1. How is the characteristics of the data? Do exploratory data analysis, calculate descriptive statistics, and visualize it.

```
<bound method DataFrame.info of             Churn  Tenure
PreferredLoginDevice  CityTier  WarehouseToHome  \
CustomerID
50027                0     8.0                1         2              6.0
50028                0     NaN                2         2             12.0
50029                0    18.0                2         0              NaN
50030                0     5.0                0         2             14.0
50031                0     2.0                0         0              6.0
...                ...     ...              ...       ...              ...
55599                1     1.0                0         2             16.0
55603                1     1.0                1         0              8.0
55605                1    20.0                2         0             14.0
55613                1    14.0                0         2              8.0
55622                1    14.0                1         2             35.0

            PreferredPaymentMode  Gender  HourSpendOnApp  DeviceRegistered
\
CustomerID
50027                          5       1             3.0                 2
50028                          5       1             2.0                 2
50029                          4       1             2.0                 2
50030                          5       0             2.0                 2
50031                          1       1             2.0                 2
...                          ...     ...             ...               ...
55599                          5       1             3.0                 3
55603                          3       1             3.0                 3
55605                          1       1             4.0                 3
55613                          4       1             4.0                 3
55622                          5       1             3.0                 4

            PreferedOrderCat  SatisfactionScore  MaritalStatus  \
CustomerID
50027                      0                  3              0
50028                      2                  2              0
50029                      2                  3              1
50030                      0                  1              2
50031                      2                  2              0
...                      ...                ...            ...
55599                      4                  4              1
55603                      4                  0              1
55605                      4                  2              1
55613                      2                  2              1
55622                      0                  4              1
```

```
            NumberOfAddress  Complain  OrderIncreaseFromLastYear
CouponUsed  \
CustomerID
50027                     1         0                       13.0
1.0
50028                     2         1                       20.0
0.0
50029                     8         0                       18.0
1.0
50030                     1         0                       14.0
2.0
50031                     1         0                       13.0
0.0
...                     ...       ...                        ...
...
55599                     2         0                       20.0
2.0
55603                    10         1                       15.0
3.0
55605                     9         0                       12.0
7.0
55613                     8         0                       13.0
2.0
55622                     5         1                       14.0
3.0

            OrderCount  DaySinceLastOrder  CashbackAmount
CustomerID
50027              1.0                6.0          172.95
50028              4.0                5.0          123.06
50029              1.0               15.0          123.48
50030              3.0                7.0          189.98
50031              1.0                9.0          143.19
...                ...                ...             ...
55599              2.0                1.0          142.90
55603              3.0                3.0          172.87
55605             10.0                9.0          148.39
55613              2.0                2.0          192.28
55622              NaN                1.0          233.54

[1896 rows x 19 columns]>
```

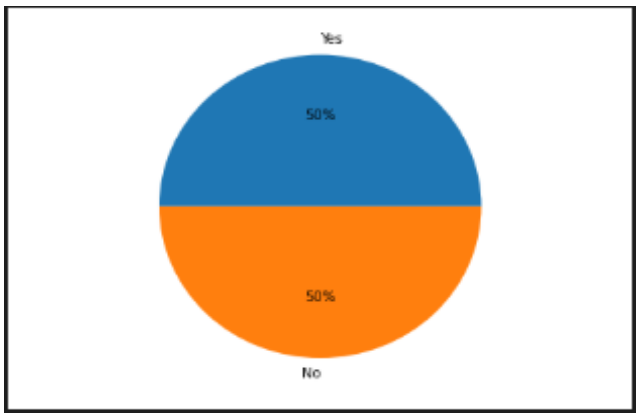
The characteristics of the eccomerce churn data are integer, float, and object types.

REGION
JAWA TIMUR

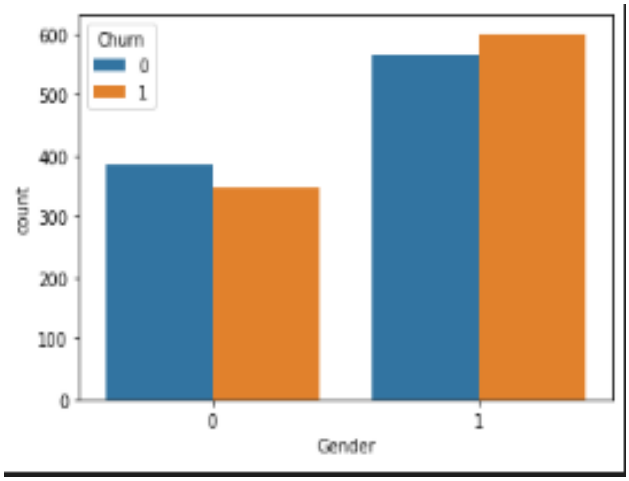
DATA SCIENCE CLUB (DSC)
UNIVERSITAS PGRI ADI BUANA SURABAYA

