IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Thomas
9.9.2021

Outline

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

Executive Summary

Summary of methodologies

- We will predict if the Falcon 9 first stage will land successfully in order to determine the cost of the launch.
- We collect data by SpaceX API and scraping from Wikipedia
- Then, we Exploratory data by SQL and Visualization
- We use some machine algorithm to predict the result such as SVM, Classification Trees and Logistic Regression and K-Nearest Neighbors

Summary of all results

Algorithm	Accuracy
Support Vector Machine	0.83
Claasifiation Tree	0.67
Logistic Regression	0.83
K-Nearest Neighbors	0.83

Introduction

SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.

Therefore if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Request Data from SpaceX API
  - Scraping Data from Wikipedia
- Perform data wrangling
  - Fill mean or drop missing value from data
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Use some algorithm such as: Logistic Regression, SVM, Tree Classification, KNN

Data Collection

- Collect Data by SpaceX API
- Collect Data by Scraping website Wikipedia

Data Collection – SpaceX API

- Request and parse the SpaceX launch data using the GET request to API
- Filter the dataframe to only include Falcon 9 launches
- Dealing with Missing Values: replace missing value with mean value

Data Collection - Scraping

- Request the Falcon9 Launch Wiki page from its URL
- Extract all column/variable names from the HTML table header
- Parse the table and convert it into a Pandas data frame

Data Wrangling

- Calculate the number of launches on each site
- Calculate the number and occurrence of each orbit
- Calculate the number and occurence of mission outcome per orbit type
- Create a landing outcome label from Outcome column

EDA with Data Visualization

- Visualize the relationship between Flight Number and Launch Site by scatter chart
- Visualize the relationship between Payload and Launch Site by scatter chart
- Visualize the relationship between success rate of each orbit type by bar chart
- Visualize the relationship between FlightNumber and Orbit type by scatter chart
- Visualize the relationship between Payload and Orbit type by scatter chart

- Visualize the launch success yearly trend y line chart

EDA with SQL

- 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 successful 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 failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- 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

Build an Interactive Map with Folium

- Mark all launch sites on a map
- Mark the success/failed launches for each site on the map
- Calculate the distances between a launch site to its proximities

Build a Dashboard with Plotly Dash

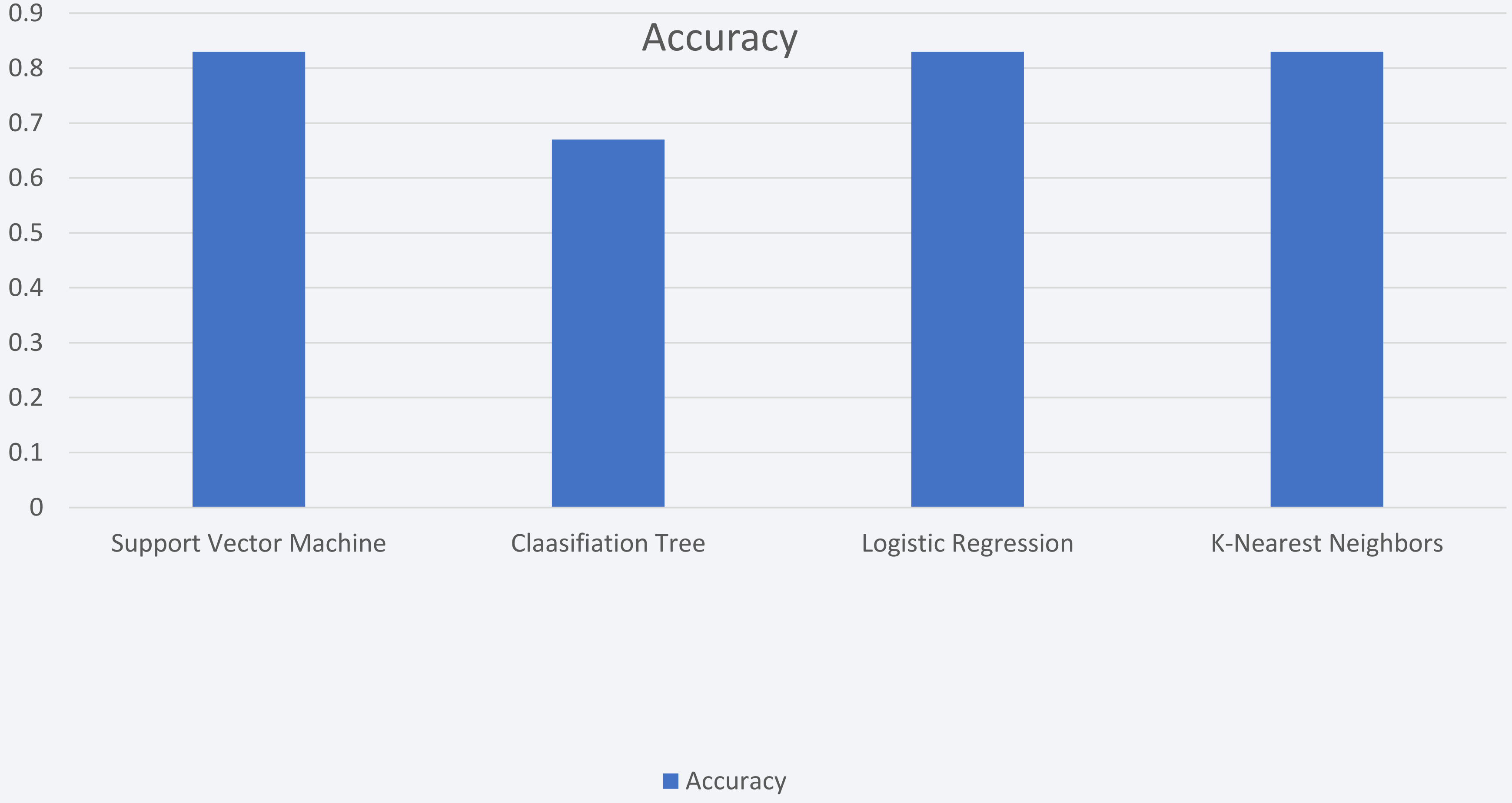
An interactive Dashboard with:

- success pie chart
- success-payload-scatter-chart

Predictive Analysis (Classification)

Algorithm	Accuracy
Support Vector Machine	0.83
Claasifiation Tree	0.67
Logistic Regression	0.83
K-Nearest Neighbors	0.83

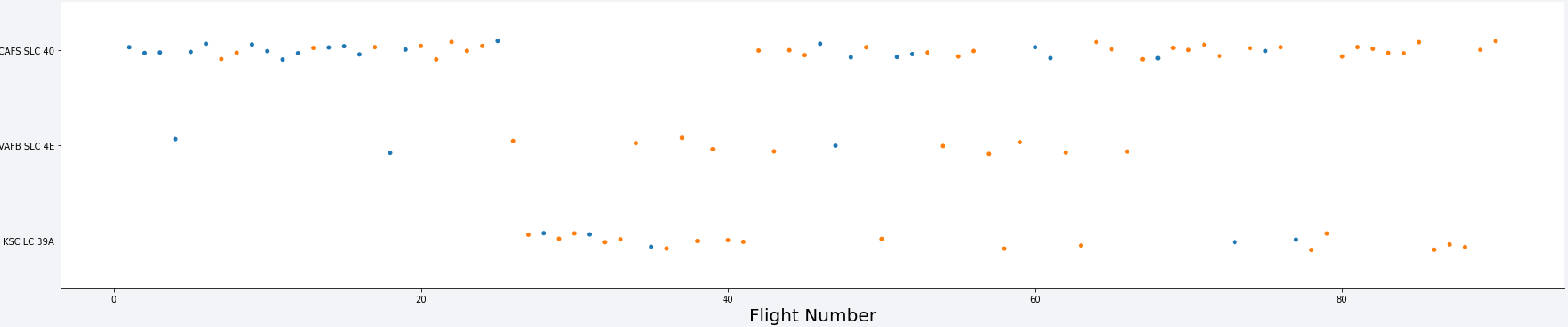
Results


Accuracy
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Support Vector Machine
Claasifiation Tree
Logistic Regression
K-Nearest Neighbors
Accuracy

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site



- For first FlightNumber from 1-25: spaceX launch at CCAFS SLC 40
- From 25-40: launch at VAFB SLC 4E and KSC LC 39A
- After 40: they launch at 3 launch sites

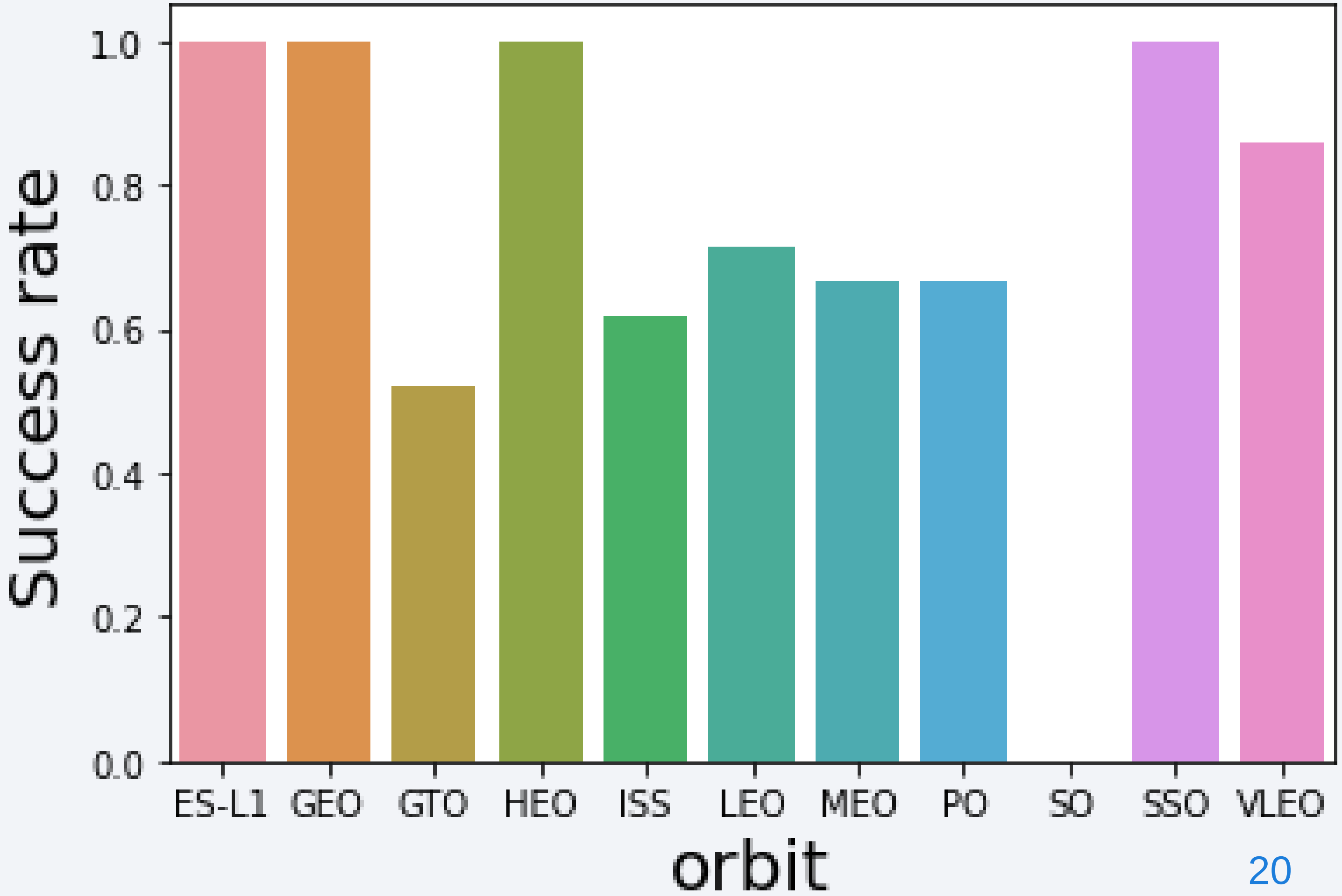
Payload vs. Launch Site

- With Pay load mass smaller than 7000kg: launch at 3 launch sites
- With Pay load mass around 10000kg: launch at VAFB SLC 4E and KSC LC 39A
- With Pay load mass around 15000kg: launch at KSC LC 39A and CCAFS SLC 40

