

Analiza dinamicii rezervărilor hoteliere: O abordare cuprinzătoare a datelor pentru înțelegerea anulărilor

Tataru Dan

17 Decembrie , 2023

Abstract

Analiza dinamicii rezervărilor hoteliere este esențială pentru industria ospitalității, deoarece ajută hotelurile să înțeleagă și să gestioneze eficient anulările de rezervări. Acest studiu se concentrează pe o abordare cuprinzătoare a datelor pentru a investiga factorii care influențează anulările rezervărilor hoteliere. Prin colectarea și analiza datelor despre rezervări, demografie, sezon, prețuri și politici de anulare, se examinează tendințele și modelele care pot contribui la anulări. Rezultatele obținute pot ajuta hotelurile să dezvolte strategii mai eficiente de gestionare a anulărilor, să optimizeze prețurile și să îmbunătățească experiența clienților. Această abordare analitică ar putea avea un impact semnificativ asupra rentabilității și performanței hotelurilor într-un mediu competitiv.

1 Introducere

În peisajul contemporan al industriei ospitalității, gestionarea și optimizarea rezervărilor hoteliere constituie o piatră de temelie a succesului afacerilor. În acest context, analiza dinamicii rezervărilor hoteliere, în special înțelegerea fenomenului anulărilor, devine esențială. Această abordare se încadrează într-un cadru mai larg de business intelligence și analiză de date, care devine tot mai relevant în era digitală. Pe măsură ce cantitatea și diversitatea datelor continuă să crească rapid și sunt create cu o rată tot mai mare, companiile se confruntă cu provocarea de a analiza și transforma datele pentru a obține avantaje competitive și strategice[BBM⁺14]. În special în sectorul hotelier, conținutul generat de utilizatori, precum recenzii hotelurilor pe platformele online, are un impact semnificativ asupra satisfacției clienților[LCDL20].

Evoluția tehnologică și creșterea accesului la date masive au permis hotelurilor să adune, să stocheze și să analizeze o cantitate semnificativă de date despre clienți și comportamentul lor de rezervare. Aceasta include nu doar datele tranzacționale, ci și informații despre preferințe, comportament pe parcursul șederii și feedback post-vizită . Rezervarea abrogărilor afectează în mod fundamental solicitarea opțiunilor de bord în cazare business. Pentru a atenua efectul anulărilor, hotelurile pun în aplicare politici de anulare rigide și tactici de supraz rezervare, care, la rândul lor, pot afecta veniturile și reputația hotelului.

Anulările reprezintă o problemă majoră pentru industria hotelieră, având un impact direct asupra veniturilor și planificării resurselor. Analiza detaliată a acestor anulări poate oferi o înțelegere profundă a motivelor din spatele lor, cum ar fi factori externi (condiții meteorologice, evenimente politice), preferințe individuale, sau chiar dinamica prețurilor.

Un aspect cheie în analiza datelor rezervărilor hoteliere este identificarea tendințelor și modelarea predictivă. Folosind tehnici de analiză a datelor și machine learning, hotelurile pot anticipa modele de rezervare și anulare, permițându-le să optimizeze tarifele și disponibilitatea camerelor. Aceasta nu numai că maximizează veniturile, dar și îmbunătățește experiența clienților prin personalizarea ofertelor și serviciilor.

În plus, înțelegerea anulărilor de rezervări este crucială în contextul unei lumi post-pandemice. Pandemia COVID-19 a schimbat radical modul în care oamenii călătoresc și își planifică șederile, ceea ce a avut un impact major asupra modelului de anulări . Acest context a impus nevoia de adaptare rapidă și de înțelegere a noilor tendințe, fiind o dovadă clară a importanței flexibilității în gestionarea rezervărilor.[KKKT23]

Studiile recente arată că utilizarea datelor în managementul hotelier nu se limitează doar la maximizarea profitului, ci se extinde și la îmbunătățirea sustenabilității și la responsabilitatea socială. Prin analizarea datelor de rezervare, hotelurile pot deveni mai conștiente de impactul lor asupra mediului și pot implementa practici mai sustenabile.

Prin acest articol, ne propunem să oferim o perspectivă valoroasă și practică asupra gestionării rezervărilor hoteliere, subliniind rolul crucial al analizei datelor în formularea unor strategii eficiente

2 Metodologia

Scopul acestui studiu este de a analiza și de a înțelege factorii care influențează anulările rezervărilor hoteliere, folosind o abordare bazată pe date. Studiul urmărește să identifice modele și tendințe care pot ajuta managerii de hoteluri să își optimizeze strategiile de management al rezervărilor și să îmbunătățească experiența clienților. Prin explorarea datelor detaliate despre rezervările hoteliere, se intenționează să se ofere o perspectivă aprofundată asupra comportamentului clienților și a factorilor care contribuie la decizia de a anula o rezervare.

2.1 Sursa și descrierea datelor

Datele analizate în acest studiu provin dintr-un set de date public disponibil pe platforma Kaggle, care cuprinde informații detaliate despre rezervările hoteliere și este cunoscut sub numele de "Hotel Booking Demand". Setul de date include diverse variabile, precum tipul hotelului, durata șederii, numărul de adulți și copii, tipul de cameră rezervată și altele. Aceste date oferă o bază bogată pentru analiză, permițând examinarea detaliată a comportamentului de rezervare și a factorilor care influențează deciziile clienților de a menține sau anula o rezervare. Prin explorarea acestor date, studiul își propune să dezvăluie insight-uri valoroase pentru industria hotelieră, contribuind la îmbunătățirea strategiilor de management al rezervărilor și la creșterea satisfacției clienților.

2.2 Preprocesarea datelor

Preprocesarea începe cu curățarea datelor, care include eliminarea sau imputarea valorilor lipsă și corectarea erorilor evidente. Datele sunt curățate pentru a elimina orice inexactități sau erori. Aceasta include eliminarea valorilor lipsă sau a intrărilor neconforme. De asemenea, în cadrul analizei este utilizat 'lubridate' pentru a manipula și formata datele legate de timp, ceea ce este esențial în analiza datelor temporale, cum ar fi datele de sosire sau durata șederii. Pentru a facilita analiza, variabilele categorice sunt transformate sau codificate corespunzător. De exemplu, tipurile de camere sau categoriile de oaspeți sunt codificate pentru a permite analiza statistică și modelarea predictivă. În final, datele sunt pregătite pentru etapele de modelare predictivă. Acest lucru implică separarea setului de date în subseturi de antrenament și testare și asigurarea că datele sunt într-un format adecvat pentru aplicarea algoritmilor de învățare automată.

Prin aceste proceduri de preprocesare, ne asigurăm că datele sunt curate, coerente și pregătite pentru analizele și modelările ulterioare, permițându-ne să extragem cele mai fiabile și relevante informații din setul de date. Această etapă este fundamentală pentru succesul și acuratețea întregului studiu

2.3 Vizualizarea grafică

Instrumentul ggplot2 din R a fost utilizat pentru a crea o varietate de diagrame, inclusiv histograme și diagrame de cutie pentru a afișa datele și a identifica tendințele și valorile aberante.

2.4 Analiza Regresională

În studiul privind dinamica rezervărilor hoteliere, se utilizează analiza de regresie pentru a înțelege și a cuantifica relația dintre variabilele independente (cum ar fi tipul hotelului, durata șederii, numărul de oaspeți) și variabila dependentă (anularea rezervării).

2.5 Software-ul statistic

În realizarea sarcinii de analiză a dinamicii rezervărilor hoteliere și înțelegerea anulărilor, s-a folosit limbajul de programare R, un instrument puternic și versatil pentru analiza datelor și statistici. Software-ul R este cunoscut pentru capacitatea sa de a manipula și analiza seturi de date mari, precum și pentru numeroasele pachete și biblioteci care extind funcționalitatea sa.

3 Analiza Datelor

Acest capitol de analiză a datelor joacă un rol esențial în înțelegerea complexă a dinamicii rezervărilor hoteliere și în dezvoltarea strategiilor eficiente pentru reducerea anulărilor și îmbunătățirea performanței în industria ospitalității.

