# **Exploring Coffee Quality Data with Power BI**

### **Project Overview**

This project focuses on exploring and visualizing coffee quality data from various countries using advanced analytical and visualization tools such as **Google Colab** (for data cleaning and analysis) and **ArcGIS Maps for Power BI** (for geographical visualization). The goal is to provide insights into how different factors such as geography, altitude, and variety impact coffee quality across the globe.

## **Table of Contents**

1. Project Description
2. Tools and Technologies
3. Dataset Information
4. Steps to Reproduce
   * Data Cleaning and Preprocessing
   * Exploratory Data Analysis (EDA)
   * Power BI Visualization
   * ArcGIS Map Configuration
5. Power BI Visualizations
6. Future Work
7. Acknowledgments

## **Project Description**

Coffee quality data from various coffee-growing countries was explored and analyzed using Python in Google Colab for data preparation. The cleaned data was then exported to **Power BI**, where a series of visualizations, including **ArcGIS Maps**, were created to highlight trends in coffee quality by country, altitude, and variety.

The main objectives of the project are:

* **Data Cleaning and Transformation**: Handling missing values, outlier detection, and data preprocessing.
* **Geographical Analysis**: Using ArcGIS Maps for Power BI to visualize the global distribution of coffee quality scores.
* **Insight Generation**: Providing key insights into coffee-growing countries, regions, and factors affecting coffee quality.

## **Tools and Technologies**

The project utilizes the following tools and technologies:

* **Google Colab**: For data cleaning, analysis, and processing using Python.
* **Pandas**: For data manipulation and cleaning.
* **Matplotlib & Seaborn**: For data visualization.
* **Power BI**: For dashboard creation and reporting.
* **GitHub**: For version control and code sharing.

## **Dataset Information**

The dataset contains information about coffee quality scores from various countries. Key columns include:

* Country\_of\_Origin: The country where the coffee is grown.
* Cup\_Score: The overall quality score of the coffee (out of 100).
* Altitude: The altitude at which the coffee is grown.
* Variety: The variety of coffee.
* Processing\_Method: The method used to process the coffee.

**Source**: You can use an open dataset from [Kaggle](https://www.kaggle.com) or a similar platform with coffee quality data.

## **Steps to Reproduce**

### **1. Data Cleaning and Preprocessing**

* **Code Location**: The data preprocessing and analysis are done in a Jupyter Notebook using Google Colab.
* **File**: coffee\_quality\_eda.ipynb
* **Steps**:
  1. Upload the raw coffee dataset to the Colab environment.
  2. Clean the data by handling missing values, removing duplicates, and handling outliers.
  3. Create exploratory visualizations (scatter plots, heatmaps) to understand relationships between variables.

python

Copy code

# Upload the dataset

uploaded = files.upload()

# Load the dataset

df = pd.read\_csv('coffee\_quality\_data.csv')

# Data cleaning steps, outlier detection, and handling missing values...

df = df.dropna(subset=['Cup\_Score'])

df = df.drop\_duplicates()

### **2. Exploratory Data Analysis (EDA)**

* **Visualizations**: Scatter plots, heatmaps, bar charts, and descriptive statistics were created to explore the dataset and identify trends.

python

Copy code

# Scatter plot for Altitude vs Cup\_Score

sns.scatterplot(x='Altitude', y='Cup\_Score', data=df)

plt.title('Effect of Altitude on Coffee Quality')

plt.show()

### **3. Power BI Visualization**

* Import the cleaned dataset into Power BI (cleaned\_coffee\_quality\_data.csv).
* Create interactive visualizations, including:
  + A bar chart for average Total\_cup\_pints by Country\_of\_Origin.
  + A scatter plot for Altitude vs. Cup\_Score.
  + Filters and slicers to explore the data by different attributes such as Variety and Processing\_Method.

### **4. ArcGIS Map Configuration**

* **Using ArcGIS Maps for Power BI**:
  1. Import the dataset into Power BI and select the **ArcGIS Maps** visual.
  2. Add the Country\_of\_Origin as the **Location** field and Total\_cup\_pints as the **Size** or **Color** field.
  3. Customize the map with reference layers, themes, and tooltips to enhance the visual presentation.

python

Copy code

# Optional Python code for exporting cleaned data

df.to\_csv('cleaned\_coffee\_quality\_data.csv', index=False)

files.download('cleaned\_coffee\_quality\_data.csv')

## **Power BI Visualizations**

The following key visualizations were created:

1. **Bar Chart**: Displaying average coffee quality scores by country.
2. **Scatter Plot**: Showing the effect of altitude on coffee quality.
3. **ArcGIS Map**: Visualizing the global distribution of coffee quality scores, with interactive layers and reference data.
4. **Slicers and Filters**: Allowing users to filter the data by Variety, Processing Method, or other attributes.

## **Future Work**

This project can be extended by:

* Incorporating more datasets, such as coffee production volume and export data, to analyze the relationship between quality and production.
* Using machine learning models to predict coffee quality scores based on various factors like altitude, country, and processing method.
* Enhancing the ArcGIS Map with custom geographic layers or deeper spatial analysis.