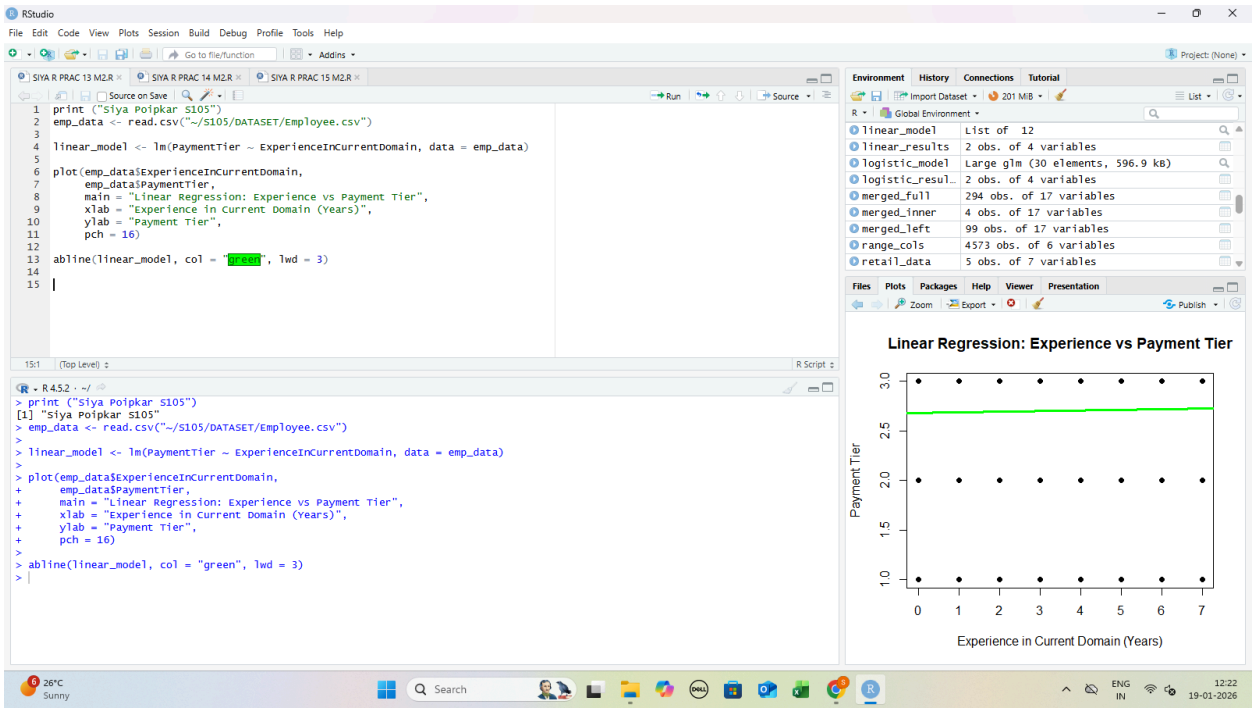
MVLU COLLEGE
PRACTICAL NO. 13 TO 15

AIM:13 Performing linear regression analysis using lm() (R).

```
print ("Siya Poipkar S105")
emp_data <- read.csv("~/S105/DATASET/Employee.csv")

linear_model <- lm(PaymentTier ~ ExperienceInCurrentDomain, data = emp_data)

plot(emp_data$ExperienceInCurrentDomain,
    emp_data$PaymentTier,
    main = "Linear Regression: Experience vs Payment Tier",
    xlab = "Experience in Current Domain (Years)",
    ylab = "Payment Tier",
    pch = 16)

abline(linear_model, col = "green", lwd = 3)
```

