CUSTOMER SEGME NTATION

INTRODUCTION

The preprocessing is a Real-world datasets are generally messy, raw, incomplete, inconsistent, and unusable. It can contain manual entry errors, missing values, inconsistent schema, etc. Data Preprocessing is the process of converting raw data into a format that is understandable and usable. It is a crucial step in any Data Science project to carry out an efficient and accurate analysis. It ensures that data quality is consistent before applying any **Machine Learning** or **Data Mining techniques**.

DIMENSTION REDUCTION

The dimension of the data is reducing to find and analysis the data set.

```
data=pd.read_csv("E:\Dataset\Mall_Customers.csv")
print(data.head())
```

```
   CustomerID   Genre  Age  AnnualIncome  SpendingScore
0           1    Male   19            15             39
1           2    Male   21            15             81
2           3  Female   20            16              6
3           4  Female   23            16             77
4           5  Female   31            17             40
```

Let find the data types of the data set.

```
print(data.dtypes)
```

```
CustomerID        int64
Genre            object
Age               int64
AnnualIncome      int64
SpendingScore     int64
dtype: object
```

We trans form the data into reduction dimentions for process.

```
sc=StandardScaler()
x=data.iloc[:,2:4]
y=data.iloc[:,4:]
scaler=sc.fit_transform(x)
```

```
tsne=TSNE(learning_rate=200,n_components=2)
x_tsne=tsne.fit_transform(scaler)
y_tsne=y
```

Now let implement the data set into the Kmeans algorithm.

```
from sklearn.cluster import KMeans
kmeans=KMeans()
predict=kmeans.fit_predict(x_tsne)
print(predict)
```

The predict value is

```
[0 0 0 0 0 0 0 0 2 0 2 0 6 0 0 0 0 0 6 0 0 0 6 0 6 0 6 0 0 0 2 0 6 0 6 0 6
 0 0 0 2 0 6 0 6 0 6 0 0 0 6 0 0 2 6 6 6 2 0 6 2 5 2 6 2 5 6 2 5 0 2 6 2 2
 2 5 6 1 5 6 2 1 2 6 5 6 2 5 1 7 2 5 7 1 1 5 7 5 7 5 5 7 2 5 7 5 2 7 2 2 2
 5 1 5 5 5 2 7 7 7 5 1 1 1 5 4 1 1 7 1 7 1 5 4 5 4 1 4 5 4 7 4 4 4 4 4 7 4
 4 4 1 1 1 1 7 4 1 4 4 4 7 4 5 4 7 4 1 4 3 4 3 4 3 3 7 4 7 4 7 3 3 3 3 3 3
 3 7 3 3 3 3 3 3 3 3 3 3 3 3 3]
```

CONCLUSION

From this we can able to predict the values from data of dimenstion reduced Data set using TSNE algorithm and implemented into KMeans clustureing algorithm.