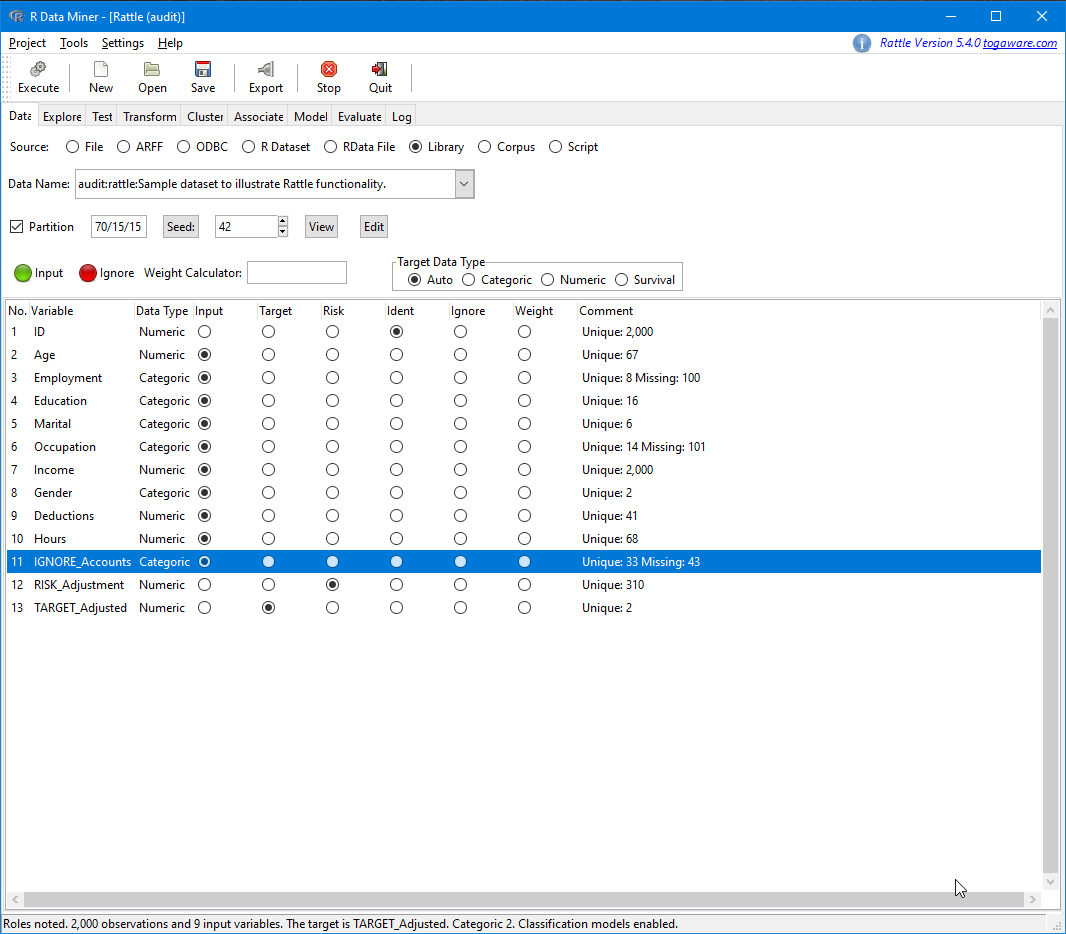
# Rattle Screenshot and Data Mining Process Comparison

# Christine Baxter, Data Analyst

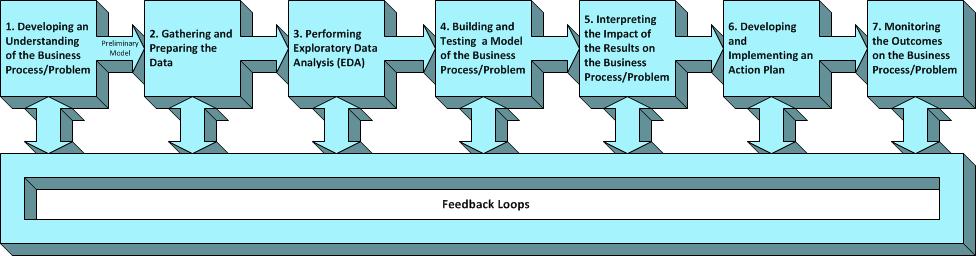
# Baxter’s, LLC

**Rattle Screenshot**



**Data Mining Process Comparison**

Process 1: Course Diagram *(source: Week 1 Lab description)*



Process 2: Rattle Tabs *(source: Rattle GUI)*

**Process Comparison**

What exactly is ‘Data Mining’ and why is it important to a business? What is ‘Rattle’ and how does it fit into data mining? These are questions on the minds of many executives, managers, and employees. The objective of this paper is to answer these questions and provide a better understanding of this concept and, most importantly, how it can help your business.

Data mining is often referred to by various names such as knowledge discovery, predictive analytics/modeling, and, most recently, business intelligence (OLAP.com, 2020) which encapsulates all the preceding terms. According to the authors of the Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management book (Berry, 2010), “Data mining is a business process for exploring large amounts of data to discover meaningful patterns and rules.”. With this definition in mind, there are two aspects to data mining…business context and technique. We will briefly discuss each topic, including a comparison/contrast between the business context data mining process and Rattle’s (or technique) data mining process.

At a high level, the data mining process covers the following: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The first three activities, designated as Items 1 and 2 (Process 1), fall under the business context aspect and are vital to producing a successful data mining project. The modeling activity, designated as Items 3 and 4 (Process 1), is where the technique (or Rattle) comes into play as this is about processing/exploring the data, ultimately, producing the best model. Evaluating (or interpreting) the data, designated as Item 5 (Process 1), really touches all activities and is where feedback loops are important among the team, both business process owners/users and analysts. Deployment, designated as Item 6 (Process 1), is just that…implementation of the action plan. Lastly and highly important, designated as Item 7 (Process 1), is monitoring outcomes on the business process or problem.

Rattle (or technique) is one of the most common data science tools (IBM, 2020) used in conjunction with business context to perform exploratory data analysis (EDA) and building/testing of a model that addresses the business process/problem. Exploratory data analysis, designated as Items 1 and 2 (Process 2) or Item 3 (Process 1), is a critical part of the data mining process. Using EDA, analysts/data scientists can manipulate data sources in the best manner, getting the answers needed to discover patterns, test a hypothesis, check assumptions, or spot anomalies (IBM, 2020). Building/testing of a model, designated as Items 3 through 7 (Process 2) or Item 4 (Process 1), is where a visual/graphical representation of a system is created (or built), communicating connections between data points and structures in support of a business need/problem. IBM simplifies this by the following statement, “A data model can be compared to a roadmap, an architect’s blueprint or any formal diagram that facilitates a deeper understanding of what is being designed.” (IBM, 2021). We now have a better understanding of what ‘Data Mining’ is, what ‘Rattle’ is, and how the two are related, now let us cover why is it important to your business.

With the electronic age, many businesses and applications used by customers and employees on a daily basis store and track data. This data can create enormous data sets and, through data mining, can hold valuable insights for marketing and advertising departments to social networking and medical outcomes. For a more personal example, when you are shopping on Amazon and scrolling through the ‘recommended for you’ section and finding all those items you did not know you needed…that is a direct result of a recommender system being used by Amazon that is analyzing the detailed purchase history and available demographic information from you and millions of other customers. For a business, the customer and the lifecycle of that customer relationship is of vital importance. No matter the business, there are five different types of customers with business relationships that evolve over time: prospects, responders, new customers, established customers, and former customers (Berry, 2010). Your business process will move your customers from one phase to the next and are important as they lead to making your customers more valuable over time.

Ultimately, data mining can be used to solve/predict a business problem or need by interpreting large amounts of data, extracting related data, examining meaningful patterns and rules among that data, and providing insights into your business and customer relationships. As we found out, there is a business process (steps depicted in Process 1) and complementary tools such as Rattle (steps depicted in Process 2) that are utilized in producing/implementing a successful data model.

**REFERENCES**

Berry, G.S.L.M. J. (2010). Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management. [devry]. Retrieved from <https://devry.vitalsource.com/#/books/9781118275603/>

OLAP.com. (2020). *Business intelligence vs predictive analytics*.  <https://olap.com/business-intelligence-versus-predictive-analytics/>

IBM. (2020). *What is exploratory data analysis?* <https://www.ibm.com/cloud/learn/exploratory-data-analysis>

IBM. (2021). *What is data modeling?*<https://www.ibm.com/cloud/learn/data-modeling>

Ledolter, J. (2021). Data Mining and Business Analytics with R. [devry]. Retrieved from https://devry.vitalsource.com/#/books/9781118593639/

R Core Team (2021). R: A language and environment for statistical computing. R citaFoundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Williams, G. J. (2011), Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery, Use R!, Springer. URL <http://www.amazon.com/gp/product/1441998896/ref=as_li_qf_sp_asin_tl?ie=UTF8&tag=togaware-20&linkCode=as2&camp=217145&creati>