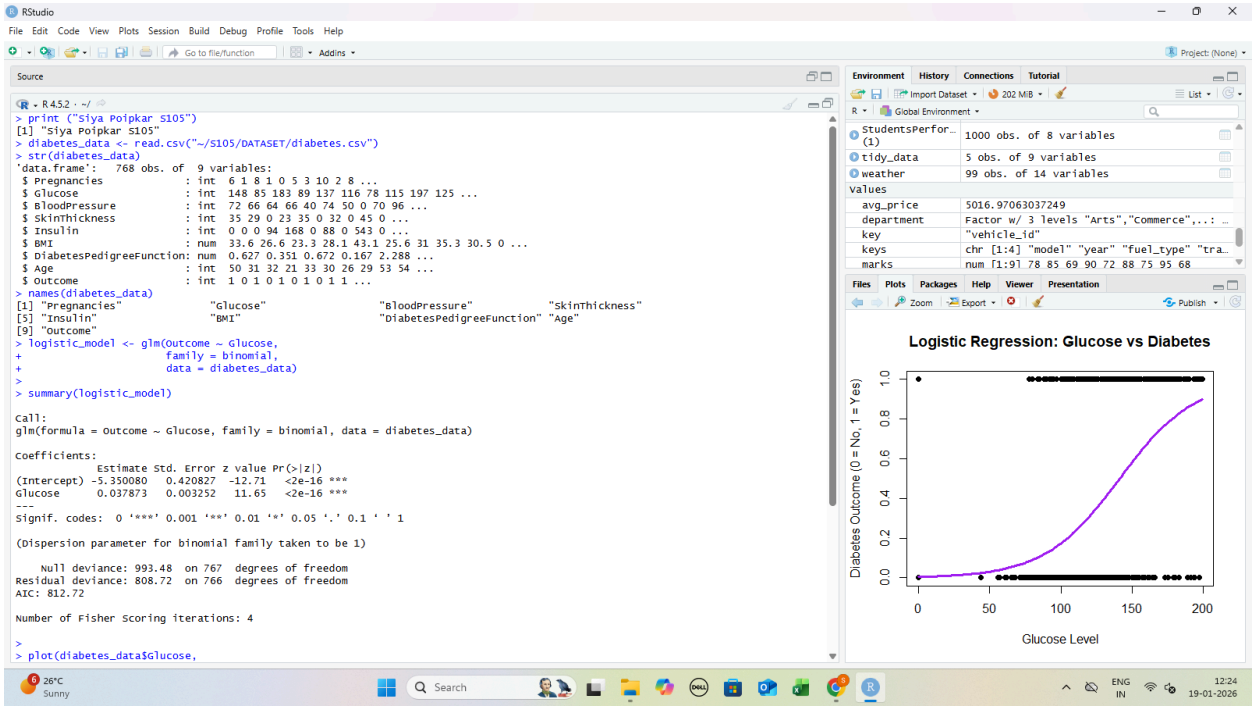


MVLU COLLEGE
PRACTICAL NO. 13 TO 15

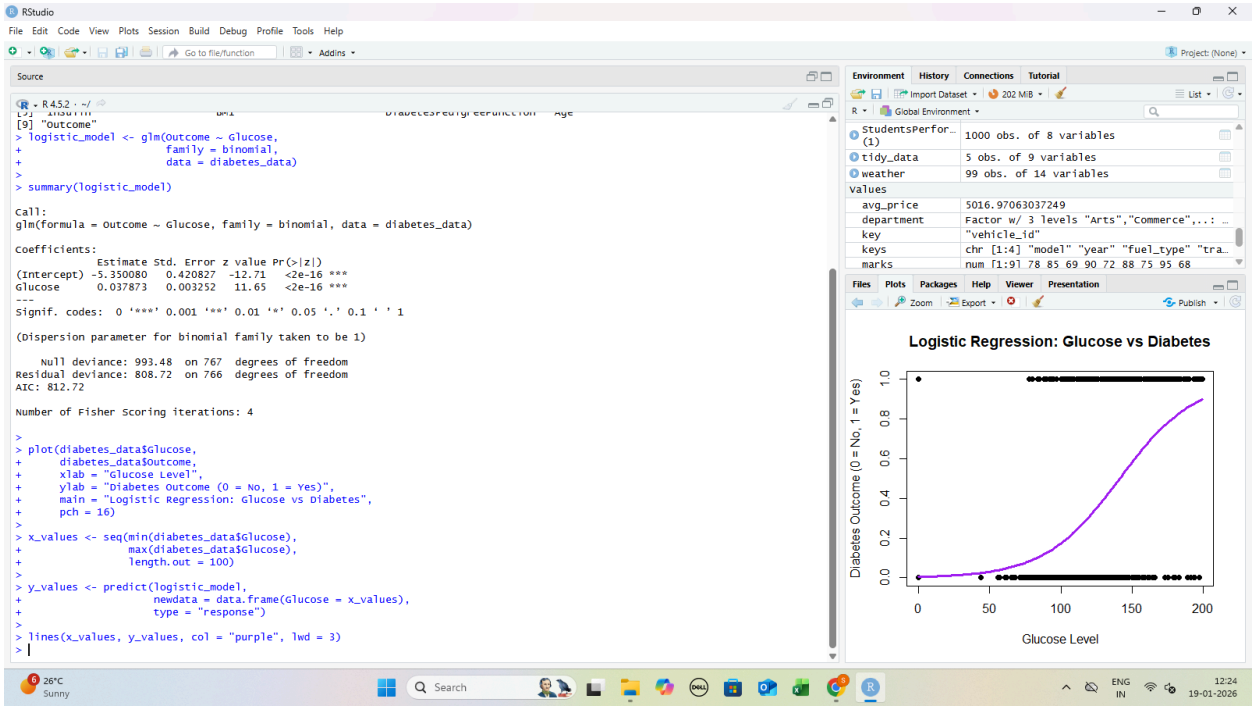
AIM:14 Performing logistic regression using glm() (R).

```
print ("Siya Poipkar S105")
diabetes_data <- read.csv("~/S105/DATASET/diabetes.csv")
str(diabetes_data)
names(diabetes_data)
logistic_model <- glm(Outcome ~ Glucose,
                      family = binomial,
                      data = diabetes_data)

summary(logistic_model)

plot(diabetes_data$Glucose,
    diabetes_data$Outcome,
    xlab = "Glucose Level",
    ylab = "Diabetes Outcome (0 = No, 1 = Yes)",
    main = "Logistic Regression: Glucose vs Diabetes",
    pch = 16)

x_values <- seq(min(diabetes_data$Glucose),
                max(diabetes_data$Glucose),
                length.out = 100)

y_values <- predict(logistic_model,
                    newdata = data.frame(Glucose = x_values),
                    type = "response")

lines(x_values, y_values, col = "purple", lwd = 3)
```

MVLU COLLEGE
PRACTICAL NO. 13 TO 15



```
> print ("Siya Poipkar S105")
[1] "Siya Poipkar S105"
> diabetes_data <- read.csv("~/S105/DATASET/diabetes.csv")
> str(diabetes_data)
'data.frame':   768 obs. of  9 variables:
 $ Pregnancies             : int  6 1 8 1 0 5 3 10 2 8 ...
 $ Glucose                 : int  148 85 183 89 137 116 78 115 197 125 ...
 $ BloodPressure           : int  72 66 64 66 40 74 50 0 70 96 ...
 $ SkinThickness           : int  35 29 0 23 35 0 32 0 45 0 ...
 $ Insulin                 : int  0 0 0 94 168 0 88 0 543 0 ...
 $ BMI                     : num  33.6 26.6 23.3 28.1 43.1 25.6 31 35.3 30.5 0 ...
 $ DiabetesPedigreeFunction: num  0.627 0.351 0.672 0.167 2.288 ...
 $ Age                     : int  50 31 32 21 33 30 26 29 53 54 ...
 $ Outcome                 : int  1 0 1 0 1 0 1 0 1 1 ...
> names(diabetes_data)
[1] "Pregnancies"              "Glucose"                  "BloodPressure"            "SkinThickness"
[5] "Insulin"                  "BMI"                      "DiabetesPedigreeFunction" "Age"
[9] "Outcome"
> logistic_model <- glm(Outcome ~ Glucose,
+                       family = binomial,
+                       data = diabetes_data)
>
> summary(logistic_model)

Call:
glm(formula = Outcome ~ Glucose, family = binomial, data = diabetes_data)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -5.350080   0.420827  -12.71   <2e-16 ***
Glucose      0.037873   0.003252   11.65   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 993.48  on 767  degrees of freedom
Residual deviance: 808.72  on 766  degrees of freedom
AIC: 812.72

Number of Fisher Scoring iterations: 4

>
> plot(diabetes_data$Glucose,
```

```
> logistic_model <- glm(Outcome ~ Glucose,
+                       family = binomial,
+                       data = diabetes_data)
>
> summary(logistic_model)

Call:
glm(formula = Outcome ~ Glucose, family = binomial, data = diabetes_data)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -5.350080   0.420827  -12.71   <2e-16 ***
Glucose      0.037873   0.003252   11.65   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 993.48  on 767  degrees of freedom
Residual deviance: 808.72  on 766  degrees of freedom
AIC: 812.72

Number of Fisher Scoring iterations: 4

>
> plot(diabetes_data$Glucose,
+      diabetes_data$Outcome,
+      xlab = "Glucose Level",
+      ylab = "Diabetes Outcome (0 = No, 1 = Yes)",
+      main = "Logistic Regression: Glucose vs Diabetes",
+      pch = 16)
>
> x_values <- seq(min(diabetes_data$Glucose),
+                 max(diabetes_data$Glucose),
+                 length.out = 100)
>
> y_values <- predict(logistic_model,
+                     newdata = data.frame(Glucose = x_values),
+                     type = "response")
>
> lines(x_values, y_values, col = "purple", lwd = 3)
>
```

