DATA SET Test\_Beneficiarydata-1542969243754

Dataset Test\_Beneficiarydata-1542969243754.csv zawiera:

* **63 968 wierszy** (czyli tylu pacjentów – beneficjentów),
* **25 kolumn** (czyli różnych zmiennych, np. płeć, wiek, choroby, koszty).

**🧾 Kolumny i ich znaczenie**

**Dane demograficzne**

|  |  |
| --- | --- |
| **Kolumna** | **Znaczenie** |
| **BeneID** | Unikalny identyfikator beneficjenta (pacjenta). |
| **DOB** | Data urodzenia pacjenta. Używane do obliczenia wieku. |
| **DOD** | Data zgonu pacjenta. Pomaga wykrywać nadużycia (np. roszczenia po śmierci). |
| **Gender** | Płeć: 1 = Male, 2 = Female. |
| **Race** | Rasa pacjenta: 1 = White, 2 = Black, 3 = Other, 4 = Asian, 5 = Native American. |
| **RenalDiseaseIndicator** | Czy pacjent ma chorobę nerek: Y = tak, N = nie, 0 = brak danych. |
| **State** | Kod stanu (USA), w którym mieszka pacjent. |
| **County** | Kod hrabstwa (lokalizacji) pacjenta. |
| **NoOfMonths\_PartACov** | Liczba miesięcy w roku, przez które pacjent był objęty ubezpieczeniem Medicare Part A (szpitalnym). |
| **NoOfMonths\_PartBCov** | Liczba miesięcy w roku, przez które pacjent był objęty Medicare Part B (ambulatoryjnym). |

**🩺 Choroby przewlekłe (wartości: 1 = ma chorobę, 2 = brak choroby)**

|  |  |
| --- | --- |
| **Kolumna** | **Znaczenie** |
| **ChronicCond\_Alzheimer** | Czy pacjent ma chorobę Alzheimera. |
| **ChronicCond\_Heartfailure** | Niewydolność serca. |
| **ChronicCond\_KidneyDisease** | Choroba nerek. |
| **ChronicCond\_Cancer** | Nowotwór złośliwy. |
| **ChronicCond\_ObstrPulmonary** | Przewlekła obturacyjna choroba płuc (POChP). |
| **ChronicCond\_Depression** | Depresja. |
| **ChronicCond\_Diabetes** | Cukrzyca. |
| **ChronicCond\_IschemicHeart** | Choroba niedokrwienna serca. |
| **ChronicCond\_Osteoporasis** | Osteoporoza. |
| **ChronicCond\_rheumatoidarthritis** | Reumatoidalne zapalenie stawów. |
| **ChronicCond\_stroke** | Udar mózgu. |

**💵**

**Roczne koszty leczenia**

|  |  |
| --- | --- |
| **Kolumna** | **Znaczenie** |
| **IPAnnualReimbursementAmt** | Łączna kwota refundacji z Medicare za leczenie hospitalizacyjne (Inpatient). |
| **IPAnnualDeductibleAmt** | Łączna kwota udziału własnego (koszty ponoszone przez pacjenta) za Inpatient. |
| **OPAnnualReimbursementAmt** | Refundacja za leczenie ambulatoryjne (Outpatient). |
| **OPAnnualDeductibleAmt** | Udział własny pacjenta w kosztach leczenia Outpatient. |

Podsumowując:

Te kolumny można pogrupować w kategorie:

* **Dane demograficzne:** Gender, Race, DOB, State, County
* **Stan zdrowia:** RenalDiseaseIndicator, ChronicCond\_\*
* **Historia ubezpieczenia:** NoOfMonths\_PartA/B
* **Koszty:** IP/OPAnnual\*Amt

Dodatkowa legenda

**🗺️ Kody lokalizacyjne**

**✅ State**

* **Opis:** Dwucyfrowy kod stanu USA, w którym zamieszkuje beneficjent.
* **Format:** Liczby całkowite (01–56)
* **Przykład interpretacji:**
  + 01 = Alabama
  + 06 = Kalifornia
  + 36 = Nowy Jork
* **Źródło kodów:** U.S. Federal Information Processing Standards (FIPS codes)

📌 Pełna lista: <https://en.wikipedia.org/wiki/Federal_Information_Processing_Standard_state_code>

**✅ County**

* **Opis:** Kod hrabstwa (powiatu) w obrębie stanu.
* **Format:** 3-cyfrowa liczba, zgodna z FIPS County Code.
* **Znaczenie:** Pozwala na bardziej szczegółowe określenie miejsca zamieszkania pacjenta.
* **Uwaga:** Kod 230 w stanie 39 może np. oznaczać Franklin County w Ohio (39 = Ohio, 230 = Franklin).

📌 Pełna lista (fips state + county): https://www.census.gov/library/reference/code-lists/ansi.html