DS785_Capstone Project (Zacharia Kibuta)

Establish connection to the ODSP

```
source('V:/R Code/R Set Up/connect.R')

## Loading required package: pacman

## Hi magu2021, Welcome to your R workspace! Have a productive day!!

pacman::p_load(pacman, DBI,RODBC, keyring, dbplyr, tidyverse, lubridate,sqldf
, tidyquery, readxl,writexl,plyr,janitor, reshape2,ggpubr,rstatix) #LOAD THE
LIBRARIES
```

#Creating the Dataframe

```
#Total number of accidents by the calendar year

year <- c(2000,2001,2002,2003,2004,2005,2006,2007,2008,2009,2010,2011,2012,20
13,2014,2015,2016,2017,2018,2019,2020,2021,2022,
          2000,2001,2002,2003,2004,2005,2006,2007,2008,2009,2010,2011,2012,20
13,2014,2015,2016,2017,2018,2019,2020,2021,2022,
          2000,2001,2002,2003,2004,2005,2006,2007,2008,2009,2010,2011,2012,20
13,2014,2015,2016,2017,2018,2019,2020,2021,2022,
          2000,2001,2002,2003,2004,2005,2006,2007,2008,2009,2010,2011,2012,20
13,2014,2015,2016,2017,2018,2019,2020,2021,2022,
          2000,2001,2002,2003,2004,2005,2006,2007,2008,2009,2010,2011,2012,20
13,2014,2015,2016,2017,2018,2019,2020,2021,2022,
          2000,2001,2002,2003,2004,2005,2006,2007,2008,2009,2010,2011,2012,20
13,2014,2015,2016,2017,2018,2019,2020,2021,2022,
          2000,2001,2002,2003,2004,2005,2006,2007,2008,2009,2010,2011,2012,20
13,2014,2015,2016,2017,2018,2019,2020,2021,2022)

count <-c( 5,1,3,1,1,3,4,4,4,2,6,1,2,3,0,0,3,1,2,1,1,0,0,
           0,0,0,0,1,1,0,0,1,2,0,0,0,0,0,0,0,0,1,0,1,1,0,
           0,0,1,0,1,0,0,0,1,0,0,1,0,0,0,0,1,0,0,0,0,0,0,
           0,0,0,2,0,0,0,0,1,2,0,0,0,0,1,0,0,0,1,1,0,1,0,
           0,1,0,0,0,0,0,0,0,0,1,1,0,0,0,0,1,0,1,0,0,0,0,
           0,4,0,0,0,0,0,0,0,0,0,0,1,1,0,2,0,0,0,0,0,0,0,
           0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0,0,0)

# PE - PILOT ERROR
# FD - FAULTY DATA
# AW - ADVERSE WEATHER
# MF - MECHANICAL FAILURE
# FO - FIRE ON-BOARD
# TR - TERRORISM
# FA - FAULTY AUTOMATION
```

```
variables<- c("PE","PE","PE","PE","PE","PE","PE","PE","PE","PE","PE","PE","PE
","PE","PE","PE","PE","PE","PE","PE","PE","PE","PE","PE",
              "FD","FD","FD","FD","FD","FD","FD","FD","FD","FD","FD","FD","FD
","FD","FD","FD","FD","FD","FD","FD","FD","FD","FD",
              "AW","AW","AW","AW","AW","AW","AW","AW","AW","AW","AW","AW","AW
","AW","AW","AW","AW","AW","AW","AW","AW","AW","AW",
              "MF","MF","MF","MF","MF","MF","MF","MF","MF","MF","MF","MF","MF
","MF","MF","MF","MF","MF","MF","MF","MF","MF","MF",
              "FO","FO","FO","FO","FO","FO","FO","FO","FO","FO","FO","FO","FO
","FO","FO","FO","FO","FO","FO","FO","FO","FO","FO",
              "TR","TR","TR","TR","TR","TR","TR","TR","TR","TR","TR","TR","TR
","TR","TR","TR","TR","TR","TR","TR","TR","TR","TR",
              "FA","FA","FA","FA","FA","FA","FA","FA","FA","FA","FA","FA","FA
","FA","FA","FA","FA","FA","FA","FA","FA","FA","FA")

df <- data.frame(year, count, variables)
df

##     year count variables
## 1   2000     5        PE
## 2   2001     1        PE
## 3   2002     3        PE
## 4   2003     1        PE
## 5   2004     1        PE
## 6   2005     3        PE
## 7   2006     4        PE
## 8   2007     4        PE
## 9   2008     4        PE
## 10  2009     2        PE
## 11  2010     6        PE
## 12  2011     1        PE
## 13  2012     2        PE
## 14  2013     3        PE
## 15  2014     0        PE
## 16  2015     0        PE
## 17  2016     3        PE
## 18  2017     1        PE
## 19  2018     2        PE
## 20  2019     1        PE
## 21  2020     1        PE
## 22  2021     0        PE
## 23  2022     0        PE
## 24  2000     0        FD
## 25  2001     0        FD
## 26  2002     0        FD
## 27  2003     0        FD
## 28  2004     1        FD
## 29  2005     1        FD
## 30  2006     0        FD
```

```
## 31  2007     0        FD
## 32  2008     1        FD
## 33  2009     2        FD
## 34  2010     0        FD
## 35  2011     0        FD
## 36  2012     0        FD
## 37  2013     0        FD
## 38  2014     0        FD
## 39  2015     0        FD
## 40  2016     0        FD
## 41  2017     0        FD
## 42  2018     1        FD
## 43  2019     0        FD
## 44  2020     1        FD
## 45  2021     1        FD
## 46  2022     0        FD
## 47  2000     0        AW
## 48  2001     0        AW
## 49  2002     1        AW
## 50  2003     0        AW
## 51  2004     1        AW
## 52  2005     0        AW
## 53  2006     0        AW
## 54  2007     0        AW
## 55  2008     1        AW
## 56  2009     0        AW
## 57  2010     0        AW
## 58  2011     1        AW
## 59  2012     0        AW
## 60  2013     0        AW
## 61  2014     0        AW
## 62  2015     0        AW
## 63  2016     1        AW
## 64  2017     0        AW
## 65  2018     0        AW
## 66  2019     0        AW
## 67  2020     0        AW
## 68  2021     0        AW
## 69  2022     0        AW
## 70  2000     0        MF
## 71  2001     0        MF
## 72  2002     0        MF
## 73  2003     2        MF
## 74  2004     0        MF
## 75  2005     0        MF
## 76  2006     0        MF
## 77  2007     0        MF
## 78  2008     1        MF
## 79  2009     2        MF
## 80  2010     0        MF
```

```
## 81  2011     0        MF
## 82  2012     0        MF
## 83  2013     0        MF
## 84  2014     1        MF
## 85  2015     0        MF
## 86  2016     0        MF
## 87  2017     0        MF
## 88  2018     1        MF
## 89  2019     1        MF
## 90  2020     0        MF
## 91  2021     1        MF
## 92  2022     0        MF
## 93  2000     0        FO
## 94  2001     1        FO
## 95  2002     0        FO
## 96  2003     0        FO
## 97  2004     0        FO
## 98  2005     0        FO
## 99  2006     0        FO
## 100 2007     0        FO
## 101 2008     0        FO
## 102 2009     0        FO
## 103 2010     1        FO
## 104 2011     1        FO
## 105 2012     0        FO
## 106 2013     0        FO
## 107 2014     0        FO
## 108 2015     0        FO
## 109 2016     1        FO
## 110 2017     0        FO
## 111 2018     1        FO
## 112 2019     0        FO
## 113 2020     0        FO
## 114 2021     0        FO
## 115 2022     0        FO
## 116 2000     0        TR
## 117 2001     4        TR
## 118 2002     0        TR
## 119 2003     0        TR
## 120 2004     0        TR
## 121 2005     0        TR
## 122 2006     0        TR
## 123 2007     0        TR
## 124 2008     0        TR
## 125 2009     0        TR
## 126 2010     0        TR
## 127 2011     0        TR
## 128 2012     1        TR
## 129 2013     1        TR
## 130 2014     0        TR
```

```
## 131 2015     2         TR
## 132 2016     0         TR
## 133 2017     0         TR
## 134 2018     0         TR
## 135 2019     0         TR
## 136 2020     0         TR
## 137 2021     0         TR
## 138 2022     0         TR
## 139 2000     0         FA
## 140 2001     0         FA
## 141 2002     0         FA
## 142 2003     0         FA
## 143 2004     0         FA
## 144 2005     0         FA
## 145 2006     0         FA
## 146 2007     0         FA
## 147 2008     0         FA
## 148 2009     0         FA
## 149 2010     0         FA
## 150 2011     0         FA
## 151 2012     0         FA
## 152 2013     0         FA
## 153 2014     0         FA
## 154 2015     0         FA
## 155 2016     0         FA
## 156 2017     0         FA
## 157 2018     1         FA
## 158 2019     1         FA
## 159 2020     0         FA
## 160 2021     0         FA
## 161 2022     0         FA
```

Normality Test Statistical Tests to be used (PARAMETRIC or NON-PARAMETRIC TEST)

```
#Factors
PE<-df%>%filter(variables=='PE')
FD<-df%>%filter(variables=='FD')
AW<-df%>%filter(variables=='AW')
MF<-df%>%filter(variables=='MF')
FO<-df%>%filter(variables=='FO')
TR<-df%>%filter(variables=='TR')
FA<-df%>%filter(variables=='FA')

#Normality Test using Shapiro Test

shapiro.test(PE$count) #Normality Test for distribution of number of accident
s due to PE
```

```
##
##  Shapiro-Wilk normality test
##
## data:  PE$count
## W = 0.91389, p-value = 0.04935
```

```
shapiro.test(FD$count) #Normality Test for distribution of number of accident
s due to FD
```

```
##
##  Shapiro-Wilk normality test
##
## data:  FD$count
## W = 0.6326, p-value = 2.099e-06
```

```
shapiro.test(AW$count) #Normality Test for distribution of number of accident
s due to AW
```

```
##
##  Shapiro-Wilk normality test
##
## data:  AW$count
## W = 0.51224, p-value = 1.124e-07
```

```
shapiro.test(MF$count) #Normality Test for distribution of number of accident
s due to MF
```

```
##
##  Shapiro-Wilk normality test
##
## data:  MF$count
## W = 0.63431, p-value = 2.197e-06
```

```
shapiro.test(FO$count) #Normality Test for distribution of number of accident
s due to FO
```

```
##
##  Shapiro-Wilk normality test
##
## data:  FO$count
## W = 0.51224, p-value = 1.124e-07
```

```
shapiro.test(TR$count) #Normality Test for distribution of number of accident
s due to TR
```

```
##
##  Shapiro-Wilk normality test
##
## data:  TR$count
## W = 0.44161, p-value = 2.483e-08
```

```
shapiro.test(FA$count) #Normality Test for distribution of number of accident
s due to FA

##
##  Shapiro-Wilk normality test
##
## data:  FA$count
## W = 0.3236, p-value = 2.573e-09

shapiro.test(df$count) #Normality Test for the Entire Data set

##
##  Shapiro-Wilk normality test
##
## data:  df$count
## W = 0.56602, p-value < 2.2e-16
```

#Normality Tes Results

```
#The p-values were less than 0.05, hence, the statistical non-parametric test
will be used
```

#Kruskal-Wallis Rank

```
kruskal.test(count ~ variables, data = df)

##
##  Kruskal-Wallis rank sum test
##
## data:  count by variables
## Kruskal-Wallis chi-squared = 47.41, df = 6, p-value = 1.55e-08
```

#Wilcox Signed Rank Tests

```
 df %>%wilcox_test(count ~ variables, p.adjust.method = "bonferroni")

## # A tibble: 21 x 9
##    .y.   group1 group2    n1    n2 statistic           p     p.adj p.adj.s
ignif
##  * <chr> <chr>  <chr>  <int> <int>     <dbl>       <dbl>     <dbl> <chr>
##  1 count AW     FA        23    23     299   0.23        1         ns
##  2 count AW     FD        23    23     239   0.472       1         ns
##  3 count AW     FO        23    23     264.  1           1         ns
##  4 count AW     MF        23    23     236.  0.43        1         ns
##  5 count AW     PE        23    23      73.5 0.00000741  0.000156  ***
##  6 count AW     TR        23    23     271   0.849       1         ns
##  7 count FA     FD        23    23     206   0.064       1         ns
##  8 count FA     FO        23    23     230   0.23        1         ns
##  9 count FA     MF        23    23     205   0.06        1         ns
## 10 count FA     PE        23    23      57   0.000000609 0.0000128 ****
## # ... with 11 more rows
```

#Concluision

```
#Kindly refer to the final paper for results interpretation
```

——————————————————THE END————————————————————————-