Winning Space Race with Data Science

Cafer Gokhan Karagoz
03-Aug-2022

Outline

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

Executive Summary

- **The following methodologies were used to analyze data:**
  - Web scraping and SpaceX API used for Data Collection
  - Exploratory Data Analysis (EDA), including data wrangling, data visualization and interactive visual analytics
  - Machine Learning Prediction
- **Summary of all results**
  - Data collected from public sources;
  - EDA enabled to identify what features are the best to predict success of launches
  - Machine Learning Prediction depicted the best model to predict what characteristics are crucial to drive this opportunity by the best way

Introduction

- Objective of the project is to evaluate and compare the abilities of Space Y to Space X and find out if Space Y can compete with Space X

- Targeted answers are;
  - Finding the best possible way to estimate the cost of launches by using predictions for successful landings of the first stage rockets
  - Identifying the best location for the launches

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected from
    - 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
  - Landing outcome label based on outcome data after summarizing and analyzing features was created in order to strenght the collected data

Methodology

Executive Summary

- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - Collected data was normalized, divided in train and test sets and following that evaluated by four different classification models.

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

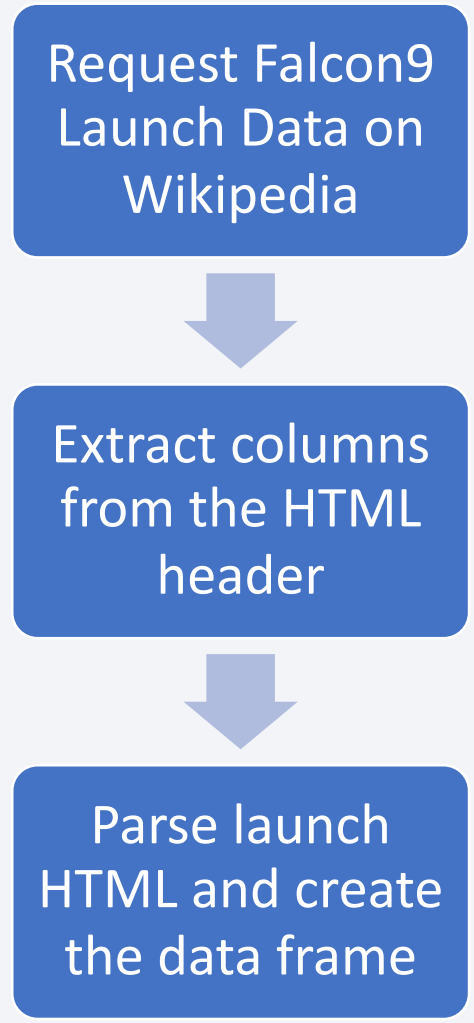
- SpaceX public data obtained and used

- Source Code:
- https://github.com/CgKaragoz/DS-Capstone-Project/blob/main/Data Collection API.ipynb

Data Collection - Scraping

- Wikipedia is another source SpaceX launches data

- Source Code:
- https://github.com/CgKaragoz/DS-Capstone-Project/blob/main/Web Scrapping.ipynb



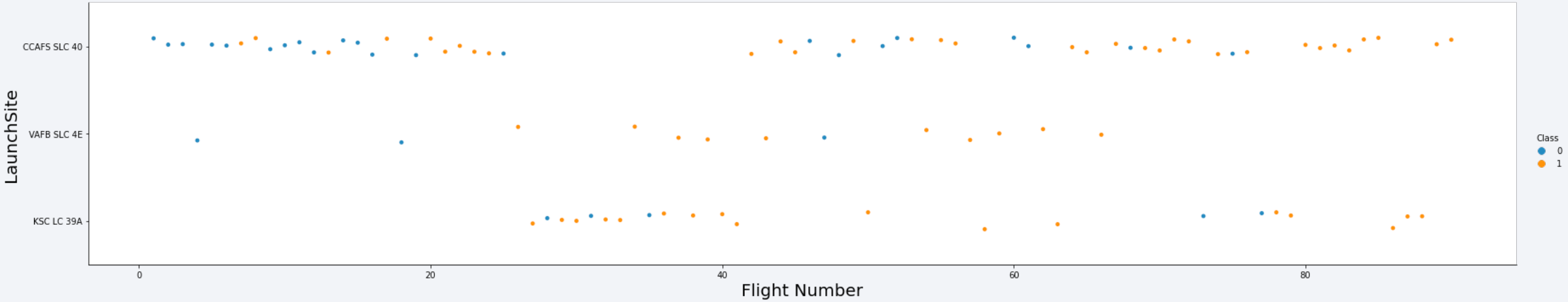
Data Wrangling

- Exploratory Data Analysis was performed
- Summaries of launches per site and occurrences of orbits were calculated
- Landing outcome label was created

Source Code: https://github.com/CgKaragoz/DS-Capstone-Project/blob/main/Data Wrangling.ipynb

EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
- Add the GitHub URL of your completed EDA with data visualization notebook, as an external reference and peer-review purpose



Source Code:https://github.com/CgKaragoz/DS-Capstone-Project/blob/main/EDA with Visualization.ipynb

EDA with SQL

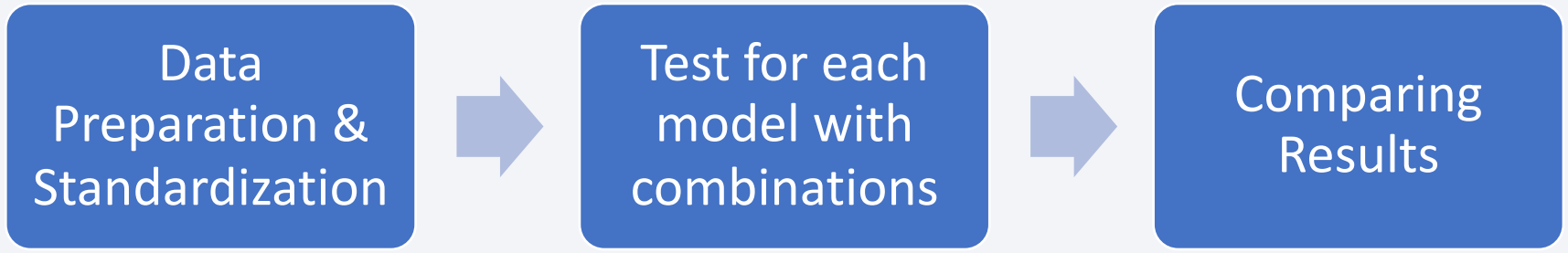
- SQL queries performed are depicted below;
  - Unique launch sites in the space mission;
  - Top 5 launch sites names begin with the string 'CCA';
  - Total payload mass carried by NASA (CRS);
  - Average payload mass carried by booster version F9 v1.1;
  - Date when the first successful landing outcome in ground pad was achieved;
  - Names of the boosters which have success in drone ship and have payload mass between 4000 and 6000 kg;
  - Total number of successful and failure mission outcomes;
  - Names of the booster versions which have carried the maximum payload mass;
  - Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015; and
  - 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
- Source Code: [https://github.com/CgKaragoz/DS-Capstone-Project/blob/main/Exploratory Analysis Using SQL.ipynb](https://github.com/CgKaragoz/DS-Capstone-Project/blob/main/Exploratory%20Analysis%20Using%20SQL.ipynb)

Build an Interactive Map with Folium

- Below items were used with Folium
  - Markers i.e. Launch sites
  - Circles (areas around specific coordinates)
  - Marker Clusters i.e. launches in any given launch site
  - Lines which indicate distances between coordinates
- Source Code: [https://github.com/CgKaragoz/DS-Capstone-Project/blob/main/Interactive Visual Analytics with Folium lab.ipynb](https://github.com/CgKaragoz/DS-Capstone-Project/blob/main/Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb)

Predictive Analysis (Classification)

- Logistic regression, support vector machine, decision tree and k nearest neighbors were compared



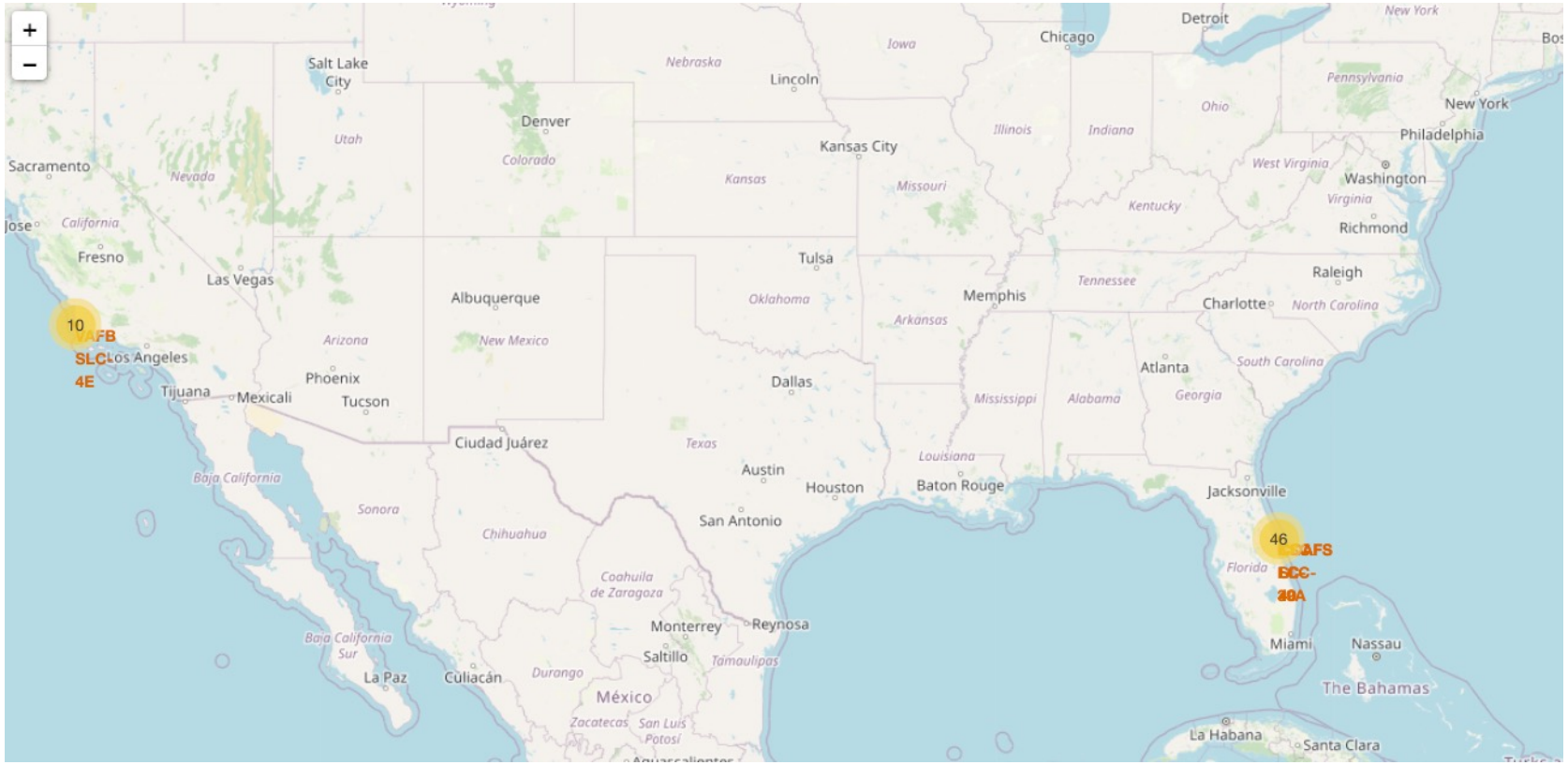
- Source Code: https://github.com/CgKaragoz/DS-Capstone-Project/blob/main/Machine Learning Prediction.ipynb

Results

- Exploratory data analysis results
  - There are 4 launch sites Space X have
  - In the beginning launches were done for Space X and NASA
  - F9 v1.1 average payload is 2928 kg
  - In 2015 first successful launched happened
  - Approximately %100 successful rate achieved
  - Failed boosters are F9 v1.1 B1012 and F9 v1.1 B1015 in 2015

