

FIFA22-23 선수분석

제출일	2022.12. 14	학과	컴퓨터공학전공
과목	인공지능	학번	2018213020 2018213019
담당교수	김진석 교수님	이름	남재식 박지석

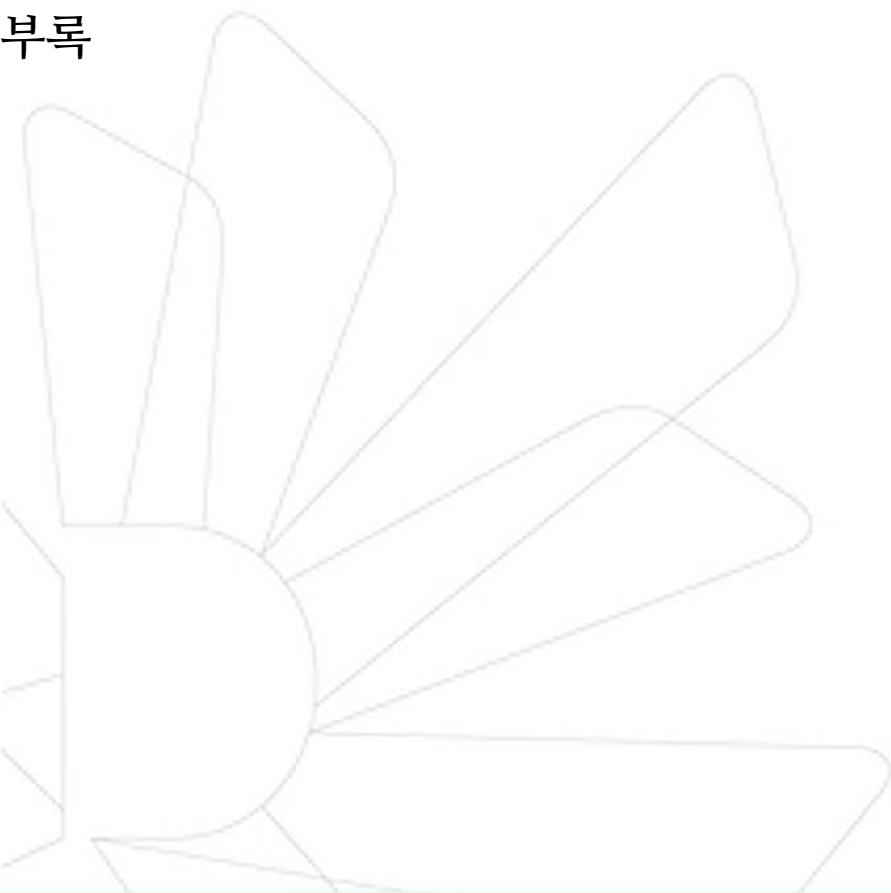


목차

1. 분석 배경 및 분석목적
2. 데이터 소개 및 분석 방법
3. 분석결과
4. 요약 및 결론

참고문헌

부록



분석 결과

1) 게임 내에서 선수 능력치 순위 1위부터 10위까지

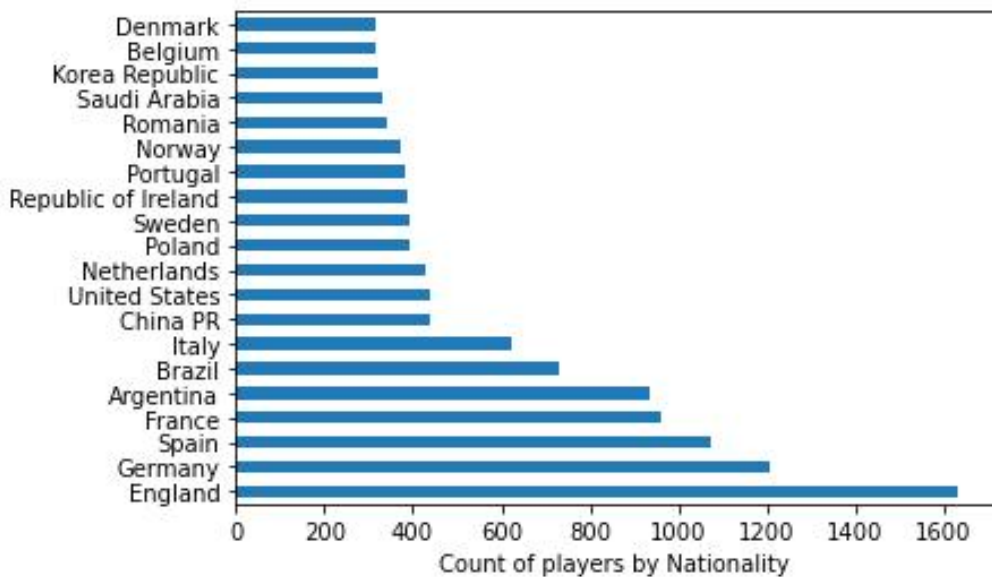
	Known As	Full Name	Overall	Potential	Value(in Euro)	Positions Played	Best Position	Nationality	Image Link	Age	...	CM Rating	R
0	L. Messi	Lionel Messi	91	91	54000000	RW	CAM	Argentina	https://cdn.sofifa.net/players/158/023/23_60.png	35	...	88	
1	K. Benzema	Karim Benzema	91	91	64000000	CF,ST	CF	France	https://cdn.sofifa.net/players/165/153/23_60.png	34	...	84	
2	R. Lewandowski	Robert Lewandowski	91	91	84000000	ST	ST	Poland	https://cdn.sofifa.net/players/188/545/23_60.png	33	...	83	
3	K. De Bruyne	Kevin De Bruyne	91	91	107500000	CM,CAM	CM	Belgium	https://cdn.sofifa.net/players/192/985/23_60.png	31	...	91	
4	K. Mbappé	Kylian Mbappé	91	95	190500000	ST,LW	ST	France	https://cdn.sofifa.net/players/231/747/23_60.png	23	...	84	
6	T. Courtois	Thibaut Courtois	90	91	90000000	GK	GK	Belgium	https://cdn.sofifa.net/players/192/119/23_60.png	30	...	35	
7	M. Neuer	Manuel Neuer	90	90	13500000	GK	GK	Germany	https://cdn.sofifa.net/players/167/495/23_60.png	36	...	53	
8	Cristiano Ronaldo	C. Ronaldo dos Santos Aveiro	90	90	41000000	ST	ST	Portugal	https://cdn.sofifa.net/players/020/801/23_60.png	37	...	81	
9	V. van Dijk	Virgil van Dijk	90	90	98000000	CB	CB	Netherlands	https://cdn.sofifa.net/players/203/376/23_60.png	30	...	79	
10	H. Kane	Harry Kane	89	89	105500000	ST	ST	England	https://cdn.sofifa.net/players/202/126/23_60.png	28	...	84	

분석결과 : 게임 내에서 나오는 선수들의 능력치 순위를 분석해본 결과, 리오넬 메시가 1위 해리 케인이 10위인걸 볼 수 있다. 그리고 선수들의 능력치 뿐만 아니라 임금, 포지션, 나이, 국적 등을 알 수 있었다.

2) 나라별 선수 인원

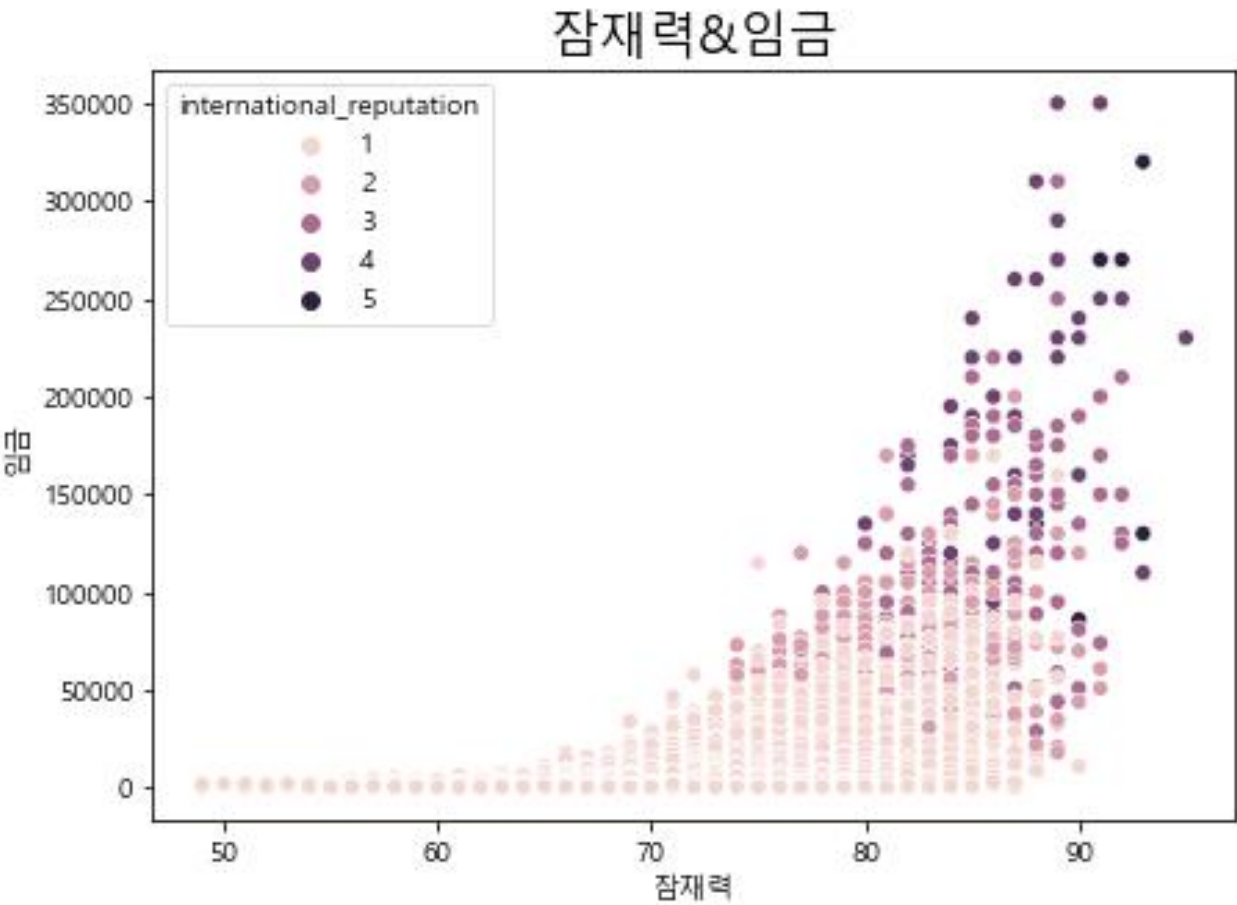
England	1632
Germany	1204
Spain	1074
France	963
Argentina	937
Brazil	728
Italy	623
China PR	439
United States	437
Netherlands	428
Poland	390
Sweden	390
Republic of Ireland	385
Portugal	382
Norway	373
Romania	341
Saudi Arabia	331
Korea Republic	319
Belgium	318
Denmark	318

Name: Nationality, dtype: int64



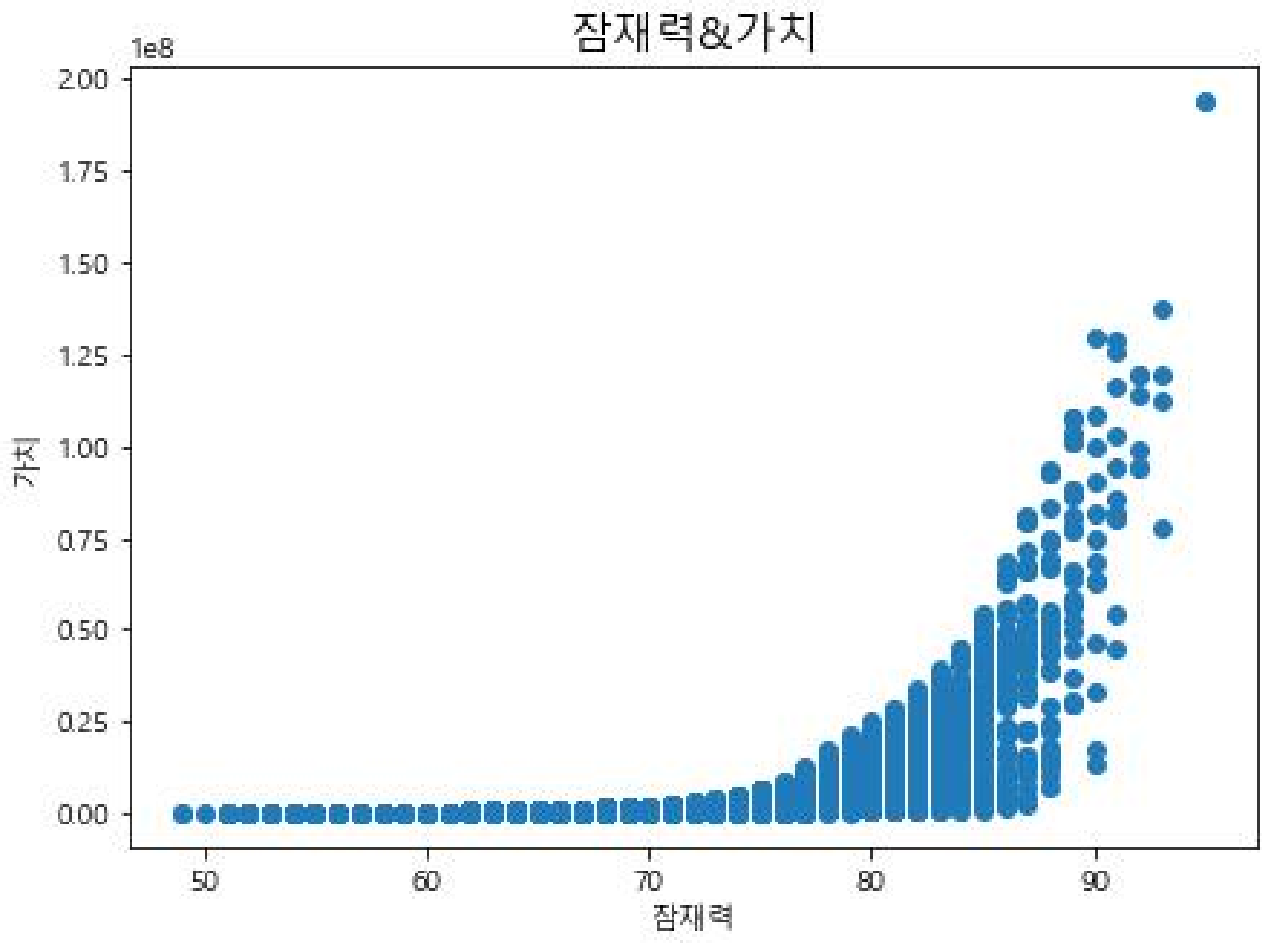
분석 결과 : 나라별 선수 인원의 수를 분석 해보았다. 잉글랜드가 1632명으로 가장 많았고 덴마크와 벨기에가 각각 318명으로 가장 적었다.

3) 선형회귀모형으로 분석한 잠재력&임금



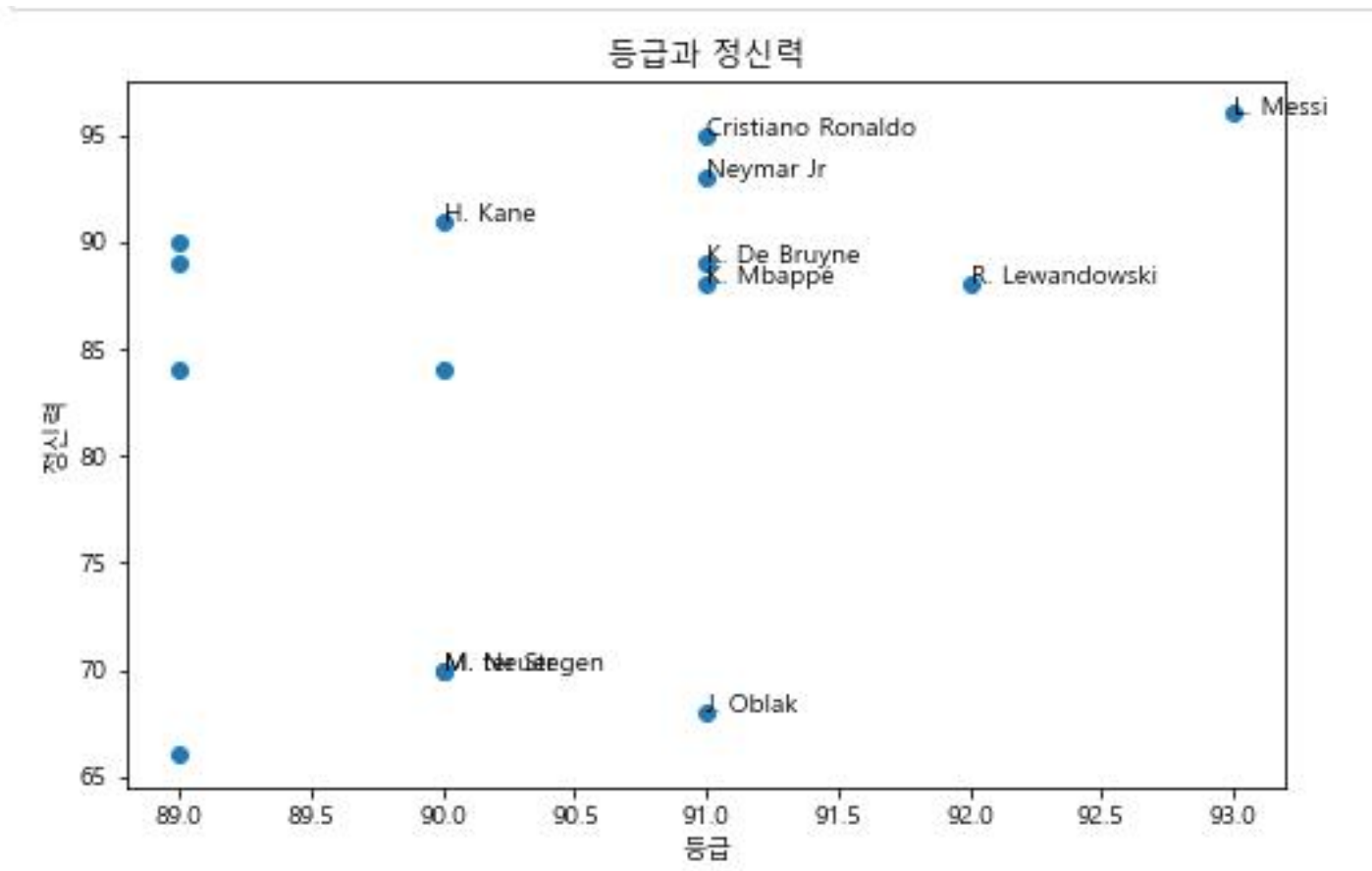
분석 결과 : 선형회귀모형을 이용해서 분석해본 결과 잠재력(능력치)이 높을수록 임금을 높게 받는다는걸 알 수 있었다.

