Sales Forecasting for Inventory Management

Tirthraj Bhatt-MS.Data Science '23

Mentor : Dr. Christelle Scharff

Pace University, Seidenberg School of CSIS

Github Qr

PACE UNIVERSITY

Abstract

The Sales Forecasting Project addresses the critical challenge of predicting sales quantity for effective inventory management. Utilizing Python and machine learning, the project offers a streamlined solution to enhance decision-making processes for businesses. Key steps include data exploration, cleaning, preprocessing, and model development, with a focus on deriving actionable insights. By leveraging Random Forest Regression and optimizing hyperparameters, the project seeks to provide businesses with a reliable tool for proactive sales forecasting. The comprehensive yet accessible documentation serves as a valuable resource for organizations aiming to implement data-driven strategies and elevate their operational efficiency.

Research Question

- Can we accurately predict sales quantity for each category in a business's inventory for the upcoming quarter and year?

  Forecasting: Develop accurate models to predict future revenue based on past sales data. This empowers our organization with the foresight needed to make informed strategic decisions.

DataSet

Data Source :- https://data.world/vineet/salesdata

Total Fields :- 15 Total Records :- 34,876

Crawled on :- 2017

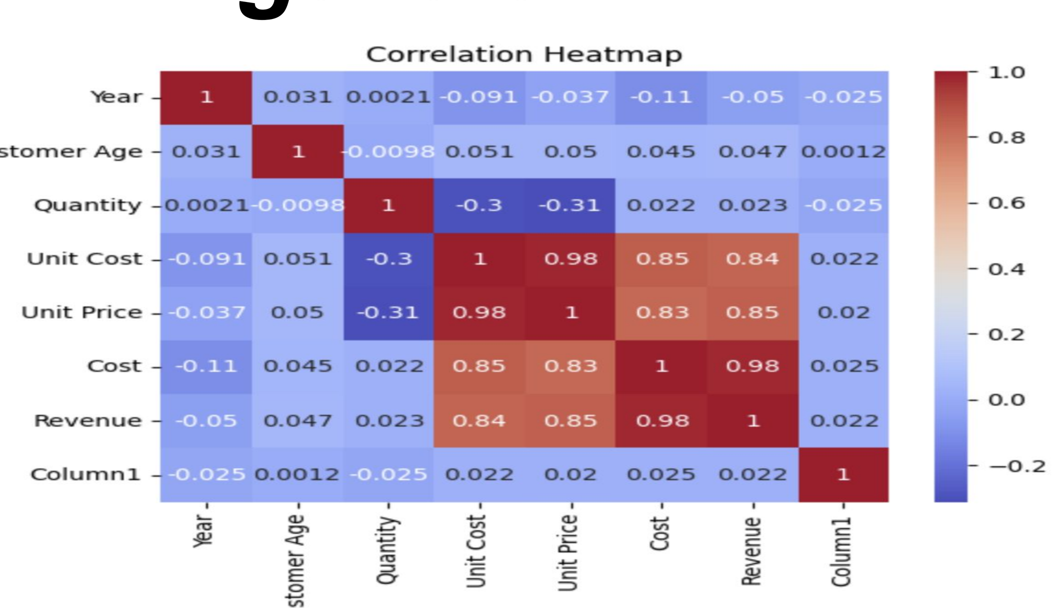
- Date: Transaction Date
- Year: Transaction Year
- Month: Transaction Month
- Customer Age: Age of Customer
- Customer Gender: Gender of Customer
- Country: Transaction Country
- State: Transaction State
- Product Category: Category of Product
- Sub Category: Subcategory of Product
- Quantity: Quantity Sold
- Unit Cost: Cost per Unit
- Unit Price: Selling Price per Unit
- Cost: Total Cost
- Revenue: Total Revenue

