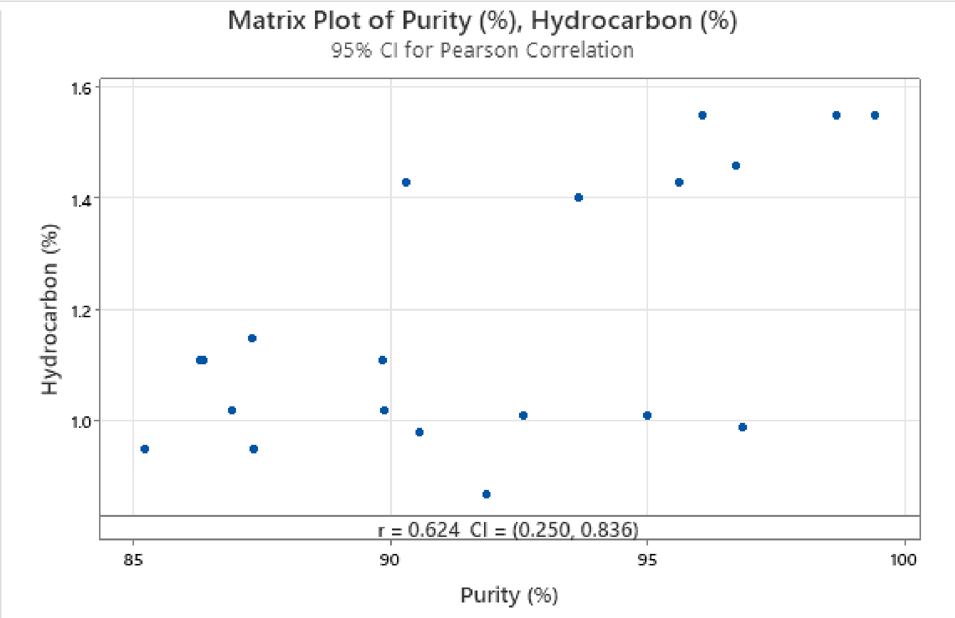
2.8)

a. From the below chart, we can see that the correlation is .624.

b. The confidence interval (0.250, 0.836) represents a range within which the true population correlation coefficient ρ is likely to fall with a 95% level of confidence. Since the interval does not contain 0 (ρ = 0), it suggests that there is a statistically significant positive correlation between oxygen purity and hydrocarbon percentage in the population. Therefore, we have evidence to reject the null hypothesis (H0: ρ = 0) and conclude that there is a statistically significant correlation between these variables.

c. 95% CI for ρ = (0.250, 0.836)

Correlation: Purity (%), Hydrocarbon (%)



Method

Correlation type	Pearson
Number of rows used	20

Correlations

	Purity (%)
Hydrocarbon (%)	0.624

2.18)

a. See regression output from MiniTab on page 3.

b. Based on the results of the analysis of variance (ANOVA) (*see page 3*), where the p-value associated with the F-statistic is less than the significance level (α = 0.05), we can conclude that there is a statistically significant relationship between the amount a company spends on advertising and the retained impressions. In other words, the amount spent on advertising significantly affects the number of retained impressions.

c. See line plot on page 3 for the 95% confidence and prediction bands.

d.

$\hat{y}_0$ =50.7 as the predicted value we're using.

$t_{\frac{a}{2},n-2} = 1.729$ using a t-distribution on StatKey (*pictured on page 3*).
MS_{RES}=552.32 from ANOVA table (*pictured on page 3*).
$n = 21$.
x_0=26.9 as the observed x-value we are using.
$\bar{x}$=50.4 as the average of all the observed x-values.
$S_{XX} = \sum(x_i - \bar{x})^2$=58556.08.

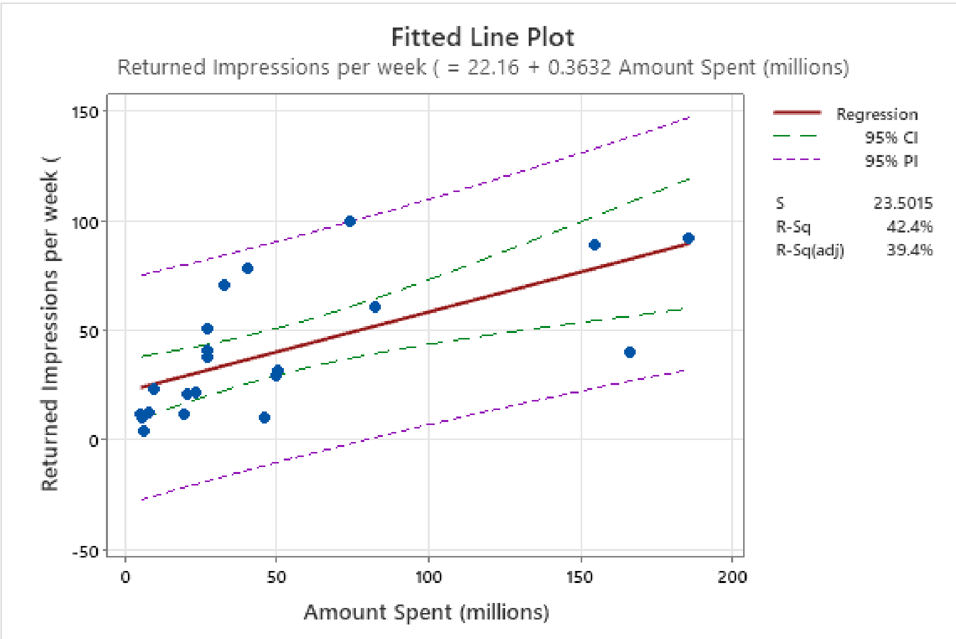
$\hat{\mu}_{y|x_0}$= 22.16+0.3632(26.9)=31.93008

Prediction interval:

$$\hat{y}_0 \pm t_{\frac{a}{2},n-2}\sqrt{MS_{Res}(1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{XX}}} = 50.7 \pm 1.729\sqrt{552.32(1+\frac{1}{21}+\frac{(26.9-50.4)^2}{58556.08}}$$
$$= (8.922910564,92.47708944)$$

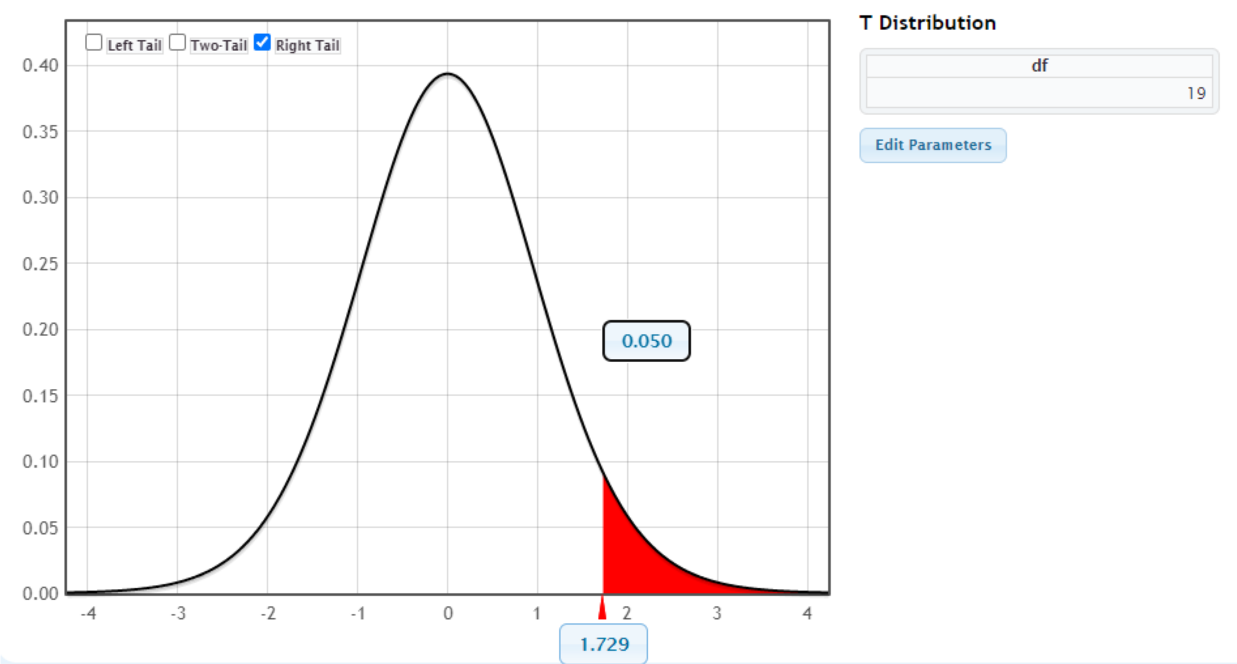
Confidence Interval:

$$\hat{\mu}_{y|x_0} \pm t_{\frac{a}{2},n-2}\sqrt{MS_{Res}(\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{XX}}} = 31.93008 \pm 1.729\sqrt{552.32(\frac{1}{21}+\frac{(26.9-50.4)^2}{58556.08}} =$$
$$(22.22455845,41.63560155)$$



Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	7723.3	7723.28	13.98	0.001
Error	19	10494.1	552.32		
Total	20	18217.4			



2.19)

a. See graph below for fitted regression equation.

b. Although the two linear models look very similar, there are substantial differences.

R-Squared:

The 'age' model with an R-squared value of approximately 0.758 explains a larger proportion of the variability in the dependent variable compared to the 'severity' model, which has an R-squared value of approximately 0.427.

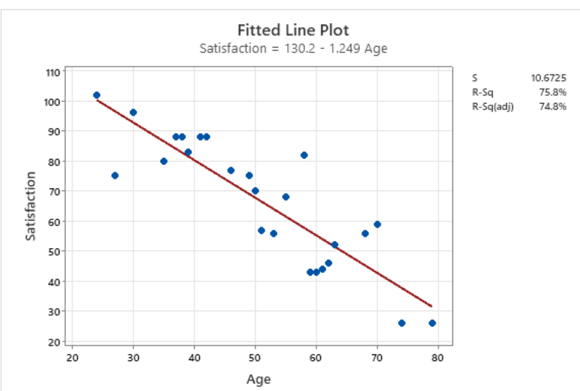
SSR/SSE:

Age model has a larger SSR (8211.3) compared to Severity model (4620) (*see ANOVA tables below*). This suggests that Age model explains more of the variability in the dependent variable compared to Severity model. Severity model has a smaller SSE (6210.558) compared to Age model (2619). This means that the residuals (unexplained variation) in Severity model are smaller than in Age model.

F-Statistic:

Age Model (F = 72.09): This high F-statistic suggests that the regression model in the Age Model is highly significant. In other words, the predictor variable(s) in the Age Model collectively have a significant effect on explaining the variability in the dependent variable.

Severity Model (F = 17.1114): The F-statistic in the Severity Model is lower than that in the Age Model. While it's still a positive value, it indicates a weaker overall significance compared to the Age Model. This suggests that the predictor variable(s) in the Severity Model may not collectively explain as much of the variability in the dependent variable as those in the Age Model.



Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	8211.3	8211.30	72.09	0.000
Error	23	2619.7	113.90		
Total	24	10831.0			

Severity Model Summary:

Summary of Fit

RSquare	0.426596
RSquare Adj	0.401666
Root Mean Square Error	16.43242
Mean of Response	66.72
Observations (or Sum Wgts)	25

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	4620.482	4620.48	17.1114
Error	23	6210.558	270.02	**Prob > F**
C. Total	24	10831.040		0.0004*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	115.6239	12.27059	9.42	<.0001*

2.36)

a.

$H_0: \beta_1 \leq 0, H_a: \beta_1 > 0$

Because $p < 0.05$, we can reject H_0 and conclude that there is a statistically significant positive linear relationship between Median Price sq/ft (x) and Rental Price (y).

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	14284629	14284629	421.02	0.000
Error	49	1662506	33929		
Total	50	15947135			

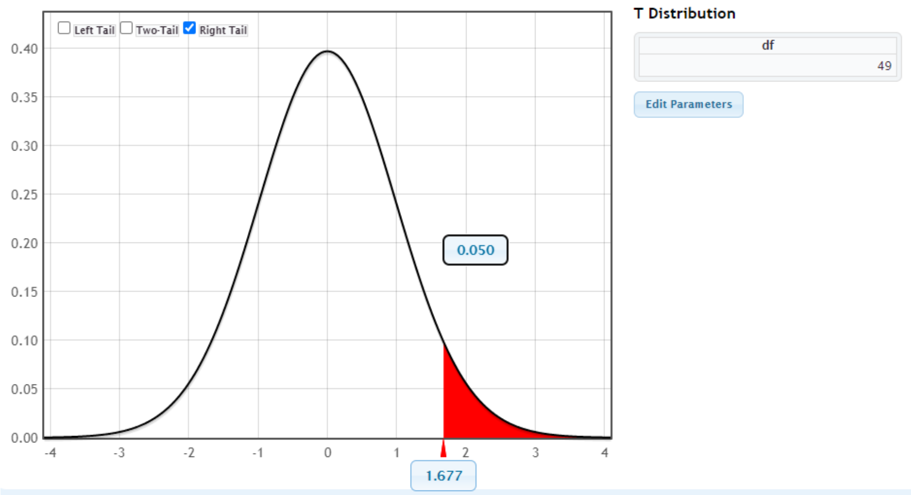
b.

(See page 6 for diagrams used to come to these conclusions)

$x\ COEF \pm t * (x\ SE\ COEF) = 6.025 \pm 1.677(0.294) = (5.531962, 6.518038)$

c.

R-squared is 89.64%, indicating that approximately 89.64% of the variability in Rental Price can be explained by the model, which includes Median Price Sq/Ft as a predictor. This suggests that the model, with Median Price Sq/Ft as a predictor, does a good job of explaining the variability in Rental Price



Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	14284629	14284629	421.02	0.000
Error	49	1662506	33929		
Total	50	15947135