CUSTOMER RETENTION PROJECT

Data Observation:

1. Data set is 47 features with 269 samples.
2. No Missing value is found. So need of handling missing values.
3. Standard deviation and mean are not applicable as all data is a categorical data.
4. Percentile values are even good and it is also categorical.
5. All data types are same and good with that.

EDA:

1. Since we have city name, we can exclude Pin code data column.
2. City name is in String format. City column is encoded using ordinal encoder.
3. From Scatter chart, data is evenly distributed. I checked with Gender data column vs other all columns.
4. **Skewness in data set:**
    - Skewness exists in some features, that need to be treated using Z-Score or Box plot outlier
        1. 8 Which device do you use to access the online shopping?
        2. 9 What is the screen size of your mobile device?
        3. 11 What browser do you run on your device to access the website?
        4. 12 Which channel did you follow to arrive at your favorite online store for the first time?
        5. 21 All relevant information on listed products must be stated clearly
        6. 23 Loading and processing speed
        7. 24 User friendly Interface of the website
5. **Multi collinearity:**
    - Multi collinearity exist on dataset. The most collinear feature with label can be retained and other collinear features can re dropped to save machine time.
6. **Feature Selection:**
    - Many features can be dropped. But here, we do not have defined target. So, dropping the collinear feature should be based on target relationship. Like, if two features have multicollinearity, the feature among the two which have highest relation with target can be retained and other one can dropped. This way, in this data set we have a chance of dropping around 7 features.

7. **Imbalanced Dataset**
    - The Data is imbalanced. Most of the features which mentioned in skewness have imbalanced data set, which need to be treated.
    - Since the data set is very small, we can treat by using oversampling technique (SMOTE).