Success Rate vs. Orbit Type

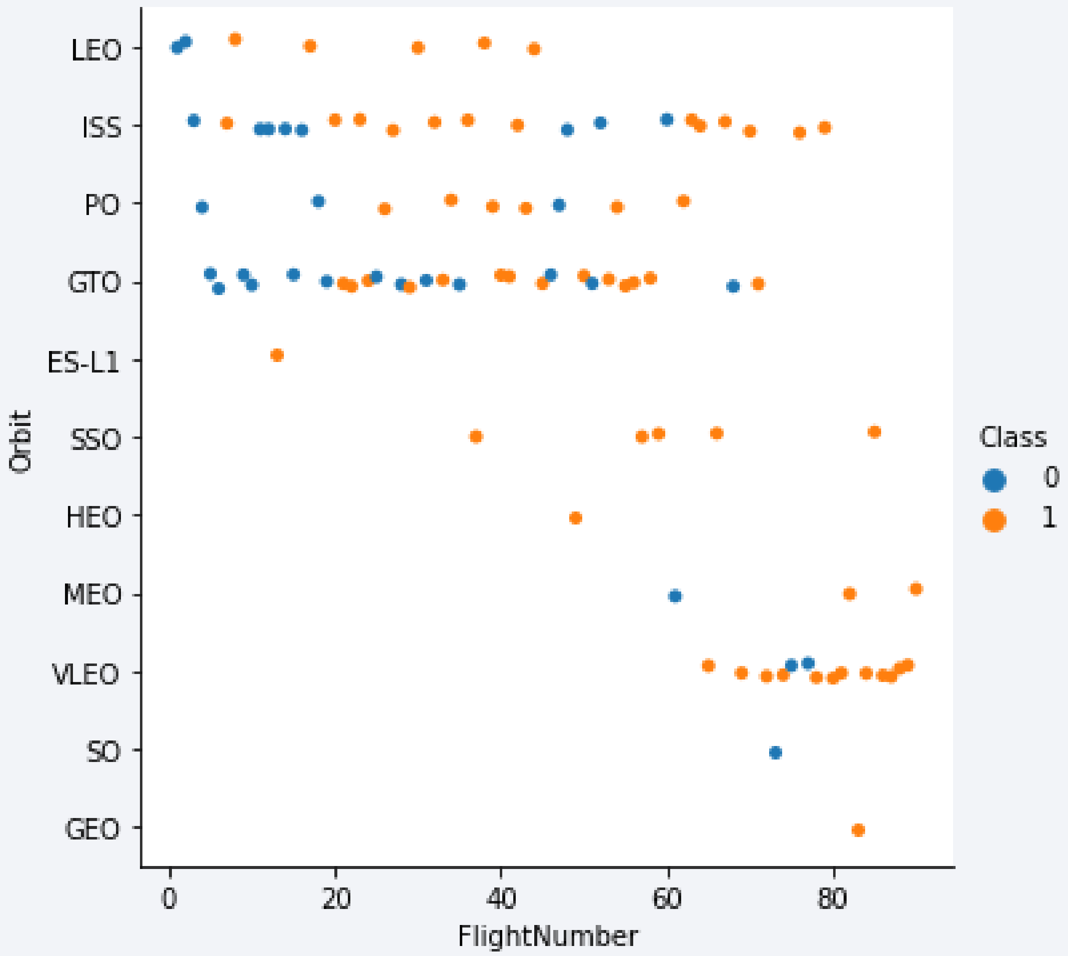
- Hightest success rate at orbit: ES-L1, GEO, SSO and HEO

- Lowest success rate at orbit: GTO (~0.5%)



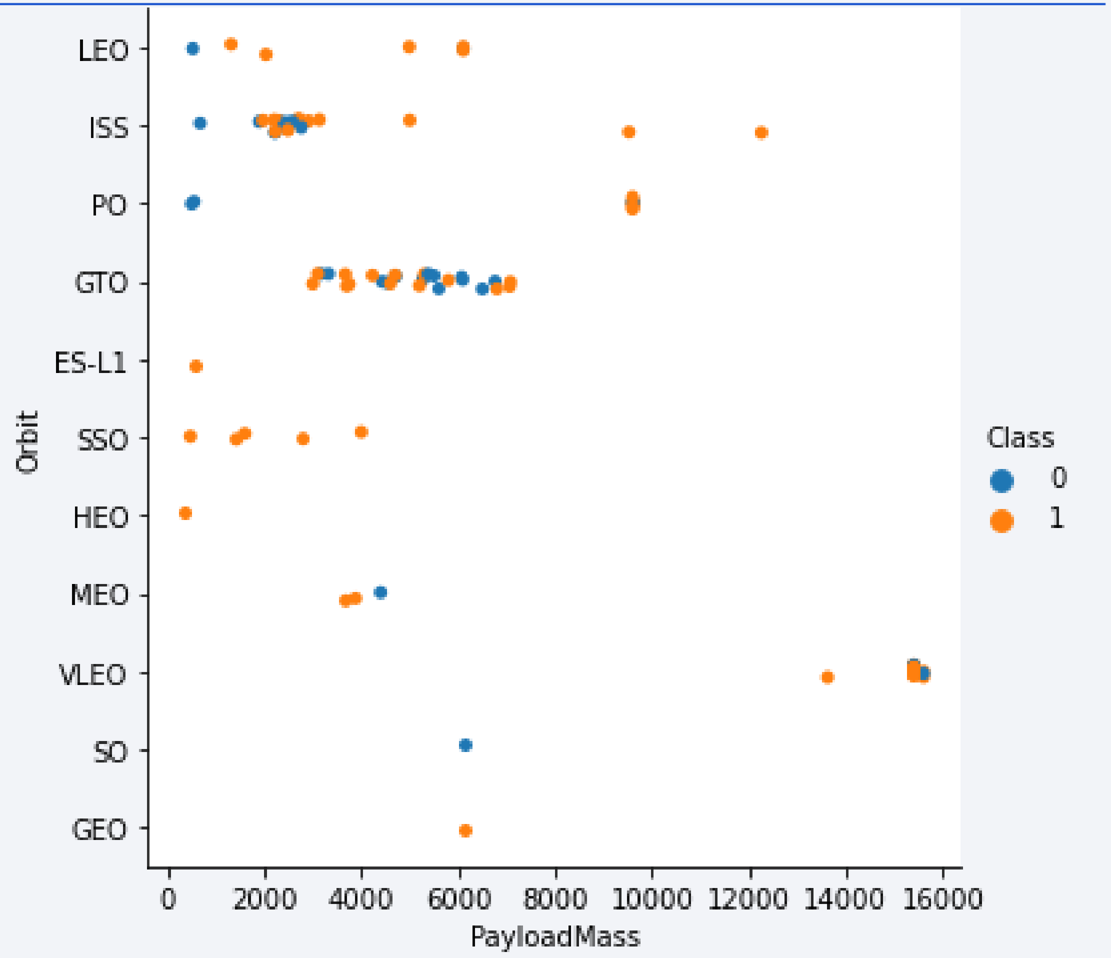
Flight Number vs. Orbit Type

- Flight number from 1-60: launch mostly at LEO,ISS,PO and GTO
- Flight number from 60: launch mostly at SSO, MEO, VLEO



Payload vs. Orbit Type

- Launch mostly with pay load mass below 10,000 in orbits such as LEO, ISS, and GTO
- Launch some Falcon9 with pay load mass more than 14,000kg at the orbit VLEO



Launch Success Yearly Trend

- Show a line chart of yearly average success rate

- Show the screenshot of the scatter plot with explanations

All Launch Site Names

- There are 4 unique launch site

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’ have the failure or No attempt landing outcomes

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

Total Payload Mass

- Calculate the total payload carried by boosters from NASA(CRS)
- 45596 KG

