IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Rob Clowting
28-09-2022

Outline

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

Executive Summary

- Summary of methodologies
  - Collection of data was done through an API and web scraping
  - Exploratory Data Analysis was performed through data wrangling, data visualization and inveractive analytics
- Summary of all results
  - We identified several characteristics that appear to predict landing success: landing site and payload mass
  - Through Machine Learning we identified the decision tree model to be the best predictor for landing success

Introduction

- The objective of this project is to evaluate if new company SpaceY can compete with existing company SpaceX
- This presentation will answer the following questions:
    - Determine the price of the each launch by estimating if the first stage will be reused
    - What is the best place to make landings

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data from SpaceX is collected from two sources: SpaceX API and Webscraping
- Data wrangling
  - First, data is summarized and analyzed. Then, data is enriched with a binary landing outcome label
- Exploratory data analysis (EDA) using visualization and SQL
- Interactive visual analytics using Folium and Plotly Dash
- Predictive analysis using classification models
  - Data is normalized, divided into training and test data and evaluated based on four classification models. These classification models are then ranked based on accuracy

Data Collection

- Data collection methodology:
  - Data from SpaceX is collected from two sources:
    - SpaceX API - https://api.spacexdata.com/v4/rockets/
    - Webscraping - https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launces)

Data Collection – SpaceX API

- SpaceX has a public API that can be used to get data on their launches
- Data was collected according to the flowchart

https://github.com/clowtmans/Coursera_IBM_Capstone/blob/42b2fdfe2432a0c5f3ab30eed4d26bf934bf9497/Data%20Collection%20with%20API.ipynb

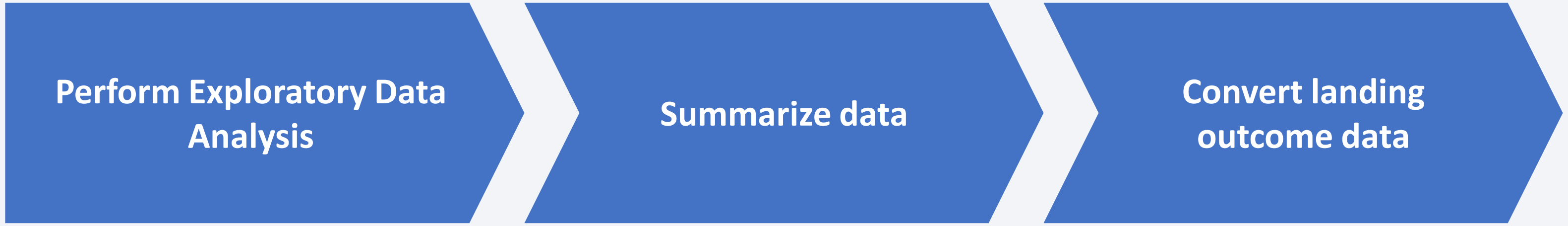
Data Collection - Scraping

- A second way to collect the data is by scraping the Wikipedia page for Falcon launces
- Data is collected according to the flowchart

https://github.com/clowtmans/Coursera_IBM_Capstone/blob/78b84bf5c8549f878588aa12da93acd716ff90e9/Data%20Collection%20with%20Web%20Scraping.ipynb

Data Wrangling

- First, Exploratory Data analysis was performed to find patterns in the data
- Then, data was summarized and number and occurrence of mission outcomes are generated
- Finally, booster landings were converted into a binary outcome label where 1 means a successful landing and 0 means unsuccessful



https://github.com/clowtmans/Coursera_IBM_Capstone/blob/f1b5ff3aa01783b1f8fd1b16c189062db1496414/Data%20Wrangling.ipynb

EDA with Data Visualization

- Summary of plotted charts to gather first insights on impact of variables on success rate:
  - Scatterplot of Flight Number vs. Payload Mass
  - Scatterplot of Flight Number vs. Launch Site
  - Scatterplot of Payload and Launch Site
  - Bar chart of Success Rate and Orbit
  - Scatterplot of Flight Number and Orbit
  - Scatterplot of Payload Mass and Orbit
  - Line chart of Success Rate over years

https://github.com/clowtmans/Coursera_IBM_Capstone/blob/ce9c5292d1a62d2df1abd66a940a94b3819286ec/EDA%20with%20data%20viz.ipynb

