Revenue Potential in Amazon UK

Data Curation by Anonymous

# General Dataset Information:

File Name: amz\_uk\_processed\_data.csv

Description: Amazon UK Products Dataset 2023

Dataset Details: 1,680,129 Rows & 10 Columns

Size: 635,804KB ( 620MB )

Source: Kaggle - [Dataset Link](https://www.kaggle.com/datasets/asaniczka/amazon-uk-products-dataset-2023)

# Data Profile:

* Import Dataset to DBeaver SQLite Database
* Examine dataset through SQL
* Query essential data associated with problem statements
* Export data into Excel
* Apply Excel filter to the dataset
* Examine each attribute's unique value for inconsistency
* Notable Features ( Prior to Export )
  + 217 Unique Product Category
  + Large Dataset for each category
  + Reviews - Concrete evidence of buyers
  + isBestSeller - Possibly gauge product popularity
  + broughtInLastMonth - Possible gauge for potential revenues

**Recorded Inconsistency:**

* Some records consist of text strings in numerical fields.
* Exaggerated unit prices are suspected to be outliers and are test listings.
* Many records have exaggerated unit price with no reviews.
* Some records have exaggerated unit price with many reviews.

# 

# Data Wrangling:

* Using SQL, sorted attributes by field types. Some attributes contain text string fields when it should be numerical. 10 Records were filtered.
* Feature Engineered “RevenueByReviews” column by multiplying Reviews and Price attributes together.
* Feature Engineered “PotentialRevenue” column by multiplying BroughtInLastMonth and Price attributes together.
* Feature Engineered “Year Recorded” column, extracted the year from the “Date Recorded” attribute.
* Filter “CategoryName” for any records that pertains to “Art Supply”. Kids' Art & Craft Supplies, Handmade Artwork, Arts & Crafts, and Handmade Gifts category filtered.
* Isolate CategoryName records and export the results to Excel. 46,324 records were isolated.

**Data Table Schema:**

| Field | Type | Description |
| --- | --- | --- |
| ASIN | STRING | Product Identifier from Amazon |
| title | STRING | Title of the Product |
| imgUrl | STRING | URL of the product image |
| productURL | STRING | URL of the product amazon page |
| stars | REAL | Product Rating |
| reviews | INTEGER | Number of Reviews |
| price | REAL | Current unit price of the product |
| isBestSeller | BOOLEAN | Whether the product had the Amazon BestSeller status |
| boughtInLastMonth | INTEGER | Number of products last month according to Amazon |
| categoryName | STRING | Name of the category this product belongs to |