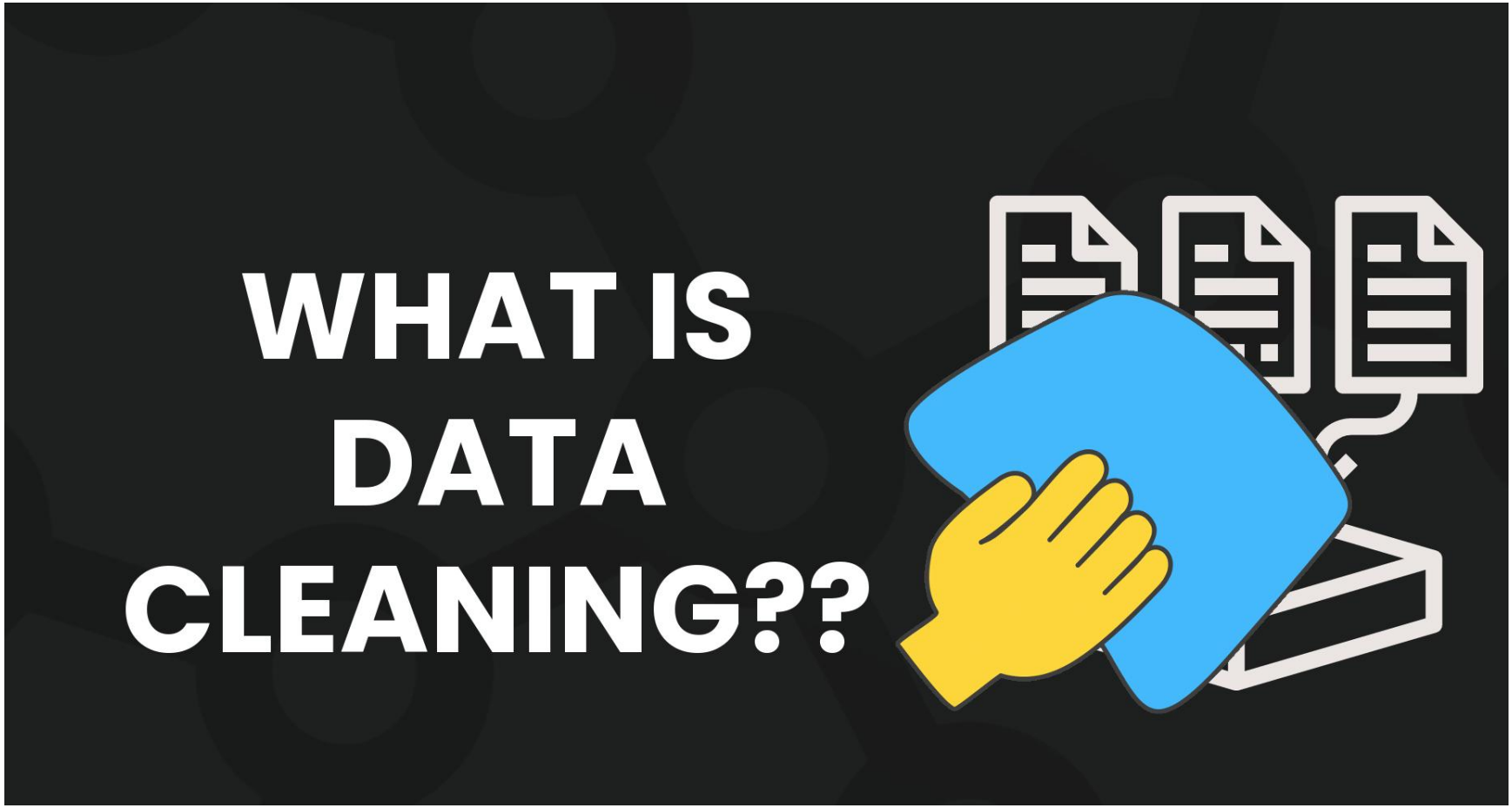
URBAN CYCLES

DATA CLEANING USING EXCEL

URBAN CYCLES

URBAN CYCLES



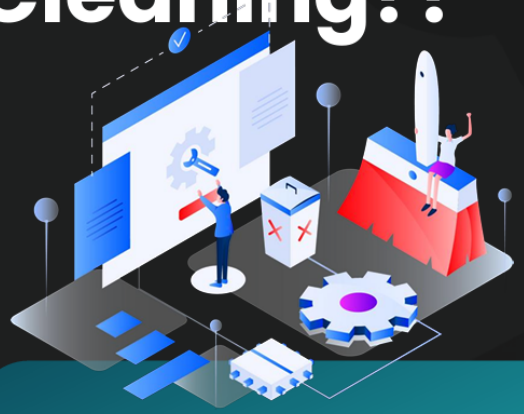
URBAN CYCLES

Data Cleaning is the process of identifying and correcting errors, inconsistencies, and inaccuracies in datasets. The goal of data cleaning is to improve the quality of data by addressing issues such as missing values, duplicate entries, outliers, and formatting errors.

- Handling Inaccuracies and Inconsistencies
- Handling Missing Data Points
- Handling Extreme Values (Outliers)
- **Handling Redundant Data (Duplicates)**

URBAN CYCLES

Why Data Cleaning??



Accuracy of Analysis

Clean data ensures accurate and reliable analysis. If the dataset contains errors, inconsistencies, or missing values, any conclusions drawn from the analysis may be flawed or misleading.

Accurate Prediction

(Machine Learning)

Accurate and reliable models depend on high-quality training data. Data cleaning is a crucial step before building models to ensure that the model is not influenced by noise, outliers, or inconsistencies in the data.

Decision Making

Businesses and organizations rely on data to make informed decisions. If the data used for decision-making is not clean, it can lead to poor choices, misguided strategies, and financial losses.

URBAN CYCLES

1.

A key reason for this project is to **enhance data accuracy** and **reliability**. Clean, well-structured sales data ensures that management can **trust the numbers** they're using to make **important business decisions**. Without this, even the most **well-designed strategies may be based on misleading or incorrect insights**.

2.

This project supports **efficiency in data analysis** and reporting. Cleaning and standardizing the dataset allows for **faster generation of insights**, automated dashboards, and **consistent reporting across** all **branches** and **platforms** – **saving time** and **reducing** human **error**.

3.

Improving data quality lays the foundation for **scaling business intelligence efforts**. As Urban Cycles grows, they'll need **cleaner data for more advanced analytics** like sales forecasting, customer segmentation, and performance tracking. This project sets the stage for deeper and more **impactful data-driven strategies.**

URBAN CYCLES

Data Description

- **Sales_Order #**: Unique identifier for each transaction.
- **Date**: Date the purchase was made.
- **Customer_Age**: Age of the customer at the time of purchase.
- **Age_Group**: Categorized age range of the customer (e.g., 18–25, 26–35).
- **Customer_Gender**: Gender of the customer (e.g., Male, Female).
- **Country**: Country where the sale occurred.
- **State**: Specific state or province of the customer.
- **Product_Category**: Broad classification of the product (e.g., Bikes, Accessories).
- **Sub_Category**: More specific type within the product category (e.g., Mountain Bikes, Helmets).
- **Product_Description**: Detailed name or description of the product sold.
- **Order_Quantity**: Number of units purchased in the transaction.
- **Unit_Cost**: Cost to Urban Cycles per unit of the product.
- **Unit_Price**: Selling price per unit to the customer.
- **Profit**: Total profit from the transaction (Revenue – Cost).
- **Cost**: Total cost of the transaction (Unit Cost × Order Quantity).
- **Revenue**: Total revenue earned (Unit Price × Order Quantity).
- **Sales_Channel**: Method through which the sale was made (e.g., In-store, Online).