

T_GAB LCBB5000

by T_gab Lcbb5000 T_gab Lcbb5000

Name: Nicoleta Claudia

Surname: Jeculescu

Student ID: 2011088

LCBB5000

Data Handling&Business

Intelligence

Assessment 2

Submission date: 31-Mar-2022 02:21AM (UTC-0500)

Submission ID: 1797685056

File name: GAB_LCBB5000.docx (123.82K)

Word count: 1715

Data Handling and Business intelligence

**STUDENT ID
STUDENT NAME
MODULE LEADER
WORD COUNT
DATE OF SUBMISSION**

Table of Contents

Part 1	3
Part 2:	4
1. Students liking towards vanilla flavour ice-cream.....	4
2. Male and female students number	4
3. Mean and median participants number for strawberry and chocolate flavour.....	4
b) Most common data mining and text mining methods for business	6
Data mining methods	6
Text mining methods	7
c) Advantages and disadvantages of SPSS and Excel with practical examples	8
Advantages and disadvantages of SPSS	8
Advantages and disadvantages of Excel	9
Reference List	10

Sum of Southern
Sum of Education
Sum of LabourForce
Sum of MoreMales
Sum of YouthUnemployment
Sum of CrimeRate10
Sum of Wage
Sum of BelowWage

	Sum of BelowW age	Sum of CrimeRa te10
	5366	3188.9
	3752	1608.4
3	9118	4797.3

Part 2:

1. Students liking towards vanilla flavour ice-cream

The mode type refers to the most common type of occurrence (or percentage). The issue isn't with the rate of occurrence per se. That's why students should get vanilla, not 1, when asking about the latent variable seen below. Depending on the operation, one or both of these may be involved. A multimodal distribution, or a bimodal distribution, is one of the most common categorizations of this characteristic of the organizational structure (if more than two modes). A uniform distribution is characterized by a lack of a clearly defined method. This photo depicts the ice cream flavour most beloved by the school's students. It's also clear that only 1% of respondents indicated vanilla ice cream over other flavours.

Favourite flavours of Ice-cream

Vanilla: 1

Chocolate: 2

Strawberry: 3

Number of male=0

Number of female=1

Then total=0+1=1

2. Male and female students number

Row Labels	Sum of female
0	0
1	109
Grand Total	109

3. Mean and median participants number for strawberry and chocolate flavour

When discussing the mean, the term "average" is commonly used. It has the highest central tendency. Only ratio or interval variables can be used to calculate the mean. Most data comes from the mean because it can account for the value of each individual observation. The mean responds more quickly to outliers than the median (Venkatesh *et al.*, 2021). Unusually, there is

only one single value that influences the dataset's mean, which is known as a single mean. The mean is often depicted as a variable with a line drawn through it, as shown here. The average score is the total score divided by the number of scores

$$\bar{X} = \frac{\sum X}{N}$$

$$\bar{X} = \frac{\sum X}{N}$$

Figure 2: Formula of mean

As shown in the equation above, the means, X, and N each represent the quantitative significance of a particular case. Summarizing, the sigma says to add up the results. A mean value can only be calculated using addition and division. Nominal and ordinal variables can't be used together in a precise way, as previously stated. For both race and the year are applicable in Excel.

Outliers are less affected by the median, which is a benefit. What is a "outlier" in statistics? An outlier is a value that deviates significantly from the norm. There are no college respondents listed on the excel sheet provided. In contrast to 1 and 2, 2 are male students. When it comes to the grand scheme of things, even the smallest number is counted as a one.

<i>Ice cream</i>	
Mean	1.815
Standard Error	0.060598978
Median	2
Mode	1
Standard Deviation	0.856998971

Sample Variance	0.734447236
	-
Kurtosis	1.544769314
Skewness	0.365980384
Range	2
Minimum	1
Maximum	3
Sum	363
Count	200

b) Most common data mining and text mining methods for business

Data mining methods

There has been a lot of data mining in the world for a long time now. As society produced more and more data, UK corporations coined the term "Big Data" in the 1990s because they couldn't keep up with it (Gola *et al.*, 2018). According to the Bureau of Labor Statistics, workforce for computer science and information research scientists is projected to increase by 15% by 2029.

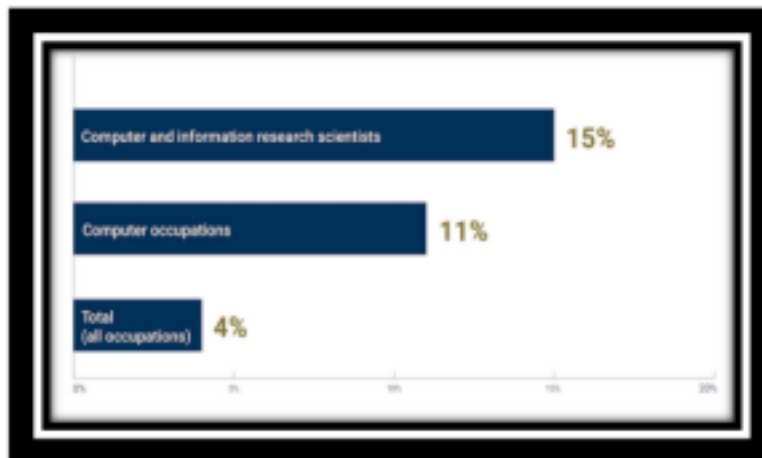


Figure: Projected UK Employment growth of computer and information research scientists, 2019-2029

Clustering

Making groups based on similarities and differences among data points is known as clustering. By seamlessly dividing the data into subgroups, researchers can now get an accurate picture of large populations' habits more quickly and easily than ever before (Rani and Kautish, 2018). Data management can be improved by utilising clustering, for example. Using clustering models,

businesses can figure out what kinds of products they should be selling in order to make more money. Because of this, businesses can better target their goods and services at specific demographics or geographical areas.

Data cleaning

This includes reorganising data, remove unnecessary or corrupted data, and filling in any gaps. The most relevant data can be collected at the end of the process for further analysis. Ninety-five percent of the time, bad data has harmed British businesses (Corrales *et al.*, 2018). As a result of incorrect data analysis, incorrect conclusions are more likely to be drawn than accurate data analysis. Regardless of how well-designed or well-executed a model or algorithm is, if an organization's data is incorrect, inadequate, or damaged, the organisation will suffer.

Machine learning

The term "machine learning" refers to the ability of computers to learn on their own through the use of various methodologies. Because of machine learning, computers are getting "smarter," which is a rising trend in modern technology (Carleo *et al.*, 2019). For example, businesses can discover sales, product use, and purchasing habits quickly using computers, and then use machine learning to develop business plans based on these insights. For a wide range of industries, this is a major concern.

Text mining methods

Text mining is a cutting-edge technology in computer science. As the amount of relevant and valuable unstructured data has increased enormously, it has grown in popularity. A wide variety of text mining techniques are available for extracting useful data from large amounts of text. Sentiment analysis, classification, and entity extraction are a few examples (Antons *et al.*, 2020). Although there is a lot of unstructured data, it is still possible to find patterns, trends, and insights. In large corporations, this system is being pushed out of the way of irrelevant data and replaced with useful data because of its growing acceptance.

Risk management

The primary cause of project failure is a lack of risk management. It is possible to gain complete control over the content of thousands of petabytes of data by using risk management software

that uses text mining technology (Shah *et al.*, 2021). The ability to link data while also having access to accurate data at the right time is also provided by this system.

Cybercrime prevention

Criminals are able to commit crimes online because of the anonymity provided by the internet and the wide variety of communication methods available (Brewer *et al.*, 2019). Businesses and law enforcement agencies can use text mining and anti-crime intelligence software to combat cybercrime.

Business intelligence

Many large corporations have their decision-making processes in place because of this procedure. However, text mining can help a company quickly decipher petabytes of internal and open-source data (Ain *et al.*, 2019). Expert.ai's Intelligence Platform application can monitor a wide range of sources and analyse massive amounts of information.

c) Advantages and disadvantages of SPSS and Excel with practical examples

Advantages and disadvantages of SPSS

SPSS	Advantages	Disadvantages
1.	Statistical Package for Social Sciences (SPSS)	SPSS is prohibitively expensive for students.
2.	There are a variety of statistical analyses available at all levels of complexity.	To make use of the available feature, you'll need to get some extra training.
3.	The interpretation of the results is straightforward.	Similar to Excel, the functionality is more advanced here.
4.	An easy-to-read table of data	It appears to take a long time for a job that needs to be done quickly.
5.	Expandable	It may not be expandable at all times.

Advantages and disadvantages of Excel

Excel	Advantages	Disadvantages
1.	Data is kept in the cloud.	Insecurities and inability to regulate
2.	Perform arithmetic functions	Not impervious to mistakes made by humans
3.	Analytical methods can be used	It is difficult to get your hands on security
4.	Simple to comprehend	Agile business practises that are not appropriate for your organisation
5.	There are free templates available for download.	Cannot be done in a short period of time
6.	Automated coding.	Collaborative work is not supported.
7.	Data that is free of errors and omissions	Hard to put to the test

Reference List

- Ain, N., Vaia, G., DeLone, W.H. and Waheed, M., 2019. Two decades of research on business intelligence system adoption, utilization and success—A systematic literature review. *Decision Support Systems*, 125, p.113113.
- Antons, D., Grünwald, E., Cichy, P. and Salge, T.O., 2020. The application of text mining methods in innovation research: current state, evolution patterns, and development priorities. *R&D Management*, 50(3), pp.329-351.
- Brewer, R., de Vel-Palumbo, M., Hutchings, A., Holt, T., Goldsmith, A. and Maimon, D., 2019. *Cybercrime prevention: Theory and applications*. Springer Nature.
- Carleo, G., Cirac, I., Cranmer, K., Daudet, L., Schuld, M., Tishby, N., Vogt-Maranto, L. and Zdeborová, L., 2019. Machine learning and the physical sciences. *Reviews of Modern Physics*, 91(4), p.045002.
- Corrales, D.C., Corrales, J.C. and Ledezma, A., 2018. How to address the data quality issues in regression models: a guided process for data cleaning. *Symmetry*, 10(4), p.99.
- Gola, J., Britz, D., Staudt, T., Winter, M., Schneider, A.S., Ludovici, M. and Mücklich, F., 2018. Advanced microstructure classification by data mining methods. *Computational Materials Science*, 148, pp.324-335.
- Rani, S. and Kautish, S., 2018, June. Association clustering and time series based data mining in continuous data for diabetes prediction. In *2018 second international conference on intelligent computing and control systems (ICICCS)* (pp. 1209-1214). IEEE.
- Shah, S.M.A., Rasheed, T., Rizwan, K., Bilal, M., Iqbal, H.M., Rasool, N., Toma, S., Marceanu, L.G. and Bobescu, E., 2021. Risk management strategies and therapeutic modalities to tackle COVID-19/SARS-CoV-2. *Journal of Infection and Public Health*, 14(3), pp.331-346.
- Venkatesh, R., Mittal, V. and Tammana, H., 2021, June. Indoor Localization in BLE Using Mean and Median Filtered RSSI Values. In *2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)* (pp. 227-234). IEEE.

4%
SIMILARITY INDEX

PRIMARY SOURCES

2%

INTERNET SOURCES

0%
PUBLICATIONS

2%
STUDENT PAPERS

2%

Submitted to University of Pretoria
Student Paper

1%

www.studymode.com
Internet Source

1%

www.experimentaleconomics.com
Internet Source

PAGE 2

PAGE 3

PAGE 4

Exclude quotes Off Exclude
bibliography On

PAGE 5

Exclude matches Off

T_GAB LCBB5000

PAGE 6

PAGE 7

GRADEMARK REPORT

PAGE 8

FINAL GRADE

PAGE 9

/0

PAGE 10

GENERAL COMMENTS **Instructor**