Data Scientist - Take Home Assignment

The assignment has 3 sections:
1. SQL
2. ML Assignment
3. Business Case Study

Please feel free to use online resources. We will be assessing your understanding, logical thinking and problem solving skills. Please make assumptions & list them wherever necessary.

SQL Test

In this assignment we ask you to answer the questions listed below. You may use data from two datasets that we have provided you. Note that we expect your answers to be based on these datasets – **there is no need to use external data sources**. Make sure to include all information and/or insights that you find relevant and, of course, feel free to explain why you chose to use a certain approach using comments.

We have two tables available: **transaction** and **item**.

Table: transaction						Table: item	
User id	Purchase time	Refund time	Item Id	Gross Transaction Value (GTV)		Item Id	Item Name
56	11/4/2020 3:38:19	11/7/2020 20:59:29	3	$34		1	cellphone
52	11/5/2020 2:40:32		6	$188		2	fan
55	11/4/2020 22:05:44	11/6/2020 11:12:19	11	$133		3	phone charger
65	11/4/2020 0:18:51		6	$176		4	mouse
66	11/3/2020 12:16:19	11/5/2020 17:02:16	1	$229		5	mug
47	11/3/2020 9:19:41	11/5/2020 14:32:04	5	$130		6	vase
61	11/2/2020 18:16:42		6	$98		7	chair
61	11/1/2020 5:44:18	11/4/2020 14:33:35	5	$222		8	sofa
61	11/2/2020 19:44:22	11/4/2020 10:20:11	2	$176		9	diary
62	11/1/2020 6:08:30	11/3/2020 9:51:08	8	$82		10	electric outlet
55	11/1/2020 2:40:47		8	$151		11	glasses
54	10/31/2020 16:59:22	11/2/2020 10:39:36	3	$114		12	water dispenser
48	10/31/2020 7:25:36	11/2/2020 2:03:04	13	$216		13	painting
54	10/31/2020 11:16:20		3	$222		14	cat cage
48	10/30/2020 5:39:46	10/31/2020 13:55:12	2	$248		15	cable extension
62	10/30/2020 2:22:58		7	$46			

Please write a SQL Queries to answer the questions below:

1. What is the most popular item that users order on their third purchase? If any.
2. How long is the shortest interval (in minute) the store has gone without purchase and refund?
3. For users who ordered at least three times in the store, how much dollar they have spent until their third purchase?

Please write your SQL queries in word doc and share it with us.

ML Assignment

Banks lend money to customers for personal or business uses at a certain interest rate. Customers pay a premium on the loan through easy monthly installments based on the term period of the loan and interest rates. Some customers fail to pay their EMIs either on time or by the amount. Given historical data for customers, predict if the customer is going to default on his next EMI. *Please use only the provided dataset as your base dataset*. The data description, training & testing datasets are provided in the google drive. Alternatively, you can access the dataset from the link below
https://drive.google.com/drive/folders/1Hdfd7s5_p8d5U4QB0e_iSmGikYxjyQ8a?usp=sharing

We would like you to explore & build models on the dataset using python jupyter notebooks. It would be great if you can structure your project by creating a simple python virtualenv and local git repo. Once you are done with your assignment, you can send us the notebook/results as a zipped file.

Note: *Please do not upload your assignment/datasets on any of the public platforms*

Business Case Study

SMEs play a major role in empowering people through entrepreneurship, employment and are also the major backbone of the economy. While running the business most of the SMEs cannot get investments from financial institutions because of various reasons. Understanding why we are doing, please list down a few points on
● How can we empower SMEs through BukuWarung & Tokoko?
● What are the current challenges faced by SMEs?

- What are the data points that you need to evaluate a SME before lending? How do you use these data points to build credibility scores? How do you differentiate between consumer lending and SME lending?

We request you to send us your solutions to all of the sections together.

All the best,
speak to you soon!