



UFRN - UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE

IMD - INSTITUTO METRÓPOLE DIGITAL

Relatório final de análise de dados das turmas da UFRN

Joamerson Islan Santos Amaral

NATAL-RN

2025

RESUMO

Com base na análise de dados das turmas da Universidade Federal do Rio Grande do Norte (UFRN) de 2023 e 2024, este estudo utiliza a linguagem Python e suas bibliotecas para examinar quase 60.000 registros do portal de dados abertos da instituição. Após um processo de tratamento que incluiu a remoção de dados dissonantes e a identificação de inconsistências na classificação de modalidades de ensino, a análise levantou discussões acerca da discrepância de oferta de turmas em diferentes campus, preferência dos alunos por professores internos, preferência de turnos de estudo e turmas com sobrecarga de solicitações de matrícula. O trabalho conclui que a análise de dados é uma ferramenta administrativa poderosa para fundamentar mudanças estruturais na universidade, podendo ser expandida com novas bases de dados para aprimorar a gestão acadêmica e promover o bem-estar da comunidade.

1. INTRODUÇÃO

No contexto da administração pública, a utilização de estudos baseados em dados como ponte para a realização de mudanças na estrutura das organizações é de suma importância. São estes estudos que possibilitam encontrar problemas na sua atual composição, dentre os quais é possível elencar: má alocação de recursos para setores que possuem demandas maiores, má otimização de horários de funcionamento e configuração pouco otimizada das suas repartições.

Dentre as instituições que mais sofrem com estes problemas estão as universidades federais, que, pelo seu grande porte e consequente complexidade administrativa, acabam por vezes falhando em entender o seu próprio funcionamento. Neste sentido, o presente projeto visa desenvolver uma análise estatística acerca das turmas ofertadas na Universidade Federal do Rio Grande do Norte (UFRN), de modo a compreender os seus fenômenos e incoerências.

2. DESCRIÇÃO DA BASE DE DADOS

Para iniciar a análise de dados das turmas, o presente projeto se valeu do repositório de dados abertos da UFRN¹, um conjunto público de dados que podem ser acessados por qualquer cidadão, a fim de cumprir com a transparência devida pelo setor público. Deste repositório, retirou-se especificamente o conteúdo relativo às turmas da UFRN, ofertadas nos anos de 2023 e 2024. Este conteúdo era formado por quase 60.000 registros únicos e dispunha de diversas informações interessantes para a análise deste trabalho, a exemplo de:

- Campus de oferta da turma, explicitando em qual dos campus da UFRN aquela turma estava sendo ofertada;
- Carga horária da turma em horas;
- Quantidade de solicitações efetivadas por discentes durante o período de matrícula;
- Ano e período em que a turma em questão foi aberta;
- Classificação das turmas conforme os modelos de ensino presencial e à distância, dentre outros.

A estrutura bruta dos dados requer que um tratamento de dados seja realizado, a fim de que as etapas de análise e utilização de algoritmos estatísticos possam ser realizadas sem maiores problemas. Os procedimentos de melhoria de qualidade das informações serão descritos a seguir.

3. METODOLOGIA

O conjunto de dados abertos da UFRN, apesar de satisfatório, carecia de um tratamento adequado, pois alguns de seus campos apresentavam informações vazias ou incompletas, o que forçou o projeto a passar por um delicado processo de limpeza da base de dados bruta, advinda do repositório supracitado. Para tal, uma análise minuciosa permitiu que um conjunto de dados mais “limpo” pudesse ser gerado, através de técnicas como o preenchimento de informações numéricas baseadas em outros valores preenchidos e o descarte consciente de informações não necessárias para a corrente análise.

¹ Disponível em: <https://dados.ufrn.br/dataset/turmas>

Para além dos valores faltantes na tabela, um estudo interno revelou que diversas turmas cadastradas na base de dados foram erroneamente categorizadas como turmas na modalidade de ensino presencial, quando, na verdade, tratavam-se de turmas de ensino à distância (EaD). Alguns dos tipos classificados erroneamente são:

- Aula remota;
- Aula online;
- Remoto;
- Sala virtual;
- Turma virtual;
- Ambiente virtual;
- Aulas remotas via plataforma virtual;
- EAD - Virtual, dentre outros.

Este fator gerou grande preocupação, pois a má classificação destes dados impacta diretamente na razão de confiabilidade dos estudos estatísticos realizados com base neles. Sendo assim, o projeto optou por não dar ênfase em análises que envolviam os tipos das turmas ofertadas.

