Name: Atul Meena

ID:240038227 ¶

1. Introduction

In [1]:
```
'''The Preliminary Data Analysis on the Carseats dataset is being done usin
g Jupyter Notebook and the following is the Exploratory Data Analysis (ED
A). The motivation for analyzing the information is therefore premised on t
he desire to understand factors that affect car seat sales in the various s
tores. Using this dataset, it is hoped that different patterns or relations
hips emerge that can assist with the formulation of business strategies wit
hin the context of retailing. The management implication of such analysis w
ould be imperative, it might aid decision-making concerning the pricing of
stock, inventory control, marketing strategies etc. This EDA will aim to sh
ow how the application of big data becomes useful in enhancing sales perfor
mance and increasing customers’ satisfaction while venturing into the car s
eat business.'''
```

Out[1]:
```
'The Preliminary Data Analysis on the Carseats dataset is being done using
Jupyter Notebook and the following is the Exploratory Data Analysis (EDA).
The motivation for analyzing the information is therefore premised on the
desire to understand factors that affect car seat sales in the various sto
res. Using this dataset, it is hoped that different patterns or relationsh
ips emerge that can assist with the formulation of business strategies wit
hin the context of retailing. The management implication of such analysis
would be imperative, it might aid decision-making concerning the pricing o
f stock, inventory control, marketing strategies etc. This EDA will aim to
show how the application of big data becomes useful in enhancing sales per
formance and increasing customers’ satisfaction while venturing into the c
ar seat business.'
```

In [3]:
```
# Importing necessary libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
import warnings
warnings.filterwarnings('ignore')
```

2. Dataset Description

In [5]:
```
'''This Carseats dataset that is present on Jupyter Notebook allows having
an adequate view of sales of car seats in various stores. It consists of da
ta on the volume of the sold product, prices and price positioning concerni
ng competitors, data on customer characteristics and the characteristics of
the store. Every row corresponds to a particular store, with columns that f
eature various aspects of a store, the volume of sales, income level of the
region, population density effective advertising budget and shelf quality.
The fact that the dataset has many variables complicates or enriches a car
seat sale analysis depending on the factors of consideration. Focusing on t
hese variables is supposed to highlight factors that could define the rate
of sales and possible issues within the reorganization of marketing and dis
tribution.'''
```

Out[5]:
```
'This Carseats dataset that is present on Jupyter Notebook allows having a
n adequate view of sales of car seats in various stores. It consists of da
ta on the volume of the sold product, prices and price positioning concern
ing competitors, data on customer characteristics and the characteristics
of the store. Every row corresponds to a particular store, with columns th
at feature various aspects of a store, the volume of sales, income level o
f the region, population density effective advertising budget and shelf qu
ality. The fact that the dataset has many variables complicates or enriche
s a car seat sale analysis depending on the factors of consideration. Focu
sing on these variables is supposed to highlight factors that could define
the rate of sales and possible issues within the reorganization of marketi
ng and distribution.'
```

In [6]:
```
# Loading the dataset into dataframe
df = pd.read_csv('Carseats.csv')
```

3. Methodology

In [7]:
```
'''Based on the excellent data analysis and visualization capability, Pytho
n is selected as the main coding language for this EDA (Nikolova et al. 202
3). Everything is done inside a Jupyter Notebook; it is an excellent place
for conducting analysis mostly because the Notebook format is designed for
it. It is very important because it enables one to easily integrate the cod
e that runs the analysis, the charts/tables that present the data, and the
text that explains results. Data handling libraries like Pandas, Libraries
for visualizations like Matplotlib and Seaborn, and libraries for implement
ing Machine learning models like scikit-learn are used. As shown in the Jup
yter Notebook in the next section, the conversion of the conclusions, recom
mendations, results, and tables generated to a PDF format of the paper will
make the presentation coherent and presentable.'''
```

Out[7]:
```
'Based on the excellent data analysis and visualization capability, Python
is selected as the main coding language for this EDA (Nikolova et al. 202
3). Everything is done inside a Jupyter Notebook; it is an excellent place
for conducting analysis mostly because the Notebook format is designed for
it. It is very important because it enables one to easily integrate the co
de that runs the analysis, the charts/tables that present the data, and th
e text that explains results. Data handling libraries like Pandas, Librari
es for visualizations like Matplotlib and Seaborn, and libraries for imple
menting Machine learning models like scikit-learn are used. As shown in th
e Jupyter Notebook in the next section, the conversion of the conclusions,
recommendations, results, and tables generated to a PDF format of the pape
r will make the presentation coherent and presentable.'
```

4. Steps of the EDA

Step - 1 Descriptive Statistics

In [8]:
```
'''EDA starts with the computation of basic statistics on the Carseats usin
g the Jupyter Notebook. This step is most important to give point estimates
of the measure of central tendencies, dispersions, and the shapes of the di
stributions. Descriptive statistics like mean, median, standard deviation,
and quartiles of the numerical variables are found with the help of pandas
functions (Silva et al. 2024). Critical analysis of categorical data includ
es looking at the frequencies as well as the proportions of the various cat
egories. It aids in finding out some of the extreme values, the range of th
e given variables, and even the noticeable patterns in the data in the firs
t instance. They include Descriptive statistics which act as a preliminary
tool and provide archetypes for further evaluations and focus on the subseq
uent adjudications of either data visualization or statistical hypothesis t
esting.'''
```

Out[8]:
```
'EDA starts with the computation of basic statistics on the Carseats using
the Jupyter Notebook. This step is most important to give point estimates
of the measure of central tendencies, dispersions, and the shapes of the d
istributions. Descriptive statistics like mean, median, standard deviatio
n, and quartiles of the numerical variables are found with the help of pan
das functions (Silva et al. 2024). Critical analysis of categorical data i
ncludes looking at the frequencies as well as the proportions of the vario
us categories. It aids in finding out some of the extreme values, the rang
e of the given variables, and even the noticeable patterns in the data in
the first instance. They include Descriptive statistics which act as a pre
liminary tool and provide archetypes for further evaluations and focus on
the subsequent adjudications of either data visualization or statistical h
ypothesis testing.'
```

In [4]:
```python
# Exploratory Data Analysis (EDA)
# Displaying basic information about the dataset
print("Shape of the dataset:", df.shape)
print("\nDataset information:\n")
df.info()
print("\nDescriptive statistics:\n")
print(df.describe())
```

```
Shape of the dataset: (400, 11)

Dataset information:

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 400 entries, 0 to 399
Data columns (total 11 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   Sales        399 non-null    float64
 1   CompPrice    399 non-null    float64
 2   Income       399 non-null    float64
 3   Advertising  400 non-null    int64  
 4   Population   400 non-null    int64  
 5   Price        400 non-null    int64  
 6   ShelveLoc    400 non-null    object 
 7   Age          400 non-null    int64  
 8   Education    400 non-null    int64  
 9   Urban        399 non-null    object 
 10  US           400 non-null    object 
dtypes: float64(3), int64(5), object(3)
memory usage: 34.5+ KB

Descriptive statistics:

            Sales   CompPrice      Income  Advertising  Population  \
count  399.000000  399.000000  399.000000   400.000000  400.000000   
mean     7.504035  124.967419   68.639098     6.635000  264.840000   
std      2.823442   15.353013   28.018750     6.650364  147.376436   
min      0.000000   77.000000   21.000000     0.000000   10.000000   
25%      5.410000  115.000000   42.500000     0.000000  139.000000   
50%      7.490000  125.000000   69.000000     5.000000  272.000000   
75%      9.320000  135.000000   91.000000    12.000000  398.500000   
max     16.270000  175.000000  120.000000    29.000000  509.000000   

            Price         Age   Education  
count  400.000000  400.000000  400.000000  
mean   115.795000   53.322500   13.900000  
std     23.676664   16.200297    2.620528  
min     24.000000   25.000000   10.000000  
25%    100.000000   39.750000   12.000000  
50%    117.000000   54.500000   14.000000  
75%    131.000000   66.000000   16.000000  
max    191.000000   80.000000   18.000000  
```

Step 2: Handling Missing Values

In [9]:
```
'''Filling in the missing values is deemed an essential component of the ED
A phase which occurs in the Jupyter Notebook (Luc et al. 2020). The use of
pandas first begins in the determination of the level of missing data and i
ts distribution concerning the given variables. For all the numerical varia
bles such as Sales, CompPrice, and Income, the blank cells are replaced wit
h relevant medians, as it is less risky than the mean and do not impact the
centrality of the data due to outliers' presence. The method of imputation
depends on the nature of the data; for categorical variables such as Urban,
the value of the most frequent occurrence (the mode) is used for imputatio
n. This approach assists in maintaining the general distribution of the ite
ms of data while at the same time making the maximum number of data items t
hat are usable.'''
```

