

데이터 과학 직업 연봉

본 분석의 목적은 데이터 과학(Data Science) 직무의 연봉 수준에 영향을 미치는 주요 요인을 파악하고, 이를 통해 다음과 같은 실무적인 인사이트를 도출하는 것입니다.

```
In [1]: import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import pandas as pd
import koreanize_matplotlib
```

```
In [6]: df = pd.read_csv("ds_salaries.csv", index_col = 0 )
df.head()
```

```
Out[6]:
```

	work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employee_residence	remote_ratio	company_size
0	2020	MI	FT	Data Scientist	70000	EUR	79833	DE	0	1
1	2020	SE	FT	Machine Learning Scientist	260000	USD	260000	JP	0	2
2	2020	SE	FT	Big Data Engineer	85000	GBP	109024	GB	50	3
3	2020	MI	FT	Product Data Analyst	20000	USD	20000	HN	0	4
4	2020	SE	FT	Machine Learning Engineer	150000	USD	150000	US	50	5

(1) 4 문제를 풀어보시오

- df 데이터프레임의 상위 5개 행을 출력하세요.
- 각 열의 데이터 타입과 결측치 여부를 확인하세요.
- experience_level 열의 고유한 값과 각 값의 개수를 확인하세요.
- salary_in_usd 열의 기술 통계량(평균, 중앙값, 표준편차 등)을 확인하세요.

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

(2) experience_level 열의 값을 'EN' (Entry-level), 'MI' (Mid-level), 'SE' (Senior-level), 'EX' (Executive-level) 값으로 바꾸시오.

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

(3) companize_siz "S"에서 직업 중에서 가장 많이 연봉 받는 Top 5 직업을 출력하시요. (salary_ind_usd)

```
In [ ]:
```

```
In [ ]:
```

(3) 연도(work_year)별 salary_in_usd의 평균을 계산하고, 이를 라인 플롯으로 시각화하여 연봉 트렌드를 보여주세요.
(x축: work_year, y축: 평균 salary_in_usd)

```
In [ ]:
```

(3) 경험 수준(experience_level)별 salary_in_usd의 분포를 상자 그림(boxplot)으로 시각화하여 경험 수준에 따른 연봉 차이를 비교하세요.

```
In [ ]:
```

(4) 회사 규모(company_size)별 salary_in_usd의 평균을 막대 그래프(bar plot)로 시각화하고, 각 막대 위에 평균값을 표시하세요.

In []:

(5) 재택근무 100%, 연봉 10만 이상 받는 직업은 무엇인가요?

In []:

(6) Analyst, Scientist 관련 데이터 필터하고, Entry-level에서 salary_ind_usd가 가장 높은 10가지 직업을 출력하시요.

In []:

(7) 가장 많이 고용된 직무 & 평균 연봉을 구하시요.

In []:

(8) 연봉 상위 10%에 속하는 직무들은 무엇인가요?

In []:

In []: