

Bem vindo à bordo!

Teste Especialista de Dados





Sobre o teste

Esse teste faz parte da etapa de avaliação técnica do processo seletivo para a vaga no time Analytics da GOL. O objetivo é avaliar como você desenvolve uma solução completa para desafios que encontramos no nosso dia a dia de trabalho.

O teste está dividido em dois desafios

- 1 Pontualidade GOL
- 2 Compra de novas aeronaves





1 Pontualidade GOL - Desafio

A pontualidade é um indicador fundamental para a GOL e demais companhias aéreas. Entre fatores como segurança e conforto, os passageiros dão grande importância ao fato de chegar ao seu destino na hora marcada, seja para encontrar a família, realizar reuniões executivas ou qualquer outro tipo de compromisso que não possa contemplar algum atraso.

As companhias aéreas de todo o mundo realizam diversos estudos, comitês, implementam planos de ação, que envolvem diferentes áreas da companhia, com o intuito de garantir ao máximo as metas de pontualidade na aviação.

Após o último Comitê de Operações da GOL o Diretor de Analytics te procurou e pediu ajuda para resolver o problema, em **duas etapas**:

Etapas 1 - Realizar uma **análise exploratória** consistente e **avaliar** como a GOL estava posicionada frente aos concorrentes em relação à **pontualidade (mês a mês) no ano de 2018**.

Etapas 2 - **Implementar um modelo de classificação** para prever se aquele voo específico da base irá atrasar ou não (utilizando apenas dados do ano de 2018).



1 Pontualidade GOL – Informações relevantes

Base de dados

Para este desafio vamos utilizar a [base de dados](#) aberta da Agência Nacional de Aviação Civil (ANAC), que contempla todos os voos regulares ativos (VRA) das companhias aéreas, para o ano de 2018.

Clique [aqui](#) ter acesso aos metadados da base.

Premissas

- O estudo deve conter apenas dados **da GOL, Azul (AZU) e LATAM (TAM)**. As demais companhias aéreas podem ser desconsideradas neste momento.
- Considerar apenas **voos nacionais**.
- A coluna “**Código Justificativa**” pode ser desconsiderada no estudo.
- Para classificar um voo como **pontual ou atrasado**, leva-se em conta apenas os dados de “**Partida Prevista somando 15 minuto**” e “**Partida Real**”, sendo:

Partida Real > Partida Prevista + 15 minutos → Voo atrasado

Partida Real <= Partida Prevista + 15 minutos → Voo pontual

- Indicador de **pontualidade**:

$$1 - \frac{\text{voos atrasados}}{\text{voos realizados}}$$





1 Pontualidade GOL - Observações

Etapa 1 – Análise Exploratória

Esta é uma das mais importantes em qualquer trabalho de um Cientista de Dados. Antes de desenvolver qualquer modelo, é necessário entender primeiramente o nosso problema. Nesta etapa, queremos que você indique as principais características presentes nos dados das aéreas selecionadas (GOL, Azul e Latam). Fique à vontade para escolher qualquer ferramenta nesta etapa.

Recomendamos que você faça um levantamento de hipóteses que você gostaria de avaliar a partir desta análise para ajudar na construção da “história” do problema que você quer resolver.

A criatividade também é um fator de muita relevância nesta etapa! Não é necessário se limitar às informações presentes nesta base.

Uma análise exploratória bem estruturada será fundamental para o desenvolvimento da próxima etapa.

Etapa 2 – Implementação de um modelo de classificação para prever atraso do voo

Não existe um algoritmo ou técnica específica para resolver este problema. Nosso interesse aqui é que toda sua implementação esteja bem documentada, decisões que foram tomadas e aquelas que podem ter dado errado ao longo do processo.

Gostaríamos que você fizesse uma avaliação de performance do modelo, mas não se preocupe em encontrar o melhor algoritmo. O principal objetivo nesta etapa é entender toda a linha de raciocínio alinhada aos conceitos teóricos.



O próximo desafio envolve conhecimentos básicos para formulação de problemas relacionados à otimização.

A área de Frotas da GOL está fazendo um estudo de viabilidade para compra de aviões a jato de passageiros para **viagens de curtas, médias e longas distâncias**, denominados **Ac, Am e AI**, respectivamente.

Sabe-se que os **custos unitários** destas aeronaves, **em milhões de euros, são 2.100, 4.000 e 5.200, respectivamente**. O conselho da companhia aprovou um **orçamento de, no máximo, 120.000 milhões de euros** para realização desta compra.

Admite-se que os **lucros anuais** com cada um dos tipos **de avião Ac, Am e AI**, sejam de **190, 220 e 310 milhões de euros, respectivamente**.

A GOL atualmente apresenta um quadro de comandantes para pilotar, **no máximo, 25 aviões novos**.

Realizando a compra apenas de **aeronaves Ac**, os serviços de manutenção seriam capazes de garantir a **manutenção de 36 aeronaves novas**. Entretanto, do ponto de vista de demanda de trabalho da manutenção, **cada avião Am equivale a 4/3 de um avião Ac e cada avião AI a 5/3 de um avião Ac**.

A diretoria de Planejamento de Malha recomenda que, **para cada aeronave Ac comprada, se comprem também pelo menos uma aeronave AI ou uma aeronave Am**.

Por outro lado, para manter as condições de preços dos aviões, o fornecedor exige que, **selecionando a compra de um avião AI, também deverão ser comprados pelo menos 7 aviões Ac ou Am**.

A partir destes dados, o Diretor de Frotas solicitou sua ajuda para tomar a decisão sobre a quantidade de aviões de cada tipo comprar, de modo a **maximar o lucro da GOL**.

Formule este problema com **um modelo de programação linear**.

* Nosso interesse aqui é que você apenas desenvolva as fórmulas da função objetivo e restrições do problema. Não é necessário resolvê-lo.



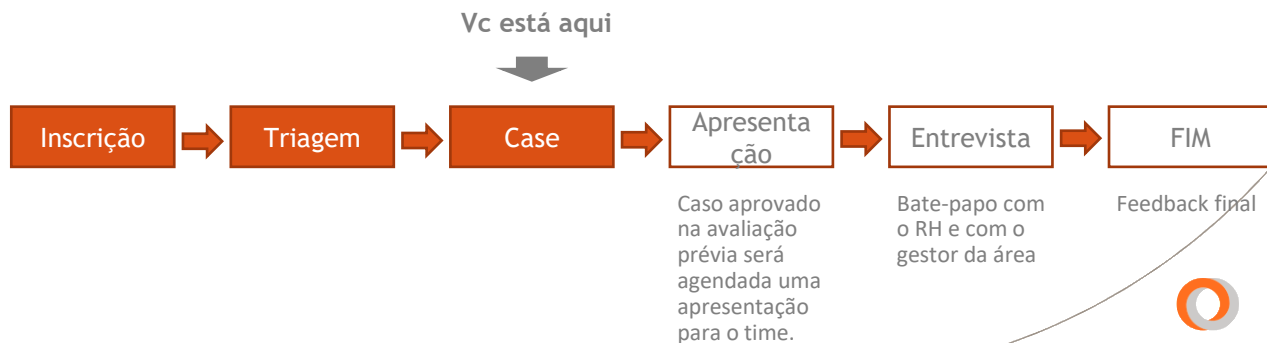
Considerações Finais

Você deve enviar o teste completo em apenas **um Jupyter Notebook** por e-mail. Lembre-se de documentar seu código e utilizar **células Markdown** para explicar detalhadamente sua solução. Lembre-se de **explicar seu raciocínio e justificar os métodos utilizados por você**. Explícite o(s) algoritmo(s) utilizado(s) e as etapas de pré-processamento que você recomenda fazer, justificando o porquê de cada uma das decisões tomadas.

Não queremos a solução ideal para o problema! Queremos entender sua **forma de pensar** 😊

Seja criativo na resolução do problema! O trabalho de um Cientista de Dados envolve conhecimento técnico, metodologia científica e muita criatividade para abordar problemas complexos. **Procure formular uma hipótese, crie o algoritmo de categorização de produtos e encontre uma métrica que teste a sua hipótese.**

Boa sorte e nos vemos na apresentação!!



GOL