

1. Resumo da ideia

Mapear sistematicamente **locais turísticos** de Angola (património cultural, áreas naturais, atracções locais, infraestrutura e serviços) e aplicar **análise preditiva** para identificar onde o desenvolvimento turístico é mais provável e mais desejável (por exemplo: zonas com alta probabilidade de aumento de visitantes, necessidade de investimento em acessibilidade, avaliação de risco ambiental).

O sistema vai integrar GIS, sensoriamento remoto, dados abertos (OpenStreetMap, UNESCO, estatísticas nacionais), machine learning espacial e modelos de previsão temporal para gerar mapas, rankings e recomendações operacionais para autoridades públicas, operadores privados e ONGs.

2. Objetivos

1. Construir um inventário georreferenciado de atracções turísticas e serviços associados em Angola.
2. Gerar indicadores espaciais (acessibilidade, proximidade a estradas/aeroportos, oferta de serviços, LULC — land use/land cover).
3. Treinar modelos preditivos que estimem:
 - (i) probabilidade de crescimento de visitantes por site;
 - (ii) impacto do investimento em infraestrutura;
 - (iii) vulnerabilidade ambiental/carrying capacity.
4. Produzir dashboards e mapas interativos para apoiar decisões de planeamento.

3. Questões de investigação (exemplos)

- Quais fatores locais explicam melhor variações em fluxos de visitantes (acessibilidade, infraestrutura, atributos naturais, património cultural)?
- Onde o investimento em estradas ou serviços teria maior retorno esperado em termos de visitas e benefícios socioeconómicos?
- Quais sítios são prioritários para conservação porque um aumento rápido de visitantes pode causar danos (capacidade de suporte)?

4. Resumo e referências chave

- **Mapeamento turístico + dados voluntários:** trabalhos e portais que usam OpenStreetMap e camadas voluntárias para construir inventories de POIs e atracções.
- **Património e inventário cultural:** UNESCO e listas nacionais úteis para priorização e rotulagem de sites históricos / protegidos.
- **Sensoriamento remoto e LULC/expansão urbana:** estudos que combinam Landsat/Sentinel com OSM, útil para medir mudanças ao redor de atracções (pressão de desenvolvimento).

- **ML aplicado a turismo / previsão:** ML (random forest, XGBoost, redes neurais, séries temporais híbridadas...) para previsão de demanda e análise de fatores explicativos;

5. Dados recomendados (fontes, tipo e utilidade)

Geoespaciais / POIs

- OpenStreetMap / HOTOSM pontos de interesse (atracções, alojamento, restaurantes, estradas). Útil para inventário inicial e atributos.

Património e sítios oficialmente reconhecidos

- UNESCO / lista de sítios e descrições. Serve para rotular e priorizar.

Sensoriamento remoto

- Sentinel-2, Landsat 8/9 — análise LULC, monitoramento de mudanças, detecção de expansão urbana e análise de cobertura vegetal.
- Ferramentas: Google Earth Engine ou processamento local com rasterio/GDAL.

Estatísticas e fluxos

- Dados de chegadas turísticas, gastos e economia (World Bank, TradingEconomics, relatórios sectoriais, EIU / Industry reports) para séries temporais e calibração.

Dados auxiliares

- Infraestrutura: redes rodoviárias (OSM), aeroportos, portos.
- População e socioeconómicos: WorldPop e INE.
- Meteorologia/clima (ERA5, CHIRPS) para modelar sazonalidade.

Coleta primária (quando possível)

- Surveys locais (contagem de visitantes, reconhecimento local), crowdsourcing via app/mobile para validar POIs e medir satisfação.

6. Metodologia proposta (pipeline técnico)

1. Ingestão e limpeza

- Extrair OSM (pontos e redes) e normalizar tags (atração, hotel, restaurante, serviços). Incluir atributos: nome, tipo, telefone/site (quando houver).

2. Gerar covariáveis espaciais

- Acessibilidade: distância de rede a estradas principais, tempo de viagem estimado a aeroportos (network analysis).

- Contexto LULC: extração de classes (urbano, vegetação, água) com Sentinel/Landsat.
- Serviços: densidade de alojamentos, restaurantes num raio (kernel density).
- Indicadores socioeconómicos: densidade populacional, renda per capita (onde disponível).

3. Modelagem preditiva

- Objetivos: previsão de visitas (séries temporais por site/região), classificação de “potencial de desenvolvimento” (alto/médio/baixo).
- Modelos candidatos:
 - Regressão espacial (SAR, SEM) para capturar autocorrelação espacial.
 - Random Forest / XGBoost (variações de importância de features).
 - Redes neurais LSTM/Transformer para séries temporais de chegadas + exógenas.
 - Graph Neural Networks (GNN) para incorporar ligações entre sites via rede de transporte.
 - Modelos híbridos para otimização de hiperparâmetros.
- Validação: cross-validation espacial (block CV), métricas RMSE/MAE para regressão, AUC/precision-recall para classificação.

4. Análise de cenários

- Simular intervenções (ex.: construção de estrada, melhoria de aeroporto) e estimar impacto nas visitas e pressão ambiental.

5. Entrega

- Mapas interativos (Leaflet/Mapbox), dashboards (Dash/Streamlit/Power BI) e relatórios com recomendações.

7. Pilha tecnológica sugerida

- **Armazenamento e GIS:** PostgreSQL + PostGIS, GeoServer (se for publicada WMS/WFS).
- **Processamento raster/vector:** GDAL, rasterio, geopandas, Fiona.
- **Scripting / ML:** Python (pandas, scikit-learn, xgboost, PyTorch ou TensorFlow, statsmodels), PySAL (análise espacial), NetworkX / pgRouting.
- **Nuvem / escala:** Google Earth Engine (para processamento Sentinel/Landsat em larga escala, digital ocean para armazenamento e deployment).

- **Visualização:** QGIS para análise exploratória; web: Leaflet/Mapbox + Dash/Streamlit para dashboards.
- **DevOps:** Docker, CI (GitHub Actions), Git.
- **Coleção mobile / campo:** app simples (React Native / Flutter) ou formulários ODK/KoBoToolbox para colecta de dados em campo.

8. Métricas de sucesso e avaliação

- **Cobertura do inventário:** % de POIs validados vs estimado (benchmark com OSM).
- **Acurácia preditiva:** RMSE/MAE para previsão de chegadas; AUC para classificações.
- **Utilidade para stakeholders:** pesquisas qualitativas com autoridades locais (aceitação das recomendações).
- **Impacto ambiental (sinal de alerta):** número de sites classificados como alto risco por capacidade de suporte.

9. Entregáveis sugeridos

- Base de dados georreferenciada (PostGIS) com POIs e atributos.
- Conjunto de mapas temáticos (acessibilidade, potencial de desenvolvimento, risco ambiental).
- Dashboard web interativo com filtros e cenários.
- Relatório técnico + recomendações de políticas (priorização de investimentos).
- Código reproduzível (notebooks, scripts Dockerizados).