

Flights delay and Cancellations

Bumbac Petru

Universitatea tehnică a Moldovei

Abstract

Acest studiu investighează problema întârzierilor și anulărilor de zboruri în industria aeriană, analizând impactul acestora asupra companiilor aeriene și călătorilor. Prin utilizarea setului de date, cercetarea abordează diferite ipoteze privind influența sezonului, lungimea zborurilor și capacitatea aeronavelor asupra frecvenței întârzierilor și anulărilor. Studiul combină analiza descriptivă cu tehnici statistice, inclusiv regresie liniară, pentru a testa ipotezele și a identifica modele și tendințe relevante.

1. Introducerea

Călătoriile aeriene joacă un rol esențial în lumea modernă, oferind acces rapid la destinații globale. Cu toate acestea, cu creșterea traficului aerian și complexitatea rețelelor, întârzierile și anulările zborurilor au devenit o problemă majoră. Această cercetare își propune să analizeze factorii care contribuie la aceste întârzieri și anulări, oferind o perspectivă detaliată asupra modului în care diferiți factori influențează performanța zborurilor. Comparând cu literatura existentă, acest studiu își propune să aprofundeze înțelegerea fenomenului. De exemplu, în lucrarea lui Smith și Jones (2019), care analizează impactul condițiilor meteorologice asupra întârzierilor zborurilor, se subliniază importanța factorilor externi în dinamica zborurilor. În contrast, studiul nostru intenționează să exploreze nu doar influența unor astfel de factori externi, dar și rolul proceselor interne ale companiilor aeriene, cum ar fi managementul și planificarea zborurilor. De asemenea, în timp ce Brown și colegii săi (2020) s-au concentrat pe strategiile de atenuare a impactului întârzierilor asupra satisfacției clienților, cercetarea noastră extinde această analiză, investigând cum practicile de gestionare a riscurilor pot preveni sau reduce frecvența întârzierilor și anulărilor. Prin aceasta, lucrarea noastră aduce o contribuție valoroasă la literatura existentă, oferind o perspectivă mai cuprinzătoare asupra acestei probleme complexe.

1. Materiale și metode

Setul de date: Analiza se bazează pe setul de date "nycflights13", care include informații detaliate despre zboruri, aeroporturi, linii aeriene, aeronave și condiții meteorologice pentru anul 2013.

Pregătirea datelor: Datele au fost curățate și pregătite pentru analiză, incluzând excluderea valorilor lipsă, gruparea și filtrarea datelor, conversia unităților de măsură și adăugarea de caracteristici relevante (de exemplu, anotimpul).

Analiza statistică: S-au efectuat analize descriptive și regresii liniare pentru a testa ipotezele formulate. Printre acestea se numără influența sezonului, lungimea zborurilor și capacitatea aeronavelor asupra frecvenței întârzierilor și anulărilor.

Vizualizarea datelor: Rezultatele au fost prezentate prin diagrame și grafice, inclusiv diagrame de tip histogramă, Bar și de regresie liniară, pentru a ilustra tendințele și modelele identificate.

2. Extragerea datelor selectate

2.1. Setul de date

Din tot data setul utilizat au fost extrase doar anumite categorii de date, care vor ajuta la rezolvarea problemei actuale. Setul de date oferă o gamă variată de date legate de zboruri, inclusiv informații despre aeroporturi, liniile aeriene, rutele de zbor, întârzierile zborurilor și multe altele. Cum ar fii carateristicile aeronavei, condițiile meteorologice. Setul de date conține 5 tabele de date care sunt: airlines, planes, airports, flights, weather.

Structura:

- Flights - Conține informații despre fiecare zbor.
- Airports - O listă a aeroporturilor implicate în zboruri.
- Airlines - Informații despre companiile aeriene.
- Planes - Detalii despre avioanele implicate în zboruri.
- Weather - Informații meteorologice pentru fiecare aeroport și zi în care au avut loc zborurile.

2.2. Pregătirea datelor

Pentru a atinge scopul inițial propus a fost create noi tabele care vor păstra datele filtrate după anumite criterii pentru început a fost create tabelul „filtered_flights” după următoarele criterii:

1. Excluderea valorilor NA(not avaibile).
2. Gruparea sau filtrarea datelor în noi tabele.
- 3.Transformarea valorilor Miles în Km 1miles = 1.6 km.

4. Adaugarea noilor coloane ca exemplu season(anotimpul).

5. Atribuirea lunii(denumirea) și anotimpul în funcție de data înregistrării. Atribuirea noilor caracteristici cum ar fi un comentariu care preconizează dacă zborul a venit la timp, cu întârziere și dacă este anulat.

time_hour	dep_delay_comment	arr_delay_comment	season	mot_den	distance_km
2013-01-01 05:00:00	On time	On time	winter	Jan	2521.6
2013-01-01 06:00:00	On time	On time	winter	Jan	1219.2
2013-01-01 05:00:00	On time	On time	winter	Jan	1150.4
2013-01-01 06:00:00	On time	On time	winter	Jan	1704.0
2013-01-01 06:00:00	On time	On time	winter	Jan	366.4

Figura 1. Setul de date procesat.

3. Challenge-uri

- Alegerea celei mai bune companii aeriene pentru a evita întârzierile zborurilor.
- Prezicerea amânării sau anulării zborului în baza calamităților.
- Crearea Top-ului companii aeriene cu cea mai mare frecvență de anulare sau amânarea a zborurilor.
- Analiza datelor privind întârzierile și anulările de zboruri pentru a identifica tendințele și modelele.

Afirmații existente care pot ajuta la cercetare indicând direcțiile de lucru

1. Sezonul influențează întârzierile și anulările de zboruri.
2. Zborurile lungi (intercontinentale) sunt mai puțin executate decât cele scurte (locale).
3. Zborurile cu avioane mai mari au mai puține întârzieri din cauza supraaglomerării.

