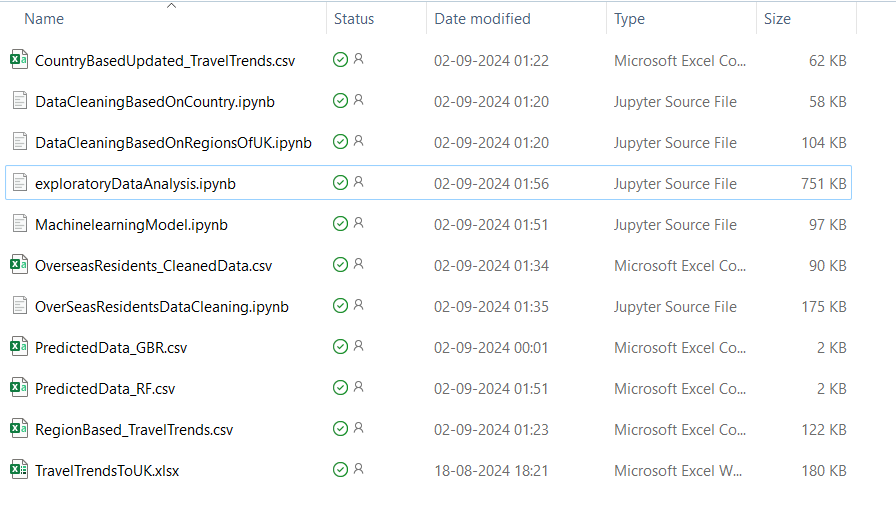
The OneDrive Link: [BEMM466BusinessProjectVaibhavi730074762](https://universityofexeteruk-my.sharepoint.com/:f:/g/personal/vr311_exeter_ac_uk/EgfmU0xsut9Ms9gUOFUoFKEBkTHHeOEWwNyIQN5kfMpKxA?e=2dx2Lm)



Description and flow of project files:

**Parent Data Source:**

File: TravelTrendsToUK.xlsx

Description: This file serves as the original dataset containing all travel trend data to the UK. It is the primary source from which the data cleaning and processing steps are derived.

**Data Cleaning Files:**

**DataCleaningBasedOnCountry.ipynb:**

Purpose: This Jupyter notebook cleans and processes the travel data by organizing it based on countries.

Output: The cleaned data is stored in CountryBasedUpdated\_TravelTrends.csv.

**DataCleaningBasedOnRegionsOfUK.ipynb:**

Purpose: This notebook cleans the data by categorizing it according to different regions within the UK.

Output: The cleaned data is stored in RegionBased\_TravelTrends.csv.

**OverSeasResidentsDataCleaning.ipynb:**

Purpose: This notebook focuses on cleaning the data specifically related to overseas residents.

Output: The cleaned data is stored in OverseasResidents\_CleanedData.csv.

**Exploratory Data Analysis (EDA):**

File: exploratoryDataAnalysis.ipynb

Purpose: After the data cleaning process, this notebook is used to carry out the Exploratory Data Analysis on the cleaned datasets. It helps in understanding the data trends, patterns, and preparing it for machine learning models.

**Machine Learning Models:**

Purpose: Implements the Random Forest algorithm to predict tourism trends and spending patterns.

The predictions generated by this model are stored in PredictedData.csv.

Purpose: Implements the Gradient Boosting algorithm to predict tourism trends and spending patterns. The predictions generated by this model are stored in PredictedData\_GBR.csv.

**Outputs:**

**PredictedData.csv:**

Description: Contains the predictions made using the Random Forest model.

**PredictedData\_GBR.csv:**

Description: Contains the predictions made using the Gradient Boosting model.

**Additional Cleaned Data Files:**

**CountryBasedUpdated\_TravelTrends.csv:**

Description: Output from DataCleaningBasedOnCountry.ipynb, contains the cleaned travel data categorized by country.

**RegionBased\_TravelTrends.csv:**

Description: Output from DataCleaningBasedOnRegionsOfUK.ipynb, contains the cleaned travel data categorized by regions within the UK.

**OverseasResidents\_CleanedData.csv:**

Description: Output from OverSeasResidentsDataCleaning.ipynb, contains the cleaned data specific to overseas residents.