

1 Project Report: Customer Retention Project

2 By:Himanshu Soni| Fliprobo Internship11| Batch 1824 – Datatrained |□

3 Problem Statement: -

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

- Customer satisfaction has emerged as one of the most important factors that guarantee the success of online store; it has been posited as a key stimulant of purchase, repurchase intentions and customer loyalty. A comprehensive review of the literature, theories and models have been carried out to propose the models for customer activation and customer retention. Five major factors that contributed to the success of an e-commerce store have been identified as: service quality, system quality, information quality, trust and net benefit. The research furthermore investigated the factors that influence the online customers repeat purchase intention. The combination of both utilitarian value and hedonistic values are needed to affect the repeat purchase intention (loyalty) positively. The data is collected from the Indian online shoppers. Results indicate the e-retail success factors, which are very much critical for customer satisfaction.

In [1]:

```
## import the data basice Labrary

import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

In [2]:

```
# read the data set
ec_df = pd.read_excel("F:/Datatrain project/Customer_retention_dataset/customer_retention.xlsx")
ec_df.head()
```

	Gender of respondent	2 How old are you?	which city do you shop online from?	where you shop online from?	3 Since Long You are Shopping Online ?	4 made an purchase in the past 1 year?	access the internet while shopping on-line?	o which device do you use to access the online shopping?	is the screen size of your mobile device?
0	Male	31-40 years	Delhi	110009	Above 4 years	31-40 times	Dial-up	Desktop	Others
1	Female	21-30 years	Delhi	110030	Above 4 years	41 times and above	Wi-Fi	Smartphone	4.7 inches
2	Female	21-30 years	Greater Noida	201308	3-4 years	41 times and above	Mobile Internet	Smartphone	5.5 inches
3	Male	21-30 years	Karnal	132001	3-4 years	Less than 10 times	Mobile Internet	Smartphone	5.5 inches

In [3]:

```
ec_df.describe(include="all")
```

Out[3]:

	1Gender of respondent	2How old are you?	3Which city do you shop online from?	4What is the Pin Code of where you shop online from?	5Since How Long You are Shopping Online ?	6How many times you have made an online purchase in the past 1 year?	7How do you access the internet while shopping on-line?
count	269	269	269	269.000000	269	269	269
unique	2	5	11	NaN	5	6	4
top	Female	31-40 years	Delhi	NaN	Above 4 years	Less than 10 times	Mobile internet
freq	181	81	58	NaN	98	114	142
mean	NaN	NaN	NaN	220465.747212	NaN	NaN	NaN
std	NaN	NaN	NaN	140524.341051	NaN	NaN	NaN
min	NaN	NaN	NaN	110008.000000	NaN	NaN	NaN
25%	NaN	NaN	NaN	122018.000000	NaN	NaN	NaN
50%	NaN	NaN	NaN	201303.000000	NaN	NaN	NaN
75%	NaN	NaN	NaN	201310.000000	NaN	NaN	NaN
max	NaN	NaN	NaN	560037.000000	NaN	NaN	NaN

11 rows × 71 columns

limit_output extension: Maximum message size of 10000 exceeded with 14596 characters

In [4]:

```
# read the data set
ec_df_1 = pd.read_excel("F:/Datatrain project/Customer_retention_dataset/customer_reten")
ec_df_1.head()
```

Out[4]:

1Gender of respondent	2How old are you?	3How you shop online from?	4What is the Pin Code of where you shop online from?	5Since How Long You are Shopping Online ?	6How many times you have made an online purchase in the past 1 year?	7How do you access the internet while shopping on-line?	8Which device do you use to access the online shopping?
0	0	3	Delhi	110009	5	4	4
1	1	2	Delhi	110030	5	5	2
2	1	2	Greater Noida	201308	4	5	3
3	0	2	Karnal	132001	4	1	3
4	1	2	Bangalore	530068	3	2	2

5 rows × 71 columns

In [5]:

```
# Let's check all the columns name and basic info
ec_df.info()

## as we found that here is 71 coulmns and 269 rows are present in the dataset
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 269 entries, 0 to 268
Data columns (total 71 columns):
 #   Column          Non-Null Count  Dtype  
 ---  --  
 0   1Gender of respondent      269 non-null    object 
 1   2 How old are you?        269 non-null    object 
 2   3 Which city do you shop online from? 269 non-null    object 
 3   4 What is the Pin Code of where you shop online from? 269 non-null    int64  
 4   5 Since How Long You are Shopping Online ? 269 non-null    object 
 5   6 How many times you have made an online purchase in the past 1 year? 269 non-null    object 
 6   7 How do you access the internet while shopping on-line? 269 non-null    object 
 7   8 Which device do you use to access the online shopping? 269 non-null    object 
 8   9 What is the screen size of your mobile device? 269 non-null    object 
 9   10 What is the operating system (OS) of your device ? 269 non-null   object  
 10  11 What browser do you run on your device to access the website? 269 non-null   object 
 11  12 Which channel did you follow to arrive at your favorite online store for the first time? 269 non-null   object 
 12  13 After first visit, how do you reach the online retail store? 269 non-null   object 
 13  14 How much time do you explore the e- retail store before making a purchase decision? 269 non-null   object 
 14  15 What is your preferred payment Option? 269 non-null   object 
 15  16 How frequently do you abandon (selecting an items and leaving without making payment) your shopping cart? 269 non-null   object 
 16  17 Why did you abandon the "Bag", "Shopping Cart"? 269 non-null   object 
 17  18 The content on the website must be easy to read and understand 269 non-null   object 
 18  19 Information on similar product to the one highlighted is important for product comparison 269 non-null   object 
 19  20 Complete information on listed seller and product being offered is important for purchase decision. 269 non-null   object
```

```

20 21 All relevant information on listed products must be stated clearly
269 non-null    object
21 22 Ease of navigation in website
269 non-null    object
22 23 Loading and processing speed
269 non-null    object
23 24 User friendly Interface of the website
269 non-null    object
24 25 Convenient Payment methods
269 non-null    object
25 26 Trust that the online retail store will fulfill its part of the transaction at the stipulated time
269 non-null    object
26 27 Empathy (readiness to assist with queries) towards the customers
269 non-null    object
27 28 Being able to guarantee the privacy of the customer
269 non-null    object
28 29 Responsiveness, availability of several communication channels (email, online rep, twitter, phone etc.)
269 non-null    object
29 30 Online shopping gives monetary benefit and discounts
269 non-null    object
30 31 Enjoyment is derived from shopping online
269 non-null    object
31 32 Shopping online is convenient and flexible
269 non-null    object
32 33 Return and replacement policy of the e-tailer is important for purchase decision
269 non-null    object
33 34 Gaining access to loyalty programs is a benefit of shopping online
269 non-null    object
34 35 Displaying quality Information on the website improves satisfaction of customers
269 non-null    object
35 36 User derive satisfaction while shopping on a good quality website or application
269 non-null    object
36 37 Net Benefit derived from shopping online can lead to users satisfaction
269 non-null    object
37 38 User satisfaction cannot exist without trust
269 non-null    object
38 39 Offering a wide variety of listed product in several category
269 non-null    object
39 40 Provision of complete and relevant product information
269 no

```

limit_output extension: Maximum message size of 10000 exceeded with 17386 characters

4 now we change the name of the columns in the dataset and clean it

In [6]:

```
▼ colnames=['Gender','Age','City','Pincode','Customer shopping online (in Years)', 'Purchase  
'Screen Size', 'Operating System', 'Browser', 'Reference', 'Source after first visit'  
'Preferred Payment Method', 'Keep items in cart', "Reason for keeping items in cart  
'Product Recommendation', 'Description of recommended product must be accurate', 'Descripti  
'Ease of Navigation', 'Loading Speed', 'User Friendly', 'Convinient Payment methods', 'Trust  
'Privacy of customers', 'Availability of communication channles', 'Discounts', 'Customer Enj  
'Return policy', 'Loyalty Programs', 'Satisfaction due to quality description', 'Customer sa  
'User satisfaction cannot exist without trust', 'Offering a wide variety of listed product  
'Provision of complete and relevant product information', 'Monetary savings', 'The Convenie  
'Shopping on the website gives you the sense of adventure', 'Shopping on your preferred e-  
'You feel gratification shopping on your favorite e-tailer', 'Shopping on the website help  
'Getting value for money spent', 'Online Retailers shopped from', 'Easy to use Website', 'Vi  
'Website with Relavent Description', 'Fast loading Website', 'Reliability of the Website', '  
'Availability of several payment options', 'Speedy order delivery', 'Privacy of customers'  
'Perceived Trustworthiness', 'Presence of online assistance through multi-channel', 'Longer  
'Longer time in displaying graphics and photos (promotion, sales period)', 'Late declarati  
'Longer page loading time (promotion, sales period)', 'Limited mode of payment on most pro  
'Change in website/Application design', 'Frequent disruption when moving from one page to  
'Which of the Indian online retailer would you recommend to a friend?']
```

In [7]:

```
len(colnames)
```

Out[7]:

71

In [8]:

```
ec_df.to_csv('E-retail_CRA.csv', header = False) # save the new dataset
```

In [9]:

```
df = pd.read_csv('E-retail_CRA.csv', names=colnames, header=None) ## read the new dataset
```

In [10]:

df

Out[10]:

					Customer shopping online (in Years)	Purchase in last year	Source	Device
	Gender	Age	City	Pincode				
0	Male	31-40 years	Delhi	110009	Above 4 years	31-40 times	Dial-up	Desktop
1	Female	21-30 years	Delhi	110030	Above 4 years	41 times and above	Wi-Fi	Smartphone
2	Female	21-30 years	Greater Noida	201308	3-4 years	41 times and above	Mobile Internet	Smartphone
3	Male	21-30 years	Karnal	132001	3-4 years	Less than 10 times	Mobile Internet	Smartphone
4	Female	21-30 years	Bangalore	530068	2-3 years	11-20 times	Wi-Fi	Smartphone
...
264	Female	21-30 years	Solan	173212	1-2 years	Less than 10 times	Mobile Internet	Smartphone
265	Female	31-40 years	Ghaziabad	201008	1-2 years	31-40 times	Mobile Internet	Smartphone
266	Female	41-50 yaers	Bangalore	560010	2-3 years	Less than 10 times	Mobile internet	Laptop
267	Female	Less than 20 years	Solan	173229	2-3 years	Less than 10 times	Wi-Fi	Smartphone
268	Female	41-50 yaers	Ghaziabad	201009	2-3 years	31-40 times	Mobile Internet	Smartphone

269 rows × 71 columns

5 check the unique value in the data set

In [11]:

```
## check the unique value in the data set  
df.nunique()
```

Out[11]:

Gender	2
Age	5
City	11
Pincode	39
Customer shopping online (in Years)	5
	..
Longer delivery period	6
Change in website/Application design	7
Frequent disruption when moving from one page to another	8
Website is as efficient as before	8
Which of the Indian online retailer would you recommend to a friend?	8
Length: 71, dtype: int64	

In [12]:

```
df.duplicated(keep="first").sum()
```

Out[12]:

166

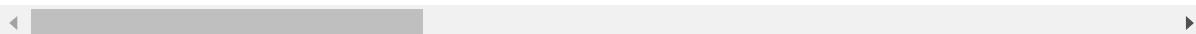
In [14]:

```
# print the duplicate dataset with (keep = False )
df.loc[df.duplicated(keep=False)]
```

Out[14]:

	Gender	Age	City	Pincode	Customer shopping online (in Years)	Purchase in last year	Source	Device
0	Male	31-40 years	Delhi	110009	Above 4 years	31-40 times	Dial-up	Desktop
1	Female	21-30 years	Delhi	110030	Above 4 years	41 times and above	Wi-Fi	Smartphone
2	Female	21-30 years	Greater Noida	201308	3-4 years	41 times and above	Mobile Internet	Smartphone
3	Male	21-30 years	Karnal	132001	3-4 years	Less than 10 times	Mobile Internet	Smartphone
4	Female	21-30 years	Bangalore	530068	2-3 years	11-20 times	Wi-Fi	Smartphone
...
264	Female	21-30 years	Solan	173212	1-2 years	Less than 10 times	Mobile Internet	Smartphone
265	Female	31-40 years	Ghaziabad	201008	1-2 years	31-40 times	Mobile Internet	Smartphone
266	Female	41-50 yaers	Bangalore	560010	2-3 years	Less than 10 times	Mobile internet	Laptop
267	Female	Less than 20 years	Solan	173229	2-3 years	Less than 10 times	Wi-Fi	Smartphone
268	Female	41-50 yaers	Ghaziabad	201009	2-3 years	31-40 times	Mobile Internet	Smartphone

226 rows × 71 columns



6 Univariate Analysis

In [15]:

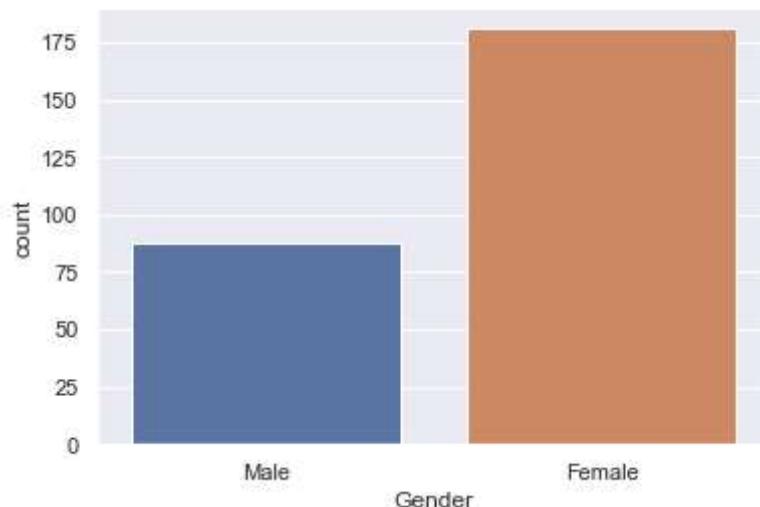
```
# lets check the count of the Female and Male in the data set  
df["Gender"].value_counts()
```

Out[15]:

```
Female    181  
Male      88  
Name: Gender, dtype: int64
```

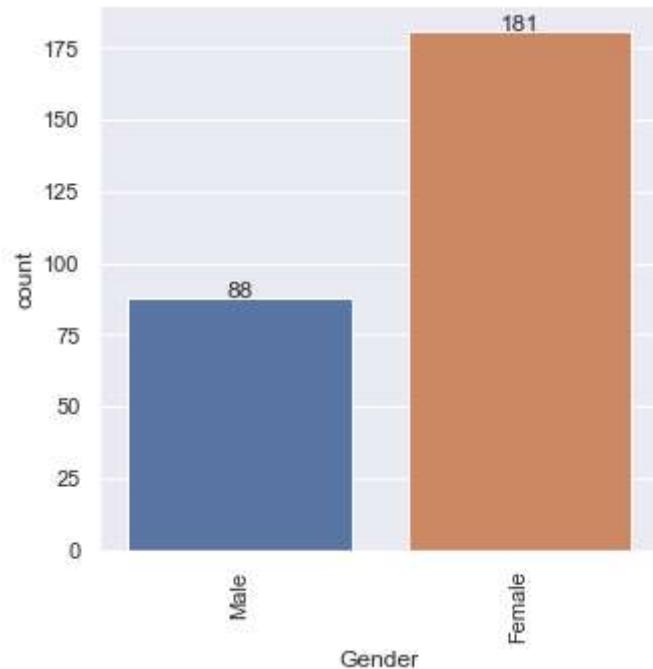
In [16]:

```
## countplot  
sns.set_theme(style="darkgrid")  
ax = sns.countplot(x="Gender", data=df)
```



In [17]:

```
# Lets Display Count on top of countplot
fig, ax1 = plt.subplots(figsize=(5,5))
graph = sns.countplot(ax=ax1,x="Gender", data=df)
graph.set_xticklabels(graph.get_xticklabels(),rotation=90)
for p in graph.patches:
    height = p.get_height()
    graph.text(p.get_x()+p.get_width()/2., height + 0.1, height ,ha="center")
```



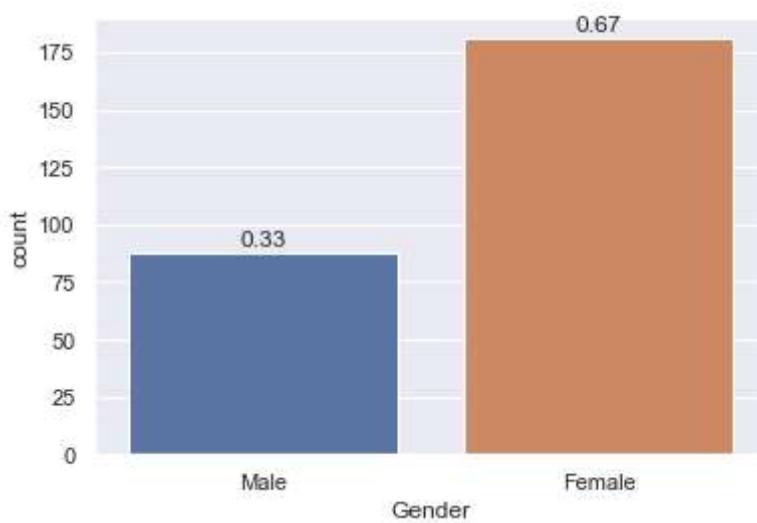
6.1 Here we check that there is more female (181) as compare with the mail (88)

7 If we want to display the percentage

In [20]:

```
# If we want to display the percentage
from matplotlib.pyplot import show

total = float(len(df)) # one person per row
#ax = sns.barplot(x="Gender of respondent", hue="who", data=ec_df)
#ax = sns.countplot(x="Gender of respondent", hue="who", data=ec_df) # for Seaborn vers
ax = sns.countplot(x="Gender", data=df) # for Seaborn version 0.7 and more
for p in ax.patches:
    height = p.get_height()
    ax.text(p.get_x() + p.get_width()/2.,
            height + 3,
            '{:1.2f}'.format(height/total),
            ha="center")
show()
```



7.1 As we see above there is 33% Male(88) and 67% of the female (181) are in dataset

7.2 * female are more active in the shopping as compare to the Mail

In [21]:

df.columns

Out[21]:

```
Index(['Gender', 'Age', 'City', 'Pincode',
       'Customer shopping online (in Years)', 'Purchase in last year',
       'Source', 'Device', 'Screen Size', 'Operating System', 'Browser',
       'Reference', 'Source after first visit', 'Time to make purchase',
       'Preferred Payment Method', 'Keep items in cart',
       'Reason for keeping items in cart', 'Ease of Content',
       'Product Recommendation',
       'Description of recommended product must be accurate',
       'Description of highlighted product must be accurate',
       'Ease of Navigation', 'Loading Speed', 'User Friendly',
       'Convinient Payment methods', 'Trust for Delivery Time', 'User Suppor
t',
       'Privacy of customers', 'Availability of communication channles',
       'Discounts', 'Customer Enjoyment', 'Customer Convinience',
       'Return policy', 'Loyalty Programs',
       'Satisfaction due to quality description', 'Customer satisfaction',
       'Net Benefit', 'User satisfaction cannot exist without trust',
       'Offering a wide variety of listed product in several category',
       'Provision of complete and relevant product information',
       'Monetary savings',
       'The Convenience of patronizing the online retailer',
       'Shopping on the website gives you the sense of adventure',
       'Shopping on your preferred e-tailer enhances your social status',
       'You feel gratification shopping on your favorite e-tailer',
       'Shopping on the website helps you fulfill certain roles',
       'Getting value for money spent', 'Online Retailers shopped from',
       'Easy to use Website', 'Visual appealing Website',
       'Website with Variety of products', 'Website with Relavent Descriptio
n',
       'Fast loading Website', 'Reliability of the Website',
       'Quickness to complete purchase',
       'Availability of several payment options', 'Speedy order delivery',
       'Privacy of customers' information',
       'Security of customer financial information',
       'Perceived Trustworthiness',
       'Presence of online assistance through multi-channel',
       'Longer time to get logged in (promotion, sales period)',
       'Longer time in displaying graphics and photos (promotion, sales peri
od),
       '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 period', 'Change in website/Application design',
       'Frequent disruption when moving from one page to another',
       'Website is as efficient as before',
       'Which of the Indian online retailer would you recommend to a frien
d?'],
      dtype='object')
```

7.3 now we check the multiple categorical column in one go by using seaborn countplot

In [22]:

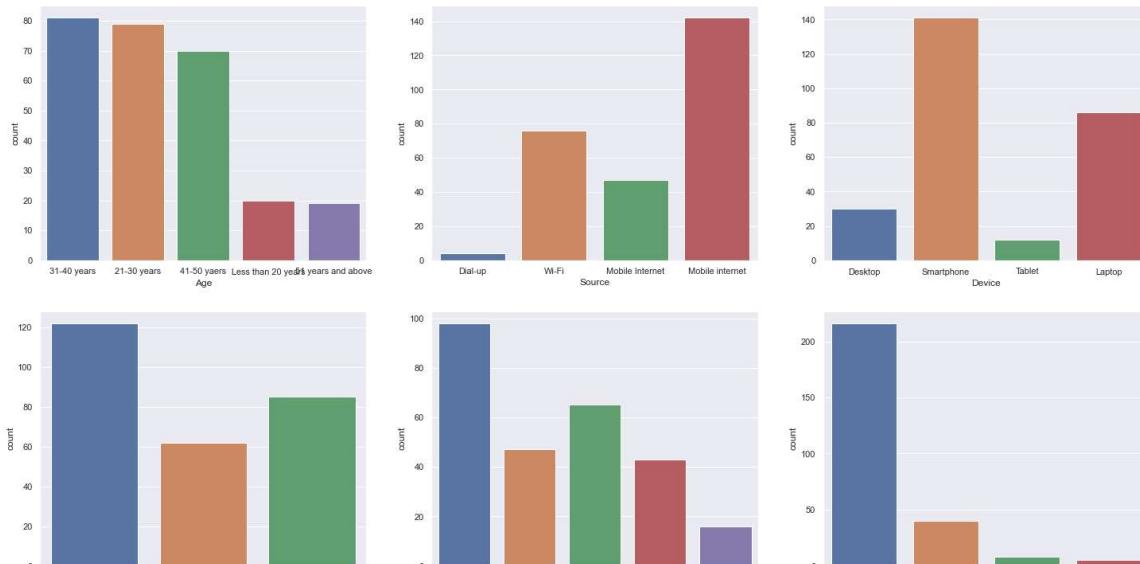
```

