# Py\_ChatGPT\_Preparation2

No arquivo chamado de **Py\_ChatGPT\_Preparation1**, vimos alguns exemplos de como preparar os dados para uma posterior análise, fizemos uma limpeza, tiramos os dados faltantes, corrigimos algumas palavras, normalizamos alguns dados, tudo conforme indicado pelo *ChatGPT da Open AI*, pois bem, este não é a continuação do primeiro, uma vez que os dois arquivos se confundem. O principal elemento aqui é entender, não só os dados, pois não estamos aqui apenas para ajustá-los, precisamos entender, termos a compreensão do *“por quê”* estamos fazendo tal análise. Eu costumo sempre dizer que todas as análises têm um único objetivo final, que é *“vender”* um determinado produto. Esta é a minha forma resumida do entendimento da necessidade de uma determinada *“Análise de Dados”.*

Porém, estamos fazendo este trabalho em conjunto com o ***ChatGPT***, então pedi ao Chat para nos explicar, segundo o seu entendimento.

**ChatGPT:**

**A análise de dados é usada para extrair informações úteis e valiosas de conjuntos de dados. Os objetivos da análise de dados podem incluir:**

**1 - Descobrir padrões e relações nos dados: A análise de dados pode ser usada para identificar tendências, relações e padrões nos dados que podem ser úteis para a tomada de decisões.**

**2 - Prever resultados futuros: Usando técnicas de análise preditiva, é possível usar dados históricos para prever resultados futuros, como a demanda de um produto ou a probabilidade de uma transação fraudulenta.**

**3 - Tomar decisões: A análise de dados pode fornecer insights valiosos que podem ser usados para tomar decisões estratégicas e operacionais.**

**4 - Identificar problemas: A análise de dados pode ser usada para identificar problemas ou desvios nos processos, como falhas de qualidade ou atrasos de entrega.**

**5 - Monitorar desempenho: A análise de dados pode ser usada para monitorar o desempenho de uma empresa, de um produto ou de um processo, permitindo acompanhar o progresso e identificar áreas de melhoria.**

Repare que em toda a sua resposta, o Chat se refere à *“venda”,* de forma direta ou não.

Dito isto, precisamos entender a natureza desta análise, ou seja, qual o seu objetivo. Precisamos responder uma pergunta (ou várias). Precisamos comparar alguma coisa. Precisamos entender um comportamento, uma tendência, descrever algo... compreender esse dado e torna-lo útil para o seu uso final (...que, normalmente, serve para vender algum produto – 95% dos casos). Vejamos um dos casos mais conhecidos, o Censo Brasileiro, por que ele é feito? Qual a sua utilidade? Onde podemos usa-lo?

**Segundo o ChatGPT:**

**O censo brasileiro é realizado a cada 10 anos pelo Instituto Brasileiro de Geografia e Estatística (IBGE) e tem como objetivo coletar informações demográficas, sociais, econômicas e habitacionais sobre a população e os domicílios do país. As informações coletadas são utilizadas pelo governo, empresas e pesquisadores para planejar políticas públicas, estabelecer metas e alocar recursos. Além disso, o censo também é utilizado para estabelecer os limites de distritos eleitorais e para a distribuição de verbas para municípios e estados.**

**Segundo o IBGE:**

**O Censo Demográfico, levantamento do Instituto Brasileiro de Geografia e Estatística (IBGE), é um levantamento muito importante para o país. Através dele, é possível analisar detalhes sobre como vive a população brasileira: renda, etnia, raça e grupo familiar.(IBGE 2022)**

Repare no final da frase da resposta do Chat, *“distribuição de verba”.* Já no fim da resposta do IBGE, *“renda”.* No fim, esta análise permite ao país investir os recursos públicos para o desenvolvimento de sua população e quando falamos nisso, estamos falando em valores que *“comprarão”* materiais, mão de obra, produtos em gerais, que serão usados ao longo de períodos determinados e, resumidamente, muitas vendas serão realizadas. Logo, por analogia simples, já entendemos que a *“Análise de Dados”* é, sobretudo, entender onde e como vender melhor. Aqui passamos por diversas etapas, entendimentos, caminhos e metodologias. É na Análise de dados que entendemos onde investir tempo e dinheiro para desenvolver determinado produto, seja físico ou abstrato, seja um produto digital, uma mão de obra específica ou um simples *“conhecimento”,* sim, a Análise de Dados pode nos dar insights cujo suas aplicações podem ter elevado valor no mercado, seja esse o conhecimento de movimentação de algumas peças ou a previsão de um acontecimento que poderia causar grande prejuízo ou grande lucratividade.

