COLLEGE CODE: 8107

COURSE: DATA ANALYTICS WITH COGNOS

PHASE II: INNOVATION SUBMISSION

TITLE: PRODUCT SALES ANALYSIS

TEAM MEMBERS:

- ARUL PALANIAPPA S
  8107243009
  arulpalaniappa.s@care.ac.in

- AZARUDEEN B 810721243011
  azarudeen.b@care.ac.in

- BALAJI C
  810721243012
  balaji.c@care.ac.in

- SARAVANAN V 810721243045
  saravanan.v@care.ac.in

- VIMAL M
  810721243057 vimal.m@care.ac.in

Project Title: Product Sales Analysis

INTRODUCTION:

The project involves using IBM Cognos to analyze sales data and extract insights about top selling products, peak sales periods, and customer preferences. The objective is to help businesses improve inventory management and marketing strategies by understanding sales trends and customer behavior. This project includes defining analysis objectives, collecting sales data, designing relevant visualizations in IBM Cognos, and deriving actionable insights.

INNOVATION:

1. Data Collection and Preparation:

Gather historical sales data, which should include information on sales volume, date/time, product details, pricing, marketing campaigns, and any other relevant factors. Clean and preprocess the data to handle missing values, outliers, and inconsistencies. Ensure that the data is in a format suitable for machine learning

2. Feature Engineering:

Create relevant features from the data that can help your machine learning model make accurate predictions. These features may include lagged sales, seasonality, holidays, and economic indicators.

3. Data Splitting:

Split your data into training, validation, and test sets. The training set is used to train the model, the validation set helps tune hyperparameters, and the test set evaluates the model's performance.

4. Selecting Machine Learning Algorithms:

Choose machine learning algorithms suitable for time-series forecasting. Common choices include:

➢ Linear Regression: Simple and interpretable, but may not capture complex patterns.
➢ Decision Trees and Random Forests: Effective for capturing non-linear relationships and feature importance.

➢ ARIMA (AutoRegressive Integrated Moving Average): A traditional time-series forecasting method.

➢ Prophet: Developed by Facebook, designed for forecasting with seasonality and holidays.

5. Model Training:

Train your selected models using the training data. Experiment with different algorithms and hyperparameters to find the best-performing model.

6. Model Evaluation:

Use the validation set to assess the model's performance. Common evaluation metrics for time-series forecasting include Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

7. Monitoring and Maintenance:

Continuously monitor the model's performance in the production environment. Update the model as needed to adapt to changing sales trends

DATASET: https://www.kaggle.com/datasets/ksabishek/product-sales-data

3. **VISUALIZATION STRATEGY**:

Python:

Matplotlib: A versatile 2D plotting library that offers a wide range of chart types and customization options.

Seaborn: Built on top of Matplotlib, Seaborn provides a high-level interface for creating attractive statistical visualizations.

Pandas Plotting: Pandas, a data manipulation library, offers built-in plotting capabilities that are convenient for quickly visualizing data.

Plotly: An interactive plotting library that supports a wide range of chart types, including interactive web-based visualizations.