

Data Cleaning Walkthrough: Analyzing and Visualizing the Data: Takeaways



by Dataquest Labs, Inc. - All rights reserved © 2021

Syntax

- Finding correlations between columns in a dataframe:

```
combined.corr()
```

- Specifying a plot type using `Dataframe.plot()`:

```
combined.plot.scatter(x='total_enrollment', y='sat_score')
```

- Converting a Pandas series to list:

```
longitudes = combined["lon"].tolist()
```

Concepts

- An r value measures how closely two sequences of numbers are correlated.
- An r value ranges for `-1` to `1`.
- An r value closer to `-1` tells us the two columns are negatively correlated while an r value closer to `1` tells us the columns are positively correlated.
- The r value is also called Pearson's correlation coefficient.

Resources

- [R value](#)
- [pandas.DataFrame.plot\(\)](#)
- [Correlation](#)
- [Guess the Correlation](#)

Takeaways by Dataquest Labs, Inc. - All rights reserved © 2021