Data Visualisation - Graded Questions

`Note` - *This stub file doesn't contain the conceptual questions asked on the platform*

I) Marks Analysis

In the **'Marks.csv'** file, you can find the scores obtained by 200 students in 4 subjects of a standardised test. The different columns - `Score A`, `Score B`, `Score C` and `Score D` indicate the score obtained by a particular student in the respective subjects A, B, C and D.

Load the dataset to your notebook and answer the following questions

In [ ]:

```
#Load the necessary Libraries
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

In [ ]:

```
#Load the dataset
df1 = pd.read_csv('Marks.csv')
df1.head()
```

Out[ ]:

	Score A	Score B	Score C	Score D
0	230.1	37.8	69.2	22.1
1	44.5	39.3	45.1	10.4
2	17.2	45.9	69.3	12.0
3	151.5	41.3	58.5	16.5
4	180.8	10.8	58.4	17.9

Q1) Load the dataset and plot a histogram for the `Score A` column by keeping the `number of bins to 6`. Which bin range among the following has the highest frequency?

(`Note` - *The bin ranges mentioned in the options are approximate values for the bin ranges that you'll actually get when you plot the histogram*)

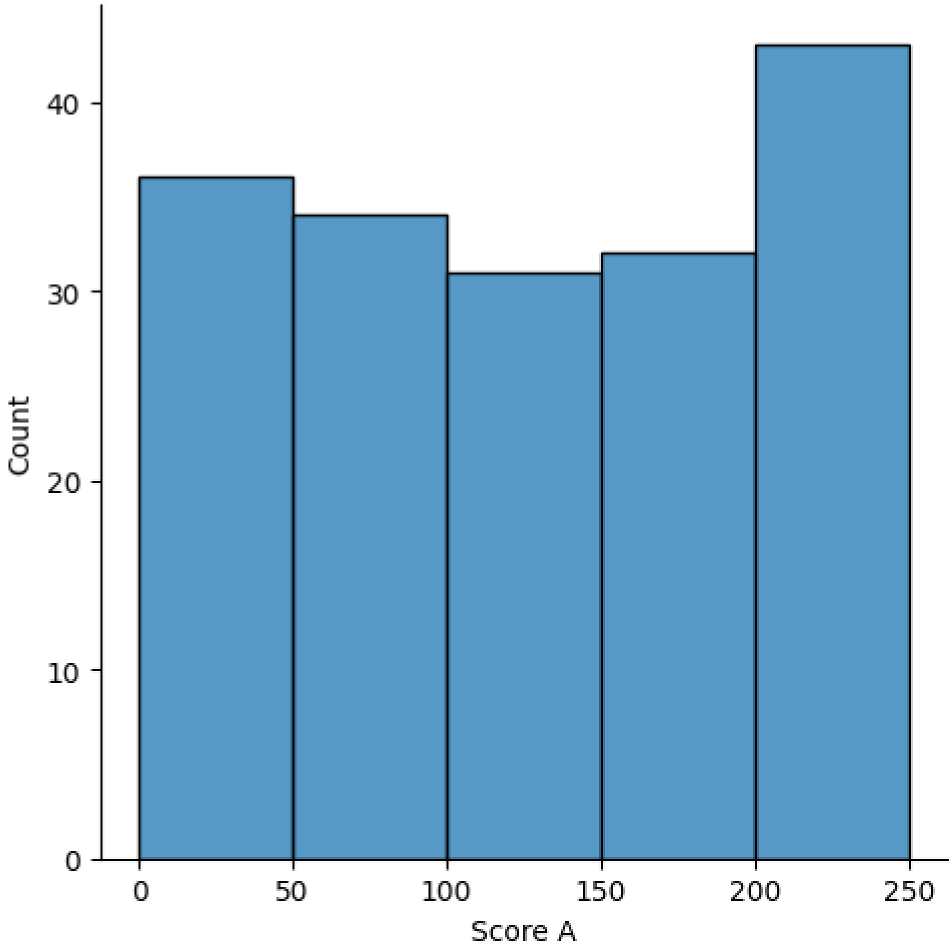
```
a)0-50
b)50-100
c)150-200
d)200-250
```

In [ ]:

```
#Your code here
sns.displot(df1, x="Score A", bins = [0,50,100,150,200,250])
```

