

**FACULDADE ESTÁCIO DO AMAZONAS
CONSTANTINO NERY**

Tratamento de Dados do Centro de Ensino Vinho Novo

JAMILY AGUIAR BEZERRA - 202201088745

LUCAS RAFAEL DE LIMA ARAGÃO - 202309855046

LYCON HEDA CASTRO DE SÁ - 202303247419

THIAGO BRUNO LIMA BENJAMIM – 202202965731

FÁBIO BRASIL E SOUZA

**2025
Manaus/Amazonas**

Sumário

1.	DIAGNÓSTICO E TEORIZAÇÃO	3
1.1.	Identificação das partes interessadas e parceiros.....	3
1.2.	Problemática e/ou problemas identificados	4
1.3.	Justificativa.....	4
1.4.	Objetivos/resultados/efeitos a serem alcançados (em relação ao problema identificado e sob a perspectiva dos públicos envolvidos)	4
1.5.	Referencial teórico (subsídio teórico para propositura de ações da extensão)	5
2.	PLANEJAMENTO E DESENVOLVIMENTO DO PROJETO	6
2.1.	Plano de trabalho (usando ferramenta acordada com o docente)	6
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.....	7
2.3.	Grupo de trabalho (descrição da responsabilidade de cada membro)	12
2.4.	Metas, critérios ou indicadores de avaliação do projeto	12
2.5.	Recursos previstos	13
2.6.	Detalhamento técnico do projeto	13
3.	ENCERRAMENTO DO PROJETO	21
3.1.	Relatório Coletivo (podendo ser oral e escrita ou apenas escrita)	21
3.2.	Avaliação de reação da parte interessada.....	22
3.3.	Relato de Experiência Individual	23
3.1.	CONTEXTUALIZAÇÃO.....	23
3.2.	METODOLOGIA	23
3.3.	RESULTADOS E DISCUSSÃO:	23
3.4.	REFLEXÃO APROFUNDADA.....	24
3.5.	CONSIDERAÇÕES FINAIS.....	24

1. DIAGNÓSTICO E TEORIZAÇÃO

1.1. Identificação das partes interessadas e parceiros

O colaborador neste projeto é o Centro de Ensino Vinho criado em 2019 pelo Jonas Sá Ribeiro e Elizabeth Aguiar Ribeiro, na época se chamava centro de formação vinho novo. No ano de 2019 iniciaram os atendimentos as crianças da comunidade. Inicialmente prestando serviço de reforço escolar, deram inicio com 06 crianças e no formato de hotelzinho onde os pais deixavam seus filhos a partir de 1 ano e 6 meses em que elas ficavam por curtos períodos, atenderam em 2019 uma média de 35 crianças. Percebendo que as famílias gostavam do serviço prestado. Os responsáveis decidiram ampliar o espaço físico na época em 2019 a escola funcionava em um espaço cedido de uma igreja localizada na rua Miraselva 1518 no bairro Redenção, a estrutura tinha 04 salas, uma cozinha e dois banheiros. Em outubro de 2019, por causa da boa aceitação da comunidade, decidiram mudar para um lugar maior localizado na rua Gurupi número 2073 bairro redenção e com o intuito de prestar um serviço de qualidade e de forma legal, a escola começou as suas atividades em 2020 em um novo endereço, que permanece até hoje: rua Gurupi 2073 Redenção. Como parte das mudanças, a escola obteve seu CNPJ e inscrição municipal provisória, dando abertura oficial a creche e escola de educação infantil, passando a se chamar Centro de Ensino Vinho Novo. O Centro de Ensino nasceu no coração do casal ainda em 2018 por causa dos trabalhos sociais que eles faziam ao observarem a carência de reforço escolar e creches que pudessem atender as crianças da comunidade e eles decidiram começar a escola.



1.2. Problemática e/ou problemas identificados

Quando se trata de escola, temos em mente que administrar informações financeiro, alunos e professores podem ser difícil. Ao entrar em contato com a escola, conversamos e perguntamos qual seria a dificuldade que estariam enfrentando atualmente, citaram falta de controle financeiro, entrada e saída.

1.3. Justificativa

O problema identificado foi:

- Falta de controle Financeiro
- Sem controle de Saídas e Entradas

Em detalhe sabemos que são duas unidades da escola, 1 unidade no bairro da Redenção e a outra se encontra no bairro da Alvorada. As informações de saída e entrada são armazenadas em um só arquivo, diferenciados na célula por uma barra e ao lado escrito “Alvorada”.

Não tem um controle específico de cada unidade, a escola da Alvorada por ser nova não tem renda própria, então toda saída é do caixa da unidade Redenção e como algumas estão com a identificação de gasto para Alvorada podemos saber o consumo da escola nesse período para que quando ela conseguir ter o seu próprio lucro, saber por mês uma média de saídas.

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

- **Organizar** os dados financeiros existentes da escola (entradas e saídas) no ambiente do Google Colab, separando corretamente as informações por unidade (Redenção e Alvorada), para facilitar a visualização e compreensão dos gastos.
- **Analigar** os dados financeiros organizados com base em critérios definidos em conjunto com os gestores da escola, identificando padrões de consumo, desequilíbrios e oportunidades de melhoria.
- **Fornecer** à equipe gestora um relatório claro e visual com os resultados obtidos a partir da análise, apoiando a tomada de decisões estratégicas e a definição de metas específicas para o equilíbrio financeiro entre as unidades.

1.5. Referencial teórico (subsídio teórico para propositura de ações da extensão)

Com o constante avanço da tecnologia, atividades que antes eram impossíveis ou difíceis de serem realizadas tornaram-se viáveis. Um exemplo claro é a comunicação: antigamente feita por meio de cartas, que levavam dias ou até meses para chegar ao destinatário, hoje pode ser realizada instantaneamente por meio de mensagens ou ligações via celular. Essa inovação, que antes era acessível apenas a uma parcela da sociedade, passou a fazer parte do cotidiano da maior parte da população mundial.

Contudo, esse avanço também traz implicações. Com um número crescente de pessoas trocando informações em tempo real ao redor do mundo, surge o desafio de como armazenar, processar e transmitir uma enorme quantidade de dados. É nesse contexto que se insere o conceito de *Big Data*.

Segundo Chiavegatto Filho (2015, p. 325), “define-se *Big Data* como uma quantidade de dados suficientemente grande que leve a uma mudança nas formas tradicionais de análise de dados”. Uma das formas tradicionais às quais ele se refere são os bancos de dados relacionais, como o SQL, que se mostram ineficientes para lidar com dados não estruturados (como vídeos, imagens e e-mails) e com volumes extremamente elevados de informações. Por outro lado, o *Big Data* permite o tratamento de diferentes tipos de dados — estruturados, semiestruturados e não estruturados —, viabiliza o processamento em tempo real (*streaming*) e contribui para a identificação de padrões e tendências por meio de técnicas como o *Machine Learning*.

Marquesone (2016) afirma que, além do grande volume de dados, é necessário considerar outras duas propriedades fundamentais: a variedade e a velocidade. O volume corresponde à característica central do conceito de *Big Data*; a variedade diz respeito aos diferentes tipos de dados; e a velocidade refere-se à rapidez com que esses dados são gerados, processados e analisados, formando o que é conhecido como os "3 V's".

