TATA 1mg

Introduction -

- About Dataset

Tata 1mg, previously 1mg, is a healthcare platform based in Gurgaon, India. It provides services, including e-pharmacy, diagnostics, e-consultation and health content. It was founded in April 2015 by Prashant Tandon.

This dataset contains list of medicines.

The dataset contains 16,944 rows and 6 columns. The columns are as follows: Column 0: index of medicines at particular page Name: Name of medicine pack_size: The description of packaging rating_count: total number of reviews rating: Rating of product price: Sell price mrp: max retail price

Thank you!

Exploratory Data Analysis (EDA) -

Importing Libraries -

In [2]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

In [3]:

```
data = pd.read_csv("D:\\Data Projects\\Python EDA\\1mg\\1mgData.csv")
```

Understanding Dataset-

In [14]:

```
1 data.head(50)
```

Out[14]:

	Name	pack_size	rating_count	rating	price	mrp
0	Tata 1mg Glucosamine HCL 1500 mg Tablets for J...	bottle of 60 tablets	284 ratings	4.1	₹995	₹399
1	Tata 1mg Multivitamin Supreme, Zinc, Calcium a...	bottle of 60 capsules	820 ratings	4.1	₹995	₹498
2	HealthVit Lgm 500mg Capsule	bottle of 60 capsules	No reviews	Nan	₹1200	₹980
3	Zincovit Tablet	strip of 15 tablets	4433 ratings	4.5	₹105	₹89
4	Tata 1mg Vitamin B Complex Capsules	bottle of 60 capsules	7 ratings	4.6	₹595	₹448
5	Shelcal 500 Tablet	strip of 15 tablets	4345 ratings	4.5	₹119.5	₹111
6	Swisse Ultiboost Co-Enzyme Q10 150mg Capsule	bottle of 50 capsules	No reviews	Nan	₹1899	₹1291
7	Depura 60000 IU Vitamin D3 Oral Solution, Help...	bottle of 5 ml Oral Solution	603 ratings	4.5	₹94.99	₹93
8	Shelcal XT Tablet	strip of 15 tablets	772 ratings	4.5	₹377.25	₹352
9	Tata 1mg Probiotics 30 Billion CFUs+ Capsule w...	bottle of 60 capsules	12 ratings	4.8	₹595	₹299
10	HealthVit Inositol 650mg Capsule	bottle of 60 capsules	No reviews	Nan	₹900	₹747
11	D-Rise 60K Capsule	strip of 4 soft gelatin capsules	1506 ratings	4.5	₹146.9	₹121
12	Evion 400mg Capsule	strip of 10 capsules	5882 ratings	4.5	₹35.9	₹32
13	TrueBasics Multivit Women Tablet	bottle of 90 tablets	17 ratings	4.2	₹1699	₹1478
14	HealthVit L-Glutathione Reduced 100mg Capsule	bottle of 60 capsules	17 ratings	3.2	₹1400	₹1161
15	Tata 1mg Calcium + Vitamin D3, Zinc, Magnesium...	bottle of 60 tablets	935 ratings	4.2	₹595	₹248
16	Tata 1mg Salmon Omega 3 Fish Oil Capsule	bottle of 60 capsules	713 ratings	4.4	₹995	₹498
17	Revital H Woman Tablet	bottle of 30 tablets	1180 ratings	4.5	₹345	₹325
18	Jointcart Plus Tablet	strip of 10 tablets	No reviews	Nan	No discounts	MRP₹310
19	Tata 1mg Melatonin 10mg Vegetarian Capsule for...	bottle of 30 capsules	34 ratings	4.1	₹510	₹256
20	Shelcal - HD Tablet	strip of 15 tablets	1449 ratings	4.5	₹113.6	₹109
21	Neurobion Forte Tablet	strip of 30 tablets	6269 ratings	4.5	₹34.7	₹33
22	Terravita Sleep Aid Vegetarian Tablet	box of 28 tablets	No reviews	Nan	₹899	₹497
23	Becosules Capsule	strip of 20 capsules	2978 ratings	4.5	₹50.23	₹43
24	Folvite 5mg Tablet	strip of 45 tablets	No reviews	Nan	₹76.54	₹71
25	Tata 1mg Senior 50+ Multivitamin & Multiminera...	bottle of 60 tablets	274 ratings	4.2	₹895	₹399

	Name	pack_size	rating_count	rating	price	mrp
26	Baidyanath (Nagpur) Rheumartho Tablet	bottle of 50 tablets	No reviews	Nan	₹225	₹191
27	Becosules Z Capsule	strip of 20 capsules	2518 ratings	4.6	₹45.96	₹42
28	Tata 1mg Immunity & Wellness Supplement Tablet	bottle of 15 tablets	24 ratings	4.5	₹99	₹90
29	Orofer XT Tablet	strip of 10 tablets	809 ratings	4.3	₹172.4	₹170
30	Becadexamin Soft Gelatin Capsule	bottle of 30 soft gelatin capsules	2786 ratings	4.5	No discounts	MRP₹46.4
31	Seacod Cod Liver Oil Capsule	bottle of 110 soft gelatin capsules	887 ratings	4.5	₹352.98	₹343
32	Cipcal 500 Tablet	strip of 15 tablets	676 ratings	4.5	₹86.5	₹83
33	Tata 1mg Calcium & Vitamin D Supplement Tablet	bottle of 15 tablets	21 ratings	4.6	₹99	₹90
34	New A to Z Gold Soft Gelatin Capsule	strip of 15 soft gelatin capsules	898 ratings	4.5	₹200	₹170
35	New Celin 500 Tablet	strip of 25 tablets	1949 ratings	4.4	₹42.84	₹36
36	Electral Powder	sachet of 21.8 gm Powder	2651 ratings	4.6	₹21.95	₹19
37	Seven Seas Original Cod-Liver Oil Capsule	bottle of 100 soft gelatin capsules	2266 ratings	4.5	₹320.9	₹311
38	Meganeuron OD Plus Capsule	strip of 10 capsules	No reviews	Nan	₹140	₹132
39	Tata 1mg Biotin + Tablet	bottle of 60 tablets	221 ratings	4.2	₹595	₹348
40	Tata 1mg Women's Multivitamin, Zinc, Vitamin C...	bottle of 60 tablets	364 ratings	4.2	₹995	₹399
41	Rejunex CD 3 Tablet	strip of 10 tablets	615 ratings	4.4	₹220	₹208
42	Cobadex Czs Tablet	strip of 15 tablets	963 ratings	4.4	₹102.7	₹93
43	Seven Seas Original Cod-Liver Oil Capsule	bottle of 500 soft gelatin capsules	2266 ratings	4.5	₹1217.8	₹1202
44	Revital H Capsule	bottle of 60 soft gelatin capsules	1741 ratings	4.4	₹550	₹540
45	Calcirol Sachet	sachet of 1 gm Granules	1791 ratings	4.5	₹56.4	₹54
46	Keraglo Men Tablet	bottle of 30 tablets	399 ratings	4.1	No discounts	MRP₹642
47	Livogen Captab	strip of 15 captabs	714 ratings	4.5	₹76.1	₹71
48	Ostocalcium Plus Chewable Tablet	bottle of 30 Chewable Tablets	863 ratings	4.6	₹196.8	₹196
49	NutriBears Multivitamin Growth & Immunity Supp...	bottle of 30 gummies	8 ratings	4.6	₹399	₹376

