



초기분양률과 특성조합에 기반한 부동산 PF 위험 관리에 관한 연구*

A Study on Real Estate Project Financing (PF) Risk Management Based on Initial Sales Rate and Bundles of Characteristics

김은미**
Eunmi Kim

Abstract

Initial sales rates are critical for gauging project financing (PF) project risk ex ante, as they determine the realizability of future cash flows. This study approaches initial sales performance not as the outcome of isolated individual factors but as the result of the market's collective choice regarding the bundles of characteristics offered by housing projects. To frame this perspective, the study adopts Lancaster's (1966) characteristics-based consumer theory as its theoretical foundation and classifies PF projects using K-means clustering. Moving beyond the question of "which factors are statistically significant on average," it provides a structural answer to a different question: "Which bundles of characteristics enable housing projects to absorb demand more rapidly and thereby exhibit greater resilience to PF risk?" The K-means results show that the cluster with high initial sales rates is characterized by a combination of mid-sized unit composition, a non-excessive supply scale, price levels comparable to nearby competing projects, and a stable market phase. By shifting the analytical focus from variable-level marginal effects to an attribute-centered perspective, this study offers a more structural interpretation of housing project sales performance and contributes to practical PF risk management.

Keywords: Real estate project finance (PF), Initial sales rate, Characteristics-based consumer theory, K-means, Bundle of characteristics

* 본 연구는 '25년 주택도시보증공사가 주최하는 주택금융과 주거안정을 위한 대토론회 발제 내용을 수정·보완한 것임.

** HUG 주택도시금융연구원 연구위원 | Research Fellow, Housing Urban Finance Research Institute, Housing & Urban Guarantee Corporation (HUG) | kem531@khug.or.kr |

I. 서론

국내 부동산 시장의 불확실성이 확대되면서, 부동산 프로젝트 파이낸싱(project financing, PF)의 잠재적 위험이 주요 정책적·금융적 이슈로 부각되고 있다.

부동산 PF는 향후 분양을 통해 창출될 현금흐름을 기초로 자금을 조달하는 구조적 특성상 분양 성과가 사업의 안정성과 직결되는 핵심 요인으로 작용한다. 특히 분양 초기 단계에서의 성과는 이후 현금흐름의 실현 가능성을 좌우하며 PF 사업장의 위험을 사전에 가늠할 수 있는 중요한 신호로 기능한다.

이러한 맥락에서 초기분양률은 단순한 판매 실적 지표를 넘어 PF 사업장의 현금흐름 안정성과 부실 가능성을 동시에 반영하는 핵심 지표로 인식되어 왔다. 실제로 PF 보증이 적용된 사업장들의 보증 사고 사례를 살펴보면, 사고가 발생한 다수의 사업장에서 초기분양 성과가 상대적으로 저조하게 나타나는 경향이 확인된다.¹⁾ 이는 분양 성과가 충분히 확보되지 않을 경우 예상 현금흐름이 초기 단계에서 실현되지 못하고, 그 결과 PF 구조 전반의 취약성이 확대될 수 있음을 시사한다.

그럼에도 불구하고 기존의 관련 연구들은 초기 분양률에 영향을 미치는 요인을 주로 개별 변수 단위로 분석하고, 각 요인의 통계적 유의성을 검증하는 접근에 집중해 왔다.

이러한 접근은 특정 요인의 평균적 효과를 파

악하는 데에는 유용하지만, 실제 PF 사업장이 지니는 복합적인 구조와 사업장 간 이질성을 충분히 설명하는 데에는 한계를 가진다.

동일한 거시경제 환경에서도 입지 조건, 단지 규모, 상품 구성, 가격 구조 등 다양한 요소가 결합된 사업장에 따라 분양 성과와 위험 수준은 상이하게 나타날 수 있음에도, 기존 연구는 이러한 구조적 차이를 체계적으로 설명하는 데에는 제한적이다.

이에 본 연구는 초기분양률을 개별 요인의 결과가 아닌, 주택 사업장이 제공하는 특성 조합에 대해 시장이 내린 집합적 선택의 결과로 이해하는 관점에서 접근한다.

본 연구의 핵심 문제의식은 “어떤 요인이 초기 분양률에 영향을 미치는가”가 아니라, “어떤 구조와 특성 조합을 가진 사업장이 시장에서 상대적으로 빠르게 수요를 흡수하며 PF 위험에 보다 강한가”에 있다. 이러한 관점 전환을 통해 본 연구는 주택 사업장의 분양 성과를 보다 구조적으로 해석하고, PF 위험 관리의 관점에서 실질적인 시사점을 도출하고자 한다.

II. 선행연구

본 연구는 초기분양률을 PF 위험 관리의 핵심 신호로 보고, 이를 단일 요인의 평균적 효과가 아니라 사업장 특성 조합에 대한 시장 선택 결과로

1) 2007년부터 PF(project financing) 보증사고 발생 이전 확인 가능한 데이터를 기준으로 총 5건의 보증사고 사업지를 검토한 바에 의하면, 그중 4곳 사업장의 실제 초기분양률 대비 예상 초기분양률이 낮고 사업장 관리 과정에서 분양률 부진에 따른 PF 위험이 최초 감지되었다.

해석하려는 문제의식에서 출발한다.

따라서 선행연구 검토는 (i) 분양성과를 나타내는 초기분양률이 무엇에 의해 좌우되는지에 대한 실증 분석, (ii) 초기분양률이 PF 부실 및 금융 안정성과 어떻게 연결되는지에 대한 위험 메커니즘, (iii) 주택을 특성 결합체로 보고 구조·유형을 식별하기 위한 방법론적 연구를 검토함으로써 기존 연구와의 차별성을 도출하는 데 목적이 있다.

먼저 초기분양률에 관한 연구는 가격, 입지, 단지 규모, 시장 여건 등 다양한 요인이 영향을 준다는 점을 반복적으로 확인해 왔다.

백민석·신종철(2011)은 '02년~'10년에 분양된 서울, 경기, 인천 아파트의 거시환경과 단지특성을 독립변수로 설정하고 다중회귀분석을 적용하였다. 그 결과 건설회사의 도급순위가 높을수록, 총세대수가 클수록 초기분양률이 높게 나타났고, 주택담보대출금리가 유의미한 거시환경 특성인 것으로 나타났다. 또한 서진호·강정규(2014)는 경남·부산 지역의 신규 아파트를 대상으로 '08년 이전·후 초기분양률 결정요인을 실증적으로 분석한 결과, 재건축, 브랜드가치, 해당 지역의 부동산 시장 영향을 받는다는 점을 제시하였다. 또한 실무 현장에서 느끼는 부동산 시장의 경기는 부동산 정책이 크게 작용함을 추가 언급하였다.

더 나아가 김리영·서원석(2016)은 경기도 사업장 중에서 주변 지역에 비해 높은 초기분양률을 보인 사업장의 특성요인을 파악하기 위해 높은 초기분양률과 낮은 초기분양률 사업장으로 구분하고, 정책 및 시장여건을 중심으로 이항로지스틱 회귀분석을 수행, 아파트 초기분양률에 영향을 미치는 특성요인 파악을 위한 다중회귀분석을 시행

하였다. 분석 결과, 주변지역에 비해 초기분양률이 상이한 사업장은 정책적 요인이 영향을 미치고 있음을 확인하였고, 분양 가격 및 초·중·고등학교 거리, 공원 거리, 신도시 여부가 초기분양률에 영향을 미치는 유의미한 변수인 것으로 나타났다.

일부 연구는 초기분양률을 예측하는 문제로 확장하여, 기존의 회귀 분석 접근을 넘어 데이터마이닝·기계학습을 적용하기도 했다.

김지영·이상경(2018)은 기존 연구들이 사업장 단위의 결정요인을 다룬 것과 달리, 하남도시공사가 위례신도시에 분양하였으나 미분양된 A 아파트 각 세대를 대상으로 미시적 관점에서 초기분양률 예측모형을 개발하였다. 이는 초기분양률 연구가 단순 설명변수의 유의성 확인을 넘어 “성과를 사전에 식별”하려는 방향으로 분석하였음을 시사한다.

