**TECHNICAL REPORT:**

**PRODUCT DATASET CLEANING, SHORT TITLE OPTIMIZATION & DATA INSIGHTS**

1. **Introduction**

This report outlines the data cleaning process and title optimization for a product dataset. The objective was to enhance data quality by addressing missing values, duplicates, and inconsistencies while creating a ***short\_title*** feature for improved SEO and readability.

1. **Data Cleaning**
   1. **Issues Identified:**
2. **Missing Values:** Found in bullet\_points and description.
3. **Duplicate Records:** Several products were duplicated.
4. **Inconsistent Formats:**
   1. Mixed case in title column.
   2. productlength had non-numeric values.
   3. Irregular spacing and special characters.

**2.2 Steps Taken:**

1. **Handled Missing Values:**
   1. Filled bullet\_points & description with "No Information Available" where missing.
2. **Removed Duplicates:**
   1. Used drop\_duplicates() to ensure uniqueness.
3. **Standardized Formats:**
   1. Converted title to lowercase.
   2. Cleaned productlength to retain only numeric values.

**3.0 Short Title Creation**

**3.1 Objective:**

Generate a concise **short\_title** to improve **SEO** and readability while retaining key product attributes.

**3.2 Methodology:**

1. Extracted **key product details** from title.
2. Removed **stop words** (e.g., "includes," "set of," "features").
3. Ensured character limit of **30–50 characters**.

**Examples:**

| **Original Title** | **Short Title** |
| --- | --- |
| \*Tulip Flowers Blackout Curtain for Door, Window & Room | Eyelets & Tie Back |
| *Marks & Spencer Girls' Pyjama Sets T86\_2561C\_Navy Mix\_9-10Y* | **Girls' Navy Pyjama Set - 9-10Y** |

**4.0 Clean Dataset Overview**

After cleaning, the dataset had:  
✅ **No missing values** in critical columns.  
✅ **Unique records**, with duplicates removed.  
✅ **Formatted text** with consistent case and spacing.  
✅ **SEO-friendly titles** added as short\_title.

**5.0 Visualizations & Insights**

1. **Bar chart** of missing values before vs. after cleaning.
2. **Sample of original vs. cleaned dataset.**

**6.0 Conclusion**

This task successfully improved data quality and created an SEO-optimized title feature. The dataset is now ready for marketing analysis with clean, structured, and readable information.