두가지 항목이상의 교차분석이 필요한 모든 항목은 시각화하여 그래프,도표등으로 나타낼 것

Ex)직업 종사 비율 연령대별 히스토그램, 직종 색 나눠서

이중 뚜렷한 경향성을 보이는 경우 추가적으로 회귀곡선 탐색??

분기별 분석은 시행하지 않으므로 가장 최근 데이터 하나만을 할 것인지? 데이터 탐색 시기 결정해야함

개인적으로는 기간별 분석도 말만 기간별 분석이지 연령대별 흐름 분석이기에 하면 좋을거라 생각함 꼭 코로나가 아니어도 소비 소득 경향성을 보여줄 거 같음

시간이 남으면 각 데이터 연령대별 기준에 추가로 지역별/성별 구분 해서 살펴보기

분석을할 때.. **소득별**로도 분석항목 추가 할 수 있을 거 같음

한 열이 한 개체는 아니지만 한 그룹을 한 개체로 보고(단 인구수 고려해서 반영할 것)

4,3/6/14(7/7)/기간별 분석(7)

이상치 처리 :

Ex) 기준시점의 직장 시군구/행정구 코드 : 99999999=정보 없음, null로 처리후 제거

WRC\_SGG\_CD/WRC\_DONG\_CD

연령구간 : AGE\_DVS : 10/21/26/31/36…/70 가 아니라 10/21/22/31/32…/70으로 되어있음, 참고

-인구수 > 급여소득자 인구수+ 자영업자 인구수+ 기타소득자 인구수 일 경우 모자란 값 기타소득자 인구수에 + 반대면 –

-월소득 평균 적당한 범위 고려해서 자릿수만 바뀌지 않으면 ㄱㅊ을 듯 한번 봐야함

분석 시

-지역별 분석 (3+1): 지역 행정 구별 연령대 분포..히스토그램 x축을 지역으로 하는게 나을 듯, 연령대별로 바 색을 다르게 하여 한 바로 합치기

아니면 청-장년층/노년층 으로 구분하여 지도에 매핑

and 각 지역별로 하나씩 원 그래프를 그림(비율)

+평균 출퇴근 거리 분석 : 각 지역별로 나누고 거기서 또 연령별로 나눔 대상은 출퇴근 거리 항목

시각화 위와 비슷하게 하되

이걸 인구수랑도 상관관계 분석하면 좋을 거 같음

인구수 많은 지역과 출퇴근 거리 분석….(연령 제외), 특징을 보이는경우 인구수 분석

++지역별 직업군 비율도 들어가야함 (연령 제외) 이것도 마찬가지로 지역별로 원 그래프…

로 나타내기 이전에 각 비율의 최대값과 평균값 차이 top3 bottom3정도만 시각화하고 두 지역의 특징을 살펴보기…이런느낌으로 가면 좋을 듯

-연령구간별 소득 대상 분석(3)

나이 기준 연령대별 평균소득, 중위소득 항목들 논리에 맞게 sum하고 각각 맞는 값들

배열로 깔끔하게 보여주기, 현황분석에 가까움

+평균 총자산도 마찬가지

+직업 종사 비율은 마찬가지로 현황, 원그래프로 나타내되

연령대별로 하나씩 원 그래프(해당 나이의 80퍼는 무슨무슨 직업)

-소비금액 분석(6)

카드 이용금액… 데이터프레임으로 깔끔하게

+체크, 신용은 그 규모 차이를 보여주면 좋을 듯 총 카드 이용금액 나타내고 연령대별 원 그래프로 신용/체크 사용량 비교?

~~(근데 이 데이터를 원하는 게 은행의 주 고객층 조사라면 카드별로 이용 연령대와 금액을 원그래프로)~~

-대출 분석(14)

연령 구간 별 대출 자 수 비율, 한 그래프에 연령x로 수와 동시에 비 대출자/대출자 색 다르게 하여 수 비율 동시에

+평균 대출 잔액은

우선 이거 대출 미보유자도 평균에 포함인지 확인 필요

X에 연령을 둔 막대그래프로 나타내기

+그외 대출 종류에 따라서 나뉜 항목도 비슷하게 나타내기

++단 연령대별로 대출 종류자의 수와 잔액을 비교할 때, 비율을 따져볼 것

(20대 70퍼가 비은행권, 30대 70퍼가 은행권 이런식으로, 단 중복(여러 종류의 대출을 받음) 인걸 생각해보면 항목별로 한번씩 (한 항목/전체 대출) 이런식으로 해야할 듯

+연령구간별 신용평점 평균 : x축 연령인 선그래프가 깔끔할 듯?

ㄴ이거 기간제로도 한 번 볼 수 있을 듯 x축 시간으로 놓고 연령대별로 그래프 하나씩

+장/단기 이상 연체자 수, 비율

연령대별로 대출자 수 나타낼 때 장/단기 중복데이터 있는지 확인해보고

색다른 선 그래프로 한번에 or 한 항목/전체로

+대출 보유자 비율과 소득 평균

+대출보유자 중, 차를 가지고 있는 경우, 집을 가지고 있는 경우, 둘 다 가지고 있는 경우를 원그래프로 표시

++시간남으면 총 대출이 아닌 은행권/비은행권 대출자 로 나누어서 추가 분석

+비은행권, 카드론 보유자비율과 신용평점

이건 연령대별로 할 수 밖에 없음, 평균 신용평점&비은행권/카드론 대출보유자수를

x연령, y 비은행권/카드론비율(%)&신용점수 로 해서

+위 그래프에 추가로 소비(카드 금액) 평균을 넣되, 소득 대비 카드 이용금액 비율로 들어가면 좋을 거 같음, 단 비은행권/카드론비율 과 구분되게 넣어야함…이건 실수치 확인하고 표기 방식을 변경하면 좋을 듯(같은 열에 있어도 카드론 비율이면 30퍼지만 소비금액 비율이면 70퍼란 식으로)

-기간별(6)

보유하고 있던 건 뺴는 게 나을 듯, 월 소득,

+신규 대출자 수,대출 잔액 등 만 살펴보기

시간 흐름에 따라 선그래프로 연령대별로 색 구분해서 대출자 수 추이가 어떻게 됐는지 살펴보기

월별 그대로 나타내면 될 듯 다만 항목들 연령 구분해서 더 해야함