

# Рубежный контроль №1. Тема: Технологии разведочного анализа и обработки данных.

Выполнил:

студент группы ПТ5-61Б Незамеддинов Тимур

## Импорт библиотек

```
In [1]: import pandas as pd
import numpy as np
import matplotlib as plt
from sklearn import LinearModel
```

## Загрузка данных

```
In [2]: data = pd.read_csv('data.csv')

In [3]: data.head()
```

Unnamed: 0	ID	Name	Age	Photo	Nationality	Flag	Overall	Head	Club	...	Composure	Marking	StandingTackle	SlidingTackle	GKDivng	GKHandling	GKCKicking	GKPositioning	GKReflexes	Release Clause	
0	0	158023	L. Messi	31	<a href="https://cdn.sofifa.org/players/4/19/158023.png">https://cdn.sofifa.org/players/4/19/158023.png</a>	Argentina	<a href="https://cdn.sofifa.org/flags/52.png">https://cdn.sofifa.org/flags/52.png</a>	94	94	FC Barcelona	--	96.0	33.0	28.0	26.0	6.0	11.0	15.0	14.0	8.0	€226.5M
1	1	200801	Cristiano Ronaldo	33	<a href="https://cdn.sofifa.org/players/4/19/200801.png">https://cdn.sofifa.org/players/4/19/200801.png</a>	Portugal	<a href="https://cdn.sofifa.org/flags/38.png">https://cdn.sofifa.org/flags/38.png</a>	94	94	Juventus	--	95.0	28.0	31.0	23.0	7.0	11.0	15.0	14.0	11.0	€127.1M
2	2	190871	Neymar Jr	26	<a href="https://cdn.sofifa.org/players/4/19/190871.png">https://cdn.sofifa.org/players/4/19/190871.png</a>	Brazil	<a href="https://cdn.sofifa.org/flags/54.png">https://cdn.sofifa.org/flags/54.png</a>	92	93	Paris Saint-Germain	--	94.0	27.0	24.0	33.0	9.0	9.0	15.0	15.0	11.0	€228.1M
3	3	193080	De Gea	27	<a href="https://cdn.sofifa.org/players/4/19/193080.png">https://cdn.sofifa.org/players/4/19/193080.png</a>	Spain	<a href="https://cdn.sofifa.org/flags/45.png">https://cdn.sofifa.org/flags/45.png</a>	91	93	Manchester United	--	68.0	15.0	21.0	13.0	90.0	85.0	87.0	88.0	94.0	€138.6M
4	4	192985	K. De Bruyne	27	<a href="https://cdn.sofifa.org/players/4/19/192985.png">https://cdn.sofifa.org/players/4/19/192985.png</a>	Belgium	<a href="https://cdn.sofifa.org/flags/7.png">https://cdn.sofifa.org/flags/7.png</a>	91	92	Manchester City	--	88.0	68.0	58.0	51.0	15.0	13.0	5.0	10.0	13.0	€196.4M
5 rows × 89 columns																					

```
In [4]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 18207 entries, 0 to 18206
Data columns (total 89 columns):
  Unnamed: 0      89 non-null int64
  ID              89 non-null int64
  Name           89 non-null object
  Age            89 non-null int64
  Photo          89 non-null object
  Nationality    89 non-null object
  Flag           89 non-null object
  Overall        89 non-null int64
  Potential      89 non-null int64
  Club           89 non-null object
  Club Logo      89 non-null object
  Value         89 non-null object
  Wage          89 non-null object
  Special        89 non-null int64
  Preferred Foot 89 non-null object
  International 89 non-null float64
  Weak Foot     89 non-null float64
  Skill Moves   89 non-null object
  Work Rate     89 non-null object
  Body Type     89 non-null object
  Real Face     89 non-null object
  Position      89 non-null object
  Jersey Number 89 non-null float64
  Joined        89 non-null object
  Loaned From    89 non-null object
  Contract Valid Until 89 non-null object
  Height        89 non-null object
  Weight        89 non-null object
  LS            89 non-null object
  ST            89 non-null object
  RS            89 non-null object
  LW            89 non-null object
  LF            89 non-null object
  CF            89 non-null object
  RF            89 non-null object
  RW            89 non-null object
  LAM           89 non-null object
  CAM           89 non-null object
  RAM           89 non-null object
  LW           89 non-null object
  LF           89 non-null object
  CF           89 non-null object
  RF           89 non-null object
  RW           89 non-null object
  LCB           89 non-null object
  CB           89 non-null object
  CCB          89 non-null object
  RB           89 non-null object
  Crossing      89 non-null float64
  Finishing     89 non-null float64
  HeadingAccuracy 89 non-null float64
  ShortPassing  89 non-null float64
  Volleys       89 non-null float64
  Dribbling     89 non-null float64
  Curve         89 non-null float64
  FkAccuracy    89 non-null float64
  LongPassing   89 non-null float64
  BallControl   89 non-null float64
  Acceleration  89 non-null float64
  SprintSpeed   89 non-null float64
  Agility       89 non-null float64
  Reactions     89 non-null float64
  Balance       89 non-null float64
  ShotPower     89 non-null float64
  Jumping       89 non-null float64
  Stamina       89 non-null float64
  Strength      89 non-null float64
  LongShots     89 non-null float64
  Aggression     89 non-null float64
  Interceptions 89 non-null float64
  Positioning   89 non-null float64
  Vision        89 non-null float64
  Penalties     89 non-null float64
  Composure     89 non-null float64
  Marking       89 non-null float64
  StandingTackle 89 non-null float64
  SlidingTackle 89 non-null float64
  GKDivng       89 non-null float64
  GKHandling    89 non-null float64
  GKCKicking    89 non-null float64
  GKPositioning 89 non-null float64
  GKReflexes    89 non-null float64
  Release Clause 89 non-null object
dtypes: float64(38), int64(6), object(45)
memory usage: 12.4+ MB
```

