

Звіт
до лабораторної роботи №1
з дисципліни “Аналіз даних”
на тему “Розвідковий аналіз даних (EDA)”

Роботу виконали студенти:

Бабич Ірина (КМ-01)

Іваник Юрій (КМ-01)

Романецький Микита (КМ-01)

Суховій Ігор (КМ-01)

Шолоп Любомир (КМ-01)

Київ, 2023

Зміст

Зміст	2
Вступ	3
Основна частина	4
Частина 1. Походження набору даних	4
Частина 2. Очищення даних	8
1. Перевірка даних на наявність пропущених значень.	8
2. Типи даних	8
Частина 3. Характеристики датасету	9
Length.of.Stay	9
Total.Charges, Total.Costs	11
Частина 4. Результати EDA	13
Розподіл пацієнтів за віком, расою та статтю	13
Середня тривалість перебування в лікарні в залежності від віку, статі та раси	15
Ризик смертності в залежності від віку і статі	16
Ризик смертності в залежності від раси	16
Висновки	18
Додаток А. Коди CSS [2]	20
Додаток Б. Коди MDC [6]	27
Джерела	29

Вступ

Для виконання лабораторних робіт із дисципліни “Аналіз даних” командою №10 було обрано набір даних, що містить інформацію про пацієнтів лікарень штату Нью-Йорк за 2015 рік. Оскільки це дані, зібрані з реального світу[1], а не згенерована навчальна вибірка, вони дають змогу дослідити справжню проблему і, можливо, використати цей досвід у майбутньому.

Отже, основним і найзагальнішим питанням, яке наша команда хотіла б дослідити протягом цього курсу лабораторних робіт, є можливість передбачити довжину та вартість лікування пацієнта, а отже, витрачені на нього ресурси і прибуток, за першими даними, які зібрані з нього. Це дозволило б планувати фінанси лікарень заздалегідь на майбутнє, прогнозувати розподіли захворювань та забезпечувати оптимальною кількістю ресурсів усі заклади, щоб кожен пацієнт отримав своєчасну і якісну допомогу.

Для цього у цій лабораторній роботі ми хотіли б розібрати розподіл пацієнтів за віком, расою та статтю, оскільки вони є характеристиками, які відомі навіть до первинного огляду. Також, ми розглянемо залежності часу перебування та ризику смертності від згаданих вище ознак.

В результаті аналізу, на нашу думку, ми отримаємо приблизно такі результати:

1. Серед пацієнтів буде більше людей старшого віку, а розподіли за статтю та расою мають відповідати таким самим посеред населення штату Нью-Йорк станом на 2015 рік.
2. Тривалість перебування в лікарні буде в середньому меншою у дітей та повільно зростатиме до 40 років, оскільки молодь хворіє менше і менша ймовірність ускладнень, а потім зупиниться на стабільному значенні. Також вона не залежатиме від статі та раси.
3. Ризик смертності, на нашу думку, зростатиме з віком і, скоріш за все, не залежатиме від статі.

Основна частина

Частина 1. Походження набору даних

Набір даних, який було використано для написання лабораторної роботи, носить назву “2015 de-identified NY inpatient discharge (SPARCS)”, тобто, він містить записи про випадки надання лікарських послуг пацієнтам у штаті Нью-Йорк за 2015 рік. Датасет було надано у відкритий доступ Департаментом здоров’я штату Нью-Йорк на його офіційному сайті у 2017 році [1]. Програма збору даних належить Відділу інформації та статистики Офісу якості і безпеки пацієнтів Департаменту здоров’я штату Нью-Йорк.

Інформація про здоров’я, що міститься у файлі, не є індивідуально розпізнаваною. Усі елементи даних, які могли сприяти визначенню особи пацієнта, були із нього видалені. Також, з цього файлу було виключено вторинні діагнози та процедури, а також коди оплати послуг. Один рядок відповідає інформації про рівно 1 клінічний випадок.

Записи про звернення пацієнток, основною причиною яких було проведення абортів, мають відредаговані дані про місце та заклад проведення процедури, а саме: назву закладу, область медичного обслуговування, округ лікарні, сертифікат експлуатації та всі номери ліцензій постачальника послуг. Поштовий індекс пацієнта обмежений першими трьома цифрами, однак у випадках, якщо розмір населення для поштового індексу менший за 20 000 або якщо наданою медичною послугою є аборт, поштовий індекс залишається порожнім. Код «OOS» означає поштові індекси за межами штату. Вік пацієнтів представлений у таких вікових групах: від 0 до 17 років, від 18 до 29 років, від 30 до 49 років, від 50 до 69 років і від 70 років і старше.

Датасет перед очисткою мав 2346760 записів та 37 змінних. Назва кожної змінної, опис її змісту та перелік можливих значень наведено в таблиці нижче. Переліки можливих значень колонок, що стосуються діагнозу за класифікацією CSS та MDC наведено в додатку А та додатку Б відповідно.

