IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Rajan Meikandan
18.05.2025

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection supported by SpaceX API and web scraping methods(requests, beautifulsoup etc)
  - Clean the collected data and replace the missing values with suitable values(mean etc)
  - Data Wrangling – Exploratory Data Analysis
  - Data Visualization with Matplotlib and Seaborn libraries
  - Launch Sites Location Analysis using Open Street Map with Folium
  - Build a Plotly Dash application for users to perform interactive visual analytics on SpaceX launch data in real-time
  - Use Machine Learning(SVM, Classification Trees, and Logistic Regression etc) to determine if the first stage of Falcon 9 will land successfully
- Summary of all results
  - Decision Tree classifier has the best prediction accuracy of 94% with only one false positive

Introduction

- Project background and context
  - The aim of this project is to predict if the SpaceX's Falcon 9 first stage will land successfully based on the data collected from previous rocket launches which contains both successful and failed launches
  - 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
- Problems you want to find answers
  - Can a Machine Learning Model predict the success rate of Falcon 9 first stage launch with high Accuracy?
  - Which ML Model is most suitable for prediction?

Section 1

Methodology

Methodology

Executive Summary

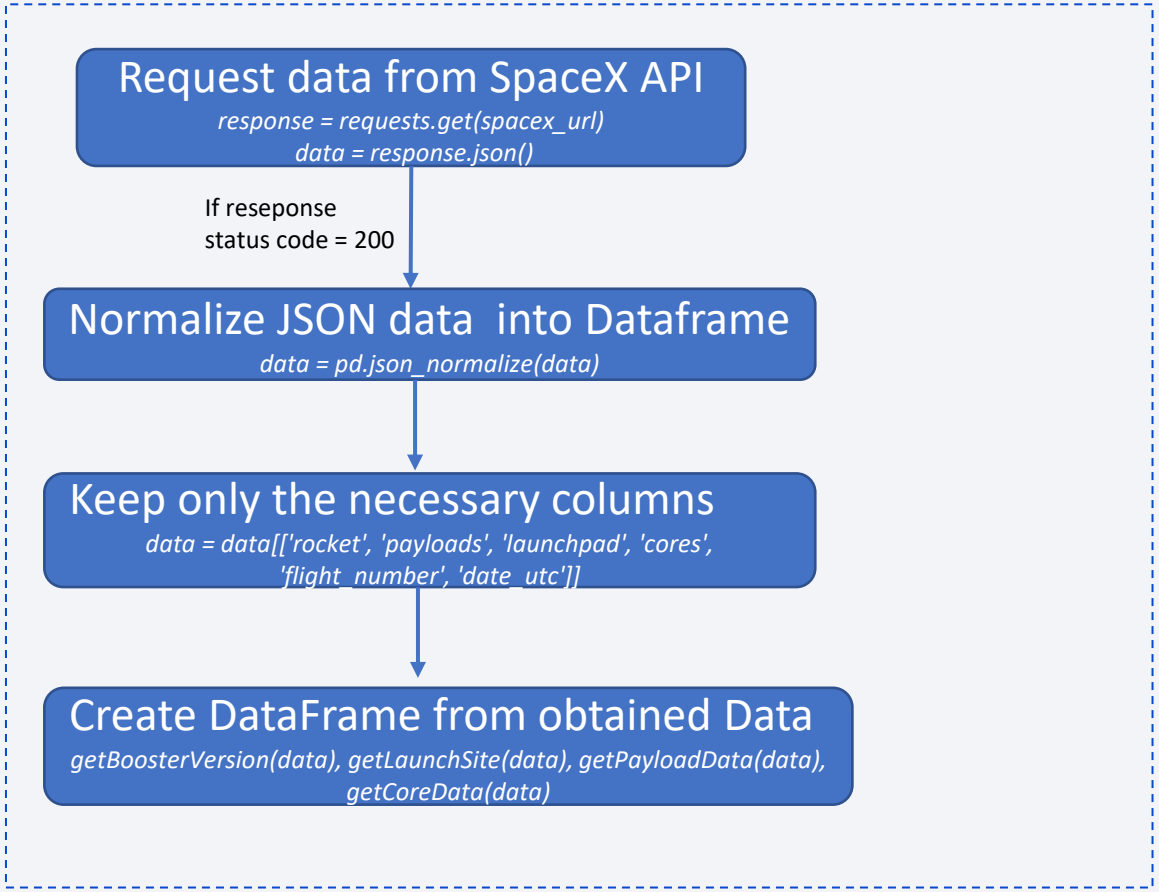
- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts

- Add the GitHub URL of the completed SpaceX API calls notebook (must include completed code cell and outcome cell), as an external reference and peer-review purpose

https://github.com/rajanme29/testrepo/blob/main/jupyter-labs-spacex-data-collection-api-v2.ipynb

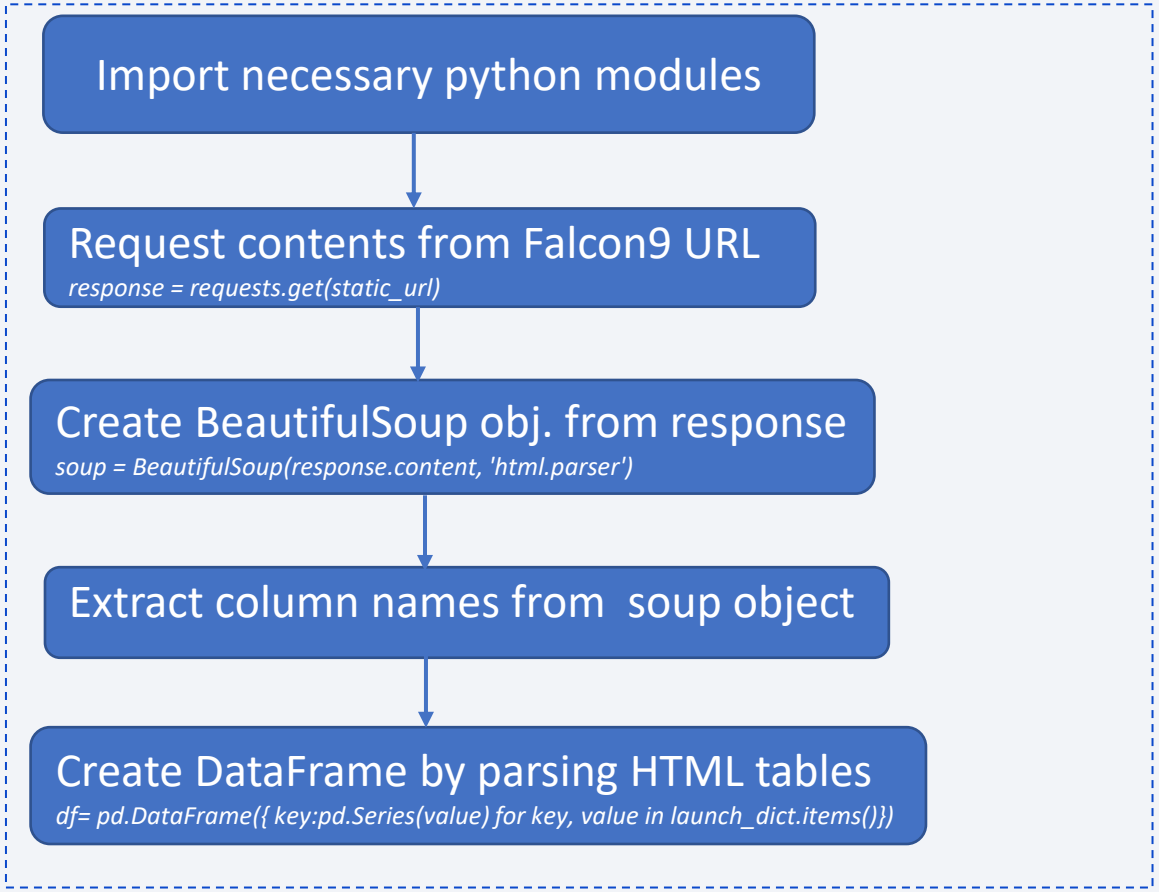


Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose

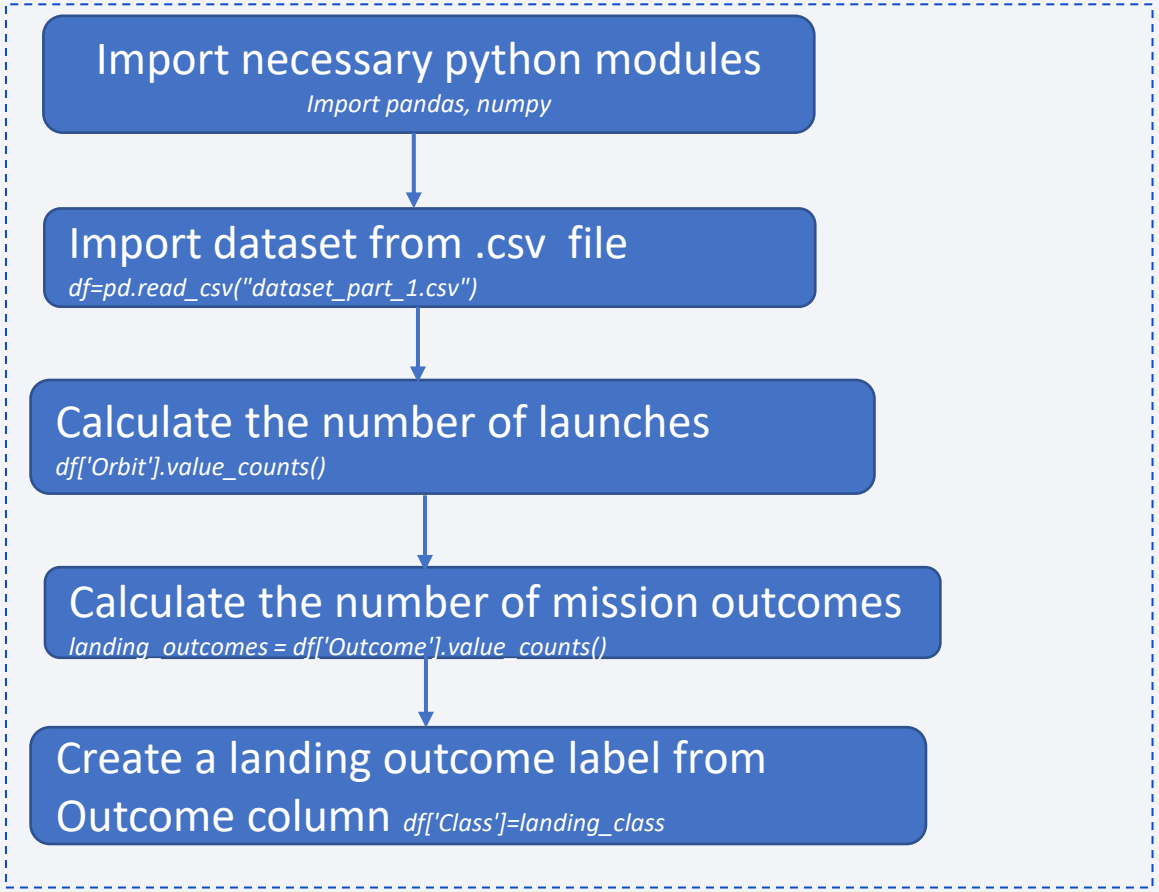
https://github.com/rajanme29/testrepo/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- Describe how data were processed
- 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/rajanme29/testrepo/blob/main/labs-jupyter-spacex-Data%20wrangling-v2.ipynb



EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
- Add the GitHub URL of your completed EDA with data visualization notebook, as an external reference and peer-review purpose

https://github.com/rajanme29/testrepo/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Charts used for Visualization:

- Scatter plot to visualize the relationship between flight number and launch site
- Scatter plot to visualize the relationship between Payload and Launch Site
- Bar chart to visualize the relationship between success rate of each orbit type
- Scatter plot to visualize the relationship between FlightNumber and Orbit type
- Scatter plot to visualize the relationship between Payload and Orbit type
- Line plot to visualize the launch success yearly trend

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose

https://github.com/rajanme29/testrepo/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

SQL Queries:

- *select distinct [Launch_Site] from SPACEXTBL*
- *select top 5 Launch_Site*
  from SPACEXTBL where Launch_Site like 'CCA%';
- *select sum(PAYLOAD_MASS__KG_)*
  from SPACEXTBL where Customer like 'NASA (CRS)'
- *select avg(PAYLOAD_MASS__KG_)*
  from SPACEXTBL where Booster_Version like 'F9 v1.1'
- *select min(Date) from SPACEXTBL*
  where Landing_Outcome like 'Success (ground pad)'
- *select Booster_Version, PAYLOAD_MASS__KG_, Landing_Outcome from SPACEXTBL*
  where (Landing_Outcome like 'Success (drone ship)') and (PAYLOAD_MASS__KG_ > 4000 and PAYLOAD_MASS__KG_ < 6000)
- *select Mission_Outcome, count(Mission_Outcome) from SPACEXTBL*
  group by mission_outcome

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- Explain why you added those objects
  - folium.Map – To visualize the geographical location of the launch site on the map
  - folium.Circle – To add a highlighted circle area with a text label on a specific coordinate
  - folium.map.Marker - To add a marker to a location on the map
  - folium.PolyLine - Calculate the distances between a launch site to its proximities
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

https://github.com/rajanme29/testrepo/blob/main/lab-jupyter-launch-site-location-v2.ipynb

Build a Dashboard with Plotly Dash

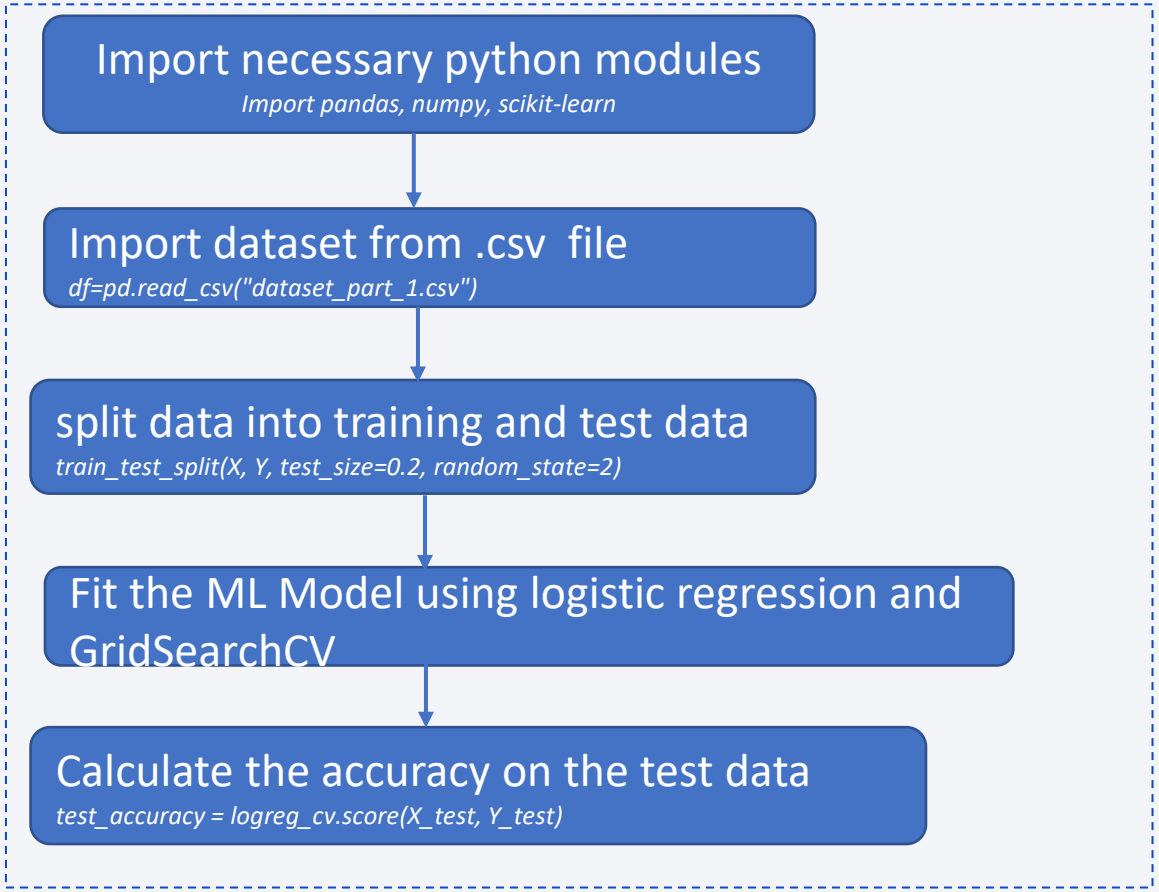
- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
  - Pie Chart – to visualize the success rate of each launch site
  - Dropdown list – to select the launch site
  - Range slider - to select payload mass range to be shown on the scatter plot
  - Scatter Plot - to show the correlation between payload mass and launch success
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

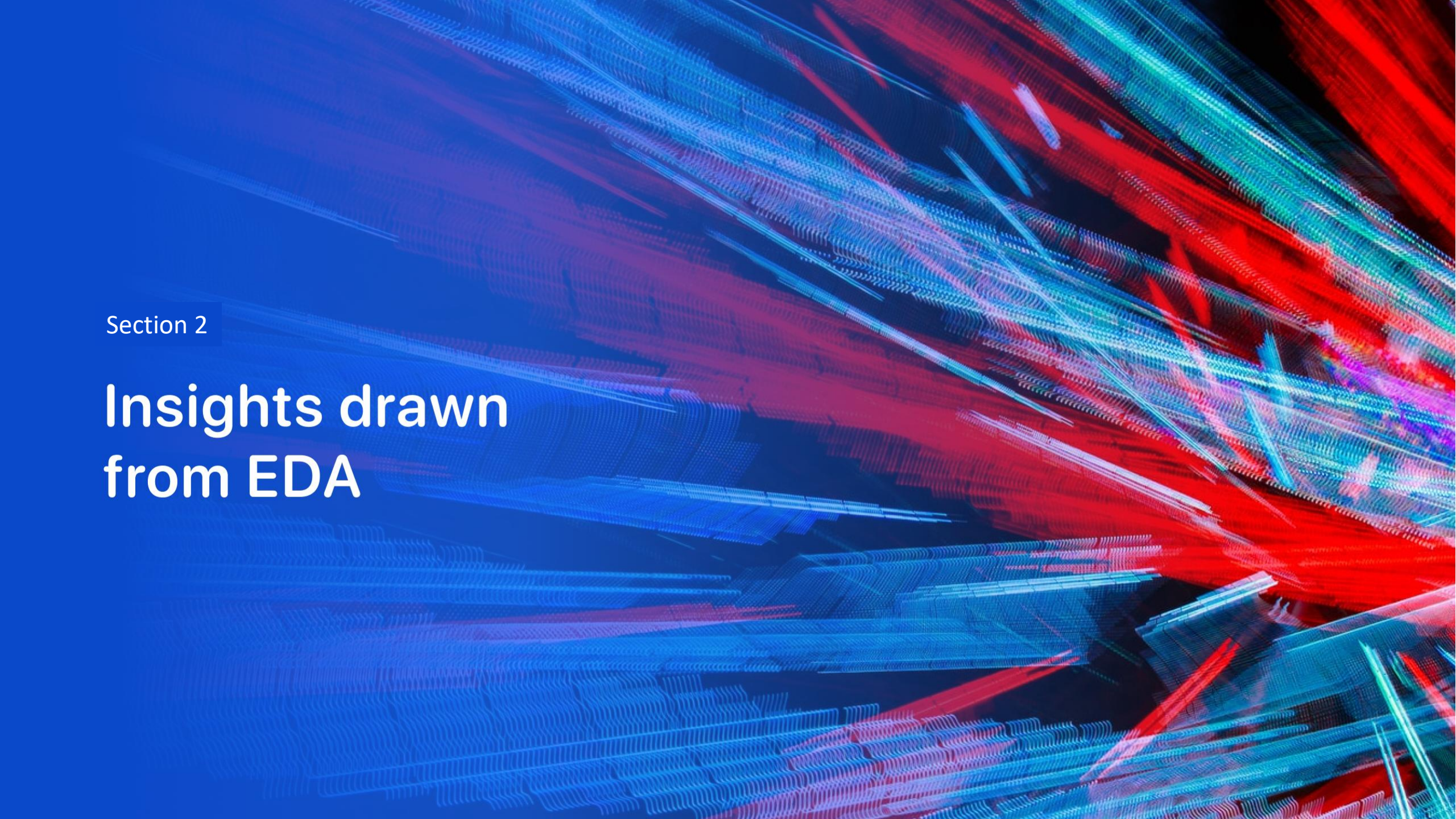
https://github.com/rajanme29/testrepo/blob/main/jupyter.labs-spacex-Dash.ipynb

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- You need present your model development process using key phrases and flowchart
- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose

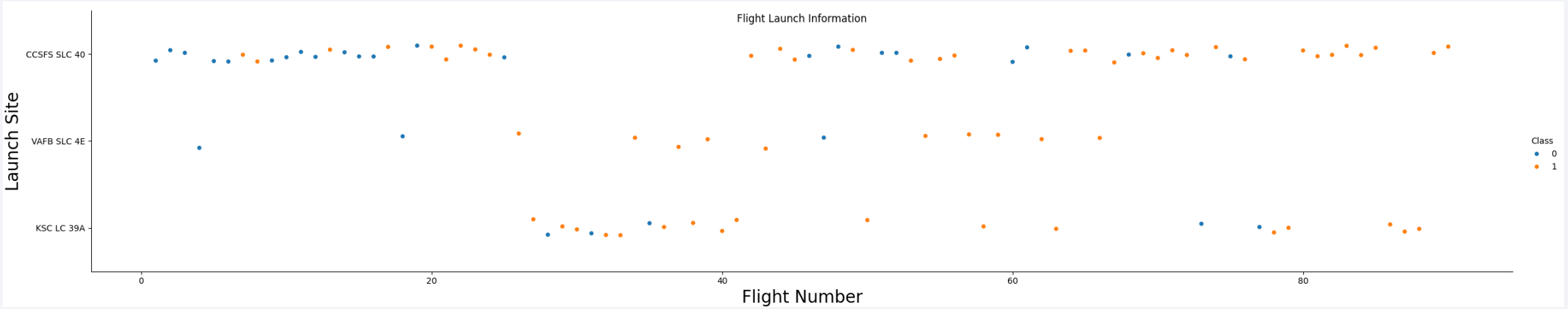
https://github.com/rajanme29/test repo/blob/main/SpaceX-Machine-Learning-Prediction-Part-5-v1.ipynb




Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

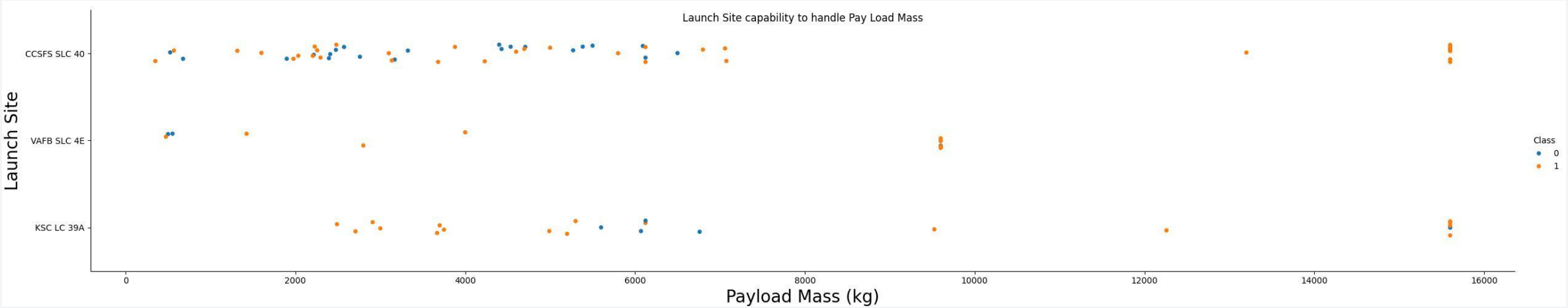
- Show a scatter plot of Flight Number vs. Launch Site



- Show the screenshot of the scatter plot with explanations
  - Around 100% success rate at launch site CCSFS SLC 40 for flight numbers > 80
  - Very few launches from launch site VAFB SLC 4E
  - Lower failure rate at launch site KSC LC-39A for flight numbers > 50

Payload vs. Launch Site

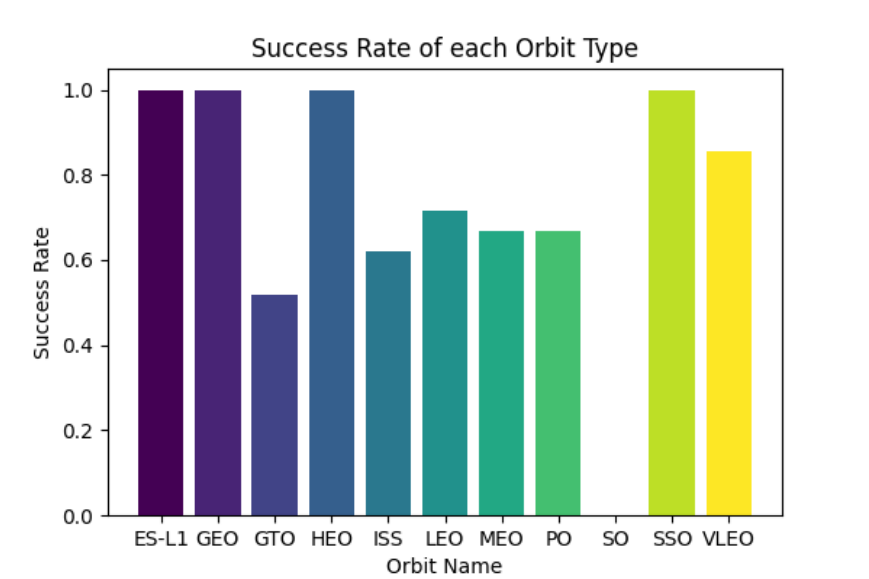
- Show a scatter plot of Payload vs. Launch Site



- Show the screenshot of the scatter plot with explanations
  - No rockets were launched from VAFB SLC-4E with payload mass > 10000kg
  - Almost 100% success rate at CCSFS SLC 40 for launches with payload mass > 7000kg
  - The launch KSC LC-39A has most success rate for launches with payload mass < 5000kg

Success Rate vs. Orbit Type

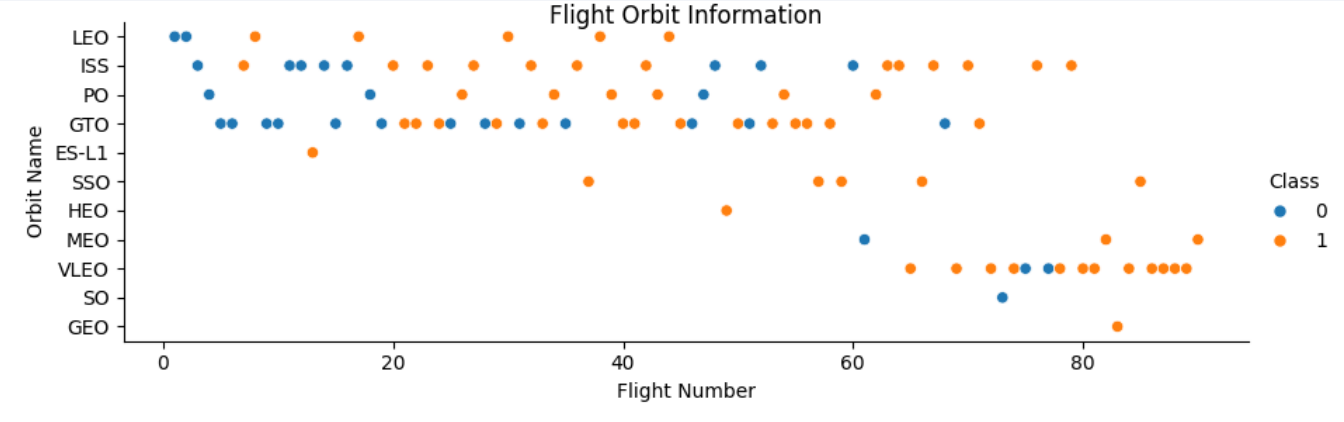
- Show a bar chart for the success rate of each orbit type



- Show the screenshot of the scatter plot with explanations
  - Almost 100% success rate for the launches to the orbits ES-L1, GEO, HEO and SSO
  - Lowest success rate for the launch to the orbit GTO (around 50% success)

Flight Number vs. Orbit Type

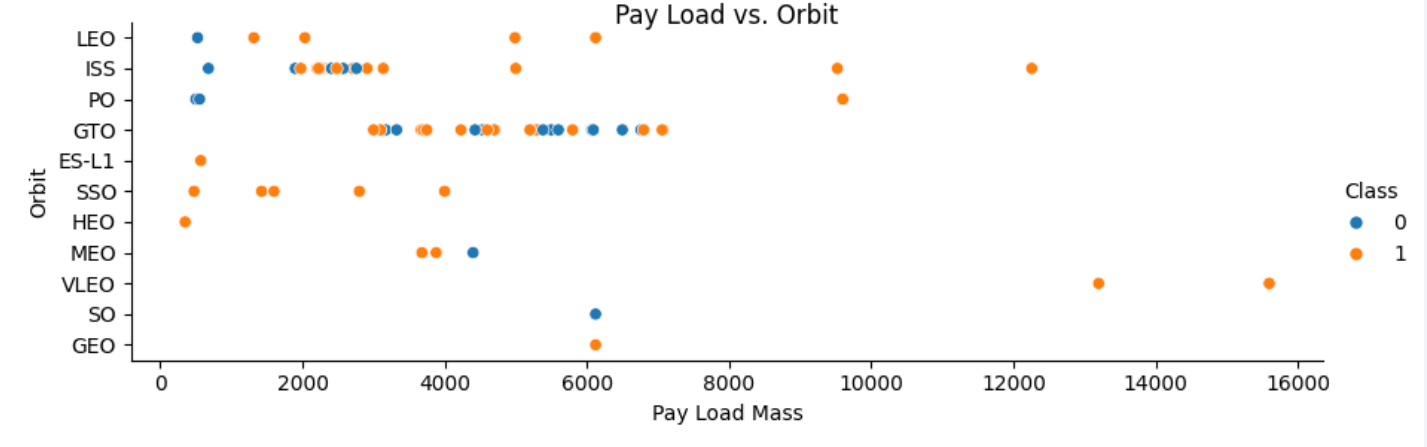
- Show a scatter point of Flight number vs. Orbit type



- Show the screenshot of the scatter plot with explanations
  - Almost 100% success rate for launches after flight number > 80, which is independent of the orbit
  - Very low successful launches below flight number 20
  - The orbit GEO has the lowest number of rocket launches, just 1

Payload vs. Orbit Type

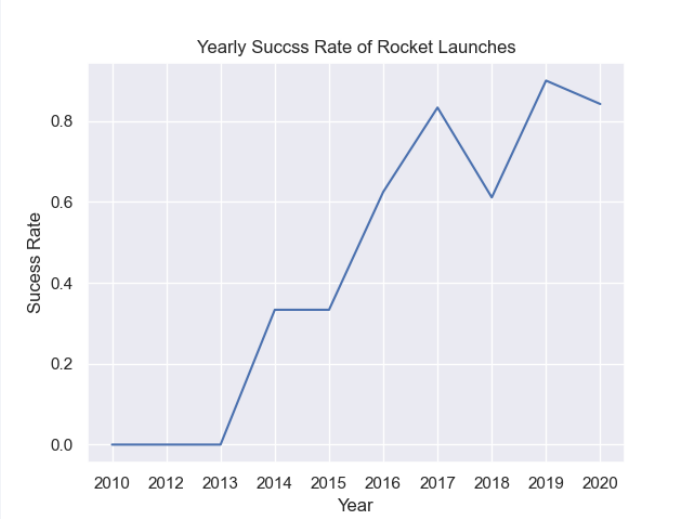
- Show a scatter point of payload vs. orbit type



- Show the screenshot of the scatter plot with explanations
  - The rockets launched to the VLEO orbit have the highest payload mass, and both of the rockets were launched successfully
  - A rocket with very low payload mass(<1000kg) was successfully launched to HEO
  - The maximum payload for the rockets which were launched to the Orbit SSO was 4000kg

Launch Success Yearly Trend

- Show a line chart of yearly average success rate



- Show the screenshot of the scatter plot with explanations
  - The rockets launched by SpaceX until 2013 were failed to reach the designated orbits
  - The ratio of successful launches increasing after 2013
  - Almost 100% successful launch rate in 2019

All Launch Site Names

- Find the names of the unique launch sites
  - CCSFS SLC 40
  - CCAFS LC-40
  - VAFB SLC 4E
  - KSC LC 39A
- Present your query result with a short explanation here
  - *select distinct [Launch_Site] from SPACEXTBL*

  This query selects the distinct launch site names from the ‘Spacextbl’ table

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
  - *select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where Customer like 'NASA (CRS)'*
- Present your query result with a short explanation here
  - Query Result: **45596 kg**
  - The query sum up the payload mass from the Spacextbl where the rocket belongs to NASA

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- CCAFS LC-40
- CCAFS LC-40
- CCAFS LC-40
- CCAFS LC-40
- CCAFS LC-40

- Present your query result with a short explanation here
  - *select top 5 [Launch_Site] from SPACEXTBL where [Launch_Site] like 'CCA%';*
  - The script selects the top 5 launch sites which begins with CCA from the table ‘Spacextbl’

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
  - 2928kg

- Present your query result with a short explanation here
  - *select avg([PAYLOAD_MASS__KG_]) from SPACEXTBL where [Booster_Version] like 'F9 v1.1';*
  - The average payload mass for the booster version F9 V1.1 shall be calculated with the avg() function

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
  - 2015-12-22

- Present your query result with a short explanation here
  - select min(Date) from SPACEXTBL where [Landing_Outcome] like 'Success (ground pad)'
  - The first successful landing outcome of rocket launched from ground pad shall be found by using min(Date) function, where Date is the launching dates of all launches in the %Y-%M-%D format

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	PAYLOAD_MASS__KG_	Landing_Outcome
F9 FT B1022	4696	Success (drone ship)
F9 FT B1026	4600	Success (drone ship)
F9 FT B1021.2	5300	Success (drone ship)
F9 FT B1031.2	5200	Success (drone ship)

- Present your query result with a short explanation here
  - select [Booster_Version], [PAYLOAD_MASS__KG_], [Landing_Outcome] from SPACEXTBL where ([Landing_Outcome] like 'Success (drone ship)') and ([PAYLOAD_MASS__KG_] > 4000 and [PAYLOAD_MASS__KG_] < 6000)

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

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

- Present your query result with a short explanation here
  - select [Mission_Outcome], count([Mission_Outcome]) from SPACEXTBL group by [mission_outcome]
  - The Mission outcome is very successful with 99% success rate

Boosters Carried Maximum Payload

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

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

- Present your query result with a short explanation here
  - select [Booster_Version], [PAYLOAD_MASS__KG_] from SPACEXTBL where [PAYLOAD_MASS__KG_] = (select max([PAYLOAD_MASS__KG_]) from SPACEXTBL)

2015 Launch Records

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

	Booster_Version	Launch_Site	Landing_Outcome
January	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
April	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

- Present your query result with a short explanation here
  - select datename(m, Date), Booster_Version, Launch_Site, Landing_Outcome from SPACEXTBL where YEAR(Date) = 2015 and (Landing_Outcome like 'Failure (drone ship)')
  - The date, booster version, launch site name and landing outcome for the year 2015 shall be extracted by fulfilling 2 conditions such as Date and String which contains ‘Failure (drone ship)’

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

- Present your query result with a short explanation here
  - select Landing_Outcome, count(Landing_Outcome) from SPACEXTBL
    where Date between '2010-06-04' and '2017-03-20'
    group by Landing_Outcome
    order by count(Landing_Outcome) desc

Section 3

Launch Sites Proximities Analysis

Launch Sites Location Analysis with Folium

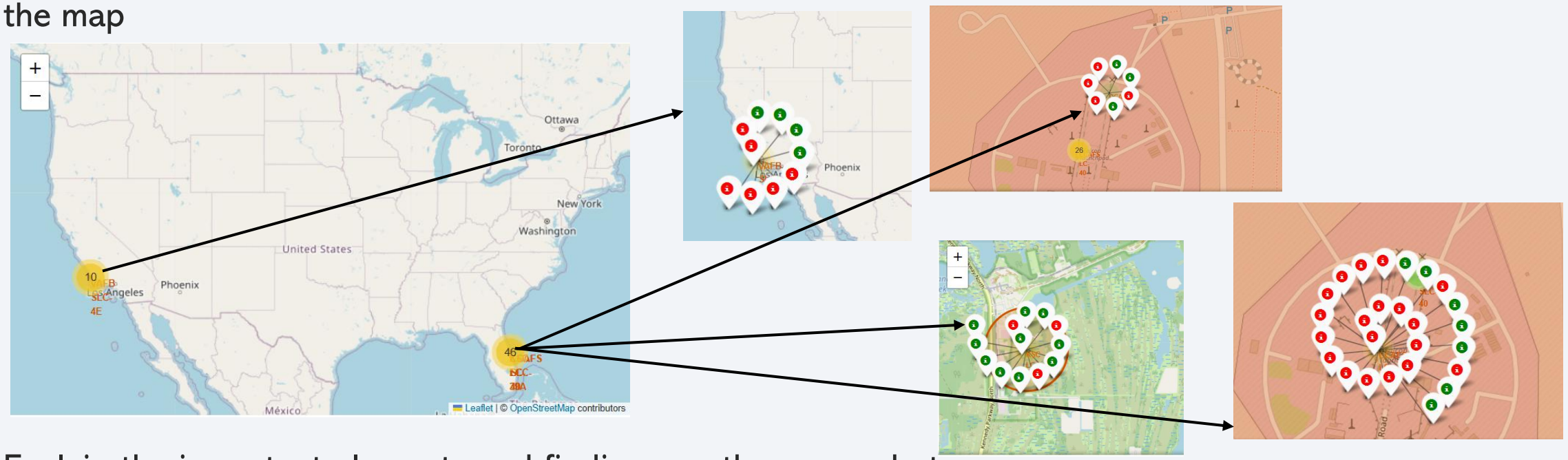
- Explore the generated folium map and make a proper screenshot to include all launch sites' location markers on a global map

	Launch Site	Lat	Long
0	CCAFS LC-40	28.562302	-80.577356
1	CCAFS SLC-40	28.563197	-80.576820
2	KSC LC-39A	28.573255	-80.646895
3	VAFB SLC-4E	34.632834	-120.610745

- Explain the important elements and findings on the screenshot
  - The launch sites identified on the folium map with following to objects folium.Circle and folium.map.Marker
  - Three launch sites are in Florida and one launch site is located in California

Launch Outcome

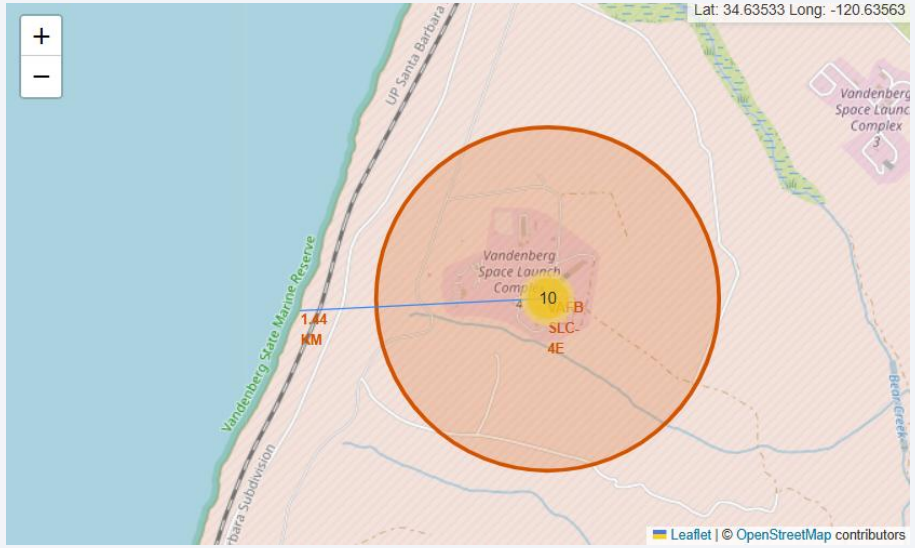
- Explore the folium map and make a proper screenshot to show the color-labeled launch outcomes on the map

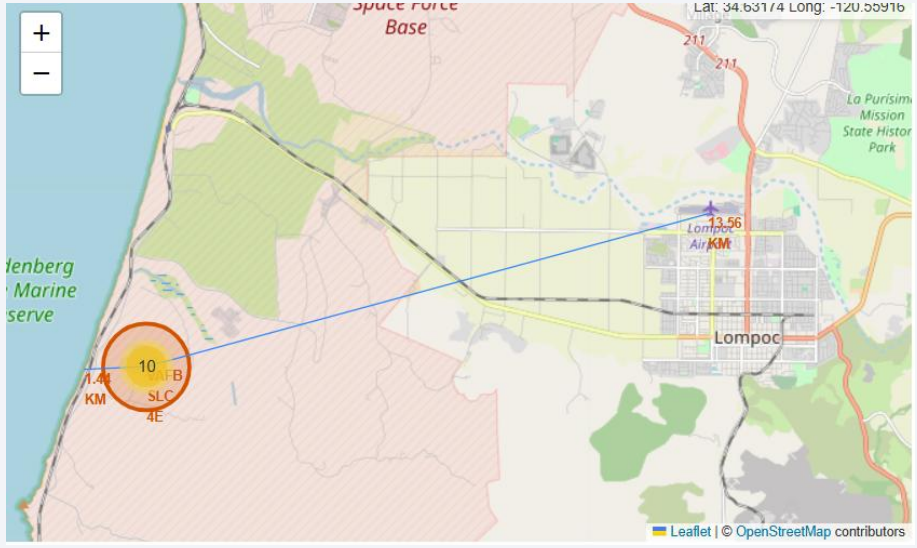


- Explain the important elements and findings on the screenshot
  - The launch sites identified on the folium map with folium.Circle and folium.map.Marker
  - Launch outcome such as success(green) or failure(red) marked on the map with folium Icon object folium.Icon(color='white', icon_color=record['marker_color'])

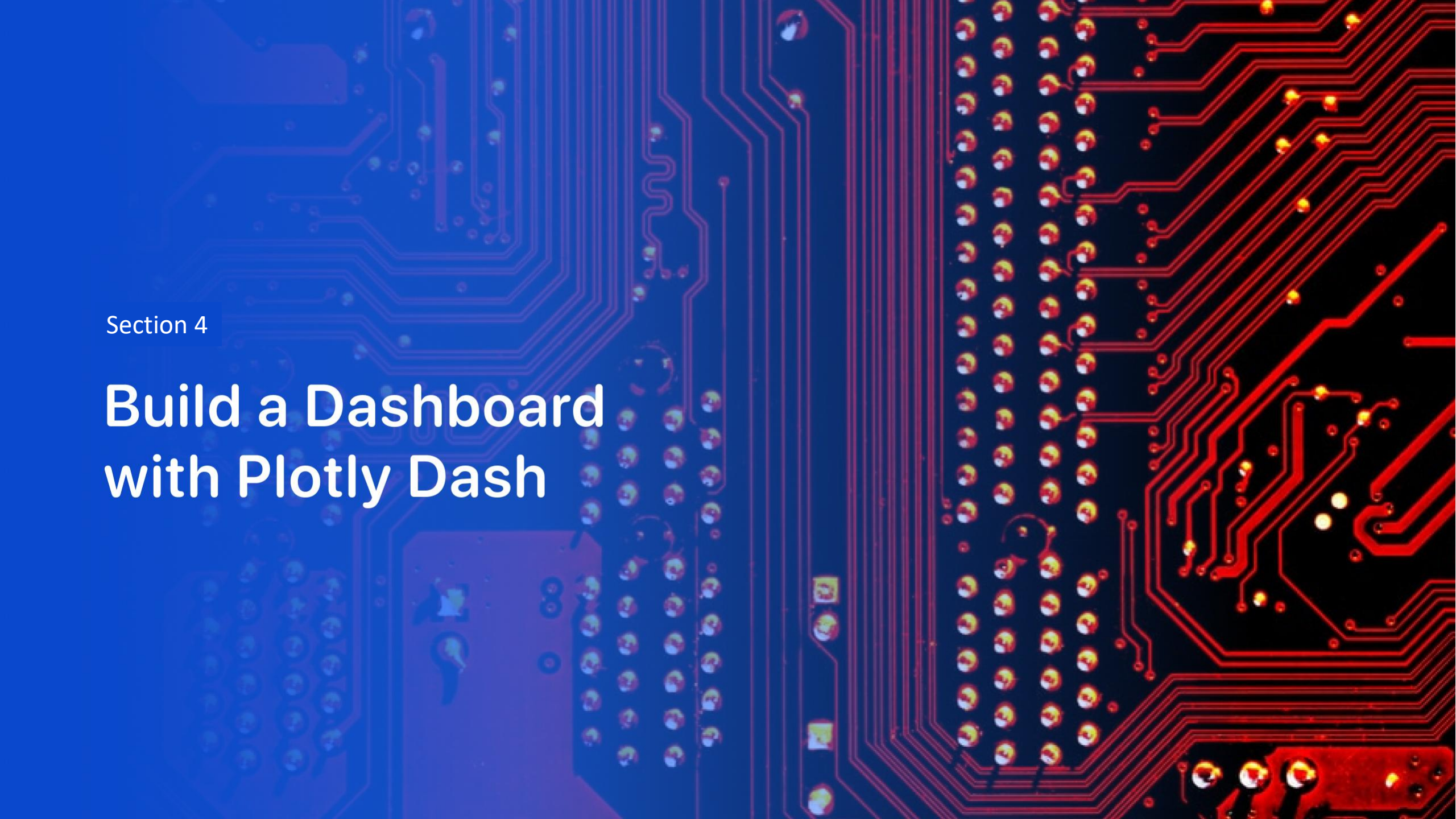
Distances between a launch site and its proximities

- Explore the generated folium map and show the screenshot of a selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed



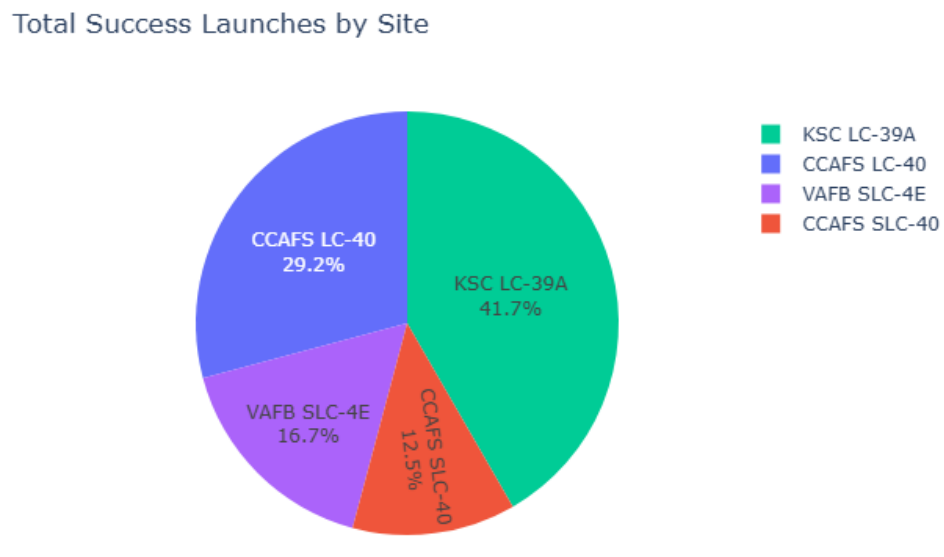


- Explain the important elements and findings on the screenshot
  - The launch site VAFB SLC-4E is located in California with lattitude 34.632834 and longitude -120.610745
  - The coastline is 1.4km away to west
  - The nearest airport in Lompoc is around 14km from the launch site in north east direction


Section 4
Build a Dashboard with Plotly Dash

SpaceX Launch Records Dashboard

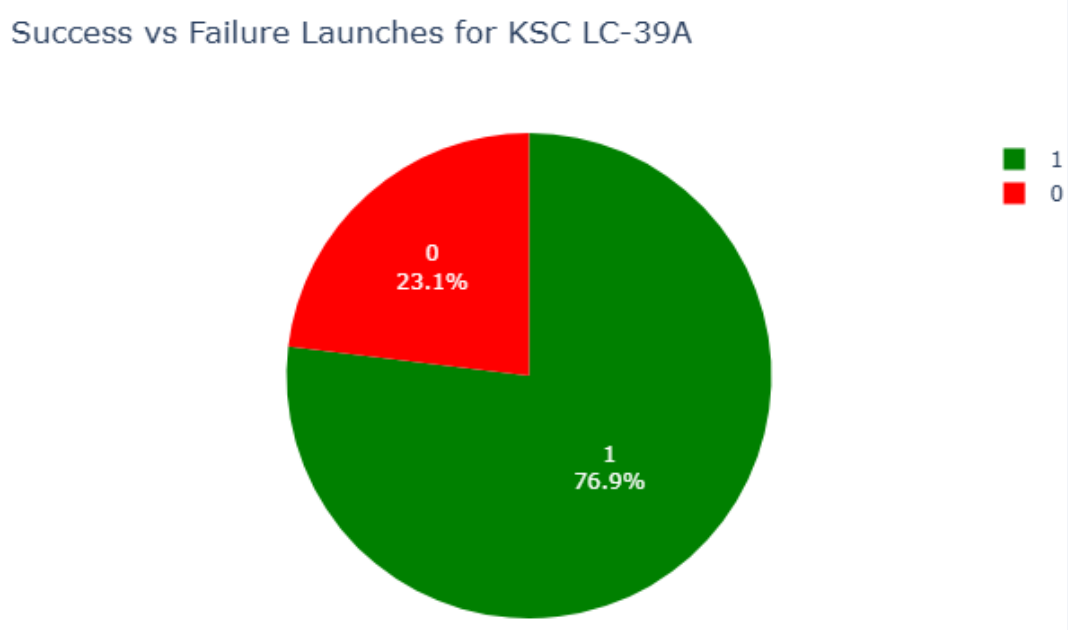
- Show the screenshot of launch success count for all sites, in a piechart



