Pharmaceutical Sales Dataset

Overview

This dataset contains transactional sales data collected over a six-year period (2014-2019) from a Point-of-Sale system in an individual pharmacy. The dataset originally consisted of 600,000 transaction records, each indicating the date and time of sale, pharmaceutical drug brand name, and sold quantity.

A selected group of 57 drugs from the dataset has been classified into the following Anatomical Therapeutic Chemical (ATC) Classification System categories:

- **M01AB** - Anti-inflammatory and antirheumatic products, non-steroids, Acetic acid derivatives, and related substances
- **M01AE** - Anti-inflammatory and antirheumatic products, non-steroids, Propionic acid derivatives
- **N02BA** - Other analgesics and antipyretics, Salicylic acid and derivatives
- **N02BE/B** - Other analgesics and antipyretics, Pyrazolones and Anilides
- **N05B** - Psycholeptic drugs, Anxiolytic drugs
- **N05C** - Psycholeptic drugs, Hypnotics and sedatives drugs
- **R03** - Drugs for obstructive airway diseases
- **R06** - Antihistamines for systemic use

Data Processing

The dataset has been pre-processed to ensure high-quality data for analysis. The pre-processing steps include:

- **Outlier Detection and Treatment:** Identifying and addressing anomalies in the data
- **Missing Data Imputation:** Filling in missing values using statistical methods
- **Resampling:** Sales data has been resampled into different time periods:
    - Hourly
    - Daily
    - Weekly
    - Monthly

Tools Used

This dataset has been analyzed and processed using the following technologies:

- **Python** (Jupyter Notebook) for data processing, analysis, and visualization
- **SQL** for querying and managing structured data
- **Entity-Relationship Diagram (ERD)** to represent the structure of the dataset and its relationships

Usage

This dataset is useful for:

- **Time-series analysis** of pharmaceutical sales trends
- **Market research** on drug demand over time
- **Predictive modeling** for inventory and supply chain optimization
- **Public health studies** on drug consumption patterns

License

Please refer to the repository for licensing details.

Citation

If you use this dataset in your research or projects, please cite it accordingly.

Contact

For questions or further information, feel free to open an issue in this repository