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
  - Select distinct launch sites in the space mission
  - Display 5 records where launch site begins with CCA
  - Display total payload mass carried by boosters launched by NASA
  - Display average mayload mass carried by booster version F9 v1.1
  - Determine date when first successful ground pad landing was achieved
  - Show booster names with mass between 4000 and 6000 kg that successfully landed on drone ship
  - Calculate total number of successful and failed mission outcomes
  - Determine booster versions that have carried the maximum payload mass
  - Rank count of successful landings between 05-06-2010 and 20-03-207

https://github.com/clowtmans/Coursera_IBM_Capstone/blob/802f547867581b7e53176d222220e4ba60888b57/EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

- A Folium map was created with markers, circles, lines and marker clusters.
  - Markers are points of interest on the map
  - Circles highlight areas around a certain point of interest (coordinates)
  - Lines are used to indicate the distance between points
  - Marker clusters are used to present a group of events around a certain coordinate

https://github.com/clowtmans/Coursera_IBM_Capstone/blob/42551da8291f85c3e3b1a638598b1880eb52a981/Data%20Viz%20with%20Folium.ipynb

Predictive Analysis (Classification)

- Four classification models were built and compared:
  - Logistic regression
  - Support vector machine
  - Decision tree
  - K-nearest neighbors

Prepare and Standardize Data → Build and test the four classification models → Compare results and determine best model

https://github.com/clowtmans/Coursera_IBM_Capstone/blob/d2448515f2740c98a3094bc5a3307ddacb1bb486/Predictive%20Analysis%20with%20Machine%20Learning.ipynb

Results

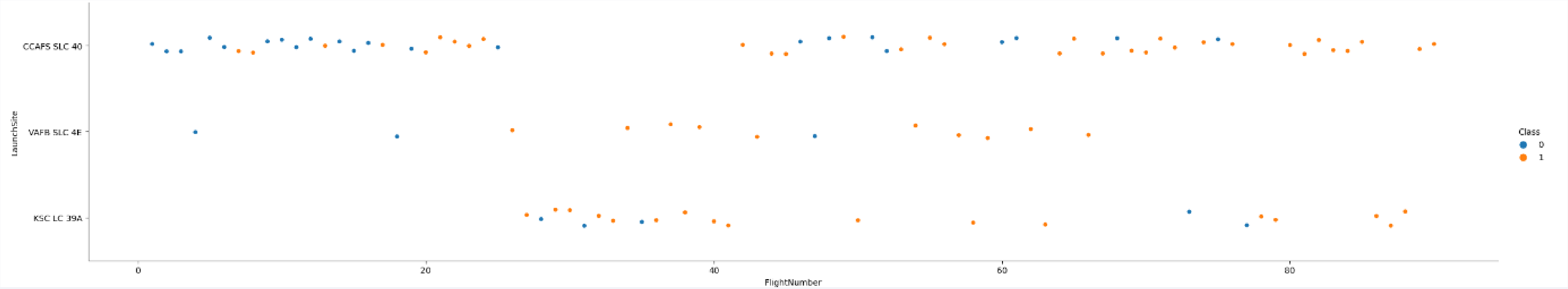
- Exploratory data analysis results
  - Higher payloads lead to higher likelihood of successful landing
  - Most flights were launched from CCAFS SLC-40
  - Orbit type is not a good predictor of landing success
  - Landing success increased over time from start in 2010 until today
- Predictive analysis results
  - Decision tree is the best model to predict successful landing with an accuracy of 89%

Section 2

Insights drawn from EDA

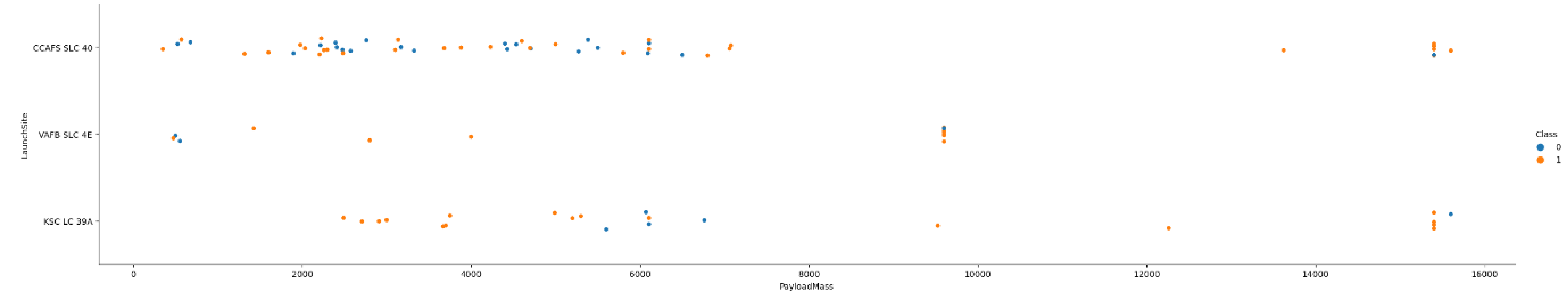
Flight Number vs. Launch Site

- Landing success increases with flight number
- Disregarding the first 25 flight numbers, there seems to be no relationship between launch site and landing success



