Question 1 and 2

Code

Find out the Top 5 cities by performance (Higher CTR and Lower CPC. Impressions and Clicks should be reasonable.)

In order to solve this query sqldf library was used. Which helps in writting queries efficiently in R.

Assumptions made:

Best Cities by performance were measured by bestindicator which shows the ratio betwen impression got clicked and the cost associated with it.

Output:

Top 5 cities by performance are:

-Claxton

-Altha

-Port Hueneme

-Dover-Foxcroft

-York Beach

For Clicks:

1. Sum of Clicks= 0 were not considered reasonable
2. Min. 1st Qu. Median Mean 3rd Qu. Max.

   1.000 1.000 1.000 2.844 2.000 190.000
3. Based on the click summary statistics, clicks>= mean were considered reasonable.

Hide

```
Data<-read.csv('D:/Rutgers Study Material/MultivariateData1.csv')
#Query:1
library(sqldf)
a<-sqldf('select City, Sum(Impressions) as Imp,Sum(Clicks) as Clk, Sum(Clicks * CPC) as totalcos
t, Sum(Clicks) * 1.000/Sum(impressions) as CTR, Sum(Clicks * CPC)/sum(Clicks) as TotalCPC, (Sum
(Clicks) * 1.000/Sum(impressions))/(Sum(Clicks * CPC)/sum(Clicks)) as bestindicator  from Data
 where impressions>=0 and Clicks>=0 group by City')
summary(a$Clk)
```

```
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
  0.0000   0.0000   0.0000   0.6288   0.0000 190.0000
```

Hide

```
b<-sqldf('select City,clk,CTR,TotalCPC,bestindicator from a where clk>=2.844 group by City order
 by bestindicator desc')
head(b)
```

	City <fctr>	Clk <int>	CTR <dbl>	TotalCPC <dbl>	bestindicator <dbl>
1	Claxton	4	0.57142857	0.000662651	862.3371
2	Altha	4	0.33333333	0.000421687	790.4757
3	Port Hueneme	3	0.23076923	0.000662651	348.2515
4	Dover-Foxcroft	3	0.11111111	0.000421687	263.4919
5	York Beach	3	0.30000000	0.002469880	121.4634
6	Jerome	3	0.07692308	0.000662651	116.0838

6 rows

Which Device has the highest and poorest performance ?

Assumptions made:

Performance of the device is measured by higher CTR and lower CPC with reasonable clicks and impression.

Output:

As per the ‘bestindicator’ we can say Roku performed the best but the impressions and clicks doesn’t seem reasonable.

Best device by performance: Mobile

Worst device by performance: Tablet

Hide

```
#Query: 2
a<-sqldf('select DeviceType, Sum(Impressions) as Imp,Sum(Clicks) as Clk, Sum(Clicks * CPC) as to
talcost, Sum(Clicks) * 1.000/Sum(impressions) as CTR, Sum(Clicks * CPC)/sum(Clicks) as TotalCPC,
 (Sum(Clicks) * 1.000/Sum(impressions))/(Sum(Clicks * CPC)/sum(Clicks)) as bestindicator  from D
ata  where impressions>=0 and Clicks>=0 group by DeviceType')
b<-sqldf('select DeviceType,clk,CTR,TotalCPC,bestindicator from a group by DeviceType order by b
estindicator desc')
head(b)
```

	DeviceType <fctr>	Clk <int>	CTR <dbl>	TotalCPC <dbl>	bestindicator <dbl>
1	Roku	8	1.000000000	0.000662651	1509.0900036
2	Mobile	741	0.008085459	0.009477068	0.8531604
3	PC	1919	0.007410869	0.013048108	0.5679650

	DeviceType <fctr>	Clk <int>	CTR <dbl>	TotalCPC <dbl>	bestindicator <dbl>
4	Tablet	568	0.003837553	0.009627560	0.3986007

4 rows