Analysis of purchases of mobile applications

Michał Szałański, Tuan Doan, Aleksandra Bednarczuk

1. INTRODUCTION

1.1. MOBILE APPLICATIONS

Mobile applications (or mobile apps) are computer programs or software applications created to be used on mobile phones, tablets or smartwatches. Originally, they were designed for basic needs like e-mails or calendar. But with the development of mobile devices, there has been an increase in demand for new and life-easing applications such as mobile games, social networking apps, navigation and location-based services. Those applications can be free of charge or can have a price. Mobile applications can be downloaded from distribution platforms like App Store or Google Play Store. According to the blogpost from the Business of Apps website, people on avarage use 9 mobile apps per day and 30 mobile apps per month. The most popular apps are Facebook, Youtube, Messenger, Google Search and Google Maps.

1.2. RESEARCH OBJECTIVE

The purpose of our research is to discover the factors affecting the number of application installs - what determines the popularity of mobile applications?

At the beginning, we carry out data cleaning process and preparation of the dataset for analysis. Then we conduct a preliminary analysis and visualize basic statistics. Next, we formulate the model to inspect the factors determining the number of installs.

1.3. DATASET

The dataset about Google play store apps was downloaded from kaggle.com website (to see the source click here (https://www.kaggle.com/lava18/google-play-store-apps?fbclid=IwAR0ovxDCXAqDkyunIhSRKqKWRbX5KXnjR33KL_RUroKBwc1dcf5D_h5572A)). The dataset includes almost 11 thousand observations - mobile applications, and 13 variables - Application name, Category, Rating, Reviews, Size, Installs, Type, Price, Content Rating, Genres, Last Updated, Current version, Android version.

2. DATA ANALYSIS

2.1. CLEANING THE DATA

Raw data downloaded from kaggle repository needs cleaning. After uploading the data from the csv file, NaN values were automatically omitted.

Based on the summary and basic statistics we can assume that most of the mobile applications are free (8719 apps) and only 648 are paid. Because of this disproportion, we do not pay attention to the level of price for the mobile app. Instead, we use the variable Type to distinguish between free and paid apps.

```
## 'data.frame':    9367 obs. of  13 variables:
##  $ App           : Factor w/ 9660 levels "- Free Comics - Comic Apps",..: 7224 2558 8998 8
113 7289 7120 8171 5585 4944 5822 ...
##  $ Category      : Factor w/ 34 levels "1.9","ART_AND_DESIGN",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ Rating        : num  4.1 3.9 4.7 4.5 4.3 4.4 3.8 4.1 4.4 4.7 ...
##  $ Reviews       : Factor w/ 6002 levels "0","1","10","100",..: 1183 5924 5681 1947 5924 1
310 1464 3385 816 485 ...
##  $ Size          : Factor w/ 462 levels "1,000+","1.0M",..: 55 30 368 102 64 222 55 118 14
6 120 ...
##  $ Installs      : Factor w/ 22 levels "0","0+","1,000,000,000+",..: 8 20 13 16 11 17 17 4
4 8 ...
##  $ Type          : Factor w/ 4 levels "0","Free","NaN",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ Price         : Factor w/ 93 levels "$0.99","$1.00",..: 92 92 92 92 92 92 92 92 92 92
...
##  $ Content.Rating: Factor w/ 7 levels "","Adults only 18+",..: 3 3 3 6 3 3 3 3 3 3 ...
##  $ Genres        : Factor w/ 120 levels "Action","Action;Action & Adventure",..: 10 13 10
10 12 10 10 10 10 12 ...
##  $ Last.Updated  : Factor w/ 1378 levels "1.0.19","April 1, 2016",..: 562 482 117 825 757
901 76 726 1317 670 ...
##  $ Current.Ver   : Factor w/ 2834 levels "","0.0.0.2","0.0.1",..: 121 1020 466 2827 279 11
5 279 2393 1457 1431 ...
##  $ Android.Ver   : Factor w/ 35 levels "","1.0 and up",..: 17 17 17 20 22 10 17 20 12 17
...
##  - attr(*, "na.action")= 'omit' Named int  24 114 124 127 130 131 135 164 181 186 ...
##   ..- attr(*, "names")= chr  "24" "114" "124" "127" ...
```

```
##                                                 App
##  ROBLOX                                           :   9
##  CBS Sports App - Scores, News, Stats & Watch Live:   8
##  8 Ball Pool                                      :   7
##  Candy Crush Saga                                 :   7
##  Duolingo: Learn Languages Free                   :   7
##  ESPN                                             :   7
##  (Other)                                          :9322
##           Category        Rating          Reviews
##  FAMILY       :1747   Min.   : 1.000   2      :  83
##  GAME         :1097   1st Qu.: 4.000   3      :  78
##  TOOLS        : 734   Median : 4.300   4      :  74
##  PRODUCTIVITY : 351   Mean   : 4.193   5      :  74
##  MEDICAL      : 350   3rd Qu.: 4.500   1      :  67
##  COMMUNICATION: 328   Max.   :19.000   6      :  62
##  (Other)      :4760                    (Other):8929
##                  Size             Installs      Type          Price
##  Varies with device:1637   1,000,000+ :1577   0   :   1   0      :8719
##  14M               : 166   10,000,000+:1252   Free:8719   $2.99  : 114
##  12M               : 161   100,000+   :1150   NaN :   0   $0.99  : 107
##  11M               : 160   10,000+    :1010   Paid: 647   $4.99  :  70
##  15M               : 159   5,000,000+ : 752               $1.99  :  59
##  13M               : 157   1,000+     : 713               $3.99  :  58
##  (Other)           :6927   (Other)    :2913               (Other): 240
##          Content.Rating           Genres             Last.Updated
##                 :   1   Tools        : 733   August 3, 2018: 319
##  Adults only 18+:   3   Entertainment: 533   August 2, 2018: 284
##  Everyone       :7420   Education    : 468   July 31, 2018 : 279
##  Everyone 10+   : 397   Action       : 358   August 1, 2018: 275
##  Mature 17+     : 461   Productivity : 351   July 30, 2018 : 199
##  Teen           :1084   Medical      : 350   July 25, 2018 : 157
##  Unrated        :   1   (Other)      :6574   (Other)       :7854
##              Current.Ver               Android.Ver
##  Varies with device:1415   4.1 and up        :2059
##  1.0               : 458   Varies with device:1319
##  1.1               : 195   4.0.3 and up      :1240
##  1.2               : 126   4.0 and up        :1131
##  1.3               : 120   4.4 and up        : 875
##  2.0               : 119   2.3 and up        : 582
##  (Other)           :6934   (Other)           :2161
```

