**Universidade Estácio de Sá - UNESA**

**Nova América**

**Projeto de Extensão de “Tópicos de Big Data em Python”**

**Nome do(s) discente(s) integrantes do grupo**

**João Victor Ferreira Gonçalves -** [**https://github.com/Haisuu/Trabalho-De-Big-Data**](https://github.com/Haisuu/Trabalho-De-Big-Data)

**Fernando Ribeiro de Almeida Junior -** [**https://github.com/NartZx/Trabalho-De-Big-Data**](https://github.com/NartZx/Trabalho-De-Big-Data)

**Rosemberg Jesus de Souza -** [**https://github.com/Maomez/Trabalho-De-Big-Data-Maome**](https://github.com/Maomez/Trabalho-De-Big-Data-Maome)

**Lucas Silva de Oliveira -** [**https://github.com/Lcs04/Trabalho-De-Big-Data**](https://github.com/Lcs04/Trabalho-De-Big-Data)

**Nome do(a) professor(a) orientador**

**Lucas Floriano**

**2024**

**Rio de Janeiro/RJ**

Sumário

[1. DIAGNÓSTICO E TEORIZAÇÃO 3](#_Toc119686561)

[1.1. Identificação das partes interessadas e parceiros 3](#_Toc119686562)

[1.2. Problemática e/ou problemas identificados 3](#_Toc119686563)

[1.3. Justificativa 3](#_Toc119686564)

[1.4. Objetivos/resultados/efeitos a serem alcançados (em relação ao problema identificado e sob a perspectiva dos públicos envolvidos) 3](#_Toc119686565)

[1.5. Referencial teórico (subsídio teórico para propositura de ações da extensão) 3](#_Toc119686566)

[2. PLANEJAMENTO E DESENVOLVIMENTO DO PROJETO 4](#_Toc119686567)

[2.1. Plano de trabalho (usando ferramenta acordada com o docente) 4](#_Toc119686568)

[2.2. Descrição da forma de envolvimento do público participante na formulação do projeto, seu desenvolvimento e avaliação, bem como as estratégias pelo grupo para mobilizá-los. 4](#_Toc119686569)

[2.3. Grupo de trabalho (descrição da responsabilidade de cada membro) 4](#_Toc119686570)

[2.4. Metas, critérios ou indicadores de avaliação do projeto 4](#_Toc119686571)

[2.5. Recursos previstos 5](#_Toc119686572)

[2.6. Detalhamento técnico do projeto 5](#_Toc119686573)

[3. ENCERRAMENTO DO PROJETO 5](#_Toc119686574)

[3.1. Relatório Coletivo (podendo ser oral e escrita ou apenas escrita) 5](#_Toc119686575)

[3.2. Avaliação de reação da parte interessada 5](#_Toc119686576)

[3.3. Relato de Experiência Individual 5](#_Toc119686577)

[3.1. CONTEXTUALIZAÇÃO 5](#_Toc119686578)

[3.2. METODOLOGIA 6](#_Toc119686579)

[3.3. RESULTADOS E DISCUSSÃO: 6](#_Toc119686580)

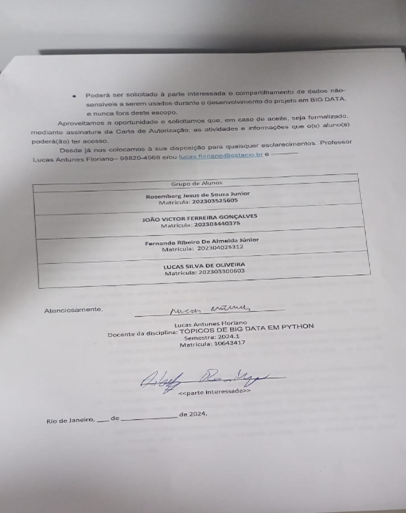
[3.4. REFLEXÃO APROFUNDADA 6](#_Toc119686581)

[3.5. CONSIDERAÇÕES FINAIS 6](#_Toc119686582)

# DIAGNÓSTICO E TEORIZAÇÃO

## Identificação das partes interessadas e parceiros

* O proprietário da Loja se chama Adolfo Rodrigues Ferreira e atualmente não tem uma equipe de desenvolvimento e Design, ele mesmo que se encontra nessa posição de desenvolver e projetar a plataforma e estrutura da loja online. Na parte de conteúdo e marketing ele mesmo que promover e cria conteúdos para sua empresa. Seus parceiros comerciais que fornecem os produtos vendidos na loja online são Alpoim, Rufer e a total distribuidora. O suporte e assistência aos clientes antes, durante e após a compra é o Adolfo Rodrigues Ferreira. Sua equipe estimada de participantes são um motorista, um ajudante e um vendedor.



## Problemática e/ou problemas identificados

* **Baixa disponibilidade e controle de dados relevantes:** Na análise inicial feita no projeto de extensão da disciplina de Tópicos de Big Data em Python, identificamos uma escassez significante de dados relevantes e precisos para o engrandecimento e desenvolvimento de nossas pesquisas com fins acadêmicas. Contudo, essa ausência de volumes estáveis de dados não-sensíveis representa para nós uma série de desafios que impactam de forma negativa a capacidade da empresa de se tomar decisões informativas de considerável proeminência para a mesma, necessitando de uma evolução criativa da parte dos integrantes do projeto, a fim de aproveitar com exímio êxito os dados adquiridos, de forma na qual a qualidade e capricho em cima da “data” seja comparável à uma quantidade considerável de dados de uma empresa de maior porte. Além de, por fim, esta problemática empecer a geração de análises corporativas que visam de maneira principal aumentar de forma significante o lucro total da empresa em seus faturamentos financeiros e sua produtividade em geral.

## Justificativa

* O Problema Citado é um fato real que não pode ser solucionado facilmente devido a falta da baixa disponibilidade de dados e com isso precisaremos nos aprimorar.
* Esse problema pode ser solucionado com diversos conhecimentos que adquirimos durante o curso, e nosso maior foco é em aprimorar a empresa com o intuito de faze-la prosperar.

* E com essas pautas em mente concluímos que big data é a melhor escolha, devido a melhoria da eficiência operacional, otimização dos processos, a análise de grandes volumes de dados financeiros e transacionais propicia a rápida identificação de padrões, oportunidades e riscos para diversos segmentos.

## Objetivos/resultados/efeitos a serem alcançados (em relação ao problema identificado e sob a perspectiva dos públicos envolvidos)

* **Análise de Dados em Tempo Real**: A análise de dados em tempo real é um processo que irá envolver a coleta, processamento e a interpretação de dados enquanto eles estão sendo gerados ou disponibilizados, permitindo uma resposta imediata às informações obtidas. Contudo, essa abordagem se difere com a análise de dados convencional, que geralmente envolve a coleta de dados em lotes e sua futura análise em um momento posterior.
* **Gerenciamento de dados:** É o processo de organização, armazenamento, processamento e proteção de conjuntos de dados de forma a garantir sua acessibilidade, confiabilidade, segurança e eficiência. Esse processo abrange uma diversificada gama de atividades e práticas que buscam garantir que os dados estejam disponíveis quando requisitados, em um formato adequado e com qualidade suficiente para suprir às necessidades de negócios ou de outros objetivos específicos.
* **Explicação Técnica dos Dados**: É um processo pelo qual os dados são analisados e interpretados com base em uma compreensão detalhada dos métodos, técnicas e algoritmos utilizados para coletar, processar e analisar os dados. Esse processo consiste em fornecer “insights” claros e precisos sobre como os dados foram gerados, transformados e interpretados, permitindo uma compreensão mais profunda de seu significado e confiabilidade.

## Referencial teórico (subsídio teórico para propositura de ações da extensão)

* **Guido van Rossum:** Embora seja mais conhecido como o criador da linguagem de programação Python, seu trabalho influenciou indiretamente o campo de Big Data em Python. A facilidade de uso e a sintaxe clara do Python o tornaram uma escolha popular para desenvolvedores e cientistas de dados que trabalham com grandes conjuntos de dados.Ele estudou matematica na universidade de Amsterdã acabando ficando familarizado com liguagens como ALGOL 60 e Pascal [1].No motivo da criação do Python em 1996 foi dito: "Há mais de seis anos, em dezembro de 1989, eu estava procurando por um projeto de programação como "hobby" que me mantivesse ocupado durante a semana próxima ao Natal. Meu escritório... estaria fechado, mas eu tinha um computador em casa e não muito mais do que isso em mãos. Eu decidi escrever um interpretador para a nova linguagem de scripting sobre a qual eu vinha pensando ultimamente: uma descendente da ABC que agradaria a hackers de Unix/C. Escolhi Python como um título provisório para o projeto, sendo que eu estava num humor um pouco irreverente (e sendo também um grande fã do Monty Python's Flying Circus)." [2] E em 1999 Van Rossum submeteu uma proposta de financiamento a DARPA chamada de Computer Programming for Everybody (Programação de computadores Para Todos), na qual ele definiu seus objetivos para a linguagem Python como uma linguagem facil e intuitiva enquanto seja uma linguagem sofisticada e de ponta, codigo aberto entre outros.
* **Travis Oliphant:** Ele é o criador do NumPy, uma das bibliotecas fundamentais para computação científica em Python. NumPy fornece suporte para “arrays” multidimensionais e funções matemáticas de alto nível para operar nesses “arrays”. Essas capacidades são essenciais para lidar eficientemente com grandes volumes de dados, tornando-o uma figura importante no ecossistema de Big Data em Python.Oliphant atuou como presidente da Enthought de 2007 a 2011. Ele fundou a Continuum Analytics em janeiro de 2012 (posteriormente renomeada para Anaconda Inc. em 2017) [3]. Ele também foi CEO da Continuum em 2012 a 2017. Continuum faz a distribuição Python Anaconda[4]. Em julho de 2015, a Continuum Analytics recebeu 24 milhões de dólares em financiamento [5]. A Continuum Analytics recebeu um prêmio de US$ 100.000 da DARPA pela viabilidade de projetar uma extensão de linguagem paralela de dados de alto nível para Python em unidades de processamento gráfico (GPUs) [6]. Em 1º de abril de 2017, Oliphant anunciou que estava deixando a Anaconda Inc. e deixando o cargo de CEO [7]. Ele posteriormente co-fundou a Quansight no mesmo ano[8][9].
* **Wes McKinney:** Ele é o criador do Pandas, uma biblioteca Python que fornece estruturas de dados e ferramentas de análise de dados de alto desempenho e fáceis de usar. O Pandas é amplamente utilizado por cientistas de dados para manipulação e análise de dados, especialmente em ambientes de Big Data.De 2007 a 2010, McKinney pesquisou estratégias globais de macro e negociação de crédito na AQR Capital Management. Durante seu tempo na AQR Capital, ele aprendeu Python e começou a construir o que se tornaria o Pandas [10]. McKinney tornou público o projetopandas em 2009 [11]. McKinney deixou a AQR em 2010 para iniciar um doutorado em Estatística na Duke University. Ele saiu de licença da Duke no verão de 2011 para dedicar mais tempo ao desenvolvimento do Pandas, [11] culminando na escrita de Python para análise de dados em 2012.Em 2012, ele foi cofundador da Lambda Foundry Inc [12]. McKinney cofundou a Datapad com Cheng She em janeiro de 2013, com McKinney como CEO. A Datapad desenvolveu um produto de visualização de dados também na pilha Python voltado para clientes corporativos. Datapad foi adquirido pela Cloudera em setembro de 2014. [13][14] McKinney juntou-se à equipe de engenharia da Cloudera após a aquisição. Ele trabalhou em um projeto de código aberto chamado Ibis, incubado no Cloudera Labs, com o objetivo de usar Python para problemas de big data. [15] em 2016, McKinney juntou-se ao fundo de investimento Two Sigma Investments para trabalhar no Apache Arrow. Em 2018, ele lançou o Ursa Labs. [16] em 2023, ingressou na Posit (anteriormente RStudio) como Arquiteto Principal. [17]

**Referencia**

01. <https://oraculoti.com.br/2017/03/04/guido-van-rossum-fala-sobre-o-passado-presente-e-futuro-do-python/> «Guido van Rossum fala sobre o passado, presente e futuro do Python | Oráculo TI». Oráculo TI. 4 de março de 2017

02. <https://www.python.org/doc/essays/foreword/> «Foreword for "Programming Python" (1st ed.)» (em inglês)

03. <https://www.anaconda.com/continuum-analytics-officially-becomes-anaconda/Collison> , Scott (2017-06-28). "Continuum Analytics Officially Becomes Anaconda". Anaconda Inc. corporate website. 31 de março de 2019

04. <https://www.continuum.io/downloads> "Download Anaconda now!". Continuum. 8 de janeiro de 2016.

05. <http://www.businesswire.com/news/home/20150723005570/en/Continuum-Analytics-Secures-24-Million-Series-Empower> Continuum Analytics Secures $24 Million Series A Round to Empower Next Phase of Data Science - Business Wire" (Press release). 23 de julho de 2015.

06. <https://www.sbir.gov/node/697137> "Data-Parallel Analytics on Graphics Processing Units (GPUs)". SBIR. 8 de janeiro de 2016.

