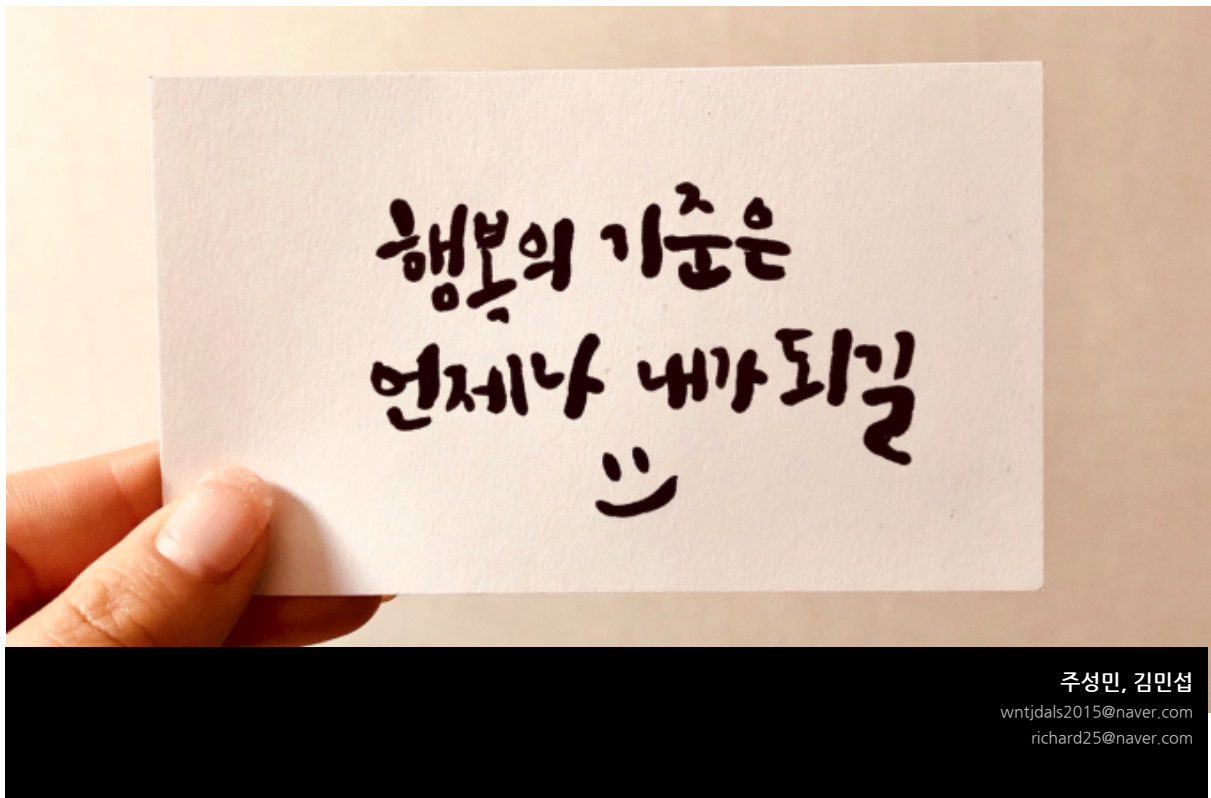


# 결혼과 행복의 기준에 관한 데이터 분석



출처 - <https://brunch.co.kr/@zannymongroad/86>

우리는 분석을 진행하기 전에 한가지 가설을 설정하려고 한다.

**‘배우자가 있는 집단은 사랑과 관련된 상황에서 배우자가 없는 집단보다 더 많이 행복감을 느낀다. 그리고 배우자가 없는 집단은 배우자가 있는 집단보다 성취와 관련된 상황에서 더 많은 행복감을 느낀다.’**

우리가 이러한 가설을 설정한 이유는 배우자가 있다면 사랑에 관해서 행복한 순간이 더 많이 올것 같다는 생각이 들었다. 20대의 우리는 아직 결혼을 하지 않았다. 물론 연애도 하며 이별도 해봤지만 ‘결혼한 후에 사랑이라는 것에서 행복감을 많이 느낄까?’ 라는 질문에 답을 구하기 위해서 위 가설을 설정하였다. 이에 대해서 우리는 결혼유무, 남녀별, 연령대별로 나누어서 분석을 진행해 보았다.

