Data Science Project

- **Space X Falcon 9 First Stage Landing Prediction**

- Lin Hung

- 2023/05/06

Outline

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

Executive Summary

- Summary of methodologies
- Data collection
- Data wrangling
- Data visualization and interactive dashboard
- Machine learning prediction

- Summary of all results
- Visualized graphs(both screenshots and interactive)
- EDA results
- ML results

Introduction

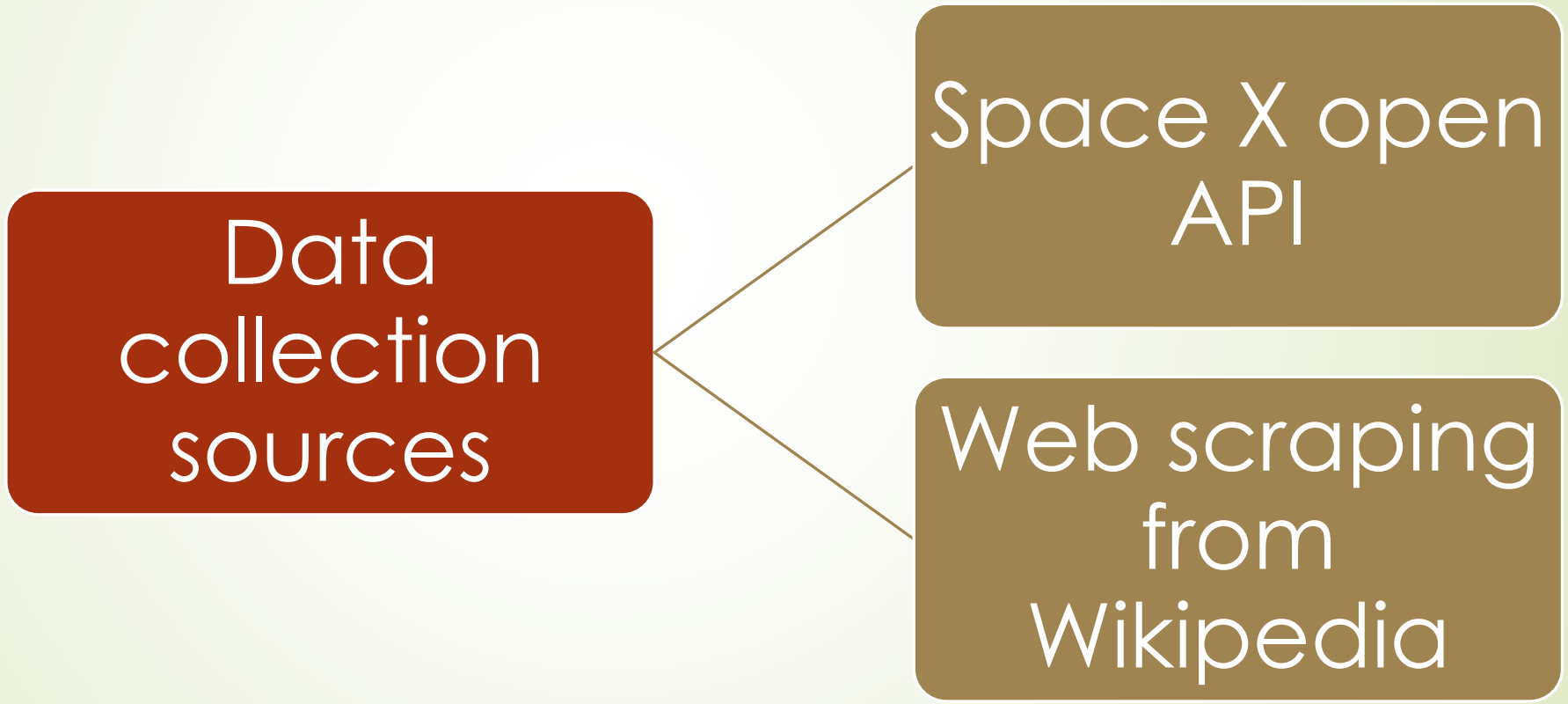
- Background
- Space X has lower rocket launch cost than other competitors
- Newly found company Space Y would like to compete with Space X by firstly predict the Falcon 9 first stage landing rate
- Problem to be answered:
- Can Space Y successfully predict Falcon 9 first stage landing using only machine learning method?

Methodology

Methodology

- Executive Summary
- Data collection methodology:
  Data was collected from Space X API and web scraping
- Perform data wrangling
  Data was processed and transformed into different labels for testing and training
- 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 was cleaned and split into test and train labels for logistic regression, SVM, decision tree, KNN models

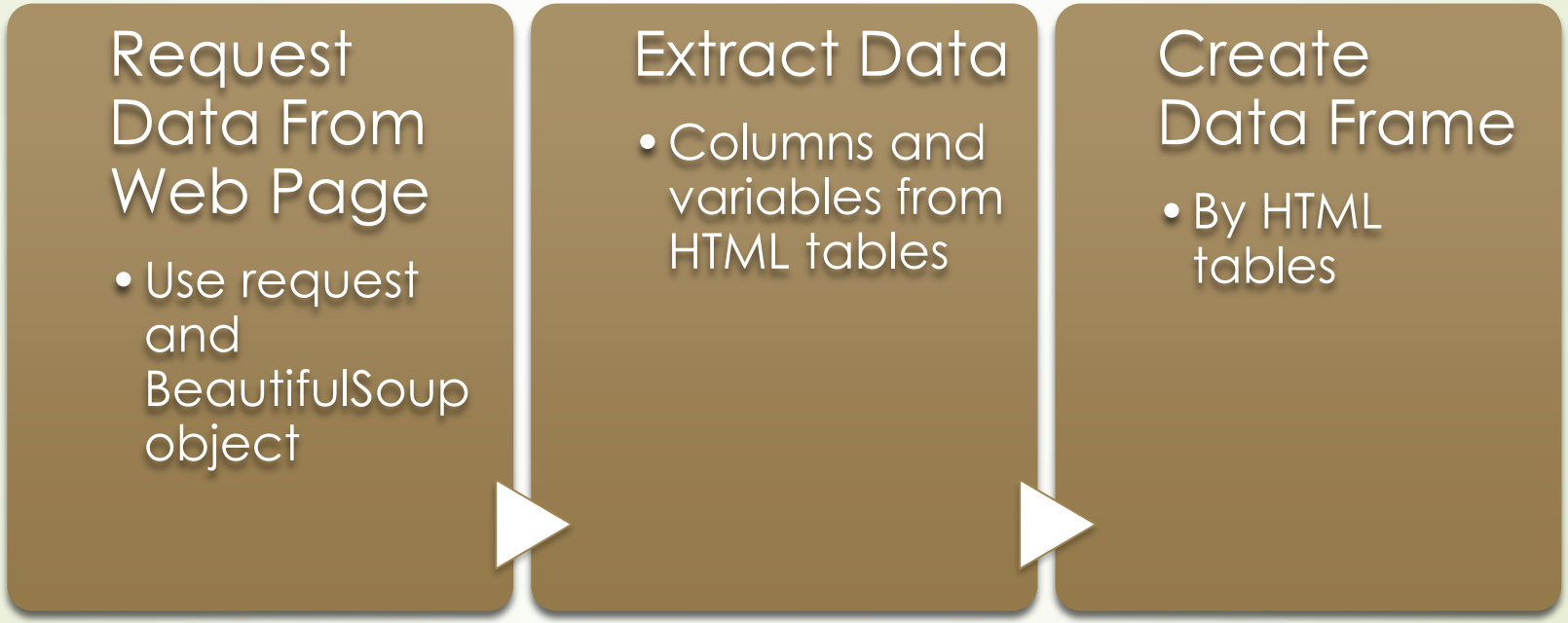
Data Collection



Data Collection – SpaceX API

- GitHub URL – Data Collection

Data Collection - Scraping



- GitHub URL - Webscraping

Data Wrangling

- Because there were failure and successful booster cases, we need to convert them into different labels or outcomes to preform EDA:

1. **Calculate the number of launches on each site**
2. **Calculate the number and occurrence of each orbit**
3. **Calculate the number and occurence of mission outcome per orbit type**
4. **Create a landing outcome label from Outcome column**

- GitHub URL – Data Wrangling

EDA with Data Visualization

- Mainly scatterplots were used to explore the following features:
  - Fight Number v. Payload Mass
  - Fight Number v. Launch Site
  - Payload Mass v. Launch Site
  - Fight Number v. Orbit
  - Payload Mass v. Orbit

- GitHub URL – Data Visualization

EDA with SQL

10 SQL queries are performed:

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first succesful landing outcome in ground pad was acheived.
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List the names of the booster_versions which have carried the maximum payload mass. Use a subquery
- List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
- Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.
- GitHub URL - SQL

Build an Interactive Map with Folium

- 3 map objects are used:
- Markers: Labeling launch site, success or failed launches
- Circles: Adding text circles for locations
- Lines: Indicates distance between specific coordinate and launch site(ex: coastline)

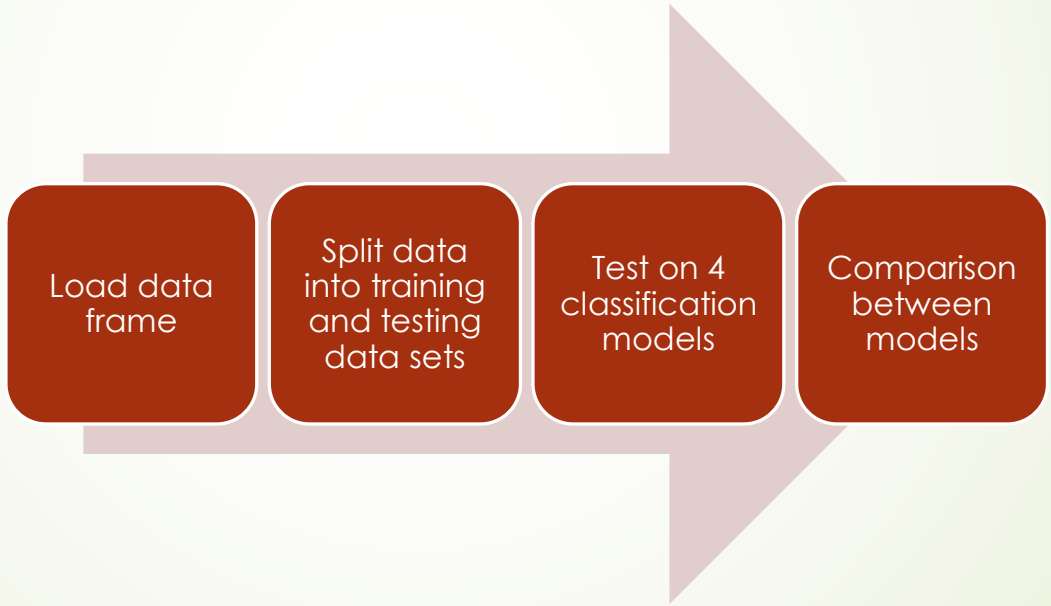
- GitHub URL – site location

Build a Dashboard with Plotly Dash

- plots/graphs used in interactive dashboard:
- Dropdown list and slider: For selection of launch site or payload range
- Pie chart: Shows successful and failed launch counts for selected launch site
- Scatter charts: Display correlation between features

