

직원 이직 여부 데이터 분석

IBM HR Analytics Employee Attrition & Performance (Kaggle)

박유진, 서승우, 송정우

Table of Contents

01	02	03	04	05
				느낀점
	이변량 분석	추가 분석	관계 정리	
가설수립				



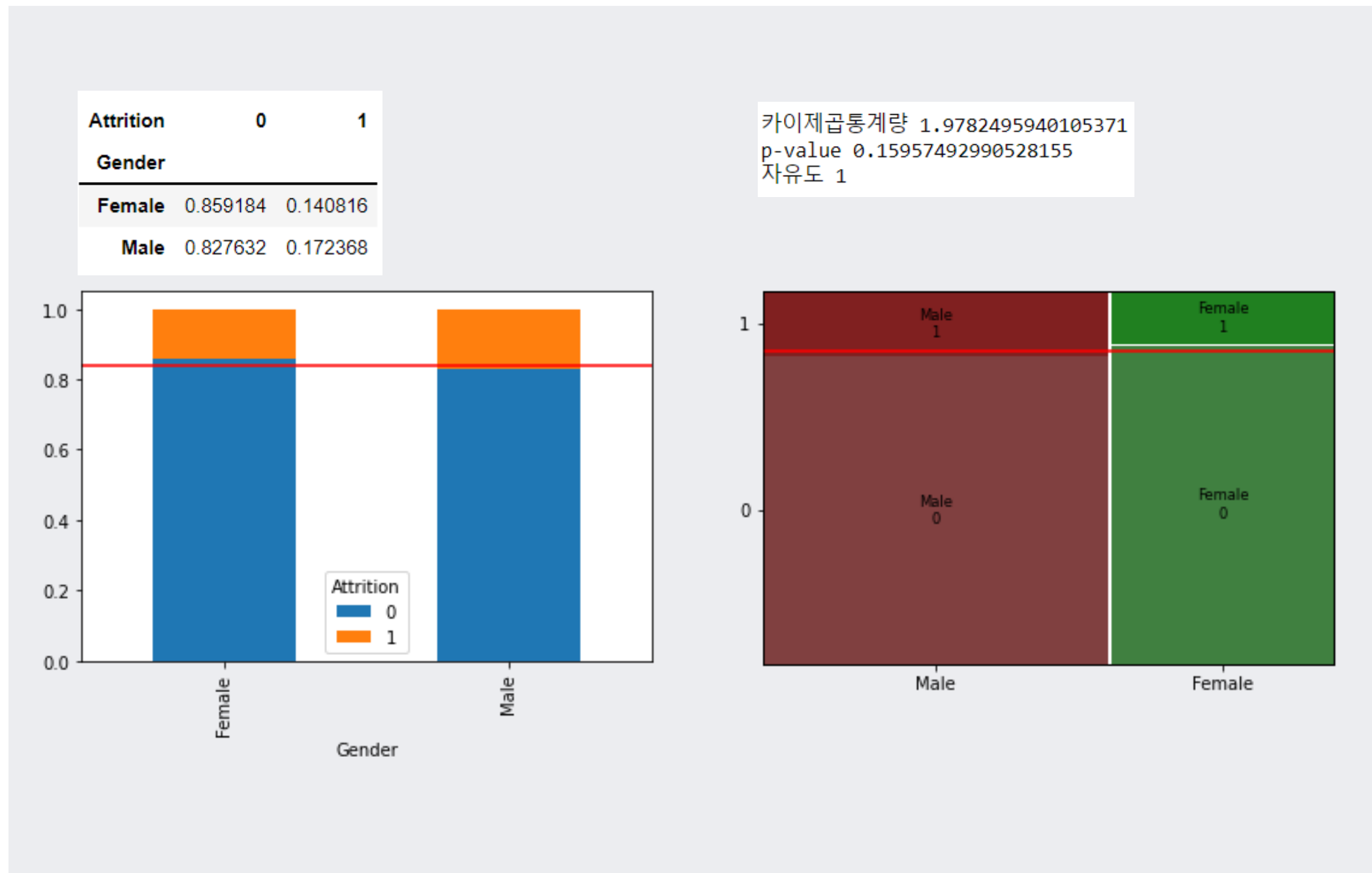
01 가설수립

01 가설 수립

1. Gender에 따라 Attrition에 차이가 있다.
2. MaritalStatus에 따라 Attrition에 차이가 있다.
3. JobSatisfaction에 따라 Attrition에 차이가 있다.
4. Age에 따라 Attrition에 차이가 있다.
5. DistanceFromHome에 따라 Attrition에 차이가 있다.
6. MonthlyIncome에 따라 Attrition에 차이가 있다.