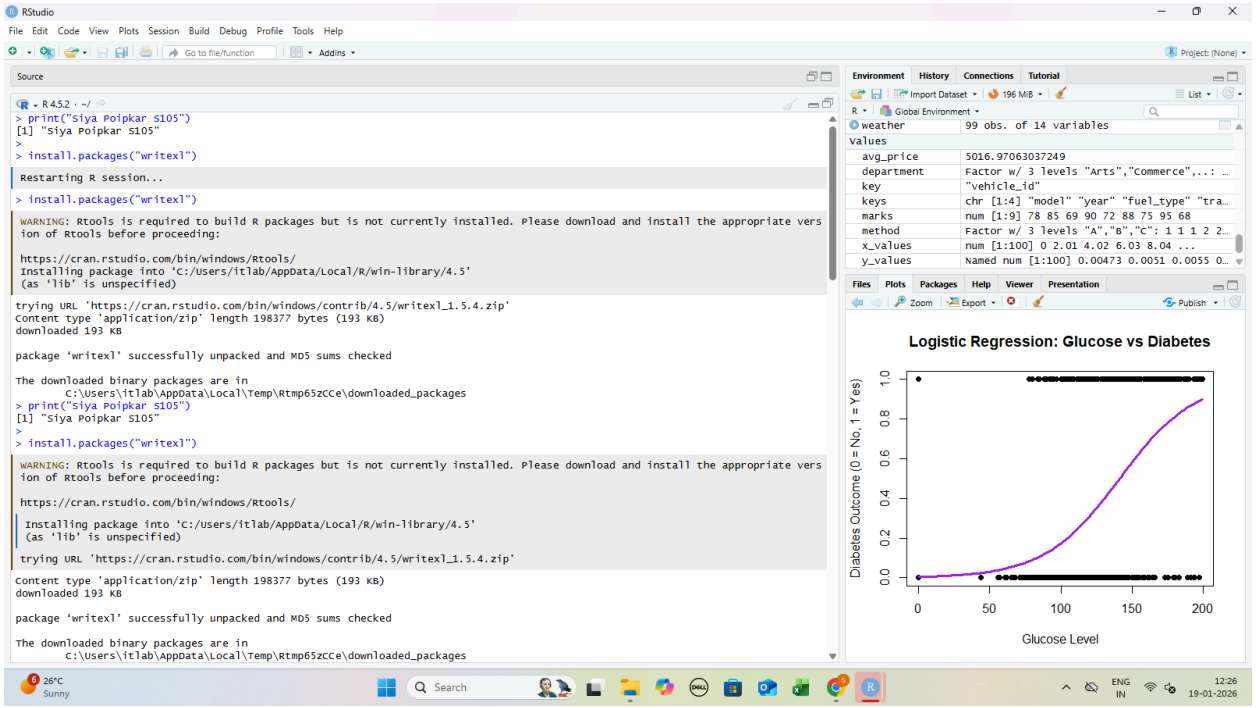
MVLU COLLEGE
PRACTICAL NO. 13 TO 15

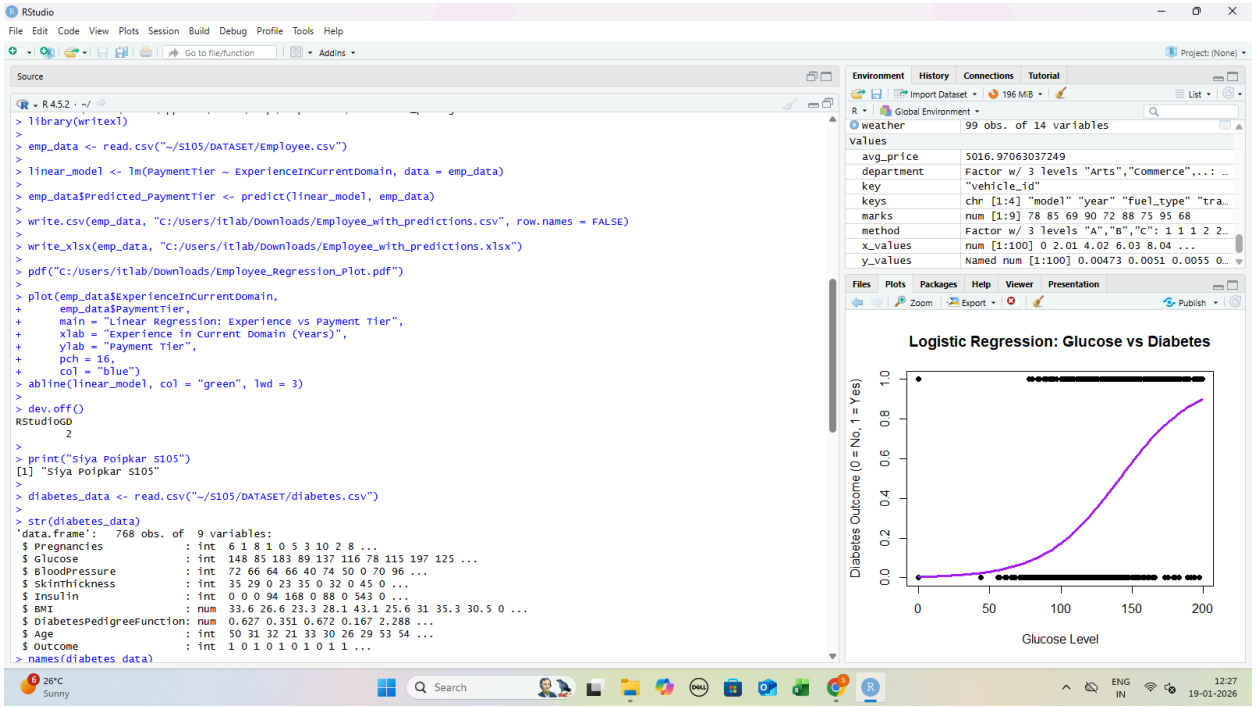
AIM:15 Exporting results into external files (Excel, CSV, PDF) using write.csv() and writexl (R).

```
print("Siya Poipkar S105")

install.packages("writexl")
library(writexl)

emp_data <- read.csv("~/S105/DATASET/Employee.csv")

linear_model <- lm(PaymentTier ~ ExperienceInCurrentDomain, data = emp_data)

emp_data$Predicted_PaymentTier <- predict(linear_model, emp_data)

write.csv(emp_data, "C:/Users/itlab/Downloads/Employee_with_predictions.csv", row.names =
FALSE)

write_xlsx(emp_data, "C:/Users/itlab/Downloads/Employee_with_predictions.xlsx")

pdf("C:/Users/itlab/Downloads/Employee_Regression_Plot.pdf")

plot(emp_data$ExperienceInCurrentDomain,
    emp_data$PaymentTier,
    main = "Linear Regression: Experience vs Payment Tier",
    xlab = "Experience in Current Domain (Years)",
    ylab = "Payment Tier",
    pch = 16,
    col = "blue")
abline(linear_model, col = "green", lwd = 3)

dev.off()

print("Siya Poipkar S105")

diabetes_data <- read.csv("~/S105/DATASET/diabetes.csv")

str(diabetes_data)
names(diabetes_data)
logistic_model <- glm(Outcome ~ Glucose,
                      family = binomial,
                      data = diabetes_data)

summary(logistic_model)

diabetes_data$Predicted_Prob <- predict(logistic_model,
```

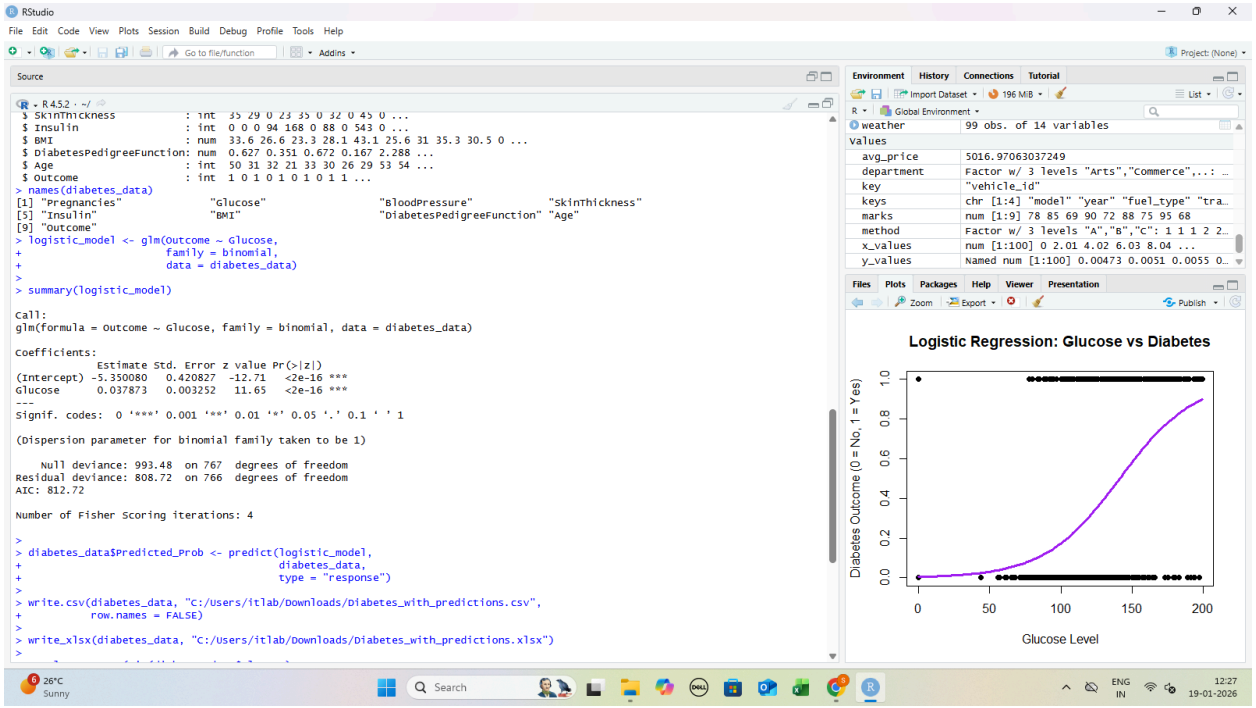
```
                        diabetes_data,
                        type = "response")

write.csv(diabetes_data, "C:/Users/itlab/Downloads/Diabetes_with_predictions.csv",
        row.names = FALSE)

write_xlsx(diabetes_data, "C:/Users/itlab/Downloads/Diabetes_with_predictions.xlsx")

x_values <- seq(min(diabetes_data$Glucose),
              max(diabetes_data$Glucose),
              length.out = 100)

y_values <- predict(logistic_model,
                  newdata = data.frame(Glucose = x_values),
                  type = "response")

pdf("C:/Users/itlab/Downloads/Diabetes_Logistic_Regression_Plot.pdf")

plot(diabetes_data$Glucose,
    diabetes_data$Outcome,
    xlab = "Glucose Level",
    ylab = "Diabetes Outcome (0 = No, 1 = Yes)",
    main = "Logistic Regression: Glucose vs Diabetes",
    pch = 16,
    col = "blue")
lines(x_values, y_values, col = "purple", lwd = 3)

abline(h = 0.5, col = "red", lty = 2)
dev.off()
```

