1. **Train-1542865627584.csv**

This file provides training labels at the provider level, which indicate whether a healthcare provider is fraudulent or not.

Columns Description **-**

|  |  |
| --- | --- |
| **Column Name** | **Description** |
| Provider | A unique identifier for a healthcare provider (e.g., hospital, clinic, or medical practice). |
| PotentialFraud | Target variable. It indicates whether the provider is suspected of fraud:  - 'Yes': Provider is labeled as potentially fraudulent.  - 'No': Provider is not labeled as fraudulent. |

The fraud label is at the provider level, not claim level. That means:

* Individual claims are not labeled as fraudulent or not.
* Instead, you must analyze patterns across multiple claims (outpatient/inpatient) submitted by each provider and determine if that provider seems suspicious.

1. **Train\_Outpatientdata-1542865627584.csv**

What is Outpatient claim? - An outpatient claim refers to a request for reimbursement made by a healthcare provider to an insurance company for medical services given to a patient who was not admitted to a hospital. It includes –

* Doctor visits
* Diagnostic tests (e.g., X-rays, blood tests)
* Minor surgeries
* Physical therapy
* Emergency room visits (if not admitted afterward)

Column explanation –

|  |  |
| --- | --- |
| **Column Name** | **Description** |
| Provider | Unique identifier for the healthcare provider (e.g., hospital or clinic). |
| BeneID | Beneficiary ID – unique identifier for the patient. |
| ClaimID | Unique identifier for the insurance claim. |
| ClaimStartDt | Start date of the claim (when the outpatient visit began). |
| ClaimEndDt | End date of the claim (when the outpatient visit ended). |
| ProcCode | Procedure code – standard CPT/HCPCS code describing the treatment provided. |
| DiagnosisCode\_1 to DiagnosisCode\_10 | Diagnosis codes (ICD codes) associated with the claim, representing the medical reasons for the treatment. |
| DeductibleAmtPaid | Amount paid by the beneficiary out-of-pocket before insurance coverage kicks in. |
| InscClaimAmtReimbursed | The amount the insurance reimbursed for this claim. |
| AttendingPhysician | ID of the primary physician attending the patient. |
| OperatingPhysician | ID of the operating physician (if surgery/procedure occurred). |
| OtherPhysician | ID of any other physician involved. |
| ClmAdmitDiagnosisCode | Primary diagnosis code provided at the time of admission. |
| ClmDiagnosisCode\_1 to ClmDiagnosisCode\_10 | Additional diagnosis codes used for claim classification. |
| ClaimType | Indicates whether this is an inpatient or outpatient claim (should be outpatient in this file). |

1. **Train\_Inpatientdata-1542865627584.csv**

What is Inpatient (claim): An inpatient claim is a request for payment submitted by a healthcare provider to an insurance company for medical services provided during a hospital admission, where the patient stayed overnight or longer.

Column explanation –

|  |  |
| --- | --- |
| **Column Name** | **Description** |
| Provider | Unique ID for the healthcare provider (e.g., hospital or clinic). |
| BeneID | Unique ID for the patient (beneficiary). |
| ClaimID | Unique identifier for the claim. |
| ClaimStartDt | Date when the inpatient care began (admission date). |
| ClaimEndDt | Date when the care ended (discharge date). |
| AdmissionDt | Actual admission date to the hospital. |
| DischargeDt | Actual discharge date from the hospital. |
| ClmAdmitDiagnosisCode | The main diagnosis code at the time of admission (ICD code). |
| DiagnosisCode\_1 to DiagnosisCode\_10 | Other diagnosis codes that describe the patient’s condition(s). |
| ProcedureCode\_1 to ProcedureCode\_6 | CPT/HCPCS codes representing procedures performed during the stay. |
| AttendingPhysician | ID of the physician overseeing care. |
| OperatingPhysician | ID of the physician who performed surgery (if applicable). |
| OtherPhysician | ID of any additional physician involved. |
| DeductibleAmtPaid | Amount paid out-of-pocket by the beneficiary. |
| InscClaimAmtReimbursed | Amount reimbursed by the insurance company. |

1. **Train\_Beneficiarydata-1542865627584.csv**

What is Benificary? - A beneficiary is a person who is enrolled in an insurance program and is eligible to receive benefits (e.g., medical treatment, reimbursement).

This file contains demographic and medical information about patients, referred to as beneficiaries, who are covered by the healthcare insurance program.

Each row corresponds to one unique patient, and it includes their

* Age, gender
* Coverage details (e.g., chronic conditions)
* Enrollment periods
* Death date (if applicable)

Column explanation –

|  |  |
| --- | --- |
| Column Name | Description |
| BeneID | Unique identifier for the beneficiary (patient). Used to join with inpatient/outpatient claim files. |
| DOB | Date of Birth |
| DOD | Date of Death (if the patient has died; NaN if still alive) |
| Gender | Patient gender (1 = Male, 2 = Female) |
| Race | Ethnicity of the patient (1 = White, 2 = Black, etc.) |
| RenalDiseaseIndicator | Indicates if patient has end-stage renal disease (Y/N) |
| ChronicCond\_\* | 13 columns indicating if the patient has specific chronic conditions (e.g., Diabetes, Heart Failure). Values are: - 1 = Present - 2 = Not Present - NaN = Unknown |
| State | State code where the beneficiary lives |
| County | County code |
| NoOfMonths\_PartACov | Number of months the beneficiary was enrolled in Medicare Part A (hospital insurance) |
| NoOfMonths\_PartBCov | Number of months enrolled in Medicare Part B (medical insurance) |