COGS 118A - Final Project

THE FIFA TRANSFERMARKET PREDICTION NOTEBOOK

Group members

- Khalid Ade
- Pablo Moreno
- Sujay Srinivasan
- Daniel Vega Lojo
- Jared Chen

Abstract

Association Football (or soccer) is a worldwide sport played by over 250 million players in over 250 countries[1]. In fact, football is the world's sport, and the most popular across the globe in terms of fans as well. Football has a huge transfer market in which players are transferred across teams for up to hundreds of millions of euros. To put the amount of money that circulates in the global football market into perspective, squad values of top teams like Manchester United surpass billions of euros[2]. The market value of a player accounts for a huge role in how teams conduct their business in regards to transfers. Our goal is to help these clubs make the right investments in players they want to obtain, especially when spending huge amounts of money. More specifically, we want to accurately predict the market value of players so that clubs aren't overpaying, or underselling their valued players. Plenty of factors play a role in determining the market value of a player. The most important factors include age, performance for club and national team (measured in stats such as goals, assists, tackles etc.) for a player in that position, experience (measured by number of seasons in top leagues), marketing value (measured by social media presence), and injury vulnerability[2].

Background

The global football transfer market involves the circulation of billions of euros. Many top European clubs have spent hundreds of millions of dollars to bolster their respective teams. For example teams like Manchester United, Manchester City, and PSG have spent almost billions of euros to sign players to help their teams' success in their respective leagues and on the European stage[6]. There is no doubt that decisions involving huge sums of such money should be carefully analyzed so that clubs can maximize success in both the business side as well as the performance side of their respective clubs. Transfermarkt is an online platform for transfers, market values, rumors, and stats. The business model consists of, in addition to sports journalistic reporting, the profiles of the players and discussion forums on the performance and market values of individual soccer players, teams and leagues[4]. Frequently being discussed in sports science and sports economics literature over the past few years, the so-called "market values" („Marktwerte“) have s to become the center of media attentioMultipleous studies have shown positive correlations between the predicted market values on Transfermarkt and the actual player income.'It's reportedly known that players who are in contract negotiations would sometimes refer to Transfermarkt values as baselines for their salary expectations[4]. The "market values" can also be used as a measure of marketability; a higher marketability helps a player secure partnerships through sponsorship contracts. The age and performance statistics on Transfermarkt are also particularly useful in that player observers can identify young players and predict the development opportunities[4].

The open forums of Transfermarkt allow users to discuss and predict individual players' market values and performance. Previous studies on collective intelligence[2] have used OLS regression models to evaluate the accuracy of predictions. It is shown that "forecasts of international soccer results based on the crowd's valuations are more accurate than those based on standard predictors."[3] This reveals a potential possibility that distributed intelligence is a contributing factor to the accuracy of predictions. We want to know if supervised machine learning algorithms, as another form of distributed intelligence, can make accurate predictions just as humans do. More particularly, we want to use machine learning models like OLS to predict market value of players across the football world.

Problem Statement

Given the considerable number of players in football across the globe, it can get tedious to know which players have potential and are worth investing in. Do they have high performance for a player in their position? Are they playing for a renowned club or in a renowned league? Is their behavior respectable and are they marketable?

These are the kinds of questions top clubs use when considering paying the big bucks for players. The problem we are trying to tackle is predicting the market value of players (in euros) using stats that are important when investing in a player such as goals, assists, and marketability.

Data

The dataset[7] is composed of 7 different subsets, we will be using 4 of the datasets. Since each feature resides in different sets.

- `Appearances.csv`
  - Player ID, Game ID, Appearance ID, Competition ID, Player club ID, Assist, Minutes Played, Yellow cards, Red Cards
- `Clubs.csv`
  - Club ID, Name, Pretty_name, Domestic_competition_id, Total_market_value, Squad_size, Average_age, Foreigners_numbers, Foreigners_percentage, National_team_players, Stadium_name, Stadium_seats, Net_transfer_record, Coach_name, URL
- `Competitions.csv`
  - Competition_id, Name, type, country_id, country_name, domestic_league_code, confederation, URL.
- `Games.csv`
  - Game_id, Competition_code, Season, Round, Date, Home_club_id, Away_club_id, Home_club_goals, away_club_goals, Home_club_postions, Away_club_postion, Stadium, Attendance, Referee, URL
- `Leagues.csv`
  - League_id, name, Confederation
- `Player_valuations.csv`
  - Player_id, Date, Market_value
- `Players.csv`
  - Player_id, Last_season, Current_club_id, Name, Pretty_name, country_of_birth, Country_of_citizenship, Date_of_birth, Position, Sub_position, Foot, Height_in_cm, Market_value_in_gbp, Highest_market_value_in_gbp, URL

- *What an observation consists of*: We are trying to use the variables we assume to be the most important and independent from each other. We decided on
    - Club, Nationality, Minutes, Goals, Assist, Age, Conduct, Years Played, Position, Physicality.
- *What some critical variables are, how they are represented*: We want variables which have the highest co-variance with each other. The metric should handle most features as unique features.
- *Any special handling, transformations, cleaning, etc will be needed*: There will be club names, and probably inferences in our data. Such as Media Presence or Potential, these are metrics which can be objective to the person. How popular is the player that we are analyzing?

We are still going to be in search of more databases that might have different descriptive data that we might like to see how organizations search for talent. We can use what they might describe as their most sought out characteristics.

For simplicity we can also assume that all players have no contracts for their evaluation and are based solely on performance and the other variables mentioned.

```
import sys
```

```
import re
_r = re.escape
def _re_replace(s : str, to_replace : dict):
    for p, r in to_replace.items():
        s = re.compile(p).sub(r, s)
    return s
```

```
import matplotlib
import matplotlib.pyplot as plt
%matplotlib inline
%config InlineBackend.figure_formats = ['svg']
```

```
!{sys.executable} -m pip install --quiet pandas
import pandas as pd
```

```
!{sys.executable} -m pip install --quiet seaborn
import seaborn as sns
```

```
# OLS using statsmodels
!{sys.executable} -m pip install --quiet statsmodels numpy
import statsmodels.api as sm
import numpy as np
```

```
/opt/conda/lib/python3.9/site-packages/statsmodels/tsa/base/tsa_model.py:
7: FutureWarning: pandas.Int64Index is deprecated and will be removed fro
m pandas in a future version. Use pandas.Index with the appropriate dtype
instead.
  from pandas import (to_datetime, Int64Index, DatetimeIndex, Period,
/opt/conda/lib/python3.9/site-packages/statsmodels/tsa/base/tsa_model.py:
7: FutureWarning: pandas.Float64Index is deprecated and will be removed f
rom pandas in a future version. Use pandas.Index with the appropriate dty
pe instead.
  from pandas import (to_datetime, Int64Index, DatetimeIndex, Period,
```

```
!{sys.executable} -m pip install --quiet sklearn
!{sys.executable} -m pip install --quiet patsy
import sklearn as skl

'''
!{sys.executable} -m pip install --quiet scikit-learn-intelex
import sklearnex as sklx
sklx.patch_sklearn()
'''

import sklearn.linear_model

from sklearn.compose import ColumnTransformer
from sklearn.datasets import fetch_openml
from sklearn.pipeline import Pipeline
from sklearn.impute import SimpleImputer
from sklearn.preprocessing import StandardScaler, OneHotEncoder
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.metrics import mean_squared_error
from sklearn.linear_model import Lasso
from sklearn.linear_model import ElasticNet
from sklearn.model_selection import KFold
import scipy.stats as stats
import patsy
```

```
_data_ = {
    name: pd.read_csv(
        file,
        engine = 'c',
        low_memory = True,
        memory_map = False, # set `False` to load into memory
        **kwargs
    ) for name, file, kwargs in [
        ('appearances', 'data/appearances.csv', {
            'dtype': {
                'player_id': 'object',
                'game_id': 'object',
                'appearance_id': 'object',
                'competition_id': 'object',
                'player_club_id': 'object'
```

```
            }
        }),
        ('clubs', 'data/clubs.csv', {
            'dtype': {
                'club_id': 'object'
            }
        }),
        #('competitions', 'data/competitions.csv', {}),
        ('games', 'data/games.csv', {
            'dtype': {
                'game_id': 'object'
            }
        }),
        #('leagues', 'data/leagues.csv', {}),
        ('players', 'data/players.csv', {
            'parse_dates': ['date_of_birth'],
            'dtype': {
                'player_id': 'object',
                'country_of_birth': 'category',
                'country_of_citizenship': 'category',
                'position': 'category',
                'sub_position': 'category'
            }
        }),
        ('player_valuations', 'data/player_valuations.csv', {
            'parse_dates': ['date'],
            'dtype': {
                'player_id': 'object'
            }
        })
    ]
}
```

```
data = {}
```

```
# clubs
data['clubs'] = _data_['clubs'].copy()

data['clubs'] = data['clubs'][[
    'club_id',
    'pretty_name'
]]
data['clubs'].rename(
    columns = {'pretty_name': 'club_name'},
    inplace = True
)
data['clubs'].set_index('club_id', inplace = True)

data['clubs']
```

	club_name
club_id	
1032	Fc Reading
2323	Orduspor
1387	Acn Siena 1904
3592	Kryvbas Kryvyi Rig
1071	Wigan Athletic
...	...
1269	Pec Zwolle
200	Fc Utrecht
317	Fc Twente Enschede
3948	Royale Union Saint Gilloise
1304	Heracles Almelo

801 rows × 1 columns

```
# games
data['games'] = _data_['games'].copy()

data['games'] = data['games'][[
    'season',
    'game_id'
]]
data['games'].set_index('game_id', inplace = True)

data['games']
```

	season
game_id	
2244388	2012
2219794	2011
2244389	2012
2271112	2012
2229332	2012
...	...
3646190	2021
3646188	2021
3655616	2021
3655629	2021
3646191	2021

56028 rows × 1 columns

```
# appearances
data['appearances'] = _data_['appearances'].copy()

data['appearances'] = data['appearances'].loc[
    :, ~data['appearances'].columns.isin([
        'appearance_id',
        'competition_id'
    ])
]
data['appearances'].rename(
    columns = {'player_club_id': 'club_id'},
    inplace = True
)

data['appearances'] = (
    data['appearances']
        .merge(
            data['games'],
            on = 'game_id',
            copy = False
        ).drop(columns = 'game_id')
        .merge(
            data['clubs'],
            on = 'club_id',
            copy = False
        ).drop(columns = 'club_id')
)
```

```python
data['appearances'] = (
    data['appearances']
        .groupby(['player_id', 'season'])
        .agg({
            **{
                c: 'sum' for c in [
                    'goals',
                    'assists',
                    'minutes_played',
                    'yellow_cards',
                    'red_cards'
                ]
            },
            'club_name': 'last'
        })
        .reset_index('season')
)

data['appearances']
```

	season	goals	assists	minutes_played	yellow_cards	red_cards	club_name
player_id							
10	2014	32	18	4578	12	0	Lazio Rom
10	2015	16	14	3428	6	0	Lazio Rom
100009	2014	0	0	5576	8	0	Kuban Krasnodar
100009	2015	2	2	4512	12	0	Kuban Krasnodar
100009	2016	0	0	1260	6	0	Anzhi Makhachkala
...	...	...	...	...	...	...	...
99923	2014	0	2	832	4	0	Cagliari Calcio
99924	2016	0	2	1824	6	0	Ca Osasuna
99977	2014	0	0	194	0	0	Rcd Mallorca
99977	2015	10	6	3046	2	0	Royal Excel Mouscron
99977	2019	0	0	716	0	0	Caykur Rizespor

54216 rows × 7 columns

```python
# player valuations
```

```python
data['player_valuations'] = _data_['player_valuations'].copy()

data['player_valuations']['season'] = (
    pd.DatetimeIndex(data['player_valuations']['date']).year
)
data['player_valuations'].drop(columns = 'date', inplace = True)

data['player_valuations'] = (
    data['player_valuations']
        .groupby(['player_id', 'season'])
        .agg({'market_value': 'mean'})
        .reset_index('season')
)
data['player_valuations'].rename(
    columns = {'market_value_in_gbp': 'market_value'},
    inplace = True
)

data['player_valuations']
```

	season	market_value
player_id		
10	2004	6300000.0
10	2005	10800000.0
10	2006	22500000.0
10	2007	20700000.0
10	2008	18000000.0
...	...	...
99977	2018	990000.0
99977	2019	720000.0
99977	2020	562500.0
99977	2021	495000.0
99977	2022	540000.0

181182 rows × 2 columns

```python
# players
data['players'] = _data_['players'].copy()

data['players'] = data['players'].loc[
    :, ~data['players'].columns.isin([
        'last_season',
        'name',
```

```
        'current_club_id',
        'market_value_in_gbp',
        'highest_market_value_in_gbp',
        'country_of_birth',
        'url',
        'foot'
    ])
]
data['players'].rename(
    columns = {
        'pretty_name': 'name',
        'height_in_cm': 'height',
        'country_of_citizenship': 'nationality'
    },
    inplace = True
)

data['players']['sub_position'] = (
    data['players']['sub_position'].cat
        .rename_categories(
            lambda s: (
                _re_replace(s, {
                    fr'''^(.*){_r(' - ')}(.*)$''': r'\2'
                })
                .title()
            )
        )
)

data['players'].set_index('player_id', inplace = True)

data['players']
```

player_id	name	nationality	date_of_birth	position	sub_position	height
254016	Arthur Delalande	France	1992-05-18	Midfield	Central Midfield	186
51053	Daniel Davari	Iran	1988-01-06	Goalkeeper	Goalkeeper	192
31451	Torsten Oehrl	Germany	1986-01-07	Attack	Centre-Forward	192
44622	Vladimir Kisenkov	Russia	1981-10-08	Defender	Right-Back	182
30802	Oscar Diaz	Spain	1984-04-24	Attack	Centre-Forward	183
...	...	...	...	...	...	...
462285	Fabian De Keijzer	Netherlands	2000-05-10	Goalkeeper	Goalkeeper	193
368612	Merveille Bokadi	DR Congo	1996-05-21	Defender	Centre-Back	186
408574	Joey Veerman	Netherlands	1998-11-19	Midfield	Central Midfield	185
364245	Jordan Teze	Netherlands	1999-09-30	Defender	Centre-Back	183
575367	Richard Ledezma	United States	2000-09-06	Attack	Attacking Midfield	174

23682 rows × 6 columns

```python
# final dataset
data['all'] = data['players'].merge(
    data['player_valuations'].merge(
        data['appearances'],
        on = ['player_id', 'season'],
        copy = False
    ),
    on = 'player_id',
    copy = False
)

data['all']['age'] = (
    pd.to_datetime(data['all']['season'], format = '%Y', utc = True)
        - pd.to_datetime(data['all']['date_of_birth'], utc = True)
).astype('timedelta64[Y]')
data['all'].drop(columns = 'date_of_birth', inplace = True)

data['all'].dropna(axis = 'index', inplace = True)

data['all']
```

player_id	name	nationality	position	sub_position	height	season	market_value
9800	Artem Milevskyi	Ukraine	Attack	Centre-Forward	189	2020	90000.0
43084	Gaetano Berardi	Switzerland	Defender	Right-Back	179	2020	360000.0
230826	Gennaro Acampora	Italy	Midfield	Central Midfield	174	2020	360000.0
198087	Matteo Ricci	Italy	Midfield	Defensive Midfield	176	2020	1530000.0
110689	Deniz Mehmet	Turkey	Goalkeeper	Goalkeeper	192	2020	68000.0
...	...	...	...	...	...	...	...
364245	Jordan Teze	Netherlands	Defender	Centre-Back	183	2019	420000.0
364245	Jordan Teze	Netherlands	Defender	Centre-Back	183	2020	1102500.0
364245	Jordan Teze	Netherlands	Defender	Centre-Back	183	2021	5400000.0
575367	Richard Ledezma	United States	Attack	Attacking Midfield	174	2020	658250.0
575367	Richard Ledezma	United States	Attack	Attacking Midfield	174	2021	765000.0

