In [1]:

```
import pandas as pd
import numpy as np
```

In [2]:

```
df=pd.read_excel('/Users/nnkundan/Downloads/Savanta Interview Task - Prin
                 sheet_name='Data for Assessment Task')
```

In [3]:

```
df.head()
```

Out[3]:

	ResponseID	Gender	Age	dAge	Employment_status	Relationship_status	Living_stat
0	1	0	42	3	1	2	
1	2	1	51	4	1	2	
2	3	0	29	2	1	1	
3	4	1	62	5	1	2	
4	5	0	36	3	1	1	

5 rows × 165 columns

In [4]:

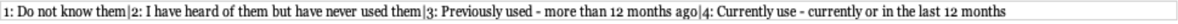
```
df.shape
```

Out[4]:

```
(73170, 165)
```

Brand Columns Analysis

1: Do not know them|2: I have heard of them but have never used them|3: Previously used - more than 12 months ago|4: Currently use - currently or in the last 12 months



1: Do not know them|2: I have heard of them but have never used them|3: Previously used - more than 12 months ago|4: Currently use - currently or in the last 12 months

1. Brand_4

Method 1: By using pandasql

In [21]:

```
val=df['Brand_4'].value_counts().sort_values(ascending=False)
val
```

Out[21]:

```
1.0    7297
2.0    3847
3.0     966
4.0     749
Name: Brand_4, dtype: int64
```

In [23]:

```
per = (val/sum(val))*100
per
```

Out[23]:

```
1.0    56.746248
2.0    29.916790
3.0     7.512248
4.0     5.824714
Name: Brand_4, dtype: float64
```

In [24]:

```
tot=sum(val)
```

In [29]:

```
df_brand_4 = pd.DataFrame({'Count':val,'Percentage':per,'Total Count':tot
df_brand_4.reset_index(inplace=True)
df_brand_4.rename(columns={'index': 'Brand_4'}, inplace=True)
df_brand_4
```

Out[29]:

	Brand_4	Count	Percentage	Total Count
0	1.0	7297	56.746248	12859
1	2.0	3847	29.916790	12859
2	3.0	966	7.512248	12859
3	4.0	749	5.824714	12859

Method 2: By using pandasql

In [30]:

```
from pandasql import sqldf
```

In [32]:

```
A=sqldf("select * from df",locals())
```

In [68]:

```
Br4= sqldf("SELECT Brand_4,COUNT(*) AS Count_Brand_4," \
          "(COUNT(*) * 100.0 / SUM(COUNT(*)) OVER ()) AS Percentage_4, (SU
    "FROM A where Brand_4 not like 'NA' GROUP BY Brand_4 order by Percent
Br4
```

Out[68]:

	Brand_4	Count_Brand_4	Percentage_4	total_Count
0	1.0	7297	56.746248	12859
1	2.0	3847	29.916790	12859
2	3.0	966	7.512248	12859
3	4.0	749	5.824714	12859

2. Brand_13

In [69]:

```
Br13= sqldf("SELECT Brand_13,COUNT(*) AS Count_Brand_13," \
         "(COUNT(*) * 100.0 / SUM(COUNT(*)) OVER ()) AS Percentage_13, (S
    "FROM A where Brand_13 not like 'NA' GROUP BY Brand_13 order by Perce
Br13
```

Out[69]:

	Brand_13	Count_Brand_13	Percentage_13	total_Count
0	2.0	6406	49.805629	12862
1	3.0	3015	23.441144	12862
2	4.0	2149	16.708132	12862
3	1.0	1292	10.045094	12862

3. Brand_82

In [70]:

```
Br82= sqldf("SELECT Brand_82,COUNT(*) AS Count_Brand_82," \
         "(COUNT(*) * 100.0 / SUM(COUNT(*)) OVER ()) AS Percentage_82, (S
    "FROM A where Brand_82 not like 'NA' GROUP BY Brand_82 order by Perce
Br82
```

Out[70]:

	Brand_82	Count_Brand_82	Percentage_82	total_Count
0	2.0	6763	52.630350	12850
1	3.0	2471	19.229572	12850
2	1.0	1912	14.879377	12850
3	4.0	1704	13.260700	12850

4. Brand_124

In [71]:

```
Br124= sqldf("SELECT Brand_124,COUNT(*) AS Count_Brand_124," \
         "(COUNT(*) * 100.0 / SUM(COUNT(*)) OVER ()) AS Percentage_124, (
    "FROM A where Brand_124 not like 'NA' GROUP BY Brand_124 order by Per
Br124
```

Out[71]:

	Brand_124	Count_Brand_124	Percentage_124	total_Count
0	2.0	6307	48.929403	12890
1	1.0	4330	33.591932	12890
2	3.0	1192	9.247479	12890
3	4.0	1061	8.231187	12890

5. Brand_128

In [72]:

```
Br128= sqldf("SELECT Brand_128,COUNT(*) AS Count_Brand_128," \
         "(COUNT(*) * 100.0 / SUM(COUNT(*)) OVER ()) AS Percentage_128, (
    "FROM A where Brand_128 not like 'NA' GROUP BY Brand_128 order by Per
Br128
```

Out[72]:

	Brand_128	Count_Brand_128	Percentage_128	total_Count
0	2.0	7177	55.735031	12877
1	1.0	3044	23.639046	12877
2	3.0	1511	11.734100	12877
3	4.0	1145	8.891823	12877

6. Brand_131

In [73]:

```
Br131= sqldf("SELECT Brand_131,COUNT(*) AS Count_Brand_131," \
         "(COUNT(*) * 100.0 / SUM(COUNT(*)) OVER ()) AS Percentage_131, (
    "FROM A where Brand_131 not like 'NA' GROUP BY Brand_131 order by Per
Br131
```

Out[73]:

	Brand_131	Count_Brand_131	Percentage_131	total_Count
0	2.0	6484	50.470927	12847
1	1.0	4010	31.213513	12847
2	3.0	1453	11.310033	12847
3	4.0	900	7.005527	12847

7. Brand_374

In [74]:

```
Br374= sqldf("SELECT Brand_374,COUNT(*) AS Count_Brand_374," \
         "(COUNT(*) * 100.0 / SUM(COUNT(*)) OVER ()) AS Percentage_374, (
    "FROM A where Brand_374 not like 'NA' GROUP BY Brand_374 order by Per
Br374
```

Out[74]:

	Brand_374	Count_Brand_374	Percentage_374	total_Count
0	1.0	3447	53.006305	6503
1	2.0	2423	37.259726	6503
2	3.0	336	5.166846	6503
3	4.0	297	4.567123	6503

In [298…

```
BrandCombined= sqldf("SELECT Brand_4 as Brand_class,Percentage_4,Percenta
            "Percentage_128,Percentage_131,Percentage_374 from "
"Br4 inner join Br13 on Br4.Brand_4=Br13.Brand_13"
" inner join Br82 on Br13.Brand_13=Br82.Brand_82"
" inner join Br124 on Br82.Brand_82=Br124.Brand_124"
" inner join Br128 on Br124.Brand_124=Br128.Brand_128"
" inner join Br131 on Br128.Brand_128=Br131.Brand_131"
" inner join Br374 on Br131.Brand_131=Br374.Brand_374"
,locals())
BrandCombined
```

Out[298]:

	Brand_class	Percentage_4	Percentage_13	Percentage_82	Percentage_124	Percent
0	1.0	56.746248	10.045094	14.879377	33.591932	23
1	2.0	29.916790	49.805629	52.630350	48.929403	5!
2	3.0	7.512248	23.441144	19.229572	9.247479	1
3	4.0	5.824714	16.708132	13.260700	8.231187	{

In [299…

```
BrandCombined1=BrandCombined.T
BrandCombined2=BrandCombined1.reset_index()
BrandCombined2.columns=BrandCombined2.iloc[0]
BrandCombined3=BrandCombined2.drop(0)
BrandCombined3.rename(columns={1.0:'1',2.0:'2',3.0:'3',4.0:'4'},inplace=T
BrandCombined3=BrandCombined3.round(2)
BrandCombined3
```

Out[299]:

	Brand_class	1	2	3	4
1	Percentage_4	56.75	29.92	7.51	5.82
2	Percentage_13	10.05	49.81	23.44	16.71
3	Percentage_82	14.88	52.63	19.23	13.26
4	Percentage_124	33.59	48.93	9.25	8.23
5	Percentage_128	23.64	55.74	11.73	8.89
6	Percentage_131	31.21	50.47	11.31	7.01
7	Percentage_374	53.01	37.26	5.17	4.57

In [365…

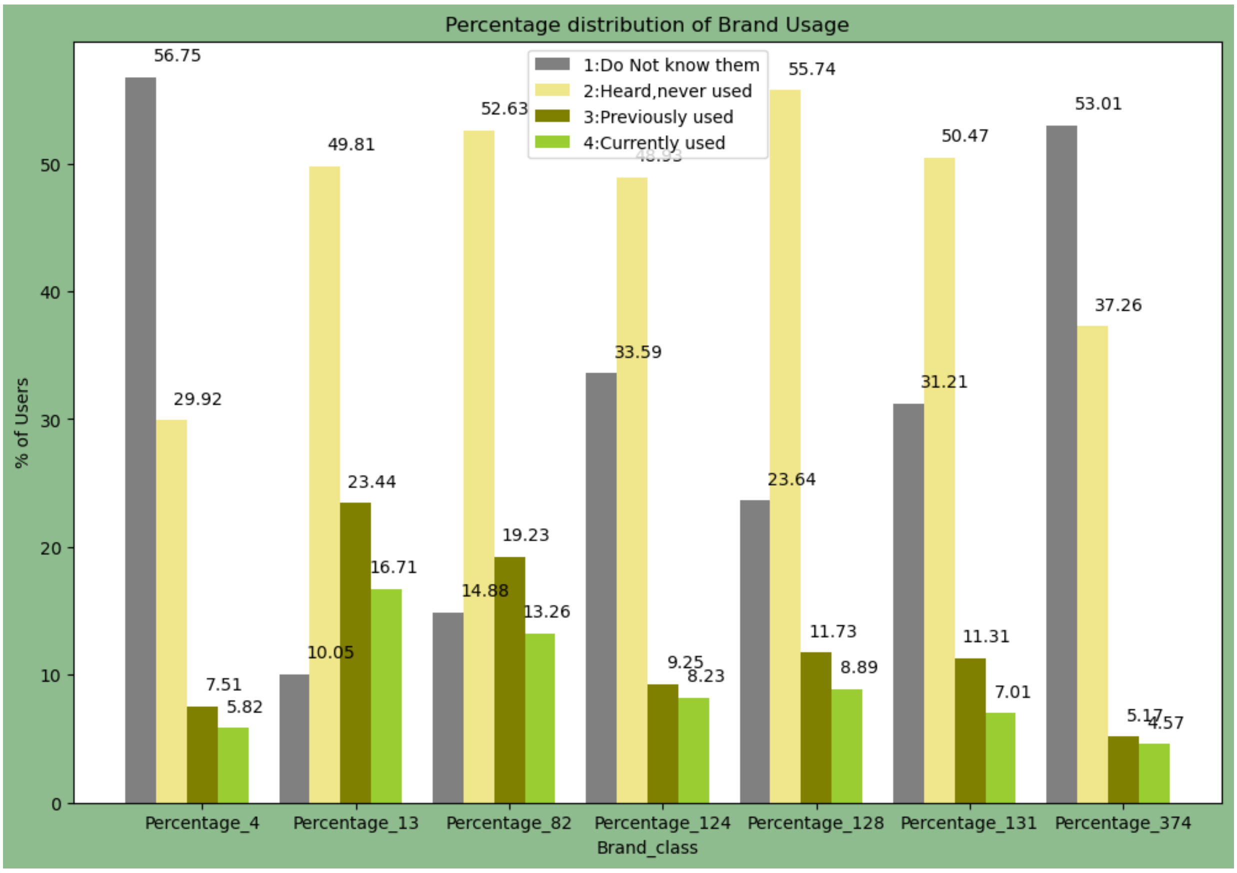
```
BrandCombined3.to_csv('usage.csv',index=False)
```

In [300…

```
bc3=BrandCombined3
```

In [464…

```
X_axis = np.arange(len(bc3['Brand_class']))

plt.figure(figsize=(12,8),facecolor='darkseagreen')
plt.bar(X_axis - 0.4, bc3['1'], 0.20, label = '1:Do Not know them',color=
plt.bar(X_axis - 0.2, bc3['2'], 0.20, label = '2:Heard,never used',color=
plt.bar(X_axis + 0.0, bc3['3'], 0.20, label = '3:Previously used',color='
plt.bar(X_axis + 0.2, bc3['4'], 0.20, label = '4:Currently used',color='y


# Add label values to the bars
bar_width = 0.135


for i, value in enumerate(bc3['1']):
    plt.text(i, value + 1, str(value), ha='right', va='bottom')
for i, value in enumerate(bc3['2']):
    plt.text(i + bar_width, value + 1, str(value), ha='right', va='bottom
for i, value in enumerate(bc3['3']):
    plt.text(i + bar_width+bar_width, value + 1, str(value), ha='right', 
for i, value in enumerate(bc3['4']):
    plt.text(i + bar_width+bar_width+bar_width, value + 1, str(value), ha




# Set labels and title
plt.xticks(X_axis, bc3['Brand_class'])
plt.xlabel("Brand_class")
plt.ylabel("% of Users")
plt.title("Percentage distribution of Brand Usage")
plt.legend()
plt.show()

# Saving the figure.
plt.savefig("output1.jpg")
```



```
<Figure size 640x480 with 0 Axes>
```

In [ ]:

In [ ]:

In [ ]:

Consider Columns Analysis

In [160…

```
df_c=df
```

In [161…

```
df_c=df_c.iloc[:,17:24]
df_c
```

Out[161]:

	Consider_general_4	Consider_general_13	Consider_general_82	Consider_genera
0	NaN	1.0	NaN	
1	NaN	NaN	NaN	
2	NaN	NaN	NaN	
3	NaN	NaN	NaN	
4	NaN	NaN	NaN	
...	...	...	...	
73165	NaN	NaN	NaN	
73166	NaN	NaN	NaN	
73167	NaN	NaN	NaN	
73168	NaN	NaN	NaN	
73169	NaN	NaN	NaN	

73170 rows × 7 columns

In [162…

```
B=sqldf("select * from df_c",locals())
```

1. Consider_general_4

In [178…

```
Cr4=sqldf("Select Consider_general_4, "
      " count(*) as Count_of_Consider_general_4, (count(*)*100.0/sum(coun
      ", (sum(count(*)) over()) as total_count "
      "from B where Consider_general_4 not like 'NA'"
      " group by Consider_general_4",locals())
Cr4
# 0: Not Selected|1: Brand_4
```

