**DOKUMEN PROYEK**

**12S4054 - PENAMBANGAN DATA**

**Segmentasi Wilayah Peserta BPJS Berdasarkan Jumlah dan Pertumbuhan Peserta Menggunakan K-Means Clustering**

**Disusun Oleh:**

| **12S22007** | **Tamara Y Sianipar** |
| --- | --- |
| **12S22018** | **Jesica A Siburian** |
| **12S22021** | **Krisnia Calysta Siahaan** |
| **12S22042** | **Ruth Septiana Simanullang** |

****

**PROGRAM STUDI SARJANA SISTEM INFORMASI**

**FAKULTAS INNFORMATIKA DAN TEKNIK ELEKTRO (FITE)**

**INSTITUT TEKNOLOGI DEL**

**(2024/2025)**

**BAB I**

**Business Understanding**

**1.1 Determine Business Objective**

BPJS Kesehatan merupakan lembaga yang bertanggung jawab dalam menjamin akses layanan kesehatan yang merata dan berkualitas bagi seluruh penduduk Indonesia. Dengan jumlah peserta yang terus meningkat setiap tahunnya, tantangan dalam pengelolaan layanan kesehatan juga menjadi semakin kompleks, terutama dalam hal distribusi sumber daya, perencanaan kapasitas layanan, dan pengendalian pembiayaan.

Permasalahan menjadi semakin signifikan ketika dikaitkan dengan penyakit kronis tidak menular seperti Diabetes Mellitus (DM) dan penyakit menular seperti Tuberkulosis (TB), yang membutuhkan pengelolaan jangka panjang dan alokasi sumber daya yang intensif dan berkelanjutan.

Oleh karena itu, BPJS Kesehatan memerlukan pendekatan analitis untuk memahami distribusi peserta penderita DM dan TB, baik secara geografis maupun berdasarkan tren pertumbuhannya. Hal ini bertujuan untuk meningkatkan efektivitas intervensi dan efisiensi penggunaan sumber daya.

Tujuan bisnis utama dari proyek ini adalah:

* Melakukan segmentasi wilayah berdasarkan jumlah peserta dan tingkat pertumbuhan peserta BPJS yang menderita DM dan TB.
* Memberikan wawasan berbasis data (data-driven insights) kepada BPJS Kesehatan untuk menentukan prioritas wilayah dalam perencanaan dan pengalokasian layanan kesehatan.
* Meningkatkan efisiensi dalam pengambilan keputusan strategis, seperti penambahan fasilitas, peningkatan kapasitas layanan, atau edukasi masyarakat di wilayah tertentu.

Dengan menggunakan metode K-Means Clustering, BPJS Kesehatan dapat mengelompokkan wilayah ke dalam kategori pertumbuhan peserta tinggi, sedang, dan rendah. Hasil segmentasi ini dapat dimanfaatkan untuk merancang strategi pelayanan kesehatan yang lebih terarah, efisien, dan responsif terhadap kondisi lokal.

**1.2 Determine Project Goal**

Agar tujuan bisnis dapat tercapai, proyek ini menetapkan beberapa sasaran teknis dan analitis sebagai berikut:

* Mengumpulkan dan menganalisis data peserta BPJS yang menderita penyakit DM dan TB berdasarkan wilayah administratif (kabupaten/kota atau provinsi).
* Mengukur pertumbuhan jumlah peserta dari waktu ke waktu untuk masing-masing wilayah.
* Menerapkan metode K-Means Clustering untuk melakukan segmentasi wilayah berdasarkan kombinasi jumlah peserta dan tingkat pertumbuhannya.
* Menghasilkan visualisasi data, seperti peta segmentasi dan grafik distribusi peserta, yang mudah dipahami dan berguna bagi pengambil keputusan.
* Memberikan rekomendasi strategi layanan dan alokasi sumber daya berdasarkan hasil segmentasi wilayah.

**1.3 Produce Project Plan**

Untuk memastikan pelaksanaan proyek berjalan dengan terstruktur dan sistematis, berikut adalah tahapan utama yang direncanakan beserta kebutuhan yang diperlukan:

#### A. Pengumpulan Data

Kebutuhan:

* Dataset peserta BPJS yang menderita penyakit Diabetes Mellitus (DM) dan Tuberkulosis (TB).
* Informasi wilayah administratif yang mencakup kode/nama provinsi, kabupaten, dan kota.
* Data historis pertumbuhan peserta (minimal per tahun atau per bulan) untuk analisis tren pertumbuhan.
* Akses ke sistem informasi BPJS atau data publik dari lembaga terkait, seperti Kemenkes, BPS, atau Open Data Pemerintah.

Aktivitas:

* Mengumpulkan data jumlah peserta penderita DM dan TB berdasarkan wilayah administratif.
* Mengintegrasikan data jumlah peserta dengan historis pertumbuhannya untuk melihat tren dari waktu ke waktu.

#### B. Eksplorasi dan Pra-pemrosesan Data

Kebutuhan:

* Alat analisis data seperti Python (pandas, matplotlib, seaborn), Excel, atau R.
* Tim analis data atau data scientist yang memiliki kemampuan dalam pengolahan data.
* Panduan kualitas data, seperti data quality checklist, untuk memastikan data yang digunakan bersih dan akurat.

Aktivitas:

* Pembersihan data, meliputi penghapusan duplikasi, penanganan nilai kosong (missing values), dan pengidentifikasian serta penanganan data anomali (outliers).
* Menghitung metrik pertumbuhan seperti growth rate tahunan dan statistik deskriptif (rata-rata, median, sebaran, dll) untuk mendapatkan gambaran awal tentang data.
* Menerapkan teknik analisis deskriptif untuk memahami distribusi dan tren data peserta BPJS.

#### C. Penerapan Algoritma K-Means

Kebutuhan:

* Library machine learning, seperti Scikit-learn (untuk Python) yang mendukung implementasi K-Means Clustering.
* Infrastruktur komputasi yang memadai, seperti laptop/PC dengan spesifikasi yang cukup untuk memproses data besar.
* Teknik validasi model, seperti Elbow Method atau Silhouette Score, untuk memastikan jumlah cluster yang optimal.

Aktivitas:

* Menentukan jumlah cluster yang optimal (misalnya 3 cluster: rendah, sedang, tinggi).
* Mengelompokkan wilayah berdasarkan jumlah peserta dan tingkat pertumbuhannya menggunakan K-Means Clustering.

#### D. Visualisasi dan Interpretasi

Kebutuhan:

* Alat visualisasi data seperti Tableau, Power BI, atau matplotlib/seaborn untuk menyajikan hasil analisis.
* Peta Indonesia (shapefile atau GeoJSON) yang digunakan untuk memvisualisasikan segmentasi wilayah secara spasial.
* Desain dashboard atau laporan interaktif untuk memudahkan interpretasi dan penyajian hasil.

Aktivitas:

* Menyusun peta segmentasi wilayah berdasarkan hasil cluster untuk memetakan hasil analisis ke dalam bentuk geografis.
* Membuat grafik pendukung (misalnya bar chart, line chart, pie chart) untuk memberikan gambaran lebih jelas mengenai distribusi peserta dan tren pertumbuhannya.

#### E. Penyusunan Insight dan Rekomendasi

Kebutuhan:

* Tim analis data dan pemangku kebijakan untuk berdiskusi mengenai hasil segmentasi.
* Template laporan dan presentasi yang telah disiapkan untuk memudahkan penyampaian hasil.
* Forum atau sesi diskusi dengan pihak BPJS Kesehatan untuk interpretasi hasil dan merumuskan tindak lanjut yang tepat.

