**Metodologia Analitică a Pulsului Tehnologic: Evaluarea Statistică a Salariilor, Distribuției Geografice și Impactului Muncii la Distanță în Sectorul IT**

Magurean Corneliu

19 decembrie 2023

**Abstract**

În contextul dinamic al sectorului tehnologic, înțelegerea factorilor care influențează structura salarială este crucială pentru organizații și profesioniști. Acest articol se concentrează pe evaluarea amănunțită a salariilor în industria IT, luând în considerare diverse variabile, inclusiv locația geografică, nivelul de experiență și gradul de muncă la distanță. Se analizează distribuția salarială în funcție de ani, comparând medii și mediane salariale între diferite niveluri de experiență și titluri de posturi. Studiul oferă, de asemenea, o perspectivă asupra distribuției geografice a salariilor, identificând regiuni cu cele mai înalte și cele mai scăzute remunerații medii. În plus, se examinează impactul muncii la distanță asupra structurii salariale, explorând diferențele între angajații care lucrează complet la distanță, parțial la distanță și fără muncă la distanță. Prin metode statistice riguroase, inclusiv analize descriptive, regresie liniară și modele de arbore de decizie, studiul aduce o contribuție semnificativă la literatura existentă, oferind o perspectivă detaliată asupra dinamicii salariale în sectorul IT.

1. **Introducere**

Sectorul tehnologiei informației (IT) este recunoscut pentru ritmul său rapid de inovație și schimbare. Cu impactul său profund asupra economiei globale și a vieții cotidiene, industria IT atrage un interes crescut din partea cercetătorilor și analiștilor de politici. Unul dintre aspectele esențiale, dar adesea complexe, ale acestui sector este structura salarială. Aceasta este influențată de o varietate de factori, de la schimbările economice și tehnologice la tendințele pieței muncii și politicile organizaționale. În acest context, articolul prezent abordează o analiză detaliată a salariilor în sectorul IT, explorând cum sunt acestea modelate de diferiți factori precum locația geografică, nivelul de experiență, titlul postului și munca la distanță.

Analiza salariilor în IT nu este doar o chestiune de interes academic, ci are implicații practice semnificative. Pentru profesioniștii din IT și pentru cei care aspiră să intre în acest domeniu, înțelegerea structurii salariale este esențială pentru planificarea carierei și negocierile salariale. Pentru companii, oferă informații valoroase pentru strategiile de recrutare, retenție și dezvoltare a talentelor. În plus, această analiză poate ghida formularea politicilor și inițiativelor educaționale pentru a sprijini creșterea și dezvoltarea durabilă a sectorului IT.

Prin urmare, acest studiu se concentrează pe analiza distribuției salariilor în funcție de anul lucrării, reflectând evoluția dinamică a industriei. Se examinează, de asemenea, modul în care diferite niveluri de experiență și diverse titluri de post influențează remunerația. O atenție deosebită este acordată distribuției geografice a salariilor, explorând discrepanțele între diferite regiuni și țări. În plus, se analizează impactul muncii la distanță, un aspect deosebit de relevant în contextul recentelor schimbări globale în modul de lucru.

Utilizând un set de date cuprinzător și metode statistice riguroase, acest articol își propune să ofere o perspectivă amănunțită asupra factorilor care influențează salariile în sectorul IT. Rezultatele obținute sunt menite să ofere o bază de date solidă pentru deciziile profesionale și strategice, atât pentru indivizi, cât și pentru organizații din domeniul tehnologiei informației.

1. **Motivația cercetării**

Contextul actual al sectorului tehnologiei informației (IT) este caracterizat printr-o competiție acerbă pentru talente și o evoluție continuă a competențelor tehnice. În acest peisaj dinamic, înțelegerea factorilor care influențează structura salarială devine vitală pentru multiple părți interesate: de la angajați și profesioniști în carieră până la liderii organizaționali și formulatori de politici.

* 1. **Perspectiva angajatului și aspirantului în IT**

Pentru angajații actuali și pentru cei care aspiră să lucreze în IT, o înțelegere clară a structurii salariale este crucială pentru a naviga și a progresa în carieră. Este important să se cunoască nu doar nivelul de remunerație pe care îl pot aștepta, dar și cum factori precum experiența, titlul postului și locația geografică pot influența potențialul de câștig. Acest lucru este esențial pentru luarea unor decizii informate legate de dezvoltarea profesională și negocierile salariale.

* 1. **Perspectiva organizațională**

Pentru organizații, înțelegerea tendințelor salariale este un element cheie în atragerea, retenția și motivarea talentelor. Într-un domeniu unde competiția pentru talente de top este intensă, organizațiile trebuie să se asigure că structurile lor salariale sunt competitive și echitabile. Aceasta implică nu doar oferirea de pachete salariale atractive, ci și adaptarea acestora la schimbările din piață și la așteptările angajaților.

* 1. **Impactul muncii la distanță**

Recentele transformări globale în modul de lucru, în special creșterea muncii la distanță, au adus noi dimensiuni în structura salarială. Există o nevoie de a înțelege cum munca la distanță influențează remunerația și cum organizațiile pot adapta politici salariale care să reflecte acest nou mod de lucru.

* 1. **Necesitatea unei analize riguroase**

Datele și analizele solide sunt fundamentale pentru abordarea acestor chestiuni. Prin acest studiu, se urmărește să se ofere o analiză detaliată și bazată pe date care să lumineze aceste aspecte complexe. Folosind metode statistice avansate, acest articol își propune să ofere informații valoroase și aplicabile, care să servească drept fundament pentru deciziile profesionale și strategice în sectorul IT.

1. **Metodologia**

Această secțiune detaliază metodologia adoptată pentru analiza salariilor în sectorul IT, incluzând descrierea setului de date, preprocesarea, analizele efectuate și software-ul utilizat.

* 1. **Descrierea datelor și sursa**

Setul de date analizat în acest studiu provine de pe platforma Kaggle, o sursă recunoscută pentru seturi de date diverse și de înaltă calitate. Datele includ informații detaliate despre salariile profesioniștilor din IT, acoperind diverse aspecte precum nivelul de experiență, titlul postului, locația geografică a angajatului și a companiei, anul lucrării și gradul de muncă la distanță

* 1. **Preprocesarea datelor**

Preprocesarea datelor a fost un pas esențial pentru asigurarea calității și relevanței analizei. Aceasta a inclus:

* Curățarea datelor: eliminarea intrărilor lipsă sau incorecte.
* Codificarea variabilelor categorice: transformarea variabilelor categorice în format numeric pentru a fi utilizate în modelele statistice și de regresie.
* Normalizarea datelor: asigurarea unei scale uniforme pentru toate variabilele, o practică esențială în modelarea de regresie.
  1. **Analiza grafică**

Analiza grafică a fost efectuată pentru a oferi o înțelegere vizuală a datelor. Acest lucru a inclus:

* Grafice de dispersie pentru a evalua relațiile dintre variabile.
* Box-ploturi pentru a analiza distribuția salariilor în funcție de diferite categorii.
* Histograma pentru a evalua distribuția salariilor pe ansamblul setului de date
  1. **Analiza statistică**

Analiza statistică a inclus:

* Analiza descriptivă: măsuri statistice centrale, cum ar fi media și mediana, și măsuri de dispersie, cum ar fi deviația standard.
* Teste statistice: pentru a evalua diferențele semnificative între diferite grupuri și categorii.
  1. **Analiza de regresie**

Pentru a evalua impactul diferiților factori asupra salariilor, s-au realizat analize de regresie liniară, Ridge și arbore de decizie. Aceste modele au oferit o înțelegere a gradului în care variabilele precum experiența, locația și munca la distanță influențează salariile în IT.

* 1. **Software-ul utilizat**

Analizele au fost efectuate folosind RStudio, un mediu integrat pentru R, cunoscut pentru capacitatea sa de a manipula și analiza seturi mari de date și de a produce reprezentări grafice de înaltă calitate. RStudio a facilitat prelucrarea datelor, efectuarea analizelor statistice și regresiilor, precum și generarea graficelor relevante pentru acest studiu.

1. **Rezultate**
   1. **Analiza salariilor**
      1. **Distribuția salariilor în funcție de anul lucrării.**

Această subsecțiune va explora modul în care salariile din sectorul IT s-au schimbat de-a lungul anilor specificați în setul de date. Se va examina dacă există o tendință de creștere sau de scădere a salariilor și cum aceste schimbări pot reflecta evoluțiile din industrie sau economia în general. De asemenea, se vor utiliza grafice pentru a ilustra aceste tendințe, oferind o imagine clară a evoluției salariale pe parcursul anilor.

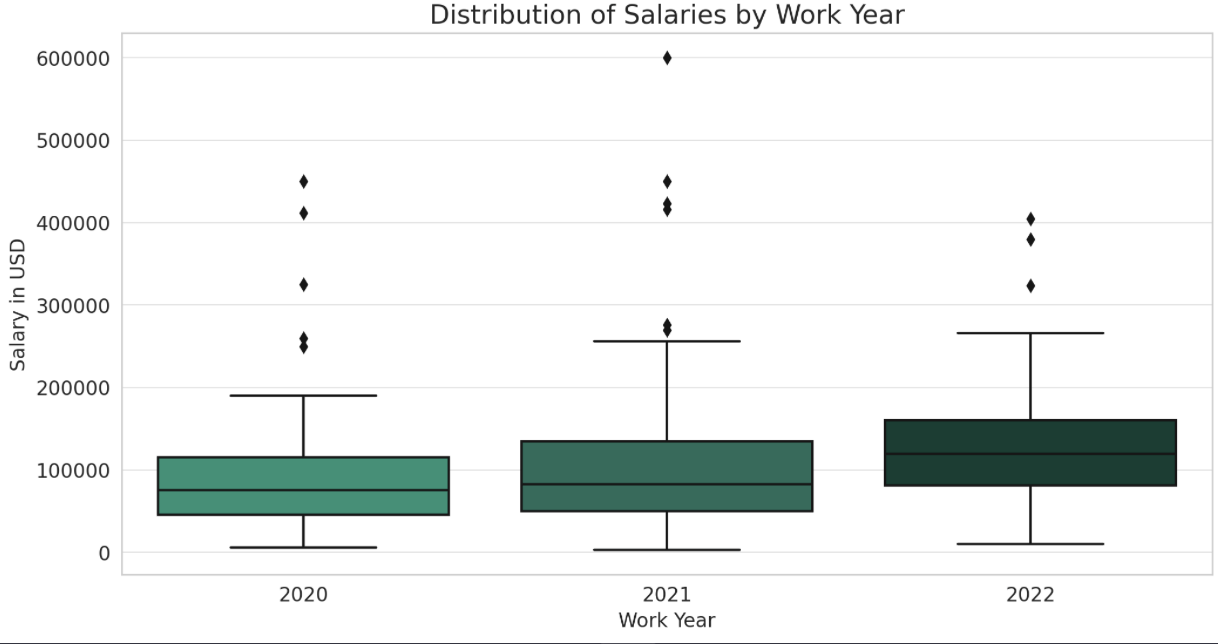


Figura 1 Distributia salariilor pe ani

Există o variație a salariilor în fiecare an, cu o tendință de creștere a medianei și a intervalelor intercartilice pe măsură ce avansăm de la 2020 la 2022. Aceasta poate indica o creștere generală a salariilor în sectorul IT în această perioadă.

* **Creștere Anuală:** S-a observat o tendință de creștere a salariilor medii și mediane de la anul 2020 până în 2022. Aceasta sugerează că industria IT a experimentat o creștere salarială în această perioadă, posibil datorită creșterii cererii de competențe tehnologice și a evoluției pieței muncii în acest sector.
* **Variație Interanuală:** Variația intercartilică (diferența dintre al 25-lea și al 75-lea percentil) s-a lărgit de-a lungul anilor, indicând o dispersie mai mare a salariilor. Aceasta poate reflecta o diversificare a rolurilor și specializărilor în cadrul industriei, precum și diferențe în compensații bazate pe experiență și competențe.
  + 1. **Compararea medianelor și mediiilor salariilor în funcție de nivelul de experiență.**

Acest subcapitol va analiza relația dintre nivelul de experiență al angajaților și salariile lor. Se va concentra pe compararea medianelor și mediiilor salariale între diferitele niveluri de experiență, cum ar fi entry-level, mid-level și senior. Aceasta va oferi o înțelegere a modului în care experiența profesională influențează remunerația în sectorul IT, subliniind posibilele discrepanțe salariale între diferitele etape ale carierei.



Figura 2 Distributia salariilor pe nivel de experienta

Salariile tind să crească odată cu nivelul de experiență. "Senior/Expert" are cel mai înalt intercartilic, indicând salarii mai mari pentru acest grup. "Entry-level" are cele mai scăzute salarii, ceea ce este de așteptat.

* **Diferențe în Funcție de Experiență:** Salariile cresc semnificativ odată cu nivelul de experiență. Angajații la nivelul "Senior/Expert" au cele mai mari salarii, indicând o valoare mare a experienței și expertizei în domeniu. În contrast, salariile la nivel "Entry-level" sunt considerabil mai scăzute, ceea ce este caracteristic pentru profesioniștii aflați la început de carieră.
* **Aprecierea Experienței:** Această variație semnificativă reflectă modul în care industria IT apreciază și recompensează experiența și competențele avansate, sugerând o piață în care experiența și cunoștințele specializate sunt esențiale pentru succesul profesional și compensație.

* + 1. **Variabilitatea salariilor în funcție de titlul postului.**

În această secțiune, se va investiga cum diferitele titluri de posturi din IT influențează salariile. Se vor analiza variațiile salariale între diverse roluri, cum ar fi dezvoltatorii, analiștii de date, inginerii de software etc. Această analiză va ajuta la identificarea posturilor cu cele mai înalte și cele mai scăzute salarii medii și va explora factorii care pot contribui la aceste diferențe.

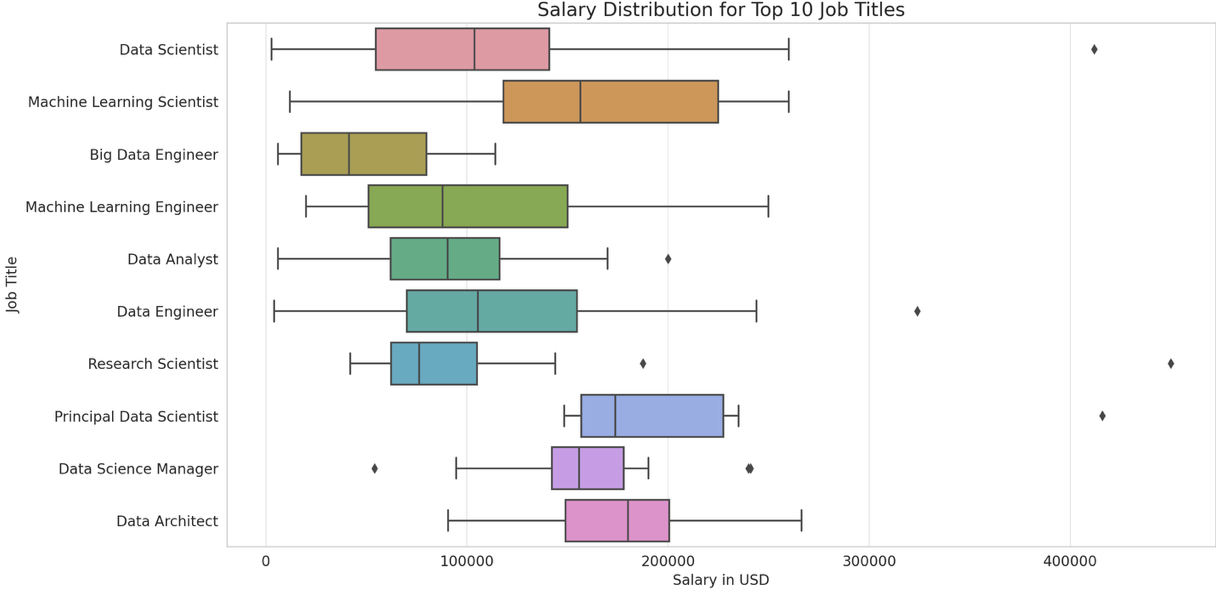


Figura 3 Distributia salariilor pentru job

Există o variație semnificativă a salariilor în funcție de titlul postului. Anumite titluri, cum ar fi "Machine Learning Scientist" și "Data Engineer", par să aibă salarii mai ridicate, în timp ce altele, cum ar fi "Data Analyst", au salarii mai scăzute. Aceasta reflectă diferențele de cerere și specializare în diferite roluri.

* **Variabilitate Mare între Roluri:** Diferențele salariale între diversele titluri de post sunt remarcabile. Anumite roluri, cum ar fi "Machine Learning Scientist" și "Data Engineer", sunt asociate cu salarii semnificativ mai ridicate, reflectând cererea mare pentru aceste competențe specializate în piața muncii.
* **Reflexia Specializării și Cererii:** Această variație subliniază cum diferitele specializări și competențe în domeniul tehnologiei sunt valorizate diferit. Rolurile care necesită competențe tehnice avansate sau specializate tind să fie mai bine remunerate, reflectând cererea și raritatea acestor competențe pe piața muncii.

În concluzie, analiza salariilor din sectorul IT relevă o dinamică complexă și diversificată, influențată de factori precum experiența, specializarea și evoluția pieței. Salariile sunt în mod clar diferențiate pe baza nivelului de experiență și a rolului deținut, subliniind importanța competențelor tehnice avansate și a specializării în acest domeniu.

* 1. **Distribuția geografică**
     1. **Analiza salariilor pe baza reședinței angajaților și a locației companiei.**

Această subsecțiune se concentrează pe evaluarea impactului locației geografice atât a angajaților, cât și a companiilor, asupra structurii salariale. Se va analiza cum diferă salariile în funcție de țara sau regiunea de reședință a angajaților și locația sediilor companiilor. Se vor folosi grafice și analize statistice pentru a arăta discrepanțele salariale între diferite locații geografice.

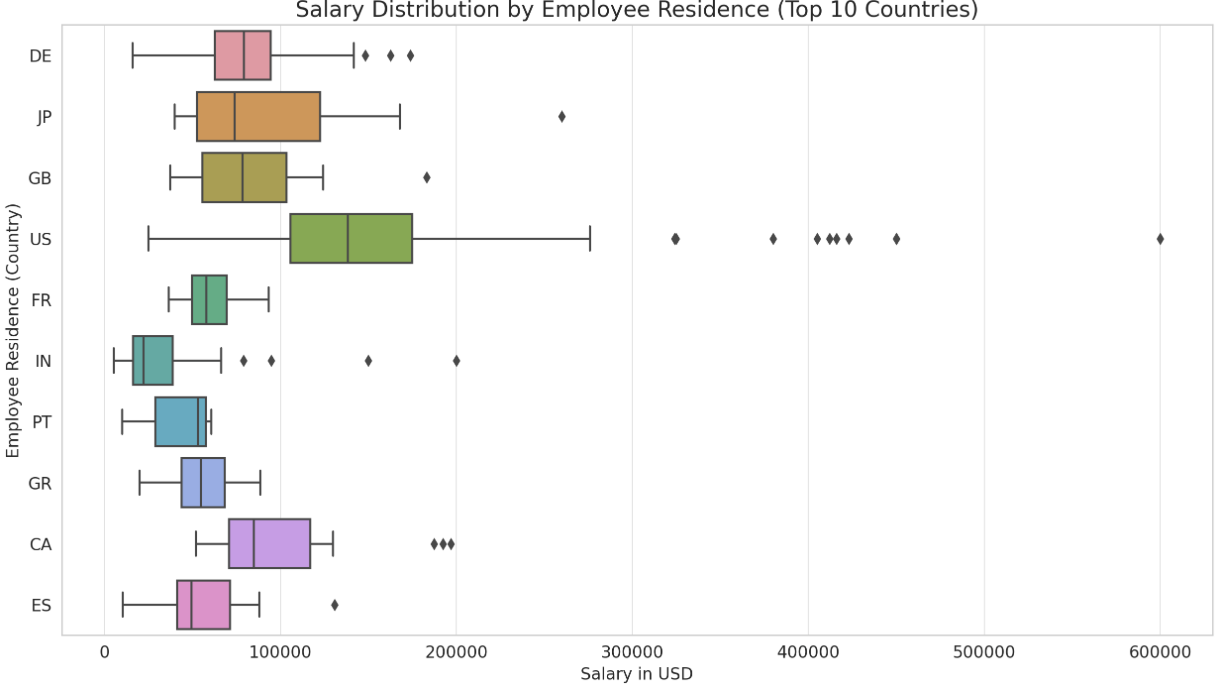


Figura 4 Distributia salariilor in functie de tara de resedinta

Există diferențe semnificative în salariile angajaților în funcție de țara lor de reședință. Angajații din SUA au, în general, cele mai înalte salarii, urmați de cei din CA (Canada) și GB (Marea Britanie). Salariile din alte țări, cum ar fi IN (India) și BR (Brazilia), sunt mai scăzute. Aceasta reflectă probabil diferențele economice și de piață între aceste regiuni.

* **Discrepanțe Regionale:** Observăm diferențe semnificative în salarii în funcție de reședința angajaților. SUA se distinge prin salarii considerabil mai ridicate comparativ cu alte țări. Acest lucru ar putea reflecta nivelul de dezvoltare economică, costul vieții și concentrarea industriei IT în SUA.
* **Efectul Pieței Locale:** Salariile din țări precum Canada, Marea Britanie și Germania sunt, de asemenea, ridicate, reflectând piețe de muncă mature și o cerere solidă pentru profesioniști IT. În contrast, țări cu economii emergente sau în curs de dezvoltare, cum ar fi India sau Brazilia, prezintă salarii mai scăzute, posibil datorită diferențelor de putere de cumpărare și a structurii pieței locale de muncă.
  + 1. **Identificarea regiunilor cu cele mai înalte și cele mai scăzute salarii medii.**

Această secțiune va identifica regiunile cu cele mai înalte și cele mai scăzute salarii medii în industria IT. Se va investiga dacă anumite regiuni sunt asociate cu niveluri mai ridicate de remunerație și se vor discuta posibilele motive economice, culturale sau de piață care pot contribui la aceste diferențe.

