

ifood





O **Data Bricks** foi escolhido como a ferramenta a ser utilizada, devido a:

- **Sua praticidade** em carregar os dados e trata-los como **datasets** a serem acessados pelas **queries**
- Possibilidade de **utilização do SQL de maneira nativa** e “pura”
- **Construção da solução no formato de jupyter notebook**, para melhor organização das soluções



- O código foi enviado no formato .dbc e .ipynb para maior facilidade no acesso e conferência dos resultados
- A solução para cada questão está indicada e separada com células de *markdown*
- Dentro da solução, existem **células adicionais** do notebook voltadas à **extrapolação da pergunta** indicada no enunciado, indicadas por “extra info”
- **Foram realizadas sanitizações nas bases, corrigindo colunas** com inputs inválidos, **adequando formatos** de valores e **filtrando erros** de sistema para não poluir a análise

O serviço A apresenta o maior número de falhas, tanto em termos absolutos quanto percentuais, apesar disso é o mais confiável por ter a menor taxa de falhas

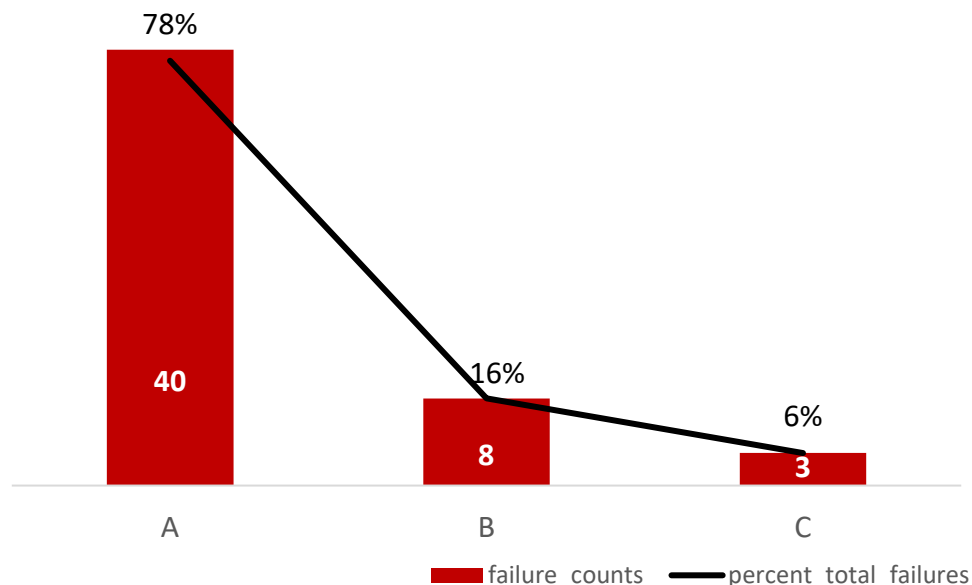


Questão A - Qual dos serviços mais falha? (status = PROVIDER_FAILED)

Serviço A é o serviço com mais falhas absolutas (40) e importância nas falhas totais (78%)

Importância de Falhas por Provider

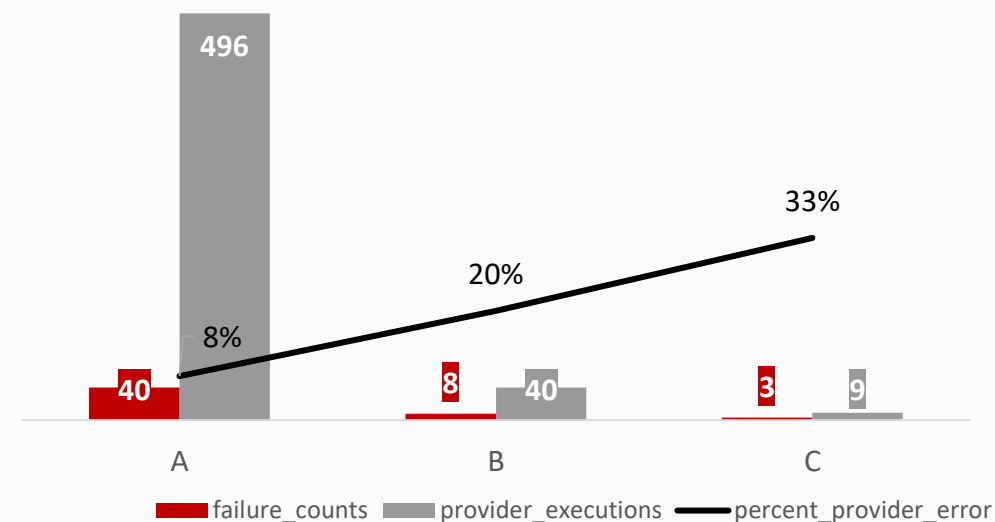
#Entregadores, #Falhas, #Índice de falha por categoria



