

Google Play poe rakenduste analüüs ja populaarseimad äpid

Kaisa Liina Keps ja Kaari Kuus

oktoober 2019

Sissejuhatus

Töö eesmärgiks on kinnistada aines Rakendustarkvara: R (LTMS.00.016) õpitud ja praktiseerida erinevaid R-i võimalusi läbi andmeanalüüsi. Selleks valisid autorid andmestiku Google Play poe rakenduste kohta ning püstitasid ka kaks hüpoteesi:

1. enim installitud äpid on mängud (keskmise järgi);
2. tasuliste äpide hinnang on kõrgem kui mittetasuliste.

Kuna andmestik on suur, siis et vältida projekti liiga suuremahulises kasvamisest ja laialivalgumist, esitasid autorid ka küsimused, millele soovisid andmestikust vastuseid leida:

1. Kui palju maksab kõige kallim äpp? Ja mis see äpp on?
2. Millised on kõige sagedasemad kategooriad? Millised on eri kategooria äppide hinnangud, kas need varieeruvad palju?

Analüüsi käigus soovitakse leida vastused hüpoteeside kehtivuse kohta ning uurida muud põnevat Google Play rakenduste kohta.

Andmestik ja selle kirjeldus

Andmestik kajastab Google Play poe rakenduste erinevaid tunnuseid (kokku 13 tunnust), milleks on äpi nimi (App), kategooria (Category), hinnang (Rating), arvustuste arv (Reviews), suurus (Size), installimiste arv (Installs), tüüp (Type), hind (Price), vanusgrupp (Content.Rating), žanrid (Genres), viimati värskendatud (Last.Updated), praegune versioon (Current.Ver) ja Androidi versioon (Android.Ver). Kõik tunnused on andmestikus toodud faktortunnustena v.a. tunnus Rating mis on numbriline. Vaatlusi ehk erinevaid äppe on tabelis kokku 10 481.

Kirjeldav analüüs

Esiteks uurime, millistesse kategooriatesse rakendused jagatud on. Selleks teeme sagedustabeli, kuid kuna meid huvitavad kategooriad, kus on äppe kõige rohkem, siis toome välja 10 suurima rakenduste arvuga kategooriat. Tabelist 2 näeme, et enim on äppe kategoorias “FAMILY” (1972) ning Google Play poes on

Table 1: Andmestiku päis

App	Category	Rating	Reviews	Size	Installs	Type	Price	Content.Rating	Genres	Last.Updated	Current.Ver	Android.Ver
Photo Editor & Camly Camera & Grid & Scrapbook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	January 7, 2018	1.0.0	4.0.3 and up
Coloring book msnm	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	2.0.0	4.0.3 and up
T Launcher Lite a€ FREE Live Cool Themes, Hide Apps	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	August 1, 2018	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	June 8, 2018	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	June 20, 2018	1.1	4.4 and up
Paper flowers instructions	ART_AND_DESIGN	4.4	167	5.6M	50,000+	Free	0	Everyone	Art & Design	March 26, 2017	1.0	2.3 and up

Table 2: Sagedustabel kategooriate kaupa

Kategooria	Sagedus
FAMILY	1972
GAME	1144
TOOLS	843
MEDICAL	463
BUSINESS	460
PRODUCTIVITY	424
PERSONALIZATION	392
COMMUNICATION	387
SPORTS	384
LIFESTYLE	382

ka palju mänge (“GAME, 1144). Kolmandalt kohalt äppide arvukuse poolest leiame kategooria”TOOLS” (843).

Uurime lähemalt äppide keskmist hinnangut kümnes kõige suuremas kategoorias (tabel 3).

Table 3: Sagedamate kategooriate keskmised hinnangud

Kategooria	Keskmine hinnang	Mitu
BUSINESS	4.3	460
COMMUNICATION	4.3	387
FAMILY	4.3	1972
GAME	4.4	1144
LIFESTYLE	4.2	382
MEDICAL	4.3	463
PERSONALIZATION	4.4	392
PRODUCTIVITY	4.3	424
SPORTS	4.3	384
TOOLS	4.2	843

Näeme, et kõige sagedamini esinevate kategooriate keskmised hinnangud on sarnased ning varieeruvad 4.2-4.4. Kõige kõrgem on see mängudel (kategooria GAME). Seega võime 10 sagedamini esinevat kategooriate põhjal öelda, et äppi kategooria ei mõjuta antavat hinnangut.

