Link to Dataset to be used for the challenge: Please request access and the BEES team will share directly: https://drive.google.com/drive/folders/1K-euQO3F8Wvt966OU0tWwUKErnJIK_Om?usp=share_link

Questions

1. **Customer Segmentation**: Can you segment customers based on their ordering behavior, such as frequency, time of day, or days since prior order, to better understand their preferences and propose ways to optimize marketing strategies?
2. **Inventory Management**: Implement a model to predict inventory levels for individual products in different departments and aisles based on historical order data, order frequency, and seasonality? Please use the designated test sample in the dataset to evaluate model performance. How can this model be used to optimize stock levels and reduce waste?
3. **Order Fulfillment (Bonus points):**How can we optimize the order fulfillment process, considering factors like day of the week and time of day when most orders are placed, to ensure timely and efficient deliveries to customers while minimizing operational costs?

Data Challenge Solving Instructions:

Problem Definition:

Start by clearly defining the problem you are going to solve. Understand the context and objectives of the challenge, and collaborate with domain experts if necessary. Establish a well-defined problem statement.

Data Collection (Restricted to Provided Data):

Collect and work exclusively with the data provided for the challenge. Identify data sources within the provided dataset and ensure data quality and integrity during analysis. Document your usage of the provided data sources.

Data Preprocessing:

Prepare and clean the provided data to make it suitable for analysis and modeling. Handle missing values, outliers, normalize or scale features, and encode categorical variables. Document your data preprocessing steps.

Exploratory Data Analysis (EDA):

Perform exploratory data analysis to gain a deeper understanding of the provided data. Visualize the data, identify patterns, correlations, and anomalies, and generate insights. Document your findings, as they will guide your modeling process.

Model Building and Evaluation:

Select appropriate machine learning or statistical models for the task at hand. Train the models using the prepared data and evaluate their performance. Utilize various metrics and cross-validation techniques. Document your model choices and evaluation results.

Conclusions and Insights to Improve the Business:

Summarize the key findings and insights gained from the analysis. Offer recommendations or actions that can be taken to improve the business based on the results. Ensure that your analysis provides valuable insights for business decision-making.

Tools and Programming Language

Tool Selection for Replicability:

Solve the problem using exclusively open source tools to ensure replicability. Python is our primary language of choice, but if you prefer to use any other programming language, please inform us in advance.

Jupyter Notebook Usage:

Create a Jupyter notebook as your primary workspace for problem-solving. You should solve the challenge, write code, and document your process entirely within this notebook.

Documentation Within the Notebook:

Document each step of your analysis, modeling, and any key insights directly within the Jupyter notebook using markdown. Clear and concise explanations are essential for replicating your work.

Single Jupyter Notebook Submission:

Submit only one Jupyter notebook containing your entire solution. The notebook should encompass all your code, comments, visualizations, and findings. Ensure it is well-structured and easy to follow.

These instructions ensure consistency and transparency in the data challenge solution. By adhering to open source tools, documenting in a Jupyter notebook, and submitting a single comprehensive notebook, you facilitate the replication of your work and enhance the clarity of your approach.

ABInBev
250 Park Avenue, New York, NY, 10177
ab-inbev.com