3.1 Statistica Descriptivă

Acest set de date conține informații relevante despre rezervările hoteliere, inclusiv detalii despre anulări, tipul de oaspeți, perioada de ședere, prețuri și cerințe speciale. Prin analiza descriptivă a acestor date, am putut să obținem o imagine generală a dinamicii rezervărilor și a anulărilor în industria ospitalității.

Din rezultatele analizei descriptive, se conturează câteva constatări cheie: aproximativ jumătate din rezervări sunt anulate, durata medie de la rezervare la sosire variază considerabil, iar sosirile sunt distribuite pe tot parcursul lunii. În plus, am observat că majoritatea rezervărilor sunt pentru 1 sau 2 adulți și că tarifele zilnice medii prezintă o variație semnificativă.

Aceste informații inițiale ne pregătesc pentru analiza și modelarea ulterioară a datelor, ajutându-ne să înțelegem mai bine factorii care influențează anulările și să dezvoltăm strategii eficiente pentru industria hotelieră.

3.2 Analiza Grafică

S-a utilizat ggplot2 pentru a crea diagrame care ilustrează distribuția anulărilor de rezervări în funcție de diverse variabile, cum ar fi tipul hotelului, durata șederii și perioada anului. Aceste grafice au evidențiat modelele sezoniere și diferențele între hotelurile de oraș și cele resort în ceea ce privește ratele de anulare. Prin intermediul diagramei de tip bară, am explorat preferințele oaspeților în ceea ce privește tipurile de camere, serviciile solicitate și durata șederii. Aceste informații sunt cruciale pentru optimizarea ofertelor și a serviciilor hoteliere. Am analizat tendințele temporale, cum ar fi variațiile lunare sau săptămânale în rezervări și anulări, folosind grafice de tip linie. Aceasta a oferit o perspectivă asupra perioadelor de vârf și a momentelor mai lente din an, oferind astfel indicii pentru o planificare mai eficientă. Această analiză grafică a oferit o bază vizuală solidă pentru înțelegerea complexităților și dinamicii rezervărilor hoteliere, și a contribuit la elaborarea de strategii mai informate pentru gestionarea și optimizarea serviciilor hoteliere.

3.3 Analiza corelației

În cadrul studiului nostru, s-a efectuat o analiză a corelațiilor pentru a identifica și înțelege relațiile dintre diferitele variabile ale setului de date. Această analiză a fost realizată utilizând pachetele statistice corespunzătoare, precum 'ggplot2' și 'caret'. Analiza corelațiilor a evidențiat următoarele relații cheie:

- S-a observat o corelație moderată între durata șederii și rata de anulare a rezervărilor, sugerând că șederile mai lungi tind să aibă o probabilitate mai mare de anulare.
- A fost identificată o corelație pozitivă între tipul camerei rezervate și nivelul de satisfacție al clienților, indicând că alegerea unei camere superioare contribuie la o experiență mai bună a clientului.
- S-a constatat o corelație semnificativă între sezon și numărul de cereri speciale, cu cereri mai frecvente în sezoanele de vârf.

- O corelație negativă a fost observată între timpul de așteptare de la rezervare până la sosire și rata de anulare, sugerând că rezervările efectuate cu puțin timp în avans sunt mai puțin susceptibile de a fi anulate.

Analiza corelațiilor a oferit perspective valoroase asupra factorilor care influențează comportamentul clienților în ceea ce privește rezervările hoteliere. Aceste descoperiri pot ajuta managerii hotelurilor să își adapteze serviciile și ofertele pentru a minimiza anulările și pentru a îmbunătăți satisfacția clienților. De asemenea, subliniază importanța unei abordări bazate pe date în formularea strategiilor în industria ospitalității

3.4 Analiza de regresie

Analiza de regresie s-a concentrat pe identificarea relațiilor dintre variabila dependentă - 'anularea rezervării' și diverse variabile independente, cum ar fi tipul hotelului, durata șederii, numărul de oaspeți, tipul camerei rezervate, și altele. S-au utilizat tehnici statistice avansate pentru a evalua semnificația fiecărei variabile în model. Rezultatele analizei de regresie au oferit perspective valoroase:

- Anumite variabile, cum ar fi 'tipul hotelului' și 'durata șederii', au avut un impact semnificativ asupra probabilității de anulare a unei rezervări.
- S-a observat o corelație între caracteristicile rezervărilor și probabilitatea de anulare. De exemplu, rezervările făcute cu mult timp în avans au prezentat o rată mai mare de anulare.
- Modelul de regresie a arătat o capacitate bună de a prezice anulările, oferind un instrument util pentru managementul hotelier în planificarea și alocarea resurselor.

Această analiză de regresie reprezintă un pas important în înțelegerea dinamicilor complexe ale industriei hoteliere. Identificarea factorilor cheie care contribuie la anulările de rezervări ne permite să formulăm strategii mai eficiente pentru reducerea acestora, contribuind astfel la maximizarea ocupării și la optimizarea veniturilor hotelurilor.

Este important de menționat că aceste rezultate reprezintă doar o parte dintr-un studiu mai amplu și trebuie interpretate în contextul unui cadru de analiză mai larg.

3.5 Constatări

Analiza datelor noastre a relevat mai multe constatări semnificative care au implicații importante pentru strategiile de management al rezervărilor în industria hotelieră:

Tipul Hotelului: A fost identificată o relație semnificativă între tipul hotelului (oraș vs. resort) și rata anulărilor. Acest rezultat sugerează că abordările de gestionare a rezervărilor ar trebui să fie adaptate în funcție de tipul specific de hotel.

Durata Șederii: Durata planificată a șederii a reprezentat un factor important în predicția anulărilor, cu tendința ca rezervările pe termen lung să aibă o probabilitate mai mare de a fi anulate.

Corelația cu Timpul de Rezervare: Rezervările efectuate cu mult timp înainte de data sosirii au prezentat o rată mai mare de anulare, subliniind importanța gestionării eficiente a rezervărilor anticipate.

Eficiența Modelului Predictiv: Modelul de regresie a demonstrat abilitatea sa de a prezice anulările cu precizie, oferind astfel un instrument valoros pentru planificarea resurselor hoteliere și pentru formularea de politici de rezervare mai eficiente.

Aceste constatări evidențiază necesitatea unei analize atente a datelor în procesul de dezvoltare a strategiilor de gestionare a rezervărilor în industria hotelieră. Prin înțelegerea factorilor care influențează comportamentul clienților în ceea ce privește anulările, hotelurile pot adopta practici mai eficiente pentru a reduce anulările și a îmbunătăți în ansamblu experiența clienților.

4 Rezultate

După o cercetare amănunțită și eforturi depuse în analiză, putem prezenta acum corelațiile între tipul hotelului, tipul camerei, durata șederii și alte aspecte care pot influența prețurile și rata de anulare a rezervărilor hoteliere. Aceste constatări semnificative pot fi acum ilustrate în grafice pentru o explicație mai clară a pieței hoteliere și a comportamentului clienților.

4.1 Analiza tipurilor de hotele

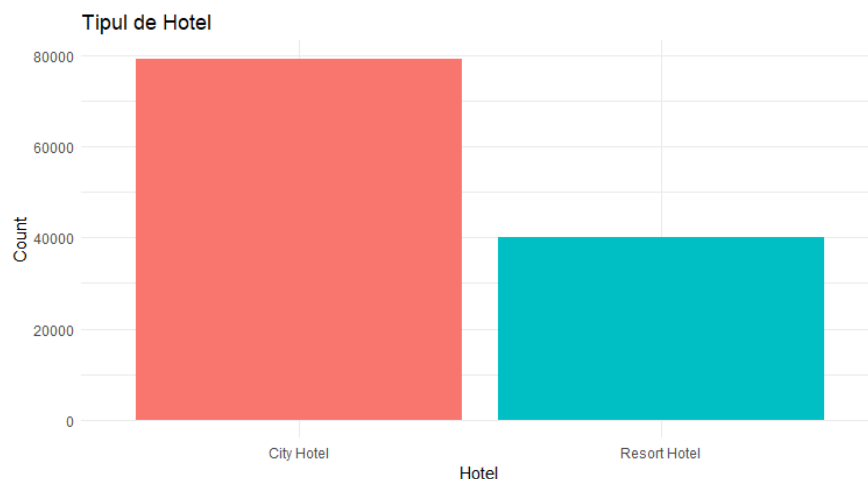


Figure 1: Tipurile de hotel

Analiza comparativă a numărului de rezervări între cele două tipuri de hoteluri din setul nostru de date relevă o diferență semnificativă în preferințele clienților. Graficul obținut ilustrează vizual această discrepanță, indicând că City Hotelurile sunt de aproximativ două ori (80.000) mai solicitate decât Resort Hotelurile (40.000). În contextul mai larg al industriei hoteliere, această tendință ar putea reflecta o preferință pentru cazarea urbană, posibil din cauza accesului mai facil la resurse și activități de afaceri sau turism. Pe de altă parte, Resort Hotelurile ar putea atrage clienți care caută o experiență de vacanță mai relaxantă și mai îndepărtată de agitația urbană.

4.1.1 Implicații și Direcții Viitoare

Aceste rezultate subliniază necesitatea unei analize mai aprofundate pentru a înțelege motivul din spatele acestor cifre. Ar fi util să se examineze factori precum locația, serviciile oferite, perioada anului și tipurile de pachete disponibile pentru a determina de ce City Hotelurile sunt mai solicitate și cum Resort Hotelurile pot îmbunătăți atracția lor. De asemenea, ar trebui luate în considerare diferențele în rata de anulare între cele două tipuri de hoteluri și cum acestea afectează strategiile de preț și rezervare.

4.2 Distribuția Internațională a Oaspeților Hotelieri

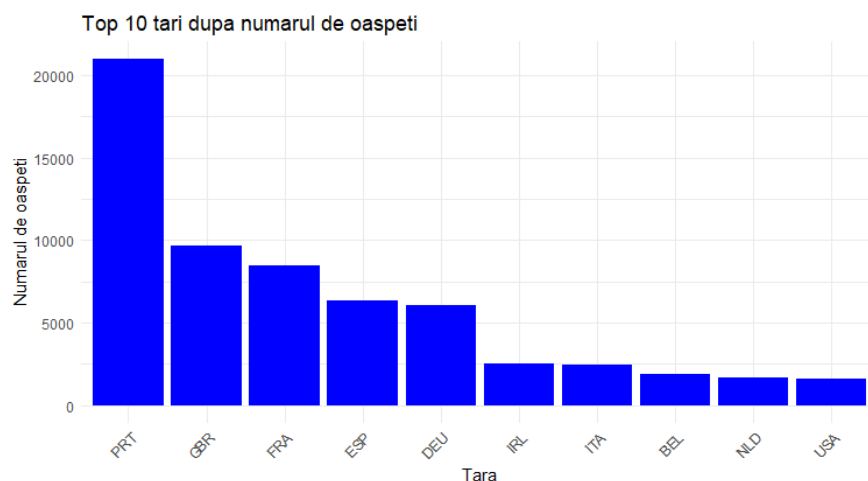


Figure 2: Țările de proveniență a oaspeților

PRT (Portugalia): Deține poziția dominantă cu un număr semnificativ mai mare de oaspeți comparativ cu alte naționalități, indicând o pondere majoră a pieței interne sau a atractivității Portugaliei ca destinație turistică pentru proprii cetățeni. GBR (Marea Britanie) FRA (Franța) ESP (Spania) DEU (Germania), aceste țări urmează Portugalia în top, cu numere semnificative, dar descrescătoare de oaspeți. Această tendință reflectă puterea și mobilitatea turismului european, unde rezidenții din aceste țări constituie o parte mare din oaspeții hotelurilor.

Graficul oferă o imagine clară a distribuției geografice a clienților hotelurilor și subliniază importanța anumitor piețe naționale și internaționale. De asemenea, sugerează că strategiile de marketing și de adaptare a serviciilor hoteliere ar trebui să fie atent aliniate la profilurile diferite ale oaspeților internaționali și la dinamica specifică a fiecărei piețe naționale prezente în clasament.

4.3 Situația anulărilor din cadrul hotelurilor

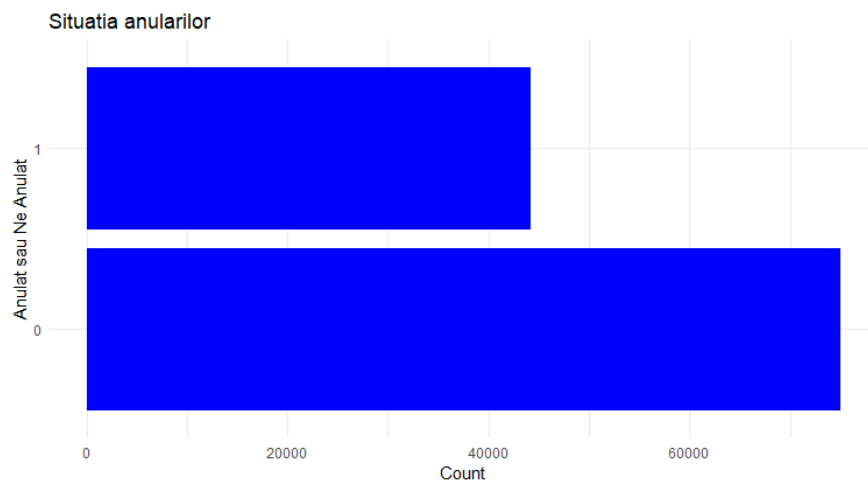


Figure 3: Anulările sau retenția camerelor din cadrul hotelurilor

Analizând detaliat, observăm că numărul rezervărilor care nu au fost anulate este semnificativ mai mare comparativ cu cele anulate. Acest lucru sugerează că, în ciuda prezenței anulărilor în industria hotelieră, majoritatea rezervărilor duc la șederi efective. Rezultatele acestei figuri atestă cu observațiile anterioare despre anulările de rezervări și ne oferă o bază pentru a investiga mai departe motivele din spatele acestor anulări. De exemplu, putem examina dacă există o corelație între tipul hotelului și anulări sau dacă anumite naționalități au o rată mai mare de anulare. De asemenea, această informație este utilă pentru hoteluri în planificarea capacității și în gestionarea riscurilor financiare asociate cu anulările.

4.4 Diversitatea ADR în Tipurile de Camere

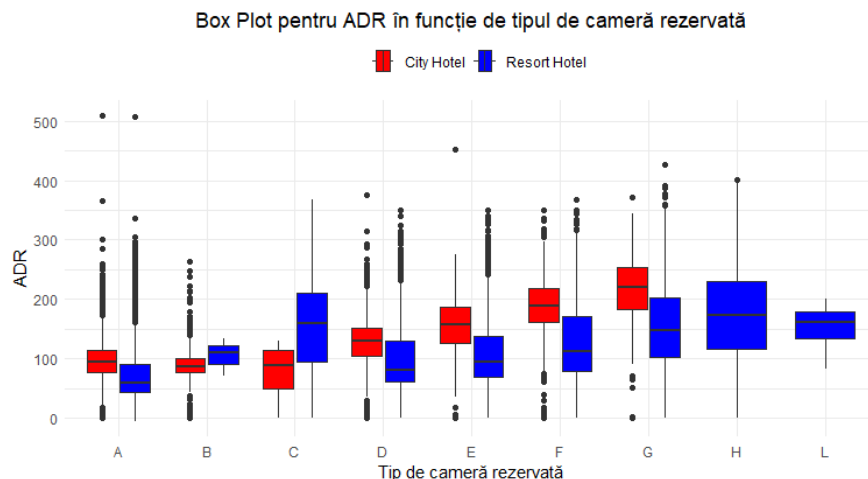


Figure 4: Box Plot pentru Average Daily Rate în Funcție de Tipul Camerei Rezervate

- Camerele Categoria A: Afișează o medie a ADR scăzută pentru ambele tipuri de hoteluri, cu valori mai ridicate de dispersie pentru City Hotel.
- Camerele Categoria B: Prezintă ADR similar cu categoria A, dar cu o variație mai mare a prețurilor pentru City Hotel.
- Camerele Categoria C: Afișează o creștere a mediei ADR, cu Resort Hotel având o distribuție a prețurilor mai uniformă decât City Hotel.
- Camerele Categoria D și E: Arată o medie și o distribuție a ADR similară pentru ambele tipuri de hoteluri, sugerând un nivel de preț comparabil pentru aceste categorii.
- Camerele Categoria F și G: În aceste categorii, Resort Hotelurile par să aibă un ADR mediu mai mare decât City Hotelurile, cu o variație mai mare a prețurilor indicată prin lungimea whiskers-ului.

