**SALES PREDICTION**

**REQUIREMENT FOR THE PROJECT**

1. Pandas: A powerful data manipulation and analysis library. It provides data structures for efficiently manipulating large datasets and tools for data cleaning, exploration, and analysis.
2. NumPy: A fundamental package for scientific computing with Python. It provides support for large, multi-dimensional arrays and matrices, along with mathematical functions to operate on these arrays.
3. matplotlib.pyplot: A plotting library in Python that is commonly used for creating static, animated, and interactive visualizations. It is part of the Matplotlib library, which provides a wide range of tools for creating various types of plots and charts. The pyplot module in Matplotlib provides a MATLAB-like interface for creating plots, making it easy to generate plots quickly.
4. Seaborn: A statistical data visualization library based on Matplotlib in Python. It provides a high-level interface for drawing attractive and informative statistical graphics. Seaborn is particularly well-suited for visualizing complex datasets with multiple variables. It comes with several built-in themes and color palettes to make it easy to create aesthetically pleasing and informative plots.

**STEPS TO EXECUTE THE PROJECT:**

1. Upload the data.
2. Handle missing values
3. Feature scaling based on general requirements
4. EDA using pandas profiling
5. Data cleaning
6. Preprocessing task before model building
7. Perform Random Forest Regression