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

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 [5]:
```
#dropping nulls
cast=cast.dropna()
cast
```

Out[5]:

| | title | year | name | type | character | n |
|---|---|---|---|---|---|---|
| **0** | Suuri illusioni | 1985 | Homo $ | actor | Guests | 22.0 |
| **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 |
| **8** | When the Man Went South | 2014 | Taipaleti 'Atu'ake | actor | Two Palms - Ua'i Paame | 8.0 |
| **9** | Little Angel (Angelita) | 2015 | Michael 'babeepower' Viera | actor | Chico | 9.0 |
| **...** | ... | ... | ... | ... | ... | ... |
| **3311641** | Niceland (Population. 1.000.002) | 2004 | Steinunn ?orvaldsd?ttir | actress | Factory Worker | 21.0 |
| **3311642** | Stuttur Frakki | 1993 | Sveinbj?rg ??rhallsd?ttir | actress | Flugfreyja | 24.0 |
| **3311643** | Foxtrot | 1988 | Lilja ??risd?ttir | actress | D?ra | 24.0 |
| **3311644** | Niceland (Population. 1.000.002) | 2004 | Sigr??ur J?na ??risd?ttir | actress | Woman in Bus | 26.0 |
| **3311645** | U.S.S.S.S... | 2003 | Krist?n Andrea ??r?ard?ttir | actress | Afgr.dama ? bens?nst?? | 17.0 |

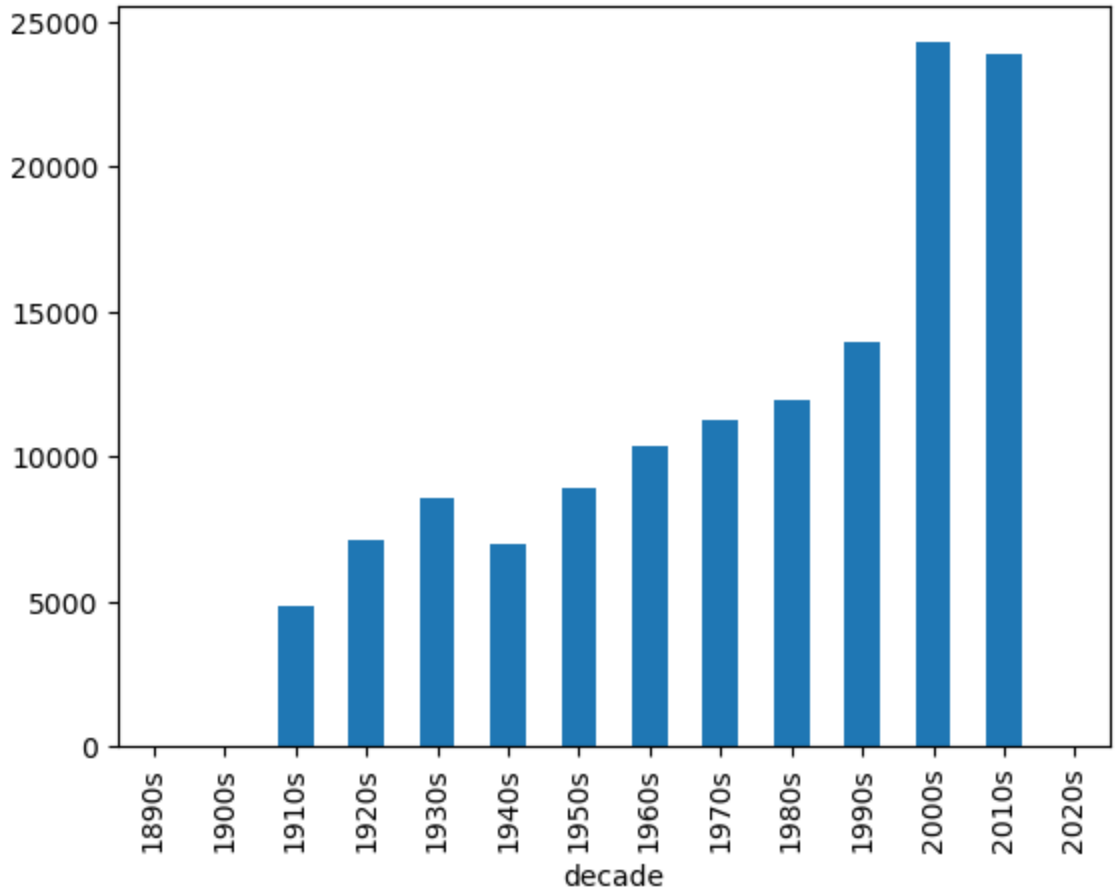
2069573 rows × 6 columns

## Using groupby(), plot the number of films that have been released each decade in the history of cinema.

In [6]:
```
cast['decade'] = cast['decade'] = (cast['year'] // 10 * 10).astype(str) + 's'

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

Out[6]:
```
<Axes: xlabel='decade'>
```

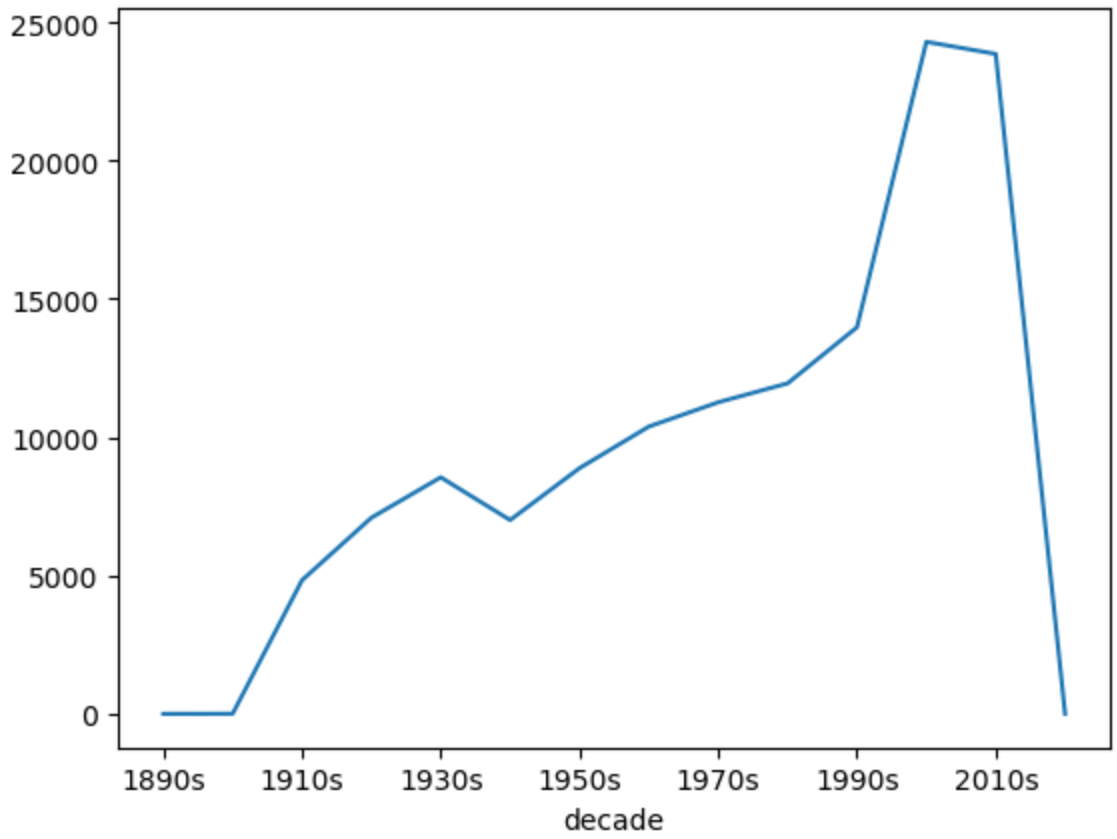



## Use groupby() to plot the number of "Hamlet" films made each decade.

In [7]:
```
hamlet = cast[cast['title'].str.contains('Hamlet', na=False, case=False)]


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

Out[7]:
```
<Axes: xlabel='decade'>
```




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

In [8]:
```
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[8]:
```
5743
```

In [9]:
```
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[9]:
```
2533
```

## In the 1950s decade taken as a whole, how many total roles were available to actors, and how many to actresses, for each "n" number 1 through 5?

In [10]:
```
# Filter the data for the 1950s
roles_50s = cast[(cast['year'] >= 1950) & (cast['year'] < 1960)]


def count_roles(roles, role_type, n):
    return roles[(roles['n'] == n) & (roles['type'].str.contains(role_type, na=False, ca


actors_roles_count = {n: count_roles(roles_50s, 'actor', n) for n in range(1, 6)}
actresses_roles_count = {n: count_roles(roles_50s, 'actress', n) for n in range(1, 6)}


print("Total roles available to actors for each 'n' from 1 to 5 in the 1950s:")
for n, count in actors_roles_count.items():
    print(f"n = {n}: {count} roles")

print("\nTotal roles available to actresses for each 'n' from 1 to 5 in the 1950s:")
for n, count in actresses_roles_count.items():
    print(f"n = {n}: {count} roles")
```

```
Total roles available to actors for each 'n' from 1 to 5 in the 1950s:
n = 1: 5743 roles
n = 2: 3981 roles
n = 3: 5043 roles
n = 4: 5075 roles
n = 5: 5116 roles

Total roles available to actresses for each 'n' from 1 to 5 in the 1950s:
n = 1: 2533 roles
n = 2: 3887 roles
n = 3: 2718 roles
n = 4: 2453 roles
n = 5: 2212 roles
```

## Use groupby() to determine how many roles are listed for each movie named *The Pink Panther*.

In [13]:
```
pinkpatherfilms=cast[cast['title'].str.contains('The Pink Panther', na=False, case=False
rolecount= pinkpatherfilms.groupby('year')['character'].nunique()
rolecount
```

Out[13]:
```
year
1963.0     1
1975.0     6
1976.0    12
1978.0    10
1982.0     3
1983.0     9
1993.0     7
2006.0    10
2009.0     6
Name: character, dtype: int64
```

## List, in order by year, each of the films in which Frank Oz has played more than 1 role.

In [13]:
```
# Filter for Frank Oz roles

frankozfilms = cast[cast['name'].str.contains('Frank Oz', na=False, case=False)]

frankozfilmsmultirole = frankozfilms.groupby(['title', 'year']).size().reset_index(name=
frankozfilmsmultirole = frankozfilmsmultirole.sort_values(by='year')
```

```
frankozfilmsmultirole=frankozfilmsmultirole[frankozfilmsmultirole.role_count>1]
frankozfilmsmultirole
```

Out[13]:

| | title | year | role_count |
|---|---|---|---|
| **20** | The Muppet Movie | 1979 | 8 |
| **0** | An American Werewolf in London | 1981 | 2 |
| **18** | The Great Muppet Caper | 1981 | 6 |
| **17** | The Dark Crystal | 1982 | 2 |
| **21** | The Muppets Take Manhattan | 1984 | 7 |
| **8** | Sesame Street Presents: Follow that Bird | 1985 | 3 |
| **19** | The Muppet Christmas Carol | 1992 | 7 |
| **6** | Muppet Treasure Island | 1996 | 4 |
| **7** | Muppets from Space | 1999 | 4 |
| **15** | The Adventures of Elmo in Grouchland | 1999 | 3 |

## List each of the characters that Frank Oz has portrayed at least twice.

In [14]:

```
# Filter for Frank Oz roles

frankozcharsmultirole = frankozfilms.groupby(['character']).size().reset_index(name='tim
frankozcharsmultirole=frankozcharsmultirole[frankozcharsmultirole.times_portrayed>1]
frankozcharsmultirole
```

Out[14]:

| | character | times_portrayed |
|---|---|---|
| **0** | Animal | 6 |
| **2** | Bert | 3 |
| **4** | Cookie Monster | 3 |
| **9** | Fozzie Bear | 4 |
| **14** | Grover | 2 |
| **17** | Miss Piggy | 6 |
| **24** | Sam the Eagle | 5 |
| **33** | Yoda | 5 |