Annals of Clinical and Medical Case Reports

Research Article ISSN 2639-8109 Volume 7

Familiarity with Big Data? A Comparison Between IT Department's Staff Vs Students of the Universities

Nazari E[1], Asgari P[2], Aldaghi T[3] and Tabesh H[1*]

[1]Department of Medical Informatics, faculty of medicine Mashhad University of Medical Science, Mashhad, Iran

[2]Department of Health and Information Technology, Mashhad University of Medicine Science, Mashhad, Iran

[3]Department of Industrial Engineering, Eqbal Lahouri Institute of Higher Education, Mashhad, Iran

[*]Corresponding author:

Hamed Tabesh,
Department of Medical Informatics, faculty of medicine, Mashhad University of Medical Science, Mashhad, Iran,
Tel: +98 51 38002536; Fax: +98 51 38002445,
E-mail: rp666600@gmail.com

Received: 16 July 2021
Accepted: 02 Aug 2021
Published: 06 Aug 2021

Citation:

Tabesh H, Familiarity with Big Data? A Comparison Between IT Department's Staff Vs Students of the Universities. Ann Clin Med Case Rep. 2021; V7(3): 1-6

Authors Contributions:

Nazari E, Asgari P, Aldaghi T, Tabesh H and these authors are contributed equally to this article

Abbreviation:

IOT: Internet of Things

Keywords:

Big Data; Data analysis; Data collection; Decision Making; Organizational

1. Abstract

1.1. Introduction: The rapid development of technology in recent decades has led to the production of a huge amount of data. One of the challenges of these analyses is the lack of specialized expertise and knowledge in this area. The purpose of this study was to compare the familiarity of IT staff and students with big data analyzes at various universities and organizations.

1.2. Materials and method: This analytical study was conducted on IT units' staff and students of different organizations and universities in Mashhad, Iran. A questionnaire was designed. The participants were 265 IT units' staff and students of different organizations, completing the designed questionnaire. Participants' opinion was evaluated using two descriptive and analytical approaches.

1.3. Results: Scores earned by students and staff were 2.66 ± 1.13 and 2.28 ± 1.21 respectively that $p = .012$ represented a significant correlation between the level of knowledge of students and staff. In other words, the level of knowledge of staff about big data was more than the level of knowledge of the students. The correlation of each of the variables was not significant

1.4. Conclusions: In general, the level of knowledge in analyzing big data in different groups of people was at a low level that implementing measures such as holding training courses in this field seems necessary. With the score of the Big Data Analysis Knowledge. But there was a significant correlation between experience and gender with the knowledge scores.

2. Introduction

Today with the emergence of various technologies such as SmartPhone, IoT and the rapid development of the Internet, data are produced in all industries. These data are called Big Data. This type of data could not be managed alone and analyzed due to having such characteristics as high volume, and the diversity that it's an analysis using classic methods impossible and require the use of related and appropriate arrangements. Review of Big Data is known for making the right decisions as to the Big Data analysis [2,3].

Today, Big Data has become a hot topic in all industries and academic environments. These analyses have many benefits such as reducing costs, discovering useful data patterns, extracting essential features, summarizing and sharing data for critical and vital decisions making and used them in essential research fields [1, 5, 7, 9]. But these analyses have challenges that have created barriers. Some of these challenges were the lack of quality and enough data, the lack of equipment and infrastructures necessary for analysis, lack of familiarity with the techniques needed and lack of expertise.

The most critical challenge, for example, lack of sufficient knowledge and expertise in this field that affects the benefits [4, 6, 8]. There are these challenges in Iran too, and it is necessary to resolve these challenges by applying measures such as informing, holding training courses, and conferences.

Therefore, we examined the importance of Big Data analysis to pay attention to the purpose of this study was to compare the level of familiarity of IT Units ' Staff and students with Big Data Analyze in Mashhad. The familiarity and awareness of students in the research stage. And staff who has a background and experience in similar environments and software applications to evaluate them.

3. Material and Methods

We conducted this cross-sectional study on IT units' staff and students of different organizations and universities in Mashhad, Iran.

Mashhad is the largest city in the eastern of Iran with about 3 million people, located on the border with Afghanistan and Turkmenistan and on the way of the Silk Road with more than 70 public organizations and private companies.

Today, information technology (IT) is vitally for organizations and necessary for the organization's improvement, in each organization, at least one IT expert provides services. In organizations, essential tasks are assigned to IT staff, including managing existing networks, software, and hardware; maintaining existing software; developing and upgrading software; monitoring databases.

People who have this job educated in software, hardware, network, and information technology. There are two major state universities in Mashhad, Ferdowsi University and the University of Medical Sciences. The former host's students from different fields of study, including engineering and basic sciences. The latter host's students from medical fields of study such as medicine and biology. We Evaluated students' knowledge and awareness of Big Data analysis in different fields of study in Mashhad universities, a questionnaire developed. Assess the level of knowledge and awareness of IT staff of different organizations in Mashhad with Big Data analyzes and designed a questionnaire. The questionnaire consisted of closed-ended items. The original items of the questionnaires were prepared based on reviewing the texts published in PubMed, google scholar, science direct, and EMBASE databases and then designed

According to the Delphi method with the attendance of ten specialists in different disciplines (medical informatics, Biostatistics, and computer). This questionnaire contains five items concerned with one's knowledge of how to analyze Big Data. The relevant items in Table 1:

Table 1: Questionnaire items

Questions	Description
Knowledge Questions	
QK1	**What is the definition of Big Data?**
QK2	**What are the hardware requirements for analysis?**
QK3	**What is the focus of Big Data analyzes?**
QK4	**What are the advantages of Big Data analyzes?**
QK5	**What are the disadvantages of Big Data analyzes?**

Ten experts confirmed the reliability and validity of the questionnaire as a panel of the validity and Cronbach's alpha was estimated to test reliability and Estimated at 81% and 73% for staff and students. Then, the required data were collected, and it was made sure that all questionnaires were completed. After that the questionnaires were provided to 30 public and private organizations and present research attempted to include students of different fields of study. These included the following within Medicine, Computer Engineering, Pharmacy, and Basic Sciences. From two major universities. The inclusion criterion for the selection of organizations was as follows: having independent IT units within their organizational chart and having staff with experience in working with different software. This study included the social security organization, hospitals, transportation, organization, and governorate. These organizations provide services in the field of health care, transfer management, supervision of other organizations. We Collected data from these organizations and the participants completed all the questionnaire items Out of 150 questionnaires sent by post to the IT staff working with these organizations, 123 questionnaires from among the initial 150 distributed questionnaires, 142 were completed and used T-test and ANOVA and GLM for the selected variables. For data entry and analysis, we used SPSS21 and Excell-2007.

4. Statistical Results

Individual characteristics of the participants were shown in Table 2:

As shown the amount of knowledge in the student group at age 25 – 44 year and the age range 35-44 year with the group of staff have had the highest score.

Also, the level of knowledge in the student group of female and male in the group of staff was higher. The work experience in the group of staff was also higher. Most of the staff have masters and most PhD students. In both groups, they won the most points at 40 points (Table 3).

As can be seen, there was a significant correlation between the mean hours of scientific study and the mean hours of the non - scientific study of students and staff. Investigating the relationship of knowledge score with each variable was studied individually (simple analysis) and once analyzed in the GLM modelling (multiple analyses). These results are as follows in Table 4:

Table 2: Individual characteristics of the participants

Variables	Items	Frequency (percentage) of student)n=142)	Frequency (percentage) of staff (n=123)
Age	18-24year	7(4.9%)	0(0%)
	25-34 year	62(43.7%)	19(15.4%)
	35-44 year	63(44.4%)	61(49.6%)
	45-54 year	10(7.0%)	36(29.3%)
	55-64 year	0(.0%)	7(5.7%)
Sex	Male	58(40.8%)	80(66.7%)
	Female	84(59.2%)	40(33.3%)
	Missing	0(0%)	3(0.02%)
Experience history	<=1 year	113(79.6%)	28(22.8%)
	>1 year	29(.20%)	95(0.67%)
Degree	BA	29(20.4%)	57(48.3%)
	MA	39(27.5%)	61(51.7%)
	Professional doctorate	45(31.7%)	0(.0%)
	PhD	29(20.4%)	0(.0%)
	Missing	0(0%)	5(0.4%)
Score	0	6(4.2%)	0(0%)
	20	37(26.1%)	18(14.6%)
	40	39(27.5%)	43(35.0%)
	60	34(23.9%)	32(26.0%)
	80	22(15.5%)	23(18.7%)
	100	4(2.8%)	7(5.7%)

Table 3: Mean and standard deviation of participants' hours of scientific and non-scientific studies across fields of study in two groups of students and staff.

Non-scientific hours studying	Scientific-hours studying	Participant
Mean ± SD	Mean ± SD	
3.13±0.92	3.79±0.59	Student(n=142)
2.70±0.82	3.13±0.89	Staff(n=123)
0.001	<0.001	p-value

Table 4: Comparison of the mean scores of knowledge of staff and students in terms of age group, gender, background, degree, field, number of hours of scientific and non-scientific study

		scores of knowledge of participant						
Variables	Items	Staff (n=123)			Students (n=143)			
		n_i	Mean	SD	n_i	Mean	SD	p-value¹
Age	18-24 year	0			7	2.2857	1.38013	-
	25-34 year	19	2.4737	1.07333	62	2.0806	1.20516	0.206
	35-44 year	61	2.5902	1.16013	63	2.4286	1.201	0.448
	45-54 year	36	2.8889	1.08963	10	2.7	1.1595	0.635
	55-64 year		2.5714	0.9759				--
p-value		0.511			0.284			
Sex								p-value³
	Male	80	2.6	1.0626	58	2.2069	1.16617	0.041
	Female	40	2.775	1.25038	84	2.3452	1.24662	0.075
p-value⁴		0.425			0.506			
experience history								Pvalue⁵
	<=1year	28	2.4643	0.99934	113	2.1947	1.17912	0.267
	>1 year	95	2.716	1.145	29	2.667	1.278	0.809
p-value⁶		0.172			0.14			

degree								Pvalue[7]
	BA	57	2.5965	1.09967	29	2.1034	1.26335	0.065
	MA	61	2.7213	1.15659	39	2.0256	1.0879	0.003
	Professional doctorate	0			45	2.6	1.21356	
	PhD	0			29	2.3448	1.2614	
p-value[8]		550			0.135			
scientific_hours_studying								Pvalue[9]
	<1	18			11			0.976
	3-Jan	35	2.7143	0.0452	12	2.3333	1.30268	0.312
	5-Mar	20	2.7	1.41793	3	2.3333	1.52753	0.683
	>5	49	2.6735	1.1436	46	2.2586	1.22383	0.044
Pvalue[10]		0.987			0.976			
Non-scientific_hours_studying								Pvalue[11]
	<1	44	2.568	1.246	33	2.879	1.166	0.27
	3-Jan	46	2.6957	1.19014	35	2.1714	1.07062	0.044
	5-Mar	22	2.7273	1.03196	15	2	1.25357	0.062
	>5	20	2.15	0.74516	48	2.2083	1.25407	0.813
Pvalue[12]		0.128			0.839			

Note that:

1. The independent sample was used to compare the mean scores of knowledge in the analysis of student data and staff data in each age group

2. We used One-way ANOVA test to compare the mean scores of knowledge of big data analysis of each group in different age groups

3-We used an independent test sample to compare the mean score of big data analysis of students and staff in any gender.

4. The independent test sample was used to compare the mean score of big data analysis of each of the groups in two different genders.

5. The independent test sample was used to compare the mean score in the big data analysis of students and staff in each category of work experience of the.

6. The independent tests were used to compare the mean scores of knowledge in the analysis of student data and staff data at each degree.

7. One-way ANOVA was used to compare the mean score in the big data analysis of each of the groups in different sections.

8. The independent test sample was used to compare the mean score in the big data analysis of students and staff in each row from the scientific study hours

9. One-way ANOVA was used to compare the mean score of big data analysis of each of the groups in the scientific study hours.

10. The independent test sample was used to compare the mean score in the big data analysis of students and staff in each row of non-scientific study.

11. One-way ANOVA was used to compare the average score of big data analysis of each of the groups in non-scientific study group's hours.

As seen in Table 4, a significant difference between the averages score of staff and students in any of the levels of age group. Also, there was no significant difference between the mean scores of the staff's knowledge at different levels of age group. The comparison of the average score of knowledge of both groups of staff and students adjusted with stratification. However, there was no significant difference between the mean scores of knowledge in these two groups. The higher average level of knowledge of the male staff is significant male students. The average of the knowledge score in the staff of the master's degree was also observed compared to the master's students. Significant differences in staff knowledge score with more than 5 hours of scientific study in comparison with the knowledge of the students with more than 5 hours of scientific study are remarkable.

5. General Linear Model Results

We used the General Linear Model, to investigate the effect of different variables, such as age and gender. On the level of knowledge of individuals about the big data. Table 5 that none of the variables studied individually did not affect the points obtained in the field of knowledge. The interaction between gender and work experience were an influential factor and could not present, no-significant for related main effects interpretation for them. Assessment of the rating according to the age group in the two groups is shown in the Figure 1.

According to the Figure 1, average score in the age group of 35 to 44 years old among men students less than other age groups, although there was no significant difference seems to have not shown the statistical tests. The average score of knowledge in other age groups of women was no much difference in students and staff.

Table 5: Investigating the relationship between the variable group (staff, students), gender, age, work experience (less than one year and more than one year) degree, the number of scientific and non-scientific study hours daily with a score of big data analysis.

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
		86.461	51	1.695	1.280	.143
		24.911	1	24.911	18.809	.000
scientific_hours_studying		.520	1	.520	.393	.532
non_scientific_hours_studying		.015	1	.015	.012	.915
group		3.635	1	3.635	2.745	.100
age		.407	2	.203	.153	.858
sex		1.095	1	1.095	.827	.365
experience history		3.600	3	1.200	.906	.441
degree		7.598	3	2.533	1.912	.132
group * experience history		.072	1	.072	.054	.817
group * degree		.892	1	.892	.673	.414
age * sex		.338	2	.169	.128	.880
age * experience history		5.093	4	1.273	.961	.432
age * degree		11.008	5	2.202	1.662	.150
sex * experience history		8.579	2	4.289	3.239	.043
sex * degree		9.245	3	3.082	2.327	.079
experience history * degree		2.248	4	.562	.424	.791
group * age * sex		.000	0	.	.	.
group * age * experience history		.000	0	.	.	.

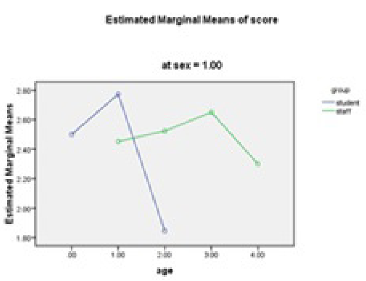

Figure 1: Evaluated scores by age group in two groups

5.1. Knowledge scores in terms of the level of work experience in two groups

According to Figure 2, the average score in people with work experience of one to three is less than the other age groups.

Although statistical tests did not show any significant difference. The average of knowledge in other groups of work experience was no difference in students and staff.

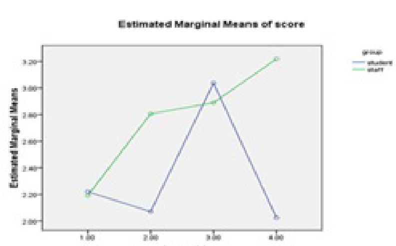

Figure 2: evaluate of knowledge in terms of the level of work experience in two groups

5.2. Knowledge scores in terms of gender and level of experience in the two groups

According to sinister Figure 3, the average score in people who have had work experiences of between one and three and were men was less than the rest; it seems that although statistical tests did not show significant differences. The average scores of knowledge in other groups of history and gender were no much difference in students and staff.

According to right Figure 3, the average score in people who have a history of between 3 and 5 and have been female. The average score of knowledge in other groups of history and gender was no much difference in students and staff.

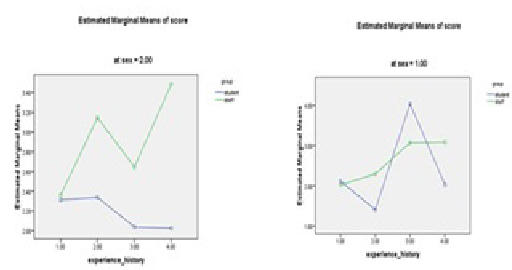

Figure 3: evaluate of knowledge in terms of gender and level of experience in the two groups

5.3. Knowledge by gender and age group in two groups

According to sinister Figure 4, the average score in people who have been ages 34-25 and female have been less than others.

Statistical tests did not show a significant difference, in any case. The average score of knowledge in other age groups and gender was no much difference in students and staff.

According to right Figure 4, the average score for people aged between 54 and 35 was less than the rest. It seems that although statistical tests do not show significant differences. The average score of knowledge in other age groups and gender was no much difference in students and staff.

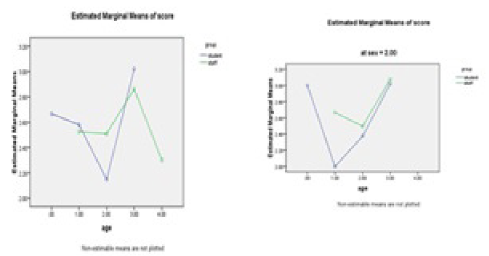

Figure 4: Evaluate of knowledge by gender and age group in two groups

6. Discussion

The rapid development of technology in recent decades has led to the production of the amount of data. These data were called big data. The familiarity with Big Data analyzes were of great importance due to many benefits, including cost and error reduction and make decisions. This study examined the challenges of lack of knowledge and expertise in the two groups of students and staff as a barrier to exploiting the benefits. On average, the staff's knowledge about the concepts of Big Data were higher than the students. In assessing the relationship between the level of knowledge of people in age groups and background, significant reported. The means that depending on the gender and experience increases the amount of knowledge. Because it seems those males more than the female is more interested in software engineering, and management jobs and their work experience was a reason for gaining a higher knowledge score has been in them. It seems to hold training courses, conferences, congress, recruiting specialist staff could overcome challenges. Scientists needed the concepts in the Big Data area, familiarized the IT staff of organizations and companies. Future studies, students, doctors and other fields of the research and the challenges of the Big Data analysis should be investigated from their viewpoint because students can provide a base for familiarizing and applying useful analysis by performing new research in this area. In other businesses, finding out the extent of their familiarity with the Big Data analyzes could be useful in applying managerial and advertising policies. Big Data could have a constructive role in all industries, and today, these analyze have become widespread in most industries and businesses. Because given the growing trend of data production, big data analysis in the coming years would become a requirement for all industries and areas.

7. Acknowledgments

The present study is the result of research project approved by the vice chancellery for research of Mashhad University of Medical Sciences (grant number 961731).

References

1. Acharjya DP & Ahmed K. A survey on big data analytics: Challenges, open research issues and tools. International Journal of Advanced Computer Science and Applications. 2016; 7(2): 511-518.
2. Archenaa J & Anita EM. A survey of big data analytics in healthcare and government. Procedia Computer Science. 2015; 50: 408-413.
3. Belle A, Thiagarajan R, Soroushmehr S, Navidi F, Beard DA & Najarian K. Big data analytics in healthcare. BioMed Research International, 2015; 2015: 1-16.
4. Bossé É & Solaiman B. Information fusion and analytics for big data and IoT. Artech House. 2016.
5. Doi K. Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. Computerized Medical Imaging and Graphics. 2007; 31(4-5): 198-211.
6. Hermon R & Williams PA. Big data in healthcare: What is it used for?. 2014.
7. Kim GH, Trimi S & Chung JH. Big-data applications in the government sector. Communications of the ACM. 2014; 57(3): 78-85.
8. Kuo MH, Sahama T, Kushniruk AW, Borycki EM & Grunwell DK. Health big data analytics: Current perspectives, challenges and potential solutions. International Journal of Big Data Intelligence. 2014; 1(1-2); 114-126.
9. Raj P, Raman A, Nagaraj D & Duggirala S. Big Data Analytics for Healthcare. InHigh-Performance Big-Data Analytics 2015 (pp. 391-424). Springer, Cham. Doi K. Computer-Aided Diagnosis in Medical Imaging: Historical Review, Current Status and Future Potential. Computerized Medical Imaging and Graphics. 2007; 31(4-5): 198-211.