

Projeto: análise estatística de dados de máquinas virtuais ociosas

Link do repositório: <https://github.com/felipesantos9/Projeto-estatisticaEprobabilidade>

Link do vídeo:  DataOPTI - Projeto Estatística - SI 2023.2

Integrantes: André Campos, Daniel Dias, Felipe Santos, Lucas Lucena e Matheus Pessoa

Relatório Detalhado da DataOpti: Transformando Dados em Insights Acionáveis

1. Quem Somos:

A DataOpti é uma empresa fundada em 2021 por cinco sócios visionários, incluindo André Campos, Daniel Dias, Felipe Santos, Lucas Lucena e Matheus Pessoa. Nosso propósito é unir os avanços tecnológicos e aplicá-los de forma inovadora ao mundo organizacional, oferecendo soluções personalizadas e eficazes em análise de dados.

- **Missão:**

Transformar dados em insights acionáveis para impulsionar o sucesso dos negócios de nossos clientes, fornecendo soluções inovadoras e personalizadas que agregam valor tangível.

- **Visão:**

Ser reconhecida como líder em análise de dados, oferecendo soluções eficazes e inovadoras que impulsionam o crescimento empresarial global, mantendo-nos na vanguarda das tendências e tecnologias emergentes.

- **Valores:**

Excelência, Inovação, Colaboração, Integridade e Paixão. Buscamos a excelência em tudo o que fazemos, valorizando a inovação, a colaboração com clientes e colegas, agindo com integridade em todas as nossas interações e sendo apaixonados pelo que fazemos.

2. Projeto Realizado:

A DataOpti realizou um projeto para a Google, que enfrentava um desafio com seus recursos de TI. A empresa possuía 22 máquinas virtuais, sendo que apenas 16 eram utilizadas em média, com um máximo de 19 em um mês. As máquinas em standby custavam \$6.000 por mês cada, enquanto as em uso custavam \$10.000 por mês cada. Isso resultava em um custo mensal significativo para a Google.

3. Detalhamento do Serviço e Metodologia:

Neste projeto, recebemos uma base de dados com informações sobre o uso de máquinas virtuais ao longo do tempo. Para obter os dados de ociosidade, realizamos um processo de tratamento e cálculo que levou em consideração o número total de máquinas virtuais disponíveis e as suas taxas de uso. O cálculo da média de máquinas virtuais ociosas envolveu a determinação do número de máquinas em standby em relação ao número total de máquinas disponíveis. Com base nessa proporção e na taxa de uso observada ao longo do tempo, conseguimos estimar o número médio de máquinas virtuais que estavam ociosas em intervalos regulares de tempo.

Esse processo nos permitiu entender a dinâmica de uso das máquinas virtuais e identificar os períodos de maior ociosidade. Com esses dados em mãos, pudemos propor uma estratégia de otimização que visava reduzir o número de máquinas virtuais em standby, mantendo um equilíbrio entre a capacidade de resposta do sistema e a eficiência operacional.

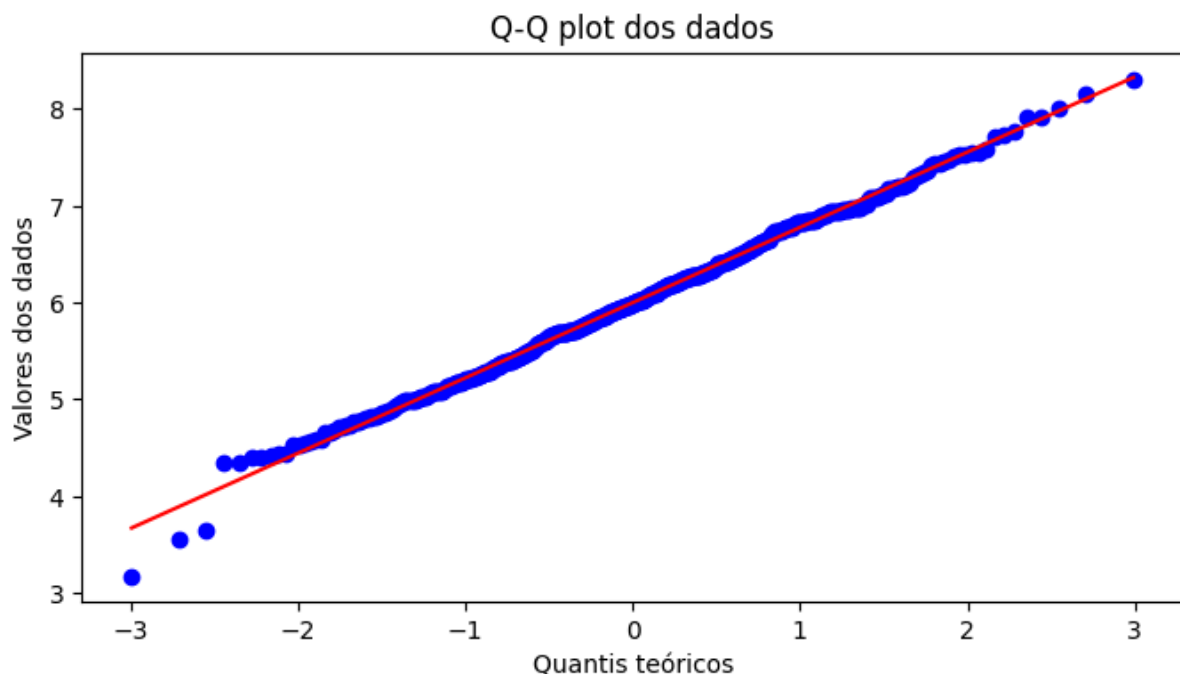
Portanto, a metodologia adotada neste projeto envolveu não apenas a análise dos dados brutos de utilização das máquinas virtuais, mas também a aplicação de cálculos e tratamentos específicos para estimar a média de máquinas ociosas e, assim, fornecer insights valiosos para a tomada de decisão estratégica.

