**Table of Contents.**

[1 Introduction 1](#_Toc414885432)

[2 Challenge. 1](#_Toc414885433)

[3 Your role. 1](#_Toc414885434)

[4 Details of the data 1](#_Toc414885435)

[5 Assignment 3](#_Toc414885436)

# Introduction

The Client is a Retail company. They want to forecast for the next four months the number of items that will be consumed by males. They wish to forecast the growth of their sales for Men's items into the future so that the company can get an idea of the future demand and use that for better production planning of its merchandize. In doing so, it can have a chance to both minimize the errors due to underproduction and having product shortages, or overproduction which will lead to excess stock that would lead to an additional cost. For this, they have engaged on a forecasting project based on time series analysis techniques.

# Challenge.

The customer has stated that the objective of the catalog prediction exercise is to build a forecast for the number of items that will be consumed by men (male customers) in the next four months. The data is historic data for ten years, from 1989 to 1998.

# Your role.

You are the lead data miner for the project. In this phase, you have to confirm that you understand the business requirements (business understanding) as well as the data (data understanding) for the project.

# Details of the data

The customer has shared the data with your team.

The data is in the form of a text file “catalog\_seasfac.txt”. It consists of a ten year time series. A time series is a field whose values represent equally spaced observations of a phenomenon over time. Examples of time series include quarterly beer sales, cigarette sales, quarterly interest rates, weekly unemployment rates, etc. Generally, time series analysis is built on an aggregated data.



The description of the various fields are given below.

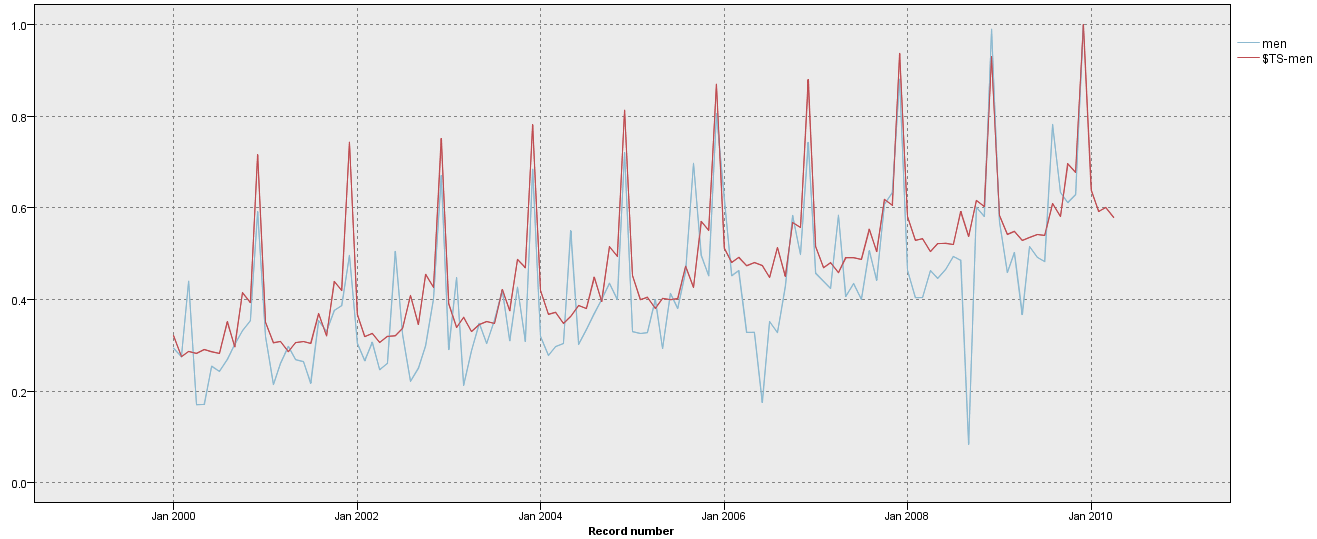
The customer has stated that the target variable is “Male”, i.e., “Sales of Men’s Clothing”.

**Data Description**

The data is described below

|  |  |  |
| --- | --- | --- |
| **SN** | **Name** | **Description** |
| 1 | date | Date |
| 2 | men | Sales of Men's Clothing |
| 3 | women | Sales of Women's Clothing |
| 4 | jewel | Sales of Jewelry |
| 5 | mail | Number of Catalogs Mailed |
| 6 | page | Number of Pages in Catalog |
| 7 | phone | Number of Phone Lines Open for Ordering |
| 8 | print | Amount Spent on Print Advertising |
| 9 | service | Number of Customer Service Representatives |
| 10 | YEAR\_ | YEAR, not periodic |
| 11 | MONTH\_ | MONTH, period 12 |
| 12 | DATE\_ | Date. Format: "MMM YYYY" |
| 13 | Seasonal\_Err\_Men | Error from Seasonal Decomposition |
| 14 | Seasonal\_AdjSer\_Men | Seasonal Adjusted Series for Sales of Men's Clothing |
| 15 | Seasonal\_Factors\_Men | Seasonal Factors for Sales of Men's Clothing |
| 16 | Seasonal\_TrendCycle\_Men | Trend-cycle for Sales of Men's Clothing |

Prior to being able to make the forecast, the company will need to collect information of its sales over time in order to gain a full picture of how sales have changed in the past. Once this information has been collected, it is possible to plot how sales change over time. See the blue colored plot for males in the figure below. Here, information for each month from January 1989 to December 1998 has been collected (note that the labels denote January 2000 to December 2009).



# Assignment

**Exercise 1.**

On the basis of the problem statement and the data provided, carry out the following steps in CRISP DM for the client –

1. Business Understanding
2. Data Understanding

The response for this exercise will be a report in the form of a word document, with separate sections for the deliverables below (refer to the lecture on CRISP-DM Methodology).

|  |  |
| --- | --- |
| **Stage** | **Deliverables** |
| **Business Understanding** | Data Mining Goals |
|  | Data Mining Success Criteria |
| **Data Understanding** | Data Exploration Report |
|  | Data Quality Report |

**Exercise 2.**

Perform initial analysis to identify possible reasons for seasaonlity, trends and cycles in your data and then understand what operations shall you apply on the data if it has seasonality or unseasonality. Apply all the necessary transformations such as date related transformations or taking aggregates of some input fields that can help in building your forecast. Provide a report of the same.