Data Acquisition Engineer  
Assessment Instructions

Thank you for taking the time to consider HealthVerity and our Data Acquisition Engineer position. Your skills and experience line up well with what we are seeking in this role and we'd like to have you move forward in the process. Below are a few questions we'd invite you to consider and respond to in a separate text document.

**Background:** We are in the process of onboarding a new data provider to our data marketplace. They have just sent over their first file of sample claims for January 2021, but it appears to have several errors. We need to generate some basic statistics as well as review the file for overall data quality. Attached you will find the sample\_claims table, a table definition document, a reference table for valid diagnosis codes and another reference table for valid procedure codes. For each question below please provide an answer as well as the code used to find the answer. Feel free to use your preferred version of SQL or Python. Your code will be reviewed for correctness, efficiency and clarity.

Question 1

1. What are the top 5 most common valid procedure codes?
2. How many patients are associated with at least one of those procedures? Please do not use the result values from 1a - provide code that will find the answer without specifying explicitly those code values.

Question 2

What are the top 5 most common valid diagnosis codes?

Question 3

We need to review this file for overall data quality and highlight any potential problems so that they can be discussed with the data provider. Write a series of tests that will identify any errors in the file and provide a list of all errors found. Do not consider any medical details of the treatments, this question does not require any healthcare knowledge.

These quality checks should include, but are not limited to

* Sample data matches the table definition
* All standardized codes are valid based on given reference material
* Date values are logical and chronological trending is consistent