PROJECT REPORT ON

CUSTOMER RETENTION PROJECT

SUBMITTED BY:

MEGHA SINGH

TABLE OF CONTENTS

ACKNOWLEDGMENT

I’m extremely grateful to Filp Robo who gave me chance to in hence my experience and get more knowledge by doing research on the topics.

I would also like to Thank:

Stack overflow, Scikit-learn.org,Kaggle.com, Google etc. Who helped me always to get information regarding my topics.

INTRODUCTION

1.Customer retention is necessary now days, if the customer is satisfied, it helps the company to distinguish itself among its competitors and thus helps to grow and develop more.

2.Using this Case study we can actually predict from e-commerce where Retailors business will work more.

3.Using EDA we can find allot of ways where a Retailors should invest more and grow their business.

4.Using EDA we will also be looking at the expectations of the customer on the goods which customers like to buy from e-commerce website.

The libraries include:

1. Pandas- Pandas library help us to read the data in the jupyter notebook.

2. Numpy- Numpy is a library in python which can be used to perform number of mathematical operations on arrays.

4. Seaborn- Seaborn is a python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphs.

5. Matplotlib- Matplotlib is python library for basic plotting. We can make bars, pies, lines, scatterplots and so on.

6. Warning- The warning module of python warns the user of certain situation that aren't fatal. It is distinct from an error and cause no stoppage in the program.

```
#importing the libraries
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
import warnings
warnings.filterwarnings('ignore')
```

The Data Sets

- Here in our datasets,we can see we have 269 rows and 71 columns.
- Using df.info().we can get the information of the datesets.
- Here all values are of Object type only pincode is Numeric. 2)Integre is only 1 and object type is 70.So we have to encode it to convert to numeric type 3)there is no null value present 4)memory storage is 149.3+ Kb.

Out[3]:

	1Gender of respondent	2 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\t	10 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)	declar of (promo sales pe
0	Male	31-40 years	Delhi	110009	Above 4 years	31-40 times	Dial-up	Desktop	Others	Window/windows Mobile	...	Amazon.in	Amazon.in	Flipkar
1	Female	21-30 years	Delhi	110030	Above 4 years	41 times and above	Wi-Fi	Smartphone	4.7 inches	IOS/Mac	...	Amazon.in, Flipkart.com	Myntra.com	snapdea
2	Female	21-30 years	Greater Noida	201308	3-4 years	41 times and above	Mobile Internet	Smartphone	5.5 inches	Android	...	Myntra.com	Myntra.com	Myntra
3	Male	21-30 years	Karnal	132001	3-4 years	Less than 10 times	Mobile Internet	Smartphone	5.5 inches	IOS/Mac	...	Snapdeal.com	Myntra.com, Snapdeal.com	Myntra
4	Female	21-30 years	Bangalore	530068	2-3 years	11-20 times	Wi-Fi	Smartphone	4.7 inches	IOS/Mac	...	Flipkart.com, Paytm.com	Paytm.com	Paytn
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
264	Female	21-30 years	Solan	173212	1-2 years	Less than 10 times	Mobile Internet	Smartphone	5.5 inches	Android	...	Amazon.in	Amazon.in	Amaz
265	Female	31-40 years	Ghaziabad	201008	1-2 years	31-40 times	Mobile Internet	Smartphone	Others	Android	...	Flipkart.com	Flipkart.com	Flipkar
266	Female	41-50 yaers	Bangalore	560010	2-3 years	Less than 10 times	Mobile internet	Laptop	Others	Window/windows Mobile	...	Amazon.in	Snapdeal.com	Amaz
267	Female	Less than 20 years	Solan	173229	2-3 years	Less than 10 times	Wi-Fi	Smartphone	5.5 inches	Android	...	Amazon.in	Amazon.in, Myntra.com, Snapdeal.com	Amaz
268	Female	41-50 yaers	Ghaziabad	201009	2-3 years	31-40 times	Mobile Internet	Smartphone	5.5 inches	Android	...	Amazon.in	Amazon.in	Amaz

269 rows × 71 columns

- The columns of the dataset itself is a brief description of the dataset.

