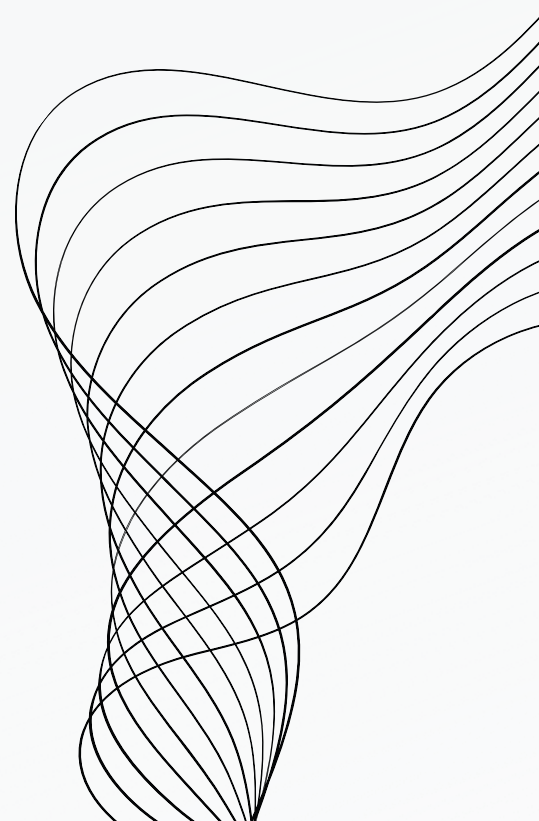


BIKECD
PROJETO



SOBRE



A ciência de dados é um campo em rápido ascensão, com um número crescente de projetos que envolvem a área, em pelo menos alguma das partes. Esse aumento torna necessário o desenvolvimento de metodologias e técnicas que tornem o processo de solução de problemas envolvendo ciência de dados mais produtivo e capaz de gerar mais valor.



A cultura de uso de bikeshare em Washington, D.C., é uma parte significativa do sistema de transporte público da cidade e se tornou uma opção popular para os moradores locais e visitantes explorarem a cidade de maneira sustentável e eficiente.

Sistemas de Bikeshare: Washington, D.C., possui um dos sistemas de bikeshare mais desenvolvidos e bem-sucedidos dos Estados Unidos, conhecido como Capital Bikeshare.

Acessibilidade: O sistema Capital Bikeshare é acessível a uma ampla gama de pessoas. Os moradores locais podem se inscrever em assinaturas mensais ou anuais, enquanto os visitantes podem optar por passes de curto prazo, como passes diários ou de vários dias.

Estações Convenientes: As estações de bicicletas compartilhadas estão estrategicamente localizadas em toda a cidade, incluindo perto de estações de metrô, pontos turísticos, áreas comerciais e bairros residenciais. Isso facilita o acesso às bicicletas e permite que as pessoas as usem como meio de transporte para uma variedade de finalidades, como ir ao trabalho, passear ou fazer compras.

COMPREENSÃO DO NEGÓCIO



O conjunto de dados está relacionado ao compartilhamento de bicicletas. Nosso objetivo é entender como as condições ambientais e sazonais afetam os padrões de aluguel de bicicletas.

OBJETIVO



A mobilidade urbana está sendo transformada por sistemas de compartilhamento de bicicletas, que têm implicações significativas para o tráfego, questões ambientais e de saúde.

CENÁRIO

DADOS



Fontes de Dados: Os dados são do sistema BikeCD e incluem informações meteorológicas extraídas de um site externo.

Dados Disponíveis: Temos dados horários e diários, com variáveis como data, estação do ano, hora do dia, condições climáticas, número de usuários registrados e casuais, entre outras.

COMPREENSÃO DOS DADOS



Os dados já foram agregados em formatos horários e diários. Também foi feita uma extração e adição de informações meteorológicas e sazonais.

Entretanto, é necessário verificar a qualidade dos dados, procurando por valores ausentes, outliers ou erros.

PREPARAÇÃO DOS DADOS



Tarefas Associadas:

Regressão: Usar variáveis como condições climáticas, hora do dia, etc., para prever a contagem de aluguel de bicicletas.

Deteção de Eventos e Anomalias: Usar os dados para validar algoritmos de detecção de anomalias ou eventos.

MODELAGEM

VALIDAÇÃO E USO DE MODELOS



A eficácia dos modelos de previsão pode ser avaliada usando métricas como RMSE ou MAE para a tarefa de regressão. Para detecção de anomalias, a eficácia pode ser avaliada com base na precisão com que eventos significativos são identificados e correlacionados com o aluguel de bicicletas.

AVALIAÇÃO



Uma vez validados, os modelos podem ser usados para prever padrões de aluguel futuros ou para detectar automaticamente eventos significativos com base na contagem de aluguel de bicicletas.

IMPLANTAÇÃO

DICIONÁRIO DE DADOS

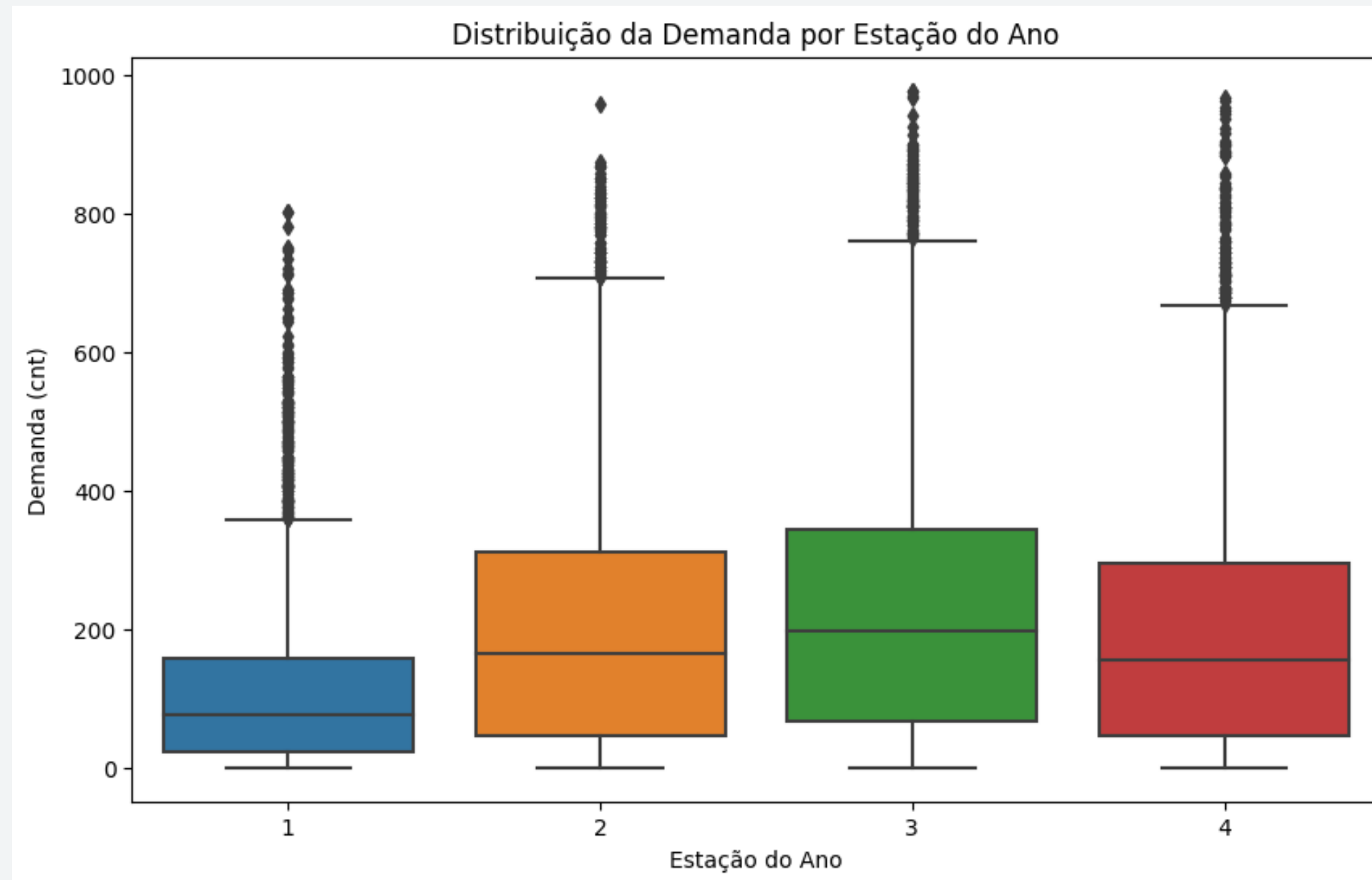
	Variável	Significado	Tipo	Subtipo
0	instant	índice do registro	Quantitativa	Discreta
1	dteday	data	Qualitativa	Ordinal
2	season	estação	Qualitativa	Ordinal
3	yr	ano	Qualitativa	Ordinal
4	mnth	mês	Qualitativa	Ordinal
5	hr	hora	Qualitativa	Ordinal
6	holiday	se o dia é feriado ou não	Qualitativa	Nominal
7	weekday	dia da semana	Qualitativa	Ordinal
8	workingday	se o dia não é fim de semana nem feriado	Qualitativa	Nominal
9	weathersit	situação climática	Qualitativa	Ordinal
10	temp	temperatura normalizada em Celsius	Quantitativa	Contínua
11	atemp	sensação térmica normalizada em Celsius	Quantitativa	Contínua
12	hum	umidade normalizada	Quantitativa	Contínua
13	windspeed	velocidade do vento normalizada	Quantitativa	Contínua
14	casual	contagem de usuários ocasionais	Quantitativa	Discreta
15	registered	contagem de usuários registrados	Quantitativa	Discreta
16	cnt	contagem total de bicicletas alugadas	Quantitativa	Discreta

Não há nenhum valor nulo encontrado no conjunto de dados.

PERGUNTAS E HIPÓTESES

1. A demanda por bicicletas é influenciada pela estação do ano?

Hipótese: O verão tem uma demanda maior devido ao clima ser mais propício.



1-Primavera 2-Verão 3-Outono 4-Inverno

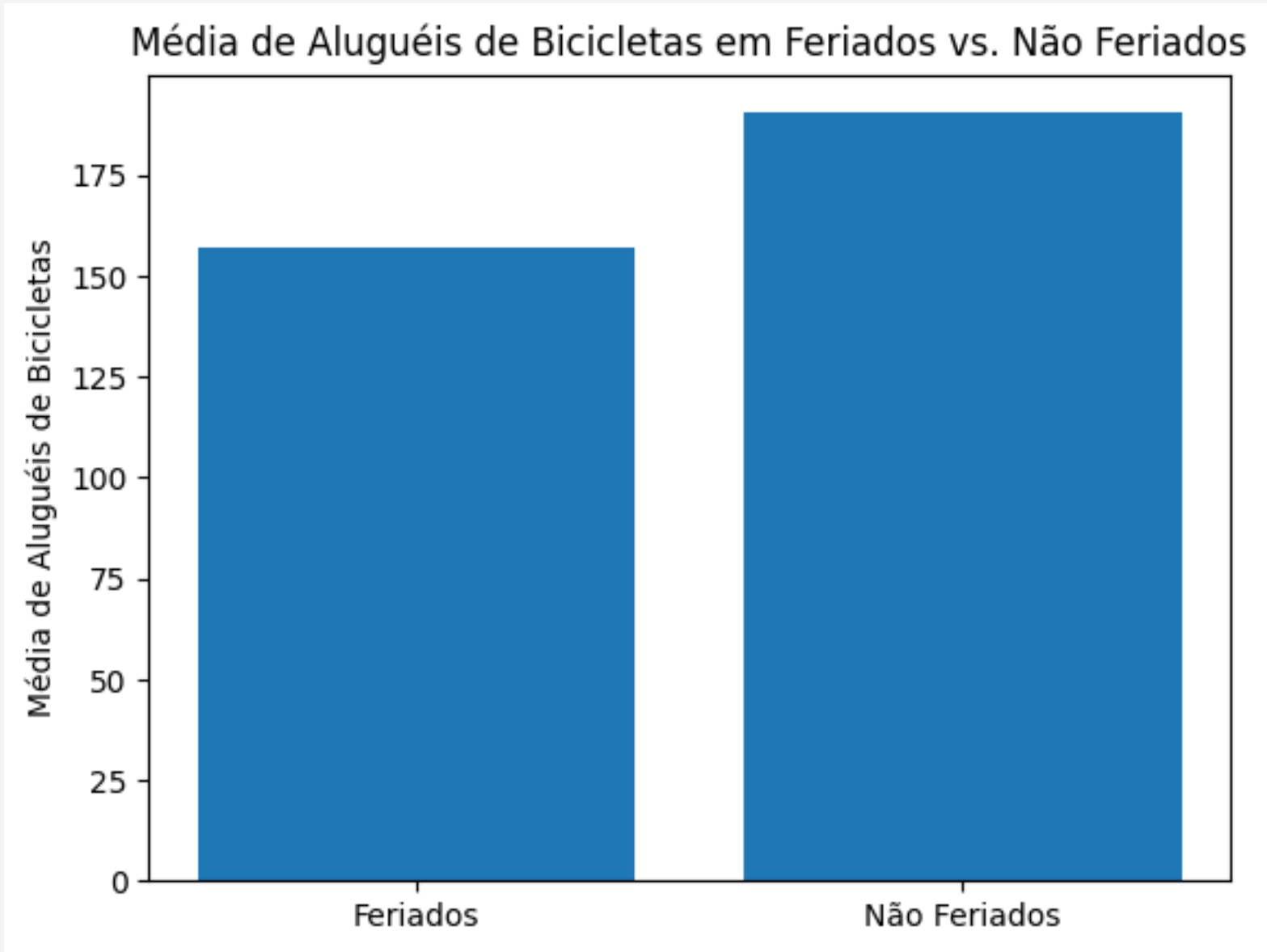
As variações na estação do ano têm um impacto significativo na quantidade de bicicletas alugadas.

A análise estatística realizada para investigar a influência da estação do ano na demanda por bicicletas revelou resultados altamente significativos. O resultado do teste ANOVA (Análise de Variância) apresentou um valor de F-statistic igual a 409.18, com um p-value extremamente baixo de aproximadamente 7.40×10^{-257} .

Esse resultado estatístico indica claramente que há uma influência estatisticamente significativa da estação do ano na demanda por bicicletas. Em outras palavras, as variações na estação do ano têm um impacto significativo na quantidade de bicicletas alugadas. O p-value extremamente baixo sugere que a probabilidade de esses resultados ocorrerem ao acaso é virtualmente nula, fortalecendo ainda mais a conclusão de que a estação do ano desempenha um papel crucial na determinação da demanda.

2. A demanda é menor em feriados?

Hipótese: Em feriados, as pessoas tendem a não trabalhar, reduzindo a necessidade de transporte.



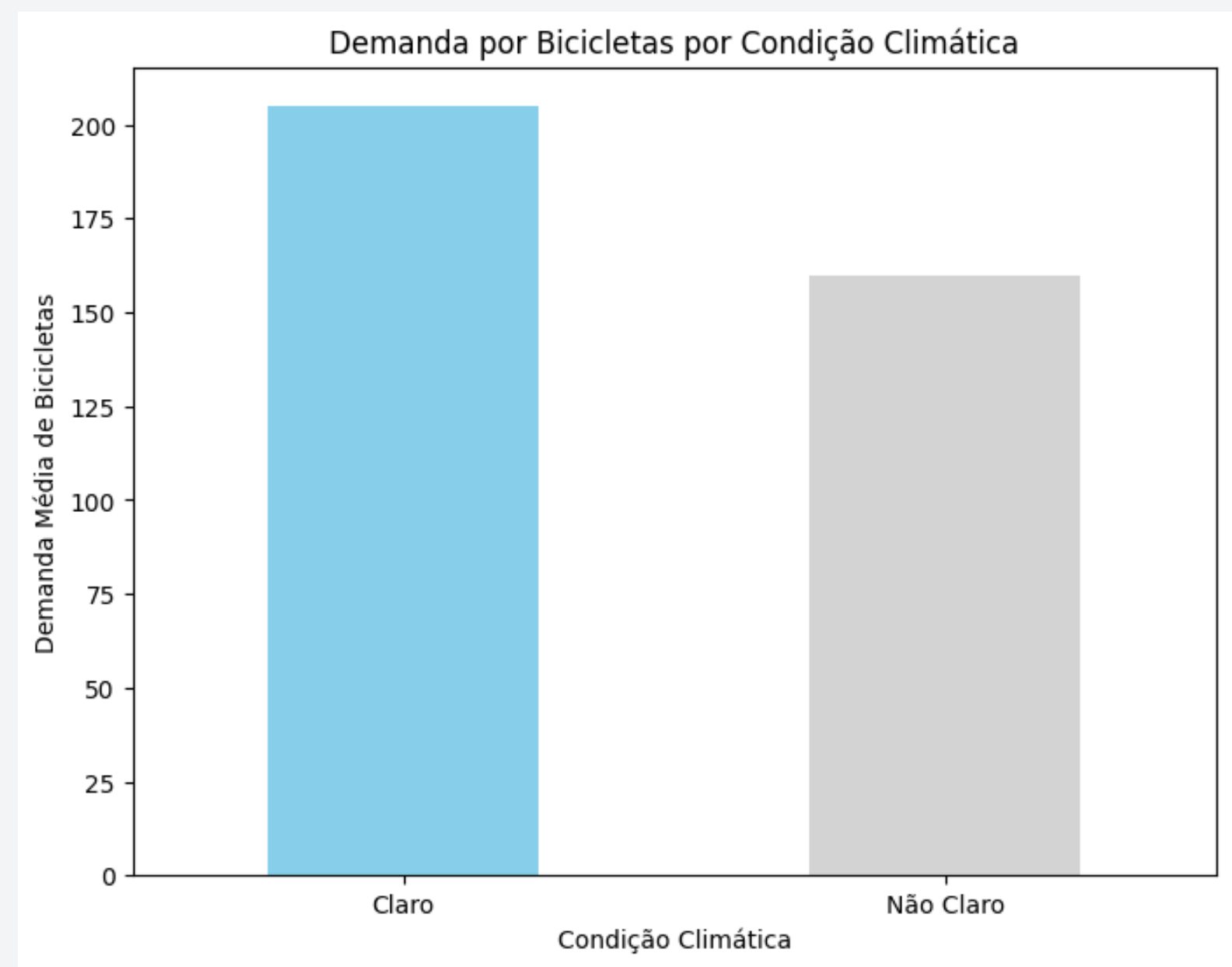
A Demanda por Bicicletas é Menor em Feriados.

