IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

- Data was gathered by SpaceX about launch success and other variables.
- Data was explored using maps and SQL to understand interest trends.
- Four predictive models were measured for accuracy in their ability tp predict launch success.
- Logistics regression, support vector machine, and k-nearest neighbor model were all equally accurate in their predictions.

Introduction

- Commercial space travel is an exciting emerging market.
- Low cost space travel can be possible when rockets can be reused.
- Thus, being able to predict rocket success allows for accurate pricing and market projections.

Section 1

Methodology

Methodology

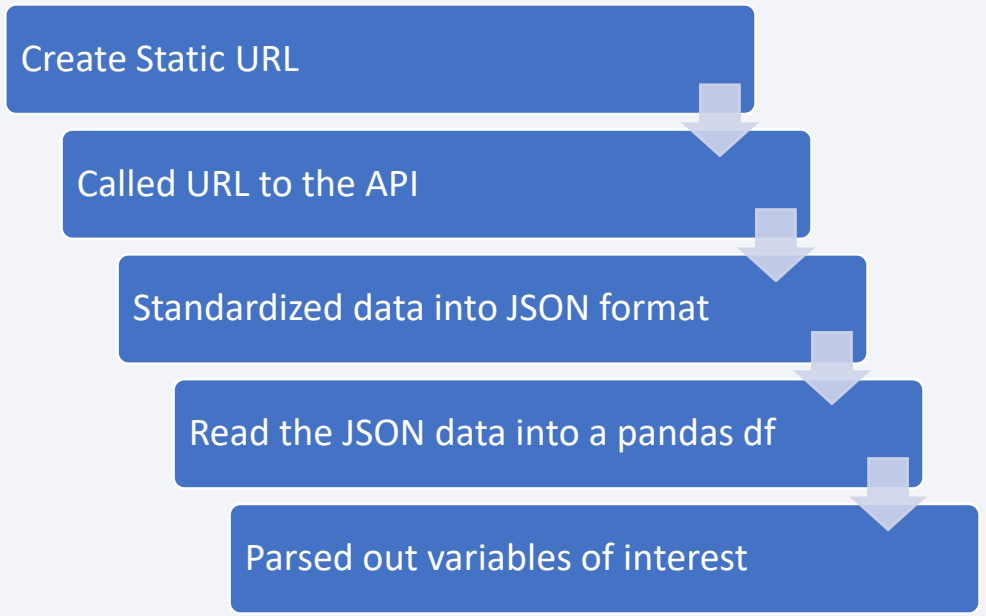
Executive Summary

- Data collection methodology:
  - API and web scraping were utilized
- Perform data wrangling
  - Replacing null values, dummy-coded all variables
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium
- Perform predictive analysis using classification models
  - Tested four predictive models and found three with identical accuracy

Data Collection – SpaceX API

- Used the SpaceX REST API
  - Called the URL to collect raw data
  - Normalized the data into a JSON
  - Read the JSON into a pandas data frame
  - Parsed the data to select only variables of interest

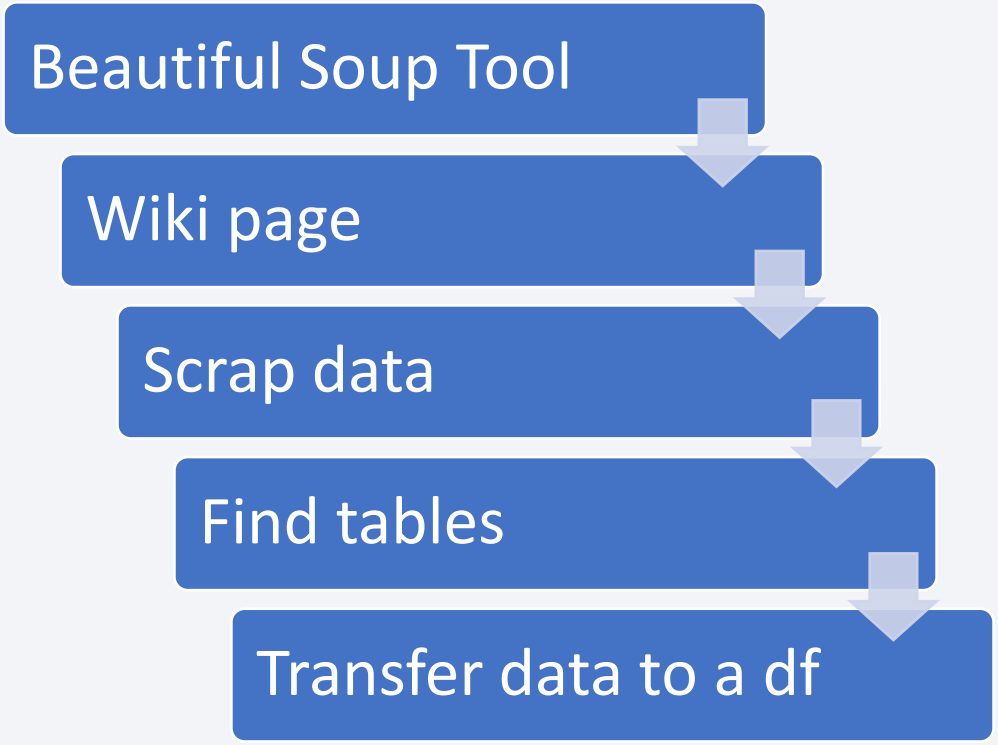
Review notebook on GitHub



Data Collection - Scraping

- Utilized Beautiful Soup
  - Scraped a Wiki Page with data
  - Found the tables of interest
  - Wrote that data into a pandas data frame

Review notebook on GitHub.



EDA with Data Visualization

- Charts Included:
  - Flight Number vs. Launch Site
  - Payload Mass vs. Launch Site
  - Success Rate vs. Orbit Type
  - Flight Number vs. Orbit Type
  - Payload mass vs. Orbit Type
  - Launch Success Yearly Trend

Review notebook on GitHub

EDA with SQL

- SQL queries were used to find:
  - Launch Site Names
    - Launch Site Names that begun with “CCA”
  - Total Payload Mass
  - Average Payload Mass of F9 v1.1
  - First Successful Ground Landing Date
  - Successful Drone Ship Landing with Payload between 4000 and 6000
  - Total Number of Successful and Failure Mission Outcomes
  - Boosters Carried Maximum Payload
  - 2015 Launch Records
  - Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

Review notebook on GitHub

Build an Interactive Map with Folium

- Folium Maps Include:
  - Maps of Launch Sites
  - Visualizing Launch Success and Failure at Site Locations
  - Finding the closest Launch site to the Florida Coast Line

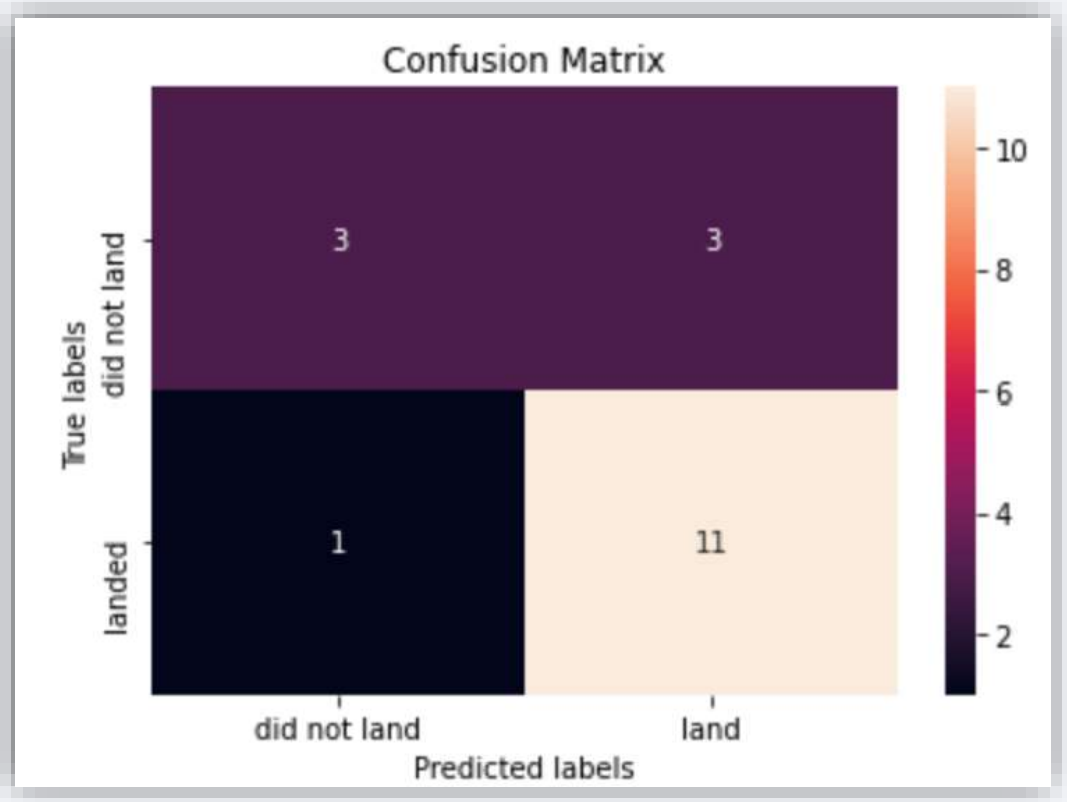
Review notebook on GitHub

Predictive Analysis (Classification)

- Data Analysis
  - Used GatherSearch() to find best parameters
  - Tested Four Models:
    - Logistic Regression
    - Support Vector Machine
    - Decision Tree
    - K-Nearest Neighbors

Results

- Key Insights Include:



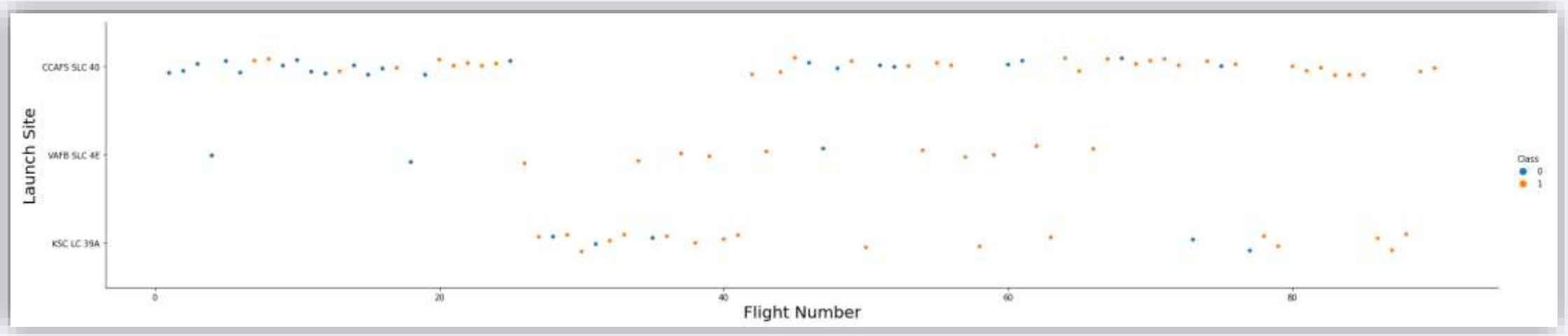
Model	Accuracy Score
Logistics Reg	0.833333
Support Vector Machine	0.833333
Decision Tree	0.888889
K-Nearest Neighbors	0.833333

Section 2

Insights drawn from EDA

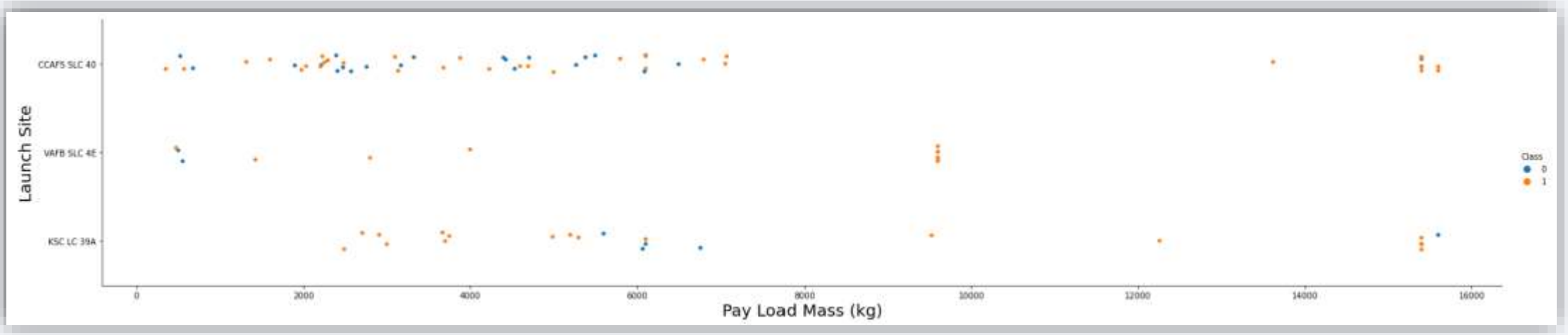
Flight Number vs. Launch Site

- Plot shows when each launch site had a successful or failed flight.