50781 rows × 14 columns

Evaluation

```
data['all'][data['all'].isna().any(axis = 1)]
```

player_id	name	nationality	position	sub_position	height	season	market_value	goals

```
data['all'].dtypes
```

```
name                 object
nationality        category
position           category
sub_position       category
height                int64
season                int64
market_value        float64
goals                 int64
assists               int64
minutes_played        int64
yellow_cards          int64
red_cards             int64
club_name            object
age                 float64
dtype: object
```

```
data['all'].describe()
```

	height	season	market_value	goals	assists	minutes
count	50781.000000	50781.000000	5.078100e+04	50781.000000	50781.000000	50781
mean	180.794628	2017.380063	3.630890e+06	3.880546	2.949883	2805
std	17.703409	2.318805	8.274637e+06	7.352176	4.793814	2103
min	0.000000	2013.000000	9.000000e+03	0.000000	0.000000	2
25%	178.000000	2015.000000	3.600000e+05	0.000000	0.000000	884
50%	182.000000	2017.000000	9.000000e+05	0.000000	2.000000	2566
75%	187.000000	2019.000000	3.150000e+06	4.000000	4.000000	4410
max	206.000000	2021.000000	1.800000e+08	122.000000	62.000000	10122

```
pd.DataFrame(data['all']['sub_position'].unique())
```

	0
0	Centre-Forward
1	Right-Back
2	Central Midfield
3	Defensive Midfield
4	Goalkeeper
5	Centre-Back
6	Attacking Midfield
7	Right Winger
8	Left Winger
9	Left-Back
10	Left Midfield
11	Midfield
12	Second Striker
13	Right Midfield
14	Attack
15	Defender

```
data['all'][data['all']['name'] == 'Cristiano Ronaldo']
```

	name	nationality	position	sub_position	height	season	market_value	gc
player_id								
8198	Cristiano Ronaldo	Portugal	Attack	Centre-Forward	187	2014	96000000.0	
8198	Cristiano Ronaldo	Portugal	Attack	Centre-Forward	187	2015	105000000.0	
8198	Cristiano Ronaldo	Portugal	Attack	Centre-Forward	187	2016	99000000.0	
8198	Cristiano Ronaldo	Portugal	Attack	Centre-Forward	187	2017	90000000.0	
8198	Cristiano Ronaldo	Portugal	Attack	Centre-Forward	187	2018	96000000.0	
8198	Cristiano Ronaldo	Portugal	Attack	Centre-Forward	187	2019	74250000.0	
8198	Cristiano Ronaldo	Portugal	Attack	Centre-Forward	187	2020	54000000.0	
8198	Cristiano Ronaldo	Portugal	Attack	Centre-Forward	187	2021	39000000.0	

One hot encoding

```
# one hot encode categorical features
data['all_onehot'] = pd.get_dummies(data['all'], columns = [
    'position',
    'sub_position',
    'nationality',
    'club_name'
])

data['all_onehot']
```

player_id	name	height	season	market_value	goals	assists	minutes_played	yell
9800	Artem Milevskyi	189	2020	90000.0	0	0	720	
43084	Gaetano Berardi	179	2020	360000.0	0	0	228	
230826	Gennaro Acampora	174	2020	360000.0	2	4	1248	
198087	Matteo Ricci	176	2020	1530000.0	0	6	4880	
110689	Deniz Mehmet	192	2020	68000.0	0	0	1080	
...	...	...	...	...	...	...	...	
364245	Jordan Teze	183	2019	420000.0	0	0	360	
364245	Jordan Teze	183	2020	1102500.0	0	2	7494	
364245	Jordan Teze	183	2021	5400000.0	2	8	5260	
575367	Richard Ledezma	174	2020	658250.0	0	2	234	
575367	Richard Ledezma	174	2021	765000.0	2	0	88	

50781 rows × 588 columns

```
data['all_onehot'].dtypes
```

```
name                                object
height                               int64
season                               int64
market_value                       float64
goals                                int64
                                  ...
club_name_Yeni Malatyaspor           uint8
club_name_Zenit St Petersburg        uint8
club_name_Zirka Kropyvnytskyi        uint8
club_name_Zorya Lugansk              uint8
club_name_Zska Moskau                uint8
Length: 588, dtype: object
```

```
data['all_onehot'].describe()
```

	height	season	market_value	goals	assists	minutes
count	50781.000000	50781.000000	5.078100e+04	50781.000000	50781.000000	50781
mean	180.794628	2017.380063	3.630890e+06	3.880546	2.949883	2805
std	17.703409	2.318805	8.274637e+06	7.352176	4.793814	2103
min	0.000000	2013.000000	9.000000e+03	0.000000	0.000000	2
25%	178.000000	2015.000000	3.600000e+05	0.000000	0.000000	884
50%	182.000000	2017.000000	9.000000e+05	0.000000	2.000000	2566
75%	187.000000	2019.000000	3.150000e+06	4.000000	4.000000	4410
max	206.000000	2021.000000	1.800000e+08	122.000000	62.000000	10122

8 rows × 587 columns

```
data['all_onehot'][data['all_onehot']['name'] == 'Lionel Messi']
```

	name	height	season	market_value	goals	assists	minutes_played	yellow_c
player_id								
28003	Lionel Messi	169	2014	108000000.0	116	62	10122	
28003	Lionel Messi	169	2015	108000000.0	82	48	8458	
28003	Lionel Messi	169	2016	108000000.0	108	40	8904	
28003	Lionel Messi	169	2017	108000000.0	90	40	8936	
28003	Lionel Messi	169	2018	156000000.0	102	44	8048	
28003	Lionel Messi	169	2019	130500000.0	60	50	7262	
28003	Lionel Messi	169	2020	95400000.0	78	30	8746	
28003	Lionel Messi	169	2021	66000000.0	22	26	5384	

8 rows × 588 columns

Exploratory Data Analysis

```
data['all_eda'] = data['all'].copy()

data['all_eda']['log_market_value'] = np.log(data['all_eda']['market_valu
```

```
df_highest_market_value_players = data['all_eda'].nlargest(n = 1, columns

df_highest_market_value_players
```

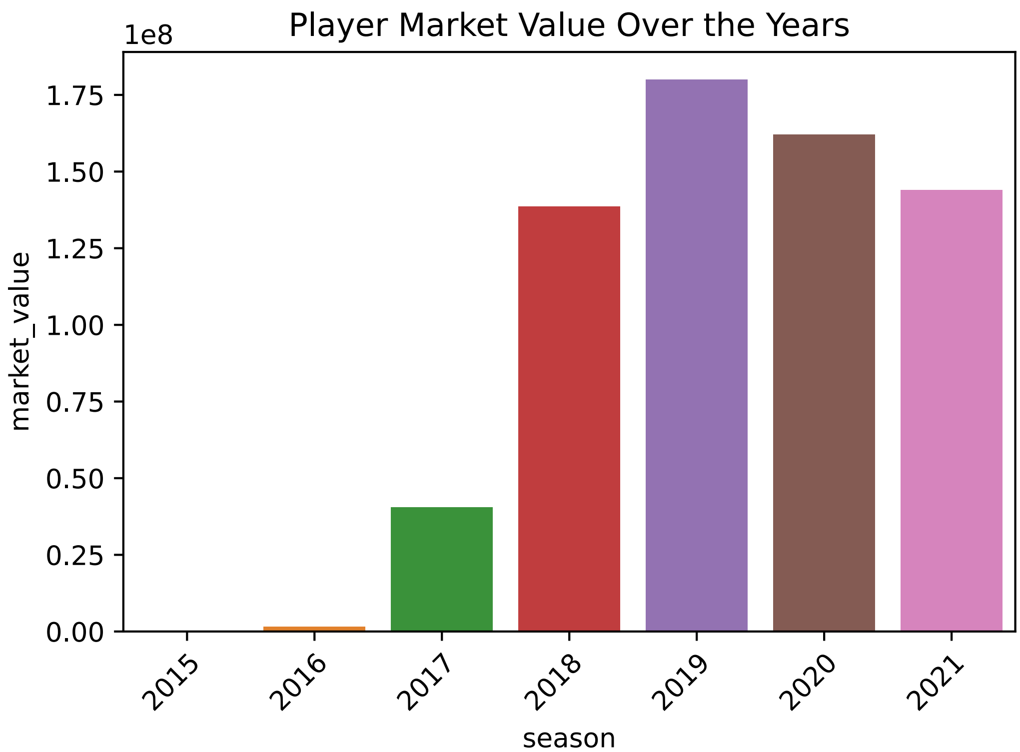
	name	nationality	position	sub_position	height	season	market_value	go
player_id								
342229	Kylian Mbappe	France	Attack	Centre-Forward	178	2019	180000000.0	

```
df_highest_market_value = data['all_eda'].loc[data['all_eda']['name'].isi

df_highest_market_value
```

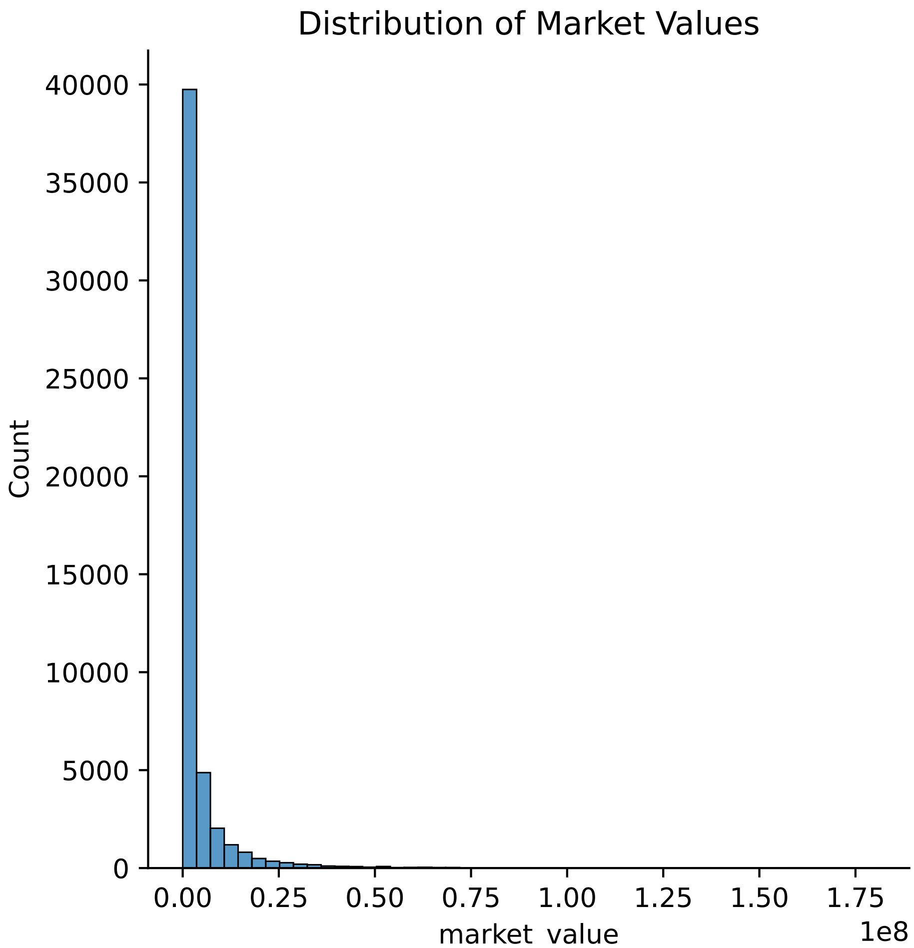
	name	nationality	position	sub_position	height	season	market_value	go
player_id								
342229	Kylian Mbappe	France	Attack	Centre-Forward	178	2015	45000.0	
342229	Kylian Mbappe	France	Attack	Centre-Forward	178	2016	1518750.0	
342229	Kylian Mbappe	France	Attack	Centre-Forward	178	2017	40500000.0	
342229	Kylian Mbappe	France	Attack	Centre-Forward	178	2018	138600000.0	
342229	Kylian Mbappe	France	Attack	Centre-Forward	178	2019	180000000.0	
342229	Kylian Mbappe	France	Attack	Centre-Forward	178	2020	162000000.0	
342229	Kylian Mbappe	France	Attack	Centre-Forward	178	2021	144000000.0	

```
_ = sns.barplot(
    data = df_highest_market_value,
    x = 'season', y = 'market_value'
).set(title = 'Player Market Value Over the Years')
plt.xticks(rotation = 45, ha = 'right', rotation_mode = 'anchor')
```

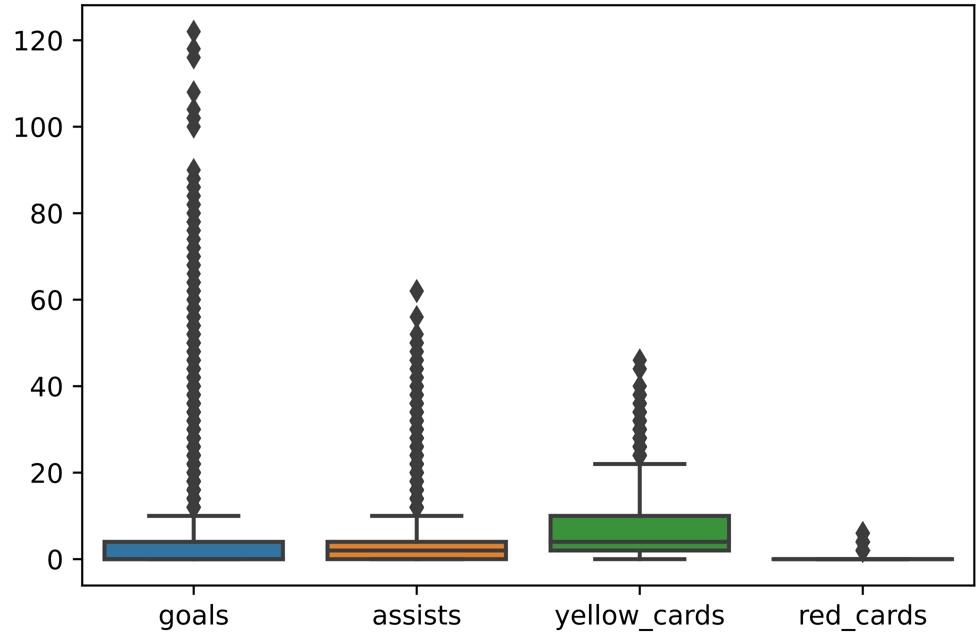
```
plt.show()
```



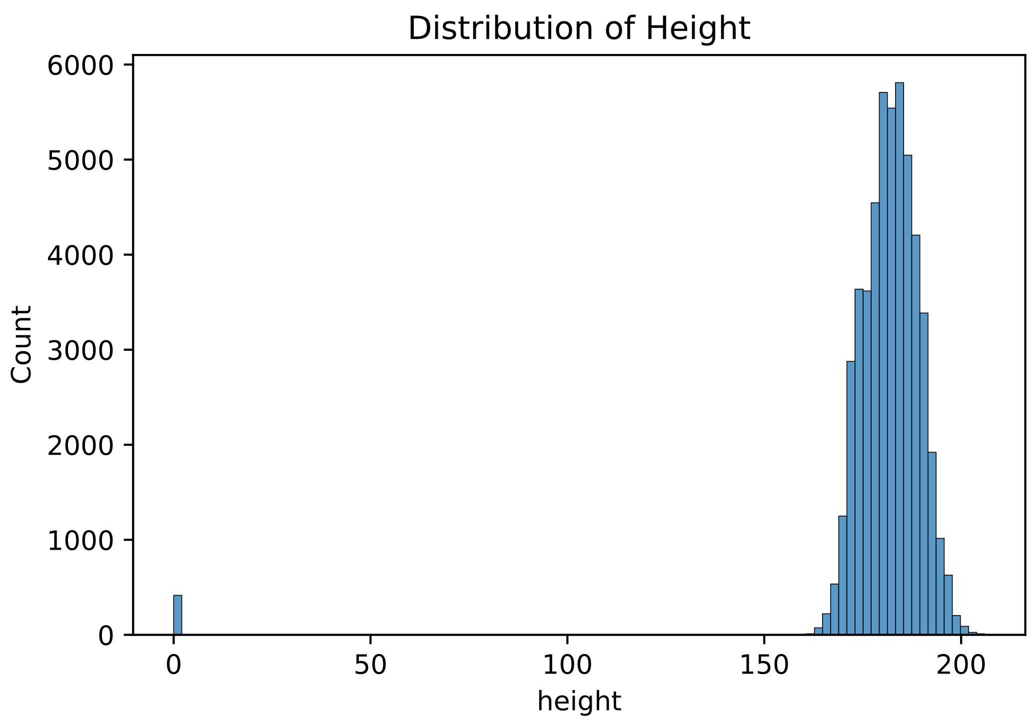

```
_ = sns.displot(
    data = data['all_eda'].reset_index(),
    x = 'market_value',
    bins = 50
).set(title = 'Distribution of Market Values')

plt.show()
```




```
_ = sns.boxplot(data = data['all_eda'][[
    'goals',
    'assists',
    'yellow_cards',
    'red_cards'
]]).set(title = '')
```




```
_ = sns.histplot(
    data = data['all_eda'].reset_index(), x = 'height',
    bins = 100
).set(title = 'Distribution of Height')
```



Proposed Solution

Packages:

- `sklearn` (scikit-learn)
- `StatsModel`
- `Seaborn`
- Ordinary Least square regression: `statsmodels.api`

We believe that we can take the characteristics football clubs may regard to be the most important in the dataset and use those features to evaluate players. Those can be our core information in order to use regression analysis. If able to determine a certain cluster for the data set depending on the position and attributes. We can compare the players in the data set with the new data. We will be using OLS (Ordinary Least Squares).

