Home assignment

Data Scientist

HomeBuddy.

We're thrilled to have reached this stage with you in our recruitment journey. It's been a pleasure getting to know you, and now we're excited to see you in action with our home assignment.

This step is designed not only to showcase your skills but also to provide you with a glimpse into your day-to-day responsibilities at HomeBuddy.

Below, you'll find the task instructions - feel free to share the task results by uploading your work to a private GitHub repository and sharing the link with us for assessment. If at any point you have questions or if you're unsure about anything, don't hesitate to reach out to Zarina Murzalieva at zarina.murzalieva@homebuddy.com.

While there's no strict time limit, we suggest allocating 8h for preparation - we want it to be both enjoyable and manageable.

Good luck, and thank you for your interest! We hope you will enjoy working on this home assignment!

Instructions:
You are tasked with building a predictive model to determine whether a lead will result in an appointment setting('is_appointment_set': true or false).
You can structure your code into files like a repository style or through a notebook-based approach. While Python is the preferred language for implementation, you can use any language you choose.

It would be best to document your solution comprehensively;

- detailing the preprocessing steps,
- feature engineering methods,
- model selection,
- evaluation procedures.

Data Description:
Three anonymized datasets were provided:

1. **Lead Dataset**: Include hashed phone numbers, zip codes, and user agent-derived information.
2. **Zip Code Statistics Dataset:** Offers statistical data related to various zip codes. Note that column names are obfuscated to protect company data.
3. **Infutor Data Service Dataset:** This contains information associated with hashed phone numbers, sourced from Infutor.

Bonus Points:

- Exploratory Data Analysis
- Clean Code
- Scalability and Efficiency