Flipkart Product Analysis Project using Python

In [40]:
```
import pandas as pd
import numpy as np
import ast
import plotly.express as px
from plotly import graph_objects as go
```

In [41]:
```
ds = pd.read_csv('flipkart_com-ecommerce_sample.csv')
```

In [42]:
```
ds.head()
```

Out[42]:

	uniq_id	crawl_timestamp	product_url	product_name	product_category_tree	
0	c2d766ca982eca8304150849735ffef9	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...	SRTEH2FF9KE
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...	SBEEH3QGU7M
2	f449ec65dcbc041b6ae5e6a32717d01b	2016-03-25 22:59:23 +0000	http://www.flipkart.com/aw-bellies/p/itmeh4grg...	AW Bellies	["Footwear >> Women's Footwear >> Ballerinas >...	SHOEH4GRSUB
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...	SRTEH2F6HUZ
4	bc940ea42ee6bef5ac7cea3fb5cfbee7	2016-03-25 22:59:23 +0000	http://www.flipkart.com/sicons-all-purpose-arn...	Sicons All Purpose Arnica Dog Shampoo	["Pet Supplies >> Grooming >> Skin & Coat Care...	PSOEH3ZYDMS

Searching any Null value

In [43]:
```
ds.isnull().sum()
```

Out[43]:
```
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
```

Data Cleaning by fillna

In [44]:
```
ds['retail_price'].fillna(ds['retail_price'].median(),inplace=True)
ds['discounted_price'].fillna(ds['discounted_price'].median(),inplace=True)
```

In [45]:
```
ds.head()
```

Out[45]:

	uniq_id	crawl_timestamp	product_url	product_name	product_category_tree	
0	c2d766ca982eca8304150849735ffef9	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...	SRTEH2FF9KE
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...	SBEEH3QGU7M
2	f449ec65dcbc041b6ae5e6a32717d01b	2016-03-25 22:59:23 +0000	http://www.flipkart.com/aw-bellies/p/itmeh4grg...	AW Bellies	["Footwear >> Women's Footwear >> Ballerinas >...	SHOEH4GRSUB
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...	SRTEH2F6HUZ
4	bc940ea42ee6bef5ac7cea3fb5cfbee7	2016-03-25 22:59:23 +0000	http://www.flipkart.com/sicons-all-purpose-arn...	Sicons All Purpose Arnica Dog Shampoo	["Pet Supplies >> Grooming >> Skin & Coat Care...	PSOEH3ZYDMS

Create a dicount_percentage Coloumn

In [46]:
```
x=ds['retail_price']-ds['discounted_price']
y= (x/ds['retail_price'])*100
ds['discount_percentage']=y
```

In [47]:
```
ds.head()
```

Out[47]:

	uniq_id	crawl_timestamp	product_url	product_name	product_category_tree	
0	c2d766ca982eca8304150849735ffef9	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...	SRTEH2FF9KE
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...	SBEEH3QGU7M
2	f449ec65dcbc041b6ae5e6a32717d01b	2016-03-25 22:59:23 +0000	http://www.flipkart.com/aw-bellies/p/itmeh4grg...	AW Bellies	["Footwear >> Women's Footwear >> Ballerinas >...	SHOEH4GRSUB
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...	SRTEH2F6HUZ
4	bc940ea42ee6bef5ac7cea3fb5cfbee7	2016-03-25 22:59:23 +0000	http://www.flipkart.com/sicons-all-purpose-arn...	Sicons All Purpose Arnica Dog Shampoo	["Pet Supplies >> Grooming >> Skin & Coat Care...	PSOEH3ZYDMS

Create Main Category column

In [48]:
```
ds['main_category']= ds['product_category_tree'].apply(lambda x:x.split('>>')[0][2:len(x.split('>>')[0])])
```

In [49]:
```
ds.head()
```

Out[49]:

