**Analiza pieței imobiliare din „Ames Housing”, localizate în USA**

IA-212 Palii Flaviu, UTM, FCIM

### ABSTRACT

Această cercetare investighează principalii factori care influențează piața imobiliară din Ames, Statele Unite, folosind Setul de Date *AMES Housing*. Studiul se lansează într-un proces meticulos de colectare a datelor, urmat de pași critici de prelucrare a acestora pentru a asigura cea mai înaltă calitate și fiabilitate a datelor. Acesta implică curățarea, organizarea și transformarea datelor, cu accent deosebit asupra gestionării valorilor lipsă, a valorilor aberante și a conversiei datelor categorice în format numeric prin tehnici precum Codificarea One-Hot și Codificarea Etichetelor. Se acordă o atenție specială aspectelor temporale ale datelor, îmbogățind analiza cu o perspectivă temporală mai profundă.

Cercetarea dezvăluie tendințe semnificative, modele și anomalii prin analiza exploratorie a datelor, contribuind la o mai bună înțelegere a caracteristicilor pieței și a comportamentului clienților. Se dezvoltă un model predictiv, cu scopul de a prognoza tendințele de vânzare pe baza datelor istorice și a factorilor relevanți identificați în faza exploratorie. Acest model trece printr-un proces continuu de rafinare pentru a îmbunătăți precizia și fiabilitatea acestuia, iar performanța sa este evaluată riguros pentru a asigura viabilitatea și precizia sa în prognozarea tendințelor de vânzare.

Principalele constatări includ o corelație pozitivă între calitatea proprietății și preț, cu prețuri mai ridicate asociate proprietăților de calitate superioară. Studiul prezintă, de asemenea, o analiză detaliată a distribuției prețurilor proprietăților și a numărului de camere, oferind informații valoroase despre preferințele pieței și tipurile de proprietăți.

Se dezvoltă și se evaluează două modele predictive. Modelul 1, concentrat pe Anul Construcției și Suprafața Terenului, explică aproximativ 37,5% din variația prețului. Modelul 2, care include Calitatea Generală, îmbunătățește semnificativ precizia predicției, explicând aproximativ 68,8% din variația prețului.

### INTRODUCERE

Setul de date AMES Housing va avea ca scop problema identificarii factorilor cheie care influenteaza vanzarea de case pentru o piata din Statele Unite, Ames.

Problema este legată cu înțelegerea și vizualizarea diferitelor aspecte ale setului de date "AMES", pentru a obține informații despre prețurile locuințelor, calitatea și numărul de camere într-un context anumit.

Se presupune ca scop, identificarea factorilor care amplifică volumul vânzărilor de locuințe, adoptarea unei strategii bazate pe date o serie de etape începând cu adunarea datelor relevante și meticuloasă, ce va servi drept fundament pentru toate analizele ulterioare. Urmează preprocesarea, o etapă critică în care datele sunt curățate, organizate și transformate pentru a asigura cea mai înaltă calitate și fiabilitate a informațiilor.

Cu o bază de date solidă, se trece la analiza exploratorie, o investigație preliminară care își propune să descopere tendințe, modele și anomalii. Această etapă este crucială pentru că ne ajută să înțelegem mai bine caracteristicile pieței și comportamentul clienților. În paralel, se lucrează la crearea unui model predictiv, care își propune să anticipeze tendințele de vânzare pe baza datelor istorice și a factorilor relevanți identificați în etapa de analiză exploratorie.

Urmează îmbunătățirea continuă a modelului prin ajustarea și finisarea acestuia, pentru a mări acuratețea și fiabilitatea predicțiilor sale. Această îmbunătățire este un proces iterativ, unde modelul este testat și retestat, ajustându-se parametrii pentru optimizarea performanței.

Finalmente, evaluarea modelului este esențială pentru a asigura că strategia propusă este viabilă și că modelul poate prezice cu acuratețe tendințele de vânzare. Odată ce modelul a fost validat și se demonstrează că poate contribui la creșterea vânzărilor, strategia poate fi implementată pentru a atinge obiectivul de creștere a numărului de tranzacții imobiliare cu un procentaj semnificativ.

### I. DATE ȘI METODE

Pentru pre-procesarea datelor din setul "ames", această secțiune de raport evidențiază procesul de curățare și pregătire a acestora. Aceste operațiuni sunt cruciale în asigurarea calității și relevanței datelor, fiind fundamentale pentru analizele viitoare și pentru construirea de modele precise.

În procesul de gestionare a datelor, am abordat mai multe etape esențiale pentru a asigura calitatea și consistența acestora, începând cu ***tratarea valorilor lipsă*** și a ***outlier-ilor***. Pentru valorile lipsă din variabilele numerice, am optat pentru înlocuirea acestora cu media sau mediana, sau, în anumite cazuri, am preferat eliminarea rândurilor cu valori lipsă. În cazul variabilelor categoricale, am utilizat strategii adaptate, cum ar fi înlocuirea cu categoriile predominante sau crearea de noi categorii, în funcție de contextul datelor.