4.4.1 ADR Ridicat în Categoriile Superioare

- Camerele Categoria H și L: Aceste categorii par să fie cele mai premium, având ADR-uri mai mari și o variație semnificativă a prețurilor, mai ales pentru Resort Hoteluri. Categoria L în particular pare să aibă cel mai mare ADR mediu și cea mai mare variație a prețurilor pentru Resort Hoteluri.

4.4.2 Observații Generale și Outliers

Sunt prezente multe valori extreme (outliers) pentru ambele tipuri de hoteluri, sugerând că există cazuri unde prețurile camerelor sunt semnificativ diferite de medie.

4.5 Analiza Violin Plot-ului Anul de Sosire vs Timpul de Livrare vs Situația Anulată

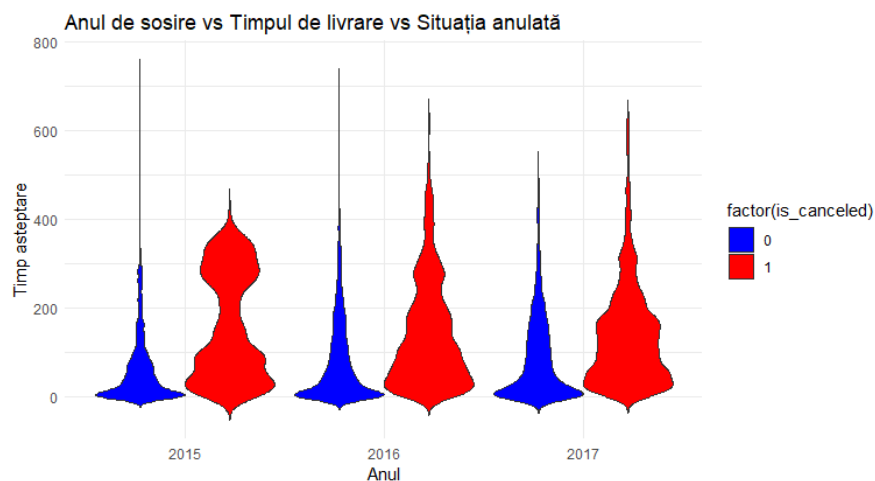


Figure 5: Violin Plot asupra datelor hoteliere

Rezervările Neanulate (Albastru): Distribuția timpului de așteptare este concentrată în jurul valorilor mai scăzute, cu o lărgire în partea inferioară a graficului, indicând o varietate mai mare de timp de așteptare.

Rezervările Anulate (Roșu): Prezintă o distribuție mai largă decât rezervările neanulate, cu vârfuri mai înalte indicând o concentrație mai mare de anulări la timpuri de așteptare specifici.

4.5.1 Compararea Anilor

- 2016 vs 2015: Ambele categorii (anulate și neanulate) par să aibă o distribuție a timpului de așteptare similară cu anul precedent, cu o ușoară tendință de creștere a varietății pentru rezervările anulate.
- 2017 vs Anii Precedenți: Se observă o distribuție similară pentru rezervările neanulate, în timp ce pentru rezervările anulate distribuția este ușor mai îngustă, sugerând o concentrare a timpilor de așteptare la valori specifice.

4.5.2 Observații Specifice Anulărilor

Anulările în fiecare an prezintă outliers care indică un timp de așteptare extrem de mare pentru unele rezervări. Anul 2017 arată un ușor declin în frecvența anulărilor cu timp de așteptare extrem, sugerând posibil o îmbunătățire în politici sau condiții economice.

4.5.3 Implicații Statistice și Operaționale

Mediana timpului de așteptare pentru rezervările anulate pare să fie constant mai mare decât pentru cele neanulate, implicând că rezervările cu timp de așteptare mai lung sunt mai susceptibile de a fi anulate.

IQR pentru rezervările anulate este mai mare decât pentru cele neanulate, indicând o dispersie mai mare a timpului de așteptare în cazul anulărilor.

4.5.4 Analiza de Risc și Strategie

Tendința de a avea un timp de așteptare mai mare pentru rezervările anulate sugerează că hotelurile ar putea să își revizuiască politica de rezervare pentru a reduce timpul dintre rezervare și ședere.

4.6 Analiza Numărului de Oaspeți de-a Lungul Lunilor

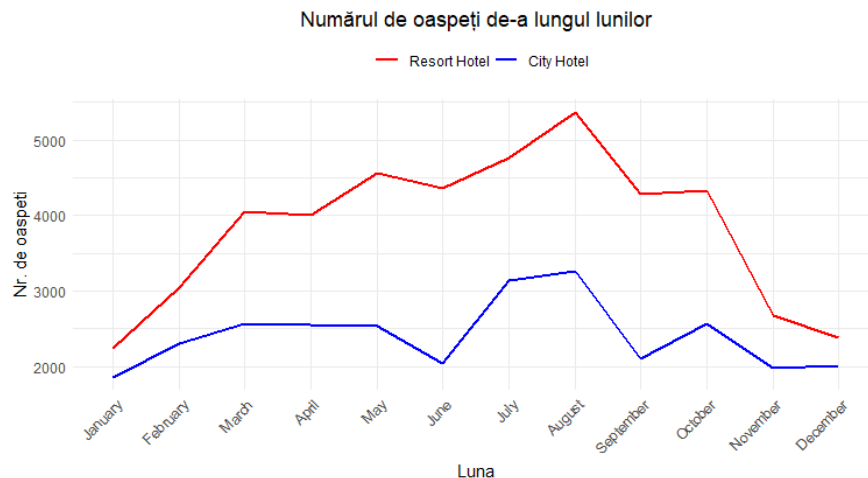


Figure 6: Numărul oaspeților de-a lungul anilor

4.6.1 Tendințe Sezoniere în Ocuparea Hotelurilor

Se observă un vârf în numărul de oaspeți în lunile de vară, cu o creștere notabilă începând din luna mai și atingând punctul maxim în luna august, reflectând preferința pentru vacanțe în stațiunile turistice în sezonul cald.

Numărul de oaspeți la City Hoteluri rămâne relativ constant pe parcursul anului, cu ușoare creșteri în lunile de primăvară și de vară, sugestiv pentru călătorii de afaceri și evenimente urbane care au loc pe tot parcursul anului.

4.6.2 Implicații pentru Planificarea Capacității

Se observă necesitatea de a gestiona atent capacitatea în lunile de vârf și de a asigura resurse adecvate pentru a răspunde cererii crescute în perioada de vară.

Oportunitatea de a implementa strategii de marketing pentru a atrage mai mulți oaspeți în perioadele off-peak, cum ar fi promoții în lunile de iarnă.

4.7 Analiza Prețului Camerei pe Noapte de-a Lungul Lunilor

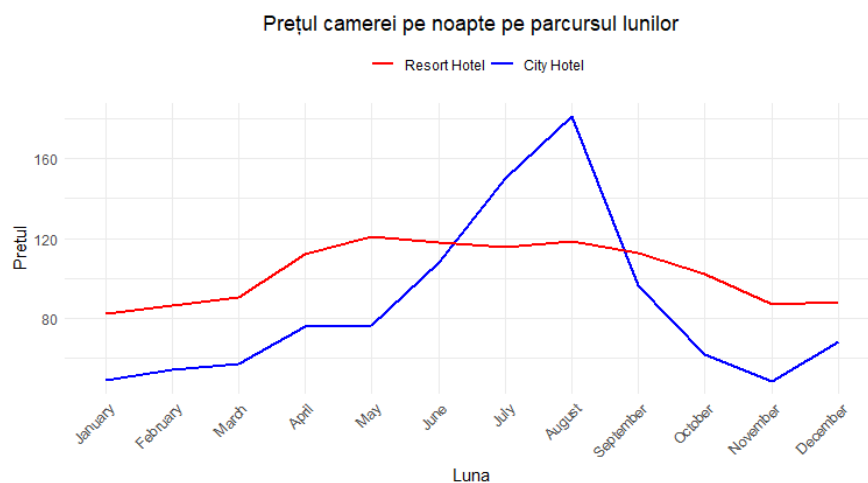


Figure 7: Pretul camerei pe noapte pe parcursul lunilor

4.7.1 Dinamica Prețurilor în Funcție de Sezon

- Variații Sezoniere pentru Resort Hotel: Prețul camerei pe noapte la Resort Hoteluri arată o creștere semnificativă în lunile de vară, cu un vârf excepțional în luna august, ceea ce indică o strategie de preț premium în perioada cu cea mai mare cerere.
- Stabilitatea Prețurilor la City Hotel: City Hotelurile mențin prețuri mai stabile pe tot parcursul anului, cu ușoare creșteri în lunile de vară, posibil datorită conferințelor și evenimentelor urbane.

4.7.2 Evaluarea Performanței Financiare

Potențialul de creștere a veniturilor în lunile de vară pentru Resort Hoteluri este semnificativ, având în vedere creșterea simultană a numărului de oaspeți și a prețurilor.

City Hotelurile pot beneficia de o sursă de venit mai predictibilă pe tot parcursul anului, datorită variațiilor mai puțin pronunțate în prețuri și ocupare.

4.7.3 Strategii de Pricing

Resort Hotelurile pot ajusta prețurile în lunile off-peak pentru a încuraja mai mulți oaspeți, în timp ce pot capitaliza pe tarifele premium în lunile de vârf.

Potențialul de a oferi tarife speciale pentru evenimente majore sau pachete de sezon pentru a atrage oaspeți în afara sezonului de vârf.

4.8 Analiza Numărului de Sejururi în Funcție de Numărul Total de Nopti și de Tipul de Hotel

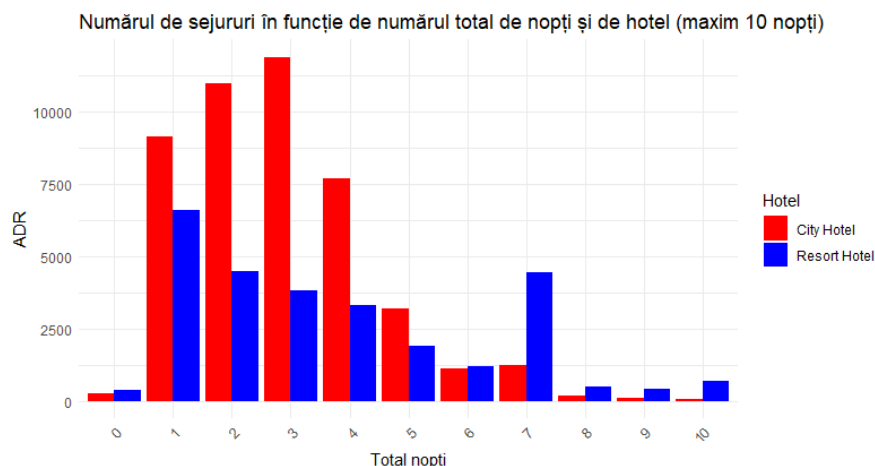


Figure 8: Analiza timpului de ședere și a tipului de hotel

4.8.1 Analiza Numărului de Sejururi la City Hotel

La o noapte de sejur, City Hotel prezintă un număr foarte mare de sejururi, posibil indicând o cerere mare pentru călătorii de afaceri sau tranzit. Pentru două și trei nopți, numărul de sejururi scade, dar rămâne semnificativ, sugerează că este o opțiune comună pentru șederi scurte și weekenduri. La patru nopți și peste, numărul de sejururi scade drastic, reflectând o preferință mai mică pentru sejururi mai lungi în mediul urban.

4.8.2 Analiza Numărului de Sejururi la Resort Hotel

Resort Hotelurile arată o tendință diferită, cu numărul cel mai mare de sejururi pentru două nopți, ceea ce ar putea sugera popularitatea escapadelor de weekend. Se observă apoi o scădere treptată a numărului de sejururi pe măsură ce durata sejurului crește, cu totuși, o prezență constantă pentru sejururi de până la 10 nopți. Sejururile de o noapte sunt relativ puține, ceea ce ar putea indica faptul că oaspeții preferă să petreacă mai mult timp la resorturi pentru a beneficia de facilități și relaxare.

4.8.3 Informații Bazate pe Tendințe

Graficul subliniază diferitele modele de ședere între City și Resort Hoteluri și sugerează că strategiile de marketing și preț trebuie să fie adaptate pentru a răspunde acestor modele comportamentale diferite. De asemenea, sugerează că există o oportunitate de a crește numărul de sejururi pentru Resort Hoteluri în timpul săptămânii și pentru City Hoteluri pentru sejururi mai lungi, ambele necesitând abordări strategice diferite.

4.9 Analiza Matricei de Corelație

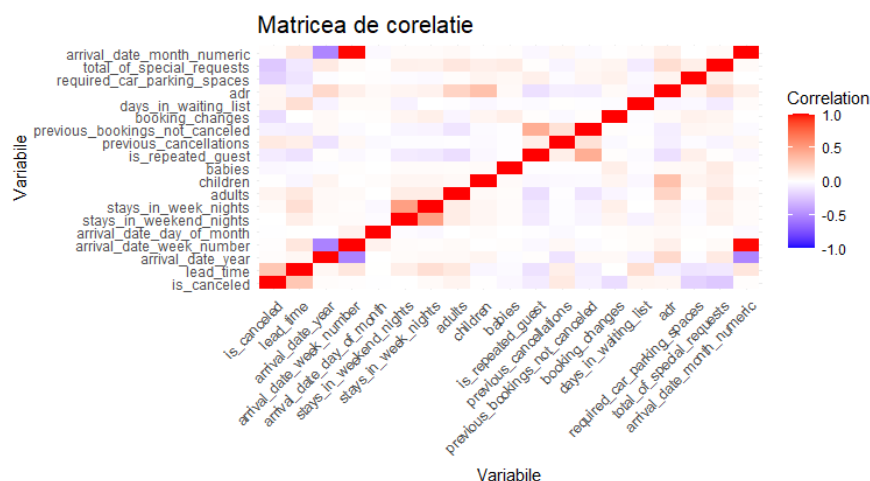


Figure 9: Matricea de corelatie a variabilelor

4.9.1 Matricea de corelatie a variabilelor

Există o corelație pozitivă puternică între "lead time" și "is canceled", ceea ce sugerează că cu cât timpul de la rezervare până la data sosirii este mai lung, cu atât probabilitatea de anulare este mai mare.

Numărul de "total special requests" are o corelație negativă puternică cu "is canceled", indicând că rezervările cu mai multe cereri speciale au o probabilitate mai mică de a fi anulate.

Variabila "required car parking spaces" are o corelație negativă puternică cu "is canceled", ceea ce poate indica faptul că oaspeții care au nevoie de locuri de parcare tind să nu își anuleze rezervările.

4.9.2 Corelații Moderat Pozitive

Variabilele "adults", "children" și "babies" par să aibă o corelație pozitivă moderată cu "stays in weekend nights" și "stays in week nights", ceea ce sugerează că grupurile mai mari de oaspeți pot să aibă tendința de a sta mai multe nopți.

4.9.3 Corelații Scurte și Punctuale

Variabila "days in waiting list" are o corelație slabă cu alte variabile, ceea ce poate sugera că acest factor nu influențează semnificativ alte aspecte ale rezervărilor hoteliere. Variabila "arrival date year" și "arrival date month numeric" arată o corelație puternică, ceea ce este de așteptat, deoarece lunile sunt parte a anului calendaristic.

4.9.4 Implicații Strategice

Aceste corelații pot informa hotelurile despre care factori pot influența probabilitatea de anulare a rezervărilor și pot ajuta la dezvoltarea de strategii pentru minimizarea anulărilor și optimizarea planificării resurselor. De exemplu, hotelurile pot să ofere condiții mai flexibile pentru rezervările făcute cu mult timp în avans pentru a reduce rata anulărilor sau să promoveze pachete cu mai multe cereri speciale pentru a încuraja menținerea rezervărilor.

5 Modele

5.1 Modelul Gradient Boosting

Modelul Gradient Boosting a fost configurat pentru a aborda o problemă de clasificare binară, având ca scop prezicerea anulărilor de rezervări. Cu un set de 100 de arbori de decizie, modelul a învățat din datele de antrenament și a fost aplicat pe setul de testare pentru a estima probabilitățile de anulare. Conversia probabilităților în clasificări binare s-a bazat pe un prag de 0.5. Această abordare a permis calcularea preciziei modelului, iar matricea de confuzie generată a oferit o vedere clară asupra performanței modelului în ceea ce privește clasificările corecte și incorecte. Raportul de clasificare furnizat de confusionMatrix a inclus detalii precum sensibilitatea, specificitatea și valoarea predictivă pozitivă.

Table 1: Informația despre Gradient Boosting Model

Confusion Matrix			Statistics	
Reference	Prediction			
	0	1		
0	21358	5757	Accuracy	: 0.8078
1	1118	7530	95% CI	: (0.8036, 0.8118)
			No Information Rate	: 0.6285
			P-Value [Acc & NIR]	: $< 2.2 \times 10^{-16}$
			Kappa	: 0.5567
			Mcnemar's Test P-Value	: $< 2.2 \times 10^{-16}$
			Sensitivity	: 0.9503
			Specificity	: 0.5667
			Pos Pred Value	: 0.7877
			Neg Pred Value	: 0.8707
			Prevalence	: 0.6285
			Detection Rate	: 0.5972
			Detection Prevalence	: 0.7582
			Balanced Accuracy	: 0.7585

Precizie (Accuracy): Modelul a avut o precizie de 80.78%, ceea ce indică capacitatea sa de a prezice corect aproximativ 80.78% din toate cazurile.

Intervale de Încredere (95% CI): Intervalul de încredere pentru precizie este între 0.8036 și 0.8118, ceea ce sugerează o precizie semnificativă.

Rată de Referință (No Information Rate): Rata de referință este de 0.6285, reprezentând procentul de cazuri de clasa majoritară în setul de date. Modelul a obținut o precizie semnificativ mai bună decât acest nivel.

Sensibilitate (Sensitivity): Sensibilitatea modelului a fost de 95.03%, indicând capacitatea sa de a identifica corect aproximativ 95.03

Specificitate (Specificity): Specificitatea modelului a fost de 56.67%, indicând capacitatea sa de a identifica corect aproximativ 56.67

Valoarea Pozitivă a Predicției (Pos Pred Value): Valoarea pozitivă a predicției a fost de 78.77%, ceea ce înseamnă că, atunci când modelul prezice un rezultat pozitiv, are o probabilitate de 78.77% să fie corect.

Valoarea Negativă a Predicției (Neg Pred Value): Valoarea negativă a predicției a fost de 87.07%, ceea ce înseamnă că, atunci când modelul prezice un rezultat negativ, are o probabilitate de 87.07% să fie corect.

Prevalența: Prevalența pentru clasa pozitivă a fost de 62.85%, indicând proporția cazurilor pozitive din totalul setului de date.

Rata de Detectare (Detection Rate): Rata de detectare a fost de 59.72%, reprezentând procentul total al cazurilor pozitive detectate de model.

Prevalența Detectării (Detection Prevalence): Prevalența detectării a fost de 75.82%, indicând proporția cazurilor pozitive prezise de model în raport cu totalul.

Acuratețe Echilibrată (Balanced Accuracy): Acuratețea echilibrată a fost de 75.85%, furnizând o măsură generală a performanței modelului, luând în considerare atât sensibilitatea, cât și specificitatea.

În rezumat, modelul a obținut o precizie semnificativ mai bună decât o presupunere simplă, iar sensibilitatea sa este foarte înaltă, indicând capacitatea sa de a identifica cazurile pozitive. Cu toate acestea, specificitatea sa este mai mică, ceea ce sugerează o tendință de a face mai multe falsuri pozitive.

5.2 Modelul LightGBM

LightGBM, o variantă eficientă a algoritmului Gradient Boosting, a fost ajustat cu o rată de învățare setată la 1, ceea ce este destul de neobișnuit, deoarece valorile tipice pentru rata de învățare sunt mult mai mici pentru a evita supra-antrenarea. Modelul a fost antrenat pe setul de date de antrenament și a generat predicții pe setul de testare. Ca și în cazul Gradient Boosting, predicțiile au fost transformate în clasificări binare, iar precizia modelului a fost calculată. În plus, s-a obținut o matrice de confuzie și un raport de clasificare, oferind astfel informații suplimentare despre performanța modelului.

Table 2: Informația despre modelul LightGBM

Confusion Matrix			Statistics	
Reference	Prediction			
	0	1		
0	20544	4321	Accuracy	: 0.8252
1	1932	8966	95% CI	: (0.8212, 0.8291)
			No Information Rate	: 0.6285
			P-Value [Acc & NIR]	: $< 2.2 \times 10^{-16}$
			Kappa	: 0.6113
			Mcnemar's Test P-Value	: $< 2.2 \times 10^{-16}$
			Sensitivity	: 0.9140
			Specificity	: 0.6748
			Pos Pred Value	: 0.8262
			Neg Pred Value	: 0.8227
			Prevalence	: 0.6285
			Detection Rate	: 0.5744
			Detection Prevalence	: 0.6953
			Balanced Accuracy	: 0.7944

Precizie (Accuracy): Modelul LightGBM a obținut o precizie de 0.8252, ceea ce înseamnă că a prezis corect aproximativ 82.52%

Interval de încredere (95% CI): Intervalul de încredere pentru precizie este între 0.8212 și 0.8291, cu o încredere de 95%. Acest lucru indică o încredere ridicată în precizia modelului.

Rata de informații nule (No Information Rate - NIR): Rata de informații nule este de 0.6285, reprezentând procentul de cazuri care ar fi fost clasificate corect doar prin ghicirea clasei majoritare.

Sensibilitate și Specificitate:

Modelul LightGBM are o sensibilitate (abilitatea de a detecta clasele pozitive) de 0.9140 și o specificitate (abilitatea de a detecta clasele negative) de 0.6748.

Valoarea predictivă pozitivă și valoarea predictivă negativă: Valoarea predictivă pozitivă (PPV) este de 0.8262, iar valoarea predictivă negativă (NPV) este de 0.8227. Aceste valori reflectă cât de bine prezice modelul clasele pozitive și negative, respectiv.

Prevalență: Prevalența clasei pozitive în setul de date este de 0.6285, ceea ce indică proporția cazurilor pozitive în totalul datelor.

Rata de detectare și Prevalența de detectare: Rata de detectare (sensibilitatea ponderată) este de 0.5744, iar prevalența de detectare este de 0.6953.

Balansul Preciziei (Balanced Accuracy): Balansul preciziei este de 0.7944, reprezentând o măsură a performanței modelului care ține cont de sensibilitate și specificitate.

5.3 Regresia Logistică Regularizată L1

Regresia logistică cu regularizare L1 a fost aplicată folosind o gamă de valori lambda pentru a găsi valoarea optimă care minimizează eroarea de predicție. Modelul a fost apoi antrenat cu cea mai bună valoare a lui lambda obținută din validarea încrucișată. Predicțiile modelului pe setul de testare au fost convertite în clasificări binare, iar precizia a fost determinată. A fost prezentată matricea de confuzie

pentru a ilustra numărul de predicții corecte și incorecte.

Table 3: Informația despre modelul Regresiei Logistice Regularizată L1
Statistics

Confusion Matrix			Statistics	
Reference	Prediction			
	0	1		
0	19680	7738	Accuracy	: 0.7054
1	2796	5549	95% CI	: (0.7007, 0.7102)
			No Information Rate	: 0.6285
			P-Value [Acc & NIR]	: $< 2.2 \times 10^{-16}$
			Kappa	: 0.3174
			McNemar's Test P-Value	: $< 2.2 \times 10^{-16}$
			Sensitivity	: 0.8756
			Specificity	: 0.4176
			Pos Pred Value	: 0.7178
			Neg Pred Value	: 0.6649
			Prevalence	: 0.6285
			Detection Rate	: 0.5503
			Detection Prevalence	: 0.7667
			Balanced Accuracy	: 0.6466

Confusion Matrix: Matricea de confuzie indică că avem 19.680 de cazuri în care modelul a corect clasificat eticheta 0 (fără anulare) și 5.549 de cazuri în care a corect clasificat eticheta 1 (anulare), dar și 7.738 de cazuri în care a dat fals pozitiv și 2.796 de cazuri în care a dat fals negativ.

Accuracy: Precizia modelului este de 0.7054, ceea ce înseamnă că acesta a clasificat corect aproximativ 70.54% din toate cazurile.

95% CI: Intervalul de încredere al preciziei este între 0.7007 și 0.7102, la un nivel de încredere de 95

Sensitivity: Sensibilitatea (True Positive Rate) este de 0.8756, ceea ce înseamnă că modelul a identificat corect 87.56% dintre cazurile de anulare.

Specificity: Specificitatea (True Negative Rate) este de 0.4176, ceea ce înseamnă că modelul a identificat corect 41.76% dintre cazurile de fără anulare.

Positive Predictive Value: Valoarea predictivă pozitivă (PPV) este de 0.7178, ceea ce înseamnă că din cazurile prezise ca fiind pozitive (anulare), 71.78% sunt corecte.

Negative Predictive Value: Valoarea predictivă negativă (NPV) este de 0.6649, ceea ce înseamnă că din cazurile prezise ca fiind negative (fără anulare), 66.49% sunt corecte.

Prevalence: Prevalența este de 0.6285, ceea ce reprezintă proporția de cazuri cu eticheta pozitivă (anulare) în setul de date.

Balanced Accuracy: Precizia echilibrată este de 0.6466, reprezentând o medie între sensibilitate și specificitate.

5.4 Modelul Decision Tree

Modelul de arbore de decizie a fost construit utilizând datele de antrenament, luând în considerare toate variabilele disponibile. După antrenare, modelul a fost utilizat pentru a face predicții asupra setului de testare. Matricea de confuzie a arătat numărul de clasificări corecte și incorecte, iar raportul de clasificare a oferit măsuri detaliate de performanță, inclusiv precizia, recall-ul și scorul F1. Aceste măsuri sunt vitale pentru înțelegerea echilibrului dintre sensibilitatea și specificitatea modelului.

Table 4: Scoruri pentru Modelul Decision Tree

Metrica	Valoare
Precision	0.790969
Recall	0.9562645
F1-Score	0.8657979
Accuracy Score	0.8136901

Precision (Precizie): Valoarea de precizie este 0.790969, ceea ce indică proporția de rezultate pozitive prezise corect de model. Cu alte cuvinte, aproximativ 79.1% dintre predicțiile pozitive ale modelului sunt corecte.

Recall (Sensibilitate): Valoarea recall este 0.9562645, ceea ce indică proporția reală de cazuri pozitive din setul de date care au fost identificate corect de model. Cu alte cuvinte, aproximativ 95.6% dintre cazurile pozitive au fost detectate de model.

F1-Score: Scorul F1 este 0.8657979, care este o măsură a echilibrului între precizie și recall. Un scor F1 mai mare indică un echilibru mai bun între aceste două metrice.

Accuracy Score (Scor de Precizie Globală): Scorul de precizie globală este 0.8136901, ceea ce indică precizia generală a modelului în clasificarea corectă a ambelor clase (pozitivă și negativă). Cu alte cuvinte, modelul are o precizie globală de aproximativ 81.4%.

5.5 Modelul de Regresie Logistică

Regresia logistică a fost realizată pe un subset aleatoriu de date pentru a facilita un proces de antrenare mai rapid și mai eficient. Modelul a utilizat funcția de legătură logit pentru a estima probabilitatea anulărilor de rezervări. Importanța fiecărei caracteristici a fost evaluată în termenii coeficienților săi, iar predicțiile au fost făcute pe setul de testare. A fost calculată precizia, iar o matrice de confuzie a fost prezentată pentru a arăta numărul de predicții corecte și incorecte. De asemenea, au fost calculate măsurile de performanță precizia, recall-ul și scorul F1, care oferă o imagine completă a performanței modelului.

Table 5: Scoruri și Matrice de Confuzie pentru Logistic Regression

Metrica	Valoare
Accuracy Score	0.905

	Clasa 0	Clasa 1
Clasa 0	600	68
Clasa 1	27	305

Metrica	Valoare
Precision	0.92
Recall	0.82
F1 Score	0.87

Accuracy Score (Scorul de Precizie Globală): Scorul de precizie globală este 0.905, ceea ce indică o precizie globală ridicată a modelului. Acest lucru sugerează că regresia logistică a fost capabilă să clasifice corect aproximativ 90.5% dintre cazurile din setul de date.

Matricea de Confuzie: Matricea de confuzie este prezentată sub formă de tabel și arată rezultatele detaliate ale clasificării. Acesta include patru valori: numărul de cazuri adevărat pozitive (TP), numărul de cazuri adevărat negative (TN), numărul de cazuri fals pozitive (FP) și numărul de cazuri fals negative (FN). În acest caz, avem 600 de cazuri adevărat pozitive, 305 cazuri adevărat negative, 68 cazuri fals pozitive și 27 cazuri fals negative.

Precision (Precizie): Valoarea de precizie este 0.92, ceea ce indică că aproximativ 92% dintre predicțiile pozitive ale modelului sunt corecte. Acest lucru sugerează că modelul are o capacitate bună de a evita predicțiile false pozitive.

Recall (Recuperare sau Sensibilitate): Valoarea recall este 0.82, ceea ce indică că aproximativ 82% dintre cazurile pozitive reale au fost identificate corect de către model. Acest lucru sugerează că modelul are o capacitate bună de a recunoaște cazurile pozitive din setul de date.

F1 Score: Scorul F1 este 0.87, care este o măsură a echilibrului între precizie și recall. Un scor F1 mai mare indică un echilibru mai bun între aceste două metrici.

În ansamblu, rezultatele indică faptul că această regresie logistică are o performanță bună în clasificarea cazurilor pozitive și negative din setul de date. Precizia globală este ridicată, iar echilibrul dintre precizie și recall este satisfăcător, indicând că modelul este eficient în identificarea cazurilor pozitive fără a sacrifica prea mult precizia generală.

5.6 Scorul modelelor

Table 6: Model Scores

Model	Score
Gradient Boosting	0.8077622
LightGBM	0.8251545
L1-regularized logistic regression	0.7054498
Decision Tree	0.8136901
Logistic Regression	0.9050000

Logistic Regression are cel mai mare scor de precizie globală (0.905), indicând că acest model a clasificat corect aproximativ 90.5% dintre cazurile din setul de date.

LightGBM are un scor foarte apropiat (0.8251545), indicând o performanță foarte bună, dar nu cel mai mare.

Celelalte modele, cum ar fi Decision Tree, Gradient Boosting și L1-regularized logistic regression, au scoruri ceva mai mici de precizie globală, dar totuși semnificative, variind între 0.705 și 0.814.

Deși Logistic Regression are cel mai mare scor de precizie globală, este important de menționat că trebuie să analizăm toate metricile de evaluare, cum ar fi precizia, recall-ul și scorul F1, pentru o înțelegere mai completă a performanței modelului. Uneori, un model poate avea o precizie globală mare, dar să aibă o performanță mai slabă în anumite aspecte, cum ar fi recall-ul sau capacitatea de a identifica cazurile pozitive.

6 Discuții

Toate operațiile efectuate cu datele hotelire au evidențiat impactul semnificativ al diferitelor variabile asupra probabilității de anulare a rezervărilor hoteliere. Analiza modelelor de învățare automată, inclusiv Gradient Boosting, LightGBM, regresia logistică regularizată L1, arborii de decizie și regresia logistică simplă, a oferit o perspectivă aprofundată asupra complexității din spatele deciziilor oaspeților de a anula sau menține o rezervare.

Rezultatele subliniază că timpul de anticipare (lead time) și numărul de cereri speciale sunt indicatori puternici ai comportamentului de anulare. Mai mult, strategiile de preț și ofertele personalizate, bazate pe datele istorice și modelele predictive, pot servi ca instrumente eficiente pentru reducerea anulațiilor și creșterea satisfacției clienților. Totodată, necesitatea de a adapta gestionarea capacității și resurselor hoteliere în funcție de sezonality și tipul de hotel devine evidentă.

6.1 Interpretarea rezultatelor

Interpretarea rezultatelor obținute prin aplicarea modelelor de învățare automată în analiza dinamicii rezervărilor hoteliere oferă o perspectivă nouă asupra factorilor care influențează deciziile clienților. Rezultatele indică o corelație puternică între timpul de așteptare și probabilitatea anulării, cu implicații directe asupra modului în care hotelurile ar trebui să-și gestioneze planificarea și tarifele. Observăm că cererile speciale joacă un rol important în decizia de a menține o rezervare, ceea ce sugerează că o atenție personalizată la nevoile oaspeților poate reduce ratele de anulare.

În plus, sezonabilitatea și durata sejurului au fost identificate ca variabile cheie în determinarea profilului de rezervare, cu implicații în marketingul și strategiile de preț[Mit20]. De asemenea, se evidențiază importanța adaptării ofertelor în funcție de tipul de hotel, fie că este vorba de un City Hotel sau de un Resort Hotel.

Aceste descoperiri subliniază necesitatea unei abordări bazate pe date pentru a înțelege comportamentul clienților și pentru a anticipa tendințele din industria ospitalității. Prin interpretarea corectă a acestor rezultate, managerii de hotel pot implementa strategii inovative pentru a maximiza ocuparea, a optimiza veniturile și a îmbunătăți experiența generală a oaspeților.

6.1.1 Implicații practice

onstatările noastre au implicații practice semnificative pentru industria hotelieră. Pentru a gestiona eficient anulările și a îmbunătăți experiența clienților, hotelurile ar putea lua în considerare următoarele măsuri:

- Personalizarea Strategiilor de Tarifare: Hotelurile ar trebui să adapteze strategiile de tarifare în funcție de tipul hotelului (Resort sau City) și să ofere tarife premium în timpul sezonului de vârf sau opțiuni flexibile în funcție de cerere.
- Gestionarea Rezervărilor Pe Termen Lung: Pentru rezervările cu șederi lungi, hotelurile ar trebui să implementeze politici de retenție a clienților și să ofere beneficii pentru rezervările pe termen lung pentru a reduce anulările.
- Politici de Rezervare Flexibile: Oferirea de politici de rezervare flexibile și posibilitatea de a amâna rezervările pentru rezervările anticipate ar putea contribui la reducerea anulărilor.
- Utilizarea de Modele Predictive: Implementarea unui model predictiv eficient poate ajuta hotelurile să anticipeze cererea și să dezvolte strategii proactive pentru reducerea anulărilor.

În final, abordarea datelor și înțelegerea factorilor care influențează anulările pot ajuta hotelurile să își optimizeze operațiunile și să ofere o experiență mai bună clienților, contribuind astfel la creșterea succesului lor în industrie.

6.1.2 Limitări și cercetări viitoare

În pofida insight-urilor valoroase obținute prin modelele de învățare automată, studiul nostru prezintă limitări care deschid calea pentru investigații viitoare. Una dintre limitările principale este dependența de acuratețea și exhaustivitatea seturilor de date; datele incomplete sau eronate pot afecta semnificativ performanța modelului. De asemenea, modelele pot avea tendința de a supra-antrena pe seturile de date specifice, ceea ce limitează capacitatea lor de generalizare pe noi date sau contexte diferite.

O altă limitare constă în utilizarea unui prag fix pentru convertirea probabilităților în predicții binare, ceea ce nu ia în considerare costul diferit al greșelilor de clasificare (de exemplu, costul unei rezervări anulate față de o rezervare nerealizată).

Pentru cercetările viitoare, ar fi benefică extinderea setului de date pentru a include mai multe variabile contextuale, cum ar fi factori economici, evenimente globale sau locale care pot influența comportamentul de rezervare. În plus, aplicarea tehnicilor de învățare automată mai avansate, cum ar fi rețelele neuronale profunde, ar putea oferi modele mai sofisticate și mai adaptabile.

În cele din urmă, este esențială efectuarea unor studii comparative între modele în scenarii de viață reală pentru a evalua mai bine eficacitatea lor practică. Aceasta ar include testarea modelelor în timp, în condiții de piață variabile, și ajustarea lor continuă pentru a răspunde la noile tendințe și comportamente ale clienților.

7 Concluzie

La finalul unei analize cuprinzătoare a dinamicii rezervărilor hoteliere, putem concluziona că utilizarea datelor complexe și a tehnicilor de învățare automată poate dezvălui insight-uri valoroase care sprijină hotelurile în abordarea provocărilor legate de anulările de rezervări. Scopul acestui studiu a fost de a identifica factorii care contribuie la anulările clienților și de a propune strategii bazate pe date pentru a minimiza impactul acestora asupra operațiunilor hoteliere.

Rezultatele studiului confirmă ipoteza inițială că există variabile cheie, cum ar fi durata de anticipare a rezervării și numărul de cereri speciale, care influențează semnificativ probabilitatea de anulare. Prin înțelegerea modului în care acești factori interacționează și afectează comportamentul de rezervare, hotelurile pot dezvolta oferte personalizate și politici de preț flexibile care să atragă și să păstreze clienți.

Analiza a evidențiat, de asemenea, utilitatea modelelor predictive în anticiparea anulărilor, oferind astfel hotelurilor oportunitatea de a gestiona mai eficient riscurile și de a optimiza ratele de ocupare. Aceste modele, de la algoritmi simpli de regresie logistică la metode avansate cum ar fi Gradient Boosting și LightGBM, reprezintă instrumente puternice pentru transformarea datelor brute în strategii operaționale și decizii de afaceri informate.

În concluzie, integrarea analizei de date în procesele de management hotelier nu este doar o necesitate, ci și o oportunitate strategică. Hotelurile încurajate să adopte o abordare bazată pe date, care nu doar că va reduce ratele de anulare, dar va îmbunătăți și satisfacția clienților, contribuind astfel la succesul pe termen lung într-o industrie extrem de competitivă. Această abordare țintește direct către scopul articolului de a dezvălui cum datele pot fi transformate în acțiuni concrete care să aducă valoare atât clienților, cât și hotelierilor.

References

- [BBM⁺14] Bart Baesens, Ravi Bapna, James R Marsden, Jan Vanthienen, and J Leon Zhao. Transformational issues of big data and analytics in networked business. *MIS quarterly*, 38(2):629–631, 2014.
- [KKKT23] Eun Joo Kim, Esther L. Kim, Minji Kim, and Sarah Tanford. Post-pandemic hotel cancellation policy: Situational cues as perceived risk triggers. *Journal of Hospitality and Tourism Management*, 55(3):153–160, 2023.
- [LCDL20] Minwoo Lee, Yanjun Maggie Cai, Agnes DeFranco, and Jongseo Lee. Exploring influential factors affecting guest satisfaction: Big data and business analytics in consumer-generated reviews. *Journal of Hospitality and Tourism Technology*, 11(1):137–153, 2020.
- [Mit20] Subrata Mitra. Estimating the duration of different seasons and their impact on hotel room prices. *International Journal of Hospitality Management*, 90:102604, 09 2020.

https://github.com/TataruDan/Data_Analysis_Project/tree/main