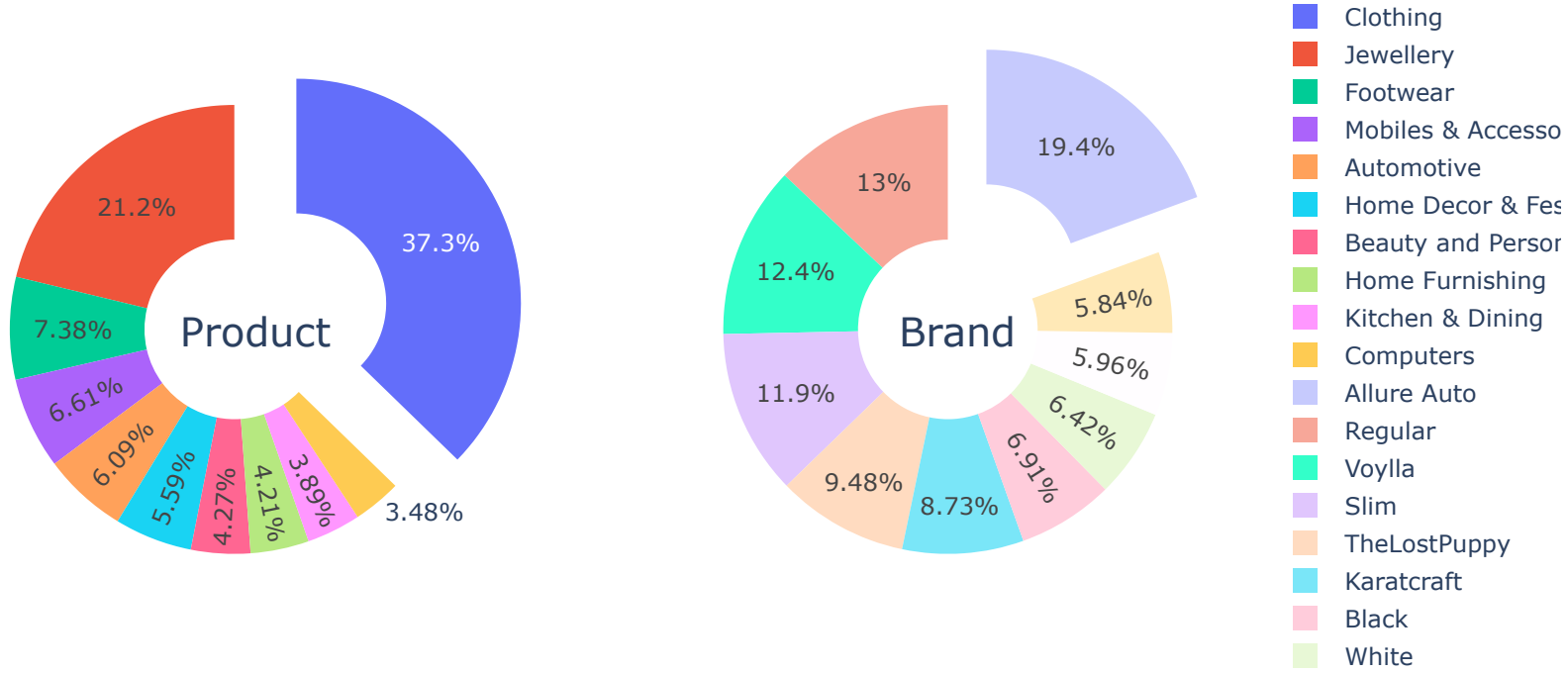
	uniq_id	crawl_timestamp	product_url	product_name	product_category_tree	
0	c2d766ca982eca8304150849735ffef9	2016-03-25 22:59:23 +0000	http://www.flipkart.com/alisha-solid-women-s-c...	Alisha Solid Women's Cycling Shorts	["Clothing >> Women's Clothing >> Lingerie, Sl...	SRTEH2FF9KE
1	7f7036a6d550aaa89d34c77bd39a5e48	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...	SBEEH3QGU7M
2	f449ec65dcbc041b6ae5e6a32717d01b	2016-03-25 22:59:23 +0000	http://www.flipkart.com/aw-bellies/p/itmeh4grg...	AW Bellies	["Footwear >> Women's Footwear >> Ballerinas >...	SHOEH4GRSUB
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...	SRTEH2F6HUZ
4	bc940ea42ee6bef5ac7cea3fb5cfbee7	2016-03-25 22:59:23 +0000	http://www.flipkart.com/sicons-all-purpose-arn...	Sicons All Purpose Arnica Dog Shampoo	["Pet Supplies >> Grooming >> Skin & Coat Care...	PSOEH3ZYDMS

In [50]:
```
n=10
top_products = pd.DataFrame(ds['main_category'].value_counts() [:n]).reset_index()
top_products.rename(columns = {'index': 'Top_Products','main_category':'Total_Count'}, inplace = True)

#Top 10 main brands being purchased

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

In [51]:

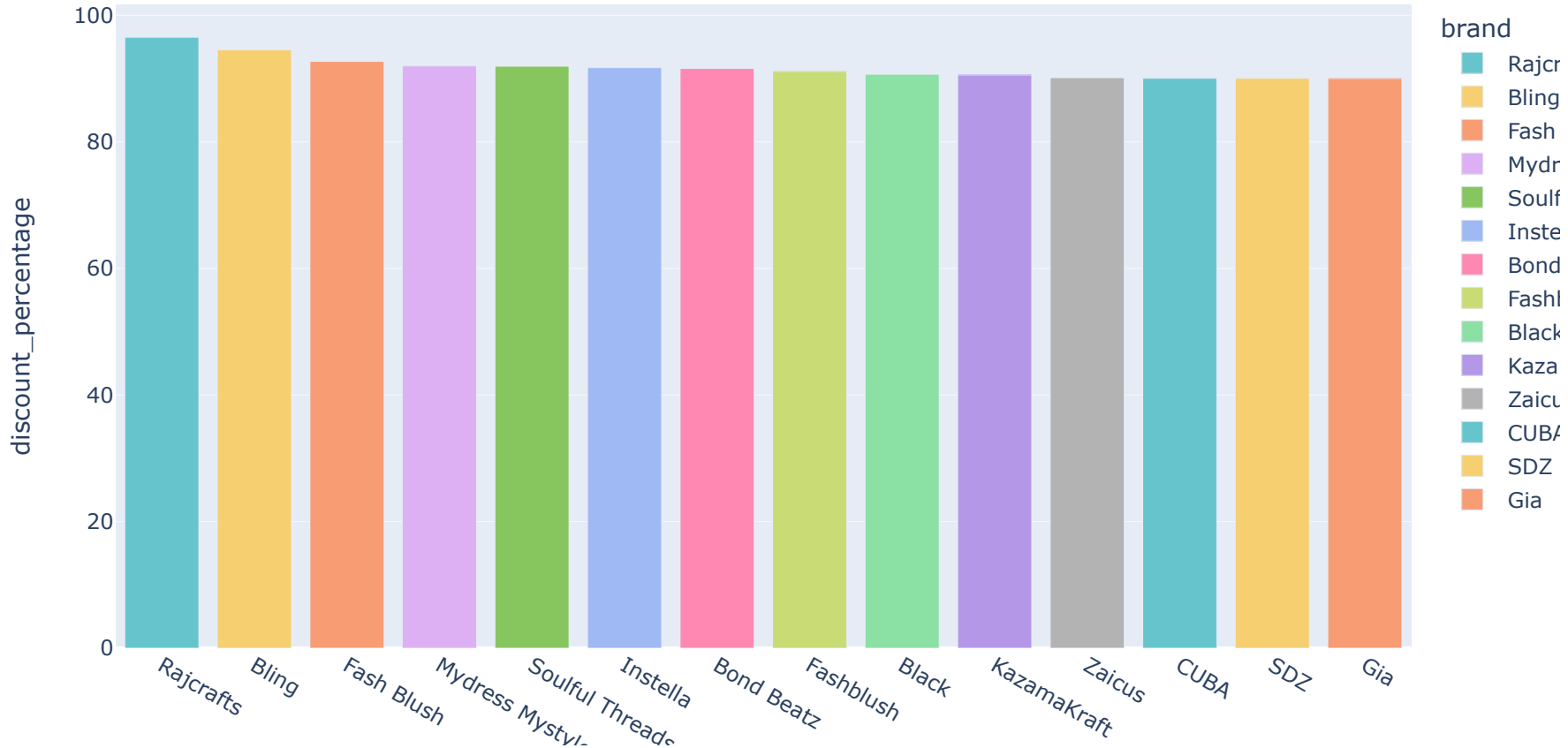
```
from plotly.subplots import make_subplots
label1 = top_products['Top_Products']
value1 = top_products['Total_Count']
label2 = top_brands['Top_Brands']
value2 = top_brands['Total_Count']

#Create Subplots

fig_both = make_subplots(rows=1, cols=2, specs=[[{'type':'domain'}, {'type':'domain'}]])
fig_both.add_trace(go.Pie(labels= label1, values= value1,
                          name="Top Products",pull=[0.3,0,0,0]),1,1)
fig_both.add_trace(go.Pie(labels= label2, values= value2,
                          name="Top Brands",pull=[0.3,0,0,0]),1,2)

#Use 'hole' to create a donut pie chart

fig_both.update_traces(hole=.4, hoverinfo="label+percent+name")

fig_both.update_layout(
title_text = "Top Products and Brands Dustribution",
#Add annotations in the center of the donut pies

annotations=[dict(text= 'Product',x=0.18,y=0.5, font_size=20,
                  showarrow=False),
             dict(text= 'Brand',x=0.83,y=0.5, font_size=20,
                  showarrow=False)])
```

Top Products and Brands Dustribution

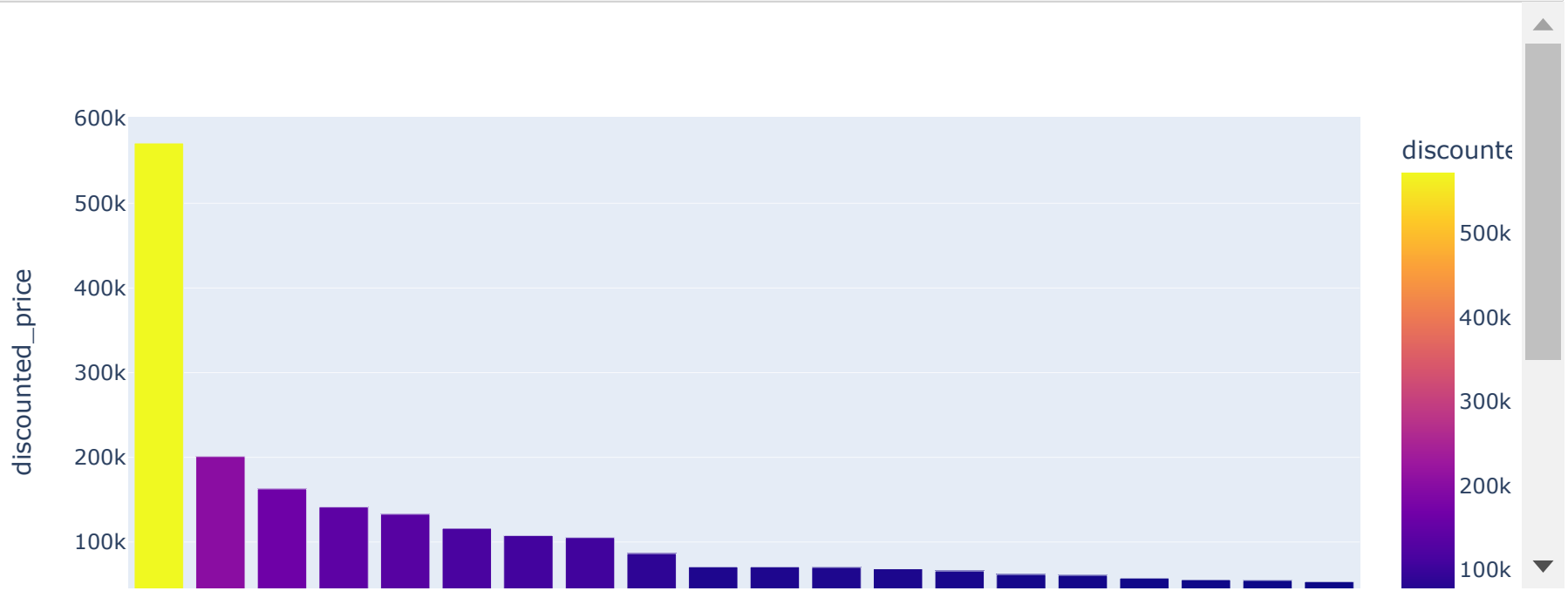


In [52]:
```
#Targeting brands giving high discount
ds_discount = ds.query('discount_percentage > 90')
#Dropping rows with NA values
ds_discount = ds_discount.dropna()
#Handling Spelling errors
ds_discount['brand'].replace('FashBlush','Fash Blush',inplace=True)
max_discount= pd.DataFrame(ds_discount.groupby('brand')[['discount_percentage']].mean().sort_values(by=['discount_perc
```

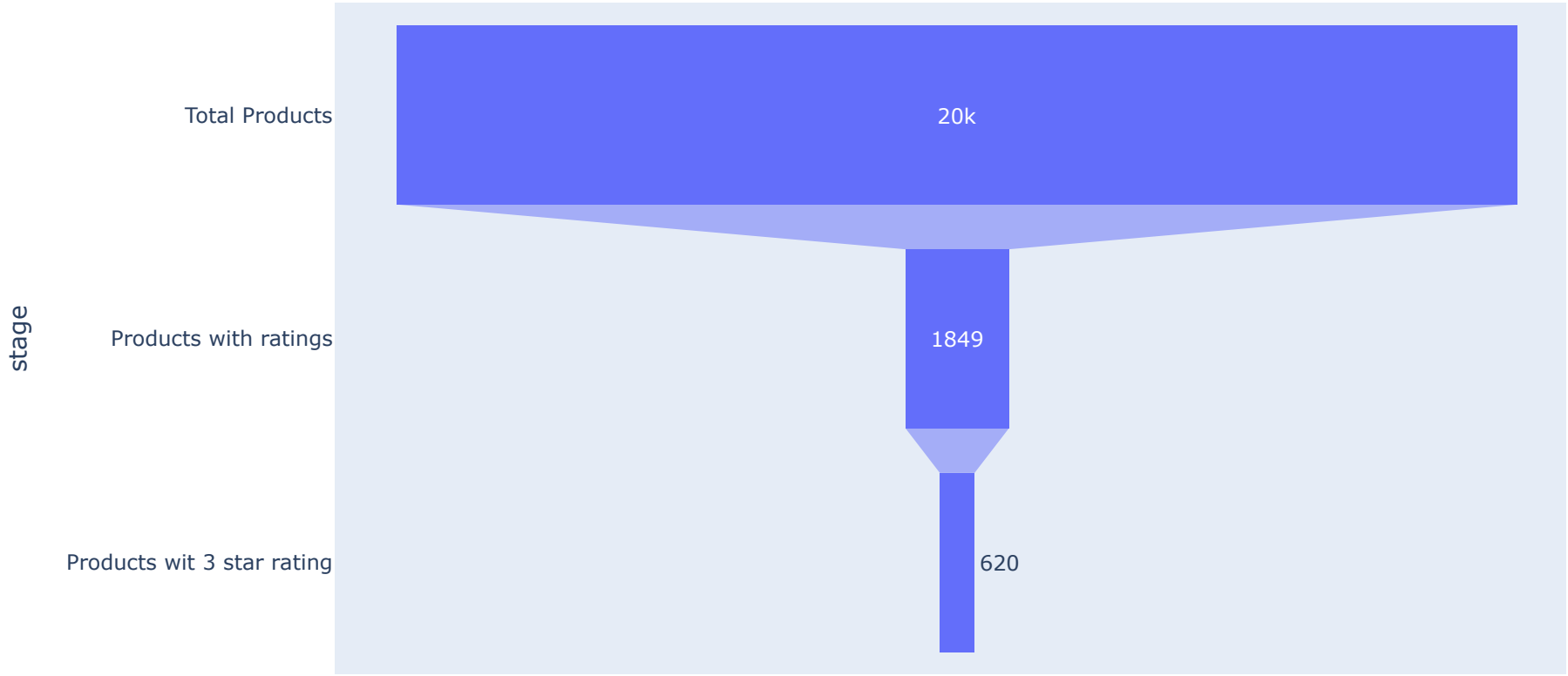
In [53]:
```
px.bar(max_discount, x= 'brand', y='discount_percentage',color='brand',
       color_discrete_sequence = px.colors.qualitative.Pastel)
```



In [54]:
```
ds_customer= ds.groupby("uniq_id")[['discounted_price']].sum().sort_values(by=['discounted_price'],ascending= False).