We have to start with the most important step which is data cleaning and EDA analysis in order to get more accurate results, Matrix transformation with the numpy library. In order to increase covariance with variables, we can use dimension reduction techniques. Next we can normalize the data sets in order to not get skewed by one particular feature. Dealing with values missing or if we need to have numeric values for non-numeric data (i.e popularity, health, position).

Our comparison of errors will be coming from the Transfermrket.com website as it is updated everyday to evaluate different players. We can get the player’s information to get a percent error or a total error for our evaluations. If we can, we will use the RMSE (Root mean square error) or Mean Absolute Values of our model prediction.

Evaluation Metrics

We will be using an OLS regression model and the evaluation techniques we are considering are RMSE and Euclidean distance. A possible evaluation metric we will use is RMSE or Mean Absolute Value of Errors. It is derived by calculating the difference between the estimated and actual value, square those results, then calculate the mean of those results. The formula for RMSE is

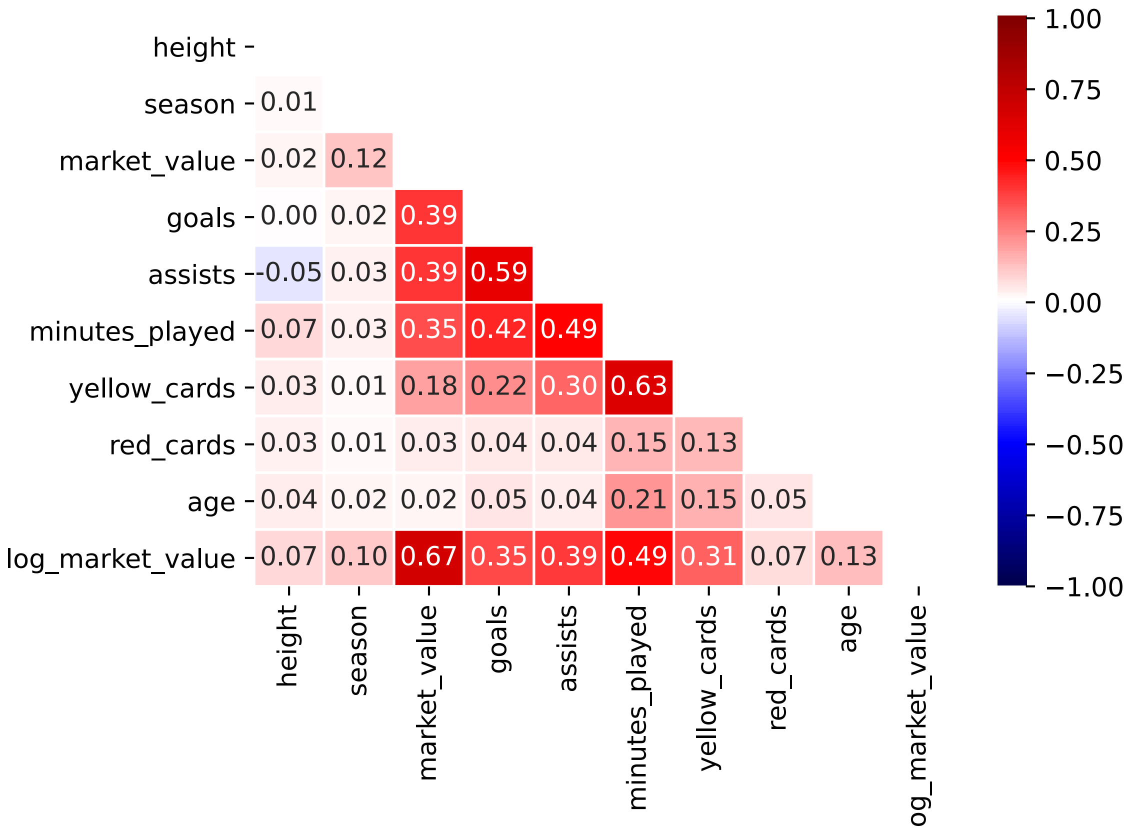
$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (\mathrm{Predicted}_i - \mathrm{Actual}_i)^2}{N}}$$

Results

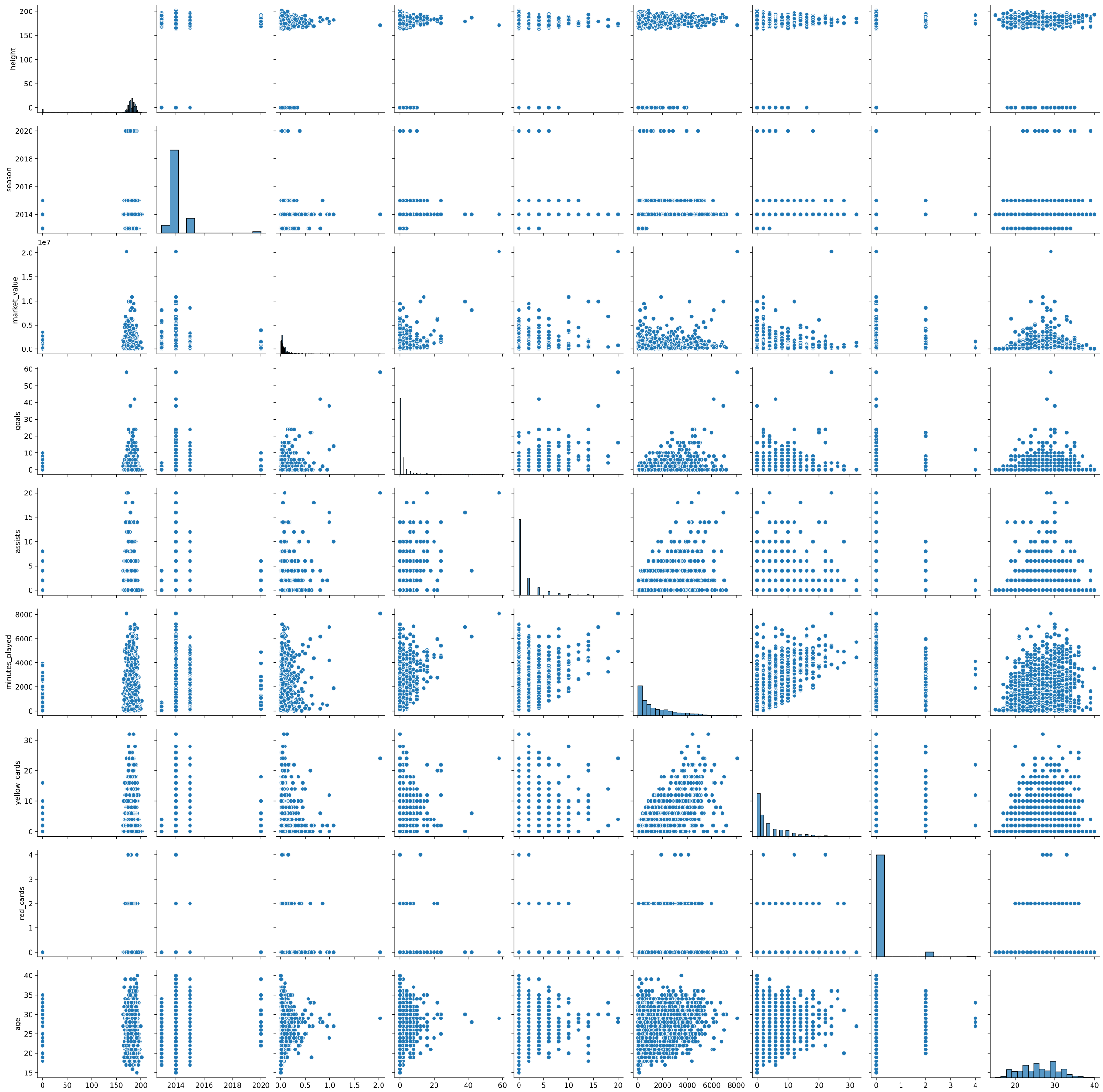
Subsection 1

We wanted to start by analyzing which data variables are important and correlate with each other using a heat map. The heat map will allow us to determine which are important to keep and also the pair plot will show the correlation between the variables.

```
corr = data['all_eda'].corr()
_ = sns.heatmap(corr,
    cmap = 'seismic',
    linewidth = 1, linecolor = 'white',
    vmax = 1, vmin = -1,
    mask = np.triu(np.ones_like(corr, dtype = bool)),
    annot = True,
    fmt = '0.2f'
)
```




```
_ = sns.pairplot(data['all'][:1500].reset_index())
```



Subsection 2

After analyzing which variables we wanted to keep and how they relate to each other. Our first intial model would be a simple OLS which will only use numerical data. We want to see the difference in our model if its classifying only numerical data. Since we believe the categorical data present can be defined as subjective in the Soccer world.

```
y, X = patsy.dmatrices('''
    market_value ~
        age + goals + assists + minutes_played
            + yellow_cards + red_cards
            + height + age
''', data=data['all_eda'], return_type="dataframe")
```

```
model = sm.OLS(y, X)
fit = model.fit()
pred = fit.predict(X)
fit.summary()
```

OLS Regression Results

Dep. Variable:	market_value	**R-squared:**	0.220
Model:	OLS	**Adj. R-squared:**	0.219
Method:	Least Squares	**F-statistic:**	2041.
Date:	Fri, 10 Jun 2022	**Prob (F-statistic):**	0.00
Time:	03:39:42	**Log-Likelihood:**	-8.7464e+05
No. Observations:	50781	**AIC:**	1.749e+06
Df Residuals:	50773	**BIC:**	1.749e+06
Df Model:	7		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	1.938e+05	3.77e+05	0.514	0.607	-5.45e+05	9.33e+05
age	-7.077e+04	7609.677	-9.300	0.000	-8.57e+04	-5.59e+04
goals	2.389e+05	5556.790	42.994	0.000	2.28e+05	2.5e+05
assists	3.135e+05	8922.514	35.136	0.000	2.96e+05	3.31e+05
minutes_played	826.8657	22.786	36.288	0.000	782.205	871.526
yellow_cards	-6.638e+04	6910.754	-9.606	0.000	-7.99e+04	-5.28e+04
red_cards	-7.781e+04	6.05e+04	-1.287	0.198	-1.96e+05	4.07e+04
height	7851.4099	1846.711	4.252	0.000	4231.837	1.15e+04

Omnibus:	53530.739	**Durbin-Watson:**	0.636
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	5620986.518
Skew:	5.193	**Prob(JB):**	0.00
Kurtosis:	53.485	**Cond. No.**	4.08e+04

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 4.08e+04. This might indicate that there are strong multicollinearity or other numerical problems.

```
print((y))
print((pred))
```

```
np.sqrt(mean_squared_error((y),(pred)))
```

```
           market_value
player_id
9800            90000.0
43084          360000.0
230826         360000.0
198087        1530000.0
110689          68000.0
...                 ...
364245         420000.0
364245        1102500.0
364245        5400000.0
575367         658250.0
575367         765000.0

[50781 rows x 1 columns]
player_id
9800     -5.312852e+05
43084    -4.060252e+05
230826    2.289005e+06
198087    5.058775e+06
110689    6.836034e+05
              ...
364245    5.837390e+05
364245    6.375011e+06
364245    6.683061e+06
575367    9.031230e+05
575367    6.952160e+05
Length: 50781, dtype: float64
7309873.9471996445
```

Taking the Log of the market value was a way to incorportate a different way of showing an error metric. The log would keep the monotonocity of each variable

```
data['all_eda']['log_market_value'] = np.log(data['all_eda']['market_valu
y, X = patsy.dmatrices('''
    log_market_value ~
        age + goals + assists + minutes_played
            + yellow_cards + red_cards
            + height + age
''', data=data['all_eda'], return_type="dataframe")
model = sm.OLS(y, X)
fit = model.fit()
pred2 = fit.predict(X)
fit.summary()
```

OLS Regression Results

Dep. Variable:	log_market_value	**R-squared:**	0.282
Model:	OLS	**Adj. R-squared:**	0.282
Method:	Least Squares	**F-statistic:**	2851.
Date:	Fri, 10 Jun 2022	**Prob (F-statistic):**	0.00
Time:	03:39:42	**Log-Likelihood:**	-84685.
No. Observations:	50781	**AIC:**	1.694e+05
Df Residuals:	50773	**BIC:**	1.695e+05
Df Model:	7		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	11.7302	0.066	177.262	0.000	11.600	11.860
age	0.0156	0.001	11.664	0.000	0.013	0.018
goals	0.0231	0.001	23.713	0.000	0.021	0.025
assists	0.0473	0.002	30.198	0.000	0.044	0.050
minutes_played	0.0003	4e-06	63.609	0.000	0.000	0.000
yellow_cards	0.0007	0.001	0.560	0.576	-0.002	0.003
red_cards	0.0093	0.011	0.880	0.379	-0.011	0.030
height	0.0047	0.000	14.436	0.000	0.004	0.005

Omnibus:	836.100	**Durbin-Watson:**	0.853
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	794.667
Skew:	0.271	**Prob(JB):**	2.76e-173
Kurtosis:	2.716	**Cond. No.**	4.08e+04

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 4.08e+04. This might indicate that there are strong multicollinearity or other numerical problems.

```
print(np.exp(y))
print(np.exp(pred2))
```

```
np.sqrt(mean_squared_error((y),(pred2)))
```

```
           log_market_value
player_id
9800                90000.0
43084              360000.0
230826             360000.0
198087            1530000.0
110689              68000.0
...                     ...
364245             420000.0
364245            1102500.0
364245            5400000.0
575367             658250.0
575367             765000.0

[50781 rows x 1 columns]
player_id
9800      6.157680e+05
43084     4.928576e+05
230826    7.211031e+05
198087    1.931418e+06
110689    6.111676e+05
              ...
364245    4.307993e+05
364245    2.970482e+06
364245    2.380843e+06
575367    4.402778e+05
575367    4.100153e+05
Length: 50781, dtype: float64
1.282369368429589
```

Subsection 3

Since our dataset contains categorical values, we want to perform again OLS but by including those values so we can compare the accuracy between categorical and non-categorical OLS.

```
y, X = patsy.dmatrices('''
    log_market_value ~
        age + goals + assists + minutes_played
            + yellow_cards + red_cards
            + height + age
            + C(nationality) + C(position) + C(sub_position) + C(club_nam
''', data=data['all_eda'], return_type="dataframe")
model = sm.OLS(y, X)
fit = model.fit()
pred3 = fit.predict(X)
fit.summary()
```

