ANA 515 Assignment 2

Vignesh Raja Nallendran

2024-06-15

The data set presented in this report is a collection of radio data from 2014. It contains a week's worth of songs played across 25 radio stations that play rock music. This data was collected to answer an analysis on what categorizes a song as classic rock because over time the definition has changed, we hope to achieve this analysis by computing the frequency of songs played by rock bands from the 1960s to early 2000s which is considered the tail end of the rock era. The data set is stored as a CSV file, delimited by a comma.

```
cr_path<- "C:/Users/atl-vigneshn/Desktop/mcdaniel/ANA515/classic-rock-raw-data.csv"
cr<-read.csv(cr_path)

#cleaning, time column is numeric and according to the raw data this is in numeric format an
cr$date<-as.POSIXct("1970-01-01 00:00:00", format = "%Y-%m-%d %H:%M:%S")
STAT_TIME<-as.POSIXct("1970-01-01 00:00:00", format = "%Y-%m-%d %H:%M:%S")
cr$newdate<-NA
cr$newdate <- cr$date+ cr$TIME #new date is the column we need
cr_new<- cr[c("Song.Clean","ARTIST.CLEAN","CALLSIGN","COMBINED","newdate","First.","UNIQUE_I
cr_new$count <- 1

#GROUP raw data to get required analysis
group_cr<-cr_new %>% group_by(ARTIST.CLEAN,Song.Clean) %>% summarise(times_song_played = sum

#change column name
colnames(group_cr)<- c('Artist','Song','Frequency','First Played','Last Played')
#after cleaning and grouping the required data is final_df/group_cr
#we will be using group_cr for the reminder of our summary and descriptive analysis
final_df<-group_cr
summary<-summary(group_cr[c("First Played","Frequency","Last Played")])
ncol<-ncol(group_cr) #number of columns
nrow<-nrow(group_cr) #number of rows
```

This dataframe has 2231 rows and 5 columns. The names of the columns and a brief description of each are in the table below:

```
#This next code chunk is to make a description table explaining the columns in our final dat
col_name<- c('Artist','Song','Frequency','First Played','Last Played')
desc_col<- c('Name of the Artist', 'Name of the Song played', 'The number of times the song
desc<-data.frame(col_name,desc_col)
```

```
colnames(desc)<-c('Column Name','Column Description')

desc %>%
  kbl(caption = "Dataset Description", bold = T) %>%
  kable_paper(c("striped", "hover"),full_width = F) %>%
  row_spec(0,bold = T, color = "white", background = "black",align = 'c') %>%
  column_spec(1, bold = T)
```

Dataset Description

Column Name	Column Description
Artist	Name of the Artist
Song	Name of the Song played
Frequency	The number of times the song was played
First Played	The First time the song was played during the sampling process
Last Played	The Last time the song was played during the sampling process

As shown in the code above, the summary of columns Frequency, First Played, and Last Played are stored in the summary variable. Attached below is the summary elements presented in a table format, note since we used the "sum(is.na())" function to check the null values for the three columns and there are no null values we will ouput the null count as NA in the summary table below

```
# creating our own summary table
sum_name<- c('Max Value','Min Value','Mean Value','Null count')
freq_values<-c(max(group_cr$Frequency),min(group_cr$`Frequency`),mean(group_cr$`Frequency`),
fp_values<-c(max(group_cr$`First Played`),min(group_cr$`First Played`),mean(group_cr$`First
lp_values<-c(max(group_cr$`Last Played`),min(group_cr$`Last Played`),mean(group_cr$`Last Pla

LP_NULL_VALUES<-sum(is.na(group_cr$`Last Played`)) # number of missing values for last playe
FP_NULL_VALUES<-sum(is.na(group_cr$`First Played`)) # number of missing values for first pla

summary_table<-data.frame(sum_name,freq_values,fp_values,lp_values) # created summary table
colnames(summary_table)<-c('Parameter','Frequency','First Played','Last Played')

#Outputing in presentable format
summary_table %>%
  kbl(caption = "Dataset Summary", bold = T) %>%
  kable_paper(c("striped", "hover"),full_width = F) %>%
  row_spec(0,bold = T, color = "white", background = "black",align = 'c') %>%
  column_spec(1, bold = T)
```

Parameter	Frequency	First Played	Last Played
Max Value	142.00000	2014-06-23 00:59:10	2014-06-23 00:59:19
Min Value	1.00000	2014-06-16 01:28:14	2014-06-16 01:59:23
Mean Value	16.88615	2014-06-17 18:54:44	2014-06-21 11:23:07
Null count	NA	NA	NA

Dataset Summary