We've finished cleaning and combining our data! We now have a clean data set on which we can base our analysis. Mapping the statistics out on a school district level might be an interesting way to analyze them. Adding a column to the data set that specifies the school district will help us accomplish this.

The school district is just the first two characters of the DBN. We can apply a function over the DBNcolumn of combined that pulls out the first two letters.

For example, we can use indexing to extract the first few characters of a string, like this:



name = "Sinbad"

print(name[0:2])

Instructions

* Write a function that extracts the first two characters of a string and returns them.
* Apply the function to the DBN column of combined, and assign the result to the school\_dist column of combined.
* Display the first few items in the school\_dist column of combined to verify the results.

# create a funtion that passes a string

def extract\_district(string):

# return only first two chars

return string[0:2]

# create a new column in combined with the first two letters of DBN (apply the function)

combined["school\_dist"] = combined["DBN"].apply(extract\_district)

# print a few lines

print(combined["school\_dist"].head())