이 연구들에서 공통적으로 확인되는 특징은, 초기분양률을 종속변수로 두고 개별 설명변수의 평균적 한계효과 즉, 다른 조건이 같을 때, 분양가·입지·규모 등 독립변수가 한 단위 변하면 성과가 얼마나 변하는가를 추정하는 방식이 여전히 중심이라는 점이다. 이 접근은 “어떤 요인이 영향을 미치는가 또는 그 정도”를 제시하는 데 유용하지만, PF 사업장처럼 다차원 속성이 동시에 결합되는 대상에서는 특성들이 함께 묶여 작동하는 구조적 패턴(예: ‘규모-입지-가격-시장여건’의 조합)이 성과를 좌우할 수 있다는 점을 충분히 포착하기 어렵다. 즉, 변수별 한계효과 중심 접근은 이런 “조합의 이질성”을 평균화할 위험이 있다.

다음으로 부동산 PF 부실 및 금융 안정성 관련 선행연구는 PF가 미래 현금흐름에 대한 기대를

기반으로 하는 구조라는 점에서, 초기분양률이 유동성 위험과 부실로 연결될 수 있음을 강조해 왔다.

손재영(2009)은 분양률이 낮아지면 주택사업자의 수익률이 급격히 하락하기 때문에 주택사업은 위험이 매우 높은 사업임을 지적하였으며, 백민석·신종철(2011)도 낮은 초기분양률이 지속될 경우 사업지연에 따른 금융비용의 증가 등으로 건설업체 및 분양계약자에게도 위험이 전이되어 파산 위험이 발생할 수 있음을 언급하였다. 최근 김용은 외(2025)는 주택개발 PF 사업의 부실화 여부에 영향을 미치는 요인을 실증적으로 분석하였다. '부실화'는 현금흐름 불일치, 금융구조상 트리거 작동, 원리금 상환지연, EOD(event of default) 발생으로 정의하고, PF 대출 원리금을 상환하기 위한 최소한의 분양률과 부실 발생 가능성 간 유의한 정(+)의 관계를 확인하였다.

또한 유정근·오동훈(2010)은 부실화된 PF 사업장에서 재무적 투자자가 채권 회수를 위한 효율적인 관리방안을 제시하고자 하였다. 이에 사례 분석 및 전문가 대상 집단면접(focus group interview, FGI)을 수행하였으며, 분양대금의 입금현황과 분양률에 따른 자금 사용분석의 관리 시스템 구축, 미분양에 따른 유동성 위기 등을 예측할 수 있도록 현금흐름 관리 업무의 중요성을 주장하였다.

이와 같이 PF 부실 및 금융안정성을 다룬 연구는 부실 사업의 관리 흐름과 자원 조달 대응을 동시 고려하며, PF 위험이 실무적으로 관리·정리되어야 하는 대상임을 보여준다.

그러나 이 범주의 연구들은 제도적 요인·관리

체계에 초점을 두어, "어떤 사업장이 왜 구조적으로 분양 성과가 낮아지고 취약해지는가"를 사업장 특성 조합의 수준에서 유형화하여 설명하는 데에는 상대적으로 약한 경우가 많다.

마지막으로 주택특성의 결합 구조 및 분석방법론 관련 선행연구는 "주택은 특성의 묶음"이라는 관점에서 발전해 왔으며, 전통적으로는 헤도닉 모형(hedonic price model)을 중심으로 축적되어왔다. Rosen(1974)은 주택이라는 복합재를 입지, 면적, 품질 등 특성으로 분해하여 가격에 반영되는 기여를 추정하는 헤도닉 모형으로 정리하였고, 이후 실증 연구들은 이 모형을 다양한 주택 시장 자료에 적용해 왔다.

다만 헤도닉 접근은 본질적으로 "가격"을 중심 결과로 두는 경향이 강하며, 특성의 조합을 식별하기보다 특성의 개별 기여(계수)를 해석하는 데 강점이 있다. 이와 연계하여 최근에는 군집분석 등 비지도 학습을 이용해 주택을 유형화하거나 하위시장을 식별하려는 방법론적 연구가 확장되고 있다. 권순재 외(2017)는 연립·다세대 주택 가격 특성을 반영한 후, K-means와 헤도닉 모델을 결합한 군집분석을 수행하여 주택 유형별 가격모형을 다르게 추정하는 방식(구조적 이질성 반영)을 제안하였다.

또한 하은혜·이건학(2024)은 시계열 군집분석을 활용해 주택 가격을 중심으로 하위시장을 유형화하였으며, 이는 '유사성 기반 분류'가 시장의 이질성을 식별하는 방법임을 보여준다.

이상의 선행연구를 종합하면, (i) 초기분양률 연구는 성과의 결정요인을 제시했으나 대체로 변수별 평균적 한계효과 추정에 집중했고, (ii) PF 위

험 연구는 분양 성과의 중요성을 전제하면서도 사업장 미시 구조의 조합을 유형화해 설명하는 데에는 제한이 있으며, (iii) 주택 특성 분석은 ‘특성의 묶음’ 관점을 공유하되 가격 중심(계수) 중심 또는 시장 유형화 중심으로 분화되어 초기분양률과의 직접 결합이 불충분하다.

즉, 기존 연구들은 “무엇이 영향을 미치는가”에는 답했지만, “어떤 특성 조합 구조를 가진 사업장이 시장에서 선택되어 초기분양률이 상이한가”라는 질문에는 상대적으로 약했다.

따라서 본 연구의 필요성은 다음과 같다.

첫째, 초기분양률을 PF 위험 관리의 실질적 신호로 활용하려면 개별 요인의 계수 해석을 넘어 사업장 수준의 구조적 패턴을 식별할 필요가 있다. 둘째, 주택을 특성의 결합체로 이해하는 관점은 존재하지만(헤도닉모형 등) 이를 분양성과이자 시장 선택 결과인 초기분양률과 연결해 PF 위험 관리 문제로 통합한 실증 설계는 제한적이다.

이에 본 연구는 선행연구가 축적해 온 “요인”들을 사업장 특성 벡터로 재구성하여 “사업장 특성 조합과 선택 결과”라는 해석 틀로 전환한다는 점에서 차별성을 갖는다.

III. 이론적 배경

본 연구는 사업장의 분양 성과를 개별 요인의 한계효과가 아닌, 특성 조합에 대한 시장의 선택 결과로 이해하는 관점에서 출발한다.

이에 Lancaster(1966)의 특성 기반 소비자이론(characteristics-based consumer theory)을 이론적 틀로 채택한다. Lancaster는 소비자가 재화 그 자체를 소비하여 효용을 얻는다는 전통적 가정과 달리, 효용의 원천은 재화가 제공하는 특성(characteristics)에 있으며, 재화는 이러한 특성들의 결합체로 이해되어야 한다고 주장하였다.²⁾

이 관점에서 소비자의 효용함수는 재화 공간(goods space, G-space)이 아닌 특성 공간(characteristics space, C-space)에서 정의된다.

$$U = U(z) \quad (\text{식 1})$$

여기서 z 는 소비자가 인식하고 평가하는 특성 벡터를 의미한다. 재화 벡터 x 와 특성 벡터 z 는 소비기술(consumption technology)을 통해 연결되며, Lancaster의 단순화된 모형에서는 (식 2)와 같이 표현된다.

$$z = Bx \quad (\text{식 2})$$

이 식에서 B 는 재화가 어떤 특성을 어떤 구조로 생성하는지를 나타내는 변환 행렬이다. 중요한 점은 이 식이 실증적으로 계수를 추정하기 위한 회귀식이 아니라, 재화와 특성 간의 구조적 관계를 정의하는 항등식이라는 점이다. Lancaster의 이론에서 핵심은 B 의 값을 추정하는 데 있지 않고, 특성 조합의 구조에 의해 이루어진다는 점에 있다.

이러한 이론적 설정은 소비자의 선택이 발생하

2) Lancaster(1966)의 특성 기반 소비자이론은 이후 Rosen(1974)에 의해 헤도닉 가격 모형으로 이어지며 시장에서 관측되는 가격은 재화가 담고 있는 특성 묶음에 대한 소비자 선호와 경쟁적 공급이 균형을 이루는 결과로 해석된다.

