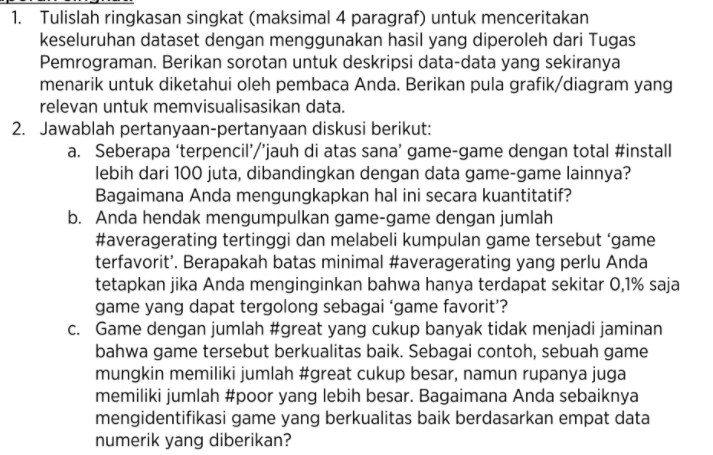
Kelompok 8:

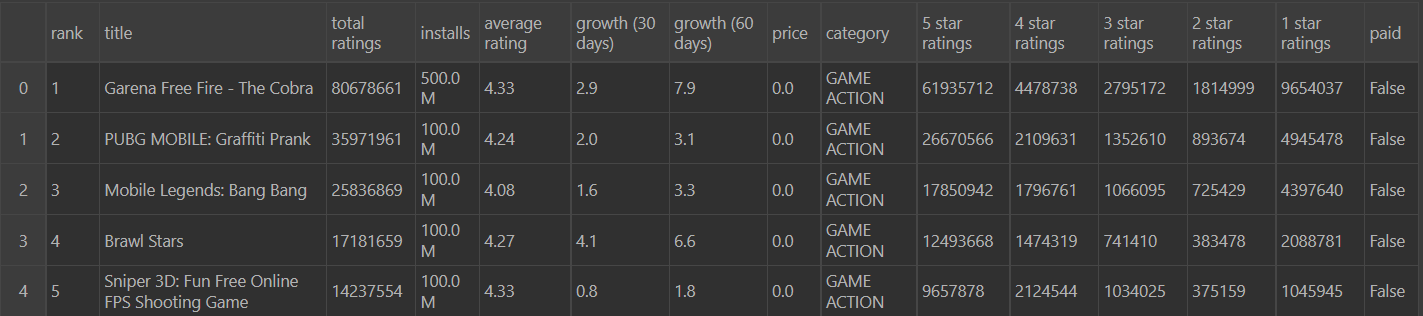
* Bryan Christopher Wijaya
* James Patrick Oentoro
* Noel Christevent Mandak
* Stefannus Christian
* Tiffany Sondakh

Link tugas pemrograman : <https://datalore.jetbrains.com/notebook/CPsQxEPW7yVpS665a3lC8b/1EU9uf8Ml4kPZtmGNMmik9>



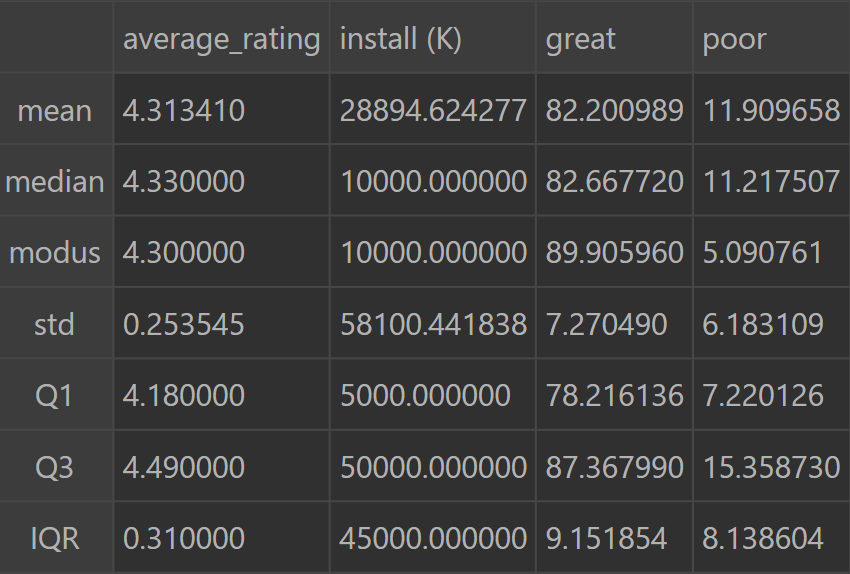
Jawaban Nomor 1

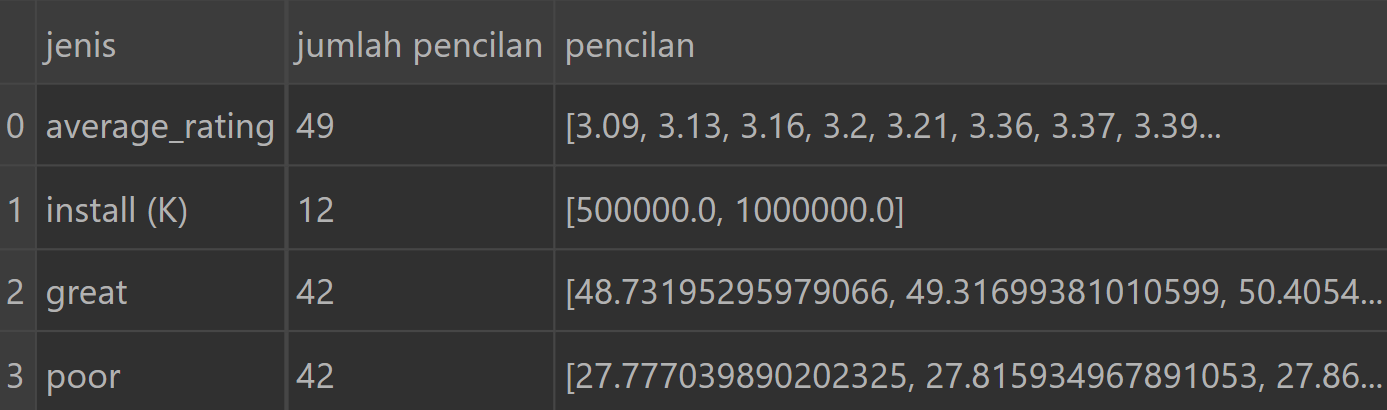
*Dataset* yang kami analisis pada Proyek ini adalah *dataset game* yang beredar di Amerika Serikat. *Dataset game* ini diambil dari layanan distribusi digital *Google Play Store* yaitu sebuah toko aplikasi resmi untuk perangkat bersertifikat yang berjalan pada sistem operasi Android. Tujuan dari analisis ini adalah untuk menjelaskan dan memberikan gambaran dari *dataset* tersebut dan agar pengguna yang melihat analisis ini dapat memiliki gambaran mengenai hal-hal penting yang terjadi dalam *dataset* ini. Dari sekian banyaknya game yang beredar pada *dataset* ini, hanya terdapat sedikit *game* yang mendapatkan *rating* yang tinggi. Maka dari itu, pada analisis ini, kami akan menjabarkan secara *detail* telaga game yang begitu banyak dan melihat beberapa terpencil nya game-game *best rating* tersebut dibanding game lain yang dapat dikategorikan biasa.*Dataset* yang dianalisis memiliki 15 kolom yang menjadi parameter atau fitur. Fitur tersebut antara lain adalah sebagaimana ditunjukkan oleh gambar dibawah ini.



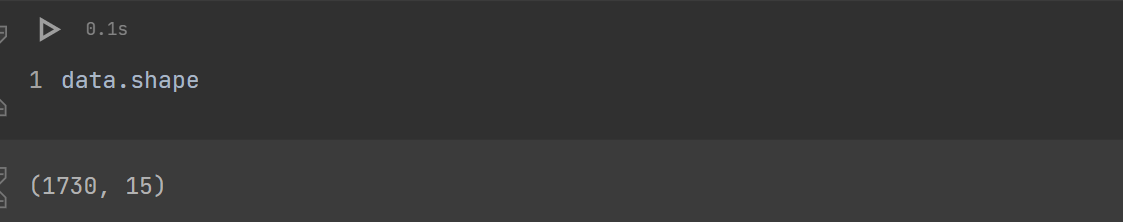


Namun, dari 15 fitur yang *dataset*, hanya beberapa fitur saja yang akan dianalisis lebih lanjut. Fitur-fitur tersebut antara lain adalah #*averagerating, #install, # great, #poor, #category.* Penjelasan singkat dari lima fitur ini dengan urut adalah sebagai berikut. Tingkat **kepuasan**pengguna, banyaknya pengguna yang **menggunakan** game tersebut, persentase pengguna yang **memberikan** star 5 dan 4, persentase yang **memberikan** star 2 dan 1, dan kategori dari *game* tersebut. Berikut merupakan rangkuman statistika kuantitatif atas parameter numerik #averagerating, #install, #great, dan #poor dari semua data *game*.



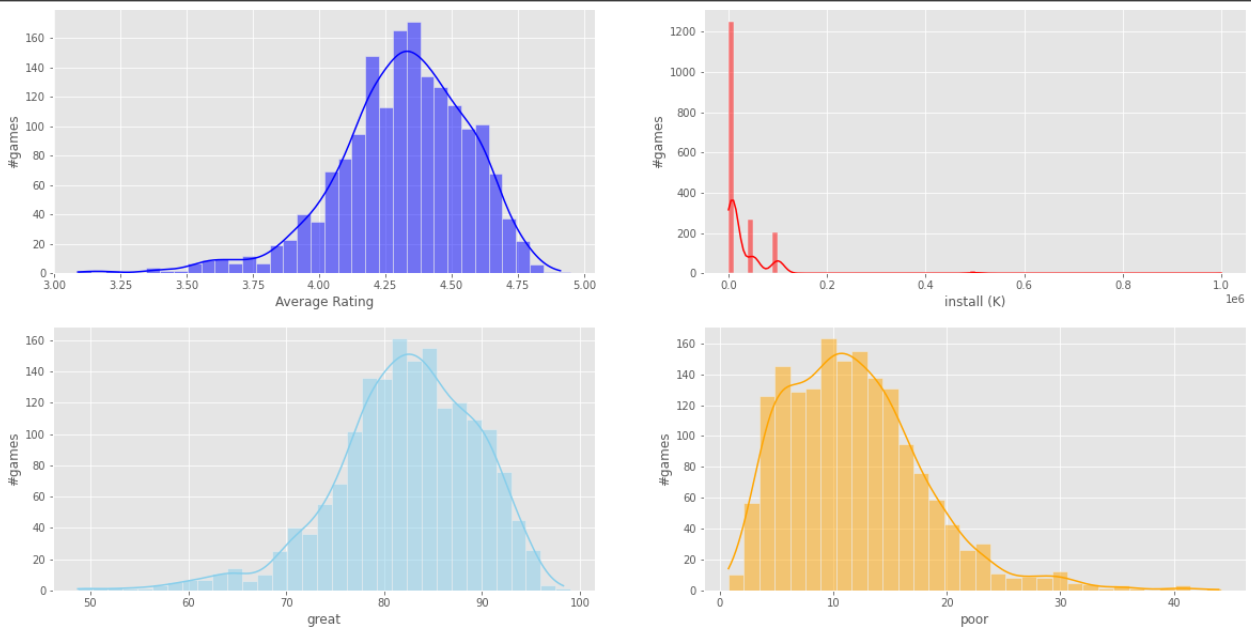




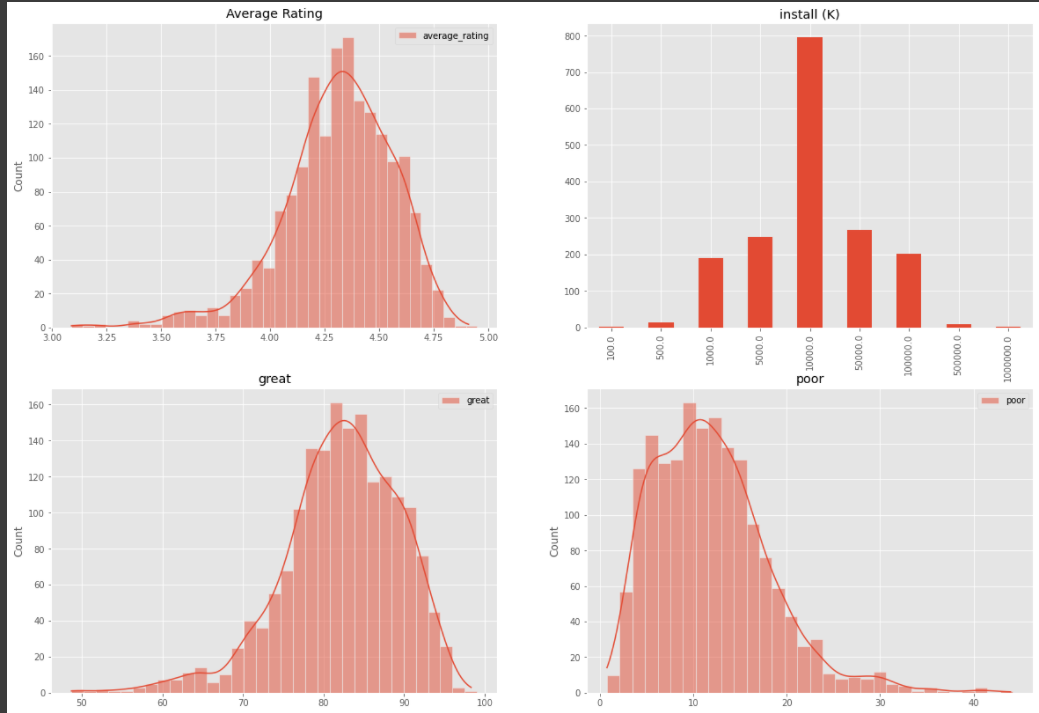


Dapat dilihat dari tabel rangkuman statistika kualitatif tersebut bahwa dari 1730 instance data yang terdapat pada *dataframe*, rata-rata dari rata-rata *rating* setiap *game* adalah 4.3. Ini artinya bahwa kita dapat *expect* *average rating* dari suatu game adalah sekitar 4.3. Selain itu, dapat dilihat dari tabel bahwa median dan modus dari *average rating* sangat serupa. Ini artinya bahwa *average rating* berdistribusi simetris/uniform. Standar deviasi dari *average rating* sangat kecil yaitu hanya 0.25. Artinya fitur *average rating* tidak memiliki keberagaman data yang besar tetapi memiliki keberagaman daya yang sangat kecil. Artinya, tidak akan banyak data yang *value* nya jauh diatas atau dibawah 4.3. Tambahan lagi, terdapat paling banyak 49 pencilan dari keempat fitur tersebut. Artinya, hanya sekitar 2.87% dari *dataset* saja yang data-data nya *extreme* atau dengan kata lain datanya cukup jauh berbeda dari rata-rata *dataset*.

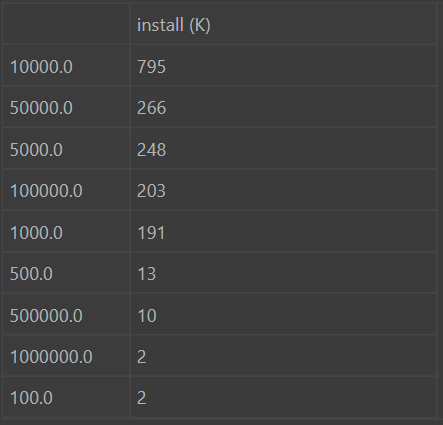
Data yang dihasilkan dengan tabel berikut tentu saja sulit untuk dipahami apa arti dari angka-angka tersebut. Maka dari itu, diperlukan suatu alat bantuan visualisasi untuk memvisualisasikan data-data pada tabel tersebut yang akan membantu pengguna untuk melihat gambaran dari data tersebut. Berikut merupakan visualisasi data dalam bentuk histogram. Dapat dilihat dari keempat parameter tersebut, parameter *average rating*, *great, poor* berbentuk mirip seperti distribusi normal. Dalam teori, jika dataset tersebut berdistribusi normal sempurna artinya bahwa rata-rata = median = modus. Namun karena data parameter *great*, *average rating* dan *poor* tidak sepenuhnya berdistribusi normal, maka tidak semua rata-rata = median = modus. Untuk kasus *great,* rata-rata = median tetapi != modus. Tetapi jarak dari rata-rata dan median ke modus tidak jauh. Untuk kasus *poor,* rata-rata = median tetapi != modus. Tetapi jarak dari rata-rata dan median ke modus juga tidak jauh, mirip dengan great. Untuk kasus *average rating,* rata-rata = median = modus. Artinya *average rating* berdistribusi normal sempurna. Karena *average rating, great, poor* mirip dengan distribusi normal, maka artinya sebagian besar titik data relatif sama, artinya data-data muncul dalam rentang nilai yang kecil dengan sedikit outlier pada ujung atas dan bawah *data range*. Untuk kasus install (K), lebih cocok divisualisasikan dengan bar chart karena data install (K) berupa data diskrit. Histogram kurang cocok untuk data install (K) karena terdapat pencilan yang ekstrim pada data install (K). Contoh dari pencilan ekstrim ini dapat dilihat pada tabel A.1. Dapat dilihat dari grafik A.1 bahwa range data dan frekuensi dari tabel A.1 cukup tinggi dan bervarian. Hal ini juga dapat dibuktikan dengan melihat standar deviasi dan IQR dari install (K) yang memiliki nilai yang tinggi. Histogram install (K) kurang jelas dan dapat dilihat dari grafik A.2 (bar chart) bahwa install (K) lebih mudah di interpretasi (lebih jelas). Terdapat alat visualisasi lain selain histogram yang dapat membantu menginterpretasikan datanya. Alat visualisasi tersebut adalah boxplot.



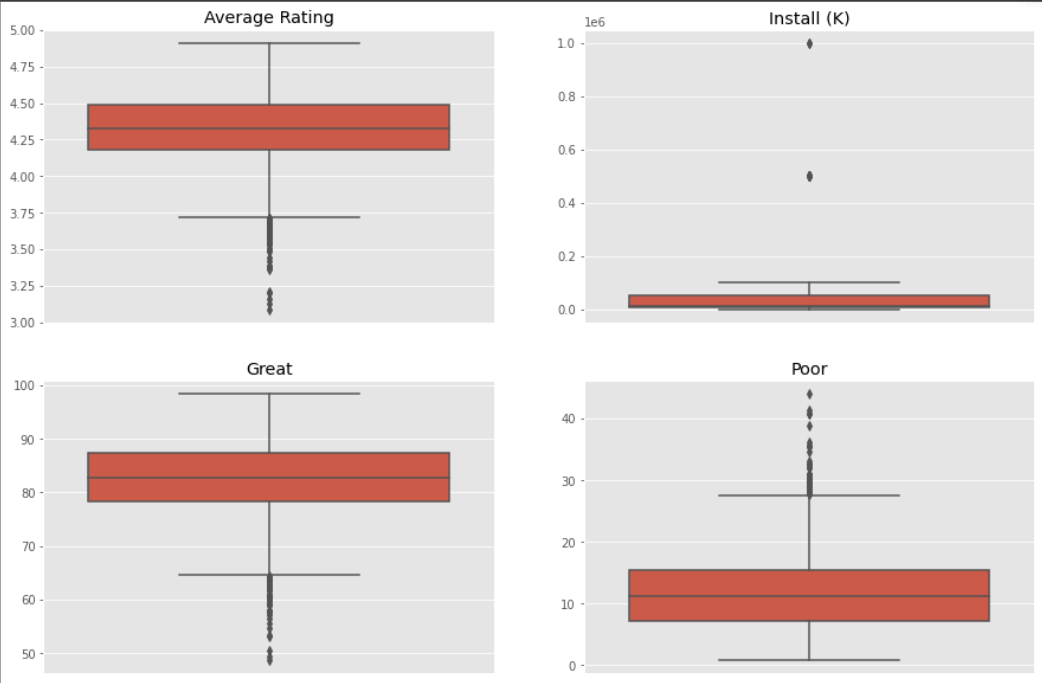
Grafik A.1



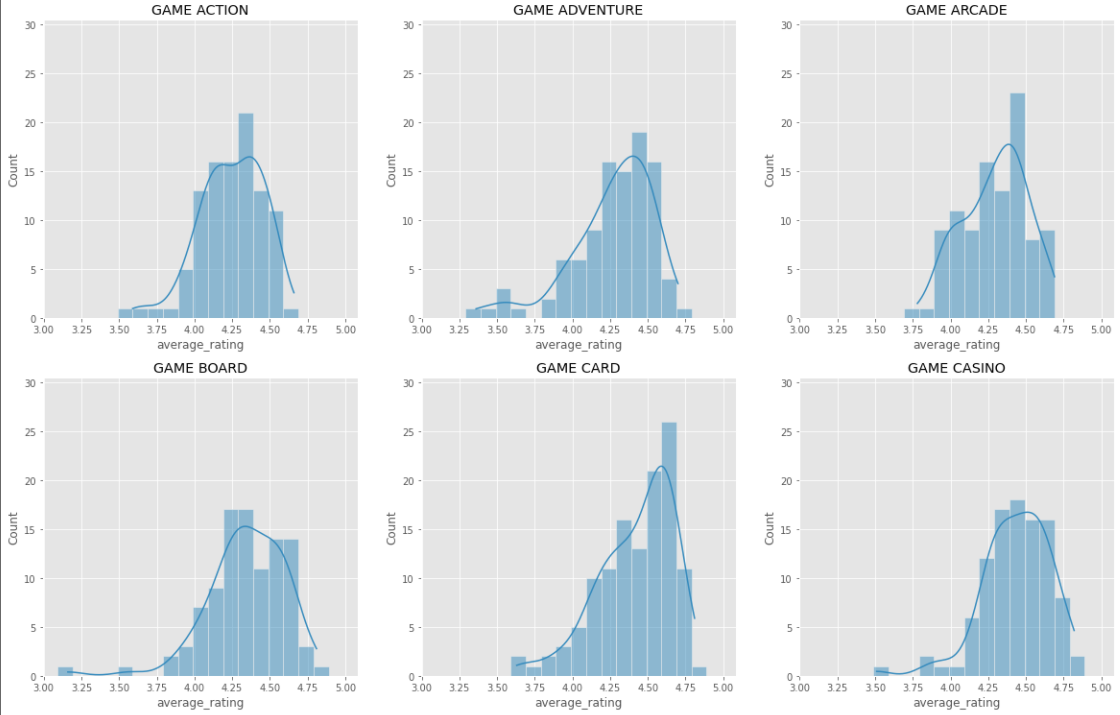
Grafik A.2

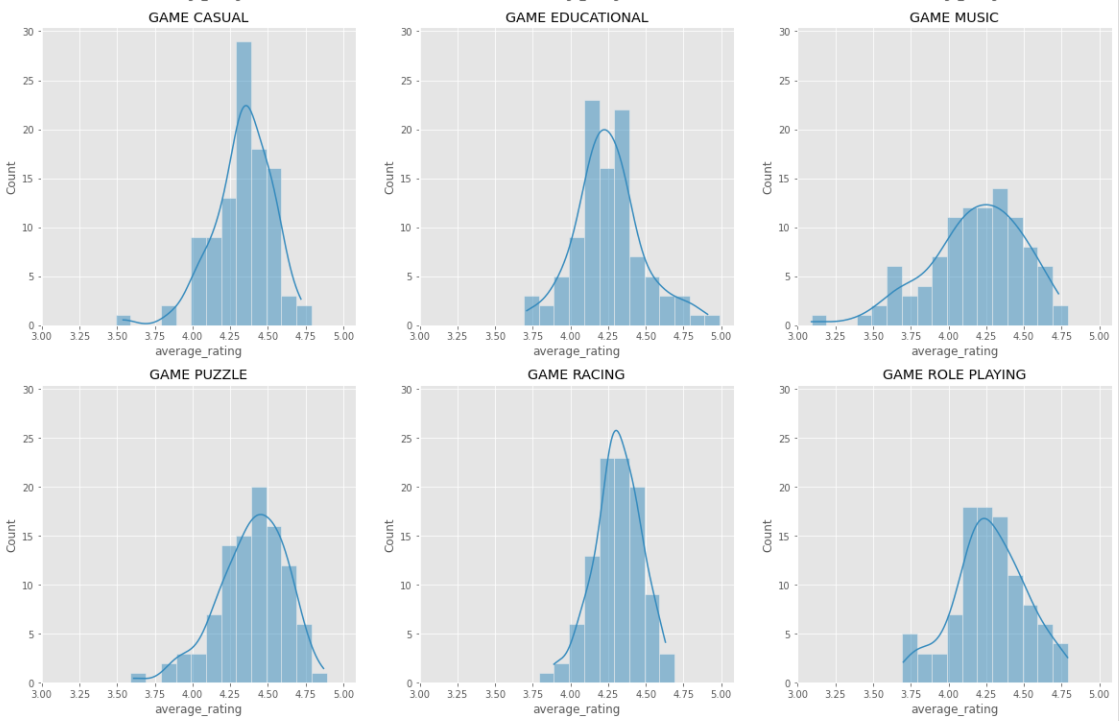


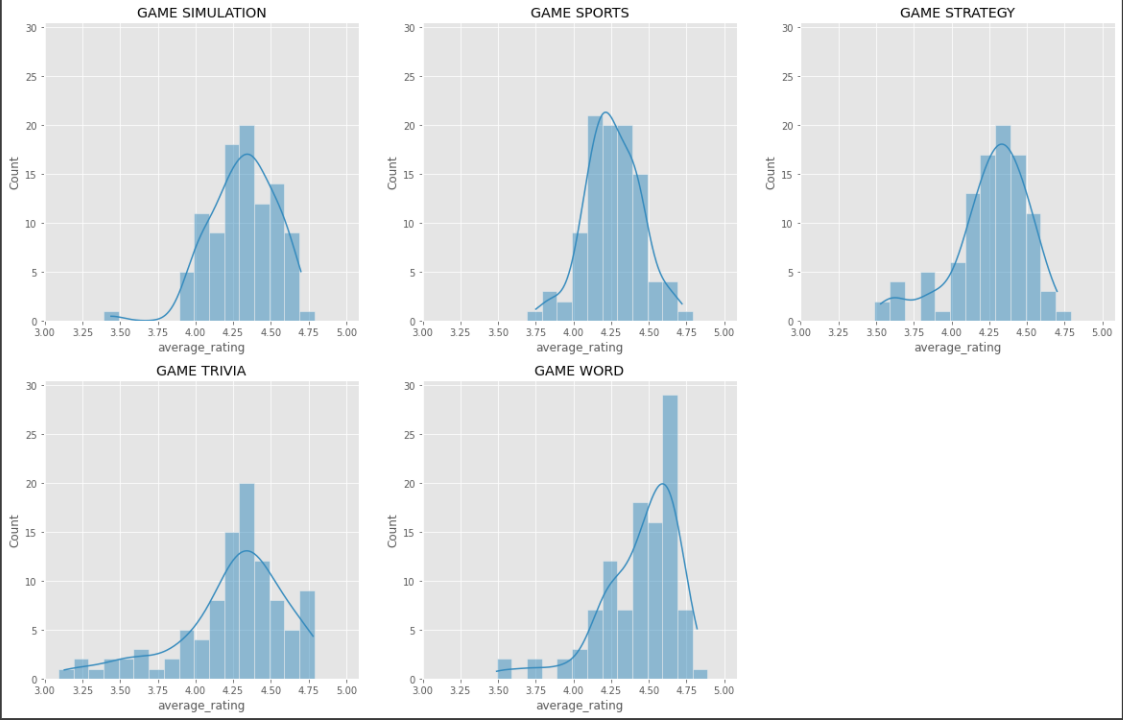
Tabel A.1

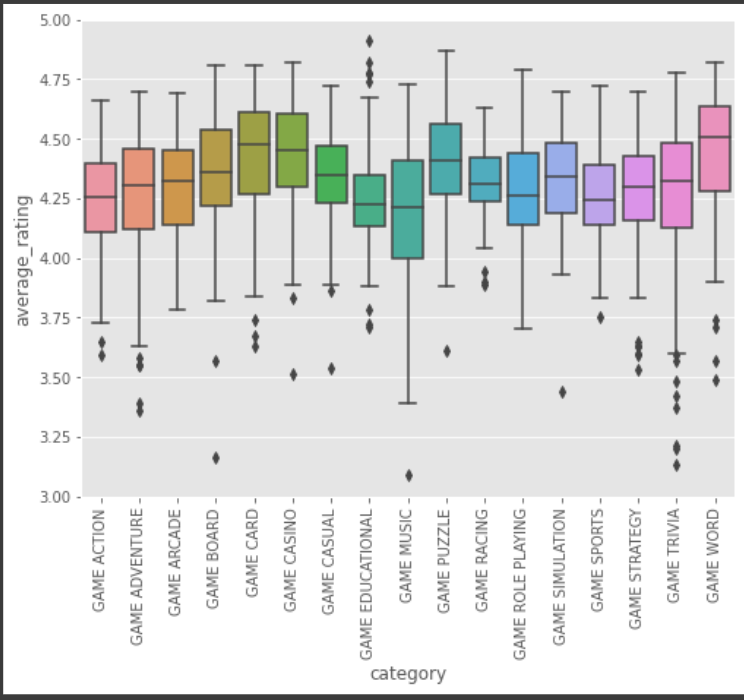


Terdapat 17 kategori game yang kami visualisasikan average ratingnya menggunakan histogram dan boxplot. Game yang ratingnya paling terdistribusi terpusat adalah game racing sedangkan yang paling bervariasi ratingnya adalah game trivia dan game music. Menurut kami ratingnya terdistribusi terpusat karena bentuk dan cara bermain dari game racing mirip-mirip. Seperti apapun bentuknya bisa racing mobil, motor, pesawat dan kapal pasti hanya mengelilingi lintasan dan berpacu pada waktu tercepat untuk memenangkan permainan. Lain hal dengan game music yang cukup beragam cara mainnya dan punya banyak variasi musik yang dipilih, begitu juga dengan game trivia yang memiliki jutaan pertanyaan yang berbeda. Salah satu yang menarik juga adalah hanya game edukasi yang memiliki pencilan di bagian atas. Game edukasi merupakan sarana pembelajaran yang efektif. Selain mendapatkan keseruan dalam memainkannya, pemain juga mendapatkan wawasan pengetahuan dari game tersebut. Karena itu, ada sebagian kelompok yang memberi rating 5 dan jarang ada yang memberi nilai dibawah 3.



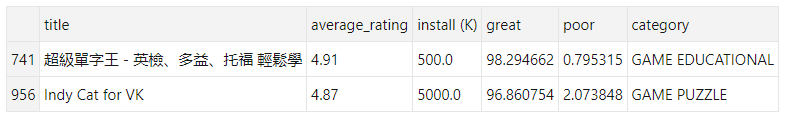






Menjawab pertanyaan:

1. Game dengan total instal lebih dari 100 juta sangat sedikit. Hanya terdapat 12 game, game tersebut terdiri 10 game dengan total install 500 juta ada 10 game dan 1000 juta ada 2 game. Jika jumlah game yang total instalnya diatas 100 juta dibandingkan dengan game lainnya, perbandingannya adalah 1:144 atau 0.69%. Rata-rata total instal game adalah 28 juta dengan standar deviasi 58 juta.
2. Kami melakukan sorting data berdasarkan “average\_rating”, “install” dan, “great” secara berturut-turut, kemudian mengambil 0,1% data tertinggi. Kami mendapatkan standar suatu game yang dapat disebut game terfavorit adalah jika game tersebut memiliki average\_rating lebih dari sama dengan 4,87.



1. Game dengan kualitas baik menurut kami adalah game yang memiliki average rating, jumlah instal, dan presentase great diatas rata-rata (average rating > 4.3 , jumlah instal > 28 M, great > 82%) selain itu juga memiliki presentase poor dibawah rata-rata (poor < 12 %). Maka melalui seleksi tersebut, terdapat 170 game yang berkualitas baik.

