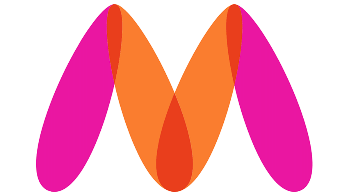
**Myntra Fashion Clothing** 

**Aim**

Myntra is a major Indian fashion e-commerce company headquartered in Bengaluru, Karnataka, India.] The company was founded in 2007 to sell personalized gift items. In May 2014, Myntra.com was acquired by Flipkart.

We will be using data science skills to identify the apparel type that customers favours and their prices. To identify the parameter that attracts customers to make purchase.

Is it number of images, or colours, or brand name or price?

**Problem Statement**

The Myntra have shared the dataset with you to identify the attributes to increase sales. You are working as Lead consultant and your key role is to identify the parameters that are extremely important while making a decision.

As a lead consultant you also have to show the results to your client and managers so it’s advised to create charts while you perform analysis and write down the insights in some separate sheet that you can refer later on.

Some of the problems can be easily identified while solving the scenarios and tasks shared here but you are also required to further share your key points in the Conclusion.

Exploratory Data Analysis (EDA) is an approach to analysing data sets to summarize the main characteristics of data by often using statistical graphs and other visualization methods such as by the use of statistical graphs.



**Learning Outcome**

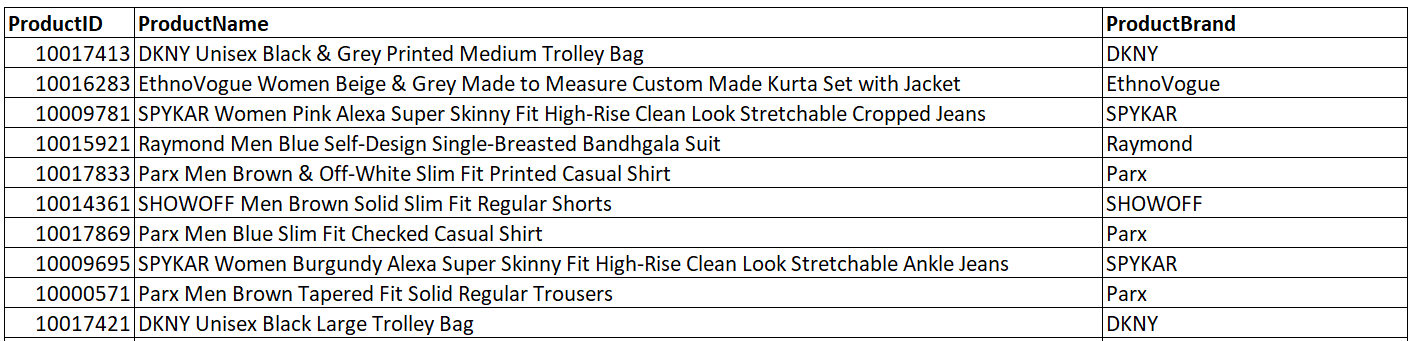
* Pandas Joins and Merge
* Data Manipulation
* Data cleaning
* Creating charts and bars
* Perform wrangling operations to draw more insights
* Evaluation of bi-columns on the basis of next attribute
* Creating new columns to drill down on analysis

**Data Information**

There are 2 csv files that are shared here.

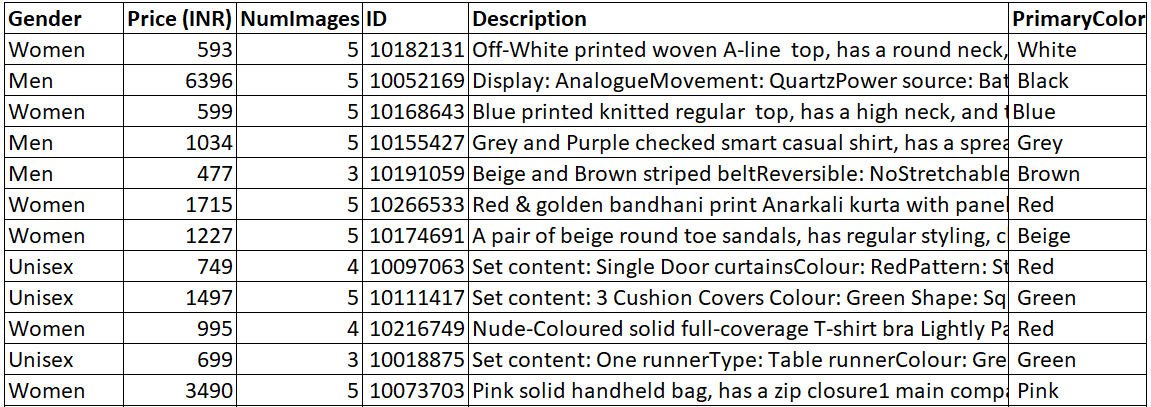
1. Product Details

* ProductID – ID assigned to the product
* ProductName – Name of the Product
* ProductBrand – Brand Name of the Product



1. Products Catalog

* Gender – gender to which specific products that have been designed
* Price (INR) – Price of the products
* NumImages – Number of images that have been clicked for specific product
* ID - ID assigned to the product
* Description – full details of the product
* PrimaryColor – Color of the product



**Skill Requirement**

* numpy
* pandas
* matplotlib
* seaborn

# **Scenario1**

You have been provided with 2 datasets. You will be learning here how to create the dataframe from 2 datasets and make some minor changes as required.

Recognize the attributes carefully and make sure they are aligned in proper format.

## Task1

1. Import all the relevant packages (Eg: Numpy, Seaborn...)
2. Import the datasets into the python environment.
3. Check the structure, statistics and other important functions. (Only observe the changes)

## Task2

1. Create a new dataframe “df” by joining the 2 datasets
2. Drop the duplicate data
3. Check for missing values

# **Scenario2**

You have successfully created the dataframe from the two input files.

Here we will be processing cleaning operations and intro to brief analysis.

Expected shape of the dataset: 12491 rows and 8 columns

## Task

1. There is a column that needs string strip operation. Identify that and apply it.
2. Fill the missing value by ‘Others’ in the column containing it
3. Since all the column names are single word so you can convert the ‘Price (INR)’ also to single name ‘Price’.
4. Analyse the Gender column and include your viewpoints how to make it useful.

# **Scenario3**

So far we have learnt the basics of the dataset and cleaned it as required. Over here you are going to perform deep analysis of the dataset with the help of data manipulation tricks as well as visualize the results.

This is the most time consuming tasks and make sure you do perform proper analysis method. While answering the question against all the tasks, it will be great if you can create charts to support it also.

Expected shape of the dataset: 12491 rows and 8 columns

## Task1

1. Univariate analysis of each variable
2. Bivariate Analysis of categorical vs numerical variables (Take target variable as fixed variable here)
3. Multivariate Analysis of categorical and numerical variables
4. Check distribution of variables

## Task2

1. Create a new Column “NewGender” to analyse further its distribution. Going forward we will consider this group for tasks

Logic Applied

* + 1. Include Boys & Men as Men
    2. Include Girls & Women as Women
    3. Include Unisex & Unisex Kids as Unisex

1. Complete the analysis of NewGender along with other categorical cols.

## Task3

1. Create a new Column “DescriptionLength” to analyse further its distribution.

Logic Applied

* + 1. Each record of DescriptionLength is equal to the number of chars in Description

1. Complete the analysis of DescriptionLength along with other categorical cols.
2. Isn’t it important to check if attribute information is also included in Description? Complete this task before answering it.