**Projeto de Análise de Dados: Etapa 2**

**Limpieza e Transformación**

**Objetivo do Projecto**

Esta etapa, buscamos realizar la limpieza y transformación de un dataset real utilizando herramientas como MySQL e Python (pandas, matplotlib, seaborn), estructurando el proceso en formato de un pipeline ETL (Extract, Transform, Load).

**ETAPA 1: Extración dos Dados (Extract)**

* 1. **Fuente de datos:**

Los datos utilizados vienen de banco MySQL con diversas tablas:

**“Productos”, “Facturasdetalles”, “Facturas”.**

* 1. **Consultas SQL:**

Fue realizada una extracción utilizando comandos JOIN para unir los datos de productos, facturasdetalle y facturas. Ejemplo:

A screenshot of a computer program

AI-generated content may be incorrect.

* 1. **Exportación para CSV:**

El resultado da la query fue exportado para .csv e cargado no Jupyter Notebook.

****

**🏋 ETAPA 2: Limpieza dos Dados (Transform)**

**2.1. Verificación de Datos Ausentes**

****

Para sacar todos los valores nulos.

**2.2. Conversión de Datas e Creación de Columna Derivada**

**A computer code on a white background

AI-generated content may be incorrect.**

Para poder sacar cálculos diferenciales.

**2.3. Padronización de Caracteres Especiales**

Corrección de strings corrompidas:

**A screenshot of a computer program

AI-generated content may be incorrect.**

**2.4. Agrupamientos e Contajes**

****

**2.5. Normalización con Z-score**

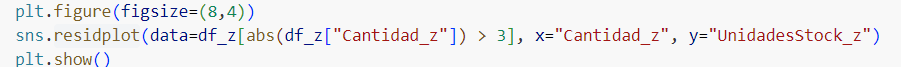
**A computer screen shot of a computer code

AI-generated content may be incorrect.**

**2.6. Detección de Outliers**

****

**2.7. Visualización con Residplot**

****

**ETAPA 3: Cargamento (Load)**

**3.1. Creación de dataset final limpio**

**A white background with red and blue text

AI-generated content may be incorrect.**

**3.2. Exportación**

**A close-up of a number

AI-generated content may be incorrect.**

**Conclusión**

El proceso de ETL, transformo un conjunto de datos relacional em una base lista para análisis. Fueron aplicadas técnicas de limpieza, normalización e detección de outliers. La documentación fue mantenida em Word y Jupyter, facilitando reproductibilidad y clareza.