R CASE STUDY 1:

CUSTOMER ANALYSIS FOR RETAIL

Ground Floor, SCO 382
(Next to Iffco Chowk Metro station)
Sector 29,
Gurgaon, Haryana

First Floor, Bldg-41,
14th Main Road
Sector 7, HSR Layout,
Bengaluru, Karnataka

ANALYTIXLABS
www.analytixlabs.co.in

Ground Floor, SCO 382, Sector 29, Gurgaon
info@analytixlabs.co.in
(+91)95552-19007

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 favor 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 build reports summarizing customer behavior using metrics such as conversion rate, average order value, recency of purchase and total amount spent in recent transactions. These measurements provided general insights into the behavioral tendencies of customers.

Customer intelligence is the practice of determining and delivering data-driven insights into past and predicted future customer behavior. To be effective, customer intelligence must combine raw transactional and behavioral 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.

DATA AVAILABLE - Retail.xlsx

This book has the following sheets:

- Customer: Customers information including demographics
- Transaction: Transactions of customers
- Product Hierarchy: Product information (category, sub category etc...)

BUSINESS PROBLEM:

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

Create an **RMarkdown report** and display the below calculated metrics, reports and inferences.

(NOTE: THE REPORT MUST CONTAIN THE CODE AND THE OUTPUT AND THE Rmd FILE SHOULD BE SENT ALONG WITH THE PDF or HTML OUTPUT)

1. Merge the datasets Customers, Product Hierarchy and Transactions as Customer_Final. Ensure to keep all customers who have done transactions with us and select the join type accordingly.
    a. Use the base merge()
    b. Dplyr merge functions

2. Prepare a summary report for the merged data set.
    a. Get the column names and their corresponding data types
    b. Top/Bottom 10 observations
    c. "Five-number summary" for continuous variables (min, Q1, median, Q3 and max)
    d. Frequency tables for all the categorical variables

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

4. Calculate the following information using the merged dataset :
    a. Time period of the available transaction data
    b. Count of transactions where the total amount of transaction was negative

5. Analyze which product categories are more popular among females vs male customers.

6. Which City code has the maximum customers and what was the percentage of customers from that city?

7. Which store type sells the maximum products by value and by quantity?

8. What was the total amount earned from the "Electronics" and "Clothing" categories from Flagship Stores?

9. What was the total amount earned from "Male" customers under the "Electronics" category?

10. How many customers have more than 10 unique transactions, after removing all transactions which have any negative amounts?

11. For all customers aged between 25 - 35, find out:

   a. What was the total amount spent for "Electronics" and "Books" product categories?
   b. What was the total amount spent by these customers between 1st Jan, 2014 to 1st Mar, 2014?