University Students’ Readiness for Job Opportunities in Big Data Analytics

Sarah Yusoff 1,*, Nur Hidayah Md Noh 1 and Norulhidayah Isa 1

1 Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Cawangan Terengganu, Kampus Kuala Terengganu, 21080 Terengganu, Malaysia

*sarahyusoff@uitm.edu.my

Abstract. This study aims to explore the students’ level of readiness in taking up job opportunities in big data analytics and determine the contributing factors to students’ readiness. In addition, the crucial factors that need to be resolved are identified. This job field requires some significant criteria such as, willing to work as a team, self-effort, and specialised skills such as data visualisations and data storytelling, big data analysis, and basic knowledge on tools for big data analytics. Intellipaat.com, a platform that offers various professional online training courses, has ranked position in big data analytics and data science as the highest paying jobs in 2019. However, from 2019 onwards, Malaysia has been predicted to suffer a shortfall of data analysis professionals of up to 7,000-15,000. Our educational institutions are being encouraged to create more graduates to meet this need. The question arises on whether students are prepared and willing to work in this sector once they graduate. An online survey was constructed and distributed to all UiTM students enrolled in various bachelor’s degrees and master’s programmes. One hundred and thirty-nine students participated in this survey. A graphical tool for data tabulation was presented using a box-and-whisker plot. Additionally, correlation analysis and multiple regression were used to determine the relationship and factors that can contribute the students’ readiness for job opportunities in big data analytics. The results from the box-and-whisker plot have discovered an excellent sign of students’ readiness towards job opportunities in big data analytics. Correlation analyses has shown a weak to moderate relationship among factors and multiple linear regression analyses revealed the data visualisation including storytelling skill (DVSS) and teamwork (TW) have significantly given some impacts on the students’ opportunity in big data analytics career. The results of this study are expected to provide insights into students’ readiness for job opportunities in big data analytics.

Keywords: Big data analytics, data scientist, box plot, higher education, student readiness

1. Introduction

Workers skilled in data science big data analytics are on demand across industries in the world. In the United States (US), the high demand has resulted in a countrywide shortage of 151,717 professionals with data science skills [1]. Previously there was a surplus of data scientist big data analysts, and the field was not considered an advantageous field in terms of job demand. In 2015, a US employer who wanted to hire a data scientist did not face any difficulty looking for someone with the skill. Employers had plenty of candidate options to select from when hiring. Employers often seek data scientists and business analysts who have excellent communication skills and business acumen. Thus, if the candidates have excellent communication skills, they are one step closer to being good data scientists. However, in 2021...
the shortage of data science skills was widespread in every large city in the US. It is estimated that data scientists in the US make an average of $120,140 per year [2]. Considering the current currency exchange rate, this is about RM500,000 per year or about RM40,000 per month. Fuelled by big data and Artificial Intelligence, the demand for data science skills is growing exponentially. However, the supply of skilled applicants increases at a slower pace [1].

In Malaysia, a similar situation of high demand for data scientists is predicted in the upcoming years. The CEO of Malaysian Digital Economy Corp (MDEC) has predicted that by 2025, the information technology (IT) industry would require one million specialised workers due to the industry rampant growth. The growing digital economy wave has seen Malaysia facing a shortage of 7,000 to 15,000 IT professionals since 2019. Many IT companies have opted to hire foreign staff to cope with the growth, especially in big data analytics and artificial intelligence (AI), because they were equipped with the required skills. It would be an advantage to the companies since no more training was needed, and these employees might contribute to the company instantly [3].

With the emerging requirements of data scientists in our country, academic institutions in Malaysia are encouraged to produce more related courses in producing more potential graduates to meet the needs of data scientists in future. Universiti Teknologi MARA (UiTM) is one of the universities working hard to accept this challenge by offering some courses related to data analysis. One of them is a postgraduate programme which is Master of Data Science. The main aim of this programme is to develop knowledgeable and technically skilled data professionals who may later be the leaders to fuel the data analytic sectors in the country. Other than UiTM, the data science courses are also available at Universiti Malaya (UM) and Universiti Sains Malaysia (USM). The data scientist profession is becoming more popular in recent years and is the sexiest career in the 21st century [4]. The main attraction of this career is the high salary offered. Malaysia has many students are local universities who are currently enrolling and studying the related courses. However, are all the students ready to work in this field once they graduate? Do they know the skills required to become an excellent data scientist? A recent article from towardsdatascience.com; emphasises that mastering data analytics is a foundation to become a data scientist. Other than that, this job also requires creativity and critical thinking, knowledge in mathematics and statistics, programming skills, especially in Phyton and R, data analysis and visualization, machine and deep learnings, databases, and education [5,6].