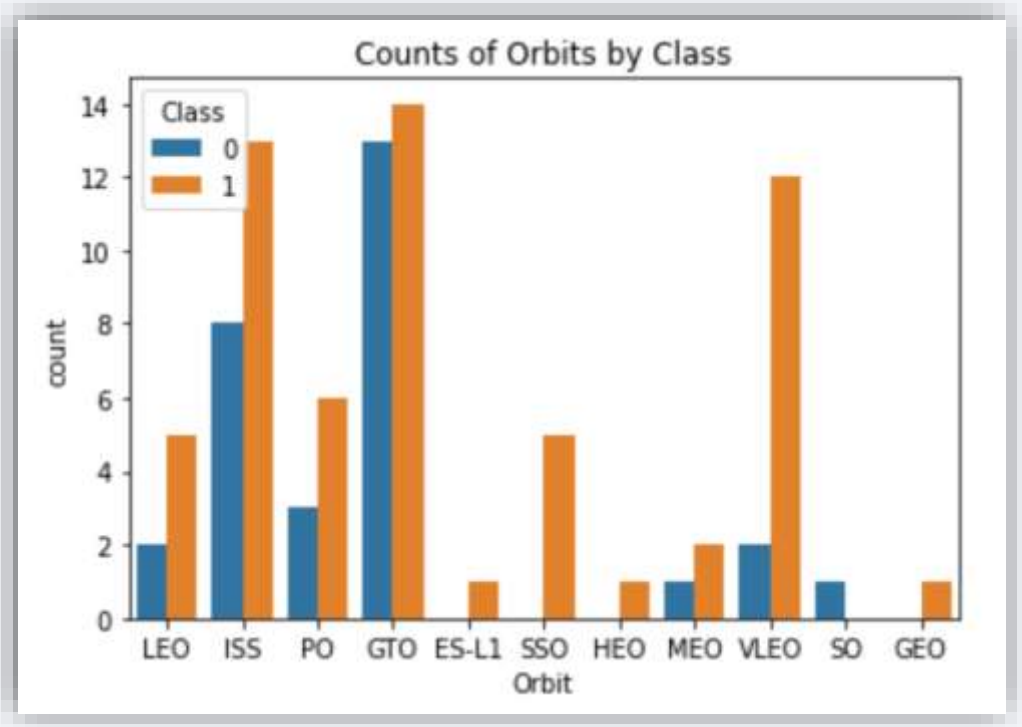
Payload vs. Launch Site

- Plot shows the payload mass of each successful and failed launch at each site



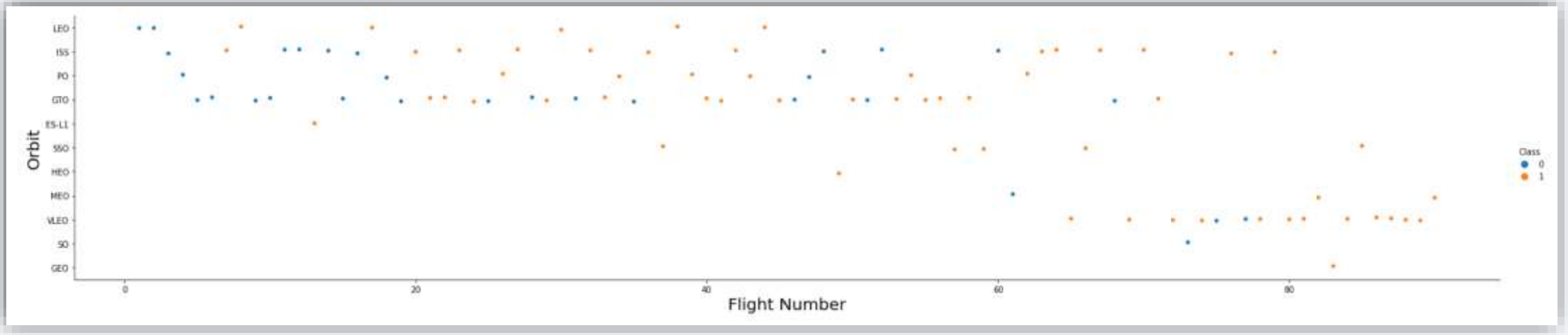
Success Rate vs. Orbit Type

- Plot shows the number of successful and failed flights for each orbit



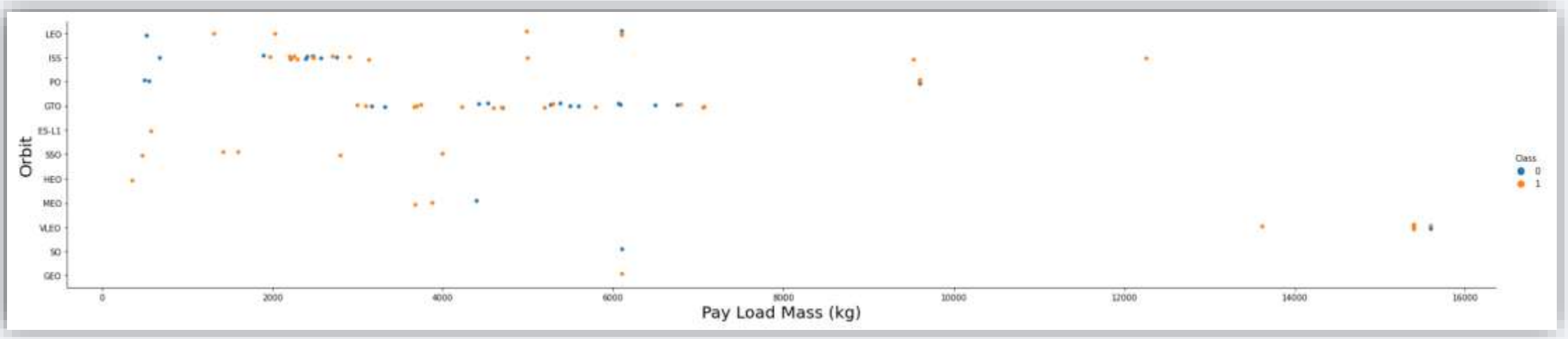
Flight Number vs. Orbit Type

- Plot shows when each orbit type had a successful or failed launch.



