Geospatial and Predictive Analysis of SpaceX Falcon 9 Launch and Landing Outcomes

Shola Awoyemi, PhD

Outline

Executive Summary

- This project explores SpaceX launch data through geospatial mapping, SQL-driven exploration, and predictive modeling. Using Python, SQL, Folium, and Plotly Dash, I built interactive dashboards and maps to uncover patterns in launch success, proximity to infrastructure, and temporal trends. A classification model was developed to predict launch outcomes based on engineered features. The final deliverables include a robust dashboard, interactive maps, and actionable insights for future launch planning.

Introduction

- SpaceX has revolutionized space travel, but understanding the factors behind successful launches remains critical. This project aims to:
  - Analyze launch patterns over time and geography
  - Identify correlations between infrastructure proximity and outcomes
  - Build predictive models to forecast launch success
  - Deliver insights through interactive dashboards and maps
- Key Project Activities Include:
  - Data Preparation: Data Collection with API and Webscrapping, Data Wrangling
  - Exploratory Data Analysis (EDA): EDA with SQL and Visualization
  - Interactive Visual Analytics and Dashboard: Interactive Visual Analytics with Folium Lab and Interactive Dashboard with Plotly Dash
  - Predictive Analysis (Classification): Machine Learning Prediction with Selected Model
- Tools/Library:
  - Jupyter Notebook, Pandas, Numpy, Folium, Scipy, Sklearn, Github
- Skills/Knowledge Required:
  - Statistical Analysis, Python, SQL

METHODOLOGY

Data Preparation

❑ Data Collection with API

- Request rocket launch data from SpaceX API with the URL
- Task 1: Request and parse the SpaceX launch data using the GET request and normalize JSON Result
- Create a Pandas data frame from the dictionary launch_dict
- Task 2: Filter the dataframe to only include Falcon 9 launches
- Task 3: Dealing with Missing Values (Data Wrangling)
- Export to CSV file

Data Preparation

- Data Collection API with Webscrapping
  - TASK 1: Request the Falcon9 Launch Wiki page from its URL
  - Perform an HTTP GET method to request the Falcon9 Launch HTML page
  - Create Bueautifulsoup object from the HTML response
  - TASK 2: Extract all column/variable names from the HTML table header
  - TASK 3: Create a data frame by parsing the launch HTML tables
  - Export to CSV file

Data Preparation

- Data Wrangling
  - Data Analysis: Import Libraries and Define Auxiliary Functions
  - TASK 1: Calculate the number of launches on each site
  - TASK 2: Calculate the number and occurrence of each orbit
  - TASK 3: Calculate the number and occurrence of mission outcome of the orbits
  - TASK 4: Create a landing outcome label from Outcome column
  - Calculate mean (success rate)
  - Export to CSV file

Exploratory Data Analysis (EDA)

❑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 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 all the booster versions that have carried the maximum payload mass, using a subquery with a suitable aggregate function.
- 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.

Exploratory Data Analysis (EDA)

❑ EDA with Visualization (Matplotlib - Plotly & Seaborn)

- Plot out the FlightNumber vs. PayloadMass and overlay the outcome of the launch
- TASK 1: Visualize the relationship between Flight Number and Launch Site
- TASK 2: Visualize the relationship between Payload Mass and Launch Site
- TASK 3: Visualize the relationship between success rate of each 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
- Features Engineering
- Features Engineering
- TASK 7: Create dummy variables to categorical columns
- TASK 8: Cast all numeric columns to float64
- Export to CSV file

Interactive Visual Analytics and Dashboard

- Interactive Visual Analytics with Folium
  - TASK 1: Mark all launch sites on a map
  - Check proximity of launch sites to the coast and equator line
  - TASK 2: Mark the success/failed launches for each site on the map
  - Create a MarkerCluster object
  - Add marker_color column based on class
  - Add marker_cluster to current site_map

Interactive Visual Analytics and Dashboard

❑ Interactive Visual Analytics with Folium

- TASK 3: Calculate the distances between a launch site to its proximities
- After you plot distance lines to the proximities, you can answer the following questions easily:
  - Are launch sites in close proximity to railways?
  - Are launch sites in close proximity to highways?
  - Are launch sites in close proximity to coastline?
  - Do launch sites keep certain distance away from cities?
  - Are launch sites in close proximity to airport?
  - Are launch sites in close proximity to seaport?

Interactive Visual Analytics and Dashboard

- ❑Interactive Dashboard with Plotly DASH
  - 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
- ❑ Find insights visually
- ❑Analyze SpaceX launch data, and answer the following questions:
  1. Which site has the largest successful launches?
  2. Which site has the highest launch success rate?
  3. Which payload range(s) has the highest launch success rate?
  4. Which payload range(s) has the lowest launch success rate?
  5. Which F9 Booster version (v1.0, v1.1, FT, B4, B5, etc.) has the highest launch success rate?

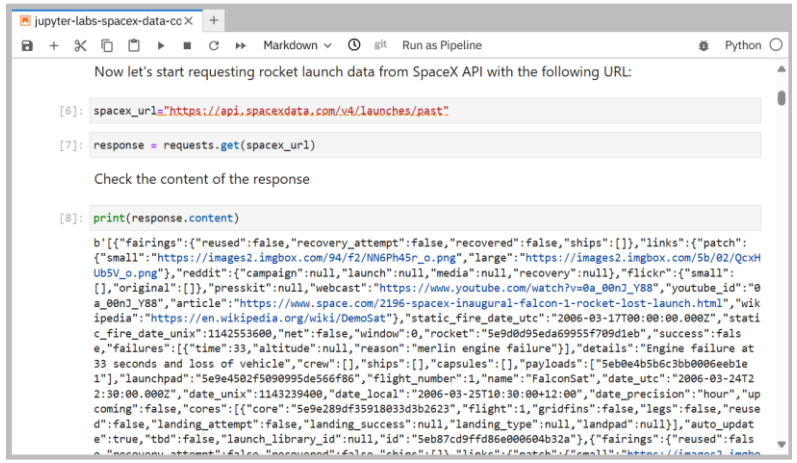
Predictive Analysis (Classification)

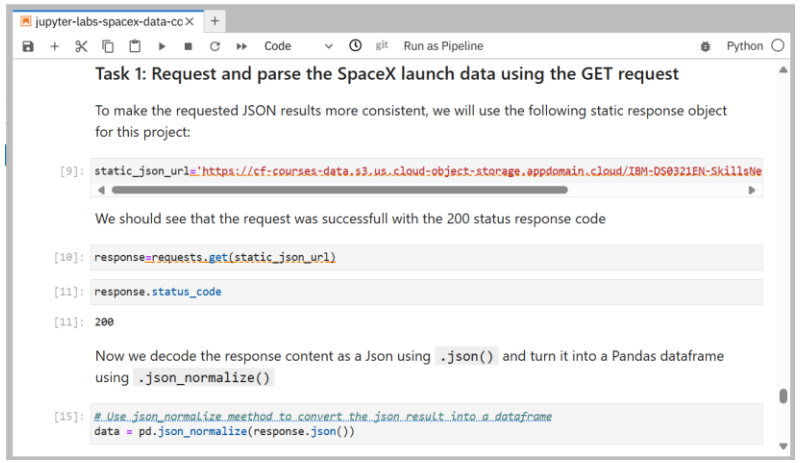
❑Machine Learning Prediction

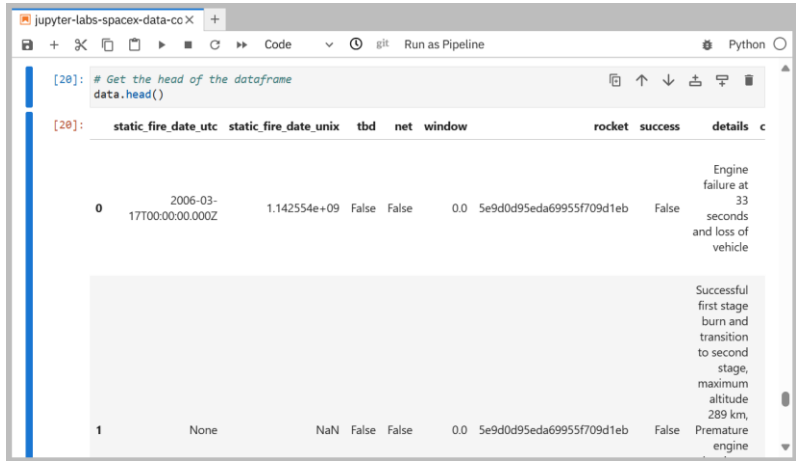
- Create NumPy array from the column Class in data
- Standardize the data
- Split the data X and Y into training and test data
- Create a logistic regression object, GridSearchCV object, Confusion matrix
- Create a support vector machine object, GridSearchCV object, Confusion matrix
- Create a decision tree classifier object, GridSearchCV object, Confusion matrix
- Create a k nearest neighbors object, GridSearchCV object, Confusion matrix

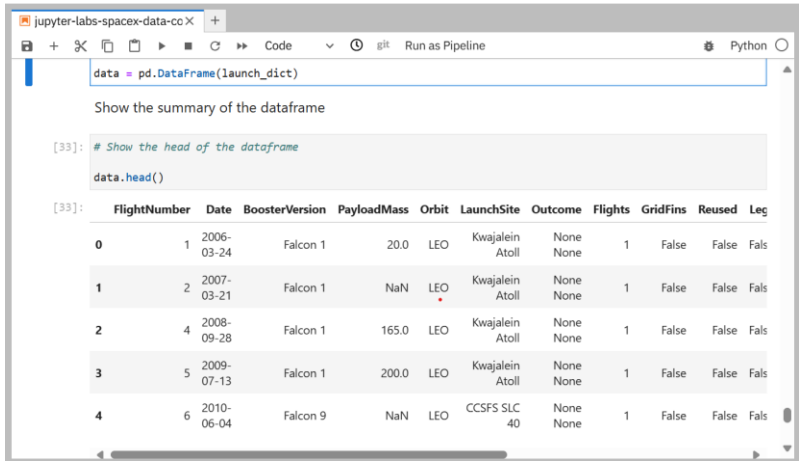
RESULTS & DISCUSSION

Data Preparation





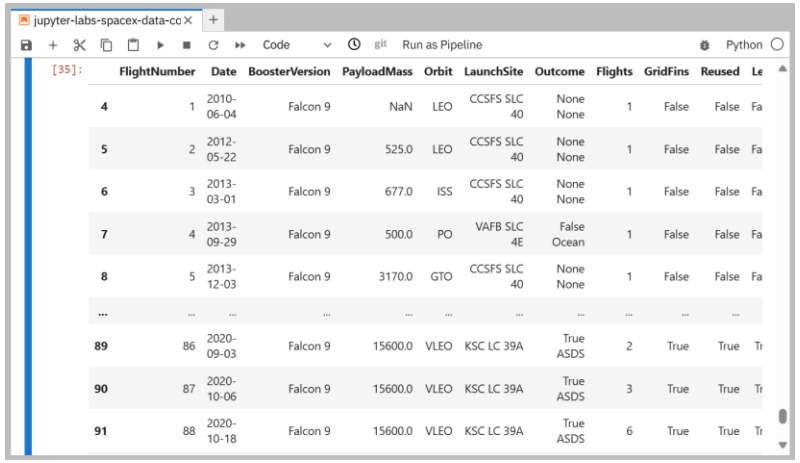




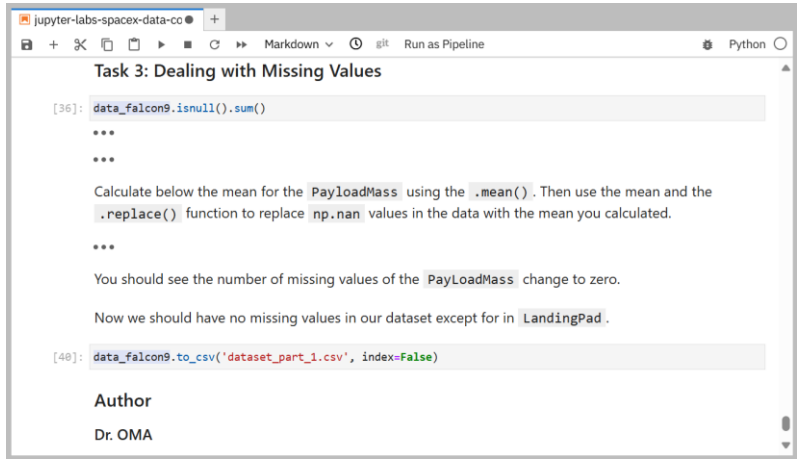
❑ Data Collection with API

- SpaceX rocket launch data collected through API with the URL
- Normalized JSON from parsing SpaceX launch data
- Pandas data frame of launch data from the dictionary