Gestionarea outlier-ilor a fost, de asemenea, un aspect crucial, eliminând sau ajustând aceste valori extreme pentru a preveni distorsionarea analizei. Aceasta etapă a fost vitală pentru menținerea integrității și acurateței analizei ulterioare.

În ceea ce privește ***pre-procesarea datelor***categorice, am implementat tehnici de encodare, cum ar fi One-Hot Encoding sau Label Encoding, pentru a transforma aceste variabile în format numeric. Această transformare a fost esențială pentru a permite utilizarea acestor date în algoritmi de învățare automată.

Pe lângă acestea, am acordat o atenție deosebită ***manipulării datelor temporale***, efectuând transformări specifice pentru a extrage informații valoroase și relevante. Am creat noi caracteristici care reflectă aspecte temporale, cum ar fi diferența dintre anul de vânzare și anul construcției, aducând astfel o perspectivă mai amplă și o adâncime mai mare în analiza temporală.

În ultima etapă, am aplicat ***standardizarea și rescalarea datelor numerice***. Această procedură a avut rolul de a uniformiza scala numerică a datelor, facilitând astfel analiza comparativă și integrarea mai eficientă a diferitelor tipuri de date în modelele de învățare automată. Această abordare sistematică și detaliată în pre-procesarea datelor asigură o bază solidă pentru analizele ulterioare și pentru dezvoltarea de modele predictive precise și fiabile.

**Explorarea variabilelor**

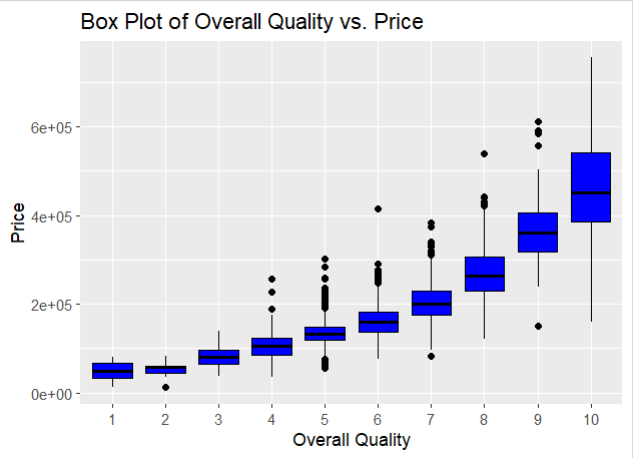
Setul de date conține un total de 2930 de rânduri (observații) și multiple coloane (variabile).

Setul de date mai conține o varietate de tipuri de variabile, inclusiv numerice (e.g., "area," "price," "Year.Built"), caracter (e.g., "MS.Zoning," "Street," "Exter.Qual"), și variabile care au NA-uri (date lipsă) în unele cazuri.

**Vizualizarea datelor**

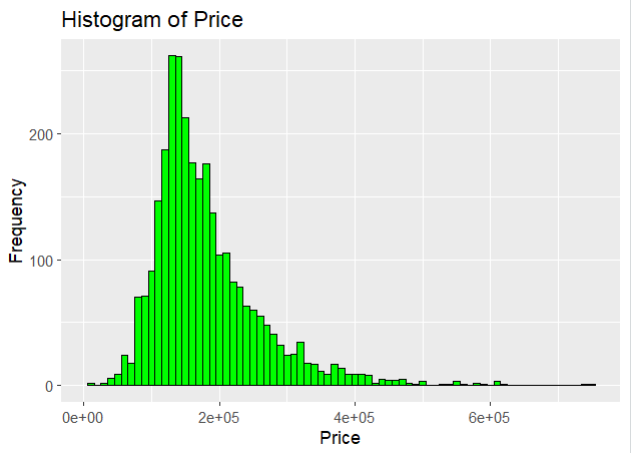
Această asociere poate fi explicată prin faptul că proprietățile cu o calitate mai înaltă sunt adesea asociate cu materiale de construcție superioare, finisaje de lux și localizare preferențială. De asemenea, dispersia mare a prețurilor în categoriile superioare de calitate poate indica o variație considerabilă în caracteristicile individuale ale locuințelor care nu sunt capturate de evaluarea calitativă generală. Totodată, este vizibil că pentru locuințele cu calitatea la nivelul cel mai înalt, există o varietate mare de prețuri, ceea ce sugerează că alți factori, posibil unici pentru fiecare proprietate, influențează prețul final.

Aceste constatări subliniază importanța evaluării amănunțite a proprietăților înainte de stabilirea prețului, având în vedere că standardele de calitate au un impact direct asupra valorii de piață. Observațiile extrase din această analiză boxplot pot fi utile pentru investitori, evaluatori și potențiali cumpărători în procesul de luare a deciziilor imobiliare, furnizându-le o perspectivă cuantificabilă asupra relației dintre calitatea percepută și preț.



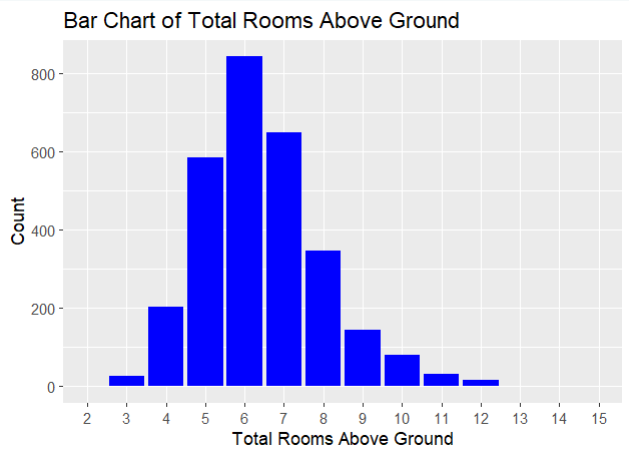
# **Figura 1.1** Distribuția Price in dependenta de OverAll.Qual

Histograma prețurilor locuințelor, ilustrată în **Figura 1.2**, evidențiază o distribuție asimetrică dreapta a prețurilor în cadrul setului de date analizat. Cea mai mare frecvență a prețurilor se situează în intervalul inferior, cu o concentrare vizibilă în jurul unui mod central. Acest lucru indică faptul că majoritatea locuințelor din setul de date sunt evaluate la prețuri mai scăzute, în timp ce un număr mai mic de locuințe ating valori extrem de înalte, așa cum sugerează coada lungă spre dreapta a distribuției.



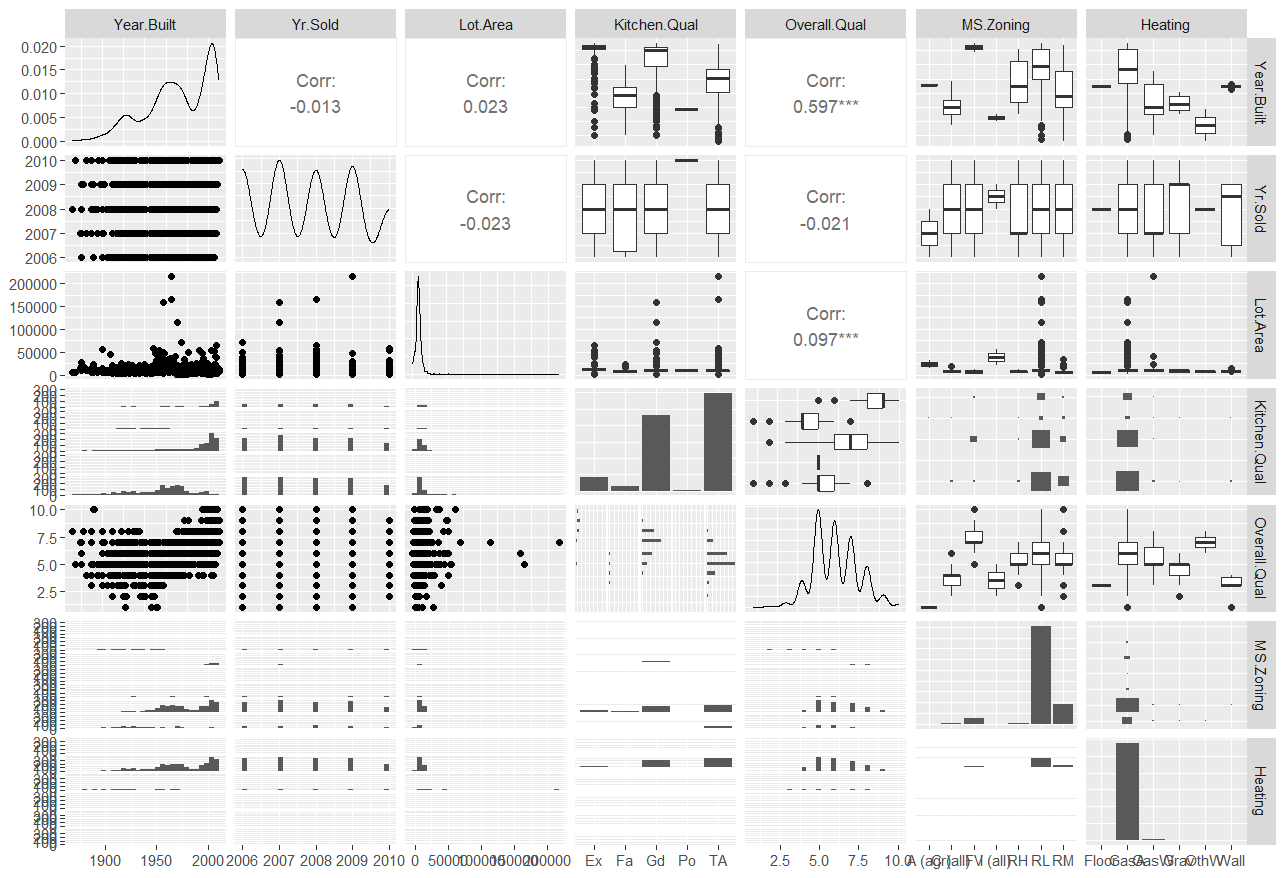
# **Figura 1.2** Distribuția Price in setul de date

