

Question 1 :

The missing values are filled with the mean of values of that feature in the same class.
When the mean itself is missing, the missing values is calculated as mean of all data of same class so that even if the data for whole state is missing it can be filled.

Question 2 :

Top 5 state representing India without normalizing are :

1. Maharashtra
2. Tamil Nadu
3. Uttar Pradesh
4. Gujarat
5. Karnataka

Top 5 state representing India with normalizing are :

1. Maharashtra
2. Karnataka
3. Rajasthan
4. Gujarat
5. Uttar Pradesh

Question 3 :

From the correlation plots we can see that :

- **Demography** does not have much linear dependency between features except for features at start and end.
- **Economy** has a lot of linear dependency between its features.
- **Education** does not have a lot of linear dependent features i.e. most of the features are linear independent.
- The correlation figure of all data suggests that is very less linear dependency among features.