**ICP-9:**

**Name: Amarnadha Reddy Ankireddypalli**

**ID: 16314889**

**Email: aa9f3@umsystem.edu**

**Description:**

Create a linear regression model in python using any dataset of your choice. For this model you can also create your own data. Find the best fit line in the data and calculate SSE (sum of square error) or MSE (Mean square error) , Y intercept, and Slope for the relationship in data. Explain your findings and understanding of these terms in detail in the report.

**Objective:**

1. Successfully executing the code with linear regression model and calculating following:
   1. SSE or MSE
   2. Y intercept
   3. Slope
2. Detail explanation of each in report
3. overall code quality
4. Wiki Report quality, video explanation

**Implementation (with screenshots):**

1. Import all the required libraries.

Text

Description automatically generated

2. In this ICP, we are going to implement Linear Regression on advertising dataset.

A picture containing graphical user interface

Description automatically generated

3. Here is the info of the dataset. All the columns are of type ‘float64’.

Table

Description automatically generated

4. Here is the statistical information of each column.

Table

Description automatically generated

5. There are no null values in any of the column.

Table

Description automatically generated with low confidence

6. Performed **Outlier analysis** on the data plotting boxplot of every column. I observed that there are no outliers present in any of the column.

Chart

Description automatically generated

7. **Heat map:** We can see that there is correlation of 0.9 between 'TV' and Sales. So, these columns are highly positive correlated. It means change in TV columns has major impact on Sales column.

Chart, treemap chart

Description automatically generated

8. **Pair plot:** From the below plot, it is quite obvious that the feature 'TV' has more linear relationship with 'Sales' column.

Chart

Description automatically generated

9. Now, split the dataset into features and target variable as follows. Declare **Linear Regression Model** and then fit the model.

Graphical user interface, text, application

Description automatically generated

10. **Slope and Intercept:** The slope indicates the steepness of a line, and the intercept indicates the location where it intersects an axis. The greater the magnitude of the slope, the steeper the line and the greater the rate of change.

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text

Description automatically generated with medium confidence

11. **Plotting Regression Line:**

Chart, scatter chart

Description automatically generated

We have positive slope and intercept which can be seen in the above plot.

12. **R squared and MSE :** R-Squared is also known as MSE's standardized form. R-squared represents the fraction of variance captured by the regression model rather than the MSE, which represents the residual error.

Graphical user interface, text, application, chat or text message

Description automatically generated

Graphical user interface, text, application

Description automatically generated

The term **mean squared error (MSE)** is commonly used in regression analysis to refer to the unbiased estimate of error variance, which is the **residual** **sum of squares** **(SSE)** divided by the number of degrees of freedom.

A picture containing text

Description automatically generated

It is recommended that R-Squared or adjusted R-Squared be used to evaluate the model performance of regression models. This is mainly because R-Squared captures the fraction of response variance captured by the regression and tends to provide a more accurate view of the regression model's quality. Furthermore, MSE values differ depending on whether the response variable's values are scaled or not.

**Difficulties faced:** No difficulty faced.

**Video Link:** [**https://youtu.be/QYAuebaPA30**](https://youtu.be/QYAuebaPA30)

**Conclusion:**

From this ICP9, I’ve learnt:

* About linear regression model
* How R squared, MSE and SSE works