는 공간과 연구자가 관측하는 공간을 구분한다. Lancaster는 소비자의 효용 극대화가 이루어지는 공간을 C-space로 정의하고, 연구자가 관측 가능한 재화의 소비량과 가격이 정의되는 공간을 G-space로 구분하였다. 소비자의 실제 선택은 C-space에서 이루어지지만, 연구자는 이를 직접 관측할 수 없으며, 오직 G-space에 나타난 재화의 특성과 선택 결과만을 관측할 수 있다.

본 연구는 이와 같은 공간 구분을 주택시장에 적용한다. 주택 사업장은 단일한 재화라기보다, 입지, 가격 구조, 공급 규모, 시장 환경 등 다양한 요소가 결합된 특성 벡터로 이해될 수 있다. 연구자가 관측하는 것은 각 사업장의 이러한 특성들로 구성된 재화 벡터이며, 이는 G-space에 해당한다. 반면, 소비자가 실제로 평가하는 주거의 쾌적성, 시장 적합성, 신뢰도와 같은 특성들은 C-space에 존재하며 직접적으로 관측되지 않는다.

이러한 맥락에서 본 연구는 개인의 사업장 선택(미시)이 사업장 초기분양률(거시)로 집계되는 구조를 설정한다. 초기분양률은 개별 소비자의 효용을 직접 측정한 값은 아니지만, 동일한 특성 조합에 대해 다수의 소비자가 내린 선택이 누적된 집합적 선택 결과로 해석될 수 있다.

따라서 초기분양률은 특성 기반 소비자이론에서 말하는 효용 극대화 행동이 시장 차원에서 집계된 결과로 이해될 수 있다.

이를 반영하여 각 주택 사업장 j 는 해당 재화가 제공하는 특성 묶음(bundle) z_j 을 제공한다고 가정하고, 특성의 차원을 r , 선택 가능한 사업장의 개수를 J , 사업장 $j(=1, \dots, J)$ 의 특성벡터를 (식 3)과 같이 정의한다.

$$z_j = (z_{1j}, z_{2j}, \dots, z_{rj})^T \in \mathbb{R}^{r \times J} \quad (\text{식 3})$$

각 사업장 특성벡터를 열(column)로 적층하여 특성행렬 B 를 (식 4)와 같이 정의한다.

$$B = [z_1, z_2, \dots, z_J] \in \mathbb{R}^{r \times J} \quad (\text{식 4})$$

B 의 j 번째 열은 사업장 j 의 특성벡터 z_j 에 대응하며 C-space는 $z \in \mathbb{R}^{r \times 1}$ 으로 정의한다.

한편, 초기 분양기간에 참여하는 잠재수요자들 $i=1, \dots, N$, 개인 i 는 사업장 중 하나를 선택하여 '1주택을 구매한다'고 가정한다. 이는 개인 i 의 선택을 구매(1), 비구매(0)로 구분하여 one-hot 벡터로 다음과 같이 표현된다.

$$x^{(i)} = (x_1^{(i)}, \dots, x_J^{(i)})^T \in \{0, 1\}^{J \times 1}, \\ \sum_{j=1}^J x_j^{(i)} = 1 \quad (\text{식 5})$$

즉, G-space는 사업장 대안에 대한 선택 $x^{(i)}$ 가 정의되는 공간으로서 $x_j^{(i)} = 1$ 은 개인 i 가 사업장 j 를 선택했음을 의미하고, 개인 i 의 선택으로 실현되는 특성은 (식 6)과 같이 정의된다.

$$z^{(i)} = Bx^{(i)} \in \mathbb{R}^{r \times 1} \quad (\text{식 6})$$

One-hot 구조상 $Bx^{(i)}$ 는 선택된 사업장에 해당하는 열벡터 z_j 를 '선택적으로 추출'하므로 $z^{(i)}$ 는 개인 i 가 실제로 획득하는 특성조합을 의미하고, 이때 간접효용(indirect utility)은 (식 7)로 나타낼 수 있다.

$$U_{ij} = u_i(z_j) + \epsilon_{ij} \quad (\text{식 7})$$

여기서 $u_i(z_j)$ 는 사업장 특성 z_j 로 설명되는 체계적 효용(deterministic component)이고, ϵ_{ij} 는 관측되지 않는 정보 등을 포괄하는 비체계적 효용(stochastic component)이다.

주목할 점은 ϵ 이 개인별로 하나만 존재하는 것이 아니라 조합(i, j)마다 존재하여 ϵ_{ij} 로 표기된다는 점이다.

개인 i 는 후보들 중 효용이 최대인 사업장을 선택한다고 가정하며, 선택된 대안을 j_i , 이에 따른 선택변수 $x_j^{(i)}$ 는 다음과 같이 표현된다.

$$j_i \in \arg \max_{\{j \in 1, \dots, J\}} U_{ij} \quad (\text{식 8})$$

$$x_j^{(i)} = \begin{cases} 1 & \text{if } j = j_i \\ 0 & \text{otherwise} \end{cases} \quad (\text{식 9})$$

선택 후 실현 효용은 선택된 대안에 대응하는 오차항을 명시하여 (식 10)과 같이 정의할 수 있다.

$$U_i := U_{i j_i} = u_i(z_{j_i}) + \epsilon_{i j_i} \quad (\text{식 10})$$

또한 $z_{j_i} = Bx^{(i)}$ 이므로,

$$U_i = u_i(Bx^{(i)}) + \epsilon_{i j_i} \quad (\text{식 11})$$

즉, ϵ_{ij} 는 사업장 j 별로 존재하며 선택된 대안 j_i 에 해당하는 $\epsilon_{i j_i}$ 가 실현된다는 구조이다.

마지막으로 사업장 초기분양률은 개별선택의 집계를 통해 정의된다. 사업장 j 의 수요 집계($Sales_j$)는 (식 12)와 같이 표현된다.

$$Sales_j = \sum_{i=1}^N x_j^{(i)} \quad (\text{식 12})$$

현실에서 초기분양률은 ‘공급된 세대 중 기간 내 계약된 비율’이므로 사업장 j 의 총 분양세대수를 H_j 라 할 때 (식 13)과 같다.

$$ISR_j = \frac{Sales_j}{H_j} = \frac{1}{H_j} \sum_{i=1}^N x_j^{(i)} \quad (\text{식 13})$$

이로써 본 연구는 Lancaster의 특성 기반 선택 모형($U_{ij}, x_j^{(i)}$)을 통해 사업장의 시장 성과지표인 초기분양률(ISR_j)을 도출하는 연결 구조를 제시한다.

Lancaster의 이론에 따르면 소비자 선택의 핵심은 개별 특성의 크기나 한계효과가 아니라, 특성들이 어떤 구조로 결합되었는가에 있다.

이는 실증 분석에서도 개별 변수의 계수를 추정하는 접근보다, 특성 벡터의 구조적 유사성을 식별하는 접근이 이론적으로 더 적합할 수 있음을 시사한다. 특히 C-space가 직접 관측되지 않는 상황에서는, G-space에서 관측 가능한 재화 벡터 간의 구조적 유사성을 통해 간접적으로 특성 조합의 유형을 파악하는 것이 합리적인 분석 전략이 된다.

이러한 이론적 판단에 따라 본 연구는 주택 사업장 특성 벡터 x_j 를 기준으로 K-means 군집분석을 적용한다.

K-means 군집분석은 사전에 종속변수를 설정하지 않고, 관측치 간 거리와 구조적 유사성을 기준으로 데이터를 분류함으로써, 재화 공간에서의 특성 조합 패턴을 식별하는 방법이다. 이를 통해 도출된 각 군집은 Lancaster의 이론적 틀에서 볼 때, 유사한 특성 조합을 제공하는 주택 사업장 유형으로 해석될 수 있다.

본 연구의 분석 전략은 먼저 G-space에서 구조적으로 유사한 사업장들을 군집으로 분류한 후, 각 군집에 속한 사업장들의 초기분양률을 비교하는 방식으로 구성된다. 군집 간 초기분양률의 차이는 곧 서로 다른 특성 조합에 대한 시장의 선택 결과 차이를 의미하며, 이는 특성 기반 소비자이론의 실증적 함의를 보여준다.