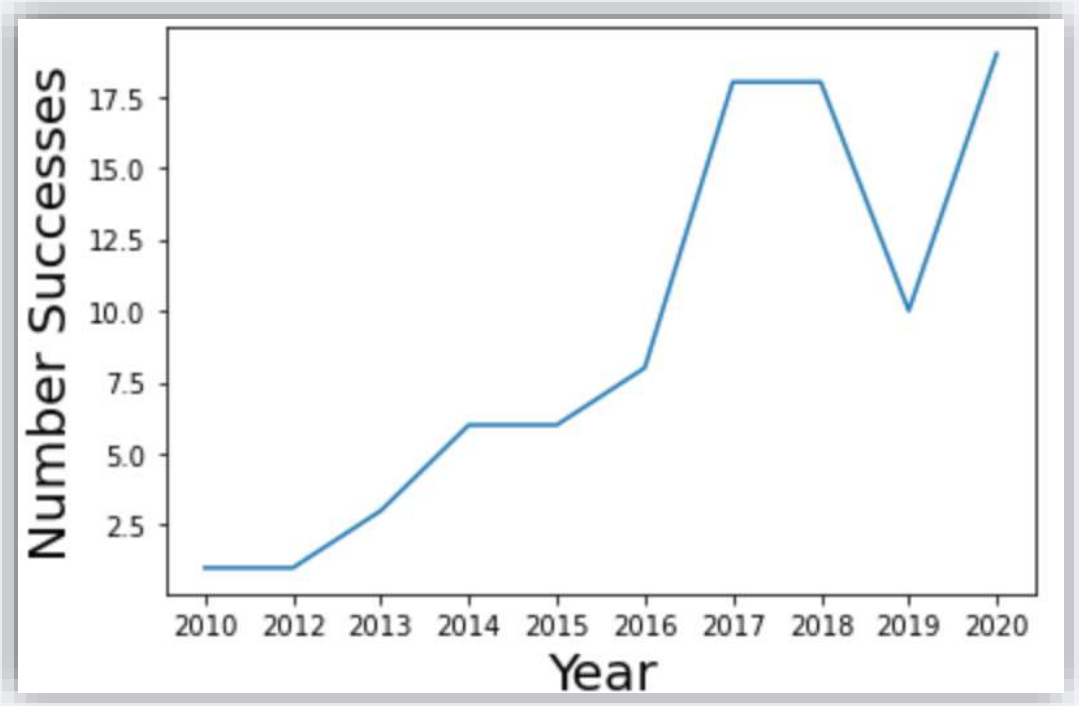
Payload vs. Orbit Type

- Plot shows successful and failed flights by payload for each orbit type



Launch Success Yearly Trend

- Plot shows the number of total successes each year.



All Launch Site Names

- Launch Site Names:

CCAFS LC-40

CCAFS SLC-40

KSC LC-39A

VAFB SLC-4E

Launch Site Names Begin with 'CCA'

- Launch Site Names Beginning with “CCA”:

CCAFS LC-40

CCAFS SLC-40

Total Payload Mass

- Total payload mass carried by boosters launched by NASA (CRS):

45,596

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1:

2,534

First Successful Ground Landing Date

- The date when the first successful landing outcome in ground pad was achieved:

December 22, 2015

Successful Drone Ship Landing with Payload between 4000 and 6000

- 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

- The total number of successful and failure mission outcomes:

Successful: 61

Failed: 10

Boosters Carried Maximum Payload

- Names of the booster which have carried the maximum payload mass:

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

- 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	Failure (drone ship)
F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

- The count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order:

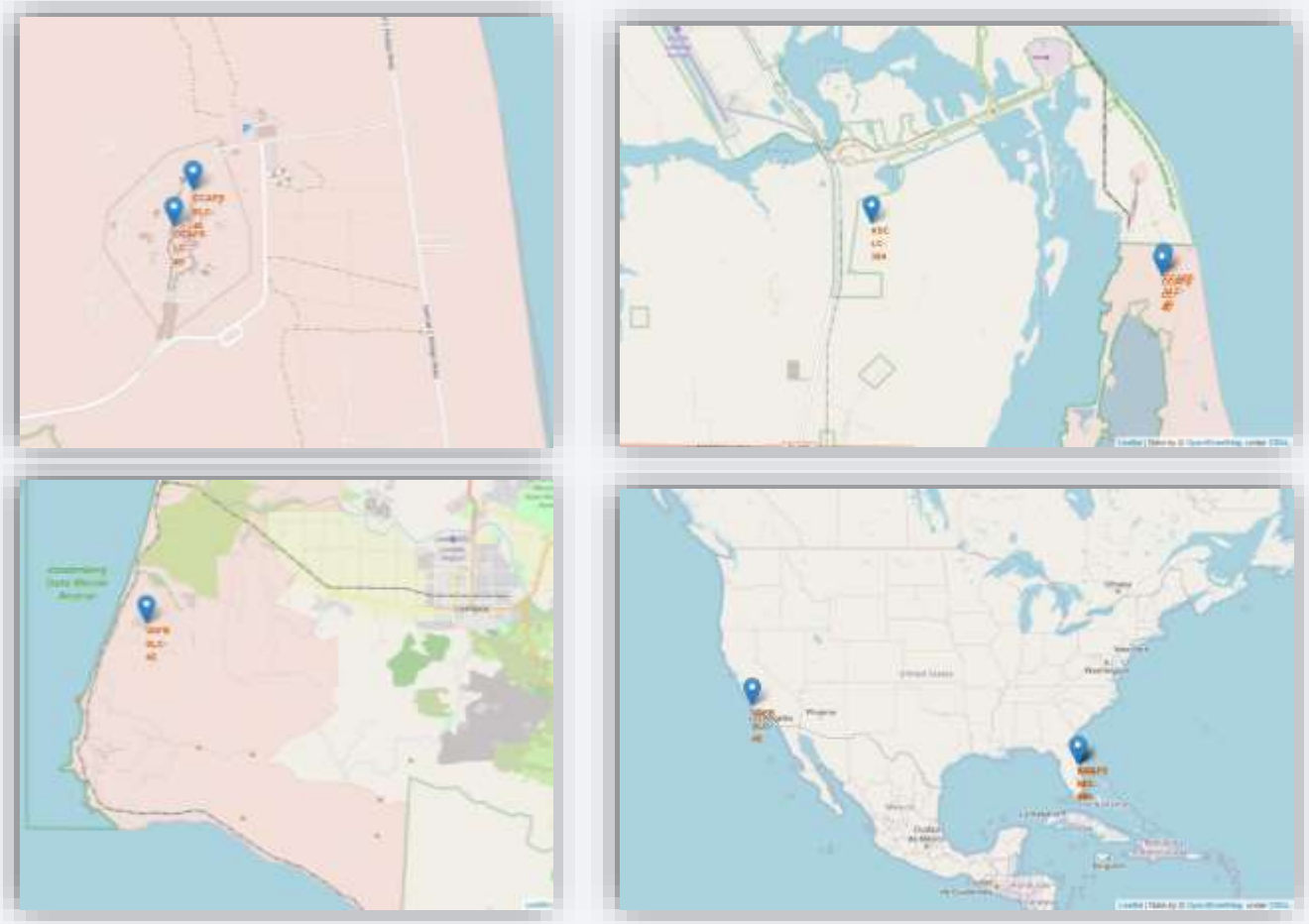
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 4

Launch Sites Proximities Analysis

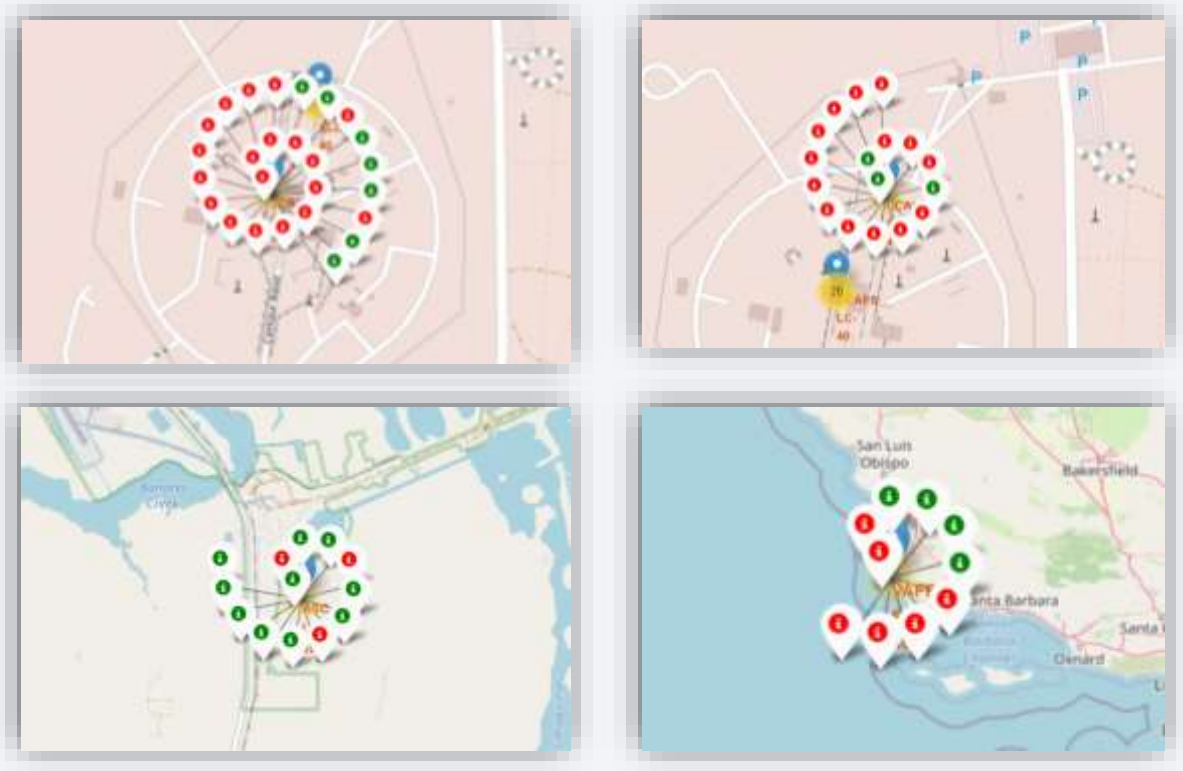
Launch Sites

- These maps display launch site locations in California and Florida.



