

# SQL 기초와 데이터 분석

하홍석

## 7. SQL을 활용한 데이터 분석

## 7. SQL을 활용한 데이터 분석

---

1. Indian Restaurant Dataset
2. Global AI, ML, Data Science Salary
3. NBA Players

# 1. Indian Restaurant Dataset

1. Data Source : <https://www.kaggle.com/datasets/abhijitdahatonde/27000-indian-restaurant-dataset>

## 2. Contents

| restaurant_name       | rating | average_price | average_delivery_time | south_indian_or_not | north_indian_or_not | fast_food_or_not | street_food | biryani_or_not | bakery_or_not | location   |
|-----------------------|--------|---------------|-----------------------|---------------------|---------------------|------------------|-------------|----------------|---------------|------------|
| # Momo                | 4.2    | 200           | 34                    | 0                   | 0                   | 0                | 0           | 0              | 0             | Durgapur   |
| #99                   | 4.1    | 100           | 31                    | 0                   | 0                   | 0                | 0           | 0              | 0             | Bareilly   |
| #BBQ                  | 3.6    | 200           | 57                    | 0                   | 0                   | 1                | 0           | 0              | 0             | Mangalore  |
| #Dilliwaala6          | 3.7    | 100           | 38                    | 0                   | 1                   | 0                | 1           | 0              | 0             | Puducherry |
| #FlyFree              | 4.0    | 200           | 26                    | 0                   | 0                   | 1                | 0           | 0              | 0             | Ajmer      |
| #Foodiz               | 3.5    | 100           | 23                    | 0                   | 1                   | 1                | 0           | 0              | 0             | Udaipur    |
| #Hashtag Momos        | 4.3    | 300           | 25                    | 0                   | 0                   | 1                | 0           | 0              | 0             | Jaipur     |
| #PB Patiala Bites     | 4.0    | 350           | 47                    | 0                   | 1                   | 1                | 0           | 0              | 0             | Udaipur    |
| '@ Pizza              | 4.0    | 200           | 71                    | 0                   | 0                   | 0                | 0           | 0              | 0             | Kolkata    |
| '@ Spirants Chai      | 4.2    | 150           | 24                    | 0                   | 0                   | 0                | 0           | 0              | 0             | Gaya       |
| '@69 Food Zone        | 3.7    | 200           | 35                    | 0                   | 1                   | 0                | 0           | 0              | 0             | Kolkata    |
| 0265 Burgers & Pizzas | 4.5    | 150           | 24                    | 0                   | 0                   | 1                | 0           | 0              | 0             | Vadodara   |
| 0612 Road Chef Online | 4.2    | 250           | 35                    | 0                   | 1                   | 0                | 0           | 1              | 0             | Patna      |
| 07 Bakery             | 2.7    | 150           | 23                    | 0                   | 0                   | 0                | 0           | 0              | 0             | Kota       |

# 1. Indian Restaurant Dataset

---

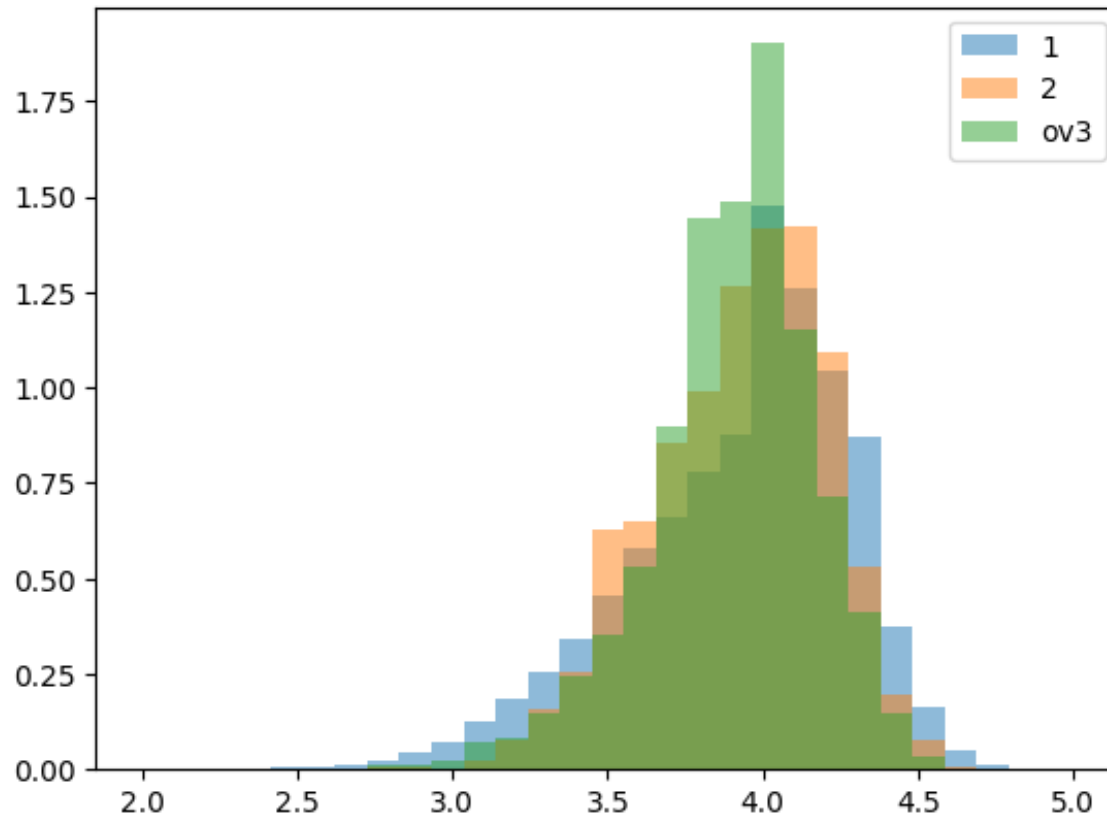
- 1. 점포 수가 1개인 식당을 단일 점포, 2개 이상인 식당을 프렌차이즈 라고 정의하자
- 2. 별점 기준 상위 20개 식당은 모두 단일 점포다
- 3. 평균적으로 프렌차이즈의 평점이 단일 점포보다 높다

