E-retail factors for customer activation and retention: A case study from Indian e-commerce customers

- The dataset is all about the online shopping website's reports and status, given by the customers.

- Our excel file contains 2 excel sheets:

1. Is the detailed excel file.

2. Encoded excel file.

TASK: To apply the data analysis and find the conclusion of data analysis.

- Importing the necessary libraries such as: pandas, NumPy, seaborn, matplotlib.

- Import our dataset.

1. df1 is the detailed sheet.

2. df2 is the encoded sheet.

*** Taking df1 for the visualization & df2 for the further analysis. ***

- **EDA**

  *** df1 ***

  - Shape of df1 is 269, 71.

  ** data.shape **

  - Information of dataset(df1)

  There are no null values present in df1.

  Range Index: 0 to 268

  Total columns: 71

  dtypes: int64(1), object(70)

  - Null Values.

  ** data.isna().sum() **

  No Null values present in any of the columns in df1.

  - Descriptive Statistic

  ** data.describe(include = 'all') **

1Gender of respondent2 How old are you?3 Which city do you shop online from?4 What is the Pin Code of where you shop online from?5 Since How Long You are Shopping Online ?6 How many times you have made an online purchase in the past 1 year?7 How do you access the internet while shopping on-line?8 Which device do you use to access the online shopping?9 What is the screen size of your mobile device?\t\t\t\t\t\t10 What is the operating system (OS) of your device?\t\t\t\t...Longer time to get logged in (promotion, sales period)Longer time in displaying graphics and photos (promotion, sales period)Late declaration of price (promotion, sales period)Longer page loading time (promotion, sales period)Limited mode of payment on most products (promotion, sales period)Longer delivery periodChange in website/Application designFrequent disruption when moving from one page to anotherWebsite is as efficient as beforeWhich of the Indian online retailer would you recommend to a friend?

count269269269269.000000269269269269269269...269269269269269269269269269269

unique2511NaN564443...1010811867888

topFemale31-40 yearsDelhiNaNAbove 4 yearsLess than 10 timesMobile internetSmartphoneOthersWindow/windows Mobile...Amazon.inAmazon.in, Flipkart.comMyntra.comMyntra.comSnapdeal.comPaytm.comAmazon.inAmazon.inAmazon.inAmazon.in

freq1818158NaN98114142141134122...57607561877296539479

meanNaNNaNNaN220465.747212NaNNaNNaNNaNNaNNaN...NaNNaNNaNNaNNaNNaNNaNNaNNaNNaN

stdNaNNaNNaN140524.341051NaNNaNNaNNaNNaNNaN...NaNNaNNaNNaNNaNNaNNaNNaNNaNNaN

minNaNNaNNaN110008.000000NaNNaNNaNNaNNaNNaN...NaNNaNNaNNaNNaNNaNNaNNaNNaNNaN

25%NaNNaNNaN122018.000000NaNNaNNaNNaNNaNNaN...NaNNaNNaNNaNNaNNaNNaNNaNNaNNaN

50%NaNNaNNaN201303.000000NaNNaNNaNNaNNaNNaN...NaNNaNNaNNaNNaNNaNNaNNaNNaNNaN

75%NaNNaNNaN201310.000000NaNNaNNaNNaNNaNNaN...NaNNaNNaNNaNNaNNaNNaNNaNNaNNaN

maxNaNNaNNaN560037.000000NaNNaNNaNNaNNaNNaN...NaNNaNNaNNaNNaNNaNNaNNaNNaNNaN

- Duplicate Values.

** data.duplicated().sum() **

There are duplicated values present in out dataset, so dropping all the duplicated values.

** data.drop_duplicates(inplace = True) **

After dropping duplicate values our dataset shape is 103, 71.

*** df2 ***

- Shape of df2 is 269, 71.

** data.shape **

- Information of dataset(df2)

There are no null values present in df2.