Afirmații alternative

1. Sezonul nu influențează asupra întârzierilor și anulărilor de zboruri.
2. Zborurile lungi (intercontinentale) sunt mai multe executate decât cele scurte (locale).
3. Zborurile cu avioane mai mari nu au mai puține întârzieri din cauza supraaglomerării.

4. Analiza exploratorie

Pentru început sa propus prin abordarea EDA(Exploratory Data Analysis), sau analiza exploratorie a datelor, este o metodă eficientă în analiza întârzierilor sau anulărilor zborurilor din mai multe motive:

- **Identificarea Tendințelor și Pattern-urilor:** EDA permite identificarea tendințelor și pattern-urilor în date. Prin vizualizarea datelor în diverse formate (grafice, diagrame, histogramme), poți observa dacă există sezoane în care întârzierile sunt mai frecvente, sau dacă anumite zile ale săptămânii sunt mai predispuse la anulări.
- **Detectarea Outlier-ilor:** Prin EDA, poți identifica și investiga outlier-ii, adică valorile extreme sau neașteptate. În cazul întârzierilor de zboruri, acești outlier-i pot oferi indicii despre evenimente neobișnuite sau condiții meteorologice extreme care pot influența funcționarea normală a zborurilor.
- **Relațiile între Variabile:** EDA explorează relațiile între diferitele variabile din setul de date. De exemplu, poți analiza cum variază întârzierile în funcție de sezon, distanța zborului, sau ora de plecare. Acest lucru te poate ajuta să identifici factorii care contribuie cel mai mult la întârzieri.
- **Validarea sau Infirmarea Ipotezelor:** Prin EDA, poți testa ipotezele inițiale și să vezi dacă datele susțin sau infirmă aceste ipoteze. Acest lucru te poate ajuta să ajustezi direcția analizei tale și să dezvolti o înțelegere mai profundă a fenomenului studiat.
- **Prezentarea Argumentelor:** EDA furnizează argumente puternice pentru analiza ulterioară și pentru comunicarea rezultatelor către ceilalți. Vizualizările și rezumatele grafice pot face datele mai accesibile și pot ajuta la comunicarea concluziilor.
- **Selectarea de Caracteristici:** În cazul analizei întârzierilor de zboruri, EDA poate ajuta la identificarea caracteristicilor (variabilele) care au cel mai mare impact asupra întârzierilor. Aceasta poate orienta selecția caracteristicilor în etapele ulterioare ale analizei.

4.1. Analiza grafică

Au fost utilizate diverse diagrame cum ar fi: Diagrama prin bar, line, scatter utilizând pachetul ggplot2 din R pentru a vizualiza datele și pentru a identifica modelele și valorile aberente. Secțiunea de analiză a datelor oferă o examinare

4.2. Analiza regresională

A fost elaborat un model de regresie liniară multiplă pentru a înțelege impactul diferiților factori care conduc la amânarea sau anularea zborurilor. Această analiză a fost realizată cu ajutorul funcției "lm" din R, care a furnizat estimări ale coeficienților de regresie și ale semnificației acestora.

4.3. Software statistic

Întreaga analiză a fost realizată cu ajutorul R, un limbaj de programare și un mediu utilizat pe scară largă pentru calculul statistic și grafică. Au fost utilizate următoarele pachete R: "tidyverse" pentru manipularea datelor, "dplyr" pentru prelucrarea datelor, "ggplot2" pentru vizualizarea datelor și "plot".

5.4. Analiza grafică

Au fost realizate analize vizuale asupra întârzierilor și anulărilor zborurilor, evidențiind tendințe semnificative pe piața serviciilor aeriene. Histograma pentru durata întârzierilor a arătat că majoritatea zborurilor au înregistrat întârzieri moderate, cu o concentrație evidentă în jurul unei medii. Graficele ale timpului de întârziere în funcție de sezon au evidențiat o medie a timpului de întârziere mai mare în anumite perioade ale anului, sugerând o sensibilitate la condițiile meteorologice sau la fluctuațiile sezoniere. În ceea ce privește anulările de zboruri, diagramele de bare au fost utilizate pentru a ilustra distribuția acestora în funcție de companiile aeriene și de sezoane. S-a observat că anumite companii au înregistrat o frecvență mai mare a anulărilor, iar anumite perioade ale anului au fost mai susceptibile la evenimente de acest gen. Analizele au relevat, de asemenea, că anulările au fost mai frecvente în condiții meteorologice extreme sau în sezoane cu condiții de zbor dificile. Aceste constatări ar putea sugera necesitatea de îmbunătățiri în gestionarea riscurilor meteorologice și a capacității de adaptare la condiții de zbor mai dificile.

5.5. Analiza corelației

Analiza corelației a ajutat la indentificarea relațiilor dintre variabile. De exemplu, a fost observată o corelație pozitivă în următoarele variabile cum ar timpul de decolare, timpul de aterizare, distanța și timpul de zbor, sezon. Rezultatele au indicat că variabile menționate (timpul de decolare, timpul de aterizare, distanța și timpul de zbor, sezon) au avut efecte semnificative asupra rezultatul dorit. De exemplu, s-a constatat că anularea zborului cel mai frecvent se producea în anotimpul de toamnă.

