IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Isha Gupta
1st Sept'22

Outline

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

Executive Summary

➢ **Methodologies Used:**

- Data Collection using Webscraping and SpaceX API.
- Exploratory Data Analysis (EDA), including data wrangling, SQL, data visualization and interactive visual analytics.
- Predictive Analysis using Machine Learning.

➢ **Summary of all results:**

- Collected rocket launch data from SpaceX API and web scraping to collect Falcon 9 historical launch records from a Wikipedia page.
- EDA allowed to identify relationship between important variables that would affect the success rate and which features are the best to predict success of launchings.
- Machine Learning identified the best model to predict which characteristics are important to drive the goal.

Introduction

Objective:

Calculate the probability of Spacy Y on competing with Space X where Space X shows capability of reusing stage1 and saving million of dollars compared to space Y

Desirable answers:

- Estimating total cost of launches can be determine from successful number of launches in first stage
- Identifying best place for launch of rockets.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:

Space X data was collected 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

Collected data was formatted by creating a landing outcome label based on outcome data after summarizing and analysing 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

Collected data in above steps was normalized, divided into training and test data sets and evaluated by four different classification models, also the accuracy of each model evaluated using different combinations of parameters.

Data Collection

- Describe how data sets were collected.

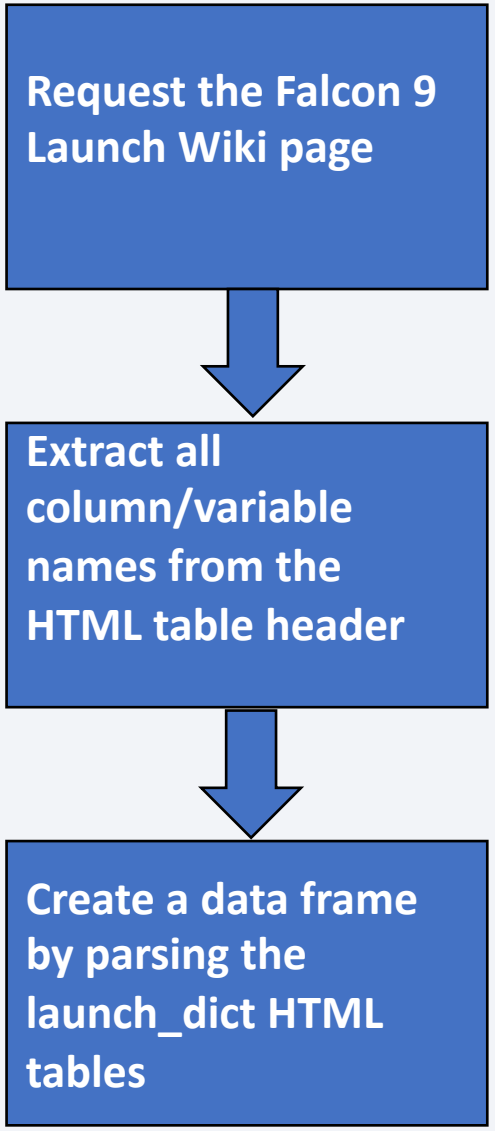
Data sets were collected from Space X API (https://api.spacexdata.com/v4/rockets/) and from Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches), using web scraping technics.

Data Collection – SpaceX API

- SpaceX offers a public API from where data can be obtained
- This API was used according to the flowchart beside and then data is persisted.
- Source code: https://github.com/ishaguptaibm/Capstone-project/blob/master/jupyter-labs-spacex-data-collection-api%20(1).ipynb

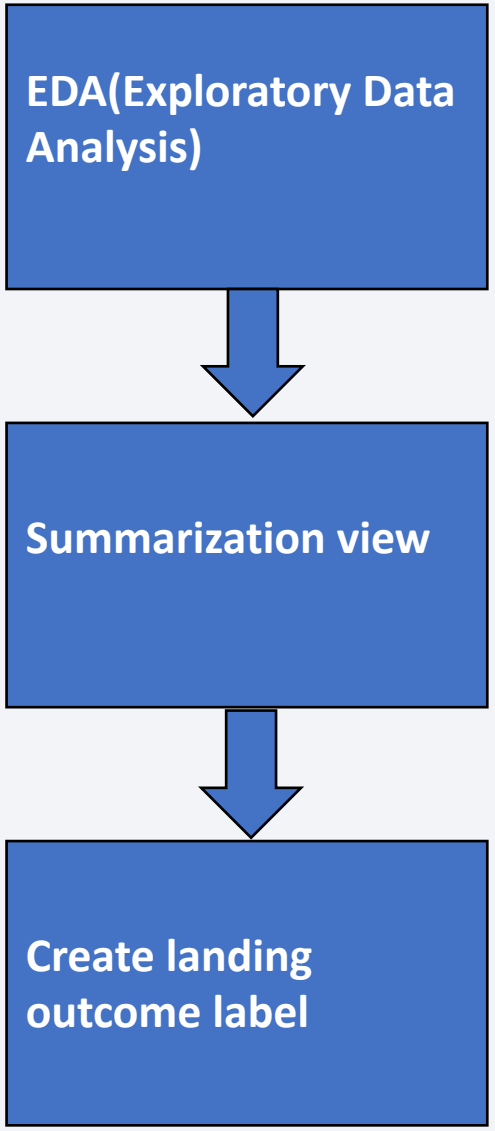
Data Collection - Scraping

- Data from SpaceX launches can also be obtained from Wikipedia
- Data was downloaded from Wikipedia according to the flowchart and then persisted.
- Source code: https://github.com/ishaguptaibm/Capstone-project/blob/master/jupyter-labs-webscraping%20(1).ipynb



Data Wrangling

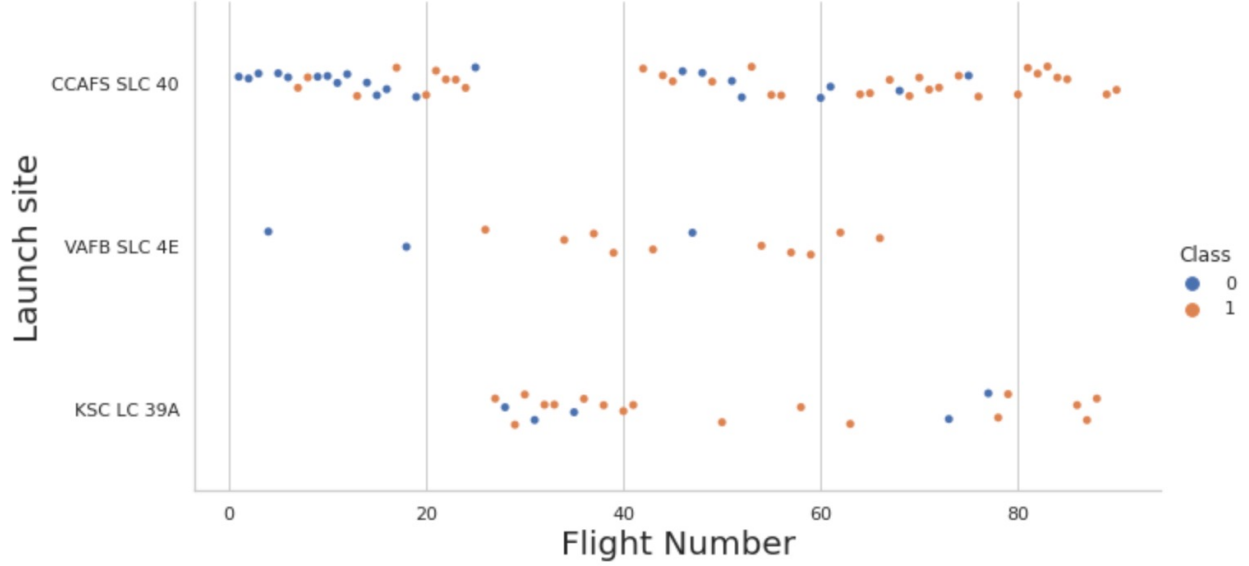
- Exploratory Data Analysis (EDA) was performed on the dataset.
- Calculated summaries-number of launches on each site, number and occurrences of each orbit and mission outcome per orbit type.
- Finally, the landing outcome label was created from Outcome column into binary form.
- Source Code: https://github.com/ishaguptaibm/Capstone-project/blob/master/labs-jupyter-spacex-Data%20wrangling%20(1).ipynb



EDA with Data Visualization

- To explore data, scatterplots and barplots were used to visualize the relationship between pair of features:

1. Payload Mass relationship with Flight Number
2. Launch Site relationship with Flight Number
3. Launch Site relationship with Payload Mass
4. Orbit relationship with Flight Number
5. Payload relationship with Orbit



Source Code:
https://github.com/ishaguptaibm/Capstone-project/blob/master/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

SQL Queries:

- Names of the unique launch sites in the space mission
- Top 5 launch sites whose name 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
- 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
- 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/ishaguptaibm/Capstone-project/blob/master/jupyter-labs-eda-sql-coursera%20(1).ipynb

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.

Source Code: https://github.com/ishaguptaibm/Capstone-project/blob/master/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

The following graphs and plots were used to visualize data

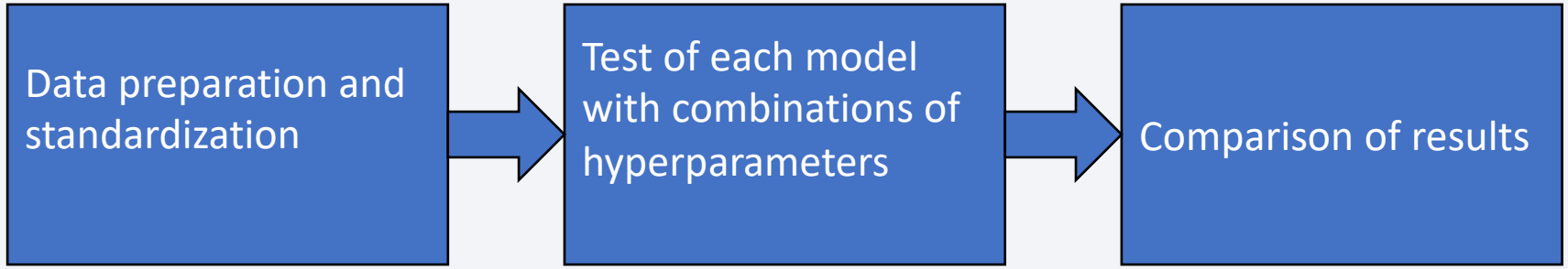
- Percentage of launches by site
- Payload range

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

Source Code: https://github.com/ishaguptaibm/Capstone-project/blob/master/spacex_dash_app.py

Predictive Analysis (Classification)

- Four classification models were compared: logistic regression, support vector machine, decision tree and k nearest neighbors.



Source Code: https://github.com/ishaguptaibm/Capstone-project/blob/master/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

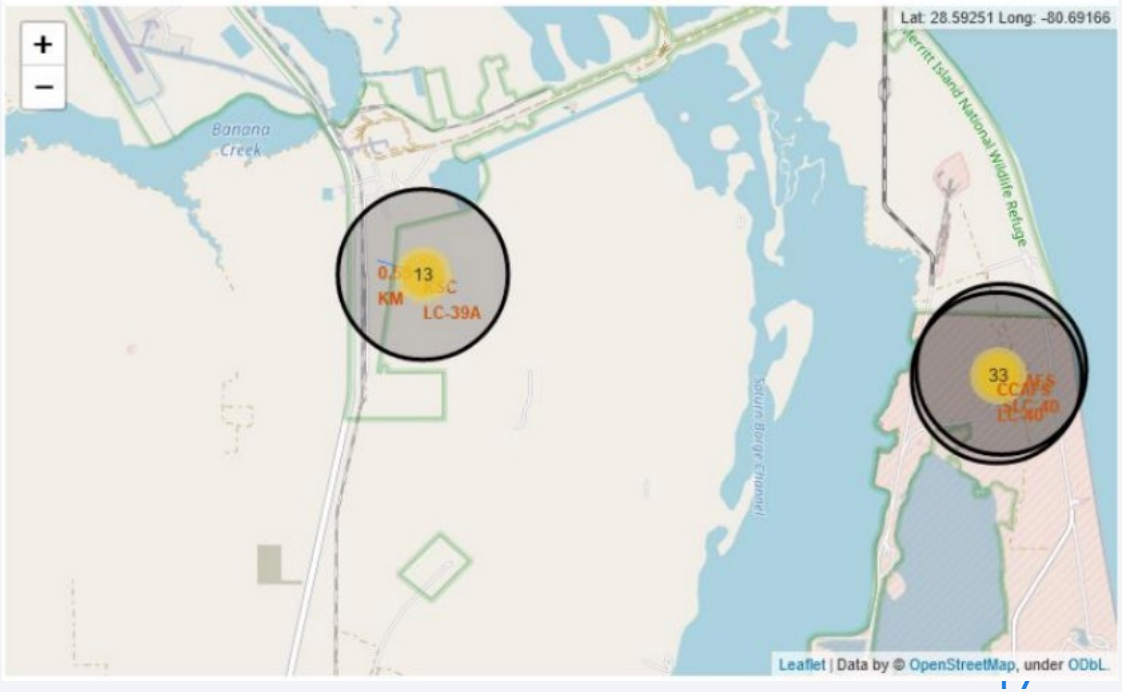
➢Exploratory data analysis results

- 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;
- 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;
- The number of landing outcomes became as better as years passed.
- Predictive analysis results

Results

➢Interactive analytics demo in screenshots

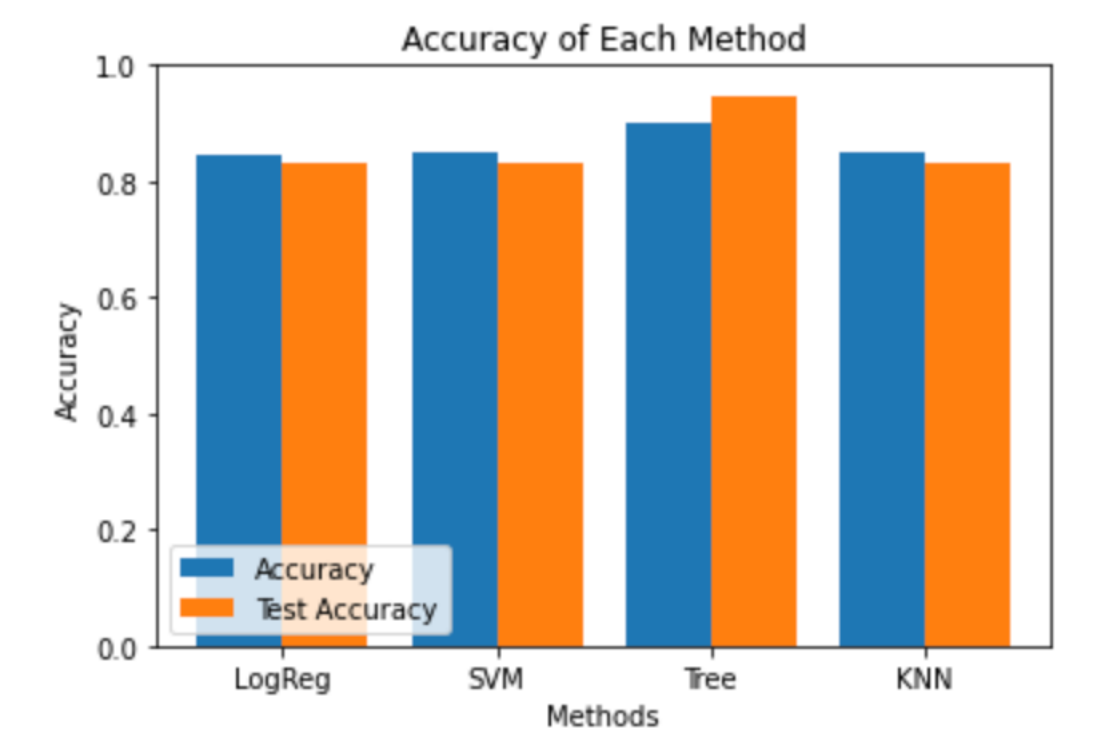
- Using interactive analytics was possible to identify that launch sites use to be in safety places, near sea, for example and have a good logistic infrastructure around.
- Most launches happens at east cost launch sites.

