Decision Science Entry Test

Applicant's name: ……..

Date: ……

Intro

This document contains list of the tasks we would like you to complete in order to evaluate fully your technical abilities.
Please read thoroughly all tasks and in case questions pop-up, do not hesitate to ask us.
You have 5 working days to complete all the tasks. However, if by some reason you cannot get all results in full, don't worry. The purpose of this exercise is not to get exact numbers, but to understand how you would approach similar assignments, what would be you thought process and how you would advise your Client (internal/external) at the end.
You are to work on one general task which is further divided into three steps: data handling, modelling and presentation. We would like to get following outputs from your side:

- Complete Python/R code for the solution. Please don't forget to include notes and comments within;
- Presentation of the results:
    - Data Storytelling;
    - If you consider to present numerical outputs of the final models and statistical calculations separately of the final presentation, please put them in a proper output form;
    - Any additional visualizations and approach of the data you consider useful to support your work (interactive charts would be considered as an advantage) are up to you.
- Any other materials that you consider useful are more than welcome.

At the end of the period, you will be required to present your work in front of the team face-to-face or via Teams.

Many thanks and good luck!

RenMoney®

Task: Cluster the customers

You are provided with a set of data (in the attached zip file) on behavior of customers for a company in the US. The data is extracted from their CRM.
Your general goal is to cluster the customers in optimal number of clusters - from a statistical point of view - in such manner that these clusters are meaningful for next step business actions. These next steps are out-of-scope for this assignment.
You are required to provide your input on the following questions:

1. Please review the data, do some data preparation and describe your findings. There are a lot of missing values. Keep in mind that different forms of missing values: N/A, NULL, etc., could be meaningful in some cases.
2. Will you use all data for clustering? If not, which variable you are going to keep? Please clearly describe how and why you have used exactly these variables.
3. Do your clustering. How many clusters you have produced? Why? Why the solution is considered optimal?

Finally: prepare a PowerPoint containing all steps: your milestones, findings and final solution. Be ready for questions and to defend your solution. There is no right or wrong approach – it's the story behind and how it sticks to the data (and to the statistical theory to some extent, of course ☺).