The Bakery Task

Reducing Food waste and enhancing product availability in the Bakery

1. Introduction:

This report presents an analysis and forecasting model aimed at helping a bakery in France optimize its daily production. The goal is to reduce waste and ensure that customers always have access to fresh products. Using 'bakery_sales.csv' and 'weather_2021.01.01-2022.10.31.csv', an accurate forecasting model is developed by combining sales and weather data.

2. Data Analysis and Preparation:

The sales dataset comprises approximately 23,000 records, while the weather dataset includes about 600 records, covering the period from January 1, 2021, to October 31, 2022. These datasets were merged based on the 'date' column using an inner join.

Sales Data:
- 'Ticket number' and 'untitled' columns are removed as they do not have relevance
- Rows with invalid or missing 'Article' values and negative 'Quantity' values were removed.
- The 'UNIT_PRICE' column was converted to float after removing currency symbols.
- New columns were created: 'total' (total sales per day), 'hour' (hour of the day), 'day_of_week' (day of the week), and month_code (month of the year).

Weather Data:

- 'tsun' column was dropped as it mostly contained null values
- NaN values were replaced with 0 in 'snow' and 'wpgt' columns

Merging Data:

- The sales data was aggregated by date to sum total sales and quantities.
- The aggregated sales data was then merged with the weather data on the 'date' column leaving us with 600 rows.

3. Exploratory Data Analysis:
- **Sales Trends Over Day and Time:**
  Sales data revealed noticeable fluctuations and trends over the period. Significant sales peaks were observed during weekends, particularly on Sundays, and around lunchtime, specifically at 1 PM and 2 PM. Analysis showed a clear pattern of higher sales during weekends. This aligns with the bakery's understanding that weekends, especially Sundays, are busier due to the holiday and leisure shopping.
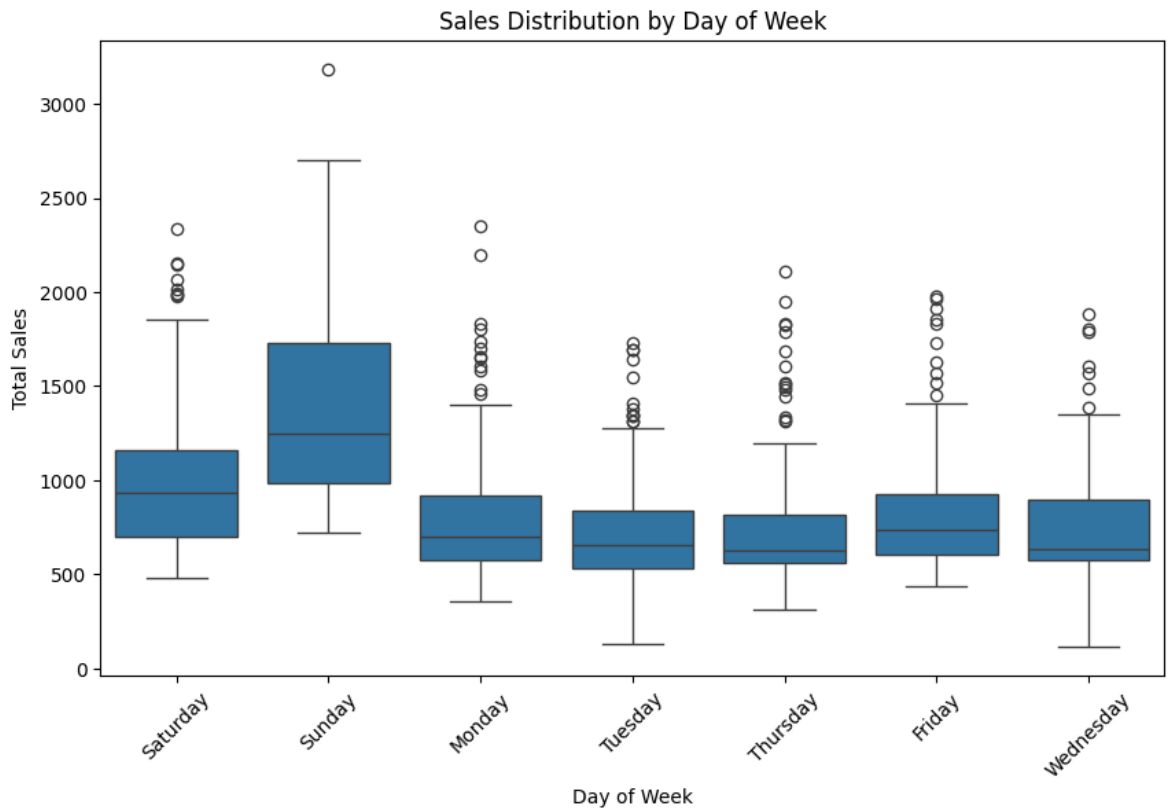
Figure 1: Total Sales Distribution by days

