IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Mark O. Worthington
PMP, SSBB, CEF, CSM

01/05/2025

Outline

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

Executive Summary

- Summary of methodologies

  - Data Collection through API (Space X)

  - Data Collection with Web Scraping (Wikipedia)

  - Data Wrangling

  - Exploratory Data Analysis with SQL

  - Exploratory Data Analysis with Data Visualization

  - Visual Analytics with Folium

  - Interactive Visual Analytics with Plotly

  - Machine Learning Prediction

- Summary of all results

  - Exploratory Data Analysis Results

  - Interactive Visual Analytics in Screenshots

  - Predictive Analytics Results

Introduction

- Project background and context

SpaceX has gained worldwide attention for a series of historic milestones. It is the only private company ever to return a spacecraft from low-earth orbit, which it first accomplished in December 2010. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars whereas other providers cost upward of 165 million dollars each, much of the savings is because Space X 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. This dataset includes a record for each payload carried during a SpaceX mission into outer space.

- Problems you want to find answers

- What factors will determine if the first stage rocket booster will land successfully

- What are the interactions between variables that will help influence the success rate of a landing

- What features need to be in place to determine consistent successful landings

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data collected using Spacex API calls
  - Data collected using web scraping of a SpaceX Wikipedia page
- Perform data wrangling
  - One-hot encoding was applied to categorical features
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Data sets generated above were used to investigate 4 different Classification Models.. The most accurate Model was identified

Space X Falcon 9 First Stage Landing Prediction

Data Collection – SpaceX API Calls

- In this capstone, we will predict if the Falcon 9 first stage will land successfully. 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. In this lab, you will collect and make sure the data is in the correct format from an API.

Most unsuccessful landings are planned. Space X performs a controlled landing in the oceans.

Objectives

In this lab, you will make a get request to the SpaceX API. You will also do some basic data wrangling and formatting.

- Request to the SpaceX API
- Clean the requested data

Data Collection – SpaceX API Calls

🚀 Open Source REST API for SpaceX launch, rocket, core, capsule, launchpad, and landing pad data.

Our Objective in this lab is to Collect and Condition data from the Spacex website for a SpaceX Falcon 9 first stage Landing Prediction and to produce a data set that will be used in the data collection module.

- We will make 6 calls [GET] to the API at -http://api.spacexdata.com/v4/~
  - Our first API [GET] call to - spacex_url=https://api.spacexdata.com/v4/launches/past will show us the response information about past SpaceX launches. This information may be dynamic.
  - Our second API [GET] call to -static_json_url='https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json will provide us with static consistent data (w/response code) that can be used to evaluate our responses for the course. The resulting data is converted to a data frame for cleaning and conditioning.
  - Our third API [GET] call is made via a function to - https://api.spacexdata.com/v4/rockets/ and will provide a list of booster names that will be appended to our data set.
  - Our fourth API [GET] call is made via a function to - https://api.spacexdata.com/v4/launchpads/ and will provide a list of launch site locations (longitude and Latitude). This will be appended to our data set.
  - Our fifth API [GET] call is made via a function to - https://api.spacexdata.com/v4/payloads/ and will provide a list of associated payload mass and orbits. This will be appended to our data set.
  - Our sixth API [GET] call is made via a function to - https://api.spacexdata.com/v4/cores/ and will provide a list of associated information that will be appended to our data set. (i.e. block, reused count, serial, outcome, flights, gridfins, reused, legs, landing pad)

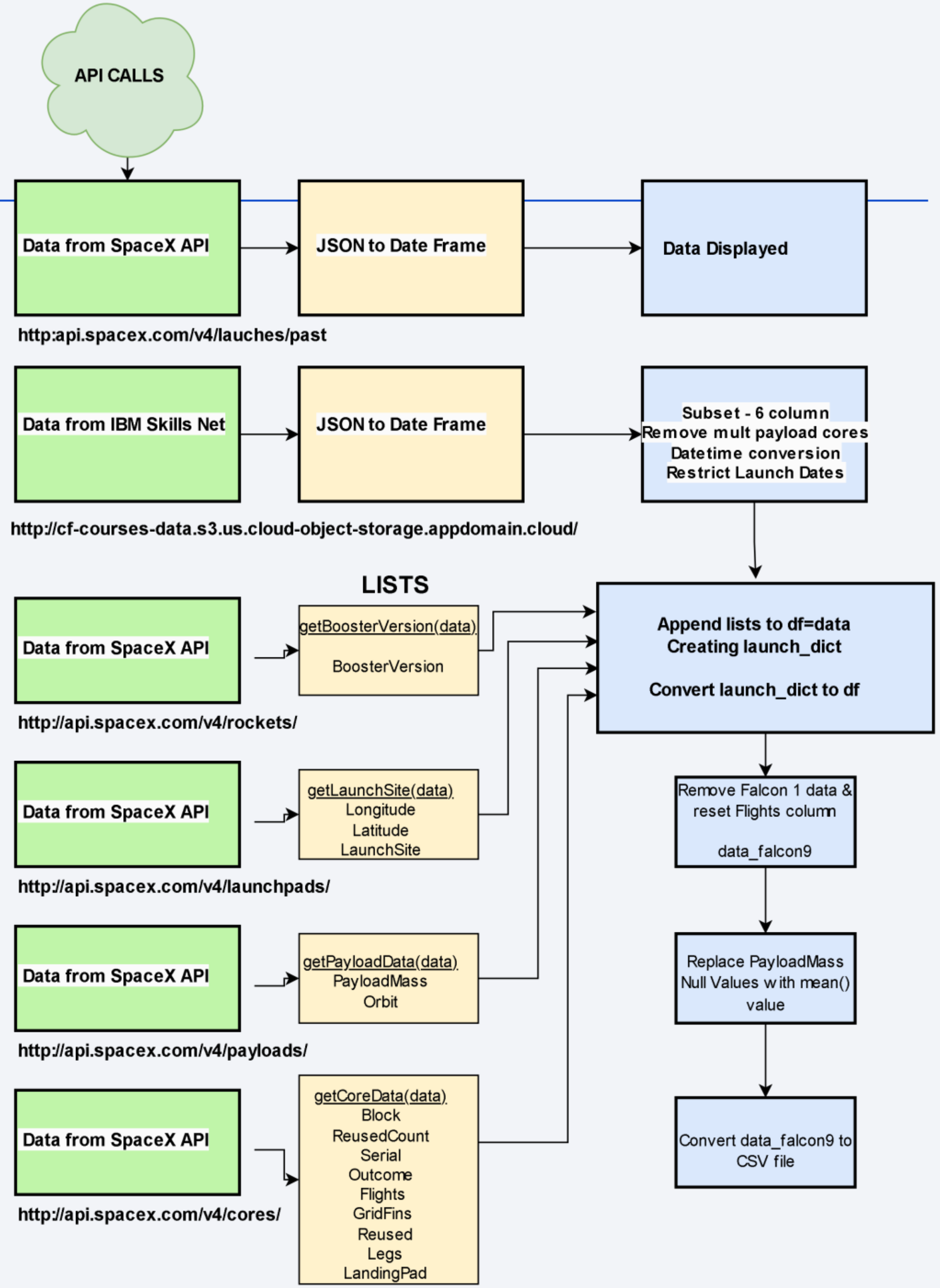
Data Collection – SpaceX API Calls

SpaceX API Calls to Final Data Set

SpaceX API calls Notebook – The data collection workspace is completed and stored at - https://github.com/flurpo/IBMspx/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

(This file can be uploaded to the Coursera workspace for Module 1A for review, it can also be reviewed in google colab)

The Final Data Set from the data collection workspace is stored at - https://github.com/flurpo/IBMspx/blob/main/Data/1A_dataset_part_1.csv

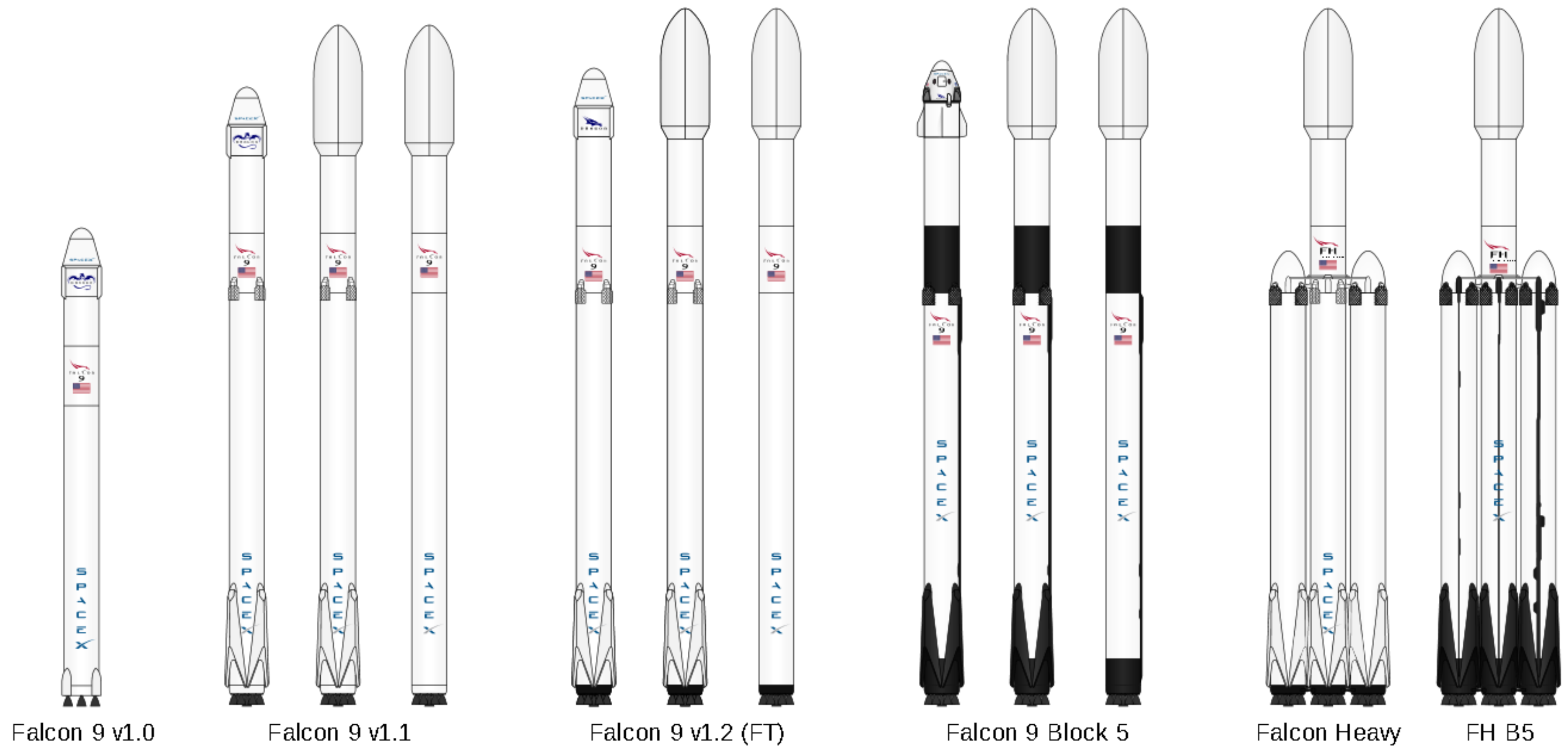


Space X Falcon 9 First Stage Landing Prediction

Data Collection – Web Scraping

- In this lab, you will be performing web scraping to collect Falcon 9 historical launch records from a Wikipedia page titled `List of Falcon 9 and Falcon Heavy launches`

https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches



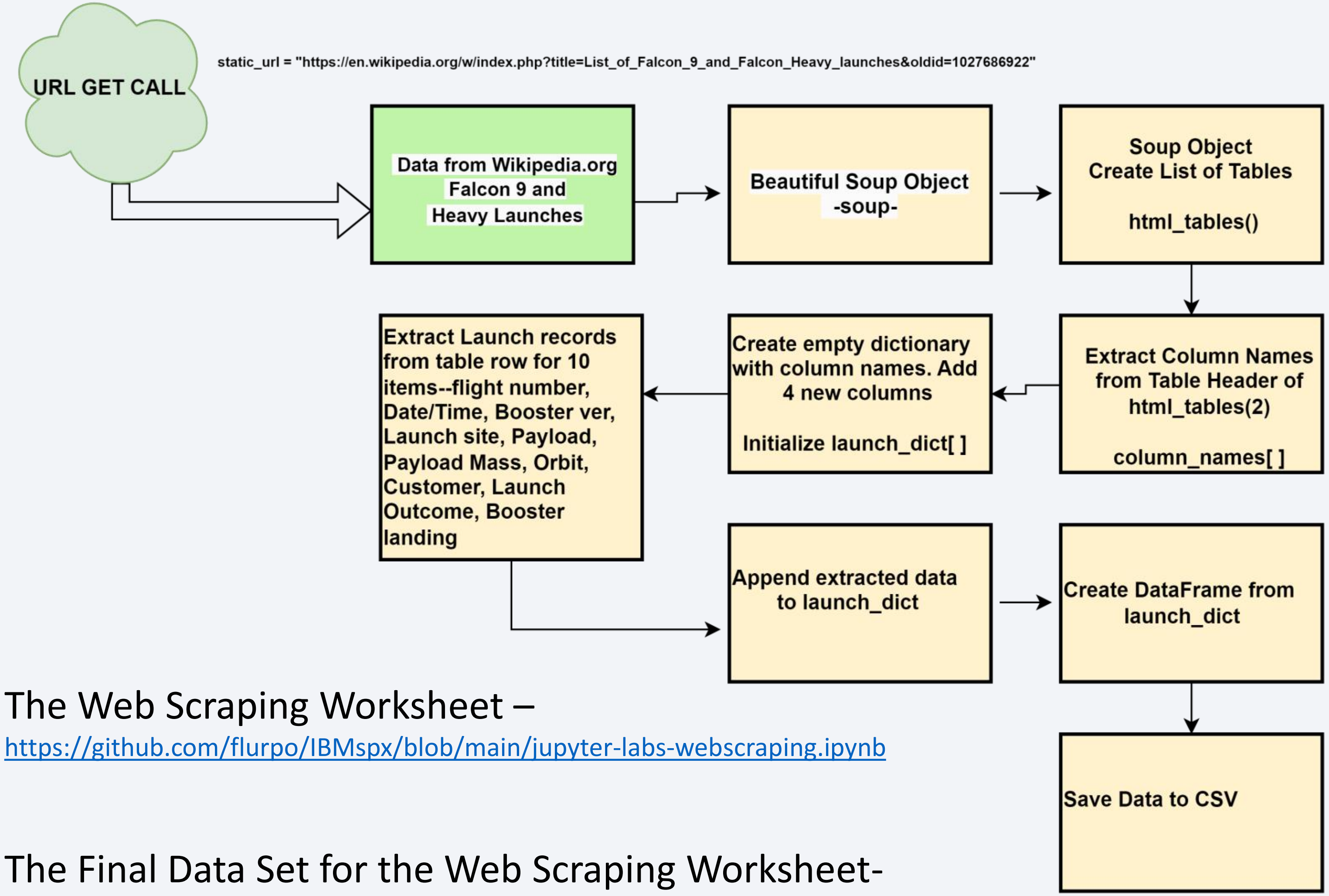
Objectives

Web scrap Falcon 9 launch records with `BeautifulSoup`:

- Extract a Falcon 9 launch records HTML table from Wikipedia
- Parse the table and convert it into a Pandas data frame

Data Collection – Web Scraping

Web Scraping Falcon 9 records from Wikipedia with Beautiful Soup




```
1. Start
2. Fetch HTML from URL
3. Parse HTML using BeautifulSoup
4. Find all tables in the HTML
5. Select the relevant table
6. Extract column headers
7. Initialize launch_dict
8. For each row in the table:
- Check if row is a launch entry
- Extract data:
- Flight Number
- Date and Time
- Version Booster
- Launch Site
- Payload
- Payload Mass
- Orbit
- Customer
- Launch Outcome
- Booster Landing
- Append extracted data to launch_dict
9. Create DataFrame from launch_dict
10. Convert numeric columns to float
11. Save DataFrame to CSV
12. End
```

The Web Scraping Worksheet –
https://github.com/flurpo/IBMspx/blob/main/jupyter-labs-webscraping.ipynb

The Final Data Set for the Web Scraping Worksheet-
https://github.com/flurpo/IBMspx/blob/main/Data/1B_spacex_web_scraped.csv

SpaceX Flacon 9 First Stage Landing Prediction

Data Wrangling

In this lab, we will perform some Exploratory Data Analysis (EDA) to find some patterns in the data and determine what would be the label for training supervised models. In the data set, there are several different cases where the booster did not land successfully. Sometimes a landing was attempted but failed due to an accident; for example, True Ocean means the mission outcome was successfully landed to a specific region of the ocean while False Ocean means the mission outcome was unsuccessfully landed to a specific region of the ocean. True RTLS means the mission outcome was successfully landed to a ground pad False RTLS means the mission outcome was unsuccessfully landed to a ground pad.True ASDS means the mission outcome was successfully landed on a drone ship False ASDS means the mission outcome was unsuccessfully landed on a drone ship.

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0003	-80.577366	28.561857
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0005	-80.577366	28.561857
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0007	-80.577366	28.561857
3	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN	1.0	0	B1003	-120.610829	34.632093
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1004	-80.577366	28.561857
5	6	2014-01-06	Falcon 9	3325.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1005	-80.577366	28.561857
6	7	2014-04-18	Falcon 9	2296.000000	ISS	CCAFS SLC 40	True Ocean	1	False	False	True	NaN	1.0	0	B1006	-80.577366	28.561857
7	8	2014-07-14	Falcon 9	1316.000000	LEO	CCAFS SLC 40	True Ocean	1	False	False	True	NaN	1.0	0	B1007	-80.577366	28.561857
8	9	2014-08-05	Falcon 9	4535.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1008	-80.577366	28.561857
9	10	2014-09-07	Falcon 9	4428.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1011	-80.577366	28.561857

In this lab we will mainly convert those outcomes into Training Labels with `1` means the booster successfully landed `0` means it was unsuccessful.

Objectives

Perform exploratory Data Analysis and determine Training Labels

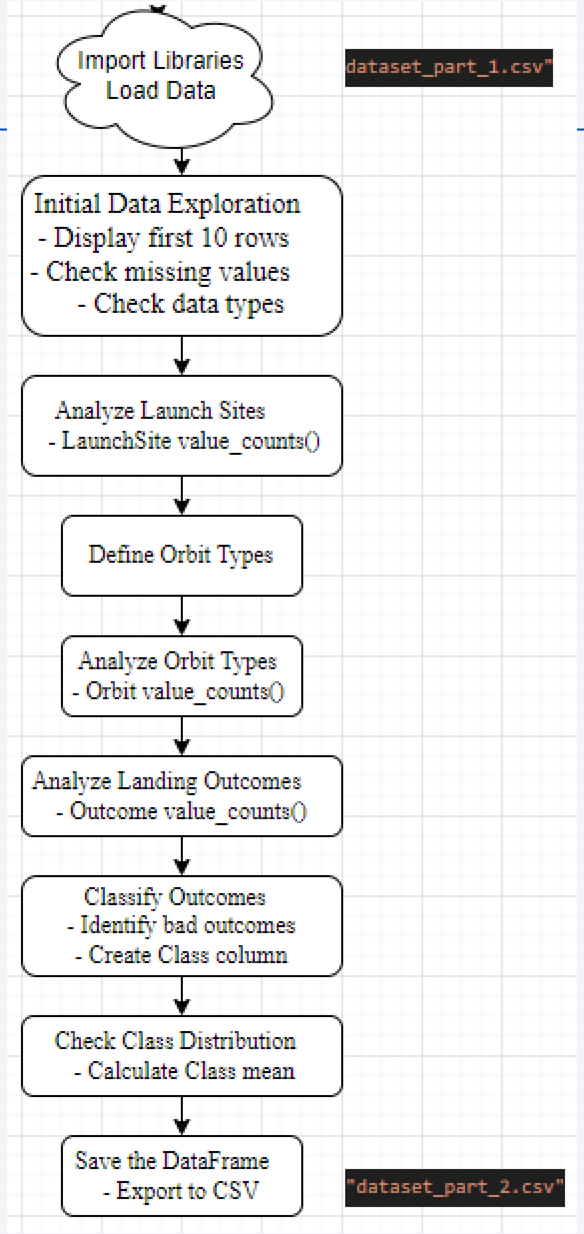
Data Wrangling

- Exploratory Data Analysis, Determine Training labels
  - *See the Data Wrangling flow chart right*

- Task 1 – Calculated the number of launches on each site
- Task 2 – Calculated the number and occurrence of each orbit
- Task 3 – Calculated the number and occurrence of mission outcome of the orbits
- Task 4 – Created a landing outcome label from the Outcome column

The Data Wrangling Worksheet –
https://github.com/flurpo/IBMspx/blob/main/labs-jupyter-spacex-Data wrangling.ipynb

The Data Set Generated from the Data Wrangling Worksheet –
https://github.com/flurpo/IBMspx/blob/main/Data/dataset_part_2.csv



Exploratory Data Analysis with SQL

EDA with SQL

Overview of the Data Set

https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/labs/module_2/data/Spacex.csv

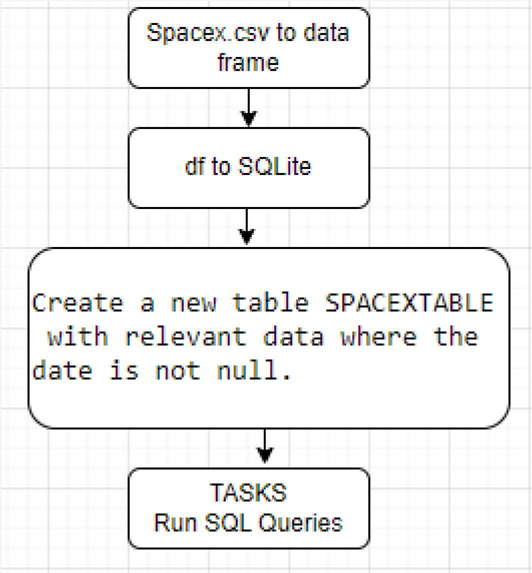
.

	Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
0	2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
1	2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of...	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2	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
3	2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
4	2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt
5	2013-09-29	16:00:00	F9 v1.1 B1003	VAFB SLC-4E	CASSIOPE	500	Polar LEO	MDA	Success	Uncontrolled (ocean)
6	2013-12-03	22:41:00	F9 v1.1	CCAFS LC-40	SES-8	3170	GTO	SES	Success	No attempt
7	2014-01-06	22:06:00	F9 v1.1	CCAFS LC-40	Thaicom 6	3325	GTO	Thaicom	Success	No attempt
8	2014-04-18	19:25:00	F9 v1.1	CCAFS LC-40	SpaceX CRS-3	2296	LEO (ISS)	NASA (CRS)	Success	Controlled (ocean)
9	2014-07-14	15:15:00	F9 v1.1	CCAFS LC-40	OG2 Mission 1 6 Orbcomm-OG2 satellites	1316	LEO	Orbcomm	Success	Controlled (ocean)
10	2014-08-05	8:00:00	F9 v1.1	CCAFS LC-40	AsiaSat 8	4535	GTO	AsiaSat	Success	No attempt
11	2014-09-07	5:00:00	F9 v1.1 B1011	CCAFS LC-40	AsiaSat 6	4428	GTO	AsiaSat	Success	No attempt
12	2014-09-21	5:52:00	F9 v1.1 B1010	CCAFS LC-40	SpaceX CRS-4	2216	LEO (ISS)	NASA (CRS)	Success	Uncontrolled (ocean)
13	2015-01-10	9:47:00	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2395	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)
14	2015-02-11	23:03:00	F9 v1.1 B1013	CCAFS LC-40	DSCOVR	570	HEO	U.S. Air Force NASA NOAA	Success	Controlled (ocean)
15	2015-03-02	3:50:00	F9 v1.1 B1014	CCAFS LC-40	ABS-3A Eutelsat 115 West B	4159	GTO	ABS Eutelsat	Success	No attempt
16	2015-04-14	20:10:00	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6	1898	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)
17	2015-04-27	23:03:00	F9 v1.1 B1016	CCAFS LC-40	Turkmen 52 / MonacoSAT	4707	GTO	Turkmenistan National Space Agency	Success	No attempt
18	2015-06-28	14:21:00	F9 v1.1 B1018	CCAFS LC-40	SpaceX CRS-7	1952	LEO (ISS)	NASA (CRS)	Failure (in flight)	Precluded (drone ship)
19	2015-12-22	1:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

EDA with SQL

- Summary of SQL Queries

- Task 1: Query for unique launch sites.
- Task 2: Query for records with launch sites that start with 'CCA'.
- Task 3: Calculate the total payload mass for NASA (CRS) missions.
- Task 4: Calculate the average payload mass for booster version F9 v1.1.
- Task 5: Use MIN() to find the earliest successful ground pad landing.
- Task 6: Query for boosters that succeeded in drone ship landings and had payloads between 4000 and 6000 kg.
- Task 7: Count successful, failed, and other mission outcomes.
- Task 8: Use a subquery to find the booster version that carried the maximum payload mass.
- Task 9: Use SUBSTR() to get month names for landing failures in 2015.
- Task 10: Count and rank landing outcomes in a specific date range.



The EDA with SQL Worksheet – https://github.com/flurpo/IBMspx/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Space X Falcon 9 First Stage Landing Prediction

EDA with Data Visualization

In this assignment, we will predict if the Falcon 9 first stage will land successfully. 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 due to the fact that SpaceX can reuse the first stage.In this lab, you will perform Exploratory Data Analysis and Feature Engineering.

Objectives

Perform exploratory Data Analysis and Feature Engineering using `Pandas` and `Matplotlib`

- Exploratory Data Analysis
- Preparing Data - Feature Engineering

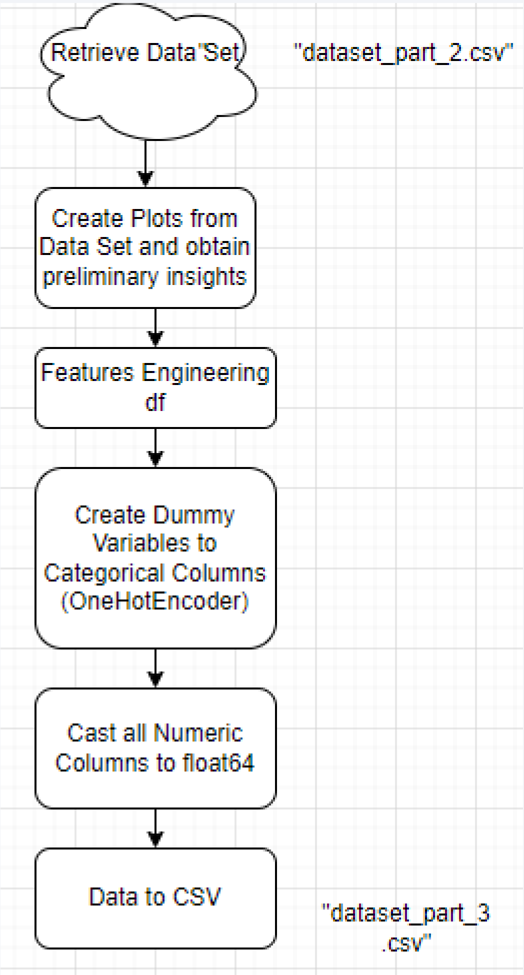
EDA with Data Visualization

Summary of the charts plotted

- Flight Number vs Payload Mass – to see how FlightNumber and Payload variables would effect the Launch outcome.
- Task 1.. Flight number vs Launch Site – to see how this relationship would effect the Launch outcome.
- Task 2.. Visualize the relationship between Payload Mass and Launch Site
- Task 3.. Visualize the relationship between Success rate and Orbit type
- Task 4.. Visualize the relationship between FlightNumber and Orbit type
- Task 5.. Visualize the relationship between Payload Mass and Orbit type
- Task 6.. Visualize the launch success yearly trend

EDA with Data Visualization Worksheet –
https://github.com/flurpo/IBMspx/blob/main/edadataviz.ipynb

The Data Set Generated from the Data Visualization Worksheet –
https://github.com/flurpo/IBMspx/blob/main/Data/dataset_part_3.csv



Hands-on Lab: Interactive Visual Analytics

Building an Interactive Map with Folium

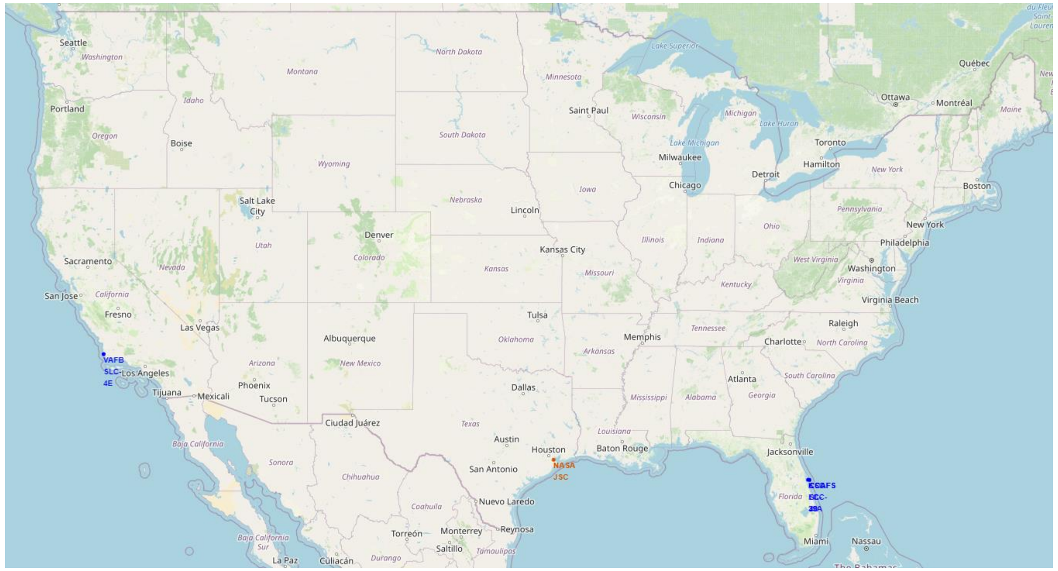
The launch success rate may depend on many factors such as payload mass, orbit type, and so on. It may also depend on the location and proximities of a launch site, i.e., the initial position of rocket trajectories. Finding an optimal location for building a launch site certainly involves many factors and hopefully we could discover some of the factors by analyzing the existing launch site locations.

Objectives

This lab contains the following tasks:

- **TASK 1:** Mark all launch sites on a map
- **TASK 2:** Mark the success/failed launches for each site on the map
- **TASK 3:** Calculate the distances between a launch site to its proximities

After completed the above tasks, you should be able to find some geographical patterns about launch sites.



Building an Interactive Map with Folium

In the Python Worksheet the following items were added to the Folium Map Generated

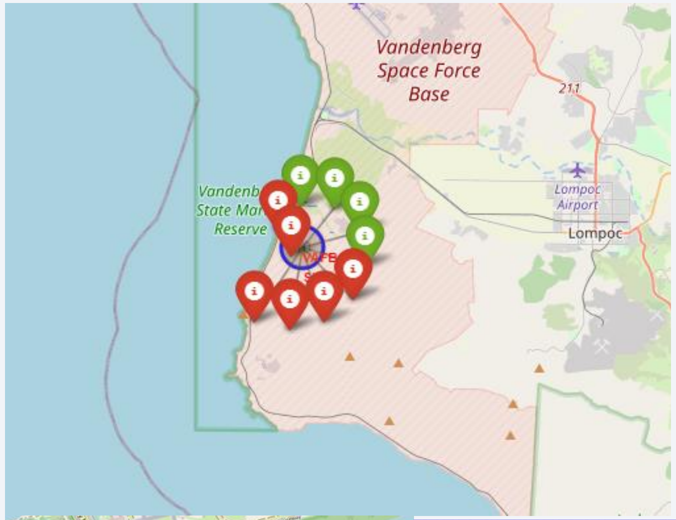
1. Pulled data from spacex_launch_geo.csv
2. Selected sub columns Launch Site, Latitude, Longitude and class
3. Create data frames spacex_df(launch site, Lat, Long, class)(56) and launch_sites_df(launch site, Lat, Long) (4)
4. NOTE- 4 launch sites, three in Florida and one in California

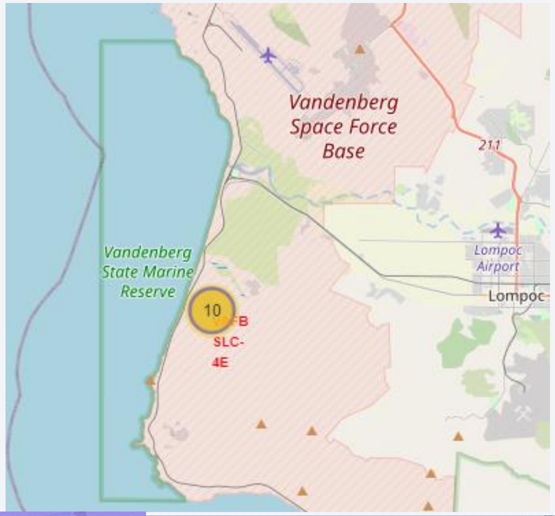
5. Created a start location at NASA Johnson Sace Center
6. Created a blue circle at NASA Johnson Space Center's coordinate with a popup label and an icon showing its name
7. Created an icon as a text label

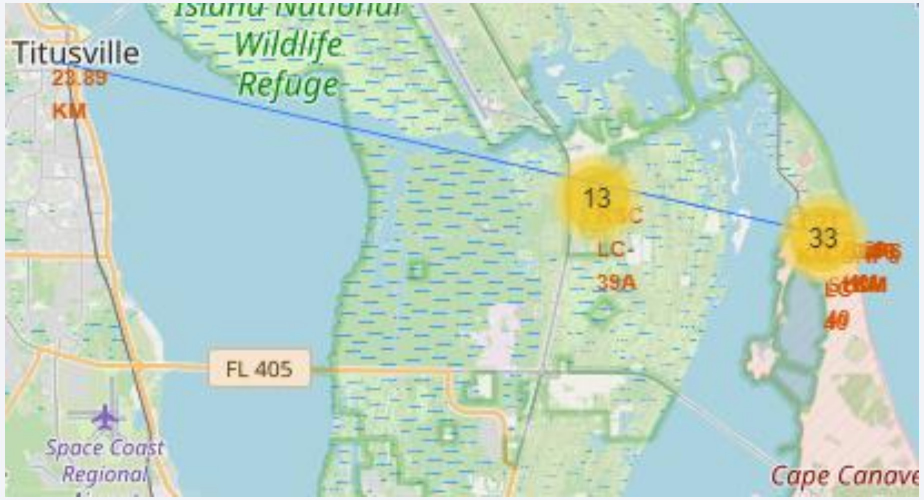
8. Added a Circle and Marker to the map for each launch site (4) with the launch site name as a popup label
9. Added markers (Green, Red) to mark the success/failed launches at each site
10. Chose a launch site and calculated distance to nearest coastline, city(titusville), railway and highway
11. Added distance markers and lines

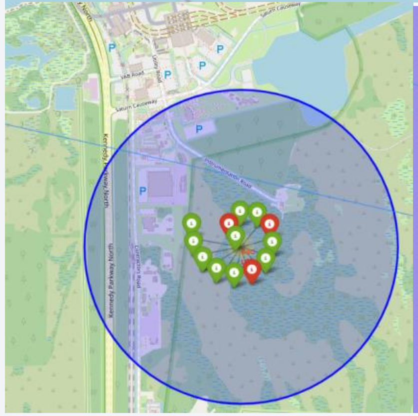
The Folium Worksheet -https://github.com/flurpo/IBMspx/blob/main/lab_jupyter_launch_site_location.ipynb

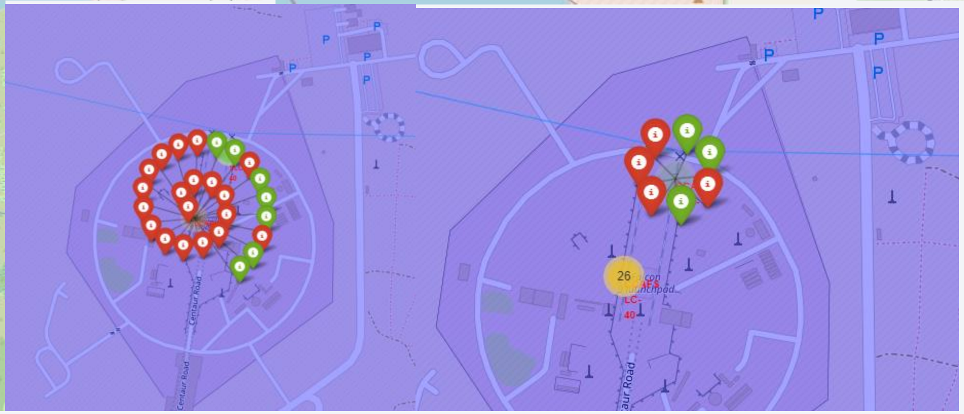
Building an Interactive Map with Folium











	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

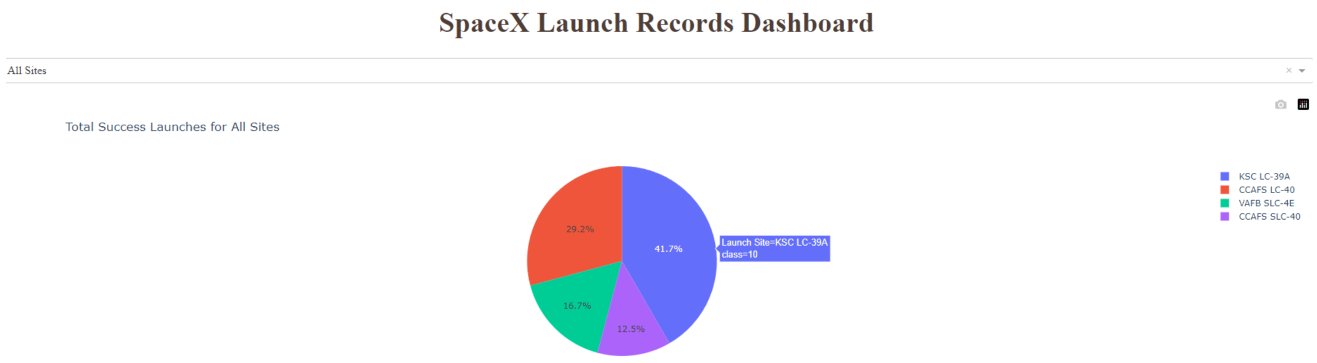
- Stand Alone Folium File (HTML) for inspection in a browser - https://github.com/flurpo/IBMspx/blob/main/Data/site_map.html

Hands-on Lab: Build an Interactive Dashboard with Ploty Dash

Building a Plotly Dash application for users to perform interactive visual analytics on SpaceX launch data in real-time.

This dashboard application contains input components such as a dropdown list and a range slider to interact with a pie chart and a scatter point chart. You will be guided to build this dashboard application via the following tasks:

- TASK 1: Add a Launch Site Drop-down Input Component
- TASK 2: Add a callback function to render `success-pie-chart` based on selected site dropdown
- TASK 3: Add a Range Slider to Select Payload
- TASK 4: Add a callback function to render the `success-payload-scatter-chart` scatter plot



Building an Interactive Dashboard with Plotly Dash

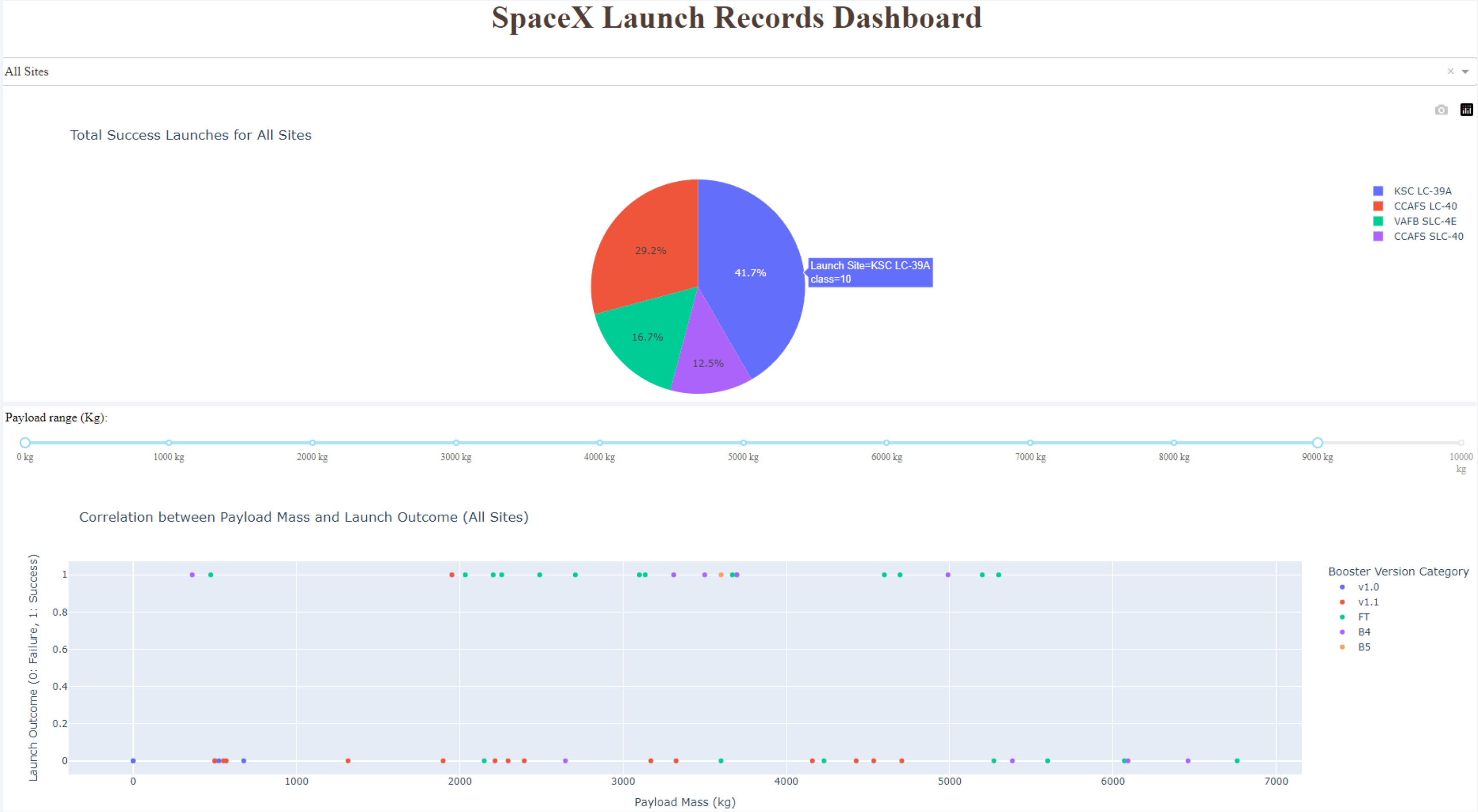
Plots/graphs and interactions added to the Plotly dashboard

1. Setup an ngrok tunnel at port '8050’
2. Loaded data from spacex_launch_dash.cvs to a data frame - spacex_df
3. Created options for a dropdown ("All Sites', unique launch sites)
4. Created the dash application and layout
5. Added dropdown list for launch site selection
6. Added pie chart to show successful launches for all sites/specific site
7. Added slider to select payload range for scatter chart
8. Added scatter chart to show correclation between payload and launch success
9. Added a callback function for \`site-dropdown\` as input, \`success-pie-chart\` as output
10. Added a callback function for \`site-dropdown\` as input, \`success-pie-chart\` as output

The Plotly Dashboard Notebook -https://github.com/flurpo/IBMspx/blob/main/spacex_dash_app_20240914.ipynb

The Plotly Dashboard App to run as a standalone Python application through a browser -
https://github.com/flurpo/IBMspx/blob/main/Python Notebooks/spacex_dash_app.py

Building an Interactive Dashboard with Plotly Dash



PDF of Plotly Dashboard -
https://github.com/flurpo/IBMspx/blob/main/Plotly_spacex_dash_app_20240914.pdf

Space X Falcon 9 First Stage Landing Prediction

Predictive Analysis (Classification)

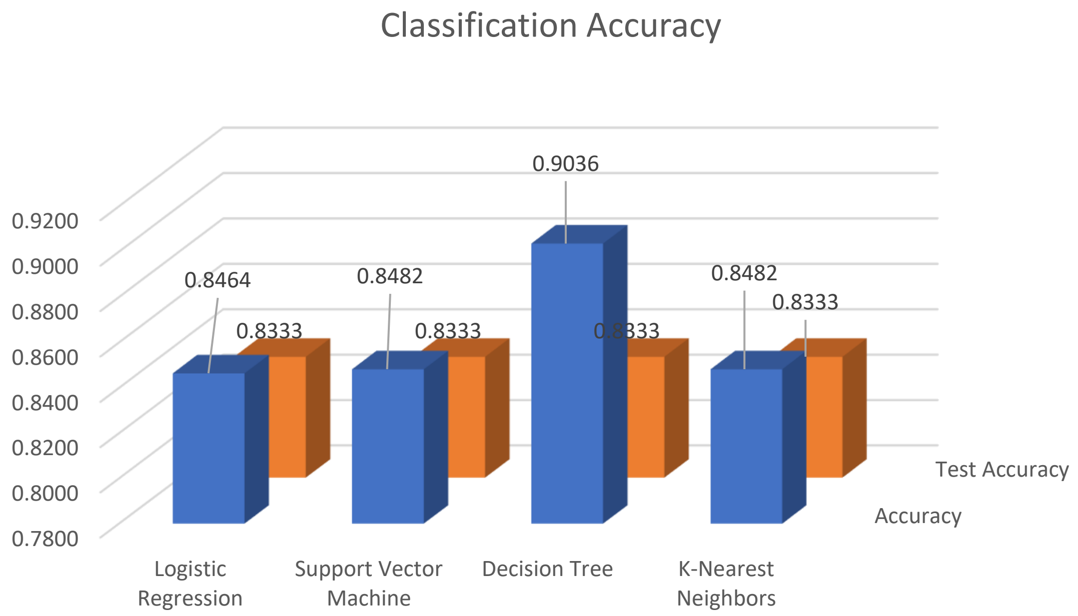
Objectives

Perform exploratory Data Analysis and determine Training Labels

- create a column for the class
- Standardize the data
- Split into training data and test data

Find best Hyperparameter for SVM, Classification Trees and Logistic Regression

- Find the method that performs best using test data



Predictive Analysis (Classification)

I built, evaluated, improved, and found the best performing classification model using the following steps

1. Created a confusion matrix plot function
2. Loaded dataset_part_2.csv to dataframe "data"
3. Loaded dataset_part_3.csv to dataframe "X"
4. Created a Training data set and a Testing data set
5. Used LOGISTIC REGRESSION estimation on data - calculated hyperparameters, calculated accuracy, ploted confusion matrix
6. Used SUPPORT VECTOR MACHINE estimation on data - calculated hyperparameters, calculated accuracy, ploted confusion matrix
7. Used DECISION TREE CLASSIFIER estimation on data - calculated hyperparameters, calculated accuracy, ploted confusion matrix
8. Used KNEIGHBORS CLASSIFIER estimation on data - calculated hyperparameters, calculated accuracy, ploted confusion matrix
9. Created a Summary Generator for the resulting data

Predictive Analysis Worksheet -https://github.com/flurpo/IBMspx/blob/main/SpaceX_Machine Learning Prediction_Part_5.ipynb

Results

Predictive analysis results

Calculated Accuracy Results for Logistic Regression
tuned hyperparameters :(best parameters) {'C': 0.01, 'penalty': 'l2', 'solver': 'lbfgs'}
accuracy : 0.8464285714285713
Logistic Regression Test Accuracy: 0.8333333333333334
Calculated Accuracy Results for Support Vector Machine
tuned hyperparameters :(best parameters) {'C': 1.0, 'gamma': 0.03162277660168379, 'kernel': 'sigmoid'}
accuracy : 0.8482142857142856
Support Vector Machine Test Accuracy: 0.8333333333333334
Calculated Accuracy Results for Decision Tree
tuned hyperparameters :(best parameters) {'criterion': 'entropy', 'max_depth': 4, 'max_features': 'log2', 'min_samples_leaf': 1, 'min_samples_split': 2, 'splitter': 'best'}
accuracy : 0.9035714285714287
Decision Tree Test Accuracy: 0.8333333333333334
removed 'auto' from 'max_features' in parameters - not a valid option in DecisionTreeClassifier
Calculated Accuracy Results for K-Nearest Neighbors
tuned hyperparameters :(best parameters) {'algorithm': 'auto', 'n_neighbors': 10, 'p': 1}
accuracy : 0.8482142857142858
K-Nearest Neighbors Test Accuracy: 0.8333333333333334

IN SUMMARY:
The test data accuracy and confusion matrix response was identical across the model results. So, its very difficult to chose a best model based on the test accuracy.
SVM is likely to perform the best overall on this dataset due to its strength in handling high-dimensional data and non-linear relationships.
However, Logistic Regression could also be a strong contender if the relationships between the features and the target are mostly linear.
KNN should be the worst performer as it is less effective with High-dimensional data.

CONCLUSION:
Although all of the models had identical test accuracy, the DECISION TREE CLASSIFIER has a higher fit accuracy - 0.9036 and better confusion matrix response.
It appears to be our best model with this data set.

Results

Exploratory Data Analysis Results

Summary of SQL Queries

1: Query for unique launch sites.

2: Query for records with launch sites that start with 'CCA'.

3: Calculate the total payload mass for NASA (CRS) missions

4: Calculate the average payload mass for booster version F9 v1.1.

5: Use MIN() to find the earliest successful ground pad landing

6: Query for boosters that succeeded in drone ship landings and had payloads between 4000 and 6000 kg.

7: Count successful, failed, and other mission outcomes.

8: Use a subquery to find the booster version that carried the maximum payload mass.

9: Use SUBSTR() to get month names for landing failures in 2015.

10: Count and rank landing outcomes in a specific date range.

Summary of the charts plotted

Flight Number vs Payload Mass – to see how FlightNumber and Payload variables would effect the Launch outcome.

1.. Flight number vs Launch Site – to see how this relationship would effect the Launch outcome.

2.. Visualize the relationship between Payload Mass and Launch Site

3.. Visualize the relationship between Success rate and Orbit type

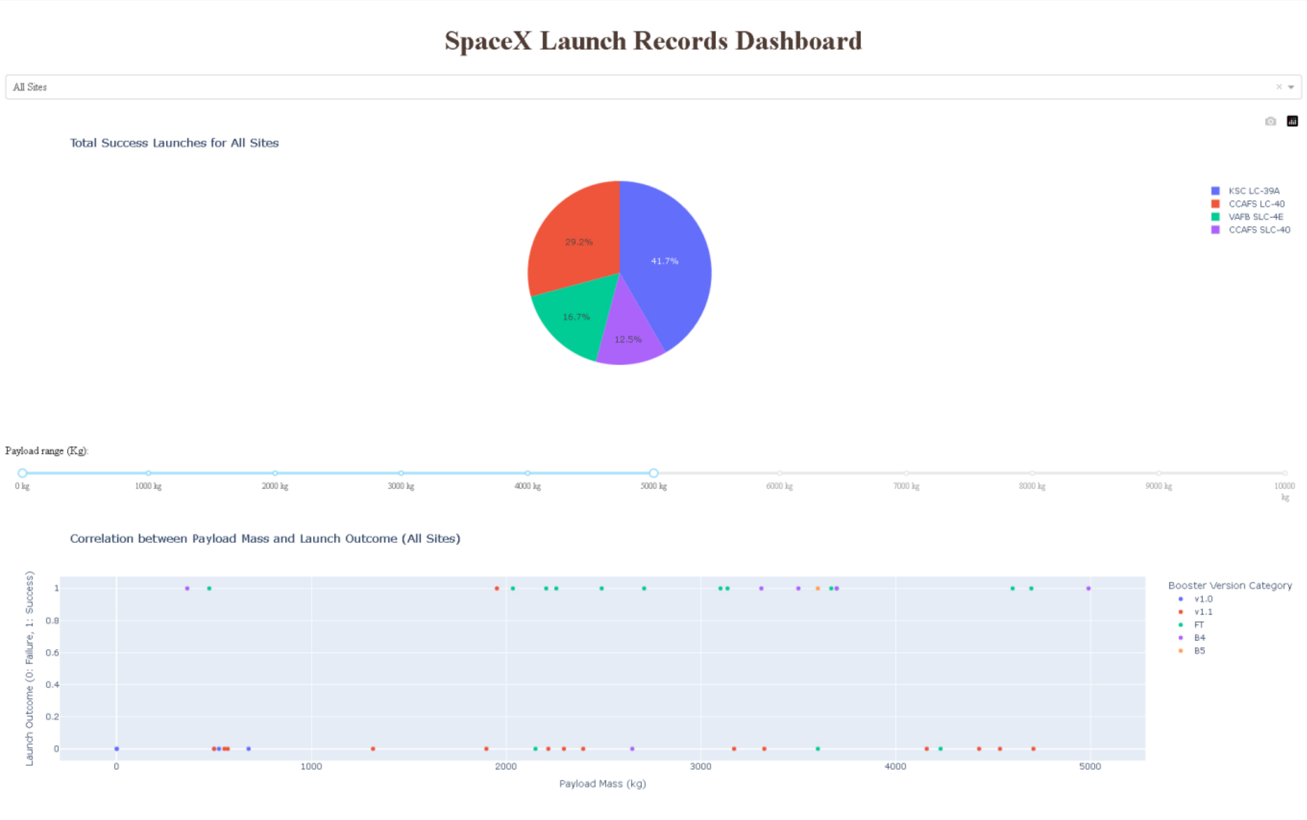
4.. Visualize the relationship between FlightNumber and Orbit type

5.. Visualize the relationship between Payload Mass and Orbit type

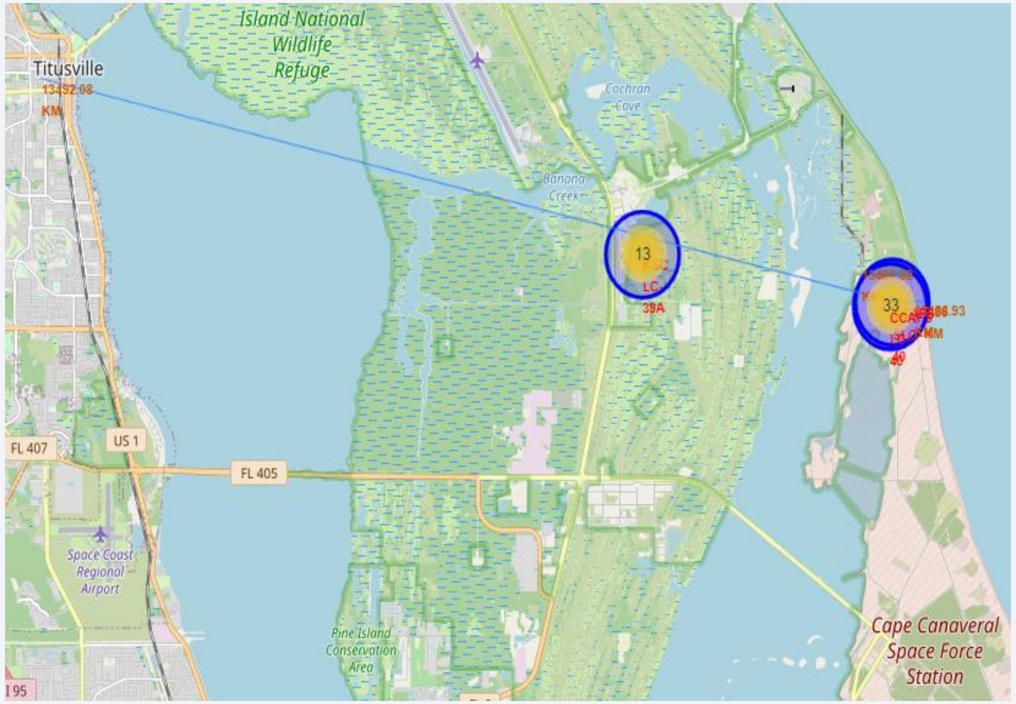
6.. Visualize the launch success yearly trend

Results

Interactive analytics demo in screenshots



PDF of Plotly Dashboard - https://github.com/flurpo/IBMspx/blob/main/Plotly_spacex_dash_app_20240914.pdf

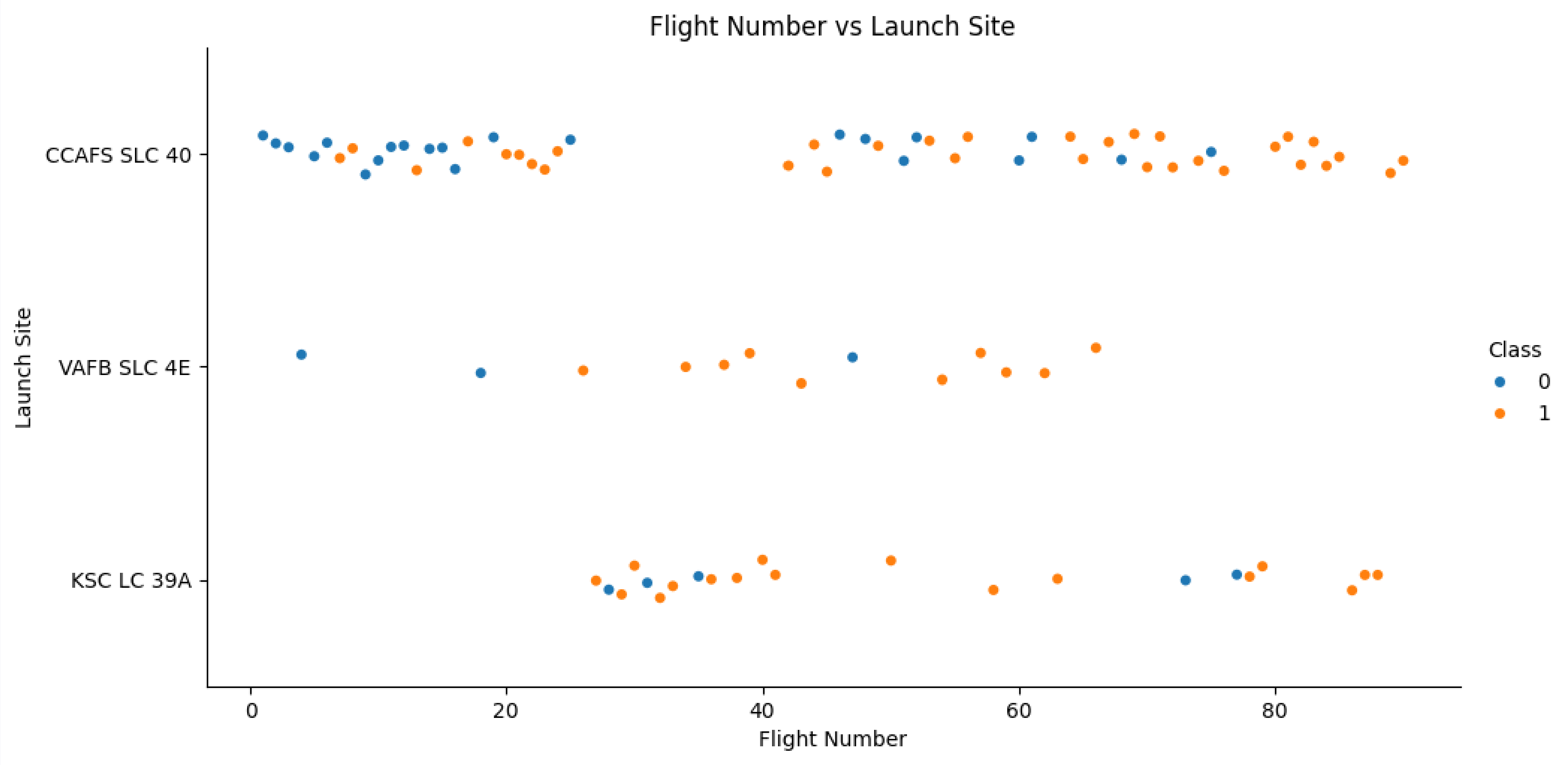


Folium App - https://github.com/flurpo/IBMspx/blob/main/site_map.html

Section 2

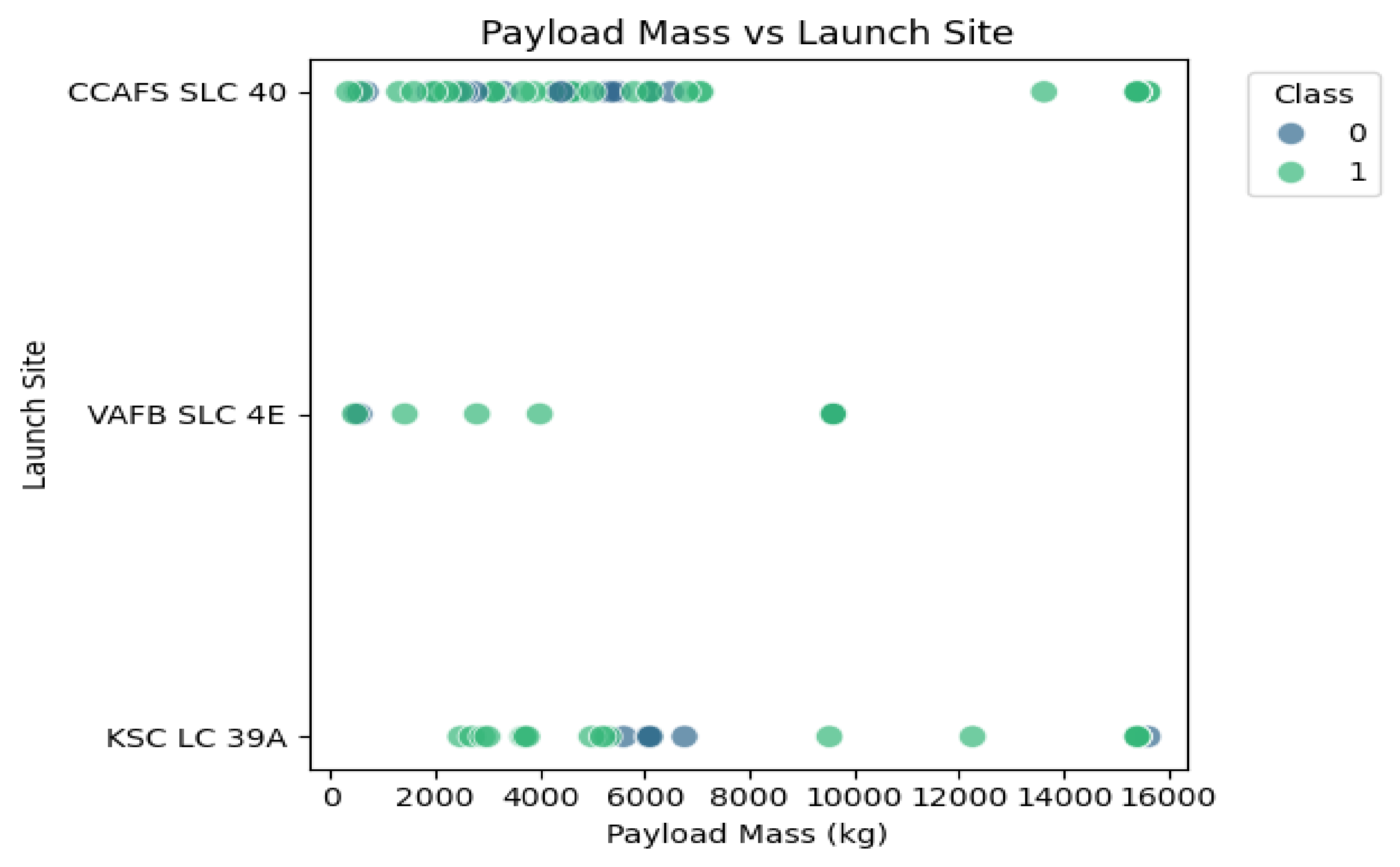
Insights drawn from EDA

Flight Number vs. Launch Site



I see that the majority moved of launches moved from Launch Site SLC40 to Site KSC LC 39A during flight numbers f#25-40 .
The overall Majority of Launches have occurred at site CCAFS SLC 40.

Payload vs. Launch Site



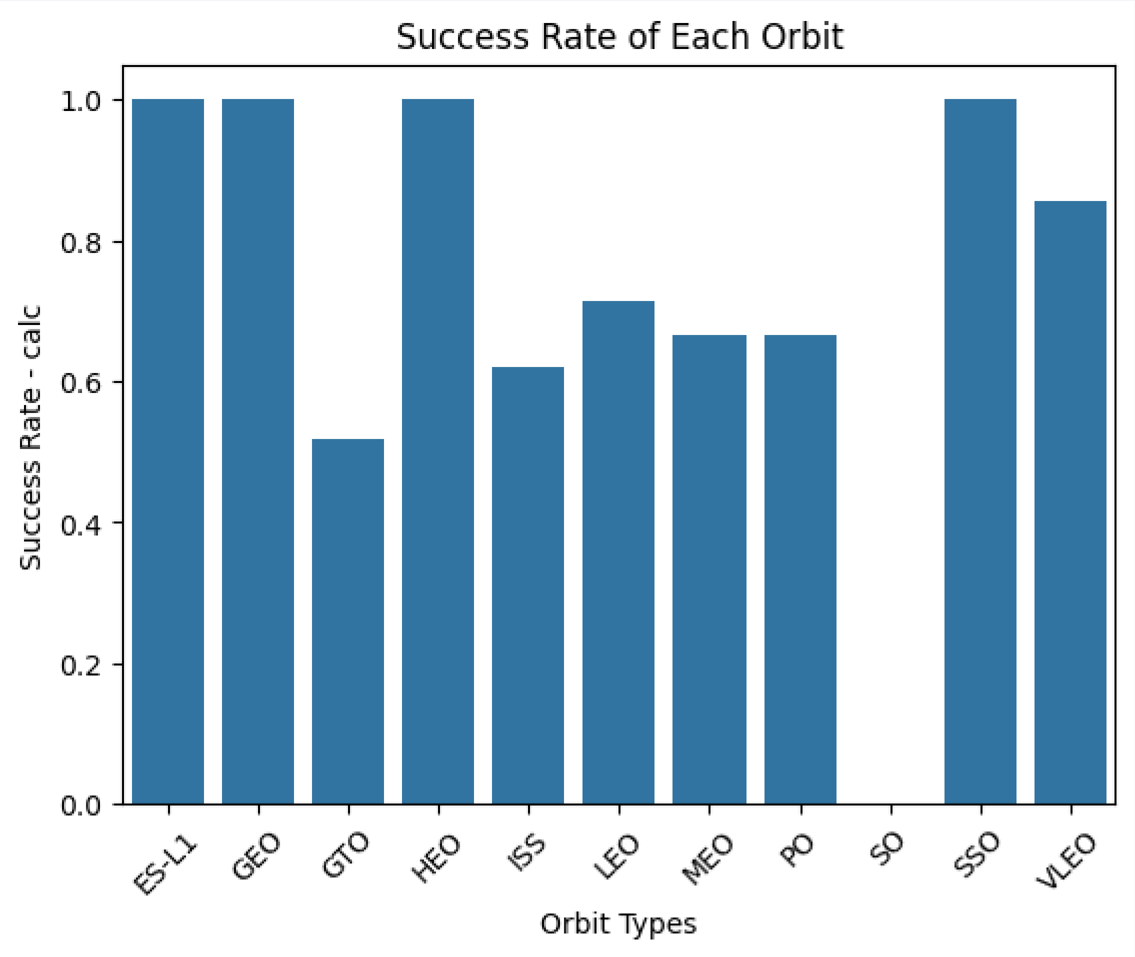
- The most launches occur at CCAFS SLC 40
- The least and lowest payload launches occur at VAFB SLC 4E
- I see the heaviest launch failure at KSC LC 39A

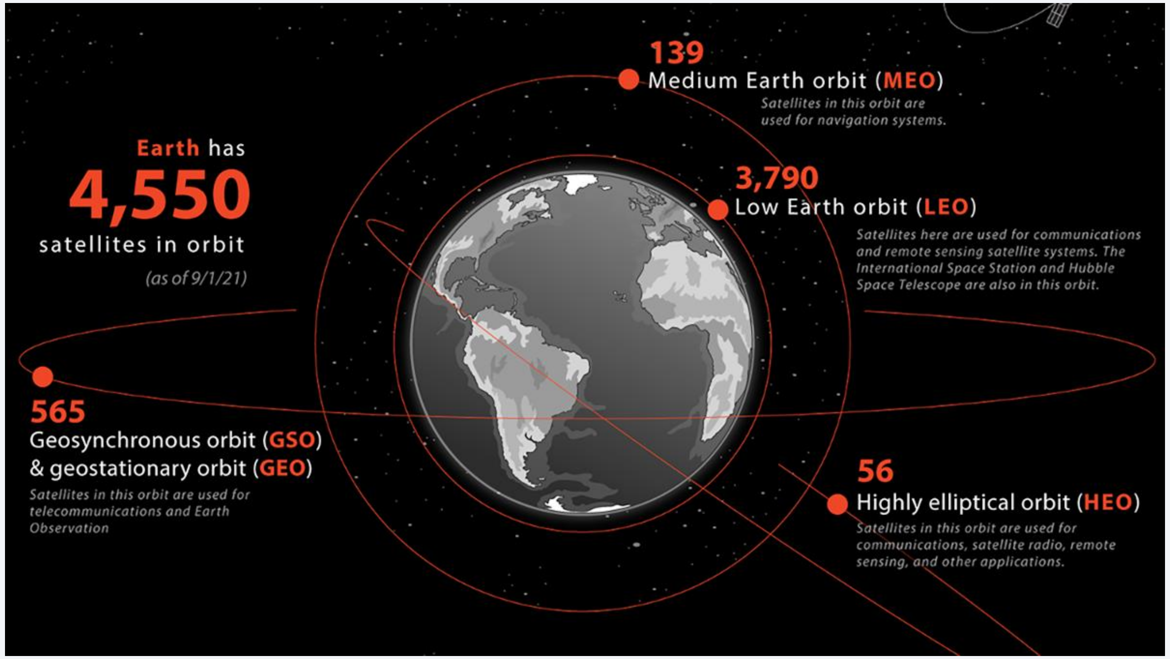
Success Rate vs. Orbit Type

Each Launch aims to a dedicated orbit. Here are some common orbit types:

- **LEO**: Low Earth orbit (LEO)is an Earth-centred orbit with an altitude of 2,000 km (1,200 mi) or less (approximately one-third of the radius of Earth),[1] or with at least 11.25 periods per day (an orbital period of 128 minutes or less) and an eccentricity less than 0.25.[2] Most of the manmade objects in outer space are in LEO [1].
- **VLEO**: Very Low Earth Orbits (VLEO) can be defined as the orbits with a mean altitude below 450 km. Operating in these orbits can provide a number of benefits to Earth observation spacecraft as the spacecraft operates closer to the observation[2].
- **GTO** A geosynchronous orbit is a high Earth orbit that allows satellites to match Earth's rotation. Located at 22,236 miles (35,786 kilometers) above Earth's equator, this position is a valuable spot for monitoring weather, communications and surveillance. Because the satellite orbits at the same speed that the Earth is turning, the satellite seems to stay in place over a single longitude, though it may drift north to south," NASA wrote on its Earth Observatory website [3] .
- **SSO (or SO)**: It is a Sun-synchronous orbit also called a heliosynchronous orbit is a nearly polar orbit around a planet, in which the satellite passes over any given point of the planet's surface at the same local mean solar time [4] .
- **ES-L1** :At the Lagrange points the gravitational forces of the two large bodies cancel out in such a way that a small object placed in orbit there is in equilibrium relative to the center of mass of the large bodies. L1 is one such point between the sun and the earth [5] .
- **HEO** A highly elliptical orbit, is an elliptic orbit with high eccentricity, usually referring to one around Earth [6].
- **ISS** A modular space station (habitable artificial satellite) in low Earth orbit. It is a multinational collaborative project between five participating space agencies: NASA (United States), Roscosmos (Russia), JAXA (Japan), ESA (Europe), and CSA (Canada) [7]
- **MEO** Geocentric orbits ranging in altitude from 2,000 km (1,200 mi) to just below geosynchronous orbit at 35,786 kilometers (22,236 mi). Also known as an intermediate circular orbit. These are "most commonly at 20,200 kilometers (12,600 mi), or 20,650 kilometers (12,830 mi), with an orbital period of 12 hours [8]
- **HEO** Geocentric orbits above the altitude of geosynchronous orbit (35,786 km or 22,236 mi) [9]
- **GEO** It is a circular geosynchronous orbit 35,786 kilometres (22,236 miles) above Earth's equator and following the direction of Earth's rotation [10]
- **PO** It is one type of satellites in which a satellite passes above or nearly above both poles of the body being orbited (usually a planet such as the Earth [11]

Success Rate vs. Orbit Type



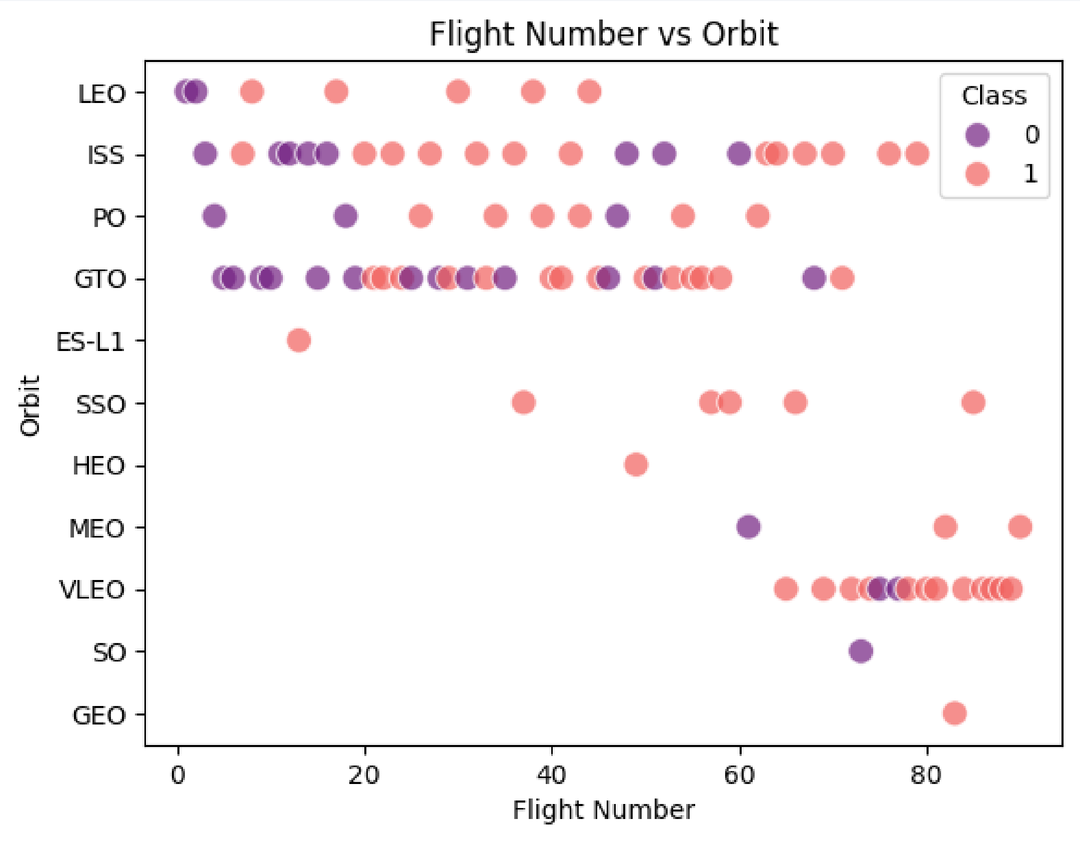


- ES-L1 (Earth-Sun Lagrange point 1), GEO, HEO and SSO (Sun-Synchronous Orbit) have perfect launch records

- GTO(Geostationary Transfer Orbit) is associated with the lowest launch success rate

- No launch data in this data set for SO

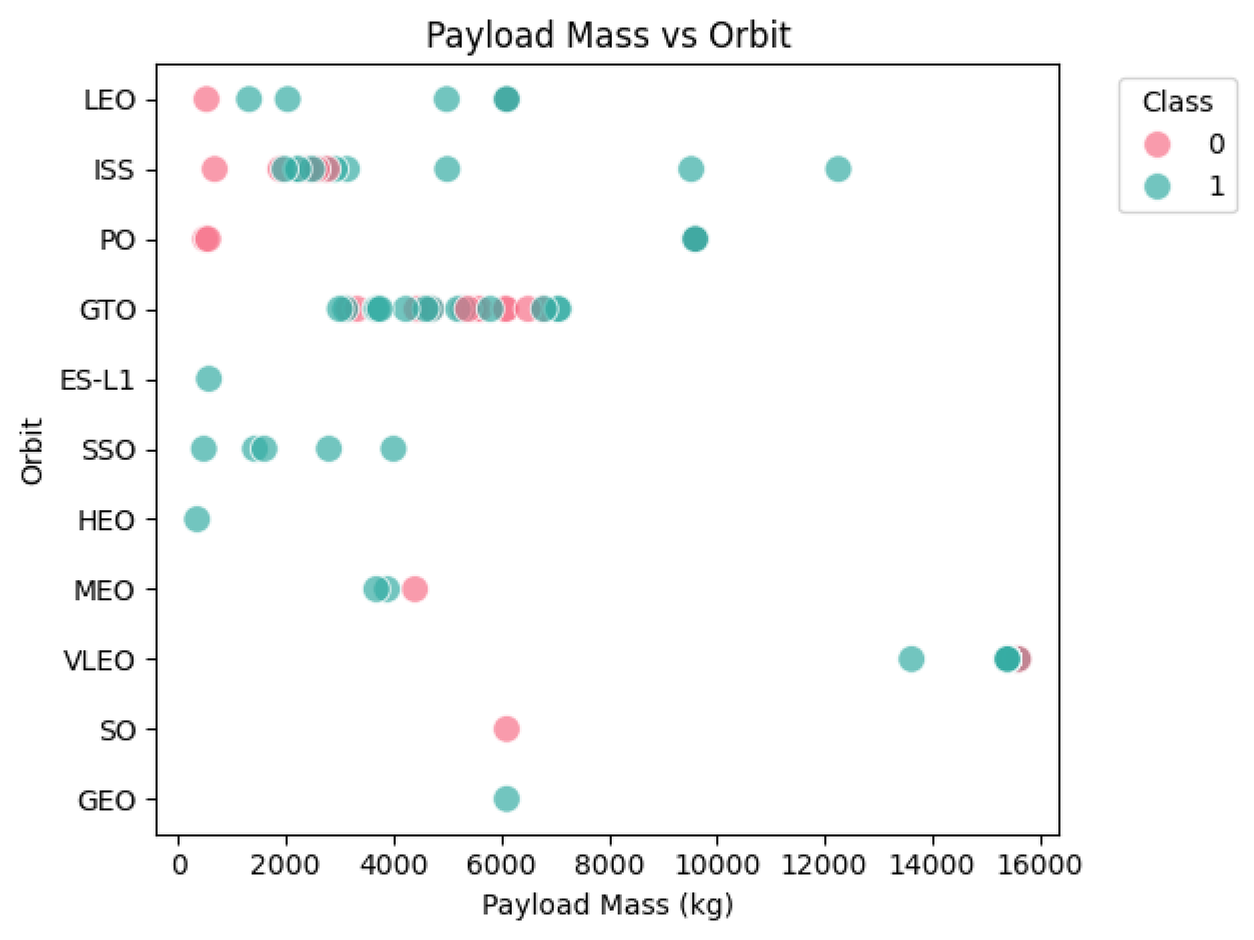
Flight Number vs. Orbit Type



- Most of the Early Flights (<20) involved launch failures.
- Most of the later Flights (>60) where associated with VLEO (very low earth orbit)

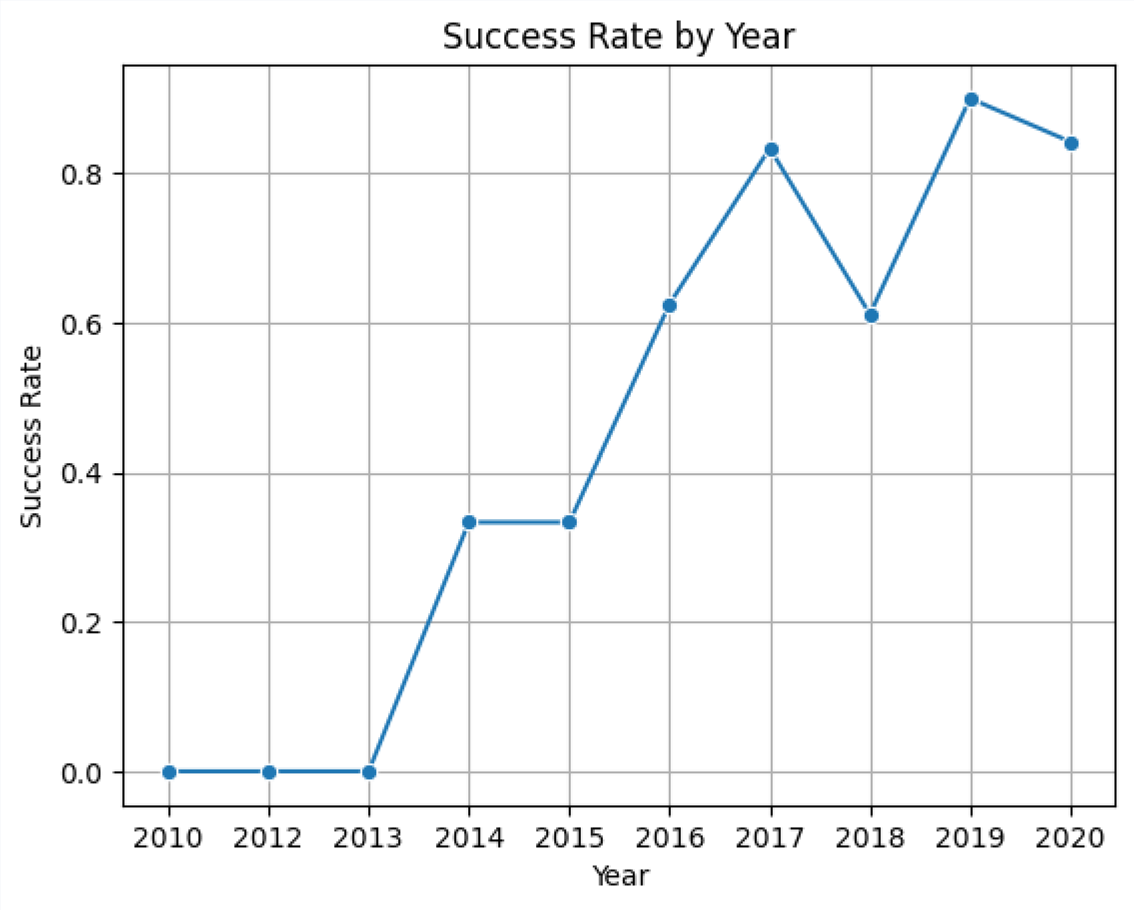
```
You can observe that in the LEO orbit, success seems to be related to the number of flights.
Conversely, in the GTO orbit, there appears to be no relationship between flight number and success.
```

Payload vs. Orbit Type



- With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.
- However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.

