

Exploratory Data Analysis (EDA)

Introduction



Understanding your data is crucial before trying to build a model



John Tukey coined the term “Exploratory Data Analysis” in 1977, he’s best known for the development of the boxplot.



Can be leveraged to suggest which hypotheses to test (in the future)

Some Methods

We'll cover all of these except for dimensionality reduction in this module

Summary statistics



Histograms



Box plots



Scatter plots



Dimensionality reduction

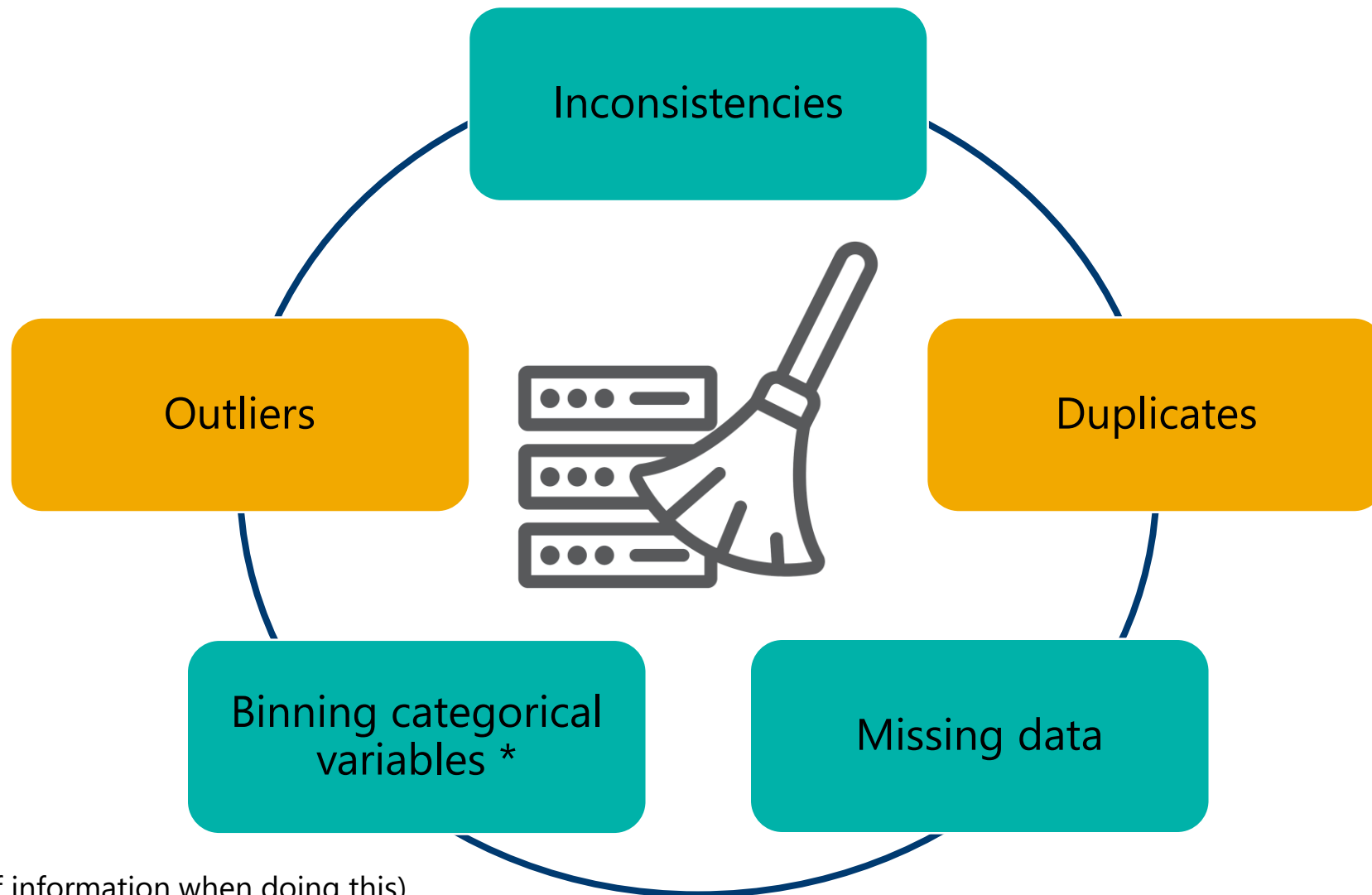


Correlations



Things You'll Find And Handle During the Process

This falls under "**data cleaning**". EDA and data cleaning are an iterative process.



*(There is a loss of information when doing this)

Summary

Summary
statistics



Histograms



Box plots



Scatter plots



Dimensionality
reduction



Correlations

