Data Science Capstone project

Monu Bhagat

04/06/2022

Outline

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

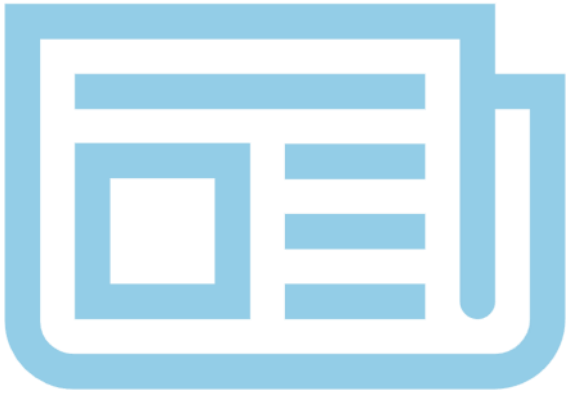
Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Project background and context
- Problems you want to find answers

Methodology

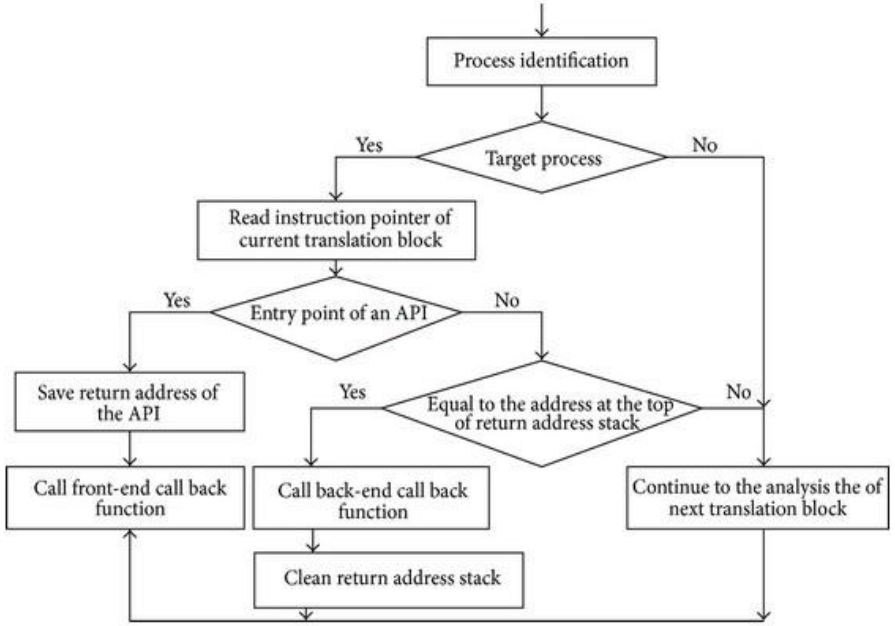


- Data collection methodology:
  - Data collection using API
  - Data collection using web scraper
- Perform data wrangling
  - Data wrangling using Auxiliary Functions
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Predictive analysis using Machine Learning

