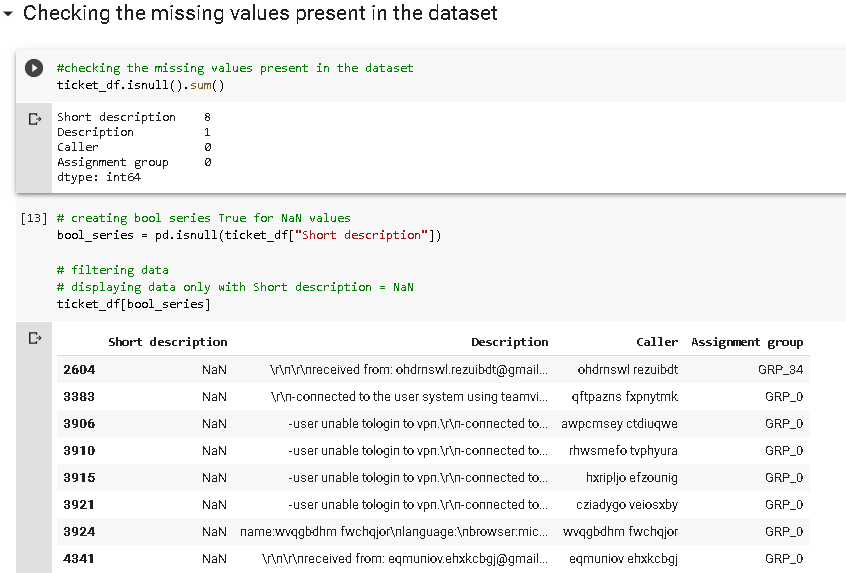
Great Job team:-)))

1. **Dealing with missing values -**

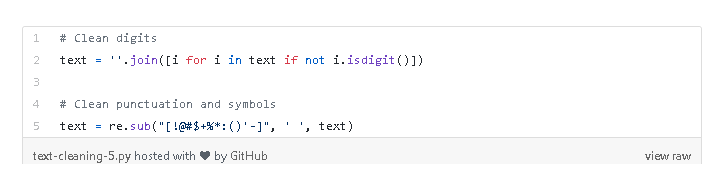


**Suggestions**:

1. Since 8 short descriptions are missing - fill them with the first 10-15 words from the Description field.
2. Since 1 description is missing - fill it with the short description field text OR drop that entry.
3. **Dealing with Junk values and data cleaning:**

**Suggestions:**

1. The re. sub() function in the re module can be used to replace substrings.



1. We require to use text-pre processing techniques.
2. Removing punctuations.
3. Remove stop words
4. Lemmatization or Stemming
5. Tokenization
6. Encoding the ‘Caller’ data since visually it is intimidating

Without this, visualizing data wont be efficient.

I suppose we can consider importing the NLTK library to do the same.

The examples from the course cover this part.

**Readings:**

[**https://www.kdnuggets.com/2018/03/text-data-preprocessing-walkthrough-python.html**](https://www.kdnuggets.com/2018/03/text-data-preprocessing-walkthrough-python.html)

[**https://medium.com/@dobko\_m/nlp-text-data-cleaning-and-preprocessing-ea3ffe0406c1**](https://medium.com/@dobko_m/nlp-text-data-cleaning-and-preprocessing-ea3ffe0406c1)

[**https://towardsdatascience.com/nlp-for-beginners-cleaning-preprocessing-text-data-ae8e306bef0f**](https://towardsdatascience.com/nlp-for-beginners-cleaning-preprocessing-text-data-ae8e306bef0f)