4) 선형회귀모형으로 분석한 잠재력&가치



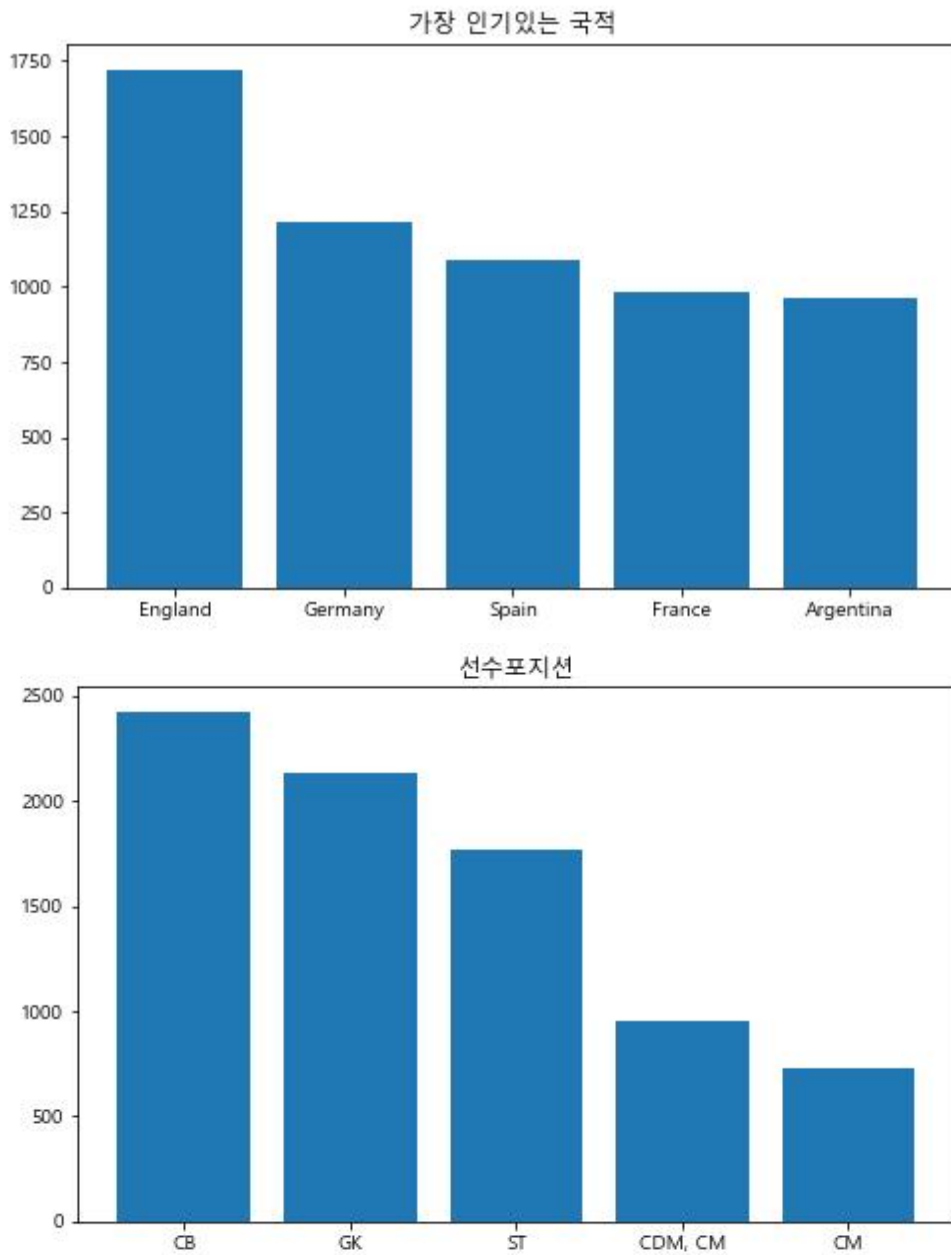
분석 결과 : 3)와 마찬가지로 선형회귀모형을 이용해서 분석해본 결과 잠재력(능력치)이 높을수록 선수의 가치가 높아진다는걸 볼 수 있었다.

