Create visualizations using Matplotib, Seaborn and Folium

Estimated time needed: **40** minutes

In this assignment, you will have the opportunity to demonstrate the skills you have acquired in creating visualizations using *Matplotlib, Seaborn, Folium*.

After each task you will be required to save your plots as an image or screenshot using the filenames specified. You will be uploading these images during your final project submission so they can be evaluated by your peers.

Table of Contents

Objectives

After completing this lab you will be able to:

- Create informative and visually appealing plots with Matplotlib and Seaborn.
- Apply visualization to communicate insights from the data.
- Analyze data through using visualizations.
- Customize visualizations

Setup

For this lab, we will be using the following libraries:

- `pandas` for managing the data.
- `numpy` for mathematical operations.

- `matplotlib` for plotting.
- `seaborn` for plotting.
- `Folium` for plotting.

Installing Required Libraries

The following required libraries are pre-installed in the Skills Network Labs environment. However, if you run these notebook commands in a different Jupyter environment (e.g. Watson Studio or Ananconda), you will need to install these libraries by removing the `#` sign before `%pip` in the code cell below.

In [1]:
```
# All Libraries required for this lab are listed below. The libraries pre-installed on
# %pip install -qy pandas==1.3.4 numpy==1.21.4 matplotlib==3.5.0 seaborn folium
# Note: If your environment doesn't support "%pip install", use "!mamba install"
```

In [2]:
```
%pip install seaborn
%pip install folium
```

Importing Required Libraries

We recommend you import all required libraries in one place (here):

In [4]:
```
import numpy as np
import pandas as pd
%matplotlib inline
import matplotlib as mpl
import matplotlib.pyplot as plt
import seaborn as sns
import folium
```

▶ Click here for a hint

▶ Click here for python solution

Scenario

In this assignment you will be tasked with creating plots which answer questions for analysing "historical_automobile_sales" data to understand the historical trends in automobile sales during recession periods.
recession period 1 - year 1980
recession period 2 - year 1981 to 1982
recession period 3 - year 1991
recession period 4 - year 2000 to 2001
recession period 5 - year end 2007 to mid 2009
recession period 6 - year 2020 -Feb to April (Covid-19 Impact)

Data Description

The dataset used for this visualization assignment contains *historical_automobile_sales* data representing automobile sales and related variables during recession and non-recession period.

The dataset includes the following variables:
1. Date: The date of the observation.
2. Recession: A binary variable indicating recession perion; 1 means it was recession, 0 means it was normal.
3. Automobile_Sales: The number of vehicles sold during the period.
4. GDP: The per capita GDP value in USD.
5. Unemployment_Rate: The monthly unemployment rate.
6. Consumer_Confidence: A synthetic index representing consumer confidence, which can impact consumer spending and automobile purchases.
7. Seasonality_Weight: The weight representing the seasonality effect on automobile sales during the period.
8. Price: The average vehicle price during the period.
9. Advertising_Expenditure: The advertising expenditure of the company.
10.Vehicle_Type: The type of vehicles sold; Supperminicar, Smallfamiliycar, Mediumfamilycar, Executivecar, Sports.
11.Competition: The measure of competition in the market, such as the number of competitors or market share of major manufacturers.
12.Month: Month of the observation extracted from Date..
13.Year: Year of the observation extracted from Date.
By examining various factors mentioned above from the dataset, you aim to gain insights into how recessions impacted automobile sales for your company.

Importing Data

For your convenience, we have already written code to import the data below.

In [5]:
```
from js import fetch
import io

URL = "https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperS
resp = await fetch(URL)
text = io.BytesIO((await resp.arrayBuffer()).to_py())
import pandas as pd
df = pd.read_csv(text)
print('Data downloaded and read into a dataframe!')
```

```
Data downloaded and read into a dataframe!
```

In [31]:
```
df.head()
```

Out[31]:

	Date	Year	Month	Recession	Consumer_Confidence	Seasonality_Weight	Price	Advertising
0	1/31/1980	1980	Jan	1	108.24	0.50	27483.571	
1	2/29/1980	1980	Feb	1	98.75	0.75	24308.678	
2	3/31/1980	1980	Mar	1	107.48	0.20	28238.443	
3	4/30/1980	1980	Apr	1	115.01	1.00	32615.149	
4	5/31/1980	1980	May	1	98.72	0.20	23829.233	

In [6]:
```
df.describe()
```

Out[6]:

	Year	Recession	Consumer_Confidence	Seasonality_Weight	Price	Advertising_Exp
count	528.000000	528.000000	528.000000	528.000000	528.000000	52
mean	2001.500000	0.214015	101.140170	0.575795	24964.991956	306
std	12.710467	0.410526	10.601154	0.454477	4888.073433	113
min	1980.000000	0.000000	73.900000	0.000000	8793.663000	100
25%	1990.750000	0.000000	94.035000	0.250000	21453.300500	208
50%	2001.500000	0.000000	100.740000	0.500000	25038.691500	307
75%	2012.250000	0.000000	108.240000	0.750000	28131.684750	406
max	2023.000000	1.000000	131.670000	1.500000	44263.657000	498

In [7]:
```
df.columns
```

Out[7]:
```
Index(['Date', 'Year', 'Month', 'Recession', 'Consumer_Confidence',
       'Seasonality_Weight', 'Price', 'Advertising_Expenditure', 'Competition',
       'GDP', 'Growth_Rate', 'unemployment_rate', 'Automobile_Sales',
       'Vehicle_Type', 'City'],
      dtype='object')
```

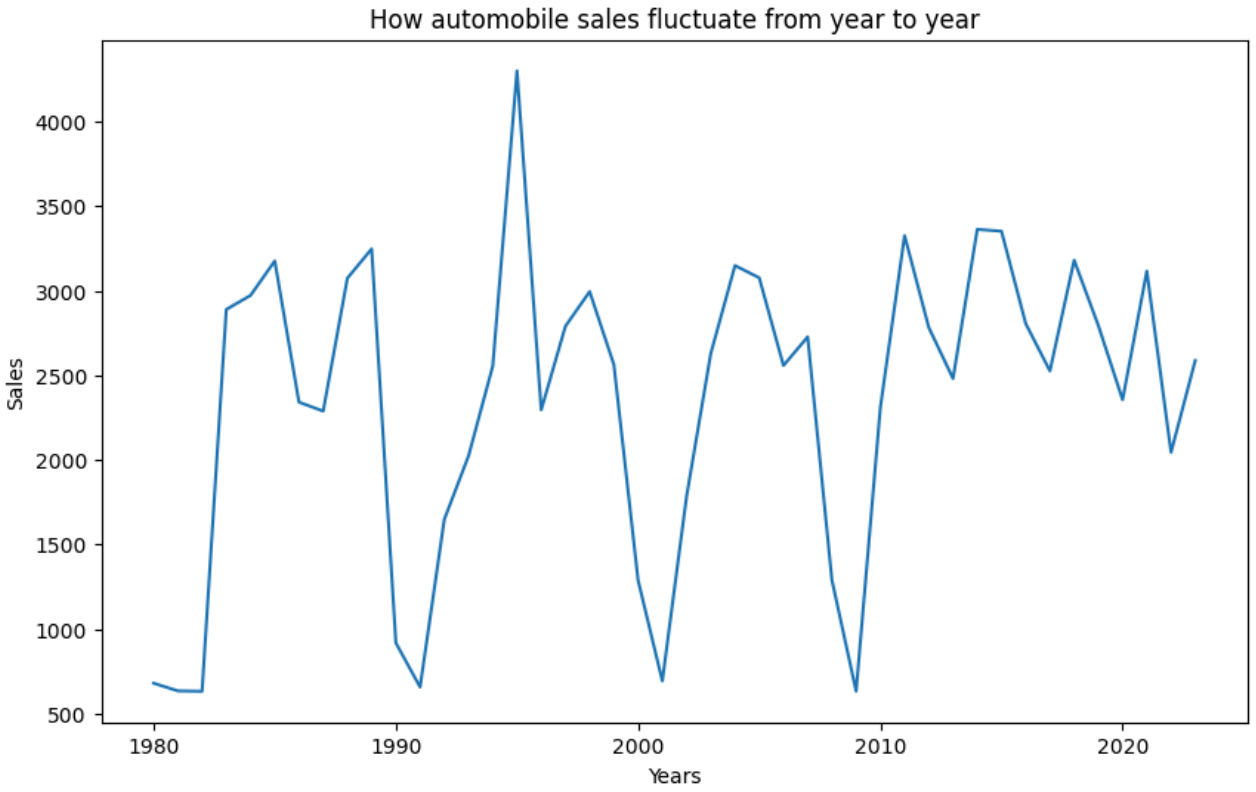
Creating Visualizations for Data Analysis

