IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Islam Ezz
<Date>

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection using web scraping and SpaceX API;
  - Exploratory Data Analysis (EDA), including data wrangling, data visualization and interactive visual analytics;
  - Machine Learning Prediction
- Summary of all results
  - It was possible to collected valuable data from public sources
  - EDA allowed to identify which features are the best to predict success of launchings
  - Machine Learning Prediction showed the best model to predict which characteristics are important to drive this opportunity by the best way, using all collected data

Introduction

- Project background and context
  - evaluate the viability of the new company Space Y to compete with Space X
- Problems you want to find answers
  - The best way to estimate the total cost for launches, by predicting successful landings of the first stage of rockets
  - Where is the best place to make launches

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - 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 enriched by creating a landing outcome label based on outcome data after summarizing and analyzing feature
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

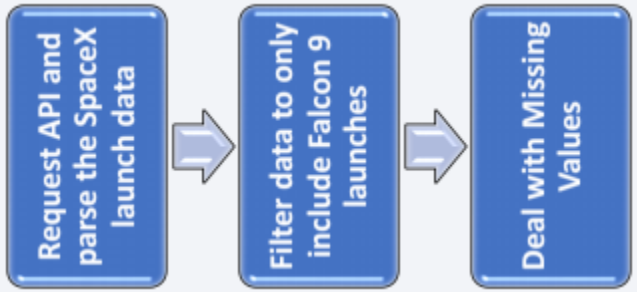
Methodology

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
    - Data that was collected until this step were normalized, divided in training and test data sets and evaluated by four different classification models, being the accuracy of each model evaluated using different combinations of parameters.

Data Collection

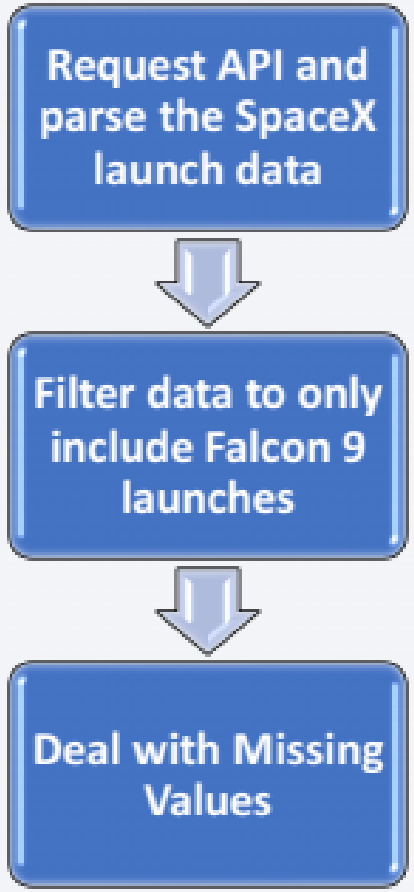
- Describe how data sets were collected.
- Data collection methodology:
- Space X API (https://api.spacexdata.com/v4/rockets/)
- WebScraping
- (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)
- You need to present your data collection process use key phrases and flowcharts



Data Collection – SpaceX API

according to the flowchart beside and then data is persisted.

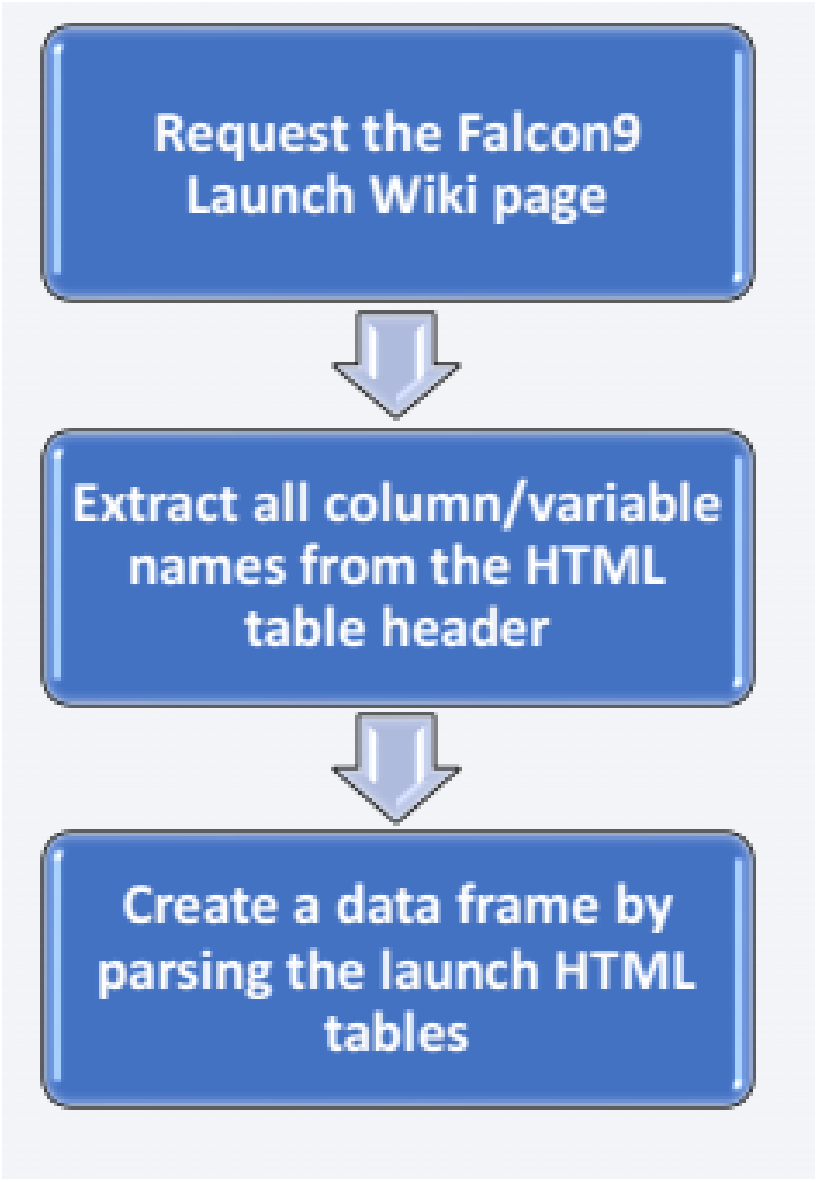
- Source code: https://github.com/islamezz26/Ibm-/blob/main/Data%20Collection%20Using%20API.ipynb



Data Collection - Scraping

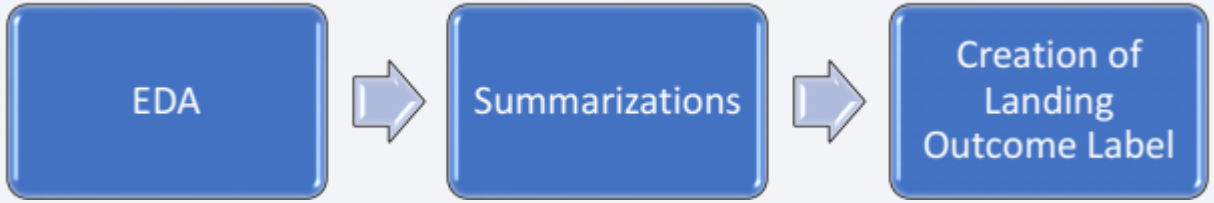
Data are downloaded from Wikipedia according to the flowchart and then persisted

https://github.com/islamezz26/Ibm-/blob/main/Data%20Collection%20with%20Using%20Web%20Scraping.ipynb



Data Wrangling

- Describe how data were processed
  - was performed on the dataset. Then the summaries launches per site, occurrences of each orbit and occurrences of mission outcome per orbit type were calculated. Finally, the landing outcome label was created from Outcome column .
- You need to present your data wrangling process using key phrases and flowcharts



- Add the GitHub URL of your completed data wrangling related notebooks, as an external reference and peer-review purpose
  - https://github.com/islamezz26/Ibm-/blob/main/Data%20Wrangling.ipynb

EDA with Data Visualization

- explore data, scatterplots and barplots were used to visualize the relationship between pair of features :
- Payload Mass X Flight Number, Launch Site X Flight Number, Launch Site X Payload Mass, Orbit and Flight Number, Payload and Orbit

Add the GitHub URL of your completed EDA with data visualization notebook, as an external reference and peer-review purpose

- https://github.com/islamezz26/Ibm-/blob/main/EDA%20with%20Data%20Visualization.ipynb

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed

  - • 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; 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.

- https://github.com/islamezz26/Ibm-/blob/main/SQL.ipynb

Build an Interactive Map with Folium

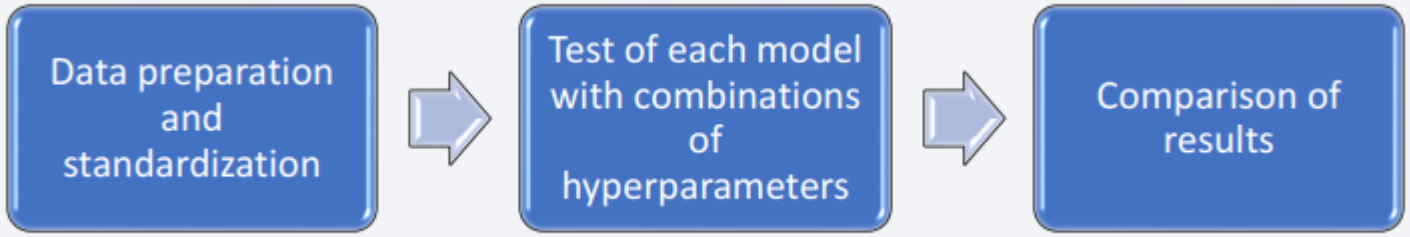
- Markers indicate points like launch sites;
- Circles indicate highlighted areas around specific coordinates, like NASA Johnson Space Center;
- Marker clusters indicates groups of events in each coordinate, like launches in a launch site; and Lines are used to indicate distances between two coordinates
- https://github.com/islamezz26/Ibm-/blob/main/Interactive%20Visual%20Analytics%20with%20Folium%20lab.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 analyze the relation between payloads
- and launch sites, helping to identify where is best place to launch according to payloads.
- https://github.com/islamezz26/Ibm-/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model : Four classification models were compared: logistic regression, support vectormachine, decision tree and k nearest neighbors.
- You need present your model development process using key phrases and flowchart



- https://github.com/islamezz26/Ibm-/blob/main/ML%20Prediction.ipynb

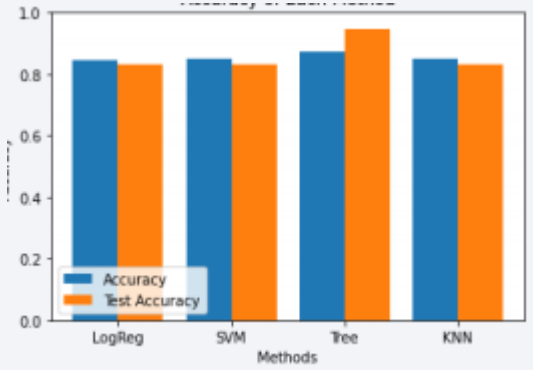
Results

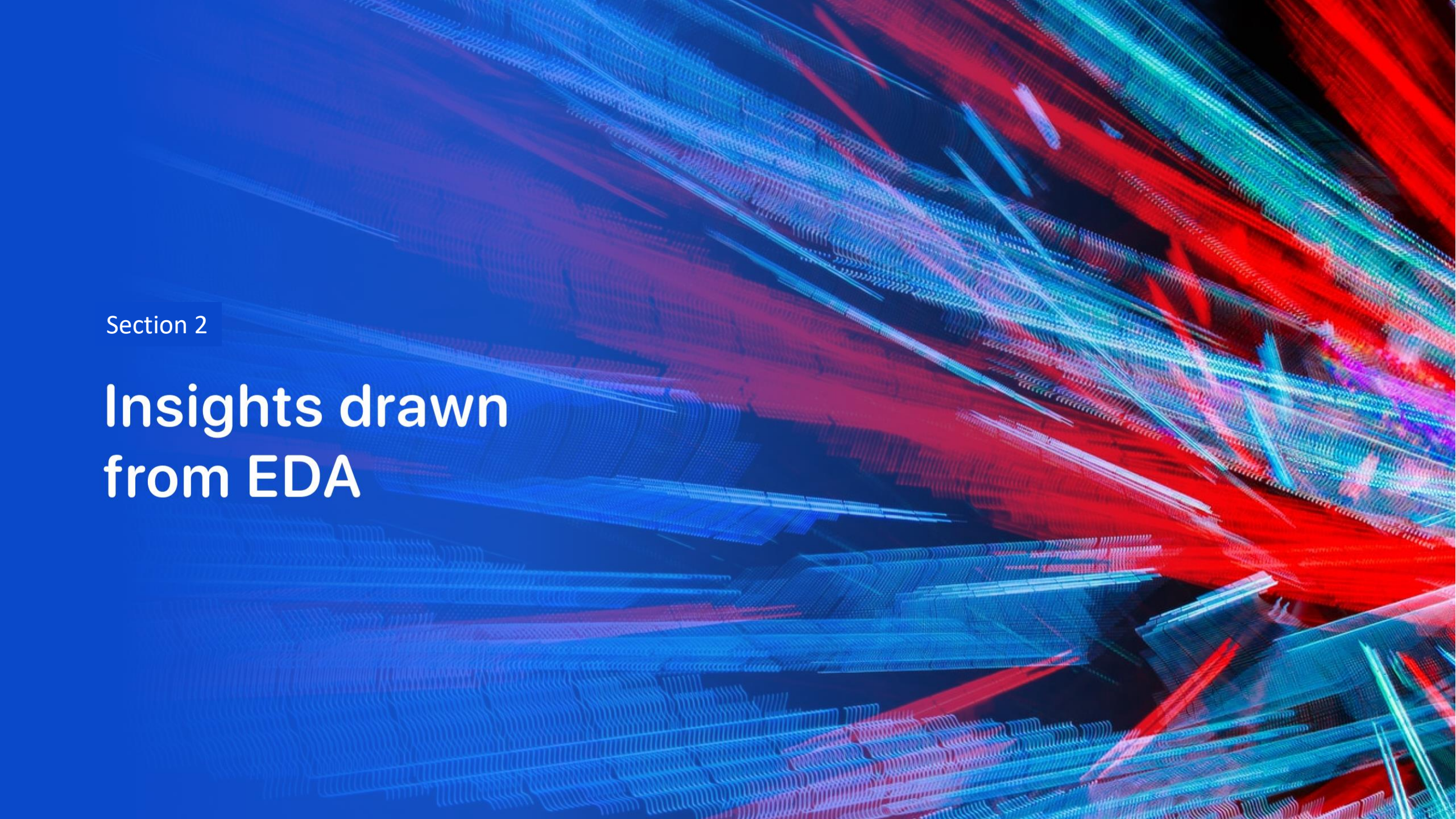
- Exploratory data analysis results
  - Space X uses 4 different launch sites;
  - The first launches were done to Space X itself and NASA;
  - The average payload of F9 v1.1 booster is 2,928 kg;
  - The first success landing outcome happened in 2015 fiver year after the first launch;
  - Many Falcon 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.
- Interactive analytics demo in screenshots
- Predictive analysis results

Results

- Interactive analytics demo in screenshots

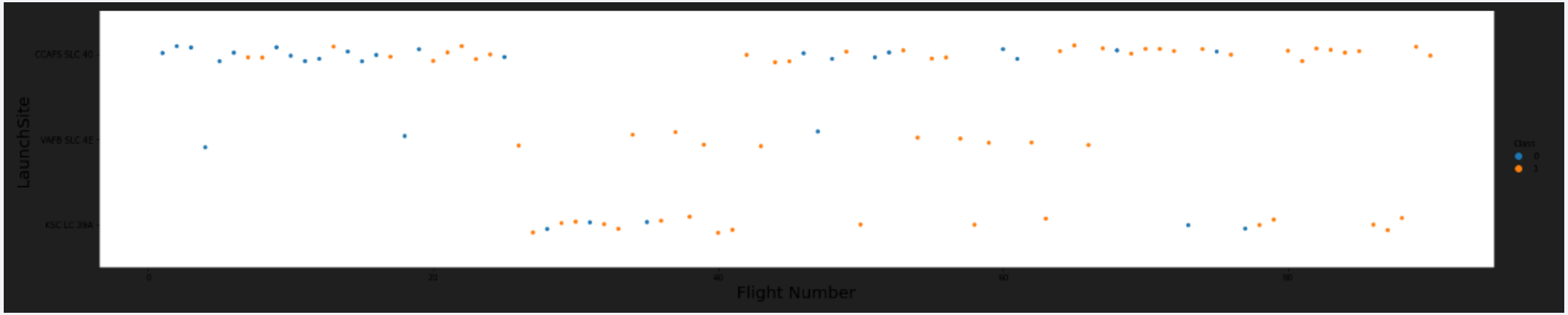
- Predictive analysis results
  - Predictive Analysis showed that Decision Tree Classifier is the best model to predict successful landings, having accuracy over 87% and accuracy for test data over 94%




Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

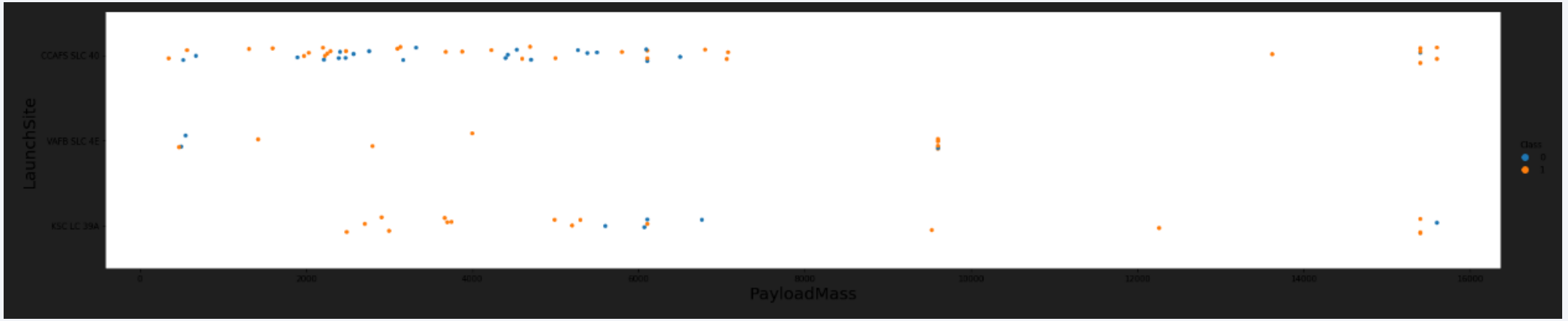
- Show a scatter plot of Flight Number vs. Launch Site



- Show the screenshot of the scatter plot with explanations
- The best launch site nowadays is CCAF5 SLC 40, where most of recent launches were successful; In second place VAFB SLC 4E and third place KSC LC 39A

Payload vs. Launch Site

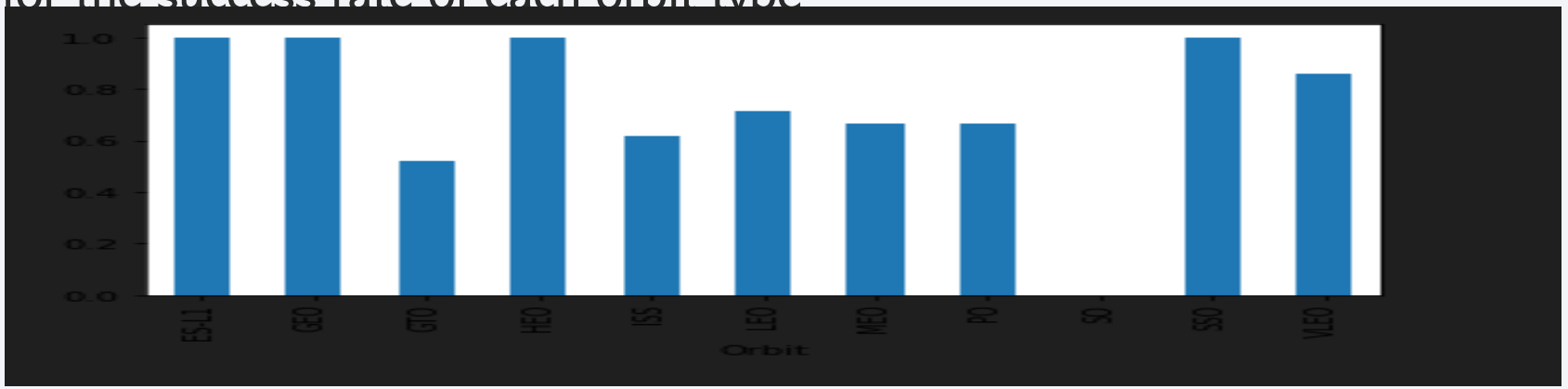
- Show a scatter plot of Payload vs. Launch Site



- Show the screenshot of the scatter plot with explanations
- Payloads over 9,000kg (about the weight of a school bus) 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

- Show a bar chart for the success rate of each orbit type

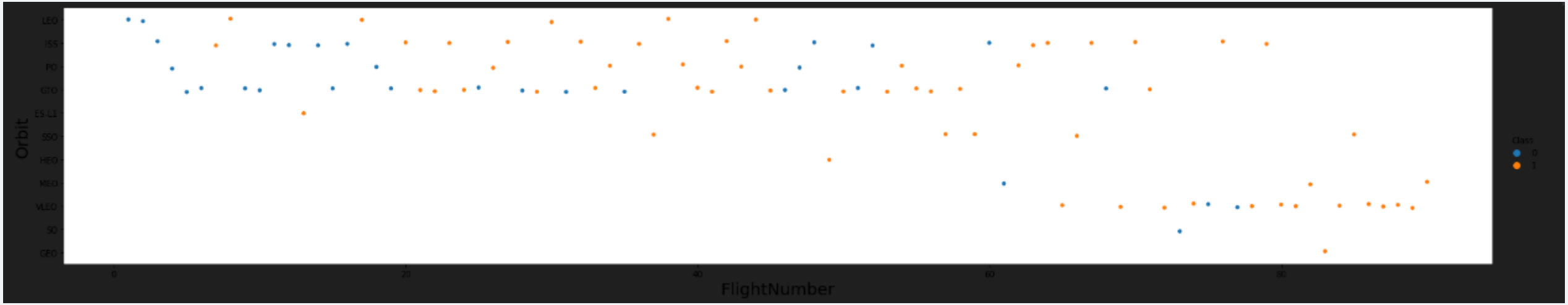


- Show the screenshot of the scatter plot with explanations

- The biggest success rates happens to orbits:
  - ES-L1;
  - GEO;
  - HEO; and
  - SSO.

Flight Number vs. Orbit Type

- Show a scatter point of Flight number vs. Orbit type



- Apparently, success rate improved over time to all orbits

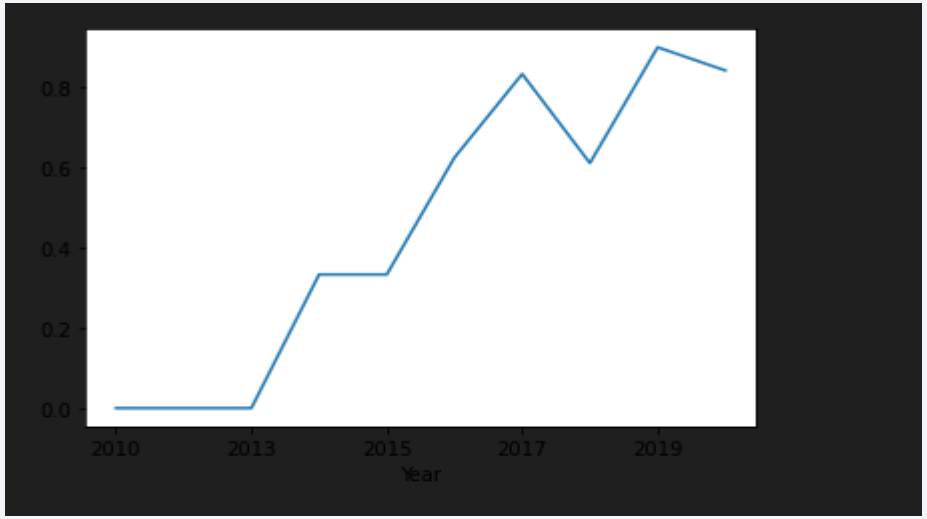
Payload vs. Orbit Type

- Show a scatter point of 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

- Show a line chart of yearly average success rate



- Success rate started increasing in 2013 and kept until 2020

All Launch Site Names

- Find the names of the unique launch sites

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

Launch Site Names Begin with 'CCA'

- Find 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
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

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
  - 111.268
- Present your query result with a short explanation here
  - summing all payloads whose codes contain ‘CRS’, which corresponds to NASA.

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
  - 2,928
- Average payload mass we obtained the value of 2,928 kg

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
  - 12/22/2015
- The minimum value for date it's possible to identify the first occurrence, that happened on 12/22/2015

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

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

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

Success	99
Success (payload status unclear)	1
Failure (in flight)	1

- the total number of successful and failure mission outcomes is above

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

Booster Version (...)	Booster Version
F9 B5 B1048.4	F9 B5 B1051.4
F9 B5 B1048.5	F9 B5 B1051.6
F9 B5 B1049.4	F9 B5 B1056.4
F9 B5 B1049.5	F9 B5 B1058.3
F9 B5 B1049.7	F9 B5 B1060.2
F9 B5 B1051.3	F9 B5 B1060.3

2015 Launch Records

- List the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

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

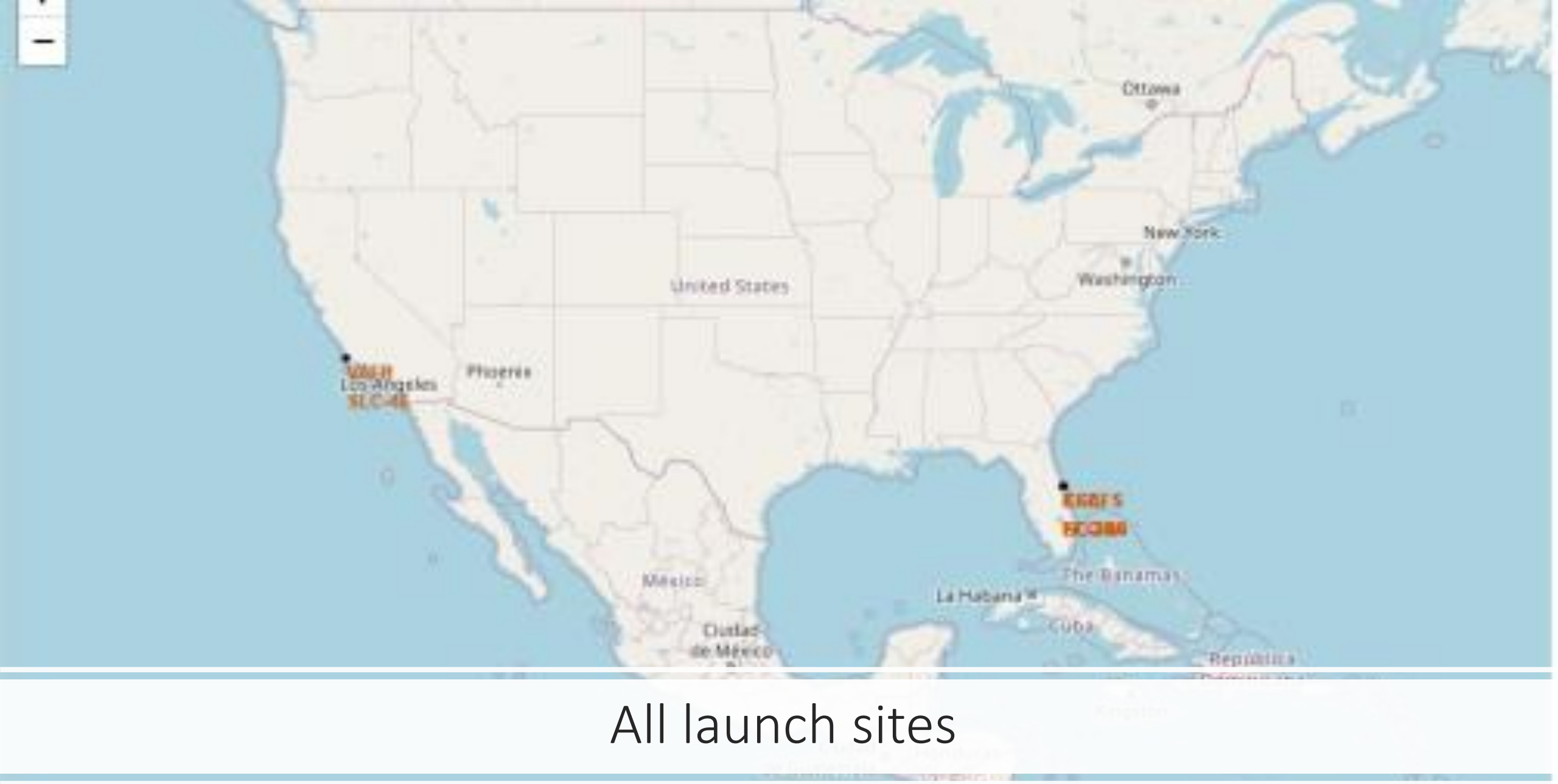
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

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

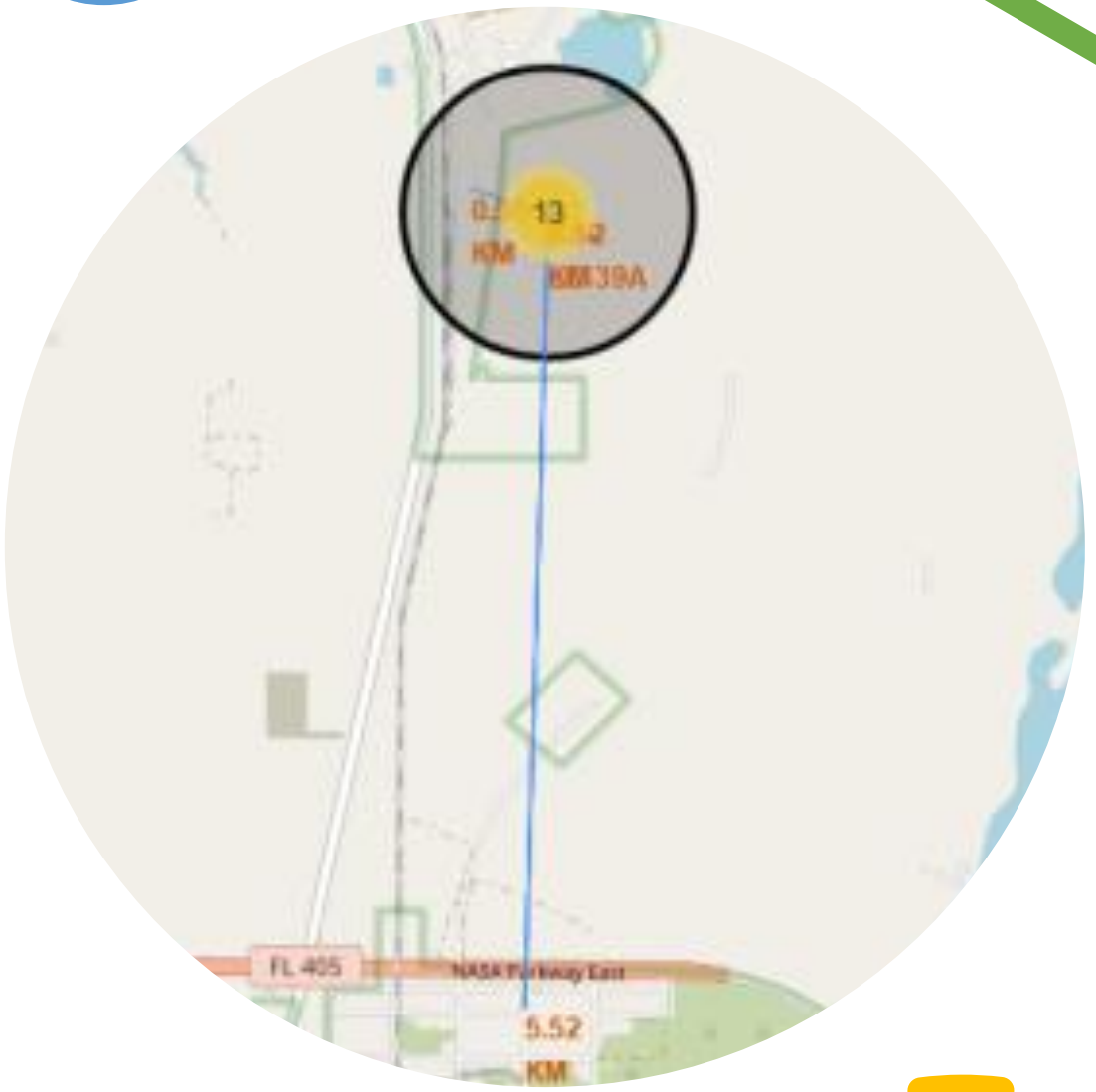
Section 3

Launch Sites Proximities Analysis



All launch sites

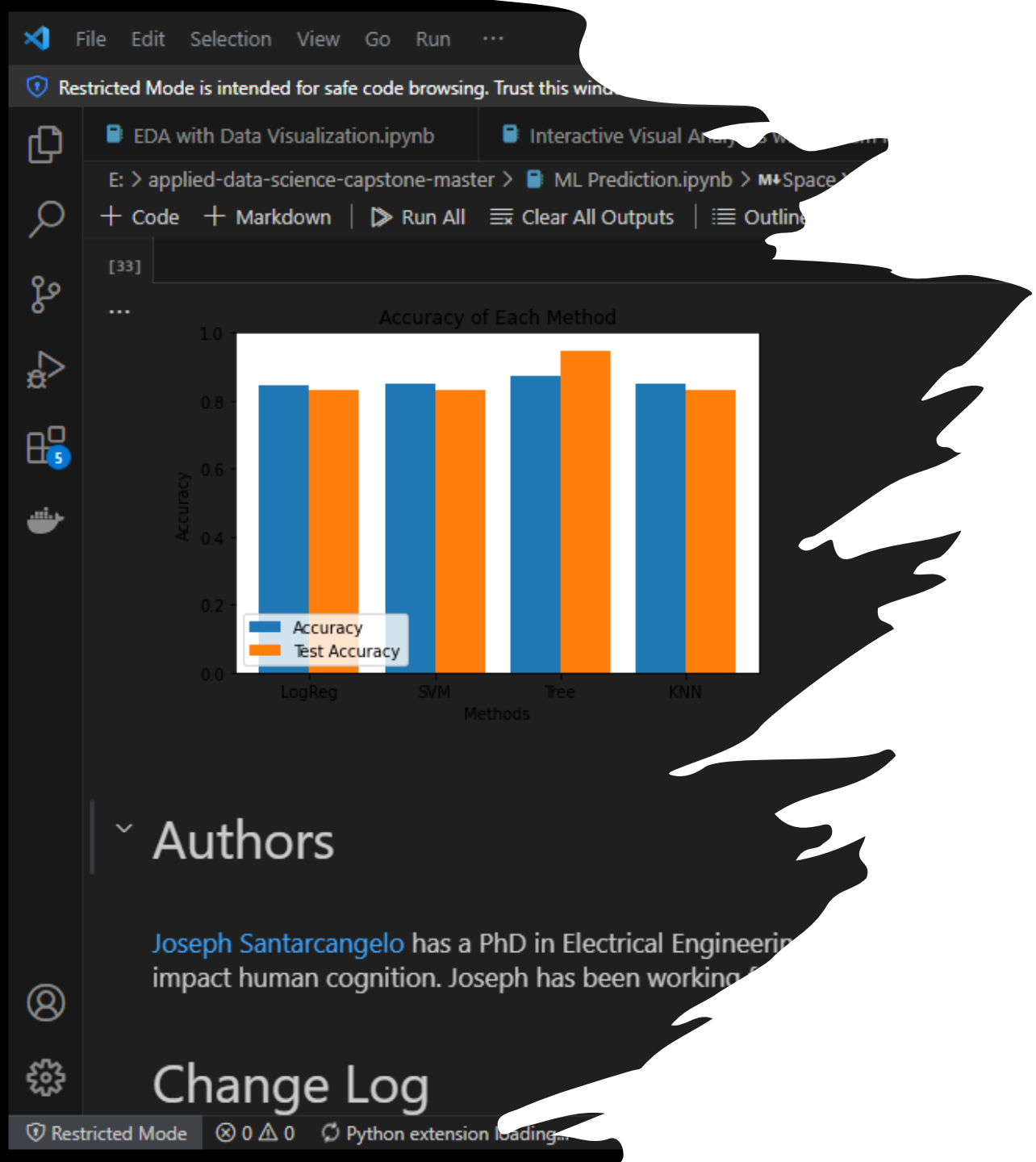
Launch Outcomes by Site



Logistics and Safety

Section 5

Predictive Analysis (Classification)

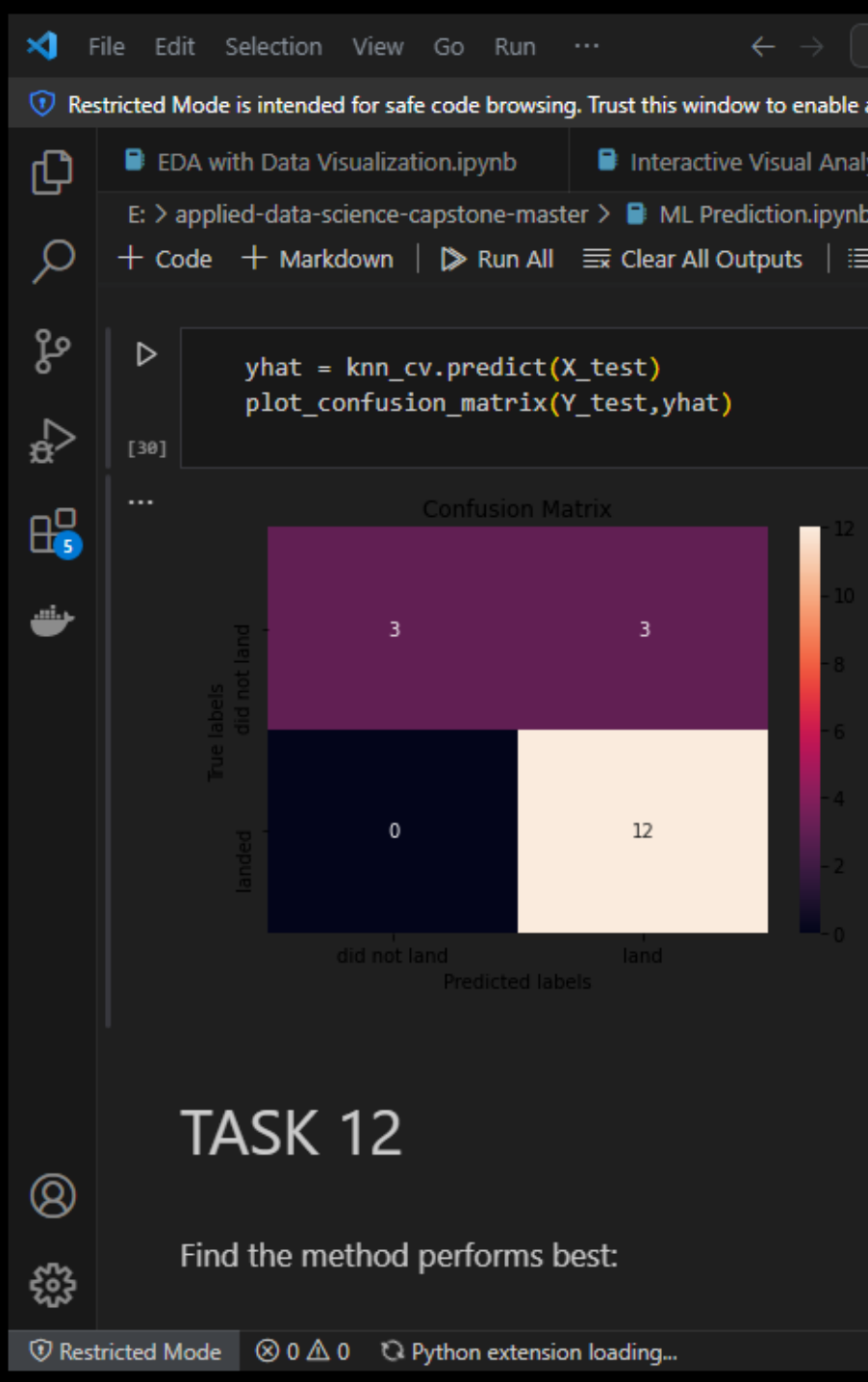


Classification Accuracy

- 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 big numbers of true positive and true negative compared to the false ones



Conclusions

- The best launch site is KSC LC-39A
- Launches KSC LC-39A 7,000kg are less risky;
- Decision Tree Classifier can be used to predict successful landings and increase profits

Thank you!