```
c:\Users\Rommel\AppData\Local\Programs\Python\Python311\Lib\site-packages\seaborn
\_oldcore.py:1498: FutureWarning: is_categorical_dtype is deprecated and will be
removed in a future version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
c:\Users\Rommel\AppData\Local\Programs\Python\Python311\Lib\site-packages\seaborn
\_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be
removed in a future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
```

```
Out[ ]: <seaborn.axisgrid.FacetGrid at 0x2a5ffa18e50>
```

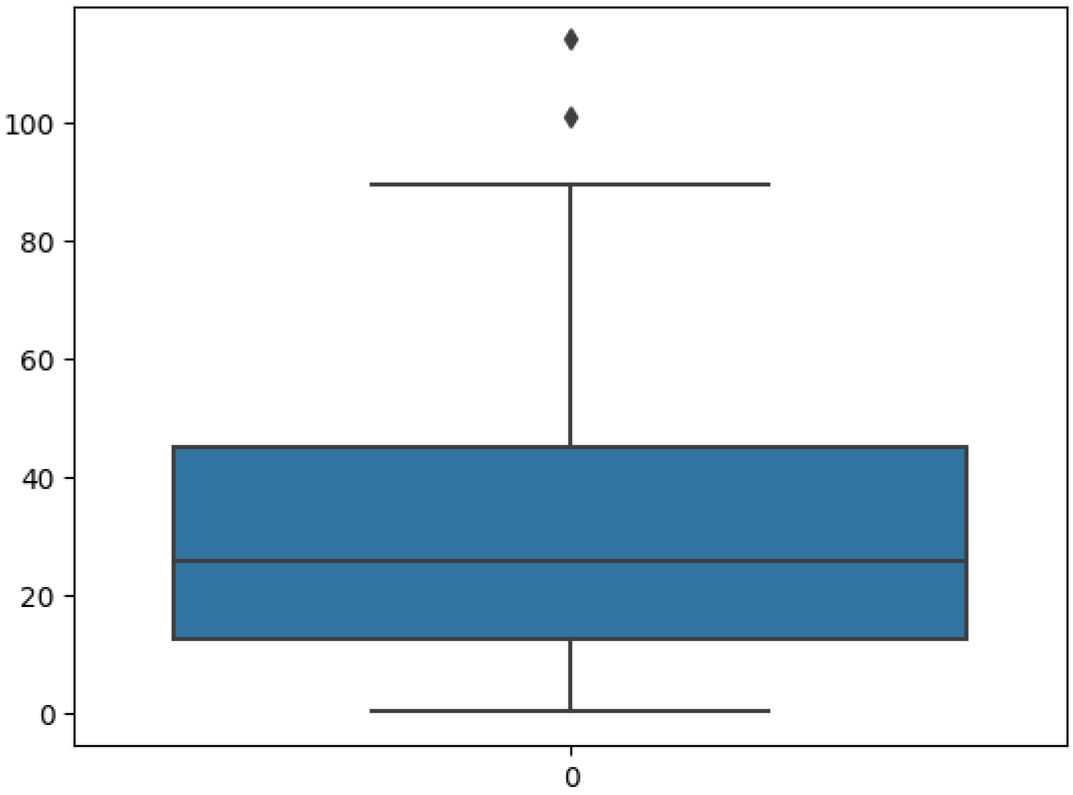


Q2) Plot a box plot for the column `Score C` and choose the correct option.

```
A - The 25th percentile lies between 20 and 40
B - The 75th percentile lies between 40 and 60
C - The 25th percentile lies between 0 and 20
D - Both B and C (Correct answer)
```

```
In [ ]: #Your code here
        sns.boxplot(df1['Score C'])
```

```
Out[ ]: <Axes: >
```



II) Superstore Data

In the `superstore.csv` file, you have the details of orders purchased in an American online retail store. Load the dataset, observe and analyse the different columns and answer the following questions.

