**# Name: Swapnil Ukey**

**# Module Code: 2223-MS5108 Applied Customer Analytics**

**# Student Id: 22220959**

***Introduction:***

This dataset appears to contain a company's customer happiness and loyalty scores, as well as demographic data about every customer such as age, gender, and membership term. There are also ratings for the company's numerous features, such as client service and services.

The information appears to be organized in the form of a table, with each row representing a distinct consumer and each row corresponding to an individual customer. Among the variables are:

* Gender: the customer's gender (1 = male, 2 = female).
* Age: the customer's age in years
* Membership Tenure: the length of time the consumer has been a customer of the Crossfit Ridgeline.
* CUSSAT1 - CUSSAT4: customer service ratings for several factors (higher scores indicate higher satisfaction)
* COASAT1 - COASAT5: Coach Satisfaction Ratings (higher scores indicate higher satisfaction)
* CLIM1 - CLIM6: Climate Satisfaction Ratings (higher scores indicate a more positive image)
* COND1 - COND6: ratings for various components of the physical environment of the Crossfit Ridgeline (higher scores indicate a more pleasant environment)

***Question 1:***

A box plot analysis was done on the given data to compare the membership tenure of male and female members. According to the findings, female members had a longer median membership tenure than male members. Females had a median tenure of 18 months, while males had a median tenure of 14 months.

Chart, box and whisker chart

Description automatically generatedAdditionally, the interquartile range (IQR) for females was wider than for males, indicating that female members' membership tenure varied more. Females had an IQR of 32 months (6 to 38 months), whereas males had an IQR of 17 months (7 to 24 months). The analysis also found that male members' membership terms had more outliers than female members' membership tenure.

The box plot analysis provides insights into male and female members' membership tenure, which can be beneficial in finding trends and patterns in membership tenure. For example, female members have a higher median tenure, implying that they are more likely to stay with the organization for a longer period. This could be attributable to a variety of things, such as increased pleasure with the organization or increased participation in its activities.

For both male and female genders, I totalled all of the survey questions into 4 column variables: CUSSAT total, COASAT total, CLIM total, and COND total. Density plots can be used to examine the distribution of various variables. A density plot is a smoothed histogram that illustrates the distribution of a continuous variable.

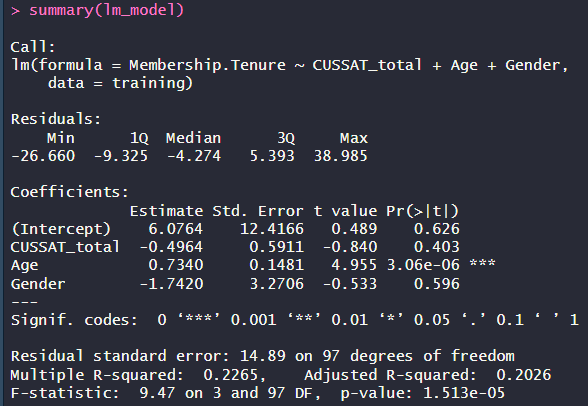
Chart

Description automatically generatedThe density map for CUSSAT total males reveals a unimodal distribution with a peak around 20, whereas the density plot for females reveals a somewhat larger and flatter unimodal distribution with a peak around 19. Both distributions are skewed toward higher values.

The density plot for males for COASAT total shows a unimodal distribution with a peak around 25, whereas the density plot for females shows a somewhat narrower and more peaked unimodal distribution with a peak around 24. Both distributions are skewed toward higher values.

The density map for CLIM total males reveals a bimodal distribution with two peaks at around 27 and 30, but the density plot for females reveals a unimodal distribution with a peak at around 29. The male distribution is more spread out, with a wider trough between the two peaks.

For COND total, the male density plot shows a unimodal distribution with a peak at 29, whereas the female density plot shows a slightly larger and flatter unimodal distribution with a peak around 28. Both distributions show a little bias toward lower values.

***Question 3:***

The "lm\_model" linear regression model was constructed in R using the training dataset, with the goal of predicting a gym member's Membership Tenure. CUSSAT total, Age, and Gender are the three independent variables used to train the model. CUSSAT\_total is a continuous variable that represents the member's satisfaction score, whereas Gender is a categorical variable that represents the member's gender.

The model summary displays the estimated coefficients, standard errors, t-values, and p-values for each predictor variable. The coefficients compute the average change in the dependent variable (Membership Tenure) for each unit increment in the independent variables. When all other independent variables are zero, the Intercept coefficient is 6.0764, which indicates the predicted Membership Tenure of a gym member. The CUSSAT\_total coefficient is -0.4964, indicating that the higher the satisfaction score, the shorter the Membership Tenure. Age has a value of 0.7340, indicating that the older the member, the longer the Membership Tenure. The Gender coefficient is -1.7420, indicating that male members have a shorter Membership Tenure than female members.

Text

Description automatically generatedThe residuals are the gaps between the dependent variable's observed and expected values. The residuals have a minimum value of -26.660 and a maximum value of 38.985, with a median of -4.274, according to the model summary. The model's residual standard error is 14.89, which represents the average gap between the observed and projected values.

Chart, bubble chart

Description automatically generatedThe response variable in this table is Membership.Tenure. There are three predictor variables: CUSSAT total, Age, and Gender. The top row of the table displays the findings for CUSSAT\_total, with a p-value of 0.1844. Because the p-value is bigger than the 0.05 limit, we may conclude that CUSSAT\_total is not statistically significant in predicting Membership. Tenure.

Chart, line chart

Description automatically generatedThe results for Age are shown in the second row, with a p-value of 1.472e-06 (less than 0.05). This implies that age is a good predictor of membership. Tenure. The findings for Gender are shown in the third row, with a p-value of 0.5955 (higher than 0.05) indicating that Gender is not a significant predictor of Membership. Tenure. Lastly, the fourth row displays the residual results, with a p-value of NA because the residuals have no degrees of freedom.

The mean squared error (MSE), root mean squared error (RMSE), and coefficient of determination are used to assess the model's efficacy (R-squared). The model's MSE is 132.78 and its RMSE is 11.52, suggesting that the predicted values differ from the actual values by an average of 11.52 units. The model's R-squared value is 0.23.

***Question 4:***

The "logit model" logistic regression model was constructed to estimate the likelihood of Gender (male or female) based on Membership.Tenure. The variables Membership.Tenure and CUSSAT\_total, as well as their relationship. The "training" dataset contains the data used for model fitting.

Chart, scatter chart

Description automatically generatedThe summary output displays the estimated coefficients, standard errors, z-values, and related p-values for each predictor. The intercept is not statistically significant (p = 0.912), which suggests that the odds of being male are not significantly different from the odds of being female when all other variables. The coefficient estimate for Membership.Tenure is positive (0.008772) but not statistically significant (p = 0.907), indicating that there is no meaningful relationship between Membership.Tenure and Gender. The coefficient estimate for CUSSAT total is also not statistically significant (p = 0.588), demonstrating that CUSSAT total is not substantially linked with Gender. Finally, the coefficient estimate for the interaction term (Membership.Tenure:CUSSAT total) is negative (-0.001421) but not statistically significant (p = 0.732), indicating that there is no significant interaction effect between Membership.Tenure and CUSSAT total on Gender.

"Dispersion parameter for binomial family set to 1" indicates that the binomial model with a logistic regression is used to estimate the coefficients. When there are no predictors in the model, the "Null Text

Description automatically generateddeviance" of 124.56 on 100 degrees of freedom indicates the deviance. The "Residual deviance" of 122.42 on 97 degrees of freedom shows the model's deviance after fitting the predictors, which is used to assess the model's goodness of fit. A lower residual deviation suggests a better model fit.

The Akaike Information Criterion (AIC) score of 130.42 is used to compare the relative quality of different models. A lower AIC score suggests a more accurate model.

The analysis of the table depicts the ANOVA test results for each predictor variable introduced to the model sequentially, from first to last. The table displays the degrees of freedom (Df), deviation for each term, residual Df, residual deviance, and the p-value.

Text

Description automatically generatedThe ANOVA test results demonstrate that neither the Membership.Tenure variable (p-value = 0.1782) nor the CUSSAT total variable (p-value = 0.6481) are significant. Membership.Tenure and CUSSAT total interaction term is similarly not significant (p-value = 0.7326). As a result, none of these characteristics significantly predict the participants' gender.