IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Francesc
25/10/2024

Outline

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

Executive Summary

- Methodology to analyze data:
    - Data collection using public Api and web scrapping
    - Exploratory data analysis(EDA), with data wrangling, data visualization and visual interactive data.
    - Machine learning prediction
- Summary of the results
    - EDA results
    - Interactive analytics visualization
    - Predictive analysis results

Introduction

- SpaceY is a company that competes against SpaceX to launch rockets, which needs to know from the existing data if its rockets are capable of landing in order to save costs.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - SpaceX Api(https://Api.spacexdata.com/v4/launches/past)
  - Web Scrapping(https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches)
- Perform data wrangling
  - Data was processed using different methodologies
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - We make a different predictive analysis to know to evaluate different variables

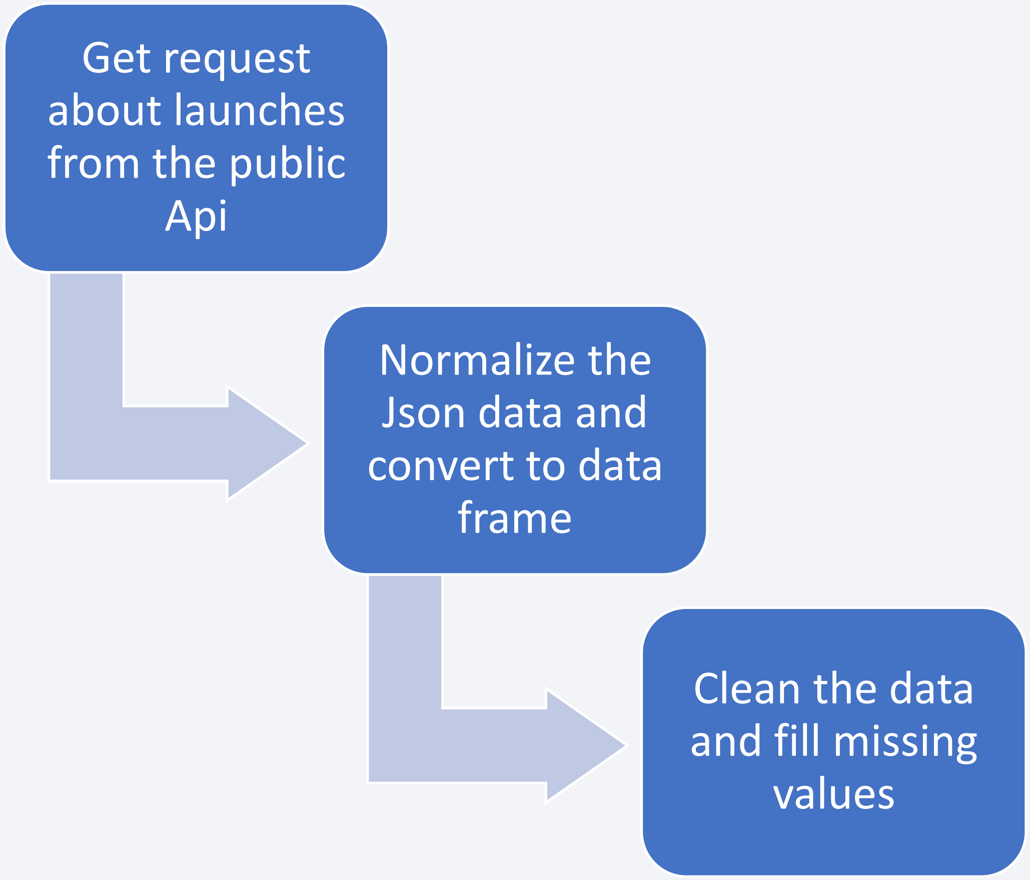
Data Collection

To collect data we use a public Api and web scrapping from the Wikipedia.

- With Api rest, we make a request to the Api and format the response as Json and normalize the data, after we do some data clean and fill missing values.
- With web scrapping we extract the launch records from html table of Wikipedia page and convert this to data frame

