```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.ticker as mtick

data =
pd.read_csv('/content/drive/MyDrive/Colab_Notebooks/DATA_211/Assignmen
ts/top_movies_data.csv')

data
```

```
           id  production_year
title  \
0     4103791             2006  Pirates of the Caribbean: Dead Man's
Chest
1     4022547             2006                          Night at the
Museum
2     3415583             2006
Cars
3     4587958             2006                       X-Men: The Last
Stand
4     4351263             2006                          The Da Vinci
Code
...       ...              ...
...
1190  3954779             2017                                Megan
Leavey
1191  4375083             2017                        The Greatest
Showman
1192  4367698             2016                                 The
Founder
1193  3342703             2017                         Battle of the
Sexes
1194  4332562             2017                                The
Beguiled

                           directors     primary_genre  \
0                    Verbinski, Gore  Action/Adventure
1                        Levy, Shawn  Action/Adventure
2           Lasseter, John|Ranft, Joe         Animation
3                      Ratner, Brett  Action/Adventure
4                        Howard, Ron           Mystery
...                              ...               ...
1190          Cowperthwaite, Gabriela         Biography
1191                  Gracey, Michael         Biography
1192                Hancock, John Lee         Biography
1193  Dayton, Jonathan|Faris, Valerie         Biography
1194                   Coppola, Sofia             Drama

                                    genres  \
0                   Action|Adventure|Fantasy
```

```
1              Adventure|Comedy|Family|Fantasy
2                Animation|Comedy|Family|Sport
3     Action|Adventure|Fantasy|Sci-Fi|Thriller
4                            Mystery|Thriller
...                                       ...
1190                     Biography|Drama|War
1191                 Biography|Drama|Musical
1192                 Biography|Drama|History
1193             Biography|Comedy|Drama|Sport
1194                               Drama|War

                                          camera_format
negative_format  \
0     Arriflex 235, Panavision Primo Lenses|Panavisi...           35
mm
1     Arriflex 435, Zeiss Ultra Prime and Angenieux ...           35
mm
2                                                    NaN
Digital
3                         Panavision Cameras and Lenses           35
mm
4     Arricam LT, Cooke S4 Lenses|Arricam ST, Cooke ...           35
mm
...                                                  ...            ..
.
1190                                      Arri Alexa
NaN
1191                                   Arri Alexa 65
Codex
1192         Arri Alexa XT, Panavision G- Series Lenses   Codex
ARRIRAW
1193  Arricam LT, Kowa Cine Prominar prime and Angen...           35
mm
1194  Arricam LT, Zeiss Master Prime and Angenieux O...           35
mm

         budget  inflation_adjusted_budget budget_source film_type
0     225000000                299587500.0   the-numbers         F
1     110000000                146465000.0   the-numbers         F
2      70000000                 93205000.0   the-numbers         D
3     210000000                279615000.0   the-numbers         F
4     125000000                166437500.0   the-numbers         F
...         ...                       ...           ...       ...
1190          0                       0.0   the-numbers         D
1191   84000000                 91988400.0   the-numbers         D
1192   10500000                 11743200.0   the-numbers         D
1193   25000000                 27377500.0   the-numbers         F
1194   10500000                 11498550.0          imdb         F

[1195 rows x 12 columns]
```

```
data.sort_values(by='production_year', ascending=True)
```

```
           id  production_year
title  \
0     4103791             2006  Pirates of the Caribbean: Dead Man's
Chest
71    3792938             2006                         John Tucker Must
Die
70    4378966             2006                          The Hills Have
Eyes
69    3862759             2006                            Lady in the
Water
68    3662671             2006                                   Glory
Road
...       ...              ...
...
1123  4359694             2017                              The Emoji
Movie
1122  3370877             2017                             Blade Runner
2049
1120  3821534             2017                 Kingsman: The Golden
Circle
1118  3328988             2017                                   Baby
Driver
1194  4332562             2017                                The
Beguiled

                directors     primary_genre  \
0         Verbinski, Gore  Action/Adventure
71          Thomas, Betty            Comedy
70         Aja, Alexandre            Horror
69    Shyamalan, M. Night             Drama
68         Gartner, James         Biography
...                   ...               ...
1123       Leondis, Tony  Action/Adventure
1122   Villeneuve, Denis           Mystery
1120     Vaughn, Matthew  Action/Adventure
1118       Wright, Edgar  Action/Adventure
1194       Coppola, Sofia             Drama

                                       genres  \
0                   Action|Adventure|Fantasy
71                            Comedy|Romance
70                                    Horror
69              Drama|Fantasy|Mystery|Thriller
68                         Biography|Drama|Sport
...                                         ...
1123  Adventure|Animation|Comedy|Family|Sci-Fi
1122                   Mystery|Sci-Fi|Thriller
1120                   Action|Adventure|Comedy
```

```
1118               Action|Crime|Music|Thriller
1194                                   Drama|War

                                            camera_format
negative_format  \
0     Arriflex 235, Panavision Primo Lenses|Panavisi...
35 mm
71                          Clairmont Cameras and Lenses
35 mm
70                           Arricam ST|Moviecam Compact
35 mm
69    Arricam LT, Cooke S4 and Angenieux Optimo Lens...
35 mm
68       Arriflex Cameras|Panavision Cameras and Lenses
35 mm
...                                                 ...
...
1123                                                NaN
Digital
1122  Arri Alexa Mini, Zeiss Master Prime Lenses|Arr...
Codex
1120  Arri Alexa XT, Hawk V-Lite and V-Plus Anamorph...
Codex
1118  Arri Alexa Mini, Panavision C-, G- and T-Serie...  35 mm|Codex
ARRIRAW
1194  Arricam LT, Zeiss Master Prime and Angenieux O...
35 mm

         budget  inflation_adjusted_budget budget_source film_type
0     225000000                299587500.0   the-numbers         F
71     18000000                 23967000.0          imdb         F
70     17000000                 22635500.0   the-numbers         F
69     75000000                 99862500.0   the-numbers         F
68     30000000                 39945000.0   the-numbers         F
...         ...                        ...           ...       ...
1123   50000000                 54755000.0   the-numbers         D
1122  185000000                202593500.0   the-numbers         D
1120  104000000                113890400.0   the-numbers         D
1118   34000000                 37233400.0   the-numbers        FD
1194   10500000                 11498550.0          imdb         F

[1195 rows x 12 columns]
```

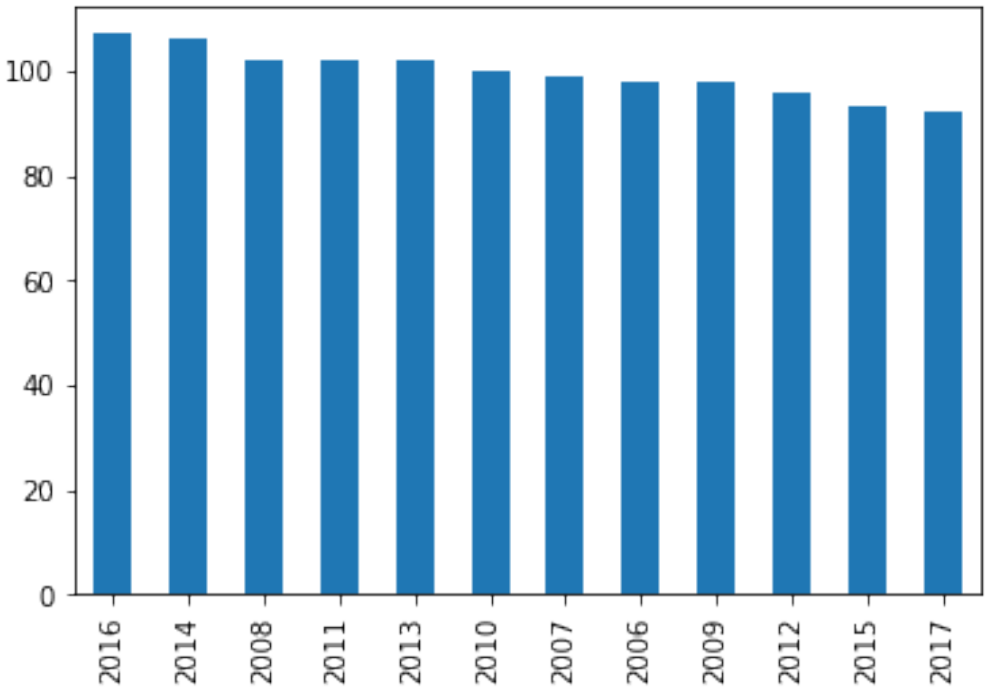
```
cols = ['production_year','film_type']
col_names = ['year', 'mediums']
df = data[cols]
df.columns = col_names
df
```

```
      year mediums
0     2006       F
1     2006       F
2     2006       D
3     2006       F
4     2006       F
...    ...     ...
1190  2017       D
1191  2017       D
1192  2016       D
1193  2017       F
1194  2017       F

[1195 rows x 2 columns]
```

```
df['year'].value_counts(dropna=False)
```

```
2016    107
2014    106
2008    102
2011    102
2013    102
2010    100
2007     99
2006     98
2009     98
2012     96
2015     93
2017     92
Name: year, dtype: int64
```

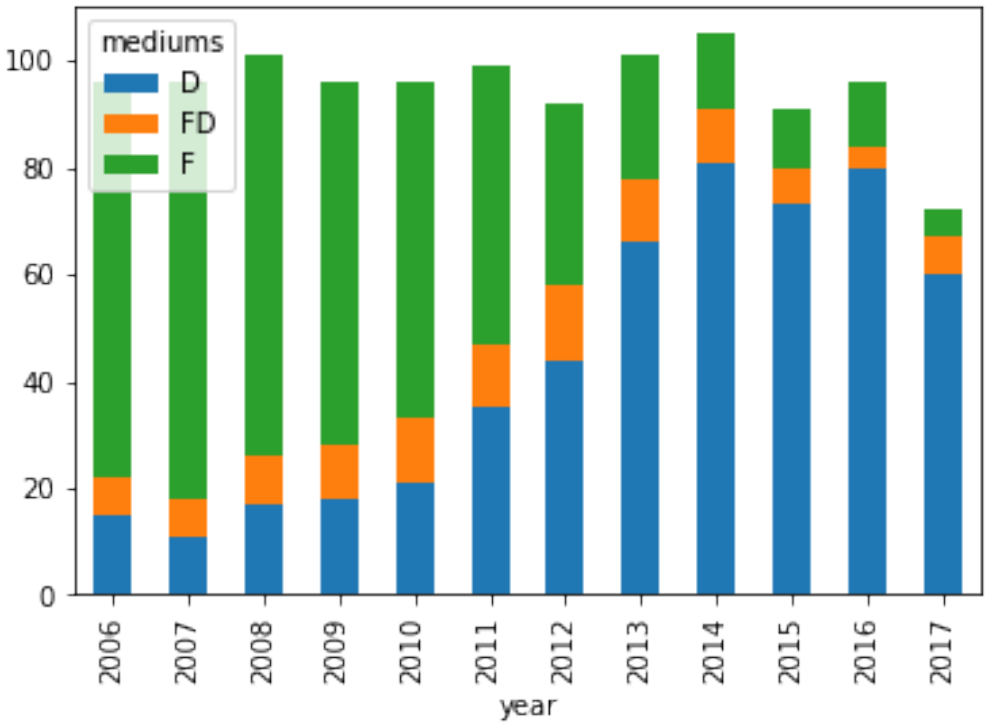
```
df['year'].value_counts(dropna=False).plot.bar()
plt.show()
```




```
medium_year = pd.crosstab(df['year'], df['mediums'])[['D','FD','F']]
medium_year
```

mediums	D	FD	F
year			
2006	15	7	74
2007	11	7	78
2008	17	9	75
2009	18	10	68
2010	21	12	63
2011	35	12	52
2012	44	14	34
2013	66	12	23
2014	81	10	14
2015	73	7	11
2016	80	4	12
2017	60	7	5

```
medium_year.plot.bar(stacked=True)
plt.show()
```

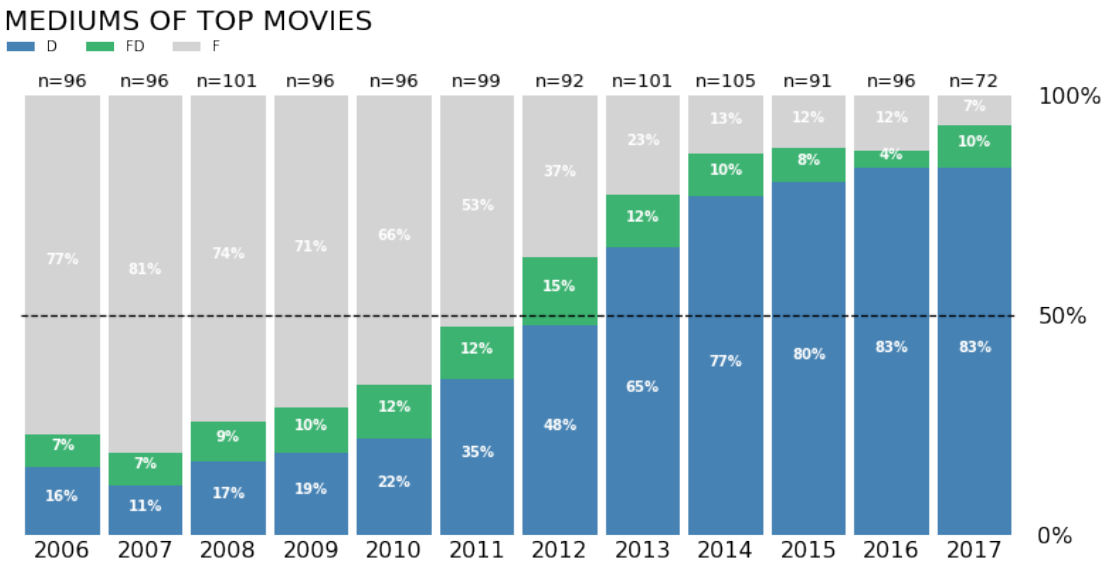



```
# Divide each number by the sum of the row
medium_year_pct = medium_year.apply(lambda r: r/r.sum(), axis=1)
medium_year_pct
```

mediums	D	FD	F
year			
2006	0.156250	0.072917	0.770833
2007	0.114583	0.072917	0.812500
2008	0.168317	0.089109	0.742574
2009	0.187500	0.104167	0.708333
2010	0.218750	0.125000	0.656250
2011	0.353535	0.121212	0.525253
2012	0.478261	0.152174	0.369565
2013	0.653465	0.118812	0.227723
2014	0.771429	0.095238	0.133333
2015	0.802198	0.076923	0.120879
2016	0.833333	0.041667	0.125000
2017	0.833333	0.097222	0.069444

```
ax = medium_year_pct.plot.bar(figsize=(13,6), width=0.9, stacked=True,
fontsize=16, rot=0, color = ['steelblue',
'mediumseagreen','lightgray'])

# switching y axis to the right
ax.yaxis.tick_right()

# making sure the tick sizes are the same
ax.tick_params(axis='both', which='major', labelsize=16)
```

```
# removing tick marks from x and y axis
ax.tick_params(axis='x', length=0)
ax.tick_params(axis='y', length=0)

# changing y-axis format to percentages
ax.yaxis.set_major_formatter(mtick.PercentFormatter(1.0))

# only showing 0,50,100%
ax.yaxis.set_ticks(np.arange(0, 1.1, 0.5))

# get rid of xlabel text
plt.xlabel('')

# dash line through middle
plt.hlines(.5, xmin=-0.5, xmax=11.5,  linewidth=1.3,
linestyles='dashed', color='black')

# title
ax.set_title('MEDIUMS OF TOP MOVIES', fontsize=20, pad=30, loc='left')

# legend
ax.legend(loc='upper left', bbox_to_anchor=(-0.01, 1.1),
frameon=False, ncol=3)

# spines
[ax.spines[i].set_visible(False) for i in ax.spines]

# add percentages for each segment
x_axis =0
for year in medium_year_pct.index:
  for medium in ['D','FD','F']:
    if medium == 'D':
      p=medium_year_pct[medium][year]/2
    elif medium == 'FD':
      p=(medium_year_pct[medium][year]/2) + (medium_year_pct['D']
[year])
    else:
      p=1-medium_year_pct[medium][year]/2
    plt.text(x_axis, p, str(round(medium_year_pct[medium][year]*100))
+ '%', color='white',weight='bold',
             horizontalalignment='center',fontsize=10)
  plt.text(x_axis,1.02, 'n=' + str(medium_year.sum(axis=1)
[year]),horizontalalignment='center',fontsize=13)
  x_axis +=1

plt.savefig('stacked_barplot.pdf')
plt.show()
```



Notes from the website regarding the data:

- This exploration only includes non-animated fiction movies in the top 100 movies at the US Box Office every year between 2006 and 2017, collected from The Numbers.
- The director shooting medium chart comprises only directors who've made movies on the above-mentioned top list, however the "path" considered all movies they made regardless the box office performance.
- Only movies with known shooting medium were counted in the calculation of the percentage breakdown.