이와 같이 본 연구는 Lancaster의 이론을 단순한 개념적 배경으로 차용하는 데 그치지 않고, 이론이 요구하는 분석 관점을 그대로 연구설계에 반영함으로써, 주택 사업장의 분양 성과를 구조적으로 해석하고자 한다.

IV. 자료 및 분석모형

1. 변수 구성

본 연구는 '21년부터 '25년 7월까지 주택도시보증공사에서 PF 보증 승인을 받은 전국 94곳 사업장을 대상으로 한다.

해당 사업장은 사업부지 매입비와 초기사업비 및 건축비, 제비용 등 단위사업의 총사업비를 주택도시보증공사로부터 보증 승인을 받았으므로 토지확보위험, 인허가 위험 등 초기사업단계의 위험은 감소한 상태이나, 이후 분양위험은 여전히 존재하는 상태이다.

즉, 사업의 미래 현금수입 및 사업성을 담보로 대출받은 사업비에 대한 원리금 상환이 결국 초기 분양률에서 결정되는 구조인 것이다.

목표변수인 초기분양률은 '사업장의 분양개시

후 6개월 차(시점) 누적분양률'로 정의한다. 사업장의 금융 사업구조를 나타내는 TPC는 해당 사업장의 현금흐름표상 지출금(공사비+제비용+기타비)에서 자본부채를 차감한 총사업비, LPR은 사업장의 PF 대출금 중 사업부지 매입비 비중을 나타낸다. 또한 APM은 KB부동산 데이터를 활용하여 해당 지역의 매매가격 종합지수 전월대비 증감률을 나타낸다. SPR은 본 사업장의 분양가격이 인근 시세 대비 상대적 수준을 정량적으로 나타내는 지표이다(〈표 1〉).

그 외 지역수요를 나타내는 주민등록 세대수와 미분양 세대수, 사업장의 전용면적별 세대수와 건폐율, 용적률 변수를 선정하였으며 이들은 각 사업장의 특성벡터 z_j 로 나타낸다.

$$z_j = \{TPC_j, LPR_j, \dots, BCR_j, FAR_j\} \in \mathbb{R}^{12} \quad (\text{식 14})$$

본 연구의 자료는 개인이 아파트를 구매한 선택 결과 ($x_j^{(i)} = 1$)를 기반으로 그에 대한 특성 벡터 z_j 를 구축하였다.

즉, 실증 분석에서 직접 관측 가능한 자료는 개인의 선택변수 $x^{(i)}$ 가 아니라 사업장 초기분양률 (ISR_j) 및 사업장 특성벡터 z_j 이다.

따라서 본 연구의 분석은 G-space의 사업장 단위의 관측 가능한 $\{(ISR_j, z_j)\}_{j=1}^J$ 를 이용하여 수행한다.

2. 모형 설명

K-means는 임의의 군집 중심점(centroid)을 선택하고 해당 중심에 가까운 포인트로 점차적으

〈표 1〉 변수설명

| 구분 | 변수명 | 내용 | 출처 |
|---------|-------|---------------------------------|--------|
| 목표변수 | ISR | 초기분양률: 분양개시 후 6개월 차 사업장 누적분양률 | HUG |
| 금융사업 구조 | TPC | 총사업비: 지출금(=공사비+제비용+기타비)-자본부채 항목 | HUG |
| | LPR | 사업부지 매입비/PF 대출금 | HUG |
| 시공평가 | CRK | 당해년도 건설사 시공 순위 | HUG |
| 시장가격 | APM | 월별 아파트 매매가격 종합지수 증감률 | KB 부동산 |
| | SPR | 본 사업장 분양가격/인근 사업장 분양가격 | HUG |
| 수요기반 | RHH | 해당 지역 주민등록세대수 | KOSIS |
| | UNS | 해당 지역 미분양 세대수 | KOSIS |
| 공급구조 | U60 | 사업장의 전용면적 60㎡ 이하 세대수 | HUG |
| | U6085 | 사업장의 전용면적 60-85㎡ 이하 세대수 | HUG |
| | U85 | 사업장의 전용면적 85㎡ 초과 세대수 | HUG |
| 토지계획 | BCR | 사업장의 건폐율 | KB 부동산 |
| | FAR | 사업장의 용적률 | KB 부동산 |

주: 1) 본 연구의 변수명은 연구자가 분석상 식별과 표기 편의를 위해 임의로 설정한 것이며, 특정한 공식 영문 명칭 또는 표준 약어를 전제하지 않음.

2) HUG, Housing & Urban Guarantee Corporation; KOSIS, Korean Statistical Information Service.

로 이동한 후 최종적으로는 가장 가까운 포인트를 선택하는 군집화 기법이다.

이는 변수 간 단위와 범위 차이, 다중공선성에 따라 군집 해석이 왜곡될 수 있으므로 변수 표준화, 상관관계 분석, 다중공선성 검정 등 전처리 과정이 선수행되어야 한다.

변수 표준화(standardization)는 변수 간 단위 및 범위가 상이할 때 특정 변수가 거리 비교를 지배할 수 있으므로 각 성분 z_{mj} 에 대해 표본 j 개 사업장의 평균, 표준편차를 이용한 Z-score 표준화를 적용하였다.

$$\bar{z}_{mj} = \frac{z_{mj} - \bar{z}_m}{s_m}, m = 1, \dots, r \quad (\text{식 15})$$

\bar{z}_m : m 번째 변수의 표본 평균

s_m : m 번째 변수의 표준편차

상관관계 분석(correlation analysis)은 $\{z_{mj}\}$ 간 상관계수 행렬을 산출하고, 절대 상관계수가 높은 변수쌍을 중심으로 중복(implicit over-weighting) 가능성을 확인하였다.

변수 간 정보 중복이 과도할 경우, 군집 구조가 왜곡될 수 있으므로 상관계수의 절댓값이 0.85 이상인 경우를 ‘매우 높은 선형 중복(near redundancy)’으로 정의하고, 변수를 정제하거나, 유사정보를 제공하는 변수의 동시사용을 지양하였다.

다중공선성(multicollinearity) 검정은 회귀 기

반 진단지표인 분산팽창요인(variance inflation factor, VIF)을 통해 수행하였다. 일반적으로 $VIF > 10$ 인 변수는 다중공선성이 존재하는 것으로 판단하고, 분석에서 제외하거나 대체가능한 변수로 조정하였다.

이를 통해 변수 간 중복 및 공선성으로 인한 결과 왜곡 가능성을 완화하고, 군집별 패턴 비교의 명료성을 확보하였다.

위와 같은 전처리 과정을 수행한 후, K-means 분석을 적용한다. 표준화된 사업장 특성벡터의 유사성에 따라 분류한 후, 각 군집에서 나타나는 초기분양률(ISR)의 차이를 비교함으로써 군집별 특성조합과 분양성과의 관계를 해석한다.

이를 위해 각 사업장 j 의 표준화된 특성벡터를 \bar{z}_j 로 정의하고, K-means의 목적함수는 (식 16)과 같이 군집 내 제곱거리합을 최소화하는 문제로 표현한다.

$$\min_{\{c(j)\}, \{\mu_k\}} \sum_{j=1}^J \|\bar{z}_j - \mu_{c(j)}\|^2, \quad c(j) \in \{1, 2, \dots, K\} \quad (\text{식 16})$$

(식 17)에서 $c(j)$ 는 사업장 j 의 군집 소속, μ_k 는 군집 k 의 중심벡터를 의미한다.

K-means는 ‘할당(assignment)-갱신(update)’의 반복 절차로 구현된다. 할당 단계에서는 군집 중심 $\{\mu_k\}_{k=1}^K$ 하에서 각 사업장 j 를 특성벡터와의 거리가 가장 가까운 군집에 배정한다.

$$c(j) \leftarrow \arg \min_{k \in \{1, \dots, K\}} \|\bar{z}_j - \mu_k\|^2 \quad (\text{식 17})$$

그 결과 $\{c(j)\}$ 가 주어지면 각 군집 중심은 해당 군집에 속한 사업장 특성벡터의 평균으로 갱신된다. 군집 k 의 원소집합을 $C_k = \{j : c(j) = k\}$ 라 하면, 군집 중심은 (식 18)과 같이 정의된다.

$$\mu_k \leftarrow \frac{1}{C_k} \sum_{j \in C_k} \bar{z}_j, \quad k = 1, \dots, K \quad (\text{식 18})$$

이를 반복하여 도출된 군집별 ISR 수준을 비교함으로써 특성 조합과 분양성과와의 관계를 분석한다.

V. 분석 결과

본 장에서는 (i) 군집별 기술통계와, (ii) 전처리 (변수표준화, 상관분석, 다중공선성 검정), (ii) 표준화 특성값 \bar{z}_j 에 기반한 K-means 분석 결과를 제시함으로써 초기분양률 군집을 결정짓는 특성 결합을 해석한다.

1. 기초통계분석

'21년부터 '25년 7월까지 전국 PF 사업장 94 곳에 관한 기초통계는 <표 2>와 같다.

목표변수인 ISR은 0%~100%까지 전 구간에 걸쳐 분포하나 전국 평균 74.53%, 중앙값은 98.96%

3) 본 연구에서 설명변수 간 선형중복(linear redundancy)을 확인하고자 상관행렬을 검토하였으며, 선행연구들(Papp et al., 2023; Sowmya et al., 2025)에서 상관계수 $|r| > 0.85$ 를 중복변수(redundant features) 또는 강한 선형 의존 기준으로 활용한 점을 참고하여 $|r| > 0.85$ 를 매우 높은 선형 중복의 기준으로 정의하였다.

〈표 2〉 기초통계

| 변수(단위) | 평균 | 최솟값 | 중앙값 | 최댓값 | 표준편차 | 변동계수 |
|-------------|---------|---------|---------|-----------|---------|------|
| ISR(% , %p) | 74.53 | 0 | 98.96 | 100 | 35.17 | 0.47 |
| TPC(백만 원) | 409,060 | 156,007 | 355,462 | 1,475,532 | 234,885 | 0.57 |
| LPR(% , %p) | 92 | 41.82 | 95.32 | 124.35 | 18.27 | 0.20 |
| CRK(순위) | 65 | 3 | 38 | 350 | 66 | 1.02 |
| APM(% , %p) | 0.27 | -1.71 | 0.07 | 3.88 | 0.92 | 3.41 |
| SPR(% , %p) | 98.93 | 57.93 | 101.44 | 135.61 | 10.45 | 0.11 |
| RHH(세대) | 235,595 | 19,445 | 240,231 | 541,349 | 134,570 | 0.57 |
| UNS(세대) | 460 | 0 | 177 | 4,071 | 800 | 1.74 |
| U60(세대) | 60 | 0 | 0 | 658 | 127 | 2.12 |
| U6085(세대) | 615 | 0 | 656 | 1,372 | 337 | 0.55 |
| U85(세대) | 129 | 0 | 0 | 1,078 | 202 | 1.57 |
| BCR(% , %p) | 21.70 | 10.24 | 16.60 | 66.33 | 13.35 | 0.62 |
| FAR(% , %p) | 253.08 | 99.11 | 219.38 | 741.12 | 126.15 | 0.50 |

주 : 본 연구의 변수명은 연구자가 분석상 식별과 표기 편의를 위해 임의로 설정한 것이며, 특정한 공식 영문 명칭 또는 표준 약어를 전제하지 않음.

로 매우 높게 확인된다.

이는 다수의 사업장에서 초기분양률 100%에 근접한 수준에 집중되어 있는 반면, 일부 사업장의 낮은 초기분양률이 평균을 하방으로 견인하는 분포 특성이 존재함을 보여준다. 또한 표준편차는 35.17%p로, 평균 대비 약 47% 수준의 변동 폭에 해당하여 사업장 간 초기분양률의 이질성이 큰 편임을 나타낸다.

TPC는 넓은 범위의 총사업비 분포를 나타내지만, 평균 409,060백만 원으로 중앙값 355,462백만 원보다 높아, 일부 대형 사업장이 평균을 끌어올린 우측 꼬리 형태 분포로 해석된다. 표준편차는 평균 대비 약 57% 수준으로 사업장 규모 차이가 큰 것으로 추정된다.

LPR은 사업부지 매입비와 PF 대출금을 동시에 반영하는 자금 구조 비율로서, 평균 92%, 중앙값 95.32%로 ‘사업부지 매입비 \approx PF 대출금’ 중심을 형성하되, 최대 124.35%까지 확장되어 사업부지매입비가 PF 대출금을 초과하는 관측치가 존재함을 확인할 수 있다. 이는 일부 사업장에서 PF 대출금이 사업부지 매입비를 충분히 커버하지 못해, PF 대출금 외 자금 투입 또는 보수적 대출 설정과 결부된다.

특히 본 연구에 적용된 사업장의 CRK 분포 특성상 시공사 순위가 낮은 건설사 위주로 형성되어 있는데, 이는 대부분 대기업 건설사는 자체 신용도와 재무 능력을 바탕으로 PF 보증을 제공하는 반면, 신용도와 재무 능력이 낮은 건설사는 HUG

(Housing & Urban Guarantee Corporation) 보증에 가입하였음을 보여준다. 앞서 언급한 LPR과 연계하자면 건설사 시공 순위는 금융기관 평가에 따라 PF 대출금의 규모 및 조건이 좌우되어 사업 부지 매입비 대비 PF 재원 조달 정도에 차이를 유발할 수 있음을 보여준다.

APM은 매매가격 변동률을 나타내는 거시경제 지표로서 대부분 0% 부근의 완만한 등락을 보여준다. 또한 SPR은 인근 사업장 분양가 대비 유사한 수준으로 형성(평균 98.93%, 중앙값 101.44%)되나 분포 범위가 넓어 일부 사업장에서 할인 또는 프리미엄 가격 책정이 존재하는 것으로 보인다.

또한 수요기반 RHH는 중·소규모 지역부터 대규모 지역까지 사업장이 고르게 분포되어 있으나, UNS는 미분양이 없는 지역(최솟값 0)부터 최대 미분양 4,071세대 지역까지 큰 변동성을 나타냈다.

또한 U60, U85는 사업장의 전용면적에 따라 절반 이상의 해당 타입 세대수가 존재하지 않는 것으로 나타났다(중앙값 0). 그러나 U6085는 중앙값이 656세대, 변동계수가 0.55로 나타나 사업장 간 세대수가 비교적 일정하고, 안정적인 비중을 차지한다.

BCR은 평균 21.7%로 중앙값을 상회하며, 최소 10.24%~최대 66.33%의 범위와 표준편차 13.35%p를 보였다. 평균이 중앙값을 상회하는 점은 일부 고건폐율 관측치가 평균을 상방으로 견인하는 우측 치우침을 시사한다. FAR 역시 평균 253%가 중앙값보다 높고, 최소 99.11%~최대 741.12%로 범위가 넓으며 표준편차 126.15%p로 관측되어 이질성이 매우 큰 분포 특성을 보여준다.

2. 데이터 전처리

K-means 분석에 앞서 (i) 변수 표준화, (ii) 상관분석, (iii) 다중공선성 검정을 수행한다.

먼저 기존 자료의 정보는 그대로 유지하되 변수 스케일을 일괄처리하고자 Z-score 표준화를 적용하였다.

상관분석은 피어슨 상관계수(Pearson correlation coefficient)의 $|r| \geq 0.85$ 를 기준으로 정보 중복 최소화를 지향하였다.

분석 결과, ISR은 UNS와 유의미한 음(-)의 상관관계($r=-0.37$, $p=0.00$)를 보인 반면, ISR과 APM, SPR, LPR 등 다수 변수와의 상관관계가 크지 않았으며 5% 유의수준에서 통계적으로 유의하지 않았다. 한편 설명변수 TPC와 U6085, U85의 유의미한 양(+)의 상관($r=0.48$, $p=0.00$, $r=0.72$, $p=0.00$)이 관측되었으며, BCR과 FAR도 유의미한 양(+)의 상관($r=0.76$, $p=0.00$)관계가 나타났다(<표 3>).

