

Московский государственный технический университет им. Н.Э. Баумана
Факультет «Информатика и системы управления»
Кафедра «Автоматизированные системы обработки информации и
управления»



Отчет
Лабораторная работа № 2
По курсу Технологии машинного обучения»

ИСПОЛНИТЕЛЬ:

Кожуро Б.Е.

Группа ИУ5-65Б

ПРЕПОДАВАТЕЛЬ:

Гапанюк Ю.Е.

"__" _____ 2021 г.

Москва 2021

Lab2

April 20, 2021

1 Lab2

dc-wikia-data

1.1

```
[1]: import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
sns.set(style="ticks")
```

```
[2]: data = pd.read_csv('dc-wikia-data.csv', sep=",")
```

```
[3]: data.shape
```

```
[3]: (6896, 13)
```

```
[4]: data.dtypes
```

```
[4]: page_id          int64
name                object
urlslug             object
ID                  object
ALIGN               object
EYE                 object
HAIR                object
SEX                 object
GSM                 object
ALIVE               object
APPEARANCES         float64
FIRST APPEARANCE    object
YEAR                float64
dtype: object
```

```
[5]: data.isnull().sum()
```

```
[5]: page_id      0
      name        0
      urlslug     0
      ID          2013
      ALIGN       601
      EYE         3628
      HAIR        2274
      SEX         125
      GSM         6832
      ALIVE       3
      APPEARANCES 355
      FIRST APPEARANCE 69
      YEAR        69
      dtype: int64
```

```
[42]: data.head()
```

```
[42]:
```

	page_id	name	urlslug
0	1422	Batman (Bruce Wayne)	\wiki\Batman_(Bruce_Wayne)
1	23387	Superman (Clark Kent)	\wiki\Superman_(Clark_Kent)
2	1458	Green Lantern (Hal Jordan)	\wiki\Green_Lantern_(Hal_Jordan)
3	1659	James Gordon (New Earth)	\wiki\James_Gordon_(New_Earth)
4	1576	Richard Grayson (New Earth)	\wiki\Richard_Grayson_(New_Earth)

	ID	ALIGN	EYE	HAIR	SEX
0	Secret Identity	Good Characters	Blue Eyes	Black Hair	Male Characters
1	Secret Identity	Good Characters	Blue Eyes	Black Hair	Male Characters
2	Secret Identity	Good Characters	Brown Eyes	Brown Hair	Male Characters
3	Public Identity	Good Characters	Brown Eyes	White Hair	Male Characters
4	Secret Identity	Good Characters	Blue Eyes	Black Hair	Male Characters

	GSM	ALIVE	APPEARANCES	FIRST APPEARANCE	YEAR
0	NaN	Living Characters	3093.0	1939, May	1939.0
1	NaN	Living Characters	2496.0	1986, October	1986.0
2	NaN	Living Characters	1565.0	1959, October	1959.0
3	NaN	Living Characters	1316.0	1987, February	1987.0
4	NaN	Living Characters	1237.0	1940, April	1940.0

1.2

```
[6]: #
      #
      total_count = data.shape[0]
      num_cols = []
      for col in data.columns:
          #
```

```

temp_null_count = data[data[col].isnull()].shape[0]
dt = str(data[col].dtype)
if temp_null_count>0 and (dt=='float64' or dt=='int64' or dt=='object'):
    num_cols.append(col)
    temp_perc = round((temp_null_count / total_count) * 100.0, 2)
    print('    {}.    {}.    {}, {}%.'.format(col, dt,
    ↪temp_null_count, temp_perc))

```

```

ID.      object.      2013, 29.19%.
ALIGN.    object.      601, 8.72%.
EYE.      object.      3628, 52.61%.
HAIR.     object.      2274, 32.98%.
SEX.      object.      125, 1.81%.
GSM.      object.      6832, 99.07%.
ALIVE.    object.      3, 0.04%.
APPEARANCES. float64.      355, 5.15%.
FIRST APPEARANCE. object.      69,
1.0%.
YEAR.     float64.      69, 1.0%.

```

1.2.1

50%, 7 %

```
[7]: data1 = data.drop(columns=['GSM', 'EYE'])
```

```
[8]: data2 = data1.dropna(subset=['ALIVE', 'YEAR', 'SEX'])
```

```
[9]: data2.isnull().sum()
```

```

[9]: page_id      0
name             0
urlslug          0
ID              1915
ALIGN           568
HAIR            2131
SEX             0
ALIVE           0
APPEARANCES     335
FIRST APPEARANCE 0
YEAR            0
dtype: int64

```

```

[10]: print(f" {data.shape[0]-data2.shape[0]} {data.shape[1]-data2.
    ↪shape[1]} ")

