

Análise Exploratória: Marmitas em Ação

Dados referentes as despesas mensais do projeto.

Dados originais em formato xlsx: limpeza e organização dos dados

As informações de cada planilha do arquivo custos.xlsx foram reorganizadas e armazenadas em um único conjunto de dados.

Total de observações com despesas: 740

Variáveis:

Num_Nota:	Número da Nota/despesa realizada
Data:	Data em que a despesa foi registrada
Semana:	Dia da semana correspondente à data da despesa
Descricao:	Descrição do tipo de despesa realizada
Local:	Estabelecimento onde foi efetuada a despesa
Valor :	Valor monetário da despesa (R\$)

Padronização dos dados:

Todas as despesas foram classificadas e padronizadas (classificação em categorias e subcategorias, com isso, foram criadas as variáveis categóricas abaixo:

Id_categoria:	identificador único da categoria principal
Categoria:	Categoria principal/primeiro nível de classificação da despesa
Subcategoria:	Segundo nível de classificação da despesa

Script inicial para criação do banco de dados e respectivas tabelas para possibilitar o aprendizado com uso de ferramentas com conexão a banco de dados para a realização das análises

Custos.sql

Linguagem escolhida para realizar as análises: Python

Diagramas/gráficos gerados: Custos.xlsx

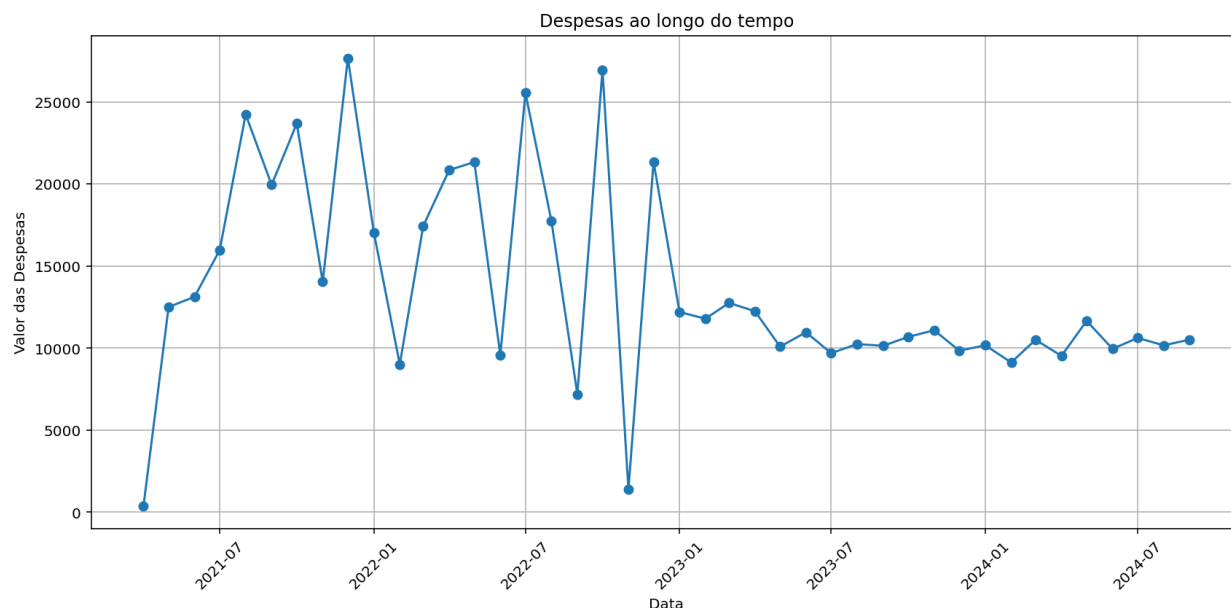


Gráfico 01 – Despesas ao longo do tempo

Pelo gráfico 01 - despesas acumuladas ao longo do tempo é possível perceber uma menor variabilidade, sugerindo um controle mais rígido das despesas a partir de janeiro de 2023.

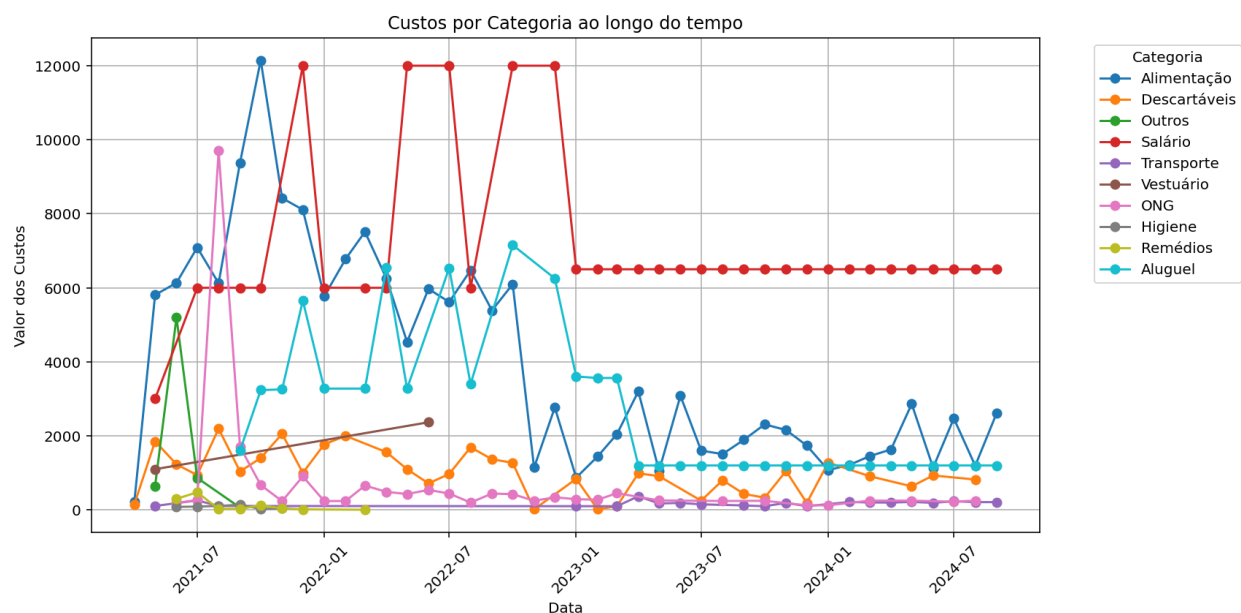


Gráfico 02 – Despesas por categoria ao longo do tempo

O gráfico 02 apresenta as despesas por categoria ao longo do tempo nos permite entender os custos operacionais antes e depois de janeiro de 2023. É visível a variabilidade dos gastos e que os maiores custos estão associados a salário, alimentação e aluguel.

Existe uma queda notável em relação aos custos de alimentação ao longo dos anos. Já com aluguel é perceptível uma tendência positiva até outubro de 2022 e após essa data uma forte tendência negativa e estabilização após abril de 2023

Existe uma oscilação salarial antes de janeiro de 2023 com o dobro dos custos em alguns meses, porém não foi possível identificar correlação com os demais custos das outras categorias.

