

# Media Mix: Impact on CoreMotion Sales

Which marketing channels are driving sales the Company?"

CoreMotion Sales Analysis



# Context, Data & Methodology

## ➤ Main Goal

Understand the impact of CoreMotion's marketing channels on sales performance in order to optimize media investments and provide actionable recommendations for executives.

## ➤ Objectives

- Identify which marketing channels (Digital, TV, OOH, etc.) drive sales growth.
- Build a model that quantifies the impact of different media channels on monetary sales.
- Provide a clear narrative of sales dynamics, incorporating seasonality and competitor effects.

## ➤ Target Audience

CoreMotion executives, Marketing Analytics team, and Media Investment managers.

## ➤ Data Context

- Weekly data (Oct 2019–Mar 2023), 4 regions.
- Size: 676 rows × 21 columns.
- Targets: Dollar Sales (main), Units Sales (secondary).
- Drivers: Marketing channels (Digital, OOH, TV), Macroeconomics (GDP, inflation, unemployment), and Competitors' spend.

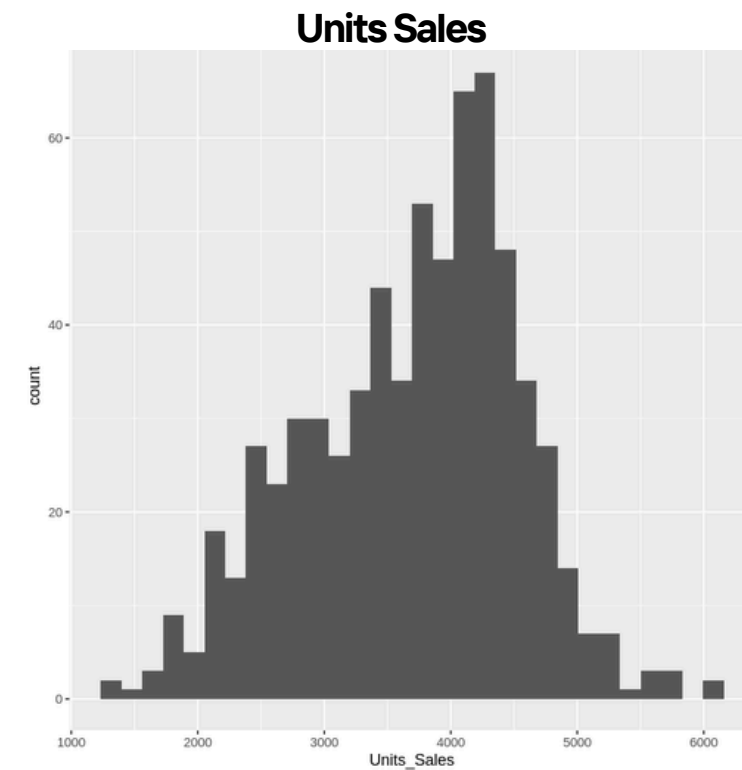
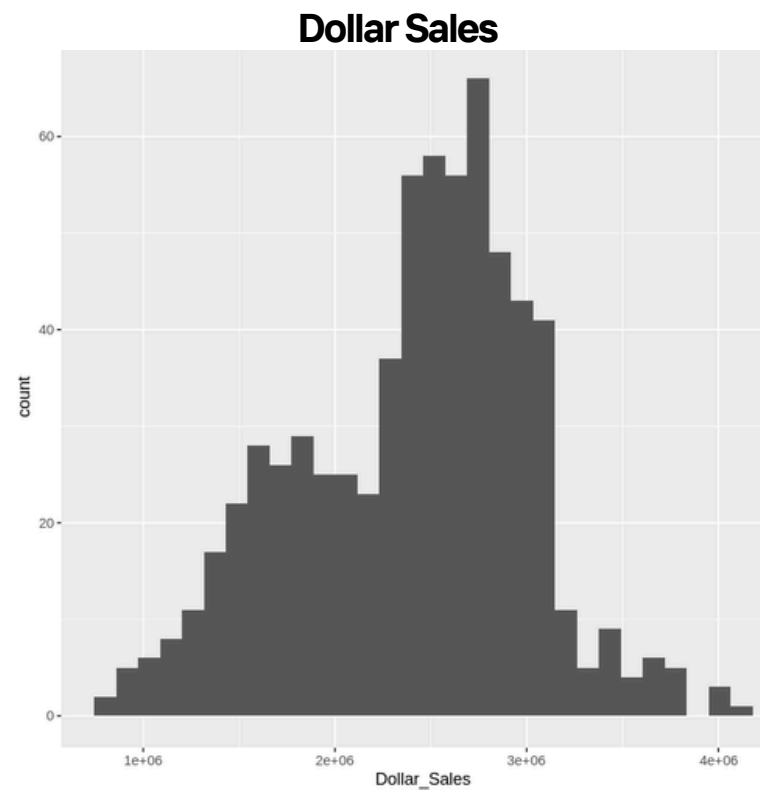
## ➤ Methodology

- CRISP-DM framework to guide the project
- Exploratory Data Analysis
- Linear regression models to quantify the impact of marketing channels and external factors on sales.
- Feature engineering to capture seasonality, lags, and efficiency ratios.
- Evaluation based on accuracy, interpretability, and business relevance

## ➤ Tools

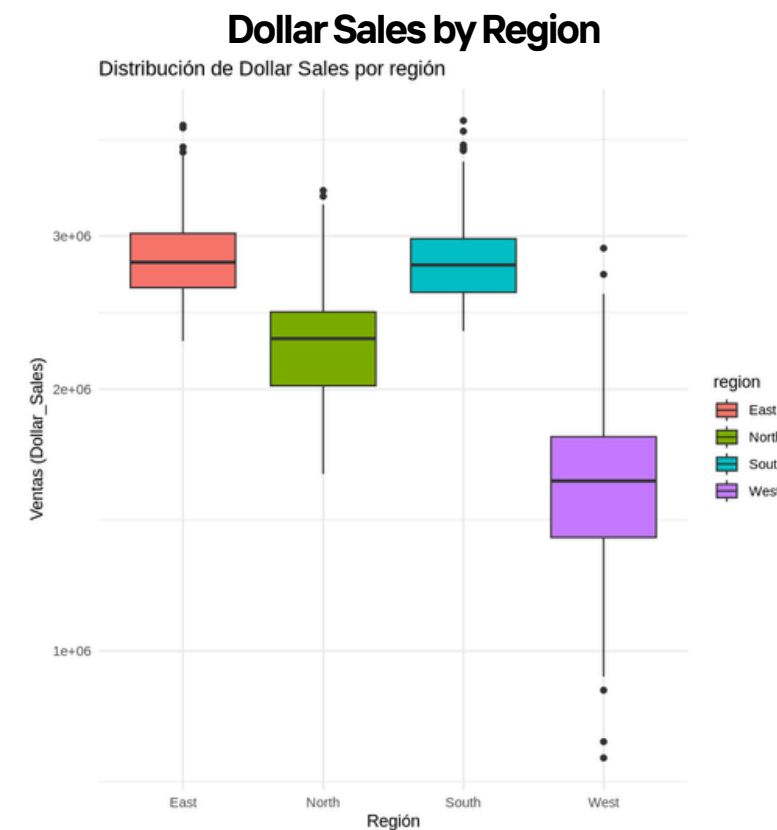
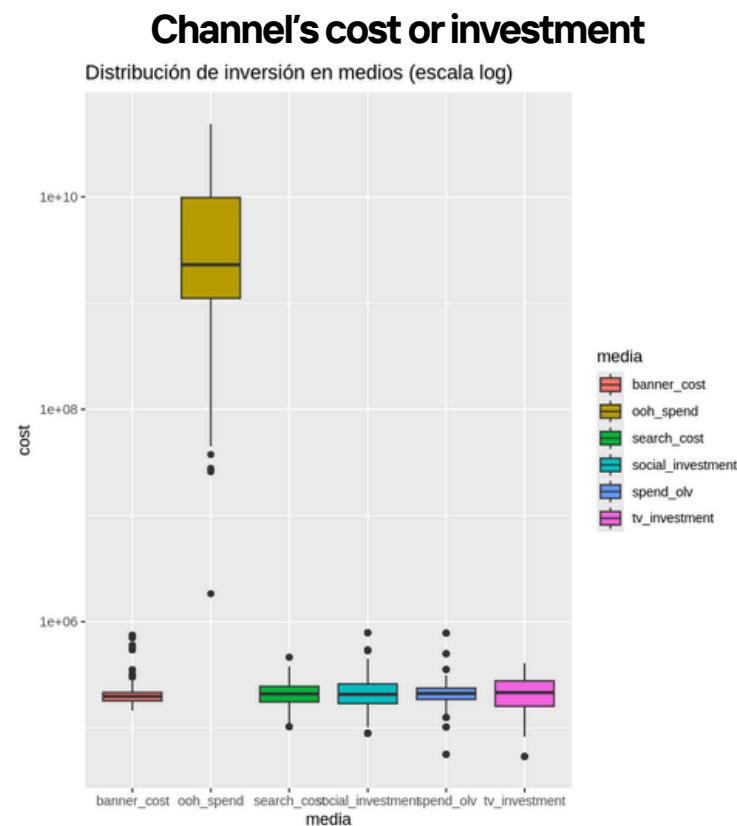
- **Environment:** R version 4.5.1 (2025-06-13), Jupyter Notebook, Google Collab
- **Libraries:** dplyr, tidyr, ggplot2, corrplot, PerformanceAnalytics, lubridate, car, MASS, glmnet, lme4

# Exploratory Data Analysis



## Principal Insights

- Sales are stable overall, with clear seasonal/campaign peaks.
- Weekly data may include natural seasonal peaks (holidays, sports events) that need explicit modeling.
- Some variables (OOH, TV, and digital channels) show spikes in spend or impressions much higher than typical weeks. Many of these spikes likely reflect real campaigns or seasonal pushes (e.g., holidays), while a few may be due to reporting inconsistencies.
- The gap between East/South and West highlights a clear regional imbalance in CoreMotion's performance. Medians align closely with means in all regions, suggesting sales are fairly symmetric with no extreme skew, though West shows more volatility with more outliers.
- Media investments vary a lot in scale. For example, OOH is by far the biggest channel (billions), used in large bursts. Digital channels and TV are much smaller (hundreds of thousands to low millions). Several outlier points appear across channels, which likely represent real campaign spikes.



# Data Preparation

## Data Treatment

- **Time Alignment & Filtering:** Considering multiple data resources and its time ranges, is necessary to restrict the dataset to the period where all key variables are consistently available (2020- 2022).
- **Outliers Treatment:** The data is generally consistent, with fewer than 5% of observations flagged as unusual. Most outliers were found in OOH spend, banner cost, and search cost, as well as a few weeks where Competitor B showed unusually high spending — likely due to aggressive campaigns. Rather than removing these points (which could erase genuine campaign effects), we applied soft capping: any value beyond the typical range was adjusted to a maximum “reasonable” limit. This preserves the signal while preventing a handful of extreme weeks from distorting the overall model.
- **Categorical Data Treatment:** Region is treated as a categorical factor (East, West, North, South). This lets the model capture differences between markets instead of averaging them. Quarter is Added as a seasonal category (Q1–Q4) to reflect recurring patterns like holidays or sports seasons. Special Weeks factor is created a flag for events (e.g., holidays, Black Friday, Christmas) to identify sales peaks that cannot be explained by media spend alone

## Feature Engineering

The process of creating new variables to capture business effects. It helps the model understand real marketing dynamics like efficiency, timing, and competition. Some variables considered are:

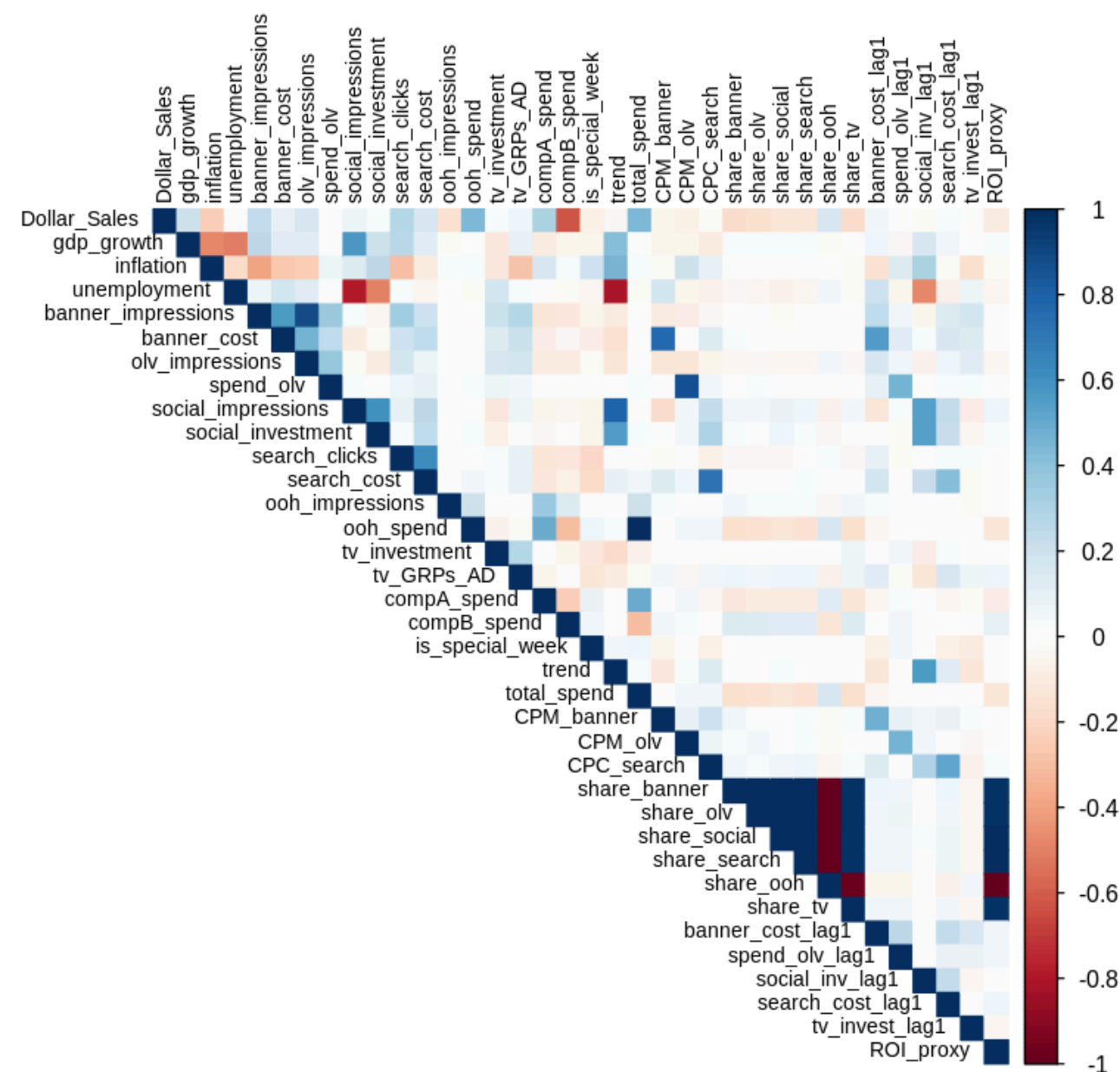
- **Total Spend:** sum of all channels.
- **Channel Shares (%):** share of each medium in the budget.
- **Efficiency Metrics:** CPM (cost per 1k impressions), CPC (cost per click).
- **Lagged Variables:** previous week’s spend to capture delayed effects.
- **ROI Proxy:**  $\text{sales} \div \text{spend}$  (for descriptive use only).

⚠ **Note:** Some variables (shares, ROI proxy) involve sales or total spend, but this is our target variable so they are useful for analysis & storytelling but excluded from the model to avoid data leakage.



# Feature Selection

Matriz de correlaciones



## Principal Insights

- Some of these variables are not included in the model because they are derived from Dollar\_Sales, which would cause data leakage. Data leakage happens when information from the target variable (in this case, sales) is used to build predictors, making the model look more accurate than it really is.

The correlation matrix shows:

- **Strong positive:** OOH spend and search clicks are most aligned with sales.
- **Strong negative:** Competitor B's spend clearly reduces CoreMotion sales.
- **High collinearity:** Digital variables (banner, OLV, social) move together — must be treated carefully in modeling.
- **Macroeconomics:** GDP growth (+) and inflation (–) align with expected consumer behavior.
- **Low correlations:** TV investment, banner cost, unemployment, etc. show little direct link with sales — but may still matter indirectly (e.g., branding effects, labor market context).

⚠ Not every driver shows up as a strong correlation — some have business relevance even if the data doesn't show an immediate effect.



# Modeling

The goal is to understand how different factors — media investments, macroeconomic conditions, competition, and time dynamics — drive Dollar sales in CoreMotion. Keep this in mind, modeling is developed using multiple regression techniques.

## Model Overview

- 1. Full Multiple Regression Model:** Includes all features and it allows to explore and detect potential drivers. Strong correlations among predictors make its results less stable and less reliable for business.
- 2. Reduced Multiple Regression (Selected Variables):** Keeps only statistically significant and theoretically meaningful predictors (What it really matters). Avoids overfitting by leaving out variables that can add noise.
- 3. Generalized Least Squares (GLS) with AR(1) by Region:** This model corrects for time correlation in weekly sales, making results more reliable, and applies AR(1) by region to reflect different sales patterns across regions.

## Model Comparison

Model	Adjusted R <sup>2</sup>	RMSE (USD)	MAPE (%)	Strengths	Limitations
1. Full Multiple Regression	869%	≈218k	≈7.5	Explains ~87% of sales variability; detects potential drivers.	Multicollinearity (overlap between channels); too many weak predictors → less interpretable.
2. Reduced Multiple Regression	869%	≈222k	7.7–8.3	Similar accuracy to Model 1 with fewer variables; simpler and easier to explain.	Mild autocorrelation & heteroscedasticity (DW=1.42); not “perfect” statistically.
3. GLS (AR(1) by Region)	~ 87 % (pseudo-R2)	221k (train), 269k (test)	7.6 (train), 8.5 (test)	Corrects for time correlation; residuals stable; statistically	More complex; harder to explain to non-technical users.

### Selected Model: Model 2 (Reduced Regression)

Delivers strong accuracy (Adj. R<sup>2</sup>≈0.87, MAPE 7–8%) with fewer variables, making it simpler and more interpretable. Mild statistical issues are not critical. Easier to explain than GLS while capturing the same main drivers → selected model for storytelling.

# Model Validation

## Selected Model: Reduced Multiple Regression (Model 2)

### Performance Metrics

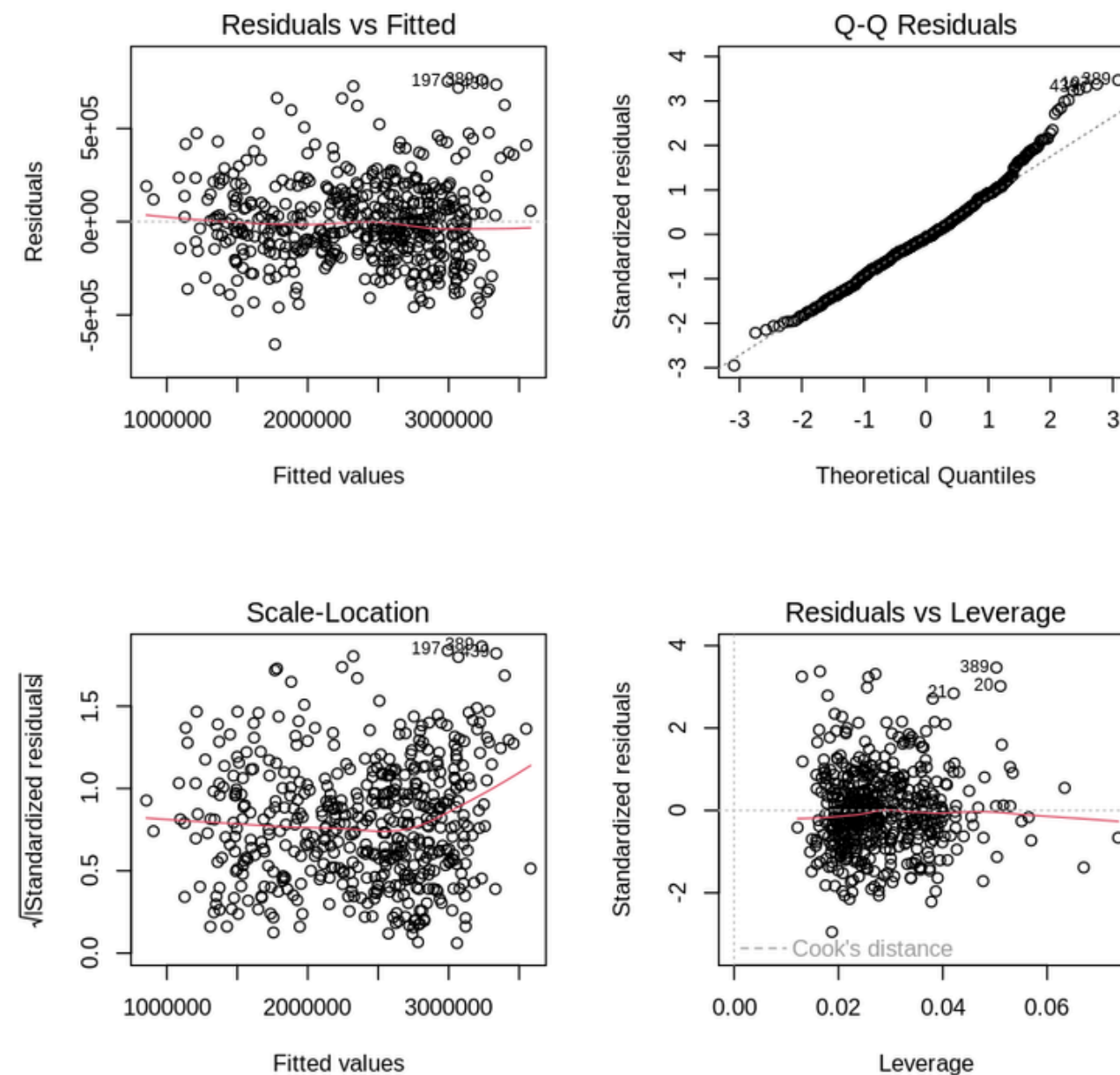
- Adjusted  $R^2 \approx 0.87 \rightarrow$  Model explains ~87% of sales variation
- RMSE  $\approx 222\text{k USD} \rightarrow$  Typical prediction error is about \$220K vs. average weekly sales of \$2.8M (~8%)
- MAPE  $\approx 7\text{--}8\% \rightarrow$  Predictions are on average within  $\pm 8\%$  of actual sales

### Residual Diagnostics

- Residuals centered  $\rightarrow$  no systematic bias
- Q-Q plot  $\rightarrow$  residuals close to normal, small deviations at extremes
- Mild heteroscedasticity  $\rightarrow$  slightly more error at higher sales levels, but acceptable
- Few high-leverage points  $\rightarrow$  no major influence on results

### Summary

- The model has reliable accuracy with interpretable results
- Errors are within a reasonable range
- The model is considered for best balance between accuracy & simplicity



# Model Insights: What Drives Sales?

Variable	Coefficient	Impact	Business Interpretation	Relative Importance
Intercept	1,448,301	— Baseline	Average weekly sales when other factors are neutral	—
Competitor A Spend	−261,730	↓ Negative	Aggressive competitor activity reduces sales	Very High
Regions (E, N, S)	+792k to +1.78M	↑ Positive	East & South outperform West (baseline)	Very High
Competitor B Spend	−89,900	↓ Negative	Competitor B pressure strongly cuts sales	High
Search Clicks	73,043	↑ Positive	Strongest digital driver, captures active demand	High
Inflation	−79,930	↓ Negative	Higher inflation reduces consumer spending	Medium
Banner Impressions	62,96	↑ Positive	Display ads support steady sales baseline	Medium
OLV Spend (lag1)	43,025	↑ Positive	Video ads build awareness, effect appears later	Medium
OOH Impressions	−58,789	↓ Negative	Outdoor campaigns less efficient / misaligned	Medium
Trend	23,888	↑ Positive	Underlying growth over time	Medium
TV Investment (lag1)	−42,139	↓ Negative	TV effect delayed, not aligned with sales cycles	Low–Medium

## Principal Model Insights

### Macroeconomics

- Inflation (−): Higher inflation reduces sales → confirms consumer sensitivity to price pressures.

### Media Effectiveness

- Search Clicks ( ↑ ): Strongest digital driver; critical conversion channel.
- Banner Impressions ( ↑ ): Steady support; cost-effective baseline channel.
- OLV Spend (Lag ↑ ): Works as brand awareness; delayed positive effect, not immediate.
- OOH Impressions (−): Inefficient or misallocated; should be reassessed.
- TV Investment (Lag −): Misaligned timing; coordination with other media may improve results.

### Competitors

- Competitor Spend (−): Strong negative impact; rival’s campaigns reduce CoreMotion’s sales.

### Time & Regions

- Trend ( ↑ ): Underlying growth trajectory in sales.
- Regions: East > South > North >>> West; clear regional imbalance.

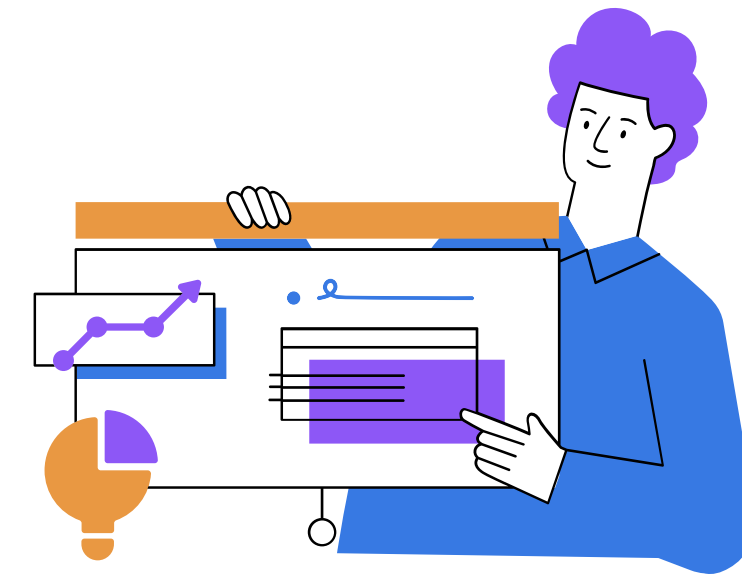
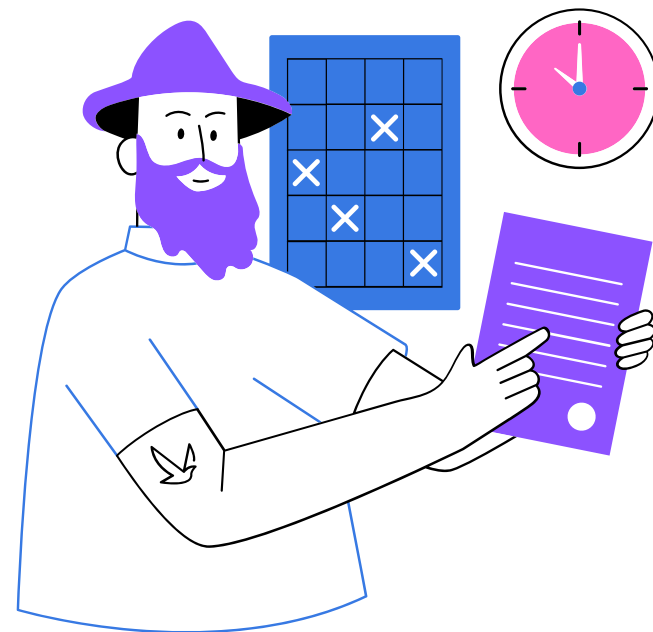
### Notes:

- *Not all variables predict sales directly — some have descriptive value, others risk biasing the model. We focused on the most robust and interpretable found predictors in the model.*
- *Coefficients are standardized — they show the impact of a 1 standard deviation change in each variable, allowing direct comparison of drivers’ relative importance.*



# Findings & Actionable Recommendations

- **Sales Drivers Identified:** Search clicks and banner impressions are the strongest positive contributors to sales. Inflation shows a strong negative effect.
- **Lagged Media Effects:** OLV (video ads) shows delayed impact, highlighting its role in long-term brand building rather than immediate sales.
- **Competitive Pressure:** Competitor spend (especially Competitor B) has a strong negative impact, confirming the importance of monitoring market dynamics.
- **Regional Gaps:** East and South dominate sales, while West lags significantly — resources should consider regional behavior
- **OOH & TV Efficiency Issues:** OOH impressions show inefficiency, and TV requires better timing/coordination.
- Other Business KPI's can be explored in Power Bi Dashboard.



- **Reinforce High ROI Channels:** Prioritize search and banner campaigns as cost-effective sales drivers.
- **Reposition OLV:** Treat OLV as a long-term awareness channel, not a direct sales driver.
- **Re-evaluate OOH:** Optimize or reallocate budget from OOH into more efficient channels.
- **Improve TV Strategy:** Adjust timing and integrate with other media campaigns to maximize effectiveness.
- **Regional Strategy:** Focus incremental investments in East and South, while experimenting with tailored strategies to improve performance in West.

# Thanks!

- 3002596060
- lobitocc7@gmail.com

