ANALYSIS OF ENVIRONMENTAL AWARENESS OF CONSUMERS WITH APPLICATION OF CLUSTER ANALYSIS

Šoše, Lara

University of Zagreb

Faculty of Economics and Business

Master in Managerial Informatics

ANALYSIS OF ENVIRONMENTAL AWARENESS OF CONSUMERS WITH APPLICATION OF CLUSTER ANALYSIS

Master thesis

Lara Šoše

Zagreb, August, 2020.

University of Zagreb

Faculty of Economics and Business

Master in Managerial Informatics

ANALYSIS OF ENVIRONMENTAL AWARENESS OF CONSUMERS WITH APPLICATION OF CLUSTER ANALYSIS

Master Thesis

Lara Šoše, 0067547945

Mentor: Mirjana Pejić Bach, PhD

Zagreb, August, 2020.

Statement on Academic Integrity

I hereby declare and confirm that the final thesis is the sole result of my own work based on my research and relies on the published literature, as shown in the listed notes and bibliography.

I declare that no part of the work has been written in an unauthorized manner, i.e., it is not transcribed from the non-cited work, and that no part of the work infringes any of the copyrights.

I also declare that no part of the work has been used for any other work in any other higher education, scientific or educational institution.

With this signature I confirm that in preparing this thesis I will comply fully with the Code of Ethics of the University of Zagreb.

Student signature: Šošo

Table of Contents

1. INTRODUCTION

1.1. Subject and Purpose

In the last decade, we have witnessed a massive rise in demand for goods and services globally, which has significantly contributed to the environmental damage and natural resource depletion. Society as a whole has become more mindful of the value of preserving the environment and one's own wellbeing, thus the importance of wiser selection of the products and services is increasingly emphasized. Markets have quickly responded to these circumstances and became filled with a variety of environmentally friendly products which are becoming increasingly popular among consumers.

The purpose of this master thesis was to inspect the extent to which consumers on the Croatian market are actually environmentally conscious, by asking them questions about their buying habits and general knowledge when it comes to environmentally responsible behavior and environmentally friendly products.

Finally, the cluster analysis on the obtained data will be conducted in order to place the consumers into groups and assign the observations to these groups.

1.2. Data and Methods

For writing the theoretical part of this master thesis, the vast majority of the literature used has been gathered from online sources, including books and various scientific articles. A collection of academic papers and scholarly publications on the subject were used as the source of information.

As for the empirical part, the research methodologies used in this master thesis are survey questionnaire and testing of existing conceptual solutions and tools for conducting cluster analysis of the collected data, in this case Weka software.

The survey questionnaire is composed of questions relating to the respondent's attitude towards environmentally friendly products and purchases. The obtained dataset is then used for conducting the cluster analysis in Weka software.

1.3. Structure and Content

The master thesis is divided into five chapters and starts with an introductory chapter containing the subject and purpose, sources and research methodology as well as the structure and content of the thesis.

The second chapter provides an insight into Knowledge Discovery in Databases (KDD), offering a detailed overview of the process, methods and practical implementations. This section provides a brief overview of the cluster analysis, which is one of the methods used in the knowledge discovery, and the method to be used in this master thesis for data processing.

The third chapter provides the reader with the theoretical aspect regarding the main research areas, describing the terminology “consumer behavior”, “environmental awareness”, “green marketing” and “green consumers”.

Furthermore, in the fourth part, a survey questionnaire results are described and analyzed in detail using summary tables. Also, the preprocessing of the gathered data is conducted, and the main attributes to be used in the cluster analysis are highlighted and defined. Finally, the results of the research are explained, i.e. the characteristics of the derived clusters are presented.

To conclude the master thesis, key findings of the research are outlined in chapter five. Following the conclusion, the literature and a list of figures and tables used for the preparation of this thesis are listed.

2. KNOWLEDGE DISCOVERY IN DATABASES

2.1. Introduction to Knowledge Discovery in Databases

Global automation and computerization, growing Internet usage, and skyrocketing popularity of smartphones have resulted in exponential growth in the amount of gathered data. Data is not recent; for thousands of years humans have been documenting the data as symbols, signs, and letters. What is really recent is the amount of data generated daily across the society, which is impossible to comprehend without the powerful and advanced IT. The vast majority of the gathered data is in its original state or data that yet needs to be turned into information and then into knowledge. Discovering trends, patterns and anomalies and transforming data and information into knowledge has been one of the major challenges of the digital age. All of this contributed to the rapid development of efficient data collecting and storage tools.

According to Matheus, Chan, & Piatetsky-Shapiro (1998), knowledge discovery can be defined as:

> "…nontrivial extraction of implicit, previously unknown, and potentially useful information from data. Given a set of facts (data) F, a language L, and some measure of certainty C, we define a *pattern* as a statement S in L that describes relationships among a subset FS of F with a certainty c, such that S is simpler (in some sense) than the enumeration of all facts in FS. A pattern that is interesting (according to a user-imposed interest measure) and certain enough (again according to the user's criteria) is called *knowledge*."[1]

The term "data mining" is oftentimes used in business, media, and the research to refer to the whole process of knowledge discovery. Hence, Han, Kamber & Pei offer the following definition the term:

> "Data mining is the process of discovering interesting patterns and knowledge from *large* amounts of data."[2]

[1] Matheus, C. J., Chan, & P.K., Piatetsky-Shapiro, G. (1998). Knowledge Discovery in Databases. IEEE Transactions on Knowledge and Data Engineering.

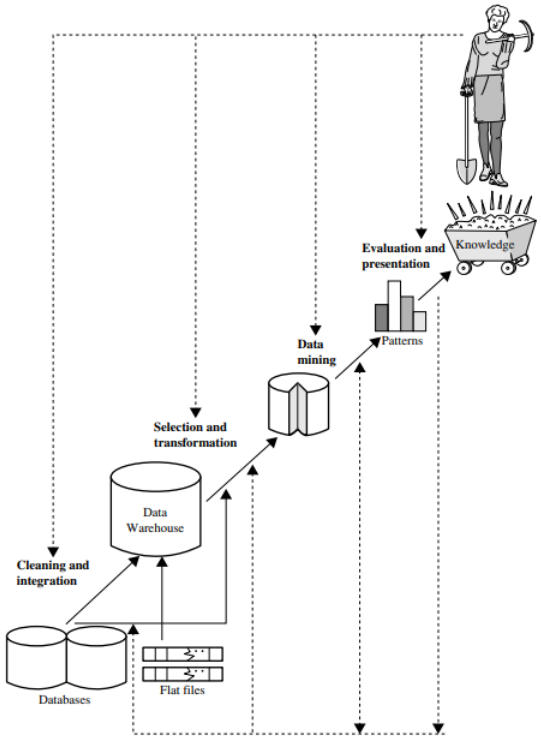
[2] Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques.

Sources of data may vary from warehouses and databases to internet and repositories. Also, data may be automatically transmitted into the system.

2.2. Process of Knowledge Discovery in Databases

Knowledge discovery in databases is the process of deriving the valuable information from the gathered data. It entails selection, preprocessing, transformation, data mining and interpretation/evaluation of the data, and finally determines what can be qualified as knowledge. Data mining is a critical step of knowledge discovery which incorporates the usage of algorithms to extract data patterns excluding the further stages of the KDD process. There is no general solution for knowledge discovery in databases to guarantee the success, although there are certain steps to help making the end result as successful as possible.

Figure 1: The process of Knowledge Discovery in Databases



Source: Han, J., Kamber, M. & Pei, J. (2012). *Data Mining as a Step in the Process of Knowledge Discovery*. Retrieved from: http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-

Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf (accessed 14 August 2020)

The figure above shows us a step-by-step process of knowledge discovery in databases. It is displayed as an iterative series of the outlined steps, meaning that there is an option to return to the previous step and to modify the existing data.[3] The steps include:

1. Data cleaning

2. Data integration

3. Data selection

4. Data transformation

5. Data mining

6. Pattern evaluation

7. Knowledge presentation[4]

2.2.1. Data Preprocessing

Cleaning, Integration, Selection & Transformation

As they are usually very big and come from multiple sources, modern databases are filled with data which tends to be noisy, missing and incomplete, i.e. poor-quality data. Using poor quality data will cause poor quality results of data mining. Data preprocessing is a quite an important step which helps us to improve the accuracy of data mining process and ensures the best possible results. As previously mentioned, there are a few different methods to be used for data preprocessing including: cleaning, integration, selection and transformation.

Data cleaning methods are used to "clean" the data and usually involve recognizing or removing deviations, filling in the missing data, smoothing the noise levels and solving errors. It is highly unlikely for the user to accept the results if he believes that the used data was not clean. Also, unclean data can cause confusion leading to inconsistent results during the mining process.

[3] Pejić Bach, M. (2005). *Data Mining in Banking*. Zbornik Ekonomskog fakulteta u Zagrebu, Vol. 3 No. 1.

[4] Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques.*

Some of the data cleaning methods for solving the problem of missing value include:

- **Ignoring the tuple** – not very effective;
- **Manually filling it in** – does not work with large datasets, time consuming;
- **Filling it in with a global constant** – simple but can be misleading for the data mining program;
- **Filling it in with an estimate of attribute middle value** – mean can be used for symmetric data and median for skewed data;
- **Using an estimate of attribute middle value for all samples belonging to the same class as the given tuple** – mean is a better choice for symmetric data and median for skewed data;
- **Using the most likely value** – can be assessed by regression, inference-based tools (Bayesian inference) or decision trees.

The last three methods are known to be biasing the data, but use of the last one is very common. It is important to stress that in some cases missing values do not mean there is an error in a dataset, but rather that the required data does not exist. Hence, there should always be an option to fill in the value as "non-existent" or something similar.

When we talk about the problem of noisy data, we must first define the term "noise". In a measured variable, namely, it refers to random error or variance. Some methods of tackling this issue include:

- **Binning methods** – clean the values by reviewing their surrounding or the values around them;
- **Outlier analysis** – detects the data with values different than expected, i.e. anomalies or outliers and levels them out;
- **Regression** – finds the best line to fit two or more variables allowing one to predict another.

Data integration entails combining the data from various databases. When applied properly, it helps minimizing and preventing redundancies and deviations, improving the overall quality and efficiency of the data mining process. Some major problems in data integration process include:

- **Entity identification problem** – metadata such as the name, meaning, data type etc. can help when combining the entities from multiple sources of data;
- **Redundancy and correlation analysis** – attributes which are extracted from some other attribute or set of attributes can be redundant;
- **Tuple duplication** – inaccurate data input or updating some but not all data instances can often result in discrepancies;
- **Data value conflict detection and resolution** – values of attributes originating from different sources may vary because of differences in displaying, scaling and encoding.

Data reduction or data selection methods may be employed in order to achieve a simplified dataset which is much smaller but maintains the quality of the original dataset. Challenging data mining usually takes a long time on massive volumes of data making the entire process exhausting or inefficient. Strategies for data reduction are:

- **Dimensionality reduction** – decreasing the number of random variables or attributes taken into account;
- **Numerosity reduction** – the original data amount is substituted with a reduced, alternative form;
- **Data compression** – transforming the data to attain it in a simplified or "compressed" form.

In **data transformation** the data is converted or combined into forms suitable for data mining. It can be done by the application of:

- **Smoothing** – cutting out the noise from data;
- **Attribute construction** – adding and creating new attributes to support the process of data mining;
- **Aggregation** – data is subject to summary or aggregation activities;
- **Normalization** – data is compressed to fit into more narrow range (e.g. between 0 and 1);
- **Discretization** – numeric values (e.g. *age*) are placed into interval labels (e.g. 15-24) or idea labels (e.g. *youth*);
- **Concept hierarchy generation for nominal data** – placing the attributes into a broader context (e.g. *chocolate* into a *food*).

2.2.2. Data Mining

Data mining is a key step in the process of knowledge discovery in databases. Hence, there will be more word on the data mining methods in the chapter 2.3.

2.2.3. Pattern Evaluation

A data mining system will produce thousands, perhaps even millions, of patterns or rules. In reality, only a small percentage of the produced patterns will be of interest to a user. To be interesting to a user, a pattern should be easy to understand, possibly useful, innovative and accurate on test data with some level of confidence. Furthermore, a pattern is interesting if it verifies a hypothesis a user wanted to test. Finally, an interesting pattern is what illustrates knowledge.

There are both subjective and objective pattern interestingness measures. Subjective measures are affected by user's personal perception. Such measures rate patterns as interesting if they are either surprising – based on user's view, or actionable – on which he can act upon. Also, they may be interesting if they support a hypothesis that the user wants to test. When it comes to objective measures of the interestingness of the patterns, they are based on the structure of the detected patterns and statistics. Some of them include rule support and confidence for association rules, as well as accuracy and coverage for classification rules. Whilst objective measurements contribute to detecting interesting patterns, they are often not adequate unless paired with subjective measurements that represent the demands and expectations of a specific user. An issue with objective measurements is that they sometimes automatically recognize certain pattern as interesting, when in reality these patterns are common sense and are uninteresting.

2.2.4. Knowledge Presentation

Presentation of knowledge involves the methods of visualization and representation used to explain the mined knowledge to users. Visualization methods are used to more efficiently and reliably interpret the data through graphical representation. They are commonly used for many tasks, such as business management, progress monitoring, and business analytics. A

major advantage of visualization methods is their ability to reveal links between the data which would otherwise not be as easily noticeable. To present data mining findings in a dynamic and functional manner, such that the discovered knowledge can be easily understood and used by humans, the system needs to adopt visualizations, expressive representations of knowledge and user-friendly interfaces.

2.3. Methods of Knowledge Discovery in Databases

There is a variety of methods and algorithms used in KDD and this chapter will cover only a portion of them. According to Pejić Bach, methods can generally be categorized into three groups including: methods of discovering groups, methods of predicting events and methods of predicting values.

2.3.1. Methods of Discovering Groups

These methods are looking for data patterns with no previous understanding of their type. An example of such methods are association rules and cluster analysis.

Association Rules

Association rules search for certain rules in datasets that tightly associate different attribute values. They enable predicting any attribute, not only the attribute class, allowing them to predict combinations of attributes as well. Because of the great number of rules which may be extracted from a very small set of data, relevance is limited to those with wide number of instances, and where they have high accuracy.

The number of instances for which the correct predictions are made is association rule coverage or support;

support $(X \Rightarrow Y) = P(X \cup Y)$.

The number of instances for which the correct predictions are made, represented as a fraction of all instances to which it refers, is accuracy or confidence;

confidence $(X \Rightarrow Y) = P(X|Y)$.

It is typical to determine minimum support and confidence values of the rules, and to pursue strictly those with values at least at specified minimums.[5]

Cluster Analysis

Cluster analysis is the method in which sets of data are divided into subsets or clusters. Objects in a cluster are different from objects in other clusters, and similar to those in the same cluster. It is performed by using clustering algorithms, and helps to find groups within the data which were not previously known.

Han, Kamber, & Pei (2012) offer the overview of basic methods of clustering, dividing them into four distinct categories including:

1. **Partitioning methods** – require for each entity to belong to only one class, i.e. clusters are mutually exclusive;
2. **Hierarchical methods** – builds a hierarchical breakdown of the objects in dataset, i.e. multiple level clustering;
3. **Density-based methods** – clusters are concentrated objects, separated by low-concentration regions;
4. **Grid-based methods** – clustering is conducted on the grid structure.

The key steps following the clustering activities include identifying the pattern of clustering, estimating the number of clusters and reviewing the quality of clustering.[6]

2.3.2. Methods of Predicting Events

Examples of methods for predicting events to be introduced in this chapter are neural networks and decision trees.

Neural Networks

After scientists came to the conclusion that the human brain is the best tool for recognition and modeling, they tried to imitate it for decades. This led to a huge interest in artificial neural networks. Thanks to great advances in neuroscience, modern algorithms can mimic parts of

[5] Witten, I.H., & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*.