| cnt_group | avg_rating        |
|-----------|-------------------|
| cnt-1     | 3.911160332242648 |
| cnt-2     | 3.92796741660202  |
| cnt-ov3   | 3.953155572606939 |

# 1. Indian Restaurant Dataset

## 4. 별점 분포 히스토그램

1. 단일 점포는 분포가 양끝으로 더 넓게 퍼져 있다. -> 양 극단의 값이 더 많다.



# 1. Indian Restaurant Dataset

## 5. 상관계수

전체 데이터

| 변수           | 별점과의 상관계수    |
|--------------|--------------|
| 점포 개수 (=cnt) | 0.017        |
| 평균 가격        | -0.007       |
| 평균 배달 시간     | <b>-0.09</b> |

점포 개수 >= 3

| 변수       | 별점과의 상관계수     |
|----------|---------------|
| 평균 가격    | -0.019        |
| 평균 배달 시간 | <b>-0.158</b> |

스트리트 푸드 여부 = 1

| 변수       | 별점과의 상관계수     |
|----------|---------------|
| 평균 가격    | -0.005        |
| 평균 배달 시간 | <b>-0.139</b> |

대도시

| 변수       | 별점과의 상관계수 |
|----------|-----------|
| 평균 가격    | 0.028     |
| 평균 배달 시간 | -0.112    |

소도시

| 변수       | 별점과의 상관계수     |
|----------|---------------|
| 평균 가격    | <b>-0.045</b> |
| 평균 배달 시간 | <b>-0.232</b> |

# 1. Indian Restaurant Dataset

## 6. 그룹별 평균 비교

그룹별 평균

| group_name | avg_price | avg_time | avg_rating |
|------------|-----------|----------|------------|
| south-0    | 174.3019  | 33.4573  | 3.917      |
| south-1    | 170.9903  | 32.7229  | 3.946      |
| north-0    | 172.1089  | 32.4662  | 3.938      |
| north-1    | 176.3204  | 34.6064  | 3.896      |
| fast-0     | 174.9690  | 32.8176  | 3.930      |
| fast-1     | 171.9505  | 34.3144  | 3.904      |
| street-0   | 174.9915  | 33.4678  | 3.921      |
| street-1   | 168.6151  | 32.8646  | 3.919      |
| biryani-0  | 172.8758  | 33.2066  | 3.929      |
| biryani-1  | 178.5484  | 34.0911  | 3.881      |
| bakery-0   | 174.0193  | 33.6083  | 3.926      |
| bakery-1   | 172.4687  | 31.0235  | 3.869      |

가격이 가장 비싼 지역과 저렴한 지역 비교

| location  | tot_cnt | south_cnt | north_cnt | fast_cnt | street_cnt | biryani_cnt | bakery_cnt |
|-----------|---------|-----------|-----------|----------|------------|-------------|------------|
| Rishikesh | 44      | 9         | 26        | 26       | 16         | 2           | 1          |
| Shimla    | 20      | 0         | 11        | 11       | 3          | 4           | 1          |

평점이 가장 높은 지역과 낮은 지역 비교

| location | tot_cnt | south_cnt | north_cnt | fast_cnt | street_cnt | biryani_cnt | bakery_cnt |
|----------|---------|-----------|-----------|----------|------------|-------------|------------|
| Junagadh | 37      | 2         | 7         | 14       | 8          | 3           | 2          |
| Motihari | 16      | 2         | 9         | 6        | 6          | 7           | 2          |



