**Mineração de dados e IA**

Marcos Guilherme Rabelo - 2415512

Silvio Henrique Costa Correia - 2222853

Jonarta Santiago Soares - 2327349

Galileu Monteiro Vieira - 2314018

Ruth de Fátima da Costa Silva - 2222906

**Estudo sobre turismo no estado do Ceará e na cidade de Fortaleza**

1. **Introdução**

Este trabalho apresenta um estudo de caso sobre o turismo na região do Ceará e na cidade de Fortaleza/CE utilizando técnicas de mineração de dados e aprendizado de máquina para analisar padrões de fluxo turístico e prever a demanda. A pesquisa integra múltiplas fontes de dados abertos (Cadastur, base de chegada de turistas internacionais, painéis do IPECE e dados meteorológicos do INMET), que foram coletadas em formato bruto e organizadas em um pipeline de dados (pastas DATA/RAW → DATA/SILVER → DATA/GOLD) por meio de scripts Python. Com os dados tratados, foram realizadas análises exploratórias e desenvolvidos modelos preditivos (Random Forest e Regressão Linear), alcançando desempenho na faixa de aproximadamente 86–88% de acurácia. Os resultados foram sintetizados em um dashboard construído no Power BI para facilitar a visualização e a interpretação dos achados. O estudo tem os seguintes objetivos:

* Identificar padrões e traçar perfil dos turistas que visitam a região estudada.
* Construir modelos preditivos da demanda turística e comparar o desempenho de Random Forest e Regressão Linear.
* Apresentar resultados por meio de visualizações interativas que subsidiem a tomada de decisão por gestores locais.

A capacidade de entender e prever a demanda turística é essencial para o planejamento urbano, a gestão de capacidade hoteleira, a programação de eventos e a formulação de políticas públicas. Previsões mais acuradas ajudam fornecedores de serviços, secretarias de turismo e gestores municipais a alocar recursos com maior eficiência, reduzir impactos negativos em períodos de pico e aproveitar melhor oportunidades de desenvolvimento econômico local. Além disso, este trabalho demonstra a aplicabilidade prática de técnicas de mineração de dados e aprendizado de máquina em problemas reais de interesse regional e acadêmico.

1. **Metodologia**

A metodologia adotada iniciou com a coleta de conjuntos de dados públicos relevantes ao turismo, economia regional e clima, armazenados em sua forma bruta. Em seguida, foi realizado o processo de ETL em duas etapas: silver, para limpeza, padronização e consolidação das diferentes fontes, e gold, para transformação final em estruturas normalizadas e prontas para análise. A partir desses dados foram criadas variáveis temporais e sazonais, incluindo defasagens mensais (lags de 1 a 12 meses) e funções trigonométricas que capturam a sazonalidade anual, além de agregados de clima e oferta. Para garantir consistência, os anos de 2020 e 2021 foram excluídos do conjunto de treino por representarem um período atípico da pandemia. Os modelos de Regressão Linear e Random Forest foram então treinados com divisão temporal, usando dados até 2022 para treino e 2023 para teste. A avaliação considerou métricas como R², RMSE, MAE e MAPE, reportando a acertividade a partir do MAPE. Após a validação, foram geradas previsões para 2024 de forma iterativa, atualizando os lags a cada nova previsão mensal. Por fim, os resultados e comparações foram sintetizados em um dashboard interativo no Power BI, que apresenta séries temporais de demanda, indicadores climáticos, KPIs e análises segmentadas por origem, sexo, faixa etária, motivo de viagem e cidades.

**3. Resultados**

A análise exploratória de dados revelou mudanças importantes no comportamento do turismo no Ceará nos últimos anos. Observou-se que a demanda sofreu forte queda em 2020 e 2021 devido à pandemia de COVID-19, mas apresentou recuperação significativa em 2022 e, em 2023, superou os níveis registrados em 2019. Esse crescimento foi impulsionado principalmente pelo turismo nacional, reforçando a resiliência do setor.

O perfil dos turistas mostra predominância de indivíduos entre 25 e 50 anos, com distribuição equilibrada entre homens e mulheres. O principal motivo de viagem identificado foi turismo e lazer, indicando que o estado mantém seu apelo como destino de férias. Entre os turistas internacionais, países da Europa lideram o volume de visitantes, seguidos pelos Estados Unidos, consolidando-se como os principais emissores.