After analysis of the data structure, we rename the columns to make data more readable:

```
columnsToRename = c(Reviews = "Reviews.Count", Current.Ver = "Current.Software.Version",
    Android.Ver = "Android.Version")
mainTable <- mainTable %>% plyr::rename(columnsToRename)
```

Furthermore, we start the process of cleaning - fixing the data and changing the data types:

```
# All - Setting 'Varies with device' as NaN
mainTable[mainTable == "Varies with device"] <- NA
# Rating - Fixing outliers
mainTable$Rating[mainTable$Rating > 5] <- NA
# Size - Delete M (megabytes)
mainTable[5] <- lapply(mainTable[5], as.character)
mainTable$Size <- substr(mainTable$Size, 1, nchar(mainTable$Size) -
    1)
# Type - Fixing the 0
mainTable$Type[mainTable$Type == 0] <- "Free"
# Price - Delete dollars, fix outliers
mainTable$Price <- substring(mainTable$Price, 2)
mainTable$Price[mainTable$Price == "" | mainTable$Price == "veryone"] <- "0"
# Content.Rating - Deleting outliers
mainTable$Content.Rating[mainTable$Content.Rating == "" | mainTable$Content.Rating ==
    "Unrated"] <- "Everyone"
# Genres - Taking only the main genre
mainTable$Genres <- gsub(";.*", "", mainTable$Genres)
# Genres - Education and Educational are the same type of
# apps
mainTable$Genres[mainTable$Genres == "Educational"] <- "Education"
# Last.Updated - Converting to date NOT WORKING YET!
mainTable <- mainTable[!mainTable$Last.Updated == "1.0.19", ]
mainTable$Last.Updated <- gsub(",", "", mainTable$Last.Updated)
# placeholder <- as.Date(mainTable$Last.Updated, format = '%B
# %d %Y', optional = TRUE) Android.Version - Take only the
# main part, eg. 4.3
mainTable$Android.Version <- substr(mainTable$Android.Version,
    1, 3)
# Convert to character
for (i in c(1, 4)) {
    mainTable[i] <- lapply(mainTable[i], as.character)
}
# Convert to numeric
for (i in c(4, 5, 8)) {
    mainTable[i] <- lapply(mainTable[i], as.numeric)
}
# Convert to factors
for (i in c(10, 13)) {
    mainTable[i] <- lapply(mainTable[i], as.factor)
}
remove(i)
# Drop all unused factors (cool function)
mainTable <- droplevels.data.frame(mainTable)
# Drop the Category column
mainTable <- mainTable[-c(2, 12)]
```

2.2. BASIC STATISTICS

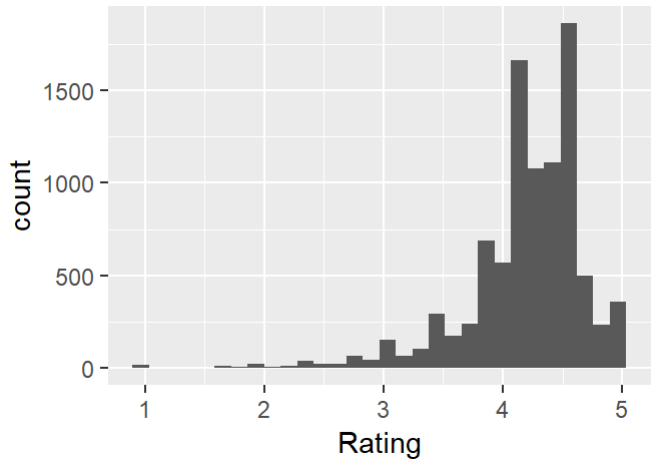
After the process of cleaning, the dataset includes 9367 observations and 11 variables. We can take another look at its structure and summary statistics:

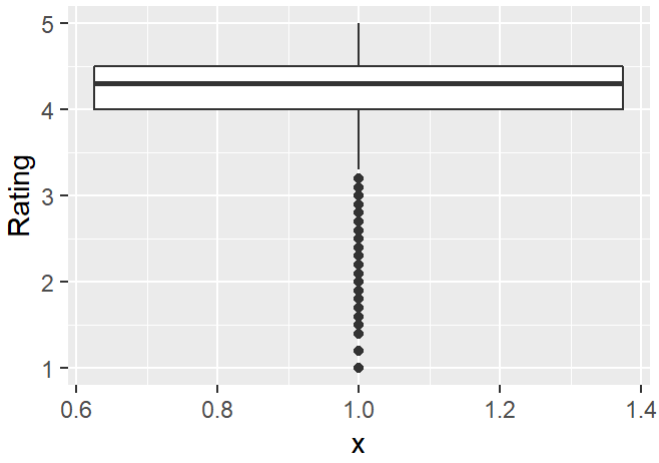
```
## 'data.frame':    9366 obs. of  11 variables:
##  $ App            : chr  "Photo Editor & Candy Camera & Grid & ScrapBook" "Coloring book m
oana" "U Launcher Lite â\200“ FREE Live Cool Themes, Hide Apps" "Sketch - Draw & Paint" ...
##  $ Rating         : num  4.1 3.9 4.7 4.5 4.3 4.4 3.8 4.1 4.4 4.7 ...
##  $ Reviews.Count  : num  159 967 87510 215644 967 ...
##  $ Size           : num  19 14 8.7 25 2.8 5.6 19 29 33 3.1 ...
##  $ Installs       : Factor w/ 19 levels "1,000,000,000+",..: 6 18 11 14 9 15 15 2 2 6 ...
##  $ Type           : Factor w/ 2 levels "Free","Paid": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Price          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Content.Rating : Factor w/ 5 levels "Adults only 18+",..: 2 2 2 5 2 2 2 2 2 2 ...
##  $ Genres         : Factor w/ 47 levels "Action","Adventure",..: 4 4 4 4 4 4 4 4 4 4 ...
##  $ Last.Updated   : chr  "January 7 2018" "January 15 2018" "August 1 2018" "June 8 2018"
...
##  $ Android.Version: Factor w/ 22 levels "1.0","1.5","1.6",..: 11 11 11 13 15 7 11 13 8 11
...
```