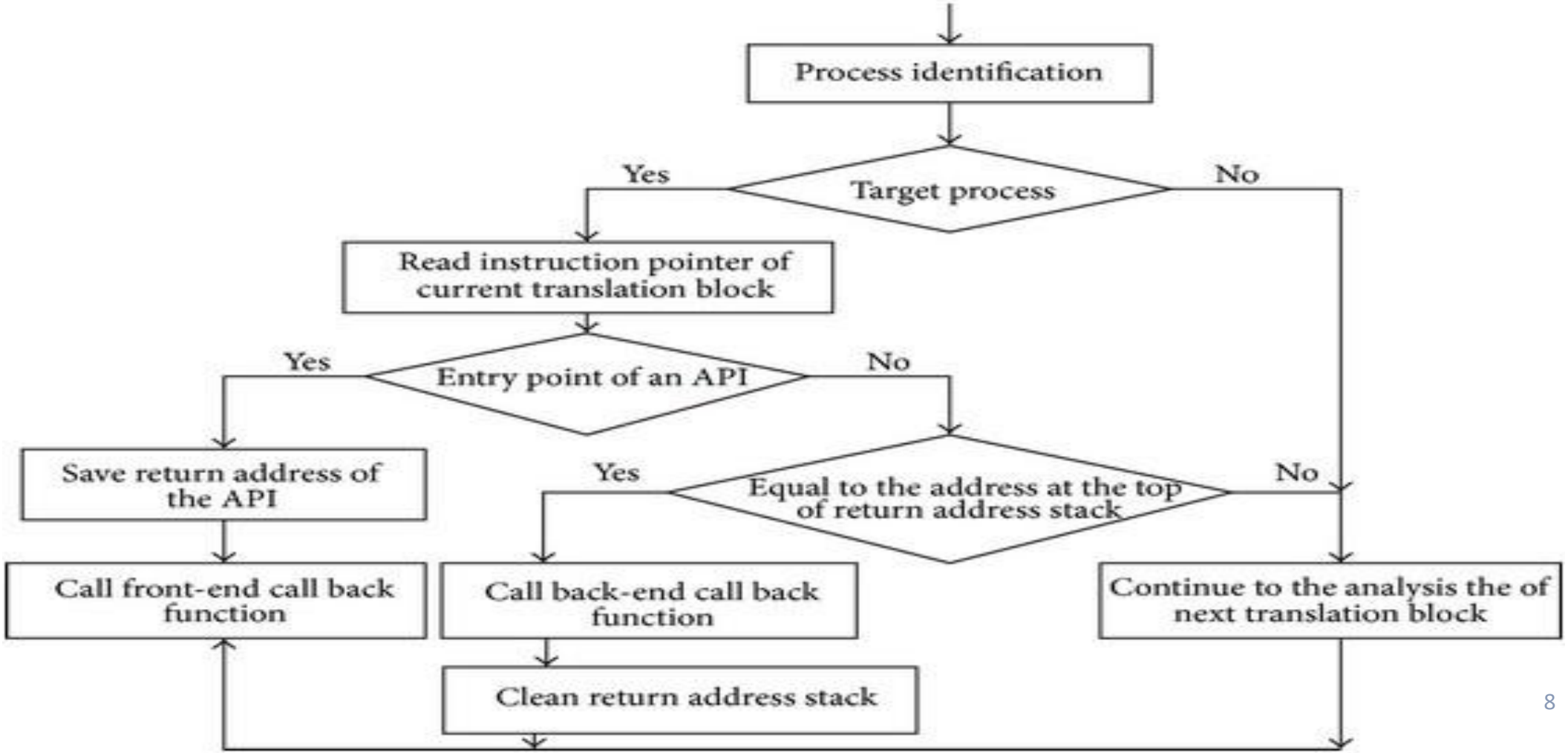
Methodology

Data collection

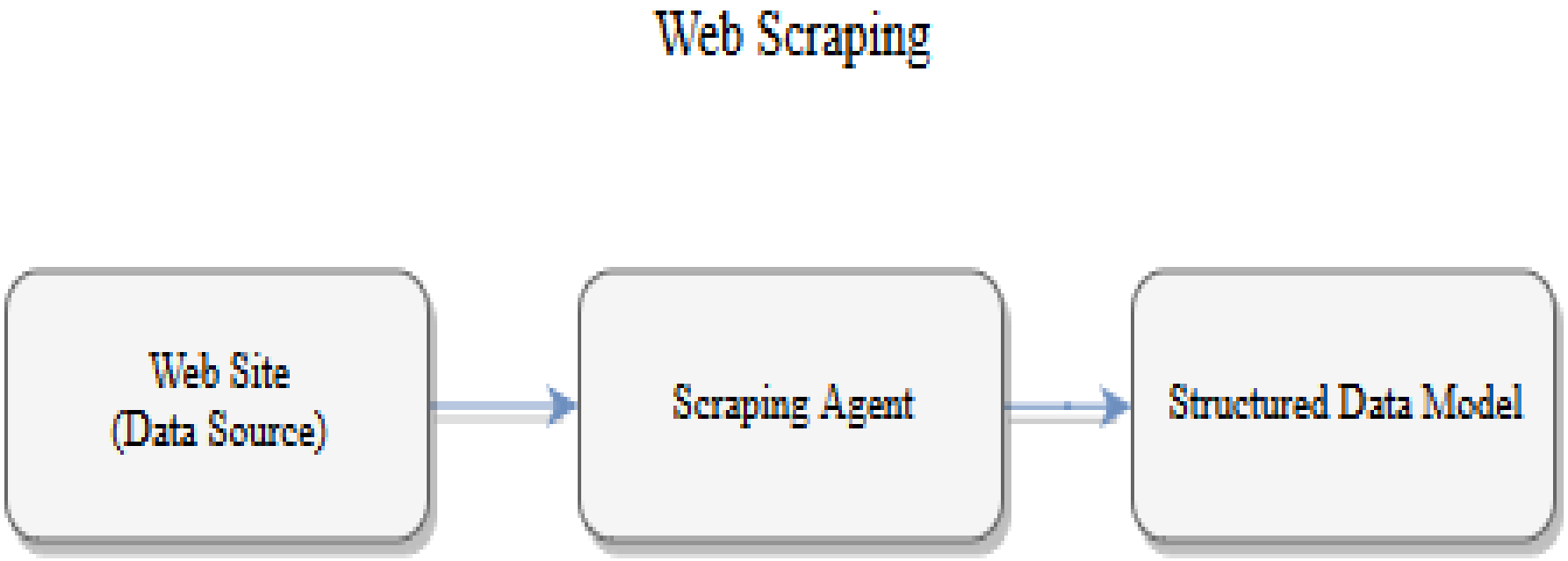
- Data collection is a systematic process of gathering observations or measurements. Whether you are performing research for business, governmental or academic purposes, data collection allows you to gain first-hand knowledge and original insights into your research problem.



Data collection – SpaceX API

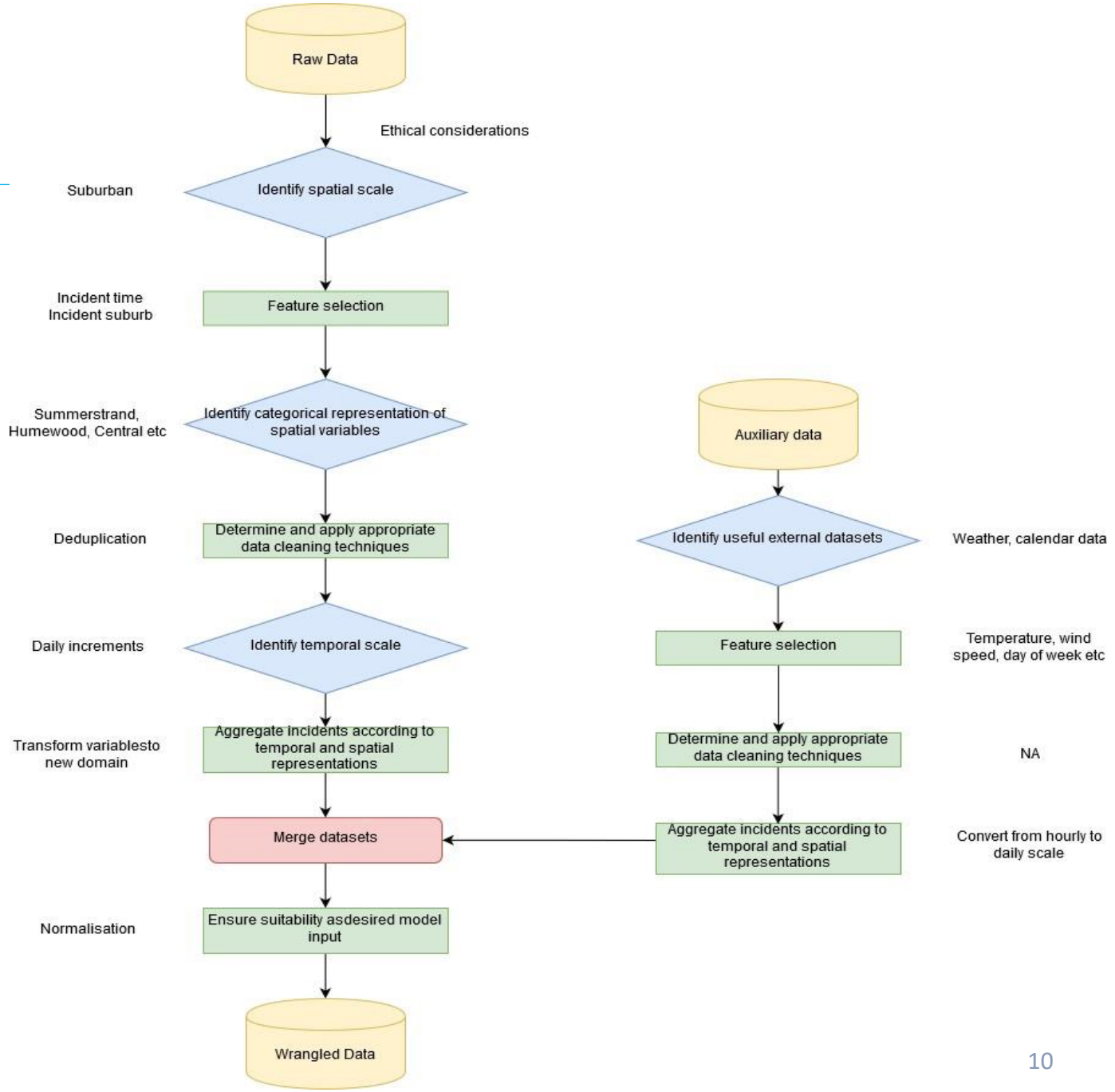


Data collection – Web scraping



Data wrangling

Data wrangling is the process of cleaning and unifying messy and complex data sets for easy access and analysis. With the amount of data and data sources rapidly growing and expanding, it is getting increasingly essential for large amounts of available data to be organized for analysis. This process typically includes manually converting and mapping data from one raw form into another format to allow for more convenient consumption and organization of the data.



EDA with data visualization

- To find out the relationship between features

EDA with SQL

- Import the data into database
- Querying the data you want to find out

Build an interactive map with Folium

- Map objects such as markers, circles, lines, etc. as a marker for the launch site

- To find out the details of the location of the launch site

Build a Dashboard with Plotly Dash

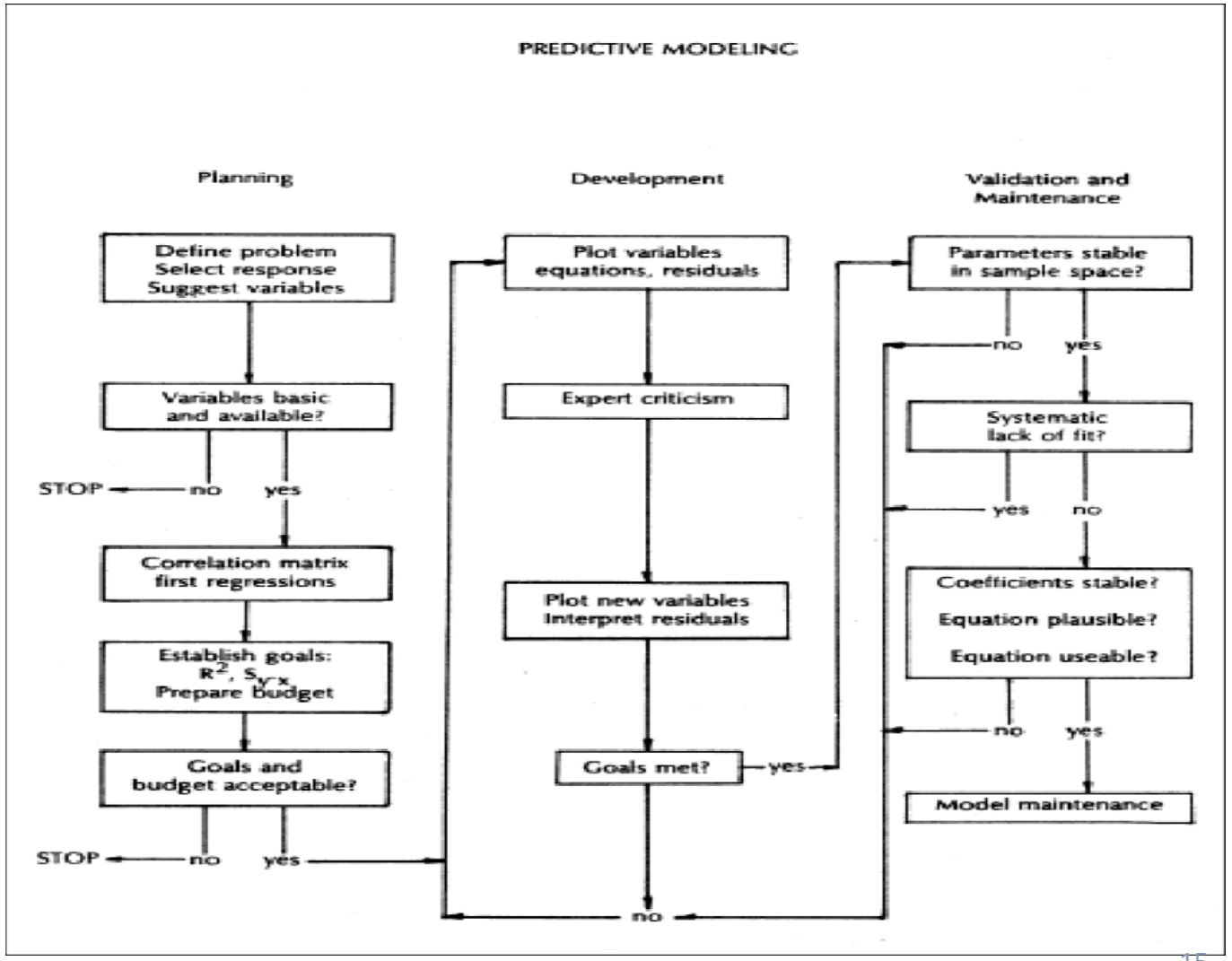
- Dropdown filter and chart

- To find out each launch site

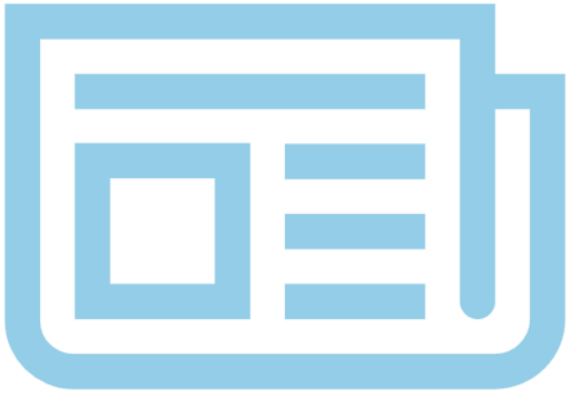
Predictive analysis (Classification)

- Using Machine Learning for predictive analysis with KNN, SVM, Decision Tree, and Logistic Regression, evaluated with Jaccard Score and F1 Score