Launch Success Yearly Trend



you can observe that the success rate since 2013 kept increasing till 2020

All Launch Site Names

- Find the names of the unique launch sites
- `[('CCAFS LC-40',), ('VAFB SLC-4E',), ('KSC LC-39A',), ('CCAFS SLC-40',)]`

```
CALCULATED BY Select DISTINCT Launch_Site FROM SPACEXTABLE
```

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`

('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')

CALCULATED BY;

```
SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE 'CCA%' LIMIT 5
```

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

Total Payload Mass for NASA (CRS) missions: **45596 KG**

CALCULATED BY;

```
SELECT PAYLOAD_MASS__KG_ FROM SPACEXTABLE WHERE Customer = 'NASA (CRS)'
```

Average Payload Mass by F9 v1.1

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

```
(3170,)(3325,) (2296,) (1316,) (4535,)

 Average Payload Mass for Booster_Version F9 v1.1: 2928.4 KG
```

Calculated by;

SELECT PAYLOAD_MASS__KG_ FROM SPACEXTABLE WHERE Booster_Version = 'F9 v1.1'

First Successful Ground Landing Date

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

Earliest Successful Ground Pad Landing: <u>2015-12-22</u>

CALCULATED BY;

```
SELECT MIN(Date) FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (ground pad)'
```

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

Names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000: <u>F9 FT B1022</u>

```
CALCULATED BY;

SELECT DISTINCT(Booster_Version) FROM SPACEXTABLE WHERE Landing_Outcome = '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

Rows in Table SPACEXTABLE: 101
Successful Mission Outcome: 98
Failure Mission Outcome: 1
Other Mission Outcome: 2

CALCULATED BY:

```
SELECT COUNT(*) FROM SPACEXTABLE WHERE Mission_Outcome = 'Success’
SELECT COUNT(*) FROM SPACEXTABLE WHERE Mission_Outcome LIKE 'Failure%’
SELECT COUNT(*) FROM SPACEXTABLE WHERE Mission_Outcome != 'Success' AND Mission_Outcome NOT LIKE 'Failure%'
```

Boosters Carried Maximum Payload

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

Booster versions MAX Payload: F9 B5 B1048.4

CALCULATED BY:

```
SELECT Booster_Version
FROM SPACEXTABLE
WHERE PAYLOAD_MASS__KG_ = (
SELECT MAX(PAYLOAD_MASS__KG_)
FROM SPACEXTABLE)
```

2015 Launch Records

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

CALC BY:

```
SELECT
    CASE SUBSTR(Date, 6, 2)
        WHEN '01' THEN 'January'
        WHEN '02' THEN 'February'
        WHEN '03' THEN 'March'
        WHEN '04' THEN 'April'
        WHEN '05' THEN 'May'
        WHEN '06' THEN 'June'
        WHEN '07' THEN 'July'
        WHEN '08' THEN 'August'
        WHEN '09' THEN 'September'
        WHEN '10' THEN 'October'
        WHEN '11' THEN 'November'
        WHEN '12' THEN 'December'
    END AS Month_Name,
    Booster_Version,
    Launch_Site,
    Landing_Outcome
FROM SPACEXTABLE
WHERE Landing_Outcome = 'Failure (drone ship)'
  AND SUBSTR(Date, 0, 5) = '2015'
```

('January', 'F9 v1.1 B1012', 'CCAFS LC-40', 'Failure (drone ship)')
('April', 'F9 v1.1 B1015', 'CCAFS LC-40', '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

('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)

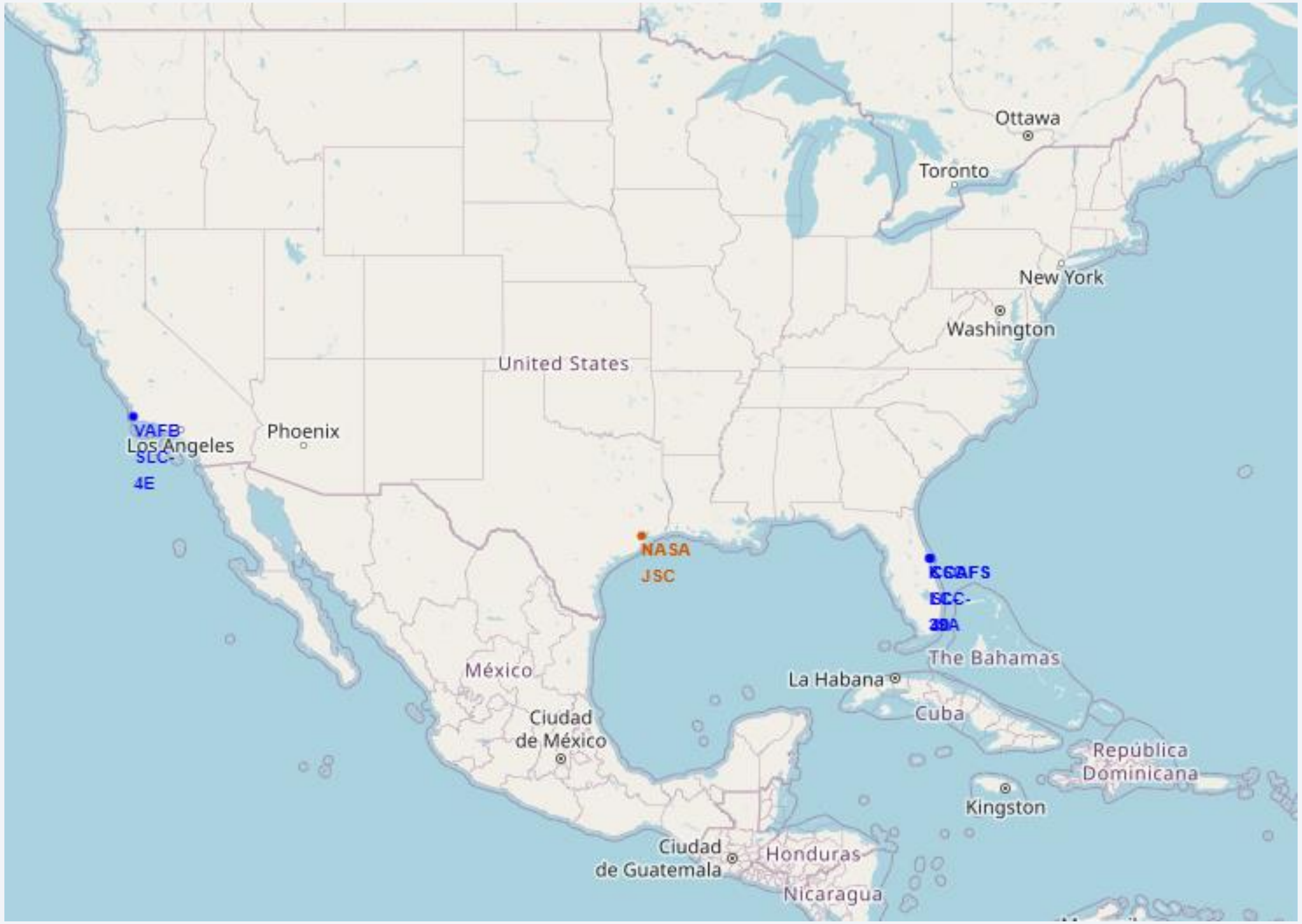
CALC BY

```
SELECT Landing_Outcome, COUNT(*) AS Outcome_Count
FROM SPACEXTABLE
WHERE Date BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY Landing_Outcome
ORDER BY Outcome_Count DESC
```

Section 3

Launch Sites Proximities Analysis

Launch Sites- Folium 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

Note:

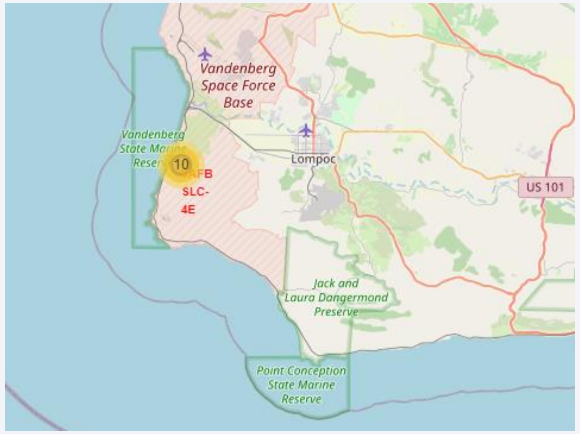
One Launch Site at Vandenburg AireForce Base in California

Two Launch Sites at Cape Canaveral Space Launch Complex in Florida

One Launch Site at the Kennedy Space Center in Florida

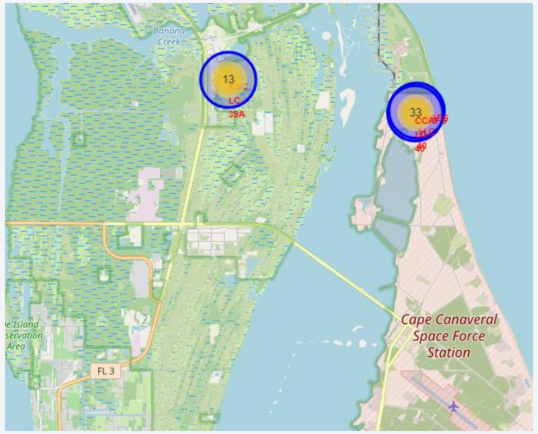
Interactive Launch Information - Folium Map

VAFB SLC 4E /10 Launches

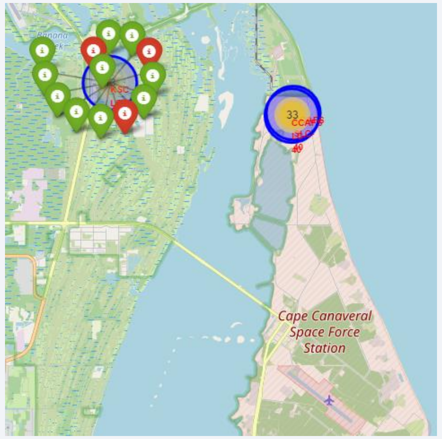


VAFB SLC 4E (4 successful Launches, 6 failures)

KSC LC 39A/13 Launches CCAFS SLC&LC 40/33 Launches



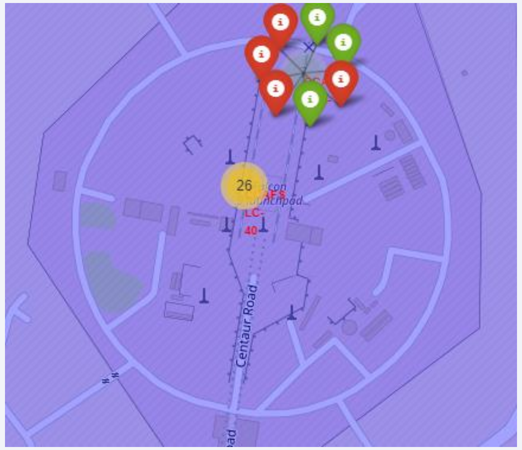
KSC LC 39A(10 successful Launches, 3 failures)



CCAFS LC 40/26 Launches CCAFS SLC 40/7 Launches

CCAFS LC 40/8 successful Launches, 18 failures)

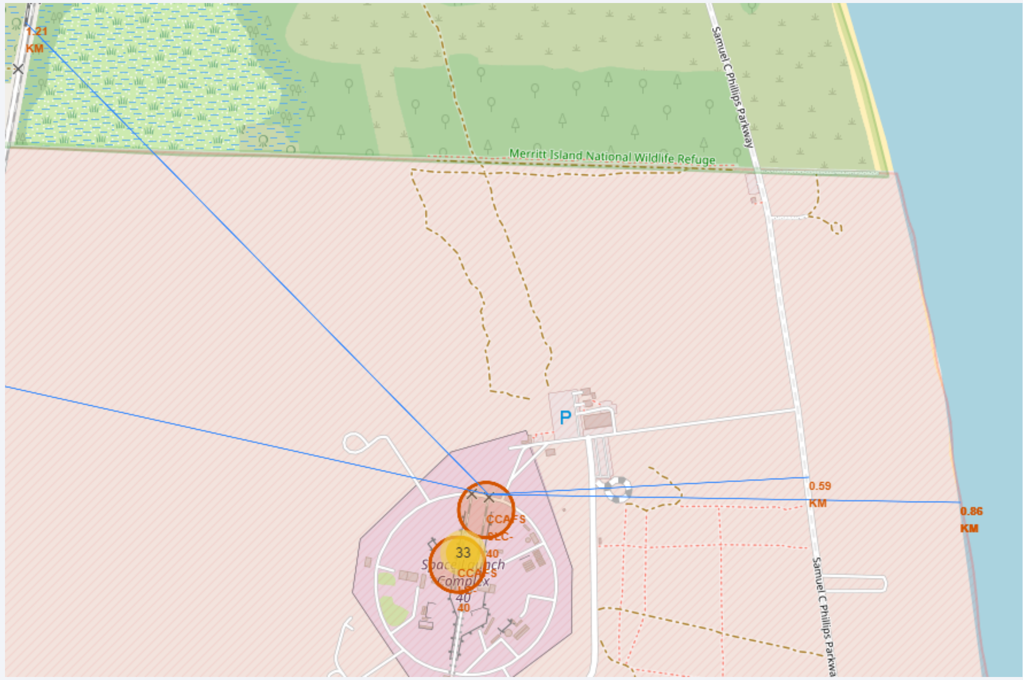
CCAFS SLC 40/7 Launches



BLUE CIRCLE - LAUNCH SITE
YELLOW CIRCLE – SITE LAUNCH COUNT MIXED
GREEN CIRCLE – SITE LAUNCH ALL SUCCESSFUL

GREEN TAGS – SUCCESSFUL LAUNCHES
RED TAGS – UNSUCCESSFUL LAUNCHES

Calculated Distances and markings - Folium Map

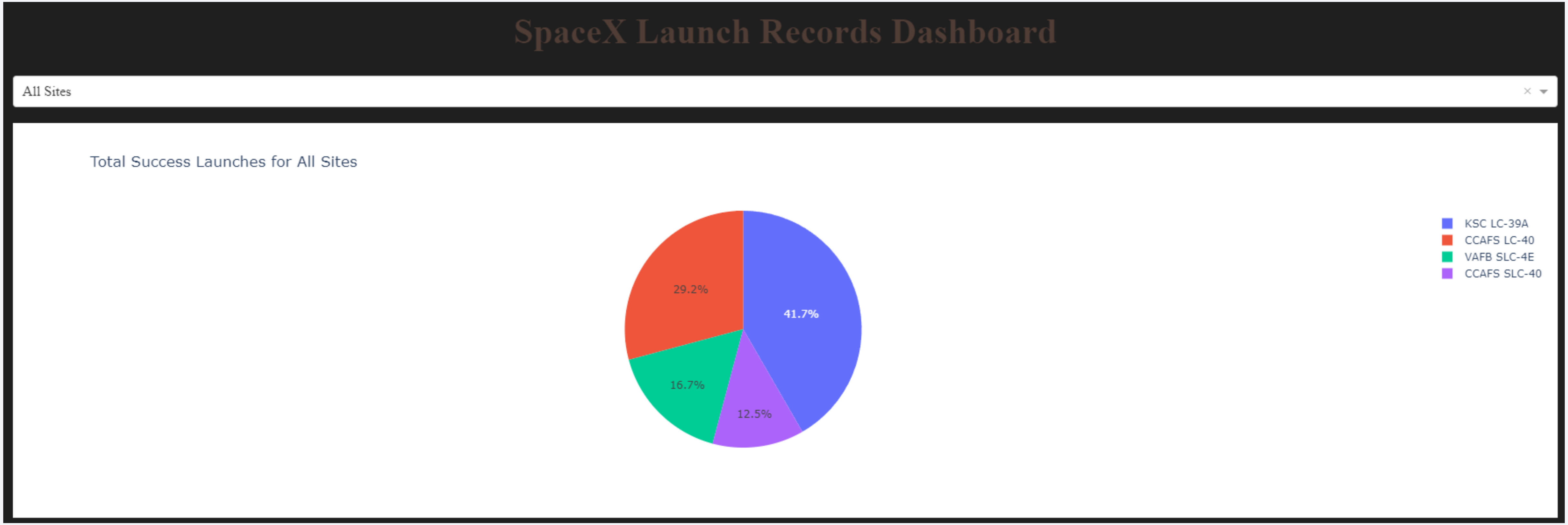


CCAFS SLC 40 to Coastline – 0.862 KM
CCAFS SLC 40 to Parkway – 0.586 KM
CCAFS SLC 40 to Railway – 1.206 KM
CCAFS SLC 40 to Titusville – 23.89 KM

Section 4

Build a Dashboard with Plotly Dash

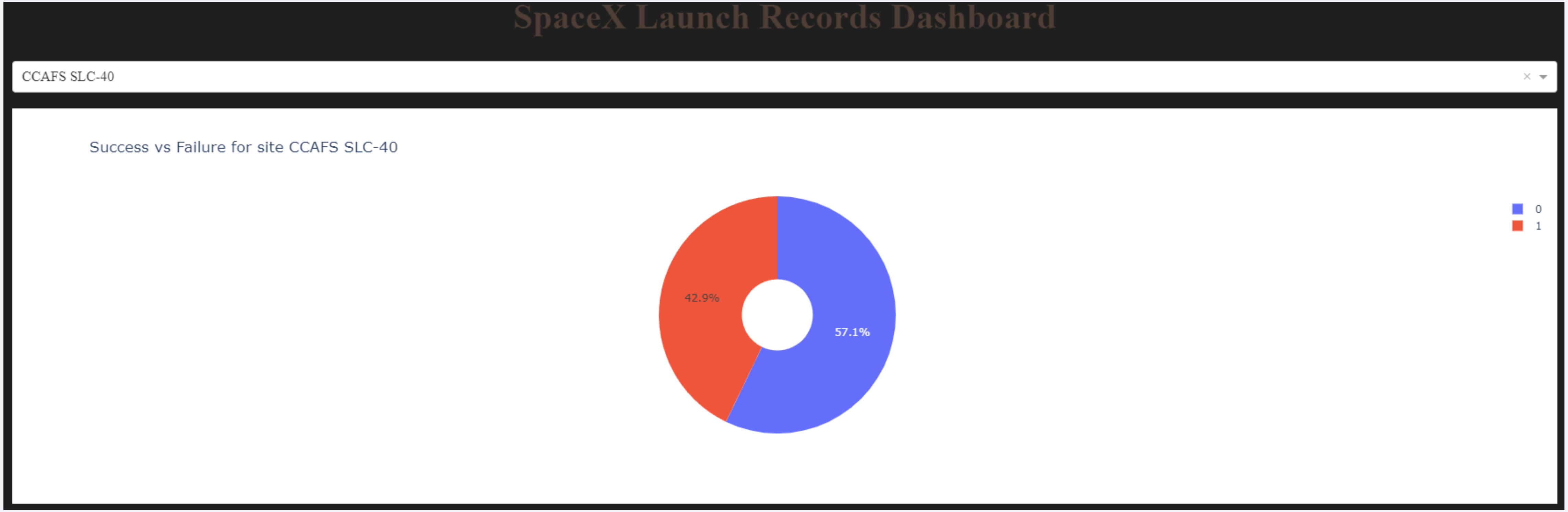
Dashboard Application – PIE CHART – Successful Launches All Sites



- Launch success in the early years (first 56 launches) was marginal.
- KSC LC – 39A represented the largest percentage of the total successful launches.

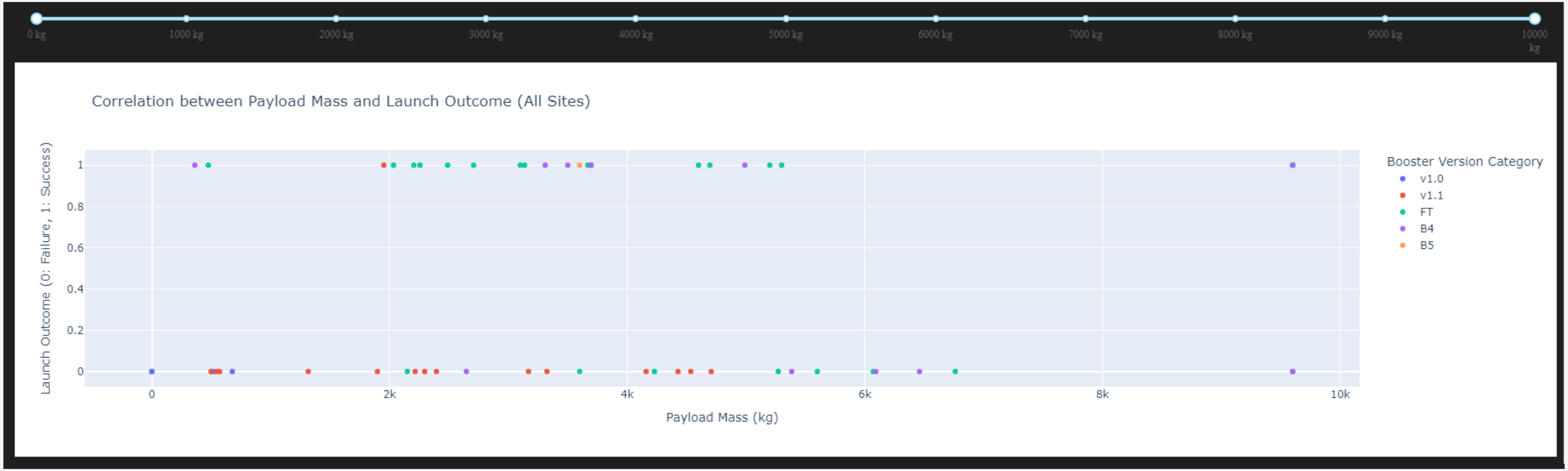
Dashboard Application – PIE CHART – Highest Launch Success Ratio

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



Though KSC LC-39A was responsible for the highest percentage of success across the launch sites, CCAFS SLC-40 had the highest launch success ratio.

Dashboard Application – SCATTER PLOT– All Site/Full Payload Range



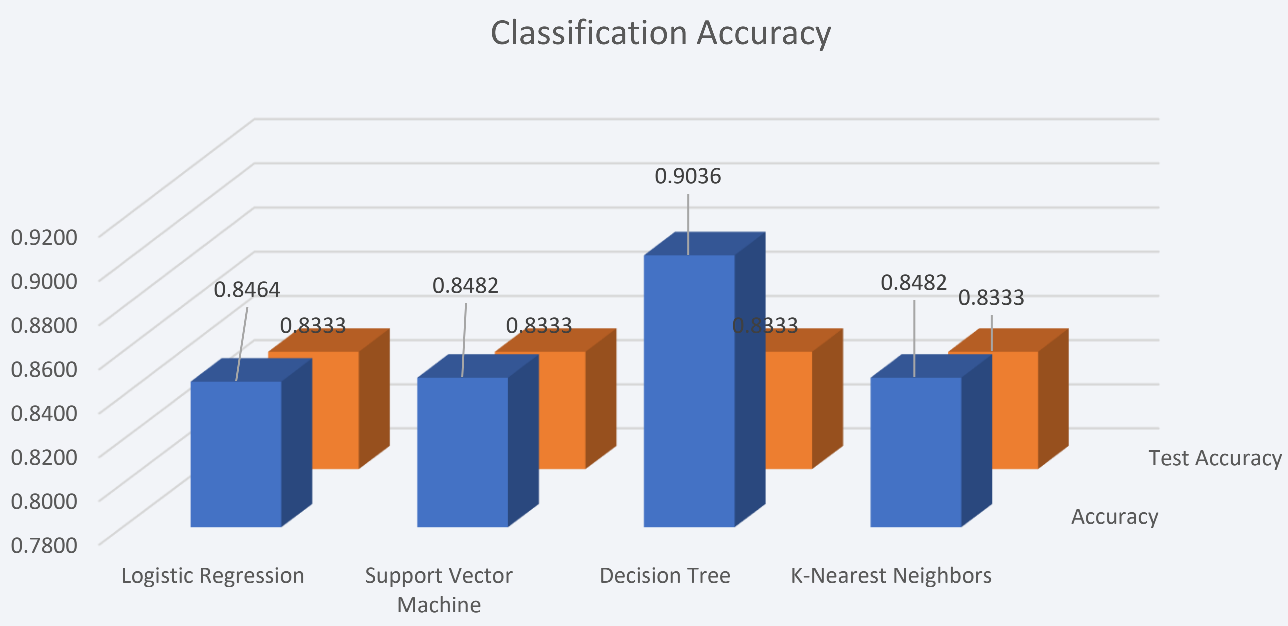
Across all Sites:

- Booster Version FT was the most successful
- Booster Version v1.1 was the least successful
- Booster Version B4 was successfully launched with the highest payload mass

Section 5

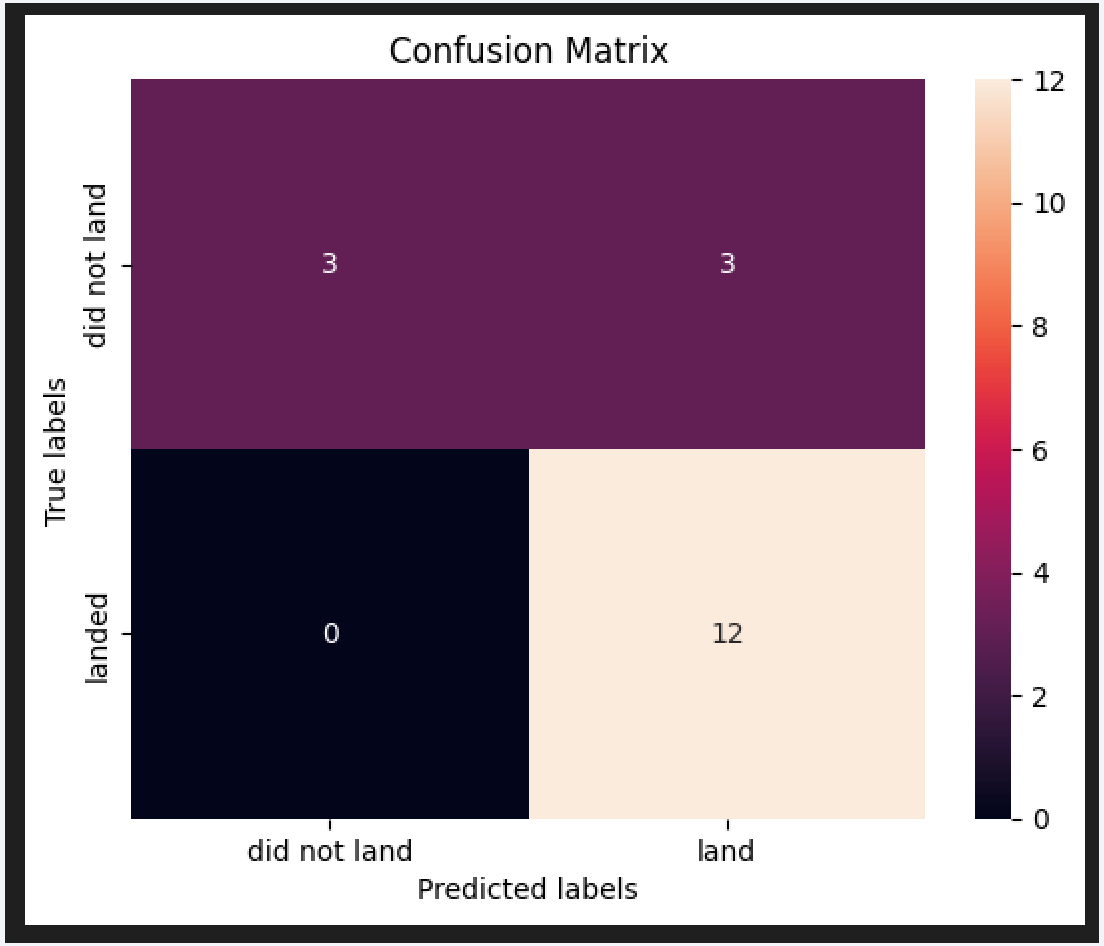
Predictive Analysis (Classification)

Classification Accuracy



The DECISION TREE CLASSIFIER has the highest fit accuracy

Confusion Matrix



The test data accuracy and confusion matrix response was identical across the model results. Likely because of the lack of data and differentiation in the data. So, its very difficult to chose a best model based on the test accuracy

1.Accuracy: (TP + TN) / (TP + TN + FP + FN) = 0.8333333
2.Precision: TP / (TP + FP) = 0.8
3.Recall (Sensitivity): TP / (TP + FN) = 0.8
4.Specificity: TN / (TN + FP) = 0.5
5.F1-Score: 2 * (Precision * Recall) / (Precision + Recall) = 0.8

Conclusions

- I would choose the DECISION TREE CLASSIFIER as the best machine learning algorithm as it has the highest GridSearchCV result! accuracy : 0.9036
- Launch success in the early years (first 56 launches) was marginal.
- KSC LC – 39A represented the largest percentage of the total successful launches
- Though KSC LC-39A was responsible for the highest percentage of success across the launch sites, CCAFS SLC-40 had the highest launch success ratio
- The overall Majority of Launches have occurred at site CCAFS SLC 40.
- you can observe that the launch success rate since 2013 kept increasing till 2020
- Orbits ES-L1, GEO, HEO, SSO, VLEO had the most success rate.

Appendix

- All files utilized for information in this document are located at-

  https://github.com/flurpo/IBMspx

- Python version of Plotly DASH app (access through browser) can be accessed at -

https://github.com/flurpo/IBMspx/blob/main/Python Notebooks/spacex_dash_app.py

Thank you!