

Estudo de Caso 1: Cyclistic

Preparação

Nesta etapa efetuei a extração e tive o primeiro contato com os dados, durante esta fase foi realizado o agrupamento dos dados em um único conjunto de dados utilizando script em Python, reunindo os dados de todo o ano de 2023. Este dataset contém mais de 5 milhões de linhas correspondentes a alugueis realizados durante o ano, com informações sobre o

- ID do aluguel
- Tipo de bike
- Data e hora de início da corrida
- Data e hora de fim da corrida
- ID estação de início da corrida
- Nome da estação de início
- ID estação de fim da corrida
- Nome da estação de fim
- Latitude e longitude da estação de inicio e fim do passeio
- Tipo de cliente que realizou a corrida (anual ou casual)

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

Observação: os conjuntos de dados têm um nome diferente porque a Cyclistic é uma empresa fictícia. Para os propósitos deste estudo de caso, os conjuntos de dados são adequados e permitem responder às perguntas de negócios. Os dados foram disponibilizados pela *Motivate International Inc.*

As colunas com as latitudes e longitudes das estações de início e fim de cada aluguel foram dispensadas da análise.

Processamento

Etapa de manipulação e limpeza dos dados. A partir dos dados de início e fim dos alugueis foi criada uma coluna com a duração de cada aluguel que norteou grande parte da limpeza dos dados. Além disso, foi realizada uma verificação para encontrar inconsistências no dataset, como:

- ID de aluguel duplicado (Não foi encontrado nenhuma ocorrência)
- Inconsistência na duração de alugueis (Duração negativa, indicando erro nas datas de início e fim da corrida)
- Corridas com tempo de duração desprezíveis (menores ou iguais a 1 minuto)

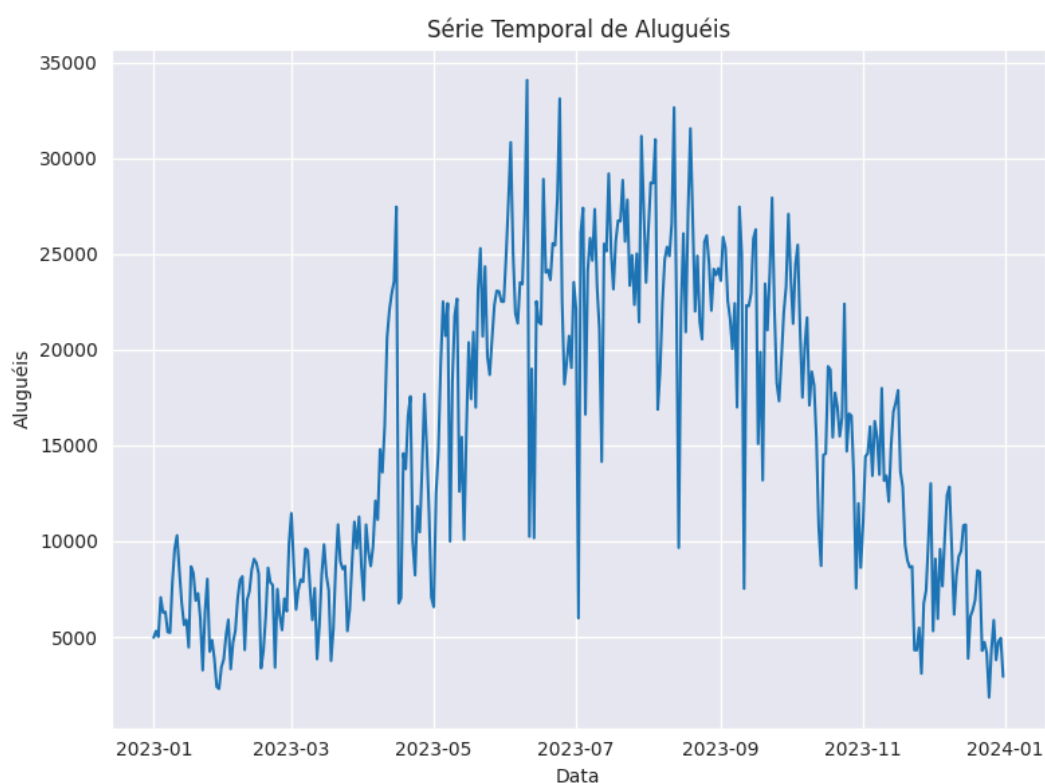
Também foi feito o ordenamento pelas datas de início do aluguel e não pelas IDs (como estava anteriormente)

Ademais, a criação de colunas contendo o número de **dias, horas, minutos e dia da semana** também foi efetuado, visando facilitar o processo de análise.