**Figura 1.3** prezintă un grafic cu bări care ne arată cât de des apare un anumit număr de camere supraterane în locuințele dintr-un set de date. Cele mai comune locuințe au între 6 și 7 camere deasupra solului, acestea fiind cele mai înalte bări din grafic. Numărul de locuințe scade pe măsură ce numărul de camere crește sau scade față de acest interval comun. Avem mai puține locuințe cu doar 2 sau 3 camere, și la fel, mai puține cu 10 sau mai multe camere. Graficul ne ajută să înțelegem preferințele oamenilor sau tipurile de locuințe disponibile când vine vorba de numărul de camere.



# **Figura 1.3** Distribuția Incaperilor de la nivelul pamantului in setul de date

După faza de preprocesare a datelor, ne angajăm într-o analiză exploratorie detaliată pentru a descoperi modele și perspective în cadrul setului de date Ames Housing. O parte esențială a acestei explorări este vizualizarea relațiilor dintre diverse caracteristici ale locuințelor, ilustrate de diagrama de perechi (sau graficul de corelație) prezentat mai jos.



# **Figura 1.4** Matricea de Perechi al Caracteristicilor Selectate

Această figură este o matrice de grafice de dispersie pentru variabilele selectate, dezvăluind relații pereche între ele. Fiecare grafic oferă o corelație vizuală unde densitatea punctelor și distribuția lor pot sugera natura unei relații. Observații notabile din această figură includ:

* **Anul Construcției vs. Calitatea Generală**: Pare să existe o corelație pozitivă între anul în care a fost construită o casă și calitatea sa generală, sugerând că locuințele mai noi au tendința de a avea evaluări de calitate mai înalte.
* **Suprafața Terenului vs. Calitatea Generală**: Corelația aici este mai puțin evidentă, cu Suprafața Terenului arătând o largă dispersie a valorilor pe diferite evaluări ale calității. Acest lucru indică faptul că, deși Suprafața Terenului este o caracteristică importantă, impactul său asupra calității nu este atât de direct și poate fi influențat de alți factori.
* **Calitatea Bucătăriei vs. Calitatea Generală**: Calitatea bucătăriei pare să coreleze bine cu calitatea generală, ceea ce se aliniază cu așteptările că locuințele de calitate superioară au bucătării mai bine dotate.
* **Variabile Categorice (Zonarea MS, Încălzirea)**: Graficele de tip boxplot pentru variabilele categorice precum Zonarea MS și Încălzirea în raport cu Calitatea Generală oferă perspective asupra modului în care acești factori sunt distribuiți pe diferite niveluri de calitate.

[Acest grafic](#_Figura_1.4_Graficul) de perechi reprezintă un pas semnificativ în analiza noastră exploratorie, oferind o vedere detaliată asupra modului în care caracteristicile individuale sunt interconectate. Aceasta informează alegerile noastre ulterioare de modelare și formularea ipotezelor. Nuanțele captate în această vizualizare subliniază complexitatea evaluării imobiliare și necesitatea unor modele analitice robuste pentru a capta astfel de relații multidimensionale.

Pentru a testa ipotezele, se vor căuta coeficienți de corelație semnificativi și modele în date care pot sugera relații puternice. De exemplu, din [matricea](#_Figura_1.4_Graficul) de mai sus, se observă că variabila "Overall Qual" are o corelație semnificativă de 0.5977\*\*\* cu una dintre variabile, posibil prețul, indicând o relație puternică. Semnele asteriscului (\*\*\*), în general, indică un nivel înalt de semnificație statistică.

In baza acestor corelații, au fost decise variabilele care urmeaza să fie incluse pentru modelele predictive de regresie liniară.

Acest lucru sugerează că în urma analizei corelațiilor, variabilele cu cele mai puternice și semnificative legături cu prețul au fost alese pentru a fi incluse în modelele de regresie liniară pentru a anticipa prețurile proprietăților. Aceste variabile selectate sunt folosite apoi pentru a construi modele care pot oferi previziuni precise cu privire la prețul de piață al proprietăților pe baza datelor istorice și a caracteristicilor relevante identificate.

În urma analizei corelației dintre variabilele studiate, s-au efectuat câțiva pași importanți pentru stabilirea datelor necesare în vederea construirii modelelor predictive de regresie liniară:

**Analiza Corelațiilor** dintre diferite variabile este un pas esențial înainte de a construi modele predictive. În imaginea furnizată, este prezentat un grafic de perechi sau matrice de corelație, diferite variabile sunt comparate între ele pentru a determina relațiile lor.

De exemplu, variabilele precum anul construcției (Year.Built), suprafața terenului (Lot.Area) și calitatea generală a proprietății (Overall.Qual) sunt analizate pentru corelațiile lor cu prețul și posibil cu alte variabile.

**Stabilirea Datelor pentru Modelele Predictive** in baza rezultatelor obținute din analiza corelațiilor, au fost pregătite și selectate date pentru a construi modele de regresie liniară. Aceasta a implicat alegerea variabilelor relevante care ar putea influența cel mai mult prețul proprietăților, pe baza forței și direcției corelațiilor lor.

**Dezvoltarea și Evaluarea Modelelor**

1. Cum variază prețul unei proprietăți în funcție de anul în care a fost construită proprietatea (*Year.Built*) și suprafața terenului (*Lot.Area*)?
2. Care este impactul asupra prețului unei proprietăți în funcție de anul construcției (*Year.Built*), calitatea generală a proprietății (*Overall.Qual*) și suprafața terenului (*Lot.Area*)?
3. Cum se schimbă prețul proprietății în funcție de anul construcției și suprafața terenului (*Lot.Area*)?
4. Care este influența pe care o au calitatea generală a proprietății (*Overall.Qual*), suprafața terenului (*Lot.Area*) și anul construcției (*Year.Built*) asupra prețului unei proprietăți?

Analiza exploratorie a datelor reflectă o tendință interesantă când se examinează distribuția prețurilor în funcție de calitatea generală a locuințelor. Boxplot-ul prezentat în [**Figura 1.1**](#_Figura_1.1_Distribuția) ilustrează o corelație pozitivă între calitate și preț, sugerând că locuințele de calitate superioară tind să aibă prețuri mai ridicate. În mod specific, prețurile cresc semnificativ pentru locuințele care sunt evaluate la o calitate generală de 7 sau mai mult, conform scalei utilizate.

### II. MODELELE DE REGRESIE LINIARĂ ȘI RELEVANTĂ VARIABILELOR

Studiul prezent a fost concentrat pe dezvoltarea și evaluarea a două modele predictive diferite pentru estimarea prețurilor proprietăților imobiliare.

Primul model, denumit *Modelul 1*, a inclus două variabile principale: Anul construcției (Year.Built) și Suprafața terenului (Lot.Area). Performanța acestui model a fost evaluată prin coeficientul de determinare R-squared, care a fost de aproximativ 0.375. Acest rezultat sugerează că Modelul 1 este capabil să explice aproximativ **37.5%** din variația prețului proprietăților. Eroarea modelului a fost măsurată prin Radicul Erorii Medie Pătratice (RMSE), care a fost aproximativ **63507.63**. Această valoare indică gradul de deviație a predicțiilor de preț față de valorile reale ale proprietăților.

Modelul 2 a extins variabilele utilizate în Modelul 1 prin includerea Calității generale a proprietății (Overall.Qual), pe lângă Anul construcției și Suprafața terenului. Performanța acestui model s-a îmbunătățit considerabil, cu un coeficient R-squared de aproximativ 0.688. Acest rezultat indică faptul că Modelul 2 explică aproximativ **68.8%** din variația prețurilor, demonstrând o capacitate superioară de a prezice prețul proprietăților în comparație cu Modelul 1. În plus, eroarea modelului, măsurată prin **RMSE**, a scăzut la aproximativ **44838.49**, indicând o precizie mai mare a predicțiilor.

**Modelele de prezicere cu StepWise**

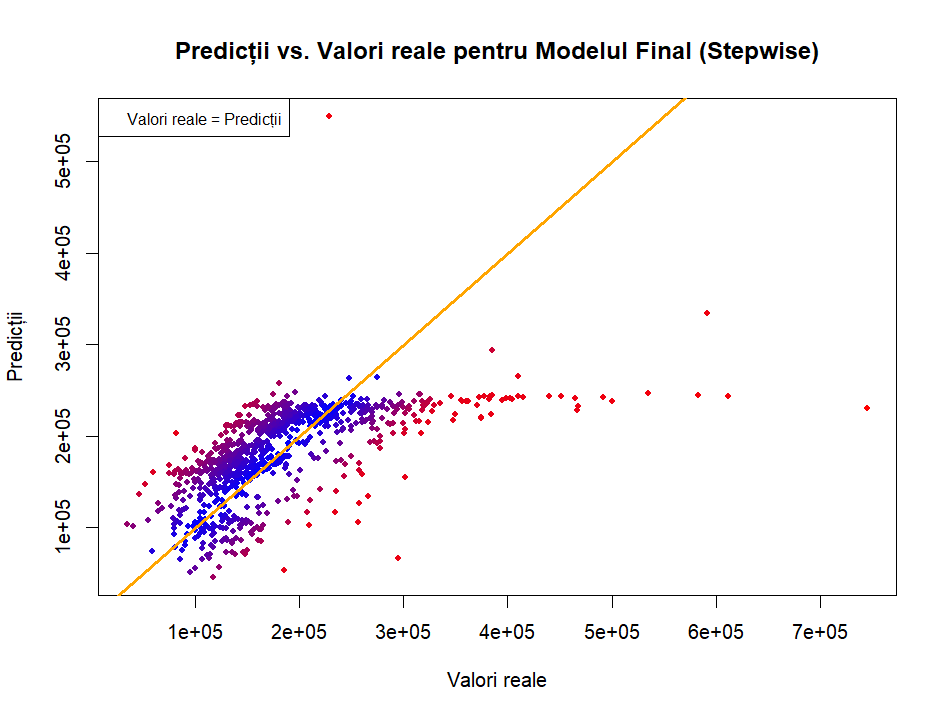
În cadrul procesului de modelare și predicție a prețurilor proprietăților, este esențial să vizualizăm și să evaluăm performanța modelului. Acest lucru se realizează adesea prin compararea valorilor prezise de model cu valorile reale, așa cum se reflectă în graficele de dispersie prezentate. În figurile discutate, fiecare punct reprezintă o proprietate individuală, cu axa absciselor (orizontală) afișând valorile reale ale prețurilor, iar axa ordonatelor (verticală) arătând predicțiile modelului.

O reprezentare grafică permite observarea rapidă a relației dintre predicții și realitate. Linia oranj reprezintă locul în care valorile prezise sunt identice cu valorile reale; punctele situate pe această linie indică **predicții perfecte**. Acest lucru ne ajută să înțelegem cât de apropiate sunt predicțiile de realitate.

Un aspect notabil al acestor grafice este utilizarea unui **gradient de culoare**, care se estompează în roșu pe măsură ce punctele se îndepărtează de această linie. Acest *fade* inspre **roșu** nu este doar o alegere estetică, ci servește un scop analitic precis: accentuează vizual punctele care reprezintă cele mai mari erori de predicție. Cu cât un punct este mai departe de linia galbenă, cu atât eroarea de predicție este mai mare și, prin urmare, punctul respectiv va fi afișat într-o nuanță mai intensă de roșu. Aceasta permite observatorilor să identifice rapid și intuitiv zonele în care modelul este mai puțin precis și să se concentreze pe îmbunătățirea predicțiilor pentru acele cazuri.

Graficul dispersiei prezentat în [Figura 3.1](#_Figura_3.1_Model1) ilustrează relația dintre valorile prezise de ***Modelul 1*** și valorile reale ale proprietăților. Analizând Modelul 1, care utilizează o metodă stepwise pentru a selecta variabile, am inclus Anul construcției (Year.Built) și Suprafața terenului (Lot.Area) ca predictorii principali. Observăm că punctele de date sunt distribuite în jurul liniei identității, unde valorile prezise sunt egale cu valorile reale, dar cu o variație semnificativă.

Evaluarea vizuală și statistică a Modelului 1, cum este ilustrată în figura respectivă, arată potențialul variabilelor selectate de a contribui la explicarea prețurilor imobiliarelor, dar totodată subliniază necesitatea de a explora și alte variabile care ar putea îmbunătăți performanța modelului.

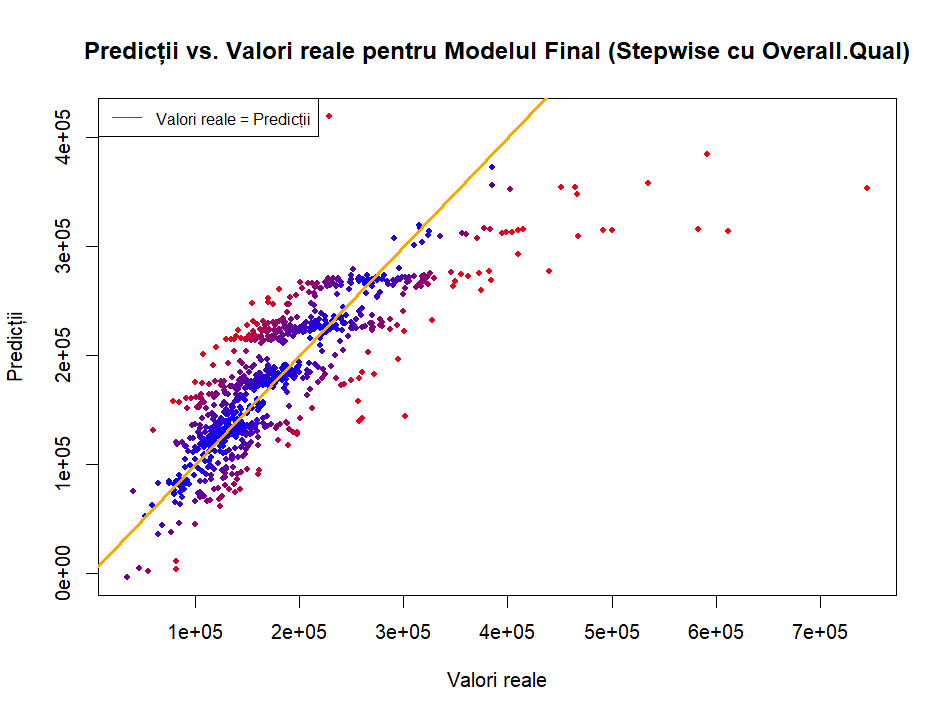


# **Figura 3.1** Model1 cu *Year.Built* si *Lot.Area*

**Modelul 2**, prezentat în [Figura 3.2](#_Figura_3.2_Model1), a fost îmbogățit prin adăugarea variabilei Calitatea generală a proprietății (Overall.Qual) la Anul construcției (Year.Built) și Suprafața terenului (Lot.Area).

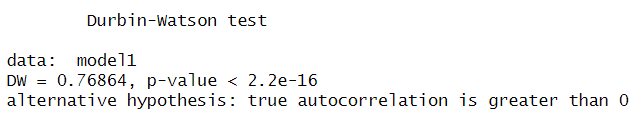
Pe lângă creșterea puterii explicative, s-a observat și **o îmbunătățire a acurateței predicțiilor** modelului. Radicul Erorii Medie Pătratice (RMSE) a scăzut la 44838.49, sugerând că predicțiile de preț generate de Modelul 2 sunt mai apropiate de prețurile reale în comparație cu Modelul 1.

Dispersia punctelor de date în jurul liniei unde valorile prezise sunt egale cu valorile reale este mai concentrată în cazul Modelului 2, ceea ce subliniază o corelare mai strânsă între valorile prezise și cele reale, și sugerează o acuratețe îmbunătățită. Aceste îmbunătățiri evidențiază importanța variabilei Calitatea generală a proprietății în predicția prețurilor imobiliarelor și confirmă utilitatea metodologiei stepwise în optimizarea modelelor de predicție prin selecția atentă a variabilelor relevante.



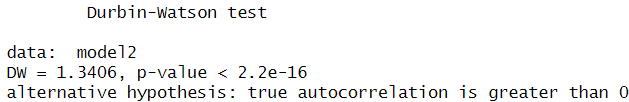
# **Figura 3.2** Model2 cu *Overall.Qual* adaugat

După prezentarea modelelor de regresie liniară și a rezultatelor statistice preliminare, am aplicat testul **Durbin-Watson** pentru a detecta prezența autocorelației în reziduurile modelelor. Autocorelația, sau corelarea erorilor dintr-un model de serie temporală, poate indica probleme potențiale cu modelul, cum ar fi omisiuni de variabile relevante sau modelarea inadecvată a dependențelor din date.



# **Figura 3.3** DW pentru *Model 1*

Rezultatele testului Durbin-Watson pentru [Modelul 1](#_Figura_3.3_DW) au arătat un DW de **0.76864**, semnificativ sub pragul de 2, sugerând astfel o autocorelație pozitivă puternică. Aceasta implică faptul că modelul ar putea suferi de omisiuni de variabile semnificative sau alte probleme de specificație. În contrast, [Modelul 2](#_Figura_3.4_DW) a prezentat un DW de **1.3406**, mai apropiat de 2, ceea ce indică o autocorelație pozitivă moderată, semnalând că acest model poate fi mai adecvat, dar totuși nu imun la probleme de autocorelație.



# **Figura 3.4** DW pentru *Model 2*

Valoarea extrem de mică a P (<< 0.05) pentru ambele modele subliniază semnificația statistică a autocorelației detectate. Aceasta sugerează că, în ciuda puterii explicative a modelelor, trebuie să acordăm o atenție deosebită la interpretarea și validitatea predicțiilor lor.

### III. Rezultate și discuții

În această lucrare am analizat piața imobiliară, am creat și evaluat modele predictive pentru a înțelege mai bine factorii care influențează prețurile locuințelor și am verificat acuratețea acestor modele în estimarea variabilelor.

În Figurile [3.1](#_Figura_3.1_Model1) si [3.2](#_Figura_3.2_Model1) sunt reprezentate niste grafice de dispersie al valorilor prezise în comparație cu valorile reale pentru un model de regresie liniară. Axa x este etichetată “Valori reale”, iar axa y este etichetată “Predicții”. Punctele de date sunt reprezentate prin cercuri albastre.

Performanța modelului de regresie liniară poate fi evaluată prin cât de aproape se află punctele de date de linia diagonală, care reprezintă cazul ideal în care valorile prezise sunt egale cu valorile reale. Dacă punctele de date sunt împrăștiate în mod aleatoriu și departe de linia diagonală, acest lucru indică o performanță slabă a modelului. Dacă punctele de date sunt concentrate în jurul liniei diagonale, acest lucru indică o performanță bună a modelului.

Din rezultate, se observa că includerea variabilei **Overall.Qual** în model a dus la o îmbunătățire semnificativă a ajustării modelului. R-squared-ul mai mare și RMSE-ul mai mic indică o mai bună potrivire a datelor și o capacitate mai mare a modelului de a face prognoze precise.

Această examinare aprofundată a relevat modul în care factori externi, precum fluctuațiile economice și schimbările demografice, exercită o influență palpabilă asupra dinamicii pieței imobiliare. Acesta a fost un demers esențial pentru a înțelege cum s-ar putea schimba piața în viitor și pentru a anticipa potențialele direcții de evoluție.

Pe de altă parte, am analizat și ipoteza că o calitate mai înaltă a locuințelor conduce la prețuri superioare pe piață. Prin aplicarea analizei corelaționale, am identificat o relație statistic semnificativă între calitate și preț, susținută de coeficienți de corelație robusti. Această descoperire confirmă presupunerea că proprietățile cu materiale de construcție superioare, finisaje de înaltă calitate și alte caracteristici atractive comandă prețuri mai mari pe piață.

Modelele noastre de regresie liniară, construite cu grijă și precizie, au oferit o imagine cantitativă a modului în care calitatea generală și alți factori influențează prețurile. Valoarea R-squared a indicat proporția de variație a prețurilor explicată de modelele noastre, în timp ce RMSE a oferit o măsură a acurateței predicțiilor. Astfel, am demonstrat nu doar că ipotezele noastre au fost înțelese, dar și că au fost validate prin aplicarea unei metodologii statistice sofisticate.

Valoarea DW sub 2 pentru Modelul 1 ridică semne de întrebare cu privire la validitatea predicțiilor acestuia. Acest lucru poate afecta încrederea în model și, prin urmare, în concluziile trase din analiză. De exemplu, o autocorelație pozitivă puternică poate duce la estimări ale coeficienților care sunt sistematic părtinite, ceea ce înseamnă că efectul anumitor predictorii asupra variabilei de răspuns este fie supraestimat, fie subestimat.

Pentru Modelul 2, deși DW este mai apropiat de 2, persistența unei autocorelații moderată necesită precauție. Chiar și o autocorelație moderată poate indica o dependență temporală care nu este captată de model, afectând astfel acuratețea predicțiilor pe termen lung.

Totuși, este important de recunoscut că orice analiză vine cu limitările sale. În cazul studiului nostru, una dintre limitări poate fi dată de natura datelor istorice folosite, care, deși extinse, nu pot acoperi toate variabilele potențiale care influențează piața imobiliară. De asemenea, modelele de regresie se bazează pe presupunerea că relațiile trecute vor continua în viitor, ceea ce nu este întotdeauna cazul într-o lume în schimbare rapidă. În plus, analiza noastră corelațională, deși puternică, nu poate stabili cauzalitate, ci doar asociații.

Studiul acesta oferă insight-uri valoroase și confirmă ipotezele propuse, trebuie de recunoscut că cercetarea continuă și analiza datelor suplimentare sunt esențiale pentru a aprofunda și a extinde înțelegerea noastră a pieței imobiliare.

### Anexa modelului

Codurile și datele sunt disponibile la:

<https://github.com/FlaviuPalii/ModelDeRegresie/blob/main/Fourth.R>

### BIBLIOGRAFIE

Set de date

1. **Ames Housing Dataset**

Autori: Dean De Cock

Sursă originală: [Ames Housing Dataset](http://jse.amstat.org/v19n3/decock.pdf)

Pachete R utilizate:

1. **tidyverse**

Autori: Wickham, H.

Sursă: [tidyverse - R Documentation](https://www.rdocumentation.org/packages/tidyverse)

1. **lattice**

Autori: Sarkar, Deepayan

Sursă: [lattice - R Documentation](https://www.rdocumentation.org/packages/lattice)

1. **rsample**

Autori: Kuhn, M., & Vaughan, D.

Sursă: [rsample - R Documentation](https://www.rdocumentation.org/packages/rsample)

1. **vip**

Autori: Greenwell, B.

Sursă: [vip - R Documentation](https://www.rdocumentation.org/packages/vip)

1. **caret**

Autori: Kuhn, M.

Sursă: [caret - R Documentation](https://www.rdocumentation.org/packages/caret)

1. **lmtest**

Autori: Zeileis, A., Hothorn, T., & Leisch, F.

Sursă: [lmtest - R Documentation](https://www.rdocumentation.org/packages/lmtest)

1. **stats**

Sursă: [stats - R Documentation](https://www.rdocumentation.org/packages/stats)