Data Scientist - Assignment

This test assesses your quantitative and programming skills.

Take the time to solve the problem listed below and submit detailed and well documented answers (code, documentation, data, etc) as well as a deck with problem statement and results presentation.

Bear in mind that all exercises and information disclosed at the interviews with DKG are confidential and must not be shared or disclosed with anyone nor in any websites.

Model development
The attached csv-file (Data.csv) contains data about 1000 UK based businesses.

Please complete the following tasks and provide your code in line with coding standards in Python or R.

1. Read the data from the csv file into a data format of your choice.

2. Preprocess data. Keep in mind that data may contain missing values.

3.Select and build a simple model that explains total funding attracted by one or more variables provided. You are free with respect to the methodology you choose and criterion. However, ensure that the approach is economically and statistically plausible.

4. Provide and explain all the steps and details of your modelling approach as well as model performance. Include a short justification for all of your choices.

Good luck!