Histograma i dispersió

Repartiment de preus, i preu contra metascore

Nom _____

Superset → **Charts**, dataset `games_clean` (base de dades `Dades classe`, esquema `public`): 26,685 jocs de Steam. Tingueu **SQL → SQL Lab** obert en una altra pestanya. Penseu en el mode competitiu: un **histograma** compta els jugadors de cada rang; un **diagrama de dispersió** posa un punt per jugador. Treballeu sols i escriviu aquí què veieu.

Referència

	Gràfic de barres	Histograma
Eix X	Categories que ja existeixen.	Columna **numèrica** tallada en **bins** iguals.
Alçada	Qualsevol mètrica.	Sempre un recompte: les files dins del bin.
Respon	Quina categoria és més gran?	Quina forma tenen els valors?

Núvol de punts	Lectura
Puja cap a la dreta	Relació positiva.
Baixa cap a la dreta	Relació negativa.
Taca rodona	Cap relació clara.

Una relació no diu mai quina de les dues coses causa l'altra.

Part A · L'histograma de preus (18 min)

Distribució de preus. On són la majoria de jocs, i com és de llarga la cua?

Pas	Feu això	Hauríeu de veure / escriviu
1	Gràfic nou sobre `games_clean`, tipus **Histogram**. No és a la primera llista (**Featured**): clic a **All charts**, o escriviu `histogram` al cercador. **Column**: `price`.	**Column**, **Bins** i **Row limit**.
2	**Row limit**: canvieu `10000` per `50000`. Si no, el gràfic llegeix només 10,000 dels vostres jocs.	`50000` a **Row limit**.
3	Deixeu **Bins** a 5. **Create chart**. Passeu el ratolí per la primera barra.	Barres que es veuen de debò: _____ Rang de la primera barra: _____

4	A SQL Lab:
	`SELECT game, price FROM games_clean`
	`ORDER BY price DESC LIMIT 6;`
		El més car: _____
		El seu preu: _____ Són jocs? _____

5	De tornada al gràfic: **Filters** → `price` → operador **Less or equal (<=)** → escriviu `60` i trieu-lo a la llista → **Save**.	Un xip `price <= 60` sota **Filters**.
6	**Bins**: escriviu `12` i premeu Enter. **Update chart**.	Barres: _____ La més alta: la _____ La cua va cap a la _____
7	**Customize** → **Y Axis Format**: `,d`. Torneu a la primera barra.	Recompte sense arrodonir: _____
8	**Save** → `UT3 · Price distribution`.	Surt a la llista de **Charts**.
9	Opcional. Marqueu **Cumulative**, **Update chart**: cada barra suma totes les anteriors. No el deseu.	Proporció per sota de 10: _____

A1 Pas 3: els bins són trams iguals des de la fila més barata fins a la més cara. Per què una sola barra gegant?

A2 Llegiu el vostre histograma final en una frase que contingui un nombre: "La majoria de jocs…".

A3 Heu deixat fora els productes de més de 60. Són dades errònies? Què hauria de dir el títol del gràfic?

Part B · Preu contra metascore (16 min)

Preu contra metascore. Els jocs més cars reben millors ressenyes?

Pas	Feu això	Hauríeu de veure / escriviu
1	A SQL Lab: `SELECT has_metascore, COUNT(*) FROM games_clean GROUP BY has_metascore;`	Amb metascore: _____ de 26,685 Aproximadament, en %: _____
2	Gràfic nou sobre `games_clean`, tipus **Scatter Plot**.	Un gràfic buit a la vista Explore.
3	**X-axis**: `price`. **Metrics**: pestanya **Simple** → columna `metascore`, agregat `AVG`.	Els dos controls omplerts.
4	**Filters** → `has_metascore` → operador **Is true** → **Save**. **Create chart**.	Un núvol de punts entre 0 i 60 a l'eix X.
5	Passeu el ratolí per tres punts: el de 9.99, el de 59.99 i el més baix.	A 9.99: _____ A 59.99: _____ El més baix: _____ a preu _____
6	**Save** → `UT3 · Price vs metascore`.	Surt a la llista de **Charts**.

B1 Molts jocs costen 9.99, però només hi ha un punt a 9.99. Mireu el control **Metrics**: què representa un punt?

B2 El núvol puja, baixa, o cap de les dues coses? Feu servir la referència.

B3 De quins punts us refiaríeu menys, i per què?

B4 Un amic llegeix el vostre gràfic i diu: "A Steam, apujar el preu d'un joc en millora les ressenyes." Trobeu els dos errors. El pas 1 us ajuda amb un.

1. _____
2. _____

Heu acabat quan Tots els espais són plens · hi ha dos gràfics desats: `UT3 · Price distribution` i `UT3 · Price vs metascore`.

Encallats?

- **Les barres sumen 10,000**: **Row limit** encara és a 10,000. Poseu-hi 50,000.
- **`price` no surt a** Column: el control només llista columnes numèriques. Aviseu el professor.
- **Encara una sola barra gegant després del filtre**: el xip `price <= 60` ha de sortir sota **Filters**, i heu de prémer **Update chart**.
- **"Must specify a value for filters with comparison operators"**: el xip diu `price <=` sense cap nombre. Feu-hi clic, escriviu `60`, trieu-lo a la llista, **Save**.
- **Dispersió amb un sol punt**: `price` ha anat a **Metrics** en lloc de **X-axis**.
- **Punts repartits uniformement, com barres**: **Force categorical** està marcat sota l'eix X. Desmarqueu-lo.
- **Superset no s'obre**: poseu-vos amb un company i aviseu el professor.

A casa Res d'obligatori per vacances. Opcional, per al projecte final: amb el vostre conjunt de dades obert, un histograma d'una columna numèrica. També té valors atípics? Els gràfics `UT3 ·` desats són els vostres apunts per al gener.