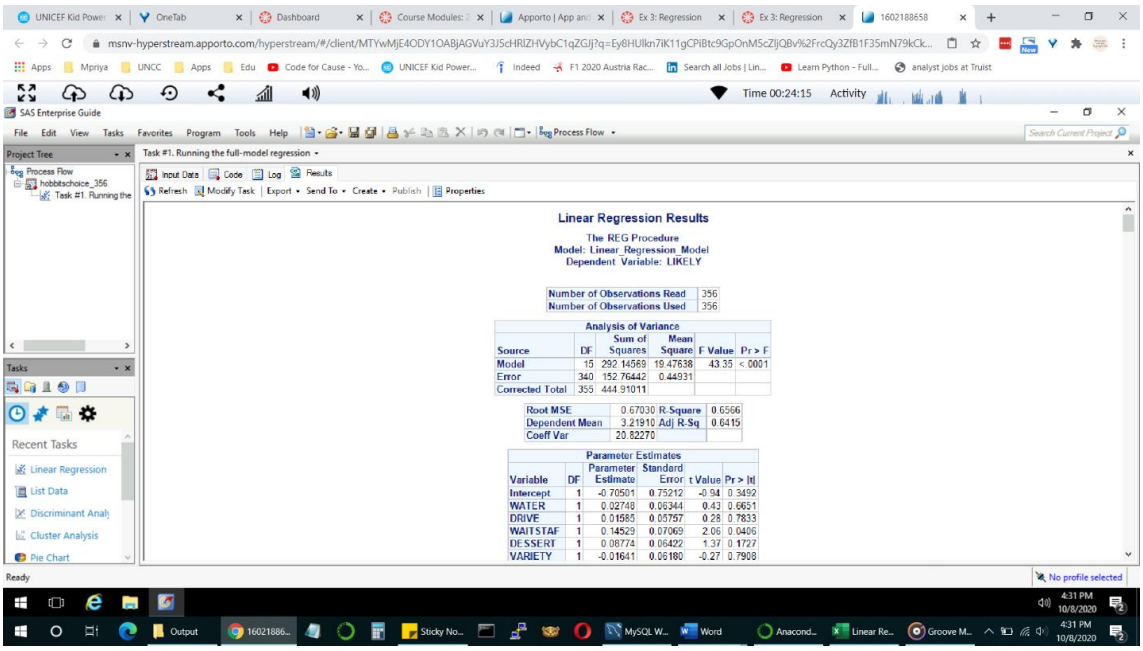
Muthu Priya Shanmugakani Velsamy
DSBA-6276 Consumer Analytics
Linear Regression

Task #1. Running the full-model regression (10 points)

Output of the linear regression model on hobbitchoice_356 data.



Task #2. Interpreting your regression results (30 points: breakout shown below)

Step 1: Identifying positive and negative significant variables (5 points)

Dependent Variable: Likely

Cut off value: 0.10; Based on this cut-off value below mentioned IV are categorized.

Check on p-values: WAITSTAF, JAZZ, AGE, EDUCA, INCOME are the only variables which is qualified from the p-value test with the given alpha (cut off) value. Other independent variables which doesn't satisfy the cut off value is neglected.

Check on parameter estimate values: All the variables WAITSTAF, JAZZ, AGE, EDUCA and INCOME have positive parameter estimate values 0.14529, 0.08071, 0.02020, 0.20524, 0.17902 respectively.

Below mentioned variables have positive significance with DV. There is no variable which passes the significance test with negative significance with DV.

Positive Significance variables with DV	Negative Significance variables with DV
WAITSTAF	NIL
JAZZ	
AGE	
EDUCA	
INCOME	

Root MSE	0.67030	R-Square	0.6566
Dependent Mean	3.21910	Adj R-Sq	0.6415
Coeff Var	20.82270		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.70501	0.75212	-0.94	0.3492
WATER	1	0.02748	0.06344	0.43	0.6651
DRIVE	1	0.01585	0.05757	0.28	0.7833
WAITSTAF	1	0.14529	0.07069	2.06	0.0406
DESSERT	1	0.08774	0.06422	1.37	0.1727
VARIETY	1	-0.01641	0.06180	-0.27	0.7908
UNUSAL	1	0.02995	0.05958	0.50	0.6154
SIMPLE	1	-0.10087	0.06736	-1.50	0.1352
ELEGANT	1	0.04265	0.06608	0.65	0.5191
STRING	1	-0.02663	0.05861	-0.45	0.6499
JAZZ	1	0.08071	0.04591	1.76	0.0796
AGE	1	0.02020	0.00730	2.77	0.0059
EDUCA	1	0.20524	0.05005	4.10	<.0001
FAMSIZE	1	0.03644	0.02644	1.38	0.1691
INCOME	1	0.17902	0.05775	3.10	0.0021
GENDER	1	0.01693	0.07389	0.23	0.8189

Step 2: Explaining the practical meaning of each positive and negative significant variable (25 points)

Dependent Variable: Likely

Cut off value: 0.10; Based on this cut-off value below mentioned IV are categorized.

Check on p-values: WAITSTAF, JAZZ, AGE, EDUCA, INCOME are the only variables which is qualified from the p-value test with the given alpha (cut off) value. Other independent variables which doesn't satisfy the cut off value is neglected.

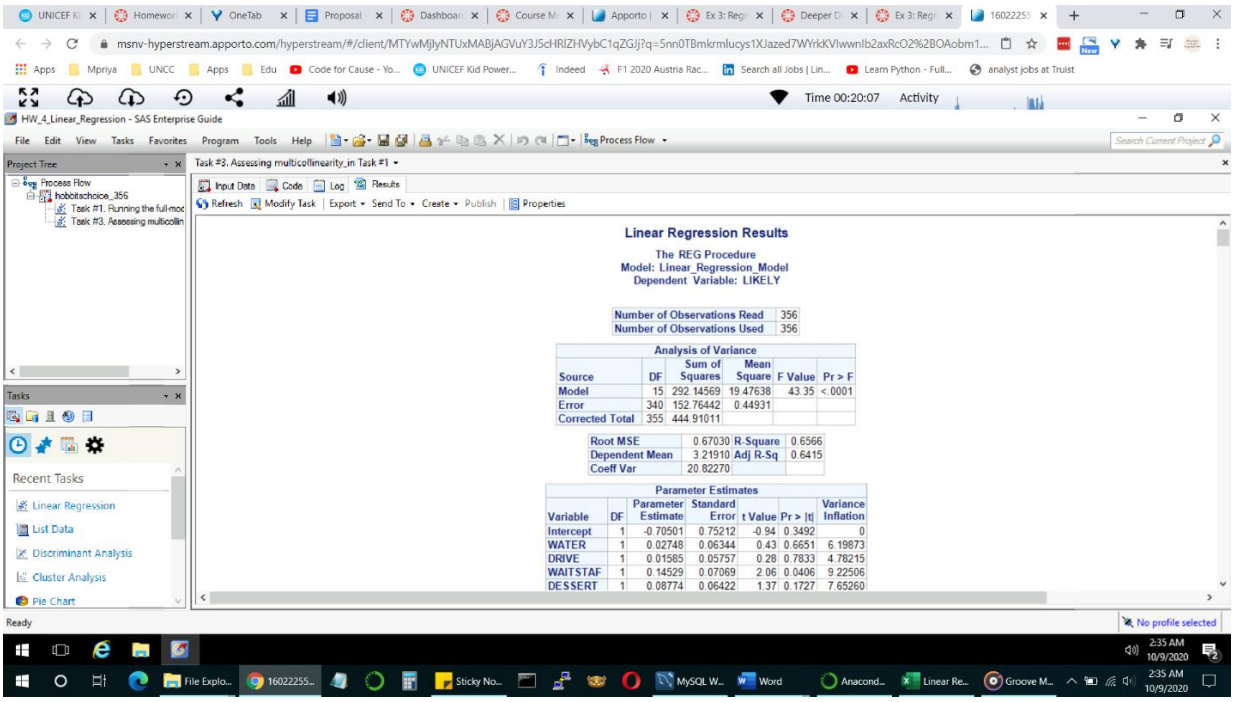
Check on parameter estimate values: All the variables WAITSTAF, JAZZ, AGE, EDUCA and INCOME have positive parameter estimate values 0.14529, 0.08071, 0.02020, 0.20524, 0.17902 respectively.

Independent Significant Variables	**Practical Reasons**
WAITSTAF	Since we are upscale restaurant the customers prefer the wait staff wearing tuxedos more and this feature have positive significance that is higher the wait staff wearing tuxedos greater is the possibility to dine in our restaurant. Wait staff wearing tuxedos provides a formal and decent look which ultimately results in royal sight of the restaurant.
JAZZ	The customers like to have the Jazz music background when they are in the restaurant for dining. This feature adds advantage to the customer health as well by providing them cool music, a relief to their mind & body which helped in reducing the fatigue. As a result, the customer will be willing to check out the restaurant so often.
AGE	Age is a factor which senses which category of people visit our restaurant and this has some important significant in the process of likelihood to dine. Because, from this value we can predict which age group we should target and send our offers or advertisement or feedback and any other communications. This variable will play an important role in attracting the people to our restaurant.
EDUCA	The Education of the people visiting our restaurant have significant effect in the likelihood of dinning because, I believe the people manners, code of respect will be different according to their standard of education.
INCOME	Since we are upscale restaurant, the foods supplied in our restaurant will be low to moderate to high. The person visiting our restaurant should be able to pay the bills without any trouble, for this to happen, customers level of Income is important. This has the second highest positive significance with the dependent variable.

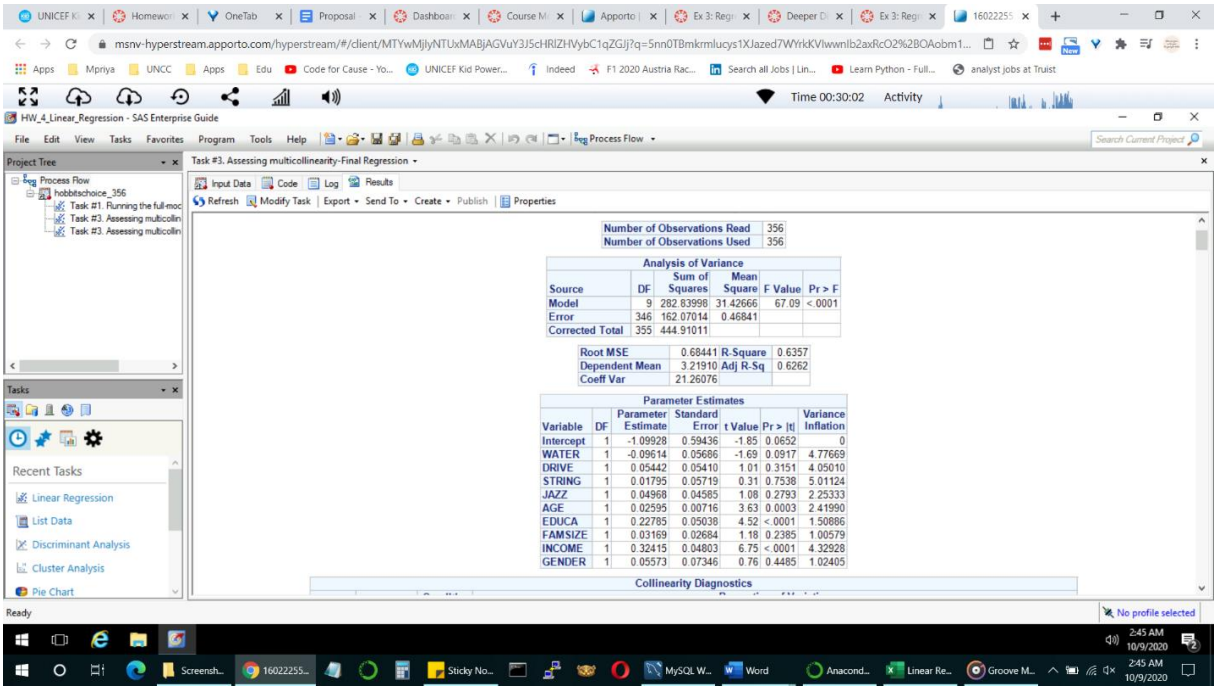
Muthu Priya Shanmugakani Velsamy
DSBA-6276 Consumer Analytics
Linear Regression

Task #3. Assessing multicollinearity (10 points)

Before Removing the IVs based on the Cut off value 6.66. Below attached screenshot will be the result.



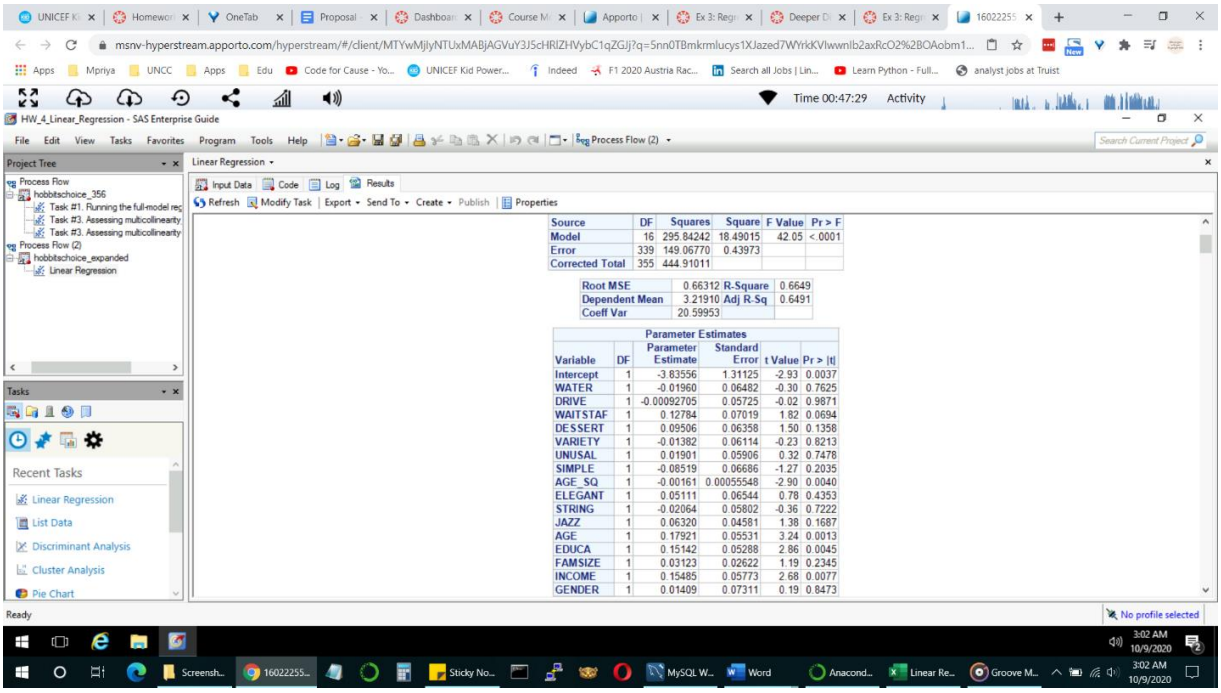
After removing the IVs based on the Cut off value 6.66 and re-running the model once again produces the output below in the screenshot.



We have removed the variables which didn't satisfy the cutoff value of 6.66. This is important because the multi-collinearity will be a severe problem when we try to modify one variable and see the effect in the dependent variable because that variable which we try to modify, will have dependency with other independent variables. So if that variable is modified then all the correlated independent variable will reflect that changes and it will result in a bad model.

Task #4. Curvilinear Effect (20 points)

Cut off value: 0.10



Interpretation of the curvilinear effect of the two age variables – AGE & AGE_SQ

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
AGE	1	0.17921	0.05531	3.24	0.0013
AGE_SQ	1	-0.00161	0.000555	-2.9	0.004

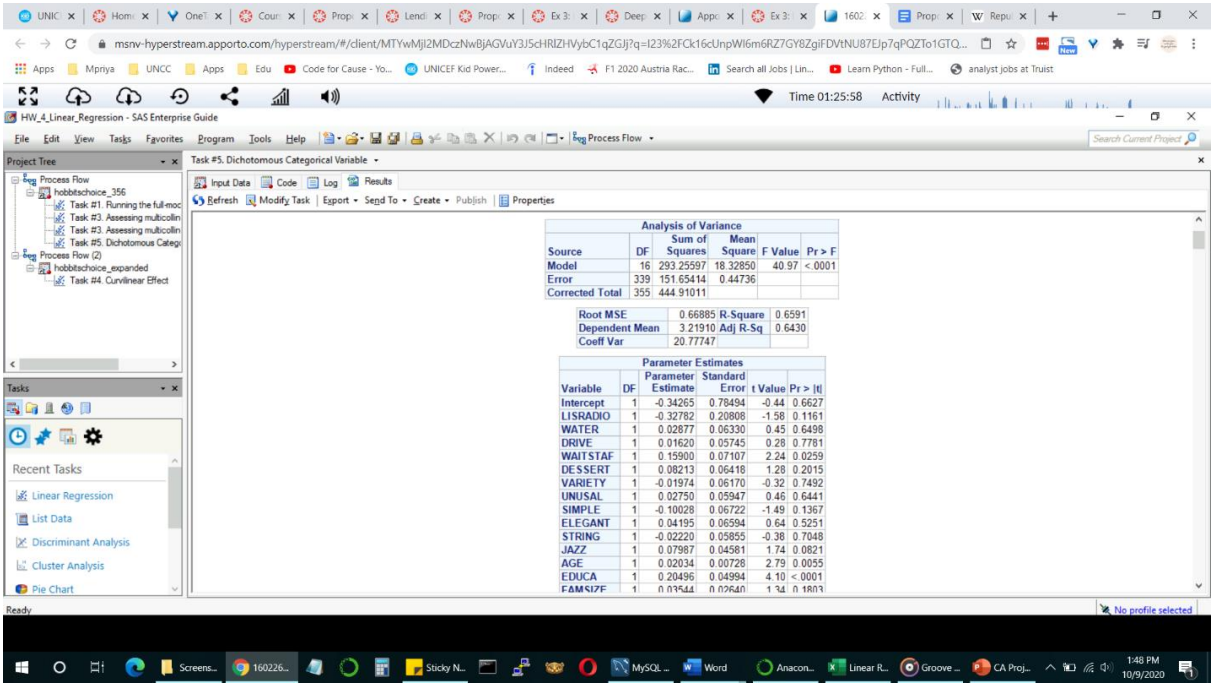
From the above results we can clearly see that both the AGE and AGE_SQ have significant value with the dependent variable LIKELY. These 2 variables also satisfy the cut off value of 0.10.

The variable AGE alone shows that there is some straight-line linear effect with the Dependent variable. i.e., the likelihood to dine in increases as the age increases. But when we consider AGE_SQ we can see the curve showing the linear effect with the DV. This is called curvilinear effect. From this curvilinear graph we can state that the number

of customers who dine-in increases as age increases but also, it drops after certain age group. Example: Middle age people have the possibility to dine in more when compared to infants or children and old age people.

We can also see that AGE_SQ have negative parameter estimate value which states that it has negative significance with DV. As said in the example above, it will have inverse effect after certain point in the graph.

Task #5. Dichotomous Categorical Variable (10 points)



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	16	293.25597	18.32850	40.97	<.0001
Error	339	151.65414	0.44736		
Corrected Total	355	444.91011			

Root MSE	0.66885	R-Square	0.6591
Dependent Mean	3.21910	Adj R-Sq	0.6430
Coeff Var	20.77747		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.34265	0.78494	-0.44	0.6627
LISRADIO	1	-0.32782	0.20808	-1.58	0.1161
WATER	1	0.02877	0.06330	0.45	0.6498
DRIVE	1	0.01620	0.05745	0.28	0.7781
WAITSTAF	1	0.15900	0.07107	2.24	0.0259
DESSERT	1	0.08213	0.06418	1.28	0.2015
VARIETY	1	-0.01974	0.06170	-0.32	0.7492
UNUSAL	1	0.02750	0.05947	0.46	0.6441
SIMPLE	1	-0.10028	0.06722	-1.49	0.1367
ELEGANT	1	0.04195	0.06594	0.64	0.5251
STRING	1	-0.02220	0.05855	-0.38	0.7048
JAZZ	1	0.07987	0.04581	1.74	0.0821
AGE	1	0.02034	0.00728	2.79	0.0055
EDUCA	1	0.20496	0.04994	4.10	<.0001
FAMSIZE	1	0.03544	0.02640	1.34	0.1803

Would you describe yourself as one who listens to the radio?

1. Yes
2. No

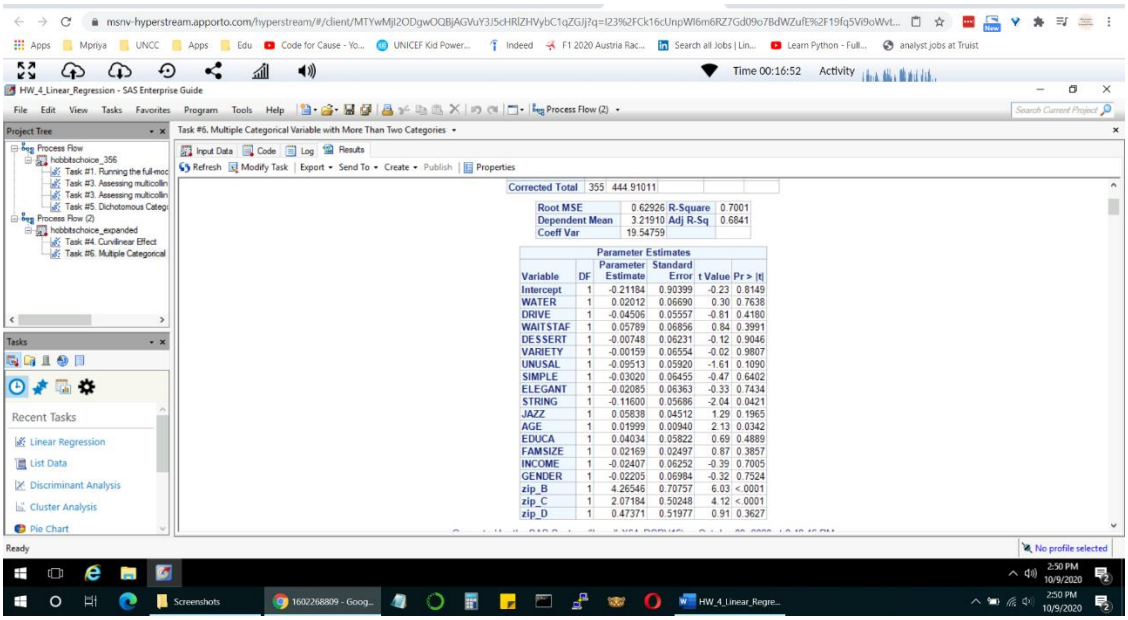
Yes and No are the two categorical variables here. These are the dichotomous categorical variables.

Yes, denotes that the customer likes to listen to radio and No, means the customer don't like to listen to radio.

According to the output above, LISRADIO variable -0.32782 as the Parameter estimate value which means it is a negative significant variable with DV.

When we speak practically, if the customer likes to listen to music then the chance of their likelihood to dine in our restaurant is much more when compared to the alternative No. When more customers like to listen to radio, then we can target those customers easily if we are able to categorize them from other customers. This will definitely have impact on increasing the customers chance of visiting the restaurant often with pleasant feeling.

Task #6. Multiple Categorical Variable with More Than Two Categories (20 points)



Corrected Total 355 444.91011

Root MSE	0.62926	R-Square	0.7001
Dependent Mean	3.21910	Adj R-Sq	0.6841
Coeff Var	19.54759		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.21184	0.90399	-0.23	0.8149
WATER	1	0.02012	0.06690	0.30	0.7638
DRIVE	1	-0.04506	0.05557	-0.81	0.4180
WAITSTAF	1	0.05789	0.06856	0.84	0.3991
DESSERT	1	-0.00748	0.06231	-0.12	0.9046
VARIETY	1	-0.00159	0.06554	-0.02	0.9807
UNUSAL	1	-0.09513	0.05920	-1.61	0.1090
SIMPLE	1	-0.03020	0.06455	-0.47	0.6402
ELEGANT	1	-0.02085	0.06363	-0.33	0.7434
STRING	1	-0.11600	0.05686	-2.04	0.0421
JAZZ	1	0.05838	0.04512	1.29	0.1965
AGE	1	0.01999	0.00940	2.13	0.0342
EDUCA	1	0.04034	0.05822	0.69	0.4889
FAMSIZE	1	0.02169	0.02497	0.87	0.3857
INCOME	1	-0.02407	0.06252	-0.39	0.7005
GENDER	1	-0.02205	0.06984	-0.32	0.7524
zip_B	1	4.26546	0.70757	6.03	< .0001
zip_C	1	2.07184	0.50248	4.12	< .0001
zip_D	1	0.47371	0.51977	0.91	0.3627

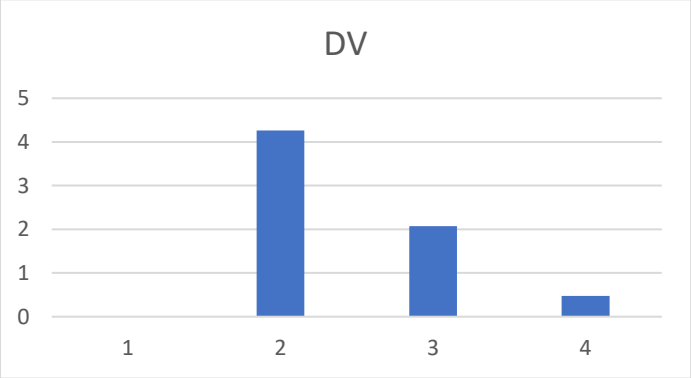
There are 4 dummy variables, zip_A, zip_B, zip_C, zip_D. We are keeping zip_A in reference level category and taking the remaining three dummy variables into account. zip_B, zip_C, zip_D have significant effect on the dependent variable likely.

The 4 categories in the original zip code variables in terms of their influence on restaurant patronization (DV) is:

zip A (Reference level variable)	1	0
zip B	2	4.26546
zip C	3	2.07184
zip D	4	0.47371

Actually, the result conveys that the zip B customers are more likely to dine in the restaurant than zip C, D and A. This may be because the customers in that zip B area may find our restaurant near to their locality or zip B people belong to high class community or they may find our food tastier than other locality people or they may like to drive to our restaurant specifically. Zip C customers comes second highly interested locality customers. Zip D customers become third highest customers out of four localities. This difference between zip B and zip D is may be because of the distance of the restaurant from their location or any other taste preferences among the community peoples or even some other upscale restaurant is nearby zip D and the distance is less when compared to ours. In that case we have to provide more competing offers and communication to the zip D customers to grab their attention towards us. On speaking about zip B and C customers, we need to provide them better ambience in the restaurant and weekly once email communication about the specials of the week or the offers of the

week, to make them visit our restaurant again. Zip A customers didn't play any role here because we have kept that as our reference level variable among the 4 zip variables. Below is the graph we obtained as a result of the variables arranged in 1, 2, 3, 4. This is a replica for the table values mentioned above.

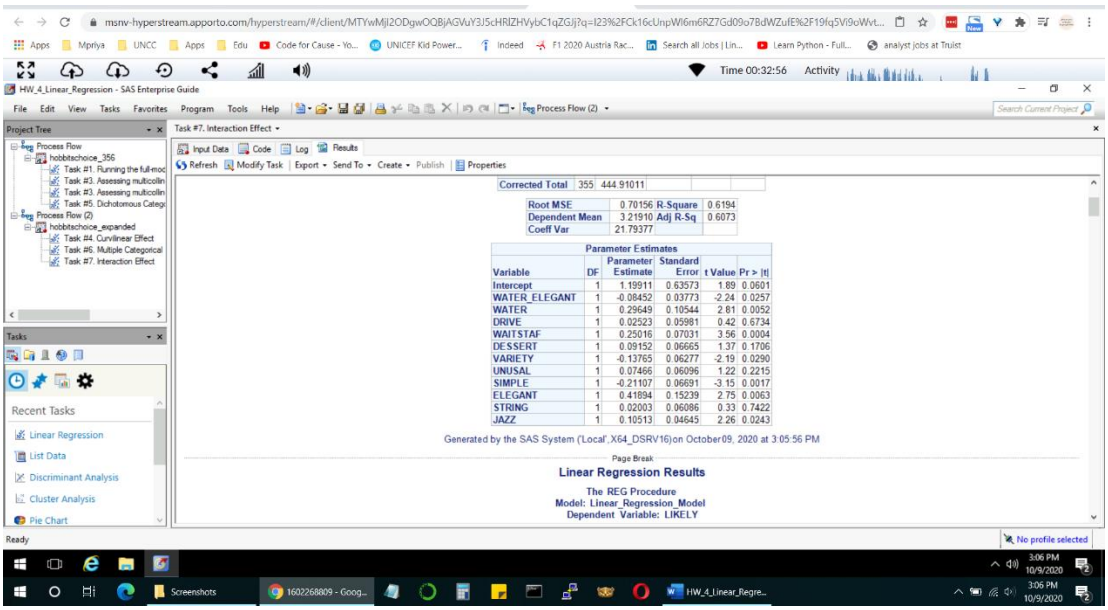


The result based on the positive and negative category values are:

In the order of increasing.

zip B > zip C > zip D > zip A

Task #7. Interaction Effect (15 points)



Cut off value: 0.10

The Interaction Effect of the two independent variables WATER and ELEGANT are taken and their values are multiplied together to get the Interaction variable WATER_ELEGANT.

From the above result we can see that all three variables WATER, ELEGANT, WATER_ELEGANT satisfy the cut off value criteria (0.10) which means they have significant relationship with the dependent variable LIKELY.

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
WATER	1	0.29649	0.10544	2.81	0.0052
ELEGANT	1	0.41894	0.15239	2.75	0.0063
WATER_ELEGANT	1	-0.08452	0.03773	-2.24	0.0257

- The main effect of the variable WATER is 0.29649 with the dependent variable likely.

If one customer prefers water front décor, and if our restaurant have water front décor, then the possibility that the customer prefer our restaurant is more since the customer wants to have a water front view. This also provides the customer a cool feeling and they will have a tension-free dinning at our restaurant.

- The main effect of the variable ELEGANT is 0.41894 with the dependent variable likely.

If one customer prefers elegant décor, and if our restaurant have elegant décor, then the possibility that the customer dinning in our restaurant will be increased since it will definitely attract the customer. Since we are upscale restaurant, the stylish and neat look of our restaurant will definitely give the customer a better feel towards visiting and revisiting the restaurant often.

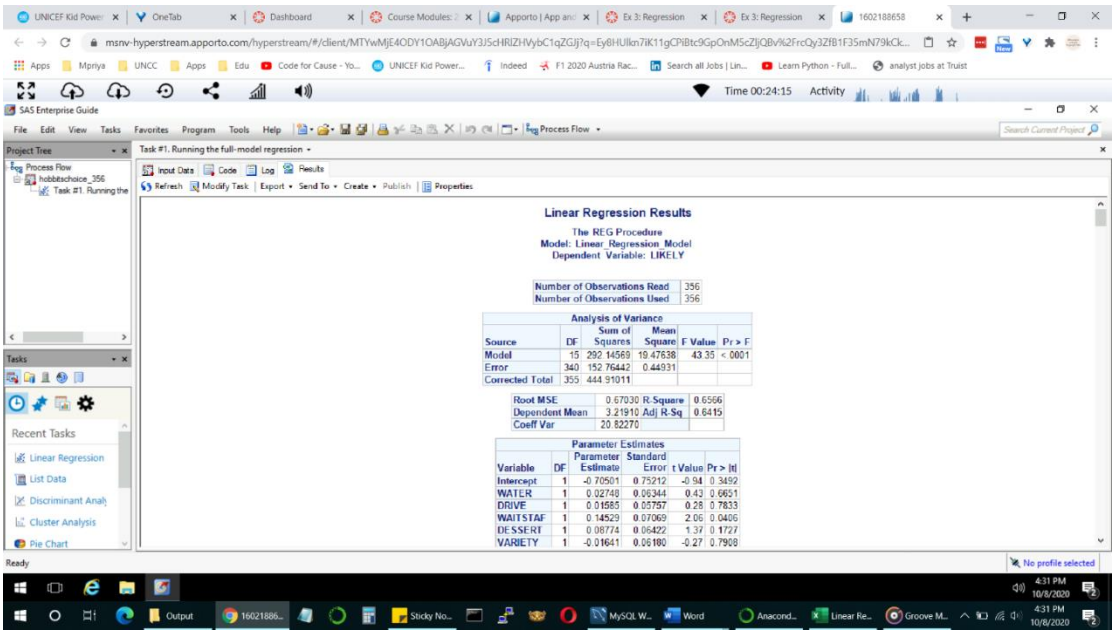
- The main effect of the variable WATER_ELEGANT (Interaction variable) is -0.08452 with the dependent variable likely. This shows that the Interaction of Water & Elegant décor have really negative significance with the dependent variable Likely.

If both the water and elegant décor is provided in our restaurant, we can't conclude that we attracted both type of customers (who like Water & Elegant décor individually) because, if they like Water décor, then there is assurance that the same customer likes Elegant décor as well or vice versa. As per the result obtained, we can clearly see that when both the decors are combined then the customer who is going to dine in our restaurant will be decreased since they don't like both the combination of the décor. Its as simple as adding beauty to a beauty is not going to add extra beauty but it will result in dominating each other which won't look that good as an end result. Two attractions at the same place won't work since they will pull each other's leg unknowingly.

Muthu Priya Shanmugakani Velsamy
DSBA-6276 Consumer Analytics
Linear Regression

Report Organization:

Task #1. Running the full-model regression

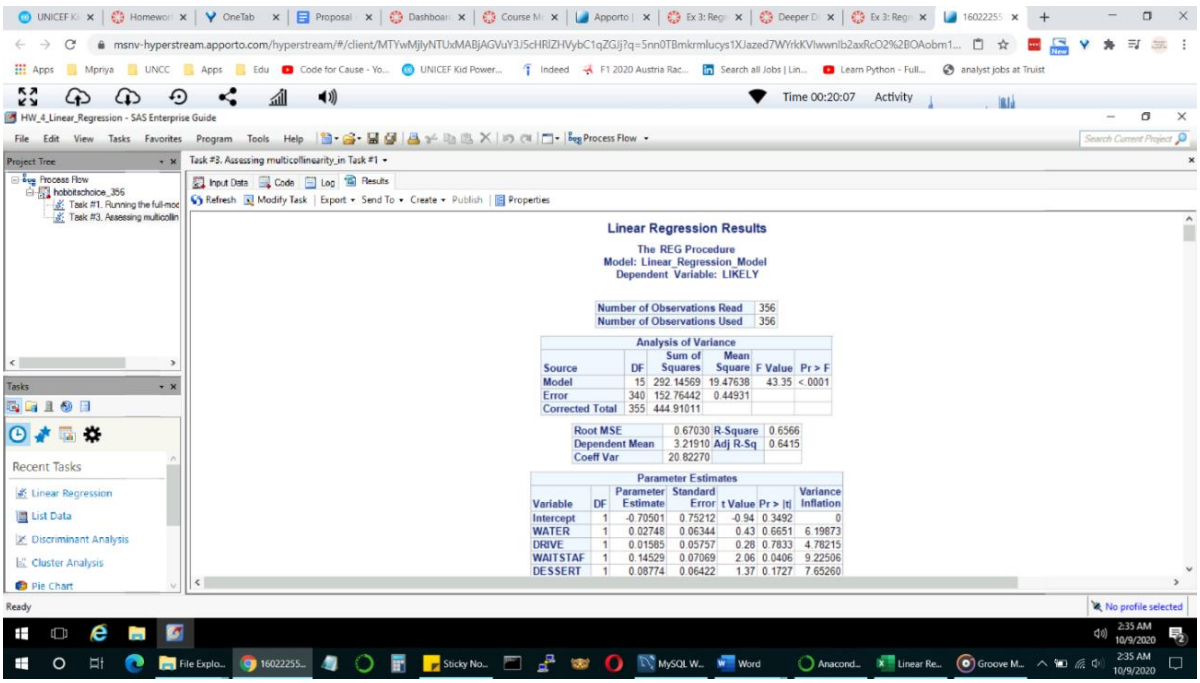


Task #2. Interpreting your regression results

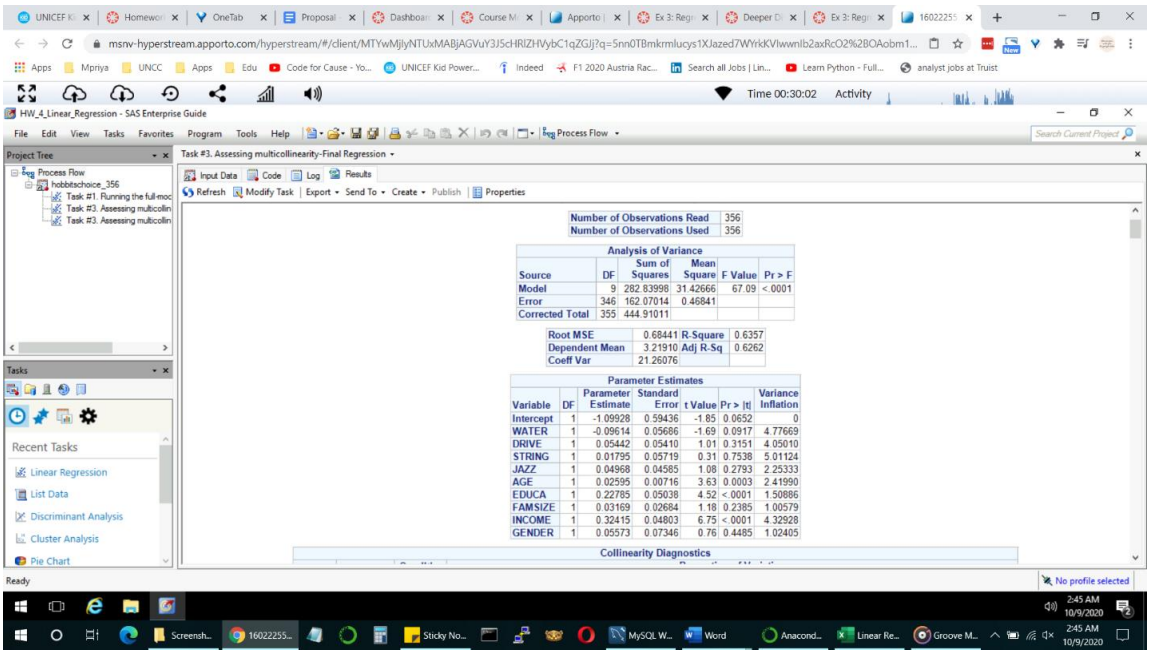
The interpretation is explained above.

Task #3. Assessing multicollinearity

The variables before satisfying the significance test



The variables after satisfying the significance test



Number of Observations Read 356
Number of Observations Used 356

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	282.83998	31.42666	67.09	<.0001
Error	346	162.07014	0.46841		
Corrected Total	355	444.91011			

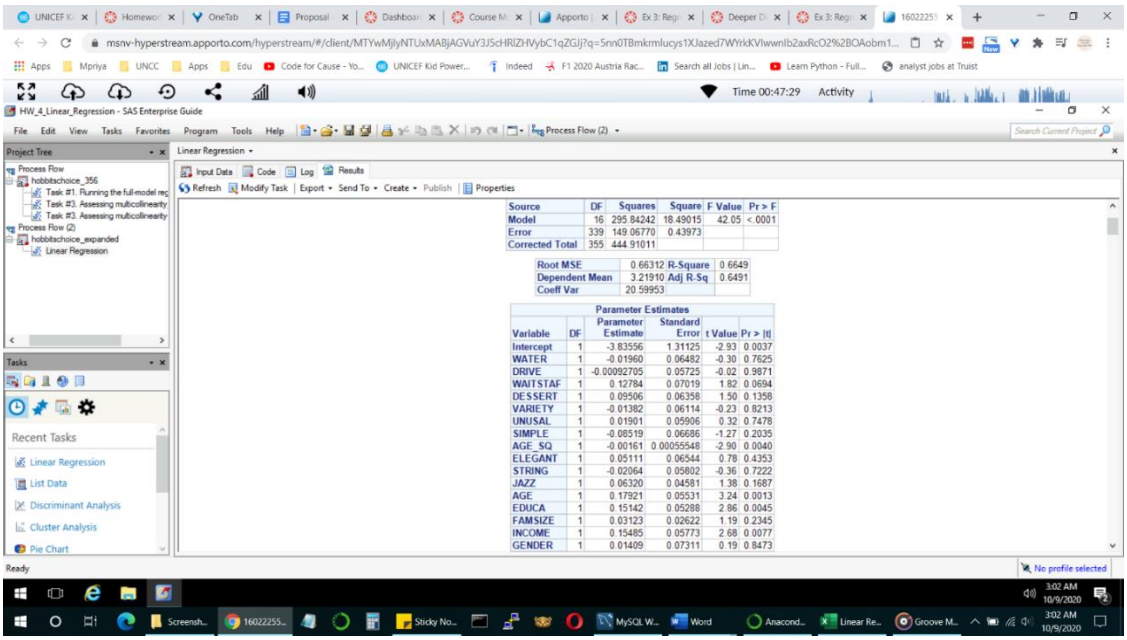
Root MSE	0.68441	R-Square	0.6357
Dependent Mean	3.21910	Adj R-Sq	0.6262
Coeff Var	21.26076		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	-1.09928	0.59436	-1.85	0.0652	0
WATER	1	-0.09614	0.05686	-1.69	0.0917	4.77669
DRIVE	1	0.05442	0.05410	1.01	0.3151	4.05010
STRING	1	0.01795	0.05719	0.31	0.7538	5.01124
JAZZ	1	0.04968	0.04585	1.08	0.2793	2.25333
AGE	1	0.02595	0.00716	3.63	0.0003	2.41990
EDUCA	1	0.22785	0.05038	4.52	<.0001	1.50886
FAMSIZE	1	0.03169	0.02684	1.18	0.2385	1.00579
INCOME	1	0.32415	0.04803	6.75	<.0001	4.32928
GENDER	1	0.05573	0.07346	0.76	0.4485	1.02405

Collinearity Diagnostics

Task #4. Curvilinear Effect



Source	DF	Squares	Square	F Value	Pr > F
Model	16	295.84242	18.49015	42.05	<.0001
Error	339	149.06770	0.43973		
Corrected Total	355	444.91011			

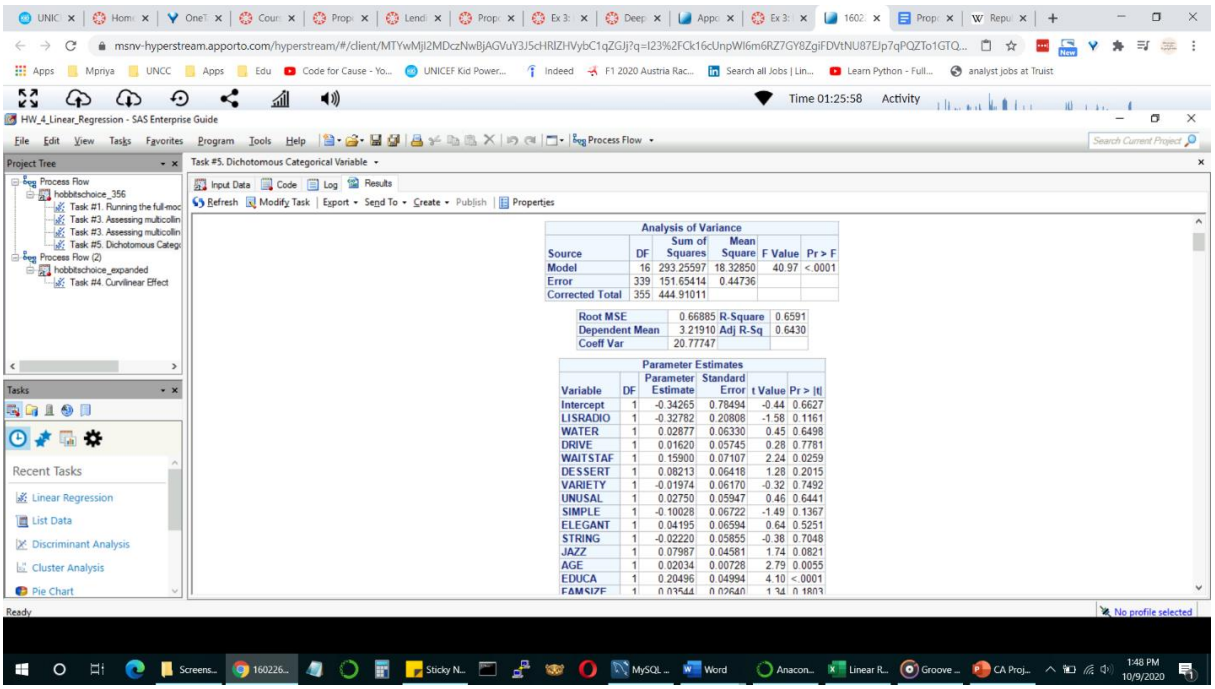
Root MSE	0.66312	R-Square	0.6649
Dependent Mean	3.21910	Adj R-Sq	0.6491
Coeff Var	20.59953		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-3.83556	1.31125	-2.93	0.0037
WATER	1	-0.01960	0.06482	-0.30	0.7625
DRIVE	1	-0.00092705	0.05725	-0.02	0.9871
WAITSTAF	1	0.12784	0.07019	1.82	0.0694
DESSERT	1	0.09506	0.06358	1.50	0.1358
VARIETY	1	-0.01382	0.06114	-0.23	0.8213
UNUSAL	1	0.01901	0.05906	0.32	0.7478
SIMPLE	1	-0.08519	0.06686	-1.27	0.2035
AGE_SQ	1	-0.00161	0.00055548	-2.90	0.0040
ELEGANT	1	0.05111	0.06544	0.78	0.4353
STRING	1	-0.02064	0.05802	-0.36	0.7222
JAZZ	1	0.06320	0.04581	1.38	0.1687
AGE	1	0.17921	0.05531	3.24	0.0013
EDUCA	1	0.15142	0.05288	2.86	0.0045
FAMSIZE	1	0.03123	0.02622	1.19	0.2345
INCOME	1	0.15485	0.05773	2.68	0.0077
GENDER	1	0.01409	0.07311	0.19	0.8473

Muthu Priya Shanmugakani Velsamy
DSBA-6276 Consumer Analytics
Linear Regression

Task #5. Dichotomous Categorical Variable



Analysis of Variance

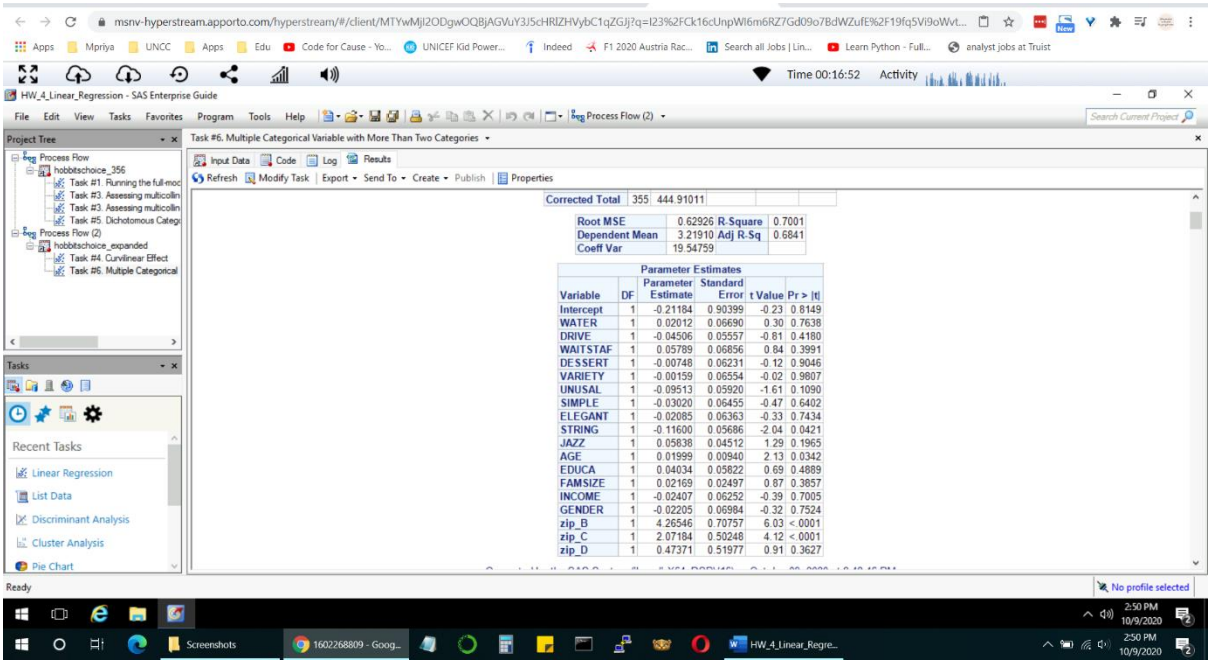
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	16	293.25597	18.32850	40.97	< .0001
Error	339	151.65414	0.44736		
Corrected Total	355	444.91011			

Root MSE	0.66885	R-Square	0.6591
Dependent Mean	3.21910	Adj R-Sq	0.6430
Coeff Var	20.77747		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.34265	0.78494	-0.44	0.6627
LISRADIO	1	-0.32782	0.20808	-1.58	0.1161
WATER	1	0.02877	0.06330	0.45	0.6498
DRIVE	1	0.01620	0.05745	0.28	0.7781
WAITSTAF	1	0.15900	0.07107	2.24	0.0259
DESSERT	1	0.08213	0.06418	1.28	0.2015
VARIETY	1	-0.01974	0.06170	-0.32	0.7492
UNUSAL	1	0.02750	0.05947	0.46	0.6441
SIMPLE	1	-0.10028	0.06722	-1.49	0.1367
ELEGANT	1	0.04195	0.06594	0.64	0.5251
STRING	1	-0.02220	0.05855	-0.38	0.7048
JAZZ	1	0.07987	0.04581	1.74	0.0821
AGE	1	0.02034	0.00728	2.79	0.0055
EDUCA	1	0.20496	0.04994	4.10	< .0001
FAMSIZE	1	0.03544	0.02640	1.34	0.1803

Task #6. Multiple Categorical Variable with More Than Two Categories



Corrected Total	355	444.91011	

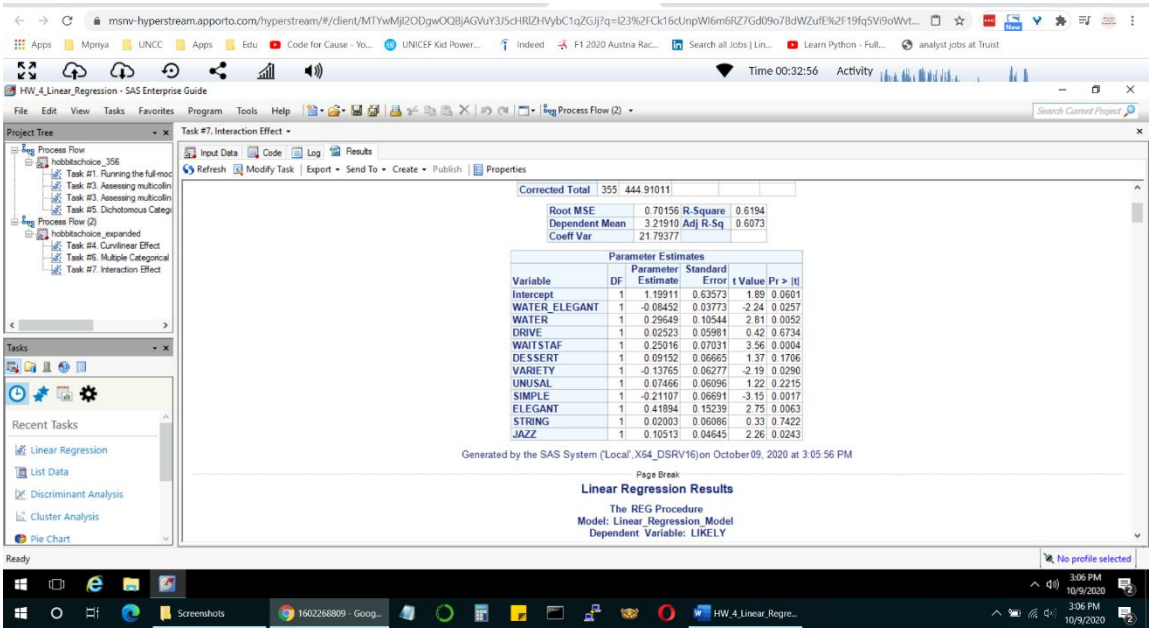
Root MSE	0.62926	R-Square	0.7001
Dependent Mean	3.21910	Adj R-Sq	0.6841
Coeff Var	19.54759		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.21184	0.90399	-0.23	0.8149
WATER	1	0.02012	0.06690	0.30	0.7638
DRIVE	1	-0.04506	0.05557	-0.81	0.4180
WAITSTAF	1	0.05789	0.06856	0.84	0.3991
DESSERT	1	-0.00748	0.06231	-0.12	0.9046
VARIETY	1	-0.00159	0.06554	-0.02	0.9807
UNUSAL	1	-0.09513	0.05920	-1.61	0.1090
SIMPLE	1	-0.03020	0.06455	-0.47	0.6402
ELEGANT	1	-0.02085	0.06363	-0.33	0.7434
STRING	1	-0.11600	0.05686	-2.04	0.0421
JAZZ	1	0.05838	0.04512	1.29	0.1965
AGE	1	0.01999	0.00940	2.13	0.0342
EDUCA	1	0.04034	0.05822	0.69	0.4889
FAMSIZE	1	0.02169	0.02497	0.87	0.3857
INCOME	1	-0.02407	0.06252	-0.39	0.7005
GENDER	1	-0.02205	0.06984	-0.32	0.7524
zip_B	1	4.26546	0.70757	6.03	< .0001
zip_C	1	2.07184	0.50248	4.12	< .0001
zip_D	1	0.47371	0.51977	0.91	0.3627

Muthu Priya Shanmugakani Velsamy
DSBA-6276 Consumer Analytics
Linear Regression

Task #7. Interaction Effect



Corrected Total	355	444.91011	

Root MSE	0.70156	R-Square	0.6194
Dependent Mean	3.21910	Adj R-Sq	0.6073
Coeff Var	21.79377		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.19911	0.63573	1.89	0.0601
WATER_ELEGANT	1	-0.06452	0.03773	-2.24	0.0257
WATER	1	0.29649	0.10544	2.81	0.0052
DRIVE	1	0.02523	0.05981	0.42	0.6734
WAITSTAF	1	0.25016	0.07031	3.56	0.0004
DESSERT	1	0.09152	0.06665	1.37	0.1706
VARIETY	1	-0.13765	0.06277	-2.19	0.0290
UNUSAL	1	0.07466	0.06096	1.22	0.2215
SIMPLE	1	-0.21107	0.06691	-3.15	0.0017
ELEGANT	1	0.41894	0.15239	2.75	0.0063
STRING	1	0.02003	0.06086	0.33	0.7422
JAZZ	1	0.10513	0.04645	2.26	0.0243

Generated by the SAS System ('Local',X64_DSRV16)on October09, 2020 at 3:05:56 PM

Page Break

Linear Regression Results

The REG Procedure
Model: Linear_Regression_Model
Dependent Variable: LIKELY