8.2.2

Prepare for the ETL Project

After you and Britta have inspected the worksheets, you both go over the tasks that you'll need to accomplish for the ETL project. It will require lots of coding, but you both feel ready for the challenge. Note that after each task, you'll add your code to the GitHub repository.

For the ETL project, you'll first help Britta with the extract phase. Specifically, you'll write the code to read each worksheet of the `crowdfunding.xlsx` file into a Pandas DataFrame. In later lessons, you'll help Britta with the transform and load phases. Specifically, you'll create the following DataFrames, export each as a CSV file, and then upload the CSV files into the appropriate tables in the PostgreSQL database:

- A `category_df` DataFrame
- A `subcategory_df` DataFrame
- A `campaign_df` DataFrame
- A `contacts_df` DataFrame

Now that we've prepared for the ETL project overall, we'll start with the extract phase. Specifically, we'll extract the crowdfunding data into Pandas DataFrames.

© 2020 - 2022 Trilogy Education Services, a 2U, Inc. brand. All Rights Reserved.