



Análise de Fraudes

Análise de Fraudes e Proposta de Soluções

Avaliação das perdas no e-commerce do **Walmart** na região central da Flórida, juntamente com soluções estratégicas fundamentadas em dados para redução de perdas no ano de 2024.

Por Pedro Kaneko

Tópicos

- Regiões do estudo
- O problema
- Análise
- Atribuição das perdas
- Entregas
- Motoristas
- Clientes
- Interação
- Solução proposta
- Teste de hipótese
- Apêndice técnico



Regiões do Estado

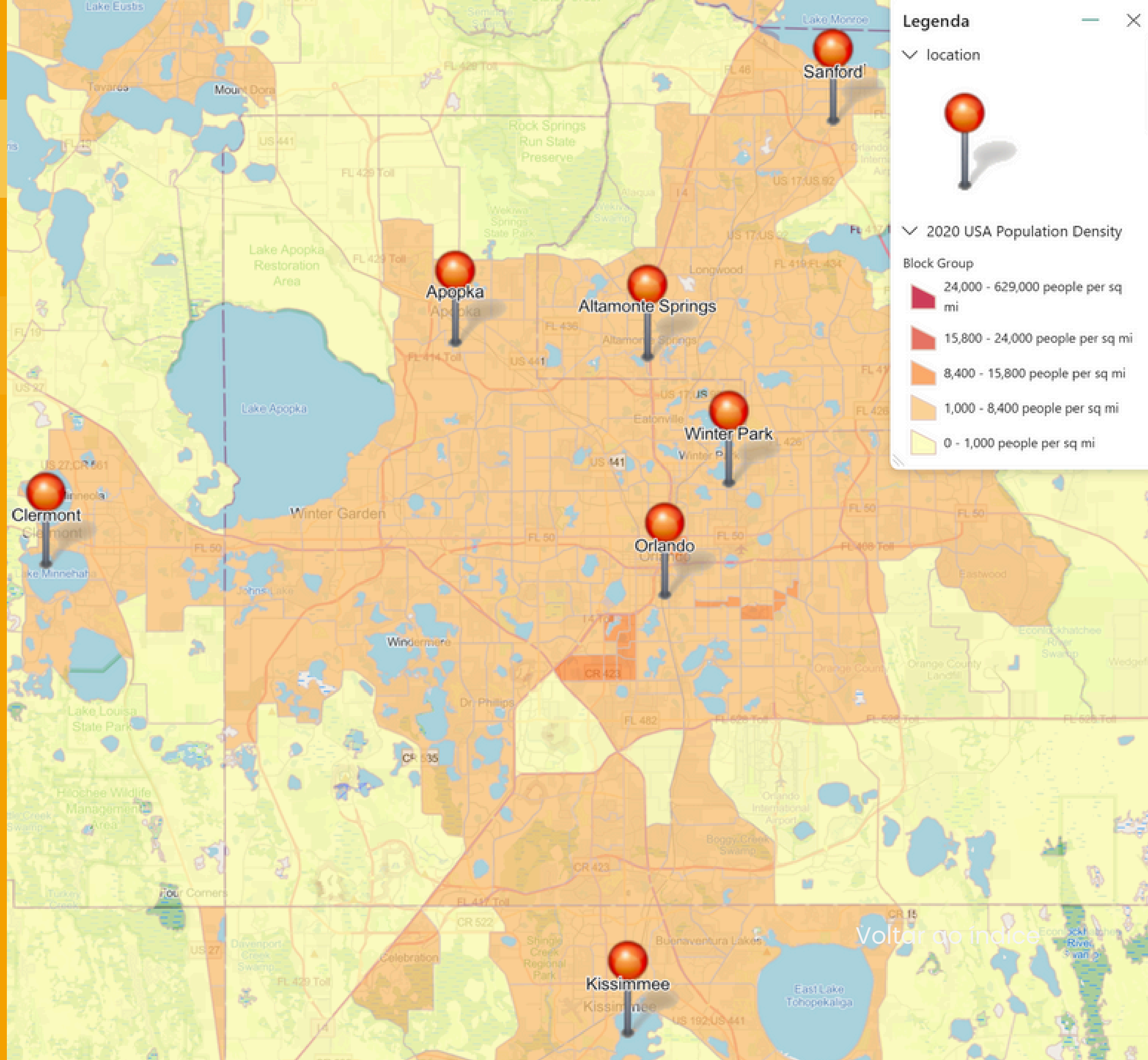
Cidades da região central do estado da Flórida, EUA.

[Voltar ao índice](#)



Flórida, EUA

- Altamonte Springs
- Apopka
- Clermont
- Kissimmee
- Orlando
- Sanford
- Winter Park





O problema

O Walmart tem enfrentado reclamações de consumidores sobre entregas incompletas e, com a análise dos dados, deseja entender se a responsabilidade pela fraude pode ser atribuída aos consumidores, aos entregadores ou a outra causa.

[Voltar ao índice](#)

O Problema

Desafios no *In-Home Delivery* e seu impacto na operação.

Pedidos	15% dos pedidos apresentaram problemas (Perda de Itens)
Itens	1,63% do total de itens foram extraviados.
Financeiro	5% da receita bruta foi perdido (prejuízo de \$149 mil)





Análise

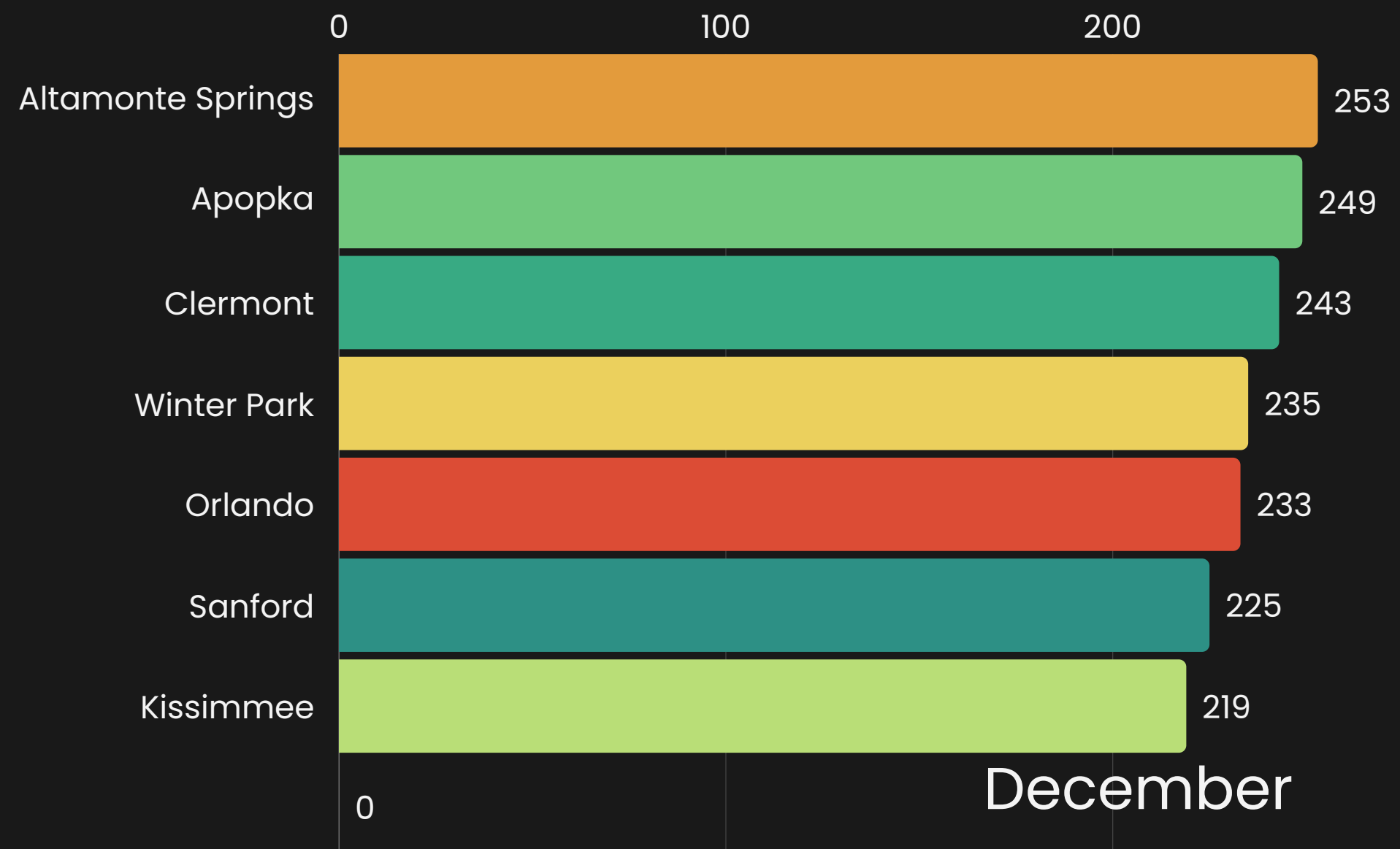
Analise das perdas e prejuízo por tipo de produto, região e motoristas ao longo do ano de 2023.

[Voltar ao índice](#)



Perdas por Região

- **Altamonte Springs e Clermont** apresentaram um aumento significativo nas perdas ao longo do ano.

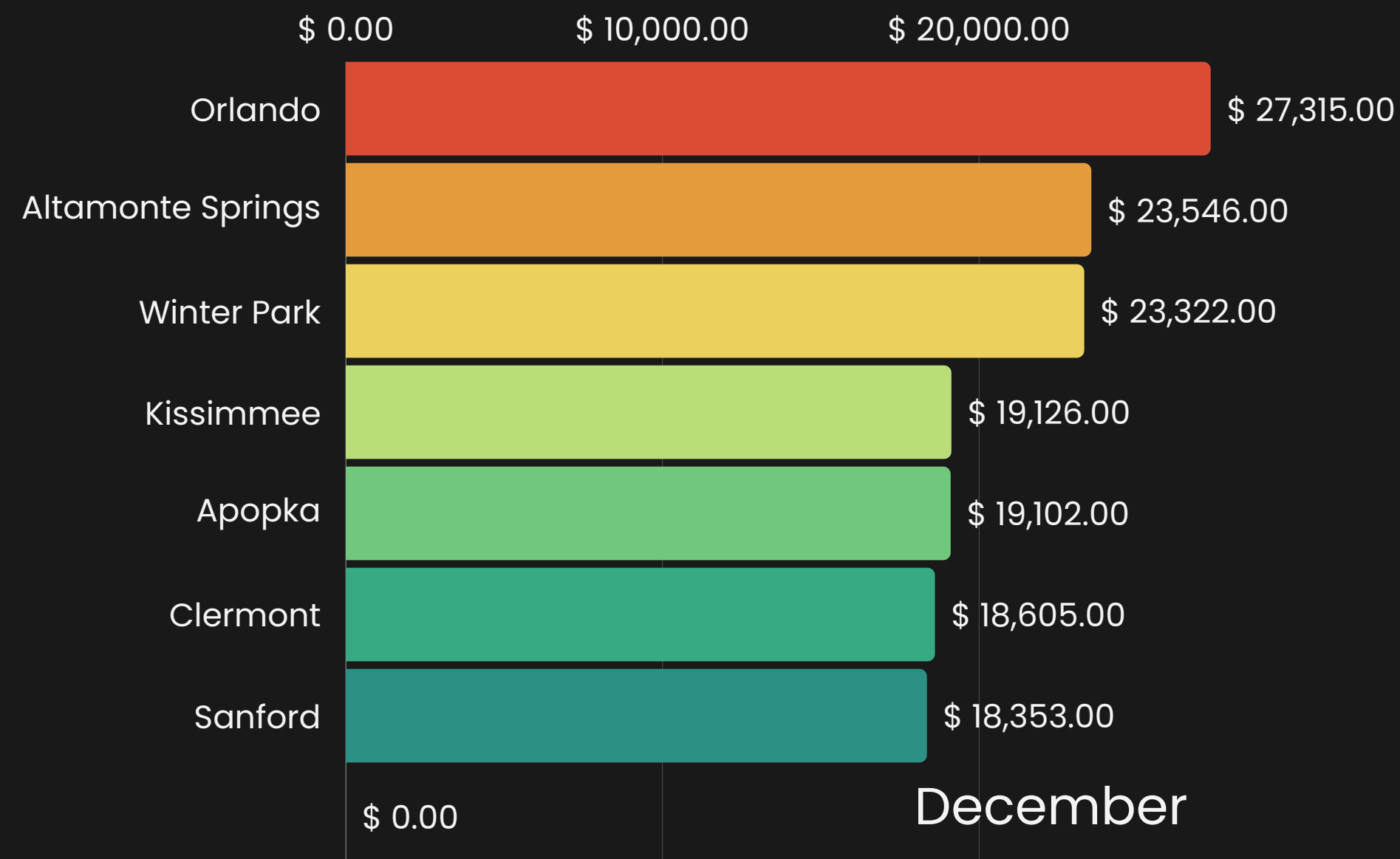


[Voltar ao índice](#)



Prejuízo por Região

- Ao longo do ano de 2023, **Orlando** avançou da quinta posição para a primeira, em termos de prejuízo total acumulado.