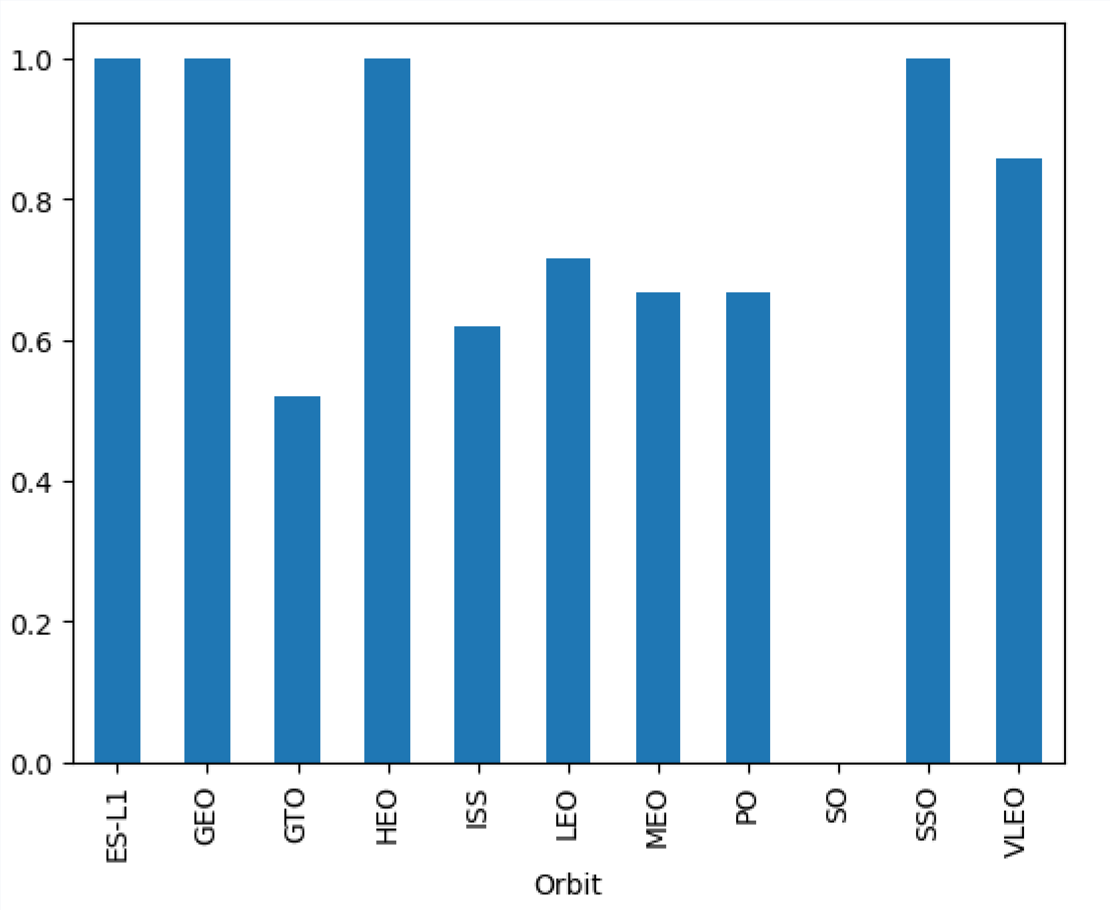
Payload vs. Launch Site

- Heavier rockets have higher success rates
- No rockets heavier than 10.000kg were launced from VAFB SLC 4E



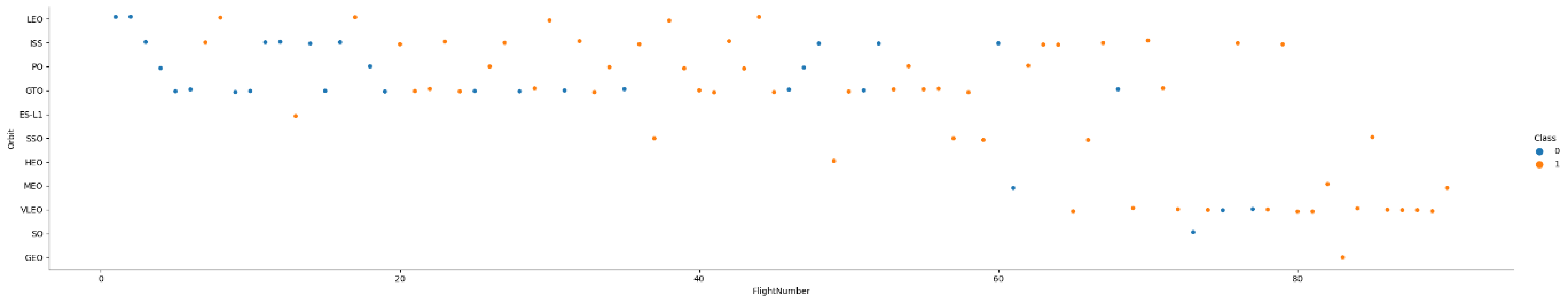
Success Rate vs. Orbit Type

- There are four Orbit types with a 100% success rate:
  - ES-L1
  - GEO
  - HEO
  - SSO
- Orbit type SO has a 0% success rate



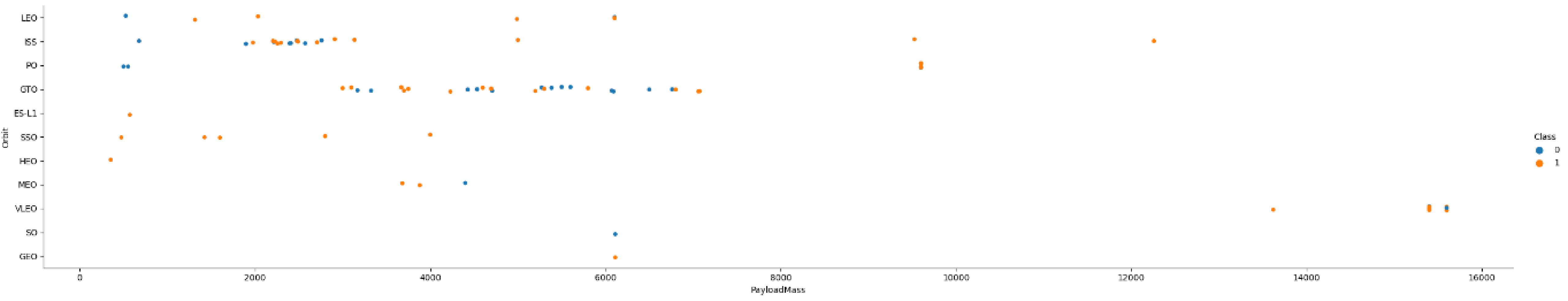
Flight Number vs. Orbit Type

- As flight number increases, orbit types LEO, PO and GTO become less common and SSO and VLEO increase