- Using df.columns we can get the names of the coluns which is present in our datsets.

Few columns are as follows:

1Gender of respondent', '2 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? ',
'10 What is the operating system (OS) of your device? ',
'11 What browser do you run on your device to access the website? ',
'12 Which channel did you follow to arrive at your favorite online st ore for the first time? ',
'13 After first visit, how do you reach the online retail store?

Etc…

Exploratory data analysis

1.Checking the shape of the data.

```
In [4]: df.shape
Out[4]: (269, 71)
```

The shape of the data set is rows- 269 and columns- 71.

2. Checking the information of the data.

```
In [7]: df.describe()
```

Out[7]:

	4 What is the Pin Code of where you shop online from?
count	269.000000
mean	220465.747212
std	140524.341051
min	110008.000000
25%	122018.000000
50%	201303.000000
75%	201310.000000
max	560037.000000

here we can see the statistic value of the dataset: Min value:-110008 Max Value;-560037 ,3rd quartile value :-201310

3. Checking the null value of the data:

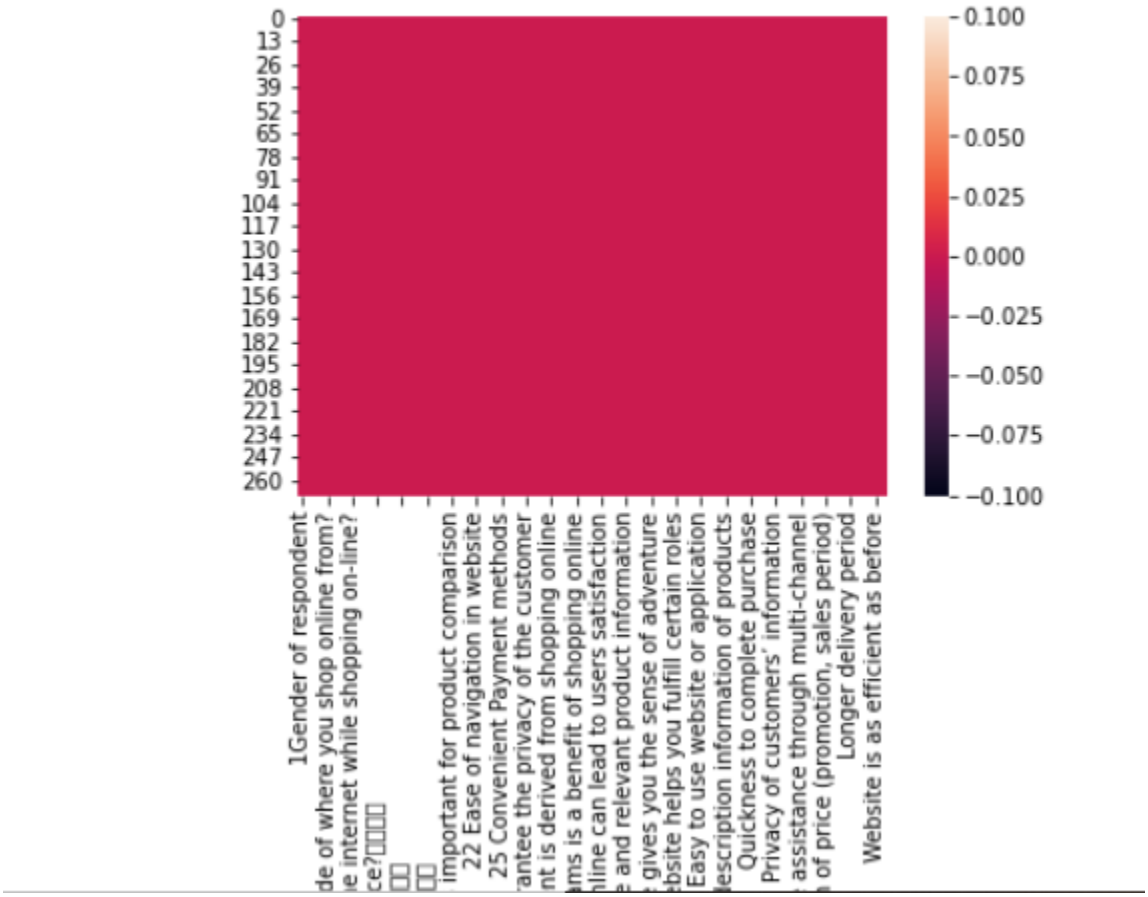
```
In [8]: df.isnull().sum()
Out[8]: 1Gender of respondent                                          0
        2 How old are you?                                             0
        3 Which city do you shop online from?                          0
        4 What is the Pin Code of where you shop online from?          0
        5 Since How Long You are Shopping Online ?                     0
                                                                      ..
        Longer delivery period                                         0
        Change in website/Application design                           0
        Frequent disruption when moving from one page to another       0
        Website is as efficient as before                              0
        Which of the Indian online retailer would you recommend to a friend?    0
        Length: 71, dtype: int64
```

