

Número USP	Número USP	Número USP	Número USP	Feedback
12692645	14654166	14670752	13661979	<p>Excelente trabalho! Mostra o cuidado e interesse do grupo com a atividade. Parabéns!</p> <p>A base de dados foi bem discutida, com uma boa motivação e análise exploratória. O tratamento de dados foi adequado. Muitas visualizações interessantes foram apresentadas e discutidas. Uma observação que não vi vocês comentando: parece que houve uma certa redução de nota ao longo dos anos, o que achei interessante.</p> <p>Sobre o teste de hipótese: faltou vocês explicarem com um pouco mais de calma os parâmetros utilizados para o teste, como ele foi realizado, etc. Mas foi bem contextualizado e seu resultado discutido.</p> <p>Os slides ficaram bons e bem apresentados. Pontos extras pela análise de correlação.</p>
				<p>Cuidado: texto da capa está "vazando" para a página abaixo.</p> <p>A introdução aos dados escolhidos é boa, assim como o tratamento. Porém, faltou apresentar a estatística descritiva dos dados: médias, desvio padrão, quartis, e outros dados de interesse que resumem os dados como um todo.</p> <p>Algumas linhas estão sendo quebradas aleatoriamente no texto. Sem divisão correta de parágrafos.</p> <p>Figuras precisam ter legendas, com seu nome e o que a figura apresenta.</p> <p>Contextualize as avaliações: Quais são os 12 grandes?</p> <p>Falem um pouco mais sobre o boxplot 2.2, não é boa prática deixar para o leitor extrair as informações. Ex: "A maioria da média dos gols concentra-se entre 1.1 e 1.3 gols, com a mediana próxima de 1.2 gols, mostrando uma distribuição similar à normal."</p> <p>Quais informações podemos extrair da figura na seção 2.3? As figuras precisam ser explicadas. Ex: observa-se uma concentração maior de média de cartões entre 0.95 e 0.96, mas com todos os times apresentando uma média próxima de 1 cartão por jogo.</p> <p>O gráfico de 2.2 e 2.3 estão cortados e sem legenda nos eixos y.</p> <p>Discussão da figura da 2.4 foi melhor apresentada na seção 2.5, mas poderia estar na própria seção. Além disso, poderia ser destacado como a média é muito menor, estando concentrada próxima de 0,05 cartões por partida.</p> <p>Análise da Seção 2.6 foi sucinta, mas suficiente para os dados. Porém... Qual a média de gols por jogador? Ou pelo menos por atacante (para ter uma comparação mais "honestas")? O dado bruto é interessante para comparar o desempenho de cada um, mas apresentar uma média como "baseline" (comparação base) ajuda nestes casos.</p> <p>Scatterplots são mais recomendados para quando existem vários pontos em "y" para cada "x". Para o dado mostrado na seção 2.7, ele é equivalente a um gráfico de barras ou de linha. No fundo, foi apresentado um box plot e variações de gráficos de coluna. Um heatmap de cartões por time e rodada poderia ser muito interessante, por exemplo: existe um aumento ou redução de cartões ao longo das rodadas? Alguns times teriam esse comportamento? Esse comportamento poderia ser relacionado ao time estar melhor ou pior no campeonato? O Heatmap poderia ser até customizado: time x rodada, com a cor sendo o número de cartões, e uma anotação mostrando a classificação do time a cada rodada.</p> <p>O teste de hipótese está certo, mas a interpretação tem alguns erros. Primeiro, um poder estatístico de 0,57 não é tão bacana, mas é aceitável. O primeiro teste usado, de teste-t, deveria ser usado depois de verificar a normalidade com o de Shapiro-Wilk que usei no exemplo em sala, ele é um teste para amostras gaussianas. Se rejeitado, aí seria aconselhável usar o teste de Mann-Whitney, não paramétrico. Outro ponto: ao ter um p-valor maior que o alfa, a hipótese não é descartada, só que isso significa que as médias das amostras comparadas NÃO pode ser rejeitada como diferentes. Ou seja, são consideradas iguais. Além disso, poderiam ter descrito como foi realizada a análise e seus parâmetros ao invés de colocar todo o código aqui, mas tudo bem nesta parte.</p> <p>A conclusão foi bem bacana, mas senti uma falta das discussões mais pontuais em cada visualização.</p>