### - 분석에 사용된 데이터

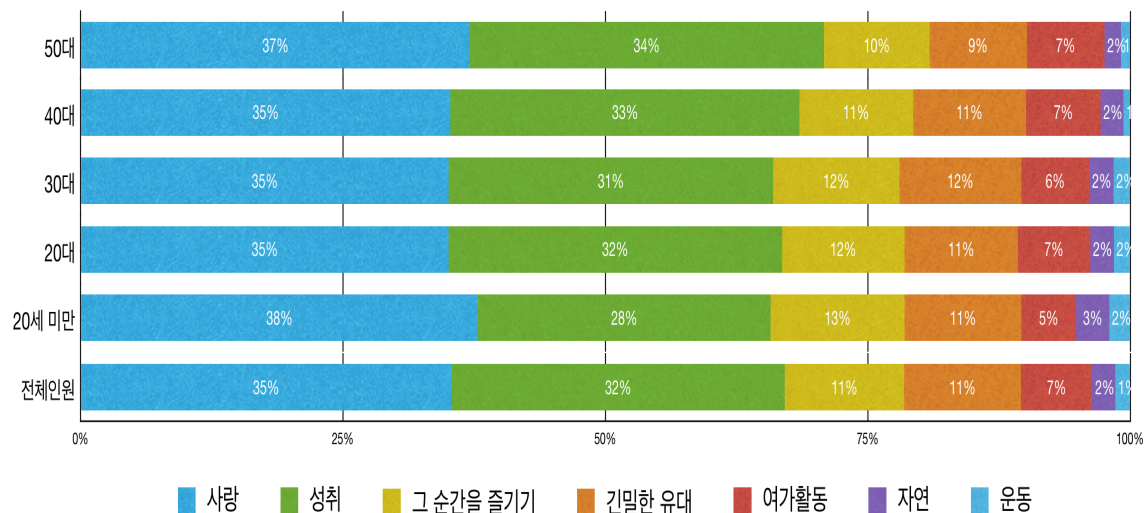
분석에 사용된 데이터 베이스는 <https://www.kaggle.com> 에서 제공하는 'HappyDB', 60,000개 정도의 행복한 순간과 행복의 기준을 보여주는 데이터 이다. 아마존의 "Mechanical Trunk"(글로벌 고객 조사를 위한 유용한 도구)를통해, 한번의 자잘한 오류들을 정리해서 제공된다.

남성	5311명
여성	5445명
무응답	88명
합계	10844명

우리가 분석하기위한 자료는 100,000개 었지만 그중 약 9만개의 데이터들이 익명으로 제공된 데이터였다. 나이와 성별등을 기준으로 미혼 과 기혼을 나누어야 하기 때문에 개인정보가 주어진 데이터(약 11,000개의 표본)만 사용했다.

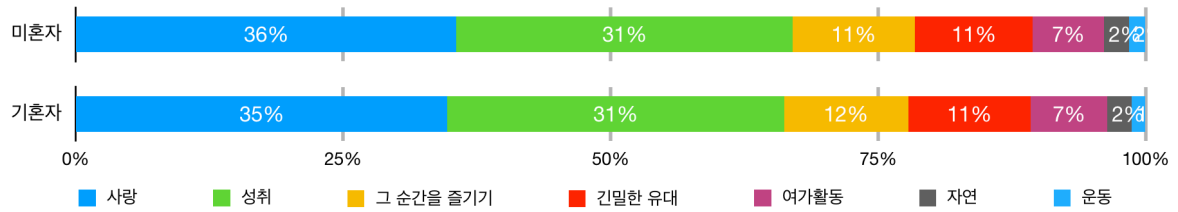
사용한 데이터에서 '국가' 라는 데이터 항목은 제외를시켰다. 표본 데이터 의 85%정도가 미국이기 때문에 국가별로 나누는건 좋은 분석이 아니라고 판단하였기 때문이다. '나이' 항목의 표본 중 20대 미만과 50대 이상의 데이터 표본들이 다른 연령대에비해 적었기 때문에 연령대를 나누는 기준은 20대 미만, 20대, 30대, 40대, 50대 이상으로 설정했다.

