# Data Challenge Solutions – (in R)

library("dplyr", lib.loc="~/R/win-library/3.2")

library("lubridate", lib.loc="~/R/win-library/3.2")

library("magrittr", lib.loc="~/R/win-library/3.2")

library("tidyr", lib.loc="~/R/win-library/3.2")

library("data.table", lib.loc="~/R/win-library/3.2")

library("stringr", lib.loc="~/R/win-library/3.2")

library("NISTunits", lib.loc="~/R/win-library/3.2")

options(digits=10)

files <- list.files(path = "C:/Users/param/Desktop/4datainc/prec\_data",pattern = ".csv")

combo<-rbindlist(lapply(files,fread,sep=","))

#1

combo %>% group\_by(TypeText) %>% summarise(cnt=n()) %>% mutate (fraction = cnt/sum(cnt)) %>% arrange(desc(cnt))

#2

TRecord<-combo %>% select(NOPD\_Item,TimeDispatch,TimeArrive,PoliceDistrict)

TRecord$TimeDispatch<-as.POSIXct(as.character(TRecord$TimeDispatch),format="%d/%m/%Y %H:%M:%S")

TRecord$TimeArrive<-as.POSIXct(as.character(TRecord$TimeArriv),format="%d/%m/%Y %H:%M:%S")

TRecord <- TRecord %>% filter(!is.na(TimeArrive),!is.na(TimeDispatch)) %>% mutate(tdiff = TimeArrive - TimeDispatch)

TRecord %>% filter(tdiff>0) %>% summarise(median\_responsetime = median(tdiff))

#3

AR<-TRecord %>% group\_by(PoliceDistrict) %>% filter(tdiff>0) %>% summarise(AvgResponse = mean(tdiff)) %>% arrange(AvgResponse)

max(AR$AvgResponse) - min(AR$AvgResponse)

#4

AR<-combo %>% group\_by(TypeText) %>% filter(n()>100,PoliceDistrict!=0) %>% summarise(pe\_count = n())

AR<-AR %>% mutate(ucprob\_event=pe\_count/nrow(AR))

BR<-combo %>% group\_by(PoliceDistrict,TypeText) %>% filter(PoliceDistrict != 0 ) %>% summarise(cevent=n())

PRecord<-inner\_join(BR,AR)

PRecord<-PRecord%>% mutate(comb\_prob\_event=cevent/nrow(PRecord))

CR<-PRecord %>% group\_by(PoliceDistrict) %>% summarise(p\_district = n()/nrow(PRecord))

PRecord <- inner\_join(PRecord,CR)

PRecord<-PRecord%>% mutate(Cond\_prob = comb\_prob\_event/p\_district,Ratio=Cond\_prob/ucprob\_event) %>% arrange(desc(Ratio))

max(PRecord$Ratio)

#5

combo <- combo %>% mutate(year=str\_sub(combo$NOPD\_Item,-2,-1))

Year1 <- combo %>% filter(year==11) %>% group\_by(TypeText) %>% summarise(cvol11 = n())

Year2 <- combo %>% filter(year==15) %>% group\_by(TypeText) %>% summarise(cvol15 = n())

CRecord<-inner\_join(Year1,Year2)

CRecord %>% mutate(cdecrease=(cvol15-cvol11)/cvol11) %>% arrange(cdecrease)

#6

DR<-combo %>% select(Disposition,TimeCreate)

DR<-DR %>% mutate(Date = as.POSIXct(as.character(TimeCreate),format="%d/%m/%Y"),TimeCreate = as.POSIXct(as.character(TimeCreate),format="%d/%m/%Y %H:%M:%S"))

DR <- DR %>% filter(strftime(Date)=="2015-01-01") %>% mutate(Hour = as.integer(TimeCreate - Date))

s1 <-DR %>% group\_by(Disposition,Hour) %>% summarise(cases=n())

s2<-DR %>% group\_by(Hour) %>% summarise(count=n())

DR <- full\_join(s1,s2)

DR %>% group\_by(Disposition) %>% summarise(variation = var(cases/count)) %>% arrange(desc(variation))

DR %>% filter(Disposition=='UNF') %>% summarise(max(cases/count)-min(cases/count))

#7

areac<-combo%>% select(PoliceDistrict,Location)%>%filter(PoliceDistrict!=0)

areac<-areac%>% separate(Location,c("Lat","Lon"),",")

areac<-areac%>% mutate(Latitude=str\_replace(Lat,"\\(",""),Longitude=str\_replace(Lon,"\\)",""))

areac %>% filter((Latitude >20 & Latitude<35)) %>% group\_by(PoliceDistrict) %>%

summarise\_each(funs(sd),Latitude,Longitude) %>%

mutate(ellipse\_area=pi\*Latitude\*110.574\*Longitude\*111.32\*cos(NISTdegTOradian(Latitude))) %>%

arrange(desc(ellipse\_area))

#8

P\_record<-combo %>% select(TypeText,Priority)

PR<-P\_record %>% group\_by(TypeText,Priority) %>% summarise(num=n()) %>% top\_n(1,num) %>% arrange(desc(num))

DR<-P\_record %>% group\_by(TypeText) %>% summarise(p\_event = n())

CallP <- full\_join(PR,DR) %>% mutate(pratio = num/p\_event) %>% arrange(pratio)

min(CallP$pratio)