

Data Annotation Class

```
1  ## Author: Wong Yee En
2  from pyspark.sql.functions import udf, col
3  from pyspark.sql.types import IntegerType
4
5  class DataAnnotator:
6      def __init__(self, spark_session):
7          self.spark = spark_session
8
9      @staticmethod
10     def annotate_sentiment(star_count):
11         if star_count in [4, 5]:
12             return 2
13         elif star_count == 3:
14             return 1
15         else:
16             return 0
17
18     def add_sentiment_column(self, df):
19         annotate_sentiment_udf = udf(self.annotate_sentiment, IntegerType())
20         df_cleaned = df.withColumn('Sentiment', annotate_sentiment_udf(col('StarCount')))
21         return df_cleaned
22
```