Project_1_part1_620118149

620118149

3/9/2021

Project Overview

A large entertainment & bar franchise (Hard Knocks) would like to determine whether now is a good time to expand and which country/region of the world is best to open next. Ultimately, they would like to gain some insights on whether to open their next franchise , based on public sentiments generally and opportunities to maximize profits. They first want to test the pulse of persons in relation to the products they sell (beers) which would indicate the current public views towards these products. In addition, they have a large dataset with sales data by region and country for different types of consumer goods (household, cosmetics etc), food and beer You will be helping Hard Knocks to make their decision in this project!

---

1. Between March 0 4 March 0 9 (max. 1 word per day) connect to twitter on two separate days and retrieve 8 ,000 or more tweets containing one of the words from a and one of the words from b (total 16 ,000 tweets Retrieve tweets for the word from a on a separate day from b. a) "beer" or "beer" b) "concert" or "concert"

- Getting Tweets

```
#tweet.beer.full <- search_tweets("beer",n=9000,include_rts = FALSE)
#tweet.concert.full <- search_tweets("concert",n=9000,include_rts = FALSE)

# load in the tweets that where retrieved

tweets_data_names <- c("beverage_2021-03-10_phillip.csv", "beverage_2021-03-13_phillip.csv", "party_202

tweets_data_dirty_names <- c("beverage_dirty_2021-03-10_phillip.csv", "beverage_dirty_2021-03-13_philli

#retrieve the dirty data for reprocessing

bev1 <- read.csv(paste("Data/usable_data/beverage",tweets_data_dirty_names[1],sep = "/",collapse = NULL]

bev2 <- read.csv(paste("Data/usable_data/beverage",tweets_data_dirty_names[2],sep = "/",collapse = NULL]

party1 <- read.csv(paste("Data/usable_data/party",tweets_data_dirty_names[3],sep = "/",collapse = NULL)

party2 <- read.csv(paste("Data/usable_data/party",tweets_data_dirty_names[4],sep = "/",collapse = NULL)
```

```
beer1 <- read.csv(paste("Data/usable_data/beer",tweets_data_dirty_names[5],sep = "/",collapse = NULL),

beer2 <- read.csv(paste("Data/usable_data/beer",tweets_data_dirty_names[6],sep = "/",collapse = NULL),

con1 <- read.csv(paste("Data/usable_data/concert",tweets_data_dirty_names[7],sep = "/",collapse = NULL)

con2 <- read.csv(paste("Data/usable_data/concert",tweets_data_dirty_names[8],sep = "/",collapse = NULL)
```

2.   For each set of tweets retrieved (a b above), retain the following features only:
text, screen_name, user_id, created_at, favourite_count, retweet_count, location, followers_count,
friends_count,account_lang, lang.

```
column.keep <- c("text", "screen_name", "user_id", "created_at", "favorite_count", "retweet_count", "lo

#tweet.beer = tweet.beer.full[column.keep]
#tweet.concert = tweet.concert.full[column.keep]

bev1.tweet = bev1[column.keep]

bev2.tweet = bev2[column.keep]

party1.tweet = party1[column.keep]

party2.tweet = party2[column.keep]

beer1.tweet = beer1[column.keep]

beer2.tweet = beer2[column.keep]

con1.tweet = con1[column.keep]

con2.tweet = con2[column.keep]

#The official data set was removed frombeing print bcause of unicode characters that are causing a issu

#kable(bev1.tweet[5:8,1:4], caption = paste("Beverage","table summary with date", tweets_data_dirty_nam
#kable(bev2.tweet[5:8,1:4], caption = paste("Beverage","table summary with date", tweets_data_dirty_nam
#kable(party1.tweet[5:8,1:4], caption = paste("Beverage","table summary with date", tweets_data_dirty_n
#kable(party2.tweet[5:8,1:4], caption = paste("Beverage","table summary with date", tweets_data_dirty_n
#kable(beer1.tweet[5:8,1:4], caption = paste("Beverage","table summary with date", tweets_data_dirty_na
#kable(beer2.tweet[5:8,1:4], caption = paste("Beverage","table summary with date", tweets_data_dirty_na
#kable(con1.tweet[5:8,1:4], caption = paste("Beverage","table summary with date", tweets_data_dirty_nam
#kable(con2.tweet[5:8,1:4], caption = paste("Beverage","table summary with date", tweets_data_dirty_nam
```

Clean The tweets

```
bev1.tweet$text <- gsub("https.*","",bev1.tweet$text)
bev1.tweet$text <- gsub("http.*","",bev1.tweet$text)
bev1.tweet$text <- gsub("#.*","",bev1.tweet$text)
bev1.tweet$text <- gsub("@.*","",bev1.tweet$text)
bev1.tweet$text <- gsub("[^[:alnum:][:blank:]?&/\\-]","",bev1.tweet$text)
bev1.tweet$text <- gsub("U00..","",bev1.tweet$text)
```

```r
bev1.tweet$text <- gsub("[^\x20-\x7E]","",bev1.tweet$text)

bev2.tweet$text <- gsub("https.*","",bev2.tweet$text)
bev2.tweet$text <- gsub("http.*","",bev2.tweet$text)
bev2.tweet$text <- gsub("#.*","",bev2.tweet$text)
bev2.tweet$text <- gsub("@.*","",bev2.tweet$text)
bev2.tweet$text <- gsub("[^[:alnum:][:blank:]?&/\\-]","",bev2.tweet$text)
bev2.tweet$text <- gsub("U00..","",bev2.tweet$text)
bev2.tweet$text <- gsub("[^\x20-\x7E]","",bev2.tweet$text)


party1.tweet$text <- gsub("https.*","",party1.tweet$text)
party1.tweet$text <- gsub("http.*","",party1.tweet$text)
party1.tweet$text <- gsub("#.*","",party1.tweet$text)
party1.tweet$text <- gsub("@.*","",party1.tweet$text)
party1.tweet$text <- gsub("[^[:alnum:][:blank:]?&/\\-]","",party1.tweet$text)
party1.tweet$text <- gsub("U00..","",party1.tweet$text)
party1.tweet$text <- gsub("[^\x20-\x7E]","",party1.tweet$text)


party2.tweet$text <- gsub("https.*","",party2.tweet$text)
party2.tweet$text <- gsub("http.*","",party2.tweet$text)
party2.tweet$text <- gsub("#.*","",party2.tweet$text)
party2.tweet$text <- gsub("@.*","",party2.tweet$text)
party2.tweet$text <- gsub("[^[:alnum:][:blank:]?&/\\-]","",party2.tweet$text)
party2.tweet$text <- gsub("U00..","",party2.tweet$text)
party2.tweet$text <- gsub("[^\x20-\x7E]","",party2.tweet$text)


beer1.tweet$text <- gsub("https.*","",beer1.tweet$text)
beer1.tweet$text <- gsub("http.*","",beer1.tweet$text)
beer1.tweet$text <- gsub("#.*","",beer1.tweet$text)
beer1.tweet$text <- gsub("@.*","",beer1.tweet$text)
beer1.tweet$text <- gsub("[^[:alnum:][:blank:]?&/\\-]","",beer1.tweet$text)
beer1.tweet$text <- gsub("U00..","",beer1.tweet$text)
beer1.tweet$text <- gsub("[^\x20-\x7E]","",beer1.tweet$text)


beer2.tweet$text <- gsub("https.*","",beer2.tweet$text)
beer2.tweet$text <- gsub("http.*","",beer2.tweet$text)
beer2.tweet$text <- gsub("#.*","",beer2.tweet$text)
beer2.tweet$text <- gsub("@.*","",beer2.tweet$text)
beer2.tweet$text <- gsub("[^[:alnum:][:blank:]?&/\\-]","",beer2.tweet$text)
beer2.tweet$text <- gsub("U00..","",beer2.tweet$text)
beer2.tweet$text <- gsub("[^\x20-\x7E]","",beer2.tweet$text)


con1.tweet$text <- gsub("https.*","",con1.tweet$text)
con1.tweet$text <- gsub("http.*","",con1.tweet$text)
con1.tweet$text <- gsub("#.*","",con1.tweet$text)
con1.tweet$text <- gsub("@.*","",con1.tweet$text)
con1.tweet$text <- gsub("[^[:alnum:][:blank:]?&/\\-]","",con1.tweet$text)
con1.tweet$text <- gsub("U00..","",con1.tweet$text)
```

```
con1.tweet$text <- gsub("[^\x20-\x7E]","",con1.tweet$text)


con2.tweet$text <- gsub("https.*","",con2.tweet$text)
con2.tweet$text <- gsub("http.*","",con2.tweet$text)
con2.tweet$text <- gsub("#.*","",con2.tweet$text)
con2.tweet$text <- gsub("@.*","",con2.tweet$text)
con2.tweet$text <- gsub("[^[:alnum:][:blank:]?&/\\-]","",con2.tweet$text)
con2.tweet$text <- gsub("U00..","",con2.tweet$text)
con2.tweet$text <- gsub("[^\x20-\x7E]","",con2.tweet$text)
```

- – The tables after the changes are made **

Table 1: Beverage table summary with date beverage_dirty_2021-03-10_phillip.csv

	text	screen_name	user_id	created_at
5		911newsnow1	1.274747e+18	2021-03-10 01:00:30
6		911newsnow1	1.274747e+18	2021-03-05 21:38:56
7		911newsnow1	1.274747e+18	2021-03-07 13:17:01
8		911newsnow1	1.274747e+18	2021-03-09 00:58:59

Table 2: Beverage table summary with date beverage_dirty_2021-03-13_phillip.csv

	text	screen_name	user_id	created_at
5	U3067U305FU3002U304CU3093U3053U3084U3002U3053U308CU306FU7F8EU5473U305FU306E U4ECAU8A71U984CU6CB8U9A30U4E2DU58F2U308AU5207U308CU7D9AU51FAU306EU4EBAU6C17U30E9U30FCU...	911newsnow1	1.506703e+18	2021-03-11 19:34:21
6	U3067U305FU3002U304CU3093U3053U3084U3002U3053U308CU306FU7F8EU5473U305FU306E U4ECAU8A71U984CU6CB8U9A30U4E2DU58F2U308AU5207U308CU7D9AU51FAU306EU4EBAU6C17U30E9U30FCU...	911newsnow1	1.506703e+18	2021-03-10 19:34:19
7		KingOfK.I.25	1.534902e+18	2021-03-13 22:33:25
8		MMPcomms	1.832960e+09	2021-03-09 22:33:11

Table 3: Beverage table summary with date party_dirty_2021-03-10_phillip.csv

	text	screen_name	user_id	created_at
5		911newsnow1	1.274747e+18	2021-03-10 01:00:30
6		911newsnow1	1.274747e+18	2021-03-05 21:38:56
7		911newsnow1	1.274747e+18	2021-03-07 13:17:01
8		911newsnow1	1.274747e+18	2021-03-09 00:58:59

