

Flipkart E-Commerce Data Analysis

Mohammad Areeb · LinkedIn · www.linkedin.com/in/mohammadareeb2544 · github · <https://github.com/areeb399>

Importing Libraries

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
from plotly import graph_objects as go
```

Importing data

```
In [2]: df = pd.read_excel("flipkart.xlsx")

In [3]: df

Out[3]:
```

	uniq_id	crawl_timestamp	product_url	product_name	product_category_tree	pid	retail_price	discounted_price	
0	c2d766ca982eca8304150849735ffe9	2016-03-25 22:59:23 +0000	http://www.flipkart.com/alisha-solid-women-s-c...	Alisha Solid Women's Cycling Shorts	["Clothing >> Women's Clothing >> Lingerie, Sl..."]	SRTEH2FF9KDEDEFGF	999.0	379.0	http://img5a.flixcart.com/images/solo-b
1	7f7036a6d550aaa89d34c77bd39a5e48	2016-03-25 22:59:23 +0000	http://www.flipkart.com/fabhomedecor-fabric-do...	FabHomeDecor Fabric Double Sofa Bed	["Furniture >> Living Room Furniture >> Sofa B..."]	SBEEH3QGUMFYJFY	32157.0	22646.0	http://img6a.flixcart.com/images/solo-b
2	f449ec65dc0c41b6ae56a32717d01b	2016-03-25 22:59:23 +0000	http://www.flipkart.com/aw-bellies/p/meh4dgr...	AW Bellies	["Footwear >> Women's Footwear >> Ballerinas >..."]	SHOEH4GRSUBJGZXE	999.0	499.0	http://img5a.flixcart.com/images/shoe7
3	0973b37acd0c664e3de26e97e5571454	2016-03-25 22:59:23 +0000	http://www.flipkart.com/alisha-solid-women-s-c...	Alisha Solid Women's Cycling Shorts	["Clothing >> Women's Clothing >> Lingerie, Sl..."]	SRTEH2FH6U2MQ6SJ	699.0	267.0	http://img5a.flixcart.com/images/shortf
4	bc940ea42ee6bef5ac7cea3fb5cbee7	2016-03-25 22:59:23 +0000	http://www.flipkart.com/sicons-all-purpose-arm...	Sicons All Purpose Amica Dog Shampoo	["Pet Supplies >> Grooming >> Skin & Coat Care..."]	PSOEH3ZYDMSYARJ5	220.0	210.0	http://img5a.flixcart.com/images/sham
...
19995	7179d2f6c4ad50a17d014ca1d2815156	2015-12-01 10:15:43 +0000	http://www.flipkart.com/walldesign-small-vinyl...	WallDesign Large Vinyl Sticker	["Baby Care >> Baby & Kids Gifts >> Stickers >..."]	STIE7KFJAKSTDY9G	1500.0	730.0	http://img5a.flixcart.com/images/sticker
19996	71ac41918639d37b8f5e3c3ffdfce09	2015-12-01 10:15:43 +0000	http://www.flipkart.com/wallmantra-large-vinyl...	Wallmantra Large Vinyl Stickers Sticker	["Baby Care >> Baby & Kids Gifts >> Stickers >..."]	STIE8Y5URNQJCGH	1429.0	1143.0	http://img6a.flixcart.com/images/sticker
19997	93e9d343837400c9d7980874ece471c	2015-12-01 10:15:43 +0000	http://www.flipkart.com/elite-collection-mediu...	Elite Collection Medium Acrylic Sticker	["Baby Care >> Baby & Kids Gifts >> Stickers >..."]	STIE7VAYDKQEBSO	1299.0	999.0	http://img5a.flixcart.com/images/sticker
19998	669e79b8fa5d9ae020841c0c97d5e935	2015-12-01 10:15:43 +0000	http://www.flipkart.com/elite-collection-mediu...	Elite Collection Medium Acrylic Sticker	["Baby Care >> Baby & Kids Gifts >> Stickers >..."]	STIE8Y5VEPPC24ZY	1499.0	1199.0	http://img5a.flixcart.com/images/sticker
19999	cb4f87a8747f715f9f5e7f7b7b3be79c	2015-12-01 10:15:43 +0000	http://www.flipkart.com/elite-collection-mediu...	Elite Collection Medium Acrylic Sticker	["Baby Care >> Baby & Kids Gifts >> Stickers >..."]	STIE8BN9ZDSZGZKY	1499.0	999.0	http://img6a.flixcart.com/images/sticker

20000 rows × 10 columns

Summary of the statistics for each numeric column

