

Análisis de las propiedades y tendencias de los juegos de mesa

¿Qué dataset se utiliza y cuáles datos contiene?

El proyecto se basará en un análisis sobre la tabla de datos: “bbg_dataset” que contiene información sobre variados juegos de mesa que salieron al mercado hasta febrero del 2021, incluye un total de 20 343 juegos de mesa recopilados. Acompañado de los datos están sus columnas con diferentes variables las cuales son: ID, nombre, año de publicación, mínimo de jugadores, máximo de jugadores, duración de juego, edad mínima, rating de los jugadores, media del rating, bbg ranking, media de complejidad, usuarios que poseen el juego, mecánicas y dominios.

Idea principal del código y resultados

La idea principal de este código es analizar las diferencias entre distintas variables por medio de gráficos con el fin de llegar a ver las correlaciones que hay de una variable a otra. Las relaciones que se tomarán en consideración para este proyecto son las siguientes:

- Edad mínima con la dificultad promedio: Revisar si esta tiene una buena correspondencia con lo que incita los juegos de mesa y llegar a una hipótesis sobre si una mayor edad mínima indica una mayor dificultad, análogamente con el caso contrario, aquí se usará un gráfico estilo histograma para ver los datos de complejidad promedio y luego se usará una de dispersión.
- Año de publicación y promedio de calificación: Visualizar en cuáles años hubo los mejores juegos de mesa acorde a la página “BoardGameGeek” y sus usuarios, está determinará si en los últimos años hubo un crecimiento o decrecimiento en la calidad de creación de juegos recientes, se usará un gráfico lineal.
- Dominio y año de publicación: Para ver qué tipo de juegos salieron con más frecuencia en el mercado y si estos siguen siendo igual de populares en la actualidad o han estado en decadencia, usando un gráfico barras apiladas será la forma más eficiente de verlo.
- Tiempo de juego y cantidad mínima de jugadores: Identificará sobre la idea de que los juegos de mesa con menor cantidad de personas tardan menos tiempo que aquellos que requieren más personas, usando un gráfico de caja de bigotes nos dirá los outliers y sus resultados.
- Juegos infravalorados y sobrevalorados: Notar cuáles son los juegos de mesa que se consideran “infravaloradas” dentro del dataset, este se hará mediante la estimación entre el ranking acorde a bbg y la calificación promedio que le dieron los usuarios, para mostrar la dispersión se usará ese tipo de gráfico y también un histograma para ver las comparativas con mayor claridad.
- Complejidad y calificación promedio: Responderá la duda sobre si aquellos juegos que poseen una complejidad más alta tienden a estar mejor calificados acorde a los usuarios o peor calificados, de igual manera se usará un gráfico de dispersión.

Este análisis exploratorio de los datos logrará ofrecer respuestas a varias dudas que puedan plantearse y de revisar bastante temas de interés, de hecho viendo los gráficos disponemos de distintas respuestas. Comenzando por el de edad mínima con dificultad promedio logramos ver con el gráfico que no hay una relación coherente sobre estas dos variables, de hecho si filtramos los valores en las dos últimas décadas daremos con resultados similares, por ello no podemos saber exactamente si es creciente o decreciente, pues la dispersión en el gráfico adhiere una poca relación entre ambas, con el año de publicación y promedio de calificación notamos que los juegos que estuvieron mejor calificados fueron en los años más recientes, de hecho ha habido un crecimiento bastante notable en la calificación brindada por los usuarios y un declive acentuado entre 1960 y 1970, esto mayormente se debe a la poca cantidad de juegos de mesa creadas a comparación con los años recientes que se ha notado una mayor cantidad de juegos entrando al mercado, ahora con año de publicación y dominio vemos que hubo una disminución de los juegos de guerra, abstractos, de niños, de familia y de fiestas recientemente, mientras que los juegos de estrategia han estado creciendo a un ritmo constante y por último los de otras categorías también han ido en creciente pues ante la gran cantidad de juegos entrando al mercado ha dado que se distribuya una mayor variedad dentro de la misma, por otro lado, con el de tiempo de juego y cantidad mínima de jugadores llegamos al mismo resultado que la edad mínima y es que en realidad no existe una relación exacta entre las categorías, el boxplot da que la media que usamos esté muy alejada del intercuantil seleccionado, y además las cajas son generalmente más largas destacando la de 8 jugadores que tienen una distribución bastante grande, esencialmente es gráfico indica que no hay una relación coherente por lo tanto no podemos llegar a una conclusión clara, la sección de juegos infravalorados es la más interesante, aquí se tomaron 10 juegos para el top construido, este dio los resultados de que: TerroriXico, Drinks & Daggers, Animator vs Animation, War Titans: Invaders Must Die!, Frontier: The Card Game, Erune, Thug Life the Game, 90°, Tomorrow Dies Today y Composition fueron los que ganaron en este top con el orden presentado, y si revisamos el gráfico vemos que la mayoría de juegos se considera infravalorados acorde a la comparación entre ranking de bbg y el promedio acreditado al juego por los usuario, y por último quedó la complejidad junto a la calificación promedio, estos dieron resultados de que a mediado de que sube la complejidad del juego, usualmente su calificación sube, entonces eso indica un crecimiento constante.

Recomendaciones:

La ventaja de este dataset es que brinda muchas formas de utilizarla, podríamos enfocarnos en un tema específico como de cuáles fueron los mejores calificados y porqué, cuáles atributos contienen a diferencia de otro juegos, qué tanto se consideraría la longitud más beneficiosa para un juego, será que aquellos juegos de mesa creados en el pasado son mejores que los actuales, y muchísimos otros enfoques y dudas, atribuye a usar mucho la imaginación y logra dar muchas salidas cautivadoras, otro detalle sería usar otros tipos de gráficos para cada tema que se quiera enfocar o seleccionar, es la parte más importante del trabajo por lo tanto debe de tomarse en cuenta cuál es el indicado que brinda la información clara y concisa.

Idea de la programación y librerías usadas

La principal biblioteca que se utilizará será “tidyverse”, el uso de tidyverse es para el enfoque de eliminar los datos que son NA dentro del dataset; también será usado con el fin de usar los filtros para algunas de las comparativas que se mencionaron en la sección anterior esto con el fin de que no haya ni repeticiones ni datos innecesarios para lo que se necesita. Uno de los paquetes principales que vienen dentro de tidyverse es “ggplot2” que será esencial para la graficación de las comparativas que se van a desarrollar, con este podremos ver los datos en un gráfico del estilo que selecciones para cada una de las partes. La idea de cada sección es la siguiente:

- *Edad mínima con la dificultad promedio:* Usaremos el la división de intercuartiles para ver los valores extremos de la dificultad promedio y poder limpiar aquellos datos, luego se filtraron los datos que cumplan tener una edad mayor y una complejidad mayor a 0 para luego graficar con un geom_point para crear la dispersión, y luego se filtraran los años para ver más claro los datos en un rango de fechas específicos y comparar los cambios.
- *Año de publicación y promedio de calificación:* Esta parte se filtraron los años de 1961 a 2021 para quitar valores innecesarios dentro del data frame, luego se usó geom_line para crear el gráfico lineal y resaltamos los puntos destacados con geom_point con el fin de acentuar los años en la gráfica.
- *Dominio y año de publicación:* Primero filtramos para quitar los valores perdidos del dominio y después agrupamos en casos específicos, filtramos denuevo los años que ocupamos y agrupamos la cuenta de juegos que cumplen el caso que necesitamos esto con el fin de hacer el gráfico de barras usando geom_bar, y scale_y_continuos para controlar los espacios/saltos. Se creo por aparte con los mismos datos nada más que con una filtración de años de 1960 a 2021 para ver los datos más recientes aparte, este usa la misma receta que la anterior.
- *Tiempo de juego y cantidad mínima de jugadores:* Aquí se sacan los intercuartiles para después filtrar los datos que tengo un tiempo de juego mayor o igual al límite inferior y menor al superior, también filtramos denuevo para hacer el conteo de los datos con n(), estos serán de ayuda para la creación del boxplot, esté marcará los outliers con rojo para verlos claramente y usará el stat_summary para mostrar la media con la comparativa que queremos hacer.
- *Juegos infravalorados:* En esta sección se usará una fórmula para la infravaloración, se decidió hacer la norma con el rating promedio y el ranking de bgg, estas si hicieron sacando el promedio menos el mínimo dividido entre el máximo menos el mínimo, de igual manera con el rating de bgg, la diferenciación entre ellas será la que determina si es sobrevalorado o infravalorado, con head(10) y arrange(desc()) ordena los datos en orden decreciente y mostrará los 10 más altos de esa diferencia que corresponde a los infravalorados de la tabla, ahora con geom_point hará el gráfico de dispersión que mostrará en azul los juegos sobrevalorados, blanco los medios y rojo los infravalorados, y geom_histogram se usará para ver la cantidad aproximada de juegos que hay en alguna de las dos áreas o en medio.
- *Complejidad y calificación mínima:* En esta parte sacaremos el promedio y mínimo del rating promedio, además dividiremos en partes la complejidad promedio para adaptarlos a alguna categoría todo con el mutate, ahora usamos el group_by para

sumar la cantidad y ver cuántas hay en cada una de las categorías. Por último, hacemos una gráfica de dispersión con las categorías y la calificación promedio, con `geom_point` será eso y con `geom_smooth` será para representar la regresión lineal.