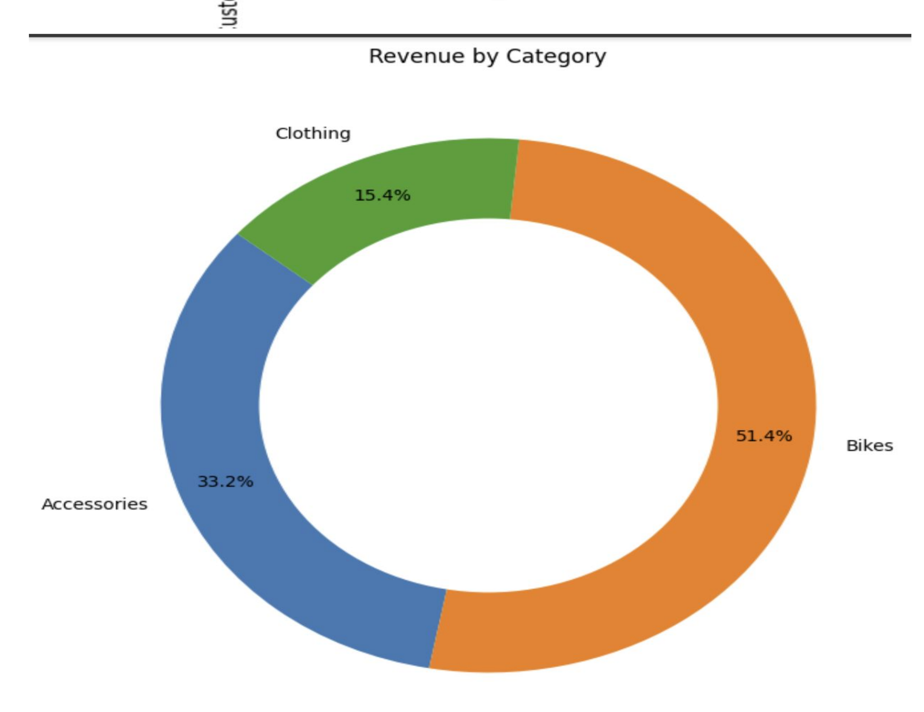
Methods

In our Exploratory Data Analysis (EDA) phase, we began by visualizing the distribution of key features, unraveling insights into customer demographics and purchase patterns. The process involved creating diverse visualizations, including bar charts for categorical features and correlation matrices to understand relationships.

Data preprocessing played a pivotal role in refining our dataset. This involved handling missing values, encoding categorical variables, and transforming the 'Revenue' column through a log transformation due to its positively skewed distribution. Outliers were addressed using the Interquartile Range (IQR) method. Additionally, we introduced feature scaling using StandardScaler to standardize the numerical features, paving the way for a more robust machine learning model To ensure a comprehensive analysis, we categorized customers into age groups, allowing for a nuanced understanding of their preferences. Leveraging label encoding, we converted categorical variables into a numerical format, facilitating seamless integration into machine learning models.