TASK 1.1: Develop a *Line chart* using the functionality of pandas to show how automobile sales fluctuate from year to year

▶ Click here for a hint

In [9]:
```
df_line = df.groupby(df['Year'])['Automobile_Sales'].mean()

plt.figure(figsize=(10, 6))
df_line.plot(kind = 'line')
plt.xlabel('Years')
plt.ylabel('Sales')
```

```
plt.title('How automobile sales fluctuate from year to year')
plt.show()
```



▶ Click here for a solution template

Include the following on the plot

ticks on x- axis with all the years, to identify the years of recession

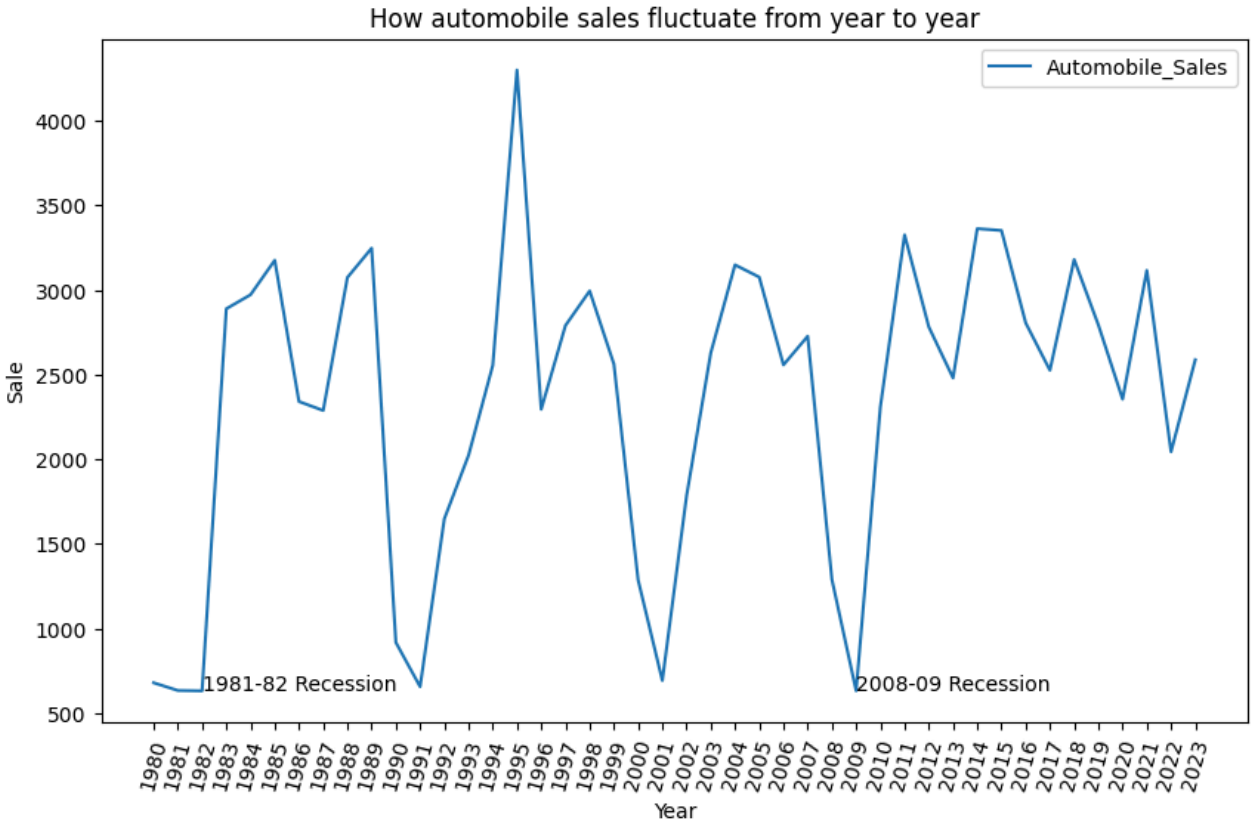
annotation for at least two years of recession

Title as Automobile Sales during Recession

▶ Click here for a hint

In [44]:

```
plt.figure(figsize=(10, 6))
df_line = df.groupby(df['Year'])['Automobile_Sales'].mean()
df_line.plot(kind = 'line')
plt.xticks(list(range(1980,2024)), rotation = 75)
plt.xlabel('Year')
plt.ylabel('Sale')
plt.title('How automobile sales fluctuate from year to year')
plt.text(1982, 631, '1981-82 Recession')
plt.text(2009, 631, '2008-09 Recession')
plt.legend()
plt.show()
```



▶ Click here for Solution template

Save this plot as "Line_Plot_1.png"

Hint: You can right lick on the plot and then click on "Save image as" option to save it on your local machine

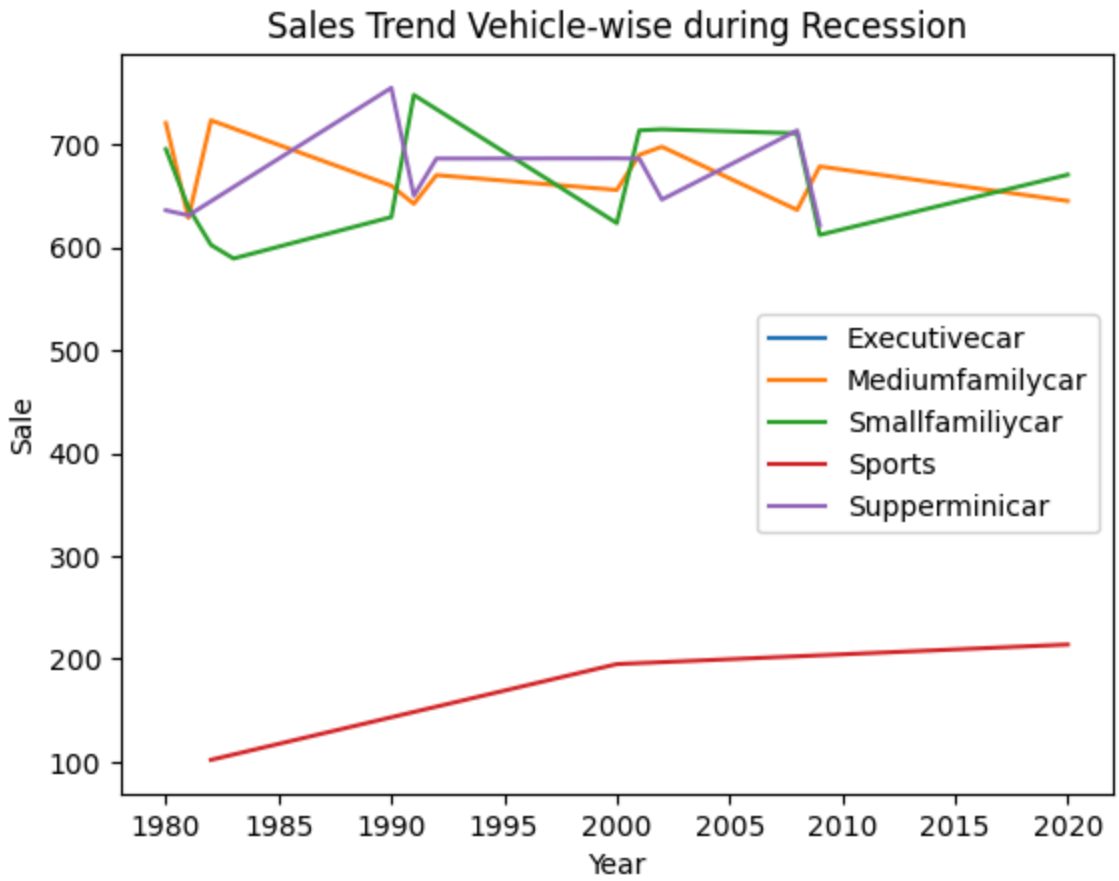
---

TASK 1.2: Plot different lines for categories of vehicle type and analyse the trend to answer the question Is there a noticeable difference in sales trends between different vehicle types during recession periods?

▶ Click here for a hint

In [43]:

```
df_rec = df[df['Recession']==1]
df_Mline = df_rec.groupby(['Year','Vehicle_Type'], as_index=False)['Automobile_Sales'].
df_Mline.set_index('Year', inplace=True)
df_Mline = df_Mline.groupby(['Vehicle_Type'])['Automobile_Sales']
df_Mline.plot(kind='line')
plt.xlabel('Year')
plt.ylabel('Sale')
plt.title('Sales Trend Vehicle-wise during Recession')
plt.legend()
plt.show()
```



▶ Click here for Solution template

From the above plot, what insights have you gained on the sales of various vehicle types?
Type in your answer below:

All car had fluctuated trend of sales in the range of around 600 to over 700, except sport car sales, which gradually rose from 100 in 1980 to almost 200 in 2000 and kept maintaining ever since. However, sport cars sales is accounted for the lowest sales across vehicle types for 40 years.

▶ Inference

Save this plot as "Line_Plot_2.png"

Hint: You can right lick on the plot and then click on "Save image as" option to save it on your local machine

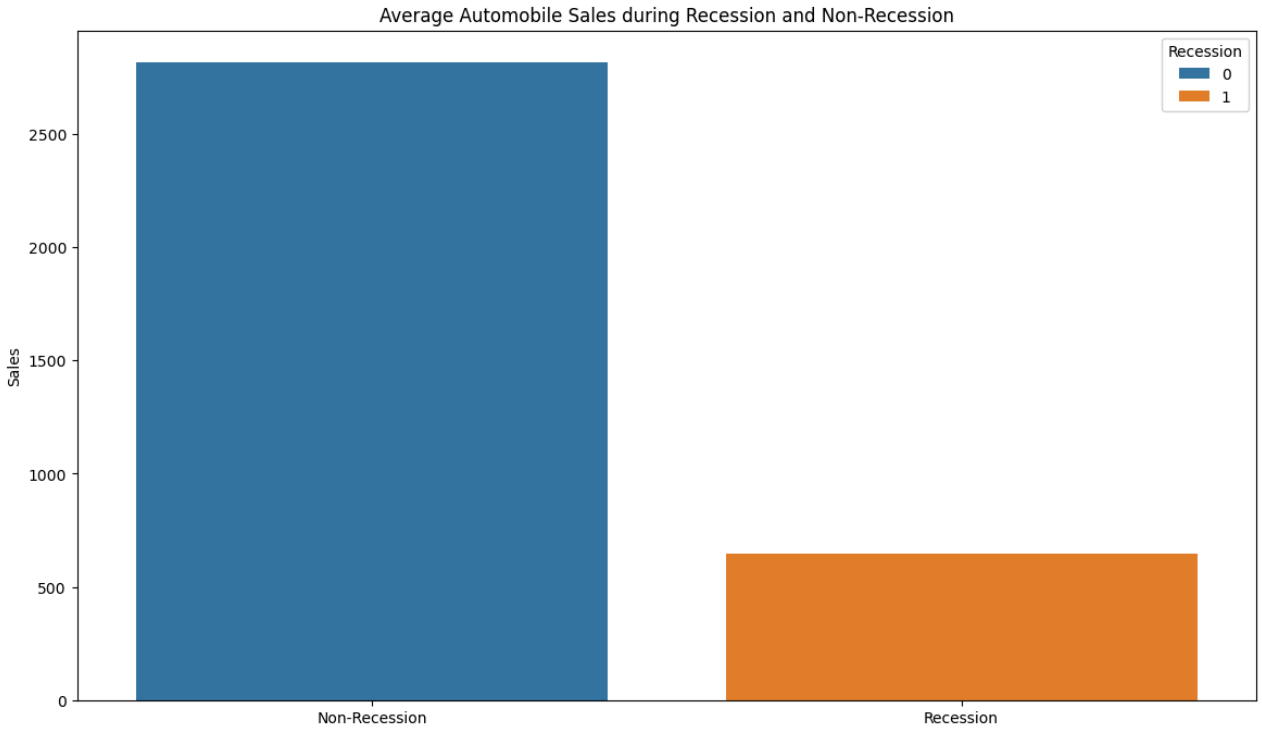
---

TASK 1.3: Use the functionality of **Seaborn Library** to create a visualization to compare the sales trend per vehicle type for a recession period with a non-recession period.

▶ Click here for a hint

In [39]:
```
new_df = df.groupby('Recession')['Automobile_Sales'].mean().reset_index()

# Create the bar chart using seaborn
plt.figure(figsize=(14,8))
sns.barplot(x='Recession', y='Automobile_Sales', hue='Recession',  data=new_df)
```

```
plt.xlabel(' ')
plt.ylabel('Sales')
plt.title('Average Automobile Sales during Recession and Non-Recession')
plt.xticks(ticks=[0, 1], labels=['Non-Recession', 'Recession'])
plt.show()
```



