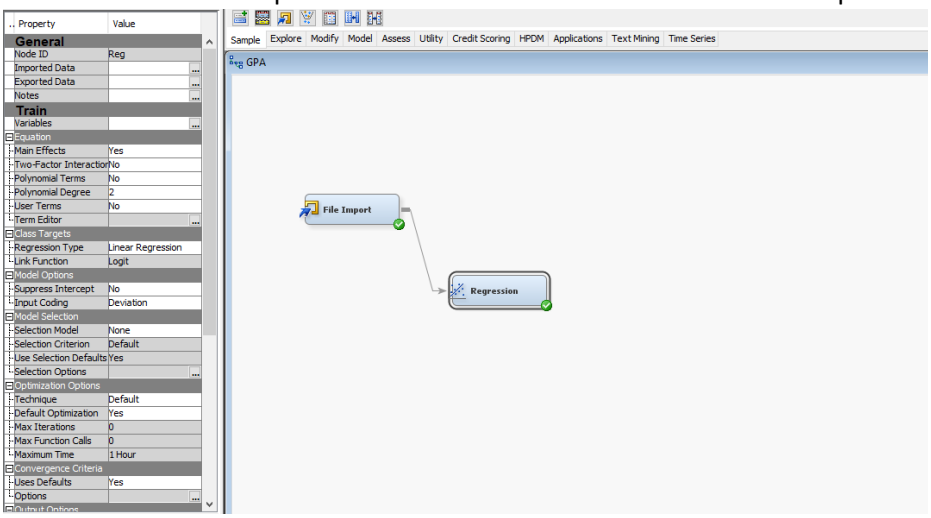
DSCI 5240- Data Mining Homework 1.

Group Members: Masters Group 9

1. Oluwaseyitan Awojobi
2. Mohamed Elgendy
3. Ramya Koya
4. Karthik Sagar Tadi
5. Vijay Tulluri

Visualization and Interpretation of Results:

1. Below is a Visual representation of the Screen before results were printed **for questions 1-5**



6. The information below tells us the regression model is significant with a P value of <.0001 and a significantly large F value of 80.50. We can reject the null hypothesis and conclude that this model provides a better fit.

Output

```
38                Model Information
39
40    Training Data Set             WORK.EM_DMREG.VIEW
41    DMDB Catalog                  WORK.REG_DMDB
42    Target Variable               GPA (GPA)
43    Target Measurement Level      Interval
44    Error                         Normal
45    Link Function                 Identity
46    Number of Model Parameters    3
47    Number of Observations        85
48
49
50                               Analysis of Variance
51
52                                    Sum of
53    Source                 DF        Squares     Mean Square    F Value    Pr > F
54
55    Model                   2       10.385763        5.192881      80.50    <.0001
56    Error                  82        5.289348        0.064504
57    Corrected Total        84       15.675111
58
59
60                 Model Fit Statistics
61
62    R-Square        0.6626     Adj R-Sq        0.6543
63    AIC          -230.0413     BIC          -227.8245
64    SBC          -222.7133     C(p)            3.0000
65
66
```

7. The F test for this model is statistically significant with a p value less of <.0001 which is less than 0.005. This shows that the co-efficient are collectively sufficient to explain a reasonable part in the variability of the GPA to be considered significant. We can infer from the F test that we have a better prediction with the model than with the mean of the response and our R-squared is significantly different from zero.

```
                               Analysis of Variance

                                    Sum of
Source                 DF        Squares     Mean Square    F Value    Pr > F

Model                   2       10.385763        5.192881      80.50    <.0001
Error                  82        5.289348        0.064504
Corrected Total        84       15.675111


                 Model Fit Statistics

R-Square        0.6626     Adj R-Sq        0.6543
AIC          -230.0413     BIC          -227.8245
SBC          -222.7133     C(p)            3.0000


              Analysis of Maximum Likelihood Estimates

                                Standard
Parameter    DF    Estimate       Error    t Value    Pr > |t|

Intercept     1     -0.5438      0.2796      -1.94      0.0552
Age           1      0.1343      0.0140       9.58      <.0001
GRE           1     0.00144    0.000369       3.90      0.0002
```

8. The predictor variables are significant as shown below with Age having a P value of <0.0001 and GRE being 0.0002 which are both less than 0.005

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.5438	0.2796	-1.94	0.0552
Age	1	0.1343	0.0140	9.58	<.0001
GRE	1	0.00144	0.000369	3.90	0.0002

9. With the same data seen above, AGE is seen to be the predictor which has the most influence on GPA since it has a higher co-efficient of 0.1343. That is a 1 unit increase in Age increases the GPA by 0.1343.

10. The R-Squared is valued at 0.6626 and Adjusted R-Squared is valued at 0.6543
11. The R-Squared tells us 66.26% of the variation is explained with the information with the variable GPA

12. GPA= Bo + B1Age + B2Gre
    GPA = -0.5438 + 0.1343(26) + 0.00144(680)
    GPA = -0.5438 + 3.4918 + 0.9792
    GPA = 3.9272
    GPA = 3.93