```
In [4]: df.describe()

Out[4]:
```

	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

Checking null(or missing) values

```
In [5]: df.isnull()

Out[5]:
```

	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
0	False	False	False	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False	False	False	False
...
19995	False	False	False	False	False	False	False	False	False	False	False	False	False	False
19996	False	False	False	False	False	False	False	False	False	False	False	False	False	False
19997	False	False	False	False	False	False	False	False	False	False	False	False	False	False
19998	False	False	False	False	False	False	False	False	False	False	False	False	False	False
19999	False	False	False	False	False	False	False	False	False	False	False	False	False	False

20000 rows × 15 columns

Total number of missing values in each columns

```
In [6]: df.isnull().sum()

Out[6]:
```

	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
uniq_id	0													
crawl_timestamp	0													
product_url	0													
product_name	0													
product_category_tree	0													
pid	0													
retail_price	78													
discounted_price	78													
image	3													
is_FK_Advantage_product	0													
description	2													
product_rating	0													
overall_rating	0													
brand	5864													
product_specifications	14													
dtype: int64														

Top 5 values

```
In [7]: df.head()

Out[7]:
```

	uniq_id	crawl_timestamp	product_url	product_name	product_category_tree	pid	retail_price	discounted_price	
0	c2d766ca982eca8304150849735ffe9	2016-03-25 22:59:23 +0000	http://www.flipkart.com/alisha-solid-women-s-c...	Alisha Solid Women's Cycling Shorts	["Clothing >> Women's Clothing >> Lingerie, Sl..."]	SRTEH2FF9KDEDEFGF	999.0	379.0	http://img5a.flixcart.com/images/solo-b
1	7f7036a6d550aaa89d34c77bd39a5e48	2016-03-25 22:59:23 +0000	http://www.flipkart.com/fabhomedecor-fabric-do...	FabHomeDecor Fabric Double Sofa Bed	["Furniture >> Living Room Furniture >> Sofa B..."]	SBEEH3QGUMFYJFY	32157.0	22646.0	http://img6a.flixcart.com/images/solo-b
2	f449ec65dc0c41b6ae56a32717d01b	2016-03-25 22:59:23 +0000	http://www.flipkart.com/aw-bellies/p/meh4dgr...	AW Bellies	["Footwear >> Women's Footwear >> Ballerinas >..."]	SHOEH4GRSUBJGZXE	999.0	499.0	http://img5a.flixcart.com/images/shoe7
3	0973b37acd0c664e3de26e97e5571454	2016-03-25 22:59:23 +0000	http://www.flipkart.com/alisha-solid-women-s-c...	Alisha Solid Women's Cycling Shorts	["Clothing >> Women's Clothing >> Lingerie, Sl..."]	SRTEH2FH6U2MQ6SJ	699.0	267.0	http://img5a.flixcart.com/images/shortf
4	bc940ea42ee6bef5ac7cea3fb5cbee7	2016-03-25 22:59:23 +0000	http://www.flipkart.com/sicons-all-purpose-arm...	Sicons All Purpose Amica Dog Shampoo	["Pet Supplies >> Grooming >> Skin & Coat Care..."]	PSOEH3ZYDMSYARJ5	220.0	210.0	http://img5a.flixcart.com/images/sham

Bottom 5 values

```
In [8]: df.tail()

