exercise2

#올리브영에 가장 많이 입점한 브랜드와 카테고리 간의 상관관계 파악하기

카테고리별로 제품들 가져오기

카테고리 1. 스킨케어 2.메이크업, 3.바디케어, 4.헤어케어, 5.향수 디퓨저, 6.미용소품, 7.남성 용품 크롤링

#'1. 스킨케어' 제품들 크롤링하기, 파일 만들기 #brand<-NULL #goods<-NULL #price<-NULL #crwal_func2<-function(x,url){ #for (i in 1:x){ #url2<-paste(url,i,sep="") #htxt<-read_html(url2) #brand<<-append(brand,html_nodes(htxt,"a.goodsList #span.tx_brand")%>%html_text()) #goods<<-append(goods,html_nodes(htxt,"p.tx_name")%>%html_text()) #price<<-append(price,html_nodes(htxt,"span.tx_cur #span.tx_num")%>%html_text()) # } #} #crwal_func2(51,"https://www.oliveyoung.co.kr/store/display/getMCateg#oryList.do?dispCatNo=100000100010001&fltDispCatNo=&prdSort=03&pageId#x= (https://www.oliveyoung.co.kr/store/display/getMCateg#oryList.do?dispCatNo=100000100010001&fltDispCatNo=&prdSort=03&pageId#x=)")
#crwal_func2(24,"https://www.oliveyoung.co.kr/store/display/getMCategoryList.do?dispCatNo=100000100010002&fltDispCatNo=&prdSort=03&pageIdx= (https://www.oliveyoung.co.kr/store/display/getMCategoryList.do?dispCatNo=100000100010002&fltDispCatNo=&prdSort=03&pageIdx=)")
#crwal_func2(25,"https://www.oliveyoung.co.kr/store/display/getMCategoryList.do?dispCatNo=100000100010006&fltDispCatNo=&prdSort=03&pageIdx= (https://www.oliveyoung.co.kr/store/display/getMCategoryList.do?dispCatNo=100000100010006&fltDispCatNo=&prdSort=03&pageIdx=)")
#crwal_func2(11,"https://www.oliveyoung.co.kr/store/display/getMCategoryList.do?dispCatNo=100000100010007&fltDispCatNo=&prdSort=03&pageIdx= (https://www.oliveyoung.co.kr/store/display/getMCategoryList.do?dispCatNo=100000100010007&fltDispCatNo=&prdSort=03&pageIdx=)")
#crwal_func2(14,"https://www.oliveyoung.co.kr/store/display/getMCategoryList.do?dispCatNo=100000100010004&fltDispCatNo=&prdSort=03&pageIdx= (https://www.oliveyoung.co.kr/store/display/getMCategoryList.do?dispCatNo=100000100010004&fltDispCatNo=&prdSort=03&pageIdx=)")

#skin_care_goods<-data.frame(brand,goods,price) #head(skin_care_goods) #write.csv(skin_care_goods,"skin_care_goods.csv")