As tecnologias utilizadas nessa análise foram:

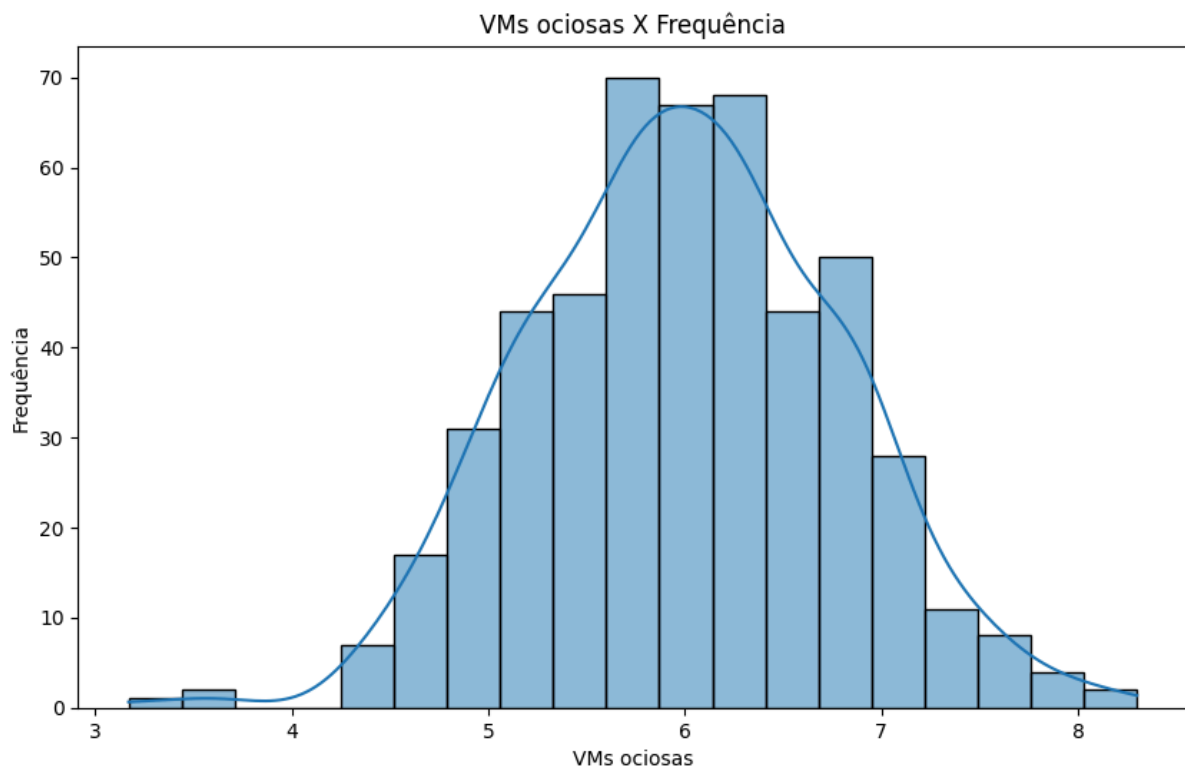
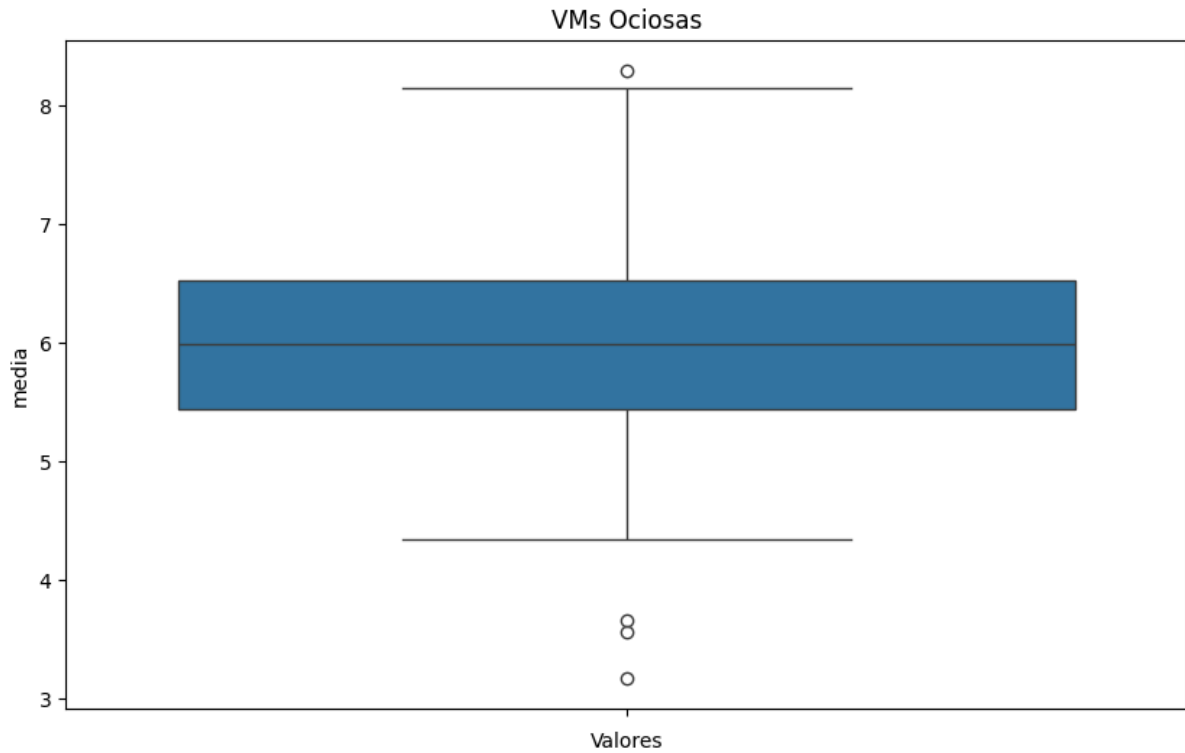
- Python 3
- Matplotlib
- Seaborn
- Pandas
- Scipy
- Numpy

4. Análise Realizada:

Neste estudo sobre a ociosidade de máquinas virtuais, iniciamos com a instalação das bibliotecas necessárias para a análise estatística. Em seguida, importamos as bibliotecas e carregamos os dados. Realizamos o teste de hipótese para verificar se os dados seguem uma distribuição normal. Com um nível de significância de 5%, concluímos que é possível afirmar que os dados são normalmente distribuídos, com base no resultado do teste e na inspeção do Q-Q plot.



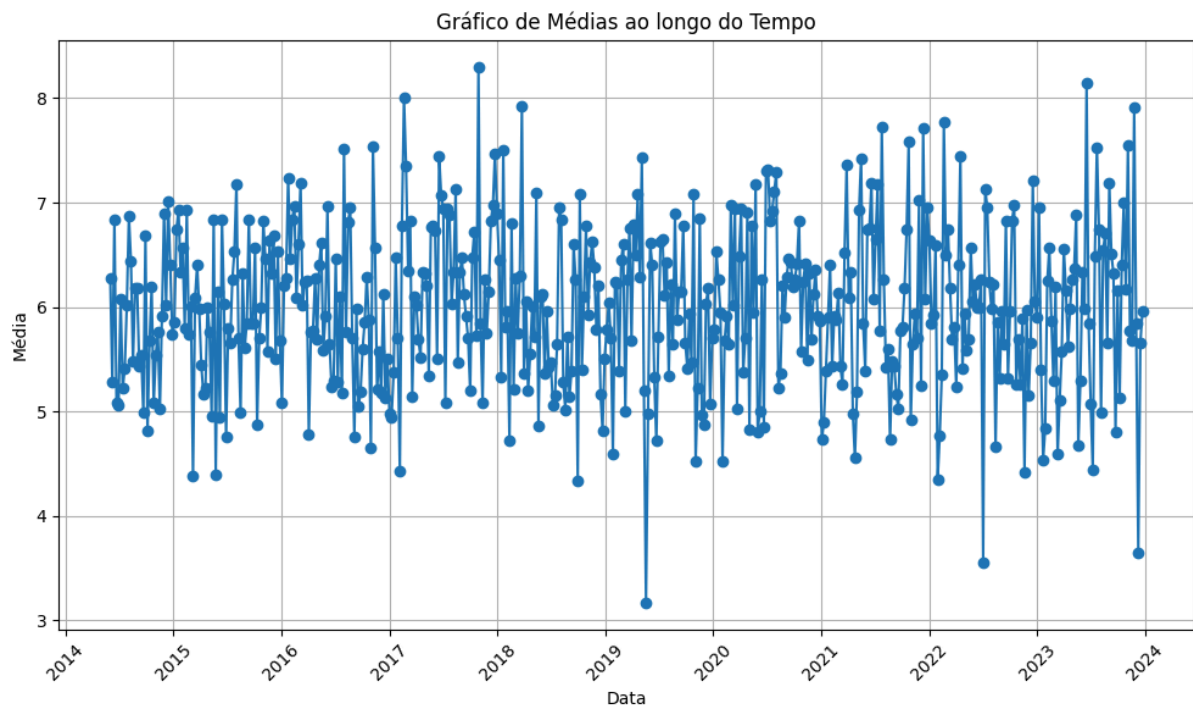
Para uma compreensão visual da distribuição dos dados, construímos um boxplot e um histograma. Ambos os gráficos indicaram uma distribuição razoavelmente simétrica das máquinas ociosas.



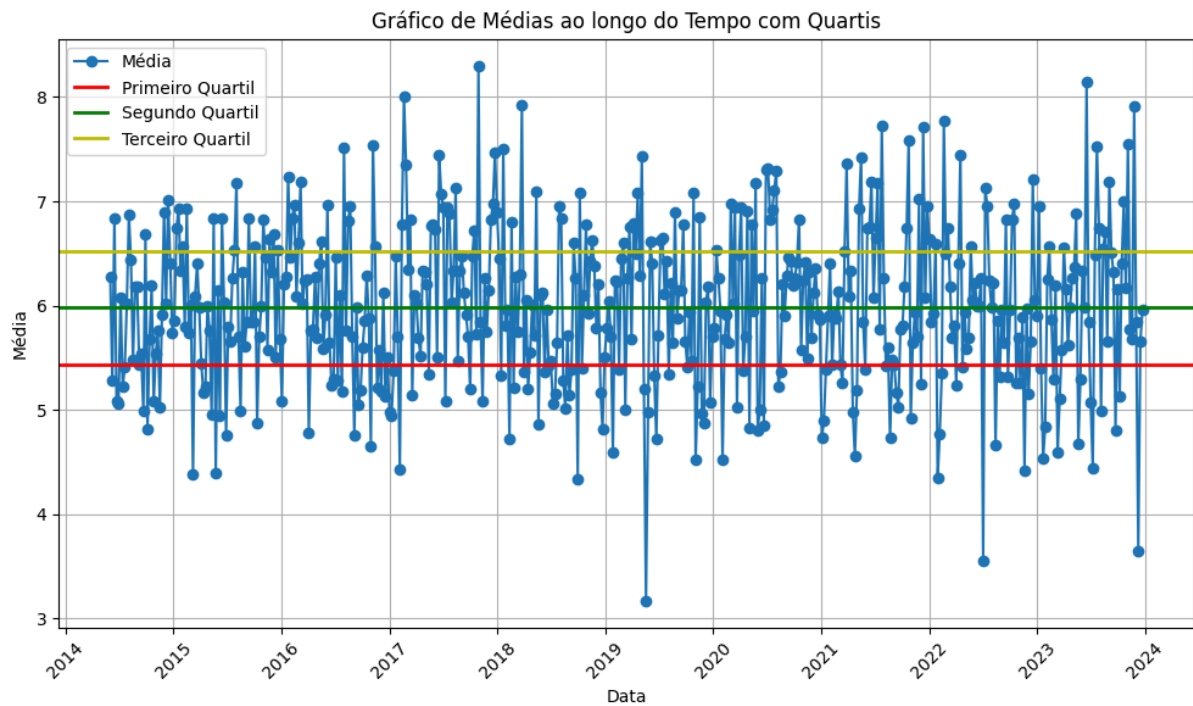
Prosseguimos com a análise descritiva das medidas de tendência central e distribuição. Calculamos a média, mediana, moda, desvio padrão, variância, coeficiente de variação, quartis e curtose:

- Média: 5.995839999999999
- Mediana: 5.985
- Moda(s): [5.69, 5.7, 6.27, 6.4, 6.84]
- Desvio padrão: 0.7760306317942391
- Variância: 0.602223541482966
- Coeficiente de variação: 12.942817550072036
- Primeiro quartil: 5.4375
- Segundo quartil: 5.985
- Terceiro quartil: 6.5225
- Curtose: 0.08692593408324312

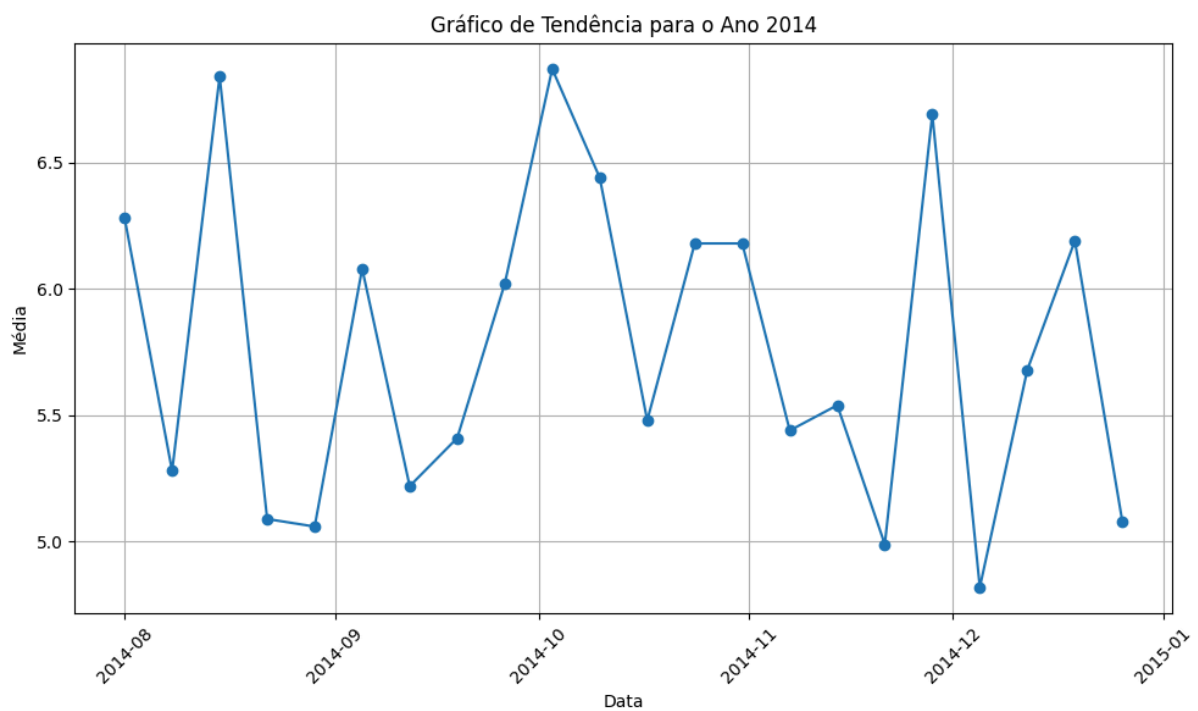
Além disso, determinamos as datas da sequência e plotamos um gráfico de médias ao longo do tempo. Isso nos permitiu visualizar padrões temporais na ociosidade das máquinas e sua evolução ao longo dos anos.



Observamos também essa distribuição temporal em relação aos quartis, fornecendo insights sobre a variação da ociosidade ao longo do tempo.



Em seguida, examinamos as tendências anuais, destacando padrões sazonais ou anomalias a cada ano.



