

Data Cleaning and Analysis of a Global Electronics Retailer Using Excel

Safiul Azam Bhuiyan

Table of Contents

About Dataset	1
Data Profile and Preparation Plan	1-3
Build a relational model.....	3-4
Enrich & explore the data	4-5
Build an interactive report	5
Analysis & Insight.....	6
Trends and Patterns:.....	7
Task Requirement	7
Summary.....	7

About Dataset

Global Electronics is a U.S.-based retail store specializing in electronic products, including audio devices, computers, home appliances, cell phones, cameras, camcorders, games, toys, and TVs. As the company grows, its customer and product databases expand, necessitating better monitoring of marketing performance and data-driven decision-making.

To address this, Global Electronics has provided access to primary data in CSV files. These files contain tables with transaction, product, customer, store, and currency exchange rate information from 2016 to 2021.

FILE TYPES: CSV

DATA STRUCTURE: Multiple tables

OF RECORDS 62,884

OF FIELDS 37

(Source: Kaggle)

Data Profile and Preparation Plan

The first objective is to ingest the data from raw CSV files, conduct data profiling, perform feature engineering and QA, and build a custom calendar to track performance by day, week, month, quarter, and year.

Task 1: Connect to the **Sales** CSV file and profile the data.

How many orders were recorded? Over what time period? Do you notice anything interesting about delivery dates?

- Connect to the Sales CSV file.
- Enable column quality and column distribution views.
- Ensure headers and data types are correct.
- Check for unique order numbers.
- Identify valid and null delivery dates.
- Calculate number of orders and date range.
- Add a TransactionKey field to uniquely identify each row.
- Name the table "Sales" and load it into the data model.

Task 2: Connect to **Products** CSV File and profile the data.

What does the company sell? How many distinct categories and subcategories do you see?

- Connect to the product CSV file.
- Enable column quality and column distribution views.
- Verify headers and data types.
- Identify distinct categories and subcategories.
- Confirm valid values and data quality.
- Name the table "Products" and load it into the data model.

Task 3: Connect to the **Stores** CSV file and profile the data.

How many stores does Global Electronics Retailer operate? Do you notice any that aren't like the others?

- Connect to the store CSV file.
- Enable column quality and column distribution views.
- Verify headers and data types.
- Identify number of stores.
- Understand store differences (e.g., online store key zero).
- Name the table "Stores" and load it into the data model.

Task 4: Connect to the **Exchange_Rates** CSV file and profile the data.

What information does this table contain, and how could it be used?

- Connect to the exchange rates CSV file.
- Enable column quality and column distribution views.
- Verify headers and data types.
- Understand the purpose of exchange rates for currency conversion.
- Name the table "Exchange Rates" and load it into the data model.

Task 5: Connect to the **Customers** CSV file and profile the data.

How many customers has the company served? Where are customers primarily located?

- Connect to the customer CSV file.
- Enable column quality and column distribution views.
- Verify headers and data types.
- Remove the **ZIP code** column.
- Transform the **City** field to capitalize each word.
- Calculate customer age from the birthday field.
- Add a new column to calculate **Customer Age** in years, then add a conditional column to calculate **Age Range**, where >60 = "Senior", >30 = "Adult" and >18 = "Young Adult".
- Name the table "Customers" and load it into the data model.

Task 6: Create a **Calendar** table that contains a list of contiguous dates (no gaps) and reflects the same date range as the Sales table, then add new columns for Day Name, Start of Week/Month/Quarter, and Year.

- Duplicate the sales query to create a calendar table.
- Retain only the order date column, rename it to "Date".
- Remove duplicates to ensure a continuous list of dates.
- Add columns for day name, start of week, start of month, start of quarter, and year.
- Name the table "Calendar" and load it into the data model.