Figure of EDA



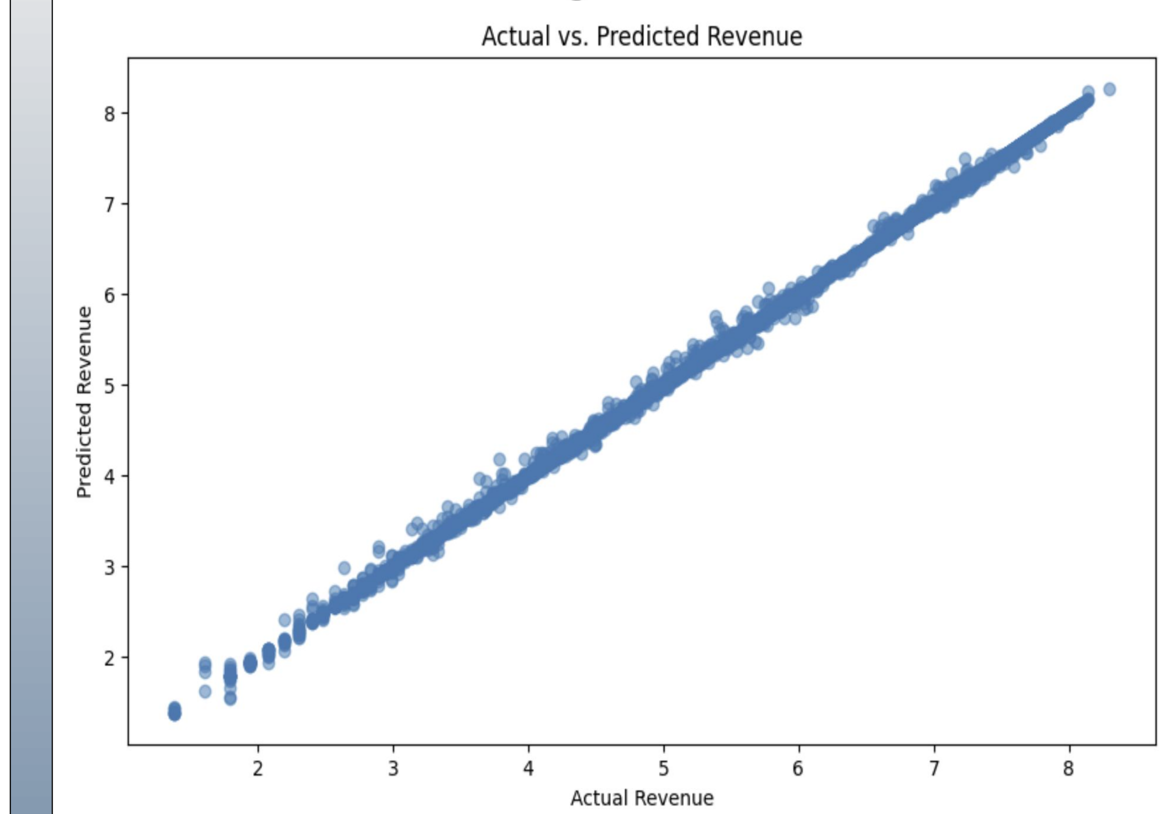


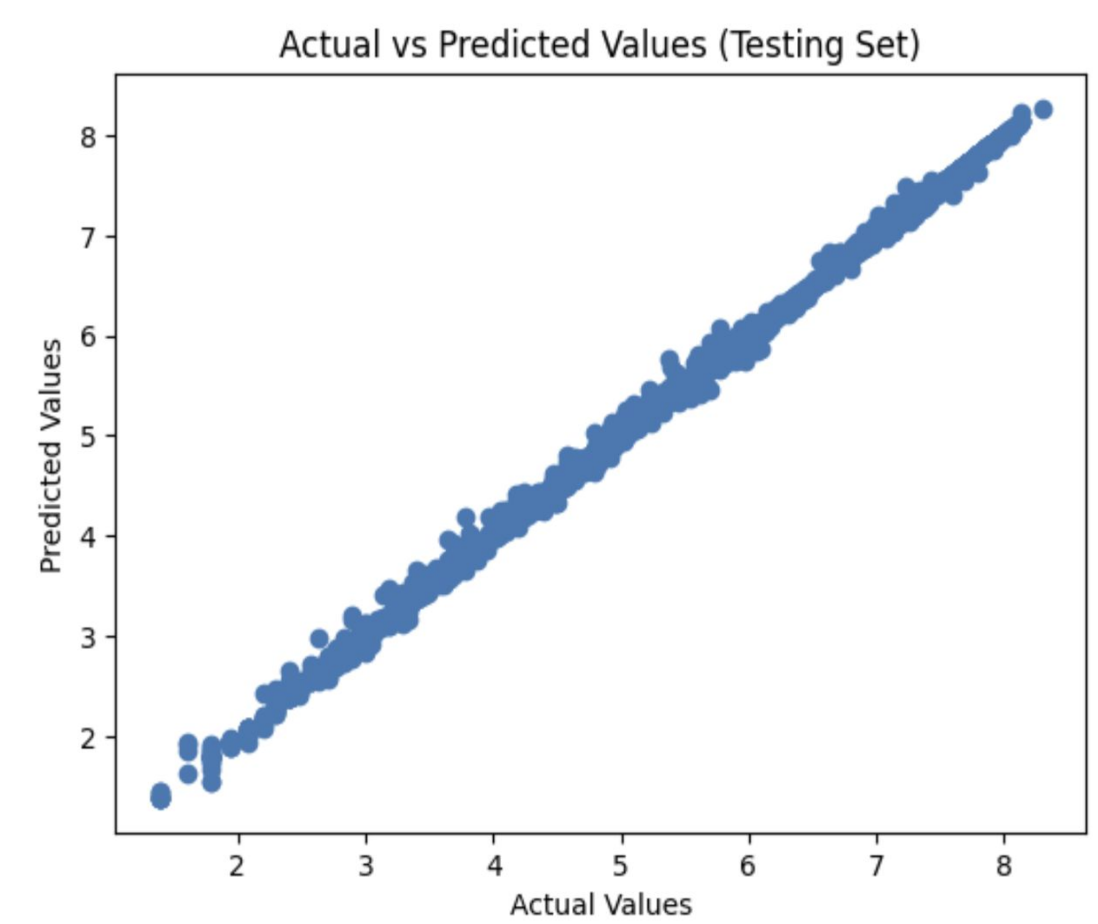
Results

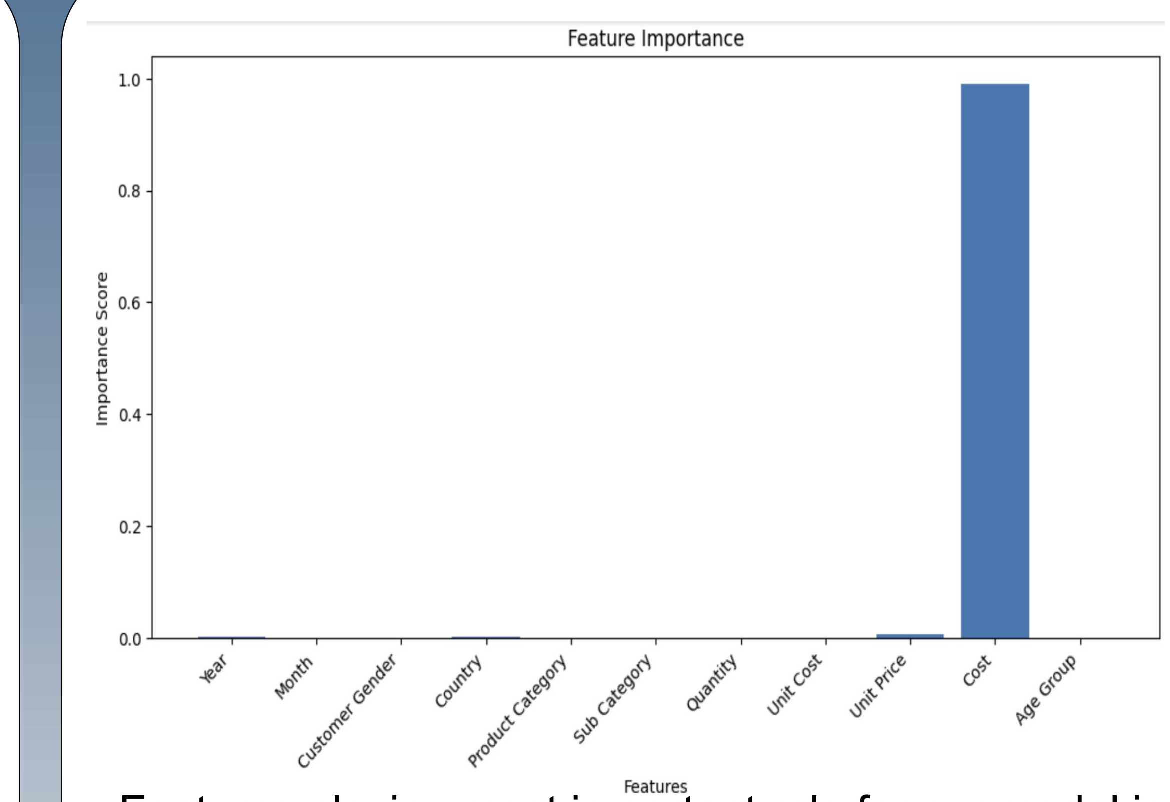
Our data-driven endeavor yielded remarkable results, with our revenue forecasting model exhibiting exceptional predictive accuracy. The Random Forest Regressor, finely tuned to our sales dataset, demonstrated outstanding performance metrics. In the training set, we achieved an extraordinarily low Mean Squared Error (MSE) of 0.0002, a Mean Absolute Error (MAE) of 0.0049, and an R-squared (R2) of 0.9999.

Cross-validation further affirmed the model's robustness, boasting consistently high R-squared scores, emphasizing its generalization capability. Post-tuning, our model maintained its stellar performance on the test set, with an MSE of 0.0011, MAE of 0.0130, and an R2 of 0.9995. These metrics collectively underscore the efficacy of our predictive model, positioning it as a reliable tool for revenue forecasting.

Figure







Features playing most important role for our model is cost,Unit price, Country and Year Respectively

Conclusions & Future work

In conclusion, our comprehensive data analysis and forecasting model have laid a solid foundation for precise revenue predictions. Leveraging sophisticated techniques like Random Forest Regression and meticulous data preprocessing, we have successfully developed a robust model. The exceptionally low MSE, MAE, and high R-squared scores obtained during training and cross-validation highlight the model's accuracy and ability to generalize.

In future iterations, enhancing the forecasting model could involve exploring additional features, incorporating external factors like economic indicators or seasonal trends, and optimizing model for further accuracy. Additionally, implementing a real-time updating mechanism based on incoming data would contribute to a more dynamic and responsive forecasting system.

Sources

1. Application of Machine Learning Model and Hybrid Model in Retail Sales Forecast. (n.d.). IEEE Xplore. Retrieved from

Authors :- Haichen Jiang, Jiatong Ruan, Jianmin Sun

Link:- [Application of Machine Learning Model and Hybrid Model in Retail Sales Forecast | IEEE Conference Publication | IEEE Xplore]

2. Time-series forecasting of seasonal items sales using machine learning – A comparative analysis

Authors :- Yasaman Ensafi , Saman Hassanzadeh Amin, Guoqing Zhang , Bharat Shah

Link:- https://www.sciencedirect.com/science/article/pii/S2667096822000027#sec0029