#분석 도중 데이터가 없데이트 돼서 일단 주석으로 바꾸고 쓴 파일을 읽어오는 방식으로 진행

```
skin_care_goods<-read.csv("skin_care_goods.csv")
skin_care_goods<-skin_care_goods[,c(2,3,4)]
head(skin_care_goods)
```

```
##      brand
## 1 아이소이
## 2 라운드랩
## 3 쏘내추럴
## 4 마녀공장
## 5 에스트라
## 6 브링그린
##                                                                  goods
## 1                [대용량 한정판] 아이소이 1등잡티세럼 40ml+마스크 20g 증정기획
## 2                   라운드랩 1025 독도 토너 대용량 500ml 기획(토너 100ml 증정)
## 3                        쏘내추럴 올 데이 메이크업 픽서 75ml [안개분사 캔타입]
## 4 [한정기획] 마녀공장 비피다 바이옴 콤플렉스 앰플 증량 80ml+앰플토너 30ml 증정
## 5                에스트라 아토베리어365 예민보습 로션 기획세트(로션 60ml 증정)
## 6                                          브링그린 당근비타토너패드60매
##    price
## 1 39,500
## 2 28,500
## 3 12,600
## 4 29,500
## 5 22,680
## 6 12,240
```

```
#같은 크롤링 방법으로 2.메이크업, 3.바디케어, 4.헤어케어, 5.향수 디퓨저, 6.미용소품, 7.남성 용품 크롤링
makeup_goods<-read.csv("makeup_goods.csv")
makeup_goods<-makeup_goods[,c(2,3,4)]
head(makeup_goods)
```

```
##           brand
## 1 메이블린 뉴욕
## 2    투쿨포스쿨
## 3    투쿨포스쿨
## 4          루나
## 5      바닐라코
## 6      지베르니
##                                                                        goods
## 1                                      [올리브영단독기획]메이블린 슈퍼스테이 파운데이션
## 2                                  [브러쉬증정 기획] 투쿨포스쿨 바이로댕 피니쉬 세팅 팩트
## 3                                        투쿨포스쿨 바이로댕 쉐이딩(브러시 미포함)
## 4                                                    루나 롱래스팅 팁 컨실러
## 5 [기획세트] 바닐라코 커버리셔스 파워 핏 파운데이션 (퍼프2매증정) (New 뉴트럴컬러 추가)
## 6                                      [브러쉬 증정 기획]지베르니 밀착 커버 파운데이션
##    price
## 1 21,000
## 2 12,800
## 3 12,800
## 4 11,200
## 5 22,000
## 6 23,400
```

```
body_care_goods<-read.csv("body_care_goods.csv")
body_care_goods<-body_care_goods[,c(2,3,4)]

hair_care_goods<-read.csv("hair_care_goods.csv")
hair_care_goods<-hair_care_goods[,c(2,3,4)]

perfume_goods<-read.csv("perfume_goods.csv")
perfume_goods<-perfume_goods[,c(2,3,4)]

beauty_item<-read.csv("beauty_item.csv")
beauty_item<-beauty_item[,c(2,3,4)]

for_man_goods<-read.csv("for_man_goods.csv")
for_man_goods<-for_man_goods[,c(2,3,4)]
```

#brand 빈도수 알아보기 #스킨케어

```
#install.packages("dplyr")
library(dplyr)
```

```
## Warning: 패키지 'dplyr'는 R 버전 4.1.1에서 작성되었습니다
```

```
##
## 다음의 패키지를 부착합니다: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
#brand 빈도수 알아보기
skin_care<-skin_care_goods%>%group_by(brand)%>%summarise(freq=n())
#빈도수를 기준으로 상위 10개의 브랜드 10개 뽑아보기
top_skin_care_brand<-skin_care%>%arrange(desc(freq))%>%head(.,10)

#메이크업
maekup<-makeup_goods%>%group_by(brand)%>%summarise(freq=n())
top_maekup_brand<-maekup%>%arrange(desc(freq))%>%head(.,10)

#바디케어
body_care<-body_care_goods%>%group_by(brand)%>%summarise(freq=n())
top_body_care_brand<-body_care%>%arrange(desc(freq))%>%head(.,10)

#헤어케어
hair_care<-hair_care_goods%>%group_by(brand)%>%summarise(freq=n())
top_hair_care_brand<-hair_care%>%arrange(desc(freq))%>%head(.,10)

#향수
perfume<-perfume_goods%>%group_by(brand)%>%summarise(freq=n())
top_perfume_brand<-perfume%>%arrange(desc(freq))%>%head(.,10)

#미용소품
beauty<-beauty_item%>%group_by(brand)%>%summarise(freq=n())
top_beauty_item_brand<-beauty%>%arrange(desc(freq))%>%head(.,10)

#남자
man<-for_man_goods%>%group_by(brand)%>%summarise(freq=n())
top_for_man_brand<-man%>%arrange(desc(freq))%>%head(.,10)

#install.packages("tidyverse")
library(tidyverse)
```

```
## Warning: 패키지 'tidyverse'는 R 버전 4.1.1에서 작성되었습니다
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.4     v stringr 1.4.0
## v tidyr   1.1.3     v forcats 0.5.1
## v readr   2.0.1
```

```
## Warning: 패키지 'ggplot2'는 R 버전 4.1.1에서 작성되었습니다
```

```
## Warning: 패키지 'tibble'는 R 버전 4.1.1에서 작성되었습니다
```

```
## Warning: 패키지 'tidyr'는 R 버전 4.1.1에서 작성되었습니다
```

```
## Warning: 패키지 'readr'는 R 버전 4.1.1에서 작성되었습니다
```

```
## Warning: 패키지 'purrr'는 R 버전 4.1.1에서 작성되었습니다
```

```
## Warning: 패키지 'stringr'는 R 버전 4.1.1에서 작성되었습니다
```

```
## Warning: 패키지 'forcats'는 R 버전 4.1.1에서 작성되었습니다
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
#카테고리를 알려줄 새로운 변수 추가하기
skin_care_goods<-skin_care_goods%>%mutate(category="스킨케어")
makeup_goods<-makeup_goods%>%mutate(category="메이크업")
body_care_goods<-body_care_goods%>%mutate(category="바디케어")
hair_care_goods<-hair_care_goods%>%mutate(category="헤어케어")
perfume_goods<-perfume_goods%>%mutate(category="향수/디퓨저")
beauty_item<-beauty_item%>%mutate(category="미용소품")
for_man_goods<-for_man_goods%>%mutate(category="남성")

#카테고리별로 브랜드의 개수를 파악하기

#스킨케어 카테고리에 입점한 브랜드의 개수
skin_care_goods%>%group_by(brand)%>%summarise(num=n())
```

```
## # A tibble: 259 x 2
##    brand          num
##    <chr>        <int>
##  1 23 years old     5
##  2 5days            1
##  3 AHC             31
##  4 DMCK             7
##  5 SRB              1
##  6 XTM              9
##  7 갸스비           5
##  8 구달            16
##  9 궁중비책         4
## 10 그라펜           6
## # ... with 249 more rows
```

```
skin_brand_num<-length(skin_care_goods$brand)

#메이크업 카테고리에 입점한 브랜드의 개수
makeup_goods%>%group_by(brand)%>%summarise(num=n())
```

```
## # A tibble: 101 x 2
##    brand            num
##    <chr>          <int>
##  1 16브랜드           1
##  2 3CE               26
##  3 AHC                2
##  4 XTM               13
##  5 그라펜             6
##  6 나인위시스         1
##  7 다슈               9
##  8 닥터원더           4
##  9 닥터자르트         8
## 10 닥터자르트포맨     1
## # ... with 91 more rows
```

```
makeup_brand_num<-length(makeup_goods$brand)

#바디케어 카테고리에 입점한 브랜드의 개수
body_care_goods%>%group_by(brand)%>%summarise(num=n())
```

```
## # A tibble: 234 x 2
##    brand            num
##    <chr>          <int>
##  1 8x4                8
##  2 ms.44              2
##  3 W피부연구소        1
##  4 XTM                2
##  5 Y.N.M              2
##  6 갸스비             4
##  7 괴솔               4
##  8 궁중비책          19
##  9 그라펜             3
## 10 그레이그라운드     1
## # ... with 224 more rows
```

```
body_brand_num<-length(body_care_goods$brand)

#헤어케어 카테고리에 입점한 브랜드의 개수
hair_care_goods%>%group_by(brand)%>%summarise(num=n())
```

```
## # A tibble: 122 x 2
##    brand            num
##    <chr>          <int>
##  1 AZH                3
##  2 CHI                1
##  3 OGX                9
##  4 XTM                9
##  5 갸스비            30
##  6 그라펜            11
##  7 그레이그라운드     1
##  8 꽃을든남자         3
##  9 니심               1
## 10 다나한             2
## # ... with 112 more rows
```

```
hair_brand_num<-length(hair_care_goods$brand)

#스킨과 메이크업 아이템에 입점한 브랜드의 개수
skin_makeup<-inner_join(skin_care_goods,makeup_goods,by="brand");head(skin_makeup)
```

```
##        brand                                                    goods.x
## 1   메디큐브                          메디큐브 제로 모공 패드 2.0 (70매)
## 2   메디큐브                          메디큐브 제로 모공 패드 2.0 (70매)
## 3   메디큐브                          메디큐브 제로 모공 패드 2.0 (70매)
## 4   메디큐브                          메디큐브 제로 모공 패드 2.0 (70매)
## 5   메디큐브                          메디큐브 제로 모공 패드 2.0 (70매)
## 6 보타닉힐보 보타닉힐보 더마 인텐시브 시카 판테놀 블레미쉬 크림 TOP 기획
##   price.x category.x
## 1  24,000   스킨케어
## 2  24,000   스킨케어
## 3  24,000   스킨케어
## 4  24,000   스킨케어
## 5  24,000   스킨케어
## 6  20,160   스킨케어
##                                                 goods.y price.y
## 1                                  메디큐브 제로 캡슐 쿠션  31,500
## 2                                   메디큐브 레드 캡슐쿠션  31,500
## 3                                  메디큐브 블루 캡슐 쿠션  31,500
## 4                                 메디큐브 레드 파운데이션  32,000
## 5                                    메디큐브 레드 컨실러  18,000
## 6 보타닉힐 보 더마 워터 세라마이드 앰플 쿠션 세트 [리뉴얼] (본품+리필)  19,500
##   category.y
## 1   메이크업
## 2   메이크업
## 3   메이크업
## 4   메이크업
## 5   메이크업
## 6   메이크업
```

```
skin_makeup_brand<-skin_makeup%>%group_by(brand)%>%summarise(num=n())
skin_makeup_brand_num<-length(skin_makeup_brand$brand)

#스킨과 바디케어 아이템에 입점한 브랜드의 개수
skin_body<-inner_join(skin_care_goods,body_care_goods,by="brand")
skin_body_brand<-skin_body%>%group_by(brand)%>%summarise(num=n())
skin_body_brand_num<-length(skin_body_brand$brand)

#스킨과 헤어케어 아이템에 입점한 브랜드의 개수
skin_hair<-inner_join(skin_care_goods,hair_care_goods,by="brand")
skin_hair_brand<-skin_hair%>%group_by(brand)%>%summarise(num=n())
skin_hair_brand_num<-length(skin_hair_brand$brand)

#메이크업과 바디케어 아이템에 입점한 브랜드의 개수
makeup_body<-inner_join(makeup_goods,body_care_goods,by="brand")
makeup_body_brand<-makeup_body%>%group_by(brand)%>%summarise(num=n())
makeup_body_brand_num<-length(makeup_body_brand$brand)

#메이크업과 헤어케어 아이템에 입점한 브랜드의 개수
makeup_hair<-inner_join(makeup_goods,hair_care_goods,by="brand")
makeup_hair_brand<-makeup_hair%>%group_by(brand)%>%summarise(num=n())
makeup_hair_brand_num<-length(makeup_hair_brand$brand)

#바디케어과 헤어케어 아이템에 입점한 브랜드의 개수
body_hair<-inner_join(body_care_goods,hair_care_goods,by="brand")
body_hair_brand<-body_hair%>%group_by(brand)%>%summarise(num=n())
body_hair_brand_num<-length(body_hair_brand$brand)

#크롤링한 모든 제품을 합쳐서 전체 제품을 담은 데이터프레임 만들기
total_product<-skin_care_goods%>%rbind(.,makeup_goods)%>%rbind(.,body_care_goods)%>%rbind(.,hair_care_goods)%>%rb
ind(.,perfume_goods)%>%rbind(.,beauty_item)%>%rbind(.,for_man_goods)

#브랜드별로 그룹핑 한 다음 빈도수를 확인하고 내림차순으로 정렬하기, 그 후 10개만 추출하기
top_total_product_brand<-total_product%>%group_by(brand)%>%summarise(freq=n())%>%arrange(desc(freq))%>%head(.,10)

#앞서 뽑은 가장 많이 나타나는 10개의 브랜드인 브랜드만 추출하고 brand열과 item열을 추출하기, 그 후 브랜드별로 정렬하고 중복되는 값
 없애기
total_top_brand_item<-total_product%>%filter(brand %in% top_total_product_brand$brand)%>%select(brand,category)%>
%arrange(brand)%>%unique()

#그래프를 그려 아이템와 많이 입점하고 있는 브랜드와 어떤 관계가 있는지 알아보기
#install.packages("igraph")
library(igraph)
```

```
## Warning: 패키지 'igraph'는 R 버전 4.1.1에서 작성되었습니다
```

```
##
## 다음의 패키지를 부착합니다: 'igraph'
```

```
## The following objects are masked from 'package:purrr':
##
##     compose, simplify
```

```
## The following object is masked from 'package:tidyr':
##
##     crossing
```

```
## The following object is masked from 'package:tibble':
##
##     as_data_frame
```

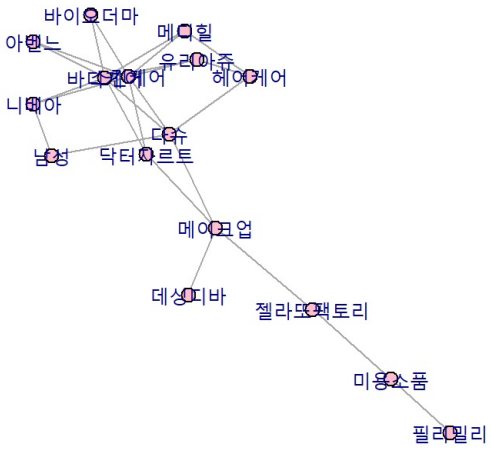
```
## The following objects are masked from 'package:dplyr':
##
##     as_data_frame, groups, union
```

```
## The following objects are masked from 'package:stats':
##
##     decompose, spectrum
```

```
## The following object is masked from 'package:base':
##
##     union
```

```
g<-graph.data.frame(total_top_brand_item,directed = F)

plot(g,layout=layout.fruchterman.reingold,vertex.size=7,edge.arrow.size=0.5,vertex.color="pink")
```




```
#네트워크 그래프는 16개의 노드와 26개의 링크로 연결이 되어 있음
# 방향의 의미가 중요치 않은 무방향 네트워크임

#무방향 이진 네트워크의 밀도 : k/{n(n-1)/2}

k<-26; n<-16
density<-k/{n*(n-1)/2}
density #네트워크 밀도
```

```
## [1] 0.2166667
```

```
#중심성 분석

#install.packages("tidygraph")
#install.packages("ggraph")

#매개 중심성 계산하기

library(tidygraph)
```

```
## Warning: 패키지 'tidygraph'는 R 버전 4.1.1에서 작성되었습니다
```

```
##
## 다음의 패키지를 부착합니다: 'tidygraph'
```

```
## The following object is masked from 'package:igraph':
##
##     groups
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
library(ggraph)
```

```
## Warning: 패키지 'ggraph'는 R 버전 4.1.1에서 작성되었습니다
```

```
total_top_brand_item %>%
  as_tbl_graph() %>%
  mutate(centrality_closeness()) %>%
  as_tibble
```

```
## Warning in closeness(graph = graph, vids = V(graph), mode = mode, weights =
## weights, : At centrality.c:2784 :closeness centrality is not well-defined for
## disconnected graphs
```

```
## # A tibble: 16 x 2
##    name         `centrality_closeness()`
##    <chr>                           <dbl>
##  1 니베아                         0.00513
##  2 다슈                           0.00606
##  3 닥터자르트                     0.00513
##  4 데싱디바                       0.00444
##  5 메디힐                         0.00513
##  6 바이오더마                     0.00476
##  7 아벤느                         0.00476
##  8 유리아쥬                       0.00513
##  9 젤라또팩토리                   0.00476
## 10 필리밀리                       0.00444
## 11 스킨케어                       0.00417
## 12 바디케어                       0.00417
## 13 남성                           0.00417
## 14 메이크업                       0.00417
## 15 헤어케어                       0.00417
## 16 미용소품                       0.00417
```

#다슈는 가장 많은 카테고리에 연결됨으로 인해 다른 노드들 간의
#네트워크 관계 형성에 있어서 중개자/ 매개자 역할을 가장 잘 수행