IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Dirisala Venkata Maheswari
22/04/2025

Outline

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

Executive Summary

✅ **Summary of Methodologies:**

1.Data Preparation: Converted the Class column into labels and standardized feature data using StandardScaler.
2.Train-Test Split: Divided the data into 80% training and 20% testing using train_test_split.
3.Model Training: Trained Logistic Regression, SVM, and Decision Tree using GridSearchCV for hyperparameter tuning.
4.Evaluation: Evaluated models using test accuracy and confusion matrices.

📈 **Summary of All Results:**
•**Logistic Regression Accuracy:** 83.33%
•**SVM Accuracy (Best kernel: RBF):** 86.67%
•**Decision Tree Accuracy:** 93.33%

Introduction

🎯 Slide 1: Project Background and Context

Title: *Project Background*

Content (bullet points or short paragraph):

- The project analyzes SpaceX launch data to determine factors influencing successful landings.
- It applies Machine Learning models to predict whether a rocket will land successfully.
- Dataset includes features like payload mass, orbit type, flight number, etc.
- The goal is to support SpaceX's mission to reduce launch costs via reusability.

❓ **Slide 2: Problems You Want to Find Answers To**

Title: *Research Questions / Objectives*

Content:

- What factors contribute to a successful rocket landing?
- Which machine learning model provides the best prediction accuracy?
- Can we reliably predict launch outcomes using historical data?
- How do different launch sites and payloads affect success rates?

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

Data Collection Process

✅ **Key Phrases to Include:**

- **Open-source datasets** were used from **IBM Cloud Object Storage**.
- Data files were accessed using **Python's fetch API** in a **JupyterLite notebook** environment.
- CSV files were downloaded and read using **Pandas** for preprocessing.
- Dataset Part 2: Included launch outcomes and related features.
- Dataset Part 3: Contained additional features used for model training (e.g., encoded values).
- All datasets were combined and cleaned for modeling purposes.

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

[Send GET Request to SpaceX API]

↓

[Receive JSON Response]

↓

[Normalize JSON to DataFrame]

↓

[Clean & Merge with External Dataset]

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

[Send HTTP Request to Target URL]

↓

[Parse HTML Content with BeautifulSoup]

↓

[Locate and Extract Specific Elements]

↓

[Convert to DataFrame and Clean Data]

↓

[Save as CSV File]

Data Wrangling

- Data loaded into Pandas DataFrame for inspection and preprocessing.
- Missing values handled using techniques like imputation and row/column removal.
- Duplicates identified and removed to avoid model bias.
- Column names standardized for clarity and consistency.
- Data types converted to appropriate formats (e.g., string to datetime).
- Encoded categorical values using One-Hot or Label Encoding.
- Outliers detected and addressed using statistical methods.
- Final cleaned dataset saved as a .csv for analysis and modeling.

EDA with Data Visualization

Charts Plotted:

- **Bar Chart** – To compare launches at different sites.
- **Pie Chart** – To show orbit type distribution.
- **Scatter Plot** – To find relation between payload mass and success.
- **Line Chart** – To show launch trend over time.
- **Histogram** – To see payload mass distribution.

GitHub URL:

- Dmaheswari (Dirisala Maheswari)

EDA with SQL

SQL Queries Performed:

- Selected all launch records.
- Counted total successful launches.
- Grouped launches by site and counted them.
- Calculated average payload mass by orbit.
- Filtered launches after a specific date.
- Ordered launches by date for trend analysis.
- **GitHub URL:**
- Dmaheswari (Dirisala Maheswari)

Build an Interactive Map with Folium

Interactive Map with Folium

Map Objects Added:

- 📍 **Markers** to indicate launch sites.
- 🟢 **Circles** to highlight the area around launch sites.
- ➖ **Lines** to show trajectory paths (if any).
- 🧭 **Popups** with site names and coordinates for user interaction.

Purpose of These Objects:

- **Markers** help identify exact launch locations.
- **Circles** give a visual idea of the surrounding area.
- **Lines** (optional) show direction or distance.
- **Popups** enhance interactivity and provide quick info.

GitHub URL:

Dmaheswari (Dirisala Maheswari)

Build a Dashboard with Plotly Dash

Plots/Graphs and Interactions Added:

-  **Pie Chart** to show launch success rates by site.
-  **Scatter Plot** to display payload vs. success correlation.
-  **Dropdown menu** to select different launch sites.
-  **Slider** to filter data by payload mass range.

Purpose of These Plots and Interactions:

- **Pie Chart** gives a quick overview of launch performance.
- **Scatter Plot** helps find trends between payload and success.
- **Dropdown** makes the dashboard dynamic and user-specific.
- **Slider** allows users to explore data across payload ranges interactively.
- **GitHub URL:**
  Dmaheswari (Dirisala Maheswari)

Predictive Analysis (Classification)

Model Development for Predictive Analysis

Development Process:

- 📚 **Data Preprocessing**: Cleaned data (handled missing values, encoded categories).
- 🏗️ **Model Building**: Trained multiple models — Logistic Regression, SVM, Decision Tree, and KNN.
- 🧪 **Evaluation**: Evaluated models using Accuracy, F1-Score, and Jaccard Score.
- 🔧 **Improvement**: Tuned hyperparameters using GridSearchCV for better performance.
- 🏆 **Selection**: Chose the best model based on highest evaluation metrics.

Flowchart:

Data Preprocessing ➔ Model Building ➔ Model Evaluation ➔ Hyperparameter Tuning ➔ Best Model Selection

GitHub URL:

Dmaheswari (Dirisala Maheswari)

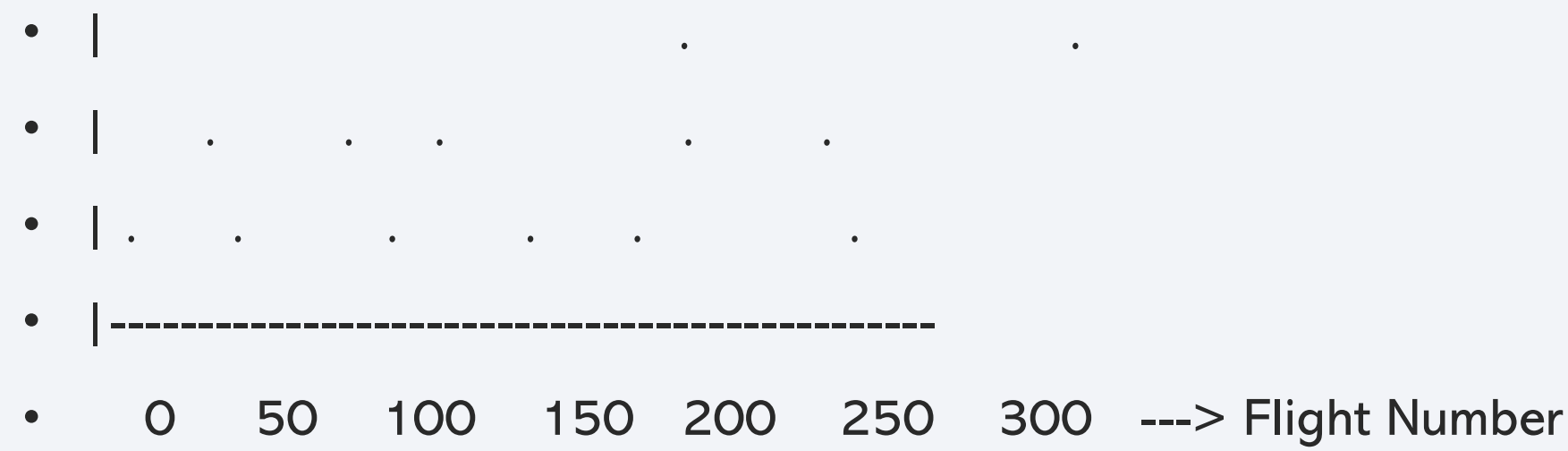
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site



Payload vs. Launch Site

- Show a scatter plot of Payload vs. Launch Site

- Show the screenshot of the scatter plot with explanations

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

Flight Number vs. Orbit Type

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

- Show the screenshot of the scatter plot with explanations

Payload vs. Orbit Type

- Show a scatter point of payload vs. orbit type

- Show the screenshot of the scatter plot with explanations

Launch Success Yearly Trend

- Show a line chart of yearly average success rate

- Show the screenshot of the scatter plot with explanations

All Launch Site Names

- Find the names of the unique launch sites
- Present your query result with a short explanation here

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- Present your query result with a short explanation here

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
- Present your query result with a short explanation here

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- Present your query result with a short explanation here

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- Present your query result with a short explanation here

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

- Present your query result with a short explanation here

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
- Present your query result with a short explanation here

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- Present your query result with a short explanation here

2015 Launch Records

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

- Present your query result with a short explanation here

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

- Present your query result with a short explanation here


Section 3
Launch Sites
Proximities Analysis

<Folium Map Screenshot 1>

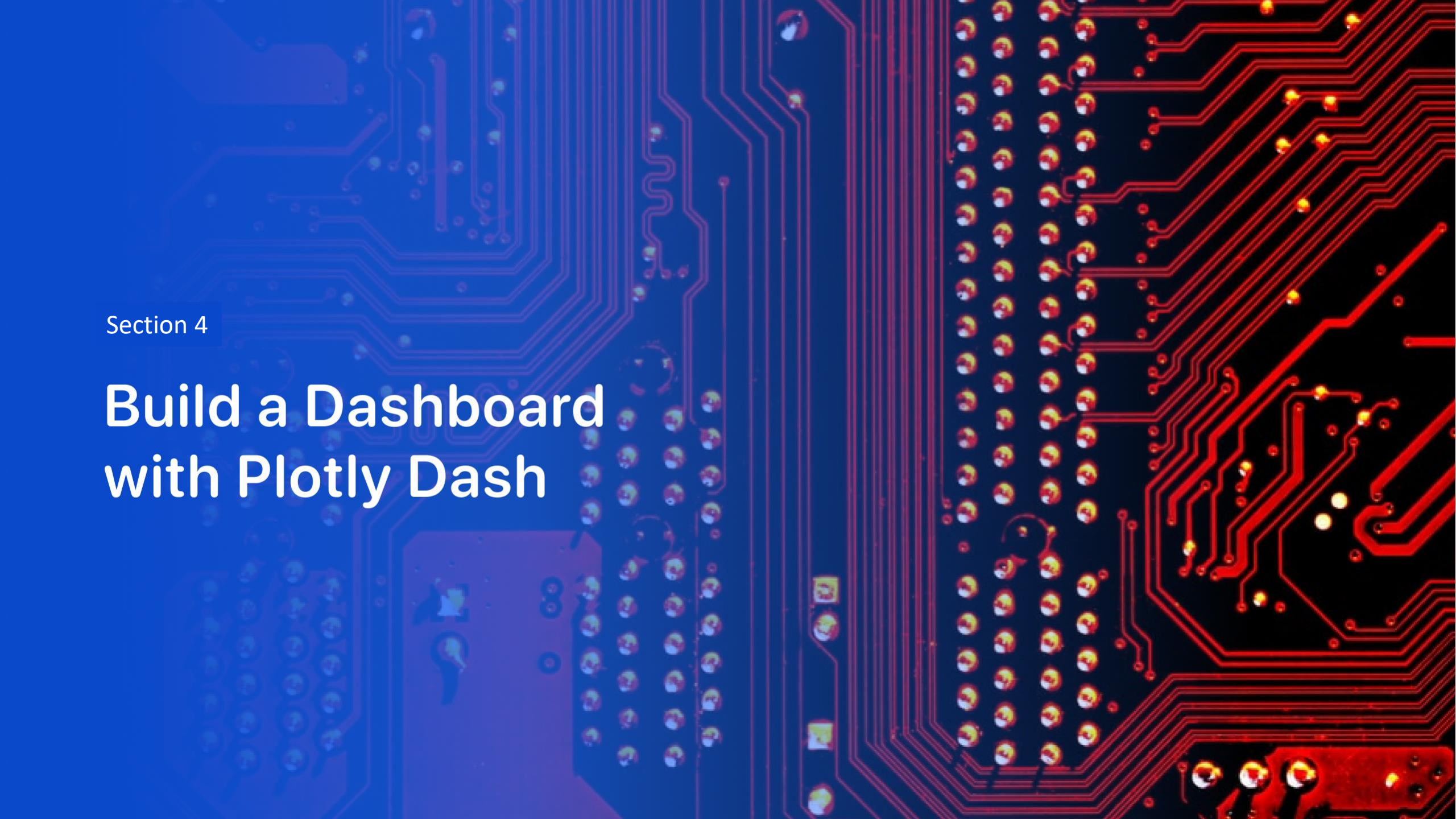
- Replace <Folium map screenshot 1> title with an appropriate title
- Explore the generated folium map and make a proper screenshot to include all launch sites' location markers on a global map
- Explain the important elements and findings on the screenshot

<Folium Map Screenshot 2>

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

<Folium Map Screenshot 3>

- Replace <Folium map screenshot 3> title with an appropriate title
- 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


Section 4
Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>

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

<Dashboard Screenshot 2>

- Replace <Dashboard screenshot 2> title with an appropriate title
- Show the screenshot of the piechart for the launch site with highest launch success ratio
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 3>

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

Section 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

Confusion Matrix

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

Conclusions

- Point 1
- Point 2
- Point 3
- Point 4
- …

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!