"Health.Service.Area"	Область знаходження госпіталю (складається з кількох округів). Може мати такі значення: Capital/Adirondack, Central NY, Finger Lakes, Hudson Valley, Long Island, New York City, Southern Tier, Western NY.
-----------------------	---

"Hospital.County"	Округ знаходження госпіталю. Значення відсутнє для випадків, що стосуються абортів.
"Operating.Certificate.Number"	Номер ліцензії закладу. Значення відсутнє для випадків, що стосуються абортів.
"Facility.Id"	Унікальний ідентифікатор закладу. Значення відсутнє для випадків, що стосуються абортів.
"Facility.Name"	Назва закладу. Значення замінене на "Abortion Record – Facility Name Redacted" для випадків, що стосуються абортів.
"Age.Group"	Вікова група.
"Zip.Code...3.digits"	Поштовий індекс. Код «OOS» означає поштові індекси за межами штату.
"Gender"	Стать пацієнта.
"Race"	Race. Black/African American, Multi, Other Race, Unknown, White. Інші раси (Other Race) включають в себе корінних американців та жителів азійських островів чи островів Тихого океану.
"Ethnicity"	Етнос. Має значення Spanish/Hispanic Origin, Not of Spanish/Hispanic Origin, Multi, Unknown. Spanish/Hispanic origin позначає людину із латиноамериканським походженням.
"Length.of.Stay"	Час перебування у лікарні.

"Type.of.Admission"	Опис того, як пацієнт потрапив у заклад медичної допомоги: Elective, Emergency, Newborn, Not Available, Trauma, Urgent. (за власним зверненням, надзвичайна ситуація, новонароджений, дані відсутні, травма, терміновий).
"Patient.Disposition"	Розміщення пацієнта (домашнє лікування, госпіталізація на довгий/короткий період часу, вирок суду, домашнє лікування з медичним наглядом і т.д.)
"Discharge.Year"	Рік випадку звернення (постійний і дорівнює 2015).
"CCS.Diagnosis.Code"	Код узагальненого діагнозу для статистичних досліджень [3][6].
"CCS.Diagnosis.Description"	Назва відповідної коду групи діагнозів.
"CCS.Procedure.Code"	Код процедури.
"CCS.Procedure.Description"	Назва відповідної коду процедури.
"APR.DRG.Code"	Код класифікації за витратами ресурсів на пацієнта та групою діагноза для оплати системою державного страхування Medicare (або іншою) [2].
"APR.DRG.Description"	Назва відповідної коду групи.
"APR.MDC.Code"	Код класифікації стану пацієнта за ураженою системою організму або для випадку, коли неможливо визначити систему – за природою ураження. Є надгрупами для кодів APR DRG (хвороби респіраторної

	системи, шлунково-кишечного тракту, імунної та кровотворної системи, опіки, вагітність, множинні травми, онкології і т. д.) [4]
"APR.MDC.Description"	Відповідна коду назва категорії.
"APR.Severity.of.Illness.Code"	Код, що відповідає важкості захворювання (4 рівні), присвоюється на основі головного і побічних діагнозів.
"APR.Severity.of.Illness.Description"	Відповідна коду характеристика словом.
"APR.Risk.of.Mortality"	Ризик смертності (4 рівні).
"APR.Medical.Surgical.Description"	Змінна індикатор того, належить код DRG і, відповідно, випадок до хірургічного чи терапевтичного класу.
"Payment.Typology.1"	Тип оплати, 1 (назва страхової компанії чи інше).
"Payment.Typology.2"	Тип оплати, 2.
"Payment.Typology.3"	Тип оплати, 3.
"Attending.Provider.License.Number"	Номер ліцензії лікаря.
"Operating.Provider.License.Number"	Номер ліцензії лікаря хірургічної спеціальності.
"Other.Provider.License.Number"	Номер ліцензії іншого лікаря чи медсестри.
"Birth.Weight"	Вага дитини при народженні (для випадків пологів).
"Abortion.Edit.Indicator"	Індикатор ознак абортів.
"Emergency.Department.Indicator"	Індикатор проведення процедур у відділення екстреної допомоги.

"Total.Charges"	Оцінена вартість перебування у лікарні, яку просять заплатити спочатку, до перемовин про знижки. Часто сплачується пацієнтом без страховки повністю. Призначається лікарнею.
"Total.Costs"	Собівартість перебування в лікарні, включно з матеріалами, препаратами і зарплатнею співробітників.

Частина 2. Очищення даних

Датасет було взято з платформи kaggle, він мав 2346760 записів та 37 змінних. За попереднім переглядом, дані не були в зручному для опрацювання форматі, тож довелося проводити очищення.

1. Перевірка даних на наявність пропущених значень.

