Relatório de Ciência de Dados e Inteligência Artificial

Karla Patrícia Oliveira Esquerre

Janaina Souza de Souza

Júlia Carolina Bijos

Sandra Pinheiro

Robson Wilson Silva Pessoa

Bruna Climaco

Luís Gustavo

Laio Luís

Nicole Príncipe

2021-11-29

# Preambulo

# Relatório - Ciência de Dados

## Informações gerais

### Eventos realizados com os jovens cientistas

### Encontro com os novos jovens cientistas

## Introdução

A disciplina de introdução à ciência de dados é constituída de um propósito transformador, aproximar estudantes e escola do universo dos dados de maneira a se apropriar dos principais dados, sistemas de informação e tomada de decisões no contexto local de um grupo de discentes, docentes, comunidade escolar e de bairro.

Coordenada por uma equipe da universidade e operada num contexto à distância, neste relatório são reveladas as principais estratégias e **insights** resultado da promoção do encontro destes vários atores. Esta disciplina não é uma ilha, em paralelo a sua execução são realizadas atividades de protagonismo racial, social e de gênero, práticas investigativas e inteligência artificial. Estes temas tão atuais, tangenciam os propósitos das ciências de dados, funcionando como um mecanismo de cooperação. Uma rede de colaboradores palestrantes também enriqueceram o repertório de todos os atores que participaram do projeto, tornando a execução da discipliana de ciência de dados uma atividade punjante e transformadora do contexto social ao qual pertencem estes jovens cientistas.

Nas seções a seguir são destacadas as principais experiências vivenciadas com os estudantes participantes do projeto Ciência de Dados na Educação Pública no contexto desta disciplina. Foram realizados com os Jovens Cientistas (JC) 11 encontros dedicados apenas para ciência de dados. Parte desses encontros é resultado da proposição de um tema em 2 dias diferentes, sendo o restante constituída apenas de um encontro, contudo guiado por retornos via whatsapp.

## Introdução à Ciência de Dados (Turma 1)

### Relato do encontro

Especialmente para o encontro de [Introdução à Ciência de Dados](https://youtu.be/FsfJ-ERe3rI) realizado em 19/02/2021 foram notadas as seguintes observações:

* Os encontros foram realizados de forma síncrona.
* É importante ressaltar que os jovens cientistas participantes deste encontro ingressaram no projeto Ciências de Dados desde de 2019. Enquanto a equipe de facilitadores é composta de pessoas que ingressaram em 2019 e outra que iniciou em 2020.
* Todos os integrantes são formados em engenharia e um integrante no nível de graduação também em engenharia.
* Antes dos encontros de Ciência de Dados (ECD) os jovens cientistas foram estimulados à leitura prévia do ebook  
  [Uma introdução gentil à Ciência de Dados](https://bookdown.org/cienciadedadosnaep/ciencia_de_dados/).
* Considerando o **background** dos estudantes neste projeto, foi utilizada uma estratégia  
  que para relembrá-los dos principais conceitos entorno da Ciência de Dados, (é recomendável testar este tipo de estratégia com as respectivas ferramentas para que o estudantes possa desenvolver habilidades e vivência no mundo virtual e síncrono). O diálogo no início da aula, o envio de **cards** pelos mecanismos de comunicação podem promover as lembraças dos estudantes. O resultado desta atividade não acaba apenas com o gráfico que contem as várias palavras. É necessário que os estudantes se envolvam na análise desta informação. Dessa maneira, ao serem impelidos a se debruçar sobre este processo, realizam a análise gráfica, uma das etapas desta ciência que seu cerne é o dado. Ao se deparar sobre os dados um dos instrutores ao longo do encontro os provocou a refletir quais das palavras daquela nuvem seria a de mais frequência entre aquelas fornecidas pelos próprios estudantes. Uma das estudantes rapidamente identificou e o instrututor completou a análise, explicando a distribuição e dimensão das palavras para todo o público alcançasse um aproveitamento próximo desta etapa do encontro. Em resumo, os estudante destacaram a própria “Ciência de Dados” como expressão ou palavra da qual mais lembravam, porém, outras palavras interessntes surgiram como “Big Data,” “Insight,” “gráficos,” “medianas” etc. Contudo, a participação no chat do google meet não foi tão expressiva até esta etapa. Mas alguns elementos foram extraídos dessa etapa, como o fato de uma das estudantes fazer o relato sobre estar familiarizada com este tipo de visualização pois assistiu no programa de televisão de “Fátima Bernardes,” evidenciando a percolação do uso de informações por outros mecanismos que não acadêmicos [EXPLORAR].