[6] Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques.*

the human brain very precisely and perform tasks outside the domain of modeling. A better understanding of how various areas of the brain communicate, restructure, learn and function has also been established. Neural networks achieved tremendous potential in tackling image processing and compression issues. As more and more data are now being gathered in form of pictures, neural networks that can analyze images are becoming particularly relevant.[7]

Decision Trees

Decision trees are one of the most common classification methods, the product of which can be written in tree shape. While the decision tree in data mining is a statistical approach that can be used to explain both classification and regression models, it refers to a network of decisions and their outcomes. The user employs the decision trees to evaluate the strategy with the best chance of accomplishing the task.[8]

There is a variety of decision tree algorithms which have been developed over the years, and some of the most well-known include:

- **CHAID (CHi-squared Automatic Interaction Detector)** – first decision tree algorithm developed in the 1980s by Gordon V Kass, easy to understand and manage;
- **CART (Classification and Regression Tree)** – based on the format of a dependent variable, it generates either regression or classification trees;
- **ID3 (Iterative Dichotomiser 3)** – at each step, the iterative algorithm splits (dichotomizes) attributes into two or more classes;
- **C4.5** – an extension of ID3, obtained decision trees can be used for classification;
- **C5.0/Sec 5** – an extension of C4.5 algorithm with better effectiveness, speed and memory;
- **Hunt's Algorithm** – creates a decision tree with the use of divides and conquers or top-down approach.[9]

[7] Cios, K.J., Pedrycz, W., Swiniarski, R.W., & Kurgan, L.A. (2007). *Data Mining: A Knowledge Discovery Approach*. Springer.

[8] Rokach, L., Maimon, O. (2014). *Data Mining with Decision Trees: Theory and Applications*.

[9] Patel, B.R., Rana, K.K. (2014). *A Survey on Decision Tree Algorithm for Classification*. IJEDR Volume 2, Issue 1.

2.3.3. *Methods of Predicting Values*

Some of the methods of predicting values include the previously mentioned neural networks, and linear regression.

Linear Regression

Linear regression is a machine learning algorithm which produces the output with steady and continuous slope, and is used not to group the values into classes, but to estimate them across a continuous range. Linear regression is one of the main methods in statistics and one to be considered if the class or outcome, and all of the attributes are in a numeric format.

The intention is to describe the class as a linear attribute combination for which weights are predefined:

$$x = w_1 + w_1 a_1 + w_2 a_2 + \ldots + w_k a_k$$

…where x is the class; a_1, a_2,…, a_k are the attribute values; and w_0, w_1,…, w_k are weights.[10]

2.4. Applications of Knowledge Discovery in Databases

Data mining applications are countless in a wide variety of areas, and some of those will be mentioned briefly in this chapter to reflect the scope of what has been achieved.

2.4.1. *Retail and Telecommunication*

Almost every key retail chain nowadays has websites which allow consumers to make online purchases. Many of them operate only through online shops, and have no physical stores at all. Since it generates massive volumes of data on sales and consumer buying behavior, it goes hand with hand with data mining, especially given the fact that the amounts of data

[10] Witten, I.H., & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*.

gathered continue to grow exponentially. Data mining can help retailing to recognize consumer buying patterns, create greater customer engagement and loyalty, develop more efficient transport and logistics strategies, reduce costs etc.

A telecommunications sector has grown steadily from providing solely telephone services to supplying numerous other complex networking services, such as Web connection, smartphones etc. The incorporation of different means of communication and computing increased the demand for data mining to explain and predict market mechanisms, recognize trends in telecommunication, detect illegal practices, enable better usage of resources and enhance the overall quality of service.

2.4.2. Science

Throughout the past, certain methods of interpreting scientific data appeared to tackle fairly limited and homogeneous sets of data, and were usually analyzed using the statistical techniques. Nowadays, it is possible to gather scientific data at much faster and cheaper rates which led to the collection of massive amounts of highly dimensional data, heterogeneous data and streaming data. Data is obtained in different fields of science such as biology, meteorology, and physics, utilizing various sophisticated technologies.

Moreover, data mining has gained attention in the social sciences because of the vast quantities of human interaction data gathered every day in the forms of websites, articles, forums, comments, messages, ads, social media, and other.

2.4.3. Cyber Attacks Prevention

The rapid expansion of the Internet and the widespread availability of tools and tactics for attacking networks made the cyber security a crucial element of networked systems. The primary tasks of cyber security systems are to detect the malicious activity, attempt to stop it, and to report it. These systems usually use either anomaly or signature–based detection. Anomaly–based detection generates profiles or standard network activity models that are used to identify new behaviors that differ substantially from the profiles. New behaviors are sometimes real intrusions, and sometimes only new trends to apply to the profiles. On the

other hand, signature–based detection uses the signatures, or patterns previously defined and configured by software engineers. The system will report the irregularity and take further measures when signature matches are discovered in network traffic.

2.4.4. Recommendation Systems

Recommendation systems support the consumers by providing them with personalized recommendations of different kinds of products and services such as books, movies, shoes and other. These systems are based on either collaborative approach or content-based approach. There is also a hybrid approach which is a combination of the two previously mentioned approaches.

The collaborative approach takes a consumer's social climate into consideration, recommending the products based on the views of others with the similar preferences and attitudes. A wide variety of techniques such as machine learning, statistics, and data mining is used to identify the similarities between consumers. On the other hand, content-based approach suggests the products which are similar to those searched before, based on product characteristics and textual descriptions of it.

2.4.5. Healthcare

Knowledge discovery in databases has been successfully employed in many healthcare institutions regardless the fact that it is not as well-established in healthcare as in other industries. It has the capacity to recognize relationships and trends that are not easily noticeable, and thus to find people at risk. A number of healthcare companies today use innovative marketing strategies to position themselves in an increasingly competitive environment. With the help of data mining, companies can use the gathered data to estimate which groups are more likely to utilize their services, as well as the most efficient marketing strategies for reaching them.

According to DeGruy (2000), one of the healthcare maintenance organizations (HMOs) utilized the knowledge discovery in databases and used past records from their patients to

assess which of the insured might be at risk with a certain illness. This is beneficial for both of the parties since the care expenses of the company are reduced, and the insured is safer.[11]

2.4.6. Other

Image Screening

Scientists have attempted to use satellite technology from its early stages to monitor oil spills from aerial photos in order to provide early notice of natural hazards and prevent unregulated dumping. When carried out manually, the process of finding oil spills from the images is very costly and demands the highly qualified experts. With the help of data mining, the program can be trained to detect the potential leaks, and to forward the case to the expert which will then decide whether it is an actual leak or a false report.

Load Estimation

In the power sector, it is essential to estimate the future power demand as soon as possible. If reliable forecasts can be made, service providers may make substantial improvements in areas such as planning the servicing, resource control and arranging the reserves. An automated assistant is able to generate two days in advance hourly forecasts. The system achieves the same results as a human expert, except that it builds a daily forecast in just a few seconds instead of hours.

Web Mining

Data mining is becoming increasingly popular among search engine corporations which use it to better evaluate websites and to find an objective measurement of popularity for each of them. They analyze the data on clicks to obtain the knowledge which can be used to enhance the further search. For example, digital bookshops use data mining to generate personalized book recommendations, and the same applies for streaming websites.[12]

[11] DeGruy, K.B. (2000). *Healthcare Applications of Knowledge Discovery in Databases*. Journal of Healthcare Information Management.

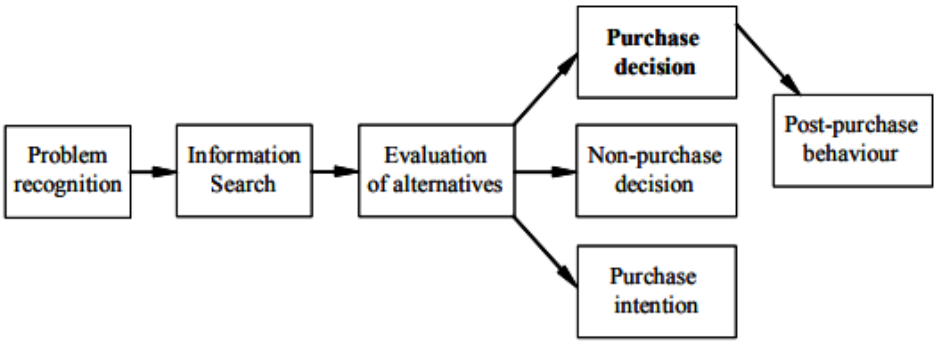
[12] Witten, I.H., & Frank, E. (2005). Data Mining: Practical Machine Learning Tools and Techniques.

3. ENVIRONMENTAL AWARENESS OF CONSUMERS

3.1. Consumer and Consumer Behavior

The action of buying an item is much more than what it seems at first, with an entire process of buying decision behind it. The buying process includes the five stages including: recognizing the need, finding the information, assessing the alternatives, purchase, and post-purchase behavior.

Figure 2: Stages of Purchase Decision Process



Source: Munthiu, M.C. (2009). *The Buying Decision Process and Types of Buying Decision Behaviour*. Retrieved from: https://pdfs.semanticscholar.org/f2d3/36232b0b67e046f38ac01c70df0f9c0dbd07.pdf (accessed 26 August 2020)

In the figure 2, all of the stages of the purchase decision process are shown starting with a problem or need recognition. It occurs when the consumer notices a significant gap between his current satisfaction level and the level of satisfaction to which he aspires, when it comes to a particular want. Also, it can be stimulated by both internal and external factors, where the internal include basic needs such as thirst, and external come from the consumer's environment. In the second and third steps, consumer searches for information about the offer on the market and assesses the product alternatives. Finally, he is ready to decide whether to make the buying decision or not. The final step is post-purchase behavior in which the consumer evaluates to what degree the decision was successful or unsuccessful, i.e. difference between his expectations and his personal impression regarding the purchase.

Wise corporations are aware of this process, and they are committed to investigating it. The primary objective of the marketing research is to get a clear idea of each of the stages, including the behavior prior to and after making the purchase decision. These stages vary significantly depending on what is being bought, and should be recognized and studied. More demanding buying decisions most often require more thoughtful consideration, as well as more participants. Hence, we distinguish between four different types of consumer behavior being:

1. Complex buying behavior
2. Dissonance-reducing buying behavior
3. Habitual buying behavior
4. Variety-seeking buying behavior

Complex buying behavior takes place when the buyer does not know enough about the product category and he buys something very expensive (e.g. buying a car). Also, in this case there are significant differences between products he can choose from. Therefore, he needs to go through a learning process, and companies must provide a large amount of data to help him in making a decision.

When it comes to dissonance-reducing purchasing behavior, it happens with expensive and not as regular purchases, where alternatives or options are limited to only few. Because the distinctions between the choices are not significant, buyers tend to make purchase decision very fast due to a reasonable price or convenient timing. Buyers often encounter confusion or dissonance after these buying decisions, therefore businesses need to provide them with the assistance to make them feel better regarding their purchases.

Habitual buying behavior occurs with more frequent purchases, and there are many options to choose from. Some of these purchases are for example milk or toilet paper. Most of the consumers tend to buy these products regularly and habitually, and are rarely highly involved in the buying process. The reason behind this is that consumers are passively exposed to information about these products through various commercials.

Variety-seeking behavior happens when buyers often switch between the brands just to experience the change. There are many marketing strategies companies use to tackle this

issue, and to try to promote habitual purchases. Some of the tactics they use are regular reminder advertising, occupying the space on the shelves and keeping them in a full stock.[13]

3.2. Environmental Awareness

Controlling and improving the overall degree of environmental awareness throughout the society has been one of the socio-cultural priorities over the last few decades. It has reached a greater degree of social and political solidarity than ever. Estimating environmental awareness on the empirical basis is becoming more and more relevant for researchers employed in various fields, as only the objects that can be measured can really be controlled and improved. In different fields such as marketing, sociology, psychology, environmental and business studies and political sciences, there have been several attempts to define the term environmental awareness over the past couple of decades. Various problems occur in various scientific areas, and it is essential to be mindful of and take into consideration the multiple perspectives. It is possible to mitigate all the negative impacts and make contributions to the standard of the findings obtained and the models produced, by taking into consideration all the relevant problems.

According to Ham, Horvat and Mrčela (2015), most of the people instinctively understand the idea of environmental awareness, but it is fair to suggest that there is no widely established definition, nor terminology. Generally defined, environmental awareness is the attitude towards environmental impacts of human activity. It is assumed that consumer practices his environmental awareness through environmentally driven behavior. Hence, we can distinguish between two types of environmentally driven behavior including environmental awareness and environmental responsibility. Environmental awareness implies the positive attitude towards the environment, and environmental responsibility the concrete positive actions and behavior. This pro-environment behavior is not only related to making conscious purchases, but also to practicing it in other activities such as reducing the daily energy consumption and separating waste.[14]

When it comes to measuring the environmental awareness, many of the research attempts

[13] Munthiu, M.C. (2009). *The Buying Decision Process and Types of Buying Decision Behavior*.

[14] Ham, M., Mrčela, D., Horvat, M. (2015). *Insights for Measuring Environmental Awareness.*

ended in contradictory results. One of the problems researchers face is that the respondents may understand the concept of environmental awareness in a different way and therefore provide different answers. Carlson argues (as cited in Ham, Horvat and Mrčela, 2015), that the large number of factors as well as conflicting research results contributed to the belief that researchers do not always have similar ideas of the concept of environmental awareness.

It is suggested that substantial change has been made over the last few years, as the usage of the expression "environmental awareness" has progressed from being only a buzzword, to actually being included in concrete policies at various levels of society. There is a great deal of room for improvement and it is necessary to determine the exact parameters and to carry out comprehensive research, but a step forward is undoubtedly noticeable.

3.3. Green Marketing

3.3.1. The Concept of Green Marketing

According to Peattie & Charter (2003), the early start of the concept of green marketing is linked to the wave of environmental movement in the 1970s. It was mainly linked to the industries that had the most significant environmental impacts, and to those that had the potential to build new technologies that would help in minimizing those negative impacts. The inclusion of environmental concerns into marketing theory is sometimes described as an enhancement of the social marketing concept. There is, however, a difference between what the idea of green marketing portrayed in the past and what it represents today in a modern society.

Peattie & Charter (2003), define the concept of green marketing as:

> "The holistic management process responsible for identifying, anticipating and satisfying the needs of customers and society, in a profitable and sustainable way."[15]

The concept of sustainability in production and consumption implies enjoying the material lifestyle to the extent at which it does not harm the lifestyle of future generations. This concept consists of two tasks; the first one is the consumption of natural resources at a level at which they could still be resupplied by human impacts or environmental processes, and the

[15] Peattie, K., Charter, M. (2003). The Marketing Book, p.p. 726.

second one is waste generation and pollution only to the extent to which ecosystems are stable and able to absorb it.

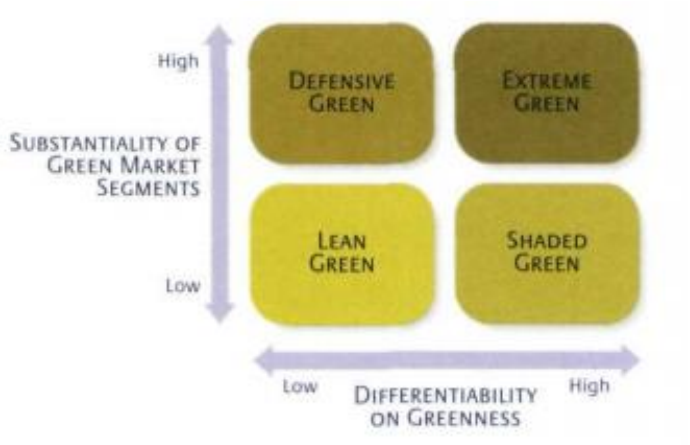
While the concept of green marketing is theoretically easy, in practice it is incredibly challenging to be enforced. It is essentially because the traditional management theory is based on economic philosophies that wrongly perceive natural capital as unlimited, worthless, or free. In the past, greening aimed to discover the efficiencies that boosted the bottom-line revenue of the company. Today, on the other hand, it is widely recognized as an opportunity for growth, improvement, and increasing the top-line revenue. Proactive companies are coming up with new clean technologies, innovative strategies and fresh solutions that catch widespread attention, as well as they attract customers and achieve a sustainable competitive advantage.[16] Companies are facing a growing number of consumers with a strong concern for the environment, and are accordingly aiming to identify the so-called "green consumers", and understand what motivates them to make environmentally friendly purchases. Rather than just trying to enhance the environmental and social impacts of conventional products and technologies, a variety of new and greener products and technologies should be built to create a substantially greener economy. Developing a fully green product at a competitive price, while delivering reasonably competitive standards of service and quality is fairly a daunting challenge. The concept of sustainability is often challenging to measure, hard to implement, and subject to different interpretations, yet, it has become clear that anything that is not sustainable cannot be sustained.

3.3.2. The Green Marketing Strategies

According to the MIT Sloan Review on the green marketing strategies, companies need to consider a few factors before pursuing a certain green marketing strategy. Before selecting one of the strategies in the matrix, they should first assess the potential scope of the green market in their field, and also their capability to distinguish their green brands from others on the market. In the figure 3, the green marketing strategies matrix is represented, showing four major strategies including: lean green, defensive green, shaded green and extreme green.