- GitHub URL - Dashboard

Predictive Analysis (Classification)



- GitHub URL – Machine learning prediction

Results

EXPLORATORY DATA ANALYSIS RESULTS

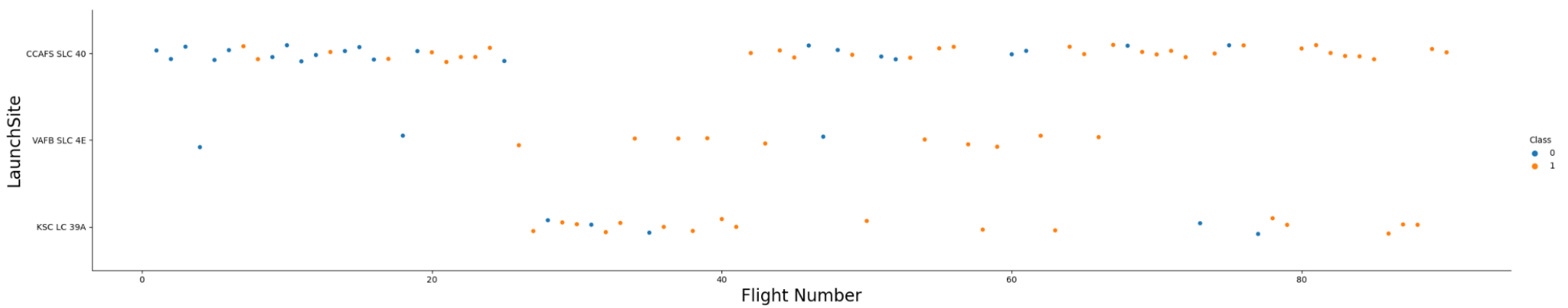
INTERACTIVE ANALYTICS DEMO IN SCREENSHOTS

PREDICTIVE ANALYSIS RESULTS

Insights drawn from EDA

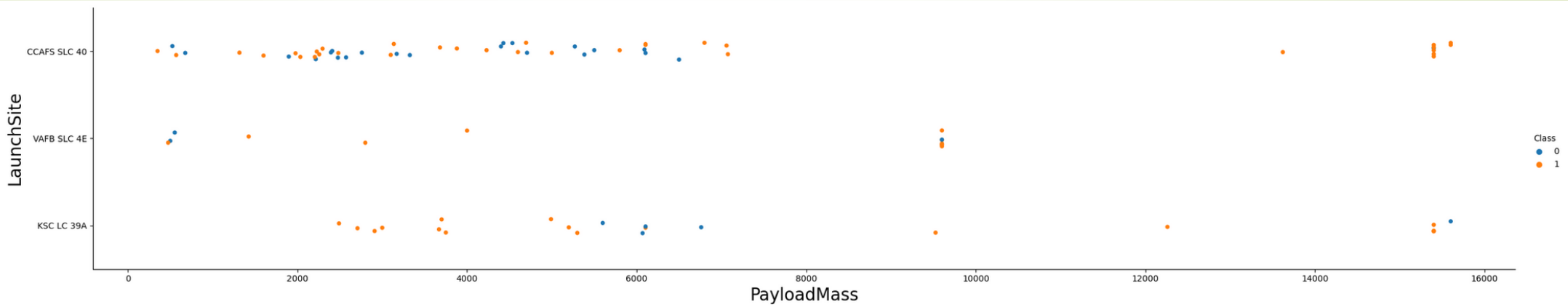
Flight Number vs. Launch Site

- Class 0 = unsuccessful launch; Class 1 = successful launch
- CCAFS SLC 40 is the most used launch site
- Successful rate has increased through times