- https://github.com/MonuBhagat11/Machine_learning_with_python



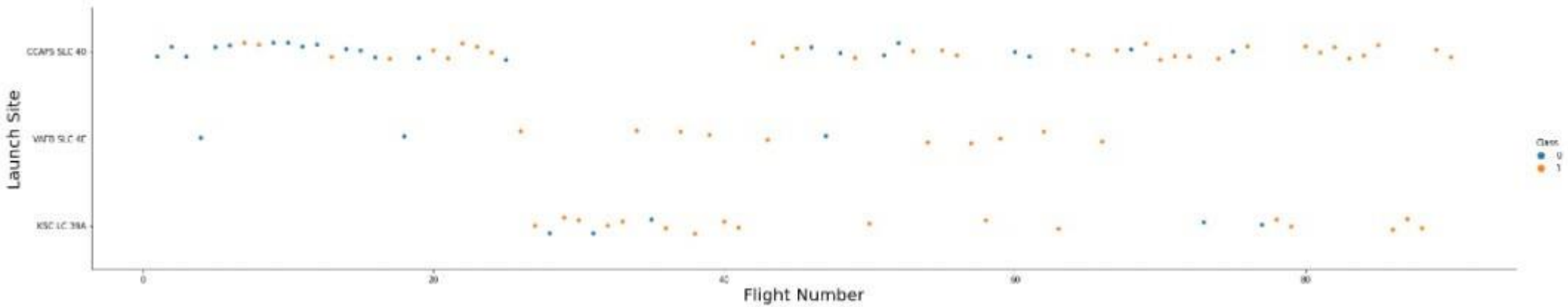
Results



- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

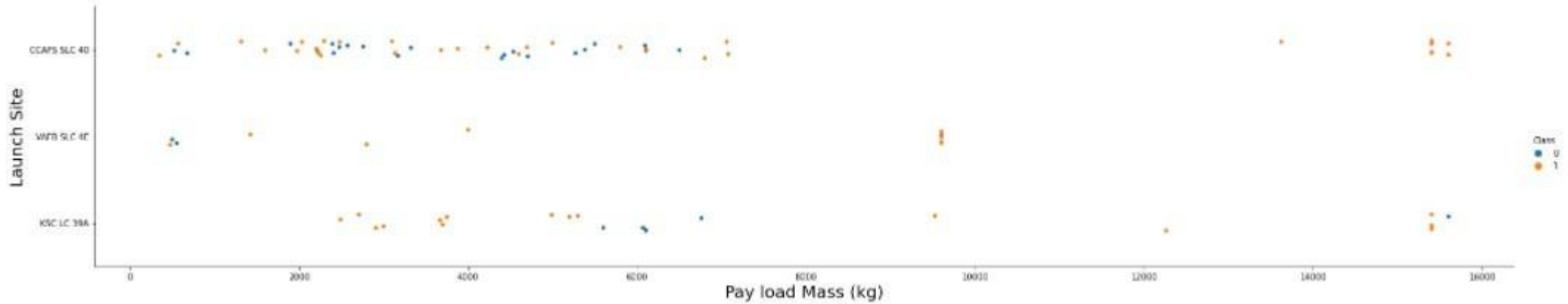
EDA with Visualization

Flight Number vs. Launch Site



The highest success rate is in CCAPS SLC-40 launch site

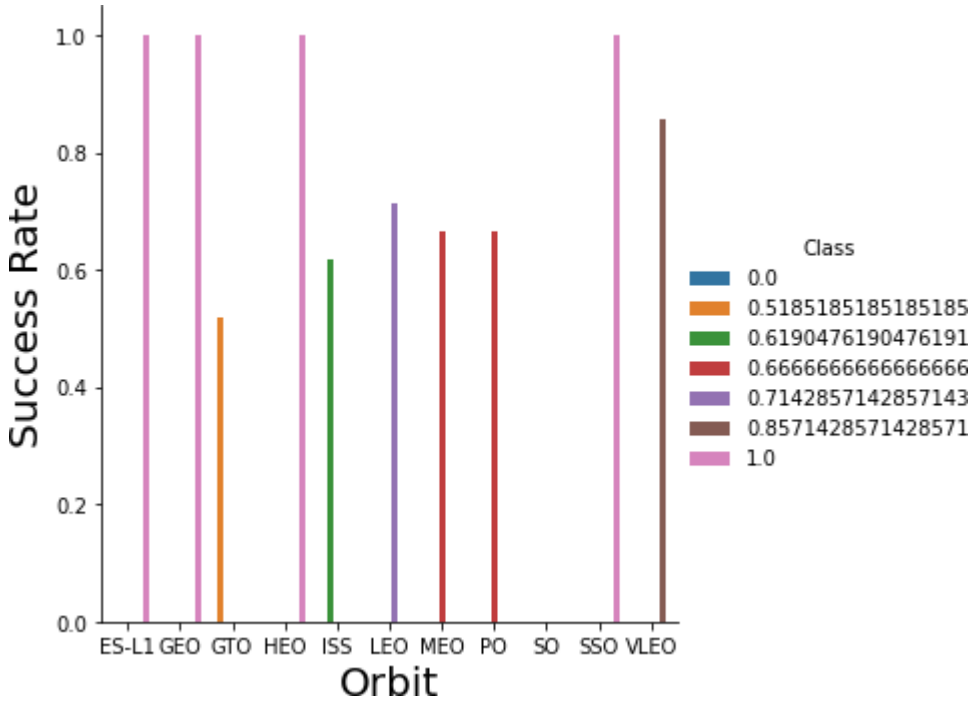
Payload vs. Launch Site



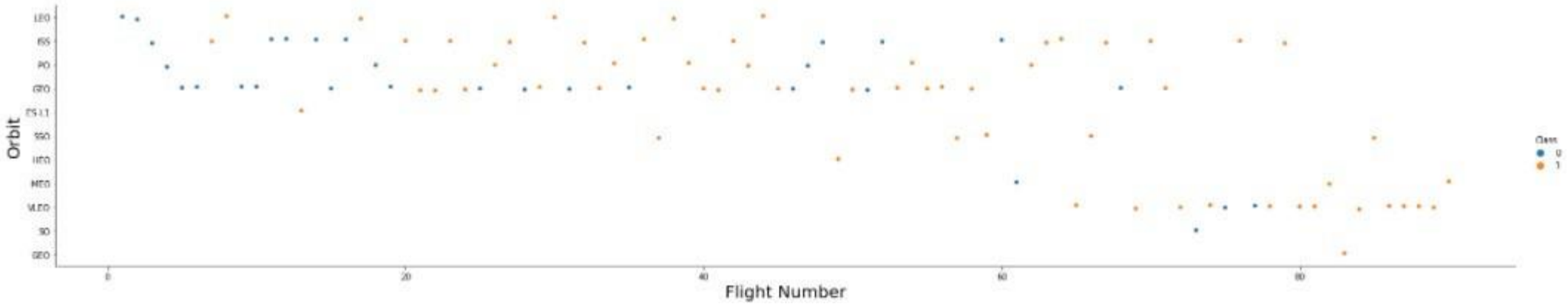
The highest pay load mass is in CCAPS SLC-40 launch site

Success rate vs. Orbit type

ES-L1, GEO, HEO, and SSO orbit has highest success rate

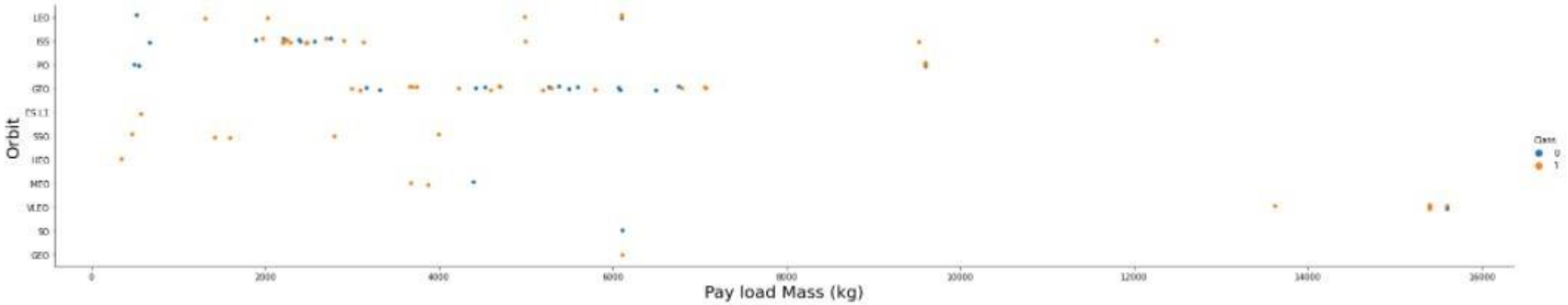


Flight Number vs. Orbit type



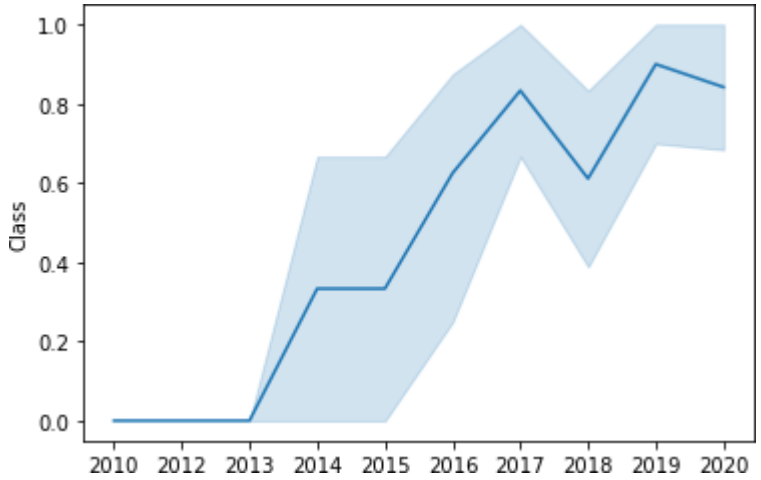
LEO, ISS, PO, and GEO orbit has highest flight number

Payload vs. Orbit type



GEO orbit has highest pay load mass

Launch success yearly trend



Every year the launch success rate increases

EDA with SQL

All launch site names

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

Theres four launch site

Launch site names begin with `CCA`

launch_site
CCAFS LC-40
CCAFS SLC-40

Theres two launch site names begin with ‘CCA’

Total payload mass

payload_mass
56479

Total pay load mass 56479 kg

Average payload mass by F9 v1.1

avg_payload_mass
3676

Average payload mass by F9 v1.1 is 3676

First successful ground landing date

first_succesful_landing
2017-01-05

First successful ground landing on 2017-01-05

Successful drone ship landing with payload between 4000 and 6000

booster_version
F9 FT B1022
F9 FT B1031.2

Successful drone ship landing with payload between 4000 and 6000 is F9 FT B1022 and F9 FT B1031.2

Total number of successful and failure mission outcomes

total_success	total_failure
46	0

Total number of successful is 46 and failure is 0

Boosters carried maximum **payload**

booster_version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3

Boosters carried maximum payload is F9 B5 B1048.4, F9 B5 B1049.4, F9 B5 B1049.5, F9 B5 B1060.2, and F9 B5 B1058.3

2015 launch records

month_name	landing__outcome	booster_version	launch_site
October	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40

In 2015 the launch was carried out in october at the ccafs lc-40 launch site with booster version f9 v1.1 B1012 and landing outcome is failure (drone ship)

Rank success count between 2010-06-04 and 2017-03-20

landing__outcome	successful_landing_outcomes
No attempt	7
Failure (drone ship)	3
Success (drone ship)	2
Success (ground pad)	2
Controlled (ocean)	1
Failure (parachute)	1