07. <https://twitter.com/teoliphant/status/947962036255895552> Oliphant, Travis [@teoliphant] (January 1, 2018). "My 2018 news is that I am leaving full time employment by Anaconda as of today. I will be working on improving OSS sustainability through non-profit work, and starting a new services/product incubation company to help organizations make better use of OSS and AI/ML. More to come" (Tweet). Archived from the original on 9 November 2020.- via Twitter.

08. <https://www.quansight.com/meet-our-team> "Meet Our Team". Quansight corprorate website. 31 de Março de 2019.

09. <https://www.linkedin.com/company/quansight/about/> "Quansight". LinkedIn. 31 de março de 2019.

10. <http://shop.oreilly.com/product/0636920023784.do> McKinney, Wes (2013). Python for Data Analysis (1st ed.). Sebastopol, Calif.: O'Reilly. ISBN 978-1449319793.

11. <https://qz.com/1126615/the-story-of-the-most-important-tool-in-data-science/> Kopf, Dan. "Meet the man behind the most important tool in data science", Quartz, 8 de dezembro de 2017.

12. <https://wesmckinney.com/pages/about.html> "wesmckinney.com".

13. <https://venturebeat.com/2014/09/29/data-startup-datapad-is-shutting-down-on-friday-now-that-its-been-acquired/> "Data startup DataPad gets acquired, says it will shut down on Friday". VentureBeat. 29 de setembro de 2014.

14. <https://gigaom.com/2014/09/30/cloudera-bought-datapad-because-data-scientists-need-tooling-too/> "Cloudera Bought Datapad". GigaOm. 30 de setembro de 2014.

15. <https://blog.cloudera.com/blog/2015/07/ibis-on-impala-python-at-scale-for-data-science/> "Ibis on Impala: Python at Scale for Data Science - Cloudera Engineering Blog". Cloudera Engineering Blog. 10-01-2016. [W]e are excited to announce a new open source project, called Ibis, that will deliver the great Python experience and ecosystem, only at any data and node scale.

16. <http://wesmckinney.com/blog/announcing-ursalabs/> "Announcing Ursa Labs: An innovation lab for open source data science". 19 de abril de 2018.

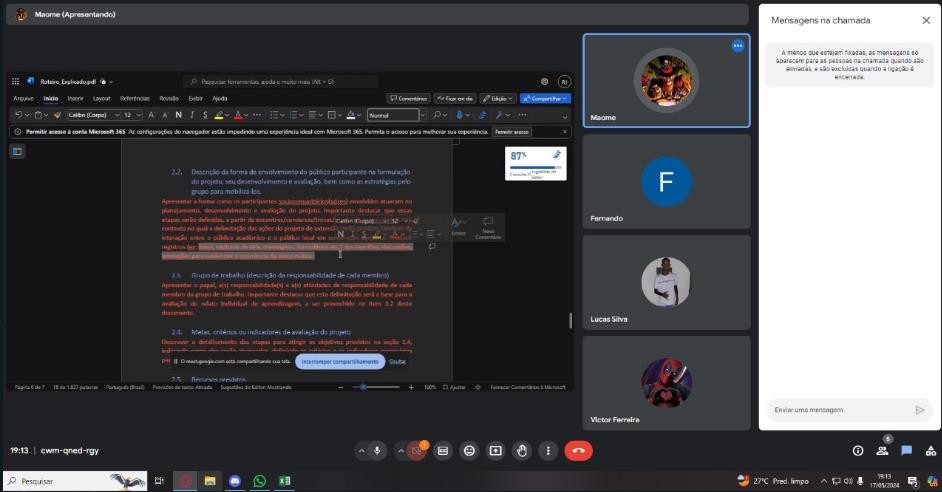
17. <https://posit.co/blog/welcome-wes/> "Welcome, Wes!". Posit. 06-11-2023.

# PLANEJAMENTO E DESENVOLVIMENTO DO PROJETO

## Plano de trabalho (usando ferramenta acordada com o docente)

* Nosso plano de trabalho consiste em trabalhar com os dados e transcrevê-los, com o objetivo de facilitar os serviços prestados pelo cliente. Início de projeto 01/04/2024 com o prazo de até 3 meses e 2 semanas. Todos estão responsáveis por todas as tarefas. Os Recursos utilizados por nossa equipe será: Google Meets (Comunicação), Google Colab (Confecção do Projeto) e ferramentas Disponibilizadas pela Estácio.
* Estamos utilizando o plano de trabalho Síncrono em nosso projeto, modelo esse que é realizado por dois ou mais agentes num momento compartilhado, no qual estarão envolvidos em uma tarefa comum. Sincronia implica em realizar junto, compartilhando o tempo comum nessa realização.

## Descrição da forma de envolvimento do público participante na formulação do projeto, seu desenvolvimento e avaliação, bem como as estratégias pelo grupo para mobilizá-los.



* Nossas reuniões e planejamentos estão sendo confeccionados tanto de forma presencial dentro da instituição, quanto à distância utilizando recursos como conferências de vídeo na interface Google Meet. Ambas ocorrem com a média de duas vezes na semana, onde as reuniões pessoais ocorrem perto do fim da semana, e as não presenciais ocorrem no início da semana. Todos os integrantes sem exceção mostram comprometimento e fazem sua presença em cada uma das reuniões, sejam elas: presenciais ou à distância.

## Grupo de trabalho (descrição da responsabilidade de cada membro)

* João e Lucas serão responsáveis pelo preenchimento dos dados não sensíveis que chegam até o grupo via Whatsapp. Já Fernando, será responsável pelo tratamento dos dados adquiridos e preenchidos por João e Lucas. Por fim, através dos dados já tratados, Rosemberg será responsável pela criação dos gráficos no PowerBI. Ademais, cabe ressaltar que todos os membros do grupo são minimamente responsáveis por todas as partes do projeto, por outras palavras, todos se ajudam ao todo no projeto extensivo, como um grupo.

## Metas, critérios ou indicadores de avaliação do projeto

* **Gerenciamento de Dados:** O gerenciamento de dados será alcançado por nossa equipe diante do tratamento dos dados recebidos no Whatsapp, onde converteremos eles em Excel, e assim, trataremos eles, filtrando apenas as informações de exímia importância para o projeto. Essa filtragem e tratação dos dados de nossa parte interessada serão feitas diante do Google Colab, na biblioteca Pandas em Python.
* **Análise em Tempo Real:** A análise em tempo real será alcançada por nossa equipe após o gerenciamento absoluto dos dados se tornar mútuo em quesito de qualidade e precisão, onde após essa difícil etapa, teremos a função de analisá-los através da criação gráfica na interface PowerBI. Com isso, visamos o entendimento completo de nossa parte interessada referente à todo material entregue no fim do projeto.
* **Explicação Técnica dos Dados:** A explicação técnica dos dados será alcançada após a análise em tempo real ser efetivamente concluída, onde com isso, nosso grupo irá explicar de forma extremamente detalhada e técnica o contexto geral dos gráficos no PowerBI e do tratamento de dados feito no Google Colab utilizando o Pandas. Cabe ressaltar que essa é a parte conclusiva do trabalho de extensão, sendo a meta final de nossas pesquisas.

## Recursos previstos

* Os recursos utilizados por nossa equipe no projeto de extensão serão: Google Meet (meio de comunicação principal entre os integrantes), Google Colab (utilizado para a confecção do projeto), PowerBI (utilizado para a análise e criação dos gráficos) e ferramentas disponibilizadas pela instituição Estácio. A linguagem utilizada para o tratamento dos dados será o Python, em sua biblioteca Pandas, pois os dados estão atualmente armazenados no formato Excel, portanto, através do Pandas, o Python irá ler o arquivo de forma similar à um banco de dados, para assim ser iniciado o tratamento. As colunas essenciais que serão utilizadas no tratamento serão as de “CEP”, “Forma de Pagamento”, “Valor do Produto” e “Produto Comprado”.

## Detalhamento técnico do projeto



* **Ingest (Ingestão):** Todos os dados não sensíveis da parte interessada foram entregues via Whatsapp para os integrantes do grupo de forma totalmente confusa e desorganizada, cabendo à nós, integrantes do projeto extensivo, organizá-lo de maneira totalmente coerente via Excel, sendo totalmente inviável a continuação daquele tipo de utilização das informações recebidas.
* **Store (Armazenamento):** Os dados não sensíveis após serem armazenados de forma organizada no Excel pelo grupo após serem obtidos de forma praticamente inutilizável, foram tratados em Python através da biblioteca Pandas, utilizando a plataforma Google Colab para tais fins, no intuito de mesclar informações de forma coerente e síncrona.
* **Apply (Aplicação):** Os dados tratados após serem tratados com o auxílio da biblioteca Pandas no Google Colab serão reutilizados para a aplicação criativa de gráficos precisos e informativos dentro da plataforma de gráficos conhecida como PowerBI.
* **Govern (Governança):** A etapa de “Ingestão” será confeccionada principalmente por João e Lucas, enquanto a parte de “Armazenamento” será feita em foco por Fernando, e por fim, a parte de “Aplicação” será utilizada em sua maioria por Rosemberg. Todavia, cabe ressaltar que todas as partes do projeto têm uma ajuda mútua de todos os membros do projeto. Os códigos serão armazenados no Git, onde a parte interessada terá acesso aos códigos utilizados no projeto de extensão, bem como, ter acesso aos gráficos feitos através do PowerBI.

# ENCERRAMENTO DO PROJETO

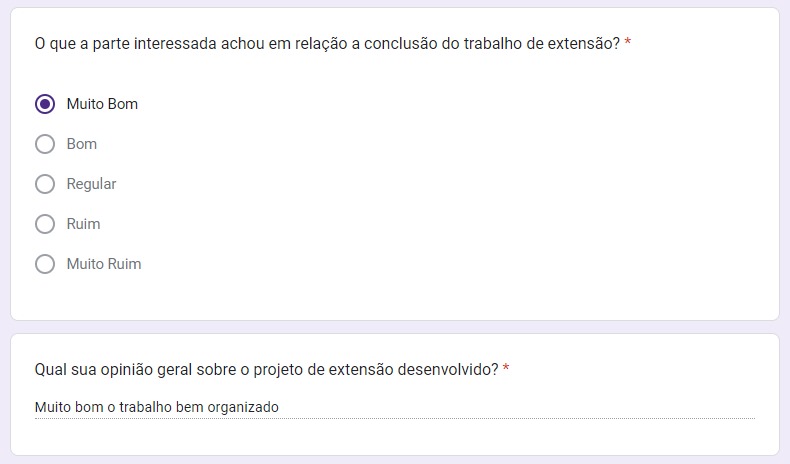
## Relato Coletivo:

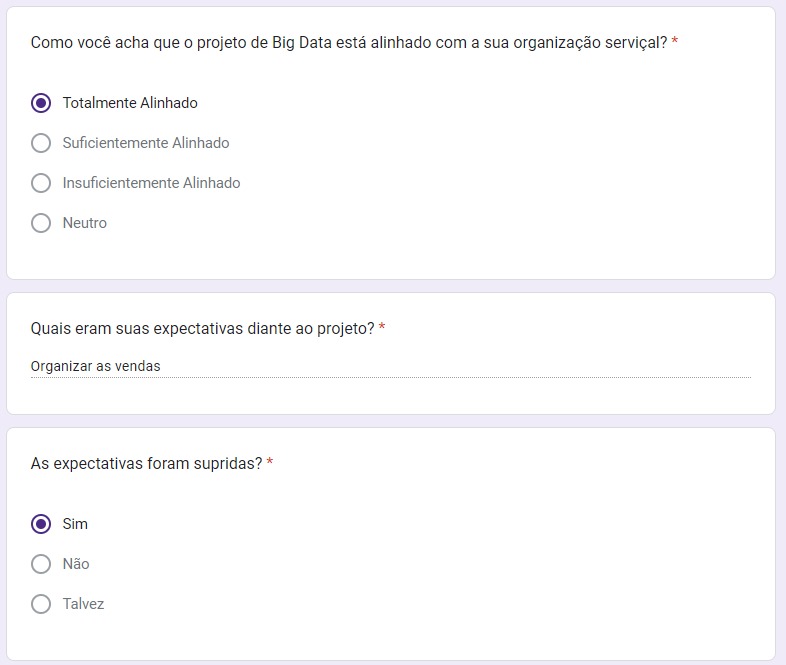
* Estamos satisfeitos com os resultados alcançados durante o projeto de extensão da disciplina de Tópicos de Big Data em Python, sentimos que o trabalho nos rendeu ótimas experiências e conhecimentos que aprimoraram nossos entendimentos pessoais, conhecimentos esses que serão utilizados durante toda nossa carreira profissional após nossa formação no campus.

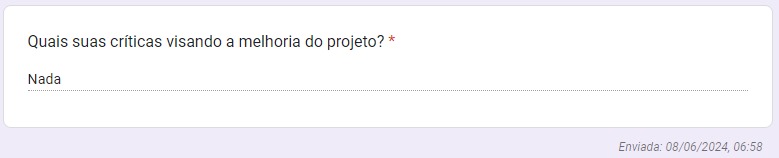
### Avaliação de reação da parte interessada

Realizar avaliação de reação com a parte interessada (ex: formulário, entrevista gravada em áudio/vídeo, depoimento em áudio/vídeo etc.), para que o efetivo atingimento dos objetivos socioncomunitários propostos fique evidente.

[**https://forms.gle/fq9EvfXdwYGSVQJ36**](https://forms.gle/fq9EvfXdwYGSVQJ36)

****

****

****

## Relato de Experiência Individual (Pontuação específica para o relato individual)

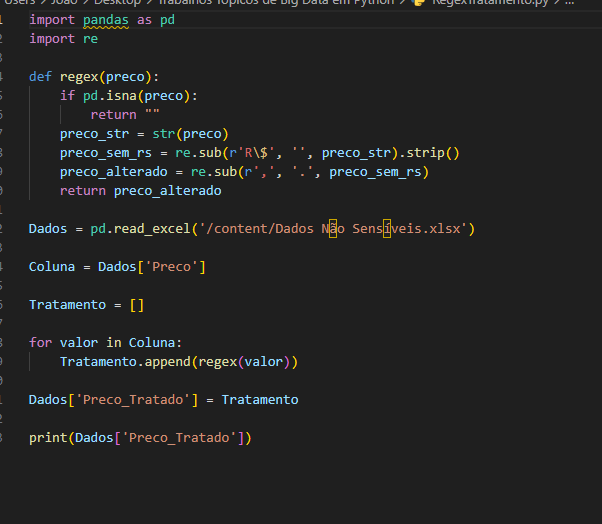
* **João Victor:** — Eu me chamo João Victor Ferreira Gonçalves e ajudei na parte de armazenamento e ingestão dos dados no projeto extensivo da disciplina de Tópicos de Big Data em Python. O projeto após ser concluído fortaleceu minhas relações comunicativas e habilidades de organização e resolução de problemáticas. Aprendi a abordar desafios de forma sistemática e destaquei a importância do aprendizado contínuo na área de TI.
* **Lucas Silva:** — Para o projeto de Big Data, comecei em 01/04/2024, com um prazo de até 2 meses. Optei por uma abordagem colaborativa e estruturada, conduzindo-o remotamente por meio do Meet. As responsabilidades foram divididas entre mim e outros colaboradores, como Lucas, Fernando, João e Rosemberg, cada um focado em suas tarefas específicas conforme o plano de trabalho.
* **Fernando:** — Me chamo Fernando e realizei a parte de armazenamento dos dados não-sensíveis da nossa parte interessada, filtrando-os na disciplina de Tópicos de Big Data em Python. Após ser concluído, pude ficar bem satisfeito com o resultado final.
* **Rosemberg:** — Meu nome é Rosemberg Jesus de Souza Junior, com o decorrer do projeto foi necessário muito esforço por todos os integrantes, devido à ausência de dados úteis, porém mesmo com as dificuldades conseguimos terminar a parte de aplicação e implementar o que tínhamos combinado no prazo estipulado.

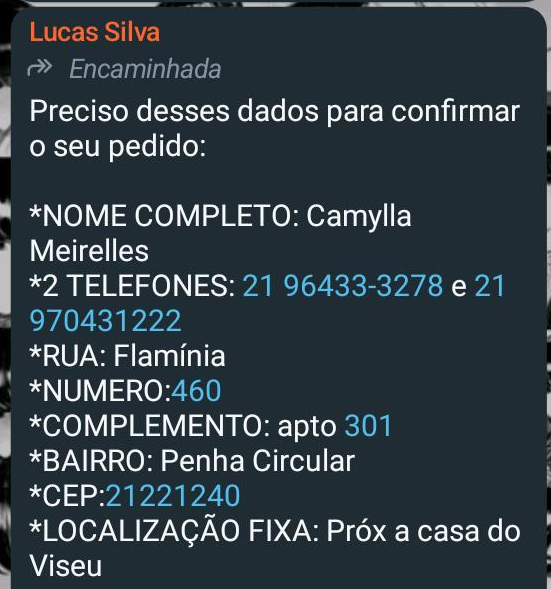
### CONTEXTUALIZAÇÃO

* **João Victor:** — Foi uma primeira experiência complicada de primeiro plano, mas gosto de desafios complicados, pois se tornam metas à serem superadas não só por mim, mas para todos os integrantes do grupo. Minha parte responsável, a parte de ingestão se tornou bastante complexa devido à desorganização referente a nossa parte interessada, o que se tornou algo um tanto quanto desesperador na hora de ingerir esses dados e passa-los para o próximo passo, a parte de aplicação.
* **Fernando Ribeiro:** — Foi uma ótima primeira experiência com o trabalho. Na minha parte, responsável pela parte de tratamento de dados no Google Colab, pensei que seria mais complicado, entretanto foi fácil e maleável fazer o projeto do esperado e tive a oportunidade de ajudar os outros membros em suas respectivas funções e ter uma melhor compreensão do trabalho ao todo.
* **Lucas Silva:** — Ao longo do desenvolvimento do projeto, desempenhei um papel ativo em vários aspectos, promovendo uma abordagem colaborativa e baseada em dados para atingir os nossos objetivos. Estou orgulhoso das contribuições que fiz e das lições valiosas que aprendi ao longo deste esforço colaborativo.
* **Rosemberg:** — Esse projeto foi muito benéfico para mim, aprendi muitas coisas em relação ao power bi e suas incríveis ferramentas, achei que seria mais complexo de mexer com os gráficos e os dados, porém foi mais fácil do que pensei, agradeço a oportunidade de participar ativamente desse projeto e a parte interessada por ter dado a chance de mostrar meus conhecimentos.

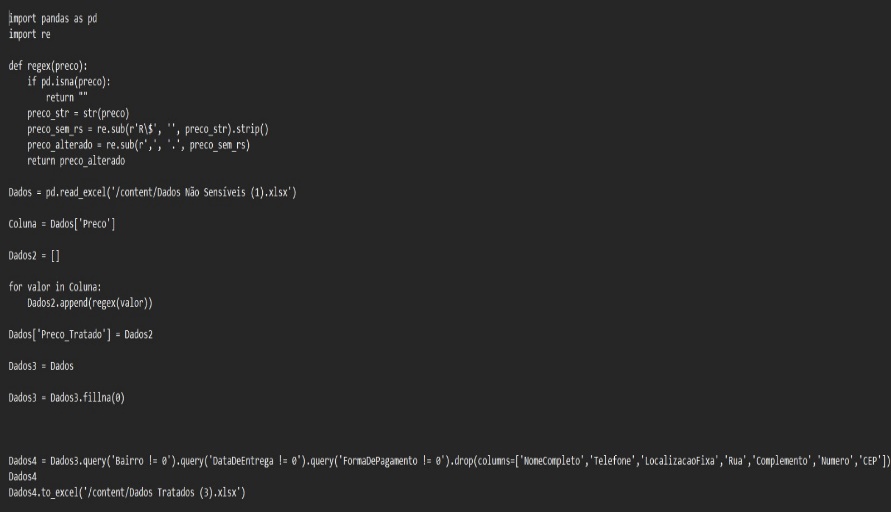
### METODOLOGIA

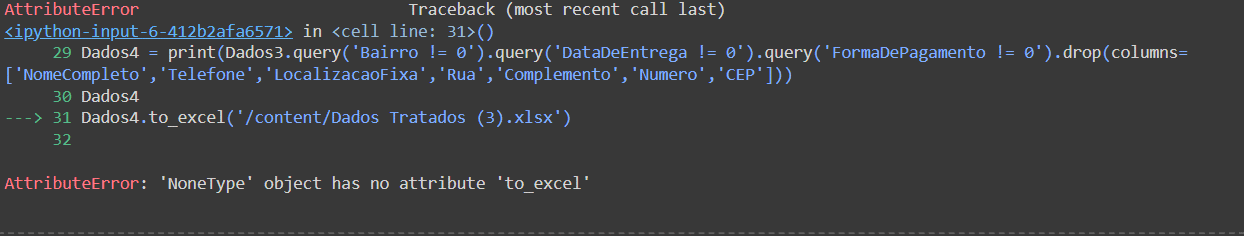
* **João Victor:** — Minha experiência vivida foi consideravelmente boa, na medida do possível. É sempre bom aprender coisas novas e ter a oportunidade de trabalhar em grupos, melhorando assim sua comunicação e desenvolvimento pessoal com novas opiniões que são divergentes da sua. O trabalho foi feito de forma online por minha parte, não tendo sido prejudicial para o aprendizado do trabalho, conseguindo separar com exímia competência o acadêmico do projeto de extensão. Provavelmente minha parte do projeto junto com a de Lucas foi a mais extensa, mas os esforços valeram a pena no fim. Ao término dessa parte, consegui ajudar meu colega Fernando com o tratamento do Regex, algo que o mesmo não estava muito familiarizado, então juntos fomos capazes de fazer algo que sozinhos não conseguiríamos, foi uma experiência incrível e espero poder trabalhar mais vezes não só com Fernando, mas com os outros membros da equipe no geral, todos foram incríveis na confecção do trabalho.

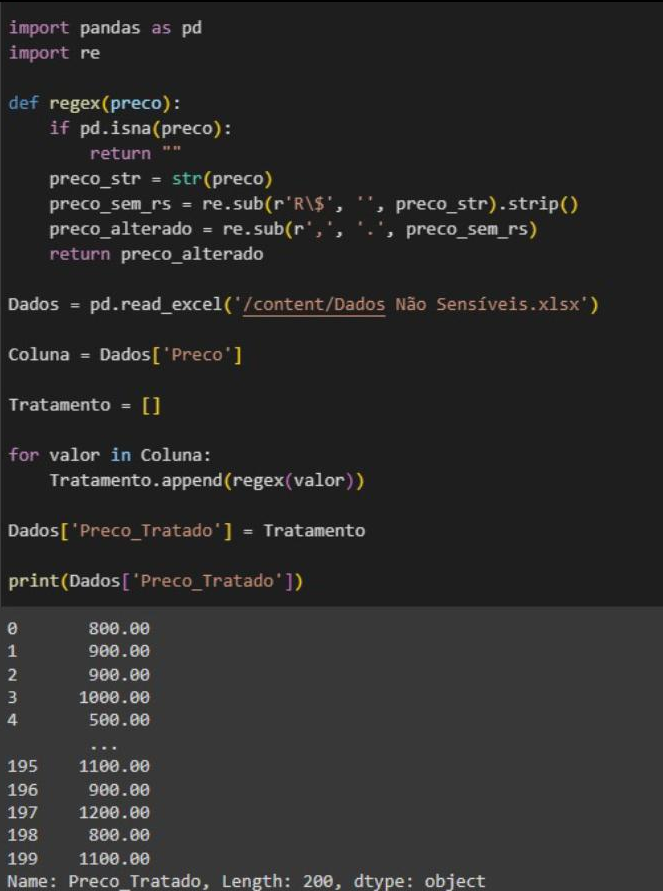




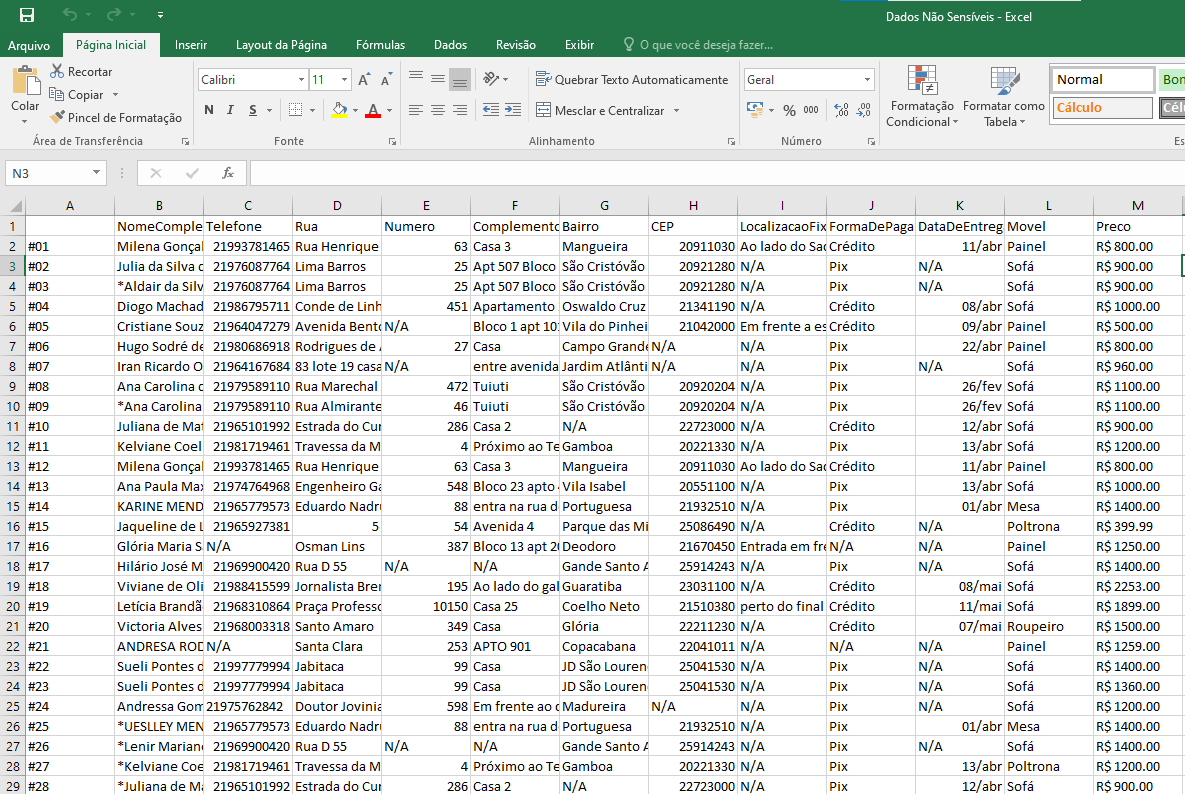
* **Fernando Ribeiro:** — A experiência foi ótima. Fizermos no online, porém não foi prejudicial para o aprendizado no aprendizado do trabalho. Somente foram envolvidos no projeto os participantes. Demorou uma semana para poder completar minha parte do trabalho. Foi feito um tratamento em regex em relação do valor do pagamento, tirar os dados sensíveis e tirar os dados contendo valor null o que pode ser visto a seguir na imagem. Houveram erros ao tratar os dados devido a lógica estar errada (Foi colocado uma print antes de Dados3 na hora de colocar em outro dataframe, Dados4, gerando um erro logico de nonetype devido a estar vazio .) Com a retirada do print foi resolvido o erro mostrado na primeira imagem. A ultima e a primeira versão do projeto.



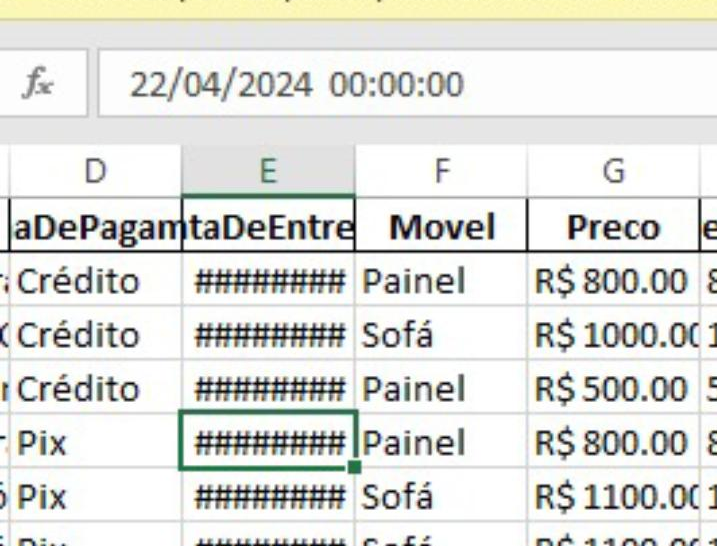




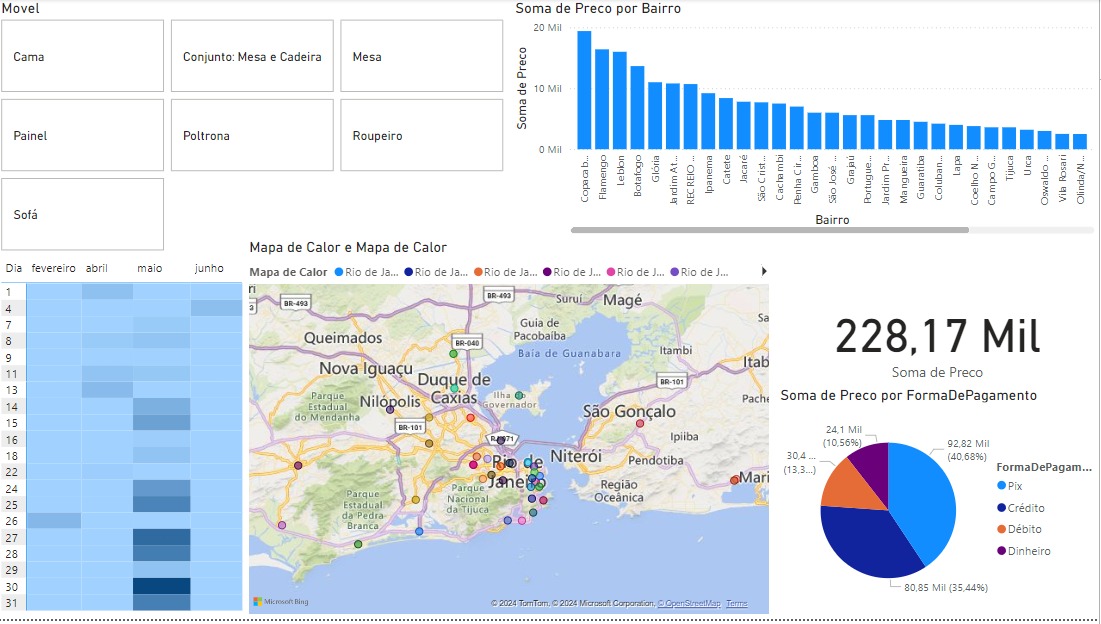
* **Lucas Silva:** — Para o projeto de Big Data, comecei em 01/04/2024, com um prazo de até 2 meses. Optei por uma abordagem colaborativa e estruturada, conduzindo-o remotamente por meio do Meet. As responsabilidades foram divididas entre mim e outros colaboradores, como Lucas, Fernando, João e Rosemberg, cada um focado em suas tarefas específicas conforme o plano de trabalho.

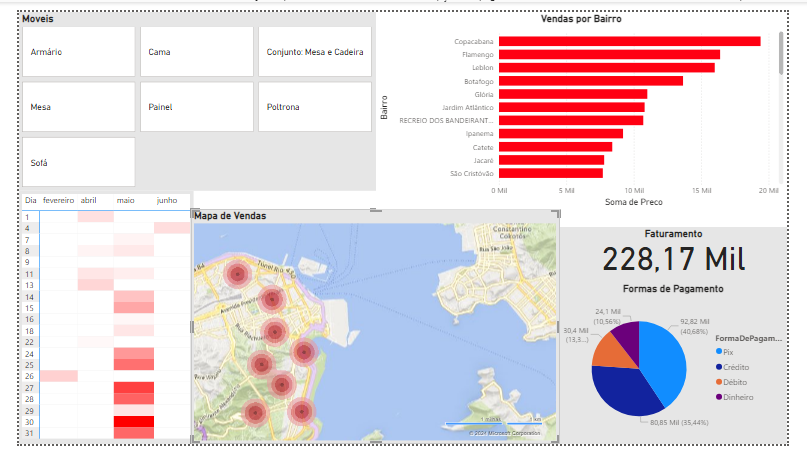


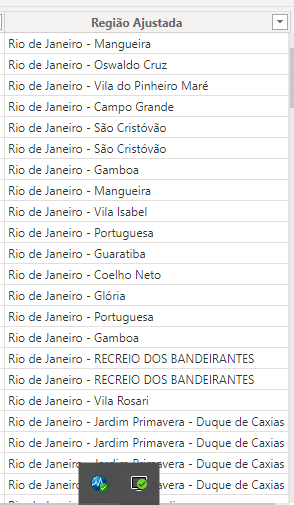


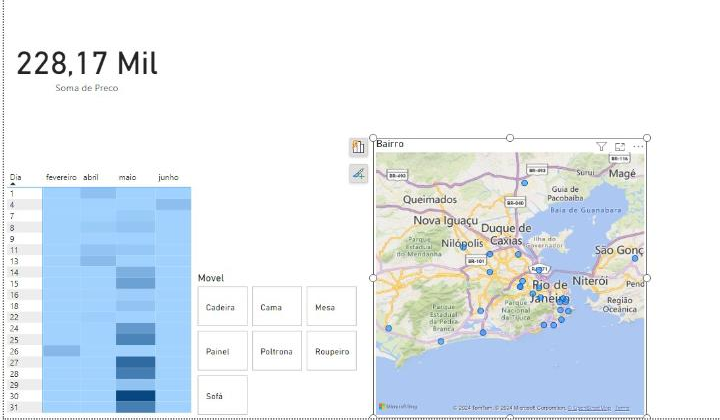


* **Rosemberg Jesus**: — Minha experiência foi ótima, nos reuníamos sempre na faculdade ou em programas de comunicação a longa distância (discord/meet), tivemos muitas conversas e troca de ideias e conhecimento, tanto dentro como fora da faculdade. E aqui estão os gráficos feitos durante o projeto. O primeiro projeto foi modificado devido ao mapa de calor mostrar lugares que não batem com a planilha de dados, adicionando um limitador na coluna Região para resolver o erro.











### RESULTADOS E DISCUSSÃO:

* **João Victor:** — Esperava uma fase de ingestão mais simples, porém por desleixo e uma leve ausência de organização de nossa parte interessada, minha parte junto com a de Lucas acabou se tornando a mais difícil do projeto sem sombra de dúvidas, não querendo desmerecer as outras partes do projeto de maneira alguma, pois tudo que eles fizeram foi de exímia importância e feito de maneira excelente. Sinto que não era a pessoa mais adequada para ficar responsável pela parte de ingestão, pois não estava acostumado com isso, porém foi uma boa primeira experiência, apesar dos desentendimentos e desorganizações diante a nossa parte interessada que mandava dado por dado após ser cobrado de forma frequente por dia, onde o mesmo não os organizava em um local na qual podia salvá-los, então o mesmo acaba sempre perdendo esses dados, o que dificultava e muito nosso trabalho.
* **Fernando Ribeiro:** — Esperava ser um trabalho difícil e árduo, todavia foi tranquilo e maleável fazê-lo. Que o início de um trabalho de Big Data e parte mais complicado pois é necessário conseguir as informações verificar sua veracidade e filtrar, mas com tempo fica mais tranquilo superar as dificuldades resultando em um trabalho difícil no início, porém ficando mais fácil. Minha recomendação é começar antes de tirar conclusões sobre o projeto.
* **Lucas Silva:** —Inicialmente, esperávamos um desenvolvimento e sem grandes contratempos para o projeto. No entanto, enfrentamos uma jornada repleta de desafios e descobertas. A comunicação entre os membros da equipe mostrou-se crucial, e o uso do Meet foi fundamental para facilitar a colaboração e a troca de ideias. Isso nos permitiu solucionar problemas de maneira ágil e eficiente. No final, conseguimos entregar um mapa de calor funcional, alinhado às expectativas da parte interessada.
* **Rosemberg Jesus:** — Minha expectativa antes e o vivido é a mesma, pois já conhecia os integrantes, com isso minha experiência foi congratulante, esse Projeto trouxe o lado curioso que eu não sabia, me senti feliz em pesquisar, ler e aprender. Senti muitas dificuldades na parte de implementação do banco de dados, porém com a ajuda do professor conseguimos fazer sem grandes dificuldades.

### REFLEXÃO APROFUNDADA

Espaço para relato sobre a experiência vivida versus teoria apresentada no relato coletivo.

* **João Victor:** — Sobre minha experiência vivida no geral, achei algo muito interessante e desafiador à ser trabalhado, uma experiência que na qual repetiria com toda certeza se fosse necessário, pois foi algo engrandecedor para mim.
* **Fernando Ribeiro:** — Em relação ao sentimento do grupo sobre o projeto e minha experiência é idêntica, foi um ótimo aprendizado e é algo que quero ter de novo em futuros trabalhos em minha carreira.
* **Lucas Silva:** —A teoria nos deu as bases do ciclo de vida do projeto, destacando o quanto é importante se comunicar bem, gerenciar o tempo e trabalhar em equipe. Mas na prática, enfrentamos desafios únicos. Isso nos fez perceber como é essencial combinar teoria e prática para aprender de verdade e estar preparado para os desafios que vêm pela frente.
* **Rosemberg Jesus:** — De fato estou muito satisfeito pelo desenvolvimento que ocorreu durante o projeto, adquiri muitos conhecimentos e um grande autodesenvolvimento pessoal.

### CONSIDERAÇÕES FINAIS

Outros aspectos que podem ser trabalhados junto à parte interessada e perspectivas de trabalhos futuros, envolvendo tanto extensão quanto pesquisa. Soluções tecnológicas alternativas que poderiam ter sido implementadas para o projeto desenvolvido.

* **João Victor:** — Em minha opinião, o principal aspecto que pode ser trabalhado juntamente à parte interessada é o quesito da organização dos dados adquiridos, pois ele não tem um controle quanto a isso. Uma solução tecnológica alternativa que poderia ter vindo a ser implementada seria a criação de um sistema de organização e backup dos dados para a nossa parte interessada.
* **Fernando Ribeiro:** — Algo notado pelo grupo no trabalho foi a mal gerenciamento de dados da parte interessada que por consequência deu mais trabalho para poder organizar os dados para poder gerar informações relevantes do projeto. Infelizmente não há uma solução tecnológica em nosso caso pois a maneira que parte interessada armazenou os dados foi em via WhatsApp em que poderia haver uma grande chance de ter alguma perda de um dado devido a correria ou falta de tratamento adequado no ciclo de vida dos dados.
* **Lucas Silva:** —No geral estou muito agradecido pela oportunidade de contribuir para o avanço da empresa. Se o cliente quiser me contatar para futuros projetos, ficarei feliz em trabalhar novamente, desde que haja uma compensação justa pelo trabalho.
* **Rosemberg Jesus:** — Em geral, me sinto grato pela parte interessada por me dar a oportunidade de utilizar meus conhecimentos para agregar na empresa, e caso o cliente esteja contente e queira me contatar para futuros projetos depois do encerramento do projeto, estaria mais do que feliz em trabalhar de novo, porém remunerado.