

D. Tenisha

20CSEG34

Msc Data Analytics

dplyr_functions

D. Tenisha 20CSEG34

29/04/2021

```
library(readxl)

## Warning: package 'readxl' was built under R version 4.0.5

data_accident<-read_excel("tamilnadu_accident_data.xlsx")
View(data_accident)
head(data_accident)

## # A tibble: 6 x 5
##   Sl.No `City/District` `Total No of Acci~` `No of Persons In~` `No of Pers
ons K~
##   <chr> <chr>          <dbl>          <dbl>
<dbl>
## 1 1      Chennai City      7486          7349
1183
## 2 2      Coimbatore City    1354          1199
288
## 3 3      Madurai City      946           926
222
## 4 4      Salem City      1255          1345
215
## 5 5      Tirunelveli City    389           320
102
## 6 6      Tiruppur City      804           741
174

df<-data.frame(data_accident)
View(df)

##eliminating the last row which contains totals

df2<-df[-c(41),]
View(df2)
```

##filter()

```
filter(df2,Total.No.of.Accidents > 1000)%>%head()
```

```
##   Sl.No   City.District Total.No.of.Accidents No.of.Persons.Injured
## 1     1   Chennai City           7486           7349
## 2     2 Coimbatore City           1354           1199
## 3     4     Salem City           1255           1345
## 4     9     Coimbatore           2607           2523
## 5    10     Cuddalore           3758           4965
## 6    11     Dharmapuri           1602           2086
##   No.of.Persons.Killed
## 1                   1183
## 2                   288
## 3                   215
## 4                   759
## 5                   570
## 6                   371
```

```
filter(df2,Total.No.of.Accidents > 5000)%>%head()
```

```
##   Sl.No City.District Total.No.of.Accidents No.of.Persons.Injured
## 1     1   Chennai City           7486           7349
##   No.of.Persons.Killed
## 1                   1183
```

```
filter(df2,Total.No.of.Accidents > 1000)%>%head()
```

```
##   Sl.No   City.District Total.No.of.Accidents No.of.Persons.Injured
## 1     1   Chennai City           7486           7349
## 2     2 Coimbatore City           1354           1199
## 3     4     Salem City           1255           1345
## 4     9     Coimbatore           2607           2523
## 5    10     Cuddalore           3758           4965
## 6    11     Dharmapuri           1602           2086
##   No.of.Persons.Killed
## 1                   1183
## 2                   288
## 3                   215
## 4                   759
## 5                   570
## 6                   371
```

```
filter(df2,City.District == "Coimbatore City")%>%head()
```

```
##   Sl.No   City.District Total.No.of.Accidents No.of.Persons.Injured
## 1     2 Coimbatore City           1354           1199
##   No.of.Persons.Killed
## 1                   288
```

##select()

```
select(df2,City.District,No.of.Persons.Killed)%>% head()
```

```
##      City.District No.of.Persons.Killed
## 1      Chennai City                1183
## 2    Coimbatore City                 288
## 3      Madurai City                 222
## 4        Salem City                 215
## 5 Tirunelveli City                 102
## 6    Tiruppur City                  174
```

```
select(df2,"City.District":"No.of.Persons.Injured") %>% head()
```

```
##      City.District Total.No.of.Accidents No.of.Persons.Injured
## 1      Chennai City                7486                7349
## 2    Coimbatore City                1354                1199
## 3      Madurai City                 946                 926
## 4        Salem City                1255                1345
## 5 Tirunelveli City                 389                 320
## 6    Tiruppur City                  804                 741
```

```
select(df2,starts_with("N")) %>% head()
```

```
##      No.of.Persons.Injured No.of.Persons.Killed
## 1                    7349                1183
## 2                    1199                 288
## 3                     926                 222
## 4                   1345                 215
## 5                     320                 102
## 6                     741                 174
```

##mutate()

```
mutate(df2,diff_injured_and_killed = No.of.Persons.Injured-No.of.Persons.Killed)%>% head()
```

```
##      Sl.No      City.District Total.No.of.Accidents No.of.Persons.Injured
## 1         1      Chennai City                7486                7349
## 2         2    Coimbatore City                1354                1199
## 3         3      Madurai City                 946                 926
## 4         4        Salem City                1255                1345
## 5         5 Tirunelveli City                 389                 320
## 6         6    Tiruppur City                  804                 741
##      No.of.Persons.Killed diff_injured_and_killed
## 1                    1183                6166
## 2                     288                 911
## 3                     222                 704
## 4                     215                1130
## 5                     102                 218
## 6                     174                 567
```

```
mutate(df2,diff_Accidents_and_killed = Total.No.of.Accidents-No.of.Persons.Killed)%>% head()
```

```
##   Sl.No   City.District Total.No.of.Accidents No.of.Persons.Injured
## 1     1     Chennai City           7486           7349
## 2     2 Coimbatore City           1354           1199
## 3     3     Madurai City            946            926
## 4     4       Salem City           1255           1345
## 5     5 Tirunelveli City            389            320
## 6     6   Tiruppur City             804            741
##   No.of.Persons.Killed diff_Accidents_and_killed
## 1                   1183                6303
## 2                   288                1066
## 3                   222                 724
## 4                   215                1040
## 5                   102                 287
## 6                   174                 630
```

```
mutate(df2,diff_injured_and_killed = No.of.Persons.Injured-No.of.Persons.Killed,df2,diff_Accidents_and_killed = Total.No.of.Accidents-No.of.Persons.Killed)%>% head()
```

```
##   Sl.No   City.District Total.No.of.Accidents No.of.Persons.Injured
## 1     1     Chennai City           7486           7349
## 2     2 Coimbatore City           1354           1199
## 3     3     Madurai City            946            926
## 4     4       Salem City           1255           1345
## 5     5 Tirunelveli City            389            320
## 6     6   Tiruppur City             804            741
##   No.of.Persons.Killed diff_injured_and_killed diff_Accidents_and_killed
## 1                   1183                6166                6303
## 2                   288                 911                1066
## 3                   222                 704                 724
## 4                   215                1130                1040
## 5                   102                 218                 287
## 6                   174                 567                 630
```

##rename()

```
rename(df2,no.of.injured = No.of.Persons.Injured,no.of.killed = No.of.Persons.Killed)%>%head()
```

```
##   Sl.No   City.District Total.No.of.Accidents no.of.injured no.of.killed
## 1     1     Chennai City           7486           7349           1183
## 2     2 Coimbatore City           1354           1199           288
## 3     3     Madurai City            946            926           222
## 4     4       Salem City           1255           1345           215
## 5     5 Tirunelveli City            389            320           102
## 6     6   Tiruppur City             804            741           174
```

##arrange()

```
arrange(df2, City.District)%>%head(10)
```

##	Sl.No	City.District	Total.No.of.Accidents	No.of.Persons.Injured
## 1	8	Ariyalur	618	803
## 2	1	Chennai City	7486	7349
## 3	9	Coimbatore	2607	2523
## 4	2	Coimbatore City	1354	1199
## 5	10	Cuddalore	3758	4965
## 6	11	Dharmapuri	1602	2086
## 7	12	Dindigul	2169	2623
## 8	13	Erode	2510	2800
## 9	14	Kancheepuram	3507	3632
## 10	15	Kanniyakumari	1243	1365
##	No.of.Persons.Killed			
## 1		162		
## 2		1183		
## 3		759		
## 4		288		
## 5		570		
## 6		371		
## 7		662		
## 8		588		
## 9		1057		
## 10		296		

```
arrange(df2, desc(City.District))%>%head(10)
```

##	Sl.No	City.District	Total.No.of.Accidents	No.of.Persons.Injured
## 1	38	Virudhunagar	1328	1554
## 2	37	Villupuram	3914	5467
## 3	36	Vellore	3245	3727
## 4	35	Tuticorin	1568	1808
## 5	7	Trichy City	657	732
## 6	34	Trichy	1735	2568
## 7	33	Tiruvavarur	1041	1204
## 8	32	Tiruvannamalai	1712	1984
## 9	31	Tiruvallur	1390	1506
## 10	6	Tiruppur City	804	741
##	No.of.Persons.Killed			
## 1		366		
## 2		915		
## 3		878		
## 4		392		
## 5		144		
## 6		567		
## 7		188		
## 8		478		
## 9		454		
## 10		174		

```
arrange(df2,desc(No.of.Persons.Killed))%>%head(10)
```

##	Sl.No	City.District	Total.No.of.Accidents	No.of.Persons.Injured
## 1	1	Chennai City	7486	7349
## 2	14	Kancheepuram	3507	3632
## 3	37	Villupuram	3914	5467
## 4	36	Vellore	3245	3727
## 5	9	Coimbatore	2607	2523
## 6	30	Tiruppur	2577	3219
## 7	12	Dindigul	2169	2623
## 8	17	Krishnagiri	1969	2280
## 9	13	Erode	2510	2800
## 10	25	Salem	2505	2683

##	No.of.Persons.Killed
## 1	1183
## 2	1057
## 3	915
## 4	878
## 5	759
## 6	753
## 7	662
## 8	624
## 9	588
## 10	582

```
arrange(df2,desc(No.of.Persons.Killed))%>%head(10)
```

##	Sl.No	City.District	Total.No.of.Accidents	No.of.Persons.Injured
## 1	1	Chennai City	7486	7349
## 2	14	Kancheepuram	3507	3632
## 3	37	Villupuram	3914	5467
## 4	36	Vellore	3245	3727
## 5	9	Coimbatore	2607	2523
## 6	30	Tiruppur	2577	3219
## 7	12	Dindigul	2169	2623
## 8	17	Krishnagiri	1969	2280
## 9	13	Erode	2510	2800
## 10	25	Salem	2505	2683

##	No.of.Persons.Killed
## 1	1183
## 2	1057
## 3	915
## 4	878
## 5	759
## 6	753
## 7	662
## 8	624
## 9	588
## 10	582

```
arrange(df2,No.of.Persons.Killed)%>%head(10)
```

```
##      Sl.No      City.District Total.No.of.Accidents No.of.Persons.Injure
d
## 1      40  Railway Police,Trichy                2
2
## 2      39  Railway Police,Chennai                1
0
## 3      21      The Nilgiris                308                50
4
## 4       5  Tirunelveli City                389                32
0
## 5       7      Trichy City                657                73
2
## 6       8      Ariyalur                618                80
3
## 7       6  Tiruppur City                804                74
1
## 8      22      Perambalur                584                96
3
## 9      33      Tiruvarur               1041               120
4
## 10     4      Salem City               1255               134
5
##      No.of.Persons.Killed
## 1                      0
## 2                      1
## 3                     55
## 4                    102
## 5                    144
## 6                    162
## 7                    174
## 8                    184
## 9                    188
## 10                   215
```

```
##summarize()
```

```
summarize(df2,mean_totalaccidents = mean(Total.No.of.Accidents))
```

```
##      mean_totalaccidents
## 1                1785.775
```

```
summarize(df2,mean_totalaccidents = mean(Total.No.of.Accidents),mean_persons_
injured = mean(No.of.Persons.Injured))
```

```
##      mean_totalaccidents mean_persons_injured
## 1                1785.775                2054.075
```