

# Cyclistic: Estudo de Caso do Google Analytics Professional Certificate

Marcelo Henrique Guidini Angeli

2025-07-31

## Contexto da análise

### A empresa

A Cyclist é uma empresa fundada em 2016 e que atua no segmento de compartilhamento de bicicletas, triciclos e similares. Atualmente a empresa conta com uma frota de aproximadamente 5.800 veículos e mais de 690 estações atuando, principalmente, na região de Chicago, EUA.

### Modelo de Receita

A Cyclistic possui três modelos para aluguel: passe único, passe diário e passe anual (membro cyclistic).

### Do Objetivo

Atualmente a Cyclistic possui uma base sólida de clientes, contudo a maioria deles assinam o passe único ou diário. Esses modelos de assinatura são significativamente menos rentáveis para a empresa. Assim, foi solicitado pela gerente de marketing uma análise que visa melhorar a taxa de conversão de usuários esporádicos para membros Cyclistic. Para guiar as futuras decisões da equipe de marketing pediu-se ao presente analista a elaboração de um relatório que responda a seguinte pergunta: *1. Como os membros anuais e os ciclistas casuais usam as bicicletas da Cyclistic de forma diferente?*

# Diferentes perfis dos clientes da Cyclistic

## Introdução

O presente relatório tem como objetivo documentar o processo de análise das diferenças entre o uso das bicicletas de usuários comuns (passe único/live) e membros Cyclistics (passe anual). Ademais, propõe-se a utilizar de tais achados para concender *insights* para as próximas campanhas de marketing da companhia.

## A análise

### Preparação dos Dados

Os dados utilizados são públicos e fornecidos por uma empresa real chamada Divvy. Para obter os dados clique aqui. Como este é um caso de estudo hipotético assumirei que os dados são da Cyclistic.

Utilizou-se os dados de julho de 2024 até junho de 2025 na presente análise. Segue código para importação dos dados e criação de um único dataframe contendo todas as informações agrupadas. Nesse dataframe adicionou-se duas colunas uma com o ano e a outra com o número do mês da corrida.

```
return_csv <- function(file_name){
  return(read.csv(paste(getwd(),file_name, sep="/")));
}

return_col_names <- function(file_name){
  if(!length(file_name)){
    stop("Sem arquivos .csv no diretorio atual. Verifique o diretorio atual.");
  }
  return(colnames(return_csv(file_name)))
}

file_names <- sort(list.files(pattern = "\\..csv$"))

# Verificando se os arquivos podem ser mergeados.

col_names <- return_col_names(file_names[1])
for (file_name in file_names) {
  if(!identical(col_names, return_col_names(file_name))){
    stop("Impossivel fazer o merge ja que os arquivos possuem colunas diferentes")
  }
}

# Criando uma coluna para indicar o ano e outra para indicar o mes da corrida.
# Vou assumir que os arquivos são nomeados conforme o padrão: YYYYMM-divvy-tripdata.csv
dataframe_list <- list()
for(file_name in file_names){
  mes <- as.numeric(substr(file_name, start=5, stop=6))
  ano <- as.numeric(substr(file_name, start=1, stop=4))
  new_dataframe <- cbind(data.frame(return_csv(file_name),
    travel_month=mes, travel_year=ano))
  dataframe_list[[length(dataframe_list) + 1]] <- new_dataframe
}
dataframe_merged <- do.call(rbind, dataframe_list)
```

Também criou-se uma coluna com o tempo de cada viagem. Veja o código:

```
# Criando a coluna duração da viagem em segundos

dataframe_merged <- dataframe_merged %>%
  mutate(travel_time = abs(
    as.numeric(difftime(
      as.POSIXct(ended_at), as.POSIXct(started_at), units = "secs"))))
```

## Limpeza dos dados

Agora que foi criado um único dataframe consolidado os dados serão limpos. Existem 43 viagens em que o horário de começo é depois do horário de término. Essas viagens serão removidas para evitar inconsistências.

```
removed_trips <- dataframe_merged %>%
  filter(started_at > ended_at)

dataframe_merged <- dataframe_merged %>%
  filter(started_at < ended_at)
```

Após a eliminação das colunas duvidosas eliminou-se as possíveis linhas repetidas.

```
# Eliminando as colunas repetidas
dataframe_merged <- dataframe_merged %>%
  distinct(ride_id, .keep_all = TRUE)
```

Além disso verificou-se se a coluna “rideable\_type” continha apenas os valores: “electric\_bike”, “classic\_bike” e “electric\_scooter”. O mesmo foi realizado para a coluna member\_casual utilizando os termos: “member” e “casual”.

```
# Fazendo a validacao de dados da coluna "tipo de veiculo" e member_casual
tipos_de_veiculo <- unique(dataframe_merged$rideable_type)
print(tipos_de_veiculo)
```

```
## [1] "electric_bike"      "classic_bike"       "electric_scooter"
```

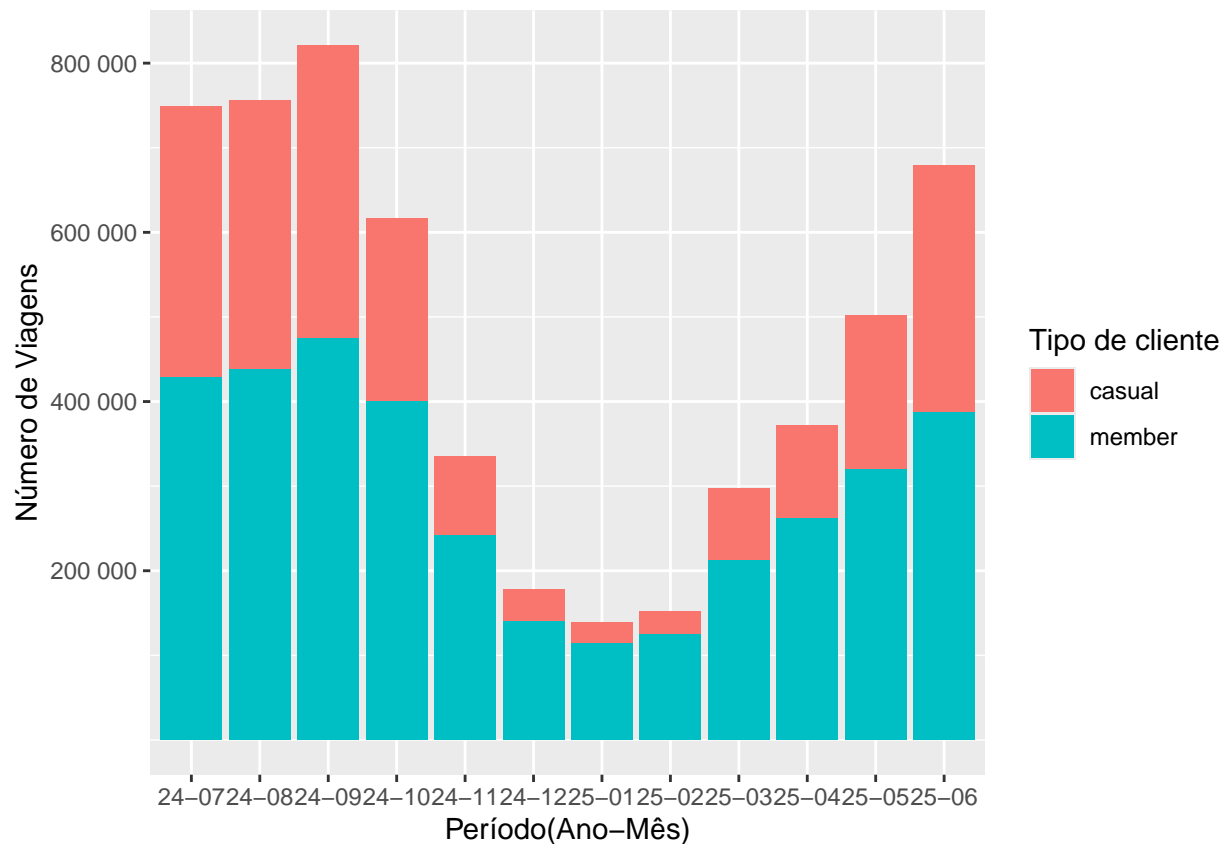
```
tipos_de_membro <- unique(dataframe_merged$member_casual)
print(tipos_de_membro)
```

```
## [1] "casual" "member"
```

## Análise

Inicialmente verificou-se a distribuição de corridas ao longo do ano agrupada por membros e não membros. No gráfico é possível observar que os membros mantem um número mais consistente de viagens ao longo do ano com significativo incremento durante os meses de maio à agosto. Já os não membros são mais sazonais, com alta representatividade nos meses de junho à outubro e quase nenhuma entre dezembro e fevereiro. Acredita-se que essa ocorrência está correlacionada com as estações do ano norte-americano, visto que o tanto o número de corridas quanto a representatividade dos não membros é muito menor durante os meses que compreendem o inverno.

```
#Encontrando o número de viagens por mês
ggplot(dataframe_merged) +
  geom_bar(mapping = aes( x = ifelse(travel_month < 10,
    paste(substr(as.character(travel_year), 3, 4),
      travel_month, sep = "-0"),
    paste(substr(as.character(travel_year), 3, 4),
      travel_month, sep = "-")),
    fill = member_casual)) +
  scale_y_continuous(breaks = c(200000, 400000, 600000, 800000),
    labels = label_number()) +
  xlab("Período(Ano-Mês)") + ylab("Número de Viagens") + labs(fill="Tipo de cliente")
```

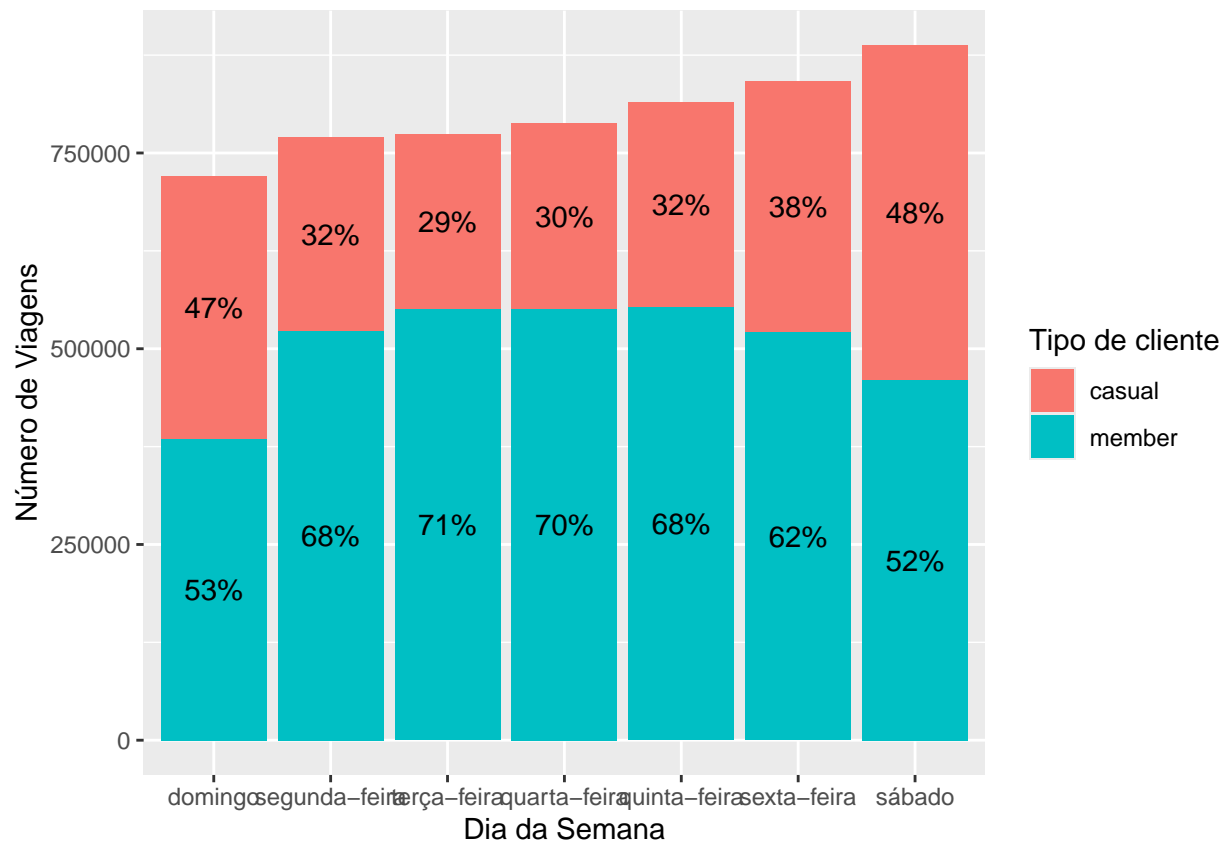


Posteriormente, verificou-se os dias da semana de maior ocorrência de viagens, novamente agrupados pela categoria do cliente. No gráfico mostrado percebe-se que os casuais tendem a preferir sábados e domingos enquanto os membros costumam optar pelos demais dias da semana. Tal situação, provavelmente, baseia-se no uso recreativo das bicicletas pelos não membros, enquanto os membros as usam para suas tarefas cotidianas.

```
#Encontrando as viagens por dia da semana
trips_per_category_per_weekday <- dataframe_merged %>%
  group_by(member_casual = dataframe_merged$member_casual,
    weekday = factor(weekdays(as.Date(started_at)),
      levels = c("domingo", "segunda-feira", "terça-feira",
        "quarta-feira", "quinta-feira", "sexta-feira", "sábado"))) %>%
  count() %>%
  group_by(weekday) %>%
```

```
mutate(daily_percentage = n/sum(n))

ggplot(trips_per_category_per_weekday, aes(x = weekday, y = n, fill = member_casual)) +
  geom_col(position = "stack") +
  xlab("Dia da Semana") + ylab("Número de Viagens") + labs(fill="Tipo de cliente") +
  geom_text(aes(label = scales::percent(daily_percentage, accuracy = 1)),
            position = position_stack(vjust = 0.5))
```



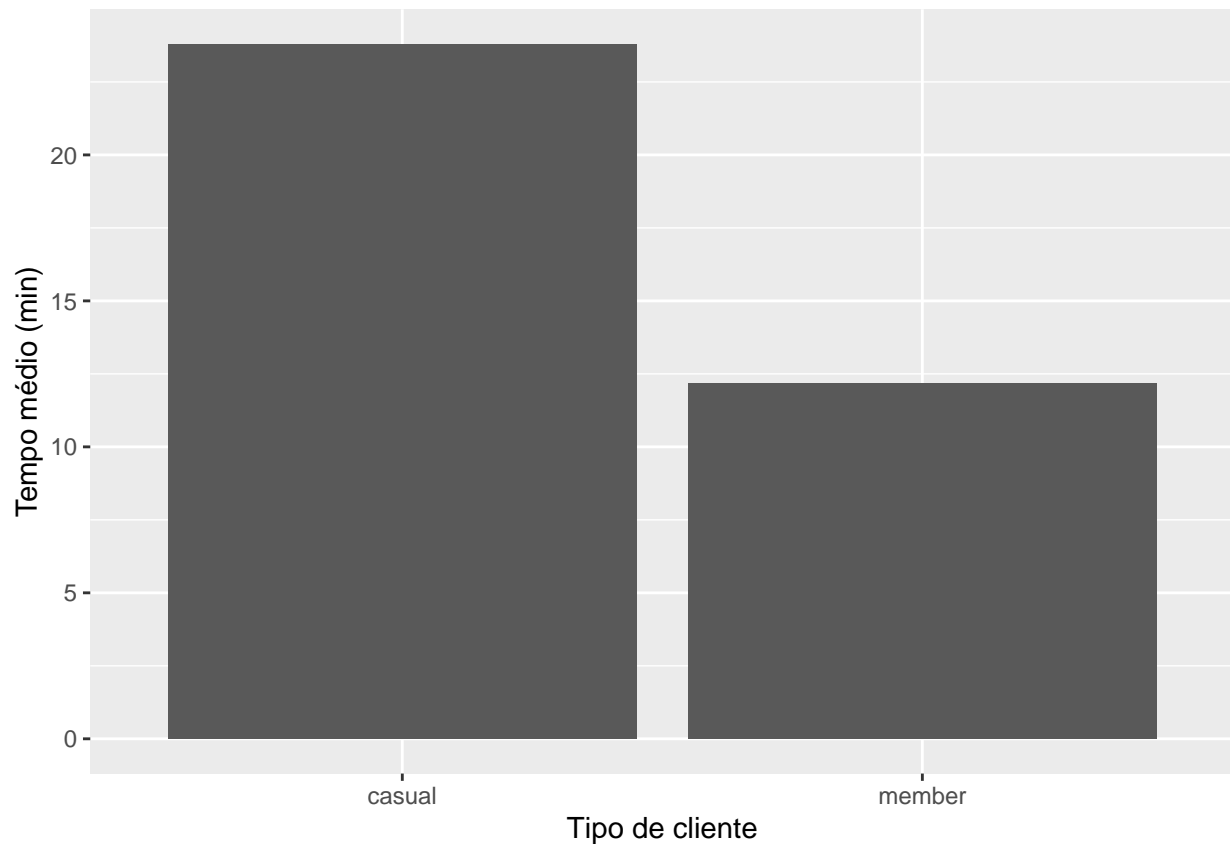
Já no que tange ao tempo médio de duração das corridas os não membros tendem a apresentar corridas de maior duração. Curiosamente, os não membros possuem um tempo médio de aproximadamente 24 minutos, enquanto nos membros esse valor cai para 13 minutos. Segue tabela que demonstra os resultados obtidos:

```
average_travel_time <- dataframe_merged %>%
  group_by(member_casual = dataframe_merged$member_casual) %>%
  summarise(average = mean(travel_time, na.rm=TRUE)/60)

print(average_travel_time)
```

```
## # A tibble: 2 x 2
##   member_casual average
##   <chr>          <dbl>
## 1 casual        23.8
## 2 member        12.2
```

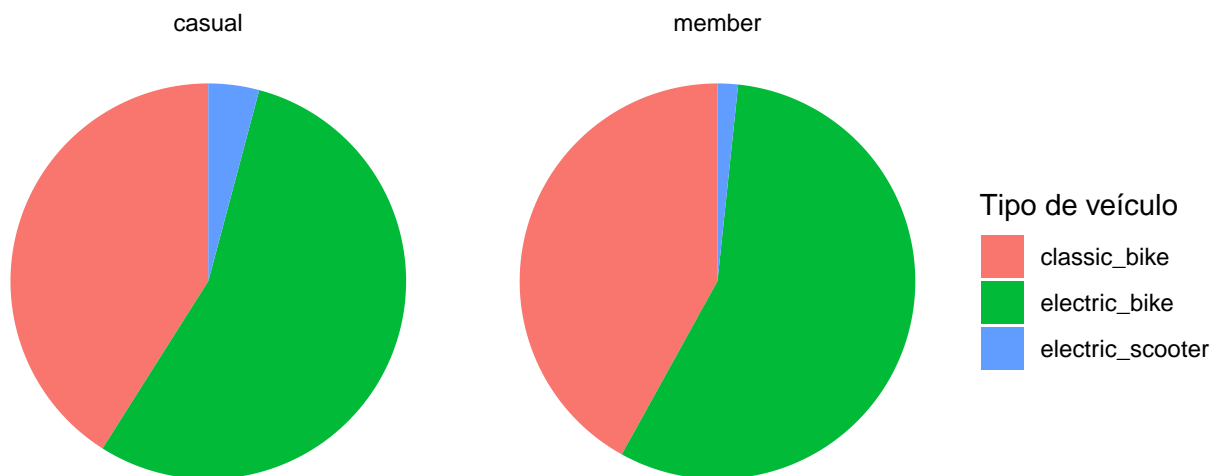
```
ggplot(average_travel_time) + geom_col(mapping = aes(x = member_casual, y = average)) +
  xlab("Tipo de cliente") + ylab("Tempo médio (min)")
```



Finalmente, verificou-se o tipo de preferência do veículo de cada grupo de clientes. Ambos apresentaram perfil similar, mas os casuais tiveram uma maior preferência pelos veículos elétricos.

```
ride_counts_percent <- dataframe_merged %>%
  count(rideable_type, member_casual) %>%
  group_by(member_casual) %>%
  mutate(percent = n / sum(n))

ggplot(ride_counts_percent, aes(x = "", y = percent, fill = rideable_type)) +
  geom_col(width = 1) +
  coord_polar("y", start = 0) +
  facet_wrap(~member_casual) +
  xlab("Tipo de veículo") +
  ylab("Proporção de Viagens") +
  labs(fill = "Tipo de veículo") +
  theme_void()
```



### ### Resultados

Os clientes, de modo geral, optam por locar os veículos principalmente durante os meses de junho à novembro. Durante esse período a proporção de corridas de não membros e membros é similar. Contudo, durante a baixa temporada (dezembro à março) os membros transformam-se na maioria esmagadora. Assim, visando converter os casuais, é recomendado que as propagandas sejam feitas durante o período do verão e da primavera.

Além disso, os não membros preferem usar as bicicletas durante os fins de semana. Assim, é ideal que os anúncios sejam exibidos durante esses dias. Curiosamente, os membros possuem viagens significativamente menores em média.

Por fim, ambos os grupos preferem os veículos elétricos. Apesar disso, os casuais demonstram uma disposição ainda maior para os veículos motorizados. Dessa maneira, é aconselhado que esse seja o tipo de bicicleta presente na maioria dos anúncios da empresa.