Take Home Challenge - Data Platform at Marley Spoon

Overview

With this Coding Challenge, we would like to get a better understanding of how you work, approach problems and communicate your results. We know that you have other things to do and want to be respectful regarding your time. We estimate the duration of this task to be around 4 hours of work.

Context

We at Marley Spoon provide our customers with a weekly **box** containing all the ingredients and recipes to cook three to five meals for two to six people each. We offer our customers more than 20 recipes per week to choose from. With a varying number of recipes per box, a unique selection of recipes and differing meal sizes, every box is unique. These boxes are packed at our **Fulfillment Centers (FCs)** around the world. In every FC there is a **production line** consisting of several **sections** each operated by one of our production associates. Every section consists of a unique set of **ingredients** from which the production associate needs to select/**pick** all the ingredients required for the box she/he is currently working on. Once a box is completed in a given section, it continues down the line to the next section where other ingredients can be added. After the box has passed all the sections a **Quality Control (QC)** team inspects a random number of boxes to verify that the right ingredients have been picked in the right amounts. If a box passes the quality control, it is shipped to the customer.

Every event in this process is tracked in the FC. The team you are applying to, will be in charge of collecting that data from the different FCs along with relevant data from other systems to be then made available to different entities within the company.

Task

For this coding challenge we have prepared some sample data in the form of two csv's representing data coming from one FC. One of the files (*data/input/picks/pick_1.csv*) contains pick events, the other file (*data/input/quality_control/qc_1.csv*) contains information about which order (= box) and ingredient was checked and if it had a defect or not.

We would like to know the defect rate per section over time.

It is your task to write an application which reads **all** input files, processes them and outputs a csv file containing a dataset consisting of the **defect rate**, **section** and **timeframe**.

Deliverable

We would like to receive a python application (source code) which:
- We can execute locally
- Is checked into version control (please ensure this is kept **private at all times**)
- Reads all files located in *data/input*
- Writes a csv file to *data/output*

Complications

We will be running the application with additional input files to verify the functionality. These files will have a unique name following this convention:
data/input/picks/pick_2.csv, data/input/picks/pick_3.csv, data/input/quality_control/qc_2.csv, etc..

Additional Information

When reviewing your solution, we will be evaluating:
- If the application functions correctly
- The readability of your solution
- The overall structure
- To what extent it is following Software Development Best Practices
- Git commit hygiene

Please remember, we want to get an understanding of how you approached this task and which decisions you made along the way. Please be transparent about your thought process, reasoning and decisions.

You can submit your solution either by granting us access to a **private git repository** or by using **git-bundle** or **zip** to attach it to an email.

Many thanks in advance from the Marley Spoon Engineering Team!