Graphical user interface, text

Description automatically generated with medium confidence

Abhilash Kumar Dikshit   
R Practice: ALY 6010

Module 2

Week 2

Initially, I began by downloading the raw dataset *“Student\_Mental\_Health.csv”* provided by the professor and using the libraries below for further data analysis report.

|  |  |  |
| --- | --- | --- |
| |  |  | | --- | --- | |  | We have 101 observations and 11 attributes including one cell with no assigned value. |   *Fig1: Top and Bottom 3 Of Raw Dataset* |

|  |  |
| --- | --- |
| *Fig3: Data set’s variable and respective skim outputs* | Graphical user interface, application  Description automatically generated  *Fig2: Top and Bottom 3 Of Cleaned Dataset*  Fig 2: In order to clean up the data, it was sorted by ascending *Age*; column headers were renamed properly, and then the blank cells were removed; final stage was to rename the regular expression using *gsub* for column 5. |

Abhilash Kumar Dikshit

R Practice: ALY 6010

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fig 3: As per the output, we had 9 character and 2 numeric data types.  We had the min, max and unique counts for individual character data type attributes. For numeric attributes, the mean of Age of students had a value pf 20.53 and SD is 2.50; for current study year the mean value was 1.98 and SD is 0.99. | |  | | |
| *Fig4: Mean, Min, Max, STD of Gender & Marital Status* | | Fig 4: Mean value for male students was 20.52 and for female students it was 20.53 with min and max value of 18 and 24 for both the genders. STD for male students was 2.48 which is 0.3 value less than the female students. | | |
| *Fig5: Health Analysis Report (Depression, Anxiety, Panic Attack)*  Fig 6: For students without any health condition, we can see the marital status of those are single and hence, were living a healthy life.    *Fig7: Health Analysis Report for Specialist Treatment*  Abhilash Kumar Dikshit  R Practice: ALY 6010    *Fig8: Health Analysis Report for Engineering students* | | | If we talk about students’ health, respective data frame (Fig 5-7) shows the data of students suffering from Depression, Anxiety and Panic attack and are they going for Specialist treatment or not.  Fig 5: For students suffering with all three-health condition, only one female student went for Specialist treatment.    *Fig6: Health Analysis Report (No Depression, Anxiety, Panic Attack)*  Fig 7: Although there were many students who suffered one or more health conditions but only 6 students went for Specialist treatment and all of them suffered from Depression. Moreover 90% of them were female between age group 18-24.  Fig 8: From a total of 17 Engineering students, only 2 students suffered from Depression, 3 students from Anxiety, 3 students from Panic attack but none went for specialist treatment. | |
| Fig 9: As shown in the Frequency table, out of 16 students in the age group who were married, most of them are in the age group of 24 followed by 19 and 18 but the highest number of students were in the age group of 18 i.e., 32, followed by 23 students in the age group of 24. | | | *Fig9: Student Marital status w.r.t Age* | |
| *Fig10: Cross tabulation for Depression and Anxiety* | Fig 10: For the cross-tabulation report for 100 observations, we can depict, 65 students had no Depression out of which only 16 were feeling Anxiety and out of 35 students who were having Depression, 18 were having Anxiety.    *Fig11: 3-way crosstabs using ggplot facets*  Fig 11: 3-way cross tabulation using ggplot was shown and considered students whose age >15 as well as their marital status. From the above plot, we can depict the count of married male students are higher than married female students who went for specialist treatment. On the other hand, more than 75 female students who were unmarried wished to not to go for specialist treatment as compared to unmarried male. | | |

Abhilash Kumar Dikshit

R Practice: ALY 6010

|  |  |  |  |
| --- | --- | --- | --- |
| Chart, scatter chart  Description automatically generated *Fig12: CGPA based on Age for the current year of study (Scatter Plot)* | Fig 12: In the given scatter plot, the students with respective CGPA’s had been shown for every year.  In year 1, few students in age group 18, 20 and 21 scored a CGPA of 3.50-4.00.  In year 2, few students in age group 21 scored a CGPA of 3.50-4.00.  In year 3, few students in age group 18, 19 and 21 scored a CGPA of 3.50-4.00.  In year 4, few students in age group 18, 22 and 23 scored a CGPA of 3.50-4.00.  Moreover, *Image* function has been used in the given plot. | | |
| Fig 13: In the given scatter plot, the number of courses were shown with regards to Year 1 to 4, and it had been differentiated based on gender.  In year 1, male students had taken BIT, Irksh and Radiography.  In year 2, male students had taken Benl, Engineering, Islamic education, Koe and TAASL.  In year 3, male students had taken Biotechnology, Kirkhs and MHSC.  In year 4, male students had taken only Mathematics.  Moreover*, Image* function has been used in the given plot. | Chart  Description automatically generated with medium confidence  *Fig13:* Courses provided for Year 1-4 | | | |
| Chart, diagram, box and whisker chart  Description automatically generated  *Fig14: Specialist Treatment for Depression, Anxiety, and Panic Attack Based on Age (Jittr Plot)* | | Fig 14: In the given jittr plot, common label “Specialist Treatment” has been chosen for all the 4 plots. Students with Depression, Anxiety (Mean and SD) and Panic attack (Median) had been shown combined using ggarrange function.  Moreover*, Image* function has been used in the given plot. |

Abhilash Kumar Dikshit

R Practice: ALY 6010

|  |  |
| --- | --- |
| Fig 15: Top 25 student’s data were considered based on CGPA for the boxplot.  Mean value for year of study was calculated and the value was 2.28 which had been denoted as red dash line. There were no male students in the top 25 list and the female students opted for BCS, BIT and Engineering courses.  Moreover*, Image* function has been used in the given plot. | Chart  Description automatically generated with low confidence  *Fig15: Courses taken by top 25 students based on CGPA* |

|  |  |
| --- | --- |
| Chart, bubble chart  Description automatically generated  *Fig16: Courses taken by top 25 students based on Depression* | Fig 16: Top 25 student’s data were considered based on Depression for the boxplot.  Mean value for Age was calculated and the value was 22.8 which had been denoted as red dash line. There were no students who were in depression in the top 25 list and the students opted for BCS (age group 21-23) and BIT courses (age group 21-24).  Moreover*, Image* function has been used in the given plot. |
| Fig 17: Top 25 student’s data were considered based on Anxiety for the boxplot.  Mean value for Age was calculated and the value was 22.8 which had been denoted as red dash line. Students who opted for BCS were in the age group 22-24 and for BIT courses in the age group was 21-23, had an anxiety whereas for students opted for Engineering did not face the same.  Moreover*, Image* function has been used in the given plot. | Chart  Description automatically generated  *Fig17: Courses taken by top 25 students based on Anxiety* |

Abhilash Kumar Dikshit

R Practice: ALY 6010

|  |  |
| --- | --- |
| Fig 18: Top 25 student’s data were considered based on Panic attach for the boxplot.  Mean value for Age was calculated and the value was 22.8 which had been denoted as red dash line. Students who opted for BCS were in the age group 21-24; for BIT courses in the age group was 21-23, and for Engineering course in the age group was 22-24 faced the Panic attack whereas for students opted for BCS did not face the same.  Moreover*, Image* function has been used in the given plot. | Chart, bubble chart  Description automatically generated  *Fig18: Courses taken by top 25 students based on Panic Attack* |

|  |  |
| --- | --- |
| *Fig19: Frequency of top 25 and total student’s age* | Fig 19: As shown in the histogram plot, the mean value of age for top 25 students was 22.8 and the highest number of students were in the age group of 24 i.e., 12.  Whereas the mean value of age for overall students was 21.5 and the highest number of students were in the age group of 18 i.e., 53. |

|  |  |
| --- | --- |
| Fig 20*:* CGPA w.r.t Age density plot was shown in the given figure. As per the density plot, in the age group of 18-20, the density of CGPA 2.50-2.99 was more than 0.20 and it went down in the age group of 20-22.  Moreover*, Image* function has been used in the given plot. | **Chart, histogram  Description automatically generated**  *Fig20* |

**References:**

1. *Descriptive Statistics and Graphics- Easy Guides* ([October 28, 2022](http://127.0.0.1:64200/rmd_output/4/#ref-R-Material)); Kervizic ([Feb 9, 2020](http://127.0.0.1:64200/rmd_output/4/#ref-R-Career));
2. *Data Visualization Tips for More Effective and Engaging Design*. n.d. Tableu. <https://www.tableau.com/learn/articles/data-visualization-tips>.
3. *Descriptive Statistics and Graphics- Easy Guides*. October 28, 2022. <http://www.sthda.com/english/wiki/descriptive-statistics-and-graphics>.
4. Kabacoff, Robert I. May 2015. *R in Action, Second Edition*. <https://www.manning.com/books/r-in-action-second-edition>.
5. Kervizic, Julien. Feb 9, 2020. *Data Science Career Path & Progression*. <https://medium.com/analytics-and-data/data-science-career-path-progression-ab5140cfbc84>.
6. *Mapping in r: Latitude and Longitude Coordinates*. October 29, 2019. <https://map-rfun.library.duke.edu/01_georeference.html>.