Data Preparation

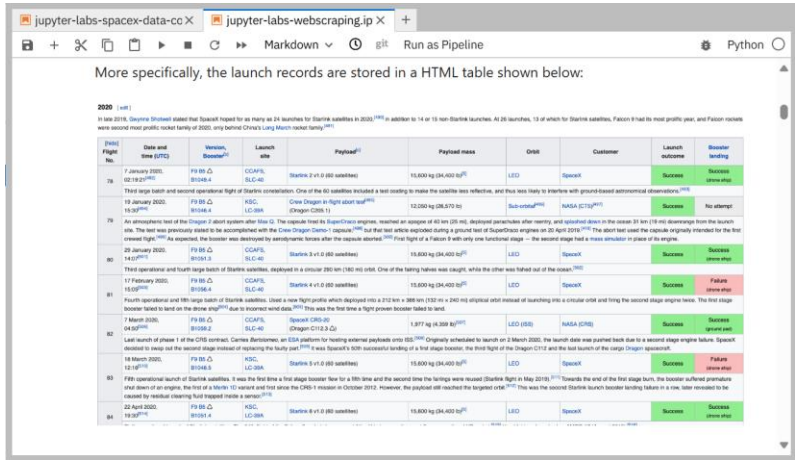


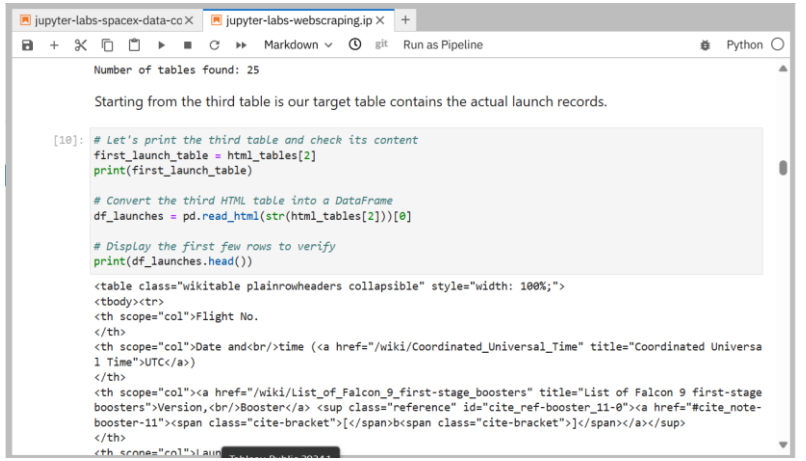
	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Le
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	Fa
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	Fa
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	Fa
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	Fa
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	Fa
...	...	...	...	...	...	...	...	...	...	...	...
89	86	2020-09-03	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	2	True	True	Tr
90	87	2020-10-06	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	3	True	True	Tr
91	88	2020-10-18	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	6	True	True	Tr

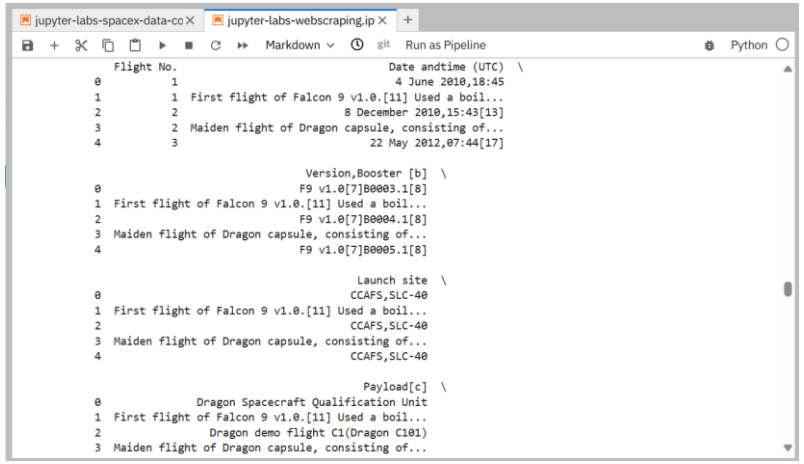


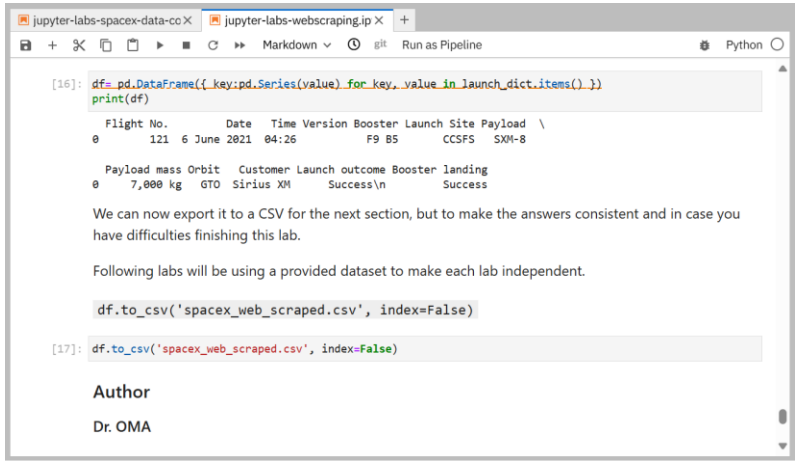
- Data Collection with API
  - Data frame includes only Falcon 9 launches
  - Data frame missing values were well handled (Data Wrangling)
  - Data frame is available as a CSV file
  - Key features of the data frame column are easily identified

Data Preparation





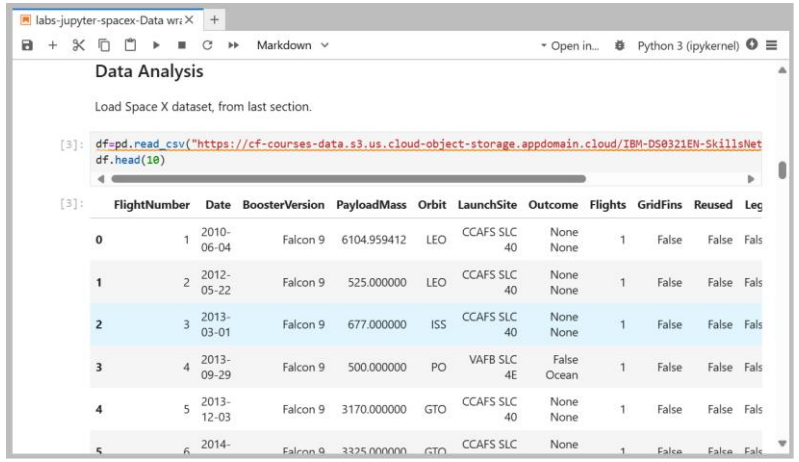


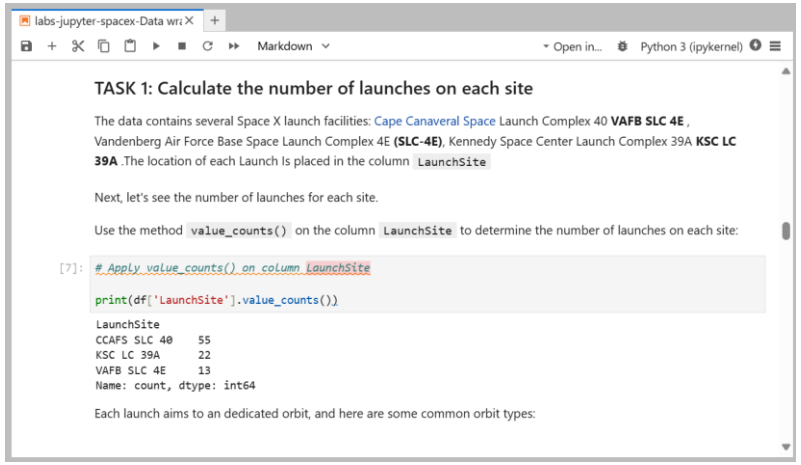


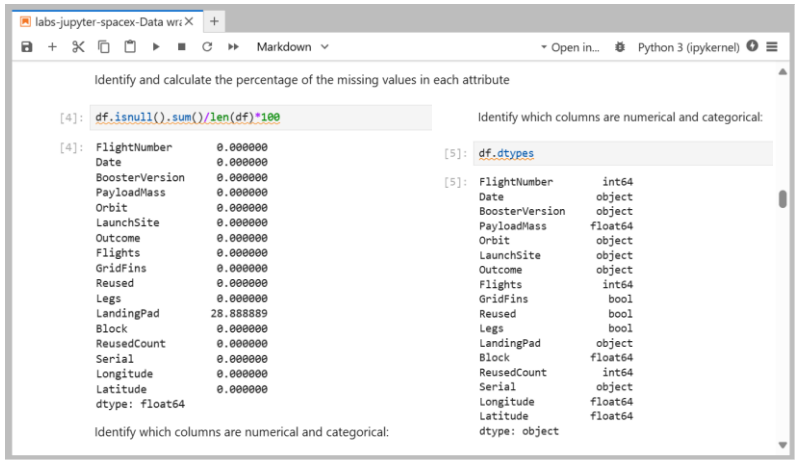
❑ Data Collection API with Webscrapping

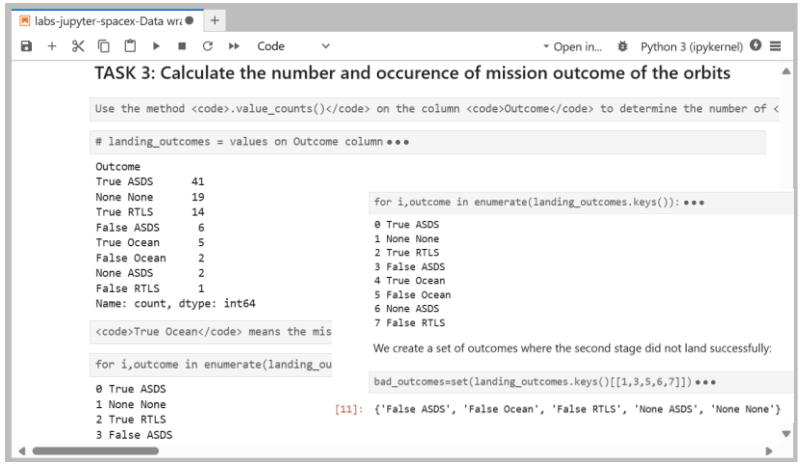
- Falcon9 Launch Wiki page from its URL
- Falcon9 Launch HTML page
- Bueautifulsoup object from the HTML response
- Data frame from parsing the launch HTML tables
- Data frame is available as a CSV file
- Column/variable names identified from the HTML table header

Data Preparation



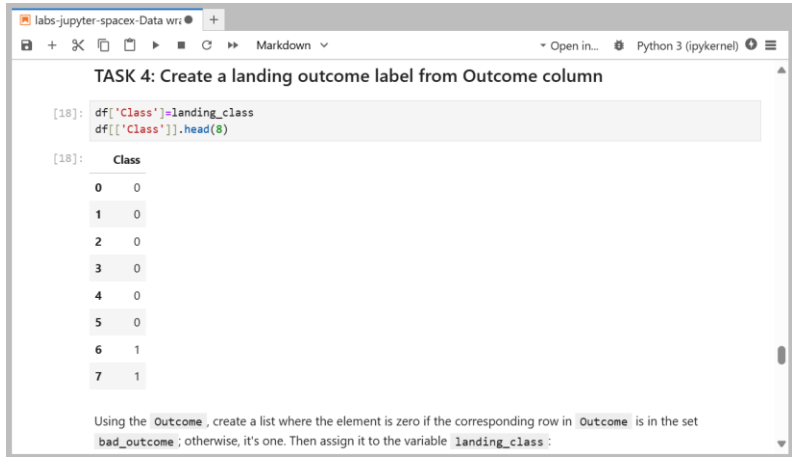


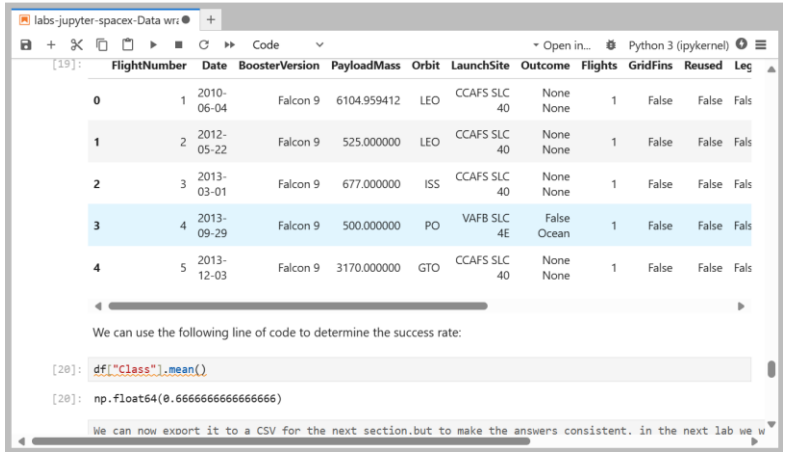




- Data Wrangling
  - Calculated number of launches on each site
  - Calculated number and occurrence of each orbit
  - Calculate number and occurrence of mission outcome of the orbits

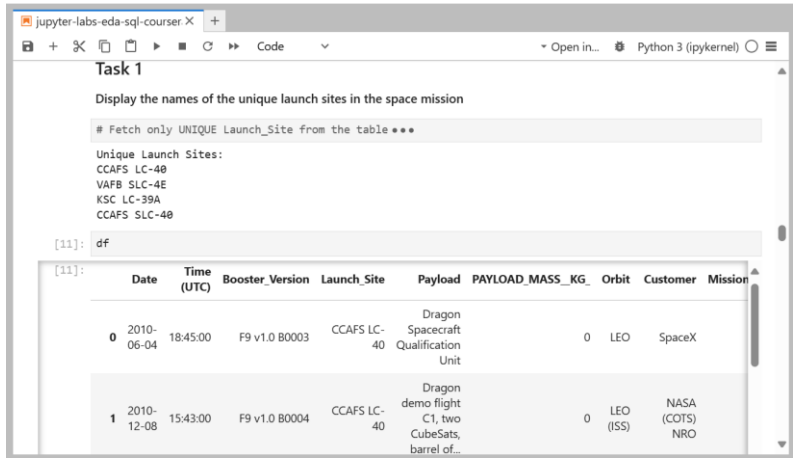
Data Preparation

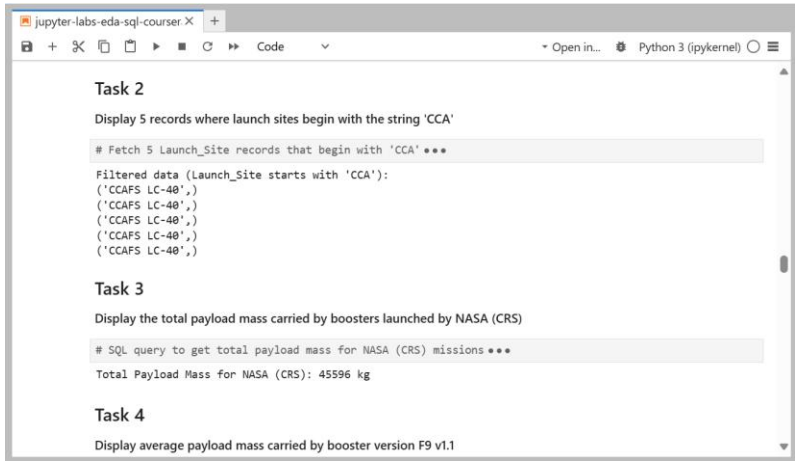


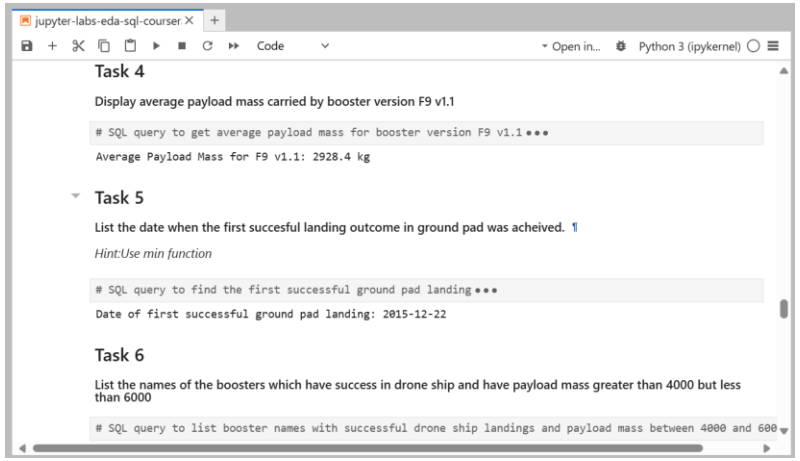


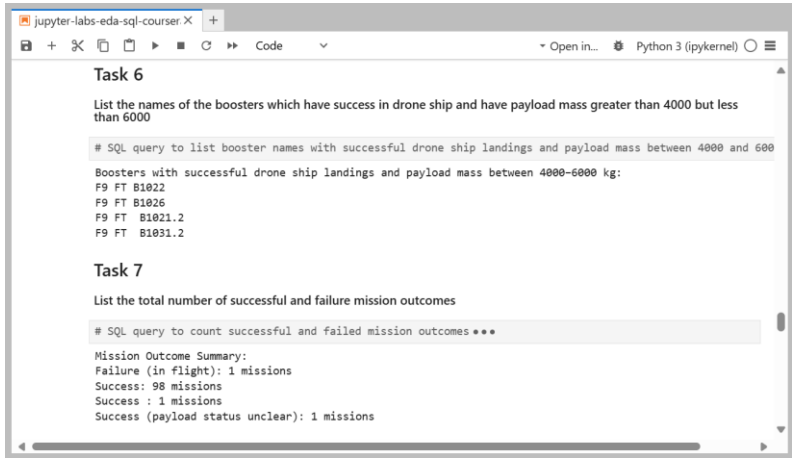
- Data Wrangling
  - Labelled landing outcomes either a failure (0) or success (1)
  - Calculated mean (success rate) of 66.67%
  - Supplementary data available as CSV file

Exploratory Data Analysis (EDA)





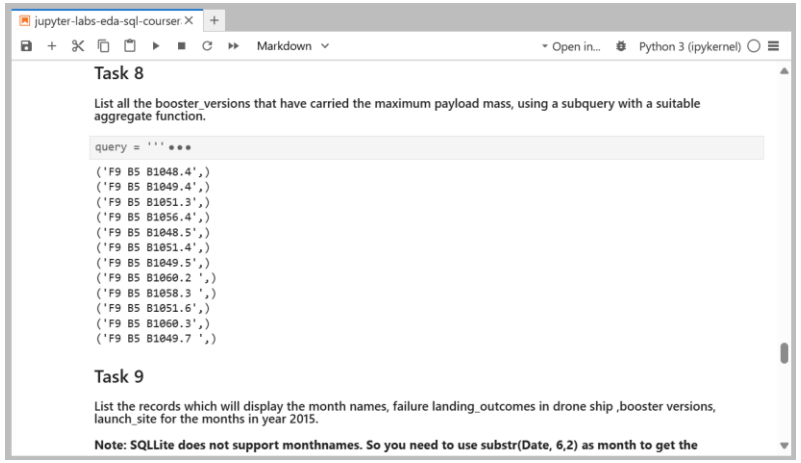


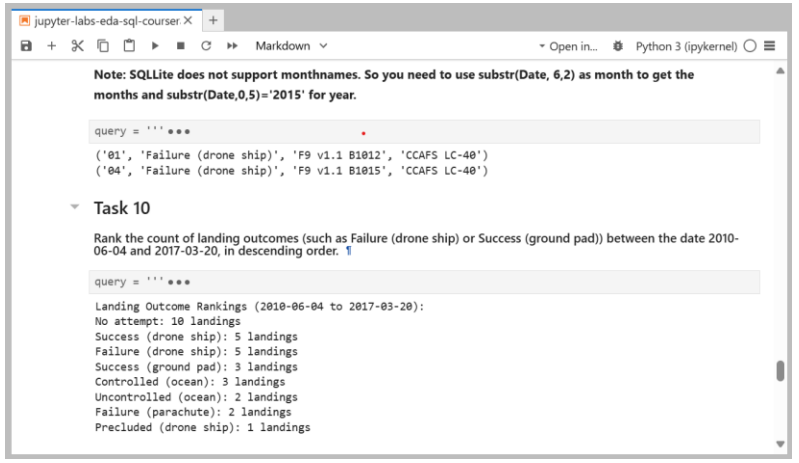


❑ EDA with SQL

- There are a total of four unique launch sites in the space mission including CCAFS LC-40, VAFB SLC-4E, KSC LC-39A, CCAFS SLC-40
- Total payload mass carried by boosters launched by NASA (CRS) was 45596 kg
- Average payload mass carried by booster version F9 v1.1 was 2928.4 kg
- Date of first successful landing outcome in ground pad was 22/12/2015.
- There are four boosters with success in drone ship and have payload mass between 4000-6000
- There were 100 successful and 1 failed missions