Деякі відсутні значення в цьому датасеті могли бути визначені автоматично, тому спершу було обраховано кількість нульових значень у кожній з колонок. У 4 із них кількості NA наближалися до $\frac{1}{3}$ і більше від загальної кількості, а саме:

Назва колонки	Кількість NA
Other.Provider.License.Number	2275424
Payment.Typology.2	762346
Payment.Typology.3	1645570
Operating.Provider.License.Number	612848

Оскільки ці колонки містять дані про номер ліцензії другого лікаря (якщо він був), який був відповідальним за надання послуги, та дані про джерело оплати у випадках, якщо воно не одне, ми вирішили, що ці колонки можна видалити без шкоди для дослідження.

Також, в деяких рядках певні змінні мали значення NA, але таких записів було небагато відносно загальної кількості, 4578, тож ми прийняли рішення їх видалити.

До всього вищесказаного, специфічне кодування відсутності даних було помічено у колонці статі: там воно позначалось як "U" (unknown). Таких записів було всього 39, тож їх також було видалено.

2. Типи даних

Деякі колонки датасету були розпізнані R як рядкові дані через їх формат запису, але насправді були числовими. До таких колонок належали `Length.of.Stay`, `Total.Charges` та `Total.Costs`.

`Total.Charges` та `Total.Costs` було просто перетворено на чисельний тип.

`Length.of.Stay` містила в собі значення “120+”, яке позначало перебування в лікарні більше за 120 діб, яке було несумісним із чисельним типом даних. Через це ми прийняли рішення позначати таку категорію термінів перебування числом 121. Після цього її тип було змінено на числовий.

Частина 3. Характеристики датасету

Після очищення було отримано датасет із такими характеристиками:

1. Кількість записів: 2342182
2. Кількість колонок: 33

Кожна колонка датасету відповідає одній змінній. Щодо пропущених значень, велика їх кількість знаходиться в колонці із *Birth.Weight*, оскільки вона за змістом заповнюється виключно у випадках пологів.

Датасет має певну кількість колонок із числовими значеннями: *Operating.Certificate.Number*, *Facility.Id*, *Length.of.Stay*, *Discharge.Year*, *CCS.Diagnosis.Code*, *CCS.Procedure.Code*, *APR.DRG.Code*, *APR.MDC.Code*, *APR.Severity.of.Illness.Code*, *Attending.Provider.License.Number*, *Birth.Weight*, *Total.Charges*, *Total.Costs*. Але якщо звернути увагу на їх зміст, то стає зрозуміло, що по своїй суті вони є не неперервними впорядкованими значеннями, а категорійними, тож обрахунок дескриптивних статистик для більшості з них не має жодного сенсу.

З них справді числовими і впорядкованими є: *Length.of.Stay*, *Birth.Weight*, *Total.Charges*, *Total.Costs*. Обчислимо для них основні статистики:

	Min	1st Q	Median	Mean	3rd Q	Max
<i>Length.of.Stay</i>	1	2	3	5.481	6	121
<i>Birth.Weight</i>	1000	2900	3300	3260	3600	9900
<i>Total.Charges</i>	0	12034	23501	43231	46645	7248391
<i>Total.Costs</i>	0	4726	8797	15992	16845	5236615

Розглянемо детальніше ті змінні, які нас цікавлять в цій лабораторній: *Length.of.Stay*, *Total.Charges*, *Total.Costs*.

Length.of.Stay

Ця змінна, як вже було сказано, відповідає за довжину перебування і не має жодних NA значень. Дескриптивні статистики вказано в таблиці вище.

Було побудовано також графік її розподілу, але оскільки у різних груп надзвичайно різняться кількості значень, що в них потрапляють:

- від 1 до 5 днів: 1711460 випадків

- від 115 до 120 днів: 160 випадків

шкала кількості випадків є логаритмічною для спрощення візуалізації закономірностей.

