발표 스크립트

안녕하세요 경기미래기술학교 부동산 프로젝트를 진행한 아파트마트 입니다. 저희는 서울시 아파트 실거래 데이터를 기반으로 아파트 거래의 추세를 탐색하고 분석하는 프로젝트를 진행했습니다

-목차

저희 팀의 발표순서는 다음과 같습니다.

--팀 소개

먼저 팀원을 소개하겠습니다.

저희 팀은 류성하 윤동호 이주희 이준형으로 구성 되어있습니다.

--주제선정배경

COVID-19로 인한 무제한적 금리인하가 종식되면서 금리인상이 예고대로 진행되고 있습니다. 이로 인해 주택을 구매한 사람들에게는 이자부담이 가중되었고 대출이자의 상승으로 대출건수도 크게 줄어 부동산 수요가 크게 줄었습니다. 게다가 올해 행정부는 주택공급 확대계획을 내놓았고, 줄어드는 수요와 확대 공급계획의 달성은 주택가격 하락에 더 가속도를 높일 것으로 보입니다.

--프로젝트 목표

저희의 프로젝트 목표는 다음과 같습니다.

먼저 유사한 면적의 아파트라도 가격차이가 나는 이유에 대해 규명하려 합니다. 다음으로 우리가 아파트를 구매한다면 어떤 아파트를 구매해야 하는지에 대해 알아보려고 합니다.

-데이터 분석

--활용도구

저희는 브이에스코드의 파이썬을 이용해 코드를 작성했고, 사소한 데이터 수정이나 단순한 계산등은 엑셀을 통해 진행했습니다. 협업도구로는 노션을 통해 의견과 개별 태스크 진행상황을 공유하고 데이터와 코드는 깃허브를 통해 공유했습니다. 멘토링은 줌을 이용한 비대면으로 진행됐습니다. 마지막으로 태블로에는 군집분석의 결과를 올려서 아파트 매물의 분기별 변동 추세를 시각적으로 나타냈습니다.

--프로젝트 설계

저희는 다음과 같은 과정을 통해 프로젝트를 진행했습니다. 먼저 국토교통부에서 제공하는 18년1분기부터, 22년2분기까지의 부동산 실거래 데이터를 가져와서 기본데이터로 삼고 아이디를 만들었습니다. 실거래 데이터에는 거래에 관련한 정보만 있고 아파트 자체에 대한 정보는 부족하기 때문에 (멘토님으로부터) 외부데이터를 받아 도로명 주소를 기준으로 실거래데이터와 조인한 뒤 추가변수로써 사용했습니다.

그렇게 완성된 데이터베이스를 회귀분석해서 유효한 변수를 추출해내고 군집알고리즘을 사용해 군집화 한 뒤 이를 시각화 함으로써 인사이트를 도출해 내려고 했습니다.

---데이터 전처리

먼저 아이디 결정입니다. 저희는 아이디를 시 구 동 단지명 면적을 결합해 만들었습니다. 로우데이터의 행들을 그대로 아이디로 쓸 경우 결측치가 지나치게 많아 분석이 불가능하고, 너무 크게 잡을 경우 분석에서 의미를 얻을 수 없기 때문입니다. 저희가 분석대상을 아파트로 정한 만큼 각 아파트를 구분할 수 있는 최대의 범주가 아파트 단지명이라고 생각했고 단지 내에서 평수에 따라 조건 및 가격의 변동이 크기 때문에 전용면적을 반올림해서 추가했습니다. 이 과정을 거쳐 만 육천여개의 아이디를 생성했습니다

다음으로 변수 설명입니다. 저희가 모은 로우 데이터의 모든 변수는 다음과 같습니다. 이중 설명이 필요한 변수는, 먼저 거래금액과 전세금은 각 분기별 평균값으로 지정했습니다. 등락율은 그렇게 구한 거래금액의 변동율을 계산한 것이며, 전세가율은 저희가 구한 평균 거래금액과 전세금을 이용해 구한 것으로 실제 전세가율과는 차이가 있을 수 있습니다.

인접 역수는 각 역의 도로명 주소를 기반으로 지오코딩해 직선상의 거리를 구하고 해당 거리가 오백 미터 이내인 경우 트루, 아니면 폴스를 적용했습니다. 거래유형의 경우 직거래는 증여 등 다양한 이유로 중개거래에 비해 가격 변동이 클 가능성이 높기 때문에 직거래 데이터는 모두 제외했습니다.

다음은 결측치 처리입니다. 일차적으로 부동산 실거래데이터와 파생변수들은 파이썬 기능을 통해 직전분기 데이터를 가져와 채워넣었고 직전분기 데이터가 없는 경우에는 직후분기 데이터를 채웠습니다. 그 외에 결측치가 전체데이터의 반을 넘는 경우엔 분석의 정확도를 저해하기 때문에 삭제했고 최종적으로 만 육천여개의 아이디 중 약 사천개의 아이디를 남기게 됐습니다.

결측치를 주변 매물의 평균치나 같은 매물의 앞뒤 분기값의 평균으로 보정하지 않은 이유는 저희가 아파트 구매를 고려할 때 기준으로 삼는 가격은 가장 최근에 거래된 가격을 기준으로 삼기 때문입니다.

