lyzing-e-commerce-trends-flipkart

February 22, 2024

1 Analyzing E-Commerce Trends (Flipkart)

1.0.1 Data Importing and Understanding

[14]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

[3]:
```
df=pd.read_csv('flipkart_com-ecommerce_sample.csv')
```

[5]:
```
df.head(2)
```

[5]:
```
                          uniq_id            crawl_timestamp  \
0  c2d766ca982eca8304150849735ffef9  2016-03-25 22:59:23 +0000
1  7f7036a6d550aaa89d34c77bd39a5e48  2016-03-25 22:59:23 +0000

                                        product_url  \
0  http://www.flipkart.com/alisha-solid-women-s-c…
1  http://www.flipkart.com/fabhomedecor-fabric-do…

                        product_name  \
0  Alisha Solid Women's Cycling Shorts
1  FabHomeDecor Fabric Double Sofa Bed

                              product_category_tree               pid  \
0  ["Clothing >> Women's Clothing >> Lingerie, Sl…  SRTEH2FF9KEDEFGF
1  ["Furniture >> Living Room Furniture >> Sofa B…  SBEEH3QGU7MFYJFY

   retail_price  discounted_price  \
0         999.0             379.0
1       32157.0           22646.0

                                              image is_FK_Advantage_product  \
0  ["http://img5a.flixcart.com/image/short/u/4/a/…                    False
1  ["http://img6a.flixcart.com/image/sofa-bed/j/f…                    False

                                        description       product_rating  \
```

```
0  Key Features of Alisha Solid Women's Cycling S…  No rating available
1  FabHomeDecor Fabric Double Sofa Bed (Finish Co…  No rating available

        overall_rating         brand  \
0  No rating available        Alisha
1  No rating available  FabHomeDecor

                              product_specifications
0  {"product_specification"=>[{"key"=>"Number of …
1  {"product_specification"=>[{"key"=>"Installati…
```

1.0.2 How Many Rows and Columns

[12]:
```
df.shape
```

[12]:
```
(20002, 15)
```

[26]:
```
df.columns
```

[26]:
```
Index(['uniq_id', 'crawl_timestamp', 'product_url', 'product_name',
       'product_category_tree', 'pid', 'retail_price', 'discounted_price',
       'image', 'is_FK_Advantage_product', 'description', 'product_rating',
       'overall_rating', 'brand', 'product_specifications'],
      dtype='object')
```

[13]:
```
df.describe()
```

[13]:
```
        retail_price  discounted_price
count   19922.000000      19922.000000
mean     2979.206104       1973.401767
std      9009.639341       7333.586040
min        35.000000         35.000000
25%       666.000000        350.000000
50%      1040.000000        550.000000
75%      1999.000000        999.000000
max    571230.000000     571230.000000
```

[10]:
```
df.isnull().sum()
```

[10]:
```
uniq_id                    2
crawl_timestamp            2
product_url                2
product_name               2
product_category_tree      2
pid                        2
retail_price              80
discounted_price          80
```

```
image                          5
is_FK_Advantage_product        2
description                    4
product_rating                 2
overall_rating                 2
brand                       5866
product_specifications        16
dtype: int64
```

[53]:
```
df.dtypes
```

[53]:
```
uniq_id                              object
crawl_timestamp           datetime64[ns, UTC]
product_url                          object
product_name                         object
product_category_tree                object
pid                                  object
retail_price                        float64
discounted_price                    float64
image                                object
is_FK_Advantage_product              object
description                          object
product_rating                       object
overall_rating                       object
brand                                object
product_specifications               object
dtype: object
```

2 Exploratory Analysis and Visualization

2.0.1 uniq_id

[33]:
```
df['uniq_id'].isnull().sum()
```

[33]:
```
2
```

[40]:
```
df['uniq_id'].nunique()  # there is 20000 unique value present in this column
```

[40]:
```
20000
```

2.0.2 crawl_timestamp

[54]:
```
df['crawl_timestamp'].isnull().sum()
```

[54]:
```
2
```

[55]:
```
df['crawl_timestamp'].nunique()
```

[55]:
```
371
```

[56]:
```
df['crawl_timestamp'].head()
```

[56]:
```
0   2016-03-25 22:59:23+00:00
1   2016-03-25 22:59:23+00:00
2   2016-03-25 22:59:23+00:00
3   2016-03-25 22:59:23+00:00
4   2016-03-25 22:59:23+00:00
Name: crawl_timestamp, dtype: datetime64[ns, UTC]
```

[175]:
```
df['crawl_timestamp']=pd.to_datetime(df['crawl_timestamp'])
```

[177]:
```
df['crawl_timestamp'].head(4)
```

[177]:
```
0   2016-03-25 22:59:23+00:00
1   2016-03-25 22:59:23+00:00
2   2016-03-25 22:59:23+00:00
3   2016-03-25 22:59:23+00:00
Name: crawl_timestamp, dtype: datetime64[ns, UTC]
```

[178]:
```
## getting date time in different column

df['date']=df['crawl_timestamp'].dt.date
df['time']=df['crawl_timestamp'].dt.time
```

[179]:
```
df.head(1)
```

[179]:
```
                            uniq_id           crawl_timestamp  \
0  c2d766ca982eca8304150849735ffef9 2016-03-25 22:59:23+00:00

                                        product_url  \
0  http://www.flipkart.com/alisha-solid-women-s-c…

                          product_name  \
0  Alisha Solid Women's Cycling Shorts

                               product_category_tree               pid  \
0  ["Clothing >> Women's Clothing >> Lingerie, Sl…  SRTEH2FF9KEDEFGF

   retail_price  discounted_price  \
0         999.0             379.0

                                               image is_FK_Advantage_product  \
0  ["http://img5a.flixcart.com/image/short/u/4/a/…                     False
```

```
                                              description  product_rating  \
0  Key Features of Alisha Solid Women's Cycling S…             NaN

        overall_rating   brand  \
0  No rating available  Alisha

                                product_specifications Product_Rating  \
0  {"product_specification"=>[{"key"=>"Number of …      No rating

   discount_perct        date      time
0       62.062062  2016-03-25  22:59:23
```

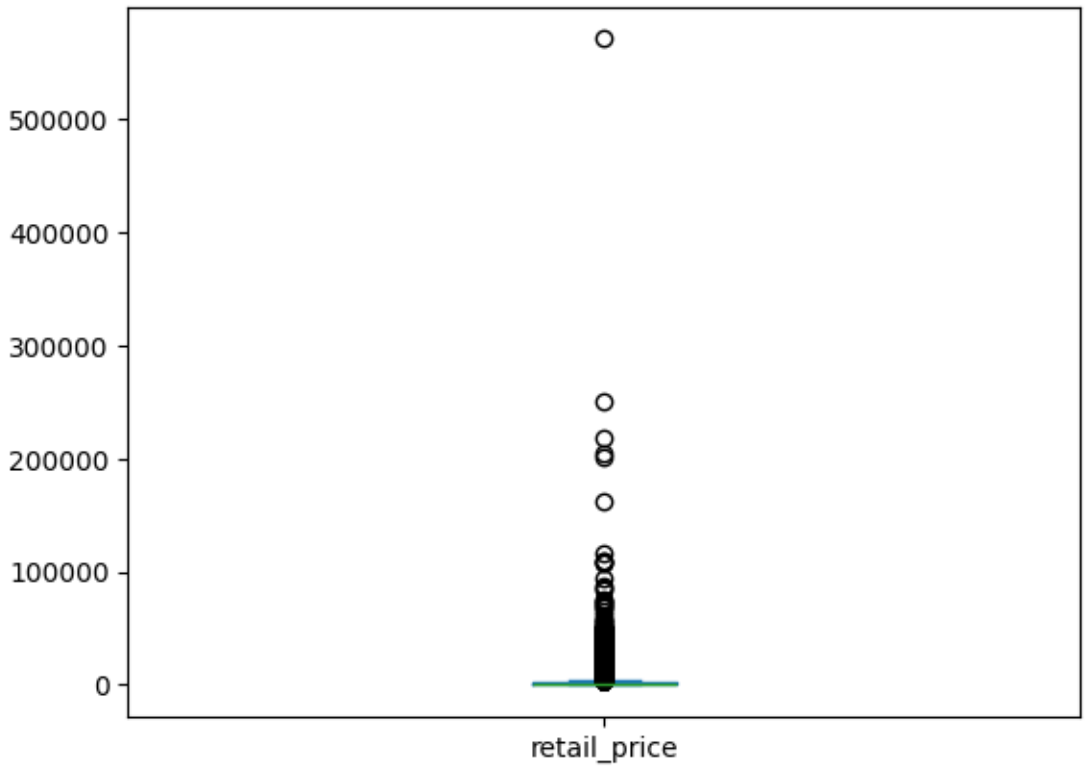
2.0.3 retail_price

[65]:
```
df['retail_price'].isnull().sum()
```

[65]: 80

[68]:
```
df['retail_price'].plot(kind='box')
```

[68]: <Axes: >



[66]:
```
df['retail_price'].mean()
```

[66]:
```
2979.2061038048387
```

[69]:
```
df['retail_price'].median()
```

[69]:
```
1040.0
```

[70]:
```
df['retail_price'].fillna(df['retail_price'].median(),inplace= True)
```

