UNIVERSITY OF ASIA PACIFIC

Department of Computer Science and Engineering

Course Title : Artificial Intelligence and Expert Systems Lab
Course Code : CSE 404

Project Report

Project Title: Implementation of Multivariable Linear Regression

Submitted By:

Sudip Ghose

Section: B

ID: 18101094

Submitted to:

Dr. Nasima Begum

Assistant Professor

Department of CSE,

University Of Asia Pacific.

Problem Statement

Implement the Multivariable Linear Regression Using Open-Source Dataset with and without SK-Learn.

Dataset: Engineering Graduate Salary Prediction

Multiple Linear Regression

Multiple Linear Regression is one of the important regression algorithms which models the linear relationship between a single dependent continuous variable and more than one independent variable. The goal of multiple linear regression is to model the linear relationship between the explanatory (independent) variables and response (dependent) variables.

By looking at a linear regression model, analysts can make predictions about the dependent variable based on data gleaned from multiple independent variables.

Formula and Calculation of Multiple Linear Regression.

$yi = \beta 0 + \beta 1 xi1 + \beta 2 xi2 + ... + \beta p xip + \epsilon$

where, for $i=n$ observations:

yi = dependent variable

xi = explanatory variables

$\beta 0$ = y-intercept (constant term)

βp = slope coefficients for each explanatory variable

ϵ=the model's error term (also known as the residuals)

Tools And Languages

- **Programing Language**: Python
- **IDE**: Jupyter Notebook

Linear Regression Without SK-Learn: Output

Cost Function & Gradient Decent:

```
Iteration number:  100
Hypothesis function value is: h0(x)=theta_0+theta_1 * x
Cost function is: j(theta)=1/(2*m) * i=1_samtionSign_m (h_theta_(x)-y)**2

Cost function is: 22532967658.318554

gradient decent:
New parameter value is:  [304789.8765748265, 6685.813441286741, 7324.264946449583, 7494.173323740315]
result coefficient is [304789.8765748265, 6685.813441286741, 7324.264946449583, 7494.173323740315]
```

Graph Plotting:

