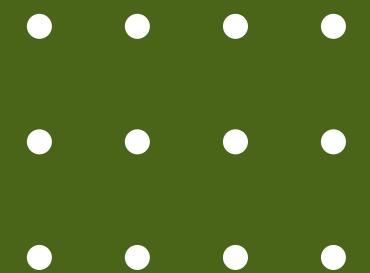


# CleanWise

Whole suite of data cleaning and wrangling

Ankita Shanbhag (Academic Researcher)



Domain  
Specificity

Privacy

Options!



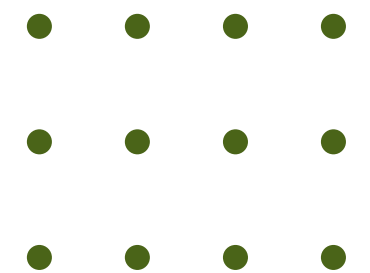
Query  
Support

Prompt  
Automation

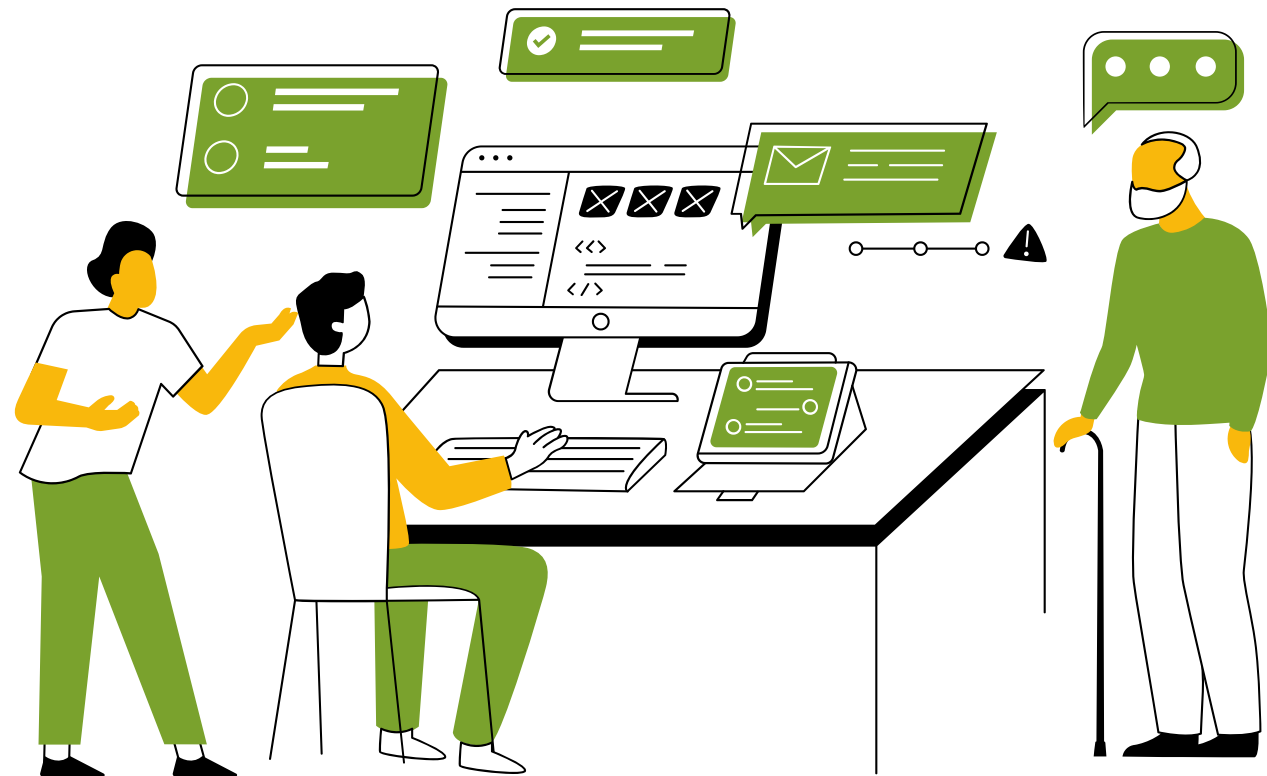
Data  
Augmentation

# Problem Setup

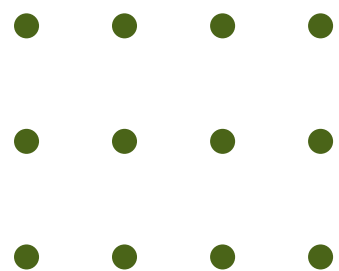
1. **Removing Duplicates** : identifying and eliminating duplicate records
2. **Filtering Outliers** : Identifying and assessing whether to keep or remove outliers
3. **Standardizing Data Formats** : Ensuring consistency in units, date formats
4. **Correcting Typos and Spelling Errors**
5. **Encoding Categorical Data** : Converting categorical data into a numerical format
6. **Data Consolidation** : Merging data from various sources and ensuring consistentancy and accuracy



# Benefits



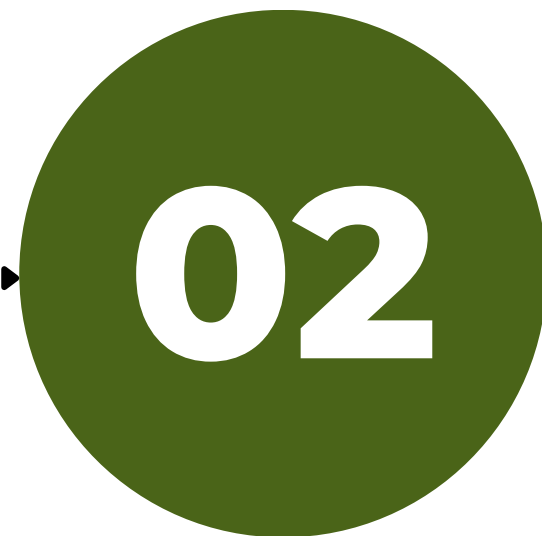
1. **Generalization & Adaptability:** Excel at handling a wide range of data types
2. **Unstructured Data Handling:** Skilled in interpreting and processing unstructured data
3. **Natural Language Processing:** Understands and executes instructions in natural language
4. **Innovative Applications:** Opens new possibilities for automating and improving data cleaning processes



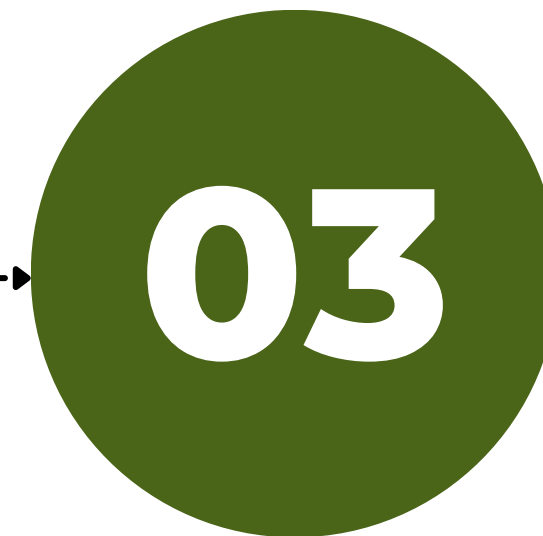
# Next Few Steps



Evaluate performance of FMs for the stated processes



Evaluate how they perform when multiple processes are combined



Automate wrangling and cleaning suggestions

**THANK  
YOU**