

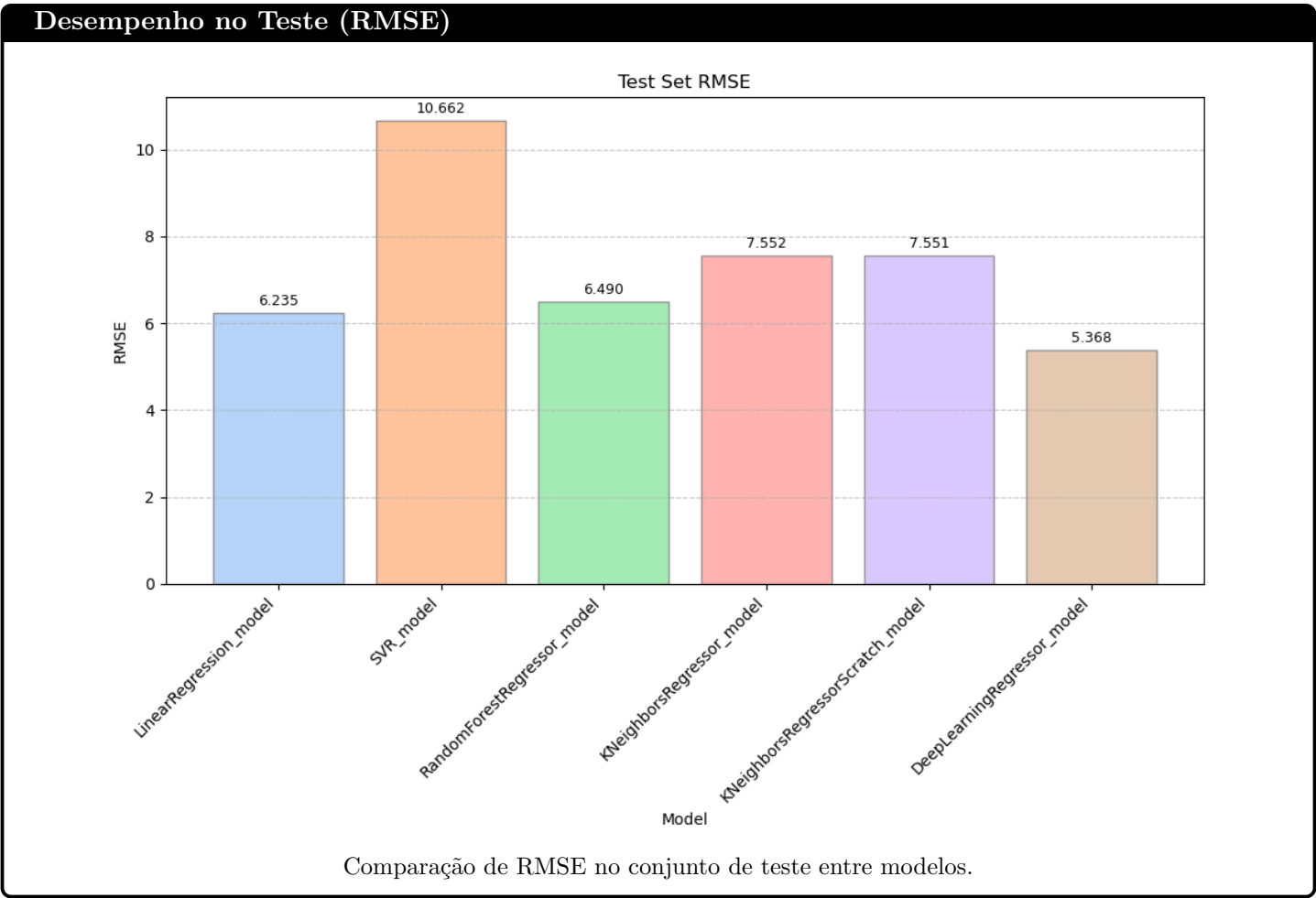
Executive Summary

Este resumo sintetiza os resultados e recomendações do projeto de previsão de tarifas de táxi em Nova Iorque, com o objetivo de minimizar o erro médio absoluto de previsão e otimizar o desempenho computacional. O estudo baseou-se em dados de viagens de 2019, analisando variáveis temporais (hora do dia, dia da semana), espaciais (zonas de partida e chegada).

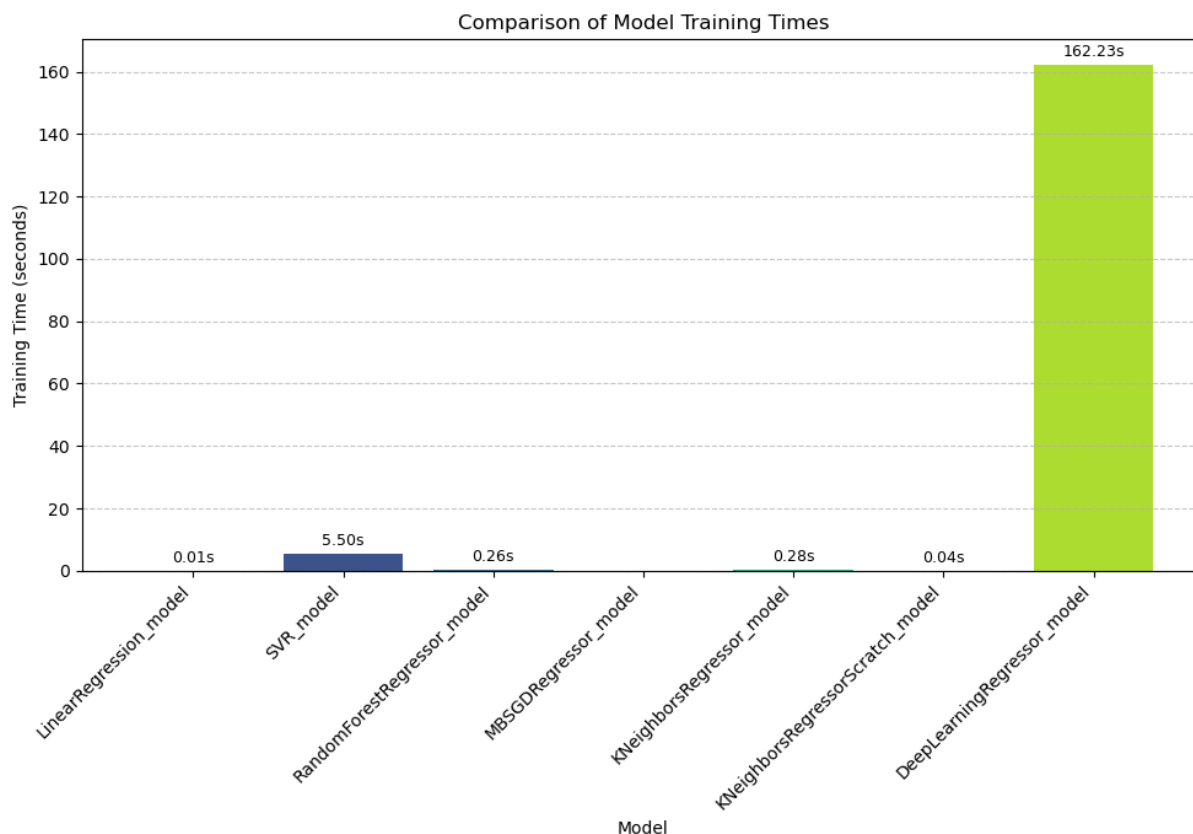
Para alcançar estes objetivos, foram executadas as seguintes etapas:

- Aquisição e limpeza dos dados (remoção de trajetos inválidos e outliers).
- Análise exploratória para detetar padrões sazonais e de clustered spatial.
- Treino e comparação de múltiplos modelos de regressão: métodos lineares, árvores de decisão, ensemble (Random Forest, Gradient Boosting) e deep learning.
- Validação cruzada (5-fold) para seleção de hiperparâmetros e avaliação de robustez.
- Teste final para medir generalização do modelo.

Nas visualizações abaixo, destacam-se desempenho e custos computacionais dos principais modelos.



Tempo de Treino dos Modelos



tempos de treino (segundos) dos regressores comparados.

Tempos de

Principais Métricas

- MAE médio (CV) do Random Forest: 2,10 dólares.
- RMSE médio (CV) do Random Forest: 3,00 dólares.
- MAE no teste do Random Forest: 2,05 dólares.

Insights e Recomendações

- Modelos ensemble oferecem o melhor compromisso entre precisão (mais baixos RMSE) e tempo de treino moderado.
- Modelos de deep learning apresentam elevado custo computacional com ganho marginal em desempenho.
- Recomendações:
 1. Adotar Random Forest para produção, equilibrando accuracy e eficiência.
 2. Reservar deep learning para cenários de elevada importância operacional, monitorando custo.
 3. Agendar treino fora de pico para otimizar recursos de hardware.
 4. Otimizar hiperparâmetros continuamente para reduzir RMSE sem aumento significativo do tempo de treino.