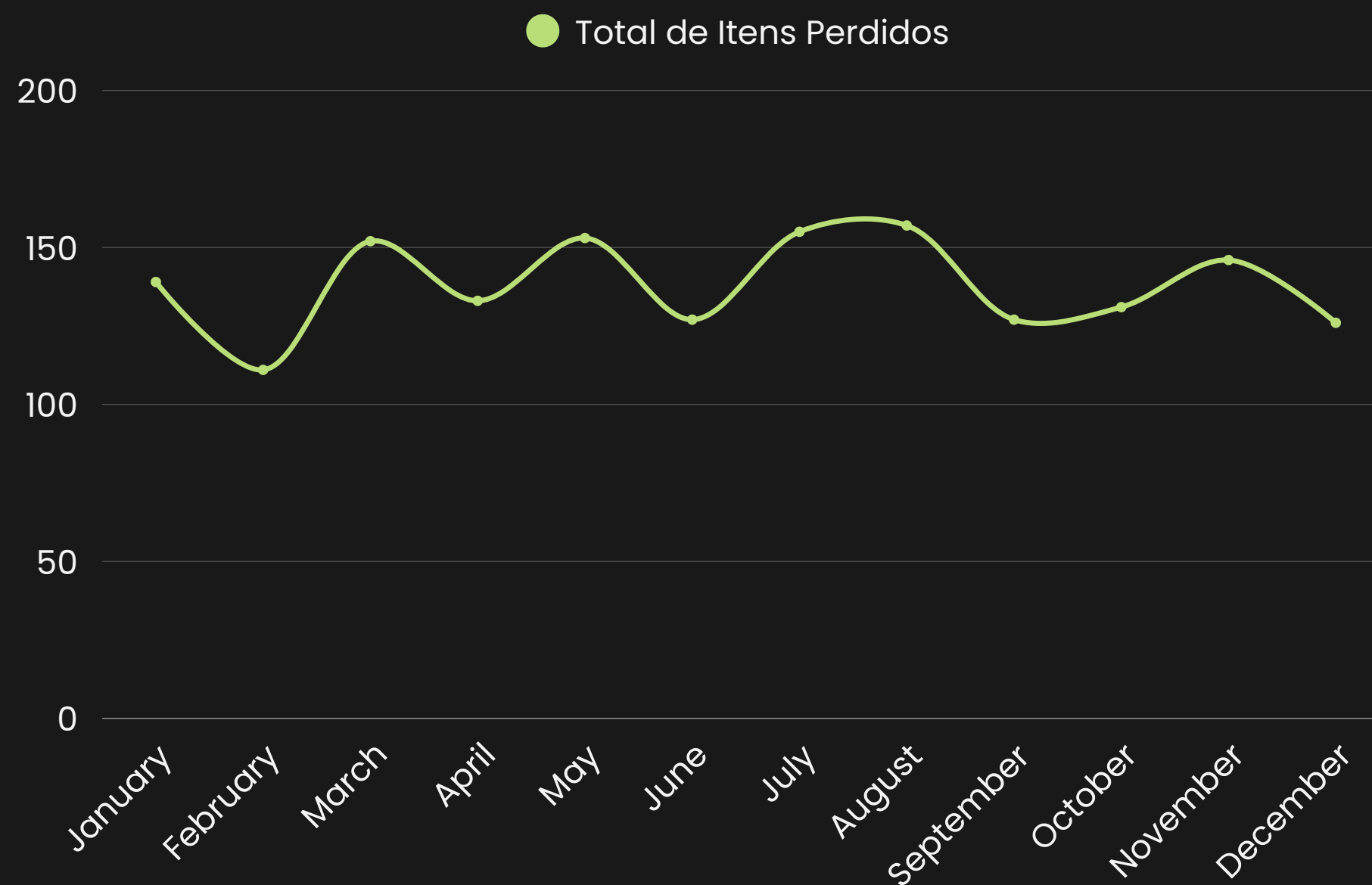
[Voltar ao índice](#)



PERDAS

Tendência Geral

- Em relação às perdas, há variações mensais, mas sem uma tendência definida, indicando um estado estacionário.



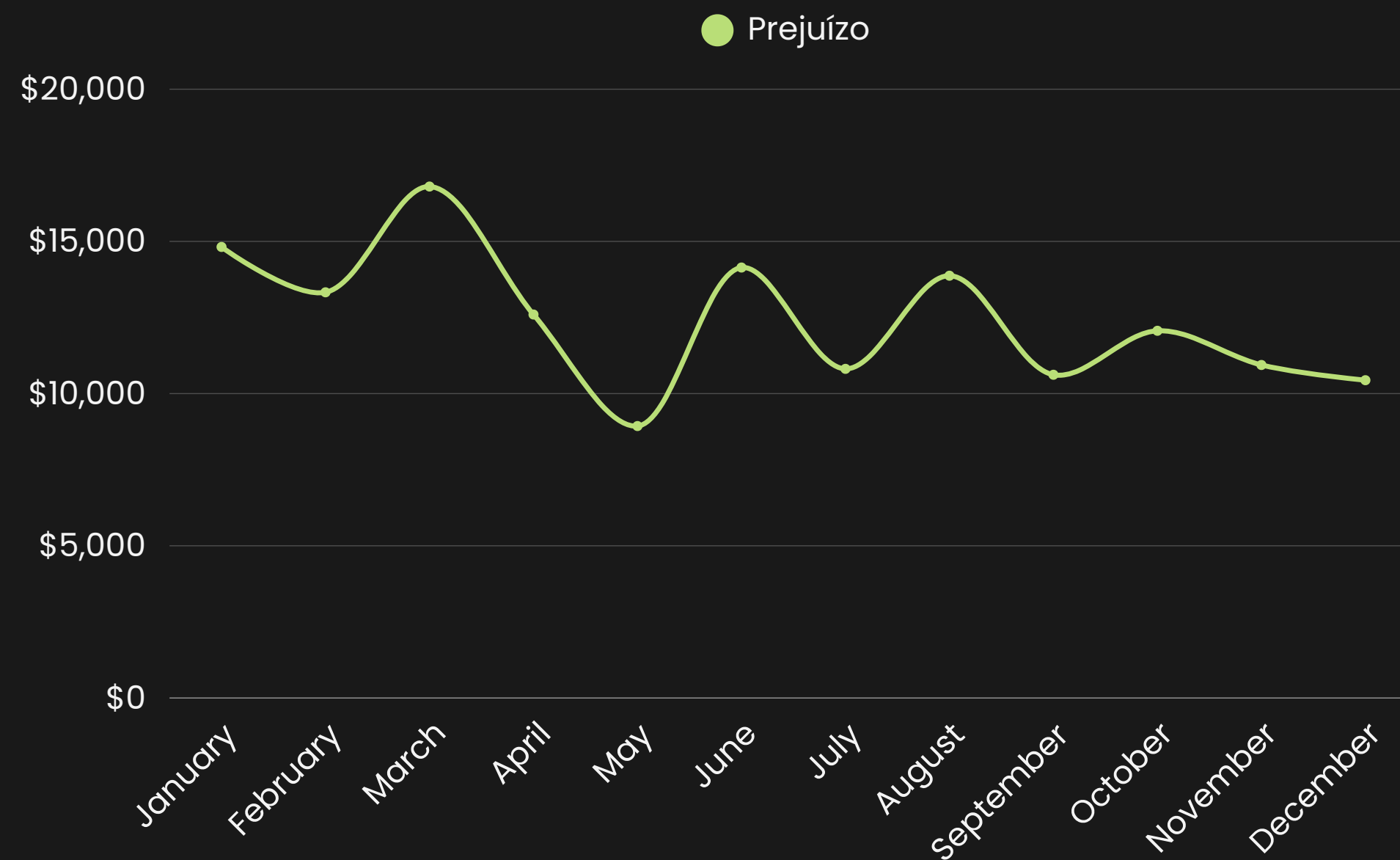
[Voltar ao índice](#)



PREJUÍZO

Tendência Geral

- O prejuízo mensal demonstra uma leve diminuição ao longo do ano.



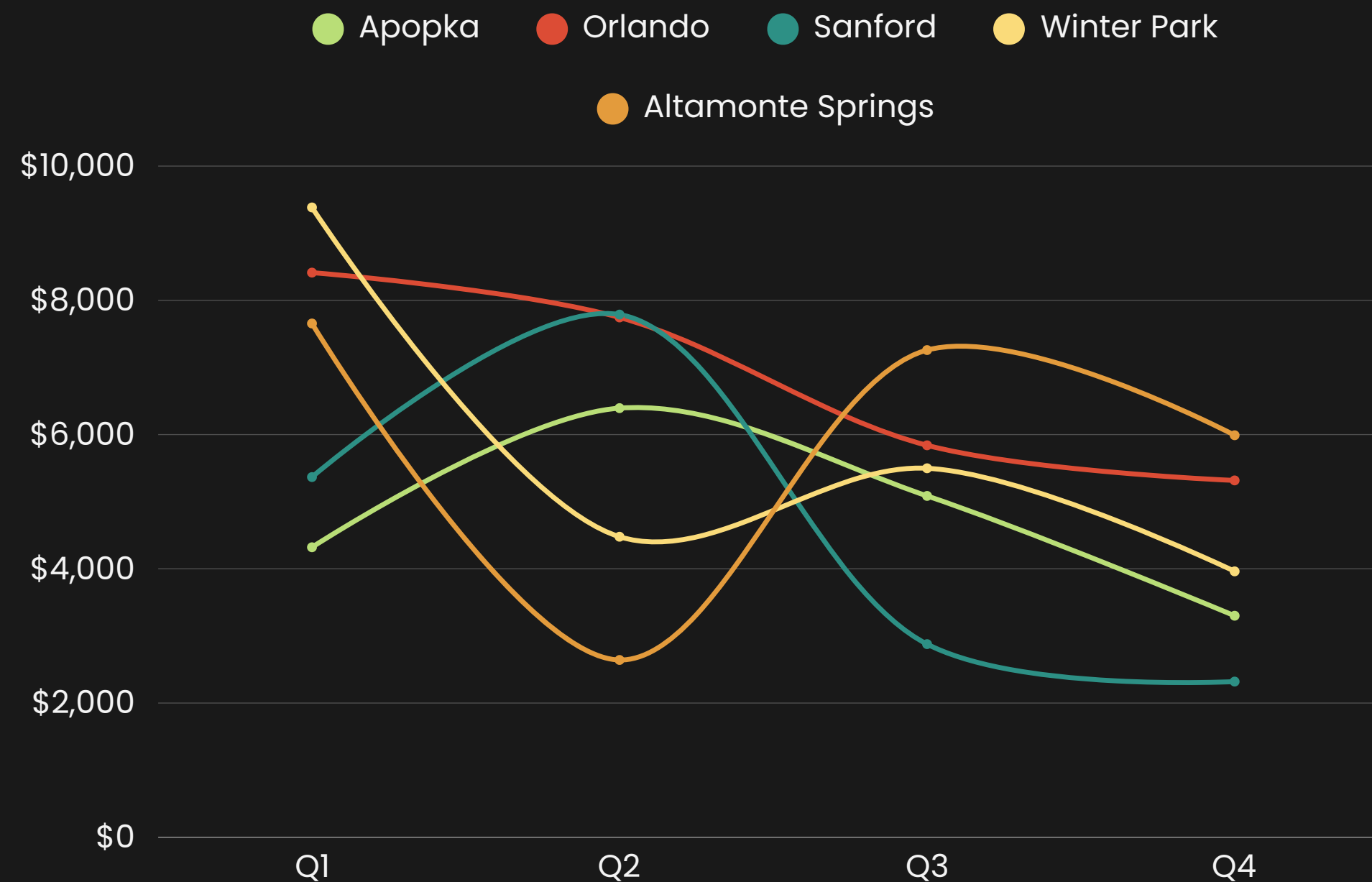
[Voltar ao índice](#)



PREJUÍZO POR REGIÃO

Tendência de Queda

- Mesmo com a queda no último trimestre, Orlando e Altamonte Springs ainda apresentam altos níveis de prejuízos.



[Voltar ao índice](#)



PREJUÍZO POR REGIÃO

Tendências de Alta

- Clermont e Kissimmee levantaram preocupações, assumindo a liderança como as cidades que apresentaram os maiores prejuízos no último trimestre.



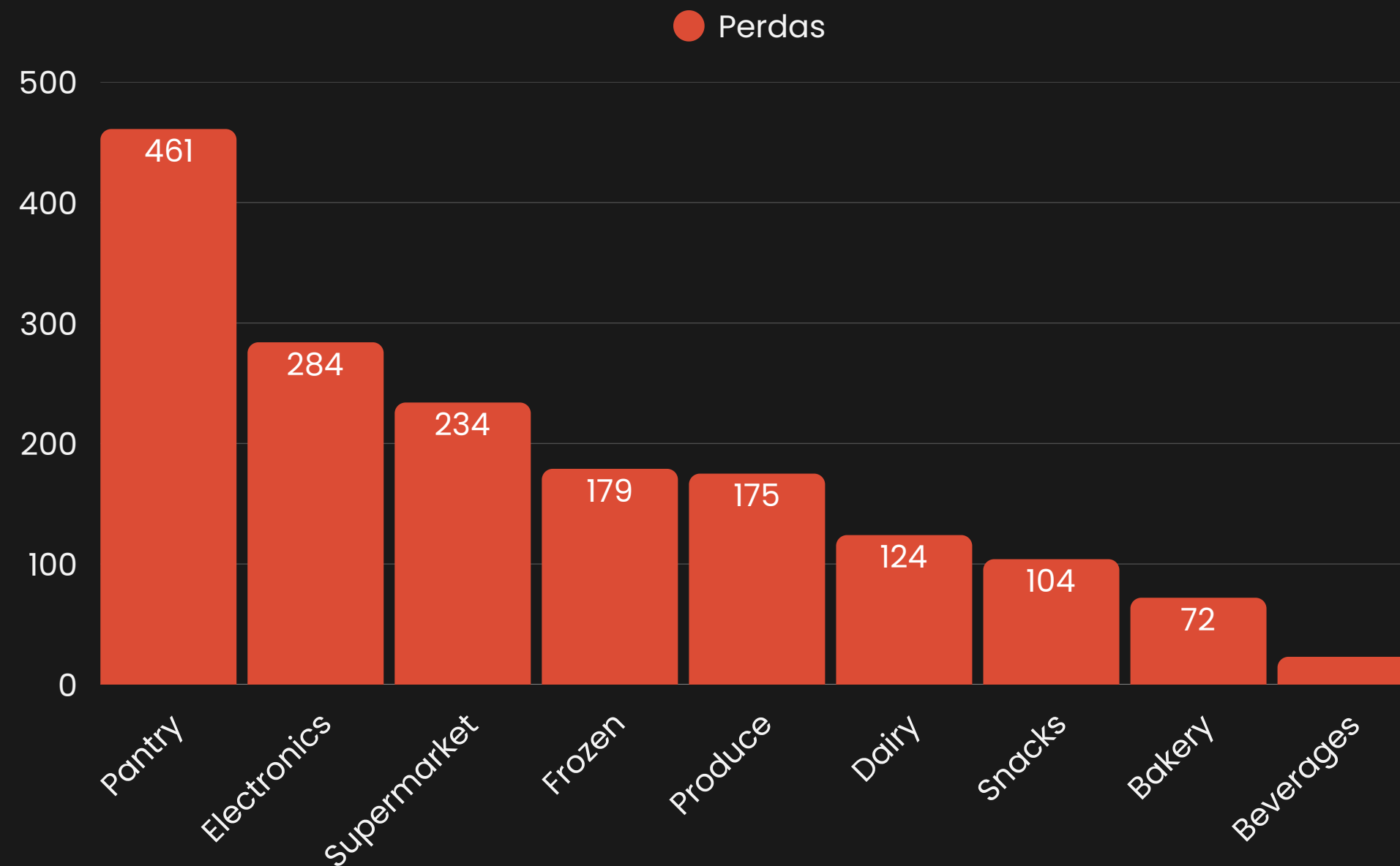
[Voltar ao índice](#)



PERDAS

Categoria de Produtos

- **Pantry** (itens de despensa) lidera em itens perdidos.



[Voltar ao índice](#)



↑ TOP 15

Produtos Extraviados

- Produtos alimentícios como laticínios congelados e itens de dispensa lideram em itens perdidos por produto.