#Top 20 customers spending the most
list1 = ds_customer[:20]

#Plotting a bar graph
px.bar(list1, x= 'uniq_id', y='discounted_price',color='discounted_price',
       color_discrete_sequence = px.colors.qualitative.Pastel)
```



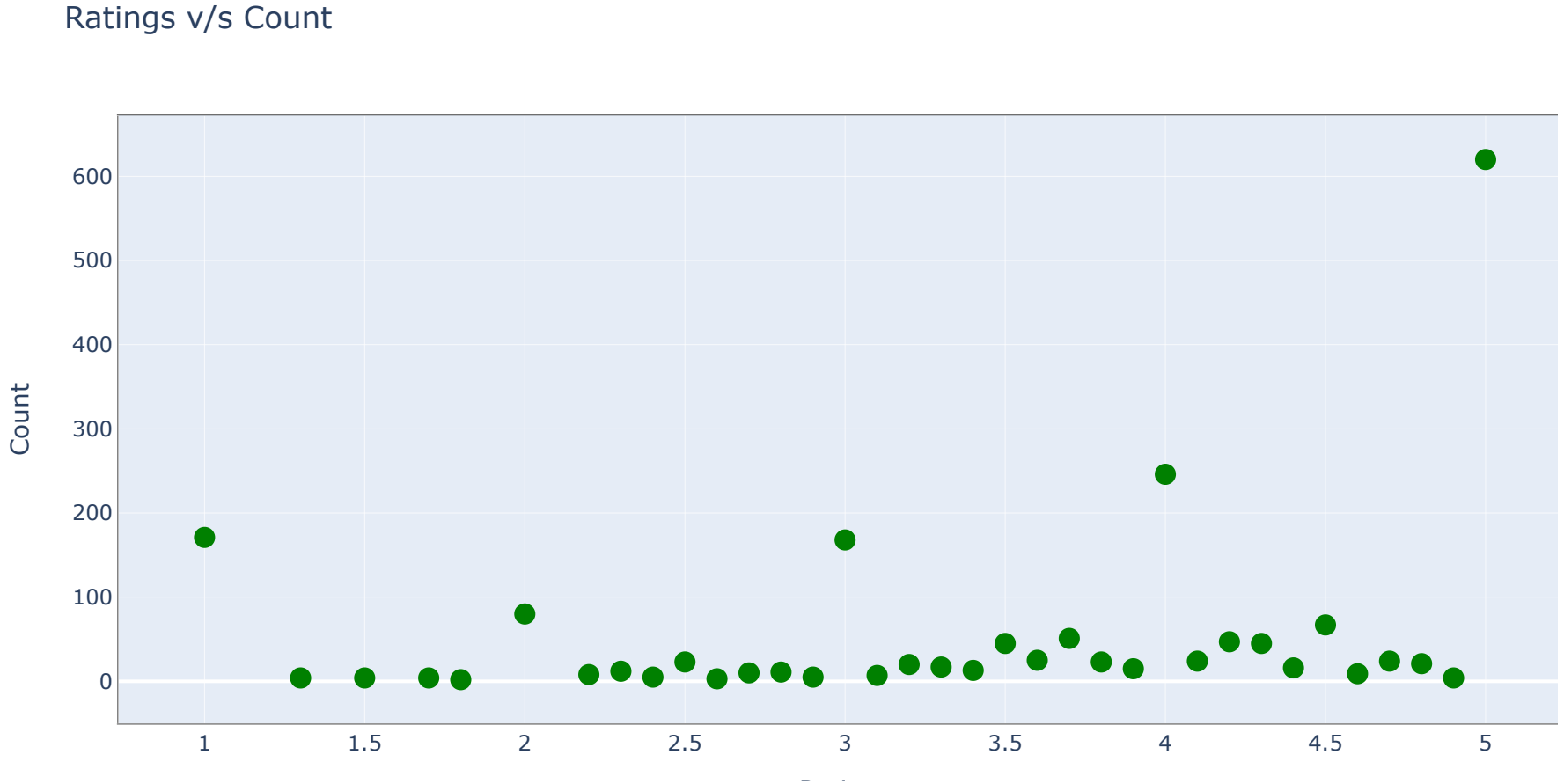
In [55]:
```
#5 star Rating