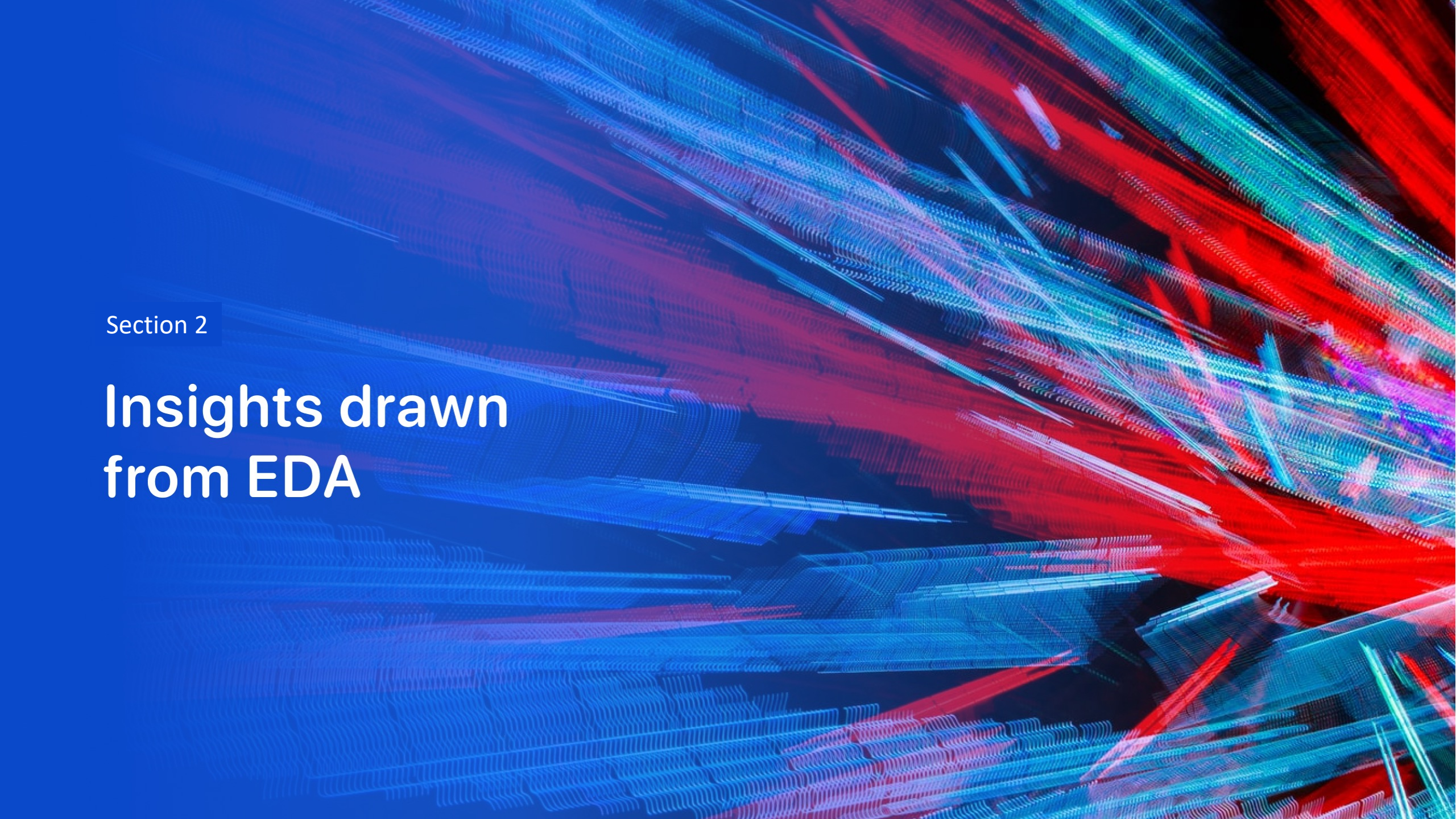


Results

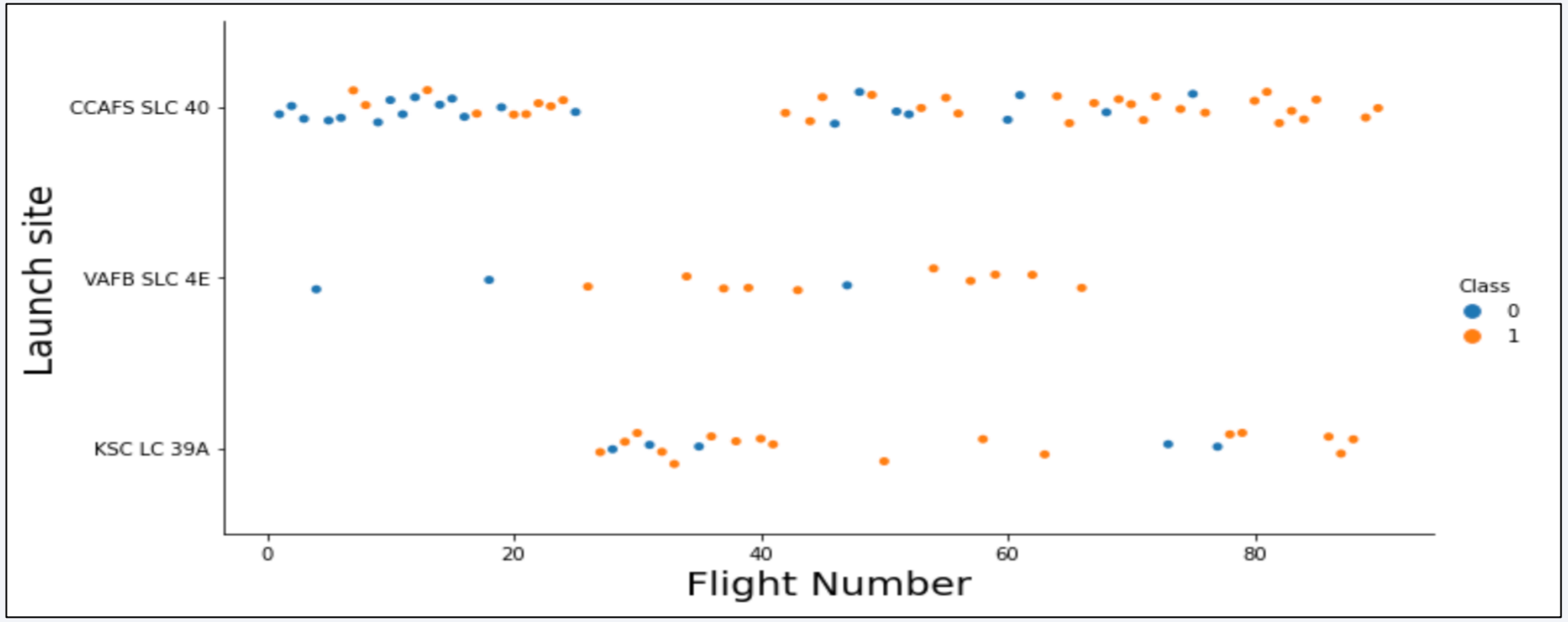
➢Interactive analytics demo in screenshots

Predictive Analysis showed that Decision Tree Classifier is the best model to predict successful landings, having accuracy over 90% and accuracy for test data over 94%.



