**📌 Data Cleaning & Preprocessing Report – Mushroom Dataset**

**📍 Overview**

The **Mushroom Dataset** was cleaned and preprocessed to ensure accuracy, consistency, and usability for further analysis. The dataset contains various features describing mushrooms, including characteristics like **cap shape, odor, gill size, and habitat**. The target variable classifies them as **edible or poisonous**.

**✅ Step 1: Loading the Dataset**

* The dataset was successfully loaded into Pandas.
* It contained **rows and multiple categorical columns**.
* No immediate missing values were detected during initial inspection.

**✅ Step 2: Handling Duplicates**

* Checked for duplicate rows using df.duplicated().sum().
* **Duplicates were found and removed**, ensuring unique records.

**✅ Step 3: Standardizing Column Names**

* Column names were **converted to lowercase** and **spaces were replaced with underscores** for consistency.
* Example: "Cap Shape" → "cap\_shape"

**✅ Step 4: Converting Data Types**

* Most columns in the dataset were categorical.
* Data types were verified, and **numeric conversions were not necessary**.
* If needed, categorical features could be encoded using Label Encoding or One-Hot Encoding for Machine Learning.

**✅ Step 5: Handling Missing Values**

* Checked for missing values using df.isnull().sum().
* **No missing values were found**, so no further imputation was required.

**✅ Final Outcome**

* The dataset is now **clean, structured, and ready for analysis**.
* It can be used for **classification tasks**, such as predicting whether a mushroom is **edible or poisonous**.
* Potential next steps include **EDA (Exploratory Data Analysis)** and **Machine Learning modeling**.