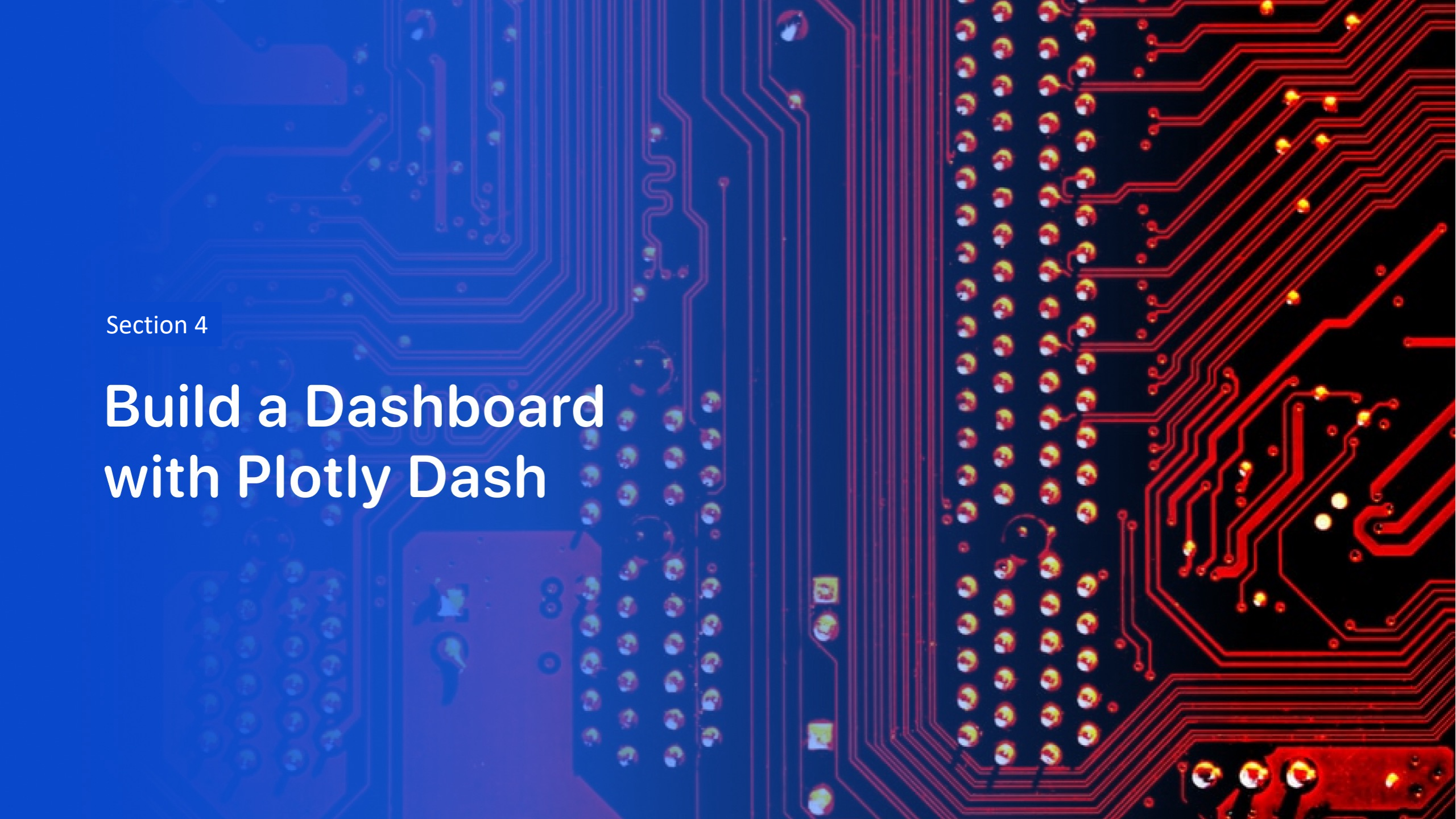
Launch Outcomes by Site

https://github.com/RuthOkolo/ProfessionalCertifications/tree/IBM-DATA-SCIENCE

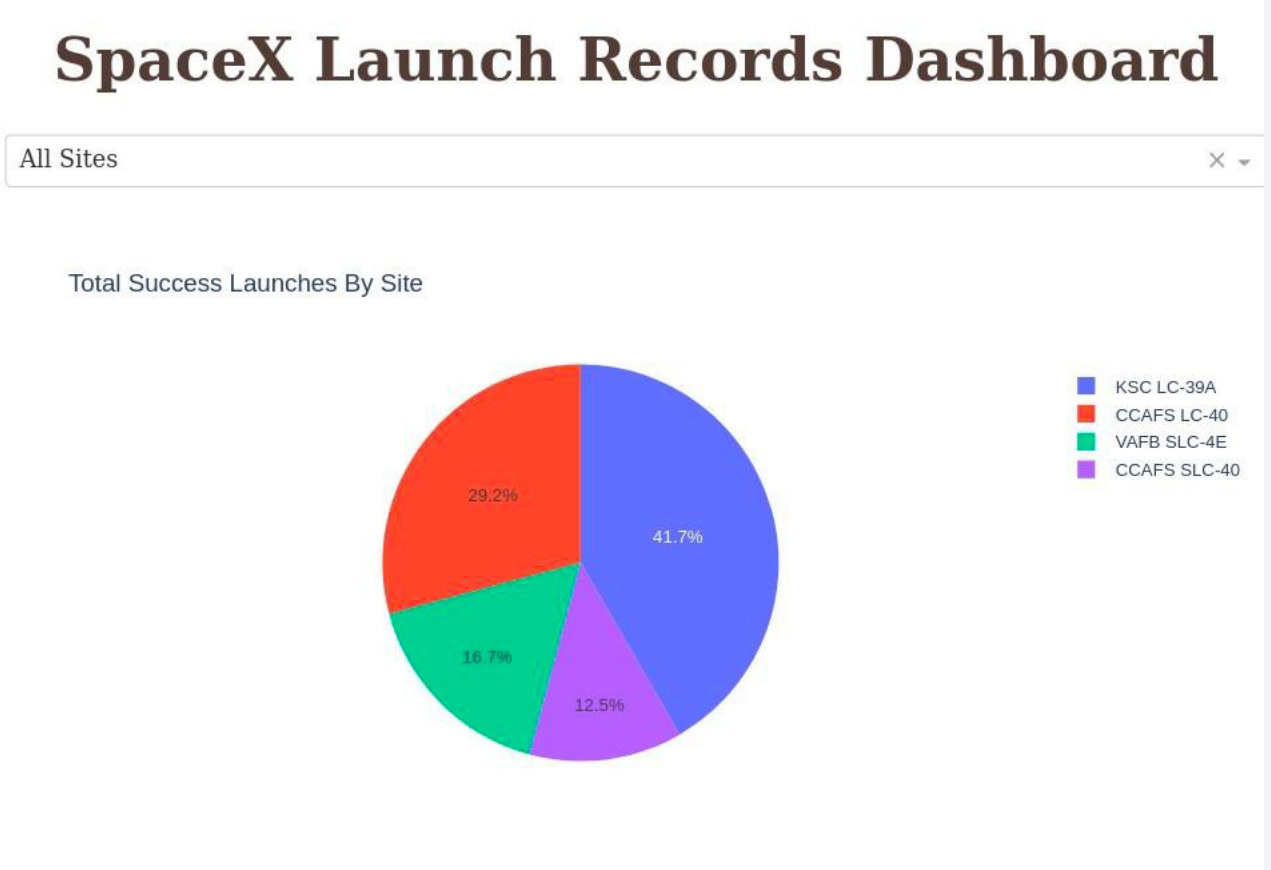
Logistics and Safety



https://github.com/RuthOkolo/ProfessionalCertifications/tree/IBM-DATA-SCIENCE

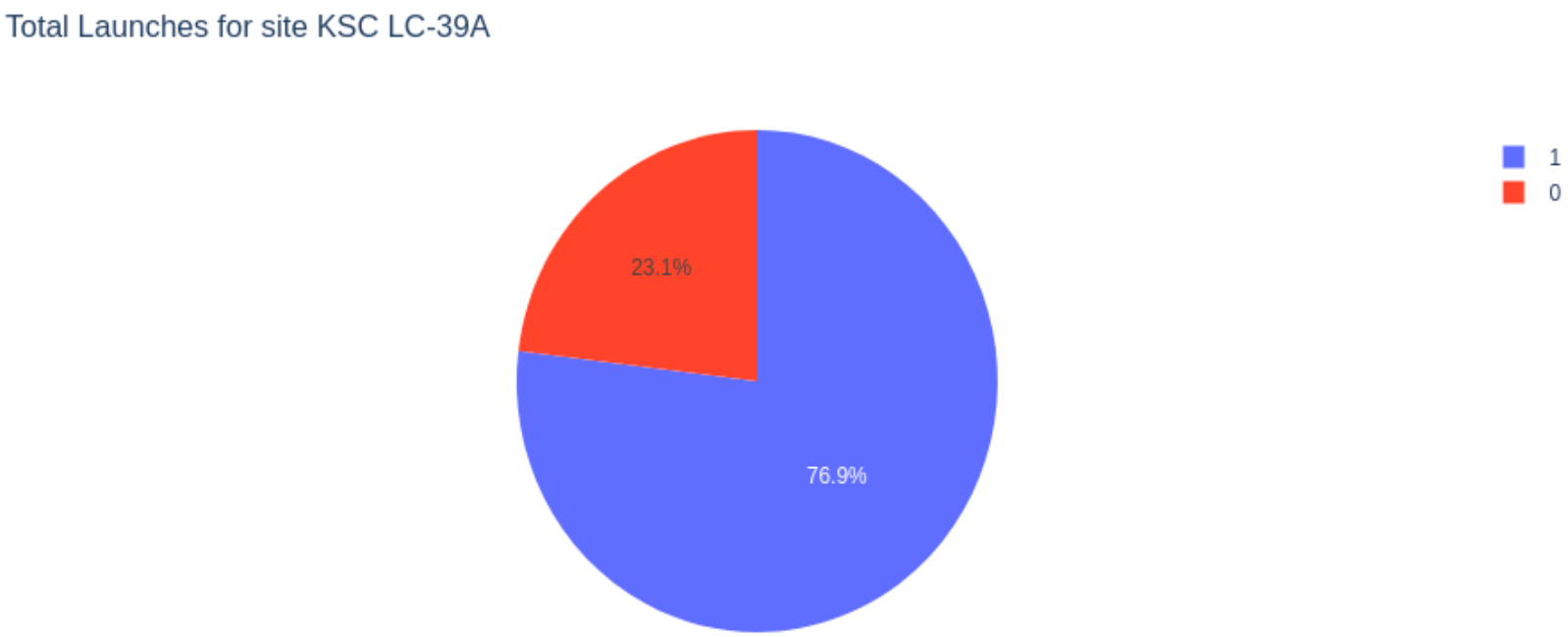

Section 4
Build a Dashboard with Plotly Dash

Sites that were successfully launched