null values filled with median value that is 1040

[71]:
```
df['retail_price'].isnull().sum()
```

[71]:
```
0
```

[78]:
```
df[df['uniq_id'].isnull()]
```

[78]:
```
       uniq_id crawl_timestamp product_url product_name product_category_tree  \
20000      NaN             NaT         NaN          NaN                    NaN
20001      NaN             NaT         NaN          NaN                    NaN

       pid  retail_price  discounted_price image is_FK_Advantage_product  \
20000  NaN        1040.0               NaN   NaN                      NaN
20001  NaN        1040.0               NaN   NaN                      NaN

       description product_rating overall_rating brand product_specifications
20000          NaN            NaN            NaN   NaN                    NaN
20001          NaN            NaN            NaN   NaN                    NaN
```

deleted 2 rows where null value was present

[79]:
```
df.dropna(subset=['uniq_id'],inplace = True)
```

2.0.4 discounted_price

[83]:
```
df['discounted_price'].head()
```

[83]:
```
0      379.0
1    22646.0
2      499.0
3      267.0
4      210.0
Name: discounted_price, dtype: float64
```

[85]:
```
df['discounted_price'].dtype
```

[85]:
```
dtype('float64')
```

[86]:
```
df['discounted_price'].isnull().sum()
```

[86]:
```
78
```

[87]:
```
df['discounted_price'].fillna(df['discounted_price'].median(),inplace = True)
```

filled median value placed of NA

[89]:
```
df['discounted_price'].isnull().sum()
```

[89]:
```
0
```

2.0.5 is_FK_Advantage_product

[92]:
```
df['is_FK_Advantage_product'].head()
```

[92]:
```
0    False
1    False
2    False
3    False
4    False
Name: is_FK_Advantage_product, dtype: object
```

[93]:
```
df['is_FK_Advantage_product'].tail()
```

[93]:
```
19995    False
19996    False
19997    False
19998    False
19999    False
Name: is_FK_Advantage_product, dtype: object
```

[99]:
```
df['is_FK_Advantage_product'].value_counts()
```

[99]:
```
is_FK_Advantage_product
False    19215
True       785
Name: count, dtype: int64
```

2.0.6 description

[103]:
```
df['description'].sample(5)
```

[103]:
```
8790     Minerva Naturals Plant Container Set (Pack of …
4088     UpTown Fabric Necklace - Buy UpTown Fabric Nec…
11924    Femella Women's Printed Casual Shirt - Buy Cor…
1993     S4S Stylish Women's Push-up Bra\n              …
```

```
7455     Flipkart.com: Buy Warner Brothers Scooby Doo C…
Name: description, dtype: object
```

[119]:
```
df.dropna(subset=['description'], inplace=True)
```

[121]:
```
df[df['description'].isnull()]
```

[121]:
```
Empty DataFrame
Columns: [uniq_id, crawl_timestamp, product_url, product_name,
product_category_tree, pid, retail_price, discounted_price, image,
is_FK_Advantage_product, description, product_rating, overall_rating, brand,
product_specifications]
Index: []
```

2.0.7 product_rating

[123]:
```
df['product_rating'].head()
```

[123]:
```
0    No rating available
1    No rating available
2    No rating available
3    No rating available
4    No rating available
Name: product_rating, dtype: object
```

[125]:
```
df['product_rating'].sample(5)
```

[125]:
```
11265    No rating available
13973                      4
15758    No rating available
7382     No rating available
14569    No rating available
Name: product_rating, dtype: object
```

[137]:
```
df['product_rating'].value_counts()
```

[137]:
```
product_rating
No rating available    18149
5                        620
4                        246
1                        171
3                        168
2                         80
4.5                       67
3.7                       51
4.2                       47
3.5                       45
```

```
4.3                         45
3.6                         25
4.1                         24
4.7                         24
2.5                         23
3.8                         23
4.8                         21
3.2                         20
3.3                         17
4.4                         16
3.9                         15
3.4                         13
2.3                         12
2.8                         11
2.7                         10
4.6                          9
2.2                          8
3.1                          7
2.4                          5
2.9                          5
1.5                          4
4.9                          4
1.3                          4
1.7                          4
2.6                          3
1.8                          2
Name: count, dtype: int64
```

[138]:
```
df['product_rating'].isnull().sum()
```

[138]:
```
0
```

[140]:
```
# Convert 'product_rating' column to numeric, coercing errors to NaN for␣
 ↪non-numeric values
df['product_rating'] = pd.to_numeric(df['product_rating'], errors='coerce')

# Define the function for assigning ratings
def assign_rating(product_rating):
    if product_rating >= 4:
        return 'Good rating'
    elif product_rating >= 2:
        return 'Average rating'
    elif product_rating < 2:
        return 'Poor rating'
    else:
        return 'No rating'
```

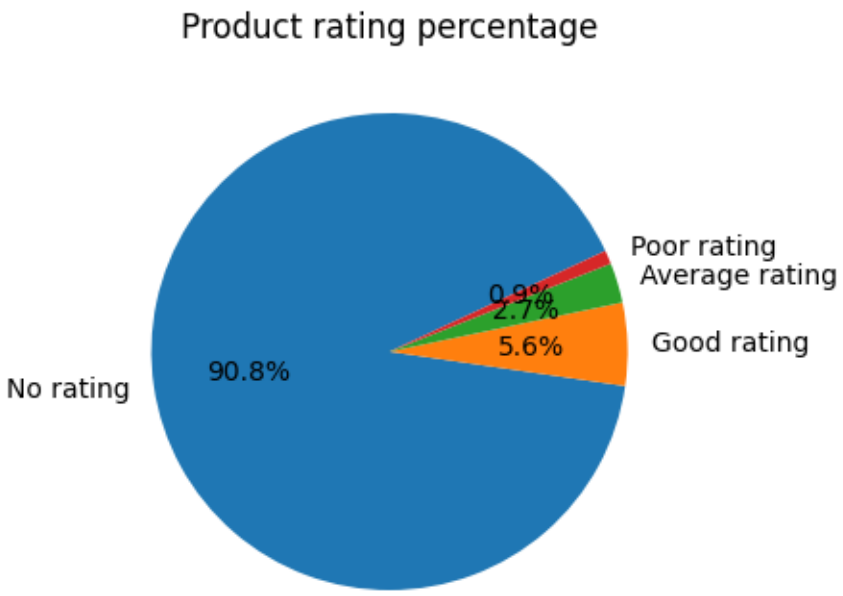
```
# Create a new column 'Product_Rating' based on the 'product_rating' column
df['Product_Rating'] = df['product_rating'].apply(assign_rating)
```