14785493	14777754	14577917	14675819	
14606961	14589789			<p>O trabalho está muito bom. Com as análises requisitas apresentadas e discutidas com cuidado e profundidade. O conjunto de dados é bem introduzido, a análise exploratória é bem completa, discutida no texto, apresentada em tabelas, e com alguns gráficos. Eu só colocaria a seção de análise da amostra junto com os resultados, para ficar tudo que é discutido com maior profundidade "perto", e a Seção 2 ficaria para a descrição geral das atividades e dos dados. A tabela que seria a 3 está sem legenda e com as letras passando as grades.</p> <p>O documento poderia ter ficado ainda mais organizado se feito em LaTeX, mas está bem formatado, e o conteúdo está excelente. Foi uma análise muito completa, objetiva, detalhada, e com grande maestria de visualizações e técnicas de análise de dados. Gostei bastante.</p> <p>Bônus por regressão, clusterização, e correlação.</p>
14572613	14581075	14748104	13827570	<p>Excelente trabalho! Mostra o cuidado e interesse do grupo com a atividade. Parabéns!</p> <p>Cuidado com as fontes: Na Figura 1, as legendas ficaram bem pequenas. Aliás, alguns dados estão em inglês, lembrem-se de padronizar para a linguagem do documento.</p> <p>Análise visual muito completa e bem discutida. O mesmo para o teste de hipótese, que inclusive foi antecedido de uma análise de correlação (um leve bônus por isso).</p> <p>Senti falta de uma seção de conclusão, para fechar a análise como um todo. Porém todas as seções e gráficos apresentados foram bem discutidos separadamente e também contextualizados com os objetivos propostos.</p>
12725411	14571981	14783601		<p>Excelente trabalho! Mostra o cuidado e interesse do grupo com a atividade. Parabéns!</p> <p>Tabela 4: começa contando no 0 e pula o 9. Alterem o nome das variáveis para facilitar a visualização das tabelas, como na Tabela 5: nomes grandes e difíceis de ler. Para as análises estatísticas, é comum apresentar uma tabela com o p-valor obtido entre os pares observados.</p> <p>Ao referenciar links, utilizem notas de rodapé ou citações de links como referido pela ABNT (ou o estilo que forem utilizar) https://canaltech.com.br/educacao/como-citar-e-referenciar-um-site-normas-abnt/</p> <p>Mantenham a consistência de termos (database ou base da dados?)</p> <p>Ao usar estrangeirismos, usem o itálico.</p> <p>Sempre comecem seções e subseções com texto, nunca com códigos, imagens, tabelas, etc.</p> <p>1094 observações é bastante. Com 1023 obs vocês teriam um poder de 0,998 para alfa 0,05 e ES 0,2. Deve ter rolado um arredondamento de 0,998 para 1. Adicionei um plot para exemplificar.</p>

14558311	14745682	11912849		<p>O trabalho, infelizmente, pareceu ter sido feito com certa pressa, sem um carinho maior com o que foi apresentado e na análise dos dados, que, como disse, era a parte principal deste projeto. É uma base com bastante potencial para ser explorada.</p> <p>O arquivo necessita de melhor formatação. Tabelas não centralizadas, linhas puladas em números aleatórios (às vezes 1, às vezes 2). Sentenças separadas por linhas puladas ao invés de parágrafos um abaixo do outro.</p> <p>Tabela 1 e 2 são quase ilegíveis sem zoom.</p> <p>Tabela 1 tem um texto "imagem 4" acima dela.</p> <p>Os dados foram expostos nas tabelas e não foram discutidos, sequer apresentados o que são cada um.</p> <p>Página 5: o que é a imagem 16? A imagem com o dado citado é a 2.