Final exam – take home

Krishna Sindhuja Kalusani

PART-1

```
map11 = function(k,flights){
if(!(is.nan(flights[[20]])&&is.nan(flights[[21]]))){
return(keyval(as.character(flights[[17]]),(flights[20]+flights[21])))}}

reduce11 = function(Origin, meanTaxi){
keyval(Origin, mean(meanTaxi,na.rm=TRUE))}

 mr11= function(input, output = NULL) { mapreduce(input = input, output = output, input.for
mat = make.input.format("csv", sep=","), map = map11, reduce = reduce11)}

options(max.print=99999)
hdfs.root = '/user/share/student'
hdfs.data = file.path(hdfs.root,'wholeEnchilada.csv')
hdfs.out = file.path(hdfs.root,'mean_time2')
mean_time2 = mr11(hdfs.data, hdfs.out)
sink('P1_output.txt')
results.df
sink()
```

PART-2

```
map22 = function(k,flights){
if(!(is.nan(flights[[9]])&&is.nan(flights[[17]])&&is.nan(flights[[20]])&&is.nan(flights[[21]]))){
return ( keyval(paste(as.character(flights[[9]]),as.character(flights[[17]]),sep="-"),(as.numeri
c(flights[[20]]+flights[[21]])))) }}

reduce22 = function(Car_Ori, min_max_taxi){
keyval(Car_Ori, c(min(min_max_taxi,na.rm=TRUE),max(min_max_taxi,na.rm=TRUE)))}

mr22= function(input, output = NULL) { mapreduce(input = input, output = output, input.for
mat = make.input.format("csv", sep=","), map = map22, reduce = reduce22)}

hdfs.root = '/user/share/student'
hdfs.data = file.path(hdfs.root,'wholeEnchilada.csv')
hdfs.out = file.path(hdfs.root,'minMax_taxi')
minMax_time = mr22(hdfs.data, hdfs.out)
results = from.dfs(minMax_time)
results.df = as.data.frame(results, stringsAsFactors=F)
colnames(results.df) = c('Airline/origin', 'Min/Max_Taxi')
sink('P2_output.txt')
results.df
sink()
```

PART-3

```
map33 = function(k,flights){
  return ( keyval(paste(as.character(flights[[9]]),apply(flights[,[17:18]],1,min),apply(flights[,17:18],1,max),sep="-"),(as.numeric(flights[[20]]+flights[[21]]))) }

reduce33 = function(airline_market, min_man_taxi){
  keyval(airline_market, c(min(min_max_taxi,na.rm=TRUE),max(min_max_taxi,na.rm=TRUE)))}

mr33= function(input, output = NULL) { mapreduce(input = input, output = output, input.format = make.input.format("csv", sep=","), map = map33, reduce = reduce33)}

hdfs.root = '/user/share/student'
hdfs.data = file.path(hdfs.root,'wholeEnchilada.csv')
hdfs.out = file.path(hdfs.root,'market_minMax')
market_minMax = mr22(hdfs.data, hdfs.out)
results = from.dfs(market_minMax)
results.df = as.data.frame(results, stringsAsFactors=F)
colnames(results.df) = c('Airline market', 'Market_Min/Max')
sink('P3_output.txt')
results.df
sink()
```