NAAN MUDHALVAN

PHASE 5 PROJECT SUBMISSION

**PRODUCT SALES ANALYSIS**

**TEAM MEMBERS:**

PARVATHY A - 2021504533

PAVITHRAN E - 2021504535

SHAMEEM AHAMED S - 2021504542

THUSHIYANTH K - 2021504555

**PROBLEM DEFINITION:**

This project involves using IBM Cognos to analyse 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 behaviour. This project includes defining analysis objectives, collecting sales data, designing relevant visualizations in IBM Cognos, and deriving actionable insights.

**DATABASE LINK:**

[**https://www.kaggle.com/datasets/ksabishek/product-sales-data**](https://www.kaggle.com/datasets/ksabishek/product-sales-data)

**OBJECTIVES:**

* Analysing data related to product sales.
* Generating valuable insights from the data.
* Based on the insights, recommendations must be formulated to address issues and optimize sales and profitability.

**DATA COLLECTION AND PREPARATION:**

* The dataset we are going to work on for this project is taken from **Kaggle** and we have mentioned it above.
* We also gather other historical sales data, including information on **products, customers, prices, promotions, and time periods of a company,** and work on it.

**FEATURE SELECTION:**

We search for relevant features in which we are going to work on that could influence sales. (**Product attributes, pricing, customer demographics, seasonality, and marketing activities**).

**DATA SPLITTING:**

We will split the data into training and testing sets. The training set is used to train the machine learning models, while the testing set is used to evaluate the model's performance.

**DATA VISUALIZATION USING IBM COGNOS:**

IBM Cognos is a powerful business intelligence and data analytics platform that empowers organizations to create compelling data visualizations.

With Cognos, you can seamlessly connect to diverse data sources, model your data for analysis, and author reports.

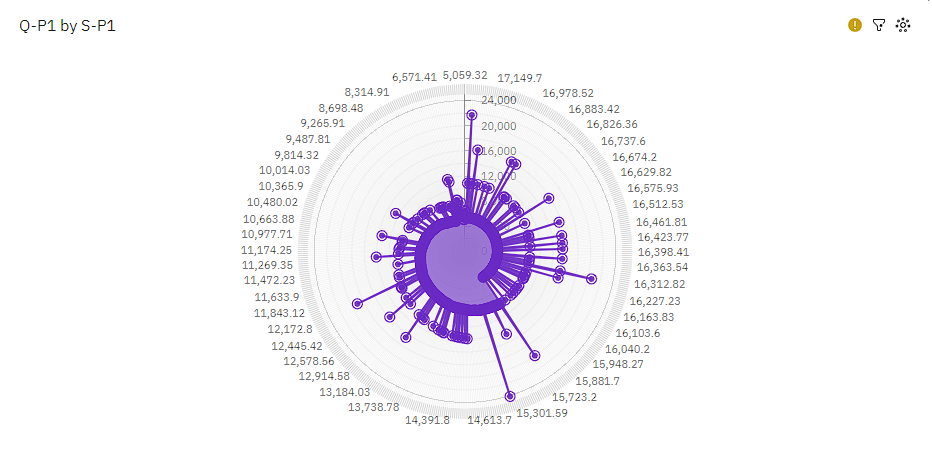
The platform offers an array of visualization types, including charts, tables, and maps, allowing you to choose the best representation for your data.

You can customize these visualizations, applying formatting, interactive features, and filters to enhance user experience and exploration.

IBM Cognos also supports the creation of interactive dashboards, consolidating multiple visualizations and reports into a single view for comprehensive insights.

Moreover, Cognos provides features for scheduling, security, and role-based access control to ensure that your visualizations are accessible to the right audience.

Its monitoring and analytical capabilities allow for continuous optimization, making it an invaluable tool for organizations looking to extract insights from their data and communicate them effectively to their stakeholders**.**

****

18601.56 Q-P1 at over 29 thousand is 81% higher than the Q-P4 of over 5500.

S-P1 15301.59 has the highest Total Q-P4 but is ranked #4 in Total Q-P1.

S-P1 18601.56 has the highest Total Q-P1 but is ranked #5 in Total Q-P4.

**Python code integration:**

**Random Forest:**

Random Forest is a robust and versatile ensemble learning algorithm in machine learning.

At its core, it leverages decision trees to make predictions. What sets Random Forest apart is its ability to combine the outputs of multiple decision trees to enhance predictive accuracy.

It does this through a technique known as bagging, where it creates several random bootstrap samples of the training data and trains individual decision trees on these subsets.

Additionally, Random Forest employs random feature selection for each split in the decision trees, promoting diversity and preventing overfitting.

**Ensemble learning algorithm:**

Ensemble methods typically consist of base learners, which can be simple models like decision trees or more complex ones, and they can be applied to various tasks, including classification, regression, and clustering.

The idea behind ensemble learning is that by aggregating the wisdom of diverse models, the collective result is often superior in terms of both accuracy and robustness.

One of the most popular ensemble techniques is the Random Forest, which combines multiple decision trees.

# Create and train a Random Forest Regressor model

model = RandomForestRegressor(n\_estimators=100, random\_state=42)

model.fit(X\_train, y\_train)

# Make predictions on the test set

y\_pred = model.predict(X\_test)

# Evaluate the model

mse = mean\_squared\_error(y\_test, y\_pred)

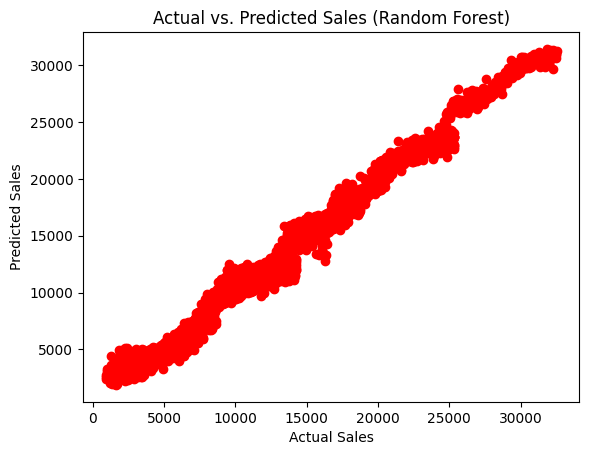
r2 = r2\_score(y\_test, y\_pred)

**Output:**

Mean Squared Error (MSE): 765893.1847328655

R-squared (R2) Score: 0.9718081814437257

Predicted Sales for New Data: [ 3787.7696 4384.2368 22258.6934 12844.4811]



**Linear Regression:**

Linear regression is a fundamental and widely used statistical and machine-learning technique that focuses on modelling the relationship between a dependent variable (the target) and one or more independent variables (predictors) by fitting a linear equation.

The goal of linear regression is to find the best-fit line that minimizes the sum of the squared differences between the predicted and actual values.

This line is characterized by a slope and an intercept and is often represented as y = mx + b.

Linear regression is commonly employed for both predictive modelling and understanding the relationships between variables.

# Create and train a Linear Regression model

model = Linear Regression()

model. Fit(X\_train, y\_train)

# Make predictions on the test set

y\_pred = model.predict(X\_test)

# Evaluate the model

mse = mean\_squared\_error(y\_test, y\_pred)

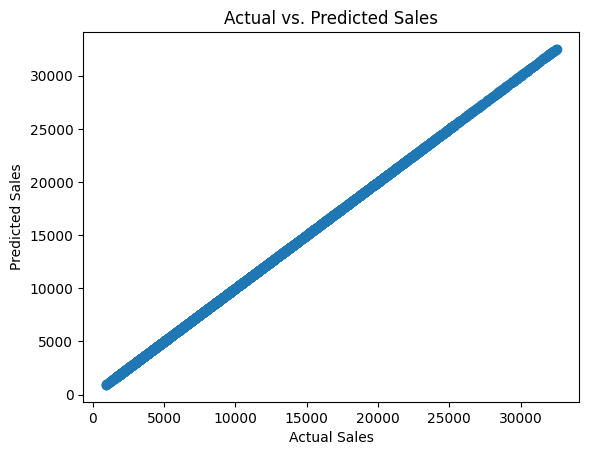
r2 = r2\_score(y\_test, y\_pred)

**Output:**

Mean Squared Error (MSE): 7.765710074562366e-22

R-squared (R2) Score: 1.0

Predicted Sales for New Data: [ 3170 4425.32 21680 24776.75]



**HOW THE INSIGHTS FROM THE ANALYSIS CAN HELP WEBSITE OWNERS IMPROVE USER EXPERIENCE:**

By analyzing historical sales data and predicting future sales, website owners can optimize various aspects of their online platforms.

Furthermore, accurate sales predictions enable dynamic pricing strategies, allowing website owners to offer competitive pricing and targeted discounts, which can lead to higher user satisfaction.

These models can also be used to customize marketing campaigns, ensuring that users receive personalized product recommendations and promotions that align with their interests.

Additionally, sales prediction models can help in load balancing and resource allocation, ensuring that the website remains responsive during peak traffic periods.

Overall, the ability to anticipate sales trends and customer behavior empowers website owners to create a more user-centric, efficient, and tailored experience, ultimately increasing customer loyalty and engagement.

**Conclusion:**

Thus, we have built a predictive model in Python using various machine learning algorithms that predict sales based on the important features, and different visualizations were carried out in IBM Cognos to analyse the import features.