

Project Report: Data Analysis and Visualization Using IBM Cognos

Team Members

Akshaya R

Ansh Bomb

Krithivarsha I

Anand Kumar

Janet Inba

Project Overview:

In this technology project, we embarked on a journey to load and preprocess a dataset, perform various analyses, and create data visualizations using IBM Cognos. The objective of this project was to gain insights from the dataset and communicate the results effectively.

The dataset to be used in our project : <https://www.kaggle.com/datasets/ksabishek/product-sales-data>

Data Preparation

Dataset Overview

We began our project by working with a dataset that contains daily sales and inventory data for a retail store. This dataset is a time series and includes the following columns:

- Date: The date of the recorded data.
- Quantity Sold for Products 1 to 4 (Q-P1, Q-P2, Q-P3, Q-P4): The quantity of each product sold.
- Sales for Products 1 to 4 (S-P1, S-P2, S-P3, S-P4): The total sales amount for each product.

The dataset consists of 4600 rows and 10 columns.

Import Libraries

We began by importing the necessary Python libraries to work with the dataset and perform data analysis and visualization. The libraries we used include:

- pandas # Library used for data manipulation and analysis
- numpy # Library used for working with arrays
- matplotlib.pyplot # Library for plots and visualizations
- seaborn # Library for visualizations

Data Loading

We used Python and the Pandas library to load the dataset into a DataFrame, allowing us to work with the data efficiently. We loaded the dataset from a CSV file using Pandas. The dataset provides information on daily sales and inventory data for a retail store. The columns in the dataset include Date, Quantity Sold for Products 1 to 4 (Q-P1, Q-P2, Q-P3, Q-P4), and Sales for Products 1 to 4 (S-P1, S-P2, S-P3, S-P4).

Observations:

- The dataset contains information from 13-06-2010 to 02-02-2023.
- We observed an "Unnamed: 0" column, which appears to be an identifier and is a repeat of the dataset index. We decided to drop this column to avoid redundancy.

Data Cleaning

To ensure data quality, we performed the following data cleaning tasks:

- Handling Missing Values: We filled missing values in numerical columns with their respective means to avoid data loss.
- Duplicate Removal: Duplicate records were identified and removed to maintain data integrity.
- Drop unnamed column: we decided to drop the "Unnamed: 0" column from the dataset, as it didn't provide additional information.
- Deriving additional time-based features such as month, day, day of the week, and year.
- Checked for duplicates (no duplicates found).

- Calculated basic statistical information for the dataset.
- Examined the correlation between columns, visualized using a heatmap.
- Checked unique values for each column to understand the data distribution.

Data Analysis and Visualization

Exploratory Data Analysis (EDA)

We conducted exploratory data analysis to better understand the dataset. Some of the key findings from our analysis include:

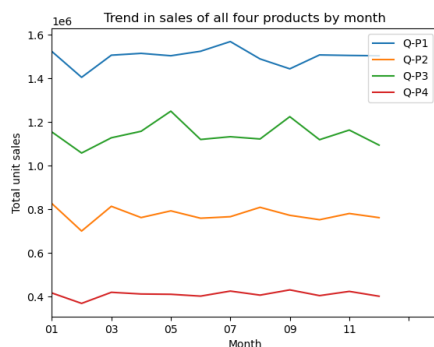
- Sales Trends: We observed that the sales data exhibit certain seasonal trends and patterns.
- Inventory Analysis: The inventory data helped us identify potential stockouts and overstock situations.

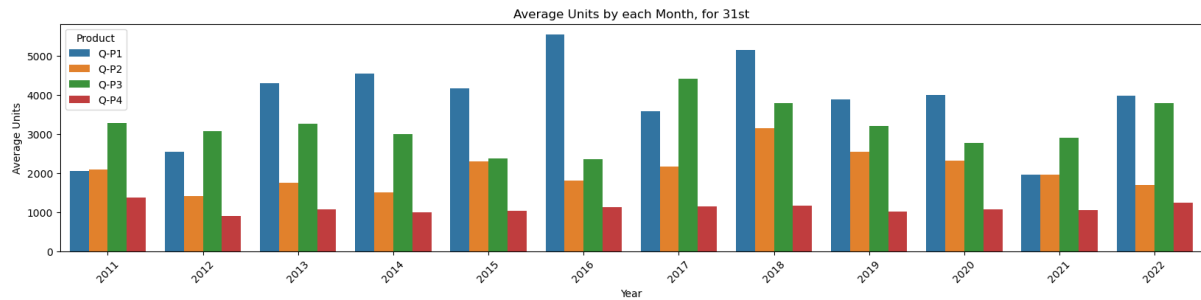
Data Visualizations

Using IBM Cognos, we created various data visualizations to enhance our understanding of the dataset. The visualizations included:

- Sales Trends Over Time: A line chart illustrating the daily sales trends for Products 1 to 4.
- Inventory Analysis: Bar charts showing the inventory levels and changes for different products over time.
- Product Sales Share: A pie chart representing the share of total sales for each product.

Below are examples of some of the visualizations created:





Insights and Conclusions

Based on our data analysis and visualizations, we arrived at the following insights and conclusions:

- **Seasonal Sales Patterns:** We observed clear seasonal sales patterns, with Products 2 and 4 showing the most significant fluctuations.
- **Inventory Optimization:** Inventory analysis highlighted the need for better inventory management, particularly for Product 3.
- **Product Sales Share:** Product 1 contributed the most to total sales, followed by Product 3.
- **October is the peak month** for unit sales and revenue, indicating a potentially profitable period for the business.
- **Fridays appear to generate the highest revenue**, followed by weekends (Saturdays and Sundays), which suggests a trend in consumer behavior.
- **The year 2016 stands out as the year with the highest revenue**, making it a notable year for the business.
- **The word clouds provide a visually appealing summary of total revenue and unit sales for each product**, with S-P3 being the highest in terms of revenue.

Conclusion

To conclude, this project seeks to harness the power of IBM Cognos Analytics to transform raw sales data into actionable insights. This project provided a valuable opportunity to work with data, perform analyses, and create data visualizations using IBM Cognos. It demonstrated the power of data analysis in uncovering meaningful insights that can be used for decision-making.