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.

Data Information

There are 2 csv files that are shared here.

A. Product Details
- ProductID – ID assigned to the product
- ProductName – Name of the Product
- ProductBrand – Brand Name of the Product

ProductID	ProductName	ProductBrand
10017413	DKNY Unisex Black & Grey Printed Medium Trolley Bag	DKNY
10016283	EthnoVogue Women Beige & Grey Made to Measure Custom Made Kurta Set with Jacket	EthnoVogue
10009781	SPYKAR Women Pink Alexa Super Skinny Fit High-Rise Clean Look Stretchable Cropped Jeans	SPYKAR
10015921	Raymond Men Blue Self-Design Single-Breasted Bandhgala Suit	Raymond
10017833	Parx Men Brown & Off-White Slim Fit Printed Casual Shirt	Parx
10014361	SHOWOFF Men Brown Solid Slim Fit Regular Shorts	SHOWOFF
10017869	Parx Men Blue Slim Fit Checked Casual Shirt	Parx
10009695	SPYKAR Women Burgundy Alexa Super Skinny Fit High-Rise Clean Look Stretchable Ankle Jeans	SPYKAR
10000571	Parx Men Brown Tapered Fit Solid Regular Trousers	Parx
10017421	DKNY Unisex Black Large Trolley Bag	DKNY

B. 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

Gender	Price (INR)	NumImages	ID	Description	PrimaryColor
Women	593	5	10182131	Off-White printed woven A-line top, has a round neck,	White
Men	6396	5	10052169	Display: AnalogueMovement: QuartzPower source: Bat	Black
Women	599	5	10168643	Blue printed knitted regular top, has a high neck, and t	Blue
Men	1034	5	10155427	Grey and Purple checked smart casual shirt, has a sprea	Grey
Men	477	3	10191059	Beige and Brown striped beltReversible: NoStretchable	Brown
Women	1715	5	10266533	Red & golden bandhani print Anarkali kurta with panel	Red
Women	1227	5	10174691	A pair of beige round toe sandals, has regular styling, c	Beige
Unisex	749	4	10097063	Set content: Single Door curtainsColour: RedPattern: St	Red
Unisex	1497	5	10111417	Set content: 3 Cushion Covers Colour: Green Shape: Sq	Green
Women	995	4	10216749	Nude-Coloured solid full-coverage T-shirt bra Lightly Pa	Red
Unisex	699	3	10018875	Set content: One runnerType: Table runnerColour: Gre	Green
Women	3490	5	10073703	Pink solid handheld bag, has a zip closure1 main compa	Pink

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
   i. Include Boys & Men as Men
   ii. Include Girls & Women as Women
   iii. Include Unisex & Unisex Kids as Unisex
2. Complete the analysis of NewGender along with other categorical cols.

Task3

1. Create a new Column “DescriptionLength” to analyse further its distribution.
   Logic Applied
   a. Each record of DescriptionLength is equal to the number of chars in Description
2. Complete the analysis of DescriptionLength along with other categorical cols.
3. Isn’t it important to check if attribute information is also included in Description? Complete this task before answering it.
4. Also check if attribute information is also included in ProductName

Task4

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

   Logic Applied

   i. Include Boys, Girls & Unisex Kids as Kids
   ii. Include Men, Women & Unisex as Adults

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