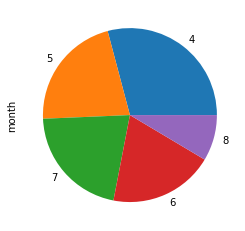
Data description “e-shop clothing 2008”

Variables:

1. YEAR (2008)

2. MONTH -> from April (4) to August (8)



تقارب نسب المبيعات في الأشهر ماعدا شهر أوت تتناقص النسبة

========================================================

3. DAY -> day number of the month

========================================================

4. ORDER -> sequence of clicks during one session

========================================================

5. COUNTRY -> variable indicating the country of origin of the IP address with the

following categories:

1-Australia

2-Austria

3-Belgium

4-British Virgin Islands

5-Cayman Islands

6-Christmas Island

7-Croatia

8-Cyprus

9-Czech Republic

10-Denmark

11-Estonia

12-unidentified

13-Faroe Islands

14-Finland

15-France

16-Germany

17-Greece

18-Hungary

19-Iceland

20-India

21-Ireland

22-Italy

23-Latvia

24-Lithuania

25-Luxembourg

26-Mexico

27-Netherlands

28-Norway

29-Poland

30-Portugal

31-Romania

32-Russia

33-San Marino

34-Slovakia

35-Slovenia

36-Spain

37-Sweden

38-Switzerland

39-Ukraine

40-United Arab Emirates

41-United Kingdom

42-USA

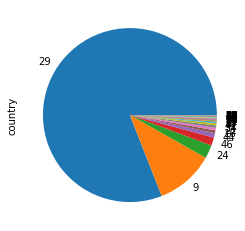
43-biz (\*.biz)

44-com (\*.com)

45-int (\*.int)

46-net (\*.net)

47-org (\*.org)



نسب المبيعات في غالب الدول ضعيفة ماعدا Poland ثم تليها Czech Republic و Lithuania ما يظهر أن دول أوروبا الشرقية أكثر إقبال

========================================================

6. SESSION ID -> variable indicating session id (short record)

========================================================

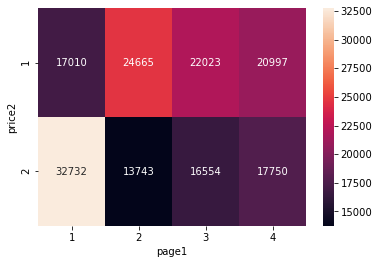
7. PAGE 1 (MAIN CATEGORY) -> concerns the main product category:

1-trousers

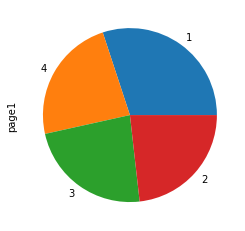
2-skirts

3-blouses

4-sale



رأي أغلب المستجوبين أن المنتج 1-trousers ليس قريب من متوسط فئة المنتج أما بقية المنتجات فهي قريبة من متوسط فئة المنتج



تقارب نسب المبيعات بالنسبة لنوعية المنتجات

========================================================

8. PAGE 2 (CLOTHING MODEL) -> contains information about the code for each product

(217 products)

========================================================

9. COLOUR -> colour of product

1-beige

2-black

3-blue

4-brown

5-burgundy

6-gray

7-green

8-navy blue

9-of many colors

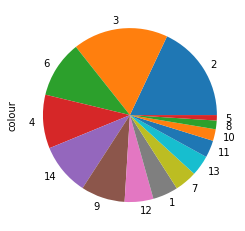
10-olive

11-pink

12-red

13-violet

14-white



الإقبال بالترتيب أكثر على اللون 2-black - 3-blue و 4-brown- 6-gray 14-white

========================================================

10. LOCATION -> photo location on the page, the screen has been divided into six parts:

1-top left

2-top in the middle

3-top right

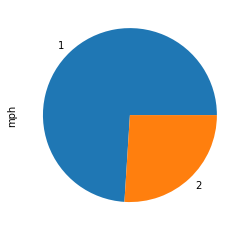
4-bottom left

5-bottom in the middle

6-bottom right

========================================================

11. MODEL PHOTOGRAPHY -> variable with two categories:



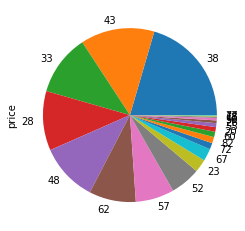
مشاهدة أكثر للصورة 1-en face

1-en face

2-profile

========================================================

12. PRICE -> price in US dollars



ترتيب السعر الأكثر أهمية 38-43-33-28-48

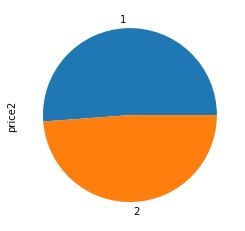
========================================================

13. PRICE 2 -> variable informing whether the price of a particular product is higher than

the average price for the entire product category

1-yes

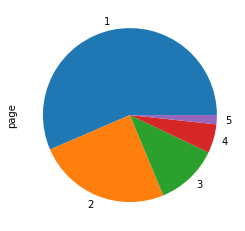
2-no



نصف الإجابات تظهر أن السعر أكبر من المتوسط لفئة المنتج و النصف الآخر العكس

===============================================================

14. PAGE -> page number within the e-store website (from 1 to 5)



الصفحة رقم 1 داخل موقع المتجر الإلكتروني الأكثر تصفحا ثم تليها الصفحة 2

=================================================================

**دراسة الارتباط :**

**بالنسبة للمتغير** **price**

price2 -0.743633

page1 -0.364791

mph -0.212395

page -0.150455

colour -0.091159

location -0.084653

order -0.064779

country -0.045345

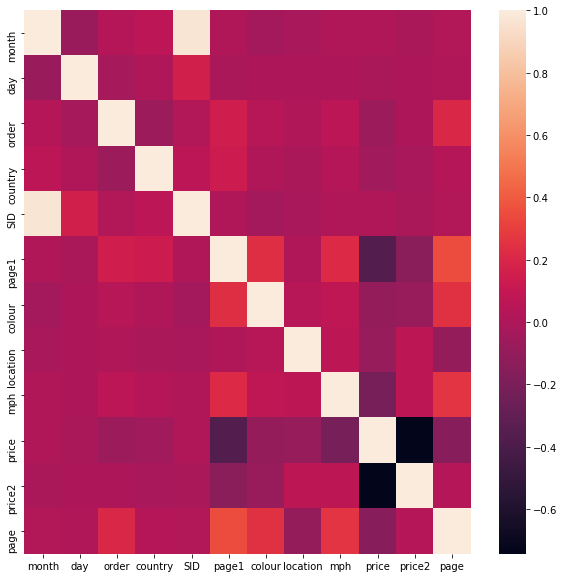
day -0.002818

SID 0.011669

month 0.012700

price 1.000000

أكبر نسبة ارتباطمع المتغير priceهي للمتغير 2price ثم تليهاpage1 و mphهو ارتباط عكسي



**تحليل البيانات :**

The mean squared error (MSE) on test set: 34.01274585975821

يعتبر ذو قيمة كبيرة لأن المتغيرات من أوزان مختلفة

The R-mean squared error (RMSE) on test set: 5.832044740891329

يعتبر ذو قيمة قليلة و هو جيد لأنه يوحد أوزان المتغيرات أي نسبة حظأ 6 دولارات في قاعدة بيانات تحتوي على (165474, 13)

The Mean Absolute\_Error(MAE) on test set: [4.38133644]

كذلك يعتبر ذو قيمة قليلة و هو جيد أي نسبة حظأ 5 دولارات في قاعدة بيانات تحتوي على (165474, 13)

يعزى هذا للارتباط القوي مع المتغيرات السالفة الذكر2price page1 و mph ما يعبر أن النموذج جيد للتنبؤ .