- Explain the important elements and findings on the screenshot
  - The launch site KSC LC-39A has the most successful rocket launches, with a success rate more than 40%
  - Rocket launches from launch site CCAFS SLC-40 are least successful, with only 12.5% success rate

Launch site with highest Launch Success Ratio

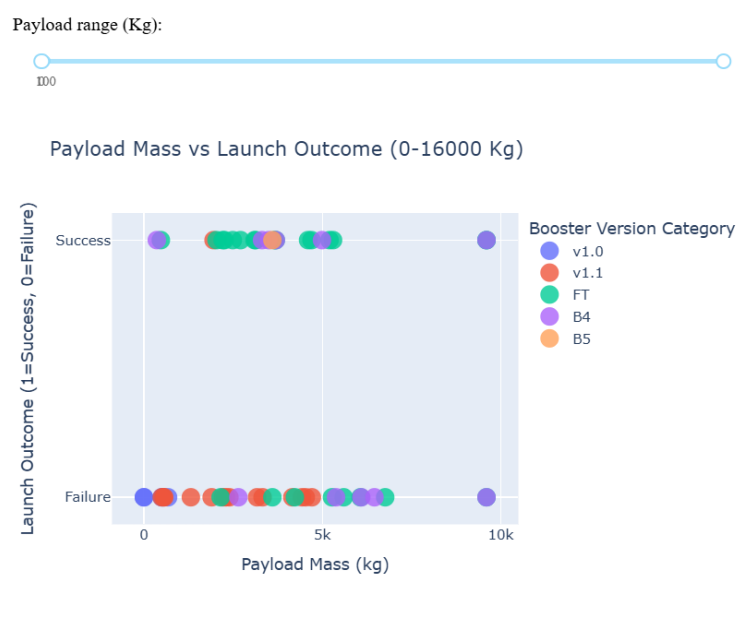
- Show the screenshot of the piechart for the launch site with highest launch success ratio



- Explain the important elements and findings on the screenshot
  - The launch site KSC LC-39A has a launch success ratio of around 77%

Influence of Payload Mass in Lauch Outcome

- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider



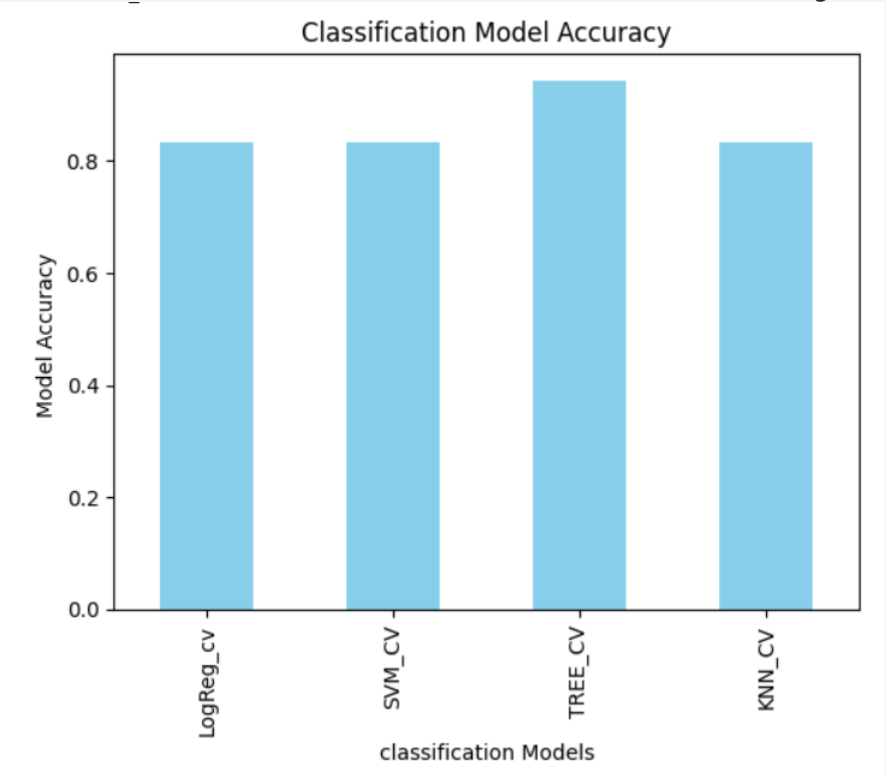
- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.
  - The Booster Version FT has the highest success rate

Section 5

Predictive Analysis (Classification)

Classification Accuracy

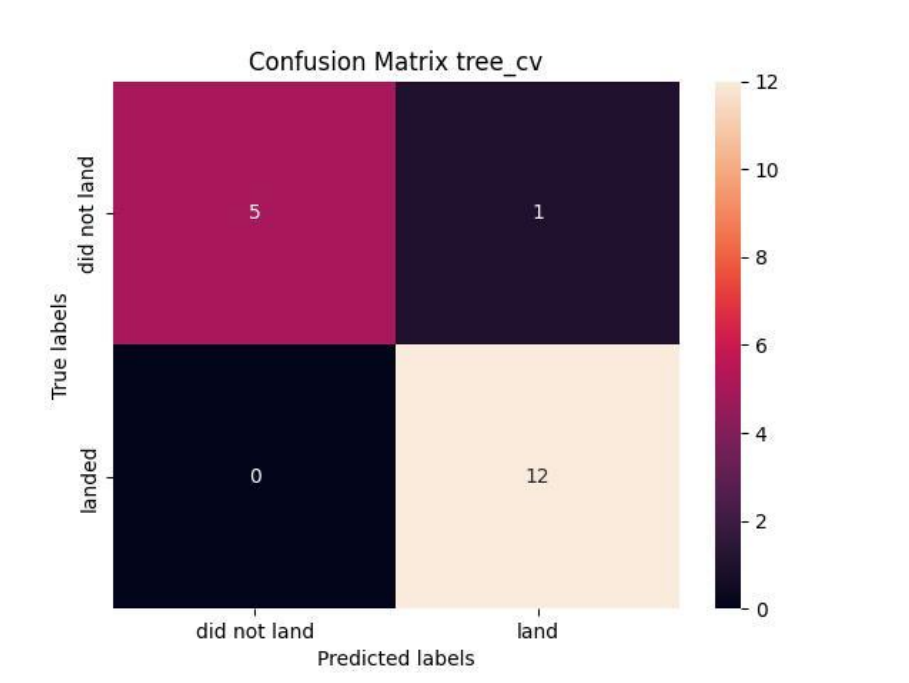
- Visualize the built model accuracy for all built classification models, in a bar chart



- Find which model has the highest classification accuracy
  - Decision Tree Classifier has the highest model accuracy

Confusion Matrix

- Show the confusion matrix of the best performing model with an explanation



- Decision Tree classifier has the best prediction accuracy of 94% with only one false positive

Conclusions

- The necessary data for the Model were collected from SpaceX API and also from Wikipedia
- Good Insights could be gained from the collected data using EDA
  - Around 100% success rate at launch site CCSFS SLC 40 for flight numbers > 80
  - No rockets were launched from VAFB SLC-4E with payload mass > 10000kg
  - Almost 100% success rate for the launches to the orbits ES-L1, GEO, HEO and SSO
- The rockets were launched from four unique sites, which could be visualized using Folium
- Success rate of each launch site could be visualized in Dash boards with the help of Dropdown lists and Pie Charts
- The success rate of each rocket launch could be predicted with linear regression models
  - Decision Tree classifier has the best prediction accuracy of 94% with only one false positive

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!