#Total Products using pid
total_prod = len(ds['pid'])
# Total Rated products
total_rating= len(ds[ds['product_rating']!='No rating available'])

# 5 star rated products
top_ratings= len(ds[ds['product_rating']=='5'])

ds_funnel_1 = dict(
    number = [total_prod,total_rating,top_ratings],
    stage=['Total Products','Products with ratings','Products wit 3 star rating'])

funnel_1_fig = px.funnel(ds_funnel_1, x='number',y='stage')
funnel_1_fig.show()
```


stage
Total Products
Products with ratings
Products wit 3 star rating
20k
1849
620

In [56]:
```
# 5 star products/brands
rating_5 = pd.DataFrame(ds.loc[ds['product_rating']=='5'])
# Top Products
top_product_type = rating_5['main_category'].value_counts()
# Top brands
top_brand_type = rating_5['brand'].value_counts()

# Top 5 products
ds_top_product = pd.DataFrame(top_product_type [:5].reset_index())
ds_top_product.rename(columns = {'index':'top_prod'}, inplace = True)
ds_top_product.drop('main_category', inplace =True, axis=1)

# Top 5 brands
ds_top_brand = pd.DataFrame(top_brand_type [:5].reset_index())
ds_top_brand.rename(columns = {'index':'top_brand'}, inplace = True)
ds_top_brand.drop('brand', inplace =True, axis=1)
ds_top_brand.head()
ds_top_product.head()

#Concating 2 tables
ds_product_brand_rate5 = pd.concat([ds_top_product,ds_top_brand], axis=1)
```

In [57]:

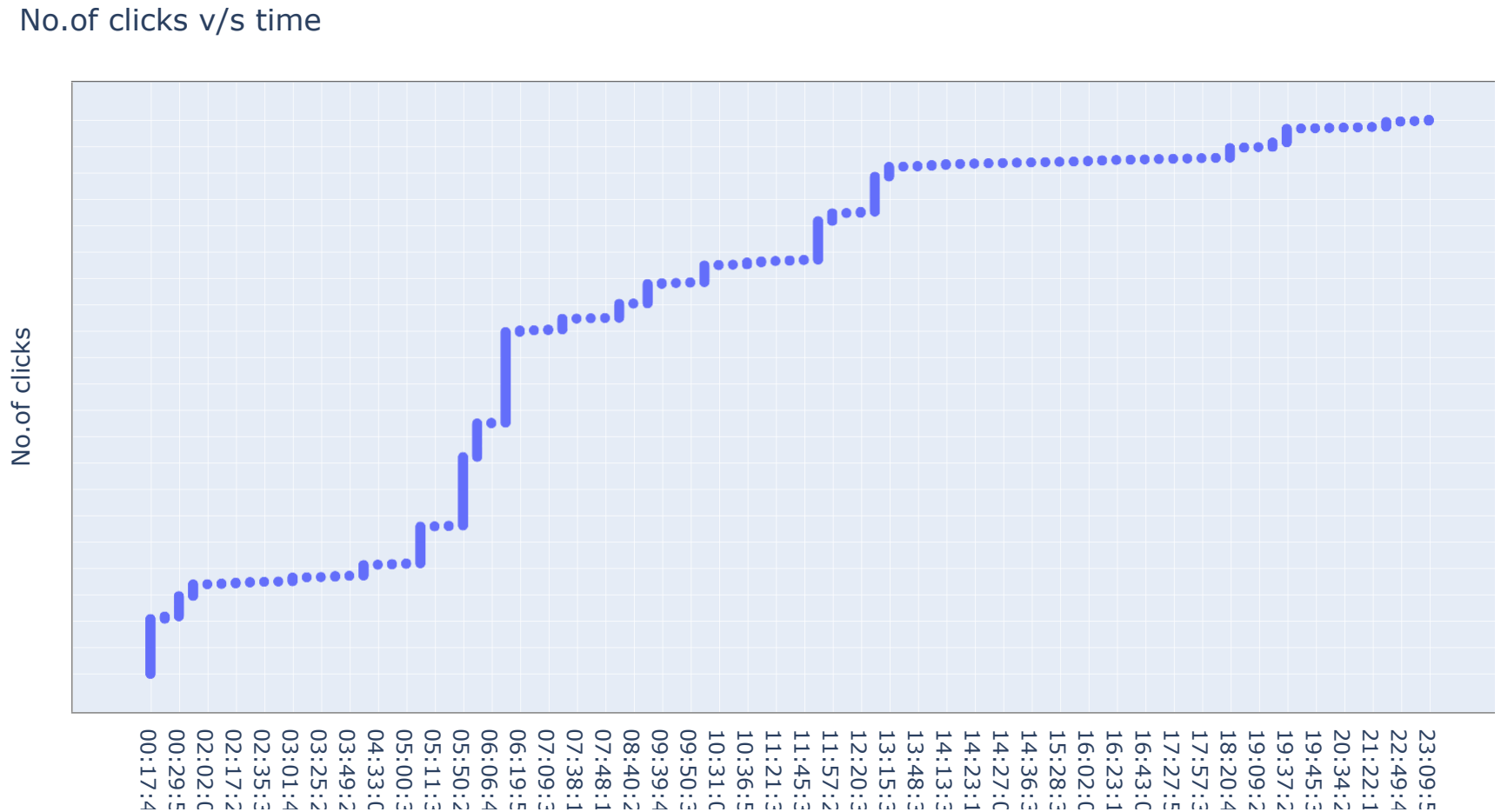
```
ds.drop(ds.index[ds['product_rating'] == 'No rating available'], inplace = True)
ratings = pd.DataFrame(ds['product_rating'].value_counts().reset_index())
ratings['index']= ratings['index'].astype(float)
ratings.head().sort_values(by= ['index'],ascending=[False])
ratings.rename(columns = {'index':'Ratings', 'product_rating':'Counts'}, inplace = True)

