

Estudio de caso: ¿Cómo logra un sistema de bicicletas compartidas alcanzar el éxito rápidamente?

Introducción: En este caso de estudio trabajaremos con una empresa ficticia de ciclismo, precisamente hablando de una compañía de bicicletas compartidas. Nuestro objetivo será responder una serie de preguntas de negocio, en conjunto con miembros clave de nuestro equipo de trabajo.



Escenario: Como un analista de datos junior en el equipo de análisis de marketing de Cyclistic, una empresa de bicicletas compartidas en Chicago, participé en un proyecto estratégico cuyo objetivo principal era incrementar la cantidad de membresías anuales.

El director de marketing considera que el éxito futuro de la compañía depende de lograr que más usuarios se conviertan en miembros recurrentes. Por ello, mi equipo y yo nos enfocamos en analizar las diferencias en el comportamiento entre los usuarios ocasionales y los miembros anuales.

A partir de este análisis, desarrollamos una propuesta de estrategia de marketing orientada a convertir a los usuarios ocasionales en miembros anuales. Antes de implementarla, nuestras recomendaciones debían ser aprobadas por el equipo directivo de Cyclistic, por lo que me enfoqué en respaldarlas con hallazgos relevantes basados en datos reales y visualizaciones profesionales que comuniquen de manera clara y efectiva las oportunidades identificadas.

Equipo:

- **Lily Moreno:** La directora de marketing y mi supervisora directa.
- **Equipo de análisis de marketing de Cyclistic:** Un equipo de analistas de datos responsable de recopilar, analizar y reportar información que ayuda a guiar la estrategia de marketing de Cyclistic.
- **Equipo ejecutivo de Cyclistic:** El equipo ejecutivo, conocido por su enfoque meticuloso en los detalles, será el encargado de decidir si aprueba o no el programa de marketing propuesto.

Fase 1: Preguntar – Planteamiento del problema

La directora de marketing de Cyclistic, Lily Moreno, ha planteado tres preguntas estratégicas para guiar el desarrollo de un nuevo programa de marketing. De estas, se me ha asignado específicamente la siguiente:

¿Cómo utilizan las bicicletas de Cyclistic los miembros anuales en comparación con los usuarios ocasionales?

El objetivo de este análisis es **explorar las diferencias clave en el comportamiento de uso** entre estos dos tipos de usuarios mediante el análisis de datos históricos de viajes. Esto permitirá **identificar patrones y tendencias** que sirvan como base para **diseñar estrategias de marketing efectivas** orientadas a **incrementar la conversión de usuarios ocasionales en miembros anuales**, lo cual es una prioridad para el crecimiento sostenible de la empresa.

Fase 2: Preparación de los datos

Se utilizó el repertorio de datos históricos de los viajes realizados por los usuarios proporcionados por **Divvy** disponibles en:

<https://divvy-tripdata.s3.amazonaws.com/index.html>

Se descargaron los **12 archivos CSV correspondientes a los 12 meses previos** (de junio 2024 a mayo 2025) para realizar un análisis completo y representativo del comportamiento de los usuarios. Los datos provienen del sistema real de bicicletas compartidas del estado de Chicago, Estados Unidos administrado por Motivate International Inc., y están disponibles bajo una licencia de uso público.

Los archivos contienen información estructurada de los viajes realizados por los usuarios con campos de nombre:

ID del viaje, hora de inicio/finalización, ID de bicicleta, duración del viaje, ID de estación de inicio y fin, nombres de las estaciones de inicio y fin, tipo de usuario (cliente casual o miembro), género y año de nacimiento

trip_id	start_time	end_time	bike_id	tripduration	from_station_id	from_station_name	to_station_id	to_station_name	user_type	gender	birth_year
---------	------------	----------	---------	--------------	-----------------	-------------------	---------------	-----------------	-----------	--------	------------

Fase 3: Procesamiento de los datos (análisis exploratorio).

Herramientas utilizadas para hacer el procesamiento de los datos.

Para el análisis del caso de Divvy, se utilizó el lenguaje de programación **R**, apoyado de las librerías tidyverse, lubridate, janitor, ggplot2 y dplyr. Estas herramientas fueron seleccionadas por su eficiencia y versatilidad en la manipulación de datos, análisis estadístico y visualización.

En particular, tidyverse ofrece una sintaxis coherente y fluida para filtrar, transformar y resumir datos, mientras que lubridate facilita el manejo de fechas y tiempos. La razón del uso de estas herramientas es porque permiten un flujo de trabajo reproducible, eficiente y legible para trabajar el análisis de datos de forma tabular.

Integridad y estructura inicial.

Se cargaron 5.6 millones de registros desde los archivos csv y se unificaron los 12 datasets de manera vertical después de estandarizar los nombres de las columnas.

Se utilizó glimpse(), summary() y colSums(is.na()) para:

- Verificar nombres, clases y resumen estadístico de las columnas.
- Identificar columnas con valores faltantes (start_station_name, end_station_name, end_lat, end_lng).

Limpieza de los datos.

Aplicación de clean_names() para normalizar nombres (snake_case).

Se eliminaron los viajes con valores nulos en estación de inicio y fin.

```
divvy_data_clean <- divvy_data %>%  
  filter(!is.na(start_station_name) & !is.na(end_station_name))
```

Se calculó la duración en minutos y se creó una nueva columna para almacenar estos datos de nombre ride_length

```
mutate(ride_length = as.numeric(difftime(ended_at, started_at, unit =  
  "mins")))
```

Eliminación de valores negativos:

```
filter(ride_length >= 0)
```

Generación de fechas derivadas:

Se añadieron columnas: date, month, day, year, day_of_week usando lubridate.

El data frame resultante (divvy_data_clean) contiene:

- Columnas clave:
 - ride_id, rideable_type, started_at, ended_at, ride_length
 - start_station_name, end_station_name, member_casual
 - date, month, day, year, day_of_week
- Registros depurados: sin valores nulos en columnas críticas ni duraciones negativas.
- Usuarios clasificados correctamente como "casual" o "member".

Documentación reproducible

Todo el proceso fue documentado en el script R, con pasos secuenciales y comentados, asegurando:

- Reproducibilidad del análisis.
- Claridad para terceros.
- Preparación adecuada para visualización, análisis estadístico o exportación a otras herramientas.

Fase 4. Analisis

Una vez que los datos fueron limpiados, estructurados y enriquecidos con variables temporales, se procedió a la fase de análisis exploratorio. El objetivo de esta etapa fue identificar **tendencias, patrones y relaciones relevantes** que ayuden a responder las preguntas de negocio clave, tales como: *¿cómo se comportan los diferentes tipos de usuarios?* y *¿cuáles son los periodos con mayor actividad en el sistema de bicicletas?*

Los datos fueron **organizados y agregados** por usuario (member_casual) y por día de la semana (day_of_week) para permitir comparaciones claras. Estas dimensiones permiten responder preguntas importantes sobre los hábitos de uso, diferenciando entre usuarios frecuentes (miembros) y esporádicos (casuales). Además, se aseguraron los formatos adecuados, como convertir la columna day_of_week en un factor ordenado para garantizar visualizaciones cronológicamente coherentes.

Se realizaron **cálculos de medidas de tendencia central** como promedio (mean()), mediana (median()), duración máxima y mínima de los viajes por tipo de usuario. Estos cálculos revelaron que:

- **Los usuarios casuales** tienden a tener viajes de mayor duración promedio que los miembros.
- **Los miembros** realizan más viajes entre semana, lo cual sugiere un uso más funcional o de rutina (por ejemplo, para transporte diario).
- **Los usuarios casuales** se concentran principalmente en fines de semana, lo cual podría estar asociado al ocio o al turismo.

Al comparar el **número de viajes por día de la semana**, se observó que la mayor actividad ocurre los **sábados y domingos**, especialmente por parte de usuarios casuales, mientras que los miembros tienen una distribución más estable a lo largo de los días hábiles.

Estas observaciones se apoyaron visualmente mediante **gráficos de barras** generados con ggplot2, mostrando tanto el número de viajes como la duración promedio por día y tipo de usuario. Las visualizaciones facilitaron la identificación clara de diferencias en los patrones de comportamiento.

Fase final. Conclusiones

La historia que emerge de los datos es clara:

- **Los usuarios casuales** tienden a usar las bicicletas más los fines de semana y realizan viajes de mayor duración, lo cual sugiere un uso recreativo o turístico.
- **Los miembros anuales** utilizan las bicicletas principalmente entre semana y sus viajes son de menor duración, indicando un uso más orientado a la rutina o transporte diario (ej. al trabajo o escuela).

Desde el inicio, el objetivo fue entender cómo difiere el comportamiento entre miembros y usuarios casuales. Los hallazgos responden directamente a esta pregunta al evidenciar diferencias en:

- **Frecuencia de uso**
- **Duración de los viajes**
- **Días preferidos de uso**

Estos insights ofrecen una base sólida para **diseñar campañas personalizadas**, como estrategias de conversión de usuarios casuales a miembros anuales mediante promociones durante fines de semana o experiencias personalizadas.

Los datos sugieren que los **usuarios casuales podrían convertirse en miembros** si se diseñan campañas centradas en **experiencias recreativas** y beneficios de membresía durante fines de semana.

El **perfil de miembro actual** valora la **eficiencia y disponibilidad**. Esto apunta a mejoras en estaciones cercanas a zonas corporativas o residenciales.

Los miembros anuales utilizan las bicicletas como medio de transporte regular, mientras que **los usuarios casuales lo hacen principalmente con fines recreativos**.

Este entendimiento profundo del comportamiento del usuario permitirá al equipo de Cyclistic diseñar **estrategias más efectivas de marketing, operación y retención**, alineadas a los hábitos reales de uso.