Figure 3: Green Marketing Strategies Matrix

[16] Ottoman, J.A. (2010). *The New Rules of Green Marketing*. Strategies, Tools, and Inspiration for Sustainable Branding. p.p. 17.



Source: 19. Ginsberg, J.M., Bloom, P.N. (2004). Choosing the Right Green Marketing Strategy. MIT Sloan Review. Available at: https://sloanreview.mit.edu/article/choosing-the-right-greenmarketing-strategy/ (accessed 30 August 2020)

The lean green strategy is used by companies that are mainly motivated in cutting costs and efficiency gains by utilizing eco-friendly practices. They usually do not promote their green initiatives, rather just enjoy reduced cost competitive advantage. They want to be in line with rules and regulations, but do not really see the green market making them significant profits. In fact, they frequently refuse to advertise their environmental responsibility, to avoid the risk of rising the expected standard to a level they may be unable to keep up with.

Furthermore, companies typically employ defensive green strategy as a reaction to a disaster or crisis, a reply to the activities of rivals, or as a precaution. They strive to enhance brand reputation and minimize harm, understanding the importance and profitability of the green markets. Usually, those companies do not have the potential to distinguish their green offerings from rivals, but put a huge amount of work into promoting and publishing their sustainable projects. If they are threatened by protestors, legislators or rivals, they will certainly protect their green impact with public relations and marketing initiatives.

Shaded green strategy is practiced by companies that view sustainability as a chance to build inventive products and services, to ultimately help them gain competitive advantage. They make significant investments in sustainable processes across the system, as they have the true ability to distinguish themselves on sustainability. However, they are mainly promoting the concrete products and services offered to the client.

Finally, the extreme green strategy is used in companies where environmental concerns are deeply incorporated into every aspect of the business, from their practices to processes and products. They mostly target niche markets and use specialty shops and boutiques to offer their products.[17]

3.3.3. Greenwashing

Many companies put a huge marketing effort to promote themselves as environmentally conscious, when in fact this is far from reality. Misleading the consumers regarding their impact on the environment, and falsely portraying themselves as green is called "greenwashing".[18] A greenwashing company, in general, has weak environmental performance and good communication regarding its environmental performance. In the recent years, markets became flooded with greenwashing companies, having a detrimental impact on customer and investor trust in green products. As a result, more and more companies are encouraged to participate in wasteful activities, producing negative environmental impacts and therefore detrimental effects on social welfare. For stakeholders and managers trying to enforce strategies or take steps to minimize the occurrence of greenwashing, it is crucial to examine the factors that primarily cause greenwashing and assess how effectively mitigate them.

3.4. Green Consumers

Rising market demand for goods considered more sustainable has, over the past few decades, resulted in increased availability of eco-friendly products and brands. A variety of different factors has been recognized as essential in encouraging green customers to purchase green products. Consumers, especially young, are progressively claiming that they want brands which support sustainability. However, the troubling dilemma lies at the core of green marketing; customers who express optimistic perceptions regarding environmentally friendly products are rarely purchasing them in reality.[19] According to a recent research, 65% of the respondents claimed they are willing to purchase sustainable products, but only 26% actually

[17] Ginsberg, J.M., Bloom, P.N. (2004). *Choosing the Right Green Marketing Strategy*. MIT Sloan Review.

[18] Delmas, M.A., Cuerel Burbano, V. (2011). *The Drivers of Greenwashing.* University of California, Berkeley.

[19] White, K., Hardisty, D.J., & Habib, R. (2019). *The Elusive Green Consumer.* Harvard Business Review.

reached in their wallets. This is a so-called "intention-action gap" that needs to be shrunk in order to facilitate sustainability objectives, but to also protect the environment.

White, Hardisty & Habib (2019) outlined five steps that companies should take to motivate consumers to make more sustainable purchasing decisions, and start behaving in a sustainable manner: create positive habits, choose between rational and emotional, utilize social influence, use the domino effect, and prioritize experiences over possession.

Creating Positive Habits

The solution to the growth of healthy buyer behavior is often to first eliminate poor habits, and then to promote new habits. Habits are stimulated by indicators that are present in recognizable settings. Many habits are a matter of the routine for customers, including what they purchase and consume, how they travel, and how they dispose waste and packaging. Making the sustainable solution a default option shown to be the quickest and perhaps most successful solution in creating positive habits among consumers. For example, setting the eco packaging as a default option, whereas the regular one must be explicitly demanded. Some of the techniques that can help companies to create positive habits are providing feedback, sending reminders, and reward systems.

Choose Between Rational and Emotional

The way businesses engage with customers seems to have a huge impact on the adoption of good habits. Companies also have to make a choice between rational thinking and emotional triggers when they are planning to create a product or a project. When positive feelings can be gained from participating in a certain behavior, consumers are more likely to take part in it. Although it has been discovered that it is very beneficial to trigger feelings of pride and optimism in sustainable campaigns, many of them prioritize upsetting messages. When employing the rational thinking in sustainable campaigns, companies illustrate the effect their operations have on the ecosystem. Also, the responses focused on local context have seen to be exceptionally effective.

Utilize Social Influence

One of the most successful ways of promoting environmental behavior is by utilizing the social influence. It was reported that people have a great need to blend in and to adapt to the behavior and attitude of those in their surroundings. For example, when online customers

were informed that others chose environmentally friendly products over the conventional ones, there was an increase to 65% in making a first green order. Nevertheless, campaigns based on social influence may sometimes cause a backlash. In case that only a small number of people take part in sustainable practices, the acceptance may be limited as the practice is not perceived as acceptable.

Use the Domino Effect

Persuading people to follow sustainable behavior can result in a domino effect, motivating them to improve other aspects of their behavior, bringing in more positive changes and improvements. Again, a sustainability initiative can contribute to less sustainable customer behavior, i.e. an adverse response can take place. An example of such situation is when consumers use more of the product as it is eco-friendly, and use more plastic as there is an option to recycle it.

Prioritize Experiences Over Possession

Many businesses have contributed to making consumers increasingly receptive towards renewable options, while also working to change their attitudes. Frequently their offerings are in the form of experiences rather than only material goods and services, which contributes to stronger connection, makes both parties satisfied, and foster more meaningful associations. This practice further entails a substantially lower carbon footprint. Similarly, the practice of sharing is very successful. It is noticeable it the fact that in recent years, some of the biggest success stories have included companies that offer access to established products and services, instead of producing and marketing new offerings. Almost anything, from vehicles to apparel and footwear, and even homes, can be leased nowadays.

4. ANALYSIS OF ENVIRONMENTAL AWARENESS OF CONSUMERS WITH APPLICATION OF CLUSTER ANALYSIS

4.1. Research Methodology

Research methodologies used in this master thesis are survey questionnaire and testing of existing conceptual solutions and tools, in this case Weka software, for conducting the cluster analysis of the gathered data. The survey questionnaire used in this thesis was created in Google Forms app and consists of 27 questions. Questions 1 to 6 refer to statistical data about the respondents including gender, age, level and field of attained education and occupation, and questions 7 to 27 refer to their attitude towards environmentally-friendly products. All questions in the survey questionnaire had closed structure, i.e. the respondent could only select between answers in previously specified categories. For easier implementation of the cluster analysis, there was a possibility to select the only one, most accurate response. Furthermore, all questions were marked as "mandatory" in order to have a complete dataset. The survey was shared through WhatsApp Messenger, E-mail as well as social media platforms such as Instagram and Facebook. Finally, 180 responses were gathered, i.e. the survey sample size is 180 people.

4.1.1. Dataset Description

This chapter will give a detailed description of the gathered dataset. Questions about the respondents' attitude towards environmentally friendly products will be presented in tables and the thorough analysis will be carried out.

Response summary to the question "Gender"

When it comes to the gender structure of the sample, out of the total of 180 respondents, 98 identified as female and 82 identified as male. We can conclude that there are slightly more female respondents with 54.4%, as compared to 45.6% of the male respondents.

Response summary to the question "Age group"

Due to the fact that the survey questionnaire was mostly shared among students, more than half of respondents, precisely 53.3%, are in the age group between 18 and 24 years old. The

least number of respondents (3) is in the age group 65 plus where people are less likely to be using social media and Internet in general. Also, there are 13 respondents or 7.2% who are under 18, 25 or 13.9% between 25 and 34, 12 or 6.7% between 35 and 49 and 31 or 17.2% between 50 and 64 years old.

Response summary to questions "Attained level of education" and "Field of education"

There are 5 distinct levels of education including primary and secondary education, undergraduate and graduate university and professional studies and postgraduate university studies. Most of the respondents (41.7%) have finished secondary education and least of them have finished postgraduate university studies (1.7%).

Educational fields include natural, technical, biotechnical, social and humanistic sciences, biomedicine and health, art and other. The highest percentage of them selected "social" due to the fact that the survey was shared mostly among economics students. Also, 20% selected "technical" and 19.4% "other". The smallest proportion of the respondents are studying/studied in the biotechnical and humanistic fields.

Response summary to the question "Amount of monthly net income expressed in HRK"

Most of the respondents have the monthly net income between 3.501,00 and 7.500,00 HRK. Also, 33.3% make up to 3.500,00 HRK of monthly net income. Only 5 or 2.8% earn over 25.000,00 HRK per month.

Table 1: Response summary to the question "Occupation"

Occupation	Number of respondents
Agricultural, hunting, forestry and fishing workers	1
Military occupations	1
Retired	1
Unemployed	1
Machine and vehicle handlers, product assemblers	2
Simple occupations	2
Members of government bodies, directors	4
Scientists and experts	6

Crafts and individual production	7
Service and retail occupations	12
Entrepreneur	18
Office clerks	24
Engineers and technicians	25
Students	76
Total	180

Source: Author's work based on the survey results

The table above illustrates how the respondents are distributed across groups of occupations. There are 14 different occupation groups including the retired, unemployed, and students. As previously mentioned, the majority of respondents are students, thus 76 or 42.2% are in the "students" category. With a total of 25 respondents, the second largest group is "engineers and technicians", and the third is "office clerks" with 24 respondents. Among the smallest groups are "military occupations", "retired" and "unemployed" with just 1 respondent per group.

Given the fact that most of the respondents are students, in the further analysis they will be viewed as one group, and all the others as another. In summary tables it is possible to observe the difference between students and respondents in other groups of occupation more clearly.

Table 2: Response summary to the question "How well are you informed about ecological state of the planet and environmental protection?"

How well are you informed about ecological state of the planet and environmental protection?	Students	%	Others	%	All	%
Not informed at all	0	0,0%	2	1,9%	2	1,1%
Below average	17	22,4%	15	14,4%	32	17,8%
Average	32	42,1%	47	45,2%	79	43,9%

Very well	24	31,6%	27	26,0%	51	28,3%
Excellent	3	3,9%	13	12,5%	16	8,9%
Total	76	100,0%	104	100,0%	180	100,0%

Source: Author's work based on the survey results

A significant number of total respondents, 79 or 43.9%, responded with "average" in self-assessment of the ecological state of the planet and knowledge on environmental protection. Only 2 or 1.1% consider themselves not informed at all on this matter.

In case we compare students and other respondents, we can notice that the results of both groups are similar with the exception of answers "below average" with 14.4% of others as opposed to 22.4% of students, and "excellent" with 12.5% of others as opposed to only 3.9% of students.

Table 3: Response summary to the question "Which sources of information do you use to inform yourself about ecological state of the planet and environmental protection?"

Which sources of information do you use to inform yourself about ecological state of the planet and environmental protection?	**Students**	**%**	**Others**	**%**	**All**	**%**
I do not inform myself	3	3,9%	5	4,8%	8	4,4%
Conversation with others	6	7,9%	12	11,5%	18	10,0%
Social media	33	43,4%	25	24,0%	58	32,2%
Portals and blogs	10	13,2%	7	6,7%	17	9,4%
TV and documentaries	21	27,6%	44	42,3%	65	36,1%
Books, scientific papers	3	3,9%	11	10,6%	14	7,8%
Total	76	100,0%	104	100,0%	180	100,0%

Source: Author's work based on the survey results

Table 3 represents the major information sources the respondents use to educate themselves about the planet's ecological state and the protection of the environment. They were able to choose between 5 groups of sources of information or could choose the answer "I do not inform myself". It is evident that the most outstanding categories include the "TV and documentary movies/shows" and "social media", which together account for 68.3%. Only 14 chose "books, scientific magazines/paper and newspaper" as their primary source of information.

Comparing the statistics for students and others we can see that twice as many students are educating themselves through social media platforms. On the other hand, almost two times more others watch TV and documentaries, and nearly three times more read books and scientific papers on the topic than students do.

Table 4: Response summary to the question "How often do you think about energy consumption during your everyday activities (e.g. turning off the tap while brushing your teeth)?"

How often do you think about energy consumption during your everyday activities?	Students	%	Others	%	All	%
Always	14	18,4%	18	17,3%	32	17,8%
Often	25	32,9%	44	42,3%	69	38,3%
Sometimes	19	25,0%	27	26,0%	46	25,6%
Rarely	16	21,1%	12	11,5%	28	15,6%
Never	2	2,6%	3	2,9%	5	2,8%
Total	76	100,0%	104	100,0%	180	100,0%

Source: Author's work based on the survey results

Table 4 shows how often the respondents think of energy consumption during their daily activities, with the possibility of choosing between "never", "rarely", "sometimes", "often"

and “always”. In general, 38.3% of the total respondents answered with “often” and only 2.8% with “never”.

There are no major discrepancies between students and others but in the response “rarely” selected by 21.1% of students and just 11.5% of others.

Table 5: Response summary to the question “How often do you sort waste?”

How often do you sort waste?	Students	%	Others	%	All	%
Always	7	9,2%	28	19,4%	35	19,4%
Often	24	31,6%	33	31,7%	57	31,7%
Sometimes	21	27,6%	19	22,2%	40	22,2%
Rarely	12	15,8%	16	15,6%	28	15,6%
Never	12	15,8%	8	11,1%	20	11,1%
Total	76	100,0%	104	100,0%	180	100,0%

Source: Author’s work based on the survey results

When asked how often do they sort waste, 19.4% claim that they do it “always”, 31.7% “often”, 22.2% “sometimes”, 15.6% “rarely” and 11.1% “never”.

The others are a bit stronger than the students in the response “always”, where twice as many of them claim that they sort the waste on a regular basis.

Table 6: Response summary to the question “How often do you buy eco products?”

How often do you buy eco products?	Students	%	Others	%	All	%
Always	1	1,3%	2	1,9%	3	1,7%
Often	17	22,4%	33	31,7%	50	27,8%
Sometimes	30	39,5%	47	45,2%	77	42,8%
Rarely	24	31,6%	19	18,3%	43	23,9%
Never	4	5,3%	3	2,9%	7	3,9%
Total	76	100,0%	104	100,0%	180	100,0%

Source: Author’s work based on the survey results

Table 7: Response summary to the question "Are you willing to set aside more money for an environmentally friendly product?"

Are you willing to set aside more money for an environmentally friendly product?	Students	%	Others	%	All	%
Yes, always	4	5,3%	15	14,4%	19	10,6%
Often	19	25,0%	28	26,9%	47	26,1%
Sometimes	29	38,2%	43	41,3%	72	40,0%
Rarely	15	19,7%	14	13,5%	29	16,1%
No, I am not	9	11,8%	4	3,8%	13	7,2%
Total	76	100,0%	104	100,0%	180	100,0%

Source: Author's work based on the survey results

Table 8: Response summary to the question "How would you evaluate the offering of eco products on the Croatian market?"

How would you evaluate the offering of eco products on the Croatian market?	Students	%	Others	%	All	%
Excellent	0	0,0%	0	0,0%	0	0,0%
Very good	2	2,6%	12	11,5%	14	7,8%
Good	19	25,0%	38	36,5%	57	31,7%
Fair	32	42,1%	31	29,8%	63	35,0%
Poor	23	30,3%	23	22,1%	46	25,6%
Total	76	100,0%	104	100,0%	180	100,0%

Source: Author's work based on the survey results

Tables 5-7 are based on questions concerning eco products i.e. the frequency of eco-friendly purchases by the respondents, their willingness to allocate more money for those purchases and their assessment of eco products offering on the Croatian market. Out of the total of 180

respondents, 77 or 42.8% stated that they buy eco products “sometimes”. Similarly, 72 or 40.0% are “sometimes” willing to give more money for an eco-friendly product. On the other hand, only 7 or 3.9% stated that they “never” buy eco products and 13 or 7.2% that they are not ready to ever set aside more money for an eco-friendly purchase.