Contudo, para uma melhor tratativa dos dados, faz-se necessária a escolha de uma linguagem adequada, que garanta a eficiência das operações, a segurança, a qualidade e a facilidade de uso, permitindo uma comunicação mais eficaz entre os responsáveis pelo tratamento dessas informações. Por esses motivos, foi selecionada a linguagem de programação Python. Segundo McKinney (2018), Python é uma escolha

bastante atrativa, devido à sua ampla e ativa comunidade, característica decorrente de ser uma das linguagens mais utilizadas no mundo. Esse fator proporciona um suporte mais robusto a bibliotecas como o *Pandas*, que contribuem significativamente para uma análise de dados eficiente.

Por fim, considerando a necessidade de uma plataforma para a criação e execução de códigos, optou-se pelo Google Colab, por ser uma ferramenta gratuita baseada em nuvem, oferecida pela Google, com acesso a GPUs. Essa característica permite a realização de operações mais complexas de forma intensiva e eficiente.

2. PLANEJAMENTO E DESENVOLVIMENTO DO PROJETO

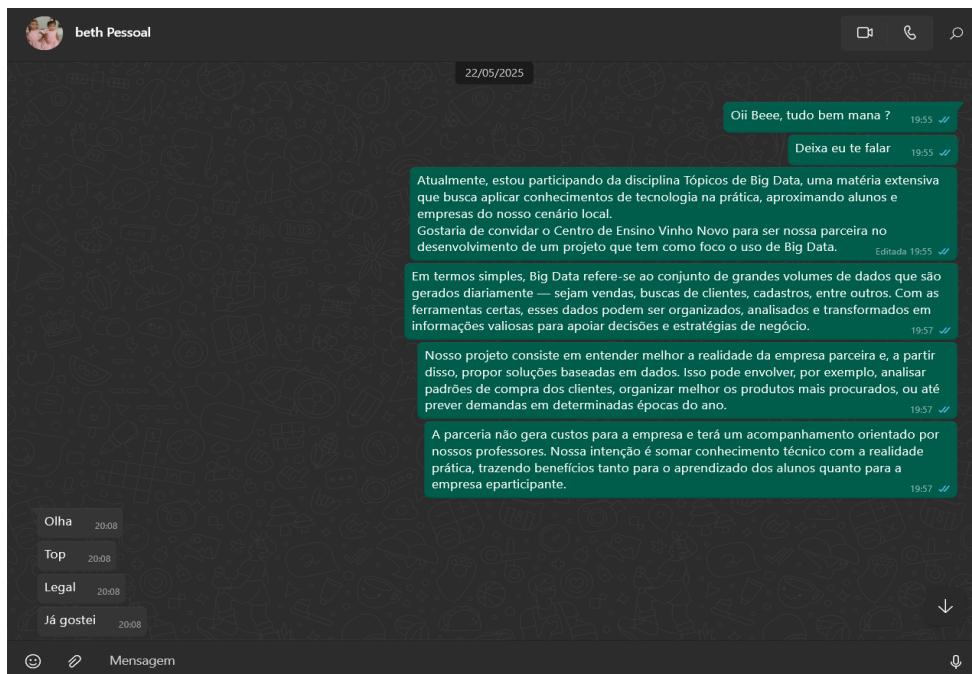
2.1. Plano de trabalho (usando ferramenta acordada com o docente)

Atividades	Abril	<th>Junho</th>	Junho
Desenvolvimento do referencial teórico	X	X	X
Documentação das etapas do projeto	X	X	X
Busca por parceiro	X	X	
Identificar as partes interessadas		X	
Identificar o problema para ser resolvido		X	
Definir o escopo do projeto		X	
Definir objetivos e funções		X	
Definir as responsabilidades de cada membro do grupo		X	
Dar início requisição dos arquivos		X	X
Observar o que de valor poderia ser extraído dos dados			X
Dar início ao tratamento dos dados			X

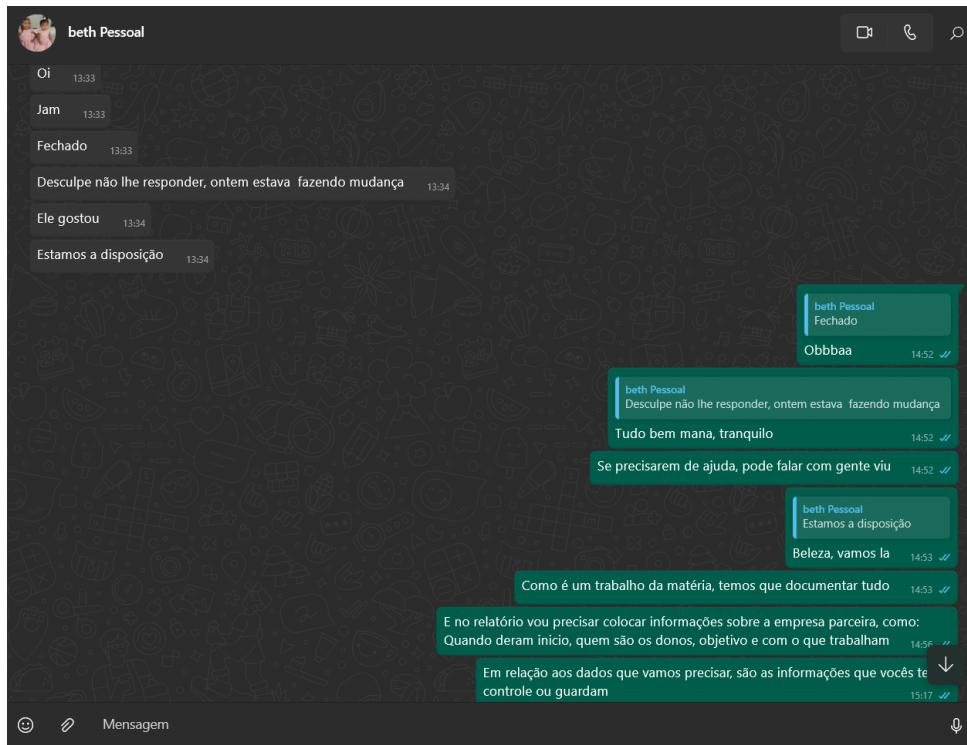
Preparar apresentação ao cliente			X
Entrega			X
Feedbacks dos resultados			X

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.

