Data Analytics Lab

FAT

S Shyam Sundaram
19BCE1560

6th December, 2021

Question set 1

Par
Part 2 Mean & linear regression imputation

Mean Imputation Algorithm

1. Let dataframe = df
2. Calculate Mean of df$1958
3. mean = Average(df$1958)
4. For i in df:
   if (df$1958[i] == NA):
   df$1958[i] = mean
5. END

Linear regression imputation Algorithm

1. Let df = dataframe
2. ~~Calculate~~
2. Fit linear model with 1958 as dependent variable and 1959 as independent variable
3. Let i = intercept , c = coefficient of independent
4. For i in df:

```
if (df$1958[i] == NA):
    df$1958[i] = i + (c * df$1959[i])
```

5. END

Explanation:-

So in mean imputation, we replace the NAs of the column with its mean.
But in linear regression, we first fit a linear model ~~on~~ with "1959" as independent variable.
Using this model, we predict the missing values.