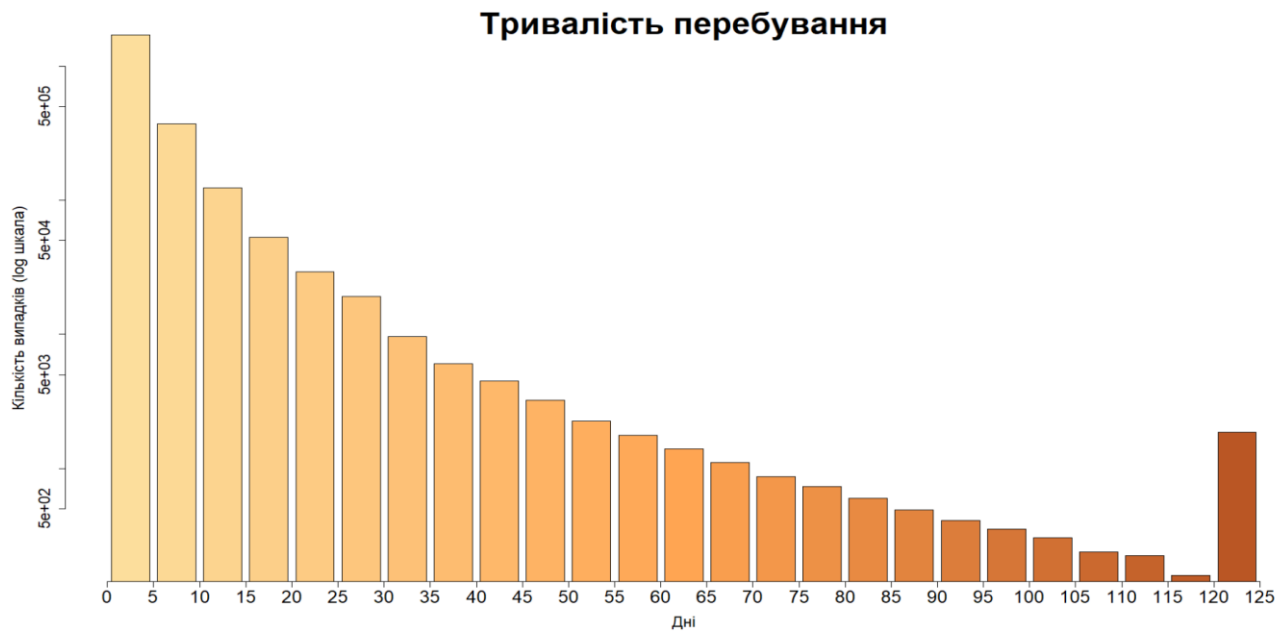


рис. 1 - Тривалість перебування за log шкалою кількості

Із графіку можна побачити очевидну закономірність розподілу - чим менша тривалість перебування, тим більше випадків, що їй відповідають. Виняток з цього правила - група з 120 та 121, але це вписується в загальну картину, оскільки за 121 позначені всі випадки перебування в лікарні, що довші за 120 днів.

Подібна скошеність розподілу наводила на думку про логаритмічно нормальний розподіл, але оскільки він має ледь помітний, але все ж спад на малих значеннях випадкової величини, природа даних у цьому випадку вказує на те, що це малоймовірно. І справді, якщо прологарифмувати цю випадкову величину, графік, що вийде, не суперечить даним міркуванням - отриманий розподіл є лише сильно скошеним ліворуч.

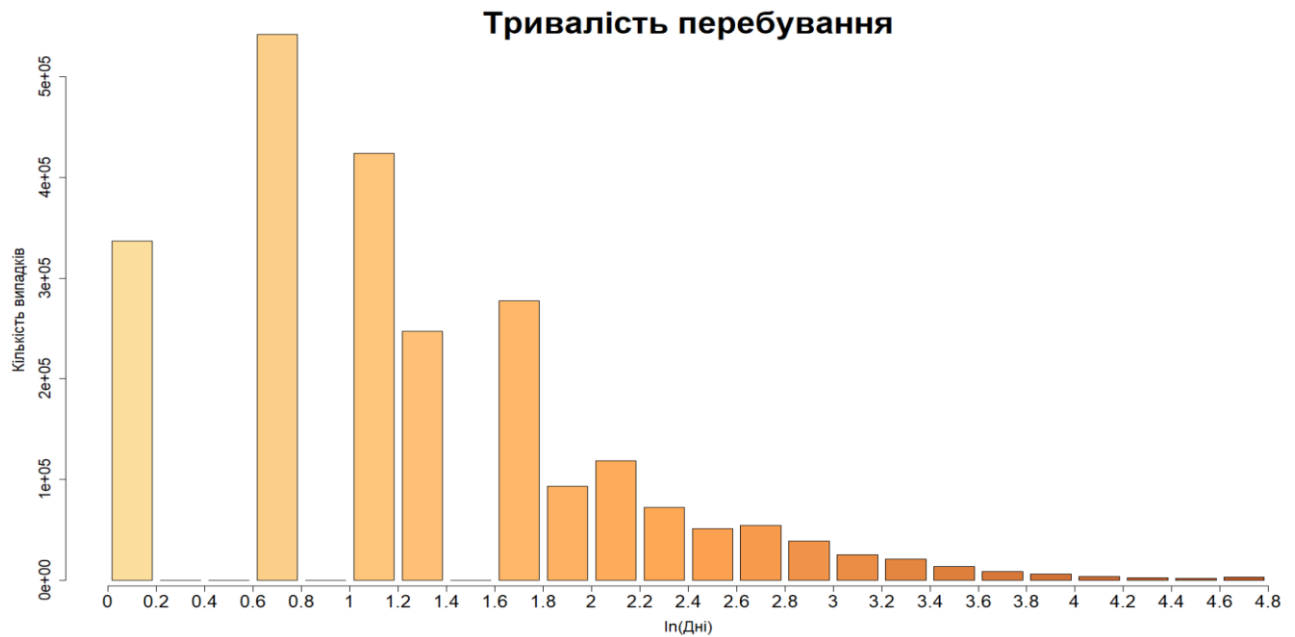


рис. 2 - Графік логарифму тривалості перебування

Total.Charges, Total.Costs

Оскільки параметри ціни та собівартості самі по собі мали надзвичайно скошений праворуч розподіл, до того ж, надзвичайно великими були різниці між кількостями елементів різних груп, було вирішено використати *логаритмічну шкалу для вісі кількості*, аби було зручніше прослідковувати загальні закономірності.

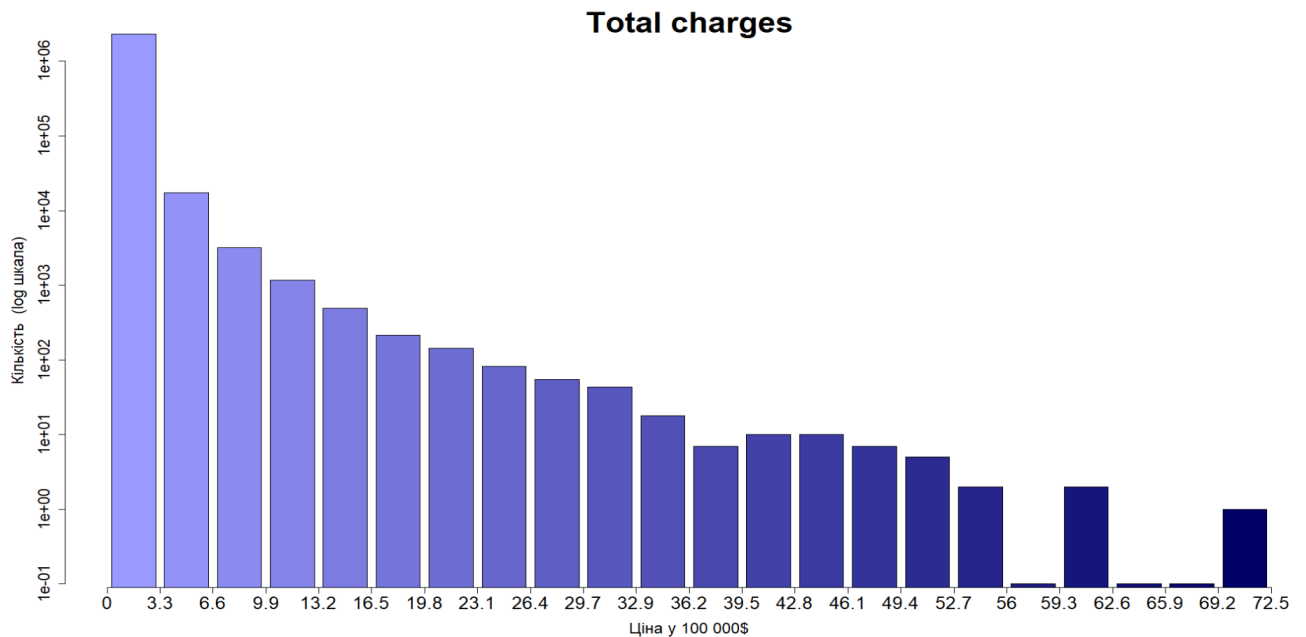


рис. 3 - Графік загальної ціни з log шкалою кількості випадків

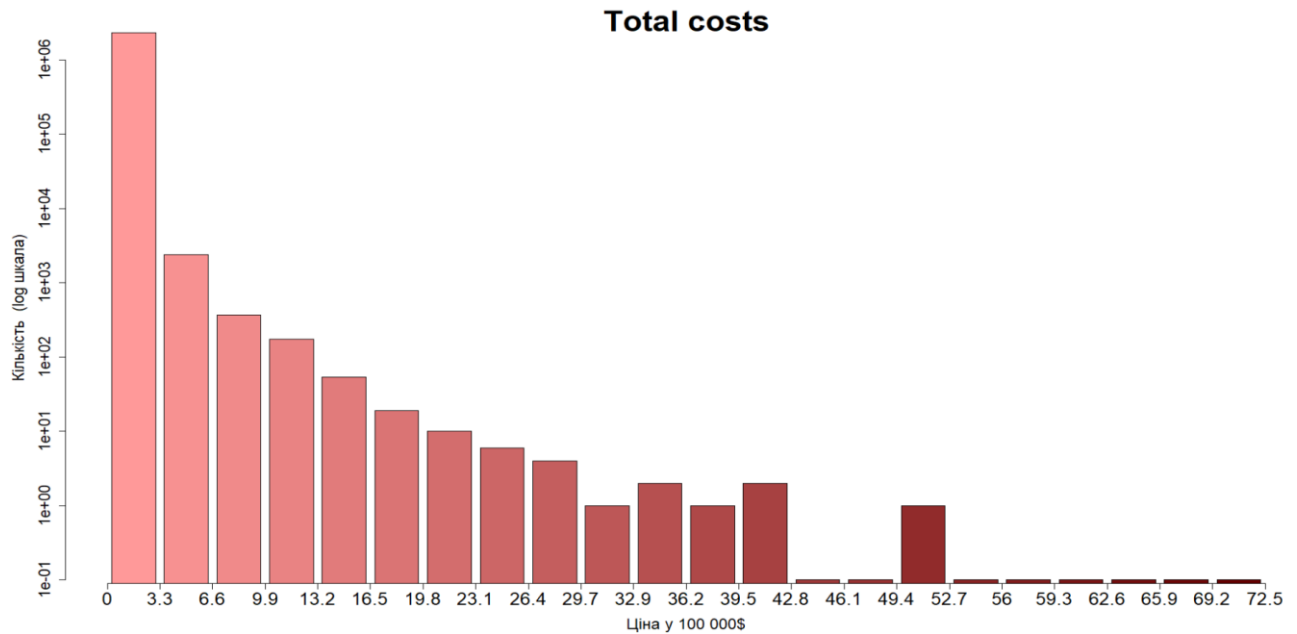


рис. 4 - Графік загальної собівартості з *log* шкалою кількості випадків

Отже, із графіків видно, що загалом ціна вища за собівартість, що обумовлюється природою даних - лікарням необхідний певний прибуток, аби функціонувати в наданій системі охорони здоров'я. Також можна зазначити, що меншим цінам відповідають більші кількості випадків, що теж зрозуміло. Великі значення, які схожі на викиди, виправдовуються типом послуг - здебільшого вони відповідають випадкам, пов'язаним з перинатальним періодом (ймовірно, через специфічність і вартість обладнання для догляду за дітьми з 28-го тижня вагітності до 7-ї доби життя), або захворюваннями серцево-судинної системи.

Сильна скошеність розподілу наводить на думку про логарифмічно нормальний розподіл, так само, як і у випадку з тривалістю перебування. Але цього разу гіпотеза не суперечить отриманим результатам - якщо прологарифмувати значення цін і побудувати гістограми, їх вигляд надзвичайно нагадує нормальний розподіл. Ця гіпотеза потребує подальших перевірок і статистичних підтверджень.

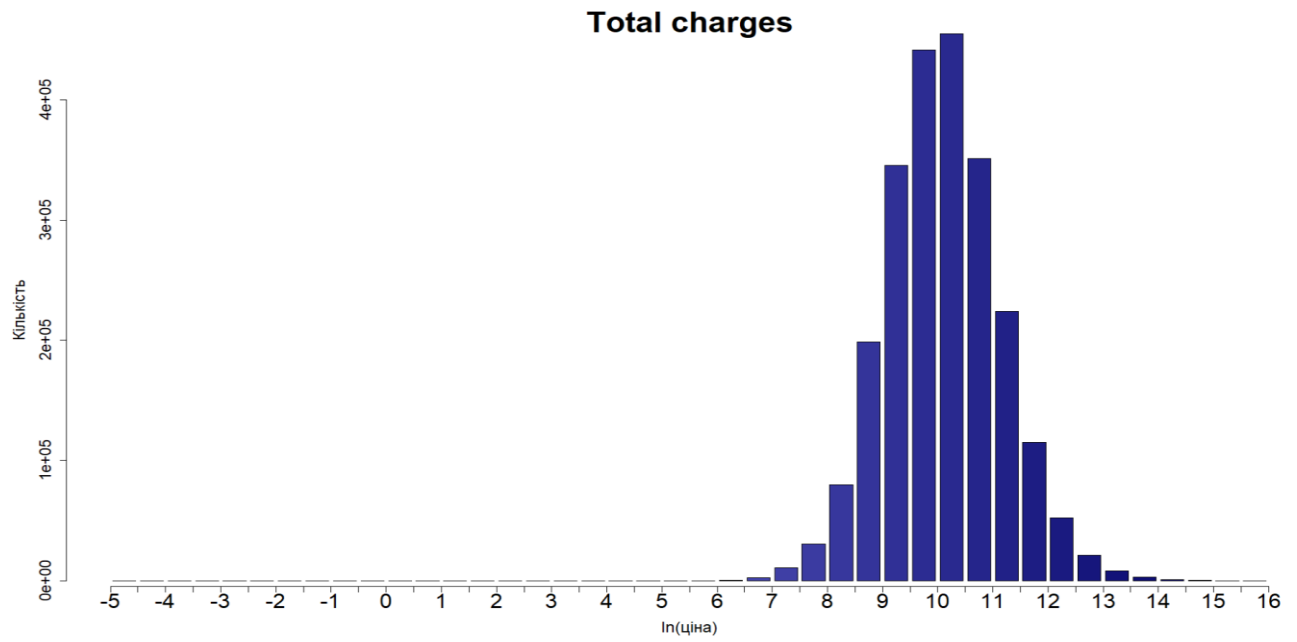


рис. 5 - Графік логарифму ціни

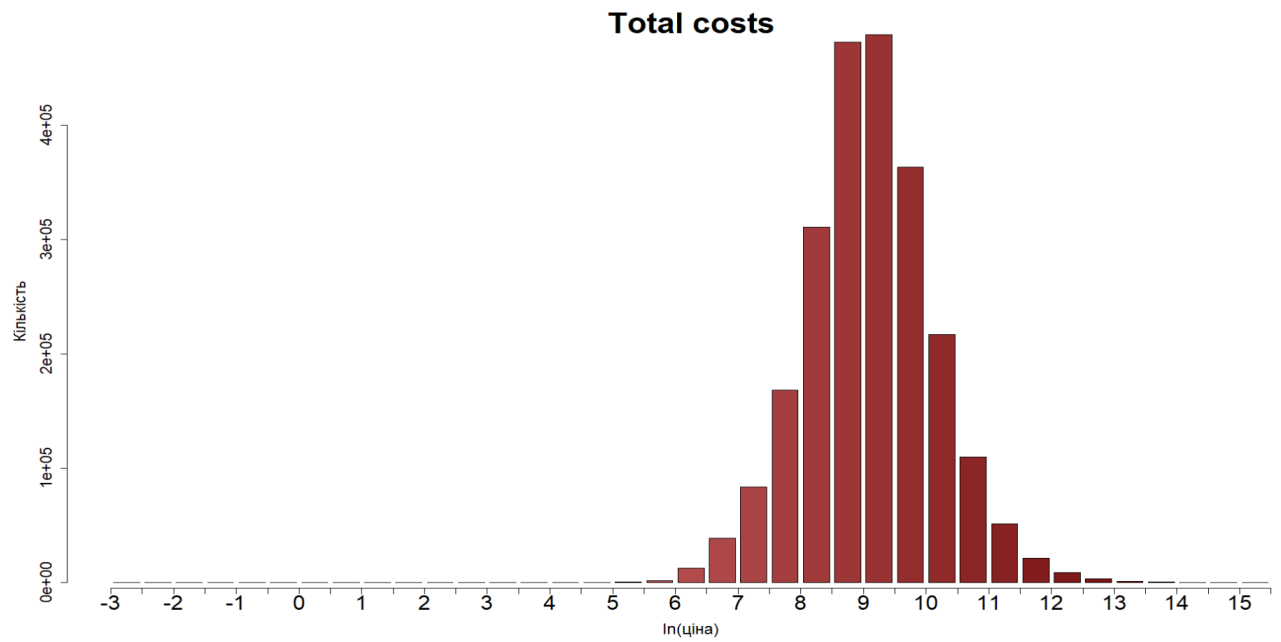


рис. 6 - Графік логарифму собівартості

Частина 4. Результати EDA

Розподіл пацієнтів за віком, расою та статтю



рис. 7 - Розподіл за віком, расою та статтю

Результати дослідження наведено на графіках, зображених на рис. 7. Розглянемо детальніше кожен із графіків.

Почнемо з розподілу за статтю. В таблиці нижче наведено дані з графіка та, для порівняння, інформацію, отриману з відкритих джерел [7] за 2015 рік щодо розподілу населення штату Нью-Йорк:

	Дані з датасету	Дані з відкритих джерел	Різниця
Чоловіки	44.343%	48.553%	-4.21%
Жінки	55.657%	51.447%	4.21%

За цими даними можемо зробити попередній висновок, що розбиття пацієнтів за статтю трохи відрізняється від відповідного розбиття серед населення, але різниця невелика, отже, вона може виявитись статистично незначущою за подальшого дослідження.

За віком пацієнти розподілені таким чином за інформацією з датасету, і додатково для порівняння наведено розподіл населення штату Нью-Йорк за віком у 2015 році[7]:

Вік	Дані з датасету	Дані з відкритих джерел	Різниця

0-17 років	15.03%	21.271%	-6.241%
18-29 років	10.495%	17.376%	-6.881%
30-49 років	19.488%	26.336%	-6.848%
50-69 років	27.529%	24.955%	2.574%
70+ років	27.458%	10.062%	17.396%

Із цієї порівняльної характеристики можемо зробити припущення, що люди, старші за 50 років, потрапляють в лікарню частіше за інших, оскільки їх відсоток серед пацієнтів переважає їх відсоток серед населення. Особливо серед них вирізняються пацієнти 70+, оскільки в них відсоток різниці між кількістю населення і кількістю пацієнтів найбільший. Серед молоді і людей середніх років різниця даними з датасету та даними про населення приблизно однакова, з чого можна сказати, що по суті вони потрапляють в лікарні приблизно з однаковою частотою.