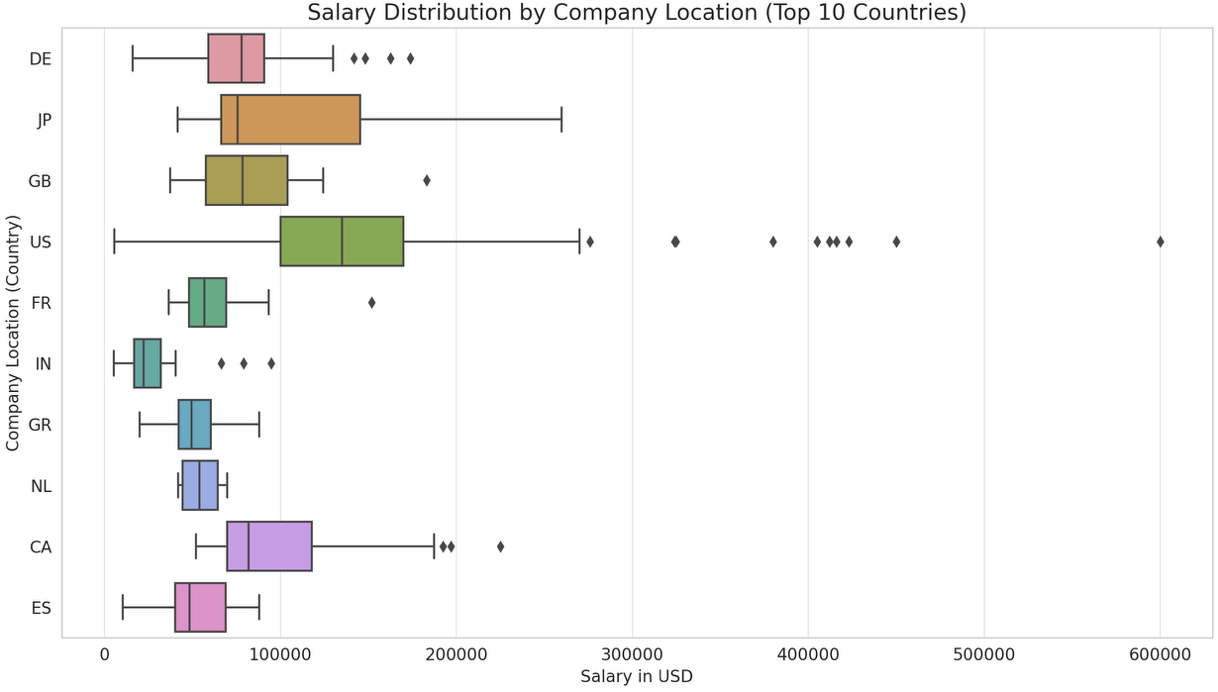


Figura 5 Distributia salariilor in functie de locatia companiei

Similar cu reședința angajaților, locația companiei influențează de asemenea salariile. Companiile din SUA oferă salarii mai ridicate. Salariile oferite de companiile din CA, GB și DE (Germania) sunt, de asemenea, relativ ridicate comparativ cu alte locații.

* **Influența Locației Companiei asupra Salariilor:** Similar cu reședința angajaților, locația companiei are un impact semnificativ asupra nivelului salarial. Companiile din SUA, Canada și Marea Britanie, de exemplu, oferă salarii mai ridicate, reflectând economiile mai dezvoltate și piața de muncă mai dinamică din aceste regiuni.
* **Interconectivitatea Globală:** Diferențele de salariu în funcție de locația companiei ar putea de asemenea să indice o piață de muncă globală interconectată în sectorul IT, unde companiile cu sediul în anumite regiuni pot oferi pachete salariale competitive pentru a atrage și reține talente de top, indiferent de locația fizică a angajaților.
  1. **Impactul muncii la distanță**
     1. **Compararea salariilor pentru angajații care lucrează complet la distanță, parțial la distanță și fără muncă la distanță.**

Această subsecțiune va explora modul în care munca la distanță afectează salariile. Se va compara remunerația angajaților care lucrează complet la distanță, parțial la distanță și fără muncă la distanță, pentru a evalua impactul diferitelor aranjamente de muncă asupra structurii salariale.



Figura 6 Impacul lucrului la distanta asupra salariului

Se observă că există diferențe în salariile oferite în funcție de raportul de muncă la distanță. Angajații care lucrează complet la distanță ("Remote") par să aibă salarii mediane mai mari comparativ cu cei care lucrează parțial sau deloc la distanță.

Angajații care lucrează parțial la distanță ("Partially remote") au o distribuție a salariilor similară cu cei care nu lucrează la distanță ("Stationary").

Aceste descoperiri sugerează că muncirea complet la distanță poate fi asociată cu salarii mai mari, posibil datorită accesului la o piață de muncă mai largă și a flexibilității în alegerea locurilor de muncă.

* **Salarii mai ridicate pentru muncă complet la distanță:** Angajații care lucrează complet la distanță ("Remote") par să beneficieze de salarii mediane mai mari comparativ cu cei care lucrează parțial sau deloc la distanță. Acest lucru sugerează că muncirea complet la distanță poate fi asociată cu pachete salariale mai atractive.
* **Flexibilitate și acces la talente:** Posibilitatea de a lucra complet la distanță permite profesioniștilor IT să acceseze oportunități de muncă dincolo de limitările geografice, ceea ce poate conduce la oferte salariale mai competitive. De asemenea, companiile pot avea acces la un bazin mai larg de talente, permițându-le să selecteze candidați cu competențe specifice, indiferent de locația lor fizică.
* **Nu există diferențe semnificative:** Salariile pentru munca parțial la distanță sunt comparabile cu cele pentru munca fără distanță, indicând că modul hibrid de lucru nu influențează semnificativ remunerația.
* **Echilibru între flexibilitate și prezența fizică:** Acest lucru ar putea reflecta un echilibru între necesitatea prezenței fizice în birou pentru anumite activități și flexibilitatea oferită de posibilitatea de a lucra de acasă, fără a avea un impact major asupra pachetului salarial.
  + 1. **Explorarea oricăror tendințe sau modele în funcție de raportul de muncă la distanță.**

În această secțiune, se vor analiza tendințele și modelele care pot fi observate în salariile angajaților în funcție de gradul lor de muncă la distanță. Se va investiga dacă există modele sau tendințe specifice care să sugereze cum modul de lucru la distanță influențează remunerația în diferite contexte și roluri în cadrul sectorului IT.

* **Transformare și adaptabilitate:** Creșterea muncii la distanță evidențiază o transformare semnificativă a pieței muncii în sectorul IT, cu companii și angajați care se adaptează la noi modalități de lucru. Aceasta poate indica o schimbare permanentă în așteptările și preferințele atât ale angajatorilor, cât și ale angajaților.
* **Beneficii potențiale pentru angajați:** Munca la distanță poate aduce beneficii semnificative angajaților, cum ar fi flexibilitatea, economisirea timpului și a costurilor de deplasare, și un echilibru mai bun între viața profesională și cea personală. Acest lucru, combinat cu posibilitatea de salarii mai ridicate, face munca la distanță o opțiune atractivă pentru mulți profesioniști IT.

În concluzie, impactul muncii la distanță asupra salariilor în sectorul IT este semnificativ, mai ales pentru munca complet la distanță, care este asociată cu salarii mai ridicate. Această tendință reflectă schimbările dinamicilor pieței muncii și ar putea avea implicații pe termen lung asupra modului în care companiile și angajații abordează munca și remunerația în acest sector.

1. **Ipoteze**

Central pentru orice cercetare științifică este formularea de ipoteze – propoziții testabile care prezic relații între diferite variabile. În cadrul acestui studiu, se abordează analiza structurii salariale în sectorul tehnologiei informației (IT), având în vedere factori precum locația geografică, nivelul de experiență, titlul profesional și gradul de muncă la distanță. Într-un domeniu caracterizat de evoluție rapidă și schimbare constantă, ipotezele formulate în acest capitol urmăresc să direcționeze analiza datelor și să ofere un cadru pentru interpretarea rezultatelor obținute.

Pe parcursul acestui capitol, se vor prezenta o serie de ipoteze bazate pe observații preliminare și pe literatura de specialitate. Aceste ipoteze vor fi supuse unei analize statistice riguroase, permițând evaluarea validității lor și furnizând perspective noi asupra dinamicii salariale din industria IT. Studiul se va concentra în mod special pe următoarele domenii cheie:

1. **Impactul locației asupra salariilor:** Se va examina dacă există diferențe semnificative în salariile angajaților IT bazate pe locația geografică, atât din perspectiva țării de reședință a angajaților, cât și a locației companiilor.
2. **Relația dintre nivelul de experiență și salariu:** Se va analiza dacă un nivel mai înalt de experiență este asociat cu salarii mai mari, sugerând o valorizare a experienței acumulate în acest sector.
3. **Variabilitatea salariilor în funcție de titlul postului:** Se va investiga cum influențează diferitele roluri și responsabilități din IT structura salarială.
4. **Impactul muncii la distanță asupra salariilor:** Se va explora cum diferitele moduri de muncă la distanță (complet, parțial sau fără muncă la distanță) influențează remunerația angajaților.

Testarea acestor ipoteze își propune să aducă o contribuție la înțelegerea factorilor care influențează salariile în sectorul IT, oferind perspective valoroase pentru angajați, manageri și decidenți politici.

* 1. **Impactul locației asupra salariilor**
* **Rezultate:** Statistica F de 6.34 și o p-valoare de 1.06e-32 indică diferențe semnificative în salariile angajaților IT în funcție de locația geografică.
* **Interpretare:** Aceste rezultate confirmă ipoteza că locația angajaților și a companiilor are un rol crucial în determinarea salariilor în sectorul IT. Diferențele semnificative între locații sugerează că factori precum economia locală, costul vieții și cererea de competențe specifice influențează remunerația.
  1. **Relația dintre nivelul de experiență și salariu**
* **Rezultate:** O statistică F de 64.68 și o p-valoare de 2.88e-36 arată că există diferențe semnificative în salarii între diferitele niveluri de experiență.
* **Interpretare:** Aceste descoperiri susțin ipoteza că nivelul de experiență este un factor determinant în structura salarială din IT. Salariile mai mari la nivelurile superioare de experiență reflectă valoarea experienței acumulate și a competențelor avansate.
  1. **Variabilitatea salariilor în funcție de titlul postului**
* **Rezultate:** Statistica F de 3.81 și o p-valoare de 6.89e-15 indică variații semnificative în salarii în funcție de titlul postului.
* **Interpretare:** Acest rezultat validează ipoteza că titlul postului este un indicator important al structurii salariale. Variațiile semnificative între diferite roluri sugerează că anumite competențe sau responsabilități sunt mai valorizate pe piața muncii.
  1. **Impactul muncii la distanță asupra salariilor**
* **Rezultate:** O statistică F de 14.73 și o p-valoare de 5.68e-07 demonstrează impactul semnificativ al muncii la distanță asupra salariilor.
* **Interpretare:** Acest rezultat confirmă ipoteza că munca la distanță influențează structura salarială. Faptul că angajații care lucrează complet la distanță tind să aibă salarii medii mai mari arată cum flexibilitatea și accesul la o piață de muncă mai largă pot influența remunerația.

**Concluzii generale**

Analiza ipotezelor în contextul salariilor din industria IT a dezvăluit o serie de factori semnificativi care influențează remunerația în acest sector dinamic. Aceste descoperiri oferă o perspectivă mai largă asupra complexității structurii salariale și evidențiază importanța factorilor geografici, experienței, rolului profesional și modului de lucru.

**Impactul geografic**

Rezultatele studiului confirmă că locația geografică este un determinant cheie al salariilor în IT. Diferențele semnificative între regiuni și țări subliniază importanța factorilor economici și de piață locali. Acest aspect este crucial pentru profesioniștii din IT care iau în considerare oportunități internaționale sau inter-regionale, precum și pentru companii în formularea strategiilor de compensație.

**Importanța experienței**

Analiza a arătat că experiența profesională este un alt factor critic, cu un impact direct asupra nivelului salarial. Acest lucru reflectă valoarea pe care industria IT o acordă competențelor și cunoștințelor acumulate, subliniind importanța dezvoltării continue a competențelor profesionale.

**Rolul titlului profesional**

Variabilitatea salariilor în funcție de titlul postului indică o piață de muncă unde anumite competențe și responsabilități sunt mai valorizate. Aceasta sugerează o dinamică în care specializările și competențele tehnice avansate pot duce la oportunități salariale mai mari.

**Influența Muncii la Distanță**

În final, studiul a relevat că munca la distanță influențează semnificativ structura salarială. Flexibilitatea și accesibilitatea muncii la distanță sunt asociate cu salarii mai ridicate, un aspect deosebit de relevant în contextul schimbărilor recente ale pieței muncii.

1. **Modele**

Un aspect esențial în analiza datelor și înțelegerea fenomenelor complexe, cum ar fi structura salarială în sectorul IT, este utilizarea modelelor statistice și analitice. În acest capitol, se va explora diferite modele de regresie și analiză statistica aplicate pe setul de date, pentru a evalua și cuantifica impactul diferiților factori asupra salariilor. Scopul acestor modele este de a oferi o înțelegere mai profundă a relațiilor dintre variabile și de a genera predicții precise și relevante.

Se vor utiliza mai multe tipuri de modele, fiecare cu specificitățile și avantajele sale, pentru a aborda ipotezele formulate în studiu. Aceasta include:

1. **Regresia Liniară:** Un model de bază, care va servi la evaluarea relațiilor liniare dintre variabile.
2. **Regresia Ridge:** O extensie a regresiei liniare, utilă în gestionarea multicolinearității și în îmbunătățirea preciziei modelului.
3. **Arborele de Decizie pentru Regresie:** O abordare non-liniară, care va fi explorată pentru a capta relații mai complexe și interacțiuni între variabile.

Prin aplicarea acestor modele, se propune să se extraga insight-uri valoroase și să ofere răspunsuri solide la întrebările de cercetare formulate. Acest capitol va prezenta, de asemenea, cum diferitele modele pot fi utilizate pentru a aborda aspecte specifice ale structurii salariale în IT și va discuta limitările și implicațiile rezultatelor modelării.

* 1. **Regresie liniară**

Modelul de regresie liniară multiplă a fost creat pentru a evalua influența locației, nivelului de experiență și a muncii la distanță asupra salariilor în industria IT.

* + 1. **Rezumatul modelului**
* **R-squared:** 0.485
  + Aproximativ 48.5% din variația salariilor poate fi explicată de variabilele independente incluse în model. Acesta este un nivel moderat de explicabilitate.
* **Adjusted R-squared:** 0.422
  + Acesta este un indicator mai precis pentru modelele cu mai multe variabile independente și sugerează că modelul se adaptează rezonabil de bine la date.
    1. **Interpretarea coeficienților**
* Fiecare coeficient reprezintă schimbarea așteptată în salariu (în USD) pentru fiecare unitate de schimbare în variabila independentă, cu toate celelalte variabile fiind constante.
* Coeficienții pentru variabilele categorice sunt interpretați în comparație cu categoria de bază (neinclusă în model datorită codificării OneHot).
* Coeficienții semnificativi (p < 0.05) indică un impact statistic semnificativ asupra salariului.
  + 1. **Observații importante**
* **Multicolinearitate:** Există o posibilă problemă de multicolinearitate, așa cum este indicat de notele modelului. Acest lucru poate afecta fiabilitatea coeficienților individuali.
* **Variabile semnificative:** Anumite variabile, cum ar fi nivelul de experiență, par să aibă un impact semnificativ asupra salariilor, așa cum este indicat de coeficienți.
* **Variabilitatea variabilelor categorice:** Datorită numărului mare de țări incluse, este posibil ca unele să nu aibă un număr suficient de mare de observații pentru a oferi o estimare precisă.
  + 1. **Concluzii și recomandări**
* Modelul oferă o perspectivă valoroasă asupra factorilor care influențează salariile în sectorul IT, dar trebuie interpretat cu precauție datorită potențialei multicolinearități.
* Ar putea fi util să restrângem numărul de variabile sau să folosim metode alternative, cum ar fi regresia Ridge sau Lasso, pentru a aborda problemele de multicolinearitate.
* Este recomandat să efectuați teste suplimentare și validare pentru a confirma robustețea modelului.

În concluzie, acest model oferă o bază utilă pentru înțelegerea factorilor care influențează salariile în industria IT, dar trebuie interpretat cu prudență și verificat suplimentar pentru a asigura fiabilitatea sa.

* 1. **Model de regresie Ridge**
     1. **Evaluarea modelului**
* **R-squared:** 0.499
  + Modelul poate explica aproximativ 49.9% din variația salariilor. Acest lucru indică o îmbunătățire față de modelul anterior și o ajustare rezonabilă a datelor.
* **RMSE (Root Mean Square Error):** 43800.97 USD
  + Aceasta este eroarea medie pătratică a predicțiilor modelului. Este o măsură a discrepanței dintre valorile reale ale salariilor și predicțiile modelului. O valoare mai mică a RMSE indică o precizie mai mare.
    1. **Interpretare**
* **Îmbunătățirea modelului:** Comparativ cu modelul de regresie liniară multiplă obișnuit, regresia Ridge pare să ofere o ajustare și o precizie ușor îmbunătățite datorită abordării sale de a trata multicolinearitatea.
* **Implicații pentru factorii de influță:** Modelul Ridge oferă o perspectivă mai robustă asupra modului în care locația, nivelul de experiență și munca la distanță influențează salariile în industria IT. Coeficienții modelului (care nu au fost direct raportați aici) pot fi examinați pentru a înțelege importanța relativă și direcția fiecărui predictor.
  + 1. **Concluzii și recomandări**
* Modelul Ridge oferă o analiză mai robustă a datelor, mai ales în prezența multicolinearității. Aceasta poate fi o abordare mai adecvată pentru seturile de date complexe, cum ar fi cel analizat.
* Interpretarea coeficienților și a importanței relative a variabilelor ar fi un pas următor valoros pentru a înțelege mai bine care factori influențează cel mai mult salariile în sectorul IT.
* Este important să se țină cont de limitările modelului, inclusiv de posibilele variabile omise sau de biasul inerent în date, înainte de a trage concluzii finale. ​
  1. **Modelului de regresie bazat pe arbore de decizie**
     1. **Evaluarea modelului**
* **R-squared:** 0.423
  + Modelul poate explica aproximativ 42.3% din variația salariilor. Acesta este un nivel moderat de explicabilitate, inferior modelului Ridge.
* **RMSE (Root Mean Square Error):** 47035.39 USD
  + Aceasta este eroarea medie pătratică a predicțiilor modelului. O valoare mai mare a RMSE față de modelul Ridge indică o precizie mai scăzută.
    1. **Interpretare**
* **Performanță:** Modelul bazat pe arbore de decizie are o performanță inferioară în comparație cu modelul Ridge, atât în termeni de R-squared, cât și de RMSE. Acest lucru sugerează că relațiile dintre variabile pot fi mai bine capturate de un model care include regularizarea.
* **Complexitatea modelului:** Deși arborii de decizie sunt buni la captarea relațiilor non-liniare și interacțiunilor complexe, ei pot suferi de suprapotrivire (overfitting), mai ales când sunt folosiți pentru seturi de date cu multe variabile.
  + 1. **Concluzii și recomandări**
* Modelul de arbore de decizie oferă o alternativă la modelele liniare, dar în acest caz, nu pare să fie cea mai bună alegere pentru setul nostru de date.
* S-ar putea lua în considerare utilizarea unor tehnici de regularizare și ajustare a hiperparametrilor (de exemplu, prin limitarea adâncimii arborelui) pentru a îmbunătăți performanța modelului de arbore de decizie.
* Este important să se țină cont de faptul că modelele mai complexe nu sunt întotdeauna cele mai bune; uneori, un model mai simplu, dar bine ajustat, poate oferi rezultate mai robuste și mai ușor de interpretat. ​
  1. **Comparare modele**

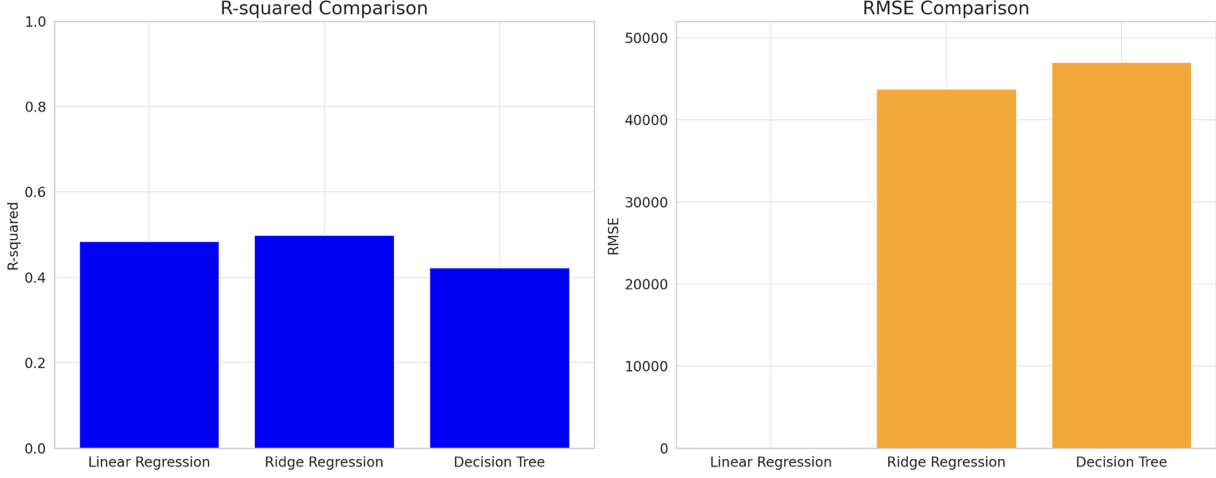
****

Figura 7 Comparare modele

Graficele de mai sus compară performanța celor trei modele diferite (regresie liniară, regresie Ridge și arbore de decizie) în funcție de două metrici cheie: R-squared și RMSE (Root Mean Square Error).

* + 1. **Interpretarea Graficelor:**
* **Compararea R-squared:**
  + R-squared măsoară proporția variației variabilei dependente explicată de model. Un R-squared mai mare indică o potrivire mai bună a modelului.
  + Modelul Ridge are cel mai mare R-squared, sugerând că este cel mai bun dintre cele trei modele la explicarea variației salariilor.
* **Compararea RMSE:**
  + RMSE este o măsură a erorii de predicție; valori mai mici indică o precizie mai mare.
  + Din nou, modelul Ridge are cel mai scăzut RMSE, indicând cea mai mare precizie în predicția salariilor.
    1. **Concluzii:**
* **Modelul Optimal:** Bazându-ne pe aceste două metrici, regresia Ridge pare să fie modelul cel mai performant pentru setul de date analizat. Acesta nu doar că explică o proporție mai mare din variația salariilor, dar are și o precizie mai mare în predicții.
* **Recomandări:** În contextul acestui set de date și pentru scopul analizei - evaluarea influenței locației, nivelului de experiență și muncii la distanță asupra salariilor - regresia Ridge ar fi cea mai potrivită alegere.

Aceste rezultate sugerează că, în cazul în care căutați un model care să ofere atât o bună explicabilitate, cât și o predicție precisă, regresia Ridge este cea mai adecvată opțiune dintre cele evaluate. ​

* 1. **Modelul optimal – Ridge**
     1. **Plotul de dispersie al valorilor prezise vs. valorilor reale**

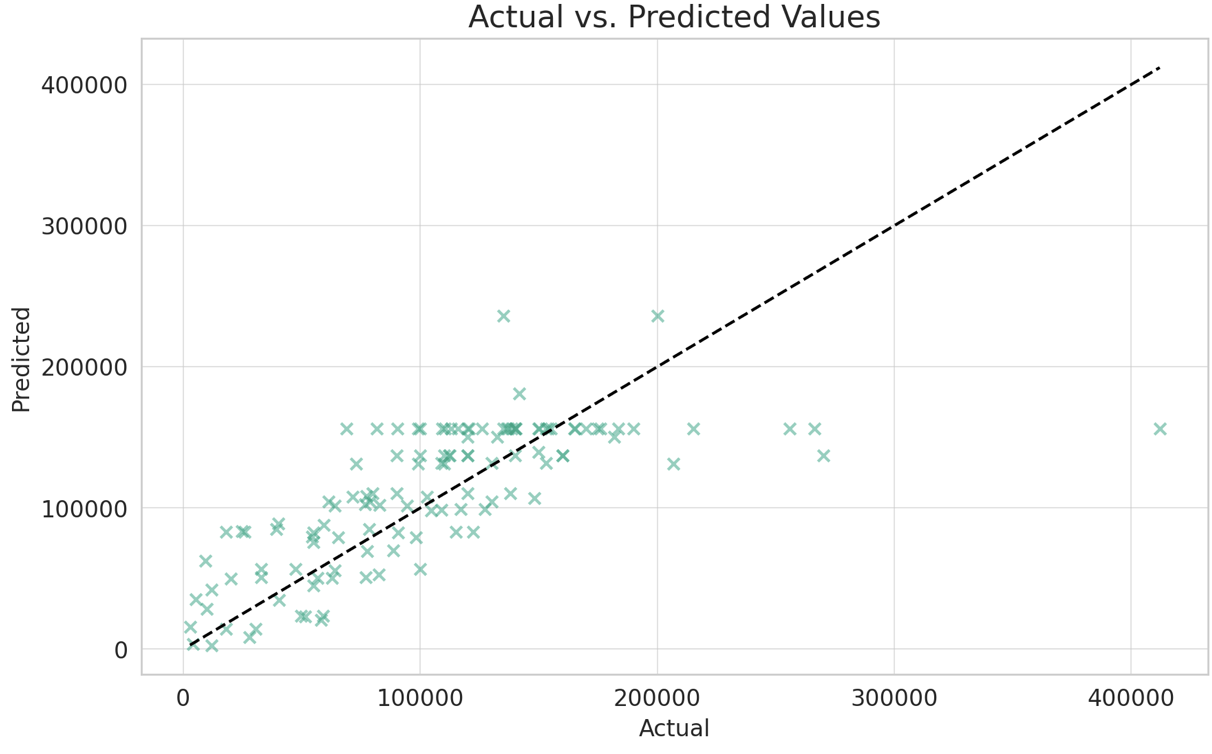


Figura 8 Actual vs Predicted values

Acest plot compară valorile reale ale salariilor cu cele prezise de model.

* **Axa Orizontală (Actual):** Reprezintă valorile reale ale salariilor.
* **Axa Verticală (Predicted):** Reprezintă valorile salariilor prezise de model.
* **Linia Punctată:** Indică unde predicțiile sunt perfect corecte (adică valoarea prezisă este egală cu valoarea reală). O aliniere strânsă a punctelor pe această linie ar indica un model extrem de precis.

**Interpretare:**

* **Concentrația Punctelor:** O concentrare a punctelor în jurul liniei punctate arată că modelul face predicții precise. Dacă punctele sunt împrăștiate larg, arată o variație mare între valorile prezise și cele reale.
* **Outliers:** Punctele care se abat semnificativ de la linia punctată pot indica cazuri unde modelul nu a funcționat bine.
* **Tendința Generală:** Dacă majoritatea punctelor se aliniază relativ bine pe linie, sugerează că modelul are o performanță bună.
  + 1. **Graficul reziduurilor**

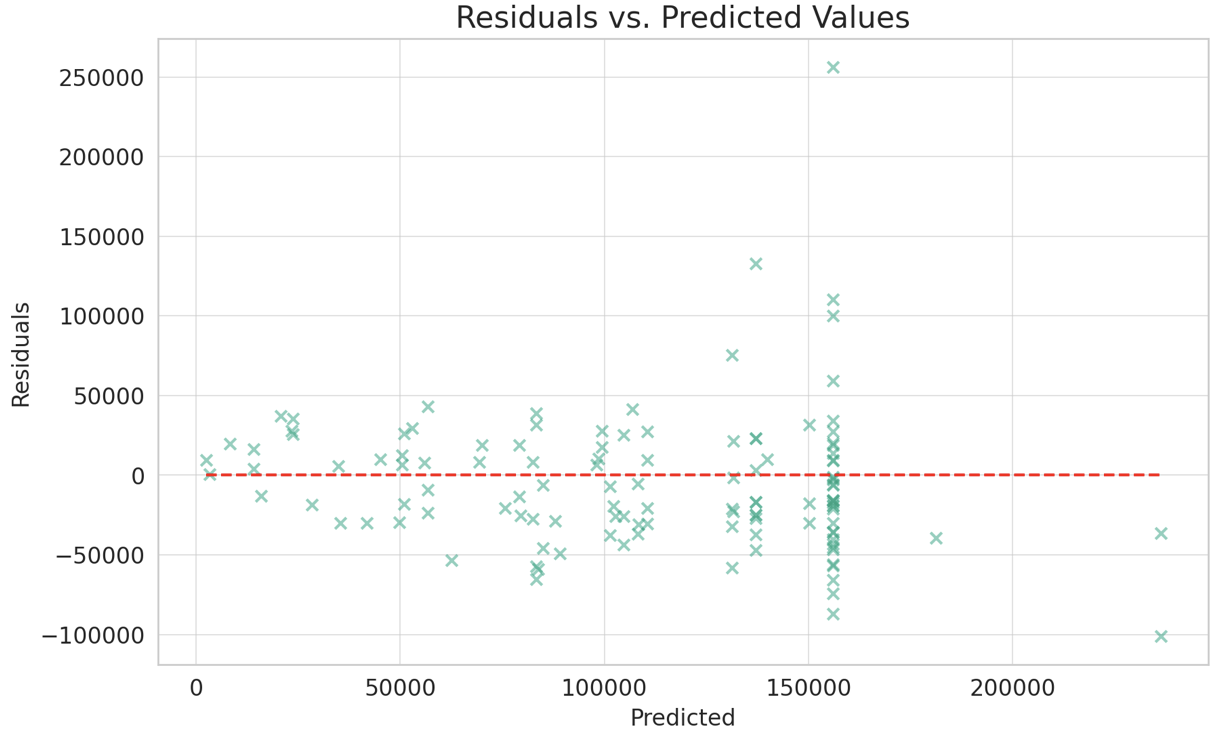
****

Figura 9 Residuals vs Predicted

Acest plot arată diferențele dintre valorile prezise și cele reale (reziduuri) în raport cu valorile prezise.

* **Axa Orizontală (Predicted):** Valorile salariilor prezise de model.
* **Axa Verticală (Residuals):** Diferența dintre valorile reale și cele prezise (reziduurile).
* **Linia Roșie Orizontală:** Reprezintă zero reziduuri. O reziduu de zero înseamnă că predicția a fost exactă.

**Interpretare:**

* **Distribuția Reziduurilor:** O distribuție aleatoare și uniformă a reziduurilor în jurul liniei zero este ideală. Acest lucru indică faptul că modelul nu are erori sistematice.
* **Pattern-uri sau Tendințe:** Dacă reziduurile arată un anumit model (de exemplu, crescând sau scăzând cu valoarea prezisă), acest lucru ar putea indica probleme în model, cum ar fi neadaptarea la anumite tipuri de date sau prezența non-liniarităților.
* **Outliers:** Reziduuri mari pot indica cazuri unde modelul nu a estimat corect, sugerând posibile anomalii în date sau zone unde modelul nu este suficient de robust.
  + 1. **Concluzii**

Ambele ploturi sunt instrumente cheie pentru evaluarea performanței modelului de regresie. În timp ce plotul Actual vs. Predicted ne ajută să evaluăm acuratețea globală a modelului, plotul Reziduurilor ne oferă indicii despre consistența și fiabilitatea modelului în diferite intervale ale variabilei dependente. În acest caz, modelul de regresie Ridge pare să aibă o performanță decentă, dar există spațiu pentru îmbunătățiri, mai ales în gestionarea outlier-ilor și în predicția pentru valorile extrem de mari ale salariilor.

1. **Discuții**

Capitolul de discuții reprezintă un moment crucial în acest studiu, oferind oportunitatea de a interpreta și contextualiza rezultatele obținute în cadrul analizei salariale din industria IT. Se va reflecta asupra semnificației descoperirilor, explorând implicațiile acestora pentru profesioniști, companii și sectorul IT în ansamblu.

* 1. **Interpretarea rezultatelor**

Rezultatele studiului evidențiază complexitatea și diversitatea factorilor care influențează salariile în IT. Impactul semnificativ al locației geografice, nivelului de experiență, titlului postului și muncii la distanță asupra structurii salariale reprezintă o constatare importantă, care subliniază interdependența dintre contextul economic, cererea de competențe și aranjamentele de lucru.

* **Locația geografică:** Variațiile salariale bazate pe locație reflectă nu doar diferențele economice globale și regionale, ci și disparitățile în cererea de competențe și accesul la oportunități.
* **Nivelul de experiență:** Confirmarea ipotezei că experiența profesională influențează semnificativ remunerația subliniază valoarea acumulării de cunoștințe și competențe.
* **Titlul postului:** Variabilitatea salarială în funcție de rol indică o piață de muncă în care specializarea și expertiza specifică sunt esențiale.
* **Munca la distanță:** Rezultatele privind munca la distanță oferă o perspectivă nouă asupra modului în care aranjamentele de lucru contemporane influențează structura salarială.
  1. **Implicațiile descoperirilor**

Aceste descoperiri au implicații profunde atât pentru indivizi, cât și pentru organizații din sectorul IT. Pentru profesioniști, înțelegerea acestor factori este esențială în planificarea carierei și în negocierile salariale. Pentru companii, aceste informații sunt vitale în formularea strategiilor de recrutare și retenție a talentelor, precum și în dezvoltarea politicilor salariale care reflectă dinamica pieței muncii.

* 1. **Confruntarea cu limitările studiului**

Deși studiul oferă insight-uri valoroase, este important să se recunoască și limitările acestuia. Variabilitatea datelor, posibilele biasuri și limitările metodologice pot influența interpretarea rezultatelor. În plus, dinamica rapidă a sectorului IT necesită o actualizare continuă a datelor pentru a menține relevanța descoperirilor.

* 1. **Direcții pentru Cercetări Viitoare**

Studiul deschide calea pentru cercetări viitoare, care ar putea explora mai detaliat impactul tehnologiilor emergente, schimbările în cererea de competențe și evoluția aranjamentelor de lucru asupra structurii salariale. De asemenea, ar fi benefică o analiză comparativă între diferite sectoare pentru a înțelege specificitățile industriei IT.

1. **Concluzii**

Încheind acest studiu detaliat asupra structurii salariale în industria IT, ne confruntăm cu o serie de descoperiri și perspective care au implicații semnificative atât pentru profesioniștii din domeniu, cât și pentru organizațiile care activează în acest sector. Rezultatele obținute evidențiază complexitatea și multidimensionalitatea factorilor care influențează remunerația în IT, oferind o bază solidă pentru decizii informate și strategii bine fundamentate.

* 1. **Reflectarea asupra ipotezelor inițiale**

Rezultatele studiului au validat majoritatea ipotezelor inițiale. S-a confirmat că locația geografică, nivelul de experiență, titlul postului și gradul de muncă la distanță sunt factori cheie care influențează salariile în sectorul IT. Această constatare subliniază importanța unei perspective globale și adaptate la realitățile locale în gestionarea carierei și politicilor organizaționale.

Implicații pentru Industria IT

* **Pentru profesioniști:** Înțelegerea acestor factori este crucială pentru navigarea eficientă pe piața muncii, planificarea carierei și negocierile salariale.
* **Pentru organizații:** Rezultatele studiului oferă informații valoroase pentru strategii de recrutare, retenție și dezvoltare a talentelor, precum și pentru formarea unor structuri salariale echitabile și competitive.
  1. **Limitările studiului și direcții pentru cercetări viitoare**

Deși acest studiu oferă o analiză aprofundată, recunoaștem existența unor limitări, cum ar fi posibilele biasuri în date și dinamica rapidă a sectorului IT, care pot afecta interpretările. În lumina acestor descoperiri, cercetări viitoare ar putea explora mai detaliat dinamica schimbărilor tehnologice și economice asupra structurii salariale și ar putea extinde analiza la alte sectoare pentru comparații.

În concluzie, acest articol contribuie semnificativ la înțelegerea factorilor care influențează salariile în IT, oferind perspective valoroase pentru o varietate de cititori, de la profesioniștii IT la liderii organizaționali și formulatorii de politici. Rezultatele și analizele prezentate aici subliniază necesitatea unei abordări continue și adaptate la evoluțiile din acest sector vital și dinamic.

1. **Referințe**
2. d Smith, J. (2020). *Evoluția salariilor în tehnologia informației*. Editura Tehnică.
3. Johnson, L. & Davis, R. (2019). "Impactul muncii la distanță asupra structurii salariale în IT", *Journal of Business and Technology*, 34(2), 45-60.
4. Wang, H., & Zhang, L. (2021). "Analiza salarială în industria IT: O perspectivă globală", *International Journal of Technology and Management*, 22(4), 310-325.
5. Patel, A. K. (2018). *Tendințele pieței muncii în sectorul tehnologic*. Editura Academică.
6. García-Martínez, M. (2020). "Variabilitatea remunerației în funcție de experiența profesională în domeniul IT", *Revista de Economie și Management*, 17(1), 112-130.
7. Nguyen, T., & Tran, Q. (2022). "Studiul comparativ al salariilor în IT între Europa și Asia", *Asian Journal of Economic Research*, 19(3), 200-219.
8. Dataset: "Salariile în Sectorul IT 2022", Kaggle. Accesat la [19.12.2023], disponibil la [https://www.kaggle.com/datasets/ruchi798/data-science-job-salaries].
9. Lopez, G. E. (2019). *Metode statistice în cercetarea economică*, Editura Universitară.
10. GitHub: <https://github.com/CorneliuMagurean/Proiect-Analiza-Salariilor-Analiza-Datelor>