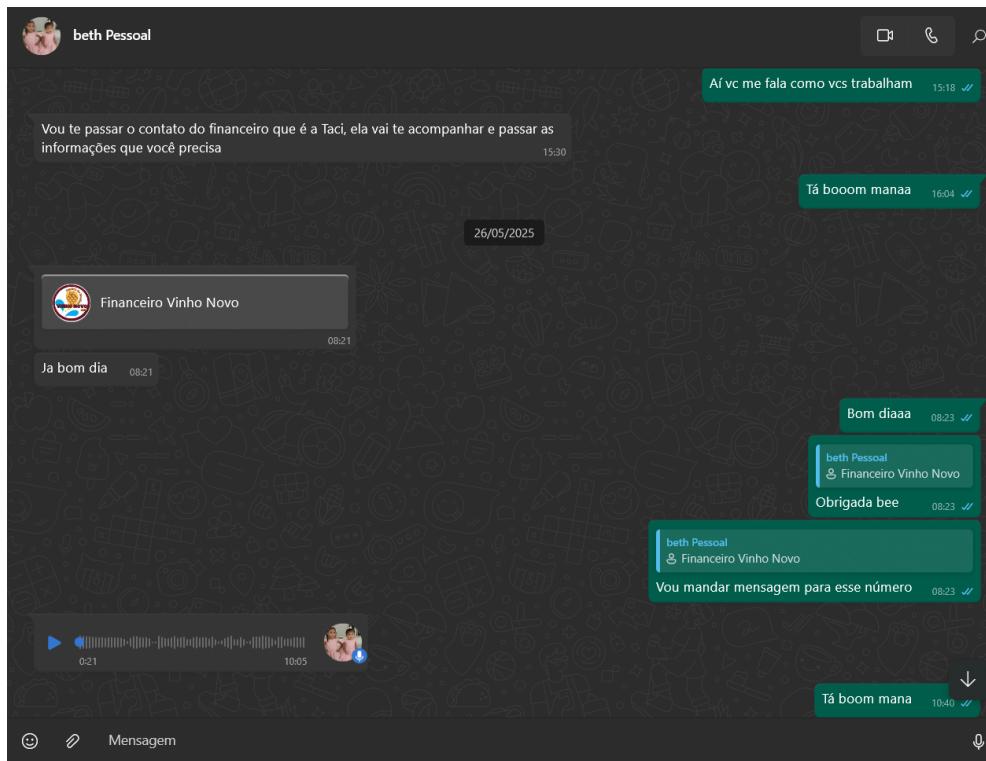
O nosso primeiro contato com a Escola parceira foi pelo WhatsApp, aonde um a Jamily Aguiar entrou em contato com a Diretora Elizabeth Aguiar e no dia 22/05 enviamos a proposta de parceria para o desenvolvimento do trabalho da matéria “Tópicos de Big Data em Python”



Após entrarmos em contato com a Diretora, ela fez uma reunião com o Diretor Jonas Sá e assim que tudo estivesse de acordo ela nos avisaria para dar início no trabalho. Recebemos a confirmação no dia 25/05. E já começamos o processo de identificação do problema e levantamento de dados necessários.



O problema identificado foi “Falta de controle Financeiro - “Entradas e Saídas” e então nos foi passado o número do Financeiro da Escola, por onde iríamos ter toda informação e material necessário.



Após recebermos a confirmação, entramos em contato com a escola para verificar o melhor dia em que pudéssemos pegar a assinatura de consentimento para o uso dos dados.



A seguir temos a Carta de Autorização assinada pelo Diretora Elizabeth Aguiar Rodrigues, autorizando o uso de dados para fins educacionais.

CARTA DE AUTORIZAÇÃO

Eu, Elizabeth Aguiar Ribeiro, Diretora
do Centro de Ensino e Pesquisa, situada no endereço Rua Grunpi 2073, autorizo
a realização das seguintes atividades acadêmicas extensionistas associadas à disciplina
TÓPICOS DE BIG DATA COM PYTHON, da **FACULDADE ESTÁCIO DO AMAZONAS**, sob
orientação do Prof. FÁBIO BRASIL E SOUZA:

Atividades:

Prezar pela confidencialidade e proteção dos dados segundo a LGPD.
Usar todo os dados somente para a aprendizagem dos alunos na matéria de Big Data.
Coleta, edição, armazenamento, análise e visualização de dados.

Conforme combinado em contato prévio, as atividades acima descritas são autorizadas para os seguintes alunos:

Nome do Aluno	Curso
JAMILY AGUIAR BEZERRA	CIÊNCIA DA COMPUTAÇÃO
LYCON HEDA CASTRO DE SÁ	CIÊNCIA DA COMPUTAÇÃO
THIAGO BRUNO LIMA BENJAMIM	ANÁLISE E DESENVOLVIMENTO DE SISTEMAS
LUCAS RAFAEL DE LIMA ARAGÃO	ANALISE E DESENVOLVIMENTO DE SISTEMAS

Declaro que fui informado por meio da **Carta de Apresentação** sobre as características e objetivos das atividades que serão realizadas na organização a qual represento e afirmo estar ciente de tratar-se de uma atividade realizada com intuito exclusivo de ensino de alunos de graduação, sem a finalidade de exercício profissional.

Desta forma, autorizo:

- (O acesso a informações e dados que forem necessários à execução da atividade;
(O registro de imagem por meio de fotografias e vídeos;

Manaus AM, 27 de 05 de 2025.



Elizabeth Aguiar Ribeiro
Diretora
Centro de Ensino e Pesquisa
Rua Grunpi 2073 - Centro
Manaus - AM - CEP: 69010-021
Fone: (92) 3619-9522

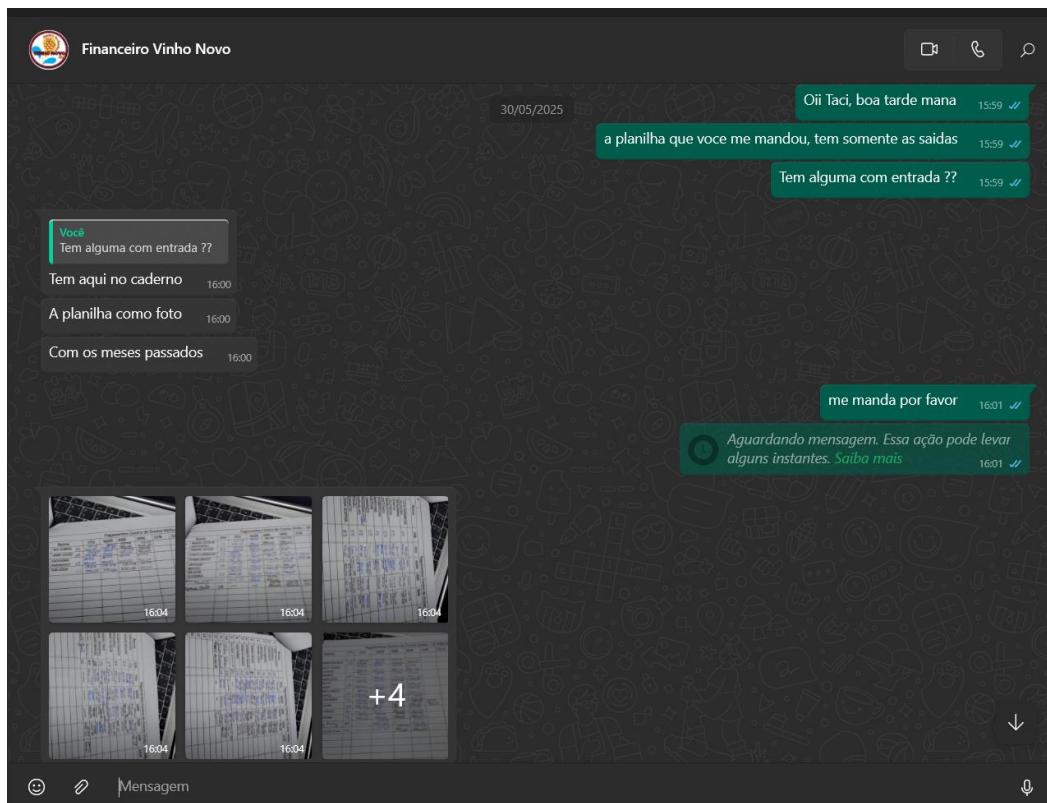
Após todas as autorizações serem concedidas, entramos em contato com o setor financeiro do Centro de Ensino Vinho Novo. Explicamos o objetivo do trabalho e perguntamos como funcionava o controle financeiro da escola, se fosse manual ou digital. Foi nos enviado um arquivo XLSX, nomeado como “Saídas”, no mesmo estava o controle de saídas do mês 01/25 ao mês 04/25.

A	B	C	D	E	F
1	FLUXO DE CAIXA				
2	ENTRADA	SAÍDA	SALDO		
3	R\$ -	R\$	56.286,61	-R\$	56.286,61
4					
5	DATA	MOVIMENTAÇÃO	DESCRIÇÃO	VALOR	OBSERVAÇÃO
6	01/04/2025	Saída	ENERGIA ESCOLA	R\$ 2.032,53	PIX
7	01/04/2025	Saída	FATURA BEMOL	R\$ 195,69	PIX
8	01/04/2025	Saída	BEMOL	R\$ 493,68	PIX
9	01/04/2025	Saída	CARNE	R\$ 77,00	
10	01/04/2025	Saída	Armador de rede	R\$ 19,00	
11	01/04/2025	Saída	material cantina (banana Chips)	R\$ 80,00	
12	01/04/2025	Saída	Concerto da impressora	R\$ 140,00	PIX
13	01/04/2025	Saída	Restante do valor da Laina	R\$ 433,00	PIX
14	01/04/2025	Saída	Sarah a penúltima parcela da sua rescisão	R\$ 833,33	PIX
15	02/04/2025	Saída	compras proteína	R\$ 105,00	
16	02/04/2025	Saída	tinta spray	R\$ 20,00	
17	02/04/2025	Saída	Negociação CDL	R\$ 380,00	PIX
18	02/04/2025	Saída	jaqueline apostilas e agendas	R\$ 300,00	
19	03/04/2025	Saída	feira	R\$ 85,53	
20	03/04/2025	Saída	Simples nacional de janeiro	R\$ 129,85	PIX
21	03/04/2025	Saída	Simples nacional de dezembro	R\$ 134,00	PIX
22	03/04/2025	Saída	Simples nacional de novembro	R\$ 285,00	PIX
23	03/04/2025	Saída	Simples nacional de outubro	R\$ 136,00	PIX
24	04/04/2025	Saída	feira almoço	R\$ 72,00	PIX
25	04/04/2025	Saída	vale Jonas	R\$ 100,00	
26	05/04/2025	Saída	Dízimo escola	R\$ 1.000,00	PIX
27	08/04/2025	Saída	compras carne	R\$ 112,50	
28	08/04/2025	Saída	vale Jonas	R\$ 300,00	
29	08/04/2025	Saída	Pagamento funcionários	R\$ 25.302,00	PIX
30	08/04/2025	Saída	parcela cartão de crédito da empresa Santander	R\$ 1.512,70	
31	08/04/2025	Saída	Energia casa	R\$ 578,03	
32	08/04/2025	Saída	Empréstimo mercado	R\$ 626,31	
33	09/04/2025	Saída	Parcela do Carro	R\$ 1.243,17	

Na reunião feita presencialmente no dia em que um dos membros da equipe foi ao Centro de Ensino para o recebimento da assinatura, aconteceu um momento de conversa em que eles falaram as dificuldades que estavam enfrentando e quais seriam os pontos fracos da empresa, não somente a parte financeira, mas “controle de mensalidades”, “controle de faltas e desempenho dos professores”. Para a realização do trabalho, focamos na área financeira pois ao longo do caminho encontramos o desafio de não encontrar uma empresa que teria os dados para serem utilizados e não podíamos perder a oportunidade de trabalhar com que tínhamos em mãos, então, focar somente no controle de saídas foi uma decisão tomada em conjunto pensando no tempo que teríamos para desenvolver tudo.

Até o momento só tínhamos os dados de saída, solicitamos os dados de entrada, porém não tivemos retorno. Algum tempo depois, no dia 30/05, entramos em contato para saber se os dados que eles guardavam seria somente os de “saídas” ou se teria os de “Entrada”.

O financeiro nos informou que teria sim os dados de “entrada” e que elas seriam as “Mensalidades”. O Centro trabalha com “Maternal I, II e III” e a series do “1º ao 5º ano” do ensino fundamental e as entradas são as mensalidades de cada aluno.



O controle de mensalidades é feito em uma planilha impressa com todas as turmas dos 2º turnos, matutino e vespertino.

Pagamentos Centro de Ensino Vinho – 3º ANO VESPERTINO										
	Nome	DIA	FEV	MAR	ABR	MAI	JUN	JULH	AGOS	SE
1	CARLOS MIGUEL	05	624.7005 623.06		04104 440,00	20105 350,00				
2	DAVID RAMOM	05	05102 350,00	05103 350,00	04103 350,00	04105 350,00				
3	DANIEL SOARES	15	05102 440,00	05103 440,00	14104 440,00	20105 440,00				
4	ESTHER LUIZA	05	03102 440,00	06103 440,00	06103 440,00	30103 440,00	30104 440,00			
5	EFRAIM MORAES	07	04102 400,00	07103 400,00	04104 400,00	08105 400,00	08105 400,00			
6	GUILHERME	05	04102 440,00	06103 440,00	06103 440,00	02104 440,00	02105 440,00			
7	JOSEPH NASC	05	05102 400,00	06103 400,00	06103 400,00	03104 400,00	05105 400,00			
8	KENNEDY	12	07102 440,00	07103 440,00	07103 440,00	14103 440,00	14104 440,00			
9	MIRELLA NEVES	05	30101 440,00	28102 440,00	28103 440,00	28103 440,00	01105 440,00	30105 440,00		
10	RAYLA VITORIA		05102 350,00	06103 350,00	06103 350,00	05104 350,00	05105 350,00			
11	ALICE NICOLE	05	05102 440,00	06103 440,00	06103 440,00	04100 440,00	04105 440,00			
12	Arthur Rulen	05	06102 440,00	10103 440,00	06103 440,00	04104 440,00	14105 539,50			
13										
4										
5										

Todas as planilhas impressas do controle de mensalidades foram passadas para o Excel à mão.

Após isso, passamos a trabalhar com os dados no Google Colab. A equipe conseguiu ajudar a empresa a digitalizar seus dados e ter uma boa organização financeira de suas entradas e saídas. A comunicação com o Centro de Ensino foi uma mão de via dupla, o assunto do qual eles tinham dúvida ou se eles gostassem de mais algum resultado, foi conversado e explicado com toda compreensão. O mesmo aconteceu conosco, o Centro forneceu total suporte à equipe em tudo, se em algum momento focemos precisar de algum arquivo, sem muito demora o financeiro disponibilizava.

