



# Relatório de Projeto

## Desafio 1

Campinas  
2022

## Conteúdo

1	Introdução	2
2	Metodologia	3
3	Resultados	4
4	Conclusão	10
5	Referências	11
6	Anexos	12

## Lista de Figuras

1	Boxplot das avaliações dos aplicativos por categoria . . . . .	4
2	Gráfico de barras da média do número de opiniões por categoria . . . .	5
3	Gráfico de barras da mediana do número de opiniões por categoria . .	5
4	Gráfico de barras da média do nível de satisfação dos jogos por gratuidade	6
5	Gráfico de barras da média do nível de satisfação dos jogos por classi- ficação de conteúdo . . . . .	7
6	Gráfico de barras da média do número de opiniões dos jogos por gratuidade	8
7	Gráfico de barras da média do número de opiniões dos jogos por clas- sificação de conteúdo . . . . .	9

## Lista de Tabelas

## 1 Introdução

Com o objetivo de escolher a melhor área para se investir em um aplicativo, analisou-se um banco de dados contendo informações de vários aplicativos divididos em 5 categorias - família, negócios, ferramentas, jogos e medicina -, em dois tipos de disponibilidade - gratuito e pago - e quatro tipos de classificação de conteúdo - acima de 10 anos, 17 anos, adolescentes e para todo mundo.

Todas as medidas, o banco de dados e os gráficos foram tratados com o Rstudio. Entretanto, por causa de algumas dificuldades encontradas para fazer tudo pelo R, primeiro, esboçaram-se todos os procedimentos no Google Planilhas e, depois, foram colocados em prática no R a fim de facilitar a inserção desses gráficos no markdown, como foi recomendado pelo desafio. O mesmo aconteceu com o relatório, primeiro foi feito o rascunho no Google Documentos e, depois, foi feita a inserção das informações no relatório criado pelo Rmd.

Os estudos dos gráficos foram de maior importância para a chegada à conclusão, pois como o banco de dados é muito grande, sem os gráficos não seria possível visualizar nada do que os dados queriam nos dizer.

Assim, analisando os gráficos e as medidas estatísticas, chegou-se à conclusão de qual tipo de aplicativo seria o ideal a se investir.

## 2 Metodologia

Primeiramente, foi feito um rascunho do que deveria ser feito no Google Planilhas pela sua facilidade em gerar gráficos, para saber quais gráficos seriam necessários e úteis para, depois serem criados usando o R.

Desse modo, no R, separou-se o banco de dados nas cinco categorias e através do sumário de cada tabela, pôde-se registrar as medidas estatísticas das avaliações e da quantidade de opiniões de cada uma das categorias.

Tomou-se uma variável muito importante para as análises o número de opiniões, uma vez que, de todas as formas, esta corresponde ao tamanho do interesse do público pelo aplicativo - se a avaliação for negativa, a pessoa se interessa pela melhora deste a fim de utilizar esse aplicativo de forma mais eficiente e se a avaliação for positiva, só demonstra o interesse pelo aplicativo mesmo. Portanto, através dessa variável, decidiu-se em qual tipo de aplicativo seria mais adequado para a investigação das outras variáveis (gratuidade e classificação de conteúdo). Assim, foi feito um gráfico da média de número de opiniões por categoria, a fim de observar em qual tipo de aplicativo foi manifestado maior interesse.

Após isso, foi feito o boxplot das avaliações por categoria para observar as avaliações dos clientes com cada tipo de aplicativo.

Daí, sabendo qual categoria de aplicativo seria o mais ideal para a investigação, através do gráfico da média de opiniões, dividiu-se o banco de dados principal para obtermos os dados apenas desta categoria para levantar mais informações através de novos gráficos.

Desse modo, foram feitos dois gráficos com as avaliações, um relacionando este com sua gratuidade a fim de saber se há diferença na satisfação do cliente com a mudança do preço a ser pago pelo aplicativo e outro relacionando este com a classificação de conteúdo para saber se há diferença na satisfação dos clientes de acordo com a classificação do aplicativo. O mesmo foi feito com o número de opiniões, foram feitos dois gráficos: um relacionado à gratuidade e outro à classificação de conteúdo.

Todos os gráficos foram feitos usando o ggplot, como ensinado no treinamento de R oferecido pela Estat Jr. Segue um exemplo base - os outros foram feitos com a mesma lógica, mas com algumas mudanças - no anexo.

Através desta metodologia, foi possível chegar a uma conclusão de qual tipo de aplicativo seria ideal para realizar um investimento.

### 3 Resultados

Foi observado que, em todo o caso, as avaliações não influenciaram no resultado, uma vez que em todas as categorias, a distribuição dessa variável está bem parecida em todas as categorias, como visto no boxplot abaixo.

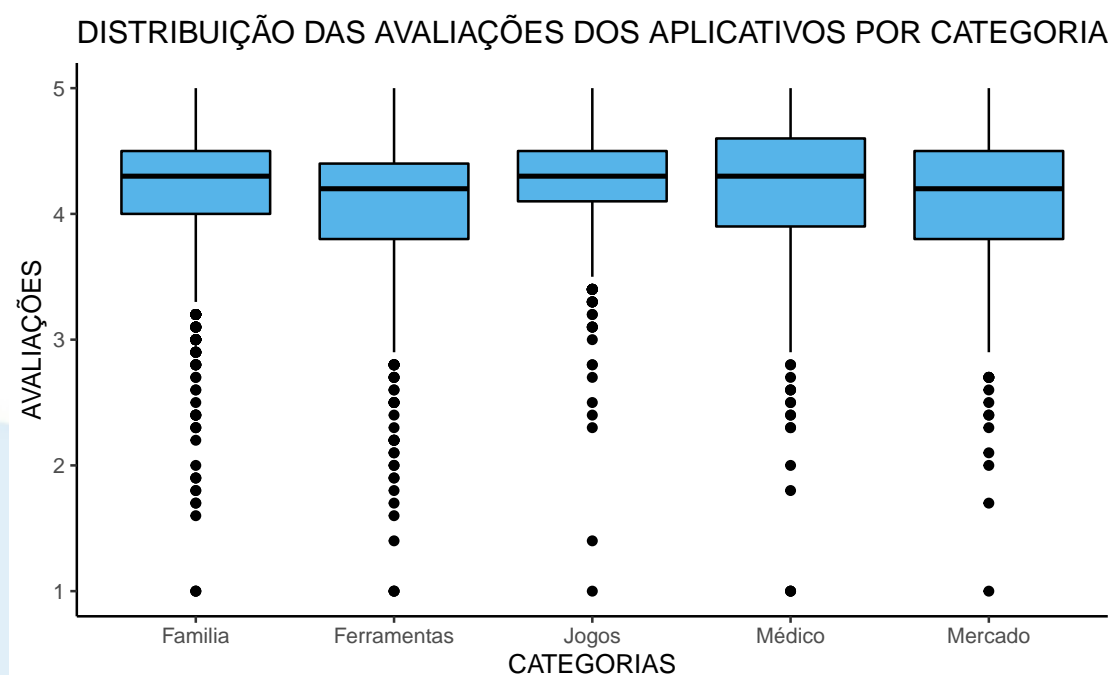


Figura 1: Boxplot das avaliações dos aplicativos por categoria

Nos outros dois gráficos que envolviam as avaliações, a média destas relacionadas à classificação de conteúdo e à gratuidade não apresentou nenhuma informação significativa para a chegada ao resultado.

Ao contrário disso, o número de avaliações foi fortemente usado para a obtenção do resultado. Desse modo, primeiro, foi observado que a média, bem como a mediana, do número de avaliações, dos jogos se apresentou muito maior do que as outras categorias.

