

Resumen del capítulo: Análisis de cohortes

Análisis de cohortes

Una **cohorte** es un conjunto de personas que comparten un cierto suceso dentro del mismo período temporal.

Para formar una cohorte es necesario determinar:

- 1) el **evento** que tienen en común;
- 2) el **periodo de tiempo** dentro del cual debe ocurrir este evento

Un evento es un caso registrado de la realización de una determinada acción por parte de un/a usuario/a.

El período es el lapso de tiempo durante el cual ocurrió el evento.

Una vez que hayas definido el evento y el período de tiempo puedes **formar una cohorte**. En otras palabras, agrupas a los usuarios y usuarias que realizaron una cierta acción durante el mismo período de tiempo.

Los/las analistas dividen la base de clientes en cohortes para identificar los patrones en el comportamiento de usuarios. Toman una cohorte y analizan su comportamiento durante un periodo de tiempo. Las conclusiones generadas a raíz del análisis de cohortes pueden sentar las bases para modificar productos y procesos, así como para plantear nuevas hipótesis. Pero esto es, en última instancia, un análisis retrospectivo que muestra cómo se comportaron los usuarios y usuarias en el pasado. En muy pocas ocasiones se utiliza para prever el comportamiento de usuarios, que se verá afectado por factores externos en el futuro.

Evaluación de cambios en valores absolutos por mes

La mayoría de las observaciones cuantitativas se registran como valores absolutos.

Un **valor absoluto** es el volumen, el tamaño o la magnitud de un evento o fenómeno observado. Puede ser, por ejemplo, un precio total en dólares o compras totales en

unidades.

La primera tarea que se puede resolver mediante el análisis de cohortes es realizar un seguimiento de los cambios en la cantidad de compradores activos dentro de una sola cohorte. Esto nos dirá cuántas personas continúan haciendo compras en los meses posteriores al mes de la primera compra.

El seguimiento de los valores absolutos por mes puede ayudarte a identificar los cambios que se producen a lo largo del tiempo. Además, revela características comunes a todas las cohortes, como caídas estacionales en el número de usuarios. También podemos sacar una conclusión sobre el porcentaje de cada cohorte en el número total de compradores por mes.

Evaluación de los cambios en valores relativos por ciclo de vida

Otra tarea del análisis de cohortes consiste en averiguar cómo cambia el volumen de **compra promedio** (ventas totales divididas entre el número de clientes) a lo largo del tiempo. En el contexto de las cohortes tenemos que calcular el tiempo del **ciclo de vida**. Muestra en qué mes ocurrió un evento respecto a la distancia desde el mes de la cohorte.

Visualización del análisis de cohortes

Un **mapa de calor** es la visualización de una tabla donde las celdas varían en color dependiendo de su proximidad a los valores máximo y mínimo.

Para crear un mapa de calor en Python necesitarás el método `heatmap()` de la librería `seaborn`. Sus parámetros son:

- **annot=True** (anotar), que significa que el valor se mostrará para cada celda;
- **fmt='.1f'** (formato), que establece el formato de impresión (aquí, un decimal);

- **linewidths=1**, que establece el ancho de la línea que separa las celdas del mapa de calor (aquí, 1 píxel)
- **linecolor='gris'**, que establece el color de la línea (aquí, gris).

```
import seaborn as sns
from matplotlib import pyplot as plt

plt.figure(figsize=(13, 9))
sns.heatmap(dataframe, annot=True, fmt='.1f', linewidths=1, linecolor='gray')
```

Tasa de retención y tasa de cancelación

El análisis de cohortes se utiliza para analizar el comportamiento de los usuarios y usuarias de productos digitales. **La tasa de retención y la tasa de cancelación** son las métricas que se evalúan con más frecuencia.

La tasa de retención te cuenta cuántos usuarios/as de una cohorte siguen activos respecto a su número inicial.
$$\text{Tasa de retención} = \frac{\text{número de usuarios activos en el periodo } n \text{ del ciclo de vida}}{\text{número de usuarios activos en el periodo } 0 \text{ del ciclo de vida}} * 100\%$$

La tasa de cancelación te señala la proporción de usuarios que dejan de utilizar el producto a lo largo del tiempo. A diferencia de la tasa de retención, la tasa de cancelación representa una comparación del número de clientes en una etapa determinada con su número en la etapa anterior:
$$\text{Tasa de cancelación} = \frac{\text{el número de usuarios activos en el periodo } n \text{ del ciclo de vida}}{\text{el número de usuarios activos en el (n-1) periodo del ciclo de vida}} * 100\%$$

Para calcular la tasa de cancelación en Python, utiliza el método **pct_change()**. Este método permite encontrar la variación porcentual de una columna con respecto a la fila anterior del marco de datos. Encuentra el cambio porcentual en una columna en comparación con el valor de la fila anterior. Si aplicas este método junto con la agrupación, se ejecutará dentro del grupo.

```
cohorts['churn_rate'] = cohorts.groupby(['column1'])['column2'].pct_change()
```

Cohortes por comportamiento

El análisis de cohortes por comportamiento es un tipo de análisis que consiste en seleccionar una cohorte de usuarios que han realizado una acción o una secuencia de acciones durante un periodo de tiempo determinado. El objetivo es estudiar cómo una métrica objetivo (por ejemplo, la tasa de retención) cambió, con el tiempo, para la cohorte por comportamiento.