

CA 1

CSE353

EDA Project

Submitted By : Prakriti Verma

Reg No. : 12018071

Roll no. : 40

Section: K20CH

Introduction to dataset.

Topic: Myntra Fashion

A single stop for all things fashion.

Myntra is a popular name among fashion enthusiasts, by fashion enthusiasts we mean quite everyone. Everyone wants to make a style statement and everyone wants to stand out. In this sort of environment, Myntra is a perfect destination for these people. It is a place that caters to a lot of demands in a single and simple setting. A single stop for all things fashion.

Basic Structure

```
: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 526564 entries, 0 to 526563
Data columns (total 13 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   URL                                    526564 non-null object
1   Product_id                           526564 non-null int64
2   BrandName                            526564 non-null object
3   Category                             526564 non-null object
4   Individual_category                  526564 non-null object
5   category_by_Gender                  526564 non-null object
6   Description                           526564 non-null object
7   DiscountPrice (in Rs)                333406 non-null float64
8   OriginalPrice (in Rs)                526564 non-null float64
9   DiscountOffer                        452258 non-null object
10  SizeOption                           526564 non-null object
11  Ratings                              190412 non-null float64
12  Reviews                              190412 non-null float64
dtypes: float64(4), int64(1), object(8)
memory usage: 52.2+ MB
```

This Dataset contains 526k rows and 13 columns:

1. URL
2. Product ID
3. Brand Name
4. Category of Product
5. Sub Category of each Product
6. Category by Gender
7. Description of the Product
8. Price after Discount
9. Original price of the Product
10. Discount Offer
11. Size Option
12. Ratings
13. Reviews

Why this dataset

Fashion, by definition, is anything that someone wears to make a style statement. The definition might look easy in theory but in practice, the word 'fashion' has a lot of potentials. This can take any shape and size, and this can make truckloads of money if it is directed in a good right direction.

The cycle of fashion mostly starts when people observe some influential person making a style statement. Then they look out for the same thing, to make the same style statement. Take the case of Luxottica. The famous luxury eyewear group enjoys a monopoly in the eyewear industry. It all started when the brand pulled in some actors and

designers to promote luxury frames. This was an instant hit and the beginning of huge cash flows.

I, myself am a fashion enthusiast and started my online clothing shopping journey with Myntra that too just a couple of years back. I even ordered from other clothing platforms but found Myntra to be the best from all.

I being a data enthusiast wanted to know that how and what factors affect online shopping hence chose this dataset.

Using data analysis tools like Pandas, NumPy, matplotlib and seaborn I can find many relations that affect a product to be sold.

For basic instance here's how top 5 rows look like:

```
df.head()
```

URL	Product_id	BrandName	Category	Individual_category	category_by_Gender	Description	DiscountPrice (in Rs)	OriginalPrice (in Rs)
://www.myntra.com/jeans/roadster/roadster...	2296012	Roadster	Bottom Wear	jeans	Men	roadster men navy blue slim fit mid rise clean...	824.0	1499.0
://www.myntra.com/track-pants/locomotive/...	13780156	LOCOMOTIVE	Bottom Wear	track-pants	Men	locomotive men black white solid slim fit tra...	517.0	1149.0
://www.myntra.com/shirts/roadster/roadste...	11895968	Roadster	Topwear	shirts	Men	roadster men navy white black geometric print...	629.0	1399.0
ww.myntra.com/shapewear/zivame/zivame...	4335679	Zivame	Lingerie & Sleep Wear	shapewear	Women	zivame women black saree shapewear zi3023core0...	893.0	1266.0
k://www.myntra.com/tshirts/roadster/roadst...	11690882	Roadster	Western	tshirts	Women	roadster women white solid v neck pure cotton ...	NaN	599.0

What are the insights we will be finding here

1. We will find which Brand sells the most.

2. What product size is sold the most.
3. How the color of the product affects the selling frequency of the item.
4. Which product(s) has/have the most rating.
5. Which product has the most reviews.
6. Which product has been the most discounted.
7. Which category of the product has been sold the most.
8. Which brand sells which product the most.
9. Which brand gives the most discount on average.
10. Finally, we will be looking at the correlations of this data and conclude our study.

We will be using powerful libraries like pandas, NumPy, seaborn, matplotlib etc to demonstrate this data on graphs and charts.