

Dataset:

https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce?resource=download&select=olist_sellers_dataset.csv

Etapa 2 do CRISP-DM - Entendimento dos Dados

1. Coleta Inicial dos Dados

Quais fontes de dados estão disponíveis para o projeto?

De onde esses dados foram obtidos? Eles são internos, externos, públicos?

Em que formato os dados foram entregues (CSV, SQL, API, Excel, etc.)?

2. Descrição dos Dados

Qual é a dimensão do conjunto de dados (número de linhas e colunas)?

Quais são os tipos de variáveis presentes (numéricas, categóricas, textuais, temporais, etc.)?

Há variáveis que claramente se relacionam com o objetivo do projeto?
Quais?

3. Exploração Inicial

Que estatísticas descritivas podem ser calculadas (médias, medianas, valores máximos/mínimos)?

Existem distribuições interessantes ou inesperadas nas variáveis?

Foi identificada alguma correlação inicial entre variáveis?

4. Qualidade dos Dados

Existem valores ausentes? Se sim, em que proporção e em quais variáveis?

Há dados duplicados ou inconsistentes?

Foram identificados outliers? Eles parecem ser erros ou observações legítimas?

Há variáveis com muitos valores únicos que podem dificultar a análise?

5. Primeiras Hipóteses

Com base nessa exploração inicial, que hipóteses podem ser levantadas sobre os dados?

Quais problemas de qualidade podem impactar diretamente os objetivos definidos na Etapa 1 (Entendimento do Negócio)?

Quais dados adicionais poderiam ser úteis, mas não estão disponíveis?

6. Conclusões e Próximos Passos

Quais etapas de pré-processamento ou limpeza provavelmente serão necessárias antes de avançar para a etapa de Preparação dos Dados?