```
In [1]:  %matplotlib inline
         import pandas as pd
```

```
In [2]:  from IPython.core.display import HTML
         css = open('style-table.css').read() + open('style-notebook.css').read()
         HTML('<style>{}</style>'.format(css))
```

Out[2]:

```
In [4]:  cast = pd.read_csv("C:\\Users\\sujoydutta\\Downloads\\cast.csv")
         cast.head()
```

Out[4]:

	title	year	name	type	character	n
0	Suuri illusioni	1985	Homo $	actor	Guests	22.0
1	Gangsta Rap: The Glockumentary	2007	Too $hort	actor	Himself	NaN
2	Menace II Society	1993	Too $hort	actor	Lew-Loc	27.0
3	Porndogs: The Adventures of Sadie	2009	Too $hort	actor	Bosco	3.0
4	Stop Pepper Palmer	2014	Too $hort	actor	Himself	NaN

```
In [8]:  cast.info()

         <class 'pandas.core.frame.DataFrame'>
         RangeIndex: 3311647 entries, 0 to 3311646
         Data columns (total 6 columns):
          #   Column     Dtype
         ---  ------     -----
          0   title      object
          1   year       int64
          2   name       object
          3   type       object
          4   character  object
          5   n          float64
         dtypes: float64(1), int64(1), object(4)
         memory usage: 151.6+ MB
```

How many movies are listed in the titles dataframe?

```
In [5]:  moviescount=len(cast.title.unique())
         moviescount
```

Out[5]: 159488

What are the earliest two films listed in the titles dataframe?

```
In [15]:  cast.title.head(2)
```

Out[15]:
```
0                   Suuri illusioni
1     Gangsta Rap: The Glockumentary
Name: title, dtype: object
```

How many movies have the single-word title "Hamlet"?

```
In [16]:  cast.title.str.contains('Hamlet').sum()
```

Out[16]: 890

How many movies are titled "North by Northwest"?

```
In [17]:  cast.title.str.contains('North by Northwest').sum()
```

Out[17]: 111

When was the first movie titled "Hamlet" made?

```
In [22]:  hamletmovies=cast[cast['title'].str.contains('Hamlet', na=False, case=False)]
          hamletmovies=hamletmovies.sort_values(by='year')
          earliestmovie=hamletmovies.year.min()
          earliestmovie
```

Out[22]: 1910

List all of the "Treasure Island" movies from earliest to most recent.

```
In [23]:  treasureislandmovies=cast[cast['title'].str.contains('Treasure Island', na=False, case=Fal
          treasureislandmovies=treasureislandmovies.sort_values(by='year')
          treasureislandmovies
```

Out[23]:

	title	year	name	type	character	n
2438589	Treasure Island	1918	Virginia Lee Corbin	actress	Louise Trelawney	2.0
1188122	Treasure Island	1918	Elmo Lincoln	actor	Prologue Player	NaN
1790003	Treasure Island	1918	Lewis Sargent	actor	Ben Gunn	5.0
311157	Treasure Island	1918	Francis Carpenter	actor	Jim Hawkins	1.0
837168	Treasure Island	1918	Edwin Harley	actor	Prologue Player	NaN
...	...	...	...	...	...	...
2439295	National Theatre Live: Treasure Island	2015	Claire-Louise Cordwell	actress	Joan the Goat	NaN
2181038	National Theatre Live: Treasure Island	2015	Roger (XIII) Wilson	actor	Shanty Singer	NaN
1548658	National Theatre Live: Treasure Island	2015	Alastair Parker	actor	Killigrew the Kind	NaN
2514237	National Theatre Live: Treasure Island	2015	Heather Dutton	actress	Red Ruth	NaN
411447	National Theatre Live: Treasure Island	2015	Daniel Coonan	actor	Black Dog	NaN

524 rows × 6 columns

How many movies were made in the year 1950?

```
In [28]:  fiftiesmovies = cast[(cast['year'] == 1950)]

          len(fiftiesmovies.title.unique())

Out[28]:  877
```

How many movies were made in the year 1960?

```
In [29]:  sixtiesmovies = cast[(cast['year'] == 1960)]

          len(sixtiesmovies.title.unique())

Out[29]:  1034
```

How many movies were made from 1950 through 1959?

```
In [30]:  vintagemovies = cast[(cast['year'] >= 1950)
                               & (cast['year'] < 1960)]

          len(vintagemovies.title.unique())

Out[30]:  9681
```

In what years has a movie titled "Batman" been released?

```
In [31]:  batmanmovies=cast[cast['title'].str.contains('Batman', na=False, case=False)]

          batmanmovies.year.unique()

Out[31]:  array([1967, 2016, 1966, 2010, 2005, 1997, 2015, 2017, 1949, 1995, 1943,
                 1989, 2012, 1992, 1993, 2014, 1991, 1973, 1965, 1964], dtype=int64)
```

How many roles were there in the movie "Inception"?

```
In [33]:  inceptionmovies=cast[cast['title'].str.contains('Inception', na=False, case=False)]
          inceptionroles=inceptionmovies.character.unique()
          len(inceptionroles)

Out[33]:  53
```

How many roles in the movie "Inception" are NOT ranked by an "n" value?

```
In [39]:  import numpy as np
          inceptionproper=inceptionmovies[(inceptionmovies['n'] != np.nan)]
          properroles=inceptionproper.character.unique()
          len(properroles)

Out[39]:  53
```

But how many roles in the movie "Inception" did receive an "n" value?

```
In [37]:  inceptionnan=inceptionmovies[(inceptionmovies['n'] == np.nan)]
          nanroles=inceptionnan.character.unique()
          len(nanroles)
```

Out[37]: 0

Display the cast of "North by Northwest" in their correct "n"-value order, ignoring roles that did not earn a numeric "n" value.

```
In [46]:  nbynwmovies = cast[cast['title'].str.contains('North by Northwest', na=False, case=False)]

          nbynwmovies = nbynwmovies[nbynwmovies['n'].notna()]

          nbynwmovies_sorted = nbynwmovies.sort_values(by='n')

          nbynwmovies_sorted=nbynwmovies_sorted[['character','n']]
          nbynwmovies_sorted
```

Out[46]:

	character	n
773496	Roger O. Thornhill	1.0
3084579	Eve Kendall	2.0
1292652	Phillip Vandamm	3.0
2776446	Clara Thornhill	4.0
315343	The Professor	5.0
2685171	Mrs. Townsend	6.0
1504777	Lester Townsend	7.0
1130557	Leonard	8.0
2167754	Valerian	9.0
1607029	Victor Larrabee	10.0
590702	Licht	11.0
2033529	Auctioneer	12.0
411247	Dr. Cross	13.0
1339004	Sergeant Flamm	14.0
180985	Captain Junket	15.0
1228229	Charley - Chicago Policeman	16.0

Display the entire cast, in "n"-order, of the 1972 film "Sleuth".

```
In [51]:  sleuthwmovies = cast[(cast['title'].str.contains('sleuth', na=False, case=False)) & (cast[
```

```
sleuthwmovies = sleuthwmovies[sleuthwmovies['n'].notna()]

sleuthwmovies_sorted = sleuthwmovies.sort_values(by='n')

sleuthwmovies_sorted = sleuthwmovies_sorted[['character', 'n']]

print(sleuthwmovies_sorted)
```

```
                          character    n
1513570                 Andrew Wyke  1.0
288502                  Milo Tindle  2.0
330750             Inspector Doppler  3.0
1299944  Detective Sergeant Tarrant  4.0
2406762             Marguerite Wyke  5.0
1285278       Police Constable Higgs  6.0
```

Now display the entire cast, in "n"-order, of the 2007 version of "Sleuth".

In [52]:
```
sleuthwmovies = cast[(cast['title'].str.contains('sleuth', na=False, case=False)) & (cast[

sleuthwmovies = sleuthwmovies[sleuthwmovies['n'].notna()]

sleuthwmovies_sorted = sleuthwmovies.sort_values(by='n')

sleuthwmovies_sorted = sleuthwmovies_sorted[['character', 'n']]

print(sleuthwmovies_sorted)
```

```
           character    n
288503        Andrew  1.0
1147041         Milo  2.0
1602268  Man on T.V.  3.0
```

How many roles were credited in the silent 1921 version of Hamlet?

In [55]:
```
Hamletwmovies = cast[(cast['title'].str.contains('Hamlet', na=False, case=False)) & (cast[

Hamletwmovies_roles = Hamletwmovies['character'].unique()

uniqueHamletwmovies_roles = len(Hamletwmovies_roles)

uniqueHamletwmovies_roles
```

Out[55]:
```
9
```

How many roles were credited in Branagh's 1996 Hamlet?

In [58]:
```
branhamlet = cast[(cast['title'].str.contains('Hamlet', na=False, case=False)) &
                  (cast['year'] == 1996) &
                  (cast['name'].str.contains('Branagh', na=False, case=False))]

branhamlet_roles = branhamlet['character'].unique()
```

```
        uniquebranhamlet_roles = len(branhamlet_roles)

        uniquebranhamlet_roles
```

Out[58]:   1

How many "Hamlet" roles have been listed in all film credits through history?

In [61]:
```
hamlet = cast[cast['title'].str.contains('Hamlet', na=False, case=False)]

hamlet_roles = hamlet['character'].unique()
hamletmovies = hamletmovies[hamletmovies['n'].notna()]

hamlet_roles = len(hamlet_roles)

hamlet_roles
```

Out[61]:   467

How many people have played an "Ophelia"?

In [62]:
```
opheliacount=cast[cast['character'].str.contains('Ophelia', na=False, case=False)]
len(opheliacount.character)
```

Out[62]:   140

How many people have played a role called "The Dude"?

In [63]:
```
dudecount=cast[cast['character'].str.contains('The Dude', na=False, case=False)]
len(dudecount.character)
```

Out[63]:   36

How many people have played a role called "The Stranger"?

In [64]:
```
dudecount=cast[cast['character'].str.contains('The Stranger', na=False, case=False)]
len(dudecount.character)
```

Out[64]:   230

How many roles has Sidney Poitier played throughout his career?

In [66]:
```
sidneypoitierfilms=cast[cast['name'].str.contains('Sidney Poitier', na=False, case=False)]
rolesunique=sidneypoitierfilms['character'].unique()
len(rolesunique)
```

Out[66]:   41

How many roles has Judi Dench played?

In [67]:
```python
jdenchfilms=cast[cast['name'].str.contains('Judi Dench', na=False, case=False)]
rolesunique=jdenchfilms['character'].unique()
len(rolesunique)
```

Out[67]:
44

List the supporting roles (having n=2) played by Cary Grant in the 1940s, in order by year.

In [72]:
```python
cgroles40s = cast[(cast['year'] >= 1940) & (cast['year'] < 1950) &
                  (cast['n'] == 2) &
                  (cast['name'].str.contains('Cary Grant', na=False, case=False))]

cgroles40s=cgroles40s.sort_values(by='year')
cgroles40s
```

Out[72]:

	title	year	name	type	character	n
773493	My Favorite Wife	1940	Cary Grant	actor	Nick	2.0
773503	Penny Serenade	1941	Cary Grant	actor	Roger Adams	2.0

List the leading roles that Cary Grant played in the 1940s in order by year.

In [73]:
```python
cgroles40s = cast[(cast['year'] >= 1940) & (cast['year'] < 1950) &
                  (cast['n'] == 1) &
                  (cast['name'].str.contains('Cary Grant', na=False, case=False))]

cgroles40s=cgroles40s.sort_values(by='year')
cgroles40s
```

Out[73]:

	title	year	name	type	character	n
773518	The Howards of Virginia	1940	Cary Grant	actor	Matt Howard	1.0
773475	His Girl Friday	1940	Cary Grant	actor	Walter Burns	1.0
773520	The Philadelphia Story	1940	Cary Grant	actor	C. K. Dexter Haven	1.0
773508	Suspicion	1941	Cary Grant	actor	Johnnie	1.0
773522	The Talk of the Town	1942	Cary Grant	actor	Leopold Dilg	1.0
773499	Once Upon a Honeymoon	1942	Cary Grant	actor	Patrick 'Pat' O'Toole	1.0
773466	Destination Tokyo	1943	Cary Grant	actor	Capt. Cassidy	1.0
773491	Mr. Lucky	1943	Cary Grant	actor	Joe Adams	1.0
773492	Mr. Lucky	1943	Cary Grant	actor	Joe Bascopolous	1.0
773500	Once Upon a Time	1944	Cary Grant	actor	Jerry Flynn	1.0
773458	Arsenic and Old Lace	1944	Cary Grant	actor	Mortimer Brewster	1.0

773495	None But the Lonely Heart	1944	Cary Grant	actor	Ernie Mott	1.0
773494	Night and Day	1946	Cary Grant	actor	Cole Porter	1.0
773497	Notorious	1946	Cary Grant	actor	Devlin	1.0
773514	The Bachelor and the Bobby-Soxer	1947	Cary Grant	actor	Dick	1.0
773515	The Bishop's Wife	1947	Cary Grant	actor	Dudley	1.0
773490	Mr. Blandings Builds His Dream House	1948	Cary Grant	actor	Jim Blandings	1.0
773470	Every Girl Should Be Married	1948	Cary Grant	actor	Dr. Madison Brown	1.0
773479	I Was a Male War Bride	1949	Cary Grant	actor	Capt. Henri Rochard	1.0

How many roles were available for actors in the 1950s?

In [74]:
```python
actorrole50s= cast[(cast['year'] >= 1950) & (cast['year'] < 1960) &

                   (cast['type'].str.contains('actor', na=False, case=False))]
rolesavail=actorrole50s.character.unique()
len(rolesavail)
```

Out[74]:
```
88802
```

How many roles were available for actresses in the 1950s?

In [75]:
```python
actressrole50s= cast[(cast['year'] >= 1950) & (cast['year'] < 1960) &

                     (cast['type'].str.contains('actress', na=False, case=False))]
rolesavail=actressrole50s.character.unique()
len(rolesavail)
```

Out[75]:
```
34206
```

How many leading roles (n=1) were available from the beginning of film history through 1980?

In [76]:
```python
leading_roles = cast[(cast['n'] == 1) & (cast['year'] <= 1980)]

number_of_leading_roles = len(leading_roles.character.unique())
number_of_leading_roles
```

Out[76]:
```
49297
```

How many non-leading roles were available through from the beginning of film history through 1980?

In [78]:
```python
non_leading_roles = cast[(cast['n'] != 1) & (cast['year'] <= 1980)]

number_of_non_leading_roles = len(non_leading_roles.character.unique())
number_of_non_leading_roles
```

537870

How many roles through 1980 were minor enough that they did not warrant a numeric "n" rank?

In [81]:
```python
minor_roles = cast[(cast['n'] == np.nan) & (cast['year'] <= 1980)]

number_of_minor_roles = len(minor_roles.character.unique())
number_of_minor_roles
```

Out[81]: 0