Build a relational model

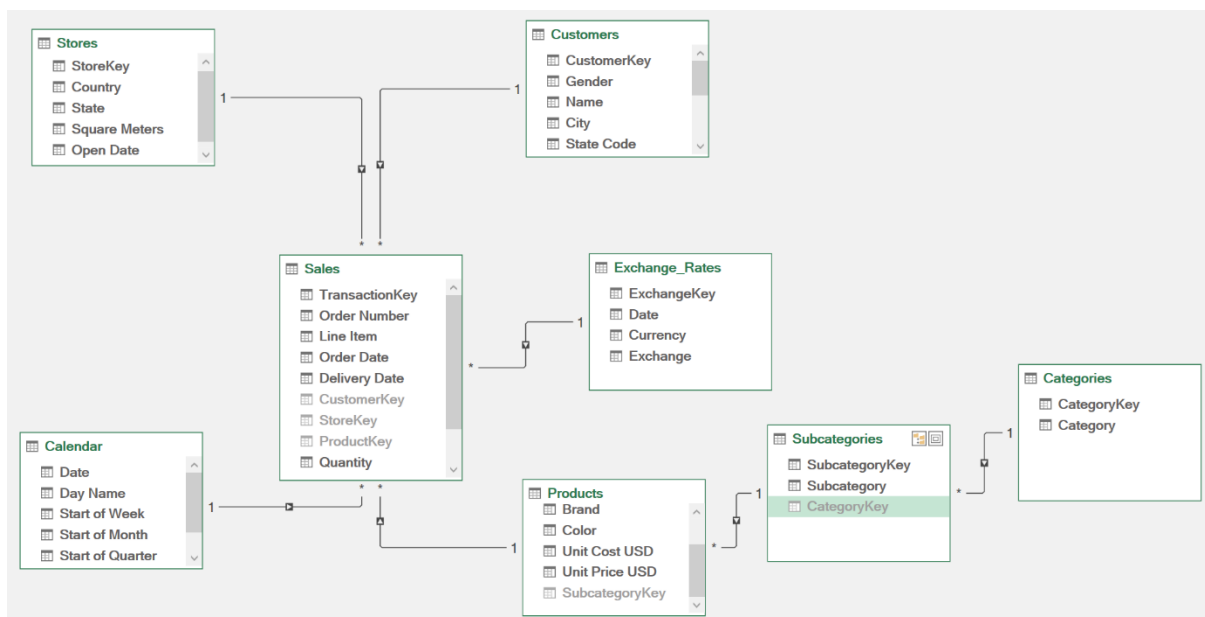
The second objective is to build a data model by connecting the foreign keys in the Sales table to the primary keys in each dimension table through one-to-many relationships.

- Create relationships from the **Sales** table to **Customers, Stores, Products** and **Calendar**

You should see 1:many relationships between the primary keys in each dimension table and the foreign keys in the Sales table. I make sure to connect the Calendar table to Order Date (not Delivery Date).

- What happens when you try to connect **Sales** to **Exchange Rates**? Add a new column to both tables named **ExchangeKey** and use that field to relate them via a 1:many relationship. You will need to create the ExchangeKey field based on both the date and currency code
- Hide all foreign keys from the **Sales** table

- Split the **Products** table into category and subcategory-level dimension tables, and update the data model to relate these new tables
- Duplicate the Products query, keep only the SubcategoryKey, Subcategory and CategoryKey fields, and remove duplicates from SubcategoryKey to create a new Subcategories table (then follow a similar process to create a new Categories table).
- In the data model diagram view, right-click and hide any fields on the "many" side of table relationships (i.e. CustomerKey, StoreKey, ProductKey and ExchangeKey in the Sales table)



Data Cardinality for Global Electronics Retailer

Enrich & explore the data

The third objective is to enrich the data model by adding calculated measures to track key business metrics, including total orders, revenue, average order value, and delivery time.

- Add a blank, dedicated measures table
- Create a new measure to calculate **Total Orders**, based on Order Number. How has order volume trended over time? Do all product categories show a similar trend?

(use a distinct count to avoid counting duplicate order numbers)

`=DISTINCTCOUNT(Sales[Order Number])`

- Calculate **Total Revenue (USD)**, based on Quantity and Unit Price (USD). Which stores generate the most revenue? Which individual products? (BONUS: Create another version to calculate revenue in local currency).

You will need to multiply the sales quantity by the related unit price from the Products table, then sum the results (use an iterator function)

=SUMX(Sales,[Quantity]*RELATED(Products[Unit Price USD]))

- Calculate **Average Order Value (AOV)**, based on Total Revenue (USD) and Total Orders. How does AOV compare across product categories? Are there differences based on customer age?

AOV can be calculated by dividing total revenue by total orders.

=DIVIDE([Total Revenue(USD)],[Total Orders])

- Calculate **Average Delivery Time**, in days. Which types of products tend to be delivered fastest or slowest?

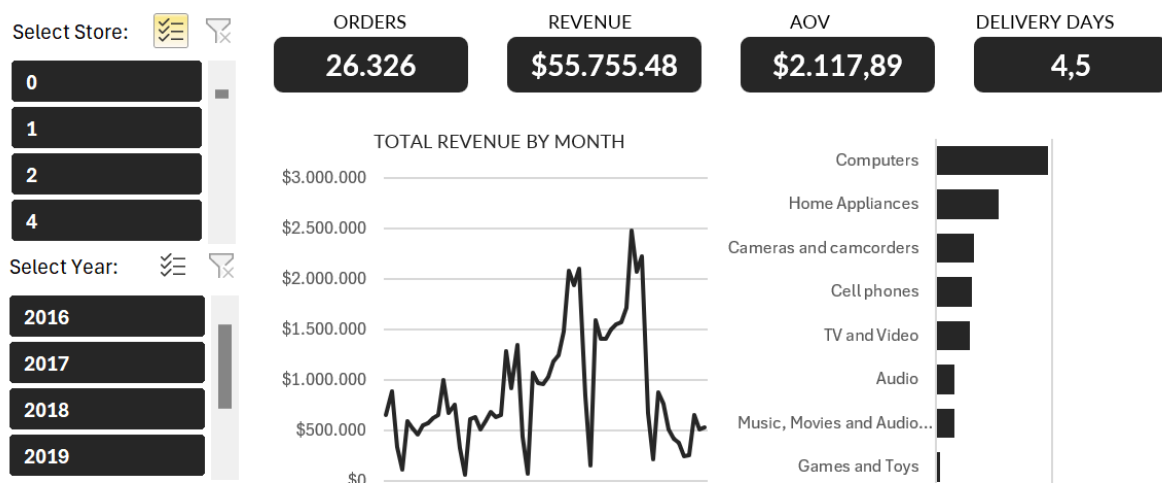
You will need to subtract the order date from the delivery date for each row and average the results.

=CALCULATE(AVERAGEX(Sales,[Delivery Date]-[Order Date]),Sales[StoreKey]=0)

Build an interactive report

The final objective is to design an interactive report that the Global Electronics Retailer leadership team can use to explore performance by store and over time.

- Show orders, revenue, AOV and average delivery time as **KPI cards**
- Show revenue by month as a **line chart**, and by Product Category as a **bar chart**. you may need to reverse the Y-axis to show the highest-volume categories at the top of the bar chart)
- Add **slicers** for StoreKey and Year, connected to all cards and visuals. use Slicer > Report Connections to connect slicers to multiple PivotTables in an Excel workbook
- Adjust formatting and polish for clarity and readability (consistent colors/themes, proper number formats, minimal clutter, etc.). Remove any unnecessary chart elements, optimize number formats for readability, and use clear and consistent colors and styles.
- What trends and patterns do you see? Where would you suggest digging deeper?
- Use the slicers to compare performance across stores, and over time. Pay attention to how delivery times have changed each year, when revenue tends to dip, and which product categories are the biggest drivers.



Global Electronics Retailer Dashboard

Analysis & Insight

After conducting an exploratory analysis of an electronics retail dataset using Excel and visualizing it with several key insights were uncovered:

Order and Sales Analysis:

- Distinct order numbers: 26,326.
- Date range of the data: January 1, 2016, to February 20, 2021.
- Only online store (store key zero) had valid delivery dates, indicating that delivery date analysis is only relevant for online sales.

Product Profiling:

- Eight distinct product categories: Home appliances, computers, cameras, camcorders, cell phones, TVs, etc.
- Most popular category: Home appliances (26%).
- Subcategories: 32 unique subcategories, with computer accessories being the most popular.

Store Analysis:

- Global Electronics Retailer operates 67 stores.
- Online store (store key zero) has no physical location, correlating with valid delivery dates.

Exchange Rates:

- Exchange rates table includes daily rates for various currencies (USD, CAD, AUD, EUR, GBP) from 2015 onwards.

Customer Data:

- Total unique customers: 15,266.
- Majority of customers are from the US (44%), followed by the UK, Canada, Germany, and Australia.

DAX Calculations:

- Key measures: Total orders, total revenue (USD and local currency), average order value (AOV), and average delivery time.
- Calculations revealed trends such as improving delivery times from over 7 days in 2016 to 3.8 days by early 2021.
- Online store (store zero) generated the most revenue (\$11.4 million), significantly higher than the next highest store (\$1.4 million).

Trends and Patterns:

- Steady upward growth in revenue until April 2020, followed by a significant decline likely due to external factors (e.g., COVID-19 pandemic).
- Seasonal trends observed, particularly a recurring dip every April.
- Home appliances, TVs, and computers have the highest average order values.
- Delivery times improved significantly over the years, showcasing better logistics and fulfillment strategies.

Task Requirement

1. What types of products does the company sell, and where are customers located?

Product Categories: The company sells eight distinct product categories: home appliances, computers, cameras, camcorders, cell phones, TVs, and more. Home appliances are the most popular category, making up 26% of the product range .

Customer Locations: Customers are primarily located in the US (44%), followed by the UK, Canada, Germany, and Australia .

2. Are there any seasonal patterns or trends for order volume or revenue?

Seasonal Trends: There is a clear seasonal trend with steady upward growth in revenue until April 2020, after which there is a significant decline that never fully recovers. A recurring dip in order volume and revenue is observed every April .

3. How long is the average delivery time in days? Has that changed over time?

Average Delivery Time: The average delivery time has improved significantly over the years. In 2016, the average was over seven days, which decreased to 3.8 days by early 2021 .

4. Is there a difference in average order value (AOV) for online vs. in-store sales?

AOV(Average Order Value) for Online vs. In-Store Sales): The AOV for the online store (store key zero) is significantly higher than for physical stores. For example, the online store generated \$11.4 million in revenue compared to \$1.4 million for the next highest store.

Summary

From the exploratory analysis and visualization using Microsoft Excel, we were able to generate insights and highlight all the key points Global Electronics Retailer needs to focus on. The results of the visualization correspond with all the queries written and generated using Excel. This action has enabled Global Electronics Retailer to make more data-driven decisions, allowing them to better understand their customer niche, identify strengths and weaknesses, and strategize inventory management and marketing efforts to capitalize on popular categories.