

Group Name: PharmaPersist

Team Member Details:

- Name : Yunxin Gan
- Email: [xin03040905@gmail.com](mailto:xin03040905@gmail.com)
- Country: United Kingdom
- College: University of Exeter
- Specialization: Data Science

Github Repo link: <https://github.com/YunxinG107112/VC/tree/main/week9>

Data cleansing and transformation done on the data

- Use two different approaches to deal with unknown values in categorical columns, for example, for the 'Unkown' values in the Ntm\_Speciality and Risk\_Segment\_During\_Rx columns, we applied imputation based on the most frequent values of them.
- All categorical columns were encoded using Label Encoder, for example the column Gender was encoded so that it has only two values 0 and 1 to represent male and female.