

Module 3: Introduction to Data Analysis

Glossary

Boolean

A data type with two possible values: true or false.

https://www.w3schools.com/python/python_booleans.asp

CRISP-DM

CRoss Industry Standard Process for Data Mining (CRISP-DM)

filter

Makes a subset the dataframe rows or columns according to the specified index labels.

<https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.filter.html>

groupby

A groupby operation involves some combination of splitting the object, applying a function, and combining the results. This can be used to group large amounts of data and compute operations on these groups.

<https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.groupby.html>

Matplotlib

A comprehensive library for creating static, animated, and interactive visualizations in Python.

<https://matplotlib.org/>

Plotly

Plotly provides online graphing, analytics, and statistics tools for individuals and collaboration, as well as scientific graphing libraries for Python, R, MATLAB, Perl, Julia, Arduino, and REST.

<https://plotly.com/>

Seaborn

A Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.

<https://seaborn.pydata.org/>