

Some hints on how to proceed in your analysis

- First work with plots to illustrate connection in the data
 - Use bar charts (for instance to compare different subgroups for instance by country, gender, age categories etc)
 - Use histograms to illustrate the distribution of values of key variables
- Start by studying the association between one key dependent variable (Y) and one a key independent variable (X_1)
 - Avoid using subjective scales from the same survey as X and Y variables (if you have subjective assessments for instance study the association between a subjective assessment – such as job satisfaction or specific values – and more objective variables such as country, age, occupation, income, education,...)
 - First use a regression to study the association between Y and X
 - Use (bin-)scatter plots/reg plots if you want to illustrate connections
 - Interpret the magnitude of the obtained coefficient and its standard error and t-value
 - Can you think of a reason for an interesting interaction effects? – How do you test for this?
 - Can you think of a reason (hypothesis) that may lead to either non-linear associations – How do you test for this?
- Under what circumstances can you give this regression a causal interpretation?
 - Can control variables help and if so how? Which ones do you add to your regression?
 - Think about good and bad control variables and your reason to include them.
 - If you have longitudinal/panel data how could this help to move closer to evaluating causal effects?
 - Note: When you don't have experimental data be careful with causal statements and if you make them, state your identifying assumptions
- If you rather want to focus on prediction use ML methods
 - Compare 2-3 methods such as a Lasso regression and a random forest
 - Check the quality of your prediction
 - Plot the feature importance
 - Potentially use partial dependence plots