INVESTIGATE A DATASET PROJECT

I analyzed the 'noshowappointment' dataset

The following questions were asked to aide my analysis

How many persons who had medical conditions, are a recipients of government scholarship, received sms text kept to their appointment?

What is the number of persons that are recipients of government scholarship, and have been diagnosed with health issues that also received an sms reminder of their appointments?

what Gender is the highest beneficiary of scholarship, and also has medical conditions which received an sms text?

Those who were on government scholarship that received sms reminder of their appointments.

To be able to provide the right-visual-answer to each of the above questions, I had to use the

GROUPBY function.

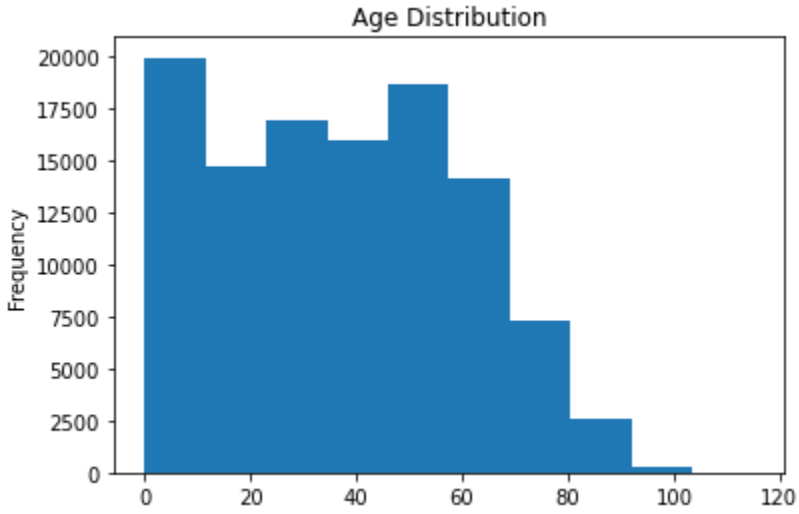
This enabled me to appropriately categorize my above questions and answer them accordingly.

- As a data analyst, I had to ensure that my data was truly clean before going on to

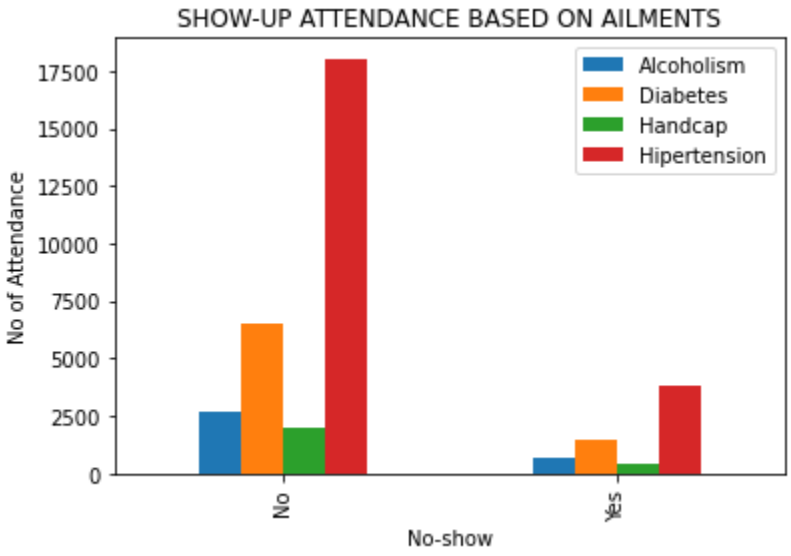
visualizing my findings. To do this

1. I had to first 'set-up' my jupyter notebook environment with the necessary libraries.
2. Next was to load the dataset.
3. I went on to drop duplicates (if there were any).
4. I described the data.
5. I checked if there was any null or NaN values.
6. I checked the info of my data.
7. I replaced age values less than zero with 0.
8. I inserted a pivot-table.
9. Finally, I had to group my data to suit the questions posed above, and Visualized each one of them using the matplotlib library.

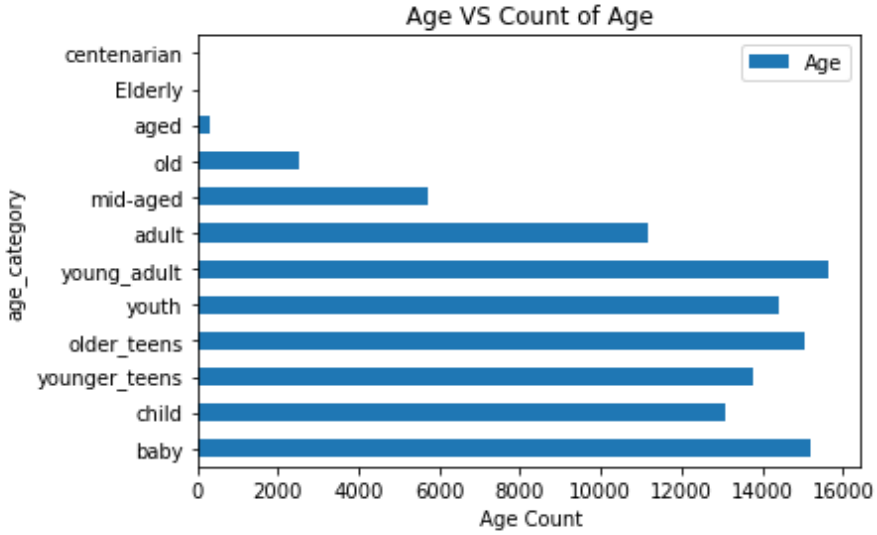
This displays the age distribution of patients with in the dataset. It shows their ages and the frequency.



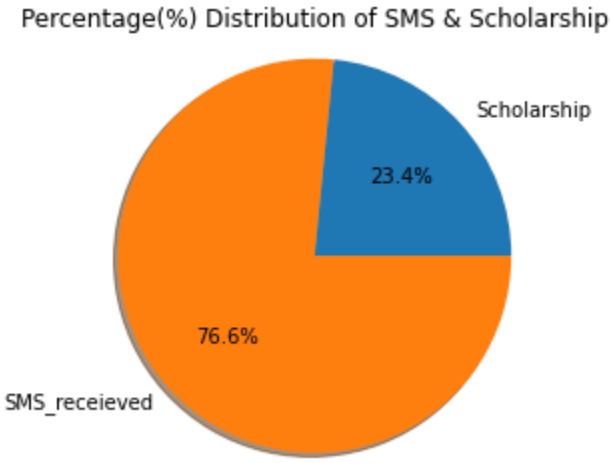
This shows the attendance of patients based on the different ailment they have. It also shows the number of different medical conditions of those who showed up and those who didn't.



This displays the age count and different age categories of patients.



This shows the percentage(%) distribution of patients who were recipients of scholarship, and also those that received sms alerts.



LIMITATION

The limitation I encountered was that matplotlib do not have so much visualization options for large datasets. I noticed that it is easy to work with matplotlib if the dataset is small but if it is a large dataset it is sort of an issue.

CONCLUSION

I realized that there were not many government scholarship given out, perhaps it affected the attendance of people to show up for the appointment. It might have been expensive. They might actually have booked for appointment but didn't show up because it was not affordable for them.

N/A (this work was done by me, I only made reference to datasets I had worked on prior to this scholarship.)