# **Case Study – Medical Retail**

## **Learning Outcomes**

## Successful completion of this case study will enable the following learning outcomes:

## Conversion of complex business problems to analytical problems.

## Analytical investigation of financial transactions

## Univariate, Bivariate and Multivariate analysis of data

## Data storytelling

## **About the Company:**

Medplus is a large medical retail chain that sells pharmaceutical prescriptions and “Over Counter” medicines in India. The company was founded in 2006 and has witnessed tremendous growth. Although it is known for online medicine ordering aggregation, its operation base includes but is not limited to manufacturing and contract manufacturing of private label pharmaceutical, wellness and FMCG products, wholesale and retail sale, import, distribution, and pathology diagnostic laboratory testing.

Medplus has a full-fledged application from where customers can place their medicine orders and upload their prescription as supporting document based on which the medicines re delivered to customer’s address. In addition, Medplus runs an extremely successful chain of brick-and-mortar outlets across the country which are based on franchise model. The owners of ships/ existing pharmacies, opt for the franchise so that they can use the brand name of Medplus along with the available process integration that includes but is not restricted to billing systems, sourcing pipeline etc. Overall, a customer gets multiple choices of ordering medicines from Medplus.

## **Problem Background:**

You have been given data of medicine sales of the year 2020 and 2021 for the country. The data includes the sale of medicines through app and website which are segmented as “Retail” channels. The company now wants to expand into “Institutional” segment that could lead to large scale procurement of medicines from them. The institutional segment could also utilize online as well as offline sales channels.

As per information received, multiple number of bulk ordering agents are already utilizing the services of Medplus. The business managers at the company have taken a ballpark approximate figure of Rs 2,000/ transaction to qualify for being a bulk order. Hence, the given data of 65,969 rows and 10 columns includes only the customer details that has each transaction value of more then 2,000.

Following is the figure that illustrates the channel distribution for customers to engage with the brand:

Graphical user interface

Description automatically generated

The highlighted part refers to the new division of “Institutional” which company plans to open soon based on a comprehensive study. Your case study submission forms the preliminary basis of study for this purpose.

## **Data Dictionary:**

| **Column** | **Definition** |
| --- | --- |
| **Customer id** | Unique Key to identify a member. Each row identifies a customer. |
| **App User** | Flag to identify if the Member uses the Medplus app or website to order medicines OR not.   * 1 – If the Medplus’s own mode of ordering is used (That is, website or app) * 0 – If the order has been placed using alternate channels of sale (That is, physical visit to the store/ phone call to the store/ orders from alternate aggregators such as Amazon from the store) |
| **Covid 19 Medicine Sales** | This column shows the sales done of any and every medicine that is used to treat Covid 19 or any of its symptoms. It includes but is not limited to prescription medicines like doxycycline, Remdesivir, Favipiravir etc and includes but is not limited to OTC medicines and immunity boosters such as Limcy, Dolo 650, Zincovit etc. |
| **Other Medicine Sales** | This column includes medicines not related to treatment of Covid or any of its symptoms. |
| **Redeemed points Flag** | Flag to identify if the Member has redeemed points |
| **Transactions** | Number of Purchases made by the Member |
| **Last Transaction date** | Last Purchase date made by the Member |
| **DOB** | Member date of birth. DOB= 00-01-1900 is the default system birth rate. |
| **EMAIL Flag** | Flag to show if we have the Member email id |
| **Gender** | Member Gender |

## **Problem Statement**

You are recruited as a data analyst with Medplus based in Managing Director’s office in Hyderabad. Your primary role is to assist Managing Director and provide him with strategic data inferences that help him navigate the company effectively. Your main task in this case study is to perform:

Univariate analysis of Data:

* Prepare histogram and boxplots of continuous data to show the extent of data spread, identify outliers and data skewness
* Prepare countplot and pie plot of categorical data to check data balance of each of the columns.

Bivariate Analysis of Data:

* Derive the Total Sales column (covid19 medicine sales + other medicines sales) and perform bivariate analysis
  + For categorical columns: Show class wise distribution of total sales and average sales
  + For continuous columns, show the scatterplot that highlights relationship between both the variables

Multivariate Analysis of Data:

* Perform multivariate analysis with a heatmap. This should show the intervariable dependencies and correlations.

Notes:

* Target column needs to be derived for above activity as indicated.
* It is preferable to draw an additional column of average order value as well