Data Visualisation and Insight: Assignment – I

E-Commerce – Sales & Returns Analysis

By

Nishant Verma (d00251774@student.dkit.ie),

Khubim Kumar Chhetri (d00251757@student.dkit.ie),

Murthy Routhula (d00243413@student.dkit.ie).

Computing Science and Mathematics.

Dundalk Institute of Technology.

Dundalk, Ireland.

1. Project Summary:

- ### Introduction:

Electronic Commerce which is popularly known as E-Commerce refers to model in business that avails companies and public to sell and buy goods, services using Internet. It has spread over almost all the countries and their markets around the globe. Some of the companies operating their sales only through E-Commerce websites. This has become a vast business in the sales sector so the companies has to maintain their records of their sales and revenue to maintain the production of goods and its improvisation based on sales respect to markets and regions. So for analysing all the factors in their business, different Data Visualization techniques are helpful through which they can interpret all the information from them. They are best ways of Data representation.

Many tools exist in the present technology world to create visualizations. Among all of them tools like Microsoft Excel, PowerBI, Tableau are famous tools which are used to make visualizations to get insights about data. Using these we can make Infographics, Graphs, Charts, etc. useful to know company customer relationships, enriching customer experience, improve in filling their needs.

The main aim of this project is to create visualizations using Excel, PowerBI and Tableau to create dashboards and reports based on the world-wide data of famous E-Commerce website. The Analysis is made according to the Orders, Markets, Regions, Countries, Shipping procedures, Segments of market, Categories (Technology, Furniture, Office Supplies), Sales, Profit, Order priorities, Heads of the Regions and Goods return based on all the above factors.

- ### Findings in the Project:
  - ✓ Sales and Profit of the company according to Regions, Market, Countries, Categories, and Segment.
  - ✓ Head of the Regions handling Regions, Segments, Shipments, Priority Orders, and Sales below them.
  - ✓ Return of Orders based on Category, Market, Segment, and Order Priority.
  - ✓ Sales, Profits, Order Returns with respect to time, Segment, and Category.

2. Data Description and Wrangling:

The Raw data consists of 3 sheets Orders, Heads, Returns. Orders Sheet consist of Order ID, Order Date, Ship Date, Ship Mode, Customer Name, Segment, Location, Market, Region Product ID, Category, Sub-Category, Product Name, Sales, Quantity, Discount, Profit, Shipping Cost, Order Priority. Return Sheet Consist of Returned Order IDs, Market. Heads sheet consist of Person, Region.

- ### Programming language: Python
- ### Libraries: Pandas, Numpy, OS, Sys
- ### Data Cleaning Activities:
  1. Column names are changed to lower cases without spaces.
  2. Location column was split into Country and City to make analysis based on them.
  3. The Null Values Market, Region, Category are checked and mapped with the correct values based on other rows.
  4. The Orders sheet was joined with Heads and Returns sheets to evaluate number of returns under specific Heads, Regions, Segments and Markets.
- ### Data Description:
  The final dataset that was cleaned contains the following columns and comes under the labelled categories.

Column Name	Category	Information
order_id	Nominal Categorical	Order IDs of Orders
order_date	Date	Date of Order
ship_date	Date	Date of Order shipped
ship_mode	Nominal Categorical	Mode of Shipment (First class/ Same day/ second class, etc.)
customer_name	Nominal Categorical	Name of Customer

segment	Nominal Categorical	Segment of Order (Consumer/ Corporate/ Home Office)
market	Nominal Categorical	Market of country belongs
region	Nominal Categorical	Region of Country belongs
product_id	Nominal Categorical	ID of Product
category	Nominal Categorical	Category of product (Technology/ Furniture/ Office supplies)
sub_category	Nominal Categorical	Sub category in categories (Accessories/ phone/ chairs, etc.)
product_name	Nominal Categorical	Different product names
sales	Continuous Numerical	Amount of sales in $
quantity	Discrete Numerical	Number of items ordered in each order
discount	Continuous Numerical	Discount on order
profit	Continuous Numerical	Profit on the order
shipping_cost	Continuous Numerical	Shipping cost of order
order_priority	Nominal Categorical	Priority of order (critical/ medium/ high/ low)
city	Nominal Categorical	Name of ordered city
country	Nominal Categorical	Country of City
returned	Nominal Categorical	Whether order returned or no (yes/no)
person	Nominal Categorical	Head of Region

3. Data Visualization Techniques:

- *Microsoft Excel* – Microsoft excel is a spreadsheet developed by the Microsoft for various operating system all around. It features calculation or computation capabilities, graphing tools, pivot tools, and macro programming language called Visual basic for applications (VBA). Excel forms part of Microsoft office suite of software. While excel isn't a typical visualization software, it's versatile, powerful tool for professionals of all levels who want to analyse and illustrates datasets. There are many options available in the Excel to generate visualisation and in fact the overall dashboard.

- *Power BI* – Power BI is an interactive data visualization software product developed by Microsoft with primary focus on business intelligence. It is part of the Microsoft power platform. Power BI is a collection of software services, apps, and connectors that work together to turn unrelated sources of data into coherent, visually immersive, and interactive insights. Data may be input by reading directly from a database, webpage, or structured files such as spreadsheets, CSV, XML and JSON.

- *Tableau* - Tableau is a Data Visualisation tool built majorly on C++ language. It can be connected to many sources like CSV files, Excel files, Databases. Quick visualisations can be made using Tableau drop downs. Univariate, Bivariate and Multivariate plots can be interpreted using this tool through different Graph, Chart, and Dashboard Visualisations. We can make two different dashboards or reports interactive using tableau. Exploratory Analysis using latitudes and longitudes are possible using this tool. We can host it either on-premise or cloud. Tableau-Prep-Builder which is part of the Tableau is used to process data and data cleaning. We have done Sales & Profit Analysis, Heads Analysis, and Return Orders Analysis using Tableau. Dashboards done are interactive with each other and the reports on dashboards consists Filters, Slicers and KPIs are added.

4. Reviews on Team Members work:

- **Nishant (Microsoft Excel):**
  Strengths-
  (1) KPIs are well presented and interactive.
  (2) Multiple interactive dashboard

(3) Great use of space

Weakness-

(1) Some plots lack clarity and are not clearly readable.

- **Khubim (PowerBI):**
  Strengths-
  (1) Great visualisation with combo plots.
  (2) Great use of pie chart as they can successfully deliver insights
  (3) Choice of colour palette is great.

  Weakness-

  (1) Space is not properly used

- **Murthy (Tableau):**
  Strengths-
  (1) Aesthetically appealing
  (2) Great insights are captured
  (3) Great use of world map visualisation

  Weakness-

  (1) Colours selection can be improved.

5. **Git-Hub Repository for Codes and Dashboards:** https://github.com/khubim/M.Sc_Group_Project.git

6. **Conclusion:** Our team was successfully able to derive the insights from the raw data through our interactive dashboards. Our team has designed full-fledged working interactive dashboards using multiple technologies – Tableau, power BI and Excel. All three selected technologies has their pros and cons in place yet our team has tried to put their best foot forward by exploiting the rich set of tool of each technology. As per Industrial standard Tableau is preferred to use for Data Visualization and simple drop downs make Tableau easy to use. As it was built on C++, it handles large datasets and works faster.