*Государственное образовательное учреждение высшего профессионального образования*

|  |  |
| --- | --- |
| **Gerb-BMSTU_01** | ***«Московский государственный технический университет  имени Н.Э. Баумана»***  ***(МГТУ им. Н.Э. Баумана)*** |

**«Рубежный контроль №1»**

«Технологии машинного обучения»

**ИСПОЛНИТЕЛЬ:**

Студен группы РТ5-61

Курьянов А.И.

**ПРЕПОДАВАТЕЛЬ:**

Гапанюк Ю.Е.

"\_\_"\_\_\_\_\_\_\_\_\_\_\_2020 г.

**Вариант: №10 (2 задача 2 датасет)**

**Задача №2.**

**Датасет: https://www.kaggle.com/fivethirtyeight/fivethirtyeight-comic-characters-dataset**

Для заданного набора данных проведите обработку пропусков в данных для одного категориального и одного количественного признака.

Дополнительные требования по группам:

Для студентов группы РТ5-61Б - для пары произвольных колонок данных построить график "Jointplot".

**Код и пояснения:**

**Подключение датасета и** **основные характеристики датасета:**

import numpy as np

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

get\_ipython().run\_line\_magic('matplotlib', 'inline')

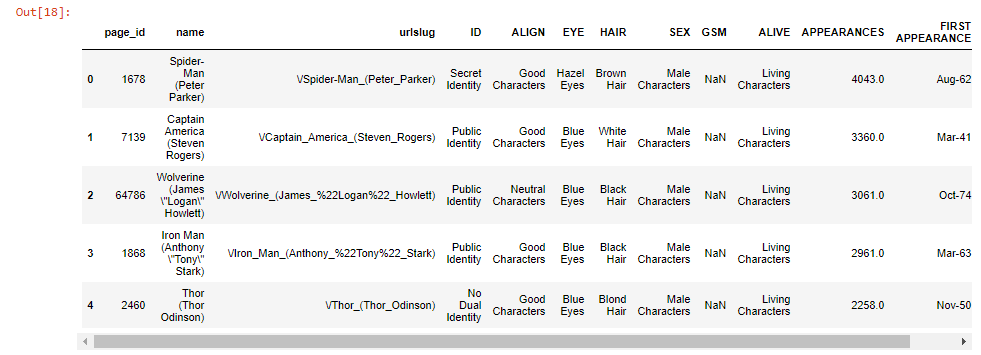
sns.set(style="ticks")

In[4]:

data = pd.read\_csv("data/marvel-wikia-data.csv")

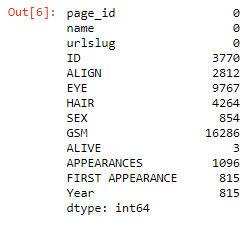
In[5]:

data.head()



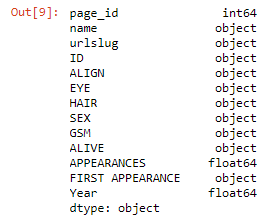
In[6]:

data.isnull().sum()



ln[7]:

data.dtypes



1. **Замена пустых значений на среднее в характеристике APPEARANCES (Количественный признак)**

In[8]:

data['APPEARANCES'].isna().sum()



In[9]:

mean = data['APPEARANCES'].mean()

mean



**Замена и проверка:**

In[10]:

data['APPEARANCES'].fillna(mean, inplace=True)

data['APPEARANCES'].isna().sum()



1. **Удаление пустых значений в характеристике ALIGN (Категориальный признак)**

In[11]:

data[data['ALIGN'].isna()] ['ALIGN'] = ' '

data['ALIGN'].isna().sum()

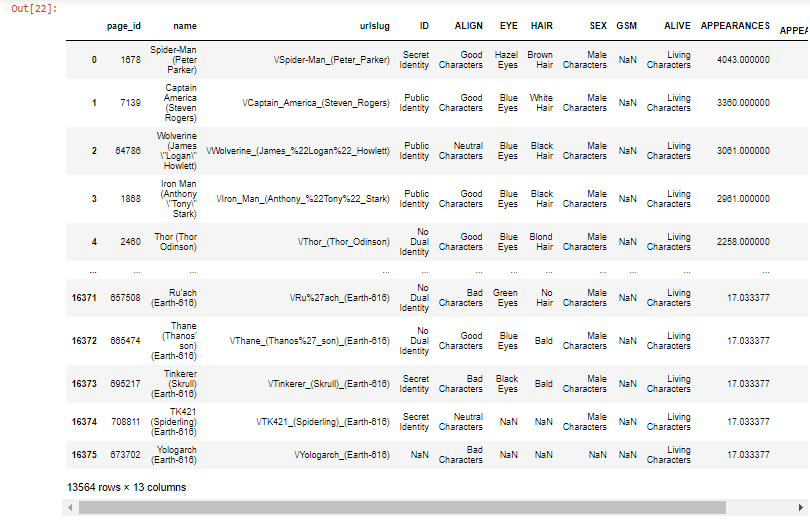


**Удаление и проверка:**

In[12]:

data=data[-data['ALIGN'].isna()]

data



In[13]:

data['ALIGN'].isna().sum()



1. **График "Jointplot" для "APPEARANCES" и "Year"**

In[14]:

jp = sns.jointplot(x="APPEARANCES", y="Year", data=data)