В загруженном датасете 89 колонок и 18207 строк. В колонках хранятся данные различного типа. В данных имеются пропуски, оценим их количество:

```
In [5]: for column in data.columns:
    if data[column].isna().sum() != 0:
        print(column, ': ', data[column].isna().sum(), ', data[column].dtype)

Club : 241 object
Preferred Foot : 48 object
International Reputation : 48 float64
Skill Moves : 48 float64
Work Rate : 48 object
Body Type : 48 object
Real Face : 48 object
Position : 68 object
Jersey Number : 68 float64
Joined : 1553 object
Loaned From : 16943 object
Contract Valid Until : 289 object
Height : 48 object
Weight : 48 object
LS : 2085 object
ST : 2085 object
RS : 2085 object
LW : 2085 object
LF : 2085 object
CF : 2085 object
RF : 2085 object
RW : 2085 object
LAM : 2085 object
CAM : 2085 object
RAM : 2085 object
LW : 2085 object
LF : 2085 object
CF : 2085 object
RF : 2085 object
RW : 2085 object
LCB : 2085 object
CB : 2085 object
CCB : 2085 object
RB : 2085 object
Crossing : 48 float64
Finishing : 48 float64
HeadingAccuracy : 48 float64
ShortPassing : 48 float64
Volleys : 48 float64
Dribbling : 48 float64
Curve : 48 float64
FkAccuracy : 48 float64
LongPassing : 48 float64
BallControl : 48 float64
Acceleration : 48 float64
SprintSpeed : 48 float64
Agility : 48 float64
Reactions : 48 float64
Balance : 48 float64
ShotPower : 48 float64
Jumping : 48 float64
Stamina : 48 float64
Strength : 48 float64
LongShots : 48 float64
Aggression : 48 float64
Interceptions : 48 float64
Positioning : 48 float64
Vision : 48 float64
Penalties : 48 float64
Composure : 48 float64
Marking : 48 float64
StandingTackle : 48 float64
SlidingTackle : 48 float64
GKDivng : 48 float64
GKHandling : 48 float64
GKCKicking : 48 float64
GKPositioning : 48 float64
GKReflexes : 48 float64
Release Clause : 1564 object
```

Согласно полученному заданию, необходимы выполнить обработку пропусков в данных для одного категориального и одного количественного признака. Выберем для этого категориальную колонку `Club`, и количественную `Weak Foot`.

## Обработка пропусков

### Обработка пропусков в категориальных столбцах

Посмотрим на уникальные значения столбца `Club`

```
In [6]: data['Club'].unique()
```

```
array(['FC Barcelona', 'Juventus', 'Paris Saint-Germain',
       'Manchester United', 'Manchester City', 'Chelsea', 'Real Madrid',
       'Atlético Madrid', 'FC Bayern München', 'Tottenham Hotspur',
       'Liverpool', 'Napoli', 'Arsenal', 'Milan', 'Inter', 'Lazio',
       'Borussia Dortmund', 'Vissel Kobe', 'Olympique Lyonnais', 'Roma',
       'Valencia CF', 'Guangzhou Evergrande Taobao FC', 'FC Porto',
       'FC Schalke 04', 'Besiktas JK', 'La Galaxy', 'Sporting CP',
       'Real Betis', 'Olympique de Marseille', 'RC Celta',
       'Bayer 04 Leverkusen', 'Real Sociedad', 'Villarreal CF',
       'Sevilla FC', 'SL Benfica', 'AS Saint-Étienne', 'AS Monaco',
       'Leicester City', 'Atalanta', 'Grêmio', 'Atlético Mineiro',
       'RB Leipzig', 'Ajax', 'Dalian Yifang FC', 'Everton',
       'West Ham United', '1. FC Köln', 'TSG 1890 Hoffenheim',
       'Shanghai SIPG FC', 'OGC Nice', 'Al Nassr',
       'Wolverhampton Wanderers', 'Borussia Mönchengladbach',
       'Hertha BSC', 'SV Werder Bremen', 'Cruzinho',
       'Athletic Club de Bilbao', 'Torino', 'Medipol Başakşehir FK',
       'Beijing Sinobo Guoan FC', 'Crystal Palace', 'RFC CSA Moscow',
       'VfL Wolfsburg', 'Shakhtar Donetsk', 'Toronto FC',
       'Lokomotiv Moscow', 'Sassuolo', 'New York City FC', 'Fluminense',
       'PSV', 'Levante UD', 'Fulham', 'Watford', 'Atlanta United',
       'Montpellier HSC', 'Galatasaray SK', 'Fenerbahçe SK', 'SD Eibar',
       'Los Angeles FC', 'Sampdoria', 'Al Hilal', 'VfB Stuttgart',
       'SC Braga', 'Silver Star', 'Deportivo Alavés', 'nan',
       'Eintracht Frankfurt', 'Girona FC', 'Guangzhou R&F FC', 'Burnley',
       'Stoke City', 'Southampton', 'Tianjin Quanjian FC', 'Getafe CF',
       'Beijing Renhe FC', 'Montreal Impact', 'Cilevo Verona', 'Genoa',
       'Portland Timbers', 'Tigres U.A.N.L.', 'RCD Espanyol',
       'Hebei China Fortune FC', 'Cagliari', 'Chicago Fire', 'DC United',
       'Sagan Tosu', 'Dynamo Kyiv', 'Santos', 'Internacional',
       'América FC (Minas Gerais)', 'Independiente', 'Boca Juniors',
       'Cruzeiro', '1. FSV Mainz 05', 'Bournemouth', 'Spartak Moscow',
       'Racing Club', 'FC Augsburg', 'Fiorentina', 'FC Nantes',
       'Feyenoord', 'Club Brugge KV', 'Brighton & Hove Albion', 'Al Ahli',
       'Jiangsu Suning FC', 'SC Freiburg', 'PAOK', 'Stade Rennais FC',
       'Trabzonspor', 'SPAL', 'Portimonense SC', 'Olympiacos CFP',
       'Club Atlético Huracán', 'Kashima S.K.', 'Newcastle United',
       'Frosinone', 'Queerstown', 'MKC GKS', 'Hannover 96',
       'Stade Malherbe Caen', 'Göteborg', 'Toulouse Football Club',
       'HSC Andorra', 'Huddersfield Town', 'CD Tenerife',
       'Seattle Sounders FC', 'Hamburger SV', 'FC Red Bull Salzburg',
       'Rio Ave FC', 'FC Girondins de Bordeaux', 'Melbourne Victory',
       'Pamir', 'FC Basel 1893', 'Al Wahda', 'BSC Young Boys', 'KAA Gent',
       'Al Ittihad', 'Standard de Liège', 'Shanghai Greenland Shenhua FC',
       'Colo-Colo', 'Junior FC', 'West Bromwich Albion',
