Pearl Omani-Darpah

Data Analytics

Week 2.5 Workshop (Capstone Project)

The dataset I have chosen to clean is Student Adaptability Level Online Education. I found this Dataset in Kaggle. Please see link: <https://www.kaggle.com/datasets/mdmahmudulhasansuzan/students-adaptability-level-in-online-education>

**IFS Formula**

First, I changed the *Gender* column A by replacing the boys/girl values with Male/Female, using the formula IFS, as shown below.

A screenshot of a computer

Description automatically generated

I then went on to change column B, by formatting the age to Numbers. However, this did not change all the values in the cells.

A screenshot of a computer

Description automatically generated

I had to change them to Text and manually enter the correct values for the cells within column B. I filtered the column to reduce human error.

A screenshot of a computer

Description automatically generated

*(I was able to change column B to Number format once I had completed the cleanup).*

I spell checked the data set and found the value Wifi. I changed this entry to Wi-Fi, in column I, *Internet*, by using the Find & Replace function.

A screenshot of a computer

Description automatically generated

I aligned the cells and placed a bold font in the title’s rows, to make them stand out and added more info in the titles to give the end-user more context. I later changed my mind on this. Instead, I created a validation, which is explained on pg4 (Data Validation).

**IFS Formula**

In column F, *Location*, I changed the values from Yes/No to Town/ Village, using an IFS formula to convey more meaning for the end-user. I chose the town as this is what is stated in the details of the web page where the dataset can be found.

A screenshot of a computer

Description automatically generated

In column K, *Class* *Duration*, which represents the hours spent in a lesson, I had to format the column to *Text* because the same error occurred as it did when I tried to format column B, *Age* to Numbers.

Below shows column K filtered and the incorrect value of 3-Jul was changed to 3-6.

A screenshot of a computer

Description automatically generated

*(I was able to change column K to the Number format, once I have completed the cleanup).*

**IFS Formula**

I wanted to make column H, *Financial Conditi*on, convey more meaning to the end-user. In a report I found discussing the average household income of families in Bangladesh, they stated the amount in USD; 259.538 USD was the poorest income, 439.888 USD was the average, and 995.873 USD was the wealthiest income belonging to families living in Bangladesh.

This research was sourced: <https://www.ceicdata.com/en/indicator/bangladesh/annual-household-income-per-capita>

I used this information to create an additional column inserted after column H, named *Financial Condition 2*, to define each student’s family economic amount. I created an IFS formula to present the research that I sourced. I rounded up the amount so Poor would become less than $438, Mid would become $439-$994 and Rich would become $995 or above.

**A screenshot of a computer

Description automatically generated**

**Conditional Formatting**

In the Home tab, I selected Conditional Formatting in the dropdown menu I selected, highlighted Cells Rules, and then Text That Contains. In ‘Text That Contains,’ I selected what colour I wanted the values to reflect, red for the poor, yellow for the average and green for the wealthiest.

A screenshot of a computer

Description automatically generated

*(The Data Validation list is saved in the Cleaned Data worksheet)*

**Data Validation**

I used Data Validation for all columns (except I) in the dataset to give the end-user more context.

The following steps are the actions I took to achieve this method:

1. I highlight (selected) the data in column A, *Gender*.

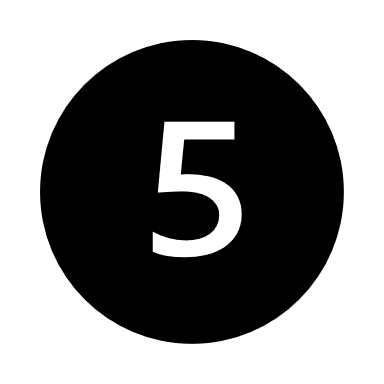
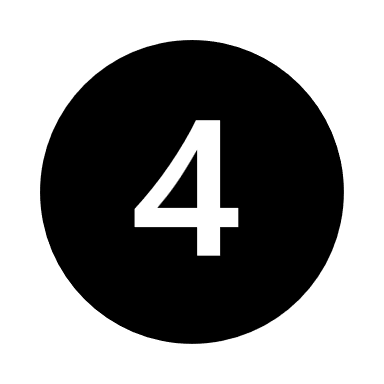
2. In the Data Tab (ribbon) select the Data Validation dropdown menu. A Data Validation table opened.

3. In the Settings Tab, I changed the entry in Allow to List.

4. Clicked on the Source.

5. Then I highlighted the cell O12:O13 because this is where I entered the valid values for the *Gender* column. The location is generated in Source.

In the Input Message Tab, under Title, I entered; ‘Instruction,’ in Message I entered; ‘Please input the valid value.’

****In the Error Alert Tab, under Style I selected ‘Stop,’ to alert the end-user to an incorrect entry. In Title, a entered a short warning; 'Invalid Value' and in Error Message, I entered; 'Please input the correct value', then pressed the OK button to save the instructions. *A screenshot of a computer

Description automatically generated (I will later change the Gender selections to include Non-Binary)*

**Unique & Count IF formulas**

I used Unique & Count IF formulas to group each column individually, so that I could create dynamic images to present my analysis. Dynamic images such as Funnel and Sunburst cannot be represented in using a Pivot table.

The formulas below are used to create the Age Range table.

=UNIQUE(H2:H1206,FALSE,FALSE)

=COUNTIF(H2:H1206, "Mid")

This table was used to create the Sunburst Chart

A screenshot of a report

Description automatically generated

*(Unique & Count IF formulas are saved in the Cleaned Data worksheet, placed in column U to AC)*

**Visualizations**

I used a Sunburst chart to present students’ financial conditions.

A graph of a student financial

Description automatically generated

I used a Picture as a dynamic image to visualise the student’s genders represented in the data.

**A person and person with numbers and symbols

Description automatically generated**

I believe this a dynamic image it is an effective way to present the data, no axis or chart names are needed.

I used a Funnel Chart to present the information available on the student's age range. **A green rectangular object with numbers

Description automatically generated with medium confidence**

*(All the above Visualizations are saved in the Worksheet named Dashboard)*

**Pivot Table, Charts and Slicers**

I created Pivot charts to represent information that has been cleaned in the worksheet Cleaned Data worksheet. I also made slicers; they are connected to all the charts in the Dashboard worksheet.

A screenshot of a computer

Description automatically generated

The Learning Management System (LMS) enhances student success and engagement in online learning environments. There is sufficient information to show that student's economic background is not a factor when it comes to their accessing LMS.

The column chart shows students with a mid-range family economic background were more likely to have access the LMS:

A graph of a bar chart

Description automatically generated with medium confidence

The bar chart below compares those students who were able to access the LMS and what their Adaptivity Levels were. The bar chart denotes that those who did not have access the LMS were highly adaptive in their institution.

A graph with a bar and a number of levels

Description automatically generated with medium confidence

The bar chart below compares the effects of Load Shedding to the location (Town/City) of the students.

A graph of a load shedding

Description automatically generated

**Dashboard**

*A screenshot of a computer

Description automatically generated (All Pivot Table/ Charts are saved in the worksheets named Pivot Tables and Charts. Dashboard is saved in the Dashboard worksheet)*

**In conclusion**

It appears that students who are in the wealthiest financial condition are not necessarily performing better than those in a poor financial condition. However, it is worth noting that those in the mid-range financial condition, who have access to Learning Management Systems (LMS) and are at an Adaptivity level, are performing well. When changing the slicers, there was not a high amount of load shedding on either those living in the village or town. Interestingly, students who reside in towns seem to be better at Adaptivity levels.

Students aged 6-10 years are educated to a school level, rated the highest at 33%, while those aged 0-1 rated low. We can hypothesize that the low rate for the 0-1 age group is due to their early age. However, when we looked at the age group of 16-25 years, they also scored low on the Adaptivity level. Further investigation is needed to determine the reasons behind this trend. If we collected data on previous students' experiences with online learning, it may help us better understand the trend and enhance the effectiveness of online learning platforms for students in all locations.

Overall, female students have an overall higher Adaptivity level rating of 2% than male students. In contrast, male students have a moderate to low adaptivity level ratio of 50/50.