Out[8]:
```

	uniq_id	crawl_timestamp	product_url	product_name	product_category_tree	pid	retail_price	discounted_price	
19995	7179d2f6c4ad50a17d014ca1d2815156	2015-12-01 10:15:43 +0000	http://www.flipkart.com/walldesign-small-vinyl...	WallDesign Small Vinyl Sticker	["Baby Care >> Baby & Kids Gifts >> Stickers >..."]	STIE7KFJAKSTDY9G	1500.0	730.0	http://img5a.flixcart.com/images/sticker
19996	71ac41918639d37b8f5e3c3ffdfce09	2015-12-01 10:15:43 +0000	http://www.flipkart.com/wallmantra-large-vinyl...	Wallmantra Large Vinyl Stickers Sticker	["Baby Care >> Baby & Kids Gifts >> Stickers >..."]	STIE8Y5URNQJCGH	1429.0	1143.0	http://img6a.flixcart.com/images/sticker
19997	93e9d343837400c9d7980874ece471c	2015-12-01 10:15:43 +0000	http://www.flipkart.com/elite-collection-mediu...	Elite Collection Medium Acrylic Sticker	["Baby Care >> Baby & Kids Gifts >> Stickers >..."]	STIE7VAYDKQEBSO	1299.0	999.0	http://img5a.flixcart.com/images/sticker
19998	669e79b8fa5d9ae020841c0c97d5e935	2015-12-01 10:15:43 +0000	http://www.flipkart.com/elite-collection-mediu...	Elite Collection Medium Acrylic Sticker	["Baby Care >> Baby & Kids Gifts >> Stickers >..."]	STIE8Y5VEPPC24ZY	1499.0	1199.0	http://img5a.flixcart.com/images/sticker
19999	cb4f87a8747f715f9f5e7f7b7b3be79c	2015-12-01 10:15:43 +0000	http://www.flipkart.com/elite-collection-mediu...	Elite Collection Medium Acrylic Sticker	["Baby Care >> Baby & Kids Gifts >> Stickers >..."]	STIE8BN9ZDSZGZKY	1499.0	999.0	http://img6a.flixcart.com/images/sticker

Filling missing values in a column 'retail_price' with the median value of that column

```
In [9]: df['retail_price'].fillna(df['retail_price'].median(), inplace = True)
```

Filling missing values in a column 'discounted_price' with the median value of that column

```
In [10]: df['discounted_price'].fillna(df['discounted_price'].median(), inplace = True)
```

Calculating and creating a new column named 'discount_percentage'

```
In [11]: x = df['retail_price'] - df['discounted_price']
y = (x / df['retail_price']) * 100
df['discount_percentage'] = y

In [12]: df.head()
```

```
Out[12]:
```

	uniq_id	crawl_timestamp	product_url	product_name	product_category_tree	pid	retail_price	discounted_price	
0	c2d766ca982eca8304150849735ffe9	2016-03-25 22:59:23 +0000	http://www.flipkart.com/alisha-solid-women-s-c...	Alisha Solid Women's Cycling Shorts	["Clothing >> Women's Clothing >> Lingerie, Sl..."]	SRTEH2FF9KDEDEFGF	999.0	379.0	http://img5a.flixcart.com/images/solo-b
1	7f7036a6d550aaa89d34c77bd39a5e48	2016-03-25 22:59:23 +0000	http://www.flipkart.com/fabhomedecor-fabric-do...	FabHomeDecor Fabric Double Sofa Bed	["Furniture >> Living Room Furniture >> Sofa B..."]	SBEEH3QGUMFYJFY	32157.0	22646.0	http://img6a.flixcart.com/images/solo-b
2	f449ec65dc0c41b6ae56a32717d01b	2016-03-25 22:59:23 +0000	http://www.flipkart.com/aw-bellies/p/meh4dgr...	AW Bellies	["Footwear >> Women's Footwear >> Ballerinas >..."]	SHOEH4GRSUBJGZXE	999.0	499.0	http://img5a.flixcart.com/images/shoe7
3	0973b37acd0c664e3de26e97e5571454	2016-03-25 22:59:23 +0000	http://www.flipkart.com/alisha-solid-women-s-c...	Alisha Solid Women's Cycling Shorts	["Clothing >> Women's Clothing >> Lingerie, Sl..."]	SRTEH2FH6U2MQ6SJ	699.0	267.0	http://img5a.flixcart.com/images/shortf
4	bc940ea42ee6bef5ac7cea3fb5cbee7	2016-03-25 22:59:23 +0000	http://www.flipkart.com/sicons-all-purpose-arm...	Sicons All Purpose Amica Dog Shampoo	["Pet Supplies >> Grooming >> Skin & Coat Care..."]	PSOEH3ZYDMSYARJ5	220.0	210.0	http://img5a.flixcart.com/images/sham

