GRT INSTITUTE OF ENGINEERING AND TECHNOLOGY, TIRUTTANI - 631209

Approved by AICTE, New Delhi Affiliated to Anna University, Chennai

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

FUTURE SALES PREDICTION

PROJECT REPORT

SUBMITTED BY

TEJA.M

3RD YEAR 5TH SEM

110321104054

tejamm52@gmail.com

ANNA UNIVERSITY:CHENNAI600025

BONAFIDECERTIFICATE

Certified that this project report**"FUTURE SALES PREDICTION"** is the bonafideworkof**"TEJA.M [110321104054]"**whocarriedouttheprojectworkundermy oursupervision.

Dr.N.Kamal M.E.,Ph.D.,
HOD

Department ofComputerScience And Engineering

GRTInstituteofEngineeringand Technology

Tiruttani

Mr.T.A. VinayagamM.Tech.,
Assistantprofessor

DepartmentofComputerScienceAnd Engineering

GRTInstituteofEngineeringand Technology

Tiruttani

Certified that the candidates were examined in Viva-voce in the

ExaminationHeldon01-11-2023

ACKNOWLEDGEMENT

WethankourManagementforprovidingusallsupport tocompletethis projectsuccessfully. Oursincerethankstohonorable**Chairman**,**Shri.G.RAJENDRANandManagingDirector,Shri.G.R. RADHAKRISHNAN** forcreatingawonderfulatmosphereinsidethecampus.

We are very grateful to **Dr.S. ARUMUGAM, M.E., Ph.D., Principal**, for providing us withconsistent guidance and motivation to execute a real time project to learn and experience theprojectwork in anenvironmentsto complete ourproject successfully.

Our sincere thanks to **Dr.N. KAMAL, M.E., Ph.D., Professor and Head, Department ofComputer Science and Engineering** for giving me this wonderful opportunity to do the projectandprovidingtherequirefacilities to fulfill our work.

Wearehighlyindebtedand thankfulto ourproject Evaluators **Mrs V. Priya M.E**., and **Mrs. Edith Esther M.E., AssistantProfessor,DepartmentofComputerScienceandEngineering**forhis immensesupport in doingtheproject.

We are very grateful to our internal guide **Mr. T.A. VINAYAGAM, M.Tech., AssistantProfessor, Department of Computer Science and Engineering** for guiding us with hervaluablesuggestions to complete ourproject.

Wealsodedicateequalandgratefulacknowledgementtoall the**respectablemembersof thefaculty and lab in-charges** of the Department of Computer Science and Engineering, friends andourfamilies fortheirmotivation, encouragement and continuous support.

Our sincere thanks **to IBM and Skill Up Team members** for giving me this wonderful opportunity to do the projectandprovidingtherequireguidance and valuable online sessions.

TABLEOFCONTENTS

ABSTRACT

Sales prediction, the art and science of forecasting future sales performance, is an imperative in today's dynamic and competitive business landscape. With the continuous evolution of technology, data analytics, and consumer behavior, accurate sales prediction has become the cornerstone of strategic decision-making for organizations across industries.

This abstract explores the multifaceted world of future sales prediction, delving into the methodologies, challenges, and implications of harnessing predictive analytics. As businesses strive to navigate the complexities of a globalized marketplace, understanding and mastering the science of sales forecasting is essential for achieving sustained growth, optimizing resource allocation, and capitalizing on emerging opportunities.

The journey begins by elucidating the foundational concepts of sales prediction, emphasizing the role of historical data as the bedrock upon which accurate forecasts are built. We traverse through various forecasting techniques, from time-series analysis to machine learning algorithms, shedding light on their strengths and limitations. The advent of artificial intelligence, big data, and advanced analytics has ushered in a new era of predictive precision, enabling organizations to dissect customer behavior, market trends, and economic indicators with unprecedented accuracy.

As we explore the practical applications of sales prediction, we unearth its significance in optimizing inventory management, streamlining production processes, and enhancing marketing strategies. The ability to anticipate customer demand, tailor product offerings, and personalize customer experiences has never been more critical.

Yet, amid the promise of predictive prowess, we confront the challenges that loom large. Ethical considerations surrounding data privacy and the responsible use of predictive analytics demand careful attention. Furthermore, the ever-changing nature of consumer preferences and global events introduces an element of uncertainty that requires adaptability and resilience in predictive models.

This abstract does not merely underscore the importance of sales prediction; it underscores the imperative of continuous learning and adaptation in an era where business strategies hinge on data-driven insights. It invites us to contemplate the future, where predictive analytics will evolve further, perhaps reaching the realm of prescriptive analytics, enabling businesses not only to predict outcomes but also to prescribe actions to shape those outcomes.

CHAPTER-1

INTRODUCTION

In today's fast-paced and ever-evolving business environment, the ability to foresee the future has become an invaluable asset. Among the myriad of forecasting endeavors, future sales prediction stands tall as a pivotal domain, offering organizations a strategic advantage in navigating the complex landscape of commerce. It is the compass that guides decision-makers, the oracle that unveils market trends, and the bedrock upon which successful business strategies are built.

Sales prediction, also known as sales forecasting, is the process of using historical data, market analysis, and advanced analytical techniques to estimate future sales performance. In essence, it empowers businesses to gaze into the crystal ball of data, discern patterns, and anticipate customer behaviors and market dynamics. Whether you are a global conglomerate, a tech startup, or a neighborhood retail store, the ability to predict sales accurately can mean the difference between prosperity and stagnation.

As we embark on this journey through the realm of future sales prediction, we shall uncover the multifaceted facets of this critical discipline. This introduction serves as a gateway to understanding why it matters, how it is achieved, and the profound impact it has on diverse sectors of the economy.

Why Sales Prediction Matters:

In an era characterized by relentless competition, consumer empowerment, and technological innovation, businesses need to be nimble and responsive. Sales prediction is the North Star that guides this agility. It allows companies to:

1. Optimize Resource Allocation: By anticipating demand, businesses can allocate resources efficiently, ensuring they have the right amount of inventory, staff, and marketing efforts in place.

2. Improve Cash Flow Management: Accurate sales forecasts enable better financial planning, helping organizations manage their cash flow effectively and invest strategically.

3. Enhance Customer Experiences: Predictive analytics enable personalized marketing and product recommendations, creating more satisfying and tailored experiences for customers.

4. Inform StrategicDecision-Making:Sales predictions inform a wide range of strategic decisions, from product development and pricing strategies to market expansion and supply chain management.

5. Mitigate Risks:By identifying potential market downturns or disruptions early, businesses can develop contingency plans and mitigate risks effectively.

How Sales Prediction is Achieved:

Sales prediction is an amalgamation of art and science, where data-driven insights meet industry expertise. It involves:

1. Data Collection:Gathering and organizing historical sales data, customer information, market trends, and external factors influencing sales.

2. Analysis: Employing various statistical and machine learning techniques to analyze data and identify patterns, seasonality, and trends.

3. Model Building: Creating predictive models that incorporate historical data and relevant variables to make accurate forecasts.

4. Validation:Rigorous testing and validation of models to ensure their accuracy and reliability in different scenarios.

5. Continuous Learning: Adapting and refining models as new data becomes available to maintain their predictive power.

As we delve deeper into the world of future sales prediction, we will explore the methodologies, technologies, and best practices that organizations employ to harness its potential. From time-series analysis and machine learning algorithms to the ethical considerations that accompany predictive capabilities, we will navigate the intricacies of this dynamic field, revealing how it shapes the future of businesses and economies alike. To grasp the essence of future sales predictions, one must delve into the methods and techniques that underpin this field. It is a dynamic blend of art and science, where data-driven insights meet human intuition and expertise. The collection and utilization of vast amounts of customer data raise concerns about privacy and security. Ethical data handling practices are essential to build trust with customers.

PROBLEM DEFINITION

The Challenge of Accurate Future Sales Prediction

The problem is todevelop a predictive model that uses historical sales data to forecast future sales for a company. The objective is to create a tool that enables the company to optimize inventory management and make informed business decisions based on datadriven sales predictions. This project involves data preprocessing, feature engineering, model selection, training, and evaluation. In the ever-evolving landscape of commerce, accurate future sales prediction is an imperative that transcends industry boundaries. The challenge lies in the intricacies of forecasting, where historical data, consumer behavior, market dynamics, and unforeseen variables converge. Businesses, regardless of size or sector, grapple with the pressing need to predict sales with precision to remain competitive, optimize resources, and foster sustainable growth.

Key Issues:

1. Data Complexity: The wealth of data available today is a double-edged sword. While it offers unprecedented insights, the sheer volume, diversity, and velocity of data pose significant challenges in terms of data collection, cleansing, and analysis. Determining which data points are relevant and how to interpret them accurately is an ongoing hurdle.

2. Changing Consumer Behavior: Consumer preferences are in a constant state of flux, influenced by factors such as economic conditions, societal trends, and technological advancements. Understanding and predicting these shifting behaviors is a complex task, as it requires capturing the essence of human decision-making.

3. Market Volatility: the global marketplace is increasingly interconnected, subject to geopolitical events, economic fluctuations, and unforeseen disruptions like the COVID-19 pandemic. Such external factors can dramatically impact sales patterns, rendering traditional forecasting models less reliable.

4. Ethical Considerations: The collection and use of consumer data for predictive purposes raise ethical concerns, particularly regarding data privacy and consent.

Problem Statement: Develop a model that uses historical sales data to predict future sales for a retail company, enabling them to optimize inventory management and make data-driven business decision.

DESIGN THINKING

1. Data Source: Utilize a dataset containing historical sales data, including features like date, product ID, store ID, and sales quantity.

2. Data Preprocessing: Clean and preprocess the data, handle missing values, and convert categorical features into numerical representations.

3. Feature Engineering: Create additional features that could enhance the predictive power of the model, such as time-based features (e.g., day of the week, month).

4. Model Selection: Choose suitable time series forecasting algorithms (e.g., ARIMA, Exponential Smoothing) for predicting sales.

5. Model Training: Train the selected model using the preprocessed data.

6. Evaluation: Evaluate the model's performance using appropriate time series forecasting metrics (e.g., Mean Absolute Error, Root Mean Squared Error).

OBJECTIVES

Enhance Predictive Accuracy: Develop predictive models that leverage the full spectrum of available data sources, from historical sales data to external market indicators, to increase the accuracy of sales forecasts.

Dynamic Adaptability: Create models that can adapt swiftly to changing consumer behavior, market dynamics, and unforeseen events, ensuring ongoing relevance and reliability.

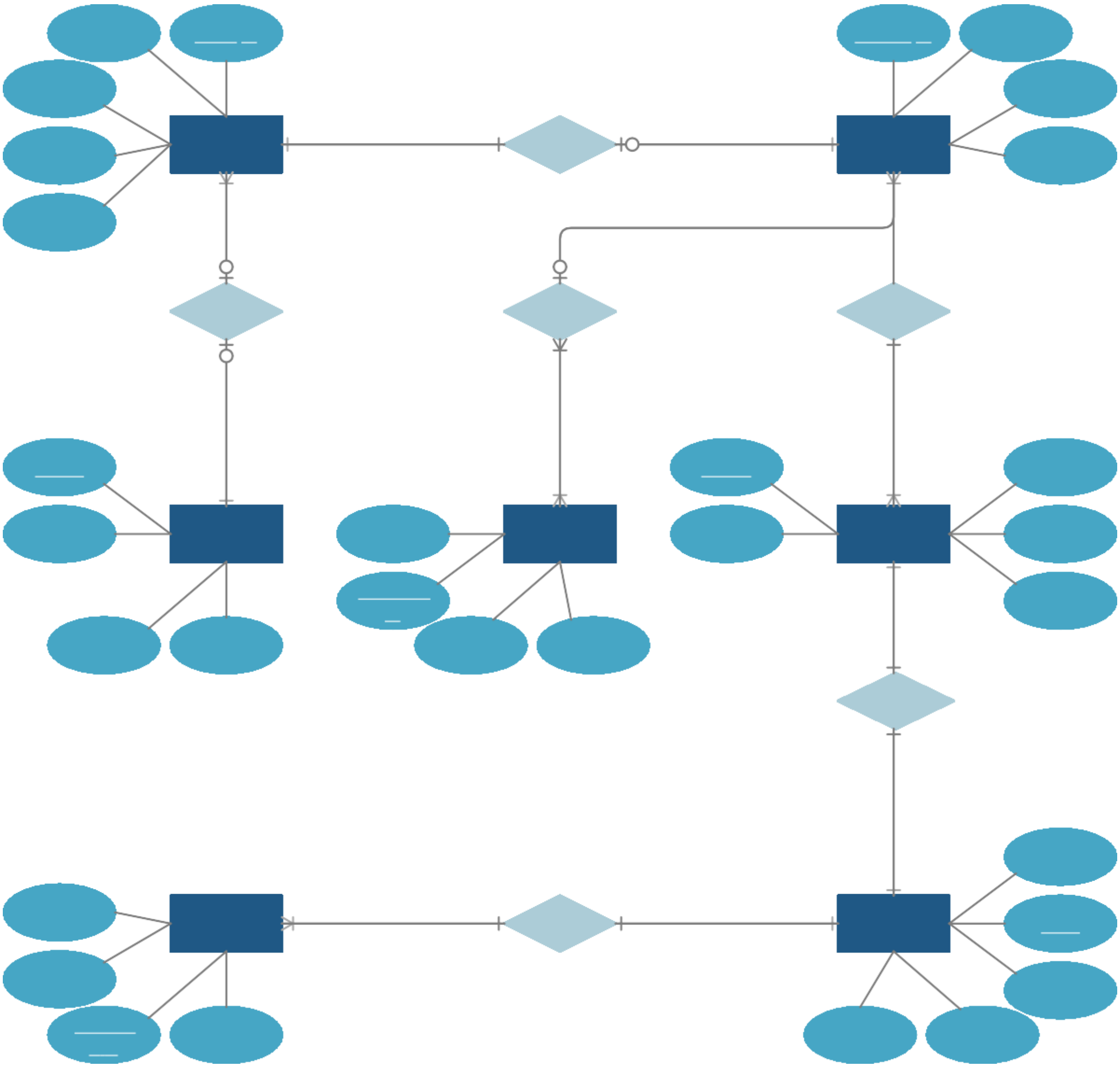
Ethical Framework: Implement robust ethical guidelines and practices for data collection, usage, and transparency, respecting consumers' privacy and consent.

User-Friendly Tools: Develop user-friendly predictive analytics tools that empower businesses of all sizes to harness the power of sales prediction without requiring extensive data science expertise.

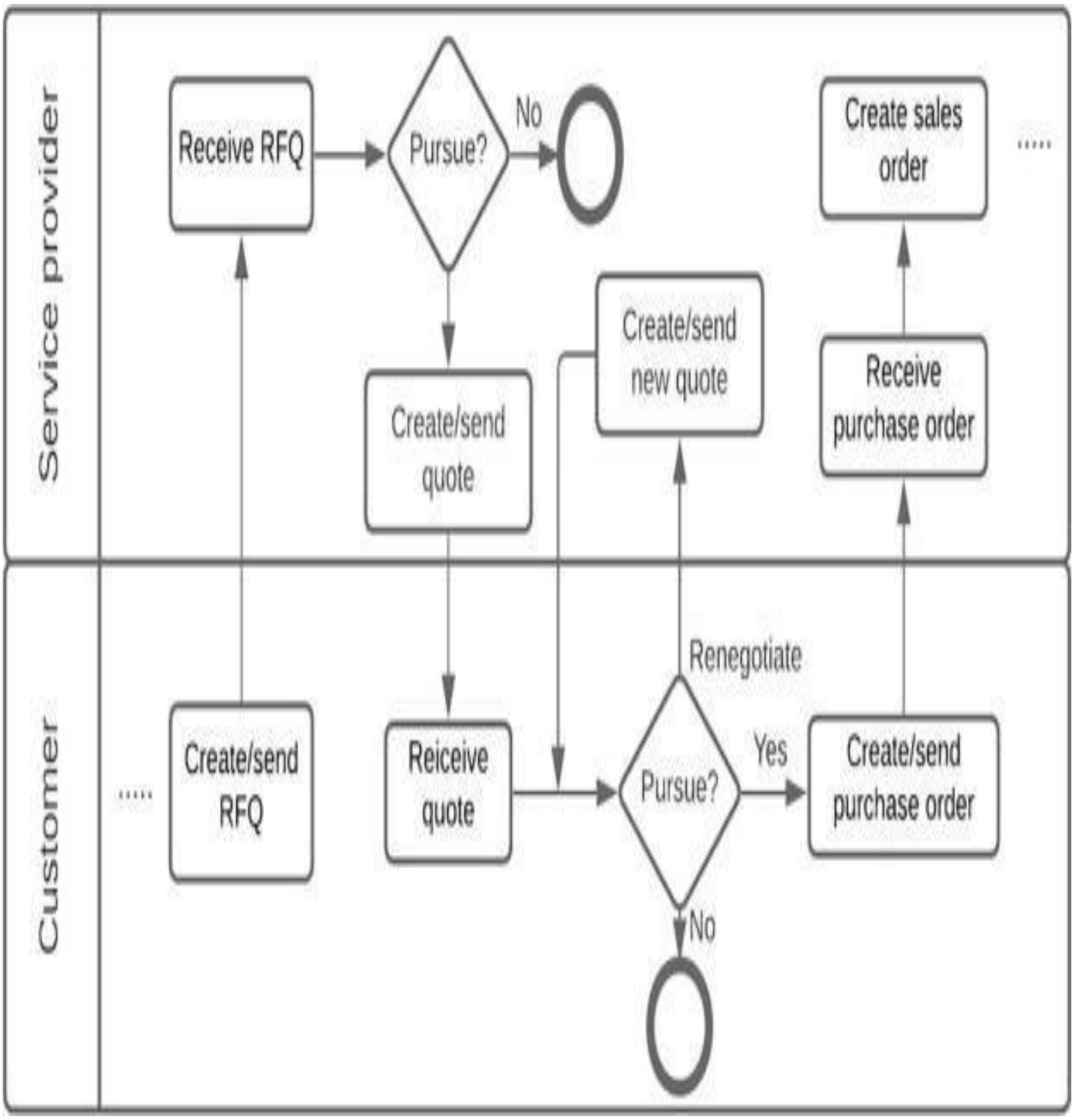
Continuous Learning: Establish a culture of continuous improvement and learning within organizations, encouraging regular model validation and refinement.

The resolution of these challenges is paramount for businesses aiming to thrive in the 21st-century marketplace. Accurate future sales prediction not only bolsters profitability and competitiveness but also contributes to more responsible and ethical business practices, fostering trust and sustainability in the relationship between businesses and their customers. It is a multifaceted challenge, but one with the potential for profound positive impact on the world of commerce. The world of future sales predictions is a dynamic and transformative arena that permeates every corner of modern commerce. It is the nexus where data, technology, and human expertise converge to shape the destinies of businesses and industries alike. From strategic decision-making to resource optimization, customer-centricity, and supply chain management, the applications of sales predictions are vast and far-reaching.

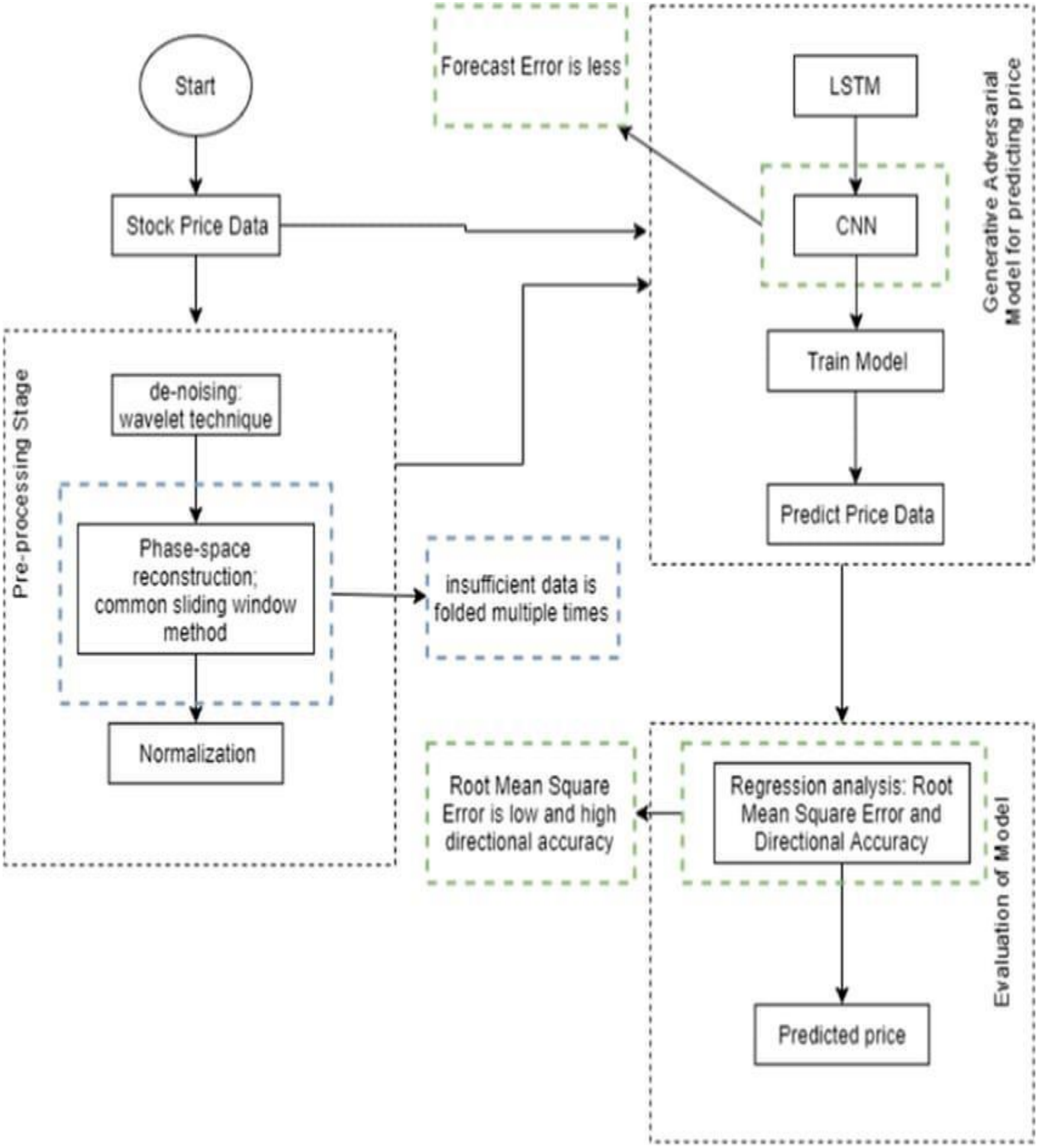
USE CASE DIAGRAM



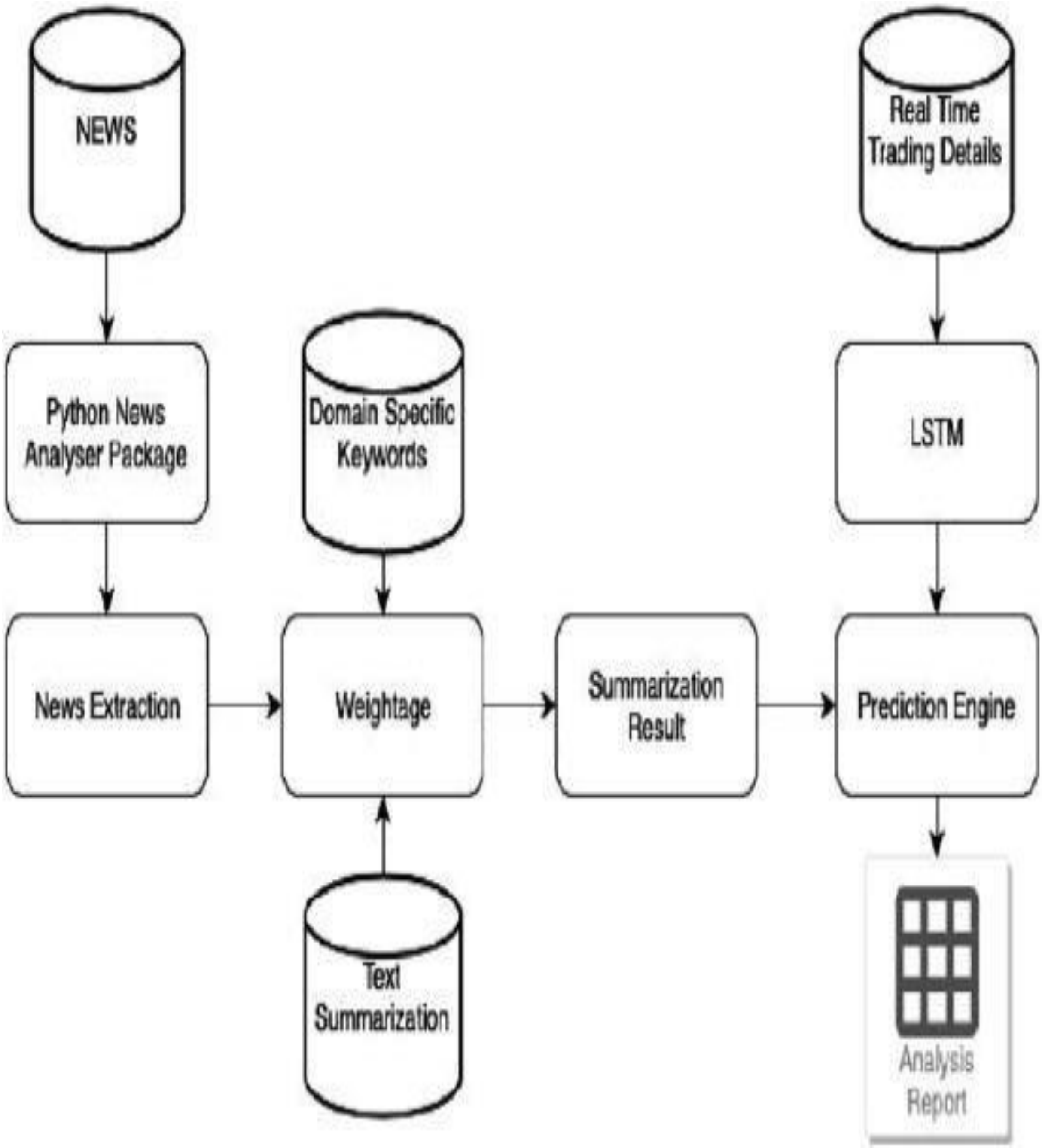
CASE STUDY DIAGRAM



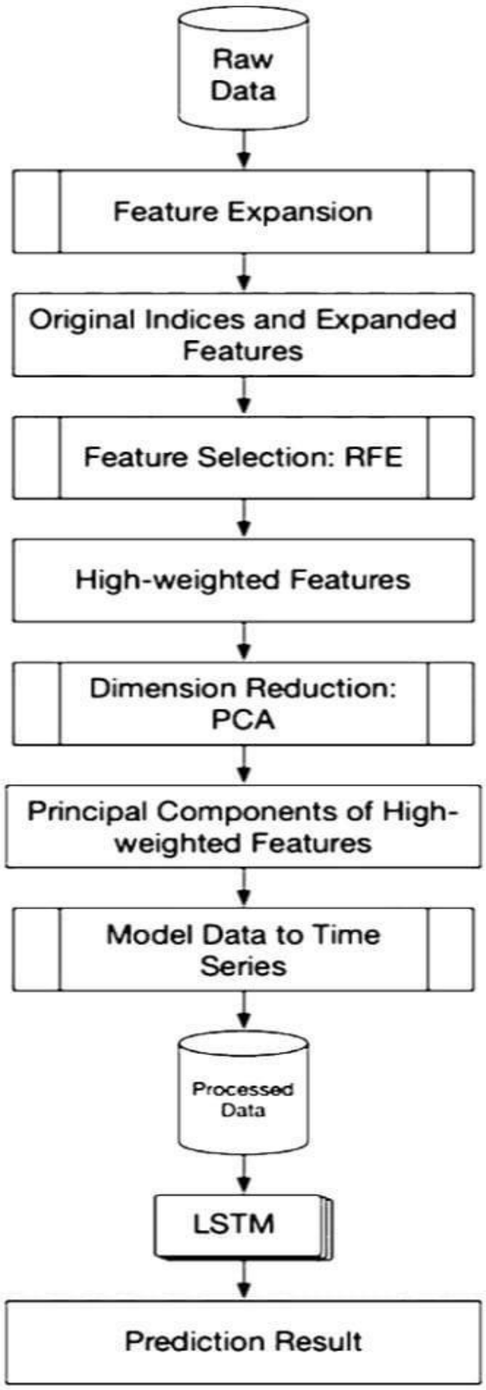
SYSTEM ARCHITECTURE



BLOCK DIAGRAM



SEQUENCE DIAGRAM



CHAPTER-2

INNOVATION

FUTURE SALESPREDICTION

Future sales prediction, also known as sales forecasting, is the process of estimating a company's future sales based on historical data, market trends, and various analytical techniques. Accurate sales predictions are crucial for businesses as they help in making informed decisions regarding inventory management, resource allocation, budget planning, and overall strategy. Here's an explanation of the key aspects of future sales prediction:

- Historical DataAnalysis
- MarketResearch
- QuantitativeMethod

1. Time SeriesAnalysis
2. RegressionAnalysis
3. Machine Learning andAI

- QualitativeFactors
- ScenarioAnalysis
- Technology andSoftware
- Continuous Monitoring and Updating
- FeedbackLoop
- Collaboration

DATASET

This dataset is designed for forecasting future sales in a retail context. It contains historical sales data along with various features that can be used to build predictive models.

When working with a dataset, it's essential to perform data exploration and preprocessing to better understand its characteristics and prepare it for modeling.

We’ve got the data set in the website called Kaggle(www.kaggle.com/data)

The data set which is respective to our project is sales.csv

(**https://www.kaggle.com/datasets/chakradharmattapalli/future-sales-prediction**)

The data set having the 4columns named TV,Radio,Newspaper,Sales and having 200 rows of datas.(numerical values).

The dataset given here contains the data about the sales of the product. The dataset is about the advertising cost incurred by the business on various advertising platforms. Below is the description of all the columns in the dataset:

TV: Advertising cost spent in dollars for advertising on TV;

Radio: Advertising cost spent in dollars for advertising on Radio;

Newspaper: Advertising cost spent in dollars for advertising on Newspaper;

Sales: Number of units sold

So, in the above dataset, the sales of the product depend on the advertisement cost of the product.

DETAILS ABOUTCOLUMNS

To predict future sales, you can create a dataset with columns for TV advertising spending, radio advertising spending, newspaper advertising spending, and sales. These columns will help you build a regression model to predict sales based on advertising expenditures. Here are some details about each of these columns:

TV:

Column Name: TV

Data Type: Numeric (continuous)

Description:

This column represents the amount of money spent on advertising through television channels. It includes expenses for television commercials, sponsorships, and other TV-related advertising efforts. Measured in dollars.

Radio:

Column Name: Radio

Data Type: Numeric

(continuous) Description:

This column represents the amount of money spent on advertising through radio channels. Itincludes expenses for radio commercials, radio show sponsorships, and other radio-related advertising efforts. Measured indollars.

Newspaper:

Column Name: Newspaper

Data Type: Numeric (continuous)

Description:

This column represents the amount of money spent on advertising in newspapers. It includes expenses for print advertisements, classified ads, and other newspaper-related advertising efforts. Measured in dollars.

Sales:

Column Name: Sales

Data Type: Numeric (continuous)

Description:

This is the target column you want to predict. It represents the actual sales revenue generated as a result of the advertising expenditures on TV, radio, and newspaper. Measured in dollars.

TV	Radio	Newspaper	Sales
203.1	37.8	69.2	22.1
44.5	39.3	45.1	10.4
17.2	45.9	69.3	12
151.5	41.3	58.5	16.5
180.8	10.8	58.4	17.9

You can collect historical data for these columns and then use regression techniques (e.g., linear regression) to build a model that predicts future sales based on advertising expenditures. Additionally, you may want to include other relevant features or variables, such as seasonality, market trends, or competitor advertising spending, to improve the accuracy of your sales prediction model.

LIBRARIES

To work with the data and build a predictive model for future sales based on the TV, radio, and newspaper advertising expenditures, you will typically use Python and several libraries.

Here are the key libraries you may need and how to download them: make sure to activate your virtual environment first and then run the pip install commands within that environment.

- Numpy
- Pandas
- Matplotlib andseaborn
- Sci-kit Learn(sklearn)

Import Libraries

```
import numpy as
np import pandas
as pd
import matplotlib.pyplot as
plt import seaborn as sns
from sklearn.model_selection import
train_test_split from sklearn.linear_model import
LinearRegression
```

Load the data

```
df = pd.read_csv('your_dataset.csv')
```

Testing for null values

```
#So ,this dataset doesn't have any null values.
print(data.isnull().sum())
```

Model Building

```
X = df[['TV', 'Radio',
'Newspaper']] y = df['Sales']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42) #for future sales prediction
model = LinearRegression()
model.fit(X_train, y_train)
model.fit(X_train, y_train)
#features = [[TV, Radio, Newspaper]]
features = np.array([[230.1, 37.8,
69.2]]) print(model.predict(features))
```

Model Evaluation

```
y_pred = model.predict(X_test)
mse = mean_squared_error(y_test,
y_pred) r2 = r2_score(y_test, y_pred)
```

TRAIN ANDTEST

Code

```
x = np.array(data.drop(["Sales"],
1)) y = np.array(data["Sales"])
xtrain, xtest, ytrain, ytest = train_test_split(x, y, test_size=0.2, random_state=42)
```

Visualization

```
#relationship between the amount spent on advertising on TV and
units sold import plotly.express as px
import plotly.graph_objects as go
figure = px.scatter(data_frame = data, x="Sales",y="TV", size="TV", trendline="ols")
figure.show()
```

REST OFEXPLANATION

Training and testing data for future sales analysis typically involves using historical sales data to build and evaluate predictive models. These models can help you forecast future sales, identify trends, and make informed business decisions. Here's a step-by-step guide on how to train and test data for future sales analysis. Split your dataset into two parts: a training set and a testing set.

A common split is 70-80% for training and 20-30% for testing. You can also use time-based splitting where the training data comes from earlier time periods, and the testing data comes from more recent periods.

METRICS USED FOR ACCURACYCHECK

Several metrics are commonly used to check the accuracy of predictive models, depending on the specific problem and the nature of the data. Here are some of the most commonly used metrics for accuracy checking:

- Mean Absolute Error(MAE)
- Mean Squared Error(MSE)
- Root Mean Squared Error (RMSE)
- Mean Absolute Percentage Error(MAPE)
- R-squared(R2)
- AdjustedR-squared
- Accuracy, Precision, Recall, F1-Score (ClassificationMetrics)
- Area Under the Receiver Operating Characteristic Curve(AUC-ROC)
- ConfusionMatrix
- Log-Loss (LogarithmicLoss)
- Cohen'sKappa

The choice of metric depends on the problem you're solving and the nature of your data. For regression problems, metrics like MAE, MSE, and RMSE are common, while classification problems typically use accuracy, precision, recall, F1-score, AUC-ROC, and log-loss.

It's essential to select the most appropriate metric based on the specific goals and characteristics of your analysis.

SUMMARY

So this is how we can train a machine learning model to predict the future sales of a product. Predicting the future sales of a product helps a business manage the manufacturing and advertising cost of the product.

In essence, future sales prediction is a dynamic and data-driven process, combining historical insights with forward-looking analysis to empower businesses to make proactive decisions and thrive in an ever-evolving marketplace. Accurate sales predictions enable organizations to streamline operations, allocate resources effectively, and remain agile in the face of market fluctuation

CHAPTER-3

Explanation:

Sales prediction is a process of using data and statistical methods to forecast future sales for a product or service. To create a sales prediction model, you typically need a dataset that includes historical sales data along with relevant features. Here's an explanation of the key elements in a sales prediction dataset:

Date/Time:

This is the timestamp when each sale occurred. It helps in capturing seasonality and trends over time.

Sales Amount:

The target variable, which is the actual sales figure for each time period. This is what you want to predict.

Features:

Product Attributes:

Information about the product being sold, such as category, price, brand, and any special promotions or discounts.

Store Information:

Details about the store where the sale took place, such as location, size, and any specific store-level promotions.

Customer Data:

If available, data on customers, including demographics, loyalty programs, or previous purchase history.

External Factors:

These could include economic indicators (e.g., GDP), weather data, and other external factors that might influencesales.

Lagged Variables:

Past sales data for the same or related products, which can help capture dependencies and seasonality.

Categorical Variables:

Variables that aren't numerical, like product category, store location, or day of the week. These need to be encoded for machine learning models.

Holidays and Special Events:

Information about holidays, special promotions, or events that might impact sales.

Competitor Data:

Data on competitors' pricing or promotions can also be relevant if it's available.

Once you have a dataset with these components, you can use various machine learning and statistical techniques to build a sales prediction model. This might include linear regression, time series analysis, or more advanced techniques like neural networks for deep learning.

The goal is to use this data to train a model that can accurately predict future sales based on the chosen features. Keep in mind that data quality and feature selection are crucial for the accuracy of your predictions.

Details about data:

To predict future sales, you can create a dataset with columns for TV advertising spending, radio advertising spending, newspaper advertising spending, and sales.

These columns will help you build a regression model to predict sales based on advertising expenditures. Here are some details about each of these columns:

TV:

Column Name: TV

Data Type: Numeric (continuous)

Description: This column represents the amount of money spent on advertising through television channels. It includes expenses for television commercials, sponsorships, and other TV-related advertising efforts. Measured indollars.

Radio:

Column Name: Radio

Data Type: Numeric (continuous)

Description: This column represents the amount of money spent on advertising through radio channels. It includes expenses for radio commercials, radio show sponsorships, and other radio-related advertising efforts. Measured in dollars

Newspaper:

Column Name: Newspaper

Data Type: Numeric (continuous)

Description: This column represents the amount of money spent on advertising in newspapers. It includes expenses for print advertisements, classified ads, and other newspaper-related advertising efforts. Measured in dollars.

Sales:

Column Name: Sales

Data Type: Numeric (continuous)

Description: This is the target column you want to predict. It represents the actual sales revenue generated as a result of the advertising expenditures on TV, radio, and newspaper. Measured in dollars

Begin building the project by loading the dataset:

The columns involved in the dataset are:

- Tv
- Radio
- Newspaper
- Sales

Importing necessary libraries:

```
i
m
p
```

loading the dataset:

preprocessing the data:

Data Collection:

Gather historical sales data, including variables like time, sales quantity, price, and any relevant features.

Future selection:

Identify the most relevant features and create new ones if needed.

```
data = data.drop(['newspaper'],axis=1)
```

```
scaler = StandardScaler()
```

train-test-split:

Split your data into a training set and a testing set. Train your selected model using the training data.

```
X
```

Modeling:

```
m
```

Evaluate the Model:

Use evaluation metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), or Root Mean Squared Error (RMSE) to assess the model's performance.

```
y_pred =
```

Make Future Sales Predictions:

Use the trained model to predict future sales based on new data.

```
future_data =
```

performing analysis:

```
#

I
m
```

```
random_state=42)

# Create and train a linear regressionmodel
model = LinearRegression()
model.fit(X_train,y_train)

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

# Evaluate the model
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
print(f"Mean Squared Error: {mse}")
print(f"R-squared: {r2}")

# Plot the predictions
plt.scatter(y_test, y_pred)
plt.xlabel("Actual Sales")
plt.ylabel("Predicted Sales")
plt.title("Actual Sales vs. Predicted Sales")
plt.show()

# Predict future sales (replace the features with your input values)
future_sales = model.predict(np.array([[value1, value2]]))

print(f"Predicted Future Sales: {future_sales[0]}")
```

CHAPTER-4

FUTURE SALES PREDICTION

Future sales prediction, also known as sales forecasting, is the process of estimating a company's future sales based on historical data, market trends, and various analytical techniques. Accurate sales predictions are crucial for businesses as they help in making informed decisions regarding inventory management, resource allocation, budget planning, and overall strategy.

IMPLEMENTATION:

- Load and preprocess your sales data.
- Split the data into features and the target variable.
- Split the data into training and testing sets for model evaluation.
- Create a linear regression model and train it on the training data.
- Use the model to make predictions on the test set and evaluate its performance.
- Visualize the results if needed.
- Finally, you can use the trained model to make predictions for future sales based on your future feature values.

In addition to these steps, it's essential to keep the business goals in mind when implementing the future sales prediction. Each business is unique, and the prediction strategy should align with the specific objectives.

DATASET AND ITS IMPLEMENTATION

We've got the data set in the website called Kaggle(www.kaggle.com/data)

The data set which is respective to our project is sales.csv

(**https://www.kaggle.com/datasets/chakradharmattapalli/future-sales-prediction**)

The data set having the4columns named TV,Radio,Newspaper,Sales and having 200 rows of datas.(numerical values).

BEGIN THE PROJECT BY LOADING THE DATASET

THE SAMPLE DATA SET:

TV	Radio	Newspaper	Sales
203.1	37.8	69.2	22.1
44.5	39.3	45.1	10.4
17.2	45.9	69.3	12
151.5	41.3	58.5	16.5
180.8	10.8	58.4	17.9

1.IMPORT NECESSARY LIBRARIES

Here are the key libraries you may need and how to download them: make sure to activate your virtual environment first and then run the pip install commands within that environment.

- Numpy
- Pandas
- Matplotlib and seaborn

PROGRAM

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

#load the dataset
data = pd.read_csv('C:/priya/Sales.csv')
print(data.head())
data['Sales'].plot()
plt.show()
data.plot(subplots=True, figsize=(4, 4))
plt.show()
print(data.describe())
print(data.isnull().sum())

# Create a scatter plot for tv vs sales
plt.figure(figsize=(8, 6))
plt.scatter(data['TV'], data['Sales'], c='b', marker='o', label='TV vs.
Sales')
plt.title('Scatter Plot of TV Advertising vs. Sales')
plt.xlabel('TV Advertising Budget')
plt.ylabel('Sales')
plt.legend()
plt.grid(True)
plt.show()

# Create a scatter plot for radio vs sales
plt.figure(figsize=(8, 6))
plt.scatter(data['Radio'], data['Sales'], c='r', marker='o', label='Radio
```

```
vs. Sales')
plt.title('Scatter Plot of Radio Advertising vs. Sales')
plt.xlabel('Radio Advertising Budget')
plt.ylabel('Sales')
plt.legend()
plt.grid(True)
plt.show()


# Create a scatter plot for newspaper vs sales
plt.figure(figsize=(8, 6))
plt.scatter(data['Newspaper'], data['Sales'], c='g', marker='o',
label='Newspaper vs. Sales')
plt.title('Scatter Plot of Newspaper Advertising vs. Sales')
plt.xlabel('Newspaper Advertising Budget')
plt.ylabel('Sales')
plt.legend()
plt.grid(True)
plt.show()

#regression
# Sample data
TV = [230.1, 44.5, 17.2, 151.5, 180.8]
Radio = [17.8, 39.3, 45.9, 41.3, 10.8]
Newspaper = [69.2, 45.1, 69.3, 58.5, 58.4]
Sales = [22.1, 10.4, 12, 16.5, 17.9]

# Calculate the mean of each feature and the target
mean_TV = sum(TV) / len(TV)
mean_Radio = sum(Radio) / len(Radio)
mean_Newspaper = sum(Newspaper) / len(Newspaper)
mean_Sales = sum(Sales) / len(Sales)

# Calculate the coefficients
numerator = 0
denominator = 0
for i in range(len(TV)):
```

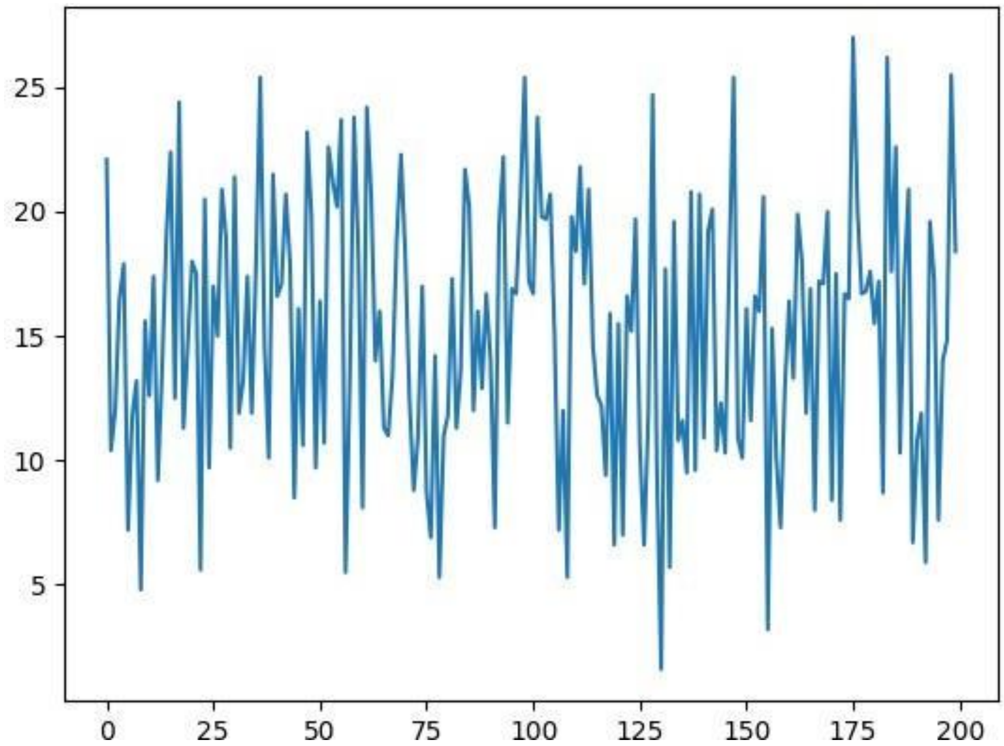
```
    numerator += (TV[i] - mean_TV) * (Sales[i] - mean_Sales)
    denominator += (TV[i] - mean_TV) ** 2

slope = numerator / denominator
intercept = mean_Sales - slope * mean_TV
# Now, you can make predictions
new_TV = 250 # Input a new value for TV
predicted_sales = intercept + slope * new_TV
print(f"Slope (Coefficient): {slope}")
print(f"Intercept: {intercept}")
print(f"Predicted Sales for TV = {new_TV}: {predicted_sales}")

# Create an area plot
data[['TV', 'Radio', 'Newspaper', 'Sales']].plot.area(stacked=True)

# Add labels and a title
plt.xlabel('Data Points')
plt.ylabel('Values')
plt.title('Area Plot for TV, Radio, Newspaper, and Sales')
# Show the plot
plt.show()

# Extract the data for the columns you want to plot
tv_data = data['TV']
radio_data = data['Radio']
newspaper_data = data['Newspaper']
sales_data = data['Sales']

# Create a bar chart
plt.bar(['TV', 'Radio', 'Newspaper', 'Sales'], [tv_data.mean(),
radio_data.mean(), newspaper_data.mean(), sales_data.mean()])
plt.xlabel('Advertising Medium')
plt.ylabel('Mean Value')
plt.title('Mean Values for TV, Radio, Newspaper, and Sales')
plt.show()
```

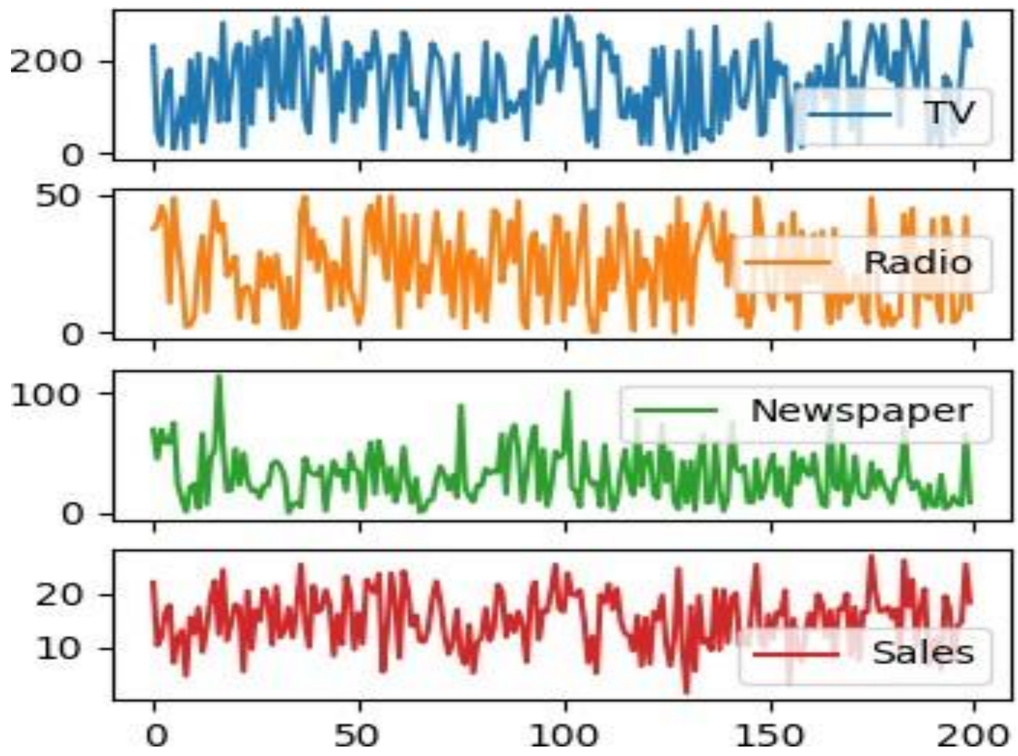
```
#LogisticRegression
import pandas as pd
import numpy as np
# Load your dataset
data = pd.read_csv('your_dataset.csv')

# Define your features and target variable
X = data[['tv', 'radio', 'newspaper']]
y = data['Target'] # Replace 'Target' with the actual column name
containing your classes

# Add a bias term (intercept) to the features
X['bias'] = 1

# Define the logistic function
def sigmoid(z):
    return 1 / (1 + np.exp(-z))

# Initialize model parameters (weights)
theta = np.zeros(X.shape[1])

# Define the learning rate and number of iterations
learning_rate = 0.01
iterations = 1000

# Gradient Descent to update model parameters

for _ in range(iterations):
    z = np.dot(X, theta)
    h = sigmoid(z)
    gradient = np.dot(X.T, (h - y)) / len(y)
    theta -= learning_rate * gradient

# Make predictions
predicted_probabilities = sigmoid(np.dot(X, theta))
predictions = [1 if p >= 0.5 else 0 for p in predicted_probabilities]
```

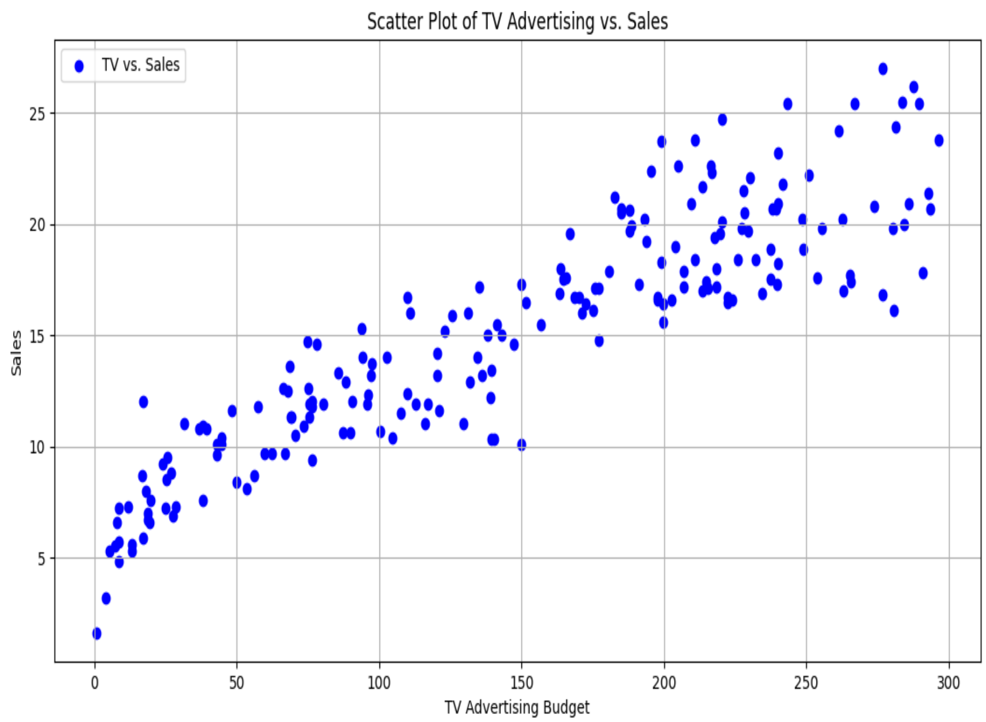
```
# Evaluate the model
accuracy = np.mean(predictions == y)
print("Accuracy:", accuracy)

#Import necessary libraries:
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.linear_model import LogisticRegression
fromsklearn.metrics import accuracy_score, jaccard_score,
log_loss, f1_score
X_train, X_test, y_train, y_test = train_test_split(TV, Radio,
test_size=0.2, random_state=42)
#Create a Logistic Regression model:
model = LogisticRegression()
#Define a range of C values to search over:
param_grid = {'C': [0.001, 0.01, 0.1, 1, 10, 100]}
#Use GridSearchCV to find the best C parameter:
grid_search = GridSearchCV(model, param_grid, cv=5)
grid_search.fit(X_train, y_train)
best_C = grid_search.best_params_['C']
#Train the Logistic Regression model with the best C value:
best_model = LogisticRegression(C=best_C)
best_model.fit(X_train, y_train)
#Make predictions on the test data:
y_pred = best_model.predict(X_test)
#Calculate accuracy on the test data:
accuracy = accuracy_score(y_test, y_pred)
#Calculate the Jaccard Index on the test data:
jaccard = jaccard_score(y_test, y_pred)
# Replace y_pred with predicted probabilities if available
logloss = log_loss(y_test, y_pred)
f1 = f1_score(y_test, y_pred)
#Return the results:
return accuracy, jaccard, logloss, f1
```

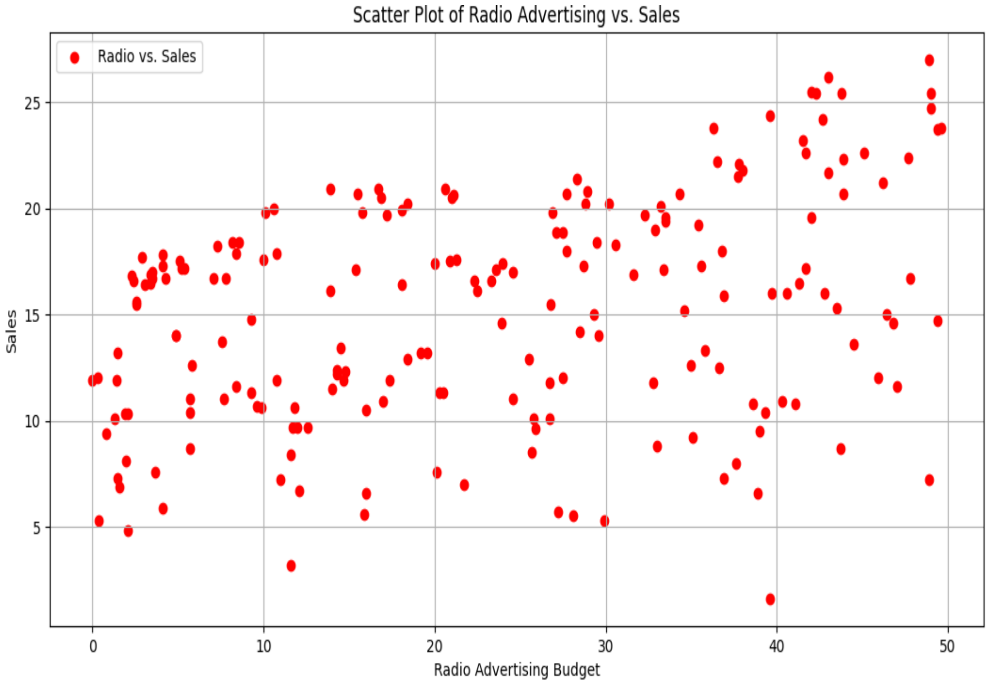
OUTPUT

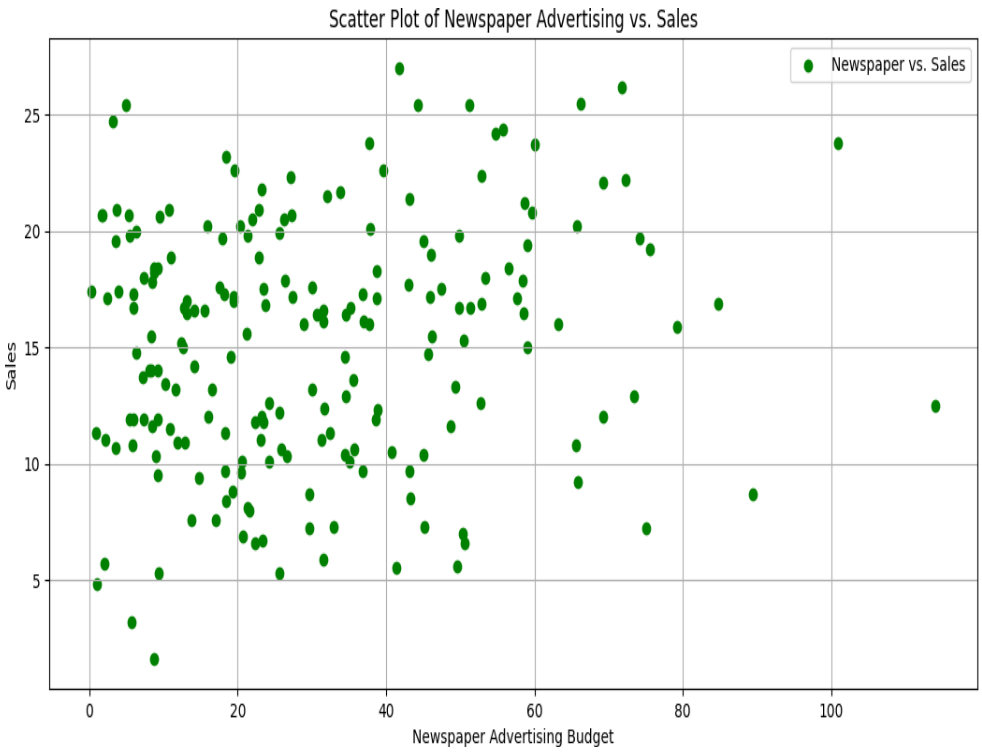
```
C:\Users\priya\PycharmProjects\pythonProjectlab\venv\Scripts\python.exe C:\Users\priya\PycharmProjects\pythonProjectlab\venv\ads1.py
      TV  Radio  Newspaper  Sales
0  230.1   37.8       69.2   22.1
1   44.5   39.3       45.1   10.4
2   17.2   45.9       69.3   12.0
3  151.5   41.3       58.5   16.5
4  180.8   10.8       58.4   17.9
               TV       Radio   Newspaper       Sales
count  200.000000  200.000000  200.000000  200.000000
mean   147.042500   23.264000   30.554000   15.130500
std     85.854236   14.846809   21.778621    5.283892
min      0.700000    0.000000    0.300000    1.600000
25%     74.375000    9.975000   12.750000   11.000000
50%    149.750000   22.900000   25.750000   16.000000
75%    218.825000   36.525000   45.100000   19.050000
max    296.400000   49.600000  114.000000   27.000000
TV           0
Radio        0
Newspaper    0
Sales        0
dtype: int64
Slope (Coefficient): 0.04981920847325357
Intercept: 9.561566398368491
Predicted Sales for TV = 250: 22.016368516681887
```

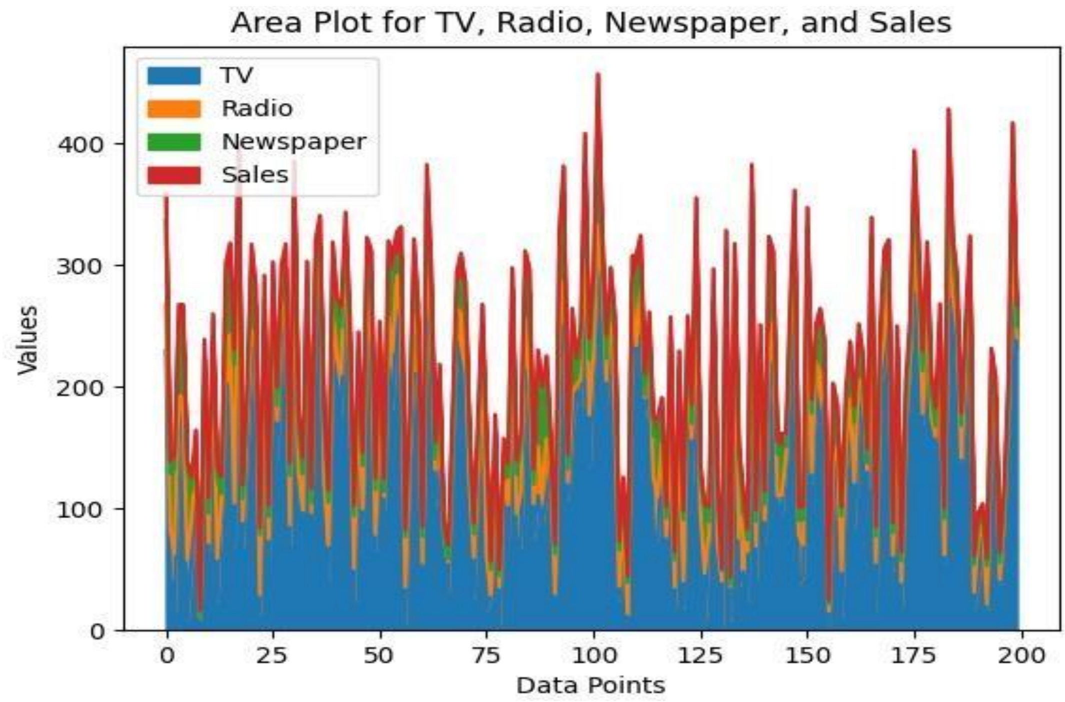


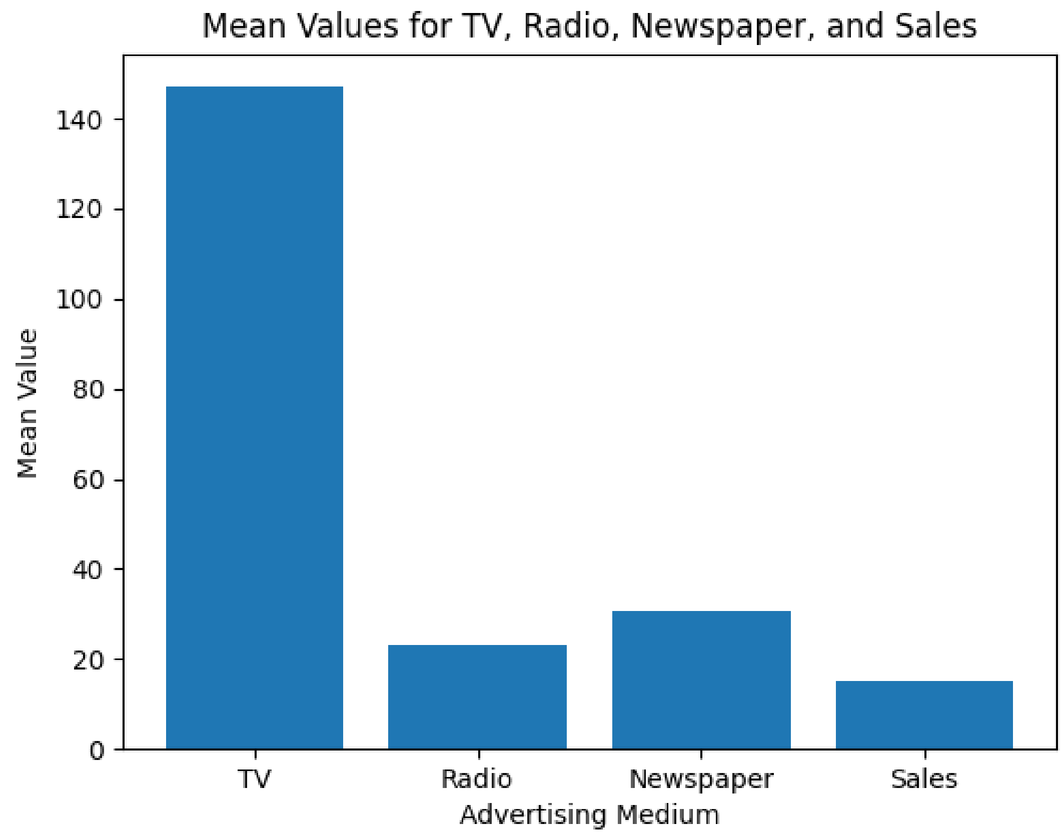

TV
Radio
Newspaper
Sales
0
50
100
150
200


Scatter Plot of TV Advertising vs. Sales
TV vs. Sales
Sales
TV Advertising Budget


Scatter Plot of Radio Advertising vs. Sales
Radio vs. Sales
Sales
Radio Advertising Budget
0
10
20
30
40
50
5
10
15
20
25


Scatter Plot of Newspaper Advertising vs. Sales
Newspaper vs. Sales
Sales
Newspaper Advertising Budget
0
20
40
60
80
100
5
10
15
20
25


Area Plot for TV, Radio, Newspaper, and Sales
TV
Radio
Newspaper
Sales
Values
Data Points
0
100
200
300
400
0
25
50
75
100
125
150
175
200


Mean Values for TV, Radio, Newspaper, and Sales
Mean Value
Advertising Medium
0
20
40
60
80
100
120
140
TV
Radio
Newspaper
Sales