

202014068_이민준

```

import pandas as pd
import re

positive_words = ['좋다', '기분', '행복하다', '맛있는', '즐겁다', '사랑해', '감사', '멋지다', '예쁘다']
negative_words = ['싫다', '별로', '지루했다', '실망했다', '어렵다', '짜증나', '화났다', '우울하다', '못생겼다']

class Aurora3:
    def __init__(self):
        self.pos_dict = set(positive_words)
        self.neg_dict = set(negative_words)

    def tokenize(self, sentence):
        return re.findall(r'\b\w+\b', sentence)

    def analyze(self, sentence):
        tokens = self.tokenize(sentence)
        pos_count = sum(1 for token in tokens if token in self.pos_dict)
        neg_count = sum(1 for token in tokens if token in self.neg_dict)
        total = pos_count + neg_count

        score = (pos_count - neg_count) / total if total != 0 else 0
        ratio = total / len(tokens) if tokens else 0

        return score, ratio

    def get_df(self, df):
        scores, ratios = [], []
        for sentence in df['sentence']:
            score, ratio = self.analyze(sentence)
            scores.append(score)
            ratios.append(ratio)
        df['score'] = scores
        df['ratio'] = ratios
        return df

df = pd.read_csv('text_data_1.txt', header=None, names=['sentence'])

analyzer = Aurora3()
df = analyzer.get_df(df)

print(df)

```

```

↻
      sentence  score  ratio
0  오늘 날씨가 너무 좋아서 기분이 좋다.    1.0  0.166667
1  이 영화는 정말 별로였어. 너무 지루했다.   -1.0  0.166667
2      맛있는 음식을 먹어서 행복하다.    1.0  0.500000
3  이 책은 이해하기 어려워서 실망했다.   -1.0  0.200000

```