

```
In [56]: import numpy as np
import pandas as pd
import warnings
import cv2
```

```
In [57]: warnings.filterwarnings('ignore')
```

```
In [58]: df=pd.read_csv('u.data',sep="\t")
```

```
In [59]: df.head()
```

Out[59]:

	196	242	3	881250949
0	186	302	3	891717742
1	22	377	1	878887116
2	244	51	2	880606923
3	166	346	1	886397596
4	298	474	4	884182806

```
In [60]: columns_name=['user_id','item_id','rating','timestamp']
df=pd.read_csv('u.data',sep="\t",names=columns_name)
```

```
In [61]: df.head()
```

Out[61]:

	user_id	item_id	rating	timestamp
0	196	242	3	881250949
1	186	302	3	891717742
2	22	377	1	878887116
3	244	51	2	880606923
4	166	346	1	886397596

```
In [62]: df.shape
```

Out[62]: (100000, 4)

```
In [63]: df['user_id']
```

Out[63]:

0	196
1	186
2	22
3	244
4	166
...	
99995	880
99996	716
99997	276
99998	13
99999	12

Name: user_id, Length: 100000, dtype: int64

```
In [64]: df["user_id"]
```

```
Out[64]: 0      196
          1      186
          2       22
          3     244
          4     166
          ...
          99995   880
          99996   716
          99997   276
          99998    13
          99999    12
          Name: user_id, Length: 100000, dtype: int64
```

```
In [65]: df['user_id'].nunique()
```

```
Out[65]: 943
```

```
In [66]: df['item_id'].nunique()
```

```
Out[66]: 1682
```

```
In [ ]:
```

```
In [78]: import numpy as np
          import pandas as pd
          import warnings
          import cv2
```

```
In [79]: warnings.filterwarnings('ignore')
```

```
In [83]: movies_title=pd.read_csv('u.item',sep="\\",header=None)
```

UnicodeDecodeError

Traceback (most recent call last)

<ipython-input-83-bee59de46505> in <module>

----> 1 movies_title=pd.read_csv('u.item',sep="\\",header=None)

```
c:\users\sony\appdata\local\programs\python\python38\lib\site-packages\pandas\io\parser
s.py in read_csv(filepath_or_buffer, sep, delimiter, header, names, index_col, usecols, squ
eeze, prefix, mangle_dupe_cols, dtype, engine, converters, true_values, false_values, ski
pinitialspace, skiprows, skipfooter, nrows, na_values, keep_default_na, na_filter, verbos
e, skip_blank_lines, parse_dates, infer_datetime_format, keep_date_col, date_parser, dayf
irst, cache_dates, iterator, chunksize, compression, thousands, decimal, lineterminator,
quotechar, quoting, doublequote, escapechar, comment, encoding, dialect, error_bad_lin
es, warn_bad_lines, delim_whitespace, low_memory, memory_map, float_precision, stora
ge_options)
```

```
603     kwds.update(kwds_defaults)
```

```
604
```

```
--> 605     return _read(filepath_or_buffer, kwds)
```

```
606
```

```
607
```

```
c:\users\sony\appdata\local\programs\python\python38\lib\site-packages\pandas\io\parser
s.py in _read(filepath_or_buffer, kwds)
```

```
461
```

```
462     with parser:
```

```
--> 463         return parser.read(nrows)
```

```
464
```

```
465
```

```
c:\users\sony\appdata\local\programs\python\python38\lib\site-packages\pandas\io\parser
s.py in read(self, nrows)
```

```
1050     def read(self, nrows=None):
```

```
1051         nrows = validate_integer("nrows", nrows)
```

```
-> 1052         index, columns, col_dict = self._engine.read(nrows)
```

```
1053
```

```
1054         if index is None:
```

```
c:\users\sony\appdata\local\programs\python\python38\lib\site-packages\pandas\io\parser
s.py in read(self, rows)
```

```
2454     def read(self, rows=None):
```

```
2455         try:
```

```
-> 2456             content = self._get_lines(rows)
```

```
2457         except StopIteration:
```

```
2458             if self._first_chunk:
```

```
c:\users\sony\appdata\local\programs\python\python38\lib\site-packages\pandas\io\parser
s.py in _get_lines(self, rows)
```

```
3256
```

```
3257         while True:
```

```
-> 3258             new_row = self._next_iter_line(row_num=self.pos + rows + 1)
```

```
3259             rows += 1
```

```
3260
```

```
c:\users\sony\appdata\local\programs\python\python38\lib\site-packages\pandas\io\parser
s.py in _next_iter_line(self, row_num)
```

```
2957         # assert for mypy, data is Iterator[str] or None, would error in next
```

```
2958         assert self.data is not None
```

```
-> 2959         return next(self.data)
```

```
2960     except csv.Error as e:
```

```
2961         if self.warn_bad_lines or self.error_bad_lines:
```

```
c:\users\sony\appdata\local\programs\python\python38\lib\site-packages\pandas\io\parser
s.py in _read()
    2441         yield pat.split(line.strip())
    2442
-> 2443     for line in f:
    2444         yield pat.split(line.strip())
    2445
```

```
c:\users\sony\appdata\local\programs\python\python38\lib\codecs.py in decode(self, input,
final)
    320     # decode input (taking the buffer into account)
    321     data = self.buffer + input
-> 322     (result, consumed) = self._buffer_decode(data, self.errors, final)
    323     # keep undecoded input until the next call
    324     self.buffer = data[consumed:]
```

UnicodeDecodeError: 'utf-8' codec can't decode byte 0xe9 in position 2892: invalid continuation byte

In []:

In []:

In []: