In [52]:
```
import numpy as np #библиотека для работы с многомерными массивами данных и математическими операциями над ними
import pandas as pd #библиотека для анализа и обработки данных
from sklearn.datasets import load_iris #берём датасет
import matplotlib.pyplot as plt #простое рисование графиков
import seaborn as sns #удобные дефолтные настройки графиков из matpotlib
import lightgbm # сожрет все сырым и построит регрессионную модель, которая покажет важные фичи
                # чтобы дальше делать лабу только на них
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.neighbors import KNeighborsRegressor

%matplotlib inline
#для сохранения в ноутбуке вывода моих графиков
"""Для заданного набора данных проведите обработку пропусков в данных для одного категориального и одного количественного
признака. Какие способы обработки пропусков в данных для категориальных и количественных признаков Вы использовали?
Какие признаки Вы будете использовать для дальнейшего построения моделей машинного обучения и почему?"""
```

Out[52]:
```
'Для заданного набора данных проведите обработку пропусков в данных для одного категориального
и одного количественного\nпризнака. Какие способы обработки пропусков в данных для категориаль
ных и количественных признаков Вы использовали?\nКакие признаки Вы будете использовать для дал
ьнейшего построения моделей машинного обучения и почему?'
```

In [53]:
```
fifa = pd.read_csv('./data/fifa19.csv')
```

In [54]:
```
fifa.head()
```

Out[54]:

	Unnamed: 0	ID	Name	Age	Photo	Nationali
0	0	158023	L. Messi	31	https://cdn.sofifa.org/players/4/19/158023.png	Argentina
1	1	20801	Cristiano Ronaldo	33	https://cdn.sofifa.org/players/4/19/20801.png	Portugal
2	2	190871	Neymar Jr	26	https://cdn.sofifa.org/players/4/19/190871.png	Brazil
3	3	193080	De Gea	27	https://cdn.sofifa.org/players/4/19/193080.png	Spain
4	4	192985	K. De Bruyne	27	https://cdn.sofifa.org/players/4/19/192985.png	Belgium

5 rows × 89 columns

In [55]:
```
fifa.columns
```

Out[55]:
```
Index(['Unnamed: 0', 'ID', 'Name', 'Age', 'Photo', 'Nationality', 'Flag',
       'Overall', 'Potential', 'Club', 'Club Logo', 'Value', 'Wage', 'Special',
       'Preferred Foot', 'International Reputation', 'Weak Foot',
       'Skill Moves', 'Work Rate', 'Body Type', 'Real Face', 'Position',
       'Jersey Number', 'Joined', 'Loaned From', 'Contract Valid Until',
       'Height', 'Weight', 'LS', 'ST', 'RS', 'LW', 'LF', 'CF', 'RF', 'RW',
       'LAM', 'CAM', 'RAM', 'LM', 'LCM', 'CM', 'RCM', 'RM', 'LWB', 'LDM',
       'CDM', 'RDM', 'RWB', 'LB', 'LCB', 'CB', 'RCB', 'RB', 'Crossing',
       'Finishing', 'HeadingAccuracy', 'ShortPassing', 'Volleys', 'Dribbling',
       'Curve', 'FKAccuracy', 'LongPassing', 'BallControl', 'Acceleration',
       'SprintSpeed', 'Agility', 'Reactions', 'Balance', 'ShotPower',
       'Jumping', 'Stamina', 'Strength', 'LongShots', 'Aggression',
       'Interceptions', 'Positioning', 'Vision', 'Penalties', 'Composure',
       'Marking', 'StandingTackle', 'SlidingTackle', 'GKDiving', 'GKHandling',
       'GKKicking', 'GKPositioning', 'GKReflexes', 'Release Clause'],
      dtype='object')
```

In [56]:
```
# Все колонки, которые не являются числами, делаем категориальными:
for column in fifa.select_dtypes(include = ['object']).columns.tolist():
    fifa[column] = fifa[column].astype('category')
```

In [57]:
```
#fifa.dtypes
```

In [58]:
```
#fifa.isna
```

In [59]:
```
# заполнение пропусков
for column in fifa.select_dtypes(include = ['int64', 'float64']).columns.tolist():
    fifa[column] = fifa[column].fillna(fifa[column].mean())
for column in fifa.select_dtypes(include = ['category']).columns.tolist():
    fifa[column] = fifa[column].fillna(fifa[column].describe(include= ['category'])['top'])
fifa.head()
```

Out[59]:

	Unnamed: 0	ID	Name	Age	Photo	Nationali
0	0	158023	L. Messi	31	https://cdn.sofifa.org/players/4/19/158023.png	Argentina
1	1	20801	Cristiano Ronaldo	33	https://cdn.sofifa.org/players/4/19/20801.png	Portugal
2	2	190871	Neymar Jr	26	https://cdn.sofifa.org/players/4/19/190871.png	Brazil
3	3	193080	De Gea	27	https://cdn.sofifa.org/players/4/19/193080.png	Spain
4	4	192985	K. De Bruyne	27	https://cdn.sofifa.org/players/4/19/192985.png	Belgium

5 rows × 89 columns

In [60]:
```
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
for column in fifa.select_dtypes(include = ['category']).columns.tolist():
    le.fit(fifa[column])
    fifa[column] = le.transform(fifa[column])
#кодирование категориальных признаков
fifa.head()
```

Out[60]:

	Unnamed: 0	ID	Name	Age	Photo	Nationality	Flag	Overall	Potential	Club	...	Com
0	0	158023	9632	31	566	6	122	94	94	212	...	96.0
1	1	20801	3153	33	6031	123	107	94	94	326	...	95.0
2	2	190871	12508	26	3131	20	124	92	93	435	...	94.0
3	3	193080	4136	27	3467	139	114	91	93	375	...	68.0
4	4	192985	8617	27	3452	13	137	91	92	374	...	88.0

5 rows × 89 columns

In [61]:
```
lgbm_regressor = lightgbm.LGBMRegressor().fit(fifa.loc[:, fifa.columns != 'Release Clause'], fifa[
'Release Clause'])
lgbm_regressor # построили сырую и простую модель, вставив на X все кроме целевой, а на y - "Ratin
g"
```

Out[61]:
```
LGBMRegressor(boosting_type='gbdt', class_weight=None, colsample_bytree=1.0,
       importance_type='split', learning_rate=0.1, max_depth=-1,
       min_child_samples=20, min_child_weight=0.001, min_split_gain=0.0,
       n_estimators=100, n_jobs=-1, num_leaves=31, objective=None,
       random_state=None, reg_alpha=0.0, reg_lambda=0.0, silent=True,
       subsample=1.0, subsample_for_bin=200000, subsample_freq=0)
```

In [62]:
```
list_of_importances = list(zip(fifa.loc[:, fifa.columns != 'Release Clause'].columns.tolist(),
                               lgbm_regressor.feature_importances_))
list_of_importances = sorted(list_of_importances, key= lambda x: x[1], reverse= True) # список фи
ч, отсортированных по важности
```

In [63]:
```
important_features = [x[0] for x in list_of_importances if x[1] > 50]
#important_features # оставим только важные фичи
```

In [64]:
```
important_features.extend(['Release Clause'])
fifa = fifa[important_features]
```

In [65]:
```
fifa.head()
```

Out[65]:

	Value	Flag	Potential	Nationality	Unnamed: 0	Club Logo	Wage	Contract Valid Until	Age	Club	Jersey Number
0	16	122	94	6	0	490	94	3	31	212	10.0
1	195	107	94	123	1	552	74	4	33	326	7.0
2	18	124	93	20	2	637	55	4	26	435	10.0
3	190	114	93	139	3	89	49	2	27	375	1.0
4	12	137	92	13	4	29	66	5	27	374	7.0