## 2. Global AI, ML, Data Science Salary

1. Data Source : <https://www.kaggle.com/dparas01/global-ai-ml-data-science-salary>

### 2. Contents

| work_year | experience_level | employment_type | job_title                      | salary | salary_currency | salary_in_usd | employee_residence | remote_ratio | company_location | company_size |
|-----------|------------------|-----------------|--------------------------------|--------|-----------------|---------------|--------------------|--------------|------------------|--------------|
| 2023      | EX               | FT              | Data Science Director          | 212000 | USD             | 212000        | US                 | 0            | US               | M            |
| 2023      | EX               | FT              | Data Science Director          | 190000 | USD             | 190000        | US                 | 0            | US               | M            |
| 2023      | MI               | FT              | Business Intelligence Engineer | 35000  | GBP             | 43064         | GB                 | 0            | GB               | M            |
| 2023      | MI               | FT              | Business Intelligence Engineer | 35000  | GBP             | 43064         | GB                 | 0            | GB               | M            |
| 2023      | SE               | FT              | Machine Learning Engineer      | 245700 | USD             | 245700        | US                 | 0            | US               | M            |
| 2023      | SE               | FT              | Machine Learning Engineer      | 132300 | USD             | 132300        | US                 | 0            | US               | M            |
| 2023      | MI               | FT              | Data Specialist                | 90000  | USD             | 90000         | US                 | 0            | US               | M            |
| 2023      | MI               | FT              | Data Specialist                | 80000  | USD             | 80000         | US                 | 0            | US               | M            |
| 2023      | SE               | FT              | Machine Learning Engineer      | 212000 | USD             | 212000        | US                 | 0            | US               | M            |
| 2023      | SE               | FT              | Machine Learning Engineer      | 93300  | USD             | 93300         | US                 | 0            | US               | M            |
| 2023      | MI               | FT              | Data Scientist                 | 212000 | USD             | 212000        | US                 | 0            | US               | M            |
| 2023      | MI               | FT              | Data Scientist                 | 93300  | USD             | 93300         | US                 | 0            | US               | M            |

## 2. Global AI, ML, Data Science Salary

3. 22년에 비해 23년에 평균 연봉이 증가한 이유는, 시니어 비중이 높아지고 중간 크기 회사 비중이 높아 졌기 때문이다.

연도별 평균 연봉

| work_year | usd_salary  |
|-----------|-------------|
| 2022      | 134449.1381 |
| 2023      | 155198.4878 |

숙련도별 평균 연봉

| experience_level | usd_salary  |
|------------------|-------------|
| EN               | 91388.4539  |
| MI               | 116621.0223 |
| SE               | 162391.0944 |
| EX               | 189309.9647 |

회사 크기별 평균 연봉

| company_size | usd_salary  |
|--------------|-------------|
| S            | 95446.0686  |
| L            | 141378.9649 |
| M            | 152651.3354 |

연도별 숙련도 비중

| experience_level | ratio_2022 | ratio_2023 |
|------------------|------------|------------|
| EN               | 7.03%      | 4.15%      |
| EX               | 2.48%      | 3.12%      |
| MI               | 21.68%     | 18.31%     |
| SE               | 68.81%     | 74.42%     |
|                  | 100.00%    | 100.00%    |

연도별 회사 크기 비중

| company_size | ratio_2022 | ratio_2023 |
|--------------|------------|------------|
| L            | 11.63%     | 5.92%      |
| M            | 85.40%     | 93.31%     |
| S            | 2.97%      | 0.77%      |
|              | 100.00%    | 100.00%    |

## 2. Global AI, ML, Data Science Salary

4. 22년에 비해 23년에 평균 원격 근무 비율이 감소한 이유는, 시니어 직원 비중이 높아지고 해외 근무자 비중이 줄어 들었기 때문이다.

연도별 평균 원격근무 비율

| work_year | avg_remote_ratio |
|-----------|------------------|
| 2022      | 55.2090          |
| 2023      | 33.4718          |



숙련도별 평균 원격 근무 비중

| experience_level | remote_ratio |
|------------------|--------------|
| MI               | 35.9356      |
| SE               | 37.5200      |
| EX               | 41.7647      |
| EN               | 44.7631      |

연도별 미국 내외 근무 비중

| residence_group | ratio_2022 | ratio_2023 |
|-----------------|------------|------------|
| In-US           | 80.25%     | 88.69%     |
| Out-US          | 19.75%     | 11.31%     |
|                 | 100.00%    | 100.00%    |

## 2. Global AI, ML, Data Science Salary

### 5. 직무 그룹별 평균

직무별 평균 연봉, 원격 근무 비중

| job_group | mean_salary | remote_ratio | cnt  |
|-----------|-------------|--------------|------|
| SP        | 100668.7639 | 31.2500      | 72   |
| A         | 108129.7664 | 44.1754      | 1528 |
| C         | 114600.7442 | 23.2558      | 43   |
| Other     | 138896.6069 | 36.2595      | 262  |
| P         | 144480.0000 | 0.0000       | 10   |
| M         | 144556.2071 | 25.8929      | 280  |
| E         | 158492.8584 | 35.9232      | 3829 |
| S         | 161875.2231 | 39.9724      | 2533 |
| AC        | 169045.2917 | 59.7222      | 216  |
| D         | 222549.5625 | 46.8750      | 32   |