▶ Click here for Solution template

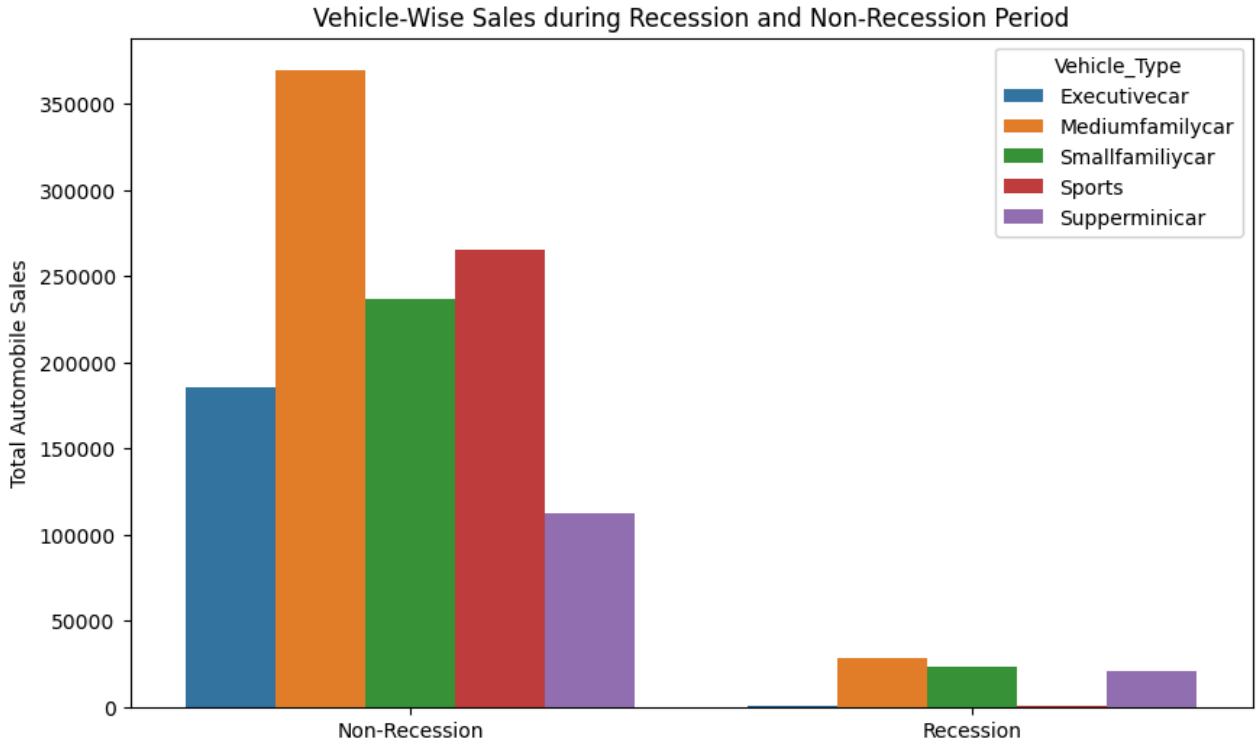
Now you want to compare the sales of different vehicle types during a recession and a non-recession period

We recommend that you use the functionality of **Seaborn Library** to create this visualization

▶ Click here for a hint

In [41]:
```
# Filter the data for recessionary periods
recession_data = df[df['Recession'] == 1]

df2=df.groupby(['Recession','Vehicle_Type'])['Automobile_Sales'].sum().reset_index()

# Create the grouped bar chart using seaborn
plt.figure(figsize=(10, 6))
sns.barplot(x='Recession', y='Automobile_Sales', hue='Vehicle_Type', data=df2)
plt.xticks(ticks=[0, 1], labels=['Non-Recession', 'Recession'])
plt.xlabel(' ')
plt.ylabel('Total Automobile Sales')
plt.title('Vehicle-Wise Sales during Recession and Non-Recession Period')
```

Out[41]:
```
Text(0.5, 1.0, 'Vehicle-Wise Sales during Recession and Non-Recession Period')
```



▶ Click here for Solution template

From the above chart what insights have you gained on the overall sales of automobiles during recession?
Type your answer below:-

The overall sales of automobiles dramatically decreased to under 500000 during recession.

▶ Inference

Save this plot as "Bar_Chart.png"

Hint: You can right lick on the plot and then click on "Save image as" option to save it on your local machine

---

TASK 1.4: Use sub plotting to compare the variations in GDP during recession and non-recession period by developing line plots for each period.

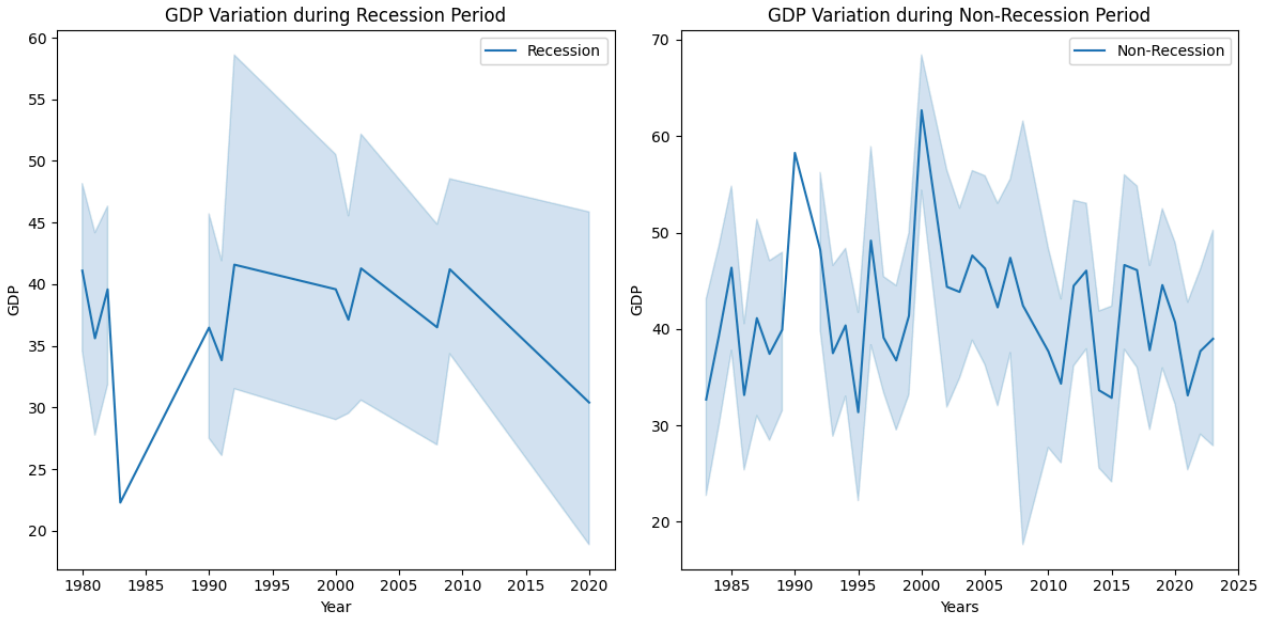
Now, you want to find more insights from the data to understand the reason.
Plot a two line charts using subplotting to answer:-

How did the GDP vary over time during recession and non-recession periods?

Make use of add_subplot() from Matplotlib for this comparision.

In [46]:
```
rec_data = df[df['Recession'] == 1]
non_rec_data = df[df['Recession'] == 0]
```

```
#Figure
fig=plt.figure(figsize=(12, 6))

#Create different axes for subploting
ax0 = fig.add_subplot(1, 2, 1) # add subplot 1 (1 row, 2 columns, first plot)
ax1 = fig.add_subplot(1 ,2 ,2 ) # add subplot 2 (1 row, 2 columns, second plot).

#plt.subplot(1, 2, 1)
sns.lineplot(x='Year', y='GDP', data=rec_data, label='Recession', ax=ax0)
ax0.set_xlabel('Year')
ax0.set_ylabel('GDP')
ax0.set_title('GDP Variation during Recession Period')

#plt.subplot(1, 2, 2)
sns.lineplot(x='Year', y='GDP', data=non_rec_data, label='Non-Recession', ax=ax1)
ax1.set_xlabel('Years')
ax1.set_ylabel('GDP')
ax1.set_title('GDP Variation during Non-Recession Period')

plt.tight_layout()
plt.show()
```



▶ Click here for Solution template

Inference

From this plot, it is evident that during recession, the GDP of the country was in a low range, might have afected the overall sales of the company

Save this plot as "Subplot.png"

Hint: You can right lick on the plot and then click on "Save image as" option to save it on your local machine

---

TASK 1.5: Develop a Bubble plot for displaying the impact of seasonality on Automobile Sales.