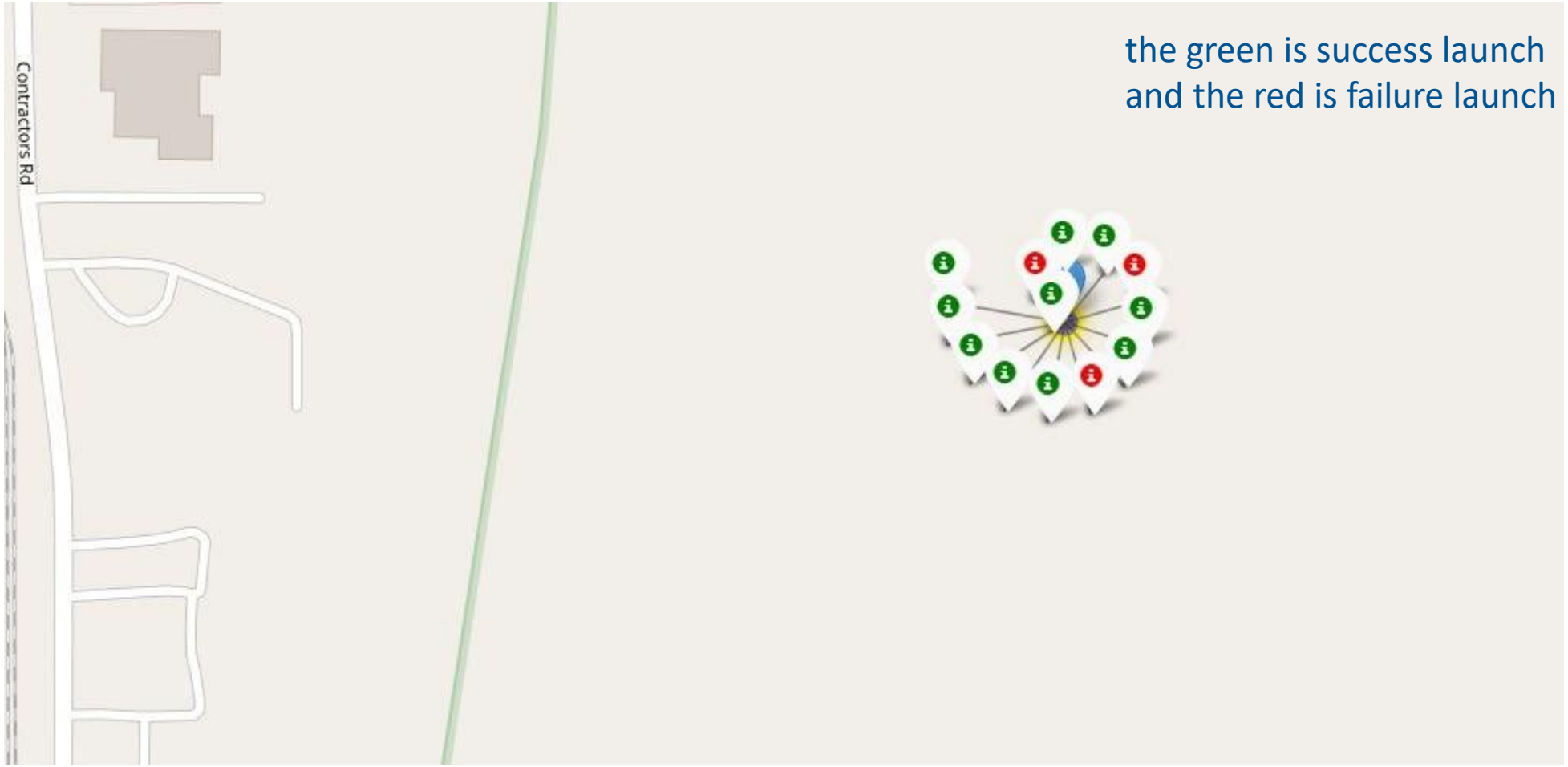
landing outcome with no attempt status was ranked first with a total of 7 times and landing outcome with controlled (ocean) and failure (parachute) status was ranked last with a total of 1 time

Interactive map with Folium

All Launch Sites

Theres 3 location launch site

Color-labeled Launch Records

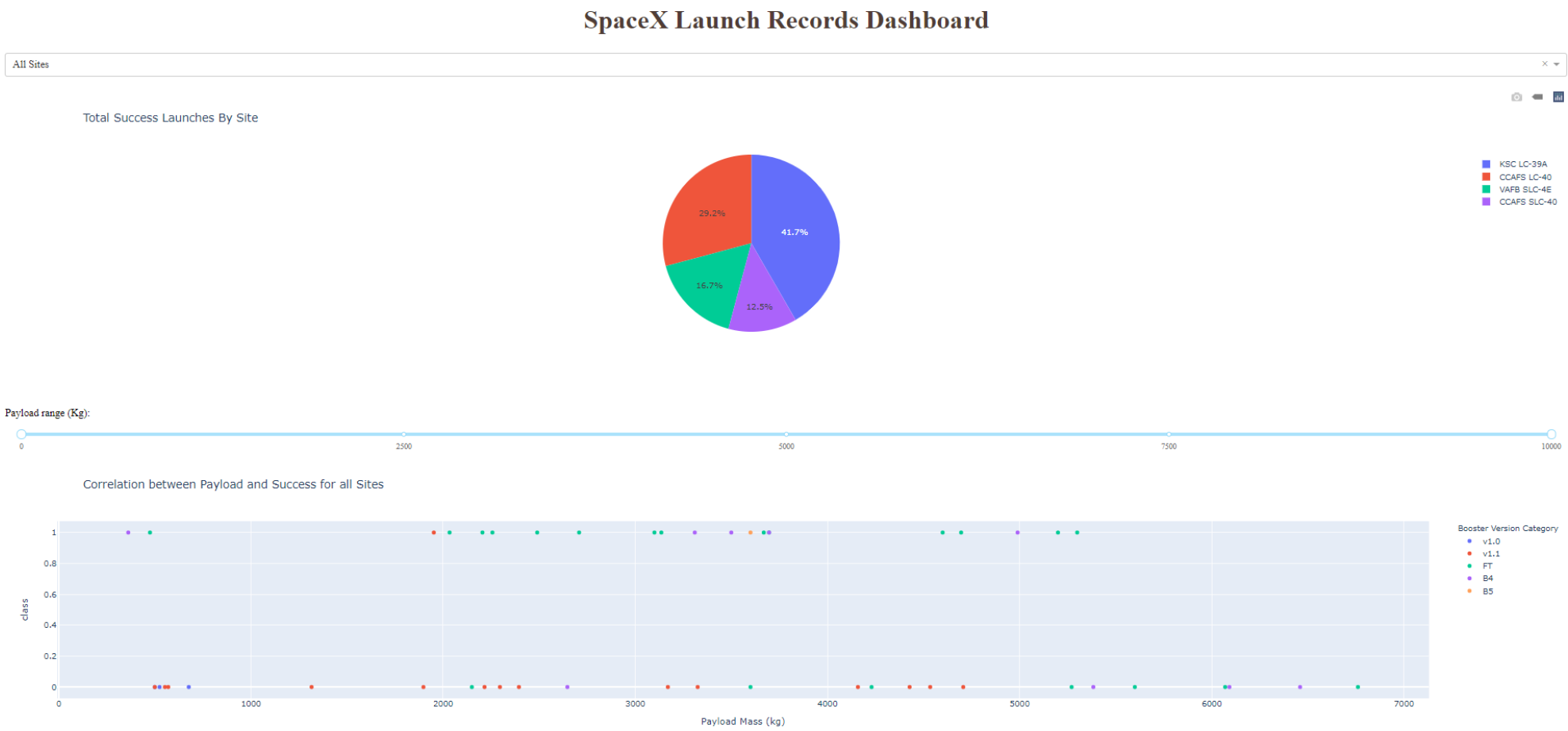


Launch Site to its Proximities Such as Railway, Highway, Coastline

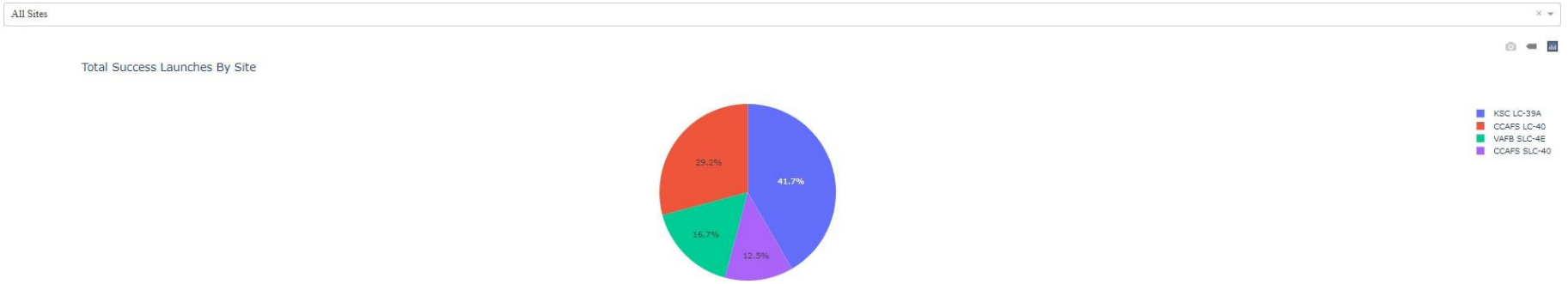
to find out how far is the distance between the launch site and the railway, highway, coastline

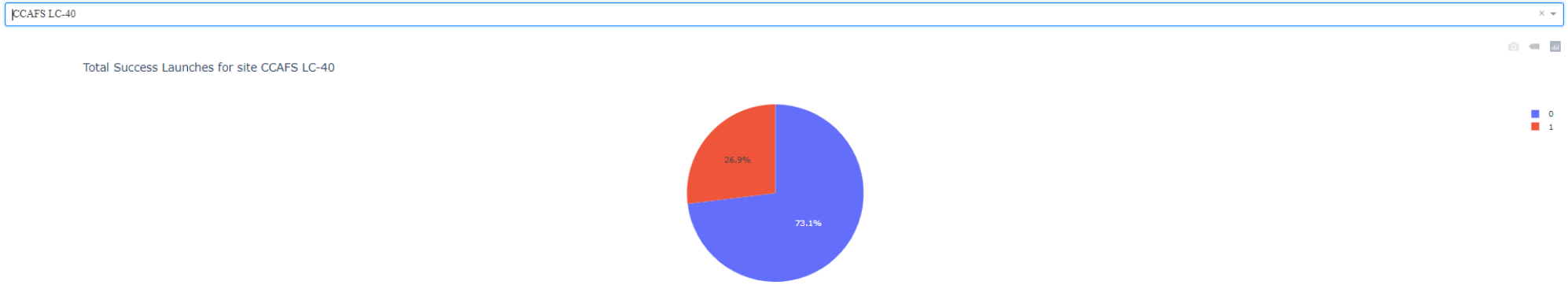
Build a Dashboard with Plotly Dash

Launch Success Count for All Sites, in a Piechart

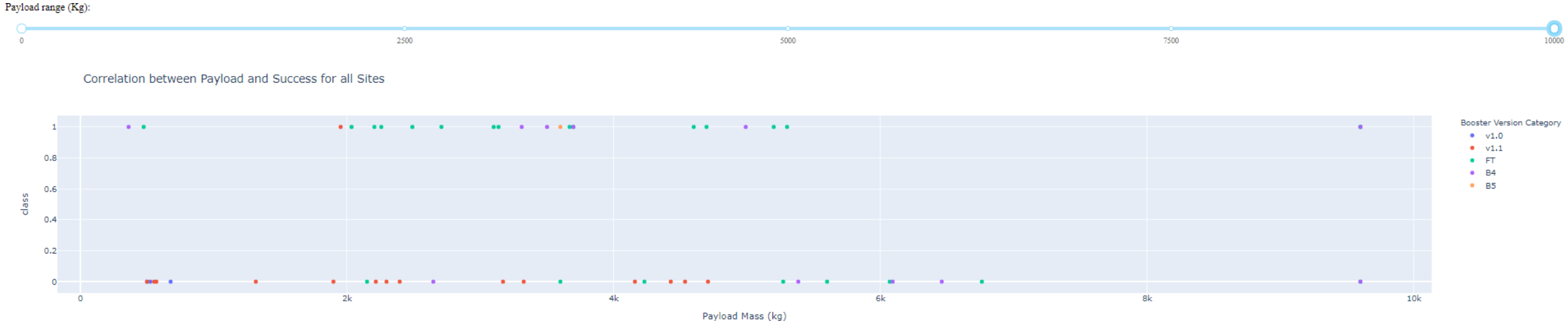


Piechart for The Launch Site With Highest Launch Success Ratio





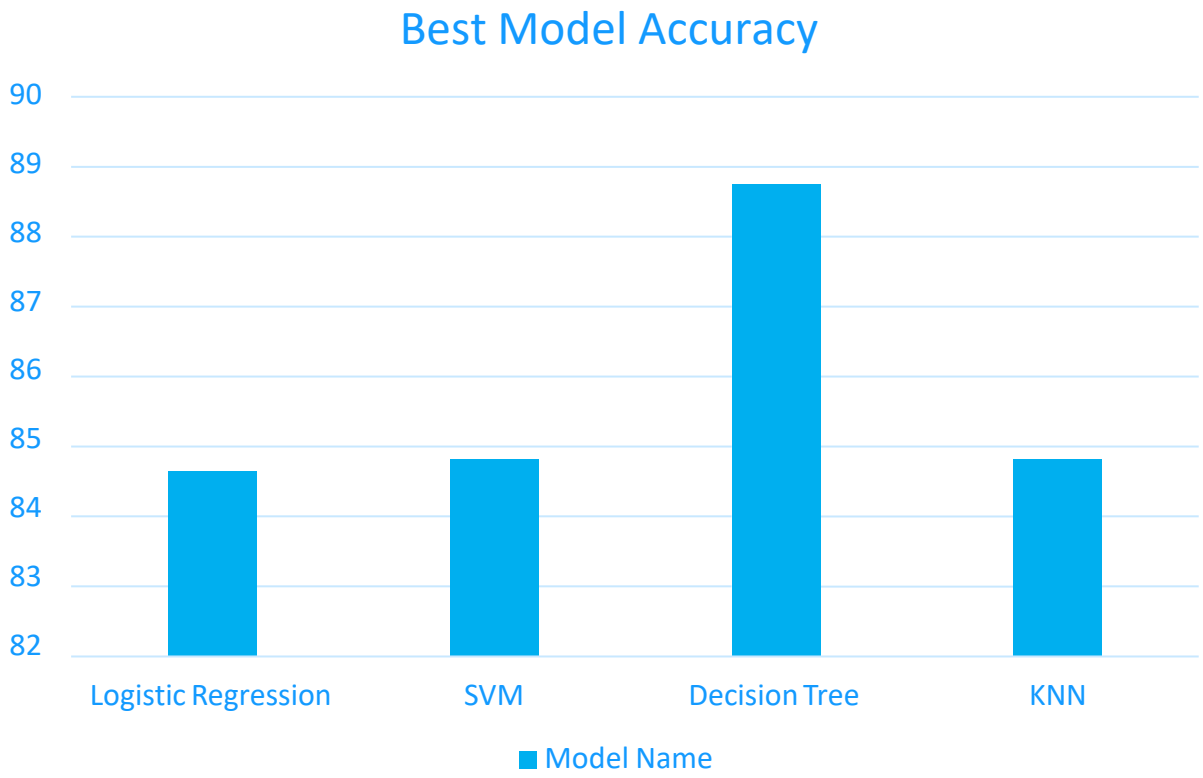
Payload vs. Launch Outcome Scatter Plot for All Sites



Predictive analysis (Classification)

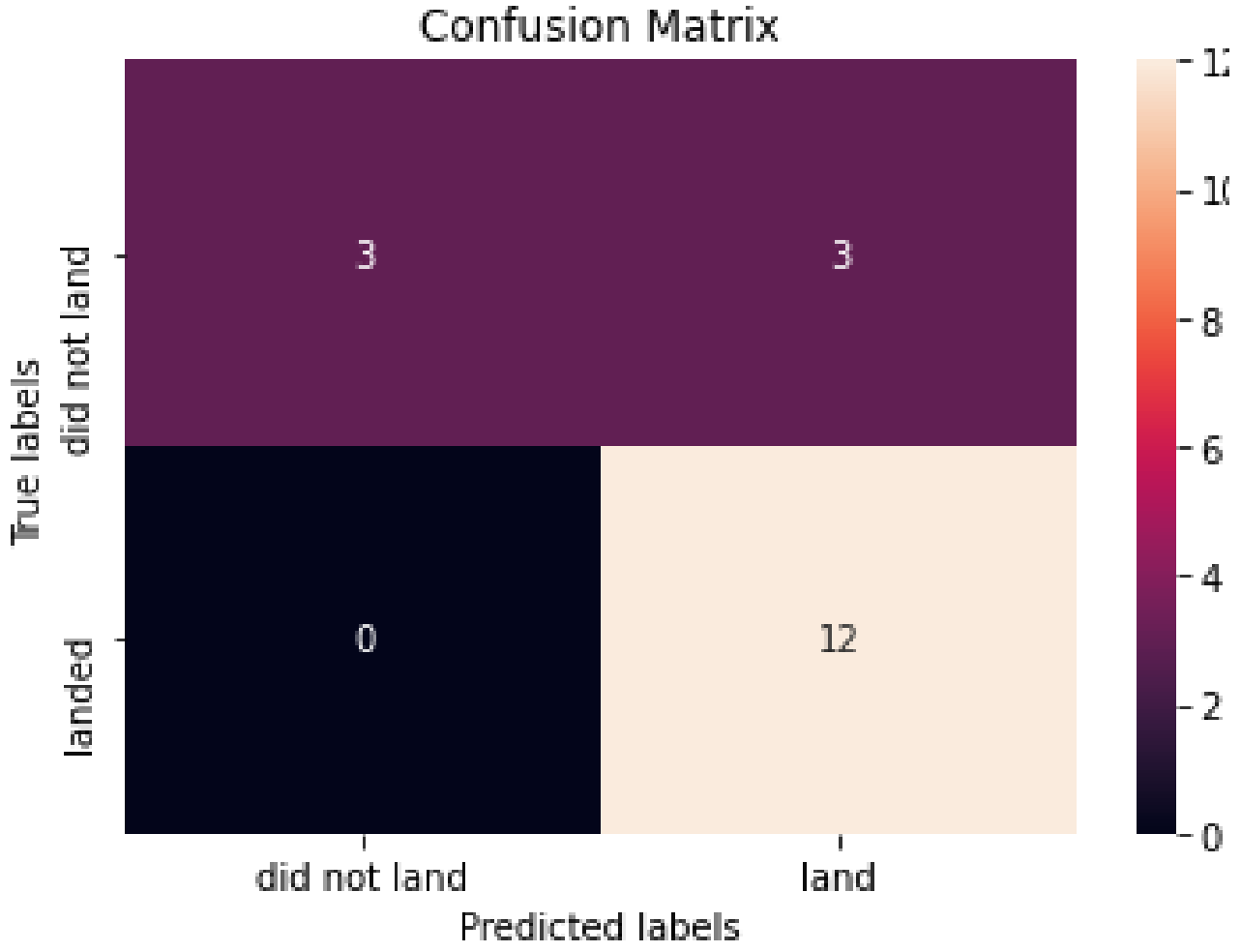
Classification Accuracy

The highest classification accuracy is Decision Tree with accuracy = 88,75%



Confusion Matrix

The best perfoming nodel is Decision Tree with accuracy = 88.75%



CONCLUSION

- The best classification model in this case is Decision Tree with 88,75% accuracy

APPENDIX

- https://github.com/MonuBhagat11/Data_Science_Capstone_Project_ppt

Thank You