5) 판다스 산점도로 분석한 선수별 등급 및 정신력



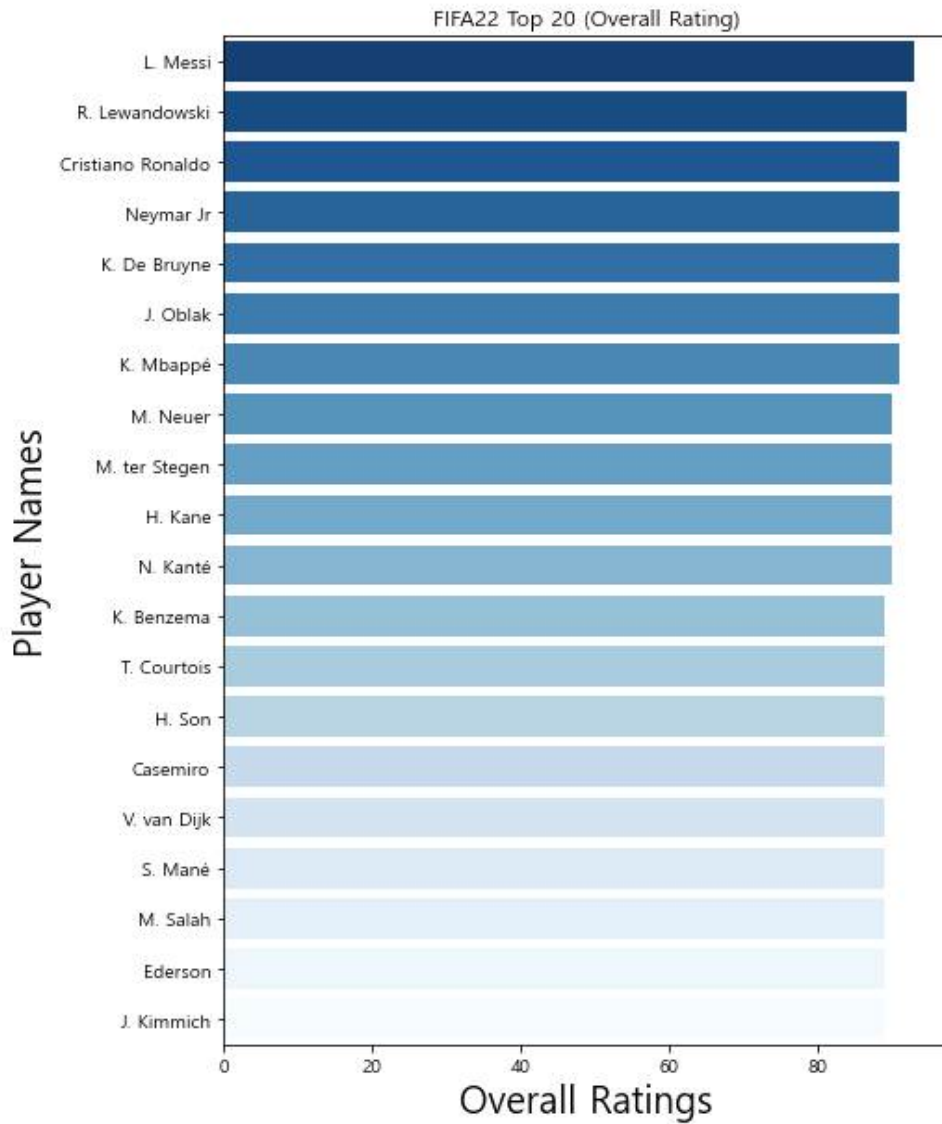
분석 결과 : 판다스 산점도를 활용하여 분석해본 결과 리오넬 메시가 등급 및 정신력이 가장 높다고 나왔고 노이어와 슈테겐, 해리 케인이 등급이 가장 낮았으며 오블락이 정신력이 가장 낮다는걸 볼 수 있었다.