To meet the shortage of data scientists, universities in Malaysia play a significant role in motivating and preparing their students to have appropriate knowledge and skills for the job. Being aware of the rising demand for data scientists is the first crucial factor that motivates the students to take up a career after graduating. The issue is, are the students, specifically UiTM students, aware and ready for the job opportunities in big data analytics? To fulfil the predicted shortage of data scientists, the government must know the approximate percentage of graduates who are fully prepared for the job. Thus, research about student-readiness towards the career as data scientist is essential in preparing the government and the universities to suitably upgrade their students and facilities.

2. Literature Review

Workers skilled in There are several factors that enable students to be ready for job opportunities in the field of big data analytics. Recognising the requisite descriptions of employees sought by employers hiring data scientists is one step closer in preparing students for the job. These criteria are also important in developing the instrument used in this research. Big data analytics is the process of examining large data sets to uncover hidden patterns. The key skills needed in data scientist are analytics, as presented by [7] on four basic classes of job roles and required skillset: (1) business analyst, (2) data scientists, (3) developers, and (4) system managers. A data scientist can make large volume of data sensible by collecting raw data, manipulating it, and drawing conclusions from it.

Data science uses various scientific tools, formulas, algorithms, and processes to understand and extract useful patterns from a voluminous amount of data. Industries depend on big data for their strategic decision-making processes. In terms of skills, data scientists are expected to have experience in tools like Hive, BigQuery, AWS, Spark and Hadoop, as well as training in statistical modelling, machine learning and programming. Data scientists are also expected to be fluent in at least one programming language, Python or R [8,9]. Along with statistical and machine learning modelling using Python or R,
skills in SQL, NoSQL databases, Apache Spark and relational database management systems (RDBMS) are also in demand [10]. SQL is a standard skill for most data analytics and data science openings, and most corporate data still sits on RDBMS legacy systems [11].

The ability to work in a team is another important criterion of a data scientist. It is important to note that data scientists do not take the combined skills of data engineer, machine learning expert and business executive. To get total contribution from data scientists, organisations must first hire a machine learning expert who can use R, Python or SAS and understand which algorithms to be applied to different kinds of situations. Then, team up the person with a data engineer and an in-house business executive [11].

A data scientist does not work alone. As mentioned earlier, this career requires a data scientist to collaborate with other people from different backgrounds and skillsets. Hence, good communication skills are required. The renewal of methods utilised within Business Intelligence and Analytics in Big Data requires developing new interdisciplinary competencies spanning from IT skills to business domain knowledge and communication skills [12]. While big data is currently the leading topic of interest within Business Intelligence and Business Analytics, survey results suggest that one of the big data skills desired by employers when making hiring decisions is communication [13]. A senior data scientist who manages big data projects with vision and leadership for data-driven business and projects communicates with business leaders, data analysts, and users. Among their primary strengths, a senior data scientist possesses communication skills [14–16]. Author [4] highlights the need for data scientists to communicate in the language that all their stakeholders understand and demonstrate the special skills involved in storytelling with data, whether verbally, visually, or ideally both. He also suggested a data scientist must be able to find a story in a data set and provide a coherent narrative about a critical data insight and must be able to communicate with numbers; visually and verbally.

Another criterion to become a data scientist is a correct attitude towards the data. When data scientists deal with big data, they should be concerned with the security and privacy of the data [17]. Data scientists work with data and appreciate data itself as a first-class product [10].

To fill the gap highlighted above, this research was conducted to evaluate students’ readiness in pursuing a career as data scientists. From the review conducted, factors that influence students’ readiness are knowledge and skills in big analytics, teamwork, and data visualisation skills. With this information, any matter or problem could be identified and solved. Whenever needed, the government and universities may also cooperate to upgrade the syllabus and facilities.

3. Data Collection Method

The scope of this research involved undergraduate and postgraduate students from the Faculty of Computer Science and Mathematics in UiTM. These students were chosen because they were the ones who fit the basic criteria that would enable them to fill job vacancies as data scientists. This research was designed as a sample survey using a self-developed questionnaire as the research instrument.

The instrument was a structured questionnaire consisting of several sections on demography and six factors of student readiness towards careers in big data analytics. This study has been designed with seven constructs which are:

1. teamwork,
2. efforts,
3. data visualization and storytelling skill,
4. skills for big data analytics,
5. campus readiness,
6. student awareness, and
7. tools for big data analytics.

Each construct in Table 1 was assessed by using items in table 2. For example, in effort construct, four items were used to measure student effort. One of the items was, “I am interested to have a career in big data analytics”. Higher rating (%) indicates higher ability to have a career in big data analytics. All twenty-nine items asked in this survey were in positive direction.

A section was also allocated to study the campus readiness for big data analytics in terms of the campus infrastructure, facilitators, and curriculum. Questionnaire items were measured using Ruler and
Options scale [18]. An online questionnaire was distributed to several UiTM branches which had undergraduate and postgraduate computer science students and had received 139 responses. Respondents comprised all bachelor’s degree students from Faculty of Computer and Mathematical Sciences across UiTM branches and students who were doing Master of Data Science at UiTM Shah Alam with ages ranging between 20 to 27 years old. The data collected was analysed with the help of the Statistical Package for Social Sciences (SPSS). The descriptive analysis such as Box and Whiskers plot were conducted. Then, reliability test using Cronbach's alpha was executed to measure the internal consistency of the items for each construct. Correlation analysis and multiple linear regression were conducted to determine the relationship between factors and factors that contribute to students' readiness for job opportunities in big data analytics based on student effort criteria.

4. Results and Discussions

4.1. Demographic Analysis

The demographic data were illustrated in figures 1, 2, and 3. In figure 1, 83.5% of the responses were female students while the remaining 16.5% were male students. Most respondents were from CS241 programme (Bachelor of Statistics) at 83.5% (in figure 2). Other programmes involved in this study were CS230 (Bachelor of Computer Science), CS249 (Bachelor of Science (Hons) Mathematics), CS253 (Bachelor of Computer Science (Hons) Multimedia Computing), and CS779 (Master of Data Science) students from different UiTM campuses. Most of the participants in this study were 22 years old and were already in their final semester.

![Figure 1. Students’ Gender](image1)

![Figure 2. Students’ Enrolled Programmes](image2)

![Figure 3. Percentage of Students from different Campuses](image3)
4.2. Reliability Test

Twenty-nine items from seven constructs in Table 1, i.e., teamwork, effort, data visualisation and storytelling skill, skills for big data analytics, campus readiness, student awareness, and tools for big data analytics were designated as the initial instrument to measure student readiness towards job opportunity in big data analytics. The reliability of all seven parameters was high (all at or above Cronbach's Alpha ≥ 0.863). The reliability coefficient (Cronbach's alpha) for student awareness was 0.907, indicating that the instrument items were highly internally accurate. The Cronbach's Alpha coefficient for each construct was shown in Table 1. The result suggested that the internal consistency of the student readiness scale was high.

| Constructs                                      | Number of items | Cronbach's Alpha |
|------------------------------------------------|-----------------|------------------|
| Teamwork (TW)                                  | 5               | 0.903            |
| Effort (EFF)                                    | 4               | 0.904            |
| Data Visualisation and Storytelling Skill (DVSS)| 5               | 0.902            |
| Skills for Big Data Analytics (SBDA)            | 4               | 0.870            |
| Campus Readiness (CR)                          | 4               | 0.867            |
| Student Awareness (SA)                         | 4               | 0.907            |
| Tools for Big Data Analytics (TBDA)             | 3               | 0.863            |
| Total                                           | 29              |                  |

Table 2 shows the minimum and maximum values, mean, standard deviation and skewness for all items in each criterion. It is observed from the skewness values that the distributions of readiness scores for each construct vary depending on the items. In general, skewness values around 0 indicate the presence of a normal distribution. For twenty-seven items in Table 2, the data distribution has been left-skewed. On the other hand, the distribution of two items in the tools for big data analytics construct are positively skewed.

| Constructs                                      | Scale items                                                                 | Min | Max | Mean | Standard Deviation | Skewness |
|------------------------------------------------|-----------------------------------------------------------------------------|-----|-----|------|--------------------|----------|
| Teamwork                                       | Teamwork is essential in big data analytics careers.                        | 17  | 100 | 82.8 | 18.482              | -1.279   |
|                                                 | I enjoy working in a group.                                                 | 14  | 100 | 80.2 | 19.052              | -1.263   |
|                                                 | I am a good listener when working in a group.                               | 22  | 100 | 82.2 | 15.977              | -1.174   |
|                                                 | I know that a data scientist must collaborate with workers from different departments. | 15  | 100 | 81.1 | 18.835              | -1.412   |
|                                                 | I know how to maintain good relationship with other team members.           | 16  | 100 | 79.5 | 18.097              | -1.058   |
| Effort                                         | I am interested to have a career in big data analytics.                     | 0   | 100 | 76.9 | 21.066              | -1.118   |
|                                                 | I am willing to upgrade myself in data communication skill.                 | 5   | 100 | 81.9 | 18.172              | -1.458   |
|                                                 | If I have the money, I am willing to spend on joining workshops to upgrade my data analytical skills. | 0   | 100 | 76.3 | 22.278              | -0.983   |
|                                                 | If I have the time, I am willing to upgrade my knowledge in big data analytic skills. | 0   | 100 | 79.1 | 20.085              | -1.395   |
| Data Visualisation and Storytelling Skill       | I find it easy to interpret results from data analysis that I have conducted. | 14  | 100 | 71.2 | 18.234              | -0.451   |
| and                                             | I enjoy explaining results from data analysis to people.                   | 6   | 100 | 71.6 | 20.274              | -0.691   |
### Storytelling Skill

| Item                                                                 | Score | Readiness Score | Z-score |
|----------------------------------------------------------------------|-------|----------------|---------|
| I know how to present research findings visually using graphs, tables and charts. | 12    | 75.5           | 19.86   |-0.967  |
| I am not afraid of being questioned by the audience during the presentation. | 8     | 60.6           | 21.565  | -0.168  |
| I enjoy describing data, how data were collected, and analysis done on the data. | 14    | 71.9           | 19.068  |-0.679  |

### Skills for Big Data Analytics

| Item                                                                 | Score | Readiness Score | Z-score |
|----------------------------------------------------------------------|-------|----------------|---------|
| I am good in Data Visualization skills.                             | 0     | 66.1           | 20.894  |-1.116  |
| I am good in Data Wrangling and Preprocessing Skills.              | 0     | 56             | 22.059  | -0.488  |
| I am good in Essential Programming Skills (e.g.: Python and R).     | 0     | 51.3           | 24.555  |-0.627  |
| I am good in Basic Machine Learning Skills (Problem Framing, Data Analysis, Model Building, Testing & Evaluation and Model Application). | 0     | 56.8           | 23.033  |-0.558  |

### Campus Readiness

| Item                                                                 | Score | Readiness Score | Z-score |
|----------------------------------------------------------------------|-------|----------------|---------|
| The university has computer laboratories for handling big data.       | 4     | 77.8           | 21.973  |-1.042  |
| The lecturers involved in teaching subjects on big data are knowledgeable in big data software. | 5     | 82.1           | 18.729  |-1.455  |
| The lecturers involved in teaching subjects on big data have teaching skills in big data analytics. | 12    | 82.1           | 17.965  |-1.493  |
| The internet connection provided to support learning big data is adequate. | 10    | 76.6           | 19.571  |-0.794  |

### Student Awareness

| Item                                                                 | Score | Readiness Score | Z-score |
|----------------------------------------------------------------------|-------|----------------|---------|
| I am aware that there are job careers relating to big data.           | 0     | 84.2           | 18.137  |-1.805  |
| I realise that jobs in the field of big data analytics are highly demanded. | 0     | 83.4           | 17.72   |-1.662  |
| I know the characteristics of big data                                | 0     | 69.5           | 20.628  |-0.687  |
| I know what kind of work involve in big data analytics careers.       | 0     | 69.2           | 20.624  |-0.79   |

### Tools for Big Data Analytics

| Item                                                                 | Score | Readiness Score | Z-score |
|----------------------------------------------------------------------|-------|----------------|---------|
| I am good in Tableau Programming Language.                           | 0     | 41.7           | 29.201  |-0.029  |
| I am good in Hadoop Programming Language.                            | 0     | 30.9           | 25.918  | 0.358   |
| I am good in Spark Programming Language.                             | 0     | 27.1           | 25.063  | 0.58    |

