|  |
| --- |
| **TECHNICAL REPORT – SUPERSTORE SALES FORECASTING** |

|  |
| --- |
| **1)MAIN OBJECTIVE** |
| **Forecasting furniture sales for the next 3 years, based on time-series data of 4 years.** |

|  |  |
| --- | --- |
| **2)PROJECT OVERVIEW** | |
| **Forecasting Method** | **SARIMA** |
| **Forecasting Period** | **3 years** |
| **Training Period** | **3 years** |
| **Testing Period** | **1 year** |
| **Level of Granularity (Daily/Monthly)** | **Weekly** |
| **Confidence Level** | **56.73%** |
| **Accuracy Metric/Metrics Used** | **RMSE and R2 Score** |
| **Accuracy Result** | **RMSE = 1906.2, R2\_Score = 56.7%** |
| **Notes** |  |

|  |  |  |
| --- | --- | --- |
| **3)EXPLORATORY DATA ANALYSIS** | | |
| **NUMERICAL EXPLORATION** | | |
| **Data Cleaning** | **Outliers Identified** | **None** |
| **Errors Identified** | **None** |
| **Missing Values** | **Yes, 565 days has no furniture sales** |
| **Data Transformation** | **Taking weekly sales instead of daily sales** |
| **Decisions Made Based on Numerical EDA** | **We can’t forecast daily sales directly** | |
| **We should take weekly sales** | |
|  | |
|  | |

|  |  |  |
| --- | --- | --- |
| **3)EXPLORATORY DATA ANALYSIS** | | |
| **GRAPHICAL EXPLORATION** | | |
| **Graph Used** | **Justification** | **Insight** |
| **2x Pie plot** | **To see how much (Data/Sales) belongs to furniture** | **21% of the data belong to furniture**  **32% of the total sales belong to furniture** |
| **boxplot** | **Identify outliers in sales column** | **No need to remove any outliers** |
| **Bar plot** | **To see sales vs profit for each year** | **Furniture profit relatively low** |
| **Scatter** | **To see missing (dates/days) in the daily sales** | **565 missing (dates/days)** |
| **X2 barplot** | **To see quarterly/monthly furniture sales over the 4 years** | **Quarterly increasing**  **Months like (9, 11, and 12) have relatively high sales** |
| **Lineplot** | **To see trend in the data** | **Data has increasingly trend** |
| **plot\_acf** | **To identify highest correlation** | **Relatively weak correlation** |
| **lag\_plot** | **To identify how strong is the correlation** | **Relatively weak correlation** |
| **X8 line plot** | **To plot prediction results** | **-** |
| **X2 barplot** | **To compare evaluation matrices (performance)** |  |
|  |  |  |
| **Notes** | | |
|  | | |

|  |  |
| --- | --- |
| **4)NAÏVE MODEL** | |
| **Forecasting Method/Methods Tested** | **Persistence model** |
| **Linear Regression with lag=1** |
|  |
| **Accuracy Metric Used & Result for each Model Tested** | **Persistence model: RMSE=** **3149, R2\_Score=** **-0.048** |
| **Linear Regression: RMSE=** **3140.5, R2\_Score=** **-0.0145** |
|  |

|  |  |
| --- | --- |
| **5)FINAL MODEL** | |
| **Forecasting Model Selected** | **SARIMA** |
| **Accuracy Metric Used & Result** | **RMSE=** **1906.243, R2\_Score=** **0.5673** |
| **Parameters Toning (changes made to parameters to reach the final model)** | **order=(1,0,26** |
| **seasonal\_order=(1,1,1,52)** |
| **trend='ct'** |
| **Seasonality Identified** | **Yearly** |