Esimese hüpoteesi kontrollimiseks peame moodustama tabeli (tabel 4), kus leiame kõikide kategoorias esinevate rakenduste installeerimiste keskmise arvu. Hindame installeerimiste arvu keskmise järgi, sest teame, et kategooriates on erinev arv äppe ja see mõjutaks väga palju installeerimiste koguarvu (summat). Hüpotees kehtib, kui kategoorial “GAME” on installimiste keskmine arv suurim. Tabelist 3 näeme, et meie hüpotees ei kehti, sest kõrgem keskmine on mitmel teisel kategoorial. Tabeliliidriks osutub tegelikult kategooria “COMMUNICATION”, teise koha võtab endale kategooria “SOCIAL” ning kolmandaks kasutavad inimesed ka kategooria “VIDEO_PLAYERS” äppe. Mängud asuvad alles viiendal positsioonil ning seega meie hüpotees kinnitust ei leidnud.

Ennist saime teada, et esimene hüpotees ei kehti, sest keskmine installeerimiste arv on suurim hoopis kategoorial “COMMUNICATION” (suhtlus).

Uurime nüüd lähemalt äppide installeerimiste arvu ja rakendusi, mida enim istallitud on (tabel 5).

On näha, et kõige suurem väärtus installeerimise arvule on “1,000,000,00+” ja sellisesse kategooriasse kuulub 58 rakendust. Veel näeme, et kõige rohkem (1578) Play poe äppe kuulub installeerimiste arvu pooles kategooriasse “1,000,000+”. Äppe millel on 0 installeerimist on vaid 1 ning äppe millel on 1 installeerimine (kategooria “0+”) on 14. See sõltub aga kindlasti ka sellest, millal anmdestik on koostatud, ning

Table 4: Installeerimiste keskmine arv kategooriate kaupa

	Kategooria	Installeerimiste_keskmine
8	COMMUNICATION	84359887+
29	SOCIAL	47694467+
34	VIDEO_PLAYERS	35554301+
27	PRODUCTIVITY	33434178+
16	GAME	30669602+
26	PHOTOGRAPHY	30114172+
32	TRAVEL_AND_LOCAL	26623594+
23	NEWS_AND_MAGAZINES	26488755+
11	ENTERTAINMENT	19256107+
31	TOOLS	13585732+

Table 5: Äppide installeerimised

Installeerimised	Mitu
0	1
0+	14
1,000,000,000+	58
1,000,000+	1579
1,000+	907
1+	67
10,000,000+	1252
10,000+	1054
10+	386
100,000,000+	409
100,000+	1169
100+	719
5,000,000+	752
5,000+	477
5+	82
50,000,000+	289
50,000+	479
50+	205
500,000,000+	72
500,000+	539
500+	330
Free	1

on suur tõenäosus, et need äpid pole jõudnud piisavalt kaua Play poes saadaval olla, et koguda suuremat installeerimiste koguarvu.

Edasi uurime 58 äppi, mida on kõige rohkem installeeritud.

Näeme enim installitud äppide kategooriate jaotust. Nagu ennist juba mainitud, siis on enim isntallitud äppid kategoorias “COMMUNICATION” ja “SOCIAL”. Jällegi leiab kinnitust, et meie püstitatud esimene hüpotees ei kehti, sest kategoorias GAME on enim installeeritud äppide hulgas vaid 6 äppi.

Uurime enim installeeritud äppide keskmisi hinnaguid ja arvustuste arvu kategooriate kaupa (tabel 7).

Tabelist (tabel 7) näeme, et kõige kõrgem keskmine hinnang on kategoorial “PHOTOGRAPHY” (fotograafia) ja “GAME” (mängud). Arvustuste poolest on aga kategooria “PHOTOGRAPHY” alla kuuluvatel äppi-

