**Case 1**

Suppose you are a data scientist working on a project for a Pharma company that is trying to understand Sales historical data of four years (2017 – 2020). This will help them to screen the products and Budget planning for next year.

To perform the task, you received a dataset containing 5017 observations representing drugs with 14 variables defining its characteristics.

**Data Set:**



Problem statement

Perform the below operations using python packages.

* Read the .xlsx file into a pandas dataframe - Done
* Find total records in columns and their datatype - Done
* Calculate statistical parameter for numerical column - Done
* Show 2020 monthly sales and profit -
* Which are the Top 20 drugs by sales of 2020
* Which are the Bottom 20 Dispensed Pharmacies by Sales of 2020
* Create visual to show comparison between 2020 Sales and previous year Sales.
* Show Region and Market segment wise sales percentage.
* Create Histogram to show number of orders by Quantity (Bins)
* Crate a visual to show State wise Sales