1. **Delivery Time:**

> df <- read.csv("delivery\_time.csv")

> View(df)

> #EDA

> boxplot(df$Sorting.Time)

> boxplot(df$Delivery.Time)

> plot(df$Sorting.Time, df$Delivery.Time, main = "Scatter Plot")

> cor(df$Sorting.Time,df$Delivery.Time)

[1] 0.8259973

> summary(df)

Delivery.Time Sorting.Time

Min. : 8.00 Min. : 2.00

1st Qu.:13.50 1st Qu.: 4.00

Median :17.83 Median : 6.00

Mean :16.79 Mean : 6.19

3rd Qu.:19.75 3rd Qu.: 8.00

Max. :29.00 Max. :10.00

> is.na(df)

Delivery.Time Sorting.Time

[1,] FALSE FALSE

[2,] FALSE FALSE

[3,] FALSE FALSE

[4,] FALSE FALSE

[5,] FALSE FALSE

[6,] FALSE FALSE

[7,] FALSE FALSE

[8,] FALSE FALSE

[9,] FALSE FALSE

[10,] FALSE FALSE

[11,] FALSE FALSE

[12,] FALSE FALSE

[13,] FALSE FALSE

[14,] FALSE FALSE

[15,] FALSE FALSE

[16,] FALSE FALSE

[17,] FALSE FALSE

[18,] FALSE FALSE

[19,] FALSE FALSE

[20,] FALSE FALSE

[21,] FALSE FALSE

> #Linear Regression Model

> model <- lm(Delivery.Time~Sorting.Time, data=df)

> summary(model)

Call:

lm(formula = Delivery.Time ~ Sorting.Time, data = df)

Residuals:

Min 1Q Median 3Q Max

-5.1729 -2.0298 -0.0298 0.8741 6.6722

Coefficients:

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

(Intercept) 6.5827 1.7217 3.823 0.00115 \*\*

Sorting.Time 1.6490 0.2582 6.387 3.98e-06 \*\*\*

---

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

Residual standard error: 2.935 on 19 degrees of freedom

Multiple R-squared: 0.6823, Adjusted R-squared: 0.6655

F-statistic: 40.8 on 1 and 19 DF, p-value: 3.983e-06

> model$residuals

1 2 3 4 5 6 7

-2.07293294 0.32118644 3.27314665 2.57608696 5.92706706 -1.12685335 0.87412675

8 9 10 11 12 13 14

-2.02979366 -5.17293294 -2.67391304 0.05510685 -2.42881356 -1.44587325 -0.02979366

15 16 17 18 19 20 21

0.50020634 1.70118644 -2.72685335 -0.01587325 -1.88077377 -0.29587325 6.67216654

> predict(model, data.frame(Sorting.Time=c(10,9,8,7,6,5,4)))

1 2 3 4 5 6 7

23.07293 21.42391 19.77489 18.12587 16.47685 14.82783 13.17881

#log transformation

> cor(log(df$Sorting.Time),df$Delivery.Time)

[1] 0.8339325

> reg\_log <- lm(Delivery.Time~log(Sorting.Time), data=df)

> summary(reg\_log)

Call:

lm(formula = Delivery.Time ~ log(Sorting.Time), data = df)

Residuals:

Min 1Q Median 3Q Max

-4.0829 -2.0133 -0.1965 0.9351 7.0171

Coefficients:

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

(Intercept) 1.160 2.455 0.472 0.642

log(Sorting.Time) 9.043 1.373 6.587 2.64e-06 \*\*\*

---

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

Residual standard error: 2.873 on 19 degrees of freedom

Multiple R-squared: 0.6954, Adjusted R-squared: 0.6794

F-statistic: 43.39 on 1 and 19 DF, p-value: 2.642e-06

> reg\_log$residuals

1 2 3 4 5 6 7

-0.9829125 -0.1965166 2.3866948 2.9699062 7.0170875 -2.0133052 0.2426465

8 9 10 11 12 13 14

-1.5948887 -4.0829125 -2.2800938 -0.1349331 -2.9465166 -2.0773535 0.4051113

15 16 17 18 19 20 21

0.9351113 1.1834834 -3.6133052 -0.6473535 0.5718999 -0.9273535 5.7855040

> predict(reg\_log,data.frame(Sorting.Time=c(10,9,8,7,6,5,4)))

1 2 3 4 5 6 7

21.98291 21.03009 19.96493 18.75735 17.36331 15.71450 13.69652

#exponential transformation

> reg\_exp <- lm(log(Delivery.Time)~Sorting.Time, data=df)

> summary(reg\_exp)

Call:

lm(formula = log(Delivery.Time) ~ Sorting.Time, data = df)

Residuals:

Min 1Q Median 3Q Max

-0.29209 -0.13364 0.02065 0.08421 0.41892

Coefficients:

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

(Intercept) 2.12137 0.10297 20.601 1.86e-14 \*\*\*

Sorting.Time 0.10555 0.01544 6.836 1.59e-06 \*\*\*

---

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

Residual standard error: 0.1755 on 19 degrees of freedom

Multiple R-squared: 0.7109, Adjusted R-squared: 0.6957

F-statistic: 46.73 on 1 and 19 DF, p-value: 1.593e-06

> reg\_exp$residuals

1 2 3 4 5 6

-0.132365397 0.059111439 0.228472049 0.106717594 0.190407996 -0.023565969

7 8 9 10 11 12

0.084205939 -0.186734850 -0.292087121 -0.140142484 0.021411304 -0.168672492

13 14 15 16 17 18

-0.046022644 0.004320387 0.049376881 0.156439783 -0.133642618 0.036231231

19 20 21

-0.253033509 0.020649391 0.418923091

> predict(reg\_exp,data.frame(Sorting.Time=c(10,9,8,7,6,5,4)))

1 2 3 4 5 6 7

3.176888 3.071336 2.965785 2.860233 2.754681 2.649130 2.543578

#polynomial transformation

> cor(df$Sorting.Time\*df$Sorting.Time,df$Delivery.Time)

[1] 0.7939063

> reg\_poly <- lm(Delivery.Time~Sorting.Time+I(Sorting.Time^2), data=df)

> summary(reg\_poly)

Call:

lm(formula = Delivery.Time ~ Sorting.Time + I(Sorting.Time^2),

data = df)

Residuals:

Min 1Q Median 3Q Max

-4.4324 -1.6951 -0.5365 0.9075 6.6676

Coefficients:

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

(Intercept) 3.5222 4.1597 0.847 0.4082

Sorting.Time 2.8130 1.4608 1.926 0.0701 .

I(Sorting.Time^2) -0.0932 0.1151 -0.810 0.4286

---

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

Residual standard error: 2.962 on 18 degrees of freedom

Multiple R-squared: 0.6934, Adjusted R-squared: 0.6594

F-statistic: 20.36 on 2 and 18 DF, p-value: 2.391e-05

> predict(reg\_poly)

1 2 3 4 5 6 7 8

22.332430 13.283069 17.045108 21.290194 22.332430 17.045108 18.646533 11.122455

9 10 11 12 13 14 15 16

22.332430 21.290194 20.061562 13.283069 18.646533 11.122455 11.122455 13.283069

17 18 19 20 21

17.045108 18.646533 8.775444 18.646533 15.257287

> reg\_poly$residuals

1 2 3 4 5 6 7

-1.3324296 0.2169308 2.7048917 2.7098061 6.6675704 -1.6951083 0.3534667

8 9 10 11 12 13 14

-1.6224550 -4.4324296 -2.5401939 -0.2315618 -2.5330692 -1.9665333 0.3775450

15 16 17 18 19 20 21

0.9075450 1.5969308 -3.2951083 -0.5365333 -0.7754444 -0.8165333 6.2427130

> predict(reg\_poly, data.frame(Sorting.Time=c(10,9,8,7,6,5,4)))

1 2 3 4 5 6 7

22.33243 21.29019 20.06156 18.64653 17.04511 15.25729 13.28307

#cubic transformation

> cor(df$Sorting.Time^3,df$Delivery.Time)

