Pharmacological Data Analysis

Investigating The Pharmacodynamic And Pharmacokinetic Profiles Of A Drug Molecule

By Emma Ezeumeh

05/29/2024

The Project

Description

Pharmacological Data Analysis aims to understand how drugs work, their efficacy, safety, metabolism, and potential side effects using data generated during the first 2 phases of drug development (drug discovery & preclinical research).

Objective

The primary objective is understanding:

- Relationships between the Compound X's properties,
- Efficacy,
- Safety, and
- Biological targets.

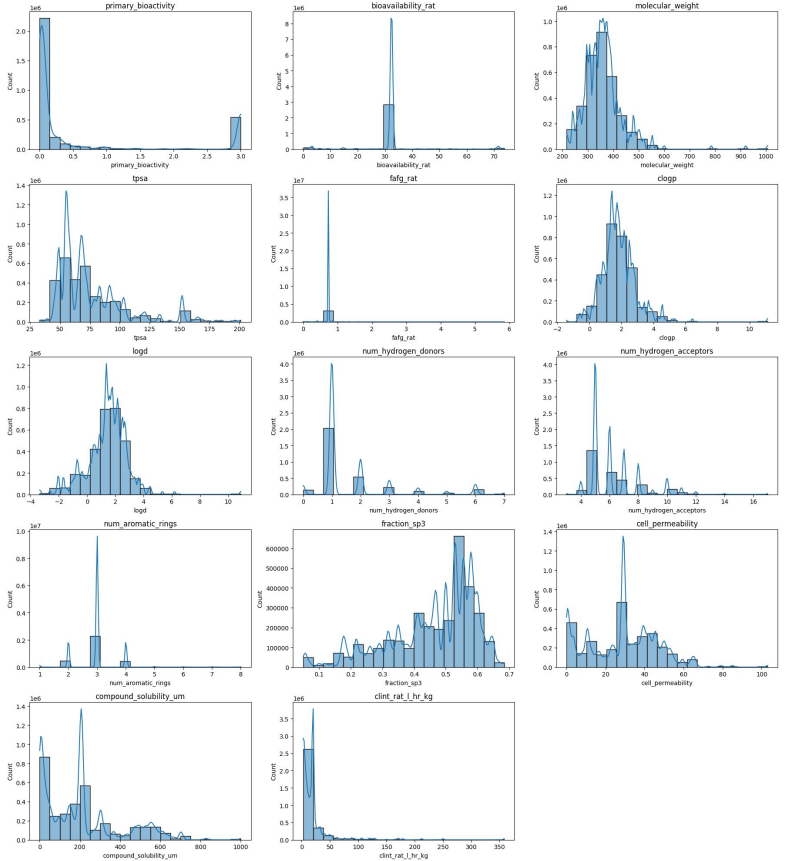
Integrating and exploring multiple datasets in this project aims to uncover patterns and insights that can inform drug development and safety evaluations.

The Dataset

Overview

- 3 datasets (Gene_Drug_Adverse_Event_Relationships.csv, Compound_Off_Target_Activity.csv and Project_Level_Data.csv)
- Almost 7 million rows across three datasets (Big Data)
- Selected features:

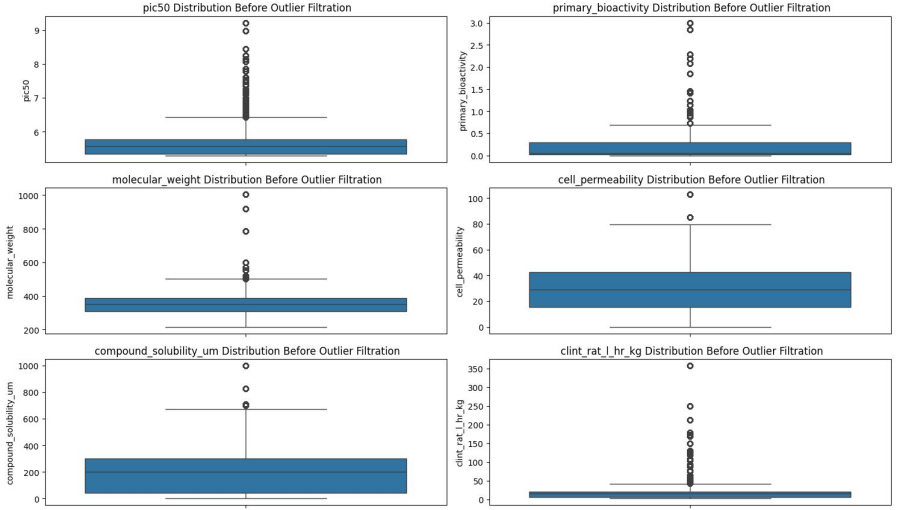
  gene symbol, adverse event, adverse event category, encoded adverse event category, compound ID, pic50, primary target assay bioactivity, molecular weight, cell permeability, compound solubility, fractional absorption, bioavailability, and clearance.