OLS Regression Results

Dep. Variable:	log_market_value	**R-squared:**	0.678
Model:	OLS	**Adj. R-squared:**	0.674
Method:	Least Squares	**F-statistic:**	185.8
Date:	Fri, 10 Jun 2022	**Prob (F-statistic):**	0.00
Time:	03:41:05	**Log-Likelihood:**	-64329.
No. Observations:	50781	**AIC:**	1.298e+05
Df Residuals:	50211	**BIC:**	1.348e+05
Df Model:	569		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	11.2022	0.559	20.054	0.000	10.107	12.297
C(nationality)[T.Albania]	0.0602	0.622	0.097	0.923	-1.159	1.280
C(nationality)[T.Algeria]	0.2460	0.621	0.396	0.692	-0.972	1.463
C(nationality)[T.Angola]	0.3028	0.625	0.485	0.628	-0.922	1.527
C(nationality)[T.Antigua and Barbuda]	0.2055	0.714	0.288	0.774	-1.194	1.606
C(nationality)[T.Argentina]	0.2909	0.620	0.470	0.639	-0.924	1.505
C(nationality)[T.Armenia]	-0.0123	0.626	-0.020	0.984	-1.238	1.214
C(nationality)[T.Aruba]	-0.0934	0.652	-0.143	0.886	-1.372	1.185
C(nationality)[T.Australia]	0.1484	0.622	0.239	0.811	-1.071	1.368
C(nationality)[T.Austria]	0.2672	0.621	0.430	0.667	-0.950	1.484
C(nationality)[T.Azerbaijan]	-0.1249	0.643	-0.194	0.846	-1.386	1.136
C(nationality)[T.Bahrain]	1.212e-12	1.44e-12	0.842	0.400	-1.61e-12	4.03e-12
C(nationality)[T.Barbados]	2.792e-13	9.28e-13	0.301	0.764	-1.54e-12	2.1e-12
C(nationality)[T.Belarus]	0.0540	0.625	0.086	0.931	-1.170	1.278
C(nationality)[T.Belgium]	0.1501	0.619	0.242	0.808	-1.064	1.364
C(nationality)[T.Benin]	-0.0736	0.633	-0.116	0.907	-1.315	1.168
C(nationality)[T.Bermuda]	0.8495	1.065	0.798	0.425	-1.238	2.937
C(nationality)[T.Bolivia]	0.2664	0.665	0.401	0.689	-1.037	1.570

C(nationality)[T.Bosnia-Herzegovina]	0.1182	0.621	0.190	0.849	-1.099	1.336
C(nationality)[T.Brazil]	0.2350	0.619	0.379	0.704	-0.979	1.449
C(nationality)[T.Bulgaria]	0.1903	0.626	0.304	0.761	-1.036	1.416
C(nationality)[T.Burkina Faso]	0.2280	0.625	0.365	0.715	-0.998	1.454
C(nationality)[T.Burundi]	0.3019	0.656	0.460	0.645	-0.984	1.588
C(nationality)[T.Cameroon]	0.2571	0.621	0.414	0.679	-0.959	1.474
C(nationality)[T.Canada]	-0.0476	0.626	-0.076	0.939	-1.275	1.180
C(nationality)[T.Cape Verde]	0.1084	0.623	0.174	0.862	-1.113	1.330
C(nationality)[T.Central African Republic]	0.4384	0.647	0.677	0.498	-0.831	1.707
C(nationality)[T.Chad]	0.1484	0.665	0.223	0.823	-1.155	1.451
C(nationality)[T.Chile]	0.2310	0.623	0.371	0.711	-0.991	1.453
C(nationality)[T.China]	-0.4184	0.656	-0.638	0.524	-1.704	0.867
C(nationality)[T.Chinese Taipei (Taiwan)]	-0.7057	0.761	-0.927	0.354	-2.198	0.786
C(nationality)[T.Colombia]	0.4487	0.621	0.723	0.470	-0.768	1.666
C(nationality)[T.Comoros]	0.3576	0.641	0.558	0.577	-0.898	1.613
C(nationality)[T.Congo]	0.1482	0.626	0.237	0.813	-1.080	1.376
C(nationality)[T.Costa Rica]	0.1354	0.627	0.216	0.829	-1.093	1.363
C(nationality)[T.Cote d'Ivoire]	0.3575	0.620	0.576	0.564	-0.858	1.573
C(nationality)[T.Croatia]	0.3805	0.620	0.614	0.539	-0.835	1.596
C(nationality)[T.Cuba]	-0.0300	1.067	-0.028	0.978	-2.121	2.061
C(nationality)[T.Curacao]	-0.0399	0.626	-0.064	0.949	-1.267	1.187
C(nationality)[T.Cyprus]	0.2628	0.636	0.413	0.680	-0.984	1.510
C(nationality)[T.Czech Republic]	0.2576	0.622	0.414	0.679	-0.961	1.476
C(nationality)[T.DR Congo]	0.3258	0.621	0.525	0.600	-0.892	1.543
C(nationality)[T.Denmark]	0.0836	0.620	0.135	0.893	-1.131	1.298
C(nationality)[T.Dominican Republic]	-0.0826	0.665	-0.124	0.901	-1.385	1.220
C(nationality)[T.Ecuador]	0.3800	0.627	0.606	0.545	-0.849	1.609
C(nationality)[T.Egypt]	0.1848	0.627	0.295	0.768	-1.044	1.413
C(nationality)[T.El Salvador]	-0.4594	0.714	-0.643	0.520	-1.859	0.940
C(nationality)[T.England]	-0.0313	0.619	-0.051	0.960	-1.246	1.183

C(nationality)[T.Equatorial Guinea]	0.2308	0.650	0.355	0.723	-1.044	1.506
C(nationality)[T.Eritrea]	0.6033	0.871	0.693	0.489	-1.104	2.311
C(nationality)[T.Estonia]	-0.0422	0.648	-0.065	0.948	-1.311	1.227
C(nationality)[T.Ethiopia]	-0.9194	1.065	-0.863	0.388	-3.007	1.168
C(nationality)[T.Faroe Islands]	-0.2503	0.646	-0.388	0.698	-1.516	1.015
C(nationality)[T.Finland]	-0.0104	0.623	-0.017	0.987	-1.232	1.212
C(nationality)[T.France]	0.2056	0.619	0.332	0.740	-1.008	1.419
C(nationality)[T.French Guiana]	-0.1264	0.654	-0.193	0.847	-1.409	1.156
C(nationality)[T.Gabon]	0.3401	0.627	0.542	0.588	-0.890	1.570
C(nationality)[T.Georgia]	0.3154	0.623	0.506	0.613	-0.905	1.536
C(nationality)[T.Germany]	0.0235	0.619	0.038	0.970	-1.190	1.237
C(nationality)[T.Ghana]	0.2310	0.620	0.372	0.710	-0.985	1.447
C(nationality)[T.Greece]	0.0206	0.620	0.033	0.973	-1.194	1.235
C(nationality)[T.Grenada]	-0.4473	0.731	-0.612	0.541	-1.881	0.986
C(nationality)[T.Guadeloupe]	0.3269	0.632	0.517	0.605	-0.912	1.566
C(nationality)[T.Guinea]	0.2698	0.623	0.433	0.665	-0.951	1.490
C(nationality)[T.Guinea-Bissau]	0.2250	0.624	0.360	0.719	-0.999	1.449
C(nationality)[T.Guyana]	0.3923	0.713	0.550	0.582	-1.006	1.790
C(nationality)[T.Haiti]	0.0332	0.635	0.052	0.958	-1.211	1.278
C(nationality)[T.Honduras]	0.2133	0.639	0.334	0.739	-1.040	1.466
C(nationality)[T.Hungary]	0.1680	0.626	0.269	0.788	-1.058	1.394
C(nationality)[T.Iceland]	0.2479	0.622	0.398	0.690	-0.972	1.468
C(nationality)[T.India]	9.933e-17	1.13e-13	0.001	0.999	-2.22e-13	2.22e-13
C(nationality)[T.Indonesia]	0.9166	1.066	0.860	0.390	-1.172	3.005
C(nationality)[T.Iran]	0.4381	0.626	0.700	0.484	-0.788	1.664
C(nationality)[T.Iraq]	0.3928	0.642	0.612	0.541	-0.866	1.651
C(nationality)[T.Ireland]	-0.2612	0.621	-0.420	0.674	-1.479	0.957
C(nationality)[T.Israel]	0.0568	0.625	0.091	0.928	-1.169	1.282
C(nationality)[T.Italy]	-0.0981	0.620	-0.158	0.874	-1.312	1.116
C(nationality)[T.Jamaica]	0.0501	0.628	0.080	0.936	-1.181	1.281
C(nationality)[T.Japan]	0.3757	0.622	0.604	0.546	-0.843	1.594

C(nationality)[T.Jordan]	0.6521	0.758	0.860	0.390	-0.833	2.137
C(nationality)[T.Kazakhstan]	0.3630	0.644	0.564	0.573	-0.898	1.624
C(nationality)[T.Kenya]	0.1721	0.651	0.264	0.791	-1.103	1.447
C(nationality)[T.Korea, North]	-0.4479	1.065	-0.421	0.674	-2.535	1.639
C(nationality)[T.Korea, South]	0.3091	0.627	0.493	0.622	-0.920	1.538
C(nationality)[T.Kosovo]	0.2812	0.625	0.450	0.653	-0.943	1.506
C(nationality)[T.Kyrgyzstan]	0.4378	0.879	0.498	0.619	-1.286	2.161
C(nationality)[T.Laos]	0.3448	0.798	0.432	0.666	-1.219	1.908
C(nationality)[T.Latvia]	0.2093	0.652	0.321	0.748	-1.069	1.488
C(nationality)[T.Lebanon]	-0.4018	0.871	-0.461	0.645	-2.110	1.306
C(nationality)[T.Liberia]	-0.0332	0.701	-0.047	0.962	-1.407	1.341
C(nationality)[T.Libya]	0.0042	0.678	0.006	0.995	-1.324	1.333
C(nationality)[T.Liechtenstein]	-0.0960	0.797	-0.120	0.904	-1.658	1.466
C(nationality)[T.Lithuania]	-0.0942	0.636	-0.148	0.882	-1.341	1.153
C(nationality)[T.Luxembourg]	-0.0390	0.633	-0.062	0.951	-1.279	1.201
C(nationality)[T.Madagascar]	-0.2638	0.641	-0.412	0.681	-1.519	0.992
C(nationality)[T.Malawi]	0.2219	1.067	0.208	0.835	-1.869	2.313
C(nationality)[T.Malaysia]	-0.1996	0.684	-0.292	0.770	-1.540	1.141
C(nationality)[T.Mali]	0.2290	0.621	0.369	0.712	-0.989	1.447
C(nationality)[T.Malta]	0.0768	0.692	0.111	0.912	-1.280	1.433
C(nationality)[T.Martinique]	0.1781	0.629	0.283	0.777	-1.055	1.411
C(nationality)[T.Mauritania]	-0.2871	0.639	-0.449	0.653	-1.540	0.966
C(nationality)[T.Mauritius]	0.5728	0.714	0.802	0.422	-0.827	1.972
C(nationality)[T.Mexico]	0.6341	0.624	1.016	0.310	-0.589	1.857
C(nationality)[T.Moldova]	0.0073	0.631	0.012	0.991	-1.230	1.245
C(nationality)[T.Monaco]	7.514e-15	1.36e-14	0.553	0.581	-1.91e-14	3.42e-14
C(nationality)[T.Montenegro]	0.1850	0.624	0.296	0.767	-1.038	1.408
C(nationality)[T.Montserrat]	-0.3983	0.756	-0.527	0.598	-1.880	1.083
C(nationality)[T.Morocco]	0.1512	0.620	0.244	0.807	-1.064	1.367
C(nationality)[T.Mozambique]	0.0046	0.636	0.007	0.994	-1.242	1.251
C(nationality)[T.Netherlands]	0.0999	0.619	0.161	0.872	-1.113	1.313

C(nationality)[T.Neukaledonien]	0.2978	0.684	0.435	0.663	-1.043	1.639
C(nationality)[T.New Zealand]	0.2391	0.636	0.376	0.707	-1.008	1.486
C(nationality)[T.Nicaragua]	0.1843	1.065	0.173	0.863	-1.903	2.272
C(nationality)[T.Niger]	0.1077	0.684	0.158	0.875	-1.232	1.448
C(nationality)[T.Nigeria]	0.1656	0.620	0.267	0.789	-1.050	1.381
C(nationality)[T.North Macedonia]	0.0101	0.626	0.016	0.987	-1.218	1.238
C(nationality)[T.Northern Ireland]	-0.2763	0.624	-0.443	0.658	-1.499	0.946
C(nationality)[T.Norway]	0.2931	0.621	0.472	0.637	-0.924	1.510
C(nationality)[T.Pakistan]	-0.3998	0.731	-0.547	0.585	-1.833	1.034
C(nationality)[T.Palästina]	-0.3134	0.872	-0.360	0.719	-2.022	1.395
C(nationality)[T.Panama]	-0.1520	0.668	-0.227	0.820	-1.461	1.157
C(nationality)[T.Papua New Guinea]	-2.589e-15	2.84e-15	-0.912	0.362	-8.15e-15	2.98e-15
C(nationality)[T.Paraguay]	0.4580	0.626	0.732	0.464	-0.768	1.684
C(nationality)[T.Peru]	0.3182	0.628	0.507	0.612	-0.912	1.549
C(nationality)[T.Philippines]	0.2035	0.654	0.311	0.756	-1.079	1.486
C(nationality)[T.Poland]	0.2277	0.621	0.367	0.714	-0.989	1.444
C(nationality)[T.Portugal]	0.2841	0.619	0.459	0.646	-0.930	1.498
C(nationality)[T.Qatar]	0.0395	0.733	0.054	0.957	-1.396	1.475
C(nationality)[T.Romania]	0.4041	0.622	0.650	0.516	-0.815	1.623
C(nationality)[T.Russia]	-0.1947	0.620	-0.314	0.753	-1.409	1.020
C(nationality)[T.Rwanda]	-0.2077	0.665	-0.312	0.755	-1.511	1.095
C(nationality)[T.Saint-Martin]	-0.4565	0.875	-0.522	0.602	-2.171	1.258
C(nationality)[T.San Marino]	1.095e-15	2.52e-15	0.435	0.664	-3.84e-15	6.03e-15
C(nationality)[T.Sao Tome and Principe]	0.4593	0.701	0.655	0.512	-0.915	1.833
C(nationality)[T.Saudi Arabia]	-0.0294	0.730	-0.040	0.968	-1.461	1.402
C(nationality)[T.Scotland]	-0.0778	0.620	-0.126	0.900	-1.293	1.137
C(nationality)[T.Senegal]	0.3065	0.620	0.494	0.621	-0.909	1.522
C(nationality)[T.Serbia]	0.2868	0.620	0.463	0.644	-0.928	1.502
C(nationality)[T.Sierra Leone]	0.0421	0.636	0.066	0.947	-1.205	1.289
C(nationality)[T.Slovakia]	0.1242	0.622	0.200	0.842	-1.095	1.344

