We would like SDE candidates to complete a few takehome tasks.

Important Note:

data used are public data and they may not be accurate interpretation of what data represents, the purpose of this exercise is to give candidate an opportunity to demonstrate his/her understanding of ELT in Snowflake and how to engineer data solution from source to data model that can be used by non-technical business users. It will also give SDE candidates to showcase his/her skills with the tools or learning to work with different tools.

Keep in mind, SDE candidate may very well be more familiar or better at the tools I have recommended so present the best you have and surprise us.

We understand that SDE candidate may find ambiguities or inaccuracies in requirements. Since this is not normal development process, when encountering such issues, SDE candidate is encouraged to make his/her own judgement and explain the reasoning behind the decision and move forward with a solution or alternative.

What's required in delivery

1. Snowflake codes checked into Github
   - ✦ codes needs to have clear comments to guide readers who can review and approve
2. ELT process if not Snowflake built and checked into Github (if can't check into Github, please explain why).
3. Snowflake objects built
4. The solution does not need to be perfect, SDE needs to demonstrate understanding and thoroughness in sharing

documentation. Brief and concise documentation that's readable by fellow engineers and QA including the following:

- problem statement
- design of ELT
- build and test
- presentation or data visualization if necessary
- Since we will not have discussion sessions: what's built to meet the requirement? what's not built based on the requirement and why?
- any potential improvements that can't be done due to time constraints.

Data set

Business wants to understand how WaFd is serving its customers in each state, by analyzing WaFd's customer composition against government's stats (https://data.census.gov/). This exercise is to extract data that we need from Census' stats and transform it to the reusable table format that analysis and reports can be built.

The questions that business needs to answer are:

1. how are we serving population segments with different languages?
2. how are our products serve the general households?
3. will online banking appealing to our customers given internet readiness and literacy?

In order to answer these questions, we need to prepare data for the following:

1. raw household stats by states: already loaded to Snowflake in table stg_state_household_data.

2. raw spoken languages stats by state: in CSV file, need to be loaded to Snowflake.
3. wafd location dimension table: Snowflake table wafd_location.
4. cleanse and transform raw data to models that can be presented to business users in a view that include the following:

- location_id
- state_code
- avg_persons_per_household
- percent_have_computer
- percent_have_high_school_degree_and_higher
- percent_english_only
- percent_bilingual percent_no_english

Development Tools

Tools: Snowflake, Talend, any scripting/coding languages

SQL IDE: recommend DBeaver for its integration with Github
Version Control: Github

Snowflake setup
Database: open_test
schema: public_test
role: sde_candidate_role
warehouse: opentest_public_vwh