Удодова Ксения ИУ**5-61**Б

Вариант **19,** задание **3,** датасет **3.**

Задание Для заданного набора данных произведите масштабирование данных **(**для одного признака**)** и преобразование категориальных признаков в количественные двумя способами **(label encoding, one hot encoding)** для одного признака**.** Какие методы Вы использовали для решения задачи и почему**?** Для пары произвольных колонок данных построить график **"**Диаграмма рассеяния**".**

In [1]:

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

In [3]:

```
# Анализируем данные только на обучающей выборке
data = pd.read_csv('dc-wikia-data.csv', sep = ',')
```

In [4]:

```
data.shape
```

Out[4]:

```
(6896, 13)
```

In [5]:

```
total_count = data.shape[0]
print('Всего строк: {}'.format(total_count))
```

```
Всего строк: 6896
```

In [6]:

```
data.dtypes
```

Out[6]:

```
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
```

In [7]:

```
data.head()
```

Out[7]:

page_id	name	urlslug	ID	ALIGN	EYE	HAIR	SEX	GSM	ALIVI

	page_id	name	urlslug	ID	ALIGN	EYE	HAIR	SEX	GSM	ALIVE
0	1422	Batman (Bruce Wayne)	\/wiki\/Batman_(Bruce_Wayne)	Secret Identity	Good Characters	Blue Eyes	Black Hair	Male Characters	NaN	Living Characters
1	23387	Superman (Clark Kent)	\/wiki\/Superman_(Clark_Kent)	Secret Identity	Good Characters	Blue Eyes	Black Hair	Male Characters	NaN	Living Characters
2	1458	Green Lantern (Hal Jordan)	\/wiki\/Green_Lantern_(Hal_Jordan)	Secret Identity	Good Characters	Brown Eyes	Brown Hair	Male Characters	NaN	Living Characters
3	1659	James Gordon (New Earth)	\/wiki\/James_Gordon_(New_Earth)	Public Identity	Good Characters	Brown Eyes	White Hair	Male Characters	NaN	Living Characters
4	1576	Richard Grayson (New Earth)	\/wiki\/Richard_Grayson_(New_Earth)	Secret Identity	Good Characters	Blue Eyes	Black Hair	Male Characters	NaN	Living Characters

In [8]:

```
# Выберем числовые колонки с пропущенными значениями
num_cols = []
for col in data.columns:
    # Количество пустых значений
    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'):
        num_cols.append(col)
        temp_perc = round((temp_null_count / total_count) * 100.0, 2)
        print('Колонка {}. Тип данных {}. Количество пустых значений {}, {}%.'.format(co
l, dt, temp_null_count, temp_perc))
```

```
Колонка APPEARANCES. Тип данных float64. Количество пустых значений 355, 5.15%.
Колонка YEAR. Тип данных float64. Количество пустых значений 69, 1.0%.
```

Выполним масштабирование колонки **"Appearances" -** кол-во появлений персонажей**.** Будем использовать метод **MinMax**

In [9]:

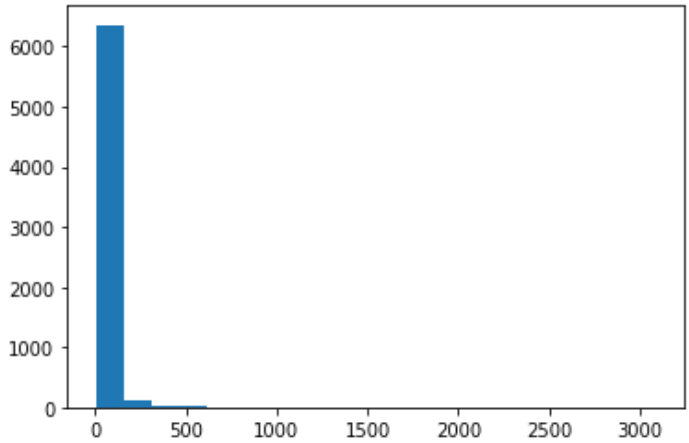
```
from sklearn.preprocessing import MinMaxScaler, StandardScaler, Normalizer
```

In [10]:

```
sc1 = MinMaxScaler()
sc1_data = sc1.fit_transform(data[['APPEARANCES']])
```

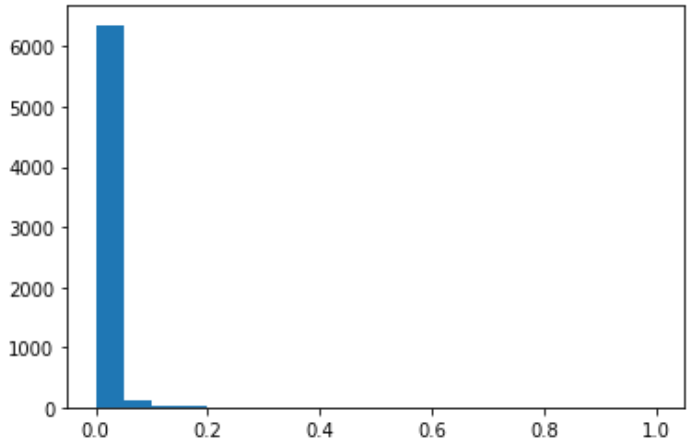
In [11]:

```
plt.hist(data['APPEARANCES'], 20)
plt.show()
```



