IBM Developer
SKILLS NETWORK

# Winning Space Race with Data Science

Julie-Ann McCalla
24 April 2024

# Outline

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

# Executive Summary

- Summary of methodologies
  - Data Collection
  - Data Wrangling
  - Data Visualization and SQL
  - Interactive visual analytics using Folium
  - Predictive analysis using classification models
- Summary of all results
  - Correlations variables
  - Build a predictive model

# Introduction

- Project background and context
  - The commercial space age is here, companies are making space travel affordable for everyone. Virgin Galactic is providing suborbital spaceflights. Rocket Lab is a small satellite provider. Blue Origin manufactures sub-orbital and orbital reusable rockets. Perhaps the most successful is SpaceX. SpaceX's accomplishments include: Sending spacecraft to the International Space Station. Starlink, a satellite internet constellation providing satellite Internet access. Sending manned missions to Space. One reason SpaceX can do this is the rocket launches are relatively inexpensive. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upwards 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.

- Problems you want to find answers
  - Use information gathered about Space X to determine the price of each launch.

Section 1

# Methodology

# Methodology

## Executive Summary

- Data collection methodology:
  - Data was collected using web scraping
- Perform data wrangling
  - Exploratory data analysis to find some patterns in the data and determine what would be the label for training supervised models
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Create a machine learning pipeline to predict if the first stage will land given the data from the preceding observation.

# Data Collection

- Describe how data sets were collected.

Import the relevant libraries
And helper functions

→

Request and parse the
SpaceX Launch data using
the GET request

→

Filter the data within the
dataframe to only use
Falcon9 launches

import requests
import pandas
import numpy
import datetime

GetBoosterVersion
getLaunchSite
GetPayloadData
getCoreData

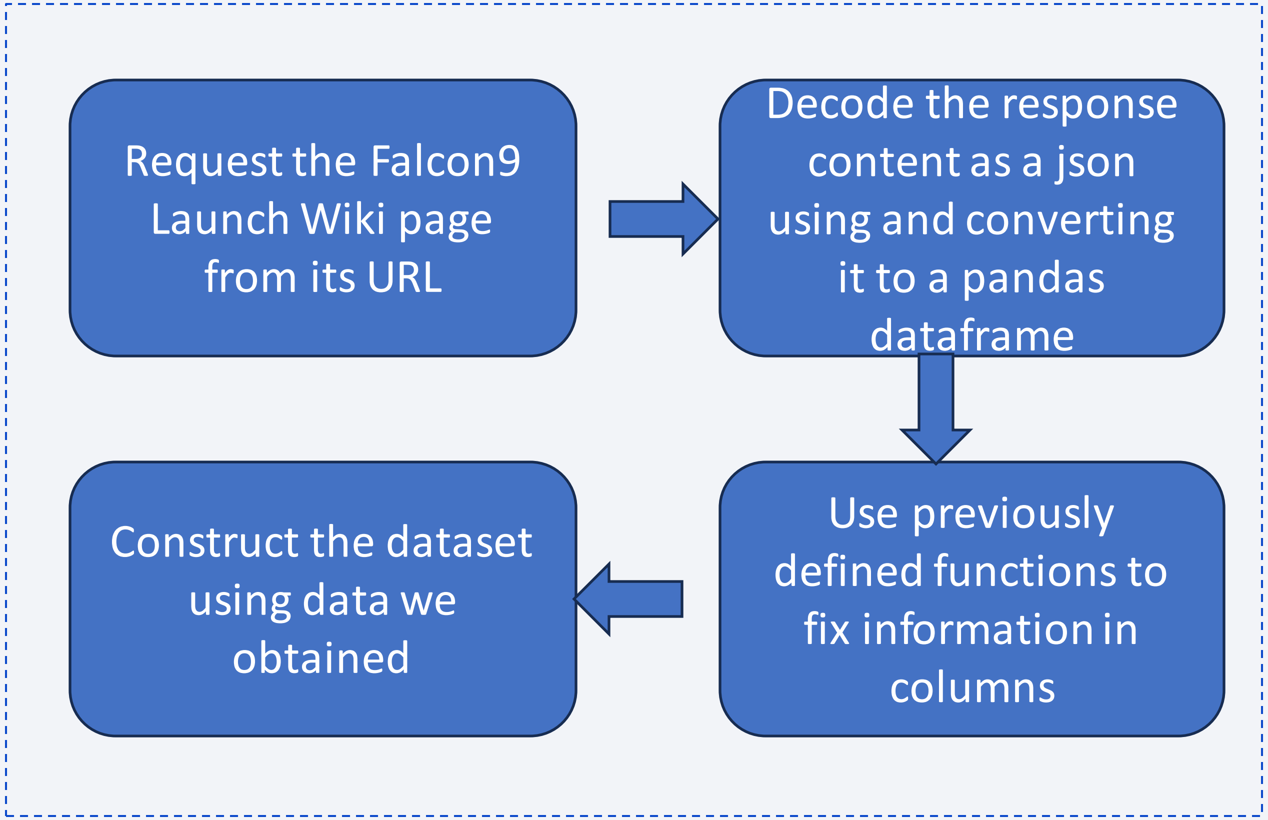
we decode the response
content as a Json
using .json() and turn it into
a Pandas dataframe
using .json_normalize()

