Data Management

Assignment 1

Background:

The data for analysis is an insurance sector data in which premiums information is provided for each policy holder for all the regions and zones.

QUESTIONS-

1. Import Premiums data and name it as Premium.
2. Check number of rows, columns in the data.
3. Display first 10 rows and last 5 rows.
4. Describe (summarize) all variables.
5. Display top 5 and bottom 5 policies in terms of premium amount.
6. Calculate the sum for variable 'Sum_Assured' by 'Region' variable.
7. Create a subset of policies of Asia Standard Plan with Sum_Assured < = 50,000. Keep variables Policy_No, Zone_name, Plan and Sum_Assured in the subset data.
8. Export the subsetted data into an xlsx file.

Data Management

Assignment Solution Sample 1

```
#Q1. Import Premiums data and name it as Premium
##A.
Premium<-read.csv(file.choose(),header=TRUE)


#Q2. Check number of rows, columns in the data
##A.
nrow(Premium)
ncol(Premium)


#Q3.  Display first 10 rows and last 5 rows
##A.
```

```
head(Premium,n=10)
tail(Premium,n=5)


#Q4. Describe (summarize) all variables
##A.
summary(Premium)


#Q5. Display top 5 and bottom 5 policies in terms of premium amount
##A.
newpremium<-Premium[order(-Premium$Premium),]
head(newpremium)
tail(newpremium)


#Q6. Calculate the sum for variable 'Sum_Assured' by 'Region' variable
##A.
agg<-aggregate(Sum_Assured~REGION,FUN=sum,data = Premium)
head(agg)


#Q7. Create a subset of policies of Asia Standard Plan with Sum_Assured < =
50,000. Keep variables Policy_No, Zone_name, Plan and Sum_Assured in the
subset data.
##A.
ASP<-subset(Premium,Plan=="Asia Standard Plan" &
Sum_Assured<=50000,select=c(POLICY_NO,ZONE_NAME,Plan,Sum_Assured))
head(ASP)


#Q8. Export the data into an xlsx file
##A.
install.packages("openxlsx")
library("openxlsx")
write.xlsx(ASP, "file.xlsx",row.names = F)
```