a. What is the median age of subjects in the sample size above?

Based on the data provided, a python code was created to identify the median age, , the code is attached in the appendix below. The median age is 27.5.

b. What is the average salary difference of a person with only a diploma vs one with a Master's degree?

Based on the data provided, a python code was created to identify the average salary difference, the code is attached in the appendix below. On average, a person with only a diploma makes RM41500 per year less than a person with a Master's degree.

c. Does age affect the salary of someone with a diploma? Please justify in mathematical terms.

Based on the data provided, a python code was created to identify whether age affects the salary of someone with a diploma, the code is attached in the appendix below. The answer is yes, a graph of Annual Income against age is plotted, if a best fit line is drawn for the graph, there is an increasing linear relationship between Annual Income and Age.

d. How much do you think a diploma holder will be earning at the age of 40? Please share your hypothesis & the calculations that lead to that outcome.

Based on the data provided, the python code for part c. was modified to identify the annual salary of a diploma holder who is 40 years old, the code is attached in the appendix below. The answer is RM 38592.5, a line equation in the form of y = mx+c was formed. The line equation was used to determine the annual salary of the 40-year-old diploma holder, y, by substituting the x value with 40, m with the slope, c with the intercept of the best fit line from part c.

e. What kind of phone do you think someone who earns RM 5,000 & above per month will possess? Kindly advise on your assumptions, and the calculations you considered.

Based on the data provided, a python code was created to identify the type of phone used by individuals with monthly income greater than or equal to RM 5000, the code is attached in the appendix below. The answer would be an iPhone, this is because 3 out of 4 individuals who earn more than or equal to RM 5000 per month use an iPhone.

f. What is the most preferred way of requesting for an interview for someone who spends more than RM 2,000 for their phone?

Based on the data provided, a python code was created to identify most preferred way of requesting for an interview for someone who spends more than RM 2,000 for their phone, the code is attached in the appendix below. The answer is Email. First the prices of the phones were found via Google Search and added into the python data frame. Then the

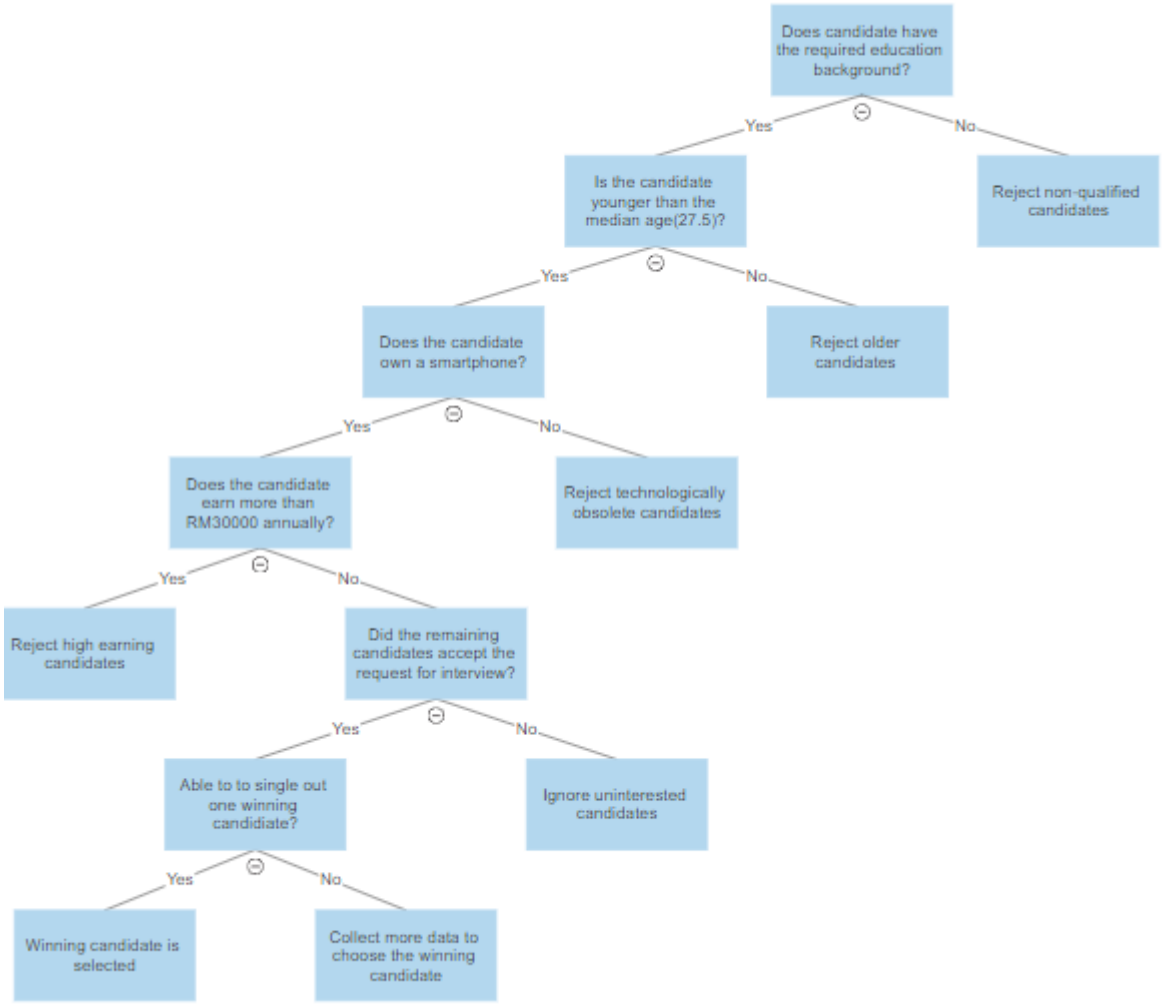
data frame was filtered so that the details of the ones who spent more than RM2000 of the phone remained. The most preferred way to request for interview was figured out by calculating the frequency of the way of requesting for interview. The way with the highest frequency will be displayed.

g. If we wanted to introduce the concept of employability (likelihood of someone being hired) in the mix, what kind of additional data would you factor into your calculations?

There are some additional data that can be factored in to determine the employability of an individual, those factors include presence of criminal records, this factor can help to filter out candidates who cannot abide by simple laws. Secondly, we can collect the data of the longest duration of employment under the same employer for the candidate, this factor can help us determine the reliability of the candidate. Next, we should factor in whether the candidate is prone to critical illnesses, this is so that the health of the candidate will not be a liability. Finally, the location of the candidate should also be factored in, provided that the job role is not remote, requires immediate hiring and relocating the candidate is unfeasible.

j. If you are required to build a decision tree data mining algorithm to predict a winning candidate based on a given job , kindly describe how you your approach would be like.

An assumption is made that the given job is a Junior Software Engineer. The decision tree is attached below. The first criteria to look at is education background, those who are underqualified will be rejected. Next, since this is a junior role, the candidates need to on the younger side, therefore, candidates that are older than the median age will be rejected. After that, because this a more technical role which requires a candidate that is not technologically challenged, candidates who do not possess a smartphone will be rejected. Furthermore, as this is a junior role, candidates who earn more than RM30000 annually are rejected. Once the candidates are short-listed, an interview request will be sent, candidates who accept the interview request will go through a series of questions and tests so that a winning candidate can be determined.



Decision tree