**PROJECT ON**

**LINEAR REGRESSION OF ‘ORANGE’ DATASET USING R-LANGUAGE**

BY KHUSHBOO KUMARI

**LINEAR REGRESSION:**

Linear regression is one of the very basic forms of machine learning where we train a model to predict the behaviour of your data based on some variables. In the case of linear regression as you can see the name suggests linear that means the two variables which are on the x-axis and y-axis should be linearly correlated.

>data("Orange")

> head(Orange)

Tree age circumference

1 1 118 30

2 1 484 58

3 1 664 87

4 1 1004 115

5 1 1231 120

6 1 1372 142

>cor(Orange$circumference, Orange$age)

1. 0.9135189

> cor(Orange$age, Orange$circumference)

[1] 0.9135189

>plot(Orange$circumference,Orange$age)

> plot(Orange$age, Orange$circumference)

> model <- lm(age ~ circumference, data= Orange)

> summary(model)

Call:

lm(formula = age ~ circumference, data = Orange)

Residuals:

Min 1Q Median 3Q Max

-317.88 -140.90 -17.20 96.54 471.16

Coefficients:

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

(Intercept) 16.6036 78.1406 0.212 0.833

Circumference 7.8160 0.6059 12.900 1.93e-14 \*\*\*

---

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

Residual standard error: 203.1 on 33 degrees of freedom

Multiple R-squared: 0.8345, Adjusted R-squared: 0.8295

F-statistic: 166.4 on 1 and 33 DF, p-value: 1.931e-14

**# Predict the circumference of an orange given its age**

> predict(model,data.frame("circumference"=1700))

1

13303.8

> predict(model,data.frame("circumference"=1500))

1

11740.6

> predict(model,data.frame("circumference"=20000))

1

156336.6

> predict(model,data.frame("circumference"=23000))

1

179784.6

> predict(model,data.frame("circumference"=60000))

1

468976.5

> predict(model,data.frame("circumference"=300))

1

2361.403

> predict(model,data.frame("circumference"=19870))

1

155320.5

> predict(model,data.frame("circumference"=14590))

1

114052

> predict(model,data.frame("circumference"=13560))

1

106001.5

> predict(model,data.frame("circumference"=148653))

1

1161888

> predict(model,data.frame("circumference"=11193))

1

87501.07

> predict(model,data.frame("circumference"=12350))

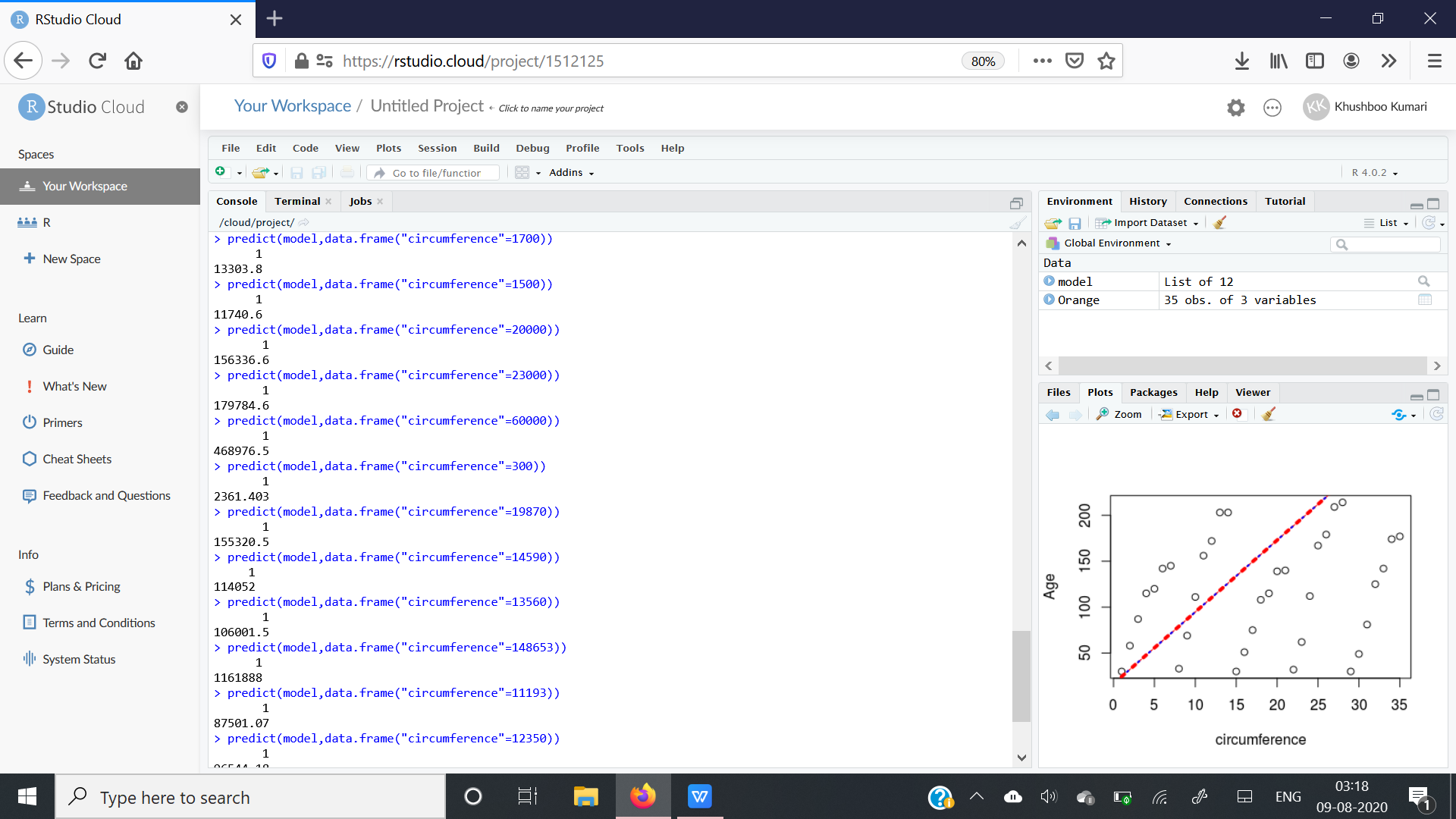
1

96544.18

>plot(Orange$circumference,Orange$Age,xlab='circumference', ylab='Age')

> model <- lm(age~circumference, data=Orange)

> abline(model, col="red",lty=2,lwd=1)



**#PREDICTING AGE :**

> predict(model,data.frame("age"=3400))

1

380.4188

> predict(model,data.frame("age"=4400))

1

487.1891

> predict(model,data.frame("age"=44000))

1

4715.294

> predict(model,data.frame("age"=31358))

1

3365.504

> predict(model,data.frame("age"=123458))

1

13199.05

> predict(model,data.frame("age"=123))

1

30.5324

> predict(model,data.frame("age"=23456))

1

2521.804

> predict(model,data.frame("age"=2446))

1

278.5599

> predict(model,data.frame("age"=89766))

1

9601.745

> predict(model,data.frame("age"=84326))

1

9020.914

> predict(model,data.frame("age"=5678))

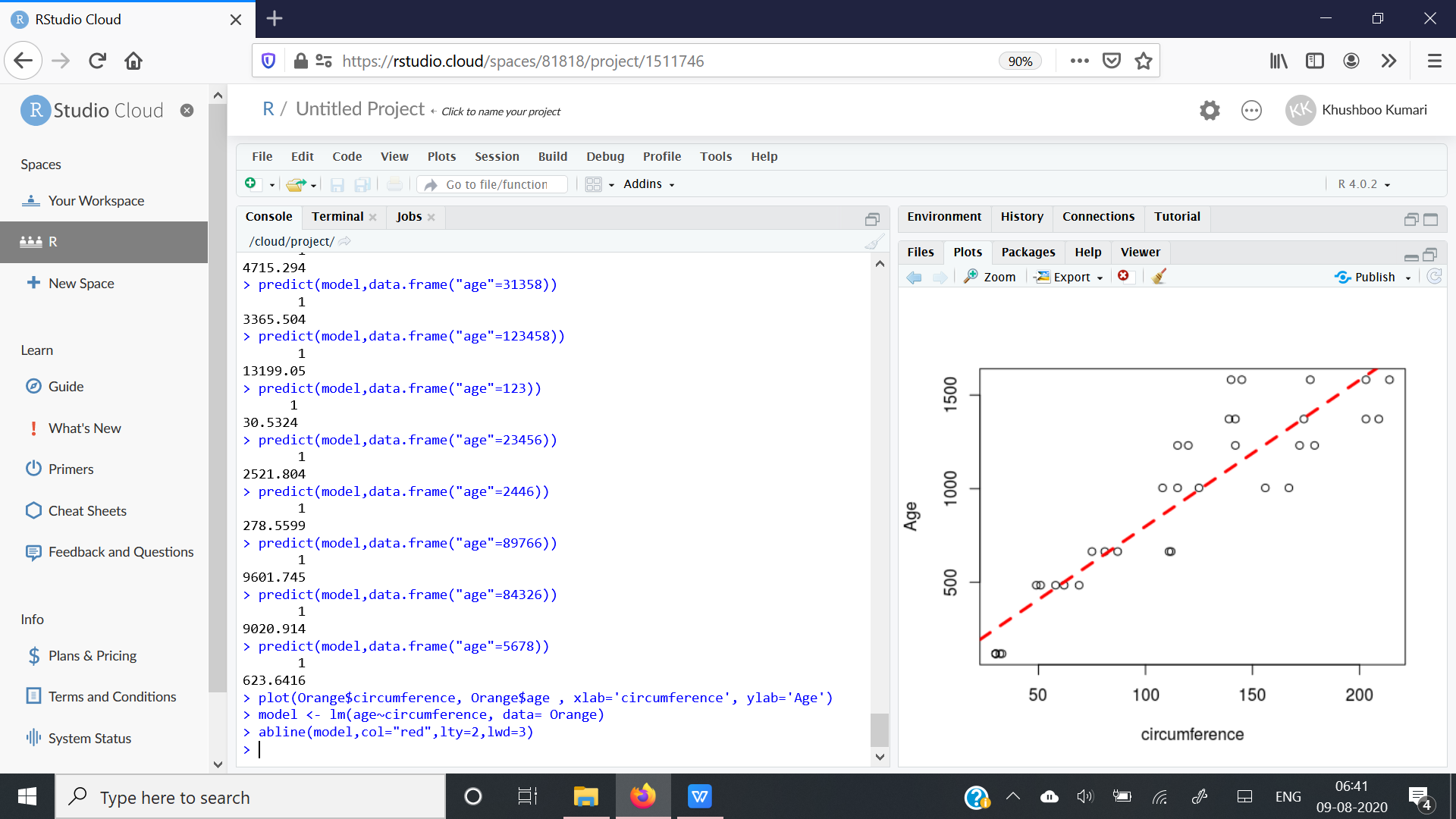
1

623.6416

>plot(Orange$circumference,Orange$age,xlab='circumference', ylab='Age')

> model <- lm(age~circumference, data= Orange)

> abline(model,col="red",lty=2,lwd=3)



**CONCLUSION:**

By using R-lang Regression was done easily in comparison to python and there is no limit for the results we want.here we have conclude with the following above graphs which shows the predicted value between age and circumference.