//to read dataset

df=pd.read\_csv("C:\\mini project\\billionaire.csv")

//print the first few rows of the Data Frame

print(df.head())

//calculate the number of missing values(NaNs) in each column of the DataFrame Data and then prints the result.

print(df.isnull().sum())

//Drops the missing values

df=df.dropna()

//it removes the $ and B sign of the networth column and converts it into float

df['NetWorth']=df['NetWorth'].str.strip('B')

df['NetWorth']=df['NetWorth'].str.strip('$')

df['NetWorth']=df["NetWorth"].astype(float)

df['NetWorth']

//Visualize the top 15 billionaires in name and networth

df2=df.sort\_values(by=["NetWorth"],ascending=False).head(15)

plt.figure(figsize=(10,5))

ax=sns.barplot(x=df2['Name'],y=df2['NetWorth'],data=df2)

ax.set\_xticklabels(ax.get\_xticklabels(),rotation=50,ha="right")

plt.show()

//Visualize the continents with most number of billionaires

a = df["Continent"].value\_counts().head(7)

index = a.index

Continent = a.values

custom\_colors = ["skyblue", "yellowgreen", 'tomato', "blue", "red"]

plt.figure(figsize=(5, 5))

plt.pie(Continent,labels=index, colors=custom\_colors)

central\_circle = plt.Circle((0, 0), 0.5, color='white')

fig = plt.gcf()

fig.gca().add\_artist(central\_circle)

plt.rc('font', size=12)

plt.title("Continents with Most Number of Billionaires", fontsize=20)

plt.show()

//Visualize the number of billionaires in industry

industry\_counts = df['Industry'].value\_counts()

plt.figure(figsize=(12, 6))

industry\_counts.plot(kind='bar', color='skyblue')

plt.title('Distribution of Billionaires by Industry')

plt.xlabel('Industry')

plt.ylabel('Number of Billionaires')

plt.xticks(rotation=60, ha='right')

plt.grid(axis='y')

plt.tight\_layout()

plt.show()

//Visualize the top 5 countries with most number of billionaires

a = df["Country"].value\_counts().head()

index = a.index

Country = a.values

custom\_colors = ["skyblue", "yellowgreen", 'tomato', "blue", "red"]

plt.figure(figsize=(5, 5))

plt.pie(Country,labels=index, colors=custom\_colors)

central\_circle = plt.Circle((0, 0), 0.5, color='white')

fig = plt.gcf()

fig.gca().add\_artist(central\_circle)

plt.rc('font', size=12)

plt.title("Top 5 Countries with Most Number of Billionaires", fontsize=20)

plt.show()

//Visualize the top 5 domains to become a billionaire

a = df["Source"].value\_counts().head()

index = a.index

sources = a.values

custom\_colors = ["skyblue", "yellowgreen", 'tomato', "blue", "red"]

plt.figure(figsize=(5, 5))

plt.pie(sources, labels=index, colors=custom\_colors)

central\_circle = plt.Circle((0, 0), 0.5, color='white')

fig = plt.gcf()

fig.gca().add\_artist(central\_circle)

plt.rc('font', size=12)

plt.title("Top 5 Domains to Become a Billionaire", fontsize=20)

plt.show()

//Performs the k-means clustering algorithm by age and networth

X=df[['Age','NetWorth']]

n\_clusters=3

kmeans=KMeans(n\_clusters=n\_clusters,random\_state=42)

kmeans.fit(X)

df['Cluster']=kmeans.labels\_

print("Cluster Centers:")

print(kmeans.cluster\_centers\_)

plt.figure(figsize=(10,6))

sns.scatterplot(x='Age',y='NetWorth',hue='Cluster',data=df,palette='viridis',s=100)

plt.title('KMeans clustering of billionaires by age and networth')

plt.xlabel('Age')

plt.ylabel('NetWorth')

plt.legend(title='Cluster')

plt.show