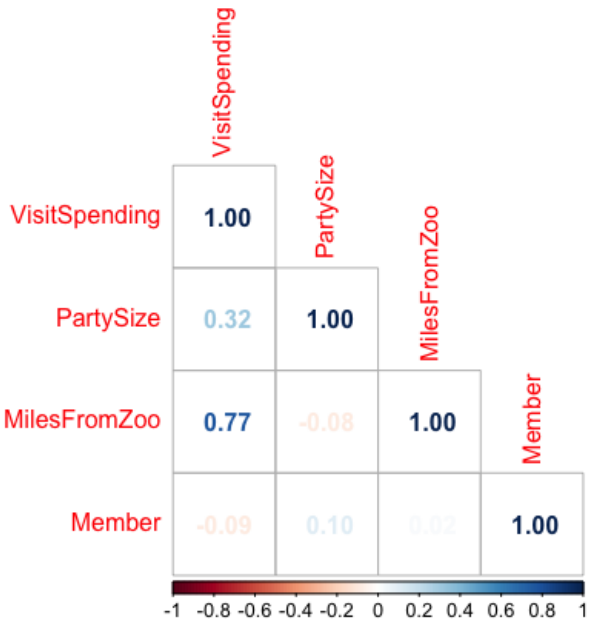
R script Code:

```
# Aadhithya Dinesh
# MIS 545 Section 02
# Lab05DineshA.R
# Import a dataset of spending by groups of people visiting a zoo and generate
# a multiple linear regression model by assigning data types, building a model
# and testing for model fit.

# install.packages("tidyverse")
# install.packages("dummies")
# install.packages("corrplot")
# install.packages("olsrr")

library(tidyverse)
library("corrplot")
library("olsrr")

# set the working directory
setwd("~/MIS/Classes/MIS545/Assignments/Lab05")

# read the csv file with column types specified
zooSpending <- read_csv(file = "ZooVisitSpending.csv",
                        col_types = "niil",
                        col_names = TRUE)

# print the zooSpending tibble
print(zooSpending)

#print the structure of zooSpending
str(zooSpending)

#print the summary of zooSpending
print(summary(zooSpending))

# define the function to display all histograms
displayAllHistograms <- function(tibbleDataset) {
  tibbleDataset %>%
    keep(is.numeric) %>%
    gather() %>%
    ggplot() + geom_histogram(mapping = aes(x=value, fill=key),
                              color = "black") +
    facet_wrap (~ key, scales = "free") +
    theme_minimal()
```

```
}

# call the function
displayAllHistograms(zooSpending)
# rounding the correlation to 2 decimal places
print(round(cor(zooSpending  %>% keep(is.numeric)),2))

# displaying the correlation plot using number method
corrplot(cor(zooSpending),
        method = "number",
        type = "lower")

# generating the linear regression model by using all the independent variables
zooSpendingModel <- lm(data = zooSpending,
                       formula = VisitSpending ~ .)

# displaying the beta coefficients for the model
print(zooSpendingModel)

# displaying the linear regression model results using the summary function
print(summary(zooSpendingModel))

# testing for multicollinearity
print(ols_vif_tol(zooSpendingModel))
```

Correlation Plot:



Model Sumary:

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	0.22141	6.49061	0.034	0.97284	
PartySize	9.13619	1.01756	8.979	4.35e-15	***
MilesFromZoo	0.88886	0.04865	18.272	< 2e-16	***
MemberTRUE	-14.90735	4.58300	-3.253	0.00148	**

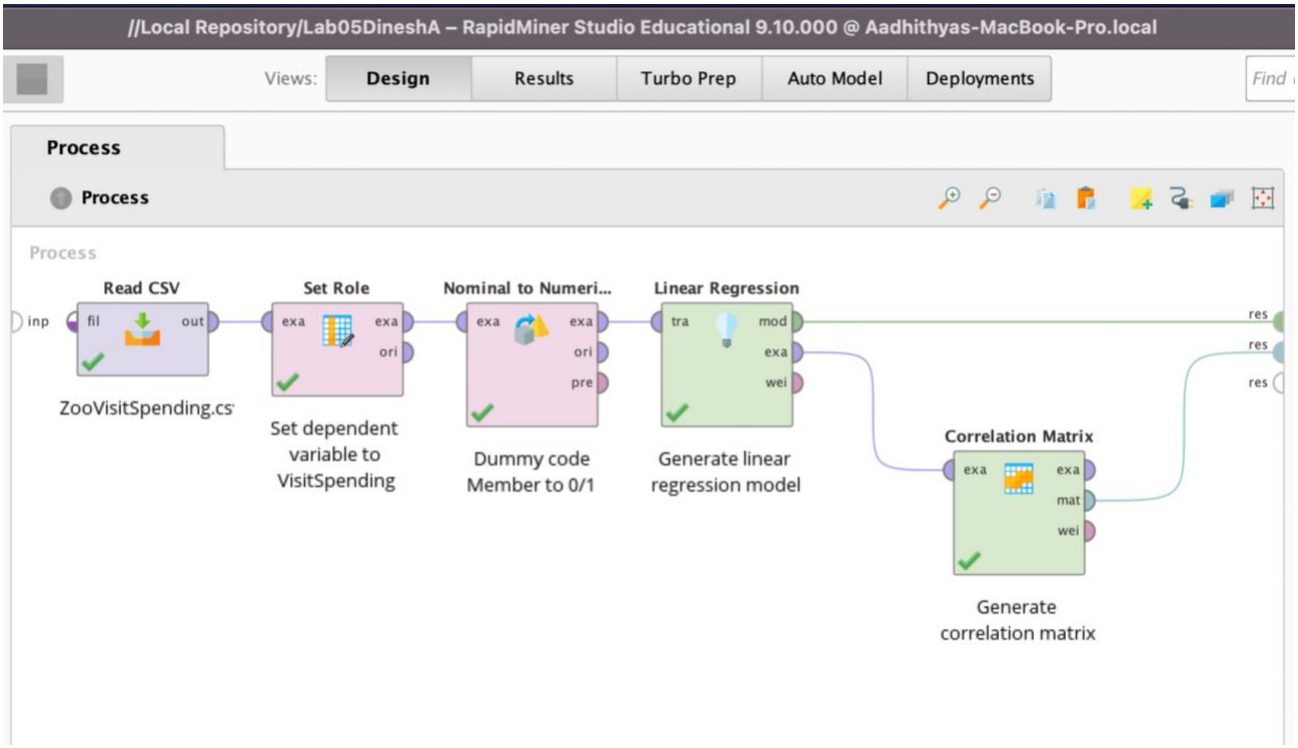
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 24.46 on 121 degrees of freedom
Multiple R-squared: 0.765, Adjusted R-squared: 0.7592
F-statistic: 131.3 on 3 and 121 DF, p-value: < 2.2e-16

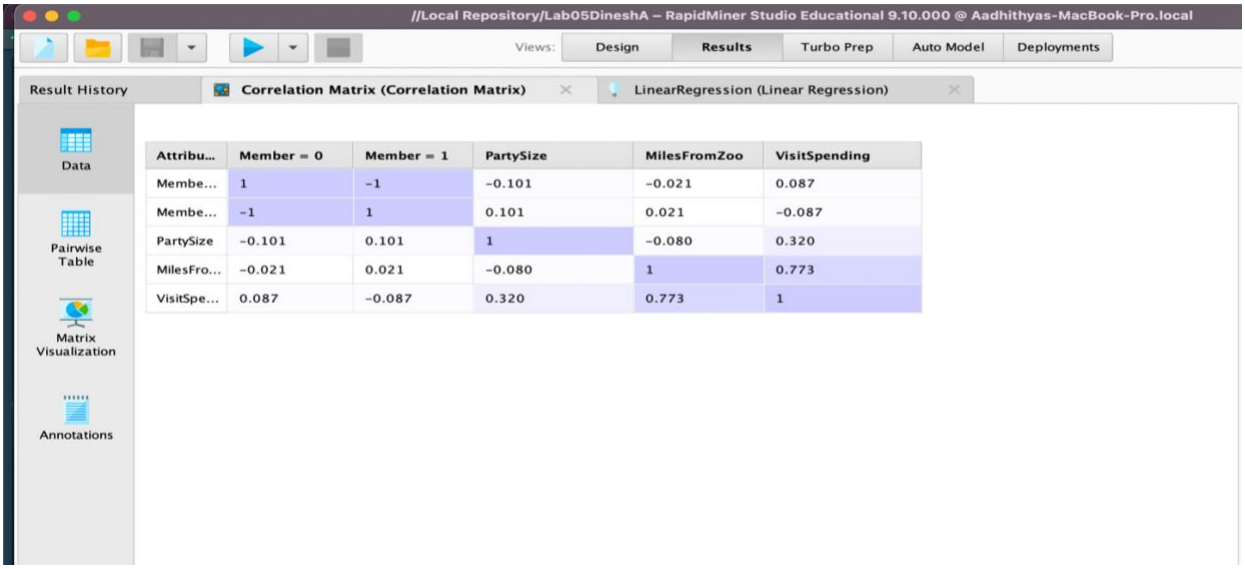
Multicollinearity results:

	Variables	Tolerance	VIF
1	PartySize	0.9831086	1.017182
2	MilesFromZoo	0.9926983	1.007355
3	MemberTRUE	0.9890274	1.011094

Rapid Miner Process:

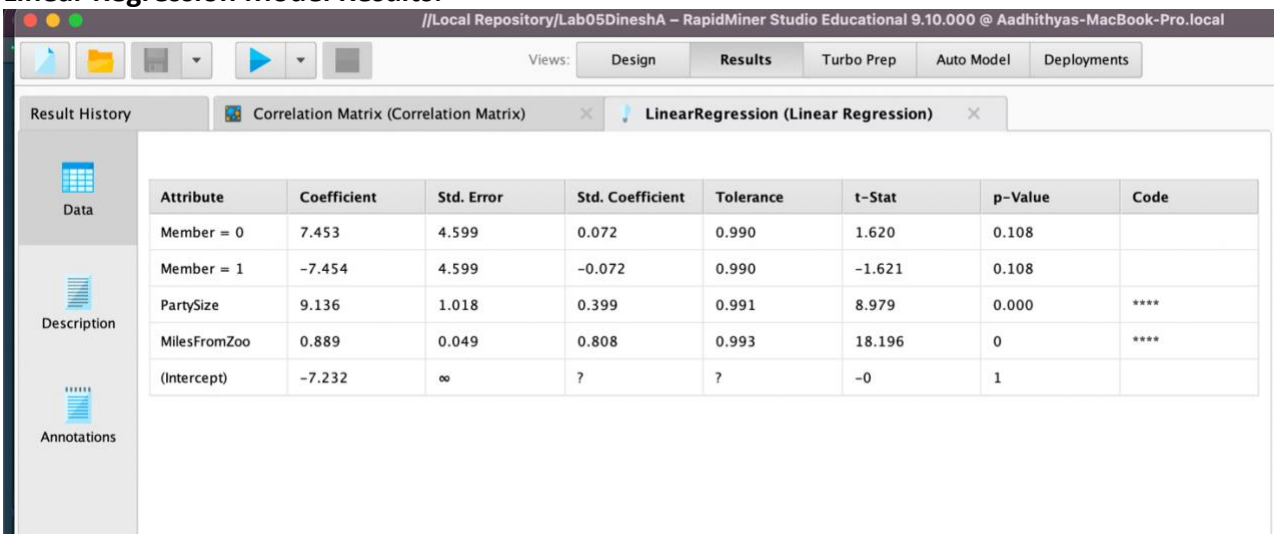


Correlation Matrix:



Attribu...	Member = 0	Member = 1	PartySize	MilesFromZoo	VisitSpending
Membe...	1	-1	-0.101	-0.021	0.087
Membe...	-1	1	0.101	0.021	-0.087
PartySize	-0.101	0.101	1	-0.080	0.320
MilesFro...	-0.021	0.021	-0.080	1	0.773
VisitSpe...	0.087	-0.087	0.320	0.773	1

Linear Regression Model Results:



Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value	Code
Member = 0	7.453	4.599	0.072	0.990	1.620	0.108	
Member = 1	-7.454	4.599	-0.072	0.990	-1.621	0.108	
PartySize	9.136	1.018	0.399	0.991	8.979	0.000	****
MilesFromZoo	0.889	0.049	0.808	0.993	18.196	0	****
(Intercept)	-7.232	∞	?	?	-0	1	

Answers:

1. MilesFromZoo seems to be the most statistically significant variable with a correlation of 0.77 followed by PartySize with 0.32.
2. The multiple R squared value of 0.765 means that 76.5% of the zoo spending can be explained by the variance in all the other variables.
3. The zoo will be spending 9 units more (as the beta coefficient suggests) for every additional guest in a party.
4. The zoo will be spending about 14 units lesser on members compared to non-members as suggested by the coefficients.
5. The zoo will be spending 0.88 units for each additional mile travelled to visit the zoo.
6. As all the tolerance values of dependent variables are >0.2 and the VIF values are < 5, there seems to be no issues of multicollinearity.