1. Write your questions of interest. (Convert these into Statistical Language or linear regression language)

2. Define the variables of interest and how to measure them.

3. Brainstorm which statistical tools you could use to answer your questions of interest. Depending on context, you may have to go back to step 2 and redefine your variables.

4. Design the study and how you're going to collect the data.

5. Collect data.

6. Enter & clean data.

7. Run some exploratory data analysis (graphs, basic numerical summaries) on variables of interest that can add insight into your questions of interest in 1.. EDA, as its name suggests, is exploratory. No hypothesis test / confidence intervals performed at this stage.

8. Based on 1. and 7. and perhaps based on automated search procedures, create an initial model. This initial model should be reasonable based on 1. and 7.. Do not expect the initial model to be perfect!

9. Carry out diagnostics, especially if the response needs a transformation.

10. Check other assumptions.

11. Consider dropping terms, adding interaction terms to your initial model.

12. Steps 9., 10., 11. need not be in that specific order, and may need to be repeated. Each step should be based on a reason, rather than trial and error. Make sure what you do is theoretically and contextually sound, and helps answer your questions of interests in 1.

13. Check for outliers in each model that is under considering. Outliers are usually interesting data points and should be investigated further.

14. You may have multiple models that are fine in terms of diagnostics. Consider which model best answers your questions of interest. Also compare them using model selection criteria.

15. Interpret and write your results.

In some data science settings, steps 2, 4, 5, 6 are already done as you harvest a database for data, and do not actually plan the study and collect the data in the traditional sense.