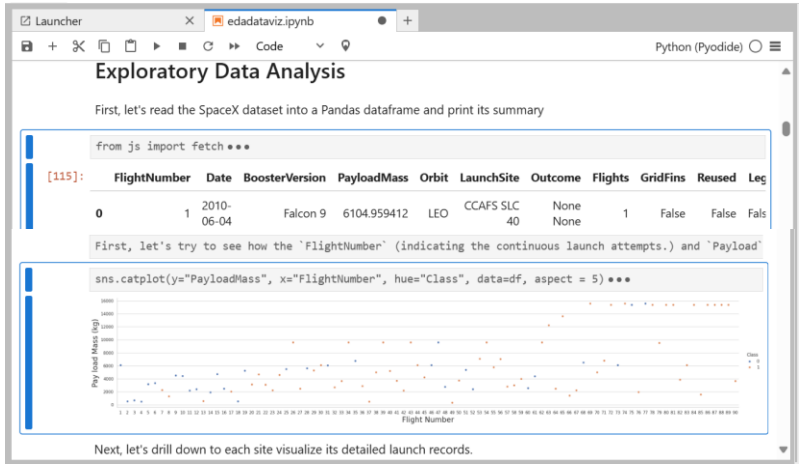
Exploratory Data Analysis (EDA)

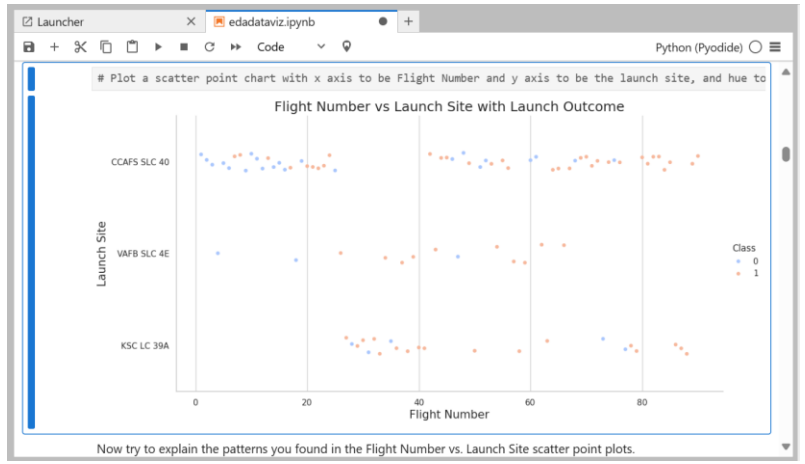


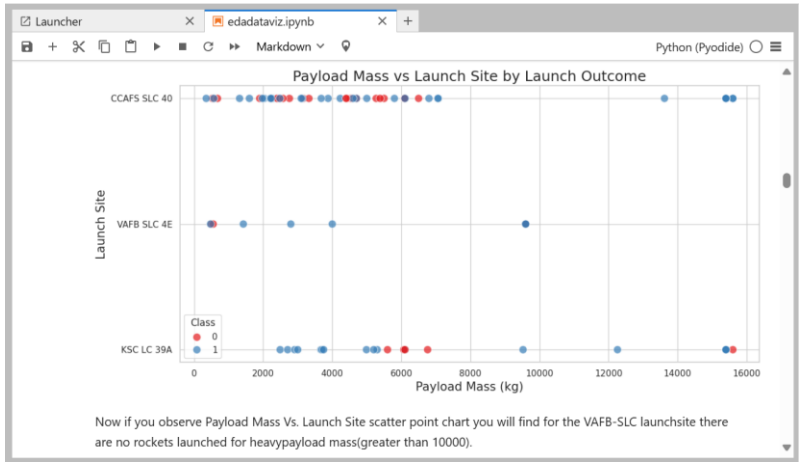


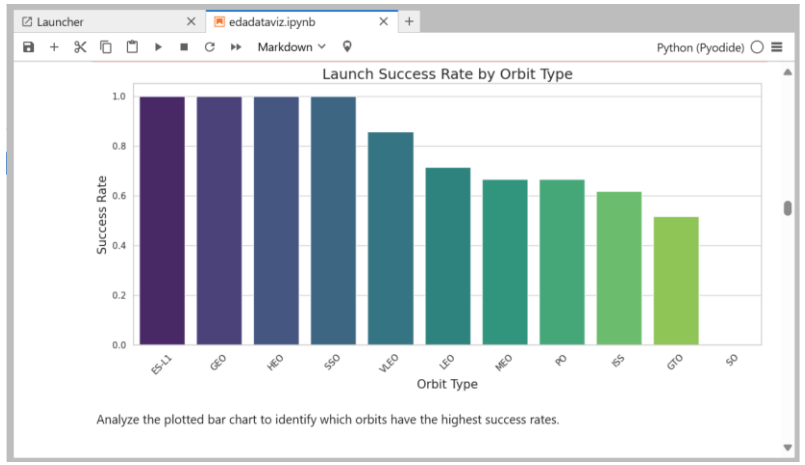
- EDA with SQL
  - There were a total of 12 booster versions that have carried the maximum payload mass
  - There were two failed landing outcomes in drone ship in 2015 (January and April) involving different boosters but at at the same launch site CCAFS LC-40.
  - Between 2010-06-04 and 2017-03-20:
    - Successful landing: Drone ship (5) > Ground pad (3)
    - Failed landing: Drone ship (5) > Parachute (2)

Exploratory Data Analysis (EDA)





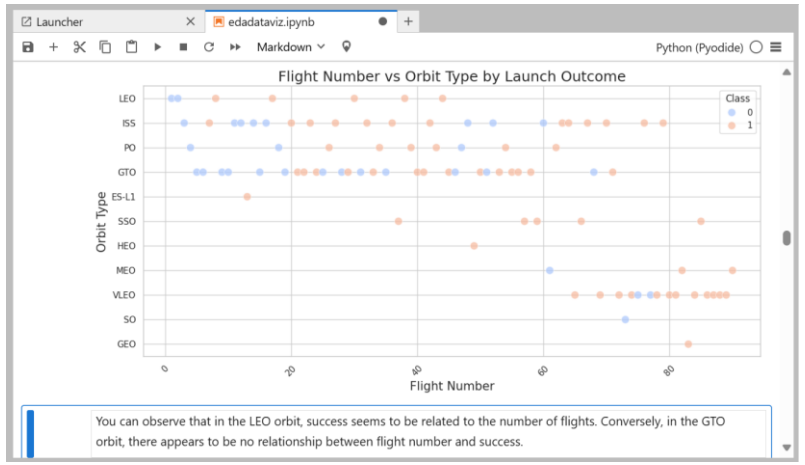


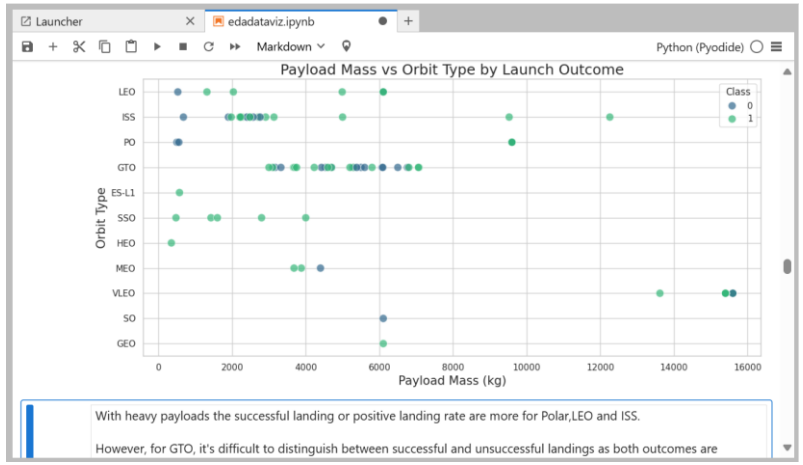


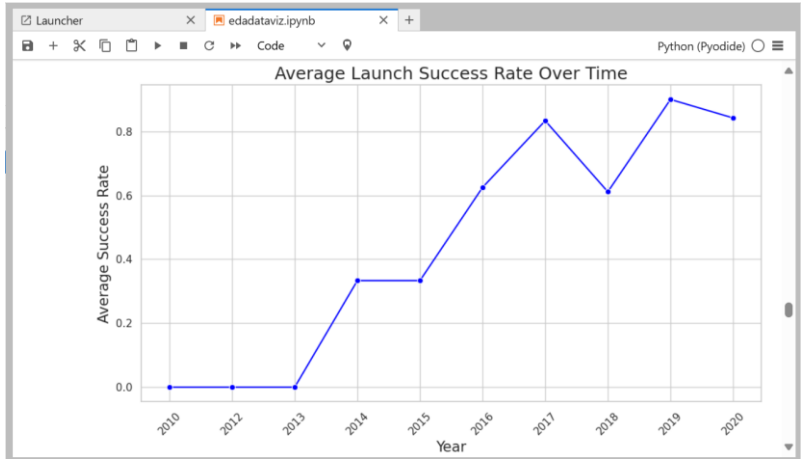
❑EDA with Visualization (Matplotlib - Plotly & Seaborn)

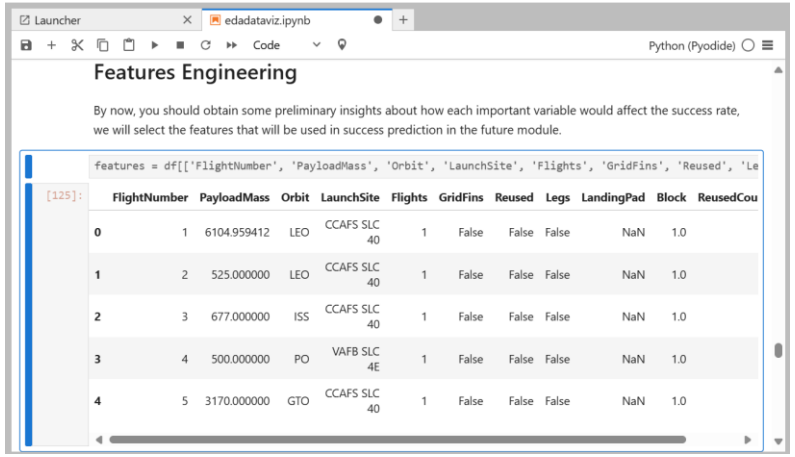
- FlightNumber and PayloadMass on launch outcomes have positive correlation
- There were more successful outcomes for all sites with increased Flight Number
- There were more successful outcomes for all sites with increased Payload Mass
- Success rate varied with Orbit Type and ranged from 1.0 (ES-L1, GEO, HEOand SSO) to 0.5 (GTO) to 0 (SO)

Exploratory Data Analysis (EDA)



