

# Case Localiza

Pipeline de Limpeza e  
Análise de Dados de Fraude

# Sumário

Análise Exploratória	<b>Pág 3</b>
Validação de Qualidade	<b>Pág 4</b>
Limpeza e Tratamento	<b>Pág 5</b>
Pós-Processamento	<b>Pág 6</b>
Qualidade da Saída	<b>Pág 7</b>
Análise de Risco por Região	<b>Pág 8</b>
Maiores Transações	<b>Pág 9</b>

# Análise Exploratória dos Dados Brutos

## Entender a estrutura, volume e características iniciais do dataset

- Exibição do formato do DataFrame e dos primeiros e últimos dados para melhor entendimento
- Listagem das colunas da Base de Dados com suas informações

	timestamp	...	sending_address	...	risk_score	...	anomaly
0	1618185002	0x9d32d0bf2c00f41ce7ca01b66e174cc4dcb0c1da	...	...	18.75	...	low_risk
1	1698642474	0xd6e251c23cbf52dbd472f079147873e655d8096f	...	...	25.0	...	low_risk
2	1619180066	0x2e0925b922fed01f6a85d213ae2718f54b8ca305	...	...	31.25	...	low_risk
3	1591413882	0x93efefc25fcacf31d7695f28018d7a11ece55457f	...	...	36.75	...	low_risk
4	1611257295	0xad3b8de45d63f5cce28aef9a82cf30c397c6ceb9	...	...	62.5	...	moderate_risk

[5 rows x 13 columns]

# Validação de Qualidade dos Dados

## Identificar e quantificar os problemas de qualidade nos dados

- Valores ausentes
- Registros duplicados
- Categorias inválidas
- Valores fora da faixa esperada

```
--- VALIDAÇÃO DE QUALIDADE INICIAL ---
✓ Nenhum valor ausente encontrado
✓ Nenhum registro duplicado encontrado
Valores categóricos inválidos encontrados:
  transaction_type: ['phishing', 'scam']
  location_region: ['0']
✓ Todos os valores numéricos estão dentro da faixa esperada
```

# Limpeza e Tratamento

## Ações tomadas para a correção de problemas identificados

- Validação de Endereços e IPs
- Validação do 'risk\_score' e 'amount' com o preenchimento de valores inválidos com a mediana
- Padronização das localizações e remoção de registros com valores inválidos
- Conversão de formato do Timestamp

# Verificações Pós-Processamento

## Checagem final dos valores obtidos

- Verificar consistência da relação entre 'risk\_score' e 'anomaly'
- Busca por padrões suspeitos
- Busca por valores inválidos ou incomuns em colunas numéricas

```
--- VERIFICAÇÕES FINAIS DE CONSISTÊNCIA ---
✓ Consistência entre risk_score e anomaly: OK
⚠ Possível anomalia: 117116 transações grandes de usuários novos
✓ Valores de transações: OK
✓ Duração de sessões: OK
```

# Qualidade dos dados após tratamento

## O Impacto da Limpeza em Números

- Após tratamento, ainda foi possível levar em consideração 99.45% dos dados. Garantindo uma análise mais completa e evitando distorções por erros.

```
== MÉTRICAS FINAIS DE QUALIDADE ==
Registros iniciais: 1,048,575
Registros após limpeza: 1,042,780
Taxa de retenção: 99.45%
Registros removidos: 5,795
```

# Resultados Finais

**Tabela 1 - Análise de Risco por Região Geográfica**

Análise de Risco por Regiões Geográficas				
	Região	Risco Médio	Desvio Padrão	Número de Registros
	América do Norte	45,13	21,78	210145
	América do Sul	45,11	21,69	207821
	Asia	44,97	21,73	208742
	Africa	44,88	21,82	206323
	Europa	44,56	21,57	209749

# Resultados Finais

**Tabela 2 - Top 3 maiores transações mais recentes por endereço**

Análise de Transações por Endereço			
	Região	Risco Médio	Desvio Padrão
	0xfe2650f030f2c966775e11009cb 015e8852ecf4b	76.757,00	2024-01-02 06:44:13
	0xe37126a5b0724737b516b97d3fb9 2311021e235d	76.716,00	2023-12-29 11:17:48
	0x646142948a3add1d05edb7789d7 3a14ef261e109	76.667,00	2024-01-01 16:14:14