Não existe uma regra de ouro de como ou onde começar uma Análise de Dados, mas, existem métodos específicos para cada segmento e a compreensão desses métodos e de onde e como aplica-los é que nos torna “Analistas de Dados”.

Não tenho a pretensão de apontar o “caminho das pedras”, estou apenas fazendo um paralelo com o *ChatGPT*. Ainda há muitas observações, muitos caminhos, entendimentos, alguns podem discordar em alguns pontos, outros podem complementar, porém, somente com a experiência e com a compreensão dos objetivos da Análise de Dados é que podemos apontar o caminho que será seguido, a técnica que será utilizada, a tecnologia que será empregada e por aí vai. Aqui, estamos seguindo, não um passo a passo, mas, a compreensão base de como começar na Análise de Dados, de forma simples e objetiva, com o apoio desta nova e fantástica ferramenta, o *ChatGPT*.

**E por falar em nisso, perguntei ao ChatGPT: Qual a principal função do analista de dados?**

**A principal função do analista de dados é transformar dados brutos em informações valiosas e significativas, através de técnicas de análise e interpretação. Isso inclui:**

**1 - Coleta e limpeza de dados: O analista de dados pode ser responsável por coletar dados de diferentes fontes, limpando-os e preparando-os para análise.**

**2 - Análise exploratória: O analista de dados pode usar técnicas estatísticas e de visualização para explorar e entender os dados, descobrindo padrões, tendências e relações.**

**3 - Análise preditiva: O analista de dados pode usar técnicas de aprendizado de máquina para prever resultados futuros, como vendas, demandas ou riscos.**

**4 - Comunicação e interpretação: O analista de dados precisa ser capaz de comunicar os resultados da análise de maneira clara e compreensível para diferentes públicos, incluindo gerentes, outros analistas de dados e executivos.**

**5 - Aplicação: O analista de dados pode trabalhar com outros departamentos para implementar as descobertas da análise de dados em processos operacionais, produtos e estratégias de negócios.**

A resposta chama a atenção para dois aspectos, o primeiro no que se resume no meu entendimento já citado, *“é tudo venda”, “tudo negócio”.* O segundo é o real objetivo desta segunda parte, que é a preparação dos dados, que não é feita apenas com técnicas ou tecnologias de APIs, Frameworks ou linguagens de programação, há uma parte da preparação que é puramente “sangue”, onde o analista precisa intervir de forma direta, uma espécie de limpeza *“no olho”* (OLHANDO), isso mesmo, é preciso fazer uma análise visual no Dataset, para entender o que pode ser transformado, utilizado, descartado, melhorado, entre outras coisas. É fundamental que o analista invista tempo para compreender o que é referente ao objetivo da análise. Quais colunas podem ser eliminadas, quais linhas podem ser renomeadas, o quê, no *Dataset*, pode ter ou não utilidade prática e nos dar a resposta que precisamos. Por exemplo, se vamos analisar o custo benefício de um automóvel em relação ao concorrente, temos que levar em conta alguns fatores, a categoria, o motor, seu consumo, seu peso, etc...será que a data de fabricação interfere nesta análise? Será que a cor do veículo importa? tipo de pneu? -> Aqui vem o quê, na minha opinião, é o primeiro *Skill (habilidade)* do analista, que é a sensibilidade de entender que NÃO entende de tudo, por isso a necessidade de se trabalhar em equipe, de ouvir os profissionais das áreas correlatas, engenheiros, vendedores, gerentes, fabricantes, clientes compradores e o que mais for determinado e só depois de *“olhar”* para o *Dataset*, analisar sua colunas, linhas, sua coleção e entender o que fazer com cada parte, aí sim, podemos continuar. Esta, ao meu ver, é a parte principal da Análise de Dados, a “compreensão dos dados”. Depois de compreender e de, junto com a equipe, decidir quais são as partes mais importantes, os objetivos desta análise, seguiremos para a “preparação dos dados”.

**Fiz esta pergunta ao ChatGPT: Qual a principal parte de uma Análise de dados?**

**Uma análise de dados pode envolver várias etapas e técnicas, dependendo do objetivo específico e do conjunto de dados em questão. No entanto, a parte principal de uma análise de dados é a análise propriamente dita, onde os dados são explorados e interpretados para extrair insights valiosos.**

Depois do entendimento do *“objetivo”,* a análise em si, em toda sua estrutura, tem uma importância singular. Por isso a interação humana é imprescindível, é de suma importância o *“olhar humano”,* aquela compreensão do que serve ou não, do que sai e do que fica, do que é ou não relevante para, só depois, continuarmos com a preparação dos dados.