       'FC Strasbourg Alsace', 'Göztepe SK', 'Deportivo Cali',
       'Deportivo Toluca', 'Bolíagaa', 'Nagoya Grampus', 'Anters SC',
       'Changchun Yatai FC', 'Club Atlético Lanús', 'BorFogo',
       'Club América', 'Udinese', 'Real Valladolid CF', 'CD Leganes',
       'Club Atlético Banfield', 'Celtic', 'Victoria Guimarães',
       'FC København', 'UD Las Palmas', 'Deportivo de La Coruña',
       'Universidad Católica', 'Caixa', 'Real Lorenzo de Almagro', 'Rayo Vallecano',
       'Monterrey', 'Columbus Crew SC', 'MKE Ankaragücü',
       'Guizhou Hengfeng FC', 'Swansea City', 'Tianjin Teda FC',
       'Chongqing Dangdai Lifan FC SC Team', 'AEK Athens', 'Al Taawoun',
       'Melbourne City FC', 'En Avant de Guingamp',
       'Akhmat Blednyesposon', 'Foggia', 'LDC Lille', '1. FC Nürnberg',
       'Clube Sport Marítimo', 'Real Sporting de Gijón', 'BB Erzurumspor',
       'Shandong Luneng Taishan FC', 'Club Atlético Colón', 'Hannu City SC',
       'Once Caldas', 'FC Groningen', 'Angers SCO', 'Paraná',
       'Antalyaspor', 'Mimense United FC', 'Club León', 'Empoli',
       'VVV-Venlo', 'Leeds United', 'Victoria Pilzen', 'Antalyaspor',
       'Atlético Paranaense', 'Derby County', 'Kawasaki Frontale',
       'Cardiff City', 'Aston Villa', 'Guadalajara', 'Bijon FCO',
       'Santos Laguna', 'Malaga CF', 'Victoria', 'Caykur Rizespor',
       'U.A.M.', 'Nottigham Forest', 'Royal Antwerp FC',
       'Club Tijuana', 'Sport Club do Recife', 'Real Salt Lake',
       'AZ Alkmaar', 'SK Slavia Praha', 'Willem II', 'Middlesbrough',
       'Dinamo Zagreb', 'Club Atlas', 'Granada CF', 'Sydney FC',
       'Sporting Kansas City', 'SV Zulte-Harogen', 'Philadelphia Union',
       'Real Oviedo', 'Pachuca', 'Boavista FC', 'Atiker Konyaspor',
       'Kaizer Chiefs', 'GD Chaves', 'Palermo', 'Atlético Nacional',
       'Puebla FC', 'Perth Glory', 'Panathinaikos FC', 'FC Sion',
       'Vitória de Setúbal', 'New York Red Bulls', 'Al Shabab',
       'Monarcas Morelia', 'Albacete BP', 'Rangers FC', 'Sparta Praha',
       'Legia Warszawa', 'Urawe Red Diamonds', 'Rosario Central',
       'Stade de Reims', 'ADO Den Haag', 'Chapcoense', 'FC Midtjylland',
       'San Jose Earthquakes', 'Helwegu de Córdoba', 'Brescia',
       'Kashima Antlers', 'CD Everton de Viña del Mar',
       'Fortuna Düsseldorf', 'SD Huesca', 'Preston North End',
       'Club Atlético Talleres', 'Benevento', 'Virtus',
       'Gimnasia y Esgrima La Plata', 'Houston Dynamo', 'Club Necaxa',
       'Monchiv', 'Holstein Kiel', 'Ittiffah FC', 'Kayserslautern',
       '1. FC Heidenheim 1846', 'Brenford', 'Vinci Halvatsaspor',
       'Lobos BUAP', 'Bursaspor', 'Ceará Sporting Club',
       'Sheffiled United', 'FC Ingolstadt 04', 'Estudiantes de La Plata',
       'Aix', 'Queens Park Rangers', 'Suwon Samsung Bluewings',
       'Heart of Midlothian', 'Reading', 'FC Dallas', 'Heracles Almelo',
       'Venezia FC', 'CD Lugo', 'Henan Jianye FC', 'Real Salt Lake',
       'CA Osasuna', 'MAC Breda', 'Livorno', 'Universidad de Chile',
       'Brandy IF', 'Aberdeen', 'Defensa y Justicia', 'Atlético Tucumán',
       'Blackburn Rovers', 'SV Darmstadt 98', 'Norrwiese FC',
       'Sanfrecce Hiroshima', 'CD Numancia', 'KV Ostende', 'FC Utrecht',