### - 일반적인 연령대에 따른 '행복을 느끼는 상황'의 비율



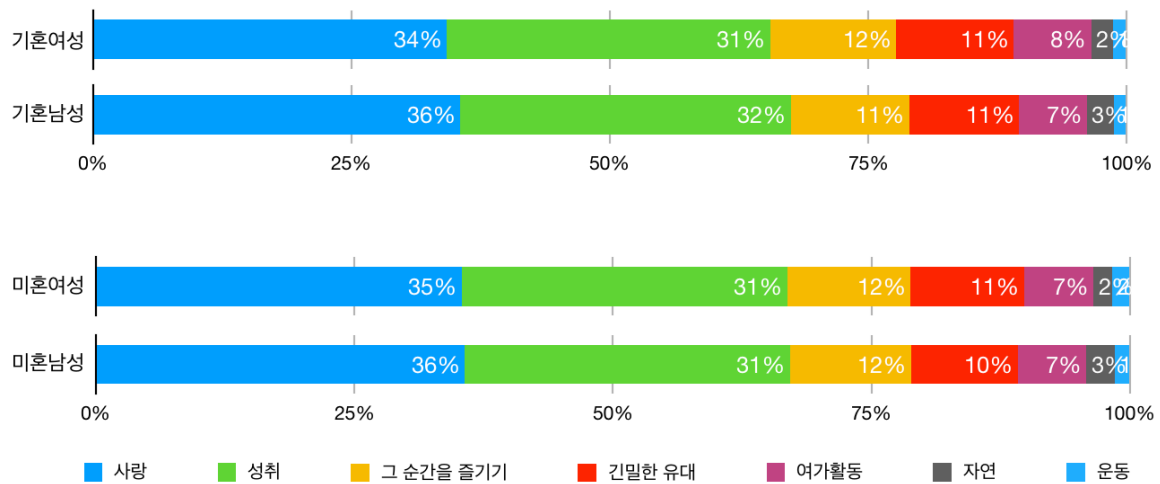
연령대별로는 전체적인 '행복을 느끼는 상황'의 분포도가 거의 차이가 없어보이지만 20대 이후 연령대에서 '운동'이라는 항목은 점점 줄어드는게 보인다. 어느정도 예측했던 결과였지만, 20세 미만이 다른 연령대에 비해 '자연'항목이 높게 측정된건 예측과 정 반대였다. 아마 여행때문일 것 같다고 추측해본다. 확실히 알아볼 수 있는건 어느 연령대이든 '사랑' 과 '성취'의 감정을 느끼는 상황에서 행복하다고 느끼고 있다는 것이다.

#### - 전체적인 결혼 유무에 대한 분석



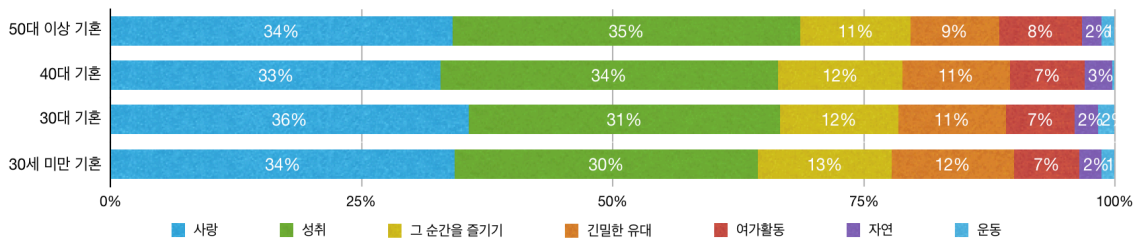
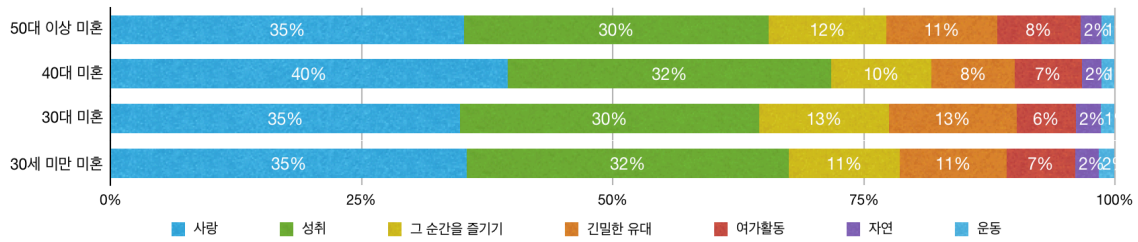
전체적인 분석으로 보았을때에 미혼자와 기혼자의 차이는 그리 많지 않아 보인다. 오히려 미혼자가 기혼자보다 사랑이라는 감정에서 1% 정도의 많은 수치를 보이고 있다. 반면 기혼자는 미혼자에 비해서 '그 순간을 즐기'라는 순간에서 약간 많은 수치를 보여주고 있다. 하지만 이러한 미미한 차이로는 정확하게 말할 수 없을 것 같았기 때문에 좀더 세분화를 시작했다.

