

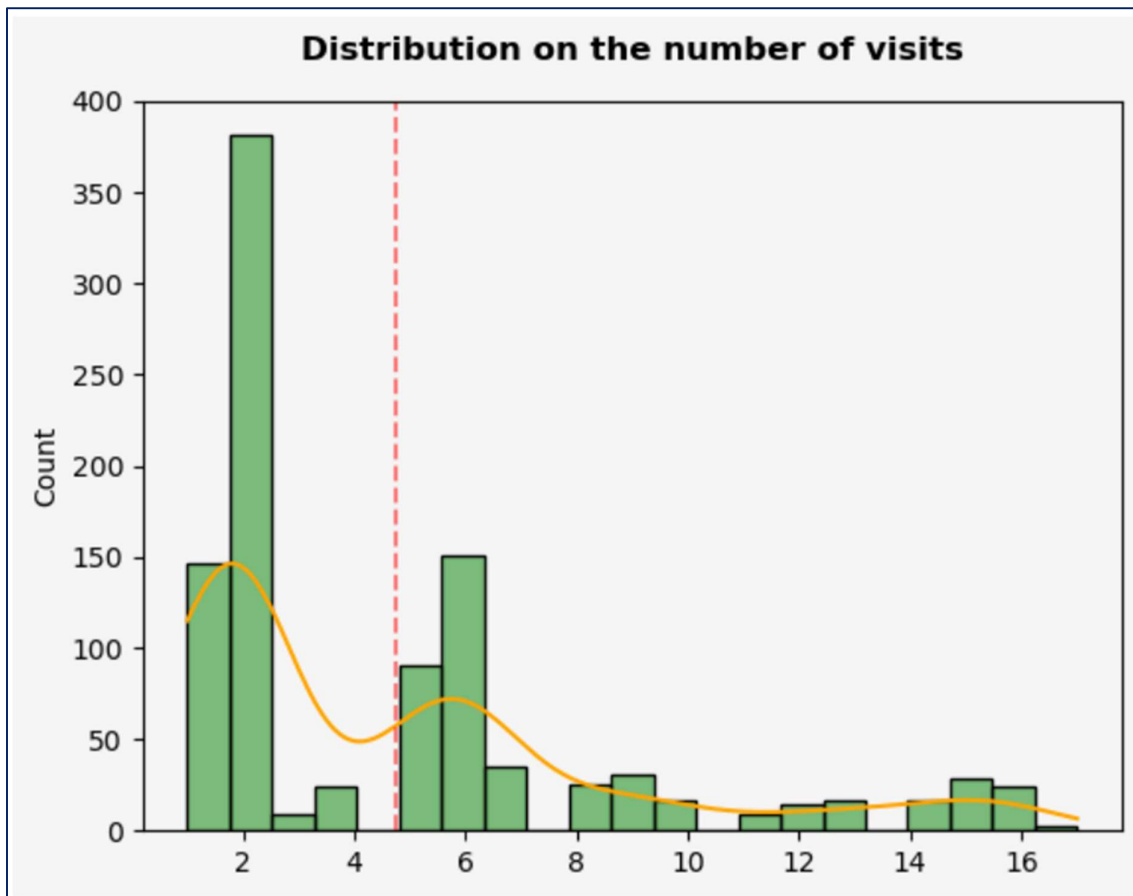
EDA Using Visualization Techniques

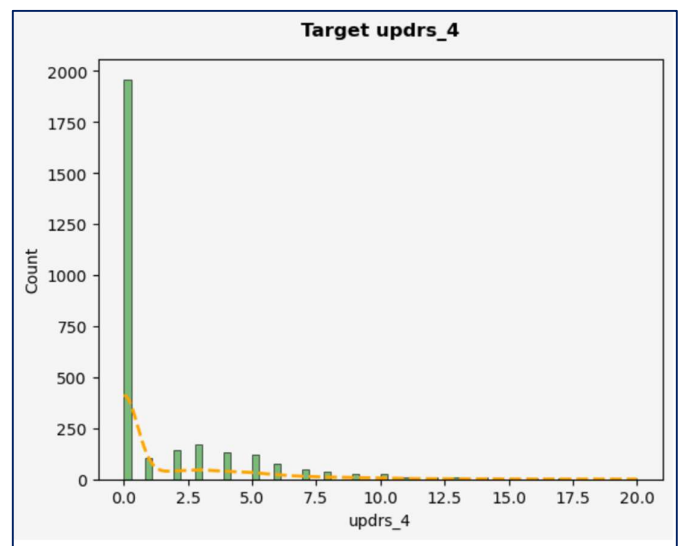
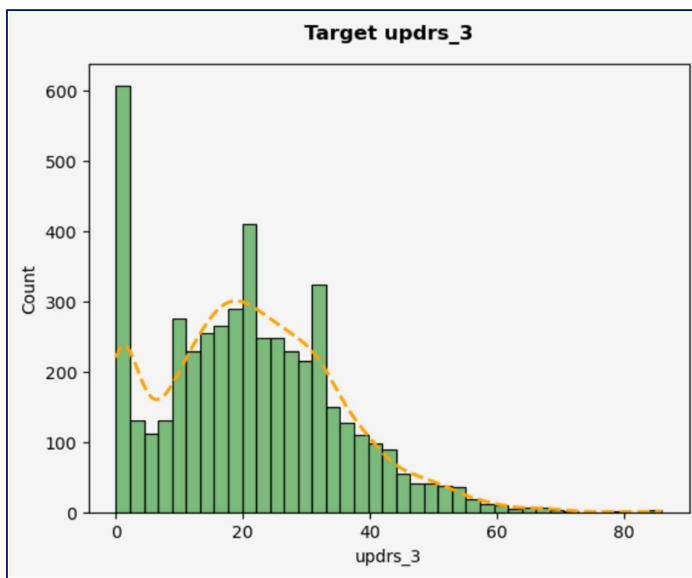
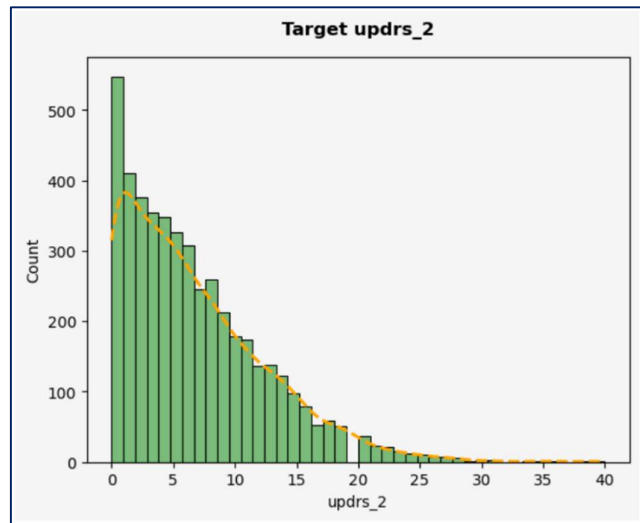
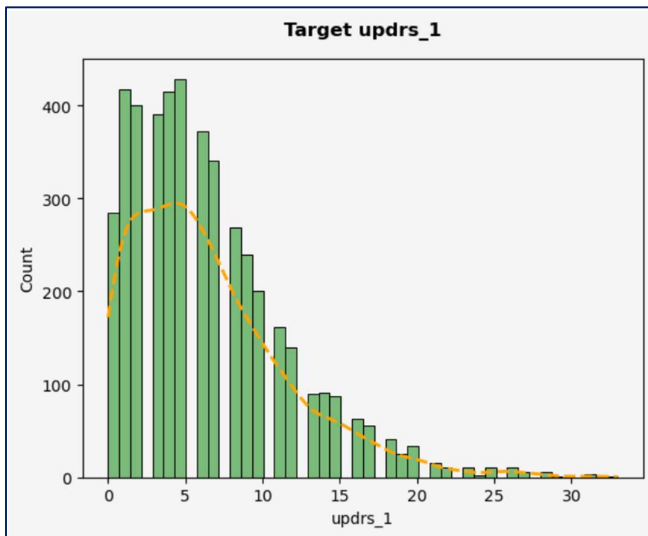
ASSIGNMENT 1
211AI016, 211AI018

TEAM 18 | Dhiraj Choudhary D, Venkata Ravi Ram G

Clinical data

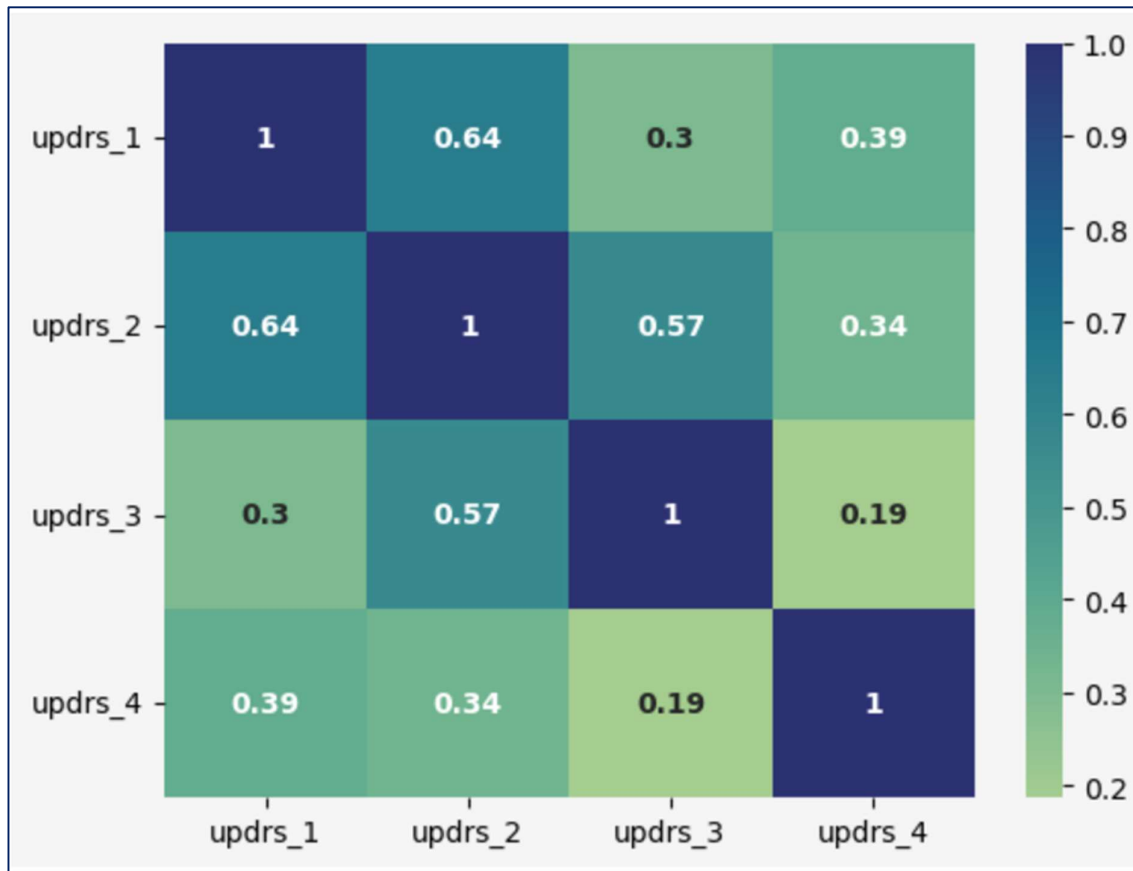
This graph visualizes the distribution of the number of visits per patient in a clinic and highlighting the mean number of visits (red line). histogram is coloured green and shows a kernel density estimate (kde) yellow line. The kde plot provides a smooth curve that approximates the shape of the distribution and can be useful for identifying features of the data such as peaks or modes.





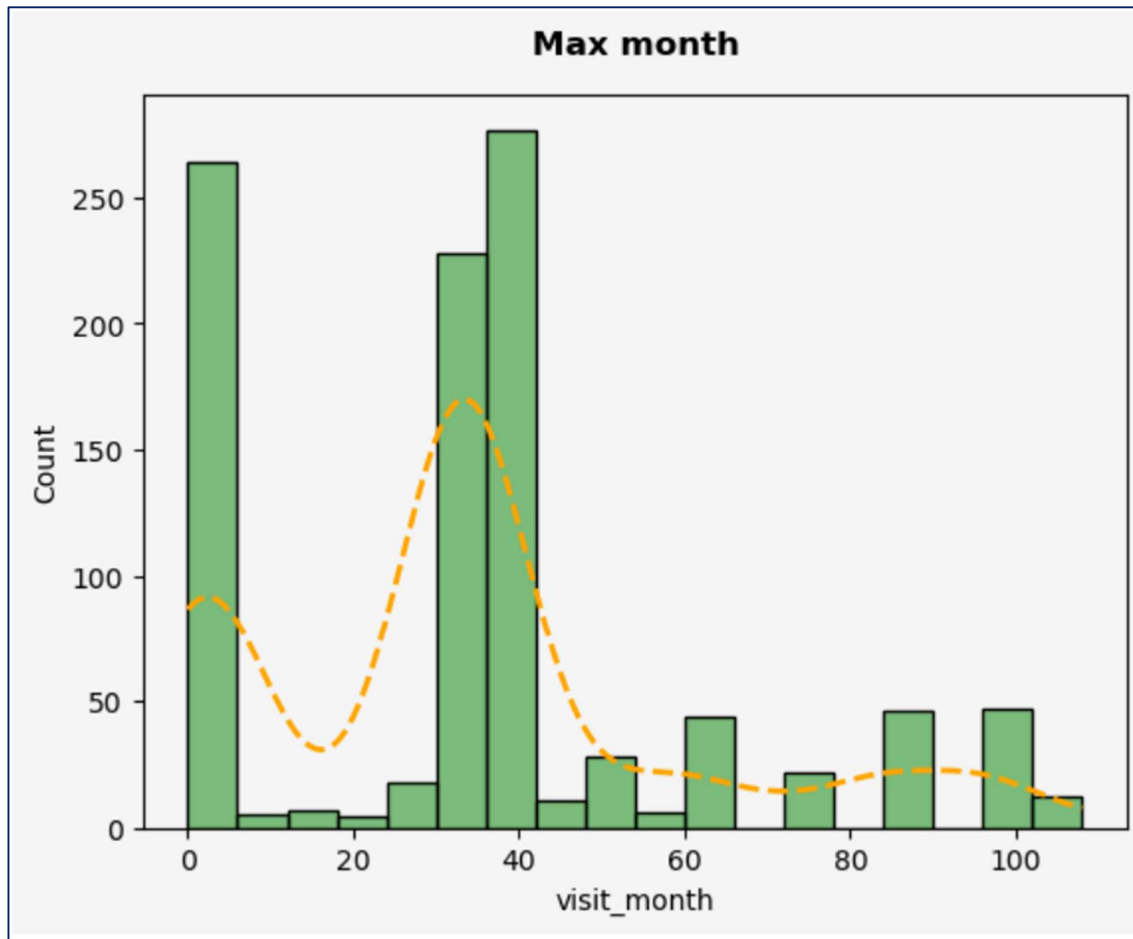
Inferences:

- All targets are not normal distribution
- updrs_4 has zeros as much as about 40% **(It seems to be good to impute with zeros)**
- all the plots are right/positive skewed



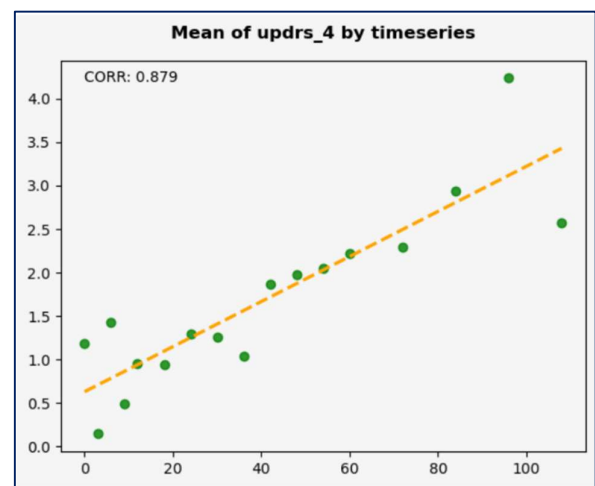
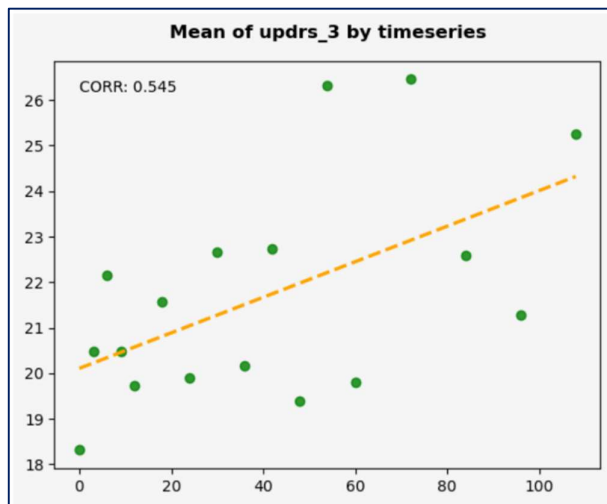
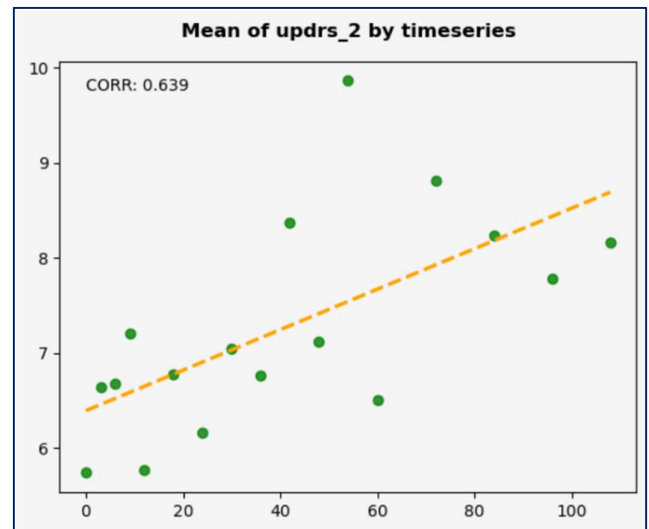
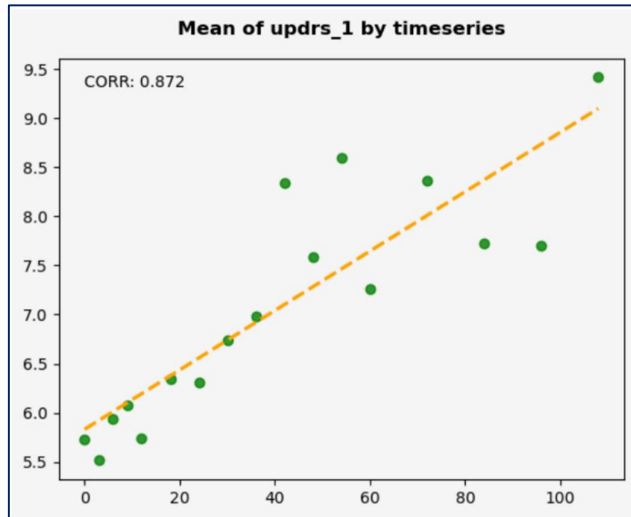
Inferences:

- updrs_2 has high correlation with **updrs_1 & updrs_3** (about 0.6)



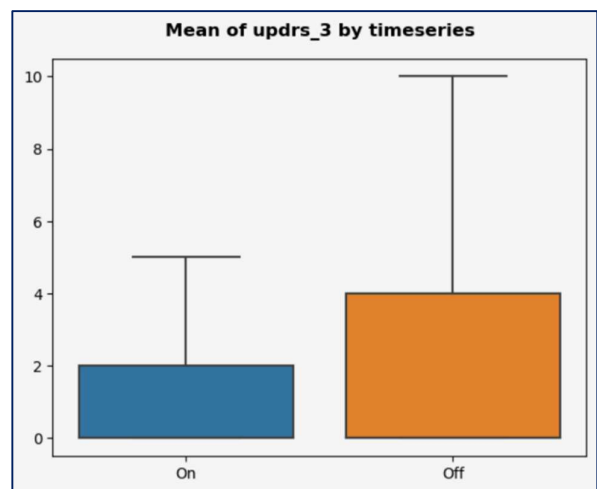
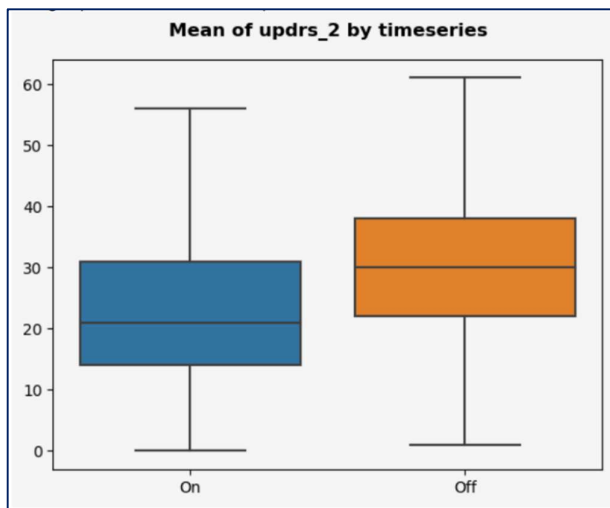
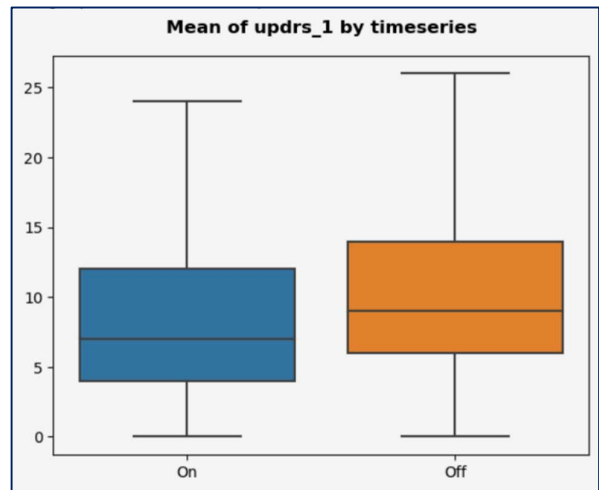
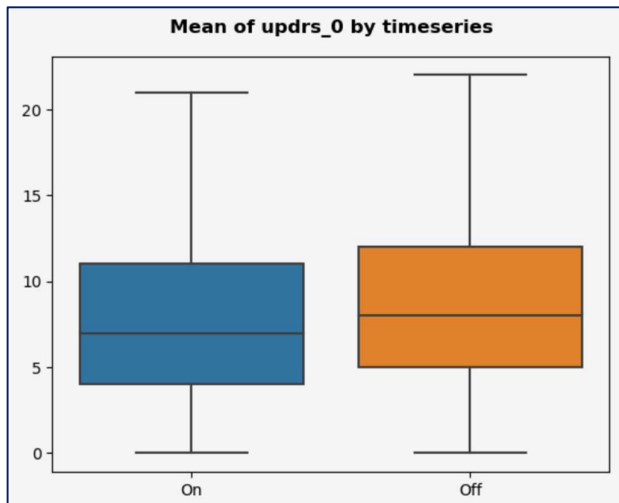
Inferences:

Short-term(0-5) and mid-term(30-36) bins show high medical examination count



Inferences:

- **Medication is effective** on reducing the updrs level



Inferences:

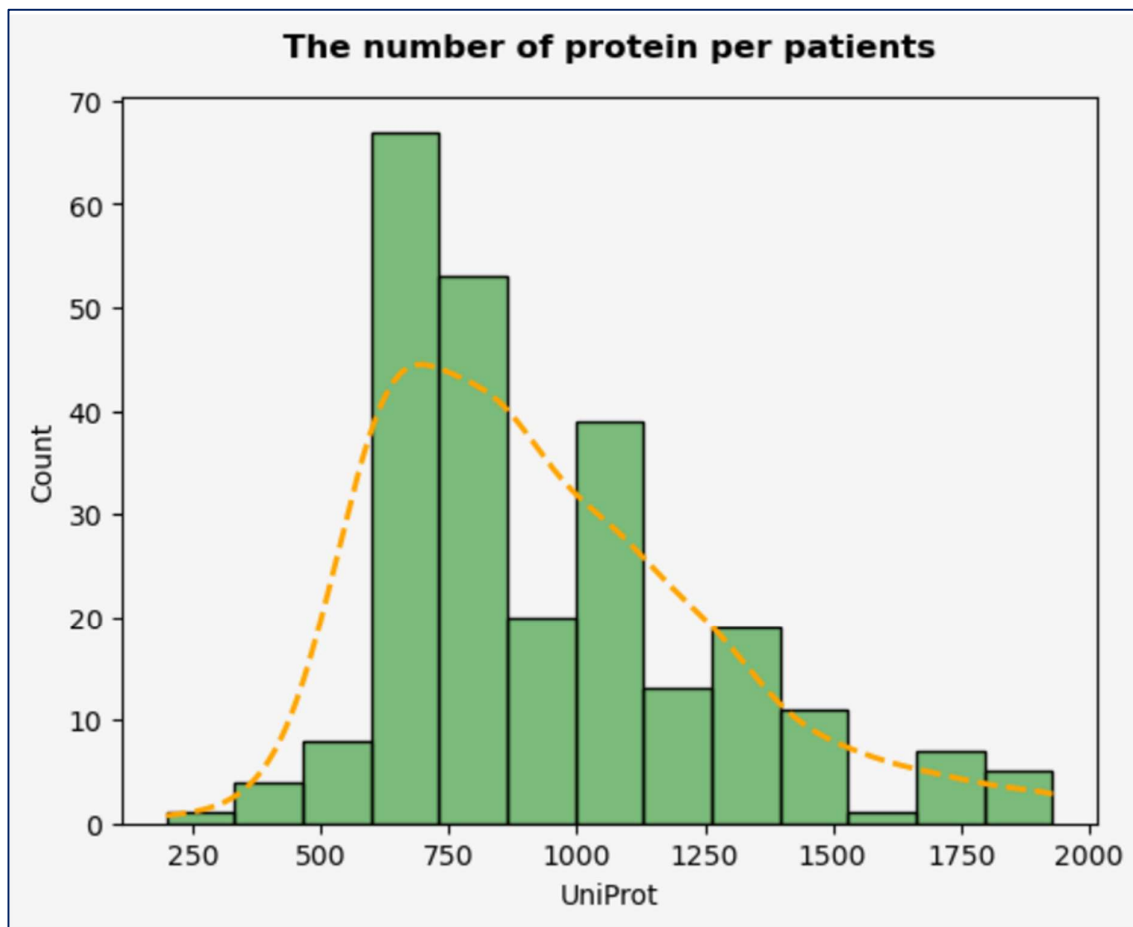
- **Medication is effective** on reducing the updrs level
- **This matches with the above inferred information from timeseries data**
- Provided original data and additional data statistically has no difference mean in all ndprs values

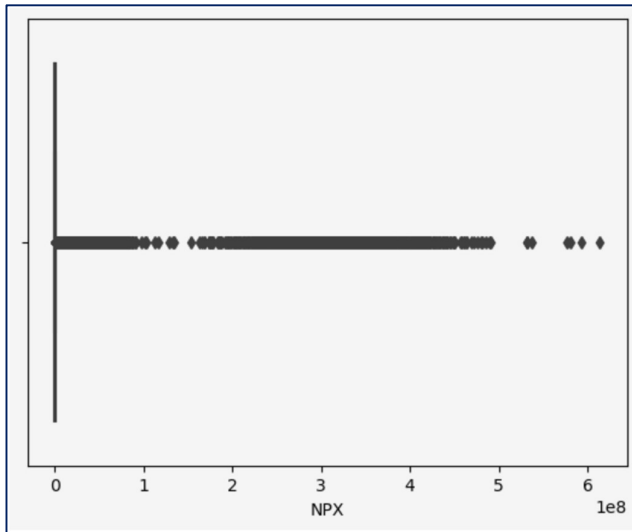
Clinic data

Summary

- All targets are not normal distribution
- updrs_4 has zeros as much as about 40% (**It seems to be good to impute with zeros**)
- **Short-term(0-5) and mid-term(30-36)** bins show high medical examination count
- **Medication is effective** on reducing the updrs level
- Some patients' medication are sometimes not applied
- Provided original data and additional data statistically has no difference mean in all ndprs values
- Therefore, We can use additional data as same way as original data **without anxiety which be able to have bias**

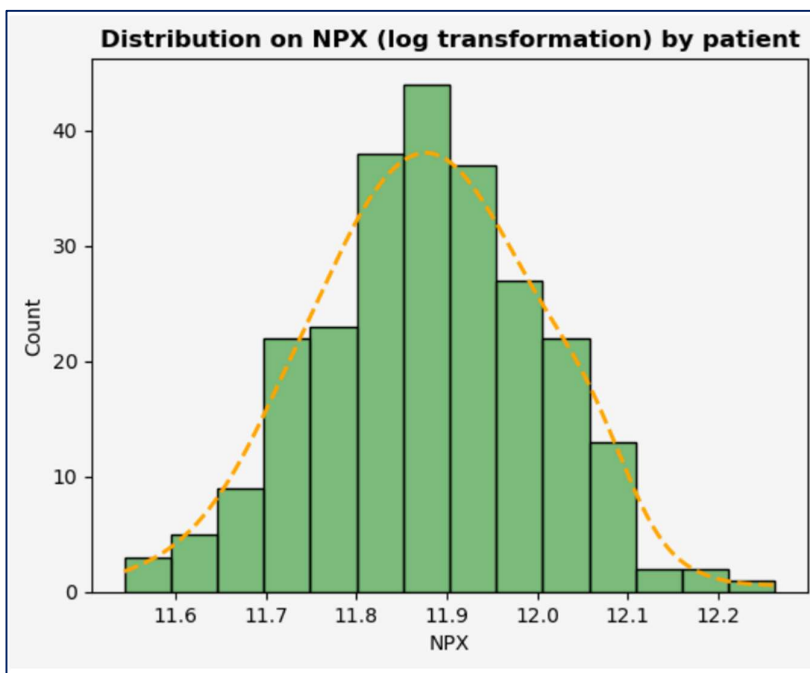
Protein data





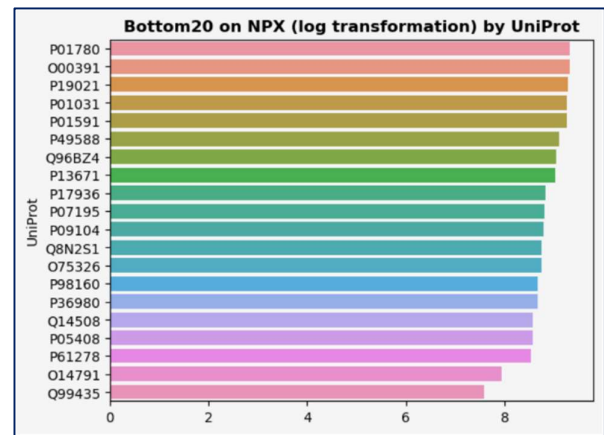
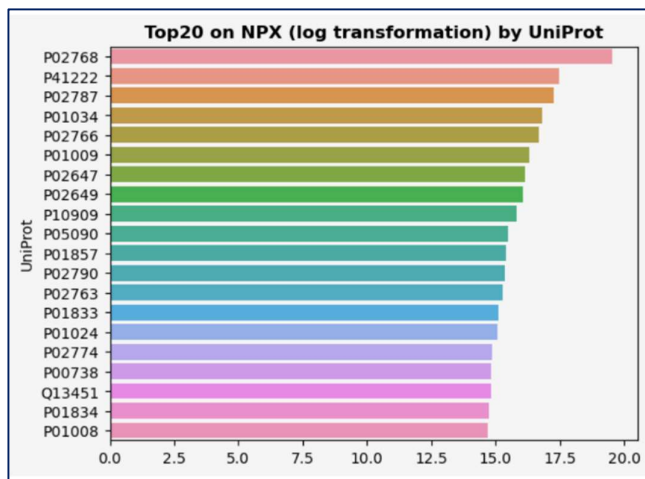
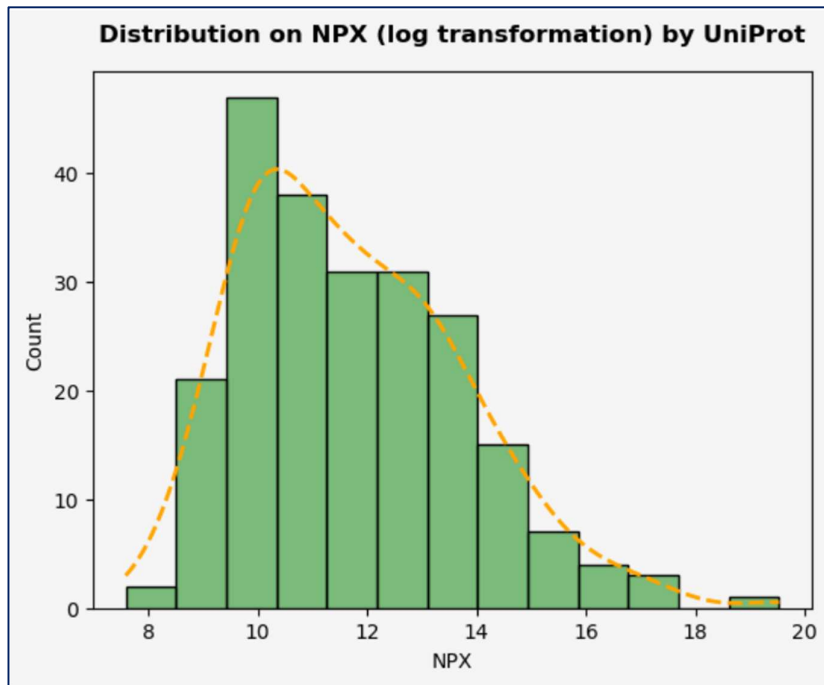
Inferences:

NPX seems to increase exponentially **(After log transformation, it shapes normal distribution)**



Inferences:

- By patient, average of mean is about 11.88
- In view of patient, It looks like perfectly normal distribution more than overall NPX

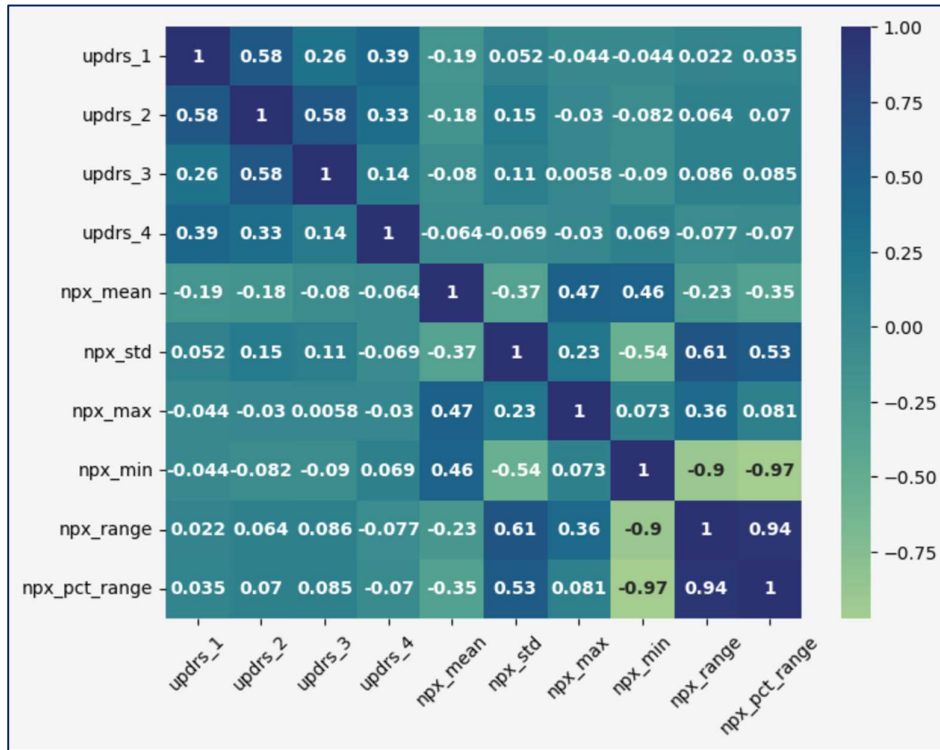


Inferences:

- Some proteins shows high NPX values than others

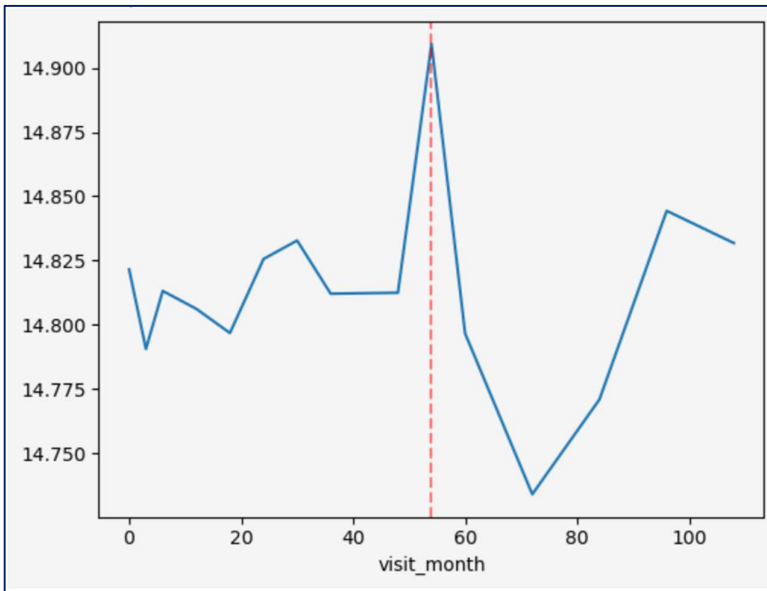
Therefore, it is important which protein the patient has

EDA USING VISUALIZATION TECHNIQUES

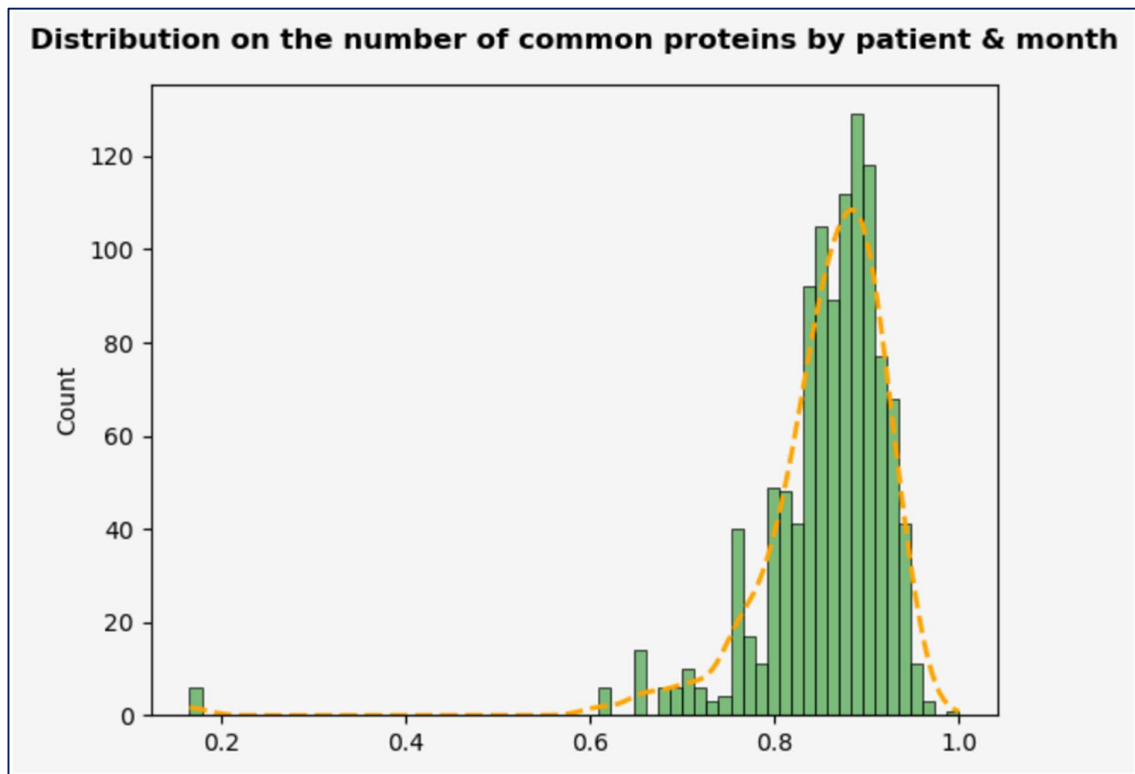


Inference:

- Mean of NPX shows **slight negative correlation** with updrs targets
- Standard deviation of NPX shows **slight positive correlation** with updrs 1-3



This plot shows the avg visit month against protein values



Inferences:

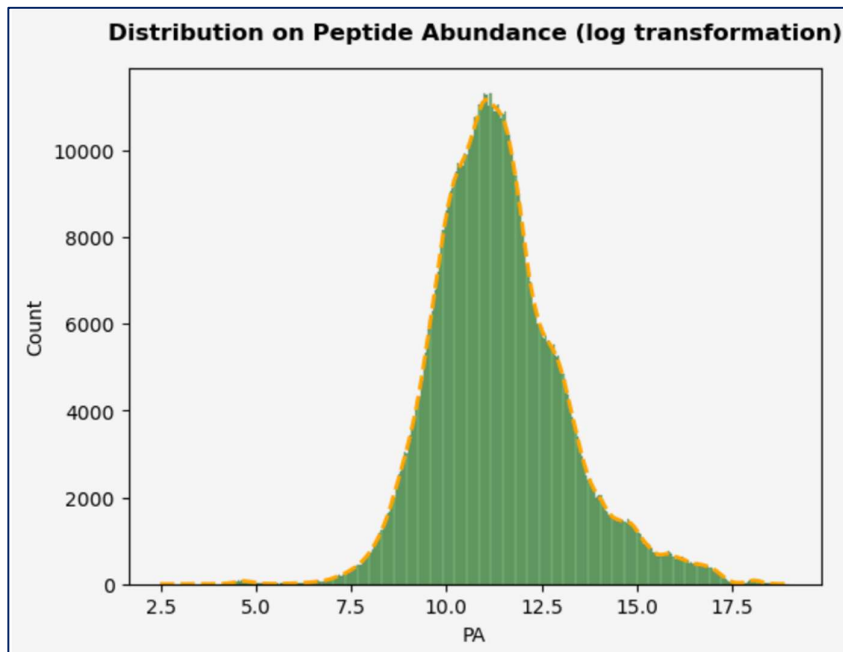
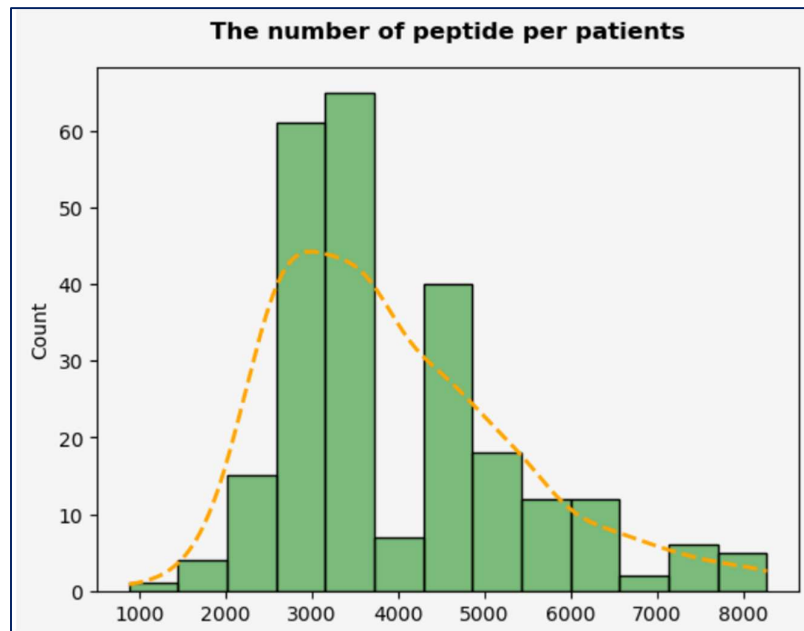
- By patient & month, common protein exists (**about 85%**)

Protein data

Summary

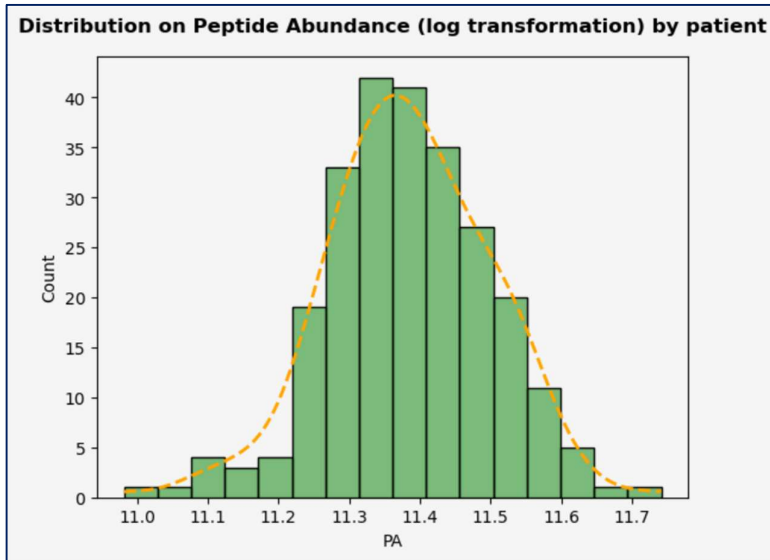
- NPX seems to increase exponentially (**After log transformation, it shapes normal distribution**)
- By patient, average of mean is about 11.88
- In view of patient, It looks like perfectly normal distribution more than overall NPX
- Some proteins show high NPX values than others
- Therefore, it is important which protein the patient has
- Mean of NPX shows **slight negative correlation** with updrs targets
- Standard deviation of NPX shows **slight positive correlation** with updrs 1-3

Peptide data



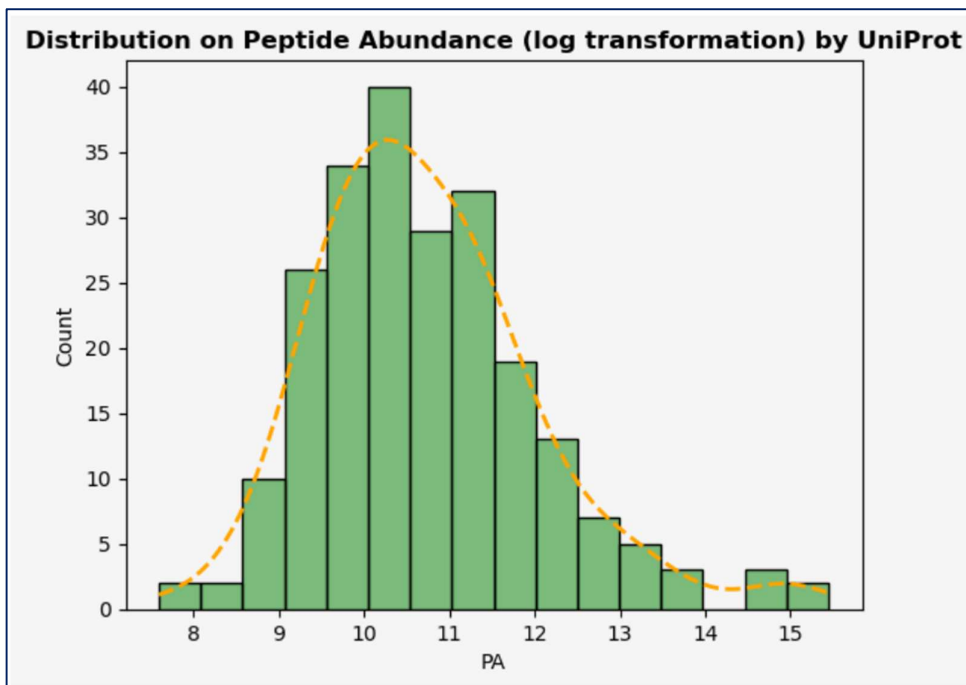
Inferences:

- Peptide Abundance seems to increase exponentially (**After log transformation, it shapes normal distribution**)



Inferences:

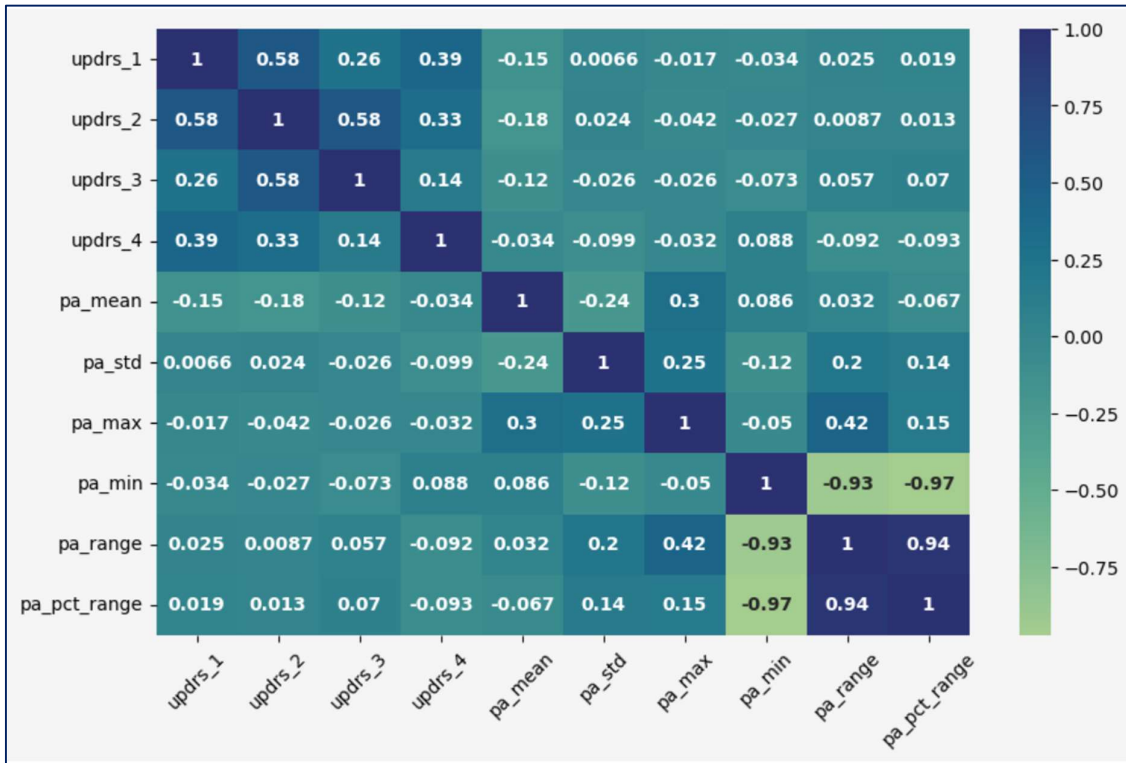
- By patient, average of mean is about 11.38



Inferences:

- Some proteins shows high Peptide Abundance values than others

Therefore, it is important which peptide the patient has



inferences:

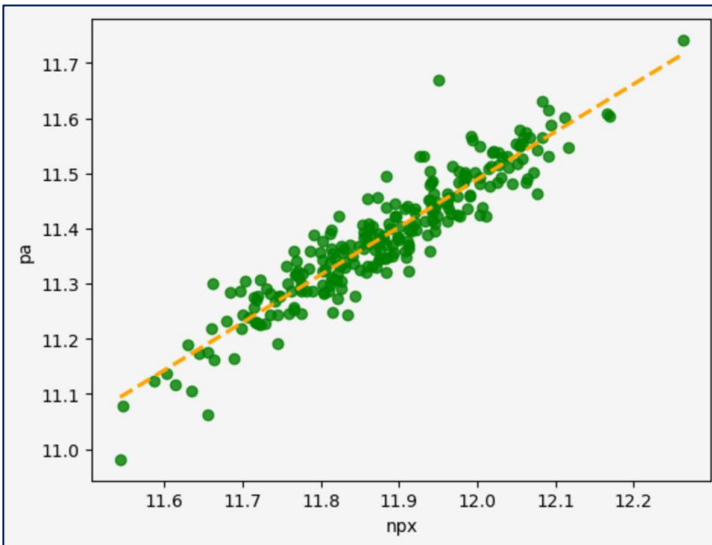
- Mean of Peptide Abundance shows **slight negative correlation** with updrs targets
- Standard deviation of Peptide Abundance **doesn't show any correlation** with updrs targets (corr < 0.1)

Peptide data

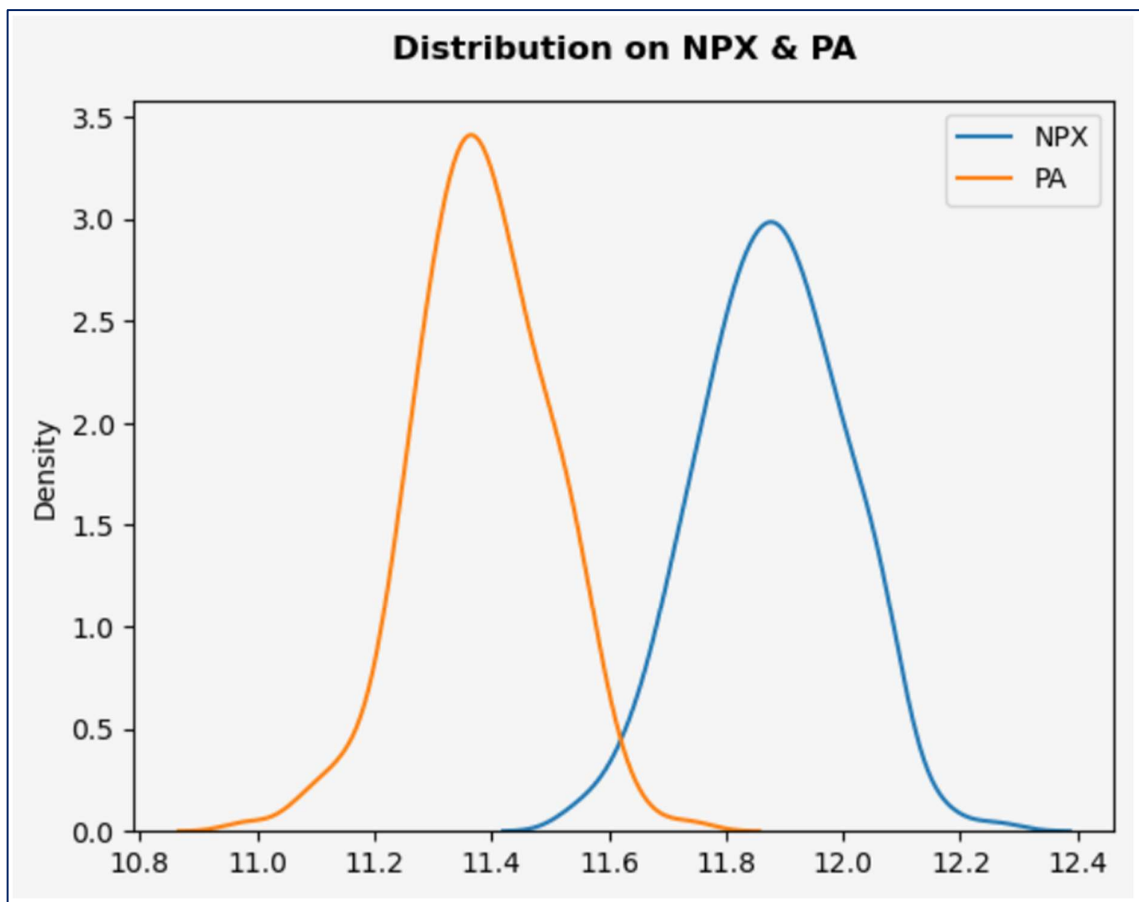
Summary

- Peptide Abundance seems to increase exponentially (**After log transformation, it shapes normal distribution**)
- By patient, average of mean is about 11.38
- Some proteins shows high Peptide Abundance values than others
- Therefore, it is important which peptide the patient has
- Mean of Peptide Abundance shows **slight negative correlation** with updrs targets
- Standard deviation of Peptide Abundance **doesn't show any correlation** with updrs targets (corr < 0.1)

Protein & Peptide joined data



scatter plot between pa and npx
protein values



Inferences:

- NPA & PA are **strong positive correlation** (mean group by patient)

Protein & Peptide join data

Summary

- NPA & PA are **strong positive correlation** (mean group by patient)
- Not all protein & peptide information are matched to clinic data (**Only about 22%**)
- It is hard to say that 'There are some common protein or peptide among patients'
- It is hard to say that 'There are some common protein or peptide among months'
- Visit ids are completely overlapped in both protein and peptide data