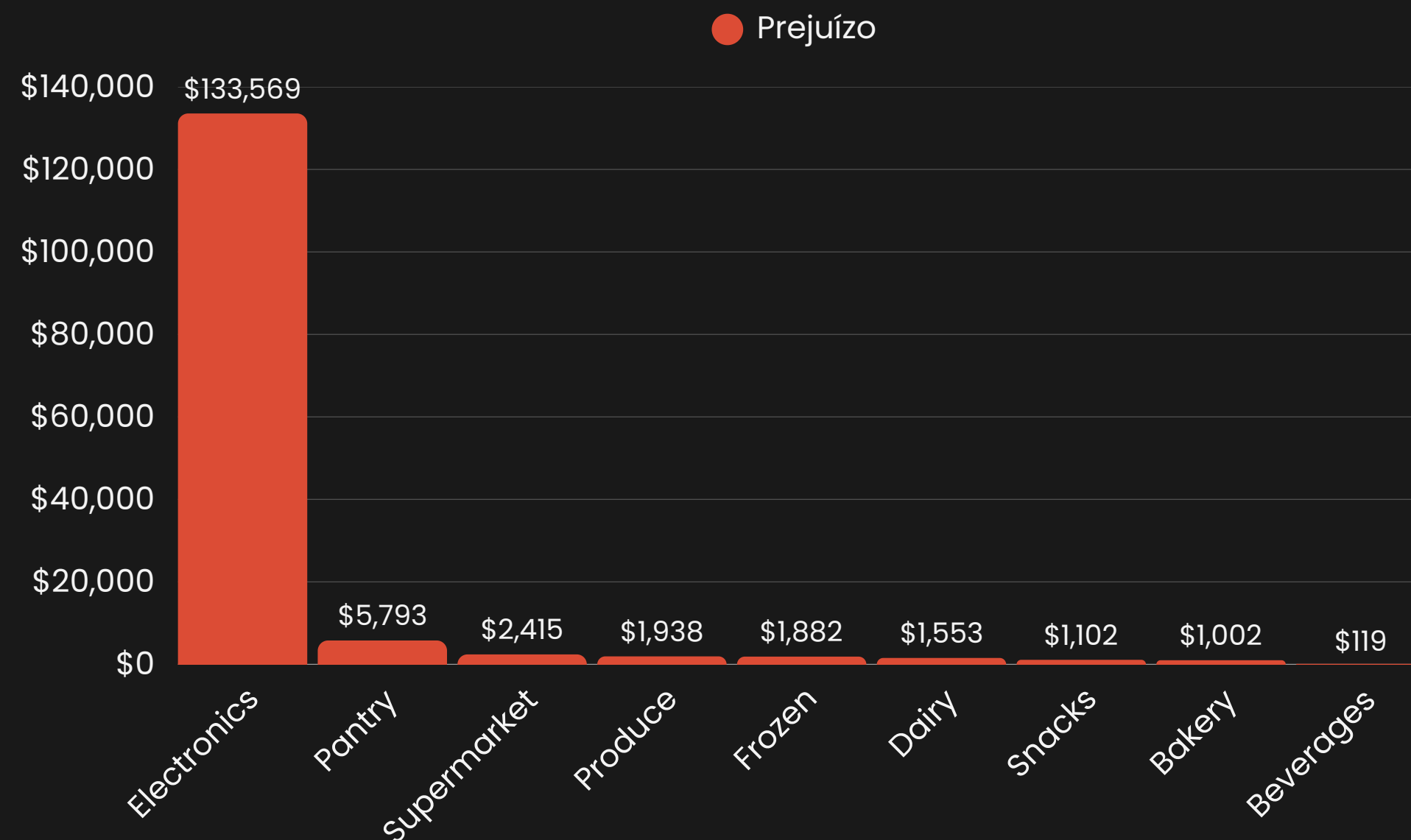
[Voltar ao índice](#)



PREJUÍZO

Categoria de Produtos

- **89 % do prejuízo** é causado pela categoria de produtos eletrônicos.



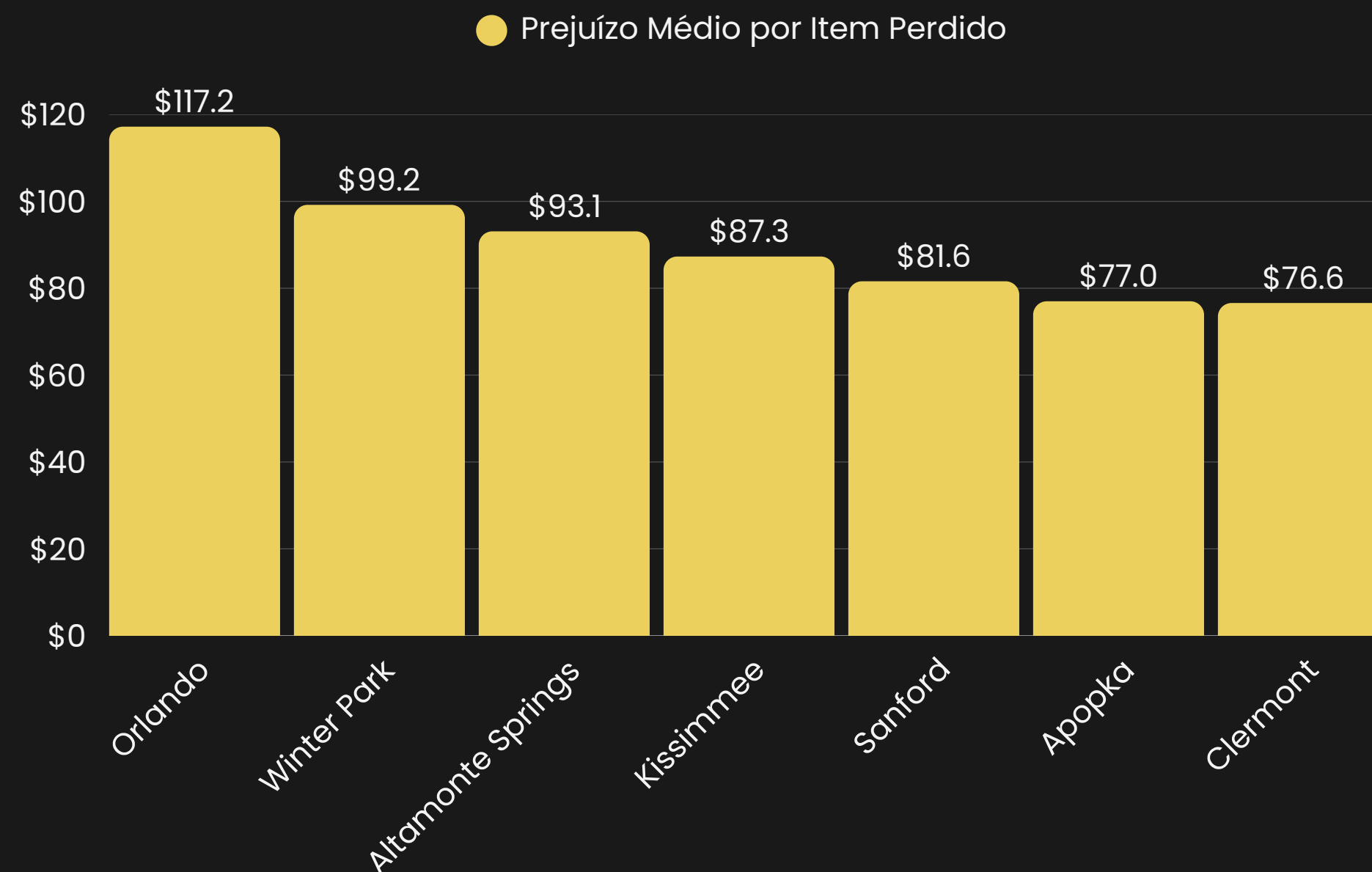
[Voltar ao índice](#)



↑ TOP 15

Prejuízo Médio por Item Perdido

- Orlando Ique lidera em prejuízo tem perdas substancialmente mais elevadas em média.



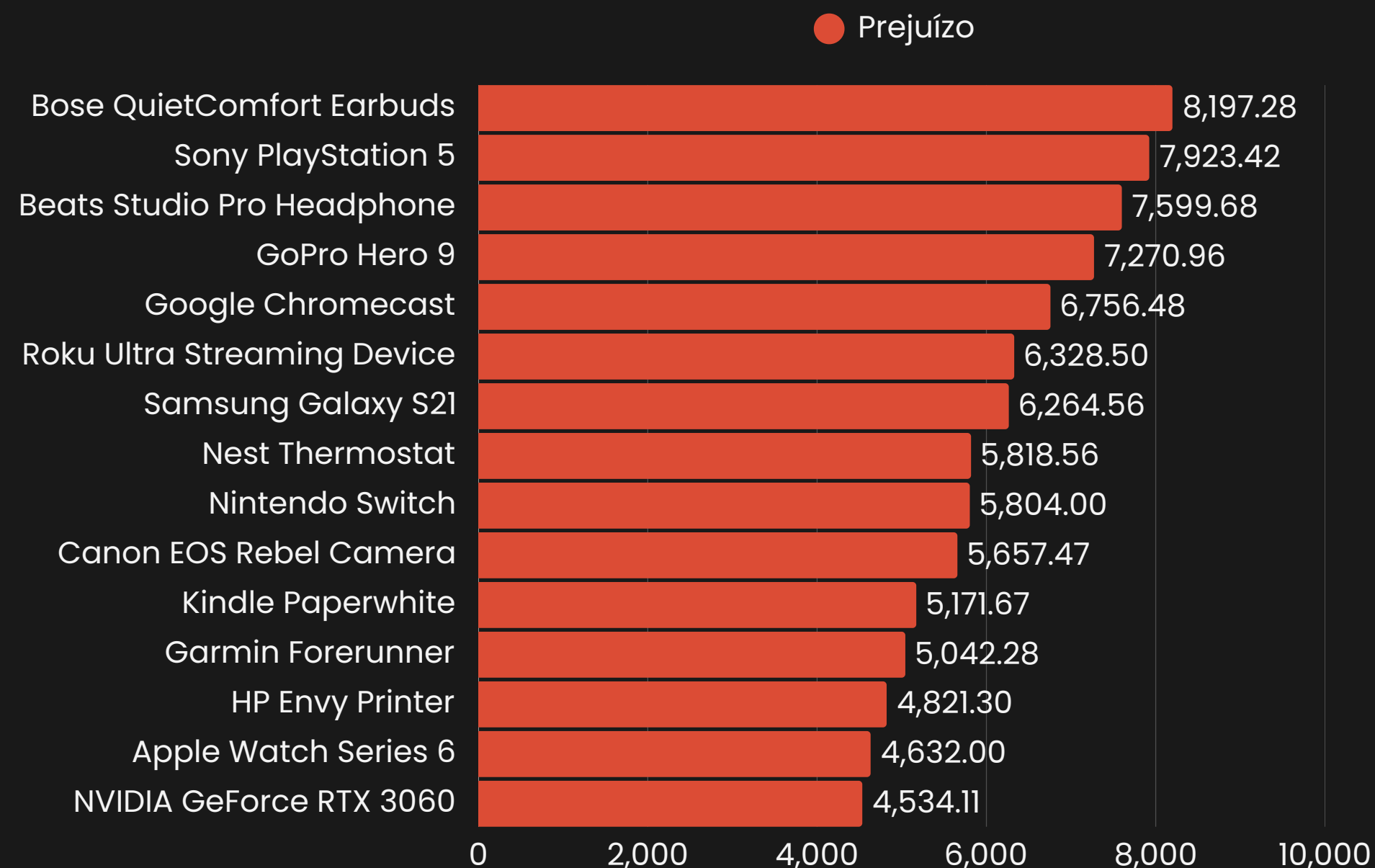
[Voltar ao índice](#)



↑ TOP 15

Prejuízo por Produto

- Os produtos da categoria eletrônicos estão no topo em termos de prejuízo acumulado.



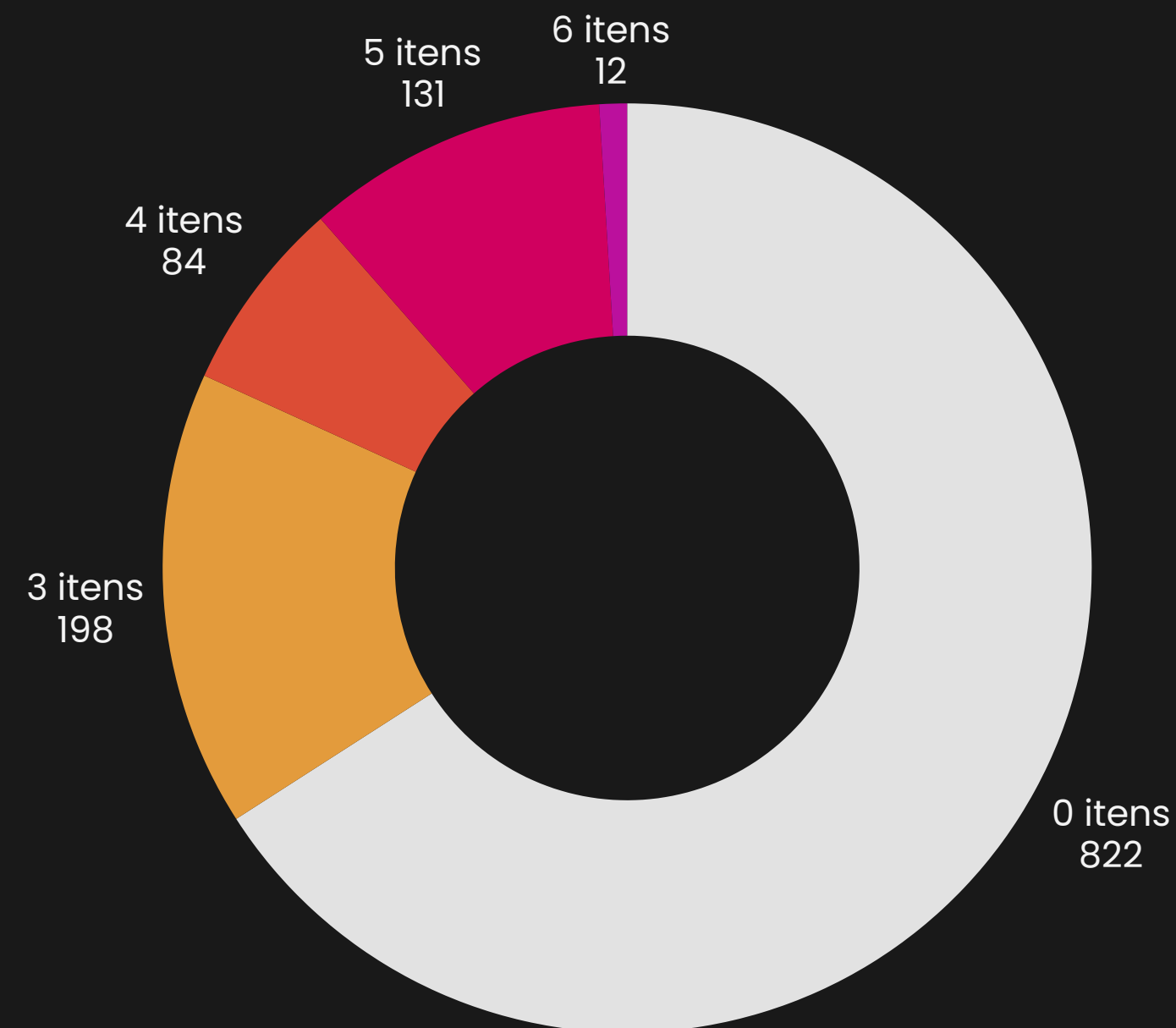
[Voltar ao índice](#)



PERDAS

Motoristas

- 822 motoristas (66%) não tiveram perdas. Os demais porém apresentam perdas recorrentes variando de 3 a 6 itens no total.



