Data Cleaning Walkthrough: Combining the Data

Syntax

- Reseting the index:

```
class_size.reset_index(inplace=True)
```

- Grouping a dataframe by column:

```
class_size=class_size.groupby("DBN")
```

- Aggregating a grouped Dataframe:

```
class_size = class_size.agg(numpy.mean)
```

- Displaying column types:

```
data["ap_2010"].dtypes
```

- Performing a left join:

```
combined.merge(data["ap_2010"], on="DBN", how="left")
```

- Displaying the shape of the dataframe (row, column):

```
combined.shape
```

- Performing an inner join:

```
combined = combined.merge(data[class_size], on="DBN", how="inner")
```

- Filling in missing values:

```
combined.fillna(0)
```

Concepts

- Merging data in Pandas supports four types of joins -- `left`, `right`, `inner`, and `outer`.

- Each of the join types dictates how pandas combines the rows.
- The strategy for merging affects the number of rows we end up with.
- We can use one or multiple aggregate functions on a grouped dataframe.

Resurces

- Data Cleaning with Python
- Dataframe.groupby()
- agg() documentation