

**Московский Авиационный Институт
(Национальный исследовательский университет)**

Кафедра 806

**Отчет
по лабораторной работе №0
по курсу «Искусственный интеллект»**

Студент: Ильиных В. М.

Группа: М8О-301Б-19

Москва, 2022

Описание

Футбол - один из самых популярных видов спорта в мире. Людей привлекают эмоции, страсть, талант игроков, умные и неожиданные решения тренеров. Многие мальчишки и девчонки видят в футболистах кумиров.

Такой популярный спорт не может остаться без денег. Как футболисты, так и сами клубы зарабатывают огромное их количество: от продаж билетов, от спонсоров, от пожертвований фанатов. У каждой команды есть свои скаутские сети, но кто сказал, что не может быть отдельного сервиса по анализу футболистов и в целом футбольного мира? На основе популярного - а, возможно, и самого главного - футбольного источника я хочу воплотить это в жизнь.

Постановка задачи

Проанализировать трансферы за 2021 год: от наиболее востребованных футболистов по позициям до их стоимости. Визуализировать данные и сделать выводы.

Ход решения

Чтобы точно проанализировать футбольные трансферы, следует вначале разобраться с типами данных.

Как видно, в выбранной базе данных все столбцы являются строками, кроме одного – возраста.

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

	player_name	position	age	team_origin	country_origin	league_origin	team_dest	country_dest	league_dest	transfer_fee	date
0	Dominik Szoboszlai	Left Winger	20.0	RB Salzburg	Austria	Bundesliga	RB Leipzig	Germany	Bundesliga	€22.00m	2021-01-01
1	Filip Stevanovic	Left Winger	18.0	Partizan	Serbia	Super liga Srbije	Man City	England	Premier League	€8.50m	2021-01-01
2	Brenden Aaronson	Attacking Midfield	20.0	Philadelphia	United States	MLS	RB Salzburg	Austria	Bundesliga	€5.45m	2021-01-01
3	Víctor Dávila	Centre-Forward	23.0	CF Pachuca	Mexico	Liga MX Clausura	León	Mexico	Liga MX Clausura	€4.50m	2021-01-01
4	Carlos González	Centre-Forward	27.0	UNAM Pumas	Mexico	Liga MX Clausura	Tigres UANL	Mexico	Liga MX Clausura	€4.48m	2021-01-01

Проверим типы данных

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 65489 entries, 0 to 65488
Data columns (total 11 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   player_name      65489 non-null  object
1   position         65486 non-null  object
2   age              65489 non-null  float64
3   team_origin      65489 non-null  object
4   country_origin   60249 non-null  object
5   league_origin    60249 non-null  object
6   team_dest        65489 non-null  object
7   country_dest     55737 non-null  object
8   league_dest      55737 non-null  object
9   transfer_fee     65485 non-null  object
10  date             65489 non-null  object
dtypes: float64(1), object(10)
memory usage: 5.5+ MB
```

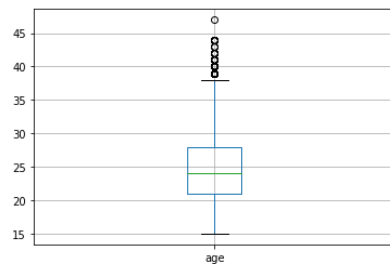
Далее, данные следует проверить на выброс. Сейчас у нас только один числовой столбец – возраст, поэтому пока что проверим его.

```
In [6]: data['age'].quantile([0.005,.01,.05,.1,.5,.9,.95,.99,.995,.999])
```

```
Out[6]: 0.005    17.0  
        0.010    18.0  
        0.050    19.0  
        0.100    19.0  
        0.500    24.0  
        0.900    32.0  
        0.950    34.0  
        0.990    37.0  
        0.995    38.0  
        0.999    40.0  
        Name: age, dtype: float64
```

```
In [7]: data.boxplot('age')
```

```
Out[7]: <AxesSubplot:>
```



Видно, что выбросы есть. Далее, при выполнении следующей лабораторной работы я избавлюсь от них.

Для удобства в данной лабораторной работе добавим в базу данных несколько столбцов – сумму трансфера в числовом формате и тип трансфера. Для этого напомним функцию.

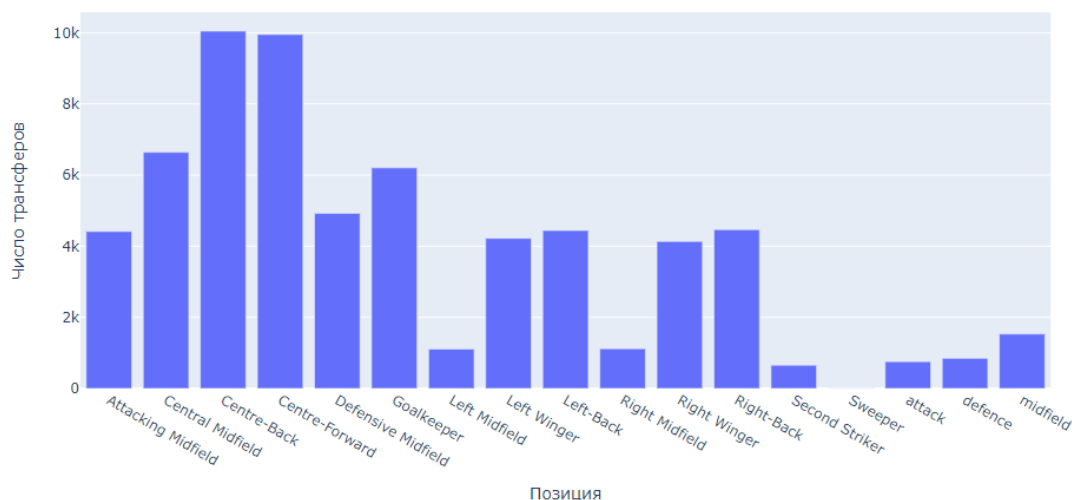
```
def clean_transfer_fee(dataframe):  
    df = dataframe.copy()  
    df = df.assign(date=lambda x: pd.to_datetime(x.date), age=lambda x: x.age.astype("uint8"))  
    x = df.transfer_fee  
    options = [  
        x == "free transfer",  
        x.str.contains("loan", flags=re.IGNORECASE, na=False),  
        x == "?",  
        x == "draft",  
        x == "-",  
    ]  
    values = ["free transfer", "loan", "unknown", "draft", "unknown"]  
    df["transfer_type"] = np.select(options, values, "transfer")  
    df["transfer_amount"] = np.where(df.transfer_fee.str.contains("\d+"), df.transfer_fee, "-1")  
    df["transfer_amount"] = df.transfer_amount.fillna("-1").str.extract(r"(-?\d+.*$)")  
    df["transfer_amount"] = df.transfer_amount.str.replace(  
        "m", "*1000000", regex=False  
    ).str.replace("Th.", "*1000", regex=False)  
    df["transfer_amount"] = df.transfer_amount.astype(str).apply(eval)  
    df["transfer_amount"] = df.transfer_amount.replace(-1, np.nan)  
    return df
```

Все подготовительные шаги готовы – настало время графиков и анализа.

Визуализации и выводы:

1)

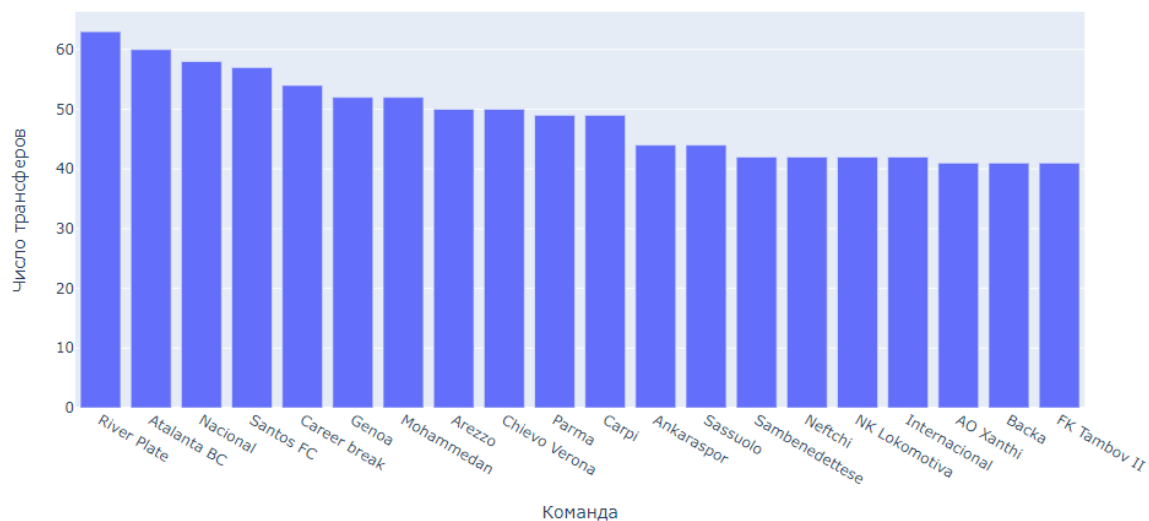
Трансферы по амплуа футболистов



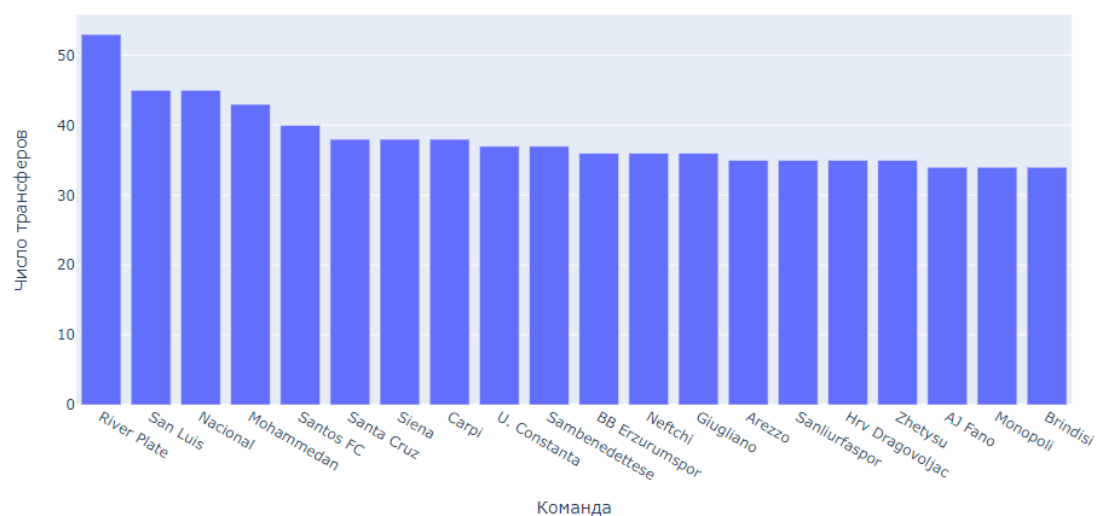
Чаще всего покупают и продают центральных защитников и центральных нападающих. Можно сделать вывод, что это важнейшие позиции. Реже всего – крайних полузащитников. Скорее всего, футболисты на этих позициях используются гораздо меньше, нежели на других.

2)

Продажи



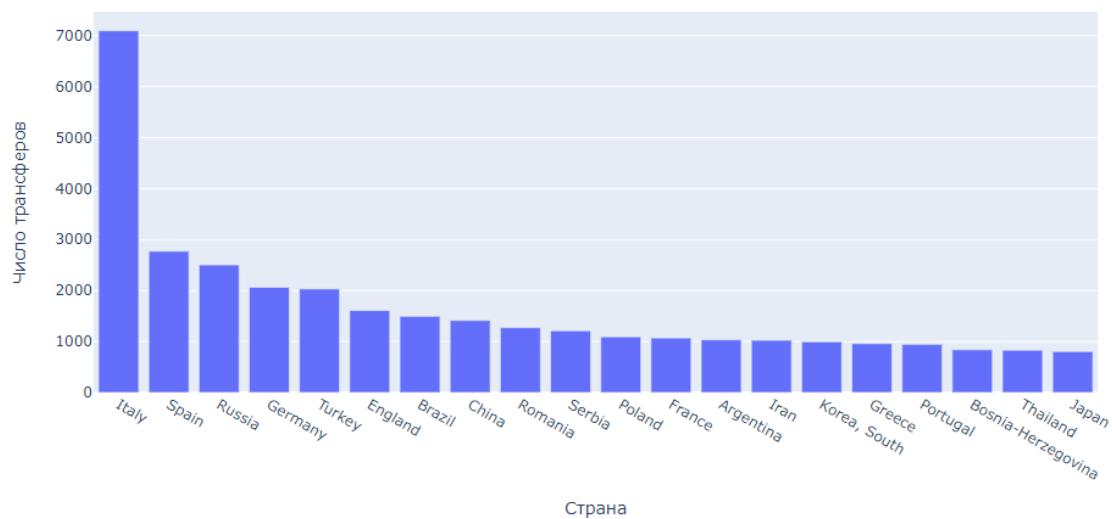
Покупки

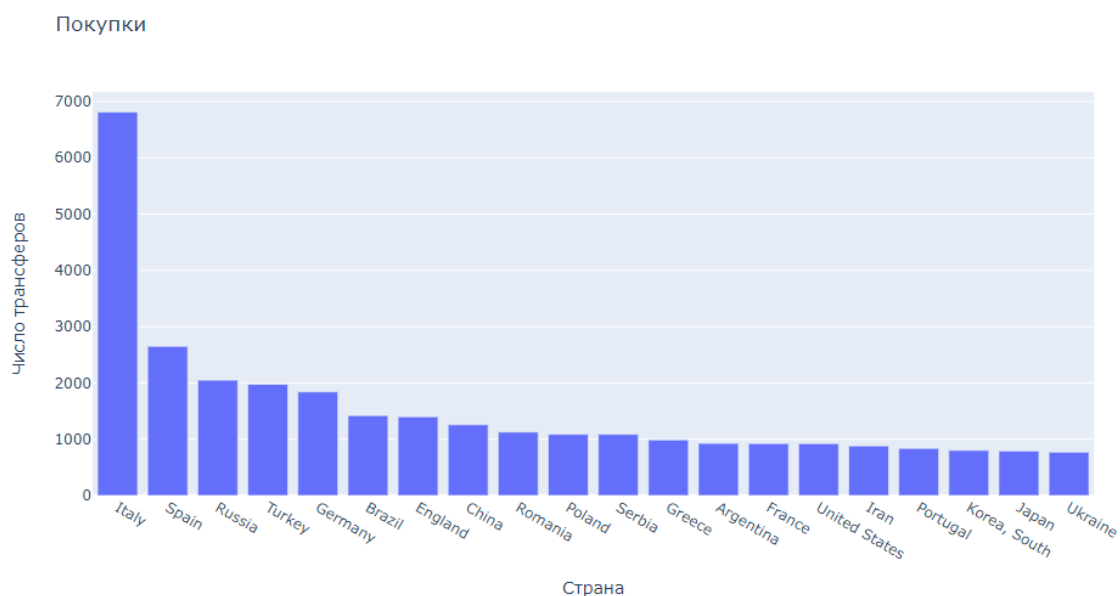


Больших выводов по данным визуализациям сделать нельзя, но можно сказать, что River Plate лидирует и по покупкам, и по продажам футболистов.

3)

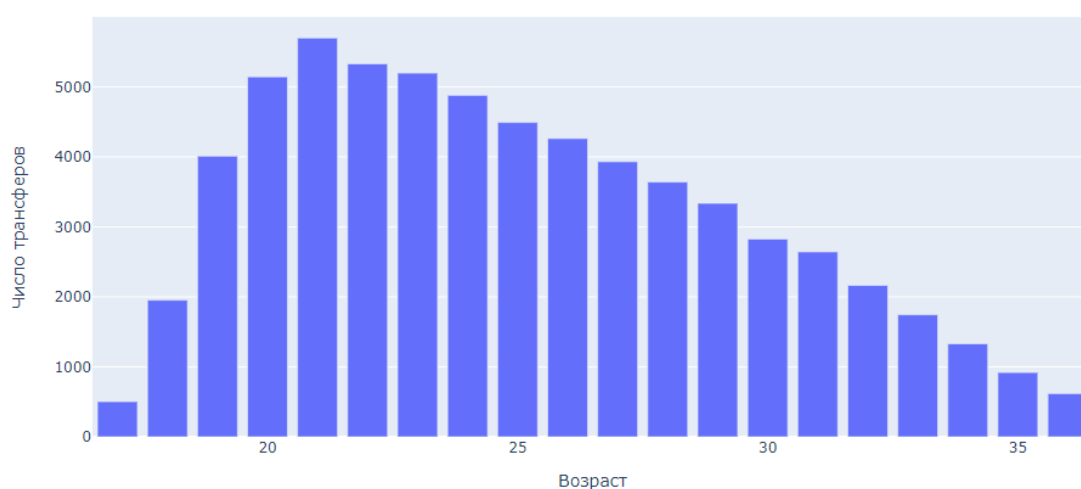
Продажи





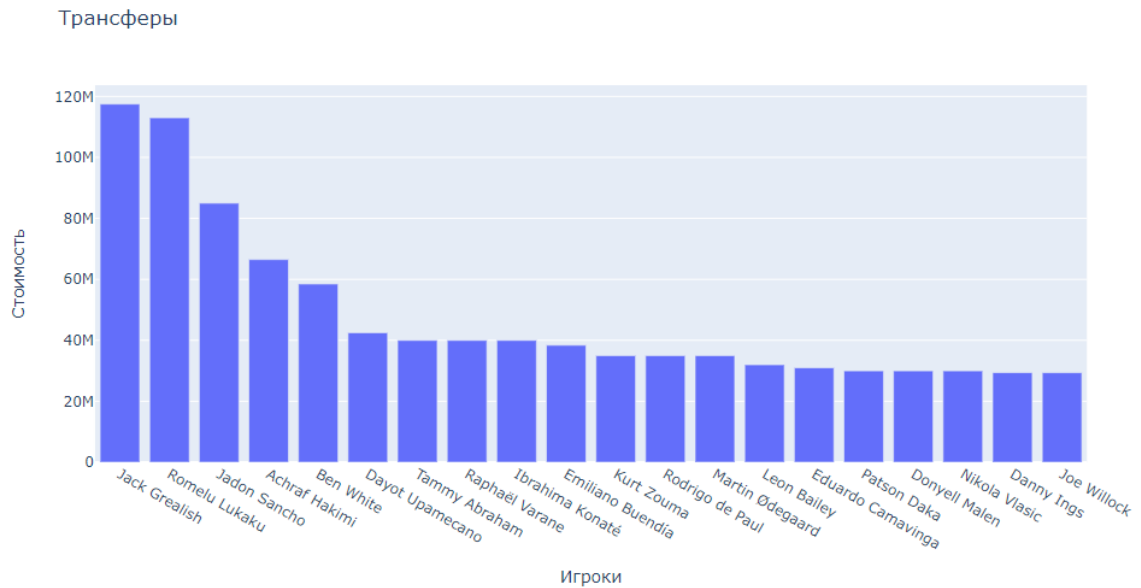
Колоссальный перевес итальянских клубов в трансферах. Вывод – итальянские клубы чаще остальных меняют своих футболистов. О причинах подобного можно только додумываться – их не устраивают нынешний состав, их футболистов скупают более богатые клубы, или есть другие причины.

4)



Больше всего – что логично – покупают молодых футболистов, от 20 до 24 лет. Они наиболее перспективные и в дальнейшем могут либо стать лидерами команды, либо быть проданными за большую сумму.

5)



Теперь самые дорогие трансферы. Видно, что вырвались вперед лишь пятеро, остальные были куплены примерно за одинаковую сумму. Это свидетельствует о том, что цены на «среднего» держатся на одном уровне.

6)



Примерное равенство покупок и продаж футболистов – что логично. Но небольшой перевес в сторону продаж присутствует.