Data Cleaning Documentation

Orders data:

- PURCHASE_TS
  - Used date function to normalize the date. One row had nonsensical data, /N, that was made blank to align with other blank rows
  - Used this column to extract PURCHASE_MONTH and PURCHASE_YEAR through the use of date functions
- USER_ID
  - Some user ids were in a scientific notation format. Changed all rows to general number format
- REFUND_TS
  - Some dates are in the future. Set these values to blank
  - Used IF function to create a helper column, 0 if REFUND_TS is null and 1 if REFUND_TS is not null
- PRODUCT_NAME
  - 27in 4k gaming monitor had typos which were manually cleaned by filtering and renaming records
- REGION
  - Used vlookup function to match the COUNTRY_CODE to the region in the country_lookup table
- USD_PRICE
  - There are 158 records with $0. I decided to keep these records as the other columns still have useful data for analysis

Country lookup:

- REGION
  - Changed North America to NA
  - Row 181 had the value 'x'. Changed this to NA
  - Countries BJ and BM were blank. Searched online to find that these countries are in EMEA and NA respectively to fill in the values