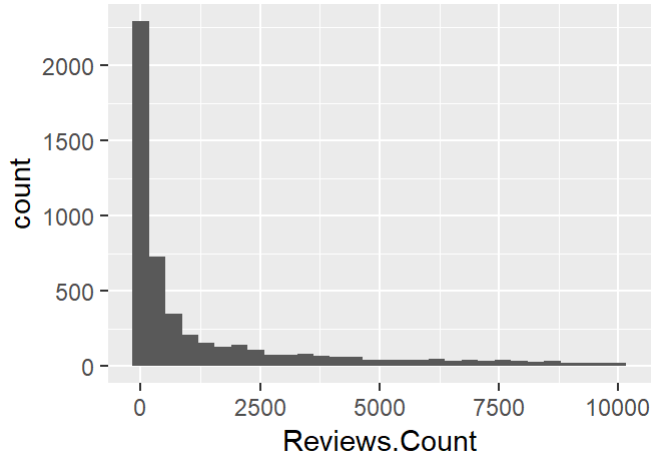
```
##      App                Rating      Reviews.Count           Size
##  Length:9366        Min.   :1.000   Min.   :       1   Min.   :  1.00
##  Class :character   1st Qu.:4.000   1st Qu.:     186   1st Qu.:  6.10
##  Mode  :character   Median :4.300   Median :    5930   Median : 16.00
##                     Mean   :4.192   Mean   :  514050   Mean   : 37.28
##                     3rd Qu.:4.500   3rd Qu.:   81533   3rd Qu.: 37.00
##                     Max.   :5.000   Max.   :78158306   Max.   :994.00
##                                                        NA's   :1637
##         Installs      Type          Price                Content.Rating
##  1,000,000+ :1577   Free:8719   Min.   :  0.0000   Adults only 18+:   3
##  10,000,000+:1252   Paid: 647   1st Qu.:  0.0000   Everyone       :7421
##  100,000+   :1150               Median :  0.0000   Everyone 10+   : 397
##  10,000+    :1010               Mean   :  0.9609   Mature 17+     : 461
##  5,000,000+ : 752               3rd Qu.:  0.0000   Teen           :1084
##  1,000+     : 713               Max.   :400.0000
##  (Other)    :2912
##            Genres     Last.Updated       Android.Version
##  Tools        : 734   Length:9366        4.0    :2373
##  Education    : 666   Class :character   4.1    :2060
##  Entertainment: 577   Mode  :character   4.4    : 881
##  Action       : 375                      2.3    : 822
##  Productivity : 351                      5.0    : 538
##  Medical      : 350                      (Other):1373
##  (Other)      :6313                      NA's   :1319
```

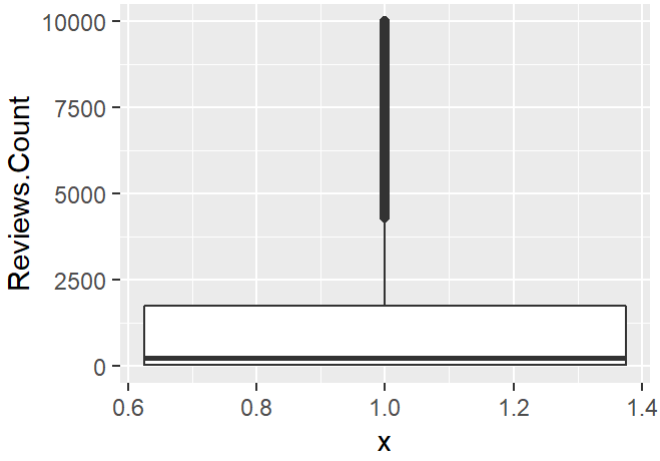
The dataset contains 11 variables - 5 factors, 4 numerics and 2 characters, and 9366 observations.

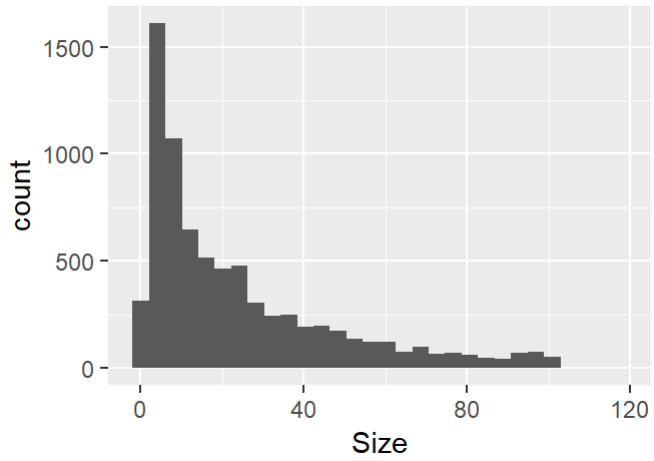
- Median rating of apps in Google Play Store is 4.3 and 50% of the apps have a rating between 4 and 4.5. We can identify outliers - apps with rating lower than 3.25.
- Most of the apps have low number of reviews - 50% of the applications in the dataset have between 186 and 81533 reviews (with median of 5930 reviews). The highest number of reviews is 78158306.
- Median size of apps is 16 MB. 50% of apps have the size between 6 and 37 MB. The largest app in the Google Play Store has 994 MB.
- As we mentioned before, most of the apps are free of charge.

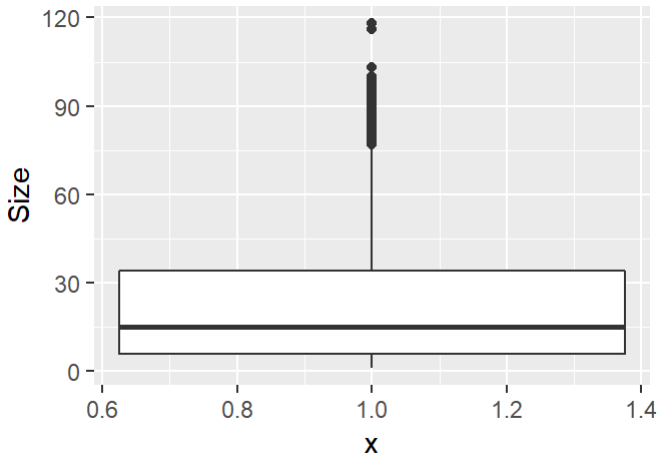


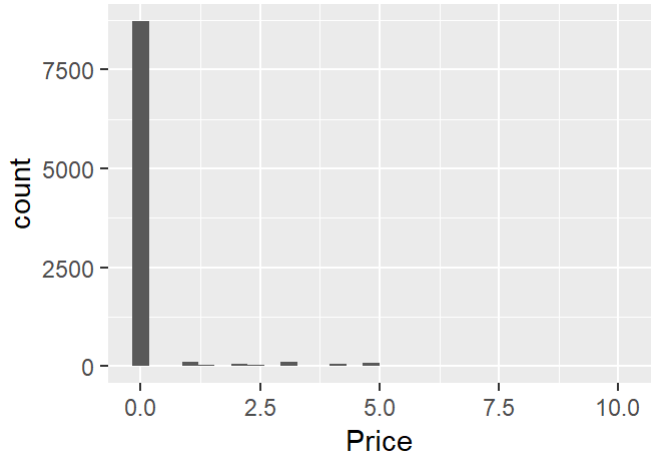


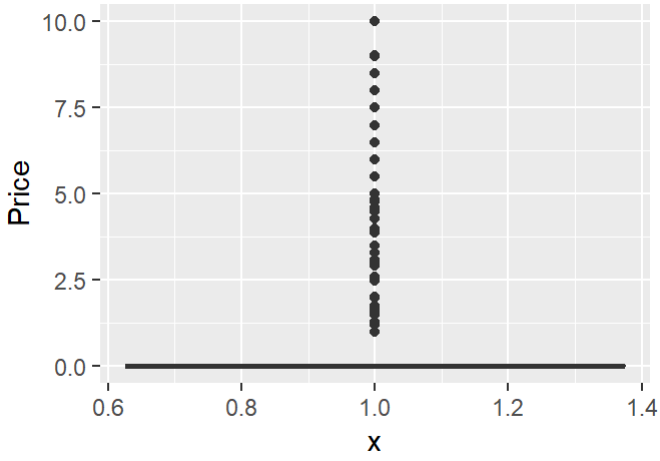






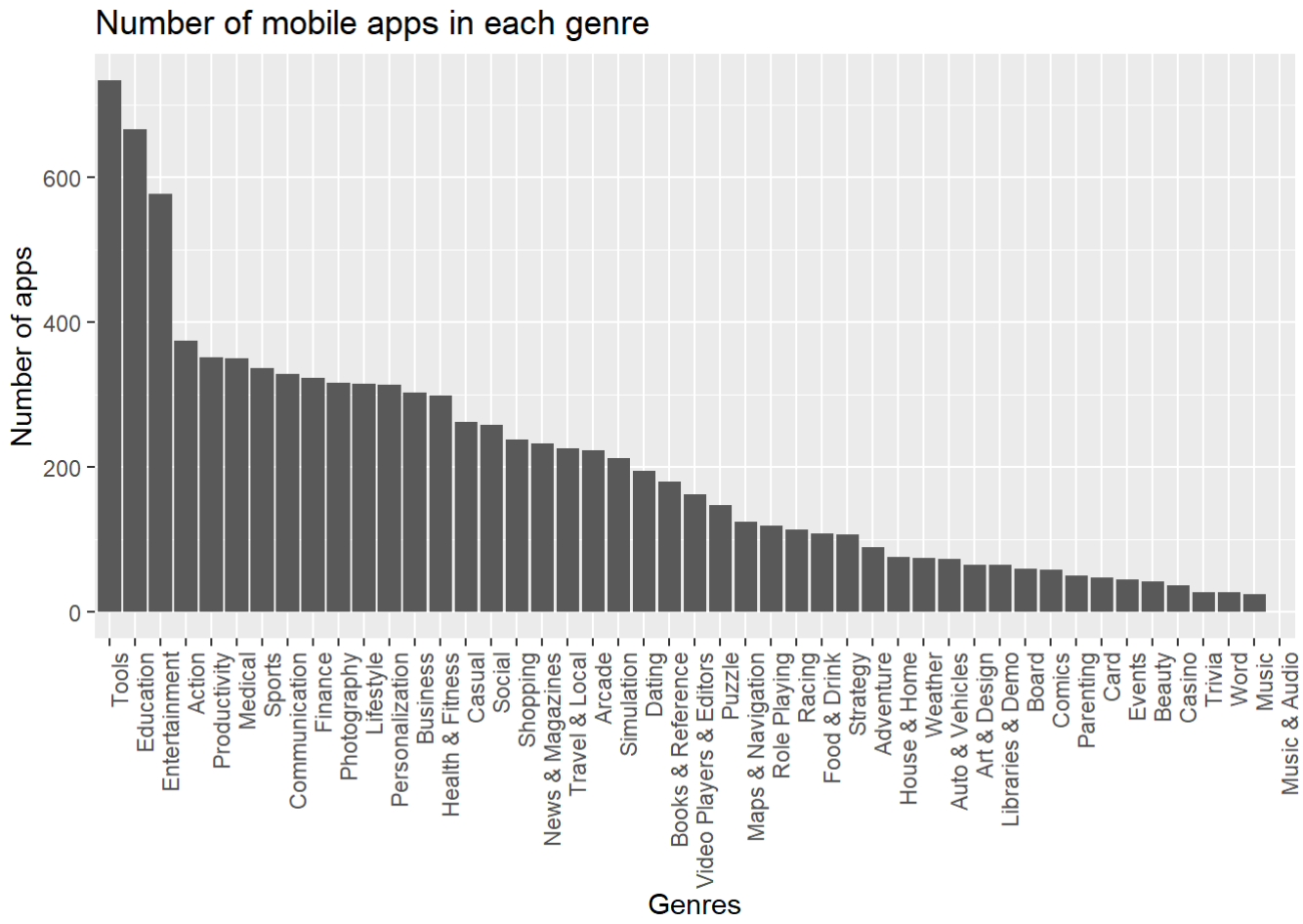




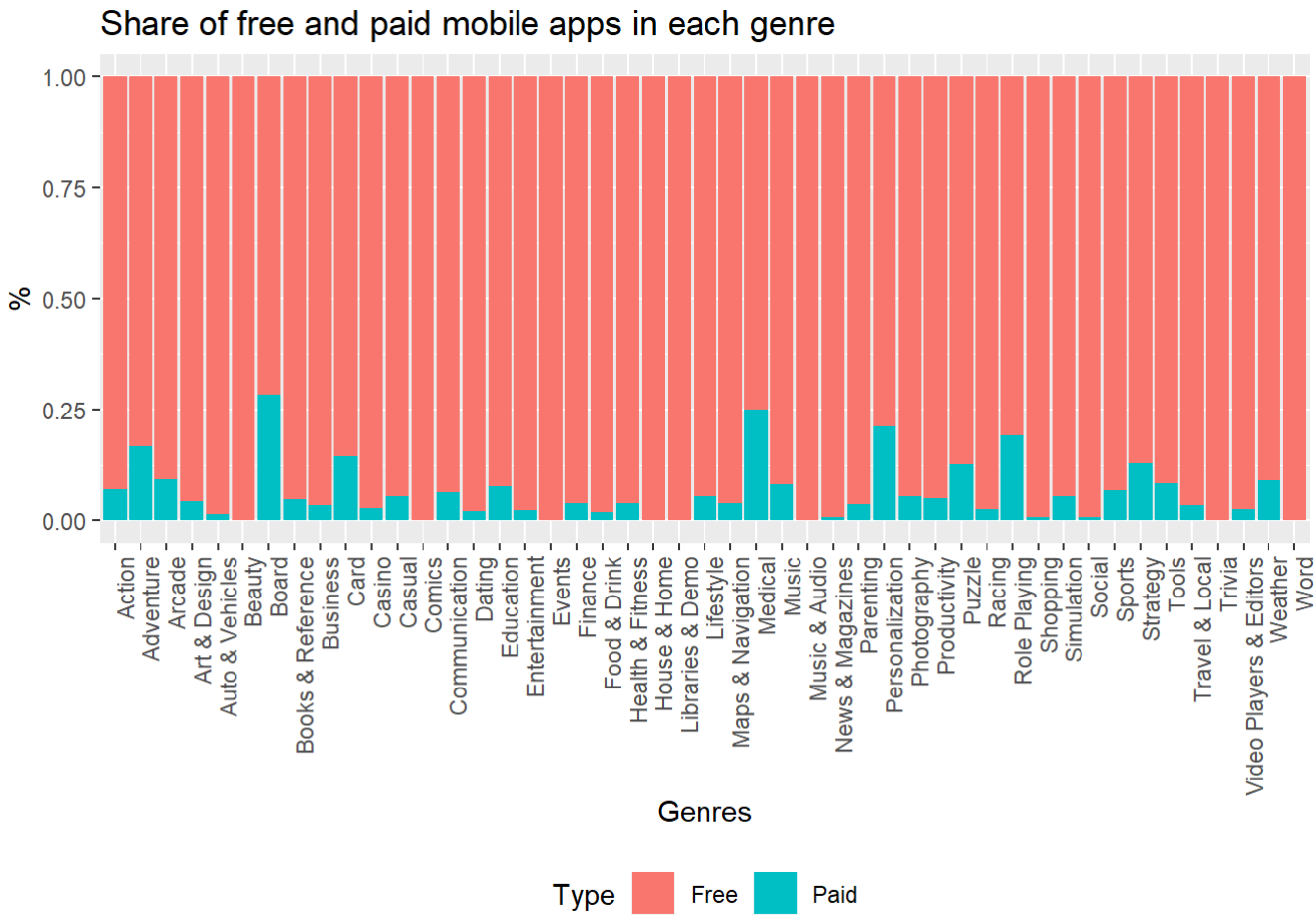


Three genres with the highest number of created mobile apps are: Tools, Education and Entertainment. Those three categories are on average 2 times bigger than the other genres. It means that these applications are the most useful for mobile devices users - developers create this kind of apps because there is a high demand

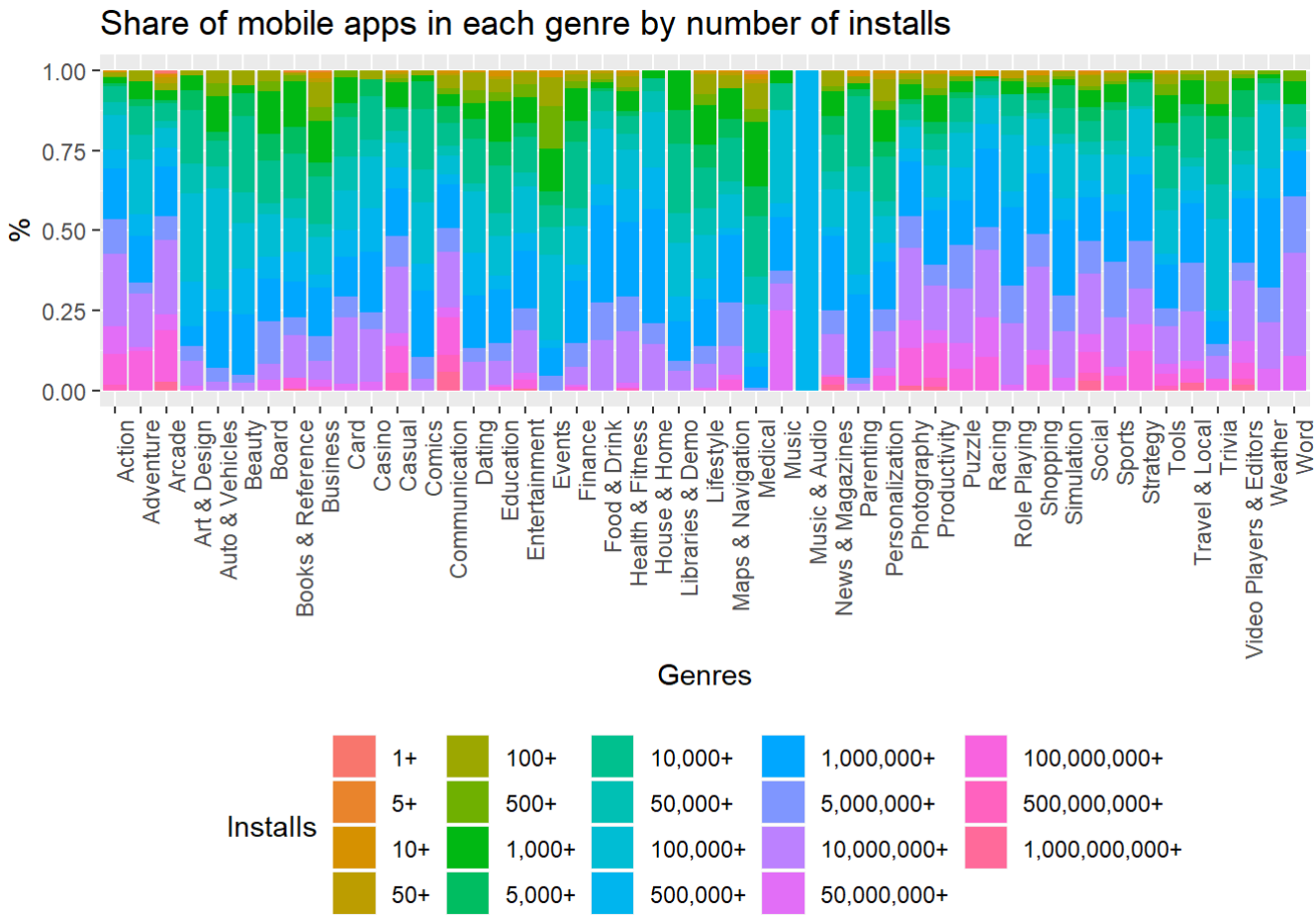
from users for them.



The highest and outstading share of paid mobile applications is in the four types of Genres: Board, Medical, Personalization and Role Playing. However, these are not the most popular types of apps - only Medical is in the top 10 of Genres. On the other hand, the categories with the highest number of applications are usually free of charge.



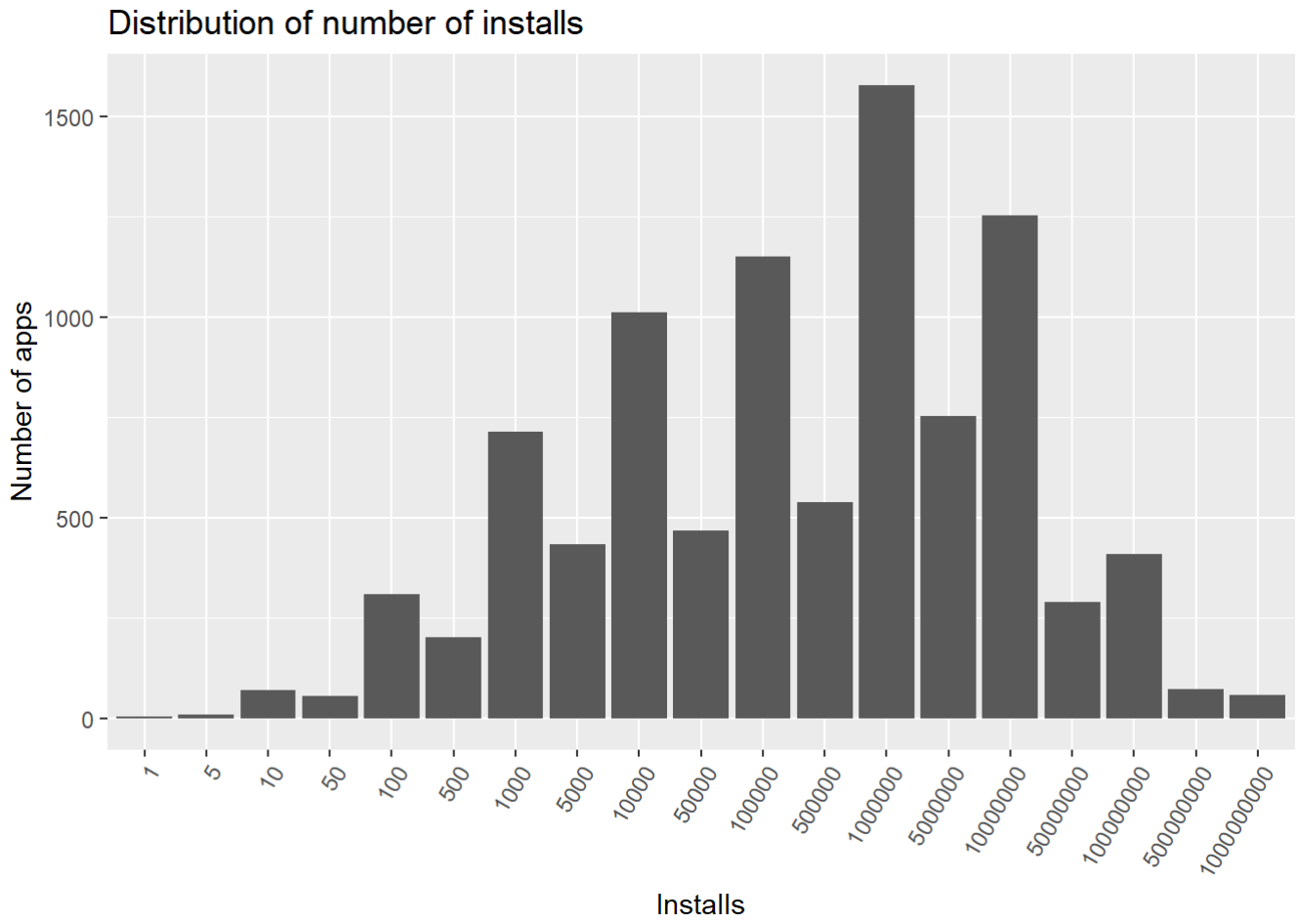
The graph below shows the number of installs of applications in each genre. The highest percentage of high numbers of installs have such applications genres like: Action, Arcade, Casual, Communication, Music, Photography, Racing, Shoopping, Social, Strategy, Travel & Loca, Video Players & Editors and Word. In the next section we provide a model to examine the factors determining the number of installs.



2.3. MODELLING

Since the purpose of our research is to find the features affecting the number of installs and Installs is not continuous but ordinal variable, we decided to use **Ordinal Logistic Regression model**. We are gonna examine the impact of the ratings, type and genres of application on the number of installs.

The most common number of installs is 1 000 000 +. The distribiution of the number of installs is plotted in the graph below. What makes the number of installs high or low? Can an app developer predict and impact it?



Ordinal Logistic Regression model

```
## Call:
## polr(formula = Installs ~ Rating + Type + Genres, data = df_selected)
##
## Coefficients:
##                                 Value Std. Error  t value
## Rating                        0.62259    0.03668  16.9736
## TypePaid                     -2.03432    0.07278 -27.9498
## GenresAdventure              -0.52513    0.20695  -2.5375
## GenresArcade                  0.11714    0.15036   0.7791
## GenresArt & Design           -1.65863    0.21737  -7.6304
## GenresAuto & Vehicles        -1.83716    0.21111  -8.7024
## GenresBeauty                 -1.90267    0.26394  -7.2086
## GenresBoard                  -1.22331    0.23470  -5.2121
## GenresBooks & Reference      -1.64750    0.15884 -10.3719
## GenresBusiness               -1.92984    0.13875 -13.9085
## GenresCard                   -0.92821    0.25612  -3.6242
## GenresCasino                 -1.16408    0.28504  -4.0840
## GenresCasual                 -0.22335    0.14278  -1.5642
## GenresComics                 -1.61782    0.22954  -7.0481
## GenresCommunication          -0.03472    0.13940  -0.2491
## GenresDating                 -1.51536    0.15207  -9.9652
## GenresEducation              -1.84630    0.11303 -16.3342
## GenresEntertainment          -1.25418    0.11705 -10.7147
## GenresEvents                 -2.76011    0.28103  -9.8213
## GenresFinance                -1.65130    0.13044 -12.6596
## GenresFood & Drink           -0.85167    0.17835  -4.7752
## GenresHealth & Fitness       -0.98743    0.13126  -7.5228
## GenresHouse & Home           -0.86986    0.20013  -4.3465
## GenresLibraries & Demo       -1.97983    0.22270  -8.8900
## GenresLifestyle              -1.76976    0.13440 -13.1680
## GenresMaps & Navigation      -1.19365    0.17991  -6.6347
## GenresMedical                -2.25318    0.12892 -17.4778
## GenresMusic                  -0.40686    0.36220  -1.1233
## GenresMusic & Audio          -1.22405    1.42069  -0.8616
## GenresNews & Magazines       -1.23788    0.14683  -8.4308
## GenresParenting              -1.69198    0.24160  -7.0033
## GenresPersonalization        -1.30349    0.13597  -9.5867
## GenresPhotography             0.04502    0.13337   0.3376
## GenresProductivity           -0.64586    0.13436  -4.8070
## GenresPuzzle                 -0.35409    0.16668  -2.1244
## GenresRacing                  0.09373    0.18263   0.5132
## GenresRole Playing           -0.52749    0.17239  -3.0599
## GenresShopping               -0.31272    0.14322  -2.1836
## GenresSimulation             -0.84630    0.14499  -5.8371
## GenresSocial                 -0.51920    0.14683  -3.5361
## GenresSports                 -0.74403    0.12962  -5.7399
## GenresStrategy               -0.03122    0.18439  -0.1693
## GenresTools                  -1.24117    0.11172 -11.1099
## GenresTravel & Local         -0.67850    0.14705  -4.6139
## GenresTrivia                 -1.78539    0.32798  -5.4436
## GenresVideo Players & Editors -0.50855    0.16645  -3.0554
## GenresWeather                -0.54554    0.20557  -2.6538
## GenresWord                   -0.30614    0.33508  -0.9136
##
## Intercepts:
##                      Value    Std. Error t value
## 1|5                   -6.9366   0.6094   -11.3823
```

```
## 5|10                  -5.7091   0.3385   -16.8648
## 10|50                 -3.8108   0.2076   -18.3551
## 50|100                -3.2700   0.1952   -16.7537
## 100|500               -2.0129   0.1823   -11.0412
## 500|1000              -1.5981   0.1805    -8.8532
## 1000|5000             -0.6864   0.1788    -3.8393
## 5000|10000            -0.3075   0.1786    -1.7218
## 10000|50000            0.3859   0.1788     2.1588
## 50000|100000           0.6591   0.1790     3.6830
## 100000|500000          1.2788   0.1796     7.1189
## 500000|1000000         1.5576   0.1800     8.6536
## 1000000|5000000        2.4074   0.1810    13.3017
## 5000000|10000000       2.8865   0.1816    15.8953
## 10000000|50000000      4.0732   0.1840    22.1410
## 50000000|100000000     4.5642   0.1859    24.5583
## 100000000|500000000    6.0698   0.2012    30.1712
## 500000000|1000000000   6.8903   0.2237    30.8011
##
## Residual Deviance: 44546.32
## AIC: 44678.32
```

We calculate p-value and filter out these variables which have p-value <= 0.05 and are statistically significant at the 5% level.

```
##                              id      Value Std. Error    t value p value
## 1                        Rating  0.6225871 0.03667970  16.973614   0.000
## 2                      TypePaid -2.0343163 0.07278475 -27.949762   0.000
## 3                GenresAdventure -0.5251321 0.20695255  -2.537452   0.011
## 4             GenresArt & Design -1.6586254 0.21737038  -7.630411   0.000
## 5           GenresAuto & Vehicles -1.8371625 0.21110924  -8.702426   0.000
## 6                    GenresBeauty -1.9026701 0.26394414  -7.208609   0.000
## 7                     GenresBoard -1.2233055 0.23470351  -5.212131   0.000
## 8         GenresBooks & Reference -1.6475037 0.15884238 -10.371941   0.000
## 9                  GenresBusiness -1.9298429 0.13875288 -13.908488   0.000
## 10                    GenresCard -0.9282061 0.25611597  -3.624163   0.000
## 11                  GenresCasino -1.1640815 0.28503661  -4.083972   0.000
## 12                  GenresComics -1.6178169 0.22953987  -7.048086   0.000
## 13                  GenresDating -1.5153634 0.15206543  -9.965207   0.000
## 14                GenresEducation -1.8462959 0.11303237 -16.334223   0.000
## 15            GenresEntertainment -1.2541776 0.11705219 -10.714687   0.000
## 16                  GenresEvents -2.7601094 0.28103168  -9.821346   0.000
## 17                 GenresFinance -1.6512960 0.13043777 -12.659646   0.000
## 18             GenresFood & Drink -0.8516658 0.17835047  -4.775237   0.000
## 19         GenresHealth & Fitness -0.9874314 0.13125765  -7.522848   0.000
## 20             GenresHouse & Home -0.8698603 0.20012751  -4.346531   0.000
## 21         GenresLibraries & Demo -1.9798288 0.22270256  -8.890013   0.000
## 22                GenresLifestyle -1.7697624 0.13439915 -13.167959   0.000
## 23       GenresMaps & Navigation -1.1936537 0.17990954  -6.634744   0.000
## 24                  GenresMedical -2.2531791 0.12891667 -17.477795   0.000
## 25         GenresNews & Magazines -1.2378785 0.14682848  -8.430779   0.000
## 26                GenresParenting -1.6919771 0.24159587  -7.003336   0.000
## 27          GenresPersonalization -1.3034913 0.13596937  -9.586654   0.000
## 28             GenresProductivity -0.6458617 0.13435954  -4.806966   0.000
## 29                   GenresPuzzle -0.3540949 0.16668063  -2.124391   0.034
## 30             GenresRole Playing -0.5274912 0.17238778  -3.059911   0.002
## 31                 GenresShopping -0.3127235 0.14321506  -2.183594   0.029
## 32               GenresSimulation -0.8462998 0.14498544  -5.837136   0.000
## 33                   GenresSocial -0.5191964 0.14682936  -3.536053   0.000
## 34                   GenresSports -0.7440328 0.12962455  -5.739906   0.000
## 35                    GenresTools -1.2411722 0.11171773 -11.109894   0.000
## 36           GenresTravel & Local -0.6784979 0.14705420  -4.613930   0.000
## 37                   GenresTrivia -1.7853906 0.32797952  -5.443604   0.000
## 38 GenresVideo Players & Editors -0.5085516 0.16644510  -3.055372   0.002
## 39                  GenresWeather -0.5455423 0.20557173  -2.653780   0.008
## 40                            1|5 -6.9365660 0.60941915 -11.382258   0.000
## 41                           5|10 -5.7090990 0.33852096 -16.864832   0.000
## 42                          10|50 -3.8108211 0.20761617 -18.355127   0.000
## 43                         50|100 -3.2700147 0.19518158 -16.753705   0.000
## 44                        100|500 -2.0128767 0.18230586 -11.041207   0.000
## 45                       500|1000 -1.5981047 0.18051234  -8.853161   0.000
## 46                      1000|5000 -0.6864149 0.17878706  -3.839288   0.000
## 47                     10000|50000  0.3859030 0.17876132   2.158761   0.031
## 48                    50000|100000  0.6591469 0.17897046   3.682993   0.000
## 49                   100000|500000  1.2788362 0.17963900   7.118923   0.000
## 50                  500000|1000000  1.5575692 0.17999176   8.653558   0.000
## 51                 1000000|5000000  2.4073653 0.18098181  13.301698   0.000
## 52                5000000|10000000  2.8864919 0.18159372  15.895329   0.000
## 53               10000000|50000000  4.0732434 0.18396854  22.140978   0.000
## 54               50000000|100000000  4.5641971 0.18585117  24.558345   0.000
## 55             100000000|500000000  6.0698011 0.20117845  30.171229   0.000
## 56           500000000|1000000000  6.8902784 0.22370249  30.801080   0.000
```

Explaining the model

The basic of proportional odds model has the following mathematical fomula:

$$logit\ [P(Y \leq j)] = \alpha_j - \sum \beta_i X_i$$

where $j = 1,\ldots,J-1$ and $i = 1,\ldots,M$

with 'J' being a number of levels of dependent variable Installs (J=18) and 'M' being a total number of independent variables (M=3).

'j' is a single level of Installs, meanwhile 'i' is a single independent variables:

- i = 1 refers to Rating
- i = 2 refers to Type
- i = 3 refers to Genres

Interpretation of the results:

Comments on Coefficients: The continuous variable Rating can be interpreted as follows: with one unit increase in rating of the free mobile app the log of odds of increasing the number of installs of this app increases by 0.62259 (62%). The ordinal variable Type can be interpreted as follows: with changing mobile app from free to paid the log of odds of increasing the number of installs of this app decreases by -2.03432 (-203%).

Comments on Intercepts: For example in case of intercept 1|5: it can be interpreted as the log of odds of having only 1 install versus having more than 1 install.

We can make an example based on our model:

Let's suppose that our mobile app developer has finished two applications with charateristic described below:

App 1 is a free app of genre Education and rating 4:

```
new_app <- data.frame(Rating = 4, Type = "Free", Genres = "Education")
round(predict(model_fit, new_app, type = "p"), 3)
```

```
##          1          5         10         50        100        500
##      0.001      0.001      0.010      0.008      0.046      0.030
##       1000       5000      10000      50000     100000     500000
##      0.113      0.069      0.157      0.068      0.150      0.060
##    1000000    5000000   10000000   50000000  100000000  500000000
##      0.140      0.050      0.065      0.012      0.015      0.002
## 1000000000
##      0.002
```

App 2 is a paid app of genre Racing and rating 3.5:

```
new_app_2 <- data.frame(Rating = 3.5, Type = "Paid", Genres = "Racing")
round(predict(model_fit, new_app_2, type = "p"), 3)
```

```
##          1          5         10         50        100        500
##      0.001      0.002      0.015      0.012      0.066      0.042
##       1000       5000      10000      50000     100000     500000
##      0.147      0.083      0.170      0.067      0.135      0.050
##    1000000    5000000   10000000   50000000  100000000  500000000
##      0.108      0.036      0.045      0.008      0.010      0.002
## 1000000000
##      0.001
```

Results: The first app has 15,7% chance to get 10 thousand downloads while the second app has 17% chances (highest) to get that number of installs. The first app has 6,5% chances to get 10 million downloads while the second app has 4,5% chances (lower) to get that number of installs.

3. CONCLUSIONS

When a mobile app developer is about to create an app and then launch it on the app stores, he/she always asks him/herself what are the features providing highest number of installs. In our report we propose a method of calculating expected number of installs knowing the type, genre and expected rating of an application.

4. REFERENCES

https://en.wikipedia.org/wiki/Mobile_app (https://en.wikipedia.org/wiki/Mobile_app)

http://www.businessofapps.com/data/app-statistics/ (http://www.businessofapps.com/data/app-statistics/)

https://en.wikipedia.org/wiki/Ordered_logit (https://en.wikipedia.org/wiki/Ordered_logit)

https://towardsdatascience.com/implementing-and-interpreting-ordinal-logistic-regression-1ee699274cf5 (https://towardsdatascience.com/implementing-and-interpreting-ordinal-logistic-regression-1ee699274cf5)