

Data Description:

A sample survey data (Sample Survey.xlsx) is provided for analysis. Samples were collected over a period of nine days where the respondents were asked questions related to elections. Survey weights are in the “weight” column. The description for each column is provided in the excel sheet under tab named “Description”. Applicants are requested to use *python-pandas* for all data manipulations and *matplotlib* for any plotting related questions. Applicants are expected to not use *for-loops* or any kind of *iterations* for solving the questions. Use in built *pandas* functions only.

****All samples are synthetic and do not correspond to any actual survey conducted by Dhruv.**

Questions:

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:

response_id	age	age_group
1	26	25-40
2	34	25-40
3	25	25-40
4	66	55+

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.
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 13 with the common column as age_group.