Wrangling Report

I began the wrangling process by first exploring the data frames visually. Starting with the columns, looking through to see if the column names were labelled as I wanted then making notes of what columns where to be dropped. I decided to do the visual assessment in jupyter notebook instead of importing it to external program, this was done by using the df.head().

I had 3 data frames to assess and clean, so I wanted to keep a similar flow throughout. However, this was a little tricky at times as certain findings need further assessing.
A tidy dataset means:

- Each variable forms a column
- Each observation forms a row
- Each type of observational unit forms a table

I split my assess and clean into 3 subsections, with each data frame having it's own. Then throughout my assessment, I made bullet points with a list of what needed to be cleaned. This was split into quality and tidiness issues to make it easier to clean. Also, there wouldn't be the need to clean certain columns if they're going to be dropped. However, some columns were cleaned even though they wasn't used as I thought there was a chance I would need them for analysis.

This was what I kept in mind and could make note of what columns needed merging through visual assessment. The merging was just done by creating a new column with “df[“new_col”] = “, then I used concatenation as the columns I was using had no overlap.

I cleaned irrelevant and unwanted data by using “df.drop()”. This was mostly decided through visually looking at what columns I wasn't going to use.

The main code I used for my assessment process were:

- df.head()
- df.info()
- df[“col_name”].value_counts()
- df.drop(columns=[“col_name”], inplace=True)

Copies were made of each data frame before cleaning. Then before analysis, I made a new data frame with all required columns I needed.