

1. Resumo

Este relatório detalha a análise estatística do valor médio dos cupons utilizados na plataforma PicMoney, com base em uma amostra robusta de 100.000 transações.

A análise conclui com 95% de confiança estatística que o **valor médio real** de todos os cupons utilizados pela população de clientes situa-se entre **R\$ 548,88 e R\$ 552,10**. A média amostral observada foi de R\$ 550,49.

A extrema precisão deste intervalo (uma variação de apenas +/- R\$ 1,61 em relação à média) fornece alta confiabilidade para o planejamento financeiro, avaliação de campanhas e estratégias de precificação.

2. Metodologia

A análise foi conduzida utilizando um script Python no ambiente Google Colab. Os dados foram extraídos de uma base CSV contendo 100.000 registros de transações ([PicMoney-Base_de_Tranca_es_-_Cupons_Capturados-100000 linhas \(1\).csv](#)).

As etapas do processo incluíram:

1. **Carga e Limpeza:** Os dados foram carregados, com o separador decimal (vírgula) sendo corrigido para o formato numérico padrão (ponto).
2. **Cálculo da Média:** A média amostral foi calculada para a coluna `valor_cupom`.
3. **Cálculo do Intervalo de Confiança (IC):** Foi calculado um IC de 95% para a média populacional, utilizando a distribuição t de Student (dada a grande amostra, os resultados são equivalentes à distribuição Normal/Z).

3. Resultados Principais

Os resultados estatísticos da análise da coluna `valor_cupom` são os seguintes:

- **Tamanho da Amostra (N):** 100.000 transações
- **Média Amostral:** R\$ 550,49
- **Nível de Confiança:** 95%
- **Intervalo de Confiança (IC 95%):** [R\$ 548,88 ; R\$ 552,10]

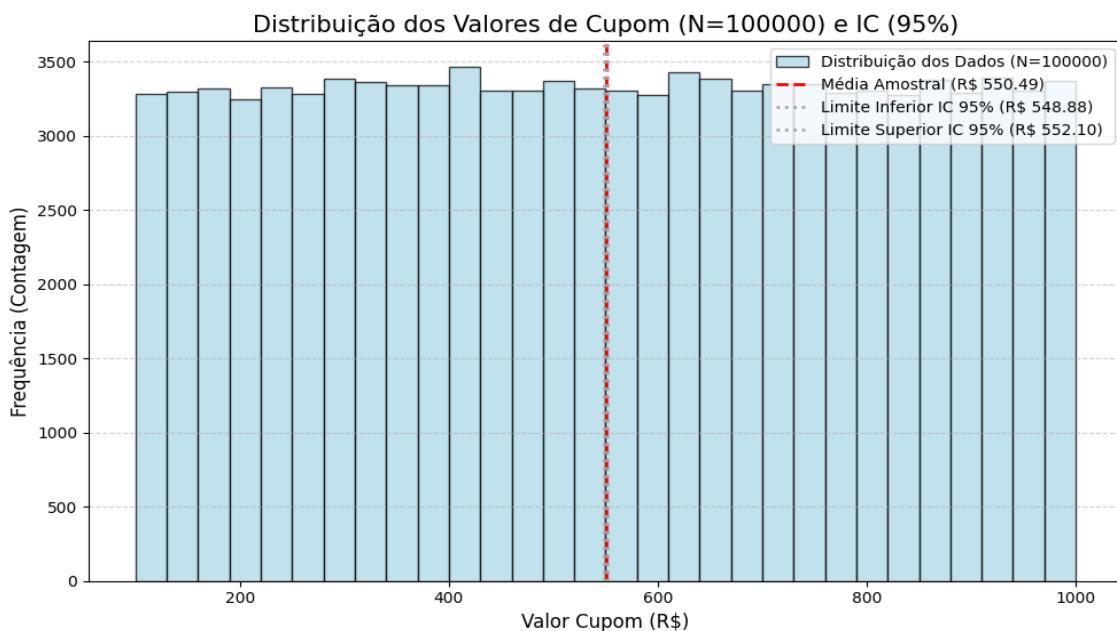
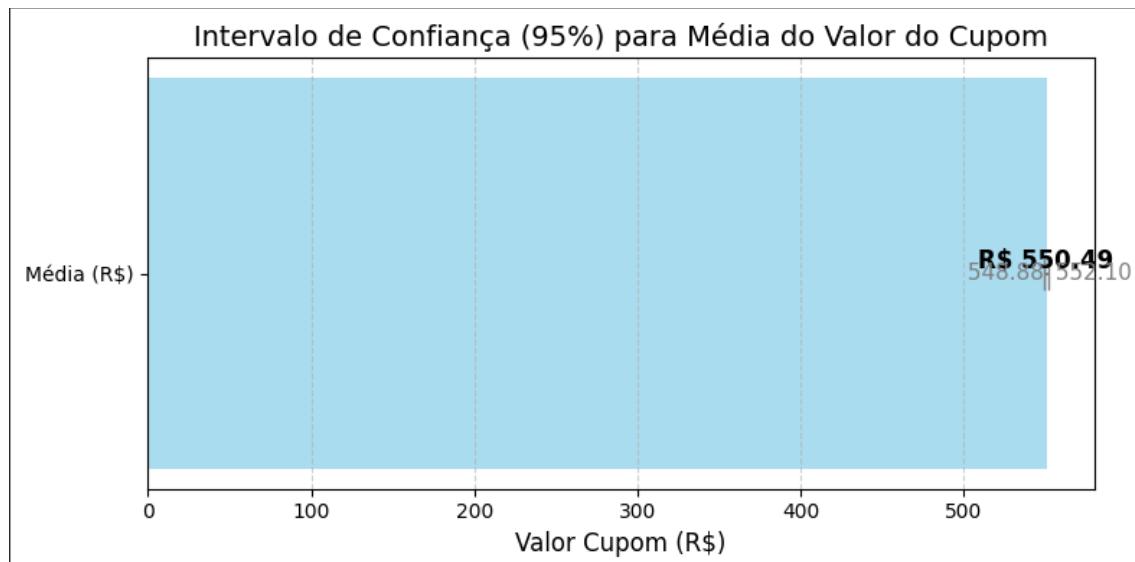
4. Interpretação e Implicações Estratégicas

A interpretação estatística padrão para este resultado é: "Se repetíssemos esta amostragem 100 vezes, 95 vezes o intervalo de confiança calculado conteria a média verdadeira da população".

- **Alta Precisão no Planejamento:** O grande tamanho da amostra ($N=100.000$) reduziu drasticamente a margem de erro. Temos uma estimativa extremamente precisa do comportamento do cliente.
- **Validação de Metas:** Qualquer meta de campanha que visava um valor médio de cupom em torno de R\$ 550 pode ser considerada um sucesso.
- **Previsibilidade de Custos (Budgeting):** Para fins de provisionamento de custos ou análise de rentabilidade de campanhas (ROI), a equipe financeira pode utilizar R\$ 550,49 como a estimativa mais provável, sabendo que o valor real dificilmente (menos de 5% de chance) estará fora da faixa de R\$ 548,88 a R\$ 552,10.

5. Visualização dos Dados

Os gráficos abaixo ilustram os resultados, mostrando a média amostral, os limites do intervalo de confiança e a distribuição dos valores analisados.



6. Conclusão

A análise estatística de 100.000 transações valida com alta precisão que o valor médio dos cupons da PicMoney está estabilizado em torno de R\$ 550,49.

7. Código

```
import pandas as pd
import numpy as np
import scipy.stats as st
import matplotlib.pyplot as plt

file_name = "PicMoney-Base_de_Transações---Cupons_Capturados-100000
linhas (1).csv"

try:
    df = pd.read_csv(file_name,
                      encoding='latin1',
                      engine='python',
                      on_bad_lines='skip',
                      sep=';')

    print("--- Dados Carregados com Sucesso ---")
    print(df.head())
    print("\n--- Informações das Colunas ---")
    df.info()

    if 'valor_cupom' not in df.columns:
        print("\n*** ERRO CRÍTICO ***")
        print("A coluna 'valor_cupom' não foi encontrada no CSV.")
        print("Verifique os nomes das colunas lidas:")
        print(df.columns)
    else:
        print("\n--- Limpando coluna 'valor_cupom' ---")

        if df['valor_cupom'].dtype == 'object':
            print("Coluna 'valor_cupom' é texto. Convertendo ',' para '...' ")
            df['valor_cupom'] =
df['valor_cupom'].astype(str).str.replace(',', '.', regex=False)

        df['valor_cupom'] = pd.to_numeric(df['valor_cupom'],
                                         errors='coerce')

        df.dropna(subset=['valor_cupom'], inplace=True)

    print("Conversão para numérico concluída.")
```

```

print(f"Total de linhas válidas para análise: {len(df)}")

print("\n--- Dados Prontos para Análise ---")

print("\n" + "="*50 + "\n")
print("--- Cálculo: IC para Média de 'valor_cupom' ---")

confidence_level_mean = 0.95

data_mean = df['valor_cupom']

mean = np.mean(data_mean)
std_err_mean = st.sem(data_mean)
n_mean = len(data_mean)
degrees_freedom = n_mean - 1

ic_mean = st.t.interval(confidence_level_mean,
                       degrees_freedom,
                       loc=mean,
                       scale=std_err_mean)

print(f"Métrica: Média da coluna 'valor_cupom'")
print(f"Tamanho da Amostra (N): {n_mean}")
print(f"Média Amostral (R$): {mean:.2f}")
print(f"Nível de Confiança: {confidence_level_mean:.0%}")
print(f"Intervalo de Confiança (IC): (R$ {ic_mean[0]:.2f},"
R$ {ic_mean[1]:.2f})")

print("\nInterpretação (Média):")
print(f"Com {confidence_level_mean:.0%} de confiança,"
podemos afirmar que a verdadeira média do 'valor_cupom' "
      f"para *toda* a população de transações (não apenas
desta amostra) "
      f"está entre R$ {ic_mean[0]:.2f} e R$ "
{ic_mean[1]:.2f}.")
print("Em outras palavras, se coletássemos 100 amostras
diferentes e calculássemos o IC para cada uma, "
      "esperaríamos que 95 delas contivessem a média
populacional real.")

print("\n" + "="*50 + "\n")
print("--- Gerando Gráficos ---")

erro_margem = mean - ic_mean[0]

plt.figure(figsize=(8, 4))

```

```

        plt.barh('Média (R$)', mean, xerr=erro_margem, capsize=8,
color='skyblue', ecolor='gray', alpha=0.7)
        plt.title('Intervalo de Confiança (95%) para Média do Valor
do Cupom', fontsize=14)
        plt.xlabel('Valor Cupom (R$)', fontsize=12)

        plt.text(mean, 'Média (R$)', f'R$ {mean:.2f}', ha='center',
va='bottom', color='black', fontsize=12, weight='bold')
        plt.text(ic_mean[0], 'Média (R$)', f'{ic_mean[0]:.2f}', ha='right', va='center', color='gray', fontsize=11)
        plt.text(ic_mean[1], 'Média (R$)', f'{ic_mean[1]:.2f}', ha='left', va='center', color='gray', fontsize=11)

        plt.grid(axis='x', linestyle='--', alpha=0.6)
        plt.tight_layout()
        plt.show()

plt.figure(figsize=(10, 6))
plt.hist(data_mean, bins=30, alpha=0.75,
label=f'Distribuição dos Dados (N={n_mean})', color='lightblue',
edgecolor='black')

plt.axvline(mean, color='red', linestyle='--', linewidth=2,
label=f'Média Amostral (R$ {mean:.2f})')
plt.axvline(ic_mean[0], color='darkgray', linestyle=':', linewidth=2, label=f'Limite Inferior IC 95% (R$ {ic_mean[0]:.2f})')
plt.axvline(ic_mean[1], color='darkgray', linestyle=':', linewidth=2, label=f'Limite Superior IC 95% (R$ {ic_mean[1]:.2f})')

plt.title(f'Distribuição dos Valores de Cupom (N={n_mean}) e IC (95%)', fontsize=15)
plt.xlabel('Valor Cupom (R$)', fontsize=12)
plt.ylabel('Frequência (Contagem)', fontsize=12)
plt.legend()
plt.grid(axis='y', linestyle='--', alpha=0.6)
plt.tight_layout()
plt.show()

print("\n--- Gráficos Gerados com Sucesso ---")

except FileNotFoundError:
    print(f"Erro: O arquivo '{file_name}' não foi encontrado.")
    print("Por favor, certifique-se de que você fez o upload do arquivo para o ambiente do Google Colab "
          "e que o nome corresponde exatamente.")
except Exception as e:

```

```
    print(f"Ocorreu um erro inesperado: {e}")  
    print("Verifique se o arquivo CSV está formatado corretamente  
(ex: separador de colunas).")
```