In [2]:
```
%matplotlib inline
import pandas as pd
```

In [5]:
```
cast = pd.read_csv("C:\\Users\\sujoydutta\\Downloads\\cast.csv")
cast.head()
```

Out[5]:

	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

What are the ten most common movie names of all time?

In [6]:
```
#top 10 actors with most number of roles
cast['title'].value_counts().head(10)
```

Out[6]:
```
title
Around the World in Eighty Days       1300
7 cajas                                689
Thelma                                 580
The Ten Commandments                   529
The Eschatrilogy: Book of the Dead     517
A Broken Code                          477
Welcome to Essex                       443
Stuck on You                           438
The Dark Knight Rises                  429
The Buccaneer                          428
Name: count, dtype: int64
```

Which three years of the 1930s saw the most films released?

In [10]:
```
the30s = cast[(cast['year'] >= 1930)
                    & (cast['year'] < 1940)]
the30smovies=the30s.title.unique()
titles_by_year = the30s.groupby('year')['title'].nunique()

titles_by_year_desc = titles_by_year.sort_values(ascending=False)

titles_by_year_desc.head(3)
```

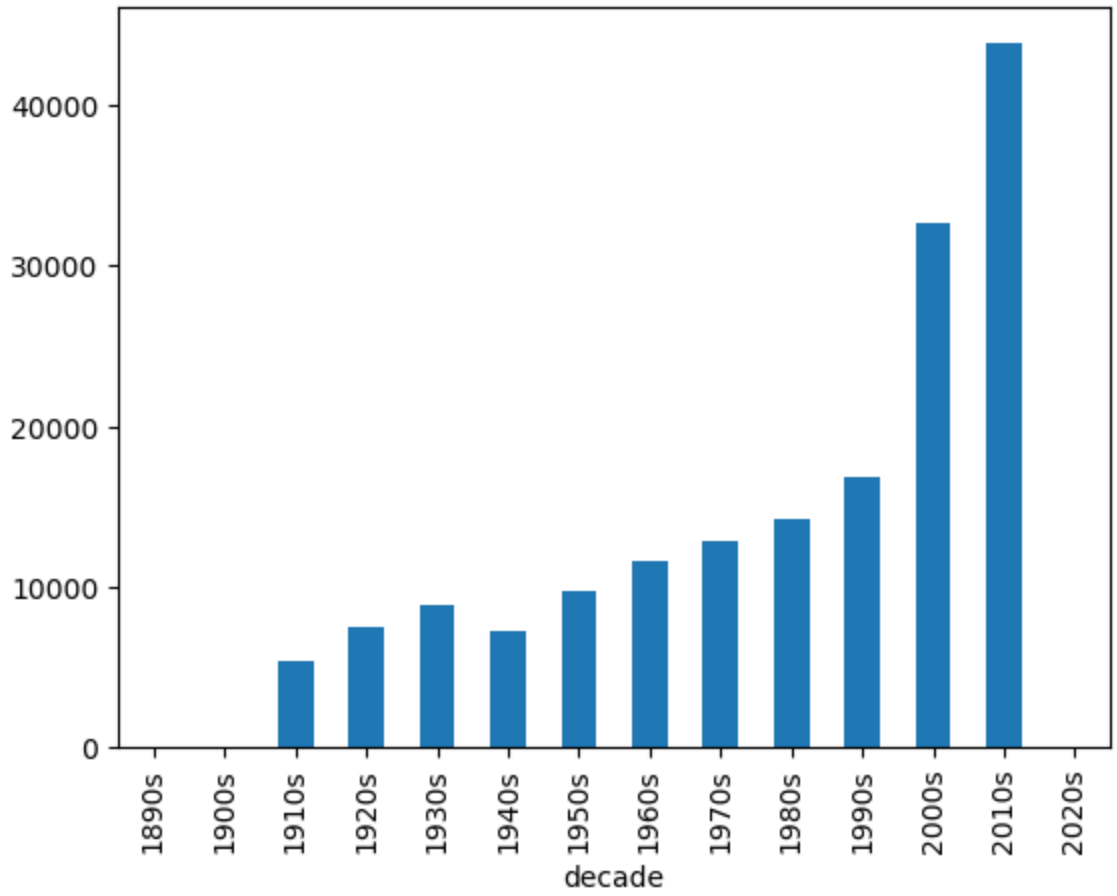
Out[10]:
```
year
1937    1044
1936    1028
1935     960
Name: title, dtype: int64
```

Plot the number of films that have been released each decade over the history of cinema.

