Introduction

This challenge represents a typical project that MarTech is working on a daily basis. Your task will be to predict the customer lifetime value for each customer within the next 12 months. Here is the schema description:

Name	Description	Example
customer_id	unique customer ID	d43cc72433881...
order_number	unique order ID	fb69e40b059fc2...
order_date	date when order was placed	2019-01-12
item_quantity_ordered	how many items were ordered	2
revenue	order value	389.40
cumulative_revenue	revenue a customer had accumulated previously	720.94
customer_region	region where purchase was placed	Berlin
customer_gender	customer gender	f
device	device where	tablet
cost_of_living_rank	zip code based ranking of costs of living	8.0
customer_lifetime_value	revenue a customer generated in the next 12 months	540.25

Please note that:

- cumulative_revenue is only including previous orders (if there were any)
- customer_lifetime_value is only including orders in the next 12 months
- 10.0 represents the highest value for “cost_of_living_rank”

We split this task into several parts to help you structure your answer, but feel free to take any approach. The goal of this exercise isn’t necessarily to obtain best predictions results, but to evaluate your reasoned take on a problem.

Please, provide the result in the Jupyter Notebook, including comments and summaries. We mostly use Python at home24, but you are free to R as well.

Lastly, you are at liberty to draw support from LLMs if you feel like it; however make sure to make the solution yours! Share your own insights about your findings.

Good luck!!

Part 1 - Exploratory Data Analysis

Question 1a:

Please guide us through an exploratory data analysis, and explain your findings, e.g.,

- What problems do you observe?
- What solutions would you suggest?

Question 1b:

Based on your analysis, how would you prepare the dataset before training a customer lifetime value predictor (regression model)?

- Provide the data preparation code
- Please explain all your decisions, and include visualizations and summaries

Part 2 - Model Interpretation

Please evaluate and interpret the model in the provided Jupyter Notebook. Here are some hints:

- How would you interpret the results?
- Is the model well-chosen and well-trained?
- Do you agree with the validation?
- Based on your analysis what would you add/change?

Question 3 - Modeling

Based on your analysis and data preparation suggestions from Part 1, train an easy customer lifetime value predictor (regression model).

Please explain each of your design choices, e.g. processing, model selection, hyper parameters, and evaluation criteria.

Question 4 - Comparing Models

Question 4a:

Compare and contrast your model with the provided one.

Question 4b:

Describe possible future improvements for your model.

We are looking forward to receiving your solution!