VIF을 활용한 다중공선성 검정 결과, 변수별 VIF는 대체로 2~7 수준에서 분포하였으며, 그중 LPR의 VIF=6.86으로 나타나 가장 높은 것으로 확인되었다. 또한 SPR(6.58), TPC(6.04), FAR(5.51), U6085(5.49), BCR(5.01) 등이 5를 소폭 상회하였으나, VIF 10 이상에 해당하는 변수는 관측되지 않았다(<표 4>).

이는 전반적으로 변수 간 과도한 선형 종속성이 지배적이지 않음을 의미한다. 본 연구의 핵심 방법론은 거리 기반 비지도 학습인 K-means 군집분석이므로, VIF 값은 변수 제거를 위한 절대적 기준이라기보다 변수 구성의 과도한 중복 여부를

〈표 3〉 상관분석 결과

| | ISR | TPC | LPR | CRK | APM | SPR | RHH | UNS | U60 | U6085 | U85 | BCR | FAR |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------|------|
| ISR | 1.00 | | | | | | | | | | | | |
| TPC | -0.11 | 1.00 | | | | | | | | | | | |
| LPR | 0.16 | 0.04 | 1.00 | | | | | | | | | | |
| CRK | -0.02 | -0.25 | -0.12 | 1.00 | | | | | | | | | |
| APM | 0.18 | -0.22 | -0.04 | 0.17 | 1.00 | | | | | | | | |
| SPR | 0.17 | 0.13 | 0.07 | -0.07 | -0.21 | 1.00 | | | | | | | |
| RHH | 0.16 | 0.27 | 0.07 | -0.22 | -0.17 | -0.13 | 1.00 | | | | | | |
| UNS | -0.37 | 0.29 | 0.08 | -0.15 | -0.29 | -0.05 | 0.15 | 1.00 | | | | | |
| U60 | -0.11 | 0.00 | 0.08 | 0.13 | 0.03 | -0.09 | -0.01 | 0.32 | 1.00 | | | | |
| U6085 | -0.02 | 0.48 | 0.03 | -0.16 | -0.29 | 0.02 | 0.10 | 0.33 | 0.01 | 1.00 | | | |
| U85 | 0.04 | 0.72 | -0.05 | -0.14 | -0.19 | 0.18 | 0.10 | 0.07 | -0.14 | 0.08 | 1.00 | | |
| BCR | -0.01 | -0.11 | 0.03 | 0.06 | 0.23 | -0.04 | -0.04 | -0.19 | -0.03 | -0.48 | -0.08 | 1.00 | |
| FAR | -0.11 | 0.11 | -0.05 | 0.07 | 0.18 | -0.01 | -0.02 | -0.06 | -0.05 | -0.30 | 0.02 | 0.76 | 1.00 |

주 : 1) ISR과 UNS, TPC와 U60, U6085, U85, BCR, FAR은 5% 이내 유의미한 것으로 나타남.

2) 본 연구의 변수명은 연구자가 분석상 식별과 표기 편의를 위해 임의로 설정한 것이며, 특정한 공식 영문 명칭 또는 표준 약어를 전제하지 않음.

〈표 4〉 다중공선성 검정 결과

| 변수 | VIF | 변수 | VIF | 변수 | VIF | 변수 | VIF |
|-----|------|-----|------|-------|------|-----|------|
| TPC | 6.04 | APM | 1.52 | UNS | 2.24 | U85 | 4.23 |
| LPR | 6.36 | SPR | 6.58 | U60 | 2.07 | BCR | 5.01 |
| CRK | 4.04 | RHH | 4.25 | U6085 | 5.49 | FAR | 5.51 |

주 : 1) 본 연구의 변수명은 연구자가 분석상 식별과 표기 편의를 위해 임의로 설정한 것이며, 특정한 공식 영문 명칭 또는 표준 약어를 전제하지 않음.

2) VIF, variance inflation factor.

점검하는 참고 지표로 해석하였다.

따라서 본 연구는 상기 수준의 VIF 분포를 고려할 때 군집분석 수행에 구조적 제약이 크지 않다고 판단하고, 동일 변수 집합을 유지한 채 후속 K-means 군집분석을 진행하였다.

3. K-means 분석 결과

군집 수(k) 결정을 위해 k=2~6 범위에서 실루엣 지수를 비교한 결과, k=3에서 실루엣 값이 0.47로 가장 높게 나타나 군집 간 분리도가 최대인 것으로 확인되었다. k=4의 실루엣 값(0.45)도

유사한 수준이나, k=5~6에서는 실루엣 값이 0.38~0.36으로 하락하여 군집이 세분화될수록 군집 간 중첩이 증가하는 경향이 관찰되었다. 이에 본 연구는 군집 구조의 분리도와 해석 가능성을 고려하여 k=3을 최적 군집 수로 설정하였다(〈표 5〉).

본 연구는 외부 위험 레이블이 부재한 상황에서 초기분양률이 50% 미만인 사업장을 저분양 위험 사건으로 정의하고, k=3 기준으로 한 K-means 분석 결과는 〈표 6〉과 같다.

먼저 군집 내 평균 초기분양률은 군집 1: 85.58%, 군집 2: 70.52%, 군집 3: 40.24% 순으로 구분되었다. 이러한 분리 패턴은 단일 변수의 크기보다는 지역 수급여건(미분양·공급 규모), 시장 국면(가격 흐름), 상품구성(평형 비중), 가격경쟁력(인근 대비 분양가 수준), 개발 밀도(용적률, 건폐율) 및 사업 규모(총사업비)의 조합에서 기인하는 것으로 해석된다.

군집 1은 평균 ISR이 85.58%로 가장 높았으며, 사업장 수는 67개로 나타났다. 그중 위험으로 정의되는 ‘초기분양률 50% 미만 사업장 수’는 5곳(7.46%)인 것으로 나타났다. 또한 UNS가 291세대로 낮고 RHH도 135,648세대로 상대적으로

작아 수급 부담이 제한적인 환경에서 형성된 군집으로 파악된다. 상품구성은 U6085가 660세대로 가장 높고, U60은 35세대로 낮아 중형 중심의 안정적 평형구성이 특징이다. 가격 수준은 SPR이 100.77로 인근과 유사한 수준이며, APM은 +16%로 월별 아파트 매매가격 지수의 완만한 상승 흐름에 해당한다. 또한 CRK가 41위로 양호한 수준을 보여, 군집 1은 전반적으로 수급·상품설계 등 여건이 상대적으로 안정적인 조합에서 높은 분양률이 관측되는 군집으로 요약된다.

군집 2는 평균 ISR이 70.52%로 중간 수준이다. 이 군집에 속한 사업장 수는 15개이며, 그중 ‘초기분양률 50% 미만 사업장 수’는 2곳으로, 해당 비율이 13.33%임을 확인하였다. 또한, 개발 밀도 측면에서 BCR(34.77), FAR(383.49)로 세 군집 중 가장 높아 고밀 개발 특성이 두드러진다. 동시에 SPR이 95.71로 나타나 인근 대비 낮은 분양가(가격 경쟁력)가 확인된다. 상품구성은 U60이 144세대(26%)로 세 군집 중 소형 비중이 가장 높게 나타났다. 군집 2는 중분양·고밀도 개발 및 상대적으로 낮은 분양가로 인한 가격경쟁형으로 군집으로 구분된다. 반면 CRK는 169위로 상대적으로 낮고, APM은 +76%로 상승 폭이 큰 국면에 해당함에도 ISR은 군집 1보다 낮다. 이는 군집 2가 ‘가격 경쟁력’과 ‘고밀·소형 중심의 상품구성’이라는 특성을 동시에 가지되, 분양성과는 이들 특성의 결합하에서 군집 1만큼의 고분양 수준에는 도달하지 않는 중간 군집으로 정리될 수 있음을 보여준다.