Out[178]:

	Consider_general_4	Count_of_Consider_general_4	Percentage_4	total_count
0	0.0	4552	81.841064	5562
1	1.0	1010	18.158936	5562

2. Consider_general_13

In [180…

```
Cr13=sqldf("Select Consider_general_13, "
      " count(*) as Count_of_Consider_general_13, (count(*)*100.0/sum(cou
      ", (sum(count(*)) over()) as total_count "
      "from B where Consider_general_13 not like 'NA'"
      " group by Consider_general_13",locals())
Cr13
# 0: Not Selected|1: Brand_13
```

Out[180]:

	Consider_general_13	Count_of_Consider_general_13	Percentage_13	total_count
0	0.0	6767	58.487468	11570
1	1.0	4803	41.512532	11570

3. Consider_general_82

In [181…

```
Cr82=sqldf("Select Consider_general_82, "
      " count(*) as Count_of_Consider_general_82, (count(*)*100.0/sum(cou
      ", (sum(count(*)) over()) as total_count "
      "from B where Consider_general_82 not like 'NA'"
      " group by Consider_general_82",locals())
Cr82
# 0: Not Selected|1: Brand_82
```

Out[181]:

	Consider_general_82	Count_of_Consider_general_82	Percentage_82	total_count
0	0.0	6646	60.760651	10938
1	1.0	4292	39.239349	10938

4. Consider_general_124

In [182…

```
Cr124=sqldf("Select Consider_general_124, "
      " count(*) as Count_of_Consider_general_124, (count(*)*100.0/sum(co
      ", (sum(count(*)) over()) as total_count "
      "from B where Consider_general_124 not like 'NA'"
      " group by Consider_general_124",locals())
Cr124
# 0: Not Selected|1: Brand_124
```

Out[182]:

	Consider_general_124	Count_of_Consider_general_124	Percentage_124	total_count
0	0.0	6476	75.654206	8560
1	1.0	2084	24.345794	8560

5. Consider_general_128

In [183…

```
Cr128=sqldf("Select Consider_general_128, "
      " count(*) as Count_of_Consider_general_128, (count(*)*100.0/sum(co
      ", (sum(count(*)) over()) as total_count "
      "from B where Consider_general_128 not like 'NA'"
      " group by Consider_general_128",locals())
Cr128
# 0: Not Selected|1: Brand_128
```

Out[183]:

	Consider_general_128	Count_of_Consider_general_128	Percentage_128	total_count
0	0.0	7339	74.636428	9833
1	1.0	2494	25.363572	9833

6. Consider_general_131

In [184…

```
Cr131=sqldf("Select Consider_general_131, "
     " count(*) as Count_of_Consider_general_131, (count(*)*100.0/sum(co
     ", (sum(count(*)) over()) as total_count "
     "from B where Consider_general_131 not like 'NA'"
     " group by Consider_general_131",locals())
Cr131
# 0: Not Selected|1: Brand_131
```

Out[184]:

	Consider_general_131	Count_of_Consider_general_131	Percentage_131	total_count
0	0.0	6829	77.277357	8837
1	1.0	2008	22.722643	8837

7. Consider_general_374

In [185…

```
Cr374=sqldf("Select Consider_general_374, "
     " count(*) as Count_of_Consider_general_374, (count(*)*100.0/sum(co
     ", (sum(count(*)) over()) as total_count "
     "from B where Consider_general_374 not like 'NA'"
     " group by Consider_general_374",locals())
Cr374
# 0: Not Selected|1: Brand_374
```

Out[185]:

	Consider_general_374	Count_of_Consider_general_374	Percentage_374	total_count
0	0.0	2421	79.221204	3056
1	1.0	635	20.778796	3056

Combined

In [192…

```
cb=sqldf("Select Consider_general_4 as Consider_type, Percentage_4, Perce
         ",Percentage_124,Percentage_128,Percentage_131,Percentage_374 "


         " from Cr4 inner join Cr13 on Cr4.Consider_general_4=Cr13.Consid
         " inner join Cr82 on Cr13.Consider_general_13=Cr82.Consider_gene
         " inner join Cr124 on Cr82.Consider_general_82=Cr124.Consider_ge
         " inner join Cr128 on Cr124.Consider_general_124=Cr128.Consider_
         " inner join Cr131 on Cr128.Consider_general_128=Cr131.Consider_
         " inner join Cr374 on Cr131.Consider_general_131=Cr374.Consider_






         ,locals())
cb
```

Out[192]:

	Consider_type	Percentage_4	Percentage_13	Percentage_82	Percentage_124	Perce
0	0.0	81.841064	58.487468	60.760651	75.654206	
1	1.0	18.158936	41.512532	39.239349	24.345794	

In [193…

```
type(cb)
```

Out[193]:

```
pandas.core.frame.DataFrame
```

In [218…

```
cb1=cb.T
cb1
```

Out[218]:

	0	1
Consider_type	0.000000	1.000000
Percentage_4	81.841064	18.158936
Percentage_13	58.487468	41.512532
Percentage_82	60.760651	39.239349
Percentage_124	75.654206	24.345794
Percentage_128	74.636428	25.363572
Percentage_131	77.277357	22.722643
Percentage_374	79.221204	20.778796

In [263…

```
import matplotlib.pyplot as plt
import numpy as np
```

In [265…

```
cb2=cb1.reset_index()
cb2.columns=cb2.iloc[0]
cb2=cb2.drop(0)
cb2.rename(columns={0.0:'0:Not Selected',1.0:'1:Considered'},inplace=True
cb2=cb2.round(2)
cb2
```

Out[265]:

	Consider_type	0:Not Selected	1:Considered
1	Percentage_4	81.84	18.16
2	Percentage_13	58.49	41.51
3	Percentage_82	60.76	39.24
4	Percentage_124	75.65	24.35
5	Percentage_128	74.64	25.36
6	Percentage_131	77.28	22.72
7	Percentage_374	79.22	20.78

In [368…

```
cb2.to_csv('brand_consideration.csv',index=False)
```

In [465…

```
X_axis = np.arange(len(cb2['Consider_type']))

plt.figure(figsize=(16,7),facecolor='darkseagreen')
plt.bar(X_axis - 0.2, cb2['0:Not Selected'], 0.4, label = '0:Not Selected
plt.bar(X_axis + 0.2, cb2['1:Considered'], 0.4, label = '1:Considered',co

# Add label values to the bars
bar_width = 0.35


for i, value in enumerate(cb2['0:Not Selected']):
    plt.text(i, value + 1, str(value), ha='center', va='bottom')
for i, value in enumerate(cb2['1:Considered']):
    plt.text(i + bar_width, value + 1, str(value), ha='center', va='botto



# Set labels and title
plt.xticks(X_axis, cb2['Consider_type'])
plt.xlabel("Consider_type")
plt.ylabel("% of users")
plt.title("Percentage distribution of Brand Consideration")
plt.legend()
plt.show()
```

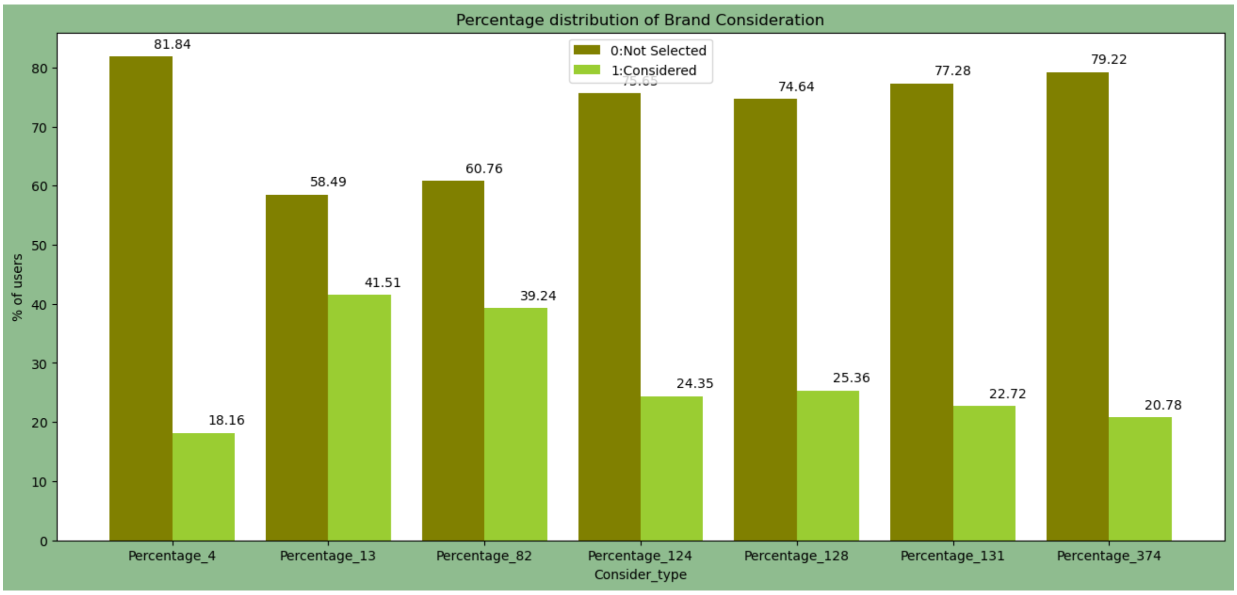


Image Analysis

In [468…

```
im_4=df
cols=im_4.columns.tolist()
```

In [469…

```
#image_brnd_4
my_new_dict={}
my_new_dict['0']='None'
my_new_dict['1']='Trustworthy'
my_new_dict['2']='Knowledgeble'
my_new_dict['3']='Friendly'
my_new_dict['4']='Expert'
my_new_dict['5']='Fun/Entertaining'
my_new_dict['6']='Innovative'
my_new_dict['7']='Helpful'
my_new_dict['8']='Accurate'
my_new_dict['9']='Caring'
my_new_dict['10']='Gets me a good deal'
my_new_dict['11']='Upmarket'
my_new_dict['12']='A market leader'
my_new_dict['13']='Straight-forward'
my_new_dict['14']='Dependable'
my_new_dict['15']='For people like me'
my_new_dict['16']='Approachable'
my_new_dict['17']='On my side'
my_new_dict['18']='Inspires confidence'
my_new_dict['19']='Socially responsible'
```

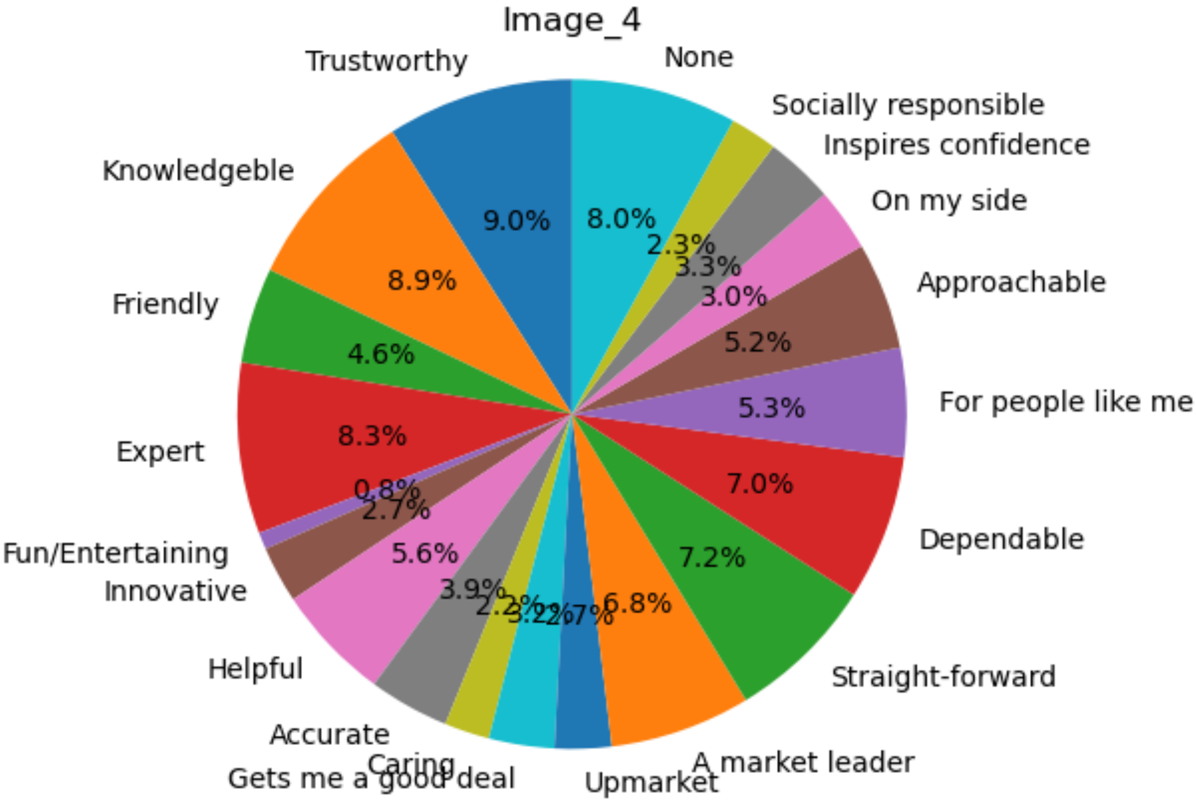
In [470…

```
my_str= ['Image_4','Image_13','Image_82','Image_124','Image_128','Image_1
final_dict={}
for i in my_str:
  x=[j for j in cols if i in j]
  image_brnd=df[x]
  image_brnd.dropna(inplace=True)
  image_brnd=image_brnd.astype(int)
  temp_dict={}
  for k in image_brnd.columns.to_list():
    temp=k.rpartition('_')[-1]
    value=image_brnd.apply(lambda x: x.value_counts(normalize=True)).fill
    temp_dict[temp]=value
  final_dict[i]=temp_dict
```

```
/var/folders/4m/mk53z65s6978g3spjfhlw2x80000gn/T/ipykernel_19446/52515004
6.py:6: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-do
cs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  image_brnd.dropna(inplace=True)
/var/folders/4m/mk53z65s6978g3spjfhlw2x80000gn/T/ipykernel_19446/52515004
6.py:6: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-do
cs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  image_brnd.dropna(inplace=True)
/var/folders/4m/mk53z65s6978g3spjfhlw2x80000gn/T/ipykernel_19446/52515004
6.py:6: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-do
cs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  image_brnd.dropna(inplace=True)
/var/folders/4m/mk53z65s6978g3spjfhlw2x80000gn/T/ipykernel_19446/52515004
6.py:6: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-do
cs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  image_brnd.dropna(inplace=True)
/var/folders/4m/mk53z65s6978g3spjfhlw2x80000gn/T/ipykernel_19446/52515004
6.py:6: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-do
cs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  image_brnd.dropna(inplace=True)
/var/folders/4m/mk53z65s6978g3spjfhlw2x80000gn/T/ipykernel_19446/52515004
6.py:6: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-do
cs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  image_brnd.dropna(inplace=True)
/var/folders/4m/mk53z65s6978g3spjfhlw2x80000gn/T/ipykernel_19446/52515004
6.py:6: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-do
cs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  image_brnd.dropna(inplace=True)
```