Table 4: Beverage table summary with date party_dirty_2021-03-13_phillip.csv

	text	screen_name	user_id	created_at
5	U3067U305FU3002U304CU3093U3053U3084U3002U3053U308CU306FU7F8EU5473U305F... U306E U4ECAU8A71U984CU6CB8U9A30U4E2DU58F2U308AU5207U308CU7D9AU51FAU306EU4EBAU6C17...			2021-03-11 ...
6	U3067U305FU3002U304CU3093U3053U3084U3002U3053U308CU306FU7F8EU5473U305F... U306E U4ECAU8A71U984CU6CB8U9A30U4E2DU58F2U308AU5207U308CU7D9AU51FAU306EU4EBAU6C17...			2021-03-10 ...
7		KingOfK12534	1.2349e+...	2021-03-13 22:33:25
8		MMPcom	1.8329e+...	2021-03-09 ... 2021-03-13 22:33:11

Table 5: Beverage table summary with date beer_dirty_2021-03-14_phillip.csv

	text	screen_name	user_id	created_at
5		issaNizulove	1.329789e+18	2021-03-14 06:44:18
6		jennyessel	2.357521e+08	2021-03-14 06:50:38
7		notmeagain	1.754464e+07	2021-03-14 06:50:30
8	Ang butter caramel sa mcdo kay murag butter beer F914	brionesmariella	2.354616e+09	2021-03-14 06:50:30

Table 6: Beverage table summary with date beer_dirty_2021-03-16_phillip.csv

	text	screen_name	user_id	created_at
5		PremiumMalts_jp	3058851294	2021-03-16 08:43:01
6		PremiumMalts_jp	3058851294	2021-03-16 10:27:07
7		PremiumMalts_jp	3058851294	2021-03-16 11:00:51
8		PremiumMalts_jp	3058851294	2021-03-16 09:51:20

Table 7: Beverage table summary with date concert_dirty_2021-03-15_phillip.csv

	text	screen_name	user_id	created_at
5		ichiban_KIRIN	2691870948	2021-03-15 09:12:13
6		ichiban_KIRIN	2691870948	2021-03-15 07:24:12
7		ichiban_KIRIN	2691870948	2021-03-15 08:18:35
8		ichiban_KIRIN	2691870948	2021-03-15 09:12:05

Table 8: Beverage table summary with date concert_dirty_2021-03-17_phillip.csv

	text	screen_name	user_id	created_at
5		KevinCmhasia	1.294871e+09	2021-03-17 21:02:16
6	Hay que ser imbcil para pelearse por ftbol Hold my beer	chijaukay	7.714309e+08	2021-03-17 21:02:13
7		_Rocklola	1.353077e+18	2021-03-17 21:02:11
8		kkodzuken5	2.416495e+09	2021-03-17 21:02:10

a) Remove all non English tweets (you must indicate how many tweets were removed).

```r
# create subset of english only tweet

#tweet.beer.english = subset(tweet.beer, tweet.beer$lang == "en")
#tweet.concert.english = subset(tweet.concert, tweet.concert$lang == "en")

#sprintf("The number of beer tweets that were removed are %d", nrow(tweet.beer)- nrow(tweet.beer.englis
#sprintf("The number of concert tweets that were removed are %d", nrow(tweet.concert)- nrow(tweet.conce

bev1.tweet.clean = subset(bev1.tweet, bev1.tweet$lang=="en");
bev2.tweet.clean = subset(bev2.tweet, bev2.tweet$lang=="en");
party1.tweet.clean = subset(party1.tweet, party1.tweet$lang=="en");
party2.tweet.clean = subset(party2.tweet, party2.tweet$lang=="en");
beer1.tweet.clean = subset(beer1.tweet, beer1.tweet$lang=="en");
beer2.tweet.clean = subset(beer2.tweet, beer2.tweet$lang=="en");
con1.tweet.clean = subset(con1.tweet, con1.tweet$lang=="en");
con2.tweet.clean = subset(con2.tweet, con2.tweet$lang=="en");

print(paste("The number of tweets that were removed for", tweets_data_dirty_names[1],":", nrow(bev1.twee
```

```
## [1] "The number of tweets that were removed for beverage_dirty_2021-03-10_phillip.csv : 1006"
```

```r
print(paste("The number of tweets that were removed for", tweets_data_dirty_names[2],":", nrow(bev2.twee
```

```
## [1] "The number of tweets that were removed for beverage_dirty_2021-03-13_phillip.csv : 897"
```

```r
print(paste("The number of tweets that were removed for", tweets_data_dirty_names[3],":", nrow(party1.tw
```

```
## [1] "The number of tweets that were removed for party_dirty_2021-03-10_phillip.csv : 1006"
```

```r
print(paste("The number of tweets that were removed for", tweets_data_dirty_names[4],":", nrow(party2.tw
```

```
## [1] "The number of tweets that were removed for party_dirty_2021-03-13_phillip.csv : 897"
```

```r
print(paste("The number of tweets that were removed for", tweets_data_dirty_names[5],":", nrow(beer1.twe
```

```
## [1] "The number of tweets that were removed for beer_dirty_2021-03-14_phillip.csv : 1733"
```

```r
print(paste("The number of tweets that were removed for", tweets_data_dirty_names[6],":", nrow(beer2.twe
```

```
## [1] "The number of tweets that were removed for beer_dirty_2021-03-16_phillip.csv : 6441"
```

```r
print(paste("The number of tweets that were removed for", tweets_data_dirty_names[7],":", nrow(con1.twe
```

```
## [1] "The number of tweets that were removed for concert_dirty_2021-03-15_phillip.csv : 7518"
```

```r
print(paste("The number of tweets that were removed for", tweets_data_dirty_names[8],":", nrow(con2.twe
```

```
## [1] "The number of tweets that were removed for concert_dirty_2021-03-17_phillip.csv : 864"
```

