

DESCRIÇÃO DE FUNÇÕES PARA O PROJETO FINAL

Cientista de Dados

Responsável: Baltazar Henriques

Responsável pelo pipeline de dados, ETL, modelagem inicial e integração com banco PostGIS.

Tarefas técnicas:

1. Aquisição e tratamento de dados

Coletar datasets de:

- OpenStreetMap (POIs turísticos, estradas, rios, parques...).
- INE Angola (dados populacionais e econômicos).
- Clima e uso do solo (Sentinel, Copernicus, WorldClim).
- Dados de mobilidade e infraestrutura (Google Maps API / Geonames).
- Limpeza e padronização: normalizar nomes, tipos e coordenadas.
- Conversão para formato geoespacial (GeoJSON, shapefile, CSV com lat/lon).

2. ETL e integração com banco de dados

Escrever **scripts Python (pandas, geopandas, psycopg2)** para:

- Criar tabelas e índices espaciais no **PostGIS**.
- Inserir dados transformados via *bulk insert*.
- Automatizar carga incremental (novos dados de turismo, infraestrutura etc.).

3. Estrutura do banco (proposta inicial)

Tabela	Descrição	Campos principais
pontos_turisticos	Locais de interesse turístico	id, nome, categoria, descricao, provincia, municipio, latitude, longitude, fonte_dados, geom (POINT)
infra_estruturas	Infraestruturas de apoio (hospedagem, transportes, serviços)	id, tipo, nome, estado_conservacao, provincia, latitude, longitude, geom (POINT)
estatisticas_turismo	Dados agregados de fluxo turístico por província	id, provincia, ano, visitantes_nacionais, visitantes_estrangeros, receita, geom (POLYGON)
clima	Dados médios climáticos por província	id, provincia, temperatura_media, precipitacao_media, umidade, ano, geom (POLYGON)
uso_solo	Tipos de cobertura do solo	id, categoria, area_km2, provincia, geom (POLYGON)
model_inputs	Variáveis derivadas para modelagem	id, provincia, densidade_pontos_turisticos, distancia_aeroporto, densidade_populacional, indice_acessibilidade, geom (POLYGON)

Entregáveis:

→ Banco PostGIS configurado e populado;

- Scripts ETL funcionais e reexecutáveis;
- Documentação do schema de dados.

2. Técnico de Machine Learning e IA

Responsável: Manuel de Deus

Responsável pela criação, treino e validação de modelos preditivos de potencial turístico.

Tarefas técnicas:

1. Preparação de features

- Usar dados da tabela `model_inputs` para criar *feature sets*.
- Normalizar variáveis (z-score, min-max).
- Calcular *spatial lag variables* (ex: influência de POIs vizinhos).

2. Modelagem preditiva

Implementar modelos:

- **Random Forest / XGBoost**: classificação de potencial turístico.
- **Regressão espacial (PySAL)**: avaliar autocorrelação geográfica.
- **Redes Neurais (PyTorch / TensorFlow)**: previsão temporal (fluxo turístico).

Cross-validation espacial (train/test split por províncias).

3. Interpretação e avaliação

- Avaliar desempenho (RMSE, R^2 , AUC).
- Analisar importância das variáveis (SHAP, feature importance).
- Gerar mapas de predição de potencial turístico (por província/município).

4. Entrega dos resultados

- Exportar resultados para PostGIS (tabela `resultados_modelo`).
- Integrar via API (com DevOps) para visualização no dashboard.

Entregáveis:

- Modelos preditivos otimizados;
- Notebook técnico com documentação e resultados;
- Arquivos `.pkl` / `.onnx` dos modelos salvos.

3. Analista GIS (Desenvolvedor Geoespacial)

Responsável: Natus Vincere

Responsável por toda a manipulação geoespacial, análise SIG e visualização.

Tarefas técnicas:

1. Preparação e análise espacial

Calcular:

- Densidade de pontos turísticos (Kernel Density).
- Distância média até estradas/aeroportos.
- Índices espaciais (Moran's I, Getis-Ord Gi*).

Classificação de uso do solo (imagens Sentinel / QGIS / Google Earth Engine).

2. Integração PostGIS + QGIS

Criar camadas e *views* espaciais no banco.

Gerar mapas temáticos de:

- Acessibilidade turística;
- Clusters de alto potencial;
- Cobertura do solo.

3. Dashboard interativo

- Usar **Leaflet / Mapbox / Streamlit / Dash** para mapa interativo.
- Filtros por província, categoria turística e acessibilidade.
- Integração com API do DevOps para exibir resultados do modelo IA.

Entregáveis:

- Mapas interativos publicados;
- Análises espaciais documentadas;
- Dashboard web conectado ao banco.

DevOps / Infraestrutura

Responsável: Jorge Tomas

Responsável pela automação, deploy e escalabilidade do projeto.

Tarefas técnicas:

1. Infraestrutura

Configurar ambiente com **Docker** (containers: API, banco, frontend).

Criar docker-compose.yml com serviços:

- PostgreSQL + PostGIS
- FastAPI (backend IA)
- Streamlit/Dash (dashboard)

Gerir volumes e persistência de dados.

2. Pipeline CI/CD

GitHub Actions para testes automáticos e deploy contínuo.

Linting (Flake8, Black) e testes unitários dos scripts ETL e modelos.

3. APIs e serviços

Implementar **API REST (FastAPI)** com endpoints:

- /predict - retorna previsões do modelo IA.
- /data/locais - consulta pontos turísticos.
- /stats - indicadores agregados.

Autenticação simples com tokens.

4. Monitorização

- Configurar logs e métricas (Prometheus + Grafana).
- Monitorar uso de CPU/memória e latência de API.

Entregáveis:

→ Ambiente Docker completo;

→ API funcional integrada ao banco e modelo;

→ CI/CD ativo e documentação de deploy.