Aktivitas:

* Menyusun laporan analitik berdasarkan hasil segmentasi dan analisis data.
* Memberikan rekomendasi berbasis data untuk strategi pelayanan kesehatan yang lebih tepat sasaran, misalnya dalam penambahan fasilitas kesehatan, peningkatan kapasitas layanan, atau program edukasi untuk masyarakat di wilayah tertentu.

**BAB II**

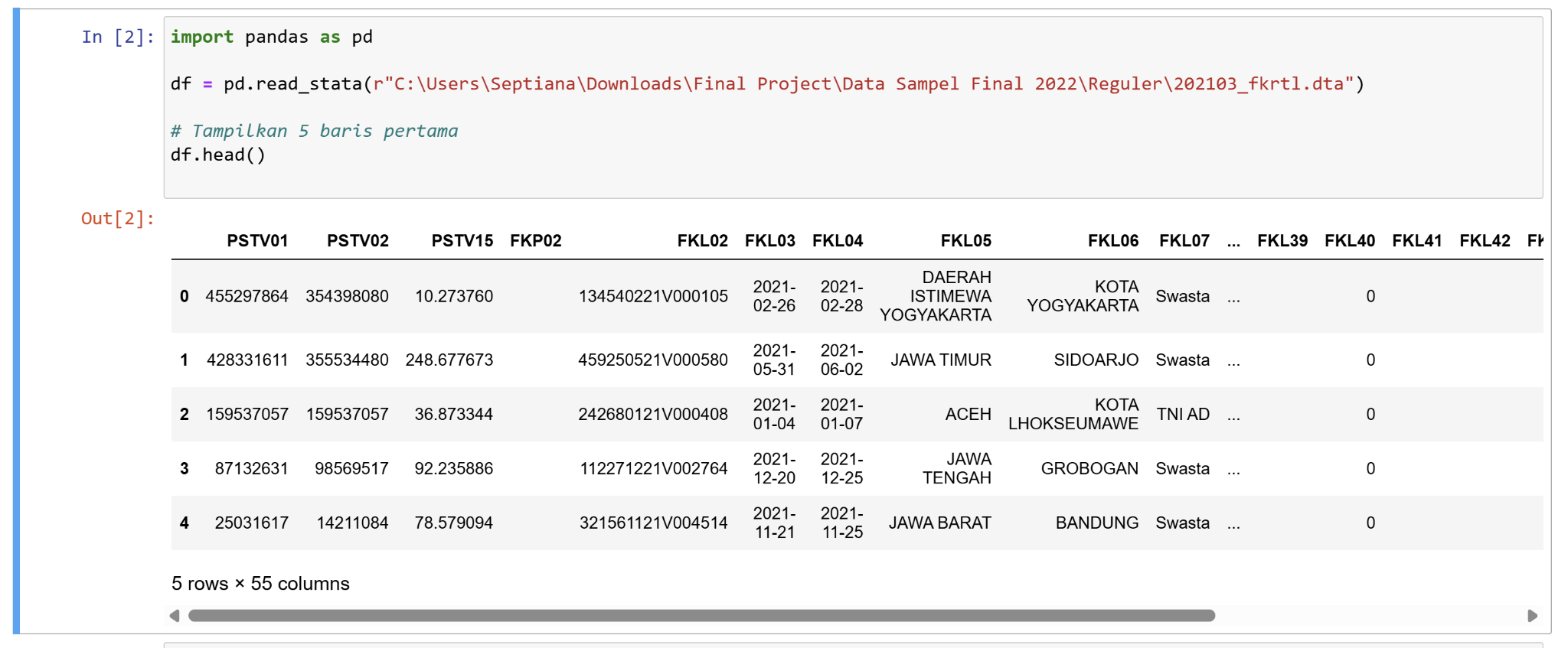
**DATA UNDERSTANDING**

**2.1 Collecting Data**

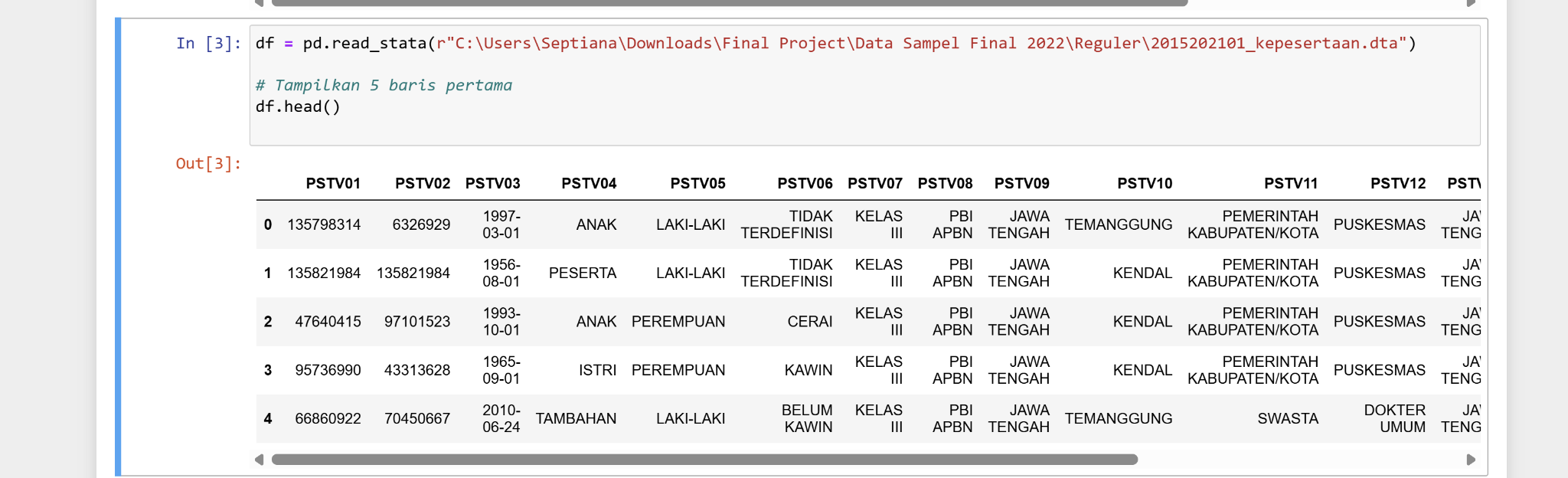
Data dibagi menjadi 3 kelompok besar yaitu:

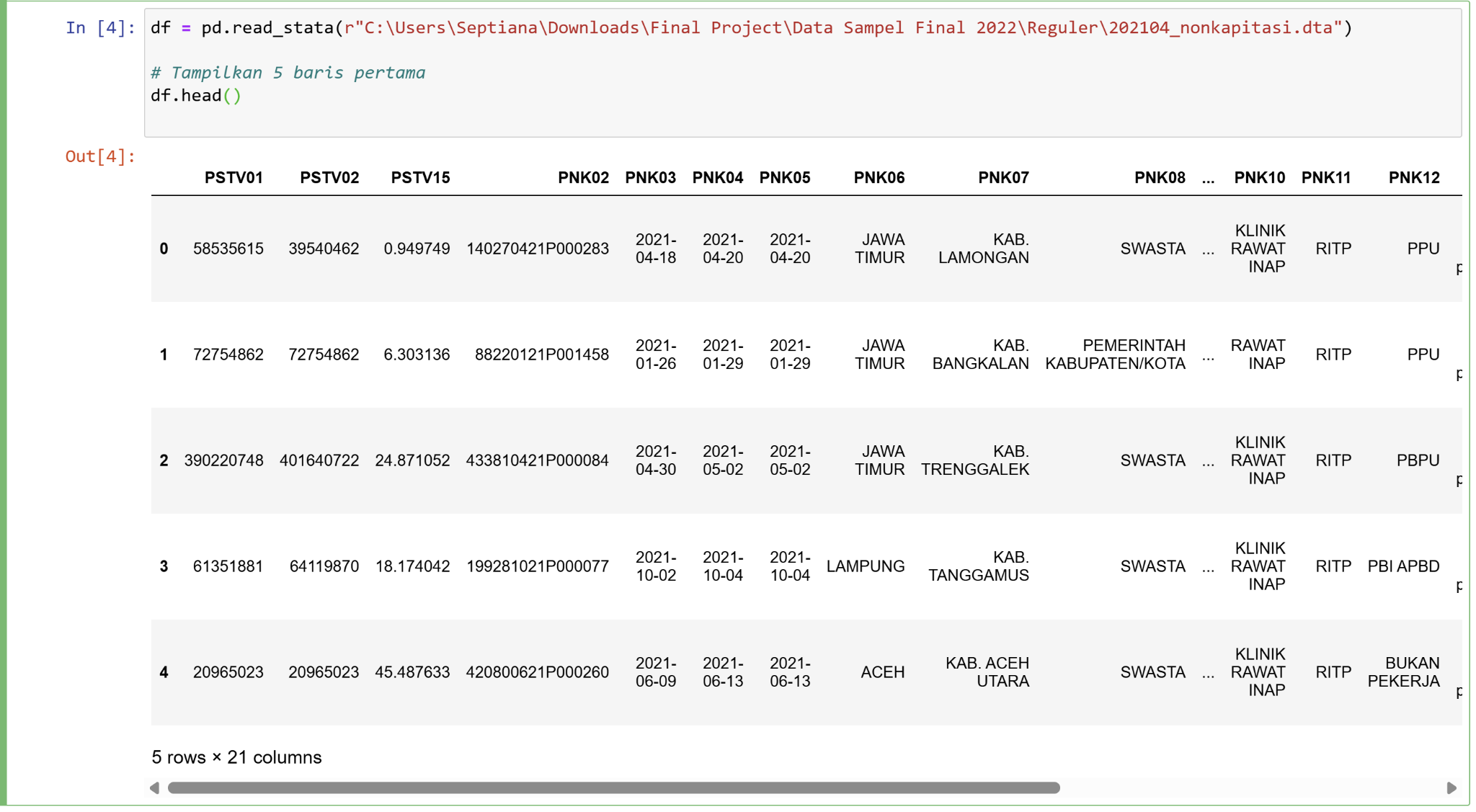
**2.1.1 Data Reguler BPJS (2015-2021)**

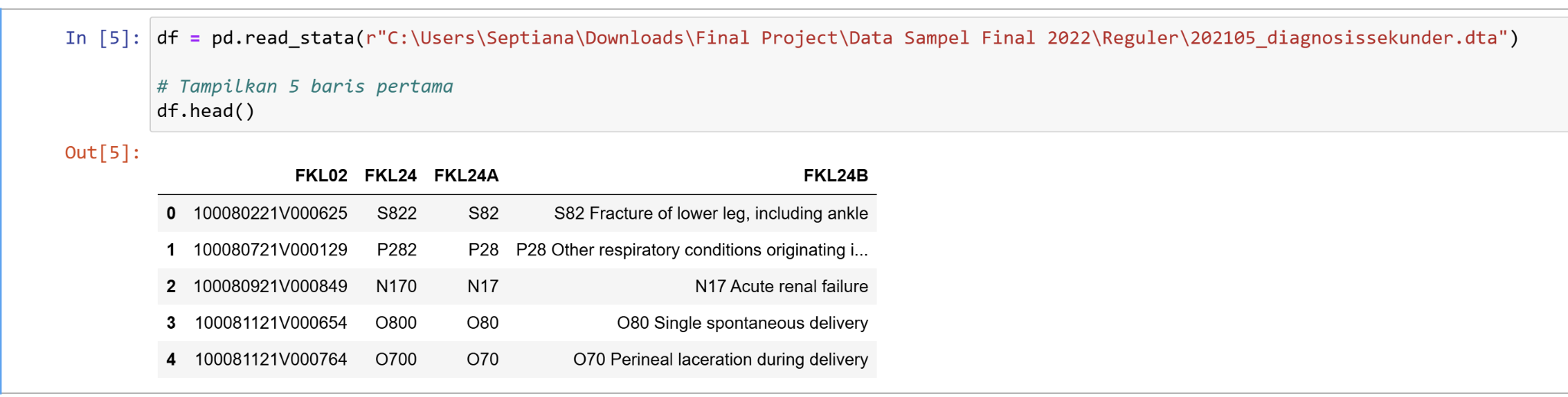
* 202103\_fkrtl.dta : Data pelayanan ke Fasilitas Kesehatan Rujukan Tingkat Lanjutan (FKRTL)



* 2015202101\_kepesertaan.dta : Informasi kepesertaan BPJS, termasuk usia, jenis kelamin, status aktif

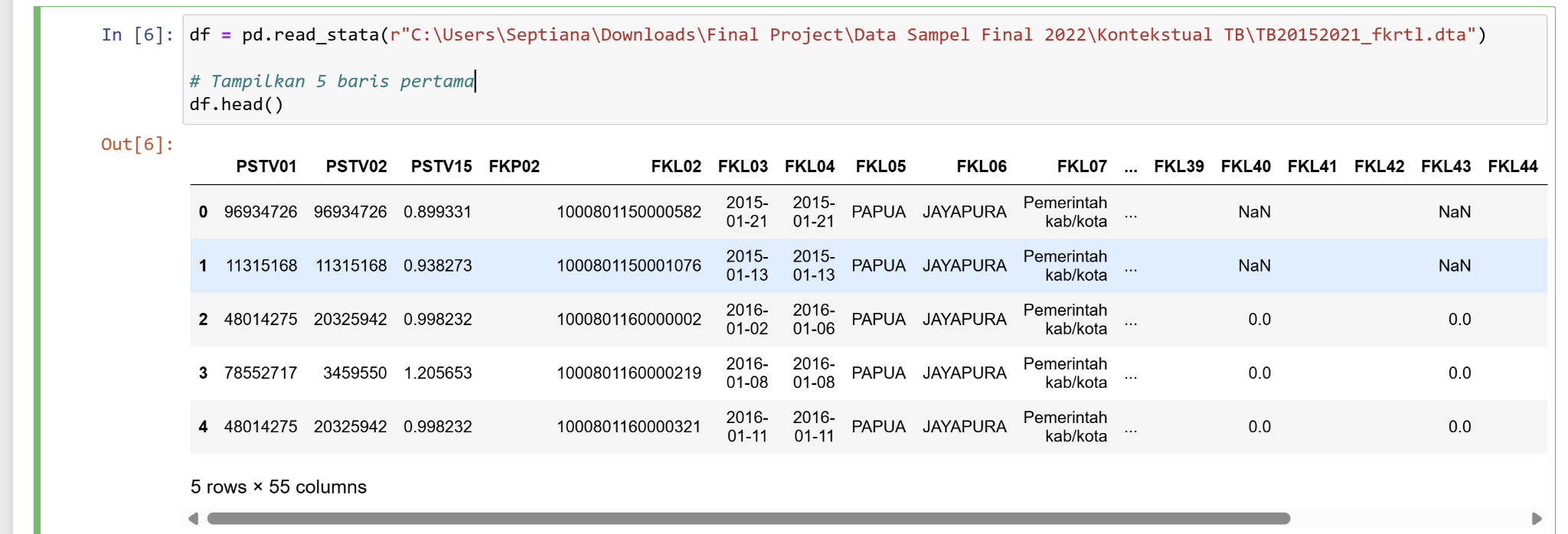


* 202104\_nonkapitasi.dta : Klaim layanan yang tidak termasuk dalam sistem kapitasi (rawat jalan/rujuk lanjut)
* 202105\_diagnosissekunder.dta : Diagnosa tambahan pasien dalam pelayanan kesehatan

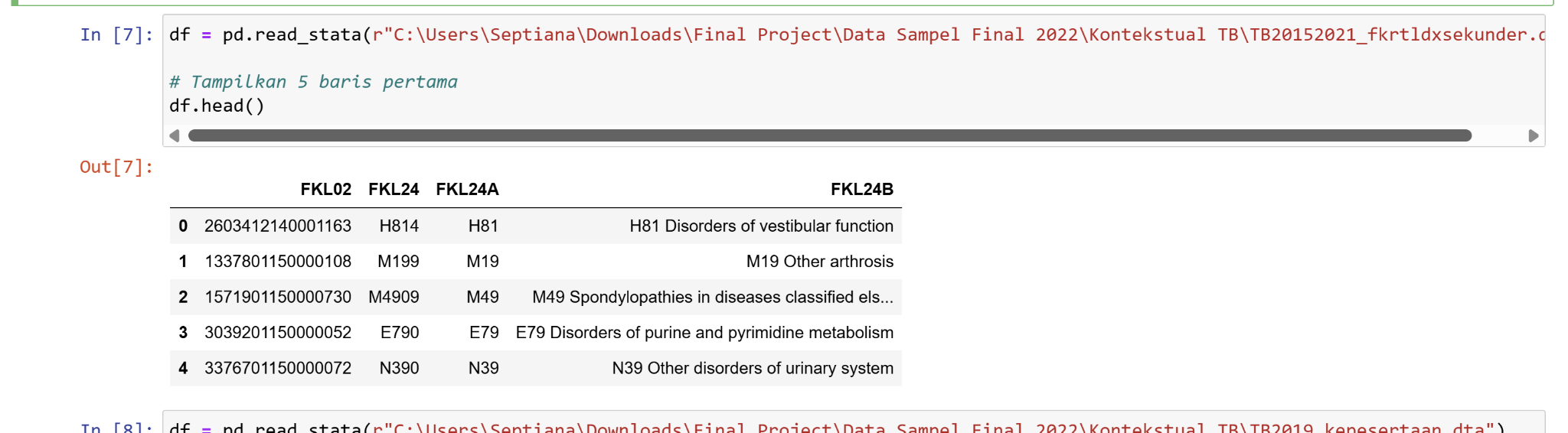


**2.1.2 Data Kontekstual TB**

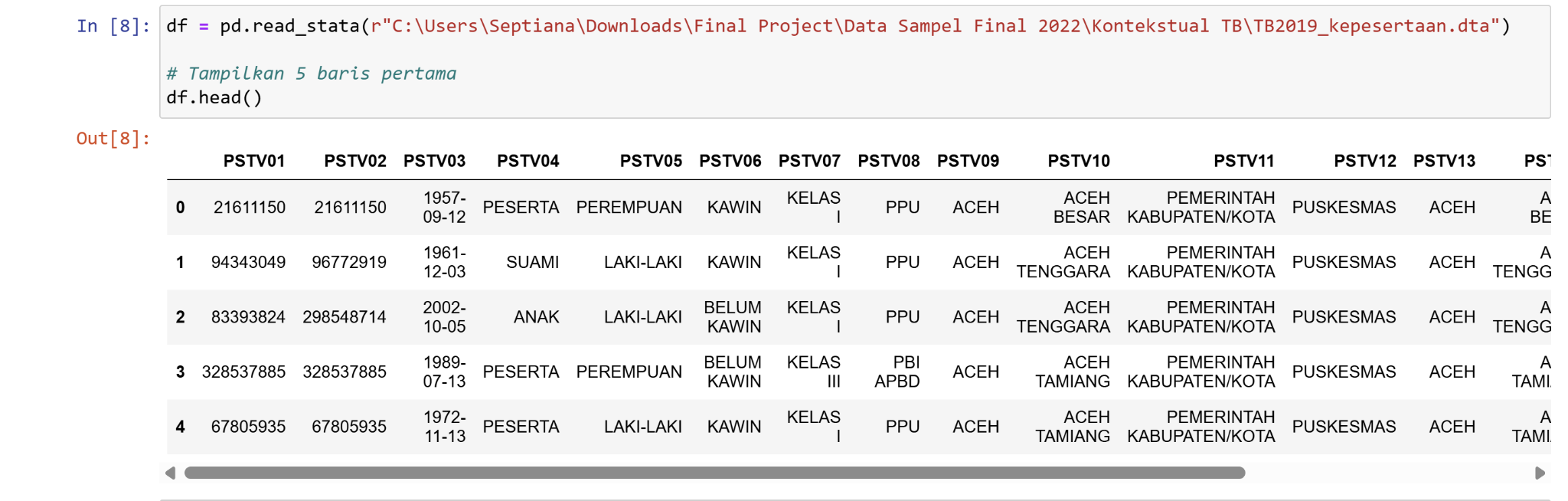
* TB20152021\_fkrtl.dta : Data kunjungan FKRTL khusus pasien TB



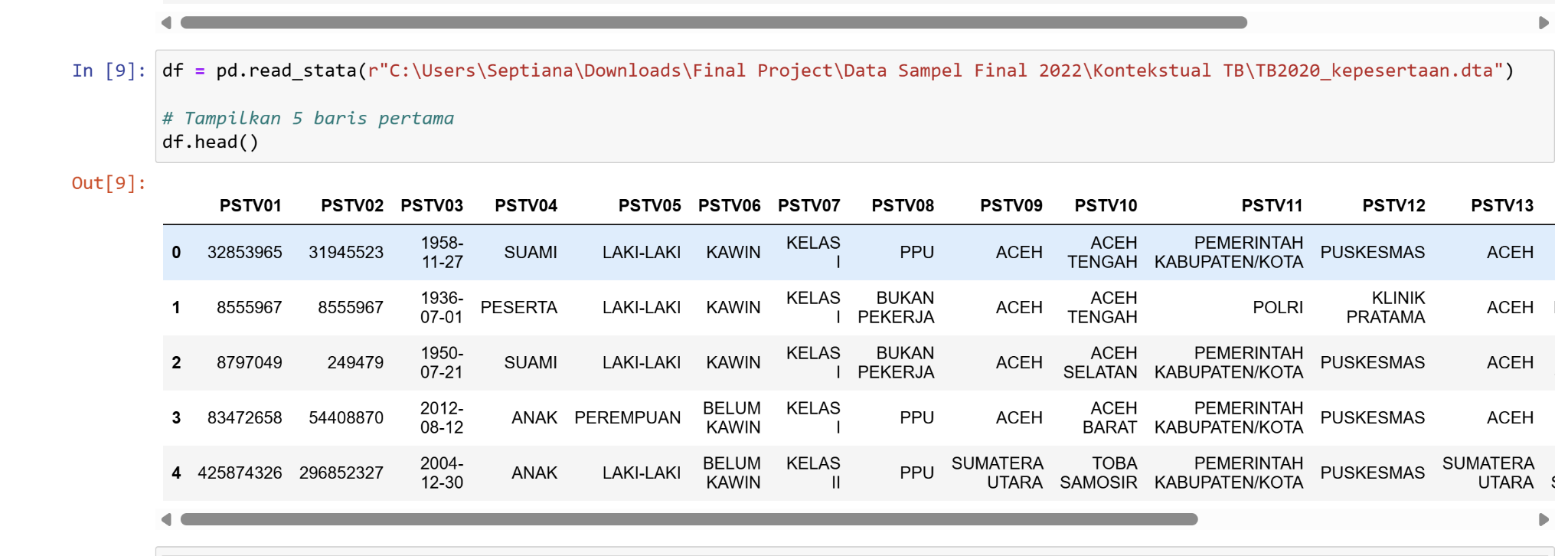
* TB20152021\_fkrtldxsekunder.dta : Diagnosa sekunder pasien TB



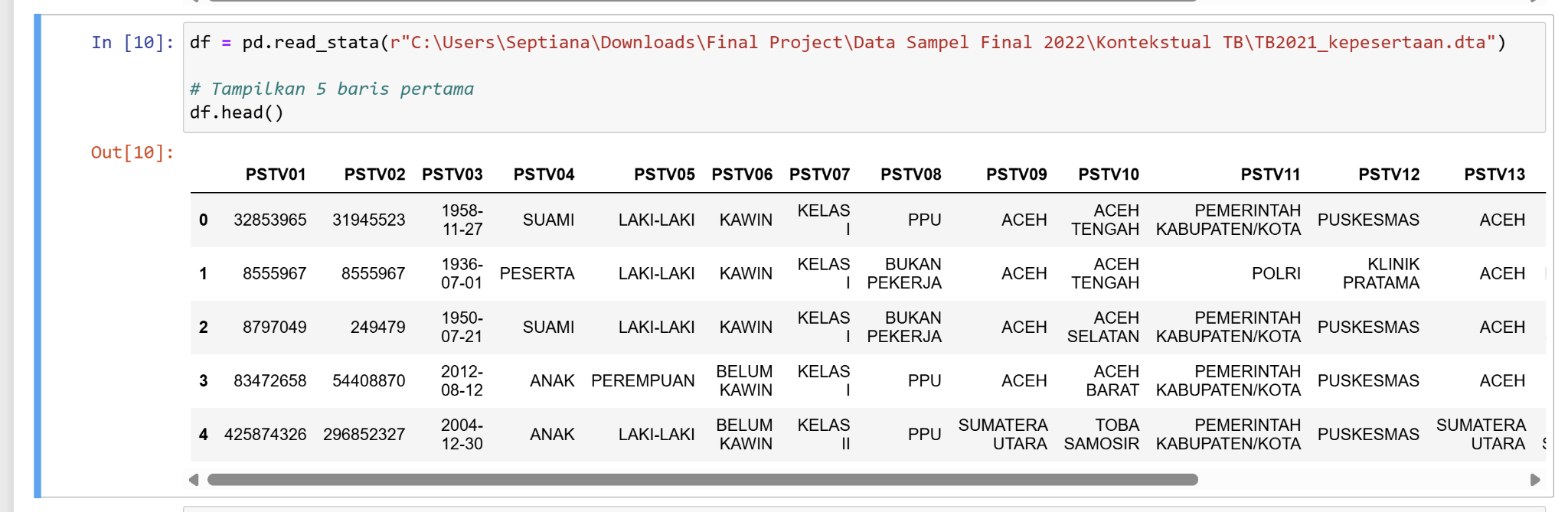
* TB2019\_kepesertaan.dta : Kepesertaan pasien TB tahun 2019



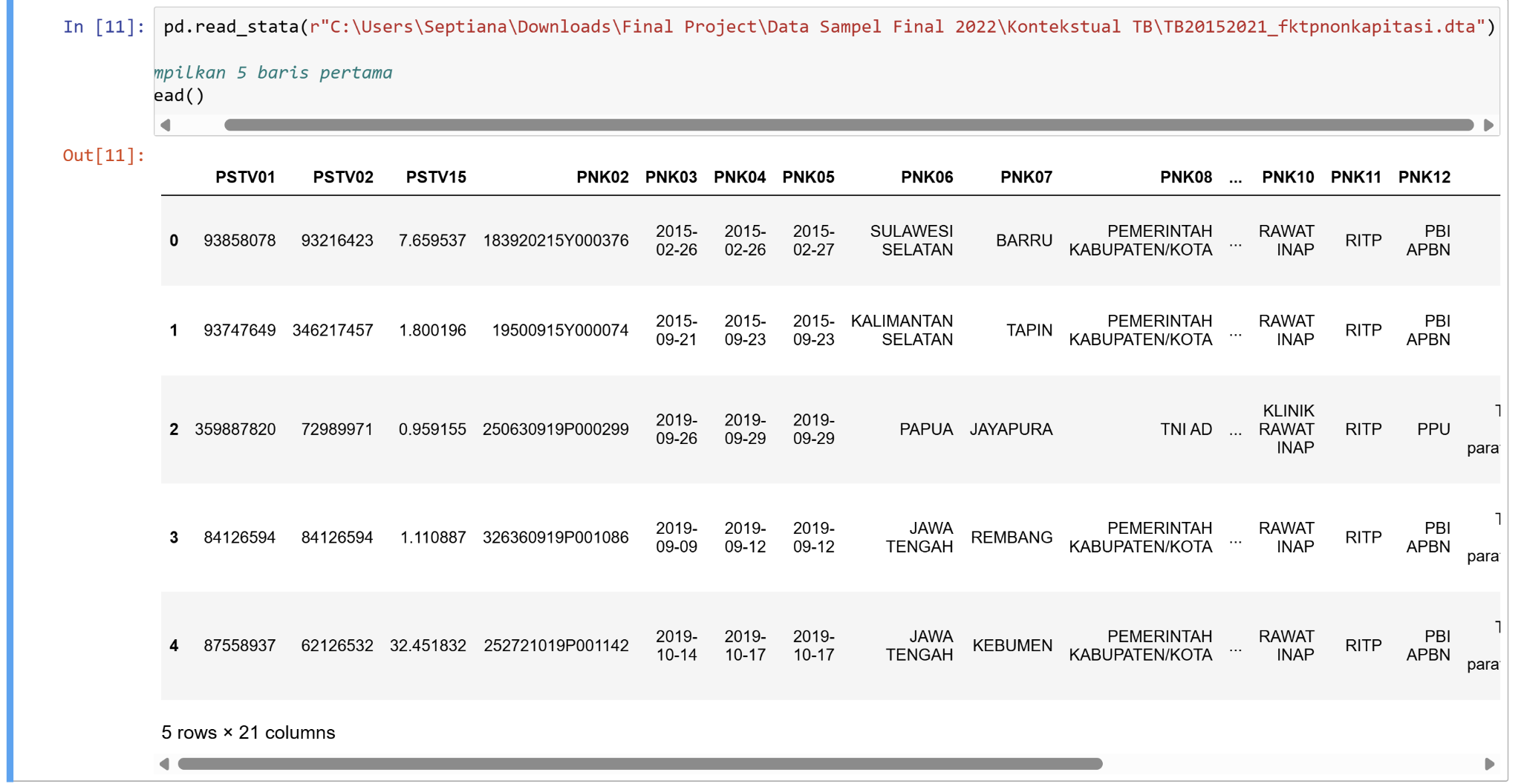
* TB2020\_kepesertaan.dta : Kepesertaan pasien TB tahun 2020



* TB2021\_kepesertaan.dta : Kepesertaan pasien TB tahun 2021

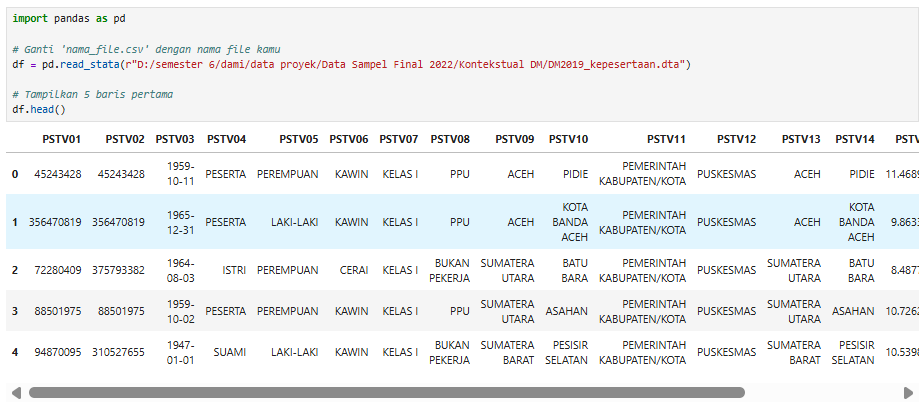


* TB20152021\_fktpnonkapitasi.dta : Layanan non-kapitasi dari Fasilitas Kesehatan Tingkat Pertama (FKTP) untuk pasien TB

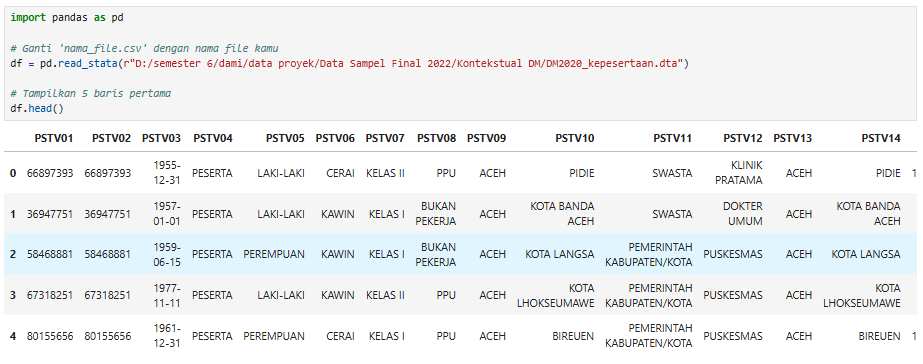


2.1.3 Data Kontekstual DM

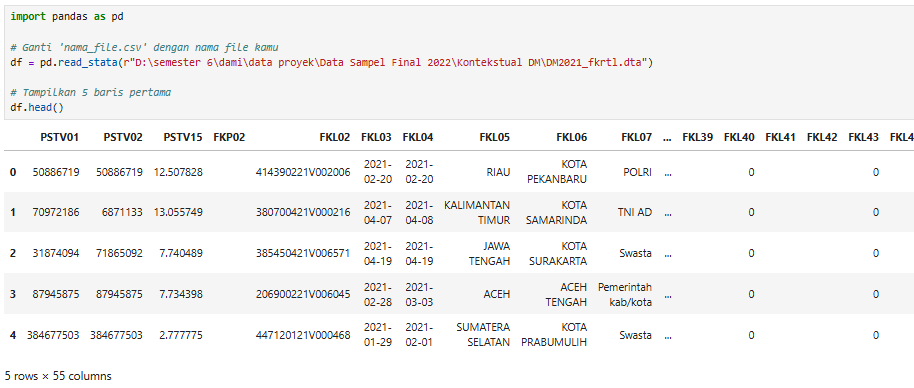
* DM2019\_kepesertaan.dta : Data kepesertaan peserta JKN tahun 2019 yang terindikasi DM.



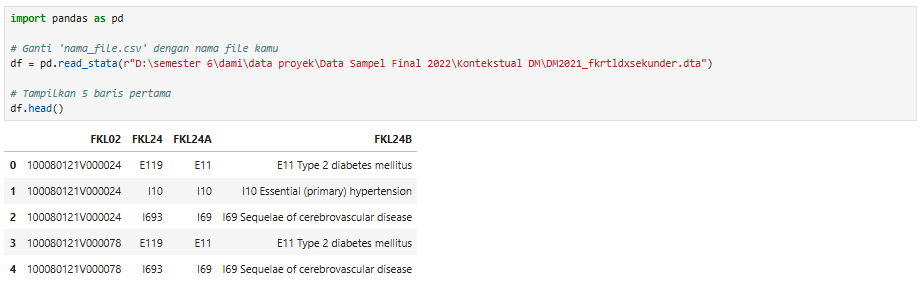
* DM2020\_kepesertaan.dta : Data kepesertaan peserta JKN tahun 2020 yang terindikasi DM.



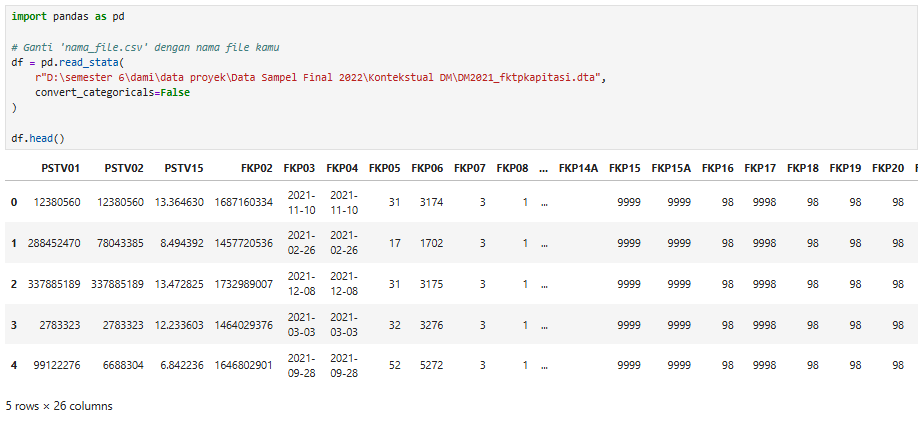
* DM2021\_fkrtl.dta : Data pelayanan peserta DM di Fasilitas Kesehatan Rujukan Tingkat Lanjutan (FKRTL) pada tahun 2021.



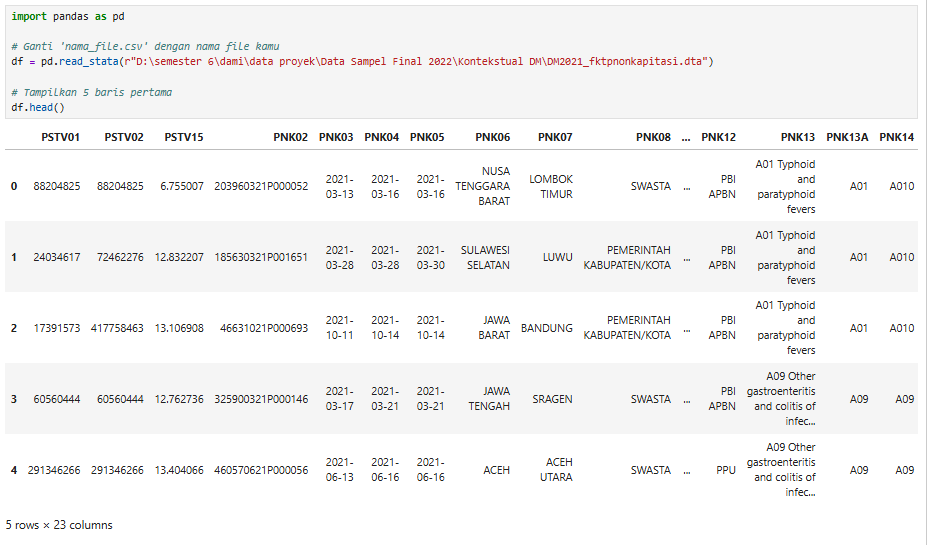
* DM2021\_fkrtldxsekunder.dta : Data diagnosis sekunder peserta DM yang mendapat pelayanan di FKRTL tahun 2021.



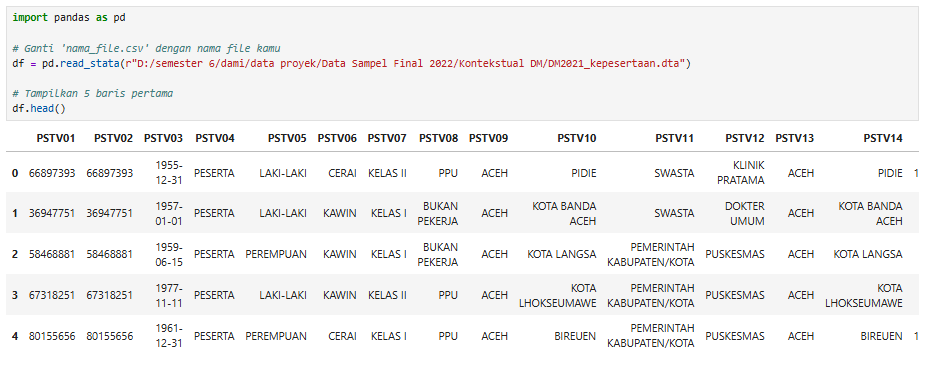
* DM2021\_fktpkapitasi.dta : Data pelayanan peserta DM di Fasilitas Kesehatan Tingkat Pertama (FKTP) dengan sistem kapitasi pada tahun 2021.



* DM2021\_fktpnonkapitasi.dta : Data pelayanan peserta DM di FKTP dengan skema non-kapitasi tahun 2021. Biasanya mencakup pelayanan yang tidak dibayar rutin per kapita.



* DM2021\_kepesertaan.dta : Data peserta JKN tahun 2021 yang memiliki indikasi DM.



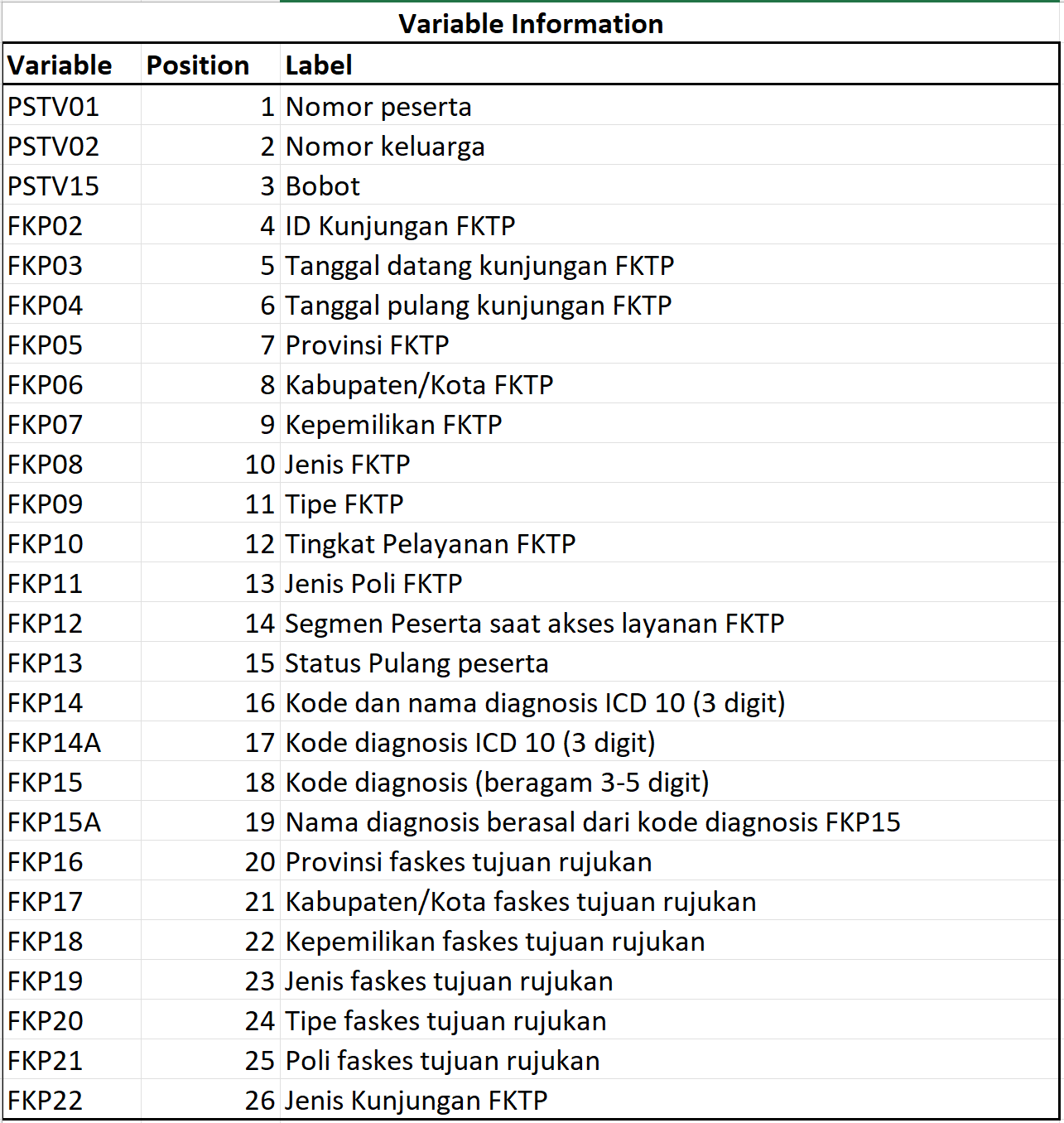
**2.2 Describe Data**

Ada 4 bagian utama file yang akan dijelaskan:

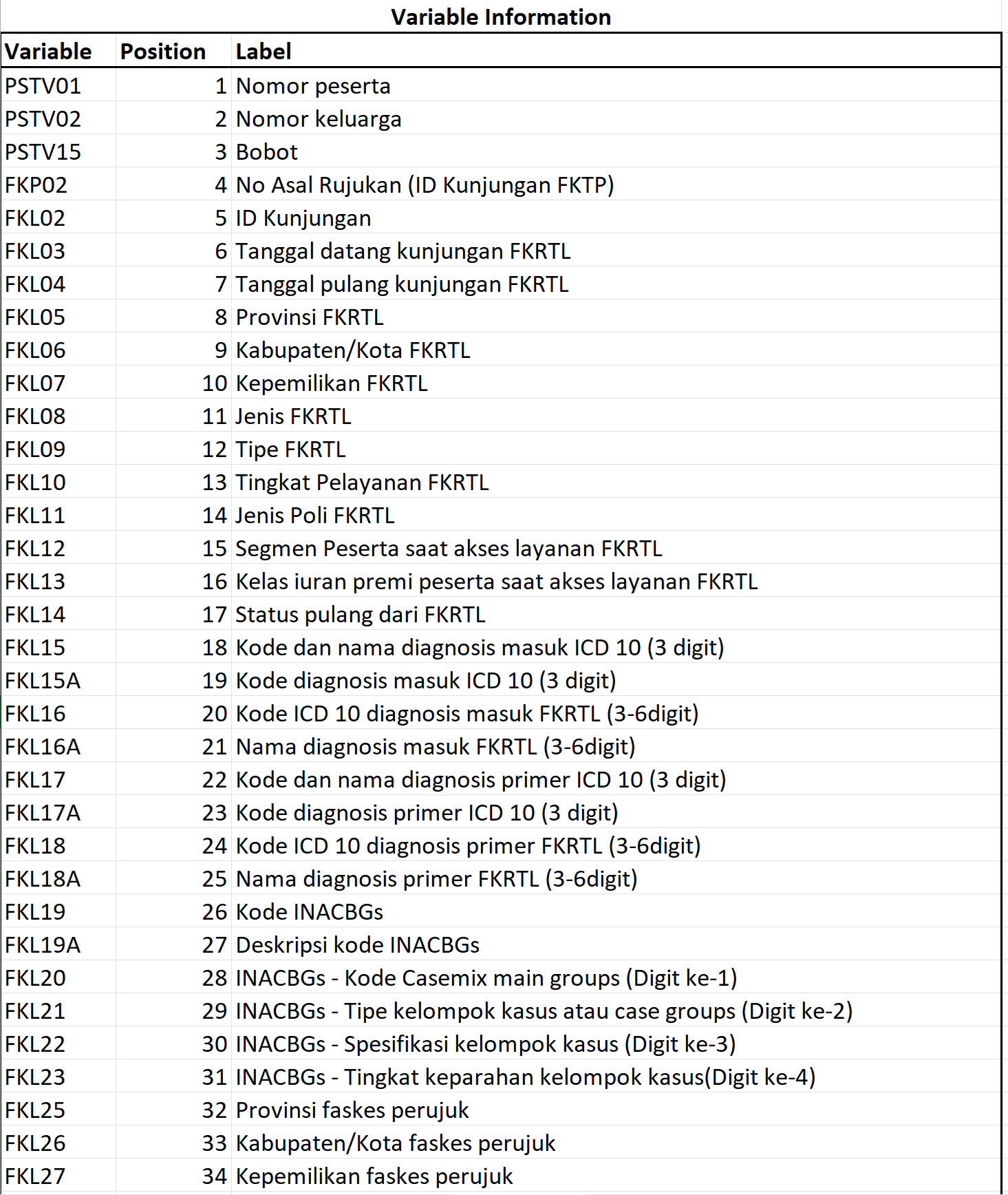
1. Kepesertaan



1. Fasilitas Kesehatan Tingkat Pertama (FKTP)

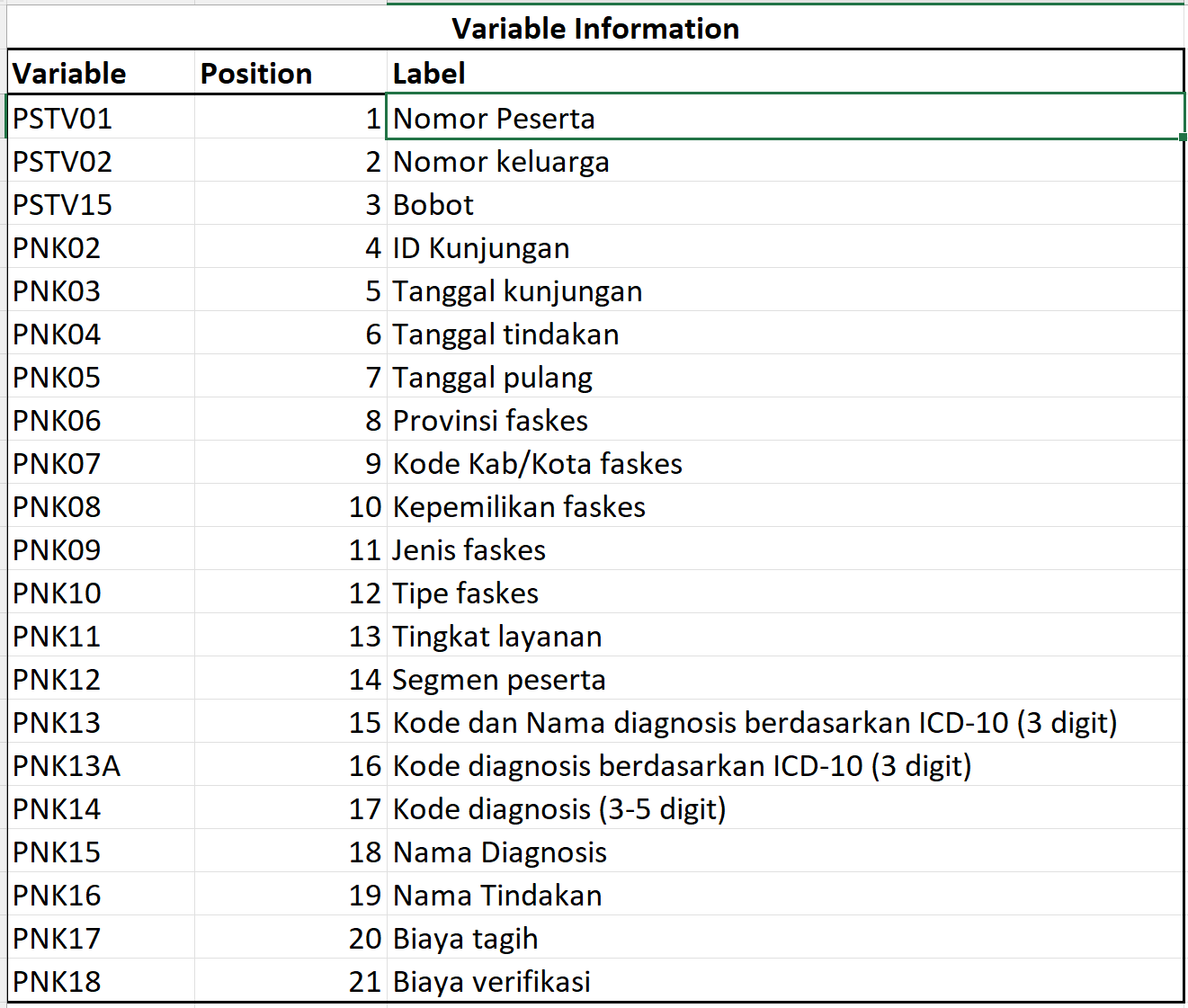
****

1. Fasilitas Kesehatan Rujukan Tingkat Lanjutan (FKRTL)

****

****

1. Non-Kapitasi

****

**2.3 Validation Data**