02 이변량 분석

1. Gender --> Attrition



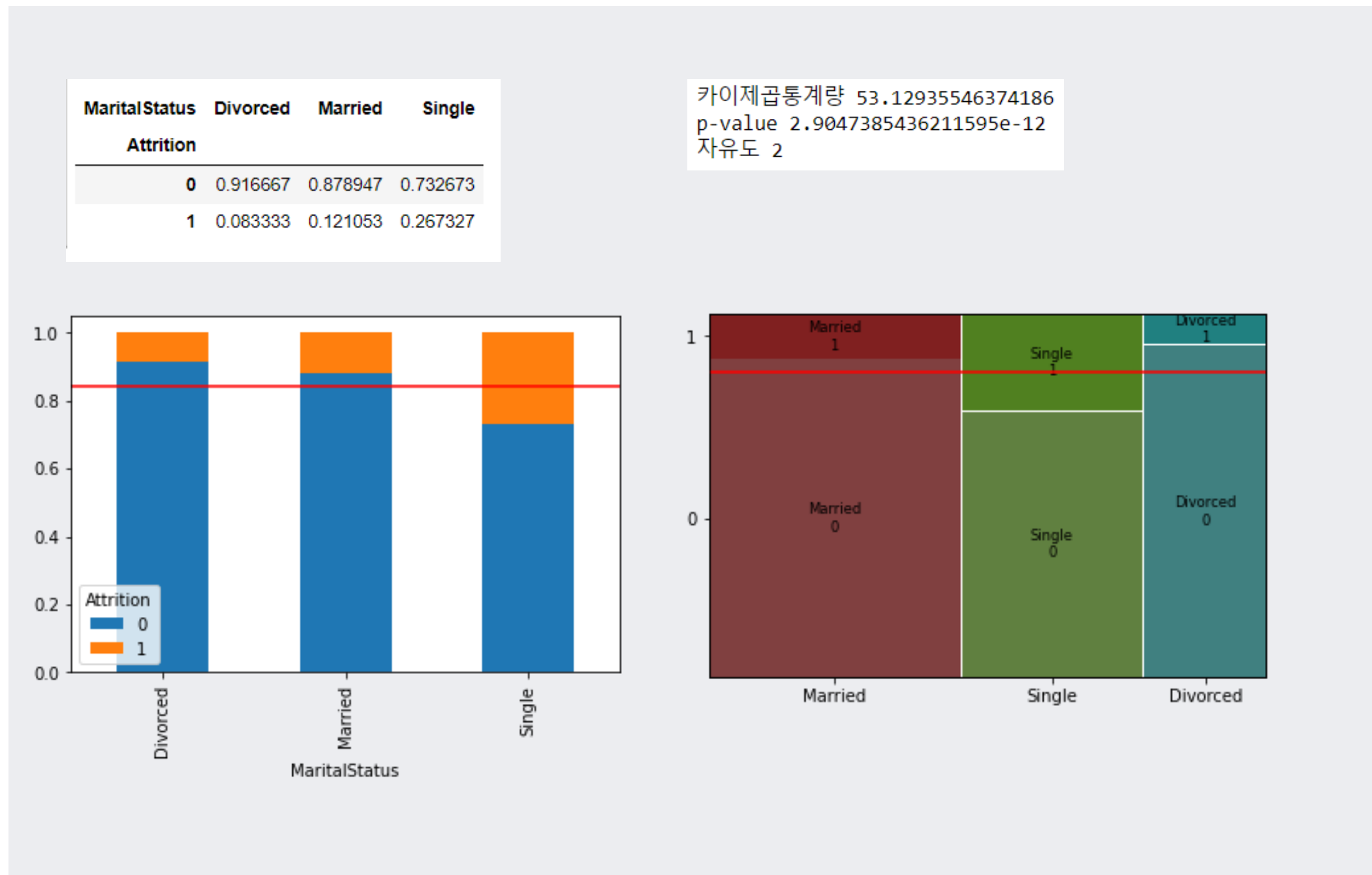
분석 결과

- 성별 이직 여부의 비율이 거의 같다
- 성별과 이직 여부 사이에 큰 연관성이 없다

추가 분석

- 여성과 남성의 비율이 약 4:6의 비율이므로, 같은 성비의 데이터로 추가 분석할 필요가 있다.

2. MaritalStatus(결혼여부) --> Attrition



분석 결과

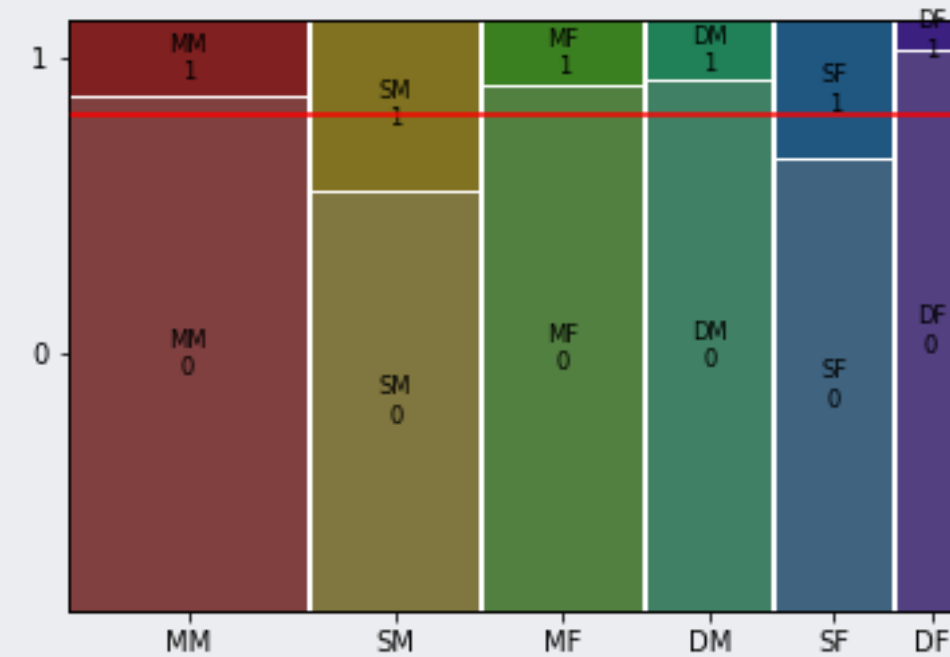
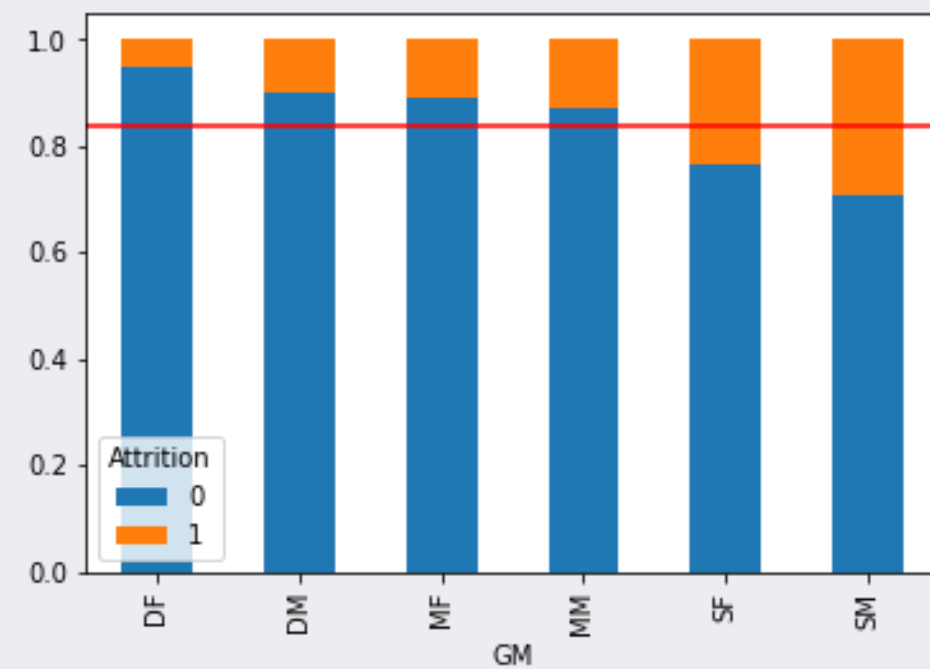
- 미혼들의 이직률이 전체 평균 이직률보다 높다.
- 그리고 기혼인 사람들의 비중이 가장 크다.
- 기혼한 사람과 이혼한 사람은 전체 평균 이직률보다 이직률이 더 낮은 편이다.
- 결혼 상태에 따라 이직 여부의 차이가 어느 정도 관계가 있어 보인다.