here we can see no null value is present

→We can also check null values using heatmap.

```
In [9]: sns.heatmap(df.isnull())
```

```
Out[9]: <AxesSubplot:>
```



4. Renaming the columns of the data.

- Here in our data sets we can change our column names to an easy to understand way.
- On checking the columns using the method ‘column’, it is observed that column names include white spaces, tabs,

unwanted special characters, etc. which needs to be removed and thus renaming the same.

```
# Instanciating the list of columns name
column_list = df_cust.columns

new_col_list = []
for col in column_list:
    # Removing the digits from the column name
    without_digits = col.lstrip(string.digits)
    # Removing the white spaces
    without_spaces = without_digits.strip()
    # Replacing space with '_'
    replace_char = without_spaces.replace("-", "_").replace(" ", "_")
    new_col_list.append(replace_char)

new_col_list
```

Here we have used for loop for removing extra space, digit and also few special char.

Lastly, the columns in the original data frame are renamed with the new or modified column list using rename().

After renaming out datasets look like this. Which is more in understable format.

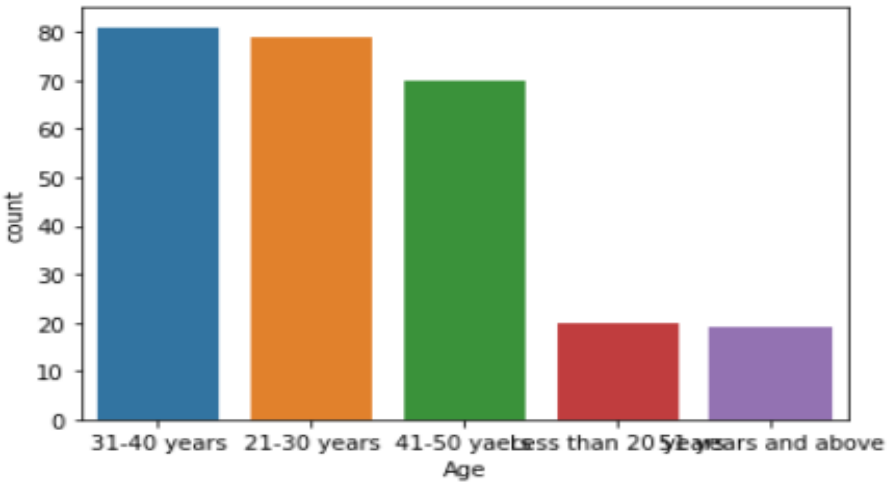
	Gender	Age	City	Pin code	Years Shopping Online	frequency of online purchase in past 1 year	Way to access the internet while shopping on-line	Device used to access online shopping	screen size of your mobile device	operating system (OS) of your device	browser to access the website	channel followed to arrive at your favorite online store for the first time	After first visit, how do you reach the online retail store	time to explore the e-retail store before making a purchase decision
0	Male	31-40 years	Delhi	110009	Above 4 years	31-40 times	Dial-up	Desktop	Others	Window/windows Mobile	Google chrome	Search Engine	Search Engine	6-10 mins
1	Female	21-30 years	Delhi	110030	Above 4 years	41 times and above	Wi-Fi	Smartphone	4.7 inches	IOS/Mac	Google chrome	Search Engine	Via application	more than 15 mins
2	Female	21-30 years	Greater Noida	201308	3-4 years	41 times and above	Mobile Internet	Smartphone	5.5 inches	Android	Google chrome	Search Engine	Via application	11-15 mins
3	Male	21-30 years	Karnal	132001	3-4 years	Less than 10 times	Mobile Internet	Smartphone	5.5 inches	IOS/Mac	Safari	Search Engine	Search Engine	6-10 mins
4	Female	21-30 years	Bangalore	530068	2-3 years	11-20 times	Wi-Fi	Smartphone	4.7 inches	IOS/Mac	Safari	Content Marketing	Via application	more than 15 mins
5	Female	31-40 years	Noida	201308	Above 4 years	41 times and above	Wi-Fi	Smartphone	5.5 inches	Android	Google chrome	Search Engine	Via application	more than 15 mins
6	Male	41-50 yaers	Delhi	110011	Above 4 years	31-40 times	Wi-Fi	Tablet	Others	Android	Google chrome	Display Adverts	Direct URL	more than 15 mins

DATA VISUALIZATION

1. ***Visualizing Age:***

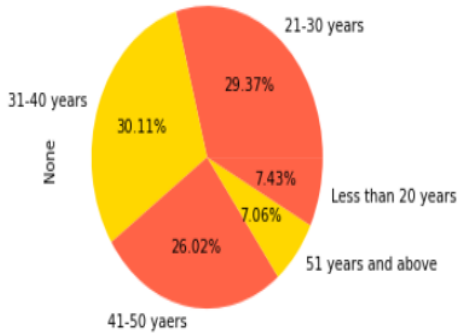
- According to the datasets we can see that age group of 31-40 yrs count is more compare to other values.
- 21-20 yrs are almost equal to 31-40 yrs.
- After 21-30yrs we can see a little flow in 41-50 yrs
- After 41-50years of age the graph drastically decreased for 50 yrs and above.

```
In [13]: sns.countplot(df["Age"])
Out[13]: <AxesSubplot:xlabel='Age', ylabel='count'>
```



2.Using pie also we can visualize our datasets. Here we can see the values of the age what percentage they are contributing in our datasets

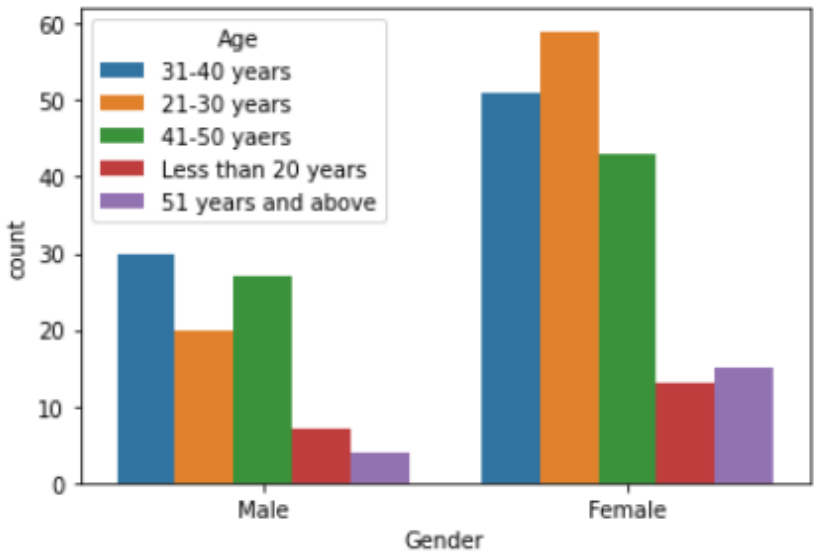
```
In [14]: df.groupby('Age').size().plot(kind='pie', autopct='%.2f%%',colors=['tomato','gold'])
Out[14]: <AxesSubplot:ylabel='None'>
```



3. ***Visualizing Gender:***

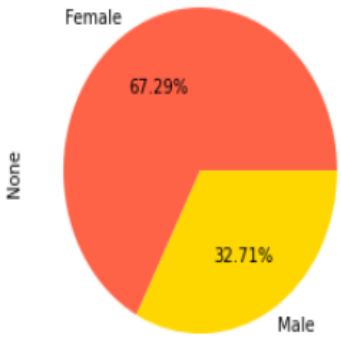
- Here we can observe Female contribution is more compare to Males in the datasets.
- Female graph shows how value for 31-40yrs has decreased and then sudden increase.
- But after 21-20yrs age group we can see fall in the graph.
- In Males we can see first count increased then fall and then increased than decreased. There is a fluctuation.

```
In [17]: sns.countplot(df["Gender"],hue=df['Age'])
Out[17]: <AxesSubplot:xlabel='Gender', ylabel='count'>
```



4. Using pie also we can visualize our datasets. Here we can see the values of the gender what percentage they are contributing in our datasets

```
In [18]: df.groupby('Gender').size().plot(kind='pie', autopct='%.2f%%',colors=['tomato','gold'])
Out[18]: <AxesSubplot:ylabel='None'>
```

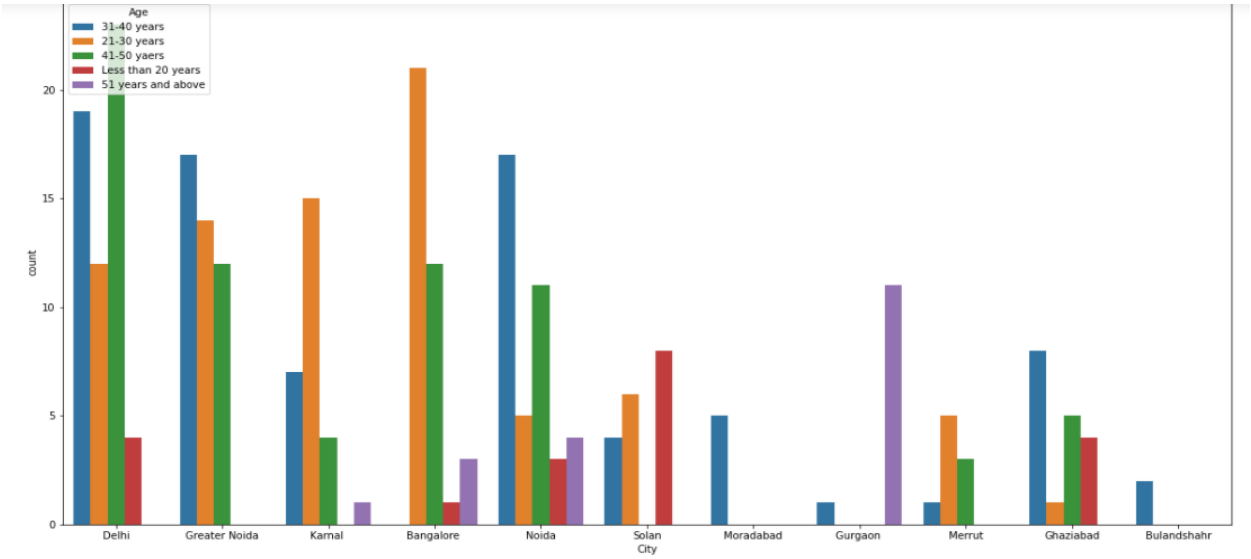


5. ***Visualizing City:***

- *From the graph we can see that Age group of 31-40 yrs give more contribution in any city except three cities like:- Kernal, Gurgao, Meerut.*

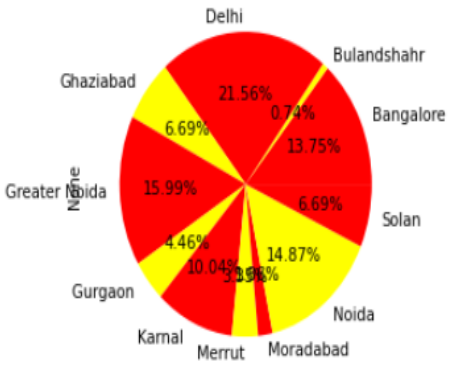
- *In Gurgao and Bulandsher we can see only one line which represent Age group of 31-40yrs.*

```
In [21]: plt.figure(figsize=(20,10))
         sns.countplot(df["City"],hue=df["Age"])
Out[21]: <AxesSubplot:xlabel='City', ylabel='count'>
```



6. Using pie also we can visualize our datasets. Here we can see the values of the City what percentage they are contributing in our datasets

```
In [22]: df.groupby('City').size().plot(kind='pie', autopct='%.2f%%',colors=['red','yellow'])
Out[22]: <AxesSubplot:ylabel='None'>
```

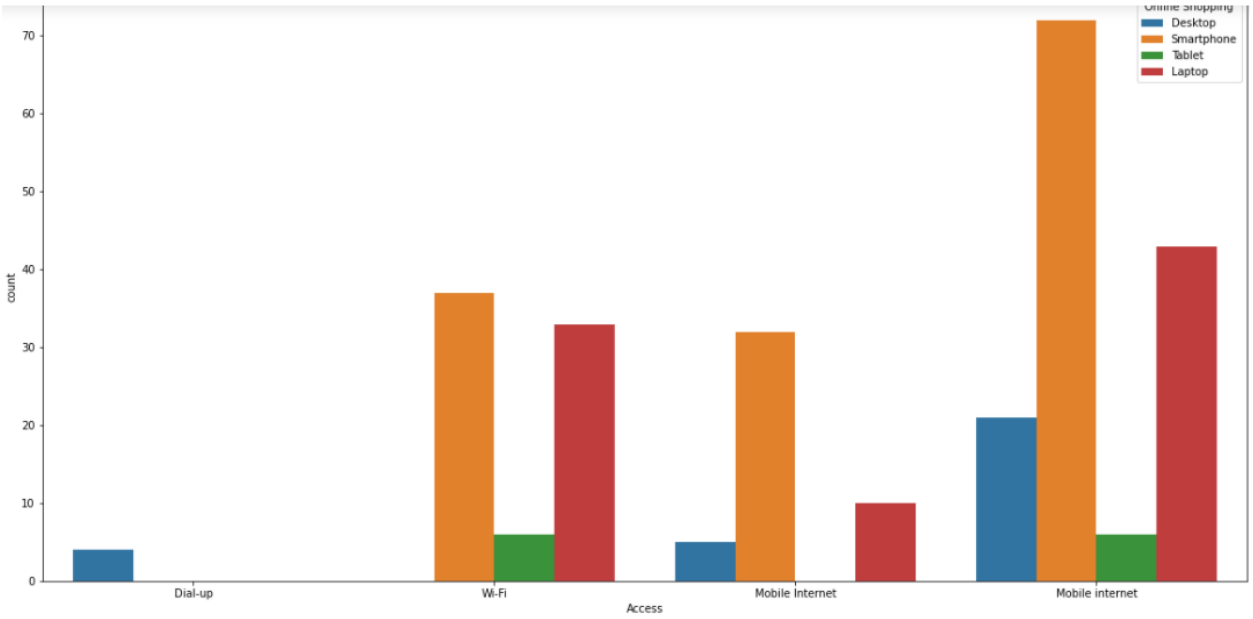


7. ***<u>Visualizing Access with Online Shopping:</u>***

- Here we can see only one line in Dial-up.
- More value is accessed in Mobile-Internet
- We can visualize that nowadays Mobile internet and also Wi-Fi is used more.
- Smart Phone is used more for Mobile Internet

```
In [29]: plt.figure(figsize=(20,10))
         sns.countplot(df["Access"],hue=df["Online Shopping"])
```

```
Out[29]: <AxesSubplot:xlabel='Access', ylabel='count'>
```

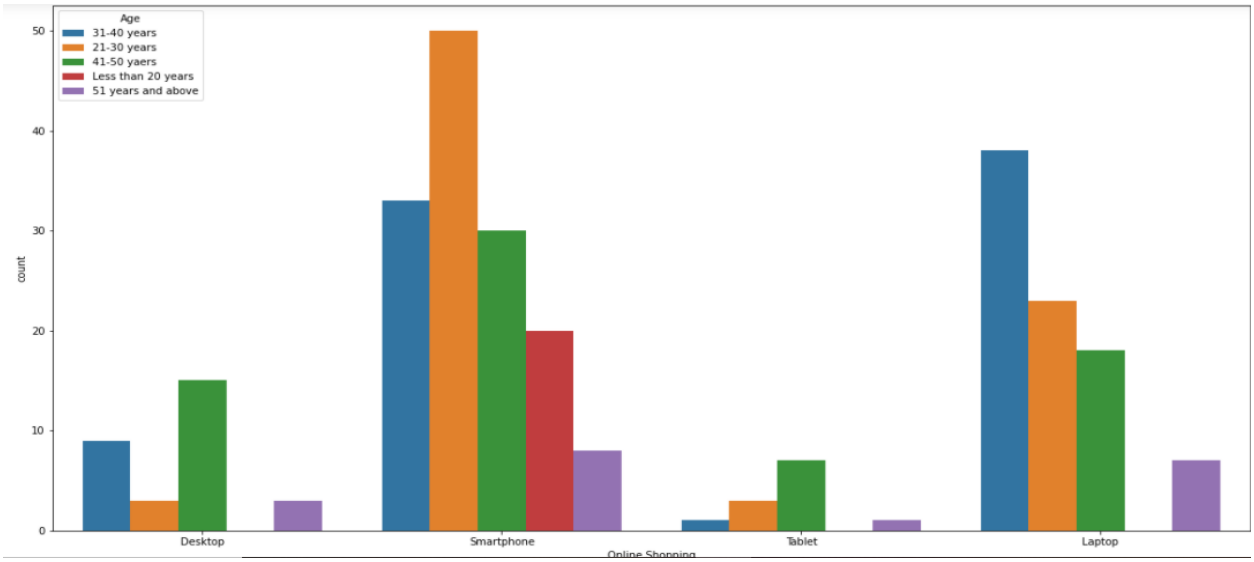


8. ***Visualizing Online Shopping with Age***:

- Here we can observe that 21-30years of age group do more shopping usinng smart phone
- And also we can find out using Tablet only age group of 41-50 yrs do shopping.But ratio of 31-40 and 50 above is same in the least cateogry in shopping.

```
In [28]: plt.figure(figsize=(20,10))
         sns.countplot(df["Online Shopping"],hue=df["Age"])
```

```
Out[28]: <AxesSubplot:xlabel='Online Shopping', ylabel='count'>
```



9. Visualizing City,Year and Gender:

- The line graph here represents Male and Female values.

  Male:-1) Here we can see Male from last 3-4 years are doing more shopping in Delhi.Greater Noida,Banaglore

- 2) And it falls in Bulandshahr

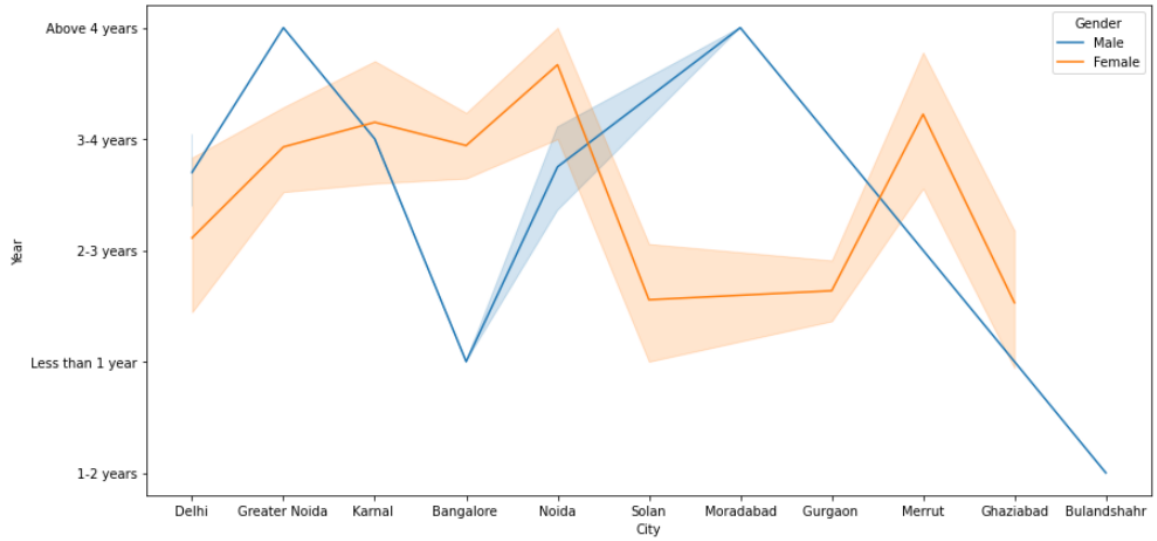
  Female:-1) from last 3-4 yrs femail shopping is increased in Banglore and noida

- 2)and drasctically it dcreased in last 2-3 yrs in solan city and gurgaon

- 3)again it increased and than decreased.

```
In [32]: plt.figure(figsize=(14,7))
         sns.lineplot(df['City'],df['Year'],hue=df['Gender'])
```

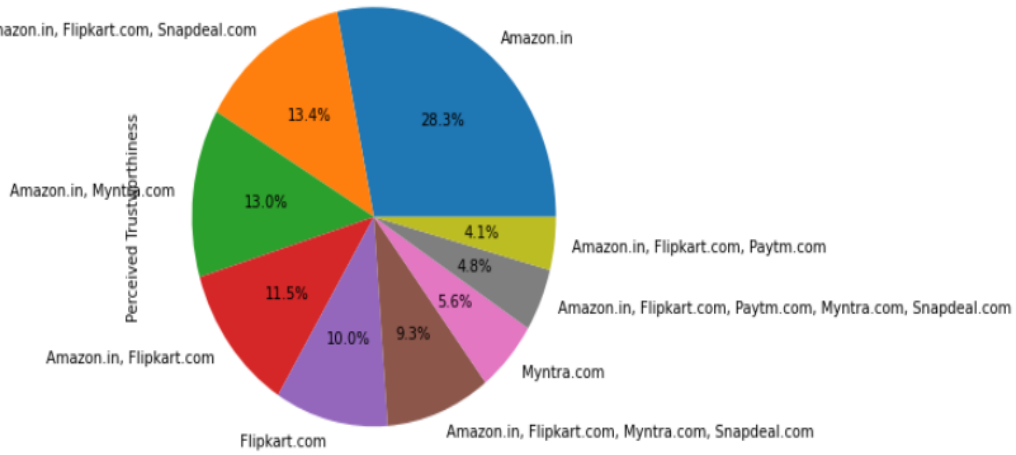
```
Out[32]: <AxesSubplot:xlabel='City', ylabel='Year'>
```



10. Using pie also we can visualize our datasets. Here we can see the values of the trustworthy what percentage they are contributing in our datasets

```
In [33]: df['Perceived Trustworthiness'].value_counts().plot.pie(figsize = (6,6), autopct = '%.1f%%')
```

```
Out[33]: <AxesSubplot:ylabel='Perceived Trustworthiness'>
```



CONCLUSION

- From this visualization here I can conclude that: Retailors can invest more in as follows:
- 1)Female do more shopping. So more things related to female can be added
- 2)People find Amzone more save and worthy for shopping. So, amazon, Flipkart Retailors can use for there products to add more
- 3)City like Banglore, Delhi, Noida is more used for shopping. So, retailors from this place can thing to start a business it will give plus point to their work.
- 4)People age group of 31-40yrs its male or female they do more shopping. So, it can be keep in mind and Retailors can thing allot of ways to in hence there business
- 5) Customers satisfaction play a major role in retention, A company should first understand what customers expect while purchasing online (e-commerce) and build a better buying experience which will in turn retain the customer. An

unforgettable experience is what drivers customers to buy again and again.