[142]:
```
df.head(2)
```

[142]:
```
                            uniq_id           crawl_timestamp  \
0  c2d766ca982eca8304150849735ffef9 2016-03-25 22:59:23+00:00
1  7f7036a6d550aaa89d34c77bd39a5e48 2016-03-25 22:59:23+00:00

                                        product_url  \
0  http://www.flipkart.com/alisha-solid-women-s-c…
1  http://www.flipkart.com/fabhomedecor-fabric-do…

                         product_name  \
0  Alisha Solid Women's Cycling Shorts
1  FabHomeDecor Fabric Double Sofa Bed

                               product_category_tree               pid  \
0  ["Clothing >> Women's Clothing >> Lingerie, Sl…  SRTEH2FF9KEDEFGF
1  ["Furniture >> Living Room Furniture >> Sofa B…  SBEEH3QGU7MFYJFY

   retail_price  discounted_price  \
0         999.0             379.0
1       32157.0           22646.0

                                              image is_FK_Advantage_product  \
0  ["http://img5a.flixcart.com/image/short/u/4/a/…                    False
1  ["http://img6a.flixcart.com/image/sofa-bed/j/f…                    False

                                       description  product_rating  \
0  Key Features of Alisha Solid Women's Cycling S…             NaN
1  FabHomeDecor Fabric Double Sofa Bed (Finish Co…             NaN

        overall_rating         brand  \
0  No rating available        Alisha
1  No rating available  FabHomeDecor

                            product_specifications Product_Rating
0  {"product_specification"=>[{"key"=>"Number of …      No rating
1  {"product_specification"=>[{"key"=>"Installati…      No rating
```

[165]:
```
plt.figure(figsize=(4,6))
df['Product_Rating'].value_counts().plot(kind='pie',autopct='%1.
 ↪1f%%',startangle=25)
plt.ylabel('')
plt.title('Product rating percentage')
```

[165]: Text(0.5, 1.0, 'Product rating percentage')



2.0.8 overall_rating

[127]:
```
df['overall_rating'].head()
```

[127]:
```
0    No rating available
1    No rating available
2    No rating available
3    No rating available
4    No rating available
Name: overall_rating, dtype: object
```

[166]:
```
df['overall_rating'].value_counts().sample(4)
```

[166]:
```
overall_rating
3.1     7
1.5     4
3.7    51
4.8    21
Name: count, dtype: int64
```

product_rating overall_rating both columns contains same values

2.0.9 brand

[145]:
```
df['brand'].sample(10)
```

[145]:
```
5822         Hotpiper
9392          Legmark
1276         Speedwav
9903              NaN
4850            Amzer
16254             NaN
15185           Pilot
1359            Zevrr
8836              NaN
17322    TheLostPuppy
Name: brand, dtype: object
```

[146]:
```
df['brand'].isnull().sum()
```

[146]:
```
5863
```

[202]:
```
df[df['brand'].isnull()].sample(3)
```

[202]:
```
                              uniq_id           crawl_timestamp  \
653    938c35bbfc9aaaf6fbdd2c1ee5d19c13 2016-01-06 18:20:45+00:00
15408  ecf913e77f9deceb90417f55e4ca1b64 2015-12-12 11:46:53+00:00
885    57df6fd62720bdc984a32e616a539821 2016-01-06 18:20:45+00:00

                                          product_url  \
653    http://www.flipkart.com/betty-girl-s-maxi-dres…
15408  http://www.flipkart.com/ne-women-s-leggings/p/…
885    http://www.flipkart.com/faballey-casual-full-s…

                                    product_name  \
653                         Betty Girl's Maxi Dress
15408                           NE Women's Leggings
885    FabAlley Casual Full Sleeve Solid Women's Top

                                 product_category_tree               pid  \
653    ["Clothing >> Kids' Clothing >> Girls Wear >> …  DREEDVHXSAAZSBTZ
15408  ["Clothing >> Women's Clothing >> Western Wear…  LJGE8GG9PKZYPEPQ
885    ["Clothing >> Women's Clothing >> Western Wear…  TOPE7ZHZCACAVAJK

       retail_price  discounted_price  \
653          1995.0            1995.0
15408        1199.0             499.0
885          1100.0            1100.0

                                                  image  \
```

```
653    ["http://img5a.flixcart.com/image/dress/z/y/y/…
15408  ["http://img6a.flixcart.com/image/legging-jegg…
885    ["http://img6a.flixcart.com/image/top/t/5/d/1-…

      is_FK_Advantage_product  \
653                     False
15408                   False
885                     False

                                            description  product_rating  \
653    Betty Girl's Maxi Dress\n                      …             NaN
15408  NE Women's Leggings - Buy Magenta-Maroon-Orang…             NaN
885    FabAlley Casual Full Sleeve Solid Women's Top …             NaN

           overall_rating brand  \
653    No rating available   NaN
15408  No rating available   NaN
885    No rating available   NaN

                                product_specifications Product_Rating  \
653    {"product_specification"=>[{"key"=>"Ideal For"…      No rating
15408  {"product_specification"=>[{"key"=>"Number of …      No rating
885    {"product_specification"=>[{"key"=>"Ideal For"…      No rating

       discount_perct        date      time main_category
653          0.000000  2016-01-06  18:20:45      Clothing
15408       58.381985  2015-12-12  11:46:53      Clothing
885          0.000000  2016-01-06  18:20:45      Clothing
```

[207]:
```
df['brand'].isnull().sum()
```

[207]:
```
5863
```

[209]:
```
# fill Null vlaues with 'not defined'
df['brand'].fillna(value='not defined', inplace=True)
```

[211]:
```
df['brand'].isnull().sum()
```

[211]:
```
0
```

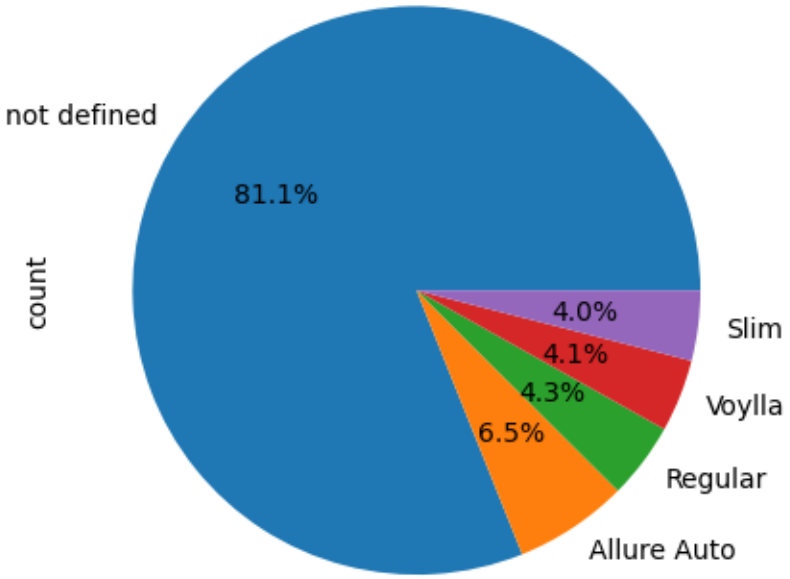
[212]:
```
df['brand'].value_counts().head(5)
```

[212]:
```
brand
not defined    5863
Allure Auto     469
Regular         313
Voylla          299
```

```
Slim            288
Name: count, dtype: int64
```