In [471…

```
for i in final_dict.keys():
    #print(i)
    j={my_new_dict[k]: v for k, v in final_dict[i].items()}
    #print(j)
    labels = list(j.keys())
    values = list(j.values())
    fig, ax = plt.subplots()
    ax.pie(values, labels=labels, autopct="%1.1f%%", startangle=90)
    plt.title(i)
    ax.axis('equal')
    plt.show()
```



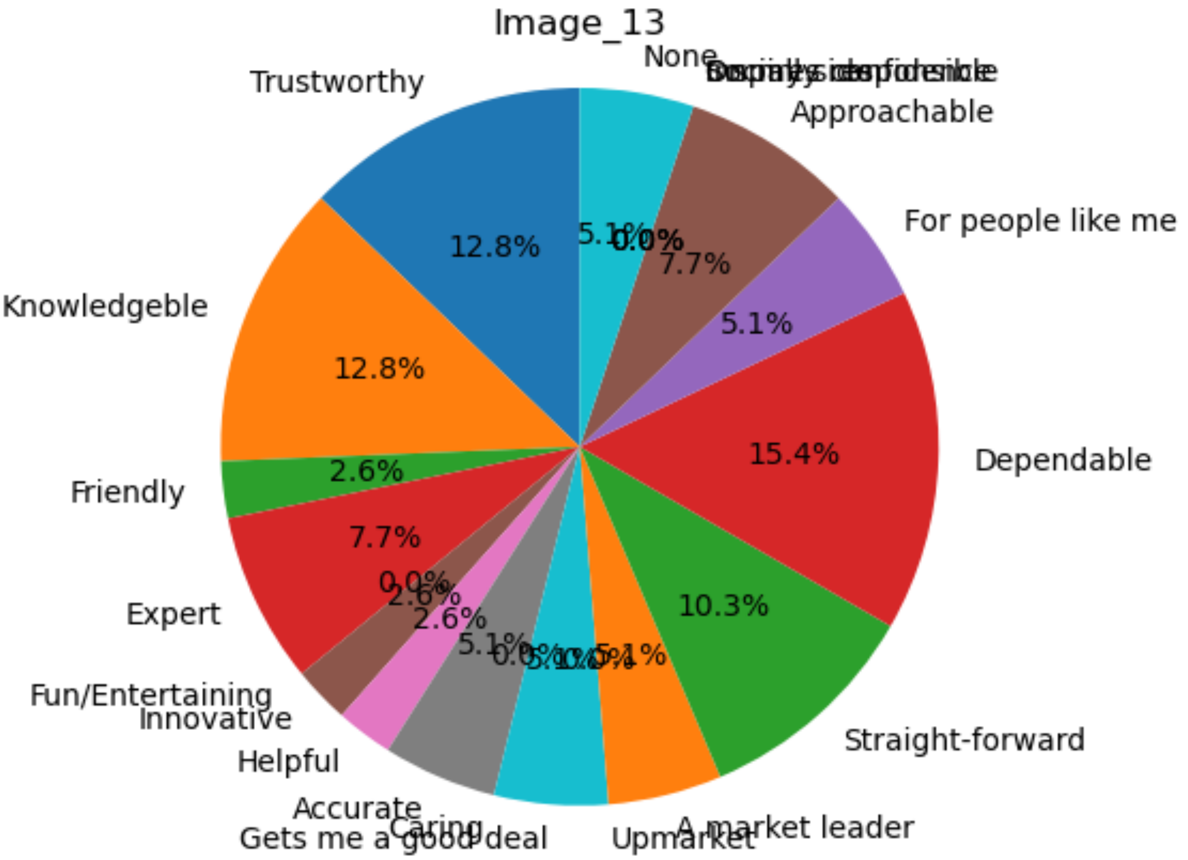


Image_13
None
Socially responsible
Inspires confidence
On my side
Approachable
For people like me
Dependable
Straight-forward
A market leader
Upmarket
Caring
Accurate
Gets me a good deal
Helpful
Innovative
Fun/Entertaining
Expert
Friendly
Knowledgeble
Trustworthy
12.8%
5.1%
0.0%
0.0%
0.0%
7.7%
5.1%
15.4%
10.3%
5.1%
0.0%
5.1%
0.0%
5.1%
2.6%
2.6%
0.0%
7.7%
2.6%
12.8%

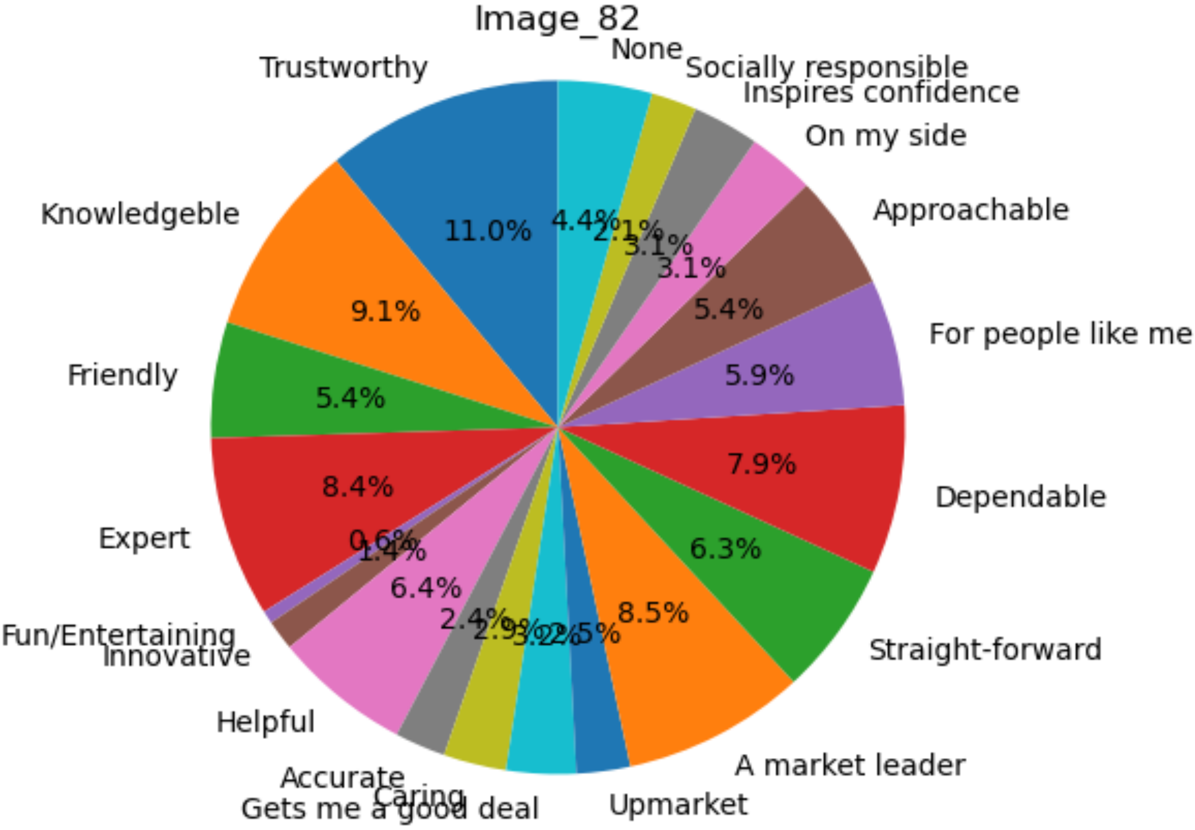


Image_82
None
Socially responsible
Inspires confidence
On my side
Approachable
For people like me
Dependable
Straight-forward
A market leader
Upmarket
Caring
Accurate
Gets me a good deal
Helpful
Innovative
Fun/Entertaining
Expert
Friendly
Knowledgeble
Trustworthy
11.0%
4.4%
2.1%
3.1%
3.1%
5.4%
5.9%
7.9%
6.3%
8.5%
2.5%
3.2%
2.9%
2.4%
6.4%
1.4%
0.6%
8.4%
5.4%
9.1%

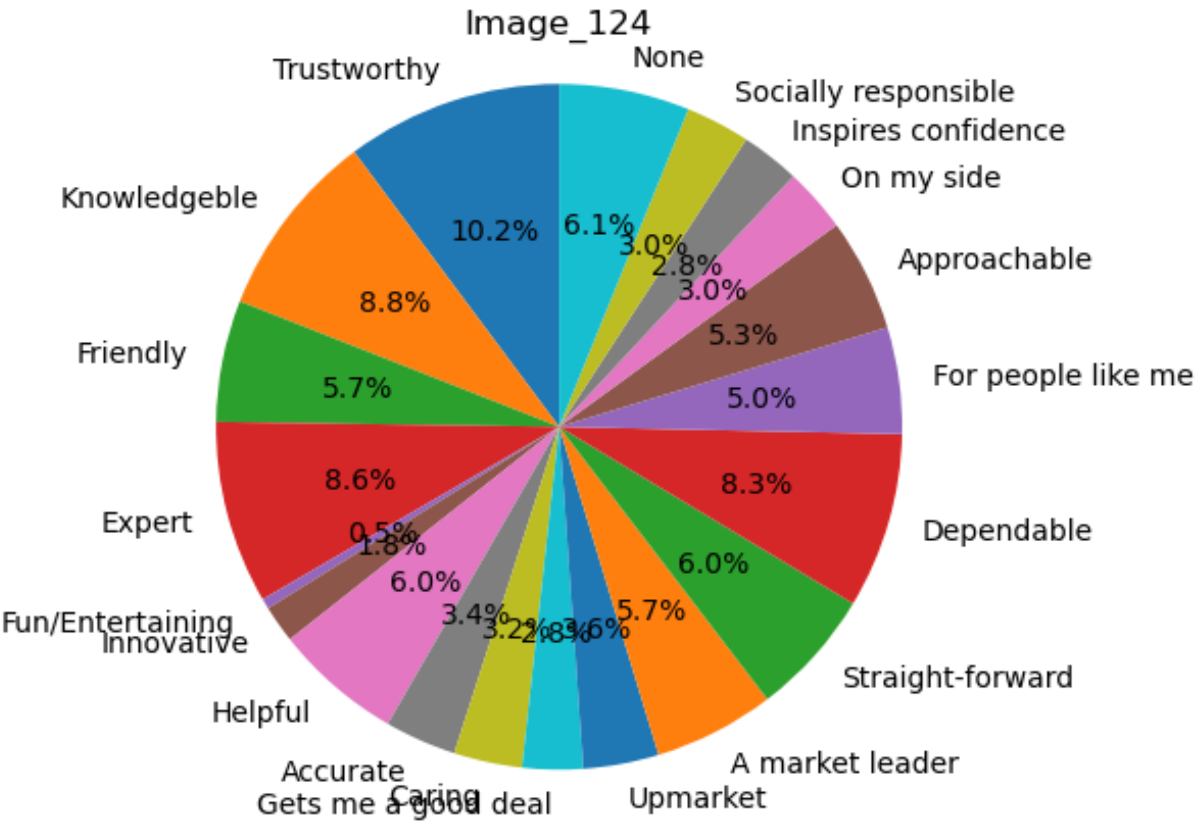


Image_124
Trustworthy
10.2%
None
6.1%
Socially responsible
3.0%
Inspires confidence
2.8%
On my side
3.0%
Approachable
5.3%
For people like me
5.0%
Dependable
8.3%
Straight-forward
6.0%
A market leader
5.7%
Upmarket
3.6%
Caring
2.8%
Gets me a good deal
3.2%
Accurate
3.4%
Helpful
6.0%
Innovative
1.8%
Fun/Entertaining
0.5%
Expert
8.6%
Friendly
5.7%
Knowledgeble
8.8%

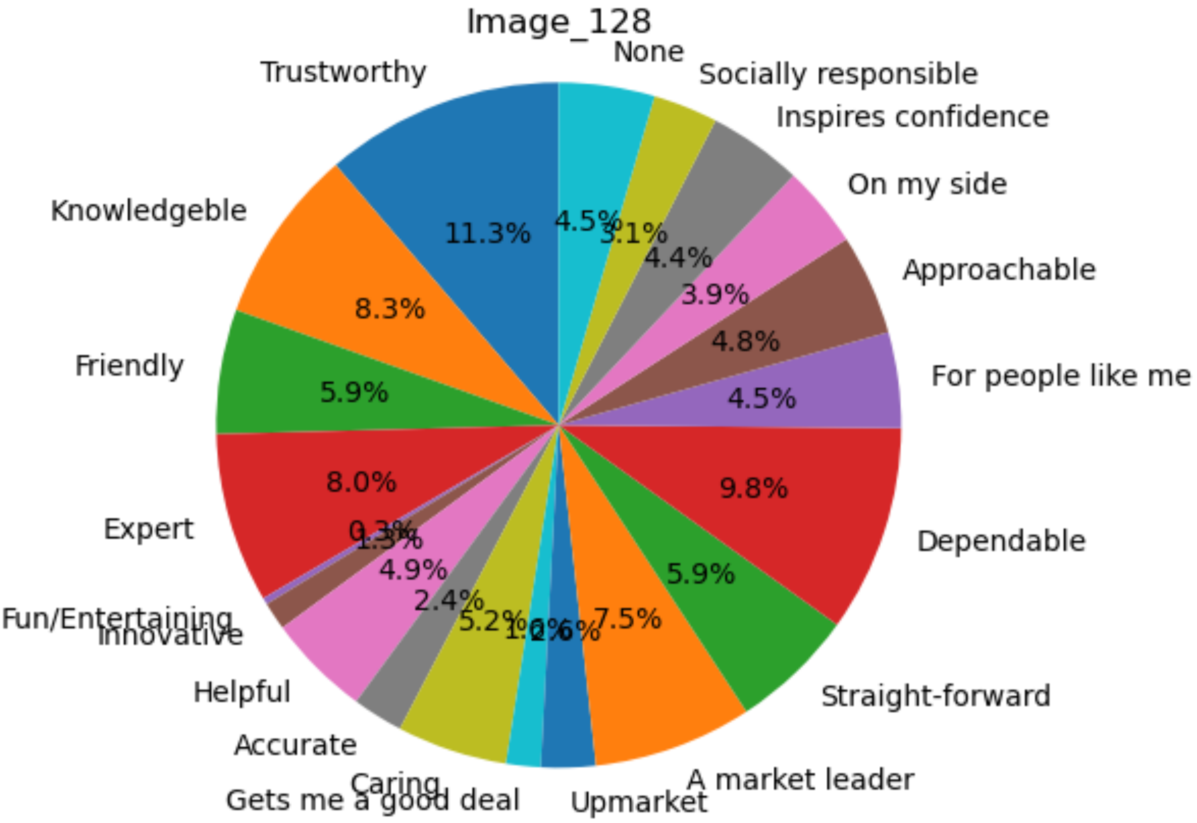
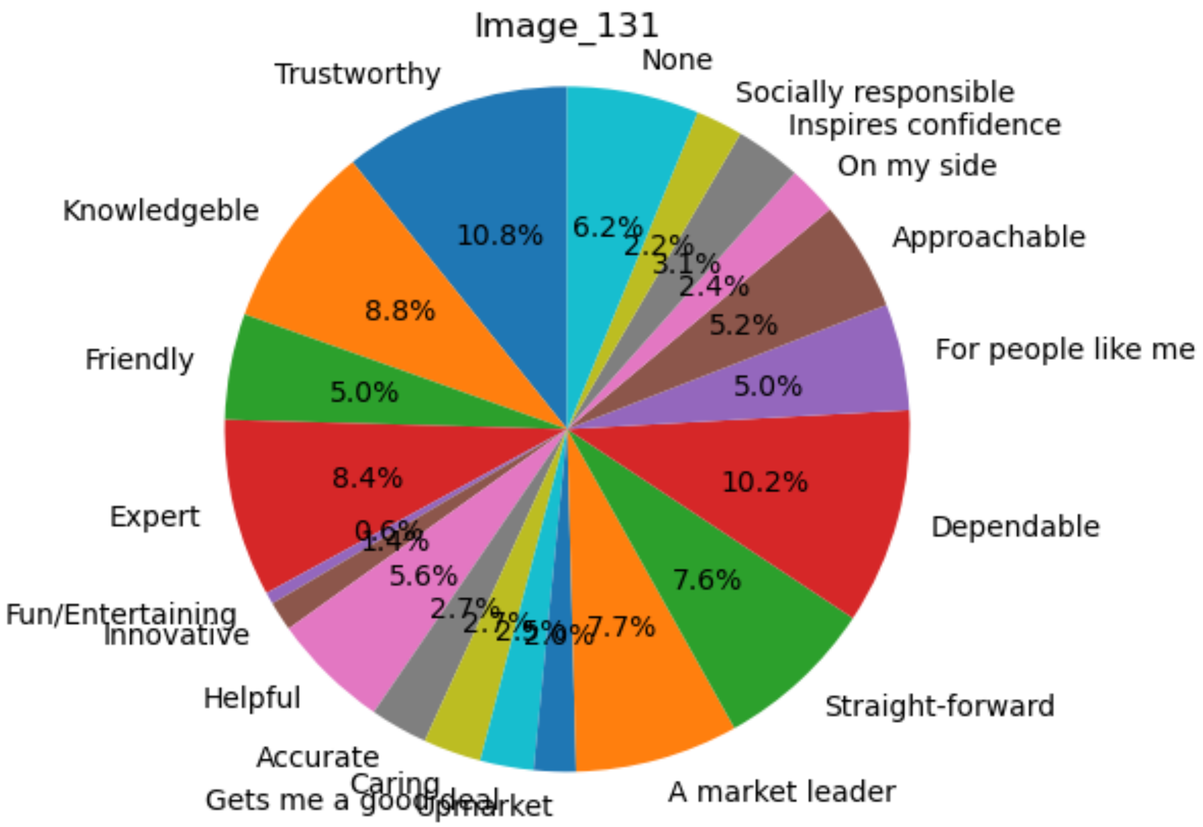
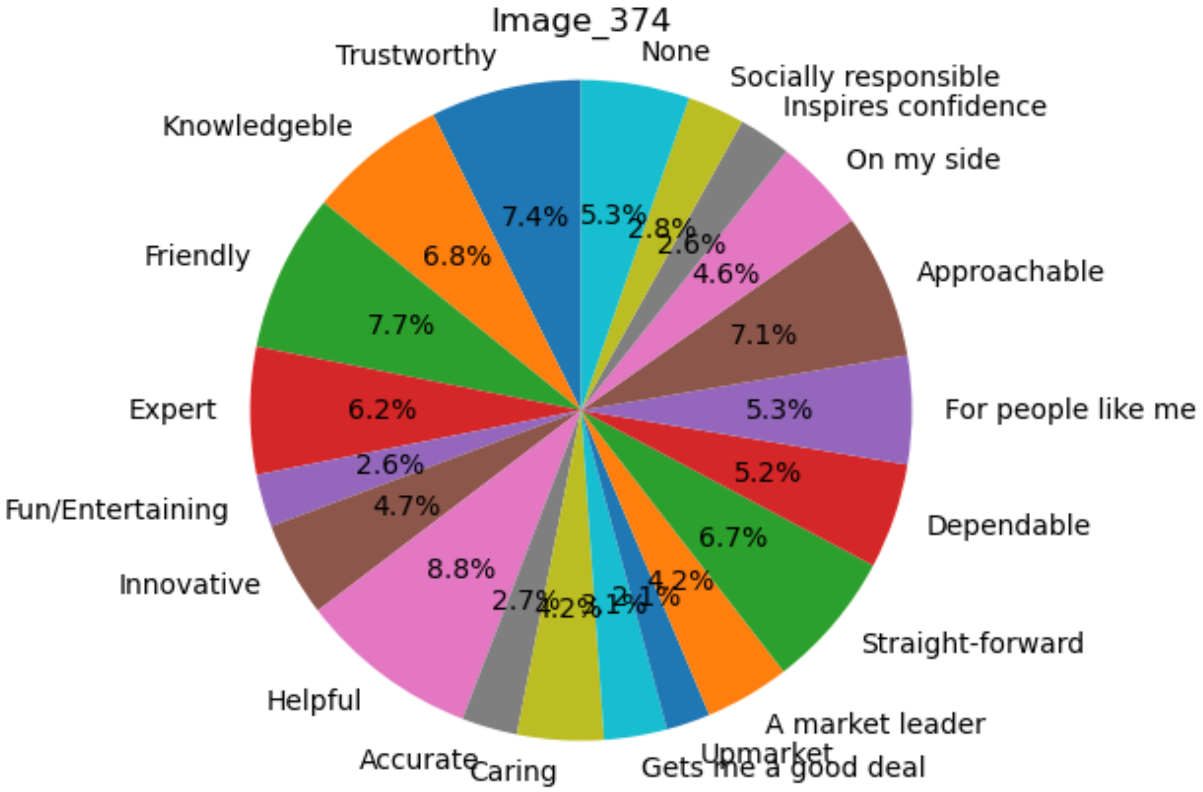


Image_128
Trustworthy
11.3%
None
4.5%
Socially responsible
3.1%
Inspires confidence
4.4%
On my side
3.9%
Approachable
4.8%
For people like me
4.5%
Dependable
9.8%
Straight-forward
5.9%
A market leader
7.5%
Upmarket
2.6%
Caring
1.2%
Gets me a good deal
5.2%
Accurate
2.4%
Helpful
4.9%
Innovative
1.3%
Fun/Entertaining
0.3%
Expert
8.0%
Friendly
5.9%
Knowledgeble
8.3%





Overall Analysis Brand 13 vs Brand 131

In [499…

```
oa1=df
```

```
In [500…  oa2=sqldf("select * from oa1",locals())
```

Brand 13

```
In [501…  oa_br13=sqldf("select * from oa2 where Consider_general_13=1",locals())
```

```
In [504…  oa_br13['Consider_general_13'].value_counts().sort_values(ascending=False
```

```
Out[504]: 1.0     4803
          Name: Consider_general_13, dtype: int64
```

```
In [505…  oa_br13.iloc[ :, 1:10].describe()
```

Out[505]:

	Gender	Age	dAge	Employment_status	Relationship_statu
count	4803.000000	4803.000000	4803.000000	4803.000000	4803.00000
mean	0.476161	47.558401	3.479492	2.854674	1.89527
std	0.499483	14.614186	1.211902	2.516607	0.69422
min	0.000000	16.000000	1.000000	0.000000	0.00000
25%	0.000000	36.000000	3.000000	1.000000	2.00000
50%	0.000000	49.000000	3.000000	1.000000	2.00000
75%	1.000000	60.000000	5.000000	6.000000	2.00000
max	1.000000	74.000000	5.000000	9.000000	5.00000

```
In [506…  oa_br13['Gender'].value_counts().sort_values(ascending=False)
```

```
Out[506]: 0     2516
          1     2287
          Name: Gender, dtype: int64
```

```
In [507…  oa_br13['Age'].describe()
```

```
Out[507]: count    4803.000000
          mean       47.558401
          std        14.614186
          min        16.000000
          25%        36.000000
          50%        49.000000
          75%        60.000000
          max        74.000000
          Name: Age, dtype: float64
```

```
In [508…  oa_br13['dAge'].value_counts().sort_values(ascending=False)
```

Out[508]:
```
3    1413
5    1289
4    1063
2     738
1     300
Name: dAge, dtype: int64
```

In [509…
```
oa_br13['Employment_status'].value_counts().sort_values(ascending=False)
```

Out[509]:
```
1    2806
6     829
4     298
5     249
8     226
2     110
7     101
0      71
9      70
3      43
Name: Employment_status, dtype: int64
```

In [510…
```
oa_br13['Relationship_status'].value_counts().sort_values(ascending=False
```

Out[510]:
```
2    3091
1    1180
3     377
4     105
5      38
0      12
Name: Relationship_status, dtype: int64
```

In [511…
```
oa_br13['Living_status'].value_counts().sort_values(ascending=False)
```

Out[511]:
```
1    3027
3     814
4     611
2     317
5      34
Name: Living_status, dtype: int64
```

In [512…
```
oa_br13['dSEG'].value_counts().sort_values(ascending=False)
```

Out[512]:
```
1    3141
2    1662
Name: dSEG, dtype: int64
```

In [513…
```
oa_br13['Personal_Income'].describe()
```

Out[513]:
```
count      4786.000000
mean      30572.133932
std       34087.896884
min        -999.000000
25%       12000.000000
50%       22958.500000
75%       35000.000000
max      200000.000000
Name: Personal_Income, dtype: float64
```

In [514…

```
oa_br13['Household_Income'].describe()
```

Out[514]:

```
count       4788.000000
mean       45704.093776
std        40336.372069
min         -999.000000
25%        20000.000000
50%        36000.000000
75%        58518.000000
max       200000.000000
Name: Household_Income, dtype: float64
```

In [515…

```
oa_br13.corr()
```

Out[515]:

	ResponseID	Gender	Age	dAge	Employment_status
ResponseID	1.000000	-0.011998	-0.035771	-0.031932	-0.006517
Gender	-0.011998	1.000000	0.059309	0.045893	-0.037713
Age	-0.035771	0.059309	1.000000	0.967171	0.325311
dAge	-0.031932	0.045893	0.967171	1.000000	0.301572
Employment_status	-0.006517	-0.037713	0.325311	0.301572	1.000000
...	...	...	...	...	...
Image_374_17	-0.009560	-0.438529	-0.115017	-0.099197	-0.097557
Image_374_18	0.416910	-0.233882	-0.029860	-0.012400	0.073168
Image_374_19	0.143194	-0.126878	0.140958	0.084968	-0.179662
Image_374_0	-0.218422	0.258199	-0.214881	-0.185326	-0.269251
Weighting	0.020745	0.069598	0.001178	0.004648	0.008281

165 rows × 165 columns

Brand 131

In [517…

```
oa_br131=sqldf("select * from oa2 where Consider_general_131=1",locals())
```

In [518…

```
oa_br131['Consider_general_131'].value_counts().sort_values(ascending=Fal
```

Out[518]:

```
1.0    2008
Name: Consider_general_131, dtype: int64
```

In [519…

```
oa_br131['Gender'].value_counts().sort_values(ascending=False)
```

Out[519]:

```
1    1050
0     958
Name: Gender, dtype: int64
```

In [520…

```
oa_br131['Age'].describe()
```

```
Out[520]: count    2008.000000
          mean       51.982570
          std        13.536133
          min        16.000000
          25%        42.000000
          50%        55.000000
          75%        63.000000
          max        74.000000
          Name: Age, dtype: float64
```

```
In [521…  oa_br131['dAge'].value_counts().sort_values(ascending=False)
```

```
Out[521]: 5    731
          4    539
          3    483
          2    187
          1     68
          Name: dAge, dtype: int64
```

```
In [523…  oa_br131['Employment_status'].value_counts().sort_values(ascending=False)
```

```
Out[523]: 1    1076
          6     448
          4     147
          5      87
          8      83
          2      58
          0      32
          7      28
          3      28
          9      21
          Name: Employment_status, dtype: int64
```

```
In [524…  oa_br131['Relationship_status'].value_counts().sort_values(ascending=Fals
```

```
Out[524]: 2    1313
          1     448
          3     172
          4      51
          5      21
          0       3
          Name: Relationship_status, dtype: int64
```

```
In [525…  oa_br131['Living_status'].value_counts().sort_values(ascending=False)
```

```
Out[525]: 1    1405
          3     277
          4     211
          2     100
          5      15
          Name: Living_status, dtype: int64
```

```
In [526…  oa_br131['dSEG'].value_counts().sort_values(ascending=False)
```

```
Out[526]: 1    1388
          2     620
          Name: dSEG, dtype: int64
```

In [527…

```
oa_br131['Personal_Income'].describe()
```

Out[527]:

```
count       2001.000000
mean       31064.037981
std        34665.946802
min         -999.000000
25%        12000.000000
50%        23000.000000
75%        35000.000000
max       200000.000000
Name: Personal_Income, dtype: float64
```

In [528…

```
oa_br131['Household_Income'].describe()
```

Out[528]:

```
count       1998.000000
mean       45476.774274
std        40375.189182
min         -999.000000
25%        20000.000000
50%        36043.500000
75%        57670.500000
max       200000.000000
Name: Household_Income, dtype: float64
```

In [ ]:

```

```