In [27]:
```
cast['decade'] = cast['decade'] = (cast['year'] // 10 * 10).astype(str) + 's'

filmbydecade= cast.groupby('decade')['title'].nunique()
filmbydecade.plot(kind='bar')
```

Out[27]: <Axes: xlabel='decade'>

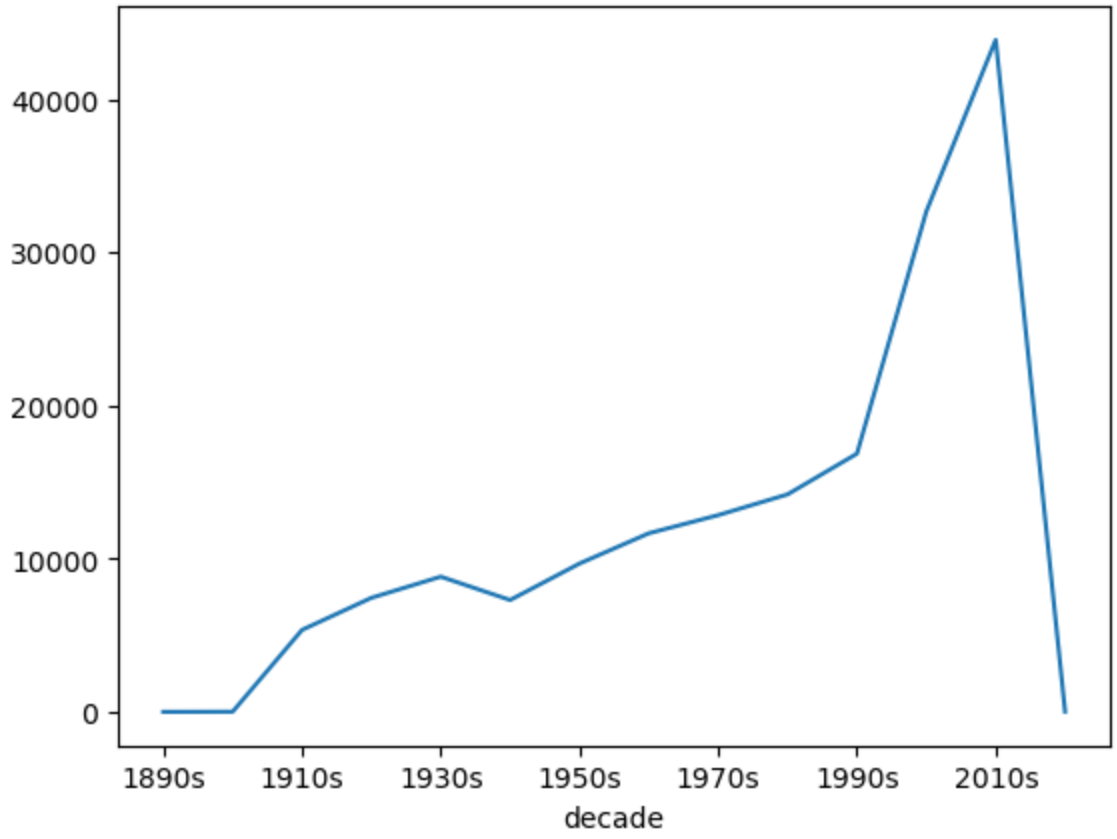


Plot the number of "Hamlet" films made each decade.

In [28]:
```
hamlet = cast[cast['title'].str.contains('Hamlet', na=False, case=False)]


hamletfilms= cast.groupby('decade')['title'].nunique()
hamletfilms.plot(kind='line')
```

Out[28]: <Axes: xlabel='decade'>



Plot the number of "Rustler" characters in each decade of the history of film.

In [36]:
```
rustlerchar = cast[cast['character'].str.contains('Rustler', na=False, case=False)]


rustlerchar= rustlerchar.groupby('decade')['character'].nunique()
rustlerchar.plot(kind='line')
```

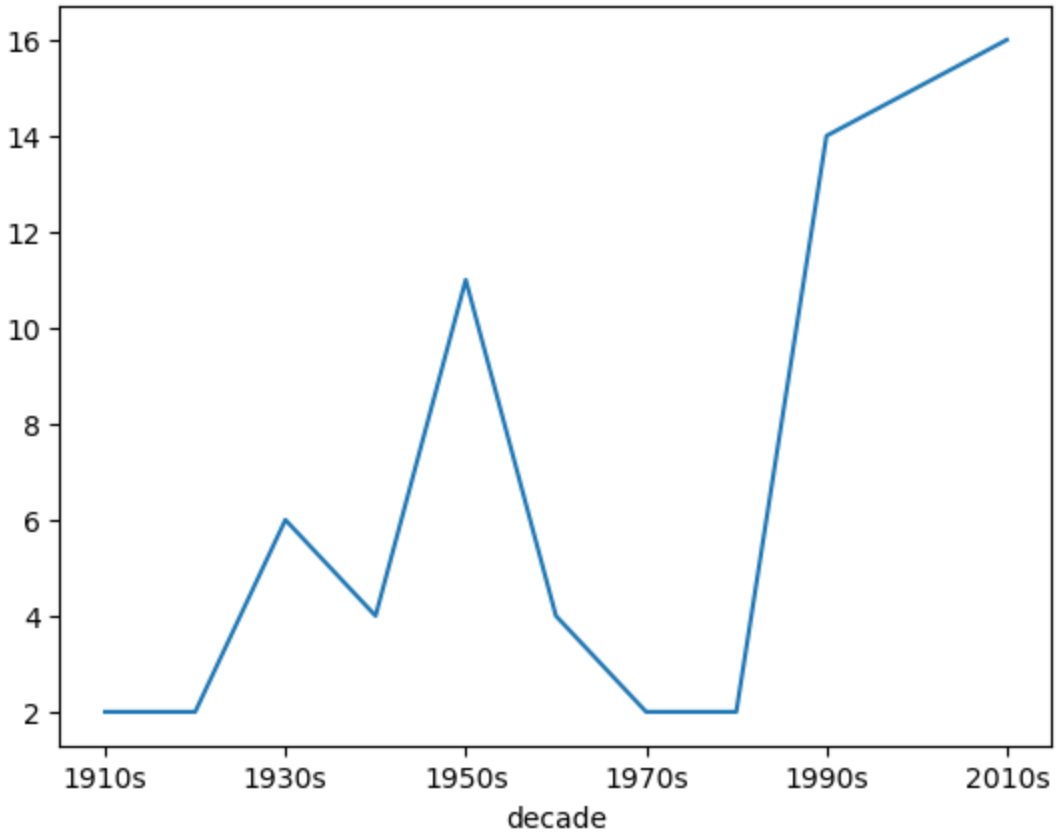
Out[36]:
```
<Axes: xlabel='decade'>
```

Plot the number of "Hamlet" characters each decade.

In [35]:
```
hamletchar = cast[cast['character'].str.contains('hamlet', na=False, case=False)]


hamletchar= hamletchar.groupby('decade')['character'].nunique()
hamletchar.plot(kind='line')
```

Out[35]:
```
<Axes: xlabel='decade'>
```



What are the 11 most common character names in movie history?

In [39]:
```
cast['character'].value_counts().head(11)
```

Out[39]:
```
character
Himself        18926
Dancer         11015
Extra           8638
Reporter        7593
Doctor          6803
Policeman       6470
Student         6390
Nurse           6127
Bartender       6123
Minor Role      5830
Party Guest     5820
Name: count, dtype: int64
```

Who are the 10 people most often credited as "Herself" in film history?

In [40]:
```
herselfcred = cast[cast['character'].str.contains('Herself', na=False, case=False)]

herselfcred= herselfcred.groupby('name')['character'].nunique()
herselfcred.sort_values(ascending=False).head(10)
```

Out[40]:
```
name
Jacqueline Kennedy    7
Queen Elizabeth II    7
Pat Nixon             5
Nellie Connally       5
Norma Shearer         4
Lady Bird Johnson     4
Princess Margaret     4
Ethel (II) Kennedy    4
Brigitte Bardot       4
```

```
Caroline Kennedy      4
Name: character, dtype: int64
```

Who are the 10 people most often credited as "Himself" in film history?

In [41]:
```
himselfcred = cast[cast['character'].str.contains('himself', na=False, case=False)]

himselfcred= himselfcred.groupby('name')['character'].nunique()
himselfcred.sort_values(ascending=False).head(10)
```

Out[41]:
```
name
Adolf Hitler             16
John F. Kennedy          13
Richard Nixon            11
Robert F. Kennedy         9
Nikita Khrushchev         8
Amitabh Bachchan          8
Joe Louis                 8
Ronald Reagan             7
Lyndon Johnson            7
Franklin D. Roosevelt     6
Name: character, dtype: int64
```

Which actors or actresses appeared in the most movies in the year 1945?

In [57]:
```
performer1945= cast[(cast['year'] == 1945)]
popularperformer45= performer1945.groupby('name')['title'].nunique()
popularperformer45.sort_values(ascending=False).head(10)
```

Out[57]:
```
name
Emmett Vogan         39
Sam (II) Harris      30
Nolan Leary          27
Harold Miller        27
Frank O'Connor       26
Bess Flowers         25
Edmund Cobb          24
Charles Sullivan     24
Pierre Watkin        24
Tom London           24
Name: title, dtype: int64
```

Which actors or actresses appeared in the most movies in the year 1985?

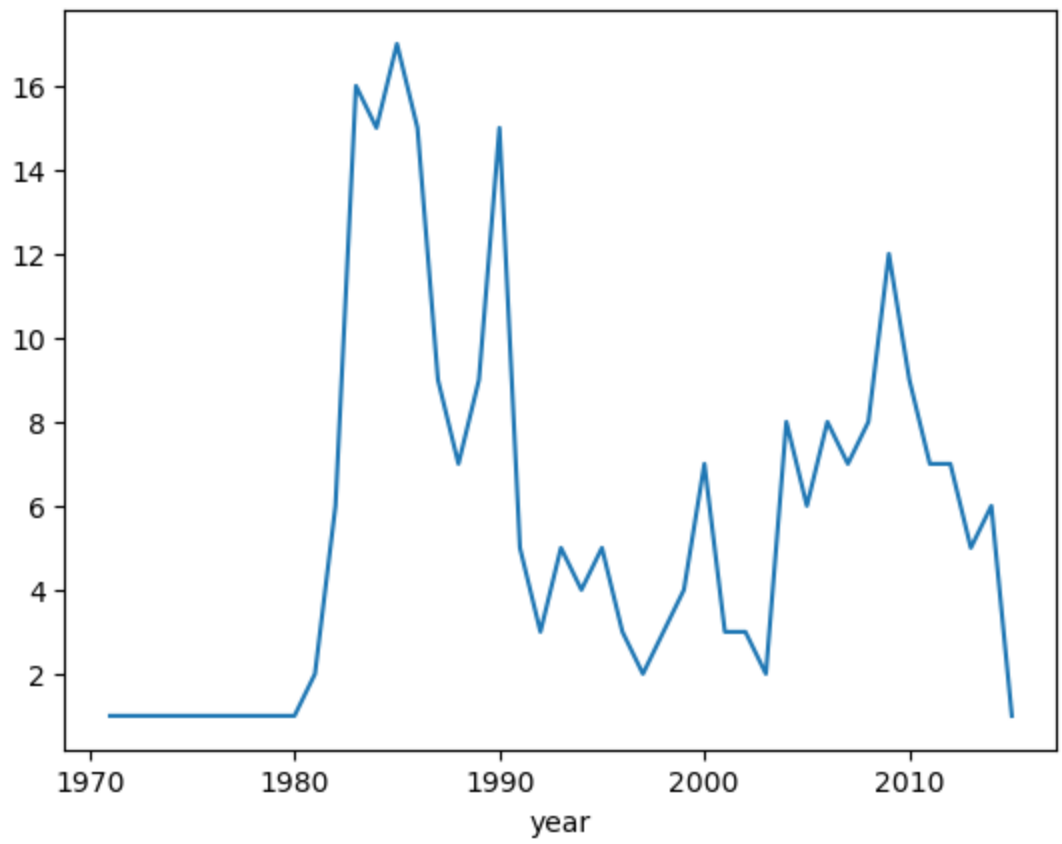
In [59]:
```
performer1985= cast[(cast['year'] == 1985)]
popularperformer85= performer1985.groupby('name')['title'].nunique()
popularperformer85.sort_values(ascending=False).head(10)
```

Out[59]:
```
name
Mammootty            17
Shakti Kapoor        16
Sukumari             16
Mohanlal             13
Aruna Irani          13
Deven Verma          13
Satyendra Kapoor     12
Asrani               12
Pinchoo Kapoor       12
Om Shivpuri          12
Name: title, dtype: int64
```

Plot how many roles Mammootty has played in each year of his career.

In [55]:
```
mamootyfilms=cast[cast['name'].str.contains('Mammootty', na=False, case=False)]
rolecount= mamootyfilms.groupby('year')['character'].nunique()
rolecount.plot(kind='line')
```

Out[55]:
```
<Axes: xlabel='year'>
```



What are the 10 most frequent roles that start with the phrase "Patron in"?

In [67]:
```
patronroles=cast[cast['character'].str.contains('^patron in', na=False, case=False)]
rolecount= patronroles.groupby('character')['title'].nunique()
rolecount.sort_values(ascending=False).head(10)
```

Out[67]:
```
character
Patron in Bar             4
Patron in restaurant      3
Patron in Restaurant      3
Patron in Coffee Shop     2
Patron In Chaps           1
Patron in Strip Club      1
Patron in Pizza Place     1
Patron in Pool Hall       1
Patron in Quiet Bar       1
Patron in Store           1
Name: title, dtype: int64
```

What are the 10 most frequent roles that start with the word "Science"?

In [68]:
```
scienceroles=cast[cast['character'].str.contains('^science', na=False, case=False)]
rolecount= scienceroles.groupby('character')['title'].nunique()
rolecount.sort_values(ascending=False).head(10)
```

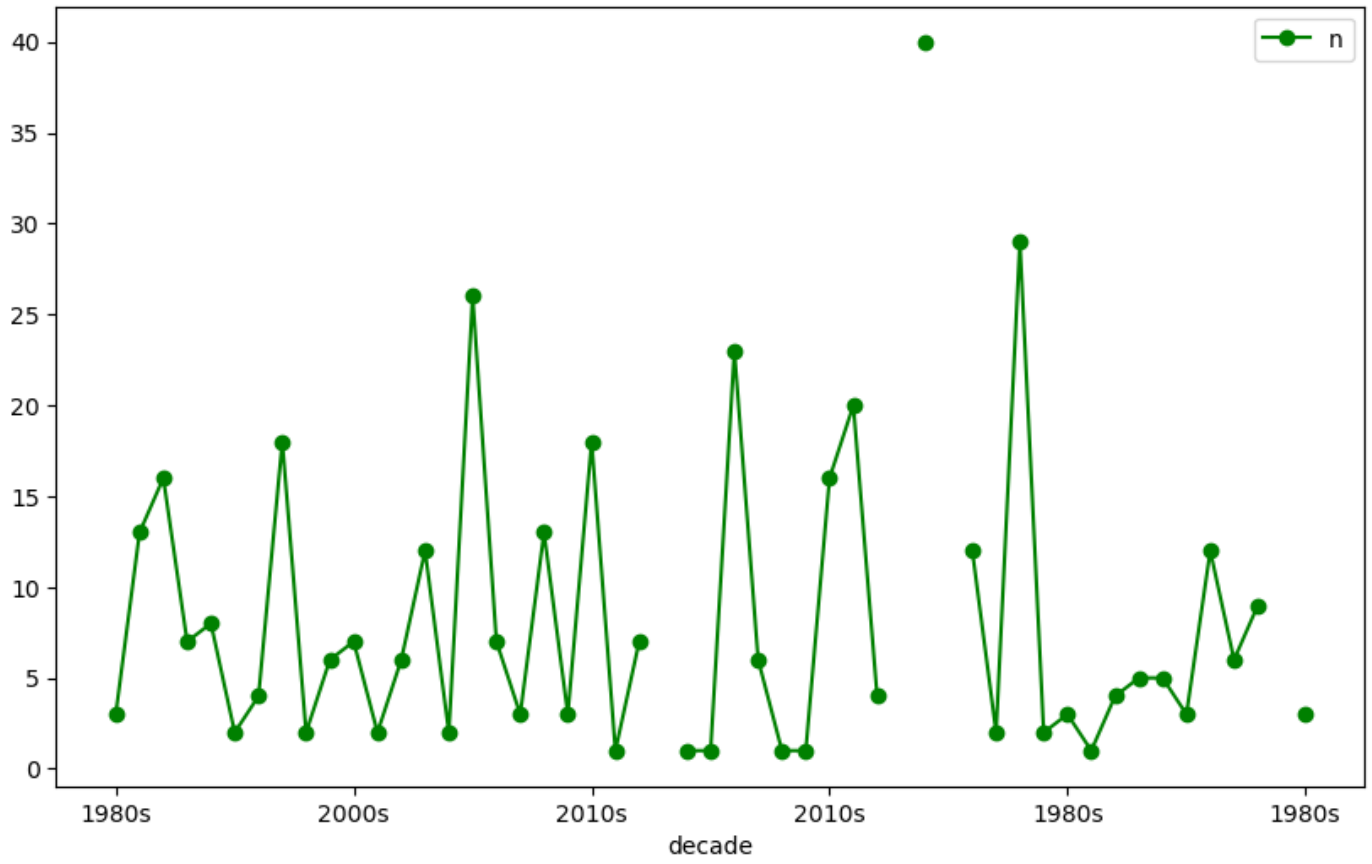
Out[68]:
```
character
Science Teacher         53
Science Student          7
```

```
Science Fair Student     5
Science Fair Judge       4
Science Reporter         4
Science Kid              4
Science teacher          3
Science Officer          3
Science Fair Kid         3
Science Officer 0718     2
Name: title, dtype: int64
```

Plot the n-values of the roles that Judi Dench has played over her career.

In [49]:
```
judidenchfilms=cast[cast['name'].str.contains('Judi Dench', na=False, case=False)]
judidenchfilms.plot(x='decade', y='n', kind='line', marker='o', linestyle='-', color='g'
```

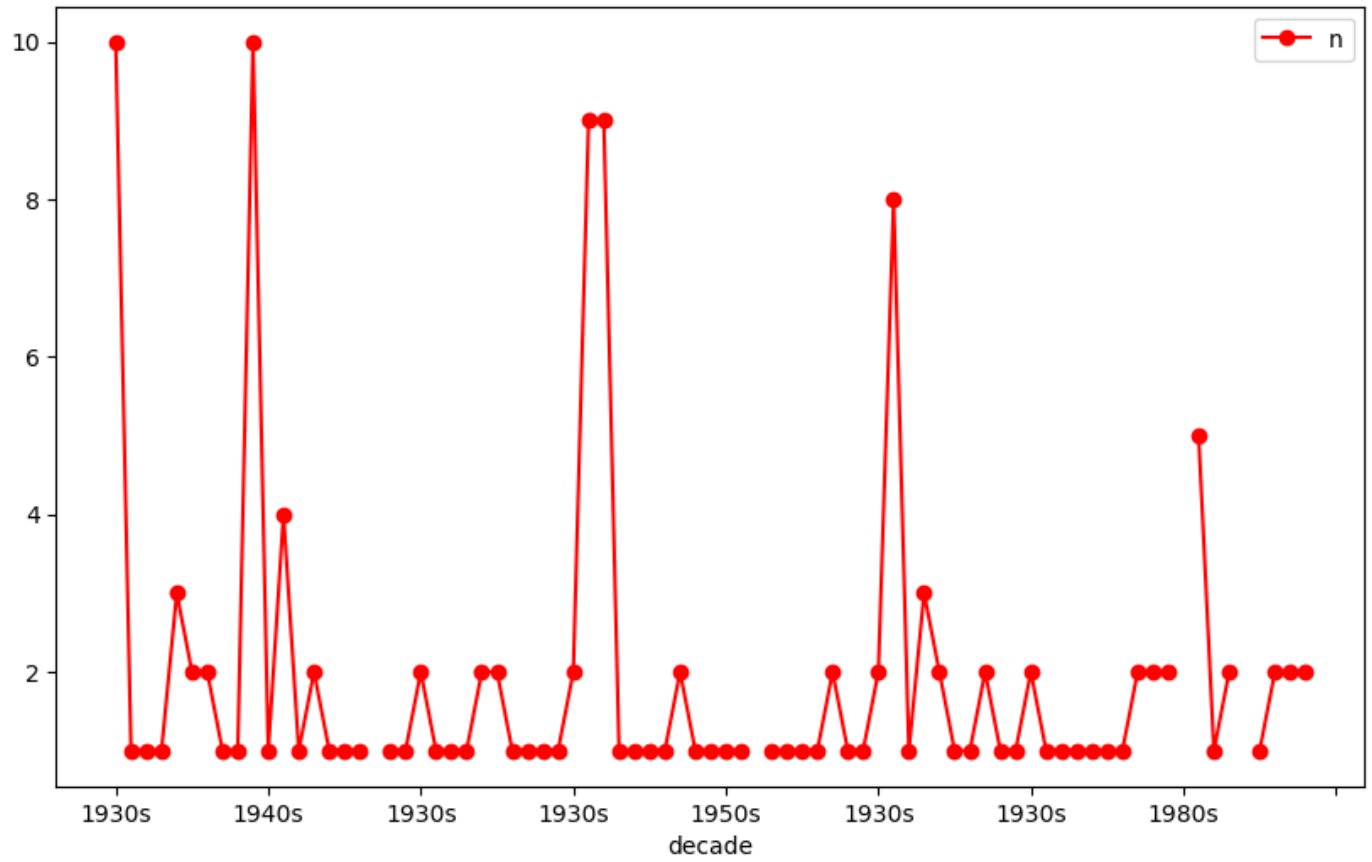
Out[49]:
```
<Axes: xlabel='decade'>
```



Plot the n-values of Cary Grant's roles through his career.

In [48]:
```
carygrantfilms=cast[cast['name'].str.contains('Cary Grant', na=False, case=False)]
carygrantfilms.plot(x='decade', y='n', kind='line', marker='o', linestyle='-', color='r'
```

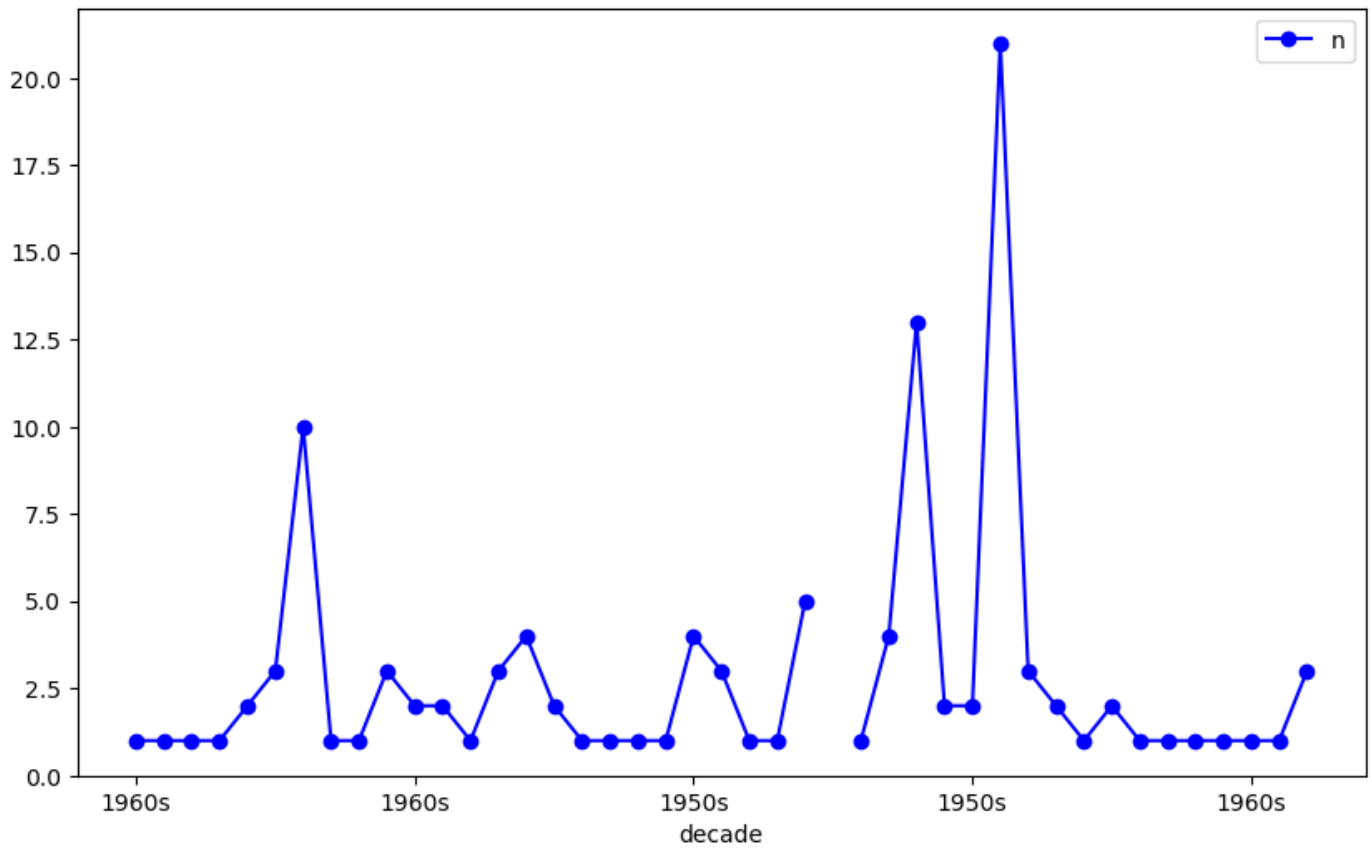
Out[48]:
```
<Axes: xlabel='decade'>
```



Plot the n-value of the roles that Sidney Poitier has acted over the years.

In [47]:
```
sidneypoitierfilms=cast[cast['name'].str.contains('Sidney Poitier', na=False, case=False
sidneypoitierfilms.plot(x='decade', y='n', kind='line', marker='o', linestyle='-', color
```

Out[47]:
```
<Axes: xlabel='decade'>
```



How many leading (n=1) roles were available to actors, and how many to actresses, in the 1950s?

In [17]:
```
actorlead50s = cast[(cast['year'] >= 1950) & (cast['year'] < 1960) &
                    (cast['n'] == 1) &
                    (cast['type'].str.contains('actor', na=False, case=False))]


rolesavail = actorlead50s['character'].unique()


number_of_roles = len(rolesavail)
number_of_roles
```

Out[17]:
```
5743
```

In [16]:
```
actresslead50s = cast[(cast['year'] >= 1950) & (cast['year'] < 1960) &
                    (cast['n'] == 1) &
                    (cast['type'].str.contains('actress', na=False, case=False))]


rolesavail = actresslead50s['character'].unique()


number_of_roles = len(rolesavail)
number_of_roles
```

Out[16]:
```
2533
```

How many supporting (n=2) roles were available to actors, and how many to actresses, in the 1950s?

In [18]:
```
actorsupport50s = cast[(cast['year'] >= 1950) & (cast['year'] < 1960) &
                    (cast['n'] == 2) &
                    (cast['type'].str.contains('actor', na=False, case=False))]


rolesavail = actorsupport50s['character'].unique()


number_of_roles = len(rolesavail)
number_of_roles
```

Out[18]:
```
3981
```

In [20]:
```
actresssupport50s = cast[(cast['year'] >= 1950) & (cast['year'] < 1960) &
                    (cast['n'] == 2) &
                    (cast['type'].str.contains('actress', na=False, case=False))]


rolesavail = actresssupport50s['character'].unique()


number_of_roles = len(rolesavail)
number_of_roles
```

Out[20]:
```
3887
```