Name: Vikram Surya R
Reg No: 19BCE1490

Essential of Data Analytics
Tasks for Week-4: Analysis of Variance (ANOVA)

Aim: Perform ANOVA test and determine the statistical differences between the means of individual groups given in the data

Algorithm:
• The path for data set is set using setwd function.
• Import the dplyr library using library(dplyr)
• Reading the dataset color-anova-example.csv using read.csv function.
• Grouping the dataset into sample data according to the type of colors we have in colors column in dataset. this is done by group_by().
• By using anova model we can predict whether the sample datasets are having almost same mean and variance or not. This can be done by aov() function.
• Using summary function the we get the statistical values of anova model if the pr(>F) value is more than 0.05 that means all the samples are having almost same mean and also all samples belongs to same population but if the value is less than 0.05 that means the samples doesn't have almost same mean in that case we have to TukeyHSD to find which combination mean has more different.
• Using TukeyHSD function we can find which combination of samples having more mean difference or significantly different then that can be said if the p adj value is less than the 0.05.

Statistic:

```
> summary(ANOVA)
            Df Sum Sq Mean Sq F value  Pr(>F)
color        2  857.2   428.6   14.81 4.44e-06 ***
Residuals   69 1996.4    28.9
---
```

Inference:
The F value is less than 0.05 that means we have to reject our null hypothesis.
As the null hypothesis is rejected, TukeyHSD will be used to check which system have significantly different. So the samples red-blue and red-green have significantly different because the p adj value is less than 0.05.

Result:

```
$color
                diff        lwr       upr      p adj
green-blue  -2.101667  -5.821045  1.617711 0.3709119
red-blue    -8.140417 -11.859795 -4.421039 0.0000049
red-green   -6.038750  -9.758128 -2.319372 0.0006628
```

CODE:

```
setwd("C:/Users/VIKRAM SURYA/Desktop/EDA_LAB""")
library(dplyr)
data<-read.csv("color-anova-example.csv")
group_by(data,color)%>%summarise(count=n(),mean=mean(response,na.ra=TRUE))
ANOVA<-aov(response~color,data=data)
summary(ANOVA)
TukeyHSD(ANOVA)
```