Range Index: 0 to 242

Total columns: 71

dtypes: int64(46), object(25).

- Null Values.

** data.isna().sum() **

No Null values present in any of the columns in df2.

- Descriptive Statistic

** data.describe(include = 'all') **

1Gender of respondent2 How old are you?4 What is the Pin Code of where you shop online from?5 Since How Long You are Shopping Online ?6 How many times you have made an online purchase in the past 1 year?7 How do you access the internet while shopping on-line?8 Which device do you use to access the online shopping?9 What is the screen size of your mobile device?\t\t\t\t\t\t10 What is the operating system (OS) of your device?\t\t\t\t11 What browser do you run on your device to access the website?\t\t\t...38 User satisfaction cannot exist without trust39 Offering a wide variety of listed product in several category40 Provision of complete and relevant product information41 Monetary savings42 The Convenience of patronizing the online retailer43 Shopping on the website gives you the sense of adventure44 Shopping on your preferred e-tailer enhances your social status45 You feel gratification shopping on your favorite e-tailer46 Shopping on the website helps you fulfill certain roles47 Getting value for money spent

count139.000000139.000000139.000000139.000000139.000000139.000000139.000000139.000000139.000000139.000000...139.000000139.000000139.000000139.000000139.000000139.000000139.000000139.000000139.000000139.000000

mean0.6258992.985612236059.9136693.5107912.8489213.8417271.7122304.2446041.81295
01.280576...4.1438854.1798564.3237414.2949643.9208633.6115113.1726623.4892093.20863
34.179856

std0.4856401.028476153177.8924851.4860071.6590271.2054160.8784420.9618950.8215250.
614189...1.1200490.8187280.7634381.0032790.6922421.0391661.2448621.1694601.2006930.
639880

min0.0000001.000000110008.0000001.0000001.0000002.0000001.0000002.0000001.0000001
.000000...1.0000002.0000002.0000002.0000003.0000001.0000001.0000001.0000001.0000003
.000000

25%0.0000002.000000132001.0000003.0000001.0000003.0000001.0000004.0000001.0000001
.000000...4.0000004.0000004.0000004.0000003.0000003.0000003.0000003.0000003.0000004
.000000

50%1.0000003.000000201306.0000004.0000003.0000004.0000001.0000004.0000002.0000001
.000000...4.0000004.0000004.0000005.0000004.0000004.0000003.0000003.0000003.0000004
.000000

75%1.0000004.000000202156.5000005.0000004.0000005.0000002.0000005.0000003.0000001
.000000...5.0000005.0000005.0000005.0000004.0000004.0000004.0000004.0000004.0000005
.000000

max1.0000005.000000560037.0000005.0000005.0000005.0000004.0000005.0000003.0000004
.000000...5.0000005.0000005.0000005.0000005.0000005.0000005.0000005.0000005.0000005
.000000

- Duplicate Values.

** data.duplicated().sum() **

There are duplicated values present in out dataset, so dropping all the duplicated values.

** data.drop_duplicates(inplace = True) **

After dropping duplicate values our dataset shape is 139, 71.

- **Visualization (Insights)**

  - Univariate analysis

  ** sns.countplot(data[columns]) **

  ** sns.countplot(data[columns]) **

Plotting countplot & histplot to visualize the contribution of each entry present in the column.

- Bivariate analysis

** plt.scatterplot(x, y, data = data) **

Plotting the scatter graph on different columns, for visualizing the relationship between two columns.

- **Correlation.**

  ** sns.heatmap(data.corr(), annot = True) **

  Plotting the heatmap to see the multicollinearity between the columns.

  - There is multicollinearity problem present in our dataset, i.e., some columns are correlated with each other's.

- **Normal Distribution and Outliers.**

  *** Taking df2 for this analysis. ***

  ** sns.distplot(data[columns]) ** (Normal Distribution)

  ** sns.boxplot(data[columns]) ** (Outliers)

  Almost all the columns are categorical columns.