

# **Análise de dados para empresa (fictícia) de bicicletas para ajudá-la a atrair mais passageiros**

*Autor: G.Corrêa - Dezembro 2023*

## **1. Tarefa de negócios**

*1.1 Atual estratégia de marketing*

*1.2 Questão-Objetivo da Análise*

*1.3 Stakeholders*

## **2. Fontes de dados**

## **3. Preparação e Processamento de Dados**

*3.1 Dados incompletos e/ou inconsistentes*

*3.2 Preparação/Limpeza e Processamento*

## **4. Análise**

*a) Cinco estações mais usadas para para início de passeios*

*b) Comparação principal entre ciclistas CAUSAIS e MEMBROS*

*c) Outras observações*

## **5. Apresentação visual**

## **6. Recomendações**

*a) Sugestão principal*

*b) Sugestões secundárias*

## **7. Códigos para revisão**

## 1. Tarefa de negócios

Desde 2016, a empresa Cyclistic tem como Tarefa de Negócios a oferta bem-sucedida de serviço de compartilhamento de bicicletas.

Dados da empresa mostram **5.824 bicicletas** com rastreamento geográfico e bloqueio dentro de uma rede de **692 estações** na cidade de Chicago, Illinois, EUA.

As bicicletas podem ser desbloqueadas em uma estação e devolvidas em qualquer outra estação a qualquer momento.

Diferencial da Cyclistic é oferecer **bicicletas reclináveis, triciclos manuais e bicicletas de carga**, tornando o compartilhamento de bicicletas mais inclusivo.

Informações preliminares indicam que a maioria dos ciclistas opta por bicicletas tradicionais e a flexibilidade de preços ajuda a atrair clientes, mas os **membros anuais são mais lucrativos** do que os passageiros casuais.

### 1.1 Atual estratégia de marketing

- Conscientização geral e no apelo a amplos segmentos de consumidores
- Três planos de preços: **passes de viagem únicos** ou **dia inteiro** (passageiros CASUAIS) e **planos anuais** (MEMBROS).

**1.2 Questão-Objetivo da Análise:** a maximização do número de membros anuais, em vez de continuar tentando atrair novos clientes casuais, pode aumentar a lucratividade?

- Como os ciclistas MEMBROS e CASUAIS os passageiros casuais diferem entre si?
- Por que passageiros CASUAIS iriam querer adquirir um plano?
- Como a mídia digital ajudaria a influenciar o marketing?

### 1.3 Stakeholders:

- Equipe de Análise de Dados
- Depto. Marketing da Cyclistic
- Equipe executiva da Cyclistic
- Ciclistas CASUAIS
- Ciclistas MEMBROS anuais
- Clientes em Potencial

## 2. Fontes de dados

Dados históricos de uma empresa real.

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

(Por questão de privacidade, informações pessoais foram extraídas.)

Dados públicos foram disponibilizados pela Motivate International Inc., sob licença:

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

(A segurança e acessibilidade dos dados também foram previamente filtrados pela entidade que os disponibilizou originalmente.)

O projeto considerou o recorte dos 12 últimos meses, para identificar padrões e tendências mais recentes.

## 3. Preparação e Processamento de Dados

### 3.1 Dados incompletos e/ou inconsistentes:

Os dados estão aparentemente incompletos, pois a descrição da rede alega haver 692 estações, enquanto os registros analisados identificaram **1593 nomes de estações**.

A descrição inicial também cita uma frota de 5.824 bicicletas. Porém, os registros não indicam quantas ou quais bicicletas únicas foram utilizadas.

Os IDs das estações também não estão padronizados.

Ciclistas CASUAIS não são discriminados “por viagem” nem “por dia”.

São discriminadas apenas três tipos de bicicletas: “**clássicas**”, “**elétricas**” ou “*docked*”, quando sabemos que um dos atrativos da empresa é disponibilizar um número mais variado de tipos de bicicletas. Além disso, não houve como estabelecer em que consiste o registro “*docked\_bike*”, fazendo com que esse tipo de bicicleta (usada em número muito menor) fosse classificada como “**indefinida**”

### 3.2 Preparação/Limpeza e Processamento

São muitos dados para serem trabalhados apenas com planilhas. Por isso, a análise foi feita no **R Studio Desktop** pois a linguagem R permite trabalhar com os mais de 6 milhões de registros originais.

#### *Tarefas:*

- checados formatação/tipos de dados
- apagados registros duplos
- apagados registros com erros. P. ex., onde a duração do passeio era “negativa”.
- criando 4 colunas a partir de outras

Identificados cerca de **1 milhão** de registros com dados ausentes ("start\_station\_name", "start\_station\_id", "end\_station\_name" ou "end\_station\_id"). Esses registros foram mantidos por representarem uma parcela significativa do total e por não comprometerem totalmente a análise sobre os períodos dos passeios.

Foram identificadas **246.945 viagens** com menos de 2 minutos de duração, o que pode não indicar “não-passeios”.

Após processamento de dados, todos registros de viagem foram unificados em uma única tabela unificada ("datatrip\_total2") **5.652.827 entradas** em **17 colunas**.

Também foi criada tabela reduzida ("tripdata\_total\_reduzido"), por meio de seleção aleatória. (\*A margem de erro estimada para uma amostra aleatória de 1% de uma população de 5,6 milhões, considerando um nível de confiança de 95%, é menor do que 0,02%)

#### 4. Análise

a) *Cinco estações mais usadas para início de passeios:*

1º NA (Dados Incompletos)	15.3%
2º Streeter Dr & Grand Ave	1.12 %
3º DuSable Lake Shore Dr & Monroe St	0.71 %
4º Michigan Ave & Oak St	0.66 %
5º DuSable Lake Shore Dr & North Blvd	0.63 %

Apesar da Estação “Streeter Dr & Grand Ave” ser o ponto de partida mais citado, o peso relativo dela na comparação com as demais estações é similar (+0.4%), não permitindo considerações mais aprofundadas sobre essas diferenças.

b) *Comparação principal entre ciclistas CAUSAIS e MEMBROS:*

Ciclistas CASUAIS, por um lado:

- são **um terço** do total de passeios
- apresentaram **queda na quantidade** de passeios nos 2 últimos meses

por outro lado:

- são mais numerosos aos  **finais de semana**
- nos finais de semana, passeiam por  **mais tempo**
- a mediana do tempo dos passeios é **+3 min** superior à dos MEMBROS
- preferem as bicicletas **elétricas** em geral
- preferem **clássicas** para passeios demorados
- têm apenas um pico de uso diário : **16h-18h**
- preservação tendência do aumento da **quantidade de passeios**
- preservação tendência do aumento do **tempo gasto**

Ciclistas MEMBROS, por um lado:

- **dois terços** do total de passeios
- tiveram **aumento constante na na quantidade** de passeios no ano
- são mais numerosos **durante os dias de trabalho**
- a mediana do tempo dos passeios é **-3 min** inferior à dos CASUAIS
- têm uma leve preferência por bicicletas **elétricas**
- têm dois picos de uso diário : **8h e 16h-18h**
- aumento acumulado da **quantidade de passeios** e do **tempo gasto**

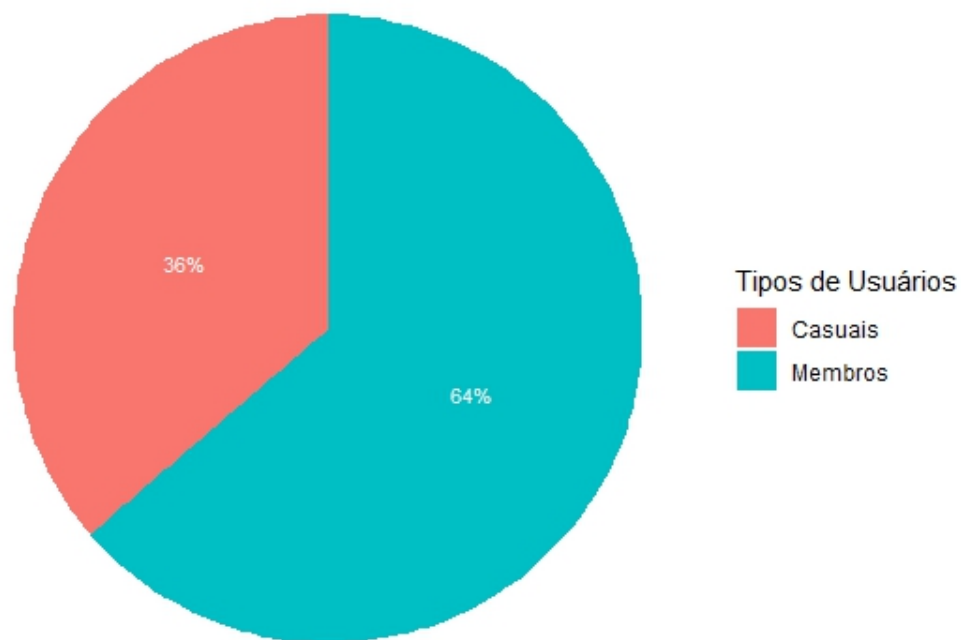
c) *Outras observações:*

- MEMBROS e CASUAIS têm tendência leve a iniciar passeios mais tarde

## 5. Apresentação Visual

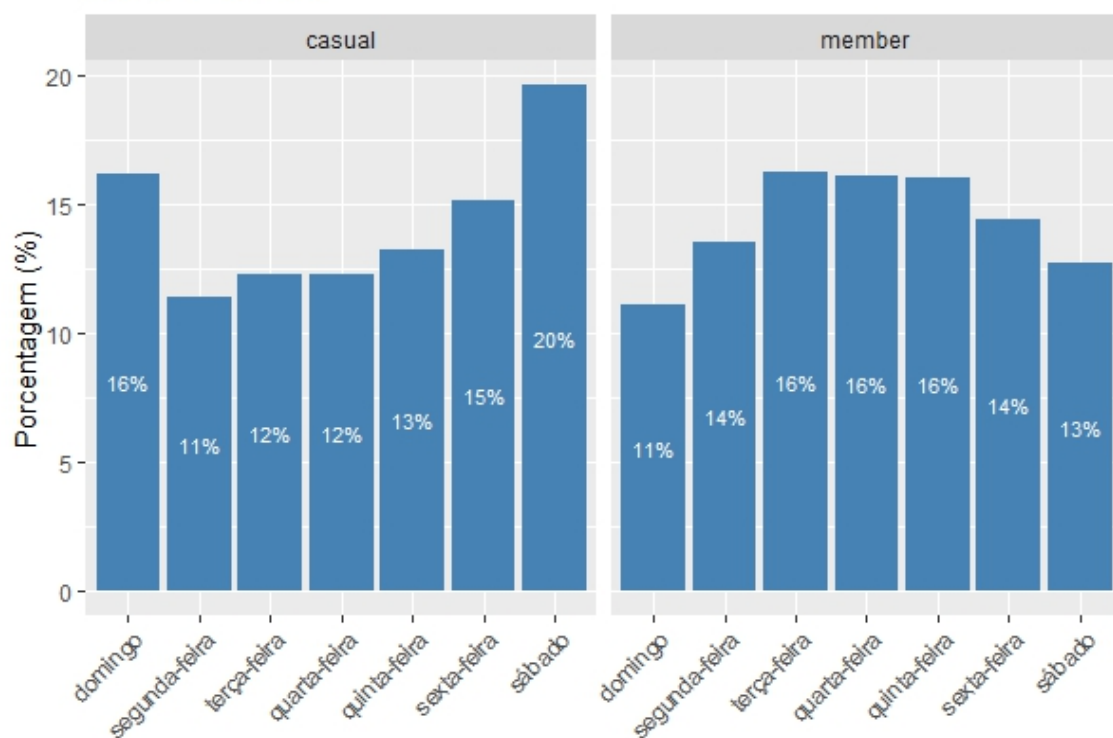
Total de passeios: 5.652.827

nov.2022 - out.2023



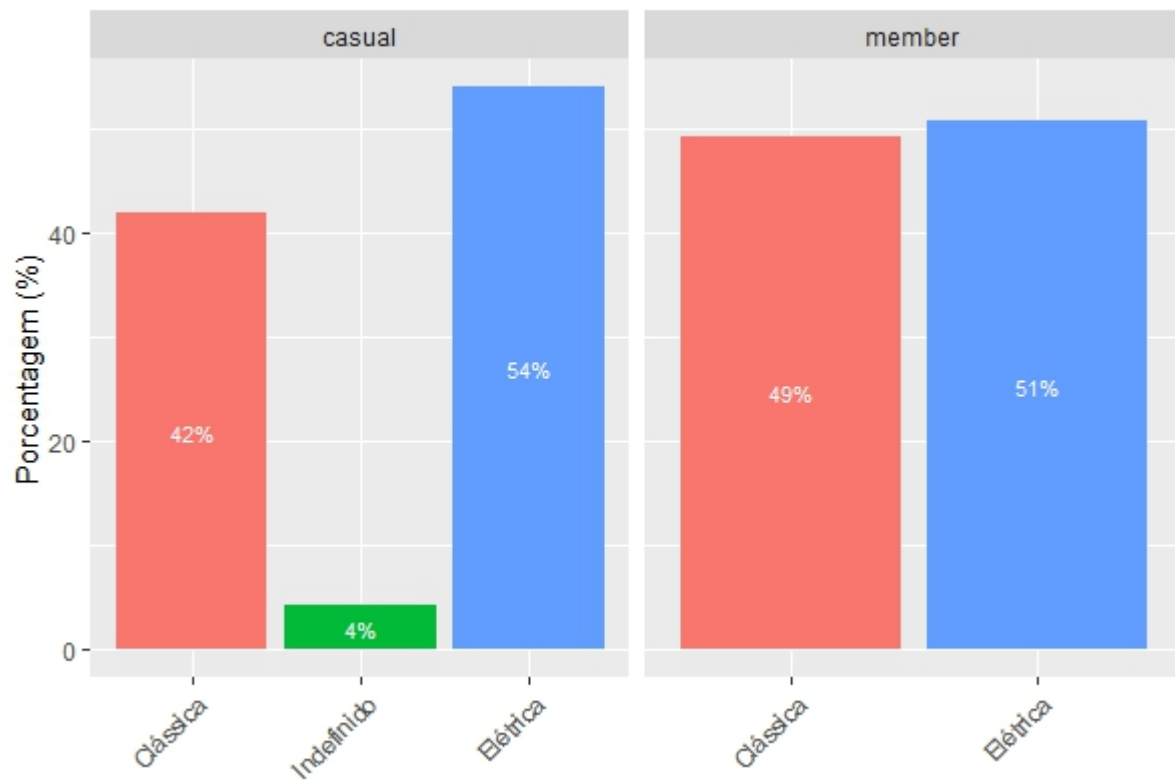
### Distribuição de usuários x Dia da Semana

nov.2022 - out.2023



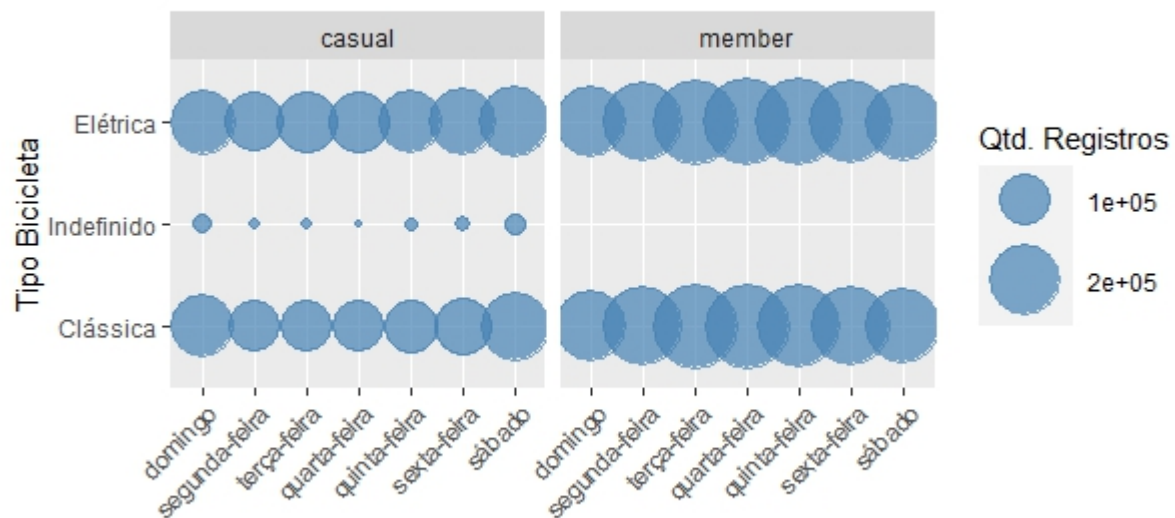
## Modelos de Bicicleta

nov.2022 - out.2023



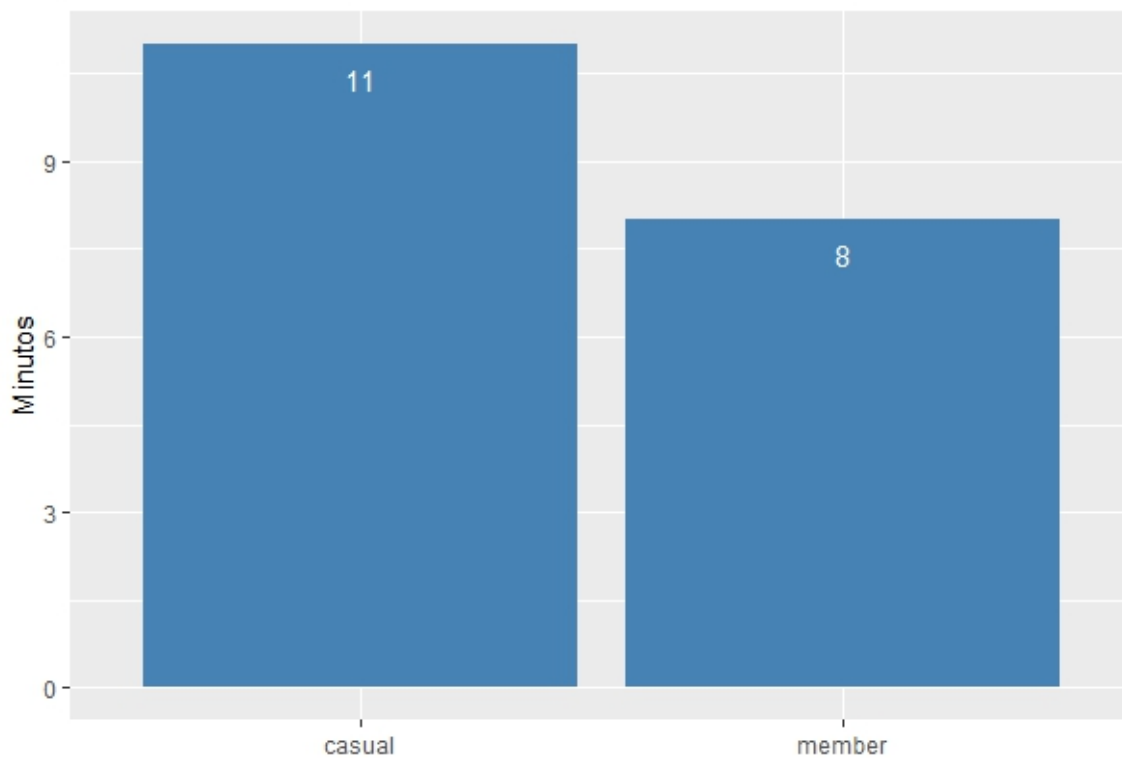
## Dia da semana x Tipo de Bicicleta

nov.2022 - out.2023



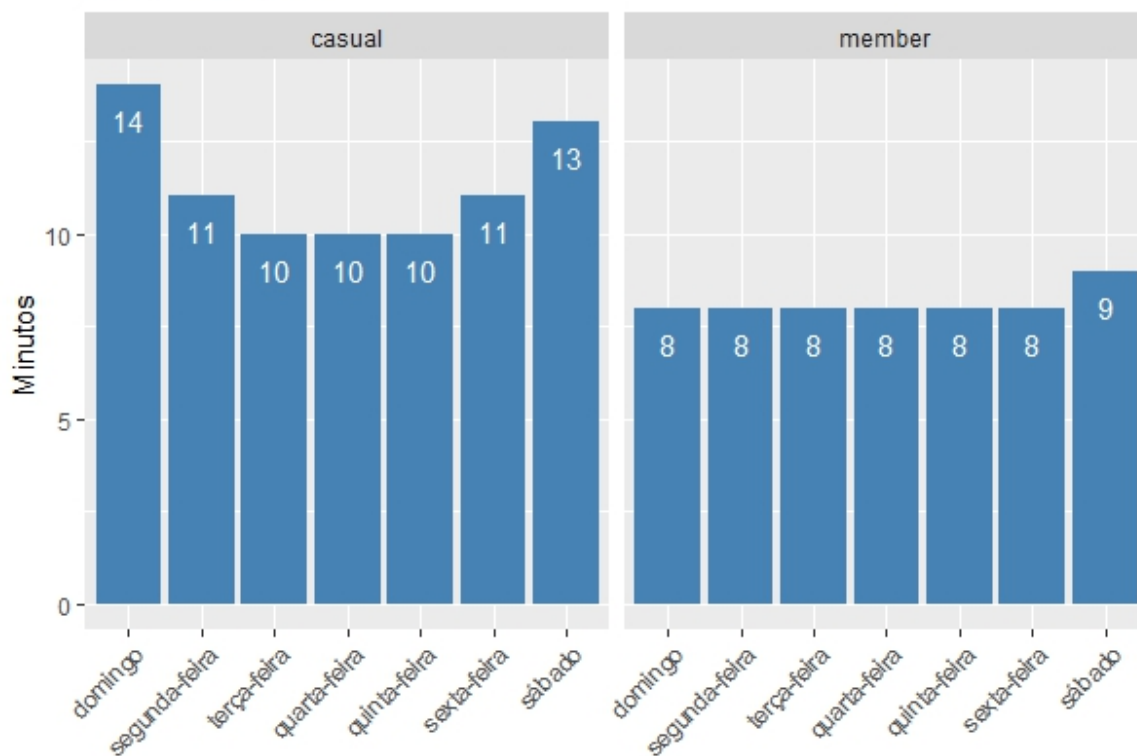
## Medianas : Duração do Passeio x Tipo de Usuário

