

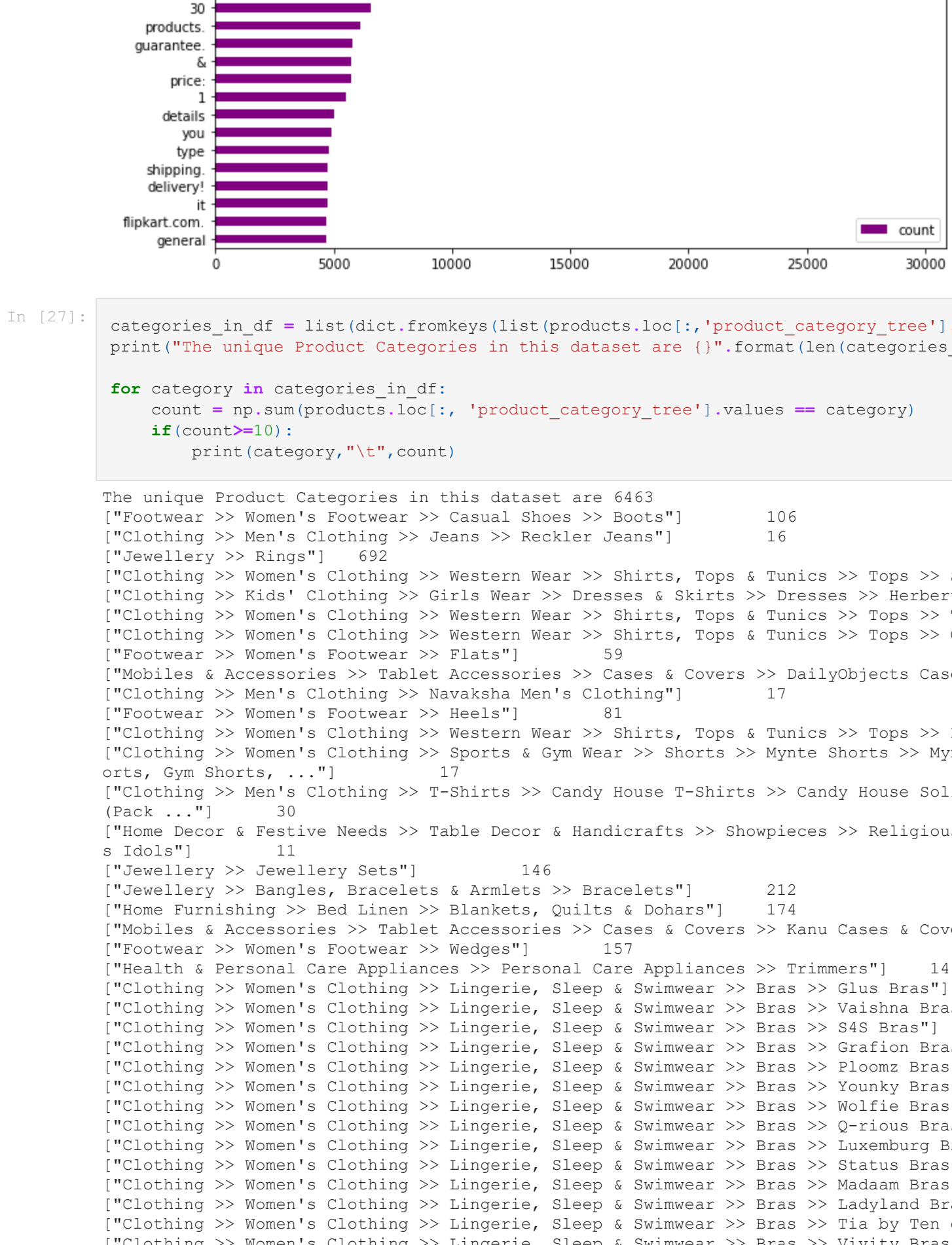
[illegible]



```
def most_frequent_words(description):  
    for i in range(len(description)):  
        description[i] = description[i].lower().split()  
  
    all_words = list(texttools.chain(*description))  
    word_counts = Counter(all_words)  
  
    #creating a pandas dataframe of the 40 most common words in the dataset  
    clean_description = pd.DataFrame(word_counts.most_common(40),  
                                     columns=['words', 'count'])  
  
    return clean_description  
  
description = products["description"].tolist()  
most_common_description = most_frequent_words(description)  
  
fig, ax = plt.subplots(figsize=(10, 10))  
# plotting the bargraph in decreasing sorted order  
most_common_description.sort_values(by='count').plot.barh(ax=ax,  
                                                         color='purple')  
  
ax.set_title('Common Words Found in Product Description (including all words)')  
plt.show()
```

Out[26]: <AxesSubplot:ylabel='words'>

Out[26]: Text(0.5, 1.0, 'Common Words Found in Product Description (including all words)')



In [27]:  
categories\_in\_df = list(dict.fromkeys(list(products.loc[:, 'product\_category\_tree'].values)))  
print("The unique Product Categories in this dataset are (%): %format(len(categories\_in\_df))")  
  
for category in categories\_in\_df:  
 count = np.sum(products.loc[:, 'product\_category\_tree'].values == category)  
 if (count > 0):  
 print(category, "\t", count)

The unique Product Categories in this dataset are 6463  
["Footwear > Women's Footwear > Casual Shoes > Boots"] 106  
["Clothing > Men's Clothing > Jeans > Reckler Jeans"] 16  
["Clothing > Kids' Clothing > Western Wear > Shirts, Tops & Tunics > Tops > Shipkale Tops"] 14  
["Clothing > Kids' Clothing > Girls Wear > Dresses & Skirts > Dresses > Hebertto Dresses"] 10  
["Clothing > Women's Clothing > Western Wear > Shirts, Tops & Tunics > Tops > Tappi Tops"] 11  
["Clothing > Women's Clothing > Western Wear > Shirts, Tops & Tunics > Tops > Cation Tops"] 39  
["Footwear > Women's Footwear > Flats"] 50  
["Mobiles & Accessories > Tablet Accessories > Cases & Covers > DailyObjects Cases & Covers"] 60  
["Clothing > Men's Clothing > Navakha Men's Clothing"] 17  
["Footwear > Women's Footwear > Heels"] 81  
["Clothing > Women's Clothing > Western Wear > Shirts, Tops & Tunics > Tops > FabAlley Tops"] 14  
["Clothing > Women's Clothing > Sports & Gym Wear > Shorts > Mynte Shorts > Mynte Solid Women's Cycling Shirts, Gym Shorts"] 17  
["Clothing > Men's Clothing > T-Shirts > Candy House T-Shirts > Candy House Solid Men's Polo Neck T-Shirt (Pack of 3)"] 30  
["Home Decor & Festive Needs > Table Decor & Handicrafts > Showpieces > Religious Idols > Divinita Religious Idols"] 11  
["Jewellery > Jewellery Sets"] 146  
["Jewellery > Bangles, Bracelets & Armbands > Bracelets"] 212  
["Home Furnishing > Bed Linen > Blankets, Quilts & Dohars"] 174  
["Mobiles & Accessories > Tablet Accessories > Cases & Covers > Manu Cases & Covers"] 20  
["Footwear > Women's Footwear > Wedges"] 157  
["Health & Personal Care Appliances > Personal Care Appliances > Trimmers"] 14  
["Clothing > Women's Clothing > Lingerie, Sleep & Swimwear > Bras > Glus Bras"] 31  
["Clothing > Women's Clothing > Lingerie, Sleep & Swimwear > Bras > Vainsha Bras"] 22  
["Clothing > Women's Clothing > Lingerie, Sleep & Swimwear > Bras > 448 Bras"] 117  
["Clothing > Women's Clothing > Lingerie, Sleep & Swimwear > Bras > Grafion Bras"] 67  
["Clothing > Women's Clothing > Lingerie, Sleep & Swimwear > Bras > Ploomz Bras"] 60  
["Clothing > Women's Clothing > Lingerie, Sleep & Swimwear > Bras > Youky Bras"] 26  
["Clothing > Women's Clothing > Lingerie, Sleep & Swimwear > Bras > Vivity Bras"] 29  
["Clothing > Women's Clothing > Lingerie, Sleep & Swimwear > Bras > Wolfie Bras"] 13  
["Clothing > Women's Clothing > Lingerie, Sleep & Swimwear > Bras > Qrious Bras"] 11  
["Clothing > Women's Clothing > Lingerie, Sleep & Swimwear > Bras > Luxemburg Bras"] 17  
["Clothing > Women's Clothing > Lingerie, Sleep & Swimwear > Bras > Status Bras"] 27  
["Clothing > Women's Clothing > Lingerie, Sleep & Swimwear > Bras > La Verite Bras"] 12  
["Clothing > Women's Clothing > Lingerie, Sleep & Swimwear > Bras > Our Rituals Bras"] 15  
["Clothing > Women's Clothing > Lingerie, Sleep & Swimwear > Bras > Tia by Ten on Ten Bras"] 49  
["Clothing > Women's Clothing > Lingerie, Sleep & Swimwear > Bras > Luxemburg Bras"] 10  
["Clothing > Women's Clothing > Lingerie, Sleep & Swimwear > Bras > Oleva Bras"] 10  
["Clothing > Women's Clothing > Lingerie, Sleep & Swimwear > Bras > Urbano Bras"] 25  
["Clothing > Women's Clothing > Lingerie, Sleep & Swimwear > Bras > R8 Paris Beauty Bras"] 13  
["Clothing > Women's Clothing > Lingerie, Sleep & Swimwear > Bras > Showtime Bras"] 14  