COLOCAR FIGURA DA NUVEM AQUI

* Antes deste encontro os estudantes tiveram uma palestra na qual viram visualização de dados aplicados a análise dos dados de irradiância solar.

A partir de uma figura do meio ujbano disponibilizada no [ebook Uma introdução gentil à Ciência de Dados](https://bookdown.org/cienciadedadosnaep/ciencia_de_dados/) os estudantes são convidados a refletir sobre diversos problemas do meio urbano, como a “Saúde,” “Meio Ambiente,” “Acessibilidade” e “Transporte” e como a ciência de dados poderia ajudar. Os estudantes indicaram por meio do chat:

* “Eu acho que no geral para fazer pesquisas,”
* “Número de Leitos,”
* “Número de vagas,”
* “Demora do atendimento,”
* “Número de trabalhadores,”
* “Tem mais tecnologias para as pessoas entenderem os dados,”
* “Quantas pessoas estão internadas,”
* “Número de infectados por doenças,”
* “O sistema de marcação de exames,”
* “Quantas pessoas estão nos ônibus,”
* “O negócio que aparece onde o ônibus vai,”
* “Ter mais elevador nos ônibus pra deficiente,”
* “Ponto de ônibus,”
* “Quais são as áreas que tem menos ônibus”
* “Tipo de problema do metrô,”
* “Quantos ônibus precisam ser renovados com ar-condicionado, elevador, cadeiras etc.”

A partir deste novo “brainstorm,” o grupo é convidado a pensar como lidar com as questões formulados no âmbito da ciência de dados.

* “Dados está em todo lugar”

Os três principais pilares são então apresentados, demonstrando a interdisciplinariedade da ciência de dados (Estatística/Matemática, Computação e conhecimento do fenômeno).

Em que situação do dia a dia são encontradas aplicações de Ciência de Dados?

Relatos dos estudantes pelo chat: - Televisão, celular, internet, comerciais, jogos, informações de jornais, recomendações do Instagram, Youtube. - Coletando meus dados.

Explicação sobre sistema de recomendações.

Relato dos estudantes por audio: - Uma das estudantes desconfia que o Youtube deve fazer propagandas e recomendações para ganhar dinheiro, “não perde uma chance de ganhar dinheiro.

* Uma das estudantes diz que o youtube faz recomendações baseado naquilo que já assistimos e demonstra familiaridade ao exemplificar o caso de vídeos de *coaching* quando se pesquisa um tem associado como *empreendedorismo*.

Ao se explorar ideias mais complexas como a operação do sistema de recomendações baseado no seu perfil de redes sociais, independemtemente do aparelho em uso, eles não demosntraram familiaridade. Um possível explicação para este fato é a falta de acesso a mais de um aparelho. Em algumas situações os estudantes não tem o parelho apenas para si.

#### Mecanismo do sistema de recomendações

A instrutora demonstrou a estrutura geral do sistema de recomendações, formada pelas etapas (Usuários, Características dos Vídeos, Pre-processamento dos dados, Análise dos dados, Predição, Recomendações, Usuários).

#### Proposta de atividade em aula

Vamos criar equipes, coletar dados e responder as seguintes perguntas:

1. Qual é o gênero de vídeos favorito da sua equipe?
2. Qual é a característica mais influencia este resultado?
3. Existe um padrão no perfil dos integrantes da equipe?

* Os estudantes levaram cerca de 30min para a realização da atividade

Os instrutores identificaram que os grupos construídos no Whatsapp para a realização da atividade trocaram cerca de 100 mensagem cada.

