**5. 結果と考察**

本プロジェクトで構築したOLS（最小二乗法）モデルによる中古マンション取引価格予測の結果についてご説明します。

**5.1. モデル評価指標**

モデルの全体的な性能を示す主要な評価指標は以下の通りです。

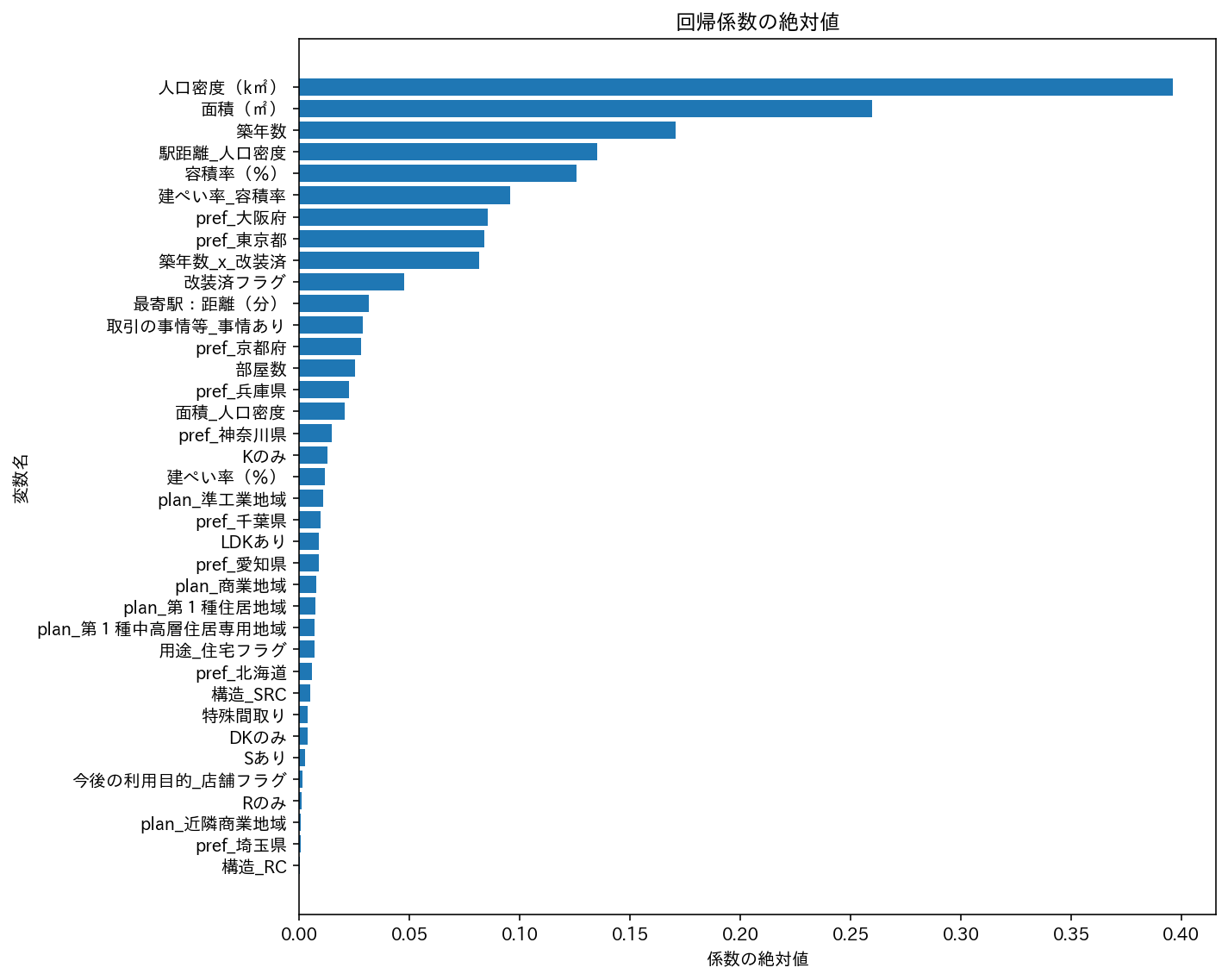
| 指標 | 値 | 解釈 |
| --- | --- | --- |
| **R-squared (決定係数)** | **0.724** | 構築したモデルが、目的変数である「取引価格（総額）の対数」の変動の約72.4%を説明できていることを示します。この値は0から1の範囲で、1に近いほどモデルの当てはまりが良いことを意味します。0.724という値は、中古マンションの価格予測モデルとして比較的高い説明力を有していると評価できます。 |
| **F-statistic** | 3.395e+04 | モデル全体の統計的な有意性を示す指標です。この値が大きいほど、モデルが統計的に有意であることを示します。 |
| **Prob (F-statistic)** | 0.00 | F-statisticに対応するp値です。0.05よりも非常に小さい（ほぼ0）ため、構築したモデルは統計的に極めて有意であり、偶然にこのような結果が得られた可能性は低いと言えます。 |

これらの指標から、提案モデルは中古マンションの取引価格の変動を比較的高い精度で捉えており、予測に有効であると判断できます。

**5.2. 主要な回帰係数と解釈**

以下に、価格予測に特に影響を与えていると見られる特徴量とその回帰係数の一部、およびその解釈を示します。回帰係数は、他の変数を一定としたときに、その変数が1単位変化した際に目的変数（取引価格の対数）がどれだけ変化するかを示します。

**回帰係数の絶対値**

上記のグラフは、各特徴量が目的変数（取引価格の対数）に対してどれだけ大きな影響力を持っているかを、係数の絶対値で示しています。

**主な特徴量の係数と考察**

・人口密度（k㎡）: 0.3959

最も大きな正の係数を持つ特徴量です。人口密度が高い地域ほどマンション価格が高くなる傾向があることを示しており、都市部の物件が高価格であることを強く裏付けています。