직무별 미국 내외 근무 비중

| job_group | residence_group | group_cnt | ratio  |
|-----------|-----------------|-----------|--------|
| A         | In-US           | 1295      | 0.8475 |
| A         | Out-US          | 233       | 0.1525 |
| AC        | In-US           | 198       | 0.9167 |
| AC        | Out-US          | 18        | 0.0833 |
| C         | In-US           | 30        | 0.6977 |
| C         | Out-US          | 13        | 0.3023 |
| D         | In-US           | 21        | 0.6563 |
| D         | Out-US          | 11        | 0.3438 |
| E         | In-US           | 3295      | 0.8605 |
| E         | Out-US          | 534       | 0.1395 |
| M         | In-US           | 257       | 0.9179 |
| M         | Out-US          | 23        | 0.0821 |
| Other     | In-US           | 194       | 0.7405 |
| Other     | Out-US          | 68        | 0.2595 |
| P         | In-US           | 10        | 1.0000 |
| S         | In-US           | 2164      | 0.8543 |
| S         | Out-US          | 369       | 0.1457 |
| SP        | In-US           | 63        | 0.8750 |
| SP        | Out-US          | 9         | 0.1250 |

### 3. NBA Players

1. Data Source : <https://www.kaggle.com/datasets/justinas/nba-players-data/data>
2. Contents

| player_name      | team_abbreviation | age  | player_height | player_weight      | college               | country | draft_year | draft_round | draft_number | gp | pts  | reb | ast | net_rating | oreb_pct             |
|------------------|-------------------|------|---------------|--------------------|-----------------------|---------|------------|-------------|--------------|----|------|-----|-----|------------|----------------------|
| Randy Livingston | HOU               | 22.0 | 193.04        | 94.800728          | Louisiana State       | USA     | 1996       | 2           | 42           | 64 | 3.9  | 1.5 | 2.4 | 0.3        | 0.042                |
| Gaylon Nickerson | WAS               | 28.0 | 190.5         | 86.18248           | Northwestern Oklahoma | USA     | 1994       | 2           | 34           | 4  | 3.8  | 1.3 | 0.3 | 8.9        | 0.03                 |
| George Lynch     | VAN               | 26.0 | 203.2         | 103.418976         | North Carolina        | USA     | 1993       | 1           | 12           | 41 | 8.3  | 6.4 | 1.9 | -8.2       | 0.106                |
| George McCloud   | LAL               | 30.0 | 203.2         | 102.0582           | Florida State         | USA     | 1989       | 1           | 7            | 64 | 10.2 | 2.8 | 1.7 | -2.7       | 0.027000000000000003 |
| George Zidek     | DEN               | 23.0 | 213.36        | 119.74828799999999 | UCLA                  | USA     | 1995       | 1           | 22           | 52 | 2.8  | 1.7 | 0.3 | -14.1      | 0.102                |
| Gerald Wilkins   | ORL               | 33.0 | 198.12        | 102.0582           | Tennessee-Chattanooga | USA     | 1985       | 2           | 47           | 80 | 10.6 | 2.2 | 2.2 | -5.8       | 0.031                |
| Gheorghe Muresan | WAS               | 26.0 | 231.14        | 137.438376         | None                  | USA     | 1993       | 2           | 30           | 73 | 10.6 | 6.6 | 0.4 | 6.9        | 0.098                |
| Glen Rice        | CHH               | 30.0 | 203.2         | 99.79024           | Michigan              | USA     | 1989       | 1           | 4            | 79 | 26.8 | 4.0 | 2.0 | 3.2        | 0.025                |
| Glenn Robinson   | MIL               | 24.0 | 200.66        | 106.59411999999999 | Purdue                | USA     | 1994       | 1           | 1            | 80 | 21.1 | 6.3 | 3.1 | -2.9       | 0.051                |
| Grant Hill       | DET               | 24.0 | 203.2         | 102.0582           | Duke                  | USA     | 1994       | 1           | 3            | 80 | 21.4 | 9.0 | 7.3 | 6.9        | 0.049                |
| Gary Trent       | POR               | 22.0 | 203.2         | 113.398            | Ohio                  | USA     | 1995       | 1           | 11           | 82 | 10.8 | 5.2 | 1.1 | 2.5        | 0.10099999999999999  |
| Grant Long       | DET               | 31.0 | 205.74        | 112.490816         | Eastern Michigan      | USA     | 1988       | 2           | 33           | 65 | 5.0  | 3.4 | 0.6 | 4.0        | 0.096                |
| Greg Anthony     | VAN               | 29.0 | 185.42        | 81.64656           | Nevada-Las Vegas      | USA     | 1991       | 1           | 12           | 65 | 9.5  | 2.8 | 6.3 | -9.4       | 0.015                |
| Greg Dreiling    | DAL               | 33.0 | 215.9         | 120.20188          | Kansas                | USA     | 1986       | 2           | 26           | 40 | 2.0  | 1.9 | 0.3 | -8.0       | 0.059000000000000004 |
| Greg Foster      | UTA               | 28.0 | 210.82        | 113.398            | Texas-El Paso         | USA     | 1990       | 2           | 35           | 79 | 3.5  | 2.4 | 0.4 | -0.9       | 0.078                |
| Greg Graham      | SEA               | 26.0 | 193.04        | 82.553744          | Indiana               | USA     | 1993       | 1           | 17           | 28 | 3.3  | 0.5 | 0.4 | 3.6        | 0.013000000000000001 |