Average Payload Mass by F9 v1.1

the average payload mass carried by booster version F9 v1.1 IS 2534KG

First Successful Ground Landing Date

- The dates of the first successful landing outcome on ground pad:

2015-12-22

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

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

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
- Mostly success at mission outcome 99%

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

Boosters Carried Maximum Payload

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

booster_version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

2015 Launch Records

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

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

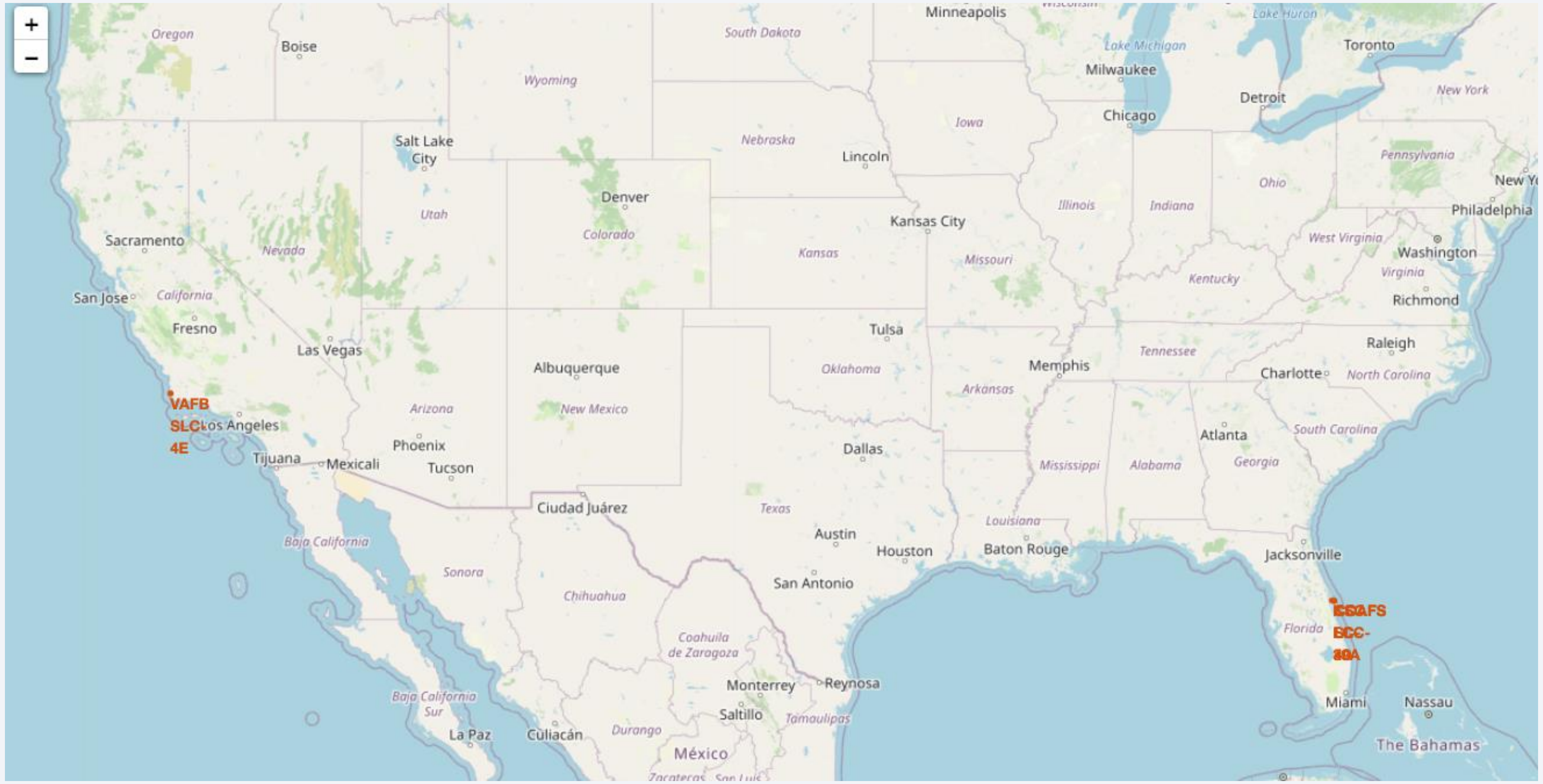
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

landing__outcome	number
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

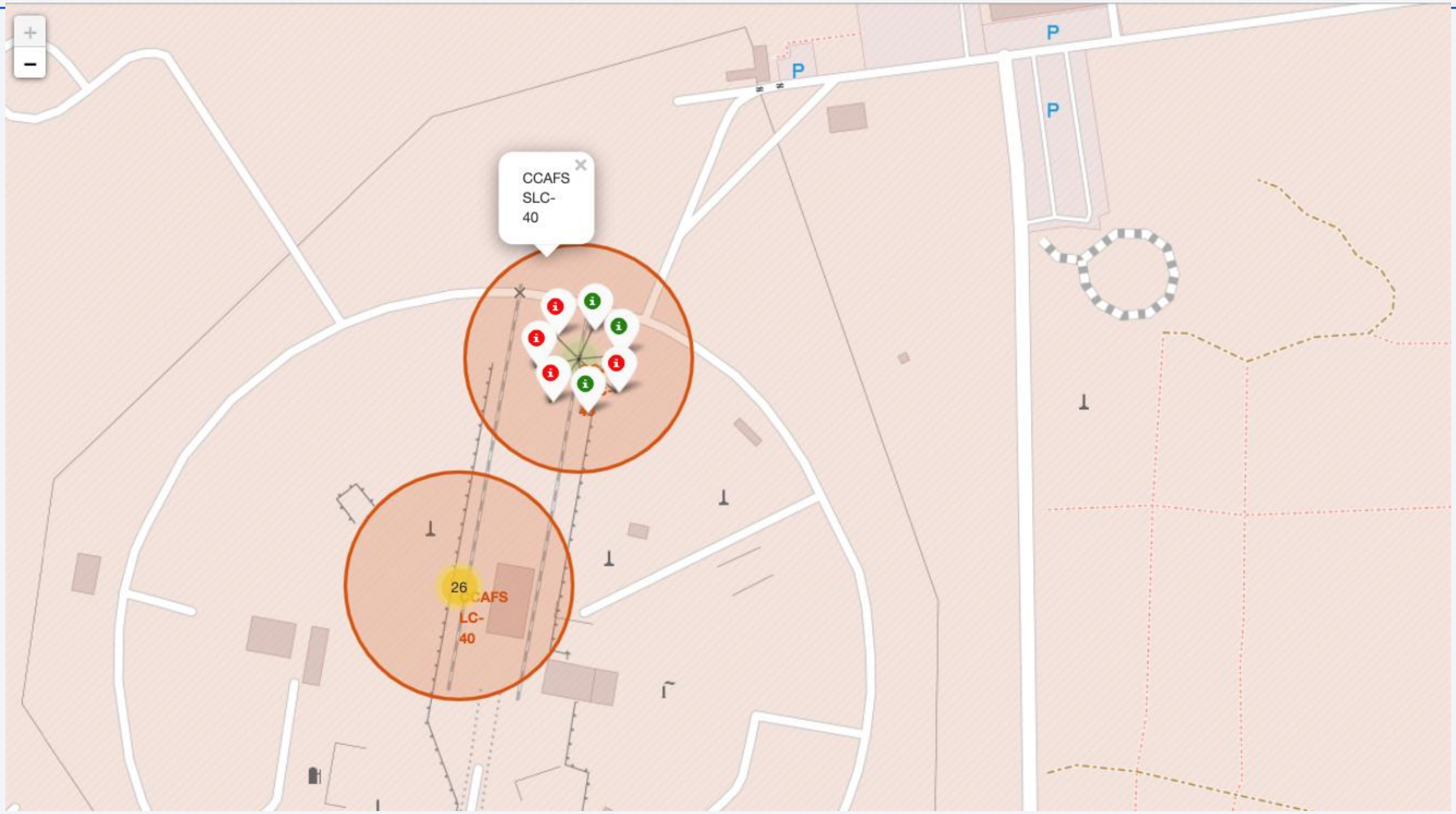
Launch Sites Proximities Analysis

<Folium Map Screenshot 1>



All launch sites are in proximity to the Equator line
All launch sites are in very close proximity to the coast

<Folium Map Screenshot 2>



- Replace <Folium map screenshot 2> title with an appropriate title

<Folium Map Screenshot 3>

- launch sites are in close proximity to coastline and far from cities

Section 5

Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>

- Replace <Dashboard screenshot 1> title with an appropriate title
- Show the screenshot of launch success count for all sites, in a piechart
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 2>

- Replace <Dashboard screenshot 2> title with an appropriate title

- Show the screenshot of the piechart for the launch site with highest launch success ratio

- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 3>

- Replace <Dashboard screenshot 3> title with an appropriate title
- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider
- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

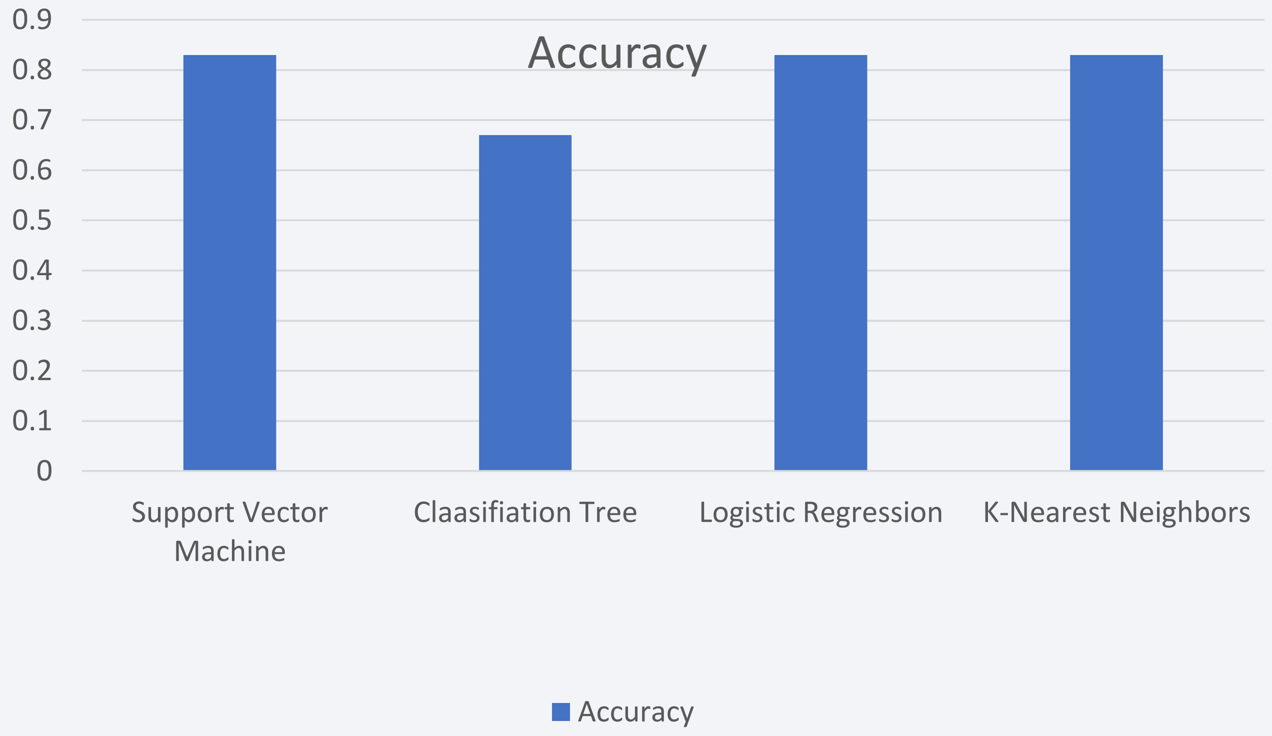
Section 6

Predictive Analysis (Classification)