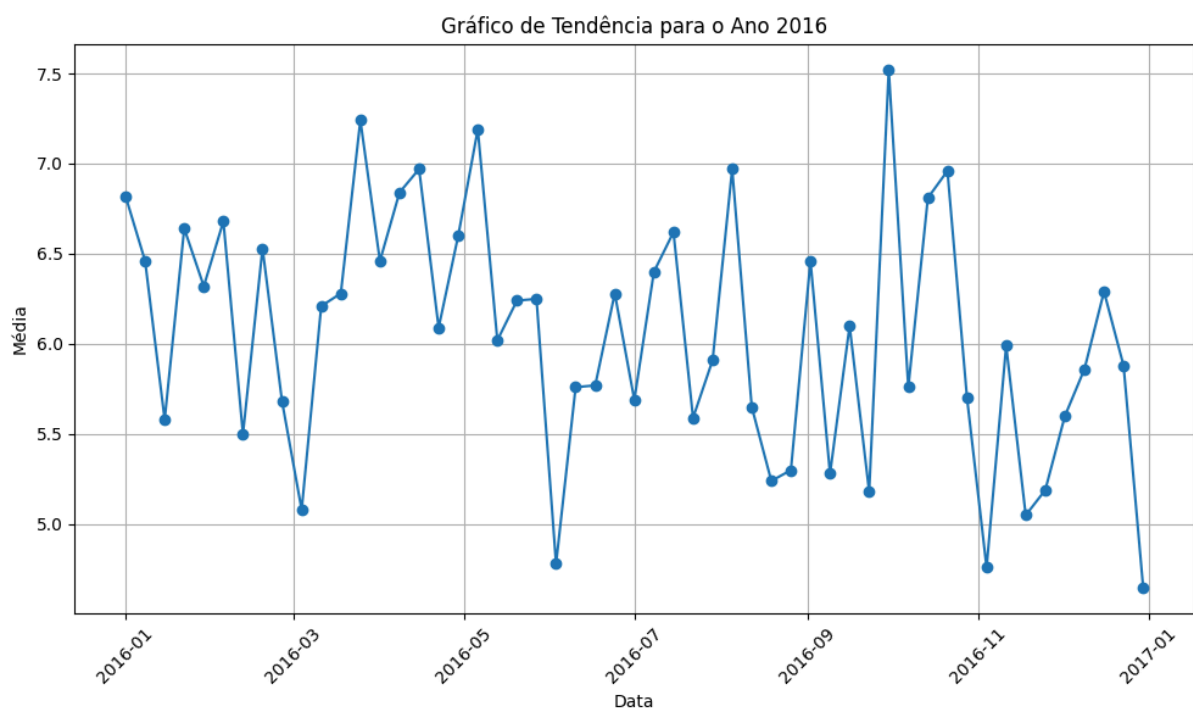
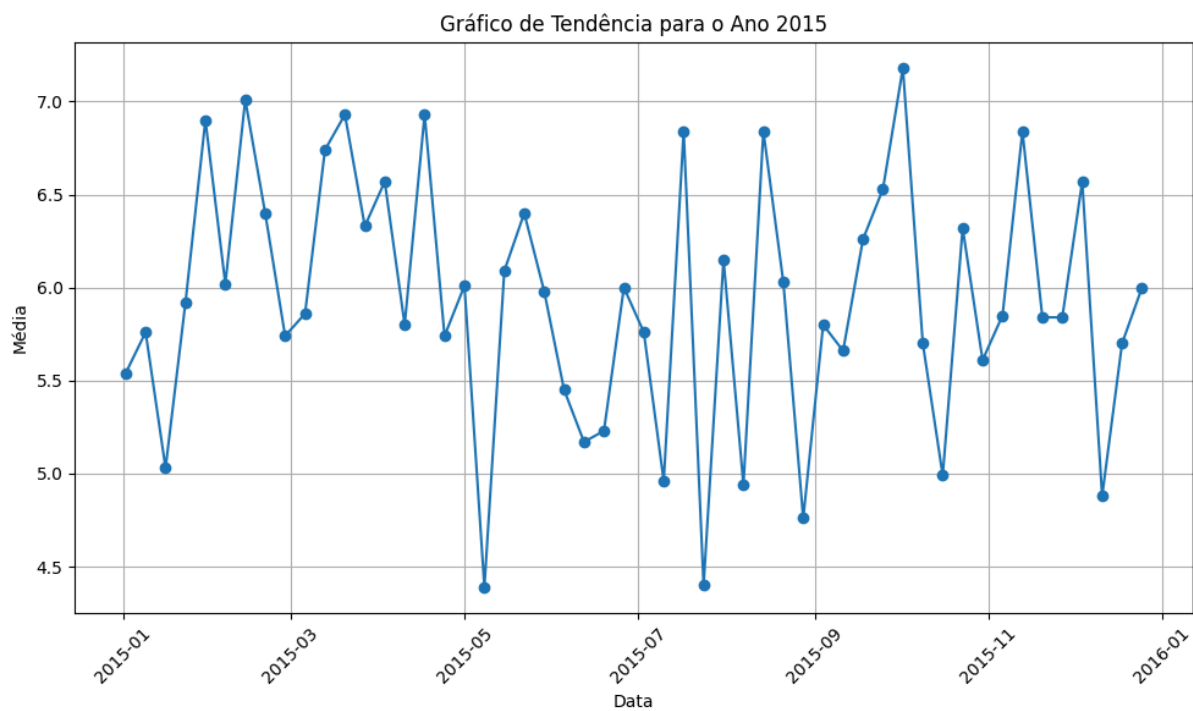


Gráfico de Tendência para o Ano 2017

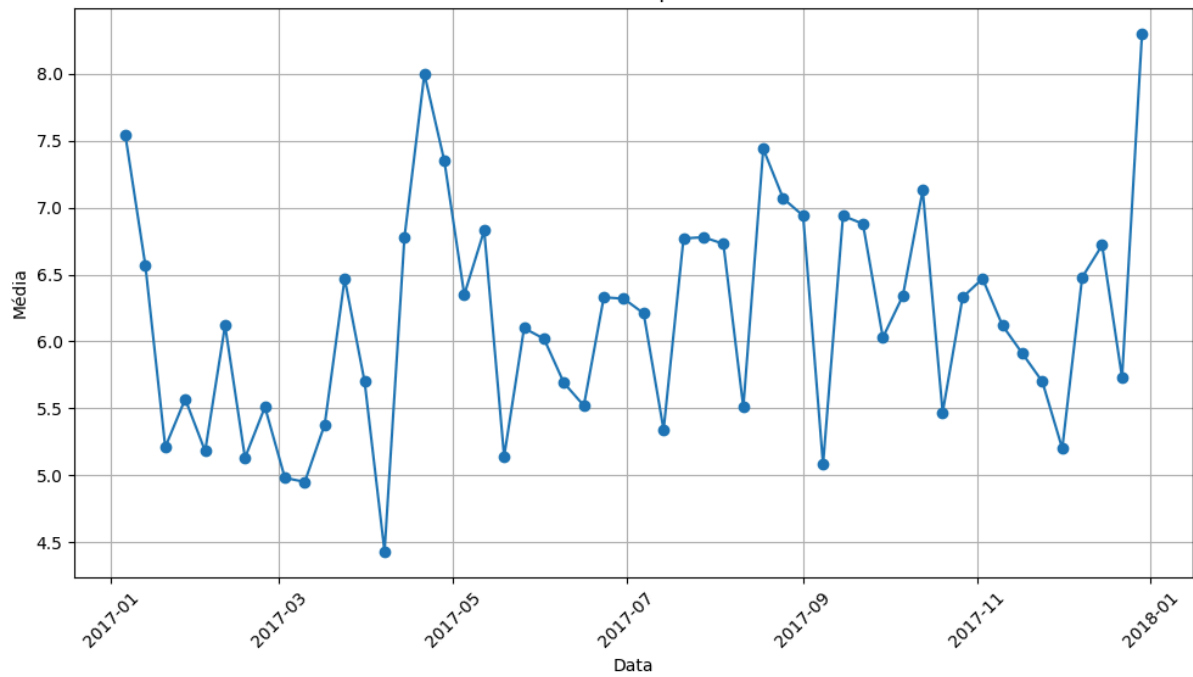


Gráfico de Tendência para o Ano 2018

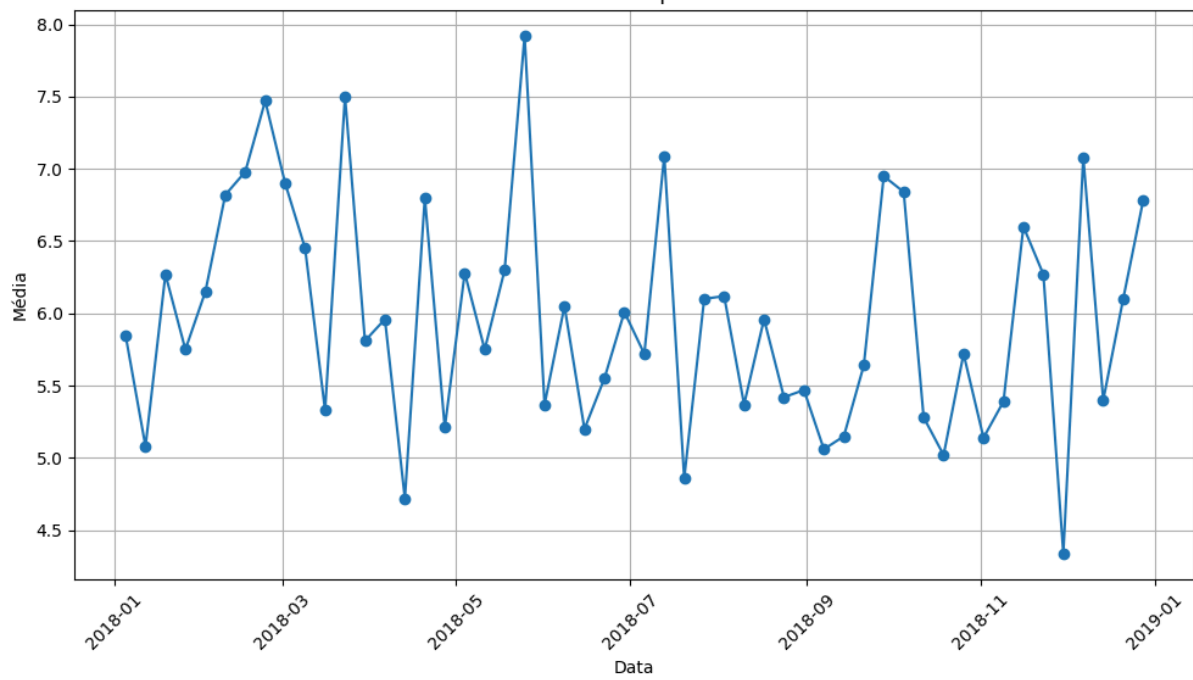


Gráfico de Tendência para o Ano 2019

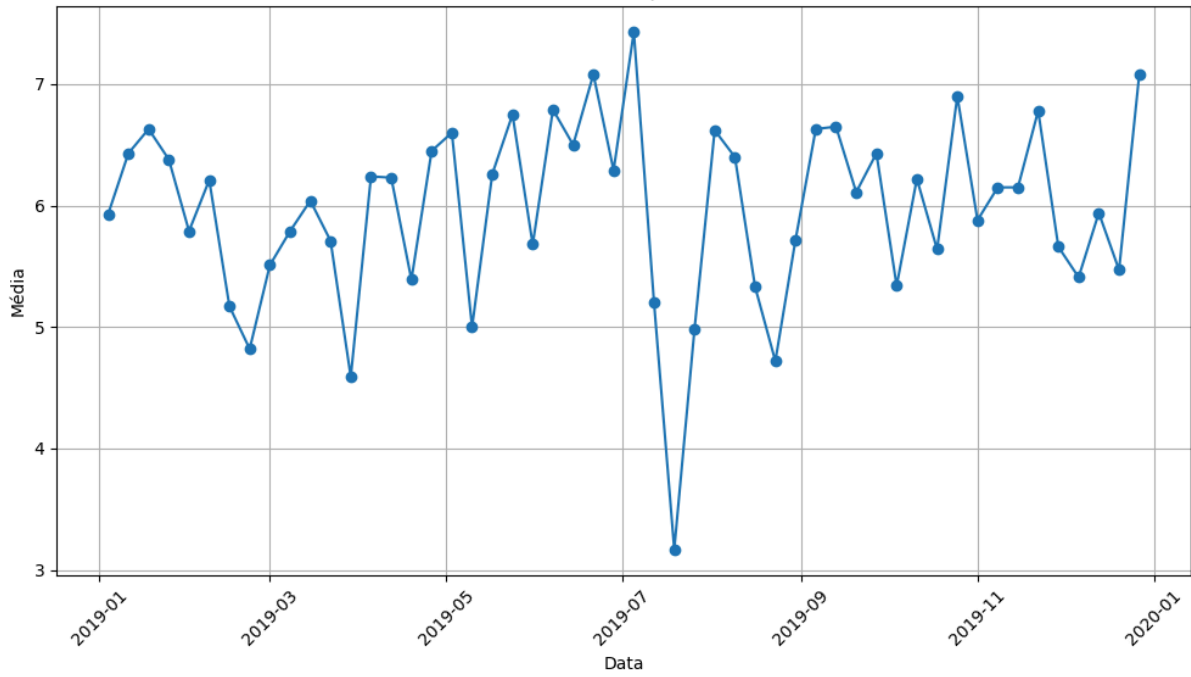


Gráfico de Tendência para o Ano 2020

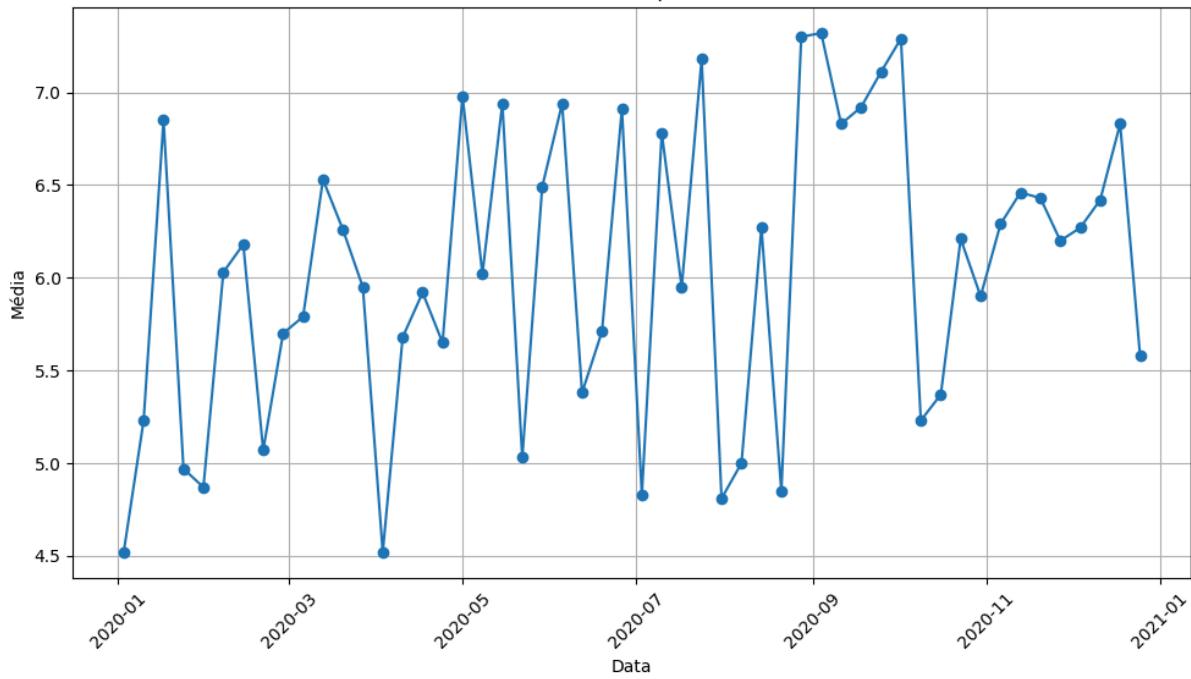


Gráfico de Tendência para o Ano 2021

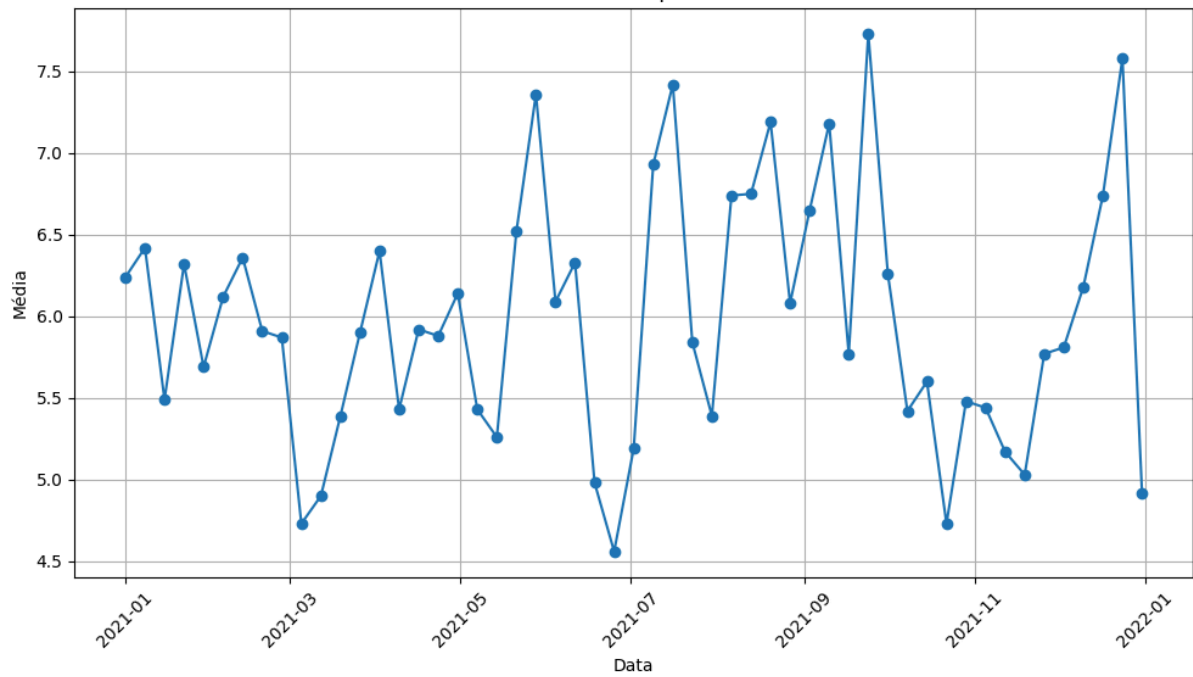
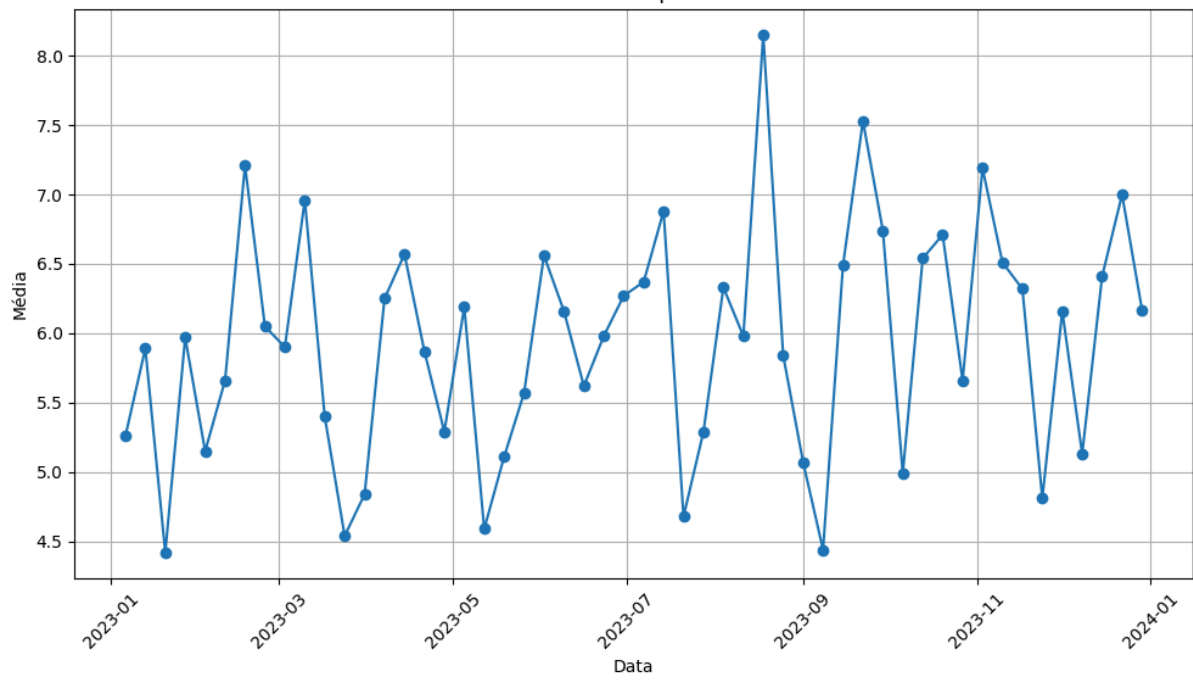


Gráfico de Tendência para o Ano 2023

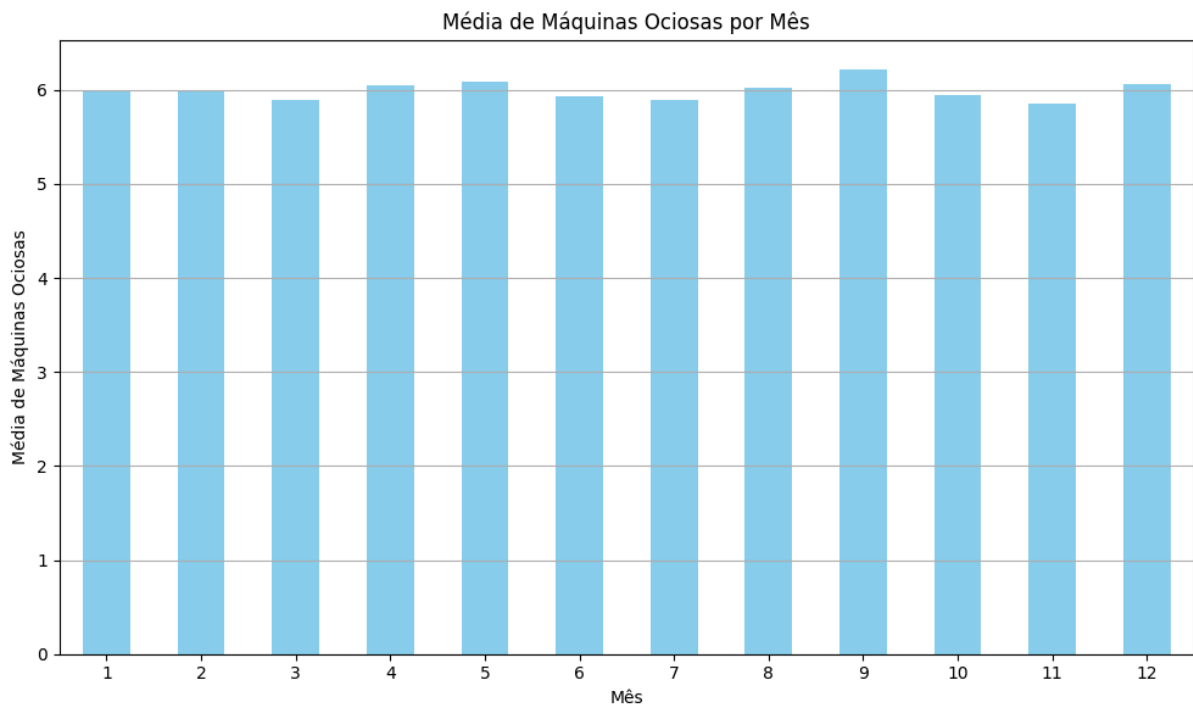




Além disso, exibimos essas tendências anuais em conjunto, destacando o comportamento ao longo dos anos.



Por fim, analisamos a média de máquinas ociosas por mês, revelando baixas variações mensais na utilização das máquinas virtuais.



5. Resultados e Conclusão:

Tendo em vista a análise realizada e as informações obtidas, recomendamos ao nosso cliente que reduzisse o número total de Máquinas Virtuais do ambiente virtual em 3, uma vez que apenas em casos outliers o número de máquinas ociosas foi menor. Com a redução para 19 máquinas, a Google economizaria aproximadamente 12% de seus gastos mensais (\$18.000,00). Isso resultaria em uma economia substancial a longo prazo e uma utilização mais eficiente de seus recursos de TI.

Portanto, a DataOpti foi capaz de fornecer uma solução eficaz e inovadora para a Google, permitindo que ela reduzisse significativamente seus custos com máquinas virtuais enquanto mantinha a eficiência operacional. A análise detalhada e a utilização de tecnologias avançadas garantiram o sucesso do projeto, destacando a expertise da DataOpti em transformar dados em insights acionáveis.