

Introdução.

O mercado de trabalho urge seguir o avanço exponencial da tecnologia. Consequentemente, o futuro do trabalho exige dos seus profissionais além do conhecimento técnico: ele requer uma capacidade contínua de autoaprendizagem, paciência, dedicação e resiliência.

Para os estudantes que estudam especificamente a tecnologia e suas diversas áreas como Ciência da Computação, Análise e Desenvolvimento de Sistemas, Engenharia da Computação, entre tantos outros, o aprendizado deve ser orientado sob a competência de “aprender a aprender”.

Nesse sentido, o desempenho acadêmico torna-se reflexo da capacidade do estudante de gerenciar um valioso recurso: o tempo. Seja para foco, estudo, atividades físicas e bem-estar psicológico em prol de uma saúde mental estável. Mas o que vai definir o sucesso e o alto desempenho? Sono, saúde mental, atividade física, dieta balanceada, tempo de tela?

Para responder tais perguntas, esse trabalho utiliza estatística descritiva. Para isso, usei conceitos de medidas de posição e dispersão, separatrizes, correlação entre variáveis, histograma e curva normal em um conjunto de 909 estudantes. Assim, irei identificar e quantificar as relações entre os hábitos de vida e desempenho acadêmico.

Compreender os padrões que vierem dessas análises é fundamental para o Futuro do Trabalho. Ao identificar quais hábitos estão fortemente correlacionados com o alto desempenho, poderei fornecer insights práticos tanto para as instituições de ensino, na criação de estratégias de apoio mais eficazes, quanto para os próprios alunos, no desenvolvimento de suas rotinas.

Metodologia.

A metodologia adotada foi quantitativa e descritiva, focada em aplicar conceitos estatísticos para analisar um conjunto de dados simulados. Todo o processamento e análise foram realizados no ambiente do Google Colab, utilizando bibliotecas como *Pandas* para manipulação e bibliotecas como *Matplotlib* e *Seaborn* para visualização gráfica.

Com o foco na *análise do desempenho de estudantes em cursos de tecnologia*, utilizei o dataset “Student Habits vs Academic Performance” (*Kaggle*). A variável dependente foi definida pela Pontuação do Exame, representado o desempenho acadêmico. As variáveis independentes foram relacionadas à base familiar, hábitos de vida, lazer, bem-estar e obrigações extras (trabalho além da faculdade).

Antes das análises, a base de dados passou por uma etapa de pré-processamento, que incluiu a tradução das variáveis para o português, remoção de dados ausentes (o que reduziu a base de dados de 1000 para 909 estudantes válidos) e exclusão da coluna de identificação do estudante, a qual não tem valor estatístico.

Na etapa de análise, medidas de tendência central e dispersão (média, mediana, desvio padrão, máximo e mínimo) foram aplicadas para descrever o conjunto de dados, o que foi feito através do método *describe()* da biblioteca Pandas. Além disso, as separatrizes (quartis) proporcionaram insights importantes e, por meio dela, consegui dividir os alunos em grupo de desempenho maior e menor.

Houve também a análise de correlação de Pearson, com o dataframe totalmente numérico para revelar relações lineares entre as variáveis. Por esse motivo, as variáveis categóricas (binárias, nominais e ordinais) foram previamente transformadas em números e os resultados foram representados graficamente em um heatmap. Por fim, a variável Pontuação do Exame também teve sua distribuição investigada por Histograma e Curva Normal, o que possibilitou avaliar a simetria e dispersão.

Resultados.

Em síntese, foi usado o método *describe()* para os cálculos de medidas de tendência central e separatrizes. Para uma análise enxuta e direta, tratei da nossa principal variável-alvo: Pontuação do Exame (marcador de desempenho acadêmico). Os resultados foram:

A amostra total foi de 909 pessoas; a média de notas foi de 69.56; o desvio padrão foi de 16.93; a menor nota foi 18.4 pontos, ao passo que a maior foi de 100 pontos; 25% dos estudantes (Q1) tirou uma nota menor ou igual a 58.4; 50% (Q2/Mediana) fez uma pontuação menor ou igual a 70.4; e 75% (Q3) fez uma nota menor ou igual a 81.3 pontos, ou seja, o quarto restante pontuou mais de 81.3 pontos.

No que diz respeito a segmentação dos grupos, através dos quartis: O grupo de alto desempenho, o qual ficou igual ou acima do terceiro quartil teve em média ao dia: 5.1 horas, 2.2 horas de Redes Sociais, 1.64 horas de Netflix, e 85.34% percentagem de Presença. Em contrapartida, o grupo de baixo desempenho (igual ou abaixo do primeiro quartil) marcou, por dia, 1.94 horas de estudo, 2.72 horas de redes sociais, 2.07 horas de Netflix e 83.19% de presença.

No Mapa de Calor (Heatmap) da Matriz de Correlação de Pearson: Os coeficientes mais fortes relacionados à Pontuação do Exame foram: Horas de Estudo/Dia (+0.82), Saúde Mental (Classificação) (+0.32), Horas de Rede Social/Dia (-0.17) e Horas de Netflix/Dia (-

0.17). Fatores como Porcentagem de Presença (0.10) e Nível de Educação dos Pais (-0.02) mostraram correlação próxima de zero.

A distribuição da Pontuação do Exame apresenta leve assimetria à esquerda, com média de 69.56 a qual é pouco menor que a mediana de 70,4. O histograma revela bimodalidade, destacando um pico principal entre 70 e 75 pontos e um segundo pico na nota máxima (100), que corresponde à moda.

Análise e Discussão

A análise inicial já apontava para uma turma complexa, pois, embora a Média (69.56) e a Mediana (70.4) estivessem próximas, sugerindo uma falsa simetria, o Desvio Padrão era muito alto. O histograma confirmou visualmente que a distribuição é "larga", espalhando-se da nota mínima à máxima.

Isso prova que a turma é heterogênea, e leva a pergunta central: o que explica, num grupo com perfil etário similar, alguns alunos irem tão bem e outros tão mal?

A resposta para essa pergunta começa com o formato da distribuição. embora a Média e a Mediana estejam próximas, a distribuição não é "normal", a análise gráfica revela dois fatos cruciais:

Assimetria Negativa: A curva de densidade mostra uma "cauda" longa à esquerda. Isso valida que as poucas notas muito baixas puxam a Média (69.6) para um valor ligeiramente *abaixo* da Mediana (70.4).

Bimodalidade Oculta: O fato de a Moda ser 100.0, enquanto a barra mais alta (classe modal) está em 70-75 revela polarização. A curva de densidade (KDE) confirma visualmente que há dois picos: a "massa" de alunos em torno de 70-75, e um segundo subgrupo de alunos de altíssimo desempenho que gabaritaram a prova (100.0).

A análise de segmentação foi a ferramenta que "dividiu" os alunos e respondeu a pergunta. O fator que realmente "divide" os grupos de alto (Q3) e baixo (Q1) desempenho é o Tempo de Estudo.

O grupo de alto desempenho (Q3) estuda, em média, 5.10 horas/dia. O grupo de baixo desempenho (Q1) estuda menos de 2 horas (média de 1.94 horas/dia). O grupo com notas melhores estuda quase o triplo de horas.

O Mapa de Correlação confirma: Pontuação do Exame e Horas de Estudo/Dia é de +0.82, uma relação positiva muito forte. Isso confirma que o tempo de estudo não é apenas *um* fator, mas o principal preditor de desempenho nesta análise.

O estudo, contudo, não é o único fator. A segmentação revelou que os hábitos pessoais de bem-estar são um fator decisivo, pois a correlação entre tempo de Saúde Mental e Desempenho foi de +0.32.

Observa-se um padrão: o grupo de alto desempenho é significativamente mais ativo fisicamente (41% exercita-se 5-6 vezes/semana) e reporta uma saúde mental muito superior (pico de notas 6-10). Em contraste, o grupo de baixo desempenho é largamente sedentário (39% exercita-se 0-1 vez/semana) e concentra-se na base da escala de saúde mental (notas 1-5).

A análise de correlação também corrigiu uma conclusão precipitada. Embora a segmentação tenha mostrado que o grupo Q1 gasta quase 1 hora a mais em lazer (Netflix e Redes Sociais), o Mapa de Correlação nega uma influência tão negativa: correlação negativa de -0.17, a qual é fraca.

Ou seja, "Tempo de Tela" não é vilão, mas sim o "Tempo de Estudo" é marcador de desempenho.

Finalmente, a análise confirma que **fatores contextuais não impactam** o desempenho nesta amostra. O Mapa de Correlação é a prova estatística: Porcentagem de Presença (0.10), Nível de Educação dos Pais (-0.02), Emprego de Meio Período (-0.02) e Qualidade da Dieta (0.01) são todos valores próximos de zero, indicando nenhuma relação linear com a nota.

Conclusão.

A análise revelou que o desempenho não é definido por fatores contextuais (como a educação dos pais ou ter um emprego), mas sim por *hábitos pessoais de autogestão*.

- O tempo dedicado ao estudo (Horas de Estudo/Dia) é o principal preditor de sucesso, com uma correlação positiva muito forte.
- O "motor" do estudo funciona melhor quando o aluno está bem. A saúde mental e a atividade física mostraram-se indicadores significativos, provando que o autocuidado é um componente do desempenho.
- A Gestão do Tempo é Mais Importante que a Abstinência: O "Tempo de Tela" (Netflix/Redes Sociais) revelou-se um distrator fraco, muito menos impactante que a ausência de estudo. A lição não é "parar" de ter lazer, mas sim "priorizar" o estudo.

Para o Futuro do Trabalho esses resultados implicam que o desenvolvimento da competência de "aprender a aprender" (o estudo ativo) e de autogestão (saúde mental e física) são essenciais para o sucesso na área da tecnologia.

Anexo Google Colab:

<https://colab.research.google.com/drive/1a1ZQ7mluoziTSB6u48cwZqVHdlSVyD6h>

Base de Dados Utilizada:

<https://www.kaggle.com/datasets/jayaantanaath/student-habits-vs-academic-performance?resource=download>