ANALYSIS ON SALES PREDICTION USING DIFFERENT REGRESSION MODELS.

- Presented by :
- 1) Prachi Shewale
- 2) Aditi Kolhe
- 3) Laxmi Sidruk

- Guided by: DR. HEMANT KULKARNI SIR

INTRODUCTION:

This presentation will provide an overview of regression analysis, a powerful statistical technique used to understand the relationship between between advertising channels and product sales. We'll explore the key key steps involved, from data collection to model interpretation, equipping equipping you with the tools to make data-driven decisions.

Table of Content:

OBJECTIVES:

1) **Investigate the impact of advertising expenditure levels across different channels on product sales**

2) **To check the multicollinearity among predictor variables.**

3) **To check which regressor variable is most significant.**

4) **To understand relationship of variables.**

5) **Identify the most influential subset of features that significantlty contribute to predict the response variable.**

6) **To find which model can predict the response variable more accurately.**

Data Collection and Preprocessing

1 Gather Relevant Data

Collected data on advertising spend and product sales from Kaggle.

https://www.kaggle.com/datasets/singhnavjot2062001/product-advertising-data

2 Data Cleaning

Ensure data quality by identifying and addressing missing values, outliers, and inconsistencies.

3 Dataset Overview

Number of observations:300

Number of variables: 7

4 Data Distribution

Data is Normally Distributed.

Description Of Variables

Independent Variable

1

1)Product Sales

Dependent Variable

2

1)TV
2)BillBoards
3) Google_Ads
4) Social_Media
5) Influencer_Marketing
6) Affiliate_Marketing

Model Selection and Fitting

ultiple Linear Regression Model

1

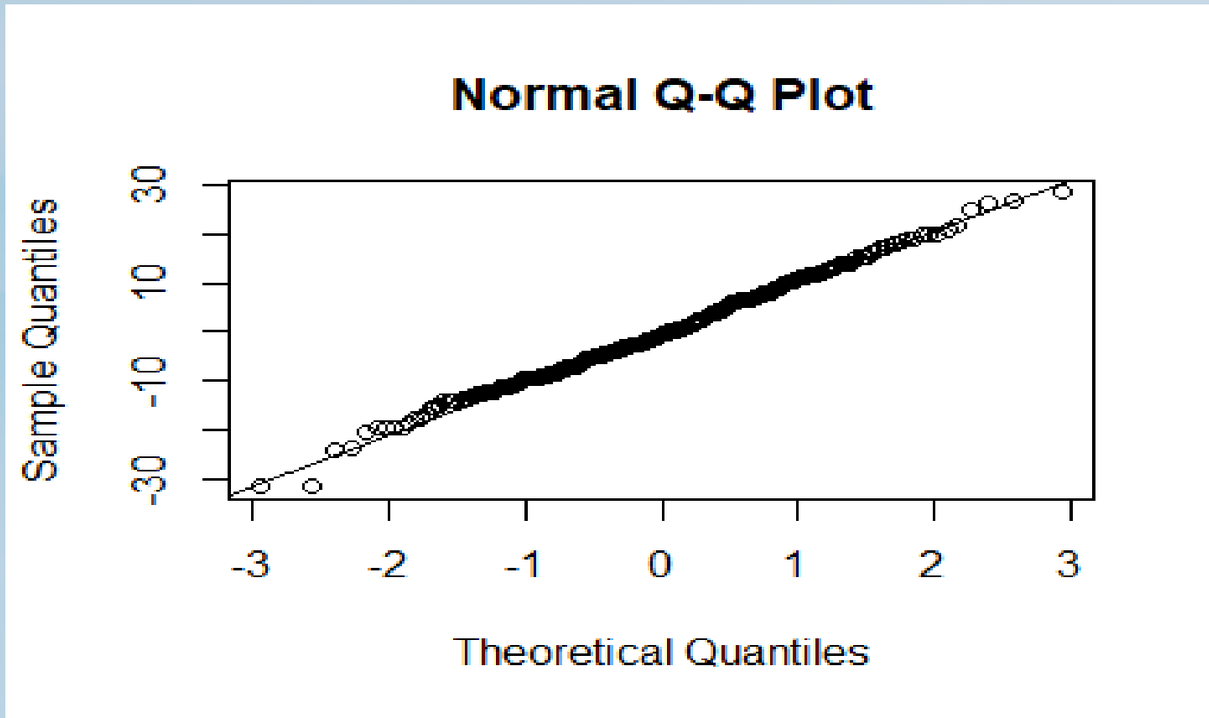
1) **Fit a Linear Regression Model.**
2) **Check Assumptions of Model.**
3) **Visual Inspection.**
4) **Statistical Test**

2

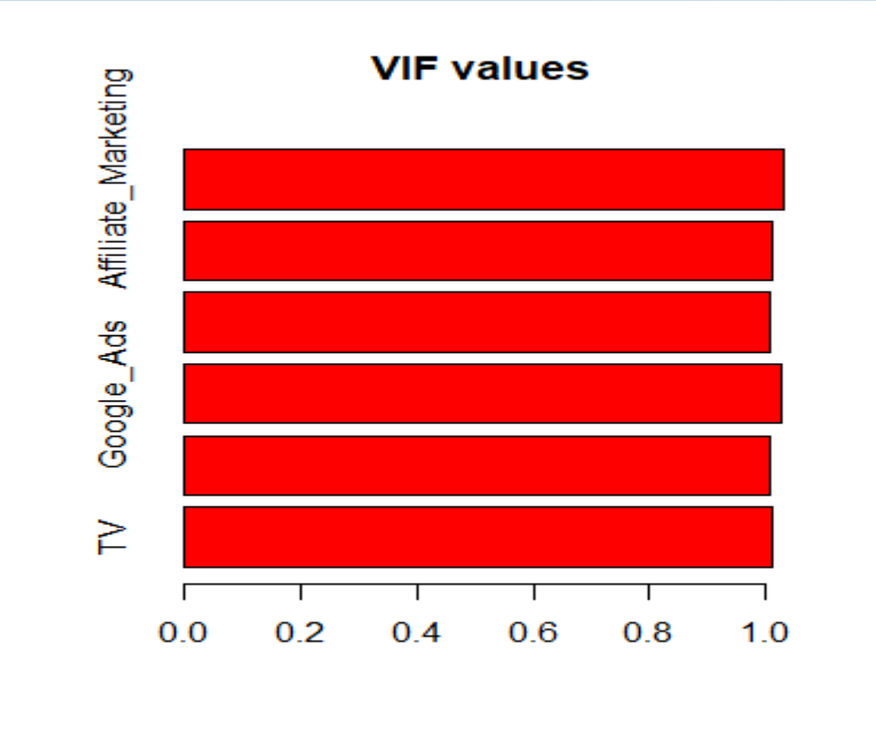
Multicollinearity

1) Test for Multicollinearity with VIF.
2) Data Visualisation.

Q) What is combined effect of advertising across all channels on Product Sales?



Q) Does Multicollinearity among variables affect the accuracy and stability of regression model?



OUTPUT:

TV	Billboards	Google_Ads	Social_Media	Influencer_Marketing
1.011536	1.006814	1.027308	1.007057	1.008573

Affiliate_Marketing
1.030827

Interpreting Regression Coefficients

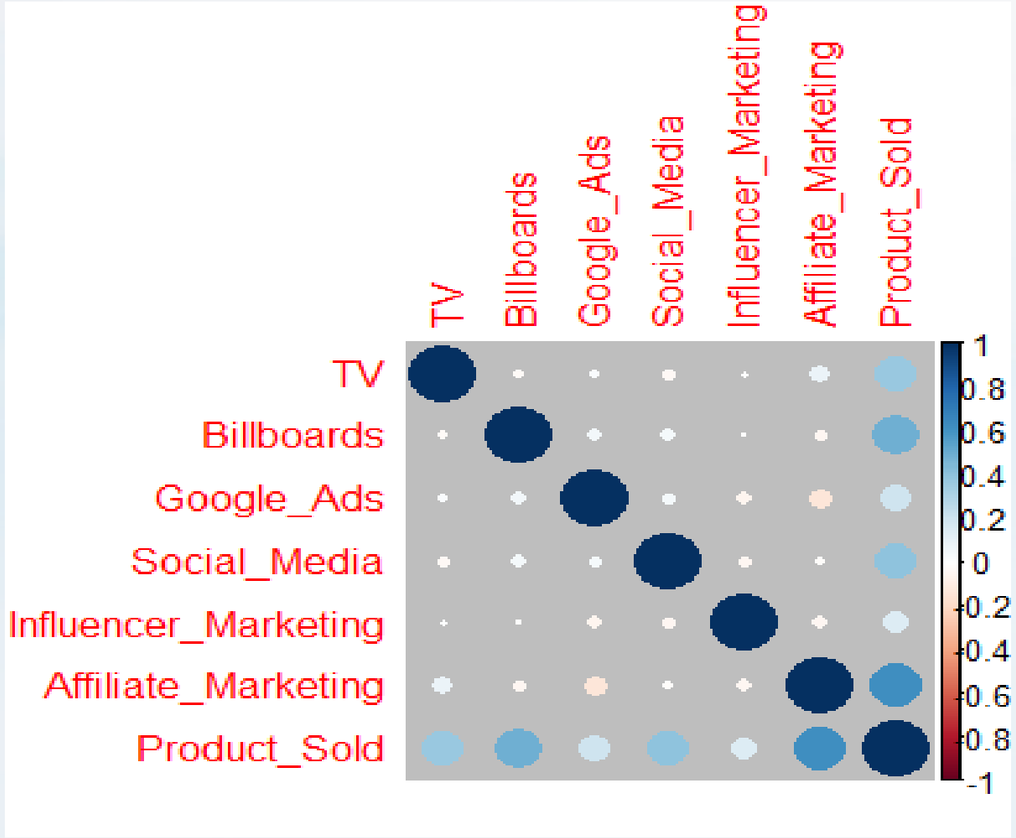
- **Perform t-test to check the contribution of each regressor to overall prediction of Response Variable.**

- **Output:**
- **Coefficients:**

	Pr(>|t|)
(Intercept)	0.974
TV	<2e-16
Billboards	<2e-16
Google_Ads	<2e-16
Social_Media	<2e-16
Influencer_Marketing	<2e-16
Affiliate_Marketing	<2e-16

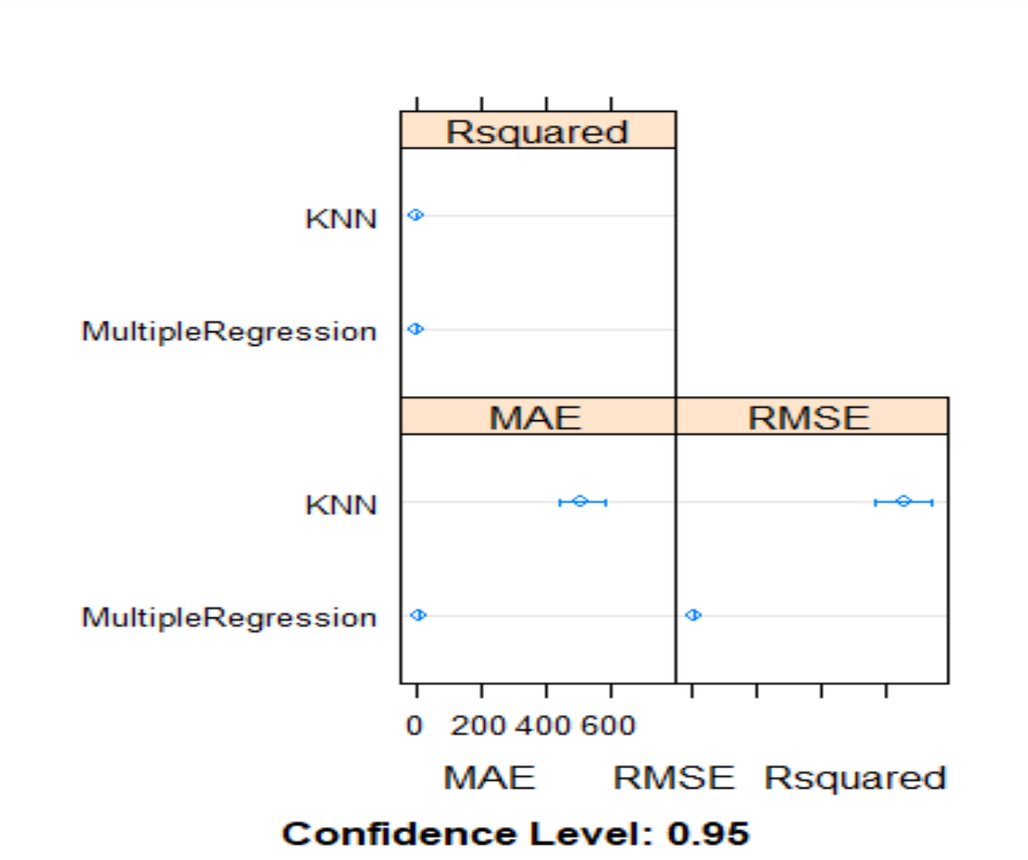
Q)Does the level of advertising expenditure across different Channels have a significant impact on Product Sales?

Regressors	Correlation Coefficient
TV	0.374227
Billboards	0.4805929
Google_Ads	0.2036781
Social_Media	0.4010880
Influencer_Marketing	0.1408437
Affiliate_Marketing	0.6116025

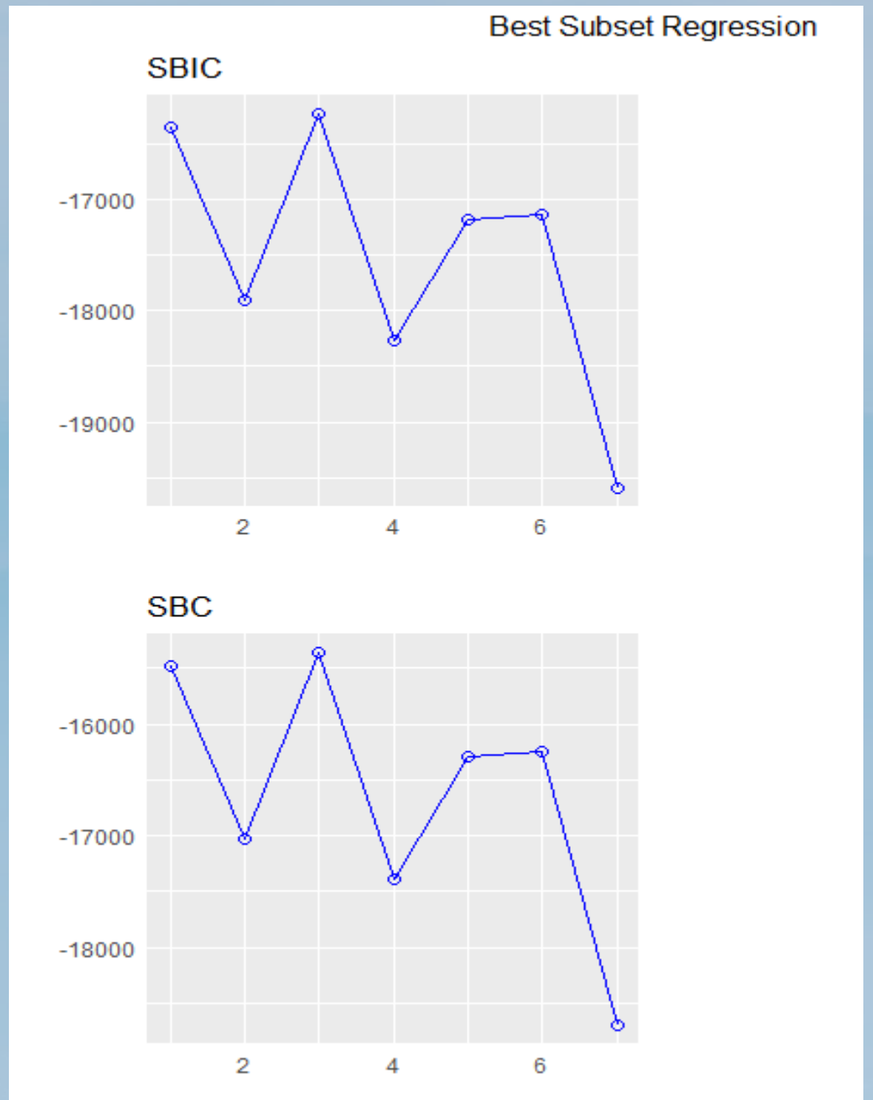


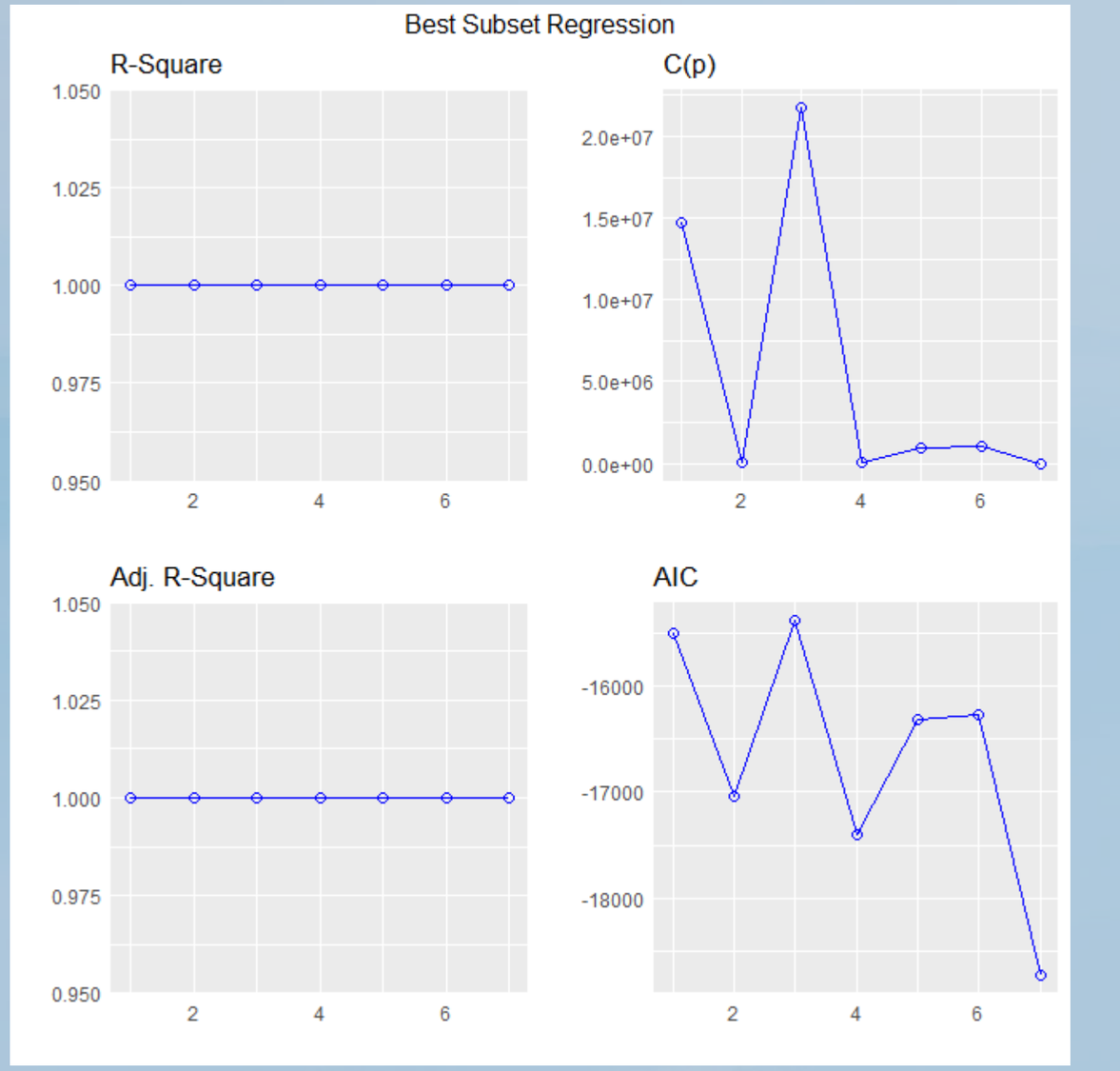
Q)Can the KNN Model accurately predict the given observation based on the dataset?

