

## Week 1 – Activity – What is Data Mining?

Student Name: Jonathan Y. Ben Avraham

Student Number: 2124450

### Data Mining Tools

Tool	Developer	Open Source?	About	Programming/GUI
RapidMiner	RapidMiner Inc	No	<ul style="list-style-type: none"><li>• datapreparation</li><li>• model creation</li><li>• data validation</li><li>• integrations with external systems.</li></ul>	Both
Oracle Data Miner	Oracle	No	<ul style="list-style-type: none"><li>• visualise data</li><li>• feature selection</li><li>• model builder</li></ul>	Both
Matillion	Matillion	No	<ul style="list-style-type: none"><li>• Data Extraction</li><li>• Data Transformation</li></ul>	GUI
Dataddo	Dataddo Inc	No	<ul style="list-style-type: none"><li>• Secure</li><li>• Datasource connectivity</li><li>• Data persistence</li><li>• Data preprocessing</li><li>• BI tool integration</li></ul>	GUI
Rivery	Rivery	No	<ul style="list-style-type: none"><li>• Datasource connectivity</li><li>• Data Transformation</li></ul>	GUI

### Applications of Data Mining and Informatics

Data Mining has many applications, however there are some industries where the application of data mining is imperative to the process of improving services or products. One such application of data mining is within healthcare organisations. Data Mining in the healthcare industry has led to improved patient care and faster diagnostic processes. Data Mining can be used in the healthcare industry for ‘characterizing patient behaviours and predicting office visits to identifying medical therapies for their illnesses.’ (Sharma, 2021)

Informatics applications in healthcare can be seen not just in quality of care but also quality of life for patients. Informatics is used for EHRs (Electronic Health Records), enabling healthcare providers from different hospitals/clinics access to historic patient records. (REZAEI-HACHESU et al., 2017)

### Histogram vs Bar Graph

Although they look similar at first glance histograms and bar graphs reflect different information about a dataset. A histogram visualises the distribution of data within a dataset, whereas a bar graph visualises data categorically. Figure 1 shows a histogram displaying the distribution of ages across our adult dataset. (Kohavi and Becker, 1996).

One other key difference is the bars in a bar chart often don't touch one another, whereas on a histogram, the bars will always touch one another. On a histogram the x-axis can only represent continuous numerical data, the example in the histogram is quantity of individuals with the same age. On a bar graph however, the x-axis can represent anything. (Cuemath, n.d.)

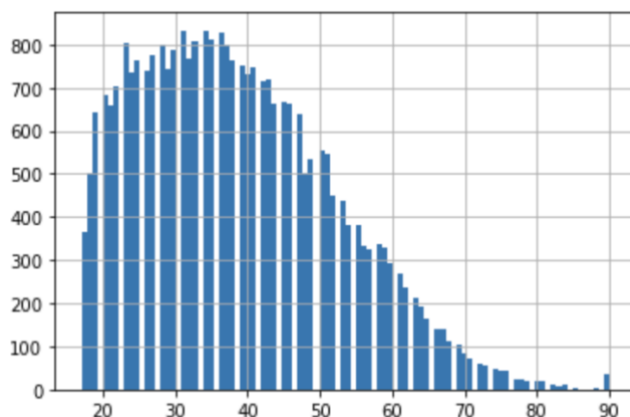


Figure 1

### Bibliography

RapidMiner, 2021. Get Started with RapidMiner [WWW Document]. RapidMiner. URL <https://rapidminer.com/get-started/> (accessed 11.3.21).

Sharma, R., 2021. 12 Most Useful Data Mining Applications of 2021 | upGrad blog [WWW Document]. URL [https://www.upgrad.com/blog/12-most-useful-data-mining-applications-of-2020/#6\\_Energy\\_Industry](https://www.upgrad.com/blog/12-most-useful-data-mining-applications-of-2020/#6_Energy_Industry) (accessed 11.7.21).

REZAEI-HACHESU, P., SAFDARI, R., GHAZISAEEDI, M., SAMAD-SOLTANI, T., 2017. The Applications of Health Informatics in Medical Tourism Industry of Iran. Iran J Public Health 46, 1147–1148.

Kohavi, R., Becker, B., 1996. UCI Machine Learning Repository: Adult Data Set [WWW Document]. URL <https://archive.ics.uci.edu/ml/datasets/Adult> (accessed 11.7.21).

Cuemath, n.d. What Is the Difference Between a Bar Graph and a Histogram [Solved] [WWW Document]. Cuemath. URL <https://www.cuemath.com/questions/what-is-the-difference-between-a-bar-graph-and-a-histogram/> (accessed 11.7.21).

About Dataddo [WWW Document], n.d. URL <https://docs.dataddo.com/about-dataddo-tool> (accessed 11.16.21).

Matillion | ETL Software [WWW Document], n.d. URL <https://info.gartnerdigitalmarkets.com/matillion-gdm-lp> (accessed 11.16.21).

Data Miner [WWW Document], n.d. URL <https://www.oracle.com/database/technologies/datawarehouse-bigdata/dataminer.html> (accessed 11.16.21).

Rivery | Data Pipeline Tool, Data Integration, ELT/ETL Platform [WWW Document], n.d. . Rivery. URL <https://rivery.io/> (accessed 11.16.21).