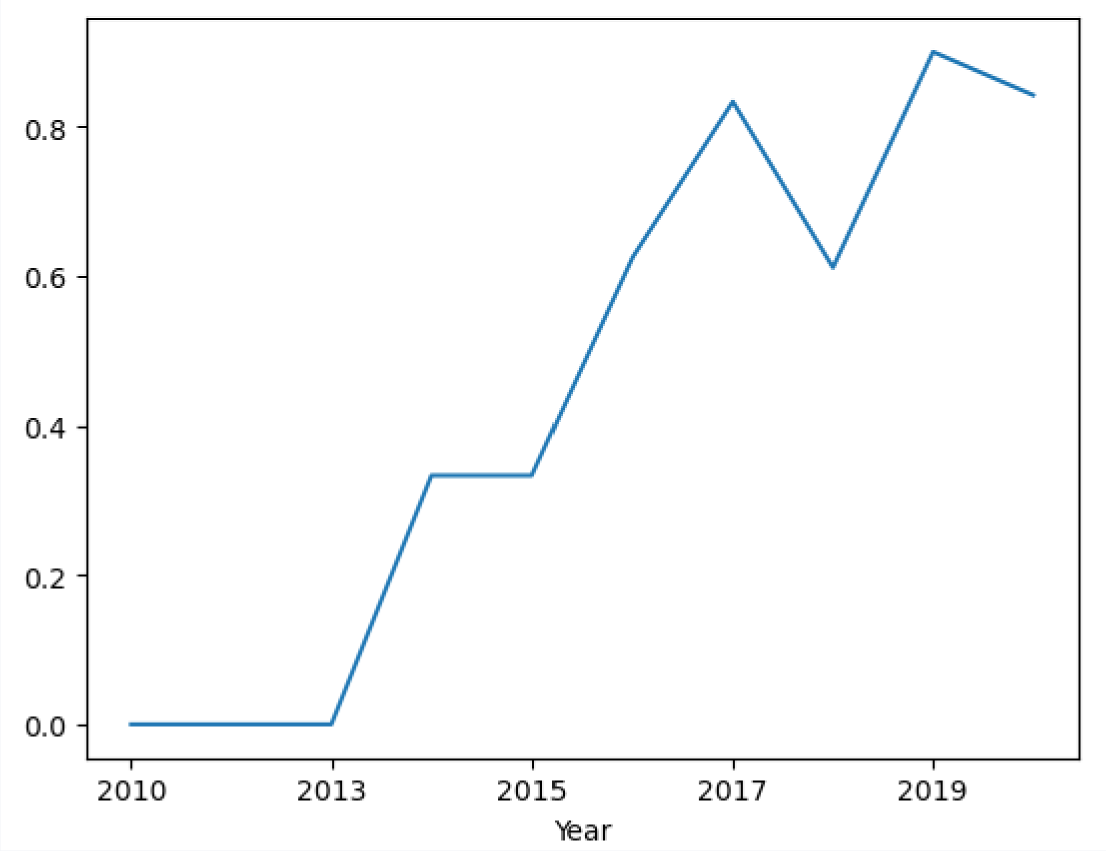
Payload vs. Orbit Type

- Orbit types ISS and GTO have lower payload mass
- High payloadmass has a higher success rate regardless of orbit type
- Orbit type VLEO has exclusively high payload rates



Launch Success Yearly Trend

- Average succes rate of the missions increased since the beginning in 2010 and today



All Launch Site Names

- Unique launch sites, found by selecting the disctinct launch sites

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

Launch Site Names Begin with 'CCA'

- 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

- By summing the total payload mass of all boosters launched by NASA we find the total mass accounts to 111.268kg

Average Payload Mass by F9 v1.1

- By selecting the average payload mass of all missions with the booster version F9 v1.1 we find the result to be 2.928kg

Total Number of Successful and Failure Mission Outcomes

- Only 1 out of 101 missions failed, the rest was successful
- Query results in table below

