

## Emoji Handler Class

```
1  ## Author: Ashantha Rosary
2  import re
3  from pyspark.sql.functions import udf
4  from pyspark.sql.types import StringType
5
6  class EmojiHandler:
7      def __init__(self, emoji_dict):
8          """
9          Initialize with the emoji dictionary.
10         """
11         self.emoji_dict = emoji_dict
12
13     def replace_emojis(self, text):
14         """
15         Replace emojis in the text using the provided dictionary and remove non-mapped
16         emojis.
17         """
18         if text:
19             # Replace emojis using the dictionary
20             for emoji, replacement in self.emoji_dict.items():
21                 text = text.replace(emoji, replacement)
22             # Remove any remaining emojis that were not in the dictionary
23             text = re.sub(r'^[\x00-\x7F]+', '', text) # This regex removes non-ASCII
24             characters (i.e., emojis)
25             return text
26
27     def get_replace_emojis_udf(self):
28         """
29         Return a UDF to be applied in a PySpark DataFrame for emoji mapping and removal.
30         """
31         return udf(lambda text: self.replace_emojis(text), StringType())
```