       'Vancouver Whitecaps FC', 'Odense Boldklub', 'SV Heerenveen',
       'Racing Club de Lens', 'Independiente Santa Fe',
       'Sporting de Charleroi', 'Millonarios FC', 'Sheffiled Wednesday',
       'Perguta', 'Daugu FC', 'Völz Sarisfeld',
       'Grasshopper Club Zürich', 'Slavasporn', 'Nîmes Olympique',
       'Rosenborg BK', 'SK Sturm Graz', 'FC Metz',
       'CD Universidad de Concepción', 'Hellas Verona', 'Brisbane Roar',
       'Real Zaragoza', 'CD Aves', 'Millwall', 'Unión de Santa Fe',
       'KAS Eupen', 'Cádiz CF', 'FC Tokyo', 'CD Tenerife',
       '1. FC Union Berlin', 'Al Fayha', 'AZ Alxerxa',
       'Patriotas Boyacá FC', 'Moldo FK', 'Bristol City', 'CD Nacional',
       'Sporting Lokeren', 'FC St. Pauli', 'Deportivo La Coruña',
       'Al Qadisiyah', 'Atlético Bucaramanga', 'Club Atlético Regal',
       'Club Atlético Alagoas', 'US Salernitana 1919', 'FC Lorient',
       'US Cremonese', 'VfL Bochum 1848', 'SK Rapid Wien',
       'KSV Cercle Brugge', 'Rionegro Águilas', 'Gimnastic de Tarragona',
       'Lecco', 'Santa Clara', 'FK Kieken', 'New England Revolution',
       'Orlando Pirates', 'Atlético Huila', 'Western Sydney Wanderers',
       'Kaiyari FF', 'Independiente Medellín', 'Fortuna Sittard',
       'Lech Poznań', 'Djurgårdens IF', 'FC Rous Deportiu', 'SV Brann',
       'Ulsan Hyundai FC', 'Sint-Truidense VV', 'Carpi', 'Al Fateh',
       'Royal Excel Mouscron', 'AC Ajaccio', 'FC Zwölle', 'Sunderland',
       'Club Atlético Alagoas', 'US Salernitana 1919', 'Harris Rovers',
       'San Martin de Tucumán', 'Incheon United FC', 'Siask Wrocław',
       'SV Kickspil', '1. FC Kaiserslautern', 'Deportivo Pasto',
       'Lincoln City', 'Motherwell', 'Rotherham United', 'Burton Albion',
       'Milla Ploot', 'FC Wacker Innsbruck', 'Preston North End',
       'Accoli', 'FC Zürich', 'Fleetwood Town', 'Padova',
       'FC Sochaux-Montbéliard', 'SV Mehen Wiesbaden', 'Unión La Calera',
       'Central Coast Mariners', 'Wigan Athletic',
       'Jagiellonia Białystok', 'Cittadella', 'Vibersian', 'FC Lugano',
       'San Martin de San Juan', 'Strensgodet IF', 'Júbilo Iwata',
       'Newell's Old Boys', 'Al Faisaly', 'Colorado Rapids',
       'IF Elfsborg', 'SV Sandhausen', 'Al Rayn', 'Stade Brestois 29',
       'UD Almería', 'Gyeongnam FC', 'Yokohama F. Marinos', 'Kilmarnock',
       'Pescara', 'Newcastle Jets', 'Coritiba CF', 'RCD Mallorca',
       'Hammarby IF', 'Cerezo Osaka', 'KFC Uerdingen 05',
       'Shizuwa S-Pulse', 'MVS Oaxaca', 'Os Belenenses',
       'TSC Arminia Bielefeld', 'Ipwich Town', 'FC Seoul',
       'Lechia Gdansk', 'Gamba Osaka', 'FC Rayo Majadahonda', 'LASK Linz',
       'Rotherham Wanderers', 'Al Raed', 'Extremadura UD', 'SC Paderborn 07',
       'Wellington Phoenix', 'Unión Española', 'Alanta Petrolina',
       'Bacovia', 'Gangwon FC', 'Elche CF', 'ESTAC Troyes', 'AS Béziers',
       'La Berrichonne de Châteauroux', 'Claremont Foot 63',
       '1. FC Magdeburg', 'Pohang Steelers', 'Orebro SK', 'Arka Gdynia',
       'SG Dynamo Dresden', 'Sovig Greuther Fürth', 'CD Huachipato',
       'Milla Kraków', 'Stade football', 'Eintracht Braunschweig',
       'Valenciennes FC', 'FC Thun', 'San Luis de Quillota',
       'SV Jahn Regensburg', 'Cosenza', 'FC Nordsjælland',
       'FC Erzgebirge Aue', 'Joanman Dragons', 'Wolfsburger AC',
       'Charlton Athletic', 'Greenhill Foot 88', 'FC Hansa Rostock',
       'San Martin de Tucumán', 'Incheon United FC', 'Siask Wrocław',
       'FC Ajaccio', '1. FC Kaiserslautern', 'Deportivo Pasto',
       'Lincoln City', 'Motherwell', 'Rotherham United', 'Burton Albion',
       'Milla Ploot', 'FC Wacker Innsbruck', 'Preston North End',
       'Accoli', 'FC Zürich', 'Fleetwood Town', 'Padova',
       'FC Sochaux-Montbéliard', 'SV Mehen Wiesbaden', 'Unión La Calera',
       'Central Coast Mariners', 'CD Hüggingen', 'CD Antofagasta',
       'Plymouth Argyle', 'Aarnen 05', 'Lillistown SK', 'Karlsruher SC',
       'GIF Sundsvall', 'FC Emmen', 'Barnsley', 'Audax Italiano',
       'V-Aves Lugano', 'Paris FC', 'Sovig Greuther Fürth', 'Harris Rovers',
       'De Graafschap', 'Mokkaido Consadole Sapporo', 'Tromsø IL',
       'FC Luzern', 'FC Huesumund', 'Zagłębie Lubin', 'VfR Aalen',
       'Dundee', 'Oxford United', 'Plant Glislice', 'Ond Club',
       'Ostersunds FK', 'Vegata Senda', 'Crawley Town',
       'FC Admira Wacker Mödling', 'Vallengren Football', 'Dundee FC',
       'Preston North End', 'Envidado FC', 'Mied Legnica', 'Odds BK',
       'SC Fortuna Köln', 'US Orléans Loiret Football', 'Sarsborg 08 FF',
       'Jaguars de Córdoba', 'Breadford City', 'Accrington Stanley',
       'St. Johnstone FC', 'Boysa Claid FC', 'Luton Town',
       'SV Mittersberg', 'Kristiansund BK', 'Sangu Sangmu FC',
       'Machado', 'Walsall', 'Korona Kielce', 'Shonan Bellmare',
       'FC Würzburger Kickers', 'FSV Zwickau', 'St. Mirren', 'AC Horsens',
       'Esbjerg fB', 'HJK Helsinki', 'Southend United', 'Bristol Rovers',
       'Hawilton Academical FC', 'TSV 1860 München', 'Curico Unido',
       'SCR Altach', 'Ranheim Football', 'Stevenage',
       'SG Sonnenhof Großaspach', 'Oldham Athletic', 'Milton Keynes Dons',
       'FC Bodogliet', 'SC Preußen Münster', 'Noyceba Wanders',
       'Vejle Boldklub', 'Bury', 'Randers FC', 'VfL Osnabrück',
       'Sonderjyske', 'IFK Göteborg', 'Mansfield Town', 'Coventry City',
       'Wateford FC', 'Shrewsbury', 'IF Start', 'Gillingham',
       'FC Energie Cottbus', 'FC Carl Zeiss Jena', 'Halleischer FC',
       'SV Meppen', 'AFC Wimbledon', 'Blackpool', 'Doncaster Rovers',
       'Sanderford Football', 'VfL Sportfreunde Lotte', 'Cheltenham Town',
       'IK Sirius', 'Vendsyssel FF', 'Swindon Town', 'Motts County',
       'SOM St. Pölten', 'Exeter City', 'Northampton Town',
       'Sharnock Rovers', 'Colchester United', 'Livingston FC',
       'TSV Hartberg', 'Frankenau Rovers', 'Cambridge United',
       'Grimsby Town', 'Port Vale', 'Itapig Lions FC',
       'Forest Green Rovers', 'Dukrud FF', 'Zagłębie Sosnowice',
       'Carlisle United', 'Trillaborg FF', 'St. Patrick's Athletic',
       'Morecambe', 'Cork City', 'IF Brommapojkarna', 'Cresce Alexandra',
       'Yeovil Town', 'Bohemian FC', 'Macclesfield Town',
       'Newport County', 'Sligo Rovers', 'Derry City', 'Limerick FC',
       'Bray Wanderers', dtype=object)
```

Для заполнения пропусков используем SimpleImputer из библиотеки sklearn. Для заполнения применим стратегию most\_frequent, которая заменит NaN значения наиболее часто встречающимися в данной колонке.

```
In [7]: #Выбираем нужную нам колонку
cat_temp_data = data['Club']

# Создаем объект класса SimpleImputer, устанавливаем тип пропущенных значений и стратегию импутирования
imp = SimpleImputer(missing_values=np.nan, strategy='most_frequent')
cat_temp_data = imp.fit_transform(cat_temp_data)

# Преобразуем полученное массив значений обратно в столбцы датасета
my_list = nap(lambda x: x[0], cat_temp_data)
data['Club'] = pd.Series(my_list)

#дополнительная проверка результата работы
data['Club'].isna().sum()
```

Out[7]: 0

Пропуски в выбранном категориальном столбце устранены

### Обработка пропусков в числовых столбцах

Посмотрим на уникальные значения столбца `Weak Foot`

```
In [8]: data['Weak Foot'].unique()
```

```
Out[8]: array([ 4.,  5.,  3.,  2.,  1., nan])
```

Построим гистограмму распределения значений в выбранном нами столбце

```
In [9]: data['Weak Foot'].hist()
```

```
Out[9]: <matplotlib.axes._subplots.AxesSubplot at 0x138349685c8>
```

Для замены воспользуемся средствами библиотеки pandas и заполним пропуски средним значением по данной колонке, поскольку выбросов(аномально больших или малых значений) нет, то данный способ уместен.

```
In [10]: data['Weak Foot'] = data['Weak Foot'].fillna(data['Weak Foot'].mean())

In [11]: data['Weak Foot'].hist()
```

```
Out[11]: <matplotlib.axes._subplots.AxesSubplot at 0x138349cd688>
```

## Ответы на контрольные вопросы:

- Какие признаки Вы будете использовать для дальнейшего построения моделей машинного обучения и почему?

Для решения задачи регрессии и предсказания параметра `Overall`, т.е. для предсказания рейтинга карточки игрока, формирующегося на основе остальных его характеристик можно применить:

- Линейную регрессию - т.к. зависимость между количественными признаками и итоговим рейтингом(аномально больших или малых значений) нет, поэтому можно использовать в качестве линейной модели
- Метод опорных векторов - в данном случае линейные выбросы из-за того, что они уже стандартизированы в рамках рейтинговой системы, из-за которых вызывают также достаточно сложный метод, который может справиться и с проблемой нелинейной зависимости целевого признака от параметров.