```

195 2

,

[illegible]

appearances

1.

```
[20]: data2.loc[:, 'APPEARANCES'] = data2.loc[:, 'APPEARANCES'].fillna(value = 1)
```

```
[21]: data2.isnull().sum()
```

```
[21]: page_id          0
      name            0
      urlslug         0
      ID              1915
      ALIGN           568
      HAIR            2131
      SEX             0
      ALIVE           0
      APPEARANCES     0
      FIRST APPEARANCE 0
      YEAR            0
      dtype: int64
```

```
[22]: from sklearn.preprocessing import LabelEncoder, OneHotEncoder
```

```
[23]: data2['ID'].unique()
```

```
[23]: array(['Secret Identity', 'Public Identity', nan, 'Identity Unknown'],
      dtype=object)
```

```
[24]: data2['ID'].value_counts()
```

```
[24]: Public Identity    2409
      Secret Identity   2368
      Identity Unknown     9
      Name: ID, dtype: int64
```

2 - Identity unknown, Identity Unknown.

```
[25]: data2.loc[:, 'ID'] = data2.loc[:, 'ID'].fillna(value = 'Identity Unknown')
```

```
[26]: data2['ALIGN'].unique()
```

```
[26]: array(['Good Characters', 'Bad Characters', 'Neutral Characters', nan,
      'Reformed Criminals'], dtype=object)
```

```
[27]: data2['ALIGN'].value_counts()
```

```
[27]: Bad Characters      2815
      Good Characters    2765
      Neutral Characters  550
      Reformed Criminals    3
      Name: ALIGN, dtype: int64
```

```
[28]: data2.loc[:, 'ALIGN'] = data2.loc[:, 'ALIGN'].fillna(value = 'Bad Characters')
```

```
[29]: data2.loc[:, 'HAIR'] = data2.loc[:, 'HAIR'].fillna(value = 'Unknown Hair')
```

```
[30]: data2.isnull().sum()
```

```
[30]: page_id      0
      name        0
      urlslug     0
      ID          0
      ALIGN       0
      HAIR        0
      SEX         0
      ALIVE       0
      APPEARANCES 0
      FIRST APPEARANCE 0
      YEAR        0
      dtype: int64
```

```
###
```

```
[32]: data2.dtypes
```

```
[32]: page_id      int64
      name        object
      urlslug     object
      ID          object
      ALIGN       object
      HAIR        object
      SEX         object
      ALIVE       object
      APPEARANCES float64
      FIRST APPEARANCE object
      YEAR        float64
      dtype: object
```

```
[34]: from sklearn.preprocessing import LabelEncoder, OneHotEncoder
      le = LabelEncoder()
      cat_enc_le = le.fit_transform(data2.loc[:, 'name'])
      print(le.inverse_transform(np.unique(cat_enc_le)))
      data2.loc[:, 'name'] = cat_enc_le
```

```
['3g4 (New Earth)' '500-ZQ (New Earth)' "A'Hwiirdh-Paan'A (New Earth)" ...
'Zyklon (New Earth)' 'Zyn (New Earth)' 'Zzlrzzzm (New Earth)']
```

```
[35]: le = LabelEncoder()
cat_enc_le = le.fit_transform(data2.loc[:, 'urlslug'])
print(le.inverse_transform(np.unique(cat_enc_le)))
data2.loc[:, 'urlslug'] = cat_enc_le
```

```
['\\wiki\\3g4_(New_Earth)' '\\wiki\\500-ZQ_(New_Earth)'
 '\\wiki\\A%27Hwiirdh-Paan%27A_(New_Earth)' ...
 '\\wiki\\Zyklon_(New_Earth)' '\\wiki\\Zyn_(New_Earth)'
 '\\wiki\\Zzlrzzzm_(New_Earth)']
```

```
[44]: cat_enc_le = le.fit_transform(data2.loc[:, 'ID'])
print(le.inverse_transform(np.unique(cat_enc_le)))
data2.loc[:, 'ID'] = cat_enc_le
```

```
[0 1 2]
```

```
[50]: cat_enc_le = le.fit_transform(data2.loc[:, 'ALIGN'])
print(le.inverse_transform(np.unique(cat_enc_le)))
data2.loc[:, 'ALIGN'] = cat_enc_le
```

```
['Bad Characters' 'Good Characters' 'Neutral Characters'
 'Reformed Criminals']
```

```
[45]: cat_enc_le = le.fit_transform(data2.loc[:, 'HAIR'])
print(le.inverse_transform(np.unique(cat_enc_le)))
data2.loc[:, 'HAIR'] = cat_enc_le
```

```
['Black Hair' 'Blond Hair' 'Blue Hair' 'Brown Hair' 'Gold Hair'
 'Green Hair' 'Grey Hair' 'Orange Hair' 'Pink Hair' 'Platinum Blond Hair'
 'Purple Hair' 'Red Hair' 'Reddish Brown Hair' 'Silver Hair'
 'Strawberry Blond Hair' 'Unknown Hair' 'Violet Hair' 'White Hair']
```

```
[46]: cat_enc_le = le.fit_transform(data2.loc[:, 'SEX'])
print(le.inverse_transform(np.unique(cat_enc_le)))
data2.loc[:, 'SEX'] = cat_enc_le
```

```
['Female Characters' 'Genderless Characters' 'Male Characters'
 'Transgender Characters']
```

```
[47]: cat_enc_le = le.fit_transform(data2.loc[:, 'ALIVE'])
print(le.inverse_transform(np.unique(cat_enc_le)))
data2.loc[:, 'ALIVE'] = cat_enc_le
```

```
['Deceased Characters' 'Living Characters']
```

```
[48]: cat_enc_le = le.fit_transform(data2.loc[:, 'FIRST APPEARANCE'])
print(le.inverse_transform(np.unique(cat_enc_le)))
data2.loc[:, 'FIRST APPEARANCE'] = cat_enc_le
```


