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

In [30]:
```
cast = pd.read_csv("C:\\Users\\sujoydutta\\Downloads\\cast.csv")
cast.head()
```

Out[30]:

	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 [31]:
```
release = pd.read_csv("C:\\Users\\sujoydutta\\Downloads\\movie_release.csv")
release.head()
```

Out[31]:

	production_countries	release_date	title
0	[{'iso_3166_1': 'US', 'name': 'United States o...	30-10-95	Toy Story
1	[{'iso_3166_1': 'US', 'name': 'United States o...	15-12-95	Jumanji
2	[{'iso_3166_1': 'US', 'name': 'United States o...	22-12-95	Grumpier Old Men
3	[{'iso_3166_1': 'US', 'name': 'United States o...	22-12-95	Waiting to Exhale
4	[{'iso_3166_1': 'US', 'name': 'United States o...	10-02-95	Father of the Bride Part II

In [39]:
```
release.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 45376 entries, 0 to 45465
Data columns (total 6 columns):
 #   Column                Non-Null Count  Dtype         
---  ------                --------------  -----         
 0   production_countries  45376 non-null  object        
 1   release_date          45376 non-null  datetime64[ns]
 2   title                 45376 non-null  object        
 3   year                  45376 non-null  int32         
 4   month                 45376 non-null  object        
 5   day_of_week           45376 non-null  object        
dtypes: datetime64[ns](1), int32(1), object(4)
memory usage: 2.3+ MB
```

In [33]:
```
release=release.dropna()
```

In [34]:
```
release['year'] = pd.to_datetime(release['release_date'], errors='coerce').dt.year
release.head()
```

```
C:\Users\sujoydutta\AppData\Local\Temp\ipykernel_4272\1911974432.py:1: UserWarning: Could not infer format, so each element will be parsed individually, falling back to `dateutil`. To ensure parsing is consistent and as-expected, please specify a format.
  release['year'] = pd.to_datetime(release['release_date'], errors='coerce').dt.year
```

Out[34]:

	production_countries	release_date	title	year
0	[{'iso_3166_1': 'US', 'name': 'United States o...	30-10-95	Toy Story	1995

1	[{'iso_3166_1': 'US', 'name': 'United States o...	15-12-95	Jumanji	1995
2	[{'iso_3166_1': 'US', 'name': 'United States o...	22-12-95	Grumpier Old Men	1995
3	[{'iso_3166_1': 'US', 'name': 'United States o...	22-12-95	Waiting to Exhale	1995
4	[{'iso_3166_1': 'US', 'name': 'United States o...	10-02-95	Father of the Bride Part II	1995

In [36]:
```
release['year'] = release['year'].astype(int)
release.head()
```

Out[36]:

	production_countries	release_date	title	year
0	[{'iso_3166_1': 'US', 'name': 'United States o...	30-10-95	Toy Story	1995
1	[{'iso_3166_1': 'US', 'name': 'United States o...	15-12-95	Jumanji	1995
2	[{'iso_3166_1': 'US', 'name': 'United States o...	22-12-95	Grumpier Old Men	1995
3	[{'iso_3166_1': 'US', 'name': 'United States o...	22-12-95	Waiting to Exhale	1995
4	[{'iso_3166_1': 'US', 'name': 'United States o...	10-02-95	Father of the Bride Part II	1995

In [37]:
```
release['month'] = pd.to_datetime(release['release_date'], errors='coerce').dt.month
release['month'] = release['month'].apply(lambda x: calendar.month_name[x])
release.head()
```

```
C:\Users\sujoydutta\AppData\Local\Temp\ipykernel_4272\184801963.py:1: UserWarning: Could
not infer format, so each element will be parsed individually, falling back to `dateutil
`. To ensure parsing is consistent and as-expected, please specify a format.
  release['month'] = pd.to_datetime(release['release_date'], errors='coerce').dt.month
```

Out[37]:

	production_countries	release_date	title	year	month
0	[{'iso_3166_1': 'US', 'name': 'United States o...	30-10-95	Toy Story	1995	October
1	[{'iso_3166_1': 'US', 'name': 'United States o...	15-12-95	Jumanji	1995	December
2	[{'iso_3166_1': 'US', 'name': 'United States o...	22-12-95	Grumpier Old Men	1995	December
3	[{'iso_3166_1': 'US', 'name': 'United States o...	22-12-95	Waiting to Exhale	1995	December
4	[{'iso_3166_1': 'US', 'name': 'United States o...	10-02-95	Father of the Bride Part II	1995	October