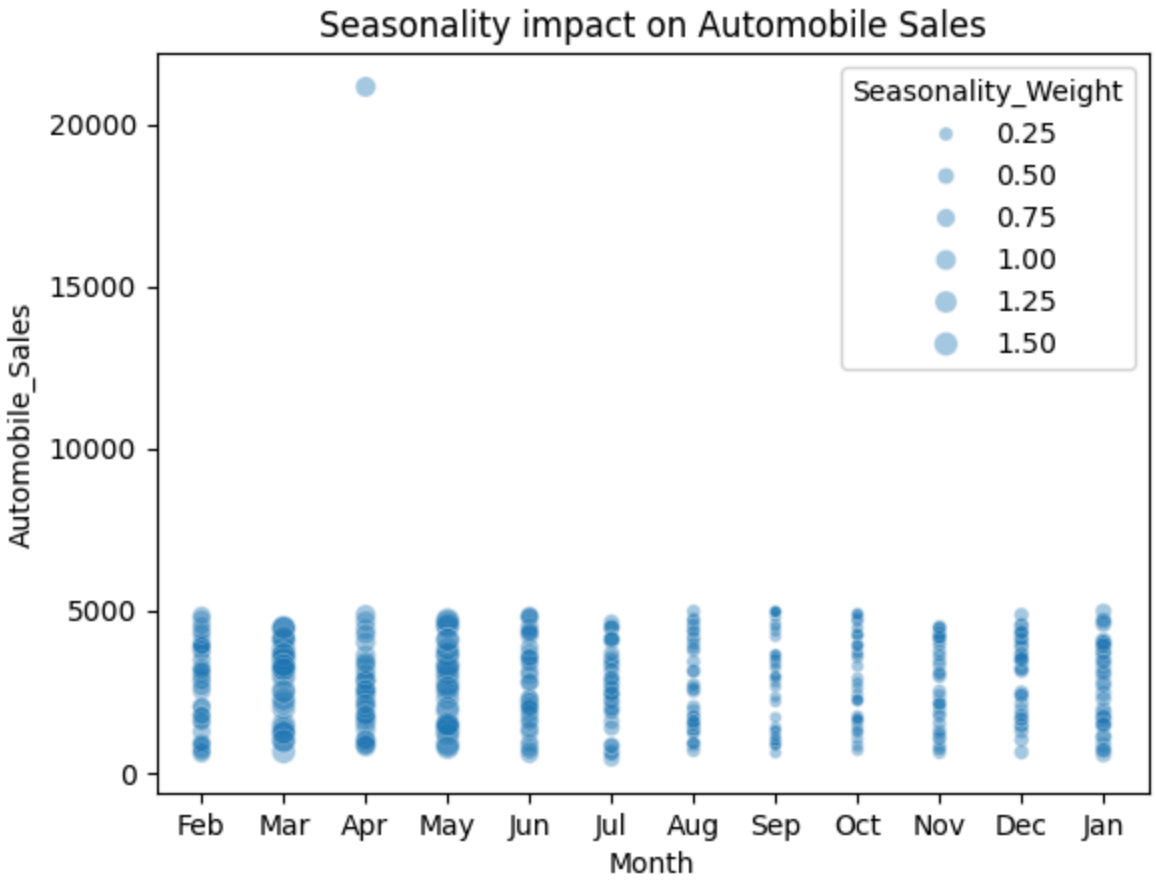
How has seasonality impacted the sales, in which months the sales were high or low? Check it for non-recession years to understand the trend

Develop a Bubble plot for displaying Automobile Sales for every month and use Seasonality Weight for representing the size of each bubble

Title this plot as 'Seasonality impact on Automobile Sales'

▶ Click here for a hint

In [51]:
```
non_rec_data = df[df['Recession'] == 0]

size=non_rec_data['Seasonality_Weight'] #for bubble effect

sns.scatterplot(data=non_rec_data, x='Month', y='Automobile_Sales', size=size, alpha=0.

#you can further include hue='Seasonality_Weight', legend=False)

plt.xlabel('Month')
plt.ylabel('Automobile_Sales')
plt.title('Seasonality impact on Automobile Sales')

plt.show()
```



▶ Click here for Solution template

Inference

From this plot, it is evident that seasonality has not affected on the overall sales. However, there is a drastic raise in sales in the month of April

Save this plot as "Bubble.png"
Hint: You can right lick on the plot and then click on "Save image as" option to save it on your local machine

---

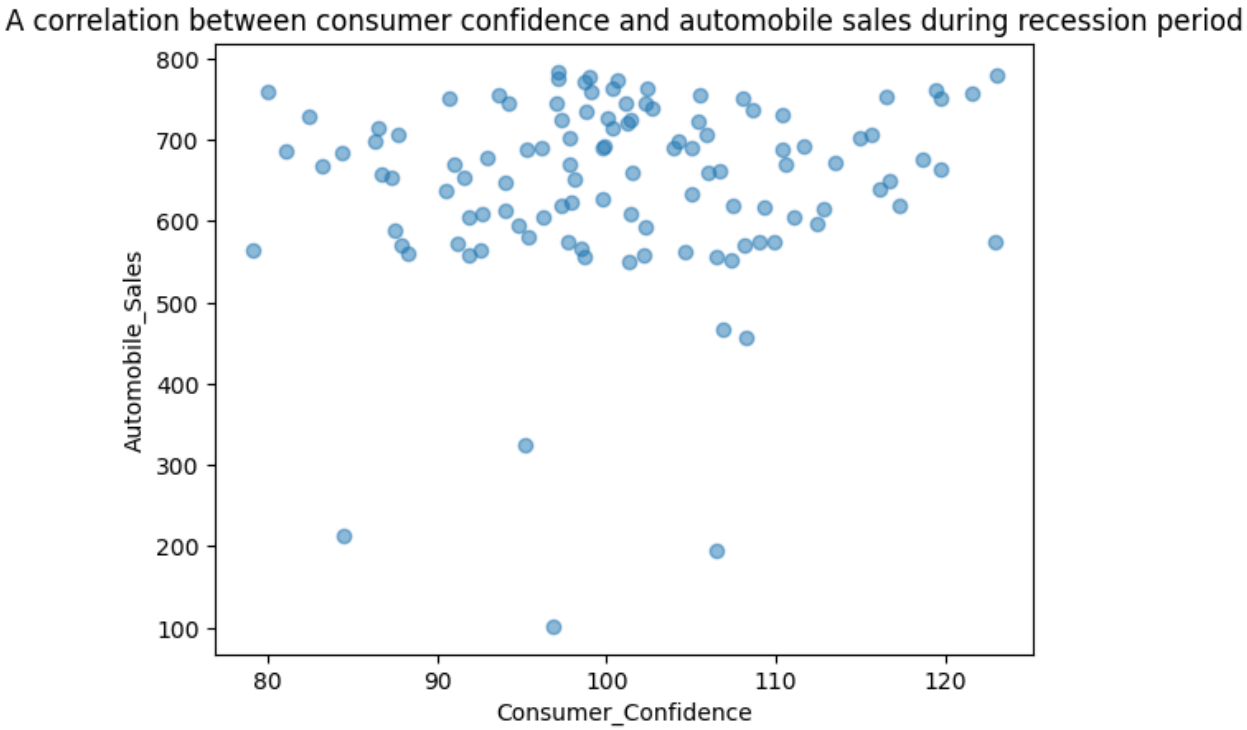
TASK 1.6: Use the functionality of Matplotlib to develop a scatter plot to identify the correlation between average vehicle price relate to the sales volume during recessions.

From the data, develop a scatter plot to identify if there a correlation between consumer confidence and automobile sales during recession period?

Title this plot as 'Consumer Confidence and Automobile Sales during Recessions'

▶ Click here for a hint

In [54]:
```
rec_data = df[df['Recession'] == 1]
plt.scatter(rec_data['Consumer_Confidence'], rec_data['Automobile_Sales'], alpha=0.5)

plt.xlabel('Consumer_Confidence')
plt.ylabel('Automobile_Sales')
plt.title('A correlation between consumer confidence and automobile sales during recess
plt.show()
```

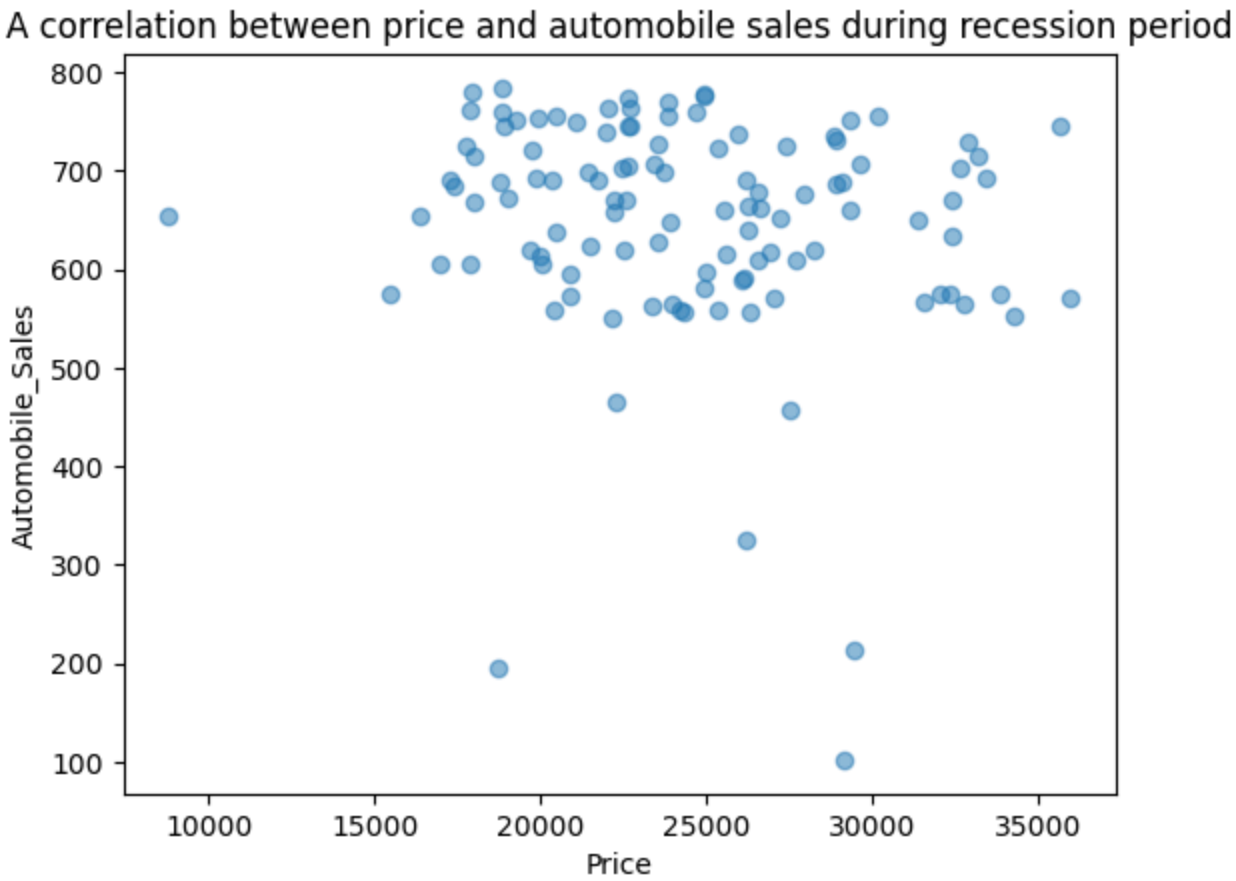


▶ Click here for Solution template

How does the average vehicle price relate to the sales volume during recessions?

Plot another scatter plot and title it as 'Relationship between Average Vehicle Price and Sales during Recessions'

In [55]:
```
rec_data = df[df['Recession'] == 1]
plt.scatter(recession_data['Price'], rec_data['Automobile_Sales'], alpha=0.5)

plt.xlabel('Price')
plt.ylabel('Automobile_Sales')
plt.title('A correlation between price and automobile sales during recession period')
plt.show()
```



▶ Click here for Solution template

Inference

There is not much relation!

Save this plot as "Scatter.png"

Hint: You can right lick on the plot and then click on "Save image as" option to save it on your local machine

---

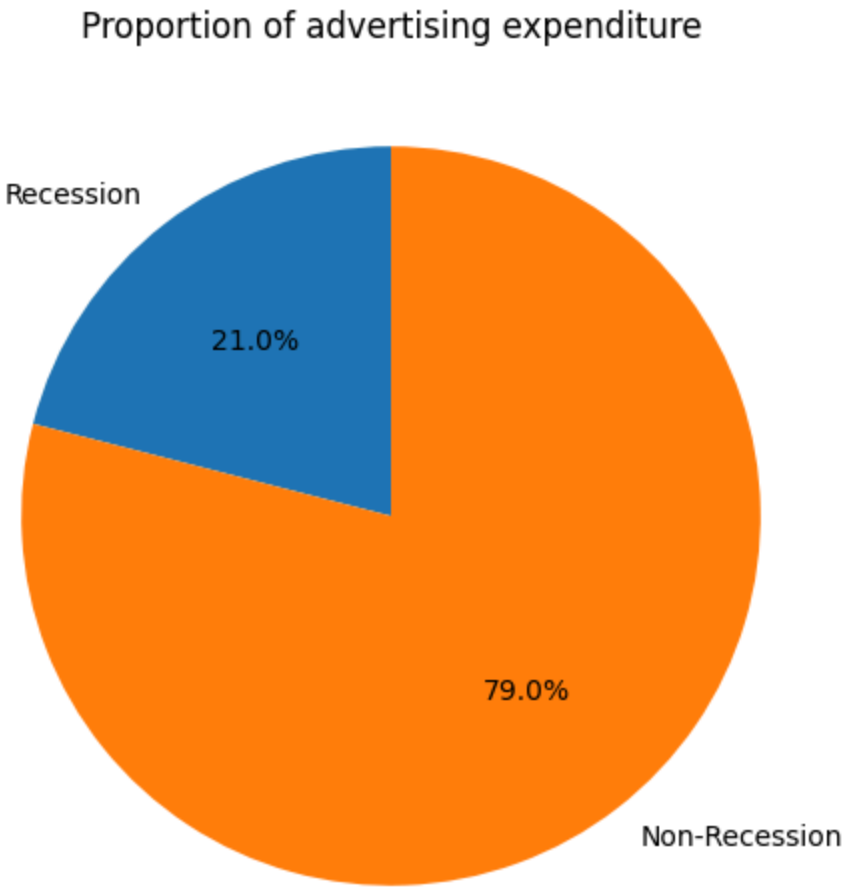
TASK 1.7: Create a pie chart to display the portion of advertising expenditure of XYZAutomotives during recession and non-recession periods.

How did the advertising expenditure of XYZAutomotives change during recession and non-recession periods?

▶ Click here for a hint

In [60]:
```
Rdata = df[df['Recession'] == 1]
NRAdata = df[df['Recession'] == 0]

# Calculate the total advertising expenditure for both periods
RAtotal = Rdata['Advertising_Expenditure'].sum()
```

```
NRAtotal = NRdata['Advertising_Expenditure'].sum()

# Create a pie chart for the advertising expenditure
plt.figure(figsize=(8, 6))

labels = ['Recession', 'Non-Recession']
sizes = [RAtotal, NRAtotal]
plt.pie(sizes, labels=labels, autopct='%1.1f%%', startangle=90)

plt.title('Proportion of advertising expenditure')

plt.show()
```



▶ Click here for Solution template

From the above plot, what insights do you find on the advertisement expenditure during recession and non recession periods? Type your answer below:-

Total exenditure during non-recession was almost 80 percent, which was four times higher than during recession, a little bit over at 20%

▶ Inference

Save this plot as "Pie_1.png"

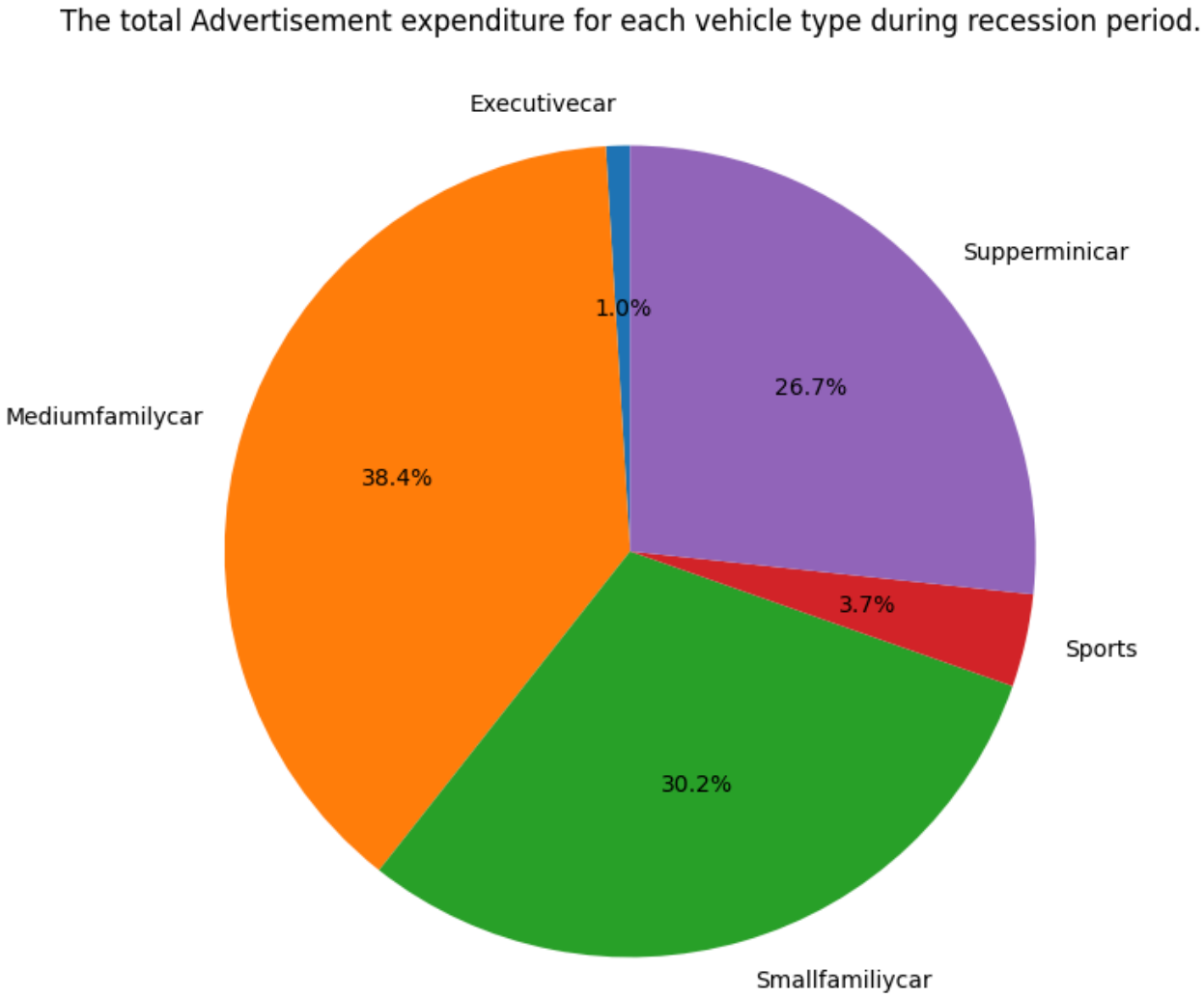
Hint: You can right lick on the plot and then click on "Save image as" option to save it on your local machine

TASK 1.8: Develop a pie chart to display the total Advertisement expenditure for each vehicle type during recession period.

Can we observe the share of each vehicle type in total expenditure during recessions?

▶ Click here for a hint

In [63]:
```
Rdata = df[df['Recession'] == 1]

# Calculate the sales volume by vehicle type during recessions
VTexpenditure = Rdata.groupby('Vehicle_Type')['Advertising_Expenditure'].sum()

# Create a pie chart for the share of each vehicle type in total expenditure during rec
plt.figure(figsize=(12, 8))

labels = VTexpenditure.index
sizes = VTexpenditure.values
plt.pie(sizes, labels=labels, autopct='%1.1f%%', startangle=90)

plt.title('The total Advertisement expenditure for each vehicle type during recession p

plt.show()
```



▶ Click here for Solution template

Inference

During recession the advertisements were mostly focued on low price range vehicle. A wise decision!

Save this plot as "Pie_2.png"

Hint: You can right lick on the plot and then click on "Save image as" option to save it on your local machine

---

TASK 1.9: Develop a lineplot to analyse the effect of the unemployment rate on vehicle type and sales during the Recession Period.

Analyze the effect of the unemployment rate on vehicle type and sales during the Recession Period

