YUSUF MUNIR ALIYU

Key EDA Findings

- pIC50 values showed clear separation between active and inactive compounds.
- LogP, NumHDonors, and NumHAcceptors showed significant differences between activity classes.
- Molecular Weight (MW) was not statistically significant.
- Boxplots and histograms indicated moderate skewness in some descriptors.

Lipinski and 2D Descriptor Statistics

- The majority of compounds complied with Lipinski's Rule of Five.
- Most compounds had:
  - MW < 500 g/mol
  - LogP < 5
  - NumHDonors ≤ 5
  - NumHAcceptors ≤ 10
- This suggests good drug-likeness and oral bioavailability potential.

Which tool or method did you use for fingerprint calculation?

PaDELpy (PubChem fingerprint module)

Why did you select this method?

- PubChem fingerprints are widely used in QSAR modelling.
- They provide standardised 881-bit structural features.
- Compatible with machine learning models.
- Easy integration with PaDELpy for batch descriptor generation.