Excel description:

'raw_marketing_data' tab contains the row data from the ifood_df.csv file, downloaded from Kaggle is on tab.

'marketing_data' tab contains the final data set for analysis after cleaning the data, adding new columns, dealing with missing values and formatting. The following is a discretion of the data cleaning and engineering done to create the final data set:

- **Arranging and formatting columns:**

Year_Birth – converted to the **Age** using: =ROUNDDOWN(YEARFRAC(D7,NOW()), 0)

Income – converted to number from text using: =VALUE(SUBSTITUTE(E2251,"$",""))

Da_Customer – converted from text to date with: data tab >> text to columns

Country – Removed for lack of relevance to project goals

- **Added columns:**

Age_Group – Added column using VLOOKUP

Income_Group - Added column using VLOOKUP

- **Dealing with missing values and outliers:**

Outliers were detected in Age and Income columns by plotting an **Age to Income** plot in the 'customer_exploration' tab. Outliers were marked in red on the plot.

Age:

- Outliers – removing 3 outliers ages 119 and above

Income:

- Missing values – using python **sklearn kNNImputer** (see python script on jupyter notebook titled **'kNNImputer income'**)
- Outliers – removing 1 outlier

'Stats_Vlookup' tab contains the statistical data about the **Age** and **Income** columns created using the Descriptive Statistic tool and the tables used to create the **Age_Group** and **Income_Group** columns using VLOOKUP.

'Customer_Explroation' tab contains initial visual analysis of each column to get a first description on the distribution of different population parameters.

'Products_Exploration' tab contains visual analysis of different product purchases by different population parameters.

'Purchases Vs Precntages' tab contains a visual comparison of pie charts showing the distribution of product purchases is identical to the general distribution of the different population parameters.

'Purchases_Exploration' tab contains visual analysis of different venues purchases by different population parameters.

'Campaign_Exploration' tab contains visual analysis of different campaign acceptance by different population parameters.