In [5]:

```
1 data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16944 entries, 0 to 16943
Data columns (total 6 columns):
 #   Column        Non-Null Count  Dtype 
---  ------        --------------  ----- 
 0   Name          16944 non-null  object
 1   pack_size     16944 non-null  object
 2   rating_count  16944 non-null  object
 3   rating        16944 non-null  object
 4   price         16944 non-null  object
 5   mrp           16944 non-null  object
dtypes: object(6)
memory usage: 794.4+ KB
```

In [7]:

```
1 data.shape
```

Out[7]:

```
(16944, 6)
```

In [8]:

```
1 data.columns
```

Out[8]:

```
Index(['Name', 'pack_size', 'rating_count', 'rating', 'price', 'mrp'], dty
pe='object')
```

In [9]:

```
1 data.isnull().sum() #Their are no null values ,But their might be some unusual value.
```

Out[9]:

```
Name            0
pack_size       0
rating_count    0
rating          0
price           0
mrp             0
dtype: int64
```

Handling columns one by one -

In [15]:

```
1 data.columns
```

Out[15]:

```
Index(['Name', 'pack_size', 'rating_count', 'rating', 'price', 'mrp'], dty
pe='object')
```

Column - "Name"

In [16]:

```
1 df = data
```

In [18]:

```
1 df["Name"].unique()
```

Out[18]:

```
array(['Tata 1mg Glucosamine HCL 1500 mg Tablets for Joint Health with Bos
wellia Serrata, Collagen Peptide, L-Arginine, and Curcuma Longa',
       'Tata 1mg Multivitamin Supreme, Zinc, Calcium and Vitamin D Immunit
y Booster Capsule',
       'HealthVit Lgm 500mg  Capsule', ..., 'Marshel Plus Capsule',
       'Bageo Fertiperk-F Softgel Capsule', 'Nuzink Syrup'], dtype=object)
```

Column - "pack_size"

In [19]:

```
1 df["pack_size"].unique()
```

Out[19]:

```
array(['bottle of 60 tablets', 'bottle of 60 capsules',
       'strip of 15 tablets', ..., 'strip of 5 capsules',
       'bottle of 112 gm Lotion', 'packet of 105 gm Sachet'], dtype=objec
t)
```

In [20]:

```
1 df["pack_size"].value_counts()
```

Out[20]:

```
strip of 10 tablets            3592
strip of 10 capsules           1486
bottle of 60 capsules           835
bottle of 60 tablets            810
bottle of 200 ml Syrup          807
                               ...
strip of 10 soflets               1
sachet of 2.25 gm Granules        1
box of 175 gm Powder              1
bottle of 200 ml Oral Drops       1
packet of 105 gm Sachet           1
Name: pack_size, Length: 1007, dtype: int64
```

In [22]:

```
def handlingpacksize(value):
    value = value.split(" ")
    if value[2] =="of":
        return float(value[3])
    else :
        return float(value[2])
df["size"] = df["pack_size"].apply(handlingpacksize)
df["size"].unique
```

Out[22]:

```
<bound method Series.unique of 0         60.0
1         60.0
2         60.0
3         15.0
4         60.0
         ...
16939     10.0
16940    100.0
16941     10.0
16942     10.0
16943     60.0
Name: size, Length: 16944, dtype: float64>
```

In [23]:

```
df.head(10)
```

Out[23]:

	Name	pack_size	rating_count	rating	price	mrp	size
0	Tata 1mg Glucosamine HCL 1500 mg Tablets for J...	bottle of 60 tablets	284 ratings	4.1	₹995	₹399	60.0
1	Tata 1mg Multivitamin Supreme, Zinc, Calcium a...	bottle of 60 capsules	820 ratings	4.1	₹995	₹498	60.0
2	HealthVit Lgm 500mg Capsule	bottle of 60 capsules	No reviews	Nan	₹1200	₹980	60.0
3	Zincovit Tablet	strip of 15 tablets	4433 ratings	4.5	₹105	₹89	15.0
4	Tata 1mg Vitamin B Complex Capsules	bottle of 60 capsules	7 ratings	4.6	₹595	₹448	60.0
5	Shelcal 500 Tablet	strip of 15 tablets	4345 ratings	4.5	₹119.5	₹111	15.0
6	Swisse Ultiboost Co-Enzyme Q10 150mg Capsule	bottle of 50 capsules	No reviews	Nan	₹1899	₹1291	50.0
7	Depura 60000 IU Vitamin D3 Oral Solution, Help...	bottle of 5 ml Oral Solution	603 ratings	4.5	₹94.99	₹93	5.0
8	Shelcal XT Tablet	strip of 15 tablets	772 ratings	4.5	₹377.25	₹352	15.0
9	Tata 1mg Probiotics 30 Billion CFUs+ Capsule w...	bottle of 60 capsules	12 ratings	4.8	₹595	₹299	60.0

- We will also create a column name Type to have an understanding of what type size is given.