["Clothing > Women's Clothing > Lingerie, Sleep & Swimwear > Bras > Loviniform Bras"] 14  
["Clothing > Women's Clothing > Lingerie, Sleep & Swimwear > Bras > Nuquam Bras"] 12  
["Fens & Stationery > College Supplies > SHOPMANIA College Supplies > SHOPMANIA DESIGNER NOTEBOOK A5 Notebook & Spiral B..."] 11  
["Footwear > Women's Footwear > Casual Shoes > Loafers"] 85  
["Kitchen & Dining > Coffee Mugs > Frutland Coffee Mugs"] 84  
["Home Decor & Festive Needs > Coffee Mugs > Frutland Coffee Mugs"] 89  
["Kitchen & Dining > Coffee Mugs > Rockmantra Coffee Mugs"] 72  
["Home Furnishing > Curtains & Accessories > Curtains"] 7  
["Automotive > Accessories & Spare Parts > Car Interior & Exterior > Car Interior > Car Mats"] 520  
["Home Decor & Festive Needs > Showpieces > Ethnic > Lal Havelli Ethnic"] 33  
["Home Decor & Festive Needs > Showpieces > Ethnic > MCB Sports Ethnic"] 11  
["Jewellery > Necklaces & Chains > Necklaces"] 1172  
["Automotive > Accessories & Spare Parts > Car Electronics & Accessories > Car Mobile Accessories > Car Mobile Holders"] 46  
["Footwear > Women's Footwear > Heels > Catalogue Women Heels"] 11  
["Baby Care > Infant Wear > Baby Boys' Clothes > Dungarees & Jumpsuits > Dungarees > Babeeworld Dungarees > Babeeworld Baby Boy's Romper"] 10  
["Beauty and Personal Care > Fragrances > Deodorants > Combos > Anna Andre Paris Combos"] 12  
["Beauty and Personal Care > Fragrances > Deodorants > Combos > Hake Combos"] 25  
["Toys & School Supplies > School Supplies > Geometry & Pencil Boxes > Disney Geometry & Pencil Boxes"] 25  
["Beauty and Personal Care > Fragrances > Deodorants > Combos > Park Avenue Combos"] 26  
["Computers > Network Components > Routers > Netgear Routers"] 16  
["Computers > Network Components > Routers > Tenda Routers"] 28  
["Beauty and Personal Care > Fragrances > Deodorants > Combos > Foggy Combos"] 11  
["Toys & School Supplies > School Supplies > Geometry & Pencil Boxes > Klassik Geometry & Pencil Boxes"] 11  
["Home Improvement > Electricals > Switches & Accessories > Switches & Dimmers"] 29  
["Beauty and Personal Care > Fragrances > Deodorants > Combos > Cinthol Combos"] 10  
["Automotive > Car Accessories > Car Interior > Steering Covers"] 12  
["Beauty and Personal Care > Fragrances > Deodorants > Combos > EMV 1000 Combos"] 15  
["Beauty and Personal Care > Fragrances > Deodorants > Combos > Adidas Combos"] 12  
["Beauty and Personal Care > Fragrances > Deodorants > Combos > Denver Combos"] 39  
["Computers > Network Components > Routers > TP-LINK Routers"] 22  
["Toys & School Supplies > School Supplies > Geometry & Pencil Boxes > Neo Gold Leaf Geometry & Pencil Boxes"] 10  
["Beauty and Personal Care > Fragrances > Deodorants > Combos > Engage Combos"] 25  
["Toys & School Supplies > School Supplies > Geometry & Pencil Boxes > Karta Geometry & Pencil Boxes"] 13  
["Computers > Network Components > Routers > D-Link Routers"] 12  
["Beauty and Personal Care > Fragrances > Deodorants > Combos > Yardley Combos"] 27  
["Beauty and Personal Care > Fragrances > Deodorants > Combos > Aero Care Combos"] 14  
["Computers > Laptop Accessories > Adapters > Smartpro Adapters > Smartpro 19.5v,3.9a Replacement Charger or Wall..."] 20  
["Home Decor & Festive Needs > Wall Decor & Clocks > Clocks > Wall Clocks > Blacksmith Wall Clocks"] 11  
["Mobiles & Accessories > Tablet Accessories > Cases & Covers > Fastway Cases & Covers"] 11  
["Tools & Hardware > Tools > Gardening Tools > Plant Containers & Sets > Plant Container Sets"] 274  
["Computers > Laptop Accessories > Batteries > Laguard Batteries"] 31  
["Computers > Laptop Accessories > Batteries > Rega IT Batteries"] 10  
["Matches > Wrist Watches > Fasttrack Wrist Watches"] 20  
["Matches > Wrist Watches > Sonata Wrist Watches"] 66  
["Matches > Wrist Watches > Times Wrist Watches"] 23  
["Matches > Wrist Watches > Pippid Wrist Watches"] 16  
["Matches > Wrist Watches > Maxima Wrist Watches"] 73  
["Matches > Wrist Watches > Timewel Wrist Watches"] 30  
["Matches > Wrist Watches > MM Wrist Watches"] 11  
["Matches > Wrist Watches > Noise Wrist Watches"] 10  
["Matches > Wrist Watches > Gift Island Wrist Watches"] 16  
["Jewellery > Bangles, Bracelets & Armbands > Bangles"] 26  
["Automotive > Accessories & Spare Parts > Car Interior & Exterior > Car Interior > Car Sun Shades > Adro Exclusive Barbie Doll Sunshade & Curtains"] 34  
["Clothing > Kids' Clothing > Girls Wear > Dresses & Skirts > Dresses > Cucumber Dresses"] 10  
["Mobiles & Accessories > Tablet Accessories > Cases & Covers > BMS Lifestyle Cases & Covers"] 15  
["Clothing > Women's Clothing > Lingerie, Sleep & Swimwear > Bras > Clavia Bras"] 22  
["Clothing > Women's Clothing > Lingerie, Sleep & Swimwear > Bras > Bralux Bras"] 10  
["Automotive > Accessories & Spare Parts > Car Interior & Exterior > Car Exterior > Door & Bumper Guards"] 17  
["Footwear > Men's Clothing > T-Shirts > Mucode T-Shirts"] 16  
["Footwear > Men's Clothing > Casual & Party Wear Shirts > Marc N' Park Casual & Party Wear Shirts > Marc N' Park Men's Solid Casual Shirt"] 18  
["Jewellery > Accessories > Brooches > Aarav Art Silver Polished Brooch (Gold)"] 18  
["Automotive > Accessories & Spare Parts > Spares & Performance Parts > Car Spare Parts > Vehicle Horns"] 15  
["Home Decor & Festive Needs > Wall Decor & Clocks > Wall Decals & Stickers > DeStudio Wall Decals & Sticker a"] 47  
["Home Decor & Festive Needs > Wall Decor & Clocks > Wall Decals & Stickers > Creative Width Decor Wall Decals & Stickers"] 15  
["Furniture > Living Room > Sofa & Sectionals"] 55  
["Beauty and Personal Care > Fragrances > Deodorants > Combos > Wild Stone Combos"] 23  
["Clothing > Men's Clothing > Accessories & Combo Sets > Ties > Alvaro Ties"] 19  
["Jewellery > Nose Rings & Studs > Radhesh Creation Crystal Yellow Gold Plated Bras..."] 12  
["Clothing > Women's Clothing > Western Wear > Shirts, Tops & Tunics > Shirts > Bombay High Shirts"] 26  
["Home Furnishing > Cushions, Pillows & Covers > Cushion Covers"] 42  
["Beauty and Personal Care > Fragrances > Deodorants > Combos > Anasaal Shirts"] 14  
["Clothing > Women's Clothing > Western Wear > Shirts, Tops & Tunics > Shirts > Jazzy Ben Shirts"] 11  
["Clothing > Women's Clothing > Western Wear > Shirts, Tops & Tunics > Shirts > Jazzy Ben Shirts"] 11  
["Home Furnishing > Bath Linen > Towels"] 80  
["Clothing > Women's Clothing > Western Wear > Shirts, Tops & Tunics > Shirts > People Shirts"] 17  
["Clothing > Women's Clothing > Western Wear > Shirts, Tops & Tunics > Shirts > Tokyo Talies Shirts"] 19  
["Clothing > Women's Clothing > Western Wear > Shirts, Tops & Tunics > Shirts > My Addiction Shirts"] 12  
["Clothing > Women's Clothing > Western Wear > Shirts, Tops & Tunics > Shirts > Kiocha Shirts"] 14  
["Clothing > Women's Clothing > Western Wear > Shirts, Tops & Tunics > Shirts > Being Fab Shirts"] 13  
["Clothing > Women's Clothing > Western Wear > Shirts, Tops & Tunics > Shirts > India Inc Shirts"] 14  
["Mobiles & Accessories > Tablet Accessories > Cases & Covers > Entopia Cases & Covers"] 22  
["Bags, Wallets & Bags > Bags > Backpaks > Cases & Covers > RMO Cases & Covers"] 12  
["Mobiles & Accessories > Tablet Accessories > Cases & Covers > Rock Cases & Covers"] 10  
["Clothing > Women's Clothing > Western Wear > Jeans & Shorts > Jeans > GALLONN Jeans > GALLONN skinny Fit Women's Jeans"] 12  
["Clothing > Men's Clothing > T-Shirts > Divyon T-Shirts"] 61  
["Clothing > Men's Clothing > T-Shirts > Numero Uno T-Shirts"] 119  
["Clothing > Men's Clothing > T-Shirts > Orange and Orchid T-Shirts"] 23  
["Clothing > Men's Clothing > T-Shirts > Khan T-Shirts"] 60  
["Clothing > Men's Clothing > T-Shirts > Northern Lights T-Shirts"] 16  
