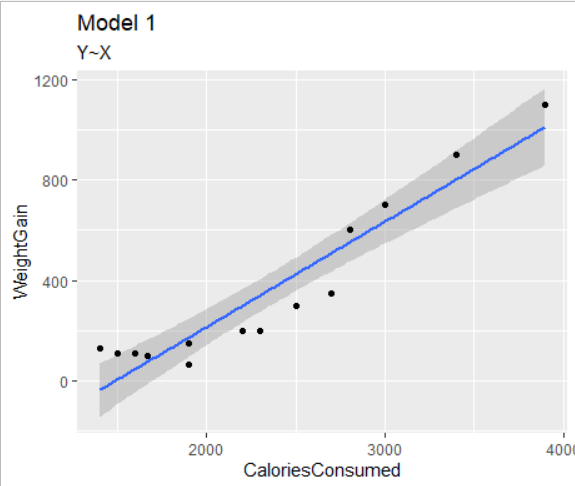
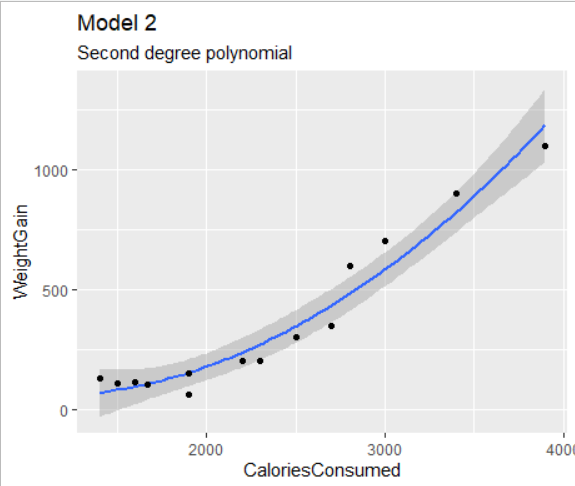
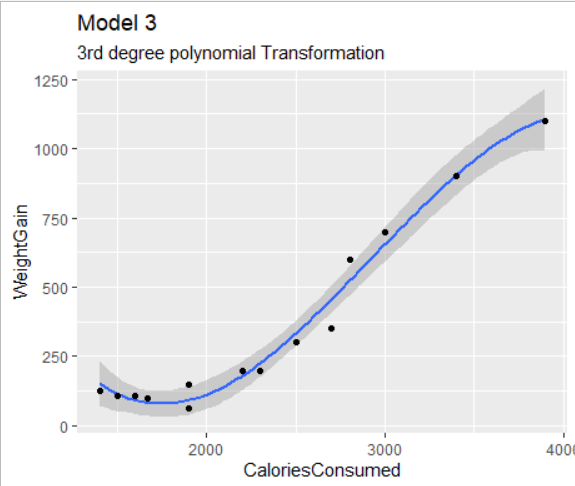
# Q1: Calories\_consumed-> predict weight gained using calories consumed

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | Y  WeightGain | X   CaloriesConsumed | Correlation  (between X and Y) | R-Square | Correlation  (between Y and Yhat) | RMSE |
| 1 | **Y** | **X** | **0.946991** | **0.8968** | **0.94699** | **103.3025** |
| 2 | **Y** | **poly(X,2)** | **--** | **0.9521** | **0.9757338** | **70.40752** |
| 3 | **Y** | **poly(X,3)** | **--** | **0.9811** | **0.9905292** | **44.15011** |

* Target variable is Weight Gain and independent variable is Calories consumed



Conclusion :

Now in our final model we are taking the 3 degree polynomial transformation of the X.