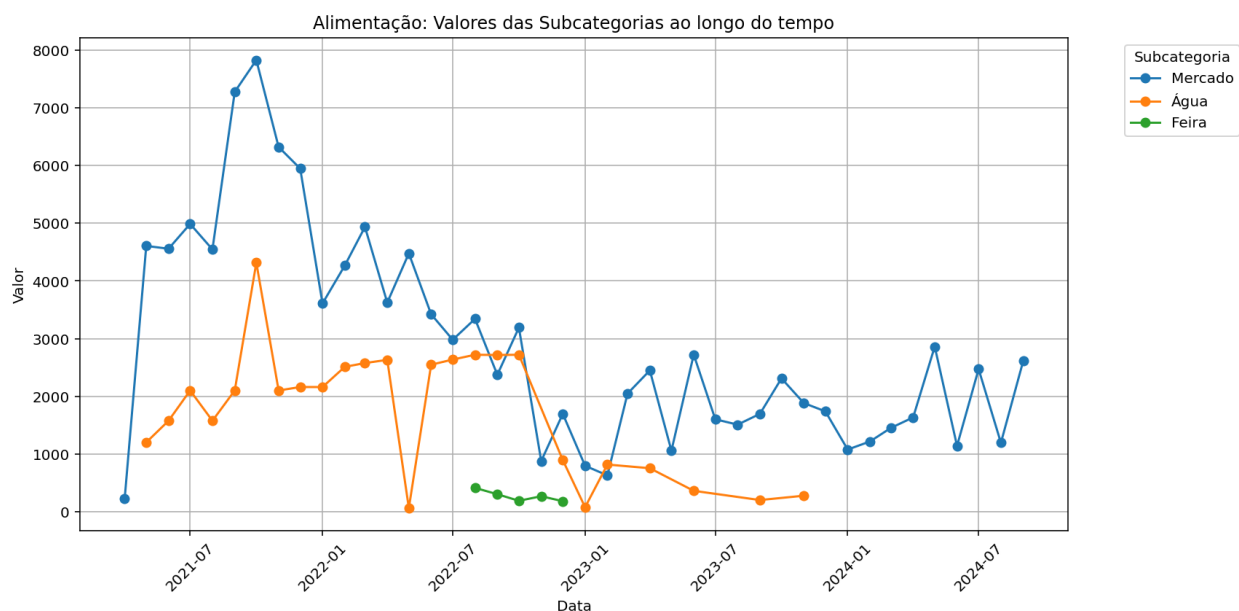


Gráfico 03 – Despesas de alimentação por subcategoria ao longo do tempo

No gráfico 03 foi possível identificar o impacto da compra de água nos custos de alimentação como um todo.

As despesas com água foram até novembro de 2023, onde a partir desta data não houve mais despesas com água. Sugerindo um foco maior em alimentação podendo estar correlacionado aos elevados custos com água ou redução na receita (doações). Contudo ainda não foram analisadas as doações para entender/confirmar esta suposição.

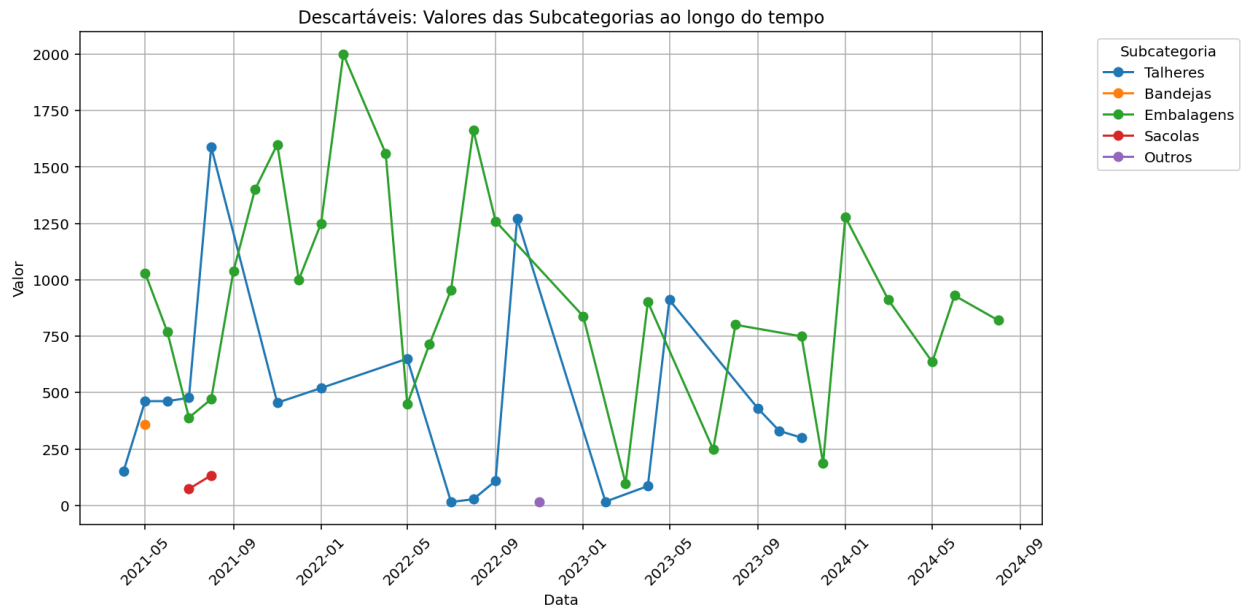


Gráfico 04 – Despesas com produtos descartáveis ao longo do tempo

Compra de talheres foi somente até novembro de 2023 como pode ser observado no gráfico 04. Sugerindo que foi mais uma decisão de negócios com foco na entrega de alimentação. Outro ponto a ser investigado.

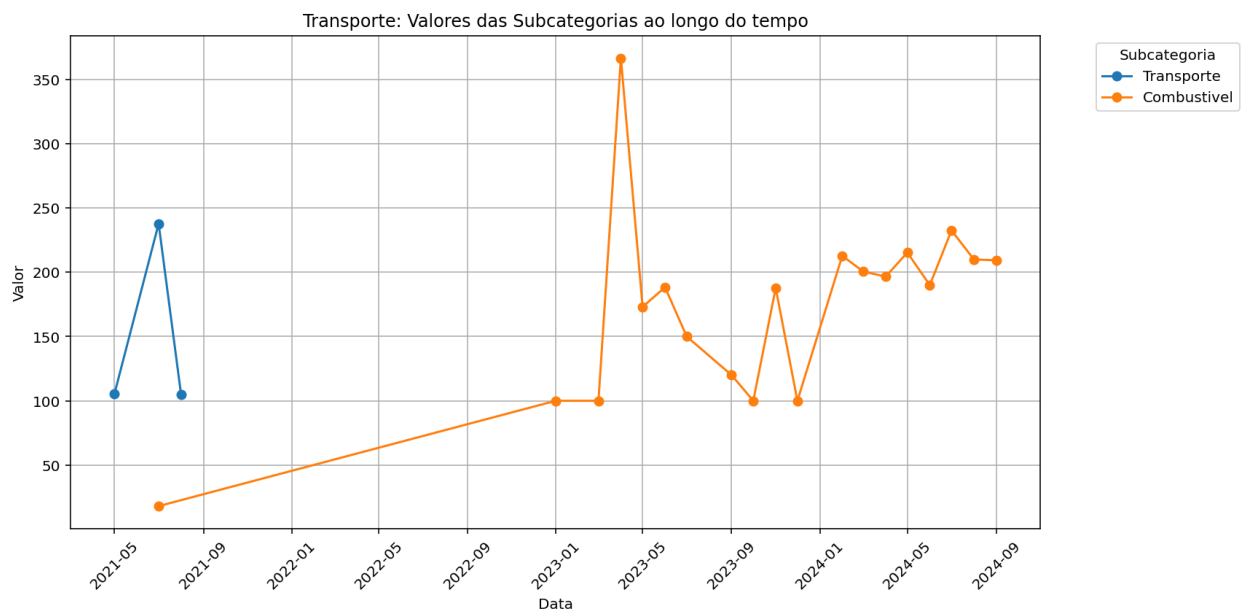


Gráfico 05 – Despesas com transporte por subcategorias ao longo do tempo

Ausência de despesas com combustíveis em alguns meses. Esse controle começou efetivamente em janeiro de 2023.

Uso de outras alternativas de transporte (Uber) no início do projeto.

