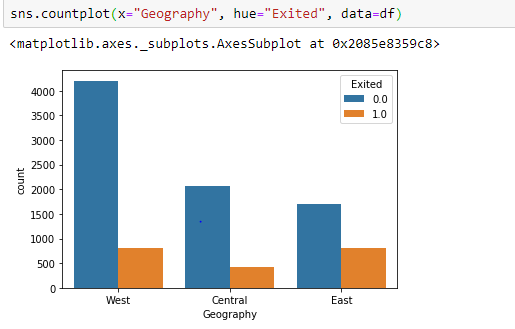
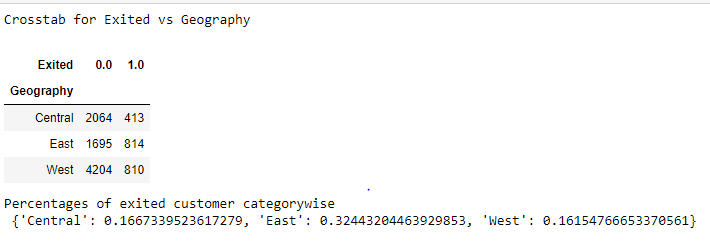
**Customer Churn Analysis**

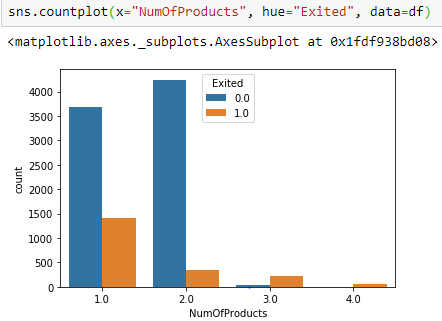
* Initially all excel files are loaded as dataframes and each of dataframes are merged using inner join with key as customerid to form dataframe df.
* Removed duplicates from final dataframe df.
* Converting some independent continuous columns into categorical columns has they discrete values.
* We have found out null values, skewness etc..
* We could see there is no much skewness, data is normally distributed for most of the columns except Age column which sightly higher compare to remaining columns.
* Replaced missing values with mean as there is no problem with skewness.
* Now plotting categorical features on countplot and finding out percentage of exited customer from various categories.

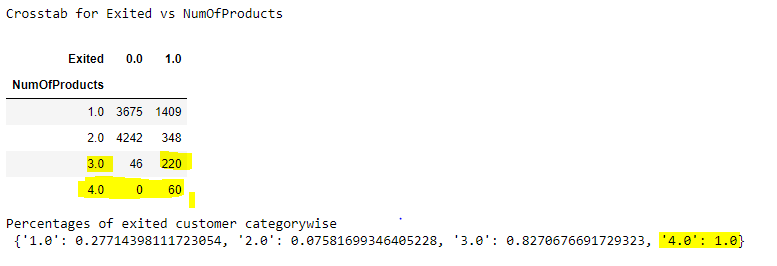
Geography column :





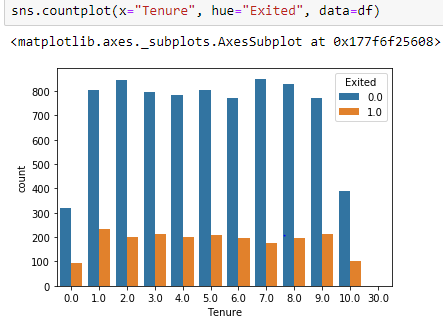
NumOfProducts:

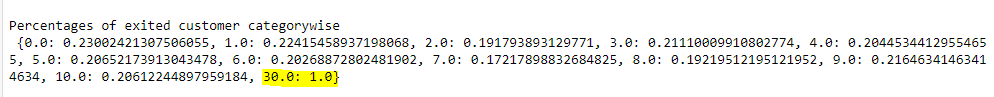




From above crosstab, We could see large number of customers exited from management unit when numofproducts exceeds 3.

Tenure:

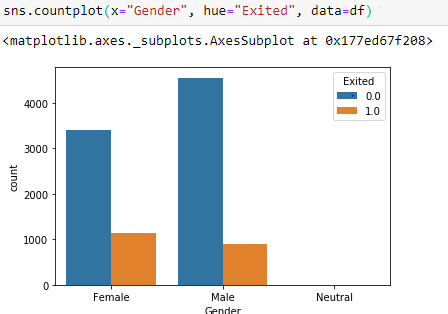


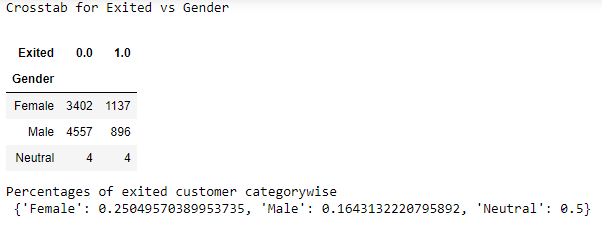


From tenure we can see, at point 30.0 there is higher rate of exit.

Remaining all other tenures, there is about 20% of exit rate.

Gender:

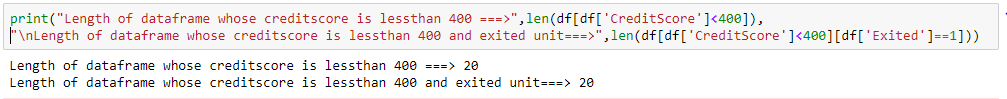




There is very low chance of male exiting unit than compared to female, where as neutrals have equal ratio of exiting and not exiting.

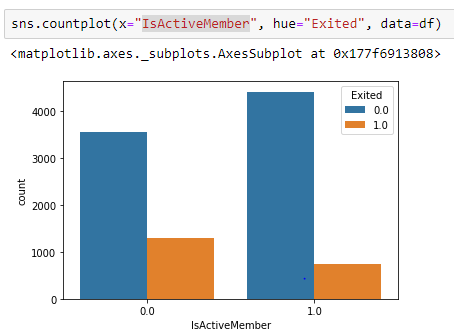
Credit Score:

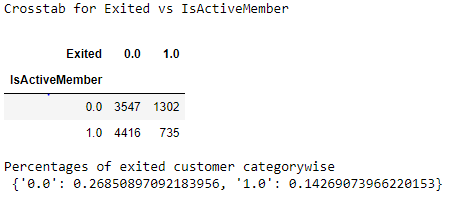




All customers whose credit score is less than 400 are exited.

ActiveMember:





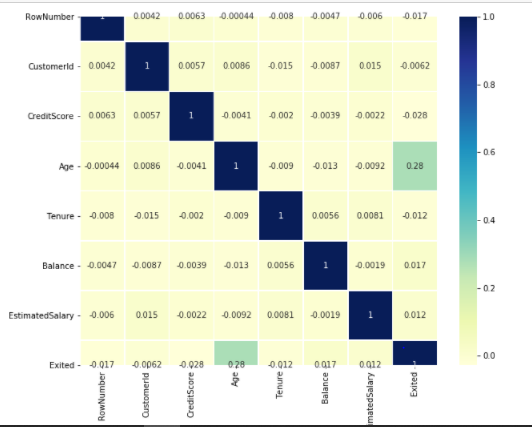
From crosstab, we can say there are 50% less chances for active member exiting unit than non active member.

Estimated Salary and Balance:

We could not find much difference in exiting and non exiting customer for fields Estimated salary and Balance.

Correlation:

* Finding out correlation between continuous variables:



There are no two features has considerable correlation between them.