Assignment: 3-Month (Weekly) Forecast

Overview

You have been given a dataset that tracks weekly sales quantities (the "quantity" column) for a specific product (or set of products) defined by fields like weekend_date, channel, brand, category, sub_category, and EAN. Your task is to:

1. Explore and understand the data (**EDA**).
2. Build a forecasting model to predict **weekly** sales for the next **3 months** (September, October, November).
3. Report the Validation Accuracy for **Last-3-months** in the data i.e. Jun-Jul-Aug.
4. Provide your forecast results and measure monthly accuracy, as specified below.
5. Share your work in **two Jupyter Notebooks** (one for EDA and one for modeling) and **include the final model objects** so your solution can be fully reproduced.

Data Description

- **weekend_date**: The date marking the end of each week (weekly time index).
- **channel**: Sales channel identifier (e.g., “Channel1”).
- **brand**: Product brand identifier.
- **category** & **sub_category**: Product classification details (e.g., “Body” → “Body Lotion”).
- **EAN**: Unique product identifier.
- **quantity**: Weekly quantity sold (target variable for forecasting).

These columns provide a clear view of **weekly sales history** for a product. You will focus primarily on **weekend_date** (as your time variable) and **quantity** (as your numeric series to forecast).

Tasks

1. **Exploratory Data Analysis (EDA) Notebook**
    - **Data Inspection**: Load and inspect the dataset (e.g., checking for missing values, outliers, or data inconsistencies).
    - **Visualizations**: Plot weekly sales over time, look for trends or seasonal patterns, and check if there are any anomalies.
    - **Initial Insights**: Summarize key findings that might influence your forecasting approach (e.g., seasonality around certain months).
2. **Forecasting & Modeling Notebook**

- **Data Preparation**:
  - Confirm the data is at weekly intervals without gaps (or address missing weeks as needed).
  - Split your data into training (historical) and validation/future periods.
- **Model Building**:
  - Select at least one machine-learning approach.
  - Fit your model(s) on the historical weekly data.
  - **Save the fitted model object(s)** (e.g., as .pkl or .joblib files) for reproducibility.
- **Validation:**
  - Perform Validation on **Last 3 months** i.e. whole of **Jun-Jul-Aug 2024**
  - Report the monthly Accuracy for the same
- **Forecast Generation**:
  - Produce **weekly forecasts** for the next 3 months (i.e., September, October, November).
  - Present a clear table or DataFrame with week-ending dates and predicted quantities.

3. **Monthly Accuracy Evaluation**
   - Once actual data becomes available (or if you perform a backtest on known actuals), aggregate weekly forecasts to **monthly** totals and calculate accuracy using:
     - Monthly Accuracy= $1 - \frac{\sum|\hat{y}-y|}{y}$
     - **Note:** '$\sum$' is summation. $\hat{y}$ is predicted target and y is the actual target
   - Report this value for **each month Jun-Jul-Aug 2024**.
   - Provide commentary on how the model performed and any observed trends or discrepancies.

Deliverables

1. **EDA Notebook**:
   - Exploration of the data (descriptive statistics, time-series plots, insights).
   - Any data-cleaning or transformation steps you performed.
2. **Modeling Notebook**:
   - Explanation of the model selection and fitting process (code + rationale).

- Weekly forecast results for Sept–Nov (e.g., a DataFrame with weekend_date and forecasted quantity).
- Monthly accuracy calculations (if you have actuals for that time period or through a backtest approach).
- **Model Objects**: Include the serialized/fitted model file(s) so that anyone can load the model and replicate your forecasts without re-fitting from scratch.

3. **Presentation/Discussion**:
    - Summary of your findings in both notebooks, focusing on data insights, forecast performance, and recommended next steps for improving or extending the model.

Notes & Suggestions

- If the dataset contains multiple products/channels, you may filter to a specific product or channel as shown in the screenshot or create a more comprehensive forecasting approach that handles multiple series.
- Take care to **document your assumptions** (e.g., handling missing data or outliers).
- For reproducibility, ensure you **version-control your code** and clearly state any library dependencies (e.g., pandas, statsmodels, scikit-learn, prophet, etc.).

Good luck, and we look forward to reviewing your **two notebooks**, your **model artifacts**, and your **forecast accuracy** findings!

Data:

Please find the data attached to the link: Data