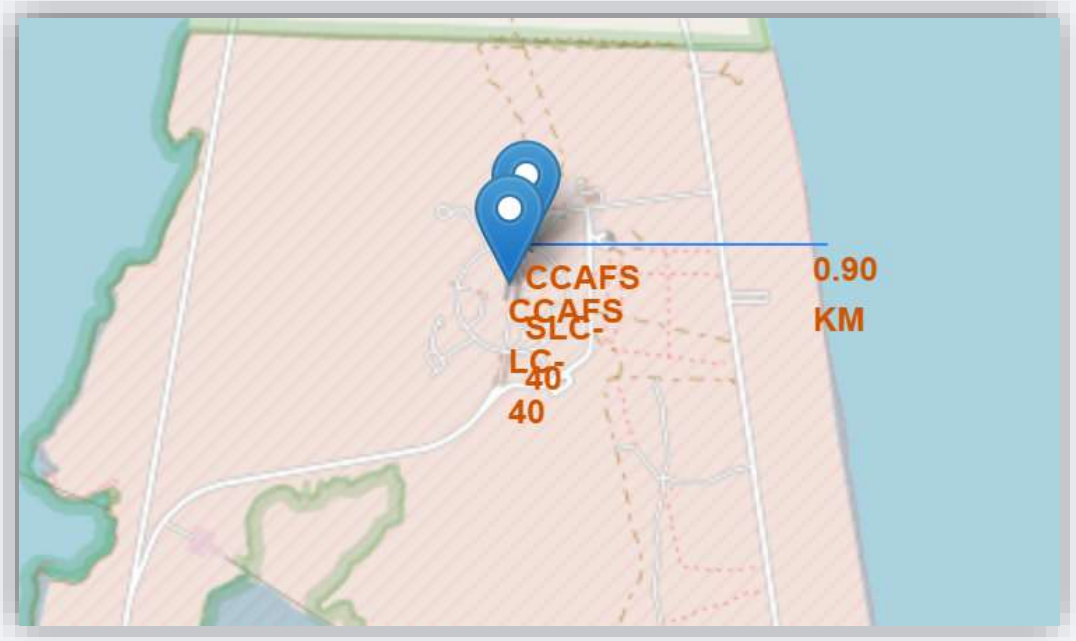
Landing Data at Launch Locations

- These maps show launch success (green) and failure (red) for each site location.



Florida Coast Line

- This map shows the distance (0.9 km) between the Florida coast line and the nearest launch site (CCAFS SLC-40).