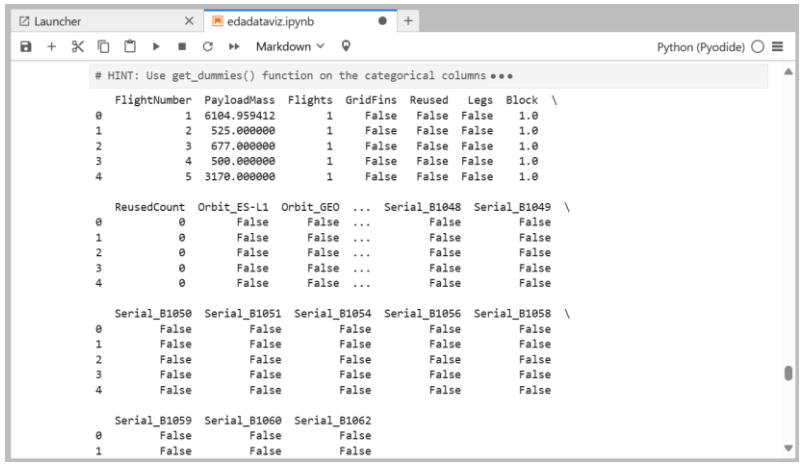


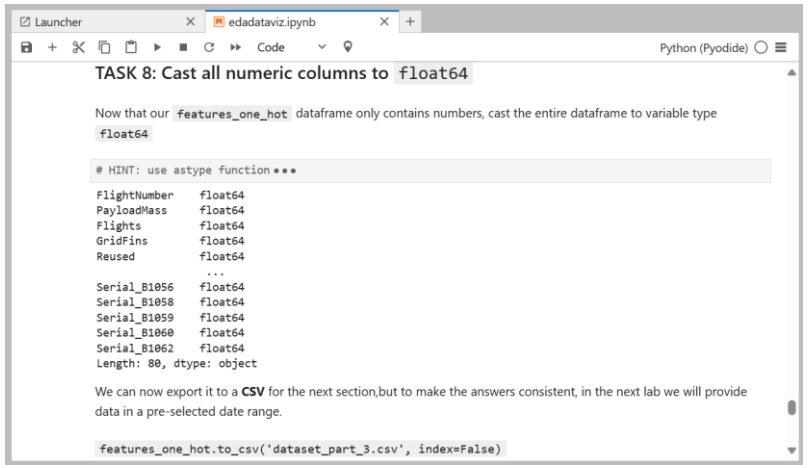




- EDA with Visualization (Matplotlib - Plotly & Seaborn)
  - There were more successful launch outcomes with increased FlightNumber for all Orbit type except for SO
  - There were more successful launch outcomes with increased Payload Mass for all Orbit type except for SO
  - Successful outcomes increased as the year increases

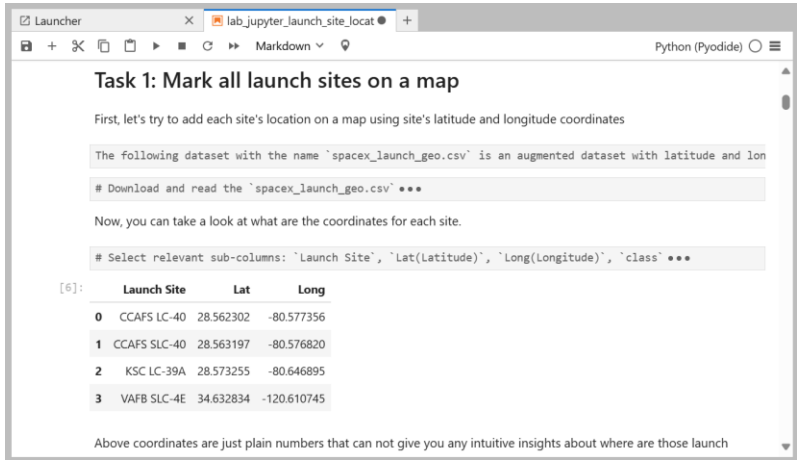
Exploratory Data Analysis (EDA)

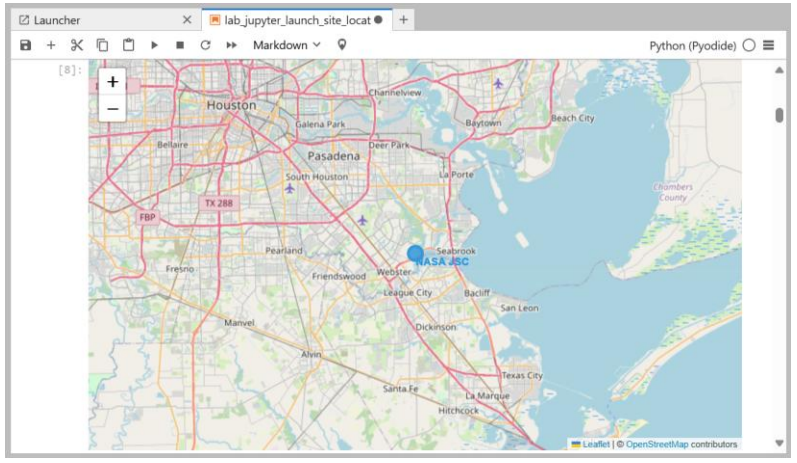


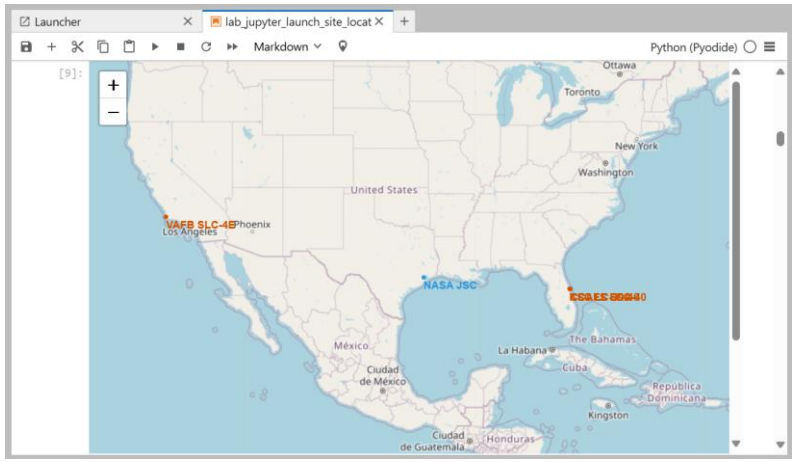


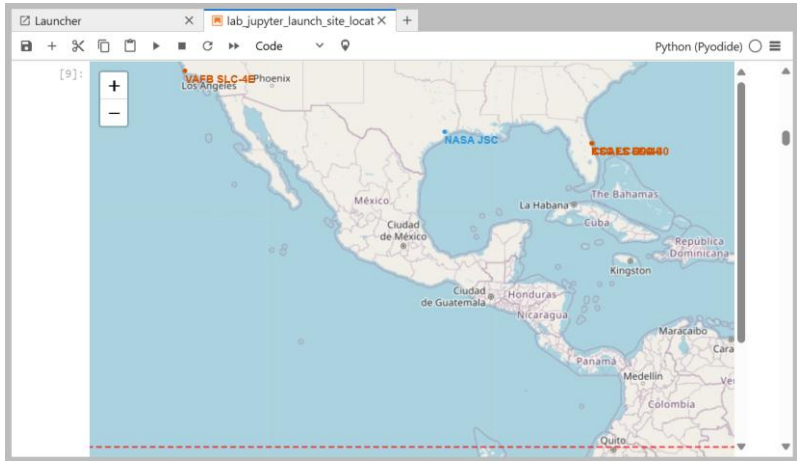
- EDA with Visualization (Matplotlib - Plotly & Seaborn)
  - Features engineering allowed for deeper dive into identifying relevant feature for predicting success rate
  - Supplementary data are available as CSV file

Interactive Visual Analytics and Dashboard



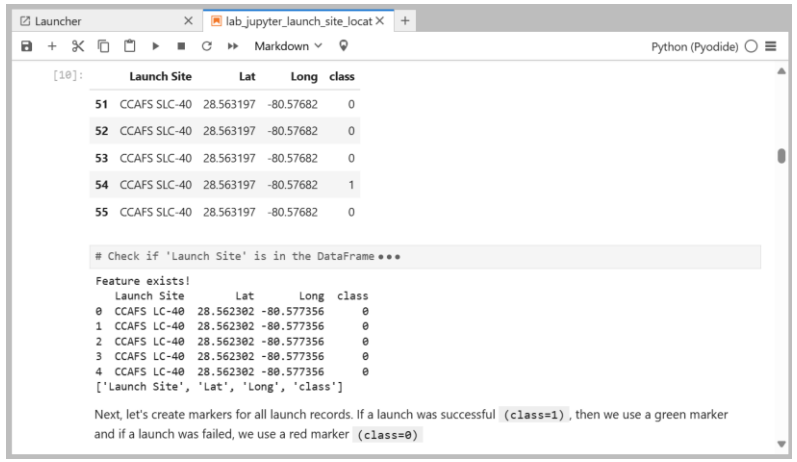


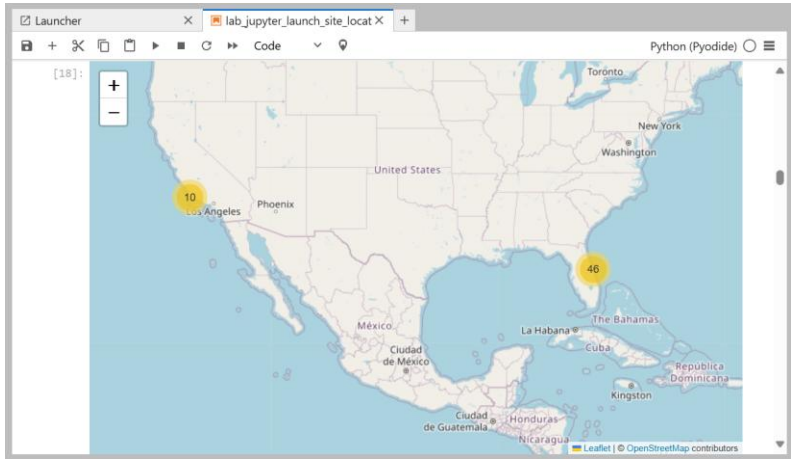


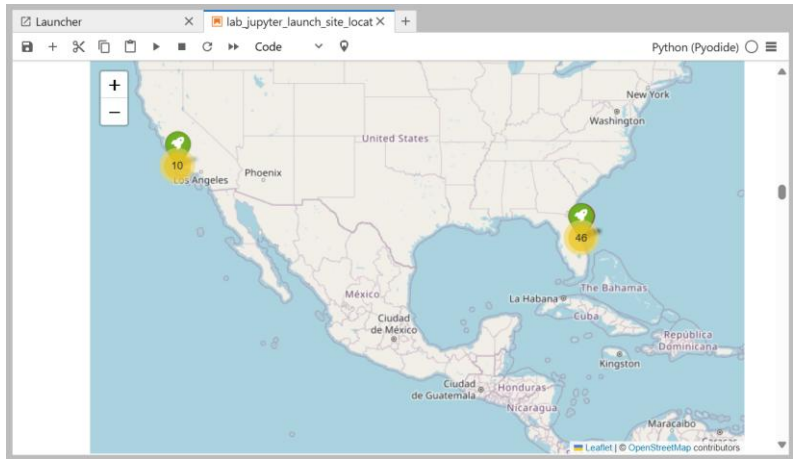


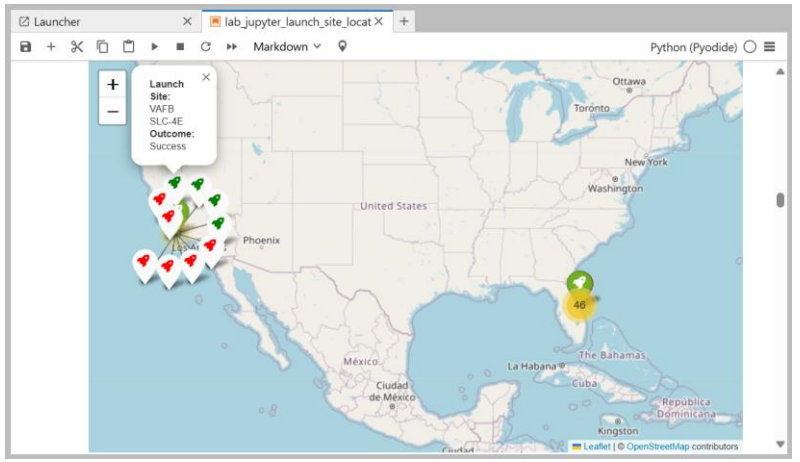
- Interactive Visual Analytics with Folium
  - Map showed all four unique launch sites and NASA JSC
  - Proximity of launch sites to the coast and equator line
    - Coast – YES, Equator – NO
    - VABF SLC-4E to the pacific ocean, CCAFS SLC-40 to the atlantic

Interactive Visual Analytics and Dashboard





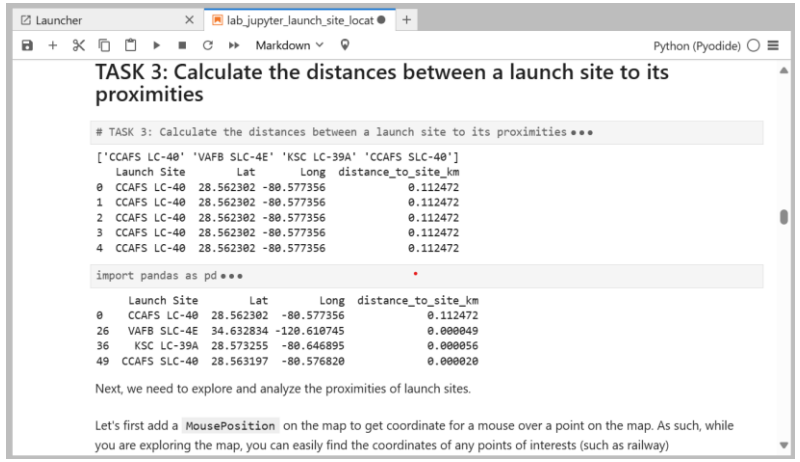


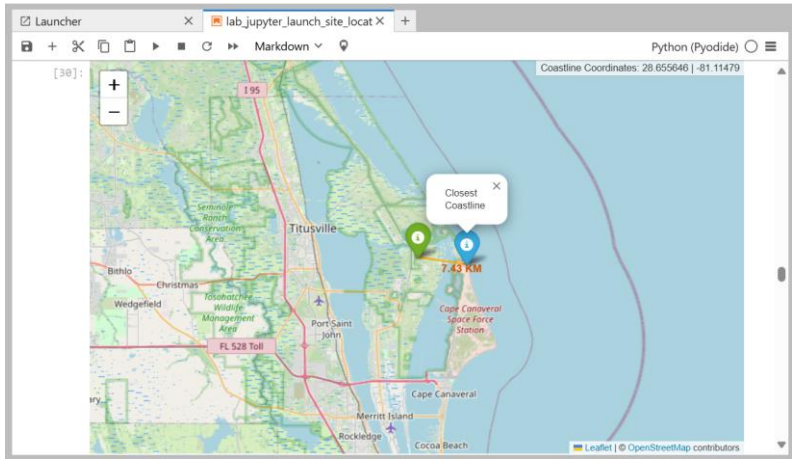


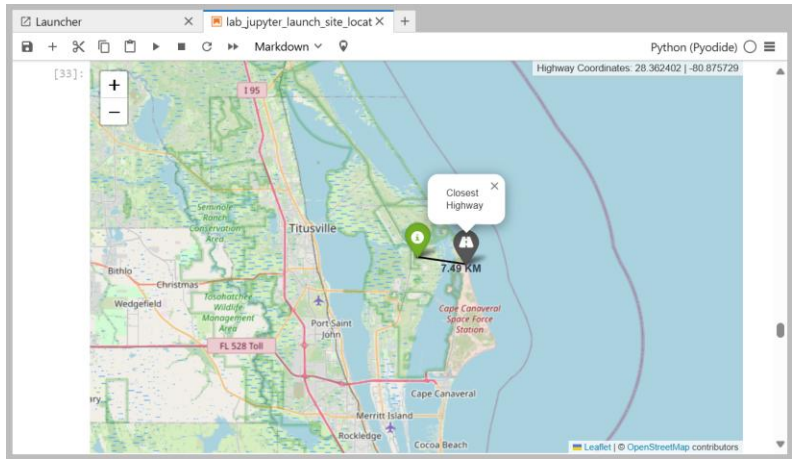
❑Interactive Visual Analytics with Folium

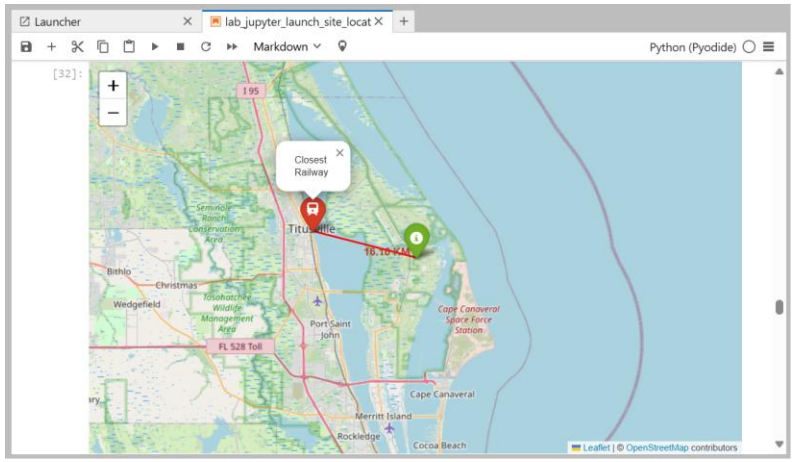
- Map showed success and failed launches for each site
- Green markers for success and red for failure
- This allowed for easy geospatial visualization and evaluation of launch outcomes

Interactive Visual Analytics and Dashboard





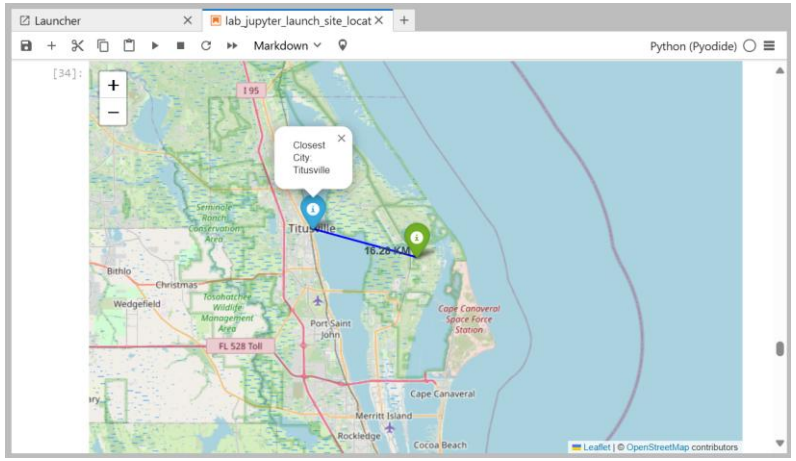


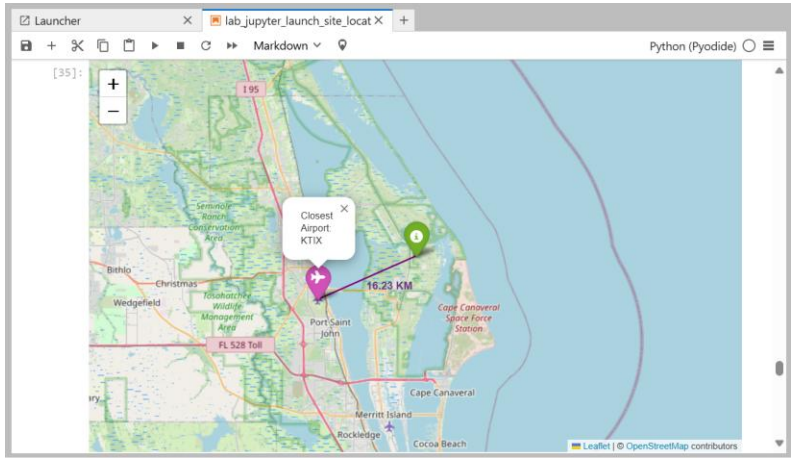


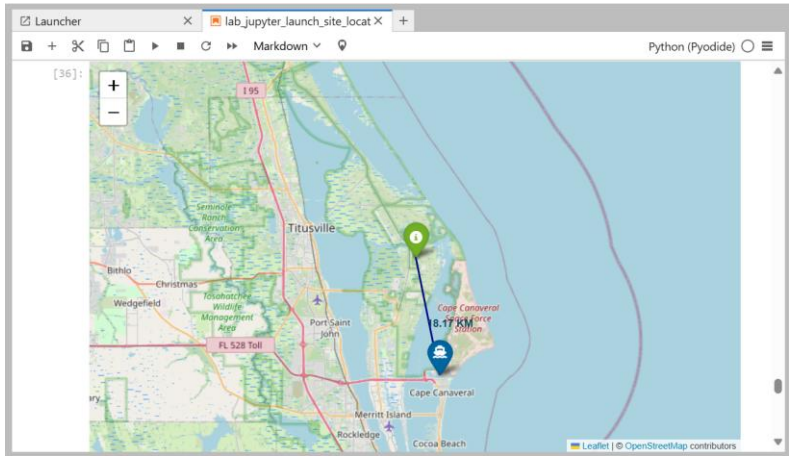
❑Interactive Visual Analytics with Folium

- Calculated distances between a launch site (e.g., CCAFS LC-40) to its proximities
  - Are launch sites in close proximity to coastline? YES, 7.43 Km
  - Are launch sites in close proximity to highways? YES, 7.49 Km
  - Are launch sites in close proximity to railways? NO, 16.10 Km