・面積（㎡） (対数): 0.2596

物件の面積が広いほど価格が上昇することを示します。対数変換されているため、面積の増加に伴う価格の上昇率は徐々に緩やかになる関係性を捉えています。

・築年数: -0.1706

築年数が古いほど価格が下落する傾向を示します。これは不動産市場における一般的な傾向と一致しています。

・駅距離\_人口密度: -0.1352

「最寄駅：距離（分）」と「人口密度（k㎡）」の交互作用項です。この係数が負であることは、駅からの距離が遠くなるほど、かつ人口密度が高い地域であるほど、価格へのマイナス影響が大きくなる可能性を示唆しています。つまり、人口密度の高い都市部では、駅からの距離が価格に与える影響がより顕著になる、と解釈できます。

・建ぺい率\_容積率: -0.0955

「建ぺい率（％）」と「容積率（％）」の交互作用項です。負の係数であることから、建ぺい率と容積率が高い物件（＝大規模な集合住宅など）が、単独の建ぺい率・容積率の効果に加えて、価格に負の影響を与える可能性があります。これは、郊外の大型団地など、高密度な開発が行われた物件の評価が相対的に低いケースを捉えていると考えられます。

・pref\_東京都 / pref\_大阪府: それぞれ -0.0838 / -0.0854

これらの都道府県ダミー変数の係数が負であることは、やや直感に反するかもしれません。しかし、これはモデルがすでに「人口密度（k㎡）」や「面積（㎡）」、「最寄駅：距離（分）」といった主要な価格決定要因を考慮した上で、それでもなお残る効果を示しています。つまり、これらの大都市圏の物件は、上記の主要因で価格の高さが説明されており、ダミー変数としては基準となる「その他の都道府県」と比較した際の相対的な差分が表れていると解釈できます。

・築年数\_x\_改装済: 0.0815

「築年数」と「改装済フラグ」の交互作用項です。この係数が正であることは、築年数が経過した物件ほど、改装されていることによる価格へのプラス効果が大きいことを示唆しています。単独の「改装済フラグ」の係数は負ですが、これは改装が必要とされる古い物件が改装される傾向を捉えている可能性があり、その上で「築年数\_x\_改装済」が正に作用することで、古い物件の改装が価格の下落を抑制し、あるいは価値を再創造する効果があることを示しています。

・改装済フラグ: -0.0475

他の条件（特に築年数）が同じであれば、単に「改装済み」であること自体は価格をわずかに押し下げる要因となっている可能性があります。これは、改装が必要とされる物件（例えば、築年数が古い、または何らかの不具合があった物件）が改装される、というデータの実態を反映しているものと考えられます。しかし、上記の「築年数\_x\_改装済」の交互作用項と併せて解釈することで、築年数に応じた改装の価値が異なることが分かります。

・Kのみ: 0.0126

キッチンのみの間取り（例: ワンルーム）の物件が、他の要因を考慮しても価格にプラスに寄与する傾向が見られます。これは、都市部の駅近など利便性の高いエリアにおいて、単身者や学生の需要が高く、コンパクトで効率的な空間利用が評価されているためと考えられます。

**5.3. 多重共線性（VIF）の確認**

モデルの安定性を評価するため、VIF（Variance Inflation Factor：分散拡大要因）を計算しました。VIFが10を超えると多重共線性の問題があると言われることがあります。

| 変数名 | VIF値 |
| --- | --- |
| 人口密度（k㎡） | 407.51 |
| pref\_東京都 | 385.26 |
| 建ぺい率\_容積率 | 299.97 |
| 容積率（％） | 213.16 |
| pref\_大阪府 | 94.43 |
| pref\_神奈川県 | 72.69 |
| 駅距離\_人口密度 | 64.91 |
| ... | ... |

複数の特徴量（特に「人口密度（k㎡）」、「pref\_東京都」、「建ぺい率\_容積率」など）でVIF値が高いことが確認されました。これは、これらの特徴量間に強い相関関係があることを示唆しています（例：人口密度が高い地域に東京都の物件が多い、建ぺい率と容積率が互いに密接に関連しているなど）。

多重共線性が高い場合、個々の係数の解釈が難しくなることがありますが、予測性能自体には直接影響しない場合もあります。今回のモデルでは高いR-squaredを維持していますが、今後の改善点として、特徴量選択（例：相関の高い片方を削除）や次元削減の手法（例：主成分分析（PCA））を検討する余地があります。

**6. 今後の展望・改善点**

* **モデルの改善**: OLS以外のより高度なモデル（例: LightGBM, XGBoost, Random Forestなど）の導入を検討し、予測精度の向上を図ります。
* **特徴量エンジニアリングの深掘り**: 更なるドメイン知識に基づいた特徴量の生成や、既存の特徴量の変換方法（例：カテゴリ変数の異なるエンコーディング手法）の検討を行います。
* **ハイパーパラメータチューニング**: 採用するモデルの性能を最大化するためのハイパーパラメータの最適化を行います。
* **アンサンブル学習**: 複数のモデルの予測結果を組み合わせることで、よりロバスト（頑健）な予測モデルを構築し、予測精度の安定化を目指します。
* **多重共線性の対応**: VIF値の高い特徴量については、特徴量選択や次元削減の手法（例: PCA）を検討し、モデルの解釈性や安定性向上を図ります。

**[GitHubへのリンク]**

**https://github.com/yotsuba-pg/nishika-mansion-price-prediction/tree/main**