군집 3은 평균 ISR이 40.24%로 가장 낮았으며, 사업장 수 12개 중 9개(75%)의 사업장이 ‘초

〈표 5〉 군집설정

| 군집 | 실루엣 | 분양률 차이 |
|-----|------|--------|
| k=2 | 0.41 | 6.3 |
| k=3 | 0.47 | 9.1 |
| k=4 | 0.45 | 11.5 |
| k=5 | 0.38 | 13.2 |
| k=6 | 0.36 | 15.7 |

〈표 6〉 K-means 분석 결과

| 변수 | | 군집 1 | 군집 2 | 군집 3 |
|-----------|------------------------|------------|----------|----------|
| 사업장 수 | | 67 | 15 | 12 |
| ISR(%) | 군집 내 평균 초기분양률 | 85.58% | 70.52% | 40.24% |
| | 군집 내 '초기분양률<50%' 사업장 수 | 5 | 2 | 9 |
| | 군집 내 '초기분양률<50%' 비율 | 7.46% | 13.33% | 75.00% |
| TPC(백만 원) | 총사업비 | 385,253 | 296,783 | 761,704 |
| LPR(%) | 사업부지 매입비/PF 대출금 | 99% | 74% | 88% |
| CRK(순위) | 당해년도 건설사 시공 순위 | 41 | 169 | 33 |
| APM(%) | 월별 아파트 매매가격 종합지수 증감률 | 16% | 76% | -31% |
| SPR(%) | 본 사업장 분양가/인근 사업장 분양가 | 100.77 | 95.71 | 98.08 |
| RHH(세대) | 해당 지역 주민등록세대수 | 135,648 | 222,279 | 313,687 |
| UNS(세대) | 해당 지역 미분양 세대수 | 291 | 351 | 2506 |
| U60(세대) | 전용면적 60㎡ 이하 세대수 | 35(4.4%) | 144(26%) | 192(14%) |
| U6085(세대) | 전용면적 60~85㎡ 이하 세대수 | 660(83.5%) | 422(76%) | 859(63%) |
| U85(세대) | 전용면적 85㎡ 초과 세대수 | 95(12%) | 58(10%) | 314(23%) |
| BCR(%) | 사업장 건폐율 | 19.38 | 34.77 | 16.04 |
| FAR(%) | 사업장 용적률 | 207.64 | 383.49 | 266.36 |

주 : 1) 각 군집의 전용면적당 세대수 비율을 괄호로 표기. 예: 군집 1의 전용면적 60㎡ 이하 세대수 비율은 $\{35/(35+660+95)\} \times 100 = 4.4\%$.
 2) 본 연구의 변수명은 연구자가 분석상 식별과 표기 편의를 위해 임의로 설정한 것이며, 특정한 공식 영문 명칭 또는 표준 약어를 전제하지 않음.

기분양률 50% 미만'으로 나타났다. 또한, UNS가 2,506세대로 매우 높아 과잉공급 및 미분양 누적 부담이 큰 군집으로 확인된다. 시장 국면 역시 APM이 -31%로 가격 하락 흐름에 해당하여, 수급 부담과 함께 하방 국면이 결합된 환경에서 낮은 분양률이 관측되는 패턴이 나타난다. 사업 규모를 나타내는 TPC는 761,704백만 원으로 가장 커 대규모 사업의 비중이 높으며, 평형구성은 U85가 314세대(23%)로 상대적으로 중대형 비중이 높은 구조가 관찰된다. 특히 군집 3은 CRK가 33위로 양호함에도 불구하고 ISR이 가장 낮아, 이 군집에서는 시공사 요인보다 지역 수급(미

분양·공급) 및 시장 하락 국면이 초기분양률을 보다 강하게 제약하는 패턴이 두드러진다.

초기분양률의 이질성은 단일 요인보다 '수급여건(미분양·공급)-시장국면(가격흐름)-상품구성(평형비중)-가격경쟁력(SPR)-개발밀도 및 사업규모'의 결합 구조에서 명확히 구분되었다.

종합하면, 군집 1은 고분양·안정형으로서 수급 부담이 낮고 중형평수 중심으로 설계되었다고 정의할 수 있다. 마지막으로 군집 3은 높은 미분양과 큰 공급 규모, 가격 하락 국면, 대규모 사업 및 중대형 비중이 동반될 때 분양률이 현저히 낮아지는 공통 패턴을 보여, 분양성과의 위험 신호

를 군집 차원에서 식별할 수 있음을 시사한다.

VI. 결론 및 시사점

본 연구는 초기분양률을 PF 위험관리의 핵심 선행지표로 설정하고, 이를 개별 요인의 평균적 한계효과가 아니라 사업장이 제공하는 특성 조합(bundle)에 대한 시장의 집합적 선택 결과로 해석하였다.

이를 위해 재화를 특성의 결합체로 보는 Lancaster(1966)의 특성 기반 소비자이론에 근거한다. 즉, 가격 자체의 한계효과 추정에 머무르기보다, 관측 가능한 사업장 특성 벡터가 어떤 조합을 이룰 때 시장에서 수요가 빠르게 집계되어 초기분양률로 나타나는지에 초점을 두었다.

이에 '21년부터 '25년 7월까지 전국 PF 보증 사업장 94곳을 대상으로 비용·시장·상품 특성 변수를 구성하고 K-means 분석을 수행하여 초기분양률이 유사하게 형성되는 구조유형(특성 조합의 공통 패턴)을 식별하였다.

분석 결과, 사업장은 세 개 군집으로 구분되며 군집별로 분양 흡수 속도(초기분양률)와 PF 위험 민감도가 달라지는 양상이 확인되었다.

군집 1은 초기분양률이 85.58%로 가장 높고 미분양이 가장 낮으며(291세대), 공급규모(군집 1의 각 전용면적 세대수 총합: 약 790세대)가 과도하지 않은 수준에서 전용면적 60~85㎡ 중형 중심의 상품 구성(83.5%), 인근 대비 유사한 가격 수준(SPR 100.77), 그리고 안정적·완만한 시장 국면(APM +16%)이 결합된 고분양·안정 흡수형

으로 해석된다. 이는 분양대금 조기 유입 가능성이 높아 현금흐름 안정성이 상대적으로 높고, PF 구조가 외부 충격에 대해 비교적 견조할 수 있음을 시사한다.

군집 2는 초기분양률 70.52%로 중간 수준이지만, 건폐율 34.77%, 용적률 383.49%에 따른 고밀도 개발 특성 및 전용면적 60㎡ 이하 소형 비중 확대, 그리고 인근 대비 상대적으로 낮은 가격 포지셔닝(SPR 95.71)이 결합되어 경쟁적 조건에서 수요를 흡수하는 가격·밀도 경쟁형인 것으로 나타났다. 다만 이 유형은 초기분양률이 일정 수준 확보되더라도 개발 밀도와 상품 구성에서 비롯되는 공급 부담이 내재할 수 있어, 시장 국면 변화에 따라 위험도가 가변적일 가능성이 있다.

반면 군집 3은 대규모 공급(군집 3의 각 전용면적 세대수 총합: 약 1,365세대) 및 사업비가 관측되고, 가격 하락 국면(APM -31%)에서 미분양이 급증(2,506세대)하면서 초기분양률(40.24%)이 가장 낮아지는 대규모·하락장 위험형 패턴을 보였다. 이는 분양대금 유입 지연과 현금흐름 불확실성이 구조적으로 확대되는 조합으로서, 분양성과 악화와 PF 위험 신호가 군집 차원에서 뚜렷하게 구분됨을 의미한다.

또한 '초기분양률 50% 미만'인 사업장을 저분양 위험으로 정의하고 군집별 분포를 비교한 결과, 군집 1은 평균 초기분양률이 가장 높고, 군집 내 '초기분양률 50% 미만 사업장' 비율이 7.46%로 가장 낮게 나타났다. 반면 군집 3은 평균 초기분양률이 40.24%로 가장 낮고, 그중 '초기분양률 50% 미만 사업장' 비율이 75%로 가장 높게 나타났다. 이는 저분양 위험이 군집 3에 집중되어 있음

을 보여주며, 본 연구의 분석 결과가 단순한 초기 분양률 수준 차이의 제시에 그치지 않고 저분양 사업장을 사전적으로 식별할 수 있음을 시사한다.

이러한 결과들은 본 연구의 핵심 질문 “어떤 구조와 특성 조합을 가진 사업장이 시장에서 상대적으로 빠르게 수요를 흡수하며 위험에 더 강한가”에 대해 다음과 같은 함의를 제공한다.

첫째, 빠른 수요 흡수는 가격, 규모, 면적 등 단일 요인의 우열로 설명되기보다, 중형 중심의 상품 구성×과도하지 않은 공급규모×인근 대비 급격히 이탈하지 않는 가격×안정적 시장 국면이 함께 결합될 때 가장 뚜렷하게 나타났다(군집 1). 이는 시장 적합성이 ‘개별 변수의 합’이 아니라 정합적인 조합의 결과임을 뒷받침한다. 둘째, 경쟁적 흡수 방식(군집 2)은 고밀도·소형평수 상품설계·총사업비 조합을 통해 성립할 수 있으나, 이는 시장 국면 변화 시 흡수 속도와 위험이 동행할 가능성이 있어, 강건성 측면에서는 군집 1과 구별되는 특성을 갖는다. 셋째, 대규모 공급·시장 하락 국면·미분양 누적·중대형 비중 확대 조합은 초기분양률 저하와 결합하여 위험이 확대되는 경로를 형성하며(군집 3), 초기분양의 ‘위험 신호’를 사업장 단위가 아니라 유형(군집) 단위로 식별할 수 있음을 시사한다.

본 연구는 “무엇이 평균적으로 유의한가”를 넘어, 시장 흡수력을 좌우하는 특성 조합의 구조를 제시함으로써 초기분양률 기반 위험 이해를 한 단계 확장하였다. 또한 주택사업의 분양성과인 초기 분양률을 소비자 선택의 집계 결과로 해석하고, Lancaster의 이론적 틀을 PF 위험관리 맥락과 연결하여 “특성 조합의 구조”를 군집 기반 실증 방법

론으로 제시했다는 데 학술적 의의가 있다.

실무적으로는 보증·금융기관에서 이와 같은 유형별 위험 신호를 별도로 반영하고, 초기분양률이 낮은(예: 군집 3) 사업장에 대한 모니터링 강도 및 심사 조건을 차등화함으로써, 위험 관리를 사후 수습이 아닌 사전 탐지·예방으로 전환할 가능성을 제시한다. 다만 본 유형화는 현재 분석에 사용한 데이터 범위와 구축 가능한 변수에 기반해 도출된 결과이므로, 이후 다른 시기·다른 사업장 데이터에서도 같은 유형 구분이 유사하게 반복되는지를 확인하고, 신규 사업장의 유형을 자동 판별할 수 있는 분류모형을 마련하여 실제 위험 식별력에 관한 추가 검증이 필요하다.

그럼에도 본 연구가 제시한 군집별 패턴은 “무엇이 평균적으로 유의한가”를 넘어, “어떤 특성 조합을 가진 사업장이 시장에서 상대적으로 빠르게 수요를 흡수하며 PF 위험에 더 강한가”라는 질문에 대해 구조적 답을 제공하며, 더 나아가 사업 기획 단계에서도 개별 요소의 단순 조정이 아니라 규모·상품 구성·가격 포지셔닝·시장 국면의 정합성을 패키지로 점검하는 방향으로 의사결정을 유도한다.

끝으로 본 연구에서 활용하고 있는 자료는 PF 보증 승인 사업장이므로 보증 승인 이후 단계에서의 분양 위험 관리로 일반화 범위를 제한한다는 점과 시장 국면의 동태적 변화 및 분양 진행 경로(시간에 따른 흡수 속도)를 충분히 반영하지 못했다는 점에서 한계가 존재한다. 또한 군집분석은 유형의 구조를 제시하는 데 강점이 있으나, 사업장의 초기분양률에 관한 예측력을 정량적으로 비교하는 데에는 추가 검증이 필요하다.

향후 연구에서는 외부 표본을 활용한 재현성 점검과 더불어, 군집 정보를 조기경보 지표·예측 모형과 결합하여 '사업장 기획-심사-사후관리' 전 과정에서 활용 가능한 실증적 기준으로 확장할 필요가 있을 것이다.

ORCID

김은미 <https://orcid.org/0000-0003-3385-8145>

참고문헌

1. 권순재, 김성현, 탁은식, 정현희. (2017). K-Means Clustering 알고리즘과 헤도닉 모형을 활용한 서울시 연립·다세대 군집분류 방법에 관한 연구. *지능정보 연구*, 23(3), 95-118.
2. 김리영, 서원석. (2016). 공동주택 초기계약률의 결정요인 분석연구: 정책 및 시장여건을 중심으로. *GRI 연구논총*, 18(3), 23-43.
3. 김용은, 노승한, 김성환. (2025). 주택개발 PF사업 부실화영향 요인 분석 연구. *부동산도시연구*, 18(1), 137-155.
4. 김지영, 이상경. (2018). 데이터 마이닝을 이용한 아파트 초기계약 예측모형 개발: 위례 신도시 미분양 아파트 단지를 사례로. *한국디지털정책학회논문지*, 16(12), 217-229.
5. 백민석, 신종철. (2011). 분양아파트의 초기계약률 결정요인에 관한 연구. *도시행정학보*, 24(1), 213-237.
6. 서진호, 강정규. (2014). 신규 아파트 초기 분양률 결정요인 분석에 관한 연구. *주거환경*, 12(2), 153-164.
7. 손재영. (2009). 주택공급제도 개선방안. *KDI 정책 포럼*, 213.
8. 유정근, 오동훈. (2010). 부실화된 부동산 PF 사업장의 효율적인 관리방안: 공동주택 PF부실사업장 관리 사례를 중심으로. *부동산학연구*, 13(3), 90-114.
9. 하은혜, 이견학. (2024). 시계열 군집 분석을 활용한 주택 하위시장 유형화: 수도권 아파트의 매매가격 지수를 중심으로. *한국지도학회지*, 24(3), 29-44.
10. Lancaster, K. J. (1966). A new approach to consumer theory. *The Journal of Political Economy*, 74(2), 132-157.
11. Papp, L., Rasul S., Spielvogel, C. P., Krajnc, D., Poetsch, N., Woehrer, A., Patronas, E. M., Ecsedi, B., Furtner, J., Mitterhauser, M., Rausch, I., Widhalm, G., Beyer, T., Hacker, M., & Traub-Weidinger T. (2023). Sex-specific radiomic features of L-[S-methyl-11C] methionine PET in patients with newly-diagnosed gliomas in relation to IDH1 predictability. *Frontiers in Oncology*, 13, 986788.
12. Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *The Journal of Political Economy*, 82(1), 34-55.
13. Sowmya, V. L., Bharathi Malakreddy, A., Natarajan, S., Prathik, N., & Rajesh, I. S. (2025). Spectral entropic radiomics feature extraction (SERFE): An adaptive approach for glioblastoma disease classification. *Frontiers in Artificial Intelligence*, 8, 1583079.

논문 접수일: 2026년 1월 29일

심사(수정)일: 2026년 3월 29일

게재 확정일: 2026년 4월 17일

국문초록

초기분양률은 현금흐름의 실현가능성을 좌우하여 PF(project financing) 사업장의 위험을 사전에 가늠할 수 있는 중요한 신호로 기능한다. 본 연구는 초기분양률을 개별 요인의 결과가 아닌, 주택 사업장이 제공하는 특성 조합에 대해 시장이 내린 집합적 선택의 결과로 이해하는 관점에서 접근한다. 이에 Lancaster(1966)의 특성 기반 소비자이론(characteristics-based consumer theory)을 이론적 틀로 채택하였으며 PF 사업장을 분류(K-means)하여 '무엇이 평균적으로 유의한가'를 넘어, '어떤 특성 조합을 가진 사업장이 시장에서 상대적으로 빠르게 수요를 흡수하여 PF 위험에 더 강인가'라는 질문에 대해 구조적 답을 제공한다. K-means 결과, 초기분양률이 높은 군집은 중형 평수 중심, 과도하지 않은 공급규모, 인근 대비 유사한 수준 가격, 안정적 시장 국면이 함께 결합됨을 확인할 수 있었으며, 변수별 한계효과가 아닌 특성 중심의 관점 전환을 통해 주택 사업장의 분양 성과를 보다 구조적으로 해석하고, 실질적인 PF 위험 관리방안에 기여할 것으로 기대한다.

주제어 : 부동산 PF(project financing), 초기분양률, 특성 기반 소비자이론, K-means, 특성조합