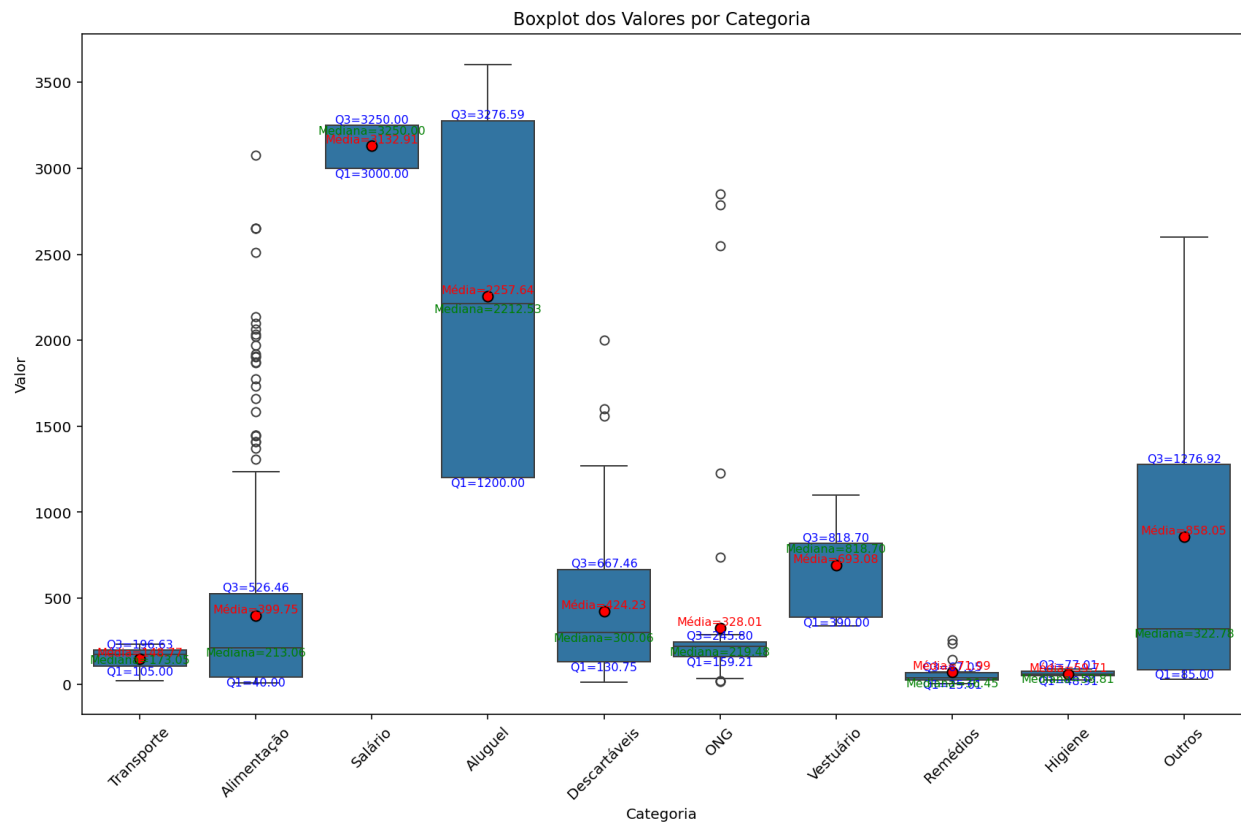


Gráfico 06 – Boxplot das despesas por Categoria

No gráfico 06 podemos observar a amplitude e os valores médios de cada categoria.

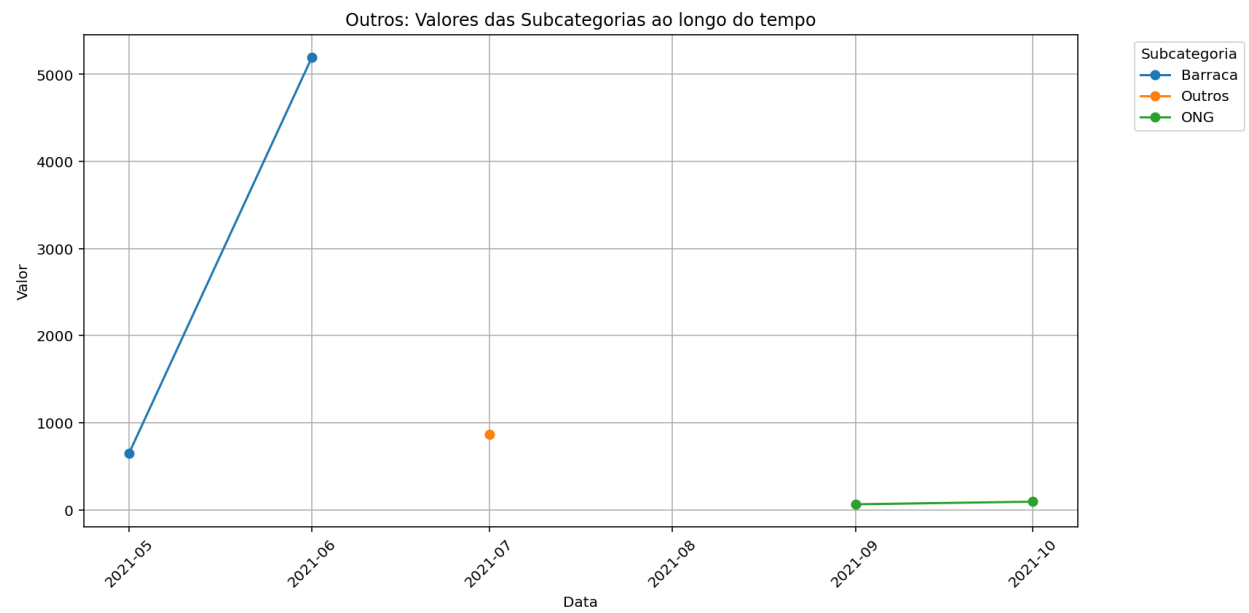


Gráfico 07 – Outras despesas

Diagramas/gráficos gerados: Descricao_Marmitas.xlsx

Total de observações: 800

Ajuste/correção em uma única linha: apresentava total de marmitas “.46”, a informação foi alterada para “46”

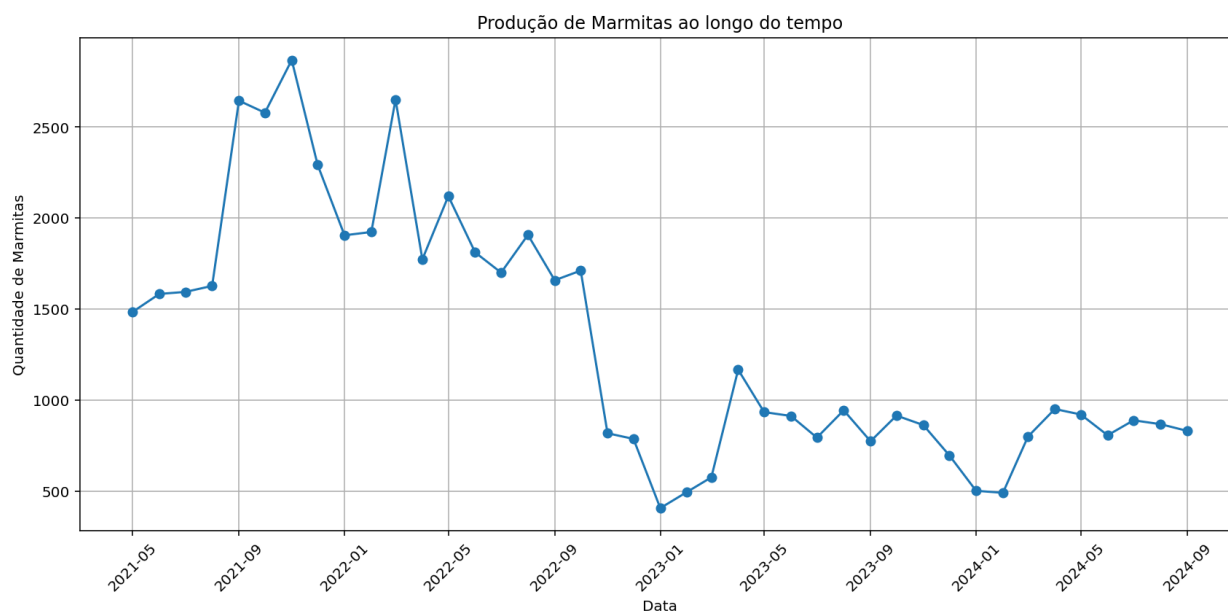
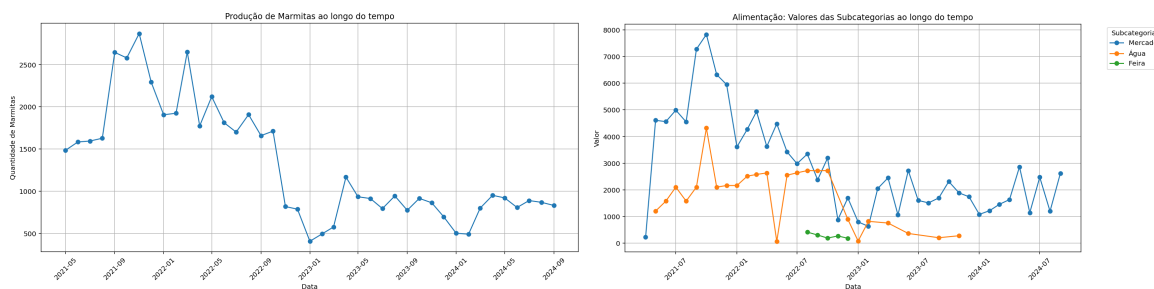


Gráfico 08 – Produção de marmitas ao longo do tempo

É visível a forte correlação entre a produção de marmitas as despesas com alimentação (subcategoria “Mercado”) presentes nos gráficos 01 e 03 cujo coeficiente de correção de Pearson entre quantidade de marmitas produzidas e as despesas com subcategoria mercado é de 90,31%



- Um coeficiente de correlação tão alto sugere que a gestão de estoque é eficiente. A quantidade de insumos adquiridos está sendo adequadamente dimensionada para atender à demanda de produção.

