Python / Pandas

1. Download the following dataset of Open Recipes.
2. Write a script in Python that reads the recipes and extracts every recipe that has “Chilies” as one of the ingredients. Your code should allow for misspelling of the word for example Chiles as well as the singular form of the word.
3. Add an extra field to each of the extracted recipes with the name difficulty. The difficulty field would have a value of "Hard" if the sum of prepTime and cookTime is greater than 1 hour, "Medium" if the total is between 30 minutes and 1 hour, "Easy" if the total is less than 30 minutes, and "Unknown" otherwise.
4. The resulting dataset should be saved as a *.csv file.
5. Place your answer in a directory called "recipes-etl" in the root of this repository along with a README.md file that outlines the instructions to run your script and a requirements.txt file with the modules you used to run it (if applicable).

Remarks

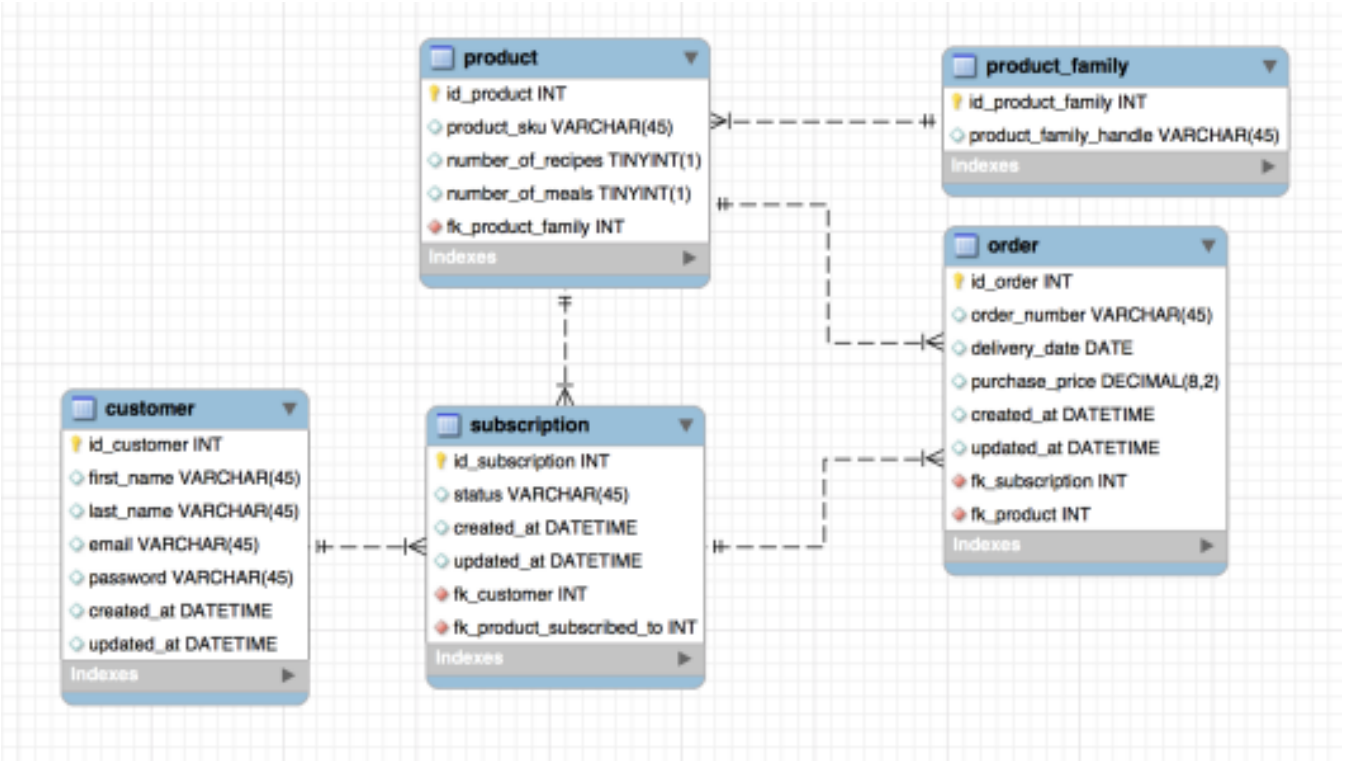
1. The README file must contain detailed instructions on how to run the code you submitted, what it does and how it does it. It should also include instructions on how to install any third- party modules you used to complete this assignment as well as the Python version you used.
2. Your submission should include a *.py file, a README.md, a requirements.txt (if applicable) and your deliverable *.csv file.
3. You are free to use any Python module you wish and any Python version that fits your needs.
4. Your code should include comments in the form of docstring and be easy to read.

Disclaimer

1. Code submitted in a format other than the one requested (*.docx, *.doc, *.txt, etc.), will not be taken into consideration.
2. If the Python version used is not clearly documented, the Hiring Manager will run the code with his pre-installed version. If the code doesn’t run, it will not be taken into consideration.

SQL

Given the following ER diagram:



And given the following information:

- created_at fields have a timestamp of the time at which a row was inserted in a table for the first time
- product_sku is unique for every product
- product_family_handle is unique for every product family
- Subscription status can be: ‘paused’, ‘cancelled’ or ‘active’
- The fk_product field in the order table indicates the product that was purchased in an order. The fk_product_subscribed_to in the subscription table indicates the product a subscription is currently subscribed to. A subscription can change the product it is subscribed to at any moment
- An order is generated when a box is shipped to a customer

Questions - Write SQL statements to retrieve the following information:

1. For the customer with email address ‘ilovefood83@hotmail.com’ show all product_skus the customer has an active subscription for
2. Get a list of all the customers (id_customer) that have an active subscription to a product that corresponds to a product family with product_family_handle = ‘classic-box’
3. How many customers have ordered more than one product?
4. Get a list of all customers which got a box delivered to them two weeks ago, and the count of boxes that had been delivered to them up to that week
5. For all our customers, get the date of the latest order delivered to them and include associated product_sku, delivery_date and purchase price. If there were two orders delivered to the same customer on the same date, they should both appear

Good luck!