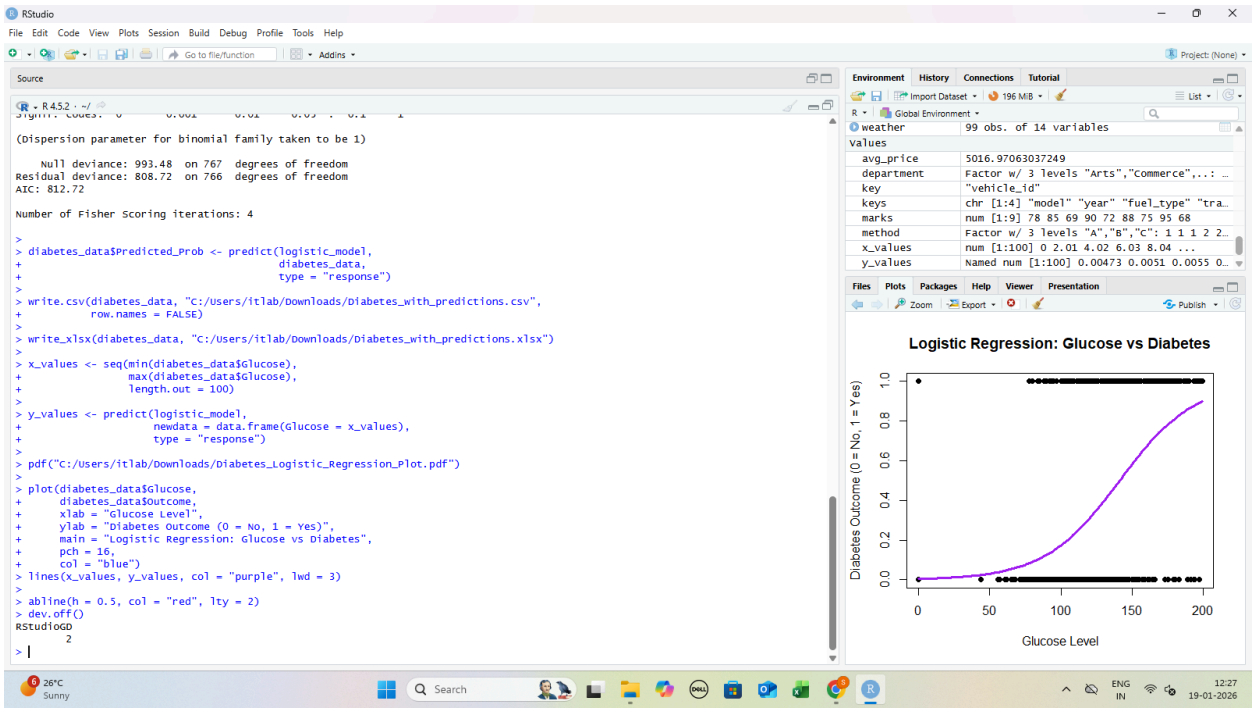
MVLU COLLEGE
PRACTICAL NO. 13 TO 15



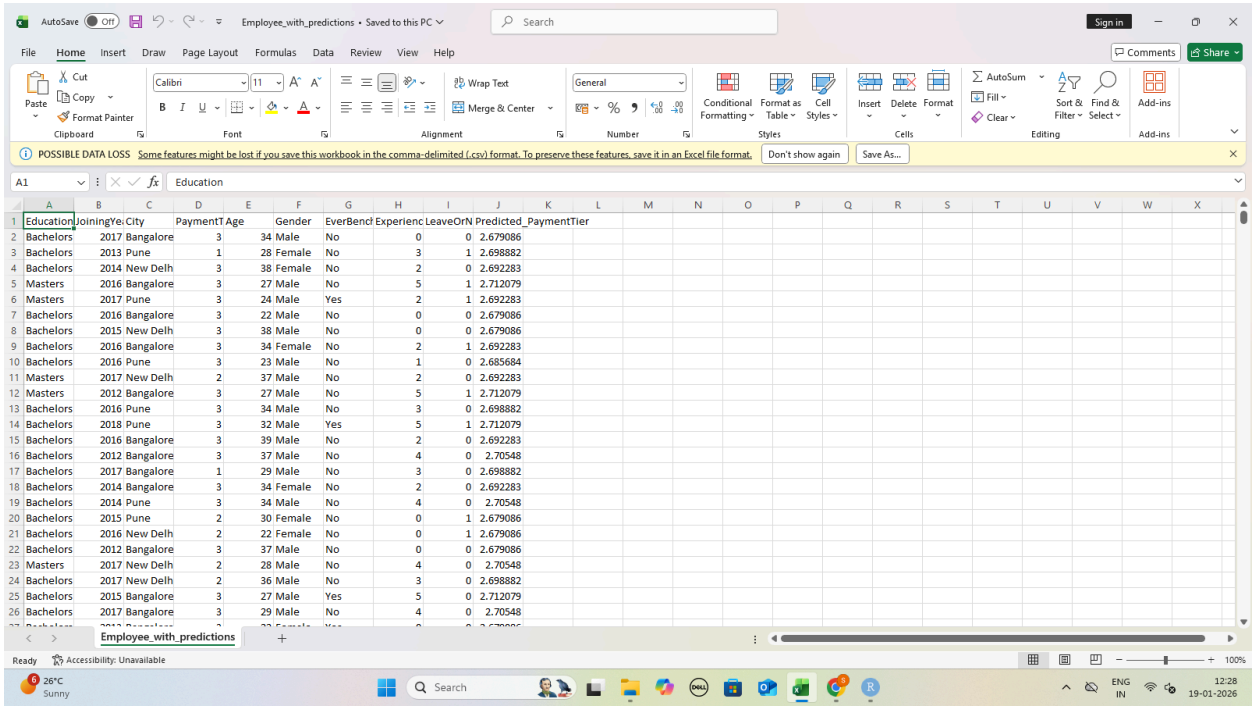


MVLU COLLEGE
PRACTICAL NO. 13 TO 15

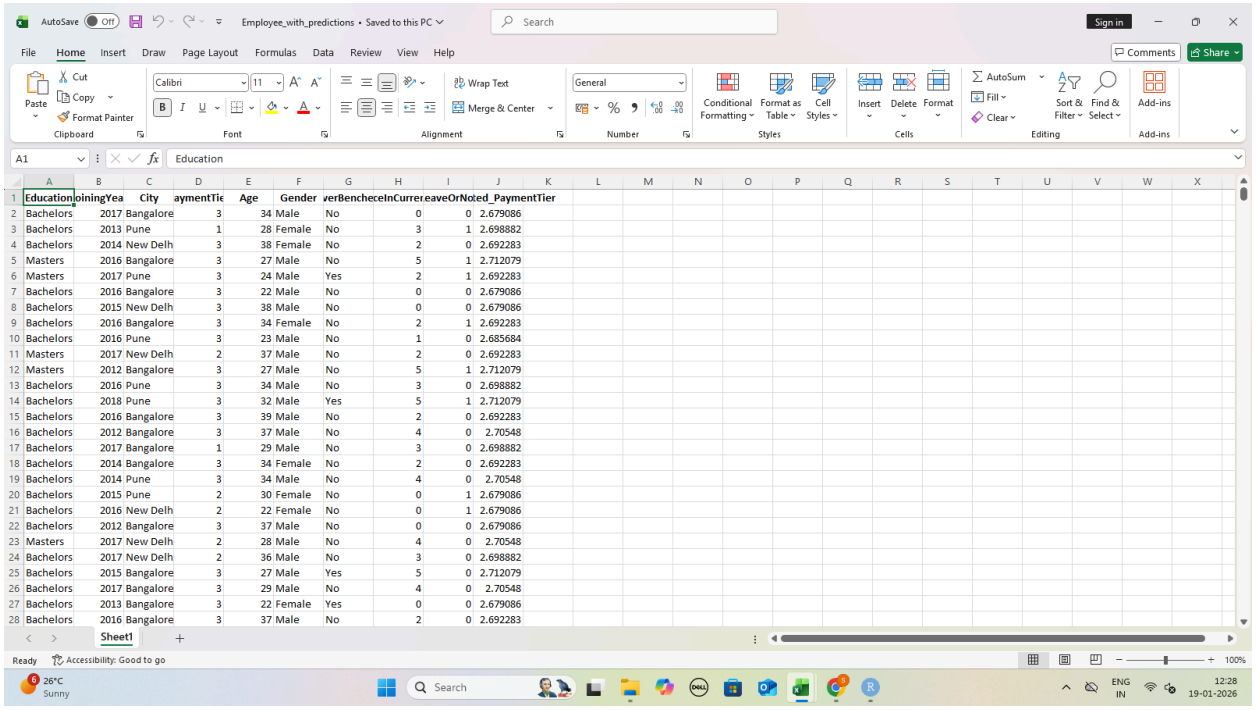




MVLU COLLEGE
PRACTICAL NO. 13 TO 15



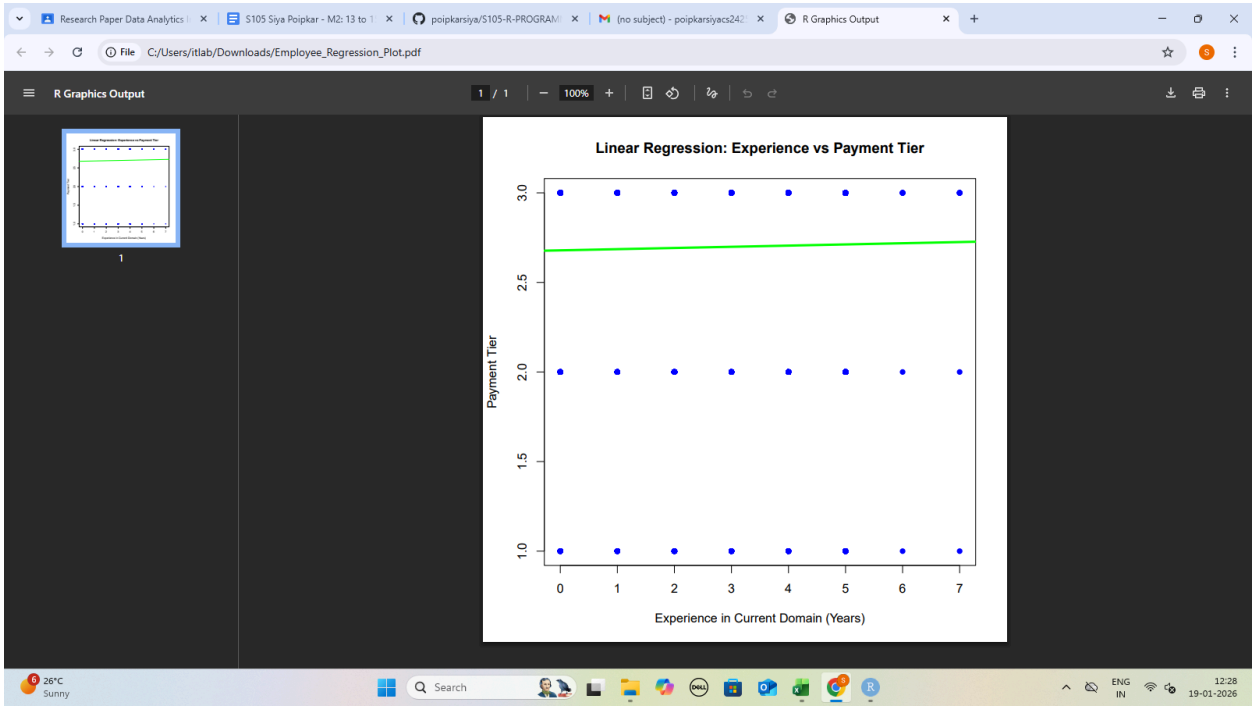
Education	JoiningYe	City	PaymentT	Age	Gender	EverBench	Experienc	LeaveOrN	Predicted_PaymentTier
Bachelors	2017	Bangalore	3	34	Male	No	0	0	2.679086
Bachelors	2013	Pune	1	28	Female	No	3	1	2.698882
Bachelors	2014	New Delh	3	38	Female	No	2	0	2.692283
Masters	2016	Bangalore	3	27	Male	No	5	1	2.712079
Masters	2017	Pune	3	24	Male	Yes	2	1	2.692283
Bachelors	2016	Bangalore	3	22	Male	No	0	0	2.679086
Bachelors	2015	New Delh	3	38	Male	No	0	0	2.679086
Bachelors	2016	Bangalore	3	34	Female	No	2	1	2.692283
Bachelors	2016	Pune	3	23	Male	No	1	0	2.685684
Masters	2017	New Delh	2	37	Male	No	2	0	2.692283
Masters	2012	Bangalore	3	27	Male	No	5	1	2.712079
Bachelors	2016	Pune	3	34	Male	No	3	0	2.698882
Bachelors	2018	Pune	3	32	Male	Yes	5	1	2.712079
Bachelors	2016	Bangalore	3	39	Male	No	2	0	2.692283
Bachelors	2012	Bangalore	3	37	Male	No	4	0	2.70548
Bachelors	2017	Bangalore	1	29	Male	No	3	0	2.698882
Bachelors	2014	Bangalore	3	34	Female	No	2	0	2.692283
Bachelors	2014	Pune	3	34	Male	No	4	0	2.70548
Bachelors	2015	Pune	2	30	Female	No	0	1	2.679086
Bachelors	2016	New Delh	2	22	Female	No	0	1	2.679086
Bachelors	2012	Bangalore	3	37	Male	No	0	0	2.679086
Masters	2017	New Delh	2	28	Male	No	4	0	2.70548
Bachelors	2017	New Delh	2	36	Male	No	3	0	2.698882
Bachelors	2015	Bangalore	3	27	Male	Yes	5	0	2.712079
Bachelors	2017	Bangalore	3	29	Male	No	4	0	2.70548

