

Cyclistic Capstone

Carlos Moreno

2025-08-14

1) Resumen Ejecutivo

Cyclistic busca maximizar membresías anuales.

Objetivo: Identificar diferencias en el uso entre *casual riders* y *annual members* para diseñar estrategias de conversión.

Preguntas clave: - ¿Cómo difieren en duración, frecuencia y días de uso? - ¿Qué patrones estacionales existen?

2) Contexto y Pregunta de Negocio

2.1 Business Task

Analizar y comparar el uso del sistema Cyclistic por parte de *casual riders* y *annual members* usando los datasets Divvy 2019 Q1 y Divvy 2020 Q1 (limitación técnica de RStudio gratuito), para identificar diferencias de comportamiento que fundamenten una estrategia de conversión.

Entregables: a) Enunciado del encargo b) Fuentes de datos c) Documentación de limpieza d) Resumen de análisis e) Visualizaciones y hallazgos clave f) Tres recomendaciones priorizadas

2.2 Stakeholders

- **Lily Moreno (Marketing Director):** espera insights accionables y visualizaciones pulidas.
 - **Marketing Analytics Team:** produce y reporta el análisis.
 - **Cyclistic Executive Team:** decide si aprueba la estrategia basada en hallazgos.
-

3) Datos y Metodología

3.1 Fuentes y metadatos

Archivo	Fuente	Período	Formato	Filas aprox.
Divvy_Trips_2019_Q1.csv	Divvy (Chicago)	Ene-Mar 2019	CSV	~365,000
Divvy_Trips_2020_Q1.csv	Divvy (Chicago)	Ene-Mar 2020	CSV	~400,000

Fuente: Motivate International Inc. *Divvy Bicycle Sharing Data*. Disponible en: <https://divvy-tripdata.s3.amazonaws.com/index.html> Licencia: Creative Commons Attribution 4.0.

4) PREPARE

Verificación de existencia de datasets crudos y carga inicial en RStudio Cloud.

```
# Carga segura del dataset procesado
# La ruta ahora sube dos niveles (../..) para llegar a la raíz del proyecto,
# y de ahí, accede a 'Cyclistic_Capstone_Files/data_processed/clean_data.rds'
# y 'Cyclistic_Capstone_Files/scripts/02_data_cleaning.R'
if (!file.exists("../..Cyclistic_Capstone_Files/data_processed/clean_data.rds")) {
  source("../..Cyclistic_Capstone_Files/scripts/02_data_cleaning.R")
}

clean_data <- readRDS("../..Cyclistic_Capstone_Files/data_processed/clean_data.rds")

cat("Dimensiones del dataset:", dim(clean_data), "\n")

## Dimensiones del dataset: 783778 17
```

5) PROCESS

Comparación de estructuras entre 2019 y 2020.

- Columnas 2019 no presentes en 2020: trip_id, start_time, end_time, bikeid, tripduration, from_station_id, from_station_name, to_station_id, to_station_name, usertype, gender, birthyear
- Columnas 2020 no presentes en 2019: ride_id, rideable_type, started_at, ended_at, start_station_name, start_station_id, end_station_name, end_station_id, start_lng, start_lng, end_lng, member_casual

Decisiones: - Renombrar columnas de 2019 para coincidir con 2020 - Agregar columnas faltantes como NA - Traducir codificación de member_casual (Subscriber → member, Customer → casual) - Calcular ride_length y day_of_week - Filtrar outliers (<1 min o >1440 min)

- Dimensiones finales: ~783k filas × 17 columnas.
-

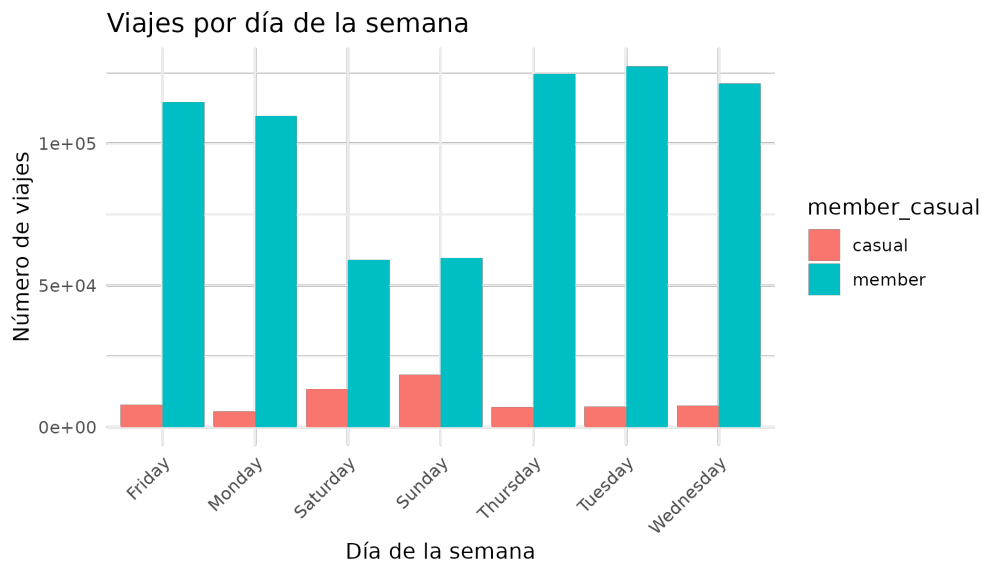
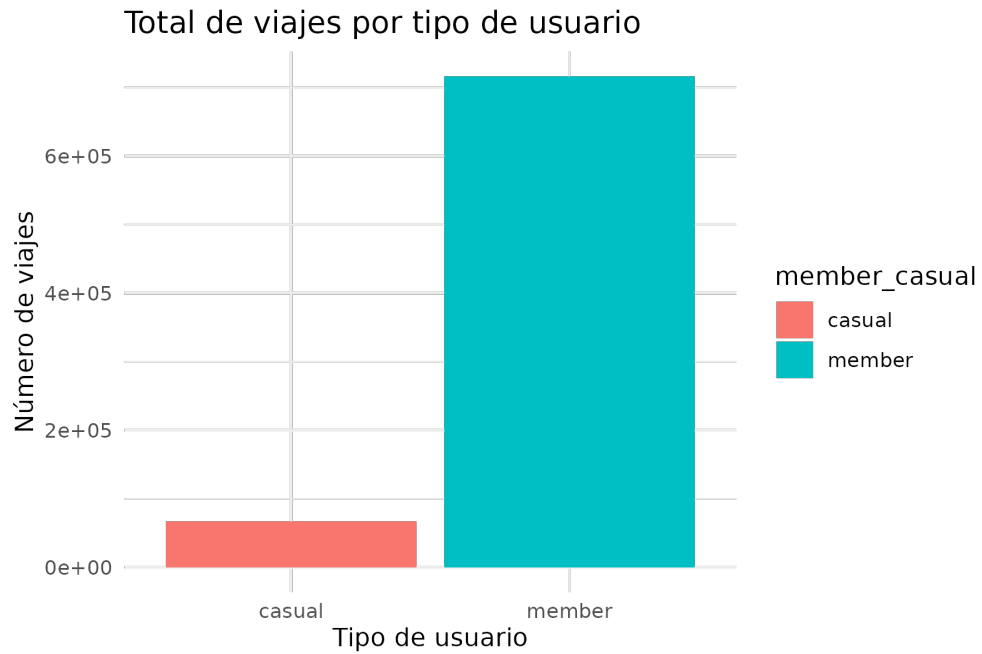
6) ANALYZE

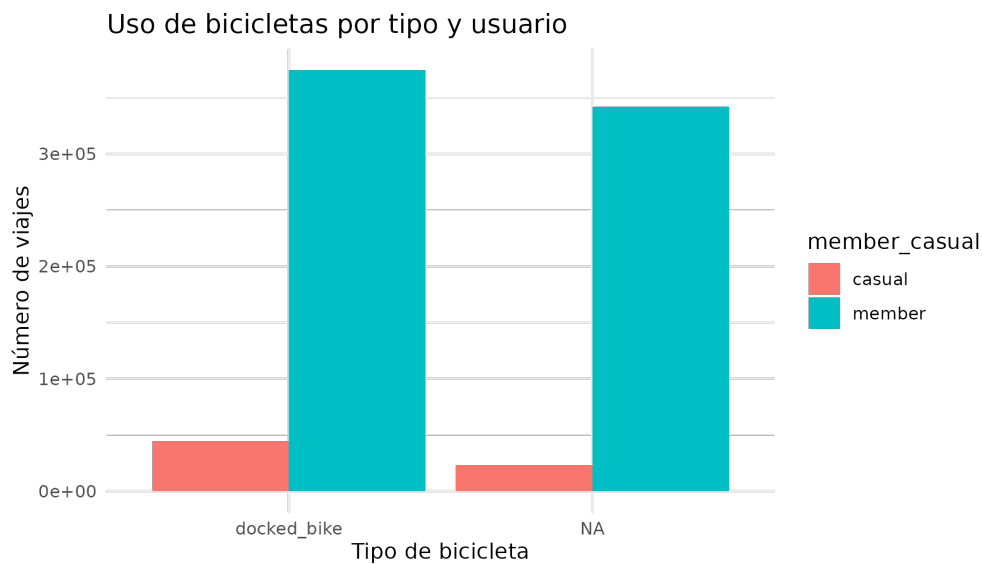
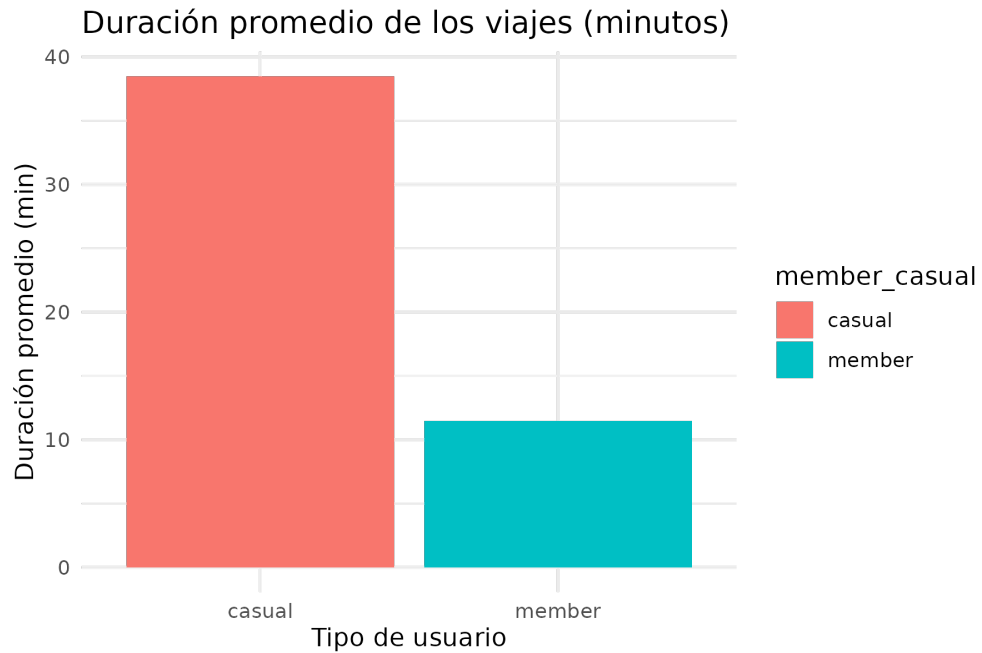
```
library(dplyr)

summary_by_user <- clean_data %>%
  group_by(member_casual) %>%
  summarise(
    total_viajes = n(),
    duracion_promedio_min = round(mean(ride_length, na.rm=TRUE), 1),
    duracion_total_horas = round(sum(ride_length, na.rm=TRUE)/60, 0),
    .groups="drop"
  )
```

La línea 'summary_by_user' se eliminó para evitar que se imprima
la tabla sin formato aquí, ya que se mostrará formateada en la Sección 8.

7) VISUALIZE





8) SHARE — Summary Table

```
## Warning in prettyNum(r, big.mark = big.mark, big.interval = big.interval, :
## 'big.mark' and 'decimal.mark' are both '.', which could be confusing
## Warning in prettyNum(r, big.mark = big.mark, big.interval = big.interval, :
## 'big.mark' and 'decimal.mark' are both '.', which could be confusing
## file:///tmp/RtmpwdgdHy/file395385d2d43.html screenshot completed
```

Tabla 1. Resumen de Uso por Tipo de Usuario Cyclistic

Comparación de métricas clave entre miembros y usuarios casuales.

Tipo de Usuario	Total de Viajes	Duración Promedio (min)	Duración Total (horas)
casual	67.397	38.5	43.228
member	716.381	11.5	136.916

Fuente: Análisis de datos de viajes Cyclistic (Q1 2019-2020)

9) Hallazgos clave

- **Members** representan >90% de viajes, cortos y constantes
- **Casuals** realizan menos viajes, más largos y en fines de semana
- Diferencia clara: **recreación vs. transporte**

10) Recomendaciones

1. **Campañas recreativas fin de semana**
 - Beneficios en ocio, descuentos familiares
2. **Conveniencia y ahorro urbano**
 - Mensajes de costo-beneficio (“5 viajes = membresía pagada”)
3. **Optimizar experiencia digital y flota**
 - Promocionar bicicletas cómodas y acceso exclusivo para miembros

11) Limitaciones

- Datos restringidos a Q1-2019 y Q1-2020
- Diferencias en esquema entre años
- Variables demográficas incompletas

12) Conclusión

- Diferencias claras permiten estrategias de conversión
- Casuals → enfocar en recreación
- Members → reforzar eficiencia semanal y ahorro
- Potencial para aumentar membresías y estabilizar ingresos

13) Anexo — Trazabilidad mínima

- Datos: Divvy (Chicago) Q1-2019 y Q1-2020; licencia CC BY 4.0
- Pipeline: unificación de esquemas, limpieza de outliers, cálculo de ride_length y day_of_week
- Artefactos: clean_data.rds y gráficos en Cyclistic_Capstone_Files/outputs/