https://github.com/RuthOkolo/ProfessionalCertifications/tree/IBM-DATA-SCIENCE

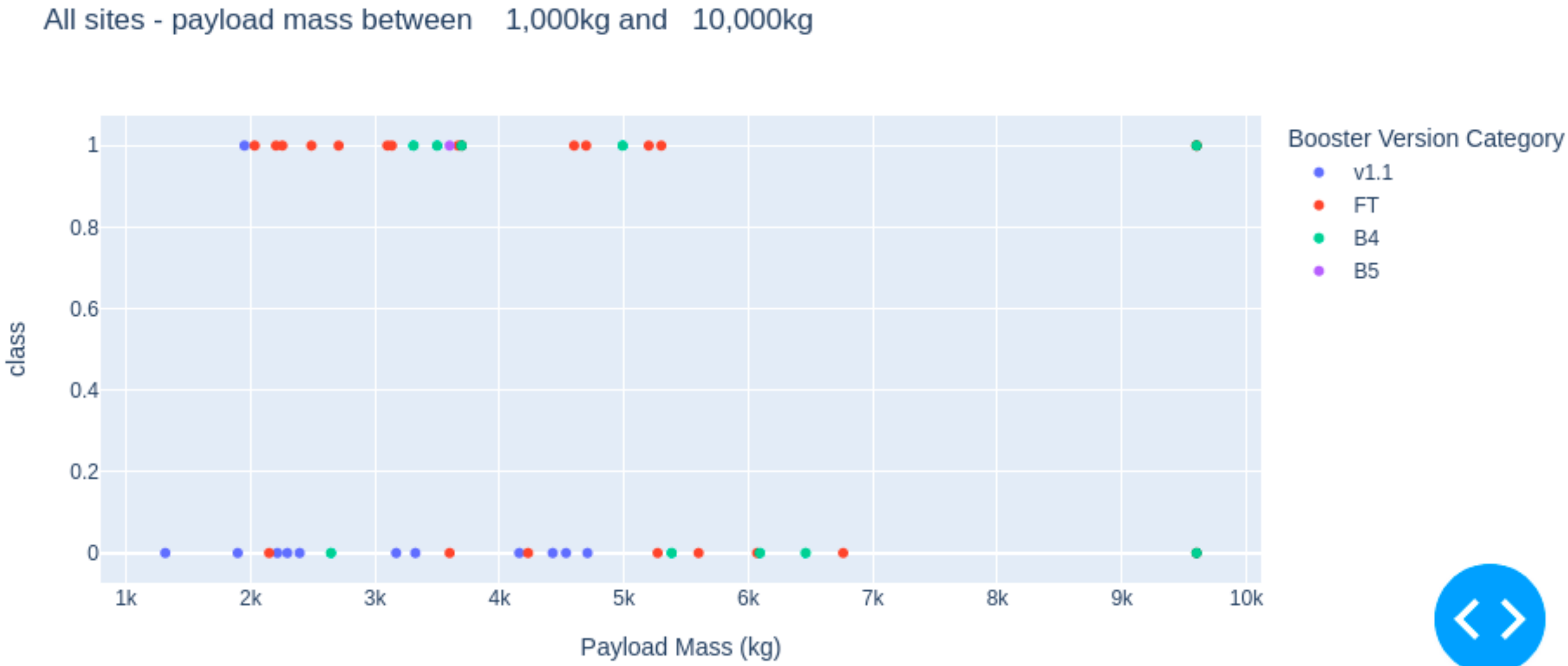
Success Ratio for KSC LC-39A Launch



https://github.com/RuthOkolo/ProfessionalCertifications/tree/IBM-DATA-SCIENCE

Payload range (Kg):

Payload vs. Launch Outcome



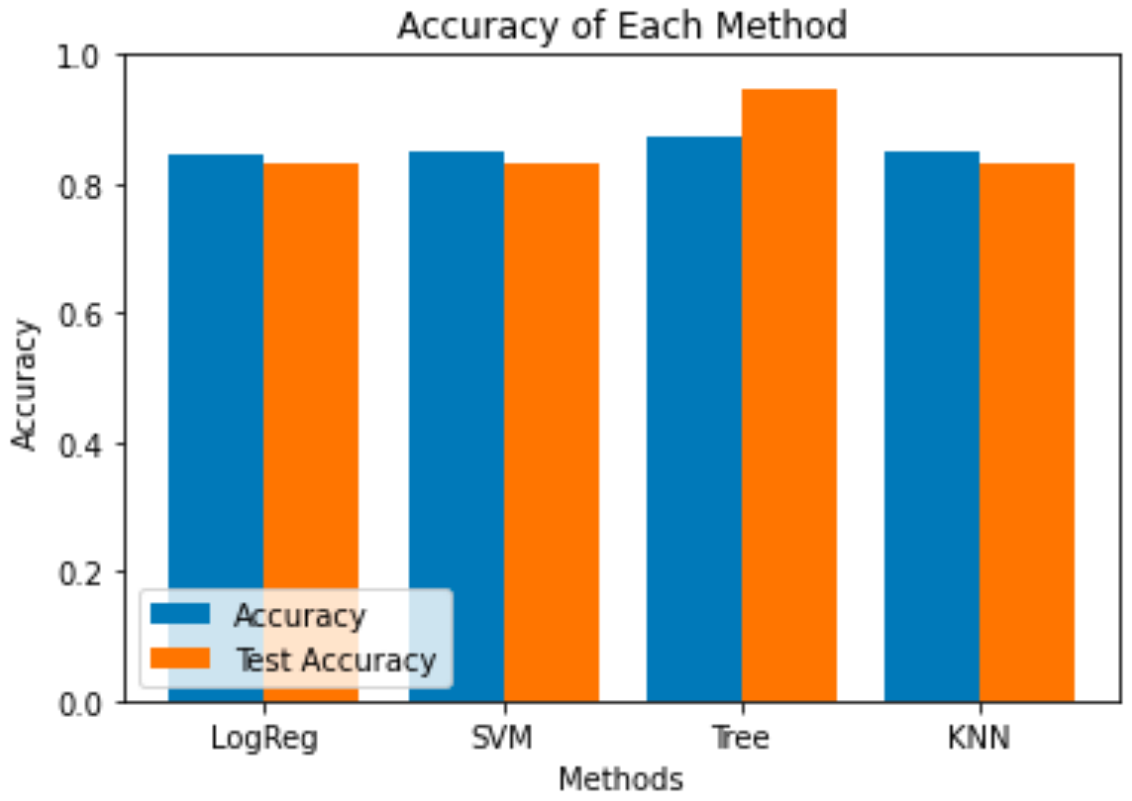
https://github.com/RuthOkolo/ProfessionalCertifications/tree/IBM-DATA-SCIENCE

Section 5

Predictive Analysis (Classification)

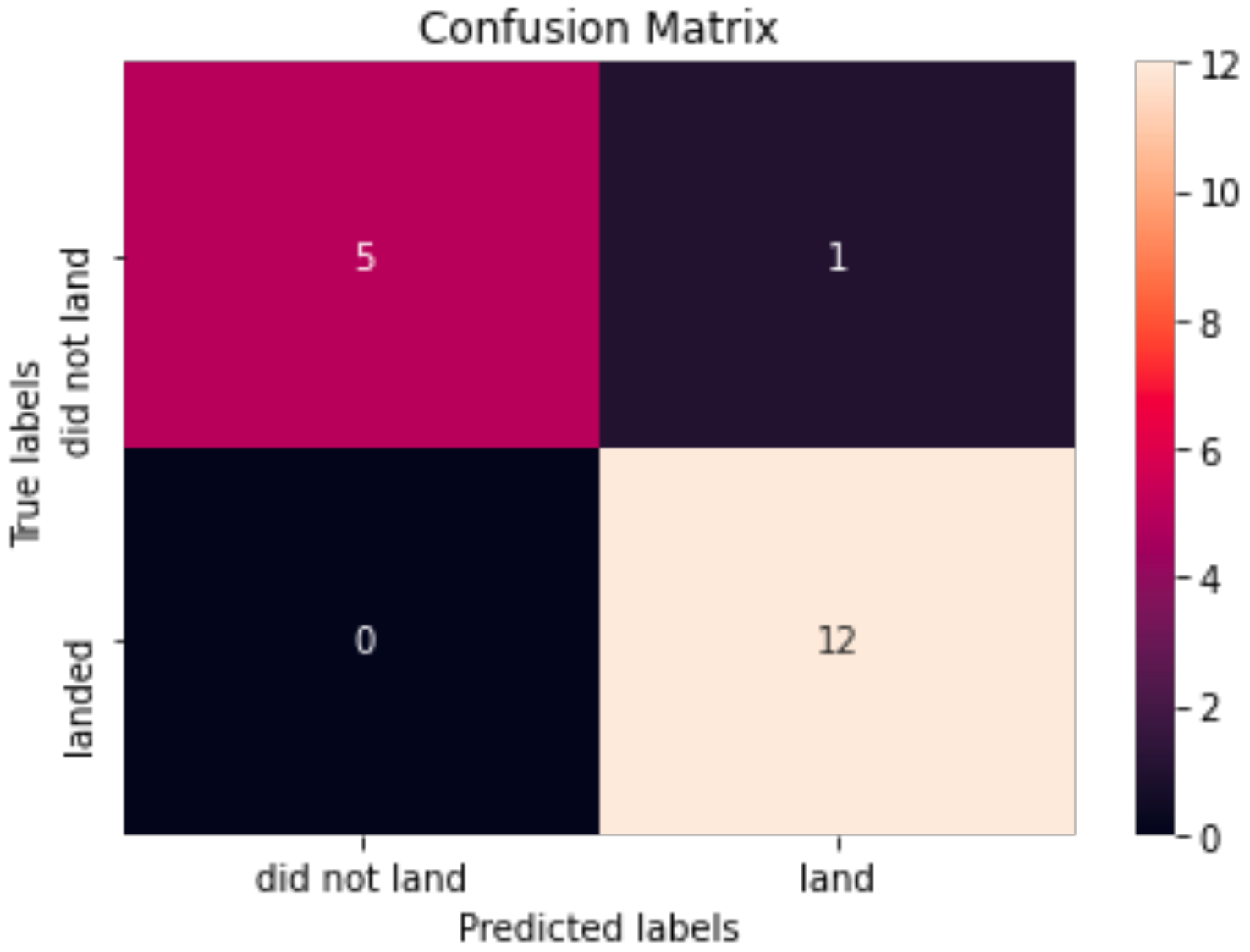
Classification Accuracy

- The model with the highest classification accuracy is Decision Tree Classifier, which has accuracies over than 87%.



https://github.com/RuthOkolo/ProfessionalCertifications/tree/IBM-DATA-SCIENCE

Confusion Matrix



https://github.com/RuthOkolo/ProfessionalCertifications/tree/IBM-DATA-SCIENCE

Conclusions

- Findings were refined through analysis of several data sources;
- KSC LC-39A is the ideal launch point.
- While the majority of mission results are successful, the evolution of processes and rocketry appears to be leading to an improvement in successful landing outcomes with time.
- Launches heavier than 7,000 kg carry less risk.
- Profit maximization and landing success prediction are two applications for decision tree classifiers.

https://github.com/RuthOkolo/ProfessionalCertifications/tree/IBM-DATA-SCIENCE

Appendix

- All resources and relevant data sets for this project were provided by the IBM data science professional certification course.

https://github.com/RuthOkolo/ProfessionalCertifications/tree/IBM-DATA-SCIENCE

Thank you!