### 4.3. Box and Whiskers Plot

The distribution of the data was demonstrated by using box and whisker plot in figures 4(a) – 4(g). In this figure, the line in the middle of the box represents median and the end whiskers represent maximum and minimum values. The data scores range from 0 to 100, indicating the percentage of readiness score for each item. A score of 0% will be deemed low, while a score of 100% will be considered high. In the box and whisker plot, outlier or extreme values were also easily noticed, except in figures 4(d) and 4(g), the presence of extremely low median scores in almost each box plot in figure 4(g).
Figures 4(a)-4(g) shows the comparison of scores using box plots. Figure 4(a) displays the box plots for teamwork as a factor in students' readiness for big data analytics job criteria. According to the median of the box (the line in the box denotes the median), 50% of the respondents scored more than 80% in
readiness score. It indicates that the students enjoy working in groups and collaborating among themselves. Each box plot shows that all five items in this factor have high scores more than or equal to 80%. One of the most crucial qualities to become a big data analyst is the willingness and ability to work in a group.

In figure 4(b), the box plot for the student efforts factor shows that the median for each box plot is approximately 80%, and the median score for each item appears to be similar. This indicates that the students in this study were interested to have a career in big data analytics as they were willing to improve their data communication skills, willing to spend on workshops to keep them updated and willing to acquire the knowledge in big data analytic skills more. Indirectly, these efforts had proven that the students were ready to take up the job opportunity in big data analytics.

Based on the median value for data visualisation and storytelling skills in figure 4(c), the median score for each box plot appears differently. The data visualisation and storytelling skills factor is between 60% to 80%, as shown by the middle line of the box. Students could interpret data, explain the analysis to others, and visually present the research findings, but they were wary of being questioned by the audience during the presentation. The outcome reflected that some courses or workshops are needed to sharpen the students’ presentation skills and boost their self-confidence. This experience will allow them to be a good presenter and, at the same time, able to convince the audience with their content.

Figure 4(d) displays the box plots for skills in BDA criteria as a factor in students’ readiness for big data analytics job criteria. Each box plot in this figure shows that all four items in this factor have scores below 80% and are considered as a moderate score. This finding reveals that students had moderate knowledge of essential programming skills like Phyton and R language. As discussed by [5], these two languages were most widely used in large companies such as Google and Facebook and learning these will help them to get hired as data scientists. Similarly, programming languages such as Phyton, SQL, JAVA, Ruby, and statistical platform such as R, SAS, MATLAB are the skills required by employers when they search for a data scientist worker [7]. Students also claimed that they were good in data visualisation skills, SQL programming, data wrangling and pre-processing skills. This has shown a good sign in equipping students for related jobs in big data analytics.

In preparing the students for big data analytics career, the university itself must have competency in terms of facilities such as a compatible computer lab, updated software, high bandwidth and stable internet connection, and knowledgeable staff in big data analytics. The box plot in figure 4(e) reveals the campus readiness factor and it is shown that the median score for each item is high, as the median scores highlighters or equal to 80%. Most of the students claimed that their campuses were well-equipped with the necessary facilities, and that their lecturers were also knowledgeable in big data analytics-based applications.

Were the students aware of the job opportunities related to big data? The median values of student awareness in figure 4(f) indicate that most of them were aware of this career opportunity and realised the high job demand in the field of big data analytics. Even though the students believed that having good knowledge in statistics, programming and machine learning is important to be hired as a data scientist, some of them were still not clear of the job scope in big data analytics. The middle line of the box demonstrates this finding. As for the faculty, exposing students to big data job families and skills required to be part of data scientist would attract students to explore more on the demanding trends in big data analytics career.

Figure 4(g) shows a box plot depicting big data analytics tools as a factor in students’ readiness for big data analytics job criteria. In this figure, the middle line denotes that the median values for each item is less than 50% and is deemed low. This study found that big data analytics tools like Tableau, Hadoop, and Spark were not extensively used and the respondents were not yet accustomed to these software.

4.4. Correlation Analysis
A correlation coefficient indicates the degree to which two variables are associated and the existence of the relationship. A correlation value less than 0.35 is regarded poor, 0.36 to 0.67 is moderate, and beyond 0.68 is considered strong [19]. The result in table 3 shows the correlation values between six (6) predictor variables and effort. DVSS (R = 0.375), SA (R = 0.580), TW (R = 0.577) and SBDA (R = 0.506) exhibit moderate positive relationship. On the contrary, TBDA (R = 0.375) and CR (R = 0.411) show a weak
positive relationship. All the six predictor variables move in the same direction; one variable will increase as the other increases. As the student effort increases, all the six predictor variables also increase.

### Table 3. Correlation Analysis

| Variable                                      | Correlation (R) | Strength       |
|-----------------------------------------------|-----------------|----------------|
| Tools for Big Data Analytics (TBDA)           | 0.375*          | Weak Positive  |
| Data Visualisation and Storytelling Skill (DVSS) | 0.595*          | Moderate Positive |
| Campus Readiness (CR)                         | 0.441*          | Weak Positive  |
| Student Awareness (SA)                        | 0.580*          | Moderate Positive |
| Teamwork (TW)                                 | 0.577*          | Moderate Positive |
| Skills for Big Data Analytics (SBDA)          | 0.506*          | Moderate Positive |

*Correlation coefficient is less than 0.01

4.5. *Multiple Linear Regression*

Table 4 shows the relationship between 6 (six) predictor variables; TBDA, DVSS, CR, SA, TW and SBDA towards student effort in a multiple linear regression analysis. It shows the existence of strong relationship \( R = 0.702 \) between dependant variables (TBDA, DVSS, CR, SA, TW and SBDA) and student effort with a coefficient of determination of 47%. This indicates that the 47% of variability in student effort is due to TBDA, DVSS, CR, SA, TW and SBDA. On the contrary, 53% of variability are due to other factors.

### Table 4. Correlation Analysis in Multiple Linear Regression

| Variable                                      | Correlation (R) | Strength       | \( R^2 \) |
|-----------------------------------------------|-----------------|----------------|----------|
| TBDA, DVSS, CR, SA, TW, SBDA                 | 0.702           | Strong Positive| 0.470    |

The relationship between 6 (six) independent variables; TBDA, DVSS, CR, SA, TW and SBDA and student effort was studied using multiple linear regression. Based on the result of the regression in table 5, significant value shown from the test \( (p\text{-value} = 0.000) \) is less than significance level of 0.05. As a result, linear regression model is significant and denotes that there is a significant relationship between six predictor variables and student effort.

### Table 5. ANOVA test in Multiple Linear Regression

| Model     | Sum of Squares | Df | Mean Square | F      | Significance |
|-----------|----------------|----|-------------|--------|--------------|
| Regression| 68.533         | 6  | 11.422      | 21.398 | 0.000        |
| Residual  | 70.461         | 132| 0.534       |        |              |
| Total     | 138.994        | 138|             |        |              |

Other concerns about the linear regression model are the existence of multicollinearity which could significantly reduce the model’s performance. The variance inflation factor (VIF) in Table 6 presents all satisfactory values for each predictor variables. The VIF is greater than one and less than five, TBDA \( (VIF = 1.524) \), DVSS \( (VIF = 1.860) \), CR \( (VIF = 2.372) \), SA \( (VIF = 2.519) \), TW \( (VIF = 2.511) \) and SBDA \( (VIF = 1.849) \). A VIF value of 1 is non-collinear, range between 1 and 5 are moderate, and VIF value more than 5 are highly collinear and indicates a serious collinearity problem [20]. As a result, there is no multicollinearity problems in the linear regression model.
Table 6. Coefficient of regression model and variance inflation factor analysis of the relationship between six predictor variables and student effort

| Variable                        | B estimation | Standard error | F value | Significance | VIF |
|--------------------------------|--------------|----------------|---------|--------------|-----|
| Constant                       | -3.095E-5    | 0.062          | 0.000   | 1.000        |     |
| Tools for Big Data Analytics   | 0.054        | 0.077          | 0.709   | 0.480        | 1.524 |
| Data Visualisation and Storytelling Skill (DVSS) | 0.291        | 0.085          | 3.442   | 0.001*       | 1.860 |
| Campus Readiness (CR)          | -0.124       | 0.095          | -1.298  | 0.196        | 2.372 |
| Student Awareness (SA)         | 0.191        | 0.098          | 1.941   | 0.054        | 2.519 |
| Teamwork (TW)                  | 0.284        | 0.098          | 2.888   | 0.005*       | 2.511 |
| Skills for Big Data Analytics  | 0.158        | 0.084          | 1.874   | 0.063        | 1.849 |

Table 6 captures multiple linear regression model that was computed to investigate predictor variables that had a significant influence on the student effort. There are two predictors which have a significant impact on student effort: DVSS and TW. DVSS increases the effort by 0.291% (p-value <0.01). Specifically, 0.291% (± 0.085) increase in the student effort for every 1% increment in the data visualisation and storytelling skill factor. Data visualisation and storytelling are important elements in Data Science. Both visualisation and storytelling approach have assisted data scientists in their effort to inform others on the conclusions drawn from their research [21]. Students who acquire data visualisation and storytelling skills are more ready to be data scientists. Thus, this might explain why data visualisation and storytelling are found to be significant contributors. TW also leads to increase student effort by 0.284% (p-value <0.01). Specifically, 0.284% (± 0.098) increase in the student effort for every 1% increase in the teamwork factor. The Data Science team mainly comprises a data engineer, data scientist, data analyst, business intelligence, software engineer, developer, project manager, and data coordinator [22]. The primary goal of teamwork is for data scientists to have a positive outlook; however, teamwork is relevant to various job types to maximise effectiveness. Since teamwork is an important element in data science, students who can work in a team are readier to be a data scientist. In contrast, TBDA, CR, SA, TW and SBDA were found to have no significant impact on student effort. Details of the multiple linear regression model can be found in table 6.

Two predictor variables, DVSS and TW, have exhibited significant results in the prediction model using multiple linear regression, while the other four predictor variables suggest the opposite. This demonstrates that the students are ready if they meet the two requirements of this field. The remaining four relevant criteria; tools for big data analytics, student awareness, campus readiness, and skills for big data analytics, are also the aspects that the university need to pay attention to so that when students graduate, they will be ready and prepared to fill up this job. According to [23], there is a gap between what education provides and what the industry requires. Companies require a strong programming and software specialisation, which is less common in university study programmes. The findings of this study can at least explore the gap in terms of skills and requirements needed for a graduate to apply for any position in the big data analytics. Interestingly, a research carried out by [24] has shown many big data job advertisements emphasise the development of analytical information systems, as well as the fact that soft skills are still highly valued. As a result, it is critical for universities and industries to collaborate by creating an effective work environment during industrial training so that students can be exposed to the industry requirements.

5. Conclusion
An expert once said that data scientist is not born, they are made. For students to be ready for a big data analytics career, the grooming stage should start since day one of their study at the university. If a law student must be trained to be a lawyer, a medical student is tailored to become a doctor, the same thing...
goes to computer science graduates. Preparing them to become a data scientist will help them master the needed skills for big data analytics and this will increase their hiring potentials. The university, especially the faculty, must expose students to activities or programmes related to big data analytics. The students can have experience in learning big data analytics tools by enrolling in related online courses such as DataCamp or Coursera. Inviting an expert in this field to share experience or the journeys to be a data scientist might be helpful to attract and prepare the students for this job opportunity. Therefore, universities must always be updated with the latest information on data analytics, such as skills needed or the latest big data analytics tools, to ensure the high employability rates of their graduates.

6. Acknowledgements
The authors would like to acknowledge the PJI of Universiti Teknologi MARA Cawangan Terengganu for providing a fund for Special Interest Group (SIG) to this research.

References
[1] Garbade MJ. LinkedIn Workforce Report: Data Science Skills are in High Demand Across Industries [Internet]. Towards Data Science. 2019 [cited 2019 Mar 14]. Available from: https://towardsdatascience.com/linkedin-workforce-report-data-science-skills-are-in-high-demand-across-industries-1518b06382a6
[2] How much does a Data Scientist make in the United States? [Internet]. indeed.com. [cited 2021 May 6]. Available from: https://www.indeed.com/career/data-scientist/salaries
[3] Shaheera AS. Local talent shortage forces IT firms to seek foreigners [Internet]. The Malaysian Reserve. [cited 2019 Apr 14]. Available from: https://themalaysianreserve.com/2017/09/19/local-talent-shortage-forces-firms-seek-foreigners/
[4] Davenport TH, Patil DJ. Data scientist: The sexiest job of the 21st century. Harv Bus Rev. 2012;
[5] Radecic D. Must-Have Skills to Get a Job as a Data Scientist [Internet]. Applison. 2021 [cited 2021 May 6]. Available from: https://applison.com/7-data-science-skills/
[6] Przybyla M. Data Science 101 A Step-by-Step Process for Every Use Case [Internet]. Towards Data Science. 2021 [cited 2021 Apr 29]. Available from: https://towardsdatascience.com/data-science-101-91b9f852a01c
[7] De Mauro A, Greco M, Grimaldi M, Ritala P. Human resources for Big Data professions: A systematic classification of job roles and required skill sets. Information Processing & Management. 2018 Sep 1;54(5):807-17.
[8] Dhar V. Data Science and Prediction [Internet]. Association for Computing Machinery. 2013 [cited 2019 Mar 5]. Available from: http://cacm.acm.org/magazines/2013/12/169933-data-science-and-prediction/fulltext
[9] Mohanty S, Jagadeesh M, Srivatsa H. Big data imperatives: Enterprise ‘Big data’ warehouse, ‘BI’ implementations and analytics. Apress; 2013 Jun 24.
[10] Woods D. EMC Greenplum’s Steven Hillion on What Is a Data Scientist? [Internet]. Forbes. 2011 cited 2019 Apr 14. Available from: https://www.forbes.com/sites/danwoods/2011/10/11/emc-greenplums-steven-hillion-on-what-is-a-data-scientist/#947531f3a070
[11] Holak B. Demand for data scientists is booming and will only increase [Internet]. TechTarget Search Business Analytics. 2019 [cited 2019 Mar 14]. Available from: https://searchbusinessanalytics.techtarget.com/feature/Demand-for-data-scientists-is-booming-and-will-increase
[12] Chen H, Chiang R, Storey V. 2012 Business Intelligence and Analytics: From Big Data to Big Impact. MIS Quarterly. 36, 1165–1188 p.
[13] Wixom, B., Ariyachandra, T., Goul, M., Gray, P., Kulkarni, U. and Phillips-Wren, G., 2011. The current state of business intelligence in academia. Communications of the Association for information Systems, 29(1), p.16.
[14] Song, I.Y. and Zhu, Y., 2016. Big data and data science: what should we teach? Expert Systems, 33(4), pp.364-373.
[15] Knopp BM, Beagley S, Frank A, Orrie R, Watson M. 2016 Defining the Roles, Responsibilities,
and Functions for Data Science Within the Defense Intelligence Agency. *RAND National Defense Research Institute Santa Monica United States.*

[16] Harris JG, Shetterley N, Alter AE, Schnell K. 2013 The Team Solution to the Data Scientist Shortage. *Accenture Institute for high performance.*

[17] Kalema BM, Mokgadi M., 2017. Developing countries organizations’ readiness for Big Data analytics. *Problems and perspectives in management, (15, Iss. 1 (cont.)), pp.260-270.*

[18] Yusoff R, Mohd Janor R., 2014, Generation of an interval metric scale to measure attitude. *SAGE Open, 4(1).*

[19] Taylor R.,1990. Interpretation of the correlation coefficient: a basic review. *Journal of diagnostic medical sonography, 6(1), pp.35-39.*

[20] James G, Witten D, Hastie T, Tibshirani R., 2013. An introduction to statistical learning. *Springer New York, 112.*

[21] Loon R Van.,2021. The Value of Mastering Data Visualization & Storytelling for Data Scientists. *Simplilearn.* Available from: https://www.simplilearn.com/value-of-mastering-data-visualization-storytelling-for-data-scientists-article

[22] Lee A., 2018. Why Team Building Is Important To Data Scientists? *Towards Data Science.* Available from: https://towardsdatascience.com/why-team-building-is-important-to-data-scientists-a8fa74dbc09b.

[23] Della Volpe M, Esposito F., 2020. How universities fill the talent gap: The data scientist in the Italian case. *African Journal of Business Management, 14(2), pp.53-64.*

[24] Gardiner A, Aasheim C, Rutner P, Williams S. Skill Requirements in Big Data: A Content Analysis of Job Advertisements. *Journal of Computer Information Systems, 58(4), pp.374-384*