```
In [ ]: #Load the dataset
        df2 = pd.read_csv('superstore.csv')
        df2.head()
```

Out[ ]:

	Order ID	Ship Mode	Segment	Region	Product ID	Sales	Quantity	Discount	Pr
0	CA-2016-152156	Second Class	Consumer	South	FUR-BO-10001798	261.9600	2	0%	41.9
1	CA-2016-152156	Second Class	Consumer	South	FUR-CH-10000454	731.9400	3	0%	219.5
2	CA-2016-138688	Second Class	Corporate	West	OFF-LA-10000240	14.6200	2	0%	6.8
3	US-2015-108966	Standard Class	Consumer	South	FUR-TA-10000577	957.5775	5	0.45%	-383.0
4	US-2015-108966	Standard Class	Consumer	South	OFF-ST-10000760	22.3680	2	0.20%	2.5

Q4) Plot a pie-chart to find the Ship Mode through which most of the orders are being delivered.

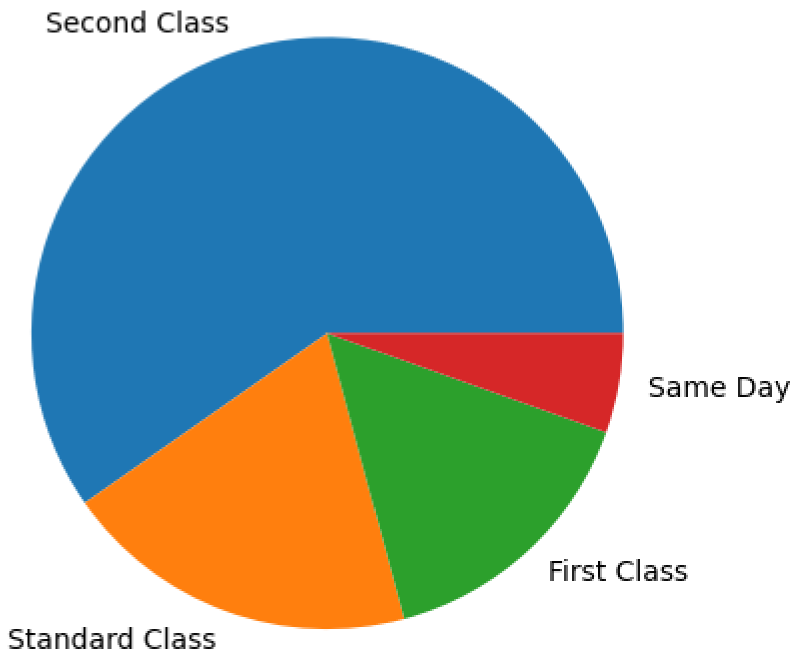
```
a)Standard Class
b)First Class
c)Second Class (Correct answer)
d)Same Day
```

In [ ]:

```
#Your code here
plt.pie(df2['Ship Mode'].value_counts(),labels=df2['Ship Mode'].unique())
```

Out[ ]:

```
([<matplotlib.patches.Wedge at 0x2a5ffb81b10>,
  <matplotlib.patches.Wedge at 0x2a590ffce90>,
  <matplotlib.patches.Wedge at 0x2a590ffe690>,
  <matplotlib.patches.Wedge at 0x2a590fffbd0>],
 [Text(-0.33056573952035373, 1.0491550370919267, 'Second Class'),
  Text(-0.37607764230951635, -1.0337144707098356, 'Standard Class'),
  Text(0.7465348771572817, -0.8078896441889587, 'First Class'),
  Text(1.0840144265772789, -0.18684946607452133, 'Same Day')])
```



Q5) Plot a bar chart comparing the average `Discount` across all the `Regions` and report back the `Region` getting the highest average discount

Note - You need to clean the `Discount` column first

```
a)Central (Correct answer)
b)South
c)West
d)East
```

In [ ]:
```
df2['Discount'] = df2['Discount'].str.replace('%','')
df2['Discount'] = df2['Discount'].astype(float)
df2.head()
```

Out[ ]:

	Order ID	Ship Mode	Segment	Region	Product ID	Sales	Quantity	Discount	Pr
0	CA-2016-152156	Second Class	Consumer	South	FUR-BO-10001798	261.9600	2	0.00	41.9
1	CA-2016-152156	Second Class	Consumer	South	FUR-CH-10000454	731.9400	3	0.00	219.5
2	CA-2016-138688	Second Class	Corporate	West	OFF-LA-10000240	14.6200	2	0.00	6.8
3	US-2015-108966	Standard Class	Consumer	South	FUR-TA-10000577	957.5775	5	0.45	-383.0
4	US-2015-108966	Standard Class	Consumer	South	OFF-ST-10000760	22.3680	2	0.20	2.5

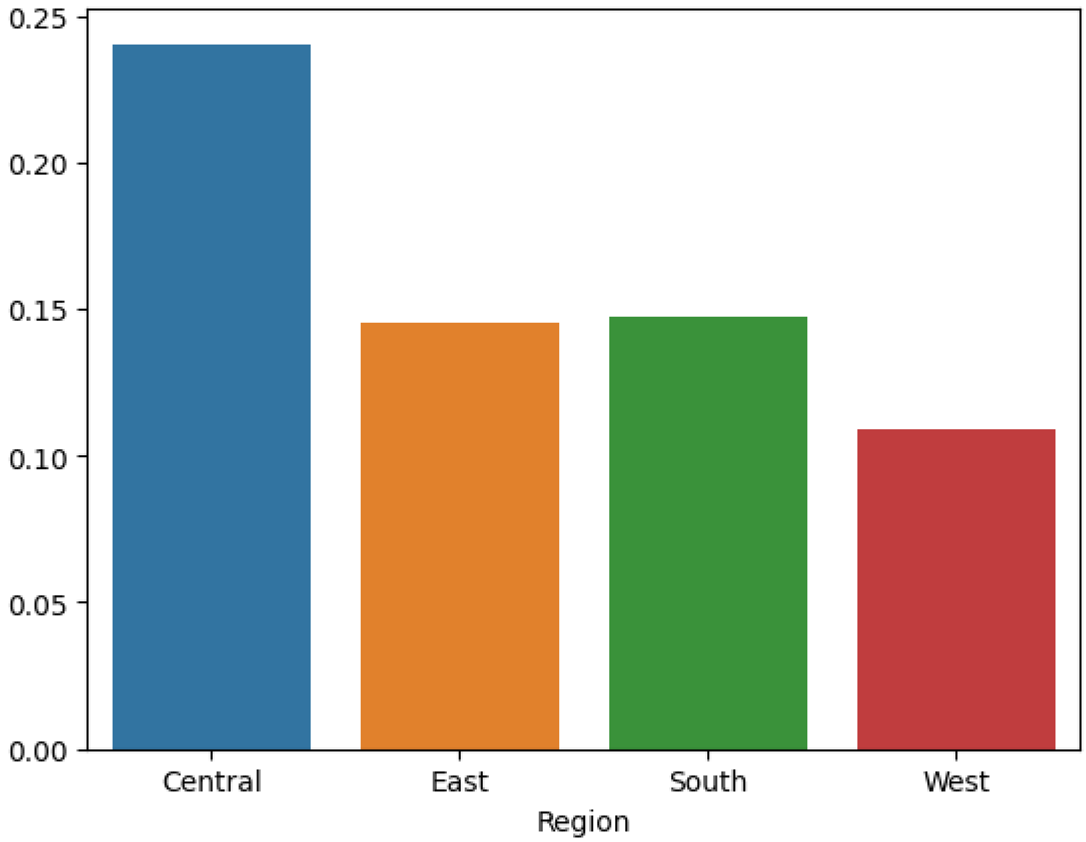
In [ ]:

```
avg = df2.groupby('Region')['Discount'].mean()
sns.barplot(x = avg.index,y = avg.values)
```

```
c:\Users\Rommel\AppData\Local\Programs\Python\Python311\Lib\site-packages\seaborn
\_oldcore.py:1498: FutureWarning: is_categorical_dtype is deprecated and will be
removed in a future version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
c:\Users\Rommel\AppData\Local\Programs\Python\Python311\Lib\site-packages\seaborn
\_oldcore.py:1498: FutureWarning: is_categorical_dtype is deprecated and will be
removed in a future version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
c:\Users\Rommel\AppData\Local\Programs\Python\Python311\Lib\site-packages\seaborn
\_oldcore.py:1498: FutureWarning: is_categorical_dtype is deprecated and will be
removed in a future version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
```

Out[ ]:

```
<Axes: xlabel='Region'>
```


0.25
0.20
0.15
0.10
0.05
0.00
Central
East
South
West
Region