Aella Data Science Activity

General guidelines

What we'd like your analysis to meet at least one of the following criteria;

Analysis and modelling: logical and creative examination of the dataset; ability to apply relevant statistical or machine learning methodology to tease out business-relevant insights.
Engineering: proficiency with Python/SQL
Communication: strong written communication skills with regards to:

- Actionable recommendations
- Clear assumptions
- Helpfully documented code

Assignment

Task 1

user_id	date_visited	purchase_amount
10001	2021-01-20	30.00
10002	2021-01-25	300.00
10003	2021-02-01	3000.00
10004	2021-01-14	15.00
10005	2021-01-13	150.00
10001	2021-01-01	1500.00
10004	2021-01-12	3200.00
10002	2021-02-23	2300.00
10005	2021-02-13	120.00
10001	2021-03-09	450.00
10001	2021-03-02	2300.00
10002	2021-02-06	120.00
10005	2021-04-01	23.00
10003	2021-03-29	5000.00
10001	2021-04-07	20.00

1. Given the above table, structure a query to determine the total amount spent by all visitors?
2. Given the above table, structure a query to return all users who visited in February AND spent more than 1000 naira
3. Given the above table, structure a query to determine the highest and lowest amount spent in each month
4. Given the above table, structure a query to determine the total monthly purchases
5. Given the above table, structure a query to determine how much each user spends on their second purchase.

Task 2

With any tooling you prefer, please build a model and conduct analysis into predicting loan defaulters based on the below data set of employment status, loan amount and bank balance

Index	Employed	Bank Balance	Annual Salary	Defaulted?
1	0	8000	300000	0

A csv can be found in this found in the below link:
https://drive.google.com/file/d/1bzK5K1HnC6uLm4aEYllPnW_PbDnZuDZr/view?usp=sharing

- Please note that for the purpose of this exercise, we're more interested in the choices you make, and how you communicate them, than in you having a perfectly optimised model.
- Please clearly document your assumptions, visualizations and findings.
- In the case where the findings are not clear, make sure to provide recommendations and document analysis based on what you've learnt so far.
- Please present the analysis in any form you prefer; however, a runnable notebook is preferred.

Submission Guidelines

- Submit your result as a link to a github repo.
- Include in your submission any scripts, notebooks, SQL queries.
- If you're using a notebook, submit them with run cells.
- Include a README that includes any packaging steps.
- Please submit your submission within 6 days from the day we sent you the exercise.