**Pharmaceutical Sales Analysis**

In this ‘Data Analysis’ project, we’ll analyze a global Pharmaceutical Manufacturing Company's raw sales data and draw meaningful insights.

**Table of Contents**

* Introduction
* Objective
* Dataset
* Solution Approach
* How To Use
* Insight And Recommendations
* Credits

**Introduction**

Platini Data Pharmaceutical is one of the leading Pharmaceutical Manufacturing companies with a global presence.

Their Markets are divided into different regions across the world. One of those regions manages the German and Poland Markets.

Company does not sell directly to customers. Instead, they work with a couple of Distributors in all their regions.

They have an agreement with each distributor to share their Sales Data. This is to enable them to gain insights up to the retail level. This data is made available to them in CSV format.

**Objectives**

The firm has asked us to perform in-depth data analysis to get insight into the company sales performance. Specifically, below are the essential requirements to be satisfied…

|Requirement ID| For Whom Requirement| Description|

PD01-REQ-1|Executive Committee| A high-level overview showing company’s overall sales performance by year by month, by customer cities, by channel, by sub-channel. Should be able to quickly see top drug class by sales, top drug by sales, top customer city by sales.

PD01-REQ-2|Sales Manager/Sales Rep| A detailed overview showing sales by distributors and product, top 5 product, customer and cities, sales numbers split by channels and sub-channels.

PD01-REQ-3|Head of Sales| A detailed report of sales by sales-team split by product and sales by sales-team split by product class A detailed analysis showing Top sales managers, Top sales reps, Top product split by sales team contributions answering. An ability to filter/slice data by year and months.

**Dataset**

The dataset is sourced from each distributor. It contains Pharmaceutical Manufacturing Company’s, Wholesale-Retail Data. The field description of the raw data is given below.

The raw dataset `pharma-data.csv` can be downloaded from https://drive.google.com/file/d/1sCdm4KKOB7IAPru1LvkihQ13SMcVZeV4/view?usp=sharing

Field Description

Distributor Name of Wholesaler

Customer Name Name of customer

City Customer's city

Country Customer's country

Latitude Customer's Geo Latitude

Longitude Customer's Geo Longitude

Channel Class of buyer (Hospital, Pharmacy)

Sub-channel Sector of the buyer (Government, Private, etc.)

Product Name Name of Drug

Product Class Class of Drug (Antibiotics, etc.)

Quantity Quantity purchased

Price Price product was sold for

Sales Amount made from sale

Month Month sale was made

Year Year sale was made

Name of Sales Rep Name of the Sales rep who facilitated the sale

Manager Sales rep's Manager Name

Sales Team Sale rep's team

**Solution Approach**

|Requirement ID|Solution ID| Proposed Solution|

|PD01-REQ-1|PD01-SOL-1|An Executive Summary PowerBI dashboard/report page will be built to show a high-level overview of sales data in interactive visuals per the requirements. A year filter will be provided to filter the data by a particular or combination of years

|PD01-REQ-2|PD01-SOL-2| A Distributor & Customer Analysis PowerBI dashboard/report page will be provided with interactive visuals showing data as per the requirement

|PD01-REQ-3|PD01-SOL-3| A Sales Team Performance PowerBI dashboard/report page will be provided with interactive visuals showing data as per the requirement. year and month slicers will be provided to slice/filter data by year and/or months

**Exploratory Data Analysis (EDA)**

In the initial phase of this project, we utilize the `pandas` python package to conduct Exploratory Data Analysis (EDA). This involved familiarizing ourselves with the data, understanding its contents, and ensuring its integrity and gaining insights into its characteristics.

* Checking of any missing values
* Any unusual value (outliers)
* Incorrect values (e.g., sales column, we see -ve numbers)
* Determine `categorical` and `numeric` columns
* Determine dimensions of categorical columns and range of numeric columns

Note that these steps can be performed using Power Query Editor and/or excel; however, pandas make it much easier and faster; on top of that, pandas can handle massive datasets.

EDA steps can be found in the data-exploration. ipynb notebook.

**Data Cleaning and Transform [Power Query Editor]**

The provided dataset was relatively clean and well organized; hence only a little work was required in this step; the following steps were carried out...

* Correct column heading provided
* Correct data type is assigned to columns

**Data Model Creation [PowerBI Desktop]**

* The provided data is in a single table format. The exploration revealed that it contains both categorical (`dimensions`) and numeric (`facts`) data.
* We build a data model where dimensions and facts are separated, then they are linked together by logical relationship to form a `star schema. ` The resultant data model is shown below...

The tables with the prefix `DIM` are dimension tables, and `FACT` is the fact table.

**Report Creation [PowerBI Desktop]**

Three interactive reports/dashboards (report pages) will be created to implement the proposed solution. Refer to [Table-3: Proposed Solution] (#solution-approach) for detailed requirements and the corresponding proposed solution.

**1. Executive Summary Report [PD01-SOL-1]**

This high-level report shows the overall sales figures and elements at a glance.

**2. Distributor & Customer Analysis Report [PD01-SOL-2]**

This granular detailed report analyses data from the company distributors' and customers' perspectives. Sales by the distributor can be drilled down to specific product levels.

**3. Sales Team Performance Report [PD01-SOL-3]**

This is another detailed report that analyses the performance of the company's sales team. Sales by the sales team can be drilled down to product class and specific product levels.

**Credits**

* Dataset sourced from [Google Drive] https://drive.google.com/file/d/1sCdm4KKOB7IAPru1LvkihQ13SMcVZeV4/view?usp=sharing

**Insight And Recommendations**

Total Sales Gh11.80bn

Top product/Drug class is Analgesics with a total sale of Gh2,371,515,114.00 and product/Drug is Ionclotide with the total sales of Gh169,083,391.00.

Top City is Butzbach with a total sale of Gh93,561,780.00.

2018 and August was the year and month with the highest sales of Gh3,506,897,354.00 and Gh1,186,627,360.00 respectively.

Retail was the sub-channel/ sector of the buyer where most sales was gathered through with an amount of Gh3,343,096,900,00.

Within the class of the buyer, Pharmacy has the highest sales in all the years. From the analysis it can be said that in the class of buyer, Pharmacy generated more sales than Hospital.

The Distributor with the most sales of product from the Manufacturing Company is Gerlach LLC with a total sale of Gh3,501,834,439.00.

Customer with the higher sales of product from distributors is Mraz-Kutch Pharma PLC with a total sale of Gh93,561,780.00

From the analysis, Distributors distribute more products to customers in the country Germany than Poland, making the total sales in Germany higher.

Delta has the highest sale of product with a total sale of Gh3,635,341,332.

The highest Sales Manager is Britanny Bold from Delta Team and the highest sales Rep who facilitate the sales is Jimmy Grey from Charlie Team with a total sale of Gh3,635,341,332 and Gh985,969,994 respectively.

**Recommendations**

From the data insight above, I recommend;

* The Manufacturers and the Distributors should foster collaboration on how to leverage expertise and resources for country with lower sales and sectors with lower sales as well.
* Branding and marketing strategies should be promoted on pharmaceutical products and build trust among health care professionals since from the insight Hospital generate low sales of drugs.
* There should be more investment plan in Digital health technologies to enhance patient engagement, improve healthcare delivery and collect real world data for drug development and post market surveillance.
* There should be Global expansion to explore opportunities into emerging markets while also maintaining a strong presence in established markets.