---회귀

저희가 가진 로우 데이터는 고정된 변수가 많아서 최적군집수가 뚜렷히 안잡히거나 군집이 완전 고정되는 문제가 발생했기 때문에 회귀분석을 통해 각 변수들의 상관관계와 피-밸류를 확인해 변수간 우선순위를 정해고 우선도가 떨어지는 변수를 제거하는 후진소거법의 방식으로 변수를 조정했습니다. 회귀분석 결과는 다음과 같습니다.

회귀 분석 결과 주택금융가능여부, 등락율, 전세가율, 전용면적, 건물연령, 세대 수, 세대당 주차 수, 지하주차비율, 관리시설 수를 최종변수로 채택했습니다.

추가적으로 이렇게 구해진 최종변수들을 기준으로 22년도2분기의 데이터를 가지고 회귀분석을 진행하면 다음과 같은 결과를 얻을 수 있는데 이 두 결과를 비교해보면 18년도 1분기에 비해 세대수와 전용면적의 영향력이 증가했고, 건물연령의 경우 변별력을 상실하는 것을 알 수 있었습니다.

---군집분석

군집분석은 케이 민즈 알고리즘을 사용했으며, 먼저 최적 군집수를 찾아내기 위해서 엘보차트를 이용했고 네개의 군집갯수를 채택했습니다. 이후 사전에 구해놓은 위경도 데이터를 군집데이터에 붙이고 태블로를 통해 지도위에 시각화했습니다. 시각화한 결과물은 태블로에서 시연하겠습니다.

(태블로에서 먼저 18/1분기부터 22/2분기까지 군집변동을 보여줌)

군집분석에서 각 군집간의 우열을 가릴 수는 없지만 비교적 고가의 아파트들이 속해 있는 노란색과 빨간색 군집이 확산하는 것을 확인할 수 있습니다. 여기서 저희 프로젝트의 목표인 비슷한 면적의 아파트임에도 가격차이가 나는 이유를 알기 위해 공릉동 동신아파트와 동부아파트를 비교군으로 보겠습니다. 지도에서 상단에 나타나는 두개의 녹색 점이 동신아파트이고 아래 파란 점이 동부아파트입니다. 두 비교군은 도로 하나를 사이에 두고 마주보고 있으며 비슷한 면적을 가지고 있습니다. 그러나 20년 1분기에는 동신아파트의 전용면적이 80제곱미터인 매물이 노란색 군집으로 변하는 것을 확인할 수 있고 시간이 더 흘러 동년 3분기에 와서야 동부아파트의 80제곱미터 매물이 빨간색 군집에 들어가는 것을 확인할 수 있습니다. 21년 4분기에는 동신아파트의 60제곱미터 아파트도 노란군집으로 변하지만 동부 아파트의 60제곱미터 매물은 마지막까지 파란색 군집을 유지하는 것을 볼 수 있습니다. 이 두 비교군의 데이터를 보면 동신아파트가 동부아파트에 비해 세대수가 더 많다는 것과 세대당 주차댓수와 지하주차비율이 차이를 보인다는 것을 확인할 수 있었습니다. 피피티에 적힌 아파트들 외에도 비슷한 면적과 건축년도를 가지고 있으나 제반시설에 따라 가격 상승이 늦거나 상승하지 못하는 경우를 쉽게 찾을 수 있습니다.

--결론

이번 프로젝트를 통해 저희가 도출한 결론입니다.

* 회귀분석 결과, 우리가 흔히 알고 있는 역세권은 아파트 가격변동에 큰 영향을 미치지 못하는 것으로 드러났습니다.
* 18년도 1분기보다 22년도 2분기에 세대수와 전용면적이 더 큰 가격 영향을 주었습니다.
* 또한, 18년도 1분기에는 건물연령이 노후된 아파트의 선호는 낮았지만 22년도 2분기 국면과 비교해보면 이 변별력을 상실하였습니다. 이는 재건축으로 인한 기대수요 및 아파트 시장이 급성장하면서 투자를 목적으로 하는 거래가 늘어난 것이 영향을 미친 것으로 보입니다.
* 비슷한 면적의 아파트에서 아파트의 제반시설이 세대 수, 관리시설이 큰 영향력을 가지는 것을 확인했습니다.
* 현 시점에선 부동산 시장 자체가 위축되고 거래량 자체가 줄어 매물을 찾기 어렵겠지만 만약 구할 수 있다면 흔히 홍보문구에 나오는 역세권 보단 실제로 사용 가능한 기반시설이나 주차공간, 세대수 등을 우선적으로 고려할 필요가 있다고 볼 수 있습니다.

--향후과제

저희의 향후과제는 다음과 같습니다.

* 본격적인 하락국면에 접어드는 22년 3분기 데이터를 추가하면 하락폭 관측 및 하락장에서의 군집변동을 탐지할 수 있을 것으로 예상되고
* 회귀분석 과정에서 빠지거나 기술적 문제로 추가하지 못한 변수 값을 사용가능 하게 조정하거나 추가함으로써 더 발전된 분석 결과와 인사이트 발견을 기대할 수 있습니다.
* 또한, 이번 프로젝트에선 서울 내 아파트만 대상으로 분석이 이뤄졌지만 향후 전국 아파트데이터를 추가하면 더 광범위한 트렌드를 관측할 수 있을 것으로 예상되며.
* '18.1Q와 '21.4Q를 하나의 페어로 비교 및 '21.4Q와 '22.3Q를 비교하여 국면을 구분하여 시장을 비교하여 각 국면에 어떤 변수가 중요한지 분석하여, 소비자에게 더 나은 부동산 정보 및 인사이트를 제공할 것입니다.