Histogram Showing Feature Distribution

The Process

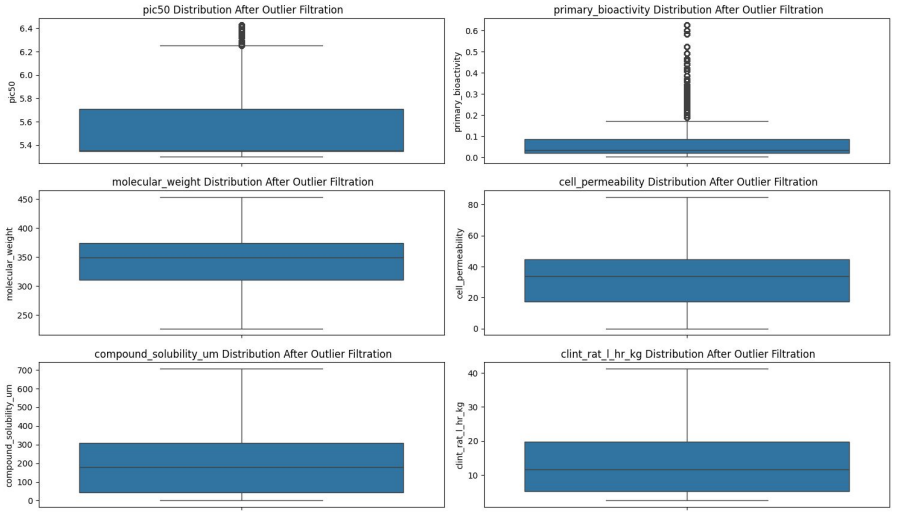
- Data Integration
- Data Cleaning and Preprocessing
- Feature Selection & Engineering
- Data Standardization
- Exploratory Data Analysis
- Statistical Analysis



Boxplot Showing Feature Distribution Before Outlier Filtration

Data Cleaning and Preprocessing

- Dropping columns with unique and >50% missing values
- Handling of missing values (Mean & Mode imputation)
- Removing duplicate rows
- Removing outliers (IQR method)
- Standardizing feature names
- Creating adverse event categories, encoding categorical features, combining columns



Boxplot Showing Feature Distribution After Outlier Filtration

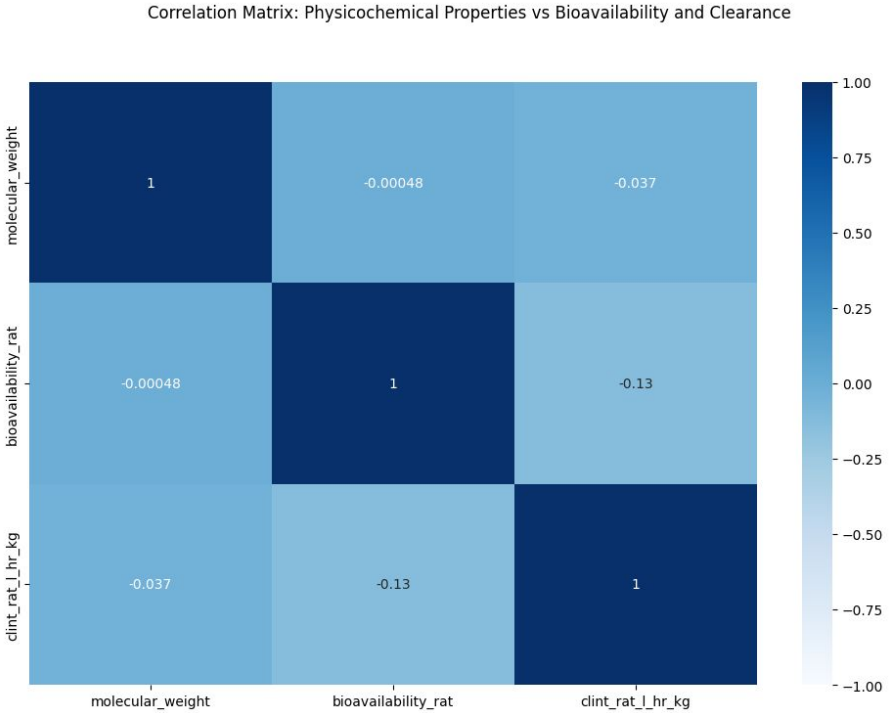
Exploratory Data Analysis

- Almost **2 million rows**
- **Compound ID**: 341
- **Adverse Event Categories**: Hematological, Dermatological, and Cardiovascular.
- **Bioavailability**: max value is 73.7%
- **Fractional Absorption:** max value 5.9
- **Intrinsic Clearance**: max value 41.2 (L/Hr/kg)
- **Compound Solubility**: max value is 705.0 (uM)
- **Cell Permeability**: max value is 85.0
- **pIC50**: max value is 6.4
- **Primary Bioactivity**: max value is 0.6
- **Molecular Weight**: max value is 453.5

Exploring The Impact Of Physicochemical Properties On Bioavailability And Clearance

Key insights

- ↑ Molecular Weight = ↓ Bioavailability and Clearance = ↓ Efficacy
- ↑ Molecular Weight = ↓ Oral Bioavailability = ↓ Absorption
- ↑ Molecular Weight = ↓ Clearance = ↓ Safety

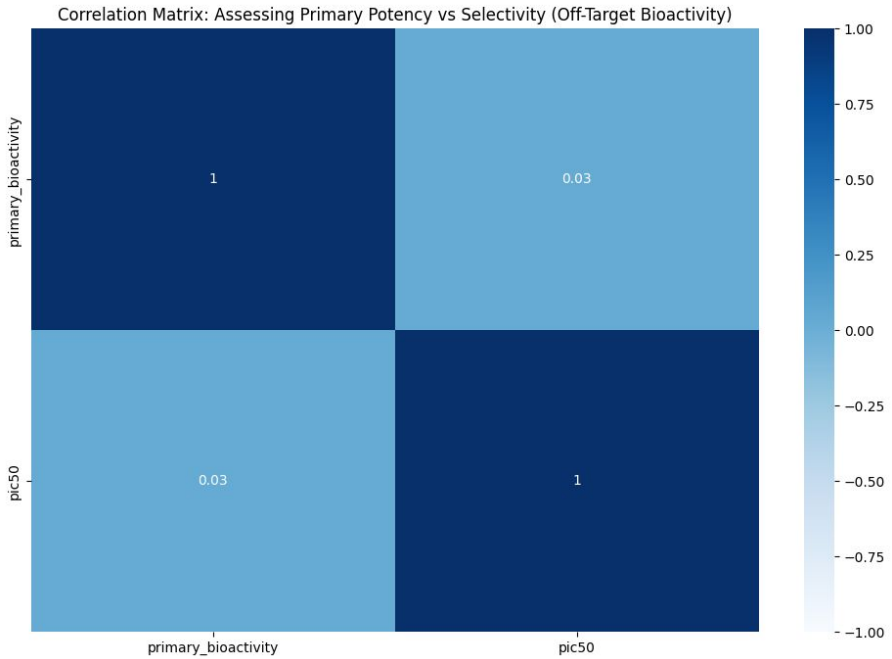


Correlation Matrix: Physicochemical Properties vs Bioavailability And Clearance

Investigating The Relationship Between Primary Potency And Off-Target Activity

Key insights

- ↑ Primary Target Assay Bioactivity = ↑ pIC50 = ↑ Efficacy
- ↑ Primary Target Assay Bioactivity = ↑ pIC50 = ↑ Safety
- ↑ Primary Target Assay Bioactivity = ↑ pIC50 = ↓ Metabolic Burden

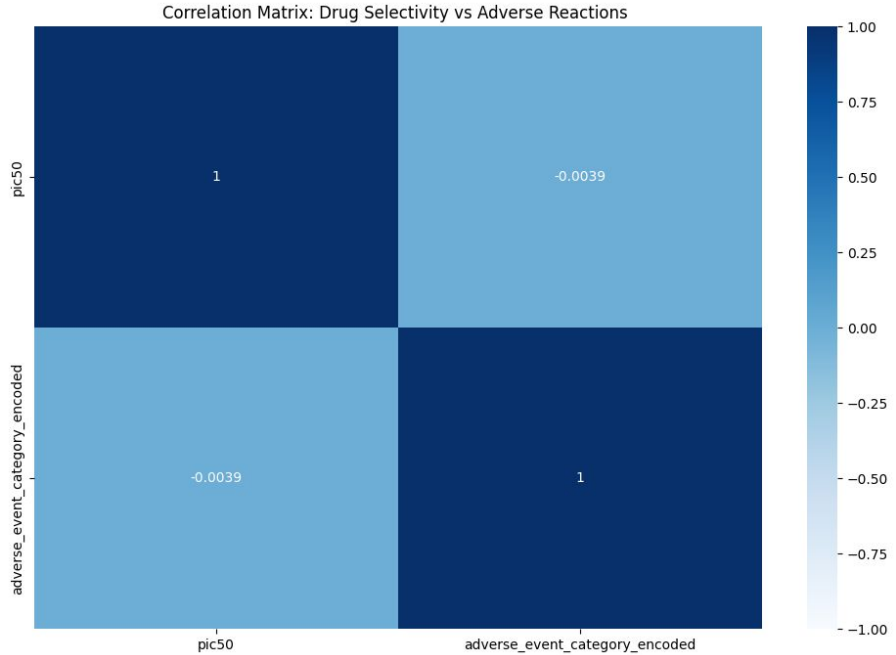


Correlation Matrix: Primary Potency vs Off-Target Activity

Is Higher Selectivity Directly Proportional To The Occurrence Of Fewer Adverse Events?

Key insights

- ↑ pIC50 = ↓ Adverse Events = ↑ Efficacy
- ↑ pIC50 = ↓ Adverse Events = ↓ Dose
- ↑ pIC50 = ↓ Adverse Events = ↑ Safety

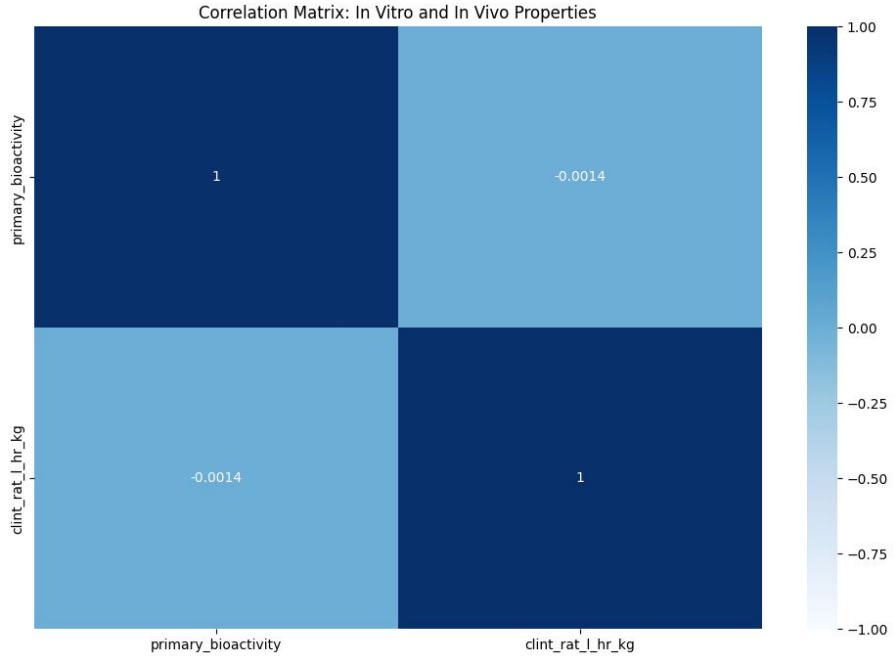


Correlation Matrix: Off-Target Activity vs Adverse Events

Comparing In Vitro Bioactivity With In Vivo Intrinsic Clearance

Key insights

- ↑ Primary Target Assay Bioactivity
  = ↓ Clearance = ↑ Efficacy
- ↑ Primary Target Assay Bioactivity
  = ↓ Clearance = ↑ Dosing Duration
- ↑ Primary Target Assay Bioactivity
  = ↓ Clearance = ↓ Safety

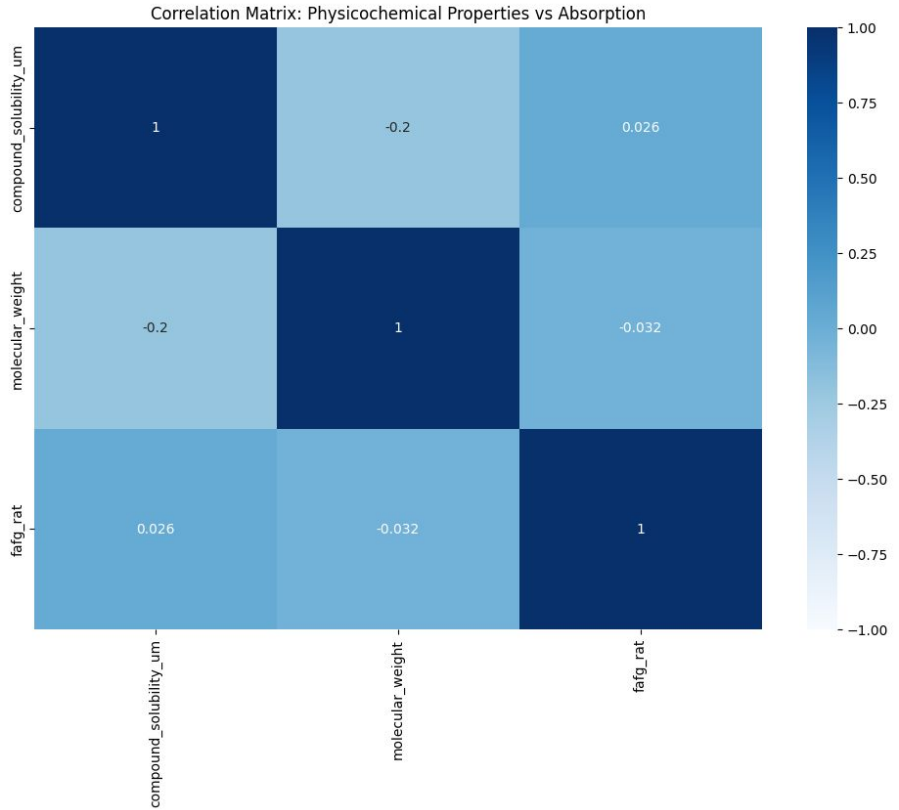


Correlation Matrix: In Vitro vs In Vivo Properties

Evaluating How Compound Solubility And Molecular Weight Affect Overall Fractional Absorption

Key insights

- ↑ Molecular Weight = ↓ Compound Solubility = ↓ Fractional Absorption
- ↑ Molecular Weight = ↓ Compound Solubility = ↓ Fractional Absorption
- ↑ Molecular Weight = ↓ Compound Solubility = ↓ Fractional Absorption



Correlation Matrix: Molecular Weight vs Fractional Absorption

Final Thoughts And Future Directions Based On Insights

Drug Development:

- Focus on optimizing molecular weight and clearance for pharmacological profile optimization.

Further Research:

- Apply machine learning algorithms to predict drug behavior, efficacy, and safety.
    a. Optimal criteria: ↑ Bioavailability, ↑ Fractional Absorption, ↑ Compound Solubility, ↑ Cell Permeability, ↑ pIC50, ↑ Primary Bioactivity, Moderate Clearance & Molecular Weight values.
    b. New target feature creation (defining a function for the optimal criteria).
    c. Model training and prediction
    d. Evaluating model performance
    e. Rank optimal compounds
- Validate findings with additional in vivo studies.

Recap!

Primary Objective:

To understand the relationships between the Compound X's properties, its efficacy, its safety and its biological targets.

Final Dataset:

About 2 million rows

Process:

Data integration, cleaning, preprocessing, standardization and visualization, feature selection and engineering, statistical and exploratory data analysis.

GitHub Link: https://github.com/emmaezeumeh/Pharmacological-Data-Analysis/

QUESTIONS / FEEDBACK