|  |  |
| --- | --- |
| Assignment Date | 20 October 2022 |
| StudentName | DEEPTHI SHERONA A |
| StudentRollNumber | 61771921008 |
| MaximumMarks | 2Marks |

**Problem Statement**: Customer Segmentation Analysis

You own the mall and want to understand the customers who can quickly

converge [Target Customers] so that the insight can be given to the

marketing team and plan the strategy accordingly.

**Clustering the data and performing classification algorithms**

1. Download the dataset: Dataset

2. Load the dataset into the tool.

3. Perform Below Visualizations.

∙ Univariate Analysis

∙ Bi- Variate Analysis

∙ Multi-Variate Analysis

4. Perform descriptive statistics on the dataset.

5. Check for Missing values and deal with them.

6. Find the outliers and replace them outliers

7. Check for Categorical columns and perform encoding.

8. Scaling the data

9. Perform any of the clustering algorithms

10. Add the cluster data with the primary dataset

11. Split the data into dependent and independent

variables.

12. Split the data into training and testing

13. Build the Model

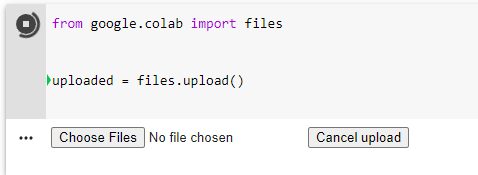
14. Train the Model

15. Test the Model

16. Measure the performance

**TASK 1**

DOWNLOAD AND LOAD THE DATASET



**TASK 2**

**import**numpy**as** np

**import** pandas **as** pd

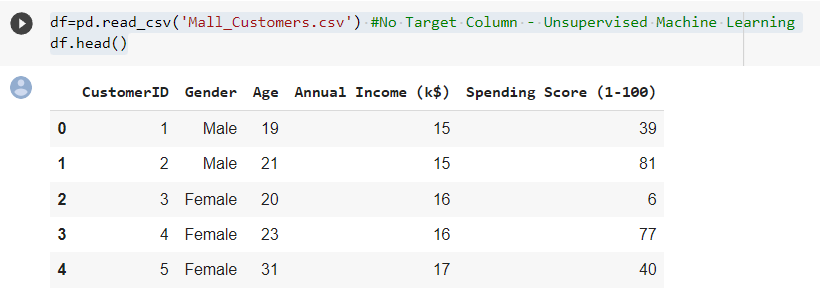
**import**matplotlib.pyplot**as**plt

**import** seaborn **as**sns

**import** matplotlib **as**rcParams

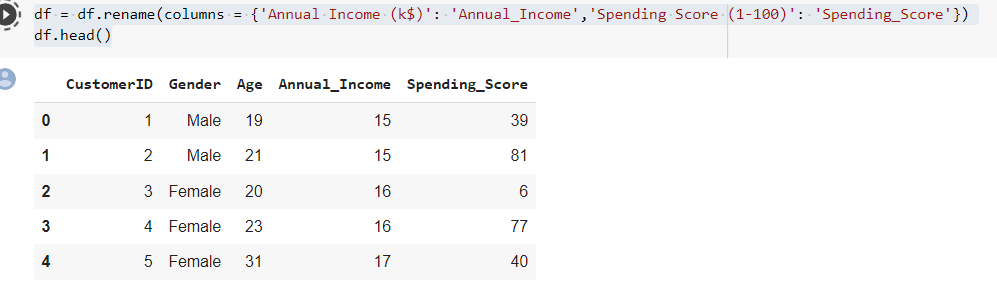
df=pd.read\_csv('Mall\_Customers.csv') #No Target Column - Unsupervised Machine Learning

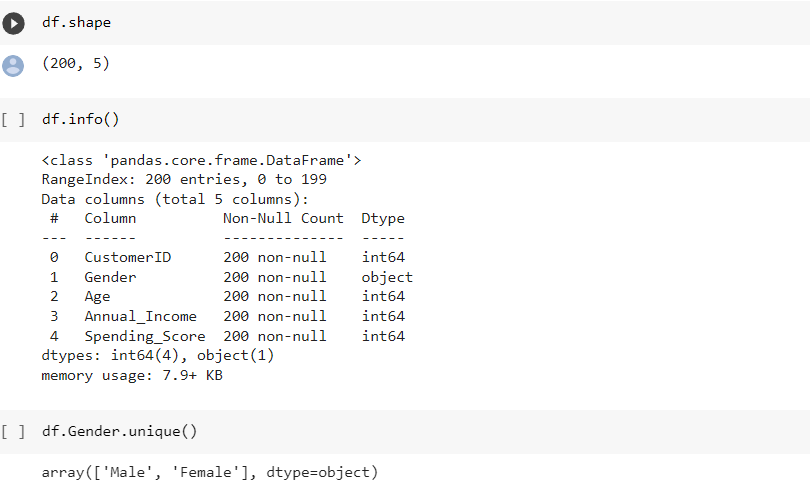
df.head()



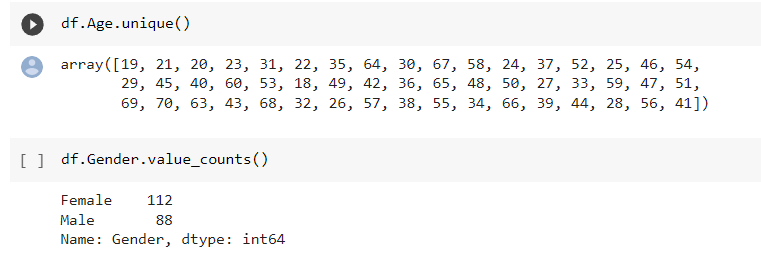
df = df.rename(columns = {'Annual Income (k$)': 'Annual\_Income','Spending Score (1-100)': 'Spending\_Score'})

df.head()

****

**df.shape()**

**df.Age.unique()**

****

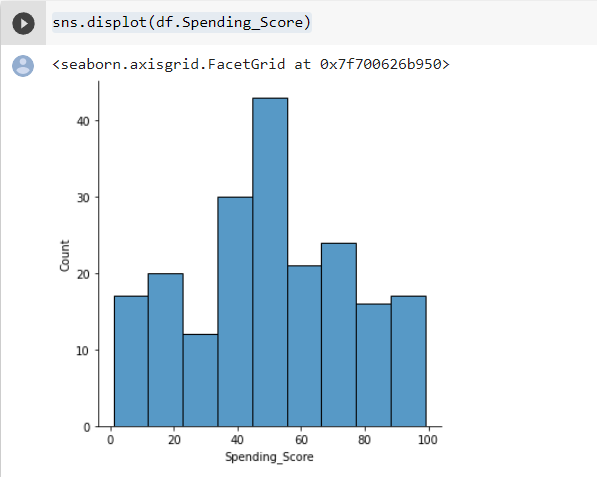
**TASK 3**

**PerformBelowVisualizations**

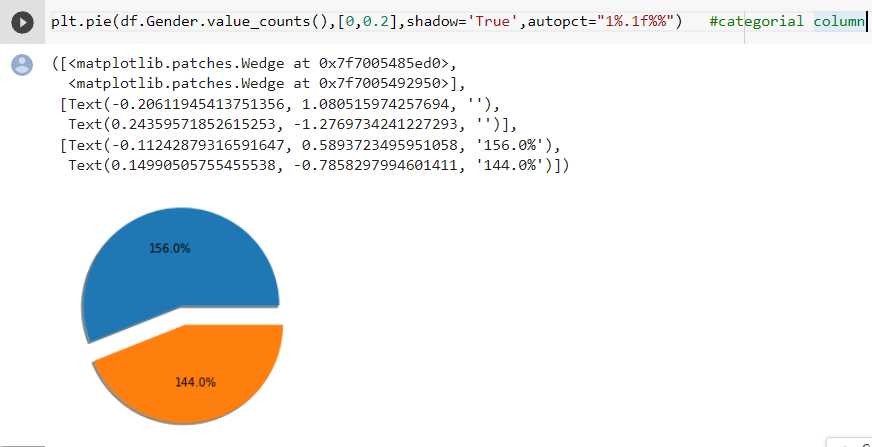
* UnivariateAnalysis
* Bi-VariateAnalysis
* Multi-VariateAnalysis

**Univariate Analysis**

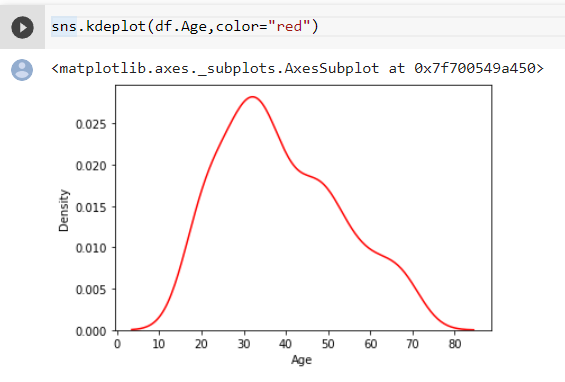
sns.displot(df.Spending\_Score)