1. grupo 4) Gênero favorito Música, o motivo é diversão, realizou a atividade chegando ao consenso, avaliação dos fatores que influenciam e inciaram a elaboração de um gráfico de setores para estudar as informações do grupo, porém ela relata que 100% das pessoas preferiram música.
2. grupo 3) Escolheram gameplay, perceberam o uso de vídeos longos 10-20min, identificaram idade entre 13-15 anos, usaram tabela para a representação.
3. grupo 2) Fizeram o descritivo das principais escolhas do grupo, e demonstram conhecimento sobre a lógica da recomendação (de acordo com o que mais assistem). Os estudantes relatam os tipos de música escutada como *Hap*. Dada a escolha distribuída entre as opções de cada um deste grupo, a única variável que os uniu era a idade. A participante do grupo ainda relata familiaridade com dados estatísticos ao dizer que seria difícil compreender o gênero de alguém como variável que definiria na recomendação devido a falta de quantidade de dados.
4. grupo 1) A equipe não podia falar, e identificaram que a eles constumam assistir vídeos com menos de 50min e a curiosidade é o que todos são

A instrutura amarra a discussão sobre as tendências dos gostos em função de algumas variáveis, quantidade de dados, identidicação de padrões por tabelas ou visualização de dados.

* Fazer perguntas
* Reuni informações (Coleta de Dados)
* Analisaram os dados
* Representação ou visualização de dados.

Nesta etapa foi amarrada a discussão do Ciclo de Dados da Ciência de Dados e novamente os estudante sãao estimulados a preencher uma nuvem de palavras com os conceitos deste encontro.

Agora, nesta etapa a nova nuvem de palavras revelou outro nível de informações. Os estudantes ficaram focados em “Coletada de Dados,” “Amostra,” “Pesquisa Trabalho em Equipe,” ““youtube’. Curiosamente, alguns estudantes relataram a palavra”protagonismos”.

### Proposta de atividade e reencontro

A construção de um formulário foi realizada para completar o ciclo e consolitar o ciclo do entendimento da Ciência de Dados. Um segundo encontro foi utilizado para debater as atividades além da proposição das atividades. Conceitos essenciais como ciclo de dados, coleta de dados, *insights* etc.

Neste [encontro 05/03/2021](https://youtu.be/v--UCfc3OQQ) desenvolvemos um sistemas de grupos que eram responsáveis pela elaboração de 3 perguntas baseadas na leitura de ebook, e em seguida seria realizado a proposta da pergunta para outra equipe responder. Estimulando tanto a leitura quanto a reflexão sobre a definição da Ciência de Dados.

Houve uma boa adesão a esta atividade. Tanto por parte da elaboração das perguntas quanto aqueles que decidiram respondê-las.

Os colegas interagiram, tiveram dúvidas e espontaneamente  
cooperaram uns com os outros. Explicando os mecanismos de recomendações e preferências, relataram novamente experiências com planbos de celulares, o que assistiram em programas de canais de televisão a respeito sobre o sistema social da internet. Eles estavam atentos sobre a assinatura dos termos e condições utilizados na entrada de redes sociais e sites. Quando questionados quais setores da sociedade pode ser aplicado a Ciência de Dados eles souberam explicar com bastante segurança, qualquer área que envolvam dados.

## Encontro sobre Coleta de Dados (Turma 1)

### Relato do encontro

Foi realizada uma apresentação inicial  
que discutiu uma das etapas ciclo de dados, a coleta de dados. A entrata neste tema exigiu a exploração dos conceitos básicos de estruturação de dados. Portanto, foi explorado com os estudantes as diferentes formas como os dados são  
encontrados, como estruturados e não estruturados. Motivados pela necessidade de maior atuação dos estudantes, para que as perguntas e formulações partissem dos estudantes a estrutura dos encontros passou para um modelo de dinâmicas.

Esta etapa envolvia a coleta de dados conduzidas por uma estrutura prévia que estavam pautadas no tema “Família.” **[Família],[Raça/Cor],[Quantidade de filhos],Quantidade de animais de estimação]**

Nesta atividade os estudantes de cada grupo coletaram os dados e em cada linha preencheram as informações sobre as suas respectivas famílias. Algumas dúvidas interessantes surgiram, como o caso de famílias que além dos pais, tem avós ou tios. Outra informação que os estudantes tiveram dificuldade foi a escolha ou definição de raça/cor, pois não conseguiam reduzir esta informação a uma das categorias conhecidas e difundidas pelo IBGE. Na opinião de uma das estudantes havia a compreensão que sua família é mista. Esta capacidade de reflexão surge com a proposta de atividades dinâmicas como esta. Por outro lado, para trabalhar habilidades específicas é necessário aperfeiçoar esta atividade para os estudantes possam alcançar melhor desenvolvimento de atividades e compor suas competências.