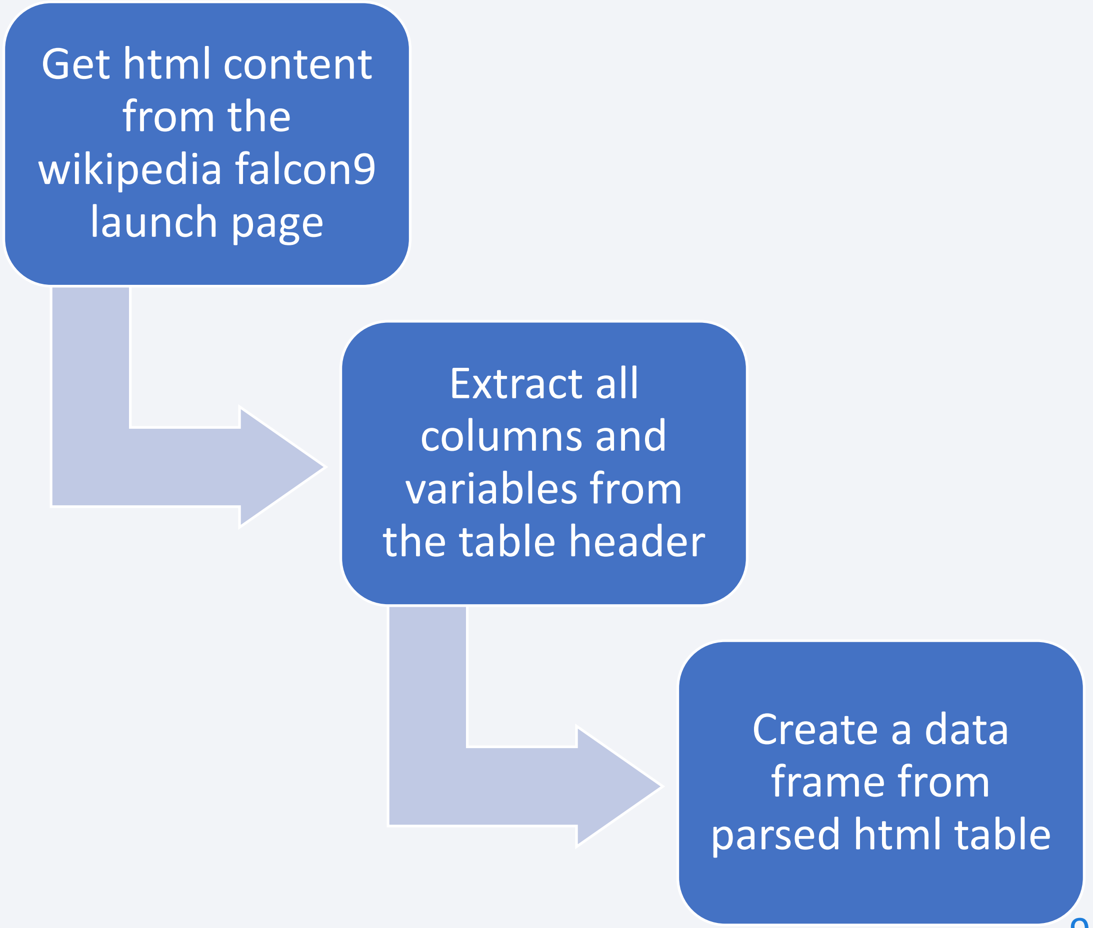
Data Collection – SpaceX API

- LINK



Data Collection - Scraping

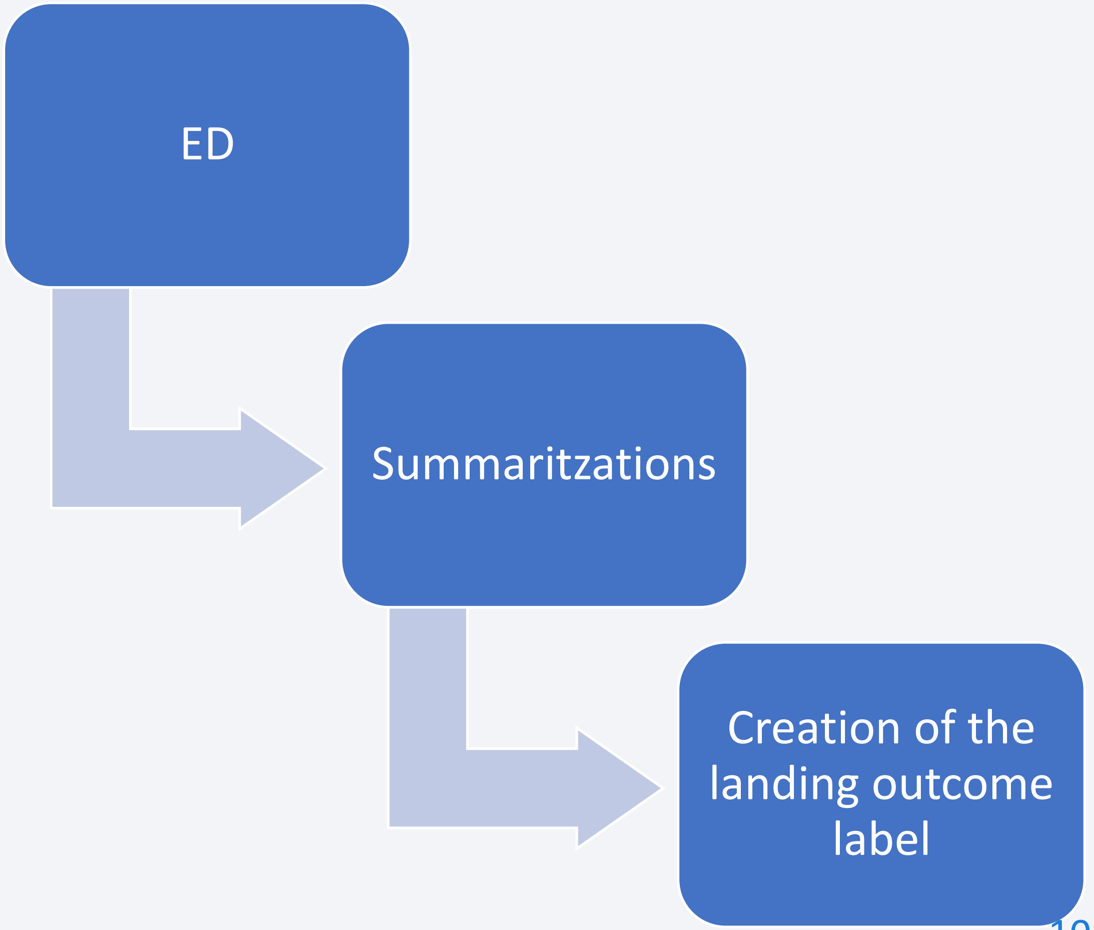
- [LINK]



Data Wrangling

- Initially some EDA was performed in the dataset
- Then the summaries launches per site, and many occurrences of the data were calculated.
- You need to present your data wrangling process using key phrases and flowcharts

- LINK



EDA with Data Visualization

- To analyze the data we use some different plots such as scatterplots or barplots to visualize the relationship between pair and features.
- LINK

EDA with SQL

- List of the queries performed:
  - Name of the unique launch sites
  - Top 5 launch sites whose start with ‘CCA’
  - Total payload mass carried by boosters launched by NASA (CRS)
  - Average payload mass carried by booster version F9 v1.1
  - Date of the first successful landing outcome in ground was achived
  - Name of the booster with a success drone ship landing with payload mass between 4000 and 6000 KG
  - Total number of successful and failure missions outcomes
  - Names of the booster versions wich have carried the maximum payload mass
  - Failed landing outcomes in drone ship, their booster versions, and launch sites names in year 2025
  - Rank of the count of landing outcomes between the date 2010-06-04 and 2017-03-20
- LINK

Build an Interactive Map with Folium

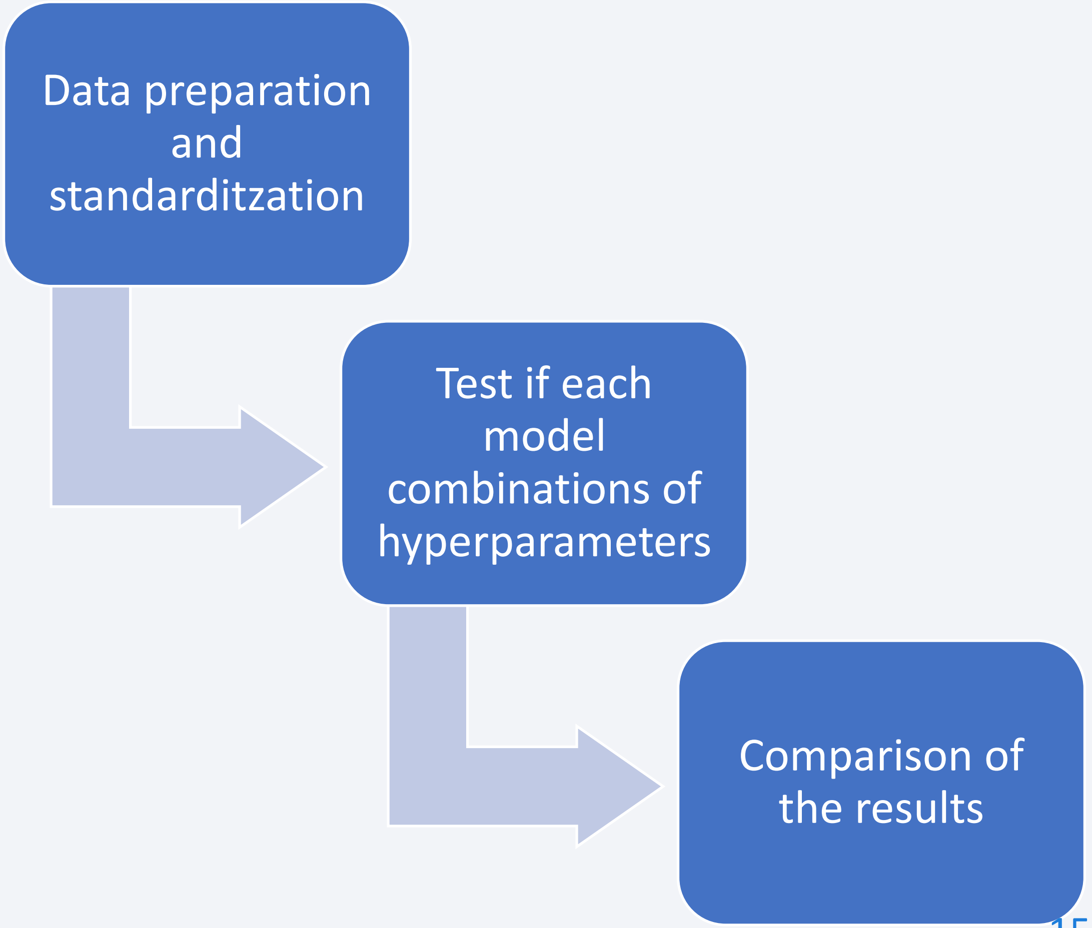
- Markers, circles and marker cluster were used with Folium Maps
  - Markers indicate launch sites
  - Circles indicate highlighted areas around specific coordinates
  - Marker cluster indicates groups of events in each coordinate such as lunches in a launch site
  - Lines are used to indicate distances between two coordinates
- LINK

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 the best place to launch according to payloads.
- LINK

Predictive Analysis (Classification)