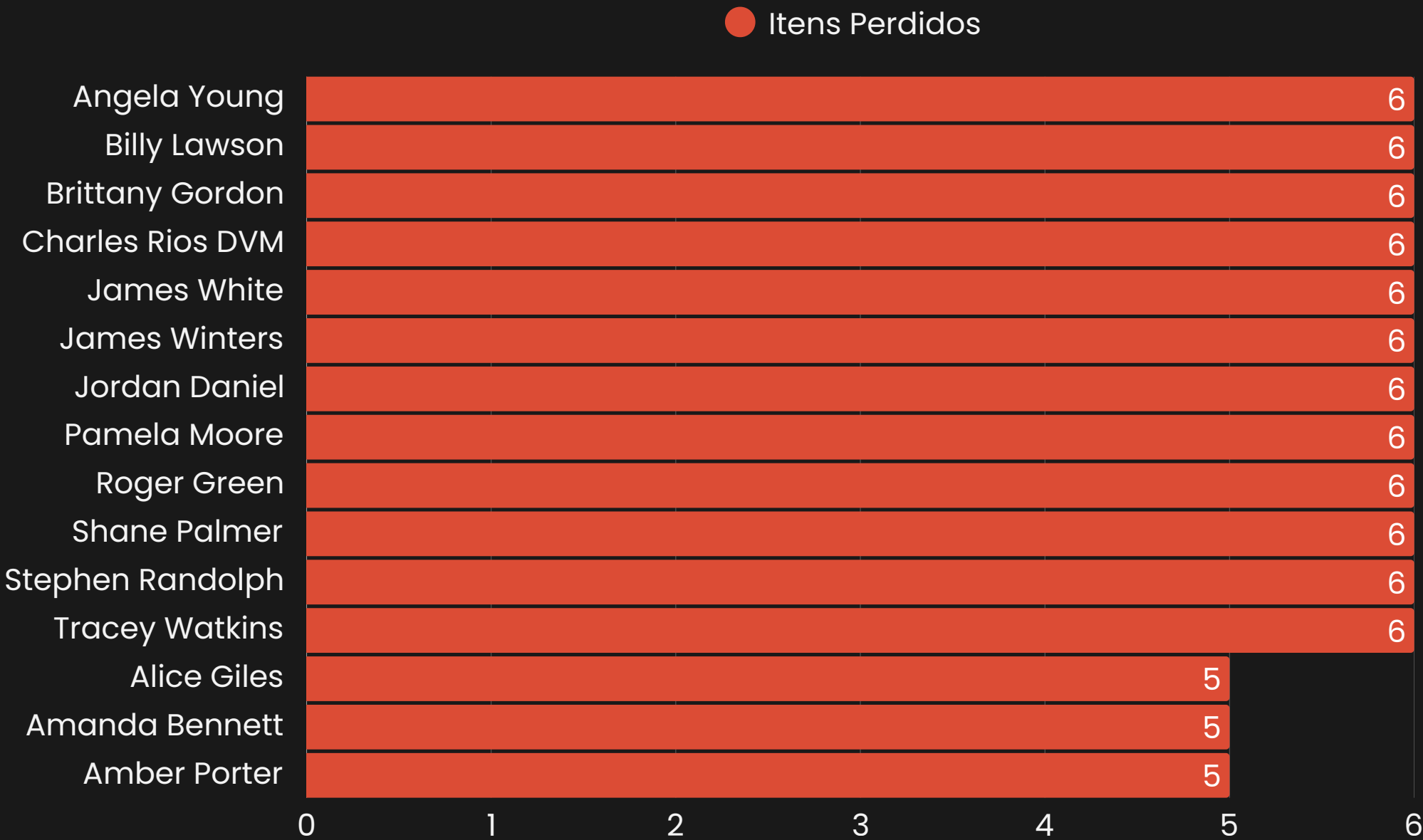
[Voltar ao índice](#)



↑ TOP 15

Perdas por Motorista

- Os produtos da categoria eletrônicos estão no topo em termos de prejuízo acumulado.



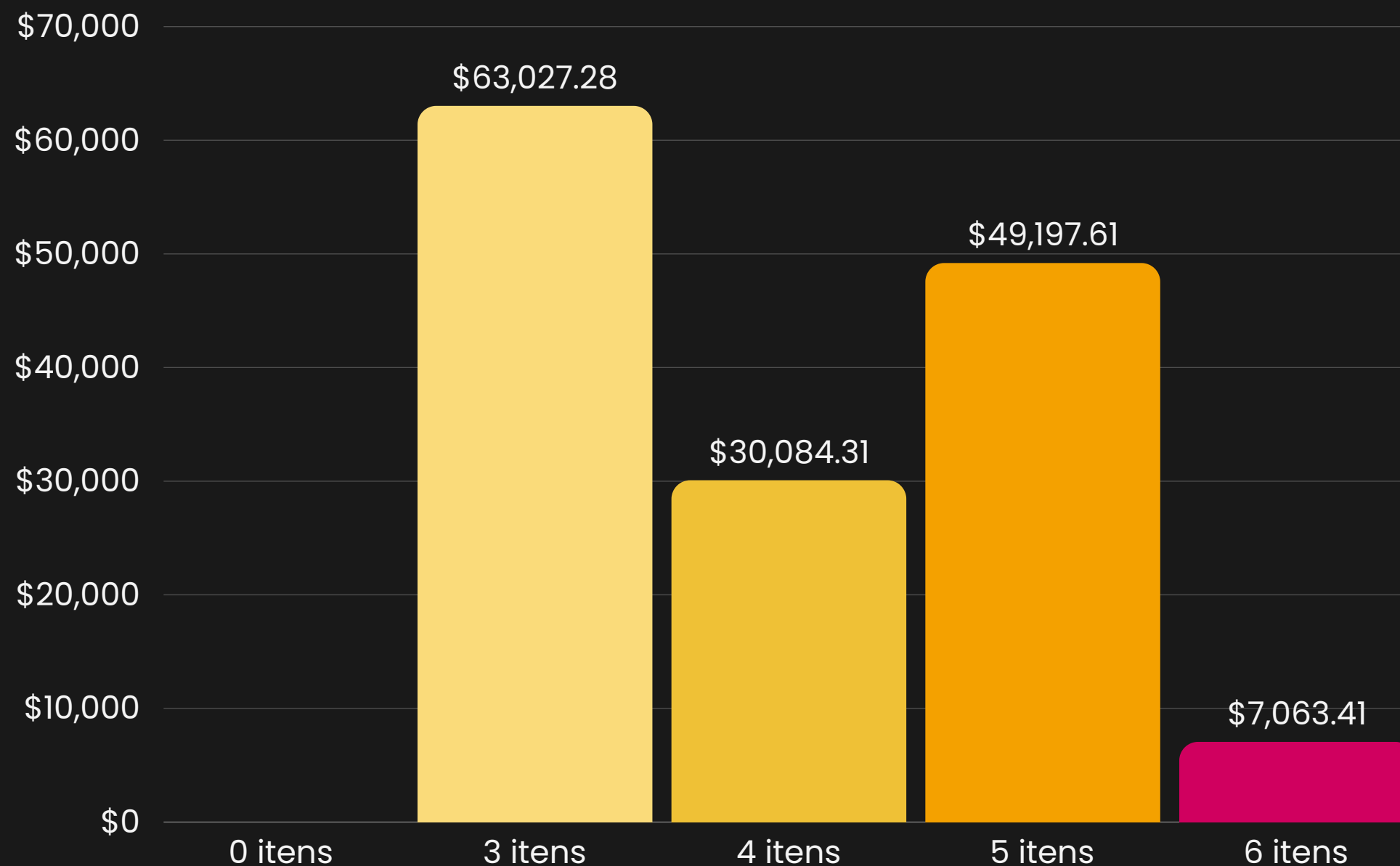
[Voltar ao índice](#)



PREJUÍZO

Motoristas

- Motoristas que perderam 3 itens respondem por **36%** do prejuízo.
- Motoristas que perderam 6 itens respondem por **40%** do prejuízo.



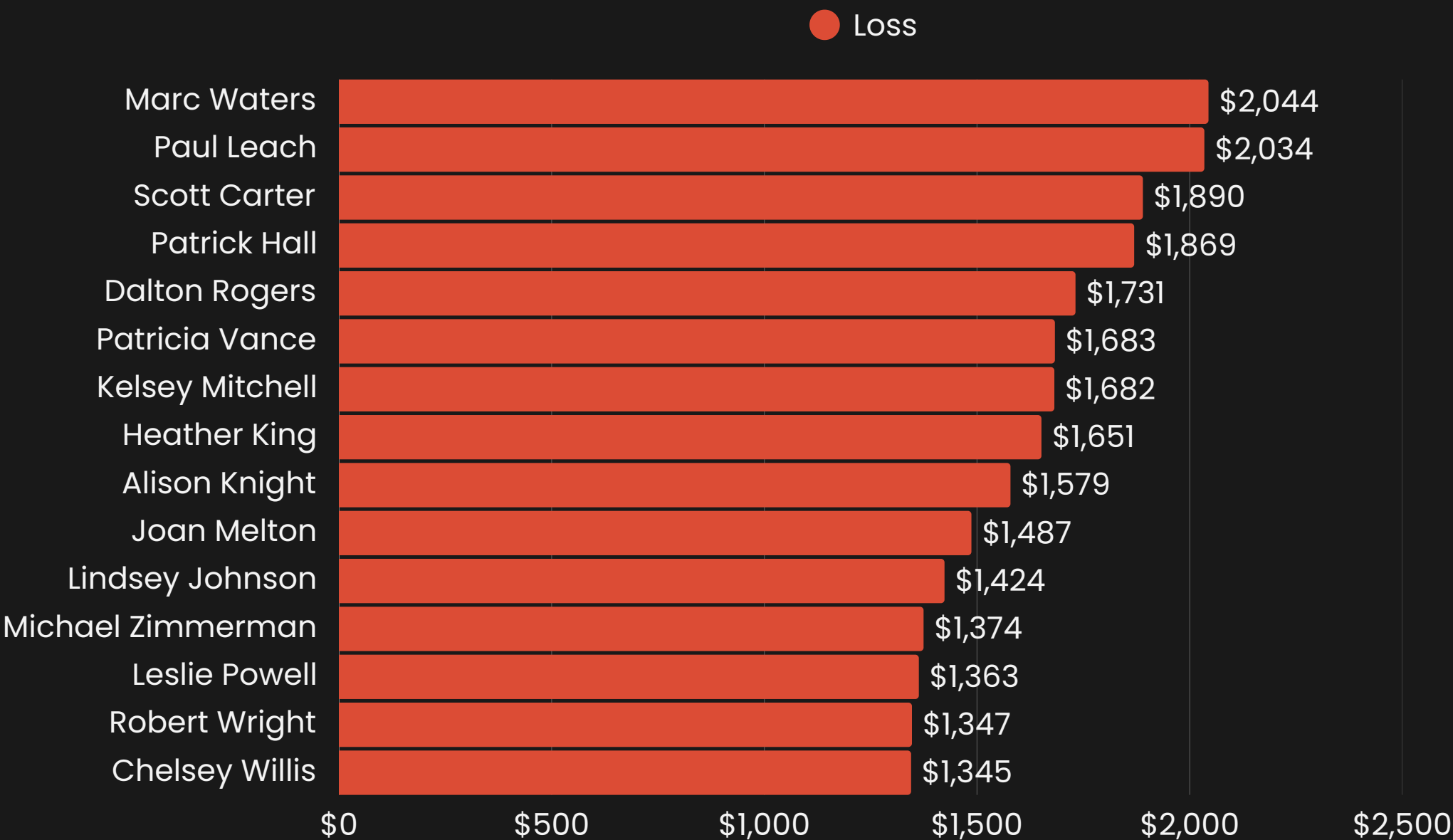
[Voltar ao índice](#)



↑ TOP 15

Prejuízo por Motorista

- Os produtos da categoria eletrônicos estão no topo em termos de prejuízo acumulado.



[Voltar ao índice](#)

Análise dos Motoristas

[Voltar ao índice](#)

Analise baseada na quantidade de itens perdidos

Motoristas agrupados pela quantidade de itens perdidos	Motoristas	Itens perdidos	Prejuízo	Prejuízo médio por motorista
0 itens perdidos	822	0	0	0
3 itens perdidos	198	594	\$63,027.28	\$318.32
4 itens perdidos	84	336	\$30,084.31	\$358.15
5 itens perdidos	131	655	\$49,197.61	\$375.55
6 itens perdidos	12	72	\$7,063.41	\$588.62



Atribuição das Perdas

A responsabilidade dos extravios pode ser atribuída a alguma das partes envolvidas ?
Motoristas, clientes, falha operacional ou a outra causa ?

[Voltar ao índice](#)

Inconsistências Detectadas

[Voltar ao índice](#)



Entregas

Há diversas entregas sendo realizadas fora do horário estabelecido, aumento o risco de furto.



Motoristas

Há motoristas que despertam suspeitas devido a um histórico que se desvia dos padrões habituais de perdas.



Clientes

Também identificamos clientes que apresentam padrões de perdas significativamente elevados, gerando um alerta.





