

Arizona vs. Nevada

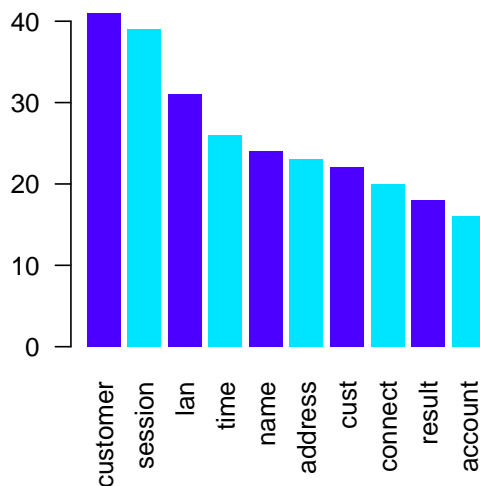
Fauzy Che Yayah

November 14, 2015

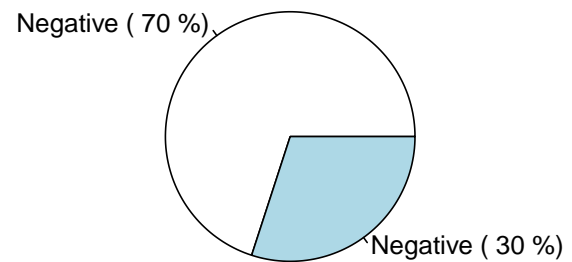
```
library(RODBC)
conn <- odbcConnect(dsn="nova", uid="", pwd="")
#sqlTables(channel)
queryResult <- sqlQuery(conn, "select * from nova.nova_trouble_ticket where length(description) > 15 limit 100")
odbcClose(conn)
#dim(queryResult)
#head(queryResult)
```

Arizona

Top 10 most frequent terms



Emotion Polarity of Review Messages



Nevada

```
# library(RODBC)
# conn <- odbcConnect(dsn="yelp", uid="", pwd="")
# #sqlTables(channel)
# queryResult <- sqlQuery(conn, "select * from consolidated_review4 where state = 'NV' limit 2000 ")
# odbcClose(conn)
# #dim(queryResult)
# #head(queryResult)
# ```
#
#
```

```

#
# ```{r echo=FALSE , fig.align='center' , fig.height=4 , fig.width=8 , message=F, warning=F}
# df_review <- as.data.frame(queryResult$txt)
# df_review <- as.data.frame(tolower(df_review$review))
# df_review <-as.data.frame(sapply(df_review,gsub,pattern="[:digit:]",replacement=""))
# df_review <-as.data.frame(sapply(df_review,gsub,pattern="[:punct:]",replacement=""))
# df_review <-as.data.frame(sapply(df_review,gsub,pattern="@\\w+",replacement=""))
# df_review <-as.data.frame(sapply(df_review,gsub,pattern="^\\s+|\\s+$",replacement=""))
# df_review <-as.data.frame(sapply(df_review,gsub,pattern="[ \\t]{2,}",replacement=""))
# names(df_review) = "review"
# #words_list = strsplit(as.character(df_review$review), " ")
#
# mach_corpus = Corpus(VectorSource(df_review$review))
#
# # create document term matrix applying some transformations
# tdm = TermDocumentMatrix(
# mach_corpus,
# control = list(
# removePunctuation = TRUE,
# stopwords = c(stopwords("english")),
# removeNumbers = TRUE, tolower = TRUE
# )
# )
# m <- as.matrix(tdm)
# v <- sort(rowSums(m), decreasing = TRUE)
#
# class_emo = classify_emotion(df_review$review[1:10], algorithm = "bayes", prior =
# 1.0)
# emotion = class_emo[,7]
# emotion[is.na(emotion)] = "unknown"
# class_pol = classify_polarity(df_review$review[1:10], algorithm = "bayes")
# # get polarity best fit
# polarity = class_pol[,4]
# pct <-
# round(table(as.data.frame(class_pol)$BEST_FIT) / sum(table(as.data.frame(class_pol)$BEST_FIT)) *
# 100)
# lbl <- as.data.frame(class_pol)$BEST_FIT
#
# par(mfrow = c(1,2))
# barplot(
# head(v, 10), border = NA, las = 2, main = "Top 10 most frequent terms", cex.main =
# 1 , col = topo.colors(2)
# )
# pie(
# table(as.data.frame(class_pol)$BEST_FIT) / length(as.data.frame(class_pol)$BEST_FIT), labels = paste(
# "Emotion Polarity of Review Messages "
# )
# )
#
# ```
#
# ```{r echo=FALSE , fig.height=3 , fig.align='center' , message=F, warning=F}
# df_review <- as.data.frame(queryResult$txt)
# df_review <-as.data.frame(sapply(df_review,gsub,pattern="[:digit:]",replacement=""))

```

```

# df_review <-as.data.frame(sapply(df_review,gsub,pattern="[:punct:]",replacement=""))
# df_review <-as.data.frame(sapply(df_review,gsub,pattern="@\\w+",replacement=""))
# df_review <-as.data.frame(sapply(df_review,gsub,pattern="^\\s+|\\s+$",replacement=""))
# df_review <-as.data.frame(sapply(df_review,gsub,pattern="[ \\t]{2,}",replacement=""))
# names(df_review) = "review"
#
# # create a corpus
# mach_corpus = Corpus(VectorSource(df_review))
#
# # create document term matrix applying some transformations
# tdm = TermDocumentMatrix(mach_corpus,
#   control = list(removePunctuation = TRUE,
#     stopwords = c("machine", "learning", stopwords("english")),
#     removeNumbers = TRUE, tolower = TRUE))
#
# # define tdm as matrix
# m = as.matrix(tdm)
# # get word counts in decreasing order
# word_freqs = sort(rowSums(m), decreasing=TRUE)
# # create a data frame with words and their frequencies
# dm = data.frame(word=names(word_freqs), freq=word_freqs)
#
# suppressWarnings(suppressMessages(wordcloud(dm$word[1:500], dm$freq[1:500], random.order=FALSE, color=

```