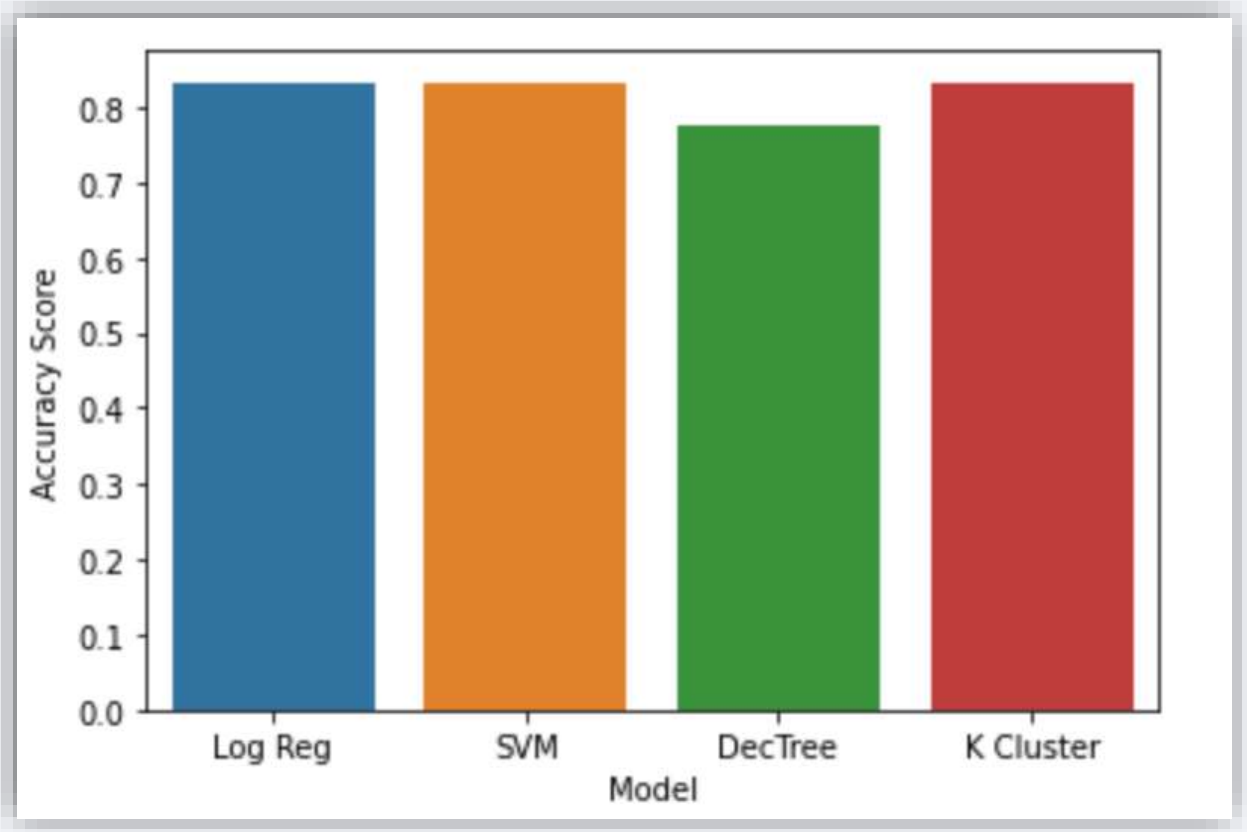


Section 6

Predictive Analysis (Classification)

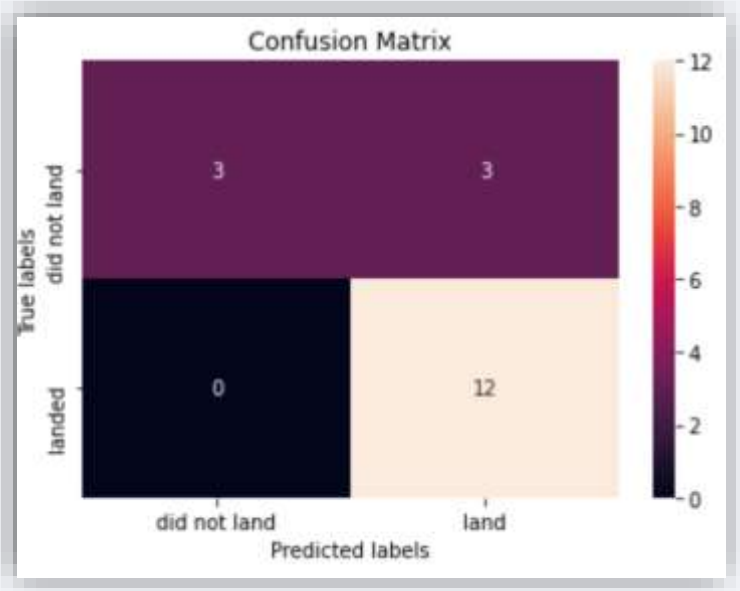
Classification Accuracy

- Each model was tested for accuracy:

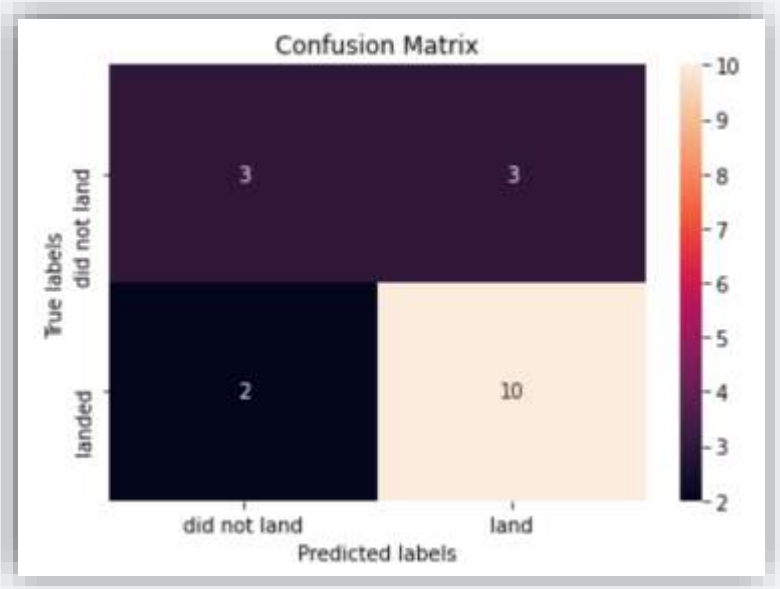


Confusion Matrix

- Two confusion matrices emerged:



Matrix for logistic regression, k-nearest neighbors, and SVC models



Matrix for decision tree model

Conclusions

- The decision tree model included more incorrect predictions.
- All other models predicted with the same accuracy.
- More data is needed to refine these models.
- The logistic regression model makes the most sense to use, as it id designed to generate a probability of an outcome event.

Thank you!