

# BCBS Data Science Engineer Coding Prompt

## Background

For this step in the interview process we want to see how you write code and tests for production. To be respectful of your time, we've created a problem that should be completable within a few hours.

You can use the language you are most comfortable working in. You can use open source libraries, but you can't outsource thinking about the solution and its data structures to these libraries.

To submit your solution please package it in a zip, tar, or git bundle and email it to back to us at (hscs.educoding.exercise@gmail.com). We will evaluate your submission using the criteria described below.

---

## Problem

At Blue Cross Blue Shield we must track our members' doctors visits and bills associated with these visits. Often times our members need to see doctors in a variety of different clinics (providers) and it can be helpful for our analysts have a monthly breakdown of the total cost of medical services rendered for each member and provider.

Your task is to write code that generates monthly total billed amounts for our members and providers.

Your code should expect to receive two arguments:

1. The (input) billing log file path
2. The (output) monthly cost report destination file path

For example your code could be run like:

```
python monthly_cost_report.py billing_log.txt cost_report.txt
```

The billing log is a text file where each line starts with a tag. There are 3 possible tags.

1. **Member** <members\_name> declares a insurance member's (i.e. patient or care recipient) name. e.g. **Member Frannie**
2. **Provider** <provider\_name> declares a provider (i.e. hospital or clinic name). e.g. **Provider EastShore Healthcare**
3. **Bill** <provider\_name> <members\_name> <date\_care\_provided> <cost\_in\_dollars> declares a bill sent by the provider to Blue Cross for medical care received by the member on the specified date. Note that bills are NOT always in chronological order.

### Example Input Billing Log File:

```
Member Edgar
Member Frannie
Provider EastShore Healthcare
Provider Alliance Medical Group
Bill Alliance Medical Group Frannie 2021-02-02 1050.00
Member Michael
Bill Alliance Medical Group Edgar 2021-01-24 299.25
Bill EastShore Healthcare Michael 2021-01-01 6014.30
Bill EastShore Healthcare Edgar 2021-04-12 3000.00
Bill EastShore Healthcare Edgar 2021-01-07 499.25
```

Your code should generate a cost report which provides the monthly total amount billed for each member and provider.

Additionally, the report should:

1. List the total billed amount on their own line, the line should start with a \* followed by a space the member/provider name, a colon (:) and, finally, the total dollar amount. e.g: \* Edgar: \$798.50
2. List each month and year on its own line followed by a colon. e.g. January 2021:
3. List the members first and then the providers; members and providers should appear in the order they are declared (i.e. their order in the billing log).
4. For a given month, omit members or providers with no bills.
5. Omit any months with no bills.

For example, given the billing log above, your code should generate the follow monthly billing report:

### Example Output Monthly Billing Report:

```
January 2021:
* Edgar: $798.50
* Michael: $6,014.30
* EastShore Healthcare: $6,513.55
* Alliance Medical Group: $299.25
February 2021:
* Frannie: $1,050.00
* Alliance Medical Group: $1,050.00
April 2021:
* Edgar: $3,000.00
* EastShore Healthcare: $3,000.00
```

For simplicity, you can make the following assumptions:

1. The billing log file is always valid - You do not need to worry about

- handling arbitrary inputs
  - 2. The format/type of care dates is always a valid date
  - 3. The billed amount is a positive amount and includes dollars and decimal cents
  - 4. A member/provider declaration will always appear before it is listed in a bill
  - 5. Member names will never contain spaces
- 

## Evaluation Criteria

The goal of this exercise is for you to showcase your knowledge of coding and testing best practices.

We are interested in how you make design decisions when coding, so we ask that you document your design decisions in the README.

We expect your solution to be somewhat efficient - we find that the xlarge log can be processed on a typical Macbook in well under one minute.

We find that using tabular data types (such as **pandas** dataframes in Python) outsources a lot of the reasoning about data structures - for this reason, **we will not review submissions in Python that use pandas.**

When reviewing your submission, we will look for:

- A README with:
  - instructions for how to install and run your code
  - documentation of your design decisions and your thought process
  - instructions for how to run your tests
- Code that:
  - Runs and generates correct reports efficiently
  - Is readable and modular
  - Uses data structures appropriate for the problem
- Tests that:
  - Verify input parsing, calculations, and output formatting
  - Are readable and easy to run

## Additional notes

Please email us if you have any questions about this prompt.

Please do not share your solution with other candidates or post your code on publicly accessible sites like GitHub or BitBucket.