



```

Pr(>|t|)
## (Intercept) -5.543e-01 1.987e-01 -2.790
0.00541
## avg_salary 1.333e-04 6.044e-04 0.221
0.82543
## age 8.591e-04 4.757e-04 1.806
0.07135
## Founded 1.466e-04 5.376e-05 2.728
0.00654
## SectorAccounting & Legal 4.540e+00 6.175e-01 7.352
5.36e-13
## SectorAerospace & Defense 4.241e+00 2.426e-01 17.479 <
2e-16
## SectorAgriculture & Forestry 4.788e+00 6.270e-01 7.637
7.16e-14
## SectorArts, Entertainment & Recreation 3.827e+00 3.661e-01 10.456 <
2e-16
## SectorBiotech & Pharmaceuticals 3.716e+00 2.189e-01 16.981 <
2e-16
## SectorBusiness Services 4.123e+00 2.186e-01 18.860 <
2e-16
## SectorConstruction, Repair & Maintenance 3.523e+00 4.050e-01 8.699 <
2e-16
## SectorConsumer Services 4.138e+00 3.576e-01 11.572 <
2e-16
## SectorEducation 3.393e+00 2.415e-01 14.049 <
2e-16
## SectorFinance 3.942e+00 2.327e-01 16.939 <
2e-16
## SectorGovernment 3.494e+00 2.789e-01 12.528 <
2e-16
## SectorHealth Care 3.719e+00 2.316e-01 16.056 <
2e-16
## SectorInformation Technology 4.145e+00 2.170e-01 19.099 <
2e-16
## SectorInsurance 3.730e+00 2.252e-01 16.565 <
2e-16
## SectorManufacturing 3.393e+00 2.332e-01 14.549 <
2e-16
## SectorMedia 3.528e+00 3.219e-01 10.959 <
2e-16
## SectorMining & Metals 4.109e+00 4.042e-01 10.165 <
2e-16
## SectorNon-Profit 4.380e+00 2.729e-01 16.046 <
2e-16
## SectorOil, Gas, Energy & Utilities 4.051e+00 2.664e-01 15.208 <
2e-16
## SectorReal Estate 4.129e+00 2.990e-01 13.810 <
2e-16
## SectorRetail 3.291e+00 2.627e-01 12.528 <

```

```

2e-16
## SectorTelecommunications      3.893e+00  3.222e-01  12.084  <
2e-16
## SectorTransportation & Logistics  4.055e+00  2.987e-01  13.574  <
2e-16
## SectorTravel & Tourism          4.024e+00  3.001e-01  13.410  <
2e-16
##
## (Intercept)                    **
## avg_salary
## age                            .
## Founded                        **
## SectorAccounting & Legal        ***
## SectorAerospace & Defense       ***
## SectorAgriculture & Forestry    ***
## SectorArts, Entertainment & Recreation ***
## SectorBiotech & Pharmaceuticals ***
## SectorBusiness Services        ***
## SectorConstruction, Repair & Maintenance ***
## SectorConsumer Services        ***
## SectorEducation                ***
## SectorFinance                  ***
## SectorGovernment               ***
## SectorHealth Care              ***
## SectorInformation Technology    ***
## SectorInsurance                ***
## SectorManufacturing            ***
## SectorMedia                    ***
## SectorMining & Metals          ***
## SectorNon-Profit               ***
## SectorOil, Gas, Energy & Utilities ***
## SectorReal Estate              ***
## SectorRetail                   ***
## SectorTelecommunications       ***
## SectorTransportation & Logistics ***
## SectorTravel & Tourism         ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5888 on 714 degrees of freedom
## Multiple R-squared:  0.4796, Adjusted R-squared:  0.46
## F-statistic: 24.37 on 27 and 714 DF,  p-value: < 2.2e-16

# Extract model coefficients
coefficients <- coef(model)

# Print the fitted model equation
cat("Fitted Model Equation: Rating =", coefficients[1], "+", coefficients[2],
    "* avg_salary +",

```

```
coefficients[3], "* age +", coefficients[4], "* Founded +",  
coefficients[5], "* Sector\n")  
  
## Fitted Model Equation: Rating = -0.554283 + 0.0001333508 * avg_salary +  
0.0008590598 * age + 0.0001466214 * Founded + 4.540153 * Sector
```

#### #Interpretation of model results

R-squared is 0.4796 this indicates that 47.96% variation in ratings is explained by the explanatory variables in the fitted model that is, average salary, age, founded and sector, therefore the model provides moderate fit to the data.

Adjusted R<sup>2</sup> penalizes the inclusion of unnecessary predictors. After accounting for the number of predictors, the model explains 46% of the variability in Rating. The small drop from R<sup>2</sup> indicates that most predictors are contributing meaningfully to the model.

Intercept: -0.554 indicates the baseline rating when all predictors are zero. avg\_salary: Coefficient (0.000133) is not significant (p-value = 0.825), implying no strong relationship with Rating.

age: Coefficient (0.000859) is marginally significant (p-value = 0.071).

Founded: Coefficient (0.000147) is significant (p-value = 0.006), suggesting older companies may have slightly higher ratings.

Sector: Most sector coefficients are highly significant, indicating that industry type strongly influences ratings.

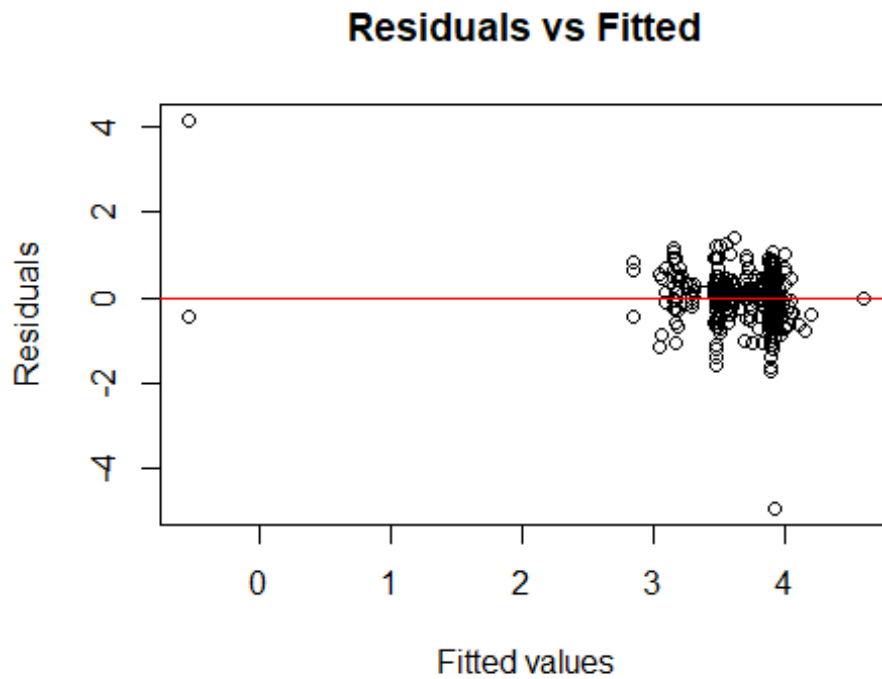
The intercept having the p\_values 0.00541 less than 0.05 suggest that it is statistically significant in the model

The predictor variables founded and Sector also has p\_values less than 0.05 indicating that they are statistically significant in the model.

F-statistic has Value: 24.37 (p-value < 2.2e-16) less than 0.05 indicates that the model is statistically significant, meaning at least one predictor is strongly associated with Rating.

Residual Standard Error (RSE) has Value: 0.5888, this indicates that the model's predictions deviate from the observed Rating by about 0.589 units on average.

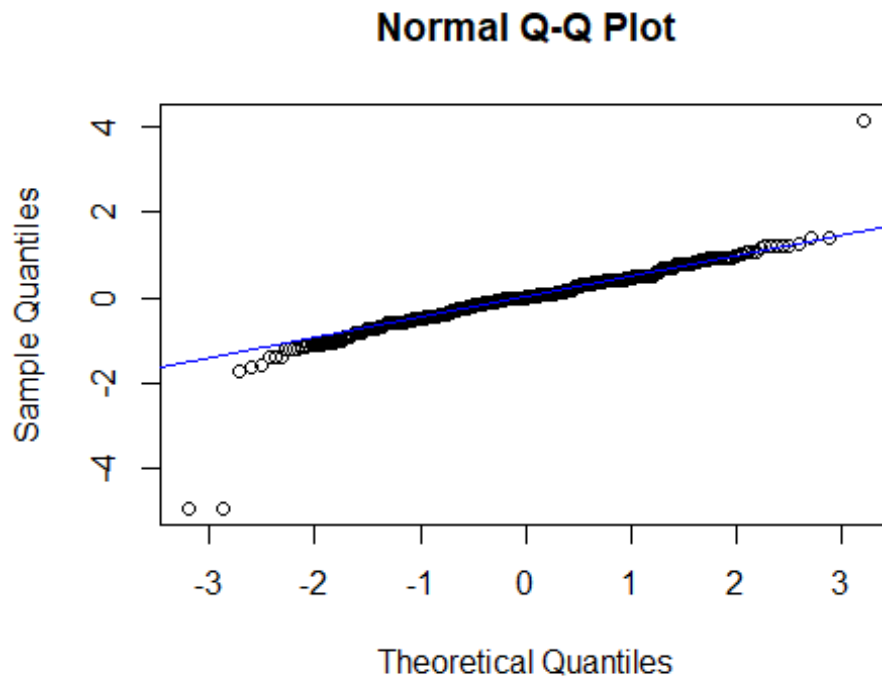
```
# Plot residuals vs fitted values  
plot(model$fitted.values, model$residuals,  
      xlab = "Fitted values", ylab = "Residuals",  
      main = "Residuals vs Fitted")  
abline(h = 0, col = "red")
```



#### #Interpretation

Residuals range from approximately -4.9 to 4.1, with a median near zero, suggesting no major skewness.

```
# Q-Q plot for normality  
qqnorm(model$residuals)  
qqline(model$residuals, col = "blue")
```



#Interpretation

Since the residuals lie close to the line in the Q-Q plot, it indicates that there is normality.

```
# Perform Breusch-Pagan test for homoscedasticity
bptest(model)

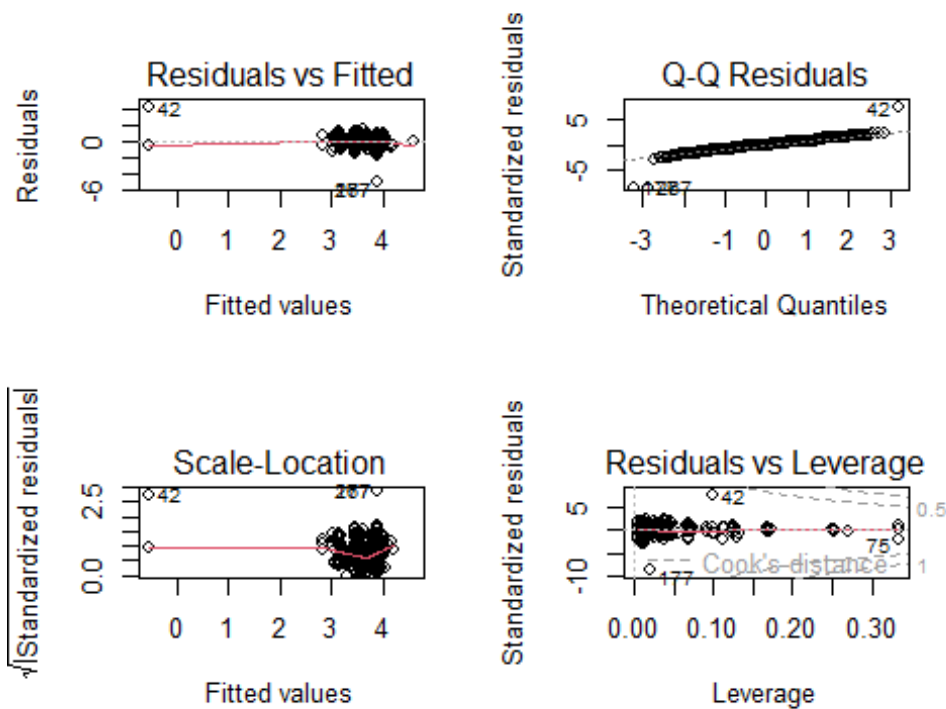
##
##  studentized Breusch-Pagan test
##
## data:  model
## BP = 36.529, df = 27, p-value = 0.1042
```

#Interpretation

BP=36.529,p-value=0.1042: Residual variance is homoscedastic (constant variance), since the BP value is higher and the the p\_value is greater than 0.05.

```
# Generate diagnostic plots
par(mfrow = c(2, 2)) # Set plot layout
plot(model) # This will generate diagnostic plots

## Warning: not plotting observations with leverage one:
## 116, 450
```



#interpretation

Residuals vs. Fitted: No clear pattern, indicating linearity assumption holds.

Q-Q Plot: Deviations suggest residuals deviate from normality.

Histogram of Residuals: Confirms lack of normality.

Scale-Location Plot: Shows some heteroscedasticity, but Breusch-Pagan suggests it's not severe.

Autocorrelation Plot: No major lags in residual autocorrelation.

*# Test assumptions of linear regression*

*# 1. Linearity assumption*

```
plot(eda_data$avg_salary, eda_data$Rating,
     main = "Scatter plot of avg_salary vs Rating",
     xlab = "avg_salary", ylab = "Rating")
abline(lm(Rating ~ avg_salary, data = eda_data), col = "red")
```

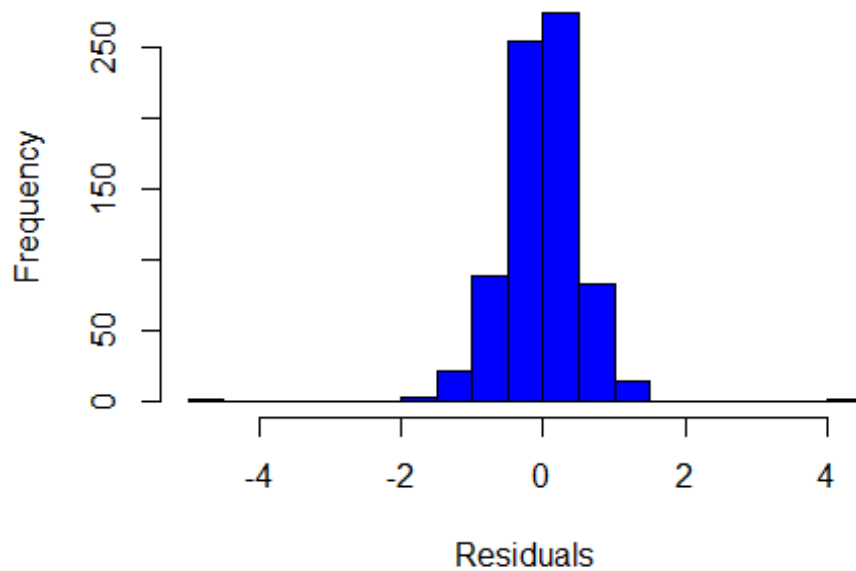
**Scatter plot of avg\_salary vs Rating**



*# 2. Normality of residuals*

```
hist(model$residuals, breaks = 20, main = "Histogram of Residuals",  
      xlab = "Residuals", col = "blue")
```

**Histogram of Residuals**





```
shapiro.test(model$residuals) # Shapiro-Wilk test for normality
```

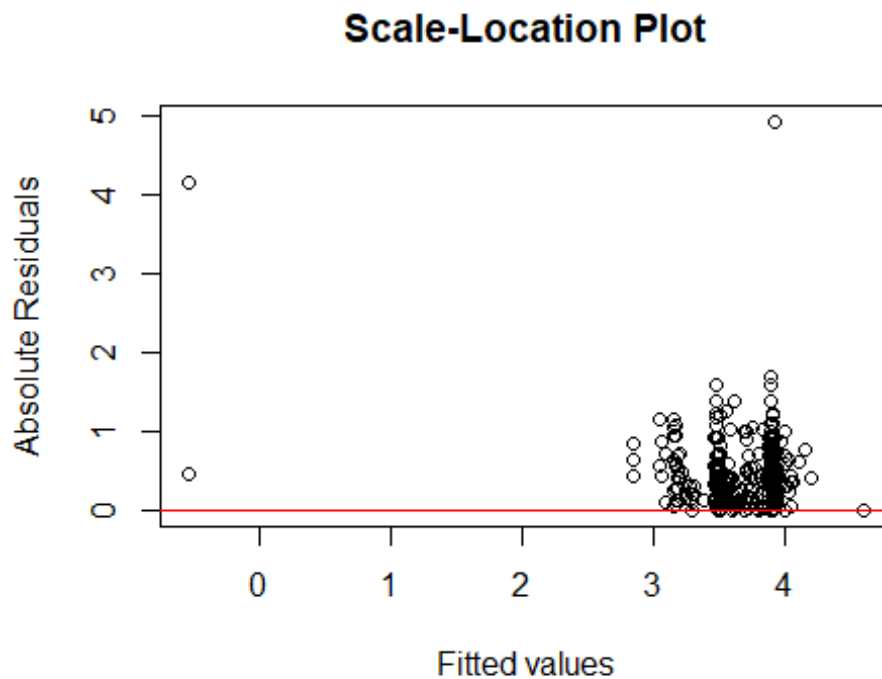
```
##  
## Shapiro-Wilk normality test  
##  
## data: model$residuals  
## W = 0.88252, p-value < 2.2e-16
```

#Interpretation

W=0.882, p-value<2.2e-16: Residuals are not normally distributed, since p-value is less than the level of significance = 0.05.

# 3. Homoscedasticity

```
plot(model$fitted.values, abs(model$residuals),  
      xlab = "Fitted values", ylab = "Absolute Residuals",  
      main = "Scale-Location Plot")  
abline(h = 0, col = "red")
```

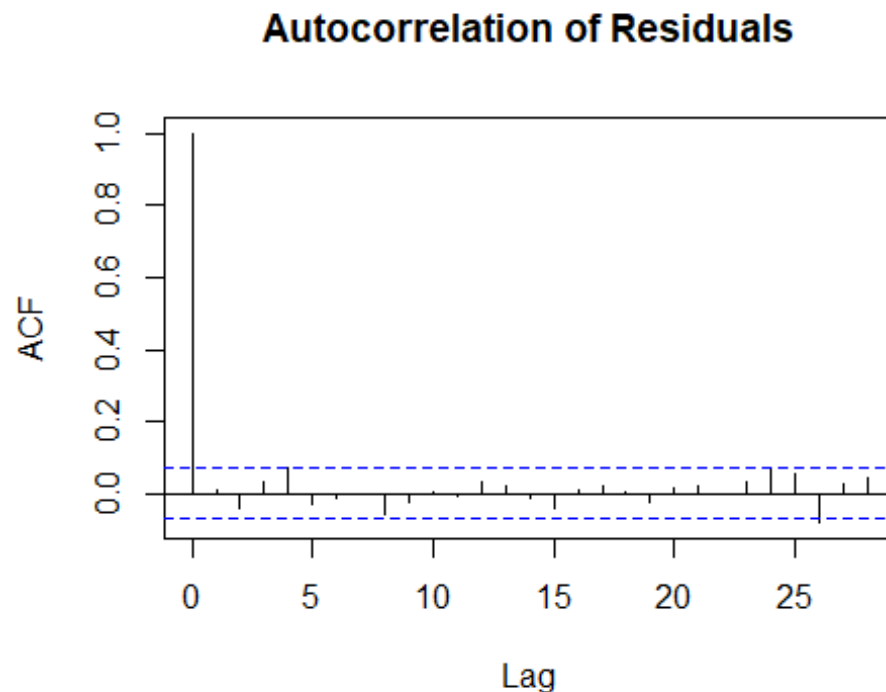


```
bptest(model) # Breusch-Pagan test
```

```
##  
## studentized Breusch-Pagan test  
##  
## data: model  
## BP = 36.529, df = 27, p-value = 0.1042
```

```
# 4. Independence of residuals
```

```
acf(model$residuals, main = "Autocorrelation of Residuals")
```



```
durbinWatsonTest(model) # Durbin-Watson test for autocorrelation
```

```
## lag Autocorrelation D-W Statistic p-value  
## 1 0.006903912 1.985587 0.812  
## Alternative hypothesis: rho != 0
```

#Interpretation

D-W statistic=1.985,p-value=0.852: No significant autocorrelation in residuals since the D-W value 1.985 is very close to 2 with p-value greater than 0.05.

```
# 5. Multicollinearity
```

```
vif_values <- vif(model)  
print(vif_values)
```

```
##          GVIF Df GVIF^(1/(2*Df))  
## avg_salary 1.178663 1      1.085663  
## age        1.398784 1      1.182702  
## Founded    1.526767 1      1.235624  
## Sector     2.424985 24      1.018626
```

#Interpretation

Since all the vif values are less than 5 it indicates that no multicollinearity that is there is no significant relationships among the predictors.