Web Mining Lab Assignment-11

Aryan Vigyat

20BCE1452

In [ ]:
```
import pandas as pd
```

Dataset Upload

In [ ]:
```
data = pd.read_csv("C/Users/ayuar/e-shop clothing 2008.csv")
data.head()
```

Out[ ]:

	year	month	day	order	country	session ID	page 1 (main category)	page 2 (clothing model)	colour	location	mod photograph
0	2008	4	1	1	29	1	1	A13	1	5	
1	2008	4	1	2	29	1	1	A16	1	6	
2	2008	4	1	3	29	1	2	B4	10	2	
3	2008	4	1	4	29	1	2	B17	6	6	
4	2008	4	1	5	29	1	2	B8	4	3	

1) When do sales peak?

In [ ]:
```
peak_sales = data.groupby(['month'])['order'].nunique()
print(peak_sales)
```

```
month
4    112
5    161
6    192
7    142
8    195
Name: order, dtype: int64
```

2) What type of clothing sells most?

In [ ]:
```
most_sold_clothing = data.groupby(['page 1 (main category)'])['order'].nunique().ic
print(most_sold_clothing)
```

```
3
```

3) What type of clothing sells most per month?

In [ ]:
```
monthly_most_sold_clothing = data.groupby(['month', 'page 1 (main category)'])['orc
monthly_most_sold_clothing = monthly_most_sold_clothing.groupby(['month']).apply(la
print(monthly_most_sold_clothing)
```

```
month
4    4
5    4
6    3
7    3
8    3
dtype: int64
```

4) Identify the sessions in the log file

In [ ]:
```
sessions = data.groupby(['session ID'])['order'].count()
print(sessions)
```

```
session ID
1         9
2        10
3         6
4         4
5         1
         ..
24022     3
24023     7
24024     1
24025     1
24026     3
Name: order, Length: 24026, dtype: int64
```

5) Does a correlation exist between price and page, and, if so, how strongly are price and product placement related?

In [ ]:
```
price_page_corr = data[['price', 'location']].corr()
print(price_page_corr)
```

```
              price  location
price     1.000000 -0.084653
location -0.084653  1.000000
```

6) Which country has the most website visitors?

In [ ]:
```
most_visitors = data['country'].value_counts().idxmax()
print(most_visitors)
```

7) Find the average number of clicks in each country

In [ ]:
```
avg_clicks_per_country = data.groupby(['country'])['order'].count().mean()
print(avg_clicks_per_country)
```

```
3520.723404255319
```

8) Does users seek cheaper products ?

In [ ]:
```
cheap_products_seeked = data.groupby(['price 2'])['order'].nunique()
print(cheap_products_seeked)
```

```
price 2
1    185
2    187
Name: order, dtype: int64
```

9) Are more clicks on the website refer to higher sales? and is the main category related?

In [ ]:
```
clicks_vs_sales = data.groupby(['page 1 (main category)'])['order'].nunique().corr(
print(clicks_vs_sales)
```

```
0.009013005332324811
```

10) What is the best selling category?

In [ ]:
```
best_selling_category = data.groupby(['page 1 (main category)'])['order'].nunique()
print(best_selling_category)
```

```
3
```

11) What do customers buy from each page?

In [ ]:
```
products_per_page = data.groupby(['page'])['page 2 (clothing model)'].nunique()
print(products_per_page)
```

```
page
1    71
2    70
3    42
4    24
5    10
Name: page 2 (clothing model), dtype: int64
```

12) What is the most attractable section in page and color of product per type?

In [ ]:
```
attractive_section_color = data.groupby(['page 1 (main category)', 'location', 'col
attractive_section_color = attractive_section_color.groupby(['page 1 (main category
print(attractive_section_color)
```

```
                         location  colour
page 1 (main category)
1                               2       3
2                               4       2
3                               6      14
4                               6       2
```

13) Are selling products affected by colour ?

In [ ]:
```
color_effect_on_sales = data.groupby(['colour'])['order'].nunique().corr(data.group
print(color_effect_on_sales)
```

```
0.07163079131878583
```

14) Find the Relationship between countries and buying days

In [ ]:
```
country_buying_days = data.groupby(['country', 'day'])['order'].nunique().reset_ind
country_buying_days = country_buying_days.groupby(['country']).apply(lambda x: x.lo
print(country_buying_days)
```

```
country
1     11
2     12
3     19
4     24
5      5
6     23
7      2
8      6
9     10
10     6
11    10
12    14
13    24
14    18
15    21
16     7
17     8
18    12
19     6
20    21
21    16
22     9
23    11
24    22
25    26
26     2
27     3
28     4
29     3
30    26
31    29
32    11
33    17
34    12
35    15
36    22
37    13
38    21
39     2
40    22
41    16
42    26
43    12
44    11
45    28
46     8
47    12
dtype: int64
```

15) Should we put model photography in the face of the product or in the inside the profile ?

In [ ]:
```
photography_placement = data.groupby(['model photography'])['order'].nunique()
print(photography_placement)
```

```
model photography
1    190
2    178
Name: order, dtype: int64
```

16) Relation between sales and buying days per month

In [ ]:
```
sales_buying_days = data.groupby(['month', 'day'])['order'].nunique().reset_index()
sales_buying_days = sales_buying_days.groupby(['month']).apply(lambda x: x.loc[x['
print(sales_buying_days)
```

```
month
4    29
5     8
6    10
7    16
8     3
dtype: int64
```