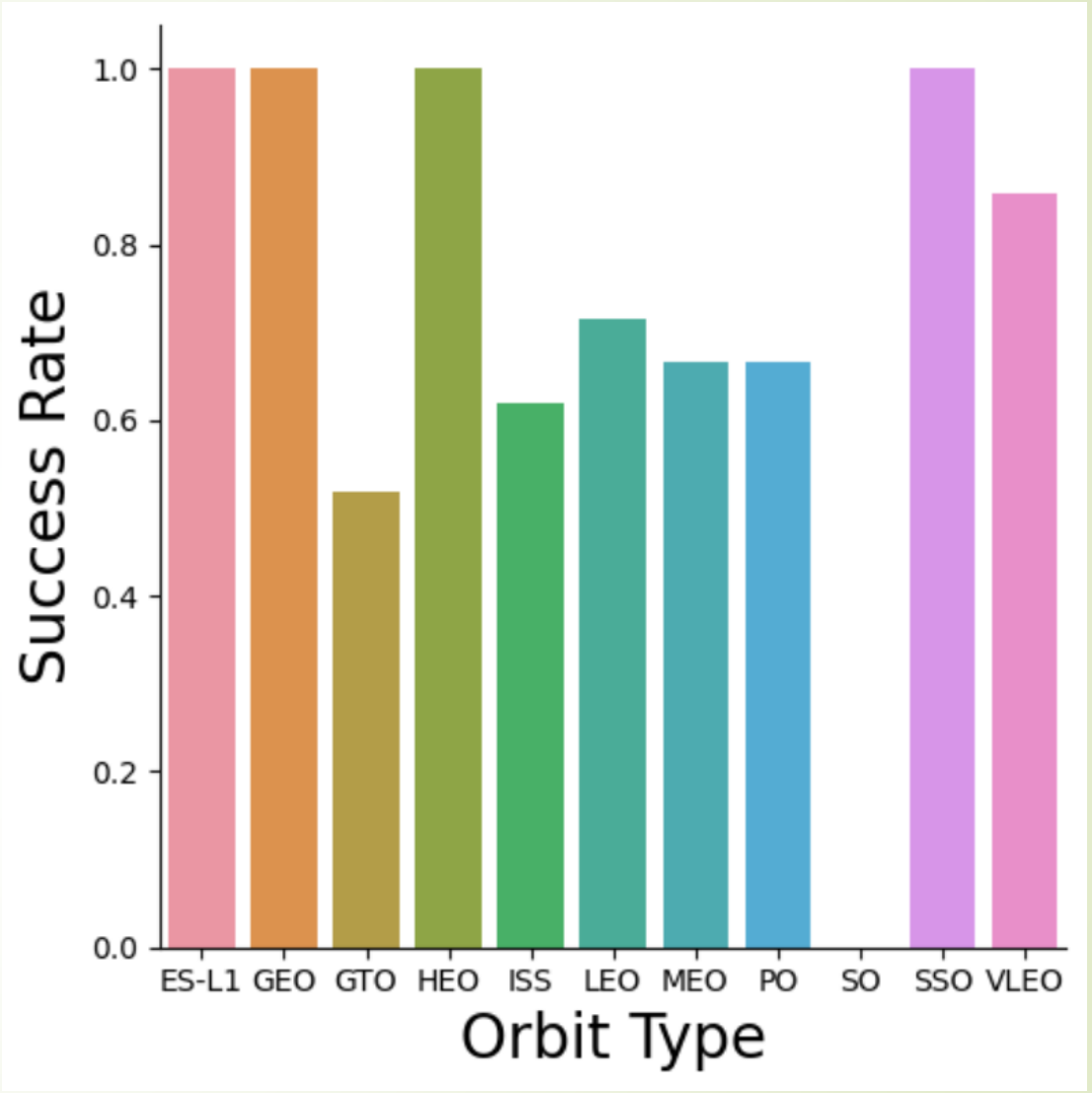
Payload vs. Launch Site

- Most sites tested payload mass under 8000 kg
- KSC LC 39A preformed 0 failure rate for payload mass under 4000kg



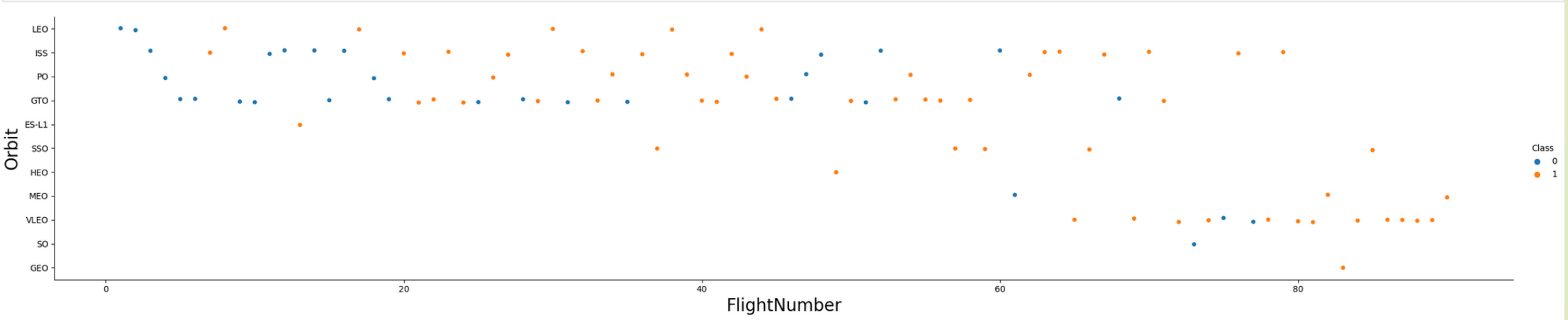
Success Rate vs. Orbit Type

- Most orbit types have over 50% successful rate
- SO is the only orbit type with 0% successful rate
- ES-L1, DEO, HEO and SSO have 100% successful rate



