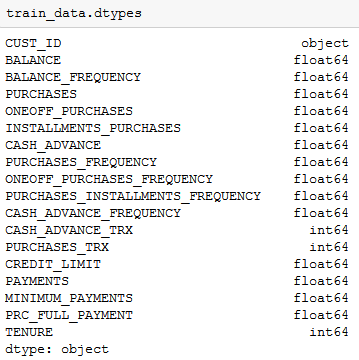
**Credit Card Segmentation**

For this project, we are given a dataset related to credit card details of a customer which consists of 9000 records.

**Data types -**

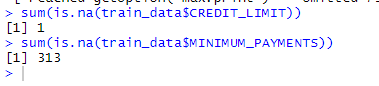
We have 18 variables present in the data out of which one variable Customer\_ID is object datatype and remaining are of numeric data. Out of the 17 numeric variables, there are 3 integer datatypes and remaining are float.

Let us list the names of the variables below -



**Handling missing values -**

There are a total of 315 missing values in the given data, 313 values in Minimum\_Payments column and 1 value in Credit limit column. Imputed these values using median



We need to derive various KPI's as follows -

* Monthly average purchase and cash advance amount
* Purchases by type (one-off, installments)
* Average amount per purchase and cash advance transaction,
* Limit usage (balance to credit limit ratio),
* Payments to minimum payments ratio etc.

Also, we need to check various behavioral segments of credit card holders using clustering algorithm.

Let us explore data before we apply any algorithm and also find the KPI's which will help us gain some insights.

**Purchase type -**

Here, we can observe that there are ones who make only One\_off transaction or Installment transaction or both.

We can find the below count for customers who made multiple types of transactions.

both\_oneoff\_installment 2741

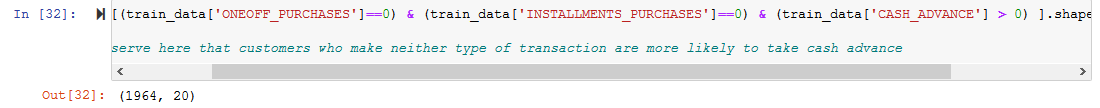
istallment 2148

none 1965

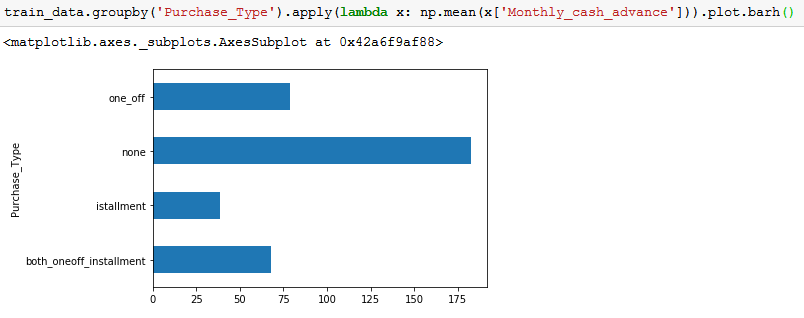
one\_off 1782

By observing the above categories, we can derive a categorical variable. I have named it as Purchase\_type

Also, we can find that customers who are not making either one\_off or installment transaction are likely to take cash advance as observed in the below screenshot



We can observe this insight in a bar plot in below screenshot -

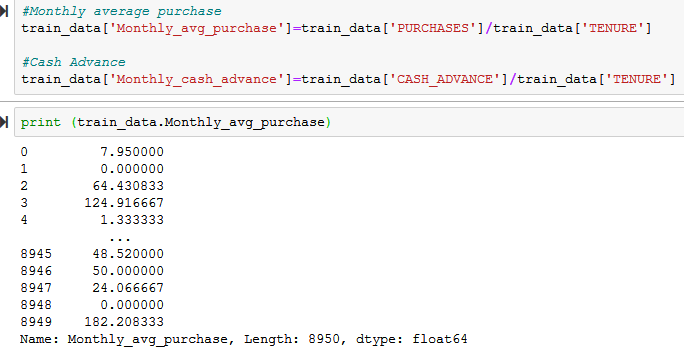


**Monthly average purchase and cash advance amount -**

Calculating monthly average purchase by dividing total purchase amount with tenure

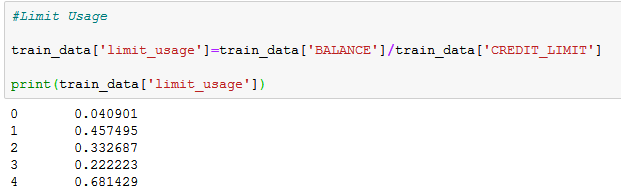
Similarly, monthly cash advance amount is calculated by dividing total cash advance amount by tenure

Find the below screenshot for output -

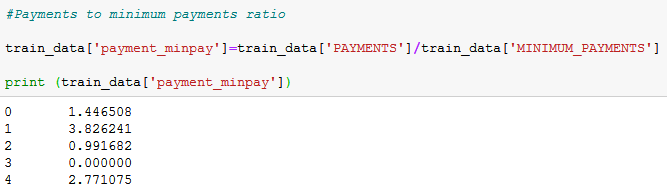


**Limit usage -**

Limit usage of a customer can be obtained by dividing Monthly average balance with credit limit

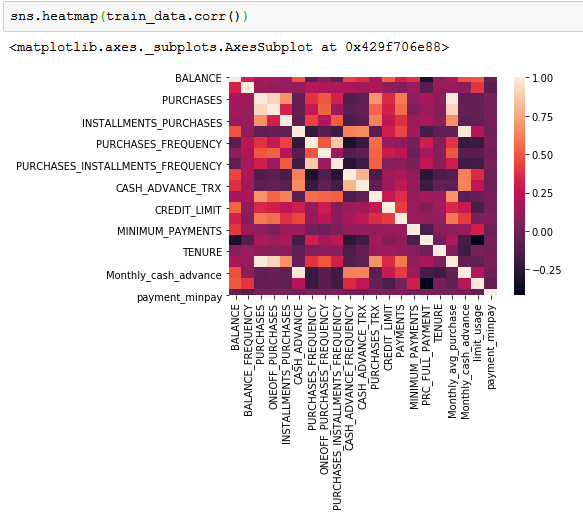


**Payments to minimum paments ratio -**

This can be obtained by calculating the ratio of total payments made with total minimum payments due in the period.

**Correlation analysis -**

Now, let us plot the correlation between the variables using the heatmap with Seaborn library



By observing the above correlation plot, we can find that the variables are correlated to each other.

One way in which we can control the correlation is by reducing number of dimensions.

Hence, we need to apply dimensionality reduction. This can be done using PCA which damps the information that's being stressed by the correlated variables

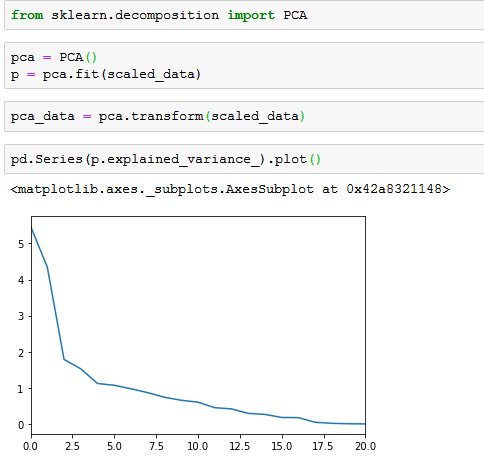
**PCA(Principle Component Analysis)**

First step in PCA is to scale data so that all variables are in equal scale. We can do this using sklearn library, importing standard scaler.

The function we used here is,

StandardScaler().fit\_transform(data)

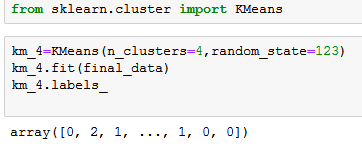
Next step is to calculate the variance between the variables to obtain the scree plot as shown below -



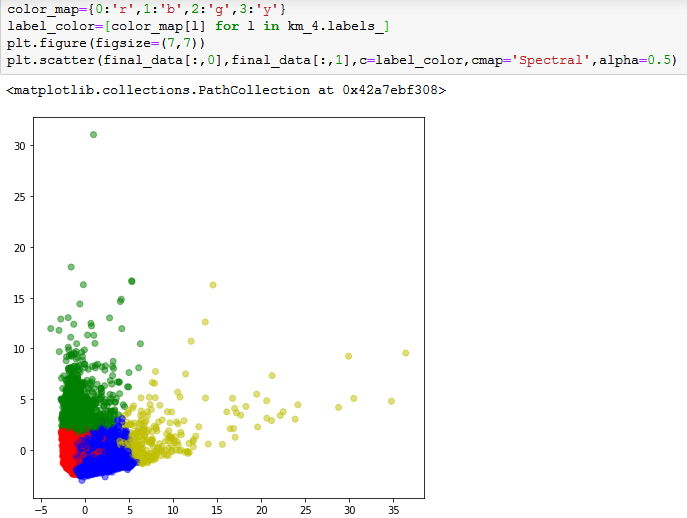
From the above plot, we can observe that 5 components are defining the total variance. Hence, we are selecting the top 5 principal components

**Clustering using K-Means**

We can perform K-Means clustering using sklearn library on the reduced data that we obtained after PCA. We are starting with 4 clusters as there are 4 types of categories we found in purchase type



Let us plot the graph using scatter plot which indicates each cluster in a different color



**Marketing Strategy:**

**Cluster 0**

This group is has minimum payment ratio and making only oneoff transactions . This group seems to be risky group.

**Cluster 1**

They have poor credit score and taking only cash on advance. We can target them by providing less interest rate on cash advance and on EMI transactions

**Cluster 2**

They are potential target customers who are paying dues and doing purchases and maintaining comparatively good credit score. We can increase credit limit or give low interest loans. Can be given premium card to increase transactions

**Cluster 3**

This group is performing best among all as cutomers are using less credit limit and paying dues on time. Giving rewards will make them use the card for more purchases.