

## Organização da Semana 1 – EDA e Estruturação dos Dados

Para iniciarmos o projeto, disponibilizarei o dataset completo contendo os dados de 2020 a 2025 com os voos operados pela ANAC, já com a variável-alvo pré-definida.

A missão da equipe de Data Science nesta primeira semana será:

- compreender o comportamento dos dados,
- responder perguntas específicas,
- e consolidar um panorama estruturado para o início da modelagem.

Para garantir produtividade e independência dos membros, dividi os dados em cinco frentes, cada uma atribuída a um integrante da equipe. Dessa forma **cada membro deverá criar um notebook individual, executar as tarefas definidas para sua dimensão e produzir um gráfico e um insight que responda a pergunta da dimensão escolhida.**

---

## DS1 – Qualidade e Estrutura dos Dados (*Data Quality*)

### Tarefas

- Verificar tipos de dados de cada coluna
- Identificar valores ausentes e calcular a porcentagem
- Detectar valores inválidos, inconsistentes e outliers
- Criar uma tabela Data Quality Report com:
  - nome da coluna
  - tipo
  - % de nulos
  - problemas encontrados
  - possíveis correções
- Descrever quais colunas exigirão limpeza futura

**Principal questão a ser respondida:**

**? “Quais problemas de qualidade podem comprometer o desempenho do modelo?”**

---

## DS2 – Distribuições e Comportamento das Variáveis

### Tarefas

- Identificar as variáveis numéricas
- Criar histogramas para variáveis numéricas
- Criar boxplots para ver variações e outliers
- Comparar valores de atraso vs. pontualidade
- Observar padrões de concentração, assimetria e dispersão

**Principal questão a ser respondida:**

**?** “Quais variáveis apresentam padrões que ajudam a diferenciar voos atrasados de pontuais?”

---

### DS3 – Correlações e Relações Entre Variáveis

**Tarefas**

- Calcular correlação para variáveis numéricas
- Construir heatmap de correlação
- Identificar variáveis redundantes (alta colinearidade)
- Destacar variáveis com maior relação com atraso

**Principal questão a ser respondida:**

**?** “Quais variáveis têm maior potencial para explicar os atrasos?”

---

### DS4 – Sazonalidade (Tempo: Mês, Dia, Horário)

**Tarefas**

- Calcular atraso por mês do ano
- Atraso por dia da semana
- Atraso por horário do dia
- Criar gráficos de linha para mostrar tendências temporais
- Identificar picos de demanda e sazonalidade

**Principal questão a ser respondida:**

**?** “Em quais períodos do ano, mês ou horários os atrasos são mais frequentes e por quê?”

---

### DS5 – Segmentação (Companhias e Aeroportos)

**Tarefas**

- Ranking das companhias com maior taxa de atraso
- Ranking dos aeroportos com mais atrasos (origem/destino)
- Comparar desempenho entre diferentes regiões do país
- Identificar gargalos e padrões específicos em companhias e aeroportos

**Principal questão a ser respondida:**

**?** “Quais companhias e aeroportos mais contribuem para os atrasos, eles apresentam padrões?”

---

Observem que, para esta primeira semana, organizei as atividades em cinco dimensões diferentes, cada uma com tarefas bem definidas. Essas frentes de pesquisa não foram atribuídas individualmente, porque a ideia é que cada pessoa escolha aquela com a qual mais se identifica, seja por afinidade, curiosidade ou por achar que faz mais sentido começar por ali.

Mas, para que tudo funcione bem e ninguém acabe fazendo a mesma coisa, é importante que cada integrante avise a equipe e a liderança qual dimensão pretende assumir antes de começar. Depois que todos comunicarem suas escolhas, seguimos com o cronograma da semana, garantindo que todas as dimensões sejam cobertas e que o trabalho avance de forma organizado e colaborativo.

---

## Cronograma da Semana 1

### Segunda-feira — 15/12

#### Reunião de alinhamento inicial

- Escolha das dimensões e ajustes nas tarefas, se necessário
- Combinação dos critérios de padronização dos notebooks
- Dúvidas técnicas

### Quinta-feira — 18/12

#### Apresentação dos achados individuais

- Cada membro apresenta seu notebook
- Discussão coletiva dos padrões encontrados
- Identificação de features potenciais

### Sexta-feira — 19/12

#### Consolidação e documentação

- Unificação dos insights
- Padronização das variáveis relevantes
- Criação da documentação formal para entrega na plataforma

### Observação

Fiz o planejamento somente das sprints que são obrigatórias, mas podemos fazer outras durante a semana, de acordo com a necessidade da equipe.