[1] 0.7540763

> reg\_poly3 <- lm(Delivery.Time~Sorting.Time+I(Sorting.Time^2)+I(Sorting.Time^3), data=df)

> summary(reg\_poly3)

Call:

lm(formula = Delivery.Time ~ Sorting.Time + I(Sorting.Time^2) +

I(Sorting.Time^3), data = df)

Residuals:

Min 1Q Median 3Q Max

-4.8972 -1.7972 -0.1601 0.8077 6.2028

Coefficients:

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

(Intercept) -4.15818 10.98653 -0.378 0.710

Sorting.Time 7.50248 6.37003 1.178 0.255

I(Sorting.Time^2) -0.92525 1.10553 -0.837 0.414

I(Sorting.Time^3) 0.04446 0.05874 0.757 0.460

Residual standard error: 2.998 on 17 degrees of freedom

Multiple R-squared: 0.7034, Adjusted R-squared: 0.6511

F-statistic: 13.44 on 3 and 17 DF, p-value: 9.586e-05

> reg\_poly3$residuals

1 2 3 4 5 6 7

-1.7972404 -0.3928561 2.5999522 3.1730001 6.2027596 -1.8000478 0.7298599

8 9 10 11 12 13 14

-1.7222837 -4.8972404 -2.0769999 0.4231213 -3.1428561 -1.5901401 0.2777163

15 16 17 18 19 20 21

0.8077163 0.9871439 -3.4000478 -0.1601401 0.4985883 -0.4401401 5.7201347

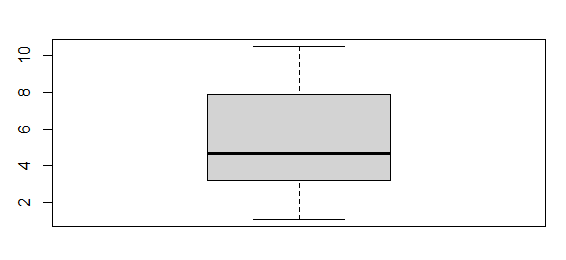
> predict(reg\_poly3,data.frame(Sorting.Time=c(10,9,8,7,6,5,4)))

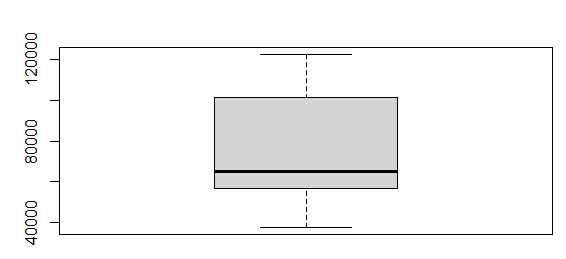
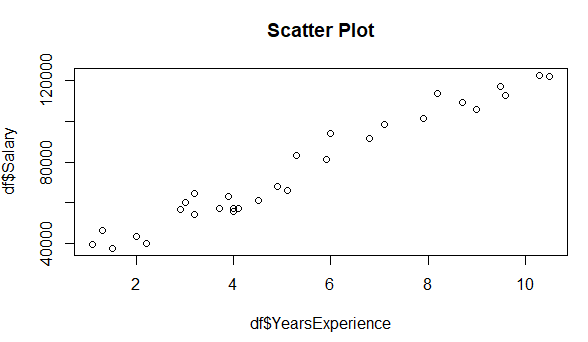
1 2 3 4 5 6 7

22.79724 20.82700 19.40688 18.27014 17.15005 15.77987 13.89286

1. **Salary Hike:**

df <- read.csv("Salary\_Data.csv")  
View(df)

#EDA  
boxplot(df$YearsExperience)  


boxplot(df$Salary)  
  
plot(df$YearsExperience, df$Salary, main="Scatter Plot")  


#Linear Regression Model  
model <- lm(Salary~YearsExperience, data=df)  
summary(model)  
Call:  
lm(formula = Salary ~ YearsExperience, data = df)

Residuals:  
 Min 1Q Median 3Q Max   
-7958.0 -4088.5 -459.9 3372.6 11448.0

Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 25792.2 2273.1 11.35 5.51e-12 \*\*\*  
YearsExperience 9450.0 378.8 24.95 < 2e-16 \*\*\*  
---  
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  
Residual standard error: 5788 on 28 degrees of freedom  
Multiple R-squared: 0.957, Adjusted R-squared: 0.9554   
F-statistic: 622.5 on 1 and 28 DF, p-value: < 2.2e-16

> predict(model,data.frame(YearsExperience=c(1.5,2.2,3.3,4.9)))  
 1 2 3 4

39967.14 46582.12 56977.08 72097.02

#log transformation  
> reg\_log <- lm(Salary~log(YearsExperience), data=df)  
> summary(reg\_log)

Call:  
lm(formula = Salary ~ log(YearsExperience), data = df)

Residuals:  
 Min 1Q Median 3Q Max   
-15392.6 -7523.0 559.7 6336.1 20629.8

Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 14928 5156 2.895 0.00727 \*\*   
log(YearsExperience) 40582 3172 12.792 3.25e-13 \*\*\*  
---  
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10660 on 28 degrees of freedom  
Multiple R-squared: 0.8539, Adjusted R-squared: 0.8487   
F-statistic: 163.6 on 1 and 28 DF, p-value: 3.25e-13

> reg\_log$residuals  
 1 2 3 4 5 6 7   
 20547.1517 20629.7648 6348.4481 467.7377 -7034.1389 -1494.0501 638.1576   
 8 9 10 11 12 13 14   
 -7685.9439 2314.0561 -10833.7185 -6941.1059 -15392.5528 -14229.5528 -15107.6281   
 15 16 17 18 19 20 21   
-14855.4226 -11484.2957 -15016.7917 481.1701 -5596.0667 6298.8670 -982.5021   
 22 23 24 25 26 27 28   
 3800.4853 2496.6286 13494.0813 6711.0792 1486.2869 10679.1316 5920.1854   
 29 30   
 12819.9928 11520.5459

> predict(reg\_log,data.frame(YearsExperience=c(1.5,2.2,3.3,4.9)))  
 1 2 3 4   
31382.55 46925.14 63379.72 79422.30

#exponential transformation  
> plot(df$YearsExperience,log(df$Salary))  
> cor(df$YearsExperience,log(df$Salary))  
[1] 0.9653844  
> reg\_exp <- lm(log(Salary)~YearsExperience, data=df)  
> summary(reg\_exp)

Call:  
lm(formula = log(Salary) ~ YearsExperience, data = df)

Residuals:  
 Min 1Q Median 3Q Max   
-0.18949 -0.06946 -0.01068 0.06932 0.19029

Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 10.507402 0.038443 273.33 <2e-16 \*\*\*  
YearsExperience 0.125453 0.006406 19.59 <2e-16 \*\*\*  
---  
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.09789 on 28 degrees of freedom  
Multiple R-squared: 0.932, Adjusted R-squared: 0.9295   
F-statistic: 383.6 on 1 and 28 DF, p-value: < 2.2e-16

> reg\_exp$residuals  
 1 2 3 4 5 6   
-0.065326727 0.070352644 -0.157343914 -0.077216910 -0.189492239 0.073290763   
 7 8 9 10 11 12   
 0.120836157 -0.003904845 0.164716284 -0.017460736 0.057676185 -0.079791838   
 13 14 15 16 17 18   
-0.059161577 -0.069532152 -0.051492737 0.004229750 -0.049362313 0.155353356   
 19 20 21 22 23 24   
 0.059101960 0.190292327 0.066210424 0.097387189 0.027381712 0.106187655   
 25 26 27 28 29 30   
 0.004207459 -0.069234719 -0.029540123 -0.079841852 -0.084590539 -0.113930641   
> predict(reg\_exp,data.frame(YearsExperience=c(1.5,2.2,3.3,4.9)))  
 1 2 3 4   
10.69558 10.78340 10.92140 11.12212   
#polynomial transformation  
> plot(df$YearsExperience\*df$YearsExperience, df$Salary, main='Scatter Plot')  
> cor(df$YearsExperience\*df$YearsExperience, df$Salary)  
[1] 0.9567235  
> reg\_poly <- lm(Salary~YearsExperience+I(YearsExperience^2), data=df)  
> summary(reg\_poly)

Call:  
lm(formula = Salary ~ YearsExperience + I(YearsExperience^2),   
 data = df)  
  
Residuals:  
 Min 1Q Median 3Q Max   
 -7835 -4026 -493 3309 11579

Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 26214.93 4554.67 5.756 4.04e-06 \*\*\*  
YearsExperience 9259.28 1811.01 5.113 2.25e-05 \*\*\*  
I(YearsExperience^2) 16.39 152.12 0.108 0.915   
---  
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5893 on 27 degrees of freedom  
Multiple R-squared: 0.957, Adjusted R-squared: 0.9538   
F-statistic: 300.3 on 2 and 27 DF, p-value: < 2.2e-16

> reg\_poly$residuals  
 1 2 3 4 5 6 7   
 2923.02004 7925.29483 -2409.74178 -1274.07072 -6773.69725 3437.28256 6009.68256   
 8 9 10 11 12 13 14   
-1567.50100 8432.49900 -3509.69730 642.52923 -7720.34929 -6557.34929 -7372.55566   
 15 16 17 18 19 20 21   
-7102.65964 -4041.00925 -7834.65116 7338.39553 -52.33285 11579.23161 1801.94462   
 22 23 24 25 26 27 28   
 5490.80245 915.66454 10568.70328 1419.54415 -5294.28554 1311.44127 -3979.79692   
 29 30   
 -933.64409 -3372.69394   
> predict(reg\_poly, data.frame(YearsExperience=c(1.5,2.2,3.3,4.9)))  
 1 2 3 4   
40140.74 46664.70 56949.08 71979.01   
#cubic transformation  
> plot((df$YearsExperience^3), df$Salary, main="Scatter Plot")  
> cor(df$YearsExperience^3,df$Salary)  
[1] 0.9133658  
> reg\_poly3 <- lm(Salary~YearsExperience+I(YearsExperience^2)+I(YearsExperience^3), data=df)  
> summary(reg\_poly3)

Call:  
lm(formula = Salary ~ YearsExperience + I(YearsExperience^2) +   
 I(YearsExperience^3), data = df)

Residuals:  
 Min 1Q Median 3Q Max   
 -7468 -4286 -1100 2639 10412

Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 38863.07 7214.75 5.387 1.21e-05 \*\*\*  
YearsExperience -718.71 4892.11 -0.147 0.8843   
I(YearsExperience^2) 2099.35 968.36 2.168 0.0395 \*   
I(YearsExperience^3) -122.92 56.52 -2.175 0.0389 \*   
---  
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5524 on 26 degrees of freedom  
Multiple R-squared: 0.9636, Adjusted R-squared: 0.9594   
F-statistic: 229.4 on 3 and 26 DF, p-value: < 2.2e-16

> reg\_poly3$residuals  
 1 2 3 4 5 6 7   
-1106.1080 4998.3895 -4362.7116 -1314.7395 -6242.9734 5205.4167 7867.6021   
 8 9 10 11 12 13 14   
 412.1234 10412.1234 -1528.9444 2517.9673 -5917.2828 -4754.2828 -5654.0203  
 15 16 17 18 19 20 21   
-5829.0938 -3347.9653 -7467.9504 7362.5647 -1093.8886 10362.2381 -663.3491   
 22 23 24 25 26 27 28   
 2677.2051 -2301.7374 7453.5384 -1139.0884 -7254.8667 851.7483 -4057.0544   
 29 30   
 2523.3688 1391.7709   
> predict(reg\_poly3,data.frame(YearsExperience=c(1.5,2.2,3.3,4.9)))  
 1 2 3 4   
42093.71 46133.97 54936.07 71285.97