

Drug Discovery data Processing project:

The goal of this project is to retrieve and clean data from various databases like DrugBank, SIDER, Cancer AI, and generate data visualizations using Python and Plotly. The project involves matching drug names with their respective CIDs, removing repeated drug names, and classifying drugs based on their common genes using a heatmap. Additionally, a backend is developed using JavaScript and Ajax to send and retrieve API data from an SQL database that is plotted in Plotly.

Data Retrieval and Cleaning: The project involves retrieving data from various databases such as DrugBank, SIDER, and Cancer AI. The retrieved data is cleaned using Python, where the drug names are pulled out from the drug description, and the respective CIDs are matched. Repeated drug names are also removed.

Data Visualization: The project generates heatmaps that classify drugs based on their common genes. Additionally, Plotly graphs are generated that provide information about respective drugs when a point on the graph is clicked.

Backend Development: A backend is developed using JavaScript and Ajax that sends and retrieves API data from an SQL database. The retrieved data is then plotted in Plotly to generate the required visualizations.

Technologies and Packages: The following technologies and packages are used in this project:

Technology/Package	Description
Python	Programming language used for data cleaning
Pandas	Data manipulation library used in Python
Matplotlib	Data visualization library used in Python
Plotly	Interactive data visualization library used in Python
JavaScript	Programming language used for backend development
Ajax	Asynchronous JavaScript and XML used for backend development
SQL	Relational database management system used to store data

Conclusion: This project involved retrieving and cleaning data from various databases, generating data visualizations using Python and Plotly, and developing a backend using JavaScript and Ajax to send and retrieve API data from an SQL database. The use of these technologies and packages allowed for efficient and effective data analysis and visualization.