Amazon Sales Data Analysis

By Unmilan Baruah

Introduction

Sales management has gained importance to meet increasing competition and the need for improved methods of distribution to reduce cost and to increase profits.Sales management today is the most important function in a commercial and business enterprise. Amazon is an e-commerce platform that sells many product lines, which are preferred by consumers all around the world. This report is a detailed analysis of Amazon Sales Data from 2010 to 2017.

Purpose

- To find and study **region/country-wise** sales trend of Amazon Sales Data.
- To find the sales trend of the Amazon Data using **time-series analysis**.
- To evaluate the overall of **Revenue**, **Profit**, and **Cost** from 2010 to 2017.
- To analyze the sales trend based on **Order Priority**.
- To analyze the sales trend based on **Item Type**.
- To analyze the sales trend based on **Sales Channel**.

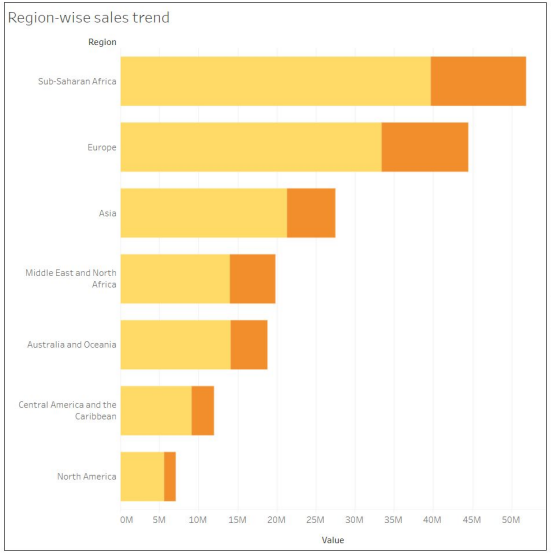
Methodology

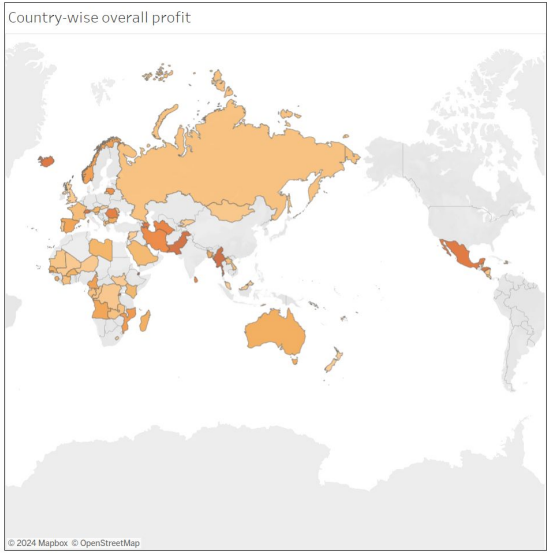
1. ETL (Extract, Transform, and Load): Collecting the data, transforming it into something which can be analyzed with accuracy, and loading the processed data for further analysis.
2. Data Analysis
3. Observations
4. Inference

ETL (Extract, Transform, and Load)

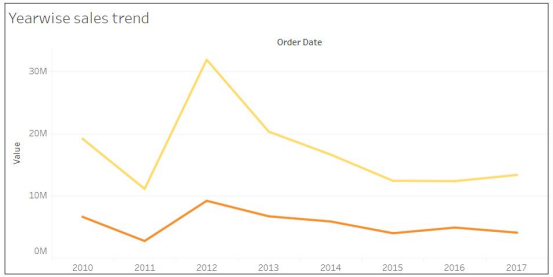
- The data source is named as “**Amazon Sales Data.csv**” and this CSV file has sales records from 2010 to 2017.
- There are total **100** entries, but no null values. Therefore, data cleaning was not required.
- For the time series analysis, in Google Colab, the date format of “**Order Date**” column, which was in **YYYY-MM-DD**, was transformed into three different columns: **order_day**, **order_month**, and **order_year**. However, it was no such transformation was needed in Tableau.

