



ASSIGNMENT 1

DATE: 2ND FEBRUARY 2025

UNIT NAME: PROGRAMMING FOR DATA SCIENCE

UNIT CODE: BSD 3203

MEMBERS:

WILLY WARIO – 19/04141

21/04744 EDWIN WAIRAGU

22/04929 FERNANDES BABULT

22/03876 IMANI MUNGAI

22/04328 CHRISTOPHER ODHIAMBO

21/05235 NOAH MBURU

21/00153 ENOS IGUNZA

21/05124 GEOFFREY ROTICH

Problem: Predicting the Demand for Groceries in a Small-Scale Delivery App

We are currently developing an app where users can order from small-scale businesses (grocers, cafes, etc.) within their building's delivery scope. One major challenge is predicting the demand for different groceries to help sellers manage inventory efficiently and minimize waste.

1. What the Data Is

The data would consist of:

Order history - items ordered and quantity

User demographics - age, location, frequency of orders

Seasonal trends - sales variations based on weather, holidays. For example items like cold soft drinks would sell more during hot weather.

Customer reviews and ratings – were the customers satisfied with the service they received and what would they want added.

2. Expected Benefits from Mining the Data

By analyzing this data, we can:

Optimize inventory management – Sellers can stock items based on predicted demand.

Reduce waste – Prevent overstocking perishable goods by predicting purchase patterns.

Personalize recommendations – Suggest relevant groceries based on user purchase history.

Improve pricing strategies – Adjust prices dynamically based on demand fluctuations.

For example, if data shows that fresh fruits sell more on weekends, sellers can stock up before the weekend to meet demand.

3. Relevant Machine Learning Algorithm: Time Series Forecasting

Since grocery demand fluctuates over time, time series forecasting (a form of regression analysis) would be suitable. This model can analyze past sales data to predict future demand patterns.

Example:

If historical data shows a spike in dairy product purchases at the beginning of the month (when most people receive salaries), a forecasting model can predict and recommend stocking more dairy products during that period.

4. Irrelevant Machine Learning Algorithm: Clustering

Clustering is used to group similar data points, such as customer segmentation. However, it is not directly useful for predicting demand over time.

Reason:

Clustering would group customers with similar shopping habits, but it wouldn't provide precise time-based forecasts.

It doesn't account for seasonality and trends that impact grocery demand.

Example:

If we clustered customers based on their order frequency, we might discover that certain users order groceries weekly, but this wouldn't help predict when the next surge in demand will occur.

Conclusion

By applying time series forecasting, we can help grocery sellers maintain optimal stock levels, reduce waste, and improve user satisfaction. Clustering, while useful for user segmentation, would not directly solve the problem of predicting demand trends over time.

References

<https://www.tableau.com/analytics/what-is-time-series-analysis#:~:text=Time%20series%20analysis%20is%20a,data%20points%20intermittently%20or%20randomly.>

https://www.datacamp.com/blog/classification-vs-clustering-in-machine-learning?utm_source=google&utm_medium=paid_search&utm_campaignid=19589720824&utm_adgroupid=152984014014&utm_device=c&utm_keyword=&utm_matchtype=&utm_network=g&utm_adposition=&utm_creative=726015683661&utm_targetid=aud-517318242147:dsa-2222697810918&utm_loc_interest_ms=&utm_loc_physical_ms=1009824&utm_content=DSA~blog~Machine-Learning&utm_campaign=230119_1-sea~dsa~tofu_2-b2c_3-row-p2_4-prc_5-na_6-na_7-le_8-pdsh-go_9-nb-e_10-na_11-na-jan25&gad_source=1&gclid=Cj0KCQjAkoE9BhDYARIsAH85cDM3G136lIVTYKUI0CgP40yo5qYtAS2Nnx60NfiwIqz0q2kgiLA_cTQaAvxJEALw_wcB