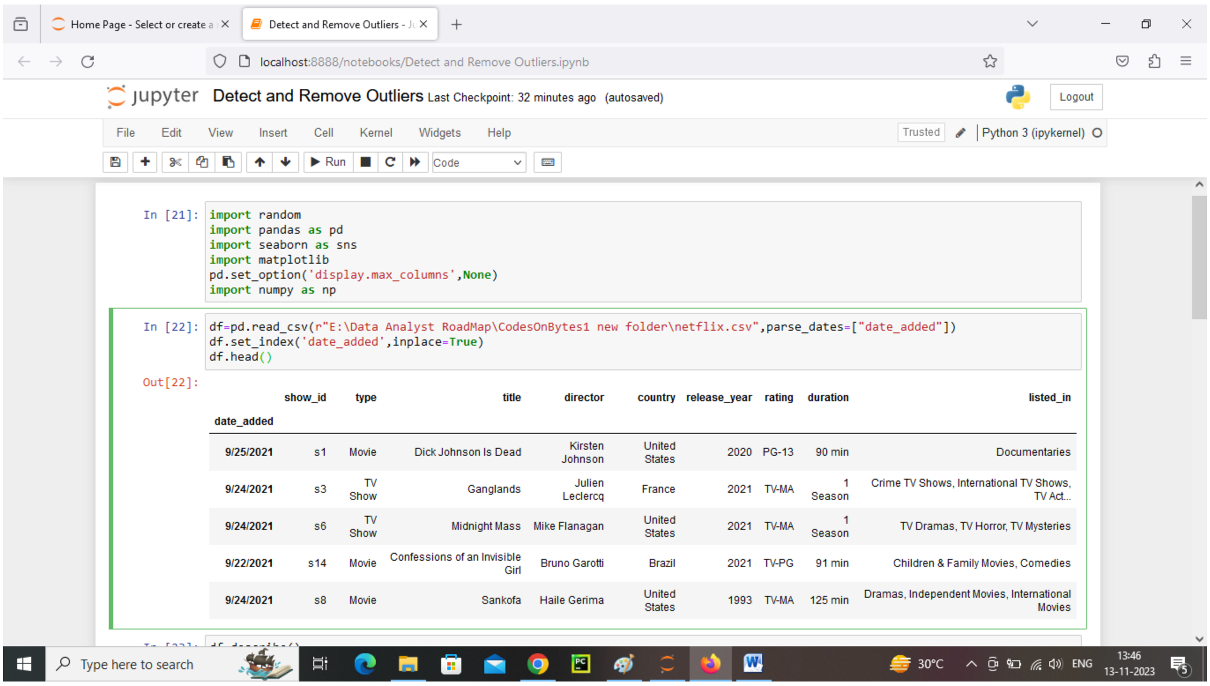
```
import random

import pandas as pd

import seaborn as sns

import matplotlib

pd.set_option('display.max_columns',None)

import numpy as np

df=pd.read_csv(r"E:\Data Analyst RoadMap\CodesOnBytes1 new folder\netflix.csv",parse_dates=["date_added"])

df.set_index('date_added',inplace=True)

df.head()
```



Home Page - Select or create a
Detect and Remove Outliers - J
localhost:8888/notebooks/Detect and Remove Outliers.ipynb
jupyter Detect and Remove Outliers Last Checkpoint: 32 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
Type here to search
30°C
ENG
13:46
13-11-2023

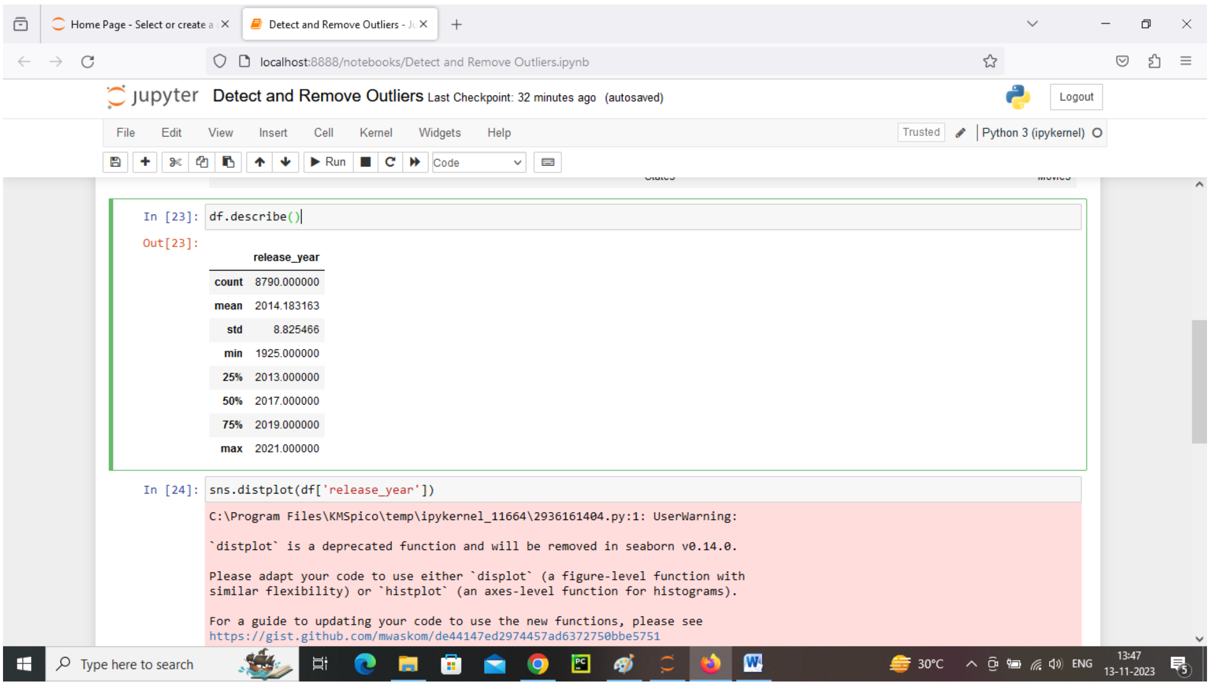
```
In [21]: import random
         import pandas as pd
         import seaborn as sns
         import matplotlib
         pd.set_option('display.max_columns',None)
         import numpy as np
```

```
In [22]: df=pd.read_csv(r"E:\Data Analyst RoadMap\CodesOnBytes1 new folder\netflix.csv",parse_dates=["date_added"])
         df.set_index('date_added',inplace=True)
         df.head()
```

Out[22]:

	show_id	type	title	director	country	release_year	rating	duration	listed_in
date_added									
9/25/2021	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	2020	PG-13	90 min	Documentaries
9/24/2021	s3	TV Show	Ganglands	Julien Leclercq	France	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...
9/24/2021	s6	TV Show	Midnight Mass	Mike Flanagan	United States	2021	TV-MA	1 Season	TV Dramas, TV Horror, TV Mysteries
9/22/2021	s14	Movie	Confessions of an Invisible Girl	Bruno Garotti	Brazil	2021	TV-PG	91 min	Children & Family Movies, Comedies
9/24/2021	s8	Movie	Sankofa	Haile Gerima	United States	1993	TV-MA	125 min	Dramas, Independent Movies, International Movies

df.describe()

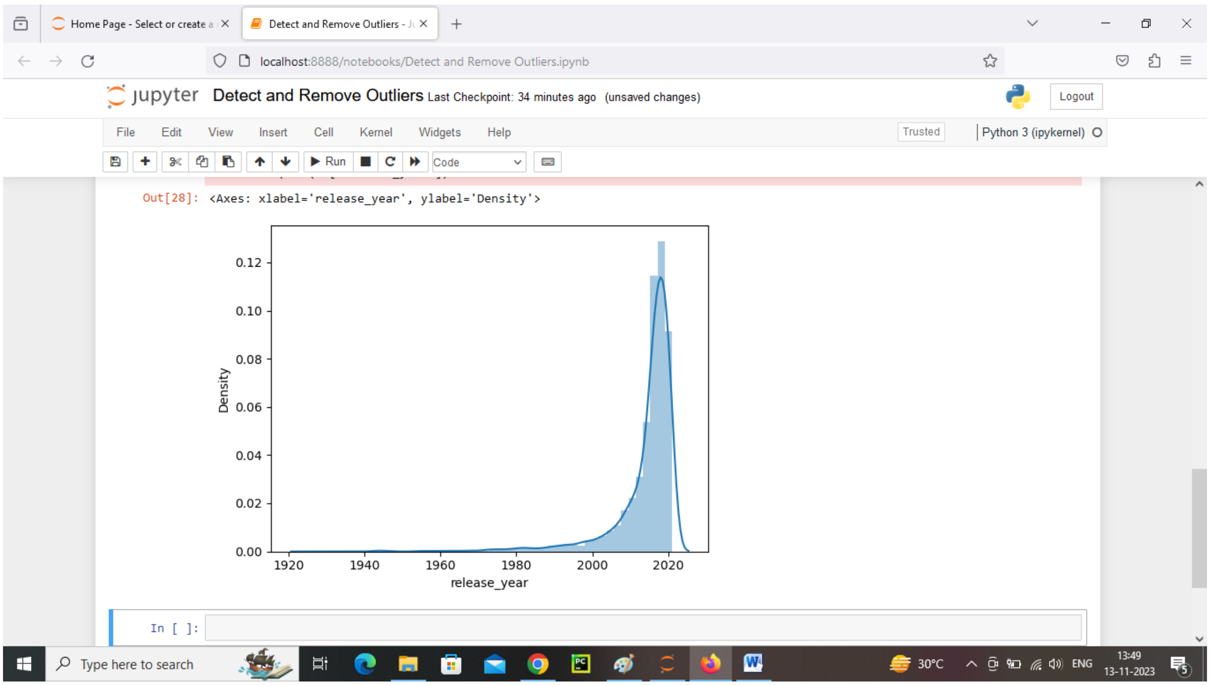

Home Page - Select or create a
Detect and Remove Outliers - J
localhost:8888/notebooks/Detect and Remove Outliers.ipynb
jupyter Detect and Remove Outliers Last Checkpoint: 32 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
Type here to search
30°C
ENG
13:47
13-11-2023

```
In [23]: df.describe()
Out[23]:
```

	release_year
count	8790.000000
mean	2014.183163
std	8.825466
min	1925.000000
25%	2013.000000
50%	2017.000000
75%	2019.000000
max	2021.000000

```
In [24]: sns.distplot(df['release_year'])

C:\Program Files\KMSpico\temp\ipykernel_11664\2936161404.py:1: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751
```

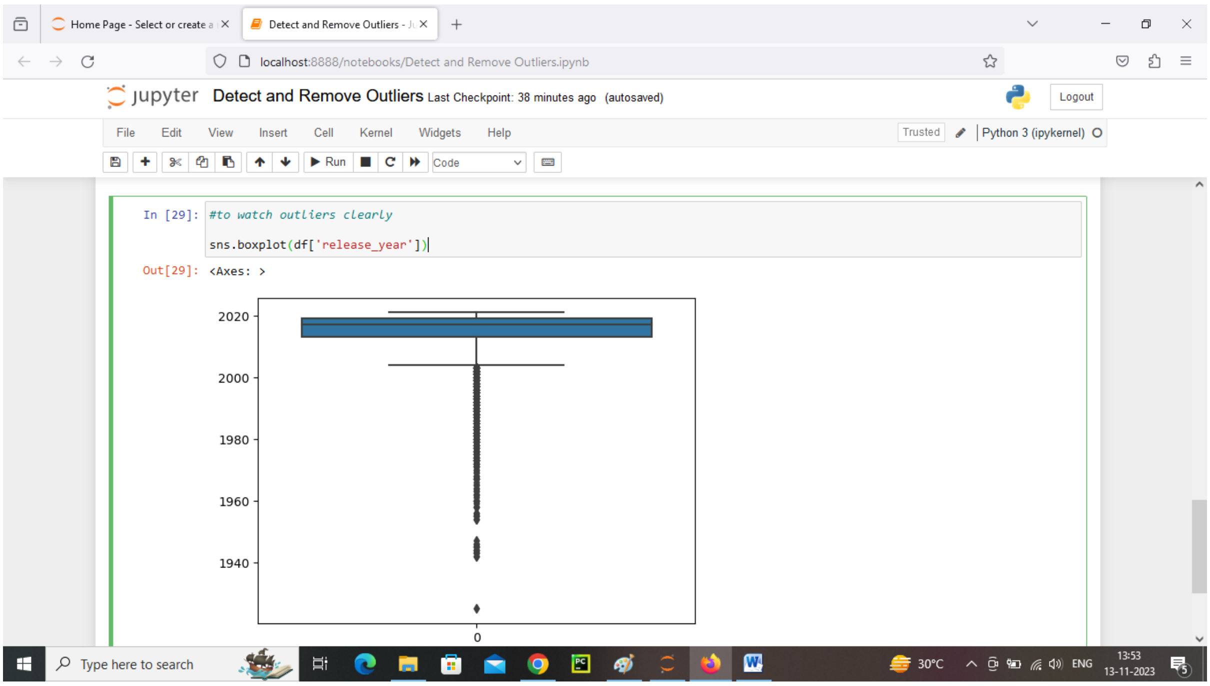
sns.distplot(df['release_year'])


Home Page - Select or create a
Detect and Remove Outliers - J
localhost:8888/notebooks/Detect and Remove Outliers.ipynb
jupyter Detect and Remove Outliers Last Checkpoint: 34 minutes ago (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
0.12
0.10
0.08
0.06
0.04
0.02
0.00
Density
1920
1940
1960
1980
2000
2020
release_year
Type here to search
30°C
ENG
13:49
13-11-2023

```
Out[28]: <Axes: xlabel='release_year', ylabel='Density'>

In [ ]:
```

#to watch outliers clearly

sns.boxplot(df['release_year'])



Home Page - Select or create a
Detect and Remove Outliers - J
localhost:8888/notebooks/Detect and Remove Outliers.ipynb
jupyter
Detect and Remove Outliers Last Checkpoint: 38 minutes ago (autosaved)
Logout
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Trusted
Python 3 (ipykernel)
Run
Code
2020
2000
1980
1960
1940
0
Type here to search
30°C
ENG
13:53
13-11-2023

```
In [29]: #to watch outliers clearly

         sns.boxplot(df['release_year'])

Out[29]: <Axes: >
```

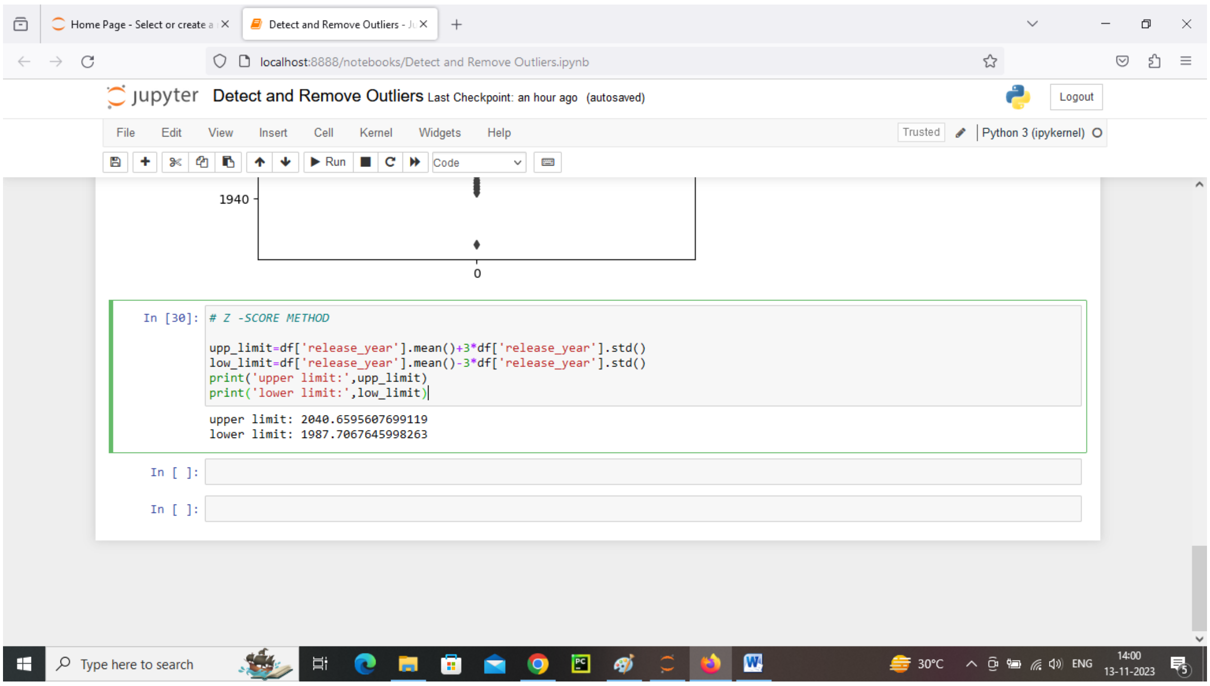
Z -SCORE METHOD

upp_limit=df['release_year'].mean()+3*df['release_year'].std()

low_limit=df['release_year'].mean()-3*df['release_year'].std()

print('upper limit:',upp_limit)

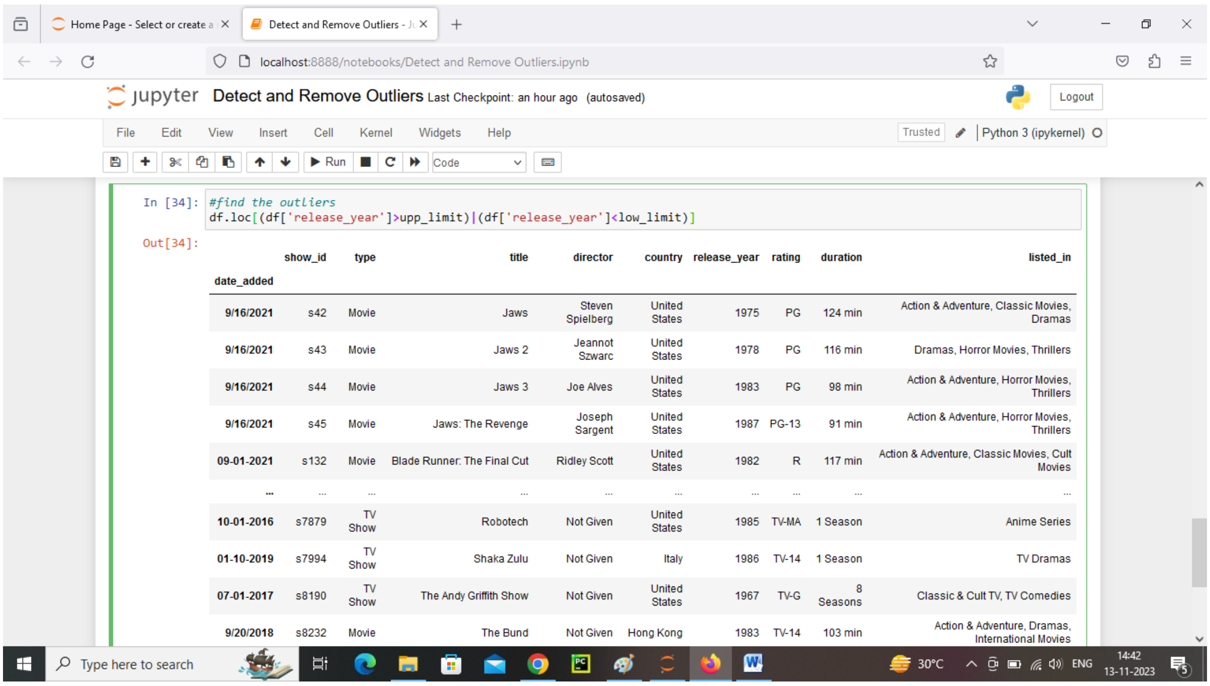
print('lower limit:',low_limit)


Home Page - Select or create a
Detect and Remove Outliers - J
localhost:8888/notebooks/Detect and Remove Outliers.ipynb
jupyter Detect and Remove Outliers Last Checkpoint: an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
1940
0
Type here to search
30°C
ENG
14:00
13-11-2023

```
In [30]: # Z -SCORE METHOD

         upp_limit=df['release_year'].mean()+3*df['release_year'].std()
         low_limit=df['release_year'].mean()-3*df['release_year'].std()
         print('upper limit:',upp_limit)
         print('lower limit:',low_limit)

         upper limit: 2040.6595607699119
         lower limit: 1987.7067645998263

In [ ]:

In [ ]:
```

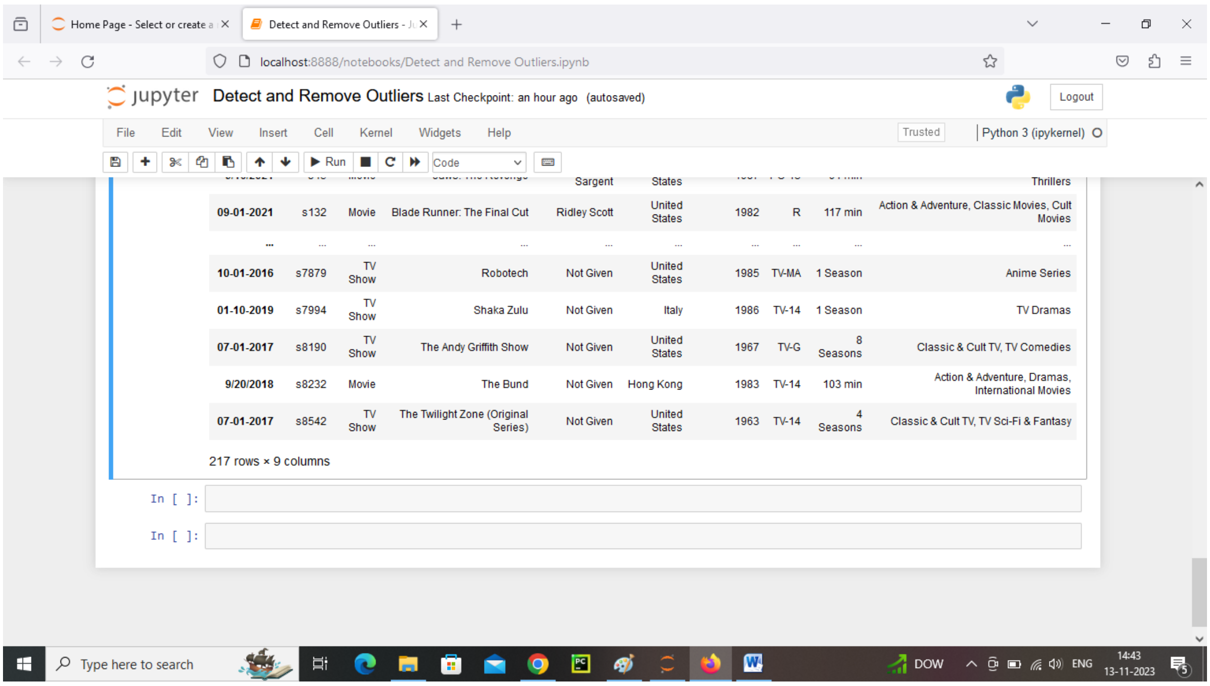
#find the outliers

df.loc[(df['release_year']>upp_limit)|(df['release_year']<low_limit)]


Home Page - Select or create a
Detect and Remove Outliers - J
localhost:8888/notebooks/Detect and Remove Outliers.ipynb
jupyter Detect and Remove Outliers Last Checkpoint: an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
Type here to search
30°C
ENG
14:42
13-11-2023

```
In [34]: #find the outliers
         df.loc[(df['release_year']>upp_limit)|(df['release_year']<low_limit)]

Out[34]:
```

	show_id	type	title	director	country	release_year	rating	duration	listed_in
date_added									
9/16/2021	s42	Movie	Jaws	Steven Spielberg	United States	1975	PG	124 min	Action & Adventure, Classic Movies, Dramas
9/16/2021	s43	Movie	Jaws 2	Jeannot Szwarc	United States	1978	PG	116 min	Dramas, Horror Movies, Thrillers
9/16/2021	s44	Movie	Jaws 3	Joe Alves	United States	1983	PG	98 min	Action & Adventure, Horror Movies, Thrillers
9/16/2021	s45	Movie	Jaws: The Revenge	Joseph Sargent	United States	1987	PG-13	91 min	Action & Adventure, Horror Movies, Thrillers
09-01-2021	s132	Movie	Blade Runner: The Final Cut	Ridley Scott	United States	1982	R	117 min	Action & Adventure, Classic Movies, Cult Movies
...	...	...	...	...	...	...	...	...	...
10-01-2016	s7879	TV Show	Robotech	Not Given	United States	1985	TV-MA	1 Season	Anime Series
01-10-2019	s7994	TV Show	Shaka Zulu	Not Given	Italy	1986	TV-14	1 Season	TV Dramas
07-01-2017	s8190	TV Show	The Andy Griffith Show	Not Given	United States	1967	TV-G	8 Seasons	Classic & Cult TV, TV Comedies
9/20/2018	s8232	Movie	The Bund	Not Given	Hong Kong	1983	TV-14	103 min	Action & Adventure, Dramas, International Movies



Home Page - Select or create a
Detect and Remove Outliers - J
localhost:8888/notebooks/Detect and Remove Outliers.ipynb
jupyter
Detect and Remove Outliers Last Checkpoint: an hour ago (autosaved)
Logout
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Trusted
Python 3 (ipykernel)
Run
Code
Type here to search
DOW
ENG
14:43
13-11-2023

				Sargent	States				Thrillers
09-01-2021	s132	Movie	Blade Runner: The Final Cut	Ridley Scott	United States	1982	R	117 min	Action & Adventure, Classic Movies, Cult Movies
...	...	...	...	...	...	...	...	...	...
10-01-2016	s7879	TV Show	Robotech	Not Given	United States	1985	TV-MA	1 Season	Anime Series
01-10-2019	s7994	TV Show	Shaka Zulu	Not Given	Italy	1986	TV-14	1 Season	TV Dramas
07-01-2017	s8190	TV Show	The Andy Griffith Show	Not Given	United States	1967	TV-G	8 Seasons	Classic & Cult TV, TV Comedies
9/20/2018	s8232	Movie	The Bund	Not Given	Hong Kong	1983	TV-14	103 min	Action & Adventure, Dramas, International Movies
07-01-2017	s8542	TV Show	The Twilight Zone (Original Series)	Not Given	United States	1963	TV-14	4 Seasons	Classic & Cult TV, TV Sci-Fi & Fantasy

217 rows × 9 columns

```
In [ ]:
In [ ]:
```

# trimming : Delete the outliers data

new_df=df.loc[(df['release_year']<upp_limit)&(df['release_year']>low_limit)]

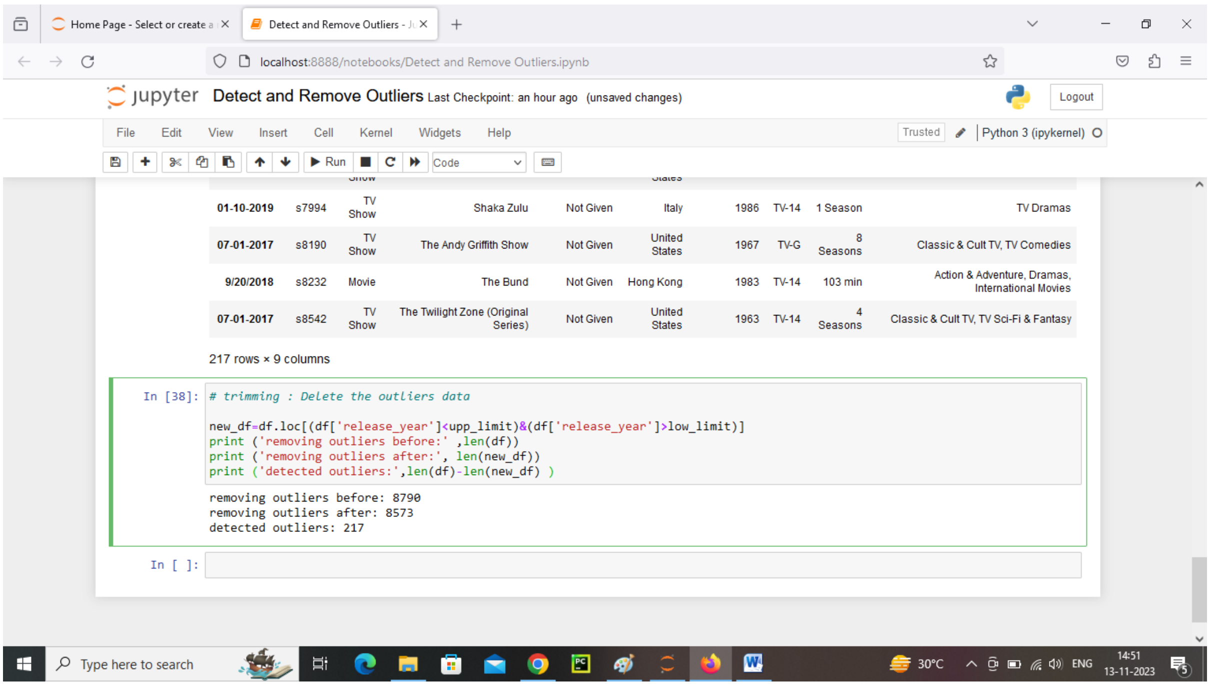
print ('removing outliers before:' ,len(df))

print ('removing outliers after:', len(new_df))

print ('detected outliers:',len(df)-len(new_df))

Output:

removing outliers before: 8790
removing outliers after: 8573
detected outliers: 217

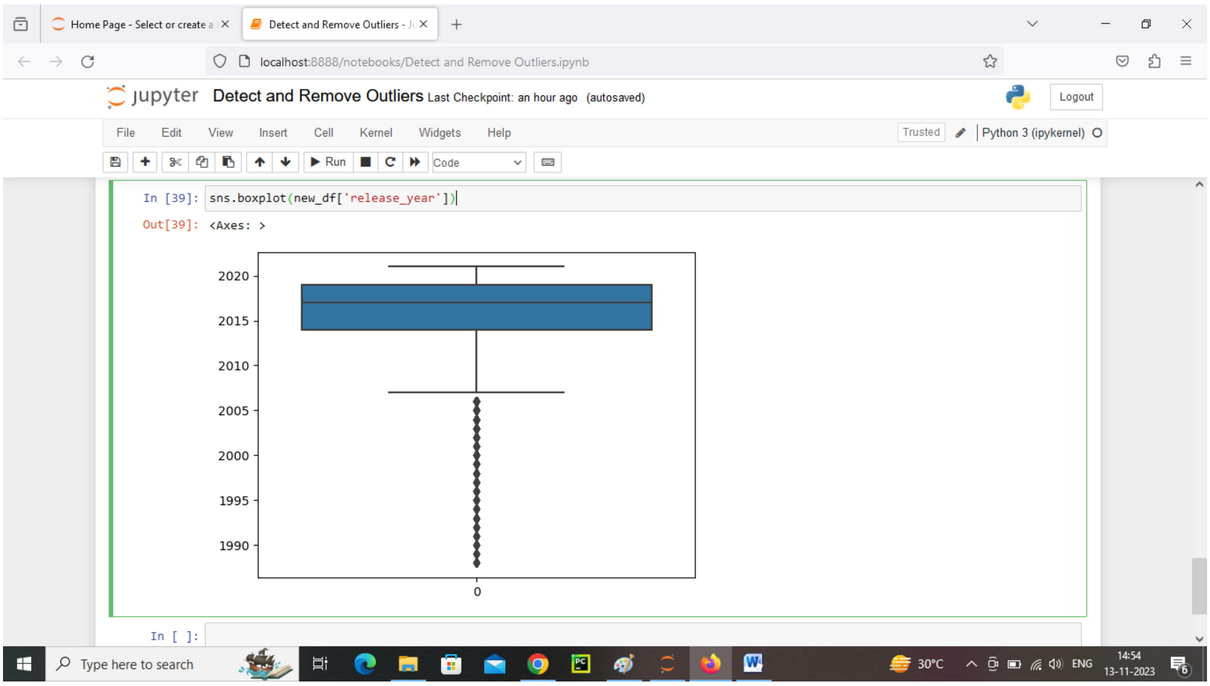


01-10-2019	s7994	TV Show	Shaka Zulu	Not Given	Italy	1986	TV-14	1 Season	TV Dramas
07-01-2017	s8190	TV Show	The Andy Griffith Show	Not Given	United States	1967	TV-G	8 Seasons	Classic & Cult TV, TV Comedies
9/20/2018	s8232	Movie	The Bund	Not Given	Hong Kong	1983	TV-14	103 min	Action & Adventure, Dramas, International Movies
07-01-2017	s8542	TV Show	The Twilight Zone (Original Series)	Not Given	United States	1963	TV-14	4 Seasons	Classic & Cult TV, TV Sci-Fi & Fantasy

217 rows × 9 columns

```
In [38]: # trimming : Delete the outliers data

         new_df=df.loc[(df['release_year']<upp_limit)&(df['release_year']>low_limit)]
         print ('removing outliers before:' ,len(df))
         print ('removing outliers after:', len(new_df))
         print ('detected outliers:',len(df)-len(new_df) )

         removing outliers before: 8790
         removing outliers after: 8573
         detected outliers: 217

In [ ]:
```

sns.boxplot(new_df['release_year'])

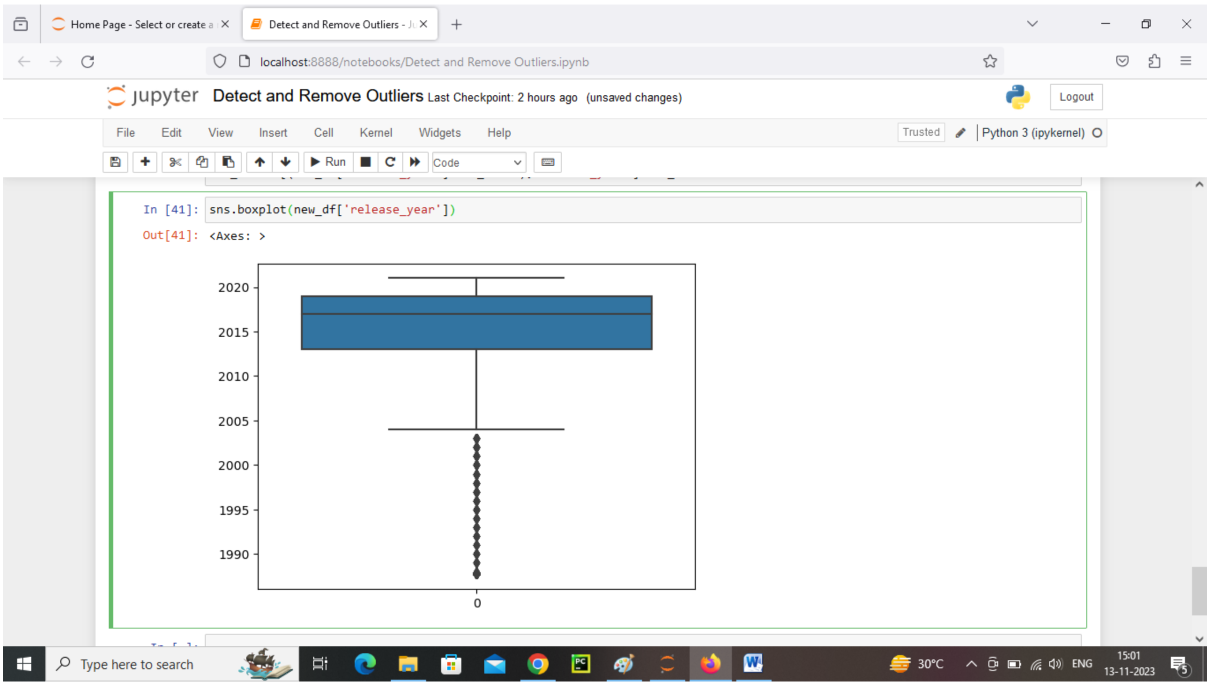


```
In [39]: sns.boxplot(new_df['release_year'])
Out[39]: <Axes: >

In [ ]:
```

#capping : Change the outlier values to Upper or lower limit values

new_df=df.copy()
new_df.loc[(new_df['release_year']>upp_limit),'release_year']=upp_limit
new_df.loc[(new_df['release_year']<low_limit),'release_year']=low_limit

sns.boxplot(new_df['release_year'])



```
In [41]: sns.boxplot(new_df['release_year'])
Out[41]: <Axes: >
```

Home Page - Select or create a
Detect and Remove Outliers - J
localhost:8888/notebooks/Detect and Remove Outliers.ipynb
jupyter Detect and Remove Outliers Last Checkpoint: 2 hours ago (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
2020
2015
2010
2005
2000
1995
1990
0
Type here to search
30°C
ENG
15:01
13-11-2023

len(new_df)

8790

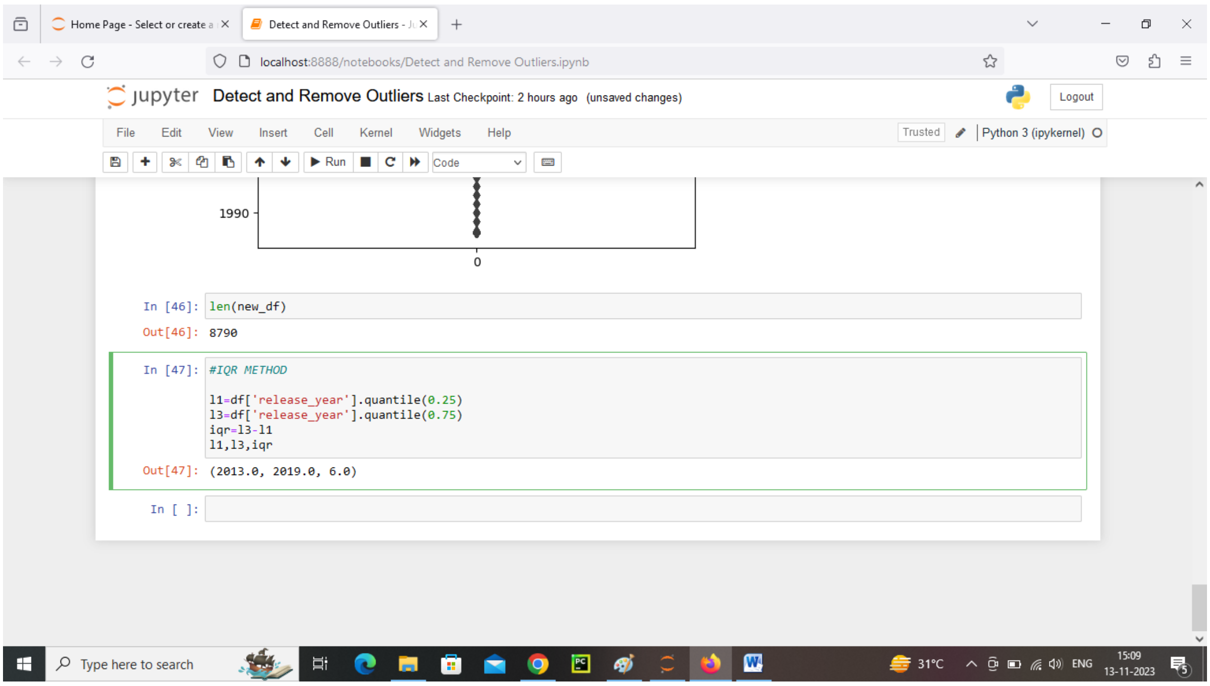
#IQR METHOD

l1=df['release_year'].quantile(0.25)

l3=df['release_year'].quantile(0.75)

iqr=l3-l1

l1,l3,iqr

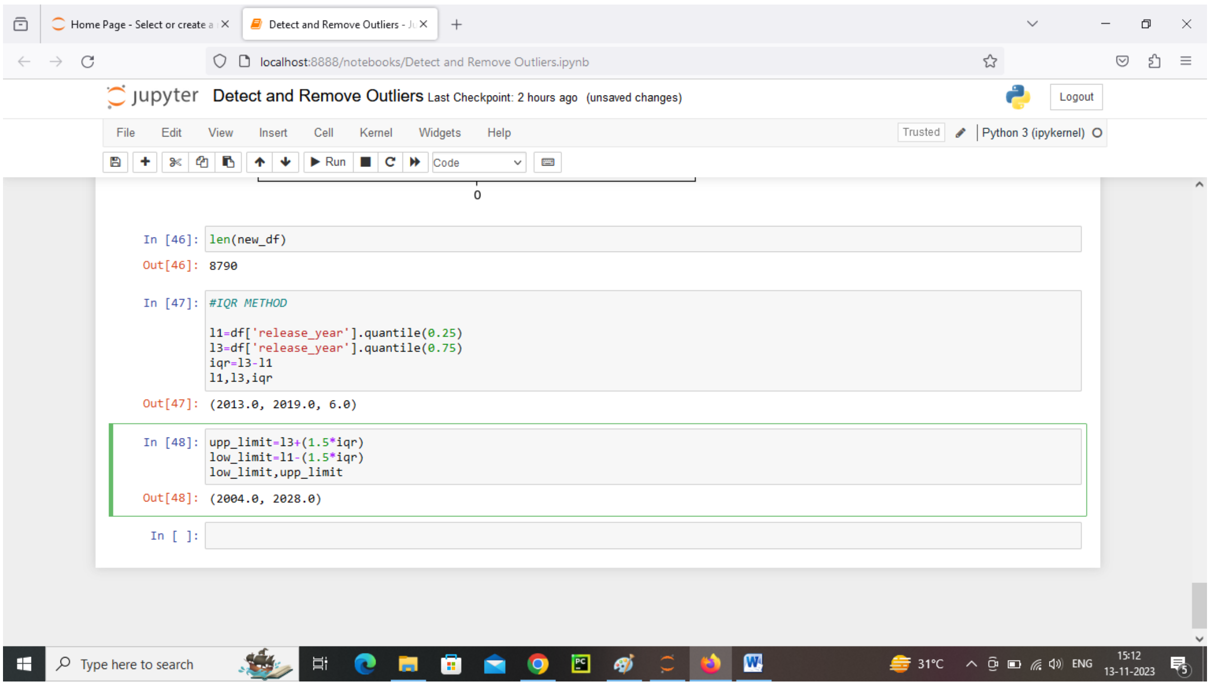

Home Page - Select or create a
Detect and Remove Outliers - J
localhost:8888/notebooks/Detect and Remove Outliers.ipynb
jupyter Detect and Remove Outliers Last Checkpoint: 2 hours ago (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
1990
0
Type here to search
31°C
ENG
15:09
13-11-2023

```
In [46]: len(new_df)
Out[46]: 8790

In [47]: #IQR METHOD

         l1=df['release_year'].quantile(0.25)
         l3=df['release_year'].quantile(0.75)
         iqr=l3-l1
         l1,l3,iqr
Out[47]: (2013.0, 2019.0, 6.0)

In [ ]:
```

upp_limit=l3+(1.5*iqr)

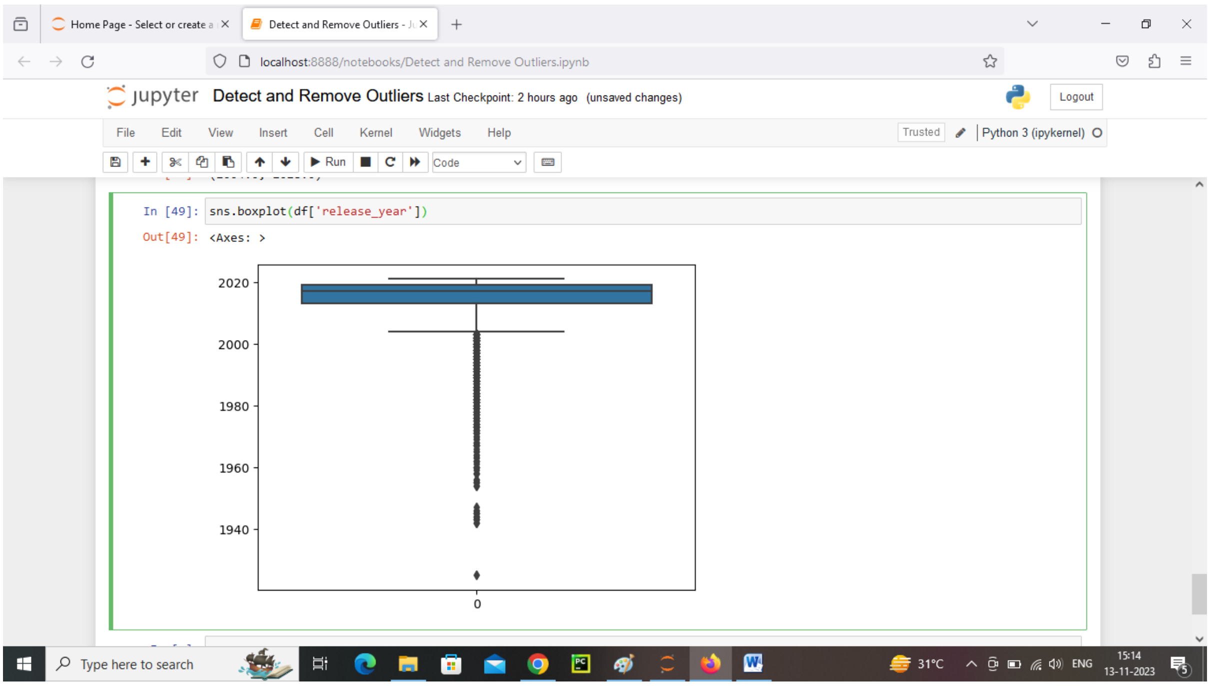
low_limit=l1-(1.5*iqr)

low_limit,upp_limit


Home Page - Select or create a
Detect and Remove Outliers - J
localhost:8888/notebooks/Detect and Remove Outliers.ipynb
jupyter Detect and Remove Outliers Last Checkpoint: 2 hours ago (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
0
Type here to search
31°C
ENG
15:12
13-11-2023

```
In [46]: len(new_df)
Out[46]: 8790

In [47]: #IQR METHOD

         l1=df['release_year'].quantile(0.25)
         l3=df['release_year'].quantile(0.75)
         iqr=l3-l1
         l1,l3,iqr
Out[47]: (2013.0, 2019.0, 6.0)

In [48]: upp_limit=l3+(1.5*iqr)
         low_limit=l1-(1.5*iqr)
         low_limit,upp_limit
Out[48]: (2004.0, 2028.0)

In [ ]:
```

sns.boxplot(df['release_year'])

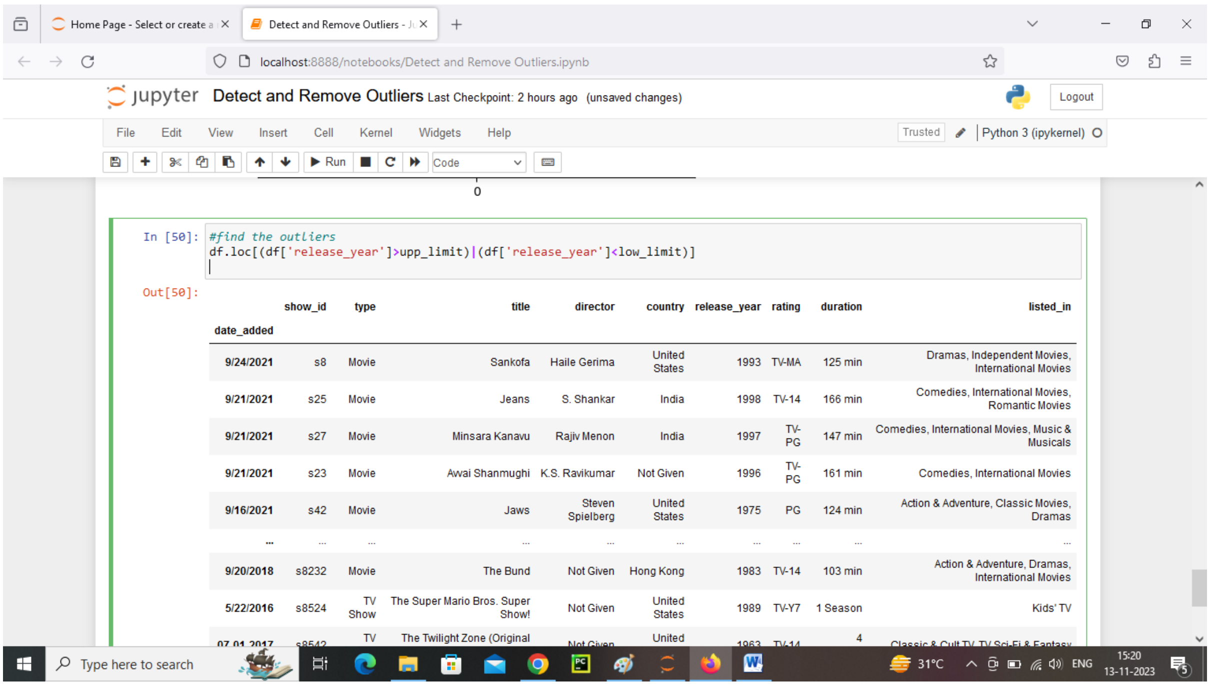


```
In [49]: sns.boxplot(df['release_year'])
Out[49]: <Axes: >
```

Home Page - Select or create a
Detect and Remove Outliers - J
localhost:8888/notebooks/Detect and Remove Outliers.ipynb
jupyter Detect and Remove Outliers Last Checkpoint: 2 hours ago (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
2020
2000
1980
1960
1940
0
Type here to search
31°C
ENG
15:14
13-11-2023

#find the outliers

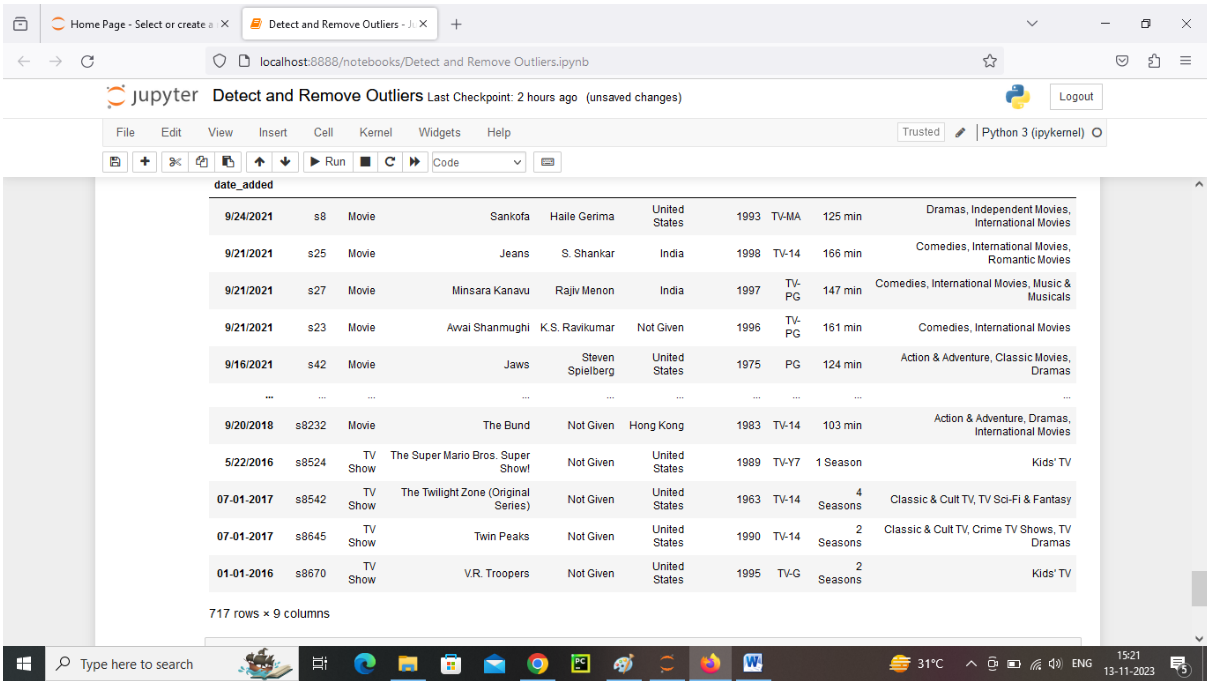
df.loc[(df['release_year']>upp_limit)|(df['release_year']<low_limit)]



```
In [50]: #find the outliers
         df.loc[(df['release_year']>upp_limit)|(df['release_year']<low_limit)]
Out[50]:
```

	show_id	type	title	director	country	release_year	rating	duration	listed_in
date_added									
9/24/2021	s8	Movie	Sankofa	Haile Gerima	United States	1993	TV-MA	125 min	Dramas, Independent Movies, International Movies
9/21/2021	s25	Movie	Jeans	S. Shankar	India	1998	TV-14	166 min	Comedies, International Movies, Romantic Movies
9/21/2021	s27	Movie	Minsara Kanavu	Rajiv Menon	India	1997	TV-PG	147 min	Comedies, International Movies, Music & Musicals
9/21/2021	s23	Movie	Awai Shanmughi	K.S. Ravikumar	Not Given	1996	TV-PG	161 min	Comedies, International Movies
9/16/2021	s42	Movie	Jaws	Steven Spielberg	United States	1975	PG	124 min	Action & Adventure, Classic Movies, Dramas
...	...	...	...	...	...	...	...	...	...
9/20/2018	s8232	Movie	The Bund	Not Given	Hong Kong	1983	TV-14	103 min	Action & Adventure, Dramas, International Movies
5/22/2016	s8524	TV Show	The Super Mario Bros. Super Show!	Not Given	United States	1989	TV-Y7	1 Season	Kids' TV
07 01 2017	s8542	TV	The Twilight Zone (Original	Not Given	United	1963	TV-14	4	Classic & Cult TV, TV Sci-Fi & Fantasy

Home Page - Select or create a
Detect and Remove Outliers - J
localhost:8888/notebooks/Detect and Remove Outliers.ipynb
jupyter Detect and Remove Outliers Last Checkpoint: 2 hours ago (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
Type here to search
31°C
ENG
15:20
13-11-2023


Home Page - Select or create a
Detect and Remove Outliers - J
localhost:8888/notebooks/Detect and Remove Outliers.ipynb
jupyter
Detect and Remove Outliers Last Checkpoint: 2 hours ago (unsaved changes)
Logout
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Trusted
Python 3 (ipykernel)
Run
Code
Type here to search
31°C
ENG
15:21
13-11-2023

date_added									
9/24/2021	s8	Movie	Sankofa	Haile Gerima	United States	1993	TV-MA	125 min	Dramas, Independent Movies, International Movies
9/21/2021	s25	Movie	Jeans	S. Shankar	India	1998	TV-14	166 min	Comedies, International Movies, Romantic Movies
9/21/2021	s27	Movie	Minsara Kanavu	Rajiv Menon	India	1997	TV-PG	147 min	Comedies, International Movies, Music & Musicals
9/21/2021	s23	Movie	Awai Shanmughi	K.S. Ravikumar	Not Given	1996	TV-PG	161 min	Comedies, International Movies
9/16/2021	s42	Movie	Jaws	Steven Spielberg	United States	1975	PG	124 min	Action & Adventure, Classic Movies, Dramas
...	...	...	...	...	...	...	...	...	...
9/20/2018	s8232	Movie	The Bund	Not Given	Hong Kong	1983	TV-14	103 min	Action & Adventure, Dramas, International Movies
5/22/2016	s8524	TV Show	The Super Mario Bros. Super Show!	Not Given	United States	1989	TV-Y7	1 Season	Kids' TV
07-01-2017	s8542	TV Show	The Twilight Zone (Original Series)	Not Given	United States	1963	TV-14	4 Seasons	Classic & Cult TV, TV Sci-Fi & Fantasy
07-01-2017	s8645	TV Show	Twin Peaks	Not Given	United States	1990	TV-14	2 Seasons	Classic & Cult TV, Crime TV Shows, TV Dramas
01-01-2016	s8670	TV Show	V.R. Troopers	Not Given	United States	1995	TV-G	2 Seasons	Kids' TV

717 rows × 9 columns

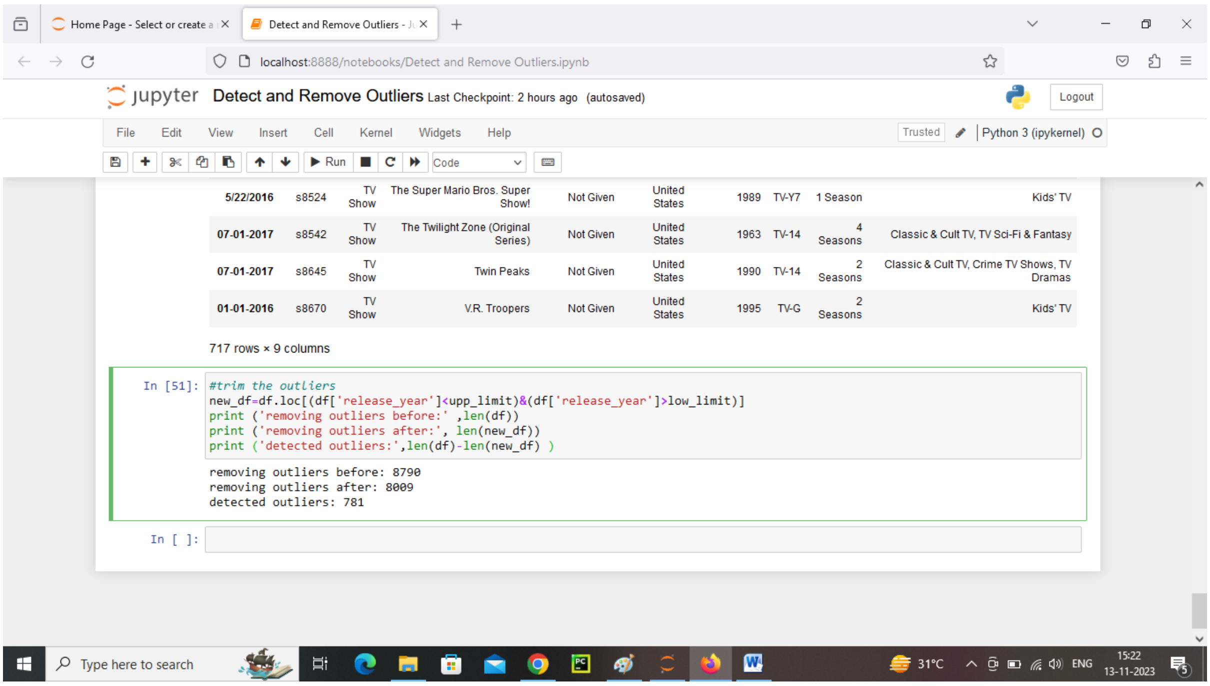
#trim the outliers

new_df=df.loc[(df['release_year']<upp_limit)&(df['release_year']>low_limit)]

print ('removing outliers before:' ,len(df))

print ('removing outliers after:', len(new_df))

print ('detected outliers:',len(df)-len(new_df))

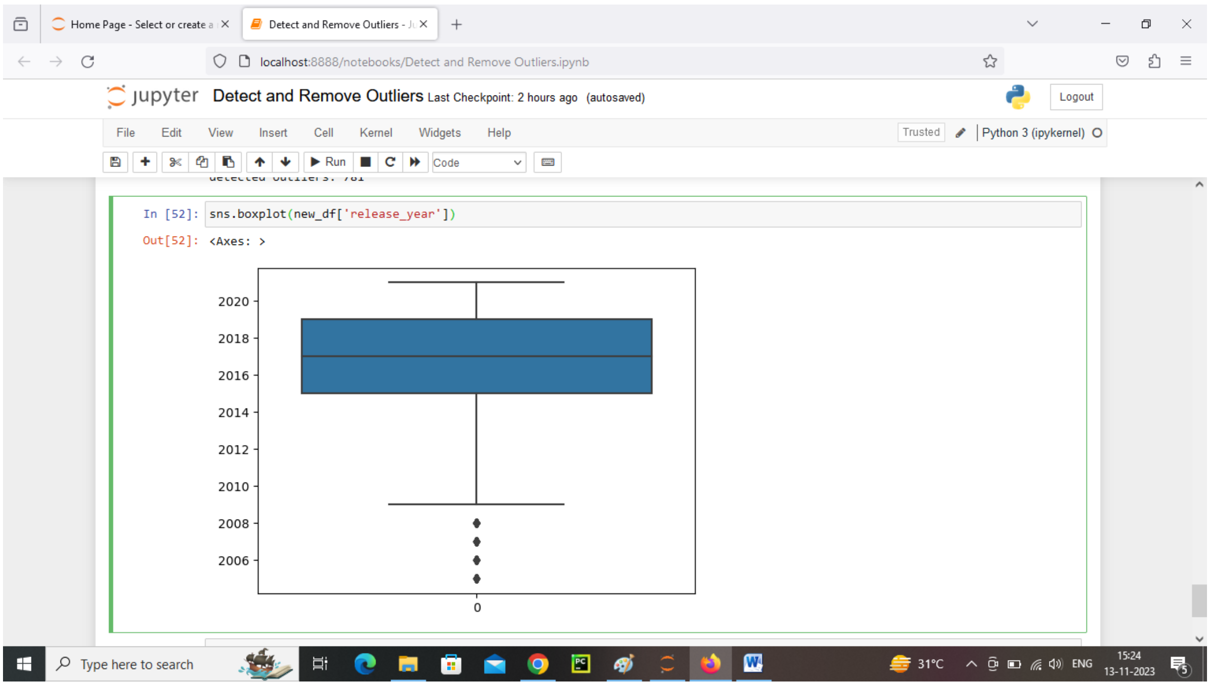

Home Page - Select or create a
Detect and Remove Outliers - J
localhost:8888/notebooks/Detect and Remove Outliers.ipynb
jupyter
Detect and Remove Outliers Last Checkpoint: 2 hours ago (autosaved)
Logout
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Trusted
Python 3 (ipykernel)
Run
Code
Type here to search
31°C
ENG
15:22
13-11-2023

5/22/2016	s8524	TV Show	The Super Mario Bros. Super Show!	Not Given	United States	1989	TV-Y7	1 Season	Kids' TV
07-01-2017	s8542	TV Show	The Twilight Zone (Original Series)	Not Given	United States	1963	TV-14	4 Seasons	Classic & Cult TV, TV Sci-Fi & Fantasy
07-01-2017	s8645	TV Show	Twin Peaks	Not Given	United States	1990	TV-14	2 Seasons	Classic & Cult TV, Crime TV Shows, TV Dramas
01-01-2016	s8670	TV Show	V.R. Troopers	Not Given	United States	1995	TV-G	2 Seasons	Kids' TV

717 rows × 9 columns

```
In [51]: #trim the outliers
         new_df=df.loc[(df['release_year']<upp_limit)&(df['release_year']>low_limit)]
         print ('removing outliers before:' ,len(df))
         print ('removing outliers after:', len(new_df))
         print ('detected outliers:',len(df)-len(new_df) )

         removing outliers before: 8790
         removing outliers after: 8009
         detected outliers: 781

In [ ]:
```

sns.boxplot(new_df['release_year'])



Home Page - Select or create a
Detect and Remove Outliers - J
localhost:8888/notebooks/Detect and Remove Outliers.ipynb
jupyter
Detect and Remove Outliers Last Checkpoint: 2 hours ago (autosaved)
Logout
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Trusted
Python 3 (ipykernel)
Run
Code
Type here to search
31°C
ENG
15:24
13-11-2023

```
In [52]: sns.boxplot(new_df['release_year'])
Out[52]: <Axes: >
```

2020
2018
2016
2014
2012
2010
2008
2006
0

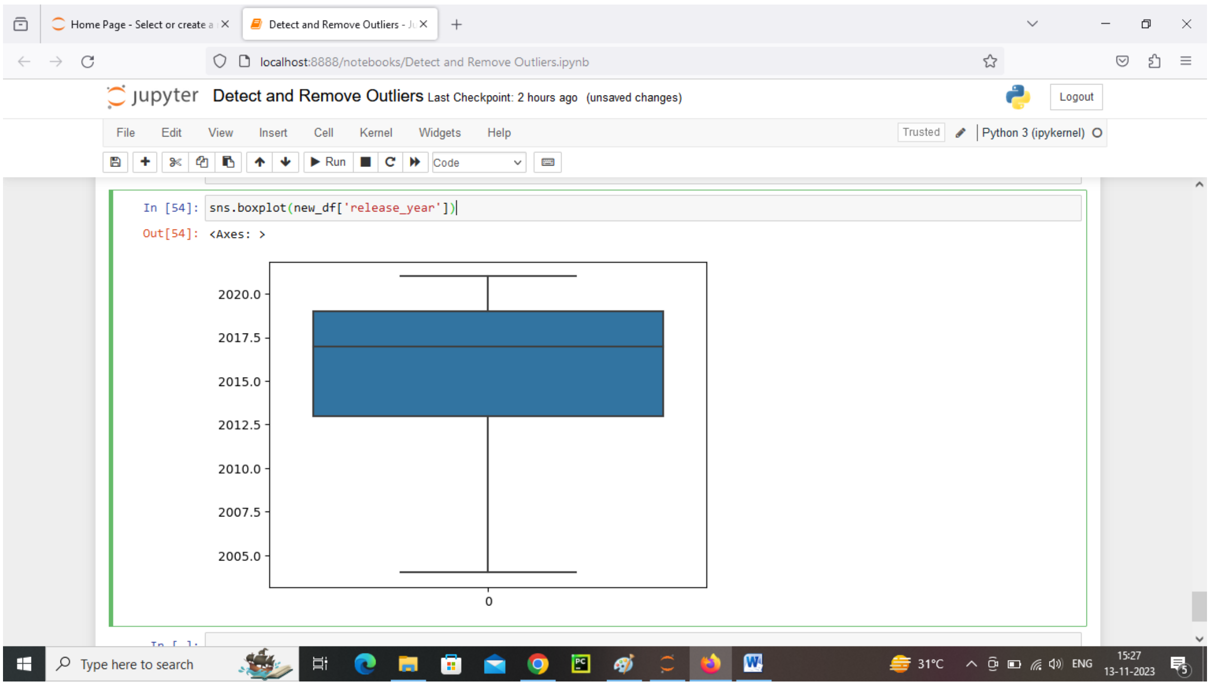
#capping the outliers

new_df=df.copy()

new_df.loc[(new_df['release_year']>upp_limit),'release_year']=upp_limit

new_df.loc[(new_df['release_year']<low_limit),'release_year']=low_limit

sns.boxplot(new_df['release_year'])



Home Page - Select or create a
Detect and Remove Outliers
localhost:8888/notebooks/Detect and Remove Outliers.ipynb
jupyter
Detect and Remove Outliers Last Checkpoint: 2 hours ago (unsaved changes)
Logout
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Trusted
Python 3 (ipykernel)
Run
Code
2020.0
2017.5
2015.0
2012.5
2010.0
2007.5
2005.0
0
Type here to search
31°C
ENG
15:27
13-11-2023

```
In [54]: sns.boxplot(new_df['release_year'])
Out[54]: <Axes: >
```

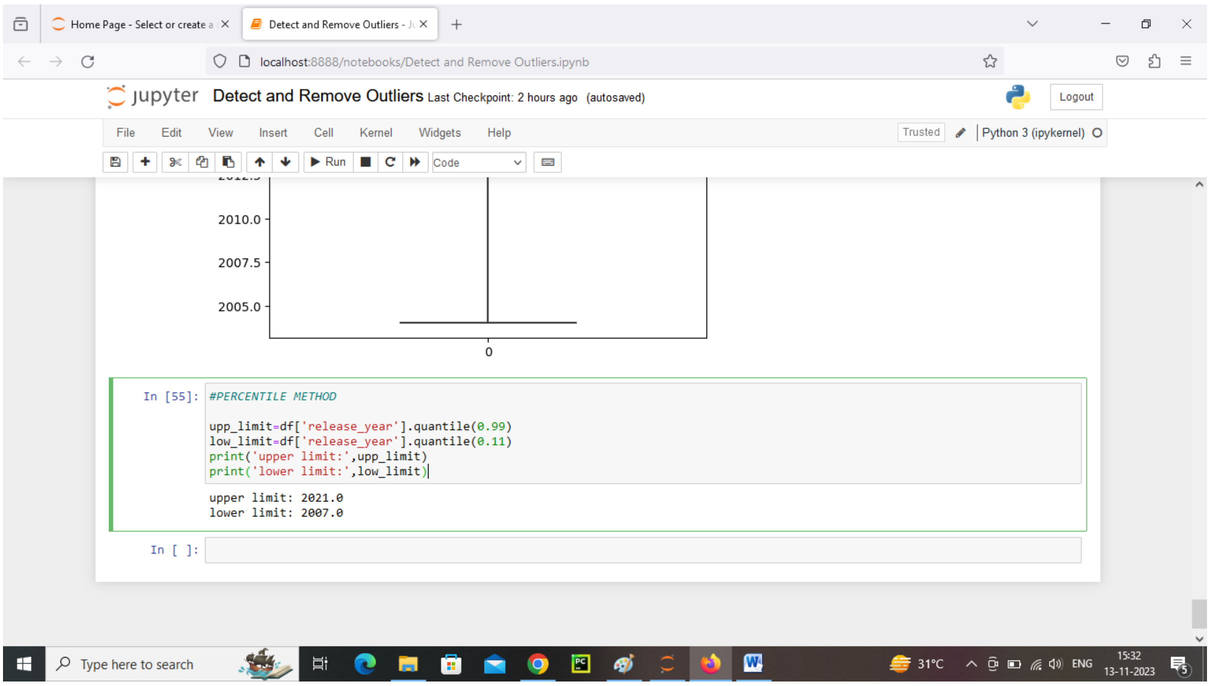
#PERCENTILE METHOD

upp_limit=df['release_year'].quantile(0.99)

low_limit=df['release_year'].quantile(0.11)

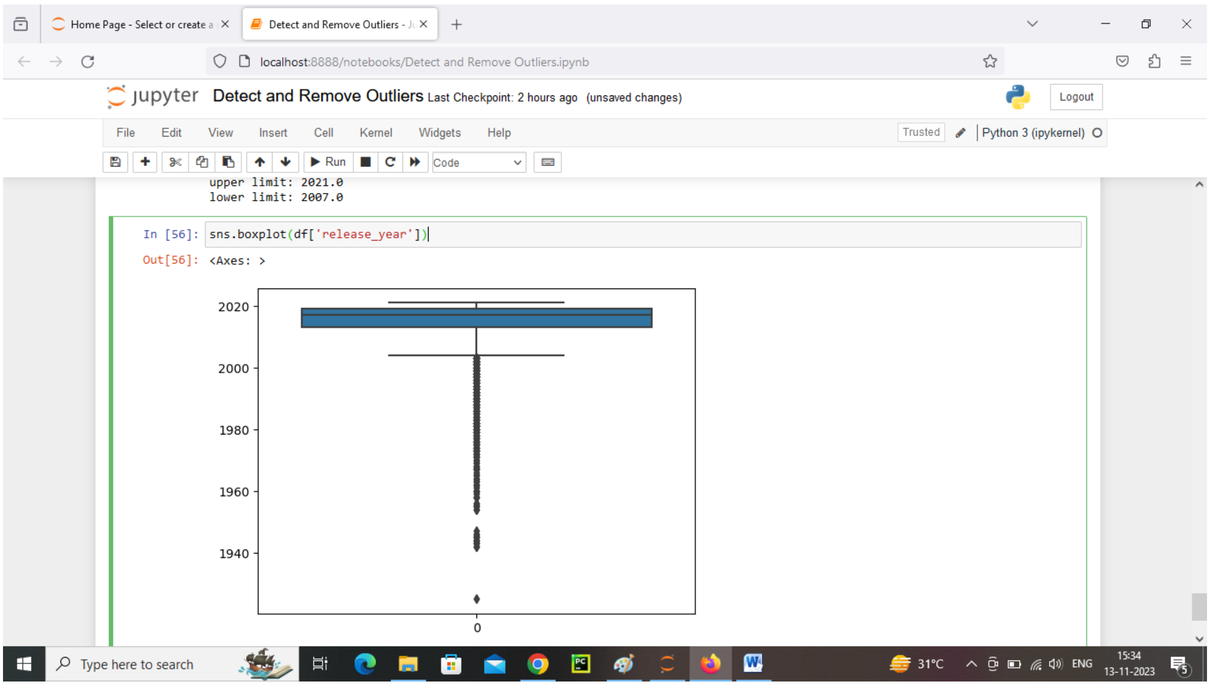
print('upper limit:',upp_limit)

print('lower limit:',low_limit)


Home Page - Select or create a
Detect and Remove Outliers - J
localhost:8888/notebooks/Detect and Remove Outliers.ipynb
jupyter Detect and Remove Outliers Last Checkpoint: 2 hours ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
2010.0
2007.5
2005.0
0
Type here to search
31°C
ENG
15:32
13-11-2023

```
In [55]: #PERCENTILE METHOD

         upp_limit=df['release_year'].quantile(0.99)
         low_limit=df['release_year'].quantile(0.11)
         print('upper limit:',upp_limit)
         print('lower limit:',low_limit)

         upper limit: 2021.0
         lower limit: 2007.0

In [ ]:
```

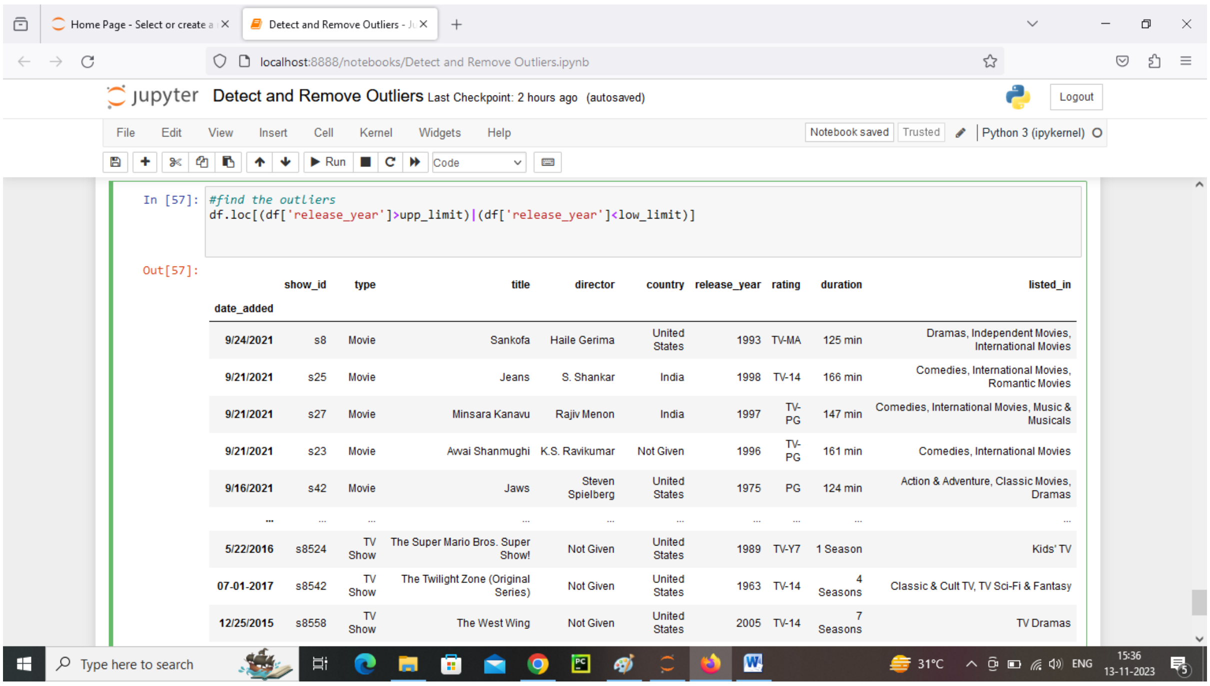
sns.boxplot(df['release_year'])


Home Page - Select or create a
Detect and Remove Outliers - J
localhost:8888/notebooks/Detect and Remove Outliers.ipynb
jupyter Detect and Remove Outliers Last Checkpoint: 2 hours ago (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
2020
2000
1980
1960
1940
0
Type here to search
31°C
ENG
15:34
13-11-2023

```
         upper limit: 2021.0
         lower limit: 2007.0

In [56]: sns.boxplot(df['release_year'])
Out[56]: <Axes: >
```

#find the outliers

df.loc[(df['release_year']>upp_limit)|(df['release_year']<low_limit)]


Home Page - Select or create a
Detect and Remove Outliers - J
localhost:8888/notebooks/Detect and Remove Outliers.ipynb
jupyter
Detect and Remove Outliers Last Checkpoint: 2 hours ago (autosaved)
Logout
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Notebook saved
Trusted
Python 3 (ipykernel)
Run
Code
Type here to search
31°C
ENG
15:36
13-11-2023

```
In [57]: #find the outliers
         df.loc[(df['release_year']>upp_limit)|(df['release_year']<low_limit)]
```

Out[57]:

	show_id	type	title	director	country	release_year	rating	duration	listed_in
date_added									
9/24/2021	s8	Movie	Sankofa	Haile Gerima	United States	1993	TV-MA	125 min	Dramas, Independent Movies, International Movies
9/21/2021	s25	Movie	Jeans	S. Shankar	India	1998	TV-14	166 min	Comedies, International Movies, Romantic Movies
9/21/2021	s27	Movie	Minsara Kanavu	Rajiv Menon	India	1997	TV-PG	147 min	Comedies, International Movies, Music & Musicals
9/21/2021	s23	Movie	Awai Shanmughi	K.S. Ravikumar	Not Given	1996	TV-PG	161 min	Comedies, International Movies
9/16/2021	s42	Movie	Jaws	Steven Spielberg	United States	1975	PG	124 min	Action & Adventure, Classic Movies, Dramas
...	...	...	...	...	...	...	...	...	...
5/22/2016	s8524	TV Show	The Super Mario Bros. Super Show!	Not Given	United States	1989	TV-Y7	1 Season	Kids' TV
07-01-2017	s8542	TV Show	The Twilight Zone (Original Series)	Not Given	United States	1963	TV-14	4 Seasons	Classic & Cult TV, TV Sci-Fi & Fantasy
12/25/2015	s8558	TV Show	The West Wing	Not Given	United States	2005	TV-14	7 Seasons	TV Dramas

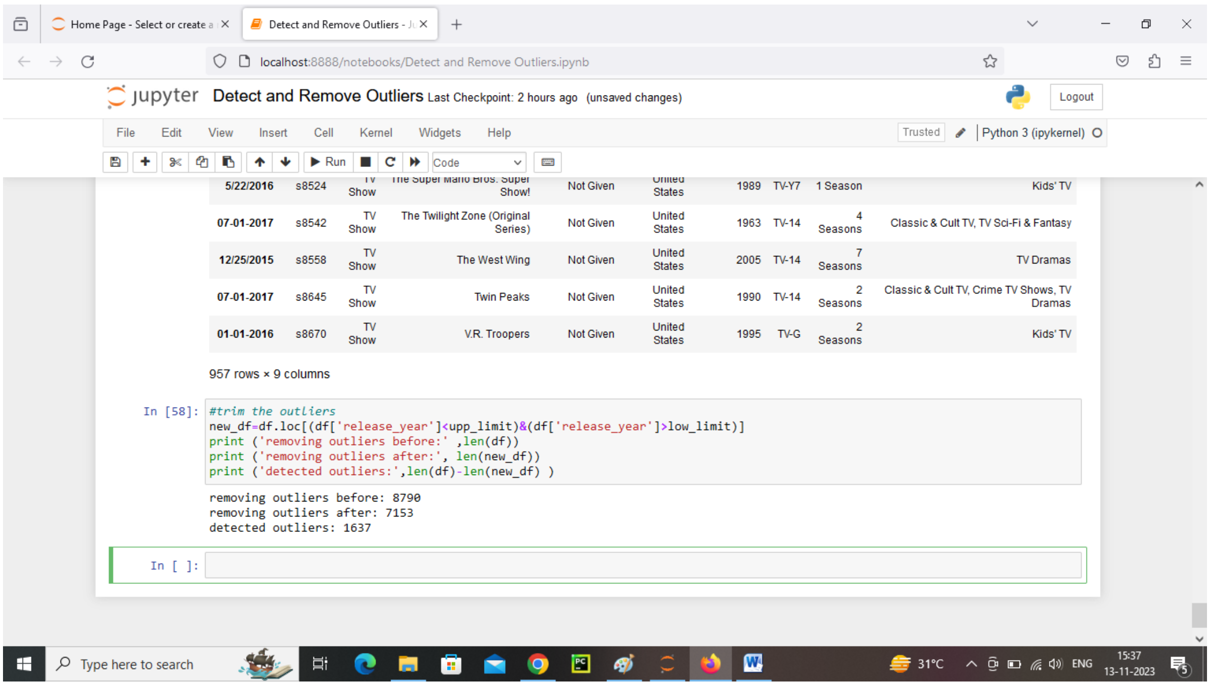
#trim the outliers

new_df=df.loc[(df['release_year']<upp_limit)&(df['release_year']>low_limit)]

print ('removing outliers before:' ,len(df))

print ('removing outliers after:', len(new_df))

print ('detected outliers:',len(df)-len(new_df))

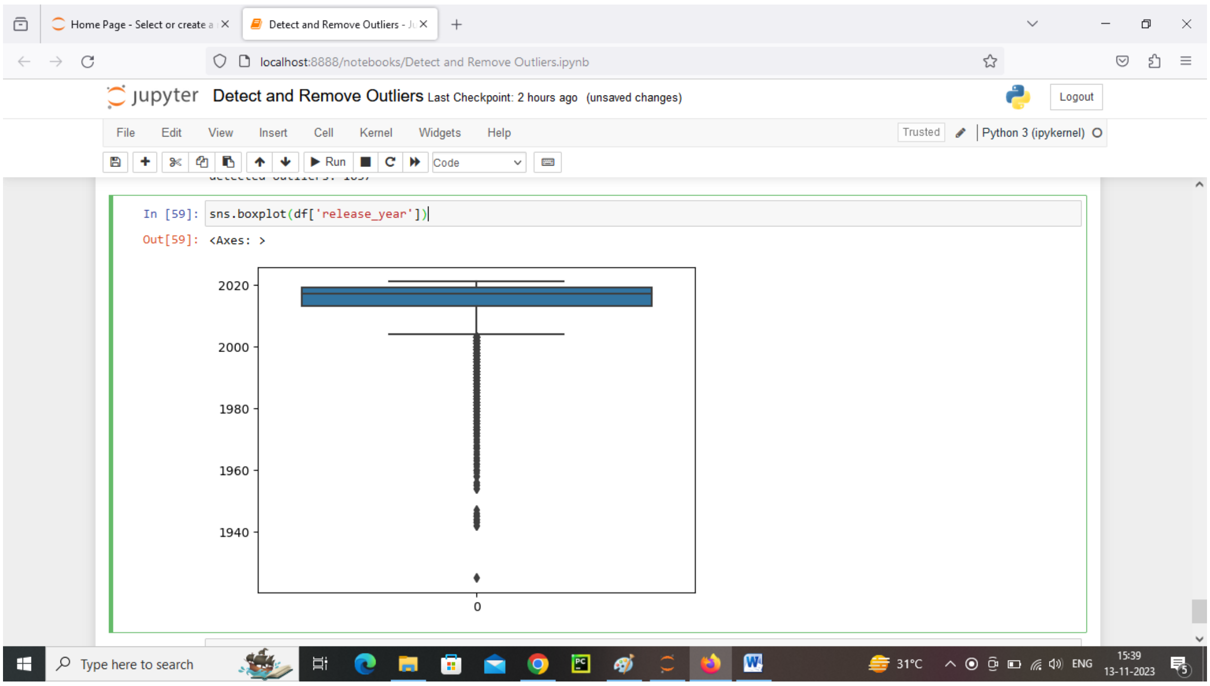

Home Page - Select or create a
Detect and Remove Outliers - J
localhost:8888/notebooks/Detect and Remove Outliers.ipynb
jupyter
Detect and Remove Outliers Last Checkpoint: 2 hours ago (unsaved changes)
Logout
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Trusted
Python 3 (ipykernel)
Run
Code
Type here to search
31°C
ENG
15:37
13-11-2023

	show_id	type	title	director	country	release_year	rating	duration	listed_in
5/22/2016	s8524	TV Show	The Super Mario Bros. Super Show!	Not Given	United States	1989	TV-Y7	1 Season	Kids' TV
07-01-2017	s8542	TV Show	The Twilight Zone (Original Series)	Not Given	United States	1963	TV-14	4 Seasons	Classic & Cult TV, TV Sci-Fi & Fantasy
12/25/2015	s8558	TV Show	The West Wing	Not Given	United States	2005	TV-14	7 Seasons	TV Dramas
07-01-2017	s8645	TV Show	Twin Peaks	Not Given	United States	1990	TV-14	2 Seasons	Classic & Cult TV, Crime TV Shows, TV Dramas
01-01-2016	s8670	TV Show	V.R. Troopers	Not Given	United States	1995	TV-G	2 Seasons	Kids' TV

957 rows × 9 columns

```
In [58]: #trim the outliers
         new_df=df.loc[(df['release_year']<upp_limit)&(df['release_year']>low_limit)]
         print ('removing outliers before:' ,len(df))
         print ('removing outliers after:', len(new_df))
         print ('detected outliers:',len(df)-len(new_df) )

         removing outliers before: 8790
         removing outliers after: 7153
         detected outliers: 1637

In [ ]:
```

sns.boxplot(df['release_year'])



Home Page - Select or create a
Detect and Remove Outliers - Ju
localhost:8888/notebooks/Detect and Remove Outliers.ipynb
jupyter Detect and Remove Outliers Last Checkpoint: 2 hours ago (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
2020
2000
1980
1960
1940
0
Type here to search
31°C
ENG
15:39
13-11-2023

```
In [59]: sns.boxplot(df['release_year'])
Out[59]: <Axes: >
```

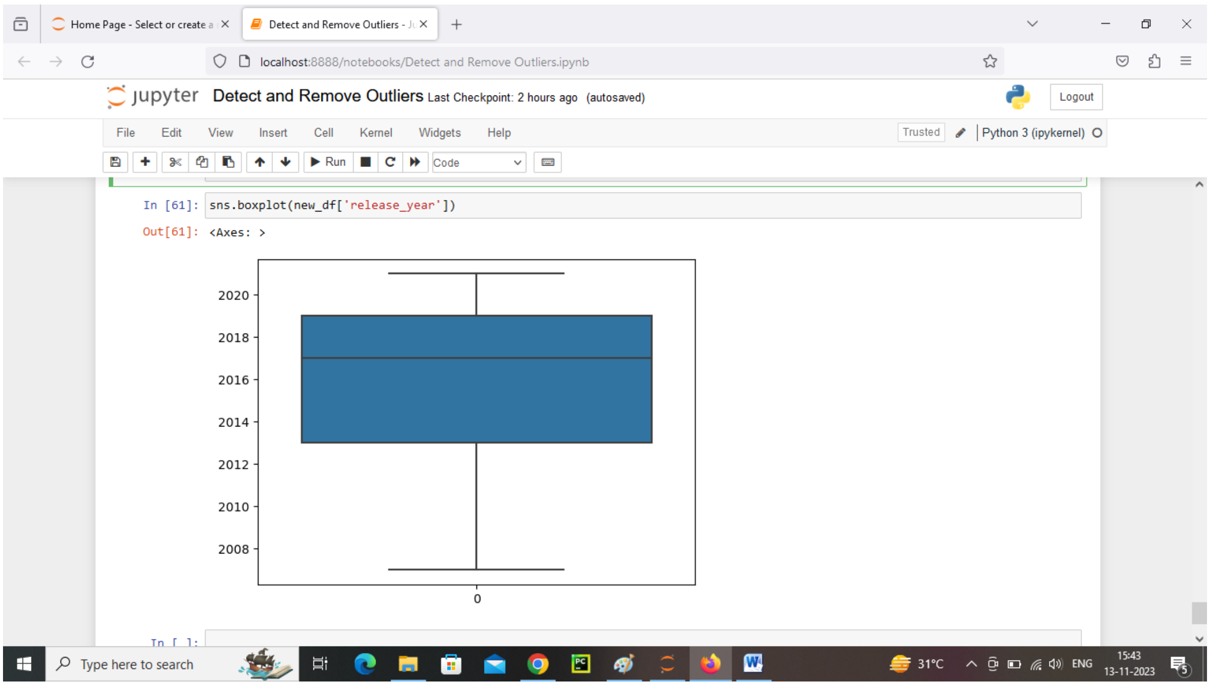
#capping the outliers

new_df=df.copy()

new_df.loc[(new_df['release_year']>upp_limit),'release_year']=upp_limit

new_df.loc[(new_df['release_year']<low_limit),'release_year']=low_limit

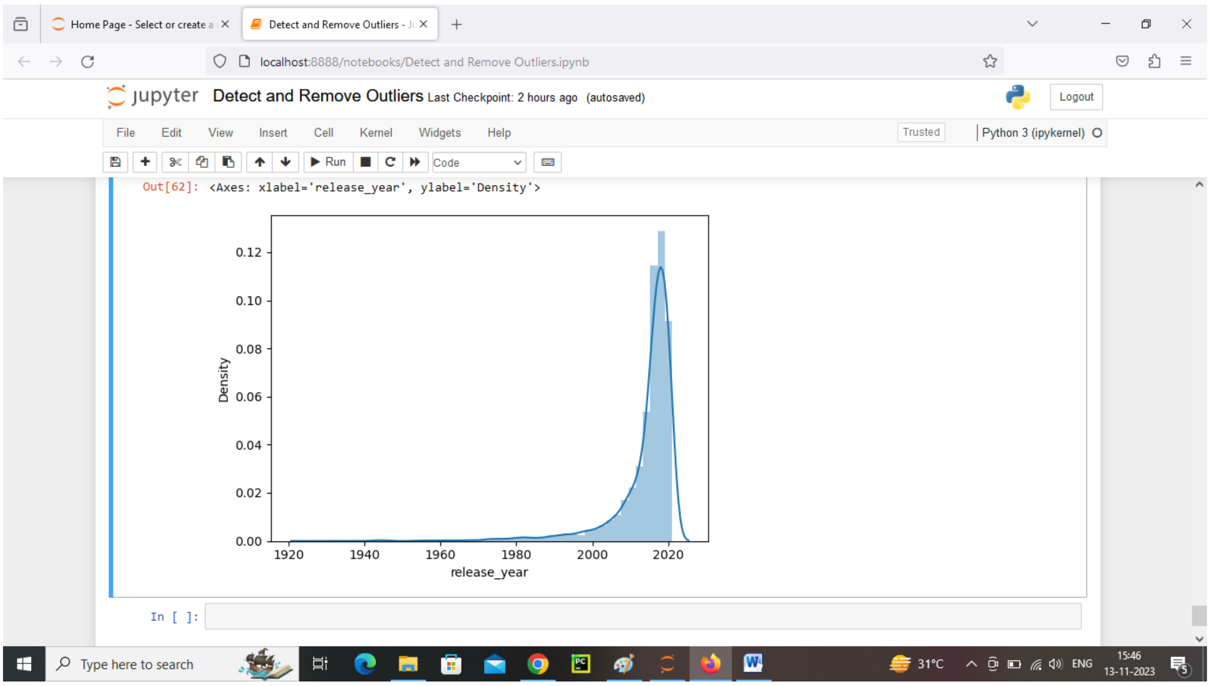
sns.boxplot(new_df['release_year'])



Home Page - Select or create a
Detect and Remove Outliers - Ju
localhost:8888/notebooks/Detect and Remove Outliers.ipynb
jupyter Detect and Remove Outliers Last Checkpoint: 2 hours ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
2020
2018
2016
2014
2012
2010
2008
0
Type here to search
31°C
ENG
15:43
13-11-2023

```
In [61]: sns.boxplot(new_df['release_year'])
Out[61]: <Axes: >
```

sns.distplot(df['release_year'])

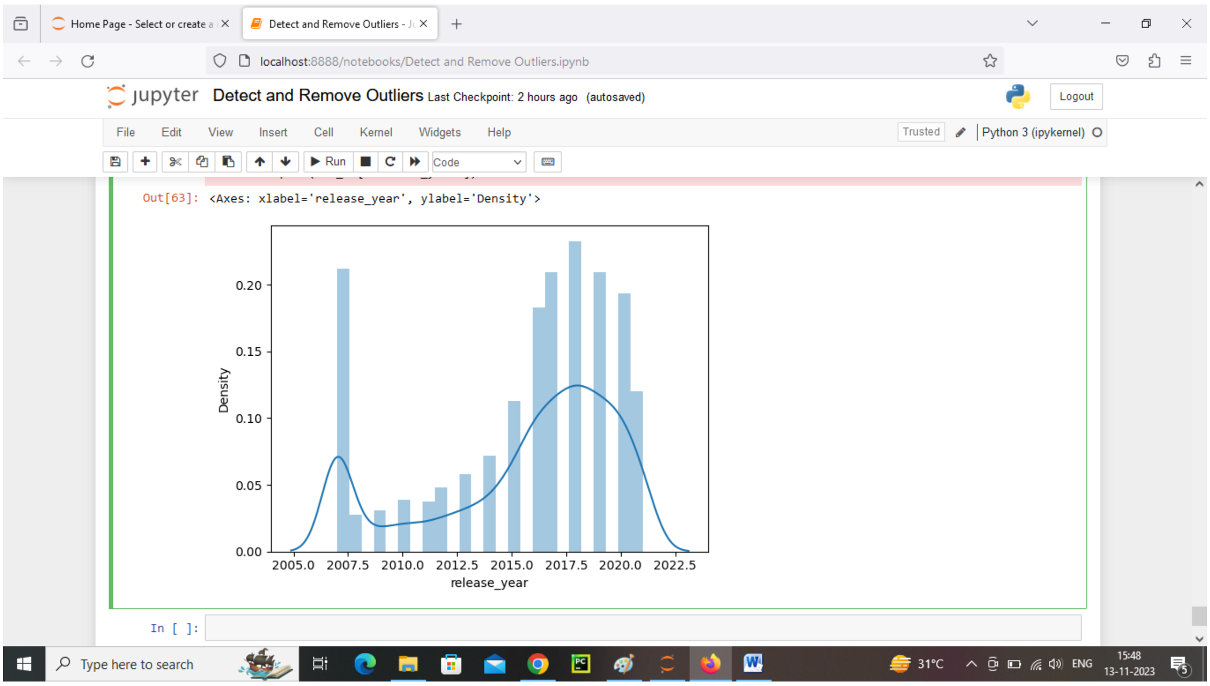


```
Out[62]: <Axes: xlabel='release_year', ylabel='Density'>
```

```
In [ ]:
```

Home Page - Select or create a
Detect and Remove Outliers - J
localhost:8888/notebooks/Detect and Remove Outliers.ipynb
jupyter Detect and Remove Outliers Last Checkpoint: 2 hours ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
0.12
0.10
0.08
0.06
0.04
0.02
0.00
Density
1920
1940
1960
1980
2000
2020
release_year
Type here to search
31°C
ENG
15:46
13-11-2023

sns.distplot(new_df['release_year'])



```
Out[63]: <Axes: xlabel='release_year', ylabel='Density'>
```

```
In [ ]:
```

Home Page - Select or create a
Detect and Remove Outliers - J
localhost:8888/notebooks/Detect and Remove Outliers.ipynb
jupyter Detect and Remove Outliers Last Checkpoint: 2 hours ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
0.20
0.15
0.10
0.05
0.00
Density
2005.0
2007.5
2010.0
2012.5
2015.0
2017.5
2020.0
2022.5
release_year
Type here to search
31°C
ENG
15:48
13-11-2023