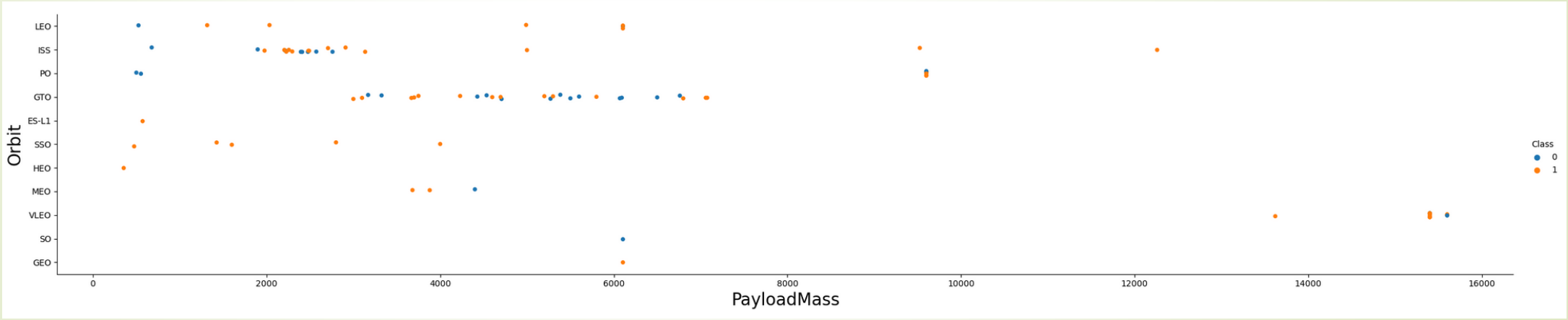
Flight Number vs. Orbit Type

- Most orbits improve their success rate over time
- LEO has 100% success rate after 10 flights
- No flights succeed before 10 flights



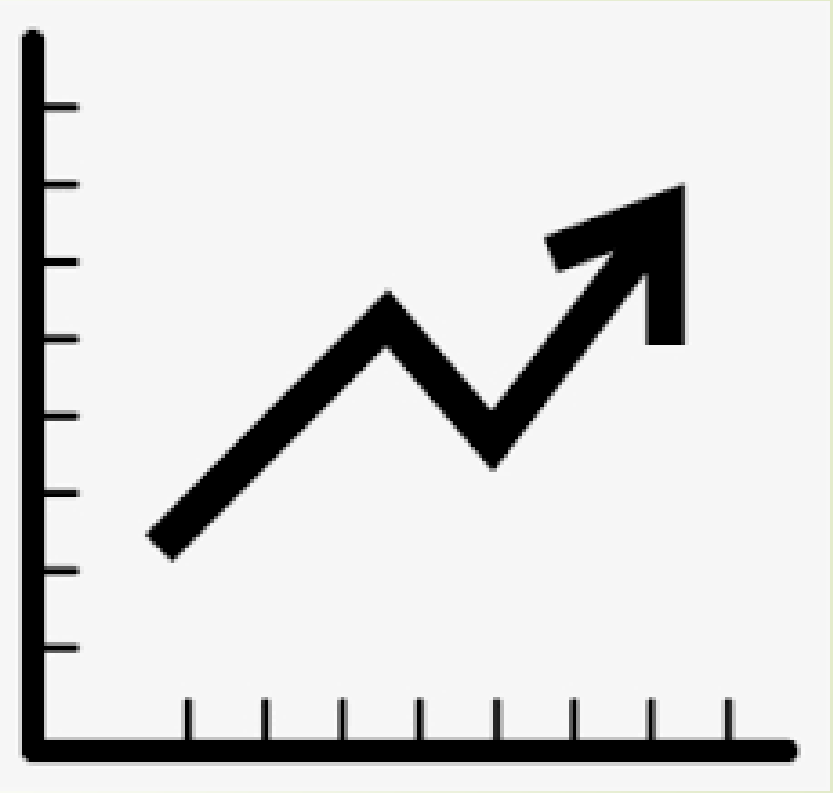
Payload vs. Orbit Type

- LEO succeed all flights with payload mass over 1000KG
- SSO and ES-L1 succeed all their payloads



Launch Success Yearly Trend

- Based on the scatter charts from previous slides, launch success rate has increased over time



All Launch Site Names

- There are total 4 launch sites:

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

Launch Site Names Begin with 'CCA'

- Samples of 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
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- Sum all payloads carried by boosters from NASA(CRS)

TOTAL_PAYLOAD
45596

Average Payload Mass by F9 v1.1

- Using AVG filter to calculate the average payload mass carried by booster version F9 v1.1

AVG_PAYLOAD
2928.4

First successful landing Date

- Date presenting in form of DD-MM-YYYY
- Select date from the table and set landing outcome as success(ground pad)

first_successful_landing
22-12-2015

Successful Drone Ship Landing with Payload between 4000 and 6000

- Filtering the payload mass kg between 4000 and 6000 will get 4 results as following:

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

Total Number of Successful and Failure Mission Outcomes

- Grouping mission outcome shows the result of 99 success number

Mission_Outcome	total_number
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- There are total 12 boosters carried maximum payload mass

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
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

- There are 2 failed landing in drone ship for in year 2015

Date	Booster_Version	Launch_Site	Landing _Outcome
10-01-2015	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
14-04-2015	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

- All landing outcomes are listed between the selected date

Landing _Outcome	numbers
Success	20
No attempt	10
Success (drone ship)	8
Success (ground pad)	6
Failure (drone ship)	4
Failure	3
Controlled (ocean)	3
Failure (parachute)	2
No attempt	1

Launch Sites Proximities Analysis

All launch sites

- All launch sites are near the coastline and are mostly located in Florida

Color labels of the map

- Successful landing in green icon and failed landing in red icon:
- The VAFB SLC-4E has higher failure rate than successful rate

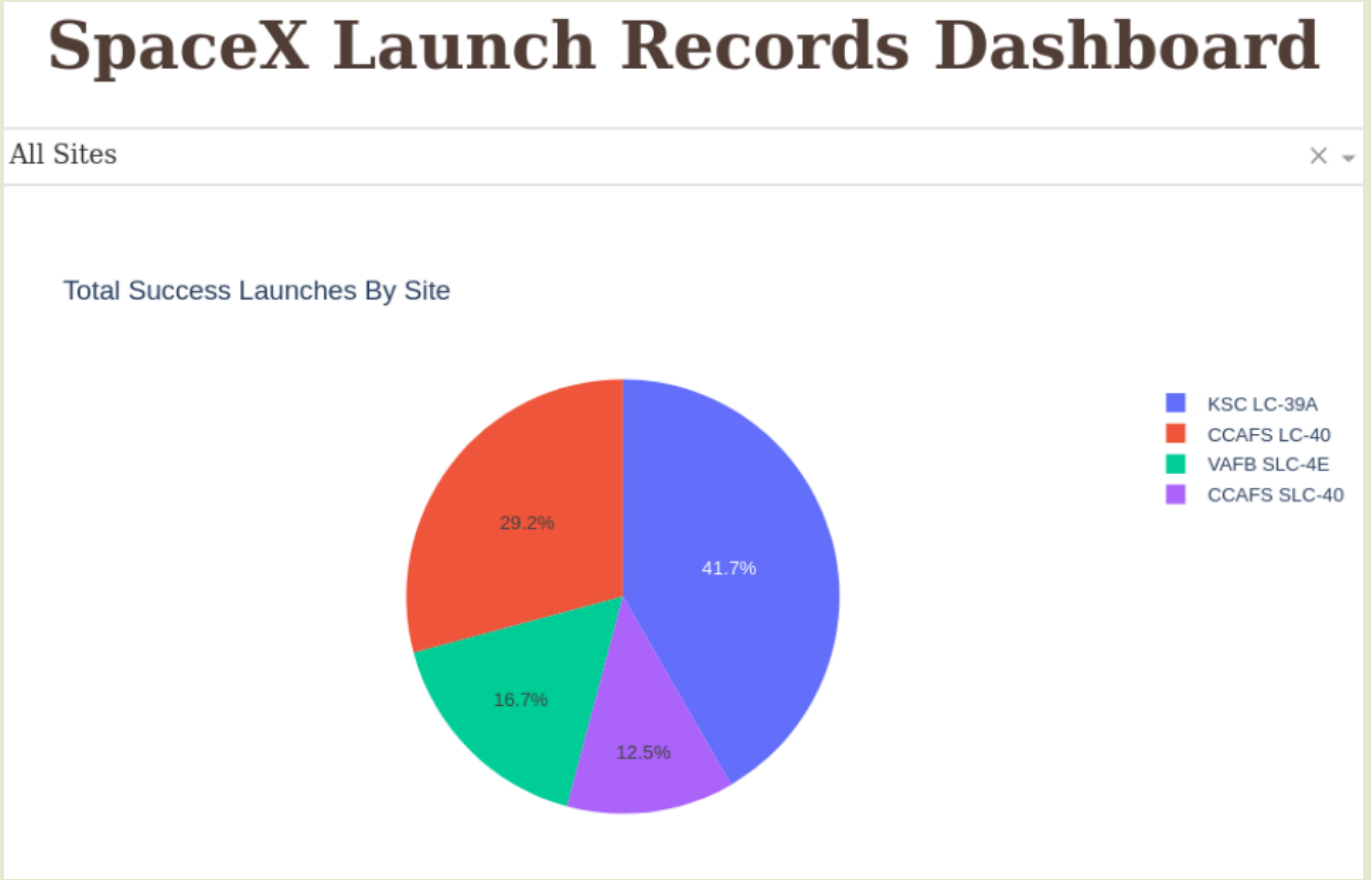
Launch site proximity to the coastline

- LC-39A is relatively close to the coastline compared to distance to city and highway as shown:

Build a Dashboard with Plotly Dash

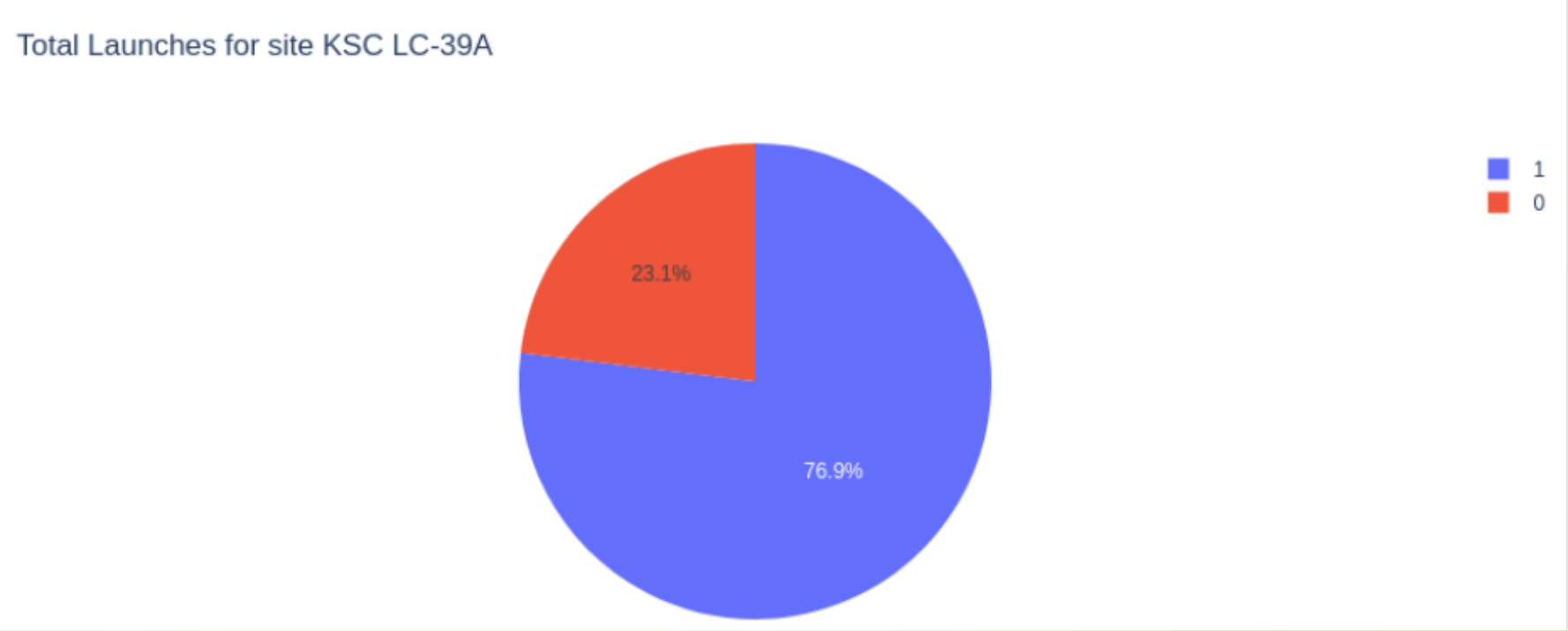
Success count in pie chart

- KSC LC-39 has the highest success rate



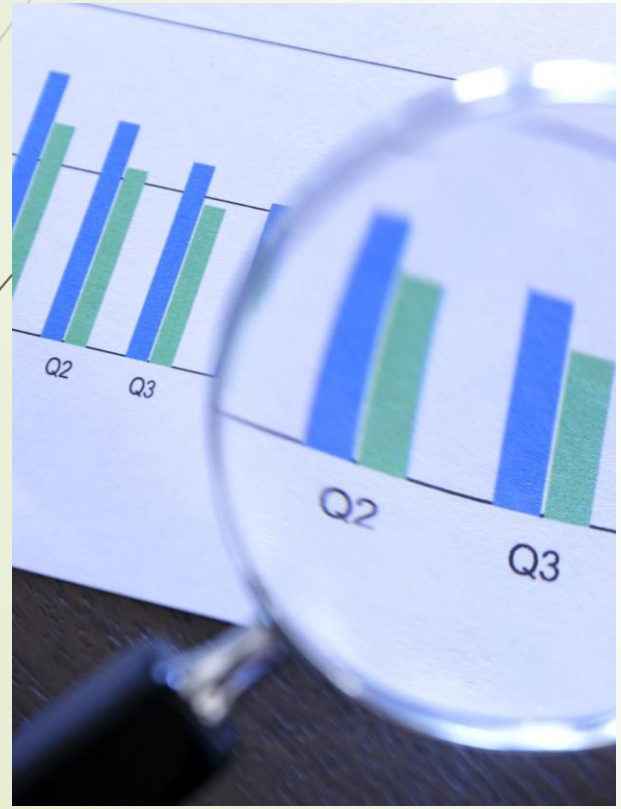
The launch site with the highest success rate

- KSC LC-39A has a success rate of 76.9%



Payload and Launch Outcome for all sites

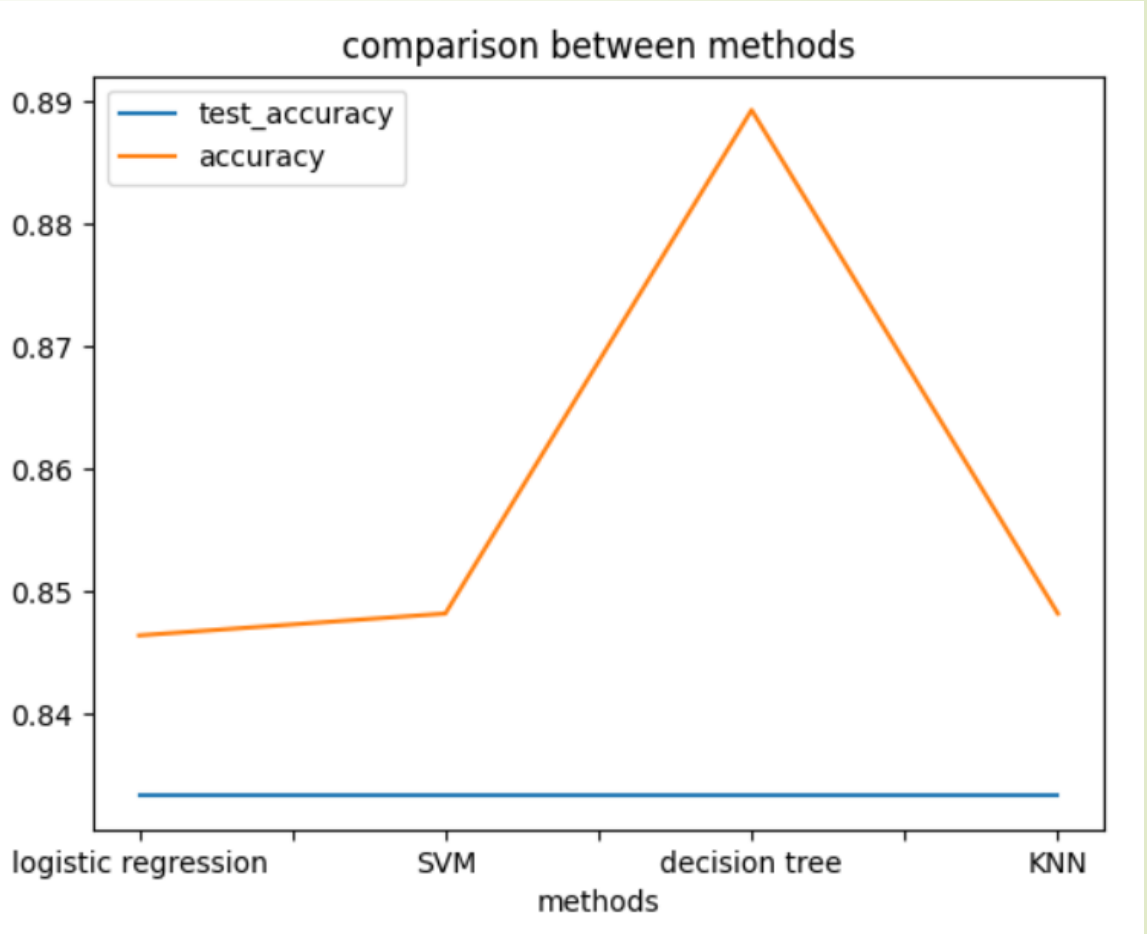
- Select different payload range slider for all sites



Predictive Analysis(Classification)

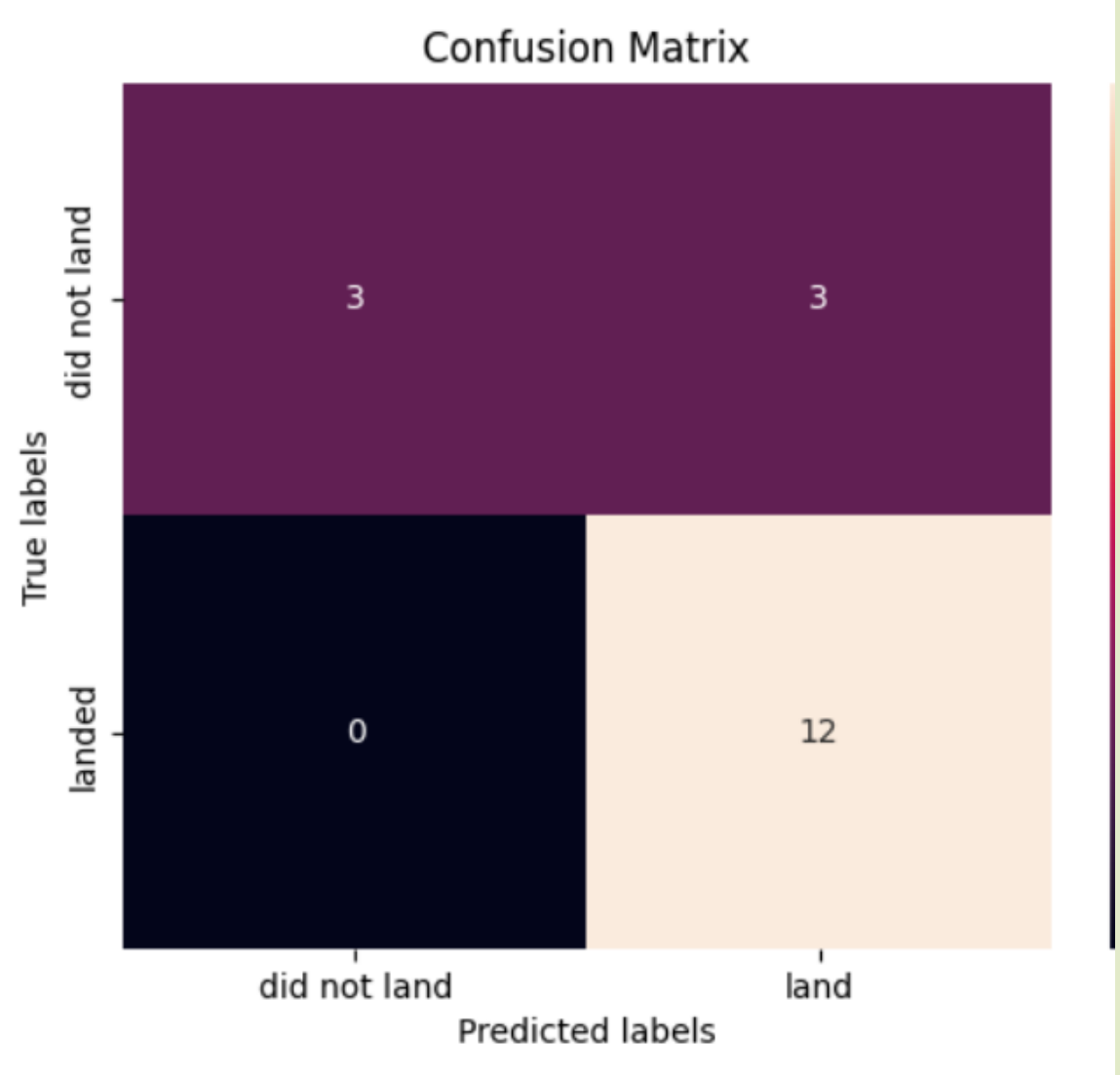
Classification Accuracy

- All 4 methods have same test accuracy result
- Decision tree has the highest actual accuracy



Confusion Matrix

- This matrix shows true positive and negative compared to the false one
- This matrix shows 12 successful landing when the true label is landed



Conclusions

- Decision Tree are best used to predict landing in this scenario
- All classification models in this project has a test accuracy of 83%
- KSC LC-39A is the most suitable launch site since it has the highest success launch rate
- Launch rate increased through times

Appendix

- GitHub URL for this project
- Special thanks to all instructors