nov.2022 - out.2023



## Medianas : Duração do Passeio x Dia da Semana

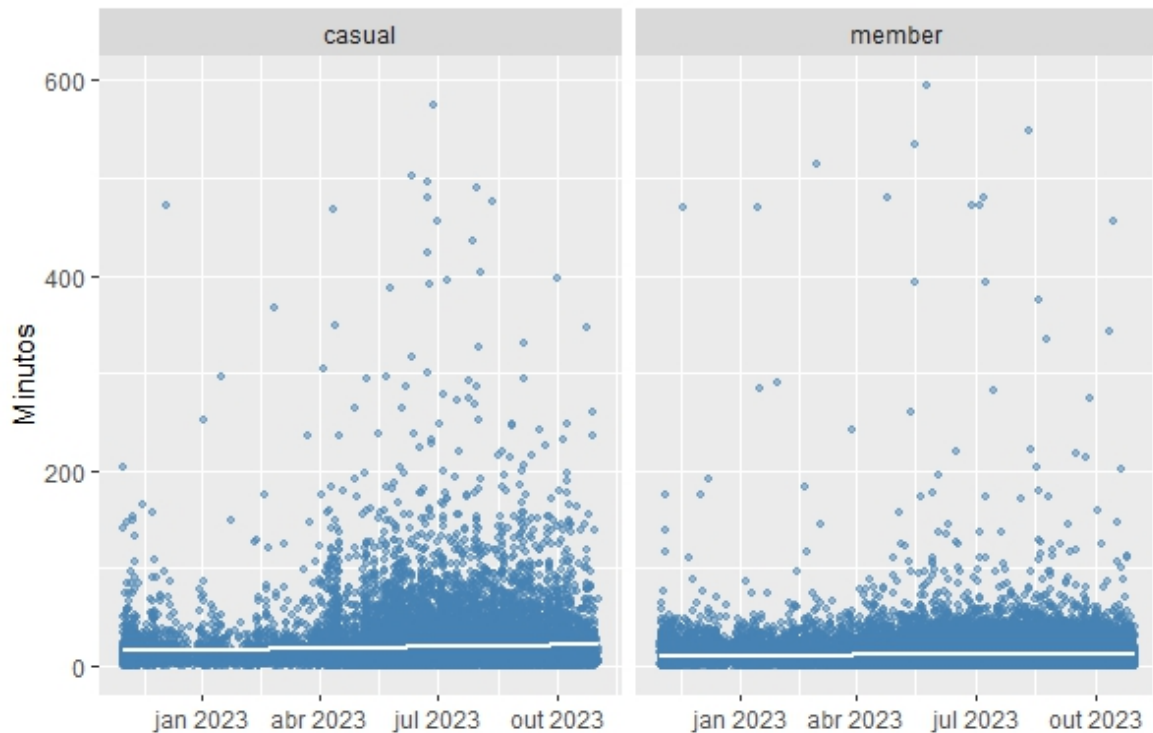
nov.2022 - out.2023





### Tempo : Duração x Início do Passeio

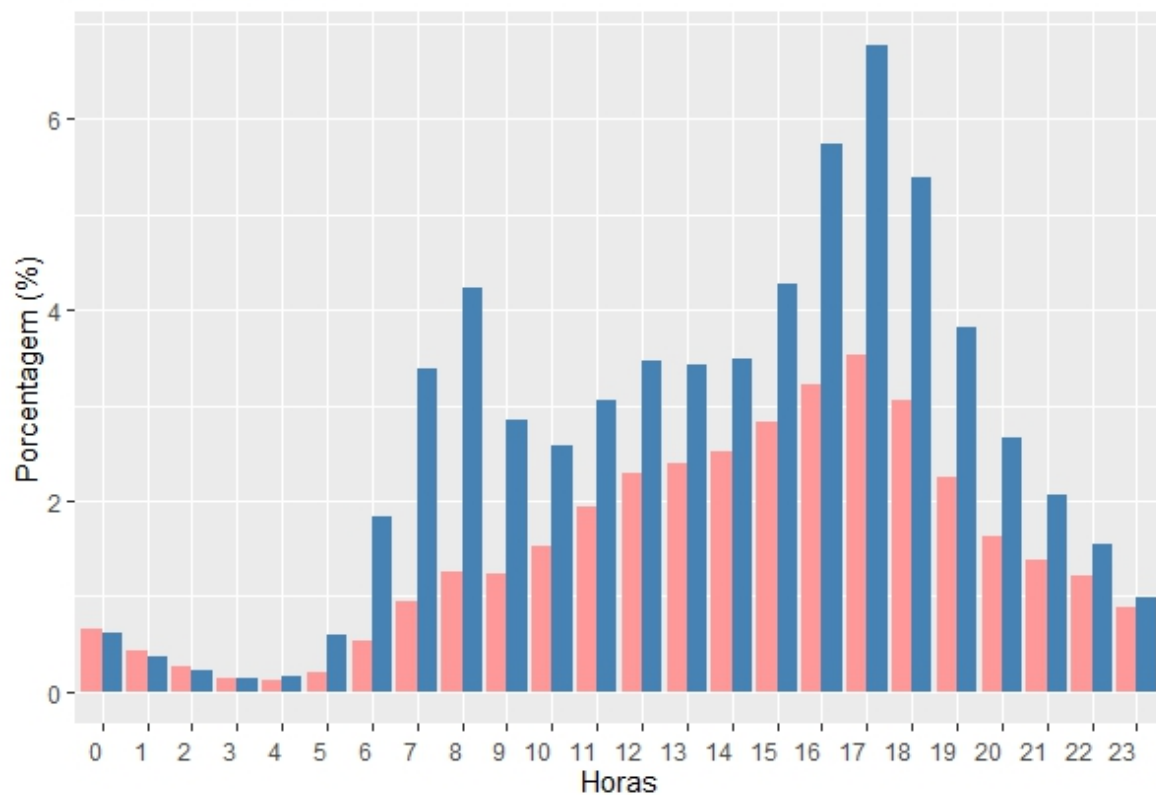
nov.2022 - out.2023



dados extraídos de 56.434 registros (~1%).

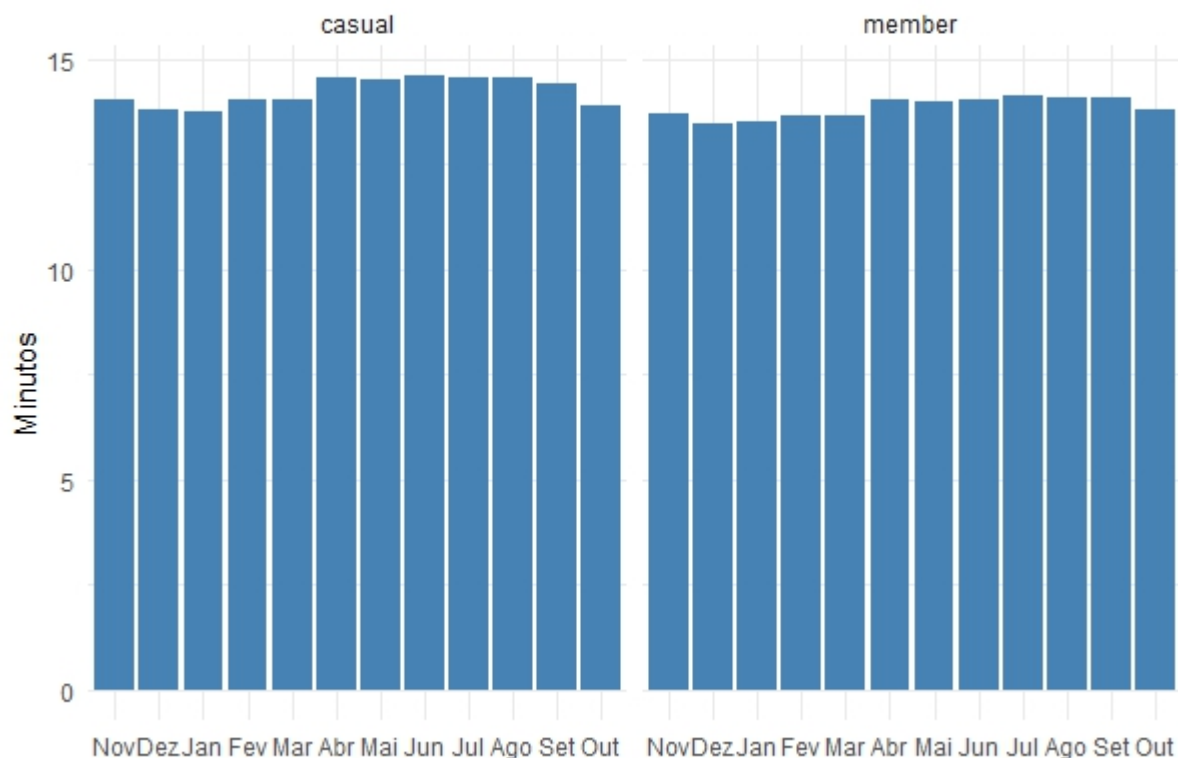
### Horário início dos passeios : Casuais x Membros

nov.2022 - out.2023



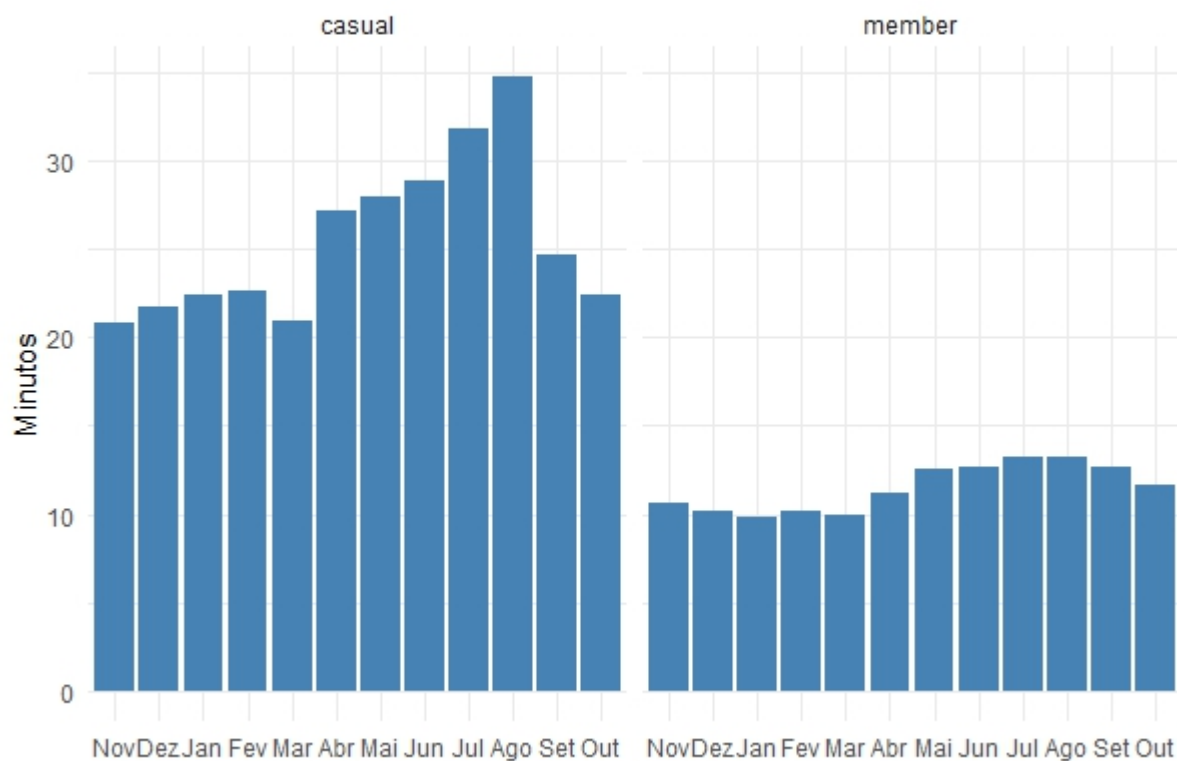
## Horário Médio de Saída por usuário

nov.2022 - out.2023



## Tempo de uso médio por usuário

nov.2022 - out.2023



## 6. Recomendações

Os dados da tabela contam que a empresa Cyclistic está em uma tendência de crescimento de demanda e tempo de passeio nos últimos 12 meses.

A maior parte dos usuários são do tipo MEMBROS anuais, e usam as bicicletas provavelmente para trabalho (entre seg-sex, com mais saídas pela manhã e pelo fim da tarde).

Por outro lado, o perfil dos CASUAIS se concentra aos fins de semana, com saídas mais longas, no fim da tarde, e com maior uso das bicicletas elétricas.

Em comum, ambos os perfis estão usando mais as bicicletas para passeios de modo relativamente uniforme por Chicago.

### a) Sugestão principal: Passeios mais longos e aos fins de semana

A intenção inicial do projeto era convencer ciclistas CASUAIS a se tornarem MEMBROS anuais pode ser colocada em prática divulgando **promoções diferenciadas e focalizadas** para tornar MEMBROS perfis que correspondam aos ciclistas CASUAIS.

Por exemplo, com promoções para aqueles que **mais passeiam aos fins de semana**, ou para ciclistas que registram **passeios mais longos**.

Outras hipóteses, como promoção para determinado tipo de bicicleta, passeios em determinados horários, saídas de determinada estação, ou quantidade mínima de passeios por semana, não apresentaram dados suficientes para levantar hipóteses concretas.

Outras estratégias de marketing podem continuar, uma vez que há alguma tendência geral de maior tempo de uso das bicicletas ao longo dos últimos meses.

### b) Sugestões secundárias:

Quanto ao registro do banco de dados, nossa sugestão é dar condições de aprofundamento da análise por meio de:

- disponibilização de dados individuais dos ciclistas, inclusive categorizando-os no mínimo em três categorias, como "viagens únicas", "diárias" e "membros anuais" (para estabelecer um perfil mais adequado do clientes, além de sua evolução ao longo do tempo);
- reclassificação dos registros de passeios muito "curtos", inferiores a determinado período, e que não constituem passeios de fato;
- rever problemas com os registros das estações de saída e chegada (1/4 dos registros estão faltando esses dados);

- revisão das razões que levam a registros de período de tempo de passeios "negativos", que exigem exclusão na hora de analisá-los;
- definição dos tipos de bicicleta, de acordo com a demanda original da empresa (bicicletas reclináveis, triciclos manuais, bicicletas de carga, clássica, elétrica etc.);
- padronização dos IDs das estações.

## 7. Códigos para Revisão

### ### DADOS ###

```
> glimpse(tripdata_total2)
Rows: 5,652,827
Columns: 17
$ ride_id          <chr> "BCC66FC6FAB27CC7", "772AB67E902C180F", "585EAD07FDEC0152", "91C4E7ED3C262FF..."
$ rideable_type    <chr> "electric_bike", "classic_bike", "classic_bike", "classic_bike", "classic_b..."
$ started_at       <dtm> 2022-11-10 06:21:55, 2022-11-04 07:31:55, 2022-11-21 17:20:29, 2022-11-25 1...
$ ended_at         <dtm> 2022-11-10 06:31:27, 2022-11-04 07:46:25, 2022-11-21 17:34:36, 2022-11-25 1...
$ start_station_name <chr> "Canal St & Adams St", "Canal St & Adams St", "Indiana Ave & Roosevelt Rd", ...
$ start_station_id  <chr> "13011", "13011", "SL-005", "SL-005", "SL-005", "13022", "331", "331", "TA13..."
$ end_station_name  <chr> "St. Clair St & Erie St", "St. Clair St & Erie St", "St. Clair St & Erie St"...
$ end_station_id    <chr> "13016", "13016", "13016", "13016", "13016", "TA1306000003", "TA1306000003",...
$ start_lat         <dbl> 41.87940, 41.87926, 41.86789, 41.86789, 41.86789, 41.89228, 41.90967, 41.909...
$ start_lng         <dbl> -87.63985, -87.63990, -87.62304, -87.62304, -87.62304, -87.61204, -87.64813,...
$ end_lat           <dbl> 41.89435, 41.89435, 41.89435, 41.89435, 41.89435, 41.88872, 41.88872, 41.898...
$ end_lng           <dbl> -87.62280, -87.62280, -87.62280, -87.62280, -87.62280, -87.64445, -87.64445,...
$ member_casual     <chr> "member", "member", "member", "member", "member", "member", "member", "membe..."
$ duracao_passeio   <int> 9, 14, 14, 15, 18, 11, 9, 8, 9, 6, 7, 13, 8, 5, 7, 9, 5, 10, 14, 4, 4, 4, 6,...
$ dia_semana        <chr> "quinta-feira", "sexta-feira", "segunda-feira", "sexta-feira", "terça-feira"...
$ mes               <dbl> 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, ...
$ horas_start       <int> 6, 7, 17, 17, 17, 14, 14, 20, 21, 20, 20, 0, 17, 15, 15, 22, 14, 16, 6, 16, ...

> glimpse(tripdata_total_reduzido)
Rows: 56,434
Columns: 17
$ ride_id          <chr> "855AAEF13EF626A", "825288580DA7FABD", "054FE726F4EA9129", "823B0A92F1C17BD..."
$ rideable_type    <chr> "classic_bike", "classic_bike", "electric_bike", "electric_bike", "classic_b..."
$ started_at       <dtm> 2023-08-12 13:18:37, 2023-07-11 20:51:42, 2022-12-21 07:34:44, 2023-06-28 1...
$ ended_at         <dtm> 2023-08-12 13:24:06, 2023-07-11 21:31:18, 2022-12-21 07:37:21, 2023-06-28 1...
$ start_station_name <chr> "Michigan Ave & Oak St", "Lakefront Trail & Wilson Ave", "Eckhart Park", "Cl..."
$ start_station_id  <chr> "13042", "639", "13289", "WL-008", "TA1307000159", "SL-005", "TA1306000011",...
$ end_station_name  <chr> "Michigan Ave & Oak St", "McClurg Ct & Ohio St", "Ogden Ave & Chicago Ave", ...
$ end_station_id    <chr> "13042", "TA1306000029", "TA1305000020", "13307", "13164", "TA1306000029", N...
$ start_lat         <dbl> 41.90096, 41.96585, 41.89631, 41.86700, 41.85790, 41.86805, 41.89991, 41.920...
$ start_lng         <dbl> -87.62378, -87.64536, -87.66103, -87.64112, -87.66875, -87.62308, -87.63447,...
$ end_lat           <dbl> 41.90096, 41.89259, 41.89636, 41.85491, 41.85817, 41.89259, 41.92000, 41.912...
$ end_lng           <dbl> -87.62378, -87.61729, -87.65406, -87.66356, -87.65650, -87.61729, -87.63000,...
$ member_casual     <chr> "casual", "member", "member", "casual", "casual", "casual", "member", "casua..."
$ duracao_passeio   <int> 5, 39, 2, 12, 5, 12, 6, 15, 10, 19, 41, 20, 8, 0, 8, 6, 0, 18, 9, 10, 18, 5,...
$ dia_semana        <fct> sábado, terça-feira, quarta-feira, quarta-feira, quinta-feira, terça-feira, ...
$ mes               <dbl> 8, 7, 12, 6, 12, 10, 5, 7, 8, 4, 4, 10, 6, 4, 5, 6, 8, 10, 11, 6, 9, 5, 8, 5...
$ horas_start       <int> 13, 20, 7, 17, 17, 18, 21, 14, 7, 12, 17, 16, 1, 14, 19, 8, 17, 7, 16, 15, 1...
```

### ### INÍCIO DAS ANÁLISES ###

```
#criar "tripdata_total_reduzido" (auxiliar para processamentos mais complexos)
```

```
tripdata_total_reduzido <- tripdata_total2 %>%
  sample_frac(0.01, replace = FALSE)
```

```
#cortar excesso de "duracao_passeio" (passeios menores que 10 horas)
```

```
tripdata_total_reduzido <- tripdata_total_reduzido %>%
  filter(duracao_passeio < 600)
```

```
# Relação entre ESTAÇÕES e CICLISTAS
```

```
frequencias <- tripdata_total2 %>%
  count(start_station_name) %>%
  arrange(desc(n)) %>%
  slice_head(n = 5)
total_registros <- nrow(tripdata_total2)
frequencias <- frequencias %>%
  mutate(porcentagem = (n / total_registros) * 100)
cat("Os 5 registros mais frequentes são:\n")
print(frequencias[, "start_station_name"])
cat("\nPorcentagens em relação ao total de registros:\n")
print(frequencias[, "porcentagem"])
```

### ### VISUALIZAÇÃO EM GRÁFICOS ###

```
# gráfico Pizza : "Total de passeios: 5.652.827"
dados_porcentagem <- tripdata_total2 %>%
  count(member_casual) %>%
  mutate(percentagem = n / sum(n) * 100)
ggplot(dados_porcentagem, aes(x = "", y = n, fill = member_casual)) +
  geom_bar(width = 1, stat = "identity") +
  geom_text(aes(label = paste0(round(percentagem), "%")),
    position = position_stack(vjust = 0.5),
    size = 3, color = "white") +
  coord_polar("y", start = 0) +
  labs(title = "Total de passeios: 5.652.827", subtitle = "nov.2022 - out.2023") +
  scale_fill_discrete(name = "Tipos de Usuários", labels = c("Casuais", "Membros"))+
  theme_void()

# Comparação Facetada em Barras : "Distribuição de usuários x Dia da Semana"
porcentagens <- tripdata_total2 %>%
  count(dia_semana, member_casual) %>%
  group_by(member_casual) %>%
  mutate(percentage = n / sum(n) * 100)
ordem_dias <- c("domingo", "segunda-feira", "terça-feira", "quarta-feira", "quinta-feira",
"sexta-feira", "sábado")
porcentagens$dia_semana <- factor(porcentagens$dia_semana, levels = ordem_dias)
ggplot(porcentagens, aes(x = dia_semana, y = percentage, fill = dia_semana)) +
  geom_bar(stat = "identity", fill="steelblue") +
  geom_text(aes(label = paste0(round(percentage), "%")),
    position = position_stack(vjust = 0.5),
    size = 3, color = "white") +
  labs(title = "Distribuição de usuários x Dia da Semana",
    y = "Porcentagem (%)", x="", subtitle = "nov.2022 - out.2023") +
  facet_wrap(~ member_casual, scales = "free_x") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  coord_cartesian(ylim = c(0, max(porcentagens$percentage))) +
  guides(fill = FALSE)

# Barras : "Modelos de Bicicleta"
porcentagens <- tripdata_total2 %>%
  count(rideable_type, member_casual) %>%
  group_by(member_casual) %>%
  mutate(percentage = n / sum(n) * 100)
porcentagens$dia_semana <- factor(porcentagens$rideable_type)
ggplot(porcentagens, aes(x = rideable_type, y = percentage, fill = rideable_type)) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = paste0(round(percentage), "%")),
    position = position_stack(vjust = 0.5),
    size = 3, color = "white") +
  labs(title = "Modelos de Bicicleta",
    y = "Porcentagem (%)", x="", subtitle = "nov.2022 - out.2023") +
  facet_wrap(~ member_casual, scales = "free_x") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  coord_cartesian(ylim = c(0, max(porcentagens$percentage))) +
  guides(fill = FALSE)+
  scale_x_discrete(labels = c("classic_bike" = "Clássica",
    "docked_bike" = "Indefinido",
    "electric_bike" = "Elétrica"))
```

```
# Dispersão : "Tempo : Duração x Início do Passeio"
ggplot(tripdata_total_reduzido, aes(x = started_at, y = duracao_passeio)) +
  geom_point(size = 1, alpha = 0.5, color="steelblue") +
  geom_smooth(method = "lm", se = FALSE, color="white") +
  labs(title = "Tempo : Duração x Início do Passeio",
       subtitle = "nov.2022 - out.2023", x = "", y = "Minutos",
       caption = "dados extraídos de 56.434 registros (~1%).") +
  facet_wrap(~ member_casual, scales = "free_x") +
  guides(fill = FALSE)

# Barra com mediana : "Medianas : Duração do Passeio x Tipo de Usuário"
mediana_por_categoria <- aggregate(duracao_passeio ~ member_casual, data = tripdata_total2,
FUN = median)
ggplot(mediana_por_categoria, aes(x = member_casual, y = duracao_passeio)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  geom_text(aes(label = duracao_passeio), vjust = 2, size = 4, color = "white") +
  labs(title = "Medianas : Duração do Passeio x Tipo de Usuário",
       subtitle = "nov.2022 - out.2023", x = "", y = "Minutos") +
  guides(fill = FALSE)

# Barra com mediana : "Mediana : Duração do Passeio x Dia da Semana"
tripdata_total_reduzido$dia_semana <- factor(tripdata_total_reduzido$dia_semana, levels =
c("domingo", "segunda-feira", "terça-feira", "quarta-feira", "quinta-feira", "sexta-feira",
"sábado"))
mediana_por_dia_e_tipo <- aggregate(duracao_passeio ~ dia_semana + member_casual, data =
tripdata_total_reduzido, FUN = median)
ggplot(mediana_por_dia_e_tipo, aes(x = dia_semana, y = duracao_passeio)) +
  geom_bar(stat = "identity", position = "dodge", fill = "steelblue") +
  geom_text(aes(label = duracao_passeio), vjust = 2, size = 4, color = "white") +
  labs(title = "Medianas : Duração do Passeio x Dia da Semana",
       subtitle = "nov.2022 - out.2023", x = "", y = "Minutos") +
  facet_wrap(~ member_casual, scales = "free_x", ncol = 2) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  guides(fill = FALSE)

# Barra com mediana : "Mediana : Duração do Passeio x Tipo de Bicicleta"
mediana_por_tipo_e_tipo <- aggregate(duracao_passeio ~ rideable_type + member_casual, data =
tripdata_total2, FUN = median)
ggplot(mediana_por_tipo_e_tipo, aes(x = rideable_type, y = duracao_passeio)) +
  geom_bar(stat = "identity", position = "dodge", fill = "steelblue") +
  geom_text(aes(label = duracao_passeio), vjust = 2, size = 4, color = "white") +
  labs(title = "Mediana : Duração do Passeio x Tipo de Bicicleta",
       subtitle = "nov.2022 - out.2023", x = "", y = "Minutos") +
  facet_wrap(~ member_casual, scales = "free_x", ncol = 2) +
  scale_x_discrete(labels = c("classic_bike" = "Clássica",
                             "docked_bike" = "Indefinido",
                             "electric_bike" = "Elétrica"))+
  guides(fill = FALSE)

# Bolha : "Dia da semana x Tipo de Bicicleta"
bubble_data <- tripdata_total2 %>%
  count(dia_semana, rideable_type, member_casual, name = "count_common") %>%
  arrange(dia_semana, rideable_type)
ordem_dias <- c("domingo", "segunda-feira", "terça-feira", "quarta-feira", "quinta-feira",
"sexta-feira", "sábado")
bubble_data$dia_semana <- factor(bubble_data$dia_semana, levels = ordem_dias)
```

```
ggplot(bubble_data, aes(x = dia_semana, y = rideable_type, size = count_common)) +
  geom_point(alpha = 0.7, color = "steelblue") +
  labs(title = "Dia da semana x Tipo de Bicicleta", subtitle = "nov.2022 - out.2023",
       x = "", y = "Tipo Bicicleta", size = "Qtd. Registros") +
  scale_size_continuous(range = c(1, 15)) +
  facet_wrap(~member_casual) +
  guides(fill = FALSE) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_y_discrete(labels = c("classic_bike" = "Clássica",
                             "docked_bike" = "Indefinido",
                             "electric_bike" = "Elétrica"))

# Barras : "Horário início dos passeios"
totals <- tripdata_total2 %>%
  count(horas_start, member_casual) %>%
  group_by(horas_start)
totals$horas_start <- factor(totals$horas_start, levels =
  unique(totals$horas_start)[order(as.numeric(as.character(unique(totals$horas_start))))])
ggplot(totals, aes(x = horas_start, y = ((n / sum(n)) * 100), fill = member_casual)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Horário início dos passeios : Casuais x Membros",
       subtitle = "nov.2022 - out.2023",
       x = "Horas", y = "Porcentagem (%)", fill = "Usuário") +
  scale_fill_discrete(name = "Tipos de Usuários", labels = c("Casuais", "Membros"))+
  guides(fill = FALSE) +
  theme(axis.text.x = element_text(angle = 0, vjust = 0.5, hjust = 1),
        legend.position = "top") +
  scale_fill_manual(values = c("member" = "steelblue", "casual" = "#FF9999"))

# Barras : "Tempo de uso médio por usuário"
meses <- c("Nov", "Dez", "Jan", "Fev", "Mar", "Abr", "Mai", "Jun", "Jul", "Ago", "Set",
"Out")
dados_agrupados <- tripdata_total2 %>%
  mutate(mes = factor(mes, levels = c(11:12,1:10), labels = meses)) %>%
  group_by(mes, member_casual) %>%
  summarise(media_duracao = mean(duracao_passeio))
ggplot(dados_agrupados, aes(x = mes, y = media_duracao)) +
  geom_bar(stat = "identity", position = "dodge", fill = "steelblue") +
  labs(title = "Tempo de uso médio por usuário", subtitle = "nov.2022 - out.2023",
       x = "", y = "Minutos") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  facet_wrap(~ member_casual) +
  guides(fill = FALSE) +
  scale_x_discrete(labels = meses) +
  theme_minimal()

# Barras : "Horário Médio de Saída por usuário"
meses <- c("Nov", "Dez", "Jan", "Fev", "Mar", "Abr", "Mai", "Jun", "Jul", "Ago", "Set",
"Out")
dados_agrupados <- tripdata_total2 %>%
  mutate(mes = factor(mes, levels = c(11:12,1:10), labels = meses)) %>%
  group_by(mes, member_casual) %>%
  summarise(media_start = mean(horas_start))
ggplot(dados_agrupados, aes(x = mes, y = media_start)) +
  geom_bar(stat = "identity", position = "dodge", fill = "steelblue") +
  labs(title = "Horário Médio de Saída por usuário", subtitle = "nov.2022 - out.2023",
       x = "", y = "Minutos") +
```



```
theme(axis.text.x = element_text(angle = 45, hjust = 1)) +  
facet_wrap(~ member_casual) +  
guides(fill = FALSE) +  
scale_x_discrete(labels = meses) +  
theme_minimal()
```