In [12]:

```
plt.hist(sc1_data, 20)
plt.show()
```



Выполним преобразование категориальных признаков в количественные

In [13]:

```
from sklearn.preprocessing import LabelEncoder, OneHotEncoder
```

In [14]:

```
# Выберем категориальные колонки с пропущенными значениями
# Цикл по колонкам датасета
cat_cols = []
for col in data.columns:
    # Количество пустых значений
    temp_null_count = data[data[col].isnull()].shape[0]
    dt = str(data[col].dtype)
    if temp_null_count>0 and (dt=='object'):
        cat_cols.append(col)
        temp_perc = round((temp_null_count / total_count) * 100.0, 2)
        print('Колонка {}. Тип данных {}. Кол-во пустых значений {}, {}%.'.format(col, d
t, temp_null_count, temp_perc))
```

```
Колонка ID. Тип данных object. Кол-во пустых значений 2013, 29.19%.
Колонка ALIGN. Тип данных object. Кол-во пустых значений 601, 8.72%.
Колонка EYE. Тип данных object. Кол-во пустых значений 3628, 52.61%.
Колонка HAIR. Тип данных object. Кол-во пустых значений 2274, 32.98%.
Колонка SEX. Тип данных object. Кол-во пустых значений 125, 1.81%.
Колонка GSM. Тип данных object. Кол-во пустых значений 6832, 99.07%.
Колонка ALIVE. Тип данных object. Кол-во пустых значений 3, 0.04%.
Колонка FIRST APPEARANCE. Тип данных object. Кол-во пустых значений 69, 1.0%.
```

Для преобразования возьмем колонку **First APPEARANCE.** Для этой колонки лучше всего использовать **Label Encoding,** так как это колонки отражают даты, то есть в них есть порядок.

In [15]:

```
#Для начала заполним пропущенные данные
cat_temp_data = data[['FIRST APPEARANCE']]
cat_temp_data.head()
```

Out[15]:

	FIRST APPEARANCE
0	1939, May
1	1986, October
2	1959, October

	FIRST APPEARANCE
3	1987, February
4	1940, April

In [16]:

```
cat_temp_data['FIRST APPEARANCE'].unique()
```

Out[16]:

```
array(['1939, May', '1986, October', '1959, October', '1987, February',
       '1940, April', '1941, December', '1941, November', '1989, August',
       '1969, November', '1956, October', '1940, July', '1967, January',
       '1940, January', '1938, June', '1943, April', '1994, January',
       '1961, October', '1976, February', '1942, January',
       '1965, November', '1968, March', '1980, October', '1993, June',
       '1960, May', '1971, December', '1940, June', '1959, April',
       '1960, February', '1965, January', '1964, November',
       '1940, February', '1986, February', '1996, January', '1940, May',
       '1974, July', '1989, April', '1939, April', '1970, December',
       '1987, March', '1978, March', '1968, August', '1984, June',
       '1940, October', '1941, April', '1983, June', '1977, April',
       '1980, December', '1952, September', '1982, June', '1963, June',
       '1972, September', '1983, September', '1972, November',
       '1992, March', '1942, August', '1999, July', '1999, August',
       '1997, June', '1986, June', '2004, May', '1972, March',
       '1940, March', '1966, June', '1999, February', '1981, July',
       '1971, April', '1967, October', '1942, June', '1983, July',
       '1941, September', '1971, March', '1961, August', '1987, June',
       '1961, March', '1992, October', '1984, May', '1965, September',
       '1941, October', '1992, August', '1950, June', '1963, October',
       '1988, April', '1960, December', '1986, November', '1957, June',
       '1987, May', '1961, May', '1987, September', '1971, October',
       '1937, March', '1985, November', '1942, November',
       '1992, February', '1951, June', '1985, July', '1968, April',
       '1943, December', '1940, November', '1973, January', '1947, June',
       '1987, April', '1983, March', '1975, November', '1959, May',
       '1971, May', '1958, December', '1971, January', '1960, October',
       '1968, June', '1994, October', '1989, July', '1944, October',
       '1988, January', '1981, May', '1942, April', '1962, April',
       '1995, April', '1994, April', '2004, July', '2006, February',
       '1942, February', '1964, February', '1949, November',
       '1960, March', '1993, January', '1982, February',
       '1966, September', '1999, October', '1975, May', '1960, January',
       '1945, December', '1987, January', '1959, February', '1962, June',
       '1957, February', '1994, September', '1967, September',
       '1938, October', '1970, July', '1960, July', '1984, October',
       '1948, October', '1979, January', '1992', '1987, August',
       '1976, September', '1963, July', '1963, November', '1947, August',
       '1970, June', '1964, December', '1964, July', '1967, June',
       '1989, May', '1942, September', '1995, May', '1935, October',
       '1989, September', '1988, October', '1996, September',
       '1996, October', '1978, February', '1964, June', '1942, December',
       '1958, May', '2000, March', '1989, February', '1981, June',
       '1971, June', '1968, December', '1962, July', '1941, May',
       '1976, March', '1991', '1966, April', '2001, May', '1955, March',
       '1949, September', '1942, March', '1963, January',
       '1985, September', '1970, February', '1980, September',
       '1965, February', '1948, February', '1982, July', '1956, May',
       '1989, November', '1978, July', '1983, August', '1959, June',
       '1988, March', '2003, August', '1994, February', '2004, October',
       '1960, November', '1989, December', '1983, February',
       '2000, September', '1990, May', '1941, June', '2007, February',
       '1989, January', '1962, March', '2005, November', '1985, August',
       '1966, October', '1962, May', '1936, February', '1981, August',
       '1948, August', '1991, July', '1990, December', '1976, January',
       '1963, September', '1947, April', '1988, February',
       '1962, November', '1952, August', '1961, January',
       '1997, November', '1997, August', '1991, February', '1964, March',
       '1959, January', '2003, January', '1998, June', '1994, March',
       '1976, December', '1952, February', '1993', '1990, October',
       '1951, February', '1951, November', '1988, August',
       '1981, January', '1975, August', '1998, January', '1970, April',
```

'1959, March', '2000, December', '1988, June', '1977, September',
'1966, March', '1939, June', nan, '1966, November', '1965, June',
'1961, April', '1992, January', '1972, February', '1993, April',
'1991, May', '1982, September', '1956, February', '1961, June',
'1959, September', '2007, March', '1943, September',
'1995, November', '1992, June', '1961, July', '1960, August',
'2005, July', '1985, May', '1984, March', '1950, August',
'2008, December', '1987, December', '1983, May', '1982, December',
'1980, April', '1948, January', '2007, January', '1984, November',
'1982, August', '1988, September', '1982, March', '1969, October',
'1965, October', '1995, January', '1991, August', '1987, July',
'1958, April', '2007, November', '1987, October', '1987, November',
'1985, March', '1947, September', '1983, April', '1973, March',
'1959, August', '1952, June', '2008, March', '1981, December',
'1970, August', '1993, July', '1961, December', '1939, December',
'1966, May', '1958, August', '1940, September', '1981, March',
'1969, May', '2000, January', '1993, August', '1992, April',
'1976, April', '2008, January', '1993, November', '1982, May',
'1980, November', '1947, February', '2001, January',
'1996, August', '1990, August', '1971, November', '1987',
'2005, January', '1947, October', '1936, March', '2006, May',
'2006, June', '1993, February', '1968, July', '1991, June', '1986',
'1985, December', '1981, February', '1943, February',
'2006, December', '1990, July', '1971, February', '2006, August',
'2003, February', '1997, February', '1978, January',
'1970, October', '1991, January', '1991, September', '1965, July',
'1945, August', '1995, June', '1989', '2006, April',
'2001, November', '1984, April', '1976, August', '2008, June',
'1997, January', '1986, March', '1980, January', '1977, June',
'2004, June', '2002, April', '1999, March', '1985, October',
'1980, May', '1963, February', '1962, December', '1997, May',
'1992, December', '1978, April', '1977, July', '1971, July',
'1957, April', '1954, April', '2007, August', '2003, November',
'2002, September', '2002, July', '2001, December',
'1994, December', '1970, November', '1969, July', '1951, May',
'2005, October', '2002, October', '2001, August', '1998, October',
'1995, September', '1977, May', '1976, October', '1972, December',
'2006, October', '1973, September', '1938, July', '2003, July',
'2000, October', '1985, June', '1984, July', '1966, January',
'1950, December', '1943, March', '2006, November', '1985, April',
'1984, September', '1973, July', '1947, May', '1943, January',
'1941, August', '2006, September', '2004, November',
'1996, November', '1993, December', '1992, September', '1979, May',
'1978, November', '1968, May', '2009, January', '2007, June',
'1990, February', '2006, January', '1998, February', '1993, May',
'1991, October', '1986, August', '1985, January', '1957, August',
'1955, September', '2007, July', '2004, February', '1992, May',
'1990', '1988, July', '1965, April', '1996, June', '1995, July',
'1994, August', '1990, March', '1989, October', '1986, July',
'1977, August', '1972, October', '1961, September',
'1959, November', '1958, September', '1954, September',
'1943, November', '1940', '2009, August', '2002, June',
'2001, April', '1993, September', '1980, June', '2009, April',
'2003, June', '1986, May', '1972, April', '2002, May',
'1996, April', '1994, November', '1976, June', '1965, December',
'1942, May', '2002, February', '1999, December', '1998, December',
'1994, May', '1993, March', '1990, June', '1982, October',
'1973, August', '1939, August', '1938, January', '2007, May',
'1988, December', '1972, August', '2006, July', '1997, March',
'1996, February', '1983, October', '1978, September',
'1969, January', '1968, September', '2009, July', '2009, May',
'2001, October', '1998, September', '1996, December', '1988, May',
'1983, November', '1983, December', '1981, April',
'1981, November', '1978, June', '1975, June', '1969, September',
'2001, June', '2001, July', '1993, October', '1982, April',
'1968, November', '1957, December', '2010, May', '2005, May',
'2005, March', '1998, July', '1978, May', '1964, October',
'1955, November', '1950, September', '1939, July',
'1995, December', '1988, November', '1988', '1978, October',
'1969, August', '1955, January', '1941, January', '2010, August',
'2009, October', '2004, April', '1974, June', '1969, March',
'1967, April', '1944, June', '2009, June', '2007, April',

'2005, September', '2003, October', '1986, January', '1972, July',
'1972, June', '1968, October', '1967, February', '2004, September',
'2004, March', '2001, March', '2000, May', '1991, March',
'1985, February', '1984, January', '1984, August',
'1981, September', '1973, April', '1969, December',
'1947, November', '1945, April', '2006, March', '2001, September',
'2000, February', '2000, August', '1998, August', '1995, March',
'1992, November', '1991, November', '1984, December', '1980, July',
'1979, June', '1967, July', '1965, May', '1946, October',
'1939, September', '2010, July', '2010, April', '2002, March',
'1998, March', '1997, October', '1990, November',
'1990, September', '1981, October', '1979, April',
'1966, February', '1959, December', '1948, June', '2008, July',
'2008, August', '2000, June', '1998, November', '1997, July',
'1977, March', '1975, April', '1974, December', '1970, May',
'1964, August', '1962, August', '1962, February', '2010, June',
'2008, February', '2007, December', '2004, August', '2003, May',
'2000, April', '2000, November', '1999, April', '1994, July',
'1991, December', '1986, September', '1984, February',
'1977, November', '1963, May', '1954, January', '1949, February',
'2011, February', '2010, March', '2009, February', '2008, April',
'1997, September', '1977, December', '1973, November',
'1970, January', '1957, September', '1957, January', '1946, June',
'1943, July', '2011, January', '2011, June', '2009, March',
'2004, December', '1997, April', '1995, October', '1977, February',
'1974, October', '1973, February', '1966, August',
'1963, December', '1945, September', '1939, October',
'1939, November', '2011, March', '2009, December',
'2009, November', '2007, September', '2004, January',
'2003, December', '2002, January', '1986, December',
'1979, February', '1979, December', '1975, December',
'1972, January', '1958, November', '1956, April', '1944, May',
'2010, January', '2010, September', '2007, October',
'2005, August', '2001, February', '1999, June', '1998, April',
'1995, February', '1979, October', '1976, July', '1971, September',
'1941, July', '2010, February', '2008, November',
'2008, September', '2008, May', '2005, December', '2003, March',
'2003, April', '2002, November', '1999', '1999, September',
'1999, January', '1995', '1986, April', '1974, March',
'1965, August', '1964, January', '1956, September',
'1953, January', '1948, April', '1944, March', '2010, November',
'2008, October', '2002, December', '2002, August', '1995, August',
'1994, June', '1992, July', '1990, April', '1989, June', '1984',
'1982, November', '1968, January', '1968, February', '1967, March',
'1958, March', '1950, July', '2011, November', '1996, July',
'1996', '1991, April', '1989, March', '1979, September',
'1969, February', '1966, December', '1963, August',
'1946, January', '1940, December', '1936, September', '2012, June',
'2011, May', '2010, December', '2010, October', '2009, September',
'2005, June', '2005, February', '1996, March', '1975, July',
'1967, December', '1946, December', '2012, December',
'2011, April', '2011, August', '2005, April', '2000, July',
'1990, January', '1980, February', '1979, July', '1972, May',
'1965, March', '1957, July', '1954, November', '2011, July',
'2000', '1999, November', '1999, May', '1997, December',
'1996, May', '1983', '1979, March', '1975, October', '1948, May',
'1946, May', '1944, January', '1944, February', '1943, August',
'2011, September', '1998, May', '1988, Holiday', '1981',
'1979, November', '1978, December', '1976, November',
'1974, April', '1967, August', '1951, October', '1941, February',
'2012, March', '1982, January', '1975, September', '1974, August',
'1961, November', '1960, June', '1959, July', '1955, February',
'1949, December', '1944, December', '2003, September', '1998',
'1997', '1983, January', '1976, May', '1974, November',
'1966, July', '1956, December', '1948, December', '1947, December',
'1945, March', '2013, October', '2011, October', '1985', '1982',
'1979, August', '1964, September', '1946, August', '2012, May',
'1975, March', '1974, February', '1946, April'], dtype=object)

In [17]:

```
from sklearn.impute import SimpleImputer
```

```
from sklearn.impute import MissingIndicator
```

In [18]:

```
# Импьютация наиболее частыми значениями
imp2 = SimpleImputer(missing_values=np.nan, strategy='most_frequent')
data_imp2 = imp2.fit_transform(cat_temp_data)
data_imp2
```

Out[18]:

```
array([['1939, May'],
       ['1986, October'],
       ['1959, October'],
       ...,
       ['2010, December'],
       ['2010, December'],
       ['2010, December']], dtype=object)
```

In [19]:

```
cat_enc = pd.DataFrame({'c1':data_imp2.T[0]})
cat_enc
```

Out[19]:

	c1
0	**1939, May**
1	**1986, October**
2	**1959, October**
3	**1987, February**
4	**1940, April**
...	**...**
6891	**2010, December**
6892	**2010, December**
6893	**2010, December**
6894	**2010, December**
6895	**2010, December**

6896 rows × 1 columns

In [20]:

```
cat_enc['c1'].unique()
```

Out[20]:

```
array(['1939, May', '1986, October', '1959, October', '1987, February',
       '1940, April', '1941, December', '1941, November', '1989, August',
       '1969, November', '1956, October', '1940, July', '1967, January',
       '1940, January', '1938, June', '1943, April', '1994, January',
       '1961, October', '1976, February', '1942, January',
       '1965, November', '1968, March', '1980, October', '1993, June',
       '1960, May', '1971, December', '1940, June', '1959, April',
       '1960, February', '1965, January', '1964, November',
       '1940, February', '1986, February', '1996, January', '1940, May',
       '1974, July', '1989, April', '1939, April', '1970, December',
       '1987, March', '1978, March', '1968, August', '1984, June',
       '1940, October', '1941, April', '1983, June', '1977, April',
       '1980, December', '1952, September', '1982, June', '1963, June',
       '1972, September', '1983, September', '1972, November',
       '1992, March', '1942, August', '1999, July', '1999, August',
       '1997, June', '1986, June', '2004, May', '1972, March',
       '1940, March', '1966, June', '1999, February', '1981, July',
       '1971, April', '1967, October', '1942, June', '1983, July',
```

'1941, September', '1971, March', '1961, August', '1987, June',
'1961, March', '1992, October', '1984, May', '1965, September',
'1941, October', '1992, August', '1950, June', '1963, October',
'1988, April', '1960, December', '1986, November', '1957, June',
'1987, May', '1961, May', '1987, September', '1971, October',
'1937, March', '1985, November', '1942, November',
'1992, February', '1951, June', '1985, July', '1968, April',
'1943, December', '1940, November', '1973, January', '1947, June',
'1987, April', '1983, March', '1975, November', '1959, May',
'1971, May', '1958, December', '1971, January', '1960, October',
'1968, June', '1994, October', '1989, July', '1944, October',
'1988, January', '1981, May', '1942, April', '1962, April',
'1995, April', '1994, April', '2004, July', '2006, February',
'1942, February', '1964, February', '1949, November',
'1960, March', '1993, January', '1982, February',
'1966, September', '1999, October', '1975, May', '1960, January',
'1945, December', '1987, January', '1959, February', '1962, June',
'1957, February', '1994, September', '1967, September',
'1938, October', '1970, July', '1960, July', '1984, October',
'1948, October', '1979, January', '1992', '1987, August',
'1976, September', '1963, July', '1963, November', '1947, August',
'1970, June', '1964, December', '1964, July', '1967, June',
'1989, May', '1942, September', '1995, May', '1935, October',
'1989, September', '1988, October', '1996, September',
'1996, October', '1978, February', '1964, June', '1942, December',
'1958, May', '2000, March', '1989, February', '1981, June',
'1971, June', '1968, December', '1962, July', '1941, May',
'1976, March', '1991', '1966, April', '2001, May', '1955, March',
'1949, September', '1942, March', '1963, January',
'1985, September', '1970, February', '1980, September',
'1965, February', '1948, February', '1982, July', '1956, May',
'1989, November', '1978, July', '1983, August', '1959, June',
'1988, March', '2003, August', '1994, February', '2004, October',
'1960, November', '1989, December', '1983, February',
'2000, September', '1990, May', '1941, June', '2007, February',
'1989, January', '1962, March', '2005, November', '1985, August',
'1966, October', '1962, May', '1936, February', '1981, August',
'1948, August', '1991, July', '1990, December', '1976, January',
'1963, September', '1947, April', '1988, February',
'1962, November', '1952, August', '1961, January',
'1997, November', '1997, August', '1991, February', '1964, March',
'1959, January', '2003, January', '1998, June', '1994, March',
'1976, December', '1952, February', '1993', '1990, October',
'1951, February', '1951, November', '1988, August',
'1981, January', '1975, August', '1998, January', '1970, April',
'1959, March', '2000, December', '1988, June', '1977, September',
'1966, March', '1939, June', '2010, December', '1966, November',
'1965, June', '1961, April', '1992, January', '1972, February',
'1993, April', '1991, May', '1982, September', '1956, February',
'1961, June', '1959, September', '2007, March', '1943, September',
'1995, November', '1992, June', '1961, July', '1960, August',
'2005, July', '1985, May', '1984, March', '1950, August',
'2008, December', '1987, December', '1983, May', '1982, December',
'1980, April', '1948, January', '2007, January', '1984, November',
'1982, August', '1988, September', '1982, March', '1969, October',
'1965, October', '1995, January', '1991, August', '1987, July',
'1958, April', '2007, November', '1987, October', '1987, November',
'1985, March', '1947, September', '1983, April', '1973, March',
'1959, August', '1952, June', '2008, March', '1981, December',
'1970, August', '1993, July', '1961, December', '1939, December',
'1966, May', '1958, August', '1940, September', '1981, March',
'1969, May', '2000, January', '1993, August', '1992, April',
'1976, April', '2008, January', '1993, November', '1982, May',
'1980, November', '1947, February', '2001, January',
'1996, August', '1990, August', '1971, November', '1987',
'2005, January', '1947, October', '1936, March', '2006, May',
'2006, June', '1993, February', '1968, July', '1991, June', '1986',
'1985, December', '1981, February', '1943, February',
'2006, December', '1990, July', '1971, February', '2006, August',
'2003, February', '1997, February', '1978, January',
'1970, October', '1991, January', '1991, September', '1965, July',
'1945, August', '1995, June', '1989', '2006, April',

'2001, November', '1984, April', '1976, August', '2008, June',
'1997, January', '1986, March', '1980, January', '1977, June',
'2004, June', '2002, April', '1999, March', '1985, October',
'1980, May', '1963, February', '1962, December', '1997, May',
'1992, December', '1978, April', '1977, July', '1971, July',
'1957, April', '1954, April', '2007, August', '2003, November',
'2002, September', '2002, July', '2001, December',
'1994, December', '1970, November', '1969, July', '1951, May',
'2005, October', '2002, October', '2001, August', '1998, October',
'1995, September', '1977, May', '1976, October', '1972, December',
'2006, October', '1973, September', '1938, July', '2003, July',
'2000, October', '1985, June', '1984, July', '1966, January',
'1950, December', '1943, March', '2006, November', '1985, April',
'1984, September', '1973, July', '1947, May', '1943, January',
'1941, August', '2006, September', '2004, November',
'1996, November', '1993, December', '1992, September', '1979, May',
'1978, November', '1968, May', '2009, January', '2007, June',
'1990, February', '2006, January', '1998, February', '1993, May',
'1991, October', '1986, August', '1985, January', '1957, August',
'1955, September', '2007, July', '2004, February', '1992, May',
'1990', '1988, July', '1965, April', '1996, June', '1995, July',
'1994, August', '1990, March', '1989, October', '1986, July',
'1977, August', '1972, October', '1961, September',
'1959, November', '1958, September', '1954, September',
'1943, November', '1940', '2009, August', '2002, June',
'2001, April', '1993, September', '1980, June', '2009, April',
'2003, June', '1986, May', '1972, April', '2002, May',
'1996, April', '1994, November', '1976, June', '1965, December',
'1942, May', '2002, February', '1999, December', '1998, December',
'1994, May', '1993, March', '1990, June', '1982, October',
'1973, August', '1939, August', '1938, January', '2007, May',
'1988, December', '1972, August', '2006, July', '1997, March',
'1996, February', '1983, October', '1978, September',
'1969, January', '1968, September', '2009, July', '2009, May',
'2001, October', '1998, September', '1996, December', '1988, May',
'1983, November', '1983, December', '1981, April',
'1981, November', '1978, June', '1975, June', '1969, September',
'2001, June', '2001, July', '1993, October', '1982, April',
'1968, November', '1957, December', '2010, May', '2005, May',
'2005, March', '1998, July', '1978, May', '1964, October',
'1955, November', '1950, September', '1939, July',
'1995, December', '1988, November', '1988', '1978, October',
'1969, August', '1955, January', '1941, January', '2010, August',
'2009, October', '2004, April', '1974, June', '1969, March',
'1967, April', '1944, June', '2009, June', '2007, April',
'2005, September', '2003, October', '1986, January', '1972, July',
'1972, June', '1968, October', '1967, February', '2004, September',
'2004, March', '2001, March', '2000, May', '1991, March',
'1985, February', '1984, January', '1984, August',
'1981, September', '1973, April', '1969, December',
'1947, November', '1945, April', '2006, March', '2001, September',
'2000, February', '2000, August', '1998, August', '1995, March',
'1992, November', '1991, November', '1984, December', '1980, July',
'1979, June', '1967, July', '1965, May', '1946, October',
'1939, September', '2010, July', '2010, April', '2002, March',
'1998, March', '1997, October', '1990, November',
'1990, September', '1981, October', '1979, April',
'1966, February', '1959, December', '1948, June', '2008, July',
'2008, August', '2000, June', '1998, November', '1997, July',