C(nationality)[T.Slovenia]	0.1211	0.622	0.195	0.845	-1.097	1.339
C(nationality)[T.Somalia]	4.92e-16	2.8e-15	0.175	0.861	-5.01e-15	5.99e-15
C(nationality)[T.South Africa]	0.3139	0.626	0.501	0.616	-0.913	1.541
C(nationality)[T.Spain]	0.0023	0.619	0.004	0.997	-1.212	1.216
C(nationality)[T.St. Kitts & Nevis]	-0.0150	1.066	-0.014	0.989	-2.104	2.074
C(nationality)[T.St. Lucia]	1.022e-15	1.7e-15	0.602	0.547	-2.31e-15	4.35e-15
C(nationality)[T.Suriname]	0.0841	0.624	0.135	0.893	-1.139	1.307
C(nationality)[T.Sweden]	0.2727	0.620	0.440	0.660	-0.942	1.488
C(nationality)[T.Switzerland]	0.2839	0.621	0.457	0.647	-0.933	1.501
C(nationality)[T.Syria]	0.1419	0.672	0.211	0.833	-1.176	1.459
C(nationality)[T.Tajikistan]	-0.2933	0.797	-0.368	0.713	-1.855	1.268
C(nationality)[T.Tanzania]	-0.0136	0.692	-0.020	0.984	-1.370	1.342
C(nationality)[T.Thailand]	0.2942	1.078	0.273	0.785	-1.819	2.407
C(nationality)[T.The Gambia]	0.1888	0.629	0.300	0.764	-1.043	1.421
C(nationality)[T.Togo]	0.0932	0.626	0.149	0.882	-1.133	1.320
C(nationality)[T.Trinidad and Tobago]	0.3490	0.647	0.539	0.590	-0.920	1.618
C(nationality)[T.Tunisia]	0.1465	0.623	0.235	0.814	-1.075	1.368
C(nationality)[T.Turkey]	-0.2210	0.619	-0.357	0.721	-1.435	0.993
C(nationality)[T.Turkmenistan]	1.34e-15	2.42e-15	0.553	0.580	-3.41e-15	6.09e-15
C(nationality)[T.Uganda]	-0.0914	0.641	-0.143	0.887	-1.347	1.164
C(nationality)[T.Ukraine]	0.0723	0.620	0.117	0.907	-1.143	1.288
C(nationality)[T.United States]	0.1875	0.622	0.302	0.763	-1.031	1.406
C(nationality)[T.Uruguay]	0.3865	0.621	0.622	0.534	-0.830	1.603
C(nationality)[T.Uzbekistan]	0.3134	0.635	0.494	0.622	-0.931	1.558
C(nationality)[T.Venezuela]	0.2504	0.624	0.402	0.688	-0.972	1.473
C(nationality)[T.Vietnam]	-0.5549	0.871	-0.637	0.524	-2.261	1.152
C(nationality)[T.Wales]	0.1269	0.623	0.204	0.839	-1.094	1.348
C(nationality)[T.Zambia]	0.1927	0.638	0.302	0.763	-1.058	1.443
C(nationality)[T.Zimbabwe]	-0.0710	0.634	-0.112	0.911	-1.314	1.172
C(position)[T.Defender]	0.6875	0.235	2.920	0.003	0.226	1.149

C(position)[T.Goalkeeper]	0.5183	0.039	13.383	0.000	0.442	0.594
C(position)[T.Midfield]	1.1611	0.071	16.262	0.000	1.021	1.301
C(sub_position)[T.Centre-Back]	0.6228	0.252	2.476	0.013	0.130	1.116
C(sub_position)[T.Left-Back]	0.6176	0.252	2.453	0.014	0.124	1.111
C(sub_position)[T.Right-Back]	0.5614	0.252	2.230	0.026	0.068	1.055
C(sub_position)[T.Goalkeeper]	0.5183	0.039	13.383	0.000	0.442	0.594
C(sub_position)[T.Attack]	1.0701	0.203	5.278	0.000	0.673	1.468
C(sub_position)[T.Centre-Forward]	1.5247	0.077	19.900	0.000	1.375	1.675
C(sub_position)[T.Left Winger]	1.5525	0.077	20.135	0.000	1.401	1.704
C(sub_position)[T.Right Winger]	1.5496	0.077	20.101	0.000	1.398	1.701
C(sub_position)[T.Second Striker]	1.5782	0.084	18.801	0.000	1.414	1.743
C(sub_position)[T.Midfield]	0.1824	0.177	1.032	0.302	-0.164	0.529
C(sub_position)[T.Attacking Midfield]	1.5602	0.077	20.233	0.000	1.409	1.711
C(sub_position)[T.Central Midfield]	0.3128	0.040	7.764	0.000	0.234	0.392
C(sub_position)[T.Defensive Midfield]	0.2723	0.041	6.663	0.000	0.192	0.352
C(sub_position)[T.Left Midfield]	0.1962	0.050	3.949	0.000	0.099	0.294
C(sub_position)[T.Right Midfield]	0.1974	0.050	3.932	0.000	0.099	0.296
C(club_name)[T.1 Fc Nurnberg]	-0.4285	0.161	-2.662	0.008	-0.744	-0.113
C(club_name)[T.1 Fc Union Berlin]	-0.2373	0.115	-2.067	0.039	-0.462	-0.012
C(club_name)[T.1 Fsv Mainz 05]	0.1497	0.090	1.666	0.096	-0.026	0.326
C(club_name)[T.Aalborg Bk]	-1.5202	0.094	-16.256	0.000	-1.704	-1.337
C(club_name)[T.Aarhus Gf]	-1.5066	0.094	-16.033	0.000	-1.691	-1.322
C(club_name)[T.Aberdeen Fc]	-1.5220	0.096	-15.839	0.000	-1.710	-1.334
C(club_name)[T.Ac Florenz]	0.6641	0.093	7.123	0.000	0.481	0.847
C(club_name)[T.Ac Horsens]	-1.7477	0.100	-17.428	0.000	-1.944	-1.551
C(club_name)[T.Ac Mailand]	1.1354	0.090	12.574	0.000	0.958	1.312
C(club_name)[T.Academica Coimbra]	-1.3883	0.135	-10.269	0.000	-1.653	-1.123
C(club_name)[T.Acn Siena 1904]	-1.3453	0.360	-3.735	0.000	-2.051	-0.639
C(club_name)[T.Adana Demirspor]	-0.7301	0.193	-3.779	0.000	-1.109	-0.351
C(club_name)[T.Adanaspor]	-1.1867	0.183	-6.467	0.000	-1.546	-0.827

C(club_name)[T.Ado Den Haag]	-1.3793	0.096	-14.413	0.000	-1.567	-1.192
C(club_name)[T.Ae Larisa]	-1.6752	0.098	-17.015	0.000	-1.868	-1.482
C(club_name)[T.Aek Athen]	-0.8059	0.093	-8.673	0.000	-0.988	-0.624
C(club_name)[T.Ael Kalloni]	-1.7709	0.131	-13.532	0.000	-2.027	-1.514
C(club_name)[T.Afc Bournemouth]	0.5318	0.103	5.166	0.000	0.330	0.734
C(club_name)[T.Afc Sunderland]	0.4600	0.125	3.688	0.000	0.216	0.704
C(club_name)[T.Ajax Amsterdam]	0.3685	0.092	4.012	0.000	0.188	0.548
C(club_name)[T.Akhisarspor]	-0.7119	0.103	-6.928	0.000	-0.913	-0.510
C(club_name)[T.Akhmat Grozny]	-0.5841	0.094	-6.199	0.000	-0.769	-0.399
C(club_name)[T.Alanyaspor]	-0.6991	0.098	-7.128	0.000	-0.891	-0.507
C(club_name)[T.Altay Sk]	-1.5116	0.194	-7.811	0.000	-1.891	-1.132
C(club_name)[T.Amiens Sc]	-0.5277	0.118	-4.475	0.000	-0.759	-0.297
C(club_name)[T.Amkar Perm]	-1.1928	0.114	-10.462	0.000	-1.416	-0.969
C(club_name)[T.Ankaraspor]	-0.4271	0.114	-3.748	0.000	-0.650	-0.204
C(club_name)[T.Antalyaspor]	-0.7002	0.093	-7.535	0.000	-0.882	-0.518
C(club_name)[T.Anzhi Makhachkala]	-0.8939	0.100	-8.913	0.000	-1.090	-0.697
C(club_name)[T.Ao Platanias]	-1.6862	0.109	-15.442	0.000	-1.900	-1.472
C(club_name)[T.Ao Xanthi]	-1.6323	0.098	-16.597	0.000	-1.825	-1.440
C(club_name)[T.Aok Kerkyra]	-1.8263	0.116	-15.749	0.000	-2.054	-1.599
C(club_name)[T.Apo Levadiakos]	-1.7009	0.100	-16.991	0.000	-1.897	-1.505
C(club_name)[T.Apollon Smyrnis]	-1.7257	0.111	-15.484	0.000	-1.944	-1.507
C(club_name)[T.Aris Thessaloniki]	-1.2442	0.108	-11.527	0.000	-1.456	-1.033
C(club_name)[T.Arminia Bielefeld]	-0.1347	0.144	-0.932	0.351	-0.418	0.148
C(club_name)[T.Arsenal Kiew]	-1.7168	0.138	-12.428	0.000	-1.988	-1.446
C(club_name)[T.Arsenal Tula]	-0.8831	0.094	-9.391	0.000	-1.067	-0.699
C(club_name)[T.As Livorno]	-0.5794	0.615	-0.942	0.346	-1.785	0.626
C(club_name)[T.As Monaco]	0.7824	0.089	8.792	0.000	0.608	0.957
C(club_name)[T.As Nancy Lorraine]	-1.0211	0.183	-5.565	0.000	-1.381	-0.661
C(club_name)[T.As Rom]	1.1245	0.092	12.264	0.000	0.945	1.304
C(club_name)[T.As Saint Etienne]	0.0619	0.092	0.674	0.500	-0.118	0.242
C(club_name)[T.Asteras Tripolis]	-1.4699	0.090	-16.323	0.000	-1.646	-1.293

C(club_name)[T.Aston Villa]	1.0315	0.106	9.764	0.000	0.824	1.239
C(club_name)[T.Atalanta Bergamo]	0.3369	0.092	3.673	0.000	0.157	0.517
C(club_name)[T.Athletic Bilbao]	0.5051	0.095	5.317	0.000	0.319	0.691
C(club_name)[T.Atletico Madrid]	1.4860	0.094	15.869	0.000	1.302	1.670
C(club_name)[T.Atromitos Athen]	-1.3711	0.091	-15.050	0.000	-1.550	-1.193
C(club_name)[T.Az Alkmaar]	-0.6281	0.093	-6.719	0.000	-0.811	-0.445
C(club_name)[T.Balikesirspor]	-0.8294	0.178	-4.659	0.000	-1.178	-0.480
C(club_name)[T.Bayer 04 Leverkusen]	1.0471	0.093	11.212	0.000	0.864	1.230
C(club_name)[T.Beerschot V A]	-1.2645	0.143	-8.845	0.000	-1.545	-0.984
C(club_name)[T.Belenenses Sad]	-1.3589	0.092	-14.849	0.000	-1.538	-1.180
C(club_name)[T.Benevento Calcio]	-0.3037	0.126	-2.417	0.016	-0.550	-0.057
C(club_name)[T.Benfica Lissabon]	0.5891	0.092	6.387	0.000	0.408	0.770
C(club_name)[T.Besiktas Istanbul]	0.3465	0.092	3.771	0.000	0.166	0.527
C(club_name)[T.Boavista Porto Fc]	-1.4923	0.092	-16.277	0.000	-1.672	-1.313
C(club_name)[T.Borussia Dortmund]	1.1845	0.089	13.236	0.000	1.009	1.360
C(club_name)[T.Borussia Monchengladbach]	0.7316	0.093	7.864	0.000	0.549	0.914
C(club_name)[T.Brescia Calcio]	0.0096	0.144	0.067	0.947	-0.272	0.291
C(club_name)[T.Brighton Amp Hove Albion]	0.8232	0.103	8.003	0.000	0.622	1.025
C(club_name)[T.Brondby If]	-1.0584	0.092	-11.511	0.000	-1.239	-0.878
C(club_name)[T.Bursaspor]	-0.3958	0.102	-3.889	0.000	-0.595	-0.196
C(club_name)[T.Buyuksehir Belediye Erzurumspor]	-1.4850	0.142	-10.456	0.000	-1.763	-1.207
C(club_name)[T.Ca Osasuna]	-0.3354	0.107	-3.124	0.002	-0.546	-0.125
C(club_name)[T.Cagliari Calcio]	0.1359	0.096	1.415	0.157	-0.052	0.324
C(club_name)[T.Cardiff City]	0.1529	0.162	0.942	0.346	-0.165	0.471
C(club_name)[T.Carpi Fc 1909]	-0.3386	0.151	-2.247	0.025	-0.634	-0.043
C(club_name)[T.Catania Calcio]	-0.0172	0.504	-0.034	0.973	-1.005	0.971
C(club_name)[T.Caykur Rizespor]	-0.7259	0.097	-7.512	0.000	-0.915	-0.537
C(club_name)[T.Cd Feirense]	-1.7222	0.113	-15.204	0.000	-1.944	-1.500
C(club_name)[T.Cd Leganes]	-0.1021	0.106	-0.965	0.335	-0.309	0.105

C(club_name)[T.Cd Nacional]	-1.3974	0.101	-13.842	0.000	-1.595	-1.200
C(club_name)[T.Cd Santa Clara]	-1.5681	0.104	-15.032	0.000	-1.773	-1.364
C(club_name)[T.Cd Tondela]	-1.5607	0.094	-16.571	0.000	-1.745	-1.376
C(club_name)[T.Celta Vigo]	0.3010	0.095	3.178	0.001	0.115	0.487
C(club_name)[T.Celtic Glasgow]	-0.2103	0.090	-2.332	0.020	-0.387	-0.034
C(club_name)[T.Cercle Brugge]	-1.0222	0.098	-10.459	0.000	-1.214	-0.831
C(club_name)[T.Cesena Fc]	-0.9259	0.162	-5.701	0.000	-1.244	-0.608
C(club_name)[T.Cf Uniao Madeira]	-1.9126	0.164	-11.645	0.000	-2.234	-1.591
C(club_name)[T.Chievo Verona]	-0.7447	0.106	-7.044	0.000	-0.952	-0.537
C(club_name)[T.Chornomorets Odessa]	-1.5272	0.102	-14.924	0.000	-1.728	-1.327
C(club_name)[T.Clermont Foot 63]	-0.5874	0.190	-3.088	0.002	-0.960	-0.215
C(club_name)[T.Crystal Palace]	0.8424	0.092	9.108	0.000	0.661	1.024
C(club_name)[T.Cs Maritimo]	-1.3668	0.093	-14.740	0.000	-1.549	-1.185
C(club_name)[T.De Graafschap Doetinchem]	-1.9424	0.140	-13.854	0.000	-2.217	-1.668
C(club_name)[T.Delfino Pescara 1936]	-0.5875	0.153	-3.840	0.000	-0.887	-0.288
C(club_name)[T.Denizlispor]	-1.3478	0.144	-9.341	0.000	-1.631	-1.065
C(club_name)[T.Deportivo Alaves]	-0.1007	0.096	-1.054	0.292	-0.288	0.087
C(club_name)[T.Deportivo La Coruna]	-0.1039	0.110	-0.948	0.343	-0.319	0.111
C(club_name)[T.Desna Chernigiv]	-1.4598	0.113	-12.935	0.000	-1.681	-1.239
C(club_name)[T.Desportivo Aves]	-1.6445	0.111	-14.765	0.000	-1.863	-1.426
C(club_name)[T.Dijon Fco]	-0.5194	0.104	-5.018	0.000	-0.722	-0.317
C(club_name)[T.Dinamo Moskau]	0.0411	0.097	0.423	0.672	-0.149	0.231
C(club_name)[T.Dnipro Dnipropetrovsk]	-0.4472	0.119	-3.762	0.000	-0.680	-0.214
C(club_name)[T.Dundee Fc]	-1.8224	0.105	-17.384	0.000	-2.028	-1.617
C(club_name)[T.Dundee United Fc]	-1.8480	0.114	-16.277	0.000	-2.071	-1.625
C(club_name)[T.Dynamo Kiew]	0.1928	0.094	2.043	0.041	0.008	0.378
C(club_name)[T.Ea Guingamp]	-0.6662	0.106	-6.307	0.000	-0.873	-0.459
C(club_name)[T.Eintracht Braunschweig]	-0.6400	0.615	-1.040	0.298	-1.846	0.566
C(club_name)[T.Eintracht	0.1776	0.091	1.954	0.051	-0.001	0.356

Frankfurt]						
C(club_name)[T.Enisey Krasnoyarsk]	-1.1658	0.147	-7.957	0.000	-1.453	-0.879
C(club_name)[T.Es Troyes Ac]	-0.8362	0.119	-7.036	0.000	-1.069	-0.603
C(club_name)[T.Esbjerg Fb]	-1.5751	0.097	-16.294	0.000	-1.765	-1.386
C(club_name)[T.Eskisehirspor]	-0.4748	0.136	-3.485	0.000	-0.742	-0.208
C(club_name)[T.Espanyol Barcelona]	0.0078	0.095	0.082	0.934	-0.178	0.194
C(club_name)[T.Fatih Karagumruk]	-0.9800	0.135	-7.266	0.000	-1.244	-0.716
C(club_name)[T.Fc Arouca]	-1.3753	0.107	-12.819	0.000	-1.586	-1.165
C(club_name)[T.Fc Arsenal]	1.6072	0.093	17.359	0.000	1.426	1.789
C(club_name)[T.Fc Augsburg]	0.0006	0.091	0.006	0.995	-0.179	0.180
C(club_name)[T.Fc Barcelona]	1.7170	0.093	18.490	0.000	1.535	1.899
C(club_name)[T.Fc Bayern Munchen]	1.5246	0.091	16.724	0.000	1.346	1.703
C(club_name)[T.Fc Bologna]	0.1517	0.092	1.645	0.100	-0.029	0.332
C(club_name)[T.Fc Brentford]	0.7550	0.171	4.428	0.000	0.421	1.089
C(club_name)[T.Fc Brugge]	0.0540	0.092	0.590	0.555	-0.125	0.233
C(club_name)[T.Fc Burnley]	0.5212	0.100	5.215	0.000	0.325	0.717
C(club_name)[T.Fc Cadiz]	-0.5253	0.137	-3.825	0.000	-0.795	-0.256
C(club_name)[T.Fc Chelsea]	1.8779	0.095	19.743	0.000	1.691	2.064
C(club_name)[T.Fc Cordoba]	-0.5106	0.175	-2.917	0.004	-0.854	-0.167
C(club_name)[T.Fc Crotone]	-0.7456	0.114	-6.525	0.000	-0.970	-0.522
C(club_name)[T.Fc Dordrecht]	-2.1142	0.176	-12.027	0.000	-2.459	-1.770
C(club_name)[T.Fc Elche]	-0.6117	0.124	-4.915	0.000	-0.856	-0.368
C(club_name)[T.Fc Emmen]	-1.6726	0.117	-14.262	0.000	-1.903	-1.443
C(club_name)[T.Fc Empoli]	-0.3508	0.104	-3.385	0.001	-0.554	-0.148
C(club_name)[T.Fc Everton]	1.3589	0.093	14.645	0.000	1.177	1.541
C(club_name)[T.Fc Famalicao]	-0.6373	0.116	-5.489	0.000	-0.865	-0.410
C(club_name)[T.Fc Fulham]	1.0195	0.128	7.986	0.000	0.769	1.270
C(club_name)[T.Fc Getafe]	0.1145	0.095	1.205	0.228	-0.072	0.301
C(club_name)[T.Fc Girona]	-0.3297	0.132	-2.498	0.012	-0.588	-0.071
C(club_name)[T.Fc Girondins Bordeaux]	0.0052	0.092	0.057	0.955	-0.175	0.186

C(club_name)[T.Fc Granada]	−0.1854	0.098	−1.888	0.059	−0.378	0.007
C(club_name)[T.Fc Groningen]	−1.0644	0.094	−11.290	0.000	−1.249	−0.880
C(club_name)[T.Fc Helsingor]	−2.1198	0.139	−15.220	0.000	−2.393	−1.847
C(club_name)[T.Fc Ingolstadt 04]	−0.6318	0.142	−4.444	0.000	−0.910	−0.353
C(club_name)[T.Fc Kopenhagen]	−0.6901	0.092	−7.513	0.000	−0.870	−0.510
C(club_name)[T.Fc Liverpool]	1.6921	0.093	18.207	0.000	1.510	1.874
C(club_name)[T.Fc Lorient]	−0.5483	0.104	−5.296	0.000	−0.751	−0.345
C(club_name)[T.Fc Malaga]	−0.0099	0.109	−0.090	0.928	−0.224	0.204
C(club_name)[T.Fc Metz]	−0.5922	0.099	−5.952	0.000	−0.787	−0.397
C(club_name)[T.Fc Middlesbrough]	0.5626	0.151	3.734	0.000	0.267	0.858
C(club_name)[T.Fc Midtjylland]	−1.1390	0.091	−12.547	0.000	−1.317	−0.961
C(club_name)[T.Fc Nantes]	−0.3154	0.093	−3.393	0.001	−0.498	−0.133
C(club_name)[T.Fc Nordsjaelland]	−1.4411	0.093	−15.547	0.000	−1.623	−1.259
C(club_name)[T.Fc Pacos De Ferreira]	−1.4240	0.094	−15.227	0.000	−1.607	−1.241
C(club_name)[T.Fc Paris Saint Germain]	1.5523	0.091	16.995	0.000	1.373	1.731
C(club_name)[T.Fc Penafiel]	−1.3852	0.173	−8.000	0.000	−1.725	−1.046
C(club_name)[T.Fc Porto]	0.6184	0.093	6.677	0.000	0.437	0.800
C(club_name)[T.Fc Reading]	0.0733	0.250	0.293	0.769	−0.416	0.563
C(club_name)[T.Fc Schalke 04]	0.7030	0.093	7.530	0.000	0.520	0.886
C(club_name)[T.Fc Sevilla]	0.7570	0.091	8.333	0.000	0.579	0.935
C(club_name)[T.Fc Southampton]	1.0955	0.096	11.432	0.000	0.908	1.283
C(club_name)[T.Fc Stade Rennes]	0.1215	0.093	1.308	0.191	−0.061	0.304
C(club_name)[T.Fc Toulouse]	−0.1991	0.099	−2.010	0.044	−0.393	−0.005
C(club_name)[T.Fc Turin]	0.2943	0.091	3.229	0.001	0.116	0.473
C(club_name)[T.Fc Twente Enschede]	−1.0740	0.096	−11.232	0.000	−1.261	−0.887
C(club_name)[T.Fc Utrecht]	−0.8458	0.091	−9.313	0.000	−1.024	−0.668
C(club_name)[T.Fc Valencia]	1.0982	0.094	11.682	0.000	0.914	1.283
C(club_name)[T.Fc Vestsjaelland]	−1.6439	0.156	−10.544	0.000	−1.950	−1.338
C(club_name)[T.Fc Villarreal]	0.5137	0.093	5.530	0.000	0.332	0.696
C(club_name)[T.Fc Vizela]	−1.6297	0.157	−10.386	0.000	−1.937	−1.322

C(club_name)[T.Fc Watford]	0.6911	0.095	7.244	0.000	0.504	0.878
C(club_name)[T.Fenerbahce Istanbul]	0.4205	0.091	4.625	0.000	0.242	0.599
C(club_name)[T.Feyenoord Rotterdam]	-0.0774	0.093	-0.832	0.406	-0.260	0.105
C(club_name)[T.Fk Khimki]	-0.8106	0.134	-6.053	0.000	-1.073	-0.548
C(club_name)[T.Fk Krasnodar]	0.1754	0.092	1.914	0.056	-0.004	0.355
C(club_name)[T.Fk Mariupol]	-1.4170	0.107	-13.245	0.000	-1.627	-1.207
C(club_name)[T.Fk Minaj]	-2.0720	0.161	-12.867	0.000	-2.388	-1.756
C(club_name)[T.Fk Nizhny Novgorod]	-0.8599	0.160	-5.376	0.000	-1.173	-0.546
C(club_name)[T.Fk Oleksandriya]	-1.4665	0.100	-14.663	0.000	-1.663	-1.270
C(club_name)[T.Fk Orenburg]	-0.9932	0.108	-9.218	0.000	-1.204	-0.782
C(club_name)[T.Fk Rostov]	-0.5281	0.091	-5.780	0.000	-0.707	-0.349
C(club_name)[T.Fk Sochi]	-0.4985	0.112	-4.448	0.000	-0.718	-0.279
C(club_name)[T.Fk Tosno]	-0.9606	0.142	-6.750	0.000	-1.240	-0.682
C(club_name)[T.Fk Ufa]	-0.9799	0.095	-10.312	0.000	-1.166	-0.794
C(club_name)[T.Fortuna Dusseldorf]	-0.0882	0.129	-0.684	0.494	-0.341	0.165
C(club_name)[T.Fortuna Sittard]	-1.4634	0.108	-13.527	0.000	-1.675	-1.251
C(club_name)[T.Frosinone Calcio]	-0.4184	0.132	-3.182	0.001	-0.676	-0.161
C(club_name)[T.Galatasaray Istanbul]	0.3332	0.092	3.606	0.000	0.152	0.514
C(club_name)[T.Gaziantep Fk]	-0.9243	0.120	-7.713	0.000	-1.159	-0.689
C(club_name)[T.Gaziantepspor]	-0.7521	0.119	-6.326	0.000	-0.985	-0.519
C(club_name)[T.Gd Chaves]	-1.2759	0.114	-11.205	0.000	-1.499	-1.053
C(club_name)[T.Gd Estoril Praia]	-1.2582	0.101	-12.488	0.000	-1.456	-1.061
C(club_name)[T.Genclerbirligi Ankara]	-0.8519	0.098	-8.669	0.000	-1.044	-0.659
C(club_name)[T.Genua Cfc]	0.1793	0.093	1.930	0.054	-0.003	0.361
C(club_name)[T.Gfc Ajaccio]	-1.5764	0.197	-7.990	0.000	-1.963	-1.190
C(club_name)[T.Gil Vicente Fc]	-1.4171	0.110	-12.899	0.000	-1.632	-1.202
C(club_name)[T.Giresunspor]	-0.6760	0.221	-3.058	0.002	-1.109	-0.243
C(club_name)[T.Glasgow Rangers]	-0.5688	0.097	-5.877	0.000	-0.759	-0.379
C(club_name)[T.Go Ahead Eagles	-1.8404	0.131	-14.100	0.000	-2.096	-1.585

Deventer]						
C(club_name)[T.Goverla Uzhgorod]	-1.5059	0.133	-11.328	0.000	-1.766	-1.245
C(club_name)[T.Goztepe]	-0.7922	0.103	-7.666	0.000	-0.995	-0.590
C(club_name)[T.Gs Ergotelis]	-1.4728	0.157	-9.365	0.000	-1.781	-1.165
C(club_name)[T.Hamburger Sv]	-0.0352	0.109	-0.323	0.747	-0.249	0.178
C(club_name)[T.Hamilton Academical Fc]	-2.1996	0.102	-21.488	0.000	-2.400	-1.999
C(club_name)[T.Hannover 96]	-0.1220	0.108	-1.134	0.257	-0.333	0.089
C(club_name)[T.Hatayspor]	-1.1647	0.149	-7.809	0.000	-1.457	-0.872
C(club_name)[T.Heart Of Midlothian Fc]	-1.3573	0.098	-13.846	0.000	-1.549	-1.165
C(club_name)[T.Hellas Verona]	-0.2078	0.096	-2.173	0.030	-0.395	-0.020
C(club_name)[T.Heracles Almelo]	-1.5233	0.093	-16.412	0.000	-1.705	-1.341
C(club_name)[T.Hertha Bsc]	0.3591	0.092	3.901	0.000	0.179	0.540
C(club_name)[T.Hibernian Fc]	-1.2933	0.106	-12.182	0.000	-1.501	-1.085
C(club_name)[T.Hobro Ik]	-1.9213	0.099	-19.329	0.000	-2.116	-1.726
C(club_name)[T.Huddersfield Town]	0.1015	0.132	0.769	0.442	-0.157	0.360
C(club_name)[T.Hull City]	0.5449	0.125	4.347	0.000	0.299	0.791
C(club_name)[T.Ingulets Petrove]	-1.8998	0.149	-12.711	0.000	-2.193	-1.607
C(club_name)[T.Inter Mailand]	1.1320	0.092	12.316	0.000	0.952	1.312
C(club_name)[T.Inverness Caledonian Thistle Fc]	-2.2548	0.139	-16.232	0.000	-2.527	-1.983
C(club_name)[T.Ionikos Nikeas]	-1.8971	0.190	-9.964	0.000	-2.270	-1.524
C(club_name)[T.Iraklis Thessaloniki]	-1.8644	0.132	-14.131	0.000	-2.123	-1.606
C(club_name)[T.Istanbul Basaksehir Fk]	-0.1475	0.092	-1.607	0.108	-0.327	0.032
C(club_name)[T.Juventus Turin]	1.5165	0.093	16.346	0.000	1.335	1.698
C(club_name)[T.Kaa Gent]	-0.3283	0.089	-3.673	0.000	-0.503	-0.153
C(club_name)[T.Kardemir Karabukspor]	-0.7373	0.117	-6.306	0.000	-0.966	-0.508
C(club_name)[T.Karpaty Lviv]	-1.4416	0.100	-14.371	0.000	-1.638	-1.245
C(club_name)[T.Kas Eupen]	-1.2221	0.094	-12.991	0.000	-1.406	-1.038
C(club_name)[T.Kasimpasa]	-0.6707	0.094	-7.168	0.000	-0.854	-0.487

C(club_name)[T.Kayseri Erciyesspor]	-0.5147	0.170	-3.020	0.003	-0.849	-0.181
C(club_name)[T.Kayserispor]	-0.8202	0.094	-8.680	0.000	-1.005	-0.635
C(club_name)[T.Kilmarnock Fc]	-1.8104	0.104	-17.431	0.000	-2.014	-1.607
C(club_name)[T.Kolos Kovalivka]	-1.7208	0.121	-14.214	0.000	-1.958	-1.484
C(club_name)[T.Konyaspor]	-0.8049	0.093	-8.666	0.000	-0.987	-0.623
C(club_name)[T.Krc Genk]	-0.2330	0.090	-2.583	0.010	-0.410	-0.056
C(club_name)[T.Krylya Sovetov Samara]	-0.5053	0.097	-5.217	0.000	-0.695	-0.315
C(club_name)[T.Ksc Lokeren]	-0.9655	0.098	-9.816	0.000	-1.158	-0.773
C(club_name)[T.Kuban Krasnodar]	-0.5086	0.149	-3.414	0.001	-0.801	-0.217
C(club_name)[T.Kv Kortrijk]	-0.9904	0.091	-10.844	0.000	-1.169	-0.811
C(club_name)[T.Kv Mechelen]	-0.9513	0.092	-10.310	0.000	-1.132	-0.770
C(club_name)[T.Kv Oostende]	-0.9321	0.090	-10.326	0.000	-1.109	-0.755
C(club_name)[T.Kvc Westerlo]	-1.3165	0.123	-10.676	0.000	-1.558	-1.075
C(club_name)[T.Lazio Rom]	0.5748	0.092	6.257	0.000	0.395	0.755
C(club_name)[T.Leeds United]	0.8682	0.133	6.546	0.000	0.608	1.128
C(club_name)[T.Leicester City]	0.9799	0.094	10.374	0.000	0.795	1.165
C(club_name)[T.Lierse Sk]	-1.0180	0.162	-6.280	0.000	-1.336	-0.700
C(club_name)[T.Livingston Fc]	-1.8763	0.110	-17.048	0.000	-2.092	-1.661
C(club_name)[T.Lokomotiv Moskau]	0.0954	0.094	1.019	0.308	-0.088	0.279
C(club_name)[T.Losc Lille]	0.3121	0.094	3.324	0.001	0.128	0.496
C(club_name)[T.Lyngby Bk]	-1.8061	0.103	-17.564	0.000	-2.008	-1.605
C(club_name)[T.Manchester City]	1.8927	0.095	20.005	0.000	1.707	2.078
C(club_name)[T.Manchester United]	1.7674	0.093	19.030	0.000	1.585	1.949
C(club_name)[T.Mersin Idmanyurdu]	-1.1230	0.140	-8.002	0.000	-1.398	-0.848
C(club_name)[T.Metalist 1925 Kharkiv]	-1.7694	0.173	-10.199	0.000	-2.109	-1.429
C(club_name)[T.Metalist Kharkiv]	-0.6346	0.129	-4.907	0.000	-0.888	-0.381
C(club_name)[T.Metalurg Donetsk]	-0.9515	0.191	-4.980	0.000	-1.326	-0.577
C(club_name)[T.Metalurg Zaporizhya Bis 2016]	-1.3840	0.139	-9.973	0.000	-1.656	-1.112
C(club_name)[T.Mke Ankaragucu]	-1.4019	0.110	-12.699	0.000	-1.618	-1.186

C(club_name)[T.Montpellier Hsc]	-0.4151	0.094	-4.431	0.000	-0.599	-0.231
C(club_name)[T.Mordovia Saransk]	-0.6939	0.146	-4.737	0.000	-0.981	-0.407
C(club_name)[T.Moreirense Fc]	-1.5195	0.091	-16.694	0.000	-1.698	-1.341
C(club_name)[T.Motherwell Fc]	-1.7634	0.100	-17.638	0.000	-1.959	-1.567
C(club_name)[T.Nac Breda]	-1.6113	0.118	-13.656	0.000	-1.843	-1.380
C(club_name)[T.Nec Nijmegen]	-1.4866	0.123	-12.132	0.000	-1.727	-1.246
C(club_name)[T.Newcastle United]	1.0884	0.095	11.426	0.000	0.902	1.275
C(club_name)[T.Niki Volou]	-1.6956	0.187	-9.060	0.000	-2.062	-1.329
C(club_name)[T.Nimes Olympique]	-0.2397	0.120	-1.991	0.047	-0.476	-0.004
C(club_name)[T.Nk Veres Rivne]	-1.5937	0.137	-11.661	0.000	-1.862	-1.326
C(club_name)[T.Norwich City]	0.6714	0.116	5.798	0.000	0.444	0.898
C(club_name)[T.Odense Boldklub]	-1.4570	0.095	-15.400	0.000	-1.642	-1.272
C(club_name)[T.Ofi Kreta]	-1.5216	0.102	-14.939	0.000	-1.721	-1.322
C(club_name)[T.Ogc Nizza]	0.3334	0.093	3.572	0.000	0.150	0.516
C(club_name)[T.Olimpik Donetsk]	-1.6456	0.098	-16.763	0.000	-1.838	-1.453
C(club_name)[T.Olympiakos Piraus]	-0.1644	0.090	-1.834	0.067	-0.340	0.011
C(club_name)[T.Olympique Lyon]	0.7555	0.093	8.161	0.000	0.574	0.937
C(club_name)[T.Olympique Marseille]	0.4870	0.093	5.262	0.000	0.306	0.668
C(club_name)[T.Oud Heverlee Leuven]	-1.1777	0.119	-9.869	0.000	-1.412	-0.944
C(club_name)[T.Palermo Fc]	-0.4999	0.115	-4.339	0.000	-0.726	-0.274
C(club_name)[T.Panathinaikos Athen]	-0.8554	0.089	-9.598	0.000	-1.030	-0.681
C(club_name)[T.Panetolikos Gfs]	-1.6353	0.092	-17.700	0.000	-1.816	-1.454
C(club_name)[T.Panionios Athen]	-1.6285	0.097	-16.719	0.000	-1.819	-1.438
C(club_name)[T.Panthrakikos Komotini]	-1.7697	0.132	-13.366	0.000	-2.029	-1.510
C(club_name)[T.Paok Thessaloniki]	-0.4713	0.090	-5.236	0.000	-0.648	-0.295
C(club_name)[T.Parma Calcio 1913]	0.0142	0.105	0.136	0.892	-0.192	0.220
C(club_name)[T.Partick Thistle Fc]	-1.9459	0.123	-15.815	0.000	-2.187	-1.705
C(club_name)[T.Pas Giannina]	-1.6537	0.097	-17.049	0.000	-1.844	-1.464
C(club_name)[T.Pas Lamia 1964]	-1.9142	0.099	-19.318	0.000	-2.108	-1.720

C(club_name)[T.Pec Zwolle]	-1.3495	0.093	-14.468	0.000	-1.532	-1.167
C(club_name)[T.Pfk Lviv]	-1.8235	0.116	-15.692	0.000	-2.051	-1.596
C(club_name)[T.Pfk Stal Kamyanske]	-1.5078	0.121	-12.496	0.000	-1.744	-1.271
C(club_name)[T.Pfk Tambov]	-1.2875	0.120	-10.716	0.000	-1.523	-1.052
C(club_name)[T.Portimonense Sc]	-1.3102	0.101	-12.930	0.000	-1.509	-1.112
C(club_name)[T.Psv Eindhoven]	0.2291	0.093	2.471	0.013	0.047	0.411
C(club_name)[T.Queens Park Rangers]	0.2684	0.184	1.462	0.144	-0.091	0.628
C(club_name)[T.Randers Fc]	-1.5948	0.094	-16.964	0.000	-1.779	-1.411
C(club_name)[T.Rasenballsport Leipzig]	1.0711	0.096	11.103	0.000	0.882	1.260
C(club_name)[T.Rayo Vallecano]	-0.3662	0.112	-3.279	0.001	-0.585	-0.147
C(club_name)[T.Rc Lens]	-0.5160	0.130	-3.965	0.000	-0.771	-0.261
C(club_name)[T.Rc Strassburg Alsace]	-0.1745	0.105	-1.667	0.095	-0.380	0.031
C(club_name)[T.Rcd Mallorca]	-0.2553	0.127	-2.006	0.045	-0.505	-0.006
C(club_name)[T.Real Betis Sevilla]	0.3661	0.094	3.880	0.000	0.181	0.551
C(club_name)[T.Real Madrid]	1.7153	0.092	18.640	0.000	1.535	1.896
C(club_name)[T.Real Saragossa]	-0.2053	0.393	-0.523	0.601	-0.975	0.565
C(club_name)[T.Real Sociedad San Sebastian]	0.6646	0.092	7.224	0.000	0.484	0.845
C(club_name)[T.Real Valladolid]	-0.1992	0.111	-1.791	0.073	-0.417	0.019
C(club_name)[T.Rfc Seraing]	-1.5597	0.165	-9.478	0.000	-1.882	-1.237
C(club_name)[T.Rio Ave Fc]	-1.1447	0.095	-11.991	0.000	-1.332	-0.958
C(club_name)[T.Rkc Waalwijk]	-1.7844	0.119	-15.036	0.000	-2.017	-1.552
C(club_name)[T.Roda Jc Kerkrade]	-1.6062	0.121	-13.251	0.000	-1.844	-1.369
C(club_name)[T.Ross County Fc]	-1.8293	0.100	-18.227	0.000	-2.026	-1.633
C(club_name)[T.Rotor Volgograd]	-0.9214	0.185	-4.984	0.000	-1.284	-0.559
C(club_name)[T.Royal Antwerpen Fc]	-0.6318	0.098	-6.446	0.000	-0.824	-0.440
C(club_name)[T.Royal Excel Mouscron]	-1.2765	0.091	-13.984	0.000	-1.455	-1.098
C(club_name)[T.Royale Union Saint Gilloise]	-0.8386	0.154	-5.435	0.000	-1.141	-0.536

C(club_name)[T.Rsc Anderlecht]	0.1837	0.088	2.086	0.037	0.011	0.356
C(club_name)[T.Rsc Charleroi]	-0.9001	0.091	-9.856	0.000	-1.079	-0.721
C(club_name)[T.Rubin Kazan]	-0.3546	0.096	-3.709	0.000	-0.542	-0.167
C(club_name)[T.Rukh Lviv]	-1.7859	0.143	-12.499	0.000	-2.066	-1.506
C(club_name)[T.Sampdoria Genua]	0.2594	0.093	2.789	0.005	0.077	0.442
C(club_name)[T.Sbv Excelsior Rotterdam]	-1.7153	0.107	-16.088	0.000	-1.924	-1.506
C(club_name)[T.Sc Bastia]	-1.0434	0.122	-8.548	0.000	-1.283	-0.804
C(club_name)[T.Sc Beira Mar]	-0.5714	0.867	-0.659	0.510	-2.271	1.128
C(club_name)[T.Sc Braga]	-0.2989	0.091	-3.282	0.001	-0.478	-0.120
C(club_name)[T.Sc Cambuur Leeuwarden]	-1.6969	0.126	-13.475	0.000	-1.944	-1.450
C(club_name)[T.Sc Farense]	-1.6438	0.178	-9.246	0.000	-1.992	-1.295
C(club_name)[T.Sc Freiburg]	0.0856	0.093	0.919	0.358	-0.097	0.268
C(club_name)[T.Sc Heerenveen]	-1.0551	0.096	-11.040	0.000	-1.242	-0.868
C(club_name)[T.Sc Olhanense]	-2.0547	0.870	-2.362	0.018	-3.759	-0.350
C(club_name)[T.Sc Paderborn 07]	-0.8829	0.143	-6.165	0.000	-1.164	-0.602
C(club_name)[T.Sco Angers]	-0.4928	0.094	-5.216	0.000	-0.678	-0.308
C(club_name)[T.Sd Eibar]	-0.3049	0.098	-3.105	0.002	-0.497	-0.112
C(club_name)[T.Sd Huesca]	-0.5592	0.133	-4.214	0.000	-0.819	-0.299
C(club_name)[T.Shakhtar Donetsk]	0.4848	0.093	5.214	0.000	0.303	0.667
C(club_name)[T.Sheffield United]	0.6197	0.139	4.467	0.000	0.348	0.892
C(club_name)[T.Silkeborg If]	-1.7335	0.100	-17.412	0.000	-1.929	-1.538
C(club_name)[T.Sivasspor]	-0.7092	0.097	-7.322	0.000	-0.899	-0.519
C(club_name)[T.Sk Dnipro 1]	-1.3228	0.121	-10.974	0.000	-1.559	-1.087
C(club_name)[T.Ska Khabarovsk]	-1.1513	0.135	-8.548	0.000	-1.415	-0.887
C(club_name)[T.Sm Caen]	-0.8034	0.110	-7.281	0.000	-1.020	-0.587
C(club_name)[T.Sonderjyske]	-1.7216	0.092	-18.650	0.000	-1.903	-1.541
C(club_name)[T.Spal]	-0.1259	0.108	-1.161	0.246	-0.339	0.087
C(club_name)[T.Sparta Rotterdam]	-1.4662	0.103	-14.212	0.000	-1.668	-1.264
C(club_name)[T.Spartak Moskau]	0.3307	0.096	3.435	0.001	0.142	0.519
C(club_name)[T.Spezia Calcio]	-0.2332	0.127	-1.833	0.067	-0.483	0.016
C(club_name)[T.Sporting Gijon]	-0.4350	0.136	-3.194	0.001	-0.702	-0.168

C(club_name)[T.Sporting Lissabon]	0.3628	0.090	4.042	0.000	0.187	0.539
C(club_name)[T.Spvgg Greuther Furth]	-0.5461	0.180	-3.041	0.002	-0.898	-0.194
C(club_name)[T.Ssc Neapel]	1.3826	0.093	14.811	0.000	1.200	1.566
C(club_name)[T.St Johnstone Fc]	-1.6149	0.100	-16.220	0.000	-1.810	-1.420
C(club_name)[T.St Mirren Fc]	-1.7823	0.112	-15.963	0.000	-2.001	-1.563
C(club_name)[T.Stade Brest 29]	-0.1555	0.124	-1.258	0.208	-0.398	0.087
C(club_name)[T.Stade Reims]	-0.1831	0.099	-1.856	0.063	-0.377	0.010
C(club_name)[T.Standard Luttich]	-0.2957	0.089	-3.327	0.001	-0.470	-0.122
C(club_name)[T.Stoke City]	0.6480	0.111	5.846	0.000	0.431	0.865
C(club_name)[T.Sv Darmstadt 98]	-0.8247	0.133	-6.198	0.000	-1.086	-0.564
C(club_name)[T.Sv Werder Bremen]	-0.0810	0.093	-0.871	0.384	-0.263	0.101
C(club_name)[T.Sv Zulte Waregem]	-0.9577	0.091	-10.493	0.000	-1.137	-0.779
C(club_name)[T.Swansea City]	0.4265	0.113	3.762	0.000	0.204	0.649
C(club_name)[T.Thonon Evian Grand Geneve Fc]	-0.7158	0.190	-3.770	0.000	-1.088	-0.344
C(club_name)[T.Tom Tomsk]	-1.2667	0.165	-7.669	0.000	-1.590	-0.943
C(club_name)[T.Torpedo Moskau]	-0.9719	0.182	-5.353	0.000	-1.328	-0.616
C(club_name)[T.Tottenham Hotspur]	1.6657	0.095	17.608	0.000	1.480	1.851
C(club_name)[T.Trabzonspor]	-0.1516	0.093	-1.634	0.102	-0.333	0.030
C(club_name)[T.Tsg 1899 Hoffenheim]	0.4987	0.092	5.446	0.000	0.319	0.678
C(club_name)[T.Ud Almeria]	-0.9034	0.194	-4.650	0.000	-1.284	-0.523
C(club_name)[T.Ud Las Palmas]	-0.5201	0.114	-4.581	0.000	-0.743	-0.298
C(club_name)[T.Ud Levante]	-0.1908	0.094	-2.027	0.043	-0.375	-0.006
C(club_name)[T.Udinese Calcio]	0.1533	0.093	1.652	0.099	-0.029	0.335
C(club_name)[T.Ural Ekaterinburg]	-0.9385	0.095	-9.919	0.000	-1.124	-0.753
C(club_name)[T.Us Lecce]	-0.3279	0.147	-2.232	0.026	-0.616	-0.040
C(club_name)[T.Us Salernitana 1919]	-0.3425	0.162	-2.117	0.034	-0.660	-0.025
C(club_name)[T.Us Sassuolo]	0.2407	0.092	2.612	0.009	0.060	0.421
C(club_name)[T.Vejle Boldklub]	-1.6151	0.113	-14.321	0.000	-1.836	-1.394
C(club_name)[T.Vendsyssel Ff]	-1.6472	0.135	-12.208	0.000	-1.912	-1.383

C(club_name)[T.Venezia Fc]	-0.2438	0.164	-1.490	0.136	-0.564	0.077
C(club_name)[T.Veria Nps]	-1.7127	0.119	-14.437	0.000	-1.945	-1.480
C(club_name)[T.Vfb Stuttgart]	0.2793	0.097	2.886	0.004	0.090	0.469
C(club_name)[T.Vfl Bochum]	-0.6721	0.174	-3.853	0.000	-1.014	-0.330
C(club_name)[T.Vfl Wolfsburg]	0.6672	0.092	7.222	0.000	0.486	0.848
C(club_name)[T.Viborg Ff]	-1.6879	0.109	-15.556	0.000	-1.901	-1.475
C(club_name)[T.Vitesse Arnheim]	-0.8675	0.094	-9.241	0.000	-1.052	-0.684
C(club_name)[T.Vitoria Guimaraes Sc]	-0.8025	0.091	-8.818	0.000	-0.981	-0.624
C(club_name)[T.Vitoria Setubal Fc]	-1.4773	0.098	-15.023	0.000	-1.670	-1.285
C(club_name)[T.Volga Nizhniy Novgorod]	-1.3473	0.617	-2.185	0.029	-2.556	-0.139
C(club_name)[T.Volos Nps]	-1.8088	0.124	-14.620	0.000	-2.051	-1.566
C(club_name)[T.Volyn Lutsk]	-1.5340	0.128	-12.000	0.000	-1.785	-1.283
C(club_name)[T.Vorskla Poltava]	-1.2073	0.097	-12.403	0.000	-1.398	-1.016
C(club_name)[T.Vv St Truiden]	-1.0456	0.092	-11.426	0.000	-1.225	-0.866
C(club_name)[T.Vvv Venlo]	-1.6141	0.109	-14.779	0.000	-1.828	-1.400
C(club_name)[T.Waasland Beveren]	-1.1518	0.092	-12.507	0.000	-1.332	-0.971
C(club_name)[T.West Bromwich Albion]	0.5443	0.107	5.077	0.000	0.334	0.754
C(club_name)[T.West Ham United]	1.0111	0.095	10.668	0.000	0.825	1.197
C(club_name)[T.Wigan Athletic]	-0.9346	0.369	-2.532	0.011	-1.658	-0.211
C(club_name)[T.Willem Ii Tilburg]	-1.3026	0.095	-13.759	0.000	-1.488	-1.117
C(club_name)[T.Wolverhampton Wanderers]	1.0977	0.110	9.976	0.000	0.882	1.313
C(club_name)[T.Yeni Malatyaspor]	-1.0827	0.103	-10.469	0.000	-1.285	-0.880
C(club_name)[T.Zenit St Petersburg]	0.6839	0.095	7.171	0.000	0.497	0.871
C(club_name)[T.Zirka Kropyvnytskyi]	-1.7463	0.129	-13.488	0.000	-2.000	-1.493
C(club_name)[T.Zorya Lugansk]	-1.1676	0.096	-12.134	0.000	-1.356	-0.979
C(club_name)[T.Zska Moskau]	0.3449	0.094	3.660	0.000	0.160	0.530
age	0.0201	0.001	20.931	0.000	0.018	0.022
goals	0.0080	0.001	10.457	0.000	0.007	0.009
assists	0.0151	0.001	13.086	0.000	0.013	0.017

minutes_played	0.0002	2.99e-06	68.984	0.000	0.000	0.000
yellow_cards	3.191e-05	0.001	0.035	0.972	-0.002	0.002
red_cards	0.0190	0.007	2.631	0.009	0.005	0.033
height	0.0027	0.000	11.838	0.000	0.002	0.003

Omnibus:	2497.905	**Durbin-Watson:**	1.351
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	3773.239
Skew:	-0.442	**Prob(JB):**	0.00
Kurtosis:	4.001	**Cond. No.**	2.37e+19

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 1.12e-27. This might indicate that there are strong multicollinearity problems or that the design matrix is singular.

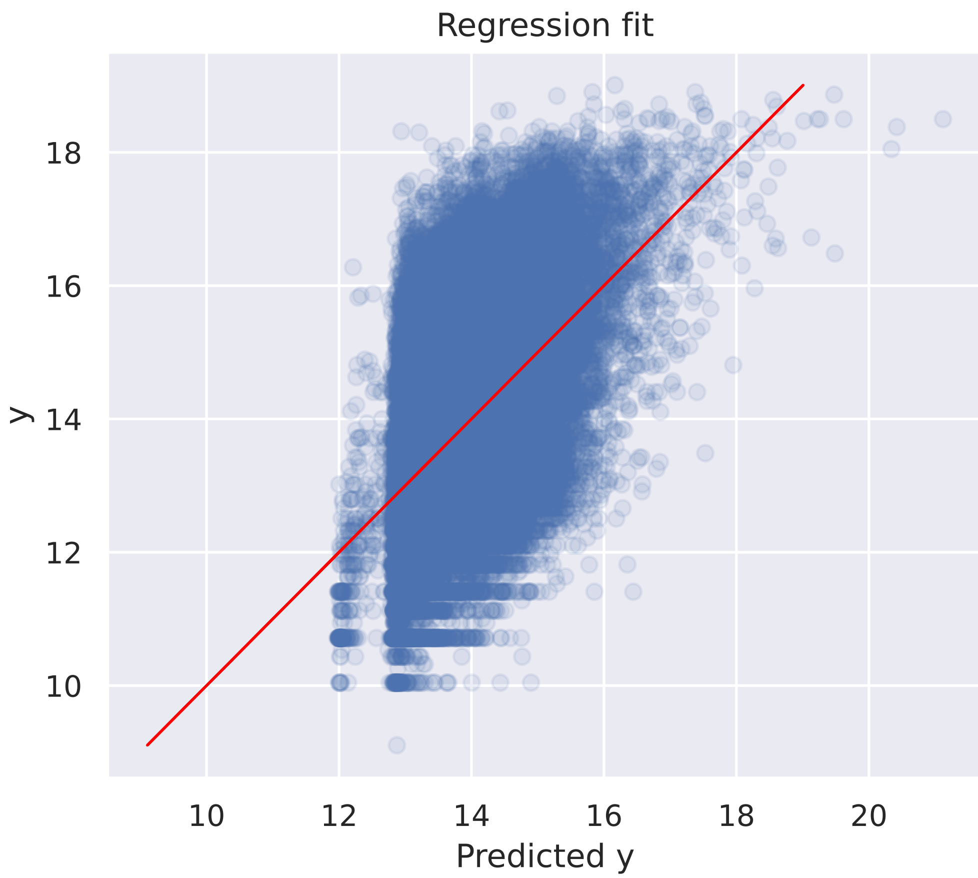
```
print(np.exp(y))
print(np.exp(pred3))

np.sqrt(mean_squared_error(y, pred3))
```

```
             log_market_value
player_id
9800                  90000.0
43084                360000.0
230826               360000.0
198087              1530000.0
110689                68000.0
...                       ...
364245               420000.0
364245              1102500.0
364245              5400000.0
575367               658250.0
575367               765000.0

[50781 rows x 1 columns]
player_id
9800      1.742980e+05
43084     2.559757e+06
230826    8.474457e+05
198087    1.757640e+06
110689    9.412867e+04
              ...
364245    9.750404e+05
364245    4.477367e+06
364245    3.204029e+06
575367    1.339464e+06
575367    1.307280e+06
Length: 50781, dtype: float64
0.8588668100888358
```

```
def diagnostic_plot(x, y):
    plt.figure(figsize=(20,5))

    rgr = LinearRegression()
    rgr.fit(x,y)
    pred = rgr.predict(x)

    plt.subplot(1, 3, 1)
    plt.scatter(pred,y,alpha=0.1)
    plt.plot(y, y, color='red',linewidth=1,)
    plt.title("Regression fit")
    plt.xlabel("Predicted y")
    plt.ylabel("y")

y = data['all_eda']['log_market_value']
X = data['all_eda'][[
    'goals',
    'assists',
    'minutes_played',
    'yellow_cards',
    'red_cards',
    'height',
    'age'
```

```
    # 'nationality',
    # 'position',
    # 'sub_position',
    # 'club_name'
]]
```

```
sns.set(style='darkgrid')
diagnostic_plot(X, y)
```




```
y = data['all_eda']['market_value']
dataset = pd.get_dummies(data['all_eda'], columns = [
    'position',
    'sub_position',
    'nationality',
    'club_name'
])

dataset = dataset.drop(columns = ['name','market_value'])

X_train, X_test, Y_train, Y_test = train_test_split(
    dataset, y, test_size = .30, random_state = 70)

regr = skl.linear_model.LinearRegression()
# Do not use fit_intercept = False if you have removed 1 column after dum
```

```
regr.fit(X_train, Y_train)

predicted = regr.predict(X_test)

regr.score(X_test, Y_test)
```

```
-96149672.18314782
```

Subsection 4

Since our features include both numerical and categorical data we will transform the data and use a pipline to run a linear regression models. Our numeric data is transformed by sklearn's standard scaler and our categorical data will be one hot encoded using sklearn's OneHotEncoder. Our model includes a regularizer. More specifically, it uses l1 regularization. The metric of evaluation here is the score method for sklearn models.

```
# feature selection
X = data['all'][[
    'goals',
    'assists',
    'minutes_played',
    'yellow_cards',
    'red_cards',
    'height',
    'age',
    'season',
    'nationality',
    'position',
    'sub_position',
    'club_name'
]]
# labels
y = data['all']['market_value']
```

```
# numeric features
num_features = ['goals','assists','minutes_played', 'yellow_cards', 'red_c
num_transformer = Pipeline(
    steps=[("scaler", StandardScaler())]
)
categorical_features = ['nationality', 'position', 'sub_position', 'club_n
cat_transformer = OneHotEncoder(handle_unknown='ignore')
preprocessor = ColumnTransformer(
    transformers=[
        ("num", num_transformer, num_features),
        ("cat", cat_transformer, categorical_features)]
)
```

```
# transform and create linear regression model
```

```python
clf = Pipeline(
    steps=[("preprocessor", preprocessor), ("model", Lasso())]
)
# split our data, train on training data, and score on our test data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, r
clf.fit(X_train, y_train)
score = clf.score(X_test, y_test)
score
```

```
/opt/conda/lib/python3.9/site-packages/sklearn/linear_model/_coordinate_d
escent.py:513: ConvergenceWarning: Objective did not converge. You might
want to increase the number of iterations. Duality gap: 3.329092090617099
e+17, tolerance: 277502341864267.94
  model = cd_fast.sparse_enet_coordinate_descent(
0.5174926228911929
```

```python
data['all']['predicted_market_value'] = np.exp(pred3)
```

```python
data['all']
```

player_id	name	nationality	position	sub_position	height	season	market_value
9800	Artem Milevskyi	Ukraine	Attack	Centre-Forward	189	2020	90000.(
43084	Gaetano Berardi	Switzerland	Defender	Right-Back	179	2020	360000.(
230826	Gennaro Acampora	Italy	Midfield	Central Midfield	174	2020	360000.(
198087	Matteo Ricci	Italy	Midfield	Defensive Midfield	176	2020	1530000.(
110689	Deniz Mehmet	Turkey	Goalkeeper	Goalkeeper	192	2020	68000.(
...	...	...	...	...	...	...	..
364245	Jordan Teze	Netherlands	Defender	Centre-Back	183	2019	420000.(
364245	Jordan Teze	Netherlands	Defender	Centre-Back	183	2020	1102500.(
364245	Jordan Teze	Netherlands	Defender	Centre-Back	183	2021	5400000.(
575367	Richard Ledezma	United States	Attack	Attacking Midfield	174	2020	658250.(
575367	Richard Ledezma	United States	Attack	Attacking Midfield	174	2021	765000.(

50781 rows × 15 columns

Discussion

In part 1, we wanted to start with showing the correlation between the variables we have. From our heatmap we can see which variables are the most important and that the most red are the most important variables which are goals, assists, minutes played, red card, and yellow card. The variables in the red staircase are the most red which means they are the most important. Our pair plot shows the relationship between all the variables and what the correlation is between them. For example we can see how red cards and minutes played are strongly correlated.

For part 2, we did our first OLS without categorical data. We wanted to compare the results from OLS done with categorical values and numerical values to see which would give better results. From our first OLS we got some negative values from the T test. We were able to see the correlation between the variables, for example, we see that the red cards variable gave the least correlation. Following that we took the log of the market value. This gave us different T variables and we got different results for the red card and yellow card variables. Our Means of Error also goes down.

We can interpret the first two results as a test ground for non-categorical data. In parts 3 and 4, we wanted to include all the possible data to determine which categories are the most important using the statsmodel OLS design.

After taking the log magnitude of our predicted value, every categorical value was one-hot encoded in the background to add weights in our model. Taking the log reduced the amount of deviation error from each of our variables. When the new data was added, our model showed a new emphasis to our age and height and less to goals and assists. Lastly, including the categorical variables made the red cards and yellow cards number less statistically significant as their p-value too high.

The regression plot showed a way to visualize the data and get an explanation of how much its centralized. Taking log helped removing the differences created because of how much our data was centered closer to 10 million pounds. The visualization of the regression plot was using a new metric of Linear regression instead of OLS. The linear regression produced a single line going through the best fit our predicted and true values. Next, we wanted to incorporate the categorical data as well. The data was split into a test and training set for both labels and samples. To simplify our problem, we used the model score to predict the metric, as our model was resulting in negative values in the market value.

Last but not least, we made another model implementing our categorical data similar to subsection 4. This time, we produced a pipeline of the linear regression which used one-hot encoding, standardized the data using the StandardScalar() module, and lastly, included a regularization lasso term. These new changes prompted our score to be lower than before. The new term reduced to almost half of an error.

The OLS Model has a higher score for our predictions, using the pipeline method lead us to be correct almost half of the time.

Limitations

One of the first limitations we ran into was that we wanted our models to be based on or inbetween specific dates but the dataset only evaluated players at the end of the season. That is why we limited our models to the results at the end of the season.

Another problem that came as a result of our data was due to outliers in regards to market values. This gave us a classic fat tail distribution, and though we tried our best to standardize and regularize our data and model, there could have been more we could've done to combat this. Due to the nature of our data, we were limited on the kind of model to use. Since we were not trying to classify the players, but to recognize the best features, our main focus was using regression. We also ran into some limitations regarding the categorical features that would help us predict market value. Because linear regression can only use numerical data, we had to find a way around this. Therefore, we one hot encoded the categorical features. While this provided a quick solution, there were some values of categories that held more weight in predicting market value. We could definitely have improved on transforming are categorical variables more effeciently since there was alot while also making it so they could be included in our linear models. Because of the vast amount of categorical data, our models ability to learn effectively was deminished. Moreover, we only had enough time to compute the best features that impact a player's worth on the market. With more time, we could have used algorithms to classify the players into certain price ranges by features such as with Neural Networks or Logistic regression. One particular classification problem we would tackle given more time would be categorizing players into those who have a higher market value then 100 million euros or those who have less (this could help clubs find players within a certain budget for example).

Ethics & Privacy

Since our data is readily available to the public and conforms to the privacy policy of the sourcing website Transfermarkt, we believe our research will not be subject to immediate concerns with neither ethics nor data privacy. Variables include `name` / `pretty_name` (name), `country_of_birth` / `country_of_citizenship` (nationality), `date_of_birth` (age) might be potentially relevant but are in no way detrimental to the ethics of our research. However, we do believe that the result of our research, once obtained and made public, could have unintended consequences. We evaluate a player's performance solely based on historical data; this implies that there will be biases. It is reasonable to expect that those biases, if not addressed and handled properly, could cause permanent damage to a person’s career. For instance, if a reliable player’s record shows that the player is unreliable, then team recruiters would use that information to make an informed decision based on false information. To prevent the chance of this being a consequence, we will put a disclaimer regarding the features that were not used in the evaluation of a player’s value. If a problem such as this comes up, we will remove said player's data from the study and make an effort to find and use the accurate data for each player.

Conclusion

The market value of players is a huge part of the football industry. Football clubs across the globe participate in the buying and selling of players and a players market value serves as their price tag. In order to conduct smarter business decisions in terms of buying and selling at the right price, it is crucial to understand a certain players value in the market. This market value is based on a number of factors ranging from performance statistics to even physical characteristics. Our analyses of tens of thousands of players indicate that the most important factors that play a role in determining market value are goals, assists, and minutes played. While yellow cards and red cards play the least role. Our models that included categorical features in training also indicate an emphasis on age, and the great impact it may have on market value. There is no doubt that the football industry involves huge amounts of money, and teams today are run more like businesses. In order to maximize success both financially and on the entertainment side of football, it is important that teams understand how the football market works. Therefore, it is important that research be done in this field. While our data and analyses covered many grounds, it is important that there is improvement. More specifically in terms of predicting market prices based on categorical features. Machine Learning is a relatively new field and there will always be room for improvement. We hope to see future work done in this field, and which improves on our analyses.

Footnotes

[1] Wikipedia contributors. (2022, April 24). *Association football*. Wikipedia. https://en.wikipedia.org/wiki/Association_football

[2] Biswas, B. (2021, July 16). Transfermarkt Market Value explained - How is it determined? *Transfermarkt*. https://www.transfermarkt.co.in/transfermarkt-market-value-explained-how-is-it-determined-/view/news/385100

[3] Peeters, T. (2018). Testing the Wisdom of Crowds in the field: Transfermarkt valuations and international soccer results. *International Journal of Forecasting*, 34(1), 17–29. https://doi.org/10.1016/j.ijforecast.2017.08.002

[4] Ackermann, P., & Follert, F. (2018). *Einige bewertungstheoretische Anmerkungen zur Marktwertanalyse der Plattform transfermarkt.de*. doi:10.22028/D291-32113

[5] Transfermarkt. (2000, May). *Fußball-Transfers, Gerüchte, Marktwerte und News*. https://www.transfermarkt.de/

[6] Summerscales, R. (2022, February 7). Man Utd, Man City And PSG Have Each Spent Over $1Billion Net On Transfers In 10 Years. *Futbol on FanNation*. https://www.si.com/fannation/soccer/futbol/news/man-utd-city-and-psg-spend-over-1b-net-on-transfers-in-10-years

[7] *Football Data from Transfermarkt*. (2022, April 22). [Dataset]. https://www.kaggle.com/datasets/davidcariboo/player-scores