In [24]:

```
def creatingcolumnsize(value):
    value = value.split(" ")
    return value[-1]
df["Type"] = df["pack_size"].apply(creatingcolumnsize)
df["Type"].unique()
```

Out[24]:

```
array(['tablets', 'capsules', 'Solution', 'Powder', 'Granules', 'captabs',
       'Tablets', 'gummies', 'bottles', 'Syrup', 'Drops', 'Liquid',
       'Suspension', 'Drop', 'cr', 'er', 'Serum', 'sr', 'soflets',
       'tablet', 'ir', 'Tablet', 'Kit', 'Sachet', 'vegicaps', 'Sachets',
       'Packs', 'Lotion', 'Tonic', 'softules', 'Bottle', 'Gel',
       'Dilution', 'strips', 'Paste', 'solutions', 'Soap', 'vials',
       'Nuts', 'Diskette', 'Bar', 'caplets', 'candies', 'Spray', 'Oil',
       'Seeds', 'jar', 'syrups', 'Gummy', 'Churna', 'Oats', 'Juice',
       'Jar', 'Globules', 'lozenges', 'Gums', 'Sanitizer', 'Cream',
       'Elixir', 'Muesli', 'md', 'DR', 'Pack', 'Butter', 'masks', 'Wash',
       'Shampoo', 'Scrub', 'Ointment', 'Salt', 'biscuits', 'dt', 'Strip',
       'Capsule', 'Pellets', 'bags', 'tr', 'boxes', 'combikit', 'jelly',
       'kit', 'bars', 'pellets', 'granules', 'Emulsion', 'diskettes',
       'gums', 'Candy', 'powders', 'Gargle', 'servings', 'Flakes',
       'Tubes', 'xl', 'gels', 'Jelly', 'Microgranules'], dtype=object)
```

Column -"rating_count"

```
- We can observe it has 'No reviews' and "ratings" written which is not required.
```

In [27]:

```
df["rating_count"].value_counts()
```

Out[27]:

```
No reviews     14014
5 ratings        291
8 ratings        213
6 ratings        200
7 ratings        132
               ...
357 ratings        1
300 ratings        1
148 ratings        1
307 ratings        1
143 ratings        1
Name: rating_count, Length: 298, dtype: int64
```

In [43]:

```
def handlingrating_count(value):
    value = value.split(" ")
    if value[0] == "No":
        return np.nan
    else :
        return int(value[0])
df["ratings"] = df["rating_count"].apply(handlingrating_count)
df["ratings"].unique()
```

Out[43]:

```
array([2.840e+02, 8.200e+02,       nan, 4.433e+03, 7.000e+00, 4.345e+03,
       6.030e+02, 7.720e+02, 1.200e+01, 1.506e+03, 5.882e+03, 1.700e+01,
       9.350e+02, 7.130e+02, 1.180e+03, 3.400e+01, 1.449e+03, 6.269e+03,
       2.978e+03, 2.740e+02, 2.518e+03, 2.400e+01, 8.090e+02, 2.786e+03,
       8.870e+02, 6.760e+02, 2.100e+01, 8.980e+02, 1.949e+03, 2.651e+03,
       2.266e+03, 2.210e+02, 3.640e+02, 6.150e+02, 9.630e+02, 1.741e+03,
       1.791e+03, 3.990e+02, 7.140e+02, 8.630e+02, 8.000e+00, 8.310e+02,
       5.430e+02, 6.000e+00, 2.110e+02, 5.090e+02, 4.760e+02, 5.100e+01,
       4.580e+02, 6.480e+02, 4.290e+02, 4.120e+02, 4.570e+02, 5.000e+00,
       3.730e+02, 4.850e+02, 1.061e+03, 3.220e+02, 2.040e+02, 8.380e+02,
       3.570e+02, 3.000e+02, 2.830e+02, 3.070e+02, 1.440e+02, 3.170e+02,
       7.060e+02, 4.790e+02, 3.420e+02, 1.670e+02, 3.430e+02, 2.300e+02,
       2.610e+02, 2.760e+02, 8.490e+02, 2.060e+02, 2.940e+02, 4.260e+02,
       1.200e+02, 4.500e+01, 2.750e+02, 2.560e+02, 4.000e+01, 2.530e+02,
       3.310e+02, 3.580e+02, 1.100e+01, 2.400e+02, 1.570e+02, 2.000e+01,
       1.600e+01, 2.010e+02, 2.920e+02, 3.550e+02, 3.080e+02, 1.310e+02,
       2.470e+02, 2.370e+02, 1.950e+02, 7.540e+02, 1.550e+02, 2.120e+02,
       5.490e+02, 1.880e+02, 1.450e+02, 3.050e+02, 2.330e+02, 1.680e+02,
       5.000e+01, 1.180e+02, 1.580e+02, 1.780e+02, 4.600e+01, 1.590e+02,
       1.830e+02, 6.490e+02, 1.370e+02, 3.810e+02, 3.090e+02, 3.800e+01,
       2.600e+01, 1.610e+02, 2.340e+02, 8.200e+01, 1.120e+02, 9.900e+01,
       1.130e+02, 1.870e+02, 8.800e+01, 1.500e+02, 1.960e+02, 1.300e+02,
       1.050e+02, 1.380e+02, 8.900e+01, 1.480e+02, 1.470e+02, 9.000e+01,
       2.310e+02, 1.340e+02, 1.490e+02, 1.220e+02, 9.800e+01, 1.280e+02,
       2.250e+02, 1.160e+02, 1.990e+02, 1.800e+02, 1.800e+01, 7.200e+01,
       1.910e+02, 1.660e+02, 4.300e+01, 1.010e+02, 1.070e+02, 4.400e+01,
       8.400e+01, 3.010e+02, 1.040e+02, 3.400e+02, 4.700e+01, 1.740e+02,
       6.900e+01, 1.900e+01, 1.760e+02, 3.600e+01, 7.000e+01, 1.360e+02,
       3.500e+01, 1.150e+02, 1.000e+02, 9.300e+01, 1.020e+02, 1.060e+02,
       4.100e+01, 1.350e+02, 2.140e+02, 9.200e+01, 1.850e+02, 9.100e+01,
       1.330e+02, 7.300e+01, 1.210e+02, 7.400e+01, 6.700e+01, 1.190e+02,
       1.110e+02, 3.700e+01, 7.900e+01, 5.800e+01, 6.600e+01, 7.600e+01,
       5.400e+01, 1.640e+02, 1.510e+02, 2.800e+01, 6.400e+01, 1.500e+01,
       6.100e+01, 7.700e+01, 9.700e+01, 4.200e+01, 7.100e+01, 5.300e+01,
       8.300e+01, 9.000e+00, 2.500e+01, 4.360e+02, 1.270e+02, 1.320e+02,
       5.600e+01, 2.900e+01, 1.400e+01, 1.980e+02, 2.300e+01, 6.000e+01,
       3.000e+01, 8.500e+01, 9.400e+01, 6.800e+01, 1.000e+01, 8.600e+01,
       3.200e+01, 3.900e+01, 5.500e+01, 1.650e+02, 6.200e+01, 7.800e+01,
       5.200e+01, 2.490e+02, 5.900e+01, 2.200e+01, 4.900e+01, 3.300e+01,
       5.700e+01, 8.100e+01, 4.800e+01, 1.080e+02, 2.700e+01, 1.300e+01,
       2.921e+03, 1.390e+02, 3.100e+01, 2.150e+02, 6.300e+01, 2.500e+02,
       8.770e+02, 8.000e+01, 1.240e+02, 9.020e+02, 2.460e+02, 1.002e+03,
       1.140e+02, 4.430e+02, 2.180e+02, 5.800e+02, 6.960e+02, 2.320e+02,
       1.620e+02, 8.170e+02, 1.710e+02, 6.500e+01, 2.860e+02, 5.700e+02,
       2.680e+02, 6.900e+02, 2.730e+02, 9.770e+02, 2.160e+02, 3.240e+02,
       2.910e+02, 7.500e+01, 1.090e+02, 2.390e+02, 1.170e+02, 4.080e+02,
       1.930e+02, 1.307e+03, 1.750e+02, 1.250e+02, 1.100e+02, 1.690e+02,
       2.030e+02, 1.230e+02, 1.820e+02, 5.870e+02, 1.030e+02, 9.500e+01,
       1.257e+03, 2.080e+02, 5.260e+02, 2.880e+02, 3.260e+02, 1.690e+03,
       2.990e+02, 3.110e+02, 1.460e+02, 1.430e+02])
```

In [35]:

```
1 df.columns
```

Out[35]:

```
Index(['Name', 'pack_size', 'rating_count', 'rating', 'price', 'mrp', 'siz
e',
       'Type', 'ratings'],
      dtype='object')
```

In [46]:

```
1 df.head()
```

Out[46]:

	Name	pack_size	rating_count	rating	price	mrp	size	Type	ratings	Avg_rati
0	Tata 1mg Glucosamine HCL 1500 mg Tablets for J...	bottle of 60 tablets	284 ratings	4.1	₹995	₹399	60.0	tablets	284.0	
1	Tata 1mg Multivitamin Supreme, Zinc, Calcium a...	bottle of 60 capsules	820 ratings	4.1	₹995	₹498	60.0	capsules	820.0	
2	HealthVit Lgm 500mg Capsule	bottle of 60 capsules	No reviews	Nan	₹1200	₹980	60.0	capsules	NaN	N
3	Zincovit Tablet	strip of 15 tablets	4433 ratings	4.5	₹105	₹89	15.0	tablets	4433.0	
4	Tata 1mg Vitamin B Complex Capsules	bottle of 60 capsules	7 ratings	4.6	₹595	₹448	60.0	capsules	7.0	

Column - "rating"

In [40]:

```
1 df["rating"].unique()
```

Out[40]:

```
array(['4.1', 'Nan', '4.5', '4.6', '4.8', '4.2', '3.2', '4.4', '4.3',
       '3.8', '4', '3.7', '4.7', '3.9', '3.4', '4.9', '5', '3.6', '3.3',
       '3.5', '2.6', '2.2', '3', '3.1', '2.7', '2.8', '2', '1.7', '2.3',
       '1.4', '2.9', '2.4'], dtype=object)
```

- We need to convert the values in the Float type.

In [41]:

```
def handlingrating(value):
    if value == "Nan":
        return np.nan
    else :
        return float(value)
df["Avg_rating"] = df["rating"].apply(handlingrating)
df["Avg_rating"].unique()
```

Out[41]:

```
array([4.1, nan, 4.5, 4.6, 4.8, 4.2, 3.2, 4.4, 4.3, 3.8, 4. , 3.7, 4.7,
       3.9, 3.4, 4.9, 5. , 3.6, 3.3, 3.5, 2.6, 2.2, 3. , 3.1, 2.7, 2.8,
       2. , 1.7, 2.3, 1.4, 2.9, 2.4])
```

In [45]:

```
df.head()
```

Out[45]:

	Name	pack_size	rating_count	rating	price	mrp	size	Type	ratings	Avg_rati
0	Tata 1mg Glucosamine HCL 1500 mg Tablets for J...	bottle of 60 tablets	284 ratings	4.1	₹995	₹399	60.0	tablets	284.0	
1	Tata 1mg Multivitamin Supreme, Zinc, Calcium a...	bottle of 60 capsules	820 ratings	4.1	₹995	₹498	60.0	capsules	820.0	
2	HealthVit Lgm 500mg Capsule	bottle of 60 capsules	No reviews	Nan	₹1200	₹980	60.0	capsules	NaN	N
3	Zincovit Tablet	strip of 15 tablets	4433 ratings	4.5	₹105	₹89	15.0	tablets	4433.0	
4	Tata 1mg Vitamin B Complex Capsules	bottle of 60 capsules	7 ratings	4.6	₹595	₹448	60.0	capsules	7.0	

Column - "price" -

In [47]:

```
1 df["price"].value_counts()
```

Out[47]:

```
No discounts    4325
₹999             196
₹120             187
₹399             180
₹500             151
                ...
₹111.61            1
₹143.9             1
₹58.02             1
₹87.44             1
₹122.27            1
Name: price, Length: 2555, dtype: int64
```

In [59]:

```
1 df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16944 entries, 0 to 16943
Data columns (total 11 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   Name          16944 non-null  object 
 1   pack_size     16944 non-null  object 
 2   rating_count  16944 non-null  object 
 3   rating        16944 non-null  object 
 4   price         16944 non-null  float64
 5   mrp           16944 non-null  object 
 6   size          16944 non-null  float64
 7   Type          16944 non-null  object 
 8   ratings       2930 non-null   float64
 9   Avg_rating    2930 non-null   float64
 10  cost          12619 non-null  float64
dtypes: float64(5), object(6)
memory usage: 1.4+ MB
```

In [ ]:

```
1 
```

In [57]:

```
1 df['price'] = df['price'].apply(lambda x: x.replace('₹', '')) #remove the '₹' in the 
2 df['price'] = df['price'].replace('No discounts', 0) #convert the "no discount" value
3 df['price'] = df['price'].astype('float')
```

In [58]:

```
1 df.head()
```

Out[58]:

	Name	pack_size	rating_count	rating	price	mrp	size	Type	ratings	Avg_rat
0	Tata 1mg Glucosamine HCL 1500 mg Tablets for J...	bottle of 60 tablets	284 ratings	4.1	995.0	₹399	60.0	tablets	284.0	
1	Tata 1mg Multivitamin Supreme, Zinc, Calcium a...	bottle of 60 capsules	820 ratings	4.1	995.0	₹498	60.0	capsules	820.0	
2	HealthVit Lgm 500mg Capsule	bottle of 60 capsules	No reviews	Nan	1200.0	₹980	60.0	capsules	NaN	N
3	Zincovit Tablet	strip of 15 tablets	4433 ratings	4.5	105.0	₹89	15.0	tablets	4433.0	
4	Tata 1mg Vitamin B Complex Capsules	bottle of 60 capsules	7 ratings	4.6	595.0	₹448	60.0	capsules	7.0	

Column - "mrp"

In [60]:

```
1 df['mrp'] = df['mrp'].apply(lambda x: x.replace('₹', ''))   #remove the '₹' in the pr
2 df['mrp'] = df['mrp'].apply(lambda x: x.replace('MRP', '')) #remove the "MRB" strnig
3 df['mrp'] = df['mrp'].astype('float')
```

In [69]:

```
1 df.head()
```

Out[69]:

	Name	pack_size	rating_count	rating	price	mrp	size	Type	ratings	Avg_ra
0	Tata 1mg Glucosamine HCL 1500 mg Tablets for J...	bottle of 60 tablets	284 ratings	4.1	995.0	399.0	60.0	tablets	284.0	
1	Tata 1mg Multivitamin Supreme, Zinc, Calcium a...	bottle of 60 capsules	820 ratings	4.1	995.0	498.0	60.0	capsules	820.0	
2	HealthVit Lgm 500mg Capsule	bottle of 60 capsules	No reviews	Nan	1200.0	980.0	60.0	capsules	NaN	
3	Zincovit Tablet	strip of 15 tablets	4433 ratings	4.5	105.0	89.0	15.0	tablets	4433.0	
4	Tata 1mg Vitamin B Complex Capsules	bottle of 60 capsules	7 ratings	4.6	595.0	448.0	60.0	capsules	7.0	

In [64]:

```
1 df.isnull().sum()
```

Out[64]:

```
Name            0
pack_size       0
rating_count    0
rating          0
price           0
mrp             0
size            0
Type            0
ratings         14014
Avg_rating      14014
dtype: int64
```

Handling NAN value-

In [84]:

```
1 df.fillna(0)
```

Out[84]:

	Name	pack_size	rating_count	rating	price	mrp	size	Type	ratings	A
0	Tata 1mg Glucosamine HCL 1500 mg Tablets for J...	bottle of 60 tablets	284 ratings	4.1	995.0	399.0	60.0	tablets	284.0	
1	Tata 1mg Multivitamin Supreme, Zinc, Calcium a...	bottle of 60 capsules	820 ratings	4.1	995.0	498.0	60.0	capsules	820.0	
2	HealthVit Lgm 500mg Capsule	bottle of 60 capsules	No reviews	Nan	1200.0	980.0	60.0	capsules	0.0	
3	Zincovit Tablet	strip of 15 tablets	4433 ratings	4.5	105.0	89.0	15.0	tablets	4433.0	
4	Tata 1mg Vitamin B Complex Capsules	bottle of 60 capsules	7 ratings	4.6	595.0	448.0	60.0	capsules	7.0	
...	...	...	...	...	...	...	...	...	...	
16939	Lycosave-4G Tablet	strip of 10 tablets	No reviews	Nan	189.0	180.0	10.0	tablets	0.0	
16940	Inovit Z Powder	box of 100 gm Powder	No reviews	Nan	30.0	26.0	100.0	Powder	0.0	
16941	Marshel Plus Capsule	strip of 10 capsules	No reviews	Nan	135.0	128.0	10.0	capsules	0.0	
16942	Bageo Fertiperk-F Softgel Capsule	strip of 10 soft gelatin capsules	No reviews	Nan	279.0	189.0	10.0	capsules	0.0	
16943	Nuzink Syrup	bottle of 60 ml Syrup	No reviews	Nan	0.0	45.0	60.0	Syrup	0.0	

16944 rows × 10 columns

In [86]:

```
1 df.isnull().sum()
```

Out[86]:

```
Name            0
pack_size       0
rating_count    0
rating          0
price           0
mrp             0
size            0
Type            0
ratings         0
Avg_rating      0
dtype: int64
```

