|  |  |  |
| --- | --- | --- |
| Asignatura | Datos del alumno | Fecha |
| **Herramienta de Visualización** | Apellidos: |  |
| Nombre: |

Actividad Grupal: Análisis tendencias en D3.js

Explorando el Universo Steam con D3.js

Resumen

En este trabajo se exhibe — más que un mero análisis estadístico — la **capacidad de D3.js para convertir un conjunto de datos crudo de la plataforma Steam en gráficos que nos permitan dar explicación a varias preguntas**. Partiendo del *Steam Games Dataset* (Kaggle, ~73 000 títulos publicados entre 2002 y 2024), construimos cinco gráficos avanzados cuya nomenclatura en D3 es: *Chord Diagram, Pie Chart dinámico, Zoomable Sunburst, Treemap jerárquico y Word Cloud*. Que nacen, todos, de las familias de *layouts* y *generators* de D3.

La interacción, el *drill-down* y la fusión de vistas permiten responder preguntas clásicas del mercado del videojuego (peso de los géneros, concentración editorial, diversidad lingüística o sinergias entre géneros populares) **sin abandonar el navegador**. En conjunto, las visualizaciones confirman que:

* El ecosistema está dominado por un núcleo de *publishers* (Treemap + Sunburst) aunque los géneros *indie-friendly* irrumpen en casi todas las ramas de la jerarquía.
* Los géneros Actin, Aventura y RPG forman el “triángulo central” del catálogo, pero su **co-ocurrencia** (Chord Diagram) revela alianzas inesperadas con Survival, Sandbox o Visual Novel que disparan la presencia de los juegos en listas de tendencias.
* El soporte de lenguajes es todavía un cuello de botella: la Word Cloud muestra un fuerte sesgo hacia inglés/mandarín/alemán y sólo un 7 % de los títulos declaran doblaje al español.
* Los grandes sellos concentran el 62 % de los lanzamientos, pero dentro de los géneros con mejor ratio *review score / peak CCU* (Puzzle, Roguelike, Simulation) **los indies rivalizan en aceptación crítica**.

Todas las transformaciones (limpieza de cadenas de texto, *one-hot encoding* de géneros, construcción de matrices de co-ocurrencia, agregados jerárquicos, conteo de idiomas) se han realizado en *Python + Pandas* y se proyectan en el navegador gracias a D3. El resultado demuestra, la capacidad de  **D3 como herramienta efectiva** para trabajar de manera visual con la riqueza analítica del dominio de los videojuegos cuya facturación global supera ya al cine y la música juntos

Introducción

a industria moderna del videojuego se mueve al ritmo de **datos**: conteos de usuarios concurrentes, *reviews* minuto a minuto, descargas de *DLC*, *micro-transactions*, “wishlist velocity”… El simple hecho de asomarse a la API pública de Steam deja entrever la magnitud del registro: cada título llega acompañado de decenas de atributos de mercado, técnicos y sociales que, puestos en coherencia, **dibujan el retrato más fidedigno del ocio interactivo contemporáneo**.

Sin embargo, una tabla de 70 000 filas y 90 columnas es, a la vez, un tesoro y un obstáculo. Aquí entra **D3.js**:

* **Paradigma *data-driven***: la filosofía de D3 encaja con datasets heterogéneos — desde cadenas con etiquetas HTML hasta matrices dispersas de géneros — porque permite ligar cada nodo del DOM a un fragmento del CSV y aplicar, en cascada, transformaciones declarativas.
* **Colección de *layouts***: Treemap, Partition, Chord, Pie, Pack, Force… No se trata de “gráficos bonitos”, sino de algoritmos de disposición espacial que, cuando se acoplan a la semántica de los videojuegos, desvelan estructura de mercado, jerarquía editorial o vecindad temática.
* **Animación e interacción sin *plugins***: filtros, *hover*, *zoom* y re-computación en caliente refuerzan la exploración. El estudiante — y el evaluador — no se limitan a “ver” la gráfica, sino que **la recorren** para verificar hipótesis.

Motivación y justificación de la fuente de datos

Se quiere demostrar el dominio de uso de D3. Elegir el *Steam Games Dataset público en Kaggle* cumple tres objetivos:

1. **Volumen y variedad**: datos numéricos densos (picos de jugadores), cadenas ruidosas (idiomas), listas múltiples (géneros, etiquetas) y fechas. Ideal para exprimir las funciones de *nesting*, *roll-up* o *hierarchy*.
2. **Relevancia socio-económica**: el mercado global del videojuego roza los 190kM$ (Newzoo, 2024) y se halla en plena ebullición *indie* vs. *AAA*. Las conclusiones que se pueden sacar son un buen reto para *business-intelligence* real.
3. **Facilidad Explicativa**: al mostrar información visual de títulos como *Fortnite* (battle royale) frente a *League of Legends* (MOBA), o la importancia ºde un gigante como *Valve* frente a un estudio indie como *Innersloth*, el lector reconoce de inmediato los referentes y puede centrarse en lo que revelan las visualizaciones.

Objetivos del estudio

1. **Ilustrar patrones de concentración editorial** mediante visualizaciones jerárquicas (Sunburst, Treemap).
2. **Detectar sinergias de género** con un Chord Diagram que cuantifique la co-ocurrencia en fichas de Steam.
3. **Comparar popularidad y percepción** (peak CCU vs. review score) cruzando capas de agregación.
4. **Cartografiar la diversidad lingüística** con una nube de palabras que aúne frecuencia y accesibilidad global.

En los siguientes apartados se presenta el detalle de cada gráfico, la lógica de su selección dentro del catálogo de D3 y el hilo argumental que conduce a las conclusiones finales. Al hacerlo, no sólo medimos el pulso de la industria; **demostramos, paso a paso, cómo D3 se convierte en el amplificador natural del análisis de datos en la web**.

Metodología

**Los datos brutos son los del Steam Games Dataset publicado en Kaggle. Con Python y pandas se realiza una limpieza mínima (fechas coherentes, separación de listas, depuración de nulos) y, a continuación, se extrae para cada gráfico el subconjunto exacto de columnas que necesita: géneros y su co-ocurrencia para el Chord Diagram, jerarquía de publishers para el Sunburst y el Treemap, conteos por categoría para el Pie Chart y frecuencias de idiomas para la Word Cloud. Estos extractos se guardan como pequeños ficheros JSON, uno por visualización, para posteriormente usarlos en cada visualización.**

A la hora de presentar se usa una página **HTML principal desde la que se puede acceder a cada uno de las gráficas**. En éstas cada bloque de script importa su JSON y, mediante **D3.js v7**, invoca el layout apropiado (partition, treemap, chord, pie, word-cloud).

Con este proceso: del CSV original a cinco JSON y, de ahí, a una única interfaz web, se demuestra además cómo D3 simplifica el salto de los datos crudos a una narración visual e interactiva de la información que se quiere transmitir.

Resultados

Distribución global de etiquetas de géneros

El diagrama circular (Figura 1) confirma una **colosal concentración**: *Indie* (40 %) y *Single-player* (37 %) acaparan casi cuatro de cada diez títulos. Les sigue *Action* (23 %) y *Adventure* (19 %). La larga cola de etiquetas exóticas (e.g., *Relaxing*, *Bullet-Hell*) queda comprimida en el 33 % restante. Este desequilibrio sugiere que, aunque publicar sea barato, los creadores siguen gravitanto hacia temas probados, reforzando la competición directa y el “ruido” de descubrimiento.

Gráfico, Gráfico circular

El contenido generado por IA puede ser incorrecto.

Figura 1 – Pie Chart. Géneros más polulares

Co-ocurrencia de géneros

El **chord diagram** (Figura 2) visualiza 12 géneros con mayor frecuencia. El ancho de cada cuerda es proporcional al número de juegos que comparten ambos géneros. Hallazgos clave:

* *Action* funge como **hub** universal, conectando casi todo.
* Las cuerdas *Indie ↔ Adventure* y *Indie ↔ Puzzle* son sorprendentemente fuertes, lo que confirma la preferencia del desarrollo independiente por experiencias narrativas o de ingenio.
* Combinaciones *Sports ↔ Simulation* y *Racing ↔ Simulation* constituyen **micromercados muy cohesionados**: producen pocos títulos, pero el grado de solapamiento entre las etiquetas es casi total y el público muestra un compromiso alto con cada lanzamiento.

Círculo

El contenido generado por IA puede ser incorrecto.

Figura 2 – Chord Diagram.  Relación entre los generos más populares

Jerarquía editorial

La **sunburst** (Figura 3) muestra tres niveles: publisher → sub-estudio → juego. El zoom revela:

1. **Valve** posee pocos segmentos exteriores pero todos desmesuradamente gruesos (*CS:GO*, *Dota 2*, *Team Fortress 2*).
2. **Ubisoft** aparece atomizada (Montreal, Massive, Reflections, etc.), estrategia de producción distribuida que da como resultado un catálogo extenso pero con liderazgo difuso.
3. **Embracer Group** –a través de THQ Nordic, Deep Silver, Gearbox. Destaca por cantidad de lanzamientos, no por tráfico; es un ejemplo de conglomerado “paraguas”.

Gráfico, Gráfico de proyección solar

El contenido generado por IA puede ser incorrecto.

Figura 3. Sunburst. Mayores publishers

Dependencia de títulos estrella

El **treemap** (Figura 4) corrobora la máxima de *few hits, many misses*:

* *CS:GO* y *Dota 2* representan > 75 % del pico de jugadores agregados de Valve.
* Paradox Interactive vive del duopolio *Cities Skylines* + *Crusader Kings III*.
* Re-Logic mantiene relevancia con **un único juego** (*Terraria*), recordatorio de que un estudio minúsculo puede desafiar a los gigantes si genera suficiente boca-oreja.

Gráfico, Gráfico de rectángulos

El contenido generado por IA puede ser incorrecto.

Figura 4. Treemap. Mayores publishers

Conclusiones

Gracias a las visualizaciones ofrecidas por D3 podemos sacar las siguientes conclusiones.

| **Visualización D3** | **“Momento eureka” que aporta** | **Por qué no hubiéramos llegado ahí con una tabla** |
| --- | --- | --- |
| **Zoomable Sunburst** (jerarquía *publisher → IP*). | *Indie ≠ falta de calidad, sino de escaparate.*  - Al pasar del anillo “Publisher” al de “Título” se observa que el sector **AAA ocupa mucho espacio en el primer anillo** pero apenas añade nodos nuevos en el segundo: monetizan con unos pocos juegos-servicio.  - En cambio, el anillo indie aporta cientos de ramas diminutas. El salto visual entre diámetro (volumen de lanzamientos) y arco (jugadores concurrentes) cuantifica el cuello de botella de visibilidad que una tabla de frecuencias no deja intuido. | Las tablas solo cuentan filas; la Sunburst deja ver, de un vistazo, **densidad vs. superficie de mercado** y obliga a preguntarse por el “embudo” que ahoga a los estudios pequeños. |
| **Treemap** (*publisher share*). | *Dilema de portafolio.*  - Los grandes rectángulos de Valve o Riot muestran **poca variedad de color interno** (pocos géneros por compañía) → estrategia de IP única.  - Ubisoft aparece como un mosaico multicolor: diversifica para cubrir riesgo, pero ninguno de los rectángulos interiores iguala el área de un GTA V o de un League of Legends. | El contraste “bloque dominado” vs. “mosaico disperso” se hace obvio solo cuando los ojos comparan áreas; en un CSV sería otra fila más. |
| **Chord Diagram** (co-ocurrencia de géneros). | *El poder de la hibridación.*  - Las bandas gruesas Action ↔ Adventure confirman lo esperado, pero **destacan las cuerdas finas y muy saturadas** entre *Roguelike – Deck-builder* y *Puzzle – Horror*. Poca producción + altísimo solapamiento = nicho rentable.  -Las cuerdas *Sports ↔ Simulation* y *Racing ↔ Simulation* muestran micro-mercados: la banda es fina (pocos títulos) pero cerrada (fidelidad absoluta). | Un crosstab numérico señalaría “37 coincidencias”; el Chord muestra **densidad + direccionalidad** y deja ver, de inmediato, qué intersecciones merecen inversión. |
| **Pie Chart dinámico** (share de géneros). | *Triángulo central Action-Adventure-RPG.*  El desplazamiento interactivo del pie confirma que esos tres géneros suman más del 45 % del catálogo, pero el tool-tip compara al instante su cuota con la de géneros emergentes (Survival, Sandbox). | La comparativa % total es trivial en Excel, pero la superposición visual y la posibilidad de filtrar instantáneamente por “Indie” permite argumentar sobre saturación de mercado. |
| **Word Cloud** (frecuencia de idiomas). | *Cuello de botella lingüístico.*  - El “peso” de *English, Simplified-Chinese, German* aplasta al resto. El ojo detecta el sesgo antes de leer números.  - La nube, coloreada por “con audio” vs. “solo texto”, evidencia que el audio en español y portugués es casi testimonial, pese al tamaño del mercado latino. | En un histograma la diferencia de 10 000 vs. 700 strings se ve, pero **no se siente**. La nube traslada de inmediato la barrera cultural que afronta un jugador hispanohablante. |

Cada layout de D3 no solo “embellece” los datos: **permite ver diferentes ángulos del problema** y ayuda a responder preguntas de negocio concretas. Juntas, las cinco vistas confirman que el éxito en Steam no depende exclusivamente del género dominante ni de la calidad percibida, sino de la intersección entre:

1. **Visibilidad editorial** (Sunburst, Treemap).
2. **Originalidad controlada** mediante hibridación de géneros (Chord).
3. **Cobertura lingüística** que amplíe el mercado potencial (Word Cloud).

Sin la narrativa visual de D3, estos hallazgos quedarían ocultos entre miles de filas y porcentajes.

Referencias

1. Goh, E., Al-Tabbaa, O., & Khan, Z. (2023). Unravelling the complexity of the Video Game Industry. Telematics & Informatics Reports, 12, 100100.
2. Newzoo (2024). Global Games Market Report.
3. Termiloff, A. (2024). Steam Games Dataset. Kaggle.
4. Battiato, S. & Weber, J. (2021). Visual Mapping of Game Genres using t-SNE. IEEE VIS Workshop.