COMPANY FINANCIALS FOR PREDICTIVE INSIGHTS AND ECONOMIC TRENDS.

Team Members :

Gangili Sai Charan(A20543155) Poojith Reddy Annachedu(A20539719)
Neeraj Vardhan Buneeti (A20545853)

INTRODUCTION AND RESEARCH GOAL

Today's market is characterized by rapid changes, making accurate financial predictions crucial for success.

This project utilizes the Company Financials Dataset to identify key indicators for forecasting future profitability and creating predictive models .

The primary objective is to identify key financial indicators that can accurately forecast a business's potential for expansion and success.

By leveraging sophisticated machine learning techniques and statistical analysis on historical data, this research aims to improve predictive models that stakeholders can use to foresee financial trends and make informed decisions.

FUTURE SCOPE

- Model Refinement: Future efforts will focus on refining the predictive models to enhance their accuracy and robustness. This includes incorporating additional features, exploring advanced machine learning algorithms, and optimizing model hyperparameters.
- Real-time Data Integration: Integrating real-time financial data will enhance the responsiveness of the models to market dynamics, allowing for more timely and informed decision-making.
- Industry-specific Analysis: Further analysis will be conducted to develop industry-specific models, considering unique economic patterns and disruptions that affect different sectors.

DATA PIPELINE

- Dataset Details: The dataset used in this project is sourced from Kaggle, specifically the Company Financials Dataset, available at https://www.kaggle.com/datasets/atharvaarya25/financials/data.
- Loading the Dataset: The dataset is loaded using the fread function from the data.table package, enabling faster loading of large datasets.
- Below are the first 10 rows of the Financials dataset..

Segment	Country	Product	Discount_	Units_Sold	Manufactu	Sale_Price	Gross_Sal	Discounts	Sales	COGS	Profit	Date	Month_Nu	Month_Na	Year
Governme	Canada	Carretera	None	$1,618.50	$3.00	$20.00	$32,370.0	$-	$32,370.0	$16,185.0	$16,185.0	01-01-2014	1	January	2014
Governme	Germany	Carretera	None	$1,321.00	$3.00	$20.00	$26,420.0	$-	$26,420.0	$13,210.0	$13,210.0	01-01-2014	1	January	2014
Midmarket	France	Carretera	None	$2,178.00	$3.00	$15.00	$32,670.0	$-	$32,670.0	$21,780.0	$10,890.0	01-06-2014	6	June	2014
Midmarket	Germany	Carretera	None	$888.00	$3.00	$15.00	$13,320.0	$-	$13,320.0	$8,880.00	$4,440.00	01-06-2014	6	June	2014
Midmarket	Mexico	Carretera	None	$2,470.00	$3.00	$15.00	$37,050.0	$-	$37,050.0	$24,700.0	$12,350.0	01-06-2014	6	June	2014
Governme	Germany	Carretera	None	$1,513.00	$3.00	$350.00	$5,29,550	$-	$5,29,550	$3,93,380	$1,36,170	01-12-2014	12	Decembei	2014
Midmarket	Germany	Montana	None	$921.00	$5.00	$15.00	$13,815.0	$-	$13,815.0	$9,210.00	$4,605.00	01-03-2014	3	March	2014
Channel Pa	Canada	Montana	None	$2,518.00	$5.00	$12.00	$30,216.0	$-	$30,216.0	$7,554.00	$22,662.0	01-06-2014	6	June	2014
Governme	France	Montana	None	$1,899.00	$5.00	$20.00	$37,980.0	$-	$37,980.0	$18,990.0	$18,990.0	01-06-2014	6	June	2014