In [108]:

```
1 df.head()
```

Out[108]:

	Name	pack_size	rating_count	rating	price	mrp	size	Type	ratings	Avg_ra
0	Tata 1mg Glucosamine HCL 1500 mg Tablets for J...	bottle of 60 tablets	284 ratings	4.1	995.0	399.0	60.0	tablets	284.0	
1	Tata 1mg Multivitamin Supreme, Zinc, Calcium a...	bottle of 60 capsules	820 ratings	4.1	995.0	498.0	60.0	capsules	820.0	
2	HealthVit Lgm 500mg Capsule	bottle of 60 capsules	No reviews	Nan	1200.0	980.0	60.0	capsules	0.0	
3	Zincovit Tablet	strip of 15 tablets	4433 ratings	4.5	105.0	89.0	15.0	tablets	4433.0	
4	Tata 1mg Vitamin B Complex Capsules	bottle of 60 capsules	7 ratings	4.6	595.0	448.0	60.0	capsules	7.0	

In [109]:

```
1 df.drop(["pack_size","rating_count","rating"],axis = 1,inplace =True)
```

In [110]:

```
1 df.head()
```

Out[110]:

	Name	price	mrp	size	Type	ratings	Avg_rating
0	Tata 1mg Glucosamine HCL 1500 mg Tablets for J...	995.0	399.0	60.0	tablets	284.0	4.1
1	Tata 1mg Multivitamin Supreme, Zinc, Calcium a...	995.0	498.0	60.0	capsules	820.0	4.1
2	HealthVit Lgm 500mg Capsule	1200.0	980.0	60.0	capsules	0.0	0.0
3	Zincovit Tablet	105.0	89.0	15.0	tablets	4433.0	4.5
4	Tata 1mg Vitamin B Complex Capsules	595.0	448.0	60.0	capsules	7.0	4.6

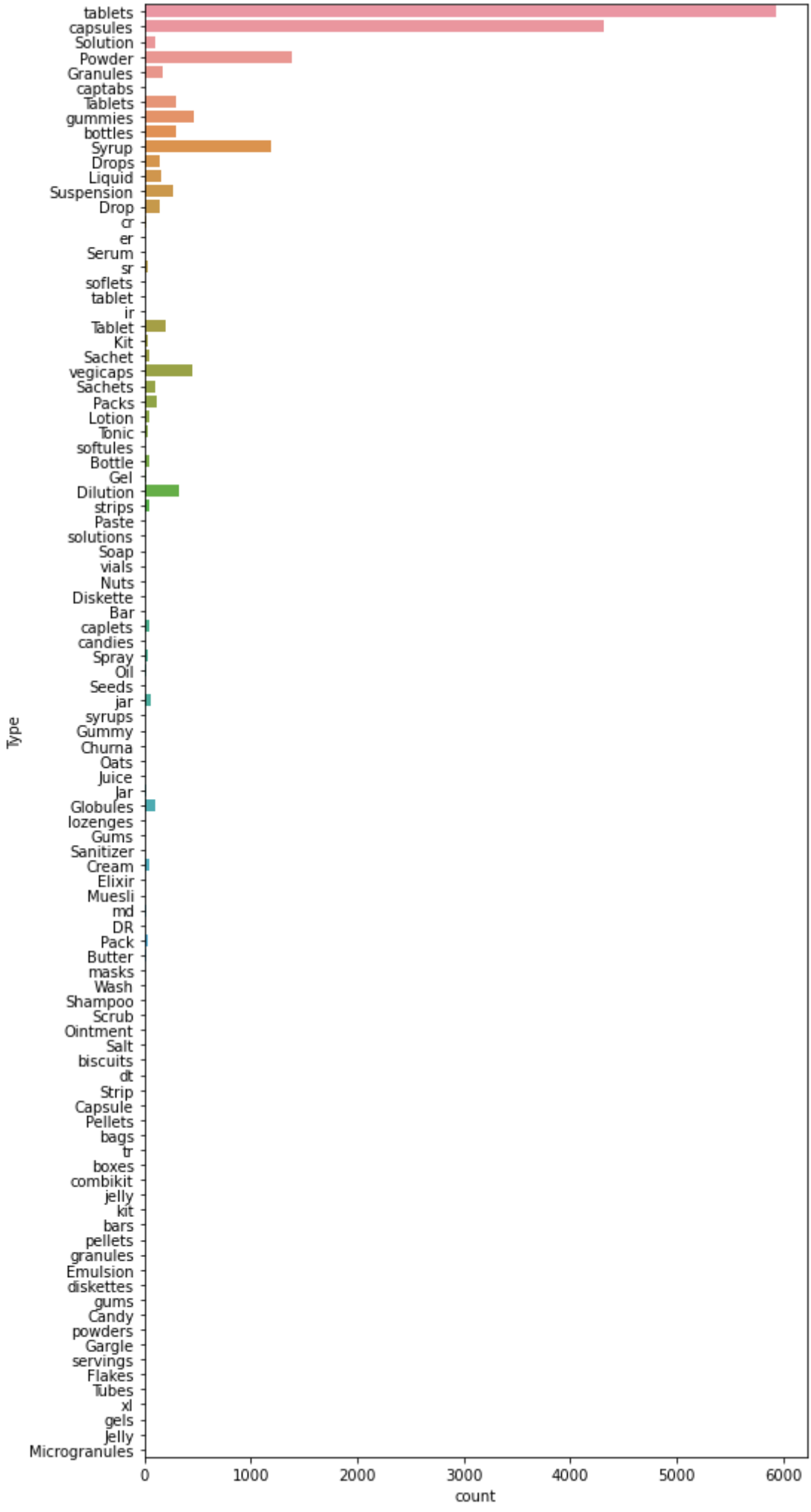
Visualisation-

In [*]:

```
1 sns.pairplot(df)
2 plt.show()
```

In [103]:

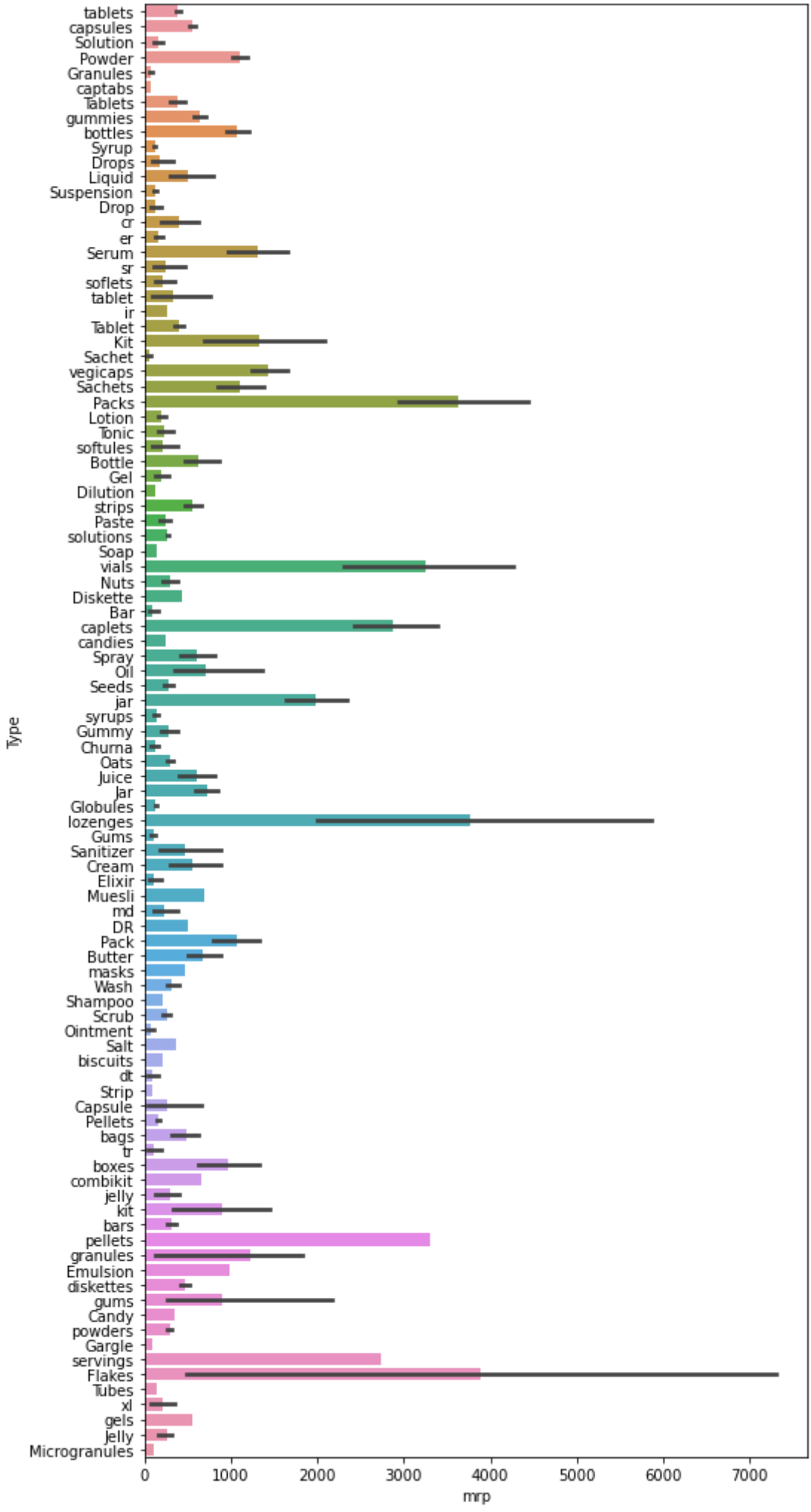
```
plt.figure(figsize = (8,18))
sns.countplot(y = "Type" , data =df )
plt.show()
```


tablets
capsules
Solution
Powder
Granules
captabs
Tablets
gummies
bottles
Syrup
Drops
Liquid
Suspension
Drop
cr
er
Serum
sr
soflets
tablet
ir
Tablet
Kit
Sachet
vegicaps
Sachets
Packs
Lotion
Tonic
softules
Bottle
Gel
Dilution
strips
Paste
solutions
Soap
vials
Nuts
Diskette
Bar
caplets
candies
Spray
Oil
Seeds
jar
syrups
Gummy
Churna
Oats
Juice
Jar
Globules
lozenges
Gums
Sanitizer
Cream
Elixir
Muesli
md
DR
Pack
Butter
masks
Wash
Shampoo
Scrub
Ointment
Salt
biscuits
dt
Strip
Capsule
Pellets
bags
tr
boxes
combikit
jelly
kit
bars
pellets
granules
Emulsion
diskettes
gums
Candy
powders
Gargle
servings
Flakes
Tubes
xl
gels
Jelly
Microgranules
Type
0
1000
2000
3000
4000
5000
6000
count

- Maximium tablets are sold by 1mg followed by capsule.