Uma outra ocorrência corriqueira no tratamento de dados foi a do tratamento de dados dissonantes (*outliers*). Estes são dados que destoam completamente da realidade dos fatos analisados, e que de nada contribuem para a análise. Sendo assim, registros deste tipo foram procurados e removidos.

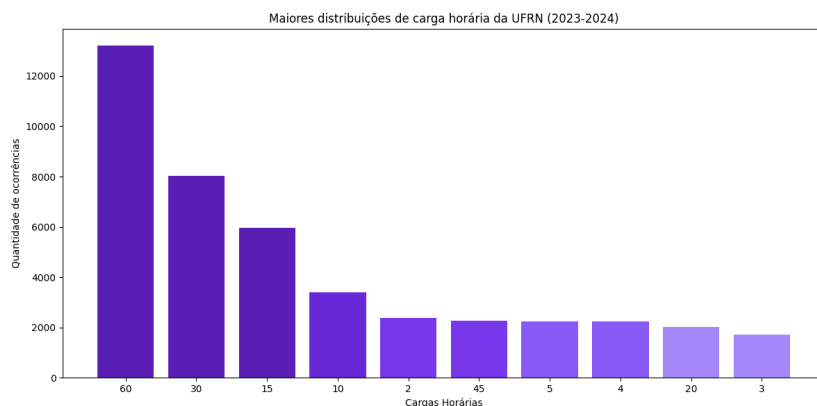
Após o processo de limpeza de dados ser concluído, a etapa de análise começou a ser efetivada. Para tal, foi escolhida a linguagem de programação Python, altamente conhecida por suas poderosas ferramentas de análise de dados. Em adição, as bibliotecas *pandas*, *matplotlib* e *seaborn* foram utilizadas, cada qual desempenhando um importante papel na obtenção dos resultados descritos na próxima seção deste documento. A função de cada uma destas bibliotecas é:

- *pandas*: converter arquivos brutos de dados “.csv” em um formato familiar para a linguagem Python, facilitando a realização de operações, como adição e remoção de informações;
- *matplotlib*: biblioteca que trabalha em cima da biblioteca *pandas*, oferecendo uma série de ferramentas de visualização de dados;
- *seaborn*: outra biblioteca de visualização de dados, construída em cima do *matplotlib*, que permite a criação de diferentes estilos de gráficos, para melhor atender cada uma das análises realizadas.

4. RESULTADOS

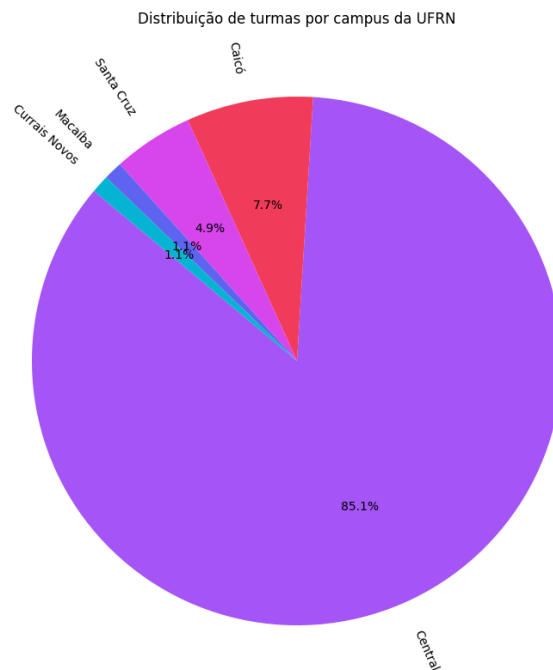
Ao fim da etapa de limpeza, foram construídos gráficos que facilitam a visualização dos dados, ajudando, assim, na tomada de decisões por parte do corpo administrativo da UFRN e abrindo portas para debates diversos.

4.1 Distribuição de carga horária



Com o gráfico acima, é possível constatar que as cargas horárias de 60 horas e 30 horas são as mais frequentes dentro da instituição, com ocorrências que ficam por volta dos 12.000 e 8.000 registros, respectivamente.

