INFLUENTIAL FACTORS AFFECTING SALES AT WALMART AND PREDICTING SALES AT WALMART USING LINEAR REGRESSION.

*ANUPAMA KADAMBI ASU ID:1222559637*

**> Goal:** To find the most influential factors that contribute to sales at Walmart supercenters.

To fit a linear regression model to the dataset and predict the sales at Walmart.

**>Dataset:** Source- Kaggle. Online community for statisticians and data scientists, who publish datasets and are validated by other on the platform.

Total number of predictors: 8. No. of numerical predictors: 4 No. of categorical predictors: 4

Response variable- Weekly sales Temperature, Fuel price, Store number, Holiday Flag

Unemployment rate, CPI. Date (Months, Years)

**>Methodology:**

**Timeline

Description automatically generated**

EXPLORATORY DATA ANALYSIS

Square

Description automatically generatedChart, scatter chart

Description automatically generated

Chart, line chart

Description automatically generated

HISTOGRAMS BOXPLOT SALES VS NUM. PREDICTORS.

1. Sales

Chart, histogram

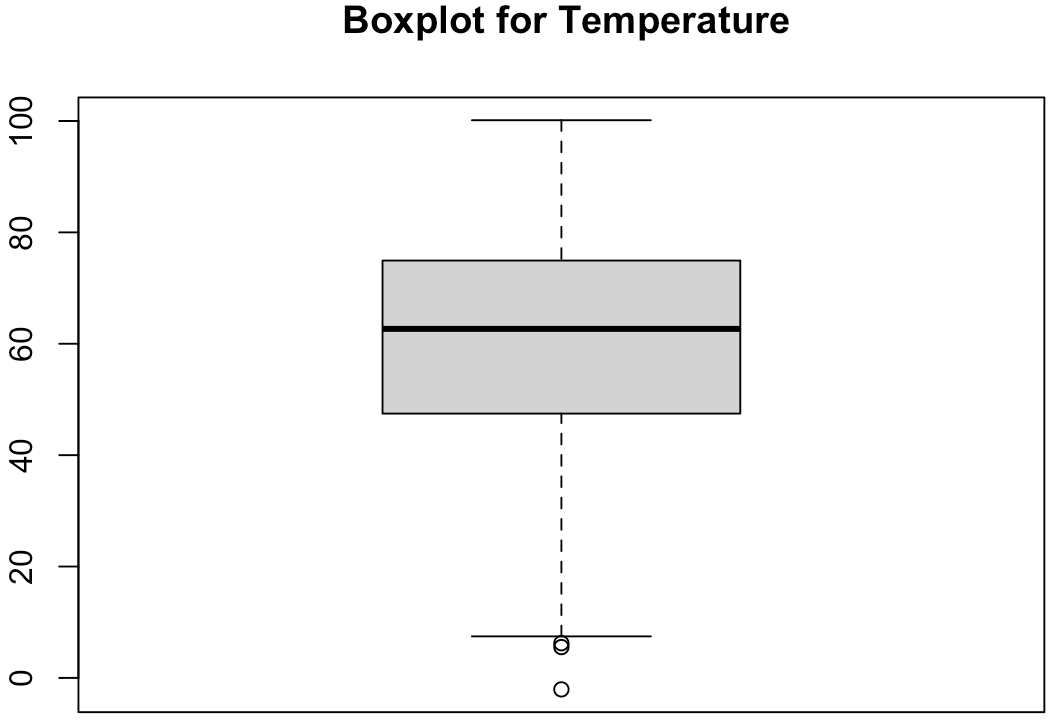
Description automatically generated Chart, box and whisker chart

Description automatically generated Chart, histogram

Description automatically generated

1. Temperature

Chart, histogram

Description automatically generated  Chart, scatter chart

Description automatically generated

1. CPI

Chart, histogram

Description automatically generated Chart, box and whisker chart

Description automatically generated Chart, scatter chart

Description automatically generated

1. Unemployment

Chart, histogram

Description automatically generated Chart, box and whisker chart

Description automatically generated Scatter chart

Description automatically generated

1. Fuel price

Chart, histogram

Description automatically generated Chart, box and whisker chart

Description automatically generated Chart, scatter chart

Description automatically generated

MODELING USING lm()

model<-lm(Weekly\_sales~ Temp+ Fuel\_price+CPI+ Unemployment+ Store\_number+ Is\_Holiday+ Month+ Year, data=walmart)

Text, letter

Description automatically generated

Chart, scatter chart

Description automatically generated

MODEL USING RANDOM FOREST

rf.fit <- randomForest(Weekly\_sales ~ ., data=walmart, ntree=170,keep.forest=FALSE, importance=TRUE)

Graphical user interface

Description automatically generated with medium confidence

Chart, histogram

Description automatically generated

Chart

Description automatically generated Chart

Description automatically generated

DIAGNOSTICS

Diagnostic plots

Graphical user interface, timeline

Description automatically generated Diagram

Description automatically generated

QQ Plot Residual Plot

Chart, line chart

Description automatically generated Chart, scatter chart

Description automatically generated