Interactive Visual Analytics and Dashboard



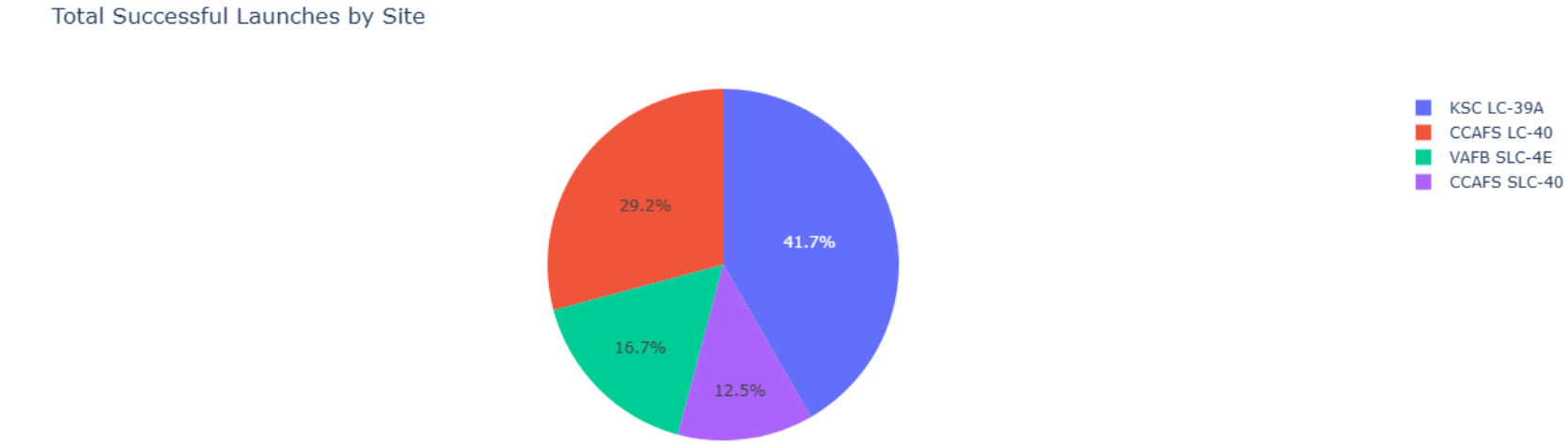


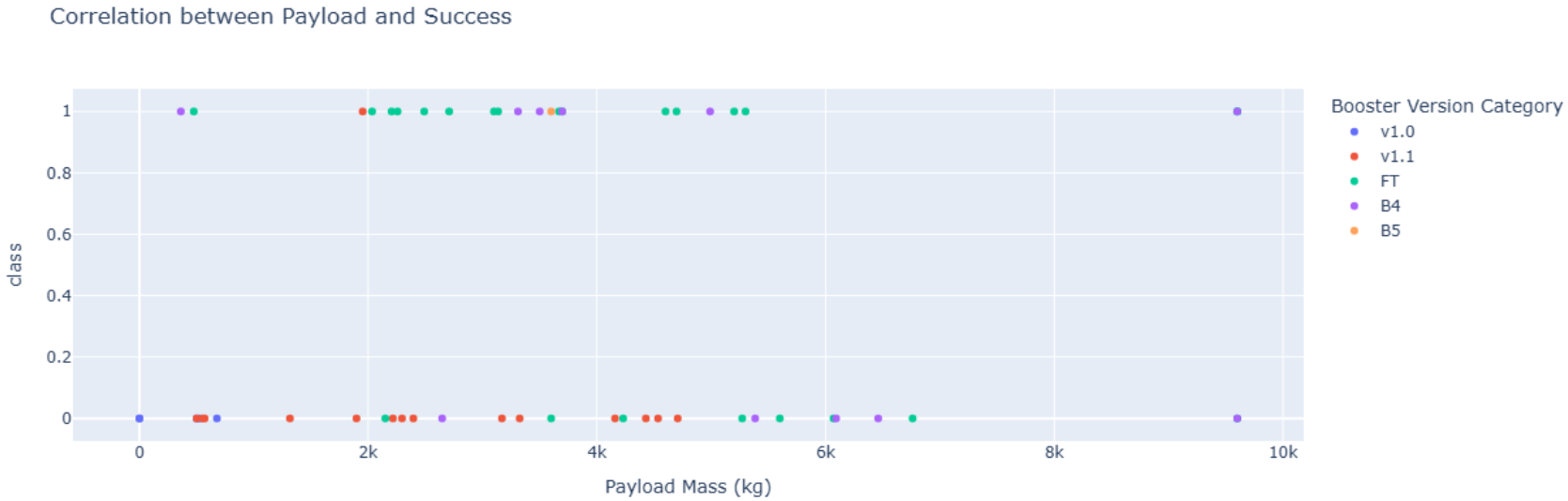


❑Interactive Visual Analytics with Folium

- TASK 3: Calculated distances between a launch site (e.g., CCAFS LC-40) to its proximities
  - Do launch sites keep certain distance away from cities? YES, 16.28Km
  - Are launch sites in close proximity to airport? NO, 16.23 Km
  - Are launch sites in close proximity to seaport? NO, 18.17 Km

Interactive Visual Analytics and Dashboard





❑Interactive Dashboard with Plotly DASH

Interactive Visual Analytics and Dashboard

- Components of the Interactive Dashboard with Plotly DASH include:
  - Launch Site Drop-down Input Component
  - Callback function to render success-pie-chart based on selected site dropdown
  - Range Slider to Select Payload on success-payload-scatter-chart scatter plot
  - Callback function to render the success-payload-scatter-chart scatter plot

Interactive Visual Analytics and Dashboard

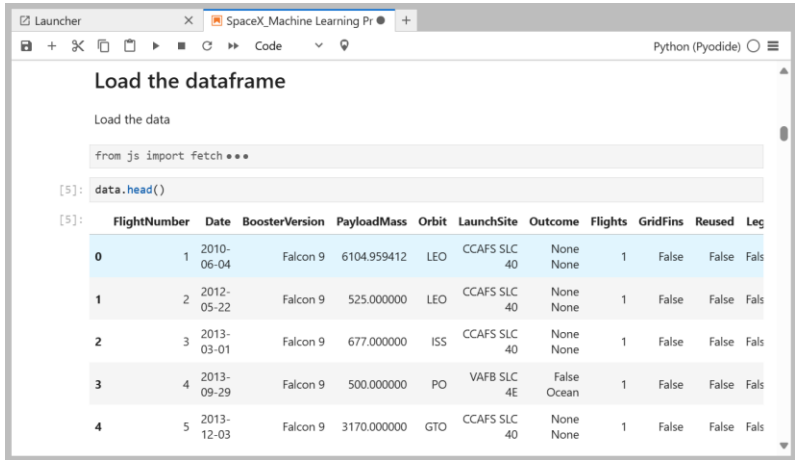
❑Visual Insights from Dashboard

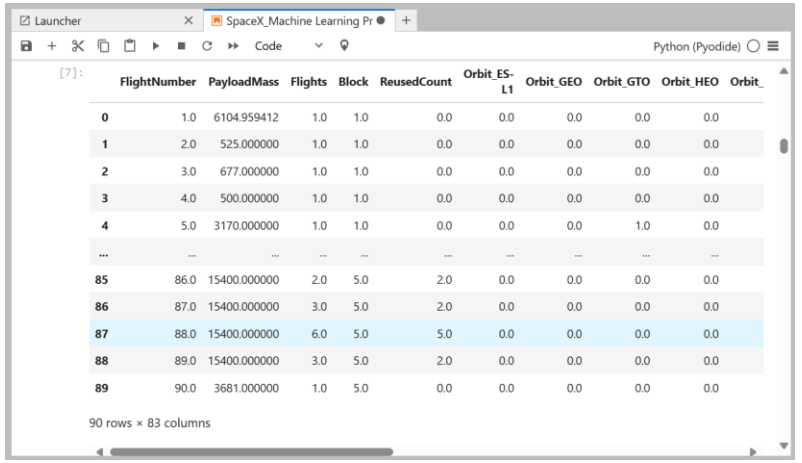
- Total successful launches by site:
  - Successful launches varied with launch sites and in the order; `
  - KSC LC-39A > CCAFS LC-40 > VAFB SLC-4E > CCAFS SLC-40
- Correlation between payload mass and success:
  - Difficult to categorically determine as this seem to be comparable outcomes between successful and failed outcomes across all payload mass. However,
  - Up to 2000 Kg, there is more failure
  - Between 2000 – 6000 Kg, relatively indistinguishable outcomes
  - Above 6000 Kg to 10000 Kg, there are more failed outcomes
- Booster Version Types:
  - Booster FT had more successful launches with payload mass, compared to others, while v1.0 and v1.1 had the least, compared to others

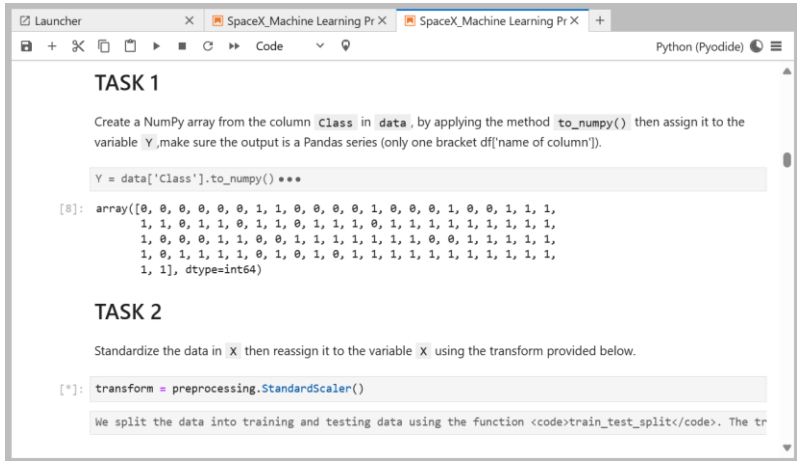
Interactive Visual Analytics and Dashboard

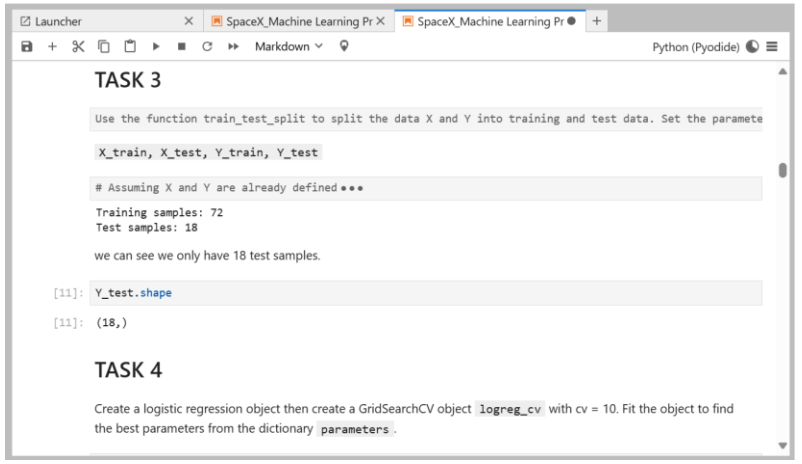
- Analyze SpaceX launch data, and answer the following questions:
  1. Which site has the largest successful launches? KSC LC-39A
  2. Which site has the highest launch success rate? KSC LC-39A
  3. Which payload range(s) has the highest launch success rate? 2000-4000 Kg
  4. Which payload range(s) has the lowest launch success rate? <2000 and >5500 Kg
  5. Which F9 Booster version (v1.0, v1.1, FT, B4, B5, etc.) has the highest launch success rate? Booster FT

Predictive Analysis (Classification)



