# DAT 500 – Roadmap – Assessment

As a data analyst, you will be expected to review data, perform analysis, and make some type of assessment that you report back to the interested party. This is the common cycle of a data analysis.

To discuss this in more detail, as a data analyst in practice, you will typically be given a project to perform an initial assessment and offer an opinion if there is something in the data that would support a business commitment to the project to further develop a data analytic project.

If a data analytic project is determined, then the full data set will be used and a business process will be developed to work with the analysis. Typically this would be in the form of some type of dashboard or analytic report that business analysts would use.

Note that this is a very generic, high-level view of data analytics, but it is a root that will be found in almost all implementations of a data analytic team.

This course is an entry-level course in the Master’s in Data Analytics program. The course introduces the tools or comparable versions of the tools you will likely use in practice. The course also simulates the data analytic implementation discussed above.

Note there are three discussions (Modules One, Four, and Six) that include a sample assessment, analysis, and the final report. Each of these samples should be used as a reference for the milestone that is assigned immediately following. Further, you will perform all the same steps in the final project as in the samples and associated milestones.

Considering this is an entry-level course, the concepts of the data analysis are being introduced as well. Over the course of the program, these concepts will be covered in much more detail.

Note, in the Sample Data Assessment in Module One, three aspects are covered that are important to understand and are the focus of the first milestone:

1. **First impressions** – first impressions are your initial thoughts based on an initial review of a data file and problem. In the first impressions, we should be able to understand what the business problem is and if the data can support researching it. That is about it. It could be thought of as a 30,000-foot view of the problem. This is a very high-level, rough order of magnitude review and thought. So a first impression can be something like reviewing a file of sales transactions and summing the total sales by region and making an assessment from this level of data. The first impression can happen in a very short period of time and should require minimal effort.
2. **Analysis** – this is the actual review of the data and performing an analysis that is specific to the business of the data being reviewed. This can be performing the descriptive statistics, including mean, max, min, percentiles, mode, standard deviations, etc. This is working with the data to form conclusions.
3. **Next Steps** – the next steps are the actions to continue working the problem based on the first impression and analysis results. If the data looks to be usable and the analysis shows that there is some path to pursue for further data analysis, then next steps would be making a recommendation from the initial review of data. Sometimes, next steps can be to end the project. Maybe there is nothing there in the findings. Alternative options are to begin a more specific project to perform the predictive analytics models.

# DAT 500 – Roadmap – Data Validation and Discovery

As we continue to move through our project and have completed the initial assessment of the data, we will perform data validation and data discovery tasks. What this means is that we have done an initial brief review of the data in Excel that led us to a decision to continue our project.

1. **Data Validation:** The first real step we take in the data analysis is to bring our data into a tool like R and perform some preliminary validation to make sure that the data we are working with is generally valid. This means running descriptive statistics on the data to learn more about its averages, minimums, maximums, and other statistical measures, such as standard deviation, variance, and percentiles. All of this information will give us a good idea of the data, and we can make some general assumptions. This is how we validate in this first review.
2. **Data Discovery:** From the descriptive statistics, we may notice values that are inconsistent with our expectation of the data. For example, maybe the average age in a data set is higher than we would expect. To further explore this, we perform data discovery tasks that are generated as a result of our analysis of the data. We may have values that we want to dig into further to answer more questions we have about the data. The questions we generate are typically “I wonder if…” or “What if…?” For example, if we see that the average retirement age is going down over time, we would want to dig in further to understand why and maybe ask “Are retirees retiring with fewer years of service?” This is something we can then study and dig into the data to analyze further.

# DAT 500 – Roadmap – Data Structures and Brief Report

As we wrap up our analysis and have completed our data validation and discovery steps, we produce quite a bit of output, which includes results of different discovery analysis as well as the final versions of the different data files we merged into our final analysis files. To wrap up the project and provide some conclusions, we need to store our data and results into data structures. In R, we can use vectors, matrices, and data frames as data structures to store our data and results. These structures then allow us to build reports to share our findings to management and decision makers regarding the data analysis project performed. Typically, the following is produced in a data analysis project:

1. **Findings:** The findings are the outputs of the different calculations and other statistics you ran during your data validation and discovery. These should be captured in your data structures designed for reporting.
2. **Informed Conclusion:** As an analyst, you are expected to make informed conclusions that come from working with the data and performing your analysis. What did you learn from the data? What is interesting or important about what you found? What conclusions were you able to come to about conditions in the data that you studied?
3. **Looking Ahead:** This is identifying the next steps. Typically the project stops here and a new project begins to develop predictive data models or dashboards based on the findings from this data analysis project. So looking ahead, we want to identify what could be. Maybe we need to add more data to gain more insight to a finding so that we can improve our business. What other next steps could be taken with the data analyzed and findings/informed conclusions discovered? This is how we look ahead as data analysts.