

## **Business Analytics Session 5 Assignment 1**

**Q1 Consider only those participants who have all the data points ?**

Answer 1 #Excluding Missing Values from Analysis using complete.cases ,

```
> dim(olympic_data) # it shows 8618 entries
>olympic_data<- olympic_data[complete.cases(olympic_data),]
>olympic_data
>dim(olympic_data) #now it has 8613 entries as the NA values have been removed
```

**# Q2 Rank the participants in terms : .**

# i) Swimming

# ii) Table Tennis

# iii) Shooting

# iv) Gymnastics

# v) Total Medal

**Answer 2** #We are creating subset data frames from the parent file ,these subsets only have the above listed sport includes and after the creation of these , we will rank them

### **Part i)**

```
>olympicdata_swim<- subset(olympic_data,Sport=="Swimming") # this command will create a new
subset in which only swimming sport is mentioned
>View(olympicdata_swim)
>rank_in_swimming<-olympicdata_swim[order(olympicdata_swim$`Total Medals`, decreasing =
TRUE),] #creating the new data frame which is ranked wrt total medals in swimming game
>View(rank_in_swimming)
```

### **Part ii)**

```
>olympicdata_table<-subset(olympic_data,Sport == "Table Tennis") # this command will create a new
subset in which only Table Tennis sport is mentioned
>View(olympicdata_table)
>rank_in_tabletennis<-olympicdata_table[order(olympicdata_table$`Total Medals`, decreasing =
TRUE),] #creating the new data frame which is ranked wrt total medals in swimming game
>View(rank_in_tabletennis)
```

### **Part iii)**

```
>olympicdata_shoot<- subset(olympic_data,Sport=="Shooting") # this command will create a new
subset in which only Shooting sport is mentioned
>View(olympicdata_shoot)
>rank_in_shooting<-olympicdata_shoot[order(olympicdata_shoot$`Total Medals`, decreasing =
TRUE),] #creating the new data frame which is ranked wrt total medals in swimming game
>View(rank_in_shooting)
```

**Part iv)**

```
>olympicdata_gym<-subset(olympic_data,Sport == "Gymnastics") # this command will create a new subset in which only Gymnastics sport is mentioned  
>View(olympicdata_gym)  
>rank_in_gymnastics<-olympicdata_gym[order(olympicdata_gym$`Total Medals`, decreasing = TRUE),] #creating the new data frame which is ranked wrt total medals in swimming game  
>View(rank_in_gymnastics)
```

**Part V)**

```
>new_olympic_data<-olympic_data  
>new_olympic_data  
>rank_totalmedals<-new_olympic_data[order(new_olympic_data$`Total Medals`,decreasing = TRUE),] # ranking wrt to total medals  
>rank_totalmedals
```

**# Q3 Rank the Categories in terms of Age.(Higher the Age,Higher the Rank)**

# Answer 3

```
>highrank_moreage<-olympic_data[order(olympic_data$Age, decreasing = TRUE) , ]  
>highrank_moreage
```

**# Q4 Identify Year wise top participants in terms of :**

#i) Swimming

#ii) Table Tennis

#iii) Shooting

#iv) Gymnastics

#v) Total Medal

#Answer 4 #we will use subset and !duplicated to find

**Part i)**

```
>yearwise_topswim<-subset(olympicdata_swim , !duplicated(Year)) # year wise top in swimming  
based on total medals  
> yearwise_topswim
```

**Part ii)**

```
> yearwise_toptable<-subset(olympicdata_table , !duplicated(Year)) # year wise top in table tennis  
based on total medals  
>yearwise_toptable
```

**Part iii)**

```
>yearwise_topshooting<-subset(olympicdata_shoot , !duplicated(Year)) # year wise top in shooting  
based on total medals  
>yearwise_topshooting
```

**Part iv)**

```
>yearwise_topgymnastics<-subset(olympicdata_gym , !duplicated(Year)) # year wise top in  
gymnastics based on total medals  
>yearwise_topgymnastics
```

**Part v)**

```
>yearwise_topoverall<-subset(olympic_data , !duplicated(Year)) #yearwise top in all the sports  
combined based on total medals  
yearwise_topoverall
```