**Introduction to dataset:**

**This dataset collected using url:** <https://www.kaggle.com/milanzdravkovic/pharma-sales-data?select=salesdaily.csv>

The dataset is built from the initial dataset consisted of 600000 transactional data collected in 6 years (period 2014-2019), indicating date and time of sale, pharmaceutical drug brand name and sold quantity, exported from Point-of-Sale system in the individual pharmacy. Selected group of drugs from the dataset (57 drugs) is classified to the following Anatomical Therapeutic Chemical (ATC) Classification System categories:

M01AB - Anti-inflammatory and ant rheumatic products, non-steroids, Acetic acid derivatives and related substances

M01AE - Anti-inflammatory and ant rheumatic products, non-steroids, Prop ionic acid derivatives

N02BA - Other analgesics and antipyretics, Salicylic acid and derivatives

N02BE/B - Other analgesics and antipyretics, Pyrazolones and Anilides

N05B - Psycholeptics drugs, Anxiolytic drugs

N05C - Psycholeptics drugs, Hypnotics and sedatives drugs

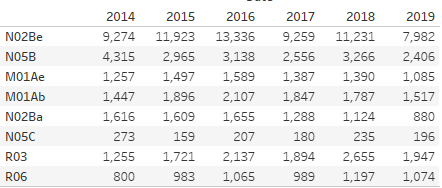
R03 - Drugs for obstructive airway diseases

R06 - Antihistamines for systemic use

**Time series analysis**

Time series analysis deals with the time series data. In time series analysis we analyzed trend based on monthly sales of the drugs. Seasonality can be used to help analyze sales and economic trends. Industries are use seasonality to help determine certain business decisions such as inventories and staffing. Seasonal or short-term cycles are explained clearly by plot diagrams Sales of N02BE sales is high in every year .Sales of N05C sales is less compared to total count of the all drugs. Here considered average value of 100 day sales or monthly mean value for plotting the graphs due to huge data. For time series analysis plotted the graphs based on monthly, weekly and yearly.

**Details of dataset show in below table**

Observe this table we can understand the sales of the drugs clearly.

Sales of the drug are high in Saturday and Sunday. This table explains the count of the sales based on year.

**ACF & PACF:**

ACF is Auto-correlation function and PACF is partial correlation function .This ACF and the PACF of the series are the following. (They start at lag 1). The PACF shows a single spike at the first lag and the ACF shows a tapering pattern. Analyzed the dataset using Autocorrelation and Partial Autocorrelation functions for this analysis considered lags=20. For lag one correlation we got positive correlation is 0.052. Correlation between the variables is less due large variation in the dales of the drugs and some variables have negative correlation.

**Autoregressive, Moving Average & differencing degrees:**

Autoregressive models predict future values based on past values. They are widely used in technical analysis to forecast future security prices. Autoregressive models implicitly assume that the future will resemble the past. Here built the auto regression model and predicted new values using test data finally we got the mean squire error is 5.58. So, predicted and dataset values has high variation.

A moving average is commonly used with time series data to smooth out short-term fluctuations and highlight longer-term trends or cycles. A moving average can smooth out the noise of random outliers and emphasize long-term trends. Analyzed the moving average here we considered average of 30 record. And plotted individual column plots and also we plotted all columns in a single plot. In moving average we analyzed long rolling with we considered Windom = 20 and 100. We calculated a 20-days span=20 EMA. Adjust=False specifies that we are interested in the recursive calculation mode. Differencing degrees is estimated using R03 drug column.

Differencing degrees is a method of transforming a non-stationary time series into a stationary one. Due to this method variation between variables is decreased then this data good fit to model building. Use of differencing degree method is Cleary shown in before not applied this method data plot and applying this method plot has variation.

**ARIMA MODEL:**

Autoregressive Integrated Moving Average Model. An ARIMA model is a class of statistical models for analyzing and forecasting time series data. It explicitly caters to a suite of standard structures in time series data, and as such provides a simple yet powerful method for making skillful time series forecasts. This ARIMA model is highly recommended model for future value prediction. To predict future observation we trained the data set using ARIMA model. We can use to train and then forecast future time points. ARIMA can capture complex relationships as it takes error terms and observations of lagged terms. These models rely on regressing a variable on past values. Using this model we predicted new values using old values and plot a graph using old values and newly predicted values. Predicted values variation is very less compared to real values this variation clearly shows in graph.

**Summary:**

These sales data is resample to hourly, daily, weekly, monthly and yearly periods. This data is mainly deals with the sales of the different drugs during the period 2014 January to 2019 November. The sales of drugs started year 2014 high sales down in the year 2016 and highly sold drug is N02BE is top sold drug in every year. Less sold drug is N05C it sold very less in every year. In this data set variation of sales is high sum of the drugs sales is high at starting and end of the year and some of the drugs sales is high in middle of the year. Correlation between variables or drug of the data set very less so they are none correlated to each even there is no high negative correlation between the variables of the dataset. For this dataset plotted the graphs based on sales by trends yearly based sales and seasonal sales of the data. Estimated the autoregressive by built AR model for this model got mean squared error is 5.5. Calculating the long-window (window=100) simple moving average and short-window (window=20) simple moving average. Finally ARIMA model fitted to dataset to predict future data using this data predicted future data and plotted a graph using old data and predicted data.