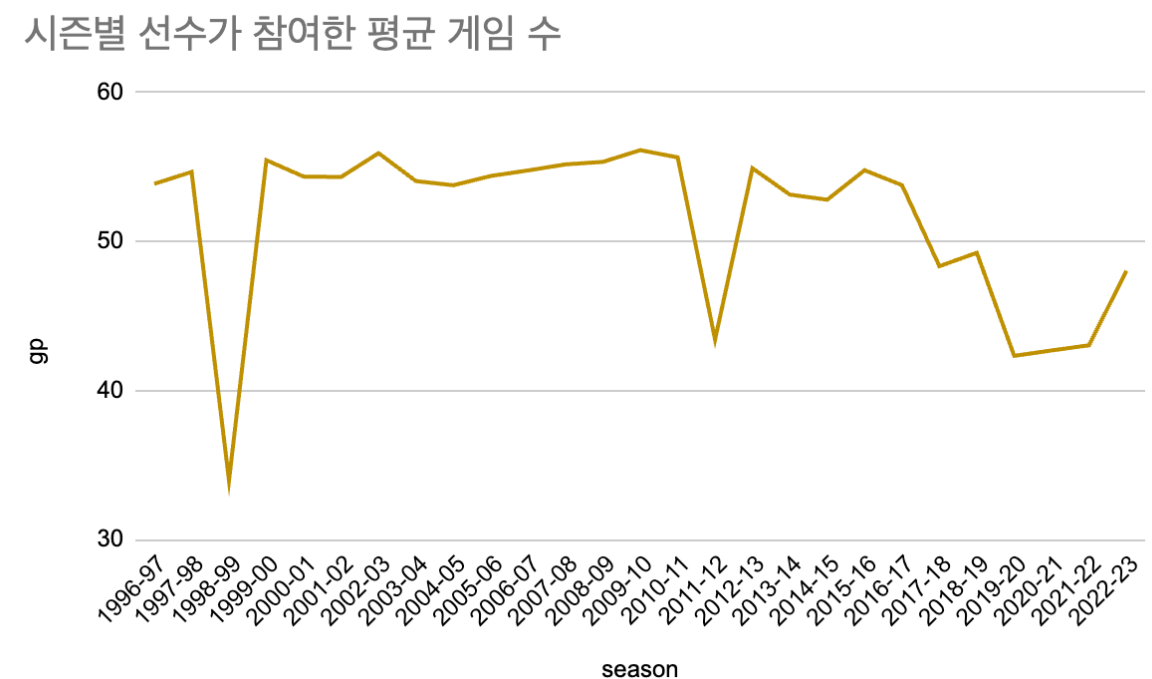
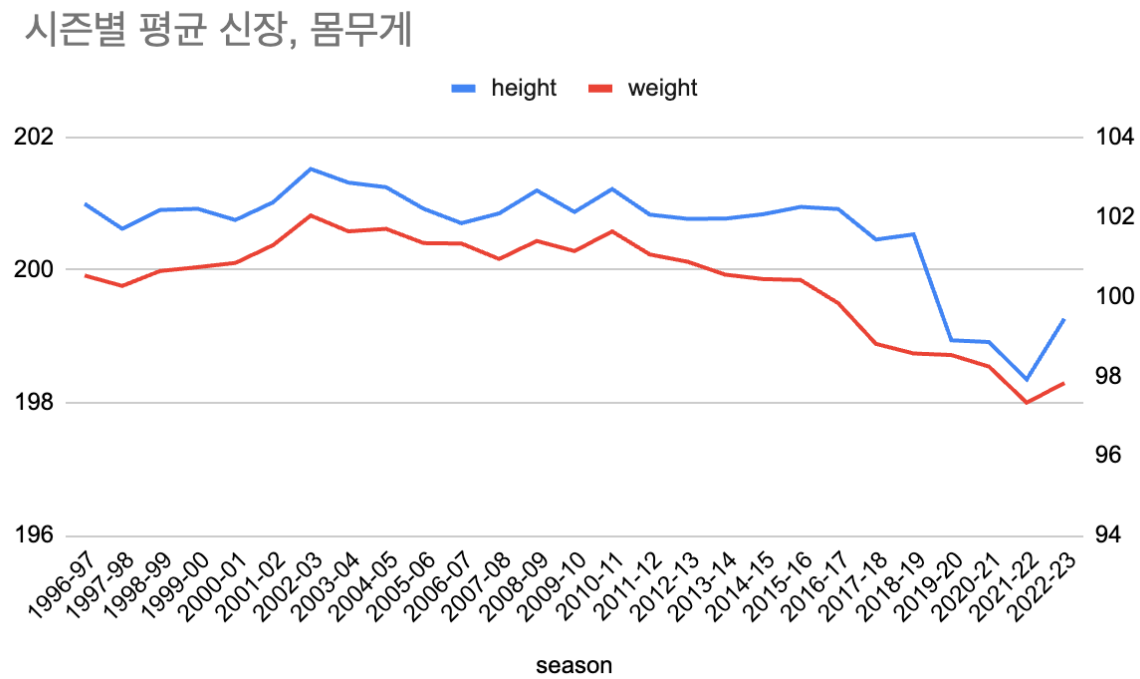
### 3. NBA Players

