

Data Cleaning & Preparation Documentation

Tools used: Excel

Orders_data Table:

- User_ID & Order_ID:
 - Converted scientific notation in ID columns to a general number format for consistency.
- Purchase_TS:
 - Replaced a nonsensical /N value with a blank to align with other missing values.
- Marketing_CHANNEL and ACCOUNT_CREATION_METHOD:
 - Updated blank values to "unknown" using filters and find-and-replace.
- Product_Name:
 - Consolidated variations of "27in 4K Gaming Monitor" into a single consistent name.
- Refund_TS:
 - Added a Region column using VLOOKUP from the Country_Lookup table. Resolved blanks by referencing the Currency_Code and manually filling regions based on corresponding countries.
- Region:
 - Pulled in Regions as a separate column from the Country_Lookup Tab using VLOOKUP. There were several rows where lookups returned blanks. For these, these were because a corresponding country code was not provided. For these, I looked at the Currency code, searched which country the currency code corresponded to, and filled in the Region where that country resides.
- Duplicate Order IDs:
 - Removed true duplicates (14% rows identical across all columns).
- USD_Price \$0 transactions:
 - Removed 158 rows with \$0 transactions.

Country_Lookup Table:

- Standardized "North America" to "NA" for consistency across the dataset.
- Updated erroneous region values (x) for the US to "NA."
- Filled blanks for country codes BJ and BM with "Benin" (EMEA) and "Bermuda" (NA) by researching their regions.