**ASSIGNMENT 1**

**7357- ANALYTICS FOR DECISION SUPPORT**

1. **1. What is the linear regression equation produced by your model? Make sure to give the coefficients for each of the independent variables.**

ANS: lm(formula = ElantraSales ~ Unemployment + Queries + CPI.Energy +

CPI.All, data = MyData)

Intercept = 95385.3 -3179.90 \* Unemployment + 19.03 \*Queries + 38.51\*CPI\_Energy – 297.65 \*CPI\_All

**2.Whats is the R2 of the model?**

ANS**:** R2 = 0.4282

**3. Which variables are significant? What does this tell you about the model?**

ANS: None of the variables are significant, and hence the model is not Proper to predict sales. As Pr(>|t|) less than 5%. We need to improve it.

We can try to add Month as an independent variable and check the model.

1. **Seasonality in the Problem:**
2. **What is the R 2 of the model and the regression equation of the model?**

ANS: The month is added as the independent variable. And re running the regression equation.

**Regression Equation:**

Intercept= 148330 – 4137 \* unemployment + 21.19 \* Queries + 54.18 \*CPI\_Energy -517.99\* CPI\_All +110.69 \*Month

**Multiple R squared:** R2 = 0.4344 **Adjusted R2**= 0.3402

None of the variables look significant for this model as we don’t have any variable having probability Pr(>|t|) less than 5%. However, Queries lis likely significant variable with probability of 8.71%.

1. **Is this the right model this variable? What would happen if we made month a categorical variable?**

Ans: It is not the right model. The R2 for the model after adding the month, is not changing significantly. It is because the month is calculated as an Integer and not a factor. In this case, the order of the months is in increasing values. Hence, we have factor the month variable. This way we could improve the model and we can predict sales with the seasonality.

1. **Month as a Factor:**
2. **Describe your new model. What is the regression equation? What is the R2  ? Which Variables are Significant**?

The month is factored using the factor keyword. This way we can predict sales with months that helps in understanding the sales better. And how the seasonality effects the sales.

Coefficients of the linear regression model

(Intercept) 312509.280  
Unemployment -7739.381  
Queries -4.764  
CPI.Energy 288.631  
CPI.All -1343.307  
MonthFebruary 2254.998  
MonthMarch 6696.557  
MonthApril 7556.607  
MonthMay 7420.249  
MonthJune 9215.833  
MonthJuly 9929.464  
MonthAugust 7939.447  
MonthSeptember 5013.287  
MonthOctober 2500.184  
MonthNovember 3238.932  
MonthDecember 5293.911

**Multiple R-squared:** 0.8193 = 81.93%

**Significant variables:** independent variables that are significant are, Unemployment, CPI.Enerfy and CPI.All are as these have Pr(>|t|) less than 5%. Only Queries is insignificant with probability 71.6%.

Feb, Oct and Nov are insignificant in seasonality that means we can predict that the sales are high in other months excluding the insignificant months.

1. **Using these predictions on the test set. Month should be factored in the test set. What is the R2 of the model on the test set?**

Ans: The month in the testing set also must be factored, for further predictions on sales with respect to other independent variables.

The month in the testing set is factored.

Factoring the Month Variable:

ElantraTesting$Month <- factor(month.name[ElantraTesting$Month] ,levels = month.name[1:12] )

**The R2 Value: 0.7426902**

1. **From what you saw in this problem, what can you conclude about predicting elantra sales? Do you think these conclusions generalize to predicting sales for other products?**

Ans: The sales of the cars are dependent on the months of the year. We could clearly see that sales in jan and December are higher when compared to other months. January introduces brand new models and December has offer to sell their stock and keeping he inventory open for the new cars that are introduce the following year in the month of January.

Yes, these predictions can be applied to sales of other products that have similar sales.

1. **If you could collect additional independent variables for this problem, which variables do you think would be useful for predicting sales**

Additional independent variables that could affect the sales are population, economic conditions, competitors in the industry.