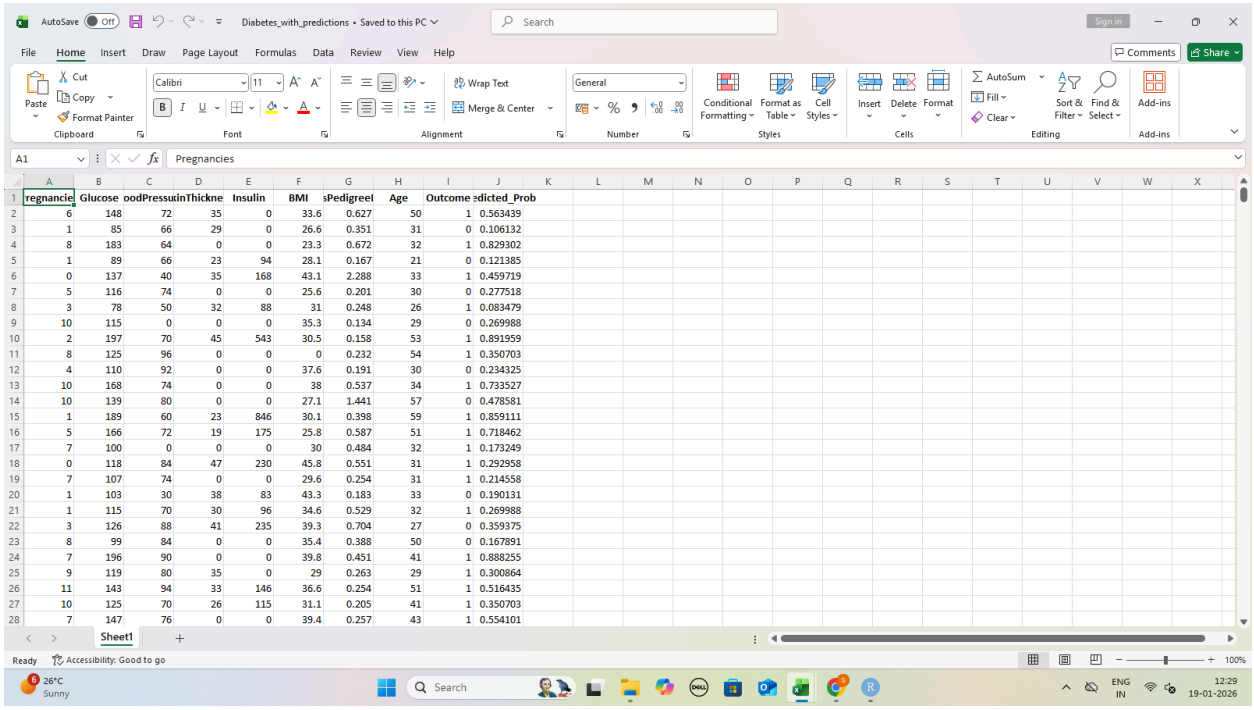


Education	JoiningYea	City	aymentTie	Age	Gender	verBenche	eInCurre	eaveOrN	ted_PaymentTier
Bachelors	2017	Bangalore	3	34	Male	No	0	0	2.679086
Bachelors	2013	Pune	1	28	Female	No	3	1	2.698882
Bachelors	2014	New Delh	3	38	Female	No	2	0	2.692283
Masters	2016	Bangalore	3	27	Male	No	5	1	2.712079
Masters	2017	Pune	3	24	Male	Yes	2	1	2.692283
Bachelors	2016	Bangalore	3	22	Male	No	0	0	2.679086
Bachelors	2015	New Delh	3	38	Male	No	0	0	2.679086
Bachelors	2016	Bangalore	3	34	Female	No	2	1	2.692283
Bachelors	2016	Pune	3	23	Male	No	1	0	2.685684
Masters	2017	New Delh	2	37	Male	No	2	0	2.692283
Masters	2012	Bangalore	3	27	Male	No	5	1	2.712079
Bachelors	2016	Pune	3	34	Male	No	3	0	2.698882
Bachelors	2018	Pune	3	32	Male	Yes	5	1	2.712079
Bachelors	2016	Bangalore	3	39	Male	No	2	0	2.692283
Bachelors	2012	Bangalore	3	37	Male	No	4	0	2.70548
Bachelors	2017	Bangalore	1	29	Male	No	3	0	2.698882
Bachelors	2014	Bangalore	3	34	Female	No	2	0	2.692283
Bachelors	2014	Pune	3	34	Male	No	4	0	2.70548
Bachelors	2015	Pune	2	30	Female	No	0	1	2.679086
Bachelors	2016	New Delh	2	22	Female	No	0	1	2.679086
Bachelors	2012	Bangalore	3	37	Male	No	0	0	2.679086
Masters	2017	New Delh	2	28	Male	No	4	0	2.70548
Bachelors	2017	New Delh	2	36	Male	No	3	0	2.698882
Bachelors	2015	Bangalore	3	27	Male	Yes	5	0	2.712079