# Plotting the Result
data= ratings
x= ratings["Ratings"]
y= ratings['Counts']
figdot2 = go.Figure()
figdot2.add_trace(go.Scatter(
x=x,
y=y,
marker = dict(color="green", size=12),
mode='markers',
name=" ratings",
))

figdot2.update_layout(title='Ratings v/s Count',
                      xaxis_title ='Ratings',
                      yaxis_title ='Count',
                      )
figdot2.update_xaxes(showline=True, linewidth=1, linecolor='black', mirror=True)
figdot2.update_yaxes(showline=True, linewidth=1, linecolor='black', mirror=True)
figdot2.show()
```



In [58]:
```
ds['timestamp']=pd.to_datetime(ds['crawl_timestamp'])
ds['Time']= ds['timestamp'].apply(lambda x :x.time)
ds['date']= ds['timestamp'].apply(lambda x :x.date)
ds.drop(['crawl_timestamp'],axis = 1, inplace=True)
```

In [59]:

```
scat1 = px.scatter(x=ds['Time'].sort_values(ascending= True), y=ds['product_url'])
scat1.update_layout(
title_text = 'No.of clicks v/s time',
xaxis_title_text='Time',
yaxis_title_text='No.of clicks',
)

#scat1.update_xaxes(showticklabels=False)
scat1.update_yaxes(showticklabels=False)
scat1.update_xaxes(showline=True, linewidth=1, linecolor='black', mirror=True)
scat1.update_yaxes(showline=True, linewidth=1, linecolor='black', mirror=True)
scat1.show()
```


No.of clicks v/s time
No.of clicks
00:17:4
00:29:5
02:02:0
02:17:2
02:35:3
03:01:4
03:25:2
03:49:2
04:33:0
05:00:3
05:11:3
05:50:2
06:06:4
06:19:5
07:09:3
07:38:1
07:48:1
08:40:2
09:39:4
09:50:3
10:31:0
10:36:5
11:21:3
11:45:3
11:57:2
12:20:3
13:15:3
13:48:3
14:13:3
14:23:1
14:27:0
14:36:3
15:28:3
16:02:0
16:23:1
16:43:0
17:27:5
17:57:3
18:20:4
19:09:2
19:37:2
19:45:3
20:34:2
21:22:1
22:49:4
23:09:5

In [ ]: