TASK 1 – WEEK 1

Datasets Provided

Data Schema

Three datasets are provided for this case study:

- Customer: Customer information including demographics
- Transaction: Transaction of customers
- Product Hierarchy: Product information

About this Dataset

Analytics in Retail

With the retail market getting more and more competitive by the day, there has never been anything more important than the ability for optimizing service business processes when trying to satisfy the expectations of customers. Channelizing and managing data with the aim of working in favour of the customer as well as generating profits is very significant for survival.

Ideally, a retailer's customer data reflects the company's success in reaching and nurturing its customers. Retailers built reports summarizing customer behaviour using metrics such as conversion rate, average order value, recency of purchase and total amount spent in recent transactions. These measurements provided general insight into the behavioural tendencies of the customers.

Customer intelligence is the practice of determining and delivering data-driven insights into past and predicted future customer behaviour be effective, customer intelligence must combine raw transactional and behavioural data to generate derived measures. In a nutshell, for big retail players all over the world, data analytics is applied more these days at all stages of the retail process – taking track of popular products that are emerging, doing forecasts of sales and future demand via predictive simulation, optimizing placements of products and offers through heat-mapping of customers and many others.

About the Data

A Retail store is required to analyse the day-to-day transactions and keep a track of its customers spread across various locations along with their purchases/returns across various categories.

Task to perform

Create a report and display the calculated metrics, reports and inferences.

CASE STUDY

1. Merge the datasets
2. Check the duplicates if there are any duplicates drop them.
3. Check for the null/missing values.
4. Prepare summary report for all the columns.
5. Find the percentage of products bought by men and women (clothes).
6. Find the percentage of products bought by men and women (footwear).
7. How many types of stores are there, and which type of stores sells maximum of products?
8. Which type of book is selling the most?
9. What is the total amount spent by men on electronics?
10. What is the total amount spent by women on clothing?
11. What is the total amount spent on furnishings?
12. Who reads books more (Men/Women)?
13. Which type of books men read more?
14. Which type of books women read more?
15. Which electronic good is selling the most?
16. What is the total amount spent by customers between 1st January 2012 and 31st December 2012?
17. Calculate the following information using the merged data. Time period of the available transaction data.
18. Count of transactions where the total amount of transaction was negative.

VISUALISATION

1. Do proper visualisations and write down what you found out (data insights) from the plots.

2. Generate histograms for all continuous variables and frequency bars for categorical variables.