```
In [13]: df['discount_percentage']

Out[13]:
```

	discount_percentage
0	62.062962
1	29.576764
2	50.050950
3	61.802575
4	4.545455
...	...
19995	51.333333
19996	20.013996
19997	23.094608
19998	20.013342
19999	33.355570

Name: discount_percentage, Length: 20000, dtype: float64

Converting the 'crawl_timestamp' column into datetime objects and storing them in a new column named 'timestamp'

```
In [14]: df['timestamp'] = pd.to_datetime(df['crawl_timestamp'])
```

Creating a new column named 'Time'. The lambda function lambda x : x.time extracts the time component from each datetime object and assigns it to the 'Time' column

```
In [15]: df['Time'] = df['timestamp'].apply(lambda x : x.time)
```

Creating a new column named 'Date'. The lambda function lambda x : x.date extracts the date component from each datetime object and assigns it to the 'Date' column

```
In [16]: df['Date'] = df['timestamp'].apply(lambda x : x.date)
```

Dropping a column named 'crawl_timestamp'

```
In [17]: df.drop(['crawl_timestamp'], axis = 1, inplace = True)
```

Creating a new column named 'main_category' based on the values in the 'product_category_tree' column

```
In [18]: df['main_category'] = df['product_category_tree'].apply(lambda x : x.split('>>')[0][2:len(x.split('>>')[0])-1])

In [19]: df.head()
```

```
Out[19]:
```

	uniq_id	product_url	product_name	product_category_tree	pid	retail_price	discounted_price	
0	c2d766ca982eca8304150849735ffe9	http://www.flipkart.com/alisha-solid-women-s-c...	Alisha Solid Women's Cycling Shorts	["Clothing >> Women's Clothing >> Lingerie, Sl..."]	SRTEH2FF9KDEDEFGF	999.0	379.0	http://img5a.flixcart.com/images/shortf
1	7f7036a6d550aaa89d34c77bd39a5e48	http://www.flipkart.com/fabhomedecor-fabric-do...	FabHomeDecor Fabric Double Sofa Bed	["Furniture >> Living Room Furniture >> Sofa B..."]	SBEEH3QGUMFYJFY	32157.0	22646.0	http://img6a.flixcart.com/images/solo-b
2	f449ec65dc0c41b6ae56a32717d01b	http://www.flipkart.com/aw-bellies/p/meh4dgr...	AW Bellies	["Footwear >> Women's Footwear >> Ballerinas >..."]	SHOEH4GRSUBJGZXE	999.0	499.0	http://img5a.flixcart.com/images/shoe7
3	0973b37acd0c664e3de26e97e5571454	http://www.flipkart.com/alisha-solid-women-s-c...	Alisha Solid Women's Cycling Shorts	["Clothing >> Women's Clothing >> Lingerie, Sl..."]	SRTEH2FH6U2MQ6SJ	699.0	267.0	http://img5a.flixcart.com/images/shortf
4	bc940ea42ee6bef5ac7cea3fb5cbee7	http://www.flipkart.com/sicons-all-purpose-arm...	Sicons All Purpose Amica Dog Shampoo	["Pet Supplies >> Grooming >> Skin & Coat Care..."]	PSOEH3ZYDMSYARJ5	220.0	210.0	http://img5a.flixcart.com/images/sham

Creating a DataFrame named 'top_products' that contains information about the top 'n' main categories and their corresponding total counts from a DataFrame (df).

```
In [20]: n = 10
top_products=pd.DataFrame(df['main_category'].value_counts()[0:n].reset_index())
top_products.rename(columns = {'index':'Top_Products','main_category':'Total_Count'}, inplace = True)

In [21]: n = 10
top_brands=pd.DataFrame(df['brand'].value_counts()[0:n].reset_index())
top_brands.rename(columns = {'index':'Top_Brands','brand':'Total_Count'}, inplace = True)
```

```
In [22]: df.head()
```

```
Out[22]:
```

	uniq_id	product_url	product_name	product_category_tree
--	---------	-------------	--------------	-----------------------