**ICT Project Assignment**

**Task 5**

**Research Sample**

The target participants are vehicle auto dealers. In this case with a population size of 116 auto dealers (Car Dealers in Malta Gozo, 2021), a 10% margin of error, and a 95% confidence, the sample size is of 53 respondents (Sample Size Calculator by Raosoft, Inc., 2021). This target group has been selected as it fit due to its relevance to the study in confirming the information discovered. These calculations have been made using Raosoft, shown in figure 1 below.

Online surveys to gather information from Auto dealers. Topics for the survey would be regarding the prices, market research, if such a system would be ideal for their use, would they suggest it etc. This will take place online. The experiment for this study will consist of understanding the idea on why decisions were made, the demographics behind these decisions, and the outcome. Upon receiving instances of these, the system will learn and adjust to scenarios, gaining capability of prediction in auto dealer sales.

The research held with respect to this study was based on deductive reasoning. This strategy is evident since the study is initialized with a hypothesis, in this case “Applying Supervised Machine Learning Technique to Predict Auto Dealer Car Sales” where a logical conclusion is reached via studying and examining possibilities. The strategy proposed has unfolded into new research with previously existing context. Deductive reasoning is based on deciding on a premises that is generally assumed to be true. As stated in Liew et al. (2018), the process of deductive reasoning is one that generates conclusions which either do or do not necessarily come from a given premises. Deductively valid arguments induce conclusions that ought to be true, provided that the premises are true. In deductive reasoning multiple means such as accepted truths, theories and laws are used to prove the accuracy of the conclusion.

The data analysis will consist of Chi squared, Cross-tabulation, T-test, and Anova. Chi squared, which is commonly used to assess tests of independence when using Cross tabulation. The test of independence assesses whether there is an association between the two variables via witnessing a repetition of reactions expected if the variables were truly objective of one other (Using Chi-Square Statistic in Research - Statistics Solutions, 2021). The T-test being an inferential statistic, it shall be used to determine any substantial difference between two group means, potentially related in certain features (T-Test Definition, 2021). Anova splits observed cumulative variability observed inside a data set into two parts: systematic factors and random factors. Systematic factors have a statistical influence on the data set given, while random factors do not (How Analysis of Variance (ANOVA) Works, 2021).

Graphical user interface, text, application, Word

Description automatically generated

Figure 1

**Ethics**

Prior to initiating the research everyone who is going to be involved is asked to agree to keep what is discussed classified and to acknowledge one another's privacy. It was also noted that confidentiality cannot be completely guaranteed so that participants are aware of the risks involved. All entities, questions or information that will cause physical or emotional harm to participants has been avoided or minimized. Participants were made aware of any potential harms prior to their participation. The researcher has tried their best to remain neutral and unbiased. Personal presumptions or opinions were disregarded so interference with the data collection process is avoided. Ethics in data analysis involves the moral obligations of assembling, safeguarding, and using personally identifiable information and how it affects individuals.

The researcher has created certainty that only himself, the supervisors and examiners have access to all the information collected by not providing admission of the study findings to anyone but the permitted personnel. The design of the survey provides an anonymity throughout both participation as well as data the input of data. Names, surnames, addresses, emails, and personal information have not been taken in this survey. Participants were not involved in any physical harm as the information was gathered online and at the convenience of the participant. Regarding moral harm, this was greatly avoided due to a short, accurate video regarding key technical words was provided to better understand the subject. Moreover, questions in the survey were straight to the point for participants to easily understand what is being asked from them. Ultimately, business harm was omitted as no specific business name or information was used throughout this research.

**Task 6**

Experimental protocols are essential information configurations that support the narrative of the processes by means of which results are produced in experimental research (Giraldo et al.). Experimental design indicates how participants are assigned to the various groups in an experiment. Types of design include repeated measures, independent groups, and matched pairs designs.

An independent variable is a variable that does not depend on any other variables and isn't changed by the other variables in the research. For example, in the research the Supervised Machine Learning Technique is the independent variable. A dependent variable is a variable that does depend on other variables and changes depending on the other variables in the research. For example, in the research the Auto Dealer car sales is the dependent variable. In an experiment, the researcher will be looking for the potential effect on the dependent variable that may be produced by altering the independent variable.

External validity is the limit to which the findings of a study to other situations, people, settings, and measures can be generalized. Essentially, can the findings of a study be applied to a broader context? The aim of scientific research is to produce generalizable knowledge about the real world. Sampling techniques are the process of selecting specific members or a subset of the population in order to make statistical conclusions from them while also approximating characteristics of the whole population. Different sampling methods are widely used by researchers in market research so that they do not need to research the entire population to collect actionable insights. Non-probability sampling was used in the research via convenience for both the participants and the researcher.

Coding is a process of identifying a section in the text or other data items (photograph, image), searching, and identifying concepts and finding relations between them. The prototype applied for the intent of this research via coding, is centred on exhibiting that a machine learning technique can eventually forecast upcoming sales figures.

Research triangulation refers to the use of multiple methods or data sources in qualitative research to develop a comprehensive understanding of phenomena. Triangulation also has been viewed as a qualitative research strategy to test validity through the convergence of information from different sources (Carter et al.). Theory in research is essentially based on research triangulation since it compares multiple sources and contrasts differences in these sources as well as the contributing research which is citing them.

**Task 7**

All the cited elements mentioned were relevant to the study. Agreement with each study as each brought its own particular information and interesting data to input into the research. I agree with all three cited studies as a result of all the studies being closely related to my study and the methods used although different were just as interesting. Another reason as to why I agree with all these studies used different programming methods and machines, ultimately arriving at the same result, which is quite interesting since different ways of tackling the research were undertaken, yet the same outcome was obtained. The concept is another reason why I agree with these papers, they followed the same concept and aims, the motivations were the same and the goals were highly similar.

All these papers are being used in the research. This is a result of the close connection made between the research and each paper individually, as well as the information that can be gathered from each paper can be both followed as well as contrasted with one another.

**Reference list:**

1. “Car Dealers in Malta & Gozo: Cars for Sale |Buy a Car from Auto Dealers.” *Www.yellow.com.mt*, www.yellow.com.mt/auto-dealers/. Accessed 6 Apr. 2022.
2. Raosoft. “Sample Size Calculator by Raosoft, Inc.” *Raosoft.com*, 2009, www.raosoft.com/samplesize.html.
3. Liew, Janice, et al. “Inductive and Deductive Reasoning in Obsessive-Compulsive Disorder.” *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 59, June 2018, pp. 79–86, 10.1016/j.jbtep.2017.12.001.
4. “Using Chi-Square Statistic in Research.” *Statistics Solutions*, www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/using-chi-square-statistic-in-research/.
5. Hayes, Adam. “How T-Tests Work.” *Investopedia*, 8 Oct. 2021, www.investopedia.com/terms/t/t-test.asp.
6. Kenton, Will. “Analysis of Variance - ANOVA.” *Investopedia*, 6 Oct. 2021, www.investopedia.com/terms/a/anova.asp.
7. Giraldo, Olga, et al. “A Guideline for Reporting Experimental Protocols in Life Sciences.” *PeerJ*, vol. 6, 28 May 2018, p. e4795, 10.7717/peerj.4795.
8. Carter, N, et al. “The Use of Triangulation in Qualitative Research.” *Oncology Nursing Forum*, 1 Sept. 2014, pubmed.ncbi.nlm.nih.gov/25158659/.