Out[9]:
```
'Filling in the missing values is deemed an essential component of the EDA
phase which occurs in the Jupyter Notebook (Luc et al. 2020). The use of p
andas first begins in the determination of the level of missing data and i
ts distribution concerning the given variables. For all the numerical vari
ables such as Sales, CompPrice, and Income, the blank cells are replaced w
ith relevant medians, as it is less risky than the mean and do not impact
the centrality of the data due to outliers' presence. The method of imputa
tion depends on the nature of the data; for categorical variables such as
Urban, the value of the most frequent occurrence (the mode) is used for im
putation. This approach assists in maintaining the general distribution of
the items of data while at the same time making the maximum number of data
items that are usable.'
```

In [10]:
```
# Data Cleaning
# Handling missing values by replacing with median
df['Sales'].fillna(df['Sales'].median(), inplace=True)
df['CompPrice'].fillna(df['CompPrice'].median(), inplace=True)
df['Income'].fillna(df['Income'].median(), inplace=True)
df['Urban'].fillna(df['Urban'].mode()[0], inplace=True)
```

In [11]:
```
# Converting float values to integers
float_cols = df.select_dtypes(include=['float64']).columns
df[float_cols] = df[float_cols].astype(int)
```

In [12]:
```
# Removing duplicate entries
df.drop_duplicates(inplace=True)
```

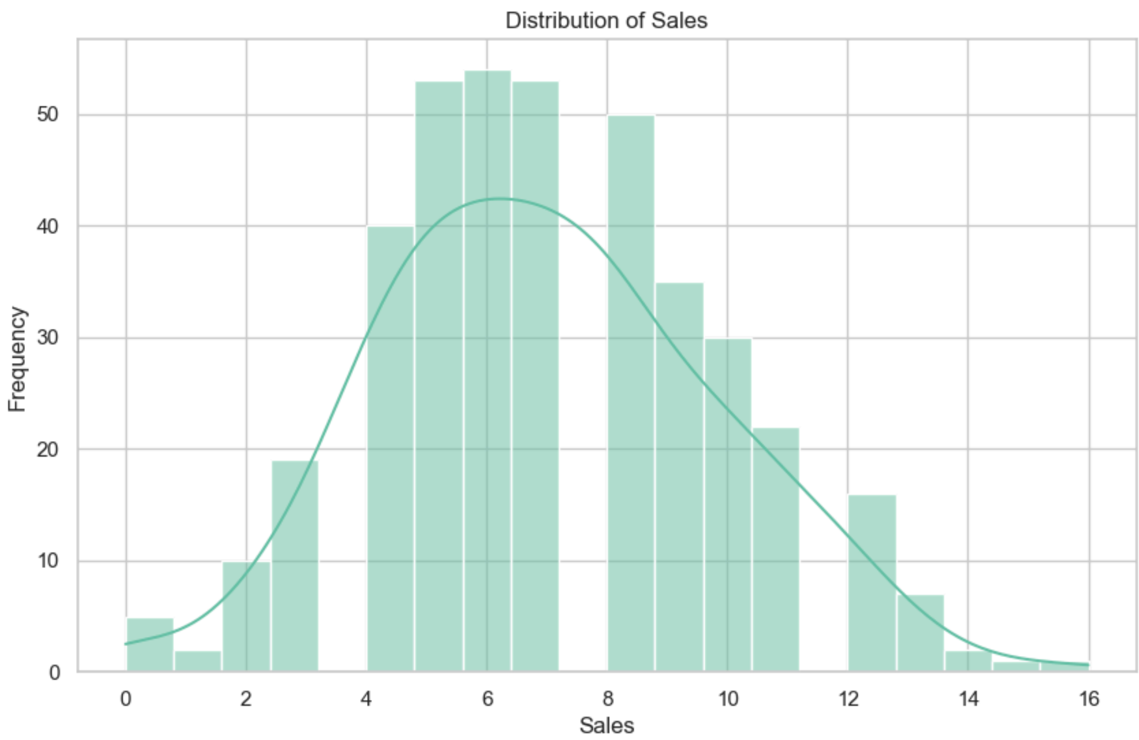
Step 3: Data Visualization

In [13]:
```
'''Descriptive analytics is used here as a form of data visualization to an
alyze the Carseats database for identifiable trends. With the help of matpl
otlib and seaborn libraries, several types of plots are made to uncover var
ious characteristics of the data. Density plots or kernel plots describe co
ntinuous items such as Sales or Price for instance. Scatter plots are usefu
l for analyzing the pairs of the elements that are continuous variables for
instance, sale and price. It can be said that two main metrics, mean and me
dian, are useful for comparing categorical variables with continuous ones,
whereas box plots help study the relation between them.'''
```

Out[13]:
```
'Descriptive analytics is used here as a form of data visualization to ana
lyze the Carseats database for identifiable trends. With the help of matpl
otlib and seaborn libraries, several types of plots are made to uncover va
rious characteristics of the data. Density plots or kernel plots describe
continuous items such as Sales or Price for instance. Scatter plots are us
eful for analyzing the pairs of the elements that are continuous variables
for instance, sale and price. It can be said that two main metrics, mean a
nd median, are useful for comparing categorical variables with continuous
ones, whereas box plots help study the relation between them.'
```

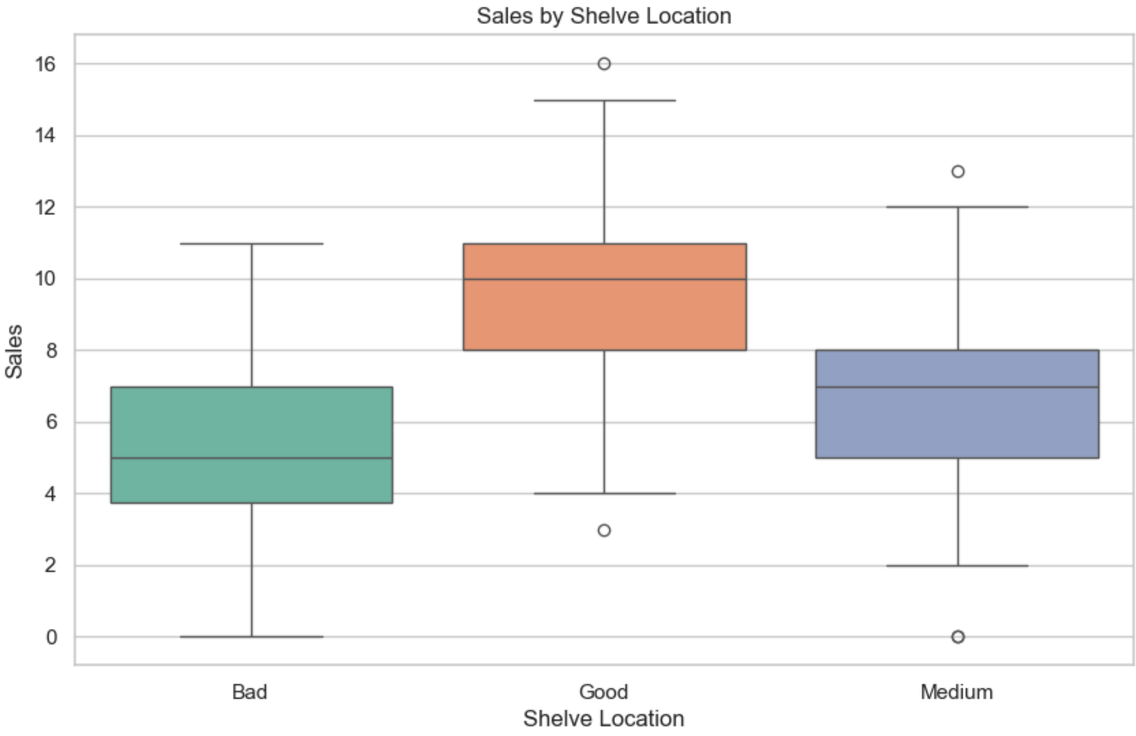
In [8]:
```
# Data Visualization
# Customizing plot aesthetics
custom_palette = sns.color_palette("Set2")
sns.set(style="whitegrid")
```

In [9]:
```
# Histogram of Sales
plt.figure(figsize=(10, 6))
sns.histplot(df['Sales'], bins=20, kde=True, color=custom_palette[0])
plt.title('Distribution of Sales')
plt.xlabel('Sales')
plt.ylabel('Frequency')
plt.show()
```



In [10]:
```
# Scatter plot of Sales vs. Price
plt.figure(figsize=(10, 6))
sns.scatterplot(x='Price', y='Sales', data=df, hue='ShelveLoc', palette=custom_palette)
plt.title('Sales vs. Price by Shelve Location')
plt.xlabel('Price')
plt.ylabel('Sales')
plt.legend(title='ShelveLoc', loc='upper right')
plt.show()
```

In [11]:
```
# Box plot of Sales by Shelve Location
plt.figure(figsize=(10, 6))
sns.boxplot(x='ShelveLoc', y='Sales', data=df, palette=custom_palette)
plt.title('Sales by Shelve Location')
plt.xlabel('Shelve Location')
plt.ylabel('Sales')
plt.show()
```



In [12]:
```
# Relationship between Sales, Income, and Price by Scatter plot
plt.figure(figsize=(10, 6))
sns.scatterplot(x='Income', y='Sales', hue='Price', data=df, palette='virid
is')
plt.title('Sales vs. Income by Price')
plt.xlabel('Income')
plt.ylabel('Sales')
plt.legend(title='Price')
plt.show()
```

Step 4: Unique Values and Frequencies

In [14]:
```
'''This step relates to the exploration of categorical variables that exist in the Carseats dataset. Pandas is used to output the unique values and counts of the different categories of the variables ShelveLoc, Urban, and US. Distribution analysis allows us to assess imbalances in the number of various categories, or categories that appear extremely seldom, which are crucial when performing subsequent analyses. Frequencies are represented for more understandable information either in the form of bar diagrams/ charts or pie diagrams/charts. This step is necessary to know the distribution of the dataset and can help in the future for feature engineering or for the use of techniques such as oversampling or undersampling in future modelling phases.'''
```

Out[14]:
```
'This step relates to the exploration of categorical variables that exist in the Carseats dataset. Pandas is used to output the unique values and counts of the different categories of the variables ShelveLoc, Urban, and US. Distribution analysis allows us to assess imbalances in the number of various categories, or categories that appear extremely seldom, which are crucial when performing subsequent analyses. Frequencies are represented for more understandable information either in the form of bar diagrams/ charts or pie diagrams/charts. This step is necessary to know the distribution of the dataset and can help in the future for feature engineering or for the use of techniques such as oversampling or undersampling in future modelling phases.'
```

In [13]:
```
# Unique values of Shelve Location and their frequencies
shelveloc_counts = df['ShelveLoc'].value_counts()
print("\nUnique values and their frequencies:\n")
print(shelveloc_counts)
```

```
Unique values and their frequencies:

ShelveLoc
Medium    219
Bad        96
Good       85
Name: count, dtype: int64
```

Step 5: Contingency Table and Statistical Test

In [11]:
```
'''Cross tabulation is done to identify contingencies between the categoric
al variables identified from the features of the Carseats dataset (Franková
et al. 2022). For example, a table with Urban and US variables is used to a
nalyze if store location impacts this country in any way. After this, a chi
-square test of independence is used to determine the significance of the a
bove-stated variables. The test yields a p-value, from which conclusions ar
e made on whether the relationship under analysis is significant or not. Th
is analysis assists in identifying a relationship in store distribution bet
ween urban and rural areas of various countries.'''
```

Out[11]:
```
'Step 5: Contingency Table and Statistical Test\nCross tabulation is done
to identify contingencies between the categorical variables identified fro
m the features of the Carseats dataset (Franková et al. 2022). For exampl
e, a table with Urban and US variables is used to analyze if store locatio
n impacts this country in any way. After this, a chi-square test of indepe
ndence is used to determine the significance of the above-stated variable
s. The test yields a p-value, from which conclusions are made on whether t
he relationship under analysis is significant or not. This analysis assist
s in identifying a relationship in store distribution between urban and ru
ral areas of various countries.'
```

In [14]:
```
# Contingency table and chi-square test between Urban and US
cont_table = pd.crosstab(df['Urban'], df['US'])
chi2_stat, p_val, _, _ = stats.chi2_contingency(cont_table)
print("\nContingency Table between Urban and US:\n")
print(cont_table)
print(f"\nChi-square test result: chi2 = {chi2_stat:.2f}, p-value = {p_va
l:.4f}")
```

```
Contingency Table between Urban and US:

US      No  Yes
Urban
No      46   72
Yes     96  186

Chi-square test result: chi2 = 0.68, p-value = 0.4082
```

Step 6: Subset Analysis and Descriptive Statistics

In [15]:
```
'''Subset analysis entails the identification of parts of Carseats and sort
ing them based on set parameters. For example, a subset can be powered abou
t stores located in urban zones, where overall sales are higher than median
s. Descriptive statistics are then computed on this subset, both numerical
and graphical means such as the measure of central tendency, measure of dis
persion, and that based on distribution. Such an approach is more conducive
to gaining a great understanding of certain aspects of certain segments of
the data. Differences between any of these subset statistics with those of
the entire data set can provide characteristics or corresponding trends in
certain groups.'''
```

Out[15]:
```
'Subset analysis entails the identification of parts of Carseats and sorti
ng them based on set parameters. For example, a subset can be powered abou
t stores located in urban zones, where overall sales are higher than media
ns. Descriptive statistics are then computed on this subset, both numerica
l and graphical means such as the measure of central tendency, measure of
dispersion, and that based on distribution. Such an approach is more condu
cive to gaining a great understanding of certain aspects of certain segmen
ts of the data. Differences between any of these subset statistics with th
ose of the entire data set can provide characteristics or corresponding tr
ends in certain groups.'
```

In [15]:
```
# Subset of rows based on criteria and descriptive statistics
median_sales = df['Sales'].median()
subset_data = df[(df['Urban'] == 'Yes') & (df['Sales'] > median_sales)]
print("\nDescriptive statistics for subset (Urban='Yes' and Sales > media
n):\n")
print(subset_data.describe())
```

```
Descriptive statistics for subset (Urban='Yes' and Sales > median):

            Sales   CompPrice      Income  Advertising  Population  \
count  110.000000  110.000000  110.000000   110.000000  110.000000
mean     9.854545  127.290909   73.627273     9.309091  268.454545
std      1.739115   16.898697   26.138256     7.261493  150.430553
min      8.000000   77.000000   21.000000     0.000000   14.000000
25%      8.000000  116.000000   55.000000     1.000000  142.000000
50%     10.000000  128.500000   75.500000    10.000000  279.500000
75%     11.000000  138.000000   94.000000    15.000000  393.250000
max     16.000000  175.000000  120.000000    29.000000  508.000000

            Price         Age   Education
count  110.000000  110.000000  110.000000
mean   107.018182   50.063636   13.572727
std     24.186956   15.003533    2.703919
min     24.000000   25.000000   10.000000
25%     89.250000   37.000000   11.000000
50%    108.000000   50.000000   13.000000
75%    123.750000   61.000000   16.000000
max    166.000000   80.000000   18.000000
```

Step 7: Statistical Test of Means

In [16]:
```
'''Inferring from the control variables, a test of means is conducted to an
alyze the sales statistics of two groups of data. For instance, a t-test mi
ght be used to compare the average sales between a store with good shelf po
sitions and one with poor shelf positions. The test outputs a t-statistic a
nd p-value associated with the result of the test. Analyzing these outcomes
requires deciding as to whether the differences in the means reached the st
atistical level of significance, or were found by chance. Conducted analysi
s can establish the correlation between some factors and the level of sales
production.'''
```

Out[16]:
```
'Inferring from the control variables, a test of means is conducted to ana
lyze the sales statistics of two groups of data. For instance, a t-test mi
ght be used to compare the average sales between a store with good shelf p
ositions and one with poor shelf positions. The test outputs a t-statistic
and p-value associated with the result of the test. Analyzing these outcom
es requires deciding as to whether the differences in the means reached th
e statistical level of significance, or were found by chance. Conducted an
alysis can establish the correlation between some factors and the level of
sales production.'
```

In [16]:
```
# T-test for Sales between Good and Bad Shelve Location
sales_good = df[df['ShelveLoc'] == 'Good']['Sales']
sales_bad = df[df['ShelveLoc'] == 'Bad']['Sales']
t_stat, p_val = stats.ttest_ind(sales_good, sales_bad, equal_var=False)
print(f"\nT-test result for Sales between ShelveLoc 'Good' and 'Bad': t-sta
t = {t_stat:.2f}, p-value = {p_val:.4f}")
```

```
T-test result for Sales between ShelveLoc 'Good' and 'Bad': t-stat = 12.7
3, p-value = 0.0000
```

Step 8: Grouped Summary Statistics

In [17]:
```
'''Tables which are carrying grouped summary statistics involve forming a t
able with categories and the data put into a table is categorized. For exam
ple, the mean Sale can be computed for each of the shelf location categorie
s namely Good, Medium and Bad. The same process is then followed for other
categorical variables and for any specific numerical measurement that may b
e of interest (Assel et al. 2022). The resulting tables offer a summarized
view of how various scissor’s KPIs differ across the identified groups. Bar
or heat maps may be included in these tables to facilitate understanding. T
his step is important to look for patterns specific to various segments of
the business and can define the advantages or possible improvements within
certain classes.'''
```

Out[17]:
```
'Tables which are carrying grouped summary statistics involve forming a ta
ble with categories and the data put into a table is categorized. For exam
ple, the mean Sale can be computed for each of the shelf location categori
es namely Good, Medium and Bad. The same process is then followed for othe
r categorical variables and for any specific numerical measurement that ma
y be of interest (Assel et al. 2022). The resulting tables offer a summari
zed view of how various scissor’s KPIs differ across the identified group
s. Bar or heat maps may be included in these tables to facilitate understa
nding. This step is important to look for patterns specific to various seg
ments of the business and can define the advantages or possible improvemen
ts within certain classes.'
```

In [17]:
```
# Grouped summary statistics of Mean Sales by ShelveLoc
mean_sales_shelveloc = df.groupby('ShelveLoc')['Sales'].mean().reset_index
()
print("\nMean Sales by ShelveLoc:")
print(mean_sales_shelveloc)
```

```
Mean Sales by ShelveLoc:
  ShelveLoc     Sales
0       Bad  5.072917
1      Good  9.694118
2    Medium  6.789954
```

Step 9: Linear Regression Model

In [18]:
```
'''The target analysis, which consists of the prediction of sales, is accom
plished by applying a linear regression model on a set of features. It star
ts with baseline data with variable encoding and data division into the tra
ining set and the test set. Then the model is trained on the training data
and a preview is made on the test set. Very commonly, the performance of mo
dels is measured using parameters like Mean Squared Error (MSE), Root Mean
Squared Error (RMSE), and R-squared (R2). Analyzing those coefficients give
s an understanding of the accuracy of the forecast and the percentage of va
riance in sales accounted for by the chosen features.'''
```

Out[18]:
```
'The target analysis, which consists of the prediction of sales, is accomp
lished by applying a linear regression model on a set of features. It star
ts with baseline data with variable encoding and data division into the tr
aining set and the test set. Then the model is trained on the training dat
a and a preview is made on the test set. Very commonly, the performance of
models is measured using parameters like Mean Squared Error (MSE), Root Me
an Squared Error (RMSE), and R-squared (R2). Analyzing those coefficients
gives an understanding of the accuracy of the forecast and the percentage
of variance in sales accounted for by the chosen features.'
```

In [18]:
```
# Linear Regression Model
# Converting categorical columns to dummy variables
df_encoded = pd.get_dummies(df, drop_first=True)

# Defining features and target variable
X = df_encoded.drop('Sales', axis=1)
y = df_encoded['Sales']

# Splitting the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, ra
ndom_state=42)

# Creating and training the linear regression model
model = LinearRegression()
model.fit(X_train, y_train)

# Predicting on the test set
y_pred = model.predict(X_test)

# Calculating and printing the accuracy metrics in a table format
accuracy_metrics = pd.DataFrame({
    'Metric': ['Mean Squared Error (MSE)', 'Root Mean Squared Error (RMS
E)', 'R-squared (R2)'],
    'Value': [mean_squared_error(y_test, y_pred), np.sqrt(mean_squared_erro
r(y_test, y_pred)), r2_score(y_test, y_pred)]
})
print("\nAccuracy Metrics:\n")
print(accuracy_metrics)
```

```
Accuracy Metrics:

                          Metric     Value
0        Mean Squared Error (MSE)  1.065700
1  Root Mean Squared Error (RMSE)  1.032327
2                   R-squared (R2)  0.893103
```

5. Conclusion

In [19]:
```
'''Exploratory Data Analysis of the Carseats dataset provides insights into the Carseats consumption patterns about various characteristics. Some of the research findings include determinants such as price factors, shelf position and demographic factors concerning sales performance. This discussion goes to support the arguments that the use of data in strategic analysis and planning of the retail business cannot be overemphasized. Some of the key findings derived include the relevance of the use of graphical and quantitative analysis to discover concealed trends and that sales forecasts cannot be determined by a single variable. Thus, these ideas can be applied to diverse operational choices, ranging from manufacturing or purchasing to communication and promotion, which might result in enhanced sales outcomes and satisfaction of the consumers in the specified area of automobiles' components, namely car seats.'''
```

Out[19]:
```
'Exploratory Data Analysis of the Carseats dataset provides insights into the Carseats consumption patterns about various characteristics. Some of the research findings include determinants such as price factors, shelf position and demographic factors concerning sales performance. This discussion goes to support the arguments that the use of data in strategic analysis and planning of the retail business cannot be overemphasized. Some of the key findings derived include the relevance of the use of graphical and quantitative analysis to discover concealed trends and that sales forecasts cannot be determined by a single variable. Thus, these ideas can be applied to diverse operational choices, ranging from manufacturing or purchasing to communication and promotion, which might result in enhanced sales outcomes and satisfaction of the consumers in the specified area of automobiles' components, namely car seats.'
```

Reference List

In [20]:

```
'''Assel, F. and Löwe, S., 2022. Digital Platforms for Textile Waste Recovery: An exploratory study about how Digital Platforms strenghten the Waste Recovery Stream in the Textile, Apparel and Clothing industry. https://www.diva-portal.org/smash/get/diva2:1660382/FULLTEXT02
Franková, A.N., Factors influencing purchasing decisions of baby care products in the Icelandic market (Doctoral dissertation). https://skemman.is/bitstream/1946/42911/1/Masters%20Thesis%20-%20Alexandra%20Nicole%20Frank%20-%20Final%20Draft.pdf
Luc, A., Lê, S. and Philippe, M., 2020. Nudging consumers for relevant data using Free JAR profiling: An application to product development. Food Quality and Preference, 79, p.103751. https://www.sciencedirect.com/science/article/am/pii/S0950329318309790
Nikolova, M. and Nterai, S., 2023. Space Utilization in Road Transportation: A Case Study of Truck Deliveries. https://gupea.ub.gu.se/bitstream/handle/2077/77662/LTM%202023-90.pdf?sequence=1&isAllowed=y
Silva, A., Simões, A.C. and Blanc, R., 2024. Supporting decision-making of collaborative robot (cobot) adoption: The development of a framework. Technological Forecasting and Social Change, 204, p.123406. https://www.sciencedirect.com/science/article/pii/S0040162524002026'''
```

Out[20]:

```
'Assel, F. and Löwe, S., 2022. Digital Platforms for Textile Waste Recovery: An exploratory study about how Digital Platforms strenghten the Waste Recovery Stream in the Textile, Apparel and Clothing industry. https://www.diva-portal.org/smash/get/diva2:1660382/FULLTEXT02\nFranková, A.N., Factors influencing purchasing decisions of baby care products in the Icelandic market (Doctoral dissertation). https://skemman.is/bitstream/1946/42911/1/Masters%20Thesis%20-%20Alexandra%20Nicole%20Frank%20-%20Final%20Draft.pdf\nLuc, A., Lê, S. and Philippe, M., 2020. Nudging consumers for relevant data using Free JAR profiling: An application to product development. Food Quality and Preference, 79, p.103751. https://www.sciencedirect.com/science/article/am/pii/S0950329318309790\nNikolova, M. and Nterai, S., 2023. Space Utilization in Road Transportation: A Case Study of Truck Deliveries. https://gupea.ub.gu.se/bitstream/handle/2077/77662/LTM%202023-90.pdf?sequence=1&isAllowed=y\nSilva, A., Simões, A.C. and Blanc, R., 2024. Supporting decision-making of collaborative robot (cobot) adoption: The development of a framework. Technological Forecasting and Social Change, 204, p.123406. https://www.sciencedirect.com/science/article/pii/S0040162524002026'
```

In [1]:

In [ ]: