**#SURVEY**

1 - Load the dataset into a pandas dataframe. Name the variable as “survey”.

2 - How many samples were collected on each day?

3 - What proportion of the total respondents were aged less than 45?

4 - Create a new column in the dataframe “age\_group”. This column should contain the age group the respondent belongs to. The age groups are 18-25, 25-40, 40-55 and 55+. The dataframe should look like this after the column creation.

5 - How many samples were collected for each age-group? Which age-group had the most samples?

6 - What proportion of the respondents had opted for the RJD party in both the Vote\_Now and the Past\_Vote questions?

7 - For each day of sample collection, determine the proportion of respondents who were fully satisfied with the performance of the CM. So if there were a total of 1000 samples on day 1 and 300 out of those said they were fully satisfied, then our answer for that day would be 0.3.

8 - In a similar fashion create a day-wise proportion of respondents that opted fully dissatisfied with their MLA. Create a line plot of the result with date on x-axis and proportions on the y-axis.

9 - Create a pivot-table (or crosstab) with index as Past\_Vote, Column as Vote\_Now and cell values as the count of samples. Answer - survey.pivot\_table(index = 'Past\_Vote', columns = 'Vote\_Now', values = 'response\_id', aggfunc = 'count')

10 - Repeat the above question with the cell values as the sum of “weight”.

11 - Create a dataframe by performing a group by over age\_group and calculate the count of total samples under each age\_group.

12 - Create a dataframe by performing a group by over age\_group and finding the count of total samples for each age\_group that opted for the JD(U) party in Vote\_Now.

13 - Join/Merge the two dataframes from questions 12 and 11 with the common column as age\_group.