</p> <p>As avaliações das imagens existem, diferente das tabelas exploratórias, mas são super sucintas. Ex: Os filmes concentram-se entre 2018 e 2020. Mas o que isso significa? Por que teve a redução em 2022? Por que um crescimento tão grande de 2015 a 2018?</p> <p>O que a Tabela 3 apresenta que instigou-os a fazer as análises sobre estes dados? Por que ela foi apresentada se não é discutida? Não seria melhor só apresentar os gráficos, neste caso?</p> <p>Não existe nenhuma análise exploratória, como verificação de média, quartis, desvio padrão, valores nulos, etc. para os dados.</p> <p>Existem 2 "Gráfico 3".</p> <p>Gráfico 4 está achatado e difícil de ler. Quais as siglas dos países apresentados? Não se pode esperar que o leitor saiba todas essas siglas de cabeça. Devem ser, ao menos, apresentadas no texto.</p> <p>Novamente, os gráficos 4 e 5 são minimamente discutidos. Apenas 1 informação é apresentada em uma única sentença.</p> <p>A escolha de paleta de cores do Gráfico 4 e 5 é ruim, pois as cores se repetem. Como comentei com vocês em diferentes aulas, o Python permite escolher paletas diferentes para estes casos. Um exemplo: aquele roxo maior em 2017 a 2019 é a Espanha ou o Canadá (provavelmente Espanha pela ordem de apresentação). O Azul maior é o Brasil ou os EUA?</p> <p>Existem 2 "Gráfico 5". O dado apresentado no segundo Gráfico 5 e do 6 e 7 é o total de votos. O mais votado não necessariamente é o que tem a melhor avaliação. Aliás, com o "hue" era possível separar as produções por país no gráfico. Como apresentado, não é possível observar a informação apresentada de que 90% dão produções americanas.</p> <p>O Gráfico 7 é sequer discutido.</p> <p>A análise do teste de hipótese não é apresentada. Quais os dados que foram comparados? Quais as médias analisadas? Como o resultado deu 1, que é quase impossível dar? Não está presente o código da análise, então só posso chutar que foi utilizado o teste de poder estatístico, e não o teste t-student em si. Mas não consigo saber quais os dados usados e parâmetros.</p> <p>A conclusão é super sucinta e rasa. Os dados encontrados foram resumidos em 3 sentenças. Mas o que esses dados podem significar? Um maior investimento na indústria cinematográfica? Maior audiência de streamings? Por que séries foram mais produzidas? Existe uma preferência deste formato para consumidores da Netflix?</p> <p>Uma hipótese bem simples e que já ajudaria a formar alguma discussão: será que a nota caiu exatamente por ter uma maior quantidade de produções, o que pode implicar que estão sendo feitas de maneira apressada?</p>
13783183	14611644	13833960	5340621	<p>O trabalho é muito bom, mas faltaram cuidados com a apresentação dos dados (imagens pequenas, texto com alguns erros, imagens colocadas todas ao final e algumas nem referenciadas pelo texto).</p> <p>Cuidado: não é uma boa prática falar sobre todos os dados e depois agrupar todas as visualizações ao final. O recomendado é discutir cada imagem separadamente, interligando texto e imagem. Do modo que foi feito, fica difícil lembrar da informação lida até chegar na imagem, ou ficar navegando entre imagem e texto.</p> <p>Seria legal ter tabelas e elementos mais gerais sobre os dados, além do que foi posto em texto e apresentado nas visualizações iniciais, para uma análise exploratória mais robusta.</p> <p>Existem alguns erros de escrita ao longo do texto, é importante revisar, ou ao menos usar um revisor automático como grammarly, language tools, ou mesmo o do google.</p> <p>Eu usaria um violin plot na Figura 2, pois, apesar de passar uma ideia similar ao scatterplot usado, como tem muitos pontos, o boxplot evitaria essa sobreposição e manteria a informação da "quantidade" de dados numa posição similar no eixo Y.</p> <p>Eu entendi a ideia das Figuras 7 e 8, mas não sei se são as melhores visualizações para os dados em questão. A 8 em especial fica muito confusa sem ter lido o texto antes. Talvez a simplicidade de um gráfico de linha com a média (das médias) seria o suficiente. Uma sugestão um pouco mais complexa, um gráfico de linha com as linhas da média, mediana, maior valor, menor valor, e quartis de cada ano. Eu já vi usarem em alguns lugares e fica bacana.</p> <p>As visualizações e suas discussões são boas. Assim como a análise estatística.</p> <p>Figuras 11, 12, e 13 estão bem pequenas. Fica difícil ler.</p> <p>As figuras de 10 a 13 não foram referenciadas. Fica difícil saber o que cada uma representa.</p> <p>Bônus pelo trabalho de agrupamento.</p> <p>Seria importante apresentar de uma maneira mais visual (ex: tabelas) o teste de hipótese. Fica difícil acompanhar todo o raciocínio pelo texto.</p> <p>Senti falta de uma conclusão um pouco mais detalhada do artigo como um todo e da parte extra.</p>
14614626	9328096	15111714	14585955	<p>Excelente trabalho! Mostra o cuidado e interesse do grupo com a atividade. Parabéns!</p> <p>Seria interessante uma análise mais geral para a exploração dos dados, como a info do dataframe que mostrei em sala. A análise exploratória gráfica e textual é interessante, mas carece de uma visão mais geral dos dados, como tabelas, para ajudar a justificar as remoções e tratamentos que fizeram, por exemplo.</p> <p>Vocês e outros grupos dividiram os filmes por décadas. Mas, será que dividir por períodos um pouco menores não ajudaria? Como triênios (assim daria para incluir 2020 a 2022, por exemplo). Neste caso, poderia-se até descartar tudo antes de 2000 e focar nos 20 anos mais recentes.</p> <p>Lembre de traduzir os nomes de colunas para padronizar a linguagem do relatório.</p> <p>Excelente aprofundamento da análise estatística, com histograma e quantis, além do teste de Anderson-Darling. Adicionei um bônus para compensar o desconto na análise exploratória.</p> <p>Senti falta de uma seção de conclusão, para fechar a análise como um todo. Porém todas as seções e gráficos apresentados foram bem discutidos separadamente e também contextualizados com os objetivos propostos.</p>
14562528				<p>Trabalho muito interessante. Mostra o cuidado e interesse do grupo com a atividade.</p> <p>Mas faltou o teste de hipótese. (Várias análises mostradas poderiam ter dado um seguimento legal para validação estatística das tendências).</p> <p>Cuidado com a consistência do espaçamento nos slides. Lembre-se de numerar slides.</p> <p>Cuidado com a escala de gráficos, especialmente de barras. Ao variar a escala, fica mais difícil mostrar o tamanho real da diferença entre grupo A e B, como nos últimos gráficos. Uma sugestão seria forçar a escala de 0 a 1 em todos os que fazem sentido, para ficar mais fácil ver a discrepância.</p> <p>No caso do scatterplot do slide 9, a questão que você comenta de serem 6 pessoas com 55 anos poderia ser mostrada no gráfico através de um bubble plot, que usa a quantidade de incidências como tamanho do círculo. Ou, menos intuitivamente, com um hue, que a cor indicaria quantas incidências tem.</p> <p>Nos gráficos de linha de correlação que você mostra, em geral, podem ser substituídos por um gráfico de barras plotado junto com um gráfico de linhas. Normalmente é melhor, pois, como você viu, as linhas do jeito que você fez "escapam" pros valores abaixo e acima dos reais, causando uma impressão de que, por exemplo, existe uma correlação negativa e outra acima de 1.</p> <p>Um exemplo simples seria o gráfico que deixei abaixo, que poderia ser feito com esses comandos:</p> <pre>plt.bar(df['Year'], df['Sample Size']) plt.plot(df['Year'], df['Sample Size'], '-o', color='orange')</pre>

13874841	10351727		<p>O conjunto é apresentado de maneira sucinta, mas suficiente. Seria interessante uma análise mais geral para a exploração dos dados, como a info do dataframe que mostrei em sala. A análise exploratória gráfica e textual é interessante, mas carece de uma visão mais geral dos dados, como tabelas, para ajudar a justificar as remoções e tratamentos que fizeram, por exemplo. Seria interessante colocar mais dados do dashboard iterativo no relatório. Lembrem-se que o foco da correção é o relatório, e ele não precisava ser tão curto como fizeram: poderia ter mais alguns exemplos para ressaltar o trabalho em código que fizeram.</p> <p>Para a análise dos dados: é preciso descrever as imagens, analisar o que está apresentado, ao invés de apenas apresentar a figura, como não foi feito nas Figuras 1 a 4, mas foi feito nas seguintes. Começaram a discussão do teste estatístico, mas apenas falaram que o Shapiro-Wilk testa normalidade e... acabou a seção (?). Precisa falar sobre o resto.</p> <p>Bônus pela clusterização, porém: o que exatamente o gráfico mostra? O que aquela relação pode significar?</p> <p>A discussão, no geral, poderia ser mais completa, assim como a conclusão, que foi bem sucinta.</p> <p>O artigo está bem formatado e apresentado.</p>
10748177	11299940		<p>Dupla não registrou os dados na planilha como foi pedido.</p> <p>Não foi enviado o arquivo com o código ou link para o Colab.</p> <p>Cuidado com pontuação adequada:</p> <p>"Esse trabalho é o relatório final da disciplina da Introdução à Ciência de Computação 1. É uma análise estatística do banco de dados das 30000 músicas do Spotify. O tópico foi escolhido pois a música é muito importante na formação de todos nós..."</p> <p>É importante deixar a referência dos dados utilizados.</p> <p>Por que a conversão para CSV trouxe tantos problemas? Não possuo o código para sequer analisar os problemas relatados.</p> <p>Os dados não foram analisados. Só foi apresentada a tabela, e dito que o valor de decibéis está errado. Mas e os que estão certos? O que a tabela 1 apresenta? Não se deve apenas mostrar a tabela, deve-se discutir sobre seus dados.</p> <p>Diz-se que apenas alguns dados foram descartados, mas poucas colunas foram de fato utilizadas, por que essa divergência de informações?</p> <p>Não se deve começar uma seção sem texto algum, e todas as subseções devem ter títulos.</p> <p>Nenhuma visualização foi discutida, apenas apresentada.</p> <p>Teste de hipótese foi baseado em que? O que é a proposta 6? Quais parâmetros foram usados? O que esses valores significam para análise?</p>
14713471	4202166		<p>Contexto muito bem apresentado. Ótima apresentação e exploração da base de dados, com tabelas e uma análise textual. Porém, senti falta de uma discussão mais específica sobre os dados da Tabela 4. Boas visualizações, mas senti falta de uma discussão um pouco maior na Figura 3. O teste estatístico poderia ter dados em tabela para ajudar a visualizar, mas está bem explicado e discutido. Extras por correlação, análises de agrupamento e PCA. Seção 2.2.1 - Referência à Tabela 4 tem um '3' antes. 2.2.2: Visualizações está com letra minúscula. Figura 2 ficou com a legenda pequena para ser vista no pdf. Sugestão: rotacionar. Figura 4 ficou difícil de ler as legendas no pdf, poderia ter sido dividida em 4 figuras separadas ou só 1 coluna. Seção 2.3: alfa de 0.01 é bem baixo, não poderia ser 0.05 e talvez teriam mais normais? Ao adicionar anexos ao documento, é importante deixá-los anexos ao pdf, e não como material suplementar no envio. É bom distinguir os dois. Lembrem-se de manter tudo no mesmo idioma, inclusive legendas de gráficos. Figura 5 - Seria importante ter 1 cor para cada aminoácido, e a legenda com as cores.</p>

Power of Test