De forma relativa à quantidade de execuções, o Provider C apresenta a maior taxa de falhas

Taxa de Falhas por Provider

#Execuções, #Falhas, #Taxa de falha por Provider



A categoria Bronze é possui maior índice e volume de falhas, podendo indicar que muitas dessas contas são destinadas para fins de empréstimo



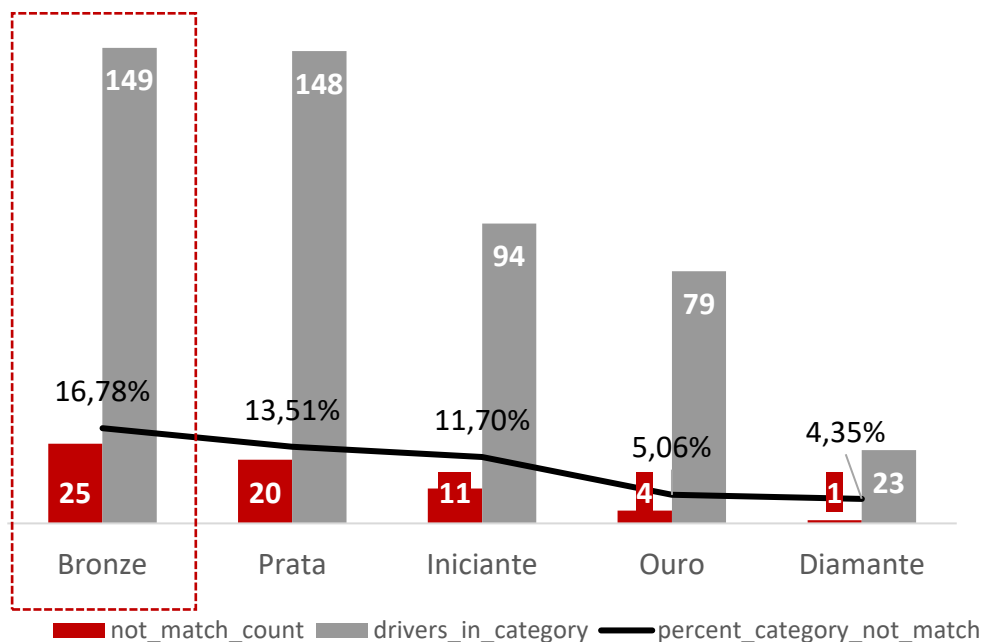
Questão B - Qual categoria de entregador possui o maior índice falha na identificação biométrica?

Ambas análises indicam que a **Categoria Bronze possui os piores números de identificação biométrica:**

- Maior índice de falha biométrica – 17%
- Maior importância das falhas totais – 40%, e portanto, maior quantidade de falhas

