Analysis of Google Playstore Apps

Project Proposal
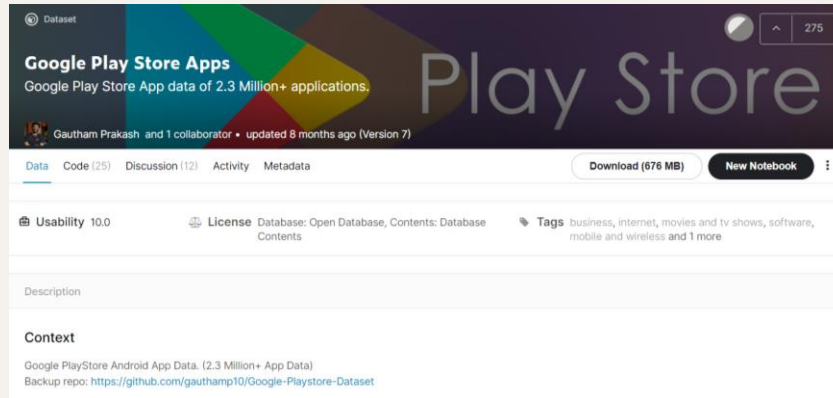
Content

1. Dataset Description

- Dataset is obtained through Kaggle , contributed by author Gautham Prakash
- Dataset was scrapped online and accurate as of June 2021
- Google PlayStore Android App Data
- Source link : https://www.kaggle.com/gauthamp10/google-playstore-apps

1. Dataset Description

Column Count	24
Row Count	2312944

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2312944 entries, 0 to 2312943
Data columns (total 24 columns):
 #   Column            Dtype
---  ------            -----
 0   App Name          object
 1   App Id            object
 2   Category          object
 3   Rating            float64
 4   Rating Count      float64
 5   Installs          object
 6   Minimum Installs  float64
 7   Maximum Installs  int64
 8   Free              bool
 9   Price             float64
 10  Currency          object
 11  Size              object
 12  Minimum Android   object
 13  Developer Id      object
 14  Developer Website object
 15  Developer Email   object
 16  Released          object
 17  Last Updated      object
 18  Content Rating    object
 19  Privacy Policy    object
 20  Ad Supported      bool
 21  In App Purchases  bool
 22  Editors Choice    bool
 23  Scraped Time      object
dtypes: bool(4), float64(4), int64(1), object(15)
memory usage: 361.8+ MB
```

1. Dataset Description

App Name	App Id	Category	Rating	Rating Count	Installs	Minimum Installs	Maximum Installs	Free	Price	Currency	Size	Minimum Android	Developer Id	Develop
Gakondo	com.ishakwe.gakondo	Adventure	0	0	10+	10	15	TRUE	0	USD	10M	7.1 and up	Jean Confident Irã…	https://
Ampere Battery Info	com.webserveis.batteryinfo	Tools	4.4	64	5,000+	5000	7662	TRUE	0	USD	2.9M	5.0 and up	Webserveis	https://
Vibook	com.doantiepvien.crm	Productivity	0	0	50+	50	58	TRUE	0	USD	3.7M	4.0.3 and up	Cabin Crew	
Smart City Trichy Public Se	cst.stJoseph.ug17ucs548	Communication	5	5	10+	10	19	TRUE	0	USD	1.8M	4.0.3 and up	Climate Smart Tech	http://w
GROW.me	com.horodyski.grower	Tools	0	0	100+	100	478	TRUE	0	USD	6.2M	4.1 and up	Rafal Milek-Horody	http://w
IMOCCI	com.imocci	Social	0	0	50+	50	89	TRUE	0	USD	46M	6.0 and up	Imocci GmbH	http://w
unlimited 4G data prank fr	getfreedata.superfatiza.unlimitedjioda	Libraries & Demo	4.5	12	1,000+	1000	2567	TRUE	0	USD	2.5M	4.1 and up	android developer779	
The Everyday Calendar	com.mozaix.simoneboard	Lifestyle	2	39	500+	500	702	TRUE	0	USD	16M	5.0 and up	Mozaix LLC	
WhatsOpen	com.whatsopen.app	Communication	0	0	10+	10	18	TRUE	0	USD	1.3M	4.4 and up	Yilver Molina Hurta	http://yi
Neon 3d Iron Tech Keyboa	com.ikeyboard.theme.neon_3d.iron.te	Personalization	4.7	820	50,000+	50000	62433	TRUE	0	USD	3.5M	4.1 and up	Free 2021 Themes	https://t
Dodge The Cars!	com.MrScratchEnterprises.CarDogeGa	Racing	4.9	55	100+	100	329	TRUE	0	USD	51M	4.4 and up	MrScratch	
Parents	com.eqra.android.BSTS.parent	Maps & Navigation	0	0	100+	100	330	TRUE	0	USD	2.7M	4.0 and up	Eqra Tech	http://w
æ¡ƒåœ'æ©Ÿå 'æ·é‹æ™"å^	com.jieapp.taoyuanmetro	Travel & Local	3.9	118	10,000+	10000	37763	TRUE	0	USD	7.6M	4.2 and up	Jie App	https://ji
be.MOBILISED	com.hastobe.bemobilised	Maps & Navigation	0	0	100+	100	274	TRUE	0	USD	15M	6.0 and up	has.to.be gmbh	http://w
Caliway Conductor	com.caliwayconductor.driver	Travel & Local	3.7	1572	10,000+	10000	42856	TRUE	0	USD	29M	4.2 and up	WayCali	https://v

Price	Currency	Size	Minimum Android	Developer Id	Developer Website	Developer Email	Released	Last Updated	Content Rating	Privacy Policy	Ad Supported	In App Purchases	Editors Choice	Scraped Time
0	USD	10M	7.1 and up	Jean Confident Irã	https://benivizibvose.	jean21101999@gm	26-Feb-20	26-Feb-20	Everyone	https://benivizibvose.tk/pr	FALSE	FALSE	FALSE	6/15/2021 20:19
0	USD	2.9M	5.0 and up	Webserveis	https://webserveis.n	webserveis@gmail.	21-May-20	6-May-21	Everyone	https://dev4phones.wordp	TRUE	FALSE	FALSE	6/15/2021 20:19
0	USD	3.7M	4.0.3 and up	Cabin Crew		vnacrewit@gmail.c	9-Aug-19	19-Aug-19	Everyone	https://www.vietnamairlin	FALSE	FALSE	FALSE	6/15/2021 20:19
0	USD	1.8M	4.0.3 and up	Climate Smart Tech	http://www.climate	climatesmarttech2	10-Sep-18	13-Oct-18	Everyone		TRUE	FALSE	FALSE	6/15/2021 20:19
0	USD	6.2M	4.1 and up	Rafal Milek-Horody	http://www.horodys	rmilekhorodyski@g	21-Feb-20	12-Nov-18	Everyone	http://www.horodyski.com	FALSE	FALSE	FALSE	6/15/2021 20:19
0	USD	46M	6.0 and up	Imocci GmbH	http://www.imocci.	info@imocci.com	24-Dec-18	20-Dec-19	Teen	https://www.imocci.com/\	FALSE	TRUE	FALSE	6/15/2021 20:19
0	USD	2.5M	4.1 and up	android developer779		aitomgharfatimezz	23-Sep-19	27-Sep-19	Everyone	https://sites.google.com/v	FALSE	FALSE	FALSE	6/15/2021 20:19
0	USD	16M	5.0 and up	Mozaix LLC		elementuser03@gr	21-Jun-19	21-Jun-19	Everyone	https://www.freeprivacypc	FALSE	FALSE	FALSE	6/15/2021 20:19
0	USD	1.3M	4.4 and up	Yilver Molina Hurta	http://yilvermolinah	yilver.mh1996@gmail.com		7-Dec-18	Teen	http://elcafedelamanana.y	FALSE	FALSE	FALSE	6/15/2021 20:19
0	USD	3.5M	4.1 and up	Free 2021 Themes	https://trendyteme8	trendyteme.888@g	22-Sep-19	7-Oct-20	Everyone	http://bit.ly/EmojiThemePr	TRUE	FALSE	FALSE	6/15/2021 20:19
0	USD	51M	4.4 and up	MrScratch		mithalaarush@gma	30-Jul-20	30-Jul-20	Everyone	https://docs.google.com/d	FALSE	FALSE	FALSE	6/15/2021 20:19
0	USD	2.7M	4.0 and up	Eqra Tech	http://www.hafizqu	hhiyassat@eqratec	10-Jan-18	27-Jun-18	Everyone	https://docs.google.com/d	FALSE	FALSE	FALSE	6/15/2021 20:19
0	USD	7.6M	4.2 and up	Jie App	https://jie-app.herol	jie.myapp@gmail.c	3-Apr-18	11-Jun-21	Everyone	http://appsetting.blogspot.	TRUE	FALSE	FALSE	6/15/2021 20:19
0	USD	15M	6.0 and up	has.to.be gmbh	http://www.has-to-I	support@has-to-be	9-Feb-20	14-May-21	Everyone	https://has-to-be.com/de/	FALSE	FALSE	FALSE	6/15/2021 20:19
0	USD	29M	4.2 and up	WayCali	https://www.facebo	waycalicol@gmail.(	5-Sep-18	30-May-20	Everyone	https://www.facebook.cor	FALSE	FALSE	FALSE	6/15/2021 20:19

1. Dataset Description

- Important columns to be included in the analysis:
  - **Category** = App Category
  - **Rating** = Average Rating of the App
  - **Rating Count** = No. of Ratings
  - **Installs** = Approximate Install Counts
  - **Free** = Whether App is Free or Paid
  - **Price** = Price of the App
  - **Size** = Size of Application Package
  - **Minimum Android** = Minimum Android Version Supported
  - **Last Updated** = Last App Update Date
  - **Ad Supported** = Whether Ads are supported in the App
  - **In-App Purchases** = Whether In –App Purchases are Applicable
  - **Editors Choice** = Whether App is rated as Editor's Choice

2. Challenges and Difficulties

o   A few columns have incorrect data type e.g. no. of installs – Object
o   Many repetitive categories e.g. 'Education' vs. 'Educational', 'Music & Audio' vs. 'Music'
o   Removal of outliers must be handled with careful consideration as some outliers reflect the reality e.g. Price of the App at $400

3. Goals to be Addressed

- **How do the features of an app affect its success ?**
  - Which app category is the most popular (ratings, no of installs) and worth making an app for?
  - Does app size affect the no. of installs and ratings ?
  - Does the Editor's Choice label really influence the ratings and the installation no.?
  - Will app maintenance affect the no. of installs, review ratings?
  - How do paid and free apps differ ?
  - Are there apps which have high no. of installs but yet low review ratings -> user disappointment ?

3. Goals to be Addressed

- **Explore the effects of different monetization methods through Apps**
  - Ad-Supported vs. Paid Apps vs. In-App Purchases
  - How do these three different ways of monetization affect the user ratings and installation numbers?