- A forte correlação indica que outros fatores que poderiam influenciar as despesas com mercado, como variações de preços ou desperdícios, estão tendo um impacto relativamente pequeno.
- Essa alta correlação pode indicar que o modelo de negócio está relativamente estável, com processos de produção e aquisição de insumos bem definidos.

Oportunidade de análise mais detalhada: Quais tipos de ingredientes têm maior impacto nas despesas?

Qual o custo médio de cada marmita? Qual custo médio de cada marmita por cardápio?

Como a inflação ao longo dos anos impactou no custo das marmitas?

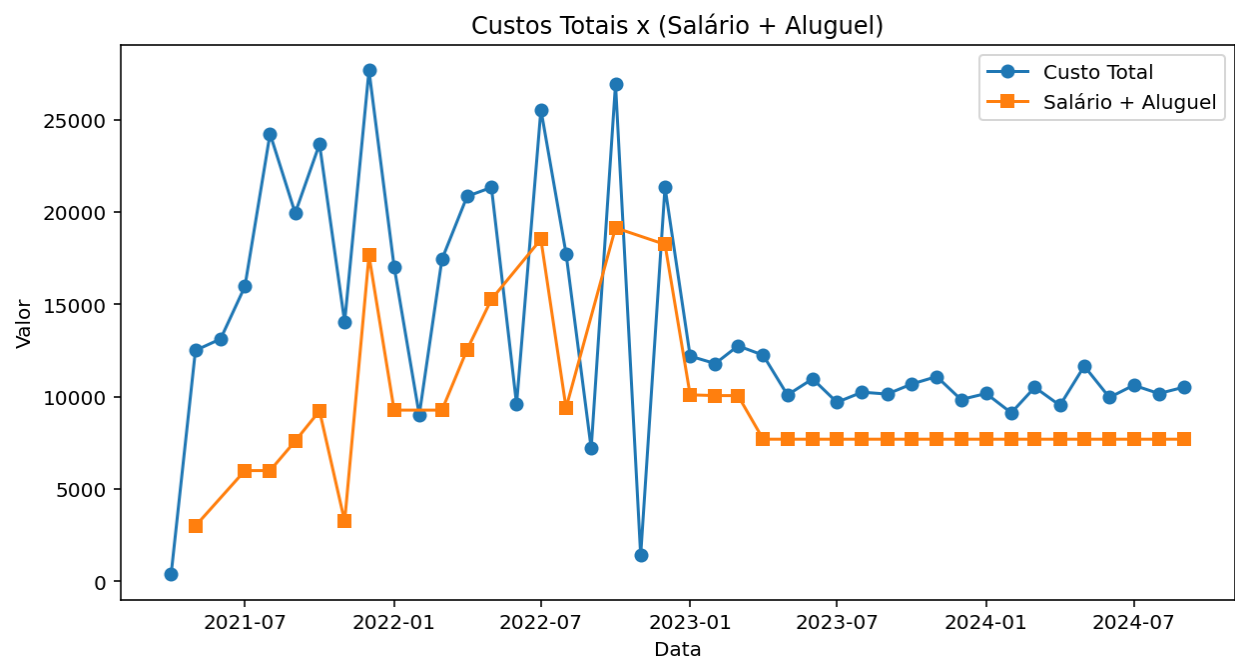


Gráfico 09 – Custos totais x Custos operacionais

De outubro de 2022 até agosto de 2024 a média do custo operacional é de 75,7245% em relação aos custos totais que podem ser vistos no gráfico 09 e tabela 01.

ano	mês	percentual
2022	10	71.0788
2022	12	85.4598
2023	1	82.8022
2023	2	85.3071
2023	3	78.8764

2023	4	62.8344
2023	5	76.2758
2023	6	70.1899
2023	7	79.3938
2023	8	75.1363
2023	9	75.8979
2023	10	72.0601
2023	11	69.4055
2023	12	78.1934
2024	1	75.6619
2024	2	84.3502
2024	3	73.2533
2024	4	80.7985
2024	5	66.0262
2024	6	77.3113
2024	7	72.4705
2024	8	75.6913
2024	9	73.1889

Tabela 01 – Percentual dos custos operacionais em relação aos custos totais

A informação de que a média dos custos operacionais é de 75,7245% em relação aos custos totais no período de outubro de 2022 a agosto de 2024, período mais recente das atividades, nos fornece uma perspectiva geral sobre a estrutura de custos da operação de produção de marmitas atualmente.

Observações/Percepções:

- O fato de os custos operacionais representarem quase 76% dos custos totais indica que a maior parte dos recursos financeiros está destinada às atividades diárias de produção.
- Isso é um indício que a margem de contribuição pode ser relativamente menor, ou seja, a diferença entre a receita e os custos variáveis (que geralmente estão inclusos nos custos operacionais) é menor do que poderia ser. A margem de contribuição é o valor que sobra das doações, após deduzir todos os custos diretamente ligados à sua produção, como os ingredientes, embalagens e mão de obra direta. Esse valor é crucial, pois é ele que cobre os custos fixos do projeto (aluguel, salários, contas) e gera saldo positivo (ter uma margem de segurança para continuidade do projeto quando as doações forem insuficientes).
- Uma alta proporção de custos operacionais implica em uma maior sensibilidade a variações nos preços dos insumos, energia, mão de obra, etc. Pequenas alterações nesses custos podem ter um impacto significativo no resultado final.

Diagramas/gráficos gerados: Extratos (Doações)

Total de observações: 3193

Variáveis:

Data: data em que foi registrada a doação

ID: identificador único do doador

Forma Pagamento: transação financeira na qual a doação foi realizada

Valor: Valor da doação (reais)

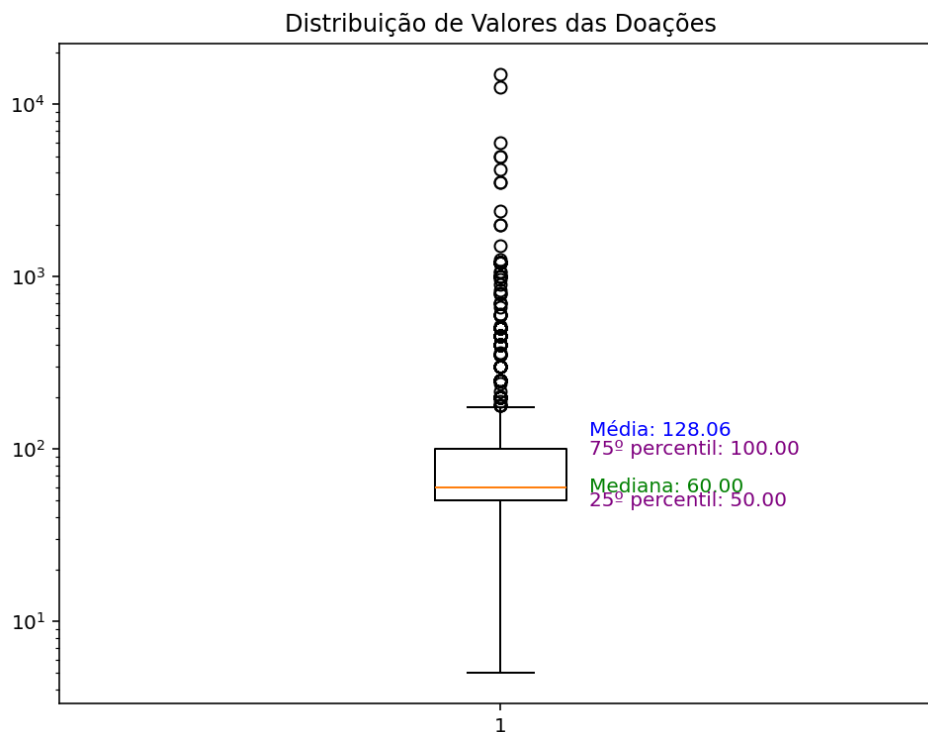


Gráfico 10 – Boxplot dos valores doados

Análise da Distribuição dos Valores Individuais das Doações, gráfico 10.

Distribuição com Outliers Significativos: Com um mínimo de R\$ 10 e um máximo de R\$ 15.000, observa-se uma grande variação nos valores doados individualmente. A mediana de R\$ 60 e o percentil 75 de R\$ 100 indicam que a maioria das doações é modesta. No entanto, o valor máximo é extremamente alto comparado aos percentis, caracterizando uma presença de outliers de alto valor que pode ser observado também

pela média muito superior a mediana, indicando uma forte assimetria positiva na distribuição.

Preponderância de Pequenas Contribuições: A mediana e os percentis 25 e 75 (R\$ 50 a R\$ 100) sugerem que a maioria das contribuições individuais é de baixo valor, confirmando uma distribuição onde doações maiores são exceções.

Curva com assimetria positiva forte média (R\$ 128,06) muito superior a mediana (R\$ 60,00)

Importância das Pequenas Doações para Sustentabilidade: A maior parte das contribuições está entre R\$ 50 e R\$ 100, indicando que, embora existam grandes doações, o montante arrecadado depende principalmente de muitas doações pequenas.

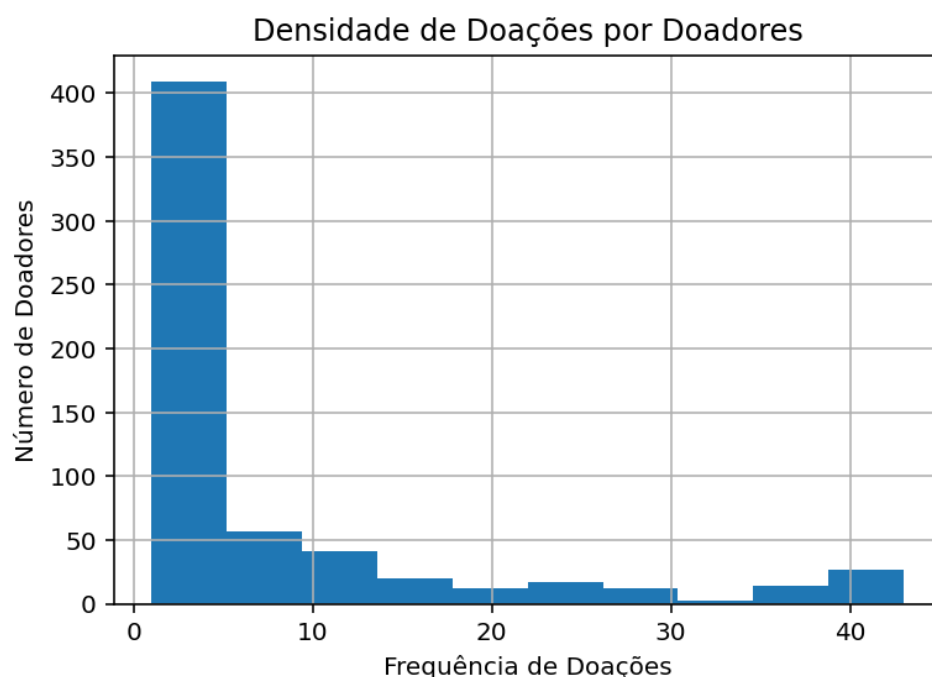
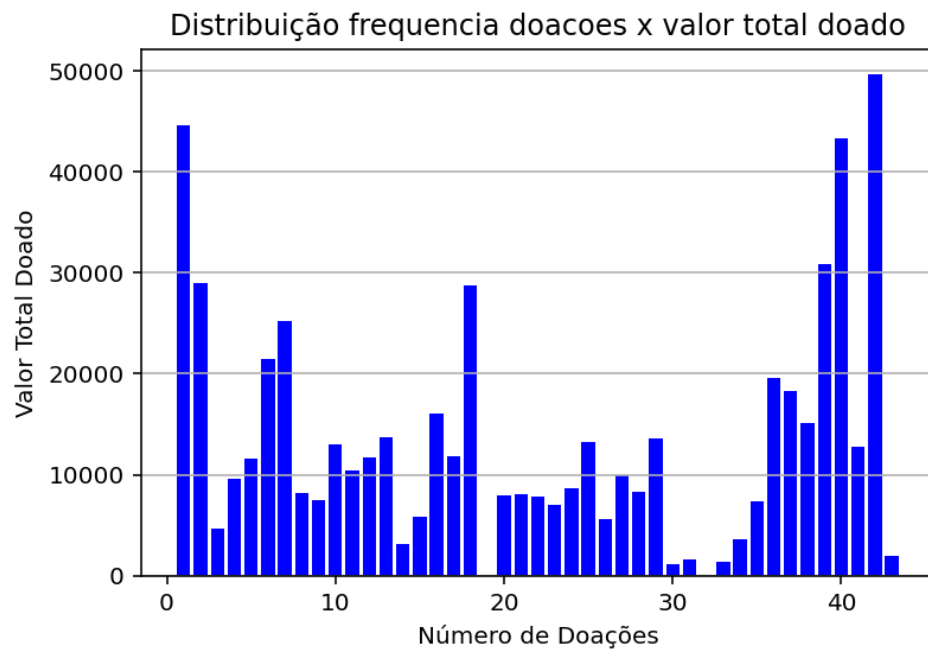


Gráfico 11 – Densidade de doações por doador

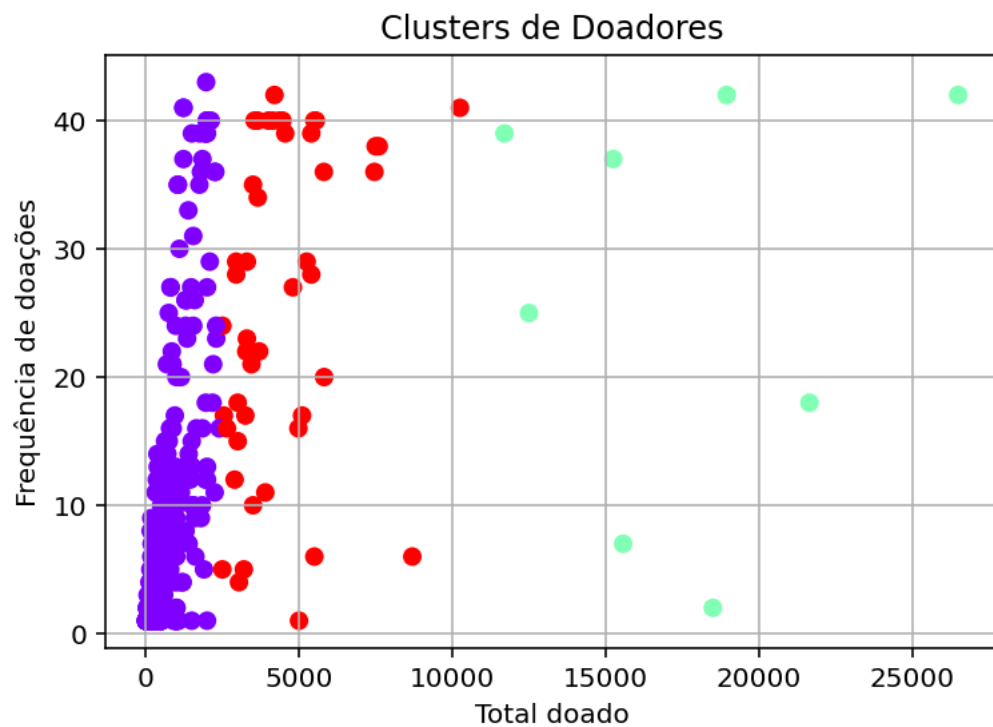
No gráfico 11 podemos observar a distribuição dos doadores por Bins de Frequência. Alta Concentração nos Bins de 1 a 5 doações: A maior densidade de doadores está na faixa de 1 a 5 doações (acima de 400 doadores), reforçando a observação de que a maioria dos doadores participa poucas vezes.

Curva de Decaimento Rápido: O número de doadores decresce significativamente à medida que a frequência aumenta, ficando próximo de 50 já no intervalo de 5 a 9 doações e abaixo em todos os demais intervalos. Existe um pico no sexto bin indicando a presença de um pequeno grupo com uma frequência mais alta e outro no décimo. Essa distribuição sugere um efeito de “calda longa”, onde poucos doadores mantêm frequências altas. A existência desses três picos sugere a existência de grupos distintos de doadores.

Comportamento de doadores ocasional versus recorrente: Os dados indicam que o grupo majoritário realiza uma ou poucas doações, enquanto uma minoria se mostra mais comprometida, doando com maior frequência.



Neste gráfico também é possível notar a possível existência dos três grupos de doadores em função do valor total doado (descontinuidade em 19 e 32).



Clusterização não hierárquica com três grupos onde é possível notar uma agrupamento por valor total doado.

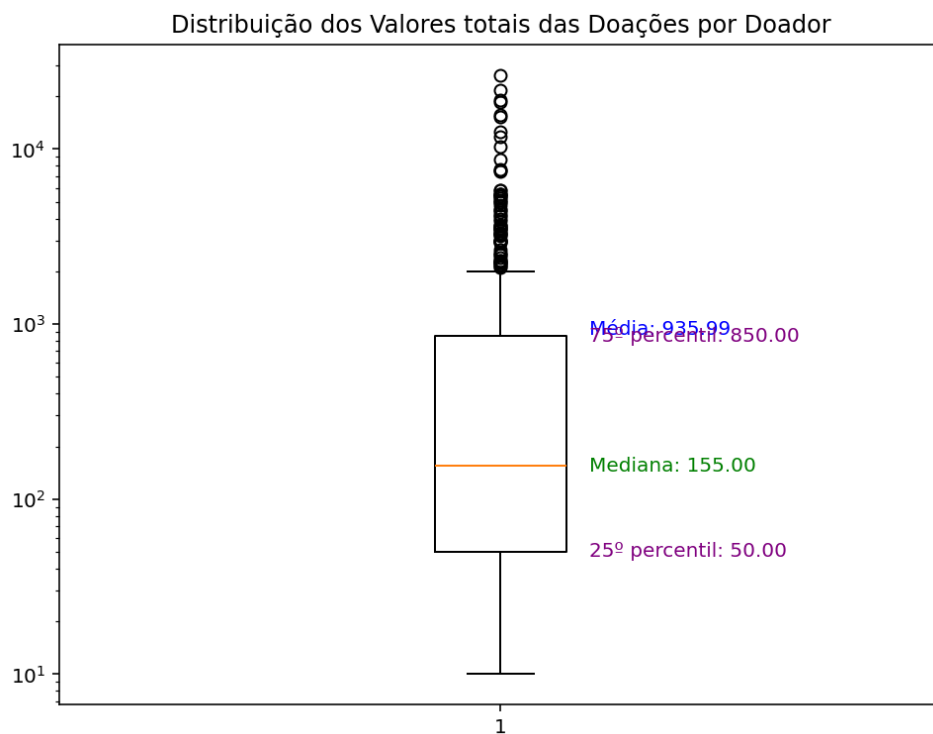


Gráfico 12 – Boxplot dos valores totais por doador

Pelo gráfico 12, podemos notar uma forte assimetria na distribuição com a média (R\$ 935,99) e percentil 75 (R\$ 850) muito acima da mediana (R\$ 155,00). Existe claramente uma desigualdade nas doações totais por doador sugerindo que poucos doadores contribuem com valores totais consideravelmente mais altos.

Perfil dos doadores por quantia total: A maior parte dos doadores parece doar quantias totais relativamente baixas, com 75% dos doadores contribuindo com até R\$ 850 no total. Esse padrão é reforçado pela diferença entre o percentil 25 (R\$ 50) e a mediana, indicando que metade dos doadores contribui com menos de R\$ 155 no total.

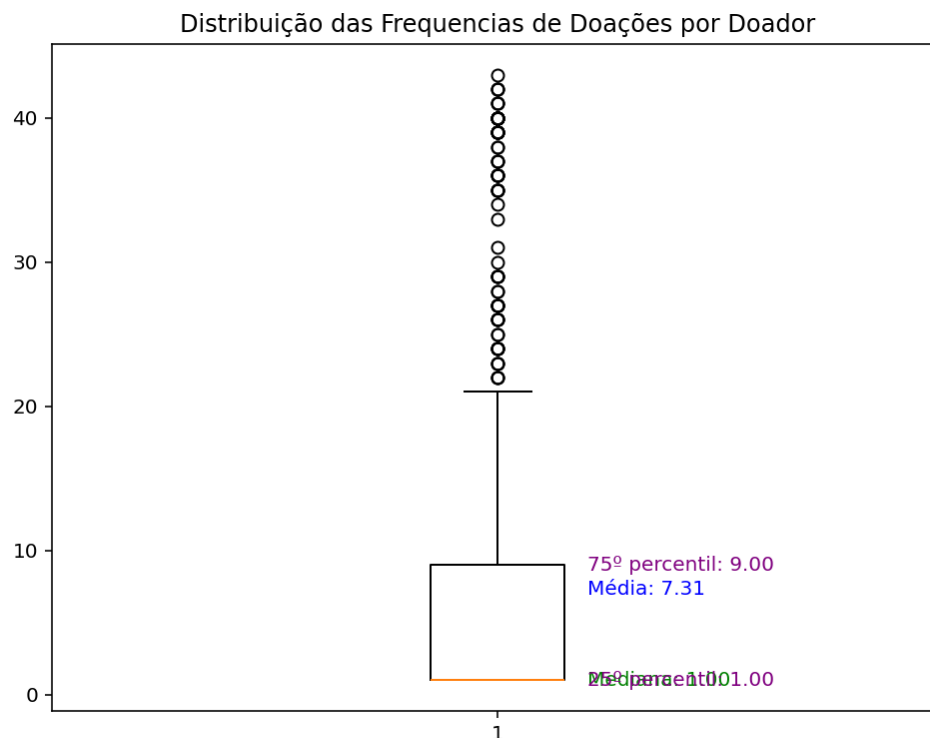
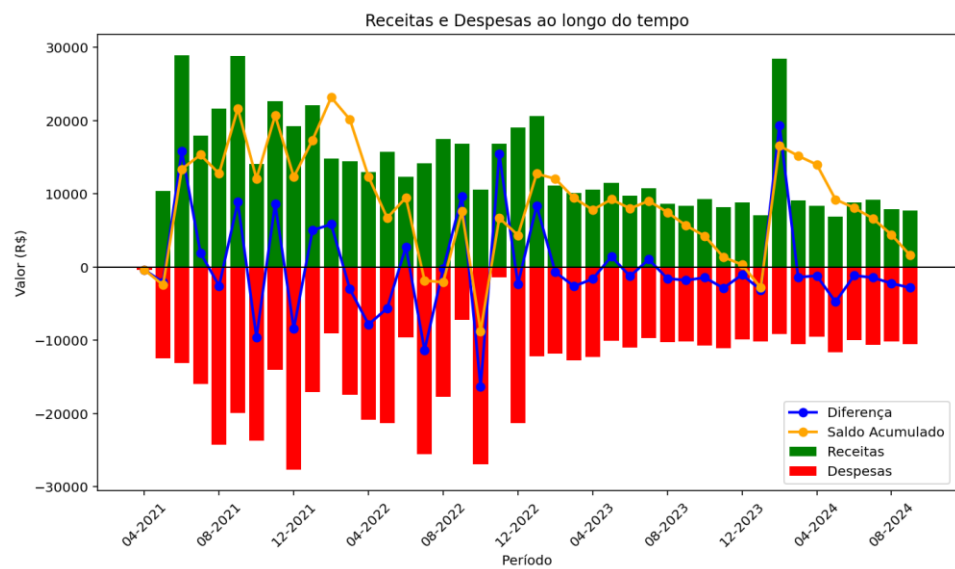
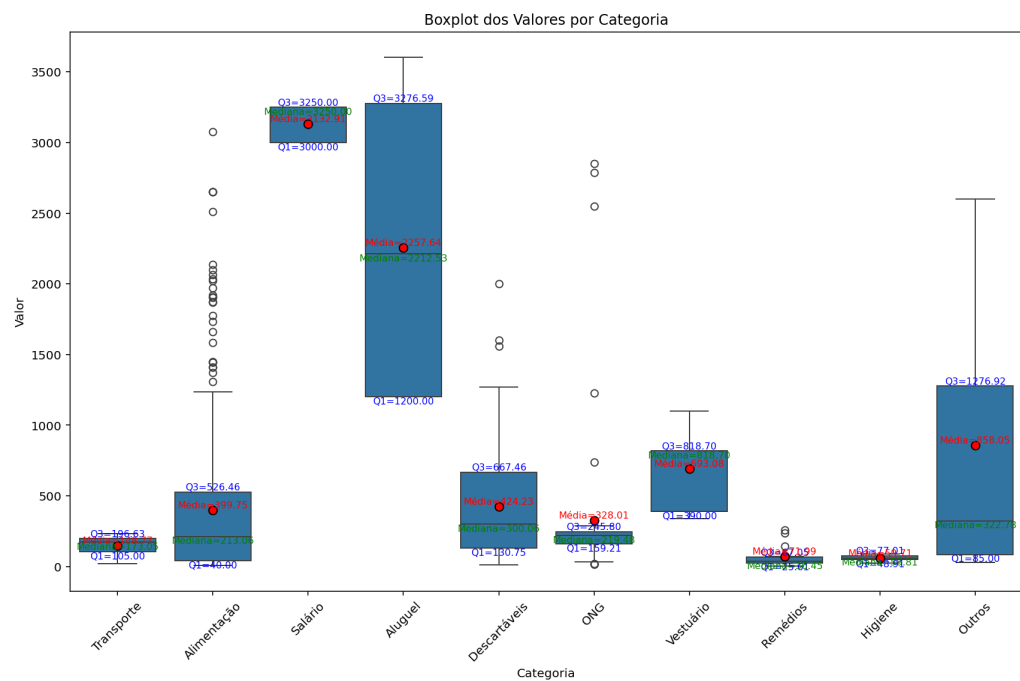


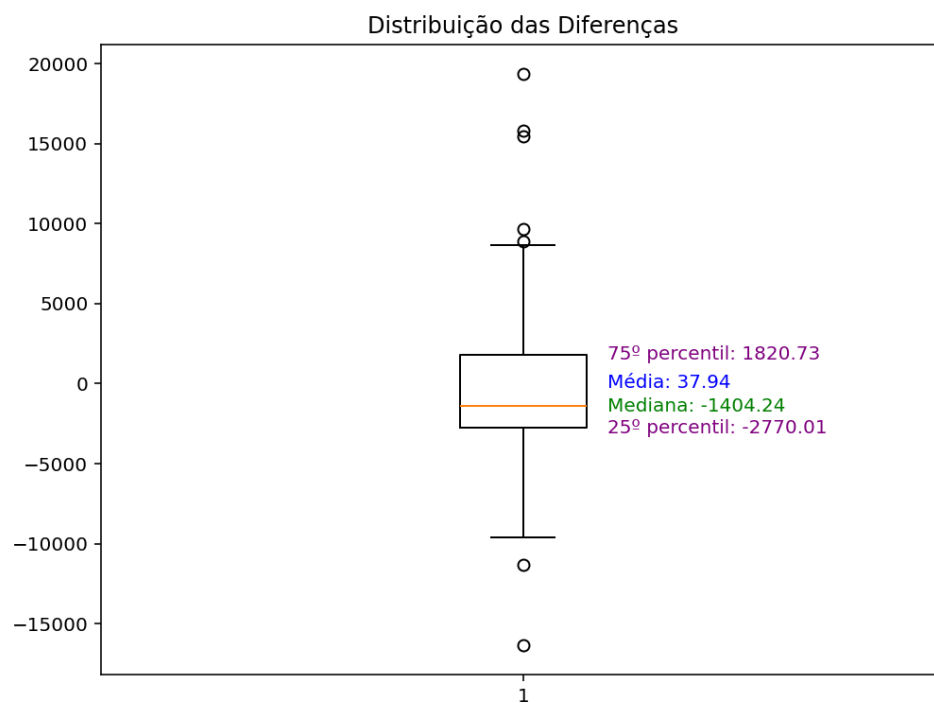
Gráfico 13 – Boxplot das frequências das doações por doador

A distribuição de frequência está concentrada em baixas recorrências, conforme podemos observar no gráfico 13. A mediana de 1 doações e o percentil 25 de apenas 1 doação indicam que muitos doadores contribuem com pouca frequência. O percentil 75 de 9, o bigode superior em torno de 20 e os outliers sugerem, porém, que há um grupo que doa com uma frequência mais alta, embora seja uma minoria.

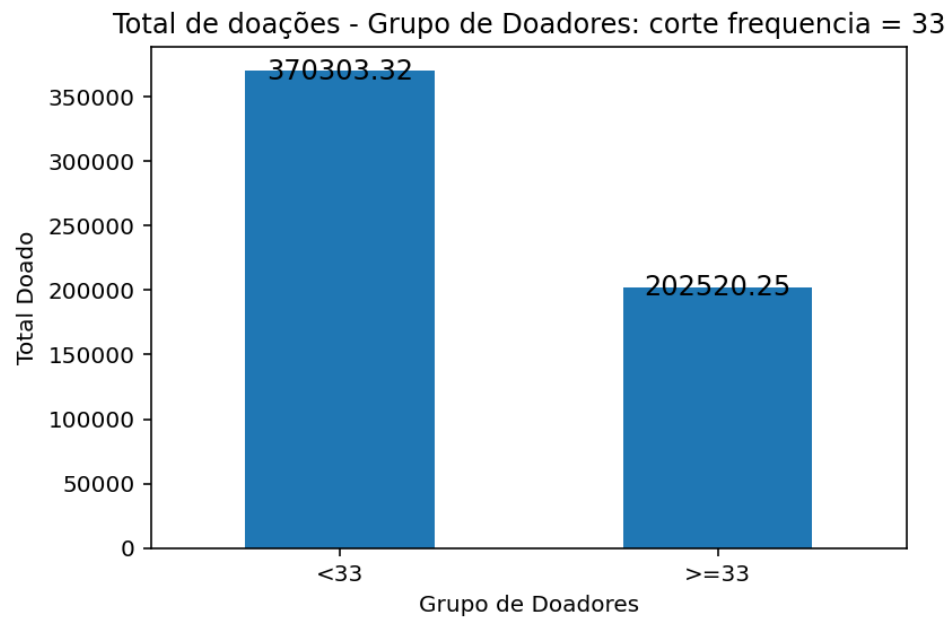
A média de 7,31 doações está muito acima da mediana de 1, confirmando uma forte assimetria. Isso aponta para um comportamento onde poucos doadores têm alta frequência de doação, enquanto a maioria realiza poucas doações.

Este grupo recorrente de doadores pode ser chave em manter uma base financeira mais estável para a projeto, já que tende a ter um impacto significativo na média das doações.





Distribuição assimétrica positiva, porém é preciso estar atento a saúde financeira do projeto tendo em vista a mediana em menos R\$ 1404.24 e uma média positiva muito baixa (R\$ 37,94)



Fazendo um corte em frequência de doações igual a 33 (80% de frequência de doação), é possível perceber que os doadores frequentes (≥ 33) representam 35% da receita de doações, indicando a importância das contribuições de doadores não frequentes no projeto.

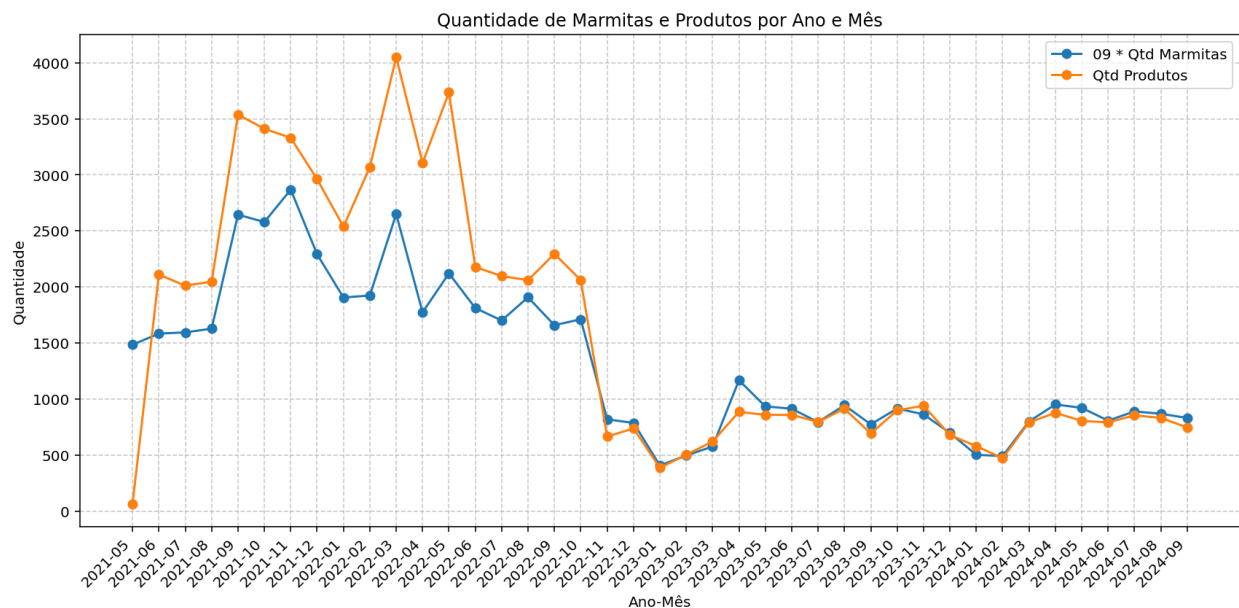
Diagramas/gráficos gerados: Descrição Marmitas

Total de observações:

Variáveis:

Data: data em que foram feitas as marmitas

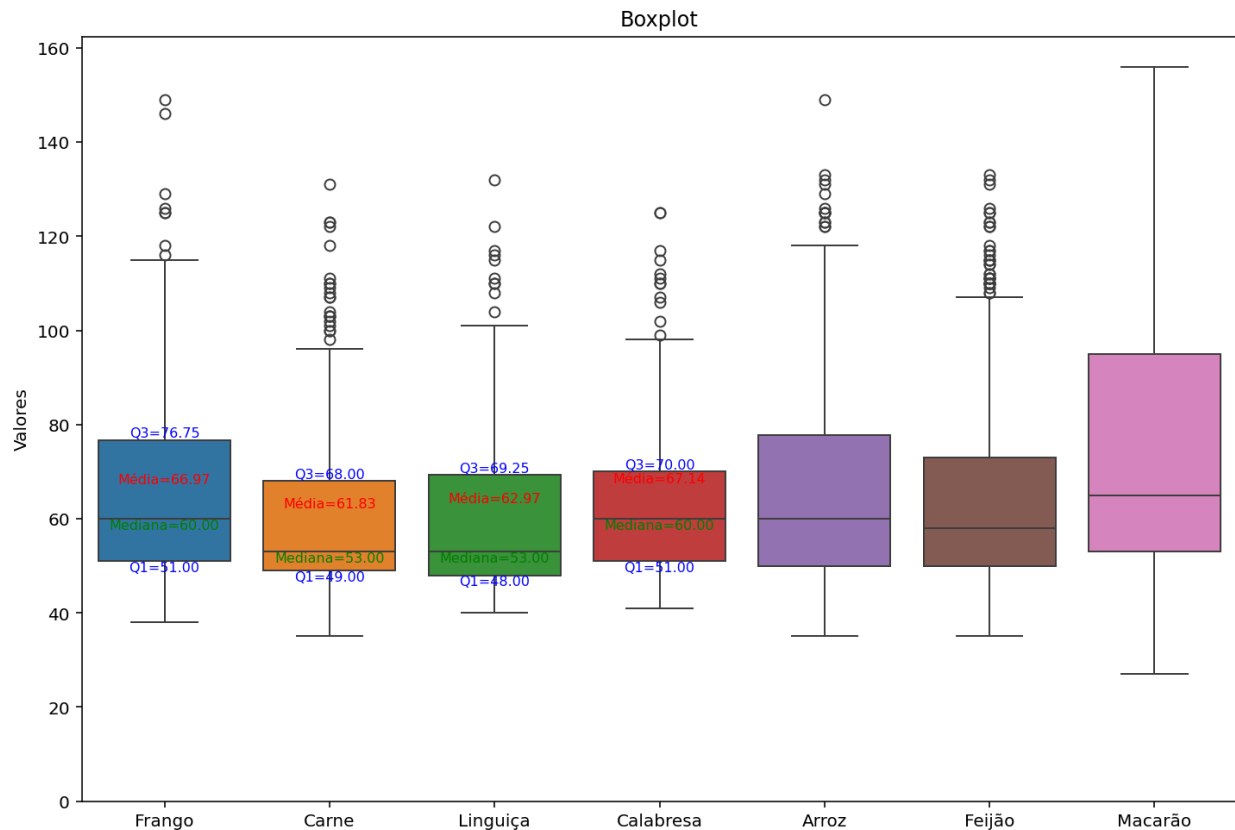
Qtd: quantidade de marmitas produzidas



Esse gráfico apresenta a produção de marmitas x [9 * qtd de produtos que foram utilizados].

A partir de novembro de 2022 é possível perceber que dada a quantidade de produtos utilizada a qtd de marmitas é da ordem de grandeza de 9 vezes maior.

Isso sugere que a produção de marmitas está mais eficiente do que do início do projeto até esta data



Boxplot de quantidade de marmitas por produto utilizado

A maior e a menor quantidade de marmitas produzida foi com Macarão. Possui a maior variabilidade

Podemos fazer um teste ANOVA (1 fator) - exige homogeneidade de variâncias e normalidade das amostras para ser aplicado. Comparação de K médias.

Verificar se a média de produção é a mesma para todos os produtos.

$$H_0 = \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$$

H_1 = pelo menos uma é diferente

ANOVA (1 fator) - exige homogeneidade de variâncias e normalidade das amostras para ser aplicado. Comparação de K médias.

Verificar se a média de produção é a mesma para todos os produtos.

$$H_0 = \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$$

H_1 = pelo menos uma é diferente

Verificando a normalidade - teste de shapiro-wilk

H0 = Os dados da amostra vêm de uma distribuição Normal

H1 = Os dados da amostra não vêm de uma distribuição Normal

Resultado dos testes de normalidade (Shapiro-Wilk) foram todos com $p_value < 0,05$.

Logo, os dados não seguem uma distribuição normal e não podemos usar o ANOVA para verificar se a média de produção é a mesma para todos os produtos

Verificando a homogeneidade - teste de Levene

H0: As variâncias dos grupos são homogêneas

H1: Pelo menos uma das variâncias dos grupos é diferente

Como não posso utilizar um teste paramétrico, pois as amostras não seguem uma distribuição normal, usei um teste não paramétrico: Kruskal-Wallis

H0: As distribuições de todos os grupos são iguais.

H1: Pelo menos uma das distribuições dos grupos é diferente.

Resultado foi:

$p_value < 0.05$, Rejeito H0, com isso, pelo menos uma das distribuições dos grupos é diferente.

O teste de Kruskal-Wallis é não paramétrico, então ele compara as distribuições dos grupos (e não as médias diretamente). Embora seja comum interpretar um resultado significativo como uma diferença entre as médias, tecnicamente o teste indica que pelo menos uma das distribuições é diferente em termos de posição central (mediana).