In [3]:
```
import numpy as np
import pandas as pd
import ast
import plotly.express as px
from plotly import graph_objects as go
```

In [6]:
```
df = pd.read_csv("flipkart_com-ecommerce_sample.csv")
```

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

Out[7]:

	uniq_id	crawl_timestamp	product_url	product_name	product_category_tree	pid	retail_
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...	SRTEH2FF9KEDEFGF	
1	7708b9a6d550aaa89ac34c77bc00a5a48	2016-03-25 22:59:23 +0000	http://www.flipkart.com/tabhomedecor-fabric-do...	FabHomeDecor Fabric Double Sofa Bed	["Furniture >> Living Room Furniture >> Sofa B...	SBEEH3QGU7MFYJFY	32
2	f439ec85ccbc041b8ae5e6a32717d01b	2016-03-25 22:59:23 +0000	http://www.flipkart.com/aw-bellies-pltmeh4prg...	AW Bellies	["Footwear >> Women's Footwear >> Ballerinas >...	SHOEH4GR5USJGZXE	
3	0973b37acd0c664e3de26e97e0071454	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...	SRTEH2HGHUZMDGSJ	
4	bd040aa42aa8bef5ec7aaa3fa5c1aaa7	2016-03-25 22:59:23 +0000	http://www.flipkart.com/scrons-all-purpose-ant...	Scrons All Purpose Anti-a Dog Shampoo	["Pet Supplies >> Grooming >> Skin & Coat Care...	PSOFH3ZYDMSYARJ5	

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

Out[8]:

	uniq_id	crawl_timestamp	product_url	product_name	product_category_tree	pid	retail_
19995	7179d2f8bfad50a17d014ca1d2616156	2015-12-01 10:15:43 +0000	http://www.flipkart.com/walldesign-small-vinyl...	WallDesign Small Vinyl Sticker	["Baby Care >> Baby & Kids Gifts >> Stickers >...	STIE7K7JAKSTDY9G	1
19996	71ac419198359d37b8fe5e3ffcfee09	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 >...	STIE9F5JRNQGJCGH	1
19997	93e6d345837400ae0d798387feee171c	2015-12-01 10:15:43 +0000	http://www.flipkart.com/elite-collection-medu...	Elite Collection Medium Acrylic Sticker	["Baby Care >> Baby & Kids Gifts >> Stickers >...	STIE7VAYDKQZE55D	1
19998	669a73b8fa5c6ee000841c0c87c5e935	2015-12-01 10:15:43 +0000	http://www.flipkart.com/elite-collection-medu...	Elite Collection Medium Acrylic Sticker	["Baby Care >> Baby & Kids Gifts >> Stickers >...	STIF8Y5VFPPC742Y	1
19999	cb41a87a874f716ff8677b7b3ba78c	2015-12-01 10:15:43 +0000	http://www.flipkart.com/elite-collection-medu...	Elite Collection Medium Acrylic Sticker	["Baby Care >> Baby & Kids Gifts >> Stickers >...	STIF88KN67DSCZKY	1

In [9]:
```
df.isnull().sum()
```

Out[9]:
```
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
```

In [11]:
```
df['retail_price'].fillna(df['retail_price'].median(),inplace=True)
df["discounted_price"].fillna(df["discounted_price"].median(),inplace=True)
```

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

Out[12]:

uniq_id	crawl_timestamp	product_url	product_name	product_category_tree	pid	retail_pri
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...	SRTEH2FFBKEDEFGF	99
7f7036a6d550aaa89d34c77bd39a5e48	2016-03-25 22:59:23 +0000	http://www.flipkart.com/sbhomedecor-fabric-do...	FabHomeDecor Fabric Double Sofa Bed	["Furniture >> Living Room Furniture >> Sofa B...	SBEEH3QGU7MFYJFY	32157
f449ec65dcbc041b6ae5e9a32717c01b	2016-03-25 22:59:23 +0000	http://www.flipkart.com/aw-bellies/p/itmeh4grg...	AW Bellies	["Footwear >> Women's Footwear >> Ballerinas >...	SHOEH4GRSUBJGZXE	999
0973b37acd0c664e3de26a97a5571454	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...	SRTEH2F6HUZMQ6SJ	699
bc840ea42ee8ae5ac7cee30b5cfbee7	2016-03-25 22:59:23 +0000	http://www.flipkart.com/sicons-all-purpose-am...	Sicons All Purpose Arnica Dog Shampoo	["Pet Supplies >> Grooming >> Skin & Coat Care...	PSOEH3ZYDMSYARJ5	220

In [14]:
```
x=df['retail_price']-df['discounted_price']
y=(x/df['retail_price'])*100
df['discount_percentage']=y
```

In [15]:
```
df.head()
```

Out[15]:

ge	is_FK_Advantage_product	description	product_rating	overall_rating	brand	product_specifications	discounted_percentage	discount_percentage
/...	False	Key Features of Alisha Solid Women's Cycling S...	No rating available	No rating available	Alisha	{"product_specification"=> [{"key"=>"Number of ...	62.062062	62.062062
)...	False	FabHomeDecor Fabric Double Sofa Bed (Finish:Co...	No rating available	No rating available	FabHomeDecor	{"product_specification"=> [{"key"=>"Installati...	29.578784	29.578784
s...	False	Key Features of AW Bellies Sandals Wedges Heel...	No rating available	No rating available	AW	{"product_specification"=> [{"key"=>"Ideal For"...	50.050050	50.050050
/...	False	Key Features of Alisha Solid Women's Cycling S...	No rating available	No rating available	Alisha	{"product_specification"=> [{"key"=>"Number of ...	61.802575	61.802575
el- /...	False	Specifications of Sicons All Purpose Arnica Do...	No rating available	No rating available	Sicons	{"product_specification"=> [{"key"=>"Pet Type",...	4.545455	4.545455

In [21]:
```
df['timestamp']=pd.to_datetime(df['crawl_timestamp'])
df['Time']=df['timestamp'].apply(lambda x : x.time)
df['date']=df['timestamp'].apply(lambda x: x.date)
df.drop(['crawl_timestamp'], axis = 1, inplace=True)
df['main_category'] = df['product_category_tree'].apply(lambda x: x.split('>>')[0][2:])
```

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

Out[22]:

roduct_rating	overall_rating	brand	product_specifications	discounted_percentage	discount_percentage	timestamp	Time	date	main_category
No rating available	No rating available	Alisha	{"product_specification"=> [{"key"=>"Number of ...	62.062062	62.062062	2016-03-25 22:59:23+00:00	22:59:23	2016-03-25	Clothing
No rating available	No rating available	FabHomeDecor	{"product_specification"=> [{"key"=>"Installati...	29.576764	29.576764	2016-03-25 22:59:23+00:00	22:59:23	2016-03-25	Furniture
No rating available	No rating available	AW	{"product_specification"=> [{"key"=>"Ideal For"...	50.050050	50.050050	2016-03-25 22:59:23+00:00	22:59:23	2016-03-25	Footwear
No rating available	No rating available	Alisha	{"product_specification"=> [{"key"=>"Number of ...	61.802575	61.802575	2016-03-25 22:59:23+00:00	22:59:23	2016-03-25	Clothing
No rating available	No rating available	Sicons	{"product_specification"=> [{"key"=>"Pet Type",...	4.545455	4.545455	2016-03-25 22:59:23+00:00	22:59:23	2016-03-25	Pet Supplies

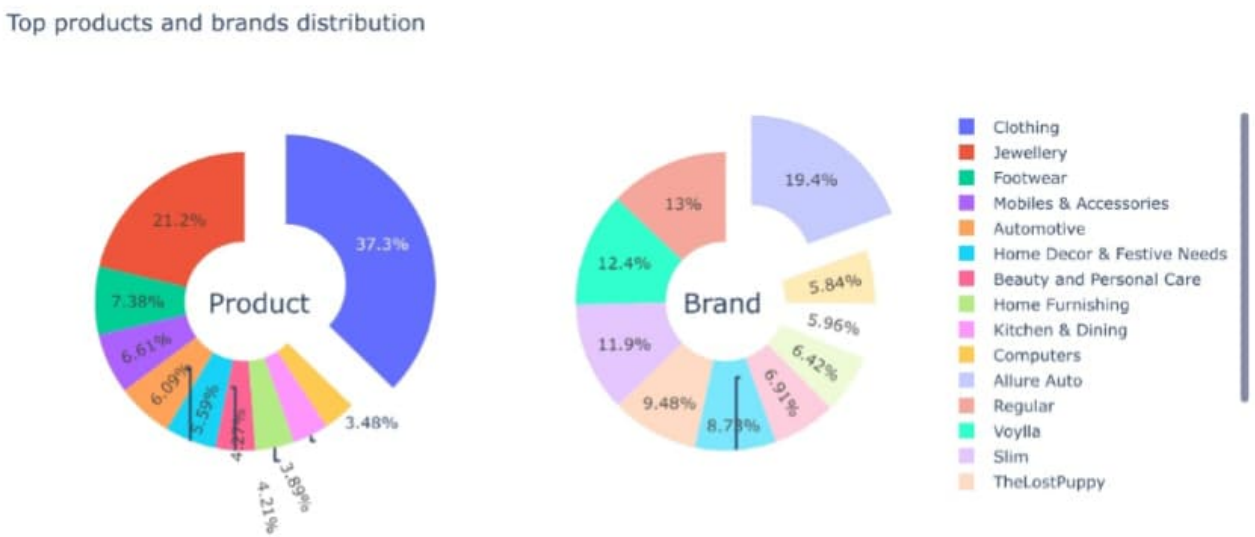
In [23]:
```
n=10
top_products=pd.DataFrame(df['main_category'].value_counts() [:n]).reset_index()
top_products.rename(columns= {'index': 'Top_Products','main_category':'Total_Count'},inplace=True)

#Top 10 main brands being purchased

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

In [24]:
```
df.head()
```

Out[24]:

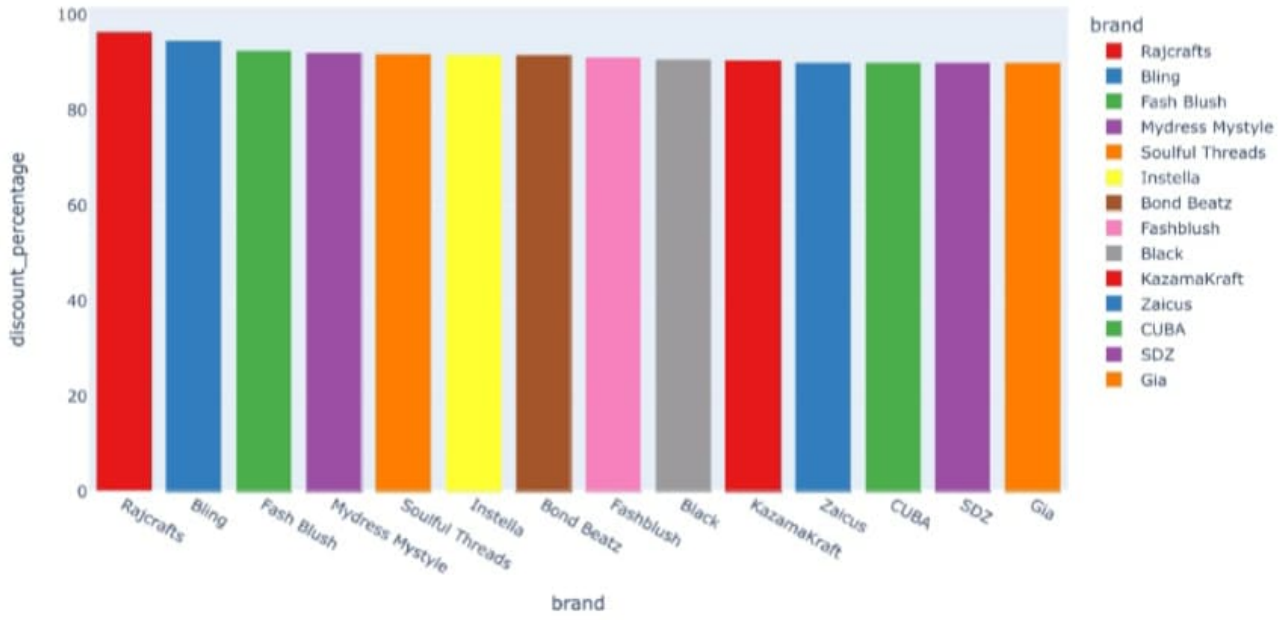
	uniq_id	product_url	product_name	product_category_tree	pid	retail_price	discounted_
0	c2d766ca982eca8304150849735ffef9	http://www.flipkart.com/alisha-solid-women-s-c...	Alisha Solid Women's Cycling Shorts	["Clothing >> Women's Clothing >> Lingerie, Sl...	SRTEH2FF9KEDEFGF	999.0	
1	7f7036a6d550aaa89d34c77bd39a5e48	http://www.flipkart.com/fabhomedecor-fabric-do...	FabHomeDecor Fabric Double Sofa Bed	["Furniture >> Living Room Furniture >> Sofa B...	SBEEH3QGU7MFYJFY	32157.0	22
2	f449ec65dcbc041b6ae5e6a32717d01b	http://www.flipkart.com/aw-bellies/p/itmeh4grg...	AW Bellies	["Footwear >> Women's Footwear >> Ballerinas >...	SHOEH4GRSUBJGZXE	999.0	
3	0973b37acd0c664e3de26e97e5571454	http://www.flipkart.com/alisha-solid-women-s-c...	Alisha Solid Women's Cycling Shorts	["Clothing >> Women's Clothing >> Lingerie, Sl...	SRTEH2F6HUZMQ6SJ	699.0	
4	bc940ea42ee6bef5ac7cea3fb5cfbee7	http://www.flipkart.com/sicons-all-purpose-arn...	Sicons All Purpose Arnica Dog Shampoo	["Pet Supplies >> Grooming >> Skin & Coat Care...	PSOEH3ZYDMSYARJ5	220.0	

In [29]:
```
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']

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)

fig_both.update_traces(hole=.4, hoverinfo="label+percent+name")
fig_both.update_layout(
    title_text="Top products and brands distribution",
    annotations=[
        dict(text='Product', x=0.18, y=0.5, font_size=20, showarrow=False),
        dict(text='Brand', x=0.82, y=0.5, font_size=20, showarrow=False)
    ]
)
```

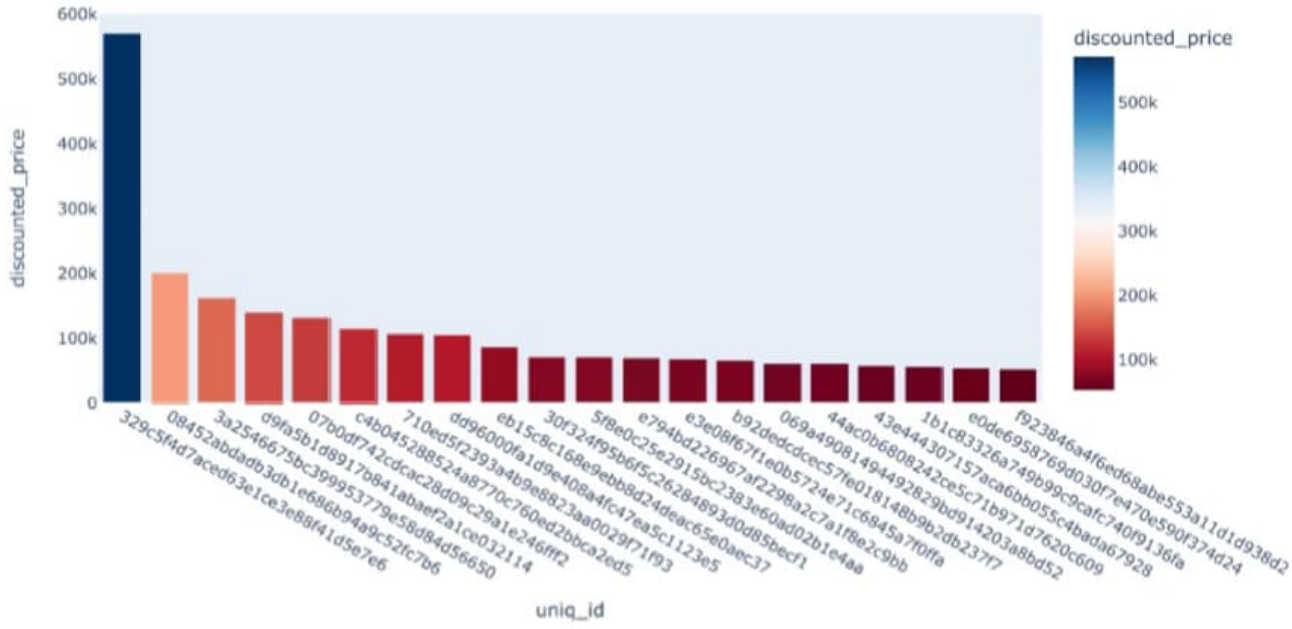


In [30]:
```
df_discount=df.query('discount_percentage > 90')
df_discount=df_discount.dropna()
df_discount["brand"].replace('FashBlush','Fash Blush', inplace=True)
max_discount=pd.DataFrame(df_discount.groupby('brand')[['discount_percentage']].mean().sort_values(by=['discount_percentage'],asc
```

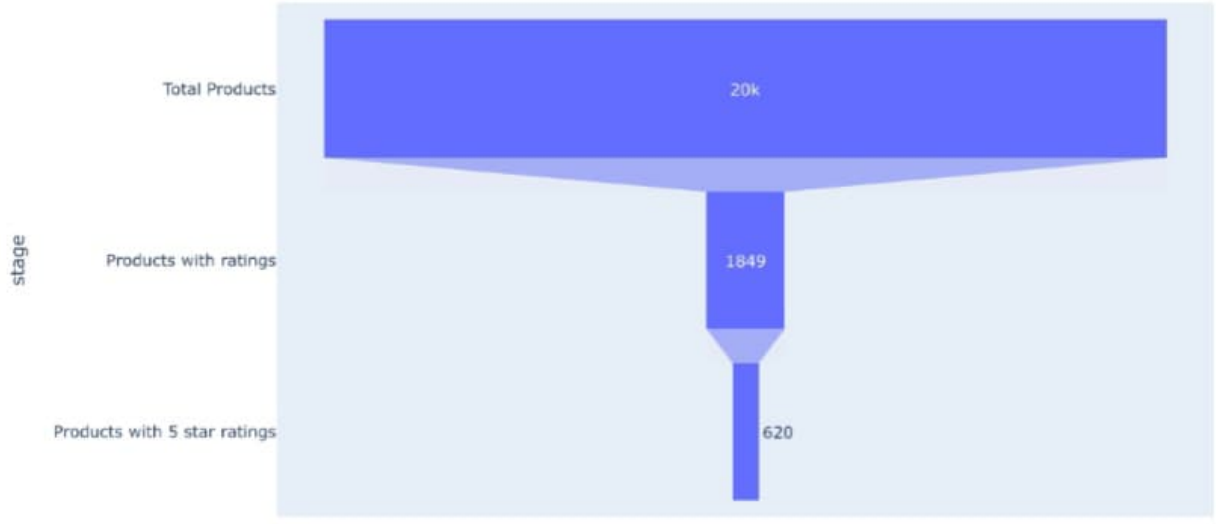
In [32]:
```
px.bar(max_discount, x='brand', y='discount_percentage', color='brand', color_discrete_sequence=px.colors.qualitative.Set1)
```



In [41]:
```
import plotly.express as px

df_customer = df.groupby("uniq_id")[["discounted_price"]].sum().sort_values(by='discounted_price', ascending=False)
list1 = df_customer[:20].reset_index()  # Resetting index to make 'uniq_id' a column

fig = px.bar(list1, x='uniq_id', y='discounted_price', color='discounted_price', color_continuous_scale='RdBu')
fig.show()
```



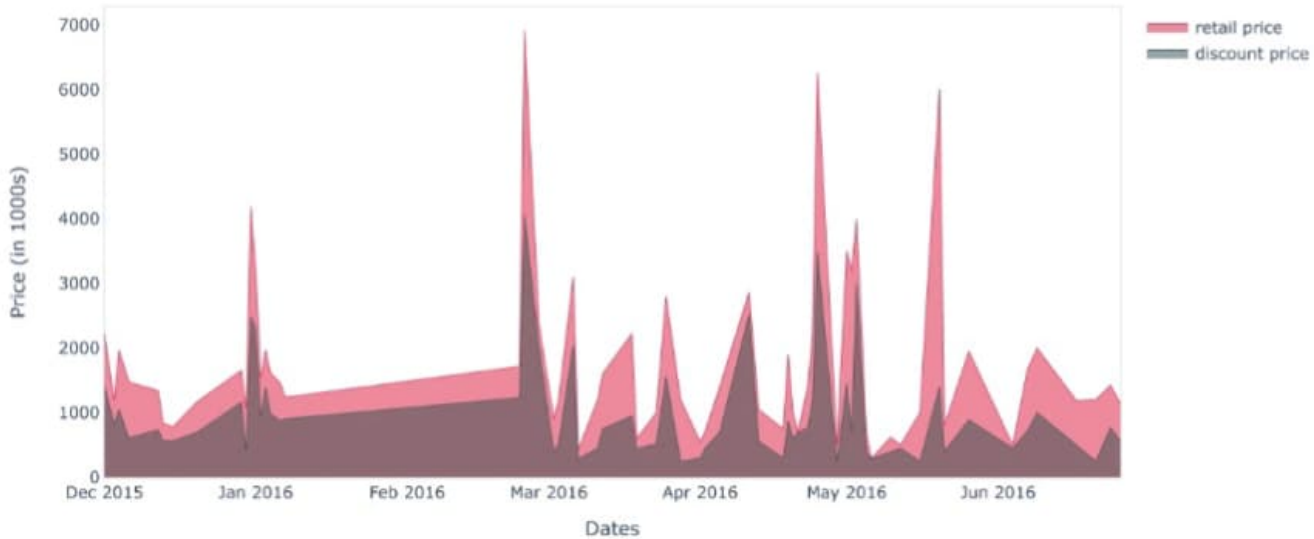
In [42]:
```
total_prod=len(df['pid'])
total_ratings=len(df[df['product_rating']!='No rating available'])
top_ratings=len(df[df['product_rating']=='5'])
df_funnel_1=dict(
    number=[total_prod, total_ratings,top_ratings],
    stage=["Total Products","Products with ratings","Products with 5 star ratings"])
funnel_1_fig=px.funnel(df_funnel_1, x='number',y='stage')
funnel_1_fig.show()
```



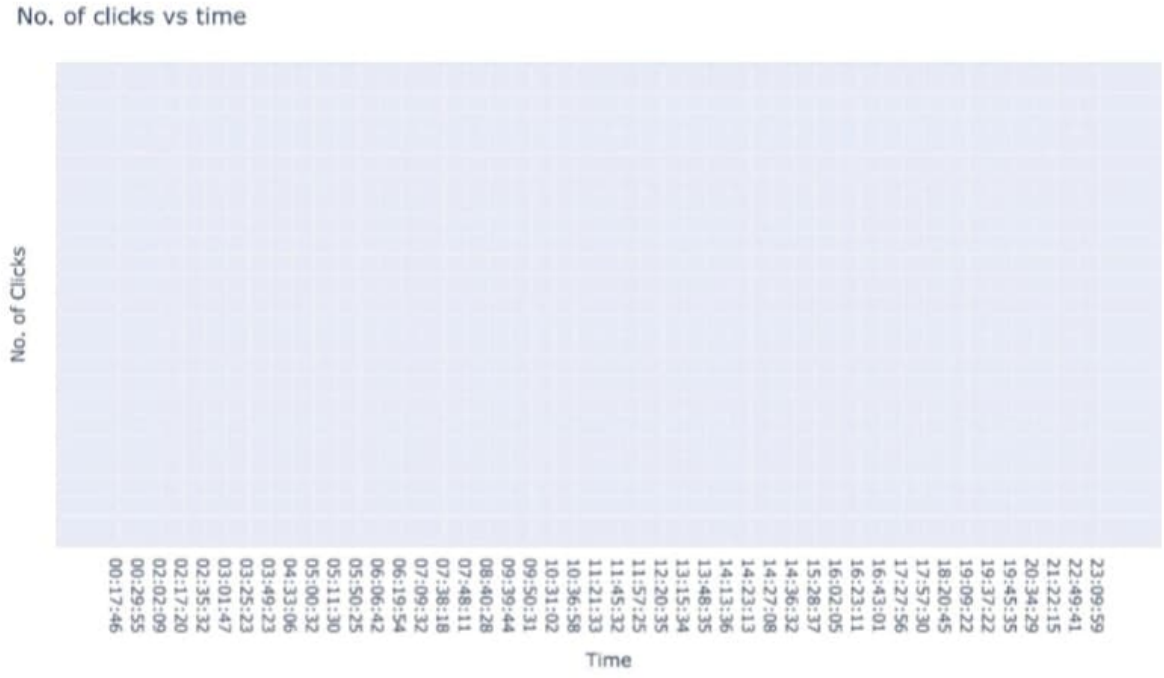
In [45]:
```
#5 star products/brands

rating_5=pd.DataFrame(df.loc[df['product_rating'] == '5'])
top_product_type=rating_5['main_category'].value_counts() #top products
top_brand_type=rating_5['brand'].value_counts() #top brands

#top 5 products

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

#top 5 brands

df_top_brand=pd.DataFrame(top_brand_type[:5].reset_index())
df_top_brand.rename(columns = {'index':'top_brands'}, inplace = True)
df_top_brand.drop('brand', inplace=True, axis=1)
df_top_brand.head()

#concatenating the 2 tables

df_product_brand_rate5=pd.concat([df_top_product,df_top_brand], axis=1)
```

In [51]:
```
import plotly.graph_objects as go

# Assuming you have a DataFrame named 'df' with a column 'product_rating'
df.drop(df.index[df['product_rating'] == 'No rating available'], inplace=True)

# Assuming you have another DataFrame named 'ratings'
ratings = df['product_rating'].value_counts().sort_index().reset_index()
ratings['index'] = ratings['index'].astype(float)
ratings.rename(columns={'index': 'Ratings', 'product_rating': 'Counts'}, inplace=True)

# Plotting the result
x = ratings['Ratings']
y = ratings['Counts']

figdot2 = go.Figure()

figdot2.add_trace(go.Scatter(
    x=x,
    y=y,
    marker=dict(color="crimson", 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 [54]:
```
df_date_retail = pd.DataFrame(df.groupby("date")[["retail_price"]].mean().reset_index())
df_date_discount = pd.DataFrame(df.groupby("date") [["discounted_price"]].mean().reset_index())
df_date_price=pd.concat([df_date_retail,df_date_discount],axis=1)
df_date_price = df_date_price.loc[:,~df_date_price.columns.duplicated()] #remove duplicate columns

#Plot
x=df_date_price['date']
y1=df_date_price['retail_price']
y2=df_date_price['discounted_price']

fig_area2 = go. Figure()
fig_area2.add_trace(go.Scatter(x=x, y=y1, fill='tozeroy', name='retail price',
                               line=dict(width=0.5, color='crimson'))) # fill down to xaxis
fig_area2.add_trace(go.Scatter(x=x, y=y2, fill='tozeroy', name='discount price',
                               line=dict(width=0.5, color='darkslategray')
                                       )) # fill to traced y
fig_area2.update_layout(
    xaxis_title="Dates",
    yaxis_title="Price (in 1000s)",
    plot_bgcolor="white"

)

fig_area2.update_xaxes(showline=True, linewidth=1, linecolor='black', mirror=True)
fig_area2.update_yaxes(showline=True, linewidth=1, linecolor='black', mirror=True)
fig_area2.show()
```



In [55]:
```
scat2 = px.scatter(x=df['Time'].sort_values(ascending=True), y=df['product_url'])

scat2.update_layout(

title_text='No. of clicks vs time', # title of plot

xaxis_title_text='Time', # xaxis label

yaxis_title_text='No. of Clicks', # yaxis label

)

#scat.update_xaxes(showticklabels=False)
scat2.update_yaxes(showticklabels=False)

scat2.update_xaxes(showline=True, linewidth=1, linecolor='black', mirror=True)

scat2.update_yaxes(showline=True, linewidth=1, linecolor='black', mirror=True)

scat2.show()
```



In [ ]: