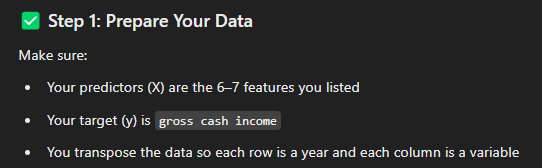
**Instructions and next steps:**

* So far with the uploaded files, we have two cleaned data sets, cleanProductionExpensesData.xlsx and cleanIncomedata.xlsx. These were made using expenseDataCleaning.py and incomeDataCleaning.py.
* Then we also have EDA (eda\_analysis.py) here we plotted the gross cash income variable at the very top of the cleanIncomeData against some independent variables from the cleanProductionExpensesData to see how different costs affect income.
* From this we found 7 of the highest correlated features such as property taxes & fees, intermediate product expenses > pesticide, etc. These 7 will be our random forest regressor’s predictor variables

**Next:**

* 
* Step 2: Train test split - be careful since this is time series data
* Step 3: Build and train the random forest
* Step 4: feature importance, and performance check (R^2 close to 1 and low RMSE)