b) A tweet is considered a duplicate if the text is the same as another tweet. Remove all duplicate tweets(you must indicate how many tweets were removed)

```r
#tweet.beer.english.unique = subset(tweet.beer.english, !duplicated(tweet.beer.english$text) )
#tweet.concert.english.unique = subset(tweet.concert.english,  !duplicated(tweet.concert.english$text) 

#sprintf("The number of duplicate beer tweets that were removed are %d", nrow(tweet.beer.english)- nrow
#sprintf("The number of duplicate concert tweets that were removed are %d", nrow(tweet.concert.english)

bev1.tweet.clean.unique = subset(bev1.tweet.clean, !duplicated(bev1.tweet.clean$text))
bev2.tweet.clean.unique = subset(bev1.tweet.clean, !duplicated(bev1.tweet.clean$text))
party1.tweet.clean.unique = subset(party1.tweet.clean, !duplicated(party1.tweet.clean$text))
party2.tweet.clean.unique = subset(party2.tweet.clean, !duplicated(party2.tweet.clean$text))
beer1.tweet.clean.unique = subset(beer1.tweet.clean, !duplicated(beer1.tweet.clean$text))
beer2.tweet.clean.unique = subset(beer2.tweet.clean, !duplicated(beer2.tweet.clean$text))
con1.tweet.clean.unique = subset(con1.tweet.clean, !duplicated(con1.tweet.clean$text))
con2.tweet.clean.unique = subset(con2.tweet.clean, !duplicated(con2.tweet.clean$text))
```

```r
print(paste("The number of tweets that were removed for", tweets_data_dirty_names[1],":", nrow(bev1.twee
```

```
## [1] "The number of tweets that were removed for beverage_dirty_2021-03-10_phillip.csv : 3434"
```

```r
print(paste("The number of tweets that were removed for", tweets_data_dirty_names[2],":", nrow(bev2.twee
```

```
## [1] "The number of tweets that were removed for beverage_dirty_2021-03-13_phillip.csv : 3301"
```

```r
print(paste("The number of tweets that were removed for", tweets_data_dirty_names[3],":", nrow(party1.tu
```

```
## [1] "The number of tweets that were removed for party_dirty_2021-03-10_phillip.csv : 3434"
```

```
print(paste("The number of tweets that were removed for", tweets_data_dirty_names[4],":", nrow(party2.tw
```

[1] "The number of tweets that were removed for party_dirty_2021-03-13_phillip.csv : 3389"

```
print(paste("The number of tweets that were removed for", tweets_data_dirty_names[5],":", nrow(beer1.twe
```

[1] "The number of tweets that were removed for beer_dirty_2021-03-14_phillip.csv : 3268"

```
print(paste("The number of tweets that were removed for", tweets_data_dirty_names[6],":", nrow(beer2.twe
```

[1] "The number of tweets that were removed for beer_dirty_2021-03-16_phillip.csv : 1242"

```
print(paste("The number of tweets that were removed for", tweets_data_dirty_names[7],":", nrow(con1.twee
```

[1] "The number of tweets that were removed for concert_dirty_2021-03-15_phillip.csv : 669"

```
print(paste("The number of tweets that were removed for", tweets_data_dirty_names[8],":", nrow(con2.twee
```

[1] "The number of tweets that were removed for concert_dirty_2021-03-17_phillip.csv : 3347"

** - The tables after the changes are made **

Table 9: Beverage table summary with date beverage_dirty_2021-03-10_phillip.csv

	text	screen_name	user_id	created_at
15	TAPNew Insider Filing on MOLSON COORS BEVERAGE COsPres&ampCEO Molson Coors EuropeSIMON COX	InsiderAlerts42	1234442043	2021-03-10 01:00:23
17	TAPNew Insider Filing on MOLSON COORS BEVERAGE COsPresident &amp CEOGAVIN HATTERSLEY	InsiderAlerts42	1234442043	2021-03-07 00:29:23
18	KDPNew Insider Filing on Keurig Dr Pepper Incs10 OwnerJAB BEVERAGE PLATFORM BV	InsiderAlerts42	1234442043	2021-03-08 11:36:24
19	TAPNew Insider Filing on MOLSON COORS BEVERAGE COsChief Financial OfficerTRACEY JOUBERT	InsiderAlerts42	1234442043	2021-03-06 05:12:23

Table 10: Beverage table summary with date beverage_dirty_2021-03-13_phillip.csv

	text	screen_name	user_id	created_at
15	TAPNew Insider Filing on MOLSON COORS BEVERAGE COsPres&ampCEO Molson Coors EuropeSIMON COX	InsiderAlerts42	1234442043	2021-03-10 01:00:23
17	TAPNew Insider Filing on MOLSON COORS BEVERAGE COsPresident &amp CEOGAVIN HATTERSLEY	InsiderAlerts42	1234442043	2021-03-07 00:29:23
18	KDPNew Insider Filing on Keurig Dr Pepper Incs10 OwnerJAB BEVERAGE PLATFORM BV	InsiderAlerts42	1234442043	2021-03-08 11:36:24
19	TAPNew Insider Filing on MOLSON COORS BEVERAGE COsChief Financial OfficerTRACEY JOUBERT	InsiderAlerts42	1234442043	2021-03-06 05:12:23

Table 11: Beverage table summary with date party_dirty_2021-03-10_phillip.csv

	text	screen_name	user_id	created_at
15	TAPNew Insider Filing on MOLSON COORS BEVERAGE COsPres&ampCEO Molson Coors EuropeSIMON COX	InsiderAlerts42	1234442043	2021-03-10 01:00:23
17	TAPNew Insider Filing on MOLSON COORS BEVERAGE COsPresident &amp CEOGAVIN HATTERSLEY	InsiderAlerts42	1234442043	2021-03-07 00:29:23
18	KDPNew Insider Filing on Keurig Dr Pepper Incs10 OwnerJAB BEVERAGE PLATFORM BV	InsiderAlerts42	1234442043	2021-03-08 11:36:24
19	TAPNew Insider Filing on MOLSON COORS BEVERAGE COsChief Financial OfficerTRACEY JOUBERT	InsiderAlerts42	1234442043	2021-03-06 05:12:23

Table 12: Beverage table summary with date party_dirty_2021-03-13_phillip.csv

	text	screen_name	user_id	created_at
12	GaaS Cloud offers a	gaascloud	9.476102e+17	2021-03-13 01:19:03
13	Recipe from GaaS Cloud	gaascloud	9.476102e+17	2021-03-13 22:31:02
14	Sometimes you just need a break and delicious	gaascloud	9.476102e+17	2021-03-12 19:35:04
15	Quench your thirst with our selection of waters	gaascloud	9.476102e+17	2021-03-10 20:21:08

Table 13: Beverage table summary with date beer_dirty_2021-03-14_phillip.csv

	text	screen_name	user_id	created_at
20	Campaign of destroying military goods such as cigarettes beer is gaining ground not only BOYCOTT military products but you also make sure no one use that have left overLOCOMOTIVE XYZ	Maysu22023	2572842018	2021-03-14 06:49:51
21	Brothers at Trappist abbeys are aging and fewer men are taking vows The world asks what does that mean for my Trappist beer?	InakiErauso	840099520	2021-03-14 06:49:44
22	Esus did drunk me not open my babalas beer and left it in the fridge what a jerk F31A	Mikka_90	1553242000	2021-03-14 06:49:44
24	I should drink all the beer in the house Prove me wrong	joerodconardy	2244596028	2021-03-14 06:49:42

Table 14: Beverage table summary with date beer_dirty_2021-03-16_phillip.csv

	text	screen_name	user_id	created_at
4172	CNNs Brianna Keilar Scolds Ron DeSantis For Drinking A Beer And Having Fun In Daytona Beach	DailyCaller	3930852	2021-03-16 11:00:35
4173	50 years ago CAMRA was born CAMRA was founded in 1971 with the simple vision of improving consumer choice of great beer and pubs 50 years on there is a lot to celebrate Never been a better time to join	LiverpoolCAMRA	173712	2021-03-16 11:00:32

	text	screen_name	user_id	created_at
4174	Check us out in the Craft Beer Travel &amp Adventure podcast on Living A Stout Life We talk about Schoolhouses history our future all things beer and our stance as a homebrew shop and brewery Go give it a listen	mariettabrews	2434517031	2021-03-16 11:00:31
4913	Beer Here Bouquets Next Door How a Bar Defied the Pandemic	Uskeus	900732607	2021-03-16 11:00:03

Table 15: Beverage table summary with date concert_dirty_2021-03-15_phillip.csv

	text	screen_name	user_id	created_at
6526	Beer Twitter is having a relaxing pint	BeerTwitt3r	1.027901e+18	2021-03-15 07:11:22
6528	A sobering strategy for pandemic times -	tbsnewsdotnet	1.149174e+18	2021-03-15 09:11:11
6529	Huge government contract announced after months of waiting	s4mAIM	1.323968e+18	2021-03-15 09:11:06
6530	Carlsberg Spore Official Store selling carton of 24 packs only costs S158 per can	GreatDealsSG	4.530807e+07	2021-03-15 09:11:04

Table 16: Beverage table summary with date concert_dirty_2021-03-17_phillip.csv

	text	screen_name	user_id	created_at
11	Free beer for Arty	batsnshit	8.093946e+17	2021-03-17 21:02:07
12	Something fun for your St Patricks Day Sing along with the Winnipeg Beer Choir D	harmer_katy	1.390629e+09	2021-03-17 21:02:06
13	The folks over at	glacierband	2.437819e+09	2021-03-17 21:02:05
14	The Asero Aluminum Artisan for ePBT Aesthetic is herefrom /u/PMMEASHOWERBEER	r_mkeyboards	1.294947e+18	2021-03-17 21:02:02

c) Write the remainin g tweets data to a file (.csv). The csv filename should have the format . For example, for tweets on b everage " retrieved on March 07 by Anderson would be: b e verage __202 1 Mar07__Anderson.csv

```
#cdate = as.Date(Sys.time())

#write.csv(tweet.beer.english.unique,paste("beer",cdate,"phillip.csv",sep = "_",collapse = NULL), row.n
#write.csv(tweet.beer,paste("beer_dirty",cdate,"phillip.csv",sep = "_",collapse = NULL), row.names = F)

#write.csv(tweet.beer.english.unique,paste("concert",cdate,"phillip.csv",sep = "_",collapse = NULL), ro
#write.csv(tweet.beer,paste("concert_dirty",cdate,"phillip.csv",sep = "_",collapse = NULL), row.names =

write.csv(bev1.tweet.clean.unique,tweets_data_names[1], row.names = F)
write.csv(bev2.tweet.clean.unique,tweets_data_names[2], row.names = F)
write.csv(party1.tweet.clean.unique,tweets_data_names[3], row.names = F)
```

```r
write.csv(party2.tweet.clean.unique,tweets_data_names[4], row.names = F)
write.csv(beer1.tweet.clean.unique,tweets_data_names[5], row.names = F)
write.csv(beer2.tweet.clean.unique,tweets_data_names[6], row.names = F)
write.csv(con1.tweet.clean.unique,tweets_data_names[7], row.names = F)
write.csv(con2.tweet.clean.unique,tweets_data_names[8], row.names = F)
```

d) Write code to review and s how details of tweets retrieved including number of tweets (after doi ng 2 a c)c), screen_name with the most followers, tweet with the most retweets, location from which the most tweets originate.

```r
#save the  data frames into another variable just incase modifiaction is needed
#tweet.beer.summary <- tweet.beer.english.unique
#tweet.concert.summary <- tweet.concert.english.unique

#print("beer Summary")
#print("-------------------------------------------------")
#print(paste("The number of Twets retrieved are: ", nrow(tweet.beer.summary)))

#print(paste("The user with the most followers: ",
#            unname(
#              tweet.beer.summary[
#                max(tweet.beer.summary$followers_count)==tweet.beer.summary$followers_count
#                ,"screen_name"
#                ]
#              )
#            )
#      )

#print(paste("The tweets with the most retweet: ",
#            unname(
#              tweet.beer.summary[
#                max(tweet.beer.summary$retweet_count)==tweet.beer.summary$retweet_count
#                ,"text"
#                ]
#              )
#            )
#      )


# count the number occurances for each location
#beers_location_count <- table(tweet.beer.summary$location)
#beers_location_count <- as.data.frame(beers_location_count)

# remove empty location row
#beers_location_count <- subset(beers_location_count, beers_location_count$Var1 != "")

#print the location with the maximum occurrences
#print(paste("The location with the most tweets: ",
#            unname(
#              beers_location_count[
#                max(beers_location_count$Freq)==beers_location_count$Freq
#                ,"Var1"
#                ]
```

```
#                )
#              )
#        )

#print("----------------------------------------------------")

#print("")
#print("concert Summary")
#print("----------------------------------------------------")
#print(paste("The number of Twets retrieved are: ", nrow(tweet.concert.summary)))

#print(paste("The user with the most followers: ",
#            unname(
#              tweet.concert.summary[
#                max(tweet.concert.summary$followers_count)==tweet.concert.summary$followers_count
#                ,"screen_name"
#                ]
#              )
#            )
#        )

#print(paste("The tweets with the most retweet: ",
#            unname(
#              tweet.concert.summary[
#                max(tweet.concert.summary$retweet_count)==tweet.concert.summary$retweet_count
#                ,"text"
#                ]
#              )
#            )
#        )


# count the number occurances for each location
#concert_location_count <- table(tweet.concert.summary$location)
#concert_location_count <- as.data.frame(concert_location_count)

# remove empty location row
#concert_location_count <- subset(concert_location_count, concert_location_count$Var1 != "")

#print the location with the maximum occurrences
#print(paste("The location with the most tweets: ",
#            unname(
#              concert_location_count[
#                max(concert_location_count$Freq)==concert_location_count$Freq
#                ,"Var1"
#                ]
#              )
#            )
#        )


#print("----------------------------------------------------")
```

```r
bev1.summary <- bev1.tweet.clean.unique

print(paste(tweets_data_names[1],"Summary", sep = " ",collapse = NULL))
```

```
## [1] "beverage_2021-03-10_phillip.csv Summary"
```

```r
print("-----------------------------------------------")
```

```
## [1] "-----------------------------------------------"
```

```r
print(paste("The number of Twets retrieved are: ", nrow(bev1.summary)))
```

```
## [1] "The number of Twets retrieved are:  4557"
```

```r
print(paste("The user with the most followers: ",
            unname(
              bev1.summary[
                max(bev1.summary$followers_count)==bev1.summary$followers_count
                ,"screen_name"
                ]
              )
            )
      )
```

```
## [1] "The user with the most followers:  XHNews"
```

```r
print(paste("The tweets with the most retweet: ",
            unname(
              bev1.summary[
                max(bev1.summary$retweet_count)==bev1.summary$retweet_count
                ,"text"
                ]
              )
            )
      )
```

```
## [1] "The tweets with the most retweet:  MEDIA F1EFF1F5 Cosmopolitan Japan reports on Jungkook causing
```

```r
# count the number occurances for each location
bev1_location_count <- table(bev1.summary$location)
bev1_location_count <- as.data.frame(bev1_location_count)

# remove empty location row
bev1_location_count <- subset(bev1_location_count, bev1_location_count$Var1 != "")

#print the location with the maximum occurrences
print(paste("The location with the most tweets: ",
            unname(
              bev1_location_count[
                max(bev1_location_count$Freq)==bev1_location_count$Freq
```

```
                ,"Var1"
                ]
              )
            )
        )
```

[1] "The location with the most tweets: United States"

```
print("------------------------------------------------")
```

[1] "--"

```
print("")
```

[1] ""

```
print("")
```

[1] ""

```
bev2.summary <- bev2.tweet.clean.unique
print(paste(tweets_data_names[2],"Summary", sep = " ",collapse = NULL))
```

[1] "beverage_2021-03-13_phillip.csv Summary"

```
print("-----------------------------------------------")
```

[1] "---"

```
print(paste("The number of Twets retrieved are: ", nrow(bev2.summary)))
```

[1] "The number of Twets retrieved are: 4557"

```
print(paste("The user with the most followers: ",
            unname(
              bev2.summary[
                max(bev2.summary$followers_count)==bev2.summary$followers_count
                ,"screen_name"
                ]
              )
            )
        )
```

[1] "The user with the most followers: XHNews"

14

```r
print(paste("The tweets with the most retweet: ",
            unname(
              bev2.summary[
                max(bev2.summary$retweet_count)==bev2.summary$retweet_count
                ,"text"
                ]
              )
            )
      )
```