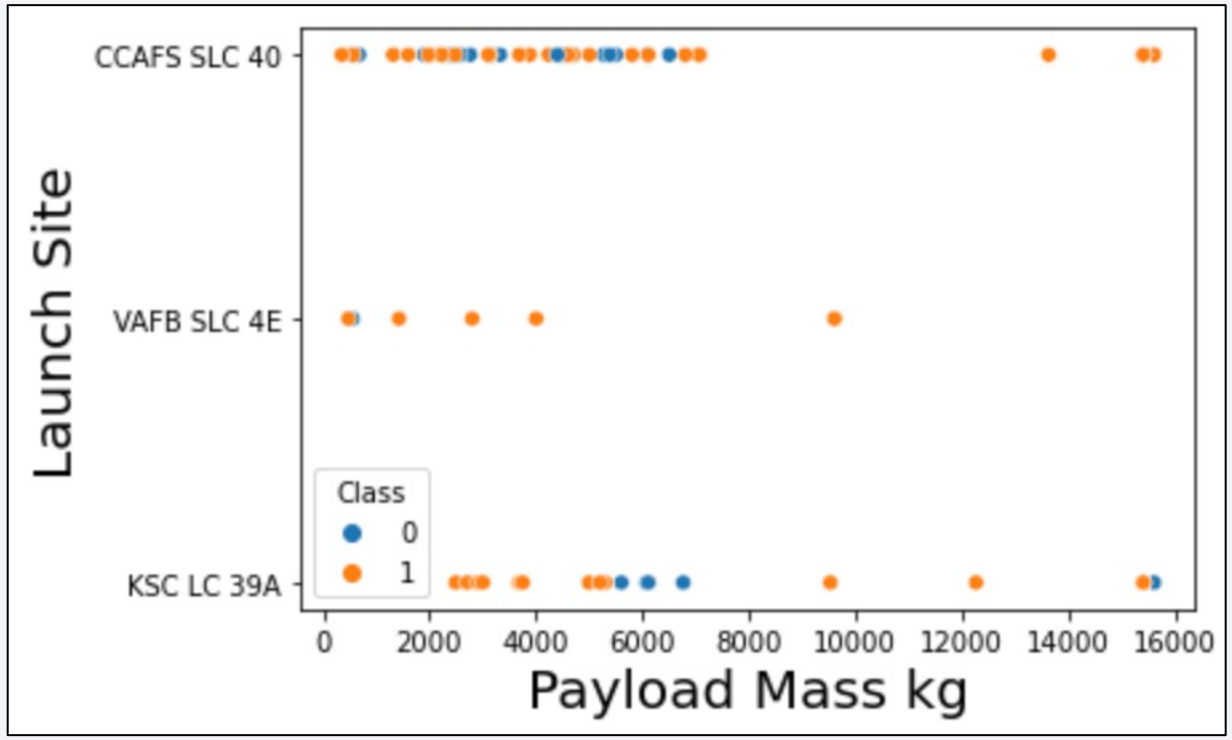

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



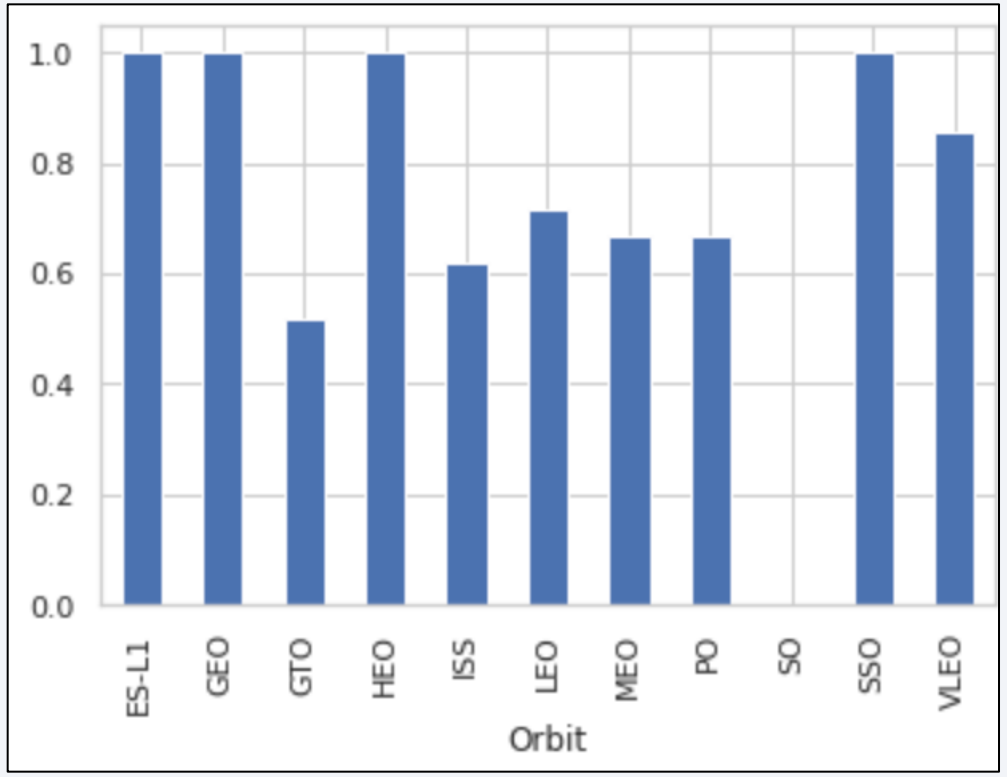
- From above graph, we can see that best launch site is CCAF5 SLC 40, where most of the recent launches were successful. On second, VAFB SLC 4E and third KSC LC 39A.
- Success rate of launch has increased with time on different launch sites.

Payload vs. Launch Site



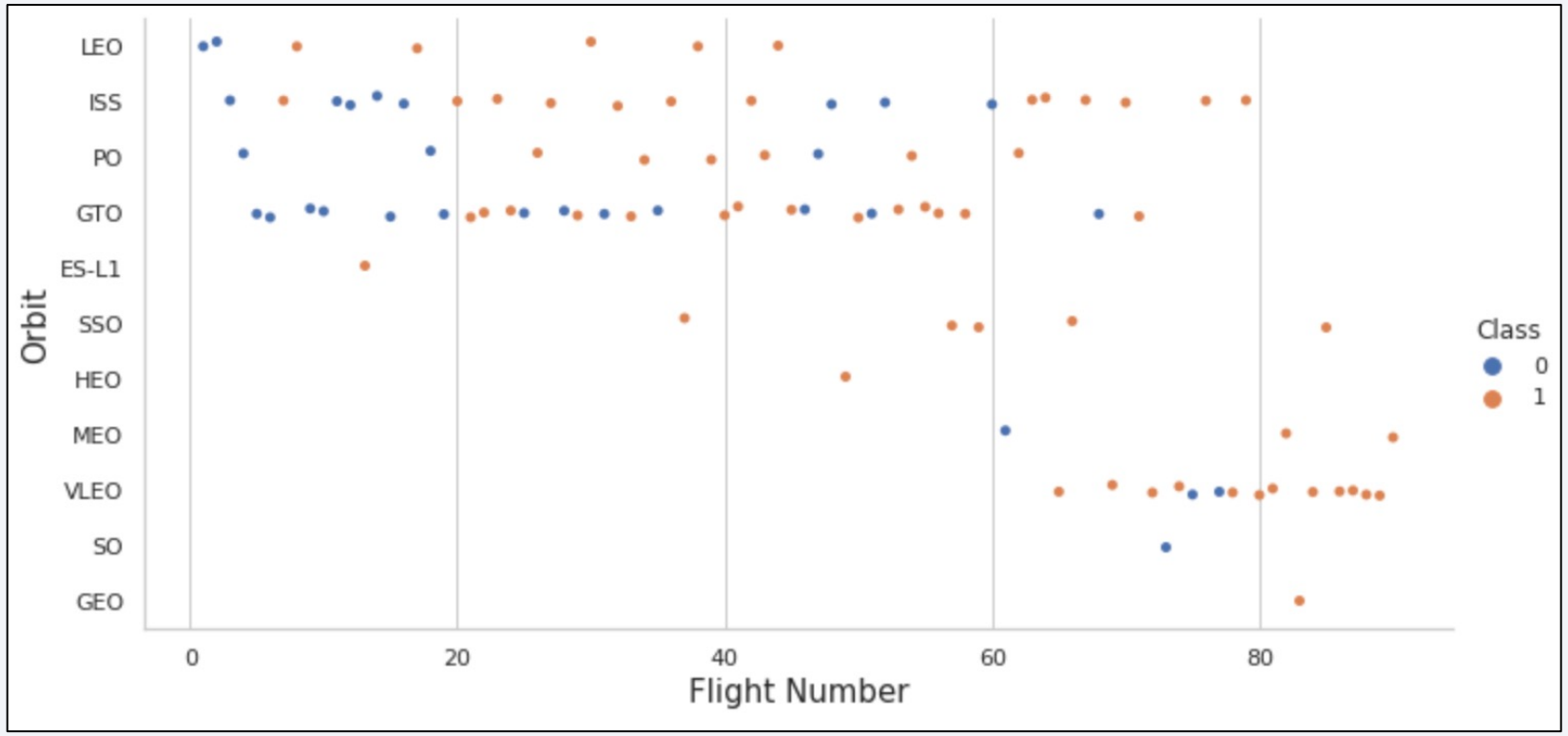
- Payload mass over 9000kg have excellent success rate
- Payload mass over 12000kg looks possible only on CCAFS SLC 40 and KSC LC 39A launch sites.

Success Rate vs. Orbit Type



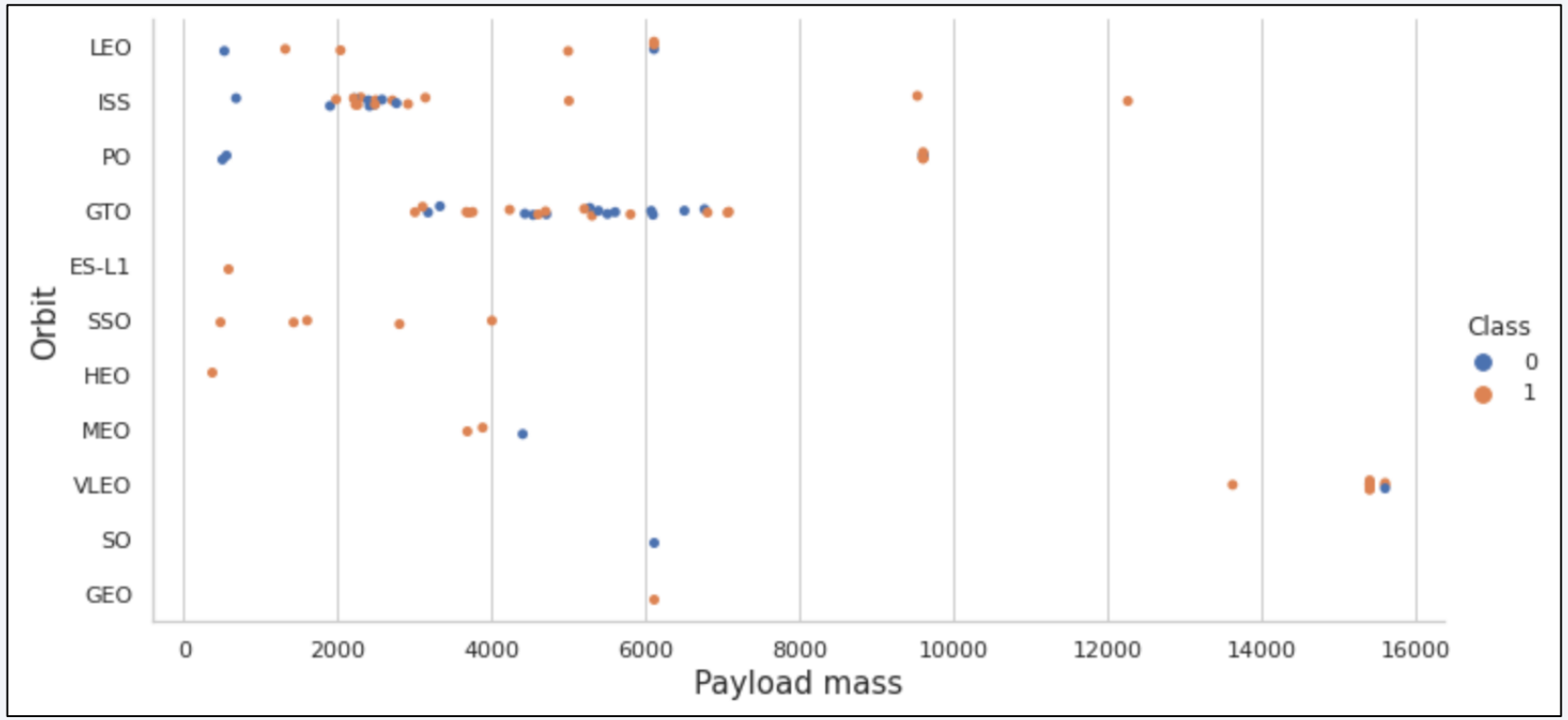
- Highest success rate is coming from ES-L1, GEO, HEO and SSO.
- 80% rate coming from VLEO and 70% from LEO

Flight Number vs. Orbit Type



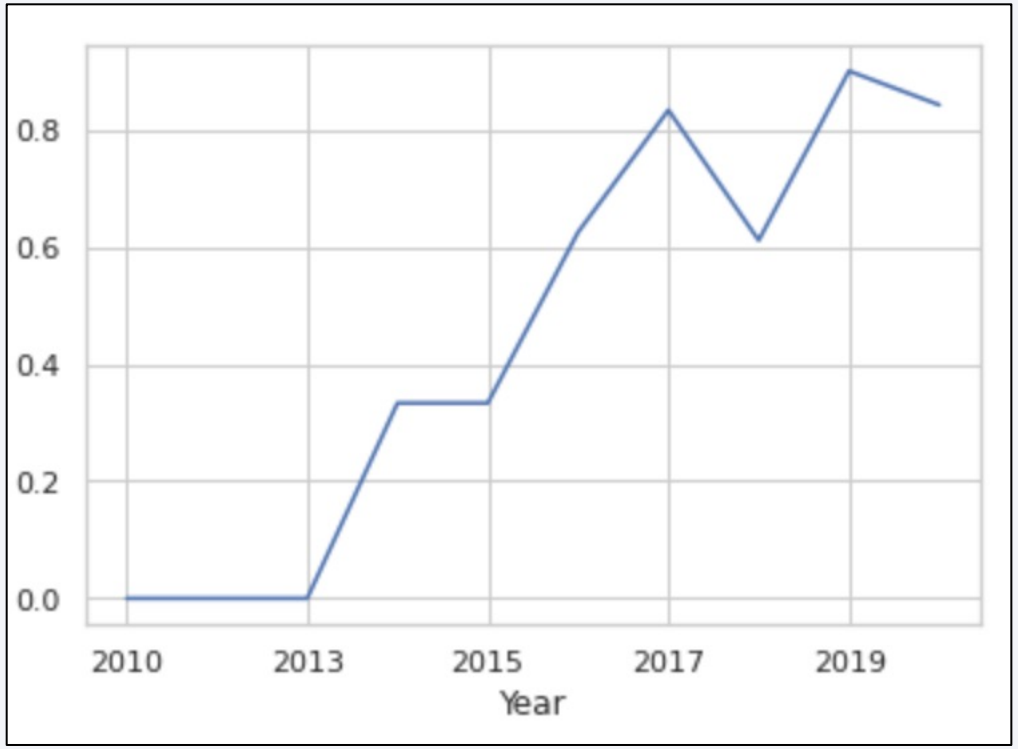
- VLEO orbit has a good success rate over the time.
- After several launches performance has increased in the orbits over the time.

Payload vs. Orbit Type



- ISS showed good rate of success with variety of payload mass.
- GTO has mixed launches(failed and successful).
- SSO was not tried with payload mass>4000 kg

Launch Success Yearly Trend



Success rate started increasing in 2013 and kept until 2020.

All Launch Site Names

There are 4 launch sites:

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

They are unique launch sites extracted from dataset.

Launch Site Names Begin with 'CCA'

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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)
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-12	22:41:00	F9 v1.1	CCAFS LC-40	SES-8	3170	GTO	SES	Success	No attempt

Above is the sample of 5 records where Launch Site starts with name “CCA”

Total Payload Mass

Total payload mass carried by boosters from NASA-CRS:

payloadmass
167747

Total payload calculated above, 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:

payloadmass
3209

Filtering data by the booster version above and calculating the average payload mass we obtained the value of 3209 kg.

First Successful Ground Landing Date

First successful landing outcome on ground pad:

firstsuccess
2015-12-22

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 22/12/2015.

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

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

There are 2 boosters that have successfully landed with payload mass b/w 4000 and 6000

Total Number of Successful and Failure Mission Outcomes

Number of successful and failed mission outcomes:

mission_outcome	missionoutcomes
Failure (in flight)	1
Success	143
Success (payload status unclear)	2

Above summary shows count of mission outcomes where 143 are success and 1 success without payload status.

Boosters Carried Maximum Payload

Boosters which have carried the maximum payload mass as per the dataset

boosterversion
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

Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

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

There are only two occurrences.

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

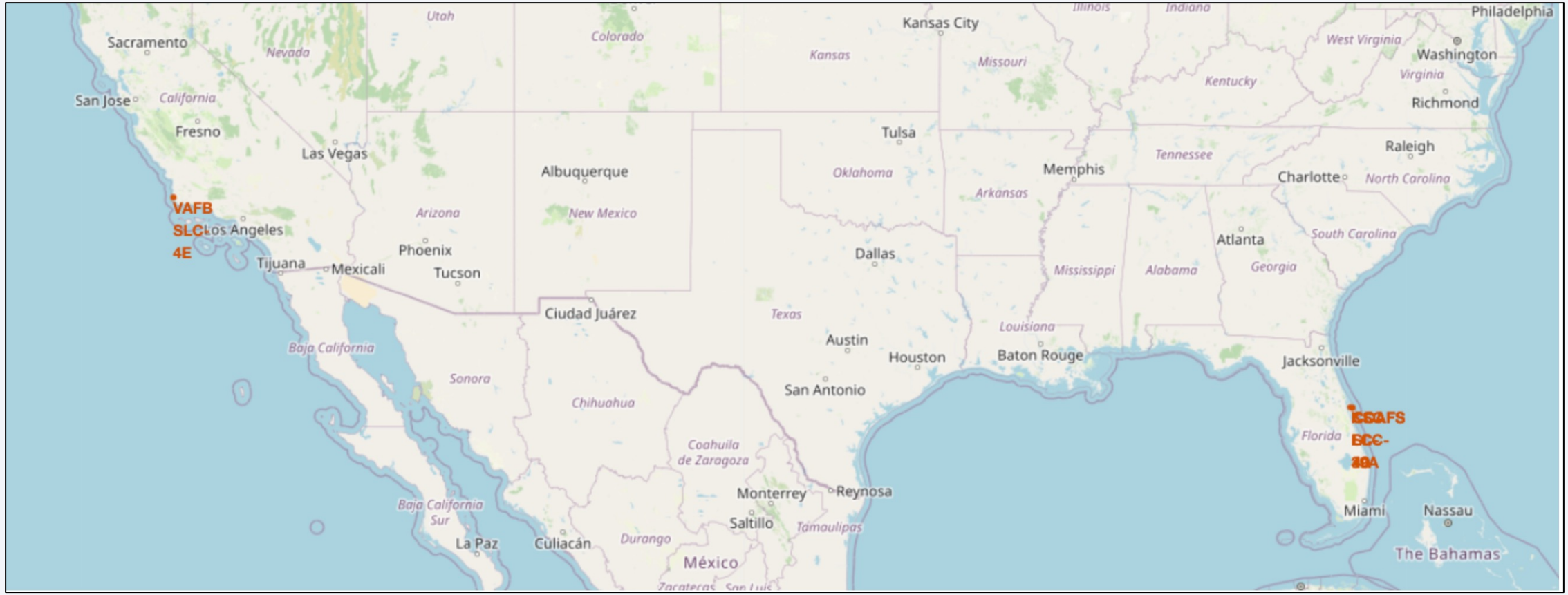
Ranking of all landing outcomes between the date 2010-06-04 and 2017- 03-20:

landing__outcome	count_launches
No attempt	17
Failure (drone ship)	7
Success (drone ship)	7
Success (ground pad)	5
Controlled (ocean)	4
Failure (parachute)	3
Uncontrolled (ocean)	2
Precluded (drone ship)	1

This view of data alerts us that “No attempt” must be taken in account.


Section 3
Launch Sites
Proximities Analysis

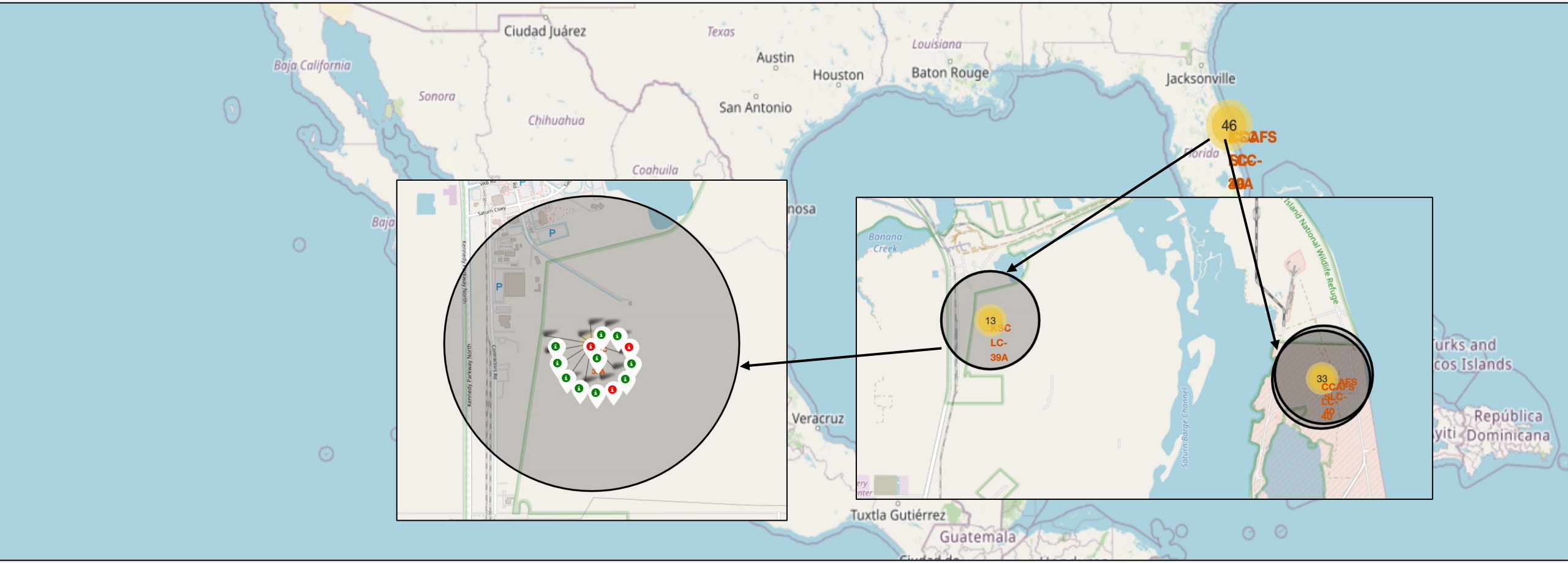
Launch Sites



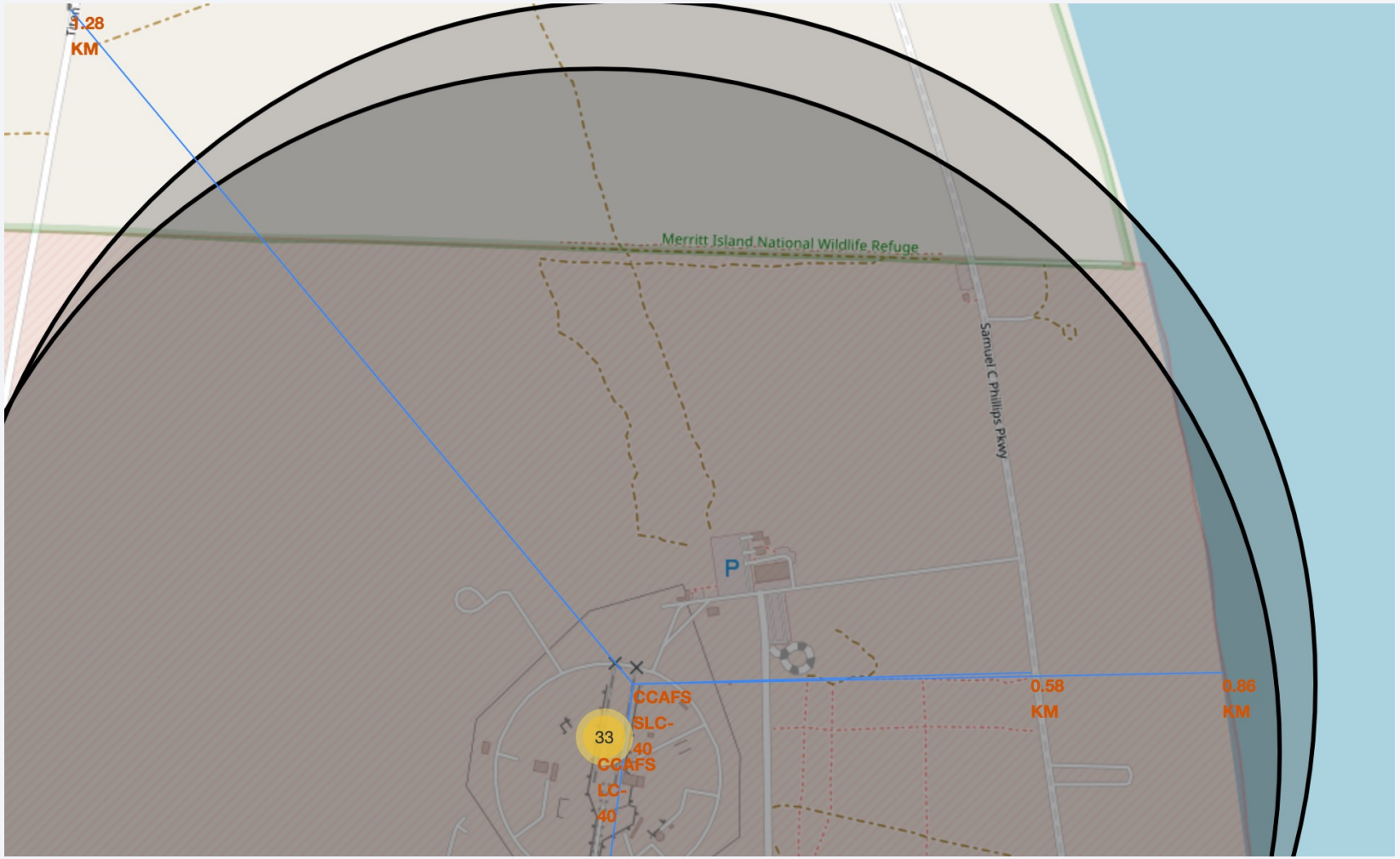
Launch sites are near sea but not too far from roads and railroads.

Launch Outcome by Site

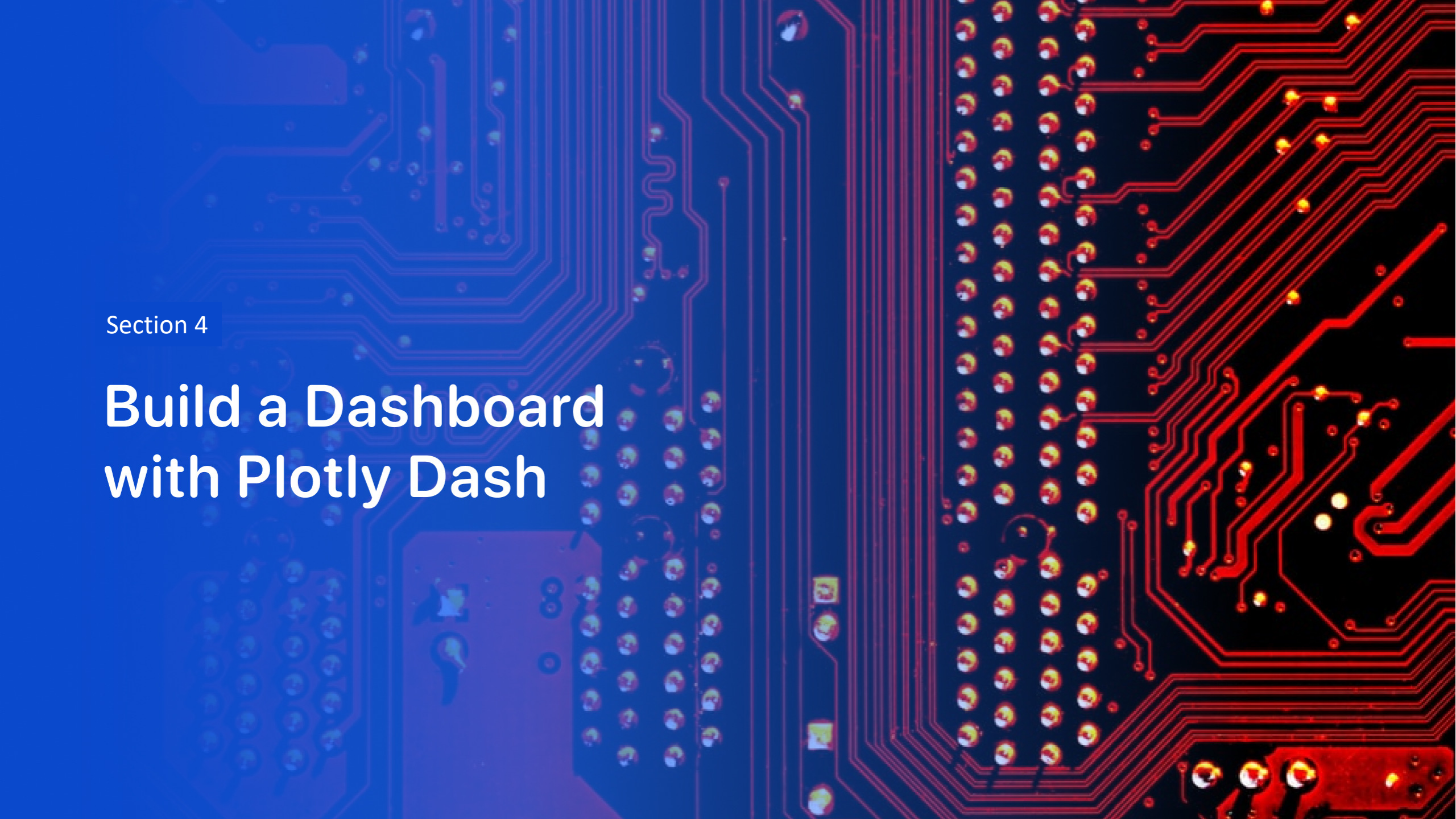
Example of KSC LC-39A launch site launch outcomes
Green markers indicate successful and red ones indicate failure.



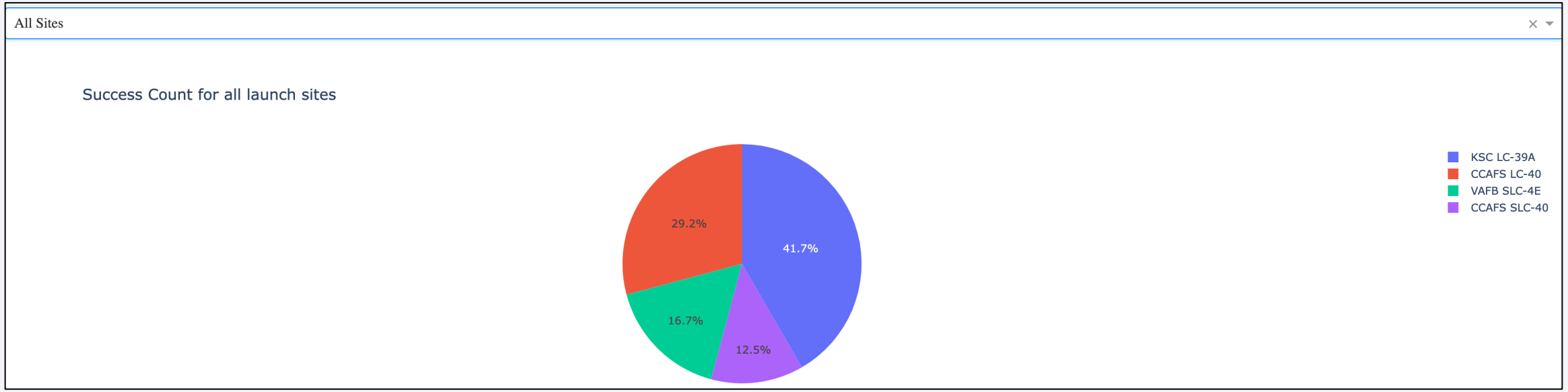
Safety and Logistics



Launch site CCAFS SLC-40 has good logistics aspects, being near railroad and Highway and relatively far from Melbourn.

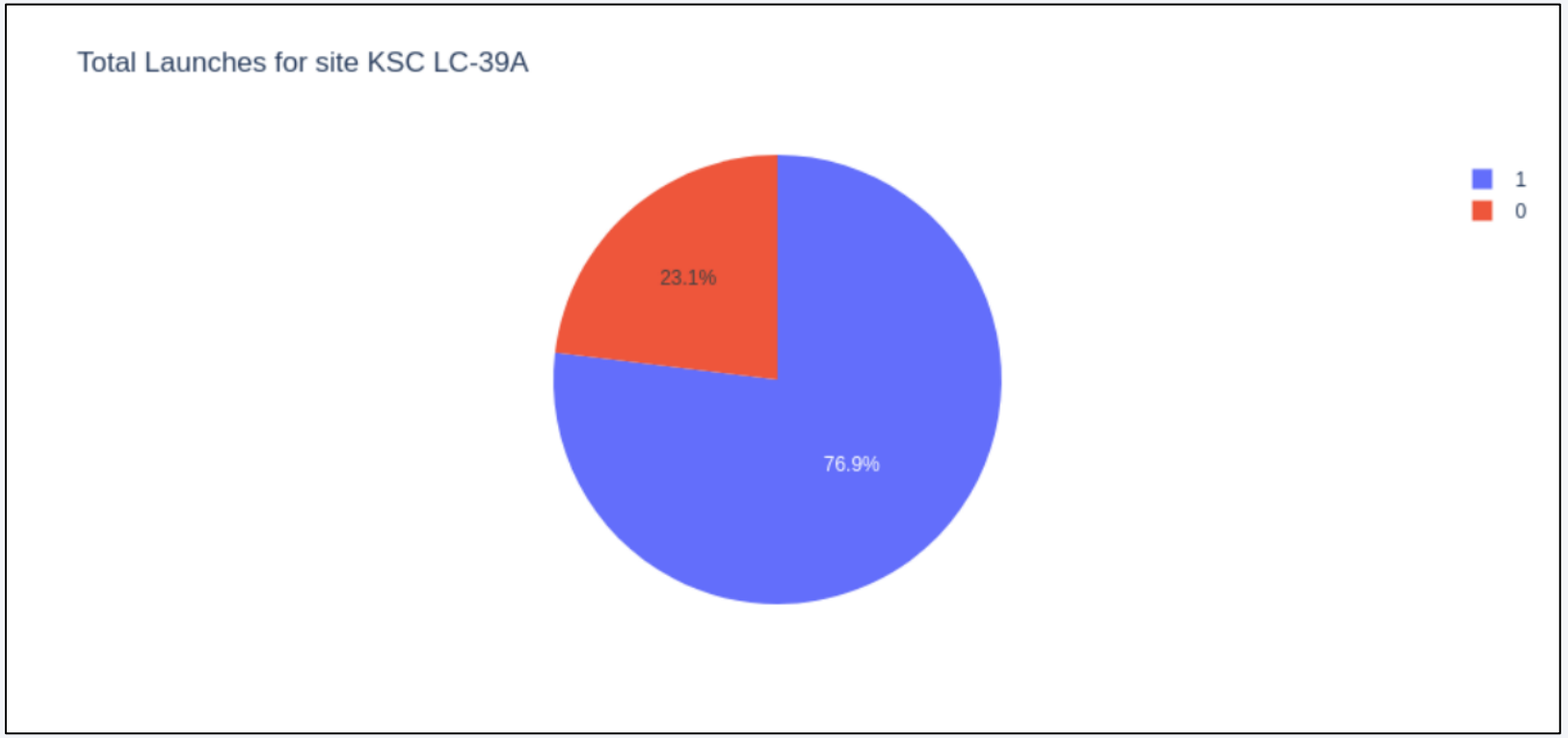

Section 4
Build a Dashboard with Plotly Dash

Successful launches by Site



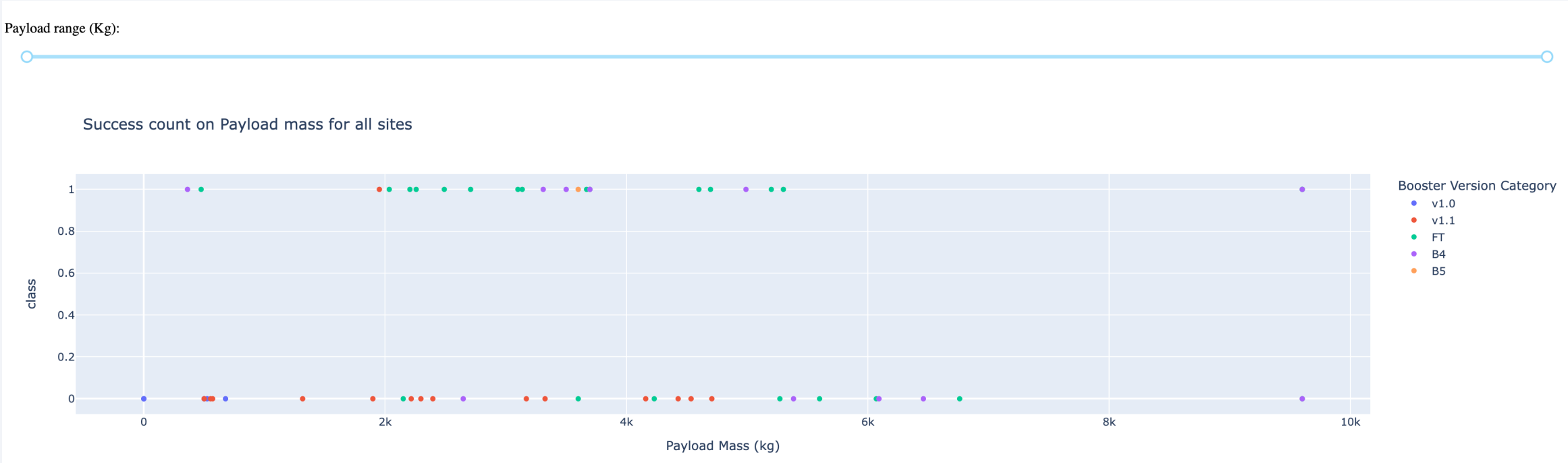
Sites with successful launches

Launch Success Ratio for KSC LC-39A



76.9% of launches are successful in this site.

Payload vs. Launch Outcome



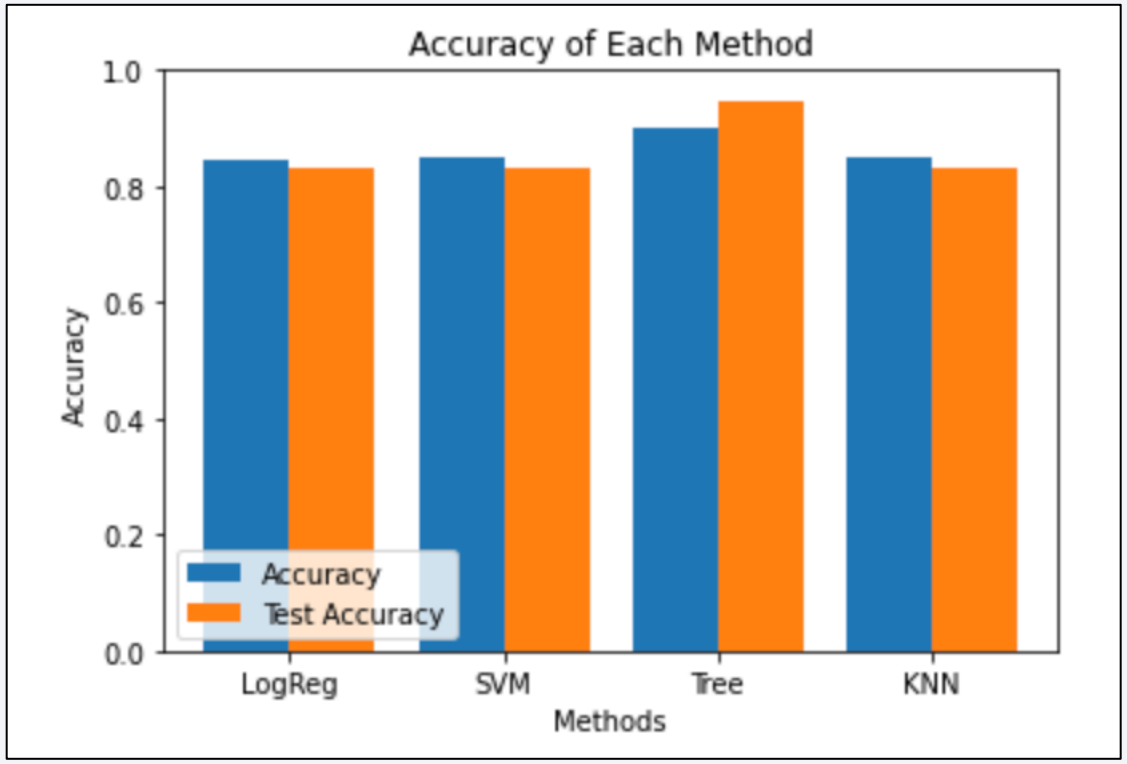
Payloads under 6,000kg and FT boosters are the most successful combination.

Section 5

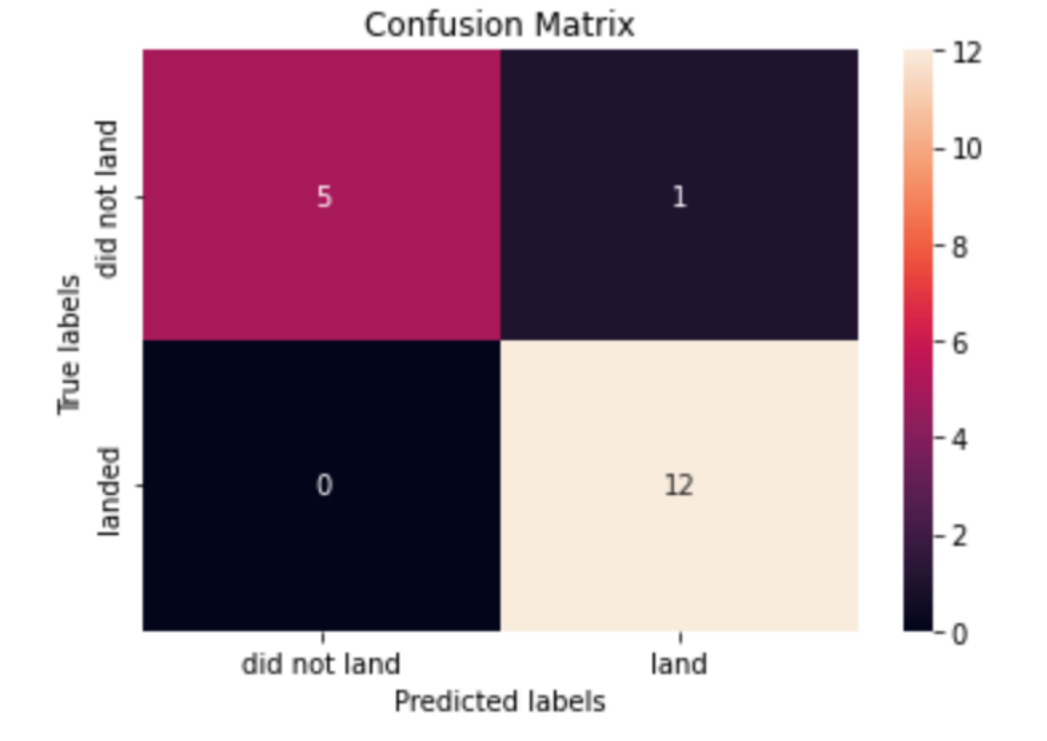
Predictive Analysis (Classification)

Classification Accuracy

- Four classification models were tested, and their accuracies are plotted beside;
- The model with the highest classification accuracy is Decision Tree Classifier, which has accuracies over than 90%.



Confusion Matrix



• Confusion matrix of Decision Tree Classifier proves its accuracy by showing the big numbers of true positive and true negative compared to the false ones.

Conclusions

- The best launch site is KSC LC-39A
- Launches above 7,000kg are less risky
- Although most of mission outcomes are successful, successful landing outcomes seem to improve over time, according the evolution of processes and rockets
- Decision Tree Classifier can be used to predict successful landings and increase profits.

Appendix

- Folium charts are not visible in the github
- Change date format in spacex csv file to use it via Watson studio.

Thank you!