

시각화 리포트

물음1

목적

1. 한눈에 상관계수를 분석할 수 있도록 합니다.
2. 상관계수를 색깔별로 시각적으로 한눈에 들어오게 합니다.

효과

1. 색깔을 통해 상관계수가 높은 관계를 파악할 수 있습니다.
2. 전체적인 경향성을 파악할 수 있습니다.

개선점

1. 음의 상관관계인 것과 양의 상관관계인 것들이 잘 구별이 되지 않았습니다. 따라서 center = 0 옵션을 통해 음의 상관관계, 양의 상관관계가 잘 구별되게 했습니다.
2. 좀더 알아보기 쉽게 각 항목을 줄로 구분했습니다.

물음2

목표

1. 음성의 비중과 인기사이의 관계를 파악합니다

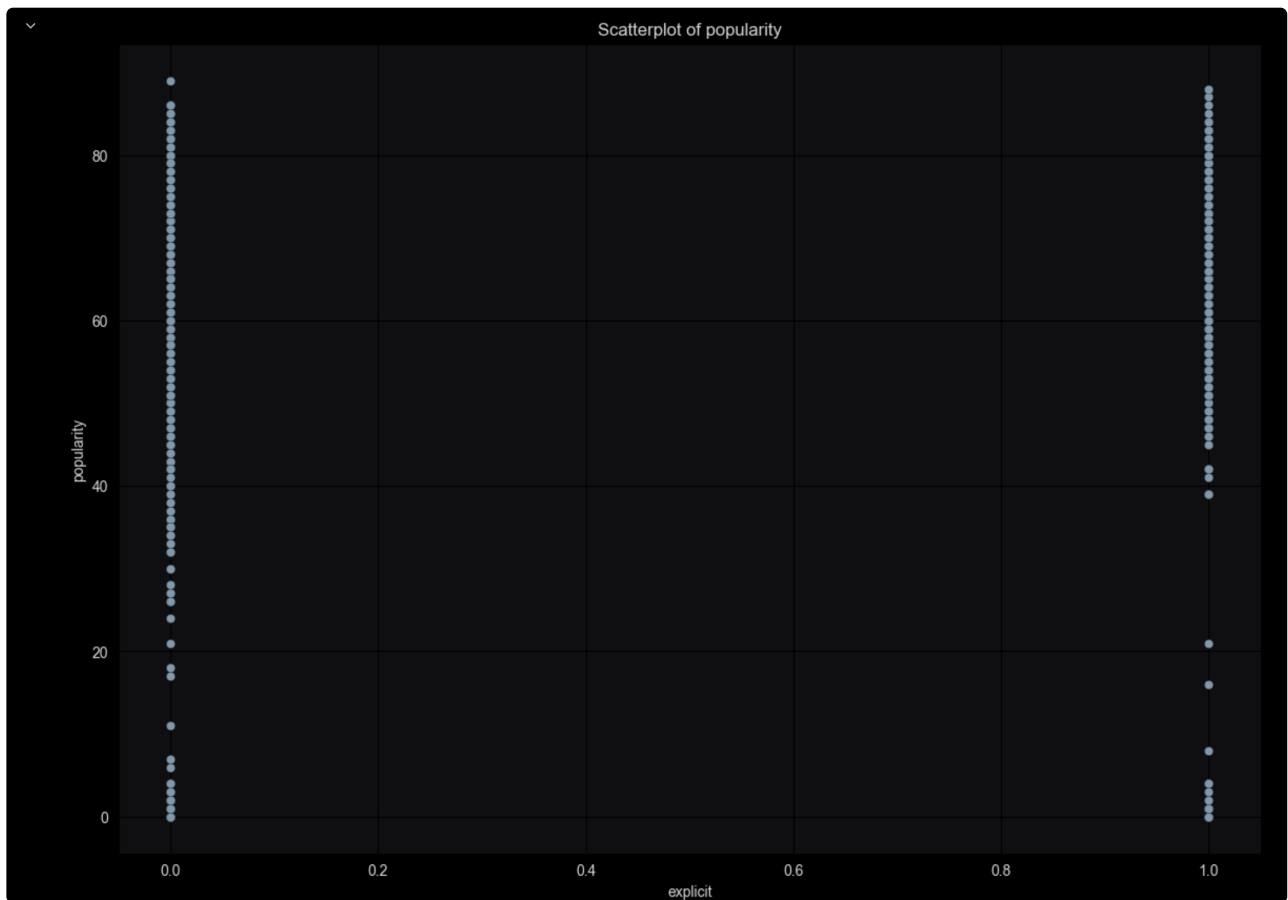
효과

1. 음성의 비중에 따른 인기의 분포를 한눈에 볼수 있습니다.

개선점

1. 0~1의 음성비중에서 0에 치우친 곡들이 많아 이들간 구별이 잘 안되어 로그스케일로 변환해봤습니다.
2. duration_ms를 점의 크기로 시각화해봤습니다

물음3

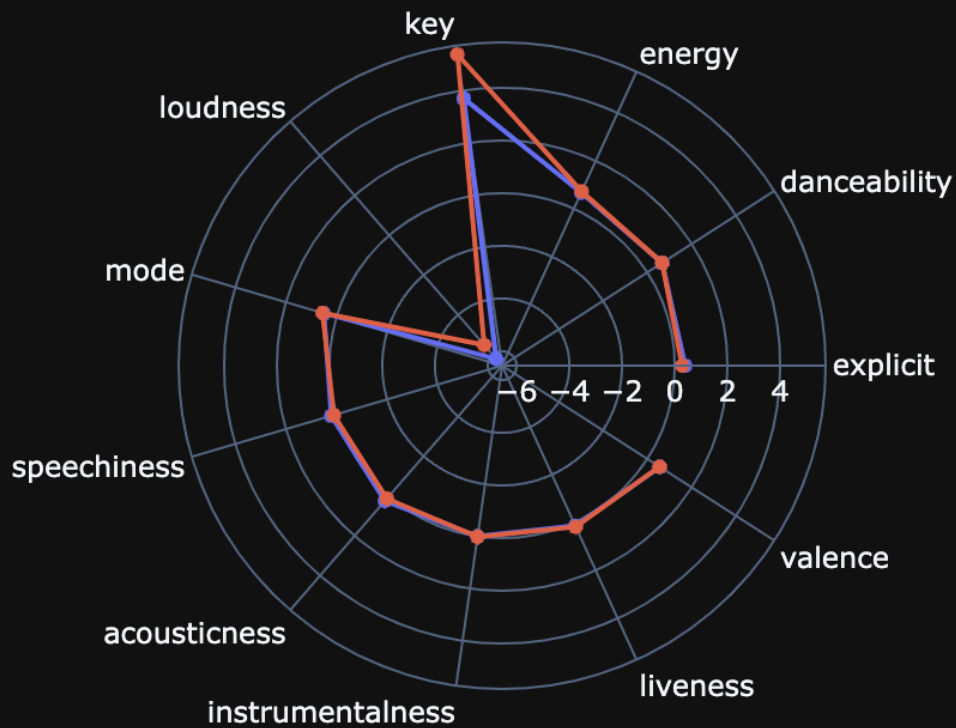


explicit가 양극단에 치우쳐져있는 것을 확인 할 수 있습니다. 양극단에서 둘다 인기도가 고르게 분포되어있으므로 explicit가 인기도에 영향을 준다고 볼 수는 없습니다.

물음4

결론적으로 현재 가지고 있는 데이터로는 인기의 정도와 상관관계/인과관계가 있는 요소를 찾기 쉽지 않은 것 같습니다. 음악이 뭔가 바이럴과 정서적인 측면이 인기도에 많은 영향을 줘서 그런 것 같습니다.

물음 5



상위 10개의 음원들과 나머지 음원들의 문자열 데이터, 시간, 인기도를 제외한 나머지 요소들을 비교해봤습니다. 저의 능력부족 혹은 음원의 인기는 정서적인 요소가 커서 그런지 인기음원만의 특징을 찾지 못했습니다.

Life Expectancy

가설

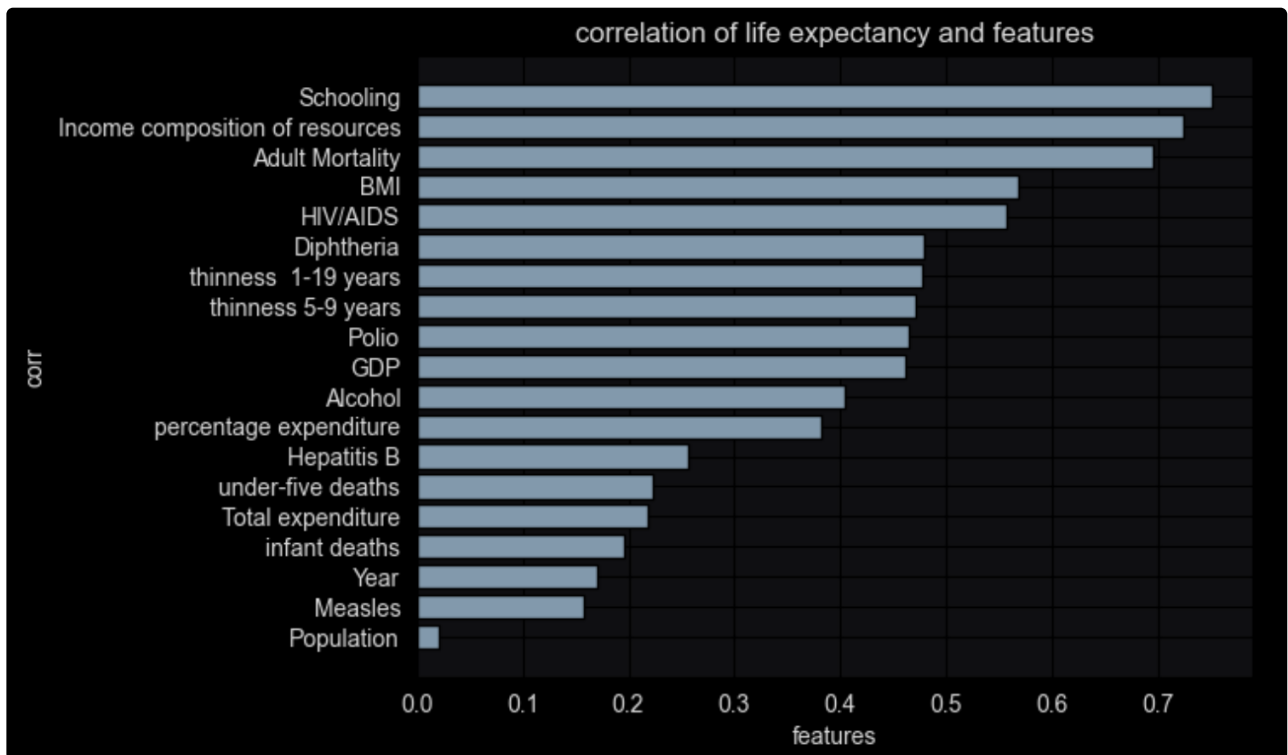
GDP, expenditure 같이 경제적인 능력에 관한 수치들이 기대수명과 높은 상관관계를 가질 것 같습니다. 경제적 능력이 있어야 병원인프라를 누릴 수 있고 식품섭취를 더 골고루 할 수 있을 거라는 판단에 근거했습니다.

고려해야할 독립변수, 종속변수, 데이터 특징

변수간 상관관계를 분석하는 것이 목표이기 때문에 텍스트 데이터들은 제외했습니다.

독립변수는 텍스트 변수와 기대수명을 제외한 나머지 변수들, 종속변수로는 기대수명을 설정하고 그래프로 시각화를 해봤습니다.

시각화에 대한 결론



교육수준, income composition(사회의 소득수준), 성인 사망률, BMI 지수들이 높은 상관관계를 가지는 것을 알 수 있었습니다.

시각화로부터 인사이트

- 경제적능력이 높은 기대수명과 상관관계가 있다는 가설과 비슷하게 높은 교육수준과 사회 소득수준 즉 사회의 인프라가 좋을 수록 기대수명이 높다는 것을 확인 할 수 있었습니다.
- adult mortality(성인 사망률) 즉 15살이 60세 이전에 사망정도가 기대수명과 높은 상관관계를 가지는 것으로 나왔습니다. 그래프를 그릴때 상관계수를 절대값으로 변환했는데 adult mortality는 기대수명과 음의 상관관계를 가지는 것이지만 양의 상관관계를 가지는 것처럼 시각화가 되었는데 시각화에서 충분히 오해의 소지가 생길 수 있는 것을 느꼈습니다.