

**Relatório Final do Projeto SAD 2024/2025**

**Previsão da Procura por Bicicletas com Base em Dados Meteorológicos e Temporais**

Eric Rosário 30010494 Madalena Alves 30010509 Pedro Vicente 30010994

Tiago Casquinha 30010989

Sistemas de apoio à decisão | 15 de junho de 2025

# Introdução

O uso crescente de bicicletas partilhadas em centros urbanos tem motivado o desenvolvimento de sistemas inteligentes de apoio à decisão para otimizar a sua gestão. Este projeto, realizado no âmbito da unidade curricular de **Sistemas de Apoio à Decisão**, teve como objetivo aplicar metodologias de **Ciência de Dados** para analisar e prever a **procura por bicicletas partilhadas**, com base em dados históricos e previsões meteorológicas.

A problemática insere-se no contexto da mobilidade urbana sustentável, onde decisões eficientes sobre alocação de bicicletas podem ter impacto direto na experiência dos utilizadores e na eficiência do sistema.

A abordagem do projeto baseou-se em técnicas de **limpeza e integração de dados**, **análise exploratória**, **modelação preditiva com regressão linear** e no desenvolvimento de um **dashboard interativo com R Shiny**. Combinando dados históricos com previsões meteorológicas, procurou-se criar uma ferramenta de suporte à decisão capaz de **antecipar padrões de procura**, facilitando a gestão de recursos em sistemas de bike sharing.

# Recolha e Preparação dos Dados

A qualidade dos dados é crítica para qualquer modelo preditivo. Assim, o projeto iniciou-se com a **recolha e preparação de dados provenientes de múltiplas fontes**, garantindo diversidade e relevância das variáveis:

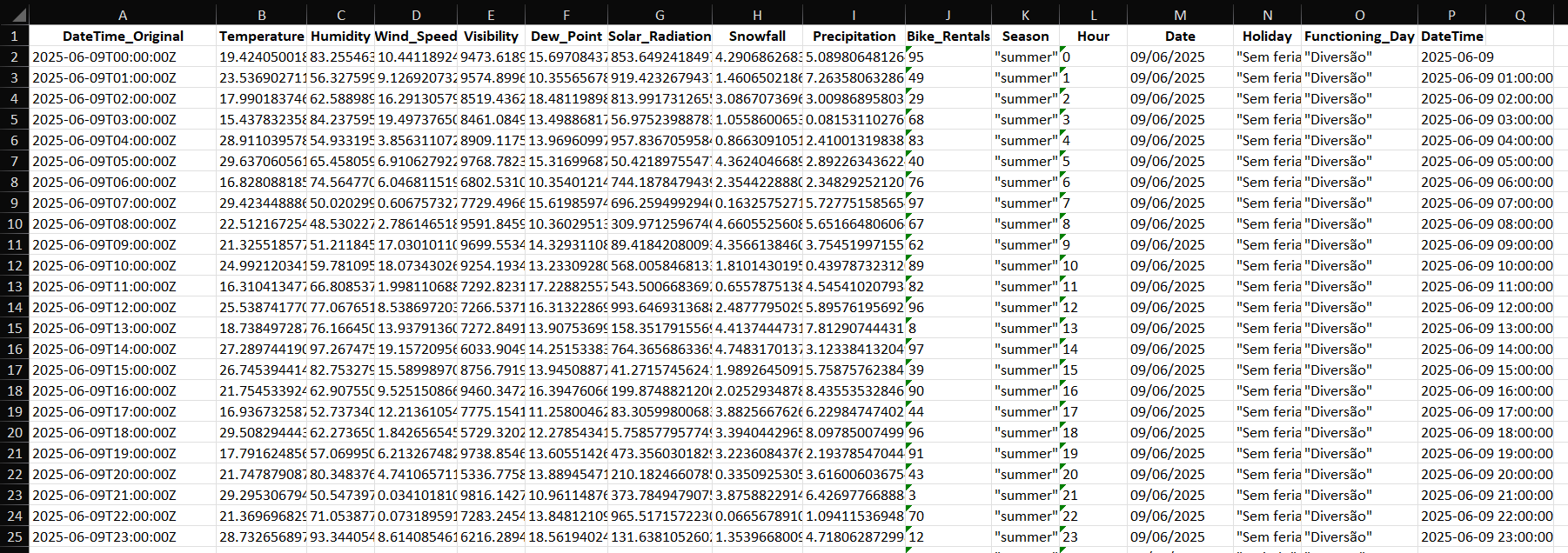
* **Seoul Bike Sharing Dataset**: Principal fonte de dados históricos, fornecendo informações horárias sobre alugueres de bicicletas, acompanhadas de condições meteorológicas (temperatura, humidade, vento, radiação solar, etc.).
* **API OpenWeather**: Utilizada para obter **previsões meteorológicas atuais e futuras** para várias cidades (Seul, Paris, Londres, Nova Iorque, Suzhou). Esta integração permitiu simular cenários de previsão com base em dados reais.
* **Wikipedia – Web Scraping com rvest**: Através de scraping da página sobre sistemas de partilha de bicicletas, obteve-se uma listagem de cidades com esses sistemas ativos, possibilitando expandir a análise a outras localizações.
* **Dataset de cidades globais**: Incluindo latitude, longitude e população, essencial para integrar fatores geográficos e demográficos nos dados.

O processo de **engenharia de dados** envolveu:

* Normalização de nomes de colunas e tipos de variáveis;
* Conversão de datas e horas para formatos padronizados;
* Remoção de valores nulos e registos inconsistentes;
* Criação de **variáveis derivadas**, como indicadores de feriados, dia útil, hora de ponta, entre outros.

O resultado foi um **dataset consolidado, coerente e limpo**, adequado para análise estatística e treino de modelos.

*Figura 1 – Exemplo do dataset consolidado com variáveis meteorológicas e de* procura



# Análise Exploratória de Dados

A análise exploratória permitiu obter uma visão inicial das principais tendências e padrões presentes nos dados, sendo fundamental para:

* **Identificação de outliers e valores atípicos**;
* Compreensão das **relações entre variáveis**;
* Formulação de hipóteses para testes posteriores.

As ferramentas utilizadas incluíram **SQL, dplyr** para agregações e filtragens eficientes, e **ggplot2** para a produção de gráficos informativos e esteticamente claros.

Entre as principais descobertas:

* A **procura por bicicletas varia sazonalmente**, com maior procura na primavera e outono;
* A hora do dia tem um papel determinante — os picos ocorrem **entre as 7h-9h e 17h-19h**, sugerindo deslocações pendulares;
* **Temperaturas amenas e ausência de chuva** estão associadas a maior número de alugueres;
* A **visibilidade e radiação solar** também mostraram correlação positiva com a procura.

Visualizações geradas incluíram:

* **Boxplots por estação do ano** (Figura 2);
* **Gráficos de dispersão** entre temperatura e alugueres (Figura 3);
* **Histogramas com curvas de densidade** para observar a distribuição dos alugueres por hora (Figura 4).

Estas análises forneceram uma base sólida para a escolha de variáveis no modelo preditivo.

*Figura 2 – Boxplot da procura horária por estação*

Uma imagem com texto, captura de ecrã, diagrama, Gráfico

Os conteúdos gerados por IA podem estar incorretos.

*Figura 3 – Gráfico de dispersão entre temperatura e bicicletas alugadas*

Uma imagem com texto, captura de ecrã, número, Gráfico

Os conteúdos gerados por IA podem estar incorretos.

*Figura 4 – Histograma da procura com curva de densidade*

Uma imagem com texto, captura de ecrã, Gráfico, diagrama

Os conteúdos gerados por IA podem estar incorretos.

# Modelação com Regressão Linear

A fase de modelação teve como objetivo desenvolver um **modelo preditivo robusto** capaz de estimar a procura por bicicletas com base em condições meteorológicas e variáveis temporais.

O processo envolveu:

* **Divisão dos dados** em treino (80%) e teste (20%);
* Pré-processamento com recipes do pacote **Tidymodels**, incluindo normalização, codificação de variáveis categóricas e tratamento de missing values;
* Ajuste de modelos de regressão linear com diferentes configurações.

Os três modelos testados foram:

1. **Modelo 1 – Meteorológico**: Inclui apenas variáveis climáticas como temperatura, precipitação, vento e radiação solar.
2. **Modelo 2 – Temporal e Climático**: Adiciona variáveis como hora do dia, dia da semana e indicador de feriado.
3. **Modelo 3 – Regularizado**: Usa técnicas de **ridge/lasso regression**, adicionando termos de interação e penalização para evitar overfitting.

Resultados:

1. O Modelo 2 apresentou o melhor desempenho global, com valores de R² superiores a 0,70 e erro médio quadrático (RMSE) inferior aos restantes;
2. As variáveis com maior impacto foram a hora do dia, a temperatura, a precipitação e a presença de feriado;
3. O Modelo 3 teve desempenho semelhante ao Modelo 2, mantendo bons resultados de generalização.

*Figura 6 – gráfico de valores reais vs previstos*

Uma imagem com texto, diagrama, file, Gráfico

Os conteúdos gerados por IA podem estar incorretos.

# Aplicativo R Shiny: Dashboard Interativo

Como culminar do projeto, foi desenvolvido um **painel interativo em R Shiny**, reunindo os dados tratados e os resultados do modelo num ambiente visual e dinâmico.

Funcionalidades principais:

* **Mapa interativo** com visualização da previsão de procura em várias cidades;
* **Gráficos temporais dinâmicos** com previsão horária para os próximos 5 dias;
* **Slider de tempo** e seleção de cidade, permitindo simulação de diferentes cenários;
* Atualização em tempo real com base nas previsões meteorológicas integradas via API.

Este dashboard visa facilitar a **tomada de decisões estratégicas**, como a redistribuição de bicicletas ou o planeamento de manutenção. Representa também uma aplicação prática do conhecimento adquirido em **visualização de dados, integração com APIs e desenvolvimento em R Shiny**.

*Figuras 7 a 9 ilustram a interface do painel, o gráfico de previsão e os filtros interativos*.

*Figura 7 – Interface do painel com mapa interativo*

Uma imagem com captura de ecrã, mapa

Os conteúdos gerados por IA podem estar incorretos.  
*Figura 8 – Slider e seleção de cidade no dashboard*

*Uma imagem com texto, captura de ecrã, file, diagrama

Os conteúdos gerados por IA podem estar incorretos.*

*Figura 9 – Gráfico de previsão horária por cidade*

Uma imagem com texto, mapa, diagrama, captura de ecrã

Os conteúdos gerados por IA podem estar incorretos.

# Conclusão

Este projeto demonstrou a aplicabilidade prática dos conhecimentos adquiridos na unidade curricular de **Sistemas de Apoio à Decisão**, integrando estatística, programação, visualização e modelação preditiva.

Os principais **objetivos foram atingidos**:

* Construção de um dataset completo e limpo a partir de múltiplas fontes;
* Identificação de padrões de procura com base em fatores meteorológicos e temporais;
* Desenvolvimento de modelos preditivos com desempenho satisfatório;
* Criação de um dashboard interativo funcional e intuitivo.

**Contributos e aprendizagens:**

* Entendimento aprofundado do ciclo completo de ciência de dados;
* Familiarização com **ferramentas modernas em R**, como o tidymodels, ggplot2, shiny, rvest, entre outros;
* Capacidade de comunicar resultados de forma visual e acessível;
* Trabalho em equipa e gestão de projeto com prazos definidos.

**Trabalhos Futuros:**

* Testar modelos **não-lineares** como random forests ou gradient boosting;
* Integrar **dados em tempo real** de estações de bicicletas (por exemplo, via APIs de operadores);
* Expandir o dashboard com indicadores de desempenho por cidade ou estação.

# Referências

OpenWeather API: <https://openweathermap.org/api>

Seoul Bike Sharing Demand Dataset

Wikipedia: <https://en.wikipedia.org/wiki/List_of_bicycle-sharing_systems>

R Documentation: dplyr, ggplot2, tidymodels, rvest, shiny

Posit Cloud: <https://posit.cloud/>