Open in Colab In [ ]: **import** sys import site import numpy as np import pandas as pd !pip install koreanize\_matplotlib Requirement already satisfied: koreanize\_matplotlib in /usr/local/lib/python3.10/dist-packages (0.1.1) Requirement already satisfied: matplotlib in /usr/local/lib/python3.10/dist-packages (from koreanize\_matplotlib) (3.7.1) Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib->koreanize\_matplotlib) (1.2.1) Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.10/dist-packages (from matplotlib->koreanize\_matplotlib) (0.12.1) Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib->koreanize\_matplotlib) (4.51.0) Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib->koreanize\_matplotlib) (1.4.5) Requirement already satisfied: numpy>=1.20 in /usr/local/lib/python3.10/dist-packages (from matplotlib->koreanize\_matplotlib) (1.25.2) Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib->koreanize\_matplotlib) (24.0) Requirement already satisfied: pillow>=6.2.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib->koreanize\_matplotlib) (9.4.0) Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib->koreanize\_matplotlib) (3.1.2) Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.10/dist-packages (from matplotlib->koreanize\_matplotlib) (2.8.2) Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.7->matplotlib->koreanize\_matplotlib) (1.16.0) In [ ]: import matplotlib.pyplot as plt %config InlineBackend.figure\_format = 'retina' import matplotlib.pyplot as plt import koreanize\_matplotlib 청년 및 고령자 고용동향 분석 과제 주제 선정 이유 현재 한국 사회는 청년층과 고령층의 고용 문제에 직면해 있습니다. 청년층의 높은 실업률과 고령층의 경제활동 참가율 증가는 사회적, 경제적 문제를 야기하고 있습니다. 이러한 문제를 심층적으로 분석함으로써 보다 효과적인 정책을 제안하고자 본 주제 를 선정하였습니다. 분석 목적 이 과제의 주요 목적은 다음과 같습니다. • 청년층(15~29세)과 고령층(50세 이상)의 고용률과 실업률 변화를 비교 분석합니다. • 각 연령대의 경제활동참가율을 분석하여 연령대별 고용 시장의 특성을 파악합니다. • 고용률과 실업률 간의 상관 관계를 분석하여 고용 시장의 상태를 평가합니다. 분석 과정 본 분석은 다음과 같은 과정으로 진행됩니다. 1. 데이터 수집: 청년 고용동향과 고령자 고용동향에 대한 데이터를 수집하였습니다. • 통계청 : https://kostat.go.kr/ansk/ • 고용노동부 : https://www.moel.go.kr/index.do • 중소벤처기업부 : https://www.mss.go.kr/site/smba/main.do 1. 데이터 전처리: 분석에 필요한 데이터만을 필터링하여 사용하였습니다. 2. 데이터 시각화: 고용률과 실업률, 경제활동참가율을 막대 그래프를 통해 시각화하였습니다. 3. 상관 관계 분석: 고용률과 실업률 간의 상관 관계를 분석하여 고용 시장의 특성을 평가하였습니다. 데이터 수집 고용률: https://www.index.go.kr/unity/potal/indicator/IndexInfo.do?cdNo=2&clasCd=10&idxCd=F0105 In []: df = pd.read\_excel('고용률.xlsx') print(df.head()) year 15~19 20~29 30~39 40~49 50~59 60~ 7.7 57.4 74.1 79.1 74.2 39.0 79.1 74.4 39.0 57.9 74.4 2016 58.2 74.7 79.3 74.3 39.5 8.0 2017 8.4 57.6 75.3 79.4 75.3 39.9 4 2018 7.4 57.9 75.7 79.0 75.2 40.1 데이터 전처리 In []: df['청년층'] = df[['15~19', '20~29']].mean(axis=1) df['중년층'] = df[['30~39', '40~49']].mean(axis=1) df['장년층'] = df[['50~59', '60~']].mean(axis=1) print(df[['year', '청년층', '중년층', '장년층']]) 청년층 중년층 2014 32.55 76.60 56.60 2015 32.85 76.75 56.70 33.10 77.00 2017 33.00 77.35 57.60 4 2018 32.65 77.35 57.65 5 2019 32.90 77.20 58.45 6 2020 31.15 76.20 58.35 7 2021 32.40 76.30 59.00 8 2022 34.20 77.70 60.80 9 2023 34.05 78.75 61.60 데이터 시각화 In [ ]: plt.figure(figsize=(10, 6)) plt.plot(df['year'], df['청년층'], label='청년층 (15~29)') plt.plot(df['year'], df['중년층'], label='중년층 (30~49)') plt.plot(df['year'], df['장년층'], label='장년층 (50~)') plt.xlabel('연도') plt.ylabel('고용률 (%)') plt.title('연령대별 고용률 변화 (2014-2023)') plt.legend() plt.grid(True) plt.show() 연령대별 고용률 변화 (2014-2023) 80 70 60 (%) 룔융ፒ 청년층 (15~29) 중년층 (30~49) **─** 장년층 (50~) 40 30 2014 2016 2018 2022 2020 연도 청년층의 고용률이 낮은 이유 자발적 요인 • 열악한 근로환경과 워라밸이 지켜지지 않음 • 연봉 및 복지 • 실업급여 • 본인의 스펙 불만족 비자발적 요인 • 경력직 선호 • 취업난 해결책 : 중소기업이 청년들에게 더 매력적인 환경을 제공할 수 있도 지속적인 노력이 필요함 관련 문헌 https://news.kbs.co.kr/datafile/2022/07/26/297951658816917588.pdf 데이터 수집 고령자 고용동향 : https://www.index.go.kr/unity/potal/main/EachDtlPageDetail.do?idx\_cd=1496 청년 고용동향: https://www.index.go.kr/unity/potal/main/EachDtlPageDetail.do?idx\_cd=1495 In []: df = pd.read\_excel('청년\_고용동향.xlsx') print(df.head()) year 생산가능인구 경제활동인구 취업자 실업자 실업률 고용률 경제활동참가율 9395 4179 3802 378 9.0 40.5 1 2015 4253 3864 389 9.1 41.2 45.3 9363 4334 3908 426 9.8 41.7 2 2016 46.3 3 2017 9282 4333 3907 426 9.8 42.1 46.7 9149 4312 304 408 9.5 42.7 47.1 4 2018 In []: df = pd.read\_excel('고령자\_고용동향.xlsx') print(df.head()) year 경제활동참가율 고용률 실업률 0 2019 68.9 66.9 3.0 1 2020 68.8 66.6 3.3 2 2021 68.6 66.3 3.3 3 2022 70.3 68.8 2.2 71.4 69.9 2.1 4 2023 데이터 전처리 In []: df\_youth = pd.read\_excel('청년\_고용동향.xlsx') df\_elderly = pd.read\_excel('고령자\_고용동향.xlsx') # 필요한 데이터 필터링 df\_youth\_filtered = df\_youth[df\_youth['year'].between(2019, 2023)] df\_elderly\_filtered = df\_elderly[df\_elderly['year'].between(2019, 2023)] 데이터 시각화 In [ ]: # 시각화 years = df\_youth\_filtered['year'].values bar\_width = 0.2 # 막대 너비 fig, ax = plt.subplots(figsize=(14, 8)) # 청년 고용률과 실업률 index = np.arange(len(years)) bar1 = ax.bar(index - 1.5 \* bar\_width, df\_youth\_filtered['고용률'], bar\_width, label='청년 고용률') bar2 = ax.bar(index - 0.5 \* bar\_width, df\_youth\_filtered['실업률'], bar\_width, label='청년 실업률') # 고령자 고용률과 실업률 bar3 = ax.bar(index + 0.5 \* bar\_width, df\_elderly\_filtered['고용률'], bar\_width, label='고령자 고용률') bar4 = ax.bar(index + 1.5 \* bar\_width, df\_elderly\_filtered['실업률'], bar\_width, label='고령자 실업률') ax.set\_xlabel('연도') ax.set\_ylabel('비율 (%)') ax.set\_title('청년 및 고령자 고용/실업률 비교 (2019-2023)') ax.set\_xticks(index) ax.set\_xticklabels(years) ax.legend() # 막대에 값 표시 def add\_labels(bars): for bar in bars: height = bar.get\_height() ax.annotate(f'{height:.1f}%', xy=(bar.get\_x() + bar.get\_width() / 2, height), xytext=(0, 3), # 3 points vertical offset textcoords="offset points", ha='center', va='bottom') add\_labels(bar1) add\_labels(bar2) add\_labels(bar3) add\_labels(bar4) plt.grid(True) plt.show() 청년 및 고령자 고용/실업률 비교 (2019-2023) %청년 고용률 68.8% 70 ■ 청년 실업률 66.9% 66.6% 66.3% 고령자 고용률 고령자 실업률 60 50 46.6% 46.5% 44.2% 43.5% 42.2% (%) 婦用 30 20 9.0% 8.9% 10 7.8% 6.4% 5.9% 3.3% 3.3% 3.0% 2.2% 2.1% 2019 2020 2023 2021 2022 연도 고용률과 실업률 간의 상관 관계 분석을 통해 도출할 수 있는 결론은 다음과 같습니다 일반적으로 고용률이 높을수록 실업률은 낮아지는 부정적 상관 관계가 나타납니다. 이는 경제가 활성화되고 일자리가 많아질 때 고용률이 증가하고 실업률이 감소함을 의미합니다. 고용률과 실업률의 상관 관계를 통해 경제의 건강 상태를 파악하고, 경제 위기나 회복기에 있는지를 이해하는 데 도움을 줄 수 있습니다. In [ ]: # 경제활동참가율 추가 분석 fig, ax = plt.subplots(figsize=(14, 8)) bar1 = ax.bar(index - 0.2, df\_youth\_filtered['경제활동참가율'], bar\_width, label='청년 경제활동참가율') bar2 = ax.bar(index + 0.2, df\_elderly\_filtered['경제활동참가율'], bar\_width, label='고령자 경제활동참가율') ax.set\_xlabel('연도') ax.set\_ylabel('비율 (%)') ax.set\_title('청년 및 고령자 경제활동참가율 비교 (2019-2023)') ax.set\_xticks(index) ax.set\_xticklabels(years) ax.legend() add\_labels(bar1) add\_labels(bar2) plt.grid(True) plt.show() # 상관 분석 correlation\_youth = df\_youth\_filtered[['고용률', '실업률']].corr() correlation\_elderly = df\_elderly\_filtered[['고용률', '실업률']].corr() print("청년 고용률과 실업률의 상관관계:\n", correlation\_youth) print("고령자 고용률과 실업률의 상관관계:\n", correlation\_elderly) 청년 및 고령자 경제활동참가율 비교 (2019-2023) ■ 청년 경제활동참가율 71.4% 70.3% 고령자 경제활동참가율 68.9% 68.8% 70 60 49.8% 49.4% 50 47.9% 47.8% 46.4% <u>%</u> 40 宁 30 20 2019 2020 2021 2022 2023 연도 청년 고용률과 실업률의 상관관계: 고용률 1.000000 -0.965725 실업률 -0.965725 1.000000 고령자 고용률과 실업률의 상관관계: 고용률 고용률 1.00000 -0.97583 실업률 -0.97583 1.00000 청년층과 고령층의 경제활동참가율 변화에서 발견할 수 있는 몇 가지 주요 패턴 1. 경제활동참가율의 일관성: 고령층의 경제활동참가율은 전반적으로 일관된 증가세를 보이는 반면, 청년층의 경제활동참가율은 경제 상황에 따라 변동합니다. 2. 취업 시장 반응: 청년층은 주로 경제 상황에 민감하게 반응하여 경제활동참가율이 변화할 수 있으며, 특히 학업과 취업 준비 기간에 따라 참가율이 달라질 수 있습니다. 3. 고령층의 증가: 고령층의 경제활동참가율이 증가하는 이유는 인구 고령화와 더불어 경제적 필요, 건강 상태, 연금 제도 등의 영향 때문일 수 있습니다. 정책적 제언 및 미래 전망 1) 청년층 고용률 제고 방안 청년층의 취업을 촉진하기 위해 직무 교육과 인턴십 프로그램을 확대하고, 신흥 산업과의 연계를 강화해야 합니다. 청년 창업 지원 프로그램을 통해 자발적 요인으로 인한 비취업 문제를 해결할 수 있도록 해야 합니다. 2) 고용 형태의 안정화 비정규직 문제를 해결하기 위해 정규직 전환을 장려하고, 비정규직의 근로 조건을 개선해야 합니다. 안정적인 일자리 창출을 위한 정책적 지원이 필요합니다. 타이타닉 분석 과제 In [ ]: **import** seaborn **as** sns # Titanic 데이터셋 불러오기 titanic = sns.load\_dataset('titanic') 1. 승객의 생존률 분석 가설 1. 객실의 등급이 높을수록 생존률 또한 높을 것이다. 2. 여성의 경우 남성보다 생존률이 높을 것이다. 3. 어린아이의 경우 청소년과 성인보다 생존률이 높을 것이다. In [ ]: # 1. 객실 등급에 따른 생존율 분석 sns.countplot(data=titanic, x='pclass', hue='alive') plt.title('객실 등급에 따른 생존자와 비생존자 수') plt.xlabel('객실 등급') plt.ylabel('인원 수') plt.show() 객실 등급에 따른 생존자와 비생존자 수 alive yes 350 no 300 250 사 200 150 100 50 0 3 객실 등급 In [ ]: # 2. 성별에 따른 생존율 분석 sns.countplot(data=titanic, x='sex', hue='alive') plt.title('성별에 따른 생존자와 비생존자 수') plt.xlabel('성별') plt.ylabel('인원 수') plt.show() 성별에 따른 생존자와 비생존자 수 alive no yes 400 300 200 100 male female 성별 In [ ]: # 3. 나이에 따른 생존율 분석 sns.histplot(data=titanic, x='age', hue='alive', multiple='stack', kde=True) plt.title('나이에 따른 생존자와 비생존자 분포') plt.xlabel('나이') plt.ylabel('인원 수') plt.show() 나이에 따른 생존자와 비생존자 분포 100 alive no no yes yes 80 60 40 20 10 20 30 40 50 60 70 80 나이 2. 가족 구성과 생존률의 관계 가설 가족구성원이 많을수록 생존률이 낮을 것이다. In [ ]: sns.countplot(data=titanic, x='sibsp') plt.xlabel('형제자매/배우자 수') plt.ylabel('인원 수') plt.show() 600 500 400 의 300 사 200 100 형제자매/배우자 수 In [ ]: # **Titanic** 데이터셋 불러오기 titanic = sns.load\_dataset('titanic') # 형제자매/배우자 수에 따른 생존율 계산 sibsp\_survival\_rate = titanic.groupby('sibsp')['survived'].mean().reset\_index() sibsp\_survival\_rate.columns = ['형제자매/배우자 수', '생존율'] # 생존율 시각화 sns.barplot(data=sibsp\_survival\_rate, x='형제자매/배우자 수', y='생존율') plt.title('형제자매/배우자 수에 따른 생존율') plt.xlabel('형제자매/배우자 수') plt.ylabel('생존율') plt.ylim(0, 1) # y축의 범위를 0에서 1로 설정 plt.show() 형제자매/배우자 수에 따른 생존율 1.0 8.0 0.6 0.4 0.2 형제자매/배우자 수 3. 승선지와 생존률의 관계 가설 부유한 지역의 승선지의 생존률이 높을것이다. In [ ]: # pclass와 embarked, 성별에 따른 생존율 시각화 g = sns.catplot( data=titanic, kind="point", x="pclass", y="survived", hue="sex", col="embarked", aspect=0.75, height=4, errorbar=None, palette="deep" # 한국어로 제목과 축 이름 설정 g.set\_axis\_labels("객실 등급", "생존율") g.set\_titles("{col\_name}에서 승선한 승객") # 각 패싯의 y축 범위 설정 for ax in g.axes.flat:  $ax.set_ylim(0, 1)$ plt.show() S에서 승선한 승객 C에서 승선한 승객 Q에서 승선한 승객 1.0 8.0 0.6 생 사 양 sex female 0.4 0.2 0.0 3 2 3 객실 등급 객실 등급 객실 등급 4. 나이와 생존률의 관계 In [ ]: # 나이에 따른 생존율 분석 sns.histplot(data=titanic, x='age', hue='survived', multiple='stack', kde=True) plt.title('나이에 따른 생존자와 비생존자 분포') plt.xlabel('나이') plt.ylabel('인원 수') plt.show() 나이에 따른 생존자와 비생존자 분포 100 survived 0 1 80 60 40 20 10 20 30 40 50 60 70 나이 In [ ]: # 나이대별 생존율 분석 bins = [0, 12, 18, 35, 60, 100] labels = ['어린이', '청소년', '성인', '중년', '노인'] titanic['age\_group'] = pd.cut(titanic['age'], bins=bins, labels=labels) sns.countplot(data=titanic, x='age\_group', hue='survived') plt.title('나이대에 따른 생존자와 비생존자 수') plt.xlabel('나이대') plt.ylabel('인원 수') plt.show() 나이대에 따른 생존자와 비생존자 수 survived 200 150 50 어린이 청소년 성인 중년 나이대 데이터분석을 통해 나이가 어리고 여성인 경우 생존률이 높다는 결과를 알수 있습니다. 이는 사회적, 문화적 요인에 기인합니다. 요약 타이타닉 사고에서 여성과 어린이의 높은 생존율은 "여성 및 어린이 우선" 규칙, 당시의 사회적 관습, 객실 등급에 따른 구명보트 접근성, 그리고 정보 접근성 등 여러 요인이 복합적으로 작용한 결과입니다. 당시 사회에서는 남성이 여성과 어린이를 보호하 는 것이 당연시되었고 이러한 관습은 구조 활동 중 남성들이 구조 기회를 양보하도록 유도했습니다. 이러한 요인들은 비상 상황에서 여성과 어린이의 안전을 우선시하는 당시 사회의 전반적인 태도를 반영하고 있습니다. 참고 문헌 • Titanic Disaster: Official Casualty Figures and Commentary https://www.anesi.com/titanic.htm • Disproportionate Devastation | Titanic https://courses.bowdoin.edu/history-2203-fall-2020-kmoyniha/reflection/ • Who Were the Titanic Survivors? | HowStuffWorks https://history.howstuffworks.com/