#####
import warnings
warnings.simplefilter(action='ignore', category=FutureWarning) # supress future warning
# Combined univariate analysis of each variables
fig,axes = plt.subplots(3,3, figsize=(25,20))
sns.countplot('Age',data=df,ax=axes[0,0])
sns.countplot('Source',data=df,ax=axes[0,1])
sns.countplot('Device',data=df,ax=axes[0,2])
sns.countplot('Operating System',data=df,ax=axes[1,0])
sns.countplot('Customer shopping online (in Years)',data=df,ax=axes[1,1])
sns.countplot('Browser',data=df,ax=axes[1,2])
sns.countplot('Time to make purchase',data=df,ax=axes[2,0])
sns.countplot('Preferred Payment Method',data=df,ax=axes[2,1])
sns.countplot('Discounts',data=df,ax=axes[2,2])

```

Out[22]:

<AxesSubplot:xlabel='Discounts', ylabel='count'>



7.4 we find blow point by above charts

- Age :- Most of the people are come under the 31-40 year active in the online shopping
- source :- Most of the people are using mobile internet for the online shopping
- Device :- Most of the people are using smartphone for the online shopping
- Operating System :- As we found that Most of the people are using Windows for the online shopping
- Customer shopping online (in Years) :- Most of the people are using online shopping more than 4 years
- Browser :- Most of the people are using Google Chrome for the online shopping
- Time to make purchase :- Most of the people are using mobile internet more than 15 mins for the online shopping
- Preferred Payment Method:- Most of the people are using Credit/Debit card for payment in the online shopping

8 column wise slice in pandas

In [23]:

```
## column wise slice in pandas
df.iloc[:, 50:71:1]
```

Out[23]:

	Website with Variety of products	Website with Relavent Description	Fast loading Website	Reliability of the Website	Quickness to complete purchase	Availability of several payment options	Sp c del
0	Flipkart.com	Snapdeal.com	Snapdeal.com	Paytm.com	Paytm.com	Patym.com	Arr
1	Flipkart.com, Myntra.com	Amazon.in, Flipkart.com, Myntra.com	Amazon.in, Flipkart.com, Myntra.com	Mynta.com	Amazon.com, Flipkart.com, Myntra.com	Amazon.in, Flipkart.com, Myntra.com	Am Flipk
2	Amazon.in, Myntra.com	Amazon.in, Paytm.com, Myntra.com	Amazon.in, Paytm.com	Amazon.in, Paytm.com, Myntra.com	Amazon.com, Paytm.com, Myntra.com	Patym.com, Myntra.com	Arr
3	Amazon.in, Flipkart.com	Amazon.in, Flipkart.com	Amazon.in, Flipkart.com, Snapdeal.com	Amazon.in, Flipkart.com, Paytm.com	Amazon.com, Flipkart.com, Paytm.com	Amazon.in, Flipkart.com, Myntra.com	Am Flipk Snapd
4	Mynta.com	Amazon.in, Flipkart.com, Paytm.com, Myntra.com...	Amazon.in	Amazon.in, Paytm.com, Myntra.com	Amazon.com, Flipkart.com, Paytm.com, Myntra.co...	Amazon.in, Flipkart.com, Patym.com, Myntra.com...	Arr
...
264	Amazon.in	Amazon.in	Amazon.in	Amazon.in	Amazon.com	Amazon.in	Arr
265	Amazon.in	Flipkart.com	Flipkart.com	Flipkart.com	Flipkart.com	Flipkart.com	Flipk
266	Amazon.in	Amazon.in	Amazon.in	Amazon.in	Amazon.com	Amazon.in, Flipkart.com	Arr
267	Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com	Flipkart.com, Mynta.com, Snapdeal.com	Flipkart.com, Myntra.com, Snapdeal.com	Flipk Myni Snapd			
268	Amazon.in	Amazon.in	Amazon.in	Amazon.in	Amazon.com	Amazon.in	Arr

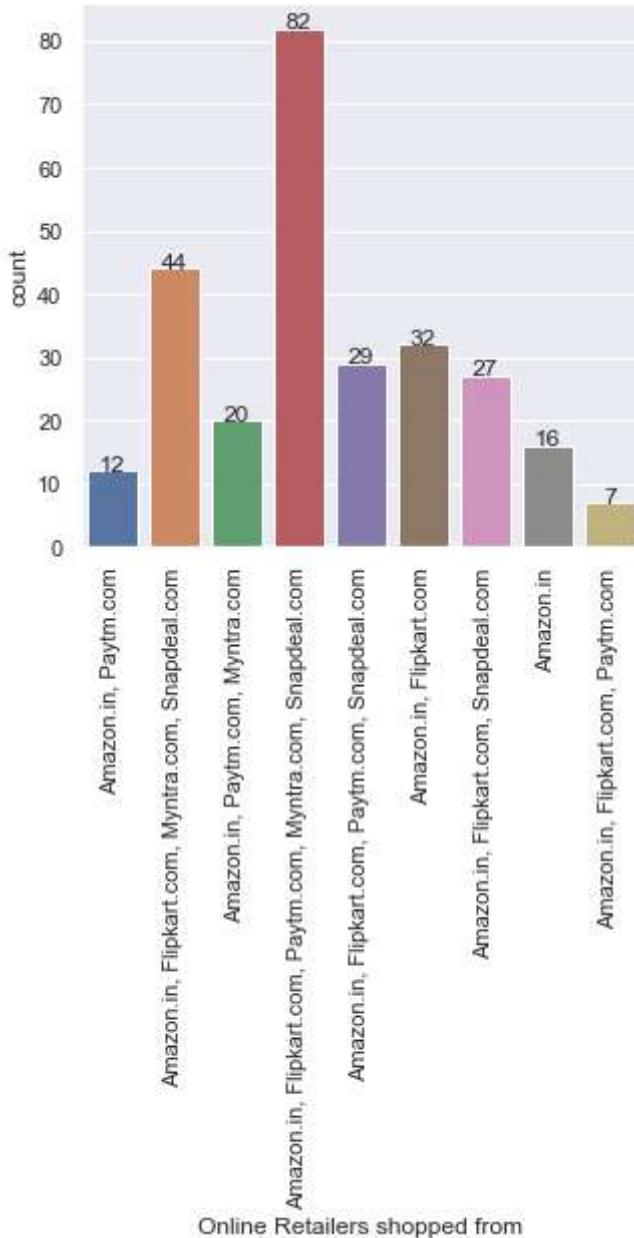
269 rows × 21 columns

limit_output extension: Maximum message size of 10000 exceeded with 13902 characters

9 Lets Display Count on top of Online Retailers shopped from website

In [24]:

```
# Lets Display Count on top of countplot
fig, ax1 = plt.subplots(figsize=(5,5))
graph = sns.countplot(ax=ax1,x="Online Retailers shopped from", data=df)
graph.set_xticklabels(graph.get_xticklabels(),rotation=90)
for p in graph.patches:
    height = p.get_height()
    graph.text(p.get_x()+p.get_width()/2., height + 0.1, height ,ha="center")
```



10 Bivariate Analysis

10.1 we are doing Bivariate Analysis by using two column "Gender" , "City"

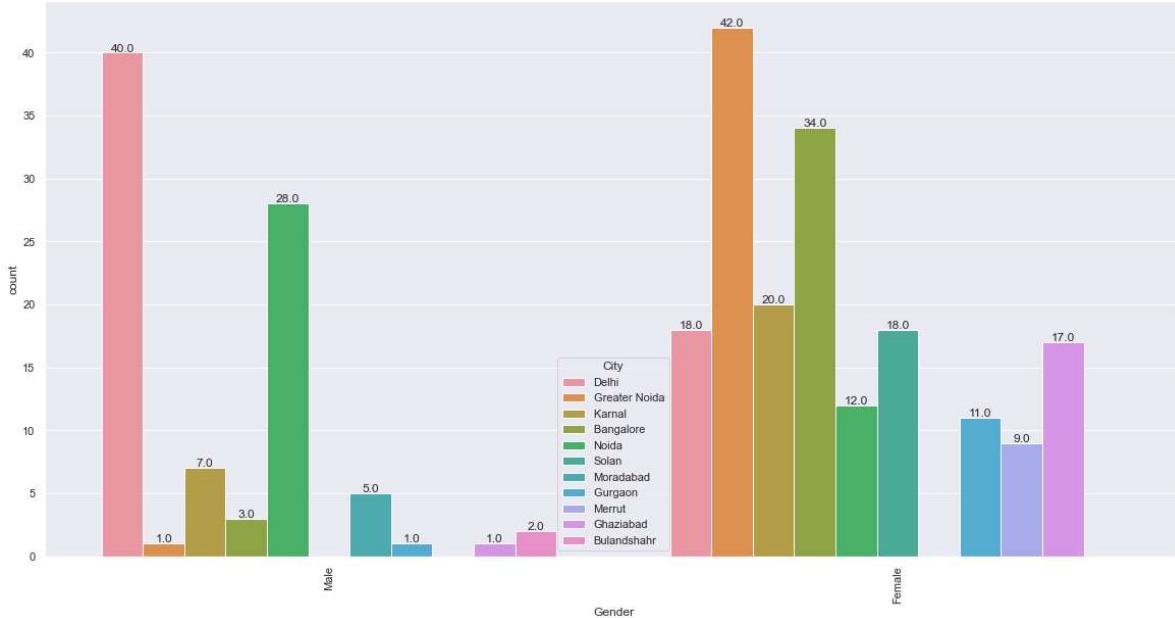
In [25]:

```

fig, ax1 = plt.subplots(figsize=(20,10))
graph = sns.countplot(ax=ax1,x='Gender',hue="City",data=df)
graph.set_xticklabels(graph.get_xticklabels(),rotation=90)
for p in graph.patches:
    height = p.get_height()
    graph.text(p.get_x()+p.get_width()/2., height + 0.1, height ,ha="center")

```

posx and posy should be finite values
 posx and posy should be finite values



10.2 As we check that there is the mostly Delhi's Male(40) and Greater Noida's Female(42) are in the dataset

"""\Categorical scatterplots:

- :func: stripplot (with kind="strip" ; the default)
- :func: swarmplot (with kind="swarm")

Categorical distribution plots:

- :func: boxplot (with kind="box")
- :func: violinplot (with kind="violin")

- :func: boxenplot (with kind="boxen")

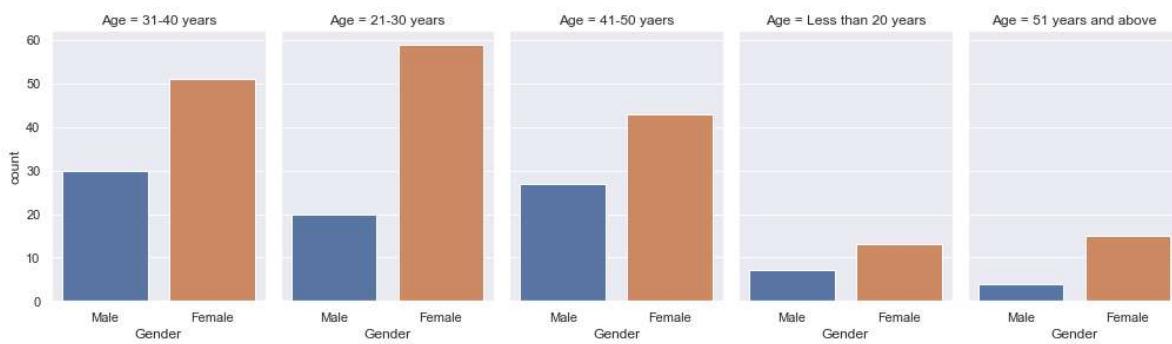
Categorical estimate plots:

- :func: pointplot (with kind="point")
- :func: barplot (with kind="bar")
- :func: countplot (with kind="count")"""

11 we are doing Bivariate Analysis by using two column "Gender" , "Age"

In [26]:

```
graph = sns.catplot(x="Gender", col="Age", data=df, kind="count", height=4, aspect=.7)
```



11.1 as we see there is Female which come under the age of 21-30 years are more active on the online shopping

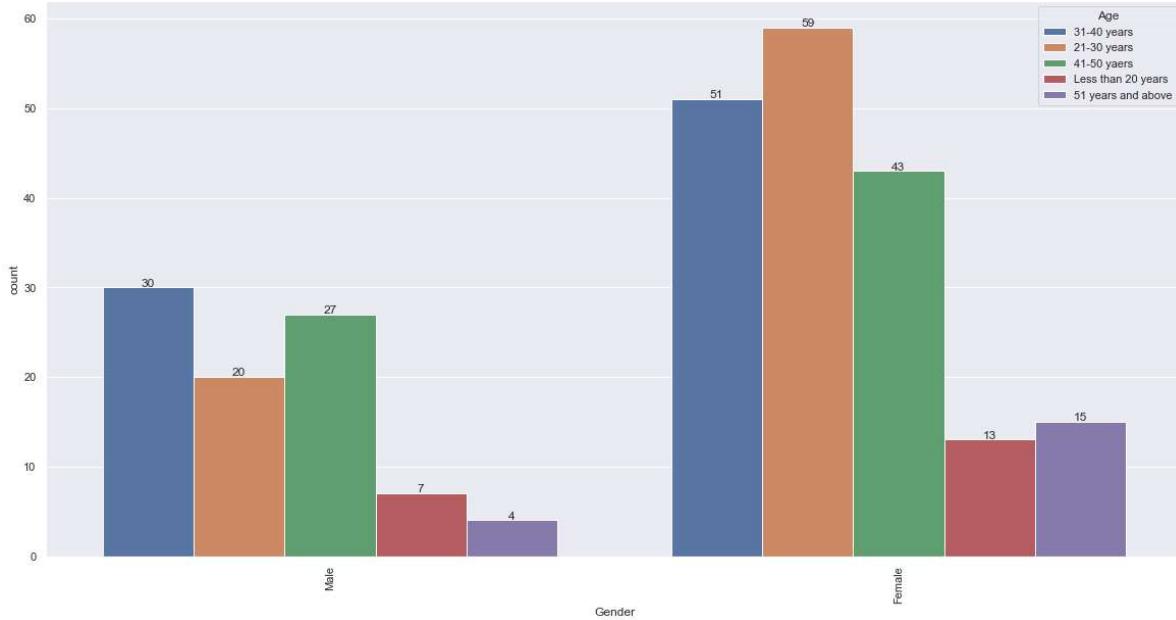
12 we are doing Bivariate Analysis by using two column "Gender" , "Age"

In [27]:

```

fig, ax1 = plt.subplots(figsize=(20,10))
graph = sns.countplot(ax=ax1,x='Gender',hue="Age",data=df)
graph.set_xticklabels(graph.get_xticklabels(),rotation=90)
for p in graph.patches:
    height = p.get_height()
    graph.text(p.get_x()+p.get_width()/2., height + 0.1,height ,ha="center")

```



12.1 As we found that there is 59 Female b/w the age of (21-30 Year) and 30 male b/w the age if 31-40 year

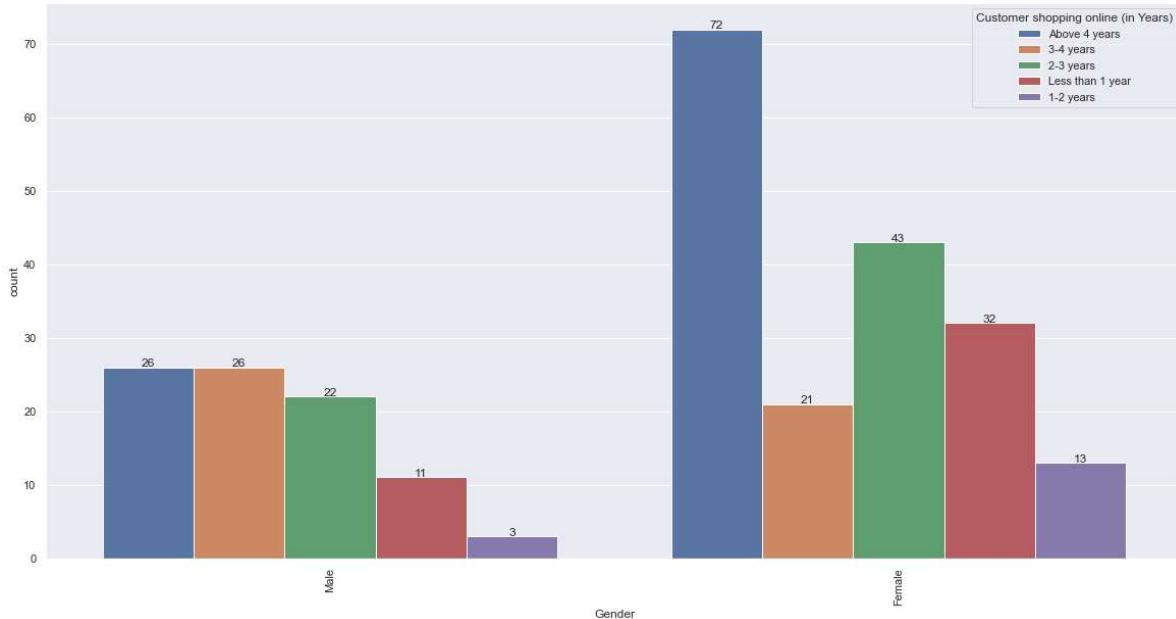
13 we are doing Bivariate Analysis by using two column " Gender" , "Customer shopping online (in Years"

In [28]:

```

fig, ax1 = plt.subplots(figsize=(20,10))
graph = sns.countplot(ax=ax1,x='Gender', hue="Customer shopping online (in Years)",data=d)
graph.set_xticklabels(graph.get_xticklabels(),rotation=90)
for p in graph.patches:
    height = p.get_height()
    graph.text(p.get_x()+p.get_width()/2., height + 0.1, height ,ha="center")

```



13.1 As we found that there is 72 Female are using online shopping last more then 4 year

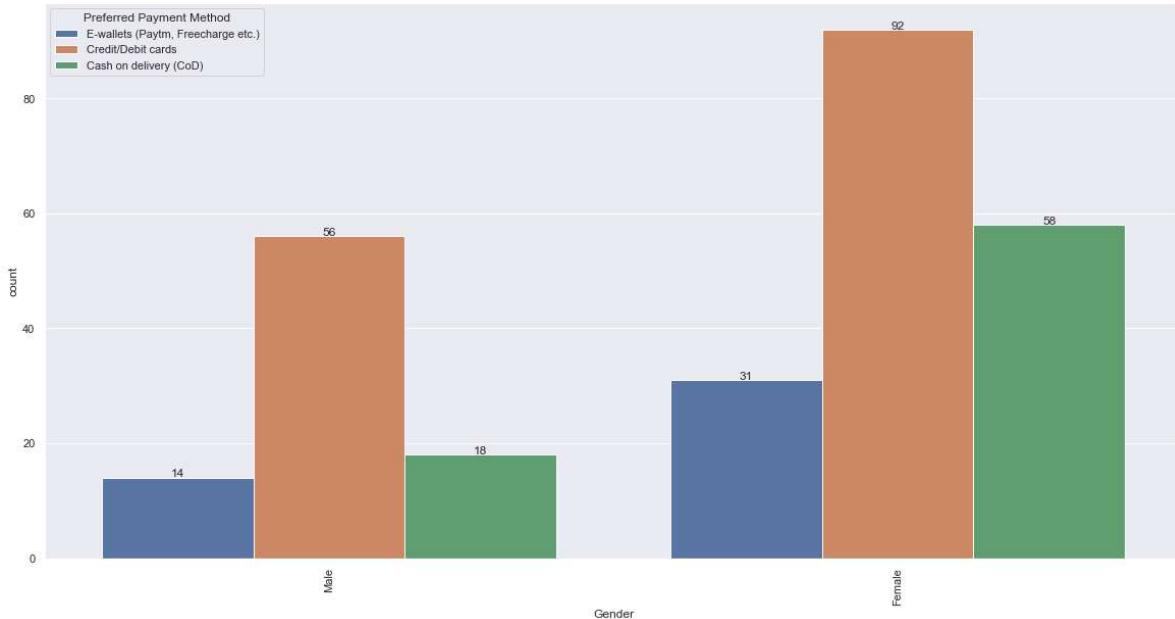
14 we are doing Bivariate Analysis by using two column " Gender" , "Preferred Payment Method"

In [29]:

```

fig, ax1 = plt.subplots(figsize=(20,10))
graph = sns.countplot(ax=ax1,x='Gender', hue="Preferred Payment Method",data=df)
graph.set_xticklabels(graph.get_xticklabels(),rotation=90)
for p in graph.patches:
    height = p.get_height()
    graph.text(p.get_x()+p.get_width()/2., height + 0.1,height ,ha="center")

```



14.1 As we check that most of the people (male(56),female(92)) as using Cards payment

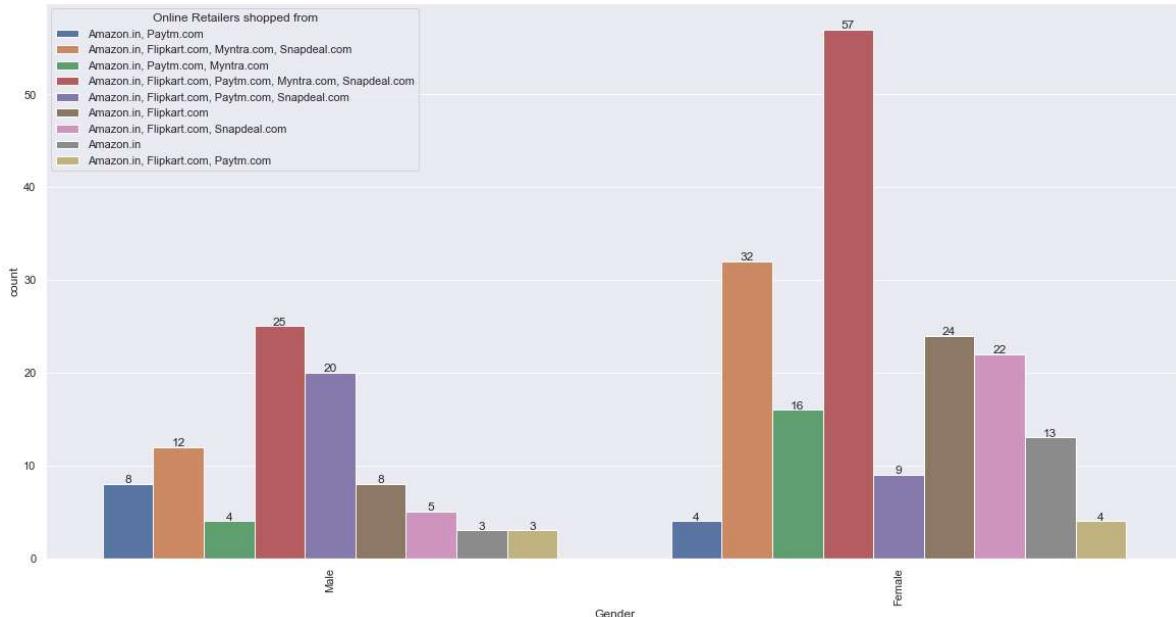
15 we are doing Bivariate Analysis by using two column " Gender" , "Online Retailers shopped from"

In [30]:

```

fig, ax1 = plt.subplots(figsize=(20,10))
graph = sns.countplot(ax=ax1,x='Gender', hue="Online Retailers shopped from", data=df)
graph.set_xticklabels(graph.get_xticklabels(),rotation=90)
for p in graph.patches:
    height = p.get_height()
    graph.text(p.get_x()+p.get_width()/2., height + 0.1, height ,ha="center")

```



16 we are doing Bivariate Analysis by using two column " Gender" , "Purchase in last year"

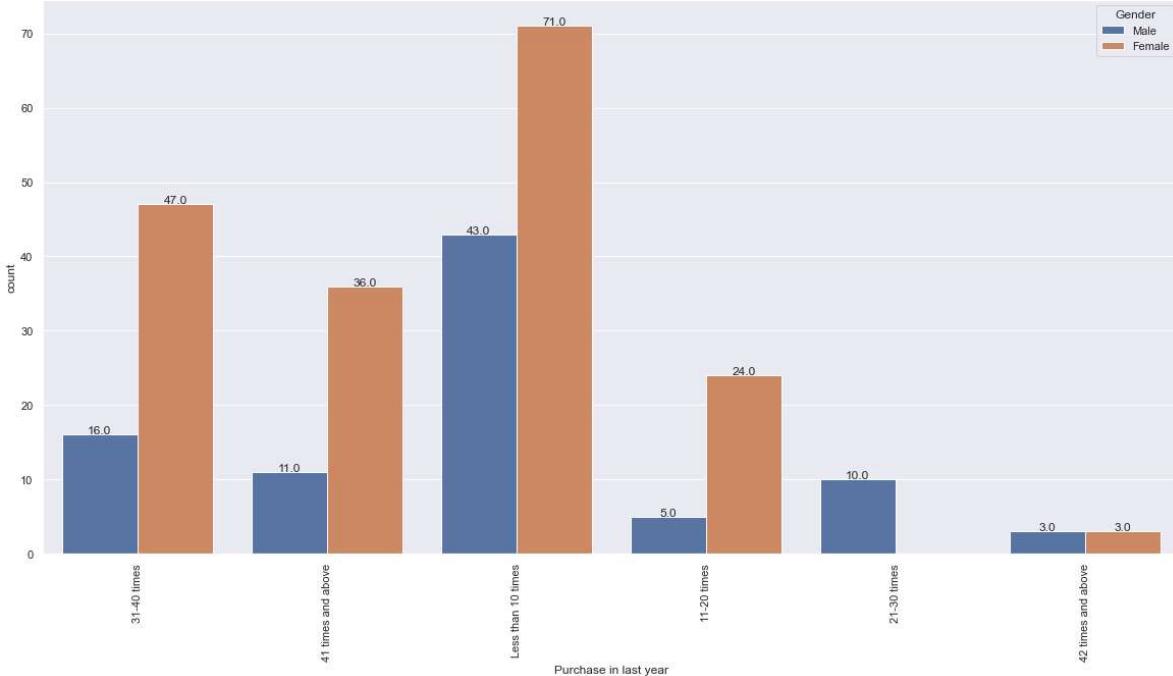
In [31]:

```

fig, ax1 = plt.subplots(figsize=(20,10))
graph = sns.countplot(ax=ax1,x='Purchase in last year', hue="Gender",data=df)
graph.set_xticklabels(graph.get_xticklabels(),rotation=90)
for p in graph.patches:
    height = p.get_height()
    graph.text(p.get_x() + p.get_width()/2., height + 0.1, height ,ha="center")

```

posx and posy should be finite values
 posx and posy should be finite values
 posx and posy should be finite values



In [32]:

```
#graph = sns.catplot(x="Purchase in Last year", col="Gender", data=df_cra, kind="count", hei
```



```
#sns.catplot(data=df_train,x="SibSp",hue = "Sex", kind = "count",height=4)
```



```
g = sns.catplot(data=df,x="Gender",hue = "Preferred Payment Method", kind = "count",height=4)
```

```
g.fig.set_size_inches(12,8)
```

```
g.fig.subplots_adjust(top=0.81,right=0.86)
```



```
ax = g.facet_axis(0,0)
```

```
for p in ax.patches:
```

```
    ax.text(p.get_x() - 0.01,
```

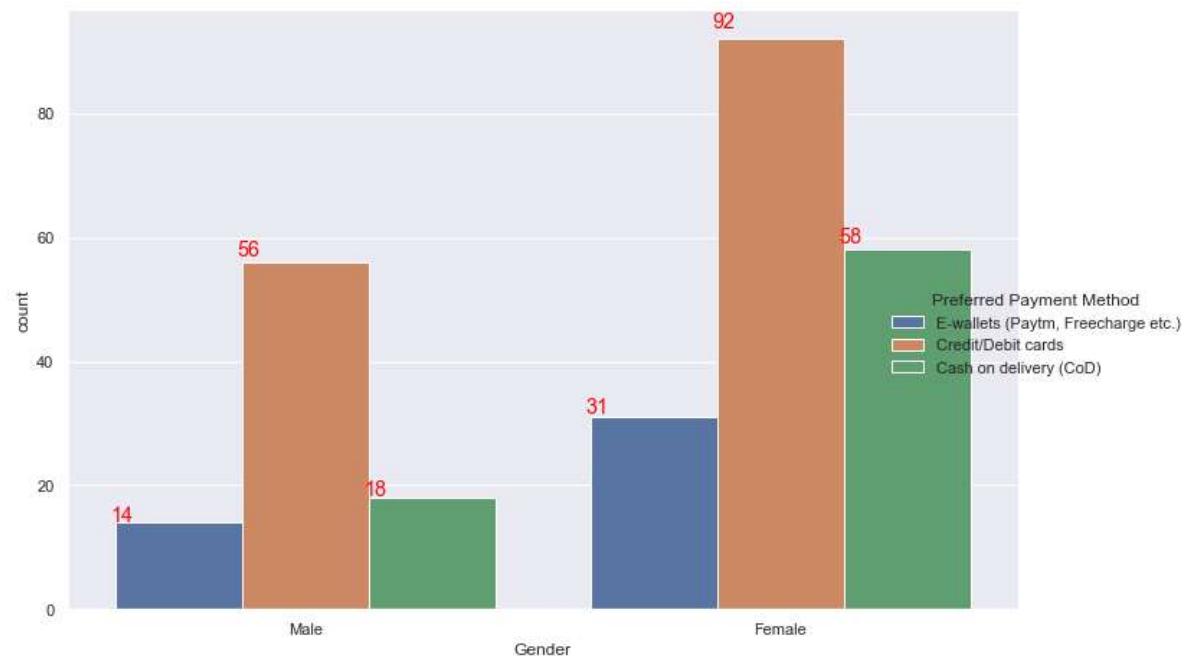
```
            p.get_height() * 1.02,
```

```
            '{0:.0f}'.format(p.get_height()),
```

```
            color='red',
```

```
            rotation='horizontal',
```

```
            size='large')
```



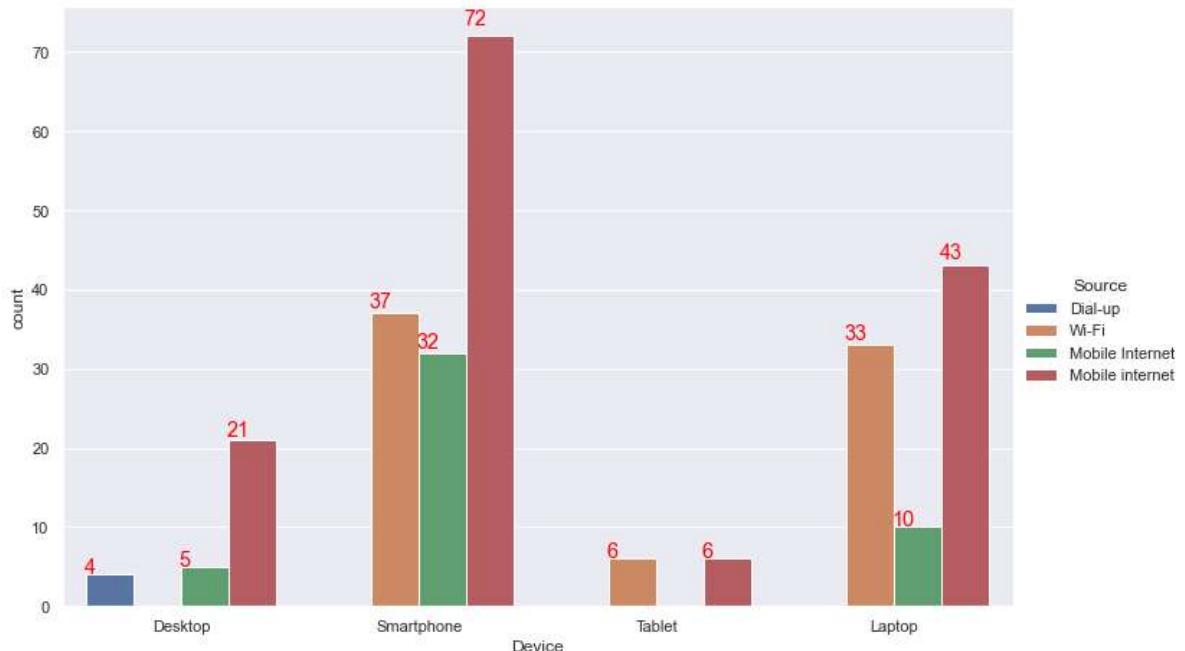
In [33]:

```
#sns.catplot(data=df_train,x="SibSp",hue = "Sex", kind = "count",height=4)

g = sns.catplot(data=df,x="Device",hue = "Source", kind = "count",height=5,legend=True)
g.fig.set_size_inches(12,8)
g.fig.subplots_adjust(top=0.81,right=0.86)

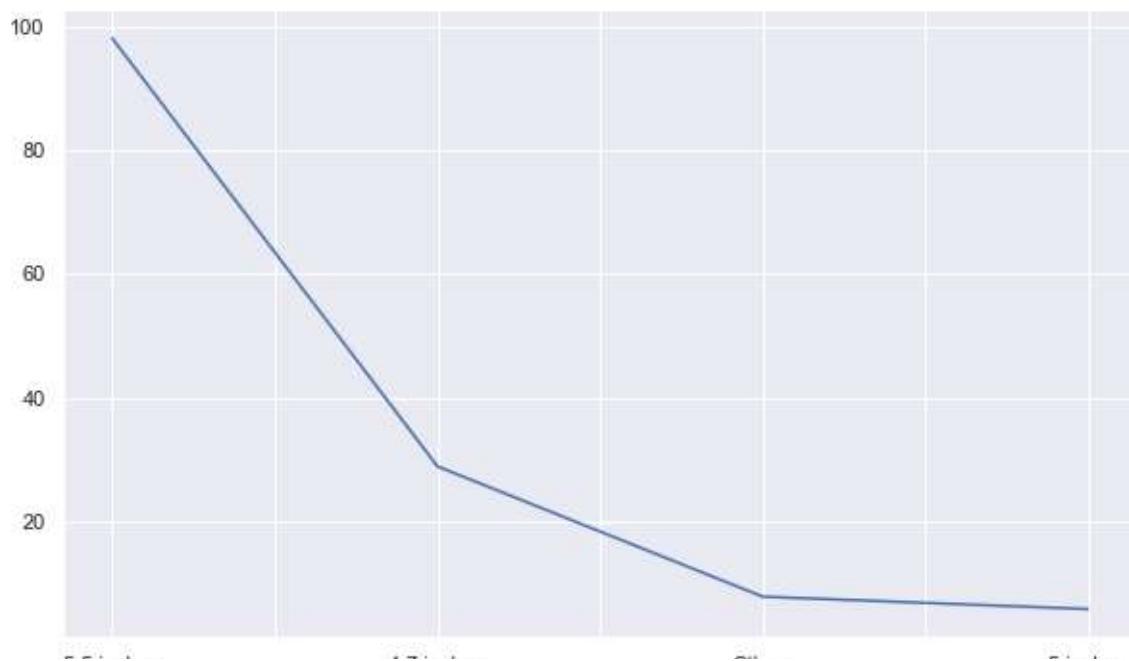
ax = g.facet_axis(0,0)
for p in ax.patches:
    ax.text(p.get_x() - 0.01,
            p.get_height() * 1.02,
            '{0:.0f}'.format(p.get_height()),
            color='red',
            rotation='horizontal',
            size='large')
```

posx and posy should be finite values
 posx and posy should be finite values



In [35]:

```
plt.figure(figsize=(10,6))
df[df['Device']=='Smartphone']['Screen Size'].value_counts().plot()
plt.show()
```



In [36]:

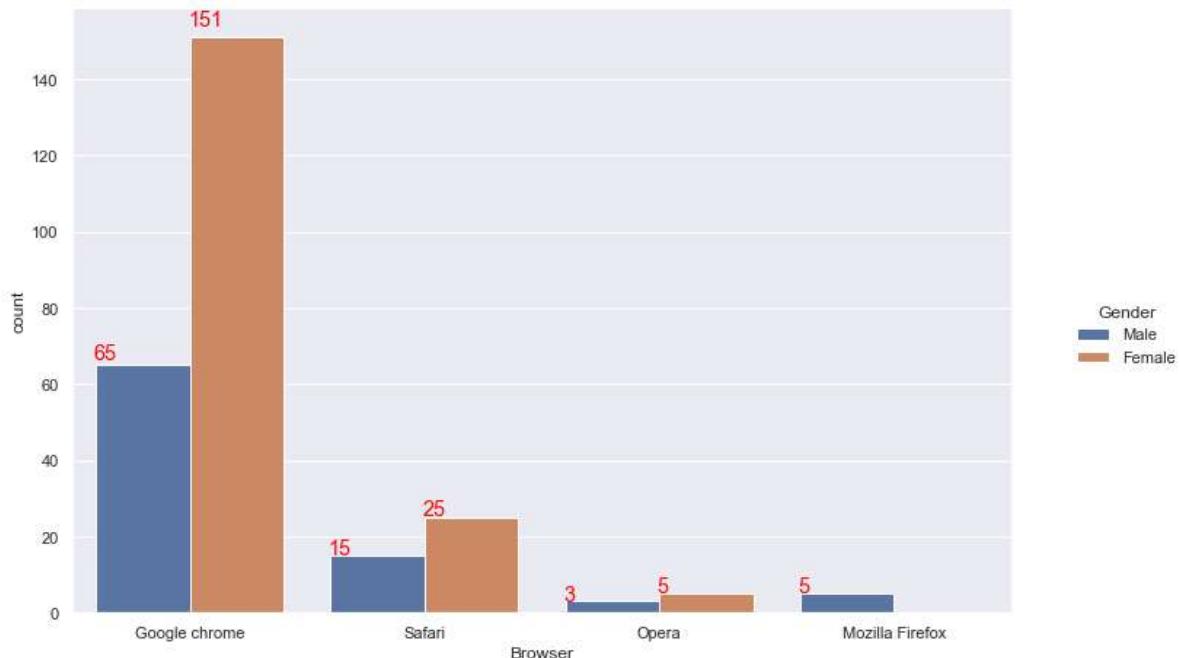
```
g = sns.catplot(data=df,x="Browser",hue = "Gender", kind = "count",height=5,legend=True)
g.fig.set_size_inches(12,8)
g.fig.subplots_adjust(top=0.81,right=0.86)

ax = g.facet_axis(0,0)
for p in ax.patches:
    ax.text(p.get_x() - 0.01,
            p.get_height() * 1.02,
            '{0:.0f}'.format(p.get_height()),
            color='red',
            rotation='horizontal',
            size='large')
```

posx and posy should be finite values

posx and posy should be finite values

posx and posy should be finite values



In []: