



Google Play Store 앱에 대하여

컴퓨터공학과
20171071
장혜정

목차



- 데이터 소개, 목적
- 변수 및 전처리 과정 •
- 상관분석
- 주성분 분석 •




01. 데이터 소개

데이터 세트

Google Play 스토어 앱

안드로이드 시장 분석을 위한 10k Play Store 앱의 웹 스크랩 데이터.

 라바 나 굽타 • 1년 전에 업데이트 됨 (버전 6)

Google play

2851

www.kaggle.com/lava18/google-play-store-apps
마지막 업데이트 : 2019



01. 목적

인기 앱/게임



ROBLOX
Roblox Corporation
★★★★★



카트라이더 러쉬플러스
NEXON Company
★★★★★



카트라이더 러쉬플러스

NEXON

★★★★★ 99,447

설치



올원뱅크(All One Bank)
농협BANK



NH농협카드 스마트
농협BANK

무승부등산가 - Draw
VOODOO

삼성 음성 녹음
Samsung Electronics

KB국민은행 스타뱅
KB국민은행

스타벅스
(주)스타벅스커피 코리아

쇼핑몰 마루온
G마켓

TikTok 틱톡
TikTok Pte. Ltd.

KB국민카드
KB국민카드



02. 변수

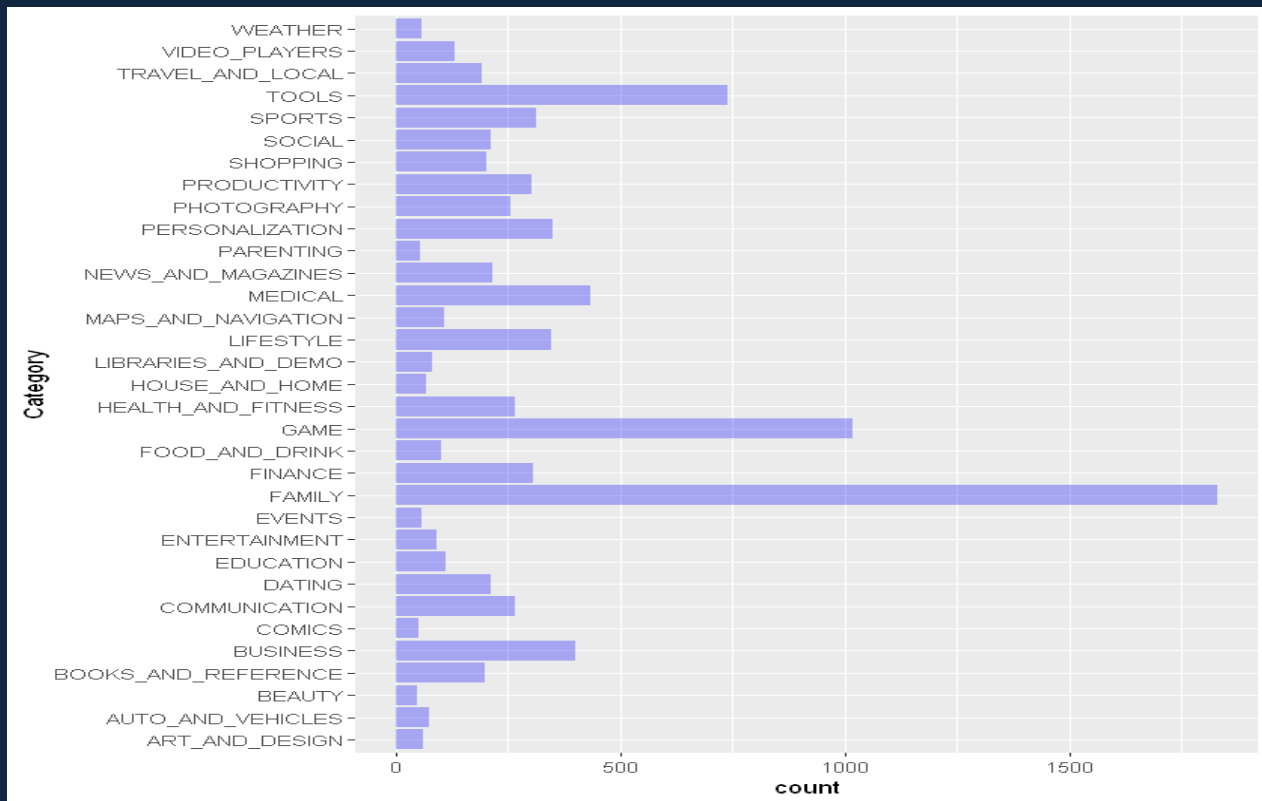
App	Category	Rating	Reviews	Size	Installs	Type	Price	Content.Rating	Genres	Last.Updated	Current.Ver	Android.Ver
Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	07-Jan-18	1.0.0	4.0.3 and up
Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	15-Jan-18	2.0.0	4.0.3 and up
U Launcher Lite ??FREE Live Cool Themes, Hide Apps	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	01-Aug-18	1.2.4	4.0.3 and up
Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	08-Jun-18	Varies with device	4.2 and up
Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design;Creativity	20-Jun-18	1.1	4.4 and up
Paper flowers instructions	ART_AND_DESIGN	4.4	167	5.6M	50,000+	Free	0	Everyone	Art & Design	26-Mar-17	1	2.3 and up

Size : “Varies with device” 제외
10804 → 9138



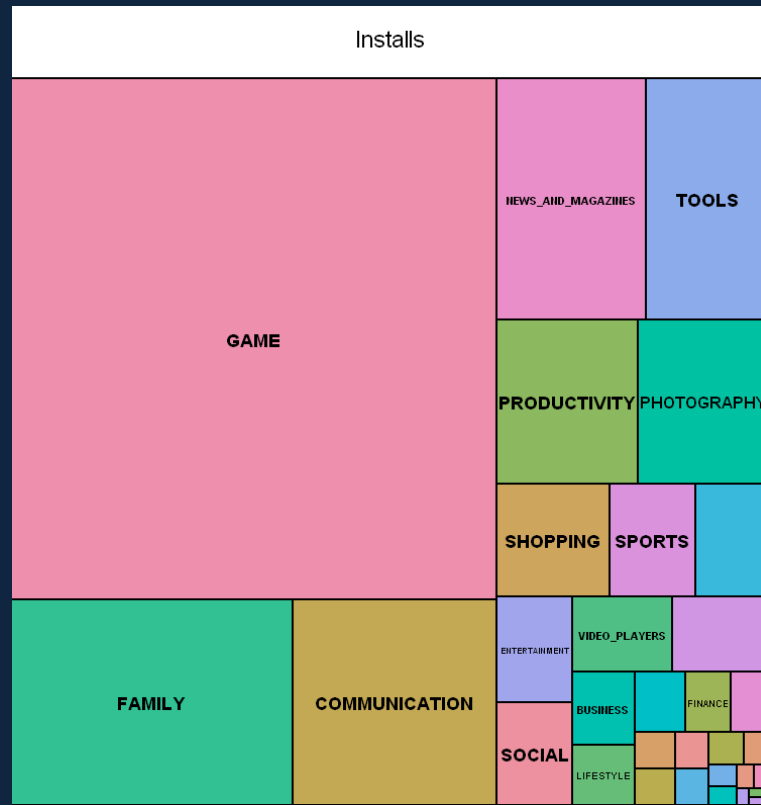
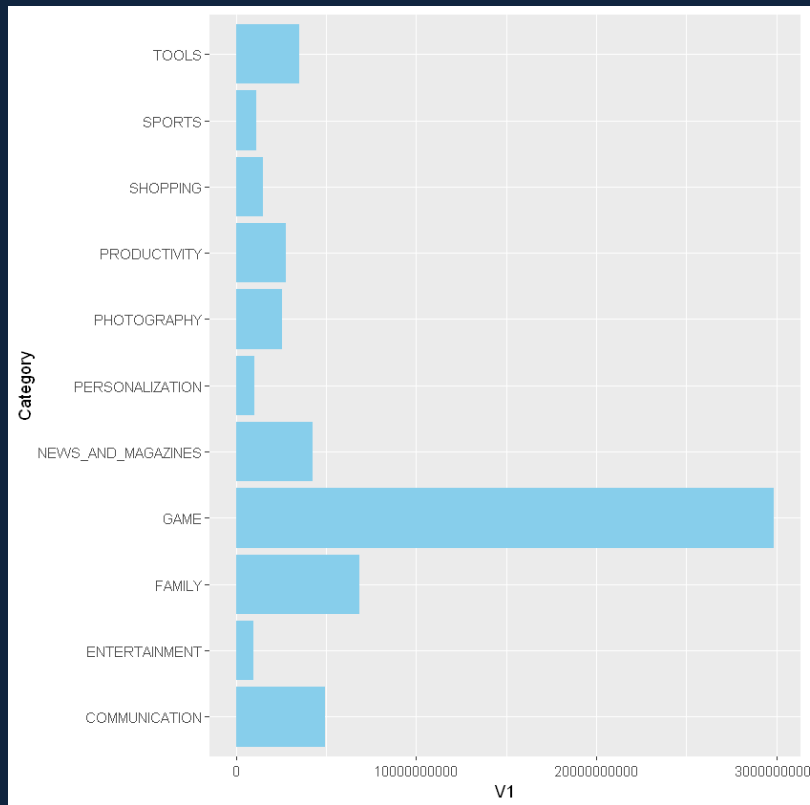
02. 변수-Category

FAMILY	1830
GAME	1016
TOOLS	739
MEDICAL	434
BUSINESS	400
PERSONALIZATION	350
Other	4369





02. 변수-Category



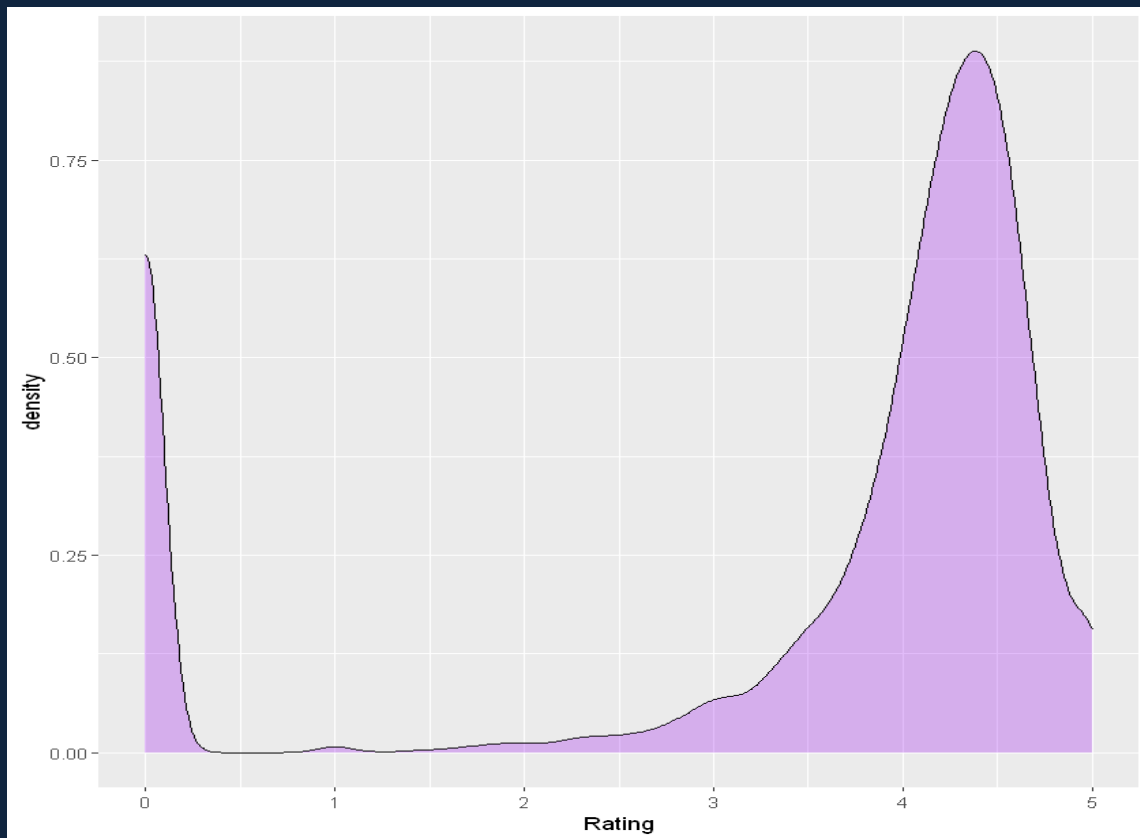


02. 변수-Rating

Rating	Rating
NaN	0
4.5	4.5
5	5
NaN	0
4.5	4.5
4.5	4.5



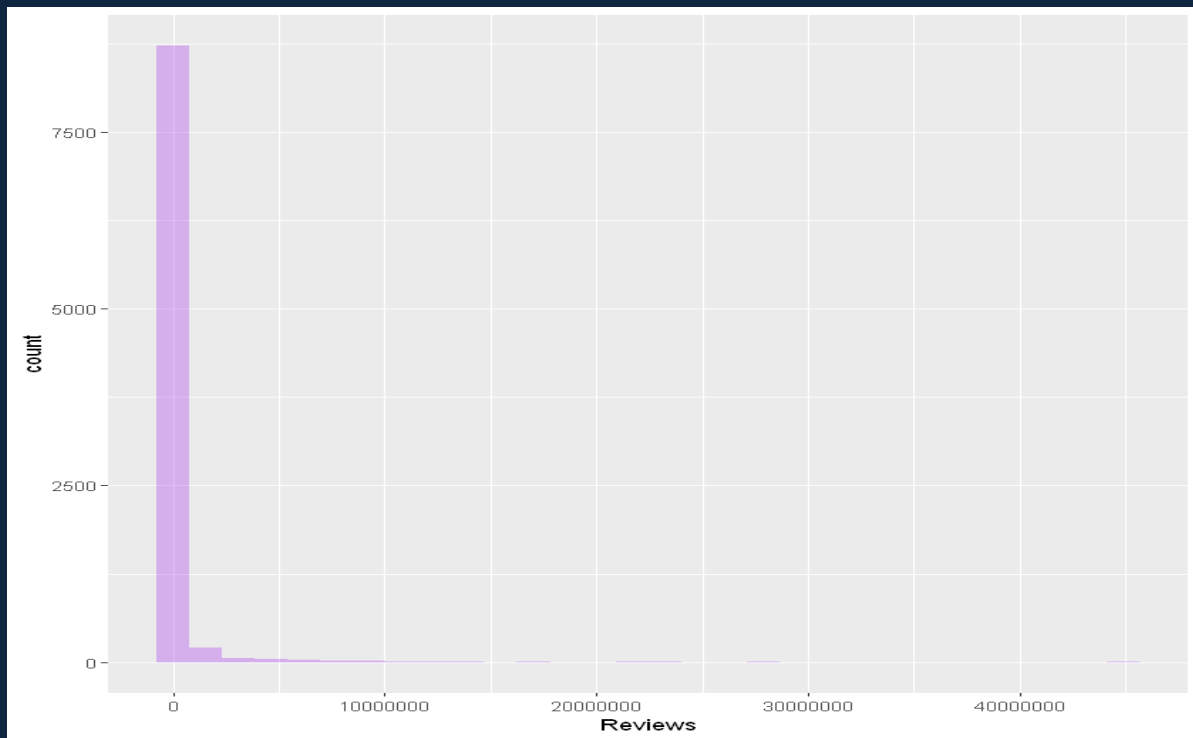
Min.	0
Median	4.0
Mean	3.529
Max.	5.0





02. 변수-Reviews

Min.	0
Median	750
Mean	249293
Max.	44893888



0~44,893,888



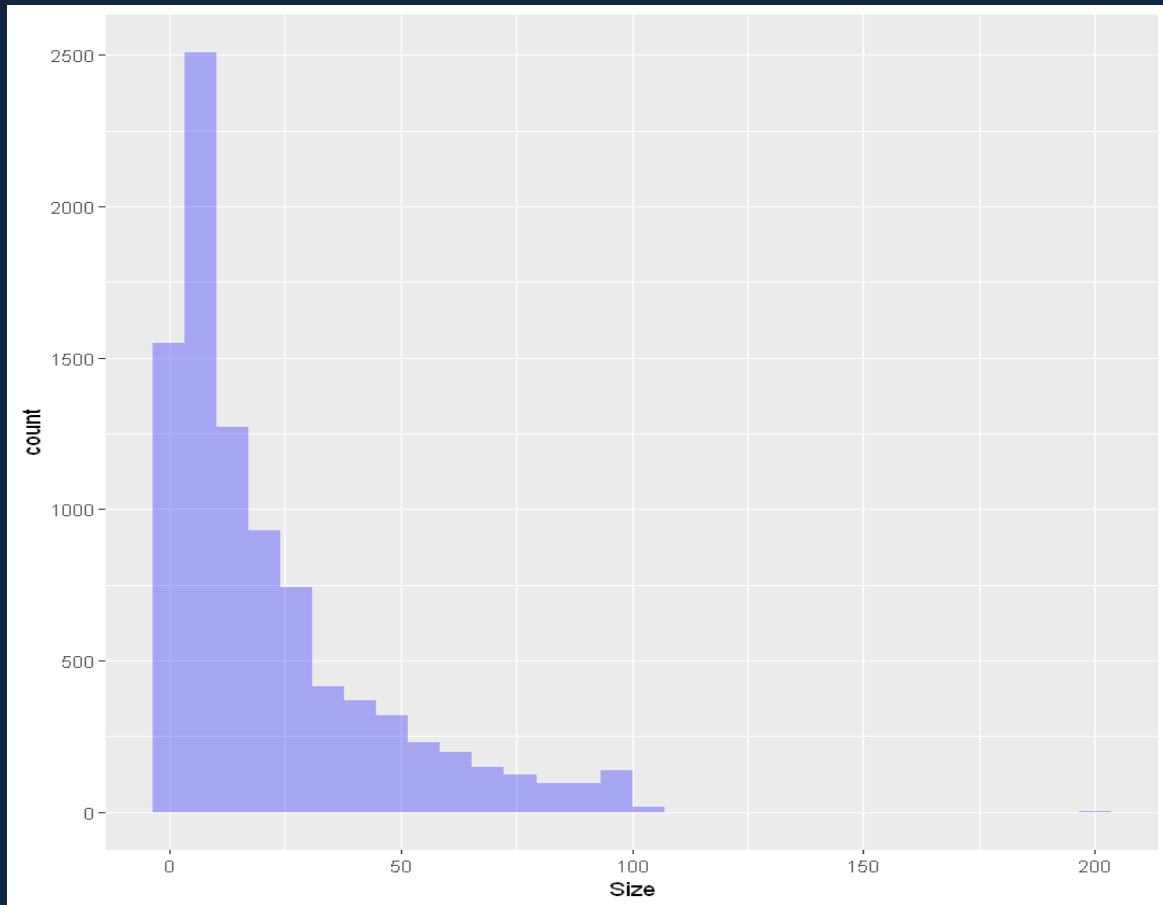
02. 변수-Size

Size
9.6M
53M
3.6M
9.5M
Varies with device
19M



Size
2.6
9.6
53.0
3.6
9.5
19.0

Min.	0
Median	13
Mean	21.5426
Max.	200



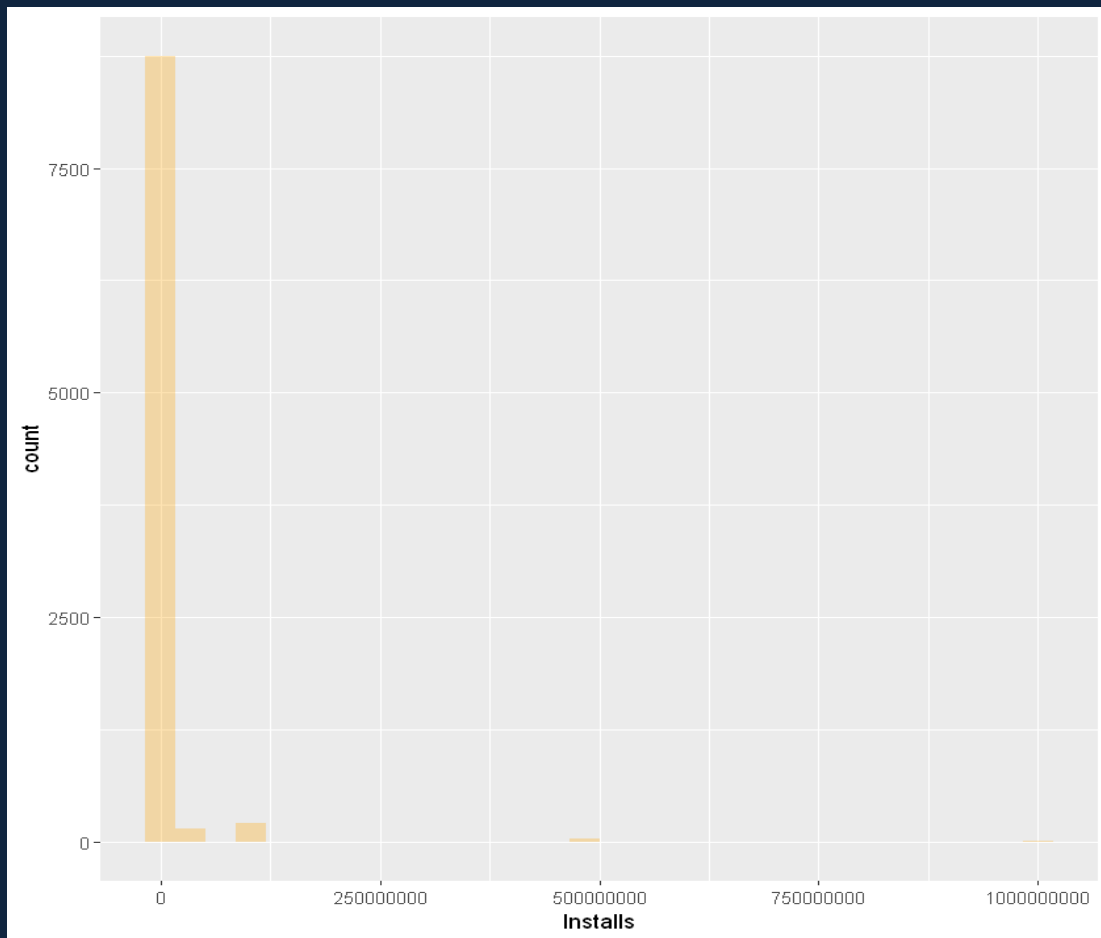


02. 변수-Installs

Installs	Installs
10,000+	10000
500,000+	500000
5,000,000+	5000000
50,000,000+	50000000
100,000+	100000
50,000+	50000



Min.	0
Median	100000
Mean	7126240
Max.	1000000000





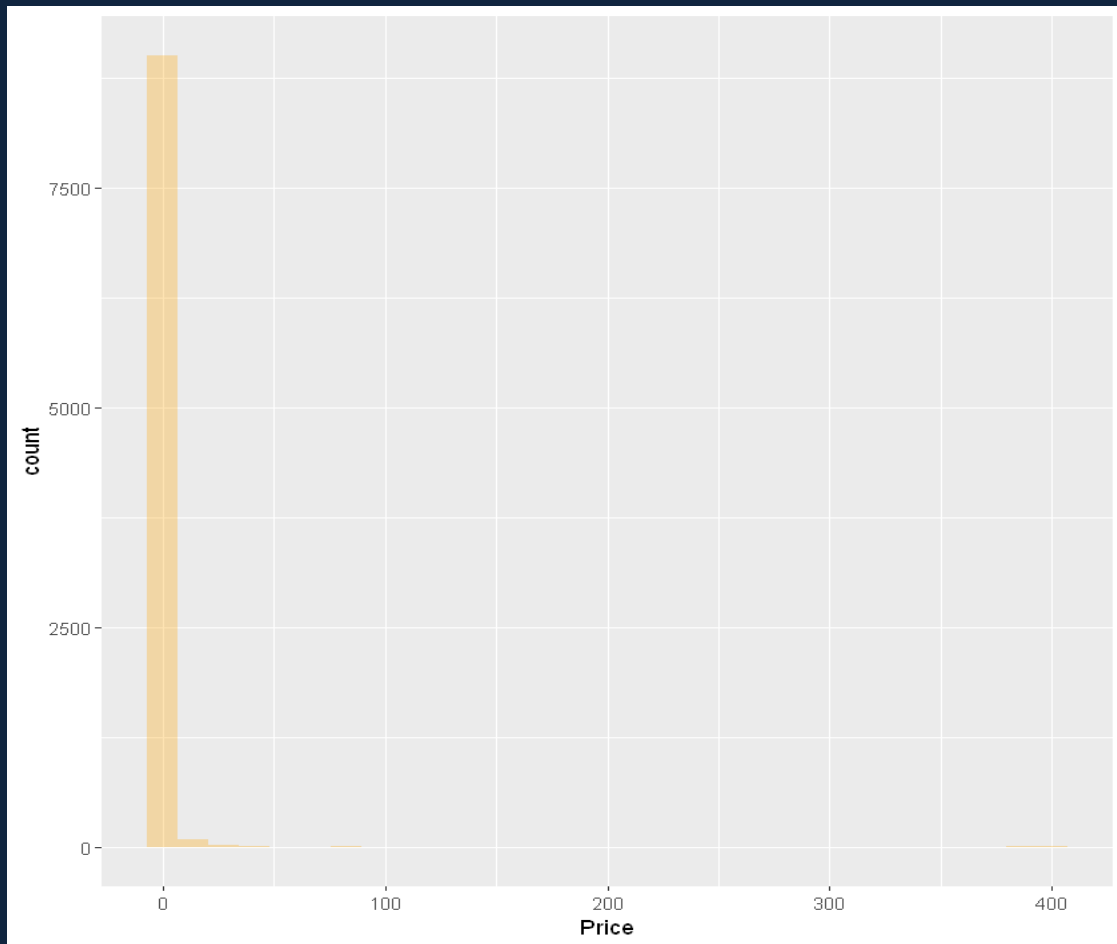
02. 변수-Price

Price
0
0
\$4.99
\$4.99
0
0



Price
0.00
0.00
4.99
4.99
0.00
0.00

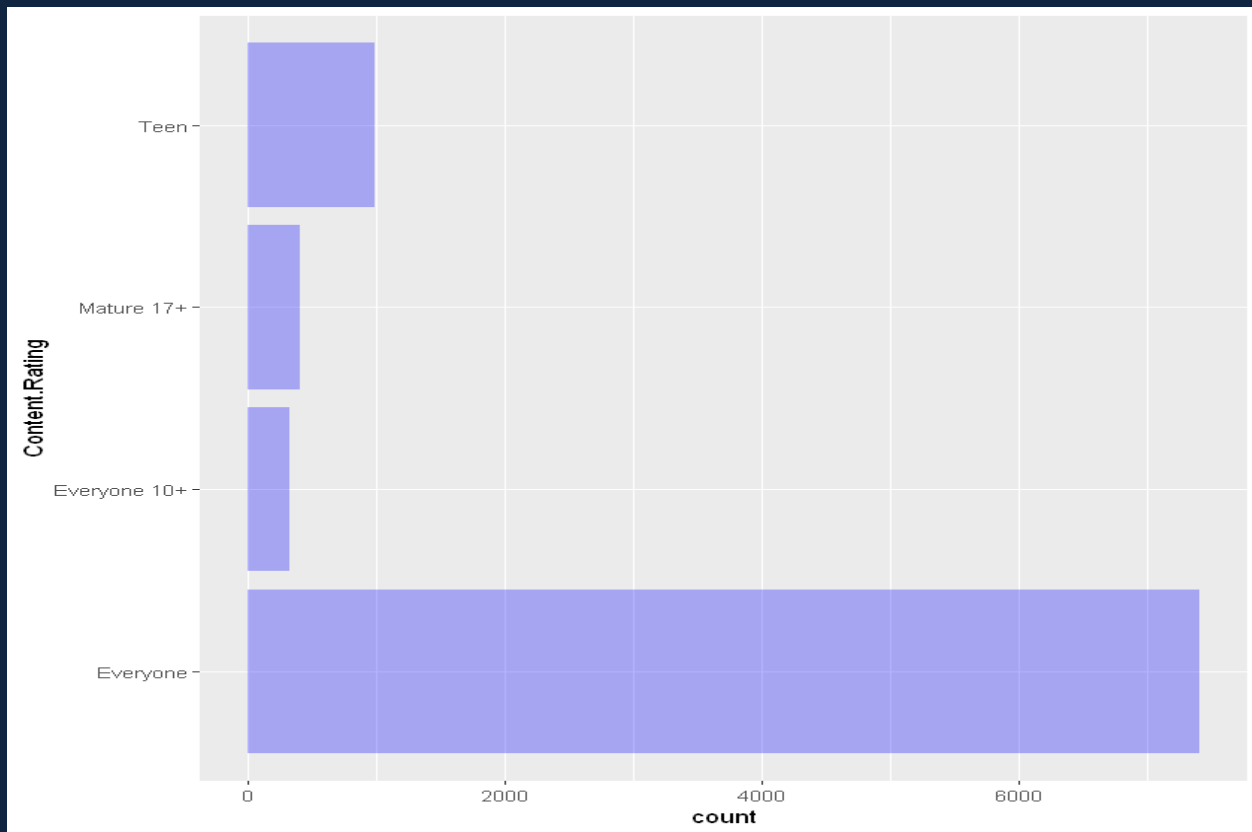
Min.	0
Median	0
Mean	1.183
Max.	400





02. 변수-Content.Rating

Everyone	7411
Everyone 10+	332
Mature 17+	408
Teen	987





02. 상관행렬

	Rating	Reviews	Size	Installs	Price
Rating	1.00000000	0.08405284	0.16665269	0.07916618	-0.01347589
Reviews	0.08405284	1.00000000	0.23738618	0.62769504	-0.00983128
Size	0.16665269	0.23738618	1.00000000	0.16405719	-0.02304647
Installs	0.07916618	0.62769504	0.16405719	1.00000000	-0.01047776
Price	-0.01347589	-0.00983128	-0.02304647	-0.01047776	1.00000000

(Installs, Reviews) : 강한 양의 상관관계
Price와 그 외 변수들 : 약한 음의 상관관계



02. 상관행렬





03. 주성분분석

	Comp. 1	Comp. 2	Comp. 3	Comp. 4	Comp. 5
Standard deviation	1.3323245	1.0248713	0.9972342	0.9014733	0.60615186
Proportion of Variance	0.3550177	0.2100722	0.1988952	0.1625308	0.07348402
Cumulative Proportion	0.3550177	0.5650900	0.7639852	0.9265160	1.00000000

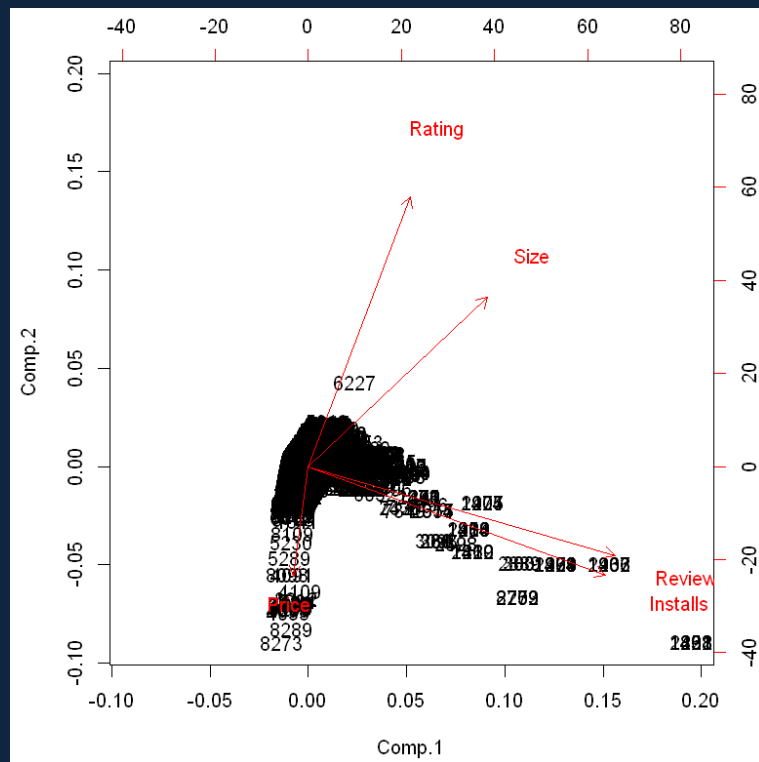
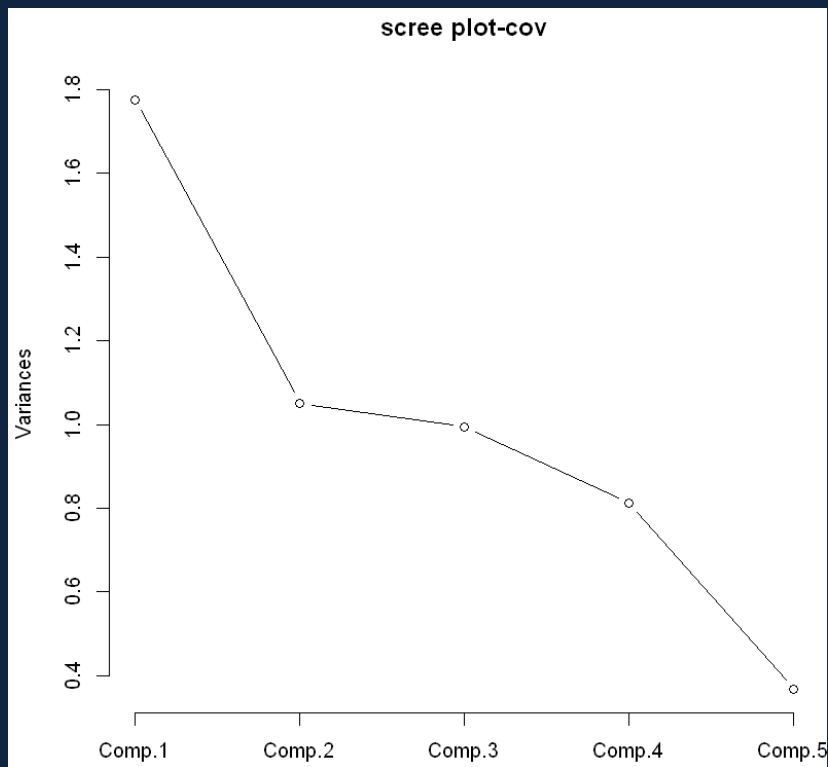
$$Y_1 = 0.216 * Rating + 0.646 * Reviews + 0.378 * Size + 0.626 * Installs$$

$$Y_2 = 0.739 * Rating - 0.243 * Reviews + 0.464 * Size - 0.299 * Installs - 0.3 * Price$$

$$Y_3 = 0.265 * Rating + 0.125 * Size - 0.952 * Price$$



03. 주성분분석





04. 계획

1. 인자분석과 군집분석을 실시한다.
- 2.. 새롭게 앱을 만들 때, 어떤 카테고리화 앱의 사이즈와 같은 어떤 앱의 특성을 가져야 인기가 있을 지 예측해본다.



감사합니다.