



Case: Biometrias



Qualidade Logística e Prevenção a Fraude





Contexto



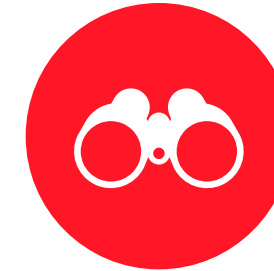
A presente análise, baseada em dados relacionados a identificação biométrica de entregadores em agosto de 2021, pretende avaliar aspectos como falhas e inconsistências no processo, com o intuito de identificar e prevenir fraudes.



Aspectos Analisados



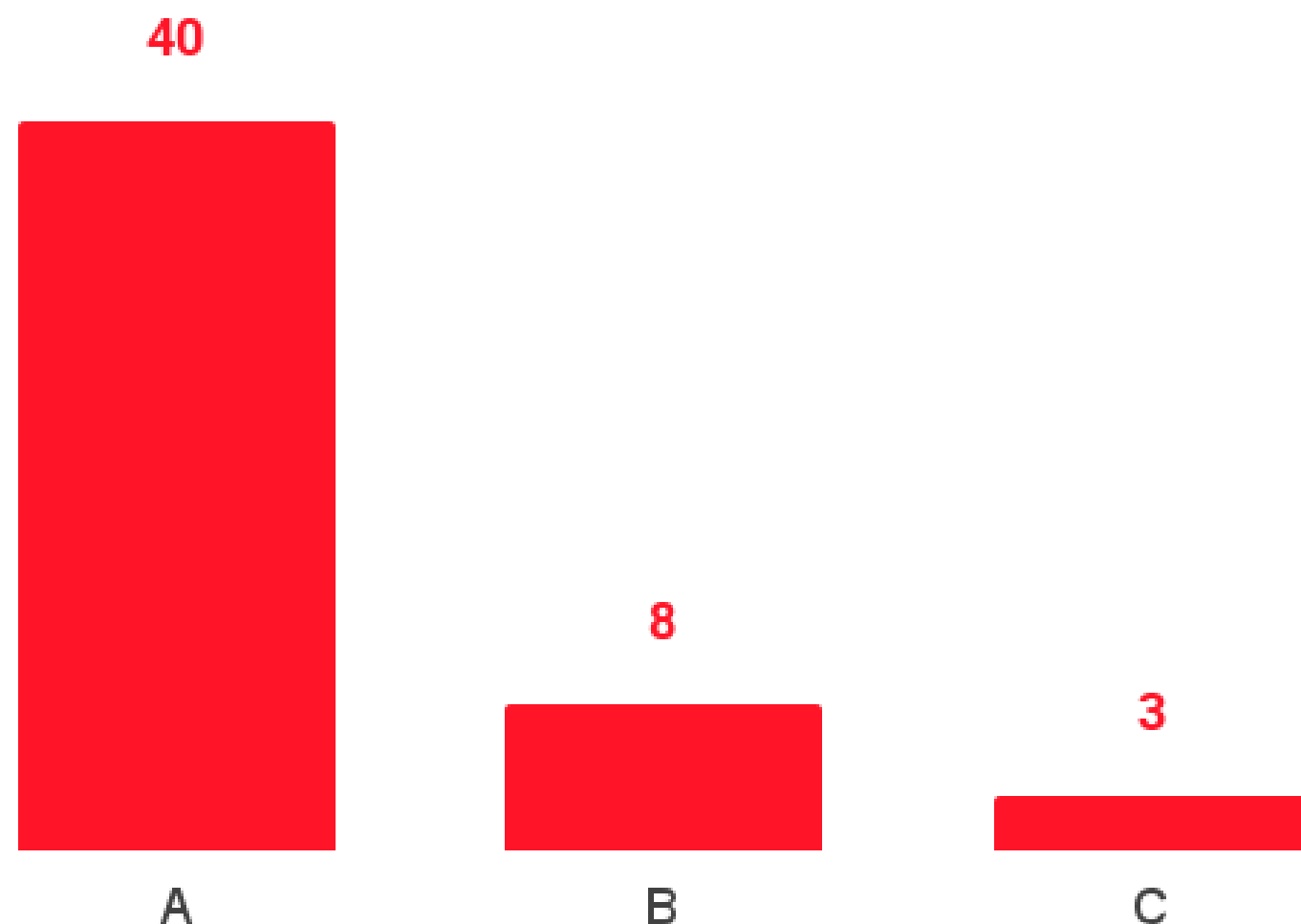
- Performance dos serviços no processo de identificação biométrica
- Índice de falha do processo de identificação biométrica por categoria de entregador
- Impacto do aumento da similaridade no índice de “MATCH”
- Relação entre cancelamentos de pedidos e status final do processo de identificação biométrica
- Análise de Comportamento Indevido - Empréstimo de Conta



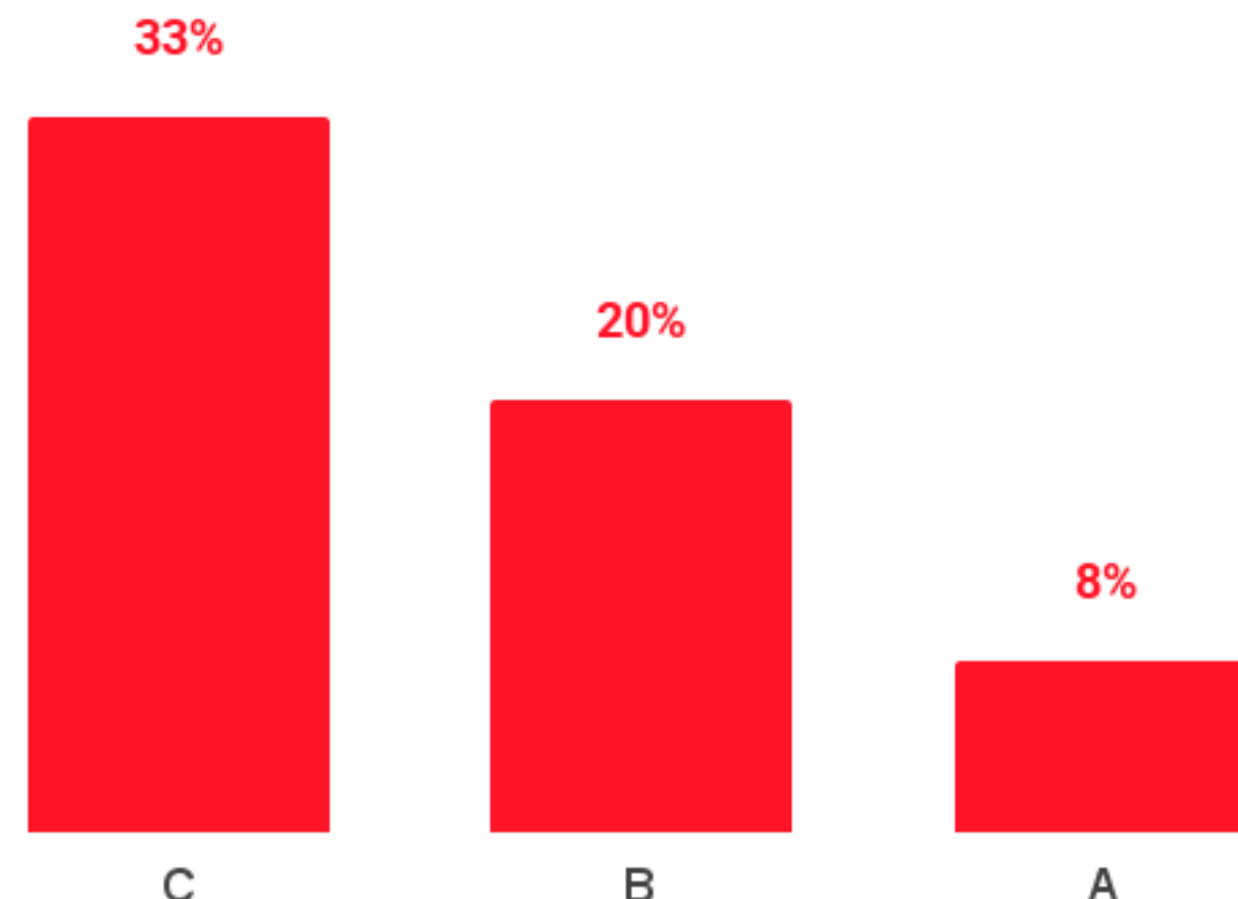


Performance dos serviços no processo de identificação biométrica

Volume de Falhas por Serviço



Taxa de Falhas por Serviço



A análise da performance dos serviços no processo de identificação biométrica revela que em relação à volumetria de falhas, **o serviço A possui maior quantidade de falhas**, representando **78,4% do total de falhas** identificadas durante o processo. Por outro lado, ao analisar o percentual de falhas, nota-se que **o serviço C apresenta maior taxa de falhas em relação ao total de sessões para esse serviço**. Dessa forma, o serviço C mesmo tendo uma volumetria menor de erros apresenta uma baixa performance ao analisarmos a taxa de falhas. Portanto, o alto volume de falhas no serviço A e a alta taxa de falhas no serviço C geram um ponto de atenção em relação à **vulnerabilidade do sistema e processo** e **impacto na identificação de fraudes**.



Query - Pergunta A

Query para Resposta Direta - Volumetria:

```
SELECT
  Provider,
  COUNT(DISTINCT Session_ID) AS Falhas
FROM biometry_ex
WHERE Status = 'PROVIDER_FAILED'
GROUP BY Provider
ORDER BY Falhas DESC
LIMIT 1
```

Resultado: A (40 falhas)

Query para o Gráfico - Volumetria:

```
SELECT
  Provider,
  COUNT(DISTINCT Session_ID) AS Falhas
FROM biometry_ex
WHERE Status = 'PROVIDER_FAILED'
GROUP BY Provider
ORDER BY Falhas DESC
-- LIMIT 1
```

Resultado: Todos os serviços e quantidades de falhas apresentados no gráfico

Query - Taxa de falhas:

```
WITH agrupado AS (
  SELECT
    Provider,
    SUM(CASE WHEN Status = 'PROVIDER_FAILED' THEN 1 ELSE 0 END) AS
    Falhas,
    COUNT(DISTINCT Session_ID) AS Total_Sessoes
  FROM biometry_ex
  GROUP BY Provider
)
SELECT *,
  ROUND(Falhas * 100 / Total_Sessoes, 1) AS Taxa_Falhas
FROM agrupado
ORDER BY Taxa_Falhas DESC;
```



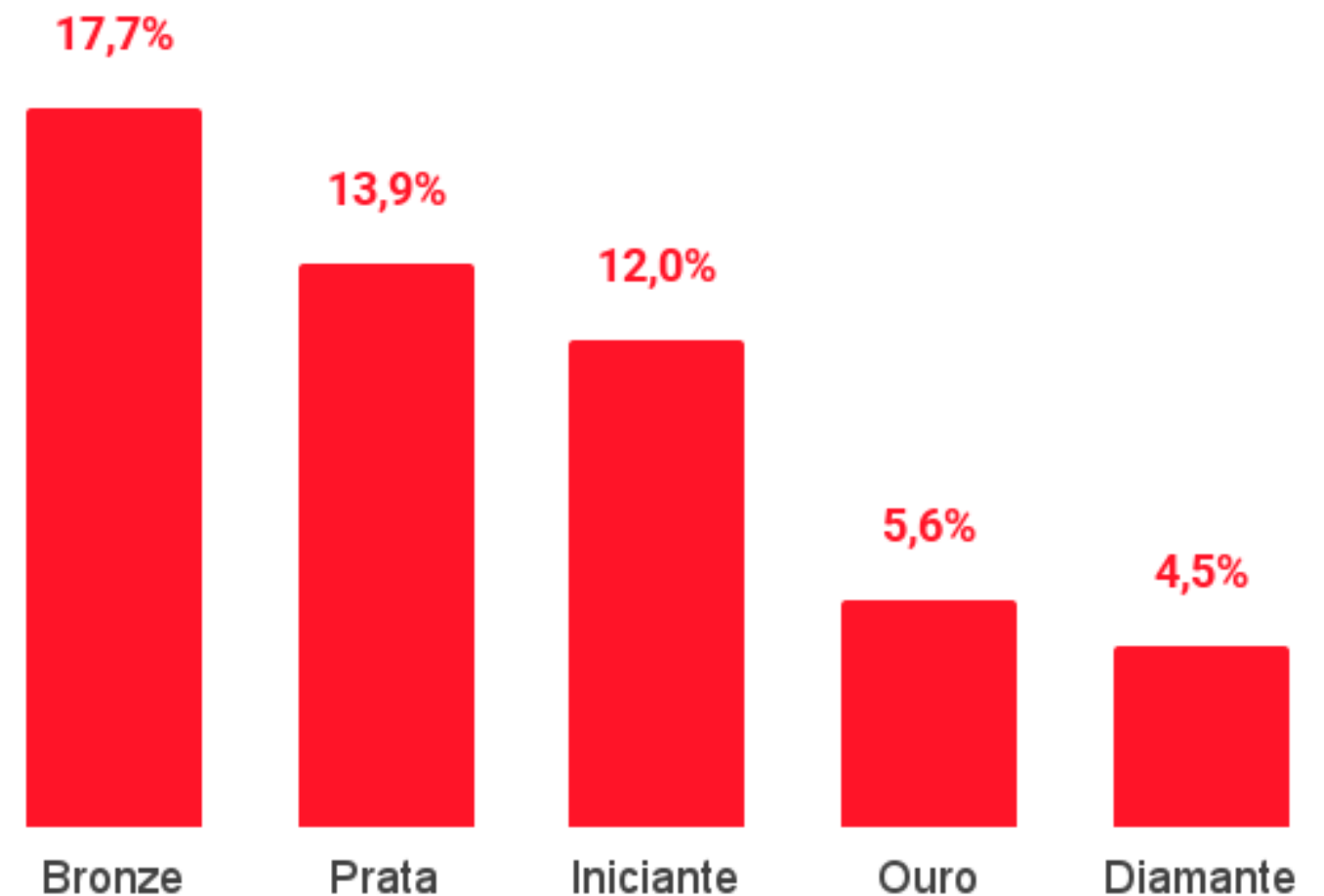


Índice de falha do processo de identificação biométrica por categoria de entregador

Volume de “NOT_MATCH” por Categoria de Entregador



Taxa de falhas (“NOT_MATCH”) por Categoria de Entregador



A análise do índice de “NOT_MATCH” por categoria de entregador evidencia que a categoria **Bronze possui a maior quantidade de falhas (NOT_MATCH)** e também o maior percentual de status “NOT_MATCH” em relação ao total de sessões para essa categoria. Dessa forma, a categoria B representa o grupo que oferece **mais riscos de fraude** e **demandas maiores de atenção e monitoramento**.





Query - Pergunta B

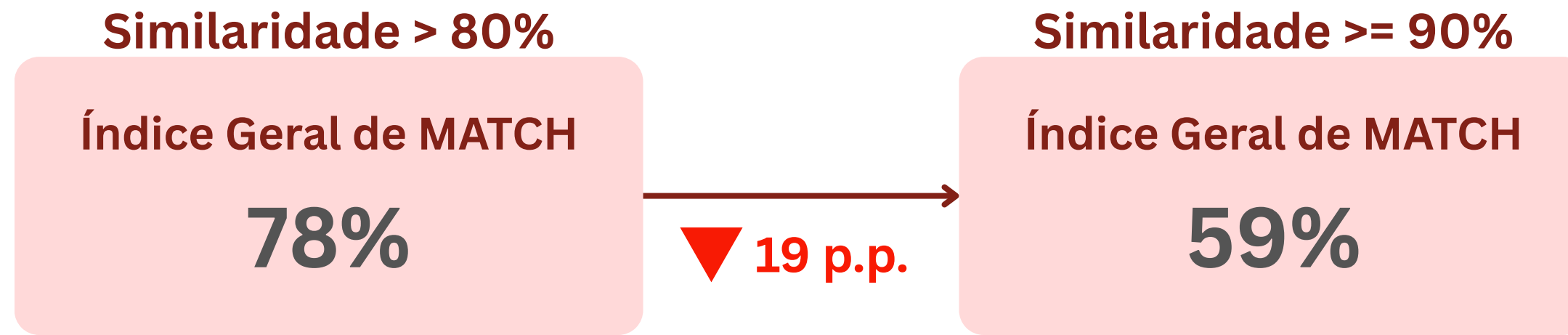
Query Volumetria e Percentual:

```
WITH agrupado AS (  
  SELECT  
    d.Category,  
    SUM(CASE WHEN b.Status = 'NOT_MATCH' THEN 1 ELSE 0 END) AS Falhas,  
    COUNT(DISTINCT b.Session_ID) AS Total_Sesses  
  FROM biometry AS b  
  LEFT JOIN drivers AS d  
    ON b.Driver_ID = d.Driver_ID  
  GROUP BY d.Category  
)  
SELECT  
  *,  
  ROUND(Falhas * 100.0 / Total_Sesses, 1) AS Taxa_Falhas  
FROM agrupado  
ORDER BY Taxa_Falhas DESC;
```





Impacto do aumento da similaridade no índice de “MATCH”



O aumento da similaridade no processo de identificação de biometria representa um **forte mecanismo para redução de fraudes**, porém, ao analisarmos os impactos desse aumento, nota-se uma **redução de 19 pontos percentuais no índice geral de MATCH** com a alteração da similaridade mínima para 90%. Essa redução de 19 p.p. representa uma melhoria no processo ao selecionar somente **casos com um maior grau de confiança**, no entanto, também representa um ponto de atenção ao **descartar usuários que não representam risco de fraude** e poderiam atuar normalmente, gerando a **insatisfação** e **baixa na fidelização** de entregadores.





Query - Pergunta C

Query:

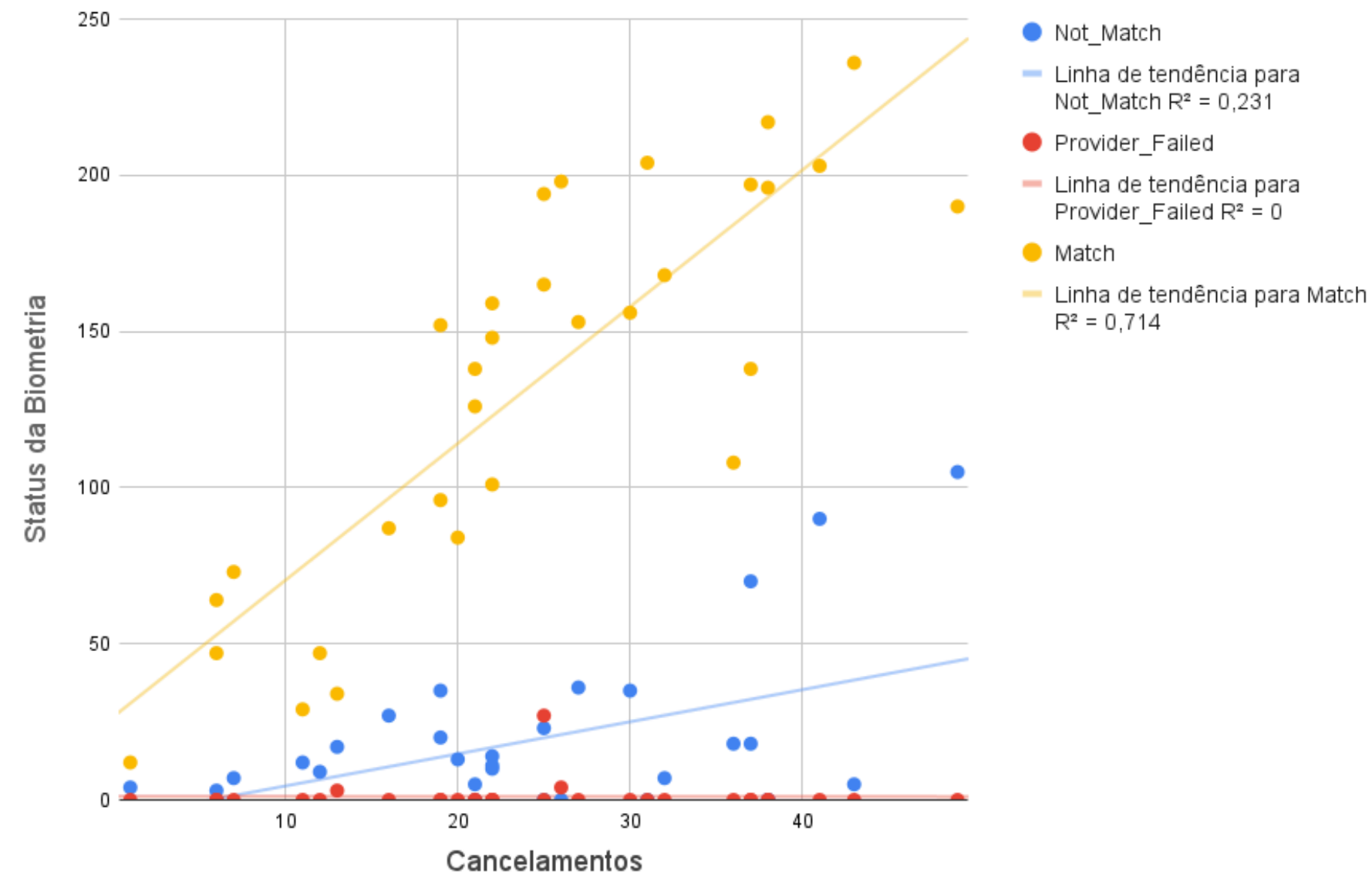
```
WITH status_new AS (  
  SELECT  
    Session_ID,  
    Status,  
    CASE  
      WHEN Similarity >= 0.9 THEN 'MATCH'  
      WHEN Similarity < 0.9 THEN 'NOT_MATCH'  
      WHEN Status = 'PROVIDER_FAILED' THEN 'PROVIDER_FAILED'  
    END AS Status_new  
  FROM biometry_ex  
)  
SELECT  
  ROUND(SUM(CASE WHEN Status = 'MATCH' THEN 1 ELSE 0 END) * 100 / COUNT(Session_ID),1) AS Indice_Match_Status,  
  ROUND(SUM(CASE WHEN Status_new = 'MATCH' THEN 1 ELSE 0 END) * 100 / COUNT(Session_ID),1) AS Indice_Match_Status_New  
FROM status_new;
```

Resultado: Indice_Match_Status = 78.0 e Indice_Match_Status_New = 59.0





Relação entre cancelamentos e status final do processo de identificação biométrica



Nota-se que, **não há correlação entre cancelamentos e falhas (PROVIDER_FAILED)** e **há uma correlação fraca entre os cancelamentos e o status de pedido NOT_MATCH**, isto é, não há relação de causa e efeito e os cancelamentos não são explicados pelos status finais PROVIDER_FAILED e NOT_MATCH. Além disso, há uma **correlação moderada entre os pedidos cancelados e o status final MATCH**, o que não é esperado em um cenário normal. Dessa forma, essa correlação representa um risco de fraude, dado que mesmo após a liberação para atuação os entregadores possuem um comportamento divergente do esperado e a entrega é cancelada. Portanto, esse comportamento pode indicar o **empréstimo de conta**, por exemplo.

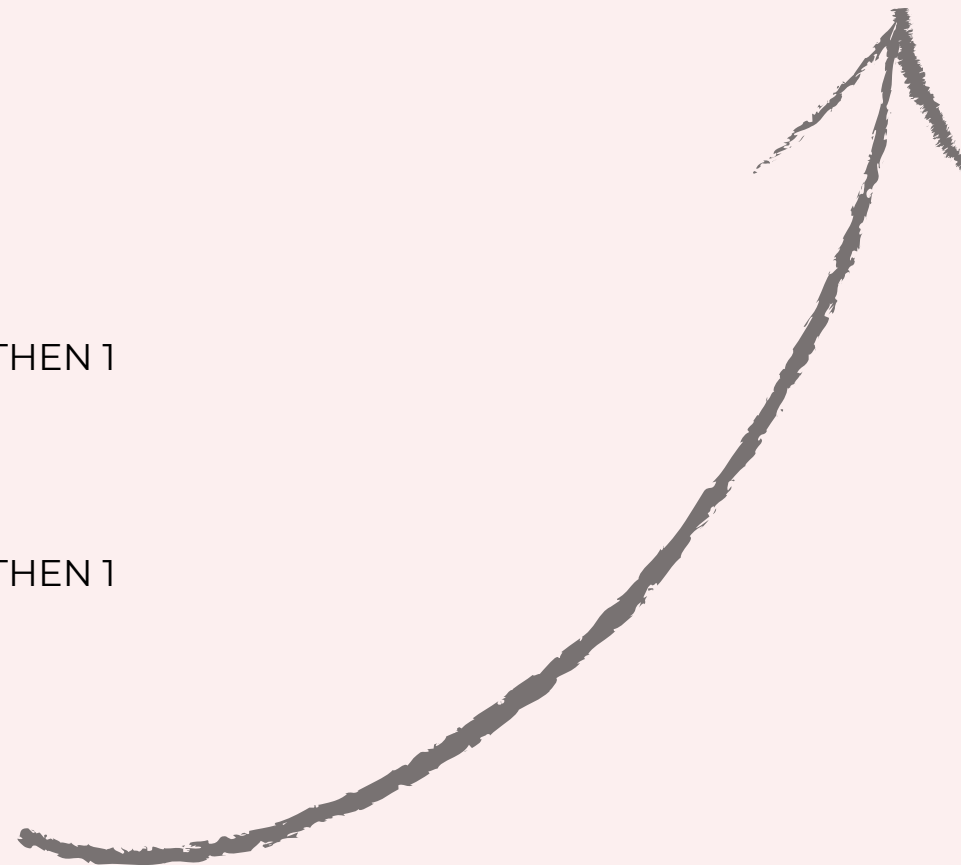


Query - Pergunta D

Query:

```
WITH validacao_bio AS (  
  SELECT  
    b.Event_Dt,  
    o.Order_ID,  
    b.Driver_ID,  
    CASE  
      WHEN Event_Dt <= Order_Dt THEN "1"  
      ELSE 0  
    END AS Biometria_Realizada,  
    CASE  
      WHEN Status == 'MATCH' THEN "1"  
      ELSE "0"  
    END AS Match,  
    CASE  
      WHEN Status == 'NOT_MATCH' THEN 1  
      ELSE 0  
    END AS Not_Match,  
    CASE  
      WHEN Status == 'PROVIDER_FAILED' THEN 1  
      ELSE 0  
    END AS Provider_Failed,  
    CASE  
      WHEN Order_Status == 'CANCELLED' THEN 1  
      ELSE 0  
    END AS Cancelados  
  FROM biometry b  
  LEFT JOIN orders o  
    ON b.Driver_ID = o.Driver_ID  
)
```

```
SELECT  
  DATE(Event_Dt) AS Event_Dt,  
  COUNT(DISTINCT Order_ID) AS Total_Pedidos,  
  SUM(Match) AS Match,  
  SUM(Not_Match) AS Not_Match,  
  SUM(Provider_Failed) AS Provider_Failed,  
  SUM(Cancelados) AS Cancelados  
FROM validacao_bio  
WHERE Biometria_Realizada == "1"  
GROUP BY DATE(Event_Dt)  
ORDER BY DATE(Event_Dt)
```