Filter the data dataframe
using
the `BoosterVersion` colu
mn to only keep the Falcon 9
launches

# Data Collection – SpaceX API

GitHub URL:

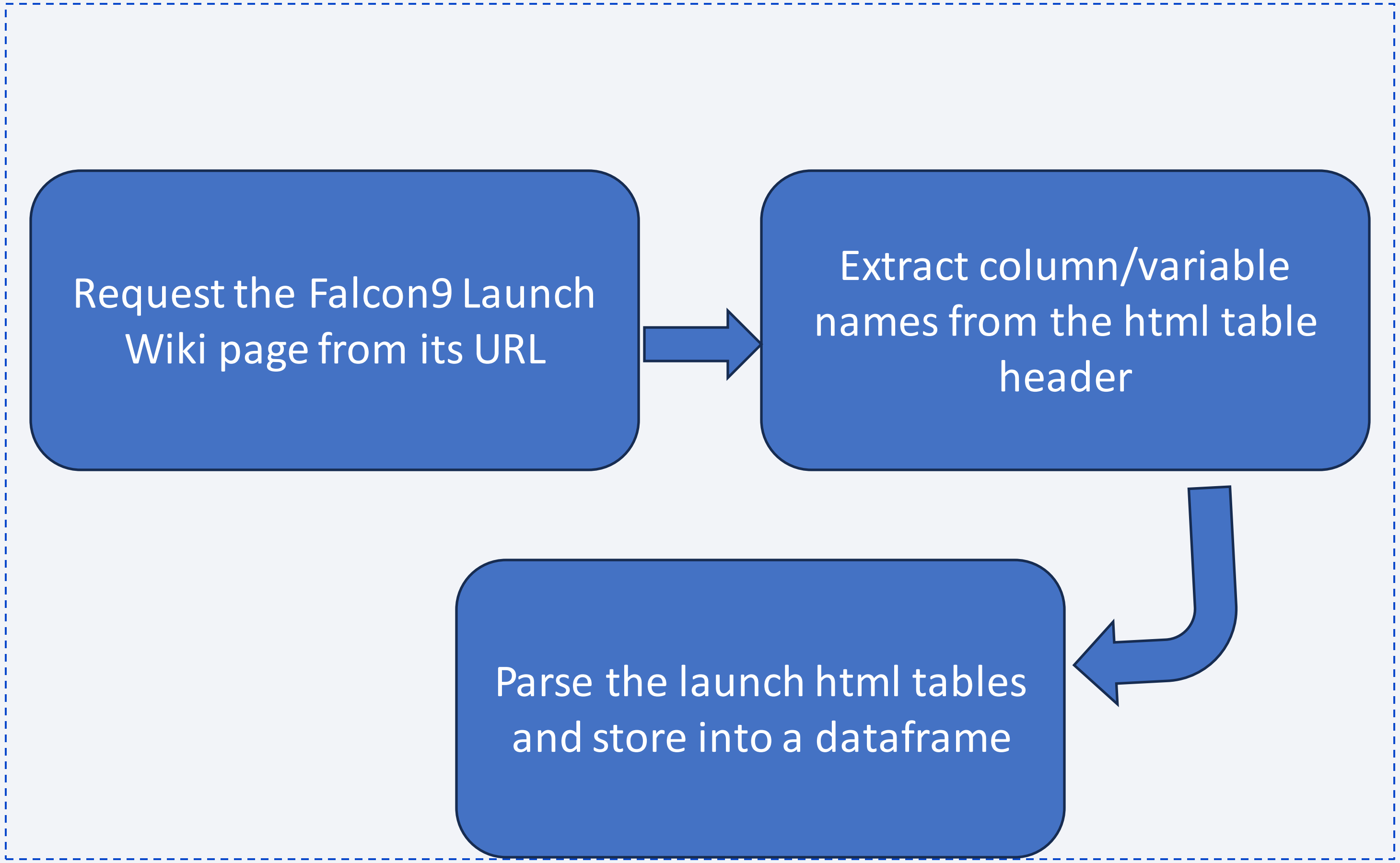
https://github.com/JulieDataSc/JulieRepos/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



