

SPAM DETECTION DATA CLEANING REPORT

1. Handling Missing Values:
 - Dropped 'Unnamed' columns.
 - Missing values count:

target 0

text 0
2. Fixing Data Types:
 - Converted columns to appropriate formats.
 - Current Types:

target object

text object

dtype: object
3. Detecting and Treating Outliers:
 - Metric: Message Length (IQR Method)
 - Outliers detected: 68
 - Action: Capped lengths at 250.00.
4. Removing Duplicate Records:
 - Initial records: 5572
 - Duplicates removed: 409
 - Final records: 5163
5. Handling Categorical Variables:
 - Method: Label Mapping
 - Mapping: ham -> 0, spam -> 1

CLEANED DATA PREVIEW:

	target	text	message_len	target_encoded
0	ham	Go until jurong point, crazy.. Available only in bugis n great world la e buffet... Cine there got amore wat...	111.0	0
1	ham	Ok lar... Joking wif u oni...	29.0	0
2	spam	Free entry in 2 a wkly comp to win FA Cup final tkts 21st May 2005. Text FA to 87121 to receive entry question(std txt rate)T&C's apply 08452810075over18's	155.0	1
3	ham	U dun say so early hor... U c already then say...	49.0	0
4	ham	Nah I don't think he goes to usf, he lives around here though	61.0	0