Test Assignment

In Raiffeisen Bank, one of the most frequent tasks in ALM (asset & liabilities management) is to retrieve some dependencies in data and model them. These models are used for managing interest rate risk and liquidity risk.

Imagine that you need to model and explain the dynamics of prepayment rate for mortgages (it is required for quantifying the pricing for the prepayment option embedded into the loans). You are given the sample (artificially created; however, it resembles the real-life dynamics of prepayments) containing 4 variables:

- MDate – date in a format of year and month
- Prep_Rate – prepayment rate on mortgage portfolio (share of prepaid volumes to outstanding amounts for a given month)
- Interest_Rate – weighted average interest rate on mortgage portfolio
- Age – weighted average age of mortgages portfolio, measured in months

Using this dataset, please, answer the following questions:

1) Briefly, give your expectations on possible factors affecting prepayments
2) Make an exploratory data analysis – provide it with plots (preferably pretty ones, try best you can), don't forget about conclusions and description of findings
3) Using the dataset (and extra factors from open sources if you see the reasoning), build a linear regression model, which will explain the dynamics of prepayments
4) Interpret the estimates of the model. Were they expectable? Are they reasonable? Why? Are they statistically significant?
5) Make sure that all statistical assumptions are satisfied (perform necessary statistical tests to ensure it)
6) Propose validation scheme for the model and conduct it. Does a model seem to be stable? What are the indicators of the model's stability?
7) Create some benchmark, and test how your model performs against it
8) Propose the ways how the model can be improved (maybe add some new factors, maybe try different models, etc.)
9) Write overall conclusion for the work done

Should you have any questions, don't hesitate to ask them!

Good luck!