### 3. 동명이인 선수들이 존재

| player_name     | team_abbreviation | age  | player_height | player_weight      | college         | country | draft_year | draft_round | draft_number |
|-----------------|-------------------|------|---------------|--------------------|-----------------|---------|------------|-------------|--------------|
| Marcus Williams | LAC               | 21.0 | 200.66        | 93.89354399999999  | Arizona         | USA     | 2007       | 2           | 33           |
| Marcus Williams | NJN               | 22.0 | 190.5         | 92.98635999999999  | Connecticut     | USA     | 2006       | 1           | 22           |
| Marcus Williams | GSW               | 23.0 | 190.5         | 95.25432           | Connecticut     | USA     | 2006       | 1           | 22           |
| Marcus Williams | SAS               | 22.0 | 200.66        | 93.89354399999999  | Arizona         | USA     | 2007       | 2           | 33           |
| Chris Johnson   | MIN               | 27.0 | 210.82        | 95.25432           | Louisiana State | USA     | Undrafted  | Undrafted   | Undrafted    |
| Chris Johnson   | MEM               | 23.0 | 198.12        | 91.171992          | Dayton          | USA     | Undrafted  | Undrafted   | Undrafted    |
| Tony Mitchell   | MIL               | 24.0 | 198.12        | 97.975872          | Alabama         | USA     | Undrafted  | Undrafted   | Undrafted    |
| Tony Mitchell   | DET               | 22.0 | 203.2         | 106.59411999999999 | North Texas     | USA     | 2013       | 2           | 37           |

### 3. NBA Players

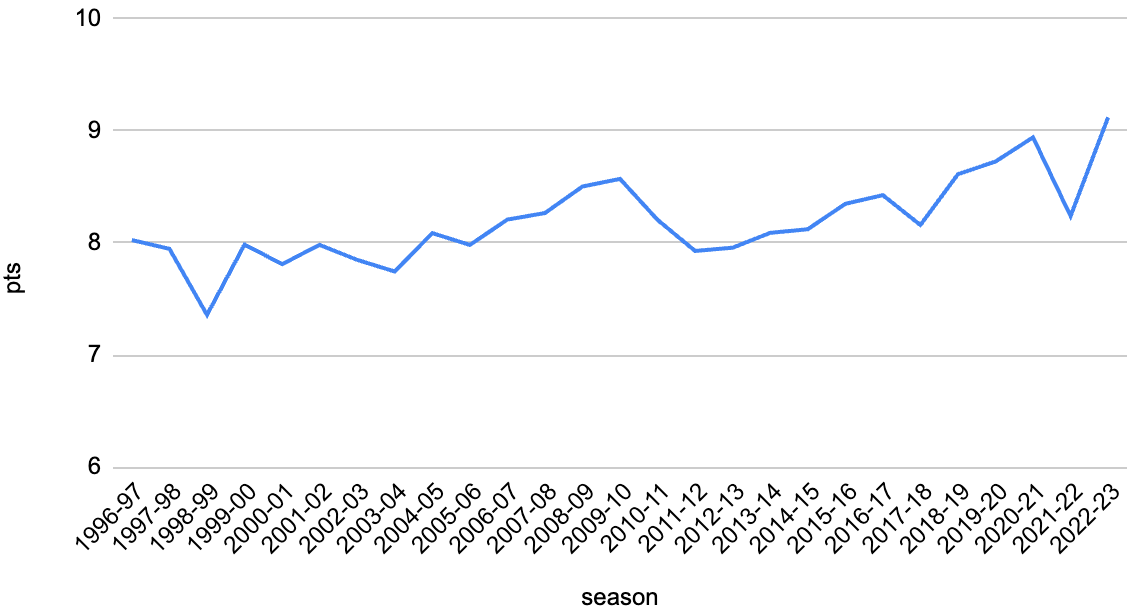
#### 4. 평균 신장과 몸무게는 감소 추세, 시즌 내에 소화하는 게임 수는 증가 추세



