Train In Data

Complete Case Analysis

Complete case analysis

- Complete-case analysis ➔ CCA
- Also called "list-wise deletion" of cases.

CCA definition

- CCA consists in **discarding** observations with missing data.
  - Remove observations with NA from dataset.

- Complete Case = observation with data in **all** of the variables.

- Suitable for categorical and numerical variables

CCA example

Gender	Price	Make	Engine
Female	100	Ford	2000
	90	Ford	2000
Male	50	Kia	1500
Male	60	Kia	
Female	120	Nissan	3000
Female		BMW	4500
Male	200	BMW	4500

Gender	Price	Make	Engine
Female	100	Ford	2000
Male	50	Kia	1500
Female	120	Nissan	3000
Male	200	BMW	4500

CCA example

Gender	Price	Make	Engine
Female	100	Ford	2000
	90	Ford	2000
Male	50	Kia	1500
Male	60	Kia	
Female	120	Nissan	3000
Female		BMW	4500
Male	200	BMW	4500

Gender	Price	Make	Engine
Female	100	Ford	2000
Male	50	Kia	1500
Female	120	Nissan	3000
Male	200	BMW	4500

- Observations with missing values are removed

CCA Assumptions

- Data is missing completely at random

CCA Advantages

- Simple
- No data manipulation required
- Preserves the distribution of the variables

CCA Limitations

- It can exclude a **large fraction** of the dataset.
- When?
  - A variable has a big proportion of NA.
  - Different variables show NA in different observations.

CCA Limitations

- Removed observations could be predictive (if data is not missing at random)

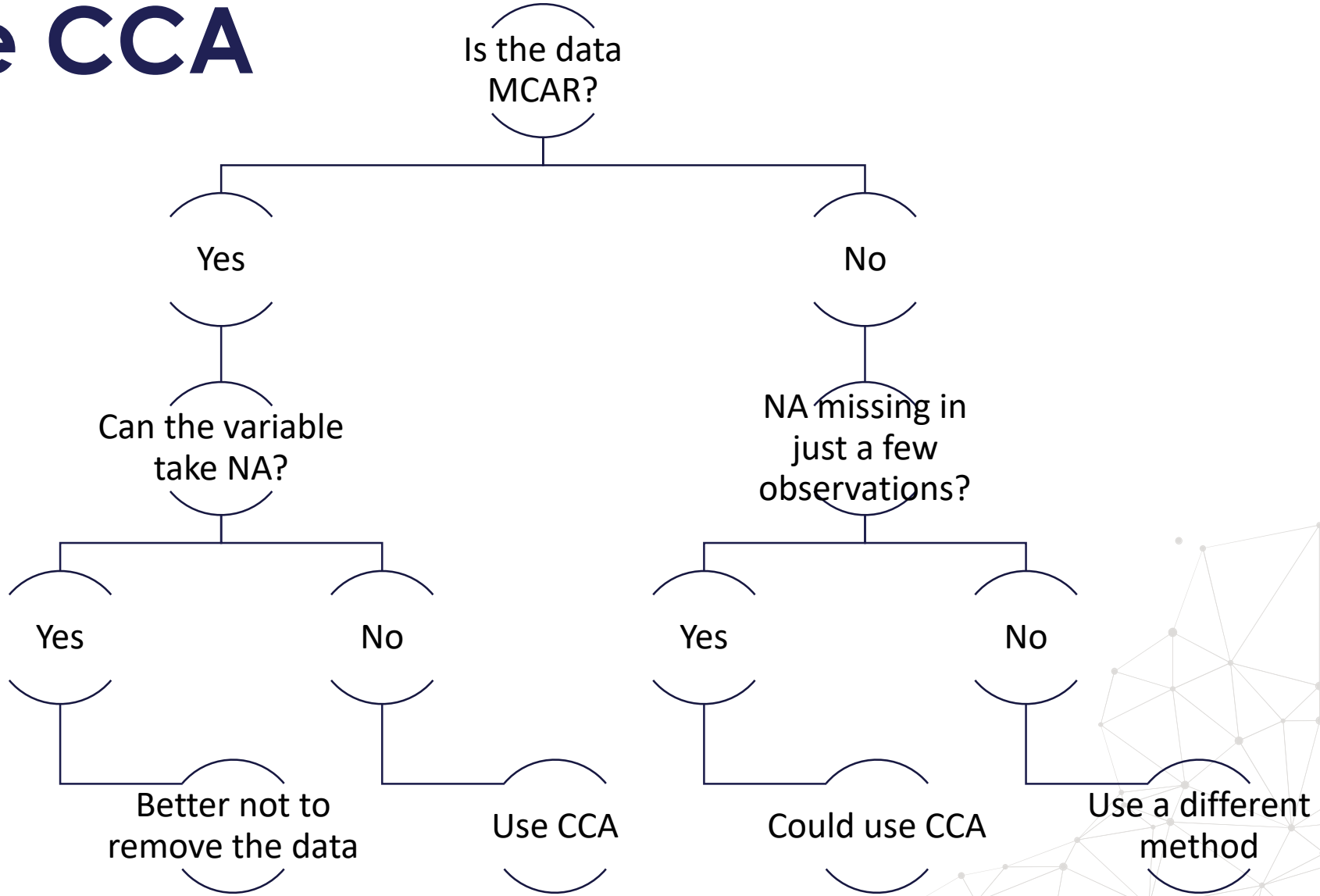
- CCA could lead to a biased dataset
  - The distribution of the remaining observations differ from the original distributions.

CCA Limitations

- When using our models in production, the model will not know how to handle missing data

When can I use CCA?

When to use CCA



Accompanying Jupyter Notebook

- Read the accompanying Jupyter Notebook

Train In Data

THANK YOU

www.trainindata.com