```python
import pandas as pd

# Replace 'rt.reviews.tsv' with the name of your TSV file
df = pd.read_csv('rt.reviews.tsv', delimiter='\t', encoding='ISO-8859-1')

# Display the contents of the file
df.head(10)
```

1 to 10 of 10 entries   Filter

index	id	review	rating	fresh	critic	top_critic	publisher	date
0	3	A distinctly gallows take on contemporary financial mores, as one absurdly rich man's limo ride across town for a haircut functions as a state-of-the-nation discourse.	3/5	fresh	PJ Nabarro	0	Patrick Nabarro	November 10, 2018
1	3	It's an allegory in search of a meaning that never arrives…It's just old-fashioned bad storytelling.	NaN	rotten	Annalee Newitz	0	io9.com	May 23, 2018
2	3	… life lived in a bubble in financial dealings and digital communications and brief face-to-face conversations and sexual intermissions in a space shuttle of a limousine creeping through the gridlock of an anonymous New York City.	NaN	fresh	Sean Axmaker	0	Stream on Demand	January 4, 2018
		Continuing along a line introduced in last year's						

```python
#drop any colums that are not necessary
df = df.drop(['critic', 'top_critic', 'publisher'], axis=1)
df.head()
```

1 to 5 of 5 entries   Filter

index	id	review	rating	fresh	date
0	3	A distinctly gallows take on contemporary financial mores, as one absurdly rich man's limo ride across town for a haircut functions as a state-of-the-nation discourse.	3/5	fresh	November 10, 2018
1	3	It's an allegory in search of a meaning that never arrives…It's just old-fashioned bad storytelling.	NaN	rotten	May 23, 2018
2	3	… life lived in a bubble in financial dealings and digital communications and brief face-to-face conversations and sexual intermissions in a space shuttle of a limousine creeping through the gridlock of an anonymous New York City.	NaN	fresh	January 4, 2018
3	3	Continuing along a line introduced in last year's "A Dangerous Method", David Cronenberg pushes his cinema towards a talky abstraction in his uncanny, perversely funny and frighteningly insular adaptation of Don DeLillo, "Cosmopolis".	NaN	fresh	November 16, 2017

```python
#convert the date column to date time data type
df['date'] = pd.to_datetime(df['date'], format='%B %d, %Y')
df.head()
```

```
#clean the rating column
df['rating'] = pd.to_numeric(df['rating'].str.replace(r'[^\d.]+', ''), errors='coerce')
```

```
<ipython-input-4-3ad2ee5a6e03>:2: FutureWarning: The default value of regex will change from True to False in a future version.
  df['rating'] = pd.to_numeric(df['rating'].str.replace(r'[^\d.]+', ''), errors='coerce')
```

life lived in a bubble in financial dealings and digital

```
df.head()
```

1 to 5 of 5 entries   Filter   ▢   ❓

index	id	review	rating	fresh	date
0	3	A distinctly gallows take on contemporary financial mores, as one absurdly rich man's limo ride across town for a haircut functions as a state-of-the-nation discourse.	3/5	fresh	2018-11-10 00:00:00
6	3	Quickly grows repetitive and tiresome, meandering toward messages that land softly, producing little impact.	C	rotten	2013-07-17 00:00:00
7	3	Cronenberg is not a director to be daunted by a scenario in which the antihero spends most of his time in a stretch limo. Turning it into a film that interests anyone … is another matter	2/5	rotten	2013-04-21 00:00:00
11	3	While not one of Cronenberg's stronger films, this anti-capitalist adaptation still merits attention.	B-	fresh	2013-02-03 00:00:00
12	3	Robert Pattinson works mighty hard to make Cosmopolis more than just an erudite slap at modern capitalism. The Twilight heartthrob ultimately fails to rescue a meandering story hitting stale versions of the same	2/4	rotten	2013-01-15 00:00:00

```
#drop any missing raws
df = df.dropna()
df.head()
```

1 to 5 of 5 entries   Filter   ▢   ❓

index	id	review	rating	fresh	date
0	3	A distinctly gallows take on contemporary financial mores, as one absurdly rich man's limo ride across town for a haircut functions as a state-of-the-nation discourse.	3/5	fresh	2018-11-10 00:00:00
6	3	Quickly grows repetitive and tiresome, meandering toward messages that land softly, producing little impact.	C	rotten	2013-07-17 00:00:00
7	3	Cronenberg is not a director to be daunted by a scenario in which the antihero spends most of his time in a stretch limo. Turning it into a film that interests anyone … is another matter	2/5	rotten	2013-04-21 00:00:00
11	3	While not one of Cronenberg's stronger films, this anti-capitalist adaptation still merits attention.	B-	fresh	2013-02-03 00:00:00
12	3	Robert Pattinson works mighty hard to make Cosmopolis more than just an erudite slap at modern capitalism. The Twilight heartthrob ultimately fails to rescue a meandering story hitting stale versions of the same	2/4	rotten	2013-01-15 00:00:00

```
# Convert the 'review' column to a numerical data type
df['review'] = pd.to_numeric(df['review'], errors='coerce')

# Calculate the correlation between the 'rating' and 'review' columns
corr = df['rating'].corr(df['review'], min_periods=1)

# Print the correlation coefficient
print("Correlation coefficient between 'rating' and 'review':", corr)
```

```
Correlation coefficient between 'rating' and 'review': nan
```

After testing the correlation between reviews and rating there was no correlation thats why we got the "NAN" output. where 'NAN' occurs if the correlation coefficient between the 'rating' and 'review' columns is NaN, it may indicate that there is no linear relationship between these variables in the dataset. It could also mean that the 'review' column contains missing or invalid values, which are being converted to NaN by the pd.to_numeric() function. In this case, we had dropped the missing values hence making sure the error doesnt occur.

✓ 0s   completed at 7:07 PM   ● ✕