Our Model becomes

**Y = ßo + ß1 X + ß2 X² + ß3 X3**

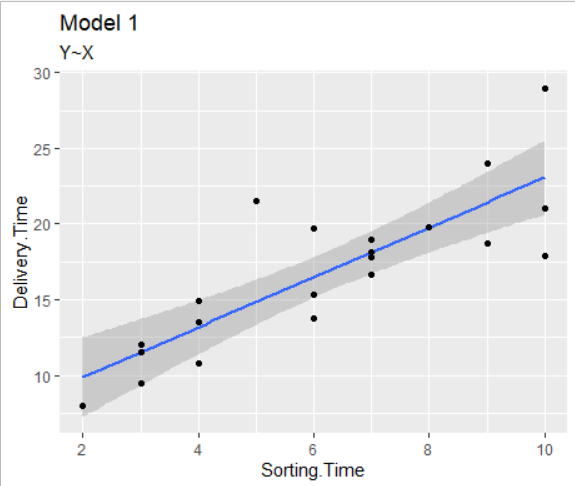
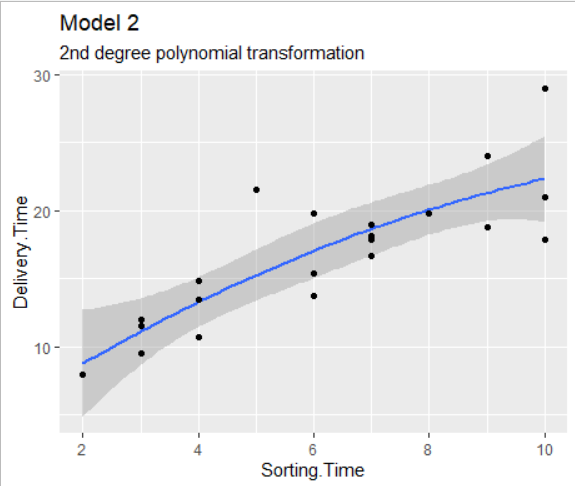
We get the High R² value and lower RMSE value in our model 3

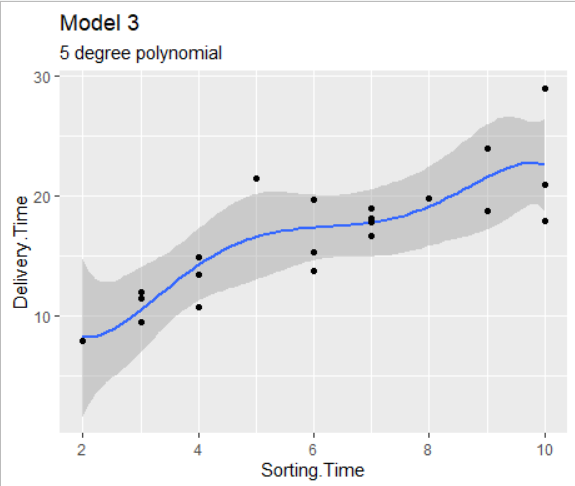
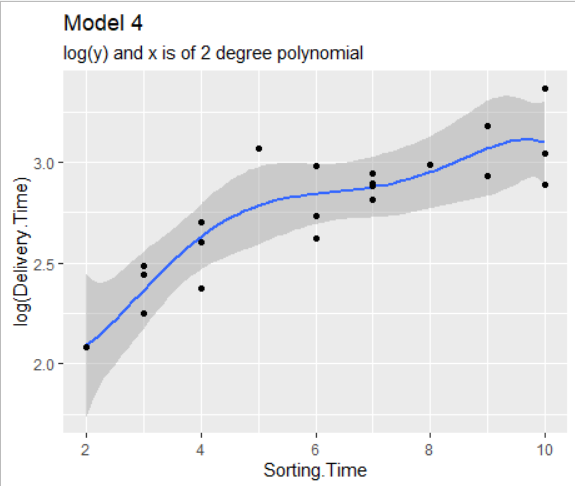
|  |  |
| --- | --- |
| Coefficients | Value |
| ß0 (Intercept) | **357.71** |
| ß1 (Slope) | **1139.37** |
| ß2 (Slope) | **282.84** |
| ß3 (Slope) | **-205.21** |

# Q2: Delivery\_time -> Predict delivery time using sorting time

## Target variable is DeliveryTime, and Independent variable is SortingTime

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | Y DeliveryTime | X SortingTime | Correlation  (between X and Y) | R-Square | Correlation  (between Y and Yhat) | RMSE |
| 1 | **Y** | **X** | **0.825599** | **0.6823** | **0.825997** | **2.79165** |
| 2 | **Y** | **poly(X,2)** | **--** | **0.6934** | **0.8327302** | **2.742148** |
| 3 | **Y** | **ploy(X,5)** | **--** | **0.7142** | **0.845121** | **2.6475** |
| 4 | **log(Y)** | **poly(X,2)** | **--** | **0.7649** | **0.8244099** | **2.799042** |





Conclusion :

Although in model 3 we consider 5 degree polynomial and get higher R² value and least RMSE value, we are not going to accept the model as the X³,X4, and X5 are insignificant in this model.

So we may consider the 4th model which has significant variable consideration.

Our model is : log(Y) = ß0 + ß1 X + ß2 X2 .

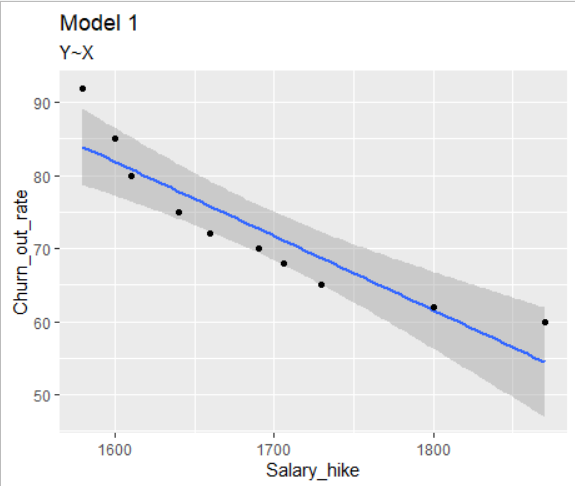
Or we can say : Y = exp(ß0 + ß1 X + ß2 X2)

|  |  |
| --- | --- |
| Coefficients | Value |
| ß0 (Intercept) | **2.77479** |
| ß1 (Slope) | **1.19994** |
| ß2 (Slope) | **-0.33045** |

# Q3: Emp\_data -> Build a prediction model for Churn\_out\_rate

## Target variable is "Churn out rate", and Independent variable is "Salary hike"

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | Y Churn out rate | X Salary hike | Correlation  (between X and Y) | R-Square | Correlation  (between Y and Yhat) | RMSE |
| 1 | **Y** | **X** | **-0.9117216** | **0.8312** | **0.9117216** | **3.997528** |
| 2 | **Y** | **poly(X,2)** | **--** | **0.9737** | **0.9867** | **1.5779** |



|  |  |
| --- | --- |
| Coefficients | Value |
| ß0 (Intercept) | **72.9** |
| ß1 (Slope) | **-28.0553** |
| ß2 (Slope) | **11.6147** |

Conclusion :

Here we will consider our second model which is with higher coefficient of determination, as well as least RMSE value. All the considered variables of the second degree polynomial i.e. X and X² are significant in this model.

Our model is : Y = ß0 + ß1 X + ß2 X2 .

# Q4: Salary\_hike -> Build a prediction model for Salary\_hike

## Target variable is " Salary", and Independent variable is " YearsExperience"

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | Y Salary | X YearsExperience | Correlation  (between X and Y) | R-Square | Correlation  (between Y and Yhat) | RMSE |
| 1 | **Y** | **X** | **0.9782416** | **0.957** | **0.9782** | **5592.044** |
| 2 | **Y** | **poly(X,2)** | **--** | **0.9636** | **0.9816** | **5142.642** |

Conclusion :

Here we will consider our First model. Although we are getting higher R² in the 2nd model, as well as lower RMSE value, but we can see that variable X² is insignificant in this model, although X³ is significant variable in this model.

Without any transformation also we are getting 0.95 coefficient of determination

Our model is : Y = ß0 + ß1 X .

|  |  |
| --- | --- |
| Coefficients | Value |
| ß0 (Intercept) | **25792.2** |
| ß1 (Slope) | **9450** |