

```
#####
```

```
## This file was used to remove char ; from reviews  
## which forces csv to read additinal columns
```

```
#####
```

```
import pandas as pd  
import json
```

```
data_file = pd.read_csv('yelp_academic_dataset_review.csv', keep_default_na=False)
```

```
def applyadj(x):
```

```
    if x[3] != '':#not pd.isna(x[3]):  
        if x[4] != '':#not pd.isna(x[4]):  
            return x[0]+x[1]+x[2]  
        else:  
            return x[0]+x[1]
```

```
    else:  
        return x[0]
```

```
def applyadj2(x):
```

```
    if x[3] != '':#not pd.isna(x[3]):  
        if x[4] != '':#not pd.isna(x[4]):  
            return x[3] #pd.DataFrame([ x[0], x[3], x[4]])  
        else:  
            return x[2]
```

```
    else:  
        return x[1]
```

```
def applyadj3(x):
```

```
    if x[3] != '':#not pd.isna(x[3]):  
        if x[4] != '':#not pd.isna(x[4]):  
            return x[4] #pd.DataFrame([ x[0], x[3], x[4]])  
        else:  
            return x[3]
```

```
    else:  
        return x[2]
```

```
def applystr(x):
```

```
    if isinstance(x,str):  
        return x.replace('\n', r' ').replace(';',' r:')  
    else: return x
```

```
data_file['text']= data_file[['text','useful', 'user_id', 'Unnamed: 9', 'Unnamed: 10']].a  
pply(applyadj, axis=1)  
data_file['useful']= data_file[['text','useful', 'user_id', 'Unnamed: 9', 'Unnamed: 10']  
.apply(applyadj2, axis=1)  
data_file['user_id']= data_file[['text','useful', 'user_id', 'Unnamed: 9', 'Unnamed: 10'  
].apply(applyadj3, axis=1)
```

```
data_file['text'] = data_file['text'].apply(lambda x: applystr(x))  
data_file.to_csv('yelp_academic_dataset_review_without_newline_text.csv')
```