Após conduzir uma análise estatística rigorosa, foi observado que a diferença na demanda por bicicletas entre dias de feriado e dias que não são feriados é estatisticamente significativa ($p\text{-valor} = 4.546e-05$). Isso indica que há uma variação notável nas médias de aluguéis de bicicletas entre esses dois grupos de dias.

Os resultados sugerem que, em média, a demanda por bicicletas é menor em feriados. Esta conclusão pode ser atribuída ao fato de que as pessoas tendem a não trabalhar em feriados, reduzindo a necessidade de transporte, o que se reflete nos números de aluguéis de bicicletas. Isso é consistente com a hipótese inicial de que os feriados podem impactar negativamente a demanda por bicicletas devido à menor atividade de trabalho e viagens relacionadas a negócios.

3. A demanda é afetada pelo clima?

Hipótese: Dias claros têm uma demanda maior do que dias chuvosos ou nevados.



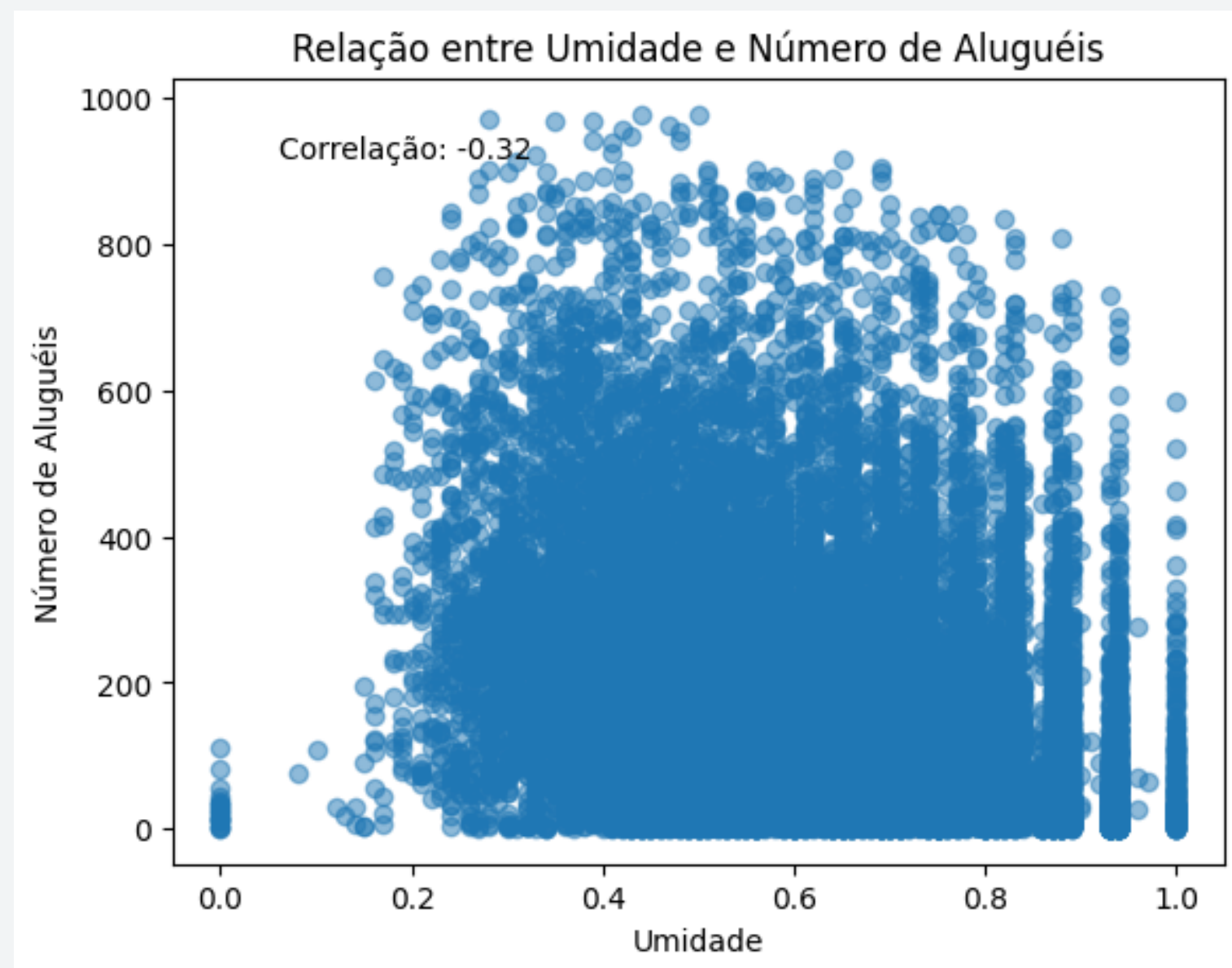
Nossas descobertas demonstram que, de fato, existe uma diferença notável na demanda média de bicicletas em dias com diferentes condições climáticas:

Dias Claros: Durante dias com condições climáticas claras, observamos uma demanda média por aluguel de bicicletas significativamente maior, com uma média de mais de 200 aluguéis. Esse resultado sugere que as pessoas estão mais propensas a optar pelo compartilhamento de bicicletas em dias ensolarados, o que é coerente com a ideia de que o clima agradável incentiva as atividades ao ar livre, como andar de bicicleta.

Dias Não Claros: Por outro lado, em dias não claros, que englobam condições climáticas adversas, como chuva ou neve, observamos uma demanda média um pouco acima de 150 aluguéis. Isso indica que, embora haja uma diminuição na demanda em comparação com dias claros, ainda existe um interesse considerável em alugar bicicletas mesmo em condições climáticas desfavoráveis. Essa demanda pode ser atribuída a necessidades de locomoção ou preferências individuais.

5. Há uma correlação entre a umidade e o número de aluguéis?

Hipótese: Uma umidade muito alta pode desencorajar os aluguéis.



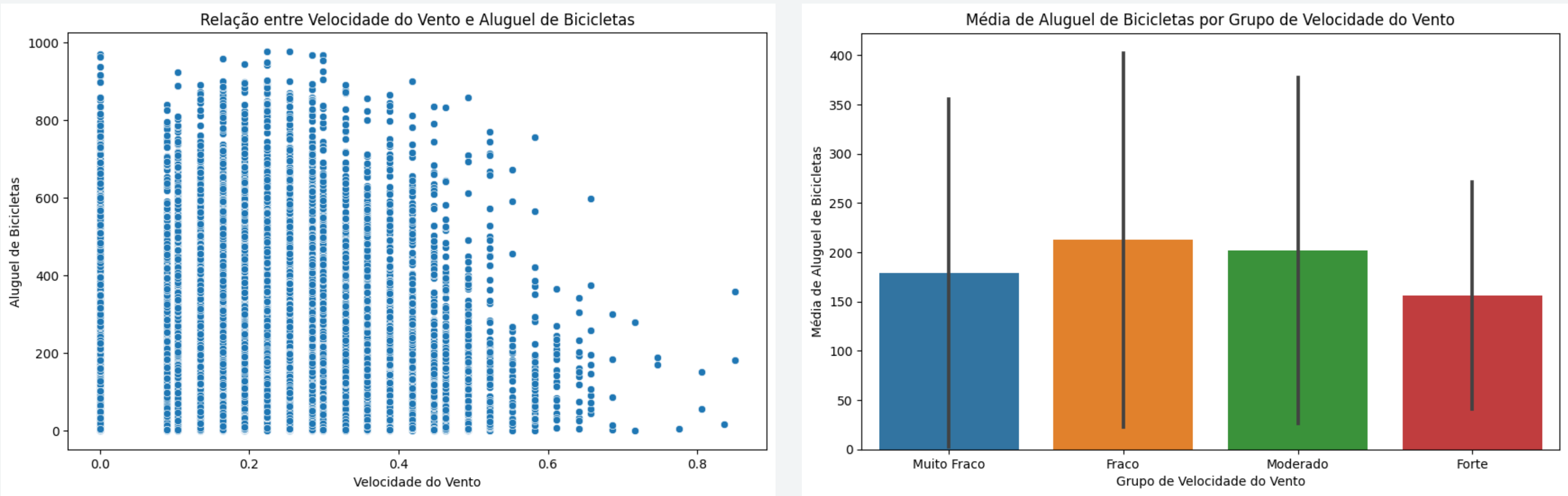
Realizamos uma análise estatística para investigar a possível relação entre a umidade e o número de aluguéis de bicicletas em nosso conjunto de dados. A correlação calculada entre essas duas variáveis foi de -0.32 .

A correlação negativa entre umidade e aluguéis sugere que há uma tendência de diminuição no número de aluguéis de bicicletas em condições de umidade mais alta. Embora a correlação não seja extremamente forte, sua natureza negativa indica que, em geral, um aumento na umidade está associado a uma redução modesta no número de aluguéis de bicicletas.

Isso pode ser interpretado como um indício de que, em dias com umidade muito alta, as pessoas podem estar menos propensas a alugar bicicletas. Existem várias razões possíveis para essa relação, como o desconforto causado pela umidade elevada ou preocupações com as condições de pilotagem em condições climáticas adversas.

6. A velocidade do vento influencia o aluguel de bicicletas?

Hipótese: Vento muito forte pode reduzir a demanda.



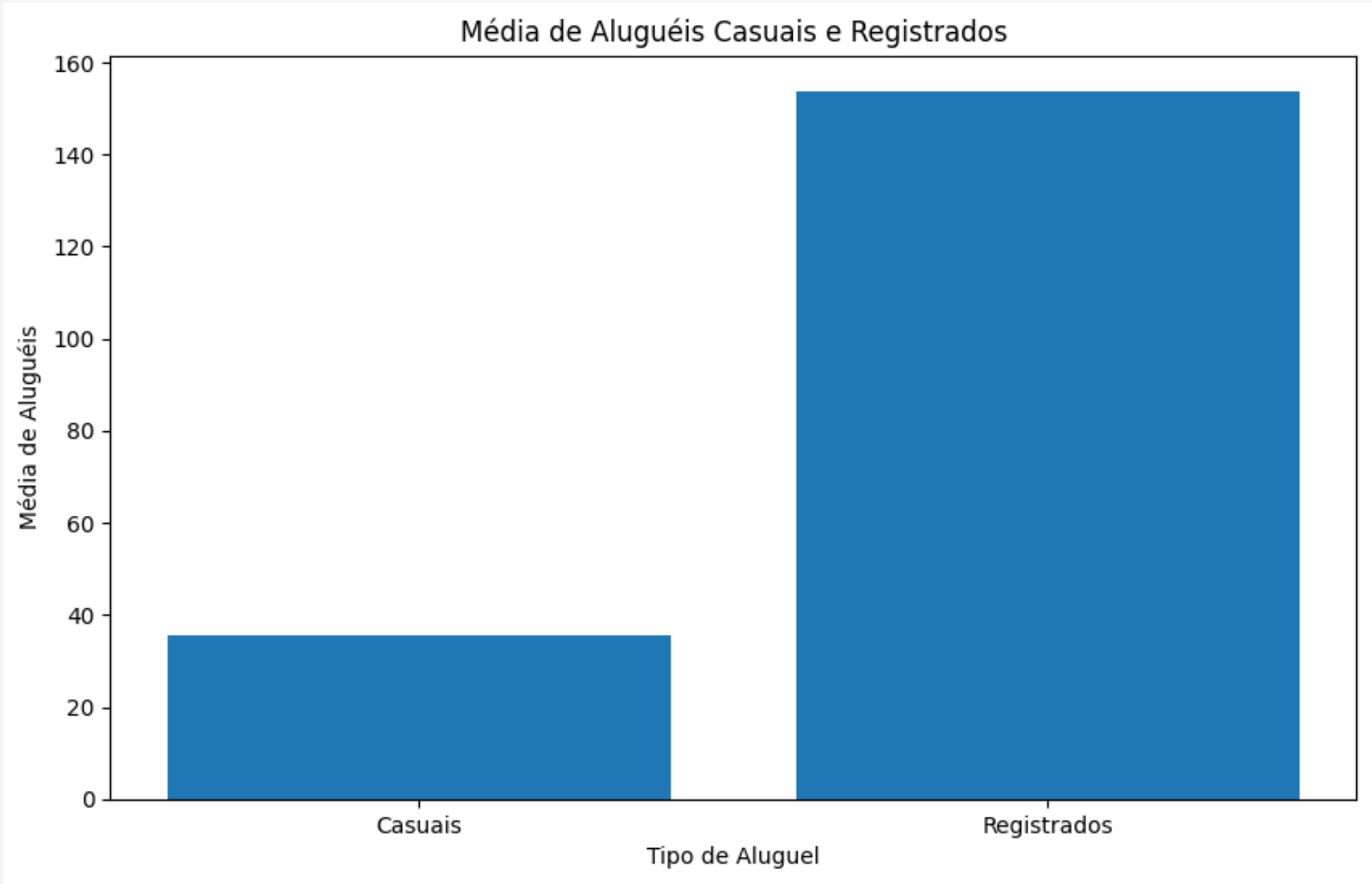
Nossa análise se concentrou na questão de como a velocidade do vento impacta o aluguel de bicicletas em um conjunto de dados de compartilhamento de bicicletas. Principais insights que podemos extrair a partir dos dados e das análises estatísticas realizadas:

Influência da Velocidade do Vento na Demanda de Bicicletas:

- Quando a velocidade do vento é igual a zero, observamos um aluguel de bicicletas alto. Isso sugere que em condições de vento calmo, as pessoas têm uma maior propensão a alugar bicicletas.
- À medida que a velocidade do vento aumenta além de 0.4, o aluguel de bicicletas diminui de forma drástica. Isso indica que ventos fortes podem reduzir significativamente a demanda por bicicletas.

7. O número de alugueis casuais difere significativamente dos alugueis registrados?

Hipótese: Usuários registrados podem ter um padrão diferente de aluguel em comparação com usuários casuais.

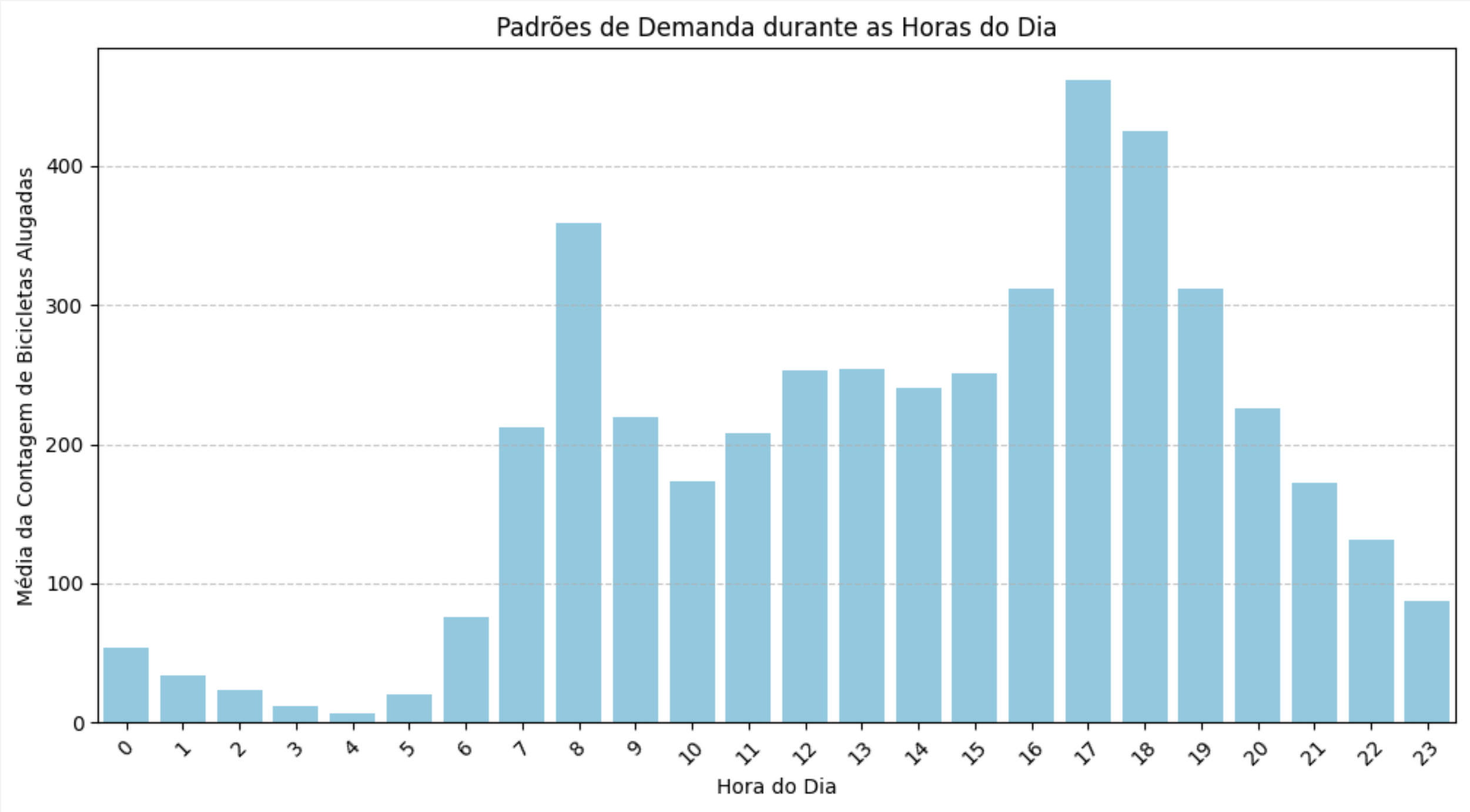


Nossa análise dos dados de aluguel de bicicletas revelou uma diferença estatisticamente significativa entre os alugueis casuais e registrados. Esta descoberta é crucial para entender o comportamento dos usuários e pode ter implicações importantes em nossa estratégia de negócios relacionada ao compartilhamento de bicicletas.

Em média, os alugueis registrados apresentam um número muito maior em comparação com os alugueis casuais. Enquanto a média de alugueis casuais se mantém próxima a 40, os alugueis registrados atingem uma média próxima a 160. Essa disparidade é evidenciada no gráfico de barras, onde observamos claramente que os alugueis registrados superam substancialmente os alugueis casuais.

8. Há padrões de demanda durante as horas do dia?

Hipótese: A demanda é maior durante as horas de pico de ida e volta do trabalho.

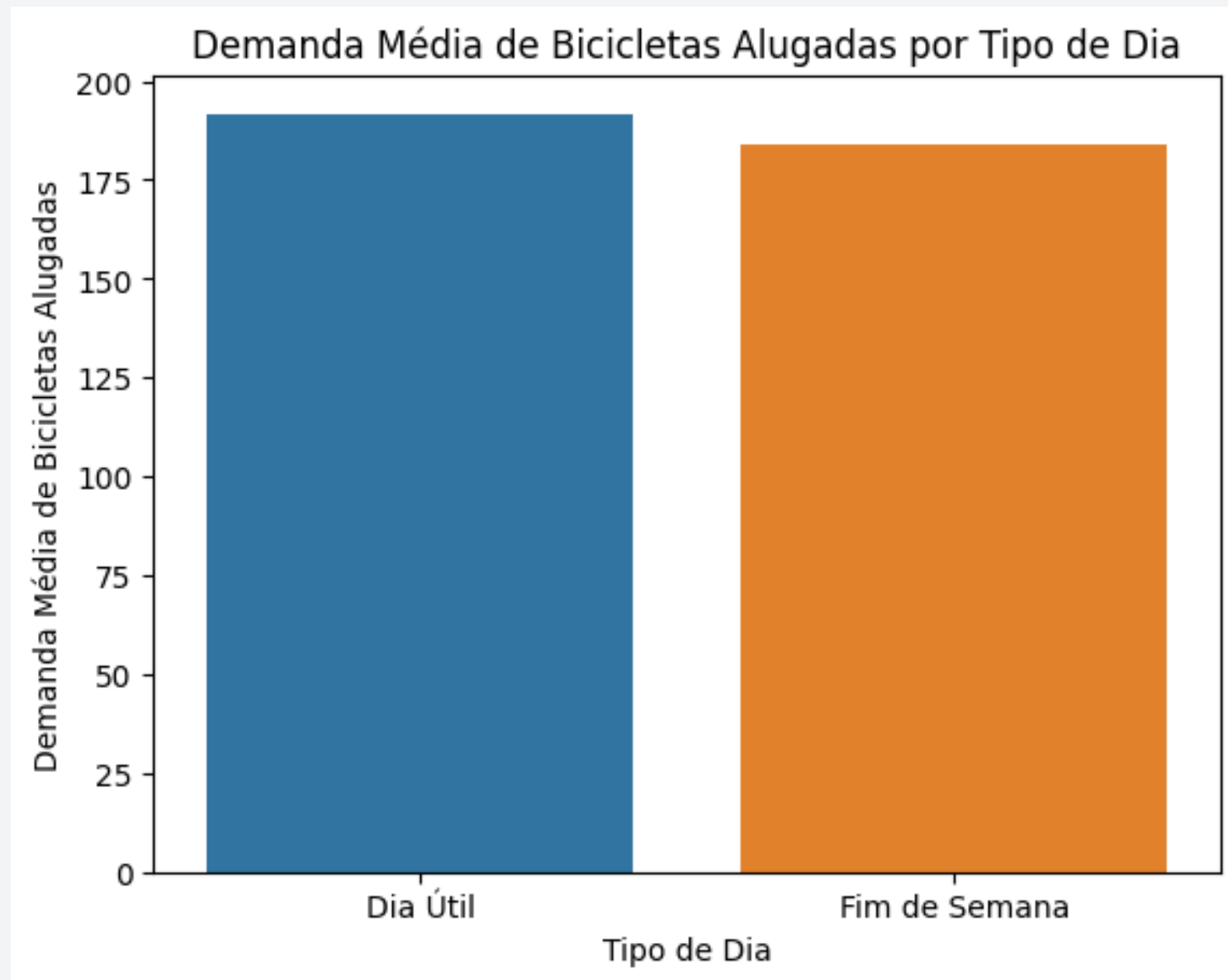


1.Picos nas Horas de Pico: Durante as horas de pico de ida e volta do trabalho, que geralmente ocorrem entre as 7h e as 9h da manhã e entre as 17h e as 19h da tarde, os aluguéis de bicicletas atingem seus níveis mais altos. Este é um achado consistente com a hipótese de que a demanda por bicicletas é maior durante essas horas, conforme as pessoas buscam alternativas de transporte eficientes para evitar o tráfego e chegar ao trabalho.

2. Aumento a partir das 6h: Os aluguéis de bicicletas começam a aumentar significativamente a partir das 6 horas da manhã. Esse aumento sugere que as pessoas estão recorrendo às bicicletas como um meio de transporte para o trabalho, escola ou outras atividades diárias. Este é um indicativo claro de que as bicicletas compartilhadas estão desempenhando um papel importante na mobilidade urbana durante as horas da manhã.

9. Há padrões semanais na demanda de bicicletas?

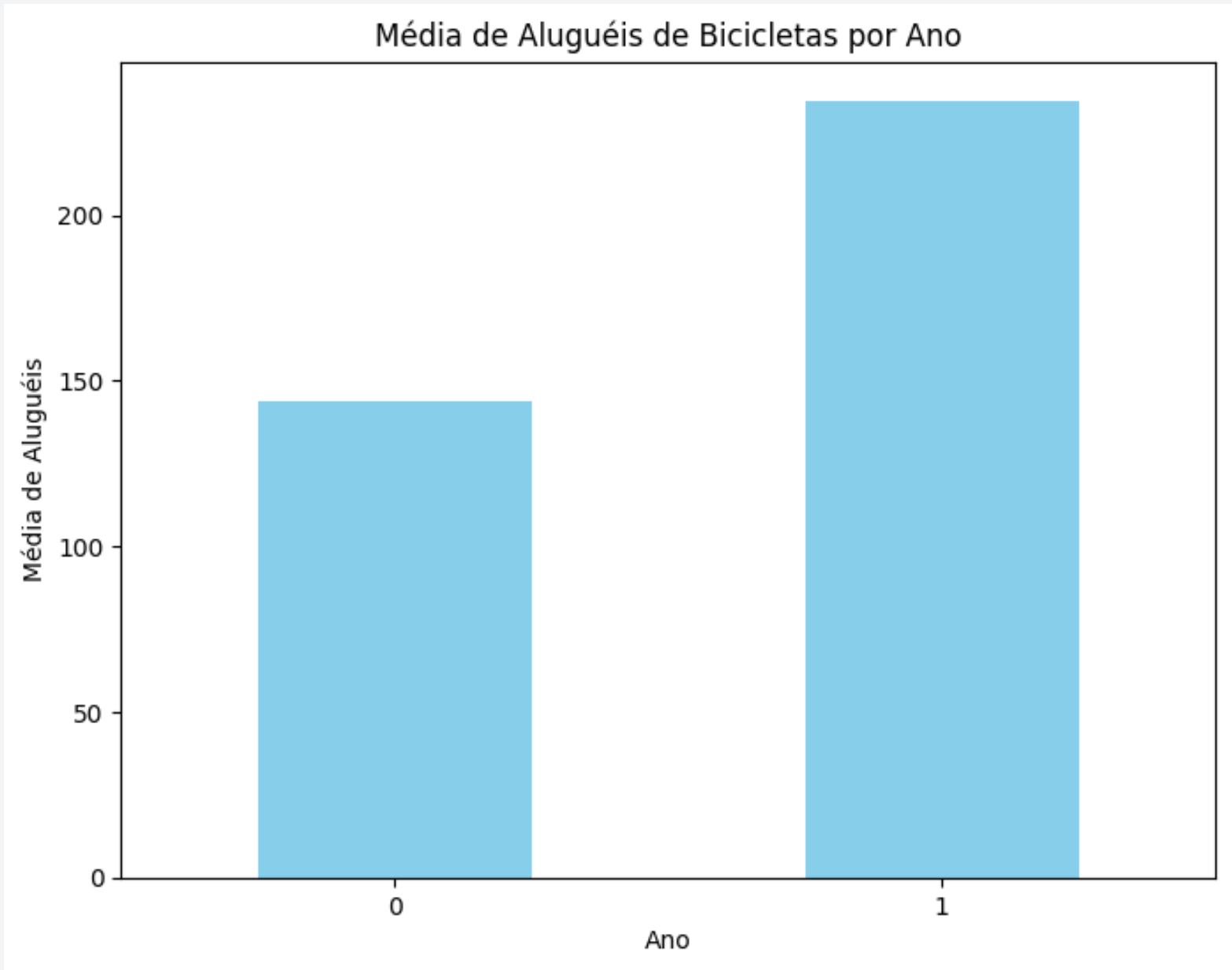
Hipótese: Os finais de semana têm um padrão de demanda diferente dos dias úteis.



Ao analisar a demanda de bicicletas ao longo dos dias da semana, observamos que a demanda média em dias úteis é maior em comparação com os finais de semana. No entanto, é importante notar que a diferença entre os dois tipos de dias não é muito significativa.

10. Aumento na Demanda por Bicicletas de Compartilhamento ao Longo dos Anos

Hipótese: Os finais de semana têm um padrão de demanda diferente dos dias úteis.



o examinar os dados, agrupamos as informações por ano e calculamos a média da contagem total de bicicletas alugadas para cada ano. Os resultados foram surpreendentes.

No primeiro ano (Ano 0), a média de aluguéis estava próxima de 150 bicicletas. No entanto, observamos um aumento notável no segundo ano (Ano 1), com a média de aluguéis aproximando-se de 250 bicicletas.

CONCLUSÃO

O conjunto de dados de compartilhamento de bicicletas é rico e tem potencial para fornecer insights valiosos sobre como diferentes fatores influenciam o comportamento de aluguel.