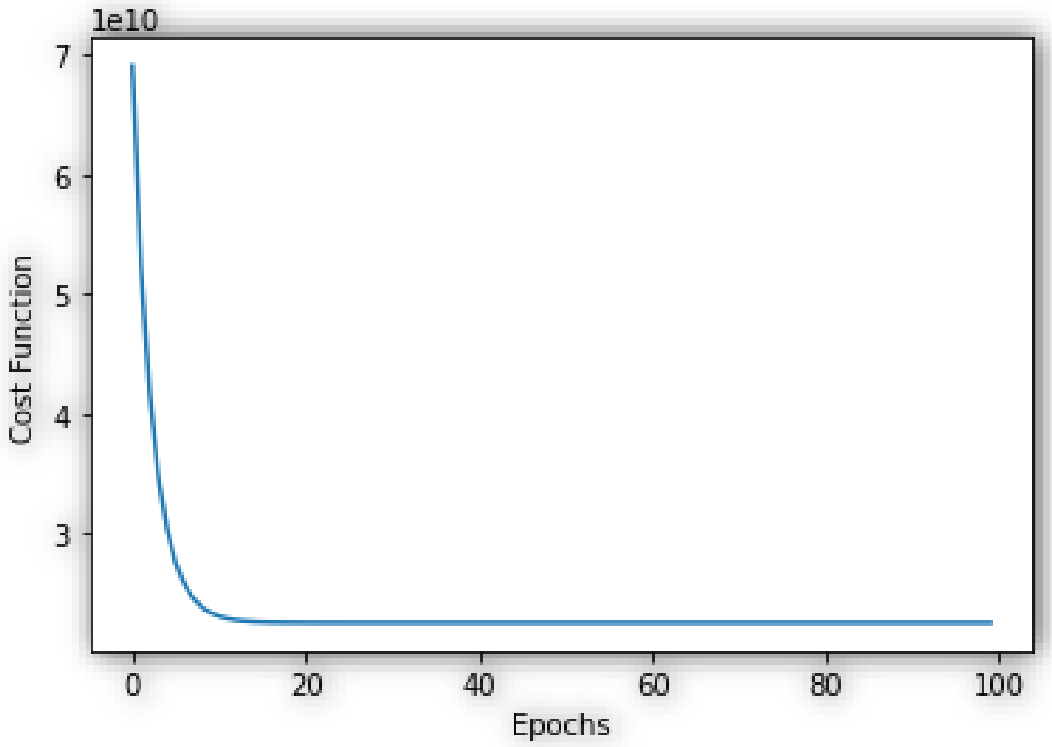
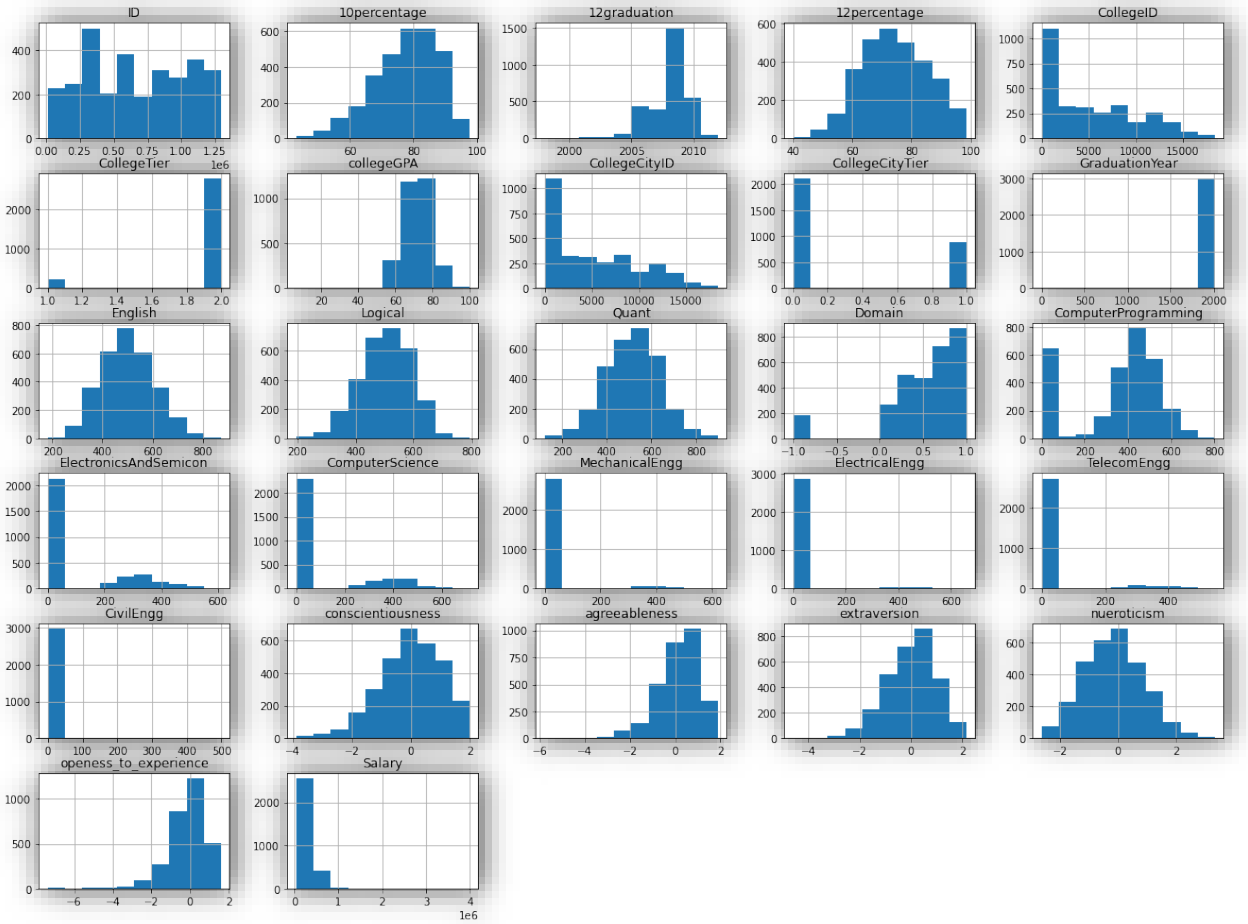
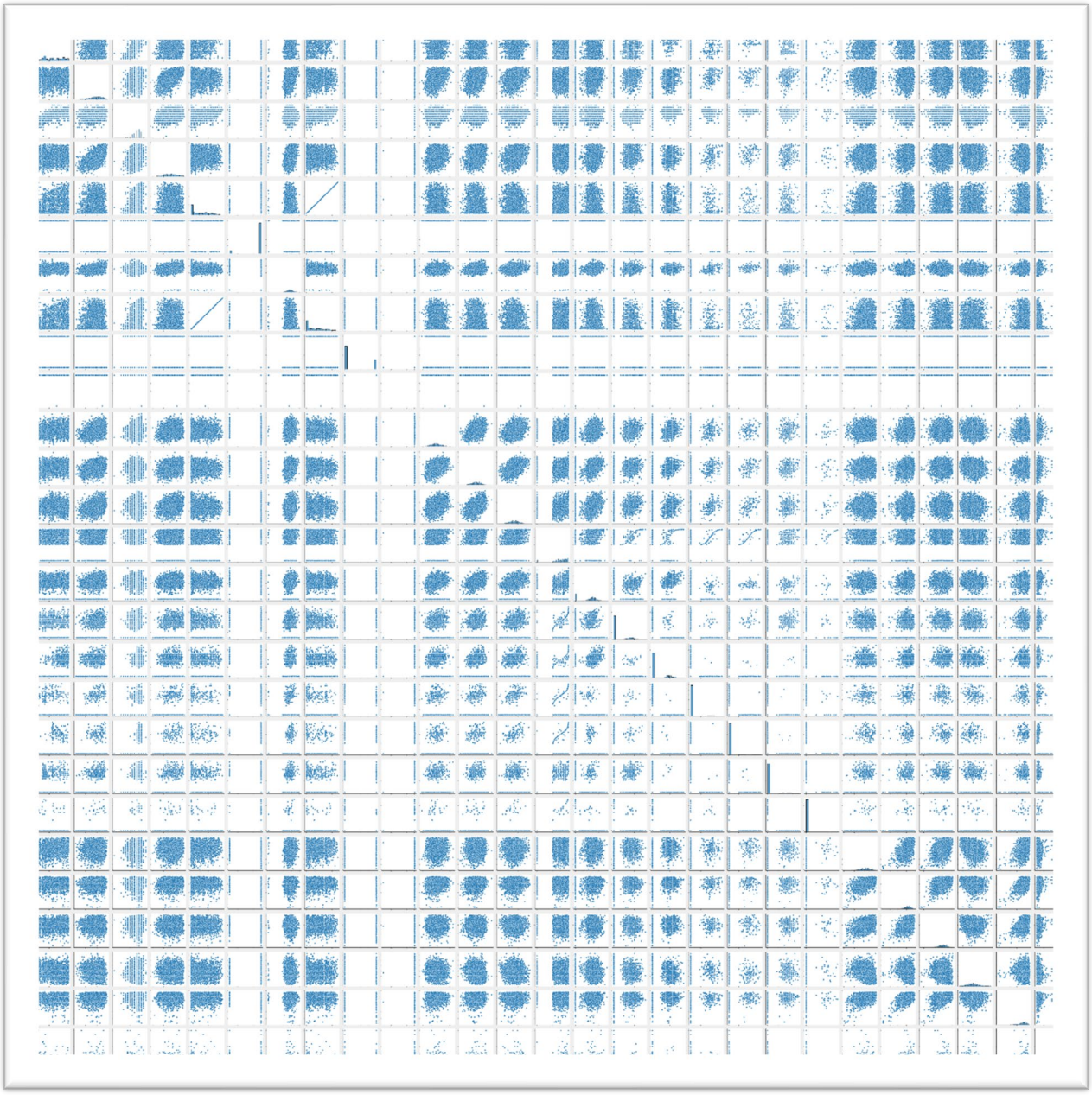


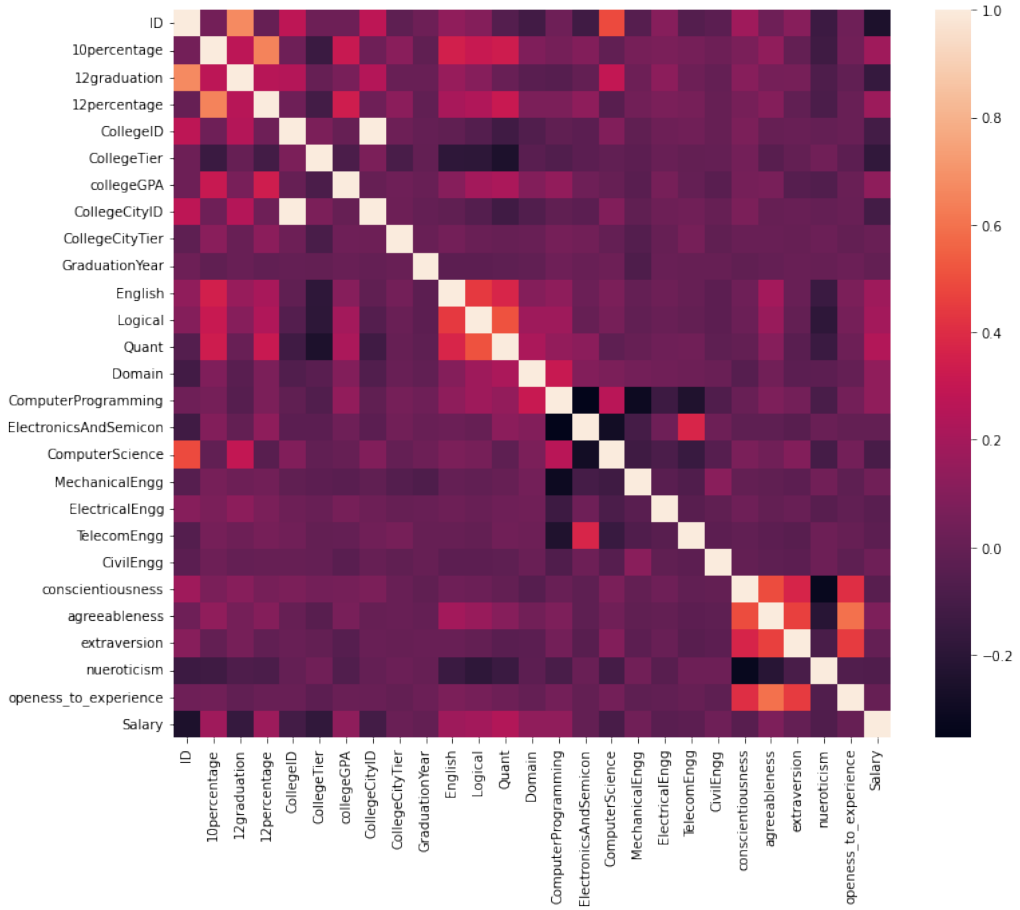
Figure Plotting:



Pairwise Figure:



HeatMap:



Pandas Profiling Report:

Overview

Overview | Alerts 58 | Reproduction

Dataset statistics

Number of variables	34
Number of observations	2998
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	2.0 MiB
Average record size in memory	692.0 B

Variable types

Numeric	25
Categorical	9

Linear Regression With SK-Learn: Output

Coefficients , Mean Error & Variance Score:

```
print('Coefficients: \n', regr.coef_)
# The Mean Squared Error
print("Mean squared error: %.2f" % np.mean((regr.predict(X_test) - y_test) ** 2))
# Explained Variance Score: 1 is perfect prediction
print('Variance score: %.2f' % regr.score(X_test, y_test))
```

```
Coefficients:
 [1877.74720053 2295.48993972 1506.06654496]
Mean squared error: 32685259466.73
Variance score: 0.05
```

```
# Predicting the Test set results
y_pred = regr.predict(X_test)

from sklearn.metrics import r2_score
score=r2_score(y_test,y_pred)
print(score)
```

```
0.050827764629067906
```

Graph Plotting:

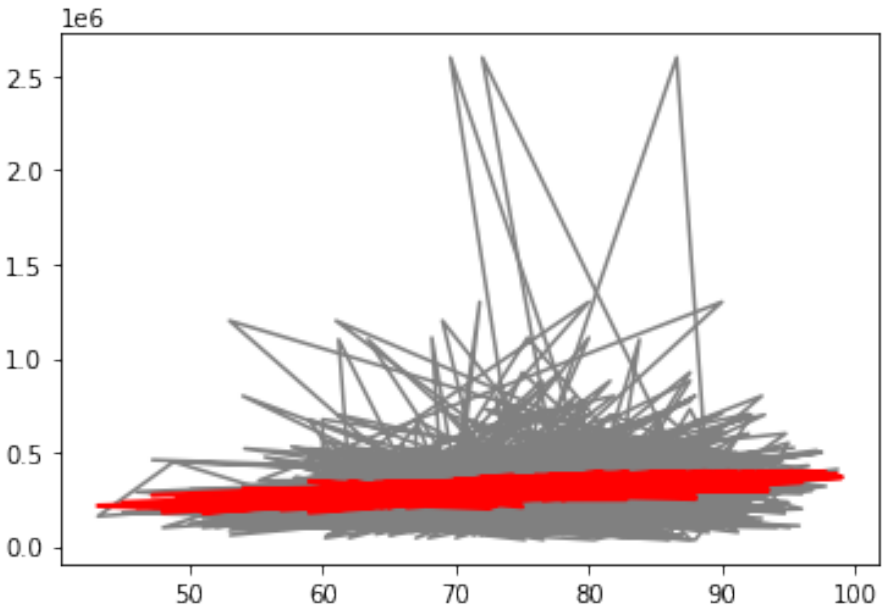
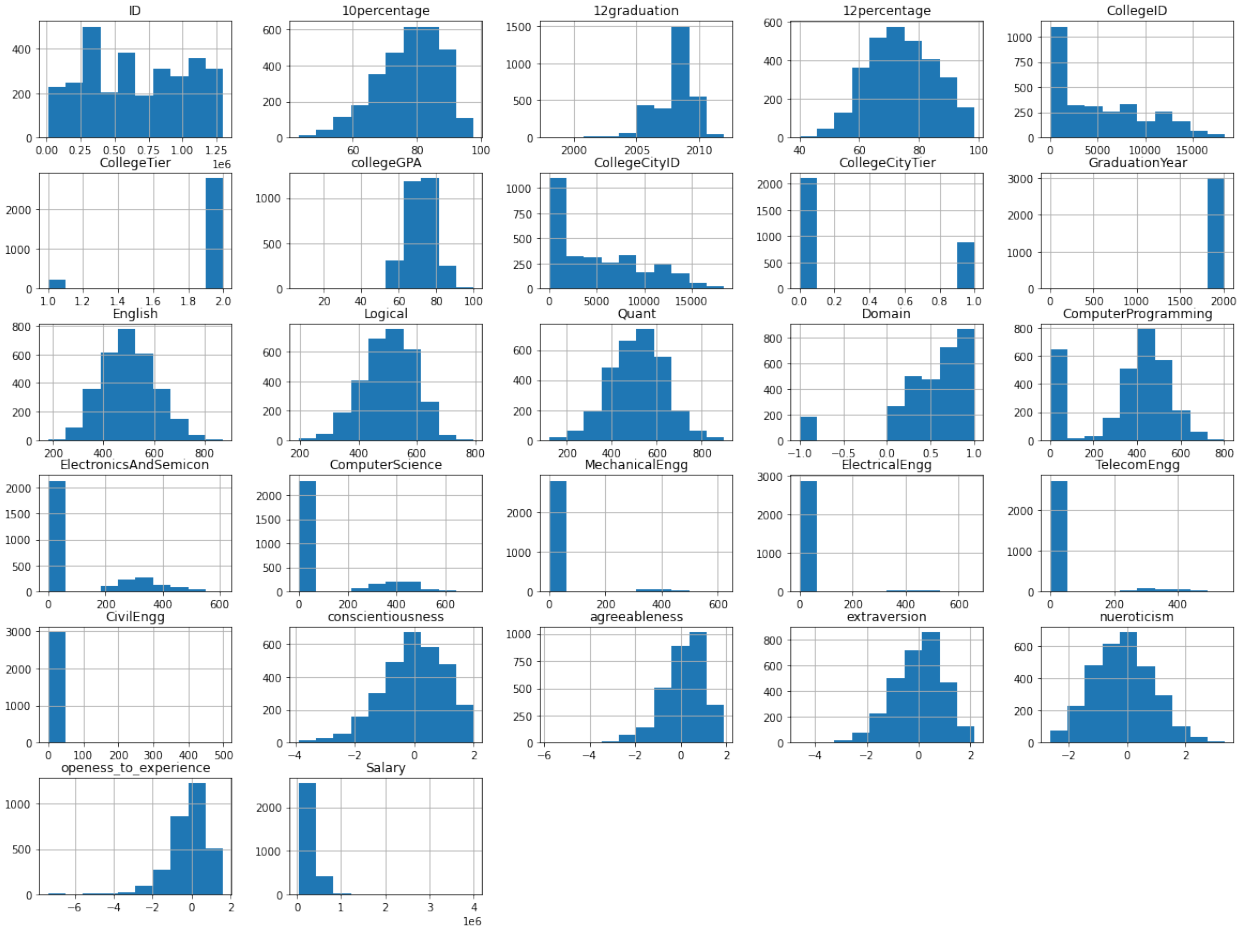
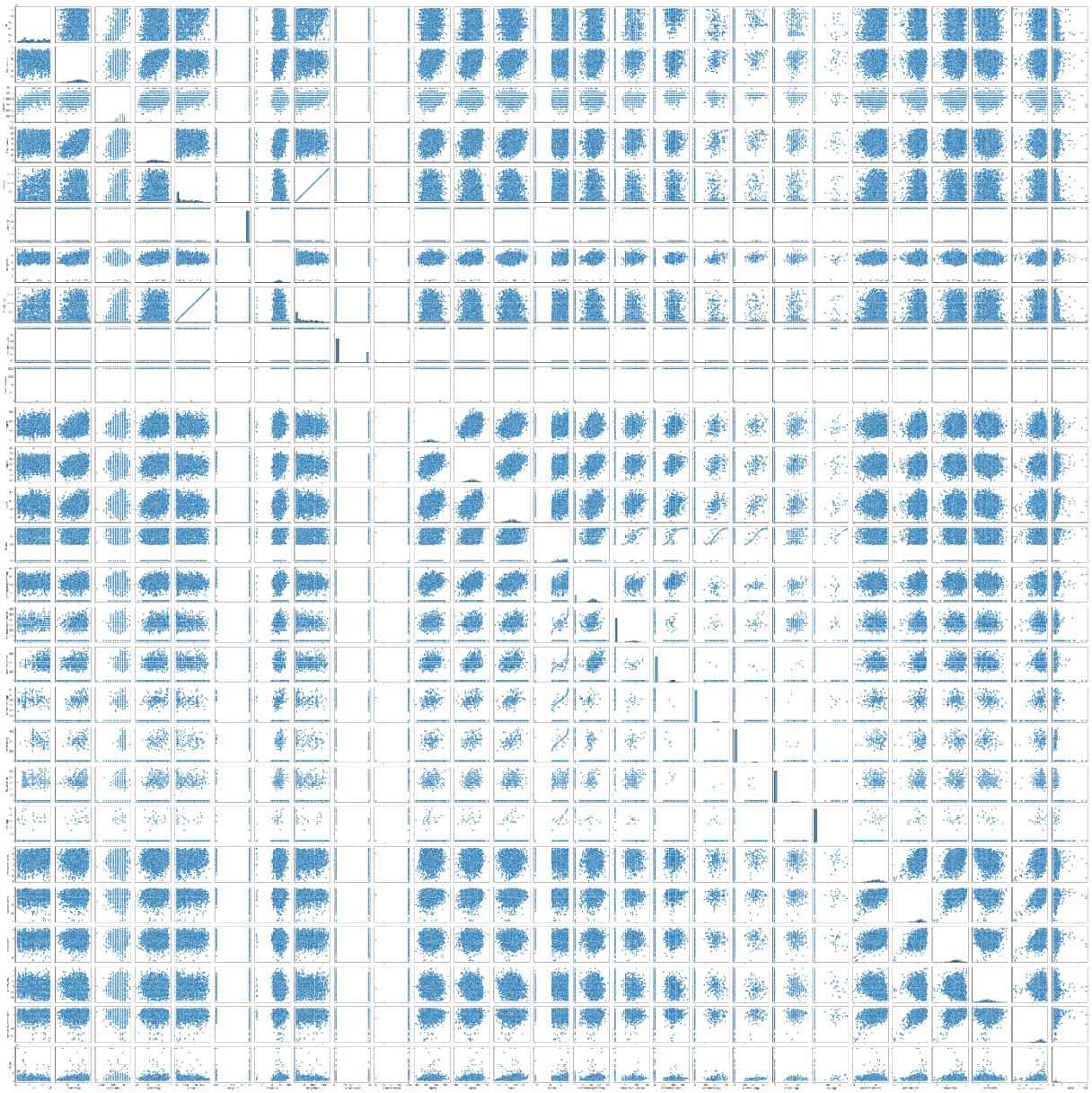


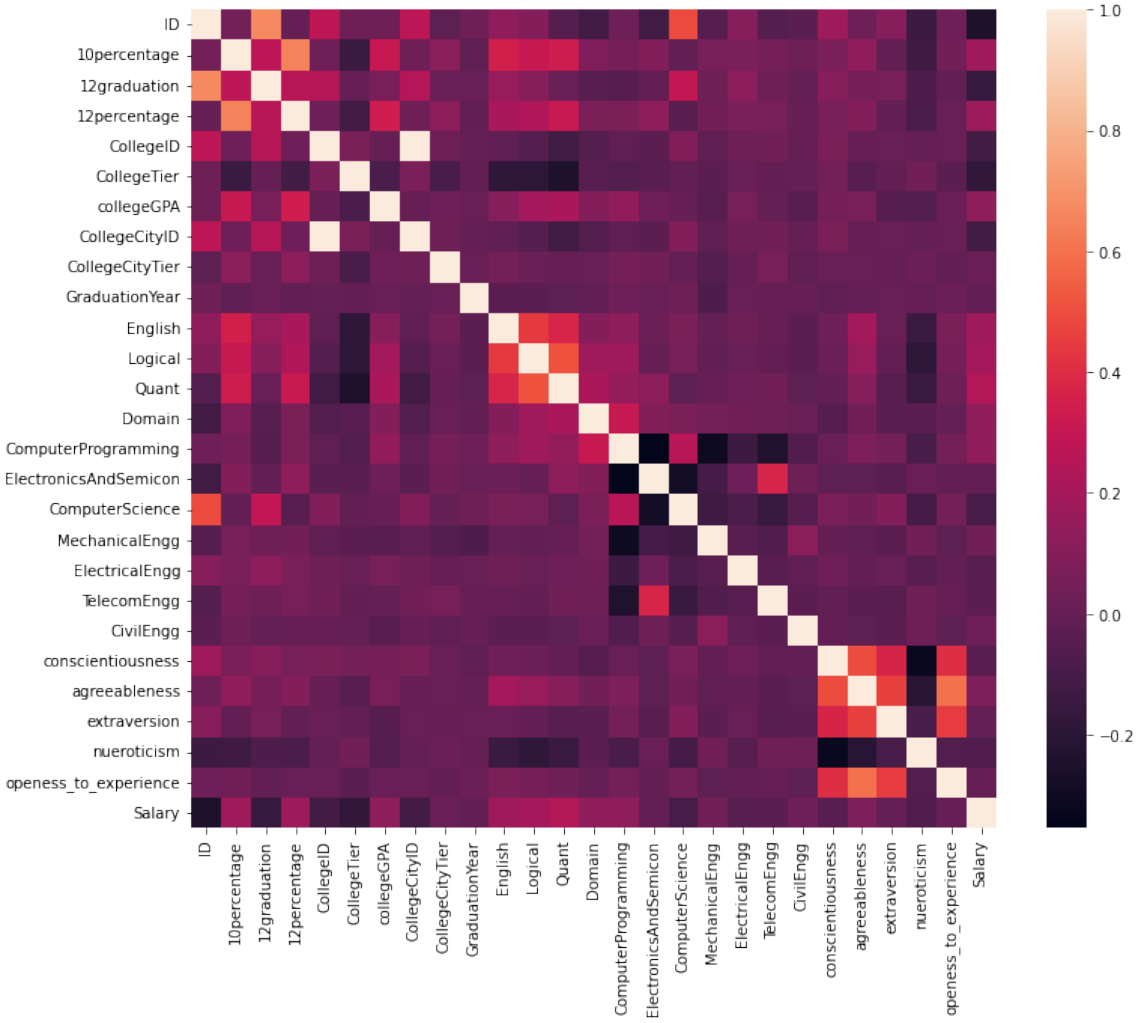
Figure Plotting:



Pairwise Figure:



HeatMap:



Pandas Profiling Report:

Overview

Overview | Alerts 58 | Reproduction

Dataset statistics

Number of variables	34
Number of observations	2998
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	2.0 MiB
Average record size in memory	692.0 B

Variable types

Numeric	25
Categorical	9

Conclusion

Multiple regression is an extension of simple linear regression. In this project, after successful implementation, I've a brief knowledge about linear regression. Multiple linear regression models are useful in helping an enterprise to consider the impact of multiple independent predictors and variables on a dependent variable and can be beneficial for forecasting and predicting results. So, the outcome knowledge of this project will help me a lot to real-life problem predicting and solving on various field.

--The End--