**Walmart Operations Optimization and Customer Experience Enhancement**

**Project Overview**

This project focuses on using advanced data analytics to optimize Walmart's operations and improve customer experience. The analysis includes identifying high-performing cities and branches, analyzing product category trends, examining payment preferences, and developing predictive and recommendation models. Additionally, an ETL pipeline was built to automate data extraction, transformation, and loading (ETL) into PostgreSQL for efficient data management and insights extraction.

**Objective**

* **High-Performing Cities & Branches**: Identified cities and branches that drive the most revenue and customer satisfaction.
* **Product Category Trends:** Analyzed product categories by sales, profit margins, and customer ratings.
* **Customer Payment Preferences:** Investigated correlations between payment methods and satisfaction levels (ratings).
* **Seasonal Sales Trends:** Analyzed sales trends over time to provide actionable inventory insights.
* **Predictive Model:** Developed a model for forecasting sales trends**.**
* **Recommendation Model**: Built a recommendation system for cross-selling and enhancing customer satisfaction.

**Workflow Breakdown**

**1. Data Cleaning**

* Missing Data: Checked for missing values in key fields like unit\_price, quantity, and rating, and handled them appropriately.
* Data Type Adjustments: Corrected data types, such as converting dates to datetime format.
* Outlier Handling: Identified and dealt with outliers in numerical columns like rating and profit\_margin.
* Data Standardization: Standardized string entries (e.g., City, Category) to maintain consistency.

**2. Exploratory Data Analysis (EDA)**

* **Segmentation Analysis:** 
  + Grouped data by branch and city to identify top performers.
  + Key metrics analyzed include sales, profits, and customer preferences by payment methods.
* **Correlation Analysis:** 
  + Explored relationships such as rating vs profit\_margin and unit\_price vs quantity.
* **Hypothesis Testing:** 
  + Example hypothesis: "Ewallet customers provide higher average ratings than Cash customers."
* **Visualizations:** 
  + Heatmap of sales by branch and product category.
  + Sales trendlines over time (daily, weekly, monthly).
  + Box plots for profit margins across branches.
  + Pie charts of payment method preferences.

**3. ETL Pipeline Implementation**

An ETL pipeline was built to automate the process of extracting data from the source, transforming it with Python, and loading it into a PostgreSQL database for efficient querying and analysis.

**1. Extract:**

* Data was extracted from CSV files, APIs, or other sources into Python using pandas for structured data manipulation.
* import pandas as pd
* data = pd.read\_csv('sales\_data.csv')

**2. Transform:**

* The data was cleaned and transformed within Python. This included handling missing values, converting data types, and performing feature engineering.
  + Missing Values Handling:
  + data.fillna(data.mean(), inplace=True)
  + Data Type Conversion:
  + data['sale\_date'] = pd.to\_datetime(data['sale\_date'])
  + Feature Engineering:
    - Created day\_of\_week, hour, and weekend flags from the sale\_date.
    - Calculated rolling averages for sales data.
  + data['day\_of\_week'] = data['sale\_date'].dt.dayofweek
  + data['hour'] = data['sale\_date'].dt.hour
  + data['is\_weekend'] = data['day\_of\_week'].isin([5, 6]).astype(int)

**3. Load:**

* The transformed data was loaded into PostgreSQL using the psycopg2 library.
* import psycopg2
* from sqlalchemy import create\_engine
* # Create engine for PostgreSQL connection
* engine = create\_engine('postgresql://username:password@localhost:5432/mydatabase')
* # Load data into PostgreSQL
* data.to\_sql('sales\_data', engine, index=False, if\_exists='replace')

**4. Advanced Insights and Business Recommendations**

**Insights:**

* Revenue Growth: Identified top-performing cities and branches contributing the most to revenue.
* Sales Trends: Analyzed monthly and seasonal sales data to provide actionable inventory insights.
* Customer Segmentation: Leveraged SQL queries to identify customer segments with higher satisfaction ratings, particularly focusing on payment method preferences.
* Profitability Insights: Found branches with the highest profit margins and product categories with the most revenue.
* Payment Method Preferences: Analyzed customer behavior based on payment method preferences and its correlation with ratings.

**Recommendations:**

* Dynamic Pricing: Implement dynamic pricing strategies for high-demand categories based on the insights from predictive analytics.
* Inventory Optimization: Optimize inventory management by forecasting demand based on seasonal trends and sales patterns.
* Payment Method Incentives: Introduce promotions for underused payment methods (e.g., mobile payments, Ewallet).

**5. Machine Learning Models**

* **Sales Forecasting:** 
  + Developed a regression model ( Random Forest, XGBoost) to forecast future sales.
  + Evaluated model performance using RMSE and MAE metrics.
* **Product Recommendation System:** 
  + Built a collaborative filtering-based recommendation system for cross-selling, using customer ratings and product categories.

**6. PostgreSQL Usage**

PostgreSQL was utilized for managing large datasets, running complex SQL queries, and storing results for further analysis. Key PostgreSQL actions included**:**

* Data Aggregation: Using SQL queries to aggregate data by city, branch, and product category for analysis.
* Customer Segmentation: Extracted customer segments based on payment preferences and spending behavior.
* Performance Analysis: Analyzed sales trends, peak hours, and product categories contributing to sales.

**7. Impact Report**

**Revenue Growth:**

* Estimated revenue increase through optimized inventory management and cross-selling strategies.

**Customer Satisfaction:**

* Highlighted improvement in customer ratings due to better product recommendations.

**Innovative Ideas:**

* Dynamic Pricing: Used predictive analytics to adjust prices based on demand patterns, increasing profitability.
* Personalized Marketing: Segmented customers and tailored promotions based on spending behavior and ratings.

**Requirements**

* Python 3.x
* PostgreSQL
* Pandas
* NumPy
* Scikit-learn
* Matplotlib
* Seaborn
* psycopg2
* SQLAlchemy

**Usage**

To run the analysis:

1. Clone this repository.
2. Install the required dependencies:
3. pip install -r requirements.txt
4. Set up your PostgreSQL database and import the dataset.
5. Follow the notebook or script instructions for running data preprocessing, EDA, machine learning models, and business recommendations**.**