['1935, October' '1936, February' '1936, March' '1936, September'
 '1937, March' '1938, January' '1938, July' '1938, June' '1938, October'
 '1939, April' '1939, August' '1939, December' '1939, July' '1939, June'
 '1939, May' '1939, November' '1939, October' '1939, September' '1940'
 '1940, April' '1940, December' '1940, February' '1940, January'
 '1940, July' '1940, June' '1940, March' '1940, May' '1940, November'
 '1940, October' '1940, September' '1941, April' '1941, August'
 '1941, December' '1941, February' '1941, January' '1941, July'
 '1941, June' '1941, May' '1941, November' '1941, October'
 '1941, September' '1942, April' '1942, August' '1942, December'
 '1942, February' '1942, January' '1942, June' '1942, March' '1942, May'
 '1942, November' '1942, September' '1943, April' '1943, August'
 '1943, December' '1943, February' '1943, January' '1943, July'
 '1943, March' '1943, November' '1943, September' '1944, December'
 '1944, February' '1944, January' '1944, June' '1944, March' '1944, May'
 '1944, October' '1945, April' '1945, August' '1945, December'
 '1945, March' '1945, September' '1946, April' '1946, August'
 '1946, December' '1946, January' '1946, June' '1946, May' '1946, October'
 '1947, April' '1947, August' '1947, December' '1947, February'
 '1947, June' '1947, May' '1947, November' '1947, October'
 '1947, September' '1948, April' '1948, August' '1948, December'
 '1948, February' '1948, January' '1948, June' '1948, May' '1948, October'
 '1949, December' '1949, February' '1949, November' '1949, September'
 '1950, August' '1950, December' '1950, July' '1950, June'
 '1950, September' '1951, February' '1951, June' '1951, May'
 '1951, November' '1951, October' '1952, August' '1952, February'
 '1952, June' '1952, September' '1953, January' '1954, April'
 '1954, January' '1954, November' '1954, September' '1955, February'
 '1955, January' '1955, March' '1955, November' '1955, September'
 '1956, April' '1956, December' '1956, February' '1956, May'
 '1956, October' '1956, September' '1957, April' '1957, August'
 '1957, December' '1957, February' '1957, January' '1957, July'
 '1957, June' '1957, September' '1958, April' '1958, August'
 '1958, December' '1958, March' '1958, May' '1958, November'
 '1958, September' '1959, April' '1959, August' '1959, December'
 '1959, February' '1959, January' '1959, July' '1959, June' '1959, March'
 '1959, May' '1959, November' '1959, October' '1959, September'
 '1960, August' '1960, December' '1960, February' '1960, January'
 '1960, July' '1960, June' '1960, March' '1960, May' '1960, November'
 '1960, October' '1961, April' '1961, August' '1961, December'
 '1961, January' '1961, July' '1961, June' '1961, March' '1961, May'
 '1961, November' '1961, October' '1961, September' '1962, April'
 '1962, August' '1962, December' '1962, February' '1962, July'
 '1962, June' '1962, March' '1962, May' '1962, November' '1963, August'
 '1963, December' '1963, February' '1963, January' '1963, July'
 '1963, June' '1963, May' '1963, November' '1963, October'
 '1963, September' '1964, August' '1964, December' '1964, February'
 '1964, January' '1964, July' '1964, June' '1964, March' '1964, November'

'1964, October' '1964, September' '1965, April' '1965, August'
 '1965, December' '1965, February' '1965, January' '1965, July'
 '1965, June' '1965, March' '1965, May' '1965, November' '1965, October'
 '1965, September' '1966, April' '1966, August' '1966, December'
 '1966, February' '1966, January' '1966, July' '1966, June' '1966, March'
 '1966, May' '1966, November' '1966, October' '1966, September'
 '1967, April' '1967, August' '1967, December' '1967, February'
 '1967, January' '1967, July' '1967, June' '1967, March' '1967, October'
 '1967, September' '1968, April' '1968, August' '1968, December'
 '1968, February' '1968, January' '1968, July' '1968, June' '1968, March'
 '1968, May' '1968, November' '1968, October' '1968, September'
 '1969, August' '1969, December' '1969, February' '1969, January'
 '1969, July' '1969, March' '1969, May' '1969, November' '1969, October'
 '1969, September' '1970, April' '1970, August' '1970, December'
 '1970, February' '1970, January' '1970, July' '1970, June' '1970, May'
 '1970, November' '1970, October' '1971, April' '1971, December'
 '1971, February' '1971, January' '1971, July' '1971, June' '1971, March'
 '1971, May' '1971, November' '1971, October' '1971, September'
 '1972, April' '1972, August' '1972, December' '1972, February'
 '1972, January' '1972, July' '1972, June' '1972, March' '1972, May'
 '1972, November' '1972, October' '1972, September' '1973, April'
 '1973, August' '1973, February' '1973, January' '1973, July'
 '1973, March' '1973, November' '1973, September' '1974, April'
 '1974, August' '1974, December' '1974, February' '1974, July'
 '1974, June' '1974, March' '1974, November' '1974, October' '1975, April'
 '1975, August' '1975, December' '1975, July' '1975, June' '1975, March'
 '1975, May' '1975, November' '1975, October' '1975, September'
 '1976, April' '1976, August' '1976, December' '1976, February'
 '1976, January' '1976, July' '1976, June' '1976, March' '1976, May'
 '1976, November' '1976, October' '1976, September' '1977, April'
 '1977, August' '1977, December' '1977, February' '1977, July'
 '1977, June' '1977, March' '1977, May' '1977, November' '1977, September'
 '1978, April' '1978, December' '1978, February' '1978, January'
 '1978, July' '1978, June' '1978, March' '1978, May' '1978, November'
 '1978, October' '1978, September' '1979, April' '1979, August'
 '1979, December' '1979, February' '1979, January' '1979, July'
 '1979, June' '1979, March' '1979, May' '1979, November' '1979, October'
 '1979, September' '1980, April' '1980, December' '1980, February'
 '1980, January' '1980, July' '1980, June' '1980, May' '1980, November'
 '1980, October' '1980, September' '1981, April' '1981, August'
 '1981, December' '1981, February' '1981, January' '1981, July'
 '1981, June' '1981, March' '1981, May' '1981, November' '1981, October'
 '1981, September' '1982, April' '1982, August' '1982, December'
 '1982, February' '1982, January' '1982, July' '1982, June' '1982, March'
 '1982, May' '1982, November' '1982, October' '1982, September' '1983'
 '1983, April' '1983, August' '1983, December' '1983, February'
 '1983, January' '1983, July' '1983, June' '1983, March' '1983, May'
 '1983, November' '1983, October' '1983, September' '1984' '1984, April'

'1984, August' '1984, December' '1984, February' '1984, January'
'1984, July' '1984, June' '1984, March' '1984, May' '1984, November'
'1984, October' '1984, September' '1985' '1985, April' '1985, August'
'1985, December' '1985, February' '1985, January' '1985, July'
'1985, June' '1985, March' '1985, May' '1985, November' '1985, October'
'1985, September' '1986' '1986, April' '1986, August' '1986, December'
'1986, February' '1986, January' '1986, July' '1986, June' '1986, March'
'1986, May' '1986, November' '1986, October' '1986, September' '1987'
'1987, April' '1987, August' '1987, December' '1987, February'
'1987, January' '1987, July' '1987, June' '1987, March' '1987, May'
'1987, November' '1987, October' '1987, September' '1988' '1988, April'
'1988, August' '1988, December' '1988, February' '1988, Holiday'
'1988, January' '1988, July' '1988, June' '1988, March' '1988, May'
'1988, November' '1988, October' '1988, September' '1989' '1989, April'
'1989, August' '1989, December' '1989, February' '1989, January'
'1989, July' '1989, June' '1989, March' '1989, May' '1989, November'
'1989, October' '1989, September' '1990' '1990, April' '1990, August'
'1990, December' '1990, February' '1990, January' '1990, July'
'1990, June' '1990, March' '1990, May' '1990, November' '1990, October'
'1990, September' '1991' '1991, April' '1991, August' '1991, December'
'1991, February' '1991, January' '1991, July' '1991, June' '1991, March'
'1991, May' '1991, November' '1991, October' '1991, September' '1992'
'1992, April' '1992, August' '1992, December' '1992, February'
'1992, January' '1992, July' '1992, June' '1992, March' '1992, May'
'1992, November' '1992, October' '1992, September' '1993' '1993, April'
'1993, August' '1993, December' '1993, February' '1993, January'
'1993, July' '1993, June' '1993, March' '1993, May' '1993, November'
'1993, October' '1993, September' '1994, April' '1994, August'
'1994, December' '1994, February' '1994, January' '1994, July'
'1994, June' '1994, March' '1994, May' '1994, November' '1994, October'
'1994, September' '1995' '1995, April' '1995, August' '1995, December'
'1995, February' '1995, January' '1995, July' '1995, June' '1995, March'
'1995, May' '1995, November' '1995, October' '1995, September' '1996'
'1996, April' '1996, August' '1996, December' '1996, February'
'1996, January' '1996, July' '1996, June' '1996, March' '1996, May'
'1996, November' '1996, October' '1996, September' '1997' '1997, April'
'1997, August' '1997, December' '1997, February' '1997, January'
'1997, July' '1997, June' '1997, March' '1997, May' '1997, November'
'1997, October' '1997, September' '1998' '1998, April' '1998, August'
'1998, December' '1998, February' '1998, January' '1998, July'
'1998, June' '1998, March' '1998, May' '1998, November' '1998, October'
'1998, September' '1999' '1999, April' '1999, August' '1999, December'
'1999, February' '1999, January' '1999, July' '1999, June' '1999, March'
'1999, May' '1999, November' '1999, October' '1999, September' '2000'
'2000, April' '2000, August' '2000, December' '2000, February'
'2000, January' '2000, July' '2000, June' '2000, March' '2000, May'
'2000, November' '2000, October' '2000, September' '2001, April'
'2001, August' '2001, December' '2001, February' '2001, January'

```

'2001, July' '2001, June' '2001, March' '2001, May' '2001, November'
'2001, October' '2001, September' '2002, April' '2002, August'
'2002, December' '2002, February' '2002, January' '2002, July'
'2002, June' '2002, March' '2002, May' '2002, November' '2002, October'
'2002, September' '2003, April' '2003, August' '2003, December'
'2003, February' '2003, January' '2003, July' '2003, June' '2003, March'
'2003, May' '2003, November' '2003, October' '2003, September'
'2004, April' '2004, August' '2004, December' '2004, February'
'2004, January' '2004, July' '2004, June' '2004, March' '2004, May'
'2004, November' '2004, October' '2004, September' '2005, April'
'2005, August' '2005, December' '2005, February' '2005, January'
'2005, July' '2005, June' '2005, March' '2005, May' '2005, November'
'2005, October' '2005, September' '2006, April' '2006, August'
'2006, December' '2006, February' '2006, January' '2006, July'
'2006, June' '2006, March' '2006, May' '2006, November' '2006, October'
'2006, September' '2007, April' '2007, August' '2007, December'
'2007, February' '2007, January' '2007, July' '2007, June' '2007, March'
'2007, May' '2007, November' '2007, October' '2007, September'
'2008, April' '2008, August' '2008, December' '2008, February'
'2008, January' '2008, July' '2008, June' '2008, March' '2008, May'
'2008, November' '2008, October' '2008, September' '2009, April'
'2009, August' '2009, December' '2009, February' '2009, January'
'2009, July' '2009, June' '2009, March' '2009, May' '2009, November'
'2009, October' '2009, September' '2010, April' '2010, August'
'2010, December' '2010, February' '2010, January' '2010, July'
'2010, June' '2010, March' '2010, May' '2010, November' '2010, October'
'2010, September' '2011, April' '2011, August' '2011, February'
'2011, January' '2011, July' '2011, June' '2011, March' '2011, May'
'2011, November' '2011, October' '2011, September' '2012, December'
'2012, June' '2012, March' '2012, May' '2013, October']

```

```
[51]: data2.dtypes
```

```

[51]: page_id          int64
      name            int64
      urlslug         int32
      ID              int64
      ALIGN           int32
      HAIR            int32
      SEX             int32
      ALIVE           int32
      APPEARANCES     float64
      FIRST APPEARANCE int32
      YEAR            float64
      dtype: object

```

```
[52]: data2.head()
```

```
[52]:
```

	page_id	name	urlslug	ID	ALIGN	HAIR	SEX	ALIVE	APPEARANCES	\
0	1422	583	583	2	1	0	2	1	3093.0	
1	23387	5830	5830	2	1	0	2	1	2496.0	
2	1458	2410	2410	2	1	3	2	1	1565.0	
3	1659	2908	2909	1	1	17	2	1	1316.0	
4	1576	5123	5123	2	1	0	2	1	1237.0	

	FIRST APPEARANCE	YEAR
0	14	1939.0
1	452	1986.0
2	155	1959.0
3	458	1987.0
4	19	1940.0

```
. ##
```

```
[55]: data2.describe()
```

```
[55]:
```

	page_id	name	urlslug	ID	ALIGN	\
count	6701.000000	6701.000000	6701.000000	6701.000000	6701.000000	
mean	146507.050739	3350.000000	3350.000000	1.066259	0.578123	
std	108109.027918	1934.556409	1934.556409	0.797625	0.640936	
min	1380.000000	0.000000	0.000000	0.000000	0.000000	
25%	43057.000000	1675.000000	1675.000000	0.000000	0.000000	
50%	139771.000000	3350.000000	3350.000000	1.000000	0.000000	
75%	212191.000000	5025.000000	5025.000000	2.000000	1.000000	
max	404010.000000	6700.000000	6700.000000	2.000000	3.000000	

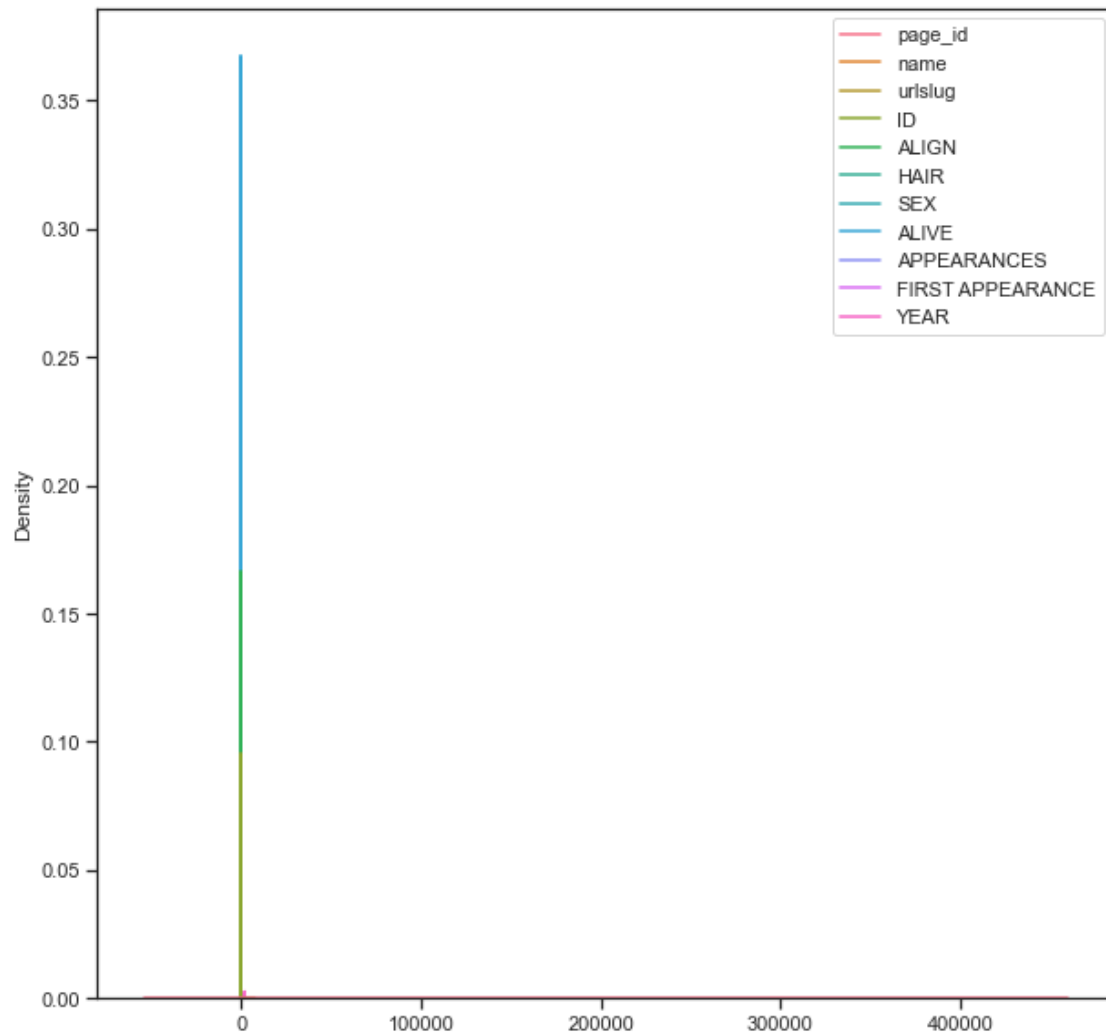
	HAIR	SEX	ALIVE	APPEARANCES	FIRST APPEARANCE	\
count	6701.000000	6701.000000	6701.000000	6701.000000	6701.000000	
mean	7.362632	1.416356	0.751977	22.884793	506.694523	
std	6.705654	0.907873	0.431897	86.403863	188.394358	
min	0.000000	0.000000	0.000000	1.000000	0.000000	
25%	1.000000	0.000000	1.000000	2.000000	413.000000	
50%	3.000000	2.000000	1.000000	6.000000	532.000000	
75%	15.000000	2.000000	1.000000	15.000000	663.000000	
max	17.000000	3.000000	1.000000	3093.000000	771.000000	

	YEAR
count	6701.000000
mean	1989.729145
std	16.853253
min	1935.000000
25%	1983.000000
50%	1992.000000
75%	2003.000000
max	2013.000000

```
[57]: def draw_kde(col_list, df1, df2, label1, label2):
    fig, (ax1, ax2) = plt.subplots(
        ncols=2, figsize=(12, 5))
    #
    ax1.set_title(label1)
    sns.kdeplot(data=df1[col_list], ax=ax1)
    #
    ax2.set_title(label2)
    sns.kdeplot(data=df2[col_list], ax=ax2)
    plt.show()
    #draw_kde(['page_id', 'name', 'urlslug', 'ID', 'ALIGN', 'HAIR', 'SEX',
    → 'ALIVE', 'APPEARANCES', 'YEAR', 'FIRST APPEARANCE'], data2, data2, 'still',
    → 'still')
```

```
[61]: plt.figure(figsize = (10, 10))
    sns.kdeplot(data=data2)
```

```
[61]: <AxesSubplot:ylabel='Density'>
```



```
[71]: from sklearn.preprocessing import StandardScaler
      from sklearn.preprocessing import RobustScaler
```

```
[75]: # StandardScaler
      cs11 = StandardScaler()
      data_cs11_scaled_temp = cs11.fit_transform(data2)
      # DataFrame
      data_cs11_scaled = pd.DataFrame(data_cs11_scaled_temp, columns=data2.columns)
      data_cs11_scaled.head()
```

```
[75]:   page_id   name  urlslug    ID  ALIGN  HAIR  SEX \
0 -1.342125 -1.430409 -1.430409  1.170739  0.65827 -1.098056  0.642918
1 -1.138936  1.282043  1.282043  1.170739  0.65827 -1.098056  0.642918
```

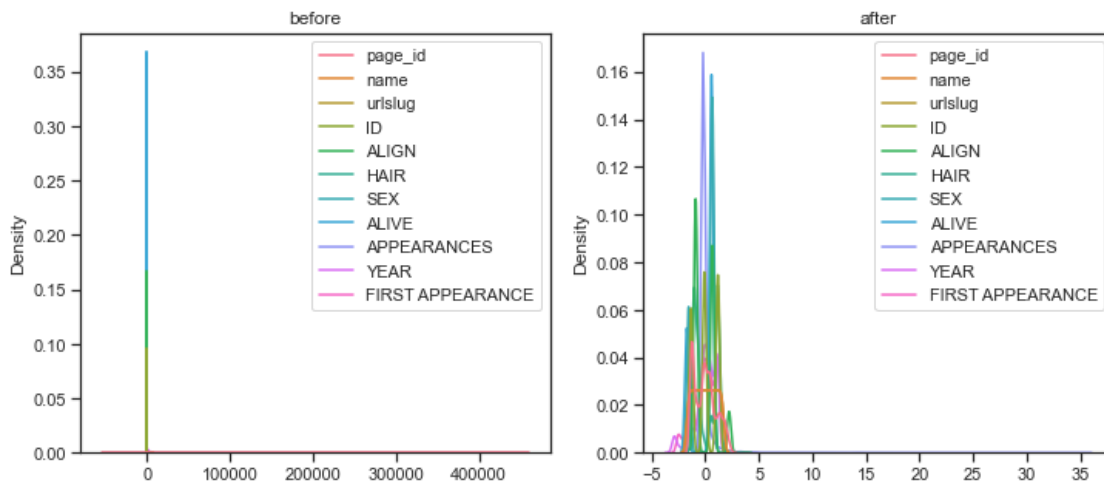
```

2 -1.341792 -0.485936 -0.485936 1.170739 0.65827 -0.650639 0.642918
3 -1.339933 -0.228493 -0.227976 -0.083076 0.65827 1.437308 0.642918
4 -1.340701 0.916558 0.916558 1.170739 0.65827 -1.098056 0.642918

```

	ALIVE	APPEARANCES	FIRST APPEARANCE	YEAR
0	0.574306	35.534804	-2.615425	-3.010275
1	0.574306	28.624875	-0.290341	-0.221288
2	0.574306	17.849089	-1.866939	-1.823472
3	0.574306	14.967059	-0.258491	-0.161948
4	0.574306	14.052680	-2.588883	-2.950935

```
[68]: draw_kde(['page_id', 'name', 'urlslug', 'ID', 'ALIGN', 'HAIR', 'SEX', 'ALIVE', 'APPEARANCES', 'YEAR', 'FIRST APPEARANCE'], data2, data_cs11_scaled, before, after)
```



```
[78]: data_cs11_scaled.corr()
```

```
[78]:
```

	page_id	name	urlslug	ID	ALIGN	HAIR	SEX	ALIVE	APPEARANCES	YEAR	FIRST APPEARANCE
page_id	1.000000	-0.005713	-0.005714	-0.164841	-0.027130	0.034827	-0.070401	0.028640	-0.236994	0.253672	0.253560
name	-0.005713	1.000000	1.000000	-0.003745	0.006856	0.006946	0.022336	0.002052	0.003869	-0.021682	-0.023463
urlslug	-0.005714	1.000000	1.000000	-0.003749	0.006852	0.006949	0.022338	0.002051	0.003871	-0.021687	-0.023468
ID	-0.164841	-0.003745	-0.003749	1.000000	-0.035527	-0.081372	0.028266	-0.034174	0.110948	0.020654	0.011161
ALIGN	-0.027130	0.006856	0.006852	-0.035527	1.000000	-0.132652	-0.106437	0.047900	0.093915	-0.059530	-0.059715
HAIR	0.034827	0.006946	0.006949	-0.081372	-0.132652	1.000000	-0.146174	-0.012848	-0.072017	0.023376	0.026616
SEX	-0.070401	0.022336	0.022338	0.028266	-0.106437	-0.146174	1.000000	0.047900	0.093915	-0.059530	-0.059715
ALIVE	0.028640	0.002052	0.002051	-0.034174	0.047900	-0.012848	0.047900	1.000000	0.093915	-0.059530	-0.059715
APPEARANCES	-0.236994	0.003869	0.003871	0.110948	0.093915	-0.072017	0.093915	0.093915	1.000000	-0.059530	-0.059715
FIRST APPEARANCE	0.253672	-0.021682	-0.021687	0.020654	-0.059530	0.023376	-0.059530	-0.059530	-0.059530	1.000000	0.026616
YEAR	0.253560	-0.023463	-0.023468	0.011161	-0.059715	0.026616	-0.059715	-0.059715	-0.059715	0.026616	1.000000

	SEX	ALIVE	APPEARANCES	FIRST APPEARANCE	YEAR
page_id	-0.070401	0.028640	-0.236994	0.253672	0.253560
name	0.022336	0.002052	0.003869	-0.021682	-0.023463
urlslug	0.022338	0.002051	0.003871	-0.021687	-0.023468
ID	0.028266	-0.034174	0.110948	0.020654	0.011161
ALIGN	-0.106437	0.047900	0.093915	-0.059530	-0.059715
HAIR	0.146174	-0.012848	-0.072017	0.023376	0.026616
SEX	1.000000	-0.070045	0.009033	-0.107070	-0.109871
ALIVE	-0.070045	1.000000	0.017544	0.048810	0.047527
APPEARANCES	0.009033	0.017544	1.000000	-0.238503	-0.247670
FIRST APPEARANCE	-0.107070	0.048810	-0.238503	1.000000	0.993906
YEAR	-0.109871	0.047527	-0.247670	0.993906	1.000000

```
[79]: data2.corr()
```

```
[79]:
```

	page_id	name	urlslug	ID	ALIGN	HAIR \
page_id	1.000000	-0.005713	-0.005714	-0.164841	-0.027130	0.034827
name	-0.005713	1.000000	1.000000	-0.003745	0.006856	0.006946
urlslug	-0.005714	1.000000	1.000000	-0.003749	0.006852	0.006949
ID	-0.164841	-0.003745	-0.003749	1.000000	-0.035527	-0.081372
ALIGN	-0.027130	0.006856	0.006852	-0.035527	1.000000	-0.132652
HAIR	0.034827	0.006946	0.006949	-0.081372	-0.132652	1.000000
SEX	-0.070401	0.022336	0.022338	0.028266	-0.106437	0.146174
ALIVE	0.028640	0.002052	0.002051	-0.034174	0.047900	-0.012848
APPEARANCES	-0.236994	0.003869	0.003871	0.110948	0.093915	-0.072017
FIRST APPEARANCE	0.253672	-0.021682	-0.021687	0.020654	-0.059530	0.023376
YEAR	0.253560	-0.023463	-0.023468	0.011161	-0.059715	0.026616

	SEX	ALIVE	APPEARANCES	FIRST APPEARANCE	YEAR
page_id	-0.070401	0.028640	-0.236994	0.253672	0.253560
name	0.022336	0.002052	0.003869	-0.021682	-0.023463
urlslug	0.022338	0.002051	0.003871	-0.021687	-0.023468
ID	0.028266	-0.034174	0.110948	0.020654	0.011161
ALIGN	-0.106437	0.047900	0.093915	-0.059530	-0.059715
HAIR	0.146174	-0.012848	-0.072017	0.023376	0.026616
SEX	1.000000	-0.070045	0.009033	-0.107070	-0.109871
ALIVE	-0.070045	1.000000	0.017544	0.048810	0.047527
APPEARANCES	0.009033	0.017544	1.000000	-0.238503	-0.247670
FIRST APPEARANCE	-0.107070	0.048810	-0.238503	1.000000	0.993906
YEAR	-0.109871	0.047527	-0.247670	0.993906	1.000000