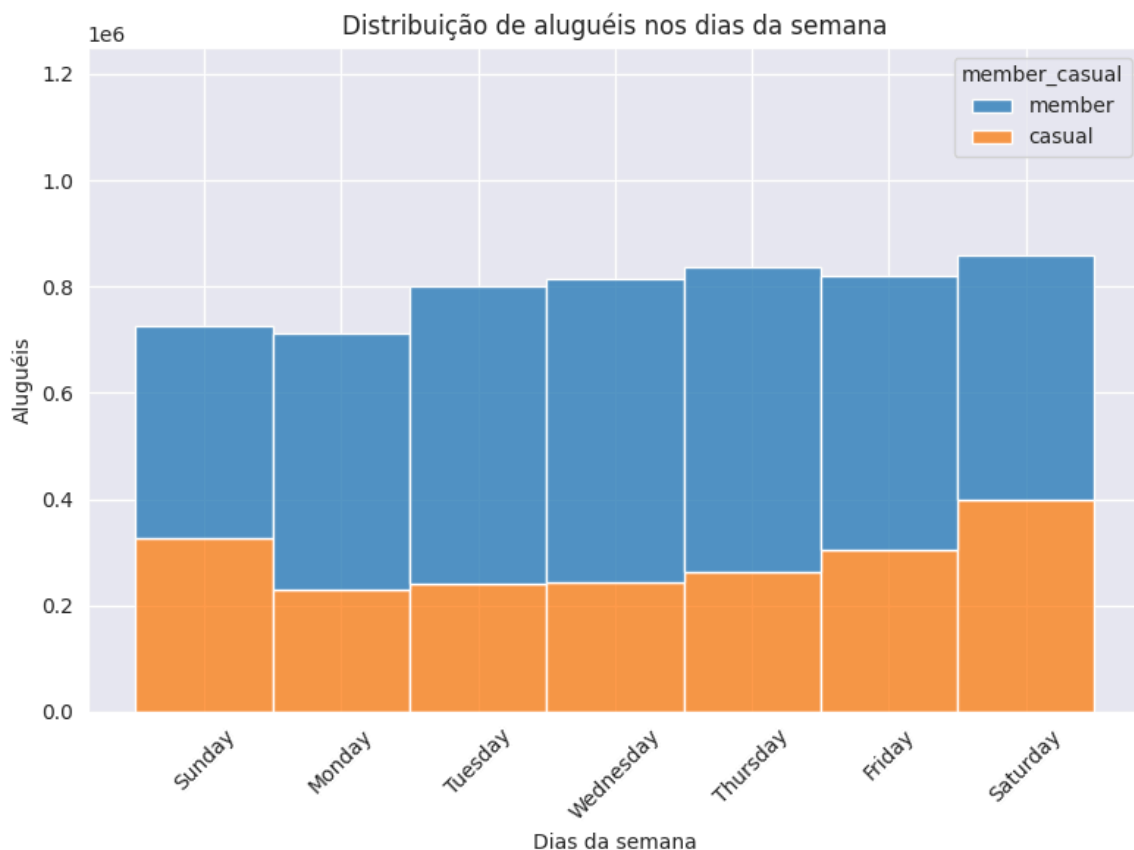
Ao final foi exportado um arquivo .csv contendo os dados limpos e manipulados.

Análise e Visualização dos Dados

O processo de análise dos dados foi orientado pelas perguntas propostas pela equipe de Marketing da *Cyclistic* que gostaria de entender o perfil de uso entre os clientes anuais e casuais.



Ao compararmos o número de alugueis entre os tipos de clientes ao longo dos dias da semana temos a seguinte distribuição:



A partir do histograma é possível observar que clientes casuais tem suas maiores concentrações de alugueis nos finais de semana (Sexta, Sábado e Domingo). Já os membros anuais, também tem uma grande concentração de uso no Sábado, porém tem um perfil de uso maior durante os dias úteis e em compensação têm menores índices aos Domingos.

Considerando o tempo médio de uso das bicicletas entre cada grupo, temos:

Para alugueis menores que 1 dia

minutes	
member_casual	
casual	21.232798
member	12.370344

Já para alugueis maiores que 1 dia

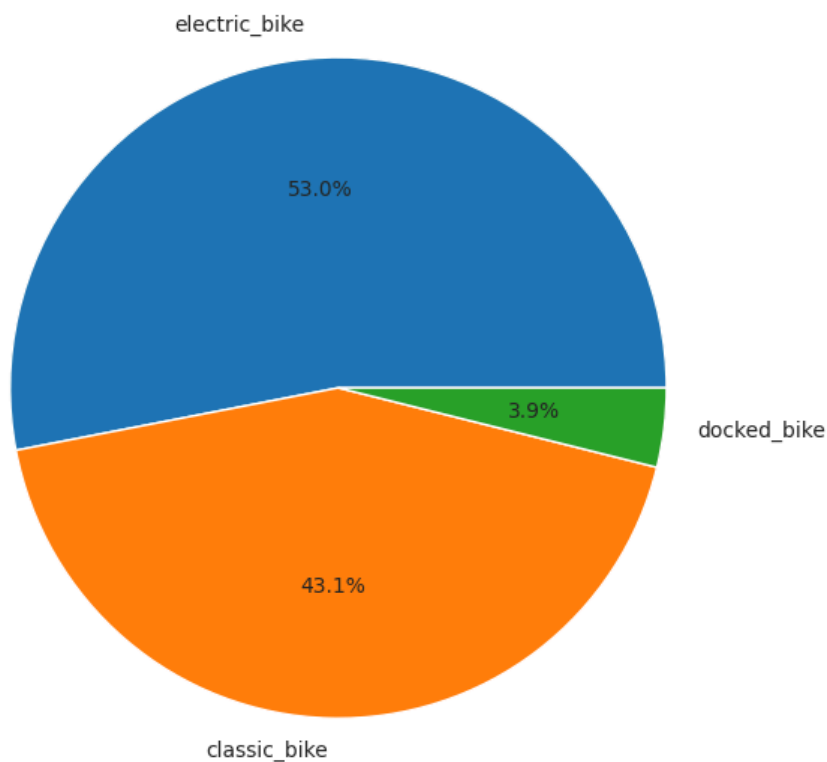
	days	
	mean	median
member_casual		
casual	1.920563	1.0
member	1.000000	1.0

A partir dessas métricas fica evidente que o uso médio entre os usuários comuns é maior se comparado ao uso de usuários anuais. O que pode indicar interesse por essa parcela dos clientes a virarem membros fidelizados.

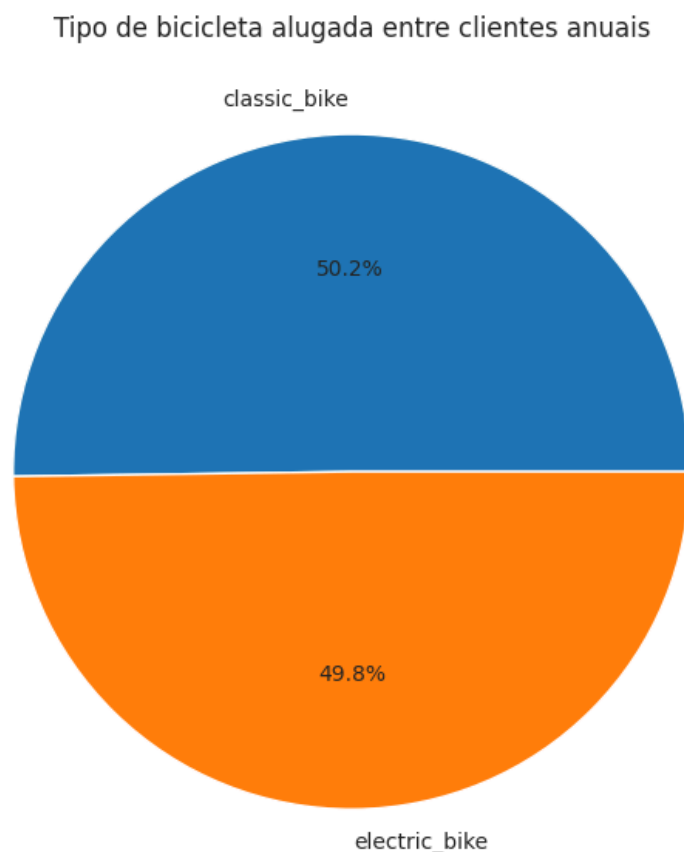
Analizando os tipos de bicicletas

No que se refere ao tipo de bicicleta alugada por consumidores casuais, as bicicletas elétricas aparecem na liderança com 53%, seguidas da bicicleta convencional com 43.1%. Há também uma pequena parcela que utiliza as bicicletas compartilhadas (docked bikes), cerca de 3.9%.

Tipo de bicicleta alugada entre clientes casuais

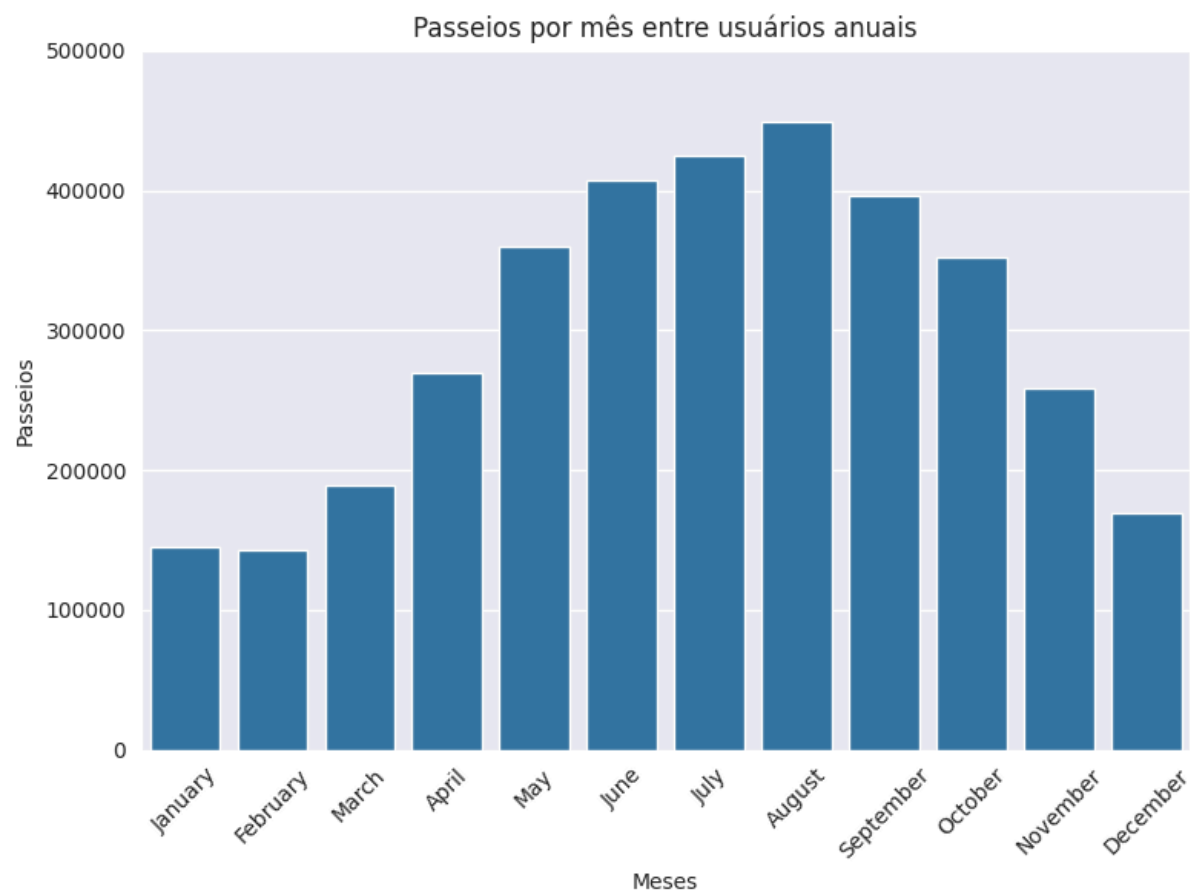
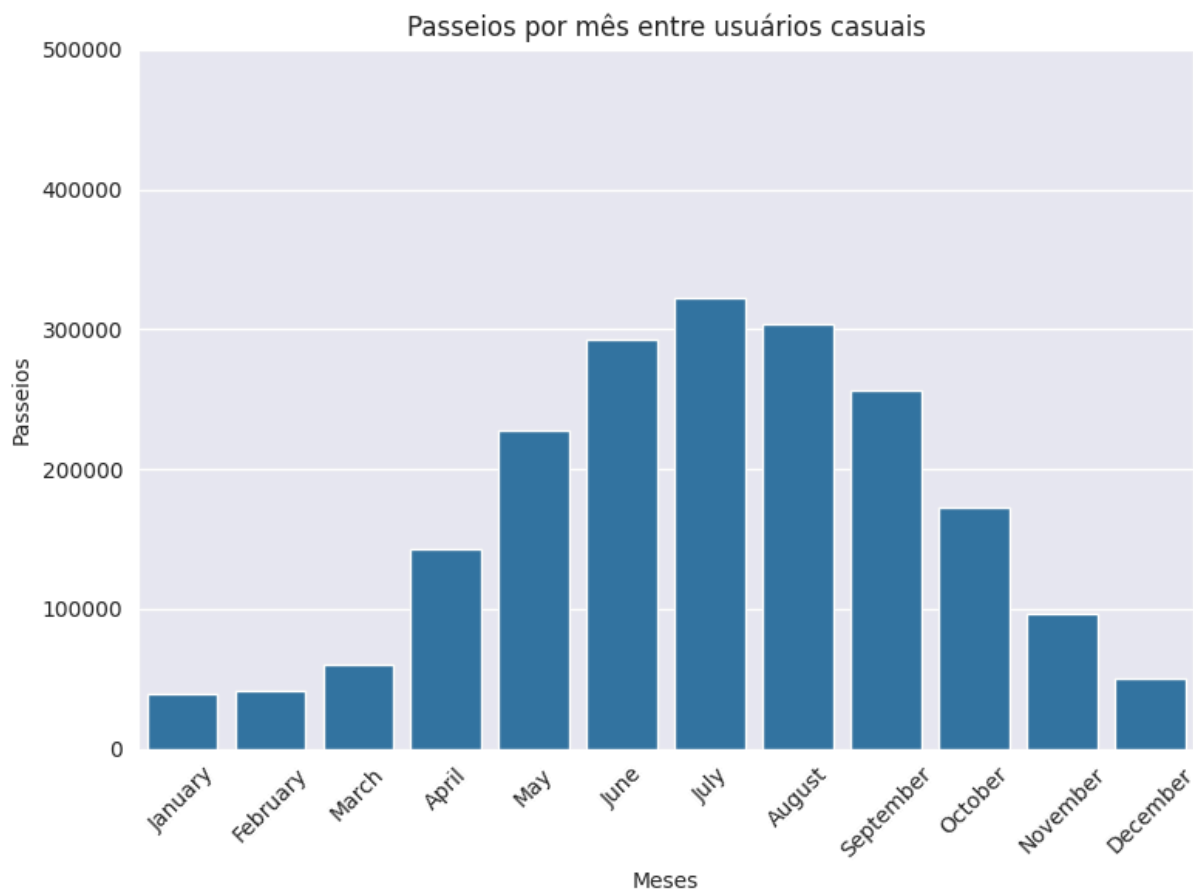


Neste quesito, existe uma diferença em relação aos usuários anuais, pois entre estes clientes a preferência é bem semelhante entre bicicletas clássicas e elétricas, com uma leve vantagem para bicicletas convencionais.



Alugueis entre meses do ano

Olhando para o número de passeios realizados entre os meses do ano, os meses que compreendem a estação do Verão (Hemisfério Norte) são onde ocorrem os maiores índices de alugueis. É interessante notar que os dados mostram uma crescente pré (Primavera) e uma baixa após o Verão (Outono), até chegar no período do Inverno onde ocorre uma baixa procura por passeios de bicicletas entre ambos os grupos.



Conclusão

Buscando aumentar a adesão de clientes anuais, é interessante movimentar campanhas de marketing nos meses que compreendem a Primavera e o Verão, tendo em vista que há uma crescente para ambos os grupos analisados. Também é apropriado buscar o público de usuários casuais que alugam bicicletas em períodos maiores que um dia, pois aparentam ter forte interesse na utilização do sistema da Cyclistic. Por fim, campanhas de marketing próximas ou durante os finais de semanas aparentam ser promissoras para a conversão de novos clientes anuais, tendo em vista que é onde ocorrem os picos de passeios entre usuários casuais.