## What is Trend analysis

Trend analysis is based on the idea that what has happened in the past gives traders/operations an idea of what will happen in the future based on that various parameter is assessed and accordingly corrective measures can be taken.

## What will be available?

Participant will be given “anonymous” file with three different sheets. Description of each field/sheet is as given below –

**Sheet -**

Trainingdata - This sheet will contain the training data to do the trend analysis and forcasting .

ValidationData - This sheet will contain the validation data

Testdata – you need to provide the result of your future model here .

**Columns –**

ProdCat - This colums contains around distinct 259 product category

Date – Date when sales data is captured .

Sales –Sales on the particular day . [-ve data means loss ]

## High level Problem statement:

In Barclays International there are multiple system which holds different types of data (e.g. Stock price , trade data , payments data , Risk data , reference data , sales data , Settlement data etc.) . Currently there is no unified framework which could do the trend analysis for all different types of data . we would like to solve this issue with creating a common framework which can do the trend analysis and forecasting for all different type of data [ to simplify this problem we will only consider the common attributes here i.e. date and value for each different data type also providing you only the sales data for different product category to solve the problem ]

Participants have to do the trend and forecasting analysis for all individual product category from given data and also create the **Generic framework** which can generate the trend (i.e. calculate different parameter like R square , MAE and RMSE and also the statement around when the significant change is being observed minimum 4 weeks also forecast the future data for each product category .

**Example -**

First Participant needs to find the right trend for each individual prodcategory data and draw the graph (e.g. graph1)

And then run the program to find the significant change with certain confidence interval and generate the statement e.g. (Product category is 60% decrease over 4 week on 24th week in below example and plot the graph of that interval , also calculate the R square , MAE and RMSE values

Ideally, participants must be able to plot this on graph so that it can be visually confirmed.

Also particpant should be able to predict the future value for each product category I,e on Rawdataset sheet .

**General guidelines, instructions and hints:**

* It should cover all the Linear/Non -linear model used to depict the comments.
* The comments should be significant with certain confidence interval.
* Identify the trend which might be partly or nearly completely hidden by noise.

Following visual will provide better explanation (similar but not limited).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Product category** | **R\_sq** | **MAE** | **RMSE** | **Comments** |
| ProductCategory | 0.078 | 11.25 | 67.8 | 60% decrease over 4 week as observed on 24th week . |

## Points to be considered for Evaluation-

## Below points will be used for Evaluation

## 1.prodcat|Comments|R\_square\_TRAINING|MAE\_VALIDATION|RMSE\_VALIDATION

## Note - R\_SQUARE TRAINING matrix can be changed to as per the model evaluation matrix.

## 2. Test data submission (Sheet – rawdatasheet)

## 3. PPT Preparation below points

## a. Thought process

## b. Data Engineering - Missing VALUE/OUTLIER

## c. Model used and reason for chosing the below model

## e. Result Evaluation

## 4. Generic framework and deployment process with any DB .

## Glossary

**R Square** - In statistics, a value is often required to determine how closely a certain function fits a particular set of experimental data.. R2 values range from 0 to 1, with 1 representing a perfect fit between the data and the line drawn through them, and 0 representing no statistical correlation between the data and a line. The R2 value (often referred to as the goodness of fit) is computed as follows:

**Mean Absolute Error (MAE):**MAE measures the average magnitude of the errors in a set of predictions, without considering their direction. It’s the average over the test sample of the absolute differences between prediction and actual observation where all individual differences have equal weight.

**Root mean squared error (RMSE):** RMSE is a quadratic scoring rule that also measures the average magnitude of the error. It’s the square root of the average of squared differences between prediction and actual observation.