Análise de Comportamento Indevido⁺ - Empréstimo de Conta

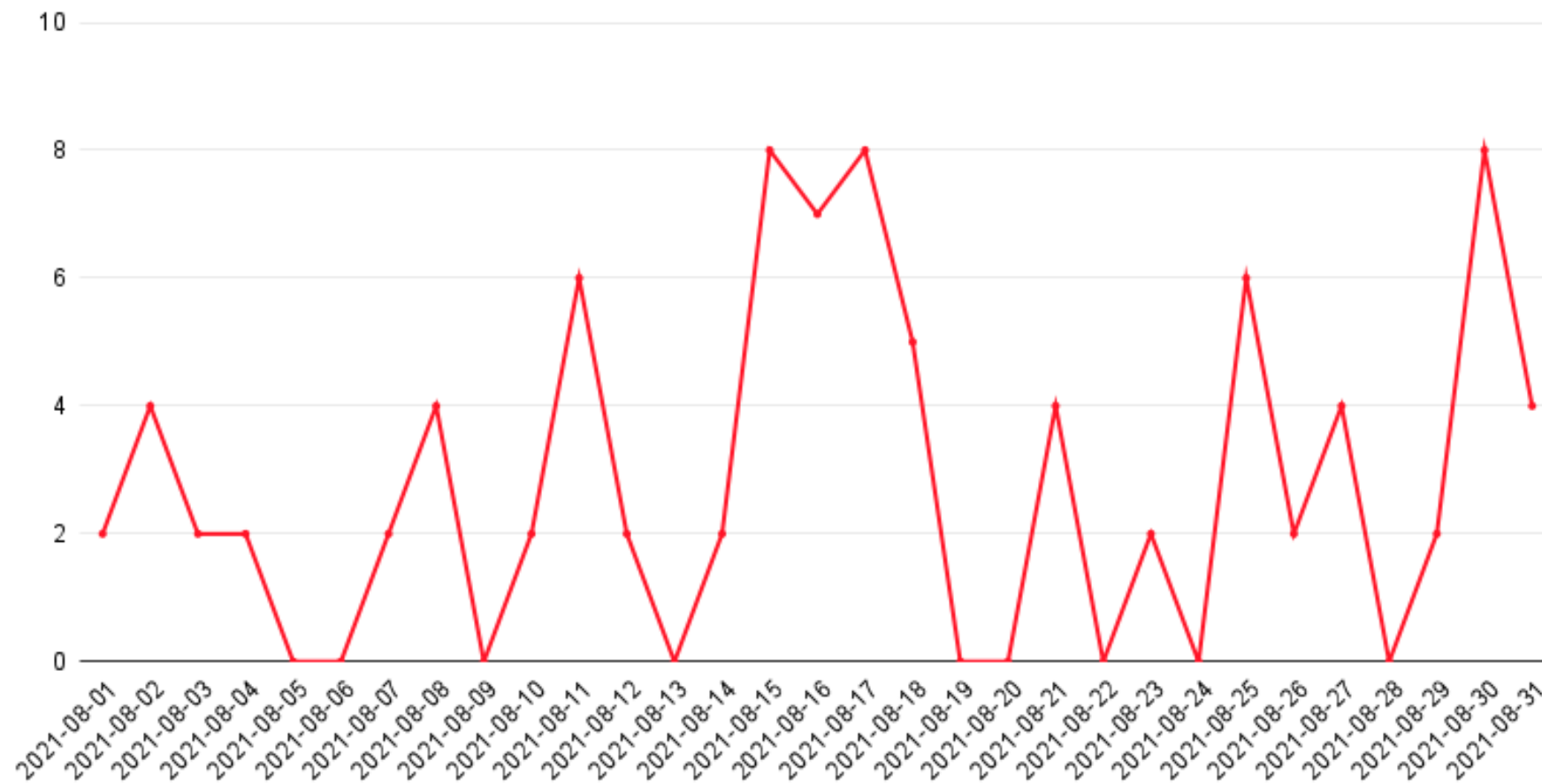
Buscando identificar o comportamento indevido em que ocorre empréstimo de conta, foi realizada a validação de situações não esperadas:

- Análise da **quantidade de sessões por drivers no registro final da biometria** para verificar se cada entregador tem somente uma sessão no resultado final. Essa análise **não retornou nenhum comportamento indevido**, foi identificado que cada driver possuía somente uma sessão.
- Identificação de casos em que o **status final é NOT_MATCH e a action é OK**, isto é, ocorre a liberação do entregador mesmo não passando na validação biométrica.
- Identificação de casos em que o **status final é MATCH e a action é FLUXO_STACK**, isto é, não ocorre a liberação do entregador mesmo passando na validação biométrica.
- Identificação de casos em que o **status final é PROVIDER_FAILED e a action é OK**, isto é, ocorre a liberação do entregador mesmo não passando na validação biométrica.
- Identificação de casos em que o **status final é MATCH e o status da entrega é CANCELLED**, isto é, ocorre a liberação do entregador, mas a entrega é cancelada.
- Análise da **quantidade de pedidos por driver** para identificação de casos suspeitos com volumes de pedidos maiores que a média. Essa análise **retornou que 49% dos usuários possuem uma quantidade de pedidos acima da média geral (cerca de 14 pedidos)**. Drivers como 6140918 e 3552604 apresentam maiores volumes de pedidos, 3 vezes maior que média, gerando um **ponto de atenção para possíveis fraudes e empréstimos de conta.**⁺

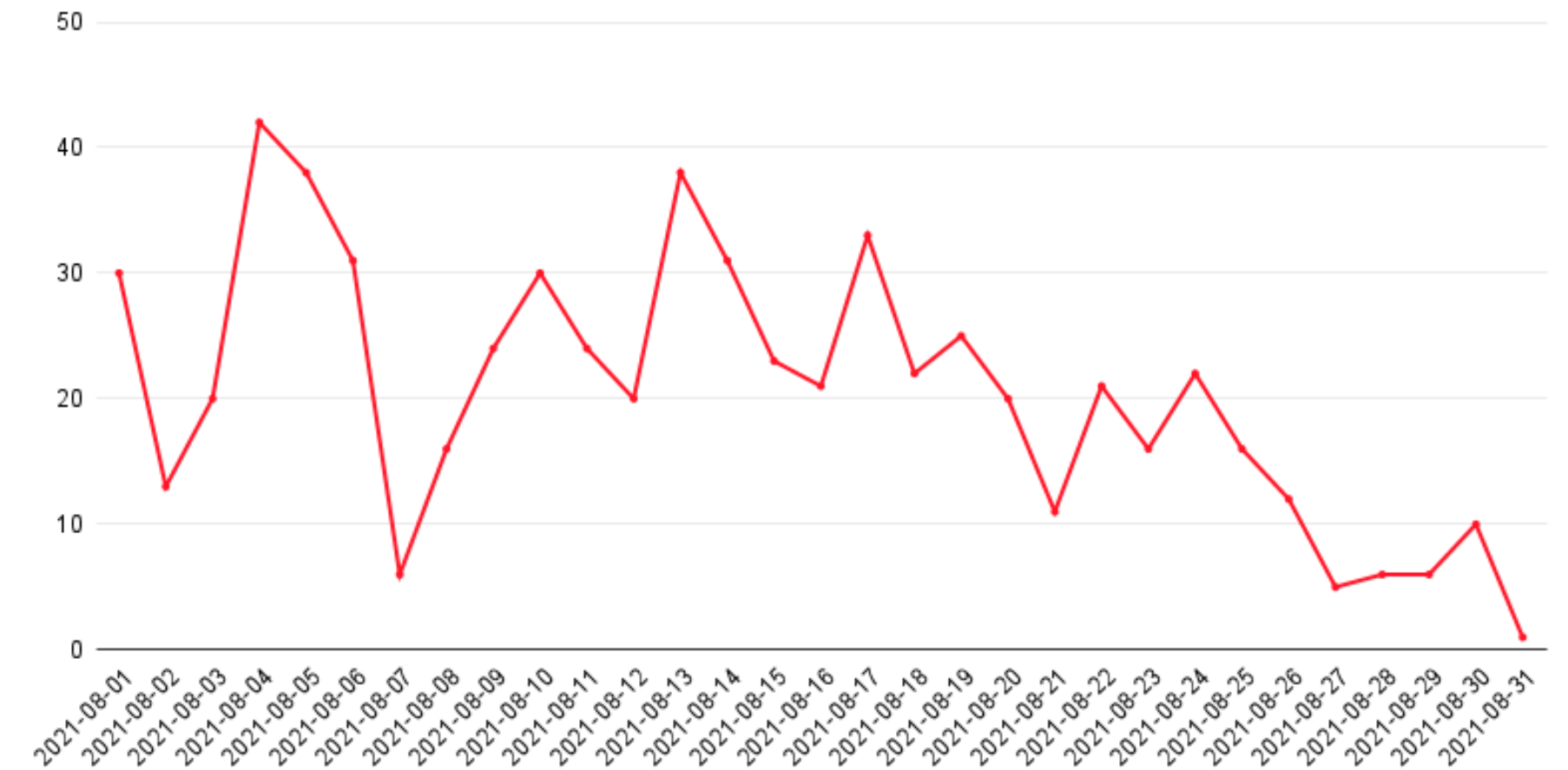


Análise de Comportamento Indevido - Empréstimo de Conta

Compotamentos indevidos (status final x action) por dia



Compotamentos indevidos (status final x status entrega) por dia



Nota-se que, os dias **04/08**, **10/08**, **13/08** e **17/08** apresentam **maior risco de empréstimos de conta**, dado que, durante esses dias tivemos um volume expressivo de comportamentos indevidos, principalmente em casos em que o status final da validação biométrica é MATCH e o status da entrega é CANCELLED. A ocorrência de casos de divergência entre o status final da ação biométrica e a ação tomada com o entregador possuem menor proporção, porém, também representam um risco de fraude principalmente entre os dias 15/08 e 17/08 e no dia 30/08.



Query - Pergunta E

Query 1 - Análise da quantidade de sessões e por drivers no registro final da biometria para verificar se cada entregador tem somente uma sessão no resultado final:

```
SELECT
Event_Dt,
Total_Sessoes_Distintos,
Total_Drivers_Distintos
FROM (
  SELECT
    DATE(Event_Dt) AS Event_Dt,
    COUNT(DISTINCT Session_ID) AS Total_Sessoes_Distintos,
    COUNT(DISTINCT Driver_ID) AS Total_Drivers_Distintos
  FROM biometry
  GROUP BY DATE(Event_Dt)
) AS aux
WHERE Total_Sessoes_Distintos != Total_Drivers_Distintos
ORDER BY Total_Sessoes_Distintos DESC;
```

Resultado: Não retornou nenhum caso divergente





Query - Pergunta E

Query 2 - Compotamentos indevidos (status final x action) por dia:

```
WITH geral AS (  
  SELECT  
    DATE(Event_Dt) AS Event_Dt,  
    CASE  
      WHEN Status == 'MATCH' AND Action == "FLUXO_STACK" THEN "1"  
      ELSE "0"  
    END AS Match_FS,  
    CASE  
      WHEN Status == 'NOT_MATCH' AND Action == "OK" THEN 1  
      ELSE 0  
    END AS Not_Match_Liberado,  
    CASE  
      WHEN Status == 'PROVIDER_FAILED' AND Action == "OK" THEN 1  
      ELSE 0  
    END AS Falha_Liberado  
  FROM biometry  
)  
SELECT  
  Event_Dt,  
  SUM(Match_FS) AS Match_FS,  
  SUM(Not_Match_Liberado) AS Not_Match_Liberado,  
  SUM(Falha_Liberado) AS Falha_Liberado,  
  SUM(Match_FS) + SUM(Not_Match_Liberado) + SUM(Falha_Liberado) AS Total_Inconsistencias  
FROM geral  
GROUP BY Event_Dt  
ORDER BY Event_Dt
```





Query - Pergunta E

Query 3 - Compotamentos indevidos (status final x status entrega) por dia:

```
WITH validacao_match AS (  
  SELECT  
    Event_Dt,  
    Order_ID,  
    CASE  
      WHEN Event_Dt <= Order_Dt THEN "1"  
      ELSE 0  
    END AS Biometria_Realizada,  
    CASE  
      WHEN Status == 'MATCH' AND Order_Status == 'CANCELLED' THEN "1"  
      ELSE "0"  
    END AS Match_Cancelado  
  FROM biometry b  
  LEFT JOIN orders o  
    ON b.Driver_ID = o.Driver_ID  
)  
SELECT  
  DATE(Event_Dt) AS Event_Dt,  
  COUNT(DISTINCT Order_ID) AS Total_Pedidos,  
  SUM(Match_Cancelado) AS Match_Cancelado  
FROM validacao_match  
WHERE Biometria_Realizada == "1"  
GROUP BY DATE(Event_Dt)  
ORDER BY DATE(Event_Dt)
```





Query - Pergunta E

Query 4 - Análise da quantidade de entregas por driver:

```
WITH pedidos_por_entregador AS (  
  SELECT  
    Driver_ID,  
    COUNT(Order_ID) AS Total_Pedidos  
  FROM orders  
  WHERE Driver_ID != '#N/D'  
  GROUP BY Driver_ID  
,  
  media_geral AS (  
    SELECT  
      AVG(Total_Pedidos) AS Media_Total_Pedidos  
    FROM pedidos_por_entregador  
  )  
  SELECT  
    p.Driver_ID,  
    p.Total_Pedidos,  
    m.Media_Total_Pedidos,  
    CASE  
      WHEN p.Total_Pedidos > m.Media_Total_Pedidos THEN 1  
      ELSE 0  
    END AS Acima_Media  
  FROM pedidos_por_entregador p  
  CROSS JOIN media_geral m  
  ORDER BY p.Total_Pedidos DESC
```

Query 4 - Análise do percentual acima da média:

```
WITH pedidos_por_entregador AS (  
  SELECT  
    Driver_ID,  
    COUNT(Order_ID) AS Total_Pedidos  
  FROM orders  
  WHERE Driver_ID != '#N/D'  
  GROUP BY Driver_ID  
,  
  media_geral AS (  
    SELECT  
      AVG(Total_Pedidos) AS Media_Total_Pedidos  
    FROM pedidos_por_entregador  
  ),  
  acima_media_final AS (  
    SELECT  
      p.Driver_ID,  
      p.Total_Pedidos,  
      m.Media_Total_Pedidos,  
      CASE  
        WHEN p.Total_Pedidos > m.Media_Total_Pedidos THEN 1  
        ELSE 0  
      END AS Acima_Media  
    FROM pedidos_por_entregador p  
    CROSS JOIN media_geral m  
  )  
  SELECT  
    ROUND(SUM(Acima_Media) * 100 / COUNT(*), 1) AS Percentual_Acima_Media  
  FROM acima_media_final;
```





+ Considerações finais



Principais insights e pontos de atenção:

- Serviço com maior número de falhas: Considerando volumetria, o serviço **A**. Considerando taxa de falhas, serviço **C**.
- Categoria de entregador com maior índice falha na identificação biométrica: **Bronze**.
- Índice geral de MATCH se aumentássemos a similaridade mínima do MATCH para 90%: **Redução em 19 p.p. atingindo um índice de 59%. Ocorre o aumento da confiança, porém, com possibilidade de falsos negativos.**
- Não há relação entre cancelamentos e status NOT_MATCH e PROVIDER_FAILED. A relação entre cancelamos e status MATCH representa um risco de fraude.
- Os dias **04/08, 10/08, 13/08, 15/08 a 17/08 e 30/08** apresentam maior risco de empréstimos de conta.

A identificação de dias, categorias de entregadores e serviços com maior incidência de falhas ou indícios de comportamentos indevidos permite a priorização de **ações contra fraude** e aplicação de **melhorias nos sistemas de validação e segurança**.



Obrigada!