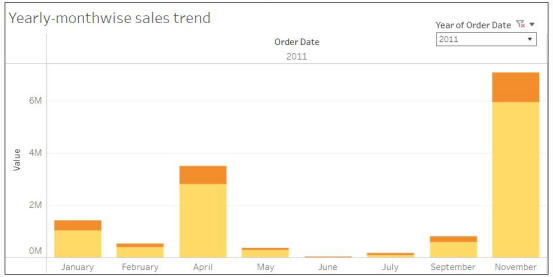
Data Analysis

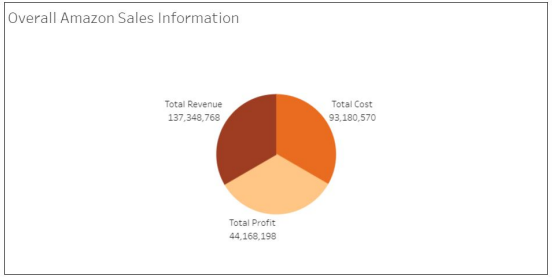




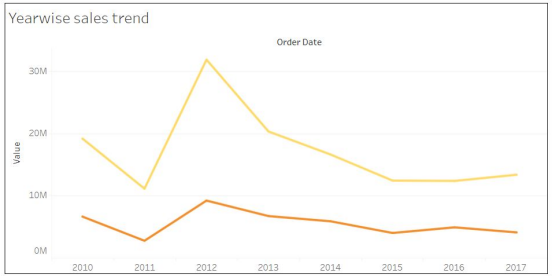
Data Analysis

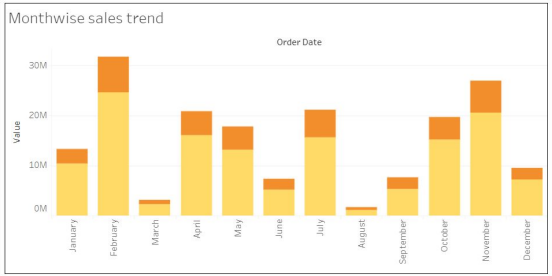


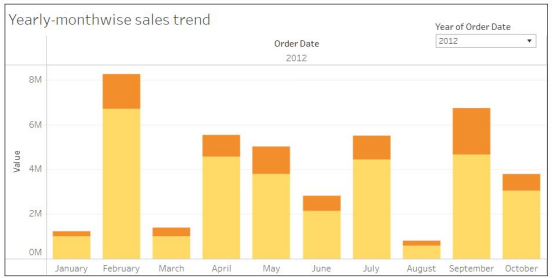


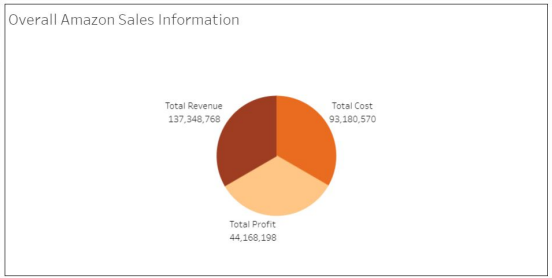


Data Analysis

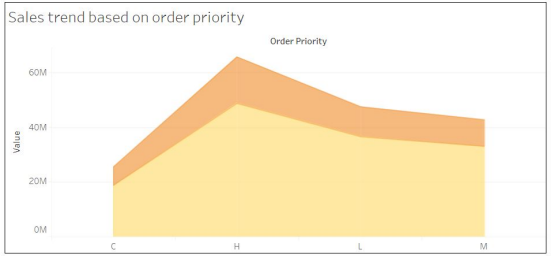


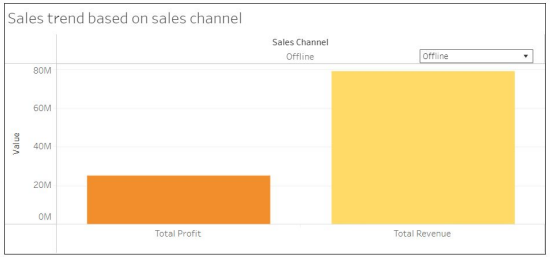


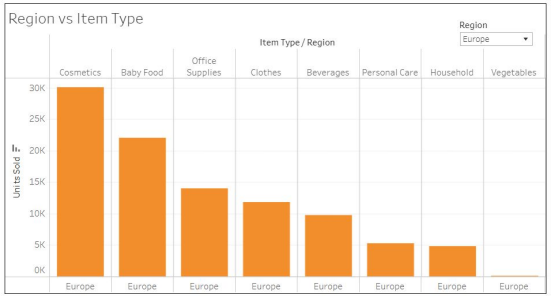


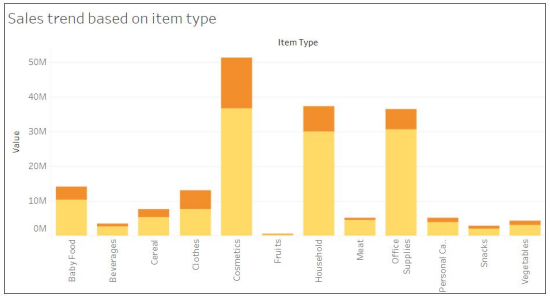


Data Analysis

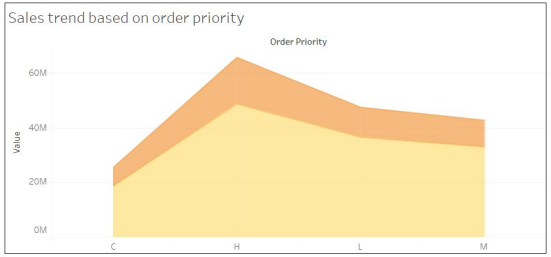


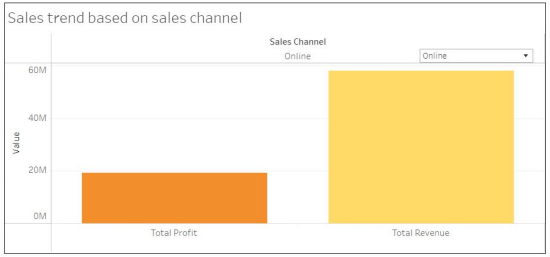


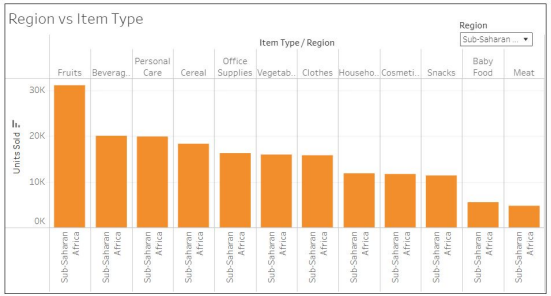


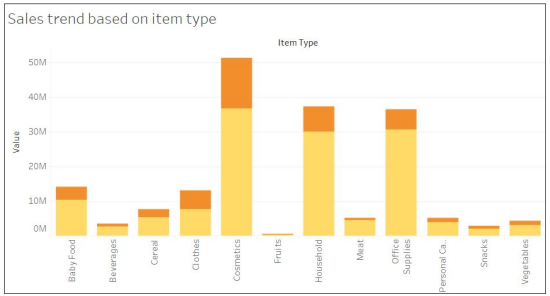


Data Analysis

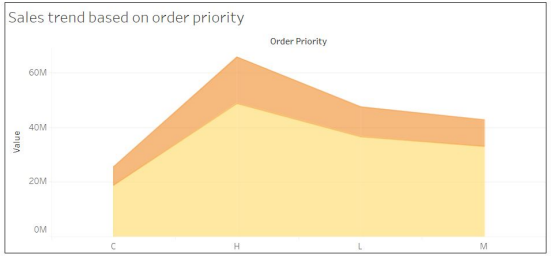


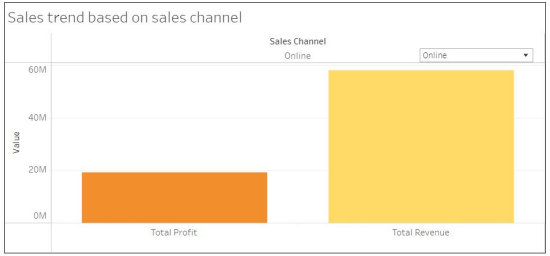


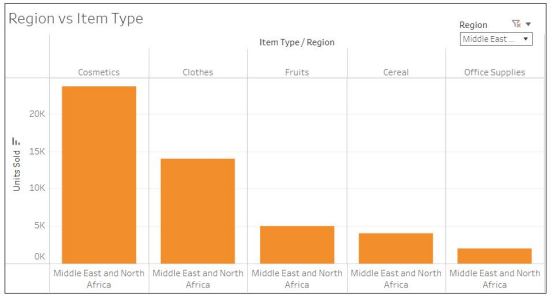


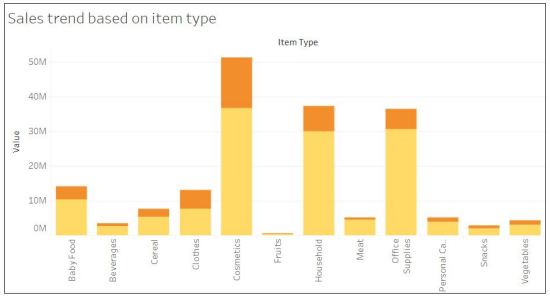


Data Analysis









Observations

- Overall **Revenue** made from 2010-17 is **$134 million**. From this revenue, overall **Cost**, and **Profit** made are **$93 million**, and **$44 million**, respectively.
- Sales trend is found out to be the highest in **Sub-Saharan Africa** region.
- The overall sales (**Profit**, and **Revenue**) is highest in the year 2012, and the lowest in the year 2011.
- People needs their products fact. This is because the order priority “**H**” is bringing the highest sales.
- Higher sales are observed for “**Offline**” sales channel, meaning people still prefer buying their products offline.

Inference

- **Cosmetic** products are sold the most in **Europe**, and **Middle East and North Africa** regions. So, it is better to promote **Cosmetic** products in those two regions.
- **Fruits** are are sold the most in **Sub-Saharan Africa** region. There, it will be better to promote **Fruits** products in the region, followed by **Beverages** as these products are the two most preferred item types in the region.
- However, necessary changes are needed to made in prices for **Fruits** products, because the sales from **Fruits** products is the lowest overall. The same is the case for **Beverages**, as it has achieved the third most lowest sales overall.

Inference

- Overall sales are highest for **Cosmetics**, followed by **Household**, and **Office Supplies**. Moreover, **Household** products are preferred the most in **Central America and the Carribean** as well as second most preferred products in **North America**, and **Asia**. So, it is also necessary to promote **Household** products with **Cosmetics** as well.
- Also, **Office Supplies** are the third most preferred products in **Asia**, and **Europe**. Thus, necessary to promote **Office Supplies**, just like **Cosmetics** and **Household** products.
- The least profit by region is observed in **North America**. Moreover, only **Personal Care**, and **Household** items are sold in this region. So, more study is necessary in the region to enhance the **Profit** in the region.

Thank You