Preparation for interviewing candidate analysts

Hi <analyst>,

Here's a link to the data we'll be using:

https://drive.google.com/drive/folders/190SEfRn6cRZ6tdGGr7qCKIpvzvsGvAkh?usp=sharing

The large 'PM' csv files are air quality sensor readings for individual participants. Each participant hosted 3 sensors at a time (two indoors and one outdoors), but each deployment might have more than 3 unique sensor IDs due to hardware failure and swapping things out. This is why I've added the 'location' column to make it a bit more user-friendly to understand where each of these measurements were collected within the home. The 'value' column represents the level of measured airborne particulate matter from that sensor at that time. Each monitor logged data every minute, and in some cases ran for ~11 months, so there's quite a lot of data per deployment.

The annotation csv file contains the small notes each participant made on their data to help make things a little more readable. We had a script that would alert participants to particulate spikes (such as when there was a sudden rise in detected levels), and we would send them a text asking them to explain what was going on.
The 'annotation_classification' column is a sorting of their responses into activity types for privacy purposes. The annotation_source column indicates whether they made their annotations through text messaging (sms), voice command (google_home), or directly through their tablet interface. Event_time is when the spike was detected and insert_time is when the participant made their annotation.

Let me know if you have any questions or if there's trouble accessing the link. I'll go over this again tomorrow, but I wanted to make sure you had some time to get familiar with the data on your own so it wasn't a complete surprise!

See you tomorrow,

<researcher>