Ansatz 1. Verwendung von sqlite3 in Python

1. Aufbereitung der Quelldaten: das Gleiche wie in Ansatz 2.
2. Erstellung der Dateien „facts\_table“ und „datawarehouse.db“.
3. INSERT INTO SQL-Anweisungen zum Auffüllen der facts\_table. Da dies nicht funktionierte und beim Generieren von Abbildungen Verknüpfungen in Abfragen neu erstellt werden mussten, hatten wir eine zweite Option: Ansatz 2.

Weitere Informationen finden Sie in der Datei Datenbank.ipynb

Ansatz 2. Verwendung von Tableau Prep

1. Aufbereitung der Quelldaten (Abbildung 1 in Wiki ). Rohdaten “csv\_data” von sechs Tabellen (Patienten, Begegnungen, Impfungen, Medikamente, Beobachtungen und Verfahren) für jede Krebsart (Brust, Darm und Lunge) wurden in einer einzigen CSV-Datei für jede Tabelle zusammengefasst, die Daten für die drei Krebsarten enthält. Weitere Informationen finden Sie im Ordner Tableau Data Flows.
2. Datenquellentabellen verbinden in Facts\_table.csv (Abbildung 2 in Wiki) Mit den Daten aus den sechs Tabellen im Ordner „data\_source“ haben wir die folgenden Verknüpfungen gestartet:
   1. Erster Join (zwischen Patienten und Begegnungen): LEFT JOIN, wobei alle Patienten in der Datei „patients.csv“ beibehalten werden, die eine oder mehrere Begegnungen in der Datei „encounters.csv“ haben.
   2. Zweiter Join (zwischen dem Ergebnis des ersten Joins und den Immunisierungen): LEFT JOIN, wobei der Patientensatz aus der Datei „patients.csv“ beibehalten wird.
   3. Dritte Join (zwischen vorheriger Verbindung und Medikamenten): same.
   4. Vierter Join (zwischen vorherigem Join und Beobachtungen): Hier wurde eine Feldberechnung hinzugefügt, um eine Spalte zu generieren, die den letzten Beobachtungswert pro Patient und Beobachtungstyp speichert. Beispielsweise wurde von 9 QALY-Werten nur der letzte mit Datum im Jahr 2020 in den Ergebnissen beibehalten.
   5. Fünfter Join (zwischen dem vorherigen Join und den Prozeduren): Derselbe Patientensatz aus der ersten Tabelle „patients.csv“ wurde bis zu diesem Join beibehalten und nur durch die anderen Spalten in den anderen Tabellen angereichert, die uns bei der Beantwortung unserer Forschungsfragen helfen. Die Datei Facts\_table.csv enthält mehr Spalten als die Facts\_table aus dem Sternschema, falls weitere Fragen beantwortet werden sollen, beispielsweise zu Geschlecht, Rasse und ethnischer Zugehörigkeit.