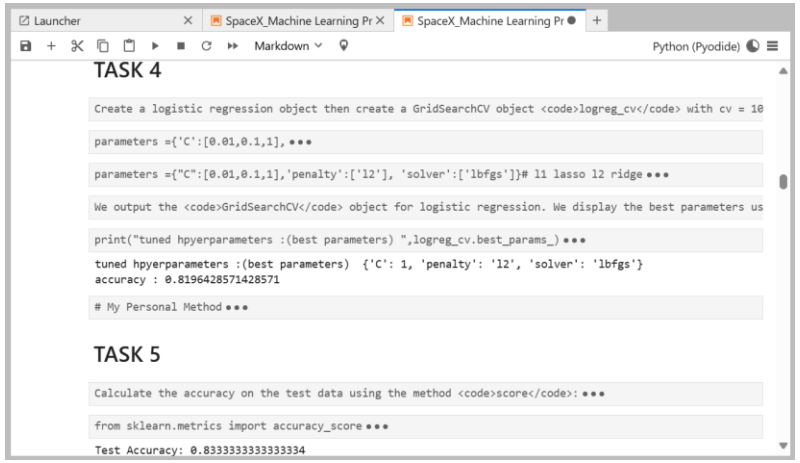


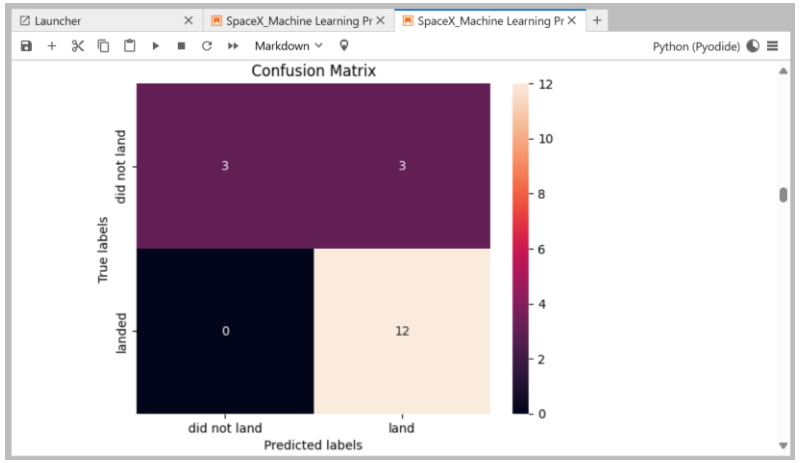


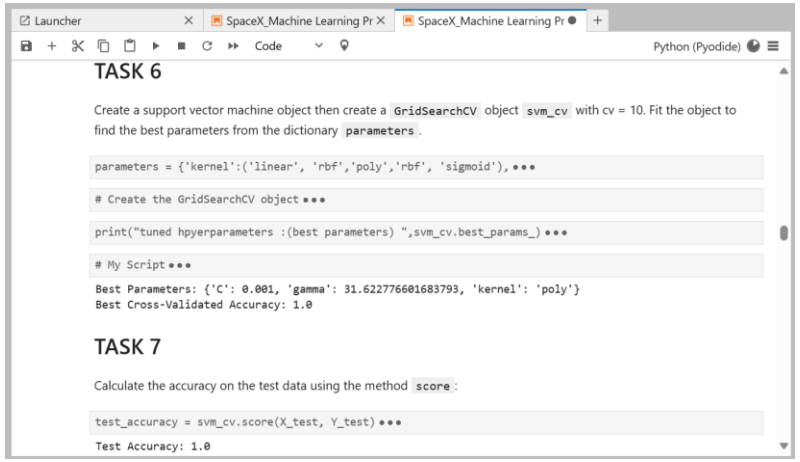


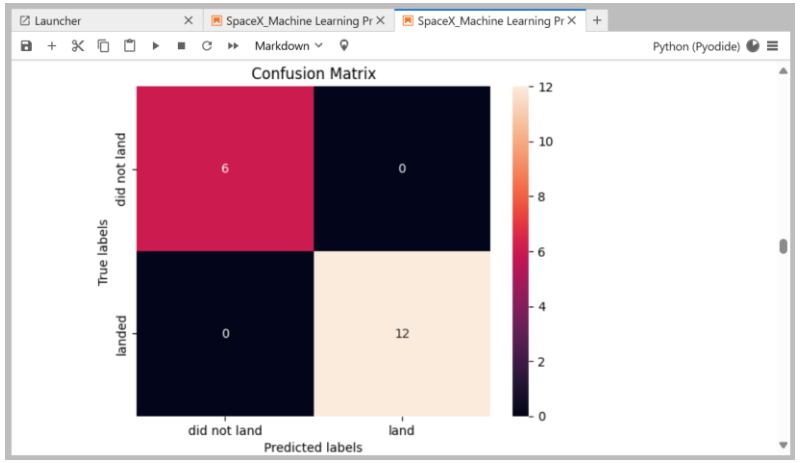
- Machine Learning Prediction
  - NumPy array from the column Class in data
  - Standardized data
  - Data X and Y split into 72 training and 18 test data

Predictive Analysis (Classification)





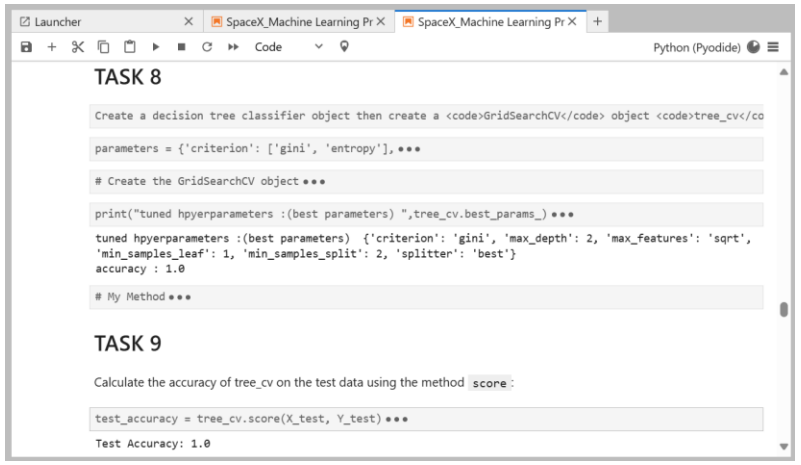


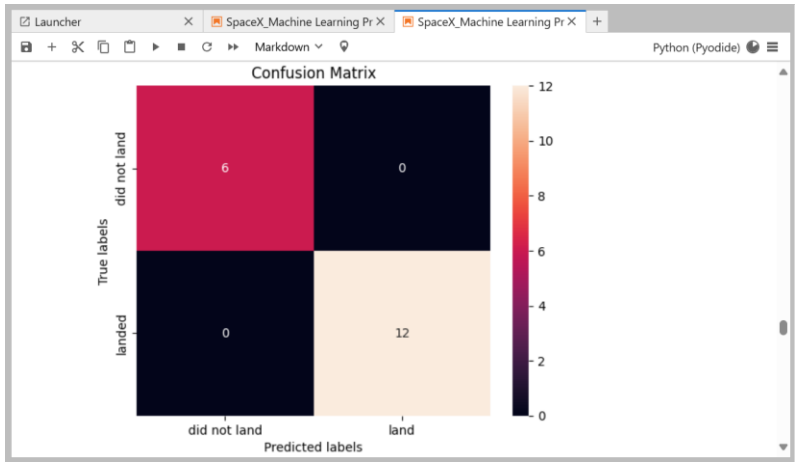


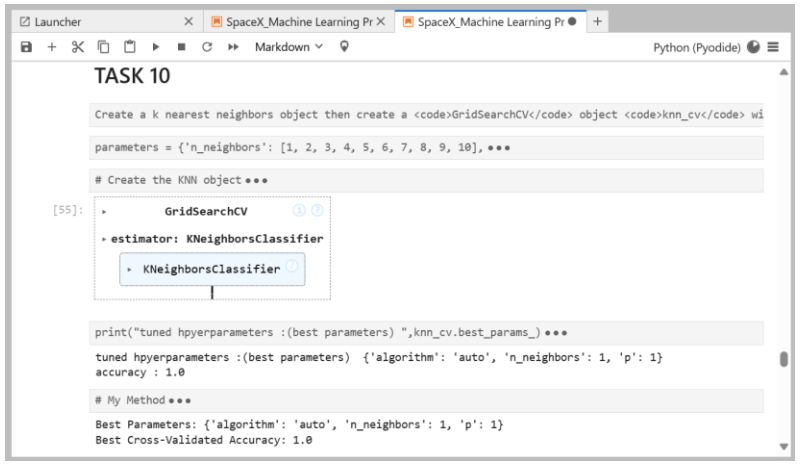
❑Machine Learning Prediction

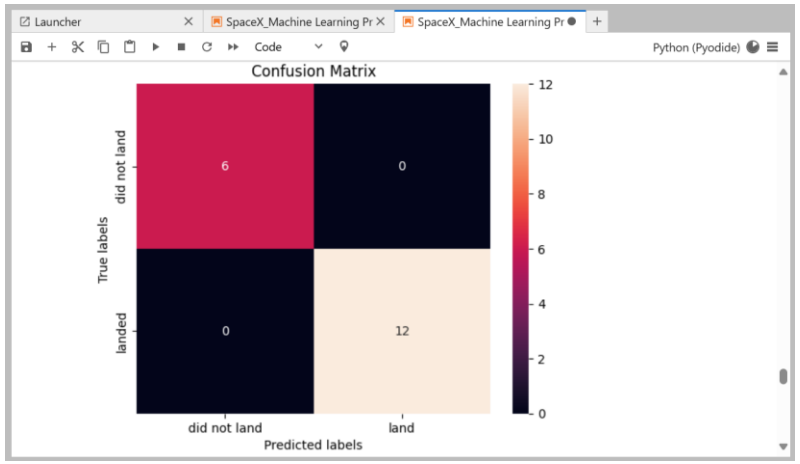
- Logistic regression object with GridSearchCV object, and Confusion matrix. Accuracy = ~0.82; Test Accuracy = ~0.83
- Support vector machine object with GridSearchCV object,and Confusion matrix. Accuracy = 1.0; Test Accuracy = 1.0

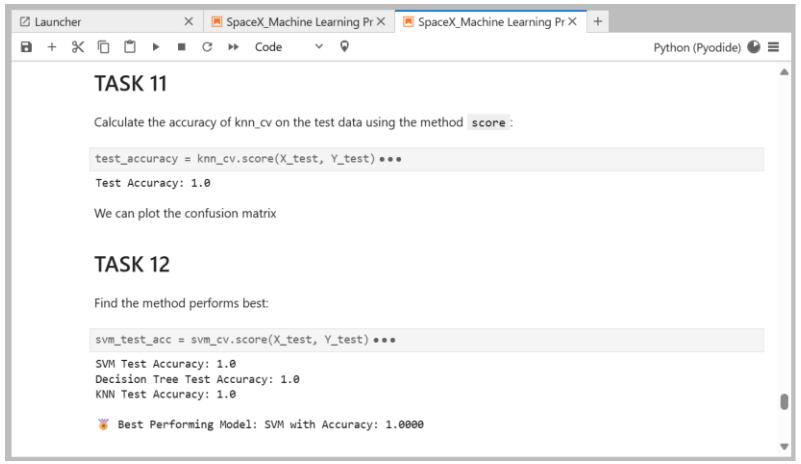
Predictive Analysis (Classification)











- Machine Learning Prediction
  - Decision tree classifier object with GridSearchCV object,and Confusion matrix. Accuracy = 1.0
  - K nearest neighbors object with GridSearchCV object, and Confusion matrix. Accuracy = 1.0
  - Best performing model is SVM

Conclusion

- This project demonstrates how geospatial and predictive analytics can uncover meaningful patterns in SpaceX launch data. The integration of SQL, Python, and interactive tools provided a rich exploration environment. Future work could include anomaly detection, launch delay prediction, and integration with weather data.

Creativity Beyond Template

- Custom dashboard theme and layout
- Animated transitions between charts
- Infrastructure overlays with dynamic tooltips
- Equator line and proximity arcs on Folium map

Innovative Insights

- Launches closer to ports and airports showed higher success rates
- Weekday launches had slightly better outcomes than weekends
- Payload mass bins revealed nonlinear success patterns
- Infrastructure proximity emerged as a strong predictor in classification model

Reference

https://github.com/Doctoroma/IBMDS_OMA.git

Thanks