- **Temperature Impact:**
  There was a noticeable increase in sales on days with moderate to high temperatures. This could be due to more people being out and about on warmer days, leading to increased foot traffic at the bakery.
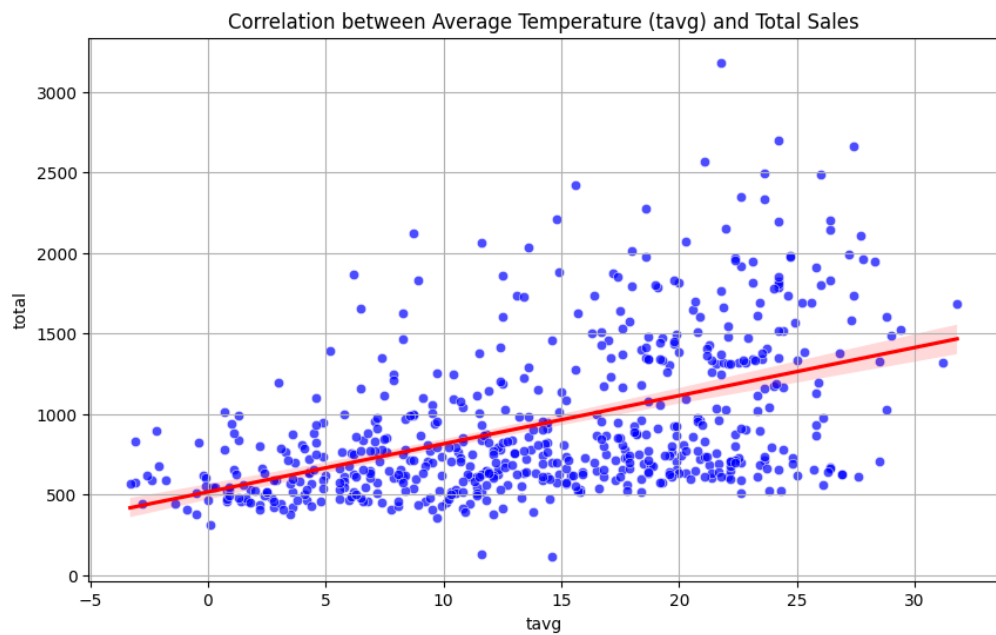

Figure 2: Mild Linear Correlation between average temperature and total sales

4. Models

Multiple forecasting models including Linear Regression, SARIMAX and LSTM were performed using the merged data of sales and weather datasets. The LSTM model performed the best, providing the following metrics on total sales forecasting:

R2 Score	0.979
Mean Squared Error (MSE)	8070.28
Mean Absolute Error (MAE)	70.49
Root Mean Squared Error (RMSE)	89.83

An R2 score of 0.979 indicates that the LSTM model explains 97.9% of the variance in sales, demonstrating high accuracy. The LSTM model's ability to retain long-term information makes it ideal for time series data. This model is also used to forecast product quantities, optimizing inventory management to meet customer demand while minimizing waste.
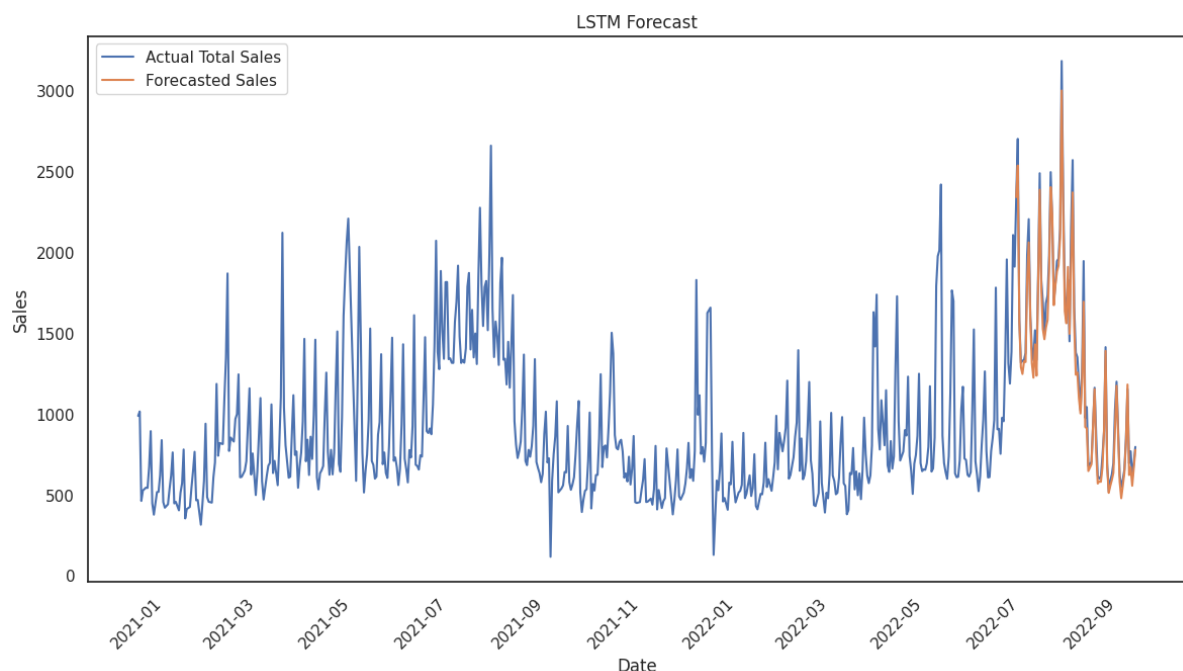

Figure 3: LSTM model test results on total sales forecast

5. Recommendations

1. **Adjust Production for Weekends and Warm Days**:
   o Increase production on weekends and days with higher temperatures to meet higher demand and reduce the risk of stockouts.
2. **Utilize LSTM Model for Daily Forecasting**:
   o Implement the LSTM model to forecast daily sales and adjust production based on weather forecasts to minimize waste and ensure product availability.

6. Conclusion

By integrating sales and weather data, the bakery can significantly enhance its production planning. The LSTM model provides a robust tool for forecasting daily sales, reducing waste, and ensuring customers have access to fresh products. Implementing these insights and

recommendations will lead to better inventory management and increased customer satisfaction.

7. Limitations

- **Data Gaps**: Some weather data points were missing (as it had only 600 records), which might slightly affect the accuracy of the model.
- **External Factors**: Factors such as local events (festivals/ holiday period) or promotions were not included in the analysis and could impact sales.

8. References:

1. https://keras.io/examples/timeseries/timeseries_weather_forecasting/