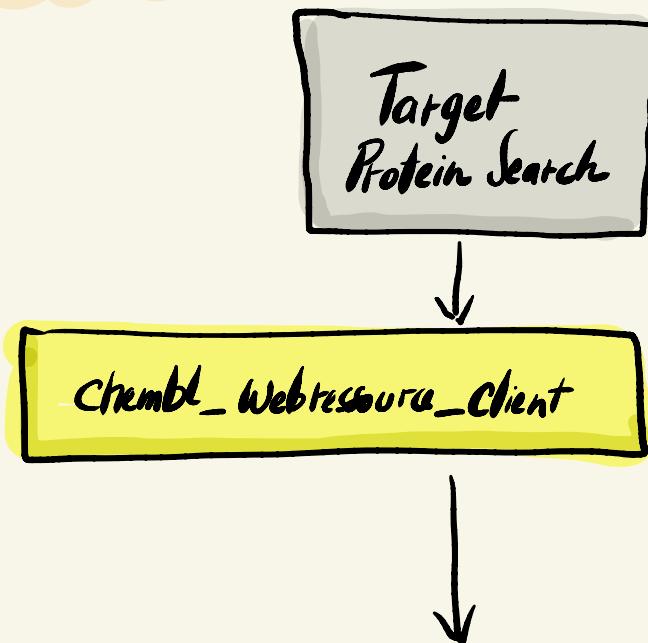


# COMPUTATIONAL MACHINE LEARNING IN DRUG DISCOVERY

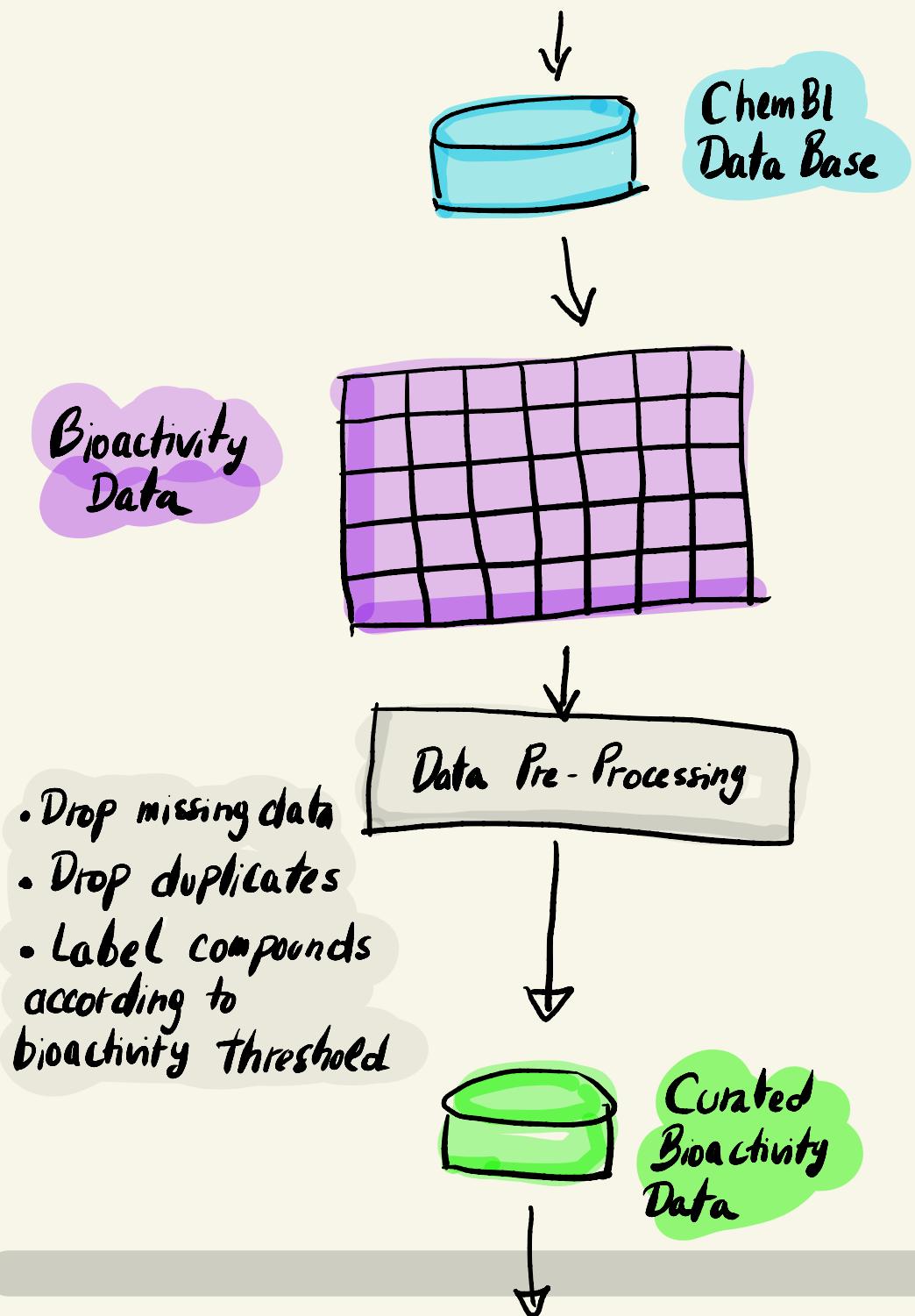
BY Khalid EL Akri

ChemCode Professor

Part 1  
Data Collection



Python libraries



## Part 02 Exploratory Data analysis

Clean SMILES notation

- Remove Small organic Compounds
- Remove Salt

CNCCACCNC

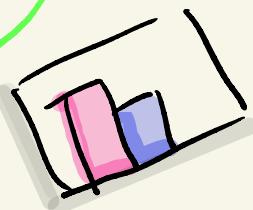
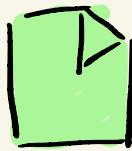
Calculate Lipinski's descriptors

- MW
- logP
- No. H-Bond donor
- No. H-Bond acceptor

Perform EDA

- Box plot
- Scatter Plot
- Statistical analysis

Export as CSV file



## Part 03 Descriptor Calculation

Read in CSV file as Data frame

↓

Prepare input file  
To PADEL-Descriptor



(Finger Print +  
Class Label)

Export as CSV  
file

↓

Calculate finger prints  
Using PADEL-Descriptor

Part 04  
Model  
Building

Read in CSV file as  
a Dataframe

↓

Remove low  
variance features

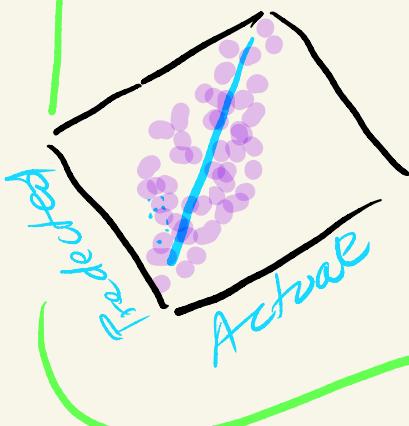
↓

Data splitting

↓

Build a regression  
model

Make a Scatter Plot



## Part 05 Model Comparison

Read in CSV  
file as a Data frame

Remove low variance  
features

Data  
Pre-processing