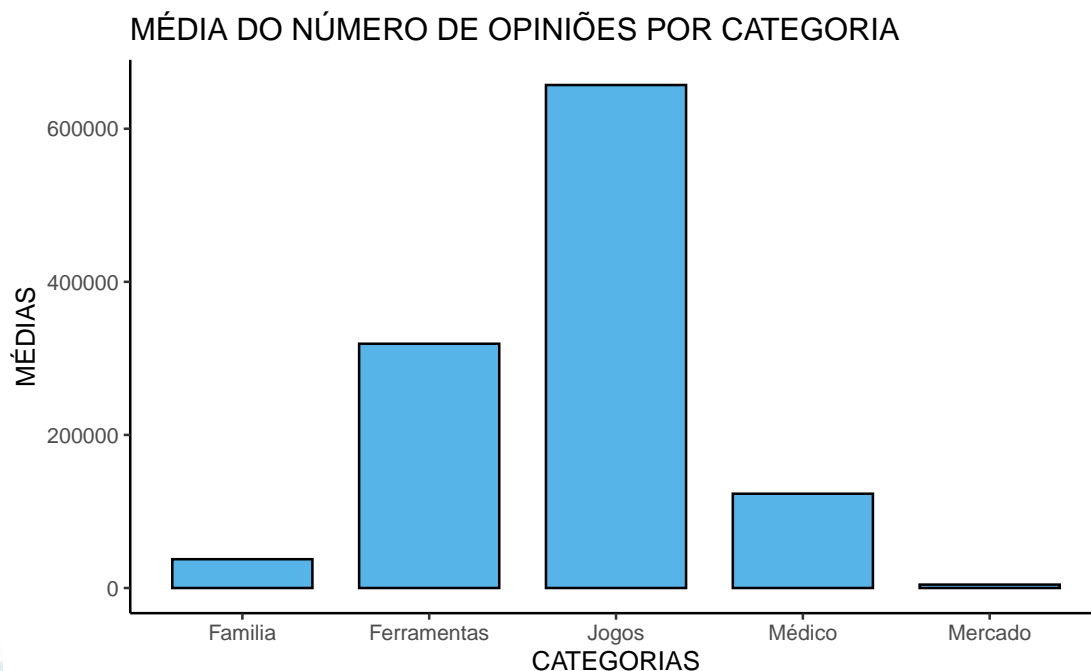


Figura 2: Gráfico de barras da média do número de opiniões por categoria

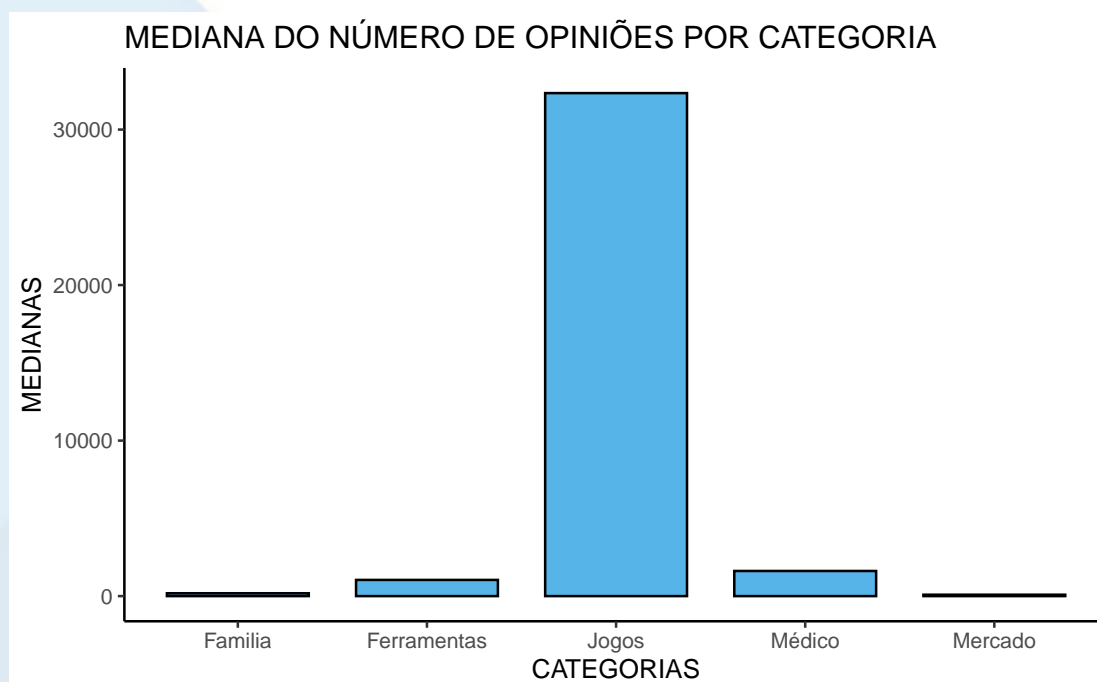


Figura 3: Gráfico de barras da mediana do número de opiniões por categoria

Portanto, decidiu-se que a categoria dos jogos é a melhor categoria para se investir e, assim, seguiu-se com os estudos em cima desta categoria.

Com isso, através dos gráficos escolhidos para análise apontados na metodologia, foi observado que não há diferença significativa no nível das avaliações dos jogos com a mudança da gratuidade nem com a mudança da classificação indicativa, ou seja, segundo os gráficos, os aplicativos pagos não satisfazem muito mais do que os gratuitos.

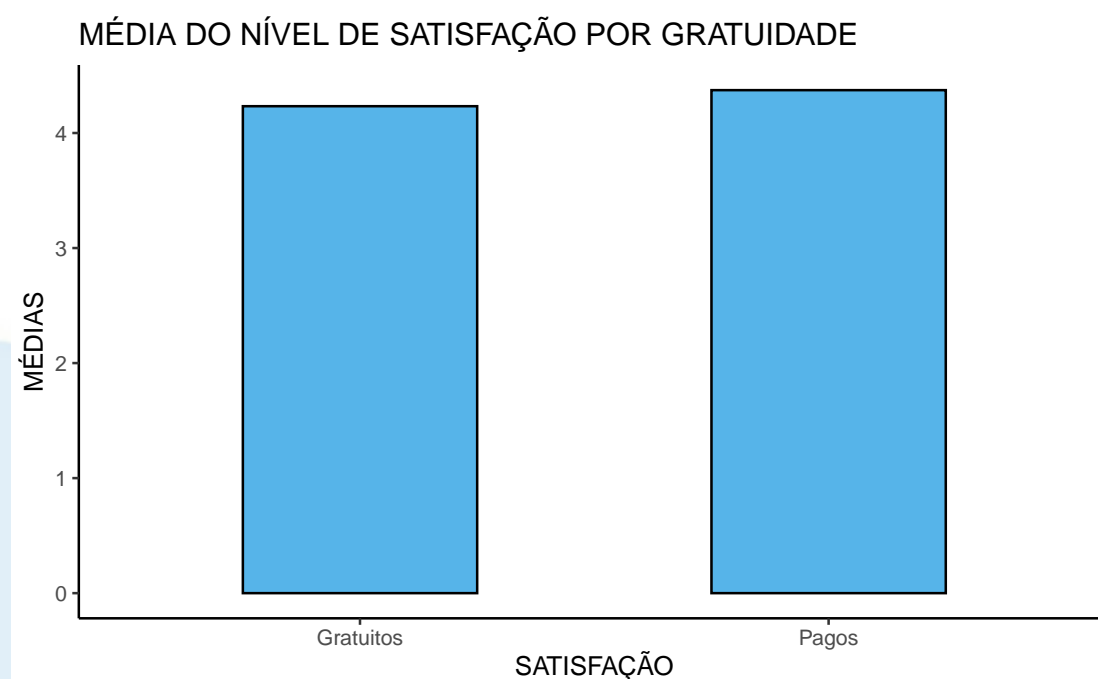


Figura 4: Gráfico de barras da média do nível de satisfação dos jogos por gratuidade

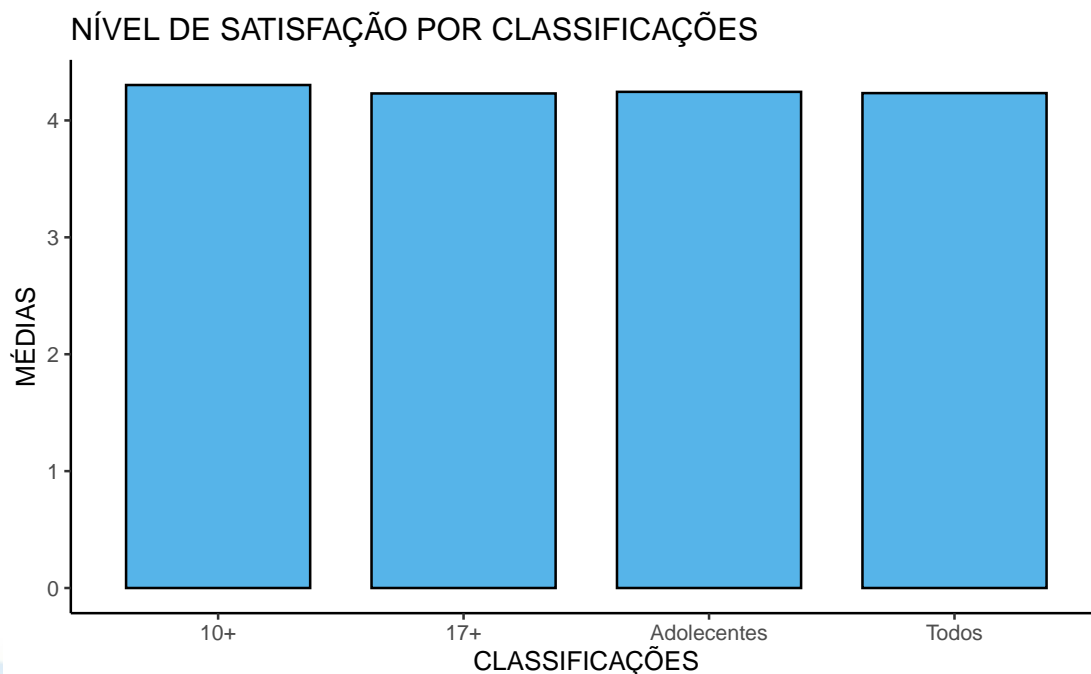


Figura 5: Gráfico de barras da média do nível de satisfação dos jogos por classificação de conteúdo

Logo, também nesse caso, o número de avaliações foi a variável mais importante para escolhermos essas a gratuidade e a classificação de conteúdo.

Desse modo, foi visto que o interesse dos clientes por jogos gratuitos é muito maior do que por jogos pagos.



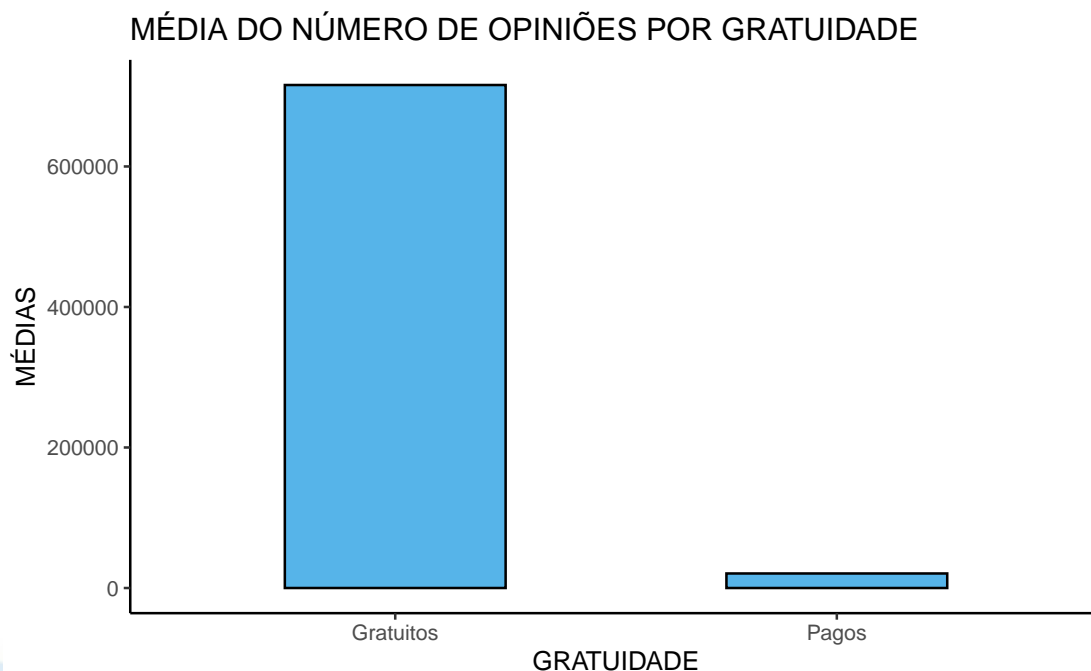


Figura 6: Gráfico de barras da média do número de opiniões dos jogos por gratuidade

Da mesma maneira, o número de avaliações por jogos para pessoas acima de 10 anos é muito maior do que para as outras classificações indicativas.

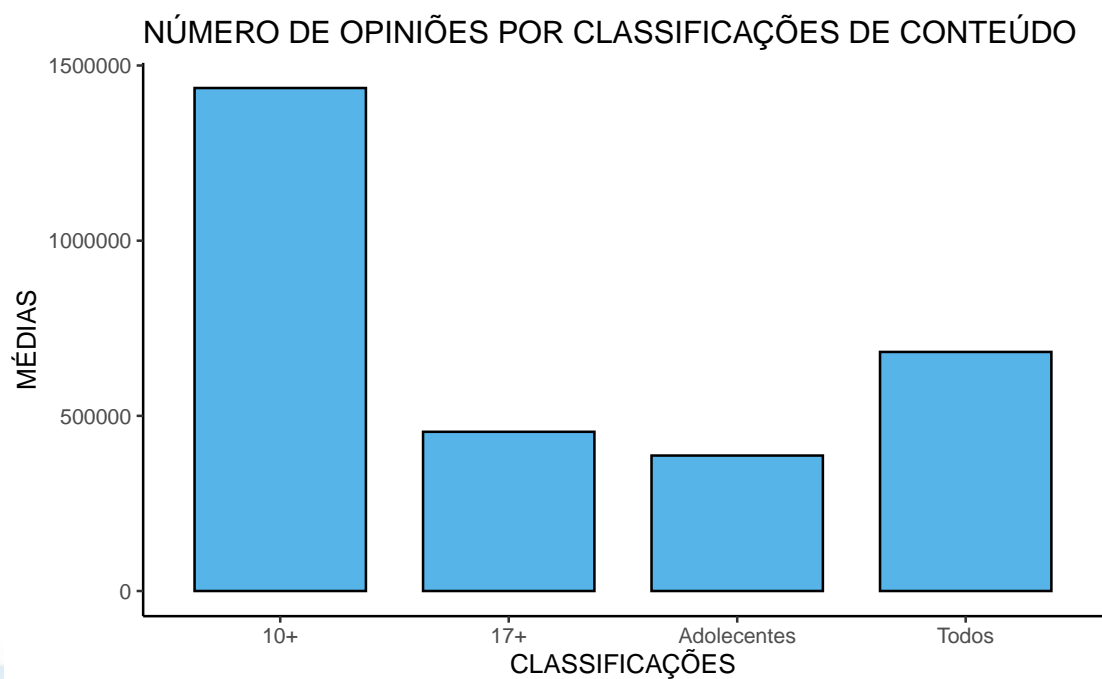


Figura 7: Gráfico de barras da média do número de opiniões dos jogos por classificação de conteúdo

## 4 Conclusão

Então, extraindo informações do banco de dados a fim de saber em qual categoria de aplicativo seria o melhor para realizar um investimento, pôde-se observar que o número de opiniões da categoria dos jogos é muito maior do que as outras categorias.

Além disso, dentro dos jogos, o público apresentou um maior interesse pelos jogos gratuitos do que os pagos, além de apresentarem uma satisfação bem parecida com os dois tipos de aplicativos.

Também, há um número maior de opiniões naqueles aplicativos com classificação de conteúdo para todos acima de dez anos.

Ademais, é possível observar que a satisfação (avaliação) dos clientes não mudam significativamente de acordo com a mudança da gratuidade ou da classificação de conteúdo.

Portanto, através dessas informações obtidas através dos gráficos, é possível afirmar que o melhor tipo de aplicativo para se investir é da categoria dos jogos, gratuitos e de classificação de conteúdo para todos acima de dez anos.

## 5 Referências

## 6 Anexos

Segue o exemplo abaixo de como foi feito o gráfico que relaciona o número de opiniões com a gratuidade dos aplicativos da categoria de jogos. Desse modo, os outros gráficos foram feitos usando a mesma lógica.

```
df_g1 = data.frame(gratuidade = c('Gratuitos', 'Pagos'),
                   medias = c(mean(df_game$Reviews
                                   [df_game$Type == "Free"]),
                              mean(df_game$Reviews
                                   [df_game$Type == "Paid"]))
                   )

ggplot(df_g1, aes(y = df_g1$medias, x = df_g1$gratuidade)) +
  geom_bar(stat = "identity",
           fill = '#56B4E9',
           color = 'black',
           width = .50) +
  xlab('GRATUIDADE') +
  ylab('MÉDIAS') +
  ggtitle('MÉDIA DO NÚMERO DE OPINIÕES POR GRATUIDADE') +
  theme_classic()
```