In [38]:
```
release['release_date'] = pd.to_datetime(release['release_date'], errors='coerce')
release['day_of_week'] = release['release_date'].dt.day_name()
release.head()
```

```
C:\Users\sujoydutta\AppData\Local\Temp\ipykernel_4272\915379251.py:1: UserWarning: Could
not infer format, so each element will be parsed individually, falling back to `dateutil
`. To ensure parsing is consistent and as-expected, please specify a format.
  release['release_date'] = pd.to_datetime(release['release_date'], errors='coerce')
```

Out[38]:

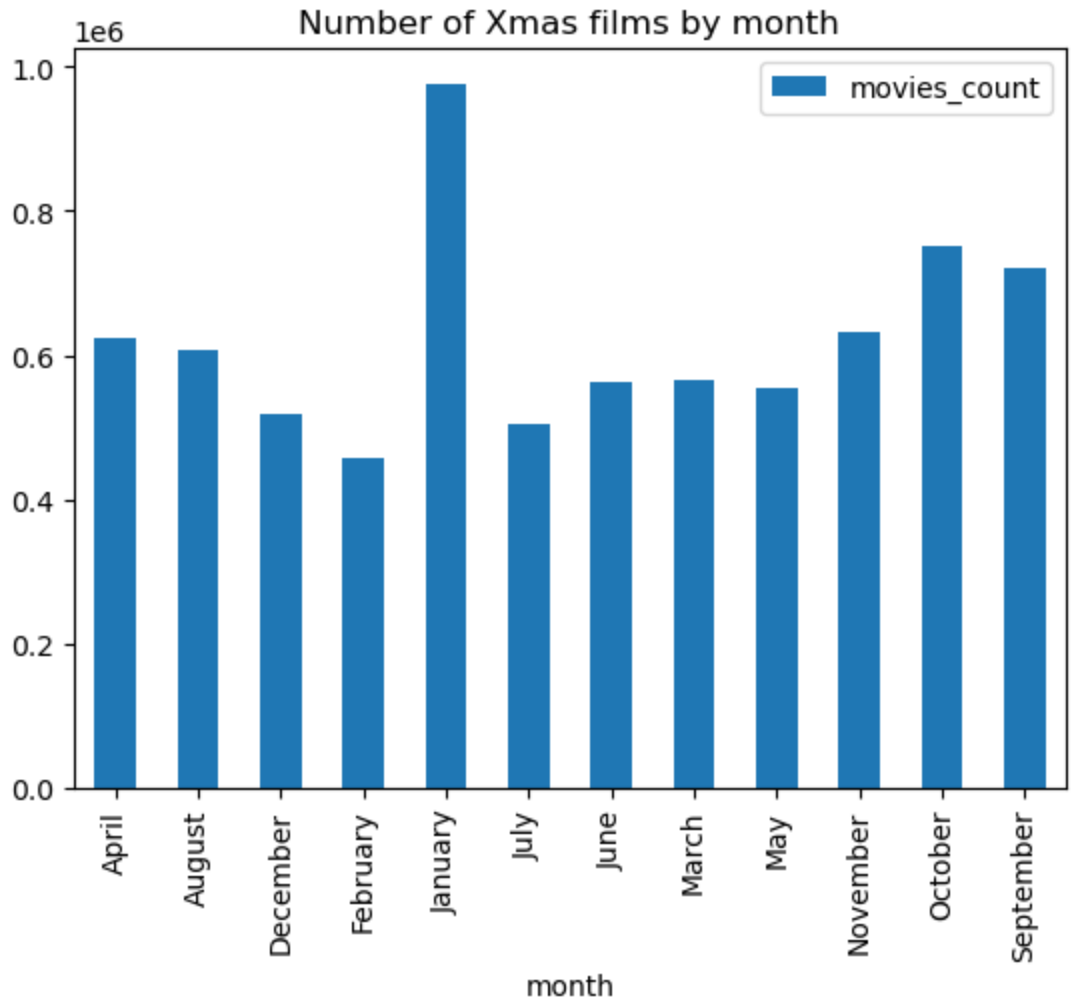
	production_countries	release_date	title	year	month	day_of_week
0	[{'iso_3166_1': 'US', 'name': 'United States o...	1995-10-30	Toy Story	1995	October	Monday
1	[{'iso_3166_1': 'US', 'name': 'United States o...	1995-12-15	Jumanji	1995	December	Friday
2	[{'iso_3166_1': 'US', 'name': 'United States o...	1995-12-22	Grumpier Old Men	1995	December	Friday
3	[{'iso_3166_1': 'US', 'name': 'United States o...	1995-12-22	Waiting to Exhale	1995	December	Friday

4	[{'iso_3166_1': 'US', 'name': 'United States o...	1995-10-02	Father of the Bride Part II	1995	October	Monday

Make a bar plot of the months in which movies with "Christmas" in their title tend to be released in the USA.