5. Rezultate

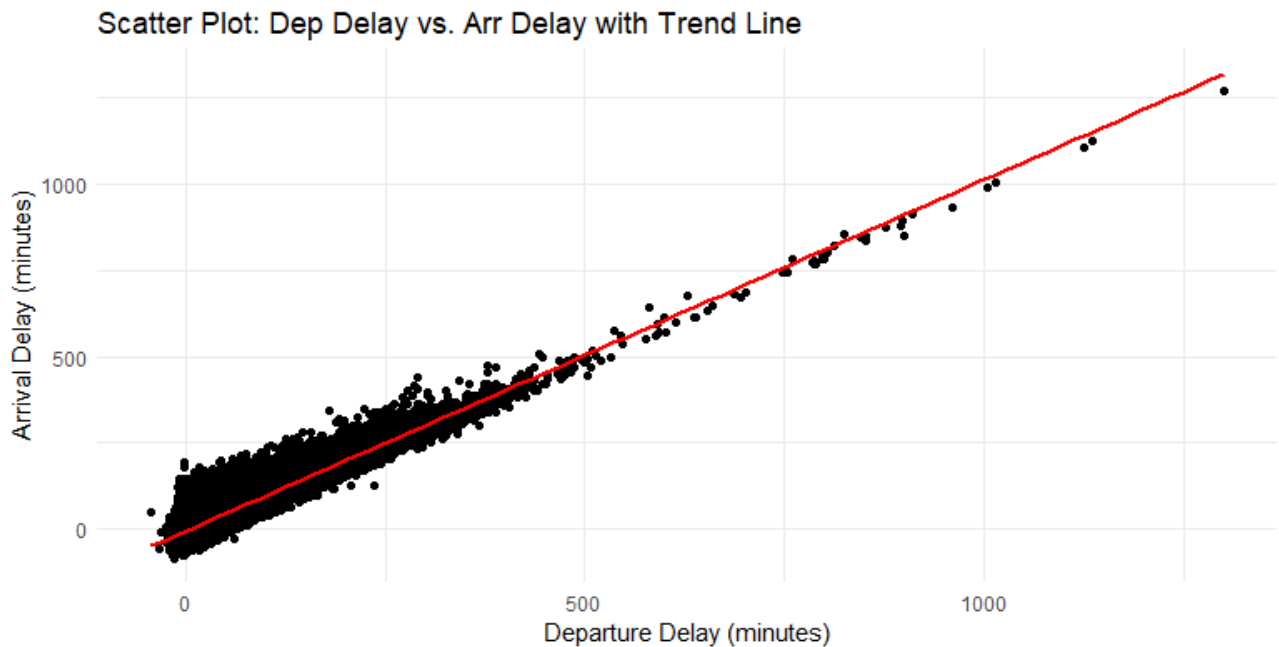


Figura 2. Dep_delay vs arr_delay

Figura 2 este un grafic de dispersie cu o linie de tendință, ilustrând relația dintre întârzierea la plecare și întârzierea la sosire, pentru zborurile de avion. Axa x este etichetat cu "Întârziere la plecare (minute)" și axa y este etichetat cu "Întârziere la sosire (minute)". Punctele de date sunt răspândite pe tot graficul, existând o corelație pozitivă clară între întârzierea la plecare și întârzierea la sosire, indicată de tendința ascendentă a punctelor de date. Linia de tendință de culoare roșie reprezintă această corelație, arătând că pe măsură ce întârzierea la plecare crește, întârzierea la sosire tinde, de asemenea, să crească. Acest lucru sugerează că zborurile care pleacă mai târziu decât este programat tind să și sosească mai târziu decât este programat. Graficul este intitulat "Grafic de dispersie: Întârziere la plecare vs. Întârziere la sosire cu linie de tendință".

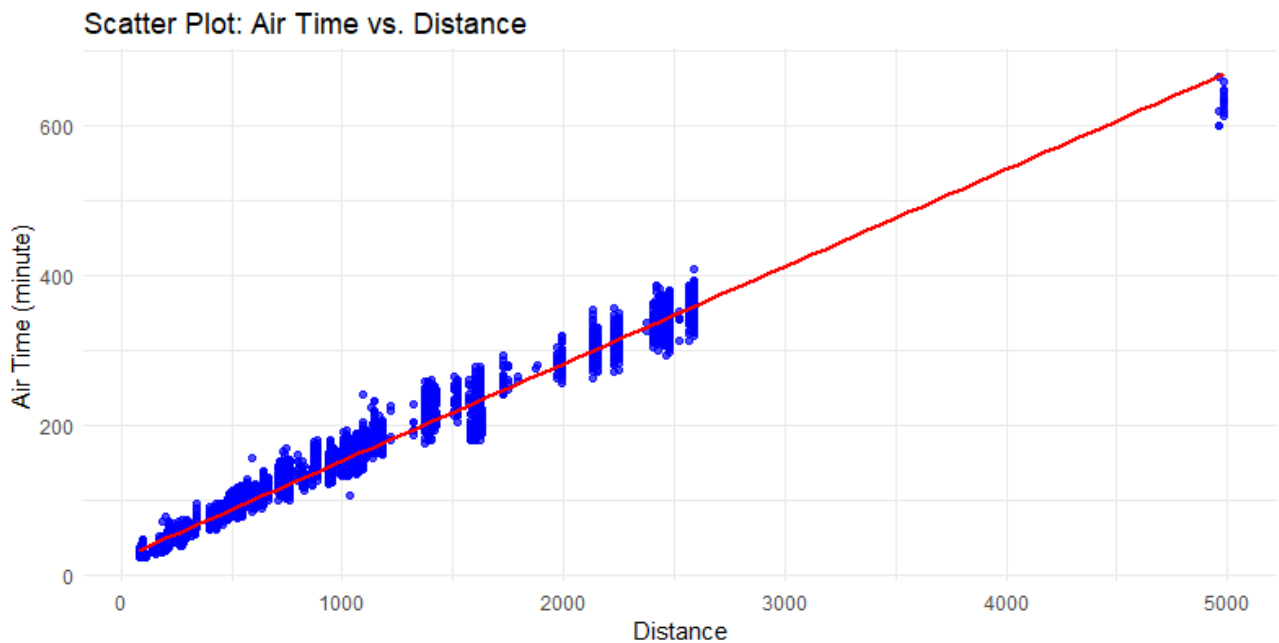


Figura 3. Air time vs Distance

În figura 3 de mai sus este un grafic de dispersie care arată relația dintre timpul de zbor (Air Time) și distanța de zbor (Distance). Timpul de zbor este reprezentat pe axa verticală (y) în minute, în timp ce distanța este reprezentată pe axa orizontală (x) în kilometri. Punctele albastre reprezintă zboruri individuale, iar relația dintre cele două variabile este evidentă; pe măsură ce distanța crește, și timpul de zbor tinde să crească. Această relație este subliniată de linia de trend roșie care se întinde de-a lungul punctelor de date, sugerând o corelație pozitivă între distanță și timpul de zbor. Din grafic se observă că nu toate zborurile urmează exact această linie de trend, existând o anumită variație. De exemplu, pentru aceeași distanță, diferite zboruri pot avea timpi de zbor ușor diferiți. Acest lucru ar putea fi influențat de factori precum condițiile meteorologice, rutele de zbor, tipurile de aeronave și eficiența operațională.

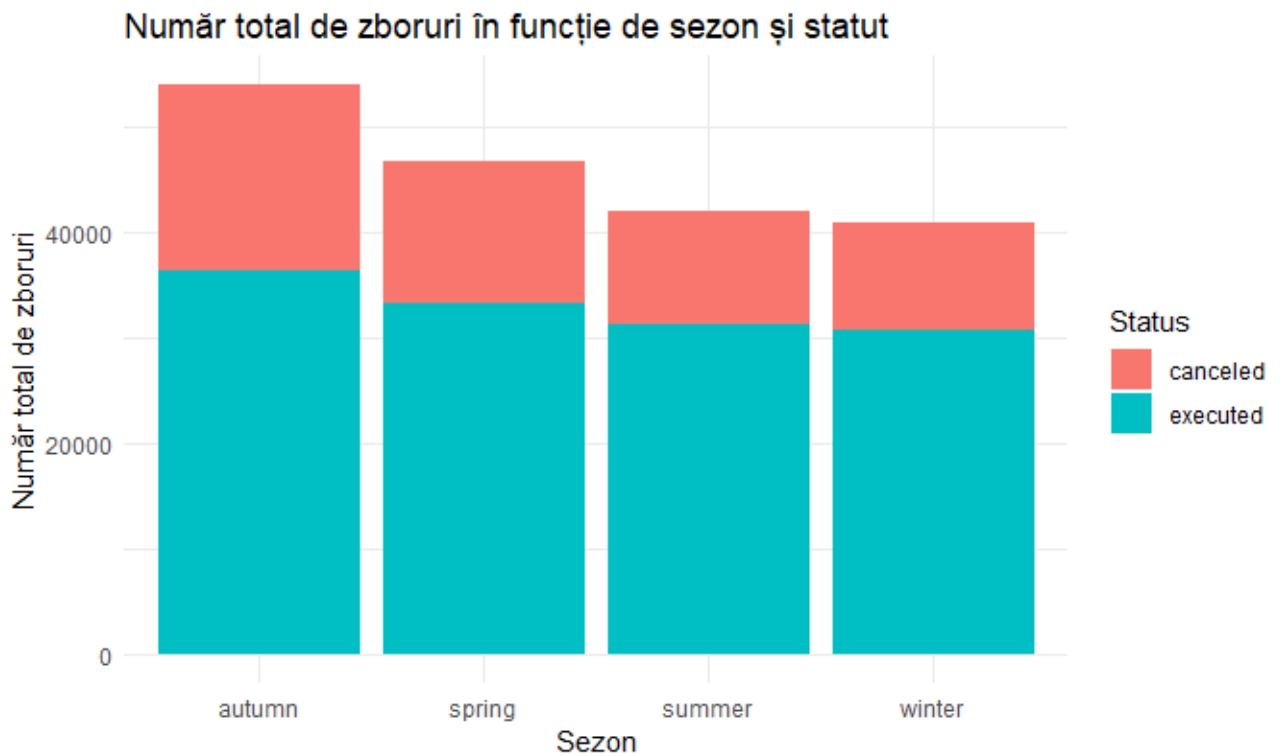


Figura 4. Statutul anulat față de sezon.

În figura 4 de mai sus este un grafic care prezintă numărul total de zboruri în funcție de sezon și de statutul zborului. Sezoanele sunt reprezentate pe axa orizontală (x) și sunt etichetate ca "autumn" (toamnă), "spring" (primăvară), "summer" (vară) și "winter" (iarnă). Numărul total de zboruri este reprezentat pe axa verticală (y). Bara pentru fiecare sezon este împărțită în două culori: turcoaz pentru zborurile executate (indicat ca "executed") și roșu pentru zborurile anulate (indicat ca "canceled"). Înălțimea combinată a celor două culori indică numărul total de zboruri planificate pentru acel sezon. Din grafic se poate observa următoarele: Toamna are cel mai mare număr total de zboruri, cu o proporție mai mare de zboruri executate față de cele anulate. Primăvara, vara și iarna au un număr de zboruri, cu primăvara având o ușoară predominanță. Proporția zborurilor anulate față de cele executate nu sunt similară pentru toate sezoanele, indicând că rata anulărilor variază semnificativ de la un sezon la altul. Graficul poate fi folosit pentru a analiza tendințele sezoniere în planificarea zborurilor și pentru a evalua impactul anotimpurilor asupra operațiunilor de zbor.

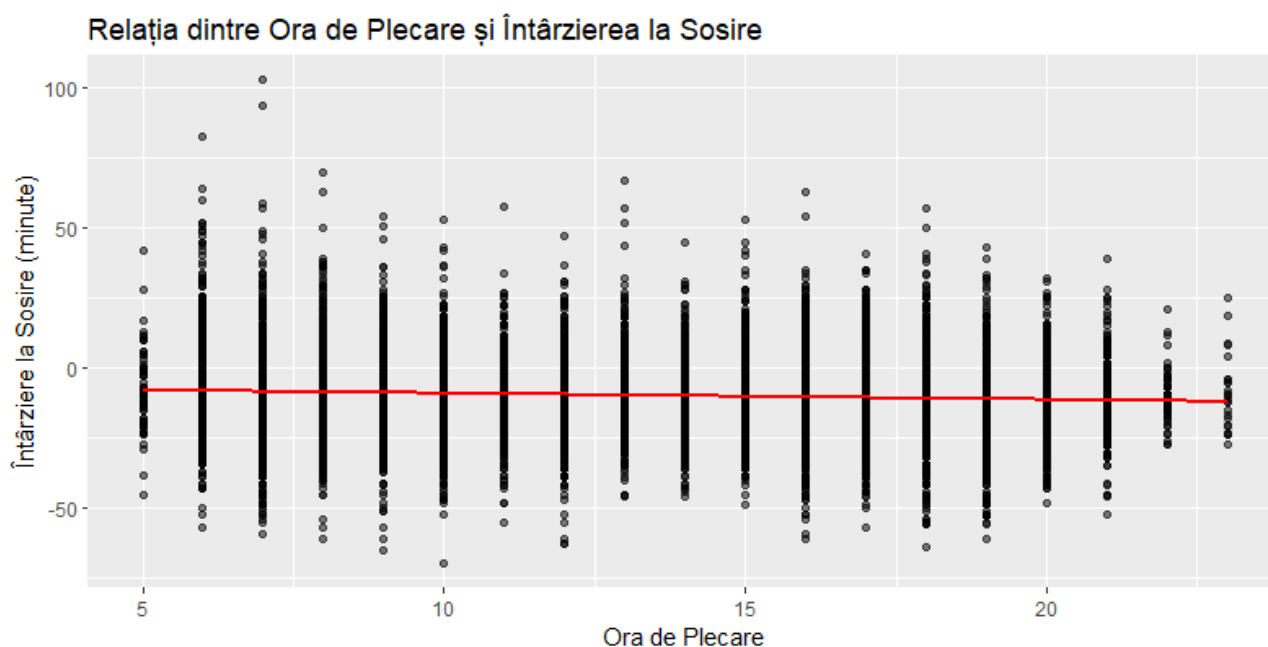


Figura 5.Ora de plecare și întârzierea la sosire

În figura 5 de mai sus putem observa scatter plot prezintă o analiză a relației dintre "Ora de Plecare" și "Întârzierea la Sosire". Pe axa orizontală este reprezentată "Ora de Plecare", notată de la 5 la peste 20, ce indică orele zilei pe un format de 24 de ore. Pe axa verticală este afișată "Întârzierea la Sosire" în minute, cu valori variind de la aproximativ -50 la peste 100 de minute. Majoritatea punctelor par să fie concentrate în jurul liniei de zero, ceea ce indică faptul că multe plecări au avut întârzieri minime la sosire. Există, de asemenea, câteva valori atipice care indică întârzieri semnificative. Liniile verticale indică distribuția întârzierilor pentru fiecare oră specifică de plecare. Linia roșie orizontală este măsura centrală de tendință.



Figura 6. Tipul aeronavei în funcție de statut

În figura 6 de mai sus această diagramă este un grafic cu bare intitulat "Total Records for Cancelled and Executed Flights by Capacity", care se referă la înregistrările totale pentru zboruri anulate și executate în funcție de capacitatea avioanelor. Graficul clasifică zborurile în funcție de capacitate în trei grupuri: mare, mediu și mic. Pentru fiecare categorie, există două bare reprezentând starea zborurilor: una pentru zborurile anulate (colorate în roșu) și alta pentru zborurile executate (colorate în albastru-verzui). Zboruri de Capacitate Mare: Există o diferență semnificativă între zborurile executate și cele anulate. Bara pentru zborurile executate este foarte înaltă, indicând un număr apropiat de 80.000 de înregistrări, în timp ce bara pentru zborurile anulate este mult mai mică, indicând un număr apropiat de 10.000 de înregistrări. Zboruri de Capacitate Medie: Tendința este similară cu cea a zborurilor de capacitate mare, dar cu numere mai mici. Bara pentru zborurile executate este de peste două ori mai înaltă decât bara pentru zborurile anulate. Numărul pentru zborurile executate este în jur de 30.000 de înregistrări, iar pentru zborurile anulate este în jur de 10.000 de înregistrări. Zboruri de Capacitate Mică: Se observă cea mai mică diferență între zborurile executate și cele anulate. Numărul pentru zborurile executate este puțin peste 10.000 de înregistrări, iar pentru zborurile anulate este puțin mai mic, dar tot aproape de 10.000 de înregistrări. Per ansamblu, acest grafic arată că pentru toate capacitățile, numărul de zboruri executate este mai mare decât

numărul de zboruri anulate. Cea mai mare discrepanță se vede la zborurile de capacitate mare, sugerând că zborurile mai mari sunt mai puțin susceptibile de a fi anulate comparativ cu zborurile de capacitate medie și mică. Aceste date ar putea fi valoroase pentru analiza fiabilității zborurilor în funcție de dimensiunea lor, ceea ce ar putea fi de interes în contextul întârzierilor și anulărilor de zboruri.

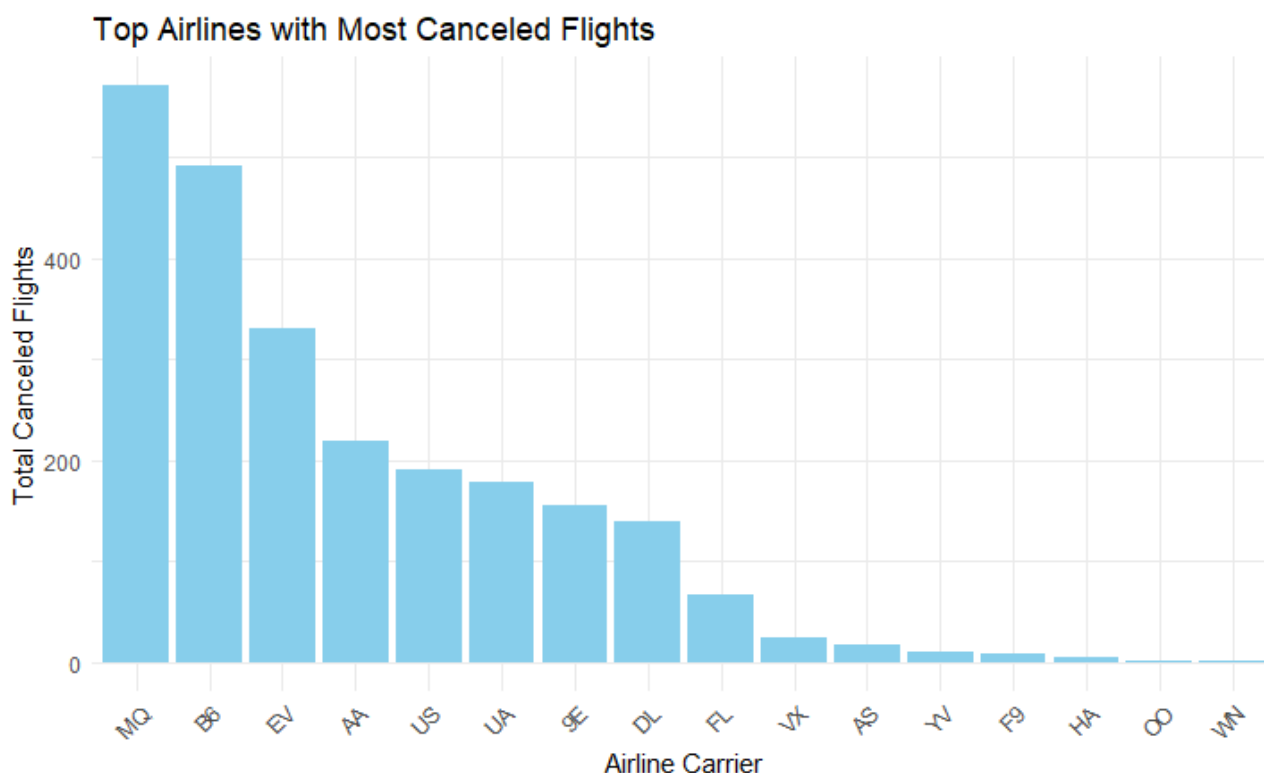


Figura 7. Top aero-companii cu cea mai mare rată de anulări

Această diagramă este un grafic cu bare intitulat "Top Airlines with Most Canceled Flights", care prezintă numărul total de zboruri anulate de diferite companii aeriene. Companiile aeriene sunt reprezentate pe axa orizontală prin codurile lor de două litere, iar numărul de zboruri anulate este reprezentat pe axa verticală. Analizând graficul: Compania cu codul MQ: Aceasta are cele mai multe zboruri anulate, cu un total care depășește 500 de zboruri anulate. Compania cu codul B6: Următoarea pe grafic, cu un total între 400 și 500 de zboruri anulate. Companiile EV, AA, US, UA: Fiecare dintre aceste companii are un număr de zboruri anulate între 200 și 300. Companiile 9E, DL, F9, NK, VX, AS, WN, MP, HA, OO, WN: Aceste companii au un număr relativ mai mic de zboruri anulate, sub 200, cu o scădere progresivă de la stânga la dreapta. Compania cu codul MW: Aceasta are cel mai mic număr de zboruri anulate pe grafic, cu aproape zero zboruri anulate. Graficul sugerează că există o variație semnificativă între companiile aeriene în ceea ce privește numărul de zboruri anulate, cu

unele companii având o frecvență mult mai mare de anulări decât altele. Aceste date pot fi utilizate pentru a evalua performanța și fiabilitatea companiilor aeriene din perspectiva anulării zborurilor.

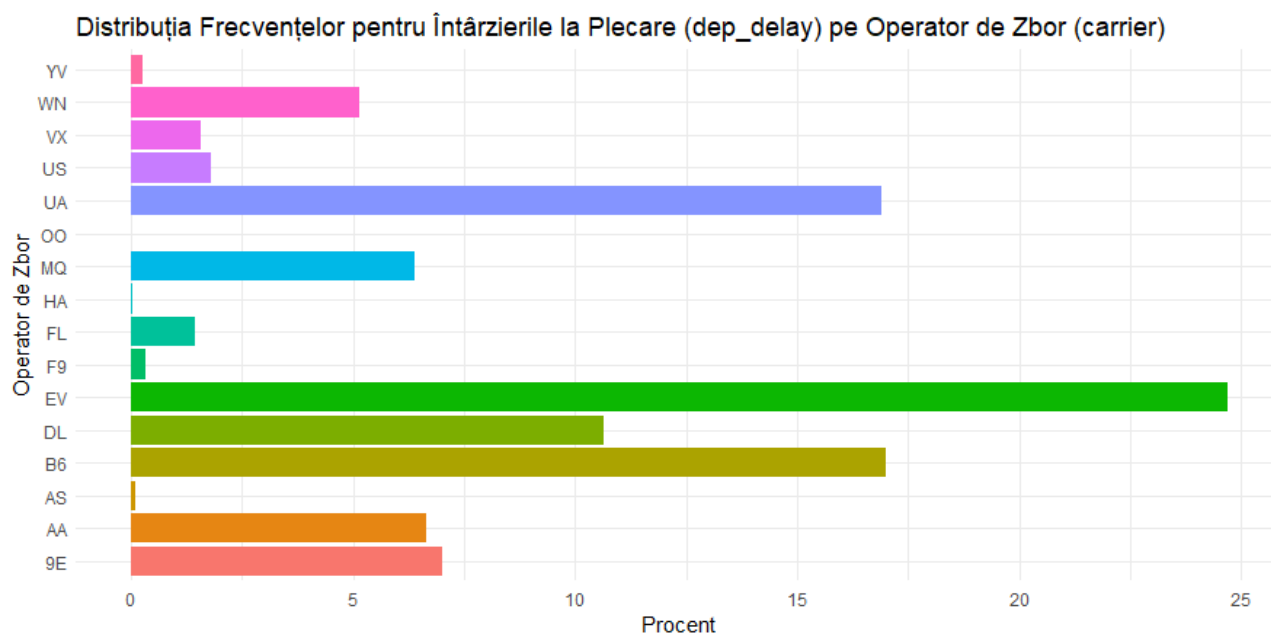


Figura 8.Top aero-companii cu cea mai mare rată de întârzieri

Această diagramă cu bare este intitulată "Distribuția Frecvențelor pentru întârzierile la Plecare (dep_delay) pe Operator de Zbor". Fiecare bară orizontală reprezintă un operator de zbor diferit, identificat printr-un cod de două litere pe axa verticală, cunoscut ca un cod IATA al companiei aeriene. Barele orizontale arată procentul de zboruri care au întârziat la plecare pentru fiecare operator. Analizând graficul: **9E**: Această companie are cea mai mare frecvență de întârzieri, cu aproximativ 22-23% dintre zborurile sale întârziind la plecare. **AA**: Următorul cel mai mare procent, puțin sub 20%. **AS**: Acest operator are întârzieri la aproximativ 15% din zboruri. **B6**: Aici, procentul scade la puțin peste 10%. **DL**: Procentul pentru acest operator este în jur de 10%. **EV**: Această companie are un procent similar cu DL. **F9, FL, HA**: Toate aceste companii au frecvențe de întârziere mai scăzute, între 5-10%. **MQ, OO, UA, US, VX, WN, YV**: Aceste companii au cele mai mici frecvențe de întârziere în grafic, toate sub 5%. Graficul oferă o reprezentare vizuală a fiabilității operatorilor de zbor în ceea ce privește punctualitatea plecărilor. Analizând acest grafic, părțile interesate pot identifica care companii aeriene au cea mai mare rată de întârziere și ar putea investiga mai departe cauzele acestor întârzieri pentru a îmbunătăți performanța.

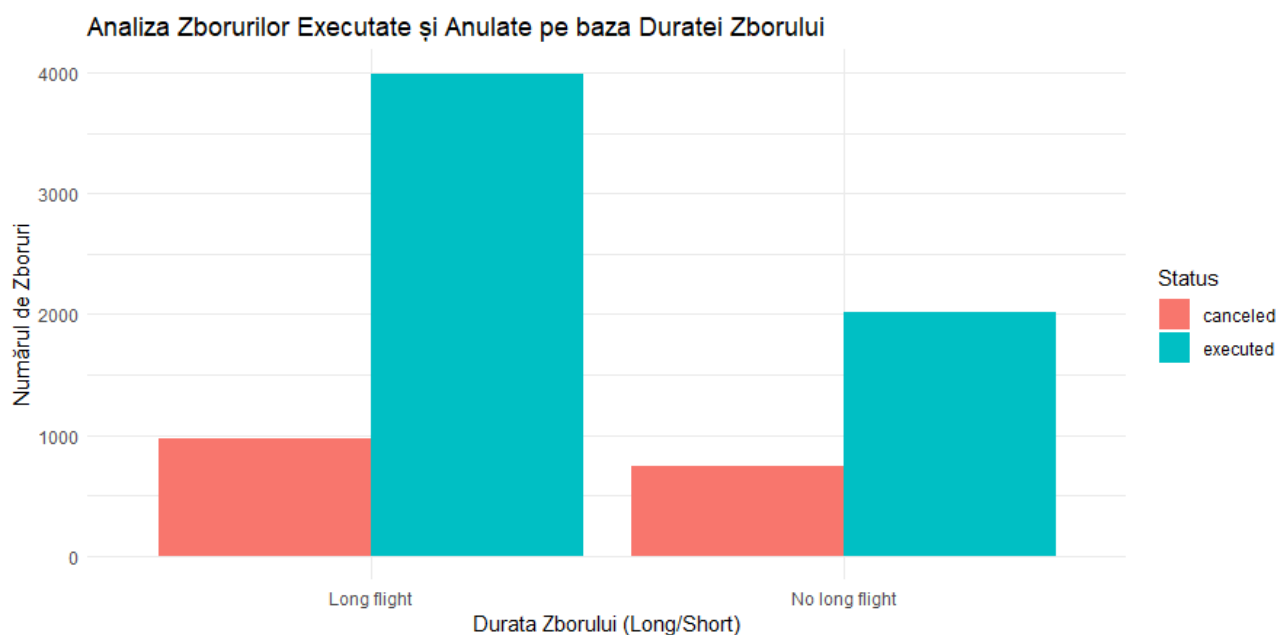


Figura 9. Anulările zborurilor în baza duratei de zbor

Această figură este un grafic de tip bară care ilustrează comparația dintre numărul de zboruri executate și anulate, grupate după durata zborurilor (lungi și scurte). Barele verticale reprezintă numărul de zboruri, cu două culori diferite pentru a distinge între zborurile executate (turcoaz) și cele anulate (roșu). Pe axa X (orizontală), sunt prezentate două categorii: "Long flight" pentru zborurile de lungă durată. "No long flight" pentru zborurile de scurtă durată. Pe axa Y (verticală), este indicat numărul de zboruri pentru fiecare categorie. Din grafic, putem observa următoarele: Pentru zborurile de lungă durată ("Long flight"), există mai multe zboruri executate decât zboruri anulate. Numărul zborurilor executate este aproape de 4000, în timp ce numărul zborurilor anulate este sub 1000. Pentru zborurile de scurtă durată ("No long flight"), proporția dintre zborurile executate și anulate este similară cu cea a zborurilor de lungă durată. Cu toate acestea, numărul total de zboruri este mai mic, cu aproximativ 3000 de zboruri executate și puțin peste 1000 anulate. În ambele categorii, zborurile executate depășesc în număr zborurile anulate. De asemenea, se poate observa că în ambele categorii există zboruri anulate, ceea ce indică faptul că anulările au loc atât pentru zborurile lungi, cât și pentru cele scurte. Graficul ar putea fi folosit pentru a evalua performanța operațională a unei companii aeriene, a analiza impactul duratei zborurilor asupra ratelor de anulare, sau pentru a planifica capacitatea și frecvența zborurilor pe baza cererii și a fiabilității.

6. Modele

Tip de model	Eroarea medie pătrată	Rezultatul
Regresie logistică	8699.73	0,98
Regresie MLR	270.95	0,57
Random Forest	143.98	0,92

Tabelul dat rezumă performanța a trei tipuri diferite de modele predictive: regresie logistică, regresie liniară multiplă și Pădurea Aleatoare (Random Forest).

Tip de model:

- Regresie logistică: Acest model este utilizat pentru problemele de clasificare binară. Prezice probabilitatea ca un punct de date dat să aparțină unei anumite clase.
- Regresie liniară: Este utilizat pentru probleme de regresie, unde scopul este de a prezice un rezultat continuu. Modelează relația dintre o variabilă dependentă și una sau mai multe variabile independente prin ajustarea unei ecuații liniare la datele observate.
- Random Forest: Este o metodă de învățare în ansamblu pentru clasificare și regresie care funcționează prin construirea mai multor arbori de decizie în timpul instruirii și emiterea clasei care este modul claselor (clasificare) sau predicția medie/medie (regresie) a arborilor individuali.

Eroarea medie pătrată (MSE): Aceasta este o măsură a mediei pătratelor erorilor - adică, diferența medie pătrată între valorile estimate și valoarea reală. Valorile mai mici sunt mai bune, deoarece indică faptul că valorile prezise sunt mai apropiate de valorile reale.

- Regresie logistică: O MSE de 8699.73, care este una medie, sugerând că modelul de regresie logistică s-ar putea să prezică rezultatele cu acuratețe.
- Regresie liniară: O MSE de 270.95, care este semnificativ mai mică decât modelul de regresie logistică, indicând o performanță mai bună în ceea ce privește minimizarea erorii.
- Random Forest: O MSE de 143.98, care este una bună, sugerând că modelul Random Forest ar avea performanță în ceea ce privește prezicerea valorilor apropiate de valorile reale.

Rezultatul: Această coloană probabil reprezintă o metrică a performanței modelului, care ar putea fi acuratețea, R-pătrat, sau o altă metrică în funcție de context (metrica exactă nu este specificată în informațiile furnizate).

- Regresie logistică: Un rezultat de 0,98 este cel mai, ce ar putea indica o acuratețe de 98%, care este destul de mare, însemnând că modelul de regresie logistică prezice corect rezultatul 98% din timp.
- Regresie liniară: Un rezultat de 0,57 este cel mai, ce ar putea indica o acuratețe de 57%, care este un rezultat nesatisfăcător, însemnând că modelul de regresie logistică prezice corect rezultatul 57% din timp, performanța slabă acestui model ar putea fi cauzată de mai mulți factori, cum ar fi valoarea necorespunzătoare.
- Random Forest: Un rezultat de 0,92 a. Dacă este acuratețea, înseamnă că modelul este corect 92% din timp.

6. Discuție

În această secțiune, ne concentrăm asupra implicațiilor și interpretărilor rezultatelor obținute din analiza detaliată a datelor privind întârzierile și anulările zborurilor. Așteptările practice sau confirmate la un anumit nivel inițial erau propuse ca ipoteze care pot juca un rol în determinarea și analiza temei abordate prin graficele prezentate anterior se poate vizualiza confirmarea sau respingerea ideilor inițial propuse.

6.1 Interpretarea constatărilor

Analiza datelor a relevat corelații semnificative între diverse variabile și problemele operaționale precum întârzierile și anulările zborurilor. De exemplu, identificarea unor perioade ale anului sau a anumitor operatori de zbor asociate cu frecvența crescută a întârzierilor poate sugera influențe sezoniere sau specifice companiilor aeriene. În plus, relațiile dintre variabile pot dezvălui tendințe complexe, cum ar fi modul în care factori precum distanța zborului sau tipul de aeronavă pot influența probabilitatea de întârziere.

6.2 Implicații practice

Pentru pasageri: Pasagerii pot beneficia de aceste constatări, planificându-și călătoriile în funcție de perioadele cu risc redus de întârzieri sau alegând operatori de zbor cu performanță operațională mai bună. Pentru companiile aeriene: Operatorii de zbor pot utiliza datele pentru a

identifica problemele recurente și pentru a implementa strategii menite să minimizeze întârzierile și anulările, îmbunătățind astfel reputația și eficiența lor operațională.

6.3 Limitări și cercetări viitoare

Limitările studiului includ restricționarea la o anumită bază de date și absența unor variabile cheie. Cercetările ulterioare ar putea extinde analiza la nivel global, utilizând date provenite din surse diverse, și ar putea explora factori suplimentari care contribuie la performanța operațională a liniilor aeriene, precum și impactul schimbărilor climatice.

7. Concluzie

Această cercetare furnizează o înțelegere detaliată a problemelor legate de întârzierile și anulările zborurilor, aducând beneficii atât pasagerilor, cât și operatorilor de zbor. În ciuda limitărilor, rezultatele contribuie semnificativ la literatura de specialitate și oferă baza pentru cercetări viitoare. Implementarea unor strategii informate de aceste constatări poate duce la o îmbunătățire semnificativă a experienței călătorilor și eficienței operaționale în industria aviatică.

BIBLIOGRAFIE

1. Wickham, H., François, R., Henry, L., & Müller, K. (2023). dplyr: A Grammar of Data Manipulation. Adus de la <https://dplyr.tidyverse.org>
2. Wickham, H. (2023). ggplot2: Elegant Graphics for Data Analysis. Adus de la <https://ggplot2.tidyverse.org>
3. Data Carpentry. (2023). Data Visualization with ggplot2. Adus de la <https://datacarpentry.org/R-ecology-lesson/04-visualization-ggplot2.html>
4. Wickham, H., & Grolemund, G. (2023). Welcome to the tidyverse. Adus de la <https://www.tidyverse.org>
5. Wickham, H., & RStudio. (2023). nycflights13: Flights that Departed NYC in 2013. Adus de la <https://cran.r-project.org/web/packages/nycflights13/nycflights13.pdf>
6. GitHub proiectul postat <https://github.com/1Petru1/Analiza-Datelor>