

Assessment 2 Coursework

Table of Contents

Introduction	3
Part 1	3
Result.....	3
Discussion	6
Conclusion.....	7
Part 2	7
a) Conjunction with SPSS	8
b) Methods of text mining and data mining	8
c) “Advantages and disadvantages of using SPSS and Excel”	9
References	11

Introduction

Data processing refers to the planning of statistics by a PC and this involves the alteration of unprocessed information to machine-readable type, the run of information throughout the remembrance to output devices and CPU, and the transformation or formatting of output. Any utilization of a computer to perform defined activities on data can be involved in data processing. Data processing takes place while information is gathered and translates into working data. This is essential that information dispensation be complete precisely or else it would have an effect on the data productivity and end users. This offered many advantage to organizations namely more dependable and precise, an increase in income and efficiency and improved results. Apart from this, it also helps to reduce costs, easy to storage, reporting, and distribution followed by better production and investigation. The necessity of progression statistics is currently broadly appreciated as well as replicate in roughly each ground of occupation. It is a transnational practice that is incorporated in approximately each meadow of human life. The foremost aim of the report will improve the evocative arithmetical examination regarding the crime set information. The another part of this study will explain the many wide-ranging text mining ways and data mining processes that could be used in the business. This would also illustrate the benefits and limits of using SPSS and excel.

Part 1

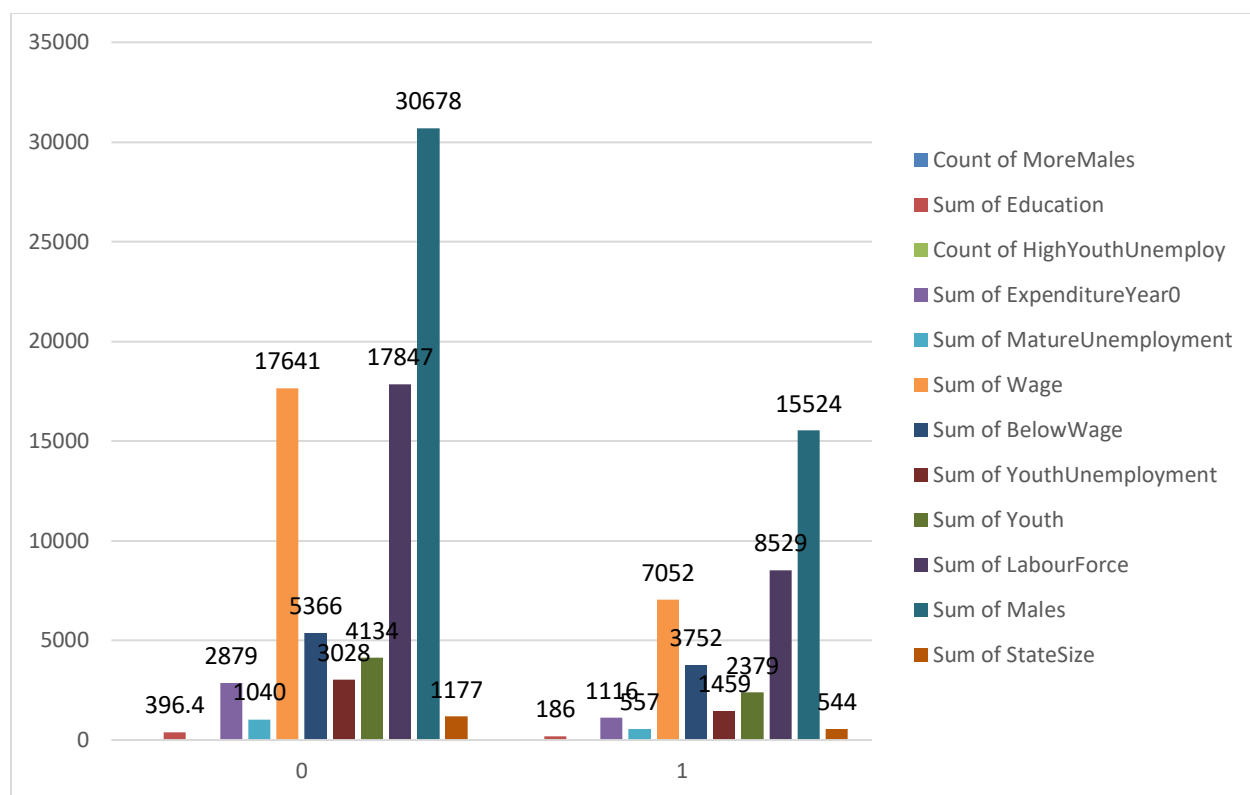
Result

		Northern	Southern
Youth	%	63.47	63.47
Education	%	68.06	68.06
Expenditure Year	%	72.07	72.07
Labour Force	%	67.66	67.66
Males	%	66.40	66.40
More Males	%	80.85	19.15
State Size	%	68.39	68.39
Youth Unemployment	%	67.48	67.48
Mature Unemployment	%	65.12	65.12

High Youth Unemployed	%	68.09	31.91
Wage	%	71.44	71.44
Below Wage	%	58.85	58.85

The above table shows the percentage of youth, education, state size, expenditure, mature employment, youth unemployment, males, labour force, high youth unemployment, below wages and wages of the Northern region and Southern region (Asif et al., 2021). As shown in the table the percentage of youth in the Northern region is 63.47 per cent while in the Southern region it is 63.47 per cent. The percentage of youth in both the Northern region and Southern regions is the same. The number of expenditure years in the Northern region is 72.07 per cent while in the Southern region it is also 72.07 per cent (Crowther et al., 2021). The percentage of males is also similar in both regions with 66.40 per cent. The percentage of labour force is also similar at 67.66 per cent.

The percentage of state size and youth unemployment are also similar at 68.39 and 67.48 respectively. The percentages of more males in the Northern region and Southern region are different (Asif et al., 2021). The percentage of more males in the Northern region and Southern region are 80.85 and 19.15 respectively. The percentage of more make in Northern region is higher than Southern region by 61.7 per cent. There are huge differences in the percentage of "more males" in the Northern region and Southern region (Crowther et al., 2021). The high youth unemployment in the Northern region is 68.09 per cent while in the Southern region it is 31.91 per cent. The percentage of high youth unemployment in the Northern region is higher than in the Southern region at 36.18. The wages and below wages of the Northern and Southern regions are similar.



The above chart shows the number of youth, education, state size, expenditure, mature employment, youth unemployment, males, labour force, high youth unemployment, below wages and wages of the Northern region and the Southern region (Crowther et al., 2021). As shown in the above diagram, the sum of education in the “Northern region and the Southern region” are 396.4 and 186 respectively. The amount of education in the Northern region is higher than the Southern region by 210.4. The sum of expenditure year of the Northern region is 2879 while in the southern region it is 1116. The sum of expenditure year of the Northern region is higher than the Southern region.

Region	(All)											
Row Labels	Count of MoreMales	Sum of Education	Count of HighYouthUnemploy	Sum of ExpenditureYear0	Sum of MatureUnemployment	Sum of Wage	Sum of BelowWage	Sum of YouthUnemployment	Sum of Youth	Sum of LabourForce	Sum of Males	Sum of StateSize
0	31	396.4	31	2879	1040	17641	5366	3028	4134	17847	30678	1177
1	16	186	16	1116	557	7052	3752	1459	2379	8529	15524	544
Grand Total	47	582.4	47	3995	1597	24693	9118	4487	6513	26376	46202	1721

The sum of the mature unemployment in the “*Northern region and the Southern region*” are 1040 and 557 respectively. There is a huge difference seen in the sum of mature unemployment in both regions. The total of wages of the Northern region is 17641 and in the Southern region is 7052 (Ghosal, Bhattacharyya and Majumder, 2020). The difference between the sums of wages of both regions is very high. The sum of below wages in the Northern region and the Southern region are 5366 and 3752 respectively. The Northern region has a higher sum of youth unemployment in comparison to the Southern region which indicates that the Northern region has a higher youth unemployment rate and there are very fewer job opportunities. It is vital for the government of the Northern region to take measures and develop policies to generate jobs and increase employment in the region. A higher rate of unemployment has many negative impacts on the growth and development of the economy as well as society. The sum of youth is also higher in the Northern region by 1755 which shows that this region has more youth than the southern nation.

The Northern region has a higher sum labor force in comparison to the Southern region. The sum of males in the Northern region and the southern region are 30678 and 15524 respectively. It shows that there are more males in the Northern region than in the Southern region (Ghosal, Bhattacharyya and Majumder, 2020). The sum of state size is also high in the Northern region in comparison to the Southern region.

Discussion

From the above analysis, it has been observed that the Northern region has a high sum of education and sum of expenditure year. The sum of mature unemployment and the sum of youth unemployment is also higher in the Northern region which indicates that the region has suffered from the unemployment problems. A higher rate of unemployment is not good for the Northern region and it will affect the growth and development of society as well as the economy (Higuchi and Maehara, 2021). Higher unemployment is one of the major reasons behind the increased crime rate. The higher rate of mature unemployment and youth unemployment leads to an increment in the crime rate. When youth do not have a source of income, they will perform crimes to obtain money to meet their needs. The sum of below wages in the Northern region is also higher which also leads to crime in the society. It indicates that labors receive fewer wages.

The Northern region has also a higher number of youths and labor force than the Southern region which can be a major reason behind the higher rate of unemployment in this region. It is vital for the Northern region to create policies and measures to reduce the unemployment rate and provide more jobs to youth as it will play a vital role in the reduction of the crime rate in this region.

The higher sum of labor force in the Northern region indicates that availability of labor is high in the Northern region. It also indicates that the Southern region has lower availability of labors. The higher availability of labor force is also a major reason behind the higher rate of youth and mature unemployment rate in the Northern region (Higuchi and Maehara, 2021). The number of males is also high in the Northern region than in the southern region which indicates that the population of males is higher than females in the Northern region. The result of the crime data shows that a higher rate of mature and youth unemployment and is higher labor force is the major reason behind the increased rate of crime in the Northern region. Creating job opportunities is one of the effective ways to minimize the crime rate in both regions, especially in the Northern region. By increasing job opportunities the Northern region can minimize the crime rate.

Conclusion

Data processing takes place while information is gathered and translates into working data. This is essential that information dispensation be complete precisely or else it would have an effect on the data productivity and end users. This offered many advantage to organizations namely more dependable and precise, an increase in income and efficiency and improved results. It gives some profit to organizations similarly addition in revenue and efficiency, more dependable and correct and good decisions. The percentages of more males in the Northern region and Southern region are different. The Northern region has a higher sum of youth unemployment in comparison to the Southern region which indicates that the Northern region has a higher youth unemployment rate and there are very fewer job opportunities. A higher rate of unemployment has many negative impacts on the growth and development of the economy as well as society. The sum of mature unemployment and the sum of youth unemployment is also higher in the Northern region which indicates that the region has suffered from unemployment problems. The higher rate of mature unemployment and youth unemployment leads to an increment in the crime rate.

Part 2

1) How many students like vanilla flavor?

Students like vanilla flavour	95
-------------------------------	----

2) How many students are male and female?

Numbers of Male	91
Numbers of Female	109

3) What are the means and median of participants who likes chocolate and strawberry flavor?

Who Like chocolates	47
Who likes strawberry	58

a) Conjunction with SPSS

This part of the study would estimate separate numerical data reached from SPSS. It will also provide knowledge about the expressive numerical border about the existing data. SPSS is usually used by different researchers to calculate hard numerical information. It is also used by the top organization to examine as well as calculate analysis data and mine data to achieve superior results. This would also calculate the mean and median of the particular data set. It will describe methods of data and text mining. The benefits and limits of data utilizing Excel and SPSS will be also discussed in this part.

b) Methods of text mining and data mining

Many data mining process are utilized by the researcher such as sequential patterns, clustering, outer, regression, classification and production. All data mining process the clustering ways is presented is persons of the many used process. As per the study of Nicoară-Farcău, Wang and Luo (2020) cluster method show to collection of data aims those aims are same enclosed by the

same covered by the same cluster that means the aims are the same to the new aims in the similar groups as well as they are quite transformed or different to the aim in different groups.

Clustering analysis is considered as the task of grouping the set of numbers and objectives in the systematic group are more same as each other as compared to those in different groups. Karagiannidis and Themelis (2021) has viewed that cluster examination is defined as the method of identifying groups along with collections in the data or information in such a method that related two aims is highest if they belong to other groups. This process is useful for the discovery of information and data by efficiently organizing field on the internet. This is also utilised in the decreasing of submissions. Operations, profits and sales expenditure are the few variables accessible in the organization context (Purwanto, Asbari and Santoso, 2021). This is utilised for the recognition of regularities and patterns in information.

Weakening is also an efficient data mining procedure and is used to identify and classify the association amid variables because of the existence of the other aspects. This data mining procedure is used to classify the chance of the particular variables. Regression is main a type of modelling and strategies (Karagiannidis and Themelis, 2021). For example, data mining process could be used to project definite expenses, based on other factors like buyers demand, accessibility along with competition (Purwanto, Asbari and Santoso, 2021).

c) “Advantages and disadvantages of using SPSS and Excel”

SPSS stands for “Statistical enclosure for the Social Science” and mention to group of software and application that are mainly used for batched and non-batched statistical information analysis (Nicolotti et al., 2021). SPSS software is described with the simplest information as well as can simply do difficult data manipulation.

The main and significant advantage of SPSS is that it mainly presents time along with attempt. This provides more effective and accurate estimate and avoid estimated and rooming of physical assess. In contrast, main limitation of SPSS is not analyze the big data or information series (Regin, Rajest and Singh, 2021). It is high price software that is not endearing for small associations. It is especially alike to Excel and contain restricted functionality.

Microsoft is launched Excel software in the year 1985. This is mainly utilized for organize information and data series in rows and columns. Around 16384 columns and 104857 rows are

existed in one excel sheet. The high number of data could be collect in this software. The excel sheet users could efficiently arrange information and data by using different charts and other hallucination tools (Nicolotti et al., 2021).

References

- Agrawal, S. and Agrawal, J., 2015. Survey on anomaly detection using data mining techniques. *Procedia Computer Science*, 60, pp.708-713.
- Asif, M.R., Bording, T.S., Barfod, A.S., Grombacher, D.J., Maurya, P.K., Christiansen, A.V., Auken, E. and Larsen, J.J., 2021. Effect of Data Pre-Processing on the Performance of Neural Networks for 1-D Transient Electromagnetic Forward Modeling. *IEEE Access*, 9, pp.34635-34646.
- Chernyshenko, V., 2021, June. Application of Data Mining Methods to Development of Qualification Framework Standards. In *2021 1st International Conference on Technology Enhanced Learning in Higher Education (TELE)* (pp. 209-215). IEEE.
- Crowther, D., Kim, S., Lee, J., Lim, J. and Loewen, S., 2021. Methodological synthesis of cluster analysis in second language research. *Language Learning*, 71(1), pp.99-130.
- Djenouri, Y., Belhadi, A. and Belkebir, R., 2018. Bees swarm optimization guided by data mining techniques for document information retrieval. *Expert Systems with Applications*, 94, pp.126-136.
- Ghosal, S., Bhattacharyya, R. and Majumder, M., 2020. Impact of complete lockdown on total infection and death rates: A hierarchical cluster analysis. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 14(4), pp.707-711.
- Higuchi, A. and Maehara, R., 2021. A factor-cluster analysis profile of consumers. *Journal of Business Research*, 123, pp.70-78.
- Jain, A., Hautier, G., Ong, S.P. and Persson, K., 2016. New opportunities for materials informatics: resources and data mining techniques for uncovering hidden relationships. *Journal of Materials Research*, 31(8), pp.977-994.

- Joseph, S.I.T. and Thanakumar, I., 2019. Survey of data mining algorithm's for intelligent computing system. *Journal of trends in Computer Science and Smart technology (TCSST)*, 1(01), pp.14-24.
- Karagiannidis, P. and Themelis, N., 2021. Data-driven modelling of ship propulsion and the effect of data pre-processing on the prediction of ship fuel consumption and speed loss. *Ocean Engineering*, 222, p.108616.
- Mughal, M.J.H., 2018. Data mining: Web data mining techniques, tools and algorithms: An overview. *International Journal of Advanced Computer Science and Applications*, 9(6).
- Nicoară-Farcău, O., Wang, X. and Luo, X., 2020. Definition of SPSS: we need to speak the same language. *Journal of Hepatology*, 73(2), pp.463-464.
- Nicolotti, L., Hack, J., Herderich, M. and Lloyd, N., 2021. MStractor: R Workflow Package for Enhancing Metabolomics Data Pre-Processing and Visualization. *Metabolites*, 11(8), p.492.
- Pallant, J., 2020. *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*. Routledge.
- Purwanto, A., Asbari, M. and Santoso, T.I., 2021. Analisis Data Penelitian Marketing: Perbandingan Hasil antara Amos, SmartPLS, WarpPLS, dan SPSS Untuk Jumlah Sampel Besar. *Journal of Industrial Engineering & Management Research*, 2(4), pp.216-227.
- Regin, R., Rajest, S.S. and Singh, B., 2021. Spatial Data Mining Methods Databases and Statistics Point of Views. *Innovations in Information and Communication Technology Series*, pp.103-109.