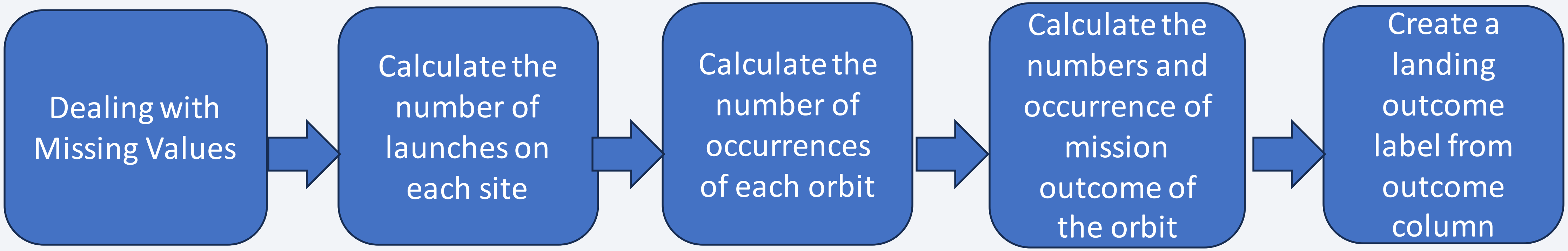
# Data Collection - Scraping

GitHub URL:

https://github.com/JulieDataSc/JulieRepos/blob/main/jupyter-labs-webscraping.ipynb



# Data Wrangling




**GitHub URL:** https://github.com/JulieDataSc/JulieRepos/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

# EDA with Data Visualization

- The following graphs were plotted for this project as they best represents the relationship between the considered variables:
  - Bar Charts
  - Scatter Charts
  - Line Charts
- GitHub URL: https://github.com/JulieDataSc/JulieRepos/blob/main/edadataviz.ipynb

# EDA with SQL

- SQL queries performed:
  - Display the names of 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 achieved.
  - 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 landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.



- GitHub URL: https://github.com/JulieDataSc/JulieRepos/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

# Build an Interactive Map with Folium

- Summarize what map objects:
    - Markers: Shows the geo location using longitude and latitude data
    - Cluster: This shows a group of markers
    - Circles: Shows a single location
    - Lines: Shows the distance between the two points

- GitHub URL:

https://github.com/JulieDataSc/JulieRepos/blob/main/lab_jupyter_launch_site_location%20(2).ipynb

# Build a Dashboard with Plotly Dash

- Summary of plots/graph:
  - Bar: Shows the difference in categories
  - Line: Provide reports on time series changes
  - Pie: Highlights percentages for events
  - Tree: Shows complex relationships for variables using interactive visualization
  - Map: Shows variables of states on maps

- GitHub URL:

  https://github.com/JulieDataSc/JulieRepos/blob/main/spacex_dash_app.py

# Predictive Analysis (Classification)

Create variable Y and standardize the variable assigned to X → Split in the train and test database → Calculate: Logis tic regression, Support Vector Machine, Decision Tree → Calculate the number and occurrence of mission outcome per orbit type → Compare score accuracies and Select the best model

- GitHub URL: https://github.com/JulieDataSc/JulieRepos/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

# Results

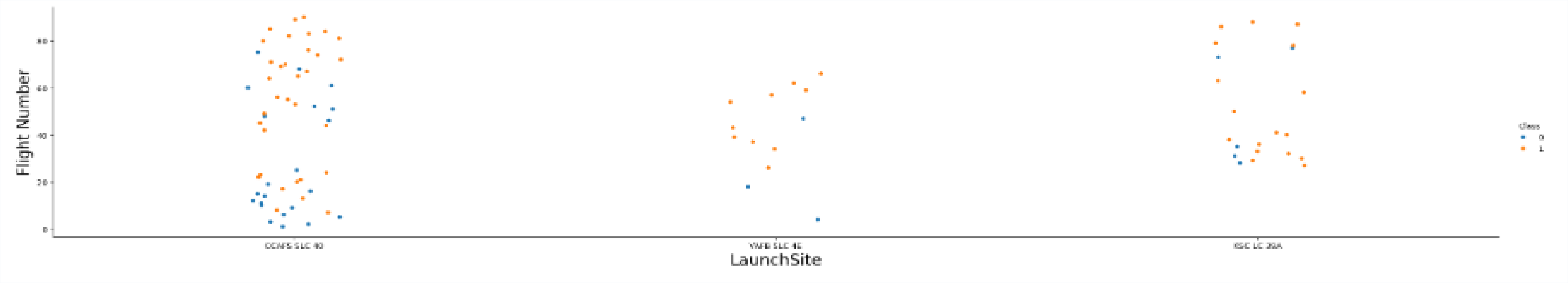
- Exploratory data analysis results
  - SpaceX data can be collected using web scraping and API
- Interactive analytics demo in screenshots
  - EDA with SQL is effective for data filtering
  - EDA with interactive visualization provides insight
  - Plotty Dash is a powerful to show instant data change
- Predictive analysis results
  - Decision Tree classifier had the best accuracy

Section 2

# Insights drawn from EDA

# Flight Number vs. Launch Site

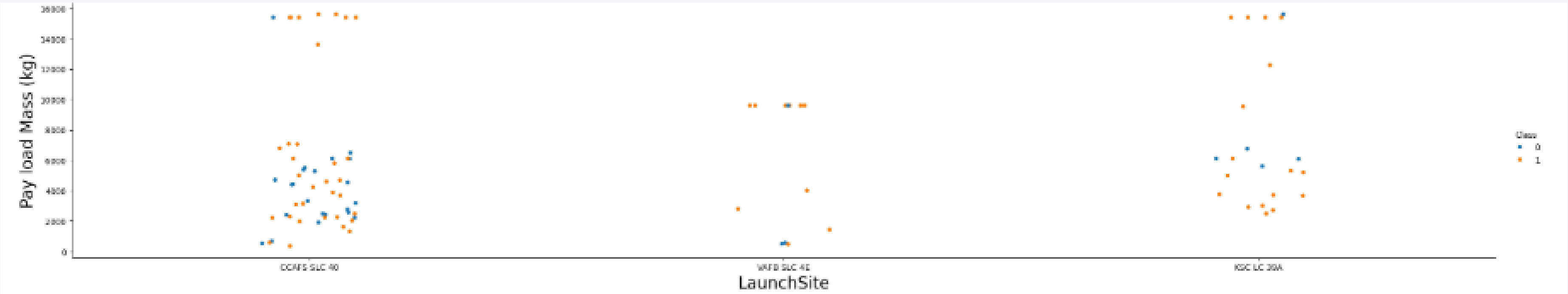
- Scatter plot of Flight Number vs. Launch Site




- CCFS SLC 40 Launch site has the greatest concentration and flight numbers.

# Payload vs. Launch Site

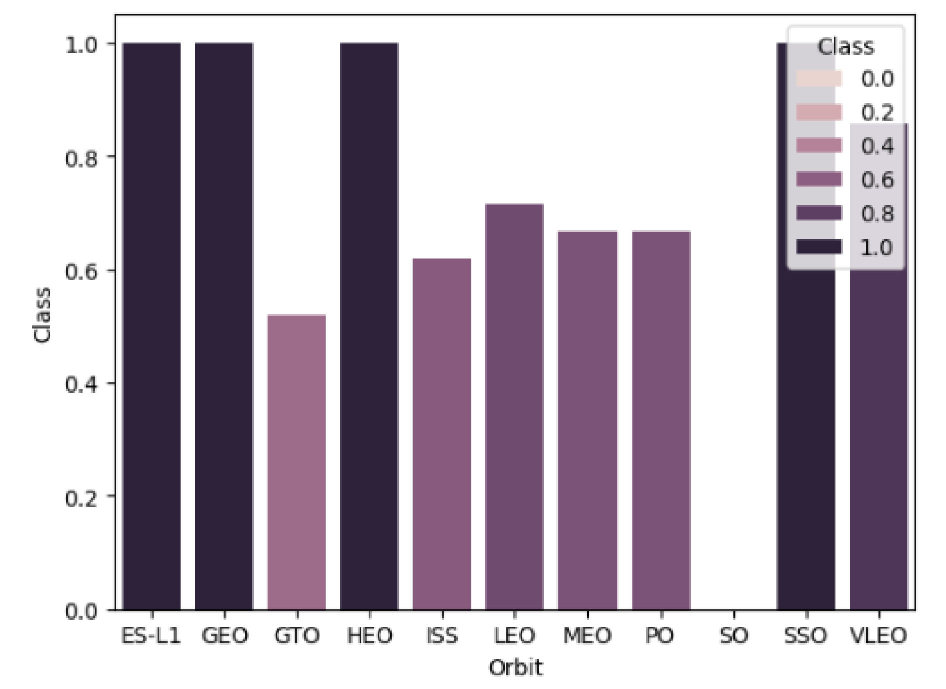
- Scatter plot of Payload vs. Launch Site




- The correlation between payload and launch site is relatively moderate.

