Below are the steps for two models, CatBoost and Sklearn Ridge, that we'd like you to perform:

1. Prepare the necessary preprocessing steps for both models, utilizing existing resources if available.

2. Determine the appropriate validation strategy for model validation (e.g., KFold, StratifiedKFold).

3. Provide initial prediction results with simple parameters for both models.

4. Perform feature selection using lofo-importance as outlined in this article: [Link to the article].

5. Implement hyperparameter optimization using techniques such as Grid Search, Random Search, or Bayesian Search. If possible, consider using Optuna (https://optuna.org/).

6. Demonstrate how your choices from step 3 to step 5 have improved model performance, documenting the pros and cons of each experiment.

7. Interpret model variables using SHAP values. You can use this resource.

8. (Optional) Explore feature engineering techniques, creating new variables and validating their impact on model performance.

You can access the dataset here.

Finally, please compile your work into a Jupyter notebook with the last 7-8 headings. Feel free to reach out if you have any questions or need clarification. We're looking forward to seeing your progress.