House Sale Price & Bonus Eligibility Prediction

```
*Question 1 - Use %LET statements to name the macro
variables and set their values. The macro variables are
referred to in the SAS code as &categorical and
&interval, to distinguish those names from
those of variables;
```

***Part 1;**

```
%let categorical = Central_Air Foundation_2 Garage_Type_2 Heating_QC Mo_Sold
House_Style House_Style2 Masonry_Veneer;

%let interval = Age_Sold Basement_Area Bedroom_AbvGr Bonus Fireplaces Full_Bathroom
Garage_Area Gr_Liv_Area
Half_Bathroom Log_Price Lot_Area Overall_Cond Overall_Cond2 Overall_Qual Overall_Qual2
SalePrice Season_Sold Total_Bathroom
Year_Built Yr_Sold;
```

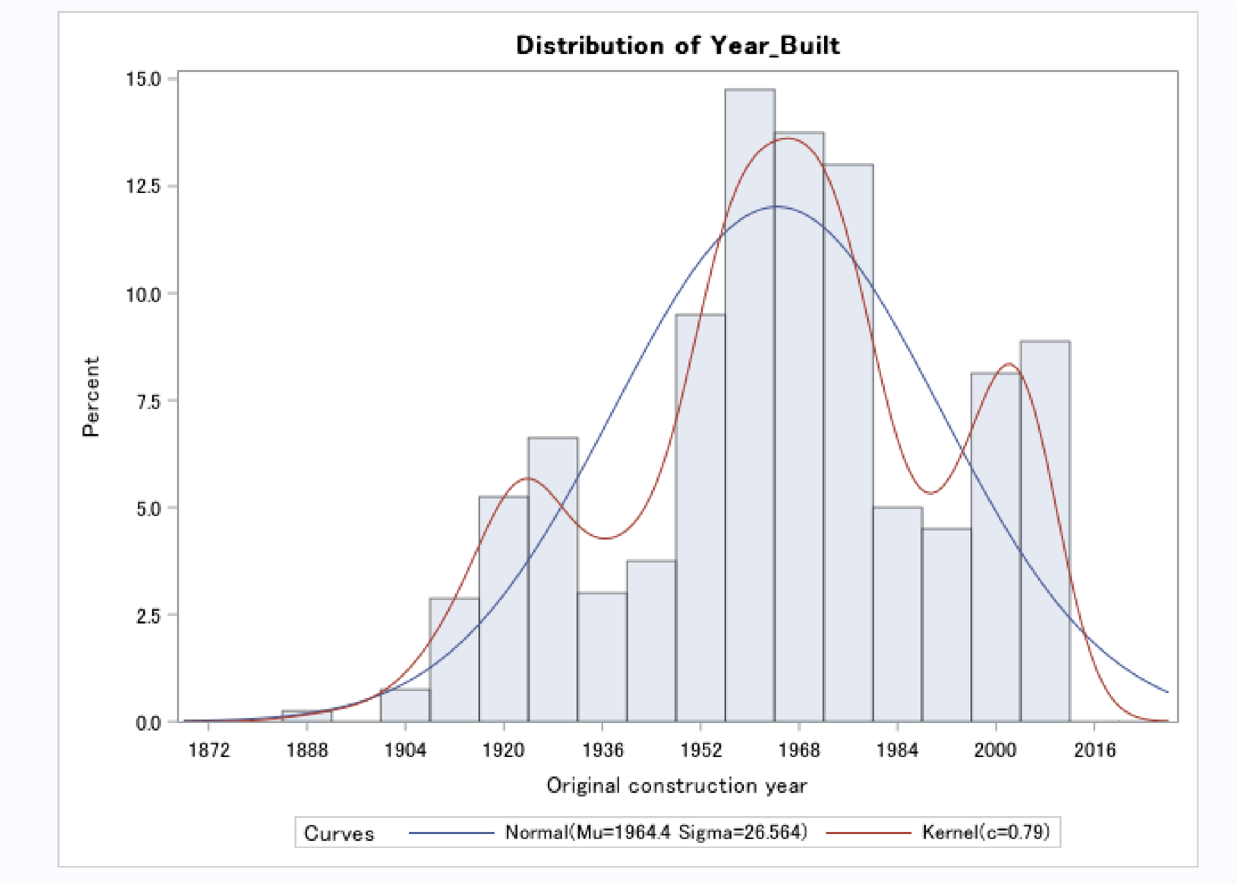
***Here, we split the character and numeric variables into the respective &categorical and &interval %let statements. In other words, we let the character variables into the categorical group, and the numeric ones into the interval group.**

```
*Question 2 - Use PROC UNIVARIATE to generate plots and
descriptive statistics for continuous variables and
PROC FREQ to generate plots and tables for categorical
variables.;
```

```
title "Categorical Variable Descriptive Statistics";
proc freq data = orion.team8;
	tables &categorical / plots = freqplot (type = bar);
run;
```

***The frequency plots and tables show the distribution of various sub-categories included within the variables part of the constructed houses. For example, the frequency table and bar graph above shows a distribution of whether or the constructed houses include masonry veneer walls or not;**

```
proc univariate data=orion.team8 noprint;
 var &interval;
 histogram &interval / normal kernel;
 title "Interval Variable Distribution";
run;
```



This plot shows a normal and kernel distribution among the construction years of the respective houses. A fitted normal distribution shows the approximate continuous probability of the construction years, whereas the kernel hits the precise data points in order to better interpret the data.

```
*Question 3 - Use the TTEST procedure to test whether
the mean of SalePrice is $135,000 in the data set. Is
the
mean value in the sample statistically significantly
different from $135,000 at an alpha level of
0.05?;
```

```
proc ttest data=orion.team8 h0=135000
 plots(only shownull)=interval;
 var SalePrice;
 title 'Testing Whether the Mean Sale Price is 135000';
run;
```

***P-value < alpha => 0.0002 < 0.05 => Yes, the mean value is statistically significant;**
***Here, the p-value, 0.0002 is less than the alpha of 0.05. Also, the mean value is greater than 135000, at 139,834. Hence we reject the null hypothesis or H0 = 135,000 for Ha != 135000, and the mean value is statistically significant;**

*Question 4 - Use the TTEST procedure to test whether the mean of SalePrice is the same for homes with masonry veneer and those without. Provide your insights.