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

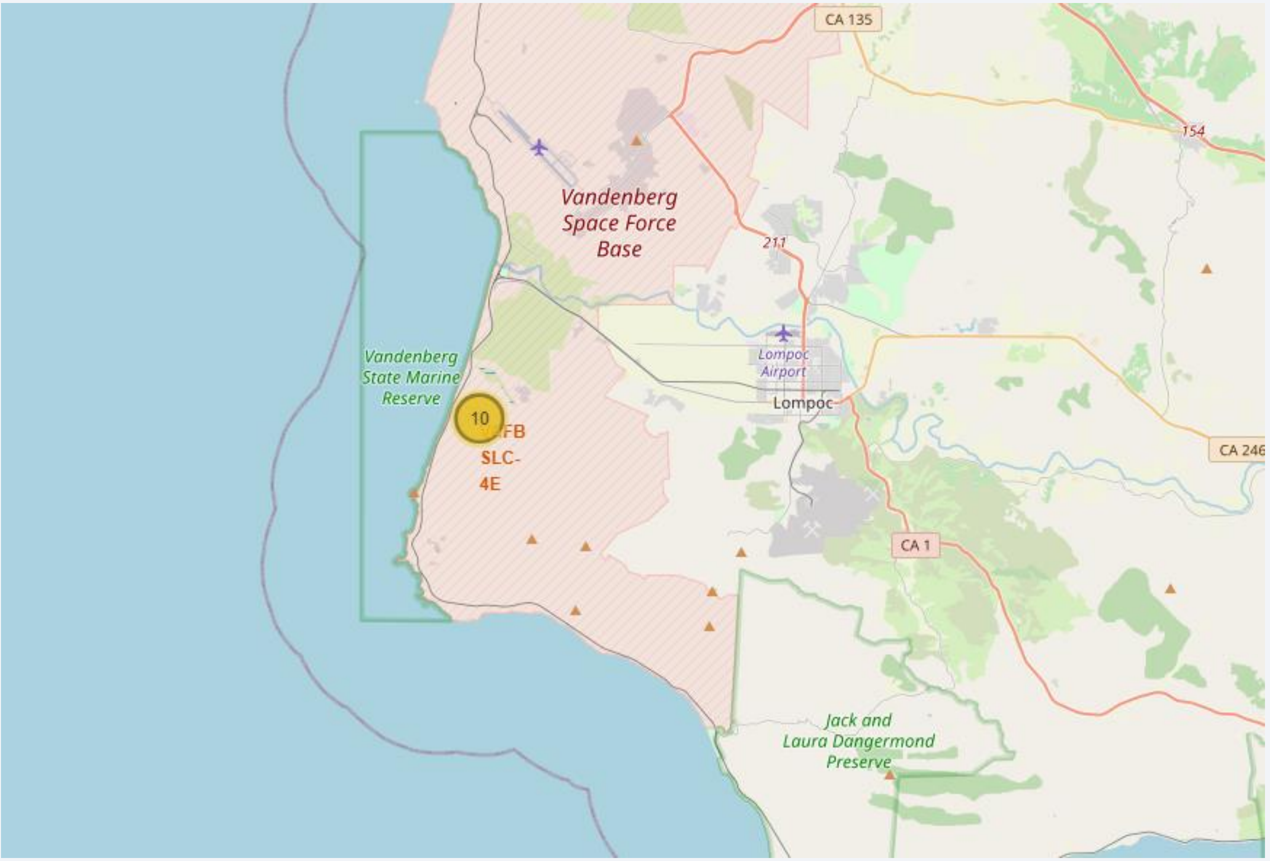
Section 3

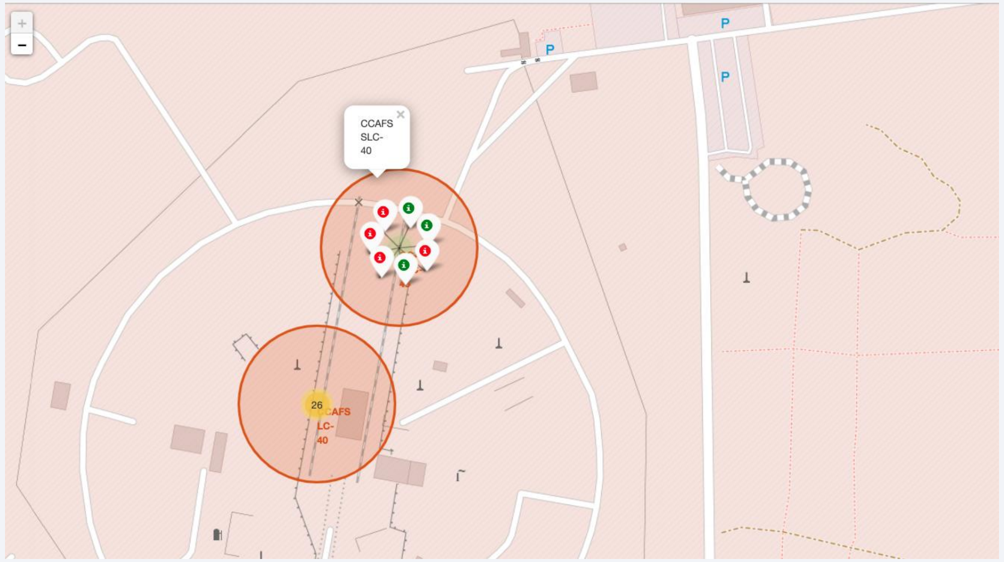
Launch Sites Proximities Analysis

SpaceX Launch Sites Map

