Arine ETL Data Engineer Take Home Evaluation
Arine Inc.
July 17, 2020

Purpose:

The purpose of this take home test is to allow a candidate to show off their coding skills and also experience working with pharmacy claims data.

Assignment:

Arine is providing 4 types of files:

1. Enrollment - has patient information including name, birthdate and enrollment period dates
2. Pharmacy claims - records of each “fill”, which is the receipt (or attempted receipt) of a set of pills for each medication on a specific date. Note that a claim has an associated status: “denied”, “paid” and “reversal”. See below for more detail on handling claim status
3. Ndc9 lookup - this is a mapping from National Drug Code (NDC) to drug information such as generic and brand name; the NDC is a national standard that assigns a code to each unique drug.
4. Example result for one patient

All personally identifiable information (PII), such as names, birthdates, ids, claim numbers etc., have been fictionalized.

We would like you to use **python** 3.6 (or greater) to process the 3 types of input files (enrollment, pharmacy claims and ndc9 lookup) to create a medication summary for each patient **enrolled as of July 1st, 2020** for which we also have pharmacy information. Not all the patients in the enrollment file have pharmacy claims data and vice versa. Only medication “fills” that occurred **between Jan 1st, 2020 and June 30th, 2020 should be included in the summary.** In the end we expect results for about 5-20 patients, each patient in their own file.

Please provide your code and output results to us when complete.

Here is an example of the expected result for one patient.

File name: “patient_ID29057550656.json”
File contents:

```
{
  "frist_name": "Elliot",
  "last_name": "Ness",
  "med_summary": {
    "glimepiride": {
```

```
      "allowed_amount": 16.0,
      "days_supply": 120.0
    },
    "metformin hydrochloride": {
      "allowed_amount": 10.0,
      "days_supply": 90.0
    },
    "rosuvastatin calcium": {
      "allowed_amount": 23.92,
      "days_supply": 120.0
    },
    "valsartan-hydrochlorothiazide": {
      "allowed_amount": 52.480000000000004,
      "days_supply": 150.0
    }
  }
}
```


Detailed Instructions

We recommend breaking up the processing into several steps. For each step we may offer guidance and will have at least one question which we would like you to answer. Please indicate clearly in your code the answers to the questions (or provide a separate document with the answers, whichever you prefer)

Step 1 - Evaluate the enrollment file
Question 1: How many patients were enrolled in the program as of July 1st, 2020?

Step 2 - Load the pharmacy claims
Question 2: how many rows are there in the initial pharmacy claims data set?

Step 3 - Clean and prepare pharmacy claims

- Note: the columns: 'card_id', 'claim_number', 'fill_date', 'ndc_11' are sufficient to identify an unique fill.
- In this step please do any necessary data cleaning.
- Please also handle the various claim statuses
    - DENIED: should be completely/removed from subsequent steps
    - PAID: retain and include all paid claims **unless there is an associated reversal**
    - REVERSAL: a reversed claim may cancel out a paid claim; the columns: 'card_id', 'claim_number', 'fill_date', 'ndc_11' are sufficient to identify an unique fill. As such, they are sufficient to associate paid and reversal claims (if a

reversal claim has the same card_id, claim_number, fill_date and ndc_11 values, then that reversal corresponds to that paid line). The paid and reversal claims should be combined together to find the net 'allowed_amount' and net 'days_supply'. If the net days_supply is <= 0, then the patient did not receive that particular medication and it should not be included in subsequent steps.

- After combining the paid and reversal claims, **all claims with days_supply <= 0 should be removed**.

Example of claims before and after preparation:

Initial Claims

card_id	claim_number	fill_date	ndc_11	claim_status	days_supply
101	2000	2020-04-05	42192033101	DENIED	7
101	2001	2020-04-06	68645054559	PAID	28
101	2001	2020-04-06	68645054559	REVERSAL	-28
101	2002	2020-04-07	59746021701	PAID	30

Properly Prepared Claims

card_id	claim_number	fill_date	ndc_11	claim_status	days_supply
101	2002	2020-04-07	59746021701	PAID	30

Question 3: How many prepared claims do you have at the end of step 3?

Step 4: Find the total days supply and amount allowed in the specified date range for each patient by generic name.

- The generic name can be found in the provided "ndc9_lookup.json" file. To convert from ndc_11 to ndc_9 just drop the last two digits (e.g. ndc_11: '12345-6789-01' becomes ndc_9: '12345-6789')
- Only include fills that occurred **between Jan 1st, 2020 and June 30th, 2020**

Question 4: What is the highest amount_allowed? Which patient and generic drug does it correspond to?

Step 5: Create a JSON summary file for each patient enrolled on July 1st, 2020 that we have pharmacy data for.

Each patient summary should conform to the following points (see above for an example):

1. Each patient should have their own individual JSON file called 'patient_<id_number>.json'
2. The json should contain the fields: "first_name", "last_name" and "med_summary"
3. The med_summary should be a json with the following properties:
    a. Only medication "fills" that occurred **between Jan 1st, 2020 and June 30th, 2020 should be included in the summary**
    b. There should be a key for each unique generic medication name in the relevant time period
    c. Each generic medication name should have two aggregates included in a json: a key, value pair for "allowed_amount" and "days_supply"
        i. The allowed_amount should be the total of all the allowed_amount values for the relevant valid claims
        ii. The days_supply should also similarly be the total of all the days_supply for the relevant valid claims

Question 5: How many unique generic names for the patient Abe Lincoln?