Page
Date / /

6. Customer Purchase Behaviour

Dataset

- Customer ID (Unique)
- Age group - 18-25 -
  25-35 -
  36-45 -
  46-60 -
- Product category - Clothing
  Electronics
  Home Decor

Interpretation

↳ we see that Product category wise distribution of products is as follows

Home Decor - 30.7%
Electronics - 29.6%
Clothing - 39.7%

↳ Age distribution wise
(18-25) - 26.9%
(26-35) 23.4%
(36-45) - 25.4%
(46-60) - 24.3%

Page
Date / /

* Clothing takes maximum proportion of sales throughout the age groups.

1. Observed Frequencies: we have them from the contingency table

2. Expected Frequencies:

H_0: Age Group and Product category are independent (no significant association b/w the age group & product category)

H_a: Age Group and Product category are dependent (significant association b/w the age group & product category)

$$E_{ij} = \frac{(\text{Row total}) \times (\text{Column total})}{\text{Total}}$$

$$E_{(18\text{-}25,\ \text{Clothing})} = \frac{269 \times 397}{1000} = 106.793$$

$$E_{(18\text{-}25,\ \text{Electronics})} = \frac{269 \times 296}{1000} = 79.624$$

Page
Date / /

$E_{(18-25,\ \text{Home Decor})} = \frac{307 \times 269}{1000} = 82.583$

$E_{(26-35,\ \text{Clothing})} = \frac{234 \times 397}{1000} = 92.898$

$E_{(26-35,\ \text{Electronics})} = \frac{234 \times 296}{1000} = 69.264$

$E_{(26-35,\ \text{Home Decor})} = \frac{234 \times 307}{1000} = 71.838$

$E_{(36-45,\ \text{Clothing})} = \frac{254 \times 397}{1000} = 100.838$

$E_{(36-45,\ \text{Electronics})} = \frac{254 \times 296}{1000} = 75.184$

$E_{(36-45,\ \text{Home Decor})} = \frac{254 \times 307}{1000} = 77.978$

$E_{(46-60,\ \text{Clothing})} = \frac{243 \times 397}{1000} = 96.471$

$E_{(46-60,\ \text{Electronics})} = \frac{243 \times 296}{1000} = 71.928$

$$E_{(46-60)\text{Home Decor}} = \frac{243 \times 307}{1000} = 74.601$$

Calculate χ^2 - Square statistics -

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

O_{ij} = observed frequency
E_{ij} = Expected frequency

~~Observed, O = 84~~
~~Expected, E = 79.821~~

Cell (18-25, Clothing):

$O = 105$
$E = 106.793$

$$\chi^2 = \frac{(105 - 106.793)^2}{106.793}$$

$$= \frac{3.214849}{106.793} = 0.0301$$

Page
Date / /

Cell (18-25, Electronics)

O = 84
E = 79.624

$\chi^2 = \frac{(84 - 79.624)^2}{79.624}$

$= 0.24049$

Cell (18-25, Home Decor)

O = 80
E = 82.583

$\chi^2 = \frac{(80 - 82.583)^2}{82.583}$

$= 0.08079$

Cell (Clothing, 26-35)

O = 90
E = 92.898

$\chi^2 = \frac{(90 - 92.898)^2}{92.898}$

$= 0.0904$

Cell (26-35, Electronics)

O = 71
E = 69.264

$\chi^2 = \frac{(71 - 69.264)^2}{69.264}$

$= 0.04351$

Page
Date / /

Cell (26-35, HomeDecor)

O = 73
E = 71.838

$\chi^2 = \frac{(73-71.838)^2}{71.838}$

$= 0.01879$

Cell (36-45, Clothing):

O = 106
E = 100.838

$\chi^2 = \frac{(106-100.838)^2}{100.838}$

$= 0.2642$

Cell (36-45, Electronics)

O = 78
E = 75.184

$\chi^2 = \frac{(78-75.184)^2}{75.184}$

$= 0.105472$

Cell (36-45, HomeDecor)

O = 70
E = 77.978

$\chi^2 = \frac{(70-77.978)^2}{77.978}$

$= 0.8162$

Page
Date / /

Cell (46-60, Clothing)

O = 96
E = 96.471

$X^2 = \frac{(96 - 96.471)^2}{96.471}$

$= 0.00229$

Cell (46-60, Electronics)

O = 63
E = 71.928

$X^2 = \frac{(63 - 71.928)^2}{71.928}$

$= 1.108$

Cell (46-60, Home Decor):

O = 84
E = 74.601

$X^2 = \frac{(84 - 74.601)^2}{74.601}$

$= 1.184$

X^2 = sum of all contributions
40 X^2

$= 3.984242$

Page
Date / /

• Degree of Freedom (df)

$df = (\text{no. of rows} - 1) \times (\text{no. of columns} - 1)$

$= (4-1) \times (3-1) = 3 \times 2 = 6$

p-value

$\{\chi^2 \approx 3.98, df = 6\}$

p-value = 0.6787

for $\alpha = 0.05$

- we fail to reject the null hypothesis.

$0.6787 \not< \alpha (0.05)$

$\Rightarrow$ there is no statistically significant association b/w the age group and the product category purchased.

Page
Date / /

Business Decisions

⇒ Ad-campaign should not be based on age-specific for each product, business could invest in a more uniform marketing approach across all age groups.

⇒ ∵ Clothing is the most demanded item across all age groups. Inventory could be kept accordingly.

X ——— X