



# Complete Case Analysis

# CCA definition

- Complete-case analysis (CCA), also called "list-wise deletion" of cases, consists in **discarding** observations where values in **any** of the variables are missing.
- In Complete Case Analysis, we analyse only those observations for which there is information in **all** of the variables in the dataset.
- Suitable for categorical and numerical variables

# CCA definition

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 definition

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 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 original dataset (if missing data is abundant)
- Excluded observations could be informative for the analysis (if data is not missing at random)
- CCA will create a biased dataset if the complete cases differ from the original data (e.g., when missing information is in fact MAR or NMAR and not missing at random).
- When using our models in production, the model will not know how to handle missing data



# When to use CCA

- Data is missing completely at random
- No more than 5% of the total dataset contains missing data



# Accompanying Jupyter Notebook



- Read the accompanying Jupyter Notebook

# THANK YOU

[www.trainindata.com](http://www.trainindata.com)