When it comes to evaluation of the eco product offering on the Croatian market, most of the respondents answered with “good” – 57 or 31.7%, “fair” – 63 or 35.0% and “poor” – 46 or 25.6%. Contrarily, neither of them would rate it with an “excellent”.

When it comes to eco product purchases, others tend to buy them more often than students, with most common responses being “sometimes” – 41.3% and “often” – 31.7% as opposed to students where most common responses are “rarely” – 31.6% and “sometimes” – 37.5%. When questioned if they're willing to give more money for an eco-friendly purchase, three times as many others replied with “yes, always” and three times fewer answered with “no, never” relative to students. Furthermore, when assessing the offering of eco products on the Croatian market, a slightly larger number of young people – 30.3% gave the response “poor”. Interestingly, none of the 180 respondents viewed it as “excellent”.

Table 9: Response summary to the question “In which of the following categories do you buy eco products most often?”

In which of the following categories do you buy eco products most often?	Students	%	Others	%	All	%
I do not buy eco products	7	9,2%	7	6,7%	14	7,8%
Food	42	55,3%	45	43,3%	87	48,3%
Cosmetic products	12	15,8%	9	8,7%	21	11,7%
Apparel and footwear	1	1,3%	4	3,8%	5	2,8%
Energy efficient appliances/technology	11	14,5%	28	26,9%	39	21,7%
Electric vehicles	1	1,3%	3	2,9%	4	2,2%
All of the above	2	2,6%	8	7,7%	10	5,6%
Total	76	100,0%	104	100,0%	180	100,0%

Source: Author’s work based on the survey results

Table 10: Response summary to the question "What do you consider the most important factor when choosing a product?"

What do you consider the most important factor when choosing a product?	Students	%	Others	%	All	%
Price	16	21,1%	16	17,8%	32	17,8%
Packaging and appearance	2	2,6%	3	2,8%	5	2,8%
Ingredients and quality	42	55,3%	54	53,3%	96	53,3%
Vegan origin/not tested on animals	2	2,6%	5	3,9%	7	3,9%
Eco-friendly production	0	0,0%	9	5,0%	9	5,0%
All of the above	14	18,4%	17	17,2%	31	17,2%
Total	76	100,0%	104	100,0%	180	100,0%

Source: Author's work based on the survey results

In tables 9 and 10 we can see in which of the product categories the respondents buy the eco products most often and what do they find most important when choosing between products to purchase.

Among the 5 product groups, including food, cosmetics, apparel and footwear, electric vehicles and energy-efficient appliances/technology, the highest number of respondents suggested that they most often purchase environmentally friendly food products, as many as 87 or 48.3%. The second largest category is energy efficient appliances/technology with a response rate of 39 or 21.7%. Merely 10 or 5.6% tend to make purchases in all mentioned product categories.

Table 9 refers to the question in which respondents were to identify the most important factor they consider when choosing the product to be purchased, and these were the price, ingredients and quality, eco-friendly production, packaging and appearance and vegan origin of the product. More than half or 53.3% responded that for them, ingredients and quality are the most essential factors. Only 5 or 2.8% said they find packaging and appearance to be the most significant factor.

The results are very similar when comparing others and students, except the fact that others are more concerned with environmentally friendly production and less concerned with prices than students.

Table 11: Response summary to the question "How often do you use plastic bags when shopping?"

How often do you use plastic bags when shopping?	Students	%	Others	%	All	%
Always	5	6,6%	6	6,1%	11	6,1%
Often	22	28,9%	22	24,4%	44	24,4%
Sometimes	19	25,0%	35	30,0%	54	30,0%
Rarely	26	34,2%	32	32,2%	58	32,2%
Never	4	5,3%	9	7,2%	13	7,2%
Total	76	100,0%	104	100,0%	180	100,0%

Source: Author's work based on the survey results

Table 12: Response summary to the question "How often do you use plastic bottles?"

How often do you use plastic bottles?	Students	%	Others	%	All	%
Always	7	9,2%	6	5,8%	13	7,2%
Often	45	59,2%	55	52,9%	100	55,6%
Sometimes	19	25,0%	23	22,1%	42	23,3%
Rarely	4	5,3%	16	15,4%	20	11,1%
Never	1	1,3%	4	3,8%	5	2,8%
Total	76	100,0%	104	100,0%	180	100,0%

Source: Author's work based on the survey results

Table 13: Response summary to the question "Which packaging do you choose when shopping online?"

Which packaging do you	Student	%	Others	%	All	%

choose when shopping online?	s					
Regular	16	21,1%	12	11,5%	28	15,6%
Eco	8	10,5%	18	17,3%	26	14,4%
I am not sure	15	19,7%	29	27,9%	44	24,4%
I do not care	37	48,7%	45	43,3%	82	45,6%
Total	76	100,0%	104	100,0%	180	100,0%

Source: Author's work based on the survey results

Tables 11-13 are based on questions regarding the product packaging, i.e. how often are they using plastic bags and plastic bottles, and which type of packaging they opt for when shopping online.

When it comes to plastic bags, the most common answers were rarely, sometimes and often with 58 or 32.2% saying that they use the plastic bags "rarely", 54 or 30% "sometimes" and 44 or 24.5% "often". Only 13 or 72.2% said that they "never" use the plastic bags.

The situation is a bit worse with disposable bottles where as many as 100 or more than half of the respondents admit to be using plastic bottles "often", and only 5 or 2.8% "do not" use them at all.

When shopping online, 82 or 45.6% of the total respondents stated that they "do not care" about the packaging when shopping online and only 26 or 14.4% opt for eco packaging.

In tables 11 and 12 it is visible that results of both groups are yet again similar, whereas table 13 shows the discrepancy between students and others when asked what kind of packaging they choose when shopping online. Students prefer regular packaging, whereas others tend to choose eco packaging more often.

The following questions (19-27) are based on Likert scale showing the level of agreement the participant has with the statement. Those questions will be summarized the in 3 individual tables, one for students, one for others and one including the both groups.

Table 14: Response summary to questions 19-27 for "All respondents"

Level of agreement participants have with the statement	1	2	3	4	5	Total
When buying cosmetic products, I make sure to check the composition and choose the ones with the most natural ingredients.	11,7%	13,9%	26,1%	28,3%	20,0%	100,0%
I avoid buying and using disposable plastic products (cutlery, plates and glasses, straws, etc.).	5,6%	13,3%	21,1%	23,9%	36,1%	100,0%
I rather rarely buy high quality products than often those of lower quality.	6,7%	9,4%	20,6%	21,1%	42,2%	100,0%
Eco products are overpriced.	0,6%	11,7%	30,6%	30,0%	27,2%	100,0%
The higher price of eco products compared to conventional ones is justified.	4,4%	12,2%	31,7%	34,4%	17,2%	100,0%
I find eco products to be of better quality compared to conventional ones.	1,1%	5,6%	38,9%	37,2%	17,2%	100,0%
I avoid buying products that I think have excess packaging compared to the size or quantity of the product.	10,0%	15,0%	32,2%	19,4%	23,3%	100,0%
I am familiar with the term ‘greenwashing’	31,7%	12,8%	22,8%	18,3%	14,4%	100,0%
I am familiar with the term ‘fast-fashion’.	14,4%	10,0%	22,8%	18,9%	33,9%	100,0%

Source: Author’s work based on the survey results

Table 14 shows us that, when asked whether they check the quality of cosmetic products before purchasing them, 51 or 28.3% of total respondents “agreed”. Only 27 or 15% “strongly disagree”.

For questions 20 and 21, we can see quite similar trend. When questioned whether they avoid buying disposable plastic goods and regular sales of poorer quality items, only around 6% of respondents replied with “strong disagreement”. Around 20% “neither agree nor disagree” and around 40% “strongly agree”.

In questions 22 and 23, the respondents were asked for their view on the price points of eco products. More than 50% "agree" and "strongly" agree that eco products are overpriced. However, more than 50% also think that these prices are justified.

Moreover, 70 or 38.9% of the respondents "neither agree nor disagree" with the statement that eco products are of better quality than conventional ones. Similarly, 67 or 37.2% "agree" with this statement and only 2 or 1.1% answers suggest a "strong disagreement".

When asked if they avoid products with excess packaging, they provided varying responses. The most common one was "neither agree nor disagree" with 58 or 32.2% of responses.

Question 26 and 27 demonstrate the extent to which respondents are familiar with terms "greenwashing" and "fast-fashion". Regarding the statement: "I am familiar with the term 'greenwashing'", 57 or 31.7% answered with "strongly disagree" and only 26 or 14.4% with "strongly agree". The statement: "I am familiar with the term 'fast-fashion'" had a bit better response with 61 or 33.9% claiming that they "strongly agree".

Table 15: Response summary to questions 19-27 for "Students"

Level of agreement participants have with the statement	1	2	3	4	5	Total
When buying cosmetic products, I make sure to check the composition and choose the ones with the most natural ingredients.	15,8%	13,2%	27,6%	30,3%	13,2%	100,0%
I avoid buying and using disposable plastic products (cutlery, plates and glasses, straws, etc.).	7,9%	13,2%	22,4%	26,3%	30,3%	100,0%
I rather rarely buy high quality products than often those of lower quality.	3,9%	10,5%	18,4%	25,0%	42,1%	100,0%
Eco products are overpriced.	0,0%	11,8%	27,6%	32,9%	27,6%	100,0%
The higher price of eco products compared to conventional ones is justified.	3,9%	17,1%	26,3%	38,2%	14,5%	100,0%
I find eco products to be of better quality compared to conventional ones.	1,3%	9,2%	48,7%	30,3%	10,5%	100,0%

I avoid buying products that I think have excess packaging compared to the size or quantity of the product.	9,2%	17,1%	31,6%	21,1%	21,1%	100,0%
I am familiar with the term 'greenwashing'	35,5%	17,1%	19,7%	13,2%	14,5%	100,0%
I am familiar with the term 'fast-fashion'.	13,2%	9,2%	21,1%	18,4%	38,2%	100,0%

Source: Author's work based on the survey results

Table 16: Response summary to questions 19-27 for "Others"

Level of agreement participants have with the statement	1	2	3	4	5	Total
When buying cosmetic products, I make sure to check the composition and choose the ones with the most natural ingredients.	8,7%	14,4%	25,0%	26,9%	25,0%	100,0%
I avoid buying and using disposable plastic products (cutlery, plates and glasses, straws, etc.).	3,8%	13,5%	20,2%	22,1%	40,4%	100,0%
I rather rarely buy high quality products than often those of lower quality.	8,7%	8,7%	22,1%	18,3%	42,3%	100,0%
Eco products are overpriced.	1,0%	11,5%	32,7%	27,9%	26,9%	100,0%
The higher price of eco products compared to conventional ones is justified.	4,8%	8,7%	35,6%	31,7%	19,2%	100,0%
I find eco products to be of better quality compared to conventional ones.	1,0%	2,9%	31,7%	42,3%	22,1%	100,0%
I avoid buying products that I think have excess packaging compared to the size or quantity of the product.	10,6%	14,4%	31,7%	18,3%	25,0%	100,0%
I am familiar with the term 'greenwashing'	28,8%	9,6%	25,0%	22,1%	14,4%	100,0%
I am familiar with the term 'fast-fashion'.	15,4%	10,6%	24,0%	19,2%	30,8%	100,0%

Source: Author's work based on the survey results

Again, tables 15 and 16 show us what the situation is if we compare students and others. There are not too many differences, in general, with the exception of some questions. The first such question is question 19 in which respondents were asked whether they check the composition of cosmetic products before purchasing them. Others pay more attention to the product composition as they replied "never" two times less, and "always" two times more than students. Also, when asked whether the higher price of eco products compared to conventional ones is justified, others agree with this to a much greater extent than students do.

4.1.2. List of Used Attributes

Table 17: List of Used Attributes

N	ATTRIBUTE NAME	ATTRIBUTE FORMAT	ATTRIBUTE MODALTY (NOMINALS)
1	Gender	Nominal	'M', 'F'
2	Age_group	Nominal	'Under_18', '18-24', '25-49', '50-65', '65_plus'
3	Lev_educ	Nominal	'Primary', 'Secondary', 'Undergraduate_univ', 'Graduate_univ', 'Postgraduate_univ'
4	Field_educ	Nominal	'Natural', 'Technical', 'Biomedicine_health', 'Biotechnical', 'Social', 'Humanistic', 'Art', 'Other'
5	Occupation	Nominal	'Government_bodies_directors', 'Scientists_experts', 'Engineers_technicians', 'Office_clerks', 'Service_retail', 'Agricult_hunting_forestry_fishing', 'Crafts', 'Handlers_assemblers', 'Simple_occupations', 'Military', 'Entrepreneur', 'Retired', 'Unemployed', 'Students'
6	Net_income_HRK	Nominal	'≤3500', '3501-7500', '7501-14000', '14001-25000', '>25000'
7	Eco_awareness_level	Nominal	'Not_at_all', 'Below_average', 'Average'. 'Very_well', 'Excellent'
8	Eco_information_sources	Nominal	'None', 'Conversation', 'Social_media',

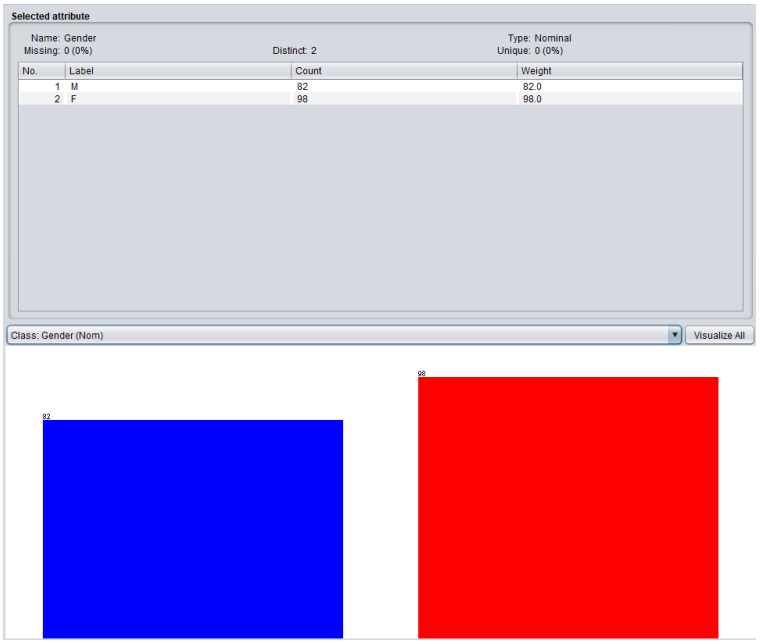
			'Portals_blogs', 'TV_documentaries', 'Books_magazines_newspaper'
9	Daily_energy_consumption	Nominal	'Never', 'Rarely', 'Sometimes', 'Often', 'Always'
10	Sort_waste	Nominal	'Never', 'Rarely', 'Sometimes', 'Often', 'Always'
11	Buy_eco_products	Nominal	'Never', 'Rarely', 'Sometimes', 'Often', 'Always'
12	More_money_eco	Nominal	'Never', 'Rarely', 'Sometimes', 'Often', 'Always'
13	Eco_offer_Croatia	Nominal	'Poor', 'Fair', 'Good', 'Very_good', 'Excellent'
14	Eco_product_categories	Nominal	'Not_buy', 'Food', 'Appliances_technology', 'Cosmetic_products', 'Clothes_footwear', 'Electric_vehicles', 'All'
15	Most_important_factor	Nominal	'Price', 'Ingredients_quality', 'Eco_production', 'Packaging_apperance', 'Vegan', 'All'
16	Plastic_bags	Nominal	'Never', 'Rarely', 'Sometimes', 'Often', 'Always'
17	Plastic_bottles	Nominal	'Never', 'Rarely', 'Sometimes', 'Often', 'Always'
18	Online_shopp_packaging	Nominal	'Regular', 'Eco', 'Not_care', 'Not_sure'
19	Cosmetic_check_composition	Nominal	‘Strongly disagree’, ‘Disagree’, ‘Neither agree nor disagree’, ‘Agree’, ‘Strongly agree’
20	Avoid_disposable_plastic	Nominal	‘Strongly disagree’, ‘Disagree’, ‘Neither agree nor disagree’, ‘Agree’, ‘Strongly agree’
21	Rarely_better_quality	Nominal	‘Strongly disagree’, ‘Disagree’, ‘Neither agree nor disagree’, ‘Agree’, ‘Strongly agree’
22	Eco_overpriced	Nominal	‘Strongly disagree’, ‘Disagree’, ‘Neither agree nor disagree’, ‘Agree’, ‘Strongly agree’

23	Eco_price_justified	Nominal	'Strongly disagree', 'Disagree', 'Neither agree nor disagree', 'Agree', 'Strongly agree'
24	Eco_better_quality	Nominal	'Strongly disagree', 'Disagree', 'Neither agree nor disagree', 'Agree', 'Strongly agree'
25	Avoid_excess_packaging	Nominal	'Strongly disagree', 'Disagree', 'Neither agree nor disagree', 'Agree', 'Strongly agree'
26	Familiar_greenwashing	Nominal	'Strongly disagree', 'Disagree', 'Neither agree nor disagree', 'Agree', 'Strongly agree'
27	Familiar_fast_fashion	Nominal	'Strongly disagree', 'Disagree', 'Neither agree nor disagree', 'Agree', 'Strongly agree'

Source: Author's work based on the survey results

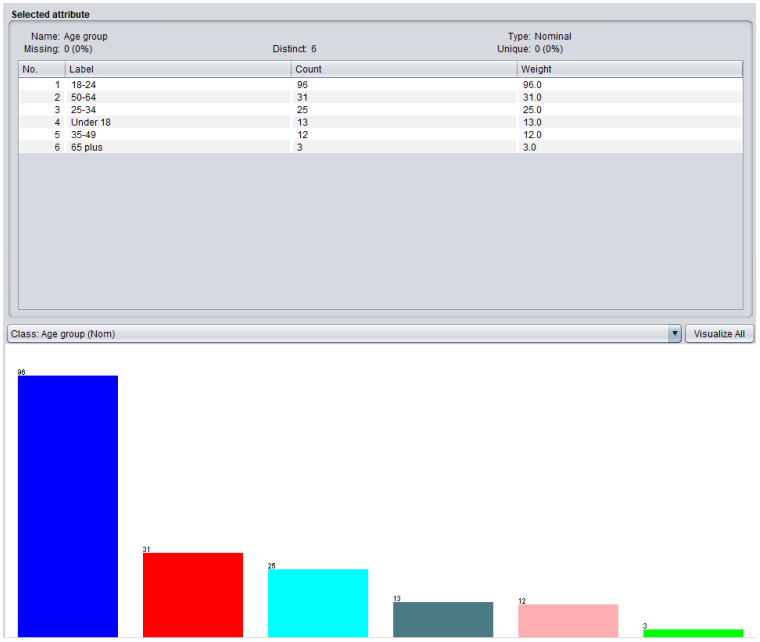
4.1.3. Graphical Representation of Attributes in Weka

Figure 4: Graphical representation of the attribute 'Gender' in Weka



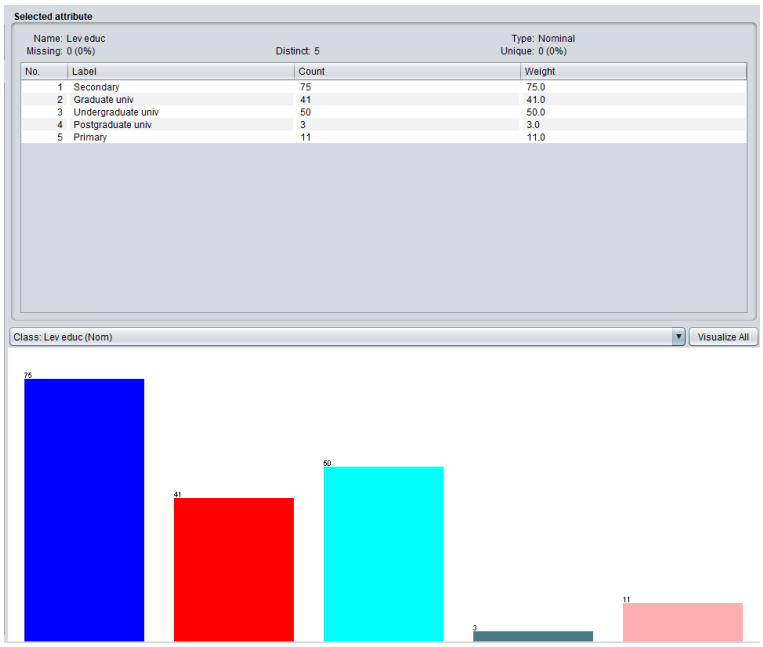
Source: Weka

Figure 5: Graphical representation of the attribute 'Age_group' in Weka



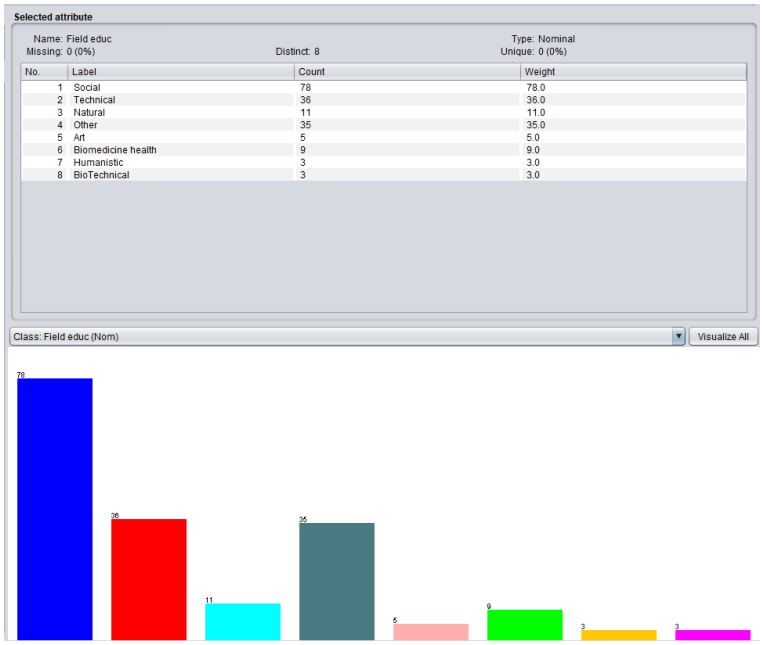
Source: Weka

Figure 6: Graphical representation of the attribute 'Lev_educ' in Weka



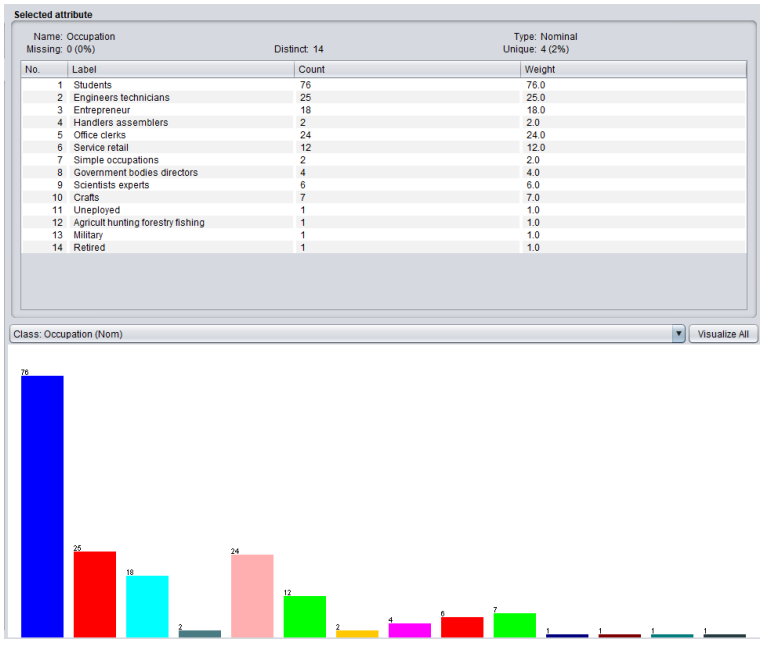
Source: Weka

Figure 7: Graphical representation of the attribute 'Field_educ' in Weka



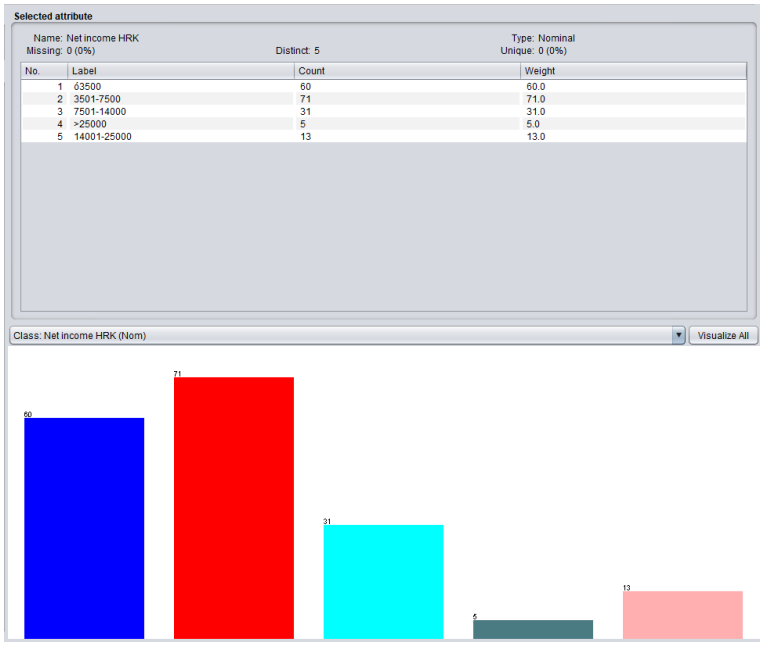
Source: Weka

Figure 8: Graphical representation of the attribute 'Occupation' in Weka



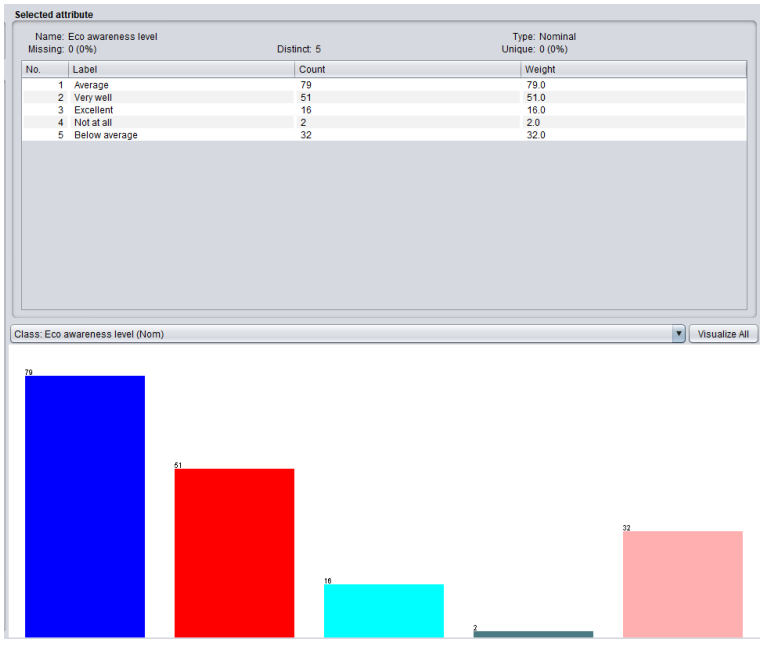
Source: Weka

Figure 9: Graphical representation of the attribute 'Net_income_HRK' in Weka



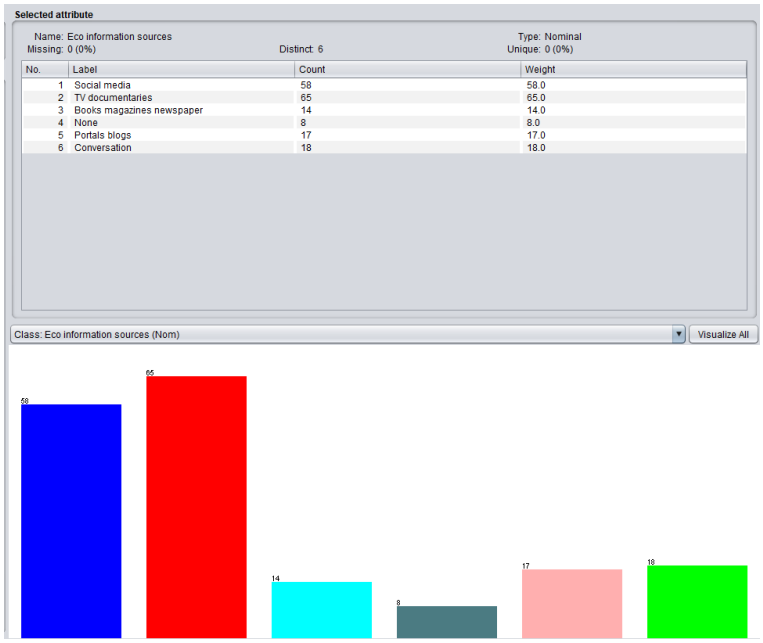
Source: Weka

Figure 10: Graphical representation of the attribute 'Eco_awareness_level' in Weka



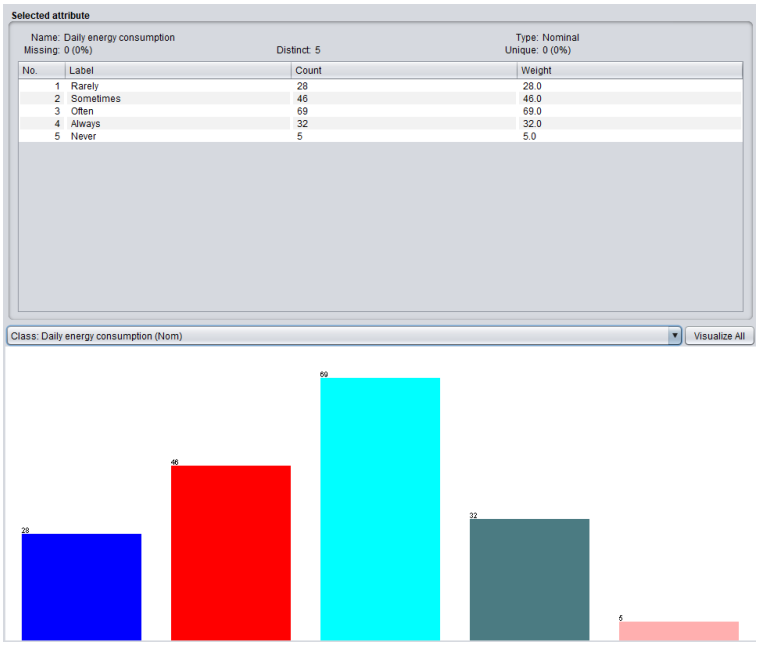
Source: Weka

Figure 11: Graphical representation of the attribute 'Eco_information_sources' in Weka



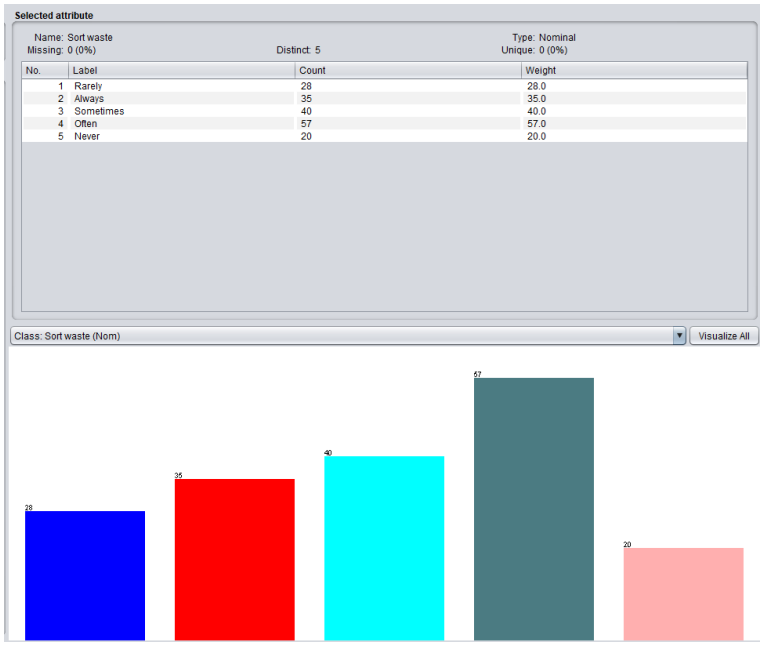
Source: Weka

Figure 12: Graphical representation of the attribute 'Daily_energy_consumption' in Weka



