# Introduction

## Background and Problem Statement

In an era of abundant advertising platforms, businesses often struggle to optimize their advertising expenditures to maximize product sales. The objective of this project is to develop a predictive model that accurately estimates product sales based on advertising budgets in various media channels, namely TV, radio, and newspaper. By doing so, businesses can allocate their advertising budgets more effectively and make data-driven decisions to enhance their marketing strategies.

## Objectives of the Project

The primary objectives of this project include:

* Exploring the relationships between product sales and advertising budgets in TV, radio, and newspaper.
* Developing machine learning models to predict sales from advertising budgets.
* Comparing the performance of different models and selecting the most suitable one.
* Providing insights into the impact of advertising media on sales through model interpretation.

# Data Understanding and Preparation

## Description of the Dataset

The dataset consists of product sales in 200 different markets and corresponding advertising budgets in TV, radio, and newspaper. The dataset was preprocessed to handle missing values and ensure data integrity. Features were analyzed to understand their distributions and relationships.

## Data Cleaning and Preprocessing

Missing values were addressed using appropriate techniques. Data types were checked and converted if needed. Outliers were examined and handled based on domain knowledge.

## Feature Analysis and Selection

Correlation analysis was conducted to identify relationships between sales and advertising budgets.

# Exploratory Data Analysis (EDA)

## Visualizing Trends in Sales and Advertising Budgets

Histograms and box plots were used to visualize the distributions of sales and advertising budgets in different media. Scatter plots revealed relationships between sales and individual advertising channels.

## Correlation Analysis and Insights

Correlation coefficients were calculated to quantify relationships between sales and advertising media. Pair plots were created to visualize correlations among all features.

## Interaction Analysis between Advertising Media

Interactions between advertising media were explored through scatter plots and visualizations, shedding light on potential combined effects.

# Machine Learning Model Building

## Data Splitting and Standardization

The dataset was split into training and testing sets. Features were standardized to ensure fair comparisons between models.

## Model Selection and Evaluation Metrics

Linear Regression, Polynomial Regression, Ridge Regression, and Random Forest Regressor were chosen as candidate models. Metrics included Mean Squared Error (MSE), R-Squared (R2), and Cross-Validation metrics (MSE CV, Avg R2 CV).

## Model Building: Linear Regression, Polynomial Regression, Ridge Regression, Random Forest Regressor

Each model was trained, evaluated, and cross-validated using the standardized training data. Performance metrics were recorded for model comparison.

# Model Evaluation and Comparison

## Comparison of Evaluation Metrics

Performance metrics (MSE, R2, MSE CV, Avg R2 CV) were compared across models. The most suitable model was selected based on the metrics' values.

## Interpretation of Model Performance

Interpretation of metrics in the context of the problem and business goals helped in assessing the models' effectiveness.

## Selection of the Final Model

The model with the best combination of evaluation metrics was identified as the final model for predicting product sales.

# Interpreting Model Coefficients

**I**nterpretation of Polynomial Regression Coefficients

Coefficients of polynomial terms and interactions were analyzed to understand the relationships between advertising media and sales.

## Interpretation of Linear Regression Coefficients

Linear coefficients provided insights into the linear relationships between advertising budgets and sales.

# Conclusions and Recommendations

## Summary of Findings and Insights

The final model's performance and its interpretability were summarized. Key findings about the impact of advertising media on sales were highlighted.

## Implications for Decision-Making

Businesses can use the final model to make informed decisions about allocating advertising budgets to maximize product sales.

## Recommendations for Future Work

Future work could involve exploring more complex model architectures, incorporating external factors, or testing the model on new data.

This document provides a comprehensive overview of the project, from data preparation and exploration to model building and interpretation. It outlines the methodology, findings, and insights gained from predicting product sales based on advertising budgets. The project's outcomes can assist businesses in making effective marketing decisions to achieve their sales goals.