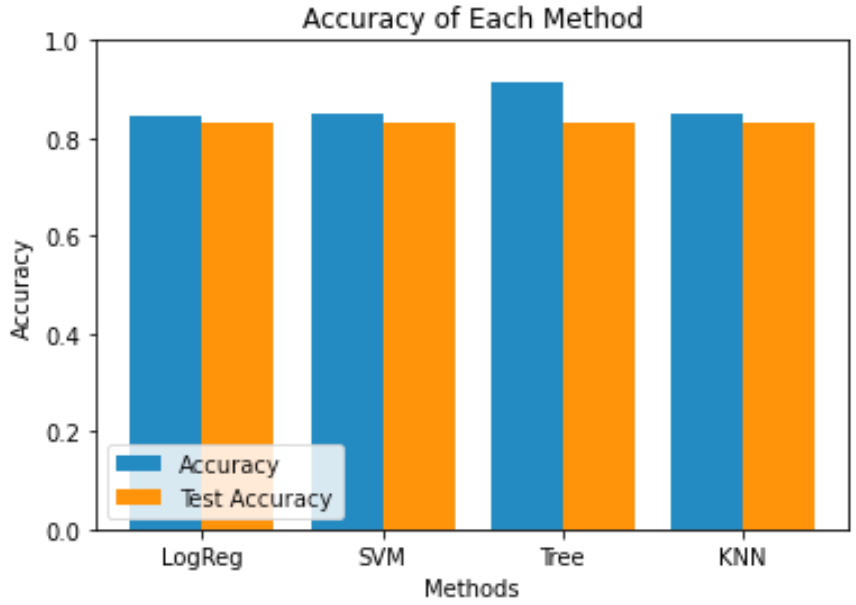
Results

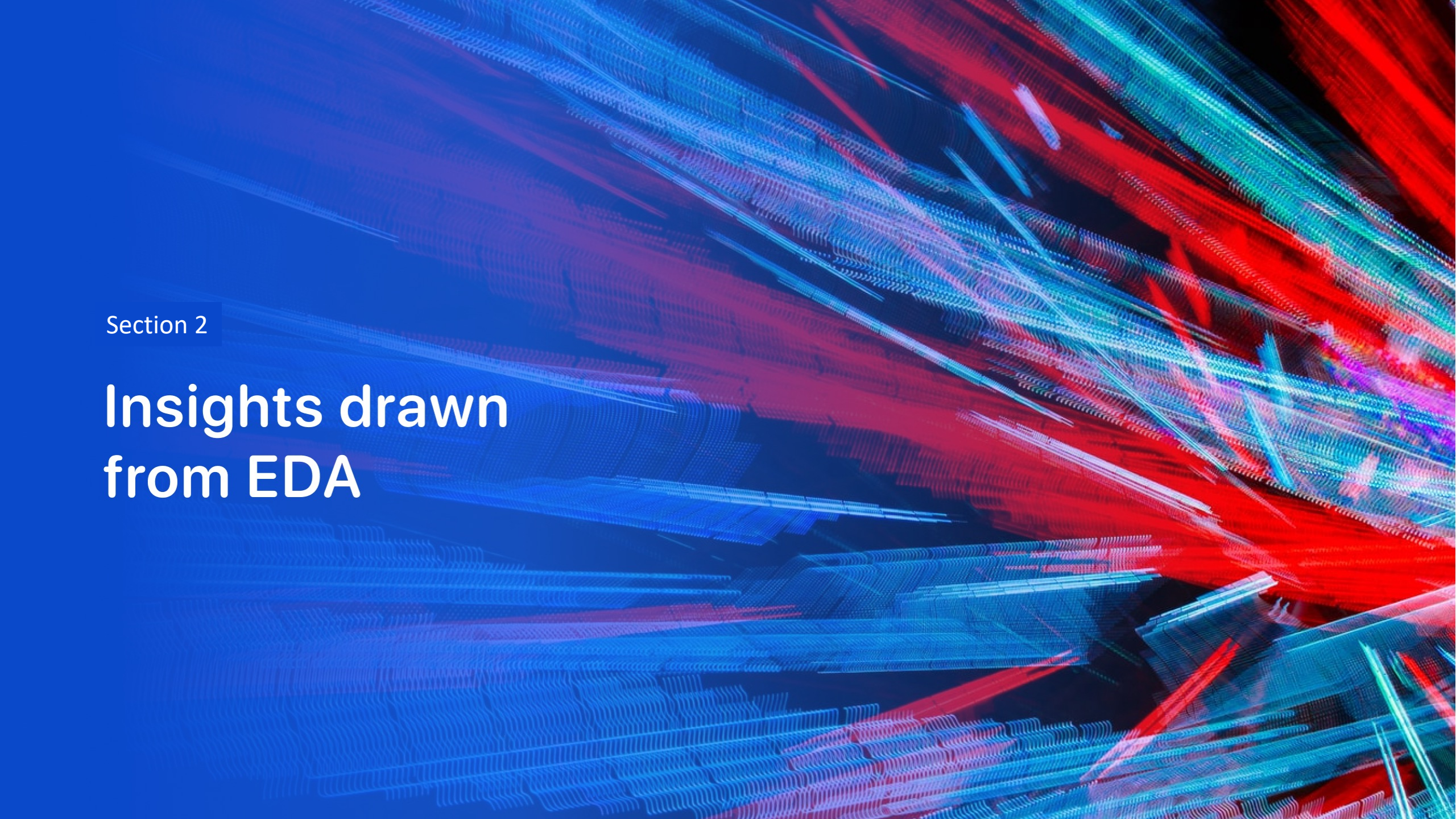
- Using interactive analytics was possible to identify that launch sites use to be in safety places, near sea, for example and have a good logistic infrastructure around.
  - Most launches happens at east cost launch sites



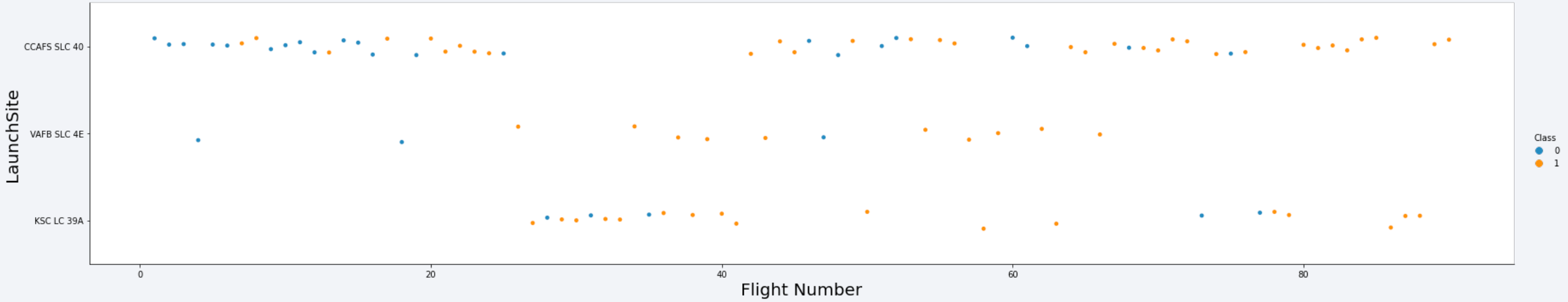
Results

- Predictive Analysis showed that Decision Tree Classifier is the best model to predict successful landings, having accuracy over 87% and accuracy for test data over 94%.



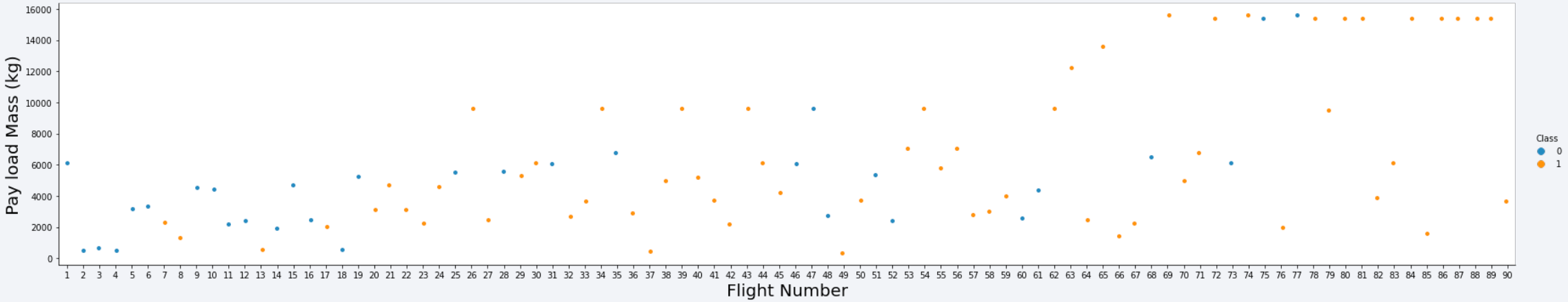

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



- it’s possible to confirm that the best launch site currently is CCAF5 SLC 40
- Second one is VAFB SLC 4E and third one is KSC LC 39A;
- Success rate improved over the period time

Payload vs. Launch Site



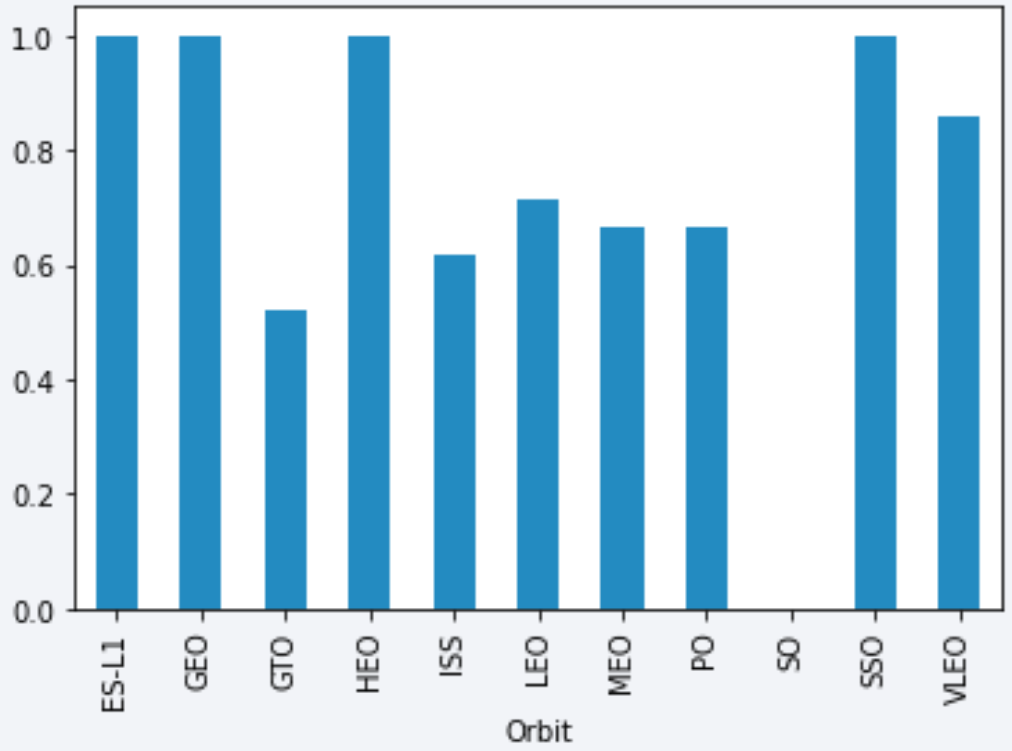
- Payloads over 9,000kg have perfect success rate;
- Payloads over 12,000kg are only possible with CCAFS SLC 40 and KSC LC 39A launch sites

Success Rate vs. Orbit Type

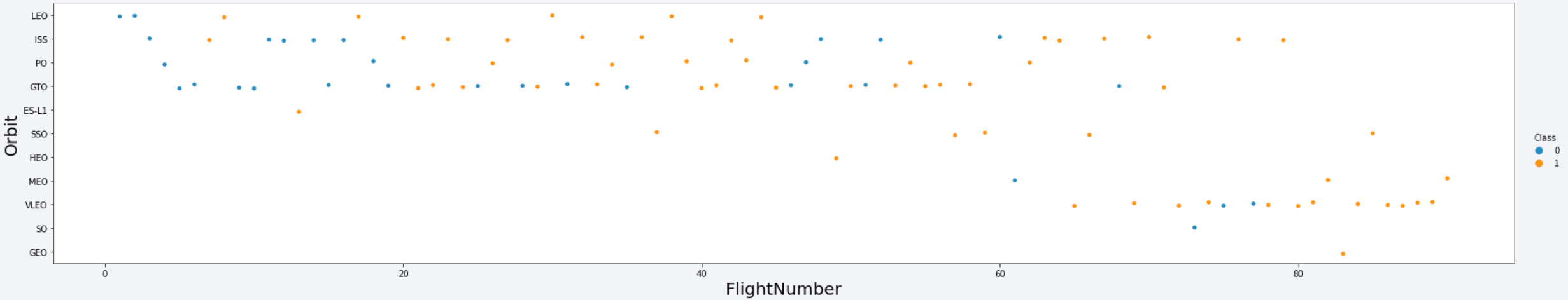
• The biggest success rates happens to orbits:

• ES-L1
• GEO
• HEO and
• SSO

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

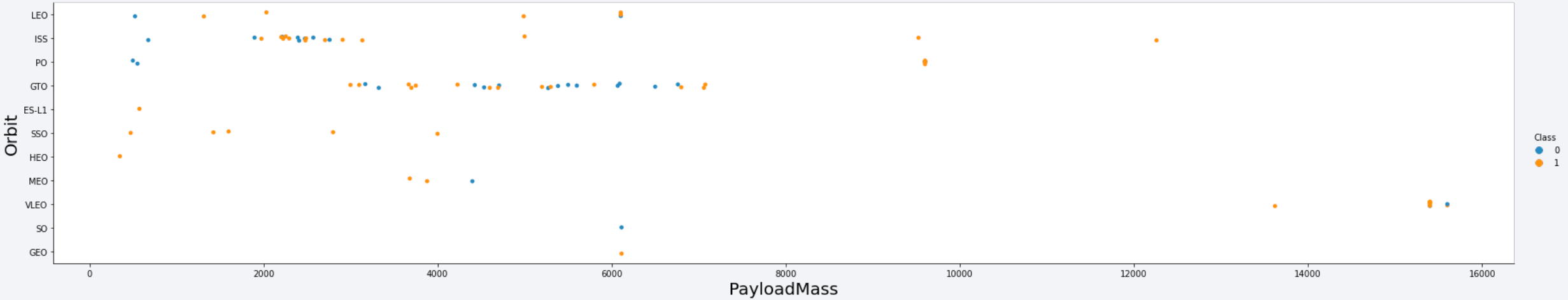


Flight Number vs. Orbit Type



- Success rate improved over time to all orbits
- VLEO orbit might be a new business opportunity. There is an increase of its frequency

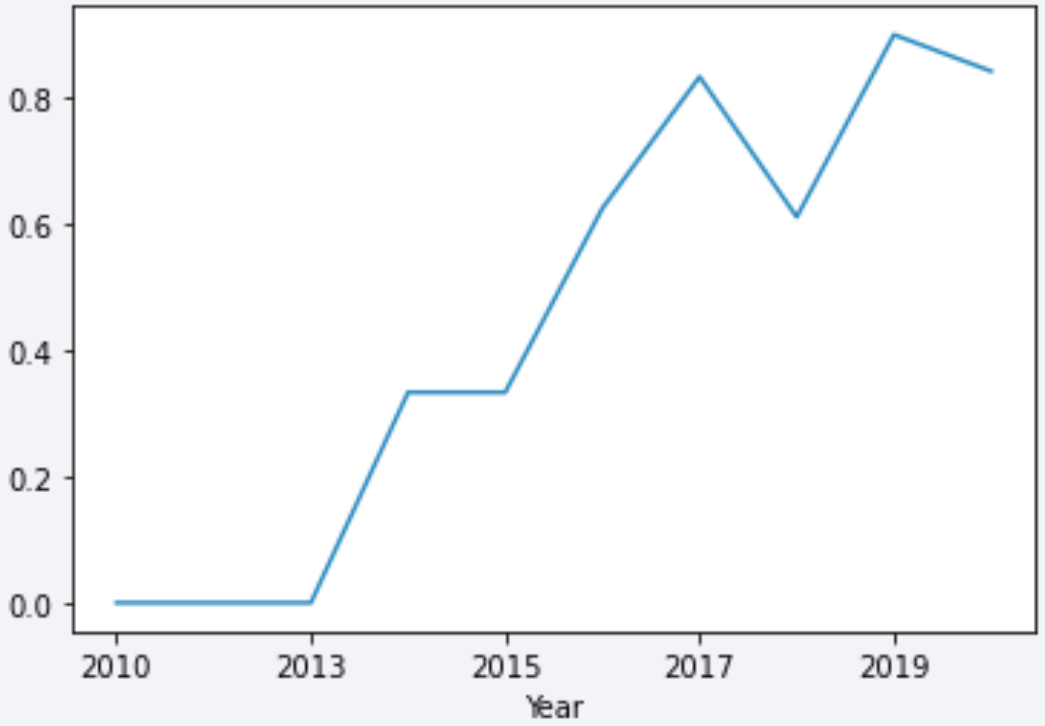
Payload vs. Orbit Type



- No correlation between payload and success rate to orbit GTO
- ISS orbit has the widest range of payload
- A few launches to the orbits SO and GEO

Launch Success Yearly Trend

- Success rate has been increasing from 2013 until 2020;
- Between 2010 and 2013 was an improvement of technology.



All Launch Site Names

- There are 4 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'

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

- 5 Cape Canaveral launches

Total Payload Mass

Total Payload (kg)
111.268

- Sum of all payloads whose codes contain ‘CRS’, which corresponds to NASA gives us total payload above.

Average Payload Mass by F9 v1.1

Avg Payload (kg)
2.928

- Booster version data filtered

First Successful Ground Landing Date

Min Date
2015-12-22

- Successful landing outcome on ground pad and getting the minimum value for date which has happened on 12/22/2015.

Successful Drone Ship Landing with Payload between 4000 and 6000

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

Total Number of Successful and Failure Mission Outcomes

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

- Number of Successful and Failed missions
- Mission outcomes & records for each group depicts the summary.

Boosters Carried Maximum Payload

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

- Maximum payloads filtered by boosters

2015 Launch Records

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

- Two occurances shown on the list above

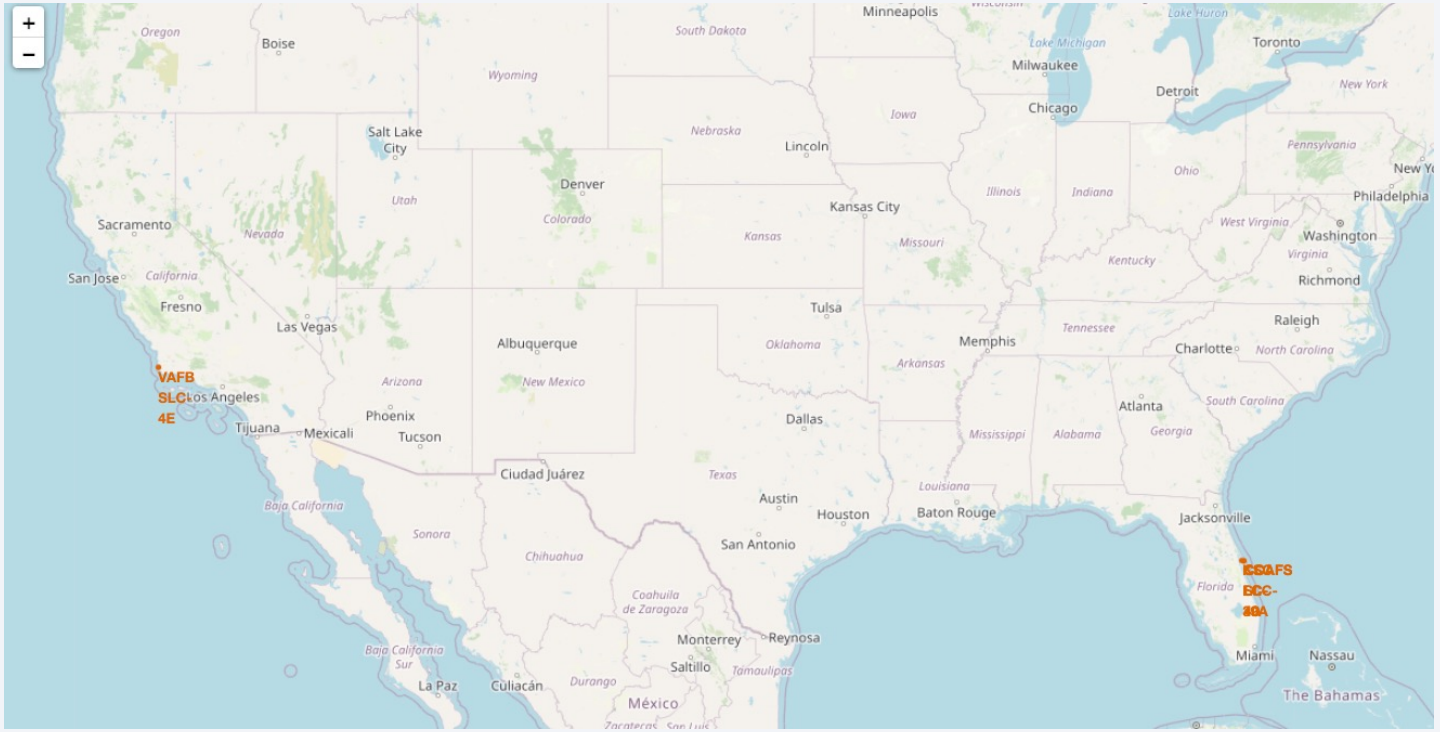
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

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

- All landings between 2010-06-04 and 2017-03-20:


Section 3
Launch Sites
Proximities Analysis

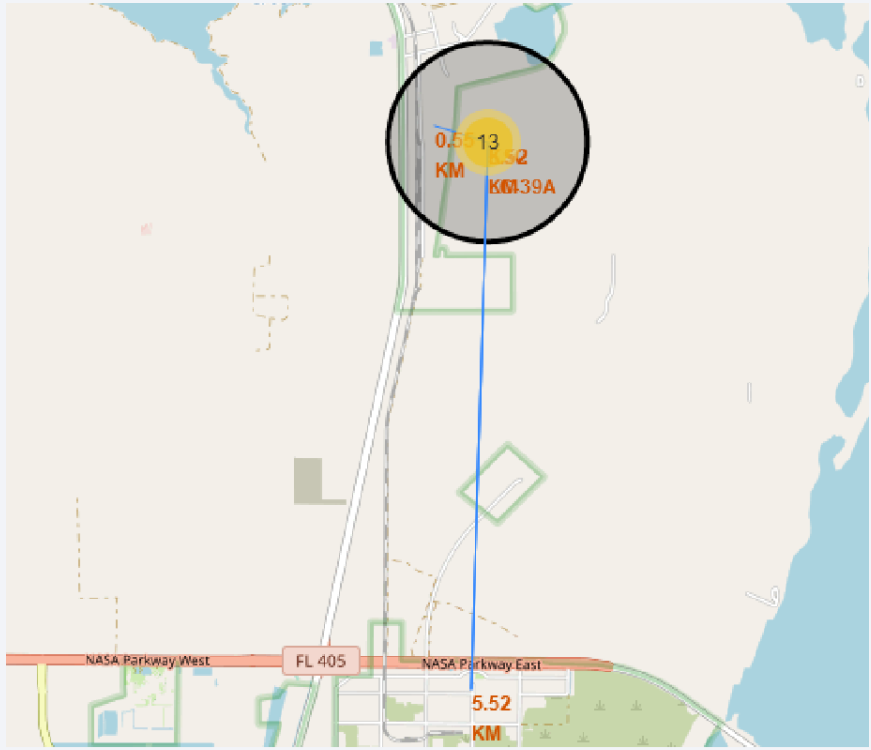
Launch Sites



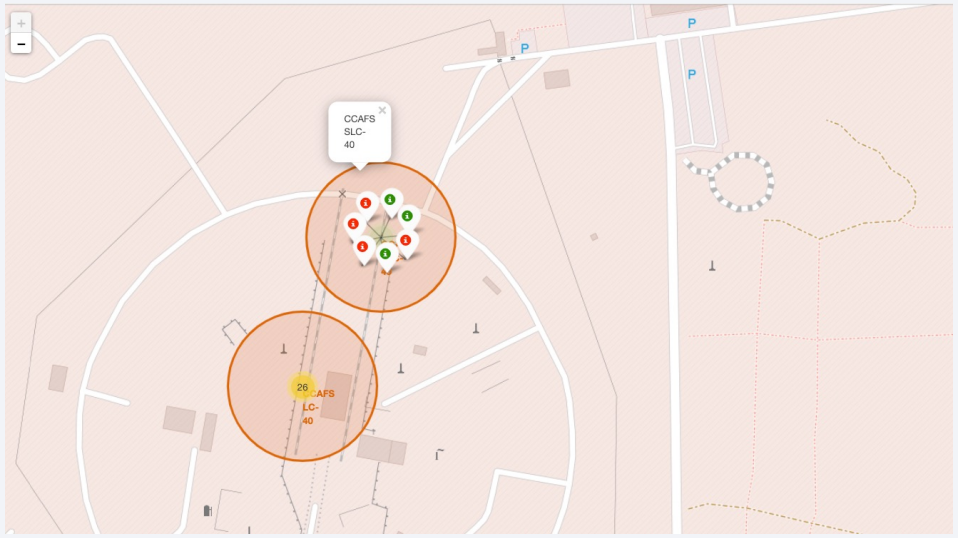
- Locations are on the coast lines and close to the roads and railways

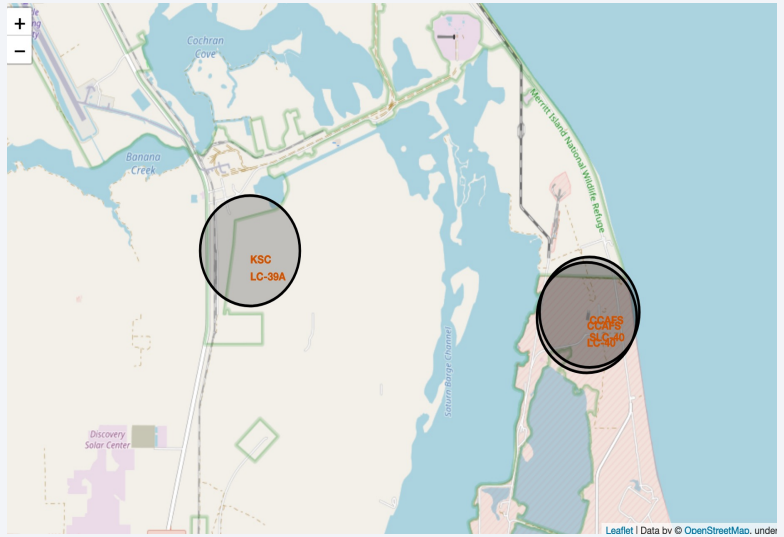
Safety and Close Proximity for Logistics

- KSC LC-39A is in the close proximity of the railroads

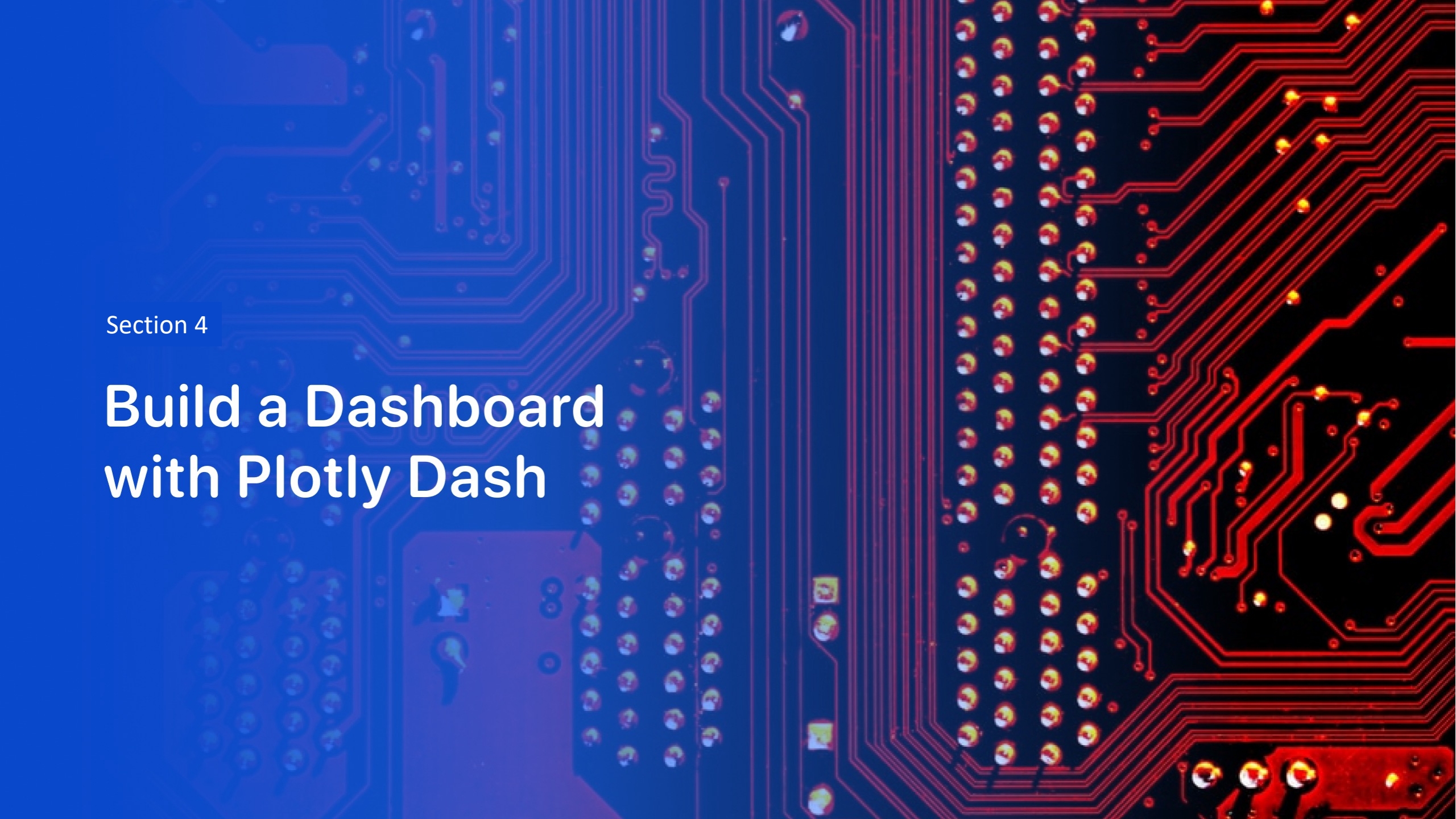


<Folium Map Screenshot 3>

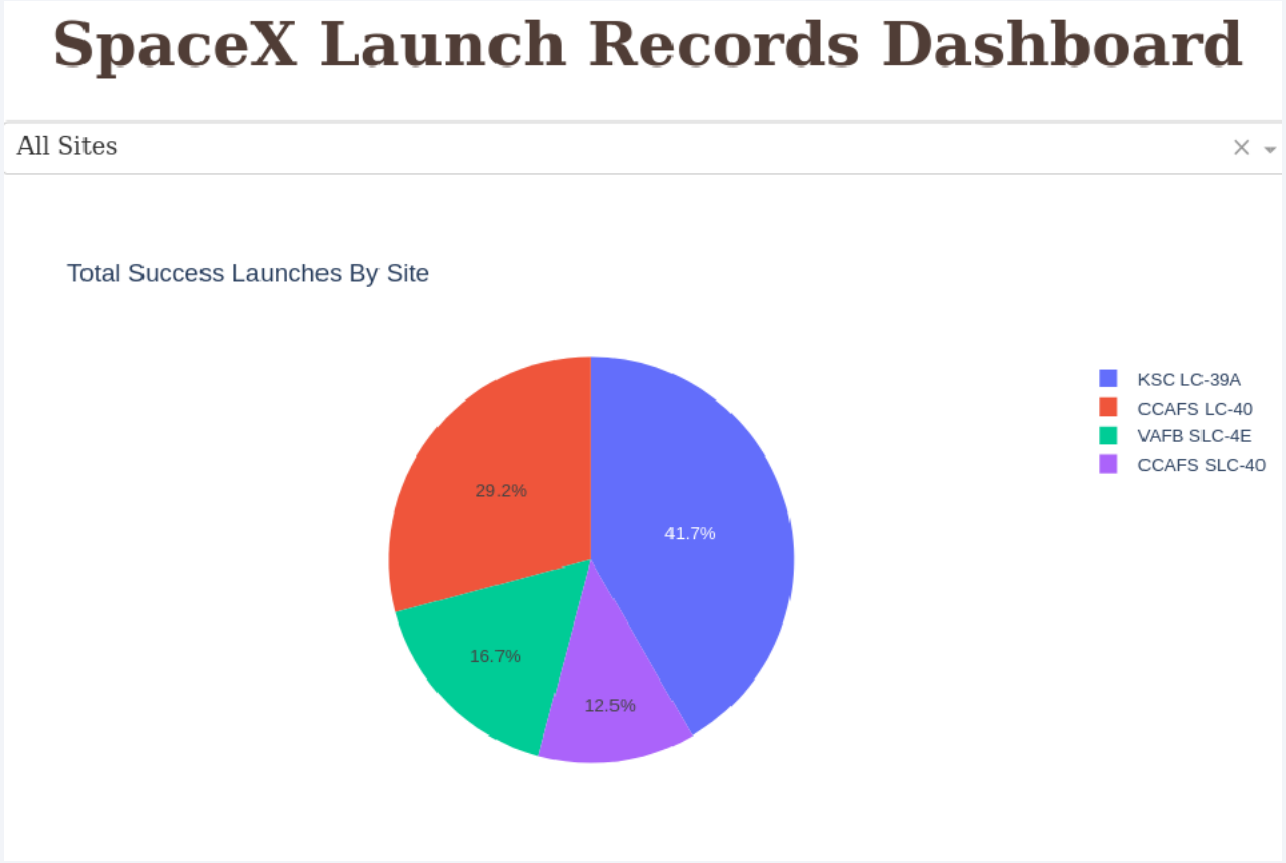




- Green ones are successful and red ones are failures

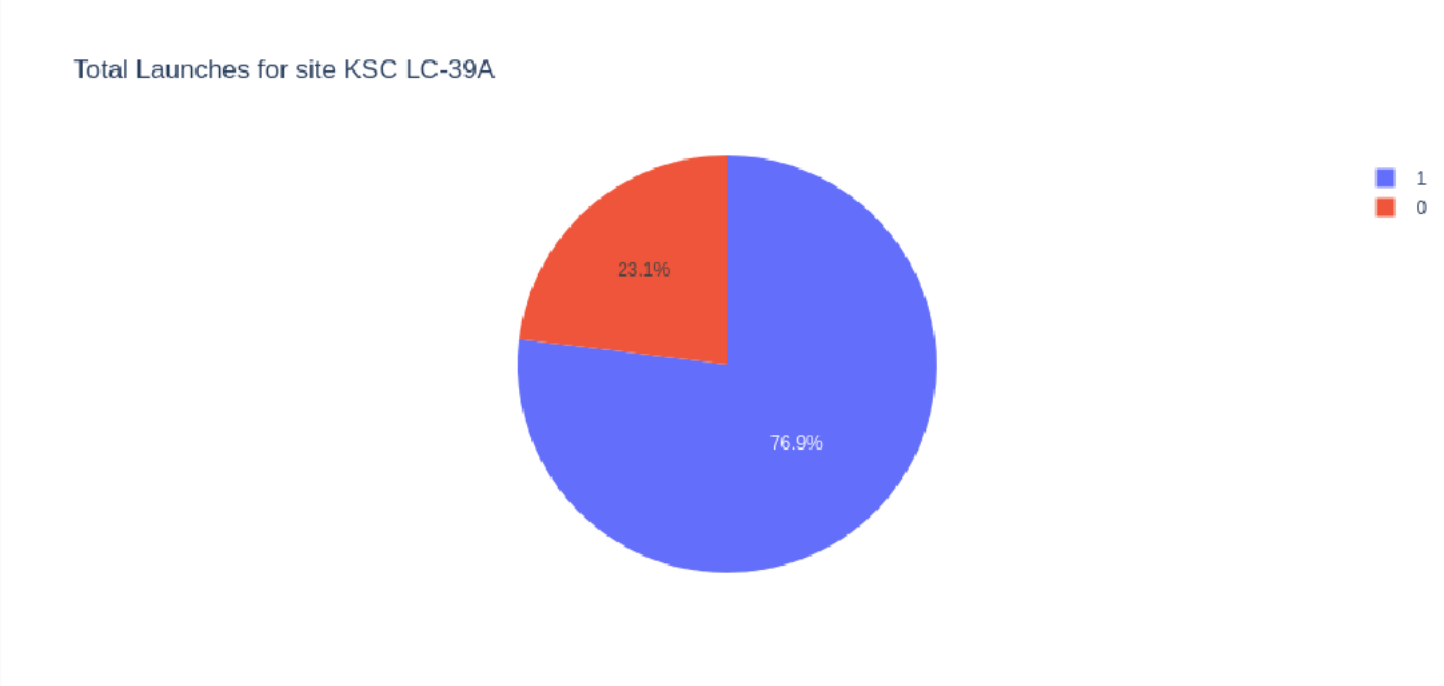

Section 4
Build a Dashboard with Plotly Dash

Launches by Sites



- Location of the lunch site is highly important for success

Success Ratio for KSC LC-39A



- Approximately %77 success rate

Payload vs Launch

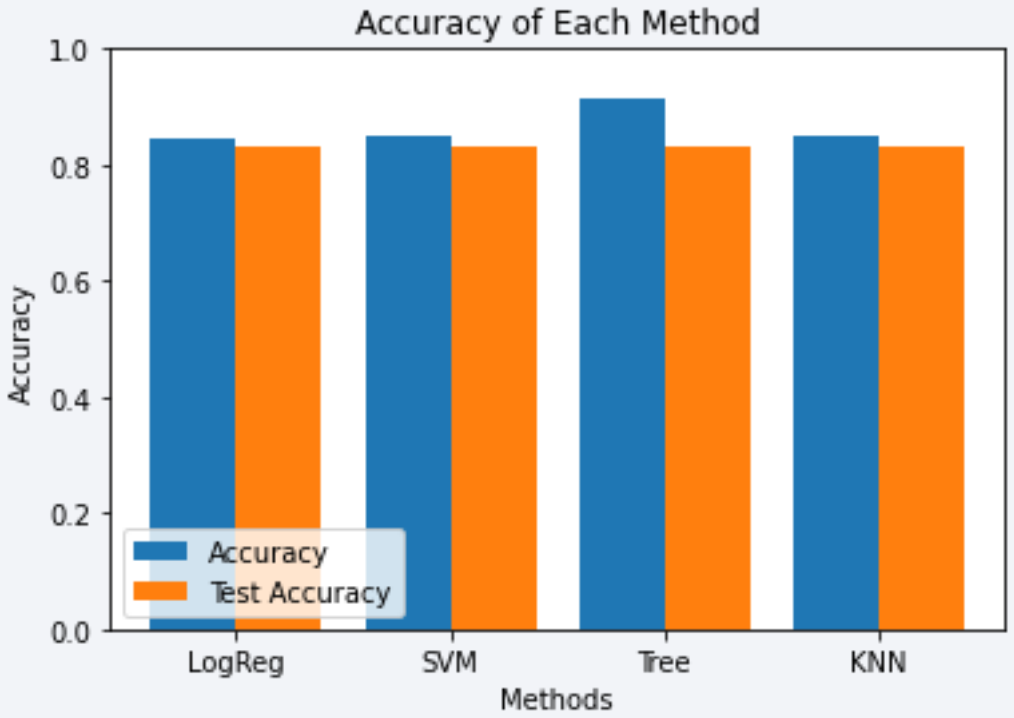
- The best combination is FT boosters with below 6000kg payloads

Section 5

Predictive Analysis (Classification)

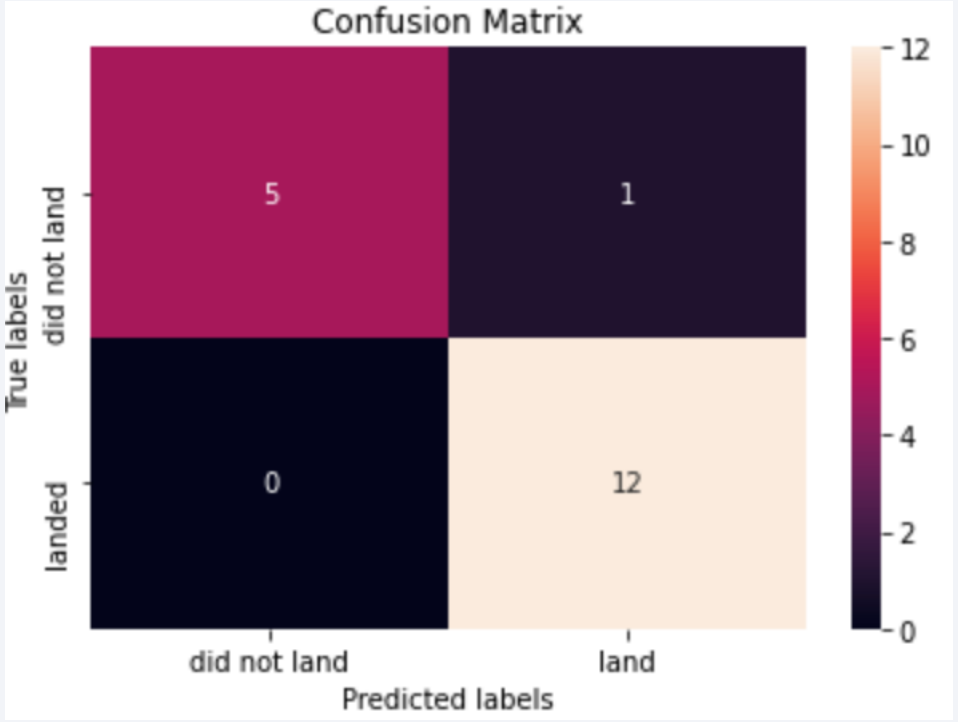
Classification Accuracy

- Decision Tree Classifier is the most accurate model of 4 tested with an accuracy rate of %87



Confusion Matrix

- Decision Tree Classifier proves its accuracy



Conclusions

- Multiple data sources were analyzed, refining conclusions along the process;
- KSC LC-39A is the best launch site
- Above 7,000kg are less risky;
- Successful landing outcomes improved over time. It might be due to the improvements of processes and rockets.
- Decision Tree Classifier might be used to predict successful landings and increase profits.

Appendix

- Folium does not show some maps on Github, that is the reason there screenshots of the maps in the presentation.

Thank you!