Entregas

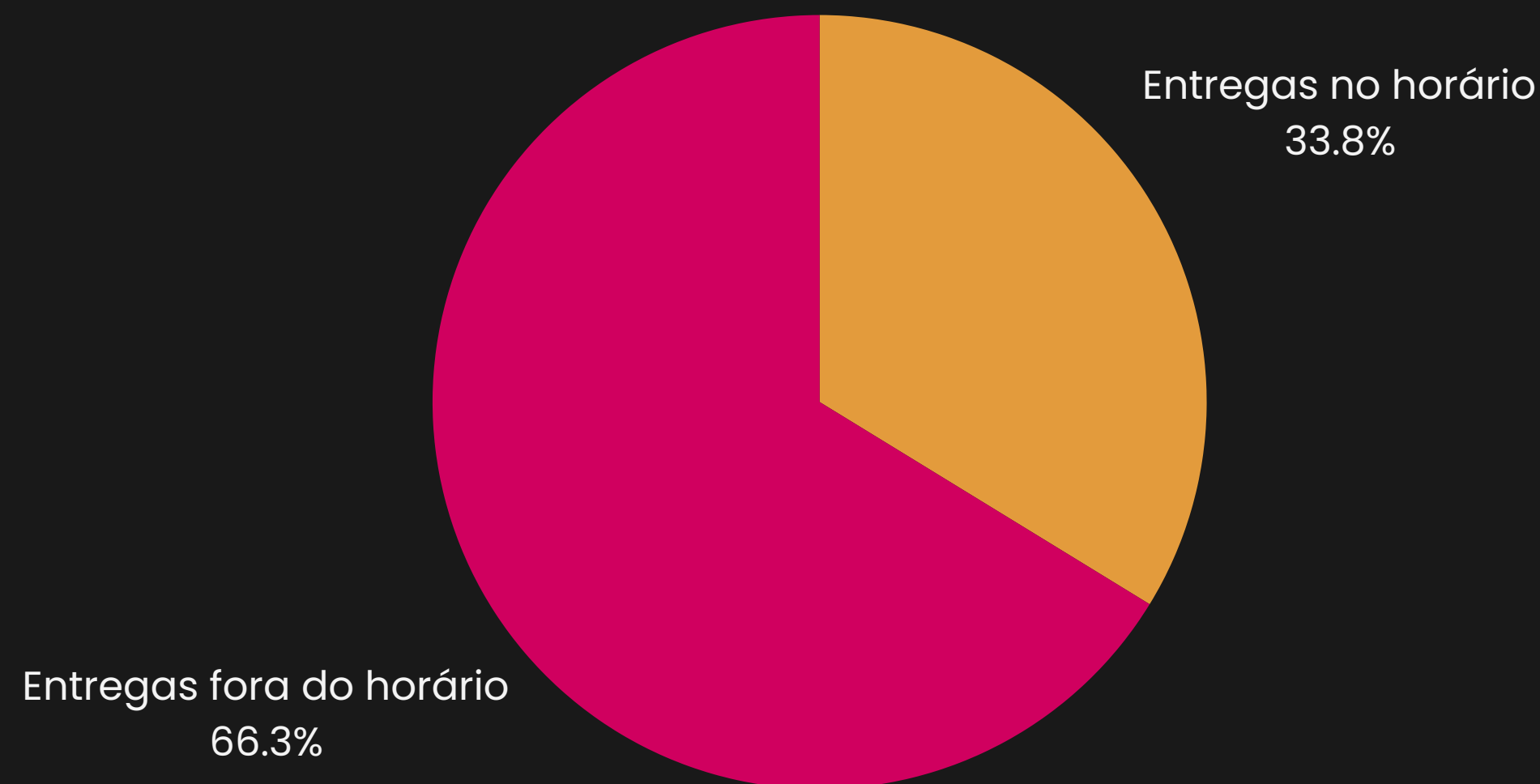
Os horários de entrega estabelecidos (das 9h às 13h ou das 14h às 18h) pelo InHome Ordering, com entregas agendadas, estão sendo respeitados?

[Voltar ao índice](#)



66% Atraso nas Entregas

- Entregas realizadas fora do horário estabelecido aumentam a exposição a furtos (porch piracy).

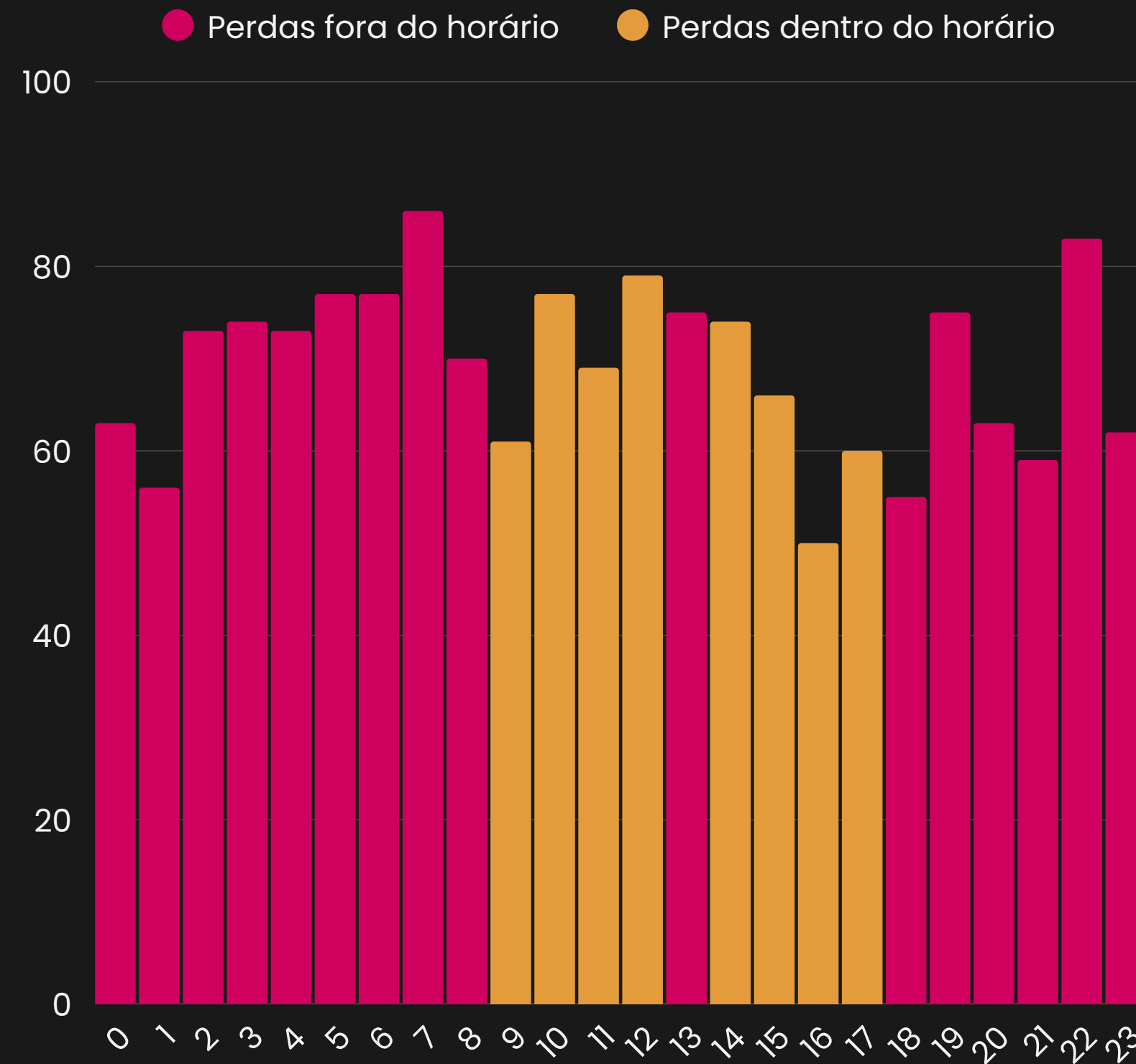


[Voltar ao índice](#)



Perdas em Horários Críticos

- Os horários das perdas indicam que os atrasos são significativos. Além disso, as perdas ocorridas dentro do horário demonstram que esse fator não é o único determinante das perdas.



[Voltar ao índice](#)



Motoristas

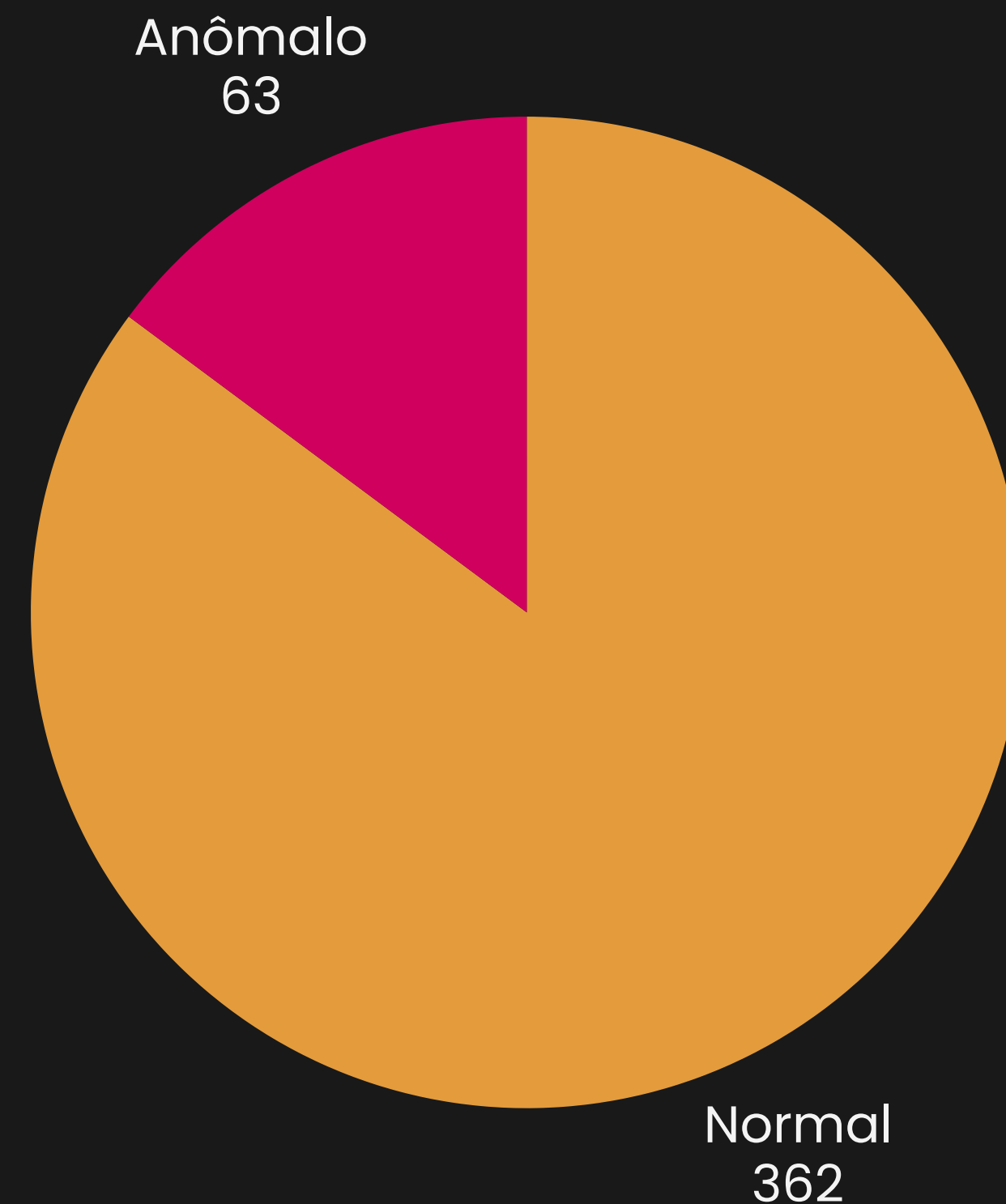
Motoristas que apresentam **padrões de perda atípicos** em comparação com os demais foram identificados por meio de métodos de detecção de anomalias e **classificados como anômalos**.

[Voltar ao índice](#)



15% apresentam padrão atípico.

- Dos 425 motoristas que registraram pelo menos um extravio, 63 apresentam padrões anormais de perda.
- Embora não possamos concluir categoricamente que esses motoristas tenham cometido fraude, levantamos uma **Red Flag**.

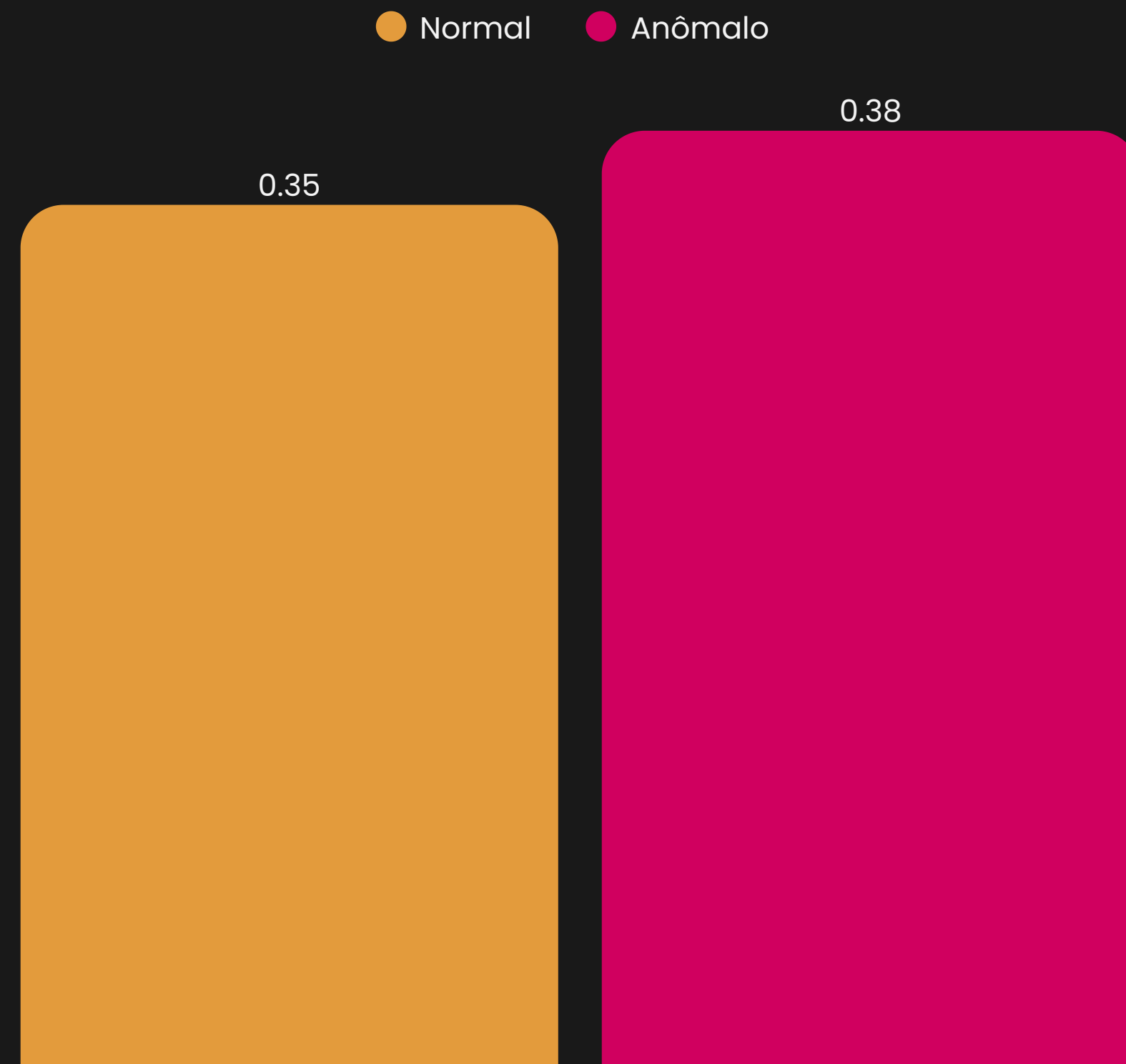


[Voltar ao índice](#)



8.6% mais perdas por pedido

- Motoristas anômalos registram em média 8.6% mais perdas em delivery de pedidos.



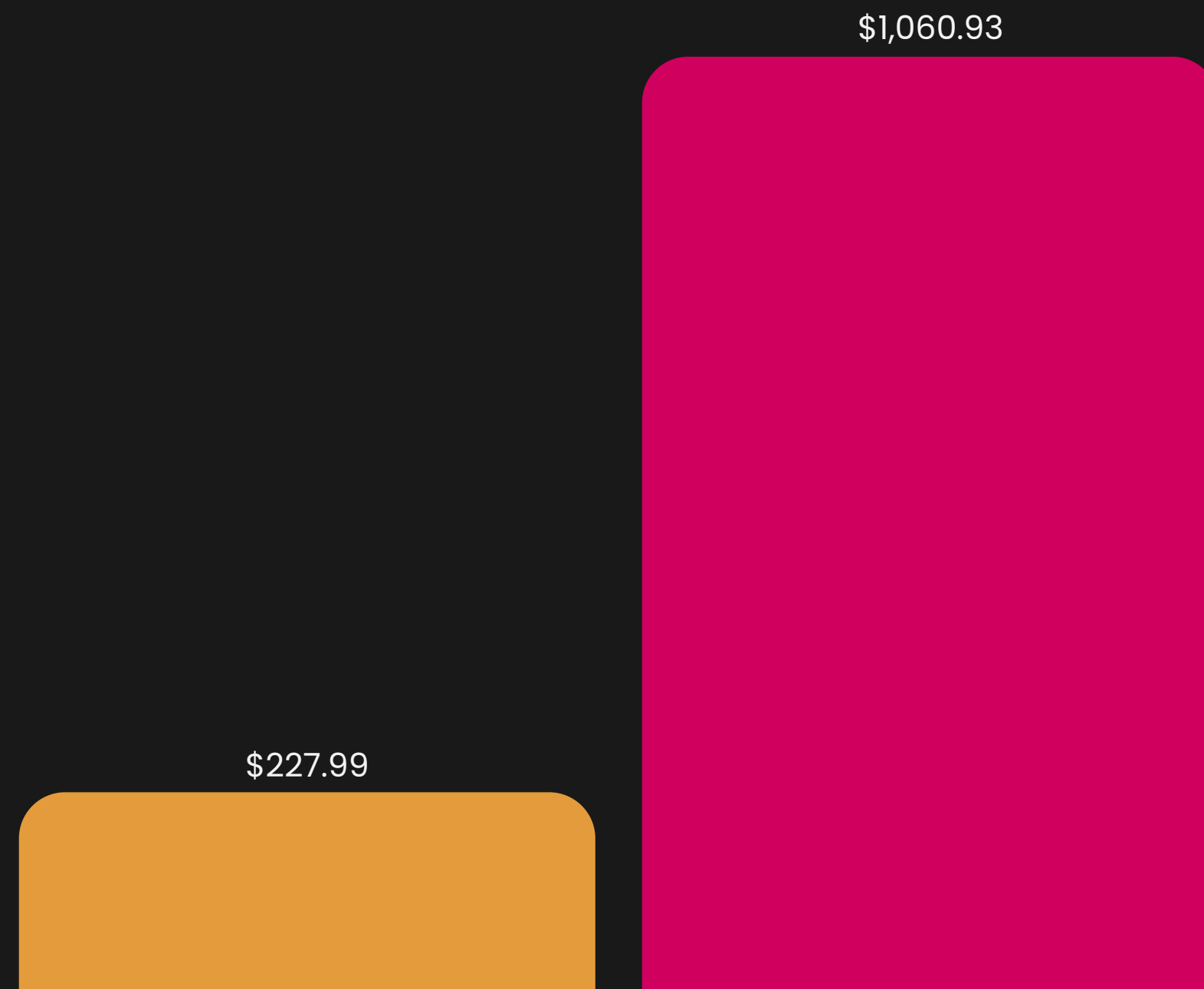
[Voltar ao índice](#)



● Normal ● Anômalo

Prejuízo Médio 4.7x maior

- Motoristas anômalos tem em média mais prejuízo acumulado do que os demais motoristas que registraram perdas.



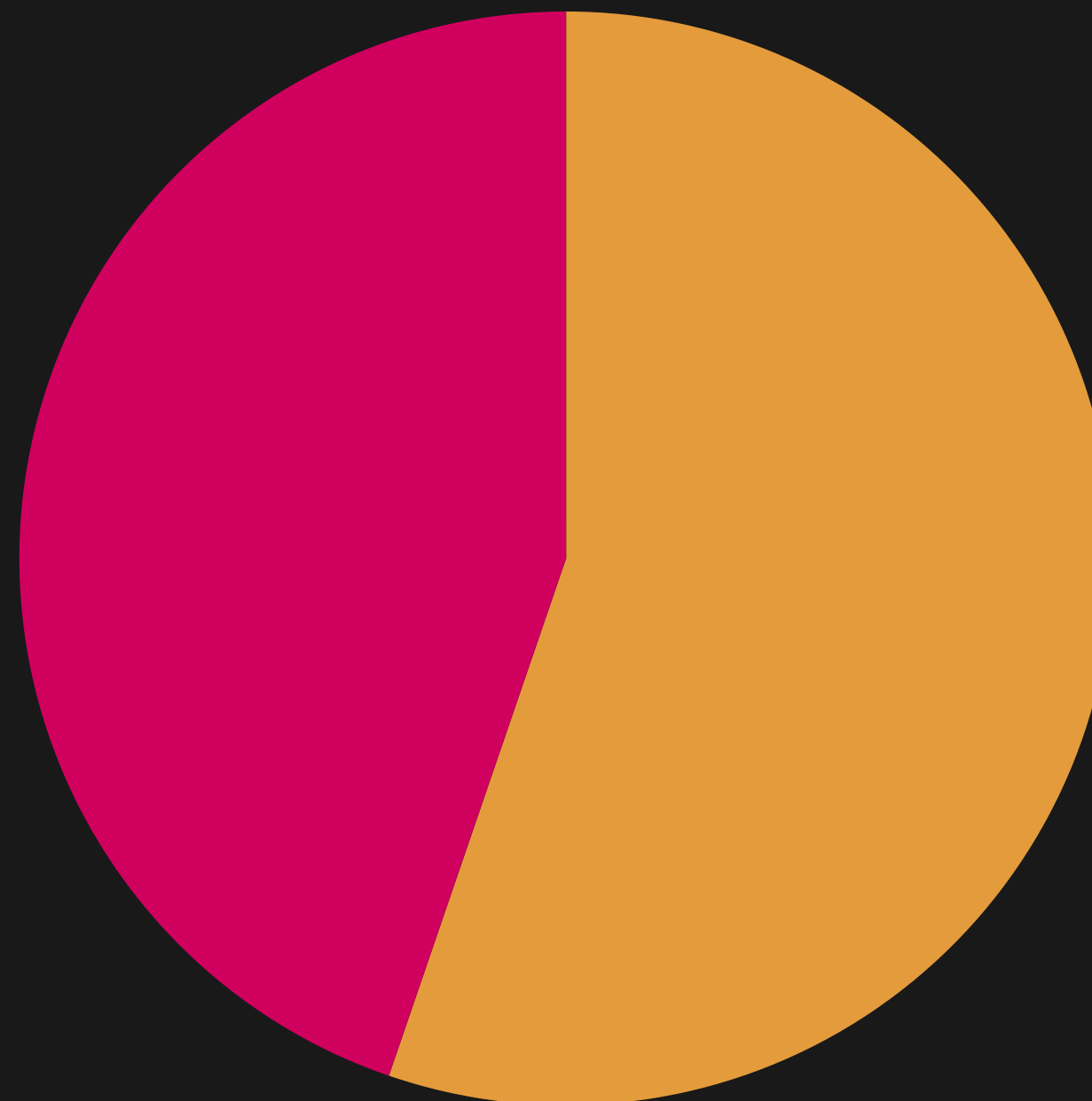
[Voltar ao índice](#)



45% Prejuízo Anual

- Os motoristas anômalos estão associados a 45% de todo o prejuízo gerado.

Anômalo
\$66,838.64



Normal
\$82,533.97

[Voltar ao índice](#)



Padrões de Perda Elevados

- Outros aspectos relacionados aos padrões de perda desse grupo foram identificados destacando o risco associado a esses motoristas.

📦 Média de itens ausentes por pedido: **+9%**

✗ Total de itens ausentes: **+11%**

⚠️ Proporção de itens ausentes total: **+33%**

[Voltar ao índice](#)



Problemas Cadastrais

- Há perfis duplicados. Motoristas com o mesmo nome e idade, mas com IDs diferentes, estão operando simultaneamente, o que pode indicar um erro de cadastro ou uma possível tentativa de fraude.

ID	Nome	Idade
WDID10082	Brittany Morgan	19
WDID10282	Brittany Morgan	19
WDID10222	Daniel Hall	20
WDID10434	Daniel Hall	20
WDID09902	Patricia Vance	18
WDID10447	Patricia Vance	18
WDID10506	Timothy Brown	37
WDID10517	Timothy Brown	37

[Voltar ao índice](#)



Clientes

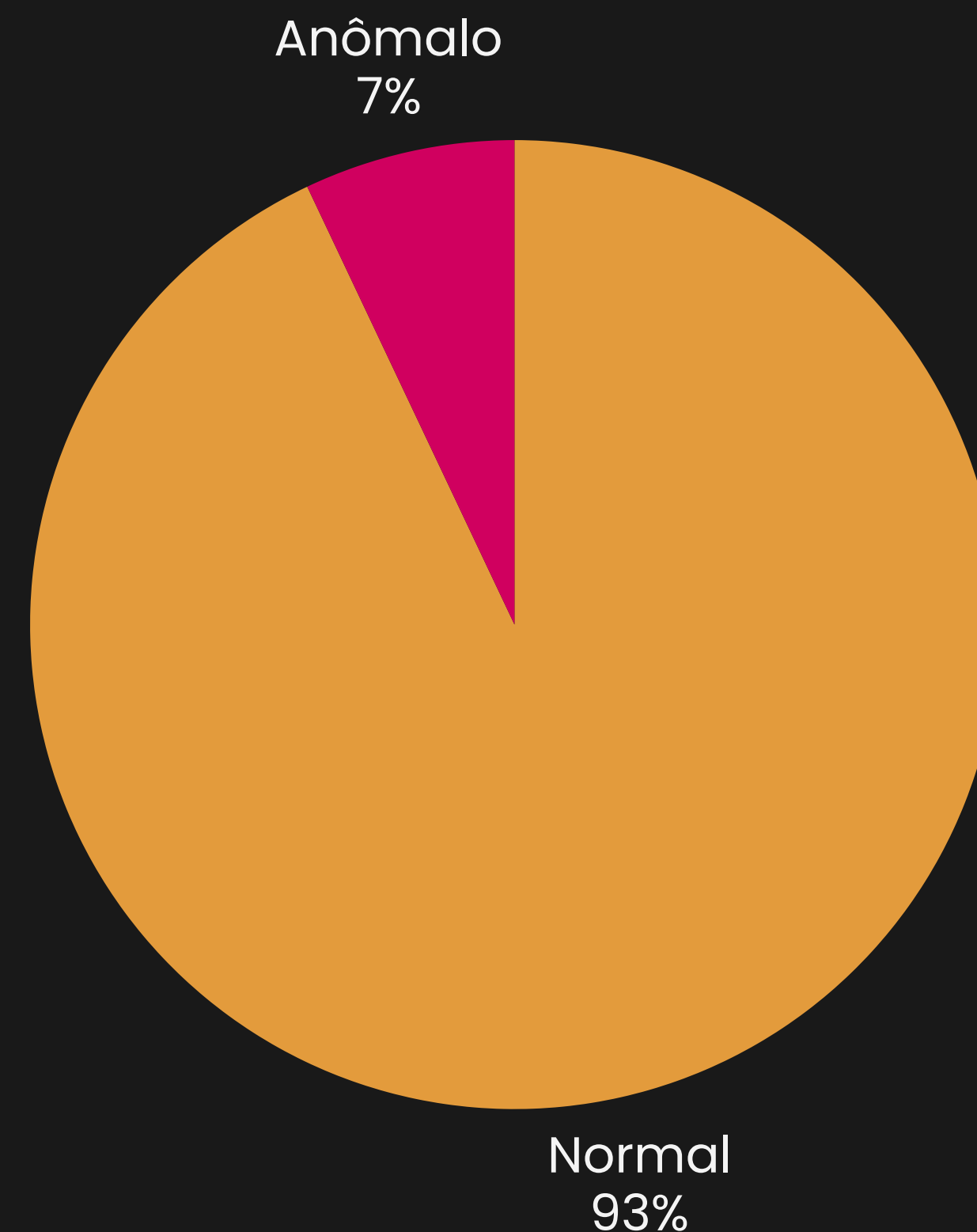
Os clientes também demonstram padrões de perda atípicos em comparação com os demais.
Através da detecção de anomalias, realizamos a classificação desses padrões.

[Voltar ao índice](#)



15% apresentam padrão atípico.

- Dos 820 clientes que registraram pelo menos um extravio, 62 apresentam padrões anormais de perda.
- Embora não possamos concluir categoricamente que esses clientes tenham cometido fraude, levantamos uma **Red Flag**.



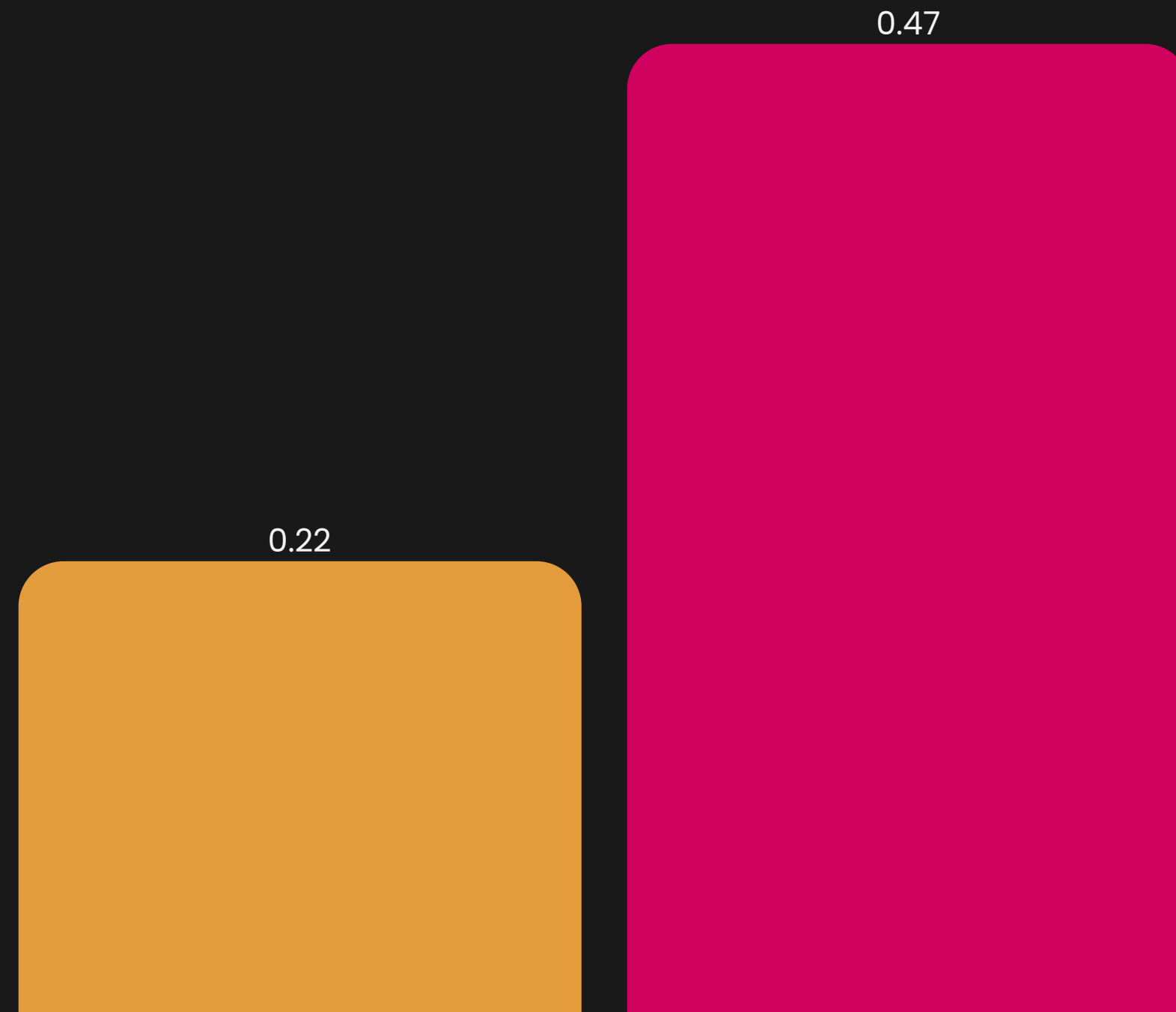
[Voltar ao índice](#)



2x mais perdas por pedido

- Clientes anômalos registram em média 2x mais perdas em seus pedidos.

● Normal ● Anômalo



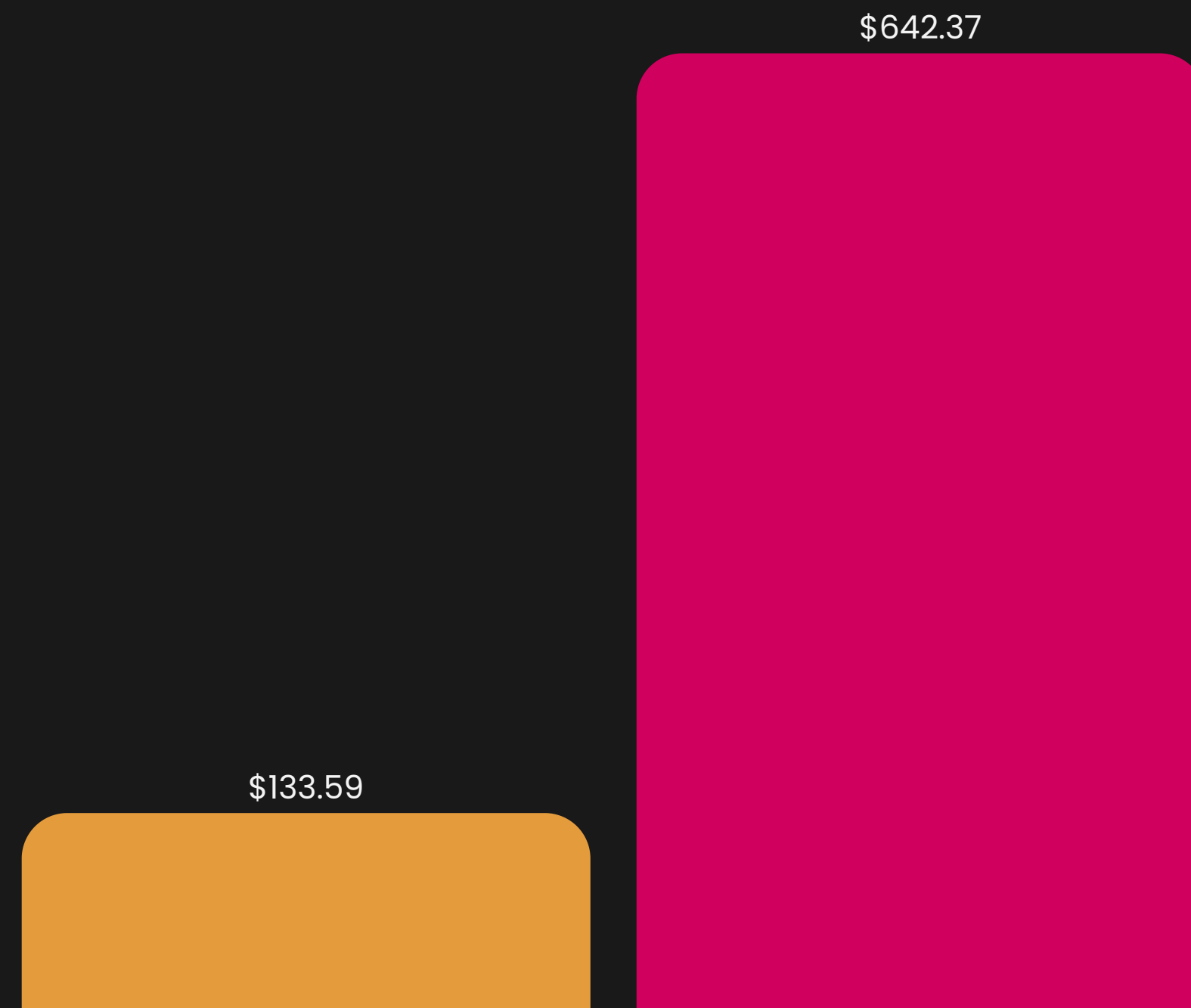
[Voltar ao índice](#)



Prejuízo Médio 4.8x maior

- Clientes anômalos tem em média mais prejuízo acumulado do que os demais clientes que registraram perdas.

● Normal ● Anômalo



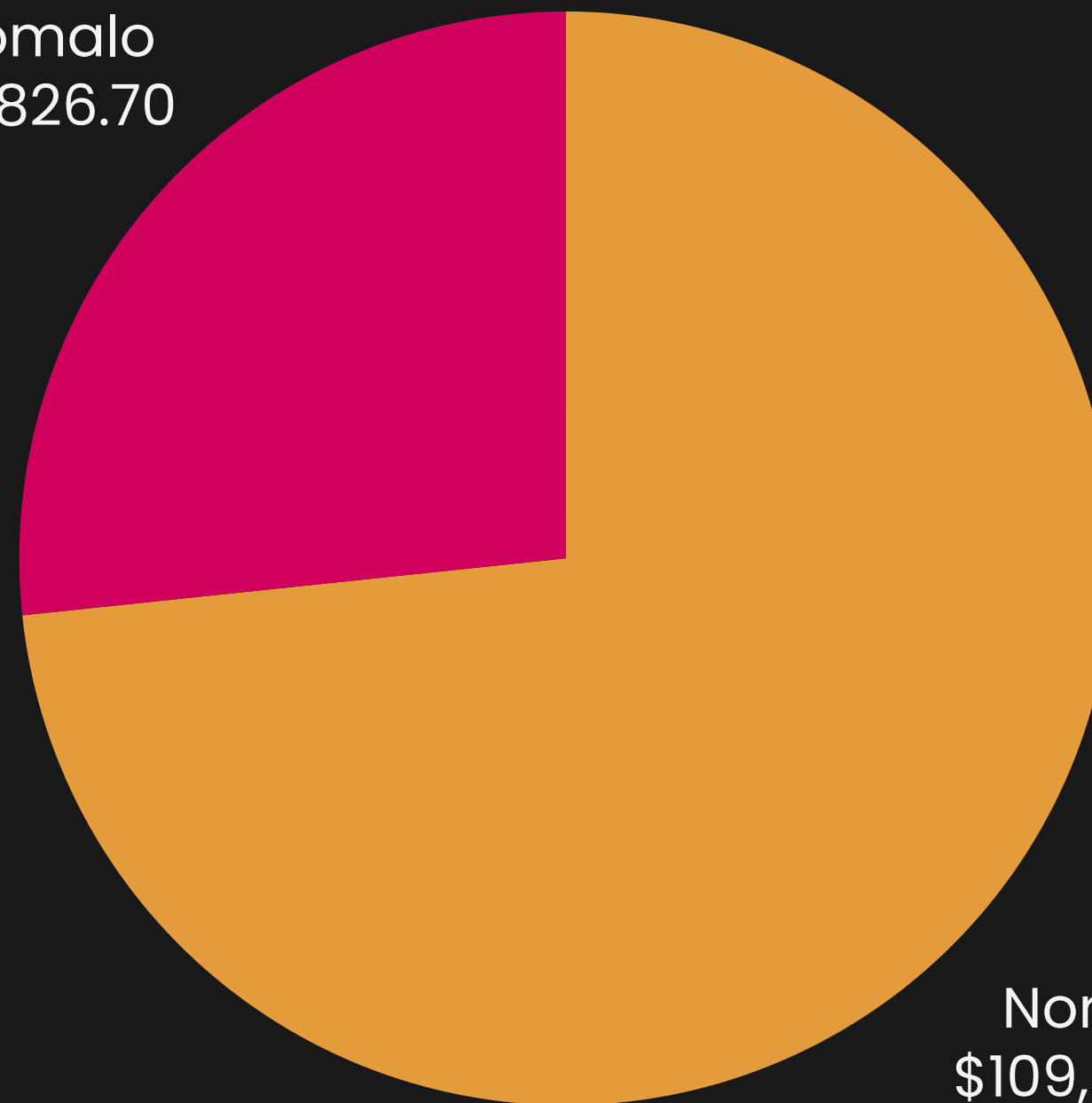
[Voltar ao índice](#)



27% Prejuízo Anual

- Os clientes anômalos estão associados a 27% de todo o prejuízo gerado.

Anômalo
\$39,826.70



Normal
\$109,545.91

[Voltar ao índice](#)



Padrões de Perda Elevados

- Outros aspectos relacionados aos padrões de perda desse grupo foram identificados destacando o risco associado a esses clientes.

📦 Média de itens ausentes por pedido: **+106%**

✗ Total de itens ausentes: **+86%**

⚠️ Proporção de itens ausentes total: **+112%**

[Voltar ao índice](#)



Interação

Há uma associação entre motoristas e clientes anômalos ?
Existe cooperação entre as partes sugerindo um sistema de fraude estruturada?

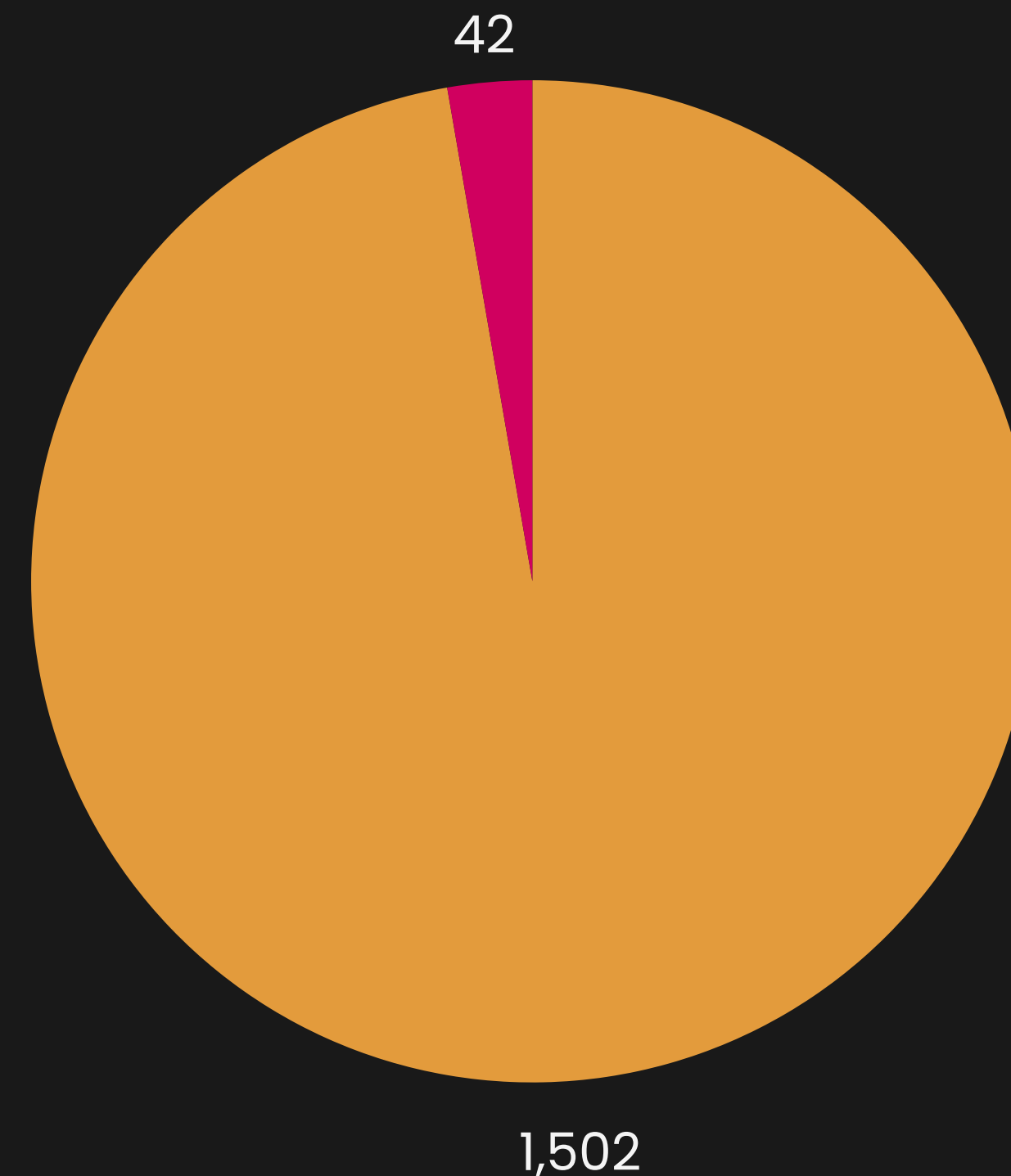
[Voltar ao índice](#)



2.7%

Co ocorrência em pedidos

- Dos 1.502 pedidos com extravios, 42 envolveram tanto clientes quanto motoristas suspeitos (anômalos). Não há registros de múltiplas transações entre pares específicos, o que reduz a hipótese de fraude estruturada.



[Voltar ao índice](#)



Solução Proposta

Estratégias para mitigar falhas e melhorar a eficiência

[Voltar ao índice](#)

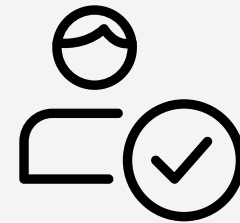
Solução Proposta

[Voltar ao índice](#)



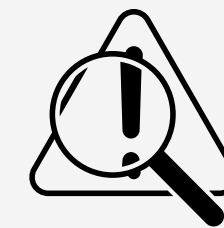
Política de Entrega

- Estabelecer regras para delivery de produtos vulneráveis como entrega em mãos com confirmação do cliente.



Elegibilidade de Entrega

- Classificar os motoristas de acordo com desempenho e alocá-los de maneira estratégica para as entregas.



Deteccção de Anomalias

- Implementar a deteção de anomalias para motoristas e clientes como uma camada extra de segurança.

Politica de Entrega

Implementar uma política de entrega **assinatura** ou **senha do cliente** para produtos mais vulneráveis e que geram mais impacto na receita em caso de extravio. Para isso foi necessário **identificar e classificar os produtos vulneráveis**.

Impacto

- 🛡️ Impossibilita a fraude e o furto.
- 💰 Reduz perdas e prejuízos.
- ✅ Melhora a qualidade do serviço.



[Voltar ao índice](#)

Classificação de Produtos Aplicada

[Voltar ao índice](#)

Classificação elaborada com base na incidência de perda e no impacto financeiro.
Método: Clusterização com KMeans.

Classe do Produto	Categoria do Produto	Produtos	Valor Médio por Item Perdido	Tipo de Confirmação
Premium	Eletronics	12	\$740.56	Senha
High Value	Eletronics	7	\$517.03	Senha
Mid Value	Eletronics	11	\$248.74	Assinatura
Low Value	Bakery, Beverages, Dairy, Frozen, Pantry, Produce, Snacks, Supermarket	115	\$11.46	Foto
Low Risk	Bakery, Beverages, Dairy, Electronics, Frozen, Household, Pantry, Personal Care, Snacks	169	Sem Perdas	Foto



Porque usar a Classificação dos Produtos ?

Vantagens da Classificação de Produtos em Relação à Categoria de Produtos.

[Voltar ao índice](#)

Eficácia	Precisão	Consistência
<ul style="list-style-type: none">A Classificação dos Produtos é multidimensional (perdas e prejuízo) e identifica padrões de risco 25 vezes melhor que as categorias, oferecendo uma base mais robusta para reduzir perdas.	<ul style="list-style-type: none">Separa os produtos em grupos mais distintos, como "alto risco" e "baixo risco", facilitando decisões sobre quais precisam de mais cuidado.	<ul style="list-style-type: none">Produtos dentro de cada grupo têm prejuízos mais parecidos, tornando as estratégias mais previsíveis e confiáveis.

Elegibilidade de entregas

Desenvolver um **sistema de delivery baseado no desempenho dos motoristas**. Para isso foi necessário **classificar os motoristas**. O objetivo é alocar motoristas mais eficientes a itens mais vulneráveis e vice-versa.

Impacto

- 🛡 Diminui a probabilidade de extravios
- 💰 Reduz perdas e prejuízos.
- ✅ Não requer mais contratações.



[Voltar ao índice](#)

Detecção de Anomalias

A implementação do sistema de detecção de anomalias permite a identificação de indivíduos que apresentam padrões atípicos e probabilidade significativamente elevada de perda. Sugere restrições e medidas de segurança adicionais.

Impacto

- 🛡 Identifica padrões de incomuns de perda.
- 💰 Reduz perdas e prejuízos.
- ✅ Melhora a qualidade do serviço.

[Voltar ao índice](#)



Classificação dos Motoristas e Detecção de Anomalias

Quais são as vantagens de usar a Classificação dos Motoristas em conjunto com a Detecção de Anomalias em relação à Quantidade de Itens perdidos por Motorista ?

[Voltar ao índice](#)

Eficácia	Precisão	Consistência	Segurança
<ul style="list-style-type: none">A Classificação dos Motoristas é multidimensional (perdas, valor, horário) e identifica padrões de risco 10 vezes melhor que a contagem de itens perdidos, oferecendo uma base mais sólida para gerenciar entregas.	<ul style="list-style-type: none">Separa os motoristas em grupos distintos, como "Elite" e "Alto Risco", facilitando a escolha de quem entrega produtos valiosos.	<ul style="list-style-type: none">Motoristas dentro de cada grupo têm desempenhos mais semelhantes, tornando as decisões mais seguras e previsíveis.	<ul style="list-style-type: none">A detecção de anomalias gera segurança adicional ao identificar padrões incomuns de perda em motoristas e clientes.

Classificação de Motoristas Aplicada

[Voltar ao índice](#)

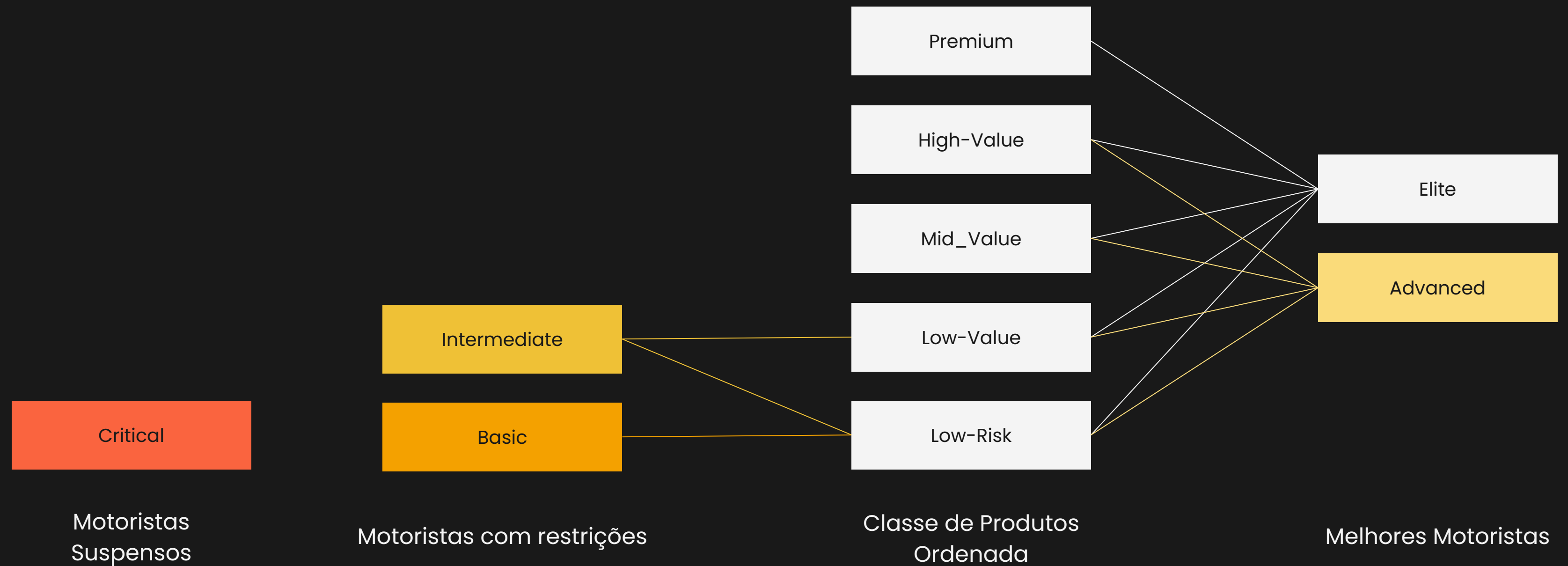
Métrica multidimensional baseada no prejuízo, perdas e a taxa de entregas atrasadas.

Método: Engenharia de atributos.

Desempenho por Classe de Motorista	Motoristas	Itens perdidos	Prejuízo Total	Prejuízo médio por Motorista
Elite	212	0	\$0	\$0
Advanced	610	0	\$0	\$0
ItermEDIATE	213	766	\$14,972.81	\$70.29
Basic	115	494	\$38,982.57	\$338.98
Critical	97	397	\$95,417.23	\$983.68

Elegibilidade de Entrega

O sistema sugerido para a alocação estratégica de motoristas e produtos propõe níveis de acesso baseados na classe de produto, de acordo com o desempenho de cada motorista.





Teste de Hipótese

As propostas de mudança na política de entrega, o sistema de elegibilidade de entregas e a detecção de anomalias são eficazes para contornar as perdas e prejuízo no ecommerce ?

[Voltar ao índice](#)



Abordagem do Teste

Tipo de abordagem e mensuração.

Experimento Randomizado Controlado – RCT

- O método é eficaz para avaliar ganhos das propostas, gerando evidências validadas por testes estatísticos sobre a eficácia das intervenções.
- O experimento será multifacetado, medindo simultaneamente duas intervenções: o sistema de elegibilidade de entregas e dispositivos de confirmação de entrega. Criar grupos distintos exigiria mais amostras e tempo, dificultando insights acionáveis. Recomenda-se esse procedimento inicial, seguido por um futuro RCT fatorial para avaliar o efeito individual de cada intervenção.

[Voltar ao índice](#)



Tipo de Amostra

Determinando o tipo de amostra para o teste

◆ Prejuízo Médio por Pedido

- Definir a quantidade de pedidos como unidade de amostra e mensurar o prejuízo médio sobre os pedidos prioriza a abundância de amostras o que é essencial no contexto.
- Possibilita um menor período para coleta e conclusão para rápida tomada de ação sem abrir mão do rigor do teste.

[Voltar ao índice](#)



Planejamento RCT

Configuração e resultado do cálculo do tamanho de amostra (Sample Size Calculator)

- **Duração:** 3 meses
- **Período:** Jan/Mar 2024
- **Tipo de amostra:** Quantidade de pedidos
- **Métrica mensurada:** Prejuízo médio por pedido
- **Estimativa de amostras:** 2,438
- **Período Referência:** Jan/Mar 2023 (baseline)

Experimento RCT – Especificações	
Total de amostras necessárias	2,178
Amostras por grupo (Controle e Teste)	1,089
Prejuízos médio por item (Baseline: 1º trimestre 2023)	\$18.44
Prejuízo médio esperado após a intervenção	\$6.45
Efeito mínimo detectável (Redução mínima esperada)	\$-11.98
Tamanho do efeito	65%
Significância estatística	95%
Poder do teste	80%

[Voltar ao índice](#)



Avaliação de resultados

Especificações e número de amostras estimado.

1. **Coleta de Amostra**
2. **Revisão do Cálculo Sample Size**
3. **Teste de Hipótese**
4. **Conclusão do Experimento**
5. **Verificação do Objetivo**

1. **Coleta de Amostras**: É fundamental monitorar constantemente o número de amostras coletadas, pois a quantidade necessária pode ser alcançada antes do prazo estabelecido.
2. **Revisão do Cálculo Sample Size**: Recalcular o Sample Size após a coleta se houver variações significativas no desvio padrão ou média, garantindo poder estatístico.
3. **Teste de Hipótese**: Verificar a normalidade dos dados e aplicar o teste estatístico apropriado, comparando a média de prejuízo por pedido entre o grupo de controle e o grupo de teste.
4. **Conclusão do Experimento**: Apresentar os resultados do experimento, confirmando ou refutando a eficácia do tratamento proposto.
5. **Verificação do Objetivo**: Comparar a redução observada no prejuízo com a meta estabelecida (30%).

[Voltar ao índice](#)

Contato

[Voltar ao índice](#)



E-mail

pedrokaneko.ds@gmail.com

Redes sociais

[linkedin.com/in/pedrokaneko](https://www.linkedin.com/in/pedrokaneko)

github.com/pedro-kaneko



Apêndice Técnico

[Voltar ao índice](#)

Comparação ANOVA – Produtos

[Voltar](#)

Evidência estatística da Superioridade da Classificação por Clusterização sobre Categorias de Produtos.

Métrica	Categorias de Produtos	Classificação dos Produtos	Comparação (Ganho)
Valor F	74,21	1840,29	25x maior: clusters separam prejuízos com mais clareza.
P-value	0	0	Ambos significativos ($P < 0,05$), sem diferença prática.
MS Between	\$45.381.286,67	\$153.339.315	3,4x maior: clusters explicam mais variabilidade entre grupos.
MS Within	\$611.522,55	\$83.323,43	7,3x menor: clusters têm maior homogeneidade interna.
SS Between	\$453.812.866,70	\$613.357.259,90	35% maior: clusters capturam 96% da variância total vs. 71%.
SS Within	\$185.291.332,40	\$25.746.939,20	7,2x menor: menos variação inexplicada nos clusters.

Comparação ANOVA – Motoristas

[Voltar](#)

Evidência da Superioridade Estatística da Classificação de Motoristas por Risco sobre Quantidade de Itens Perdidos.

Métrica	Quantidade de Itens Perdidos	Classe de Motorista	Comparação (Ganho)
Valor F	151,49	1533,47	10x maior: classes separam riscos com mais clareza.
P-value	0	0	Ambos significativos ($P < 0,05$), sem diferença prática.
MS Between	\$8.894.855,90	\$22.558.757,81	2,5x maior: classes explicam mais variabilidade entre grupos.
MS Within	\$58.717,07	\$14.710,95	4x menor: classes têm maior homogeneidade interna.
SS Between	\$35.579.423,58	\$90.235.031,23	2,5x maior: classes capturam 83% da variância total vs. 33%.
SS Within	\$72.926.607,10	\$18.270.999,46	4x menor: menos variação inexplicada nas classes.

Classificação dos Motoristas

[Voltar](#)

Desenvolvimento da Classificação dos Motoristas com engenharia de atributos.

$$DriverRiskScore = (\overline{\text{prejuízo por item perdido}}_{\text{norm}} \times 0.5) + (\overline{\text{taxa itens perdidos por pedido}}_{\text{norm}} \times 0.3) + (\overline{\text{taxa entrega fora horario}}_{\text{norm}} \times 0.2)$$

Pontuação (Driver Risk Score)	Classificação (Driver Risk Label)
<= 0.0	Elite
<= 0.1	Advanced
<= 0.2	Intermediate
<= 0.4	Basic
<= 0.5	Critical

- Driver Risk Score** é uma métrica que avalia o risco associado aos motoristas, levando em consideração o prejuízo por item perdido, a taxa de itens perdidos por pedido e a taxa de entregas realizadas fora do horário. Essa pontuação é obtida através da soma das médias das três métricas normalizadas e ponderadas de acordo com sua relevância no contexto.
- Driver Risk Label** classifica os motoristas em diferentes categorias com base em intervalos da pontuação de risco atribuída a eles.