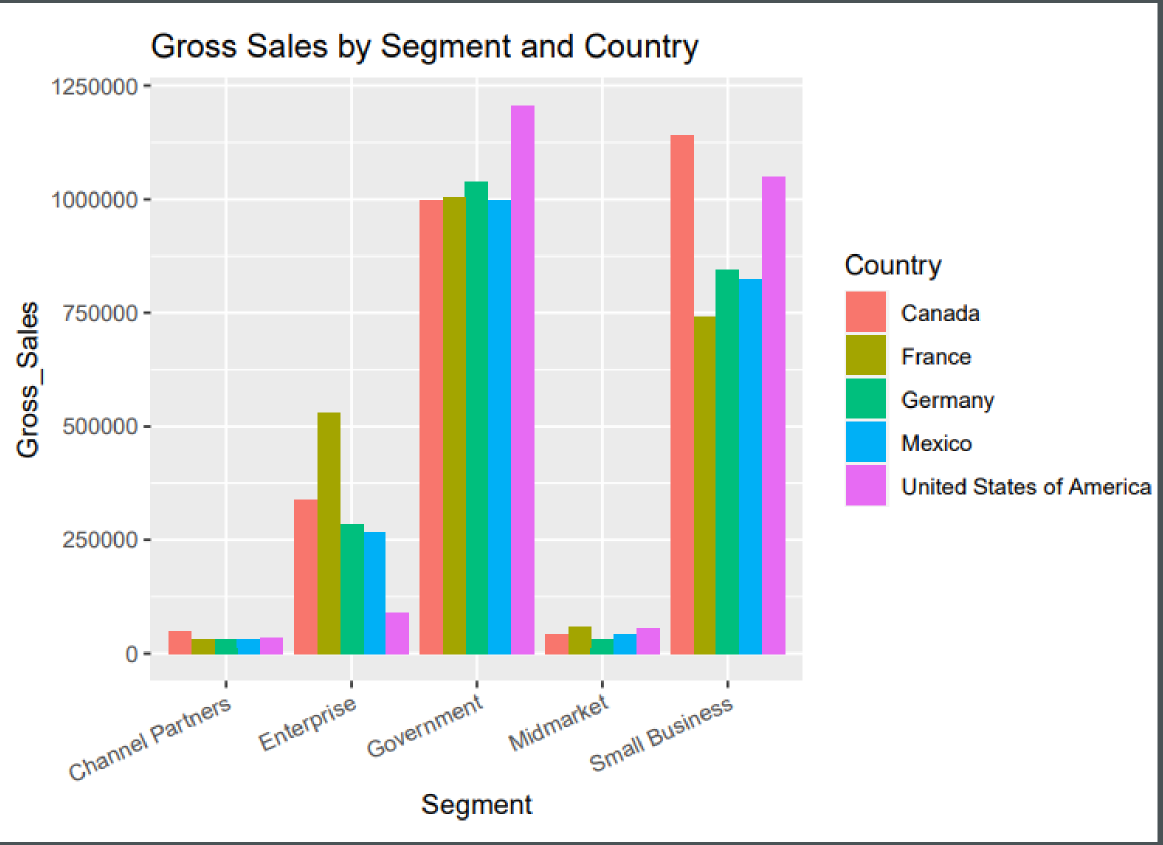
DATA PREPROCESSING

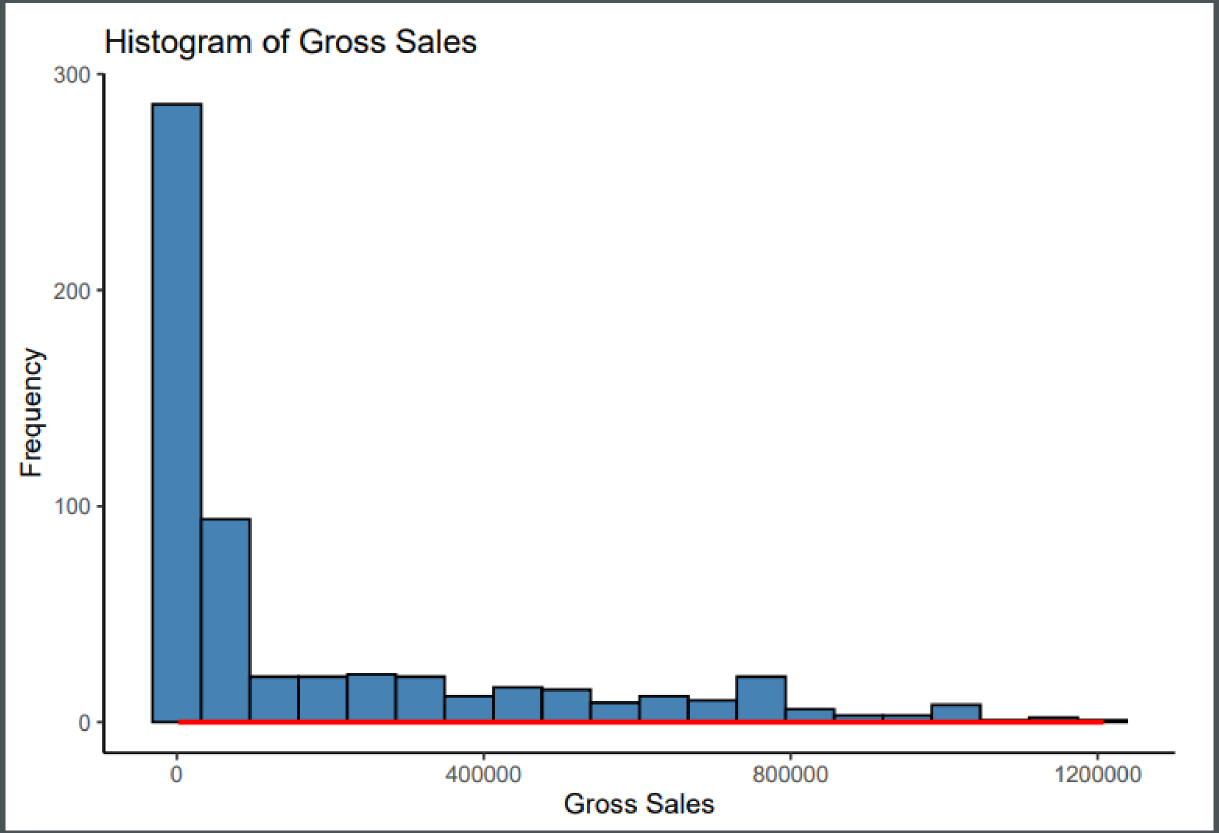
- Cleaning Numeric Columns: Numeric columns are processed to remove unwanted characters like commas and spaces.
- Handling Missing Values: "None" values in the Discount_Band column are converted to NA and rows with NA values are removed.
- Imputing Missing Values: Missing values in the Discount_Band column are imputed with the mode value.
- Ensuring Data Consistency: Levels are set for Discount_Band and it is converted to a factor.
- Summary and Validation: Checks are performed for NA, NaN, and infinite values in key columns to ensure data integrity.
- Final Dataset: The cleaned dataset is ready for analysis

FINAL DATASET

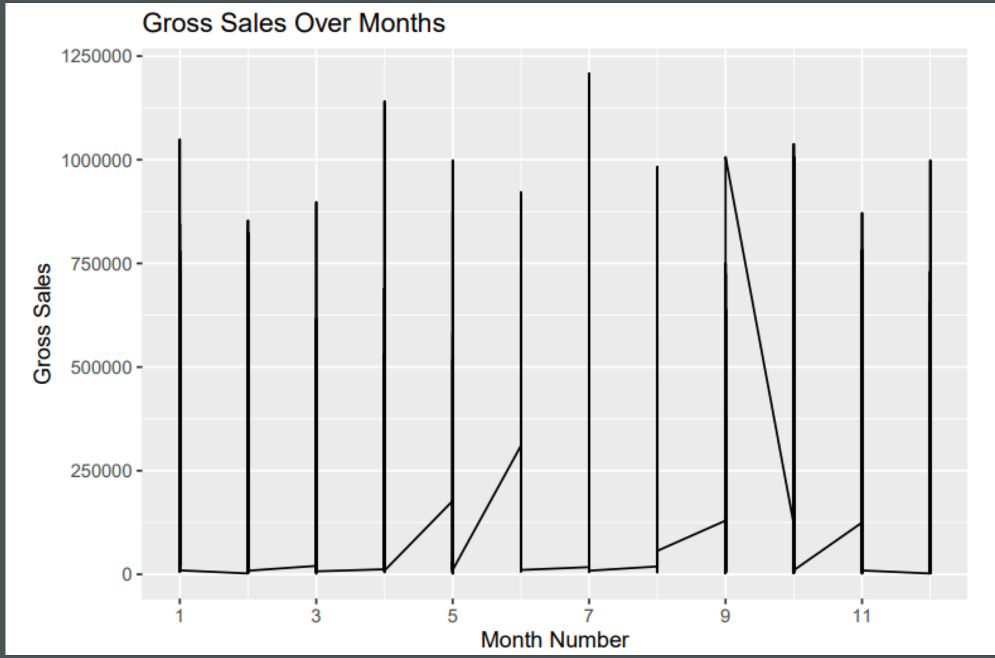
Segment	Country	Product	Discount_Band	Units_Sold	Manufacturing_Price	Sale_Price	Gross_Sales	Discounts	Sales	COGS
<chr>	<chr>	<chr>	<fctr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
Government	France	Paseo	Low	3945.0	10	7	27615.0	276.15	27338.85	19725.0
Midmarket	France	Paseo	Low	2296.0	10	15	34440.0	344.40	34095.60	22960.0
Government	France	Paseo	Low	1030.0	10	7	7210.0	72.10	7137.90	5150.0
Government	France	Velo	Low	639.0	120	7	4473.0	44.73	4428.27	3195.0
Government	Canada	VTT	Low	1326.0	250	7	9282.0	92.82	9189.18	6630.0
Channel Partners	United States of America	Carretera	Low	1858.0	3	12	22296.0	222.96	22073.04	5574.0
Government	Mexico	Carretera	Low	1210.0	3	350	423500.0	4235.00	419265.00	314600.0
Government	United States of America	Carretera	Low	2529.0	3	7	17703.0	177.03	17525.97	12645.0

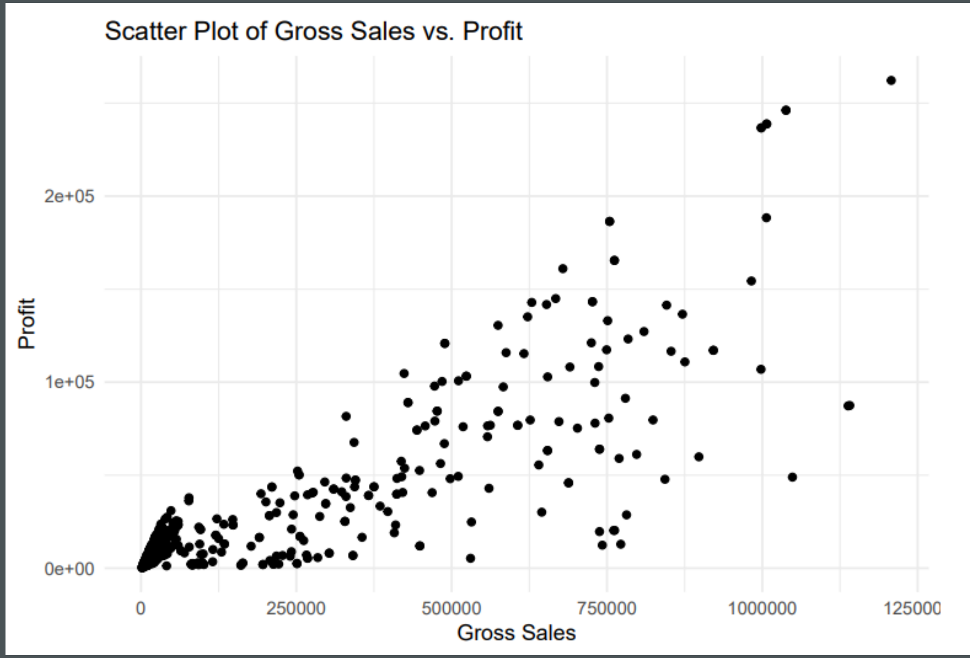
DATA VISUALIZATION



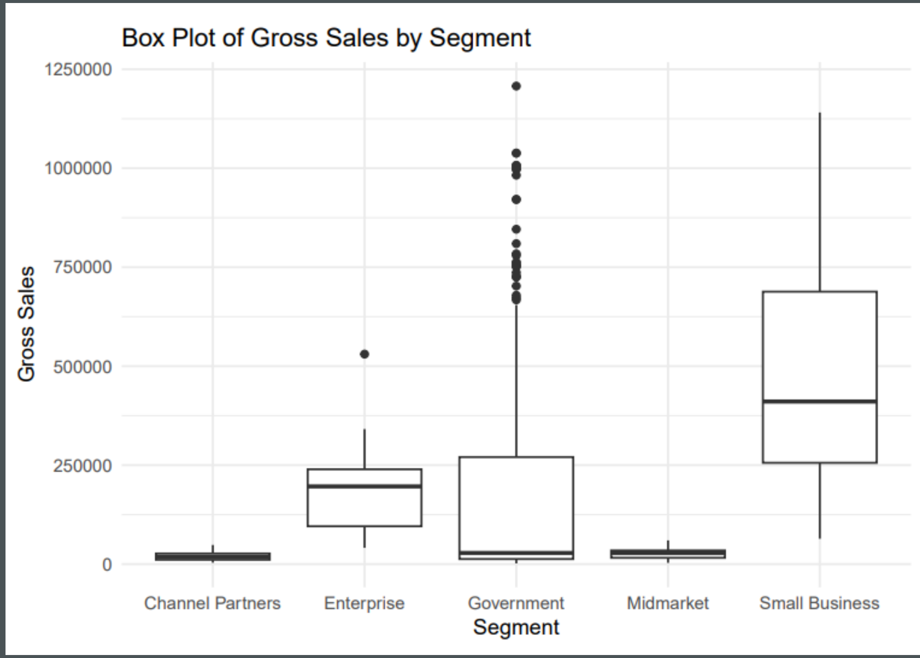


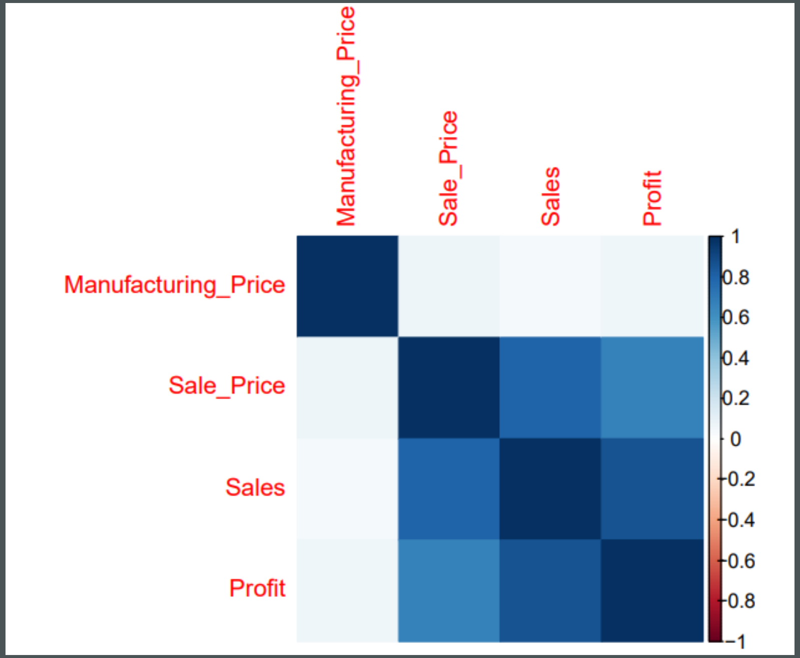
DATA VISUALIZATION





DATA VISUALIZATION





HYPER PARAMETER TUNING

- Hyperparameter are tuned using an iterative process of either:
- Validation
- Cross-Validation
- Training, Validation and Test Sets.
- Training Set - Data set used to learn the optimal model parameters
- Validation Set - Data set used to perform model selection
- Test Set – Data set used to assess the fully trained model

COMPARING OF RESULT

	Without Hyper tuning	With Hyper tuning
Linear Regression	9.36154412732311e-07	9.36154465761407e-07
KNN	517698159.815887	108967409.627813
Navies Bayes	1548398233.49873	-
SVM	54650290.9365748	642180448.014468
Decision Tree	156115495.460309	155115495.460309
Random Forest	80394140.4680553	-

CONCLUSION

- Hyperparameter tuning plays a crucial role in optimizing the performance of machine learning algorithms.
- The impact of hyperparameter tuning varies across different models.
- For some models like KNN and SVM, hyperparameter tuning leads to significant improvements in performance, highlighting the importance of fine-tuning these parameters.
- On the other hand, models such as Linear Regression and Decision Tree may show minimal to no improvement with hyperparameter tuning, suggesting that default parameters or inherent simplicity already provide near-optimal results.

CONCLUSION

- Understanding the effect of hyperparameter tuning on specific models is vital for efficient model selection and deployment.
- While hyperparameter tuning can enhance model performance, it's essential to consider the computational cost and potential overfitting.
- Future research could explore advanced hyperparameter optimization techniques and their impact on a wider range of machine learning algorithms.
- Overall, the judicious selection and tuning of hyperparameters are essential for building accurate and robust machine learning models.

PROJECT REPOSITORY LINKS

- Kindly refer to this location for all code files and documentation.
- Link :
  https://drive.google.com/drive/folders/17FLJ4N_NFPNtBk6VGagocSRY2gJp_OgS?usp=sharing