OBJECTIVE: The major aim of this project is to gain insight into the sales data of Amazon to understand the different factors that affect sales of the different branches.

About Data: This dataset contains sales transactions from three different branches of Amazon, respectively located in Mandalay, Yangon and Naypyitaw. The data contains 17 columns and 1000 rows

Methodology Used

Data Wrangling: This is the first step where inspection of data is done to make sure NULL values and missing values are detected and data replacement methods are used to replace missing or NULL values.

1. Build a database

2. Create a table and insert the data.

3. Select columns with null values in them. There are no null values in our database as in creating the tables, we set NOT NULL for each field, hence null values are filtered out.

Feature Engineering: This will help us generate some new columns from existing ones.

1. Add a new column named timeofday to give insight of sales in the Morning, Afternoon and Evening. This will help answer the question on which part of the day most sales are made.
2. Add a new column named dayname that contains the extracted days of the week on which the given transaction took place (Mon, Tue, Wed, Thur, Fri). This will help answer the question on which week of the day each branch is busiest.
3. Add a new column named monthname that contains the extracted months of the year on which the given transaction took place (Jan, Feb, Mar). Help determine which month of the year has the most sales and profit.

Exploratory Data Analysis (EDA): Exploratory data analysis is done to answer the listed questions and aims of this project.

Data Visualization: Utilized Excel to create visualizations for customer trends, product performance, and sales distribution. Incorporated SQL query results to generate actionable insights through dynamic charts and tables.

Business Questions To Answer:

1) What is the count of distinct cities in the dataset?
2) For each branch, what is the corresponding city?
3) What is the count of distinct product lines in the dataset?
4) Which payment method occurs most frequently?
5) Which product line has the highest sales?
6) How much revenue is generated each month?
7) In which month did the cost of goods sold reach its peak?
8) Which product line generated the highest revenue?
9) In which city was the highest revenue recorded?
10) Which product line incurred the highest Value Added Tax?

11) For each product line, add a column indicating "Good" if its sales are above average, otherwise "Bad."
12) Identify the branch that exceeded the average number of products sold
13) Which product line is most frequently associated with each gender?
14) Calculate the average rating for each product line.
15) Count the sales occurrences for each time of day on every weekday.
16) Identify the customer type contributing the highest revenue.
17) Determine the city with the highest VAT percentage.
18) Identify the customer type with the highest VAT payments.
19) What is the count of distinct customer types in the dataset?
20) What is the count of distinct payment methods in the dataset?
21) Which customer type occurs most frequently?
22) Identify the customer type with the highest purchase frequency.
23) Determine the predominant gender among customers.
24) Examine the distribution of genders within each branch.
25) Identify the time of day when customers provide the most ratings.
26) Determine the time of day with the highest customer ratings for each branch.
27) Identify the day of the week with the highest average ratings.
28) Determine the day of the week with the highest average ratings for each branch.

Product Analysis:

- **Food and Beverages and Fashion Accessories** have the highest average ratings, indicating strong customer satisfaction and potential market strengths.

- **Food and Beverages** have the highest total sales, amounting to **$56,144.96**. This indicates a strong market demand and customer preference for this category.

- Categories with relatively lower sales and VAT, such as **Health and Beauty**, might require targeted marketing strategies or product enhancements to boost their performance.

- **Branch A** has the highest total quantity of products sold **(1859)**, **exceeding the average number of products sold (1836.67)**. This indicates strong performance and effective sales strategies at Branch A.

- **Fashion Accessories** are more popular among **females (96)** compared to **males (82)**.

- Focus on enhancing the performance of lower-rated categories like **Health and beauty** by addressing customer feedback, improving product quality, and implementing effective marketing campaigns.

Customer Analysis:

- The predominant gender among customers is **Female**, with **501 counts** compared to **499 counts** for males.
- Customers tend to provide the most ratings in the **Evening (432 ratings).**
- **The lower number of ratings in the morning** indicates a potential area for improvement. Strategies such as morning discounts or engaging content could help increase morning activity.
- **Monday** has the highest average rating of **7.15**, followed by **Friday** (7.08) and **Sunday** (7.01). This indicates that customers tend to be more satisfied with their purchases at the start and end of the week.
- **Members** contribute the highest total revenue of **$164,223.81**, surpassing normal customers who contribute $158,743.62. This highlights the significant impact of loyalty programs and the importance of retaining members.
- **Evening** is the peak time for ratings and **Monday** has the highest average rating, businesses should consider focusing their promotional activities and customer engagement efforts during these times.

Sales Analysis:

- **January** generates the highest revenue **$116,292.11**. This indicates strong sales performance at the start of the year, possibly due to new year promotions.
- The popularity of **Ewallet (345 occurrences)** indicates a shift towards digital transactions.
- **January** has the highest cost of goods sold (COGS) at **$110,754.16**, suggesting that this month has the most inventory turnover or highest sales volume.
- **Evenings** are consistently high-performing across multiple days, including **Monday, Thursday, and Friday**. Focusing on enhancing evening sales strategies could further boost revenue.
- The **high revenue** in **January** suggests that this month are critical for business performance.