****

plt.pie(df.Gender.value\_counts(),[0,0.2],shadow='True',autopct="1%.1f%%")   #categorial column

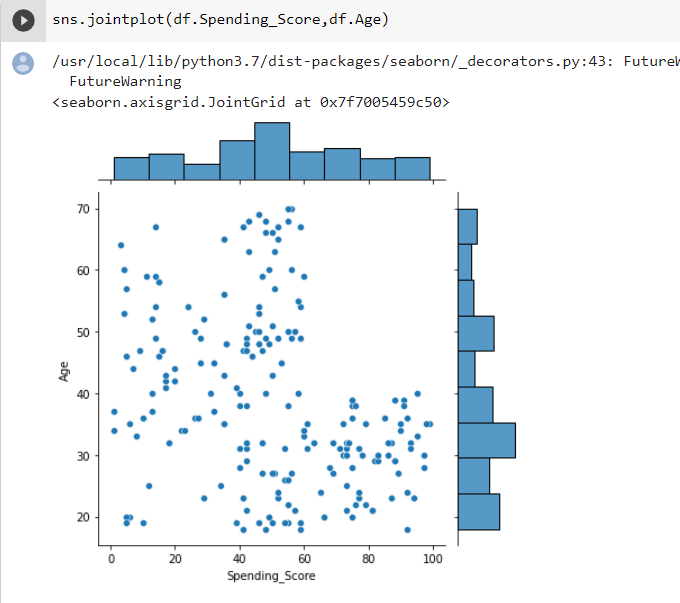
****

sns.kdeplot(df.Age,color="red”)

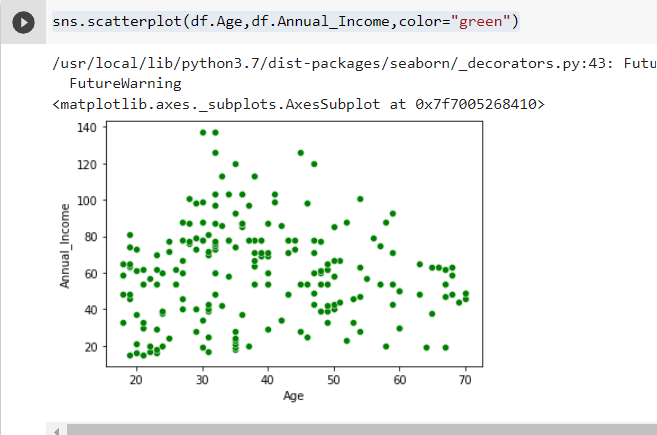


**Bi-Variate Analysis**

sns.jointplot(df.Spending\_Score,df.Age)



sns.scatterplot(df.Age,df.Annual\_Income,color="green")

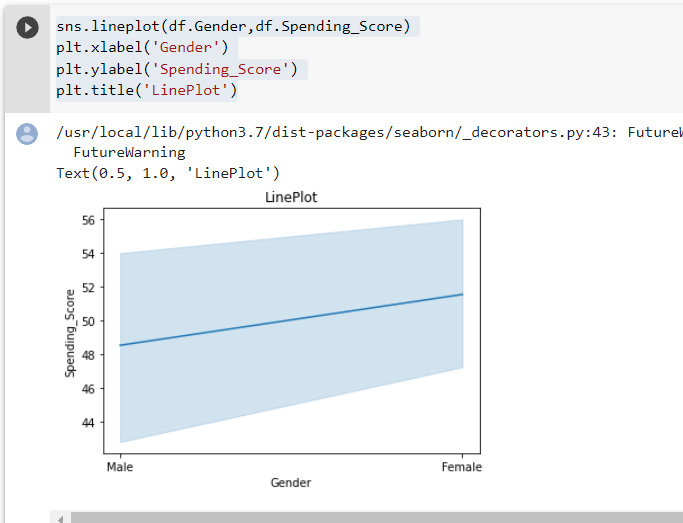


sns.lineplot(df.Gender,df.Spending\_Score)

plt.xlabel('Gender')

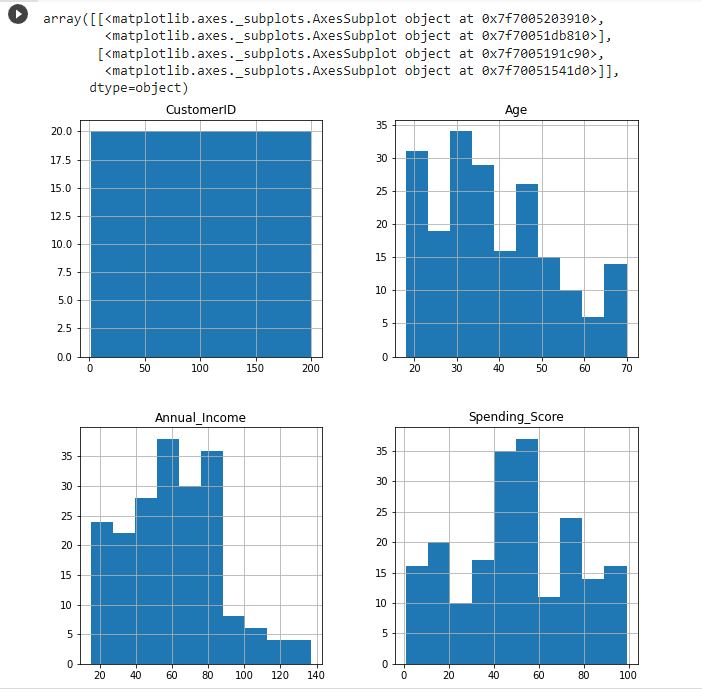
plt.ylabel('Spending\_Score')

plt.title('LinePlot')

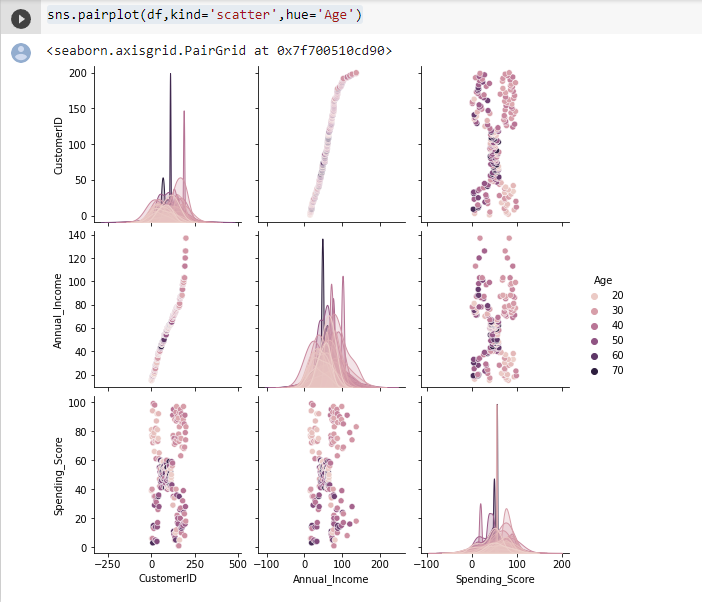


**Multi-Variate Analysis**

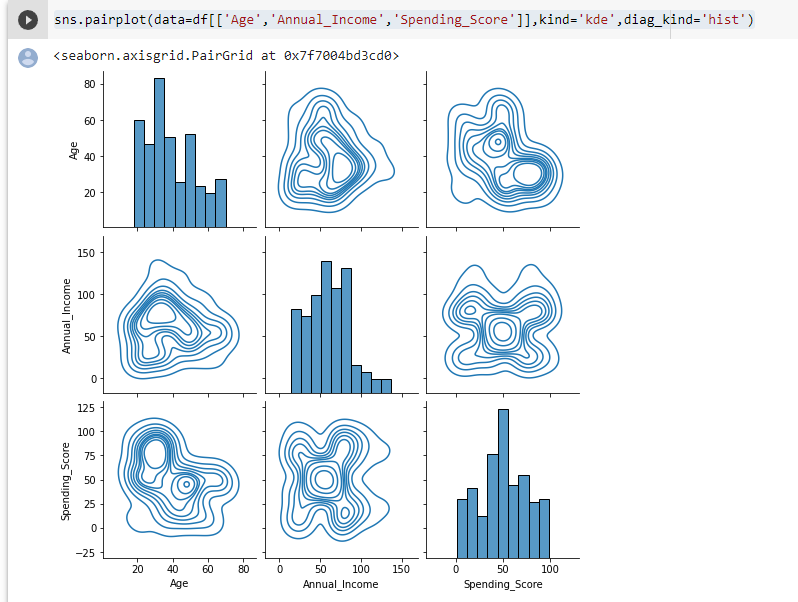
df.hist(figsize=(10,10))



sns.pairplot(df,kind='scatter',hue='Age')



