SEIS632-02-Assignment2
Shivali Dalmia

Question1: Which variables are included in the final model?
Answer1: Total 6 variables are included in the final model, namely IMP_DemAffl,
IMP_DemAge, IMP_DemGender, M_DemAffl, M_DemAge, M_DemGender.

```
                        Summary of Stepwise Selection

              Effect           Number    Score      Wald              Validation
     Step     Entered      DF     In    Chi-Square  Chi-Square  Pr > ChiSq  Error Rate

       1      IMP_DemAffl   1      1    1379.1010              <.0001      11096.5
       2      IMP_DemAge    1      2     783.0549              <.0001      10338.0
       3      IMP_DemGender 2      3     526.8359              <.0001       9619.3
       4      M_DemAffl     1      4      10.3319              0.0013       9618.9
       5      M_DemAge      1      5       5.7905              0.0161       9610.1
       6      M_DemGender   1      6       4.4363              0.0352       9605.8
```

The selected model, based on the error rate for the validation data, is the model trained in Step 6. It consists of the following effects:

Intercept IMP_DemAffl IMP_DemAge IMP_DemGender M_DemAffl M_DemAge M_DemGender

Question2: What is the validation ASE?
Answer2: 0.138204

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
TargetBuy	Organics Purchase Indicator	_AIC_	Akaike's Information Criterion		9758.609	
TargetBuy	Organics Purchase Indicator	_ASE_	Average Squared Error	0.139545		0.138204

Question3: Do the selected variables change?
Answer3: Yes, now IMP_LOG_DemAffl and M_LOG_DemAffl is considered in place of IMP_DemAffl and M_DemAffl respectively.

```
                        Summary of Stepwise Selection

               Effect            Number     Score        Wald                 Validation
     Step      Entered       DF     In     Chi-Square   Chi-Square   Pr > ChiSq   Error Rate

       1       IMP_LOG_DemAffl  1     1    1131.8911                 <.0001       11205.0
       2       IMP_DemAge       1     2     798.6031                 <.0001       10426.1
       3       IMP_DemGender    2     3     532.9943                 <.0001        9703.8
       4       M_LOG_DemAffl    1     4      12.8524                 0.0003        9702.1
       5       M_DemAge         1     5       5.8363                 0.0157        9693.0
       6       M_DemGender      1     6       4.9880                 0.0255        9688.6
```

Question4: How about the validation ASE?
Answer4: The validation ASE remains same, i.e., 0.138204.

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
TargetBuy	Organics Purchase Indicator	_AIC_	Akaike's Information Criterion		9758.609	
TargetBuy	Organics Purchase Indicator	_ASE_	Average Squared Error	0.139545		0.138204

Polynomial Regression Model

Assumptions:
1. There is a direct relation between how much a member spends and the total time spend as loyalty card member.(PromTime ~ PromSpend)
2. There is a direct relation between time spend as loyalty card member and loyalty status i.e., tin, silver, gold, or platinum.(PromTime ~ PromClass)
3. A parabola-shaped relationship between organic purchase indicator and Age.()

Adding above mentioned interaction terms to the regression model.

Question5: How does the validation average squared error for the polynomial model compared to the original model?
Answer5: The validation ASE reduces to 0.134909 from 0.138204 after addition of few interaction terms as compared to the original regression model.

Fit Statistics						
Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
TargetBuy	Organics Purchase Indicator	_AIC_	Akaike's Information Criterion	9585.364		
TargetBuy	Organics Purchase Indicator	_ASE_	Average Squared Error	0.137208	0.134909	

The selected input variables also change.

```
                    Summary of Stepwise Selection

              Effect            Number      Score       Wald                  Validation
     Step    Entered         DF     In    Chi-Square  Chi-Square   Pr > ChiSq  Error Rate

      1      IMP_LOG_DemAff1   1      1    1131.8911                  <.0001     11205.0
      2      IMP_DemAge        1      2     798.6031                  <.0001     10426.1
      3      IMP_DemGender     2      3     532.9943                  <.0001      9703.8
      4      IMP_DemAge*IMP_DemAge  1   4   159.1181                  <.0001      9502.9
      5      M_DemAge          1      5      30.7118                  <.0001      9459.5
      6      M_LOG_DemAff1     1      6      11.2845                  0.0008      9455.9
      7      M_DemGender       1      7       5.3626                  0.0206      9451.6
```