#### - 성별에 따른 결혼 유무에 대한 분석



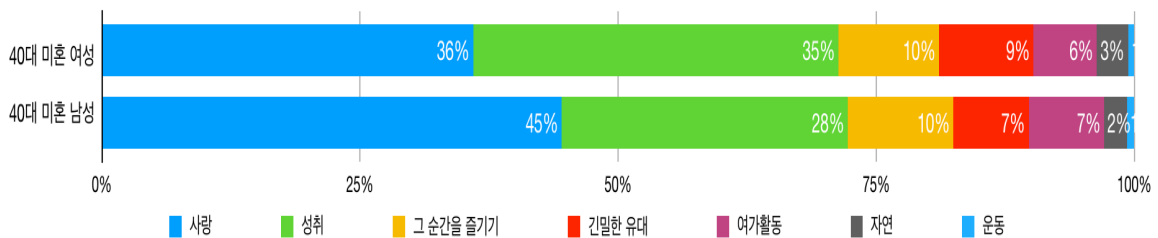
성별을 나눈 후에 결혼 유무를 분류하여 나타내어 보였다. 아직까지도 결혼 유무가 크게 관련이 없어 보인다. 연령대를 나눠 좀더 세분화 해 보도록 하자.

#### - 연령대에 따른 결혼 유무에 대한 분석



연령대에 따른 분석을 진행해 보았더니 기존의 것과는 다소 다른점이 보이기 시작했다. 40대 미혼자의 경우 '사랑'으로 행복감을 느낀 상황이 '성취'로 행복감을 느낀 상황보다 월등히 많아졌다. 30대 이후의 미혼자는 '사랑'이 '성취'보다 평균 6%의 격차를 보이고 있다. 이에 반해 30대 이상의 기혼자의 경우 '사랑'이 '성취'보다 약 1.3% 높을뿐이다. 게다가 40대 이상의 기혼자부터 '사랑'보다 '성취'에서 더 많은 사람들이 행복감을 느낀다고 분석되었다. 그렇다면 40대 미혼자들은 다른 집합들보다 왜 '사랑'이라는 상황에서 행복감을 더 많이 느끼고 있는 것일까?

#### - 40대 미혼자들을 성별에 따라 분류



질문을 좀 더 구체화 하기 위해서 40대 미혼자들을 성별에 따라 다시 분류를 해 보았다. 그 결과 여성보다 남성이 더 '사랑'과 관련된 상황에서 행복감을 더 많이 느끼고 있다.

흥미로운 사실은 분석된 데이터의 수치는 최종 우리팀의 가설은 잘못된 가설이라고 보여주고 있다는 것이다. 당연히 결혼 후 배우자와의 행복한 순간이 훨씬 많았을 것이고, '사랑'이 행복의 기준이 될 것이라고 예측 했었는데 예측이 잘못되었다고 분석이 되었다. 결혼이 사랑과 관련된 행복의 기준은 전체적으로는 미미한 차이였지만, 연령대 와 성별로 더 세세하게 분석해보니 차이가 많이 나는 부분을 찾을 수 있었다.

'40대 미혼 남성이 왜 다른 그룹에 비해서 '사랑'과 관련된 상황에서 더 많은 행복감을 느끼는걸까?'이란 질문에 대해서 더 분석을 진행해보고 싶은 생각이 들었다. 현재 사용중인 데이터로는 위의 질문에 대해서 답을 하지 못하였지만 이번 'Data Science Competition 2018'을 통하여 답을 얻고싶다.

이번 과제를 진행해보니 데이터분석하는 과정이 쉽지만은 않다는 것을 느꼈다. 과제를 진행하면서 강의에서 없던 것들도 배울 수 있었으며 프로그래밍과 알고리즘 능력도 진행하기 전보다 다소 향상된것 같았다.