'1977, March', '1975, April', '1974, December', '1970, May',
'1964, August', '1962, August', '1962, February', '2010, June',
'2008, February', '2007, December', '2004, August', '2003, May',
'2000, April', '2000, November', '1999, April', '1994, July',
'1991, December', '1986, September', '1984, February',
'1977, November', '1963, May', '1954, January', '1949, February',
'2011, February', '2010, March', '2009, February', '2008, April',
'1997, September', '1977, December', '1973, November',
'1970, January', '1957, September', '1957, January', '1946, June',
'1943, July', '2011, January', '2011, June', '2009, March',
'2004, December', '1997, April', '1995, October', '1977, February',
'1974, October', '1973, February', '1966, August',
'1963, December', '1945, September', '1939, October',

```
       '1939, November', '2011, March', '2009, December',
       '2009, November', '2007, September', '2004, January',
       '2003, December', '2002, January', '1986, December',
       '1979, February', '1979, December', '1975, December',
       '1972, January', '1958, November', '1956, April', '1944, May',
       '2010, January', '2010, September', '2007, October',
       '2005, August', '2001, February', '1999, June', '1998, April',
       '1995, February', '1979, October', '1976, July', '1971, September',
       '1941, July', '2010, February', '2008, November',
       '2008, September', '2008, May', '2005, December', '2003, March',
       '2003, April', '2002, November', '1999', '1999, September',
       '1999, January', '1995', '1986, April', '1974, March',
       '1965, August', '1964, January', '1956, September',
       '1953, January', '1948, April', '1944, March', '2010, November',
       '2008, October', '2002, December', '2002, August', '1995, August',
       '1994, June', '1992, July', '1990, April', '1989, June', '1984',
       '1982, November', '1968, January', '1968, February', '1967, March',
       '1958, March', '1950, July', '2011, November', '1996, July',
       '1996', '1991, April', '1989, March', '1979, September',
       '1969, February', '1966, December', '1963, August',
       '1946, January', '1940, December', '1936, September', '2012, June',
       '2011, May', '2010, October', '2009, September', '2005, June',
       '2005, February', '1996, March', '1975, July', '1967, December',
       '1946, December', '2012, December', '2011, April', '2011, August',
       '2005, April', '2000, July', '1990, January', '1980, February',
       '1979, July', '1972, May', '1965, March', '1957, July',
       '1954, November', '2011, July', '2000', '1999, November',
       '1999, May', '1997, December', '1996, May', '1983', '1979, March',
       '1975, October', '1948, May', '1946, May', '1944, January',
       '1944, February', '1943, August', '2011, September', '1998, May',
       '1988, Holiday', '1981', '1979, November', '1978, December',
       '1976, November', '1974, April', '1967, August', '1951, October',
       '1941, February', '2012, March', '1982, January',
       '1975, September', '1974, August', '1961, November', '1960, June',
       '1959, July', '1955, February', '1949, December', '1944, December',
       '2003, September', '1998', '1997', '1983, January', '1976, May',
       '1974, November', '1966, July', '1956, December', '1948, December',
       '1947, December', '1945, March', '2013, October', '2011, October',
       '1985', '1982', '1979, August', '1964, September', '1946, August',
       '2012, May', '1975, March', '1974, February', '1946, April'],
      dtype=object)
```

In [21]:

```
le = LabelEncoder()
cat_enc_le = le.fit_transform(cat_enc['c1'])
```

In [22]:

```
le.classes_
```

Out[22]:

```
array(['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', '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', '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'], dtype=object)
```

In [23]:

```
np.unique(cat_enc_le)
```

Out[23]:

```
array([  0,   1,   2,   3,   4,   5,   6,   7,   8,   9,  10,  11,  12,
        13,  14,  15,  16,  17,  18,  19,  20,  21,  22,  23,  24,  25,
        26,  27,  28,  29,  30,  31,  32,  33,  34,  35,  36,  37,  38,
        39,  40,  41,  42,  43,  44,  45,  46,  47,  48,  49,  50,  51,
        52,  53,  54,  55,  56,  57,  58,  59,  60,  61,  62,  63,  64,
        65,  66,  67,  68,  69,  70,  71,  72,  73,  74,  75,  76,  77,
        78,  79,  80,  81,  82,  83,  84,  85,  86,  87,  88,  89,  90,
        91,  92,  93,  94,  95,  96,  97,  98,  99, 100, 101, 102, 103,
       104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116,
       117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129,
       130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142,
       143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155,
       156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168,
       169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181,
       182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194,
       195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207,
       208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220,
       221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233,
       234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246,
       247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259,
       260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272,
       273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285,
       286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298,
       299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311,
       312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324,
       325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337,
       338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350,
       351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363,
```

```
364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376,
377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389,
390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402,
403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415,
416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428,
429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441,
442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454,
455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467,
468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480,
481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493,
494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506,
507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519,
520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532,
533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545,
546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558,
559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571,
572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584,
585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597,
598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610,
611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623,
624, 625, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636,
637, 638, 639, 640, 641, 642, 643, 644, 645, 646, 647, 648, 649,
650, 651, 652, 653, 654, 655, 656, 657, 658, 659, 660, 661, 662,
663, 664, 665, 666, 667, 668, 669, 670, 671, 672, 673, 674, 675,
676, 677, 678, 679, 680, 681, 682, 683, 684, 685, 686, 687, 688,
689, 690, 691, 692, 693, 694, 695, 696, 697, 698, 699, 700, 701,
702, 703, 704, 705, 706, 707, 708, 709, 710, 711, 712, 713, 714,
715, 716, 717, 718, 719, 720, 721, 722, 723, 724, 725, 726, 727,
728, 729, 730, 731, 732, 733, 734, 735, 736, 737, 738, 739, 740,
741, 742, 743, 744, 745, 746, 747, 748, 749, 750, 751, 752, 753,
754, 755, 756, 757, 758, 759, 760, 761, 762, 763, 764, 765, 766,
767, 768, 769, 770, 771, 772, 773])
```

Для этой колонки не подойдет метод **One-Hot encoding,** т.к. в данных изначально есть порядок и присутствует много уникальных значений**.** Для метода **One-Hot encoding** подойдет колонка **ALIGN.**

In [24]:

```
#Для начала заполним пропущенные данные
cat_temp_data2 = data[['ALIGN']]
cat_temp_data2.head()
```

Out[24]:

	ALIGN
0	Good Characters
1	Good Characters
2	Good Characters
3	Good Characters
4	Good Characters

In [25]:

```
cat_temp_data2['ALIGN'].unique()
```

Out[25]:

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

In [26]:

```
# Импьютация наиболее частыми значениями
imp4 = SimpleImputer(missing_values=np.nan, strategy='most_frequent')
data_imp4 = imp4.fit_transform(cat_temp_data2)
data_imp4
```

Out[26]:

```
array([['Good Characters'],
       ['Good Characters'],
       ['Good Characters'],
       ...,
       ['Good Characters'],
       ['Good Characters'],
       ['Bad Characters']], dtype=object)
```

In [27]:

```
cat_enc4 = pd.DataFrame({'c2':data_imp4.T[0]})
cat_enc4
```

Out[27]:

	c2
0	Good Characters
1	Good Characters
2	Good Characters
3	Good Characters
4	Good Characters
...	...
6891	Good Characters
6892	Good Characters
6893	Good Characters
6894	Good Characters
6895	Bad Characters

6896 rows × 1 columns

In [28]:

```
pd.get_dummies(cat_enc4).head()
```

Out[28]:

	c2_Bad Characters	c2_Good Characters	c2_Neutral Characters	c2_Reformed Criminals
0	0	1	0	0
1	0	1	0	0
2	0	1	0	0
3	0	1	0	0
4	0	1	0	0

In [29]:

```
ohe = OneHotEncoder()
cat_enc_ohe = ohe.fit_transform(cat_enc4[['c2']])
```

In [30]:

```
cat_enc_ohe
```

Out[30]:

```
<6896x4 sparse matrix of type '<class 'numpy.float64'>'
 with 6896 stored elements in Compressed Sparse Row format>
```

In [31]:

```
cat_enc_ohe.todense()[0:10]
```

Out[31]:

```
matrix([[0., 1., 0., 0.],
        [0., 1., 0., 0.],
        [0., 1., 0., 0.],
        [0., 1., 0., 0.],
        [0., 1., 0., 0.],
        [0., 1., 0., 0.],
        [0., 1., 0., 0.],
        [0., 1., 0., 0.],
        [0., 1., 0., 0.],
        [0., 1., 0., 0.]])
```

Построение диаграммы рассеяния

In [32]:

```
import seaborn as sns
```

In [33]:

```
fig, ax = plt.subplots(figsize=(10,10))
fig.suptitle("Диаграмма рассеяния для колонок YEAR и HAIR")
sns.scatterplot(ax=ax, x='YEAR', y='HAIR', data=data)
```

Out[33]:

```
<AxesSubplot:xlabel='YEAR', ylabel='HAIR'>
```