TASK 1

Saim Ali

1. The covariance matrix below is shown for each of the following pairs of attributes: Age Gap, Economic Similarity, Common Interests, and Divorce Score. After evaluating the covariance matrix, we can start seeing that the four attributes we are evaluating have relatively difference relationships when comparing pairs to each other. For example, the relationship between Age Gap and Economic Similarity has the covariance of -4.847… which shows us that there is a strong negative relationship between the two attributes where one s increasing and the other is decreasing. On the other hand, the relationship between Common Interests and Economic Similarity has the covariance of 48.213... which shows us there is a very strong positive covariance relation as both attributes increase with respect to each other. Using this covariance matrix, we can see that there is many different relationships going on in our dataset and we would have to evaluate each one separately to determine if two attributes are truly linked with each other or not.  
   Text, letter

   Description automatically generated
2. The scatter plot for the attributes Common Interests and Love is shown in the image below. Each dot in the scatter plot represents a data point and as you can see clearly that the data points are scattered without any clear pattern present. This obvious “random” scattering leads us to believe that there is no trend in the data between the attributes Common Interests and Love.  
     
   **Chart, scatter chart

   Description automatically generated**
3. Using the three selected attributes (Divorce, Desire to Marry, and Common Interests) from the data set we evaluate the data through histograms for BOTH the Marry and Divorce recommendations. The six histograms formulated with the data are shown below.

a) Divorce score with Marry recommendation  
Chart, histogram

Description automatically generated  
When evaluating the Divorce Score with the Marry recommendation we can see that the highest frequency of Divorce Score is from 1.3-1.4 roughly meaning that the most common Divorce Score is close to that value. We can also determine that any value below 2.0 on the Divorce Score leads to the recommendation of Marry instead of Divorce.  
  
  
b) Divorce score with Divorce recommendation  
Chart, histogram

Description automatically generated  
On the other hand, when evaluating Divorce Score with the Divorce recommendation we can see that there are three frequency spikes at 2.0, 2.3, and 2.8. We can also determine that any Divorce Score above 2.0 leads to a Divorce recommendation.  
  
  
  
c) Desire to Marry with Marry recommendation  
Chart, histogram

Description automatically generated  
Now we are looking at the Desire to Marry with the Marry recommendation. When looking at this histogram we can see that the frequency is pretty evenly spread while the recommendation is still marry which leads me to hypothesize that the Desire to Marry does not affect the fiunal recommendation too much.

d) Desire to Marry with Divorce recommendation   
Chart, histogram

Description automatically generated  
Building upon what is previously said we can see that the Desire to Marry is clearly lower on this histogram which could also mean that the lower the Desire to Marry is the more likely a Divorce recommendation is issued. In both histograms we can see that the highest frequency is around 40 for the Desire to Marry which is pretty close to 50 which is the halfway mark for the data.  
  
e) Common Interests with Marry recommendation   
Chart, histogram

Description automatically generated

Similarly, to the histograms before this one we can see the frequencies are pretty evenly spread other than the spike at 70 which we can also use to assume that there is not a strong connection between a higher Common Interest value and the recommendation to Marry.  
  
f) Common Interests with Divorce recommendation  
  
Chart, histogram

Description automatically generated  
Finally, looking at Common Interests with he Divorce recommendation we can see that there is actually more spikes in the data for Common Interests with the Divorce recommendation when compared to the other. So, we can use this also to determine that the Common Interests do not affect the recommendation in the end.

1. We will now be creating box plots for the Self Confidence attribute as well as two more for the Marry and Divorce recommendation.  
     
   a) First we do the Self Confidence Boxplot for all instances in the dataset. (shown below)  
   When evaluating this box plot, we can see the median to be slightly above 70 and no outliers present.  
   Chart, box and whisker chart

   Description automatically generated  
     
     
     
   b) Next we will show the box plot for Self Confidence with the Marry recommendation. When comparing this box plot with the previous one for the entire dataset we see little difference other than the median being slightly higher now but with little change nonetheless.  
   Chart, box and whisker chart

   Description automatically generated  
     
   c) Finally we look at the box plot for Self Confidence with the Divorce recommendation which also shows us there is little to no difference between the three boxplots. This means since there is little difference between the three box plots there is little correlation between the recommendation and Self Confidence value.  
   Chart, box and whisker chart

   Description automatically generated
2. For this task I chose to create supervised density plots for the following 3 pairs for attributes: Common Interests and Economic Similarity, Common Interests and Loyalty, and Economic Similarity and Loyalty. A density plot shows us the distribution of data over a period of values and we will use it here to visualize the shape of the distribution for each of the attributes in each pair then compare the overlap.
3. Common Interests and Economic Similarity  
   For this supervised density plot we can see that the density for Common Interests is far more condensed from around 50-100 while on the other hand Economic Similarity is vastly spread out from 0-125. Although we cannot really see any skewness for Common interests, for Economic Similarity we can see the data is skewed left. This means that the relationship between the mean and median is Mean<Median.  
   Chart, histogram

   Description automatically generated  
     
   b) Common Interests and Loyalty  
   This density plot is very similar to the one above but the difference lies in the Loyalty value as the min and max values are less spread so we can still see the left skewed pattern but it is less prominent.  
   **Chart, histogram

   Description automatically generated**  
     
   c) Economic Similarity and Loyalty  
   For this pair we see that the spread of the data is almost uniform to each other when compared in this density plot. This means that the values for Loyalty and Economic Similarity are very similar in spread as well as mean, mode, and median.  
   **Chart, line chart, histogram

   Description automatically generated**
4. For this question must first create   
   Density plot for Age Gap from main dataset  
   Chart, line chart

   Description automatically generated  
   Density plot for Age gap with Marry and Divorce recommendations:  
   Chart, histogram

   Description automatically generated  
   This histogram shows us a better understanding in the difference between the Age gap values for both Marry and Divorce recommendations. We can see for Marry the Age gap values highest frequency is right at the middle at 5 which creates an almost symmetrical shape while Divorce is a more asymmetrical shape we can see. We can assume that the average age gap between the datasets is from 4-6.  
     
   Density for Social Gap  
   Chart, line chart

   Description automatically generated  
     
   Density plot for Social Gap with Marry and Divorce recommendation  
   Chart, line chart

   Description automatically generated  
     
   When looking to compare the Marry and Divorce values for this density plot we can see they are mostly different in the middle 50% of their data. We can assume this to lead to a different mean and mode. We also see that the spike in Divorce is way higher at 40 compared to the low for Marry.
5. For this task I created a new dataset which contained the first 30 attribute’s z scores and then fit a linear model to predict the given divorce score as shown below. The R^2 value I calculated was .2113012 which tells us that only approximately 21.30% of the data is explained by the linear regression model. This means that this model is very weak when trying to predict the recommendation on whether a couple should marry or divorce. If the value was higher, we would be able to somewhat accurately recommend people on what to do but since it is low we can determine that the given attributes and data are not too useful when trying to come up with the recommendation. This data is backed up by the z scores given as the z score tells us how much each attribute differ from the given standard deviation value making it hard to predict the outcomes.  
     
   **Table

   Description automatically generated**
6. For this task I was assigned to create three decision trees with 20 or less nodes without including Independency and Education. With the accuracy ratings I got (.8, .7, .74) I found this data analysis portion to be very helpful when analyzing this data as it had high accuracy ratings meaning it would be quite accurate. I created and evaluated the three trees below:  
     
   a) Tree for all attributes excluding Education and Independency:   
   In this tree if the engagement time is greater than or equal to 3.282 we recommend to Marry so that is a strongly accurate way to tell the recommendation. The other attributes present are from 50-70 so those branches are more evenly spread making it harder to predict whether to Marry or Divorce.  
   A screenshot of a computer

   Description automatically generated with low confidence  
   **Timeline

   Description automatically generated**  
     
   b) Tree for Age Gap and Common Interests:  
   For this tree we evaluate the affects of Age Gap and Common Interests and can see that Age gap is a better attribute when predicting the outcome compared to the attribute of Common interests. A screenshot of a computer

   Description automatically generated with low confidence  
   **Diagram, timeline

   Description automatically generated**  
     
   c) Tree for Loyalty and Desire to Marry:  
   For Loyalty and Desire to Marry the tree showed that loyalty is a strong factor in the recommendation. If is low then the Divorce rate goes up greatly. On the other hand we can see that the Desire to Marry does not as clean cut at the Loyalty attribute.   
   A screenshot of a computer

   Description automatically generated with low confidence**Diagram, schematic

   Description automatically generated**
7. To conclude this study, I will go over the observations I came up with in the previous tasks. Throughout this project I was set out to see what will and won’t affect the recommendation on predicting a divorce/marry rating. I started with the hypothesis that Loyalty, Economic Similarities, and Desire to Marry will be the most important when coming up the predictions for this dataset. I was surprised to say the least that the Economic Similarities attribute did not play a bigger factor when trying to decide what to recommend. I have learned that in a marriage maybe economic factors are outweighed by other factors such as Loyalty. Loyalty was one I predicted to be a heavy influencer and was partially correct but still surprised when seeing the numbers. This study has shown me that it is far more difficult to predict the outcome of a marriage from just looking at some values than it originally seemed to me. From personality traits to Religion compatibility, this dataset proves there are so many factors when coming up with your prediction in the end. I believe in the future even if a more extensive study were to occur with a more detailed dataset, we will still be unable to correctly recommend whether a couple should Marry or Divorce despite the values given to us.