

**Elysson Alves**  
**Gustavo Almeida**  
**Kalmax dos Santos**

## Entrega 1

### 1. Abordar o Problema e Analisar o Panorama em Geral

#### **Objetivos em termos de negócios**

O objetivo é prever o preço de voos com base em dados históricos, permitindo que consumidores e empresas tomem decisões financeiras e operacionais mais informadas. Essa solução pode ajudar consumidores a economizar dinheiro ao identificar os melhores momentos para compra de passagens e auxiliar empresas no ajuste de estratégias de precificação e planejamento de demanda.

#### **Como a solução será usada**

A solução será utilizada por:

- **Consumidores:** Para prever os melhores momentos para adquirir passagens aéreas com base em padrões históricos.
- **Empresas (companhias aéreas e agências de viagens):** Para otimizar estratégias de precificação dinâmica e aumentar eficiência operacional.

Por exemplo:

- Consumidores podem programar compras de passagens antecipadas para evitar preços mais altos.
- Companhias podem ajustar as tarifas de voos com base em previsões de demanda e sazonalidade.

#### **Soluções/alternativas atuais**

- Monitoramento manual de preços por consumidores em aplicativos e sites de comparação.
- Modelos de precificação dinâmica utilizados por empresas com base em sazonalidade e regras fixas.
- Estudos estatísticos que consideram apenas médias e variâncias históricas, sem aplicar técnicas de predição avançadas.

#### **Como o problema deve ser abordado**

- **Tipo de aprendizado:** Aprendizado supervisionado para prever os preços com base em variáveis históricas e contextuais.
- **Modo de operação:** Offline inicialmente, com atualizações regulares no modelo conforme novos dados forem inseridos.

### Como o desempenho será medido

- **Métricas primárias:**
  - **Erro absoluto médio (MAE):** Para avaliar precisão das previsões em valores absolutos.
  - **Erro quadrático médio (RMSE):** Para penalizar erros maiores, particularmente indesejáveis, já que grandes diferenças entre os valores previstos e reais têm um impacto maior na métrica.
- **Métricas secundárias:**
  - **R<sup>2</sup>:** Para medir o ajuste do modelo aos dados, variando entre 0 e 1, onde 1 indica que o modelo explica a variabilidade do modelo.

### A medida de desempenho alinhada com o objetivo do negócio

Sim, as métricas como MAE e RMSE refletem diretamente a precisão do modelo e ajudam a prever com maior segurança os preços de voos, garantindo aplicações práticas, como recomendações de compra.

### Qual seria o desempenho mínimo necessário para alcançar o objetivo do negócio?

- Um **MAE** inferior à média das variações de preço em voos semelhantes, indicando que o modelo é mais eficaz do que estimativas simples como a média histórica.
- O **RMSE** deve ser significativamente menor do que a variância dos preços para garantir previsões mais precisas e úteis.
- Um **R<sup>2</sup>** acima de 0,7, o que indica que mais de 70% da variabilidade dos preços foi explicada pelo modelo.

---

## 2. Análise Comparativa e Planejamento

### Problemas que podem ser comparados

- Previsão de tarifas de transporte (como ônibus ou transporte por aplicativo), onde padrões históricos e variáveis contextuais também são relevantes.
- Previsão de demanda para reservas de hotéis, considerando sazonalidade e eventos locais.

### **Reutilização de experiências e ferramentas**

- Modelos de machine learning aplicados em previsão de demanda e precificação dinâmica podem ser adaptados.
- Estudos sobre sazonalidade e fatores que influenciam decisões de compra podem ser aproveitados para enriquecer o modelo.

### **Expertise Humana**

- Consultores em ciência de dados e especialistas no setor de aviação podem fornecer insights valiosos sobre padrões de preços e comportamento de consumidores.
- Colaboração com profissionais de marketing de companhias aéreas para entender as estratégias atuais de precificação.

### **Como o problema poderia ser resolvido de forma manual**

- Monitorar e analisar tendências históricas de preços, considerando sazonalidade, feriados e eventos especiais.
  - Utilizar métodos estatísticos simples para identificar padrões de variação e prever flutuações de preços.
- 

## **3. Discussões**

### **Suposições feitas:**

1. Dados históricos de criminalidade são confiáveis e completos, eles abordam as reais ocorrências dos anos que foram captadas.
2. O comportamento criminal é parcialmente previsível com base em padrões históricos.
3. As intervenções propostas podem impactar positivamente para os índices de criminalidade, visando reduzi-los.

4. A distribuição da criminalidade não é igualitária, existem crimes que quase não ocorrem em alguns casos temporais.
  5. Ser uma espécie de [Policimento Preditivo](#), onde é possível prever ocorrências criminais.
  6. Vamos tratar muito a parte de pré-processamento.
- 

## 4. Obter os Dados

- **Liste os dados de que você precisa e de quanto precisa:**  
Conseguimos os dados e listamos o que precisamos para o treinamento do modelo.
- **Documente onde você pode obter esses dados:**  
Repositórios como o Kaggle (ex.: [Flight Price Prediction Dataset](#)) e APIs de comparação de preços aéreos.
- **Verifique quanto espaço esses dados ocuparão:**  
Aproximadamente 500 MB ou menos.
- **Verifique as obrigações legais e obtenha autorização, se necessário:**  
Dados provenientes de fontes públicas ou com licença aberta. Garantir anonimização dos dados pessoais, se presentes.
- **Crie um workspace:**  
Escolhemos fazer localmente com capacidade suficiente para processar os dados.
- **Converta os dados em um formato manipulado com facilidade:**  
Arquivos CSV ou bases relacionais, com limpeza inicial para excluir inconsistências.
- **Amostre um conjunto de teste e deixe-o de lado:**  
Conjunto de 20% dos dados para avaliação final do modelo, evitando data snooping.

## Entrega 2

## 5. Explore os Dados

- Nota: procure um especialista de campo para estas etapas.
  - Crie uma cópia dos dados para exploração (amostragem até um tamanho gerenciável, se necessário).

- Crie um Jupyter notebook para manter um registro de sua exploração de dados.
- Estude cada atributo e suas propriedades:
  - Nome.
  - Tipo (categórico, int/float, bounded/unbounded, texto, estruturado etc.).
  - % de valores ausentes.
  - Ruído e tipo de ruído (estocástico, outliers, erros de arredondamento etc.).
  - Utilidade para a tarefa.
  - Tipo de distribuição (gaussiana, uniforme, logarítmica etc.).
- Para tarefas de aprendizado supervisionado, identifique o(s) atributo(s)-alvo.
- Visualize os dados.
- Estude as correlações entre os atributos.
- Estude como você resolveria o problema manualmente.
- Identifique as transformações promissoras que você pode querer aplicar.
- Identifique dados extras que seriam úteis (confira o Capítulo 2).
- Documente o que você aprendeu.

## 6. Prepare os Dados

- Notas:
  - Trabalhe em cópias dos dados (mantenha o conjunto de dados original intacto).
  - Escreva funções para todas as transformações de dados que você aplicar, por cinco motivos:
    - Desse modo, você consegue preparar facilmente os dados da próxima vez que obtiver um novo conjunto de dados.
    - Assim, você pode aplicar essas transformações em projetos futuros para:
    - Limpar e preparar o conjunto de teste.
    - Limpar e preparar novas instâncias de dados assim que sua solução estiver em produção.
    - Facilitar o tratamento de suas escolhas de preparação como hiperparâmetros.
- Limpeza de dados:
  - Corrige ou remove outliers (opcional).
  - Preenche os valores ausentes (por exemplo, com zero, média, mediana...) ou elimina suas linhas (ou colunas).
- Seleção de características (opcional):
  - Dropa os atributos que não fornecem informações úteis para a tarefa.

- Feature engineering, quando apropriado:
  - Discretize características contínuas.
  - Decompõe as características (por exemplo, categóricas, data/hora etc.).
  - Adiciona transformações promissoras de características (ex.:  $\log(x)$ ,  $\sqrt{x}$ ,  $x^2$  etc.).
  - Agrega características em novas características promissoras.
- Escalonamento de características
  - Padronizar e normalizar as características.