추가 분석

- 성별에 따른 결혼에 대한 인식 차이가 있을 수 있기 때문에 따라서 결혼여부를 분석해볼 필요가 있어 보인다.
- 미국의 결혼 문화에 대한 분석이 필요하다.

2. MaritalStatus(결혼여부) --> Attrition

카이제곱통계량 56.81229721285164
p-value 5.528072204832141e-11
자유도 5



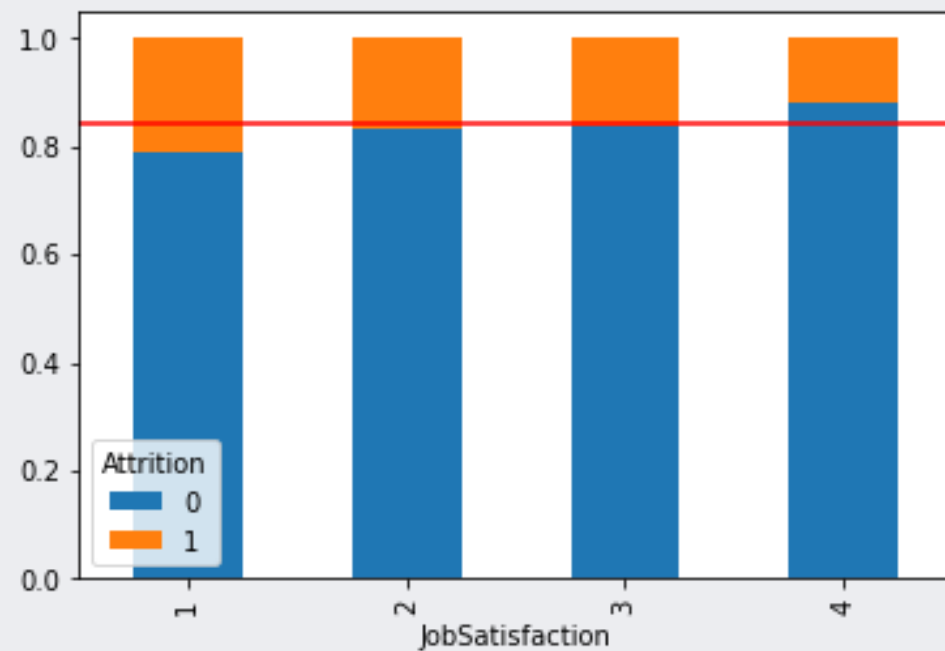
분석 결과

- 성별과 결혼 상태를 묶어 이직 여부를 분석하니 범주간의 이직 여부의 차이가 생겼다.
- 싱글인 남성과 싱글인 여성의 이직율이 높은 편이며, 이혼한 여성의 이직율이 가장 낮은 편이다.

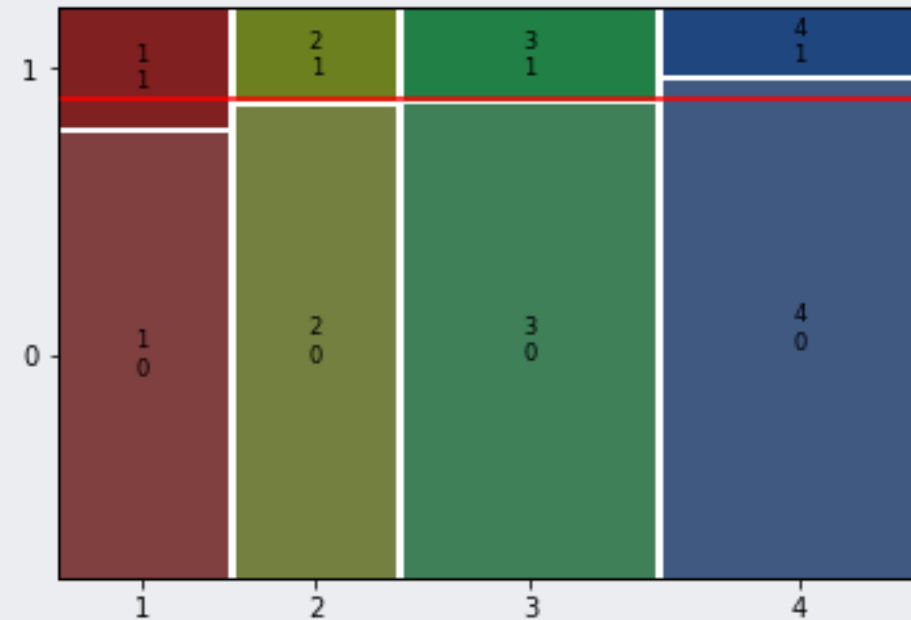
3. JobSatisfaction--> Attrition

JobSatisfaction(직업 만족도) : 1 (Low), 2 (Medium), 3 (High), 4 (Very High)

