

# Nota Conceitual E Plano De Implementação Do Projecto Capstone

## Título do projecto: Angola Tourism Insight

### Membros da equipe

1. Ângelo Rocha Garcia
2. Azevaldo Miguel Caluaco
3. Dorivaldo Catala Mandele
4. Francisco Ramos Cadete
5. Francisco Benguela José
6. Reinaldo Sagrado Paulo

### Nota conceitual

#### 1. Visão geral do projecto

O projecto Angola Tourism Insight propõe o uso de Inteligência Artificial (IA) e *Machine Learning* (ML) para analisar dados turísticos em Angola, gerando insights preditivos e prescritivos que apoiem autoridades, empresas e turistas na tomada de decisões mais informadas. A iniciativa está alinhada com os Objectivos de Desenvolvimento Sustentável (ODS) da ONU, especialmente:

- ODS 8: Trabalho Decente e Crescimento Económico: ao impulsionar o turismo, promove-se geração de emprego e renda.
- ODS 9: Indústria, Inovação e Infraestrutura: ao aplicar tecnologias avançadas para modernizar o sector turístico.
- ODS 11: Cidades e Comunidades Sustentáveis: ao fomentar práticas turísticas sustentáveis e integradas às comunidades locais.

#### Situação Problemática

Angola possui um enorme potencial turístico ainda sub-explorado, em parte pela falta de dados organizados e análises preditivas que orientem políticas públicas e investimentos. O projecto visa resolver essa lacuna, transformando dados brutos em informações estratégicas.

A solução procura revolucionar o sector turístico angolano, promovendo inovação, atraindo investimentos, fortalecendo o ecossistema local e contribuindo directamente para o desenvolvimento sustentável do país.

#### 2. Objectivos

- Identificar padrões no comportamento turístico em Angola;
- Prever fluxos turísticos com base em variáveis como clima, eventos e feriados.
- Criar *dashboards* interativos para apoiar decisões em tempo real;
- Sugerir soluções sustentáveis e baseadas em dados para impulsionar o turismo local.

Com o projecto, o grupo pretende desenvolver e implementar um sistema inteligente que permita antecipar tendências e otimizar estratégias para o sector turístico, isso por meio de relatórios analíticos que orientem políticas públicas e

investimentos, modelos de *machine learning* validados para previsão de fluxo turístico, e um painel interativo acessível a gestores, empreendedores e pesquisadores.

Ao fornecer esta ferramenta tecnológica robusta, o projecto permitirá que o sector turístico angolano planeie com base em evidências, melhore a experiência dos visitantes, reduza desperdícios e maximize o impacto económico e social do turismo, promovendo um crescimento sustentável e inclusivo.

### **3. Histórico**

O sector turístico em Angola possui um potencial significativo ainda sub-explorado, devido à sua diversidade cultural, paisagens naturais e localização estratégica. Todavia, o crescimento sustentável do turismo enfrenta diversos desafios, entre eles, a falta de dados organizados e acessíveis sobre o comportamento dos turistas, a ausência de ferramentas analíticas que permitam prever fluxos turísticos e planejar acções estratégicas, a baixa digitalização dos serviços turísticos e pouca integração entre os actores do sector por fim, a desconexão entre políticas públicas e dados reais sobre o turismo.

Estes factores dificultam a tomada de decisões informadas, o planeamento de investimentos e a promoção eficaz do turismo sustentável.

Algumas iniciativas pontuais têm sido implementadas por órgãos como o Ministério da Cultura e Turismo e o Instituto Nacional de Estatística (INE Angola), que disponibilizam relatórios e dados estatísticos. No entanto, essas iniciativas são limitadas em escopo e profundidade, pois, não utilizam tecnologias avançadas como inteligência artificial ou *machine learning*, não oferecem ferramentas interativas para análise em tempo real e não conseguem prever tendências ou simular cenários futuros.

A abordagem baseada em *machine learning* é essencial para superar essas limitações, pois permite analisar grandes volumes de dados de forma eficiente, detectar padrões ocultos no comportamento turístico, prever fluxos futuros com base em variáveis como clima, eventos e sazonalidade, gerar insights prescritivos que orientem acções concretas e automatizar a tomada de decisão com base em evidências.

Além disso, o uso de *machine learning* promove inovação tecnológica, fortalece o ecossistema de dados em Angola e contribui directamente para os Objectivos de Desenvolvimento Sustentável (ODS), especialmente os relacionados à infraestrutura, crescimento económico e sustentabilidade urbana.

### **4. Metodologia**

O projecto Angola Tourism Insight utilizará técnicas de aprendizado de máquina supervisionado e não supervisionado para analisar dados turísticos de Angola, com o objectivo de gerar previsões e insights estratégicos. A metodologia será dividida em três etapas principais: coleta de dados, modelagem preditiva e visualização interativa.

#### **4.1 Coleta e Pré-processamento de Dados**

Serão integradas fontes públicas e APIs abertas, como:

INE Angola, Ministério da Cultura e Turismo, Banco Mundial, Google Trends, OpenWeatherMap. Os dados serão tratados com Pandas para limpeza, normalização e estruturação.

Segundo Han et al. (2011), o pré-processamento é essencial para garantir a qualidade dos dados e a eficácia dos modelos de aprendizado de máquina.

## 4.2 Modelagem Preditiva com Machine Learning

A modelagem será realizada com algoritmos como:

- Random Forest: para previsão de fluxo turístico com base em variáveis como clima, feriados e eventos. É robusto contra *overfitting* e eficaz em dados heterogêneos (Breiman, 2001);
- K-Means Clustering: para segmentação de perfis turísticos e identificação de padrões sazonais;
- Regressão Linear e Regressão Ridge: para estimar impacto de variáveis socioeconômicas no turismo;
- Redes Neurais: para modelos mais complexos e não lineares, especialmente em séries temporais.

A escolha dos algoritmos será baseada em testes de desempenho com métricas como RMSE, MAE e  $R^2$ , conforme recomendado por Géron (2019).

## 4.3 Visualização e Dashboard Interativo

Os resultados serão apresentados em um *dashboard* interativo desenvolvido com Streamlit ou Power BI, permitindo a visualização em tempo real de métricas e previsões, filtros por região, período e tipo de turista, exportação de relatórios para gestores e investidores. A visualização de dados é fundamental para a tomada de decisão baseada em evidências (Few, 2009).

## 4.4 Justificativa da Abordagem

O uso de aprendizado de máquina é justificado pela complexidade e volume dos dados turísticos, que exigem técnicas avançadas para detectar padrões ocultos, prever comportamentos futuros e apoiar decisões estratégicas com base em evidências.

Essa abordagem promove inovação, eficiência e sustentabilidade, alinhando-se aos ODS da ONU e às necessidades do setor turístico angolano

## 5. Diagrama de Projecto de Arquitetura

O projecto de previsão de turistas utiliza uma arquitetura composta por cinco principais componentes: Coleta de dados, armazenamento, processamento de dados, modelo de machine learning e otimização para turismo sustentável. O objectivo é prever o fluxo de turistas e, com base nisso, otimizar estrategicamente os recursos de alojamento, transporte e saneamento, promovendo uma gestão eficiente e sustentável do turismo.

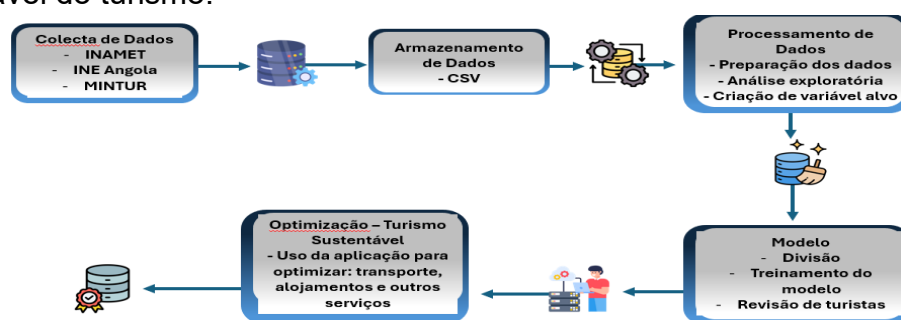


Figura 1: Diagrama funcional do projecto

## Componentes do Diagrama e Suas Funções

- **1.Coleta de Dados**
  - Função: Capturar dados relevantes como entradas em fronteiras, reservas de hotéis, eventos locais, clima, redes sociais e mobilidade urbana. Utilizada para fornecer as informações brutas essenciais para alimentar todo o sistema;
- **2. Armazenamento**
  - Função: Organizar e guardar os dados coletados em bancos de dados estruturados em CSV. Serve para garantir que os dados estejam acessíveis e prontos para análises e treinamentos futuros;
- **3.Processamento de Dados**
  - Função: Limpar, transformar e preparar os dados para o modelo de Machine Learning. Ela garante a qualidade e consistência dos dados, aplicando engenharia de atributos e normalização;
- **4.Modelo de Machine Learning**
  - Função: Utilizar algoritmos preditivos (como regressão ou redes neurais) para prever o número e o perfil dos turistas em períodos futuros. Utilizada para gerar previsões confiáveis que servirão de base para tomadas de decisão estratégicas;
- **5.Otimização – Turismo Sustentável**
  - Função: Usar as previsões para ajustar a alocação de recursos (capacidade hoteleira, transporte, coleta de lixo, etc.). Útil para promover decisões sustentáveis, reduzindo desperdícios, prevenindo sobrecargas e melhorando a experiência turística

## 6. Fontes de dados

O projecto utilizará duas principais categorias de dados: dados turísticos e dados meteorológicos históricos, ambos essenciais para a construção de modelos preditivos sobre o fluxo turístico em Angola.

### 1. Dados Turísticos

A principal fonte será o Anuário Estatístico do Turismo de Angola, publicado pelo Instituto Nacional de Estatística (INE Angola). Este documento, convertido para formato CSV, contém informações sobre:

- Número de turistas por ano;
- Países de origem;
- Distribuição por continente e tipo de visitante.

Esses dados são fundamentais para identificar padrões históricos e tendências de entrada de turistas no país. O pré-processamento incluirá limpeza de dados, padronização de datas, tratamento de valores ausentes e codificação categórica para uso em modelos de *machine learning*.

### 2. Dados Meteorológicos

Para dados climáticos históricos, serão utilizadas fontes confiáveis e gratuitas:

- Projecção da Precipitação DJF e JFM 2022/2023 – INAMET: possui mapas e análises por província sobre chuvas abaixo, dentro ou acima da média climatológica.

Esses dados serão utilizados para correlacionar condições climáticas com o fluxo turístico, permitindo prever sazonalidades e comportamentos. O pré-processamento incluirá interpolação de dados ausentes, normalização de variáveis e sincronização temporal com os dados turísticos.

## **7. Revisão de Literatura**

A aplicação de *machine learning* no turismo tem sido amplamente estudada nos últimos anos, com resultados promissores em diversas regiões do mundo. Por exemplo, Gretzel et al. (2015) destacam que o turismo inteligente depende da coleta e análise de grandes volumes de dados para melhorar a experiência do visitante e a eficiência dos serviços. Li et al. (2018) demonstraram que algoritmos como Random Forest e redes neurais são eficazes para prever o número de visitantes com base em variáveis como clima, feriados e eventos, o que é directamente aplicável ao contexto angolano. Já Zheng et al. (2020) reforçam que a visualização interativa dos dados por meio de dashboards facilita a tomada de decisão por gestores públicos e privados, promovendo transparência e agilidade.

Esses estudos mostram que a combinação de dados históricos turísticos e meteorológicos, aliada a técnicas de aprendizado de máquina, pode gerar modelos preditivos robustos e sistemas de apoio à decisão. O projecto Angola Tourism Insight se inspira nessas abordagens, mas propõe uma adaptação ao contexto angolano, onde há escassez de dados organizados e ferramentas analíticas. Ao integrar fontes locais (como o INE e o INAMET) com tecnologias abertas (como Python e Streamlit), o projecto busca democratizar o acesso à inteligência turística, promovendo inovação, sustentabilidade e alinhamento com os ODS.

# Plano de Implementação

## 1. Pilha de tecnologia

O projecto Angola Tourism Insight será implementado com uma pilha tecnológica leve, acessível e eficiente, voltada para análise de dados e desenvolvimento de soluções preditivas. As tecnologias e ferramentas escolhidas são:

### Linguagem de Programação

- Python: principal linguagem para manipulação de dados, desenvolvimento de modelos de Machine Learning e construção de visualizações interativas.

### Bibliotecas e Frameworks

- Pandas e NumPy: para leitura, limpeza e transformação de dados em formato CSV;
- Scikit-learn: para aplicação de algoritmos de aprendizado de máquina como Random Forest, K-Means e regressão;
- TensorFlow: para redes neurais e modelos mais avançados;
- Matplotlib e Plotly: para geração de gráficos e visualizações;
- Streamlit: para criação de dashboards interativos acessíveis via navegador.

### Fontes de Dados

- INE Angola e MINTUR: dados turísticos em CSV.
- INAMET, CHIRPS, Meteoblue, AClimate Angola: dados meteorológicos históricos em CSV.

### Outros Componentes

- Jupyter Notebook e Visual Code Studio: para prototipagem e documentação técnica;
- GitHub: para controle de versão e colaboração entre membros da equipe.

## 2. Principais Funcionalidades

### 2.1. Previsão de Turistas

Esta é a funcionalidade central da plataforma. Ela utiliza modelos de Machine Learning (como Random Forest, Regressão Linear, Redes Neurais) para prever o número de turistas em determinada localidade e período, com base em:

- Dados históricos de visitantes;
- Temperatura e precipitação;
- Eventos e feriados;
- Localidade e época do ano.

A previsão é acompanhada por indicadores visuais e analíticos que classificam o fluxo turístico em três níveis:

1. Pico Turístico: número previsto acima da média máxima histórica;
2. Médio: valor previsto entre a média mínima e máxima;
3. Baixo: número de visitantes abaixo da média mínima histórica.

Esses padrões são definidos por análise estatística regional, permitindo planeamento estratégico para cada província.

## **2.2. Sugestões Inteligentes**

Com base nas previsões, o sistema gera sugestões automáticas que orientam acções estratégicas:

- Sugestão de Pico: indica alto potencial turístico, sendo ideal para acções de marketing, eventos culturais e promoções;
- Sugestão Média: indica estabilidade, esta sugestão é propícia para manutenção de serviços e monitoramento contínuo;
- Sugestão Baixa: indica retração e recomenda essencialmente, promoções, revisão de preços ou actividades locais alternativas.

Essas sugestões estão relacionadas aos ODS, destacando como o turismo pode contribuir para o crescimento económico e emprego, comunidades sustentáveis e acção climática.

## **2.3. Gestão de Arquivos e Publicações**

A plataforma possui um módulo robusto para gestão documental e permite:

- Upload de documentos técnicos e oficiais (relatórios, estudos, pautas e dados estatísticos);
- Classificação por título, descrição, categoria e meta dados;
- Validação por administradores, com três estados possíveis:
  - a) Aprovado: publicado e visível ao público;
  - b) Pendente: aguardando revisão;
  - c) Arquivado: disponível apenas para consulta interna.

A plataforma reconhece automaticamente diversos tipos de ficheiros (PDF, DOCX, imagens e planilhas), com pré-visualização e download disponíveis.

## **2.4. Histórico e Meta dados**

Cada previsão e documento armazenado inclui informações complementares para garantir rastreabilidade e transparência:

- Data de criação e atualização;
- Nome do utilizador que realizou o upload;
- Tipo e tamanho do arquivo;
- Estado de aprovação (pendente, aprovado, arquivado).

Esses dados permitem auditorias, controle de versões e monitoramento de alterações, fortalecendo a governança digital da plataforma.

## **3. Linha do tempo**

Nesta secção abordaremos sobre as datas para o cumprimento e realização de tarefas, bem como, as tarefas distribuídas por cada membro. Para mais detalhes, acompanhe o cronograma apresentado em anexo para as diferentes etapas do seu projecto, desde coleta e pré-processamento de dados, desenvolvimento de modelo, treinamento e avaliação e implantação.

## 4. Marcos

O projecto Angola Tourism Insight já alcançou um marco fundamental: a coleta e tratamento dos dados turísticos e meteorológicos, que exigiu esforço significativo devido à dispersão, inconsistência e escassez de fontes organizadas. Os dados foram estruturados e convertidos para formato CSV, prontos para uso em modelagem preditiva. Esta etapa representa uma conquista crítica, pois estabelece a base técnica e analítica do projecto.

Os próximos marcos planeados incluem:

- Desenvolvimento dos modelos preditivos: implementação de algoritmos de *machine learning* para prever fluxos turísticos com base em variáveis como clima, eventos e sazonalidade;
- Validação dos modelos: obtenção de métricas de desempenho satisfatórias, garantindo confiabilidade nas previsões;
- Construção do dashboard interativo: visualização dos dados e previsões em tempo real, com filtros por região, período e tipo de visitante;
- Implantação e apresentação final: entrega do sistema funcional, com relatório técnico e pitch para *stakeholders* do sector turístico.

## 5. Desafios e mitigações

Durante o desenvolvimento do projecto, alguns desafios são esperados. Abaixo estão os principais riscos e as estratégias propostas para mitigá-los:

### Qualidade dos Dados

- Desafio: dados incompletos, inconsistentes ou com lacunas temporais.
- Mitigação: uso de técnicas de interpolação, imputação de valores ausentes e validação cruzada com múltiplas fontes (INE, MINTUR e INAMET).

### Desempenho dos Modelos

- Desafio: baixa precisão nas previsões devido à variabilidade dos dados.
- Mitigação: testes com diferentes algoritmos (Random Forest, regressão, redes neurais), ajuste de parâmetros e validação com métricas como RMSE, MAE e  $R^2$ .

### Restrições Técnicas

- Desafio: limitação de recursos computacionais para treinamento de modelos mais complexos.
- Mitigação: uso de bibliotecas optimizadas (Scikit-learn, TensorFlow), execução local com amostragem de dados e, se necessário, uso de plataformas gratuitas como Google Colab.

### Integração e Visualização

- Desafio: dificuldade na criação de dashboards acessíveis e responsivos.
- Mitigação: uso do Streamlit, que permite desenvolvimento rápido e visualização interativa com baixo custo técnico.

## **6. Considerações éticas**

A ética no uso de dados e inteligência artificial é um pilar essencial do projecto. Desde a fase inicial, foi adoptada uma postura responsável na coleta, tratamento e uso dos dados, priorizando fontes públicas e agregadas, como os anuários do INE, relatórios do MINTUR e dados climáticos do INAMET. Nenhum dado pessoal ou sensível foi utilizado, o que garante conformidade com os princípios de privacidade e proteção de dados.

Além disso, há uma preocupação activa com o viés algorítmico. Sendo que os modelos de Machine Learning podem, inadvertidamente, reproduzir desigualdades se forem treinados com dados desbalanceados ou mal representados. Para mitigar esse risco, o projecto prevê a análise crítica dos dados de entrada, testes com diferentes algoritmos e validação contínua dos resultados. A segmentação de perfis turísticos será feita com cautela, evitando estereótipos ou exclusões.

Outro ponto importante é o impacto social. O projecto visa beneficiar comunidades locais, gestores públicos e empreendedores do sector turístico, promovendo decisões baseadas em evidências e incentivando práticas sustentáveis. A transparência será garantida por meio de documentação acessível, visualizações abertas e comunicação clara dos objectivos e limitações do sistema.

Por fim, o projecto se compromete com os princípios de ética digital, inclusão, sustentabilidade e responsabilidade social, alinhando-se às diretrizes dos ODS e às boas práticas internacionais em ciência de dados.

## 7. Referências

- Breiman, L. (2001). *Random forests*. Machine Learning, 45(1), 5–32.
- Few, S. (2009). *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. Analytics Press.
- Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (2nd ed.). O'Reilly Media.
- Gretzel, U., Sigala, M., Xiang, Z., & Koo, C. (2015). *Smart tourism: foundations and developments*. Electronic Markets, 25(3), 179–188.
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
- Li, X., Pan, B., Zhang, L., & Smith, S. (2018). *Forecasting tourism demand with machine learning models*. Journal of Travel Research, 57(6), 818–832.
- Zheng, Y., Capra, L., Wolfson, O., & Yang, H. (2020). *Urban computing: concepts, methodologies, and applications*. ACM Transactions on Intelligent Systems and Technology, 5(3), 1–55.

## **Anexos**

Etapas	Dias da execução do projecto (meses de <u>Outubro</u> e <u>Novembro</u> )																														
	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	01	02	03	04	05	06	07	08	09	10	11	
Estabelecimento dos ODS relacionados																															
Identificação das fontes de dados																															
Recolha de dados turísticos																															
Recolha de dados meteorológicos																															
Limpeza e estruturação dos dados																															
Conversão para formato CSV																															
Verificação da qualidade e consistência dos dados																															
Análise exploratória dos dados																															
Escolha dos algoritmos																															
Implementação do modelo com dados tratados																															
Validação com métricas																															
Comparação entre modelos																															
Desenvolvimento com Streamlit																															
Testes de usabilidade																															
Preparação de material para apresentação																															
Apresentação final e entrega																															

Figura 2: Cronograma oficial do projecto