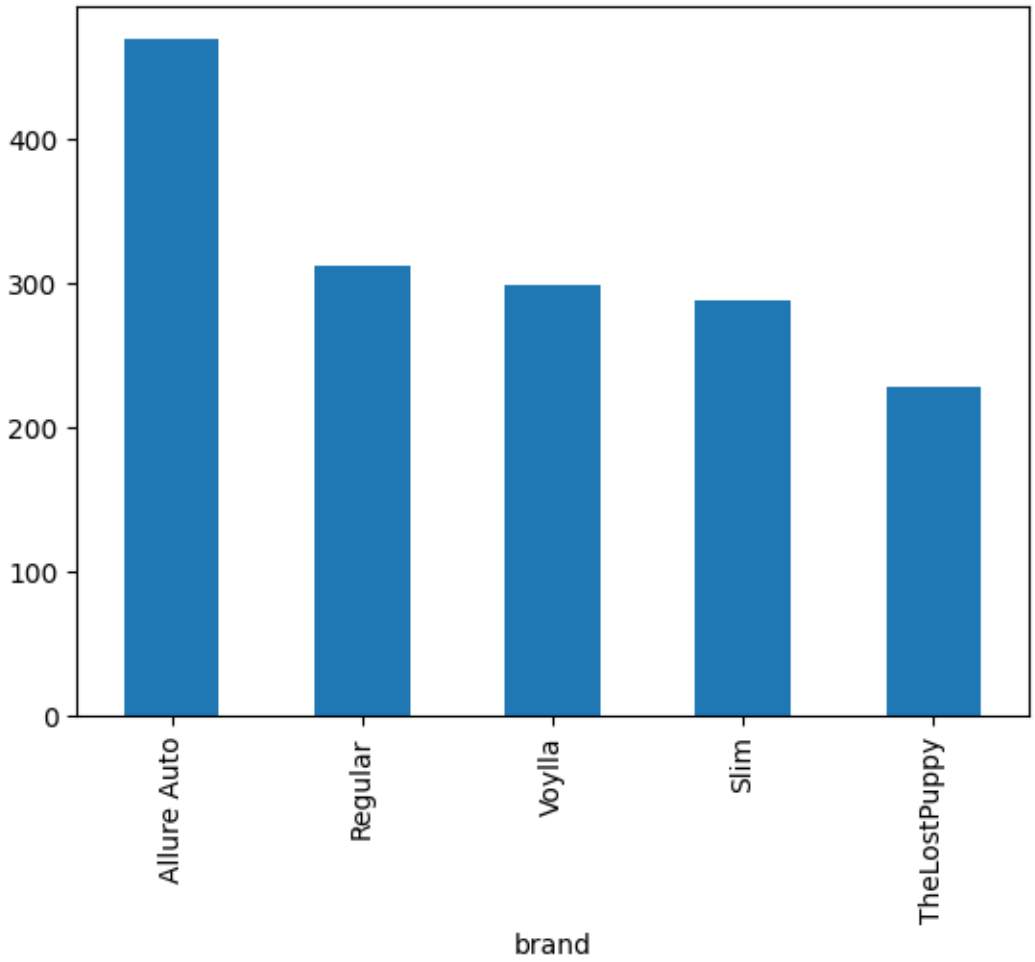
[229]:
```
df['brand'].value_counts().head(5).plot(kind='pie',autopct='%1.1f%%')
```

[229]:
```
<Axes: ylabel='count'>
```



[289]:
```
# not defined value is huge
# if i exclude not define

top_5_brands = df['brand'].value_counts().iloc[1:6]
top_5_brands.plot(kind='bar')
plt.ylabel("")
```

[289]:
```
Text(0, 0.5, '')
```



2.0.10 product_specifications

[168]:
```
df['product_specifications'].sample(5)
```

[168]:
```
11110    {"product_specification"=>[{"key"=>"Brand", "v…
6107     {"product_specification"=>[{"key"=>"Number of …
2784     {"product_specification"=>[{"key"=>"Brand Colo…
16882    {"product_specification"=>[{"key"=>"Brand", "v…
12189    {"product_specification"=>[{"key"=>"Stretchabl…
Name: product_specifications, dtype: object
```

[169]:
```
df['product_specifications'].isnull().sum()
```

[169]:
```
14
```

[122]:
```
df.isnull().sum()
```

[122]:
```
uniq_id                       0
crawl_timestamp               0
product_url                   0
product_name                  0
product_category_tree         0
pid                           0
retail_price                  0
discounted_price              0
image                         3
is_FK_Advantage_product       0
description                   0
product_rating                0
overall_rating                0
brand                      5863
product_specifications       14
dtype: int64
```

2.0.11 Drop columns like image and is_FK_Advantage_Product and product_specification

[ ]:
```
pd.drop(df['image'],df['is_FK_Advantage_product'],df['product_specifications'])
```

2.0.12 Added a column discount_percentage

[172]:
```
df['discount_perct']=((df['retail_price'] - df['discounted_price'])/
↪df['retail_price'])*100
```

[173]:
```
df.head(2)
```

[173]:
```
                             uniq_id           crawl_timestamp  \
0  c2d766ca982eca8304150849735ffef9 2016-03-25 22:59:23+00:00
1  7f7036a6d550aaa89d34c77bd39a5e48 2016-03-25 22:59:23+00:00

                                        product_url  \
0  http://www.flipkart.com/alisha-solid-women-s-c…
1  http://www.flipkart.com/fabhomedecor-fabric-do…

                         product_name  \
0  Alisha Solid Women's Cycling Shorts
1  FabHomeDecor Fabric Double Sofa Bed

                               product_category_tree               pid  \
0  ["Clothing >> Women's Clothing >> Lingerie, Sl…  SRTEH2FF9KEDEFGF
1  ["Furniture >> Living Room Furniture >> Sofa B…  SBEEH3QGU7MFYJFY

   retail_price  discounted_price  \
```

```
0         999.0             379.0
1       32157.0           22646.0

                                                   image is_FK_Advantage_product  \
0  ["http://img5a.flixcart.com/image/short/u/4/a/…                    False
1  ["http://img6a.flixcart.com/image/sofa-bed/j/f…                    False

                                          description  product_rating  \
0  Key Features of Alisha Solid Women's Cycling S…             NaN
1  FabHomeDecor Fabric Double Sofa Bed (Finish Co…             NaN

        overall_rating         brand  \
0  No rating available        Alisha
1  No rating available  FabHomeDecor

                               product_specifications Product_Rating  \
0  {"product_specification"=>[{"key"=>"Number of …      No rating
1  {"product_specification"=>[{"key"=>"Installati…      No rating

   discount_perct
0       62.062062
1       29.576764
```

2.0.13 Added new column main_product

[184]:
```
df['product_category_tree'].value_counts()
```

[184]:
```
product_category_tree
["Jewellery >> Necklaces & Chains >> Necklaces"]
1567
["Jewellery >> Rings"]
710
["Automotive >> Accessories & Spare parts >> Car Interior & Exterior >> Car
Interior >> Car Mats"]
522
["Jewellery >> Bangles, Bracelets & Armlets >> Bangles"]
430
["Tools & Hardware >> Tools >> Gardening Tools >> Plant Containers & Sets >>
Plant Container Sets"]
332
…
["Clothing >> Kids' Clothing >> Infants Wear >> Baby Boys >> Accessories >>
Socks >> Ole Baby Socks >> Ole Baby Baby Boy's, Baby Girl's Woven Ankle Len…"]
1
["Clothing >> Women's Clothing >> Lingerie, Sleep & Swimwear >> Babydolls >>
Heart 2 Heart Babydolls >> Heart 2 Heart Printed Babydoll"]
1
```

```
["Clothing >> Women's Clothing >> Western Wear >> Dresses & Skirts >> Skirts >>
Indigocart Skirts >> Indigocart Printed Women's Regular Skirt"]
1
["Clothing >> Men's Clothing >> Shirts >> Formal Shirts >> Hoffmen Formal Shirts
>> Hoffmen Men's Self Design Formal Shirt"]
1
["Baby Care >> Baby & Kids Gifts >> Stickers >> Fun To See Stickers"]
1
Name: count, Length: 6466, dtype: int64
```

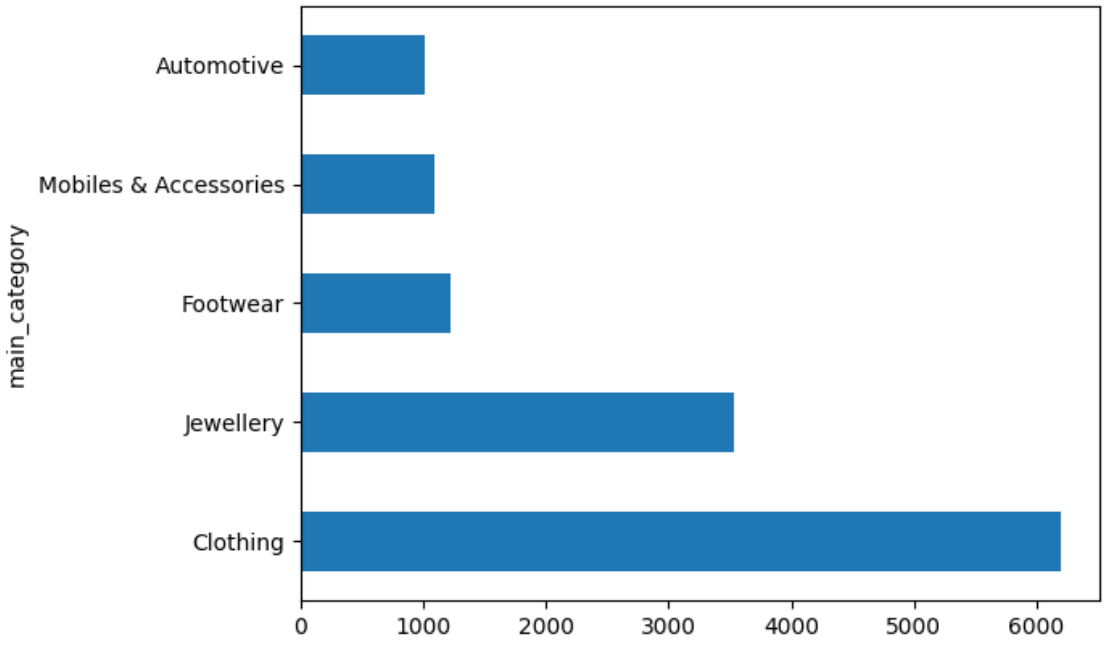
[188]:
```
# need to make main_category

df['main_category'] = df['product_category_tree'].str.split('>>').str[0].str.
 ↪strip().str.replace('["', '')
```

[192]:
```
df['main_category'].value_counts().head()
```

[192]:
```
main_category
Clothing                 6197
Jewellery                3531
Footwear                 1227
Mobiles & Accessories    1099
Automotive               1012
Name: count, dtype: int64
```

[288]:
```
# plot pie chart on it
df['main_category'].value_counts().head().plot(kind='barh')
```

[288]:
```
<Axes: ylabel='main_category'>
```



[201]:
```
df['brand'].value_counts()
```

[201]:
```
brand
Allure Auto              469
Regular                  313
Voylla                   299
Slim                     288
TheLostPuppy             229
                        ...
Hariom enterprises         1
PerfectBlue                1
morpankh Enterprise        1
Lakiya                     1
Fun To See                 1
Name: count, Length: 3499, dtype: int64
```

3 Drop some columns

[232]:
```
df.drop(['crawl_timestamp'],axis=1,inplace = True)
```

[234]:
```
# delete product_url
df.drop(['product_url'],axis=1,inplace = True)
```

[237]:
```
df.drop(['image'],axis=1, inplace=True)
```

[265]:
```
df.drop(['product_rating'],axis=1, inplace= True)
```

3.0.1 Add year column

[253]:
```
# make year column form date
df['date1']= pd.to_datetime(df['date'])
```

[255]:
```
df['year']=df['date1'].dt.year
```

3.0.2 Add discount amount column

[281]:
```
df.head(2)
```

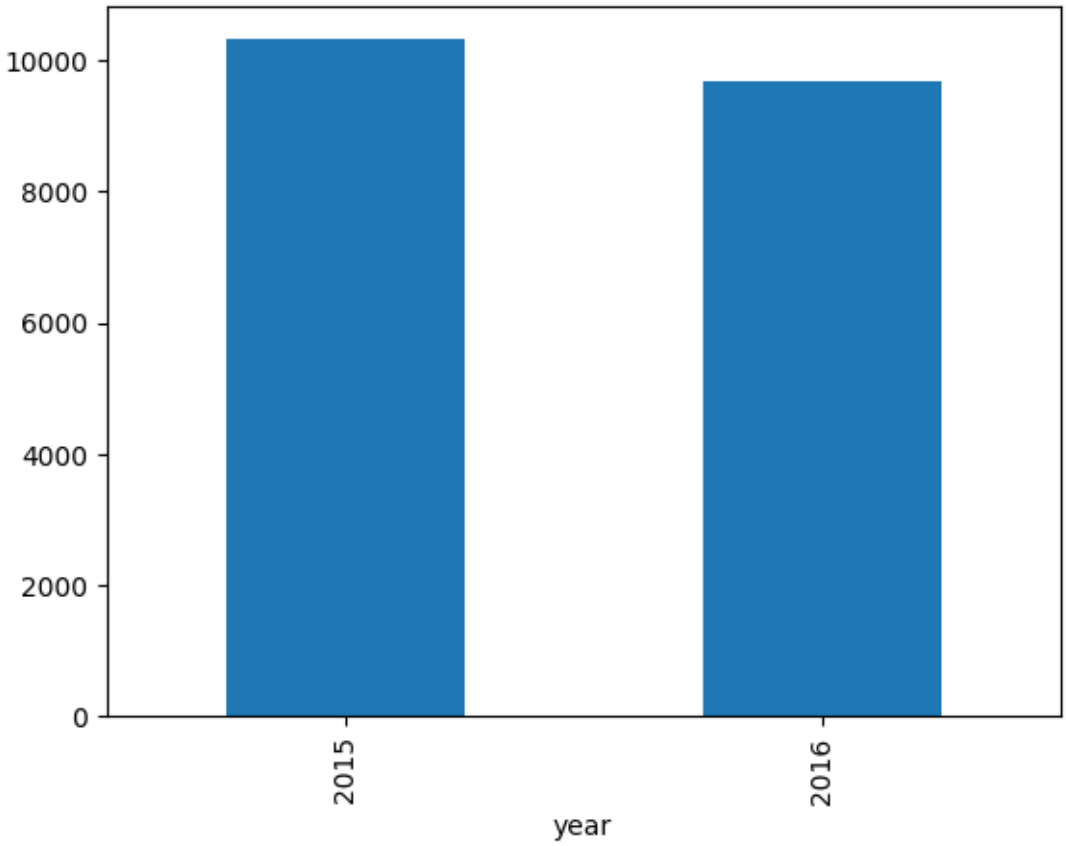
[281]:

	uniq_id	product_name	product_category_tree	pid	retail_price	discounted_price	is_FK_Advantage_product	description	overall_rating	brand	Product_Rating	discount_perct	date	time	main_category	year
0	c2d766ca982eca8304150849735ffef9	Alisha Solid Women's Cycling Shorts	["Clothing >> Women's Clothing >> Lingerie, Sl…	SRTEH2FF9KEDEFGF	999.0	379.0	False	Key Features of Alisha Solid Women's Cycling S…	No rating available	Alisha	No rating	62.062062	2016-03-25	22:59:23	Clothing	2016
1	7f7036a6d550aaa89d34c77bd39a5e48	FabHomeDecor Fabric Double Sofa Bed	["Furniture >> Living Room Furniture >> Sofa B…	SBEEH3QGU7MFYJFY	32157.0	22646.0	False	FabHomeDecor Fabric Double Sofa Bed (Finish Co…	No rating available	FabHomeDecor	No rating	29.576764	2016-03-25	22:59:23	Furniture	2016

4 Bivariate analysis

[258]:
```
df['year'].value_counts().plot(kind='bar')
```

[258]:
```
<Axes: xlabel='year'>
```



[ ]:
```
plt.scatter([df['year'],df['retail_price']])
```

[262]:
```
df.drop(['product_specifications'],axis=1, inplace= True)
```

4.0.1 Top 10 Brnads who gives highest discount

[389]:
```
brand_discount_mean = round(df.groupby('brand')['discount_perct'].mean(),2)
```

[390]:
```
brand_discount_mean.nlargest(10)
```

[390]:
```
brand
Rajcrafts          96.53
Bling              94.55
Instella           91.72
Bond Beatz         91.60
KazamaKraft        90.57
Mydress Mystyle    90.52
CUBA               90.05
SDZ                90.05
```

```
Kaizer Jewelry     89.62
Fash Blush         89.26
Name: discount_perct, dtype: float64
```

[283]:
```
df['main_category'].value_counts()
```

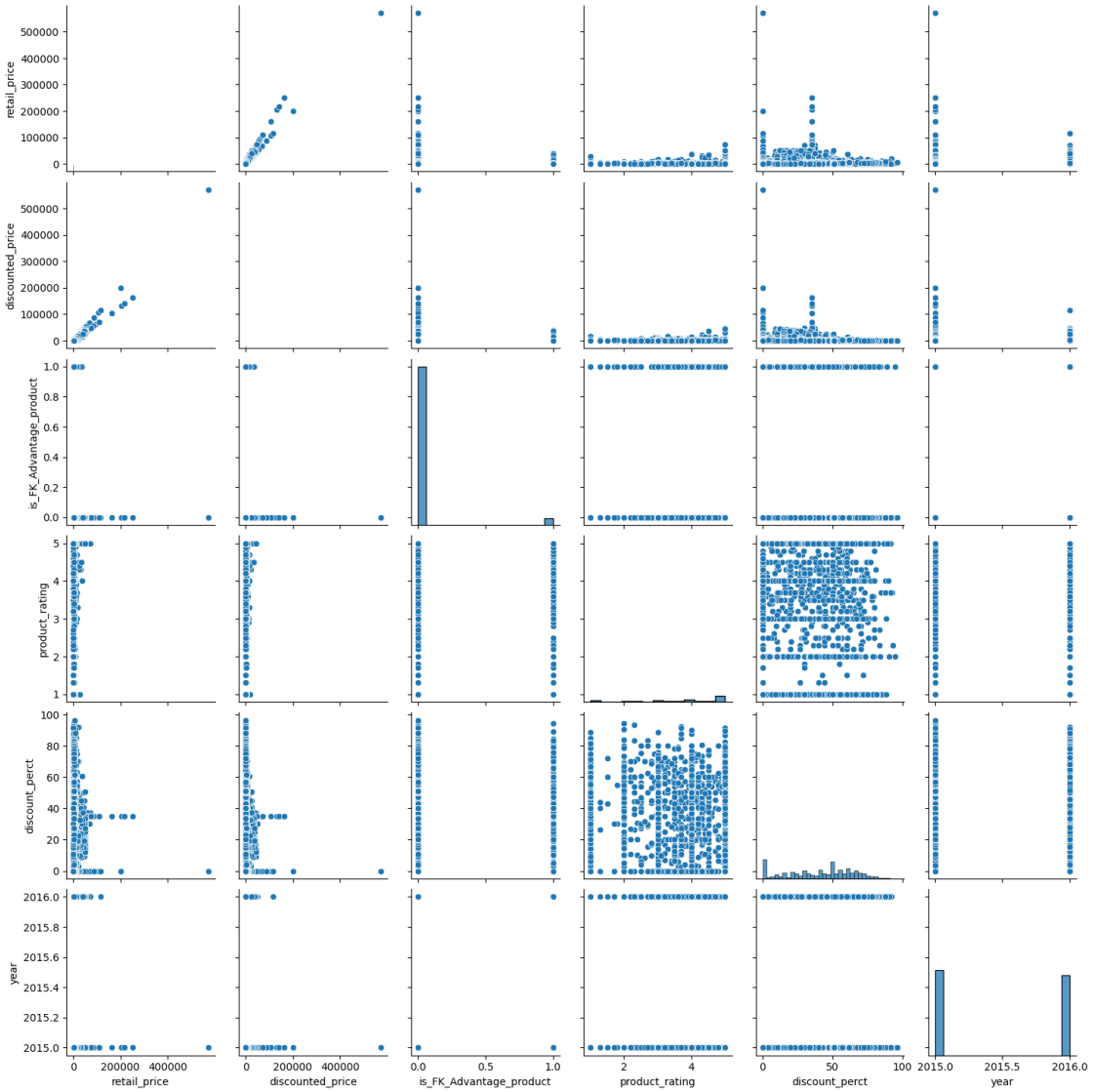
[283]:
```
main_category
Clothing                                                  6197
Jewellery                                                 3531
Footwear                                                  1227
Mobiles & Accessories                                     1099
Automotive                                                1012
                                                          …
Mast & Harbour Gold Synthetic  Clutch"]                      1
Asics Gel-Cumulus 17 Running Shoes"]                         1
Glacier Running Shoes"]                                      1
Starsy Solid Women's Round Neck Green T-Shirt"]              1
Areon Luxurious Fragrance Long Lasting Car,Home,…"]       1
Name: count, Length: 266, dtype: int64
```

[274]:
```
df.columns
```

[274]:
```
Index(['uniq_id', 'product_name', 'product_category_tree', 'pid',
       'retail_price', 'discounted_price', 'is_FK_Advantage_product',
       'description', 'overall_rating', 'brand', 'Product_Rating',
       'discount_perct', 'date', 'time', 'main_category', 'year'],
      dtype='object')
```

[259]:
```
sns.pairplot(df)
```

[259]:
```
<seaborn.axisgrid.PairGrid at 0x1c3f5694980>
```



5 Task and Questions

5.0.1 1. Which are the top 5 product categories based on the number of listings?

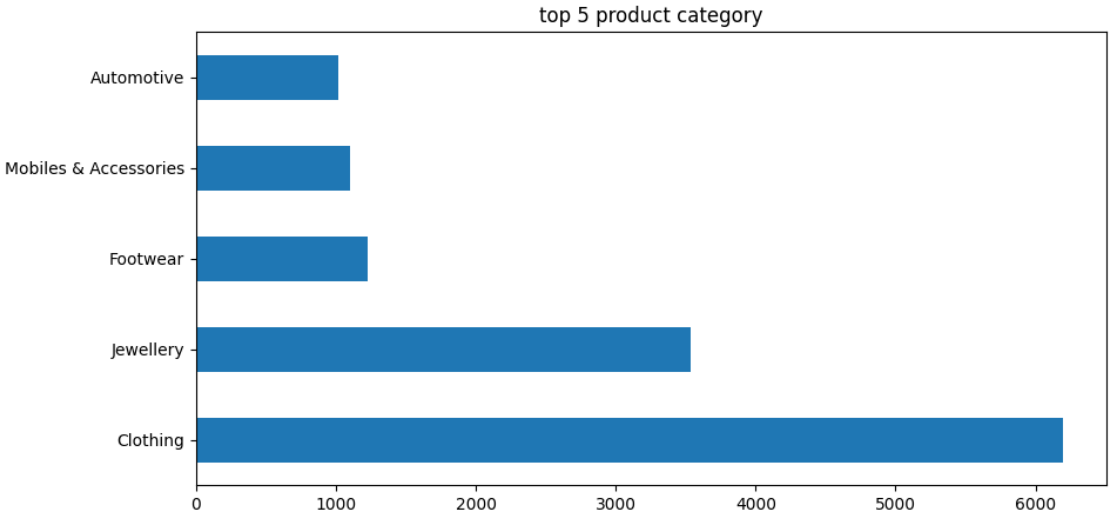
[364]:
```
top_5_product_category = df['main_category'].value_counts().nlargest(5)
top_5_product_category
```

[364]:
```
main_category
Clothing                  6197
Jewellery                 3531
Footwear                  1227
Mobiles & Accessories     1099
```

```
Automotive                1012
Name: count, dtype: int64
```

[385]:
```
plt.figure(figsize=(10,5))
top_5_product_category.plot(kind='barh')
plt.title('top 5 product category')
plt.ylabel("")
plt.show
```

[385]:
```
<function matplotlib.pyplot.show(close=None, block=None)>
```

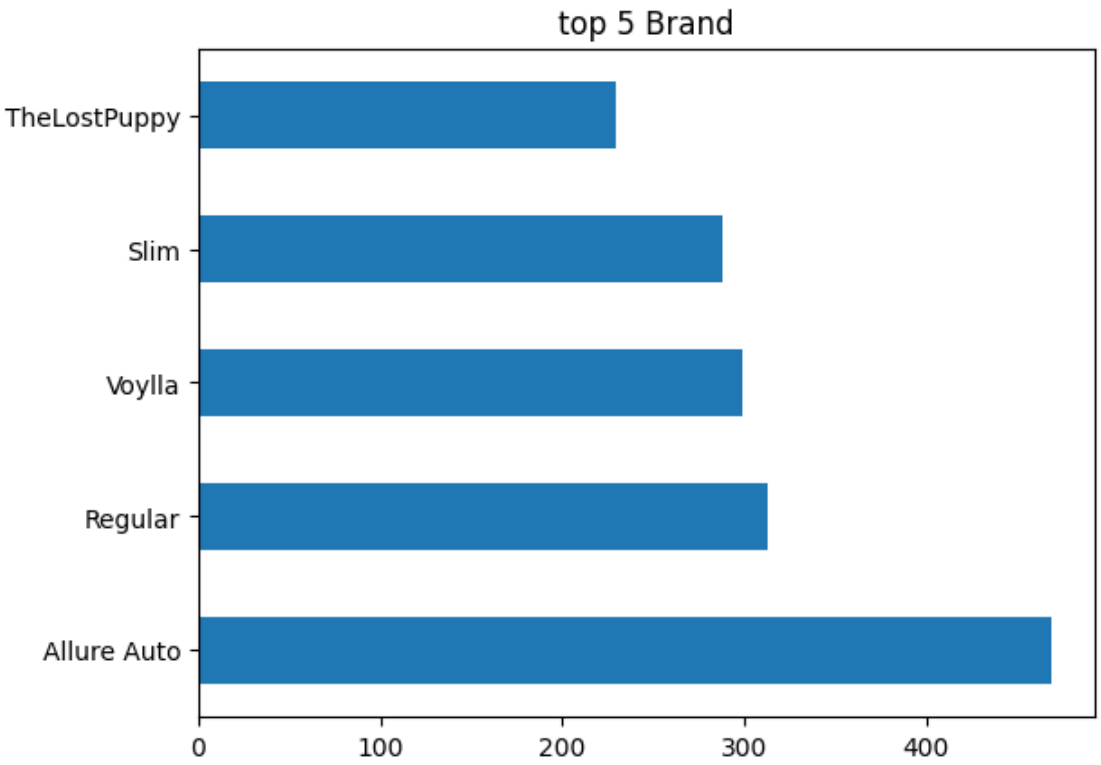


5.0.2 2. Which are the top 5 brands with the most product listings?

[380]:
```
top_5_brands = df['brand'].value_counts().iloc[1:6]
```

[382]:
```
top_5_brands.plot(kind='barh')
plt.title('top 5 Brand')
plt.ylabel("")
plt.show
```

[382]:
```
<function matplotlib.pyplot.show(close=None, block=None)>
```



5.0.3 3. What is the average discount offered by Flipkart across all products?

[295]:
```
#  discount offered by Flipkart across all products
A=round((df['discounted_price']/df['retail_price'])*100,2)
A
```

[295]:
```
0        37.94
1        70.42
2        49.95
3        38.20
4        95.45
         ...
19995    48.67
19996    79.99
19997    76.91
19998    79.99
19999    66.64
Length: 19998, dtype: float64
```

[300]:
```
Avg_Discount_offered_by_flipkart=round(np.mean(A),2)
Avg_Discount_offered_by_flipkart
```

[300]: 59.45

Average discount offered by Flipkart across all products is 59.45%

5.0.4 4. How many products have customer ratings?¶

[305]:
```
df['overall_rating'].value_counts()
```

[305]:
```
overall_rating
No rating available    18149
5                        620
4                        246
1                        171
3                        168
2                         80
4.5                       67
3.7                       51
4.2                       47
3.5                       45
4.3                       45
3.6                       25
4.1                       24
4.7                       24
2.5                       23
3.8                       23
4.8                       21
3.2                       20
3.3                       17
4.4                       16
3.9                       15
3.4                       13
2.3                       12
2.8                       11
2.7                       10
4.6                        9
2.2                        8
3.1                        7
2.4                        5
2.9                        5
1.5                        4
4.9                        4
1.3                        4
1.7                        4
2.6                        3
1.8                        2
Name: count, dtype: int64
```

[309]:
```
count_of_customers_rating = df['overall_rating'].count() - df['overall_rating'].
↪value_counts()['No rating available']
count_of_customers_rating
```

[309]: 1849

1849 products have customer ratings.

5.0.5 5. What percentage of products are part of the Flipkart Advantage program?¶

[320]:
```
round((df['is_FK_Advantage_product'].value_counts()[True] /␣
↪df['is_FK_Advantage_product'].count())*100,2)
```

[320]: 3.93

3.93 Percentage of products are part of the Flipkart Advantage program

5.0.6 6. Which are the top 3 most expensive products listed on Flipkart?

[327]:
```
df.nlargest(3,'retail_price')
```

[327]:
```
                              uniq_id  \
116    329c5f4d7aced63e1ce3e88f41d5e7e6
11631  3a2546675bc399953779e58d84d56650
11507  d9fa5b1d8917b841abaef2a1ce032114

                                          product_name  \
116    Breitling AB011010/BB08 131S Chronomat 44 Anal…
11631                      Durian Leather 2 Seater Sofa
11507              Durian Laze/3 Leather 3 Seater Sofa

                                 product_category_tree               pid  \
116    ["Watches >> Wrist Watches >> Breitling Wrist …  WATE9DX9E4YSRJSY
11631  ["Furniture >> Living Room >> Sofas & Sectiona…  SOFEE7ACBVAGKYUV
11507  ["Furniture >> Living Room >> Sofas & Sectiona…  SOFECYFA8R8T5HNG

       retail_price  discounted_price is_FK_Advantage_product  \
116        571230.0          571230.0                   False
11631      250500.0          162825.0                   False
11507      217500.0          141375.0                   False

                                           description       overall_rating  \
116    Breitling AB011010/BB08 131S Chronomat 44 Anal…  No rating available
11631  Durian Leather 2 Seater Sofa (Finish Color - B…  No rating available
11507  Durian Laze/3 Leather 3 Seater Sofa (Finish Co…  No rating available

             brand Product_Rating  discount_perct        date      time  \
```

```
116    not defined      No rating             0.0  2015-12-04  07:25:36
11631       Durian      No rating            35.0  2015-12-31  09:19:31
11507       Durian      No rating            35.0  2015-12-31  09:19:31

      main_category  year
116         Watches  2015
11631     Furniture  2015
11507     Furniture  2015
```

5.0.7 7. Which brands have the highest average product rating?¶

[340]:
```
brand_rating =df[df['overall_rating'] != 'No rating available']
```

[342]:
```
brand_rating['overall_rating']=  brand_rating['overall_rating'].astype(float)
```

```
C:\Users\Akshat\AppData\Local\Temp\ipykernel_3852\3284945968.py:1:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  brand_rating['overall_rating']=  brand_rating['overall_rating'].astype(float)
```

[347]:
```
brand_rating.groupby('brand').mean("overall_rating").
↪sort_values(by='overall_rating',ascending=False)
```

[347]:
```
              retail_price  discounted_price  overall_rating  discount_perct  \
brand
zDelhi.com          4499.0            3999.0             5.0       11.113581
Perfect              998.0             598.0             5.0       40.080160
Pu-Good              999.0             378.0             5.0       62.162162
PrixCracker          699.0             299.0             5.0       57.224607
Prime               1900.0             874.0             5.0       53.941667
…                      …                 …               …               …
Grey Melange        1199.0             259.0             1.0       78.398666
PURPLE               349.0             249.0             1.0       28.653295
Grey                1199.0             255.0             1.0       78.732277
Mz Nova              699.0             198.0             1.0       71.673820
Peora               2700.0            1080.0             1.0       60.000000

                year
brand
zDelhi.com    2015.0
Perfect       2016.0
Pu-Good       2016.0
PrixCracker   2015.0
```

```
Prime          2016.0
…               …
Grey Melange   2016.0
PURPLE         2016.0
Grey           2016.0
Mz Nova        2016.0
Peora          2015.0

[398 rows x 5 columns]
```

5.0.8 8. Which product category has the highest average retail price?

[349]:
```
df.groupby('main_category').mean('retail_price').
  ↪sort_values(by='retail_price',ascending=False)
```

[349]:
```
                                                  retail_price  \
main_category
Furniture                                         23262.966667
Automation & Robotics                             19999.000000
Rasav Jewels Yellow Gold Diamond 18 K Ring"]      13903.000000
Asics Gel-Kayano 22 Running Shoes"]               12499.000000
BALAJI EXPORTS Bottled Wine Cooler (9 Bottles)"]  10000.000000
…                                                        …
Oddy RS 1.5 X 2 100 Sheets Self Stick Repositio…  280.000000
JUSF2 Black Color Hair Band (Multicolor)"]          249.000000
Naaz 2 in 1 Paper Quilling Board Game"]             230.000000
Siemens 5SL Betagard 5SL MCB (1)"]                  197.000000
SUPERMOD Men's Brief"]                              139.000000

                                                  discounted_price  \
main_category
Furniture                                            17114.472222
Automation & Robotics                                17000.000000
Rasav Jewels Yellow Gold Diamond 18 K Ring"]          9733.000000
Asics Gel-Kayano 22 Running Shoes"]                  12499.000000
BALAJI EXPORTS Bottled Wine Cooler (9 Bottles)"]      8000.000000
…                                                           …
Oddy RS 1.5 X 2 100 Sheets Self Stick Repositio…     260.000000
JUSF2 Black Color Hair Band (Multicolor)"]             189.000000
Naaz 2 in 1 Paper Quilling Board Game"]                199.000000
Siemens 5SL Betagard 5SL MCB (1)"]                     120.000000
SUPERMOD Men's Brief"]                                 139.000000

                                                  discount_perct     year
main_category
Furniture                                            29.721585  2015.25
Automation & Robotics                                14.995750  2016.00
```

```
Rasav Jewels Yellow Gold Diamond 18 K Ring"]              29.993527  2016.00
Asics Gel-Kayano 22 Running Shoes"]                       0.000000  2016.00
BALAJI EXPORTS Bottled Wine Cooler (9 Bottles)"]         20.000000  2016.00
…                                                              …       …
Oddy RS 1.5 X 2 100 Sheets Self Stick Repositio…         7.142857  2016.00
JUSF2 Black Color Hair Band (Multicolor)"]               24.096386  2016.00
Naaz 2 in 1 Paper Quilling Board Game"]                  13.478261  2016.00
Siemens 5SL Betagard 5SL MCB (1)"]                       39.086294  2016.00
SUPERMOD Men's Brief"]                                    0.000000  2016.00

[266 rows x 4 columns]
```

5.0.9 9. Which products have the longest and shortest descriptions?¶

[352]:
```
df['description'].str.len().sort_values(ascending=False)
```

[352]:
```
439      5309
457      5300
1481     4692
18752    4494
18267    4467
         …
10809      84
10797      83
10768      79
19241      79
10952      74
Name: description, Length: 19998, dtype: int64
```

[353]:
```
df.iloc[439]
```

[353]:
```
uniq_id                                    cb3f1acf06c356feed26973d51f78bcc
product_name               Asian Gems & Jewels 6.25 Ratti 5 Dhaatu Sterli…
product_category_tree                                  ["Jewellery >> Rings"]
pid                                                        RNGEBPU7PG8W88TV
retail_price                                                         1656.0
discounted_price                                                      828.0
is_FK_Advantage_product                                               False
description                Asian Gems & Jewels 6.25 Ratti 5 Dhaatu Sterli…
overall_rating                                        No rating available
brand                                                 Asian Gems & Jewels
Product_Rating                                                  No rating
discount_perct                                                       50.0
date                                                           2016-01-06
time                                                             18:20:45
main_category                                                   Jewellery
year                                                                2016
```

```
Name: 439, dtype: object
```

This have Logest product description

[354]:
```
df.iloc[10952]
```

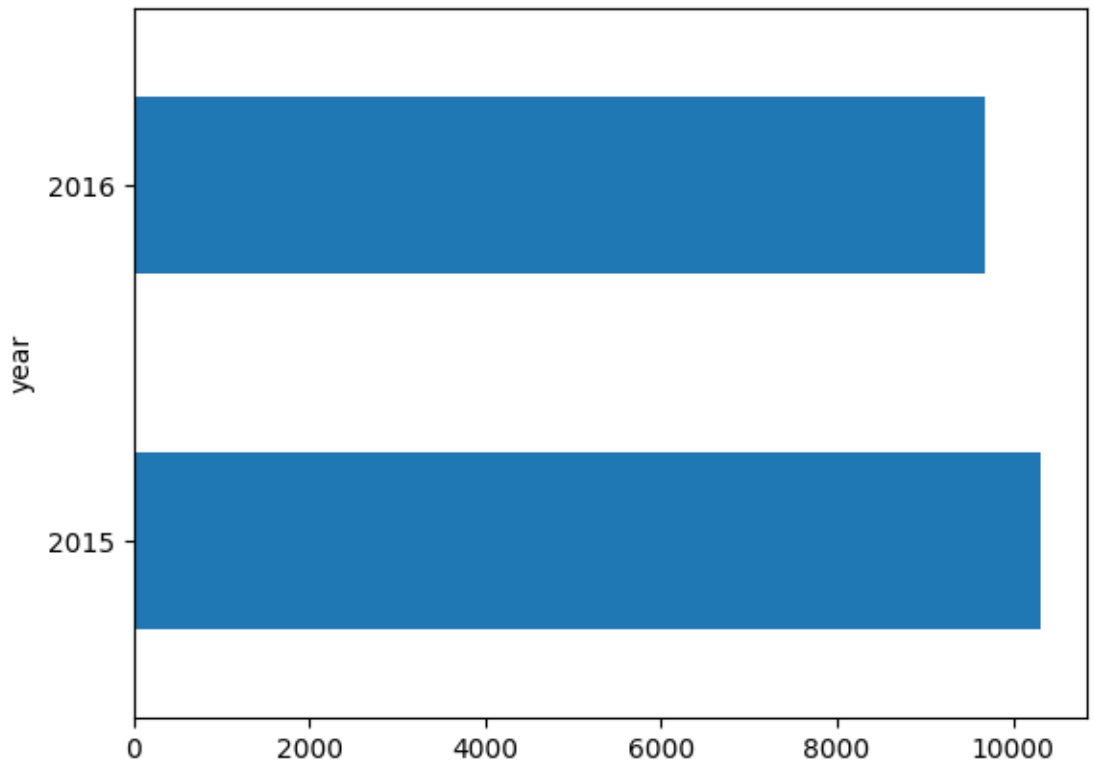
[354]:
```
uniq_id                                    ad012998a8b40e95924fe09e972c33d5
product_name                                          Kyron Women's Jeggings
product_category_tree      ["Clothing >> Women's Clothing >> Western Wear…
pid                                                          LJGED2A4CGQAZTSZ
retail_price                                                            999.0
discounted_price                                                        399.0
is_FK_Advantage_product                                                 False
description                Kyron Women's Jeggings - Buy Blue Kyron Women'…
overall_rating                                        No rating available
brand                                                          not defined
Product_Rating                                                   No rating
discount_perct                                                    60.06006
date                                                           2015-12-31
time                                                             09:19:31
main_category                                                    Clothing
year                                                                 2015
Name: 10953, dtype: object
```

This product have shortest product discription

5.0.10 10. In which year was the data mostly crawled?

[359]:
```
df['year'].value_counts().plot(kind='barh')
```

[359]:
```
<Axes: ylabel='year'>
```



6 Conclusion

6.0.1 Top 5 product category : Clothing,Jewellery, Footwear, Mobile & Accessories, Automotive.

6.0.2 Top 5 brands: Allure Auto, Regular, Voylla, Slim, TheLostPuppy.

6.0.3 Average discount offered by Flipkart across all products is 59.45%.

6.0.4 Out of all only 1849 products have customer ratings.

6.0.5 3.92% of products are part of the Flipkart Advantage program.

6.0.6 The top most expensive products belong to Watches and Furniture Category.

6.0.7 Furniture category has the highest average retail price.

6.0.8 In year 2015 data was mostly crawled.