**Assignment 1 Visual Analytics Steps -to-do**

**STEP 1 – READING**

1. Read carefully the exercise question
2. Write down what it asks, which is the objective (problem statement) -> eventually come up with an hypothesis and how to test it

Tools: Dashbuilder, Databox

**STEP 2 – DATA COLLECTION**

We’ve already got datasets. Instructor provided them.

Otherwise -> Come up with a strategy to collect and aggregate data

**STEP 3 – DATA CLEANING**

1. Remove major errors, duplicates, outliers
2. Remove unwanted data points -> extract irrelevant observations
3. Bring structure to data -> fix typos, layout issues
4. Fill in major gaps -> insert values for important missing data

Tools: OpenRefine, Pandas, RStudio

+

STEP 3 – EXPLORATORY ANALYSIS

To identify trends, characteristics and patterns (i.e., correlation analysis)

**STEP 4 – DATA ANALIYSIS**

Data to carry out depends on the goal

* *Descriptive analysis* -> identifies what has already happened. It is a common first step that companies carry out before proceeding with deeper explorations
* *Diagnostic analysis* -> focuses on understanding why something has happened. It is literally the diagnosis of a problem, just as a doctor uses a patient’s symptoms to diagnose a disease
* *Predictive analysis* -> allows you to identify future trends based on historical data. In business, predictive analysis is commonly used to forecast future growth, for example. But it doesn’t stop there. Predictive analysis has grown increasingly sophisticated in recent years. The speedy evolution of machine learning allows organizations to make surprisingly accurate forecasts
* *Prescriptive analysis* -> allows you to make recommendations for the future. This is the final step in the analytics part of the process. It’s also the most complex. This is because it incorporates aspects of all the other analyses we’ve described

**STEP 5 – SHARE RESULTS**

You’ve finished carrying out your analyses. You have your insights. The final step of the data analytics process is to share these insights with the wider world (or at least with your organization’s stakeholders!) This is more complex than simply sharing the raw results of your work—it involves interpreting the outcomes, and presenting them in a manner that’s digestible for all types of audiences. Since you’ll often present information to decision-makers, it’s very important that the insights you present are 100% clear and unambiguous. For this reason, data analysts commonly use reports, dashboards, and interactive visualizations to support their findings.

Tools: Tableau, Seaborn, Matplotlib

**STEP 6 – EMBRACE FAILURES**

The last ‘step’ in the data analytics process is to embrace your failures. The path we’ve described above is more of an iterative process than a one-way street. Data analytics is inherently messy, and the process you follow will be different for every project. For instance, while cleaning data, you might spot patterns that spark a whole new set of questions. This could send you back to step one (to redefine your objective). Equally, an exploratory analysis might highlight a set of data points you’d never considered using before. Or maybe you find that the results of your core analyses are misleading or erroneous. This might be caused by mistakes in the data, or human error earlier in the process.

While these pitfalls can feel like failures, don’t be disheartened if they happen. Data analysis is inherently chaotic, and mistakes occur. What’s important is to hone your ability to spot and rectify errors. If data analytics was straightforward, it might be easier, but it certainly wouldn’t be as interesting. Use the steps we’ve outlined as a framework, stay open-minded, and be creative. If you lose your way, you can refer back to the process to keep yourself on track.