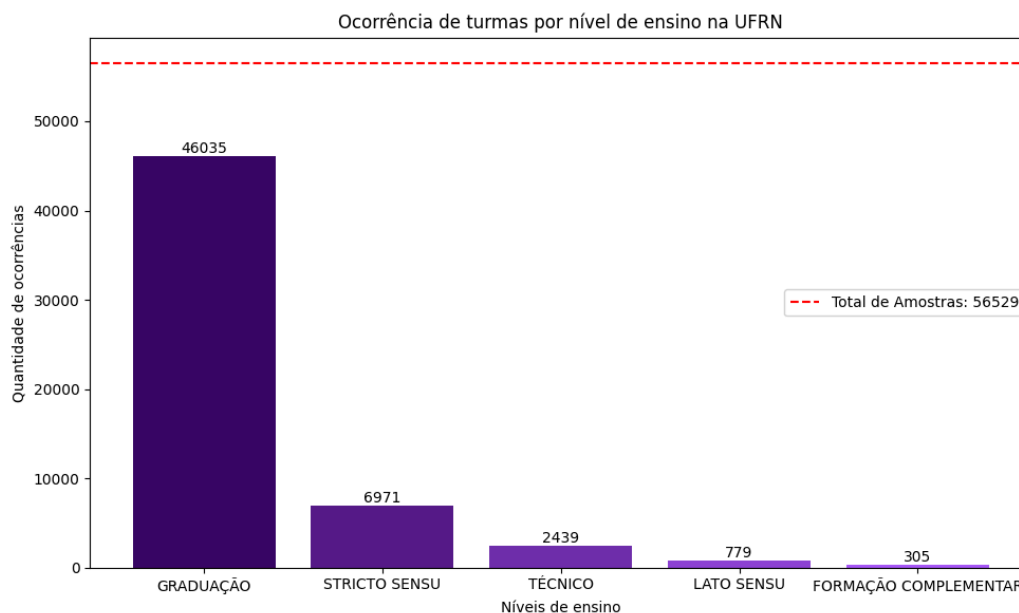
4.2 Distribuição de turmas por campus da UFRN



O gráfico acima exibe a porcentagem de ocorrência das turmas presentes na base de dados da UFRN para cada campus da UFRN. Notavelmente, o campus Central lidera de forma isolada, contendo mais de 85% das turmas cadastradas, seguidas pelo campus Caicó, Santa Cruz e finalmente Macaíba e Currais Novos, que disputam o último lugar com 1,1% cada. Esse gráfico ilustra bem a assimetria que existe na oferta por cursos nas unidades da UFRN presentes no interior do estado e na capital, onde a segunda concentra uma maioria esmagadora dos cursos, o que reflete também na quantidade de turmas ofertadas. É uma reflexão interessante refletir sobre a democratização do acesso ao ensino superior e se ela vem se concretizando de fato, uma vez que estudantes residentes no interior do estado possuem um nítido menor acesso à educação.

Por se tratar de um conjunto limitado de dados, essa análise pode ser ainda mais aprimorada com a utilização de bases de dados de outras instituições de ensino, a exemplo da UFERSA e UERN.

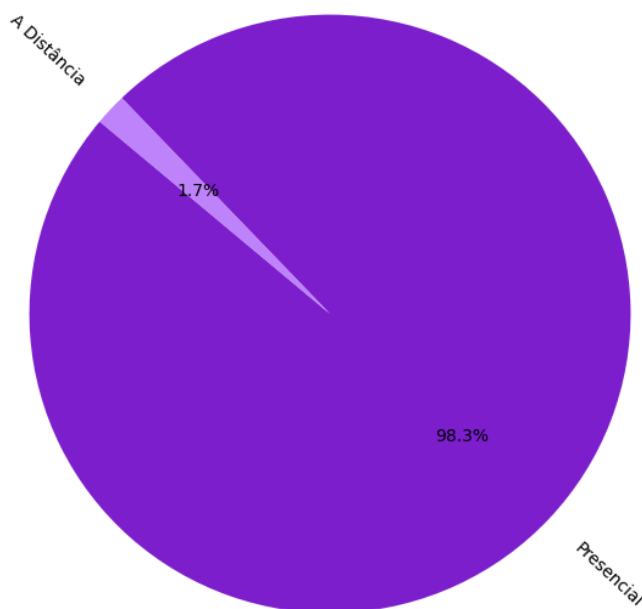
4.3 Distribuição de turmas por nível de ensino na UFRN



O gráfico acima exhibe a quantidade de turmas existentes entre os anos de 2023 e 2024 na UFRN, separados por níveis de ensino. As turmas de graduação lideram com folga a comparação, com 46.035 turmas de um total de 56.529 amostras, sendo seguida pelas turmas de *stricto sensu*, com 6.971 ocorrências, técnico, com 2.439 ocorrências, *lato sensu*, com 779 ocorrências e formação complementar, com 305 ocorrências. É notável que há uma oferta bem maior de turmas de pós-graduação *stricto sensu* do que *lato sensu*, o que pode denotar uma tendência da UFRN como um todo. Um montante maior de dados de outros anos reforçaria ainda mais esta hipótese.

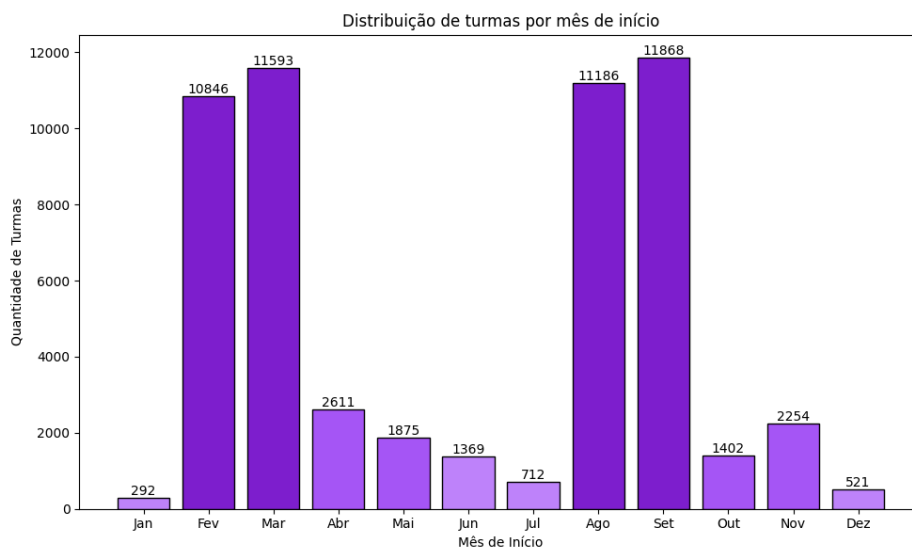
4.4 Distribuição de turmas por nível de ensino na UFRN

Distribuição de turmas por modalidade na UFRN



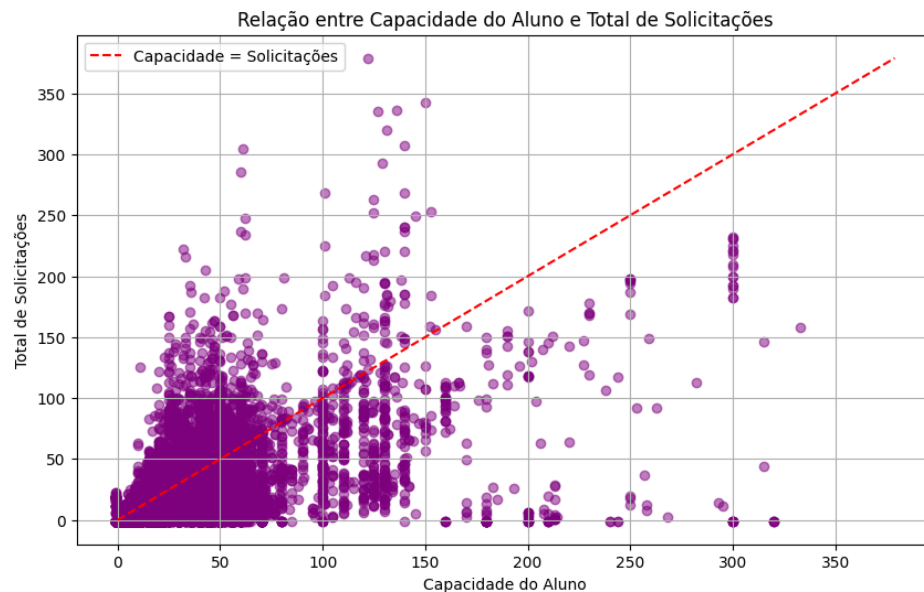
O gráfico abaixo exibe a distribuição das turmas por modalidade de ensino na UFRN. A modalidade presencial toma uma fatia de 98,3%, contra apenas 1,7% das turmas EaD. No entanto, como já explicitado anteriormente, uma investigação interna revelou que uma fração destas turmas categorizadas como presencial deveriam, na verdade, serem contabilizadas como sendo turmas de educação à distância. Sendo assim, a correção desses dados em semestres futuros se faz necessária para que o nível de confiabilidade deste tipo de análise aumente.

4.5 Distribuição de turmas por mês de início



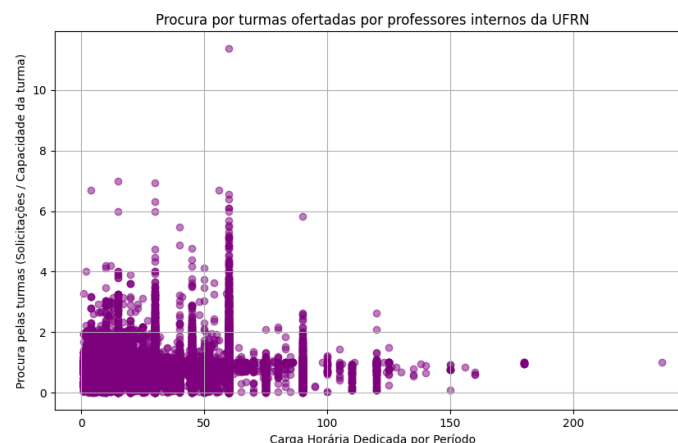
O gráfico acima exibe a distribuição das turmas por mês de início, considerando os registros dos anos de 2023 e 2024. É nítido que há uma maior criação de turmas durante os meses de fevereiro, março, agosto e setembro. Estes meses correspondem às datas normais de início de semestre, sendo os dois primeiros relativos ao primeiro semestre e os dois últimos relativos ao segundo semestre. Sendo assim, conclui-se que a grande maioria das turmas da UFRN são criadas antes ou pouco depois do início do período letivo.

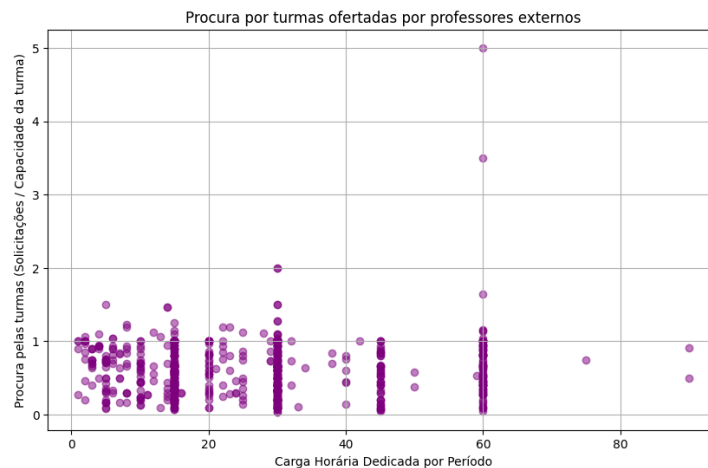
4.6 Relação entre capacidade de alunos de uma turma e total de solicitações



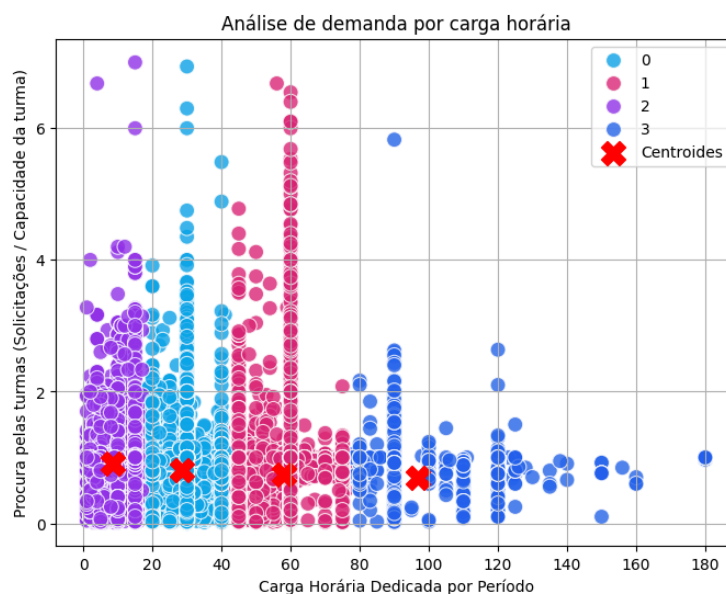
O gráfico acima relaciona as capacidades de alunos de cada turma e o total de solicitações que obteve durante o período de matrícula. É visível, por exemplo, que salas que possuem capacidade para até 150 alunos geralmente possuem um número de solicitações maior do que a quantidade de vagas. No entanto, ao ultrapassar esse limite, a frequência de turmas com mais capacidade do que solicitações é recorrente. Isto pode indicar que estes dados de capacidade também foram gerados de forma incorreta.

4.7 Comparação entre procura por turmas ofertadas por docentes internos e externos





Analisando ambos os gráficos acima, é possível constatar uma grande diferença entre a procura por turmas ofertadas por docentes internos da UFRN (que possuem SIAPE) e a procura por turmas ofertadas por docentes externos (que não possuem SIAPE). Enquanto as turmas dos professores internos têm uma razão de procura concentrada entre 2 e 4, as turmas ofertadas por docentes externos raramente ultrapassam a razão de procura 1. Este fato pode indicar uma certa preferência por docentes internos, possivelmente ocasionada pela familiaridade ou contato frequente dos discentes com certos profissionais da instituição.

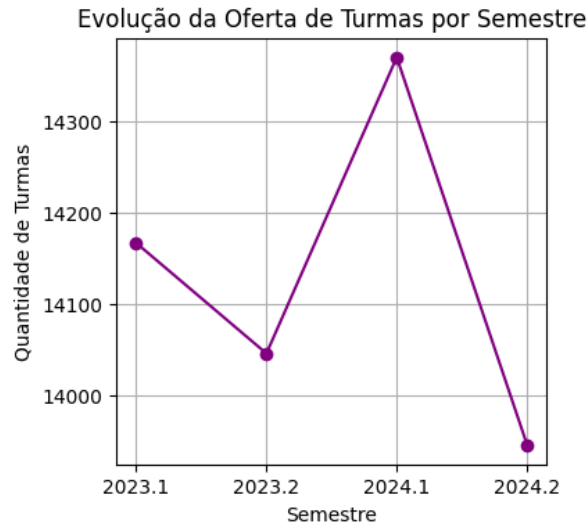


Aplicando um algoritmo estatístico, é possível estimar uma separação de grupos baseado na relação entre carga horária dedicada por período e a procura pelas turmas:

- **Grupo 2 (roxo):** é o grupo das turmas com carga horária máxima de 20 horas que possui uma alta procura;
- **Grupo 0 (ciano):** é o grupo das turmas com carga horária entre 20 e 45 horas que possui uma alta procura, com destaque para as disciplinas de 30 horas;
- **Grupo 1 (rosa):** é o grupo das turmas com carga horária entre 45 e 80 horas que possui uma alta procura, com destaque para as disciplinas de 60 horas;

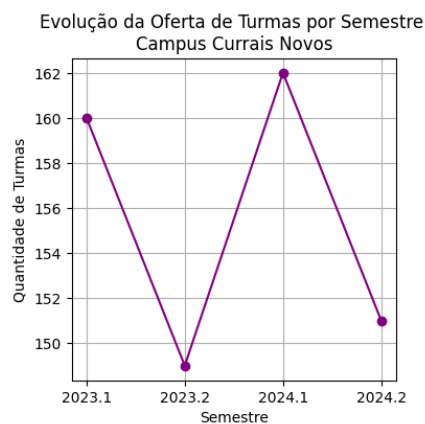
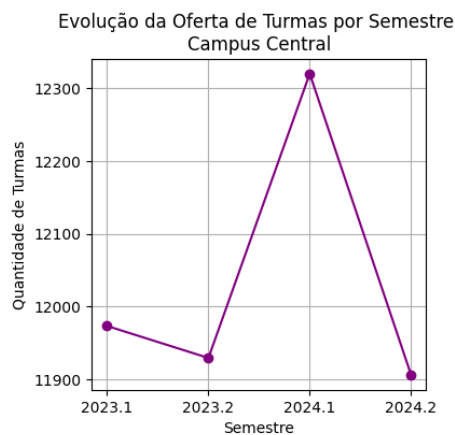
- **Grupo 3 (azul):** é o grupo das turmas com carga horária acima de 80 horas que possui procura moderada para baixa, com um leve destaque para as disciplinas de 90 horas.

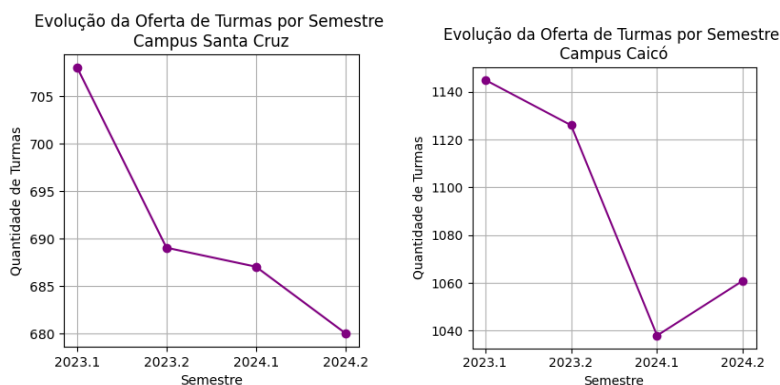
4.8 Evolução da oferta de turmas por semestre



O gráfico acima mostra a oferta de turmas de toda a UFRN nos últimos 4 semestres. Nota-se que há uma oferta maior de turmas nos dois semestres iniciais de cada ano. Um motivo plausível para isso seria o fato de que mais estudantes da UFRN adentram durante o primeiro semestre letivo do ano. Sendo assim, há uma demanda maior por turmas e uma consequente oferta também maior. O segundo fato que pode ser observado é a discrepância entre a quantidade de turmas ofertadas no semestre de 2023.2 e 2024.2. Este fato pode ser um reflexo da paralisação realizada pelos professores da rede federal de ensino, que ocorreu durante o ano de 2024.

Apesar deste gráfico apresentar bons resultados, é natural que a situação de cada campus da UFRN seja diferente. Sendo assim, uma alternativa é separar agrupar essas informações para cada campus, gerando assim um novo gráfico que represente a situação real de cada um.



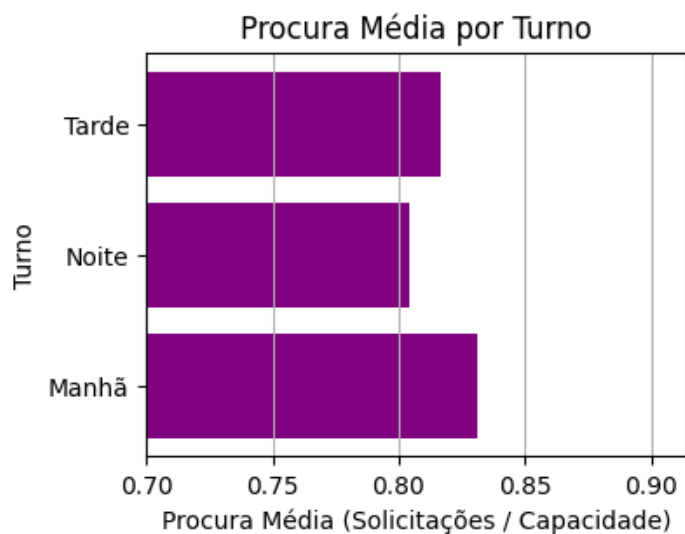


Observando estes novos gráficos, é possível notar que cada campus possui sua própria dinâmica de oferta por turmas em certos semestres. Enquanto a oferta do Campus Central é menor em 2024.2 se comparado a 2023.2, em Currais Novos esses valores não se alteraram muito durante os dois últimos anos, o que pode indicar uma certa estabilidade de oferta, esperada para um campus de menor porte.

Já em Santa Cruz, observa-se uma queda gradual e cronológica da oferta de turmas, o que pode estar indicando uma baixa procura pelos cursos. Esse dado pode ser cruzado com outras bases de dados, como a do INEP, para provar ou refutar essa hipótese através da observação da nota de corte nos cursos deste campus ao longo dos anos. Em Caicó, diferentemente de Natal e Currais Novos, observa-se uma oferta maior de turmas nos segundos semestres de cada ano, em comparação com os primeiros semestres. Isso pode indicar que nesse campus ingressam mais alunos no meio do ano do que no início.

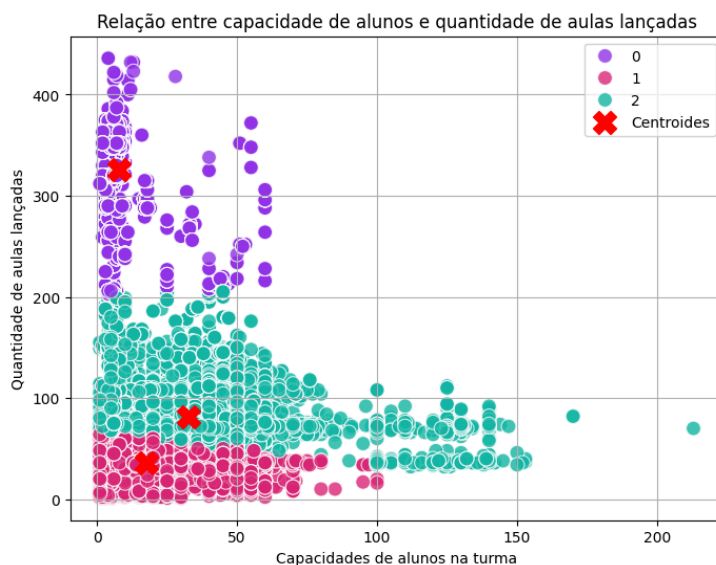
É interessante fazer essa análise separadamente para perceber que cada campus da UFRN possui suas próprias nuances, demandas e necessidades. Observar esses dados de forma conjunta pode dar a falsa impressão de homogeneidade que é majoritariamente causada pelo Campus Central, a quem se deve a maioria dos dados presentes nas bases estudadas.

4.9 Procura média de turmas por turno



O gráfico acima mostra a procura média por turmas em cada um dos turnos de ensino. Apesar de não parecer uma diferença expressiva, é notório que há uma preferência dos alunos por alunas durante o período da manhã e da tarde, sendo o período da noite aquele com a menor procura média. Um teste de hipótese feito utilizando o algoritmo ANOVA confirmou que existe uma diferença estatisticamente significativa na média da quantidade de aulas lançadas por turno, considerando o volume completo de dados.

4.10 Relação entre capacidade de alunos e quantidade de aulas lançadas



O gráfico acima mostra novamente o algoritmo estatístico k-means que realizou uma separação em grupos baseado nas suas diferenças. Analisando essa representação visual, é possível classificar os dados em questão em três grupos:

- **Grupo 2 (verde):** É o cluster que representa as turmas mais volumosas da UFRN, com aulas que provavelmente são lecionadas em auditórios para dar conta da grande demanda de alunos. Sua baixa quantidade de aulas pode indicar se tratar de uma disciplina de curta duração. Caso seja optativa, trata-se de uma disciplina muito visada mesmo por alunos de outros departamentos;
- **Grupo 0 (roxo):** São turmas que possuem uma relativa baixa capacidade de alunos e uma quantidade alta de aulas lançadas. Este é, provavelmente, o grupo no qual estão inseridos os componentes curriculares mais "convencionais", ministrados em salas de aula que raramente ultrapassam 50 alunos;
- **Grupo 1 (rosa):** Este cluster representa turmas de atividades complementares: TCC's, projetos de extensão e outras atividades com menos encontros formais.

Análises desse tipo podem auxiliar a gestão da UFRN a tomarem uma maior decisão sobre quais disciplinas precisam de mais atenção, ajudando, por exemplo, na tomada de decisão para alocar salas de aula grandes que contém poucos alunos e salas pequenas que contém muitos alunos. Apesar disso, a análise feita ainda é bastante inconclusiva, pela

proximidade entre as amostras estudadas. É possível que, com a utilização de um algoritmo mais robusto, os resultados acabem se tornando mais agradáveis.

5. DISCUSSÃO

Tendo em vista as análises feitas e expostas, fica claro que a ciência de dados, unida aos campos da estatística e matemática, podem constituir uma poderosa ferramenta administrativa, facilitando o processo de visualização de dados e compreensão da realidade. O presente estudo apontou algumas possíveis abordagens que podem ser levadas em consideração para melhorar questões como a da assimetria no acesso à educação no interior do estado e a das turmas que não conseguem atender a todos os alunos. Setores administrativos da UFRN podem utilizar deste estudo para fundamentar uma mudança estrutural na universidade, priorizando sempre o bem estar de todos aqueles que compõem seu corpo acadêmico.

6. CONCLUSÃO

Conclui-se que o estudo realizado é satisfatório no que tange a análise das turmas da UFRN. Este trabalho pode ser amplamente aprimorado através da inserção de novas bases de dados advindas da UFRN e de outras instituições do estado. Também pode analisar diversas outras questões relativas às instituições, como notas de alunos, avaliações docentes, infraestrutura dos campus e gestão financeira. As possibilidades trazidas pela ciência de dados são inúmeras, e as melhorias que podem ser realizadas através da sua inserção no meio acadêmico são inúmeras.