6) 막대그래프를 활용하여 분석한 가장 인기있는 국적 및 포지션별 인원 분포



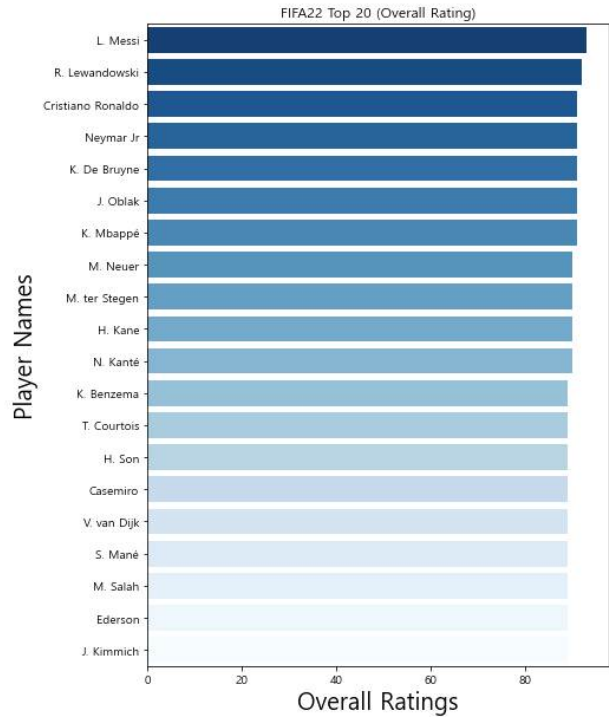
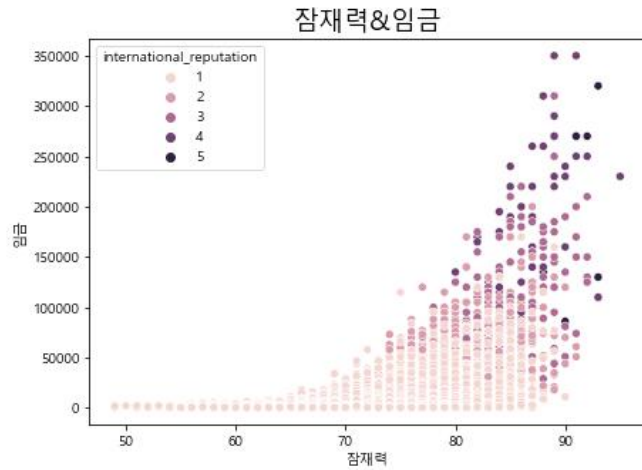
분석 결과 : 막대그래프를 활용하여 분석해본 결과 가장 인기있는 나라는 잉글랜드, 가장 인기없는 나라는 아르헨티나라는걸 볼 수 있다. 그리고 CB라는 포지션에 가장 많은 인원이 있었고 CM이라는 포지션에 가장 적은 인원이 있다는걸 볼 수 있었다.

7) 선수별 능력치



분석 결과 : 막대그래프를 활용하여 분석해본 결과 리오넬 메시가 가장 능력치가 높았고, 조슈아 키미치가 가장 능력치가 낮았다.

요약 및 결론



우리는 중간고사 때 게임별 판매량을 분석했었다. 그리고 이번에는 무엇을 분석할지 고민하다가 중간고사 때 했던 게임과 최근 열린 월드컵이 생각나서 그와 관련된 게임인 FIFA라는 게임을 분석해보기로 결정하였고 이러한 분석 결과가 나왔다. 이 분석을 하면서 느꼈던 한계점은 선수들이 굉장히 많아서 데이터의 양이 엄청났다는 것이다.

참고문헌

1. <http://bigdata.dongguk.ac.kr/> 인공지능 강의노트
2. <https://www.kaggle.com/search?q=fifa+data>

부록

1) 게임 내에서 선수 능력치 순위 1위부터 10위까지

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.metrics import mean_absolute_error
from sklearn.model_selection import train_test_split
from category_encoders import OneHotEncoder
from sklearn.pipeline import make_pipeline
from sklearn.linear_model import LinearRegression, Ridge

def wrangle(filepath):
    df = pd.read_csv(filepath)

    df['Nationality Count'] = df.groupby('Nationality')['Nationality'].transform('count')

    mask_nation = df['Nationality Count'] > 300
    df = df[mask_nation]

    return df

df = wrangle('Fifa 23 Players Data.csv')
df.head(10)
```

2) 나라별 선수 인원

```
df_nationality_counts = df["Nationality"].value_counts()
df_nationality_counts

file = "players_22.csv"
df = pd.read_csv(file)
df.head(5)
```

3) 잠재력&임금, 잠재력&가치

```
plt.figure(figsize=(7, 5))
ax = sns.scatterplot(x=df['potential'], y=df['wage_eur'], hue=df['international_reputation'])
plt.xlabel("잠재력")
plt.ylabel("임금")
plt.rc('font', family='Malgun Gothic')
plt.title("잠재력&임금", fontsize=18)
plt.show()

fig, ax = plt.subplots(figsize=(7, 5))
plt.scatter(x=df['potential'], y=df['value_eur'])
plt.xlabel("잠재력")
plt.ylabel("가치")
plt.title("잠재력&가치", fontsize=15)
plt.show()
```

4) 선수별 등급 및 정신력

```
fig, ax = plt.subplots(figsize=(8,5))

plt.scatter(top_15['overall'], top_15['mentality_composure'])

plt.text(top_15.iloc[0]['overall'], top_15.iloc[0]['mentality_composure'], top_15.iloc[0]['short_name'])
plt.text(top_15.iloc[1]['overall'], top_15.iloc[1]['mentality_composure'], top_15.iloc[1]['short_name'])
plt.text(top_15.iloc[2]['overall'], top_15.iloc[2]['mentality_composure'], top_15.iloc[2]['short_name'])
plt.text(top_15.iloc[3]['overall'], top_15.iloc[3]['mentality_composure'], top_15.iloc[3]['short_name'])
plt.text(top_15.iloc[4]['overall'], top_15.iloc[4]['mentality_composure'], top_15.iloc[4]['short_name'])
plt.text(top_15.iloc[5]['overall'], top_15.iloc[5]['mentality_composure'], top_15.iloc[5]['short_name'])
plt.text(top_15.iloc[6]['overall'], top_15.iloc[6]['mentality_composure'], top_15.iloc[6]['short_name'])
plt.text(top_15.iloc[7]['overall'], top_15.iloc[7]['mentality_composure'], top_15.iloc[7]['short_name'])
plt.text(top_15.iloc[8]['overall'], top_15.iloc[8]['mentality_composure'], top_15.iloc[8]['short_name'])
plt.text(top_15.iloc[9]['overall'], top_15.iloc[9]['mentality_composure'], top_15.iloc[9]['short_name'])

ax.set_title("등급과 정신력")
ax.set_ylabel('정신력')
ax.set_xlabel('등급')

plt.show()
```

5) 가장 인기있는 국적 및 포지션별 인원 분포

```
from collections import Counter
bar_plot = dict(Counter(df['nationality_name'].values).most_common(5))
bar_plot

fig, ax = plt.subplots(figsize = (8,5))
plt.bar(*zip(*bar_plot.items()))
ax.set_title('가장 인기있는 국적')
plt.show()

def plot_most_common(category):
    bar_plot = dict(Counter(df[category].values).most_common(5))
    plt.bar(*zip(*bar_plot.items()))
    ax.set_title('선수포지션')
    plt.show()
```

```
fig, ax = plt.subplots(figsize = (8,5))
plot_most_common('player_positions')
plt.show()
```

6) 선수별 능력치

```
Overall = df["overall"]
footballer_name = df["short_name"]

data = pd.DataFrame({'short_name': footballer_name, 'overall': Overall})

x = df['short_name'].head(20)
y = df['overall'].head(20)

plt.figure(figsize=(7,10))

ax = sns.barplot(x=y, y=x, palette = 'Blues_r', orient='h')
plt.xticks()
plt.xlabel('Overall Ratings', size = 20)
plt.ylabel('Player Names', size = 20)
plt.title('FIFA22 Top 20 (Overall Rating)')

plt.show()
```