[1] "The tweets with the most retweet: MEDIA F1EFF1F5 Cosmopolitan Japan reports on Jungkook causin

```r
# count the number occurances for each location
bev2_location_count <- table(bev2.summary$location)
bev2_location_count <- as.data.frame(bev2_location_count)

# remove empty location row
bev2_location_count <- subset(bev2_location_count, bev2_location_count$Var1 != "")

#print the location with the maximum occurrences
print(paste("The location with the most tweets: ",
            unname(
              bev2_location_count[
                max(bev2_location_count$Freq)==bev2_location_count$Freq
                ,"Var1"
                ]
              )
            )
      )
```

[1] "The location with the most tweets: United States"

```r
print("------------------------------------------------")
```

[1] "--"

```r
print("")
```

[1] ""

```r
print("")
```

[1] ""

```r
party1.summary <- party1.tweet.clean.unique

print(paste(tweets_data_names[3],"Summary", sep = " ",collapse = NULL))
```

[1] "party_2021-03-10_phillip.csv Summary"

```r
print("-------------------------------------------------")
```

```
## [1] "-------------------------------------------------"
```

```r
print(paste("The number of Twets retrieved are: ", nrow(party1.summary)))
```

```
## [1] "The number of Twets retrieved are:  4557"
```

```r
print(paste("The user with the most followers: ",
            unname(
              party1.summary[
                max(party1.summary$followers_count)==party1.summary$followers_count
                ,"screen_name"
                ]
              )
            )
      )
```

```
## [1] "The user with the most followers:  XHNews"
```

```r
print(paste("The tweets with the most retweet: ",
            unname(
              party1.summary[
                max(party1.summary$retweet_count)==party1.summary$retweet_count
                ,"text"
                ]
              )
            )
      )
```

```
## [1] "The tweets with the most retweet:  MEDIA F1EFF1F5 Cosmopolitan Japan reports on Jungkook causing
```

```r
# count the number occurances for each location
party1_location_count <- table(party1.summary$location)
party1_location_count <- as.data.frame(party1_location_count)

# remove empty location row
party1_location_count <- subset(party1_location_count, party1_location_count$Var1 != "")

#print the location with the maximum occurrences
print(paste("The location with the most tweets: ",
            unname(
              party1_location_count[
                max(party1_location_count$Freq)==party1_location_count$Freq
                ,"Var1"
                ]
              )
            )
      )
```

```
## [1] "The location with the most tweets:  United States"
```

```r
print("-------------------------------------------------")
```

```
## [1] "-------------------------------------------------"
```

```r
print("")
```

```
## [1] ""
```

```r
print("")
```

```
## [1] ""
```

```r
party2.summary <- party2.tweet.clean.unique

print(paste(tweets_data_names[4],"Summary", sep = " ",collapse = NULL))
```

```
## [1] "party_2021-03-13_phillip.csv Summary"
```

```r
print("-------------------------------------------------")
```

```
## [1] "-------------------------------------------------"
```

```r
print(paste("The number of Twets retrieved are: ", nrow(party2.summary)))
```

```
## [1] "The number of Twets retrieved are:  4469"
```

```r
print(paste("The user with the most followers: ",
            unname(
              party2.summary[
                max(party2.summary$followers_count)==party2.summary$followers_count
                ,"screen_name"
                ]
              )
           )
      )
```

```
## [1] "The user with the most followers:  ABSCBNNews"
```

```r
print(paste("The tweets with the most retweet: ",
            unname(
              party2.summary[
                max(party2.summary$retweet_count)==party2.summary$retweet_count
                ,"text"
                ]
              )
           )
      )
```

```
## [1] "The tweets with the most retweet:  Erica Nlewedim bags new endorsement deal with beverage compa
```

```r
# count the number occurances for each location
party2_location_count <- table(party2.summary$location)
party2_location_count <- as.data.frame(party2_location_count)

# remove empty location row
party2_location_count <- subset(party2_location_count, party2_location_count$Var1 != "")

#print the location with the maximum occurrences
print(paste("The location with the most tweets: ",
            unname(
              party2_location_count[
                max(party2_location_count$Freq)==party2_location_count$Freq
                ,"Var1"
                ]
              )
            )
      )
```

```
## [1] "The location with the most tweets:  United States"
```

```r
print("--------------------------------------------------")
```

```
## [1] "--------------------------------------------------"
```

```r
print("")
```

```
## [1] ""
```

```r
print("")
```

```
## [1] ""
```

```r
beer1.summary <- beer1.tweet.clean.unique

print(paste(tweets_data_names[5],"Summary", sep = " ",collapse = NULL))
```

```
## [1] "beer_2021-03-14_phillip.csv Summary"
```

```r
print("--------------------------------------------------")
```

```
## [1] "--------------------------------------------------"
```

```r
print(paste("The number of Twets retrieved are: ", nrow(beer1.summary)))
```

```
## [1] "The number of Twets retrieved are:  3822"
```

```r
print(paste("The user with the most followers: ",
            unname(
              beer1.summary[
                max(beer1.summary$followers_count)==beer1.summary$followers_count
                ,"screen_name"
                ]
              )
            )
      )
```

```
## [1] "The user with the most followers:  ABC"
```

```r
print(paste("The tweets with the most retweet: ",
            unname(
              beer1.summary[
                max(beer1.summary$retweet_count)==beer1.summary$retweet_count
                ,"text"
                ]
              )
            )
      )
```

```
## [1] "The tweets with the most retweet:  Good people drink good beer "
```

```r
# count the number occurances for each location
beer1_location_count <- table(beer1.summary$location)
beer1_location_count <- as.data.frame(beer1_location_count)

# remove empty location row
beer1_location_count <- subset(beer1_location_count, beer1_location_count$Var1 != "")

#print the location with the maximum occurrences
print(paste("The location with the most tweets: ",
            unname(
              beer1_location_count[
                max(beer1_location_count$Freq)==beer1_location_count$Freq
                ,"Var1"
                ]
              )
            )
      )
```

```
## [1] "The location with the most tweets:  Los Angeles, CA"
```

```r
print("-----------------------------------------------")
```

```
## [1] "-----------------------------------------------"
```

```r
print("")
```

```
## [1] ""
```

```
print("")
```

```
## [1] ""
```

```
beer2.summary <- beer2.tweet.clean.unique
```

```
print(paste(tweets_data_names[6],"Summary", sep = " ",collapse = NULL))
```

```
## [1] "beer_2021-03-16_phillip.csv Summary"
```

```
print("-------------------------------------------------")
```

```
## [1] "------------------------------------------------"
```

```
print(paste("The number of Twets retrieved are: ", nrow(beer2.summary)))
```

```
## [1] "The number of Twets retrieved are:  1317"
```

```
print(paste("The user with the most followers: ",
            unname(
              beer2.summary[
                max(beer2.summary$followers_count)==beer2.summary$followers_count
                ,"screen_name"
                ]
              )
            )
      )
```

```
## [1] "The user with the most followers:  Londonist"
```

```
print(paste("The tweets with the most retweet: ",
            unname(
              beer2.summary[
                max(beer2.summary$retweet_count)==beer2.summary$retweet_count
                ,"text"
                ]
              )
            )
      )
```

```
## [1] "The tweets with the most retweet:  Believing that the best way to prepare for marriage is to go
```

```
# count the number occurances for each location
beer2_location_count <- table(beer2.summary$location)
beer2_location_count <- as.data.frame(beer2_location_count)

# remove empty location row
beer2_location_count <- subset(beer2_location_count, beer2_location_count$Var1 != "")
```

```r
#print the location with the maximum occurrences
print(paste("The location with the most tweets: ",
            unname(
              beer2_location_count[
                max(beer2_location_count$Freq)==beer2_location_count$Freq
                ,"Var1"
                ]
              )
            )
      )
```

```
## [1] "The location with the most tweets:  London"
```

```r
print("------------------------------------------------")
```

```
## [1] "------------------------------------------------"
```

```r
print("")
```

```
## [1] ""
```

```r
print("")
```

```
## [1] ""
```

```r
con1.summary <- con1.tweet.clean.unique

print(paste(tweets_data_names[7],"Summary", sep = " ",collapse = NULL))
```

```
## [1] "concert_2021-03-15_phillip.csv Summary"
```

```r
print("------------------------------------------------")
```

```
## [1] "------------------------------------------------"
```

```r
print(paste("The number of Twets retrieved are: ", nrow(con1.summary)))
```

```
## [1] "The number of Twets retrieved are:  813"
```

```r
print(paste("The user with the most followers: ",
            unname(
              con1.summary[
                max(con1.summary$followers_count)==con1.summary$followers_count
                ,"screen_name"
                ]
              )
            )
      )
```

```
## [1] "The user with the most followers:  Konami"
```

21

```r
print(paste("The tweets with the most retweet: ",
            unname(
              con1.summary[
                max(con1.summary$retweet_count)==con1.summary$retweet_count
                ,"text"
                ]
              )
            )
      )
```

[1] "The tweets with the most retweet: Give a man a fish and he will eat for a day Teach him how to

```r
# count the number occurances for each location
con1_location_count <- table(con1.summary$location)
con1_location_count <- as.data.frame(con1_location_count)

# remove empty location row
con1_location_count <- subset(con1_location_count, con1_location_count$Var1 != "")

#print the location with the maximum occurrences
print(paste("The location with the most tweets: ",
            unname(
              con1_location_count[
                max(con1_location_count$Freq)==con1_location_count$Freq
                ,"Var1"
                ]
              )
            )
      )
```

[1] "The location with the most tweets: London"

```r
print("-----------------------------------------------")
```

[1] "---"

```r
print("")
```

[1] ""

```r
print("")
```

[1] ""

```r
con2.summary <- con2.tweet.clean.unique

print(paste(tweets_data_names[8],"Summary", sep = " ",collapse = NULL))
```

[1] "concert_2021-03-17_phillip.csv Summary"

```r
print("---------------------------------------------------")
```

```
## [1] "---------------------------------------------------"
```

```r
print(paste("The number of Twets retrieved are: ", nrow(con2.summary)))
```

```
## [1] "The number of Twets retrieved are:  4789"
```

```r
print(paste("The user with the most followers: ",
            unname(
              con2.summary[
                max(con2.summary$followers_count)==con2.summary$followers_count
                ,"screen_name"
                ]
              )
            )
      )
```

```
## [1] "The user with the most followers:  cnni"
```

```r
print(paste("The tweets with the most retweet: ",
            unname(
              con2.summary[
                max(con2.summary$retweet_count)==con2.summary$retweet_count
                ,"text"
                ]
              )
            )
      )
```

```
## [1] "The tweets with the most retweet:  WIN Want to win yourself a Warsteiner glass &amp a case of be
```

```r
# count the number occurances for each location
con2_location_count <- table(con2.summary$location)
con2_location_count <- as.data.frame(con2_location_count)

# remove empty location row
con2_location_count <- subset(con2_location_count, con2_location_count$Var1 != "")

#print the location with the maximum occurrences
print(paste("The location with the most tweets: ",
            unname(
              con2_location_count[
                max(con2_location_count$Freq)==con2_location_count$Freq
                ,"Var1"
                ]
              )
            )
      )
```

```
## [1] "The location with the most tweets:  Houston, TX"
```

```
print("---------------------------------------------")
```

```
## [1] "---------------------------------------------"
```

```
print("")
```

```
## [1] ""
```

```
print("")
```

```
## [1] ""
```