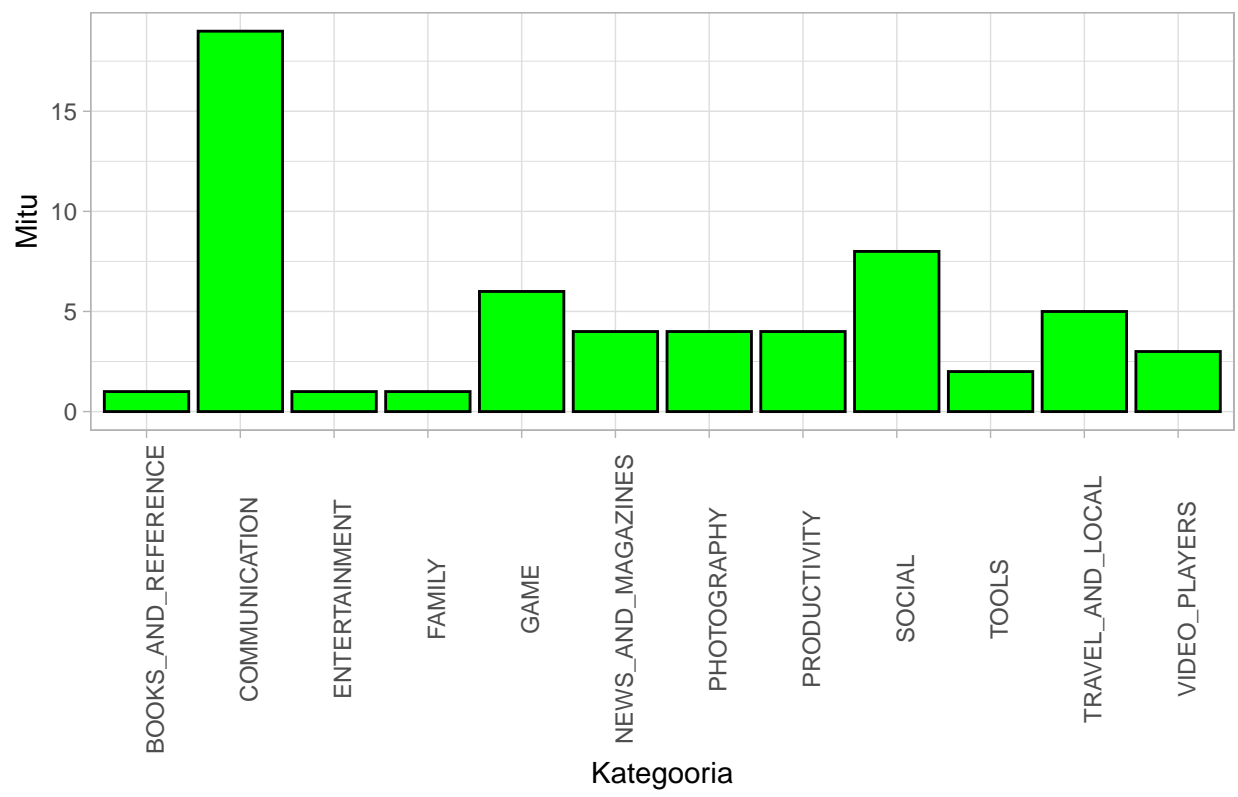


Figure 1: Populaarseimate äppide arv kategooriates

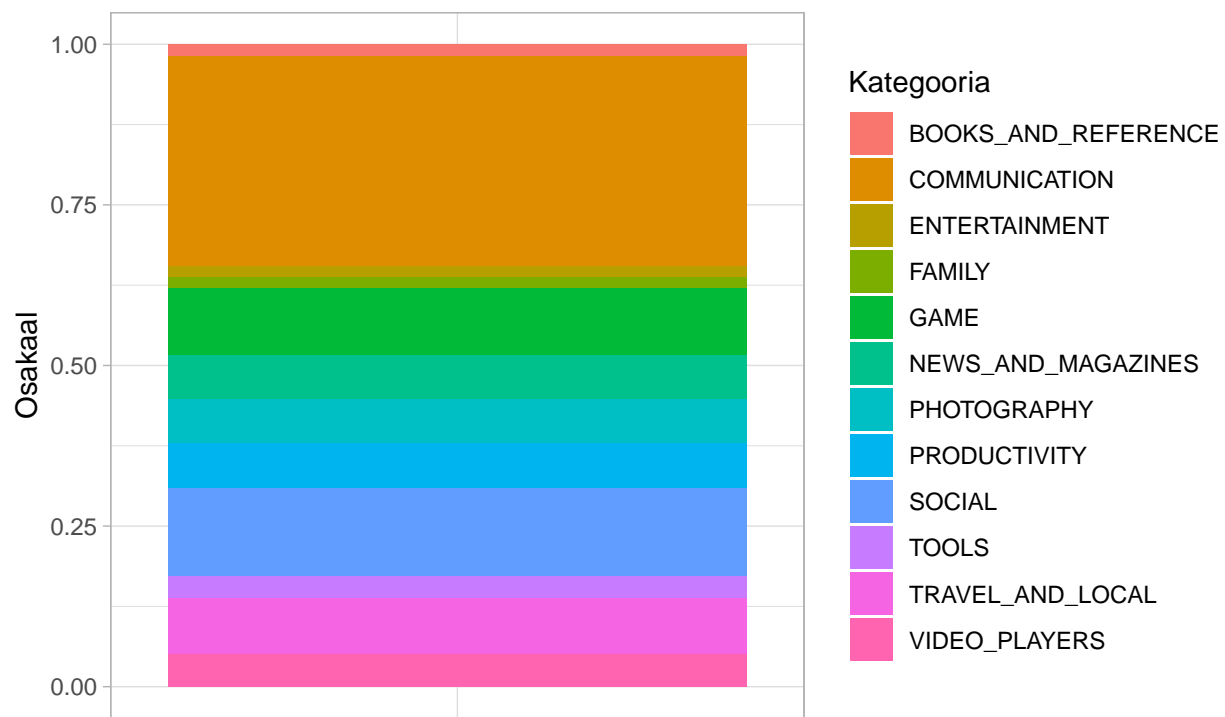


Figure 2: Populaarseimate rakenduste osakaal kategooriate kaupa

Table 6: Enim installitud äppide kategooriate sagedused

Kategooria	Mitu
BOOKS_AND_REFERENCE	1
COMMUNICATION	19
ENTERTAINMENT	1
FAMILY	1
GAME	6
NEWS_AND_MAGAZINES	4
PHOTOGRAPHY	4
PRODUCTIVITY	4
SOCIAL	8
TOOLS	2
TRAVEL_AND_LOCAL	5
VIDEO_PLAYERS	3

Table 7: Enim installeeritute äppide keskmine hinnang ja arvustuste arv kategooriate kaupa

Kategooria	Keskmine hinnang	Keskmine arvustuste arv	Mitu
BOOKS_AND_REFERENCE	3.90	911.0	1
COMMUNICATION	4.10	3992.0	19
ENTERTAINMENT	4.30	5165.0	1
FAMILY	4.30	5166.0	1
GAME	4.50	2633.5	6
NEWS_AND_MAGAZINES	3.90	5688.5	4
PHOTOGRAPHY	4.50	231.5	4
PRODUCTIVITY	4.40	2596.5	4
SOCIAL	4.35	4957.0	8
TOOLS	4.40	5453.5	2
TRAVEL_AND_LOCAL	4.30	5808.0	5
VIDEO_PLAYERS	4.30	2421.0	3

del kõige vähem arvustusi. Kõige rohkem arvustusi on kirjutatud kategooria “TRAVEL_AND_LOCAL” (reisimine ja paiksus) rakendustele.

Uurime, kas hinnangu ja arvustuste arvu vahel võib olla sõltuvus.

Näeme, et hajuvusdiagrammi (joonis 3) punktid on üsna laiali seega mingit tugevat seost me hinnangu ja arvustuste arvu vahel ei saa järeldada.

Kontrollime ka teist hüpoteesi ning kasutame selleks joonist 4.

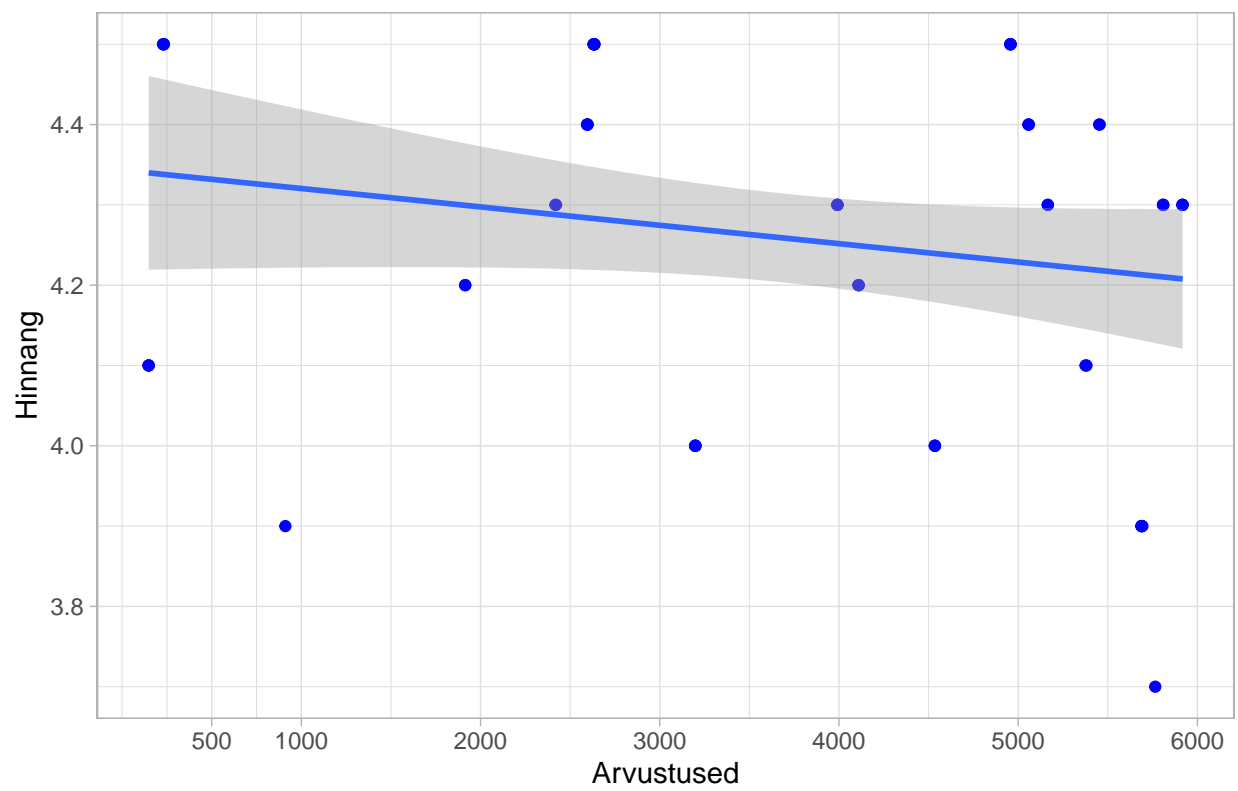


Figure 3: Arvustuste ja hinnangute hajuvusdiagramm

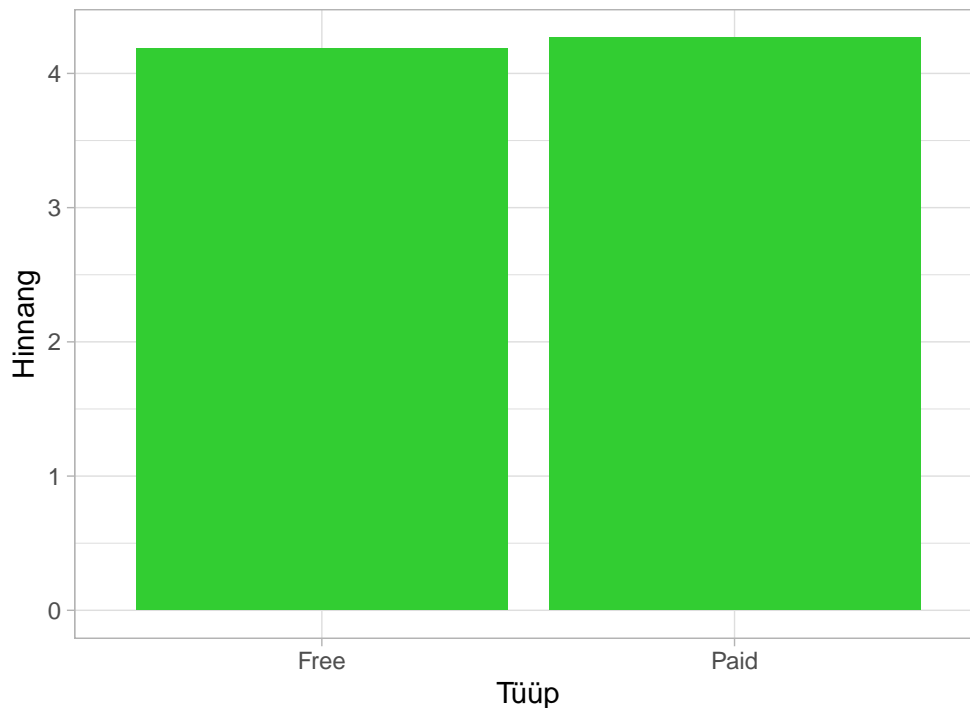


Figure 4: Tasuliste ja tasuta äppide keskmine hinnang

Jooniselt on näha, et kuigi tasuliste (Paid) ja tasuta (Free) rakenduste keskmised hinnangud on üpris sarnased, on tasuliste äppide oma natuke kõrgem. Seega oli meie teine hüpotees õige, sest tasuliste äppide hinnang on veidi parem kui tasuta rakenduste oma.

Uurime lisaks, kellele on rakendused suunatud ning andmete uurimise lihtsustamiseks teeme sagedustabeli vanusgruppide (Content.Rating) kaupa (tabel 8).

Table 8: Sagedustabel vanusgruppidele

Vanusgrupp	Äppide arv
Everyone	8714
Teen	1208
Mature 17+	499
Everyone 10+	414
Adults only 18+	3
Unrated	2
	1

Näeme, et rakenduste arv, mille sisu on hinnatud sobivaks kõigile vanustele, on ülekaalukalt suurim (8714 äppi). Järgnevad teismelistele (Teen) suunatud rakendused (1208) ning siis üle 17-aastastele (Mature 17+) mõeldud äpid (499). Lisaks sellele saame teada, et leidub kaks telefonirakendust, mille sisu sobivust erinevatele vanusgruppidele pole hinnatud, ning eksisteerib ka äpp, mille vanusgrupi kohta puudub info, sest tõenäoliselt on tunnus Content.Rating valesti sisestatud.

Uurime lähemalt tasulisi äppe ja nende maksuvust.

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.99   1.49    2.99   13.92   4.99  400.00
```


Table 9: Kõige kallim äpp

	Category	Rating	Reviews	Installs	Price
216	LIFESTYLE	3.6	275	10,000+	\$400.00

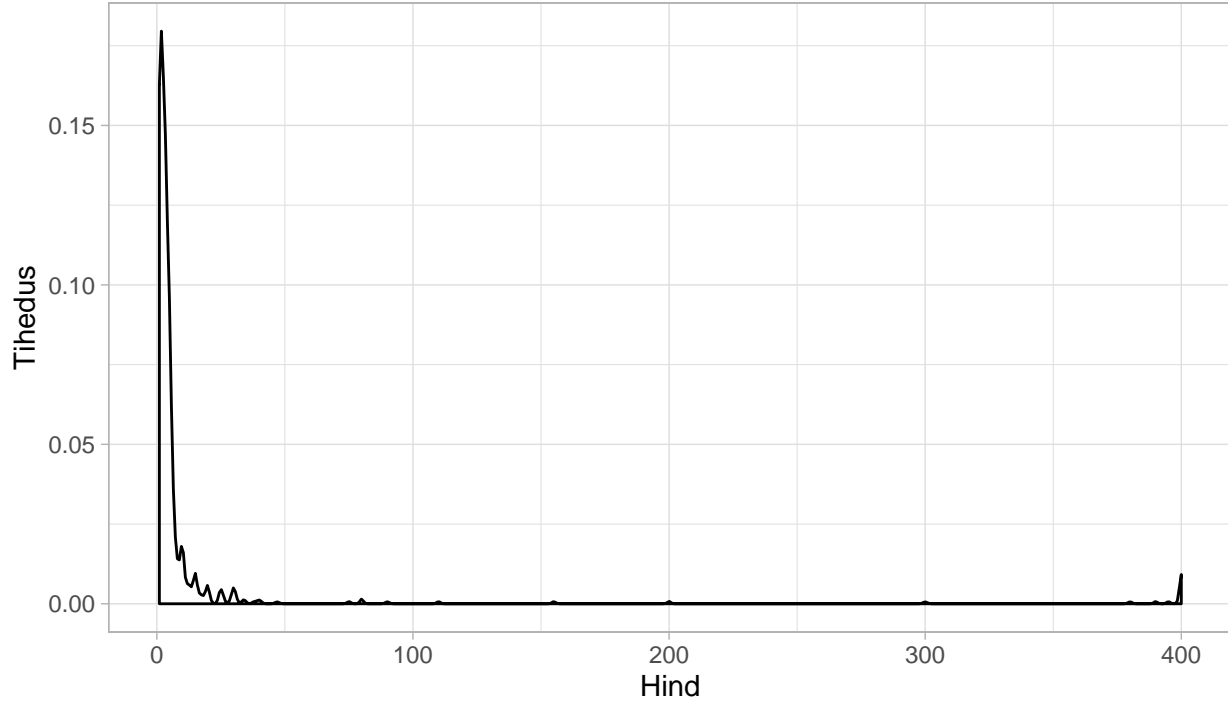


Figure 5: Tasuliste äppide hinna tihedus

Kõige odavam tasuline äpp maksab 0.99\$ ja kõige kallim 400.00\$. Kõige kallima rakenduse nimi on “I’m Rich - Trump Edition”, millel on 10,000+ installeerimisi. Keskmine tasulise äppi maksumus tuleb lausa 13.92 dollarit. Autorite jaoks on suur üllatus, et leidub niivõrd palju äppe, mille eest inimesed on nõus välja käima suurt raha. Näeme, et enamus äppid on odavamad aga leidub ka väikseid “võnkeid” suuremate väärtuste poolel. Uurime edasi, selleks jagame objektid kahte gruppi: hind 10 dollarit või madalam ning hind üle 10 dollari.

```
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  10.99  14.99   24.99   98.98  79.99  400.00
```

```
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.990   1.490   2.990   3.274   4.490  10.000
```

On näha, et suurem osa üle 10\$ maksvate äppide hinnast kuulub vahemikku 10-100 dollarit. Samas on palju ka äppe, mille hind jääb 400 dollari lähistele.

Vaatame, kas 10 dollarit või madalama hinnaga äppide hinna ja hinnangu vahel on seos.

Näeme, et hajuvusdiagrammi (joonis 3) punktidest ei tule välja tugevat seost.

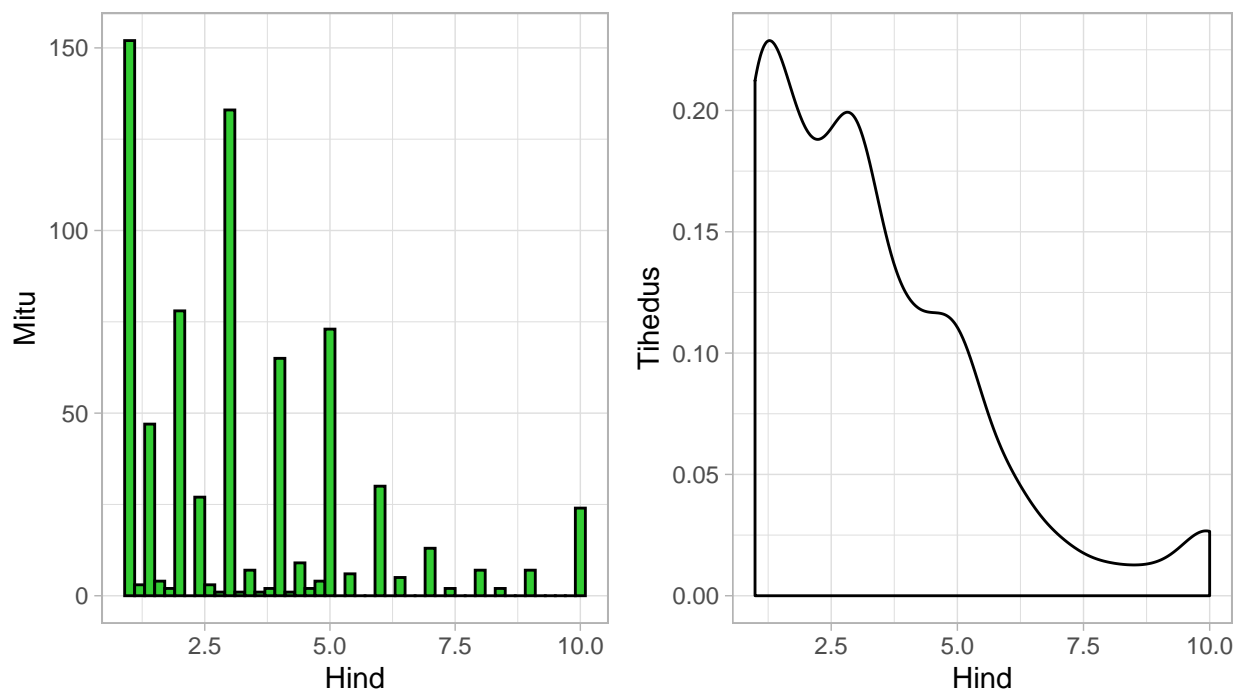


Figure 6: Alla 10\$ maksvad äpid

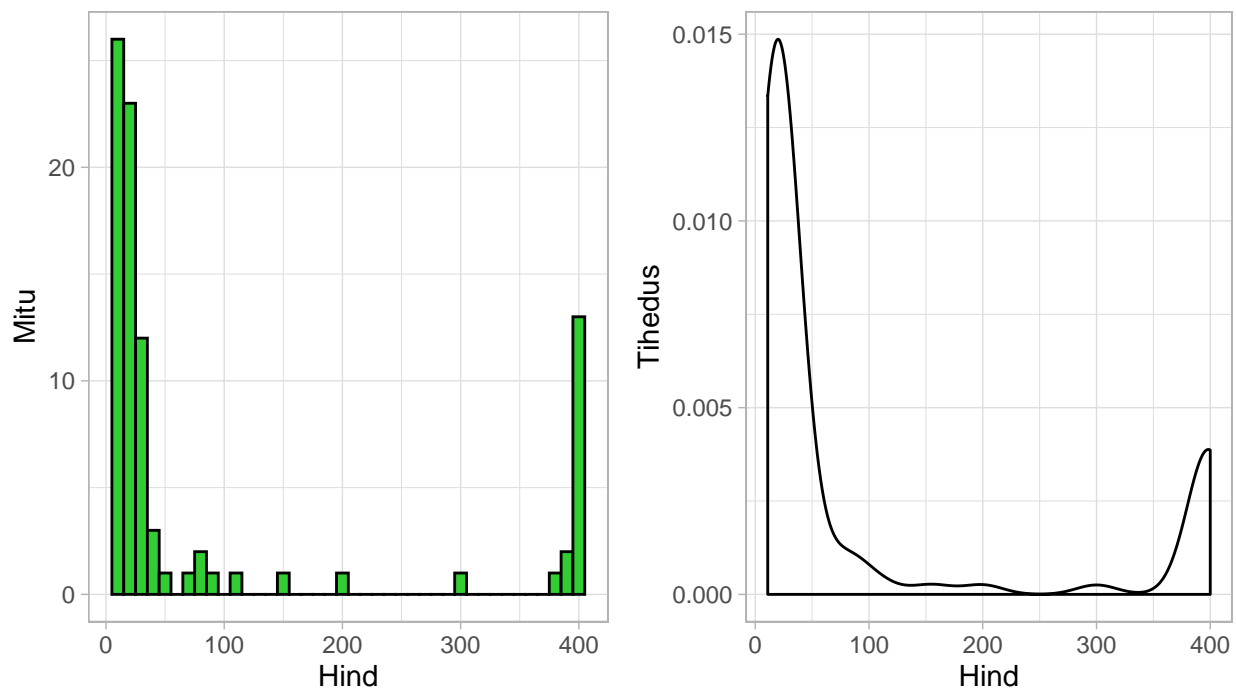


Figure 7: Hind ja üle 10\$ maksvad äpid

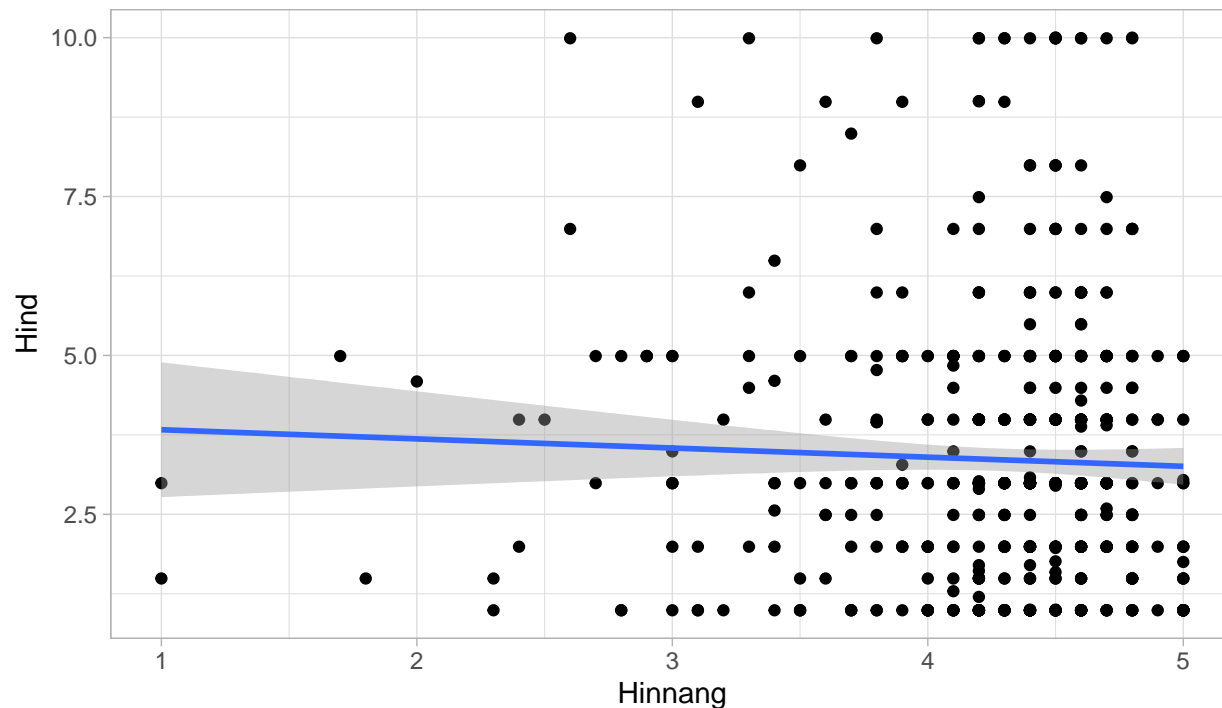


Figure 8: Hinna ja hinnagu hajuvusdiagramm

Kokkuvõte

Kahest püstitatud hüpoteesist said autorid kinnitust vaid teisele ehk tasuliste äppide hinnang on andmete põhjal kõrgem kui tasuta rakenduste oma. Esimene hüpotees kinnitust ei leidnud, sest kõige kõrgem keskmine installimiste arv ei kuulunud kategooriale “GAME”, vaid hoopis kategooriale “COMMUNICATION”. Selgus ka, et kõige kallim rakendus maksab lausa 400\$ ja kannab nime “I’m Rich - Trump Edition” ja omab üllatavalt palju isntalleerimiste ja arvustuste arvu. Autorid said teada huvitavat infot, et Play poes leidub ka väga kalleid äppe, mis maksavad sadu dollareid, mida nad ei oodanud. Saadi teada ka, et erinevates kategooriates keskmine hinnang ei varieeru märgatavalt ja üldiselt ei ole kategooria hinnangu osas määrav.