Glossary terms from module 1

Terms and definitions from Course 3, Module 1

Bias: In data structuring, refers to organizing data results in groupings, categories, or variables that are misrepresentative of the whole dataset

Cleaning: The process of removing errors that might distort your data or make it less useful; one of the six practices of EDA

Data visualization: A graph, chart, diagram, or dashboard that is created as a representation of information

Discovering: The process data professionals use to familiarize themselves with the data so they can start conceptualizing how to use it; one of the six practices of EDA

Exploratory data analysis (EDA): The process of investigating, organizing, and analyzing datasets and summarizing their main characteristics, often by employing data wrangling and visualization methods; the six main practices of EDA are: discovering, structuring, cleaning, joining, validating, and presenting

Joining: The process of augmenting data by adding values from other datasets; one of the six practices of EDA

PACE: A workflow data professionals can use to remain focused on the end goal of any given dataset; stands for plan, analyze, construct, and execute

Presenting: The process of making a cleaned dataset available to others for analysis or further modeling; one of the six practices of EDA

Structuring: The process of taking raw data and organizing or transforming it to be more easily visualized, explained, or modeled; one of the six practices of EDA

Validating: The process of verifying that the data is consistent and high quality; one of the six practices of EDA