Using TTEST to find whether the mean of SalePrice is the same for homes with masonry veneer and those without */

```
proc ttest data=orion.team8 plots(only shownull)=interval;
	class Masonry_Veneer;
	var SalePrice;
	title "Two-sample t-Test Comparing homes with masonry veneer and those without";
run;
```

The above code gives the following output where looking at the F-test, the P-value(0.8387) is > 0.05 which tells that we accept the variances are equal.
While looking at the t-test table we can see the P-value is significant for the pooled method and we reject the null hypothesis. Hence, there is significant difference between Saleprice with homes with and without masonry veneer

Two-sample t-Test Comparing homes with masonry veneer and those without

The TTEST Procedure

Variable: SalePrice (Sale price in dollars)

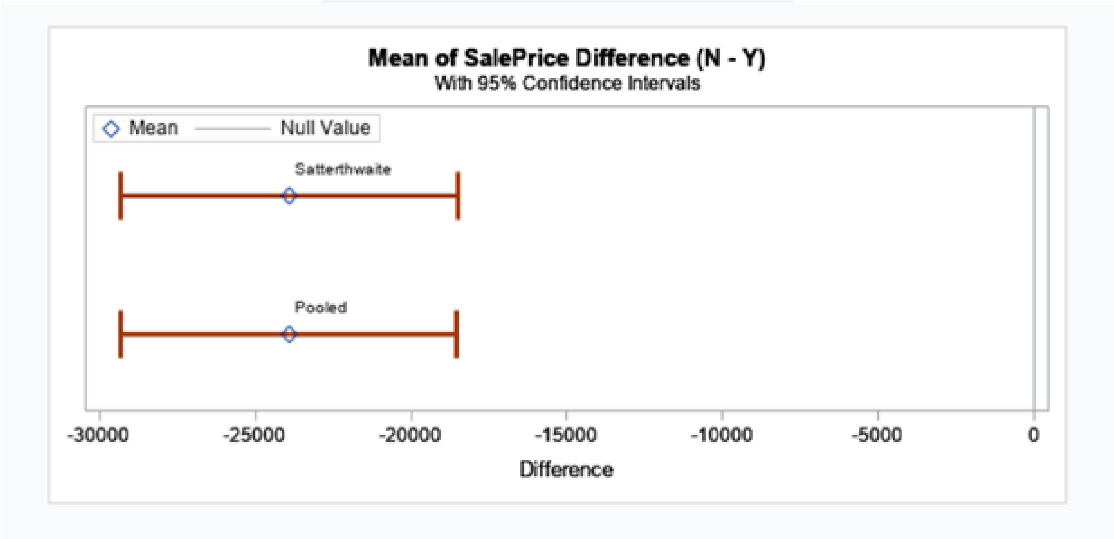
Masonry_Veneer	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
N		565	132759	35140.5	1478.4	39300.0	271500
Y		233	156687	35506.9	2326.1	61500.0	285000
Diff (1-2)	Pooled		-23927.8	35247.7	2744.3		
Diff (1-2)	Satterthwaite		-23927.8		2756.2		

Masonry_Veneer	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
N		132759	129856	135663	35140.5	33204.1	37318.5
Y		156687	152104	161270	35506.9	32549.1	39060.5
Diff (1-2)	Pooled	-23927.8	-29314.7	-18540.9	35247.7	33598.1	37068.8
Diff (1-2)	Satterthwaite	-23927.8	-29345.1	-18510.5			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	796	-8.72	<.0001
Satterthwaite	Unequal	428.51	-8.68	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	232	564	1.02	0.8387

The same thing can be inferred from the means from the confidence interval plot. The null value is not included(which is our hypothesized value) in the plot and with 95% confidence we can say that there is significant difference around (-24000)




```
*Question 5 - Create scatter plots to show
relationships between continuous predictors and
SalePrice and
comparative box plots to show relationships between
categorical predictors and SalePrice using Macro
program;
```

```
 %let interval = Lot_Area Year_Built Gr_Liv_Area Garage_Area Basement_Area
Deck_Porch_Area Age_Sold;

ods graphics / reset=all imagemap;
ods select scatterplot;
proc corr data=orion.team8 rank
          plots(only)=scatter(nvar=all ellipse=none);
  var &interval;
  with SalePrice;
  title "Scatter Plots";
run;

%macro Boxplot(name);
proc sgplot data=orion.team8;
vbox SalePrice/ category=&name connect=mean;
run;
%mend Boxplot;

%Boxplot(House_Style);
```

```
*Question 6 - Examine the relationships between
SalePrice and the continuous predictor variables in the
dataset;
```

```
* continuous predictor variables in our dataset are considered as follows;
%let variable= Lot_Area Gr_Liv_Area Garage_Area Basement_Area Deck_Porch_Area
Age_Sold;

proc corr data=orion.Team8 PLOTS=SCATTER(NVAR=all);
 var &variable;
 with SalePrice;
 title "Correlations Plots";
run;
```

/* Output of question 6 is as follows
Pearson Correlation Coefficients, N = 800

	Lot_Area	Gr_Liv_Area	Garage_Area	Basement_Area	Deck_Porch_Area	Age_Sold
SalePrice	**0.22170**	**0.59629**	**0.50223**	**0.61098**	**0.40526**	**-0.63645 */**

```
*Question 7 - Perform a simple linear regression
analysis with SalePrice as the response variable, and
one of
the significant predictors. Explain why you have chosen
that variable. What's the prediction
equation?;
```

*** Selecting Age_Sold as the most significant predictors as its negatively correlated with SalePrice with highest corr coefficient as -0.63 as compared to other independent variables**
Regression analysis code of Sale Price with Age sold --;

```
proc reg data=orion.Team8;
model SalePrice=Age_Sold  / R P VIF SLENTRY = 0.01 ;
run;
```

***Answer- The prediction equation is SalePrice = 178230 - 883.45296(Age_Sold);**

```
*Question 8 - Perform a regression model of SalePrice
with Lot_Area and Basement_Area as predictor
variables;
```

***Regression analysis code considering only independent variables (Lot_Area & Basement_Area)--;**

```
proc reg data=orion.team8;
model SalePrice=Lot_Area Basement_Area;
plot RESIDUAL. * Lot_Area Basement_Area;
run;
```

***70892 = 1.07517(Lot_Area)+66.19581(Basement_Area)**
R squared value is 0.3835 and both p-values are less than 0.01;

```
*Question 9 - Call to macro to run SELECT for the
options SL, AIC, BIC, AICC, and SBC
and compare the selected models from the output. Does
the significance level for entry
into and staying in the model have any impact when you
use options other than SL?
Which variables stay in the model for each 5 options?
Which selection methods and
criteria would you recommend?
```

Regression of salePrice with all the interval variables using GLMSELECT with select options SL, AIC, BIC, AICC, SBC */

```
%let interval = Lot_Area Year_Built Gr_Liv_Area Bedroom_AbvGr Fireplaces
Garage_Area                                    Mo_Sold Yr_Sold Basement_Area
Total_Bathroom Deck_Porch_Area Age_Sold;

%macro modelsel(mod, type, slent=0.05, slst =0.05);

        title "&mod and &type";
        proc glmselect data=orion.team8 plots=all;
                model SalePrice = &interval / selection = &mod
                                              details = steps
                                              select = &type
                                              slentry = &slent
                                              slstay = &slst;
        run;
%mend modelsel;

%modelsel(stepwise,SL)
%modelsel(stepwise, AIC)
%modelsel(stepwise, BIC)
%modelsel(stepwise, AICC)
%modelsel(stepwise, SBC)
```

When SL is used as select option:

stepwise and SL

The GLMSELECT Procedure
Selected Model

The selected model is the model at the last step (Step 9).

Effects:	Intercept Lot_Area Year_Built Gr_Liv_Area Bedroom_AbvGr Fireplaces Garage_Area Basement_Area Total_Bathroom Deck_Porch_Area

Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Value
Model	9	8.288547E11	92094969633	282.83
Error	790	2.57241E11	325621540	
Corrected Total	799	1.086096E12		

Root MSE	18045
Dependent Mean	139834
R-Square	0.7632
Adj R-Sq	0.7605
AIC	16493
AICC	16493
SBC	15738

Parameter Estimates				
Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	-910066	57462	-15.84
Lot_Area	1	0.885539	0.183098	4.84
Year_Built	1	480.373552	29.875897	16.08
Gr_Liv_Area	1	53.669608	3.721637	14.42
Bedroom_AbvGr	1	-4837.903390	1152.960650	-4.20
Fireplaces	1	6600.151305	1232.804966	5.35
Garage_Area	1	26.029497	4.328398	6.01
Basement_Area	1	25.789570	2.320196	11.12
Total_Bathroom	1	5688.414830	1270.363108	4.48
Deck_Porch_Area	1	26.960426	5.263665	5.12

The R-square value is : 0.7632 and a total of 10 variables are considered including intercept

```
Lot_Area Year_Built Gr_Liv_Area Bedroom_AbvGr Fireplaces Garage_Area
Basement_Area Total_Bathroom Deck_Porch_Area
```

All the select option produces the same result.So any of the select option can be used.

/* *Question 10 - Invoke PROC REG with the plots option using rsquare adjrsq cp to produce a regression of SalePrice on all the other interval variables in the data set. Which model you would suggest, and why?
Regression of salePrice with all the interval variables and comparing the models with selection options R-square, Adjusted R-square and Cp   */

```
ods graphics on;
        title "Regression of SalePrice with selection using R-square, Adjusted R-square and Cp";
        proc reg data =orion.Team8 plots(only) = (rsquare cp adjrsq);
        model SalePrice = Lot_Area Year_Built Gr_Liv_Area Bedroom_AbvGr
                          Fireplaces Garage_Area Mo_Sold Yr_Sold
                          Basement_Area Total_Bathroom Deck_Porch_Area
        Age_Sold                                   /selection = rsquare cp adjrsq;
run;
quit;
ods graphics off;
```

The following diagrams shows the output of the model with different selection methods. It is obvious that the values of r-square and adjusted r-square are maximum when all the variables are used and the value of cp is minimum for the same, however after 8 variables minimum value of information is added to the model. Hence, we can suggest we can use any of the three models as all of them are producing the same results.

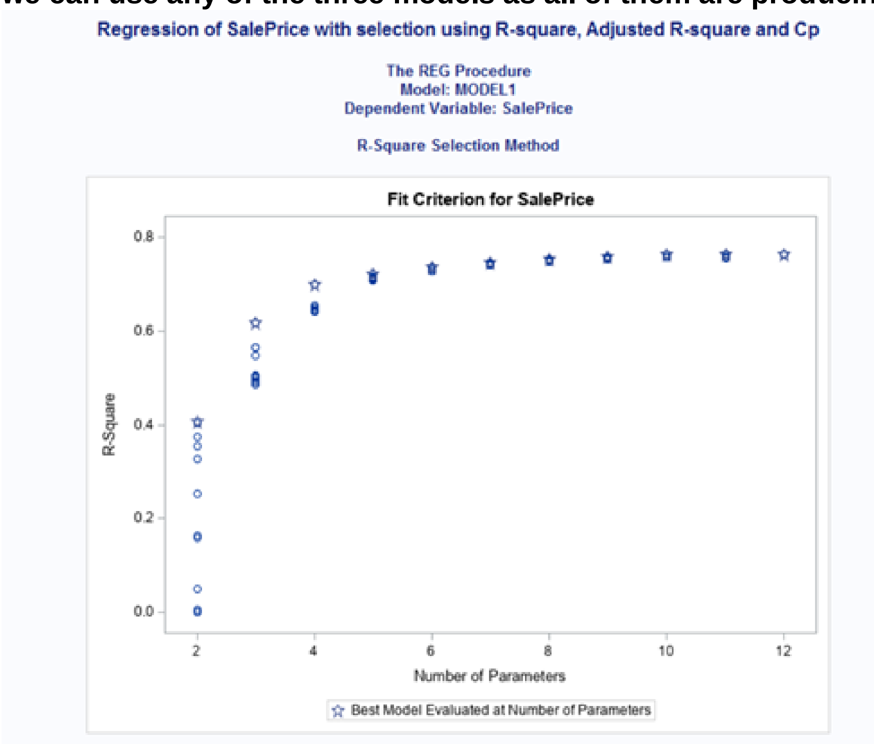


Fig 1: Plot with R-square selection method

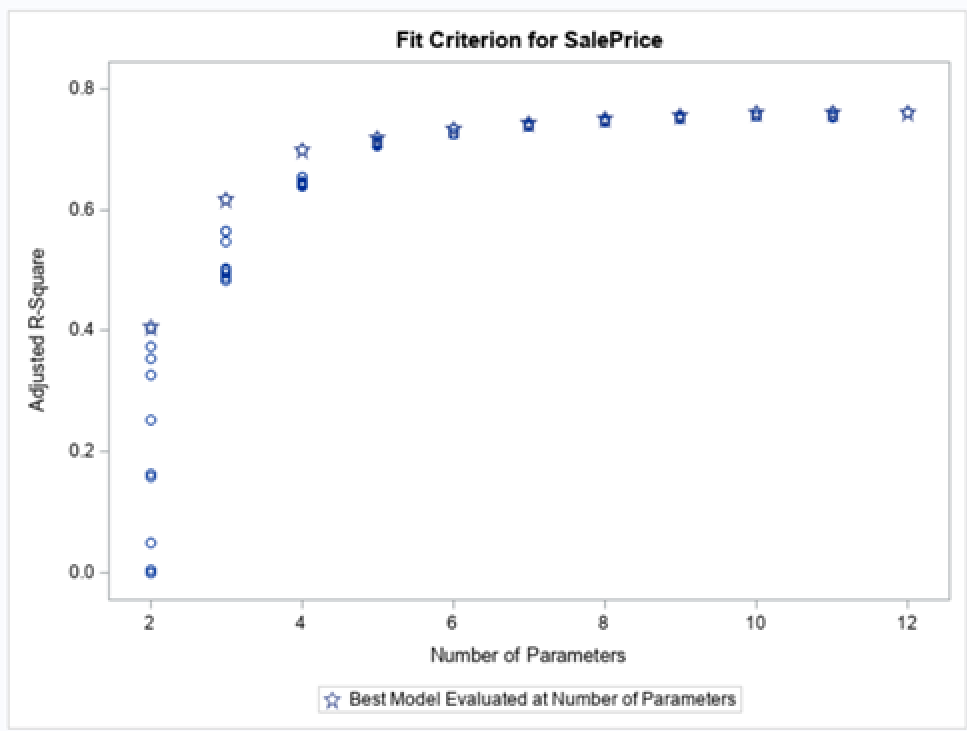


Fig 2: Plot with adjusted R-square selection method

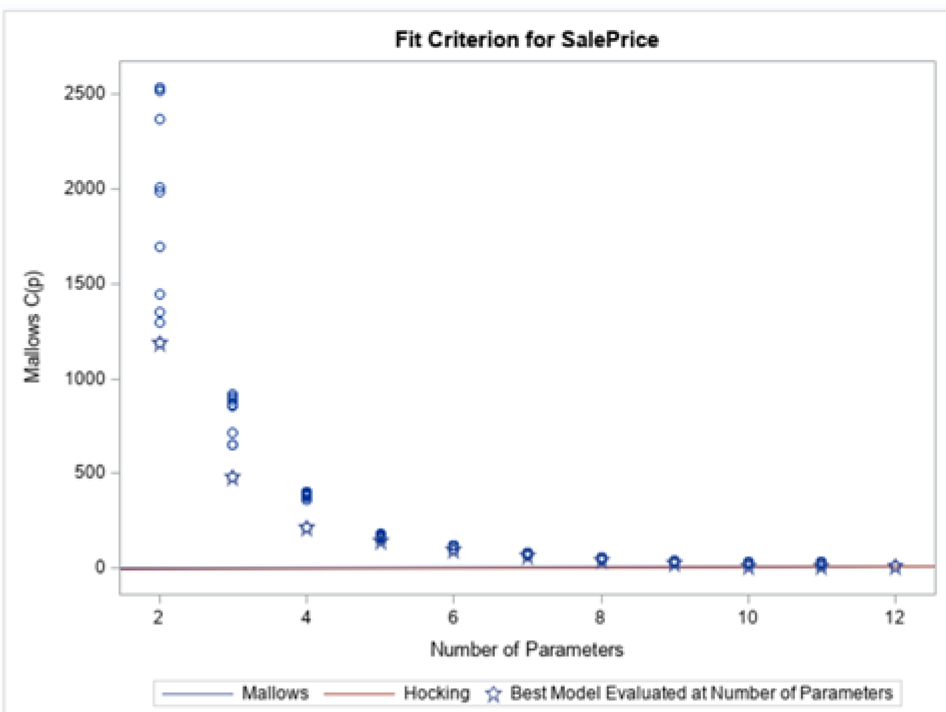


Fig 3: Plot with cp selection method

```
*Question 11 - create one-way frequency tables for the variables Bonus,
Fireplaces, and Lot_Shape_2 and create two-way frequency tables for the variables
Bonus by Fireplaces, and Bonus by Lot_Shape_2. For the continuous variable,
Basement_Area, create histograms for each level of Bonus?;

proc freq data = orion.team8;
tables Bonus;
run;
proc freq data = orion.team8;
tables Fireplaces;
run;
proc freq data = orion.team8;
tables Lot_Shape_2 ;
run;
proc freq data = orion.team8;
tables Bonus *Fireplaces;
run;
proc freq data = orion.team8;
tables Bonus *Lot_Shape_2;
run;

proc univariate data=orion.team8;
class Bonus;
  histogram;
  var Basement_Area;
run;
```

***Answer a There are missing values in the dataset. There are present in the table since misssing option wasn't added in the tables statement.**
***Asnwer b Basement area is normally distributed when Bonus = 0 and it is left-skewed when there is a bonus for the sale. It is also evident that incresase in basement area increases the chance of bonus.**

```
*Question 12 - Use PROC FREQ to test whether an ordinal association exists between Bonus and
Fireplaces.;
proc freq data=orion.team8;
tables Fireplaces*Bonus / chisq cl;
Run;
```

***Answer a;**

There is Bonus and Fireplaces have a significant ordinal association as p value is less than 0.001

***Answer b;**

For the Spearman correlation statistic, the relationship is significant at the 0.05 significance
Value - 0.2898
95% Confidence Limit - 0.2203,0.3593

```
*Question 13 - Fit a binary logistic regression model
in PROC LOGISTIC. Select Bonus as the outcome
variable and VARIABLE assigned to your team as the
predictor variable. ;
```

***Answer a;**

```
proc logistic data =orion.Team8;
	model Bonus(event='1') = Total_Bathroom  / CTABLE PPOB = (0 to 1 by .1)
/*classification table */
	LACKFIT clodds=pl/*Goodness-of-fit test – Hosmer & Lemeshow*/
	RISKLIMITS /*odds ratios for each varb with 95% CI*/
    OUTROC=ROC ALPHA=.10 ;
run;
```

***Answer b - BETA=0 mean that the intercept value is 0 and we'll look at the importance of independent variables when intercept is 0;**
***Answer c - Since all the 3 statistics in Global Null Hypothesis has a p-value <0.0001, wecan say that we reject the null hypothesis.**
***Answer d - The logistic regression equation is P = 1/(e^(-6.0946+2.2333(Total_Bathroom))) or it can be written as LOG(Bonus)=-6.0946+2.2333(Total_Bathroom;**
***Answer e - p-value of Total_Bathroom is <.0001 & hence is significant at the 0.10 significance level ;**
***Answer f - odd ratio of Total_Bathroom calculation is**
ODDS(Bonus)when Total_Bathroom is present= exp(-6.0946+2.2333* 1) = 0.021
ODDS(Bonus)when Total_Bathroom is absent= exp(-6.0946+2.2333* 0) = 0.0023
Odds Ratio = 0.021/0.0023 = 9.13;

```
*Question 14 - Run an analysis of variance with
SalePrice as the response variable and Heating_QC as
the
categorical predictor variable. Output diagnostic plots
and look at Levene's test of homogeneity
of variances.

#Levene's test of homogeneity of variances;
proc glm data=mylib.Team8
plots(only)=diagnostics;
 class Heating_QC;
 model SalePrice=Heating_QC;
```

```
 means Heating_QC / hovtest=levene;
 title "One-Way ANOVA with SalePrice as Heating_QC using levene";
run;
quit;
```

Solution – Please find below the output screenshots of my code

One-Way ANOVA with SalePrice as Heating_QC

The GLM Procedure

Dependent Variable: SalePrice Sale price in dollars

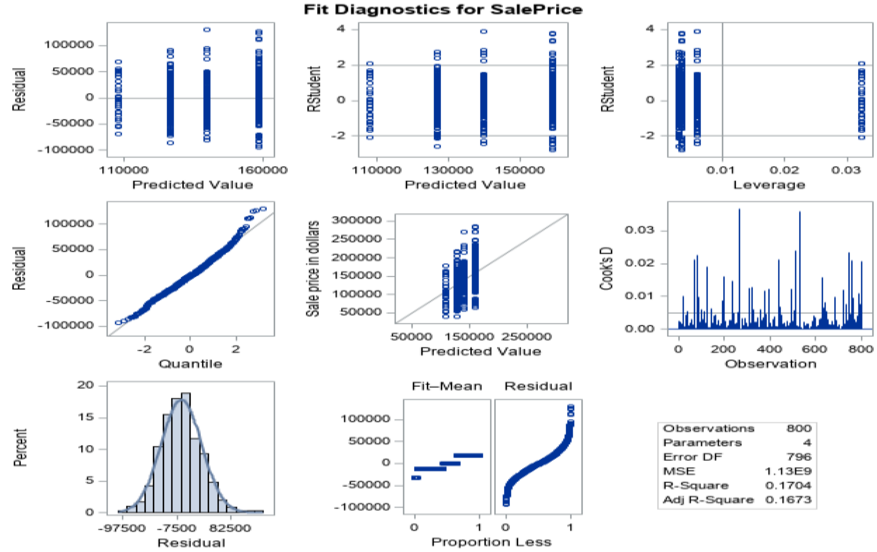
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	185076246528	61692082176	54.50	<.0001
Error	796	901019496537	1131934040.9		
Corrected Total	799	1.0860957E12			

R-Square	Coeff Var	Root MSE	SalePrice Mean
0.170405	24.06005	33644.23	139834.4

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Heating_QC	3	185076246528	61692082176	54.50	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Heating_QC	3	185076246528	61692082176	54.50	<.0001

Fit Diagnostics for SalePrice



One-Way ANOVA with SalePrice as Heating_QC

The GLM Procedure

Levene's Test for Homogeneity of SalePrice Variance ANOVA of Squared Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Heating_QC	3	6.153E19	2.051E19	5.99	0.0005
Error	796	2.725E21	3.424E18		

One-Way ANOVA with SalePrice as Heating_QC

The GLM Procedure

Level of Heating_QC	N	SalePrice	
		Mean	Std Dev
Ex	279	158621.487	38020.3081
Fa	31	107975.419	38968.0813
Gd	167	139629.982	33727.3522
TA	323	126769.892	28675.3232

Answer:According to anova analysis, SalePrice & Heating_QC seem to be significant to each other with p-value less than 0.0001 with both 95 & 99% confidence.

*Question 15 - Use the LSMEANS statement in PROC GLM to produce comparison information about
the mean sale prices of the different heating system quality ratings.
#Using lsmeans;

```
proc glm data=mylib.Team8
plots(only)=diagnostics;
 class Heating_QC;
 model SalePrice=Heating_QC;
 lsmeans Heating_QC / slice=Heating_QC;
 title "One-Way ANOVA with SalePrice as Heating_QC including lsmeans";
run;
quit;
```

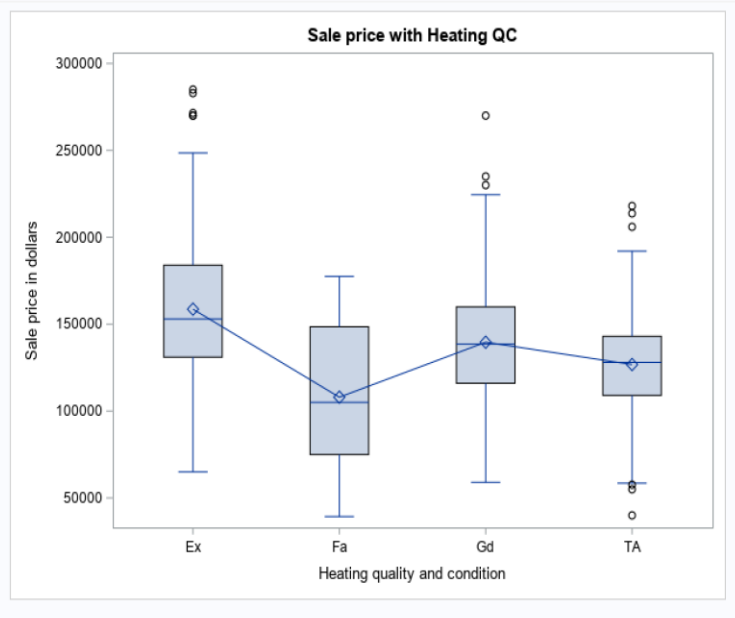
One-Way ANOVA with SalePrice as Heating_QC

The GLM Procedure
Least Squares Means

Heating_QC	SalePrice LSMEAN
Ex	158621.487
Fa	107975.419
Gd	139629.982
TA	126769.892

***Answer : Using lsmeans produces the above table in addition to Q1 output. Looking at the above table, we can say FA in Heating_QC has the lowest Salepeice mean, whereas TA & GD come next with mean sale price close to each other.**
Ex has the highest SalePrice mean among all the heating_QC;

```
*Question 16 - Perform a two-way ANOVA of SalePrice
with Heating_QC and Season_Sold as predictor
variables. Before conducting an analysis of variance,
you should explore the data. To further
explore the numerous treatments, examine the means
graphically. Include the interaction between
the two explanatory variables. Store the output to a
dataset and adjust p-values using PROC PLM.

*Performing Bivariate analysis between variables and exploring the data before jumping into two-way anova analysis;
********bivariate analysis between Heating_QC & SalePrice;
proc sgplot data=mylib.Team8;
 vbox SalePrice / category=Heating_QC connect=mean;
 title "Sale price with Heating QC";
run;
```



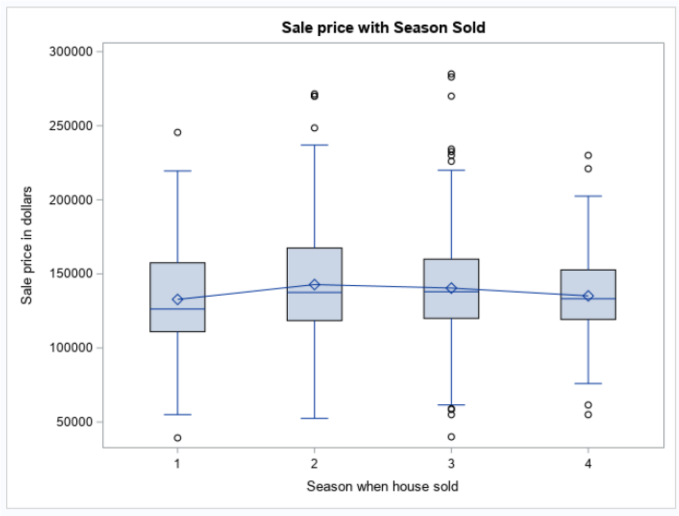

```
#Descriptive statistics
proc means data=mylib.Team8;
 var SalePrice;
 class Heating_QC ;
 title 'Descriptive Statistics of SalePrice by Heating_QC';
run;
```

Descriptive Statistics of SalePrice by Heating_QC

The MEANS Procedure

Analysis Variable : SalePrice Sale price in dollars						
Heating quality and condition	N Obs	N	Mean	Std Dev	Minimum	Maximum
Ex	279	279	158621.49	38020.31	65000.00	285000.00
Fa	31	31	107975.42	38968.08	39300.00	177500.00
Gd	167	167	139629.98	33727.35	59000.00	270000.00
TA	323	323	126769.89	28675.32	40000.00	218000.00

```
#bivariate analysis between Season_Sold & SalePrice; #Boxplot
*******bivariate analysis between Season_Sold &
SalePrice;
proc sgplot data=mylib.Team8;
 vbox SalePrice / category=Season_Sold connect=mean;
 title "Sale price with Season Sold";
run;
```



#Descriptive Statistics between Season_Sold & SalePrice

```
proc means data=mylib.Team8;
 var SalePrice;
 class Season_Sold ;
 title 'Descriptive Statistics of SalePrice by Season_Sold';
run;
```

Descriptive Statistics of SalePrice by Season_Sold

The MEANS Procedure

Analysis Variable : SalePrice Sale price in dollars						
Season when house sold	N Obs	N	Mean	Std Dev	Minimum	Maximum
1	80	80	132753.13	36982.39	39300.00	245500.00
2	289	289	142795.08	38496.58	52500.00	271500.00
3	327	327	140430.22	36821.41	40000.00	285000.00
4	104	104	135180.81	31209.66	55000.00	230000.00

```
********bivariate analysis between Heating_QC & Season_Sold;
proc freq data = mylib.Team8;
tables Heating_QC *Season_Sold;
run;
```

Descriptive Statistics of SalePrice by Season_Sold

The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

Table of Heating_QC by Season_Sold					
Heating_QC(Heating quality and condition)	Season_Sold(Season when house sold)				
	1	2	3	4	Total
Ex	24 3.00 8.60 30.00	107 13.38 38.35 37.02	117 14.63 41.94 35.78	31 3.88 11.11 29.81	279 34.88
Fa	4 0.50 12.90 5.00	10 1.25 32.26 3.46	14 1.75 45.16 4.28	3 0.38 9.68 2.88	31 3.88
Gd	16 2.00 9.58 20.00	59 7.38 35.33 20.42	66 8.25 39.52 20.18	26 3.25 15.57 25.00	167 20.88
TA	36 4.50 11.15 45.00	113 14.13 34.98 39.10	130 16.25 40.25 39.76	44 5.50 13.62 42.31	323 40.38
Total	80 10.00	289 36.13	327 40.88	104 13.00	800 100.00

```
********Performing GLM where I have included an interaction
between Heating_QC & Season_Sold and also storing the output to
a new dataset
proc glm data=mylib.Team8
 plots(only)=(intplot);
 class Heating_QC Season_Sold;
 model SalePrice=Heating_QC Season_Sold Heating_QC*Season_Sold;
 lsmeans Heating_QC*Season_Sold / slice=Heating_QC;
 store out=new;
 title "Model with Heating QC and Season Sold along with their interaction";
run;
```

Model with Heating QC and Season Sold along with their interaction

The GLM Procedure

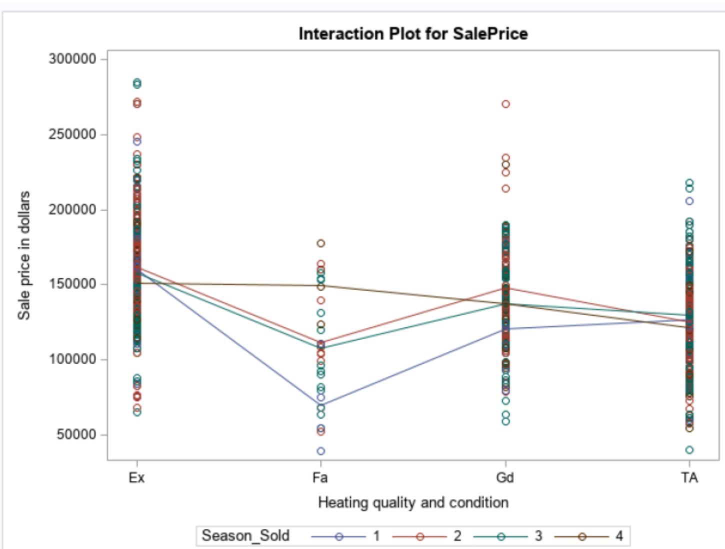
Dependent Variable: SalePrice Sale price in dollars

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	15	211797395550	14119826370	12.66	<.0001
Error	784	874298347515	1115176463.7		
Corrected Total	799	1.0860957E12			

R-Square	Coeff Var	Root MSE	SalePrice Mean
0.195008	23.88129	33394.26	139834.4

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Heating_QC	3	185076246528	61692082176	55.32	<.0001
Season_Sold	3	5197053124.7	1732351041.6	1.55	0.1993
Heating_Q*Season_Sol	9	21524095898	2391566210.9	2.14	0.0239

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Heating_QC	3	119916208462	39972069487	35.84	<.0001
Season_Sold	3	10855638542	3618546180.8	3.24	0.0215
Heating_Q*Season_Sol	9	21524095898	2391566210.9	2.14	0.0239



The GLM Procedure
Least Squares Means

Heating_QC	Season_Sold	SalePrice LSMEAN
Ex	1	160520.833
Ex	2	161361.766
Ex	3	157744.291
Ex	4	151003.355
Fa	1	70075.000
Fa	2	111213.800
Fa	3	107521.429
Fa	4	149833.333
Gd	1	120868.750
Gd	2	148069.610
Gd	3	137571.970
Gd	4	137248.077
TA	1	126487.500
TA	2	125255.088
TA	3	129842.692
TA	4	121812.500

Model with Heating QC and Season Sold along with their interaction

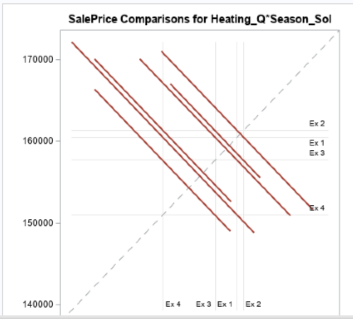
The GLM Procedure
Least Squares Means

Heating_Q*Season_Sol Effect Sliced by Heating_QC for SalePrice					
Heating_QC	DF	Sum of Squares	Mean Square	F Value	Pr > F
Ex	3	2779199876	926399959	0.83	0.4771
Fa	3	11109778636	3703259545	3.32	0.0194
Gd	3	10261200946	3420400315	3.07	0.0273
TA	3	2570969564	856989855	0.77	0.5118

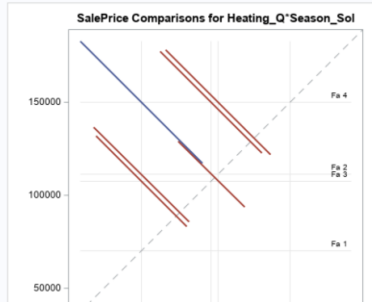
Answer : Heating_QC seems to be significant with Sale Price (p-value <0.0001) whereas Season Sold alone is not significant with Sale Price but when Season Sold is made to interact with Heating_QC, it acts significant with p-value of 0.02

```
********Adjusted p-values using PROC PLM on new dataset created;

#Below are the plots of how interaction between Heating_QC &
Season_Sold are significant with Sale_Price at individual level

proc plm restore=new plots=all;
	slice Heating_QC*Season_Sold / sliceby=Heating_QC adjust=tukey;
	effectplot interaction(sliceby = Heating_QC) / clm;
title "Model with PLM";
run;
```

Slice	Season when house sold	Season when house sold	Estimate	Standard Error	DF	t Value	Pr > |t|	Adj P
Heating_QC Ex	1	2	-840.93	7542.41	784	-0.11	0.9113	0.9995
Heating_QC Ex	1	3	2776.54	7483.12	784	0.37	0.7107	0.9826
Heating_QC Ex	1	4	9517.48	9079.60	784	1.05	0.2949	0.7210
Heating_QC Ex	2	3	3617.48	4466.95	784	0.81	0.4183	0.8498
Heating_QC Ex	2	4	10358	6811.43	784	1.52	0.1287	0.4255
Heating_QC Ex	3	4	6740.94	6745.73	784	1.00	0.3180	0.7498



Adjustment for Multiple Comparisons: Tukey-Kramer								
Slice	Season when house sold	Season when house sold	Estimate	Standard Error	DF	t Value	Pr > |t|	Adj P
Heating_QC Fa	1	2	-41139	19756	784	-2.08	0.0376	0.1598
Heating_QC Fa	1	3	-37446	18933	784	-1.98	0.0483	0.1972
Heating_QC Fa	1	4	-79758	25505	784	-3.13	0.0018	0.0099
Heating_QC Fa	2	3	3692.37	13827	784	0.27	0.7895	0.9933
Heating_QC Fa	2	4	-38620	21983	784	-1.76	0.0793	0.2952
Heating_QC Fa	3	4	-42312	21246	784	-1.99	0.0468	0.1920

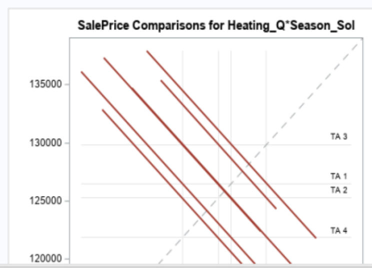


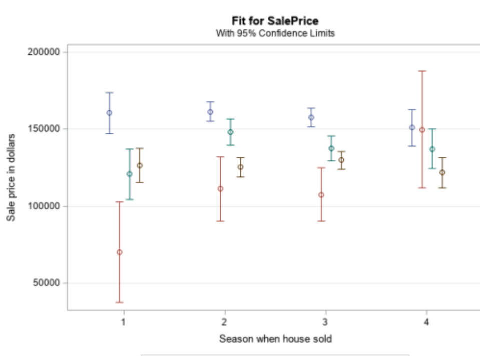
Slice	Num DF	Den DF	F Value	Pr > F
Heating_QC Gd	3	784	3.07	0.0273

Simple Differences of Heating_Q*Season_Sol Least Squares Means Adjustment for Multiple Comparisons: Tukey-Kramer								
Slice	Season when house sold	Season when house sold	Estimate	Standard Error	DF	t Value	Pr > |t|	Adj P
Heating_QC Gd	1	2	-27201	9412.75	784	-2.89	0.0040	0.0206
Heating_QC Gd	1	3	-16703	9305.65	784	-1.79	0.0730	0.2765
Heating_QC Gd	1	4	-16379	10611	784	-1.54	0.1231	0.4119
Heating_QC Gd	2	3	10498	5983.14	784	1.75	0.0797	0.2963
Heating_QC Gd	2	4	10822	7860.83	784	1.38	0.1690	0.5144
Heating_QC Gd	3	4	323.89	7732.27	784	0.04	0.9666	1.0000

Slice	Num DF	Den DF	F Value	Pr > F
Heating_QC TA	3	784	0.77	0.5118

Simple Differences of Heating_Q*Season_Sol Least Squares Means Adjustment for Multiple Comparisons: Tukey-Kramer								
Slice	Season when house sold	Season when house sold	Estimate	Standard Error	DF	t Value	Pr > |t|	Adj P
Heating_QC TA	1	2	1232.41	6391.08	784	0.19	0.8471	0.9975
Heating_QC TA	1	3	-3355.19	6289.31	784	-0.53	0.5939	0.9509
Heating_QC TA	1	4	4675.00	7504.80	784	0.62	0.5335	0.9247
Heating_QC TA	2	3	-4587.60	4295.01	784	-1.07	0.2858	0.7091
Heating_QC TA	2	4	3442.59	5934.12	784	0.58	0.5620	0.9380
Heating_QC TA	3	4	8030.19	5824.36	784	1.38	0.1684	0.5131





Answer:Used Tukey to get better insights into the interaction between Season Sold & Heating_QC. Anova gives us the overall result but Tukey test exactly tells us where the difference lie that is basically which Season_Sol with Heating_QC is actually significant with SalePrice