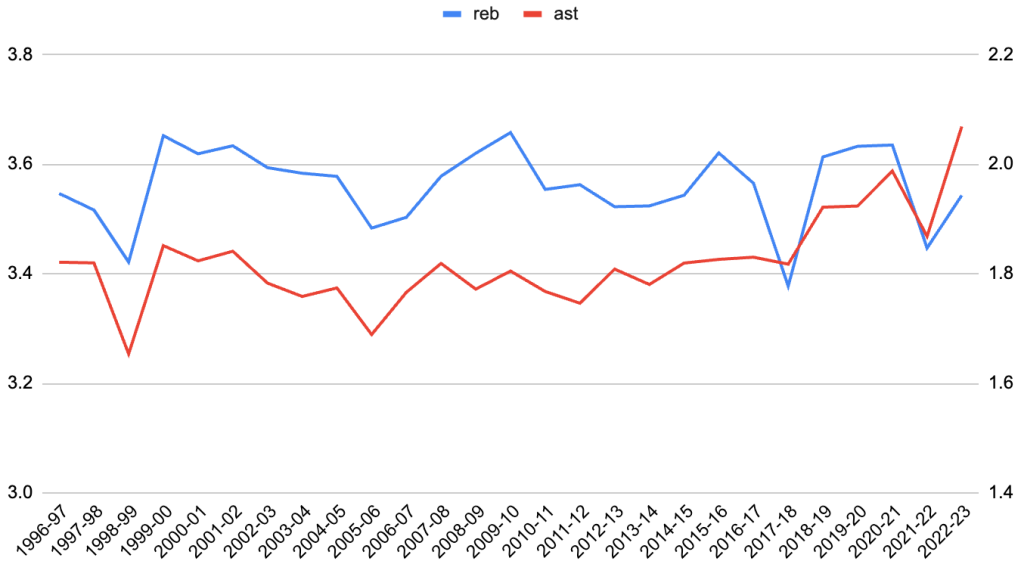
### 3. NBA Players

## 5. 평균 득점과 어시스트는 증가 추세, 리바운드는 유지

시즌별 선수들의 평균 득점



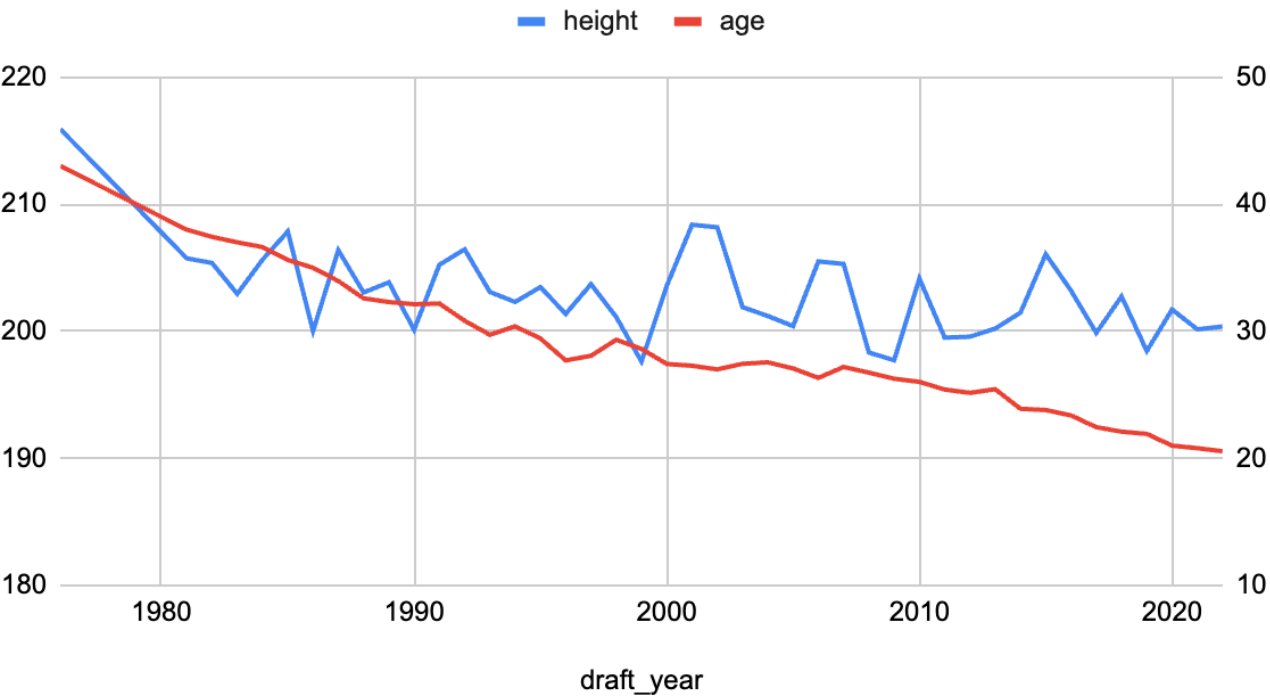
시즌별 평균 리바운드, 어시스트





6. 드래프트 1라운드, 10순위 이내에 뽑힌 선수들의 키와 나이 추세는 감소

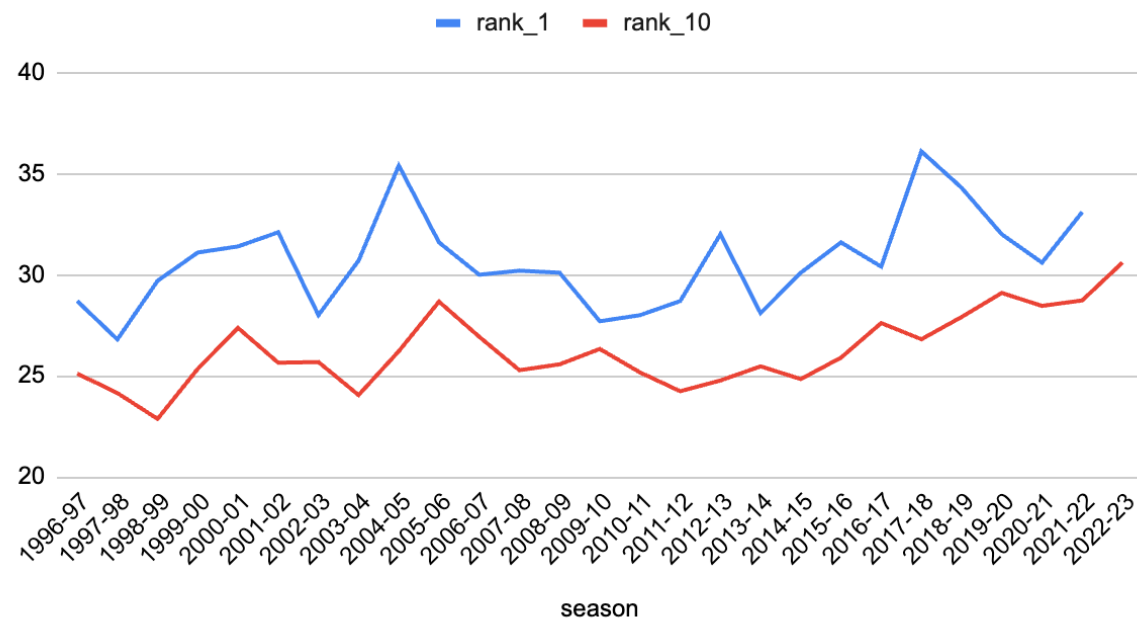
연도별 상위 드래프트 선수들의 키와 나이



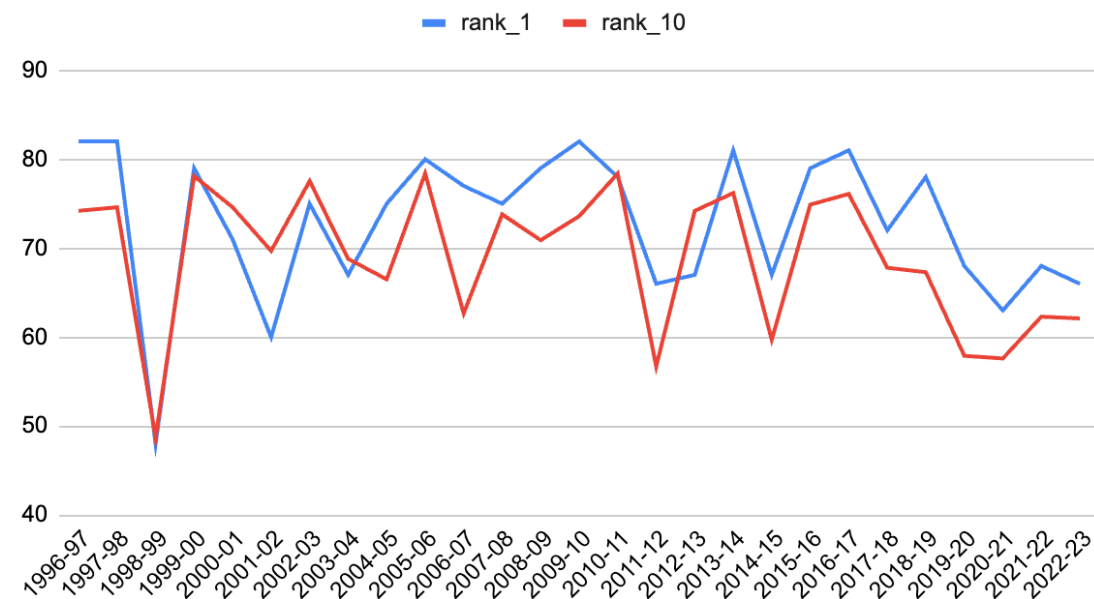
### 3. NBA Players

## 7. 시즌별 10위 이내 선수들의 평균 득점

시즌별 득점 1위 / 10위 이내 선수들의 평균 득점



시즌별 10위 이내 선수들의 평균 시합 수



**End of Document**