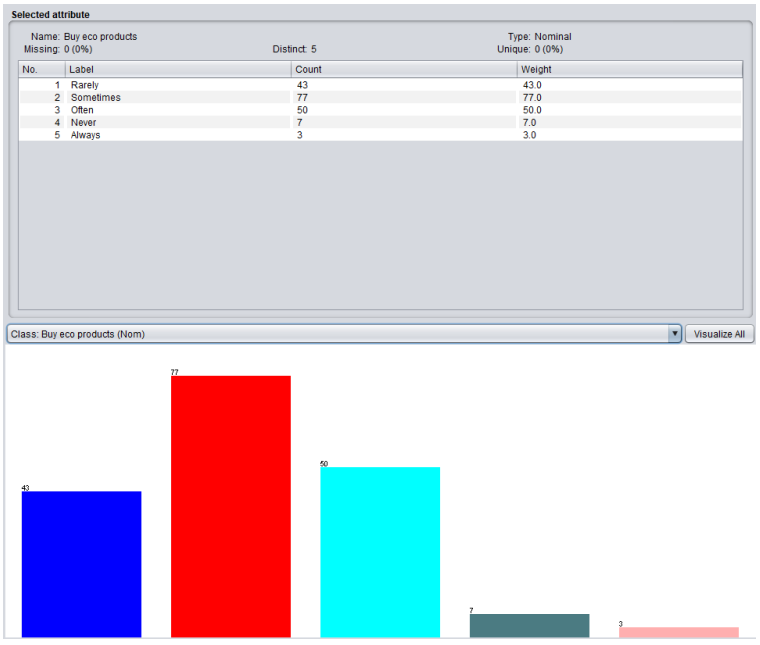
Source: Weka

Figure 13: Graphical representation of the attribute 'Sort_waste' in Weka



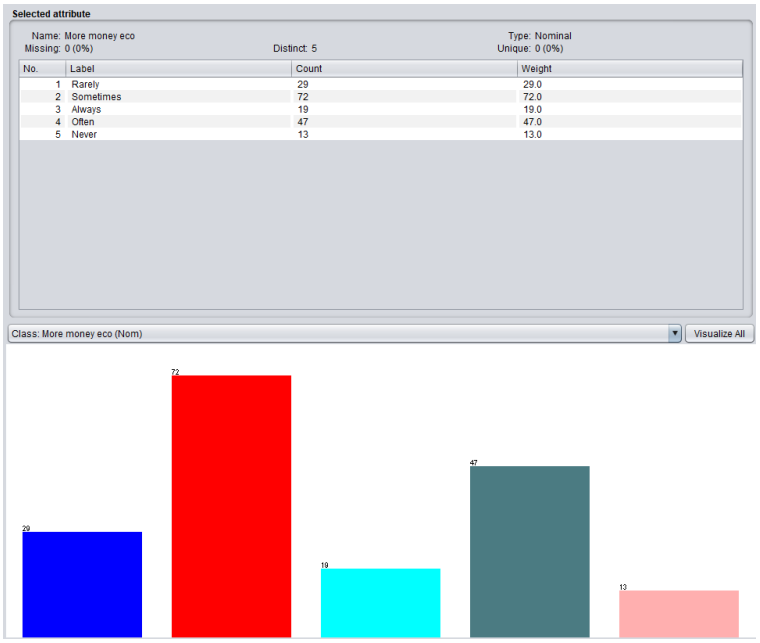
Source: Weka

Figure 14: Graphical representation of the attribute 'Buy_eco_products' in Weka



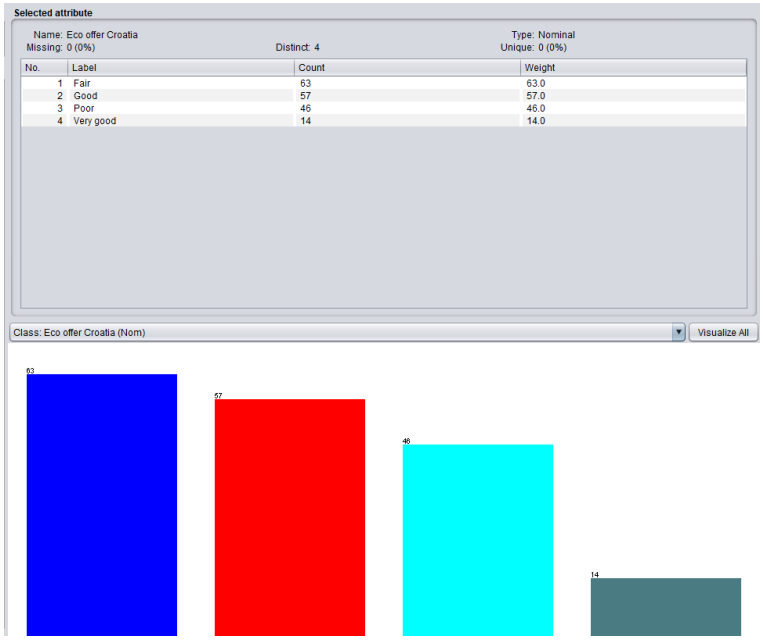
Source: Weka

Figure 15: Graphical representation of the attribute 'More_money_eco' in Weka



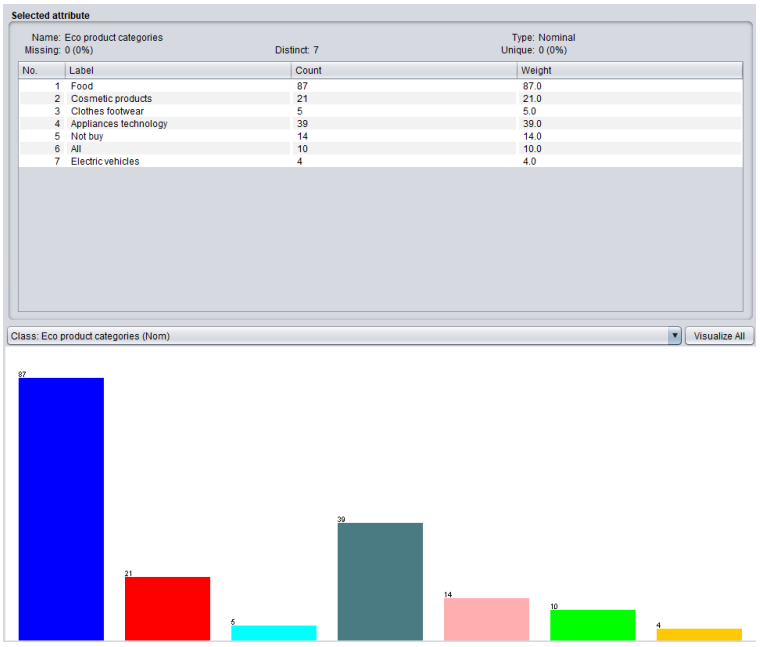
Source: Weka

Figure 16: Graphical representation of the attribute 'Eco_offer_Croatia' in Weka



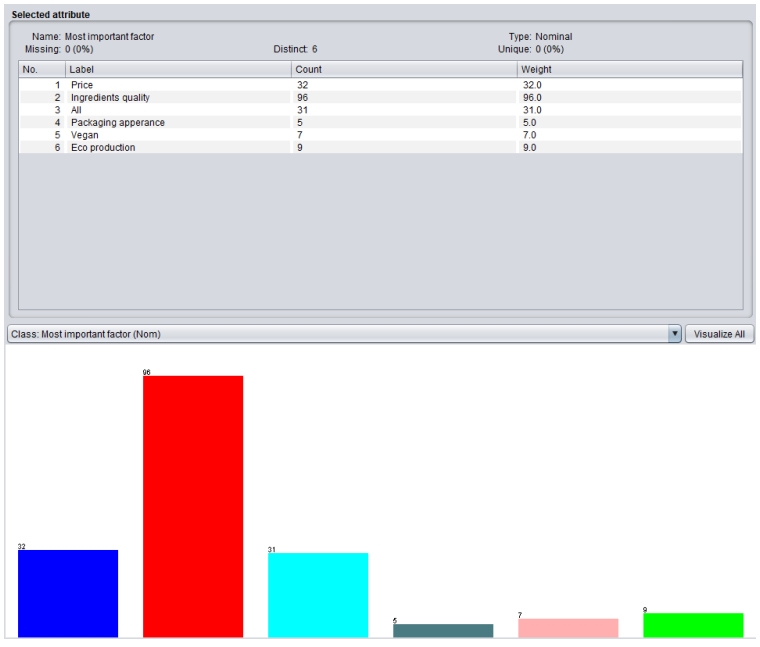
Source: Weka

Figure 17: Graphical representation of the attribute 'Eco_product_categories' in Weka



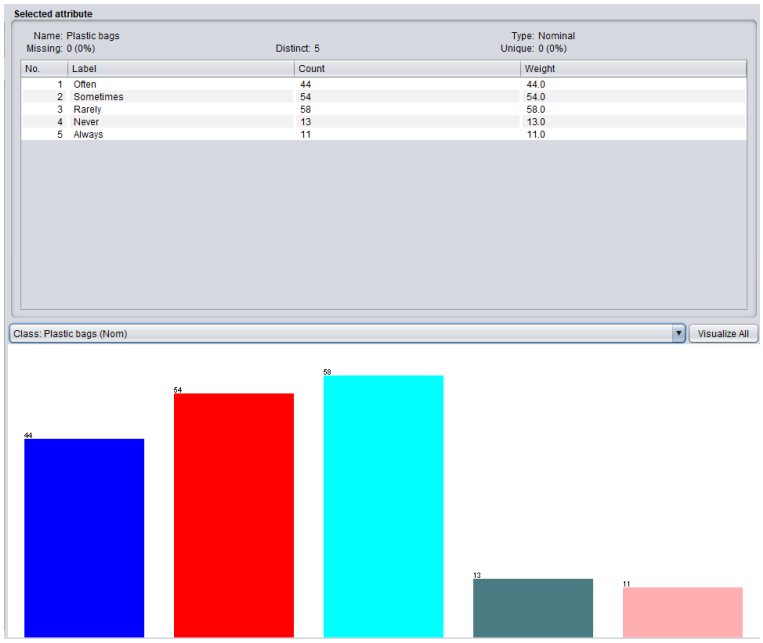
Source: Weka

Figure 18: Graphical representation of the attribute 'Most_important_factor' in Weka



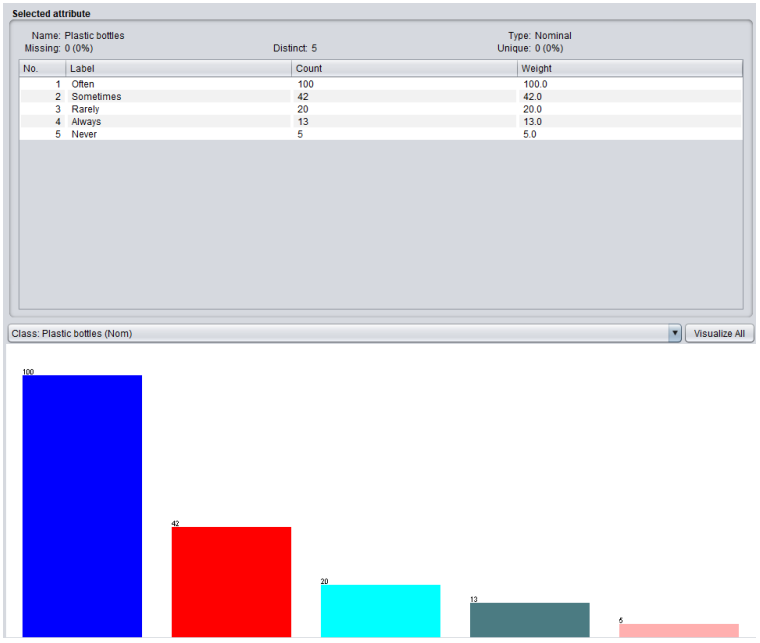
Source: Weka

Figure 19: Graphical representation of the attribute 'Plastic_bags' in Weka



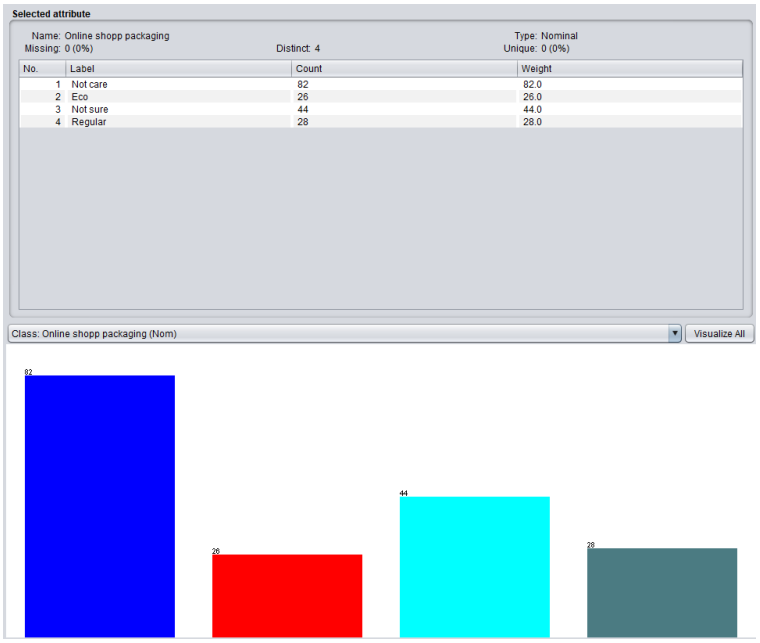
Source: Weka

Figure 20: Graphical representation of the attribute 'Plastic_bottles' in Weka



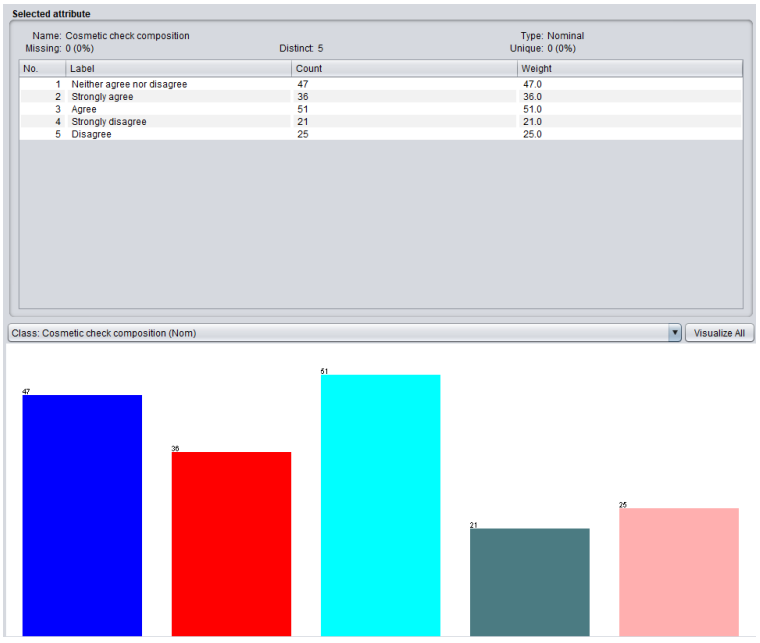
Source: Weka

Figure 21: Graphical representation of the attribute 'Online_shopp_packaging' in Weka



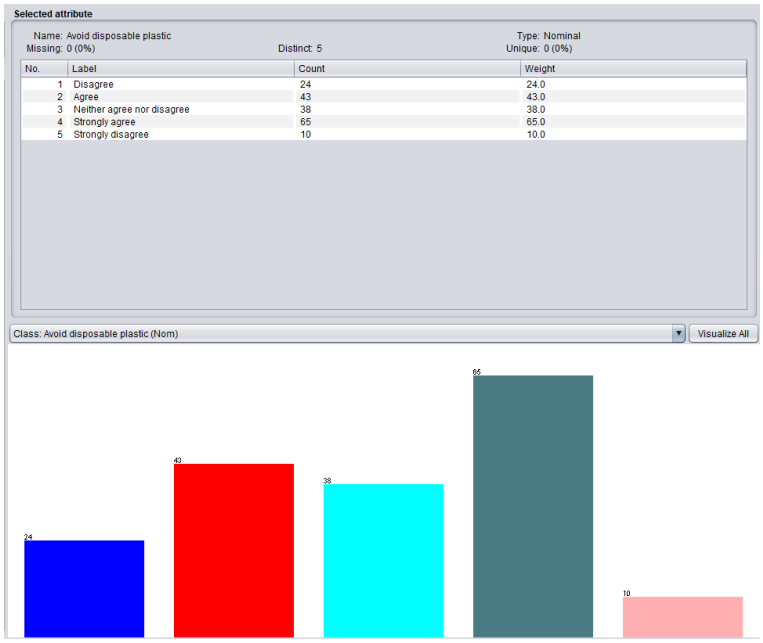
Source: Weka

Figure 22: Graphical representation of the attribute 'Cosmetic_check_composition' in Weka



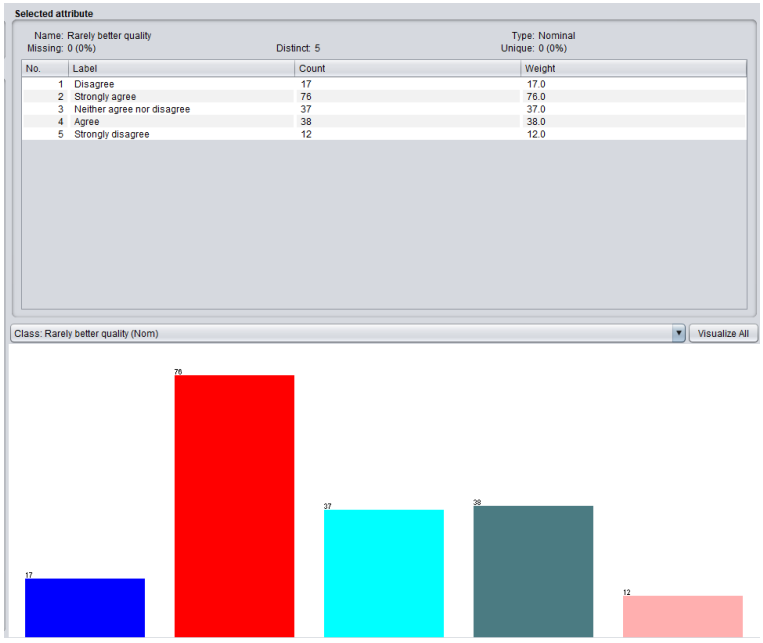
Source: Weka

Figure 23: Graphical representation of the attribute 'Avoid_disposable_plastic' in Weka



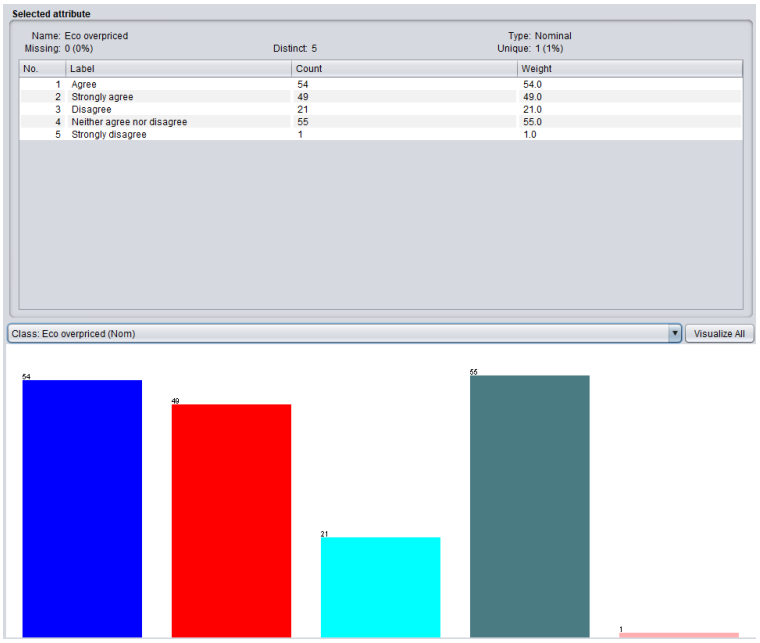
Source: Weka

Figure 24: Graphical representation of the attribute 'Rarely_better_quality' in Weka



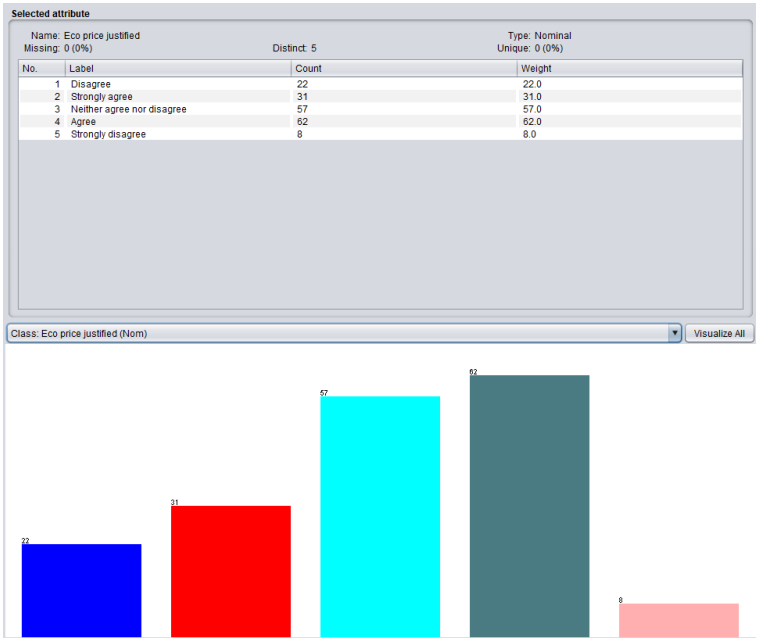
Source: Weka

Figure 25: Graphical representation of the attribute 'Eco_overpriced' in Weka



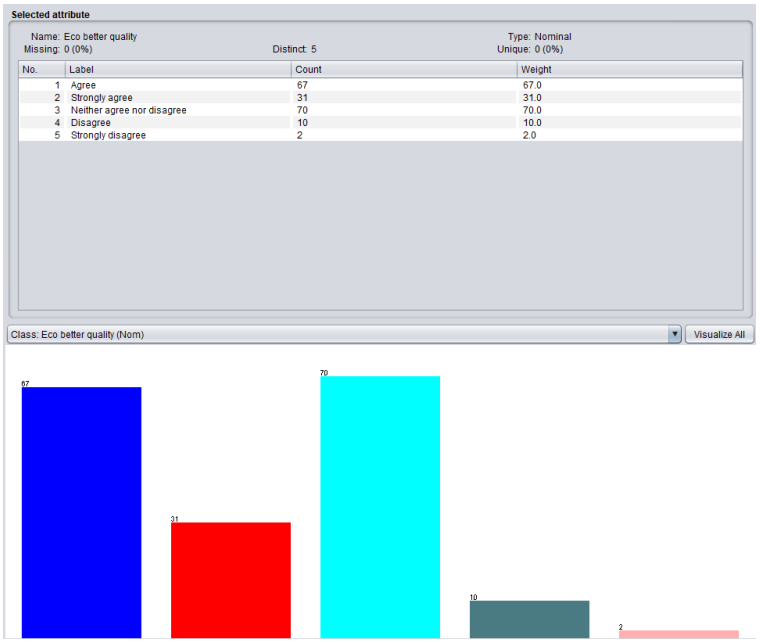
Source: Weka

Figure 26: Graphical representation of the attribute 'Eco_price_justified' in Weka



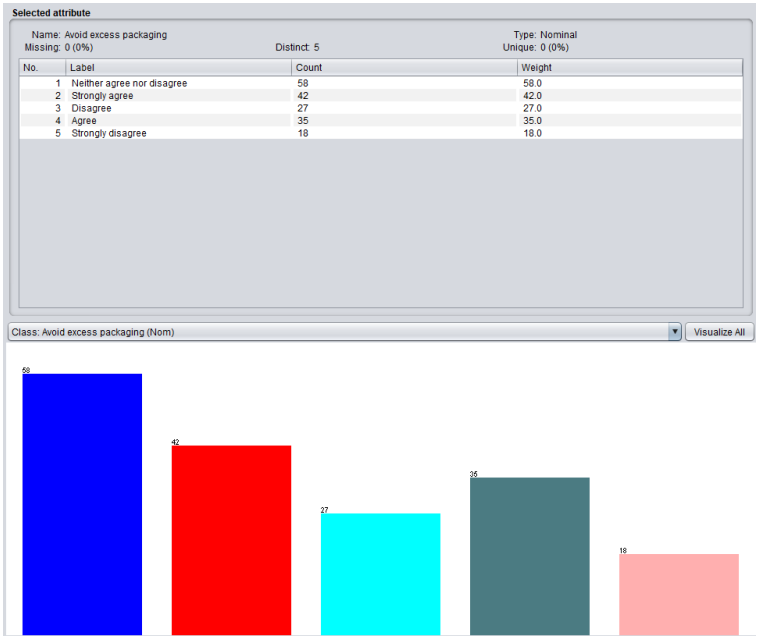
Source: Weka

Figure 27: Graphical representation of the attribute 'Eco_better_quality' in Weka



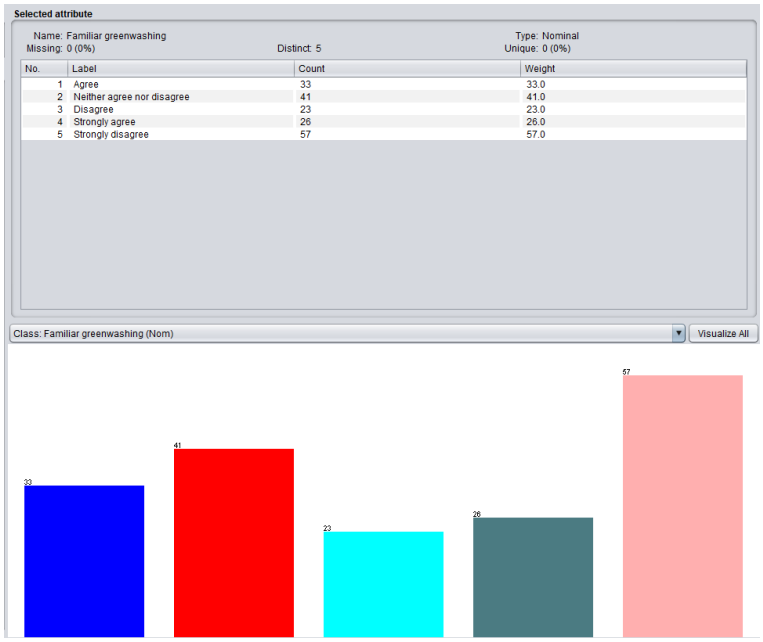
Source: Weka

Figure 28: Graphical representation of the attribute 'Avoid_excess_packaging' in Weka



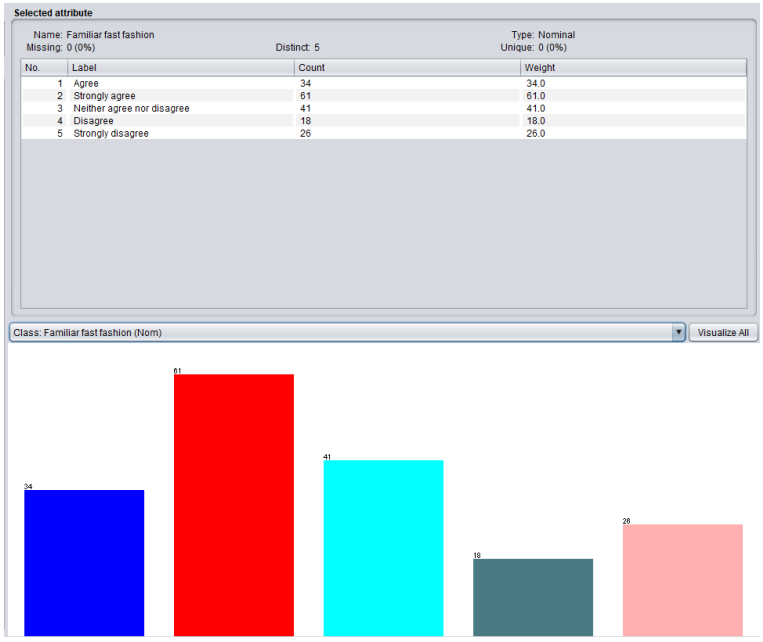
Source: Weka

Figure 29: Graphical representation of the attribute 'Familiar_greenwashing' in Weka



Source: Weka

Figure 30: Graphical representation of the attribute 'Familiar_fast_fashion' in Weka



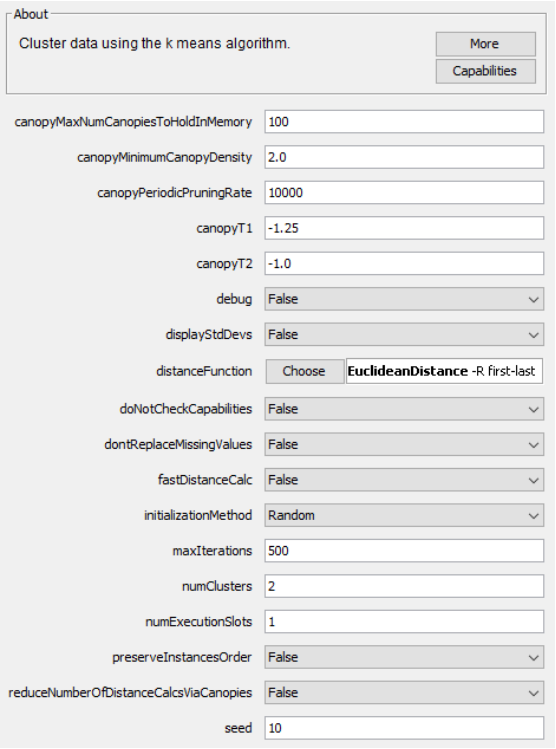
Source: Weka

4.2. Research Results

As previously mentioned, almost half of the total respondents are students. In the cluster analysis they will be viewed as one group, and the others as another group. Hence, three cluster analysis will be conducted, one for students, one for other occupations and one with all respondents. Cluster analysis will be conducted in Weka – open source machine learning workbench.

4.2.1. All Respondents

Figure 31: Selection of parameters in Weka



Source: Weka

Figure above shows the Weka screen when selecting the parameters of previously chosen algorithm – in this case, it was SimpleKMeans algorithm. For the distance we used the default, Euclidian Distance. In this step we also chose the number of clusters (2) and seed number (10), where the seed is the initialization value for the random number generator.[20]

[20] Boucheart, R.R. (2014). *Bayesian Network Classifiers in Weka.*

Seed determines the randomness of the initial values the algorithm takes for the process of clustering, and the better the randomness at the start of the process, the better the final results.

Next, we used the option “ignore attributes” to select only those we want to use in the analysis. Here we used 9 attributes including: cosmetic check composition, avoid disposable plastic, rarely better quality, eco overpriced, eco price justified, eco better quality, avoid excess packaging, familiar greenwashing and familiar fast fashion. These attributes are related to the survey questionnaire questions 18-27 in which the respondents were asked to which extent they agree with the given statements. Then we selected “start” button which initiates the clustering.

Figure 32: Results of cluster analysis for all respondents with 2 clusters

```
Missing values globally replaced with mean/mode

Final cluster centroids:
                                                          Cluster#
Attribute                                    Full Data           0                      1
                                               (180.0)     (108.0)                (72.0)
=============================================================================================================
Cosmetic check composition                        Agree       Agree Neither agree nor disagree
Avoid disposable plastic                 Strongly agree Strongly agree Neither agree nor disagree
Rarely better quality                    Strongly agree Strongly agree Neither agree nor disagree
Eco overpriced               Neither agree nor disagree Neither agree nor disagree                  Agree
Eco price justified                               Agree       Agree Neither agree nor disagree
Eco better quality           Neither agree nor disagree       Agree Neither agree nor disagree
Avoid excess packaging       Neither agree nor disagree Strongly agree Neither agree nor disagree
Familiar greenwashing              Strongly disagree Neither agree nor disagree          Strongly disagree
Familiar fast fashion                Strongly agree Strongly agree Neither agree nor disagree



Time taken to build model (full training data) : 0 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      108 ( 60%)
1       72 ( 40%)
```

Source: Weka

Figure 32 shows the results of the cluster analysis for all respondents with 2 clusters. There are 108 or 60 % of respondents in the cluster 0 and 72 or 40% of respondents in the cluster 1.

4.2.2. Students

In the second cluster analysis only the data on the students was used. The attributes chosen for this analysis are those that represent concrete positive actions and behavior through which a person can directly influence the preservation of the environment. In this way we wanted to check whether the respondents are only environmentally aware or environmentally responsible.

Figure 33: Results of cluster analysis for students only with 3 clusters

```
Missing values globally replaced with mean/mode

Final cluster centroids:
                                                    Cluster#
Attribute                                Full Data          0                1                2
                                            (76.0)     (44.0)           (22.0)           (10.0)
=============================================================================================================================
Daily energy consumption                    Often      Often            Rarely           Sometimes
Sort waste                                  Often      Often             Never           Sometimes
Buy eco products                         Sometimes  Sometimes            Rarely           Sometimes
Plastic bags                                Rarely      Often            Rarely           Sometimes
Plastic bottles                             Often      Often             Often               Often
Cosmetic check composition                  Agree      Agree Neither agree nor disagree Neither agree nor disagree



Time taken to build model (full training data) : 0 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      44 ( 58%)
1      22 ( 29%)
2      10 ( 13%)
```

Source: Weka

Figure 33 shows the results of the cluster analysis for students only with 3 clusters. There are 44 or 58% of students in the cluster 0 and 22 or 29% in the cluster 1, and 10 or 13% in the cluster 2.

4.2.3. Other Occupations

The third cluster analysis is based on the data on the rest of the respondents, i.e. those with other occupations. The chosen attributes include those which are related to eco products in a

certain way. The main idea was to group the respondents into clusters associated with eco product opinions and purchases.

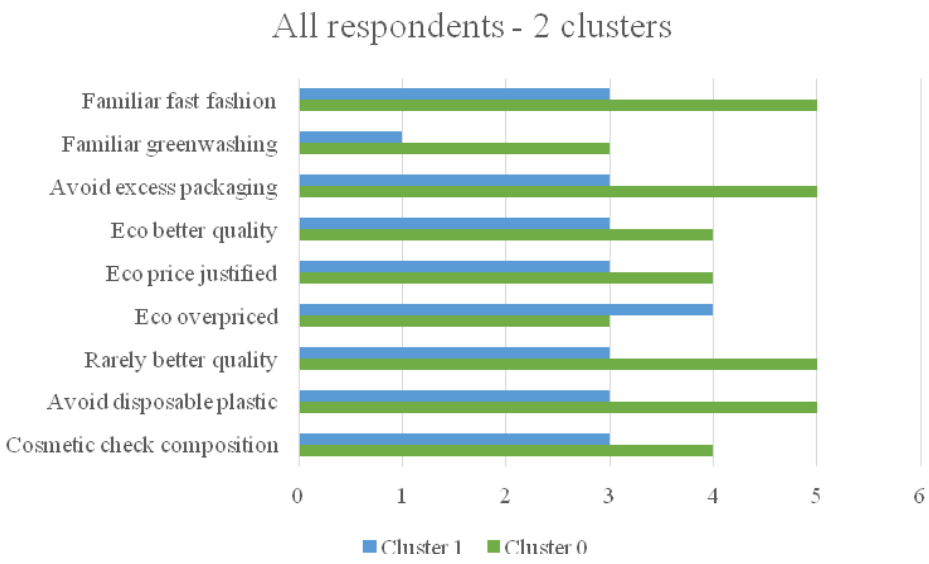
Figure 34: Results of cluster analysis for other occupations with 3 clusters

```
Missing values globally replaced with mean/mode

Final cluster centroids:
                                                    Cluster#
Attribute                                Full Data          0                  1                  2
                                           (104.0)    (54.0)             (27.0)             (23.0)
=============================================================================================================================
Buy eco products                         Sometimes  Sometimes              Often              Rarely
More money eco                           Sometimes  Sometimes              Often              Rarely
Eco offer Croatia                             Good       Fair          Very good                Good
Eco price justified  Neither agree nor disagree      Agree              Agree Neither agree nor disagree
Eco better quality                           Agree      Agree  Strongly agree Neither agree nor disagree



Time taken to build model (full training data) : 0.01 seconds

=== Model and evaluation on training set ===

Clustered Instances

0       54 ( 52%)
1       27 ( 26%)
2       23 ( 22%)
```

Source: Weka

4.3. Discussion

Figure 35: Graphical representation of results of cluster analysis for all respondents with 2 clusters



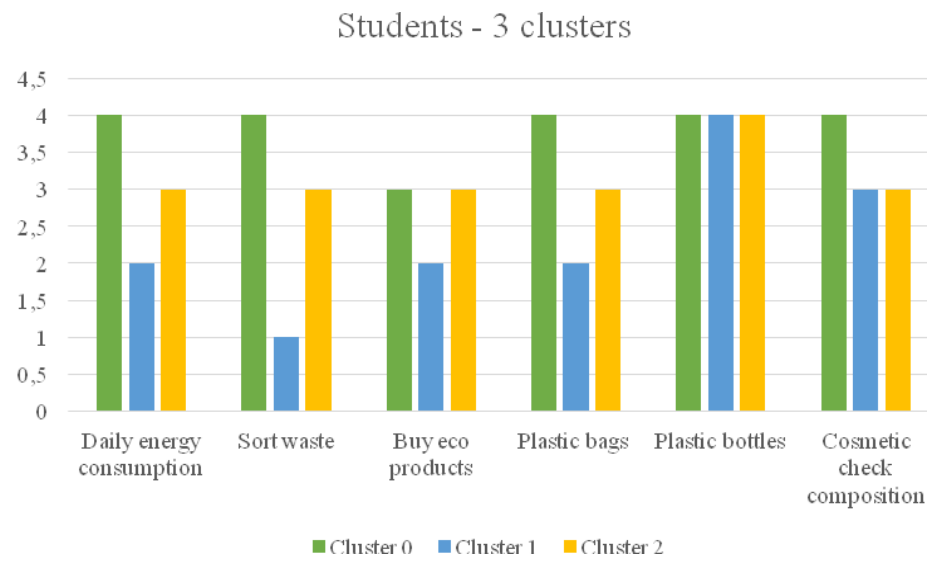
Source: Author's work based on the survey results

What can be concluded is that:

- Cluster 0 – is a group of respondents who are better acquainted with terms related to ecology, who are willing to pay more for a better and more sustainable product and who avoid excessive packaging and plastics;

- Cluster 1 – is a group of respondents who are somewhat less familiar with terms related to ecology, who believe that eco products are not necessarily of better quality than conventional ones and are not always willing to pay more for them, and who only sometimes avoid excessive packaging and plastics.

Figure 36: Graphical representation of results of cluster analysis for students only with 3 clusters

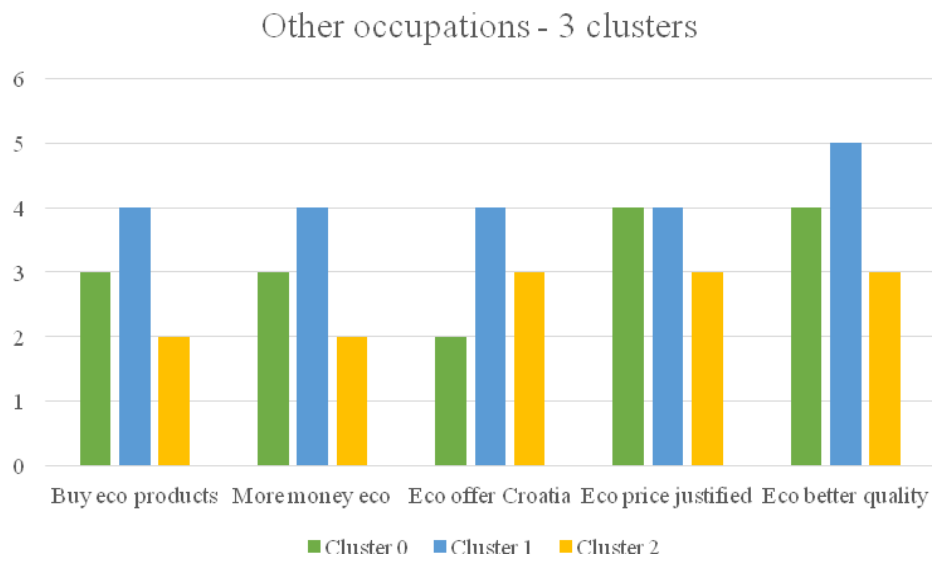


Source: Author's work based on the survey results

What can be concluded is that:

- Cluster 0 – is a group of students with the most environmentally responsible behavior if we compare these 3 clusters. They think about daily energy consumption and separate waste most often, as well as they most often check to the composition of their cosmetic products. When it comes to the plastic bags usage, they are the worst group, which uses them most often;
- Cluster 1 – is the group that acts least responsibly towards the environment. They are rarely careful regarding their daily energy consumption, and they rarely buy eco products. Also, they do not ever sort the waste. When it comes to the plastic bottles usage, all three clusters are the same and use them often;
- Cluster 2 – is a group somewhere in between clusters 0 and 1.

Figure 37: Graphical representation of results of cluster analysis for other occupations with 3 clusters



Source: Author's work based on the survey results

What can be concluded is that:

- Cluster 0 – is a group in between clusters 1 and 2;
- Cluster 1 – is a group which purchases eco-friendly products most often and thinks these products are mostly of better quality than conventional ones. Also, they are willing to give up more money for eco-friendly products and think that the prices of such products are justified;
- Cluster 2 – is a group which engages in purchasing of eco-friendly products least often. They would rarely give up more money for these purchases, however, they think the higher prices are justified.

5. CONCLUSION

The huge popularity of the internet and the great usage of smartphones has contributed to the collection of huge amounts of data in all layers of society. In order for this huge amount of data to be processed, advanced IT and tools were developed and implemented throughout the various scientific fields. Today, various companies use different data mining tools to turn the obtained data into useful information that could help them gain a competitive advantage and stand out in the market. This is actually the essence of data mining; converting a huge amount of data into useful information, and then deriving the knowledge from this information. The process of knowledge discovery in databases is extensive and requires various steps including data selection, data preprocessing, data transformation, data mining and interpretation/ evaluation of the data, and finally concluding what can be qualified as knowledge. The data mining process can produce thousands and even millions of rules and patterns that are mostly uninteresting to the user. Therefore, it is important that the pattern is easy to understand, possibly useful, innovative and what in the end represents a knowledge.

In order to make a practical application of data mining, the data related to environmental awareness of consumers was collected for the purposes of this thesis. Data for 180 respondents were collected through a survey questionnaire method, which consisted of 27 questions related to environmental awareness and environmental responsibility. The act of buying itself is much more than it seems at first glance - behind it is the whole process and the companies that analyze it in order to make a profit. The buying process includes the five stages including: recognizing the need, finding the information, assessing the alternatives, purchase, and post-purchase behavior.

Although the definition of the term environmental awareness has not yet been established, broadly defined it is the attitude towards environmental impacts of human activity. There are two types of environmentally driven behavior; environmental awareness and environmental responsibility. Environmental awareness is the positive attitude towards the environment, and environmental responsibility are the concrete positive actions and behavior.

The obtained dataset was used for conducting the cluster analysis in Weka program, which is an open-source machine learning software. Dataset was observed in terms of three distinct groups of respondents including students, those in other occupations and everyone. Hence, three cluster analysis were conducted. For the cluster analysis based on data for all

respondents, chosen attributes were those related to the survey questionnaire questions 18-27 in which the respondents were asked to which extent they agree with the given statements. Two clusters were formed; first one with 60% of respondents who are better informed on the topic and are leaning towards environmentally conscious behavior, and the second one with 40% of respondents who are less interested in the topic and participation in such behavior.

The second analysis was concerning only students, and the chosen attributes related to environmental responsibility, i.e. direct positive actions towards the preservation of the environment. This analysis resulted in three clusters; first one with 58% of students who represent the most responsible cluster, second one with 29% of students with the least responsible behavior, and the third one with 13% of students who are somewhere in the middle.

Finally, the third cluster analysis was based on the data on respondents with other occupations, i.e. not students. They were compared based on attributes regarding survey questionnaire questions 19-27 in which they were asked to which extent they agree with given statements. Three clusters were obtained in cluster analysis; first one with 52% of respondents who are in between clusters two and three, second one with 26% of respondents who purchase eco products most often and consider such products of better quality than conventional ones, and the third one with 22% of respondents who rarely give up more money for eco-friendly purchases and rarely engage in buying them.

Bibliography

1. Boucheart, R.R. (2014). *Bayesian Network Classifiers in Weka.* Available at: https://researchcommons.waikato.ac.nz/bitstream/handle/10289/85/content.pdf?sequence=1

2. Cios, K.J., Pedrycz, W., Swiniarski, R.W., & Kurgan, L.A. (2007). *Data Mining: A Knowledge Discovery Approach*. Springer. Available at: https://www.springer.com/gp/book/9780387333335

3. Dangelico, R.M., Potrandolfo, P. (2010). *From Green Product Definitions and Classifications to the Green Option Matrix.* Journal of Cleaner Production. Available at: https://www.academia.edu/316323/From_Green_Product_Definitions_and_Classifications_to_the_Green_Option_Matrix

4. DeGruy, K.B. (2000). *Healthcare Applications of Knowledge Discovery in Databases*. Journal of Healthcare Information Management. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.92.8227&rep=rep1&type=pdf

5. Delmas, M.A., Cuerel Burbano, V. (2011). *The Drivers of Greenwashing.* University of California, Berkeley. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1966721

6. Frawley, W. J., Piatetsky-Shapiro, G., & Matheus, C. J. (1992). *Knowledge Discovery in Databases: An Overview.* AI Magazine, 13(3), 57. Available at: https://www.aaai.org/ojs/index.php/aimagazine/article/view/1011

7. Gilchrist, M., Lehmann Mooers, D., Skrubbeltrang, G., & Vachon, H. (2012). *Knowledge Discovery in Databases for Competitive Advantage*. Available at: https://www.researchgate.net/publication/220038481_Knowledge_Discovery_in_Databases_for_Competitive_Advantage

8. Ginsberg, J.M., Bloom, P.N. (2004). *Choosing the Right Green Marketing Strategy*. MIT Sloan Review. Available at: https://sloanreview.mit.edu/article/choosing-the-right-greenmarketing-strategy/

9. Ham, M., Mrčela, D., Horvat, M. (2015). *Insights for Measuring Environmental Awareness.* Available at: https://hrcak.srce.hr/ojs/index.php/ekonomski-vjesnik/article/view/3661

10. Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*. Available at: https://www.sciencedirect.com/book/9780123814791/data-mining-concepts-and-techniques

11. Holmes, G., Donkin, A., & Witten, I.H. (1994). *WEKA: A Machine Learning Workbench*. Proceedings of ANZIIS '94 - Australian New Zealand Intelligent Information Systems Conference, Brisbane, Queensland, Australia, 1994, pp. 357-361, doi: 10.1109/ANZIIS.1994.396988. Available at: https://ieeexplore.ieee.org/abstract/document/396988

12. Matheus, C. J., Chan, & P.K., Piatetsky-Shapiro, G. (1998). *Knowledge Discovery in Databases. IEEE Transactions on Knowledge and Data Engineering*. Available at: https://www.researchgate.net/publication/2247895_Knowledge_Discovery_in_Databases

13. Munthiu, M.C. (2009). *The Buying Decision Process and Types of Buying Decision Behaviour*. Available at: https://pdfs.semanticscholar.org/f2d3/36232b0b67e046f38ac01c70df0f9c0dbd07.pdf

14. Ottoman, J.A. (2010). *The New Rules of Green Marketing*. Strategies, Tools, and Inspiration for Sustainable Branding. p.p. 17. Available at: https://www.bkconnection.com/static/The_New_Rules_of_Green_Marketing_EXCERPT.pdf

15. Patel, B.R., Rana, K.K. (2014). *A Survey on Decision Tree Algorithm for Classification*. IJEDR Volume 2, Issue 1. Available at:

https://www.ijedr.org/papers/IJEDR1401001.pdf

16. Peattie, K., Charter, M. (2003). The Marketing Book, p.p. 726. Available at: http://digilib.stiem.ac.id:8080/xmlui/bitstream/handle/123456789/38/Marketing%20Book.pdf?sequence=1&isAllowed=y#page=765

17. Pejić Bach, M. (2005). *Data Mining in Banking*. Zbornik Ekonomskog fakulteta u Zagrebu, Vol. 3 No. 1. Available at: https://hrcak.srce.hr/26220

18. Rokach, L., Maimon, O. (2014). *Data Mining with Decision Trees: Theory and Applications*. Available at: https://www.researchgate.net/publication/236005680_Data_mining_with_decision_trees_Theory_and_applications

19. White, K., Hardisty, D.J., & Habib, R. (2019). *The Elusive Green Consumer*. Harvard Business Review. Available at: https://hbr.org/2019/07/the-elusive-green-consumer

20. Witten, I.H., & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*.

List of Figures

List of Tables