["Clothing > Men's Clothing > T-Shirts > Rod T-Shirts"] 11  
["Clothing > Men's Clothing > T-Shirts > Nimya T-Shirts"] 41  
["Clothing > Men's Clothing > T-Shirts > Horwood T-Shirts"] 12  
["Clothing > Men's Clothing > T-Shirts > Rivia T-Shirts"] 11  
["Clothing > Men's Clothing > T-Shirts > Ninja Turtle T-Shirts"] 14  
["Clothing > Men's Clothing > T-Shirts > Ocean Race T-Shirts"] 28  
["Clothing > Men's Clothing > T-Shirts > Kimbus T-Shirts"] 14  
["Matches > Wrist Watches > Luba Wrist Watches"] 11  
["Kitchen & Dining > Containers & Bottles > Bottles & Slippers > Water bottle > Nutcase Water bottle"] 17  
["Jewellery > Earrings"] 83  
["Clothing > Women's Clothing > Western Wear > Leggings & Jeggings > Leggings & Jeggings > NF Leggings & Jeggings"] 24  
["Beauty and Personal Care > Combs and Kits > Combs > Wild Stone Combs"] 38  
["Beauty and Personal Care > Combs and Kits > Moxfere Combs and Kits"] 31  
["Matches > Wrist Watches > Perucci Wrist Watches"] 20  
["Computers > Women's Clothing > Ethnic Wear > Kurtas & Kurtis > Kurtas > Parijat Kurtas"] 13  
["Computers > Laptop Accessories > Adapters > Rega IT Adapters"] 11  
["Jewellery > Pendants & Lockets"] 16  
["Fens & Stationery > Office Supplies > Punches & Punching Machines > Toopado Punches & Punching Machines"] 14  
["Footwear > Men's Footwear > Casual Shoes > Zoot24 Casual Shoes"] 13  
["Furniture > Pet Furniture"] 76  
["Footwear > Men's Footwear > Casual Shoes > Arden Casual Shoes"] 12  
["Footwear > Men's Footwear > Casual Shoes > Knotty Derby Casual Shoes"] 15  
["Tools & Hardware > Tools > Tool Accessories > Safety Gear > Safety Helmets"] 14  
["Cameras & Accessories > Camera Accessories > Lens Hoods > JJC Lens Hoods"] 41  
["Footwear > Men's Footwear > Casual Shoes > Antoreindis Casual Shoes"] 10  
["Tools & Hardware > Tools > Tool Accessories > Screwdriver Bit Sets"] 11  
["Computers > Laptop Accessories > USB Gadgets > Teatronics USB Gadgets"] 20  
["Sports & Fitness > Team Sports > Basketball > Basketball > Ripsta Basketball"] 10  
["Clothing > Women's Clothing > Ethnic Wear > Kurtas & Kurtis > Kurtis > Shopping Karega Kurtis"] 11  
["Clothing > Women's Clothing > Ethnic Wear > Kurtas & Kurtis > Kurtis > Shop Karega Kurtis"] 15  
["Home Furnishing > Living"] 14  
["Fens & Stationery > Office Supplies > AMT Office Supplies"] 17  
["Beauty and Personal Care > Pyjamas & Pajamas > Pyjamas > Antashrike Pyjamas & Lounge Pants > Antashrike Men's Pyjama (Pack of 3)"] 11  
["Home Decor & Festive Needs > Decorative Lighting & Lamps > Series Lights > Dilight Series Lights"] 27  
["Clothing > Women's Clothing > Lingerie, Sleep & Swimwear > Night Dresses & Nighties > Simrit Night Dresses & Nighties > Simrit Women's Nighty (2 Nighties)"] 1  
["Clothing > Women's Clothing > Western Wear > Dresses & Skirts > Dresses > LiENE Camisoles & Slips > LiENE Women's Camisole"] 21  
["Home Furnishing > Floor Coverings > Carpets & Rugs > Lookmat Ethnic Cotton Durrrie (120 cm X 18...") 12  
["Clothing > Women's Clothing > Western Wear > Shirts, Tops & Tunics > Shirts > Anasaal Shirts"] 17  
["Clothing > Women's Clothing > Western Wear > Dresses & Skirts > Dresses > Meiro Dresses"] 13  
["Clothing > Women's Clothing > Western Wear > Dresses & Skirts > Dresses > Indi Bargain Dresses"] 15  
["Clothing > Women's Clothing > Western Wear > Dresses & Skirts > Dresses > Ladybug Dresses"] 15  
["Clothing > Women's Clothing > Western Wear > Dresses & Skirts > Dresses > Hugo Chavez Dresses"] 20  
["Clothing > Women's Clothing > Western Wear > Dresses & Skirts > Dresses > Miss Chase Dresses"] 19  
["Clothing > Women's Clothing > Western Wear > Dresses & Skirts > Dresses > Latin Quarters Dresses"] 1  
["Baby Care > Baby & Kids Gifts > Stickers > Wallmanta Stickers"] 38  
["Baby Care > Baby & Kids Gifts > Stickers > WallDesign Stickers"] 14

In [28]:  
primary\_category = products["product\_category\_tree"].tolist()  
  
#overprocessing the categories to prevent duplicate categories and removal of punctuations like ">", etc  
for i in range(len(primary\_category)):  
 primary\_category[i] = primary\_category[i].lower()  
 primary\_category[i] = primary\_category[i].replace('(','')  
 primary\_category[i] = primary\_category[i].replace(')','')  
 primary\_category[i] = primary\_category[i].replace(' ','')  
 primary\_category[i] = primary\_category[i].split('>')  
 primary\_category[i] = primary\_category[i][0]  
  
print(primary\_category)  
print(len(primary\_category))



[illegible]







[illegible]

```
Out[29]: Text(0.5, 1.0, 'Number of products in categories v/s Category')

Out[29]: Text(0.5, 0, 'Product Categories')

Out[29]: Text(0, 0.5, 'Frequency')

Out[29]: (array([5., 0.2, 0.4, 0.6, 0.8, 1. ]),
          [Text(0, 0, ''),
           Text(0, 0, ''),
           Text(0, 0, ''),
           Text(0, 0, ''),
           Text(0, 0, ''),
           Text(0, 0, '')])

Out[29]: (array([5., 0.2, 0.4, 0.6, 0.8, 1. ]),
          [Text(0, 0, ''),
           Text(0, 0, ''),
           Text(0, 0, ''),
           Text(0, 0, ''),
           Text(0, 0, ''),
           Text(0, 0, '')])
```

```

Text(0, 0, ''),
Text(0, 0, ''),
Text(0, 0, ''),
Text(0, 0, ''),
Text(0, 0, ''),
Text(0, 0, '')]

Out[28]: {'AxesSubplot': ('center': 'Number of products in categories w/s Category'), 'xlabel': 'Product Categories', 'ylabel': 'Frequency'}

Number of products in categories w/s Category

4000
3000
2000
1000
0
-1000
-2000
-3000
-4000
-5000
-6000
-7000
-8000
-9000
-10000
-11000
-12000
-13000
-14000
-15000
-16000
-17000
-18000
-19000
-20000
-21000
-22000
-23000
-24000
-25000
-26000
-27000
-28000
-29000
-30000
-31000
-32000
-33000
-34000
-35000
-36000
-37000
-38000
-39000
-40000
-41000
-42000
-43000
-44000
-45000
-46000
-47000
-48000
-49000
-50000
-51000
-52000
-53000
-54000
-55000
-56000
-57000
-58000
-59000
-60000
-61000
-62000
-63000
-64000
-65000
-66000
-67000
-68000
-69000
-70000
-71000
-72000
-73000
-74000
-75000
-76000
-77000
-78000
-79000
-80000
-81000
-82000
-83000
-84000
-85000
-86000
-87000
-88000
-89000
-90000
-91000
-92000
-93000
-94000
-95000
-96000
-97000
-98000
-99000
100000
101000
102000
103000
104000
105000
106000
107000
108000
109000
110000
111000
112000
113000
114000
115000
116000
117000
118000
119000
120000
121000
122000
123000
124000
125000
126000
127000
128000
129000
130000
131000
132000
133000
134000
135000
136000
137000
138000
139000
140000
141000
142000
143000
144000
145000
146000
147000
148000
149000
150000
151000
152000
153000
154000
155000
156000
157000
158000
159000
160000
161000
162000
163000
164000
165000
166000
167000
168000
169000
170000
171000
172000
173000
174000
175000
176000
177000
178000
179000
180000
181000
182000
183000
184000
185000
186000
187000
188000
189000
190000
191000
192000
193000
194000
195000
196000
197000
198000
199000
200000
201000
202000
203000
204000
205000
206000
207000
208000
209000
210000
211000
212000
213000
214000
215000
216000
217000
218000
219000
220000
221000
222000
223000
224000
225000
226000
227000
228000
229000
230000
231000
232000
233000
234000
235000
236000
237000
238000
239000
240000
241000
242000
243000
244000
245000
246000
247000
248000
249000
250000
251000
252000
253000
254000
255000
256000
257000
258000
259000
260000
261000
262000
263000
264000
265000
266000
267000
268000
269000
270000
271000
272000
273000
274000
275000
276000
277000
278000
279000
280000
281000
282000
283000
284000
285000
286000
287000
288000
289000
290000
291000
292000
293000
294000
295000
296000
297000
298000
299000
300000
301000
302000
303000
304000
305000
306000
307000
308000
309000
310000
311000
312000
313000
314000
315000
316000
317000
318000
319000
320000
321000
322000
323000
324000
325000
326000
327000
328000
329000
330000
331000
332000
333000
334000
335000
336000
337000
338000
339000
340000
341000
342000
343000
344000
345000
346000
347000
348000
349000
350000
351000
352000
353000
354000
355000
356000
357000
358000
359000
360000
361000
362000
363000
364000
365000
366000
367000
368000
369000
370000
371000
372000
373000
374000
375000
376000
377000
378000
379000
380000
381000
382000
383000
384000
385000
386000
387000
388000
389000
390000
391000
392000
393000
394000
395000
396000
397000
398000
399000
400000
401000
402000
403000
404000
405000
406000
407000
408000
409000
410000
411000
412000
413000
414000
415000
416000
417000
418000
419000
420000
421000
422000
423000
424000
425000
426000
427000
428000
429000
430000
431000
432000
433000
434000
435000
436000
437000
438000
439000
440000
441000
442000
443000
444000
445000
446000
447000
448000
449000
450000
451000
452000
453000
454000
455000
456000
457000
458000
459000
460000
461000
462000
463000
464000
465000
466000
467000
468000
469000
470000
471000
472000
473000
474000
475000
476000
477000
478000
479000
480000
481000
482000
483000
484000
485000
486000
487000
488000
489000
490000
491000
492000
493000
494000
495000
496000
497000
498000
499000
500000
501000
502000
503000
504000
505000
506000
507000
508000
509000
510000
511000
512000
513000
514000
515000
516000
517000
518000
519000
520000
521000
522000
523000
524000
525000
526000
527000
528000
529000
530000
531000
532000
533000
534000
535000
536000
537000
538000
539000
540000
541000
542000
543000
544000
545000
546000
547000
548000
549000
550000
551000
552000
553000
554000
555000
556000
557000
558000
559000
560000
561000
562000
563000
564000
565000
566000
567000
568000
569
```

```
In [30]: all_categories=[]
         count={}

         for keys,vals in primary_category_count.items():
             #excluding only those categories which have at least 10 products
             if (vals>=10):
```

```

all_categories.append(keys)
count.append(vals)

plt.figure(figsize=(40,6))
plt.title("Number of products in categories v/s Category", fontsize=30)
plt.xlabel("Product Categories having frequency at least 10")
plt.ylabel("Frequency")
plt.xticks(fontsize=11, rotation=30)
plt.yticks(fontsize=18)
sns.barplot(x=all_categories, y=count)

Out[30]: <Figure size 2880x432 with 0 Axes>

Out[30]: Text(0.5, 1.0, 'Number of products in categories v/s Category')
```

```
Out[30]: Text(0.5, 0, 'Product Categories having frequency at least 10')

Out[30]: Text(0, 0.5, 'Frequency')

Out[30]: (array([0. , 0.2, 0.4, 0.6, 0.8, 1. ]),
          [Text(0, 0, ''),
           Text(0, 0, ''),
           Text(0, 0, ''),
           Text(0, 0, ''),
           Text(0, 0, ''),
           Text(0, 0, '')])

Out[30]: (array([0. , 0.2, 0.4, 0.6, 0.8, 1. ]),
          [Text(0, 0, ''),
           Text(0, 0, ''),
           Text(0, 0, ''),
           Text(0, 0, ''),
           Text(0, 0, ''),
           Text(0, 0, '')])
```

```

Out[30]: <AxesSubplot: {'center': 'Number of products in categories w/s Category', 'label': 'Product Categories having frequency at least 10', 'ylabel': 'Frequency'}>

```

```

In [31]: revised_categories = [None]*len(primary_category)

for i in range(len(primary_category)):
    if primary_category[i]=="homeimprovement" or primary_category[i]=="homedecorfestiveneeds" or primary_category[i]=="homefurnishings/kitchen":
        revised_categories[i] = "homefurnishings/kitchen"
    elif primary_category[i] == "pet/stationery":
        revised_categories[i] = "toys/schoolsupplies"
    elif primary_category[i] == "gaming" or primary_category[i] == "cameras/accessories" or primary_category[i] == "":
        revised_categories[i] = "electronics"
    elif primary_category[i] == "health/personal care appliances" or primary_category[i] == "beautyand/personalcare":
        revised_categories[i] = "personalcare/accessories"
    else:

```

```

        revised_categories[i] = primary_category[i]

print("The length of the revised grouped categories are {}".format(len(Counter(revised_categories))))
print(revised_categories)

```



[illegible]



[illegible]



[illegible]

7	FabHomeDecor Fabric Double Sofa Bed	["Furniture >> Living Room Furniture >> Sofa B...	FabHomeDecor Fabric Double Sofa Bed (Finish Co...	FabHomeDecor	homefurnishing/kitchen
8	dilli bazaar Belies, Corporate Casuals, Casuals	["Footwear >> Women's Footwear >> Ballerinas >...	Key Features of dilli bazaar Belies, Corpora...	dilli bazaar	footwear
9	Alisha Solid Women's Cycling Shorts	["Clothing >> Women's Clothing >> Lingerie, SL...	Key Features of Alisha Solid Women's Cycling S...	Alisha	clothing
10	Ladela Belies	["Footwear >> Women's Footwear >> Ballerinas >...	Key Features of Ladela Belies Brand: LADSLA C...	Ladela	footwear
11	Carrel Printed Women's	["Clothing >> Women's Clothing >> Sports & Gym...	Specifications of Carrel Printed Women's Fabric...	Carrel	clothing
12	Sicons All Purpose Tea Tree Dog Shampoo	["Pet Supplies >> Grooming >> Skin & Coat Care...	Specifications of Sicons All Purpose Tea Tree ...	Sicons	petsupplies
13	Alisha Solid Women's Cycling Shorts	["Clothing >> Women's Clothing >> Lingerie, SL...	Key Features of Alisha Solid Women's Cycling S...	Alisha	clothing
14	Freelance Vacuum Bottles 350 ml Bottle	["Pens & Stationery >> School Supplies >> Wate...	Specifications of Freelance Vacuum Bottles 350...	Freelance	toys&schoolsupplies
15	Alisha Solid Women's Cycling Shorts	["Clothing >> Women's Clothing >> Lingerie, SL...	Key Features of Alisha Solid Women's Cycling S...	Alisha	clothing
16	FabHomeDecor Fabric Double Sofa Bed	["Furniture >> Living Room Furniture >> Sofa B...	FabHomeDecor Fabric Double Sofa Bed (Finish Co...	FabHomeDecor	homefurnishing/kitchen
17	Style Foot Belies	["Footwear >> Women's Footwear >> Ballerinas >...	Key Features of Style Foot Belies ballerina s...	Style Foot	footwear
18	Carrel Printed Women's	["Clothing >> Women's Clothing >> Sports & Gym...	Key Features of Carrel Printed Women's Fabric...	Carrel	clothing
19	FabHomeDecor Fabric Double Sofa Bed	["Furniture >> Living Room Furniture >> Sofa B...	FabHomeDecor Fabric Double Sofa Bed (Finish Co...	FabHomeDecor	homefurnishing/kitchen

In (33):

```

temp_categories = list(dict.fromkeys(list(products.loc[:, 'primary_categories'].values)))

considered_category = []
revised_category_count = {}
noise = []
all_categories = {}

print("\n")
print("***** THE FINAL CONSIDERED CATEGORIES AND THEIR COUNTS ARE AS FOLLOWS *****")
print("\n")

for category in temp_categories:
    count = np.sum(products.loc[:, 'primary_categories'].values == category)
    all_categories[category]=count
    #a category is considered to be a main category only if the frequency of a product in that category is at
    if (count>=10):
        print(category,count)
        revised_category_count[category] = count
    else:
        noise[category] = count

for i in range(len(revised_category_count)):
    if revised_category_count[i] in revised_category_count:
        considered_category.append(i)
    else:
        considered_category.append(0)

print("\n")
print("***** THE NOISE IN THE DATASET IS AS FOLLOWS *****")
print("\n")
for keys,vals in noise.items():
    print(keys,vals)

print("\n")
print("***** THE NOISE + PRIMARY CATEGORIES IN THE DATASET ARE AS FOLLOWS *****")
print("\n")
for keys,vals in all_categories.items():
    print(keys,vals)

```