Bachelors	2017	Bangalore	3	29	Male	No	4	0	2.70548
Bachelors	2013	Bangalore	3	22	Female	Yes	0	0	2.679086
Bachelors	2016	Bangalore	3	37	Male	No	2	0	2.692283

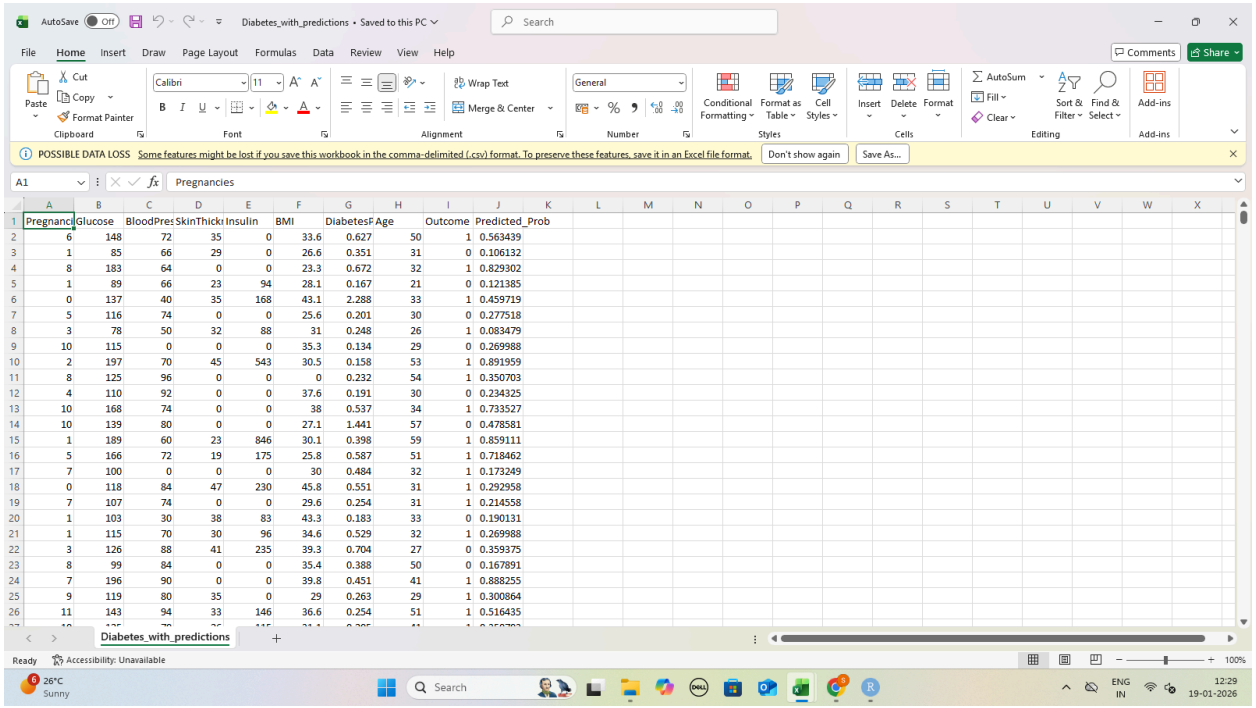
SIYA POIPKAR S105

MVLU COLLEGE
PRACTICAL NO. 13 TO 15





MVLU COLLEGE
PRACTICAL NO. 13 TO 15



Pregnanci	Glucose	BloodPres	SkinThick	Insulin	BMI	DiabetesF	Age	Outcome	Predicted_Prob
6	148	72	35	0	33.6	0.627	50	1	0.563439
1	85	66	29	0	26.6	0.351	31	0	0.106132
8	183	64	0	0	23.3	0.672	32	1	0.829302
1	89	66	23	94	28.1	0.167	21	0	0.121385
0	137	40	35	168	43.1	2.288	33	1	0.459719
5	116	74	0	0	25.6	0.201	30	0	0.277518
3	78	50	32	88	31	0.248	26	1	0.083479
10	115	0	0	0	35.3	0.134	29	0	0.269988
2	197	70	45	543	30.5	0.158	53	1	0.891959
8	125	96	0	0	0	0.232	54	1	0.350703
4	110	92	0	0	37.6	0.191	30	0	0.234325
10	168	74	0	0	38	0.537	34	1	0.733527
10	139	80	0	0	27.1	1.441	57	0	0.478581
1	189	60	23	846	30.1	0.398	59	1	0.859111
5	166	72	19	175	25.8	0.587	51	1	0.718462
7	100	0	0	0	30	0.484	32	1	0.173249
0	118	84	47	230	45.8	0.551	31	1	0.292958
7	107	74	0	0	29.6	0.254	31	1	0.214558
1	103	30	38	83	43.3	0.183	33	0	0.190131
1	115	70	30	96	34.6	0.529	32	1	0.269988
3	126	88	41	235	39.3	0.704	27	0	0.359375
8	99	84	0	0	35.4	0.388	50	0	0.167891
7	196	90	0	0	39.8	0.451	41	1	0.888255
9	119	80	35	0	29	0.263	29	1	0.300864
11	143	94	33	146	36.6	0.254	51	1	0.516435