What am I doing?
Health Gen
Error check
Insanely open ended

Where did i **TRY** get the data from
HealthData.govcensus.gov : https://www.archives.gov/research/census/download <- ABHIK SIR RECOMMENDATION

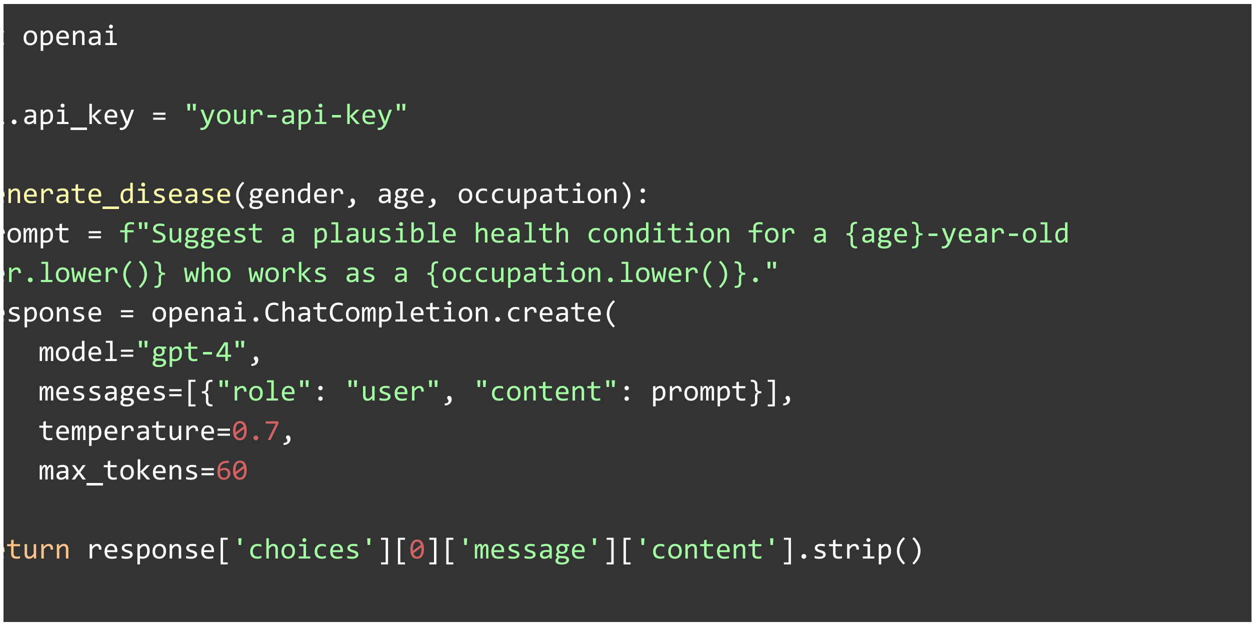
ICPSR: https://www.icpsr.umich.edu/web/pages/ICPSR/
https://catalog.archives.gov/
🏥 1. CDC National Youth Tobacco Survey (NYTS) – Smoking Campaigns
🌡️ 2. CDC Chronic Disease & Health Behavior
🍏 3. Nutrition, Physical Activity, Obesity (HealthData.gov
📊 4. PLACES: Local Health Data (GIS-Friendly)

Some trackbacks
Hashmaps?
I manually created 300 first names and 300 last names and did the cross product giving me 90000 names **but gender???**
Occupations : *["Software Engineer", "Data Scientist", "Web Developer", "Network Administrator", "Cybersecurity Analyst", "AI Researcher", "UX Designer", "Graphic Designer", "Mechanical Engineer", "Civil Engineer", "Electrical Engineer", "Chemical Engineer", "Biomedical Engineer", "Architect", "Urban Planner", "Doctor", "Nurse", "Pharmacist", "Dentist", "Surgeon", "Psychiatrist", "Psychologist", "Veterinarian", "Paramedic", "Medical Lab Technician", "Radiologist", "Physical Therapist", "Occupational Therapist", "Speech Therapist", "Nutritionist", "Teacher", "Professor", "School Counselor", "Principal", "Librarian", "Curriculum Designer", "Researcher", "Scientist", "Chemist", "Biologist", "Astronomer", "Physicist", "Geologist", "Environmental Scientist", "Economist", "Statistician", "Financial Analyst", "Accountant", "Auditor", "Investment Banker", "Marketing Manager", "Sales Executive", "SEO Specialist", "Social Media Manager", "Business Analyst", "Project Manager", "Human Resources Manager", "Recruiter", "Lawyer", "Judge", "Paralegal", "Police Officer", "Detective", "Firefighter", "Soldier", "Pilot", "Flight Attendant", "Air Traffic Controller", "Train Conductor", "Bus Driver", "Truck Driver", "Taxi Driver", "Chef", "Baker", "Waiter/Waitress", "Bartender", "Hotel Manager", "Housekeeper", "Real Estate Agent", "Construction Worker", "Plumber", "Electrician", "Carpenter", "Welder", "Mechanic", "Factory Worker", "Warehouse Manager", "Delivery Driver", "Farmer", "Fisherman", "Photographer", "Videographer", "Actor", "Musician", "Dancer", "Painter", "Sculptor", "Fashion Designer", "Tailor", "Content Creator"]*

```
import requests

def get_gender_api(name):
    response = requests.get(f"https://api.genderize.io/?name={name}")
    if response.status_code != 200:
        return response.json().get("gender", "Unknown")
    return "Unknown"
NUMBER_OF_CALLS_PER_DAY = 100
```

ojasb> java -cp ".;.\json-20250517.jar" RandomUserCSV

OPEN AI : api calls
Bitter Gourd : did not give me even 5 calls before i ran out of token
I will upgrade if i have to : BUT I AM BROKE

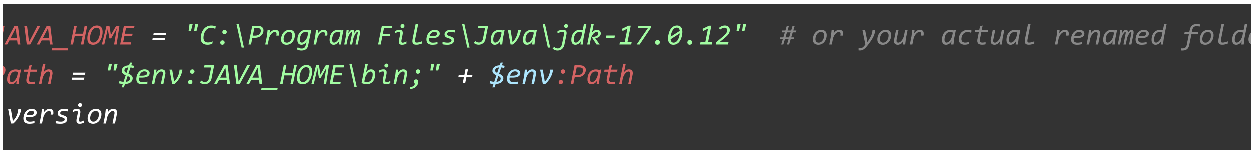


```
 openai

.api_key = "your-api-key"

nerate_disease(gender, age, occupation):
ompt = f"Suggest a plausible health condition for a {age}-year-old
r.lower()} who works as a {occupation.lower()}."
sponse = openai.ChatCompletion.create(
  model="gpt-4",
  messages=[{"role": "user", "content": prompt}],
  temperature=0.7,
  max_tokens=60

turn response['choices'][0]['message']['content'].strip()
```

<u>SYNTHEA MY SAVIOR</u>

https://github.com/synthetichealth/synthea

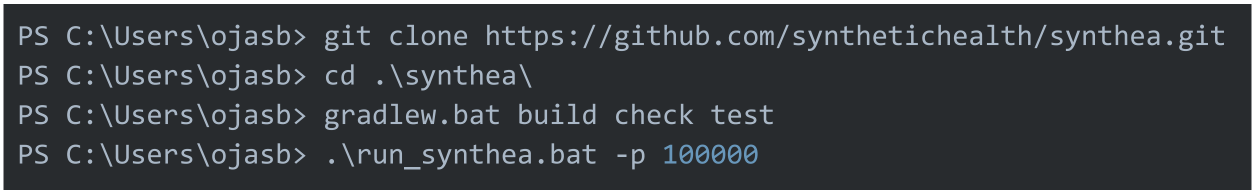
Synthea build code
Synthea - bitter gourd
For the better or worse it is not random i have all the 100k from massachusetts
Uses only Java17 not any other version of jdk
What do my IDEs use ? - java 21
Switching between java 17 and java 21



```
AVA_HOME = "C:\Program Files\Java\jdk-17.0.12"  # or your actual renamed fold
ath = "$env:JAVA_HOME\bin;" + $env:Path
version
```

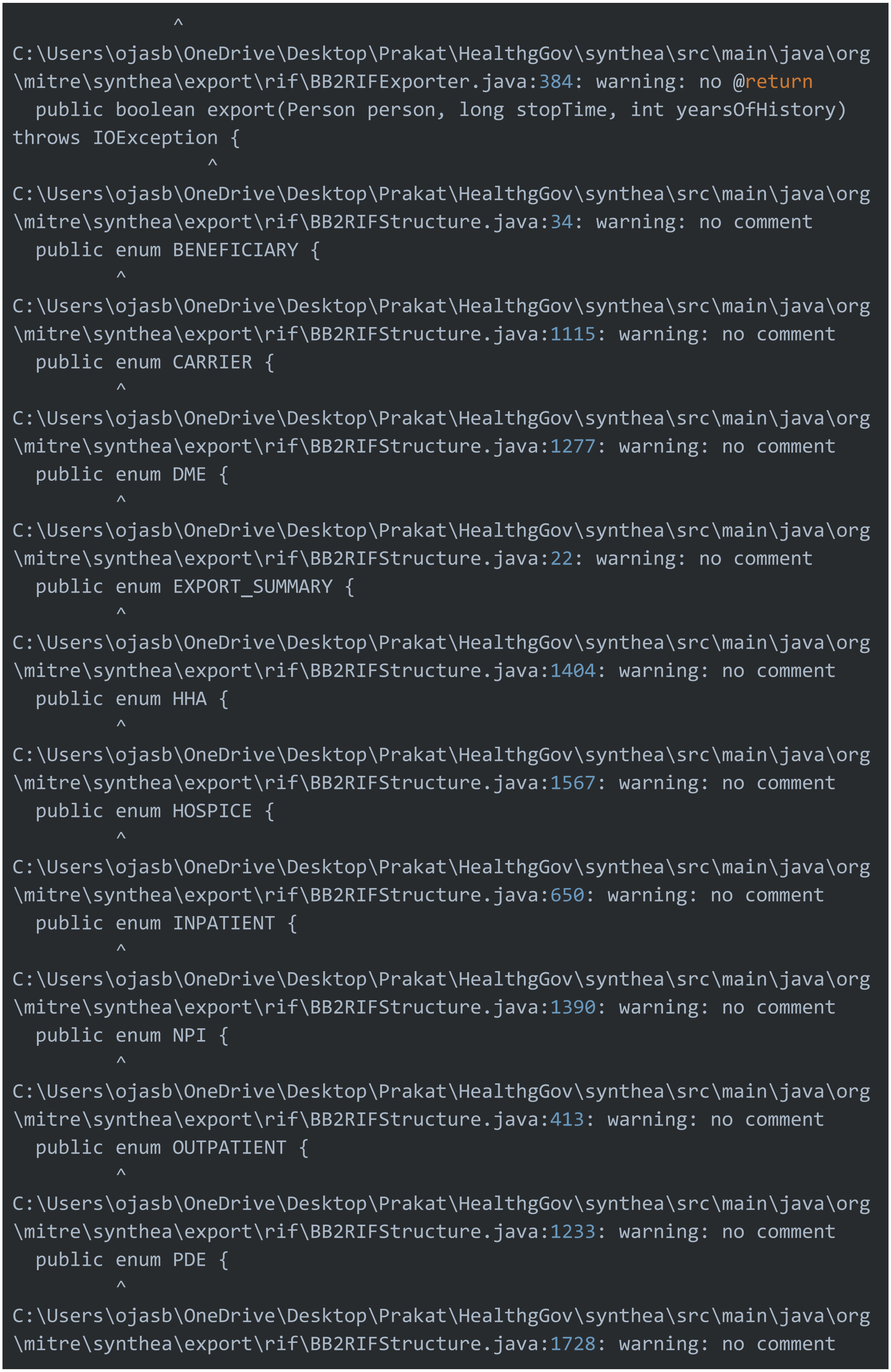
½ -> 1 seconds per patient data : 100,000* 1
Total time : 22.4 hours approx
Synthea - chocolate cake
Gives any number of data requested
Output in json format

Build and run SYNTHEA

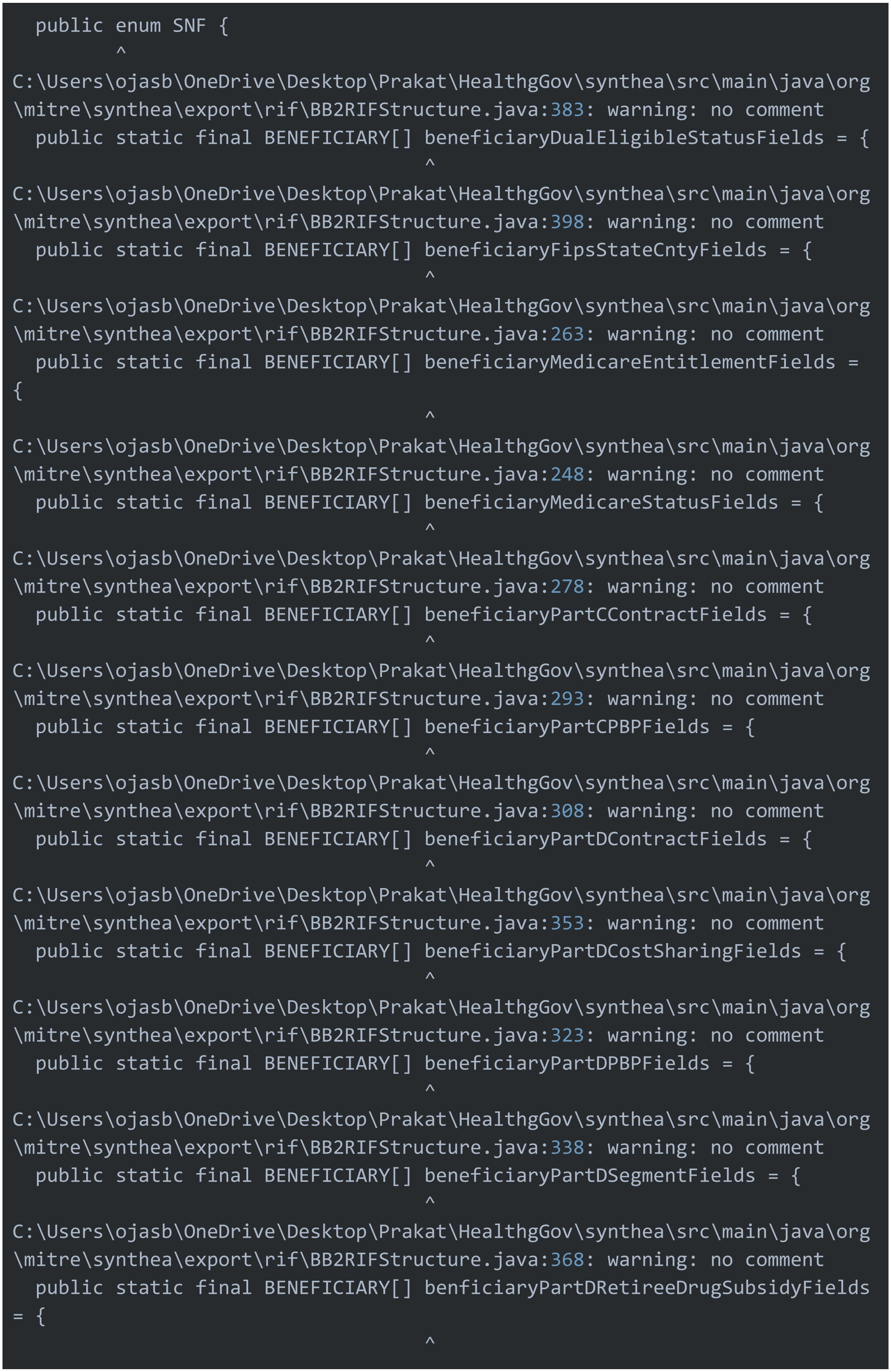


```
PS C:\Users\ojasb> git clone https://github.com/synthetichealth/synthea.git
PS C:\Users\ojasb> cd .\synthea\
PS C:\Users\ojasb> gradlew.bat build check test
PS C:\Users\ojasb> .\run_synthea.bat -p 100000
PS C:\Users\ojasb> py .\ExtractInfo.py
```

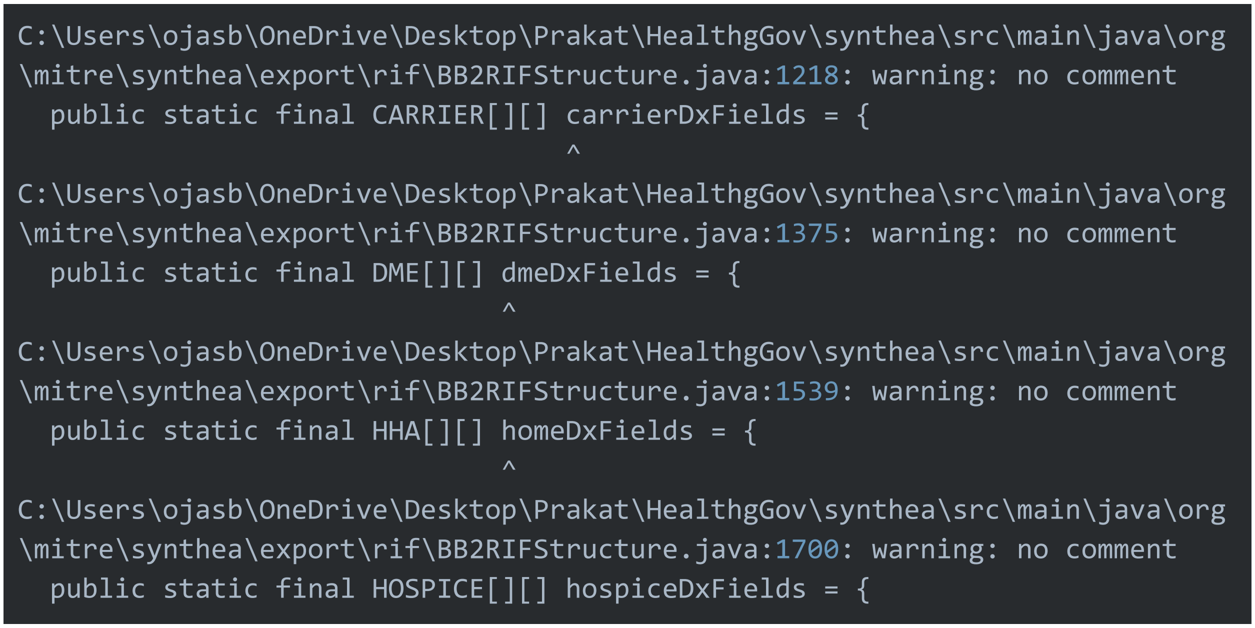
```
PS C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthGov\synthea> .\gradlew.bat
build check test
Starting a Gradle Daemon, 1 incompatible Daemon could not be reused, use
--status for details

> Task :compileJava
Note: Some input files use unchecked or unsafe operations.
Note: Recompile with -Xlint:unchecked for details.

> Task :javadoc
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\export\CDWExporter.java:44: warning: empty <p> tag
 * <p></p>
      ^
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\world\concepts\ClinicianSpecialty.java:8: warning: empty <p>
tag
 * <p></p>
      ^
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\world\concepts\ClinicianSpecialty.java:10: warning: empty
<p> tag
 * <p></p>The numeric value in each cell identifies how many of clinicians
should be
      ^
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\App
.java:24: warning: no comment
public class App {
       ^
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\modules\calculators\ASCVD.java:12: warning: no comment
public class ASCVD {
       ^
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\modules\calculators\ASCVD.java:14: warning: no comment
  public static final long TEN_YEARS_IN_MS = TimeUnit.DAYS.toMillis(3650);
                           ^
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\helpers\Attributes.java:54: warning: no comment
  public class Inventory {
         ^
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\export\rif\BB2RIFExporter.java:354: warning: no @throws for
java.io.IOException
  public void exportMissingCodes() throws IOException {
```



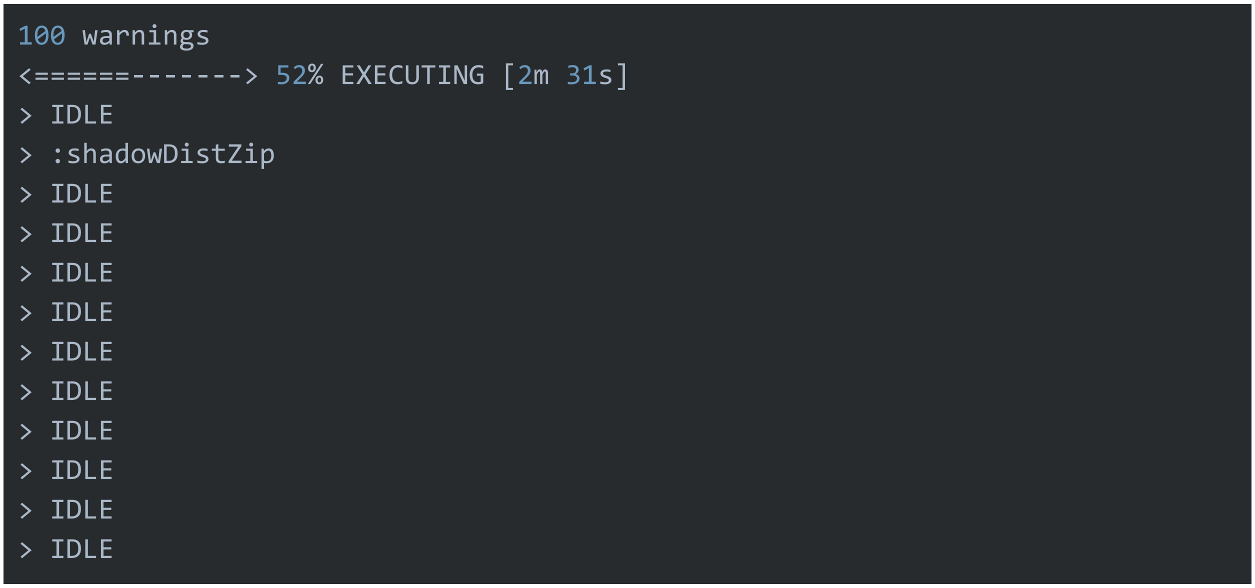
```
              ^
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\export\rif\BB2RIFExporter.java:384: warning: no @return
  public boolean export(Person person, long stopTime, int yearsOfHistory)
throws IOException {
                 ^
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\export\rif\BB2RIFStructure.java:34: warning: no comment
  public enum BENEFICIARY {
         ^
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\export\rif\BB2RIFStructure.java:1115: warning: no comment
  public enum CARRIER {
         ^
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\export\rif\BB2RIFStructure.java:1277: warning: no comment
  public enum DME {
         ^
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\export\rif\BB2RIFStructure.java:22: warning: no comment
  public enum EXPORT_SUMMARY {
         ^
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\export\rif\BB2RIFStructure.java:1404: warning: no comment
  public enum HHA {
         ^
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\export\rif\BB2RIFStructure.java:1567: warning: no comment
  public enum HOSPICE {
         ^
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\export\rif\BB2RIFStructure.java:650: warning: no comment
  public enum INPATIENT {
         ^
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\export\rif\BB2RIFStructure.java:1390: warning: no comment
  public enum NPI {
         ^
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\export\rif\BB2RIFStructure.java:413: warning: no comment
  public enum OUTPATIENT {
         ^
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\export\rif\BB2RIFStructure.java:1233: warning: no comment
  public enum PDE {
         ^
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\export\rif\BB2RIFStructure.java:1728: warning: no comment
```



```
  public enum SNF {
         ^
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\export\rif\BB2RIFStructure.java:383: warning: no comment
  public static final BENEFICIARY[] beneficiaryDualEligibleStatusFields = {
                                    ^
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\export\rif\BB2RIFStructure.java:398: warning: no comment
  public static final BENEFICIARY[] beneficiaryFipsStateCntyFields = {
                                    ^
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\export\rif\BB2RIFStructure.java:263: warning: no comment
  public static final BENEFICIARY[] beneficiaryMedicareEntitlementFields =
{
                                    ^
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\export\rif\BB2RIFStructure.java:248: warning: no comment
  public static final BENEFICIARY[] beneficiaryMedicareStatusFields = {
                                    ^
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\export\rif\BB2RIFStructure.java:278: warning: no comment
  public static final BENEFICIARY[] beneficiaryPartCContractFields = {
                                    ^
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\export\rif\BB2RIFStructure.java:293: warning: no comment
  public static final BENEFICIARY[] beneficiaryPartCPBPFields = {
                                    ^
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\export\rif\BB2RIFStructure.java:308: warning: no comment
  public static final BENEFICIARY[] beneficiaryPartDContractFields = {
                                    ^
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\export\rif\BB2RIFStructure.java:353: warning: no comment
  public static final BENEFICIARY[] beneficiaryPartDCostSharingFields = {
                                    ^
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\export\rif\BB2RIFStructure.java:323: warning: no comment
  public static final BENEFICIARY[] beneficiaryPartDPBPFields = {
                                    ^
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\export\rif\BB2RIFStructure.java:338: warning: no comment
  public static final BENEFICIARY[] beneficiaryPartDSegmentFields = {
                                    ^
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\export\rif\BB2RIFStructure.java:368: warning: no comment
  public static final BENEFICIARY[] benficiaryPartDRetireeDrugSubsidyFields
= {
                                    ^
```



```
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\export\rif\BB2RIFStructure.java:1218: warning: no comment
  public static final CARRIER[][] carrierDxFields = {
                                  ^
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\export\rif\BB2RIFStructure.java:1375: warning: no comment
  public static final DME[][] dmeDxFields = {
                              ^
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\export\rif\BB2RIFStructure.java:1539: warning: no comment
  public static final HHA[][] homeDxFields = {
                              ^
C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\src\main\java\org
\mitre\synthea\export\rif\BB2RIFStructure.java:1700: warning: no comment
  public static final HOSPICE[][] hospiceDxFields = {
```



```
100 warnings
<======-------> 52% EXECUTING [2m 31s]
> IDLE
> :shadowDistZip
> IDLE
> IDLE
> IDLE
> IDLE
> IDLE
> IDLE
> IDLE
> IDLE
> IDLE
> IDLE
```

What info can i ask for and what can it give

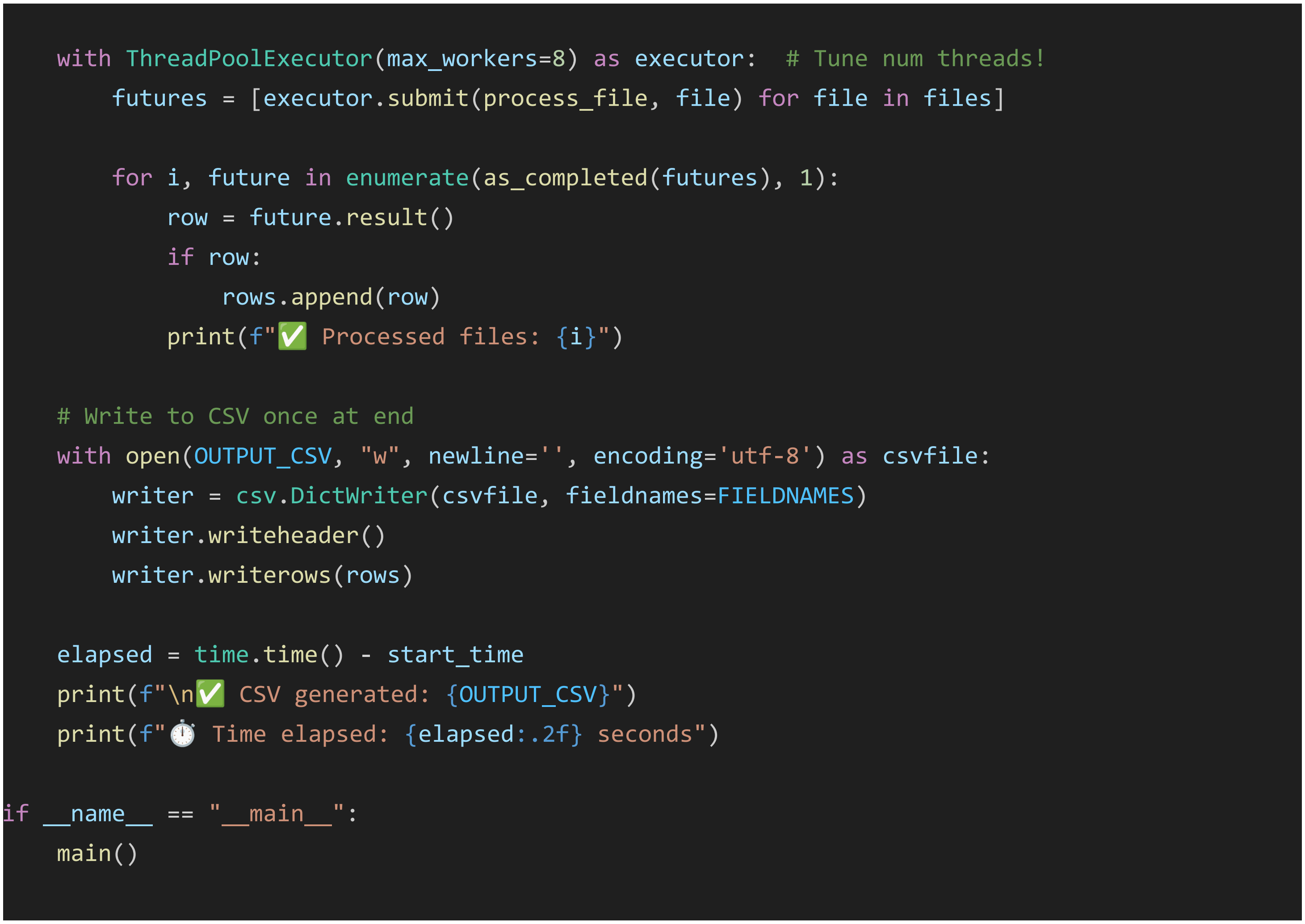
Output?
PatientID, FirstName, LastName, Gender, BirthDate, MaritalStatus, Race, Ethnicity, BirthSex, MotherMaidenName, BirthPlace_City, BirthPlace_State, BirthPlace_Country, Address_Line, Address_City ,Address_State, Address_PostalCode, Address_Country ,Phone,Language, SSN, DriverLicenseNumber, PassportNumber, DisabilityAdjustedLifeYears, QualityAdjustedLifeYears, EncounterCount, FirstEncounterDate, LastEncounterDate, ConditionCount, Conditions, MedicationCount, Medications, ObservationCount, Observations, ProcedureCount, Procedures, ImmunizationCount, Immunizations, CarePlanCount, CarePlans, ClaimCount, TotalClaimedAmount, InsuranceProvider

Need?
PatientID, FirstName, LastName, Gender, BirthDate, MaritalStatus, Ethnicity, BirthSex, BirthPlace_City, Address_Line, Address_City, Address_PostalCode, Phone, SSN, DriverLicenseNumber, PassportNumber, DisabilityAdjustedLifeYears, QualityAdjustedLifeYears, ConditionCount, Conditions, MedicationCount, Medications, ObservationCount, Observations, ProcedureCount, Procedures, ImmunizationCount, Immunizations,CarePlanCount, CarePlans, ClaimCount, TotalClaimedAmount, InsuranceProvider

How can i extract the info from json: I DONNO JSON!!!

// not my code

```
import json
import os
import csv
import time
from concurrent.futures import ThreadPoolExecutor, as_completed

# === CONFIG ===
FHIR_DIR = r"C:\Users\ojasb\OneDrive\Desktop\Prakat\HealthgGov\synthea\output\fhir"
OUTPUT_CSV = "synthea_output.csv"

# === CSV columns ===
FIELDNAMES = [
    "PatientID", "FirstName", "LastName", "Gender", "BirthDate", "MaritalStatus", "Ethnicity",
    "BirthSex", "BirthPlace_City", "Address_Line", "Address_City", "Address_PostalCode",
    "Phone", "SSN", "DriverLicenseNumber", "PassportNumber",
    "DisabilityAdjustedLifeYears", "QualityAdjustedLifeYears",
    "ConditionCount", "Conditions", "MedicationCount", "Medications",
    "ObservationCount", "Observations",
    "ClaimCount", "TotalClaimedAmount", "InsuranceProvider"
]

def process_file(filepath):
    # Same logic you already have inside the for loop
    with open(filepath, "r", encoding='utf-8') as f:
        data = json.load(f)

    patient_resource = None
    for entry in data.get("entry", []):
        res = entry["resource"]
        if res.get("resourceType") == "Patient":
            patient_resource = res
            break

    if not patient_resource:
        return None  # Skip

    row = {field: "" for field in FIELDNAMES}
    row["PatientID"] = patient_resource.get("id", "")
    name = patient_resource.get("name", [{}])[0]
    row["FirstName"] = name.get("given", [""])[0] if name.get("given") else ""
    row["LastName"] = name.get("family", "")
    row["Gender"] = patient_resource.get("gender", "")
    row["BirthDate"] = patient_resource.get("birthDate", "")
    row["MaritalStatus"] = patient_resource.get("maritalStatus", {}).get("text", "")
    row["Address_Line"] = patient_resource.get("address", [{}])[0].get("line", [""])[0] if
patient_resource.get("address") else ""
    row["Address_City"] = patient_resource.get("address", [{}])[0].get("city", "")
    row["Address_PostalCode"] = patient_resource.get("address", [{}])[0].get("postalCode", "")
    row["Phone"] = patient_resource.get("telecom", [{}])[0].get("value", "")

    for ident in patient_resource.get("identifier", []):
        sys = ident.get("system", "")
        if "us-ssn" in sys:
            row["SSN"] = ident.get("value", "")
        elif "4.3.25" in sys:
            row["DriverLicenseNumber"] = ident.get("value", "")
        elif "passport" in sys:
            row["PassportNumber"] = ident.get("value", "")

    for ext in patient_resource.get("extension", []):
        url = ext["url"]
        if "ethnicity" in url:
            row["Ethnicity"] = ext["extension"][1]["valueString"]
        elif "birthsex" in url:
            row["BirthSex"] = ext["valueCode"]
        elif "birthPlace" in url:
            row["BirthPlace_City"] = ext["valueAddress"]["city"]
        elif "disability-adjusted-life-years" in url:
            row["DisabilityAdjustedLifeYears"] = ext["valueDecimal"]
        elif "quality-adjusted-life-years" in url:
            row["QualityAdjustedLifeYears"] = ext["valueDecimal"]

    conditions = []
    medications = []
    observations = []
    total_claimed_amount = 0.0
    insurance_providers = set()
    claim_count = 0

    for entry in data.get("entry", []):
        res = entry["resource"]
        rtype = res.get("resourceType")
        if rtype == "Condition":
            code = res.get("code", {}).get("text") or \
                   res.get("code", {}).get("coding", [{}])[0].get("display", "")
            if code:
                conditions.append(code)

        elif rtype == "MedicationRequest":
            med = res.get("medicationCodeableConcept", {}).get("text") or \
                  res.get("medicationCodeableConcept", {}).get("coding",
[{}])[0].get("display", "")
            if med:
                medications.append(med)

        elif rtype == "Observation":
            obs = res.get("code", {}).get("text") or \
                  res.get("code", {}).get("coding", [{}])[0].get("display", "")
            if obs:
                observations.append(obs)

        elif rtype == "Claim":
            claim_count += 1
            amt = res.get("total", {}).get("value", 0.0)
            total_claimed_amount += amt
            for ins in res.get("insurance", []):
                insurance_providers.add(ins.get("coverage", {}).get("display", ""))

    row["ConditionCount"] = len(conditions)
    row["Conditions"] = "; ".join(conditions)
    row["MedicationCount"] = len(medications)
    row["Medications"] = "; ".join(medications)
    row["ObservationCount"] = len(observations)
    row["Observations"] = "; ".join(observations)
    row["ClaimCount"] = claim_count
    row["TotalClaimedAmount"] = round(total_claimed_amount, 2)
    row["InsuranceProvider"] = "; ".join(list(insurance_providers))

    return row


def main():
    start_time = time.time()
    files = [os.path.join(FHIR_DIR, f) for f in os.listdir(FHIR_DIR) if f.endswith(".json")]
    rows = []
```



```
    with ThreadPoolExecutor(max_workers=8) as executor:  # Tune num threads!
        futures = [executor.submit(process_file, file) for file in files]

        for i, future in enumerate(as_completed(futures), 1):
            row = future.result()
            if row:
                rows.append(row)
            print(f"✅ Processed files: {i}")

    # Write to CSV once at end
    with open(OUTPUT_CSV, "w", newline='', encoding='utf-8') as csvfile:
        writer = csv.DictWriter(csvfile, fieldnames=FIELDNAMES)
        writer.writeheader()
        writer.writerows(rows)

    elapsed = time.time() - start_time
    print(f"\n✅ CSV generated: {OUTPUT_CSV}")
    print(f"⏱️ Time elapsed: {elapsed:.2f} seconds")

if __name__ == "__main__":
    main()
```

If one thousand files = 1 min 20 secs
Total time : 2 hours 37 mins
Simple maths ?? NO NO NO unpredictable code compilation

I looked at the output and there were some mismatches so i either will have to manually change things or write code to clean

Solution : XLSX
Output XLSX DONE

Clean output - DONE

Next step : plotting , prolly pinch of genAI in it R