Assignment NO.7

```python
[1]: import pandas as pd

     # csv
     df = pd.read_csv('Dummy Data HSS.csv')

     # excel
     # df_excel = pd.read_excel('sales_data.xlsx')

     # json
     # import json
     # with open('sales.json') as f:
     #     data_json = json.load(f)
     # df_json = pd.DataFrame(data_json)
```

```python
[2]: df.head()
```

[2]:

	TV	Radio	Social Media	Influencer	Sales
0	16.0	6.566231	2.907983	Mega	54.732757
1	13.0	9.237765	2.409567	Mega	46.677897
2	41.0	15.886446	2.913410	Mega	150.177829
3	83.0	30.020028	6.922304	Mega	298.246340
4	15.0	8.437408	1.405998	Micro	56.594181

```python
[3]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4572 entries, 0 to 4571
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   TV            4562 non-null   float64
 1   Radio         4568 non-null   float64
 2   Social Media  4566 non-null   float64
 3   Influencer    4572 non-null   object
 4   Sales         4566 non-null   float64
dtypes: float64(4), object(1)
memory usage: 178.7+ KB
```

```
[4]: df.isnull().sum()
```

```
[4]: TV               10
     Radio             4
     Social Media      6
     Influencer        0
     Sales             6
     dtype: int64
```

```
[5]: value = df['TV'].mean()
     df['TV'].fillna(value, inplace=True)
     # df.drop_duplicates(inplace=True)
```

```
[6]: df.isnull().sum()
```

```
[6]: TV                0
     Radio             4
     Social Media      6
     Influencer        0
     Sales             6
     dtype: int64
```

```
[7]: # merged_df = pd.merge(df1, df2, on='common_column', how='inner')
     # df[['day', 'month', 'year']] = df['date'].str.split('/', expand=True)
```

```
[8]: df.describe()
```

[8]:

	TV	Radio	Social Media	Sales
count	4572.000000	4568.000000	4566.000000	4566.000000
mean	54.066857	18.160356	3.323956	192.466602
std	26.096461	9.676958	2.212670	93.133092
min	10.000000	0.000684	0.000031	31.199409
25%	32.000000	10.525957	1.527849	112.322882
50%	53.000000	17.859513	3.055565	189.231172
75%	77.000000	25.649730	4.807558	272.507922
max	100.000000	48.871161	13.981662	364.079751

```
[9]: category_sales = df.groupby(['TV', 'Radio', 'Social Media'])['Sales'].sum()
     category_sales
```

```
[9]: TV      Radio       Social Media
     10.0    0.573244    1.072542              33.719607
             0.688749    0.982756              33.459886
             0.758569    0.527881              35.547998
             0.858810    3.621606              36.882298
             1.179967    0.939789              34.205170
                                                   ...
     100.0   42.225232   8.977117             364.079751
             42.832653   3.965113             354.869546
             43.760694   6.420971             350.087078
             44.560410   8.470340             357.092487
             45.082921   4.511628             352.657695
     Name: Sales, Length: 4562, dtype: float64
```

```
[10]: total_sales = df['Sales'].sum()
      print('total sales',total_sales)
      avg_order_value = df['Sales'].mean()
      print('avg order value',avg_order_value)
```
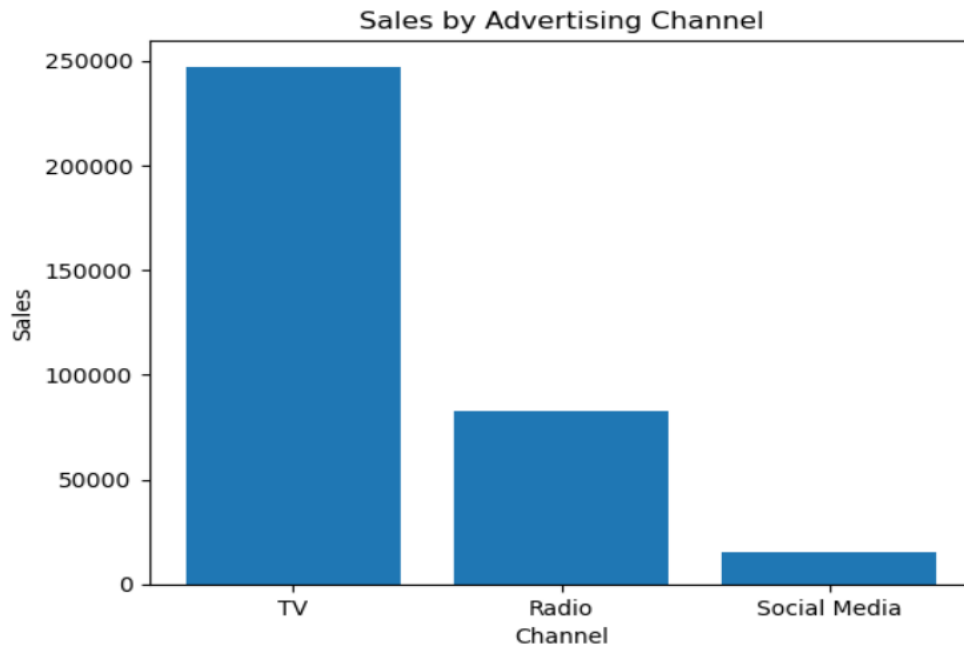
```
total sales 878802.50521883
avg order value 192.46660210662066
```

```
[11]: import matplotlib.pyplot as plt

      channels = ['TV', 'Radio', 'Social Media']

      sales = df[['TV', 'Radio', 'Social Media']].sum()

      plt.bar(channels, sales)
      plt.xlabel('Channel')
      plt.ylabel('Sales')
      plt.title('Sales by Advertising Channel')
      plt.show()
```

Sales by Advertising Channel

```
[12]: import plotly.express as px

[13]: channels = ['TV', 'Radio', 'Social Media']

      sales = df[['TV', 'Radio', 'Social Media']].sum()

      px.bar(x=channels, y=sales)
```