

Chrome File Edit View History Bookmarks Profiles Tab Window Help

colab.research.google.com/drive/1frTzkwNPV5c8NyX8dxLao-GPZH4uewPI?authuser=0#scrollTo=bq2VJaKCTc8V

Predictive Analytics Project ★

All changes saved

+ Code + Text

RAM Disk Gemini

1. Import the dataset in R and check its dimensions and characteristics

2. Check the class of different variables in the dataset

[1] #We import the dataset file from Google drive
system("gdown --id 1zy0xb12ju8TxbtDP-zg3S8z1IergRno")
df1 = read.csv("heart_disease.csv")

[2] df1

	25.139771	5.836373	11.482937
1	11.658029	26.7341762	17.473940
2	6.295051	19.4351915	17.762115
3	60.587355	8.1735868	4.938059
4	56.151543	28.3524443	10.161202
5	74.221991	26.0782411	4.631398
6	25.443047	13.2003989	12.358035
7	36.368318	23.6979059	11.282946
8	25.887693	8.1590235	10.887413
9	59.350253	20.3678154	7.665346
10	11.883447	1.1427967	12.788451
11	4.817058	27.5778350	18.503664
12	7.582143	29.6442147	18.081268
13	54.181002	5.9596483	5.351619
14	33.282015	6.8121297	10.372329
15	36.086220	13.7865545	9.731545
16	55.663943	14.9632192	6.916259
17	22.719661	1.0247563	10.893894
18	41.178991	7.5752679	7.831385
19	22.021893	19.4992577	13.770639
20	69.927963	25.9843899	6.756015
21	41.642072	0.5330593	5.925680
22	70.194411	16.2928029	4.707009
23	50.831908	6.0260661	5.585493
24	68.884065	10.5429959	3.109338
25	21.599230	7.6043297	12.373691
26	47.660440	27.5624638	11.294392
27	45.097203	21.3856201	9.616762
28	8.279743	6.4237195	13.495168
29	42.345863	20.7413275	10.115865
30	30.774254	23.6101750	11.843556

Chrome File Edit View History Bookmarks Profiles Tab Window Help

colab.research.google.com/drive/1frTzkwNPV5c8NyX8dxLao-GPZH4uewPI?authuser=0#scrollTo=bF7hUCn1UBGm

Predictive Analytics Project ★

All changes saved

+ Code + Text

RAM Disk Gemini

3. Summarize the data and display the head and tail of the data

	7.582143	29.6442147	18.081268
1	54.181002	5.9596483	5.351619
2	33.282015	6.8121297	10.372329
3	36.086220	13.7865545	9.731545
4	55.663943	14.9632192	6.916259
5	22.719661	1.0247563	10.893894
6	41.178991	7.5752679	7.831385
7	22.021893	19.4992577	13.770639
8	69.927963	25.9843899	6.756015
9	41.642072	0.5330593	5.925680
10	70.194411	16.2928029	4.707009
11	50.831908	6.0260661	5.585493
12	68.884065	10.5429959	3.109338
13	21.599230	7.6043297	12.373691
14	47.660440	27.5624638	11.294392
15	45.097203	21.3856201	9.616762
16	8.279743	6.4237195	13.495168
17	42.345863	20.7413275	10.115865
18	30.774254	23.6101750	11.843556

Chrome File Edit View History Bookmarks Profiles Tab Window Help Mon 27 May 9:10 PM

colab.research.google.com/drive/1frTzkwNPV5c8NyX8dxLao-GPZH4uewPI?authuser=0#scrollTo=bF7hUCn1UBGm

Predictive Analytics Project ★

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

3. Summarize the data and display the head and tail of the data

[3] summary(df1)

```
cycling      smoking      heart_diseases
Min. : 1.119  Min. : 0.5259  Min. : 0.5519
1st Qu.:20.205 1st Qu.: 8.2798  1st Qu.: 6.5137
Median :35.824  Median :15.8146  Median :10.3853
Mean  :37.788  Mean  :15.4350  Mean  :10.1745
3rd Qu.:57.853 3rd Qu.:22.5689  3rd Qu.:13.7240
Max.  :74.907  Max.  :29.9467  Max.  :20.4535
```

[4] head(df1)

```
A data.frame: 6 × 3
#> #>   cycling   smoking   heart_diseases
#> #>   <dbl>     <dbl>       <dbl>
#> #> 1 30.801246 10.896608 11.769423
#> #> 2 65.129215 2.219563  2.854081
#> #> 3 1.959665 17.588331 17.177803
#> #> 4 44.800196 2.802559  6.816647
#> #> 5 69.428454 15.974505 4.062224
#> #> 6 54.403626 29.333178 9.550046
```

[5] tail(df1)

```
A data.frame: 6 × 3
```

Chrome File Edit View History Bookmarks Profiles Tab Window Help Mon 27 May 9:10 PM

colab.research.google.com/drive/1frTzkwNPV5c8NyX8dxLao-GPZH4uewPI?authuser=0#scrollTo=bF7hUCn1UBGm

Predictive Analytics Project ★

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

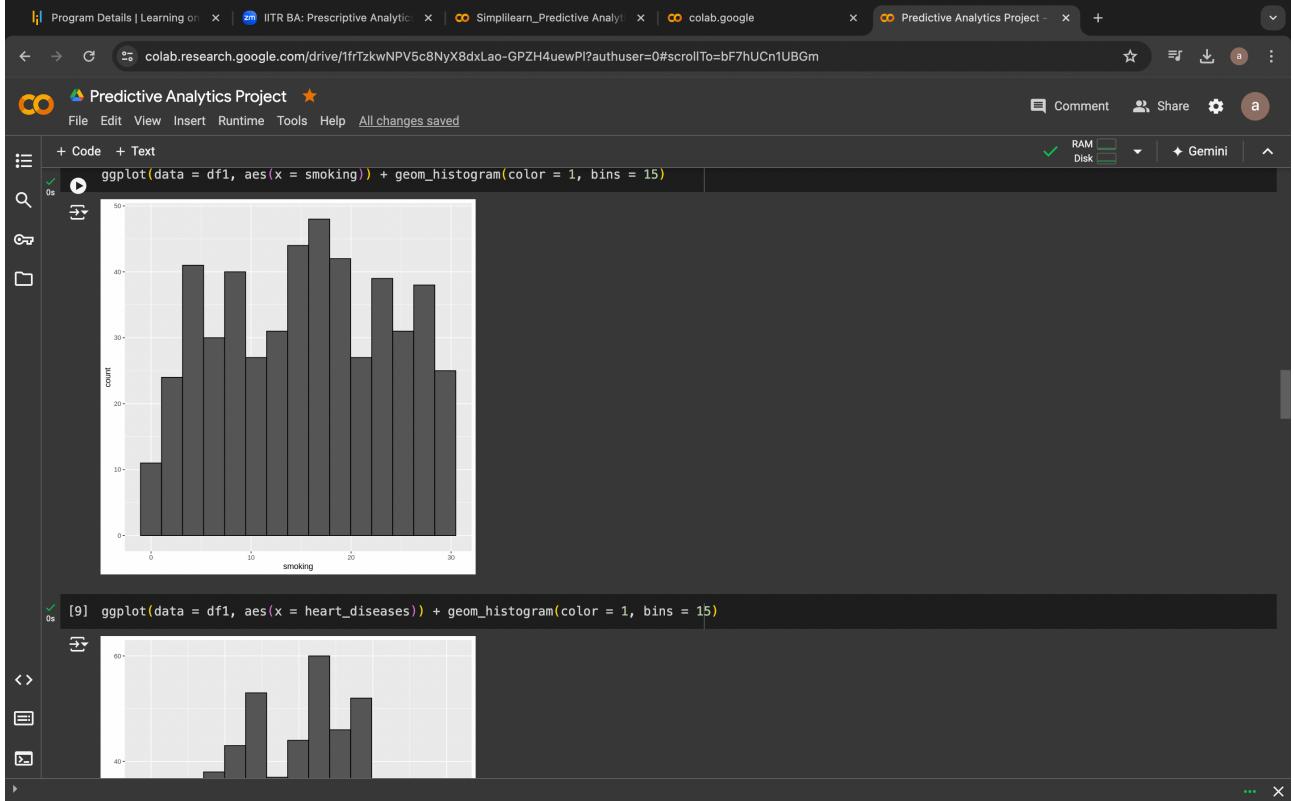
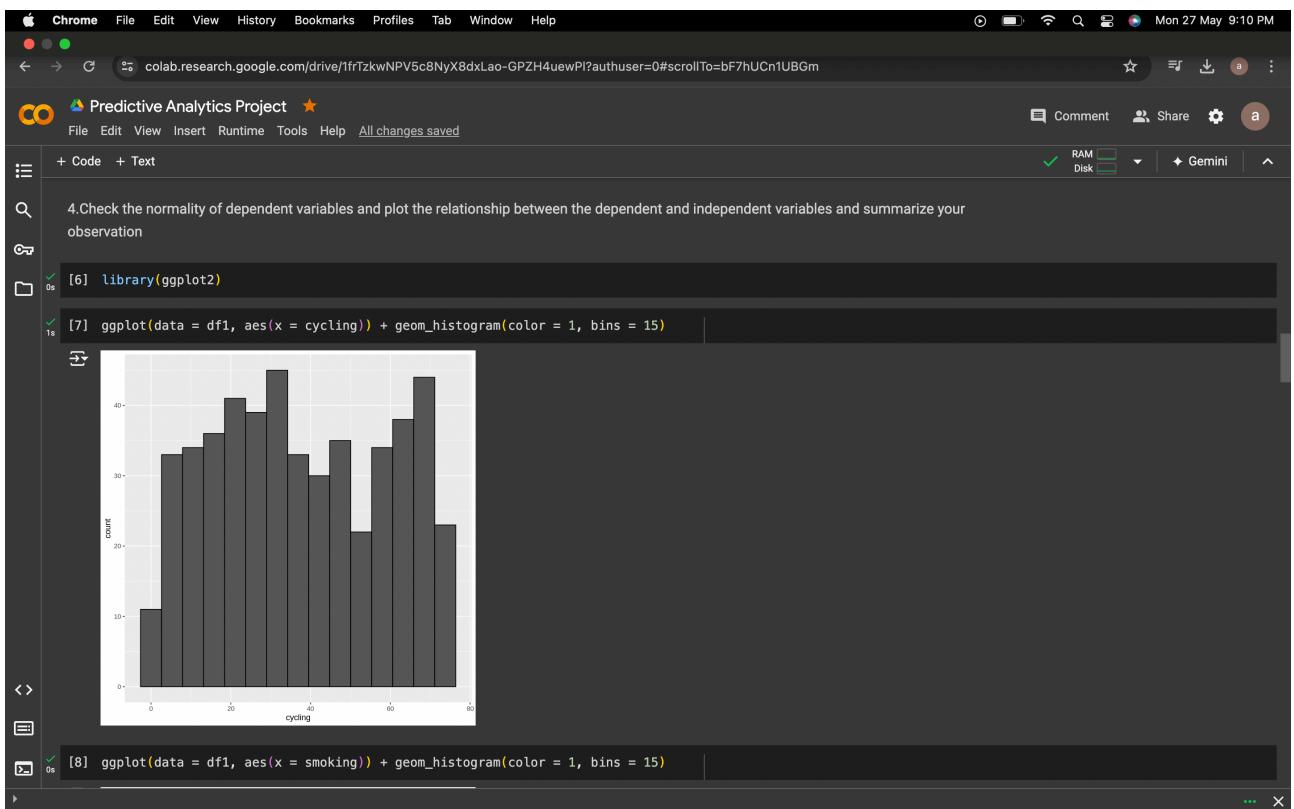
[5] tail(df1)

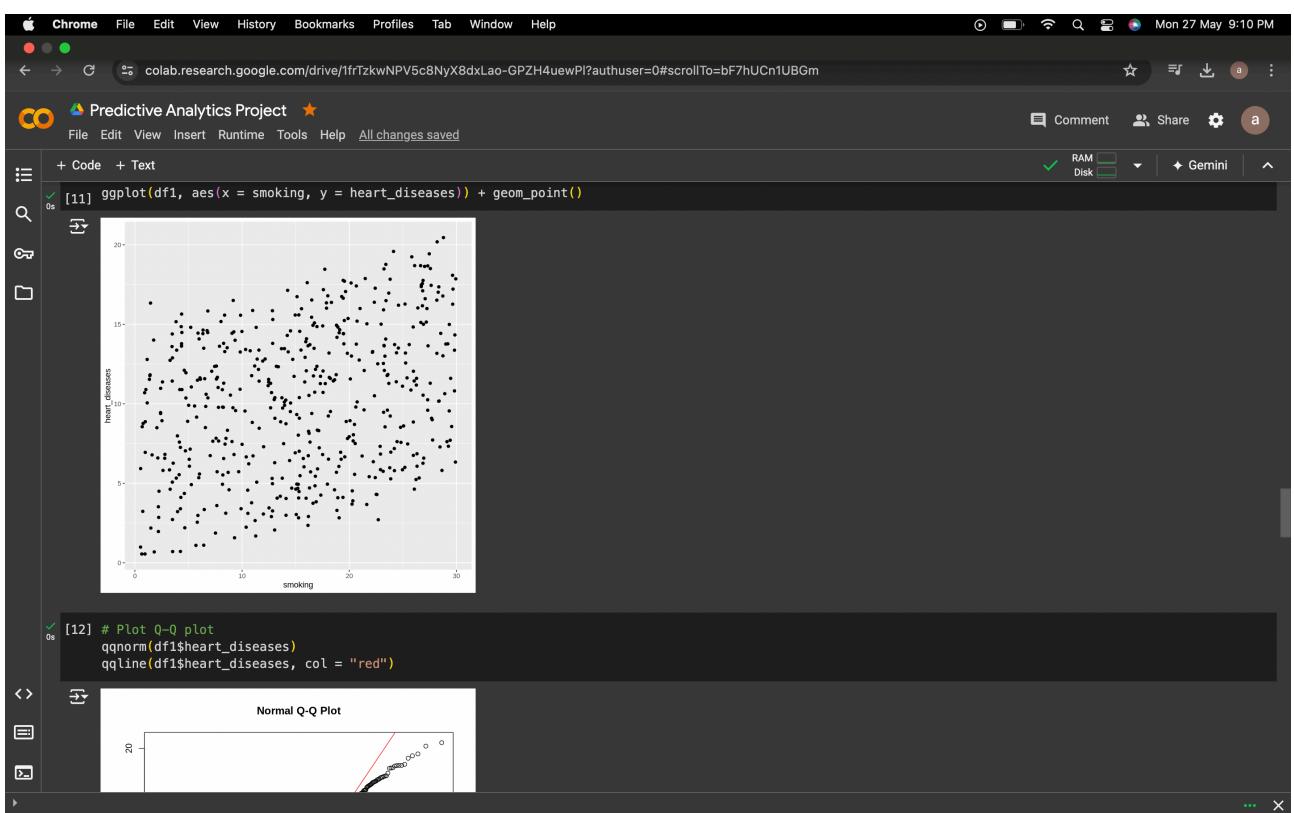
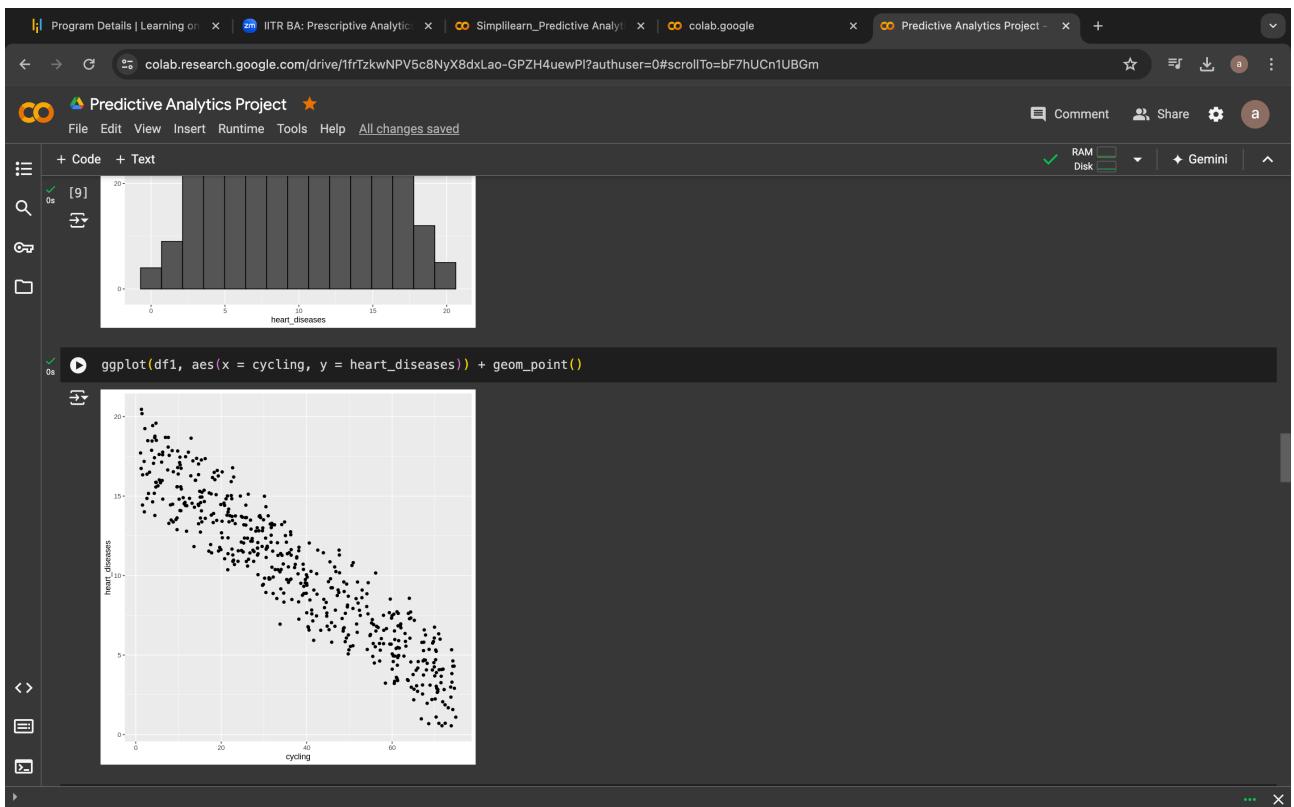
```
A data.frame: 6 × 3
#> #>   cycling   smoking   heart_diseases
#> #>   <dbl>     <dbl>       <dbl>
#> #> 493 21.599230 7.60433 12.373691
#> #> 494 47.660440 27.56246 11.294392
#> #> 495 45.097203 21.38562 9.616762
#> #> 496 8.279743  6.42372 13.495168
#> #> 497 42.345863 20.74133 10.115865
#> #> 498 30.774254 23.61017 11.843556
```

