Apresentação: Análise de Dados - Clustering, Regras de Associação e Regressão Logística

# 1. Análise de Clustering

a. Existe um grupo incomum de passageiros que não se enquadra no perfil típico de cliente do aeroporto?

Resposta: Sim, o algoritmo K-Means identifica três clusters, e um deles tem um perfil incomum de clientes.

i. Qual é o tamanho do cluster em percentagem aos passageiros do aeroporto e qual é o perfil do grupo?

Tamanho: O cluster incomum representa cerca de 6.19% dos passageiros.

Perfil: Esse grupo possui uma satisfação mais baixa (NETPRO), menor experiência em voos (Q5TIMESFLOWN) e menor tempo de uso do aeroporto (Q6LONGUSE).

# 2. Regras de Associação

a. Problema:

Resposta: Identificar padrões de comportamento de clientes no aeroporto, como a relação entre a satisfação (NETPRO) e o tempo de uso do aeroporto (Q6LONGUSE).

b. Dados utilizado para modelagem do problema (dataset):

Resposta: Utilizamos o dataset "sfo\_2018\_data\_file\_final\_Weightedv2", com variáveis como NETPRO, Q6LONGUSE, Q23FLY.

c. Passos utilizados para geração de regras:

Resposta: Binarização dos dados, aplicação do algoritmo FP-Growth para encontrar itemsets frequentes, e extração de regras de associação com base no lift.

d. Regras Geradas:

Resposta: Exemplo de regras geradas:  
 - Se um cliente tem NETPRO > 8, então tem maior probabilidade de ser usuário de longo prazo.  
 - Clientes com frequência de voos alta (Q23FLY > 5) têm maior probabilidade de ser satisfeitos (NETPRO > 8).

# 3. Regressão Logística

a. Problema:

Resposta: Prever se um cliente está satisfeito com o aeroporto com base em variáveis demográficas e comportamentais.

b. Dados utilizado para modelagem do problema (dataset):

Resposta: Dataset "sfo\_2018\_data\_file\_final\_Weightedv2", com variáveis como Q20Age, Q21Gender, Q22Income, Q23FLY, Q5TIMESFLOWN, Q6LONGUSE.

c. Treinamento do modelo:

Resposta: Utilizamos Regressão Logística para treinar o modelo, com validação cruzada e divisão dos dados em treino (75%) e teste (25%).

d. Resultado do modelo:

Resposta: O modelo obteve uma precisão de 55% e um ROC AUC de 0.55, com coeficientes que indicam que variáveis como idade e tempo de uso do aeroporto têm impacto sobre a satisfação.