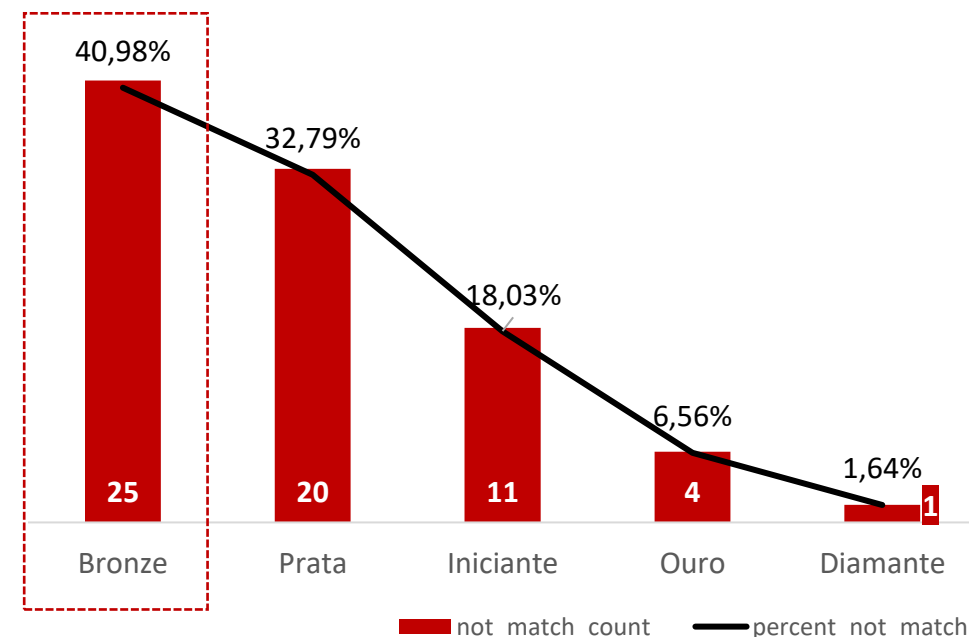
Índice de Falhas por categoria

#Entregadores, #Falhas, #Índice de falha por categoria



Importância de falhas por categoria

#Falhas, %Importância do total de falhas



O Índice geral de MATCH perderia 20 p.p com o aumento do *threshold* para 0.9, logo o valor atual traz mais equilíbrio entre segurança e taxa de aprovação

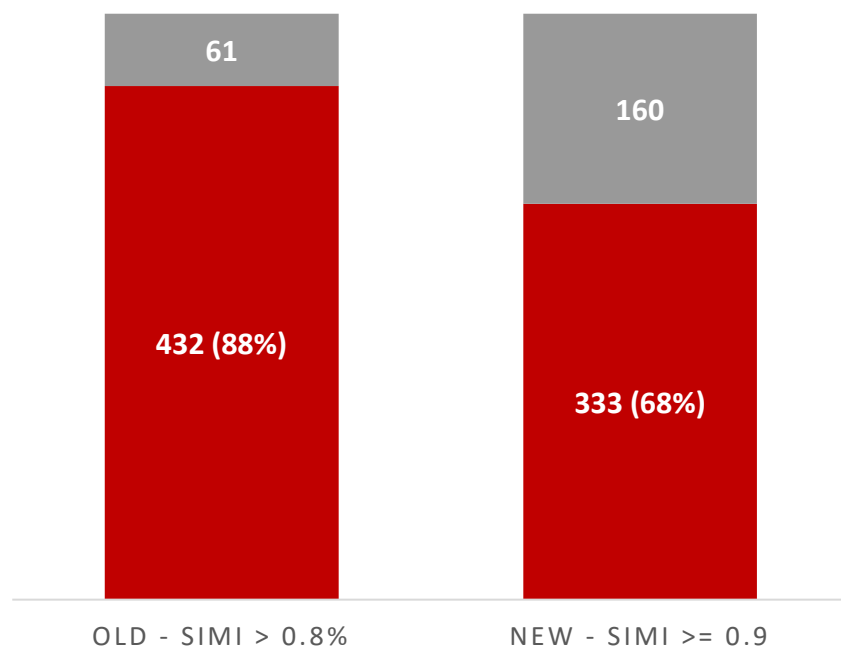


Questão C – Calcule quanto o índice geral de MATCH seria se aumentássemos a similaridade mínima do MATCH para 0.90.

Provider B sofre com maior queda em % de match por execuções – cerca de 32 p.p

Match – similarity 80% vs 90%

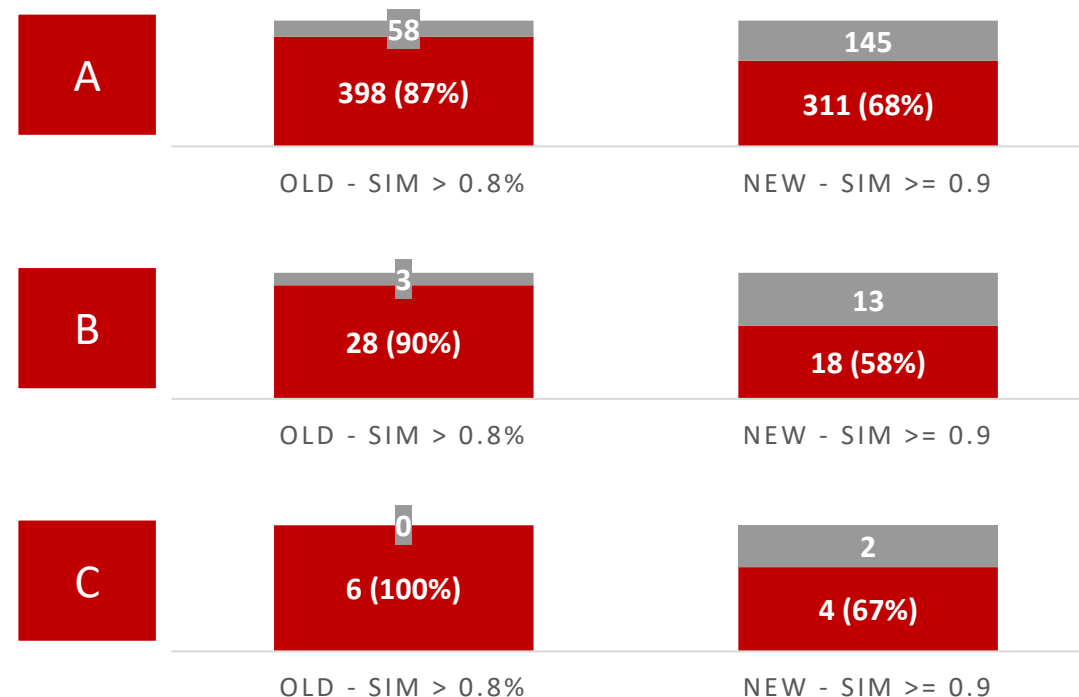
#Match # Not Match #%Match



MATCH NOT MATCH

Match – similarity 80% vs 90% por Provider

#Match # Not Match #%Match



Não há relação direta entre status de biometria de um entregador e seu volume de pedidos cancelados, porém aqueles com falha biométrica recebem menos pedidos



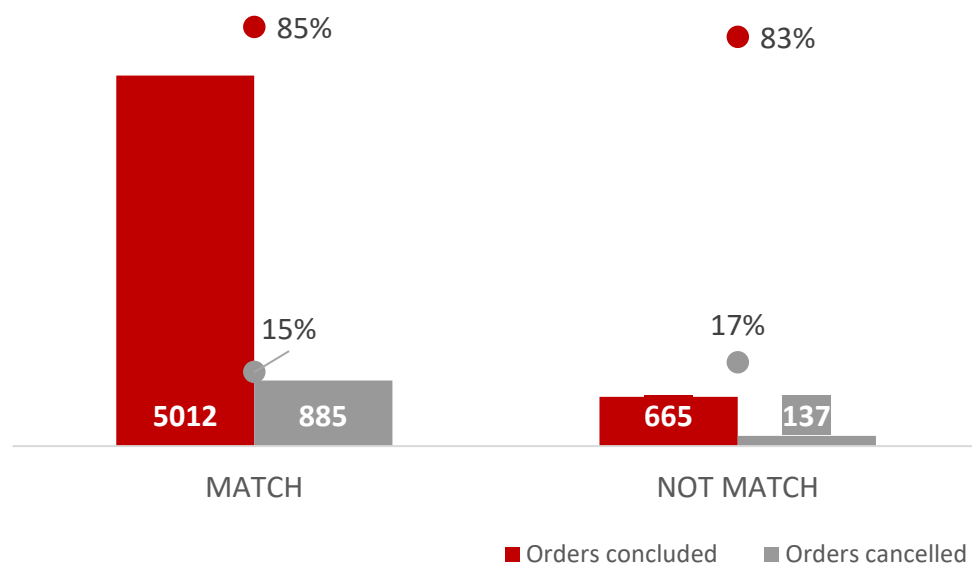
Questão D - Há alguma relação entre volume de pedidos cancelados de um entregador e o status final do processo de identificação biométrica dele?

Todas as verificações abaixo indicam que **não é possível apontar relação direta entre o volume de pedidos cancelados de um entregador com seu resultado biométrico.**

- Percentual de cancelados de match e not match se mantém próximo, com singela variação de 2 p.p
- Média de pedidos cancelados por entregador permanece similar próximo 2 pedidos por entregador

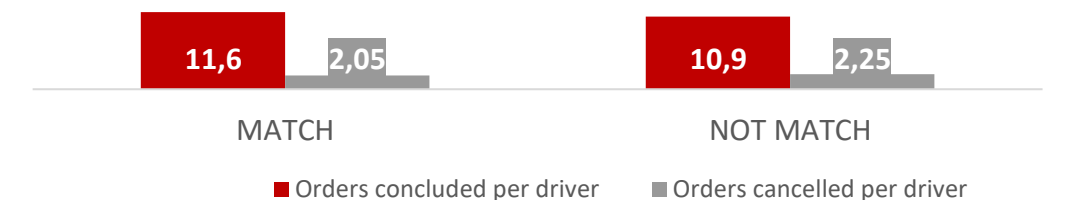
Pedidos por status de biometria do entregador

e % de pedidos, por status e por biometria



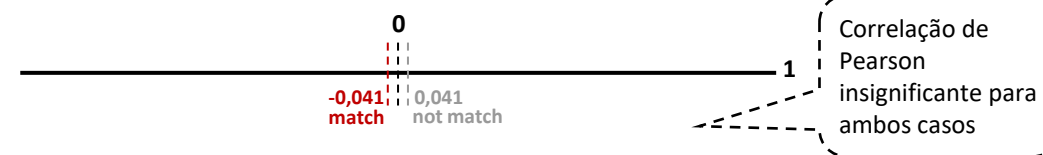
Média de pedidos por entregador por biometria

#Pedidos por status do pedido e por biometria (média por entregador)



Correlação de Pearson – pedidos e biometria

Correlação entre #pedidos cancelados e dummy de biometria



Os dias iniciais da semana indicam um possível aumento no empréstimo de conta, possivelmente pois o usuário principal tem outra atividade remunerada



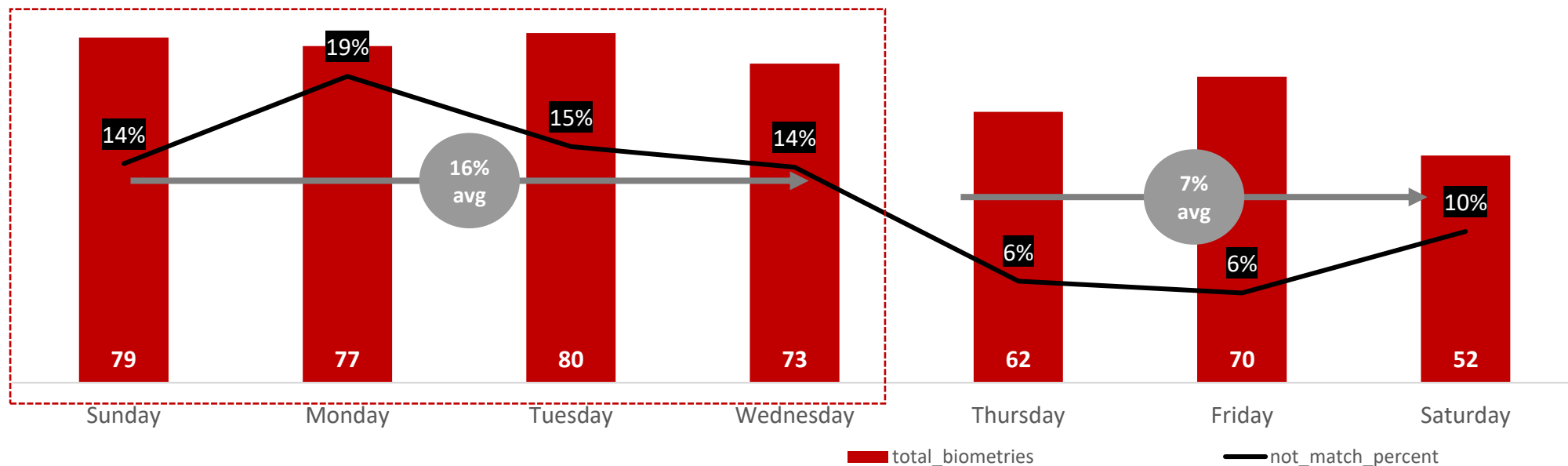
Em quais dias provavelmente tivemos um aumento do comportamento indevido “empréstimo de conta”?

O início da semana (domingo até quarta-feira) se mostraram como intervalo com suspeita de aumento do comportamento de empréstimo da conta, pois:

- Apresentam uma média de falhas na biometria bem superior aos dias do fim da semana
- Padrão se repete durante todas as semanas, não sendo comportamento exclusivo de datas específicas

Total de biometrias realizadas e percentual de match

#Biometrias %Not match, por dia da semana



OBRIGADO,



Cristiano Araújo Abreu Diniz
<https://www.linkedin.com/in/cristianoaadiniz>