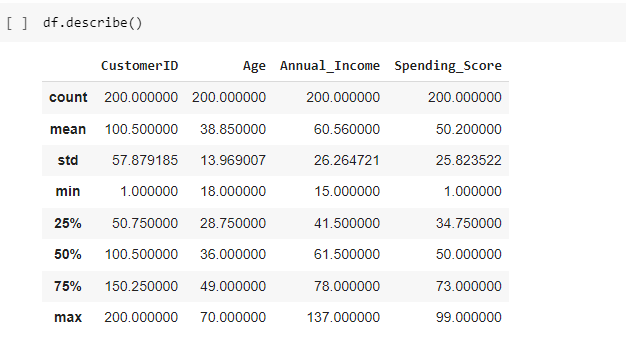
sns.pairplot(data=df[['Age','Annual\_Income','Spending\_Score']],kind='kde',diag\_kind='hist')



**TASK 4**

# Descriptivestatistics

df.describe()



**TASK 5**

# Handlemissingdata

df.isnull().any() #no missing data

# 

**TASK 6**

**Outliers Replacement**

sns.boxplot(df.Age) #no outliers

# 

# TASK 7

# ****Check for Categorical column and perform encoding****

# 

**TASK 8**

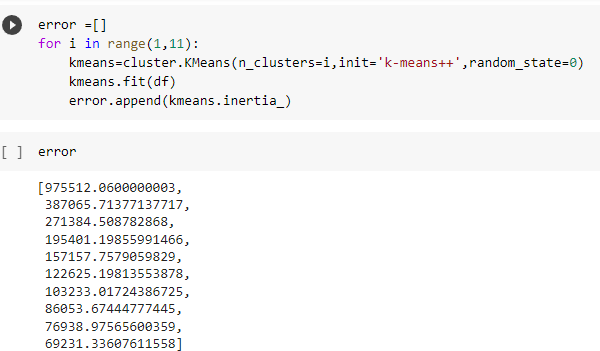
# Scaling the data

# 

**TASK 9**

# Perform any of the clustering algorithms

from sklearn import cluster



# 

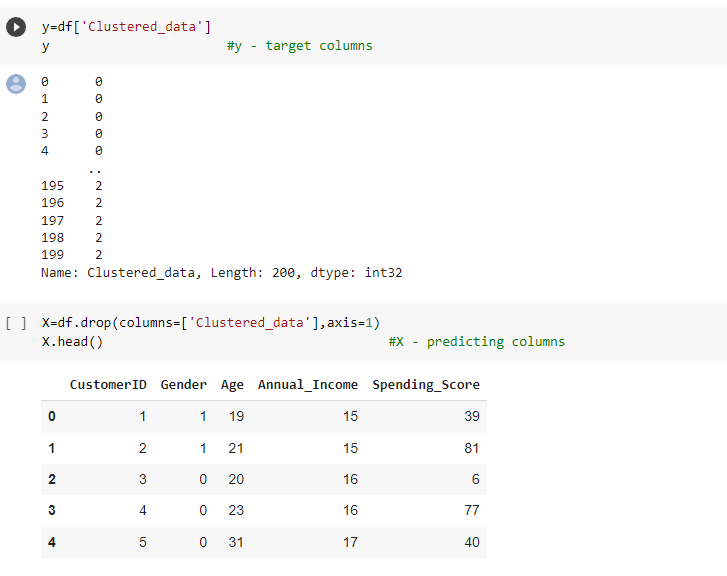
**TASK 10**

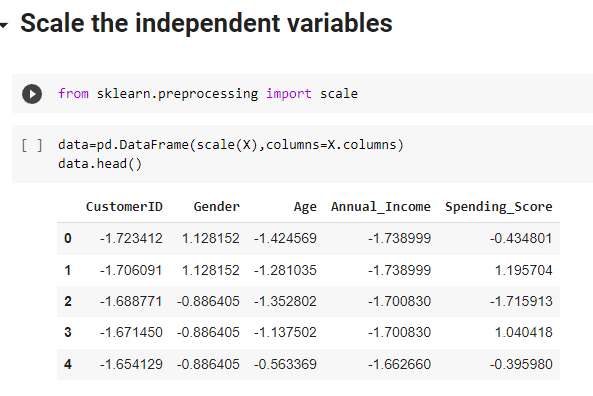
# Add the cluster data with the primary dataset

# 

**TASK 11**

**Split the data into dependent and independent variables**

****

****

**TASK 12**

# ****Split the data into training and testing****

# 

**TASK 13**

# ****Build the model****

# 

**TASK 14**

# ****Train the model****

# 

**TASK 15**

# ****Test the model****

y\_test

# 

# 

**TASK 16**

# ****Measure the performance using metrics****

# 