**20 Problem statement and its solution**

**1)Problem Statement:** What are the unique game genres in the dataset?

Solution: import pandas as pd

data = pd.read\_csv('vgsales.csv')

genres = data['Genre'].unique()

print(genres)

**2) Problem Statement:** What is the average global sales for each genre?

solution: import pandas as pd

data = pd.read\_csv('vgsales.csv')

avg\_sales = data.groupby('Genre')['Global\_Sales'].mean()

print(avg\_sales)

**3) Problem Statement:** Which game has the highest global sales?

Solution: import pandas as pd

data = pd.read\_csv('vgsales.csv')

highest\_sales = data.loc[data['Global\_Sales'].idxmax()]

print(highest\_sales)

**4) Problem Statement:** How many games were released each year

Solution: import pandas as pd

data = pd.read\_csv('vgsales.csv')

games\_per\_year = data['Year'].value\_counts().sort\_index()

print(games\_per\_year)

**5) Problem Statement:** What are the top 10 best-selling games?

Solution: import pandas as pd

data = pd.read\_csv('vgsales.csv')

top\_10\_sales = data.nlargest(10, 'Global\_Sales')

print(top\_10\_sales)

**6) Problem Statement:** Which publisher has the most games in the dataset ?

Solution: import pandas as pd

data = pd.read\_csv('vgsales.csv')

most\_games\_publisher = data['Publisher'].value\_counts().idxmax()

print(most\_games\_publisher)

**7) Problem Statement:** What is the total sales for each region (NA, EU, JP, Other)?

Solution: import pandas as pd

data = pd.read\_csv('vgsales.csv')

regional\_sales = data[['NA\_Sales', 'EU\_Sales', 'JP\_Sales', 'Other\_Sales']].sum()

print(regional\_sales)

**8) Problem Statement:** What is the correlation between North American sales and European sales?

Solution:import pandas as pd

data = pd.read\_csv('vgsales.csv')

correlation = data['NA\_Sales'].corr(data['EU\_Sales'])

print(correlation)

**9) Problem Statement:** What is the median global sales value?

Solution: import pandas as pd

data = pd.read\_csv('vgsales.csv')

median\_sales = data['Global\_Sales'].median()

print(median\_sales)

**10) Problem Statement:** How many games were released for each platform?

Solution: import pandas as pd

data = pd.read\_csv('vgsales.csv')

platform\_counts = data['Platform'].value\_counts()

print(platform\_counts)

**11) Problem Statement:** Find the average year of release for each genre.

Solution: import pandas as pd

data = pd.read\_csv('vgsales.csv')

data['Total\_Sales'] = data[['NA\_Sales', 'EU\_Sales', 'JP\_Sales', 'Other\_Sales']].sum(axis=1)

print(data['Total\_Sales'])

**12) Problem Statement**: Calculate the total sales for each game.

Solution: import pandas as pd

data = pd.read\_csv('vgsales.csv')

data['Total\_Sales'] = data[['NA\_Sales', 'EU\_Sales', 'JP\_Sales', 'Other\_Sales']].sum(axis=1)

print(data['Total\_Sales'])

**13) Problem Statement:** Identify the top 5 publishers by global sales.

Solution: import pandas as pd

data = pd.read\_csv('vgsales.csv')

top\_5\_publishers = data.groupby('Publisher')['Global\_Sales'].sum().nlargest(5)

print(top\_5\_publishers)

**14) Problem Statement:** Determine the number of games released before the year 2000. Solution: import pandas as pd

data = pd.read\_csv('vgsales.csv')

games\_before\_2000 = data[data['Year'] < 2000].shape[0]

print(games\_before\_2000)

**15)Problem Statement:** Calculate the percentage of games in each genre.

Solution:import pandas as pd

data = pd.read\_csv('vgsales.csv')

genre\_percentage = data['Genre'].value\_counts(normalize=True) \* 100

print(genre\_percentage)

**16)Problem Statement:** Find the game with the highest sales in Japan.

Solution: import pandas as pd

data = pd.read\_csv('vgsales.csv')

highest\_jp\_sales = data.loc[data['JP\_Sales'].idxmax()]

print(highest\_jp\_sales)

**17)Problem Statement:** Group games by platform and find the average global sales for each.

Solution: import pandas as pd

data = pd.read\_csv('vgsales.csv')

avg\_sales\_by\_platform = data.groupby('Platform')['Global\_Sales'].mean()

print(avg\_sales\_by\_platform)

**18) Problem Statement**: Calculate the cumulative global sales over the years

Solution: import pandas as pd

data = pd.read\_csv('vgsales.csv')

cumulative\_sales = data.groupby('Year')['Global\_Sales'].sum().cumsum()

print(cumulative\_sales)

**19) Problem Statement:** Identify the games with global sales above 10 million.

Solution: import pandas as pd

data = pd.read\_csv('vgsales.csv')

high\_selling\_games = data[data['Global\_Sales'] > 10]

print(high\_selling\_games)

**20) Problem Statement**: Determine the range of years in which games were released. Solution: import pandas as pd

data = pd.read\_csv('vgsales.csv')

year\_range = data['Year'].agg(['min', 'max'])

print(year\_range)

**Name=Prashant Ahire**

**Class=ET2**

**Roll no=88**

**Prn=202401070045**