I’m writing this report to explain how I did wrangling the data , first of all I gathered dada from 3 different resources the first one about csv file so it was so easy to gather data from this file because I practiced a lot with the same file then the second recourse (tsv) was the same just I need to add \t so the result would be the same as csv then the third resource was so difficult to gather because I did work with json file before but I follow the instructions in the lessons and ask my session lead about it then I open it and loaded the file and I take this information from it('tweet_id' ,'created_at' , 'entities', 'truncated', 'full_text', 'favorite_count' , 'retweet_count') after that I converted to dataframe and now after gathering these files I’m ready to go to the next phase which assessment but first of all I should be make copies for these data then started the assessment phase so I had 2 ways to do the assessments , first one is visually and the second by coding so I looked in excel for csv file and the other I just assessed by coding and use these codes (info,head,sample,tail) and I found a lot of mistakes for quality and tidiness and quality issues like data type and missing value and writing mistakes for names and duplicated rows and also I decided to rename some columns with appropriate names, after that I looked for tidiness issues like the dogs category should be in one column and also I should extract the tweet id from expanded_urls and finally join the df2 and df3 together so it would be easy to work with just 2 files and now I’m ready to move on to the next phase which is cleaning so I cleaned all the mistakes I notify in assessments phase by using pandas libraries then I iterate these two phases (assess and clean) until I satisfied with results to I store as csv file .

So now after I gathered data that I need from three different source s then I did assessments and cleaned the quality and tidiness issu es now I assured that my data is ready to the next stage which is a nalysis and visualization and here I'm going to explain the insights and visuals that I made in my analysis step by step so let's begin , first of all I deleted the columns that I don't need in my analysis th en I asked some question like what is the name the most dogs ha d and I used value _ counts code so I found these names (Charlie, Cooper ,Oliver) for the most dogs then I asked what is the most br eed and I used the same code which is value _ counts and I found the results like that pupper by 137 ,doggo by 48,puppo by 22, and f loofer by 2 then I made pia chart for that and asked the third quest ion which was what is the most category had ratings and it was do ggo then floofer then puppo the pupper and I made bar chart so it could be easy to see the difference between them and the other in sight I show the most prediction done in these(golden _ retriever ,, Labrador _ retriever ,,Pembroke ,,Chihuahua)
And I asked what the images are had the most like the I answer thi s question visually and I show them visually. And also, I did the sa me thing with the images had the most retweet and also, I show th em visually and that was the end of my analysis.