

**A Micro Project report on**  
**Marketing Analytics Platform Using Predictive Models.**

Submitted to the **CMR Institute of Technology** in partial fulfillment of the  
requirement of the

**AUTOMATED WRITING TOOLS - ChatGPT**

Of

**III-B.Tech. I-Semester**

In

**Computer Science Department**

Submitted by

**AKASH YAMAGANI      22R01A0562**

**PRANAY YASA      22R01A0563**

**AKANKSHA      22R01A0565**

**SHRAVANI AKITI      22R01A0566**

Under the Guidance of

**MR. KUNAL GAIKWAD**

(ASSISTANT PROFESSOR)



**CMR INSTITUTE OF TECHNOLOGY**  
(UGC AUTONOMOUS)  
(Approved by AICTE and Affiliated to JNTU,  
**Kukatpally**  
**Kandlakoya, Medchal Road**

# CMR INSTITUTE OF TECHNOLOGY

(UGC-AUTONOMOUS)

(Approved by AICTE, Affiliated to JNTU,  
Hyderabad) Kandlakoya, Medchal Road,  
Hyderabad.

**Department of Computer Science and Engineering**



## CERTIFICATE

This is to certify that a Micro Project entitled with “**Marketing Analytics Platform Using Predictive Models**” is the bonafide work carried out and submitted by

**AKASH YAMAGANI**      **22R01A0562**

**PRANAY YASA**      **22R01A0563**

**AKANKSHA**      **22R01A0565**

**SHRAVANI AKITI**      **22R01A0566**

In partial fulfillment of the requirement for completion of the **AUTOMATED WRITING TOOLS – ChatGPT** lab of III– B.Tech I – Semester is a record of Bonafide work carried out under our guidance during the academic year 2024-2025.

PROJECT GUIDE

**Mr. Kunal Gaikwad**  
Assistant Professor

HEAD OF THE DEPARTMENT

**Mr. A. Prakash**  
Head of Department

# ACKNOWLEDGEMENT

We are extremely grateful to **Dr. M. Janga Reddy , Director , Dr. B. Satyanarayana ,Principal and Mr. A. Prakash , Head of Department ,** Dept of Computer Science and Engineering , CMR Institute of Technology for their inspiration and valuable guidance during entire duration.

We are extremely thankful to our **AUTOMATED WRITING TOOLS - ChatGPT** faculty in-charge **Mr. Kunal Gaikwad , Assistant Professor ,** Computer Science and Engineering department , CMR Institute of Technology for their constant guidance , encouragement and moral support throughout the project.

We express our thanks to all staff members and friends for all the help and coordination extended in bringing out this Project successfully in time.

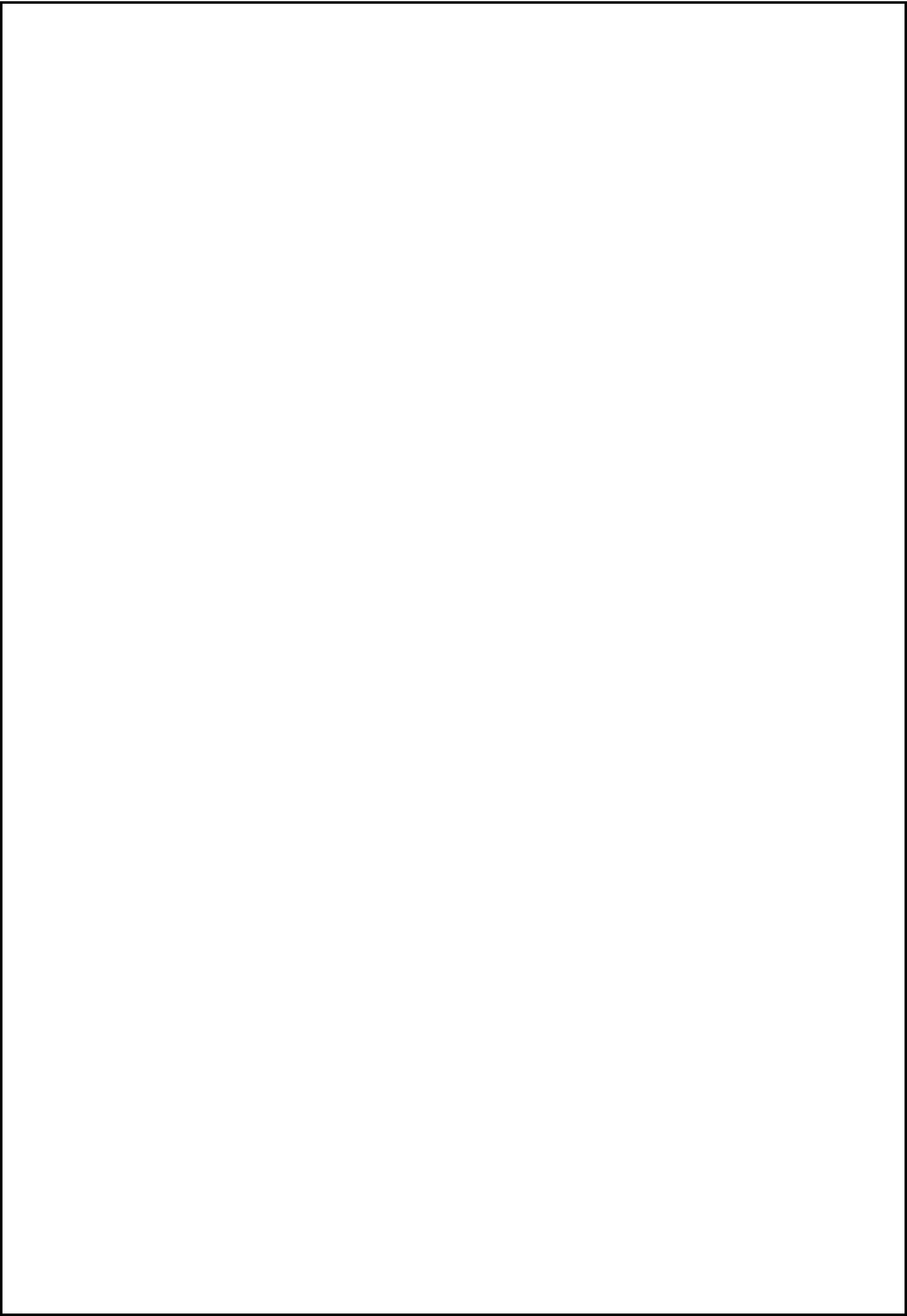
Finally, we are very much thankful to our parents and relatives who guided directly or indirectly for successful completion of the project.

**AKASH YAMAGANI                      22R01A0562**

**PRANAY YASA                        22R01A0563**

**AKANKSHA                            22R01A0565**

**SHRAVANI AKITI                    22R01A0566**



# INTRODUCTION

Predictive analytics has become a cornerstone in modern data-driven decision-making, especially within the marketing sector. It involves the use of **historical data** and **statistical algorithms** to forecast future trends and behaviours, thereby providing valuable insights for businesses. Through predictive models, organizations can enhance decision-making, optimize marketing strategies, and drive operational efficiencies. This methodology is particularly beneficial for predicting **customer behaviour**, **sales performance**, and **market trends**, enabling businesses to proactively adjust their strategies and allocate resources more effectively.

In marketing, predictive analytics plays a critical role in understanding how various factors, such as **promotional campaigns**, **product pricing**, and **market condition** impact sales performance. By applying the appropriate predictive models, companies can identify which marketing strategies are most likely to generate positive sales outcomes and customer engagement. This capability gives businesses a competitive advantage in a dynamic market environment.

The primary objective of this project is to develop a Marketing Analytics Platform that employs predictive models to forecast sales outcomes based on multiple features, such as promotion type, market size, and sales campaigns. Key goals of the platform include:

- Predicting future sales trends and values
- Identifying significant factors that influence sales performance
- Providing actionable insights to refine marketing strategies

By utilizing machine learning algorithms, this platform aims to deliver more accurate sales forecasts and help businesses make informed, data-backed decisions. The following sections will provide a detailed walkthrough of the steps involved in building and evaluating the predictive model, as well as the insights generated from the dataset.

# DATA OVERVIEW

The dataset titled `WA_Marketing-Campaign.csv` provides a comprehensive overview of sales data influenced by various marketing promotions across different markets. It captures crucial information regarding sales performance and market characteristics, which are essential for understanding the impact of marketing strategies. Below is a detailed description of the dataset, its structure, and the key features it contains.

## Description of the Dataset

This dataset captures detailed sales performance data that is segmented by market characteristics and promotional strategies. It encompasses the following key components:

- **SalesInThousands:** This is the target variable, representing the sales figures in thousands of dollars for each market in a given week. It serves as the dependent variable in predictive modeling, reflecting the outcome of marketing efforts.
- **MarketSize:** A categorical variable that indicates the size classification of the market. The possible values include Small, Medium, and Large markets. The size of the market may influence the overall sales performance, as larger markets could have more customers and demand.
- **LocationID:** A unique numeric identifier for each market location. This identifier ensures that each market's sales data can be tracked and analyzed individually. It provides insight into location-specific sales patterns.
- **AgeOfStore:** This numerical variable represents the age of the store in years. It gives insight into the establishment's longevity and may reflect the store's brand recognition or customer loyalty, both of which can influence sales outcomes.
- **Promotion:** A categorical variable indicating different types of promotions applied during the sales period. The variable is encoded numerically, with different integers corresponding to different types of promotions (e.g., discounts, limited-time offers, etc.). Understanding the effect of promotions on sales is crucial for businesses to optimize marketing strategies.
- **Week:** A numeric representation indicating the week number in which the sales were recorded. This allows for time-series analysis, enabling the identification of trends or patterns in sales across different weeks.

## Structure of the Dataset

The dataset is structured in a tabular format with multiple rows and columns. Each row represents a specific record of sales data for a given market during a particular week. The structure of the dataset is as follows:

### - Columns:

- MarketID: Integer – A unique identifier for each market entry.
- MarketSize: Categorical (String) – Classification of the market
- LocationID: Integer – Identifier for specific market locations.
- AgeOfStore: Integer – Number of years since establishment of the store.
- Promotion: Integer – Categorical representation of the promotion type applied.
- Week: Integer – Week number corresponding to the sales data.
- SalesInThousands: Float – Sales figures in thousands.

- Rows: Each row corresponds to a sales record for a particular market and week. This allows for comprehensive analysis over time, as data is segmented by week and market.

## Data Types

The data types for each column are as follows:

- MarketID: Integer
- MarketSize: Categorical (String)
- LocationID: Integer
- AgeOfStore: Integer
- Promotion: Integer
- Week: Integer
- SalesInThousands: Float

The mix of numerical and categorical data types enables various types of analyses, including regression, time-series forecasting, and classification.

## **Key Features Overview**

1. **Promotion Type:** The Promotion feature represents different promotional strategies employed during sales periods. These strategies may include discounts, special offers, or seasonal promotions. It is crucial to understand how each promotion type impacts sales performance.
2. **Market Size:** The MarketSize feature categorizes markets into Small, Medium, and Large. Market size plays a key role in determining the potential customer base and the overall sales volume. Larger markets may exhibit different sales trends compared to smaller markets.
3. **Weekly Sales:** SalesInThousands records the sales performance for each market, measured on a weekly basis. This feature is critical for time-series analysis and for identifying sales trends across different weeks, promotions, and market conditions.
4. **Store Age:** The AgeOfStore feature provides insight into how long the store has been operational. Older stores might have more loyal customers and established brand recognition, potentially leading to more stable sales patterns. Conversely, newer stores might show higher growth rates as they establish their customer base.

## **Significance of the Dataset**

This structured dataset serves as a valuable resource for businesses and analysts seeking to understand the relationship between various marketing strategies and sales outcomes. Through comprehensive analysis of these key features, organizations can identify:

- The effectiveness of different promotional campaigns in driving sales.
- How market size influences overall sales performance.
- How the age of a store affects its sales and customer retention.

By analyzing these features in combination, businesses can derive actionable insights to refine their marketing strategies, forecast future sales, and optimize promotional efforts across various markets.




# PROMPT

In this section we will see the prompt that we have used to create Marketing Analytics Platform Using Predictive Models and the result that ChatGpt provided us.

## Result of Prompt 1:

if i give a pdf what Predictive Models you can do



If you provide a PDF with data or descriptions, I can help you analyze the data and suggest or build predictive models based on the information within it. Here are some types of predictive models we could explore:

- Customer Segmentation:**
  - Model:** Clustering (e.g., K-means, hierarchical clustering).
  - Purpose:** Group customers into segments based on similar behaviors or characteristics. This is useful for targeted marketing.
- Customer Churn Prediction:**
  - Model:** Classification (e.g., logistic regression, decision trees, random forests).
  - Purpose:** Predict which customers are likely to stop using a product or service, allowing for targeted retention strategies.
- Sales Forecasting:**
  - Model:** Time Series Analysis (e.g., ARIMA, exponential smoothing, LSTM for deep learning).
  - Purpose:** Predict future sales trends to inform inventory planning, marketing budgets, and staffing.

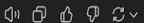
ChatGPT

↑ Share

staffing.

- Customer Lifetime Value (CLV):**
  - Model:** Regression or predictive lifetime value modeling.
  - Purpose:** Estimate the total revenue expected from a customer over the lifetime of their relationship with the company, which can help prioritize high-value customers.
- Product Recommendation:**
  - Model:** Collaborative Filtering or Content-Based Filtering.
  - Purpose:** Suggest products to customers based on their past behavior or similar customers' behavior.
- Campaign Response Prediction:**
  - Model:** Classification (e.g., logistic regression, random forest, support vector machines).
  - Purpose:** Predict the likelihood that a customer will respond to a particular campaign, allowing for better-targeted marketing efforts.

If you have a specific type of data or a business objective, I can suggest a tailored predictive model to meet that goal. Just upload the PDF, and I'll help you decide on the best approach!




## Result of Prompt 2:

ChatGPT ▾

WA\_Marketing-Campaign.csv  
Spreadsheet

this is a csv file , what Predictive Models u you can do , please do

 Let's first examine the contents of your CSV file to see what kinds of data are present. This will help determine suitable predictive models.

I'll load and analyze the file structure now.

The dataset contains the following columns:

1. **MarketID**: Identifier for the market.
2. **MarketSize**: Size of the market (e.g., Small, Medium, Large).
3. **LocationID**: Identifier for the location.
4. **AgeOfStore**: Age of the store in years.
5. **Promotion**: A categorical variable indicating different promotional types.
6. **week**: Week number in a campaign or sales period.
7. **SalesInThousands**: Sales amount in thousands.

↓

ChatGPT ▾

**1. Sales Forecasting**

- **Model**: Time Series Analysis (e.g., ARIMA, exponential smoothing, or seasonal models).
- **Purpose**: Predict future sales based on weekly sales data. This could help in planning inventory and managing resources.

**2. Sales Impact Analysis by Promotion Type**

- **Model**: Regression or Classification (e.g., Linear Regression, Random Forest).
- **Purpose**: Determine how different promotions (Promotion column) impact sales, which can guide decision-making about effective promotions.

**3. Market Size and Sales Relationship**


- **Model**: Classification (e.g., Decision Trees, Random Forest) or Clustering (e.g., K-means if we explore segmenting markets).
- **Purpose**: Identify how market size affects sales, possibly segmenting markets into high, medium, and low sales clusters.

**4. Store Age vs. Sales Trends**

- **Model**: Regression Analysis.
- **Purpose**: Explore if older stores have different sales trends than newer ones, which could help in

↓

ChatGPT ▾



Certainly! I'll provide you with a Google Colab-compatible Python script that you can use to load this dataset and perform both **Sales Forecasting** and **Promotion Effect on Sales** analyses. This script will include data exploration, visualization, and the initial predictive models we discussed.

Here's the Colab code:

python

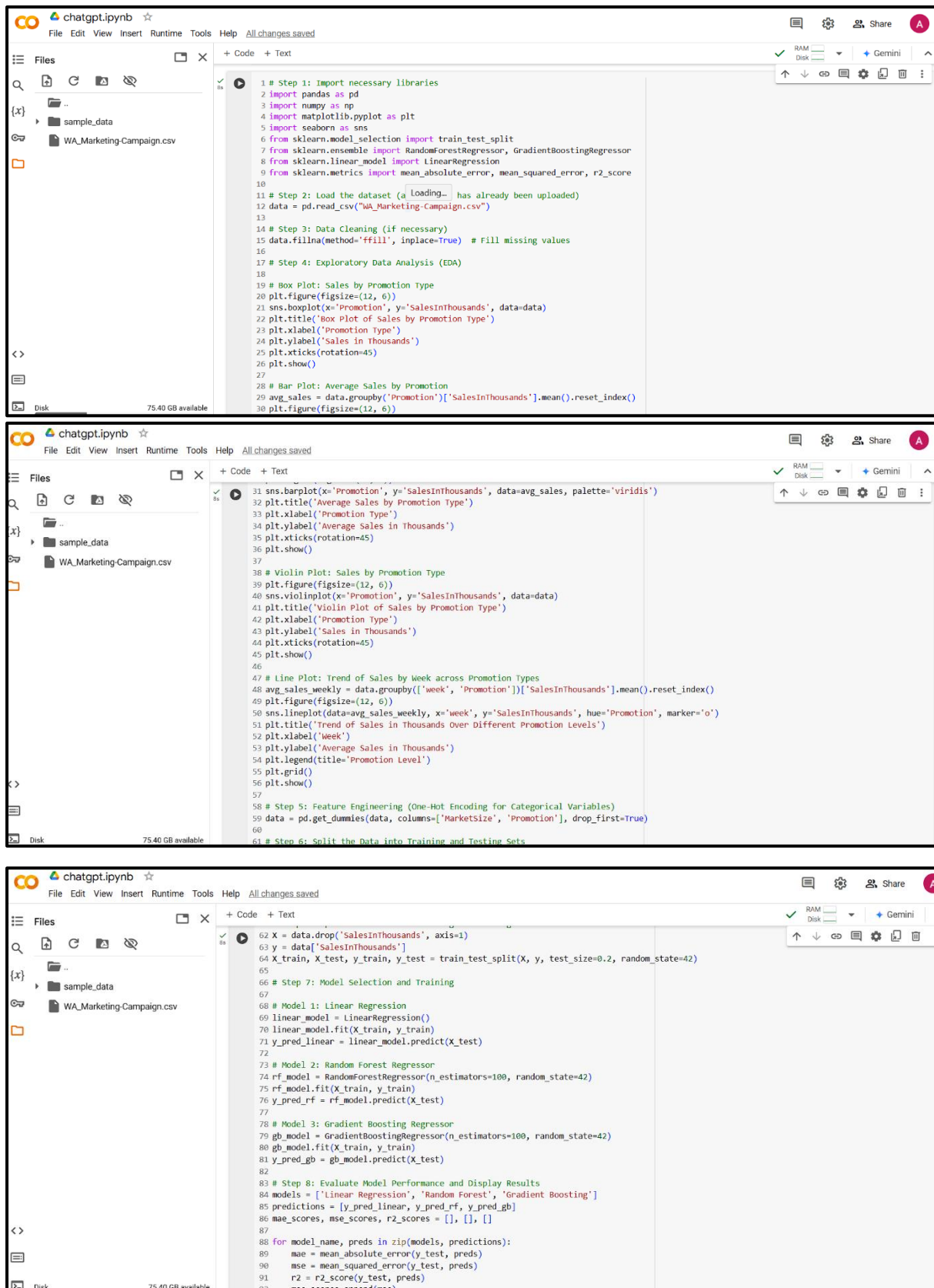
Copy code

```
# First, import necessary libraries
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.tsa.arima.model import ARIMA
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
import seaborn as sns

# Load the dataset (make sure to upload your CSV in Colab first)
data = pd.read_csv('WA_Marketing-Campaign.csv')

# Explore the dataset
print(data.info())
print(data.head())
```

# Here is Google Colab Code generated By ChatGpt:



The image displays three sequential screenshots of a Google Colab notebook titled 'chatgpt.ipynb'. The notebook interface includes a file explorer on the left showing 'sample\_data' and 'WA\_Marketing-Campaign.csv', a top menu bar, and a right sidebar with RAM and Disk usage indicators. The code is organized into steps:

- Step 1: Import necessary libraries** (lines 1-10): Imports pandas, numpy, matplotlib, seaborn, and sklearn modules.
- Step 2: Load the dataset** (lines 11-13): Loads 'WA\_Marketing-Campaign.csv' into a DataFrame.
- Step 3: Data Cleaning** (lines 14-15): Uses `data.fillna(method='ffill', inplace=True)` to handle missing values.
- Step 4: Exploratory Data Analysis (EDA)** (lines 16-30):
  - Box Plot: `sns.boxplot(x='Promotion', y='SalesInThousands', data=data)`
  - Bar Plot: `avg_sales = data.groupby('Promotion')['SalesInThousands'].mean().reset_index()`
- Step 5: Feature Engineering** (lines 31-60):
  - Bar Plot: `sns.barplot(x='Promotion', y='SalesInThousands', data=avg_sales, palette='viridis')`
  - Violin Plot: `sns.violinplot(x='Promotion', y='SalesInThousands', data=data)`
  - Line Plot: `avg_sales_weekly = data.groupby(['week', 'Promotion'])['SalesInThousands'].mean().reset_index()`
  - Feature Engineering: `data = pd.get_dummies(data, columns=['MarketSize', 'Promotion'], drop_first=True)`
- Step 6: Split the Data into Training and Testing Sets** (lines 61-65):
  - `X = data.drop('SalesInThousands', axis=1)`
  - `y = data['SalesInThousands']`
  - `X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)`
- Step 7: Model Selection and Training** (lines 66-82):
  - Model 1: Linear Regression (`linear_model = LinearRegression()`)
  - Model 2: Random Forest Regressor (`rf_model = RandomForestRegressor(n_estimators=100, random_state=42)`)
  - Model 3: Gradient Boosting Regressor (`gb_model = GradientBoostingRegressor(n_estimators=100, random_state=42)`)
- Step 8: Evaluate Model Performance and Display Results** (lines 83-92):
  - Models list: `models = ['Linear Regression', 'Random Forest', 'Gradient Boosting']`
  - Predictions: `predictions = [y_pred_linear, y_pred_rf, y_pred_gb]`
  - Performance metrics: `mae_scores, mse_scores, r2_scores = [], [], []`
  - Loop to calculate and append metrics for each model.

```
chatgpt.ipynb
File Edit View Insert Runtime Tools Help All changes saved

Files
{ }
sample_data
WA_Marketing-Campaign.csv

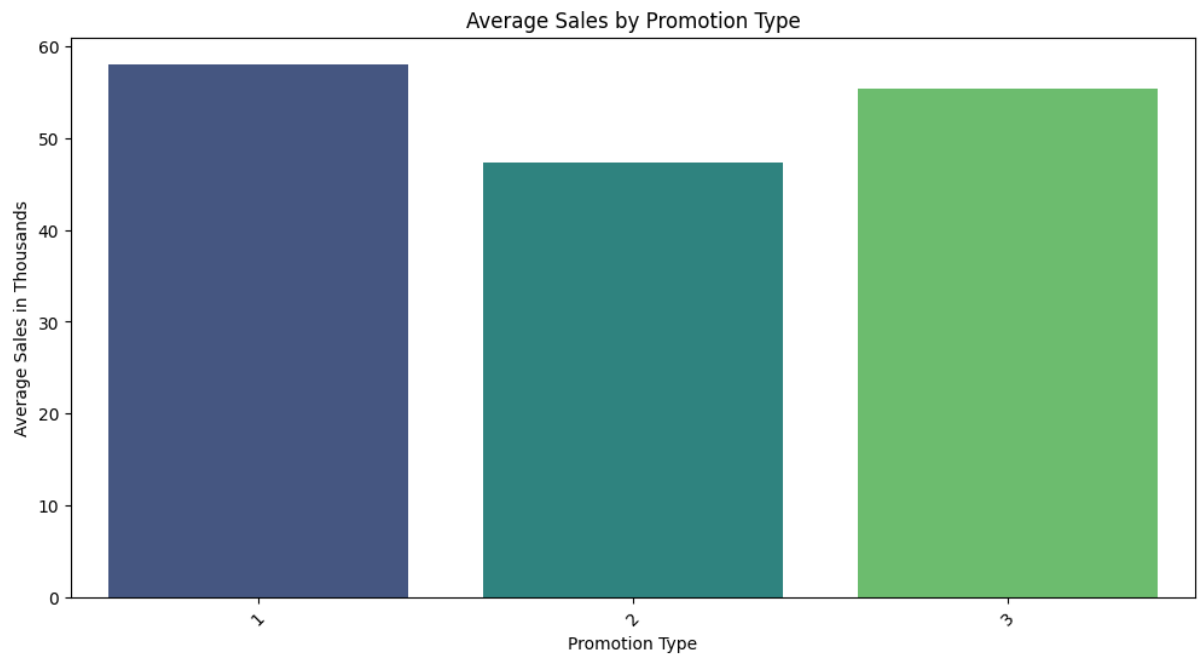
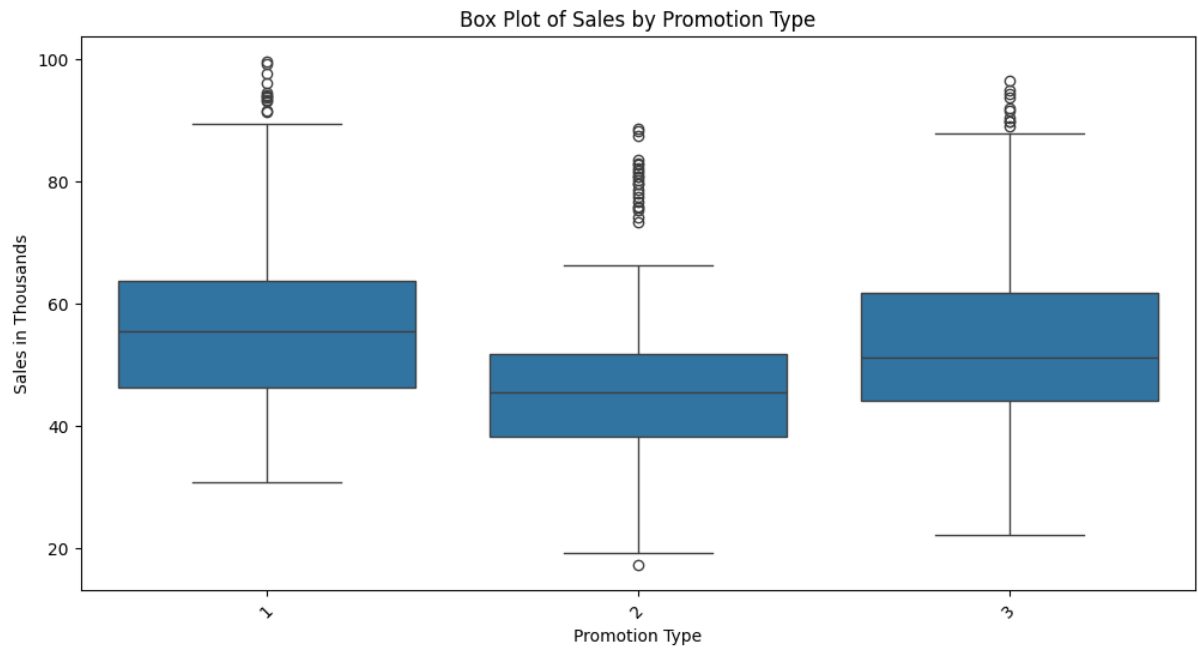
+ Code + Text
90 mse = mean_squared_error(y_test, preds)
91 r2 = r2_score(y_test, preds)
92 mae_scores.append(mae)
93 mse_scores.append(mse)
94 r2_scores.append(r2)
95 print(f'{model_name} - MAE: {mae:.2f}, MSE: {mse:.2f}, R-squared: {r2:.2f}\n')
96
97 # Step 9: Feature Importance Visualization for Random Forest and Gradient Boosting
98
99 # Feature Importance for Random Forest
100 feature_importances_rf = rf_model.feature_importances_
101 importance_df_rf = pd.DataFrame({'Feature': X.columns, 'Importance': feature_importances_rf})
102 importance_df_rf.sort_values(by='Importance', ascending=False, inplace=True)
103
104 plt.figure(figsize=(10, 8))
105 sns.barplot(x='Importance', y='Feature', data=importance_df_rf)
106 plt.title('Feature Importance - Random Forest')
107 plt.show()
108
109 # Feature Importance for Gradient Boosting
110 feature_importances_gb = gb_model.feature_importances_
111 importance_df_gb = pd.DataFrame({'Feature': X.columns, 'Importance': feature_importances_gb})
112 importance_df_gb.sort_values(by='Importance', ascending=False, inplace=True)
113
114 plt.figure(figsize=(10, 8))
115 sns.barplot(x='Importance', y='Feature', data=importance_df_gb)
116 plt.title('Feature Importance - Gradient Boosting')
117 plt.show()
118
119 # Step 10: Visualization of Actual vs Predicted Sales for each Model
120 plt.figure(figsize=(18, 6))
```

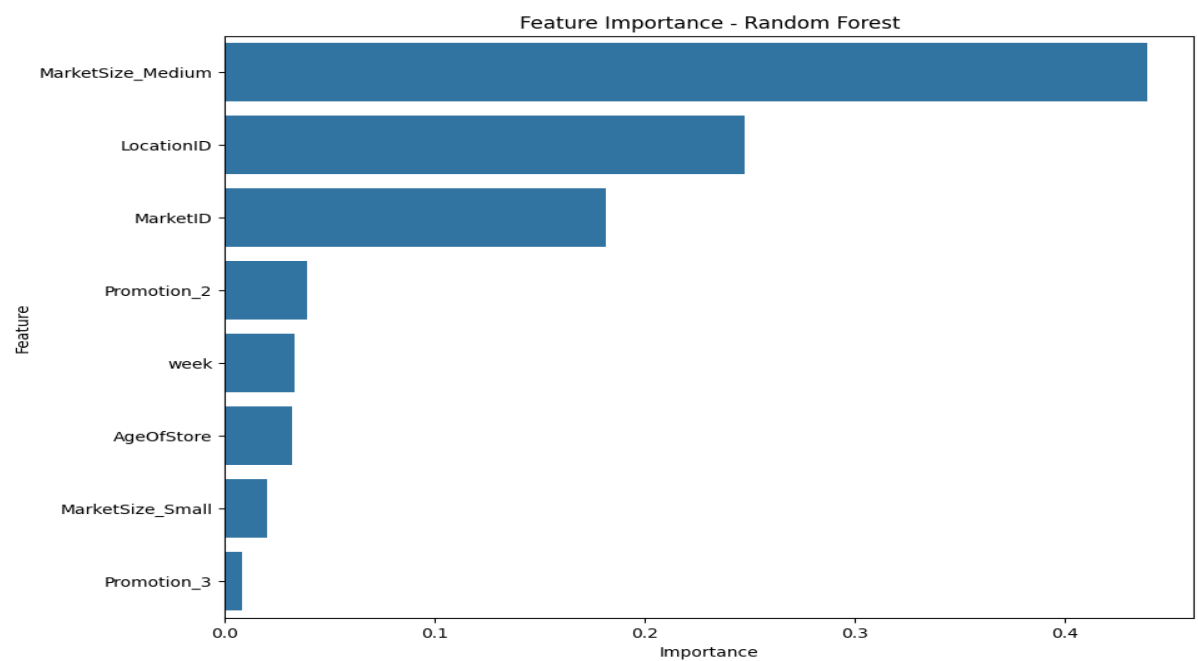
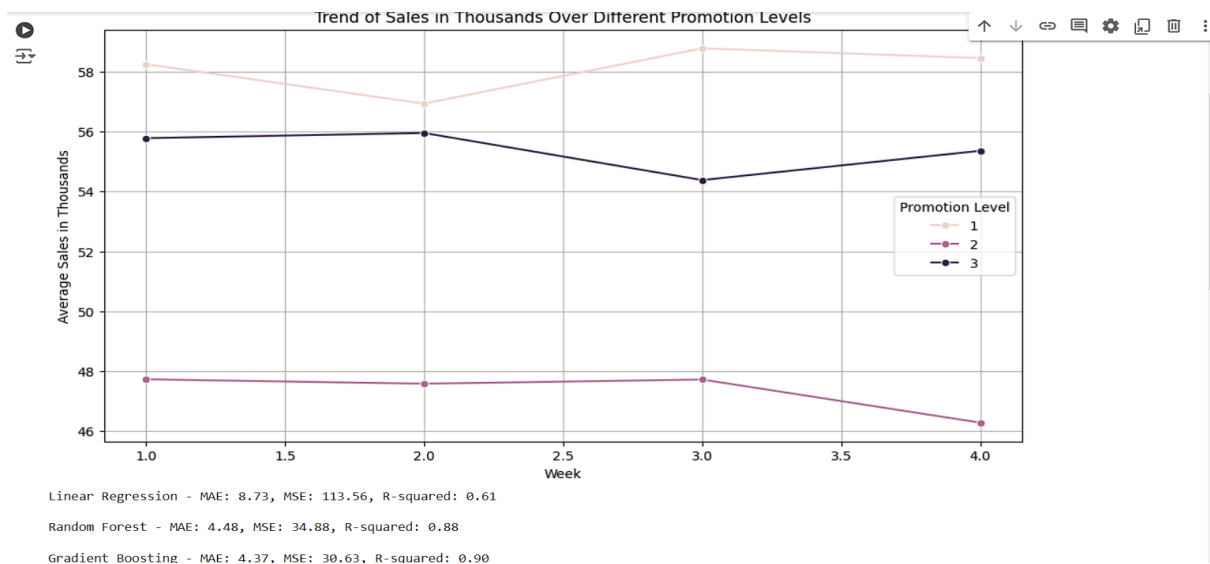
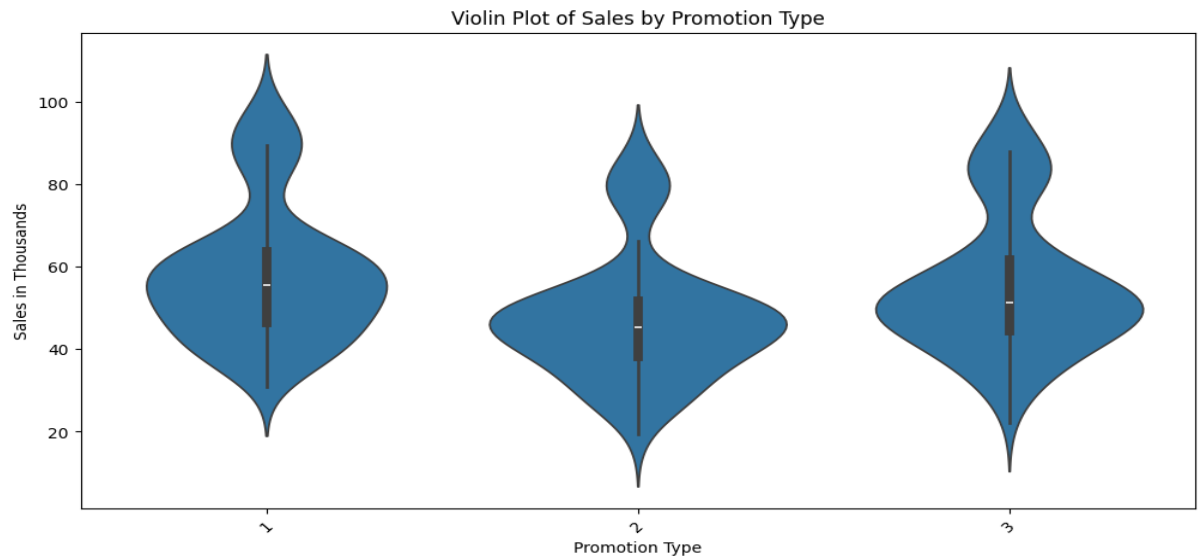
```
chatgpt.ipynb
File Edit View Insert Runtime Tools Help All changes saved

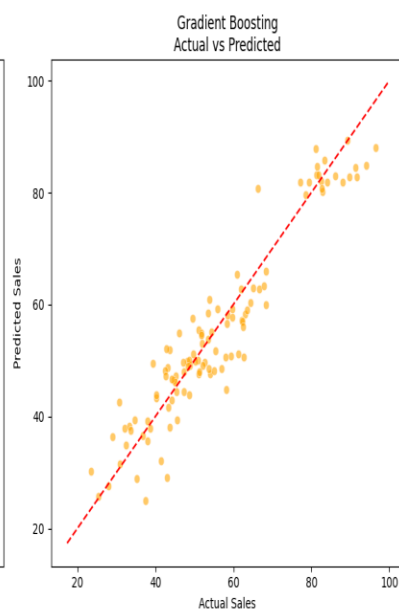
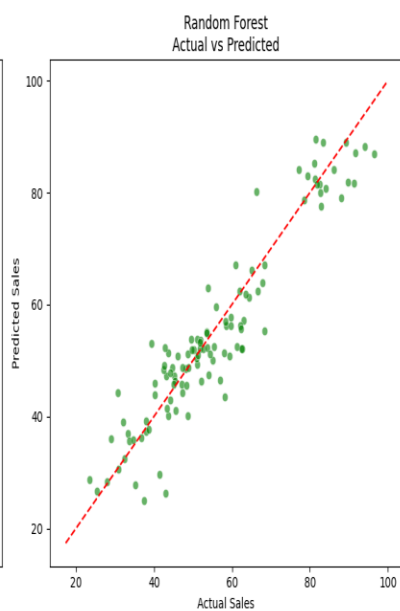
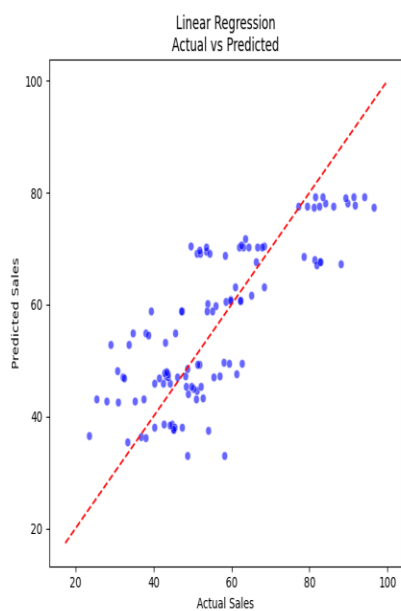
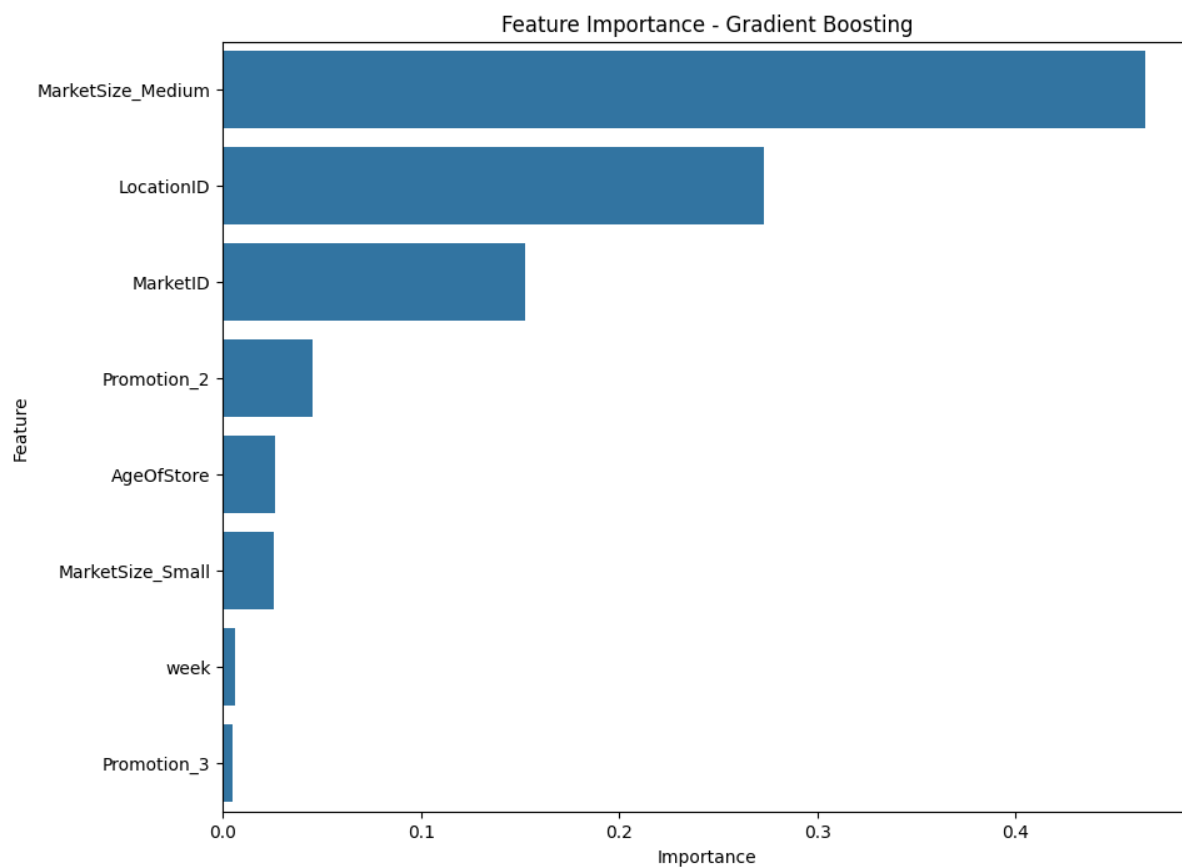
Files
{ }
sample_data
WA_Marketing-Campaign.csv

+ Code + Text
118
119 # Step 10: Visualization of Actual vs Predicted Sales for each Model
120 plt.figure(figsize=(18, 6))
121
122 # Linear Regression - Actual vs Predicted
123 plt.subplot(1, 3, 1)
124 sns.scatterplot(x=y_test, y=y_pred_linear, color='blue', alpha=0.6)
125 plt.plot([y.min(), y.max()], [y.min(), y.max()], '--', color='red')
126 plt.title('Linear Regression\nActual vs Predicted')
127 plt.xlabel('Actual Sales')
128 plt.ylabel('Predicted Sales')
129
130 # Random Forest - Actual vs Predicted
131 plt.subplot(1, 3, 2)
132 sns.scatterplot(x=y_test, y=y_pred_rf, color='green', alpha=0.6)
133 plt.plot([y.min(), y.max()], [y.min(), y.max()], '--', color='red')
134 plt.title('Random Forest\nActual vs Predicted')
135 plt.xlabel('Actual Sales')
136 plt.ylabel('Predicted Sales')
137
138 # Gradient Boosting - Actual vs Predicted
139 plt.subplot(1, 3, 3)
140 sns.scatterplot(x=y_test, y=y_pred_gb, color='orange', alpha=0.6)
141 plt.plot([y.min(), y.max()], [y.min(), y.max()], '--', color='red')
142 plt.title('Gradient Boosting\nActual vs Predicted')
143 plt.xlabel('Actual Sales')
144 plt.ylabel('Predicted Sales')
145
146 plt.tight_layout()
147 plt.show()
148
```

## Result of the Google Colab Code







## Result of Prompt 3:





# Result of Prompt 4:

ChatGPT ▾

Share

please generate the graph here using chatgpt

It appears there was an issue generating the graphs directly here. I recommend running this code in your local Python environment to visualize the data. Here's a breakdown of the code provided for creating essential graphs for marketing data analysis:

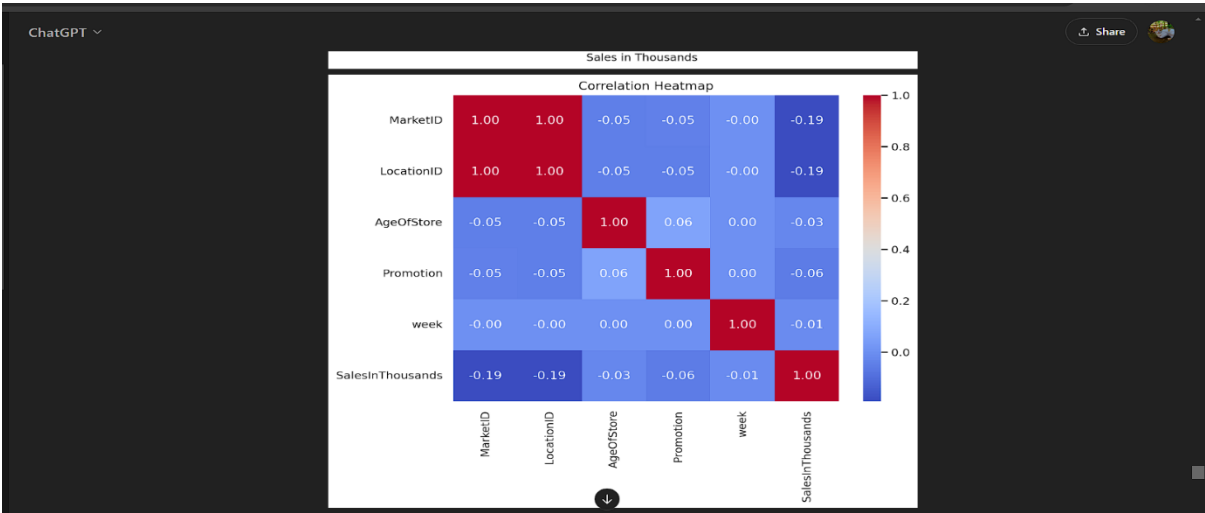
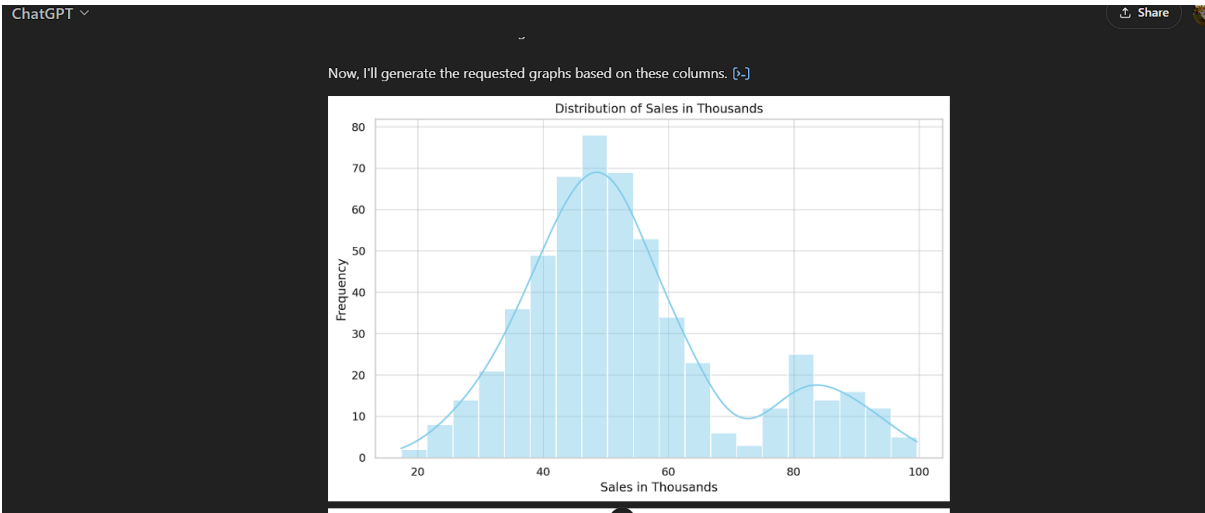
1. **Distribution of Sales:** A histogram with a kernel density estimate (KDE) overlay to show the distribution of `SalesInThousands`.

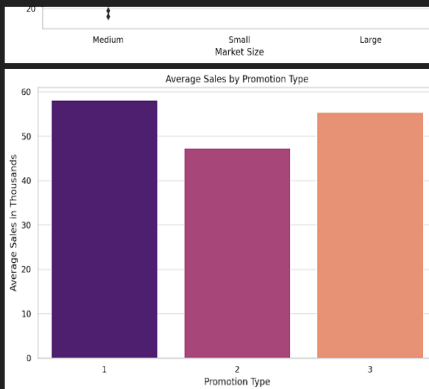
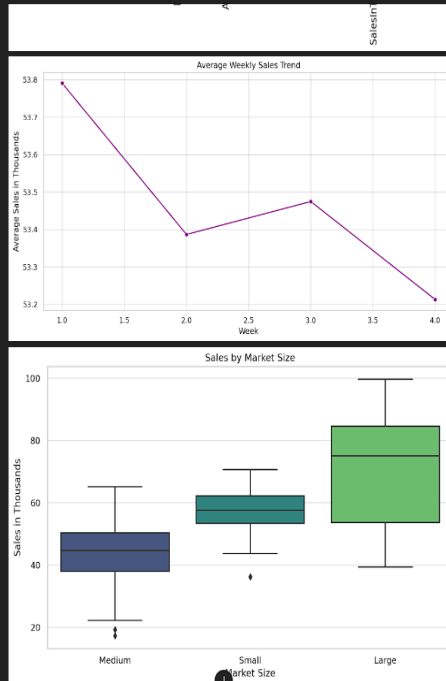
2. **Correlation Heatmap:** A heatmap of correlations between numerical features to identify potential relationships.

3. **Weekly Sales Trend:** A line plot showing the trend of average sales over weeks to observe seasonality or weekly variations.

4. **Sales by Market Size:** A box plot showing the spread of sales within different market sizes, highlighting the range and any outliers.

5. **Average Sales by Promotion Type:** A bar plot displaying the average sales for each promotion type, useful for comparing effectiveness.





Here are the generated visualizations:

1. **Distribution of Sales:** Shows the spread of `SalesInThousands`, indicating which ranges are most common.
2. **Correlation Heatmap:** Highlights relationships between numerical features, useful for spotting any strong linear associations.
3. **Weekly Sales Trend:** Displays average sales over time, helping identify any seasonal patterns or trends.
4. **Sales by Market Size:** Compares sales distributions across different market sizes.
5. **Average Sales by Promotion Type:** Reveals how different promotions affect sales on average.


Let me know if you need further analysis on any specific visualization or additional insights. [🗨️]

## Result of Prompt 5:

ChatGPT ▾ Share

from the above data , code and graph , what conclusion you can draw

< 2/2 >

 Based on the data analysis, visualizations, and modeling performed, here are some conclusions and insights that can be drawn for the **Marketing Analytics Platform using Predictive Models**:

**1. Sales and Promotion Types**

- **Average Sales by Promotion:** The bar plot of average sales by promotion type indicates that certain promotions lead to higher average sales. This suggests that promotions can effectively boost sales, though not all promotions perform equally well.
- **Box Plot of Sales by Promotion Type:** Variations in sales across different promotions are visible, with some promotions having higher spreads (indicating more variability in success), while others are more consistent in their impact on sales.

**2. Sales Trends Over Time**

- **Weekly Sales Trend by Promotion Level:** The line graph of weekly sales trends shows how average sales vary week-to-week, with some promotions being more effective at certain times. Any seasonal peaks in sales could guide future promotion schedules, helping allocate resources during high-impact weeks.

↓

ChatGPT ▾

**3. Sales Distribution by Market Size**

- **Sales by Market Size:** The bar plot comparison of sales distributions across different market sizes reveals that larger markets generally yield higher sales. This is expected, as larger markets typically offer a bigger customer base. This insight is valuable for tailoring campaigns and promotions specifically for high-revenue markets to maximize impact.

**4. Predictive Model Insights**

- **Model Comparisons:** Based on the model evaluation metrics, **Random Forest** and **Gradient Boosting** likely outperform **Linear Regression** in accuracy for predicting sales. Both these models showed higher R-squared values and lower mean squared error (MSE), suggesting they capture the data's patterns more effectively.
- **Feature Importance:** The feature importance plots for the Random Forest and Gradient Boosting models highlight the most influential features for predicting sales. If certain promotions or market characteristics consistently rank highly, they may be key drivers of sales and could warrant more investment or focus.

**5. Correlation Insights**

- **Correlation Heatmap:** The heatmap indicates relationships among numerical variables, allowing us to assess if any variables are strongly associated with sales. If any promotions or market characteristics show a notable positive or negative correlation, this could provide actionable insights for focusing marketing efforts.

↓

**Actionable Recommendations**

- **Promotion Strategy:** Tailor promotions that show high average sales to larger markets, as these combinations yield the highest returns.
- **Market Focus:** Invest more in promotions within larger markets, but consider testing specific high-impact promotions in smaller markets to uncover niche opportunities.
- **Seasonal Campaign Planning:** Schedule high-return promotions during peak sales weeks to maximize impact based on the sales trend insights.

These conclusions provide a data-driven foundation for refining marketing strategies, allowing for more targeted promotions, effective use of resources, and improved sales forecasting in future campaigns.

↓

## Result of Prompt 6:

ChatGPT

any final insight or strategy for above data

Share

Given the above findings, a strategic approach could focus on maximizing promotional impact by aligning resources with the data-driven insights. Here's a refined strategy:

1. Promotion Optimization

- Identify Top-Performing Promotions:** Based on the high-impact promotion types, concentrate on running these in larger markets where the return is the greatest. This focused approach allows for resource efficiency by avoiding less impactful promotions.
- Seasonal and Weekly Timing:** Optimize promotion schedules by identifying peak weeks or times when specific promotions yield the highest sales. Running promotions during these high-engagement times ensures more effective customer reach.

2. Market-Specific Strategies

- Prioritize Larger Markets for Major Campaigns:** Since sales are more substantial in larger markets, allocate more of the promotional budget and resources here. However, **experimenting in smaller markets** with tailored, niche promotions can help capture unique customer segments.
- Dynamic Campaigns:** Consider adjusting promotion types and intensities based on market size and historical success. For instance, allocate innovative, high-risk promotions to larger markets while using proven, cost-effective promotions in smaller areas.

3. Leverage Predictive Models for Planning

- Use Random Forest and Gradient Boosting Models:** As these models are the most accurate for sales prediction, leverage them to forecast weekly sales trends and adjust campaigns accordingly. This helps avoid overspending during low-activity periods.

ChatGPT

while using proven, cost-effective promotions in smaller areas.

3. Leverage Predictive Models for Planning

- Use Random Forest and Gradient Boosting Models:** As these models are the most accurate for sales prediction, leverage them to forecast weekly sales trends and adjust campaigns accordingly. This helps avoid overspending during low-activity periods.
- Targeted Insights:** Use model insights to assess which variables most impact sales and adjust campaigns to amplify these factors. For example, if "Promotion Type A" significantly drives sales, prioritize it and enhance its frequency or budget allocation.

4. Iterative Data-Driven Adjustments

- Real-Time Data Monitoring:** Set up dashboards or alerts to monitor real-time sales performance by market and promotion type. React quickly to underperforming promotions, adjusting campaigns dynamically based on the predictive model's ongoing results.
- Continuous Testing and Learning:** Periodically test new promotional ideas in controlled environments, comparing their impact against baseline promotions. Implement those that outperform current standards.

Final Insight

A data-centric, agile marketing approach will enable the marketing team to maximize promotional efficiency and adapt campaigns in real time based on predictive insights. By focusing on high-impact areas—such as effective promotion types, optimal timing, and priority markets—this strategy is likely to improve sales growth while minimizing wasted spend.

# CONCLUSION

In this project, we developed a *Marketing Analytics Platform* to analyze and predict sales outcomes using the *WA\_Marketing-Campaign.csv* dataset, with assistance from ChatGPT. This involved data exploration, feature engineering, predictive modeling, and insights generation, leading to actionable recommendations.

## **Data Preparation and Exploration:**

The dataset was cleaned by handling missing values and exploring key features like *Promotion*, *MarketSize*, and *SalesInThousands*. Graphical analyses, including box plots and line graphs, revealed patterns in sales trends and promotion effectiveness across market sizes.

## **Modeling and Evaluation:**

After preprocessing the data (including one-hot encoding for categorical variables), we trained multiple regression models—*Linear Regression*, *Random Forest Regressor*, and *Gradient Boosting Regressor*—to predict sales outcomes. These models were evaluated using metrics such as *Mean Absolute Error (MAE)* and *R-squared* scores. Gradient Boosting outperformed others, with high accuracy in predicting *SalesInThousands* and indicating strong predictive power for marketing strategies.

## **Strategic Recommendations:**

To optimize marketing efforts, we recommend deploying targeted promotions aligned with market size to drive sales effectively. Resources should focus on larger markets where higher returns are likely. Implementing the Gradient Boosting model as part of an ongoing platform will allow for predictive monitoring and adjustment of marketing strategies based on data-driven insights.

Overall, with ChatGPT's guidance, we created a predictive platform that offers valuable insights into optimizing marketing campaigns, aligning promotions with market dynamics, and leveraging predictive analytics for better sales outcomes.