As we saw when we displayed class\_size on the last screen, DBN still isn't completely unique. This is due to the CORE COURSE (MS CORE and 9-12 ONLY) and CORE SUBJECT (MS CORE and 9-12 ONLY) columns.

CORE COURSE (MS CORE and 9-12 ONLY) and CORE SUBJECT (MS CORE and 9-12 ONLY) seem to pertain to different kinds of classes. For example, here are the unique values for CORE SUBJECT (MS CORE and 9-12 ONLY):

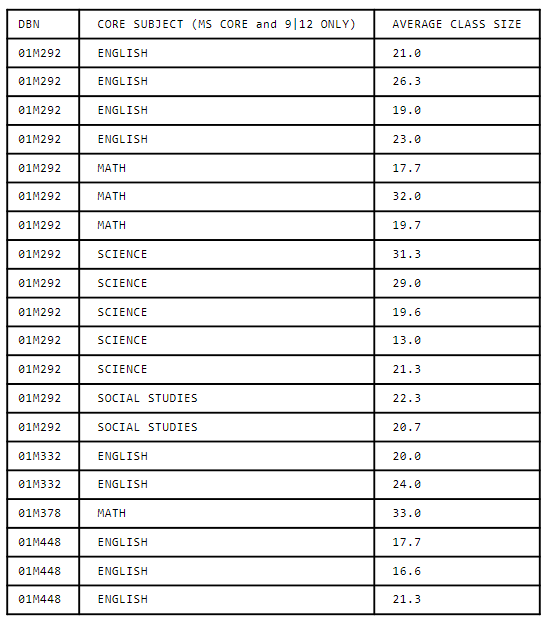


array(['ENGLISH', 'MATH', 'SCIENCE', 'SOCIAL STUDIES'], dtype=object)

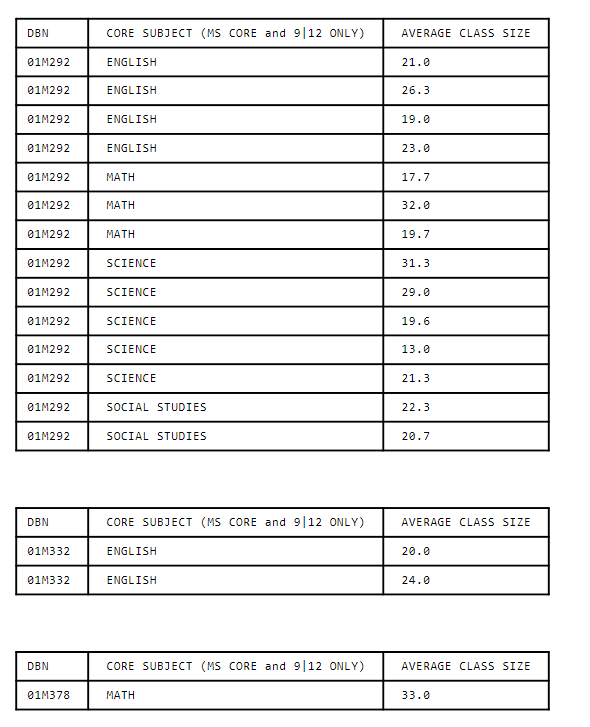
This column only seems to include certain subjects. We want our class size data to include every single class a school offers -- not just a subset of them. What we can do is take the average across all of the classes a school offers. This will give us unique DBN values, while also incorporating as much data as possible into the average.

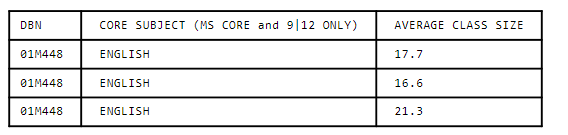
Fortunately, we can use the [pandas.DataFrame.groupby()](http://pandas.pydata.org/pandas-docs/stable/groupby.html" \t "_blank) method to help us with this. The DataFrame.groupby() method will split a dataframe up into unique groups, based on a given column. We can then use the [agg()](http://pandas.pydata.org/pandas-docs/stable/groupby.html" \l "aggregation" \t "_blank) method on the resulting pandas.core.groupby object to find the mean of each column.

Let's say we have this data set:



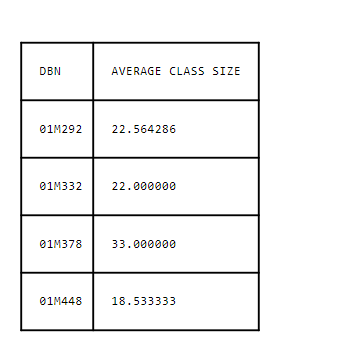
Using the groupby() method, we'll split this dataframe into four separate groups -- one with the DBN 01M292, one with the DBN 01M332, one with the DBN 01M378, and one with the DBN 01M448:





Then, we can compute the averages for the AVERAGE CLASS SIZE column in each of the four groups using the agg()method:

DBN



After we group a dataframe and aggregate data based on it, the column we performed the grouping on (in this case DBN) will become the index, and will no longer appear as a column in the data itself. To undo this change and keep DBN as a column, we'll need to use [pandas.DataFrame.reset\_index()](http://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.reset_index.html" \t "_blank). This method will reset the index to a list of integers and make DBN a column again.