4. Check the normality of dependent variables and plot the relationship between the dependent and independent variables and summarize your observation

[6] library(ggplot2)

[7] ggplot(data = df1, aes(x = cycling)) + geom_histogram(color = 1, bins = 15)





Chrome File Edit View History Bookmarks Profiles Tab Window Help

colab.research.google.com/drive/1frTzkwNPV5c8NyX8dxLao-GPZH4uewPl?authuser=0#scrollTo=bF7hUCn1UBGm

Predictive Analytics Project ★

All changes saved

```
[12] # Plot Q-Q plot
qqnorm(df1$heart_diseases)
qqline(df1$heart_diseases, col = "red")
```

5.Create and run the regression model for the dataset and summarize the observation
6.Store the output of the regression model and print its coefficients
7.Print the estimate, Std. Error, t-value, and p-value for the independent variables (i.e., cycling and smoking)
8.Display residual standard error, r-squared, adjusted r-squared, f-statistic, and p-value from the output

Chrome File Edit View History Bookmarks Profiles Tab Window Help

colab.research.google.com/drive/1frTzkwNPV5c8NyX8dxLao-GPZH4uewPl?authuser=0#scrollTo=bF7hUCn1UBGm

Predictive Analytics Project ★

All changes saved

```
[13] # Create the linear regression model
model <- lm(heart_diseases ~ cycling + smoking, data = df1)

[14] # Summarize the model
summary(model)
```

Call:
`lm(formula = heart_diseases ~ cycling + smoking, data = df1)`

Residuals:

	Min	1Q	Median	3Q	Max
	-2.1789	-0.4463	0.0362	0.4422	1.9331

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	14.984658	0.080137	186.99	<2e-16 ***
cycling	-0.200133	0.001366	-146.53	<2e-16 ***
smoking	0.178334	0.003539	50.39	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.654 on 495 degrees of freedom
Multiple R-squared: 0.9796, Adjusted R-squared: 0.9795
F-statistic: 1.19e+04 on 2 and 495 DF, p-value: < 2.2e-16

```
[15] # Compute the confidence intervals for the model coefficients
confint(model)
```

A matrix: 3 x 2 of type dbl

	2.5 %	97.5 %
(Intercept)	14.8272075	15.1421084
cycling	-0.2028166	-0.1974495
smoking	0.1742890	0.1852878

Chrome File Edit View History Bookmarks Profiles Tab Window Help Mon 27 May 9:11 PM

colab.research.google.com/drive/1frTzkwNPV5c8NyX8dxLao-GPZH4uewPI?authuser=0#scrollTo=bF7hUCn1UBGm

Predictive Analytics Project ★

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

[15] smoking 0.1713800 0.1852878

10.Create the diagnostic plots and point out your observation for it

[16] plot(model, which = 1)

[17] plot(model, which = 2)

Chrome File Edit View History Bookmarks Profiles Tab Window Help Mon 27 May 9:11 PM

colab.research.google.com/drive/1frTzkwNPV5c8NyX8dxLao-GPZH4uewPI?authuser=0#scrollTo=bF7hUCn1UBGm

Predictive Analytics Project ★

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

[17] plot(model, which = 2)

11.Check autocorrelation and heteroscedasticity using an appropriate statistical test

[18] install.packages('lmtest')

Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)

also installing the dependency 'zoo'

Chrome File Edit View History Bookmarks Profiles Tab Window Help Mon 27 May 9:11 PM

colab.research.google.com/drive/1frTzkwNPV5c8NyX8dxLao-GPZH4uewPI?authuser=0#scrollTo=bF7hUCn1UBGm

Predictive Analytics Project ★

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text Comment Share Gemini

RAM Disk

11.Check autocorrelation and heteroscedasticity using an appropriate statistical test

[18] `install.packages('lmtest')`

Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)

also installing the dependency 'zoo'

[19] `library(lmtest)`

Loading required package: zoo

Attaching package: 'zoo'

The following objects are masked from 'package:base':
as.Date, as.Date.numeric

[20] `# Perform Durbin-Watson test for autocorrelation
dwtest(model)`

Durbin-Watson test

data: model
DW = 1.9174, p-value = 0.1773
alternative hypothesis: true autocorrelation is greater than 0

Chrome File Edit View History Bookmarks Profiles Tab Window Help Mon 27 May 9:11 PM

colab.research.google.com/drive/1frTzkwNPV5c8NyX8dxLao-GPZH4uewPI?authuser=0#scrollTo=bF7hUCn1UBGm

Predictive Analytics Project ★

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text Comment Share Gemini

RAM Disk

[20] `DW = 1.9174, p-value = 0.1773
alternative hypothesis: true autocorrelation is greater than 0`

[21] `# Perform Breusch-Pagan test for heteroscedasticity
bpptest(model)`

studentized Breusch-Pagan test

data: model
BP = 5.7775, df = 2, p-value = 0.05564

Colab paid products - Cancel contracts here