**Question 1:**

Model1:

*Call:*

*lm(formula = log(Price) ~ Weight + ColourPurity + Clarity + Certifier,*

*data = diamonds)*

*Residuals:*

*Min 1Q Median 3Q Max*

*-0.31236 -0.11520 0.01613 0.10833 0.36339*

*Coefficients:*

*Estimate Std. Error t value Pr(>|t|)*

*(Intercept) 6.086094 0.041873 145.346 < 2e-16 \*\*\**

*Weight 2.855013 0.036968 77.230 < 2e-16 \*\*\**

*ColourPurityD 0.416557 0.041382 10.066 < 2e-16 \*\*\**

*ColourPurityE 0.387047 0.030824 12.557 < 2e-16 \*\*\**

*ColourPurityF 0.310198 0.027479 11.288 < 2e-16 \*\*\**

*ColourPurityG 0.210207 0.028359 7.412 1.32e-12 \*\*\**

*ColourPurityH 0.128681 0.028523 4.511 9.31e-06 \*\*\**

*ClarityIF 0.298541 0.033303 8.964 < 2e-16 \*\*\**

*ClarityVS1 0.096609 0.024919 3.877 0.00013 \*\*\**

*ClarityVVS1 0.297835 0.028102 10.598 < 2e-16 \*\*\**

*ClarityVVS2 0.201923 0.025344 7.967 3.56e-14 \*\*\**

*CertifierHRD -0.008856 0.020864 -0.424 0.67155*

*CertifierIGI -0.182711 0.024952 -7.323 2.33e-12 \*\*\**

*---*

*Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1*

*Residual standard error: 0.1382 on 295 degrees of freedom*

*Multiple R-squared: 0.9723, Adjusted R-squared: 0.9712*

*F-statistic: 863.6 on 12 and 295 DF, p-value: < 2.2e-16*

* For the weight coefficient: Since the dependent variable is expressed in log, then we can say either:
* In log scale: If you increase one unit of cartage , log(price) will increase by 2.855
* In Price scale: If you increase one unit of cartage the price of diamond is multiplied by a factor of exp (2.885) =17.37466
* For the color purity coefficient: Since the dependent variable is expressed in log and reference category is “I”, then we can say that the price of “ColourPurityD” is exp(0.41)=1.53 times the price of “ColourPurityI” , we can conclude that “ColourPurityD” is much better than the reference category.
* For the clarity coefficient Since the dependent variable is expressed in log and reference category for clarity is “VS2 ”, then we can say that the price of “ClarityIF” is exp(0.298541)=1.34 times the price of “ClarityVS2”, we can conclude that “ClarityIF” is much better than the reference category.
* For the certifiers coefficient, we can see that the HRD and GIA are better than IGI but with a very small difference ?(YES/NO)

Question 2:

Model 2:

*Call:*

*lm(formula = log(Price) ~ Weight + ColourPurity + Clarity + Certifier +*

*Carat\_Size + Carat\_Size \* Weight, data = diamonds)*

*Residuals:*

*Min 1Q Median 3Q Max*

*-0.188079 -0.033598 -0.000428 0.043654 0.141349*

*Coefficients:*

*Estimate Std. Error t value Pr(>|t|)*

*(Intercept) 5.494410 0.029288 187.600 < 2e-16 \*\*\**

*Weight 4.413138 0.070905 62.240 < 2e-16 \*\*\**

*ColourPurityD 0.428864 0.017996 23.832 < 2e-16 \*\*\**

*ColourPurityE 0.350425 0.012922 27.118 < 2e-16 \*\*\**

*ColourPurityF 0.275029 0.011563 23.785 < 2e-16 \*\*\**

*ColourPurityG 0.190772 0.011868 16.074 < 2e-16 \*\*\**

*ColourPurityH 0.111002 0.011916 9.315 < 2e-16 \*\*\**

*ClarityIF 0.310467 0.014231 21.817 < 2e-16 \*\*\**

*ClarityVS1 0.067862 0.010495 6.466 4.29e-10 \*\*\**

*ClarityVVS1 0.213581 0.011994 17.808 < 2e-16 \*\*\**

*ClarityVVS2 0.132259 0.010757 12.295 < 2e-16 \*\*\**

*CertifierHRD -0.004912 0.008805 -0.558 0.5773*

*CertifierIGI -0.021370 0.011505 -1.858 0.0643 .*

*Carat\_SizeMedium 1.057243 0.033215 31.831 < 2e-16 \*\*\**

*Carat\_Sizelarge 2.333086 0.404712 5.765 2.10e-08 \*\*\**

*Weight:Carat\_SizeMedium -2.034513 0.075729 -26.866 < 2e-16 \*\*\**

*Weight:Carat\_Sizelarge -3.334498 0.399954 -8.337 3.17e-15 \*\*\**

*---*

*Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1*

*Residual standard error: 0.05757 on 288 degrees of freedom*

*Multiple R-squared: 0.9952, Adjusted R-squared: 0.995*

*F-statistic: 3760 on 16 and 288 DF, p-value: < 2.2e-16*

The question: interpret the interaction parameter med\*weight (carat)

My answer: using the exponential  
Interpretation of carat\*carat\_med is given by exp(weight) + exp(Weight:Carat\_SizeMedium) = (82.52804) + (0.1307445) = 82.6 , this means that a small sized carat is 82.6 (0.82) of the price of the medium

NOT using exponential  
The coefficient for carat\_med = 4.4 - 2.04 =2.38, since the diamonds carat weight is referenced to “small” carat weight, so we can say that a small sized carat is 2.3(0.023%) of the price of the medium sized carat(for the large sized , we can say that a small sized carat is equal to 1.08 (0.108%) of the price of the medium sized carat

**Arminda Notes:**

Model 1:  
  
In the log scale: if you increase one unit of caratage,  log(price) will increase in average by 2.855.  
  
In the price scale:  if you increase one unit of caratage, the price is multiplied by a factor of exp(2.885)=17.9. Is multiplied by, not is higher by 17.35.  
  
Color purity D: Compared to a diamond of color "I" (the reference class), the price of a diamond "D" is exp(0.416)=1.53 times the price of an "I". Your conclusion is Ok.  
  
For clarity: explain a little bit.  
  
For certifier: Ok.  
  
  
Model 2:  
  
In the log scale (not using exponential): the effect on the response variable is additive. So, weight has a coefficient of 4.4 for small diamonds. For medium size it is 4.4- 2.04=2.38, which means that for small diamonds the rate of change in log(price) is higher as a function of carats than for medium size.  
  
Using exp: the effect on the response variable is multiplicative. All other things being equal, for a 1 unit incerease in carat, the increase in the price of a medium diamond is exp(-2.04) times the increase in the price of a small one. Price increases faster for small diamonds than medium ones when you increase carat by one unit.  
  
Note 1: perhaps to increase carat in one unit is too much. May be you want to study the effect on price when you increase carat by, say, 0.1 units and see how everything changes.  
  
Note 2: Are you using the whole data set? My coefficients have small differences with yours.