Maji Ndogo Agricultural Innovation Project 4

Overview of Project

The Maji Ndogo farming automation project aims to revolutionize agriculture by leveraging data science to make informed decisions about crop cultivation. The project focuses on analyzing various factors such as rainfall, soil type, climate, and geographical data to optimize farming practices in Maji Ndogo.

Project Goal

The goal of the project is to optimize farming practices, increase crop yields, and contribute to sustainable agriculture in the Maji Ndogo . To achieve this we explore, analyse and visualise our agricultural dataset to find patterns in our data and draw out insights.

Key Features of Project

- Importing and cleaning the dataset
- Exploratory analysis (Univariate and Multivariate) of the features in our datasets.
- Validating the real-time data on weather conditions in MAji Ndogo

SQLite Database Tables

- **Geographic features:** Contains information about the geographic features of each field, such as its elevation, latitude, longitude, location (province), and slope.
- **Weather features:** Contains weather-related information for each field, such as rainfall, minimum temperature, maximum temperature, and average temperature.
- **Soil and crop features**: Provides information about the soil and crop features of each field, including soil fertility, soil type, pH level, chosen crop for cultivation, and annual yield.
- **Farm management features:** This table contains farm management-related information, such as pollution level in the area, plot size, chosen crop for cultivation, annual yield, and standardized yield expected from the field.Weather_station_data (CSV):

Weather Data (CSV files)

- **Weather_station_data**: Contains information that identifies a particular weather station and the weather data captured by sensors at that weather station in text fromat messages.

- **Weather_data_field_mapping**: Provides information that helps us identify and map a particular field in Maji Ndogo to the weather station that is connected to that field (closest to it).

Tools Used

- Python
- Jupyter notebooks/VS code/Google Collab
- Pandas library
- SQLite Database Files
- SQLalchemy package
- Re package for Regular expressions
- Seaborn and Matplotlib packages

What I did

1. The project began by importing data from an SQLite database into a single DataFrame in Python using pandas and sqlalchemy. A comprehensive dataset was created by joining multiple tables using a single SQL query. The data was then cleaned to address issues such as swapped column names, spelling errors in crop types, and negative elevation values.

2. With our data cleaned, I began the exploratory analysis by using the **.info()** and **.describe()** methods to perform univariate analysis on all features of our datasets. The subplot method of matplotlib was used to generate a series of visualisations of the **Kernel density estimation** (KDE) plots for each numerical feature in our dataset. I used the hue function in Seaborn to take a closer look at the Rainfall data distribution segregating it by Location, and Crop_ type.

3. I performed some multivariate analysis to gain an understanding of tthe relationships between features in our datasets. I made use:

    a. Violin plots to compare the distribution of a continuous variable across different levels of a categorical variable e.g. **Rainfall** and **Crop_type.**

b. Pair plots to visualise relations among continuous variables e.g **Standard_yield** and **Elevation.**

c. Crosstabs to visualise relations among categorical variables e.g **Location** and **Crop_type.**

d. Correlations to identfy the strength and direction of relationships between numeric varables

NB: See insights for a description of the insights and observations drawn from our analysis.

4. To validate our weather data. We imported two new csv files (See Weather Data (CSV files) section for info on the CSV files:
    a. Weather_station_data.csv
    b. Weather_data_field_mapping.csv

5. Once the data has been imported I used Regex to extract all of the measurement types (Rainfall, Temperature, and Poluution levels) and their values from the weather station data. (Temp. Reading [2023-05-23 09:41:36]: Current 14.53 C is an example of text messages to extract measurement and values from)
6. With the measurement value extracted I calculated the mean Temperature, Rainfall and Pollution_level for each weather station from the weather data.
7. To be able to compare and validate our data we need to:
    a. First merged our main dataset with the Weather_data_field_mapping.csv mapping data to create a Weather_station_ID column.
    b. Calculated the mean values for each measurement type per weather station using the groupby function.
    c. Used the already created **within_tolerance_percentage** and **check_means** functions to check the means calculated from our merged datasets and compare it to the means calculated from the weather data to determine if they are within 1.5% of one another.

Conclusion

The Maji Ndogo Farming Automation Project aims to improve agriculture in the region using information from an agricultural dataset. By analyzing factors like rainfall, soil type, and climate, the project aims to gain valuable insights into optimizing farming practices and increasing crop

yields to make informed decisions about crop cultivation. Validating our data is also an essential step in determining the accuracy of the data used for analysis.