In [106]:

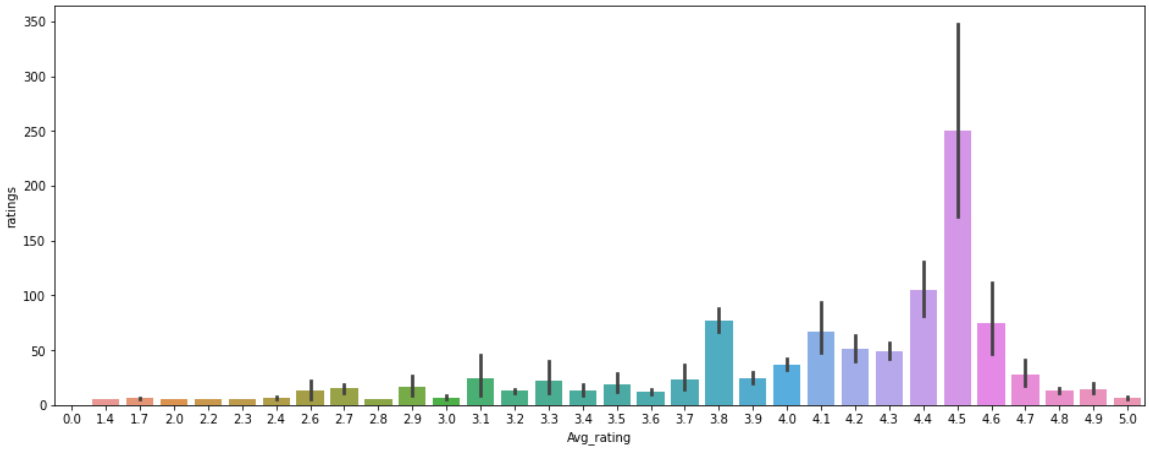
```
plt.figure(figsize = (8,18))
sns.barplot(x = "mrp" , y = "Type" , data =df )
plt.show()
```


tablets
capsules
Solution
Powder
Granules
captabs
Tablets
gummies
bottles
Syrup
Drops
Liquid
Suspension
Drop
cr
er
Serum
sr
soflets
tablet
ir
Tablet
Kit
Sachet
vegicaps
Sachets
Packs
Lotion
Tonic
softules
Bottle
Gel
Dilution
strips
Paste
solutions
Soap
vials
Nuts
Diskette
Bar
caplets
candies
Spray
Oil
Seeds
jar
syrups
Gummy
Churna
Oats
Juice
Jar
Globules
lozenges
Gums
Sanitizer
Cream
Elixir
Muesli
md
DR
Pack
Butter
masks
Wash
Shampoo
Scrub
Ointment
Salt
biscuits
dt
Strip
Capsule
Pellets
bags
tr
boxes
combikit
jelly
kit
bars
pellets
granules
Emulsion
diskettes
gums
Candy
powders
Gargle
servings
Flakes
Tubes
xl
gels
Jelly
Microgranules
Type
0
1000
2000
3000
4000
5000
6000
7000
mrp

- The most expensive product sold by 1mg is flakes.

In [116]:

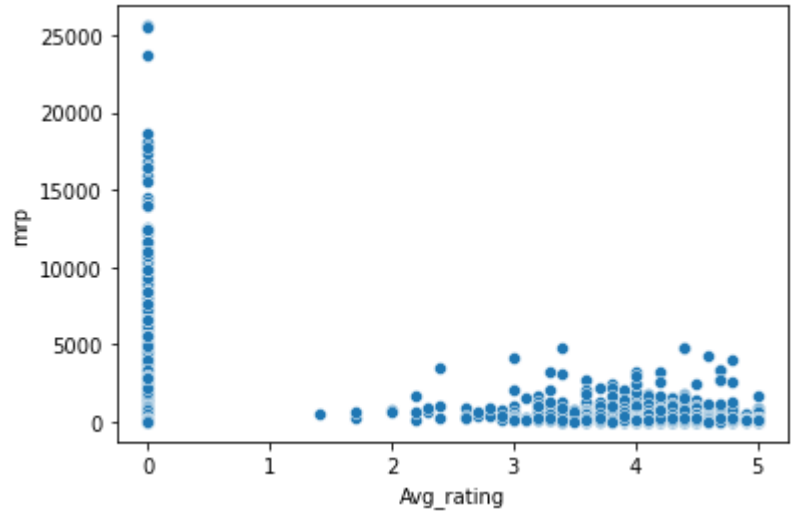
```
plt.figure(figsize = (16,6))
sns.barplot(x = "Avg_rating" , y = "ratings" , data =df )
plt.show()
```



- Avg_rating is directly related to the user rating .

In [123]:

```
sns.scatterplot(x ="Avg_rating" , y ="mrp" , data =df )
plt.show()
```



- Expensive products has very low rating.

Tata 1mg Dataset EDA Report

Introduction: Tata 1mg, previously known as 1mg, is a healthcare platform based in Gurgaon, India. Founded in April 2015 by Prashant Tandon, it offers a range of healthcare services such as e-pharmacy, diagnostics, e-consultation, and health content. This report presents an exploratory data analysis (EDA) of the Tata 1mg dataset, which contains information about medicines available on the platform.

Dataset Overview: The dataset consists of 16,944 rows and 6 columns, providing valuable insights into the medicines listed on Tata 1mg. The columns in the dataset are as follows:

1. Column 0: Index of medicines at a particular page.
2. Name: Name of the medicine.
3. pack_size: Description of the packaging of the medicine.
4. rating_count: Total number of reviews received for the medicine.
5. rating: Rating of the medicine.
6. price: Sell price of the medicine.
7. mrp: Maximum retail price of the medicine.

Data Analysis: The EDA of the Tata 1mg dataset involved several key observations and analyses:

1. Distribution of Medicines: The dataset provided insights into the number of medicines available on the platform, with 16,944 unique entries.
2. Packaging Description: The 'pack_size' column allowed us to understand the various packaging sizes available for different medicines. This information can be useful for inventory management and customer preferences.
3. Rating and Reviews: The 'rating' and 'rating_count' columns provided an understanding of how users rated different medicines and the total number of reviews received. This data can be utilized to assess the popularity and effectiveness of medicines.
4. Price Analysis: I believe their were some data error in the 'price' and 'mrp' columns. Which could have been resolved by the data provider.

Conclusion: The EDA of the Tata 1mg dataset revealed important insights into the medicines available on the platform. The dataset allowed for an analysis of packaging descriptions, ratings, reviews, and pricing information. This analysis can aid in understanding customer preferences, popularity, and pricing strategies. Further analysis, such as correlation studies or segmentation based on medicine categories, could provide deeper insights for business decision-making in areas such as inventory management, marketing, and pricing strategies. Overall, the Tata 1mg dataset offers valuable information for understanding the medicine landscape on the platform and has the potential to drive data-driven insights for improved healthcare