

Final Project Report

December 08, 2025

David Lewis

lewis3d7@mail.uc.edu

Contents

1. Introduction	1
2. Methods	1
2.1. Data source	1
3. Results	2
4. References	3

1. Introduction

According to Drugbank, there are around 20,000 unique drugs used in the healthcare system.¹ To contrast, the average 20 year old American knows around 40,000 words, though not necessarily what each word means.² Therefore, if a physician were to learn all of the drug names, they would take up around 1/3 of the physician's entire vocabulary. Having full comprehension of all 20,000 drugs would likely be extremely difficult, if not impossible. Free and open access tools typically do not provide interactive visualizations of drug relationships.³⁻⁵ Instead, they may provide comparison charts or articles.⁶

This project explores the development of an interactive tool meant to aid physicians in medicine selection and comparison. The tool provides interactive visualizations meant to make the comparison of medications easier, showing how different medications relate among their attributes including: user reviews, composition, side effects, and usage.

2. Methods

2.1. Data source

The data were collected from Kaggle using the search term "medicine". The collected dataset is entitled "11000 Medicine details".⁷

After collection of the data, which were contained within a single csv file, the data were processed into an SQLite database using python and the pandas package.⁸ The database uses the database architecture defined in Figure 1. The database is then queried by the visualization tool, an R shiny app.⁹

The tool is broken up into two web pages, a "home" page, and a "data viewer" page. The home page provides some simple statistics and vizualizations

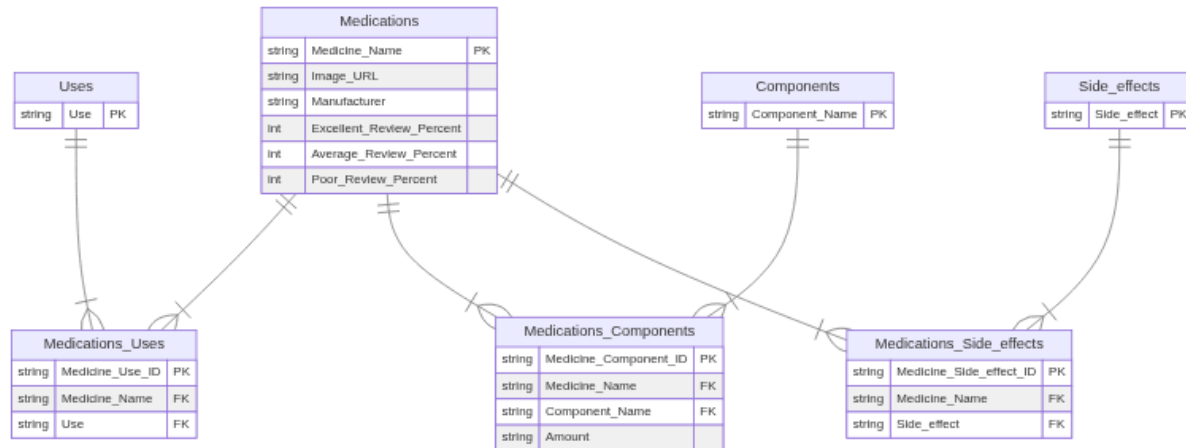


Figure 1: Entity relationship diagram showing architecture of generated SQLite database

3. Results

A live demo of the shiny app can be found on shinyapps.io at the following URL: <https://illustratedman-code.shinyapps.io/project/>. The source code for the project can be found on Github at the following URL: <https://github.com/IllustratedMan-code/ph8093-database-project>.

4. References

1. Statistics | DrugBank Online. Accessed December 2, 2025. <https://go.drugbank.com/stats>
2. Brysbaert M, Stevens M, Mandera P, Keuleers E. How Many Words Do We Know? Practical Estimates of Vocabulary Size Dependent on Word Definition, the Degree of Language Input and the Participant's Age. *Front Psychol.* 2016;7:1116. doi:[10.3389/fpsyg.2016.01116](https://doi.org/10.3389/fpsyg.2016.01116)¹
3. Knox C, Wilson M, Klinger C, et al. DrugBank 6.0: the DrugBank Knowledgebase for 2024. *Nucleic Acids Research.* 2024;52(D1):D1265-D1275. doi:[10.1093/nar/gkad976](https://doi.org/10.1093/nar/gkad976)²
4. WebMD - Better information. Better health. WebMD. Accessed December 2, 2025. <https://www.webmd.com/>
5. RxList - The Internet Drug Index for prescription drug information, interactions, and side effects. RxList. Accessed December 2, 2025. <https://www.rxlist.com/>
6. Drugs Comparison Index: Find Drug vs. Drug, Side Effects, Uses, Interactions, More. RxList. Accessed December 2, 2025. https://www.rxlist.com/drugs_comparison/article.htm
7. 11000 Medicine details. Accessed December 2, 2025. <https://www.kaggle.com/datasets/singhnavjot2062001/11000-medicine-details>
8. team T pandas development. pandas-dev/pandas: Pandas. Published online September 30, 2025. doi:[10.5281/zenodo.17229934](https://doi.org/10.5281/zenodo.17229934)³
9. Shiny. Shiny. Accessed December 2, 2025. <https://shiny.posit.co/>

¹<https://doi.org/10.3389/fpsyg.2016.01116>

²<https://doi.org/10.1093/nar/gkad976>

³<https://doi.org/10.5281/zenodo.17229934>