	Churn	Tenure	CityTier	WarehouseToHome	HourSpendOnApp	DeviceRegistered	SatisfactionScore	NumberOfAddress	Complain	OrderIncreaseFromLastYear	CouponUsed	OrderCount	DaySinceLastOrder	CashbackAmount
count	1896.000000	1739.000000	1896.000000	1744.000000	1766.000000	1896.000000	1896.000000	1896.000000	1896.000000	1855.000000	1817.000000	1837.000000	1812.000000	1896.000000
mean	0.500000	7.347901	1.719409	15.922018	2.682899	3.543776	3.275316	4.088080	0.385549	15.391914	1.470556	2.694066	3.786976	164.907252
std	0.500132	8.149302	0.936148	8.498368	0.679286	1.015023	1.269551	2.694888	0.486853	3.695976	1.862077	2.866878	3.540237	44.698011
min	0.000000	0.000000	1.000000	5.000000	0.000000	1.000000	1.000000	1.000000	0.000000	11.000000	0.000000	1.000000	0.000000	0.000000
25%	0.000000	1.000000	1.000000	9.000000	2.000000	3.000000	2.000000	2.000000	0.000000	12.000000	0.000000	1.000000	1.000000	132.940000
50%	0.500000	4.000000	1.000000	14.000000	3.000000	3.000000	3.000000	3.000000	0.000000	14.000000	1.000000	2.000000	3.000000	150.870000
75%	1.000000	13.000000	3.000000	22.000000	3.000000	4.000000	4.000000	6.000000	1.000000	18.000000	2.000000	3.000000	7.000000	181.610000
max	1.000000	50.000000	3.000000	36.000000	4.000000	6.000000	5.000000	21.000000	1.000000	26.000000	16.000000	16.000000	46.000000	323.590000

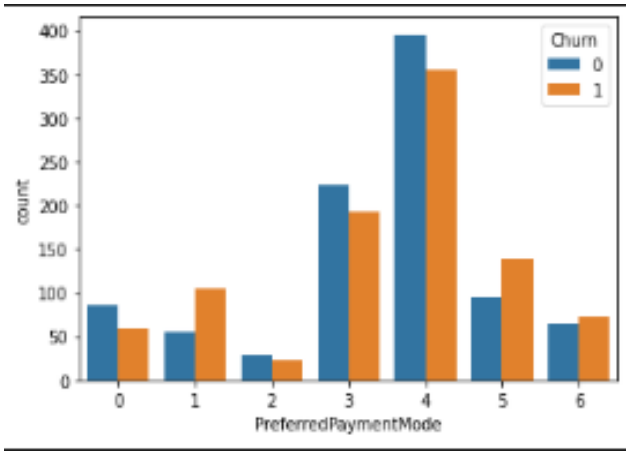
From the table above, it can be seen descriptive statistics by knowing count, mean, standard deviation, minimum and maximum values and we can see that the longest tenure is 50 months and the maximum cashback amount is $323,59. The minimum cashback amount is about $0. The customer can expect to have a cashback amount of about $164,91. I am assuming the charges are in United States Dollars (USD).



Based on the pie chart above, it can be concluded that the distribution of customer data is balanced between those who do not churn and churn, with churn details as much as 50% and no churn as much as 50%.



From the plot above, it looks like gender does not play a role in customer churn. Let's visualize the churn count for Preferred Payment Mode



The chart above is interesting, as it helps me to differentiate between retained and churned customers, it shows that most of the customers make the preferred payment method is CC while the less used one is COD method.

2. Please Do preprocessing data. Is there any missing values or outliers? If yes, solve it and give some explanation. Do variable selection or dimension reduction if needed and give some explanation.

   In the data there are missing values and outliers. The way handle missing values by means of missing values will be filled with the average of the column, while handling outliers is by normalizing the data

3. Find the best model and evaluate the model.