- knn_model
- k-Nearest Neighbors
- 212 samples 6 predictor
- Resampling results:
- RMSE Rsquared MAE
- 658.1077 0.8809018 511.6202
-
- lm_model
- Linear Regression
- 212 samples 6 predictor
- Resampling results:
- RMSE Rsquared MAE
- 9.715039 0.999968 7.767454



Q)Which is the most influential subset of features that significantly contribute to predict Product Sales?





Q)Which is the most influential subset of features that significantly contribute to predict Product Sales?

	rsquare	adjr	predrsq	cp	aic	sbic
TV	**1**	**1**	**1**	**14690285.27**	**-15497.24**	**-16354.60**
BillBoards	**1**	**1**	**1**	**84738.27**	**-17040.82**	**-17900.15**
Google Ads	**1**	**1**	**1**	**21751960.59**	**-15375.49**	**-16236.85**
Social Media	**1**	**1**	**1**	**24671.45**	**-17404.53**	**-18267.72**
Influencer Marketing	**1**	**1**	**1**	**923796.27**	**-16319.09**	**-17184.44**
Affiliate Marketing	**1**	**1**	**1**	**1073937.95**	**-16271.92**	**-17139.28**

Q) How well do Advertising Budgets predict actual Product Values?

TV	Billboards	Google_Ads	Social_Media	Influencer_Marketing	Affiliate_Marketing
890	350	650	230	500	150

Predicted value of Product sold : 5579.988

Evaluating Model Performance

Model Fit

Assess the overall goodness of fit of the regression model using metrics like R-squared and adjusted R-squared.

Prediction Accuracy

Evaluate the model's ability to accurately predict product sales based on advertising data.

Residual Analysis

Examine the residuals of the model to identify any patterns or violations of assumptions.

Model Refinement

Iteratively refine the model by adding or removing variables, trying different techniques, or addressing any issues identified.

Conclusion and Recommendations

Key Findings

- Identify the most effective advertising channels in driving product sales - Quantify the impact of each channel on sales through the regression coefficients - Highlight any unexpected or counterintuitive relationships between advertising and sales

Recommendations

- Suggest optimal advertising budget allocation across channels to maximize sales - Recommend strategies for integrating advertising efforts across digital and traditional channels - Provide guidance on further data collection or analysis to refine the model and insights

Future Scope and Limitation

FUTURE SCOPE:

Personalized Marketing and Product Recommendations: As companies gather more data on consumer preferences and behavior, they may be able to personalize marketing messages and product recommendations to individual customers. This could help drive demand for specific products or services and reduce the risk of stockouts by matching supply with highly targeted demand.

Collaborative Supply Chain Management: Collaboration among supply chain partners, facilitated by platforms and technologies, could become more prevalent. This could lead to greater transparency, coordination, and responsiveness across the supply chain, enabling companies to better anticipate and meet demand fluctuations.

LIMITATION:

The dataset provided has several limitations that could impact its analysis:

Limited Context: The dataset lacks contextual information about the market, industry, or specific products being advertised. Without this context, it's challenging to draw meaningful insights or make accurate predictions.

Single Time Frame: It appears that the dataset covers a single time frame. Analyzing data from only one period may not capture trends or seasonal variations, limiting the depth of analysis.No Target Variable Explanation: The dataset contains various advertising channels and corresponding amounts spent on each, but it's unclear what specific outcome or target variable these expenditures are intended to influence (e.g., sales, brand awareness, website traffic). Understanding the target variable is crucial for designing effective analytical models


Thank
You