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



Results

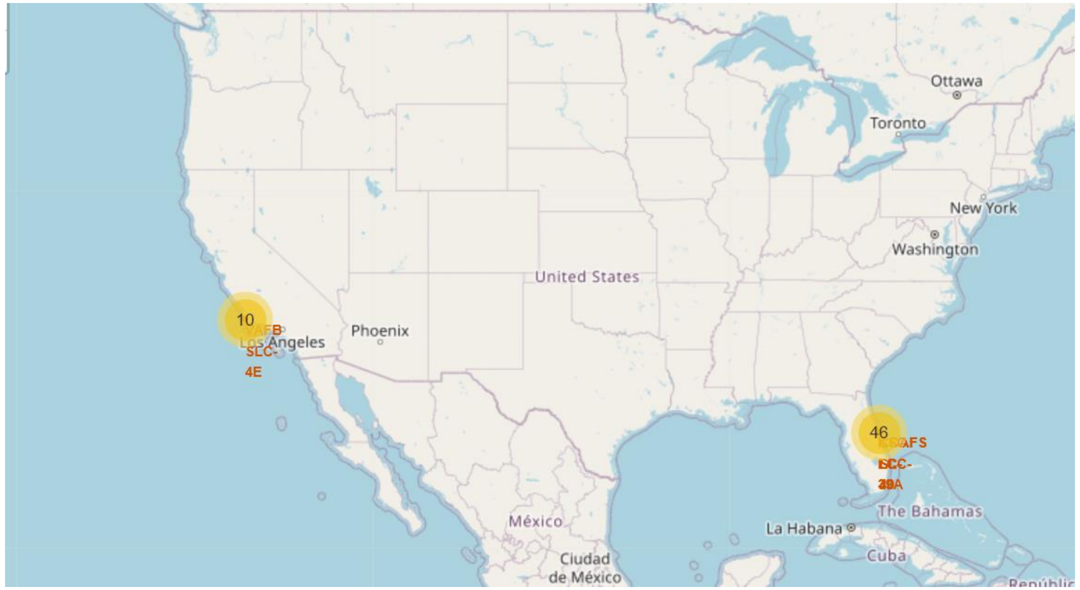
Exploratory data analysis results:

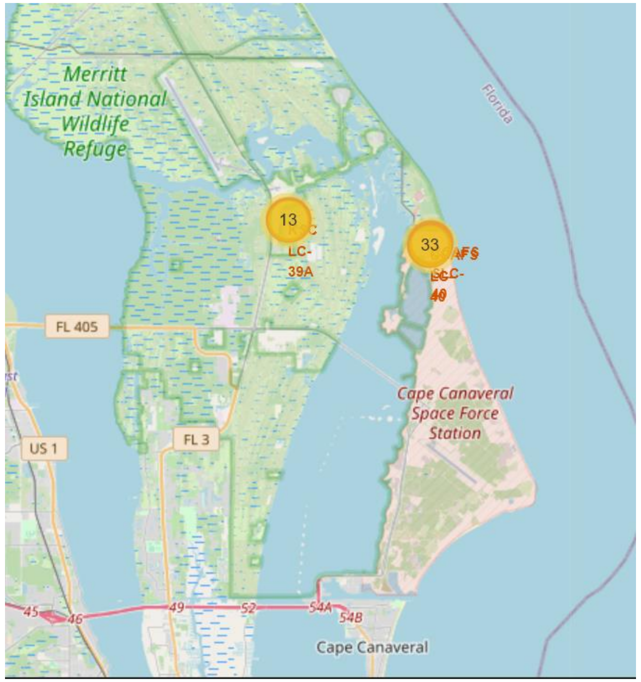
- Space X uses 4 different launch sites.
- The first launches were done to Space X itself and NASA.
- The average payload of F9 v1.1 booster is 2,928 kg.
- The first success landing outcome happened in 2015 fiver year after the first launch.
- Many Falcon 9 booster versions were successful at landing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

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

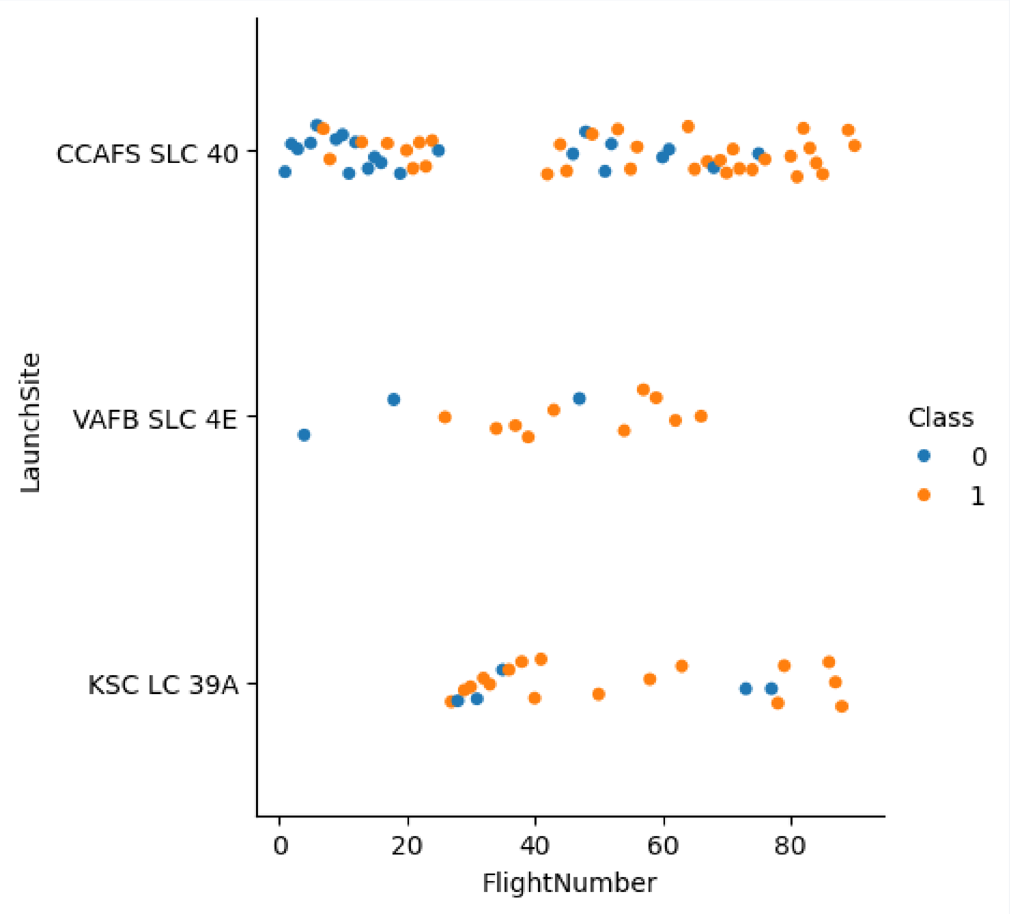
- Predictive Analysis showed that Decision Tree Classifier is the best model to predict successful landings, having accuracy over 87% and accuracy for test data over 94%.

Section 2

Insights drawn from EDA

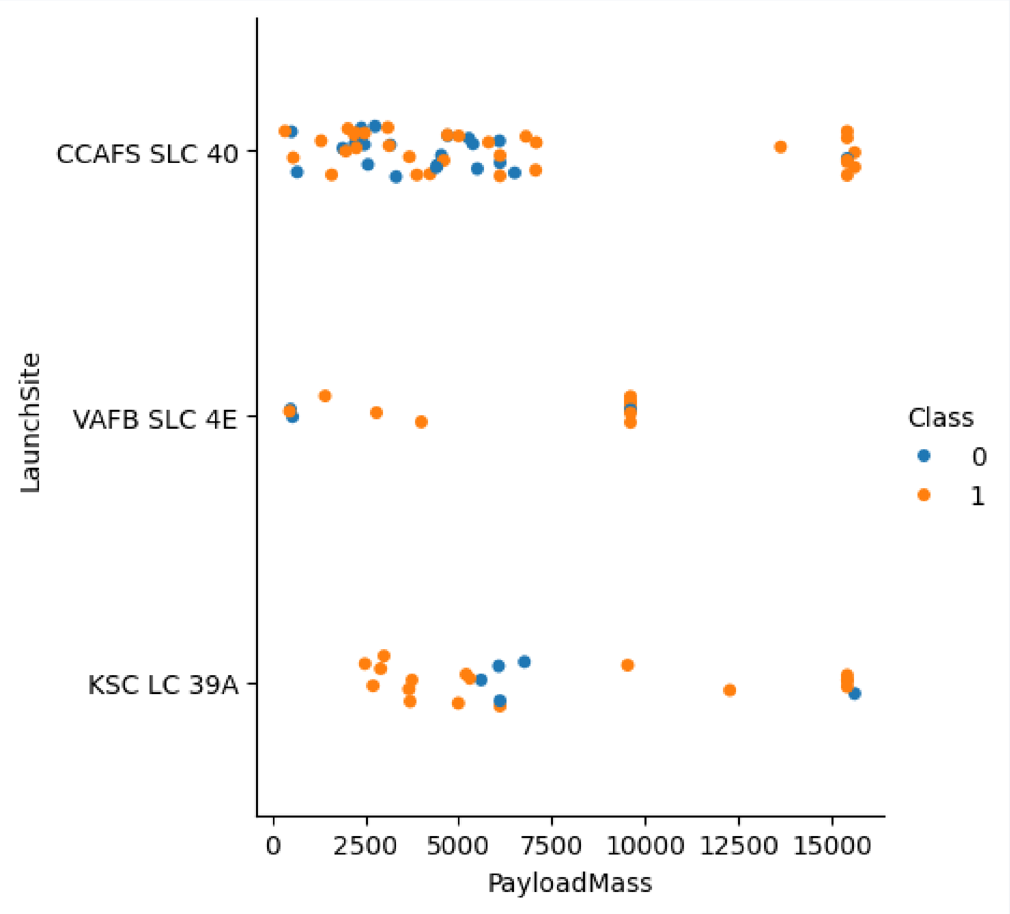
Flight Number vs. Launch Site

- the best launch site nowadays is CCAF5 SLC 40, where most of recent launches were successful
- It’s also possible to see that the general success rate improved over time.



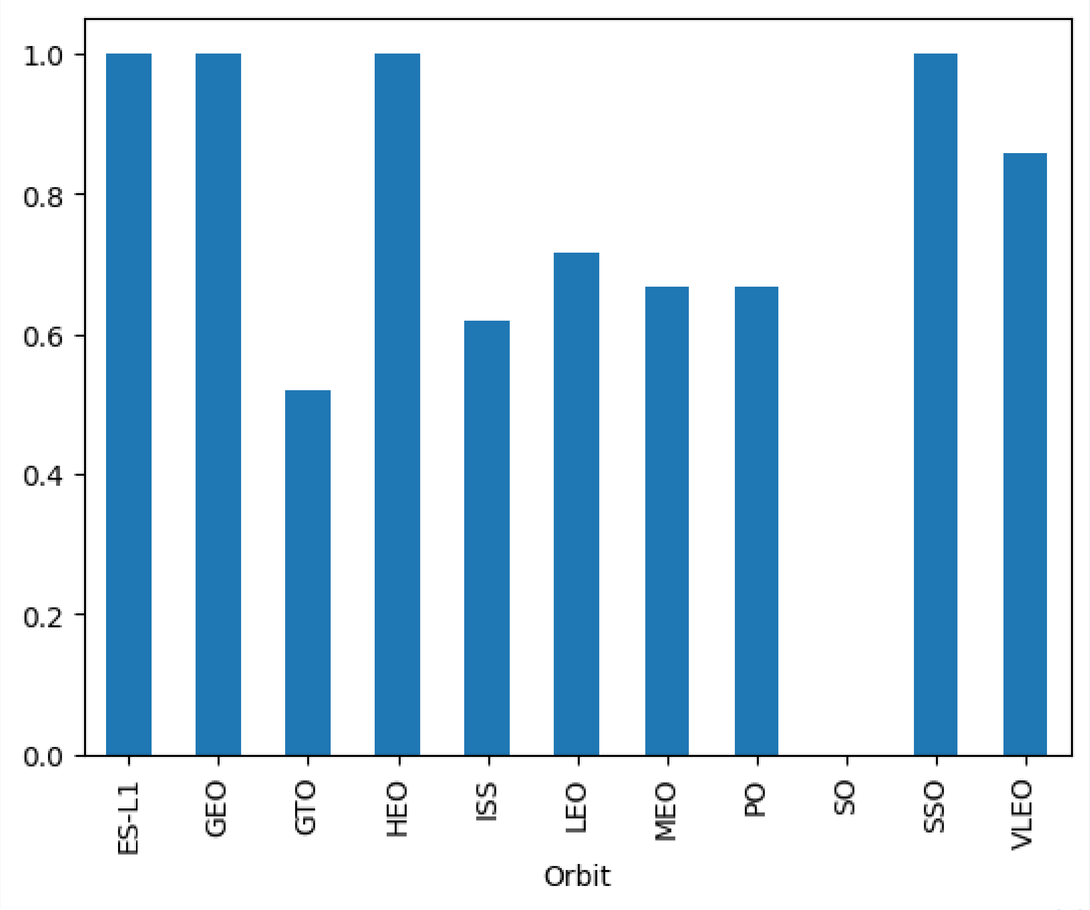
Payload vs. Launch Site

- Payloads over 9,000kg have excellent success rate.
- Payloads over 12,000kg seems to be possible only on CCAFS SLC 40 and KSC LC 39A launch sites.



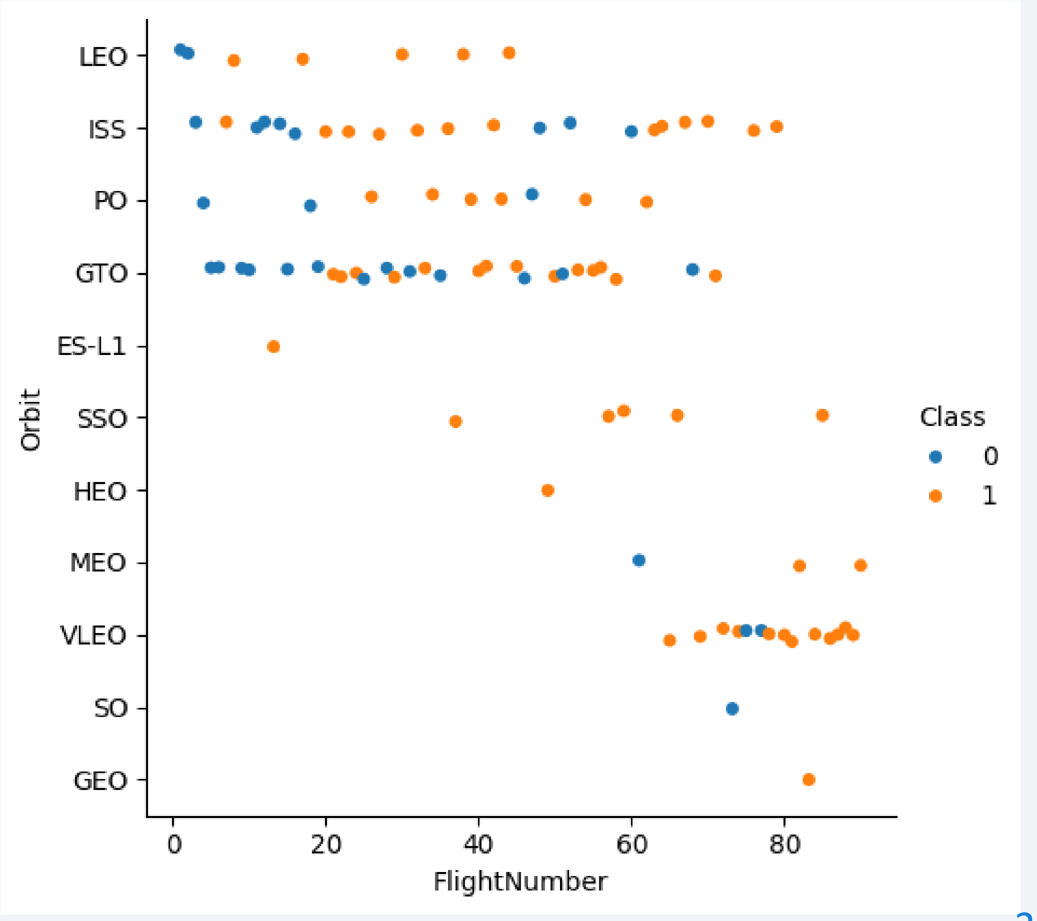
Success Rate vs. Orbit Type

- •The biggest success rates happens to orbits:
    - ES-L1
    - GEO
    - •HEO
    - SSO
- •Followed by:
    - VLEO (near 80%)
    - LFO (near70%)



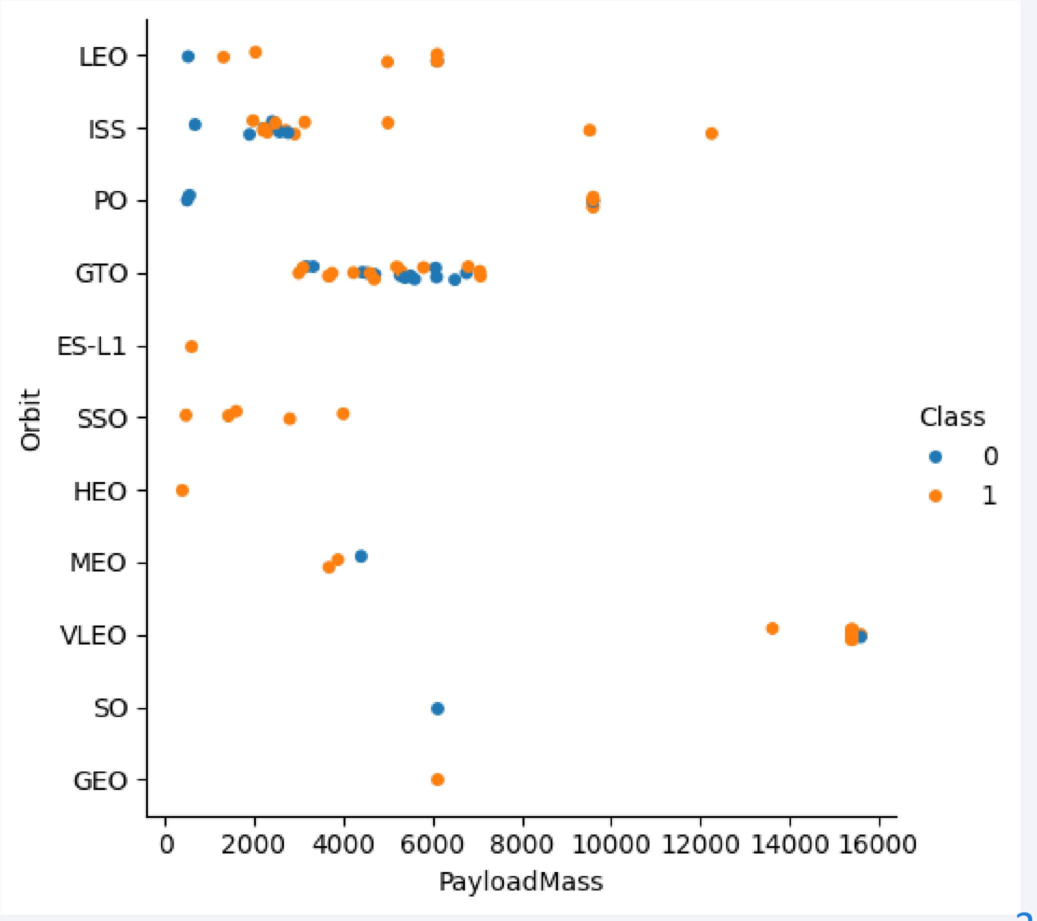
Flight Number vs. Orbit Type

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



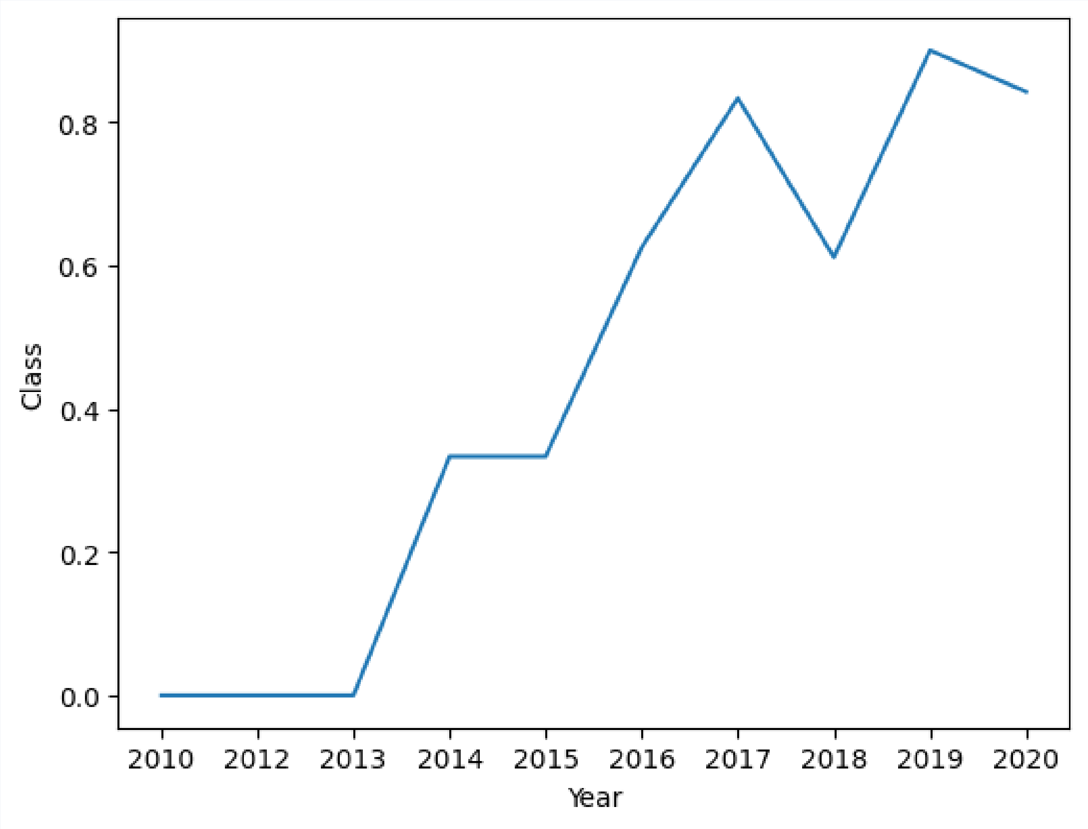
Payload vs. Orbit Type

- Apparently, 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.



Launch Success Yearly Trend

- Success rate started increasing in 2013 and kept until 2020;
- the first three years were used for adjusts and improvement of technology.



All Launch Site Names

- Search all unique launch site names
- Launch_Site
  - CCAFS LC-40
  - VAFB SLC-4E
  - KSC LC-39A
  - CCAFS SLC-40

Launch Site Names Begin with 'CCA'

- Search the five first elements where launch site name start wit 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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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

- The result of sum all payload whose codes was Nasa CRS code
- total payload(Kg)
  - 45596

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- avg(PAYLOAD_MASS__KG_)
  - 2928.4

First Successful Ground Landing Date

- Search the lowest date were landig outcome is success in groud
- 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
- Using distinct for unique values
- Booster version
  - F9 FT B1022
  - F9 FT B1026
  - F9 FT B1021.2
  - F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- Group by mission outcome and count the records

Mission_Outcome	count(*)
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- First get the maximum payload and make other select to get witch booster carried this payload
- Booster version
    - F9 B5 B1048.4
    - F9 B5 B1049.4
    - F9 B5 B1051.3
    - F9 B5 B1056.4
    - F9 B5 B1048.5
    - F9 B5 B1051.4
    - F9 B5 B1049.5
    - F9 B5 B1060.2
    - F9 B5 B1058.3
    - F9 B5 B1051.6
    - F9 B5 B1060.3
    - F9 B5 B1049.7

2015 Launch Records

- failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

- We have only to failed landing this year

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

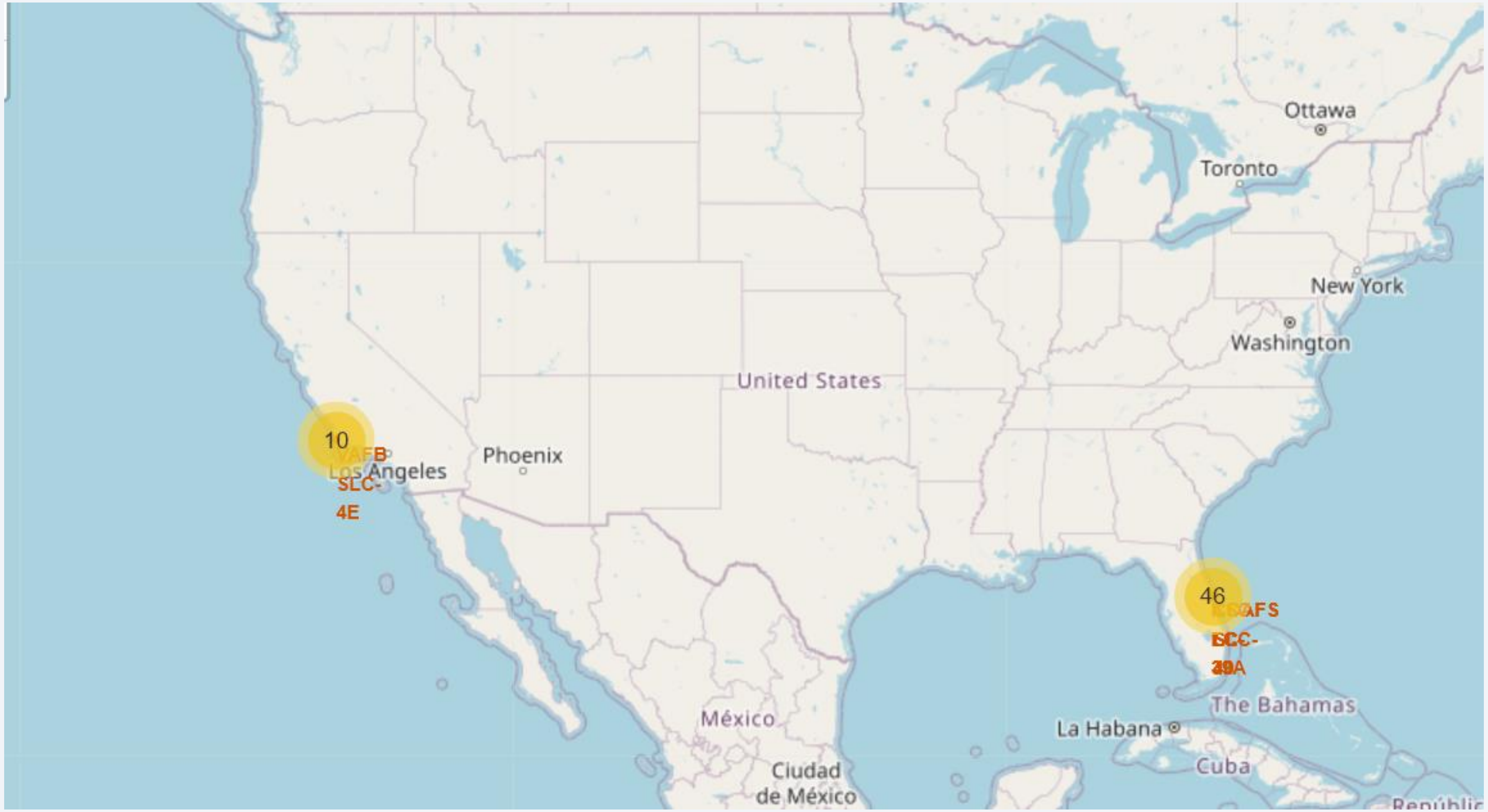
- We must be take “no attempt” in consideration

Landing_Outcome	count(*)
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

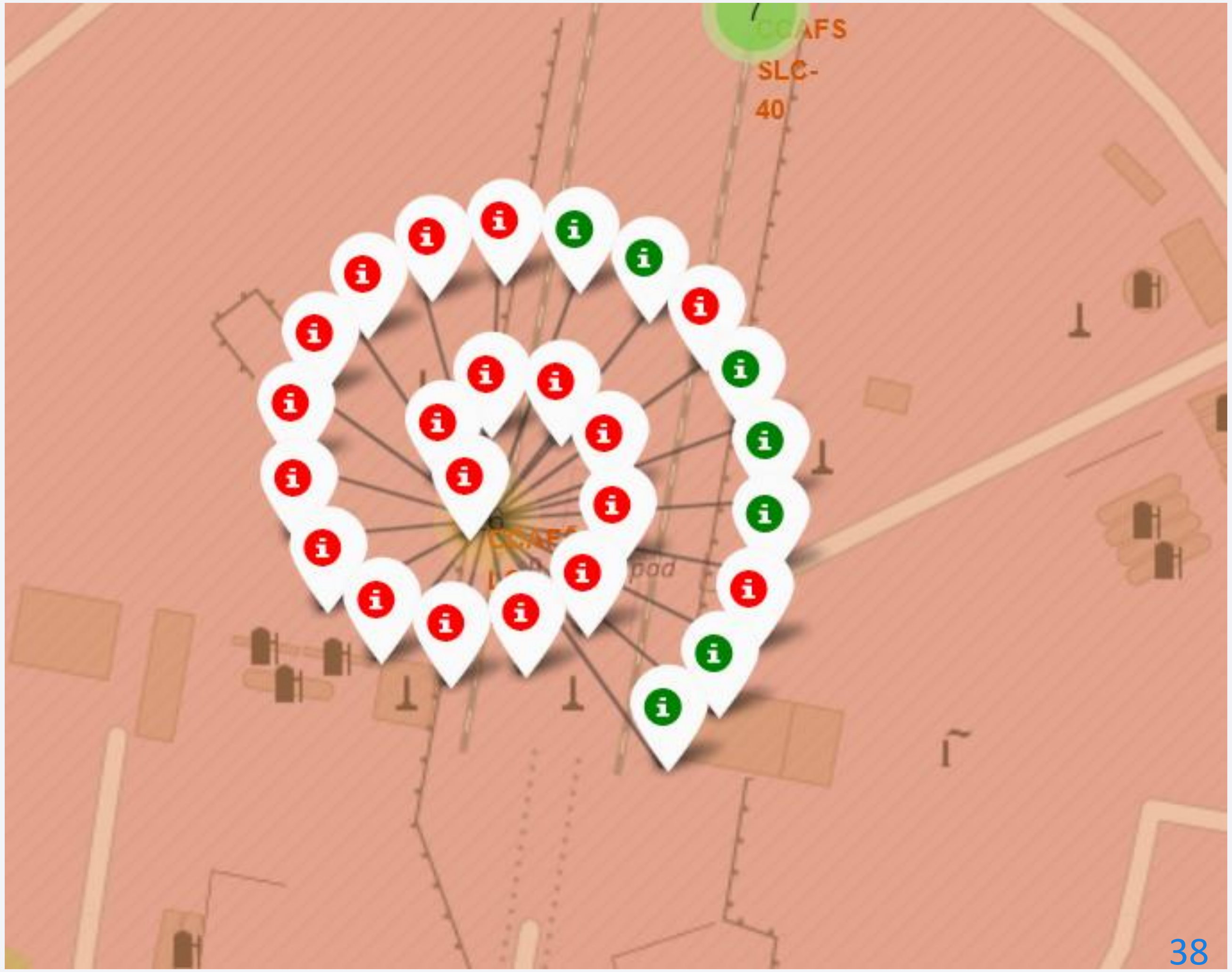
Laucnh outcomes by site



color-labeled launch outcomes on the map

•Green markers indicate successful and red ones indicate failure.

Platform: CCAFS SLC-40



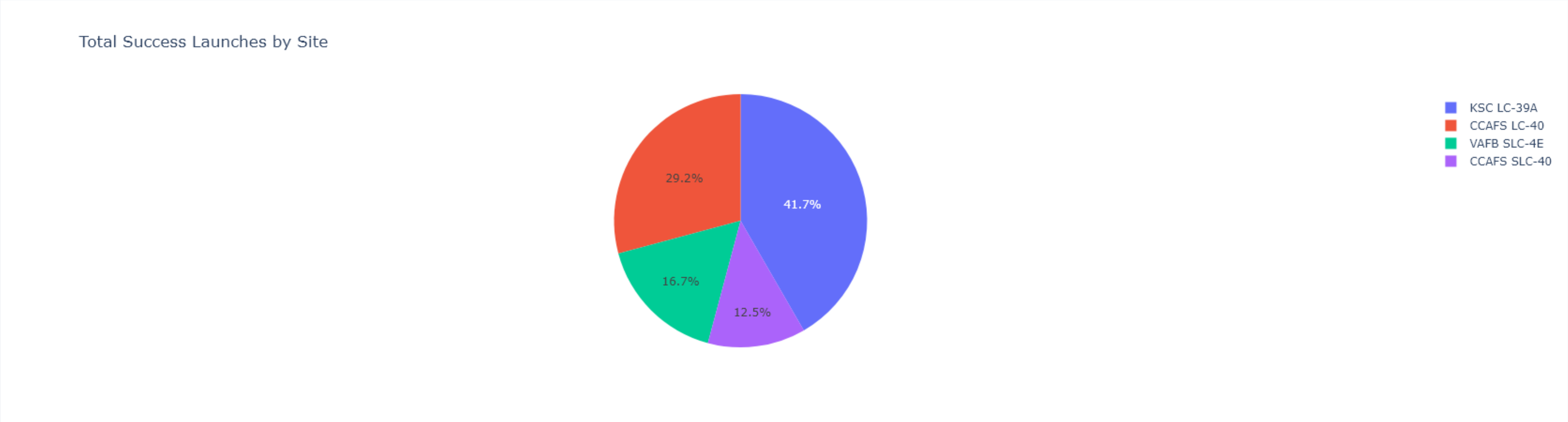
Distance of important points

Launch site CCAFS SLC-40has good logistics aspects, being near railroad and road and relatively far from inhabited areas.