2.3. Grupo de trabalho (descrição da responsabilidade de cada membro)

TAREFAS	PARTICIPANTES
CONTATO COM A EMPRESA	JAMILY AGUIAR
COLETA DE DADOS	JAMILY AGUIAR
MANIPULAÇÃO DOS DADOS "SAÍDAS"	LUCAS RAFAEL
DESENVOLVIMENTO DO RELATÓRIO	JAMILY AGUIAR LUCAS RAFAEL
MANIPULAÇÃO DOS DADOS "ENTRADA"	JAMILY AGUIAR
SLIDE DE APRESENTAÇÃO	JAMILY AGUIAR LINKON HEDA
APRESENTAÇÃO	EQUIPE

2.4. Metas, critérios ou indicadores de avaliação do projeto

O trabalho mostra de forma simples, como podemos organizar, analisar e fornecer os dados de uma empresa. A meta da equipe é mitigar a desorganização de dados sensíveis do Centro de Ensino Vinho Novo, mostrando ao público-alvo como é enriquecedor saber os pontos positivos e negativos da situação atual da empresa e através disso tomar uma decisão que leva para um futuro engrandecedor. Após recebermos os dados e começar a manipulação deles, conseguimos digitalizar “entrada” e “saída” e dá os resultados de “gastos por mês” e “lucro por mês”. Dando a eles uma visão geral do rendimento do Centro de Ensino.

2.5. Recursos previstos

Na parte de programação e ajuste na manipulação de dados, utilizamos vídeos no YouTube, pesquisas no Google Acadêmico e aulas na Faculdade como referências, que podem ser consultadas nos seguintes links:

- <https://m.youtube.com/playlist?list=PL3ZsII15yo2pfkf7EGNR14xTwe-wZ2bNX>
- <https://m.youtube.com/watch?v=C0aj3FjN5e0&pp=ygUTUGFuZGFzIHB5dGhbibJdXJzbw%3D%3D>

2.6. Detalhamento técnico do projeto

Após o recebimento dos dados de saída fornecidos pelo cliente, foi criada uma conta no google com o objetivo de ser para livre acesso para cada membro do grupo que queira ver o andamento do código ou modificá-lo e somente então iniciar a integração do google drive com a plataforma google Colab.

```
✓ [94] # Importa a biblioteca Pandas com apelido pd
      import pandas as pd

✓ [95] # Carrega o arquivo Excel inteiro e armazena na variável df
      df = pd.ExcelFile('/content/drive/MyDrive/Saídas-1.xlsx')

✓ [96] # Exibe o nome das planilhas existentes no arquivo
      print(df.sheet_names)

→  ['JAN', 'FEV', 'MAR', 'ABR', 'MAI', 'JUN', 'JUL', 'AGO', 'SET', 'OUT', 'NOV', 'DEZ']

✓ [97] # Armazena na variável 'meses_preenchidos' os meses que possuímos os dados
      meses_preenchidos = ['JAN', 'FEV', 'MAR', 'ABR', 'MAI']

✓ [98] # Lê as planilhas do arquivo e as concatena em um único dataframe criando uma coluna mês
      # para identificar a qual mês os dados pertencem
      df_valido = pd.concat(
          [pd.read_excel(df, sheet_name=nome, header=4).assign(Mês=nome) for nome in meses_preenchidos],
          ignore_index=True)
```

Primeira parte do código, focada em filtrar as partes de interesse do arquivo.

```
[99] # Mostra as 5 primeiras linhas do dataframe df_valido
print(df_valido.head())

DATA MOVIMENTAÇÃO DESCRIÇÃO VALOR \
0     NaN     Saída      Comprar HD 200.00
1     NaN     Saída conta energia/ ref dezembro escola 3245.98
2     NaN     Saída          agua/ novembro 2024 1180.97
3     NaN     Saída      Material de expediente 132.06
4     NaN     Saída J s Ribibeiro / grafica 150.00

OBSERVAÇÃO Mês Unnamed: 0 DATA
0     pix JAN      NaN  NaN
1     pix JAN      NaN  NaN
2     pix JAN      NaN  NaN
3     pix JAN      NaN  NaN
4     pix JAN      NaN  NaN
```

```
[100] # Exibe os nomes das colunas do dataframe df_valido
print(df_valido.columns)
```

```
Index(['DATA', 'MOVIMENTAÇÃO', 'DESCRIÇÃO', 'VALOR', 'OBSERVAÇÃO', 'Mês',
       'Unnamed: 0', 'DATA'],
      dtype='object')
```

Demonstração de como estava as colunas e linhas do arquivo antes do tratamento.

```
✓ [101] # Remove colunas com nome 'Unnamed'
0s df_valido = df_valido.loc[:, ~df_valido.columns.str.contains('Unnamed')]

✓ [102] # Uniformiza os nomes do dataframe, removendo os espaços em brancos dos lados
# e convertendo todos os caracteres para maiúsculo
df_valido.columns = df_valido.columns.str.strip().str.upper()

✓ [103] # Demostrar o resultado após a remoção da coluna 'Unnamed' e padronização dos nomes
# das colunas
print(df_valido.columns)

Index(['DATA', 'MOVIMENTAÇÃO', 'DESCRIÇÃO', 'VALOR', 'OBSERVAÇÃO', 'MÊS',
       'DATA'],
      dtype='object')

✓ [104] # Cria uma lista com índices das colunas com nome DATA
data_cols = [i for i, col in enumerate(df_valido.columns) if col == 'DATA']
```

Eliminação das colunas totalmente vazias e preparação para a junção das colunas DATA.

```

✓ [105] # Usa a primeira coluna DATA como base e preenche os valores da primeira com
  # com os valores da segunda
  if len(data_cols) > 1:
      data_completa = df_valido.iloc[:, data_cols[0]]
      for i in data_cols[1:]:
          data_completa = data_completa.combine_first(df_valido.iloc[:, i])

✓ [106] # Substitui os valores da coluna DATA do dataframe df_valido pelos valores data_completa
df_valido.drop(columns=df_valido.columns[data_cols], inplace=True)
df_valido['DATA'] = data_completa

✓ [107] # Substitui as colunas DATA pelo resultado unificado
df_valido = df_valido.drop(columns='DATA')
df_valido.insert(0, 'DATA', data_completa)

✓ [108] # Resultado após a unificação das colunas DATA
print(df_valido.columns)

→ Index(['DATA', 'MOVIMENTAÇÃO', 'DESCRIÇÃO', 'VALOR', 'OBSERVAÇÃO', 'MÊS'], dtype='object')

```

Unificação das colunas DATA e demonstração dos resultados.

```

[109] # Copia o dataframe df_valido para df_backup
df_backup = df_valido.copy()

[111] # Atribui a df_valido as colunas relevantes para a análise
df_valido = df_valido[['DATA', 'DESCRIÇÃO', 'VALOR', 'MÊS']]

```

```
[112] df_valido.head()
```

	DATA	DESCRIÇÃO	VALOR	MÊS	
0	NaN	Comprar HD	200.00	JAN	
1	NaN	conta energia/ ref dezembro escola	3245.98	JAN	
2	NaN	agua/ novembro 2024	1180.97	JAN	
3	NaN	Material de expediente	132.06	JAN	
4	NaN	J s Ribibeiro / grafica	150.00	JAN	

Próximas etapas: [Gerar código com df_valido](#) [Ver gráficos recomendados](#)

Cópia do que foi feito até aqui e filtragem das colunas necessária para análise.

```

[113] # Preenche os dados vazios da coluna DATA com a str 'Data não informada'
df_valido['DATA'] = df_valido['DATA'].fillna("Data não informada")

[114] df_valido.head()

```

	DATA	DESCRIÇÃO	VALOR	MÊS
0	Data não informada	Comprar HD	200.00	JAN
1	Data não informada	conta energia/ ref dezembro escola	3245.98	JAN
2	Data não informada	agua/ novembro 2024	1180.97	JAN
3	Data não informada	Material de expediente	132.06	JAN
4	Data não informada	J s Ribibeiro / grafica	150.00	JAN

Próximas etapas: [Gerar código com df_valido](#) [Ver gráficos recomendados](#) [New interactive sheet](#)

```

[115] # Verifica quantos dados ausentes há na coluna VALOR
df_valido['VALOR'].isna().sum()

[116] np.int64(2)

```

Preenche dados ausentes da coluna DATA e verifica quantas linhas estão vazias na coluna VALOR.

```

[116] # Elimina as linhas na qual a coluna valor está ausente
df_valido = df_valido.dropna(subset=['VALOR'])

[117] # Importa a biblioteca matplotlib para criação de gráficos
import matplotlib.pyplot as plt

[118] # Cria uma lista com os meses ordenados
meses_ordenados = ['JAN', 'FEV', 'MAR', 'ABR', 'MAI']

[119] # Converte a coluna MÊS em uma categoria ordenada segundo a lista meses_ordenados
df_valido['MÊS'] = pd.Categorical(df_valido['MÊS'], categories=meses_ordenados, ordered=True)

[120] # Agrupa os dados por mês e soma os valores da coluna VALOR para cada mês
gastos_por_mes = df_valido.groupby('MÊS')['VALOR'].sum()

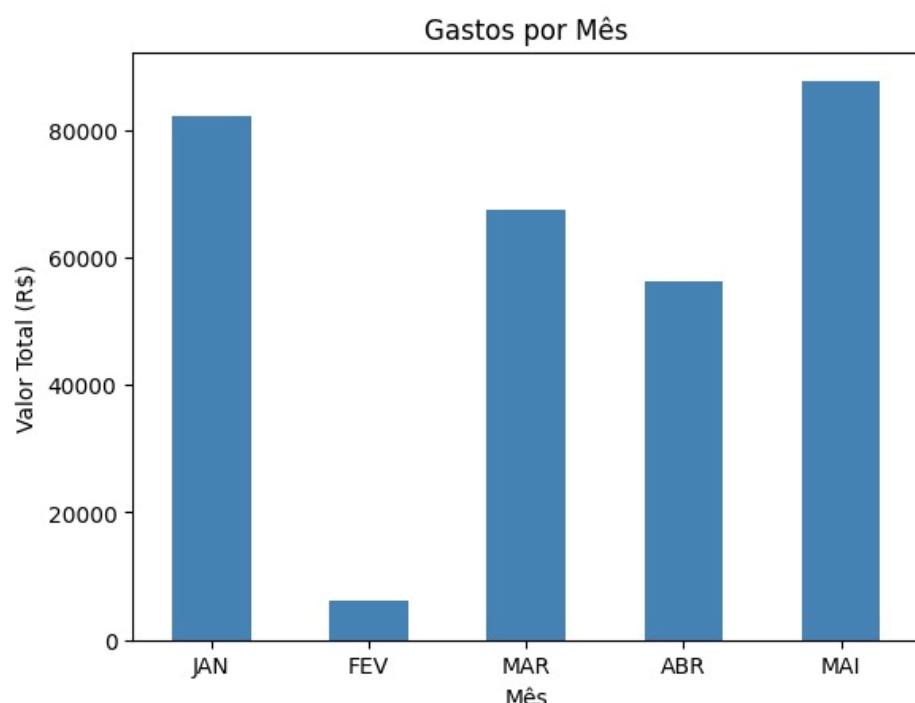
```

Elimina as linhas vazias da coluna VALOR e importa a biblioteca matplotlib para gerar um gráfico.

```
# Cria um gráfico de barras com dados dos gastos de cada mês  
gastos_por_mes.plot(kind='bar', color='steelblue')  
  
plt.title('Gastos por Mês')  
plt.xlabel('Mês')  
plt.ylabel('Valor Total (R$)')  
plt.xticks(rotation=0)  
plt.show()
```

Gastos por Mês

Gera o gráfico.



Representação gráfica dos meses com maior gasto.

```
# Salva o dataframe em um novo arquivo excel  
df_valido.to_excel('/content/drive/MyDrive/df_valido_tratado.xlsx', index=False)
```

Salva os dados tratados em um novo arquivo.

Já a manipulação dos dados de “Entrada”, foi da seguinte forma:

```
1 # └── Etapa 1 – Upload do arquivo  
2 from google.colab import files  
3 uploaded = files.upload()
```

Essa parte abre a caixinha pra você enviar o arquivo **.xlsx** da sua planilha. O arquivo será salvo no Colab para que a gente possa trabalhar com ele.

```
9  
10 file_name = list(uploaded.keys())[0]  
11 xls = pd.ExcelFile(file_name)  
12
```

- Captura o **nome do arquivo enviado**
- Usa o `pd.ExcelFile()` para **ler todas as abas (sheets)** da planilha

```
12  
13 # ✗ Ignorar a aba "LUCRO"  
14 sheet_names = [name for name in xls.sheet_names if name.upper() != 'LUCRO']  
15
```

Essa linha **cria uma lista com o nome das abas**, mas exclui qualquer aba chamada "LUCRO".

```
15  
16 # 📈 Etapa 3 – Função para contar turnos em todas as colunas  
17 def contar_turnos(sheet_name):  
18     df = pd.read_excel(xls, sheet_name=sheet_name, header=None)  
19     turnos = []  
20  
21     for col in df.columns:  
22         col_data = df[col].dropna().astype(str)  
23         for val in col_data:  
24             val_limpo = val.strip().upper()  
25             if val_limpo in ['MANHÃ', 'TARDE', 'INTEGRAL']:  
26                 turnos.append(val_limpo)  
27  
28     return dict(Counter(turnos))
```

Essa função:

- Lê uma aba da planilha inteira (sem cabeçalho)
- Percorre todas as células de todas as colunas
- Procura por palavras que representem **turnos** (MANHÃ, TARDE, INTEGRAL)
- Conta quantas vezes cada turno aparece usando `Counter` (como um "contador de palavras")

Retorna um dicionário, por exemplo:

```
{'MANHÃ': 5, 'TARDE': 7}
```

```
29  
30 # 🏷️ Etapa 4 – Aplicar função em todas as abas e montar DataFrame  
31 dados_turnos = {}  
32 for aba in sheet_names:  
33     dados_turnos[aba] = contar_turnos(aba)
```

Aqui a gente **roda a função `contar_turnos()` para cada aba** da planilha (exceto "LUCRO") e guarda os resultados em um dicionário:

```
{
```

```
'1ºANO': {'MANHÃ': 6, 'TARDE': 4},
```

```
'2ºANO': {'TARDE': 7},
```

```
...
```

```
}
```

```
34  
35 df_turnos = pd.DataFrame(dados_turnos).fillna(0).astype(int).T
```

Isso transforma o dicionário anterior em uma **tabela organizada com Pandas**, com turmas nas linhas e turnos nas colunas, por exemplo:

Turma MANHÃ TARDE

1ºANO	6	4
2ºANO	0	7

O .fillna(0) preenche valores vazios com 0.

O .astype(int) garante que todos os dados são números.

O .T (transpose) troca linhas por colunas para facilitar o gráfico.

```
36  
37 # ❸ Etapa 5 – Verificar dados coletados  
38 print("DataFrame de turnos por turma:")  
39 print(df_turnos)  
  
40  
41 # ❹ Etapa 6 – Gráfico por turma (empilhado)  
42 if not df_turnos.empty:  
43     ax = df_turnos.plot(  
44         kind='bar',  
45         stacked=True,  
46         figsize=(14, 7),  
47         color=['#4CAF50', '#2196F3', '#FF9800']) # Cores: manhã, tarde, integral  
48     )
```

Esse trecho cria um **gráfico de barras empilhadas**, onde cada barra representa uma turma, e os blocos empilhados são os turnos.

Depois usamos um `for` para adicionar **números sobre os blocos** da barra:

```
55     # Adicionar rótulos com os valores nas barras  
56     for i, turma in enumerate(df_turnos.index):  
57         y_offset = 0  
58         for turno in df_turnos.columns:  
59             valor = df_turnos.loc[turma, turno]  
  
72     # ❺ Etapa 7 – Gráfico total geral por turno  
73     totais_gerais = df_turnos.sum()  
74     cores = {  
75         'MANHÃ': '#2196F3',  
76         'TARDE': '#FF7043',  
77         'INTEGRAL': '#FFD54F'  
78     }
```

Esse comando soma todos os valores por coluna (ou seja, por turno). Exemplo:

MANHÃ: 46

TARDE: 51

```

79     plt.figure(figsize=(8, 6))
80     bars = plt.bar(totais_gerais.index.map(
81         lambda x: 'Matutino' if x == 'MANHÃ' else 'Vespertino' if x == 'TARDE' else x.title()
82     ), totais_gerais.values, color=[cores.get(x, '#90A4AE') for x in totais_gerais.index])
83
84     # Adicionar os valores acima das barras
85     for bar in bars:
86         yval = bar.get_height()
87         plt.text(bar.get_x() + bar.get_width()/2, yval + 1, int(yval), ha='center', fontsize=12, weight='bold')
88
89     plt.title('Total Geral de Alunos por Turno', fontsize=16, weight='bold')
90     plt.ylabel('Número de Alunos', fontsize=12)
91     plt.ylim(0, max(totais_gerais.values) + 10)
92     plt.grid(axis='y', linestyle='--', alpha=0.5)
93
94     plt.tight_layout()

```

Aqui o gráfico mostra **apenas uma barra para cada turno**, com o total de alunos em cada um. Os números são colocados acima de cada barra para facilitar a leitura.

```

95     plt.show()
96
97 else:
98     print("⚠ Nenhum dado numérico encontrado para gerar os gráficos.")
99

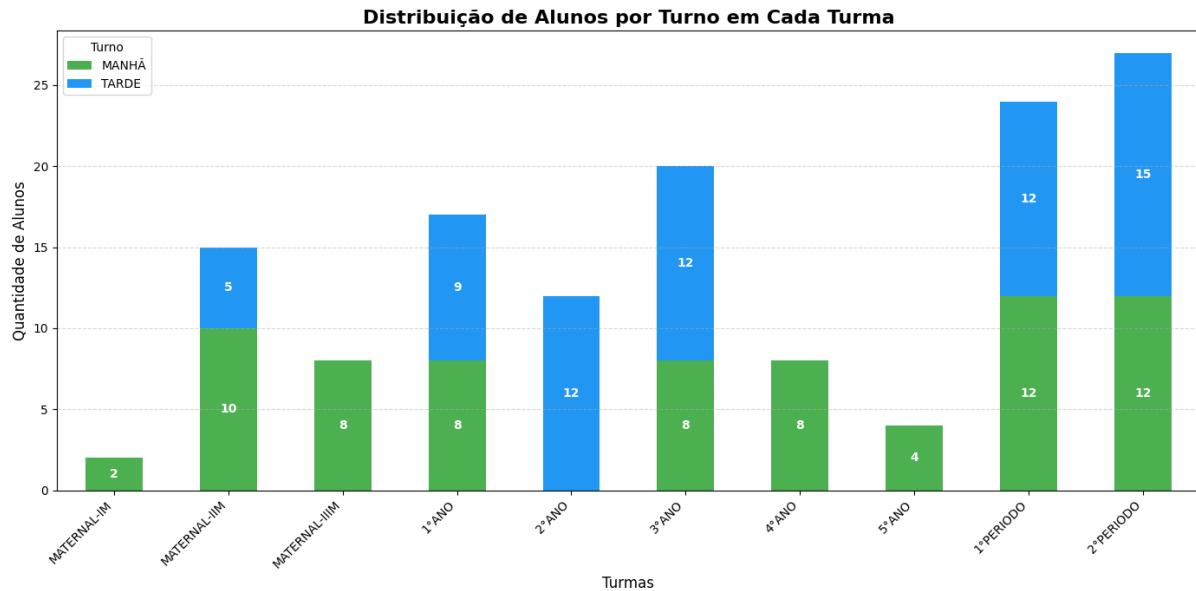
```

Exibe os dois gráficos finais na tela:

1. Por turma e turno
2. Total geral por turno

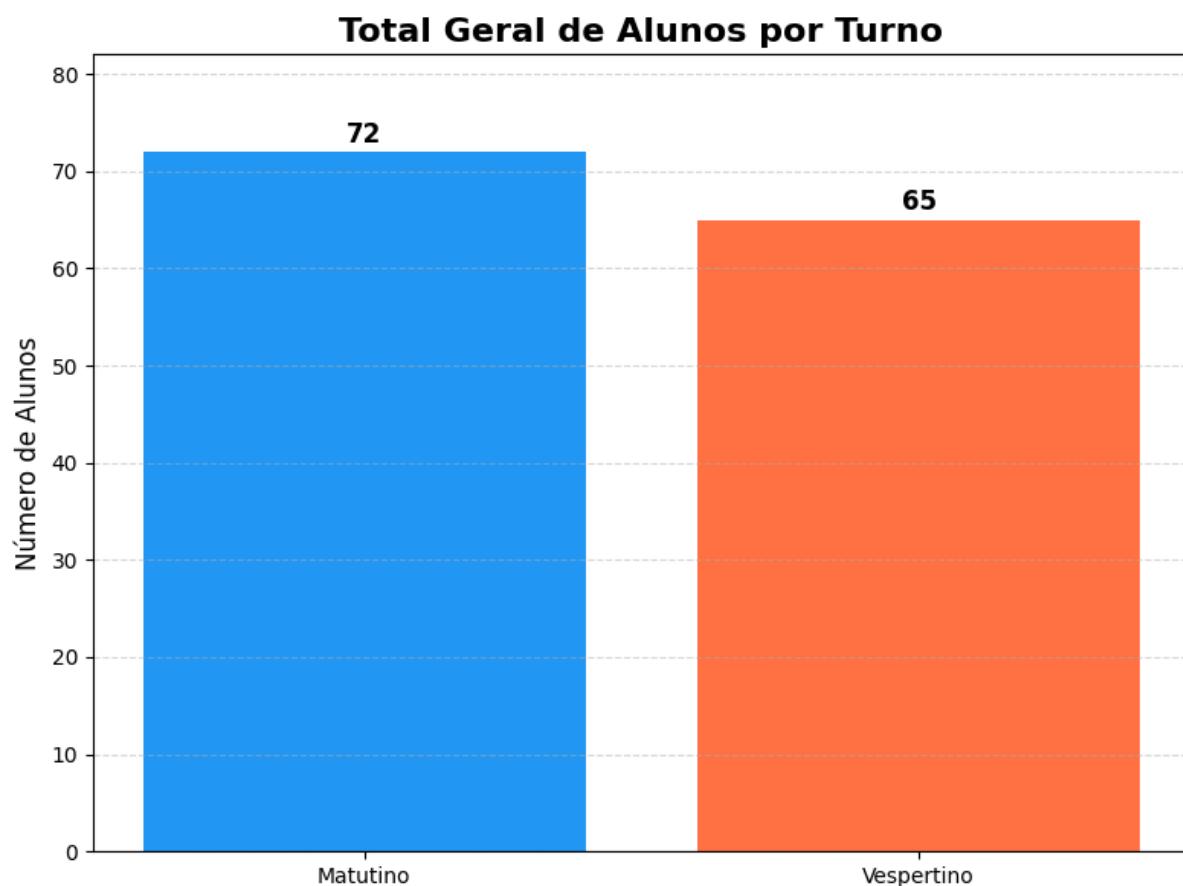
Tendo como resultado 2 gráficos:

1 – “Distribuição de Alunos por Turno em Cada Turma”



2 – “Total Geral de Alunos por Turno”

[com](#)



Link para acesso ao código:

<https://github.com/milyjam/BigGr-fico>

3. ENCERRAMENTO DO PROJETO

3.1. Relato Coletivo:

Durante o desenvolvimento do projeto de extensão, tivemos a oportunidade de aplicar os conhecimentos adquiridos ao longo da disciplina “Tópicos de Big Data em Python” em um contexto real e relevante para a comunidade. A proposta inicial de apoiar o Centro de Ensino Vinho Novo surgiu da necessidade expressa pela própria instituição: a falta de controle e organização nos dados financeiros, especialmente nas informações de entradas e saídas de recursos.

A partir dessa demanda, estruturamos nosso plano de ação, mantendo contato direto com os gestores da escola, o que nos permitiu entender de forma mais sensível e precisa os desafios enfrentados por eles. Esse vínculo com a realidade do parceiro

nos motivou a trabalhar de maneira colaborativa e dedicada, superando obstáculos como a limitação no acesso aos dados e o tempo reduzido para execução.

O trabalho em equipe foi essencial para o bom andamento das atividades. Cada integrante assumiu responsabilidades específicas, contribuindo com suas habilidades para alcançar os objetivos traçados. O uso do Google Colab como plataforma de desenvolvimento nos possibilitou trabalhar de maneira integrada e eficiente, mesmo fora do ambiente físico da faculdade.

Ao final do projeto, conseguimos não apenas organizar os dados financeiros da instituição, mas também gerar relatórios visuais e acessíveis, que facilitaram a compreensão dos gestores sobre a situação financeira de cada unidade escolar. A reação da parte interessada foi extremamente positiva, o que reforçou o impacto social do nosso trabalho.

Concluímos que a tecnologia, quando bem direcionada, tem o poder de transformar a realidade de pequenas instituições e contribuir para uma gestão mais consciente, estratégica e eficaz. O projeto nos proporcionou crescimento acadêmico, técnico e humano, além de fortalecer nosso compromisso com ações que gerem benefícios reais para a sociedade.

3.1.1. Avaliação de reação da parte interessada

A avaliação de reação foi realizada por meio de uma conversa informal com os gestores da escola, onde foram apresentados o relatório final e os gráficos gerados. A diretora Elizabeth Aguiar demonstrou satisfação com o resultado e destacou a importância da organização dos dados para melhorar a gestão das finanças escolares. Foi sugerido, inclusive, que a metodologia utilizada no projeto pudesse ser aplicada futuramente a outras áreas administrativas da instituição. A equipe recebeu retorno positivo quanto à clareza e utilidade dos resultados apresentados.

3.2. Relato de Experiência Individual – JAMILY AGUIAR BEZERRA

3.2.1. CONTEXTUALIZAÇÃO

Durante o desenvolvimento do projeto de extensão, atuei de forma ativa e essencial em diversas etapas. Fui responsável por encontrar a escola parceira — o Centro de Ensino Vinho Novo — e manter contato direto com os diretores para garantir sua participação no projeto. Realizei reuniões presenciais, inclusive para obter a assinatura de autorização para uso dos dados, e conduzi o processo inicial de alinhamento da parceria, o que foi fundamental para o avanço do trabalho. Todas essas ações estão descritas na seção “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.”

3.2.2. METODOLOGIA

O projeto foi desenvolvido em parceria com a escola mencionada, localizada em Manaus-AM, durante os meses de abril a junho de 2025. O público envolvido foi composto pelos diretores e o setor financeiro da instituição. Após o estabelecimento da parceria, coletei os dados necessários — tanto os de saída (gastos) quanto os de entrada (mensalidades) — e, posteriormente, realizei a manipulação dos dados de entrada. Atuei ainda no desenvolvimento do relatório e na construção do slide de apresentação do projeto. Além disso, fui o canal de comunicação direta com a escola, tirando dúvidas e garantindo que eles estivessem sempre cientes do andamento do trabalho.

3.2.3. RESULTADOS E DISCUSSÃO:

A experiência foi desafiadora, mas extremamente enriquecedora. Assumi responsabilidades importantes e me esforcei em dobro para que o projeto tivesse um bom resultado. Apesar das tarefas terem sido divididas entre os membros, nem todos conseguiram cumprir com suas obrigações, o que me levou a assumir também parte do trabalho de outro colega para garantir a continuidade do projeto. Mesmo com essas dificuldades, superei minhas próprias expectativas e pude ver, na prática, como o uso do Big Data pode ajudar uma empresa a entender seu status atual — no nosso caso, na área financeira. Organizando e filtrando os dados de entradas e saídas, conseguimos fornecer à escola uma visão clara e útil para sua gestão.

3.2.4. REFLEXÃO APROFUNDADA

A teoria aprendida em sala se conectou com a prática de forma muito significativa.

Compreendi de maneira mais concreta como conceitos como organização de dados, uso de Python, bibliotecas como Pandas e visualizações com Matplotlib podem ser aplicados para resolver problemas reais. A experiência também me mostrou como é importante o comprometimento de todos os integrantes em um projeto em grupo. O aprendizado foi além da técnica — foi também sobre responsabilidade, comunicação e superação de desafios.

3.2.5. CONSIDERAÇÕES FINAIS

Acredito que o projeto cumpriu sua missão de forma satisfatória, entregando à escola resultados práticos e aplicáveis. Se tivesse que mudar algo, talvez pensasse melhor na divisão de tarefas e no acompanhamento mais próximo das entregas individuais dos membros do grupo. Para trabalhos futuros, vejo potencial em explorar outras áreas da escola com o uso de Big Data, como desempenho de professores e frequência de alunos. O projeto me inspirou a continuar explorando soluções tecnológicas que ajudem instituições a evoluir por meio dos dados.