JobSatisfaction	1	2	3	4
Attrition				
0	0.788	0.834043	0.836957	0.879093
1	0.212	0.165957	0.163043	0.120907



카이제곱통계량 9.63151465436614
p-value 0.021972649369306214
자유도 3



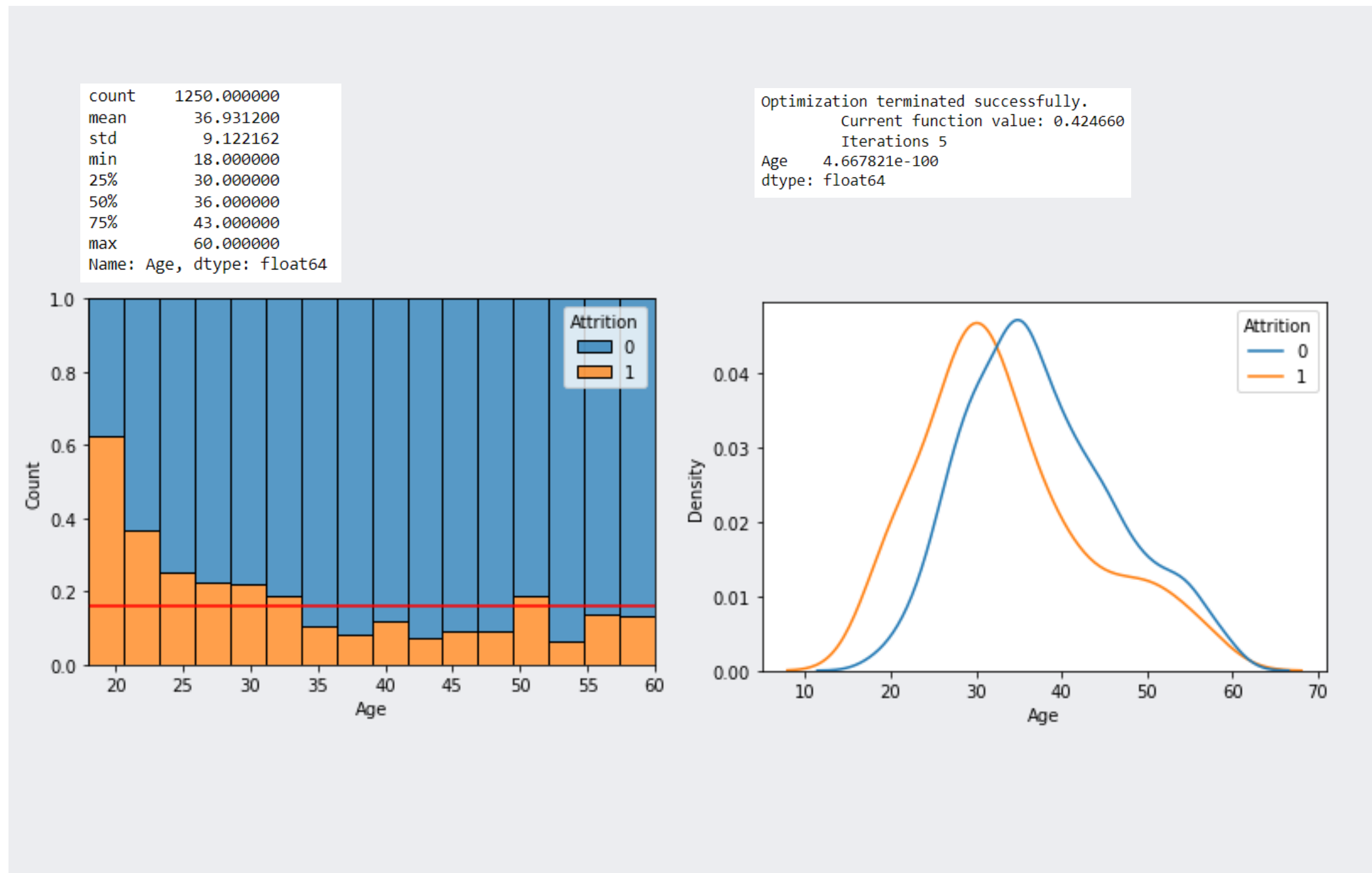
분석 결과

- 만족도 2,3을 표시한 직원들은 평균에 가깝고, 1,4를 표시한 이들의 이직 여부는 평균과 차이를 보인다
- 만족도 1을 표시한 직원들의 이직 여부는 평균보다 높고, 4를 표시한 직원은 낮다
- P-value는 0.021, 카이제곱통계량은 9.631이다. (자유도 3)
- 만족도 조사결과와 이직 여부 사이에 큰 연관이 없다.

추가 분석

- 설문조사 결과는 그 방식과 질문에 따라 큰 차이를 보인다. 따라서 조사 방법에 대한 확인이 필요하다.

