**Data Analysis for CDM Claims data with Socio-Economic Status Data from 2007 to 2020 Q2**

**Addiction Prediction data load**

**Folder location** **–** optum -> data -> parquet -> ses\_80 (version – 80)

**Tables loaded –**

1. Confinement
2. Medical diagnosis
3. Lab result
4. Medical
5. Procedure
6. RX

**Look-up tables**

1. LU diagnosis
2. LU NDC
3. LU Procedure

**Member and provider details**

1. Member continuous enrollment
2. Member enrollment
3. Provider bridge
4. Provider

Data load issues - LU NDC, LU Procedure returned an error - ArrowInvalid: Casting from timestamp[ms] to timestamp[ns] would result in out of bounds timestamp: 253402214400000 – while loading, used engine – fastparquet instead of pyarrow

**Data load process:**

1. Read **MEMBER\_ENROLLMENT** table (119,543,915 rows) to get a list of all **Patient IDs** (PATID, 68,233,646 patients)
2. Created a **YEAR** column from **ELIGEFF** column (The date this member coverage row of information is effective) and counted the number of occurrences each patient has over the years. Filtered patients with top 10 occurrences.
3. Read the tables – **Confinement, Medical diagnosis, Lab result, Medical, Procedure, RX** – for top 2 patients from 2007 to 2020 Q2.
4. Merged all the above tables based on their respective unique keys.

|  |  |  |
| --- | --- | --- |
| **Table** | **Merged On** | **Table Description** |
| CONFINEMENT | PATID, PAT\_PLANID | In patient stays |
| MED\_DIAGNOSIS | PATID, PAT\_PLANID | Claim diagnosis codes |
| LABRESULT | PATID, PAT\_PLANID | Claim lab results |
| MEDICAL | PATID, PAT\_PLANID, CLMID, PROV | Claim detail information |
| PROCEDURE | PATID, PAT\_PLANID, CLMID | Claim procedure codes |
| RX | PATID, PAT\_PLANID, CLMID | Claim prescriptions |
| LU\_DIAGNOSIS | ICD\_VER\_CD | Lookup detail on ICD/9/10 |
| LU\_NDC |  | Lookup detail on drug codes |
| LU\_PROCEDURE |  | Lookup detail on procedures |
| MEMBER\_CONTINUOUS\_ENROLLMENT | PATID | Member continuous enrollment |
| MEMBER\_ENROLLMENT | PATID, PAT\_PLANID, | Member detailed enrollment |
| PROVIDER\_BRIDGE | PROV\_UNIQUE | Anonymized provider ids |
| PROVIDER | PROV\_UNIQUE | Anonymized provider details |

**Data description**

1. **Member eligibility tables -** info on every member enrolled with the health plan during the specified data period.
2. ***Member Enrollment –*** contains columns like year of birth, gender, state, race. Also includes enrollment specific info such as eligibility dates, product, encrypted policy number, etc. New row is added each time any information about the member changes, such as state or product.
3. ***Member Continuous Enrollment -*** rollup of above table. Each row – span of continuous enrollment (less than 30-day break in coverage) regardless of changes in coverage. May have multiple rows per member. Breaks in eligibility will trigger a new row, representing different coverage periods. It contains columns that generally do not change over time (Year of Birth, Gender, Race).

Both tables have same info, but continuous enrollment will have info of period with continuous plan, whereas enrollment table has same info but adds new whenever any info changes, so one row in continuous enrollment can be represented by multiple rows in enrollment table.

**Example:**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Member Continuous Enrollment** | | | | |  | **Member Enrollment** | | | | |
| PATID | ELIGEFF | ELIGEND | GDR\_CD | YRDOB |  | PATID | ELIGEFF | ELIGEND | STATE | PROD |
| A1234 | 1/1/2005 | 12/31/2007 | F | 1950 |  | A1234 | 1/1/2005 | 8/31/2006 | LA | PPO |
| A1234 | 6/1/2008 | 12/31/2010 | F | 1950 |  | A1234 | 9/01/2006 | 12/31/2006 | TX | PPO |
|  |  |  |  |  |  | A1234 | 1/1/2007 | 12/31/2007 | TX | HMO |
|  |  |  |  |  |  | A1234 | 6/1/2008 | 12/31/2010 | TX | HMO |

Joining with other tables:

1. ***Member Enrollment –*** PATID
2. ***Member Continuous Enrollment -*** PAT\_PLANID (identifies both member and policy of the eligibility table)
3. **Medical claims table -** medical claims data for inpatient and outpatient professional services
4. **Medical diag, medical proc tables –** info about diagnosis and procedure codes. A medical claim can have up to 25 diag codes and 25 facility proc codes. To simplify searching for specific diagnosis/procedure codes and to reduce the width of the medical claim record, the diagnosis and procedures are in separate tables (***Med Diag*** and ***Med Proc***). One code per row.

Joining with medical claim table:

1. ***Med Diag -*** PATID, PAT\_PLANID, CLMID, FST\_DT
2. ***Med Proc -*** PATID, PAT\_PLANID, CLMID, FST\_DT, LOC\_CD
3. **Pharmacy claims table (RX) -** claims submitted by pharmacies for prescriptions filled on an outpatient basis.
4. **Laboratory Test** **results table** - lab test results for all available lab tests. It only contains lab tests performed within certain lab networks. So, it will not contain every lab result for every member.
5. **Inpatient Confinement Table -** a derived, summarized record for each inpatient episode occurring in an acute care hospitalization or skilled nursing facility setting. Unique record for every hospitalization.
6. Provider tables

Graphical user interface, diagram

Description automatically generated

1. **Reference tables**

|  |  |
| --- | --- |
| LU\_Diagnosis | Diagnosis Lookup Table – Descriptions and Categories for Diagnosis Codes |
| LU\_NDC | National Drug Code (NDC) Lookup table – Information about drugs from the FirstData Bank NDC table. Each AHFSCLSS code can have multiple rows due to different brands, drug strength and other fearures. |
| LU\_Procedure | Procedure Lookup Table – Descriptions and Categories for Procedure Codes |

**Database structure**



Most tables have *both* member ID fields, PAT\_ID and PAT\_PLANID. Exceptions: MEMBER\_CONTINUOUS\_ENROLLMENT, SES, DOD. The tables with only PATID allows for information to be gathered by member.