In [43]:
```
xmasfilm = cast[cast['title'].str.contains('Christmas', na=False, case=False)]
xmasfilmrel = pd.merge(xmasfilm, release, on='year', how='inner').fillna(0)
xmasfilmrel_count = xmasfilmrel.groupby('month').size().reset_index(name='movies_count')
xmasfilmrel_count.plot(x='month', y='movies_count', kind='bar', title='Number of Xmas fi
```

Out[43]:
```
<Axes: title={'center': 'Number of Xmas films by month'}, xlabel='month'>
```

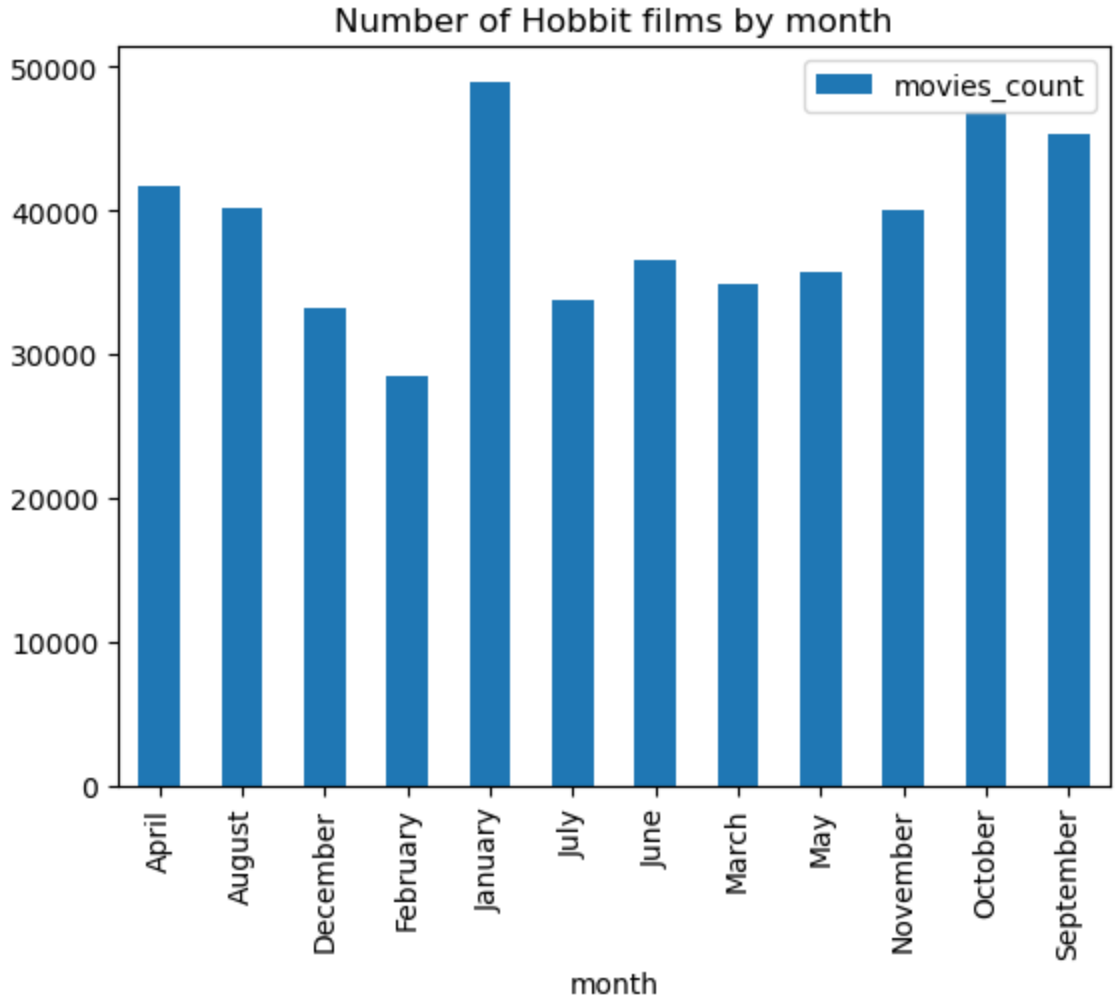


Make a bar plot of the months in which movies whose titles start with "The Hobbit" are released in the USA.

In [46]:
```
hobbitfilm = cast[cast['title'].str.contains('The Hobbit', na=False, case=False)]
hobbitfilmmrel = pd.merge(hobbitfilm, release, on='year', how='inner').fillna(0)
hobbitfilmrel_count = hobbitfilmmrel.groupby('month').size().reset_index(name='movies_co

hobbitfilmrel_count.plot(x='month', y='movies_count', kind='bar', title='Number of Hobbi
```

Out[46]:
```
<Axes: title={'center': 'Number of Hobbit films by month'}, xlabel='month'>
```

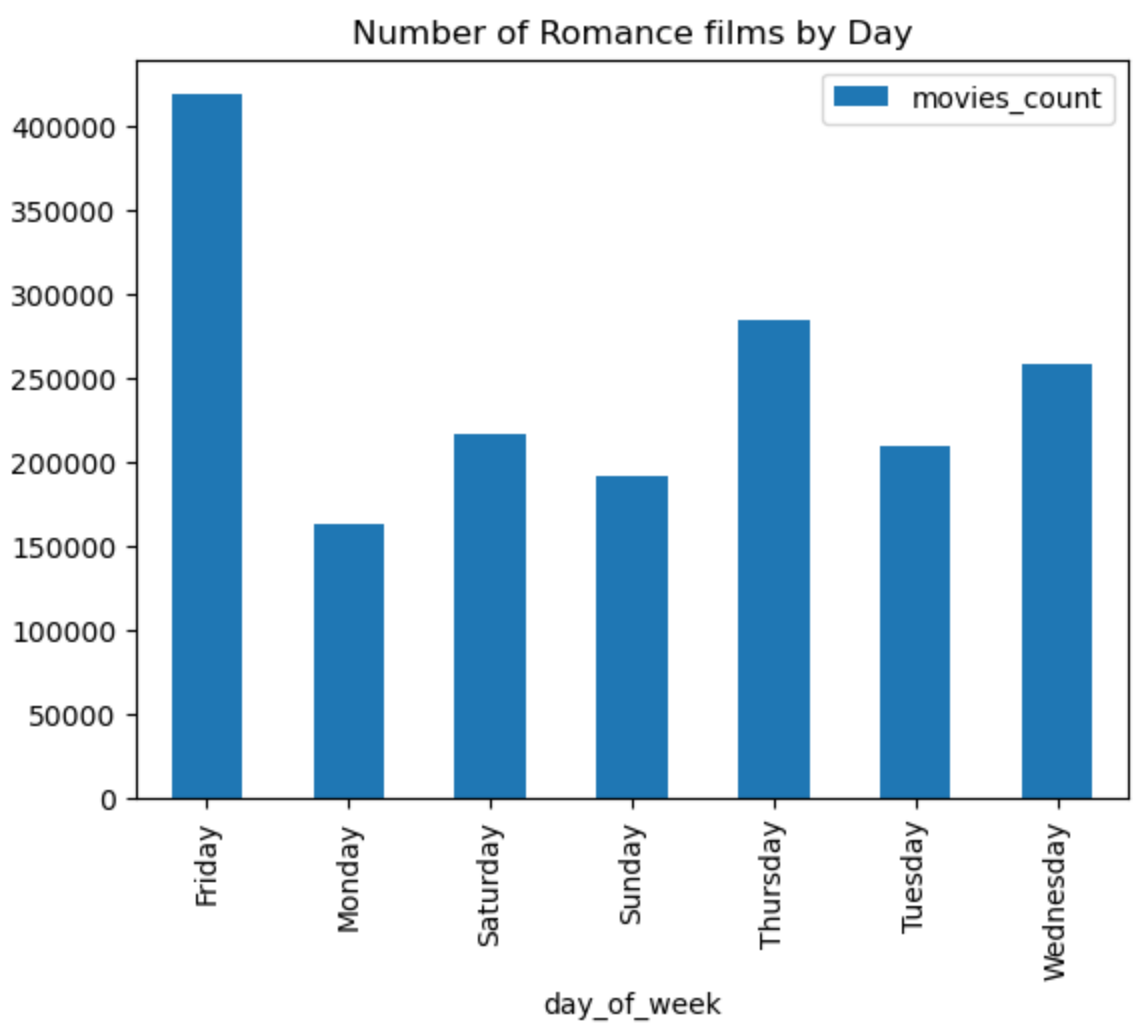


Make a bar plot of the day of the week on which movies with "Romance" in their title tend to be released in the USA.

In [48]:
```
romfilm = cast[cast['title'].str.contains('Romance', na=False, case=False)]
romfilmmrel = pd.merge(romfilm, release, on='year', how='inner').fillna(0)
romfilmmrel_count = romfilmmrel.groupby('day_of_week').size().reset_index(name='movies_c

romfilmmrel_count.plot(x='day_of_week', y='movies_count', kind='bar', title='Number of R
```

Out[48]:
```
<Axes: title={'center': 'Number of Romance films by Day'}, xlabel='day_of_week'>
```

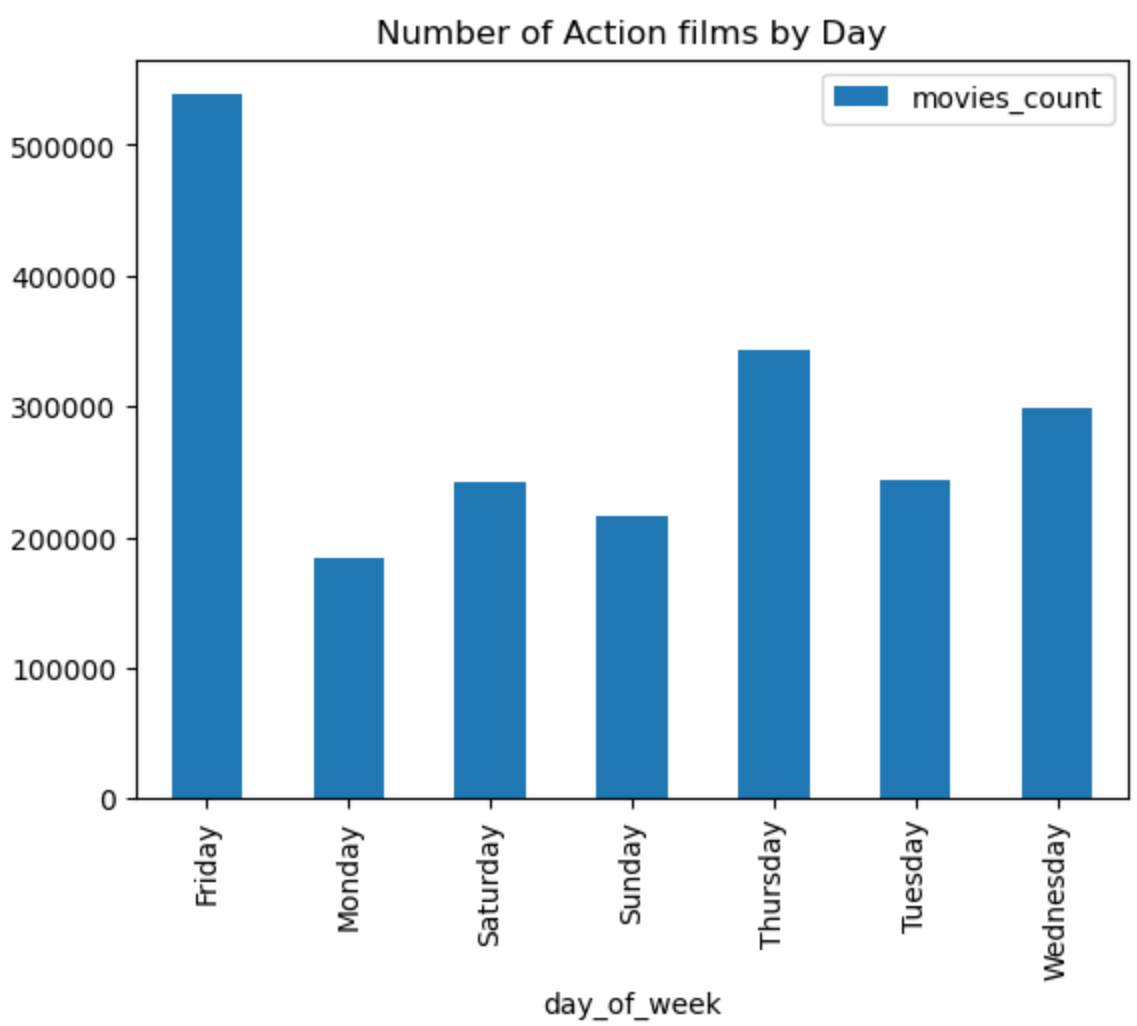


Make a bar plot of the day of the week on which movies with "Action" in their title tend to be released in the USA.

In [49]:
```
actfilm = cast[cast['title'].str.contains('action', na=False, case=False)]
actfilmmrel = pd.merge(actfilm, release, on='year', how='inner').fillna(0)
actfilmmrel_count = actfilmmrel.groupby('day_of_week').size().reset_index(name='movies_c

actfilmmrel_count.plot(x='day_of_week', y='movies_count', kind='bar', title='Number of A
```

Out[49]:
```
<Axes: title={'center': 'Number of Action films by Day'}, xlabel='day_of_week'>
```



On which date was each Judi Dench movie from the 1990s released in the USA?

In [69]:
```
judidenchfilm = cast[cast['name'].str.contains('Judi Dench', na=False, case=False)]
judidenchfilmrel = pd.merge(judidenchfilm, release, on='year', how='inner').fillna(0)
judidenchfilmus=judidenchfilmrel[judidenchfilmrel['production_countries'].str.contains('


judidenchfilmus_90s = judidenchfilmus[(judidenchfilmus['release_date'].dt.year >= 1990)


judidenchfilmus_90s = judidenchfilmus_90s.sort_values('release_date')
judidenchfilmus_90s_count = judidenchfilmus_90s.groupby('release_date').size().reset_ind
judidenchfilmus_90s_count
```

Out[69]:

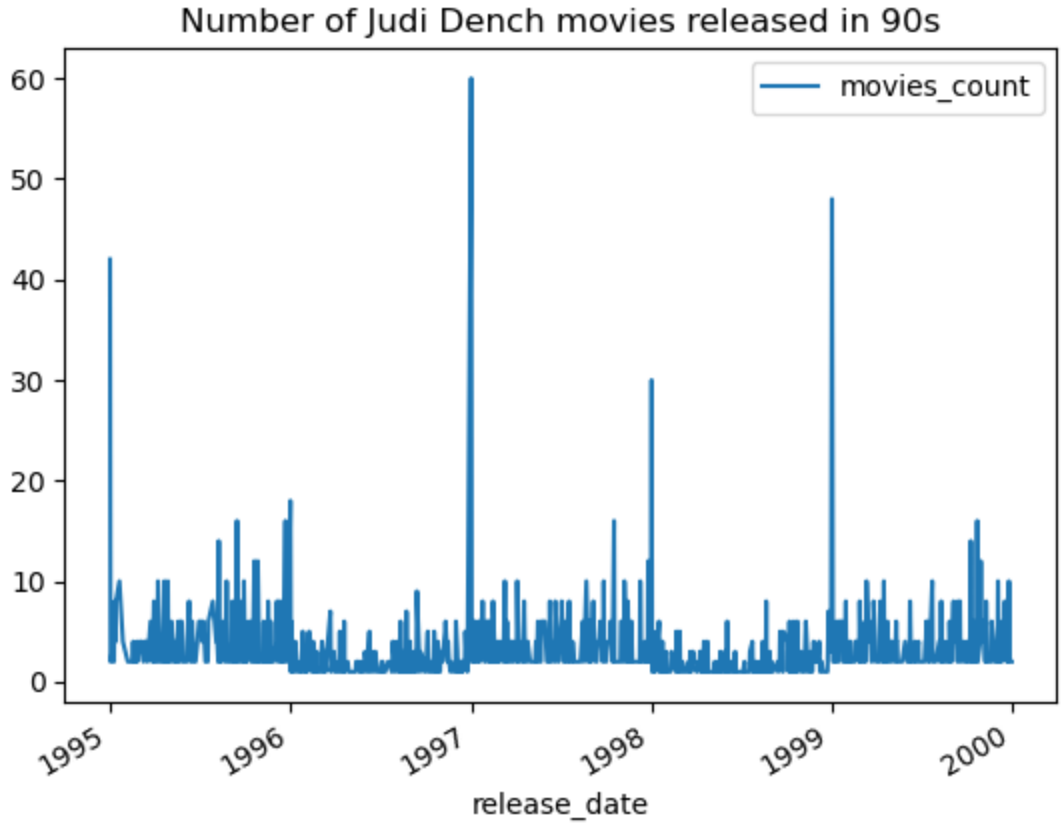
	release_date	movies_count
0	1995-01-01	42
1	1995-01-02	2
2	1995-01-03	2
3	1995-01-08	4
4	1995-01-09	2
...	...	...
865	1999-12-22	6
866	1999-12-23	2

867	1999-12-25	10
868	1999-12-27	2
869	1999-12-31	2

870 rows × 2 columns

In [70]:
```
judidenchfilmus_90s_count.plot(x='release_date', y='movies_count', kind='line', title='N
```

Out[70]:
```
<Axes: title={'center': 'Number of Judi Dench movies released in 90s'}, xlabel='release_
date'>
```

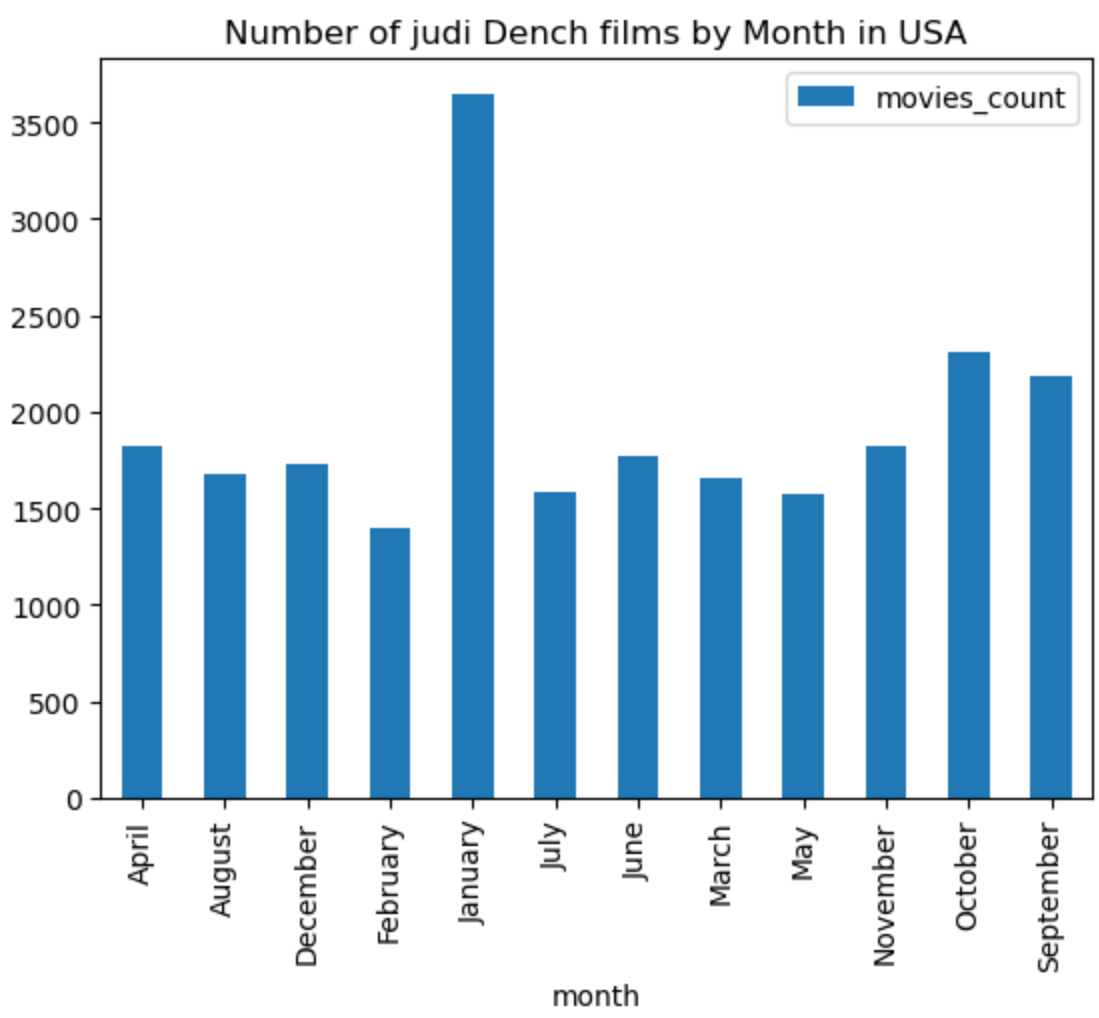


In which months do films with Judi Dench tend to be released in the USA?

In [55]:
```
judidenchfilm = cast[cast['name'].str.contains('Judi Dench', na=False, case=False)]
judidenchfilmrel = pd.merge(judidenchfilm, release, on='year', how='inner').fillna(0)
judidenchfilmus=judidenchfilmrel[judidenchfilmrel['production_countries'].str.contains('
judidenchfilmus_count = judidenchfilmus.groupby('month').size().reset_index(name='movies

judidenchfilmus_count.plot(x='month', y='movies_count', kind='bar', title='Number of jud
```

Out[55]:
```
<Axes: title={'center': 'Number of judi Dench films by Month in USA'}, xlabel='month'>
```

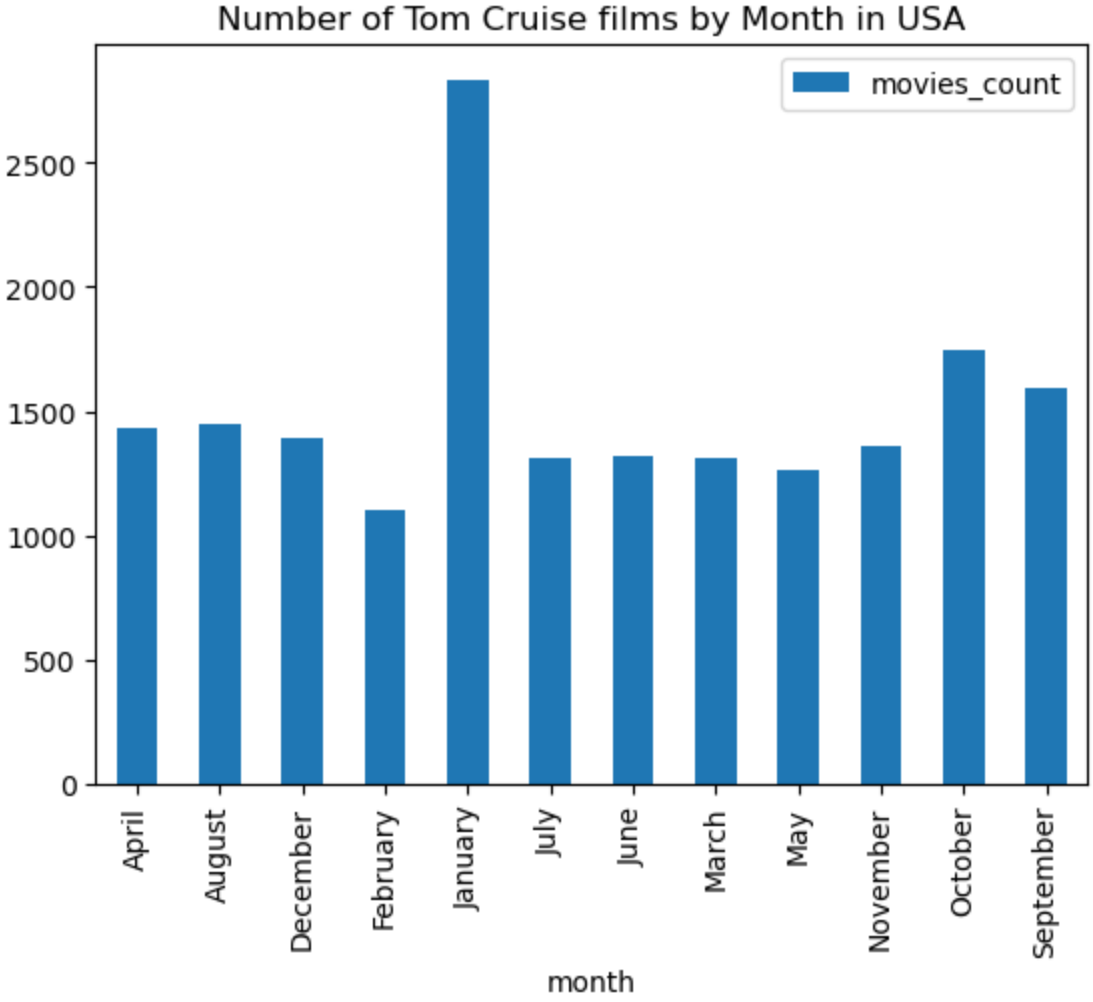


In which months do films with Tom Cruise tend to be released in the USA?

In [54]:
```
tomcruisefilm = cast[cast['name'].str.contains('Tom Cruise', na=False, case=False)]
tomcruisefilmrel = pd.merge(tomcruisefilm, release, on='year', how='inner').fillna(0)
tomcruisefilmus=tomcruisefilmrel[tomcruisefilmrel['production_countries'].str.contains('
tomcruisefilmus_count = tomcruisefilmus.groupby('month').size().reset_index(name='movies

tomcruisefilmus_count.plot(x='month', y='movies_count', kind='bar', title='Number of Tom
```

Out[54]:
```
<Axes: title={'center': 'Number of Tom Cruise films by Month in USA'}, xlabel='month'>
```


Number of Tom Cruise films by Month in USA
movies_count
2500
2000
1500
1000
500
0
April
August
December
February
January
July
June
March
May
November
October
September
month