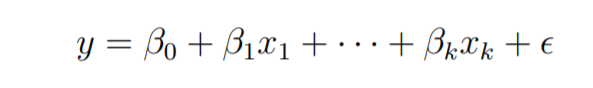
|  |  |
| --- | --- |
| Algorithm: Linear Regression | |
| USN: 1MS18CS024 | NAME: Aravind S |
| USN: 1MS18CS029 | NAME: Arun J Kennedy |

**Description of the Algorithm: <<Write 2-3 Paragraphs about the Algorithm>>**

* Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models a target prediction value based on independent variables.
* Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x).
* In the given dataset of Seaborn Tips there are many variables on which the tip can depend on, so Multivariable Linear Regression is used to determine the tip more accurately.



**Algorithm Pseudocode:**

1. Start

2. Read Number of Data (n)

3. For i=1 to n:

Read Xi and Yi

Next i

4. Initialize:

sumX = 0

sumX2 = 0

sumY = 0

sumXY = 0

5. Calculate Required Sum

For i=1 to n:

sumX = sumX + Xi

sumX2 = sumX2 + Xi \* Xi

sumY = sumY + Yi

sumXY = sumXY + Xi \* Yi

Next i

6. Calculate Required Constant a and b of y = a + bx:

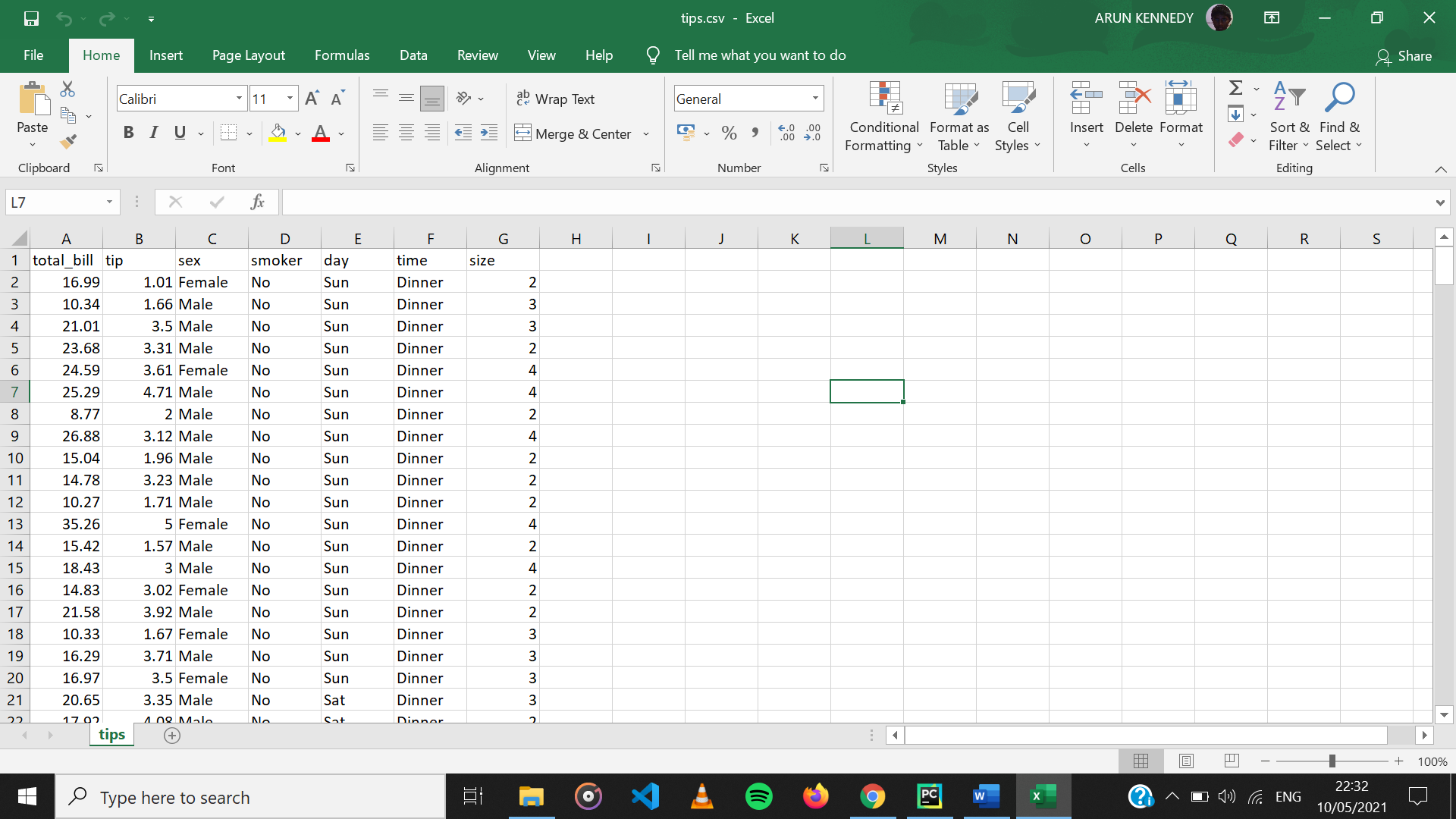
b = (n \* sumXY - sumX \* sumY)/(n\*sumX2 - sumX \* sumX)

a = (sumY - b\*sumX)/n

7. Display value of a and b

8. Stop

**Data set Used: (Attach Screen shot of the few rows and also the Kaggle/Dataset link)**



Kaggle link: <https://www.kaggle.com/ranjeetjain3/seaborn-tips-dataset>

**Challenges faced during the implementation of the program:**

**Problem 1:** The data set contained Non-numerical values which is not suitable for Linear Regression.

**Solution:** Map the non-numerical data values to numerical data values.

**Problem 2:** The Tips value of the dataset seem to rely on more than one parameter.

**Solution:** Use the Multi-Variable Linear Regression Model.

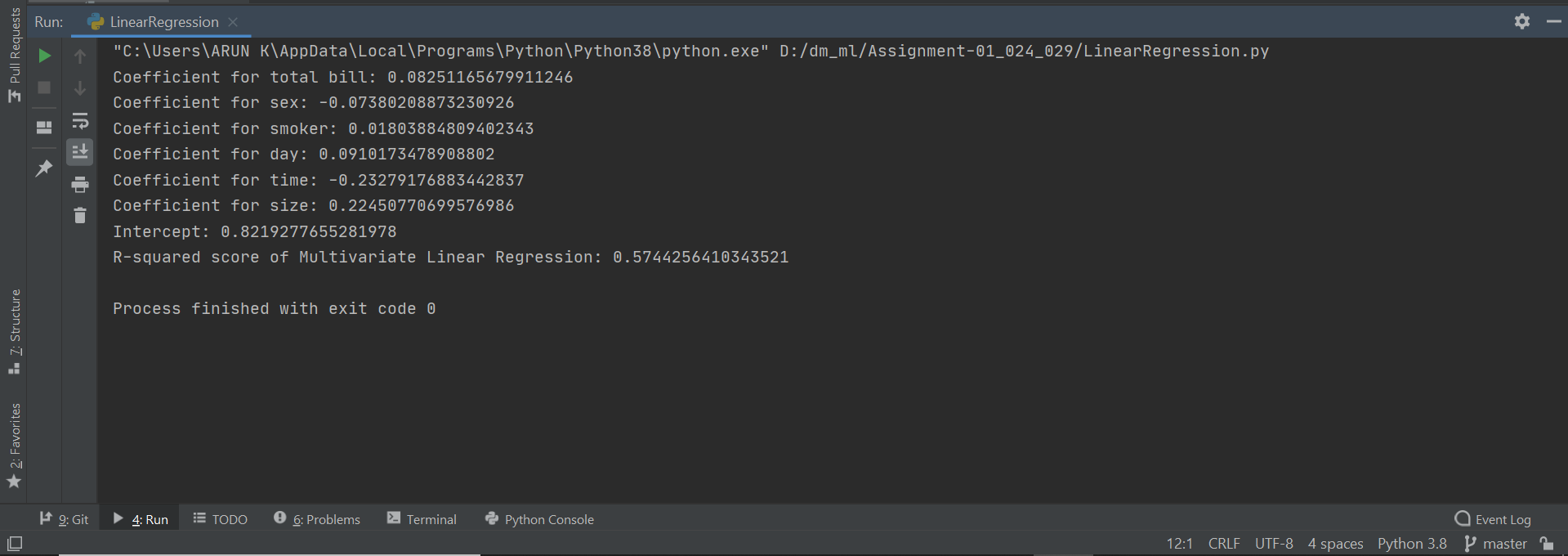
**Advantages & Disadvantages of the Algorithm:**

|  |  |
| --- | --- |
| Advantages | Disadvantages |
| Linear Regression is simple to implement and easier to interpret the output coefficients. | On the other hand in linear regression technique outliers can have huge effects on the regression and boundaries are linear in this technique. |
| When you know the relationship between the independent and dependent variable have a linear relationship, this algorithm is the best to use because of it’s less complexity to compared to other algorithms. | Diversely, linear regression assumes a linear relationship between dependent and independent variables. That means it assumes that there is a straight-line relationship between them. It assumes independence between attributes. |
| Linear Regression is susceptible to over-fitting but it can be avoided using some dimensionality reduction techniques, regularization (L1 and L2) techniques and cross-validation. | But then linear regression also looks at a relationship between the mean of the dependent variables and the independent variables. Just as the mean is not a complete description of a single variable, linear regression is not a complete description of relationships among variables. |

**Applications of the Algorithm:**

* Linear Regression can be also used to assess risk in financial services or insurance domain. For example, a car insurance company might conduct a linear regression to come up with a suggested premium table using predicted claims to Insured Declared Value ratio.
* Linear regression can be used to analyze the marketing effectiveness, pricing and promotions on sales of a product. For instance, if company XYZ, wants to know if the funds that they have invested in marketing a particular brand has given them substantial return on investment, they can use linear regression.
* In the credit card industry, a financial company maybe interested in minimizing the risk portfolio and wants to understand the top five factors that cause a customer to default. Based on the results the company could implement specific EMI options so as to minimize default among risky customers.

**Output: (Screen shots)**



**References:**

* [**https://www.geeksforgeeks.org/ml-linear-regression/#:~:text=Linear%20Regression%20is%20a%20machine,value%20based%20on%20independent%20variables.&text=Linear%20regression%20performs%20the%20task,given%20independent%20variable%20(x)**](https://www.geeksforgeeks.org/ml-linear-regression/#:~:text=Linear%20Regression%20is%20a%20machine,value%20based%20on%20independent%20variables.&text=Linear%20regression%20performs%20the%20task,given%20independent%20variable%20(x))**.**
* [**https://www.geeksforgeeks.org/ml-advantages-and-disadvantages-of-linear-regression/**](https://www.geeksforgeeks.org/ml-advantages-and-disadvantages-of-linear-regression/)
* [**https://www.codesansar.com/numerical-methods/linear-regression-method-pseudocode.htm**](https://www.codesansar.com/numerical-methods/linear-regression-method-pseudocode.htm)