- Launch sites are in California and Florida and located near the sea

Launch Sites and Success Rates





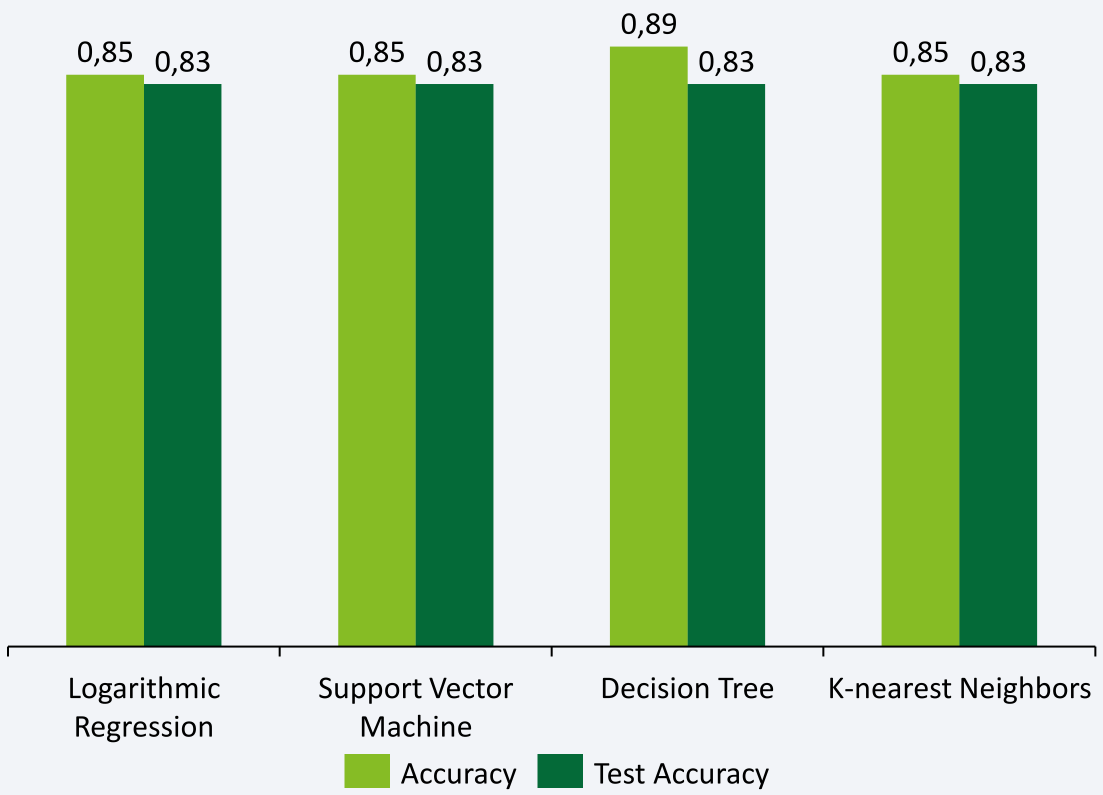
Distance to Sea

Section 5

Predictive Analysis (Classification)

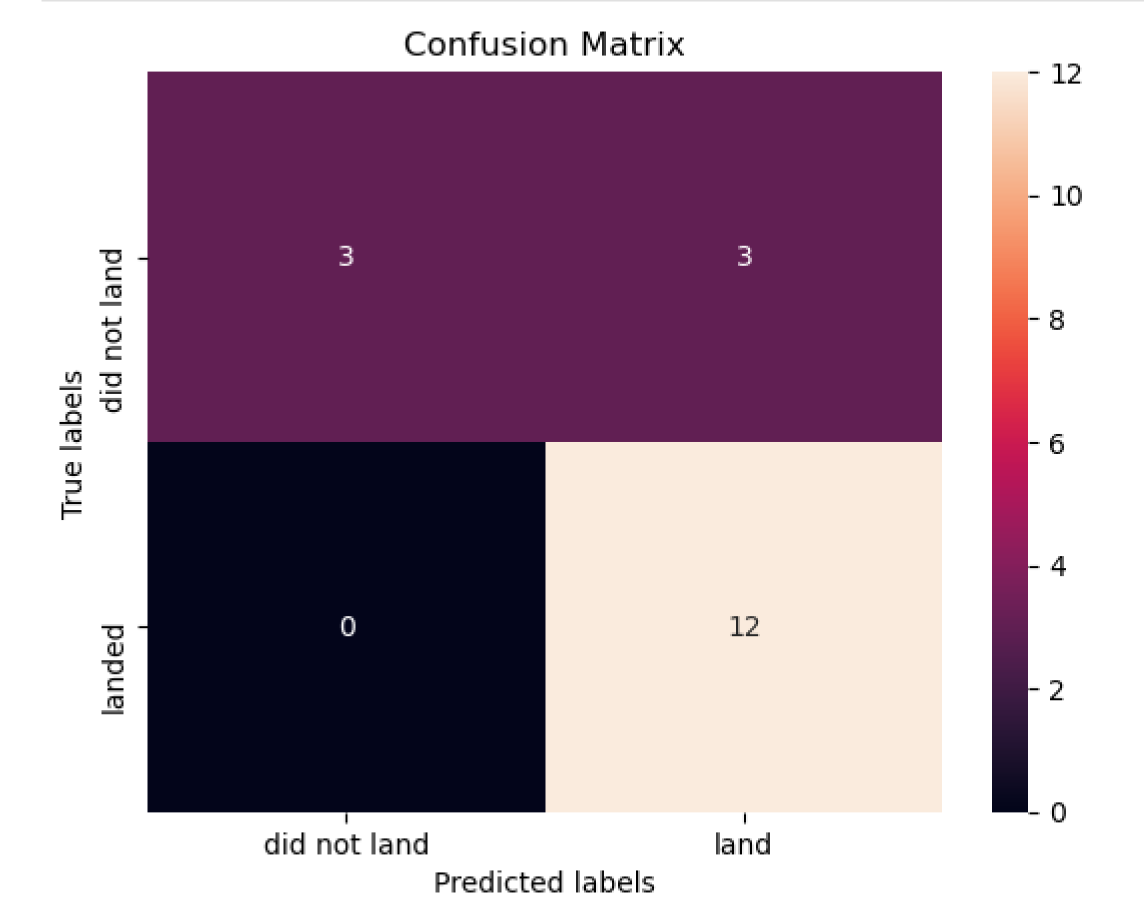
Classification Accuracy

- Decision Tree has the highest accuracy with 89%
- The other models score similarly



Confusion Matrix

- The confusion matrix shows the amount of true positives is good, the amount of true negatives can be improved



Conclusions

- Through the decision tree Machine Learning Models, we can predict the landing outcome with a 89% accuracy

Thank you!