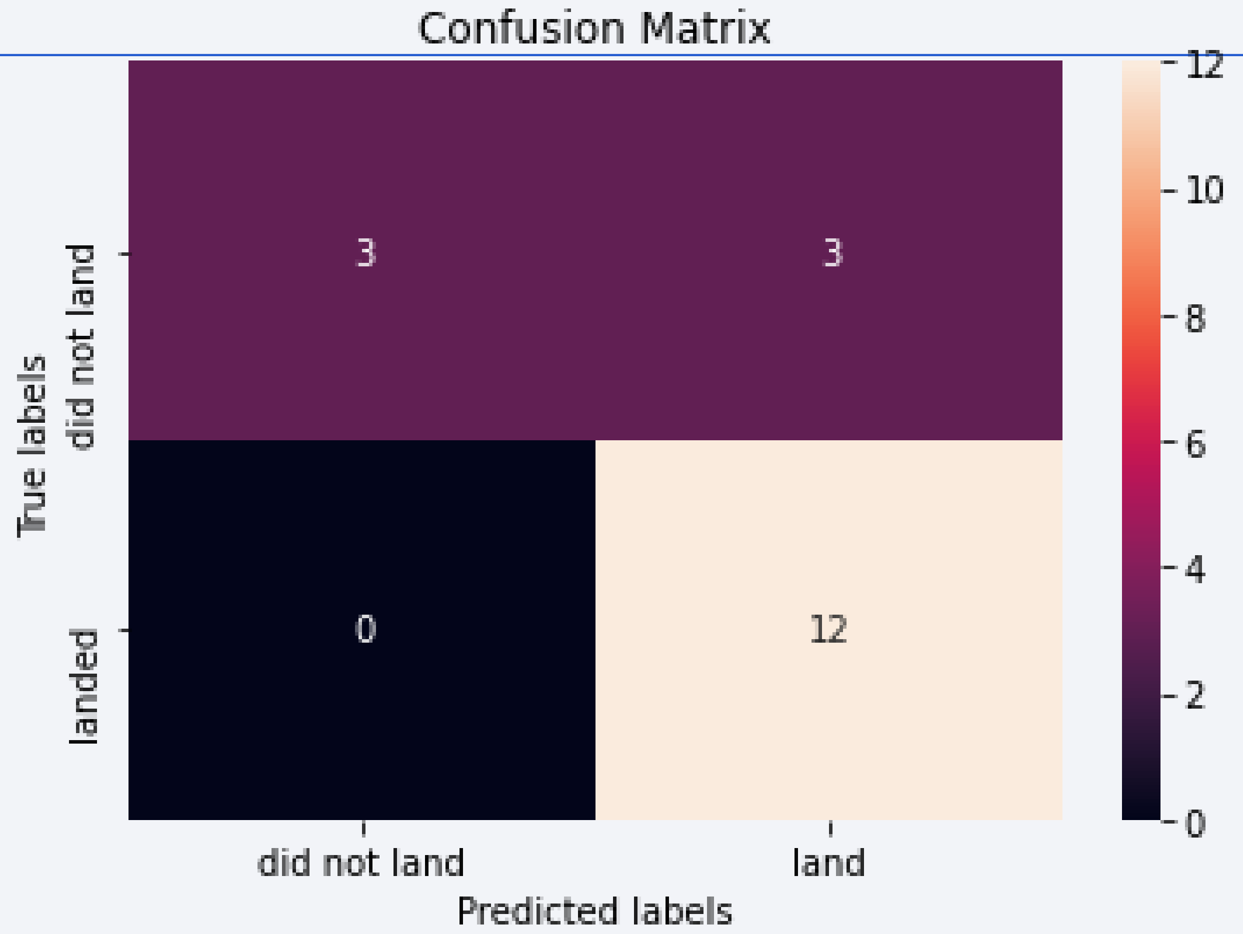
Classification Accuracy

Algorithm	Accuracy
Support Vector Machine	0.83
Claasifiation Tree	0.67
Logistic Regression	0.83
K-Nearest Neighbors	0.83

The best model are SVM, Logistic Regression and KNN



Confusion Matrix



Examining the confusion matrix, we see that logistic regression can distinguish between the different classes. We see that the major problem is false positives.

Conclusions

- The first stage of the Falcon 9 land successfully affect greatly to cost of the launch
- With data we get from previous launching of SpaceX, we can predict the successful rate of the first stage landing
- Based on model we build with the higest accuracy of 83.33%, we can predict the cost of the launch and we can bid against the SpaceX

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!