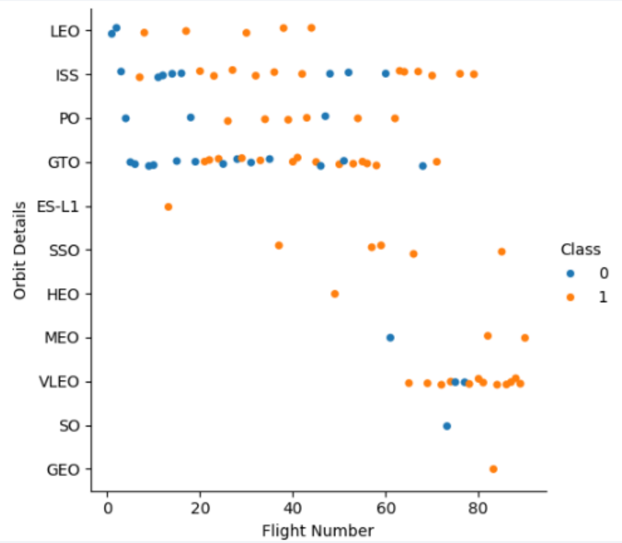
# Success Rate vs. Orbit Type

- Bar chart for the success rate of each orbit type



# Flight Number vs. Orbit Type

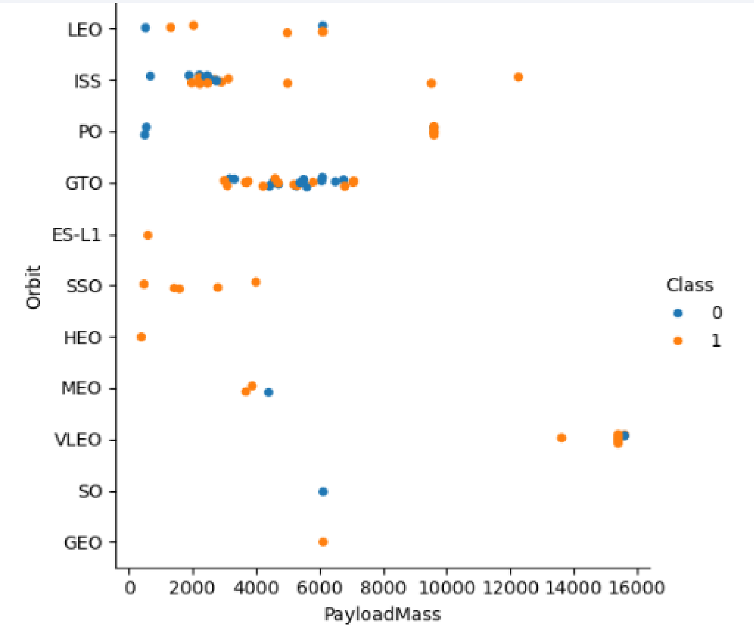
- Scatter point of Flight number vs. Orbit type




- VLEO has a high flight numbers for all its orbit showing a strong relationship, whilst could be seen as moderately or evenly distributed.

# Payload vs. Orbit Type

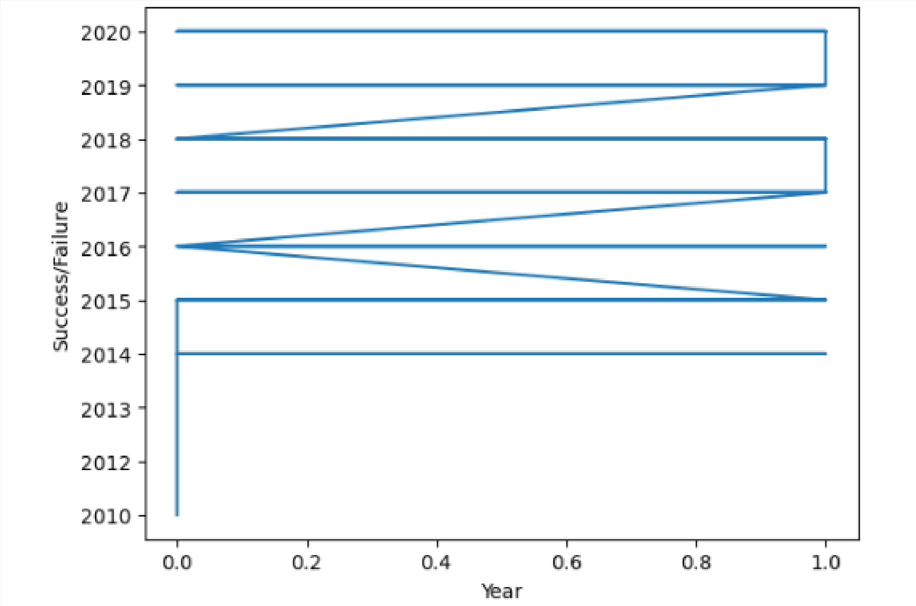
- Scatter point of payload vs. orbit type




- For VLEO payload negatively impacts the orbit.

# 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

- The names of the unique launch sites present in the database

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

# Launch Site Names Begin with 'CCA'

- Displaying five(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 | 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 | 7:44:00 | F9 v1.0 B0005 | CCAFS LC-40 | Dragon demo flight C2 | 525 | LEO (ISS) | NASA (COTS) | Success | No attempt |
| 2012-10-08 | 0:35:00 | 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

- Payload carried by boosters from NASA

| payloadmass |
|---|
| 619967 |

# Average Payload Mass by F9 v1.1

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

| payloadmass |
|---|
| 6138.287128712871 |

# First Successful Ground Landing Date

- Date of the first successful landing outcome on ground pad

| min(DATE) |
|---|
| 2010-06-04 |

# 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

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

# Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes

| Mission_Outcome | missionoutcomes |
|---:|---:|
| Failure (in flight) | 1 |
| Success | 98 |
| Success | 1 |
| Success (payload status unclear) | 1 |

# Boosters Carried Maximum Payload

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

| boosterversion |
|---|
| 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

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

| | | | |
|---|---|---|---|
| 06 | Failure (in flight) | F9 v1.1 B1018 | CCAFS LC-40 |

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

- The rank by 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 | COUNT |
|---:|---:|
| No attempt | 10 |
| Success (drone ship) | 5 |
| Failure (drone ship) | 5 |
| Success (ground pad) | 3 |
| Controlled (ocean) | 3 |
| Uncontrolled (ocean) | 2 |
| Failure (parachute) | 2 |
| Precluded (drone ship) | 1 |

Section 3

# Launch Sites Proximities Analysis

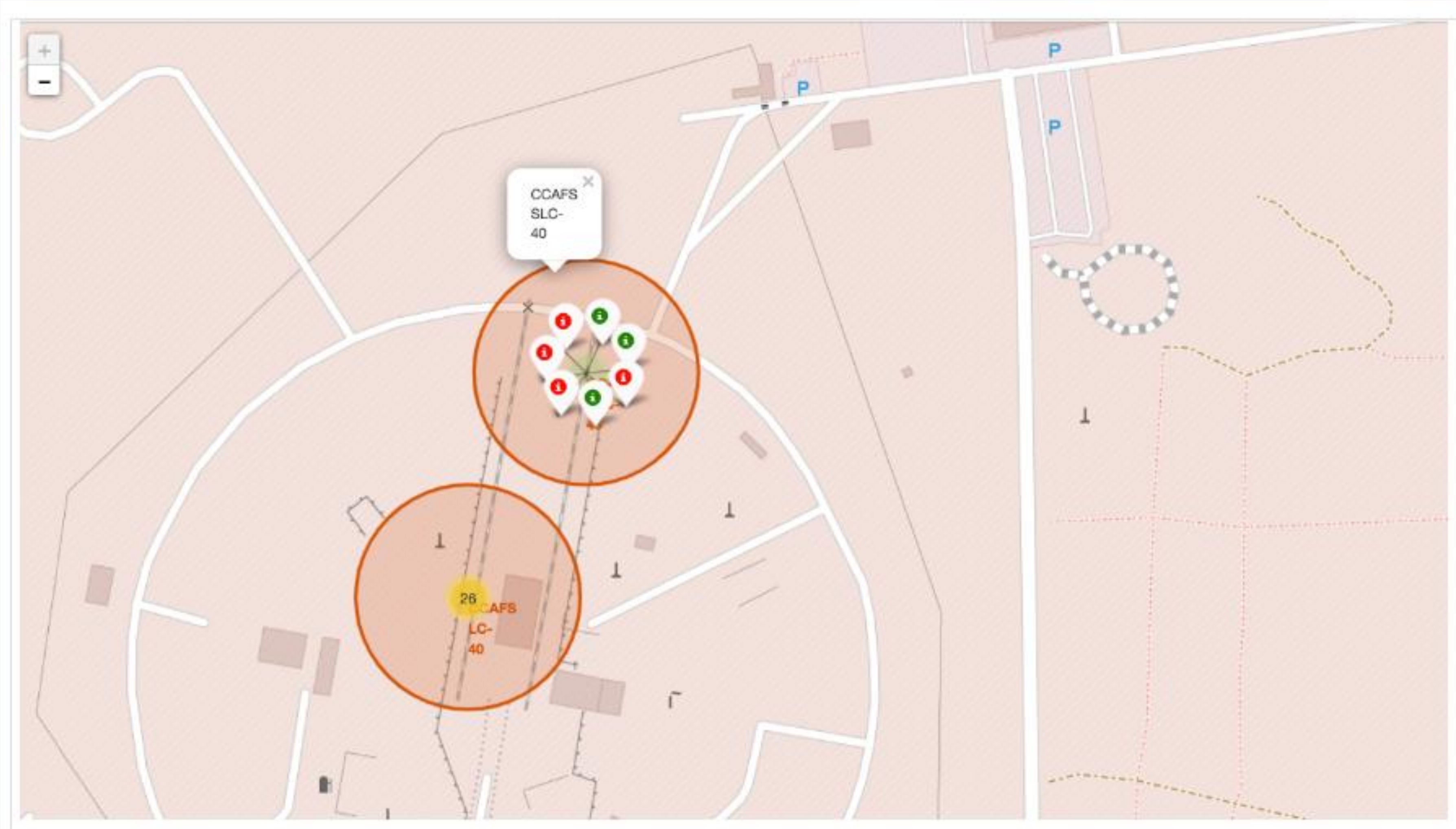
# Launch Sites

Both sites at ground and sea surface are necessary.

