Data Science Project 7PAM2002-0509-2023
Semester C 2023
Logbook (Activities and GitHub submissions)

Student Name and ID: Vinoth Rajendran - 22022031

Project Title: Analysing Historical Weather Data for Climate Trends and Abnormalities in the UK (2020-2024)

Supervisor: Dr Calum Morris

Student GitHub URL: https://github.com/vr22abb/MSc-Data-Science-Project

Number of versions of the code submitted on GitHub: 10

User documentation has been submitted on GitHub: YES

Student GitHub URL has been shared with markers: YES

Log of Activities
Must record attendance at lectures and supervisions

Week	Date	Activity incl. lectures & supervisions	Reason if not attend lecture or supervision	Weekly project progress. How lecture/supervision was helpful to your project.
1	14/05/2024	Lecture 1		• Overview of the Module • How to choose a Project • Supervisor Allocation
2	21/05/2024	Lecture 2		• Project and Data Management Plan • Data Selection • Module Scoring
3	28/05/2024	Supervision 1		• Project Dataset Discussion • How to Choose data for project • Module Queries • Discussed the reason to change the topic
4	03/06/2024	Lecture 3		• Data Ethics lecture (part 1) • Knowledge on Ethics • Basic idea on PDM Plan
5	17/06/2024	Lecture 4		• Data Ethics lecture (part 2) • UK GDPR law • UH Ethical Policy
6	18/06/2024	Supervision 2		• PDM Plan Presentation • Q&A - Got Feedback
7	25/06/2024	Supervision 3		• Explained project status • Discussed best practices

8	09/07/2024	Supervision 4		• Explained the Project Progress • Discussed the changes on code
9	15/07/2024	Lecture 5		• Results and analysis lecture • Got knowledge on Report preparation
10	23/07/2024	Supervision 5		• Showed the code progress and the report • Got clarified on doubts
11	30/07/2024	Supervision 6		• Mock VIVA • Got feedback

Log of GitHub Submissions

Record the versions of code and user documentation submitted on GitHub

Date	**Filename and version submitted to GitHub**	**Description of code and/or documentation submitted (what has been added since the previous version).**
27/08/2024	USER DOCUMENTATION OF THE CODE.docx - **653773c**	Detailed description document of the code.
10/06/2024	MSc Project - 22022031.ipynb - **1fd248c** (version 1)	**1. Code Description (Version 1):** • Mount Google Drive and Import Libraries • Load Datasets • Inspect Data • Identify Missing Values • Handle Missing Values • Check for Remaining Missing Values 2. **Key Features in Version 1:** • Data Loading from CSV Files • Basic Data Inspection and Information Display • Missing Data Identification • Missing Data Imputation using Mean and Mode • Data Preparation for Further Analysis
24/06/2024	MSc Project - 22022031.ipynb - **b005914** (version 2)	**1. Code Description (Version 2):** • Mount Google Drive and Import Libraries • Load and Preprocess Datasets • Display Summary Statistics • Plot Histograms for Key Variables • Filter and Aggregate Data • Plot Aggregated Data with Dual Y-Axes • Load Processed Monthly and Yearly Data • Plot Grouped Parameters **2. New Features and Enhancements Added in Version 2:** • Data Saving and Reloading • Forward Fill and Data Type Conversion • Advanced Visualization Techniques • Aggregated Data Analysis • Expanded Analysis to More Parameters

24/06/2024	MSc Project - 22022031.ipynb - **87d2b7f** (version 3)	**1. Code Description (Version 3):** • Plot Monthly and Yearly Aggregated Data • Load Processed Monthly Dataset • Filter and Group Monthly Data by Month • Function to Plot Grouped Parameters • Define Grouped Parameters and Their Full Names • Plot Grouped Parameters by Month • Load Processed Yearly Dataset • Filter and Group Yearly Data by Year • Plot Grouped Parameters by Year **2. New Features and Enhancements Added in Version 3:** • Refined Data Loading and Indexing • Dynamic Plotting Function (plot_grouped_parameters) • Detailed Grouping and Analysis by Month and Year • Improved Visual Representation of Trends • Separation of Monthly and Yearly Analysis
12/07/2024	MSc Project - 22022031.ipynb - **b158762 (**version 4)	**1. Code Description (Version 4):** • Mount Google Drive and Import Libraries • Set Global Plot Parameters and Define Column Metadata • Load and Preprocess Datasets • Display Summary Statistics for Each Dataset • Plot Histograms with Detailed Statistics • Plot Aggregated Data with Dual Y-Axes • Function to Plot Grouped Parameters by Month and Year • Filter and Group Monthly and Yearly Data • Plot Grouped Parameters by Month and Year • Categorize Data by Seasons and Plot Seasonal Trends • Detect Anomalies Using the IQR Method • Plot Anomalies for Daily, Monthly, and Yearly Data **2. New Features and Enhancements Added in Version 4:** • Global Plot Settings for Consistent Visualization • Detailed Histogram Plots with KDE, Bell Curve, and Statistics • Seasonal Trend Analysis and Plotting • Anomaly Detection Using the Interquartile Range (IQR) Method • Enhanced Plotting Functions for Anomaly Visualization
18/07/2024	MSc Project - 22022031.ipynb - **7fcee0a** (version 5)	**1. Code Description (Version 5):** 1. Mount Google Drive and Import Libraries 2. Global Definitions for Column Names and Units 3. Load and Preprocess Datasets with Error Handling 4. Display Summary Statistics for Each Dataset 5. Plot Histograms with Detailed Statistics

		6. Calculate and Plot Monthly Variability for Parameters 7. Calculate and Plot Yearly Variability for Parameters 8. Categorize Data by Seasons and Plot Seasonal Trends 9. Plot Grouped Parameters by Month and Year 10. Detect Outliers Using Z-Scores and Plot Outliers 11. Summarize and Visualize Outlier Information 12. Perform Trend Analysis on Temperature and Precipitation 13. Forecast with Prophet and Plot Forecasts for Daily, Monthly, and Yearly Data **2. New Features and Enhancements Added in Version 5:** 1. Error Handling in Data Loading and Preprocessing 2. Advanced Variability Analysis (Monthly and Yearly) 3. Outlier Detection Using Z-Scores 4. Comprehensive Outlier Summary and Visualization 5. Trend Analysis with Linear Regression 6. Time Series Forecasting with Prophet Model 7. Enhanced Forecasting Visualization with Prophet
18/07/2024	MSc Project - 22022031.ipynb - **cde7630** (version 6)	**1. Code Description (Version 6):** 1. Mount Google Drive and Import Libraries 2. Global Definitions for Column Names and Units 3. Load and Preprocess Datasets with Error Handling 4. Display Summary Statistics for Each Dataset 5. Plot Histograms with Detailed Statistics 6. Calculate and Plot Monthly Variability for Parameters 7. Calculate and Plot Yearly Variability for Parameters 8. Categorize Data by Seasons and Plot Seasonal Trends 9. Plot Grouped Parameters by Month and Year 10. Detect Outliers Using Z-Scores and Plot Outliers 11. Summarize and Visualize Outlier Information 12. Perform Trend Analysis on Temperature and Precipitation 13. Prepare Data for Prophet and Forecast 14. Forecast with Prophet and Plot Forecasts for Daily, Monthly, and Yearly Data **2. New Features and Enhancements Added in Version 6:** 1. Enhanced Data Loading with Error Handling 2. Detailed Histogram Plots with KDE, Bell Curve, and Statistics 3. Advanced Variability Analysis (Monthly and Yearly) 4. Outlier Detection Using Z-Scores and Summary Visualization 5. Trend Analysis with Linear Regression 6. Time Series Forecasting with Prophet Model 7. Improved Forecasting Visualization with Prophet

07/08/2024	MSc Project - 22022031.ipynb - **d644310** (version 7)	**1. Code Description (Version 7):** 1. Import Libraries and Set Up Environment 2. Global Definitions for Column Names and Units 3. Data Loading and Preprocessing with Error Handling 4. Fill Missing Values with Mean 5. Display Summary Statistics for Each Dataset 6. Visualizations: Plot Histograms with Detailed Statistics 7. Categorize Data by Seasons and Plot Seasonal Trends 8. Calculate and Plot Monthly Variability for Parameters 9. Calculate and Plot Yearly Variability for Parameters 10. Aggregate Data and Plot Correlation Matrices 11. Perform Clustering and Dimensionality Reduction (PCA) for Daily, Monthly, and Yearly Data 12. Plot Clusters for Daily, Monthly, and Yearly Weather Data **2. New Features and Enhancements Added in Version 7:** 1. Data Aggregation and Correlation Matrix Visualization 2. K-Means Clustering and PCA for Dimensionality Reduction 3. Comprehensive Seasonal and Variability Analysis 4. Enhanced Visualizations for Clusters and Correlation 5. Detailed Data Exploration and Preparation for Advanced Modelling
15/08/2024	MSc Project - 22022031.ipynb - **3befcaa** (version 8)	**1. Code Description (Version 8):** 1. Importing Libraries and Setting Up Environment 2. Global Definitions for Column Names and Units 3. Data Loading and Preprocessing with Error Handling 4. Filling Missing Values with Mean 5. Displaying Summary Statistics for Each Dataset 6. Visualization: Plotting Histograms with Detailed Statistics 7. Categorizing Data by Seasons and Plotting Seasonal Trends 8. Calculating and Plotting Monthly Variability for Parameters 9. Calculating and Plotting Yearly Variability for Parameters 10. Seasonal Decomposition for Trend Analysis 11. Identification and Visualization of Extreme Events 12. Correlation Matrix Visualization for Daily, Monthly, and Yearly Data 13. K-Means Clustering and Dimensionality Reduction (PCA) for Weather Data 14. Outlier Detection and Visualization using Z-Scores

		15. Forecasting Using Prophet Model (Daily, Monthly, and Yearly Data) 16. Forecasting Using ARIMA Model (Daily, Monthly, and Yearly Data) 17. ARIMA Model Parameter Tuning with Grid Search 18. Visualization of Forecasting Results and Future Predictions 19. Comparison of RMSE Values Across Different Forecasting Models **2. New Features and Enhancements Added in Version 8:** 1. Incorporation of ARIMA Model with Parameter Tuning using Grid Search 2. Detailed Visualization of Forecasting Results and Confidence Intervals 3. Comprehensive Analysis of Extreme Weather Events 4. Advanced Seasonal Decomposition for Enhanced Trend Analysis 5. Enhanced Forecasting Techniques with Prophet and ARIMA 6. Comparison of Forecasting Performance (RMSE) Across Models 7. Improved Visualizations for Clustering, Outlier Detection, and Forecasting
27/08/2024	MSc Project - 22022031.ipynb- **9ca92c3** (version 9)	**1. Code Description (Version 9):** 1. Importing Libraries and Setting Up Environment 2. Global Definitions for Column Names and Units 3. Data Loading and Preprocessing with Error Handling 4. Filling Missing Values with Mean 5. Displaying Summary Statistics for Each Dataset 6. Visualization: Plotting Histograms with Detailed Statistics 7. Categorizing Data by Seasons and Plotting Seasonal Trends 8. Calculating and Plotting Monthly Variability for Parameters 9. Calculating and Plotting Yearly Variability for Parameters 10. Seasonal Decomposition for Trend Analysis 11. Identification and Visualization of Extreme Events 12. Correlation Matrix Visualization for Daily, Monthly, and Yearly Data 13. K-Means Clustering and Dimensionality Reduction (PCA) for Weather Data 14. Outlier Detection and Visualization using IQR and Z-Scores 15. Forecasting Using Prophet Model (Daily, Monthly, and Yearly Data) 16. Forecasting Using ARIMA Model with Parameter Tuning

		17. Advanced LSTM Modelling with Cross-Validation and Hyperparameter Tuning 18. Ensemble Learning: Combining Multiple Models for Enhanced Forecasting 19. Future Predictions and Visualization with Confidence Intervals 20. Comparison of RMSE and MAE Across Different Forecasting Models **2. New Features and Enhancements Added in Version 9:** 1. Bidirectional LSTM and Hyperparameter Tuning using Bayesian Optimization 2. Advanced Feature Engineering with Rolling Statistics 3. Enhanced Forecasting with Ensemble Learning Techniques 4. Detailed Visualization of Predictions and Confidence Intervals 5. Advanced Outlier Detection Methods (IQR, Z-Scores) 6. Model Performance Comparison and Visualization 7. Extended Future Predictions with Confidence Intervals
27/08/2024	MSc Project - 22022031.ipynb- **3b34c3f** (version 10)	**1. New Features and Enhancements Added in Version 10** 1. Improved Doc string and comments