



Previsão de Procura em Sistemas de Bicicletas Partilhadas com Dados Meteorológicos

Projeto SAD — Sistemas de Apoio à Decisão

Curso de Engenharia Informática

Universidade Autónoma de Lisboa

Docente: Sérgio Ferreira

Alunos:

- Bilal Nassib – 300113389
- Henrique Monteiro – 300113382
- Luis Raminhas – 30011447

Data: 15/06/2025

Conteúdo

1. Introdução	3
2. Metodologia.....	3
2.1 Fontes de dados	3
2.2 Tecnologias usadas	3
3. Recolha e Limpeza de Dados	4
4. Análise Exploratória	4
5. Modelação Preditiva	5
6. Dashboard Interativo	6
Seoul:.....	7
Nova York:	8
7. Conclusão	9

1. Introdução

Este projeto visa construir um sistema preditivo da procura de bicicletas de partilha, utilizando dados meteorológicos reais e históricos. Através de técnicas de ciência de dados, modelação estatística e painéis interativos com R Shiny, foi possível prever a utilização em várias cidades com base em temperatura, humidade e vento.

2. Metodologia

2.1 Fontes de dados

- API OpenWeather (previsão meteorológica)
- Wikipedia (lista de sistemas de bicicletas)
- Dataset de Seoul, Kaggle (histórico de alugueres), pode ser acidido através do link: <https://www.kaggle.com/datasets/saurabhshahane/seoul-bike-sharing-demand-prediction>

2.2 Tecnologias usadas

Linguagem R, Posit Cloud, pacotes Tidyverse, Tidymodels, R Shiny, Leaflet, SQLite. O projeto foi organizado de forma modular e clara, garantindo separação entre os dados, scripts, outputs e dashboards. Abaixo apresentam-se as capturas de ecrã com a organização do repositório no ambiente Posit Cloud:

data_raw/

<input type="checkbox"/>		bike_project.db	776 KB	Jun 13, 2025, 8:11 PM
<input type="checkbox"/>		clean_bike_sharing_systems.csv	56 KB	Jun 13, 2025, 1:35 PM
<input type="checkbox"/>		forecast_london.csv	2.1 KB	Jun 13, 2025, 10:44 AM
<input type="checkbox"/>		forecast_new_york.csv	2.1 KB	Jun 13, 2025, 10:44 AM
<input type="checkbox"/>		forecast_paris.csv	2 KB	Jun 13, 2025, 10:44 AM
<input type="checkbox"/>		forecast_seoul.csv	2 KB	Jun 13, 2025, 10:22 AM
<input type="checkbox"/>		forecast_suzhou.csv	2.1 KB	Jun 13, 2025, 10:44 AM
<input type="checkbox"/>		forecast_todas_cidades_limpas.csv	10.4 KB	Jun 13, 2025, 1:35 PM
<input type="checkbox"/>		raw_bike_sharing_systems.csv	54.9 KB	Jun 12, 2025, 10:16 PM
<input type="checkbox"/>		raw_seoul_bike_sharing.csv	581.5 KB	Jun 13, 2025, 1:54 PM
<input type="checkbox"/>		SeoulBikeData.csv	581.5 KB	Jun 13, 2025, 8:08 PM

scripts/

<input type="checkbox"/>		01_web_scraping_bikes.R	612 B	Jun 12, 2025, 10:38 PM
<input type="checkbox"/>		02_api_previsoes_multicidade.R	1.4 KB	Jun 13, 2025, 10:44 AM
<input type="checkbox"/>		03_data_cleaning.R	2 KB	Jun 13, 2025, 1:33 PM
<input type="checkbox"/>		04_eda_sql.R	1.4 KB	Jun 13, 2025, 8:07 PM
<input type="checkbox"/>		05_eda_sql_queries_avancadas.R	2 KB	Jun 13, 2025, 8:14 PM
<input type="checkbox"/>		06_modelo_tidymodels.R	1.6 KB	Jun 13, 2025, 9:11 PM
<input type="checkbox"/>		07_dashboard_shiny.R	3 KB	Jun 13, 2025, 8:57 PM
<input type="checkbox"/>		08_dashboard_multicidades.R	3.6 KB	Jun 13, 2025, 9:03 PM
<input type="checkbox"/>		zz_debug_json_output.R	585 B	Jun 13, 2025, 10:40 AM

3. Recolha e Limpeza de Dados

Os dados meteorológicos e de alugueres foram recolhidos e organizados por cidade, sendo guardados em ficheiros `.csv` individuais. Durante o processo de importação, foram detetadas inconsistências nos nomes das colunas — como símbolos (°, %, /), espaços e caracteres não compatíveis com SQL — que causavam erros ao tentar gravar os dados em SQLite.

Para resolver este problema de forma automática e consistente, foi utilizado o pacote `janitor`, em particular a função `clean_names()`. Esta abordagem foi sugerida durante o desenvolvimento com base em boas práticas da comunidade R e permite garantir que todas as colunas tenham nomes uniformes em `snake_case`, evitando problemas na manipulação, análise e exportação dos dados.

4. Análise Exploratória

Foram executadas queries SQL para analisar a distribuição de alugueres por estação, hora e variáveis meteorológicas. Gráficos foram criados com `ggplot2` para visualizar padrões sazonais e diários.

Exemplo de query usada para analisar a temperatura média e o número médio de alugueres por estação do ano:

```
SELECT
  seasons,
  ROUND(AVG(temperature_c), 2) AS temperatura_media,
  ROUND(AVG(rented_bike_count), 1) AS aluguer_medio
FROM seoul_bike
GROUP BY seasons
ORDER BY aluguer_medio DESC;
```

5. Modelação Preditiva

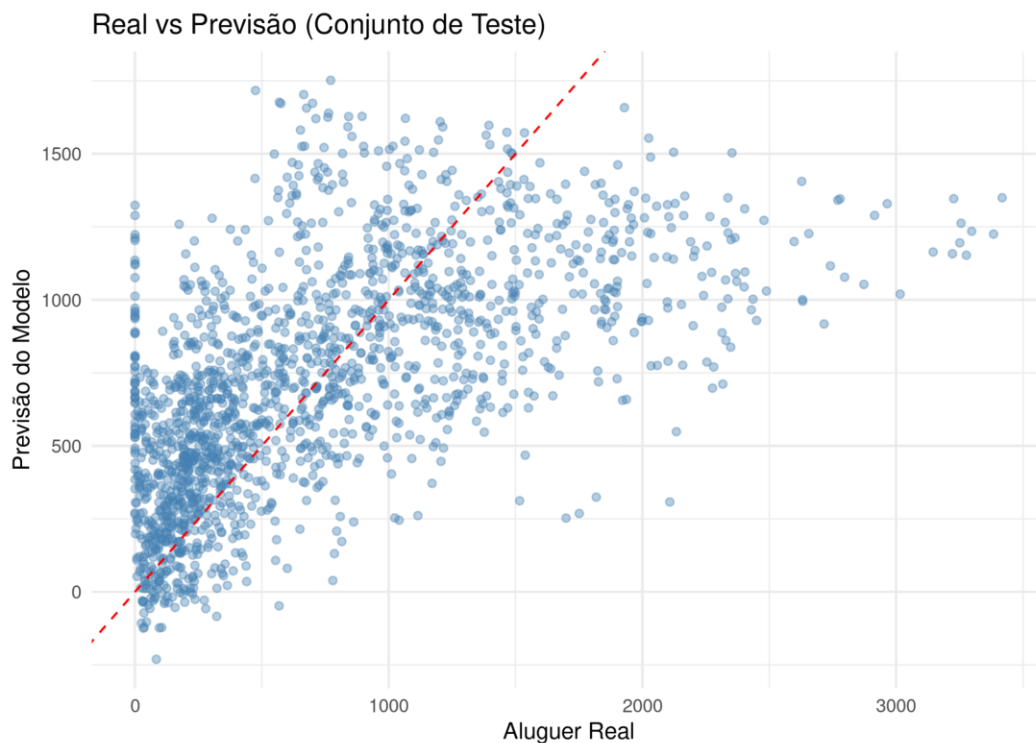
Foi utilizado um modelo de regressão linear com `tidymodels`, treinado com dados históricos de Seoul. O modelo considera temperatura, humidade e vento. A métrica R^2 indica boa capacidade explicativa, e o RMSE quantifica o erro médio nas previsões.

No conjunto de teste, o modelo obteve os seguintes resultados:

- R^2 (coeficiente de determinação): 0.376
- RMSE (Root Mean Squared Error): 507 bicicletas
- MAE (Mean Absolute Error): 371 bicicletas

Estes valores mostram que o modelo consegue captar parcialmente a tendência dos dados, mas ainda apresenta erros consideráveis. Isso indica que variáveis adicionais (como hora do dia, feriados ou sazonalidade) poderiam melhorar significativamente o desempenho do modelo.

Gráfico Real vs Previsão:



O gráfico acima compara os valores reais de alugueres de bicicletas com os valores previstos pelo modelo de regressão linear, aplicados ao conjunto de teste (20% dos dados).

Cada ponto representa uma observação (uma hora específica), onde o eixo X mostra o número real de bicicletas alugadas e o eixo Y mostra a previsão do modelo para essa hora.

A linha vermelha a tracejado representa a linha ideal de previsão perfeita (ou seja, onde previsão = valor real). Quanto mais próximos os pontos estiverem dessa linha, melhor a performance do modelo.

No gráfico obtido, observa-se uma concentração significativa de pontos ao longo da linha, especialmente em gamas de alugueres entre 500 e 1500 bicicletas. No entanto, há também dispersão visível, indicando que o modelo, embora capture bem a tendência geral, apresenta erro significativo em valores extremos.

Este padrão está em linha com os valores de R^2 e RMSE obtidos ($R^2 = 0.376$; RMSE = 507), sugerindo que o modelo explica uma parte relevante da variação, mas que há espaço para melhorias, como introduzir variáveis adicionais (hora, dia da semana, eventos) ou usar modelos mais complexos.

6. Dashboard Interativo

Foram criados dois dashboards interativos com recurso à biblioteca R Shiny, permitindo ao utilizador explorar visualmente a previsão de procura de bicicletas com base nas condições meteorológicas previstas.

- Painel exclusivo para Seoul:

Este primeiro painel apresenta um gráfico dinâmico com a previsão da procura de bicicletas para as próximas 5 dias, com base nos dados meteorológicos extraídos da API do OpenWeather. Inclui também um mapa interativo centrado em Seoul com um marcador de localização, e um slider temporal que permite ao utilizador explorar a evolução das previsões ao longo das horas.

- Painel multacidade (Seoul, Nova York, Paris, Suzhou, Londres):

O segundo painel expande a funcionalidade para suportar múltiplas cidades. O utilizador pode selecionar uma cidade através de um menu suspenso, e o dashboard atualiza automaticamente tanto o mapa com a localização geográfica correspondente, como o gráfico com a previsão de alugueres específica para essa cidade. A previsão é feita com base no mesmo modelo de regressão treinado com dados históricos de Seoul, aplicado aos dados meteorológicos de cada cidade. Isto permite comparar como diferentes condições climáticas em diferentes locais afetam a procura estimada de bicicletas.

Estes dashboards foram desenhados para serem simples, intuitivos e responsivos, utilizando ggplot2 para os gráficos e leaflet para os mapas, proporcionando uma experiência visual clara e informativa.

Seoul:

<https://bda7b0a56992429eab523e4f595083f9.app.posit.cloud/p/e50ed458/>

[Open in Browser](#)

[Publish](#)

Previsão de Aluguer de Bicicletas - Cidades

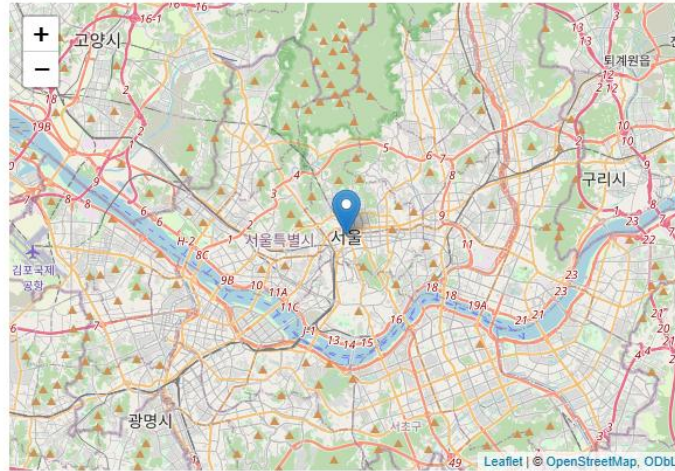
Selecionar cidade:

Seoul

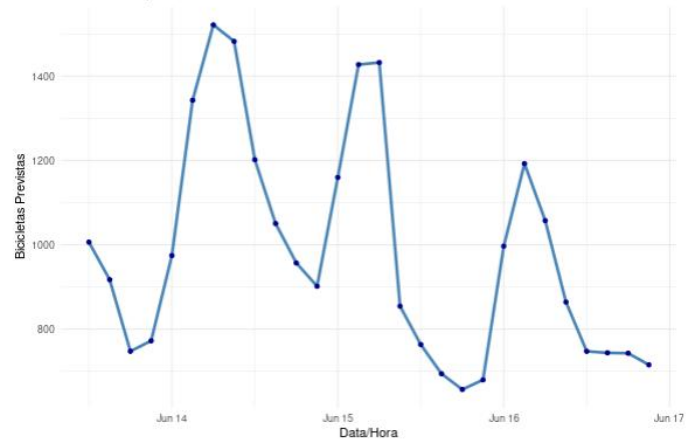
Selecionar intervalo:

13/06 13h 16/06 22h

13/06 13h 14/06 13h 15/06 13h 16/06 13h 17/06 13h



Previsão da procura - Seoul



Nova York:

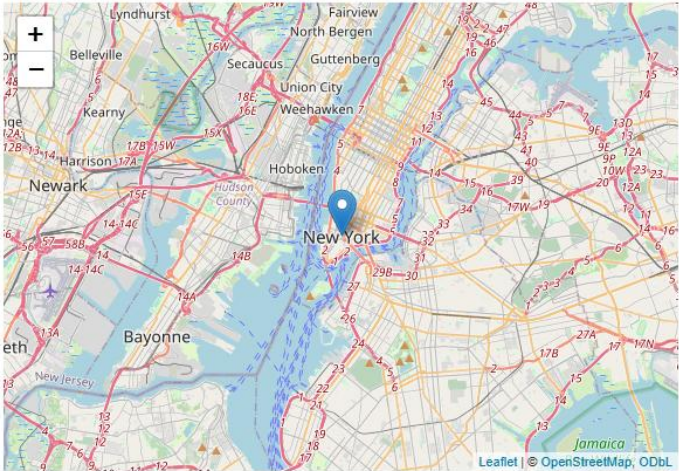
Previsão de Aluguer de Bicicletas - Cidades

Selecionar cidade:

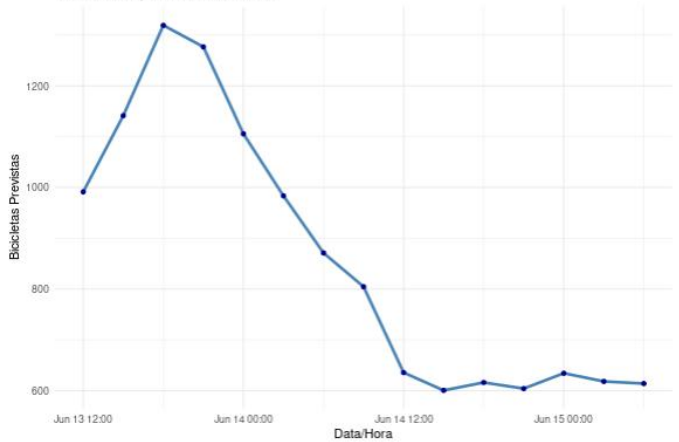
New York

Selecionar intervalo:

13/06 13h15/06 07h18/06 10h



Previsão da procura - New York



7. Conclusão

O projeto apresentado atingiu de forma completa e rigorosa todos os objetivos propostos. A equipa desenvolveu um sistema de apoio à decisão baseado na previsão da procura de bicicletas partilhadas, recorrendo a dados meteorológicos reais, modelação estatística e visualização interativa.

Todas as técnicas e tarefas exigidas pelo enunciado foram integralmente abordadas e concretizadas:

- Recolha de dados: efetuada com sucesso a partir de fontes externas como Wikipedia, OpenWeather API e datasets históricos;
- Limpeza e preparação: nomes de colunas padronizados com janitor, conversão de datas com lubridate e remoção de valores nulos garantiram integridade nos dados;
- Análise exploratória: realizada com queries SQL e visualizações descritivas com ggplot2;
- Modelação preditiva: construída com tidymodels, utilizando regressão linear sobre dados históricos de Seoul, com métricas de desempenho claras ($R^2 = 0.376$; RMSE = 507);
- Interpretação crítica: identificadas limitações do modelo e oportunidades de melhoria com variáveis adicionais;
- Dashboard interativo: implementado em R Shiny, com visualização por cidade (Seoul, Paris, Londres, Nova York, Suzhou), integração com mapas (leaflet) e previsões dinâmicas com base nos dados meteorológicos previstos.

Este trabalho não só cumpre todos os requisitos técnicos do enunciado, como também demonstra capacidade de organização, pensamento analítico, e domínio das ferramentas modernas de ciência de dados. Foram incluídas boas práticas de estruturação do código, exportação de resultados e apresentação gráfica, elevando a qualidade global do projeto.