The generated map with marked launch sites should look similar to the following:

# Success of launch

- Adding colour elements is an impressive way of displaying successful launch



# Distance

- The plot of distance between locations allows for estimation of travel time. And decision making.

Section 4

# Build a Dashboard with Plotly Dash

# The Most Successful Launch Site

- 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

# Payload and Successful launch

- 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 5

# Predictive Analysis (Classification)

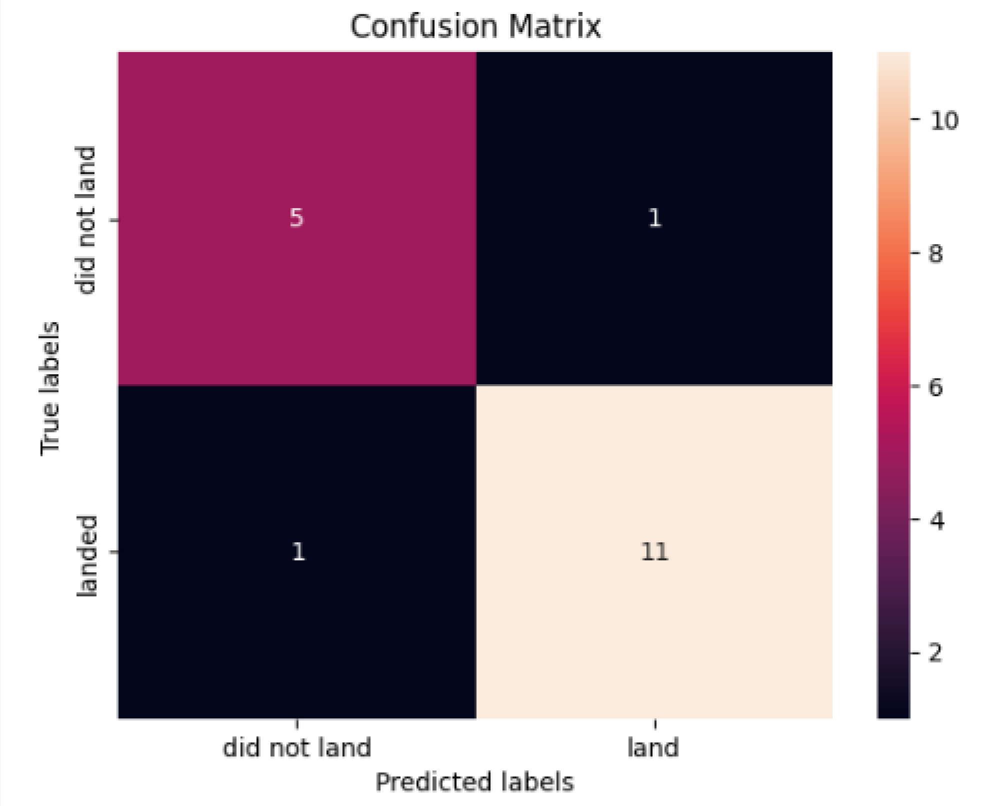
# Classification Accuracy

- The model accuracy for Decision Tree is the best whilst the other classification models are the same.

| Method | Test Data Accuracy |
|---|---|
| Logistic_Reg | 0.833333 |
| SVM | 0.833333 |
| Decision Tree | 0.888889 |
| KNN | 0.833333 |

# Confusion Matrix

- The confusion matrix of the best performing model is the Decision Tree. It shows a 88% accuracy for predictions.



# Conclusions

- Modeling can provide insight for data through visualization and aide in the decision making process
- The success rate for SpaceX launches has been increasing moderately and with time and correct payload the required target will be met
- Using a Decision tree model classifier for this dataset is the best.

# 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!