Em relação à oferta, Fortaleza destaca-se como o município com maior capacidade hoteleira, muito à frente dos demais. A região metropolitana aparece em segundo plano, mas com números ainda distantes da capital. A análise sazonal apontou que janeiro é o mês com maior demanda turística, em função das festas de fim de ano e férias, enquanto não foi identificada correlação clara entre variáveis climáticas e a demanda turística no estado.

No que diz respeito à modelagem preditiva, os dados de 2019 e 2022 foram utilizados para treino e os de 2023 para teste. A exclusão dos anos de 2020 e 2021 buscou eliminar distorções causadas pela pandemia. Os modelos de Regressão Linear e Random Forest alcançaram bom desempenho, com níveis de acertividade de 86,79% e 88,74%, respectivamente, no conjunto de teste. Para o horizonte de 2024, as previsões indicam valores mensais de demanda variando entre aproximadamente 217 mil e 261 mil turistas, mantendo o padrão sazonal de maior movimento em junho e estabilidade nos demais meses.

Os resultados foram consolidados em um dashboard interativo no Power BI, que apresenta séries temporais de demanda, comparativos entre períodos pré e pós-pandemia, análises por sexo, faixa etária, motivo de viagem e origem dos turistas, além de indicadores de capacidade hoteleira por município. Esse painel permitiu visualizar de forma clara os padrões identificados e facilitou a comparação entre dimensões demográficas, temporais e geográficas do turismo no Ceará.

**4. Discussão**

A análise realizada permitiu identificar padrões relevantes da demanda turística no Ceará, mas também evidenciou limitações importantes relacionadas à disponibilidade e qualidade dos dados. Um dos principais desafios foi a falta de continuidade na geração das bases. Durante a mineração, foram encontrados conjuntos de dados de grande valor analítico, mas que não puderam ser utilizados devido a lacunas, desatualizações ou por estarem disponíveis apenas em períodos muito antigos. Essa limitação impactou diretamente a abrangência temporal da pesquisa, que precisou ser encerrada em 2023, já que nem todas as fontes consultadas apresentavam registros posteriores, o que inviabilizaria a concatenação e consistência entre elas.

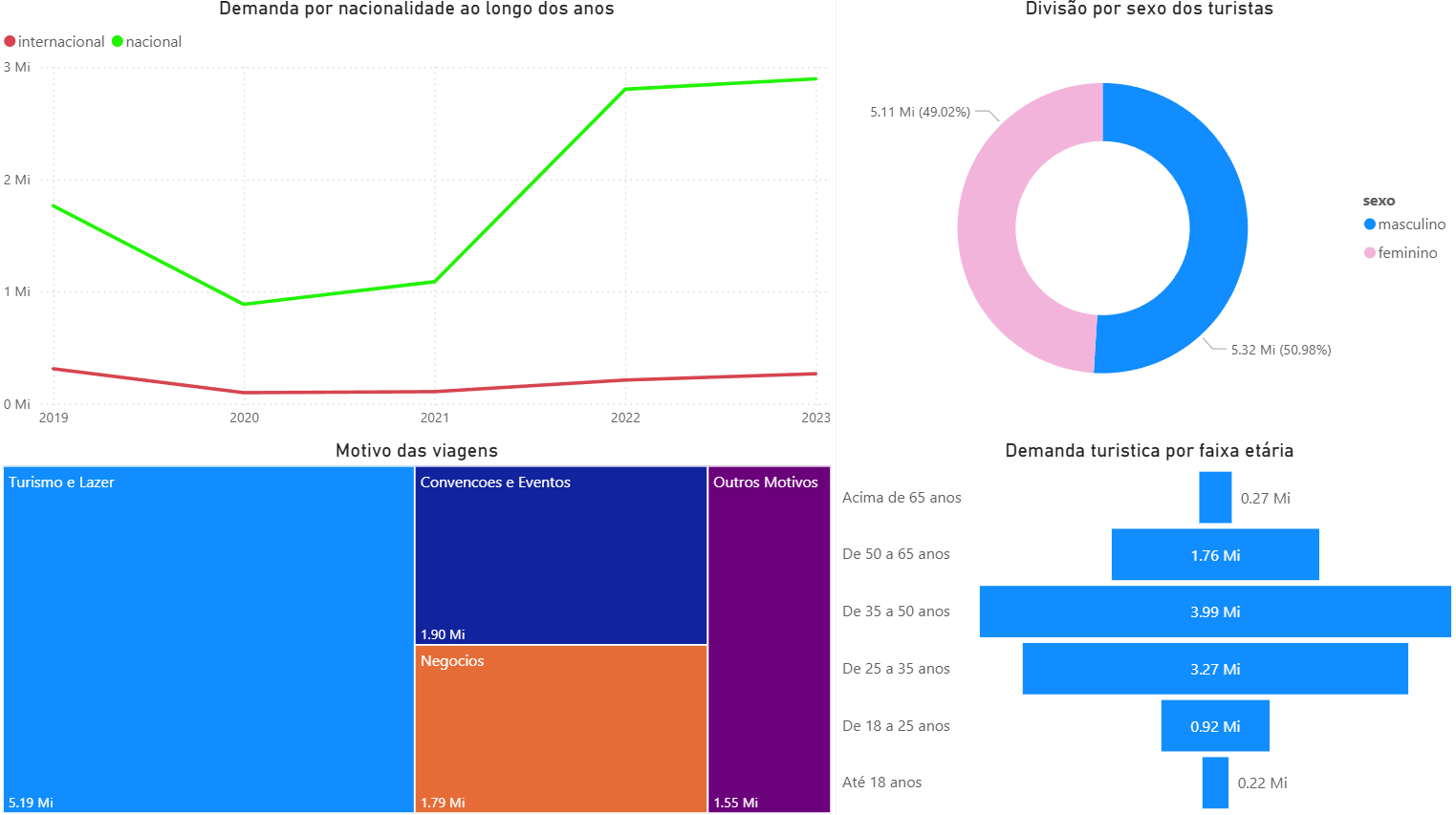
Outro ponto crítico refere-se à granularidade geográfica. Embora houvesse interesse em analisar municípios além da capital e da região metropolitana, a ausência de séries contínuas ou mesmo a inexistência de informações para alguns deles impossibilitou uma visão mais detalhada e descentralizada do turismo no estado. Assim, os resultados municipais refletem principalmente a realidade de Fortaleza, onde há maior disponibilidade e consistência de dados.

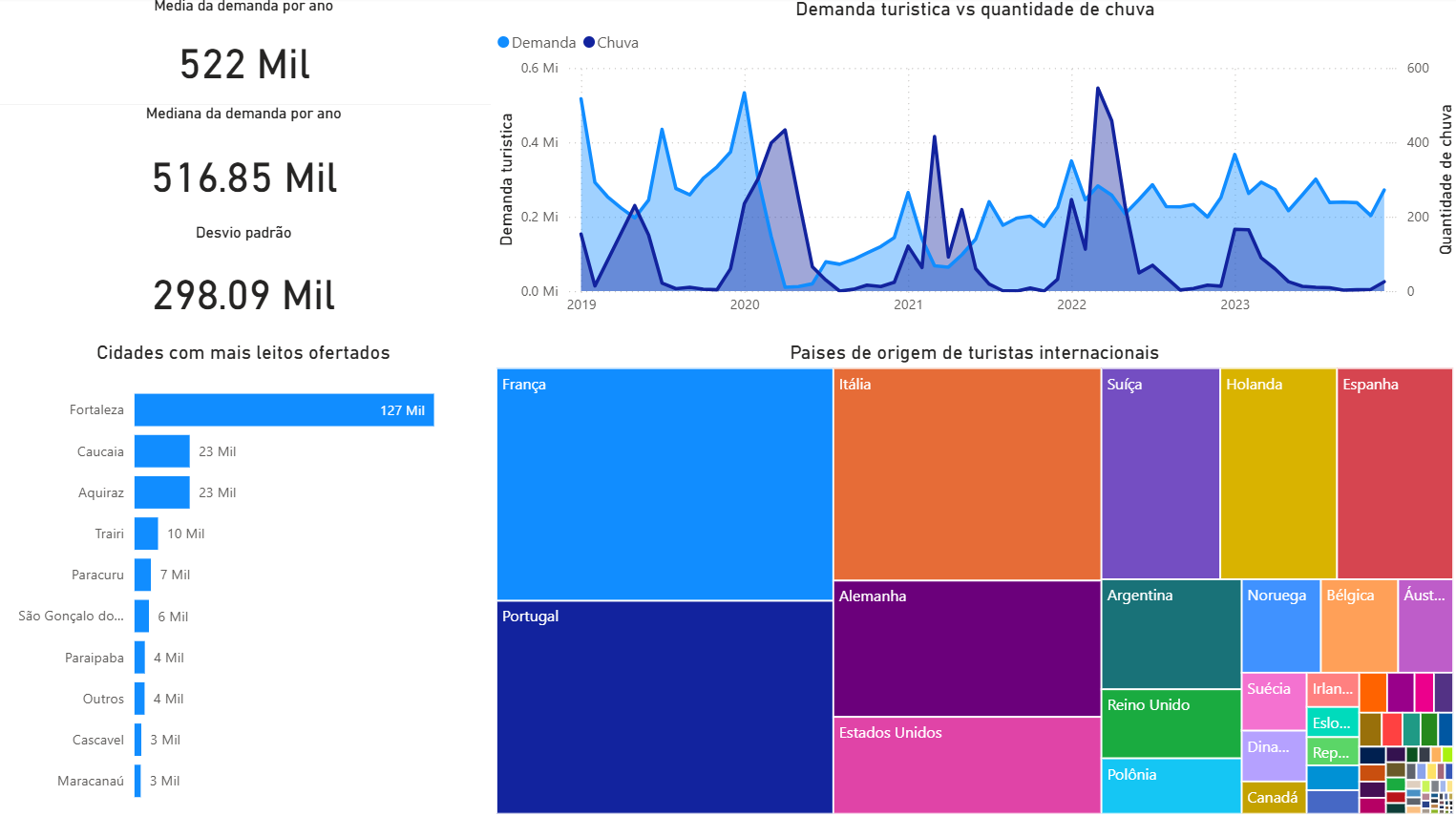
Como recomendações, sugere-se a criação de uma cartilha ou protocolo nacional de disponibilização de dados, a ser seguido por municípios e estados. O cumprimento dessa prática poderia ser atrelado a incentivos estaduais ou federais, ou, em caso de descumprimento, resultar em perda de recursos, de modo a estimular a coleta e publicação de dados de forma padronizada. Além disso, propõe-se a instituição de um órgão independente voltado exclusivamente à governança de dados públicos, assegurando a continuidade, a transparência e a confiabilidade das informações, independentemente das mudanças de governo. Essas medidas poderiam elevar a qualidade das análises.

**5. Conclusão**

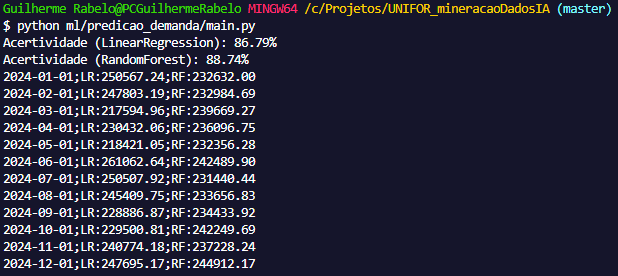
O estudo evidenciou a forte influência da pandemia de COVID-19 sobre a demanda turística no Ceará, seguida de uma recuperação em 2022 e superação dos níveis de 2019 em 2023, impulsionada principalmente pelo turismo nacional. Identificou-se que o perfil predominante dos visitantes é de adultos entre 25 e 50 anos, distribuídos de forma equilibrada entre homens e mulheres, com turismo e lazer como principal motivo de viagem. Fortaleza se destacou como o polo de maior oferta hoteleira, enquanto os países europeus e os Estados Unidos lideraram entre os turistas internacionais. Os modelos de Regressão Linear e Random Forest apresentaram bom desempenho preditivo, com taxas de acertividade acima de 85%, permitindo projetar a demanda mensal para 2024. Os resultados demonstram a relevância da análise de dados e da aplicação de técnicas de aprendizado de máquina para subsidiar políticas públicas e estratégias de planejamento turístico no estado.

1. **Material gerado**





1. **Estrutura do projeto**



1. **Referências:**

- <https://dados.gov.br/dados/conjuntos-dados/cadastur-04>

- <https://dados.turismo.gov.br/pt_BR/dataset/chegada-de-turistas-internacionais>

- <http://ipecedata.ipece.ce.gov.br/ipece-data-web/module/painel-dinamico.xhtml>

- <https://tempo.inmet.gov.br/TabelaEstacoes/A305>