Cada grupo apresentou seus resultados, sendo que os instruturores direcionavam perguntas em relação às tabelas alcançadas pelos estudantes. Um dos principais debates surgiu entorno da necessidade de verificar a consistências dos dados por meio do cruzamento de variáveis e observações. Discutiu-se sobre variáveis numéricas e categóricas. Também foi observado um outro fenômeno, a duplicidade das linhas, demonstrando mais um possível problema tipicamente encontrado nos bancos de dados.

Em seguida, foi realizado um fechamento dos conceitos tabelas, variáveis, observações, base de dados etc.

### Proposta de atividades e reencontros

Ao final do encontro, foi proposta uma atividade na qual os estudantes teriam que escolher um candidato para vaga de emprego de motorista. Para isto deveriam definir variáveis na seleção e organizar uma base de dados para auxiliar na decisão, explicando suas decisões. Sendo importante que a principal referência seria o uso do ebook de ciência de dados.

## Reencontro sobre base dos dados e encontro de visualização de dados (Turma 1)

[encontro 30/03/2021](https://youtu.be/tT6_xB47Hssv)

### Relato do reencontro

Um dos grupos dedicaram a atividade de contratação de um engenheiro ambiental. O grupo debateu sobre a seleção de variáveis a partir de um conjunto previamente disponibilizado na aula anterior. Discutiu-se a conexão entre as variáveis e a pergunta realizada na solução de contratação de um profissional. Assim, foi discutido o encadeamento da metodologia utilizada na ciência de dados. Em outra equipe foi explorado o excesso de variáveis como fator de confusão da análise conduzindo os estudantes a reflexão sobre a necessidade de planejar uma coleta de dados de acordo com a pergunta que pretende-se responder. Além disso, é importante notar que não é ética a atitude de coleta de dados em entrevistas de emprego que não sirvam para este fim. Também foram debatidos temas transversais às ciências de dados como elementos da candidatura envolverem preconceitos e como algumas variáveis podem conduzir para este viés. Também foi realizada um debate a respeito do aproveito de bancos de dados prévios e para quais fins ele foi construído.

### Encontro de visualização de dados

O encontro de visualização de dados utilizou o tema de educação, especificamente dedicado aos dados do ENEM.

Inicialmente foi relembrado aos estudantes que no ano anterior havia uma das disciplinas do projeto que envolvia apenas a visualização de dados.

Uma das instrutoras usou recursos de handwriting em um tablet para a construção de gráficos de barras. Primeiro demonstrou os valores na tabela da evolução do número de inscritos no ENEM. Os estudantes opinaram sobre o comportmento dos dados, o que os induziu a construção de um gráfico para explorar de forma mais direta o comportamento geral das informações.

Foram debatidos alguns conceitos básicos como eixos,representação da variável, valores de referência, legenda, título de gráficos etc.

Na etapa de análise foi explorado o comportamento do número de inscritos, exatamente oposto ao senso comum, segundo a resposta dos próprios estudantes.

Em seguida, a discussão é conduzido aos problemas de pessoas que tem dificuldade de acesso a computadores e as implicações na realização da prova do ENEM. Nesta etapa foi explorado de forma detalhada a construção de um gráfico de setores. Que envolveu o cálculo de frequência, percentual e a ideia de proporcionalidade da área do setor como representação da quantidade de uma determinada categoria.

Neste encontro a participação dos estudantes foi menos pronunciada,tanto verbalmente quanto pela escrita no chat. Este comportamento deve ser investigado, sendo que o modelo expositivo da aula ou pela dificuldade nas habilidades com cálculos com frações ou percentuais. Poucos estudantes demonstraram conhecimento prévio sobre este conteúdo.

Um dos instrutores esclarece que a motivação do uso de um gráfico ou outro depende da informação que se pretende comunicar. Também são exploradas as fragilidades dos gráficos.

## Reencontro de visualização de dados (Turma 1)

Neste [encontro](https://youtu.be/pbioaYwR89U) os conteúdos da aula anterior foram revisitados. Porém, devido as dificuldades encontradas pelos estudantes no cálculo de frações foi utilizado um recurso de lousa matemática [Mathigon](pt.mathigon.org) que contém elementos geométricos, recursos que simulam realidades físicas como balanças, dados ou sistemas de cartas.

A apresentação do recurso digital prendeu um pouco mais a atenção dos estudantes além de melhorar a participação e compreesão sobre as ideias de frações, fundamentais para aprofundar na análise de gráficos de setores. Uma das estudantes compreendeu detalhadamente o conceito de frações e sua ligação com o gráfico de setores, verbalizando uma ideia da visualização a qual foi estimulado a refletir.

Algumas ressalvas foram realizadas sobre aspectos de visualização de dados, como a proporcionalidade, aspectos técnicos dos eixos, percentuais de áreas ocupadas, legenda e títulos.

### Proposta de atividades e reencontros

Numa atividades proposta através de um formulário web os estudantes foram estimulados a responder informações sobre frações, gráficos de barras e gráficos de setores. Além disso, foi uma nova oportunidade para que os estudantes pudessem familiarizar com informações transversais ao tema do ENEM, como a compreensão a respeito sobre os recortes de raça/cor adotadas pelo IBGE.

## Capoeira, esporte e liberdade. Correlação e causalidade (Turma 1)

O [encontro](https://youtu.be/ja7g6Z88E74) discutiu os conceitos de correlação e causalidade suportado no debate de sobre a expressão cultural local. Inicialmente foi proposta uma pequena dinâmica que envolvia a produção de uma nuvem de palavras estimulados pela experiências que os estudantes tinham a respeito da capoeira. A palavra que mais se destacou foi a “Dança,” mas os estudantes lembraram da capoeira através de vários vocábulos como:

* A arte da dança e da luta;
* Arte;
* História;
* Autodefesa;
* Cultura;
* Exercício Físico;
* Au;
* Escravos;
* Berimbau e
* África.

Na sequência, com o propósito de lembrá-los dos movimentos da capoeira foram apresentados vídeos com animações que representam a capoeira, sendo um deles construído com figuras baseada em carvão e um segundo que explora as relações de som, roda, instrumentos, movimento e a geometria com ideia de roda. Nesta etapa a participação dos estudantes se tornou mais intensa, devido a aproximação com a capoeira conhecendo desde os movimentos da capoeira e detalhes especiais como o nome de instrumentos como atabaque, berimbau etc.

Numa etapa seguinte, os estudantes foram convidados à leitura de trechos do livro “José Moçambique e a Capoeira por Thereza de Almeida (Autor), Joaquim de Almeida (Autor), Laurabeatriz (Ilustrador) - 2007.” A leitura foi realizada por três estudantes, eram trechos que relembraram as “Instrumentos e música,” “Movimento e cooperação” e “Dinâmica, continuidade e multifuncioalidade.”

Em seguida foi demonstrado a importância da capoeira e seu destaque como patrimônio histórico cultural e as conexões da com outros patrimônios reconhecidos pela UNESCO.

A discussão é guiada inicialmente entorno da relação causal, a estrutura do berimbau e seu uso são explorados. A relação do material, da corda, a ação de tocá-la e a propagação gerada pela cabaça são o fio de construção para o conceito de causalidade. O texto é usado como base para ajudar na diferenciação entre os conceitos de causalidade e correlação. Os estudantes se sentiram confortáveis para opinar devido a reflexão baseados na vivência e experiência cultural ofericida pela cidade. Esses elementos facilitaram a entrada nas estruturas formais da matriz de causalidade. Baseado no jogo da capoeira foi construída uma matriz constituída de variáveis como movimento dos jogadores 1 e 2, movimento dos outros participantes da roda, instrumento berimbau, música. Foi definido uma legenda de cores para estabelecer três níveis de correlação, baixo (vermelho) , médio (amarelo) e alto (verde). Houve bastante cooperação nesta etapa, os estudntes transportaram as ideias discutidas nos elementos objetivos da representação da matriz de correlação.

Um segundo problema foi apresentada, baseado no futebol, que também está presente no cotidiano da maioria dos estudantes. Foi construído um [dashboard com dados de futebol](https://cienciadedadosnaep.github.io/dbesporte/), no qual foram disponibilizados gráficos de dispersão com dados de faltas, gols, posição dos jogadores etc. Esta estratégia trouxe o estudante do campo das reflexões, do debate baseado na fenomenalogia do problema (caso da capoeira) para a discussão a partir dos dados (dados de futebol). Adicionalmente, foi ampliada a discusão a respeito de correlação positiva, nula e negativa.

### Proposta de atividades e reencontros

Ao final do encontro, foi proposta uma atividade que envolvia a medição do comprimento das pernas, a altura e abertura da perna num movimento de ginga da capoeira. Assim, o estudante é estimulado a produção de suas próprias medidas, seguida da organização na forma de tabela, suportados nas experiências previamente desenvolvidas em outros encontros. Especialmente nesta atividade foram criados grupos de whatsapp para acompanhamento próximo do desenvolvimento dos estudantes, muitas dúvidas surgiram, sendo uma grande oportunidade para  
oirientá-los de acordo com suas respectivas dificuldades.

## Média e Mediana e Segurança Pública (Turma 1)

Inicialmente foi realizada uma breve revisão em relação ao último encontro, na qual os dados medidos pelos estudantes foram reunidos e disponibilizados na forma de dashboard. Alguns comentários sobre as formas como realizadas as medidas e as consequências sobre a análise final. Houve uma sequência na qual os dados originais dos estudantes foram apresentados como banco de dados, explorando mais uma vez as informações e os conceitos de observação, variáveis categoricas, variáveis numéricas e contínuas. Foi usado a estratégia de visualização de dados destas medidas para auxiliar no debate com os estudantes. Estimulando a análise crítica dos dados. Consolidando a abordagem do ciclo de dados desenvolvido no primeiro semestre. A partir da visualização dos dados os estudantes reconheceram os próprios dados e foi possível a identificação das divergências de medição.

O tema de segurança pública tão caro aos estudantes do projeto, foi apresentado numa das suas linhas de recorte racial, especialmente por meio de dados de ocorrências de crimes contra a mulher.

A estratégia utilizada se constituiu num vídeo de 3 min que explicava brevemente os principais indicadores de segurança pública utilizada pela Secretaria de Segurança Pública da Bahia (SSP-BA) e a possibilidade de avaliá-los por meio de média e mediana.

Na sequência, utilizando os dados da Rede Observatório de Segurança Pública sobre o número de mulheres vítimas de violência sexual na Bahia. Foi apresentado dois comportamentos destes dados entre 2009 até 2017. Alguns aspectos do contexto foram apontadas como a lei nacional no enfrentamento dessa violência é a Lei nº 11.340/2006 (Lei Maria da Penha) e Lei do Feminicídio (nº 13.104/2015) e trazidos para a compreensão dos dados e a ligação com os casos de subnotificação. O aprofundamento da questão se deu ao enxergar os dados desagregados por raça/cor.

A evolução da análise avançou pela aplicação de média e mediana, apontando as principais vantagens e desvantagens em relação aos dados. Foi proposto ao longo do encontro que os participantes escolhessem uma das categorias (Ignorados, Negras, Não Negras e Total) e fizem o cálculo da mediana. Esta atividade gerou um bom envolvimento dos estudantes, que além de realizar os cálculos souberam julgar a escolha entre média e mediana. Na sequência, os estudantes foram apresentados a mais parâmetros para a descrição dos dados. Com foco na amplitude, quartis e sua representação pelo boxplot.

Para atividade assíncrona foi proposto que os estudantes escolhessem um dos indicadores de segurança pública da SSP-BA e selecionassem apenas os valores mensais de um ano específico e calculassem a mediana, para fortalecer a habilidade de cálculo de indicadores.

## Pesquisa Amostral - Transporte (Turma 1)

Quanto ao tema de pesquisa amostral foi utilizado um texto com o título: **Em Salvador, mulheres não se sentem seguras para circular pela cidade**. Previamente, os estudantes foram questionados a respeito da segurança de andar em ônibus em Salvador. Algumas estudantes relataram a experiência de serem importunadas pelo olhar dentro do ônibus (ou no ponto) ou de outras formas.

Como a entrevista foi construída baseada em Pesquisa Amostral e era acomapanhada de excelente qualidade da visualização dos dados. Eles foram relembrados como a pesquisa amostral faz parte do ciclo dos dados apresentado no primeiro encontro.

Dessa forma, foram estimulados a pensar sobre quais foram as estratégias utilizadas para lacançar um bom resultado dessa pesquisa. Gradativamente, ao longo do rico debate com os estudantes, sendo apresenta todas as etapas de planejamento e execução da pesquisa. Os estudantes foram bem assertivos na compreensão dos viéses (de seleção, medição, questão e de não respondentes) gerados pela mã conduta da pesquisa amostral. Também foi discutido a importância do tamanho da amostra e a forma como a pesquisa foi realizada. Mostramos como dados de instituições públicas foram utilizados para ajudar na contextualização.

Os dados dos resultados da pesquisa são apresentados  
por meio de gráfico de setores e barras dentro de infográficos. Esta forma de apresentaçãoi traz um benefício ao estudante na compreensão do dado contextualizado, suportado num discussão. À medida que este debate avançou, algumas questões tornaram-se emergentes e os estudantes reconheceram a possibilidade de serem vítimas de situações presentes no texto.

As fragilidades da pesquisa também foram exploradas ao longo da discussão, estimulando ao grupo a reflexão sobre a qualidade do dado na formação da opinião. Seguindo por uma breve apresentação das políticas públicas para auxiliar na solução destes problemas.

Na conclusão do encontro foi proposta uma atividade que ajudava na leitura dirigida da reportagem que traz o foco paras as questões da pesquisa amostral.

## Saúde e Probabilidade - Pesquisa Nacional de Saúde do Escolar (PeNSE) (Turma 1)

Dado pelos encontros anteriores, os estudantes foram apresentados aos seguintes temas:

* Ciclo dados;
* Coleta de Dados;
* Visualização de Dados;
* Indicadores de tendência central etc.

O tema de saúde deu suporte para introdução dos estudantes aos conceitos de probabilidade. Particularmente, foi adotado o tema de saúde próximo da realidade vivida pelos estudantes. Utilizamos os dados da Pesquisa Nacional da Saúde do Escolar (PeNSE) como estratégia. As perguntas do PeNSE são adequadamente planejadas para o público dos anos finais do ensino fundamental e ensino médio tornou efetivo o desenvolvimento das atividades do projeto.

O vídeo de promoção da pesquisa foi proposto aos estudantes para uma reflexão prévia. Os dados apresentados chamaram atenção dos estudantes gerando comentários a respeito de drogas, sedentarismo e hábitos alimentares. Entretanto, não foram suficientes para alcançar a noção completa de probabilidade. Assim, foi utilizado uma estratégia de construção dos conceitos como experimento aleatório, evento, espaço amostral, possibilidade e probabilidade por meio de uma estratégia tradicional baseada em elementos que os estudantes tem mais controle como dados e moedas.

Apoiado nesta discussão, avançamos na análise dos dados do PeNSE. Primeiro, os estudantes tiveram que responder um formulário elaborado a partir das perguntas do próprio PeNSE, o que permitiu avaliar a consolidação dos conceitos apresentados neste encontro.

Os estudantes também foram estimulados durante o encontro a leitura de dois trechos de teses que se pautaram na PeNSE. Nesta etapa do curso os estudantes tem mostrado elementos de argumentação baseado nas informações precedentes do curso. Também faz parte do vocabulário dos estudantes os outros temas debatidos em outros ambientes do projeto, especialmente aqueles ligados aos temas de protagonismo racial, social e de gênero.

## Introdução à Ciência de Dados (Turma 2)

A Turma 2, que iniciou as atividades no segundo semestre de 2021 teve um número de encontros bem menos que a Turma 1. Diferente da Turma 1, que teve a experiência com muitos encontros com a equipe desde 2020 e em alguns casos que tinham experiências presenciais em 2019.

### Relato do encontro

A proposta do encontro de introdução à ciência de dados é a mesma realizada na Turma 1. À respeito da nuvem de palavras proposta houve um resultado um pouco diferente, particularmente pela confusão entre ciência de dados e inteligência artificial.

O debate sobre decisões nos ambientes urbanos foi muito rico, houve cooperação dos estudantes com suas experiências, como a exemplificação dos problemas enfrentado ao ir para a escola, como eles usariam as informações de um celular para apoiá-los em decisões em problemas reais.

A estratégia de demonstrar a previsão do tempo coloca um contexto no qual as informações são resultado de um grande processo de análise de dados, modelagem, apresentação e tomada de decisões. Problemas emergentes como o da pandemia, trazidos pelos estudantes deram mais elementos para o debate e ajudou na definição das ciências de dados.

A qualidade da participação dos estudantes foi em média muito diferenciada, sendo necessário e uma oportunidade de pesquisa as diferenças entre os diferentes grupos  
de participantes de projetos similares a este.

Com muita facilidade foi possível conectar o sistema de recomendações com o ciclo dos dados da ciência de dados.

Portanto um excelente nível de aproveitamento de todas as partes.

## Tabela resumo de habilidades dos encontros

Descrição das atividades realizadas pelo equipe de Ciência de Dados.

Encontros

Tema

Cód. Habilidade

Descrição

Introdução à Ciência de dados

Sistemas de recomendações

EMIFCG01

Compreensão do funcionamento das tecnologias de digitais baseada em ciências de dados.

Coleta de dados

Programas de assistência social e distribuição de renda

EMIFMAT07 e EMIFMAT08

Compreensão dos sistemas de coleta de dados e informações ligados aos principais programas de assistência social e redistribuição de renda.

Visualização de dados

Exame Nacional do Ensino Médio

EMIFMAT07

Identificação de fatores correlacionados ao desempenho no ENEM por meio de visualização de dados.

Correlação e Causalidade

Capoeira e dados de futebol

EMIFMAT05 e EMIFMAT06

Desenvolvimento de pensamento matemático e estatístico por meio do diálogo sobre manifestações culturais como capoeira. E de conceitos de causalidade e correlação por meio de análise de dados de futebol

Indicadores de tendência central - Média e Mediana

Indicadores criminais

EMIFCG07

Análise crítica de indicadores criminais suportado em visualização de dados e desenvolvimento da capacidade de síntese por meio de estatísticas.

Pesquisa Amostral

Sistema de transporte público e violência contra mulher

EMIFCG02 e EMIFCNT02

Desenvolvimento de habilidades ligadas a formulação de hipóteses, planejamento e execução de pesquisas amostrais e compreensão da mobilidade e segurança da mulher.

Introdução à probabilidade

Pesquisa Nacional da Saúde do Escolar (PeNSE)

EMIFMAT03 e EMIFMAT06

Representação e aquisição de vocabulários ligados a probabilidade promovendo a habilidade de transposição de conhecimento para diversas áreas. Debate ligado a saúde do estudantes por meio da pesquisa PeNSE IBGE.

## Dicussão e conclusão

O principal estrato do material dos relatos dos encontros de ciências de dados constitui-se dos pilares sobre os quais este projeto se baseou para sua realização. Os estudantes buscam a todo momento caminhos para dialogar sobre suas experiências, sendo elas as mais profundas quanto a desigualdade social. Muito embora isto seja verdade, esse também é o elemento que os posicionam num local de autoridade para tratar destes problemas. Incluímos na nossa pequena trajetória com estes incríveis jovens, informações a respeito de sistemas de redistribuição de renda e auxílio social, violência contra mulher no Estado da Bahia e a perda de mobilidade urbana em Salvador e indicadores de segurança com recortes de gênero e raça/cor. Em resumo, eles reconhecem estes elementos pelo lado em que os dados não revelam com tanto clareza, trazendo uma forte base argumentativa. Por outra lado, a Ciência de Dados dá um caminho diferente, para acessar outras visões sobre estes assuntos. Esta ponte é ainda mais reforçada quando trazemos para os encontros as experiências culturais que são ainda mais íntimas a estes estudantes, como a capoeira ou o debate sobre o que são os seus alimentos, seus hábitos e sua saúde. Este entralace entre desigualdade social, escancaradamente revelado nas justificativas de faltas dos estudantes participantes do projeto,  
emergem aquilo que é de conhecimento tácito, a maioria dos jovens de Salvador não estão incluídos de forma qualificada numa sociedade digitalizada. Também foi alvo de observação do projeto a necessidade que os estudantes tiveram de gerenciar o tempo devido ao retorno as aulas online e depois presencial. Alguns relataram como o volume de atividades começaram a intereferir na sua ampla capacidade de participar de todas as atividades. Os avanços observados na argumentação de alguns estudantes ao longo dos encontros, revelados na frequente utilização de vocábulos da área, contrasta com o que tipicamente se observa nos principais indicadores da educação brasileira a respeito dos avanços em disciplinas ligadas a área de exatas. Embora dados recentes revelem o crescimento do IDEB,  
por exemplo, os indices das escolas participantes estão aquém do que se espera para que o jovem alcance os elementos idealizados como uma boa educação.  
Por todo o contexto complexo da realização das atividades durante a pandemia, aliado às dificuldades relatadas pelos estudantes, a frequente participação e a exitosa realização das avaliações do projeto, demonstram que é possível alcançar resultados neste tecido de interface entre tantos atores. Num momento da história que a tecnologia digital tem que ser planejado para evitar mais um braço de desigualdade atividades como esta são exemplos de atividades que devem ser ampliadas para um número maior de estudantes.