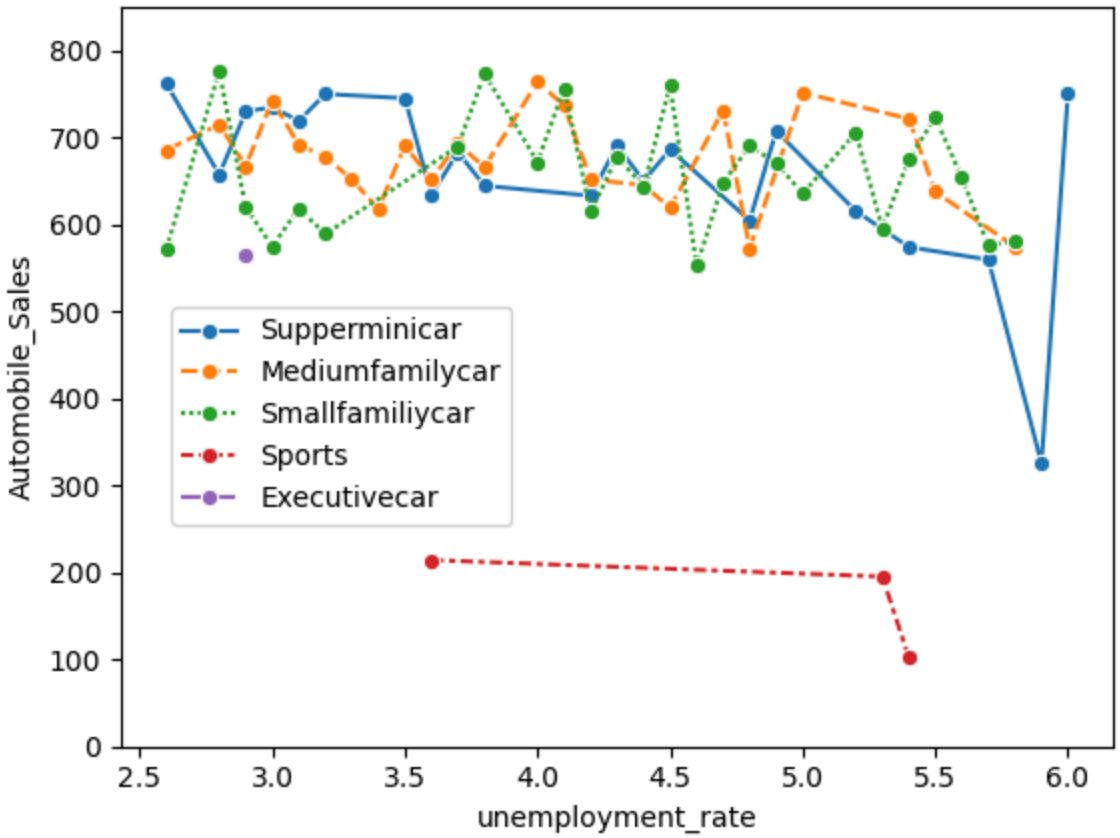
You can create a lineplot and title the plot as 'Effect of Unemployment Rate on Vehicle Type and Sales'

▶ Click here for a hint

In [66]:
```
df_rec = df[df['Recession']==1]
sns.lineplot(data=df_rec, x='unemployment_rate', y='Automobile_Sales',
             hue='Vehicle_Type', style='Vehicle_Type', markers='o', err_style=None)
plt.ylim(0,850)
plt.legend(loc=(0.05,.3))
```

```
<ipython-input-66-59a453784ef1>:2: UserWarning:
The markers list has fewer values (1) than needed (5) and will cycle, which may produce
an uninterpretable plot.
  sns.lineplot(data=df_rec, x='unemployment_rate', y='Automobile_Sales',
```

Out[66]: <matplotlib.legend.Legend at 0xce37658>



▶ Click here for Solution template

From the above plot, what insights have you gained on the sales of superminicar, smallfamilycar, mediumminicar?
Type your answer below:-

There were fluctuation.

▶ Inference

Save this plot as "line_plot_3.png"

Hint: You can right lick on the plot and then click on "Save image as" option to save it on your local machine

---

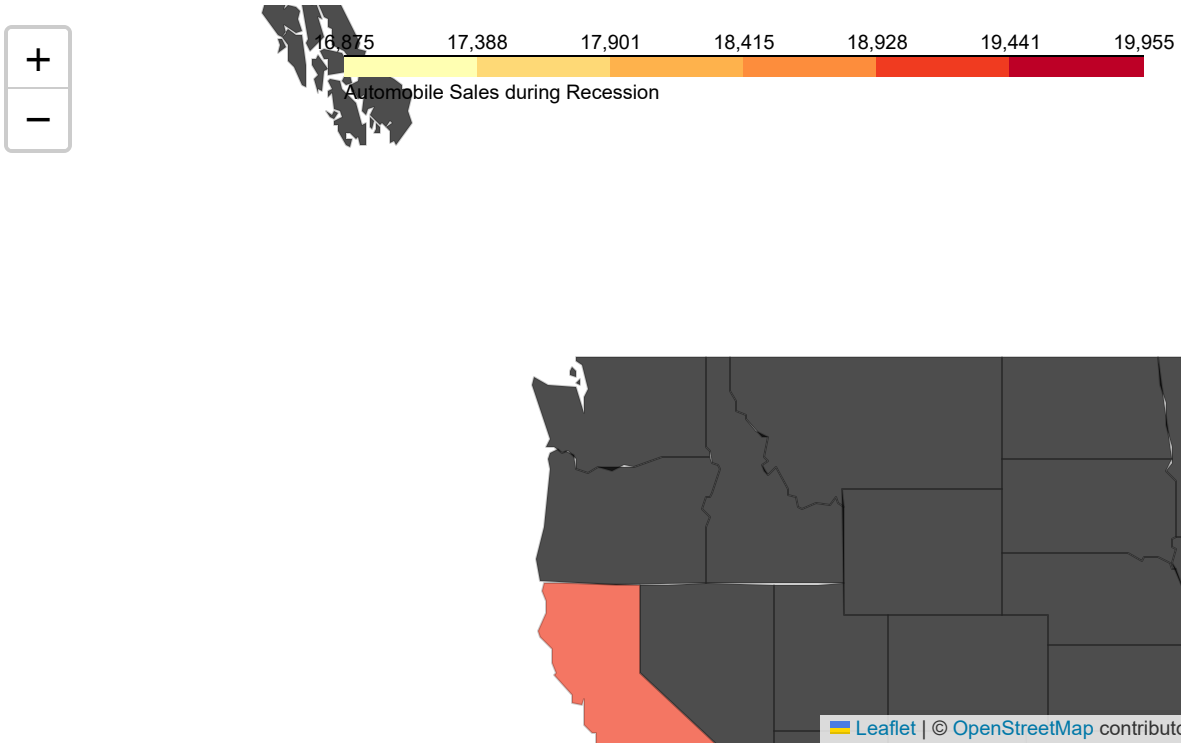
OPTIONAL : TASK 1.10 Create a map on the hightest sales region/offices of the company during recession period

In [67]:
```
from pyodide.http import pyfetch

async def download(url, filename):
    response = await pyfetch(url)
    if response.status == 200:
        with open(filename, "wb") as f:
            f.write(await response.bytes())

path = 'https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloper
await download(path, "us-states.json")

filename = "us-states.json"
```

You found that the datset also contains the location/city for company offices. Now you want to show the recession impact on various offices/city sales by developing a choropleth

In [70]:
```
data = df

# Filter the data for the recession period and specific cities
recession_data = data[data['Recession'] == 1]

# Calculate the total sales by city
sales_by_city = recession_data.groupby('City')['Automobile_Sales'].sum().reset_index()

# Create a base map centered on the United States
map1 = folium.Map(location=[37.0902, -95.7129], zoom_start=4)

# Create a choropleth layer using Folium
choropleth = folium.Choropleth(
    geo_data= 'us-states.json',  # GeoJSON file with state boundaries
    data=sales_by_city,
    columns=['City', 'Automobile_Sales'],
    key_on='feature.properties.name',
    fill_color='YlOrRd',
    fill_opacity=0.7,
    line_opacity=0.2,
    legend_name='Automobile Sales during Recession'
    ).add_to(map1)


# Add tooltips to the choropleth layer
choropleth.geojson.add_child(
    folium.features.GeoJsonTooltip(['name'], labels=True)
    )

# Display the map
map1
```

Out[70]:



▶ Click for Solution

Congratulations! You have completed the lab

Authors

Dr. Pooja

```
toggle ## Change Log
```