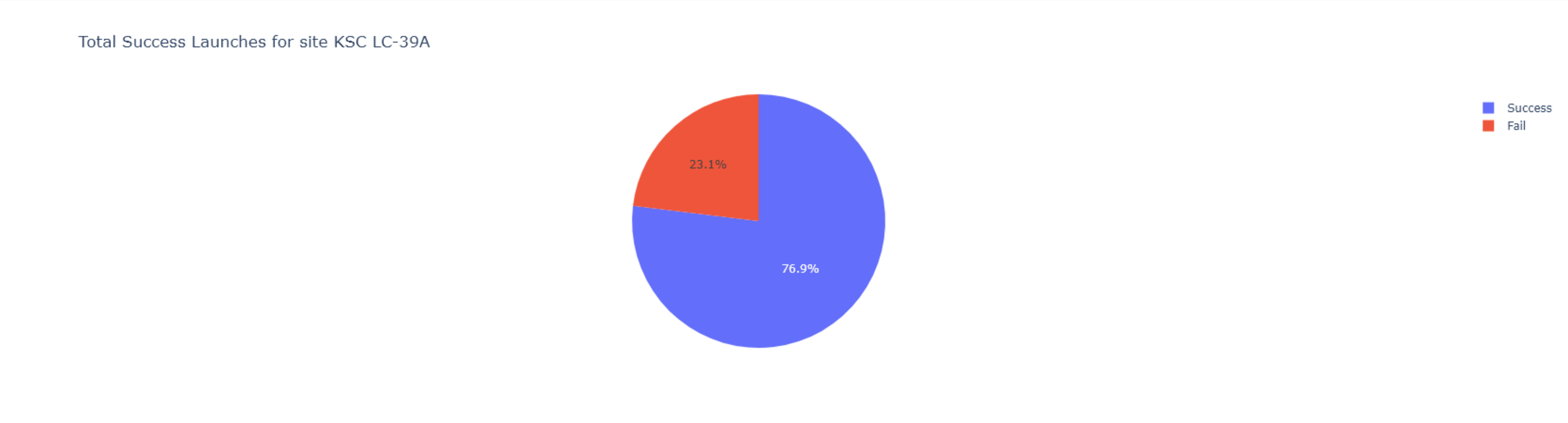
Section 4

Build a Dashboard with Plotly Dash

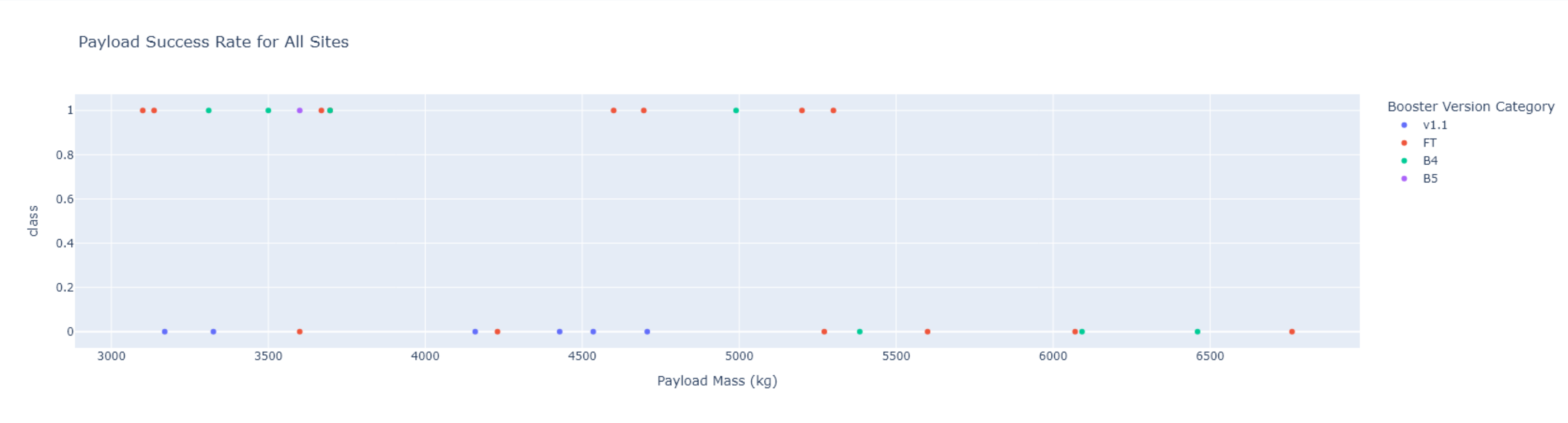
launch success count for all sites



Launch Success Ratio for KSC LC-39A



Payload vs. Launch Outcome

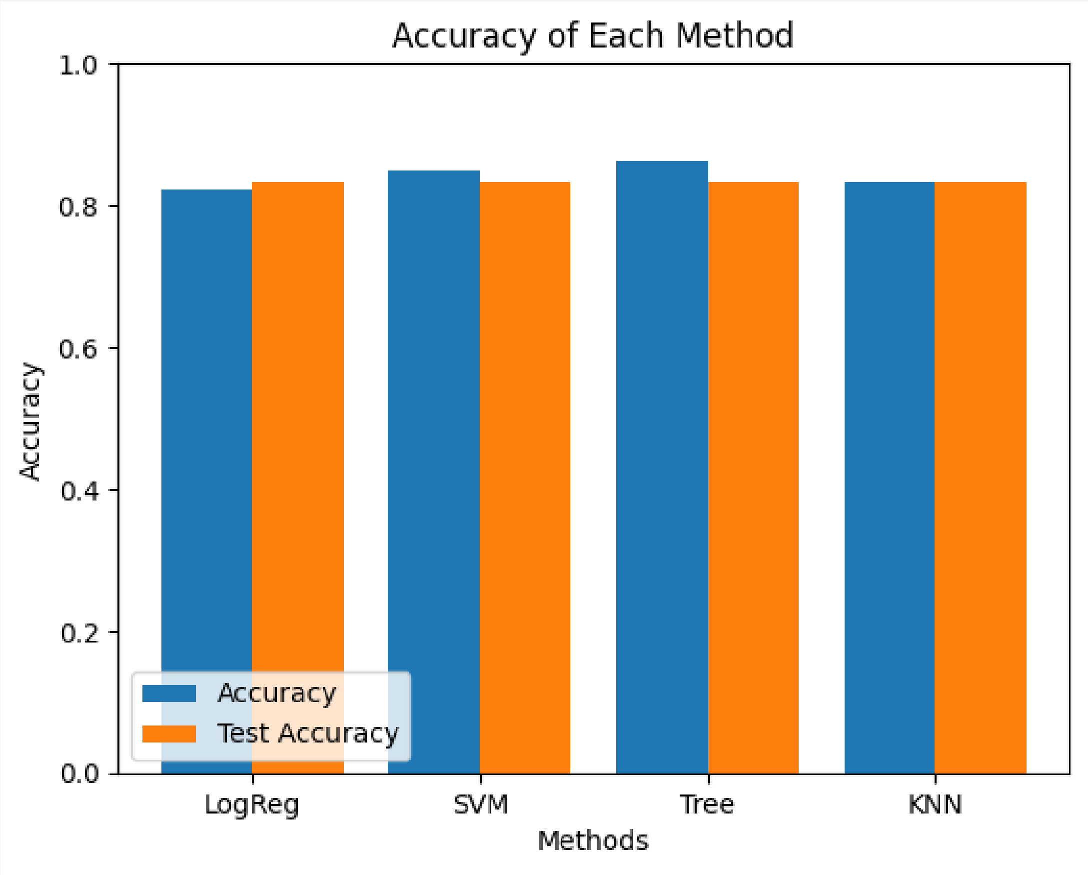


Section 5

Predictive Analysis (Classification)

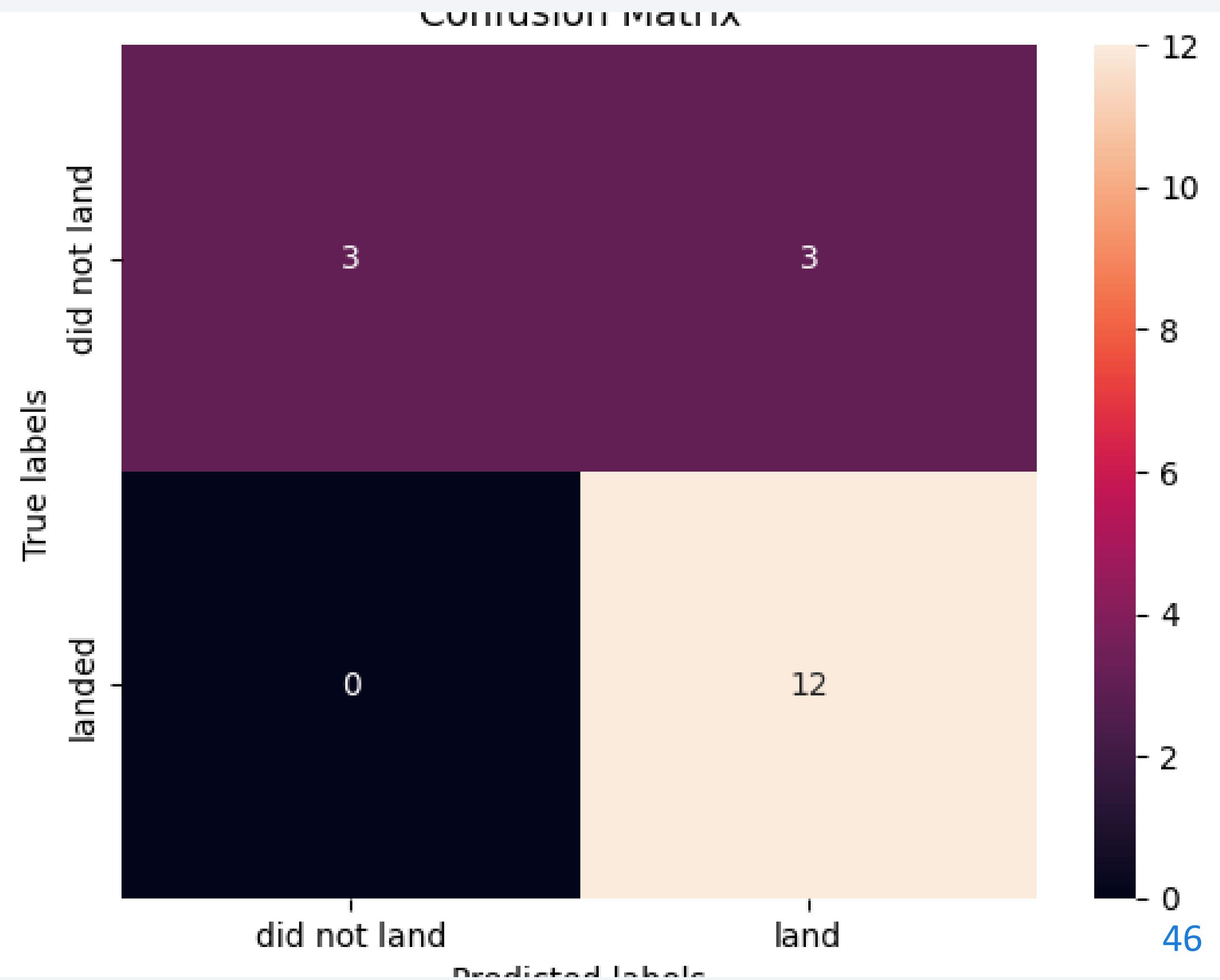
Classification Accuracy

- 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 87%.



Confusion Matrix of Decision Tree Classifier

Confusion matrix of Decision Tree Classifier proves its accuracy by showing the same numbers.



Conclusions

- Different data sources were analyzed, refining conclusions along the process.
- 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 didn’t show maps on Github

Thank you!