4. Age--> Attrition

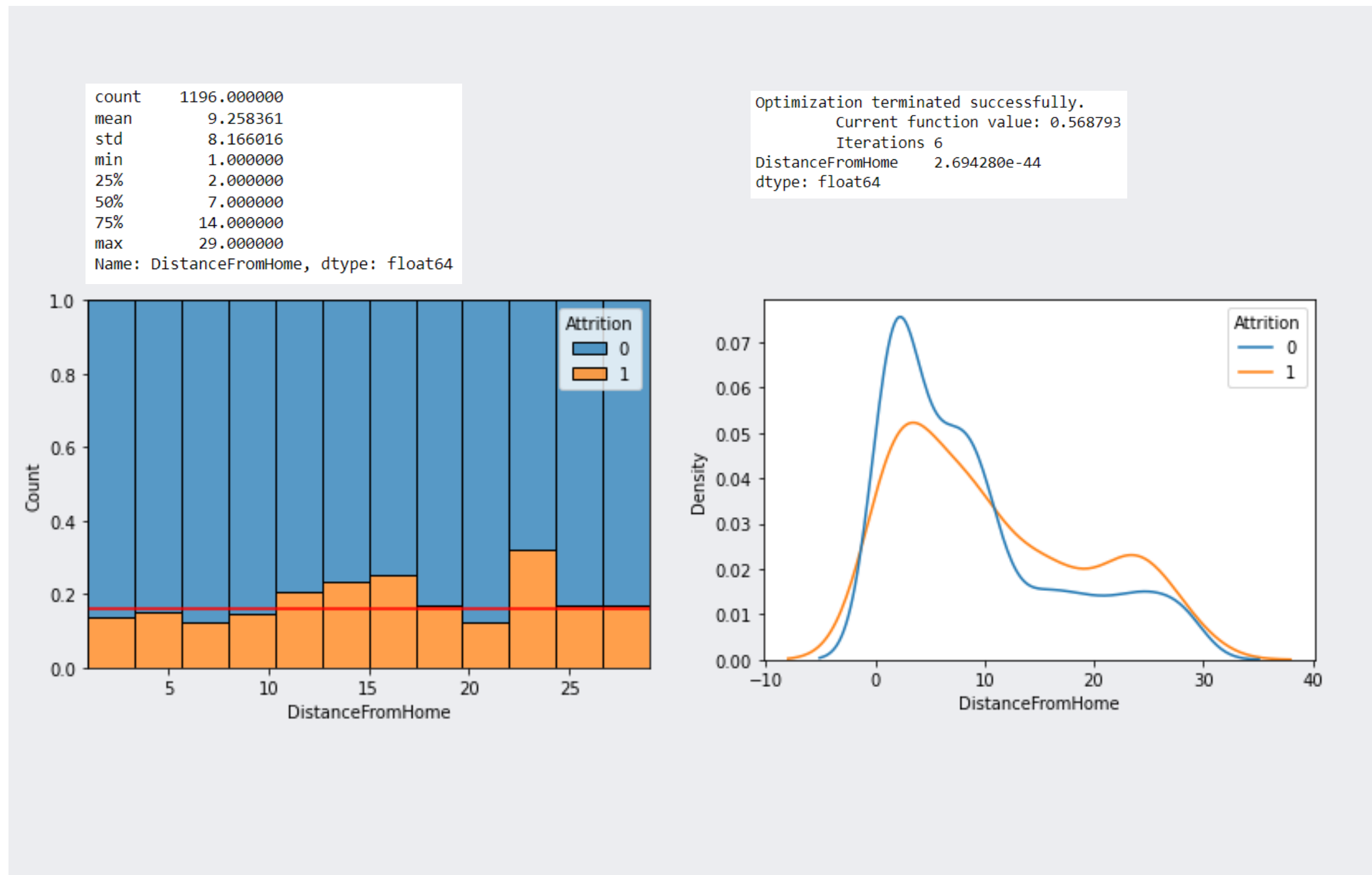


분석 결과

- 32세 전까지의 이직률이 높은 편이다. 보통 미혼일 가능성이 높으며, 또 나이가 젊기 때문에 이직도 더 많이 할 가능성이 높다.
- 50세 초반의 나이에서도 이직률이 높은데, 보통 대기업에서 40대에 승진에서 누락되는 사람들이거나 새로운 분야로 진출하려는 사람들 혹은 창업자들일 확률이 높다.
- 나이와 이직 여부 간의 관계가 강해 보인다.

추가 분석

5. DistanceFromHome--> Attrition



분석 결과

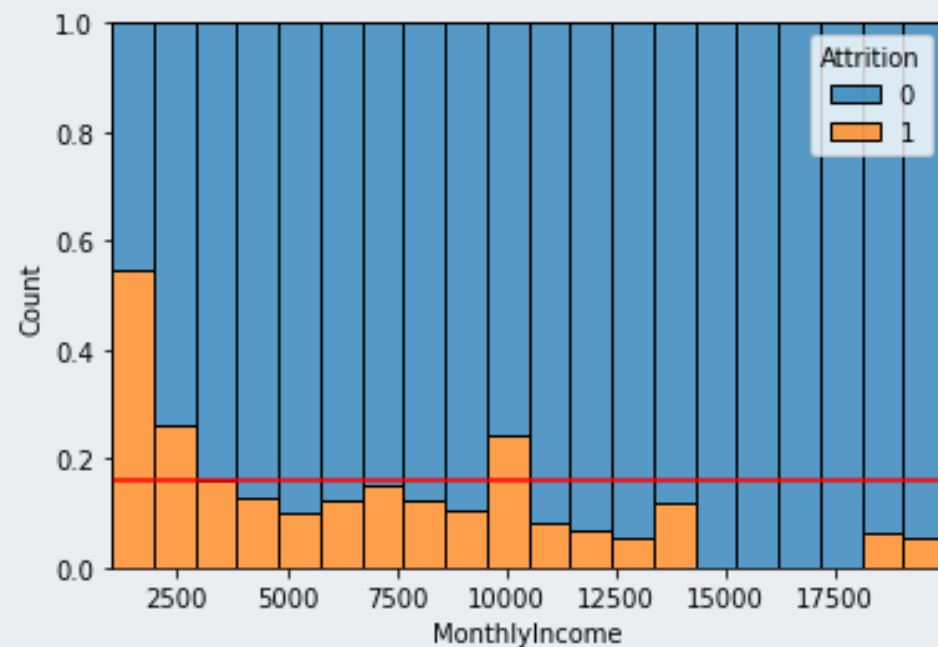
- 거리가 12마일 이상부터 대체적으로 이직률이 높다가, 18-29 마일에서는 이직률이 낮다.
- 10마일 이내에 대부분의 데이터가 몰려 있다. 회사에서 주거 복지에 신경을 쓰는 것으로 볼 수 있다(기숙사 제공)
- 혹은 단순히 출퇴근을 편히 하기 위해 20분 거리내에 주거하고 있는 것을 의미할 수도 있다.

추가 분석

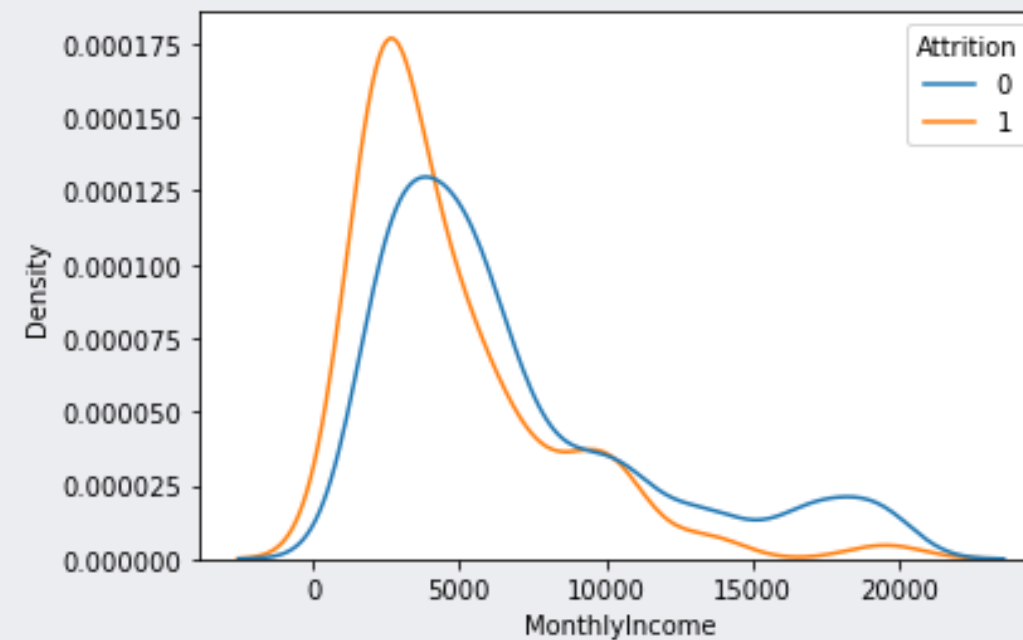
- 거리별로 차이가 있는 구간을 나눠 분석한다면 더 큰 차이가 있을 것으로 보인다.

6. MonthlyIncome--> Attrition

```
count    1250.000000
mean     6525.548000
std      4676.979554
min      1009.000000
25%      2932.250000
50%      4988.000000
75%      8437.500000
max      19999.000000
Name: MonthlyIncome, dtype: float64
```



```
Optimization terminated successfully.
Current function value: 0.437401
Iterations: 6
MonthlyIncome    1.922717e-71
dtype: float64
```



분석 결과

- 최소값이 1009달러인데, 이 구간에서 이직율이 높은 것으로 보아 비정규직 직원 구간이라고 예상해볼 수 있다.
- 대략 4,000달러를 기준으로 이하일 때는 전체 평균 비율보다 이직율이 더 높고, 이상일 때는 이직율이 더 낮다.
- 10,000달러 부근에서 이직율이 평균보다 높아지는데, 왜 그런지에 대해 알아볼 필요가 있다.
 - 비즈니스 상황을 살펴보면, 퇴직금 등의 이유를 고려해 볼 수 있다.
- 약 14,000~18,000달러 구간을 보면 이직률이 0이다. 왜 그런지에 대해 파악해볼 필요가 있다.
- 이후 구간에서 다시 이직률이 발생하는데, 이는 이사 등 임원들의 퇴직 등과 관련된 데이터일 것으로 예상 된다.
- 로지스틱 모델을 통해 p-value를 파악해보면 0에 가깝다.
 - 이는 미봉책이긴 하지만, 시각화와 수치화를 종합해보면 Age와, Attrition은 강한 관계라고 볼 수 있다.

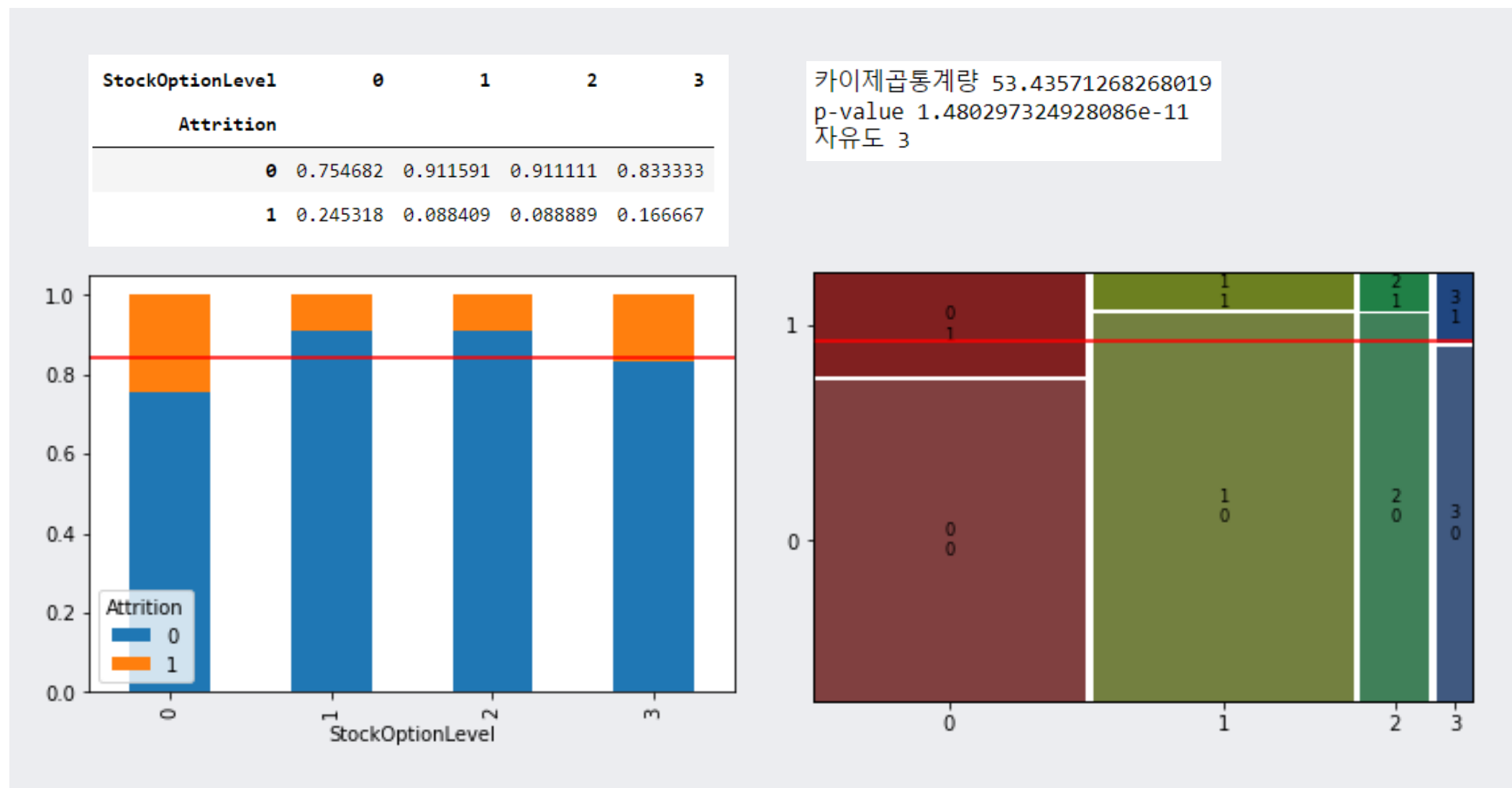


03 추가 가설수립

1. StockOptionLevel에 따라 Attrition에 차이가 있다.
2. YearsWithCurrManager에 따라 Attrition에 차이가 있다.
3. EnvironmentSatisfaction에 따라 Attrition에 차이가 있다.

03 추가 가설수립

1. StockOptionLevel--> Attrition

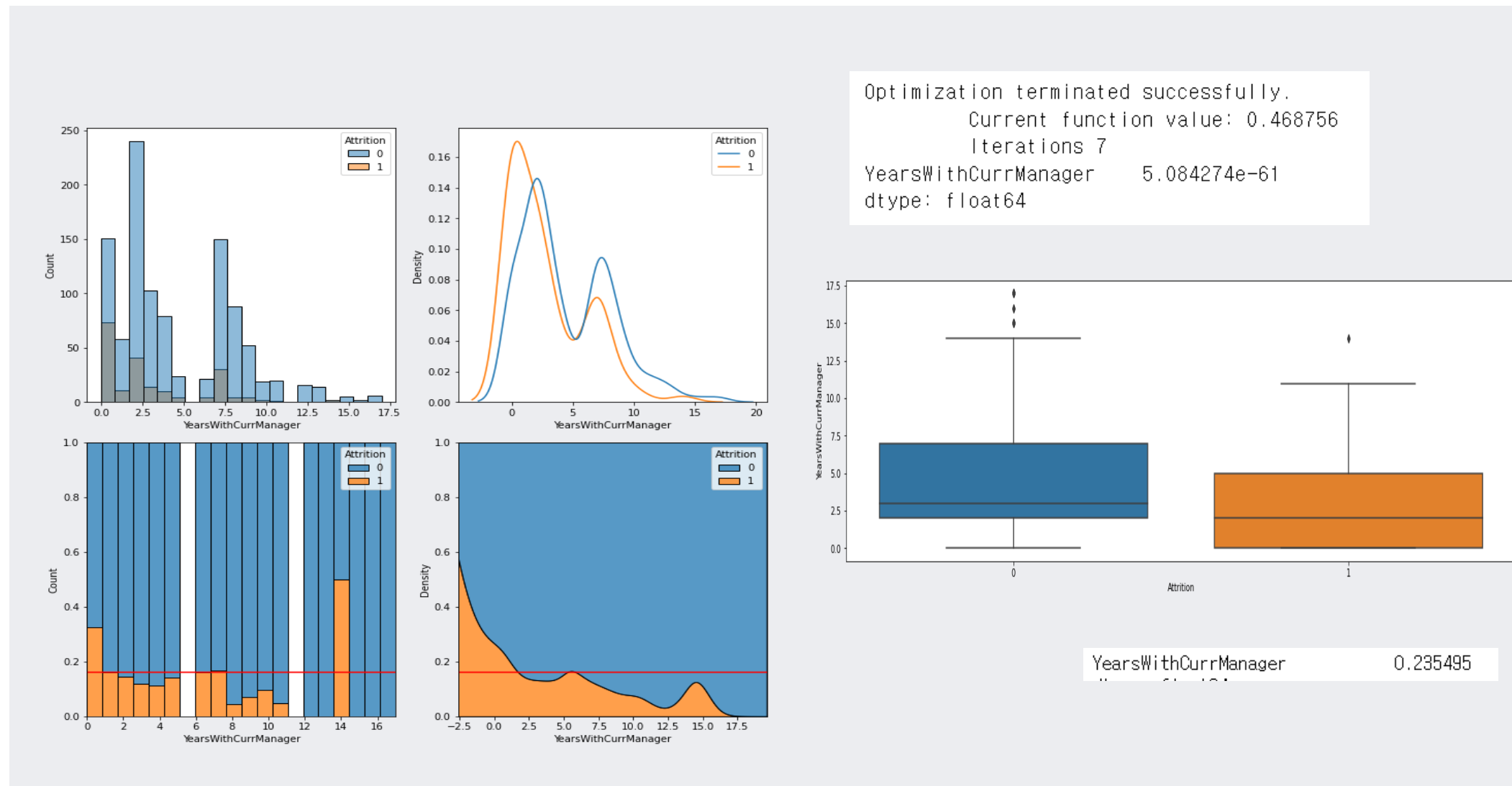


분석 결과

- 시각화를 통해 StockOptionLevel이 0일 때 이직률이 평균보다 높음을 확인할 수 있다.
- 카이제곱통계량은 약 53.4로 자유도(3)의 약 18배 크고, p-value는 0에 수렴한다.
 - 따라서 StockOptionLevel과 Attrition은 강한 관계를 갖는다고 볼 수 있다.

2. YearsWithCurrManager--> Attrition

YearsWithCurrManager(현재 상사와의 근무 연수)



분석 결과

- 상사와의 근무 연수가 2년 이내인 사람들의 이직율이 매우 높은 편이다.
- 매우 관계가 좋지 않더라도 경력을 쌓기 위해 2년 정도는 근무를 계속하는 것으로 보인다.

추가 분석

이직여부별 상사와의 근무 연수

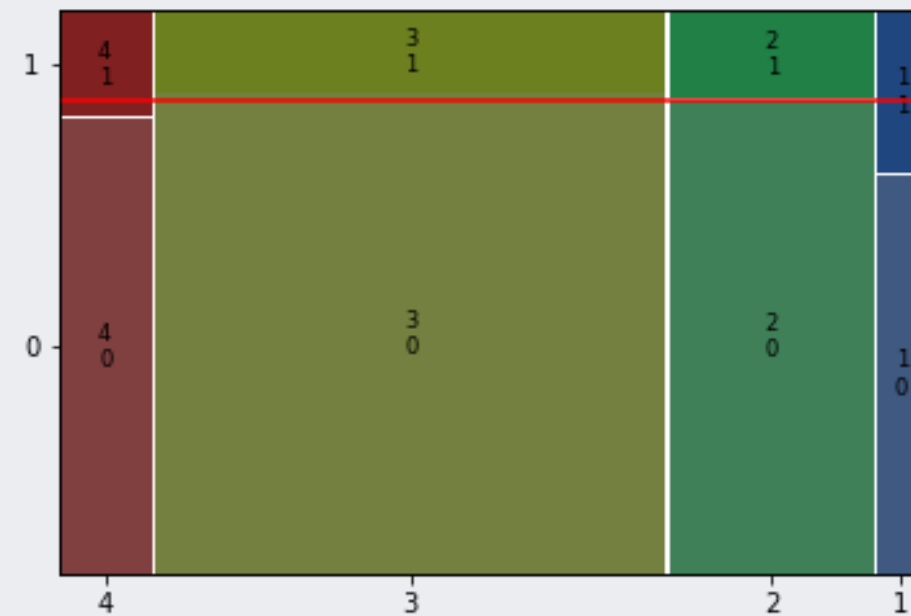
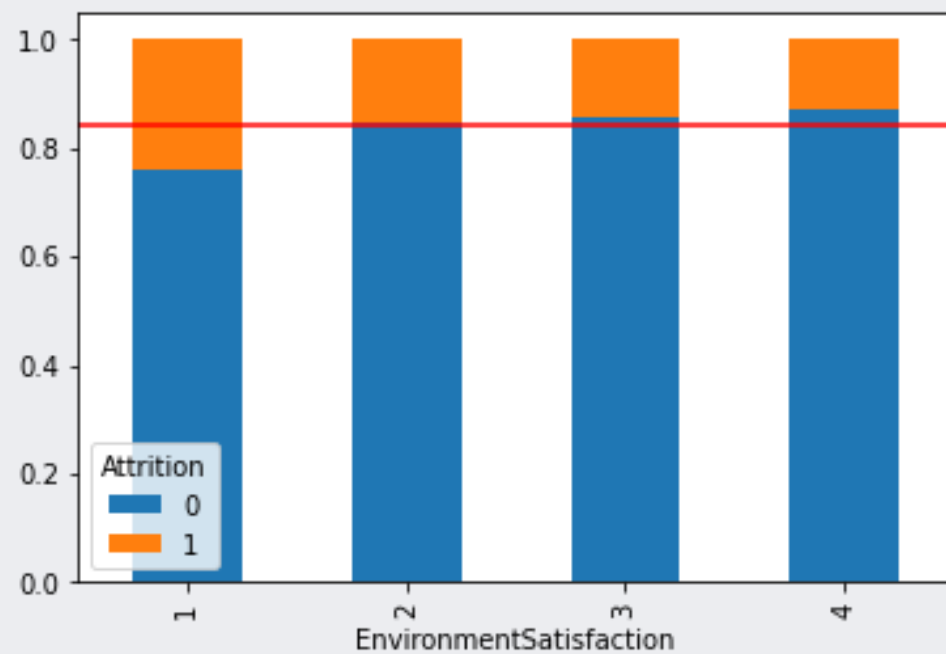
- 이직을 한 사람들의 절반 정도가 2.5년만 근무를 하다가 이직을 한다.
- 상사와 근무연수가 2년 이내인 사람들의 근무 연수 평균은 상사와 3개월도 안되는 것을 알 수 있다.

3. EnvironmentSatisfaction--> Attrition

EnvironmentSatisfaction(환경 만족도) : 1 (Low), 2 (Medium), 3 (High), 4 (Very High)

Attrition		0	1
EnvironmentSatisfaction			
1	0.759336	0.240664	
2	0.846154	0.153846	
3	0.857143	0.142857	
4	0.869231	0.130769	

카이제곱통계량 : 15.054623429327041
p-value : 0.0017705594620230482
자유도: 3



분석 결과

- 환경 만족도가 증가함에 따라, 이직 여부의 비율이 줄어든다.
- P-value는 0.00177, 카이제곱통계량은 15.054이다.
- 환경 만족도와 이직 여부에 관련은 있으나, 강하지 않다.

추가 분석

- 설문조사 결과는 그 방식과 질문에 따라 큰 차이를 보인다.
따라서 조사 방법에 대한 확인이 필요하다.
- 근무 환경은 수시로 변할 수 있기 때문에
그 기준에 대한 명시가 필요하다.



04 관계정리

관계 없음	중간 관계	강한 관계
Gender	JobSatisfaction DistanceFromHome WorkLifeBalance	MaritalStatus Age MonthlyIncome StockOptionLevel YearsWithCurrManager



05 느낀점

- 모든 변수를 해보진 않아서 정확성은 떨어질지 몰라도, 데이터 인사이트를 뽑아 수업시간엔 못해본 추가분석까지 해본 점이 좋았다.
- 데이터를 분석하는데 있어서 직관적인 느낌과 수치로의 결과는 차이가 있을 수 있다.
- 수업 시간에 사용되지 않은 다른 요인들을 분석해보니, 이직 여부에 차이를 주는 요인들이 많이 있었다. 데이터 분석 기법을 더 많이 익혀서 캐글 대회에 나가보고 싶다.



Thank you