За расою розподіл має такий вигляд:

	Дані з датасету	Дані з відкритих джерел [8]	Різниця
Білі	56.914%	70.3%	-13.386%
Чорні/афроамериканці	18.949%	17.6%	1.349%
Мультирасові	0.95%	2.4%	-1.45%
Інша раса (азіати, корінні американці, жителі островів Тихого океану)	23.187%	9.7%	13.487%

Із цих даних можемо зробити орієнтовні висновки, що найменше відносно їх частки в популяції в лікарню потрапляють білі пацієнти, а найбільше - представники інших рас, тобто азіати, корінні американці та жителі Тихоокеанських островів. Пояснення, чому так відбувається, потребує подальшого дослідження.

Середня тривалість перебування в лікарні в залежності від віку, статі та раси

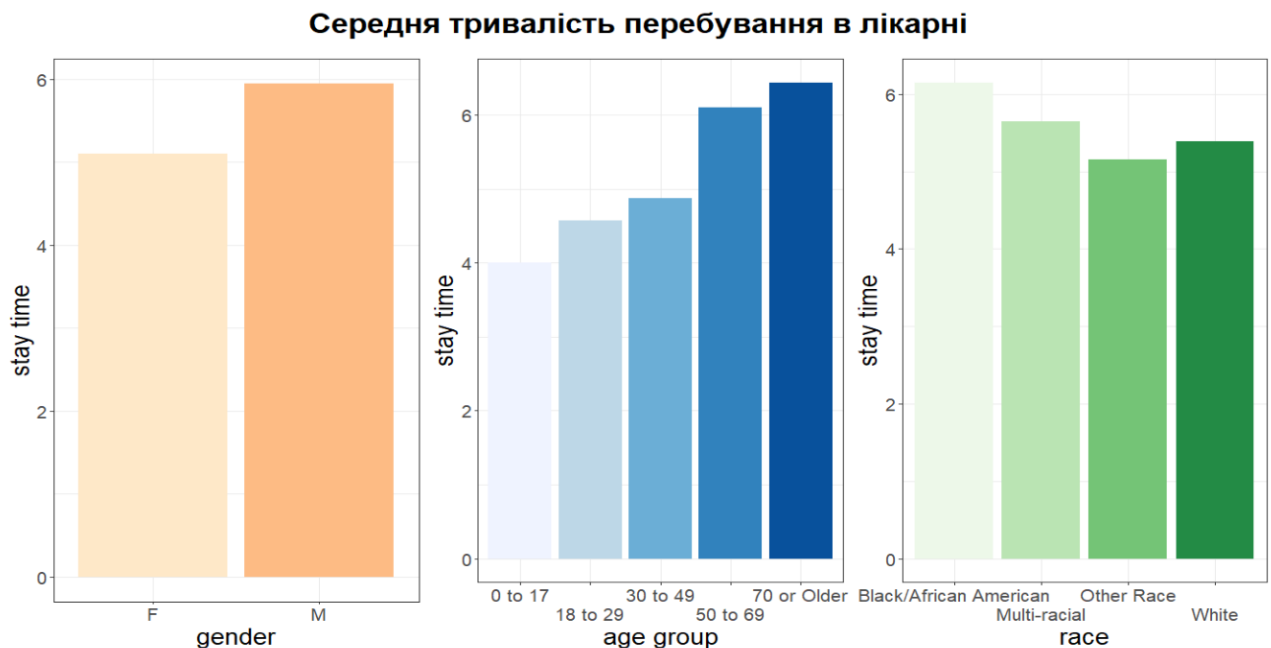


рис. 8 - Залежності тривалості лікування від віку, статі та раси

Із цих графіків можна зробити такі орієнтовні твердження:

1. Як можна побачити, в середньому чоловіки лежать в лікарні майже на день довше (на 16% довше). Ми маємо припущення, що однією з причин може бути те, що чоловіки рідше ходять в лікарню і більш схильні ігнорувати ранні симптоми, відповідно з цим, час на лікування збільшується.

2. Старше покоління проводить в лікарні більше часу за молодших. Цікавим є той факт, що залежність візуально дуже схожа на лінійну.

3. Виявилось, що довжина перебування в лікарні не є незалежною від раси. Найбільше серед усіх часу у лікарні, якщо припускати справедливості спостережень з графіку, проводять афроамериканці, а найменше - представники інших рас, тобто азіати, корінні американці та жителі Тихоокеанських островів, що в парі з тим фактом, що вони дуже часто звертаються в лікарні, наводить на

подальші питання. Представники білих та мультирасових пацієнтів не виділяються на тлі інших.

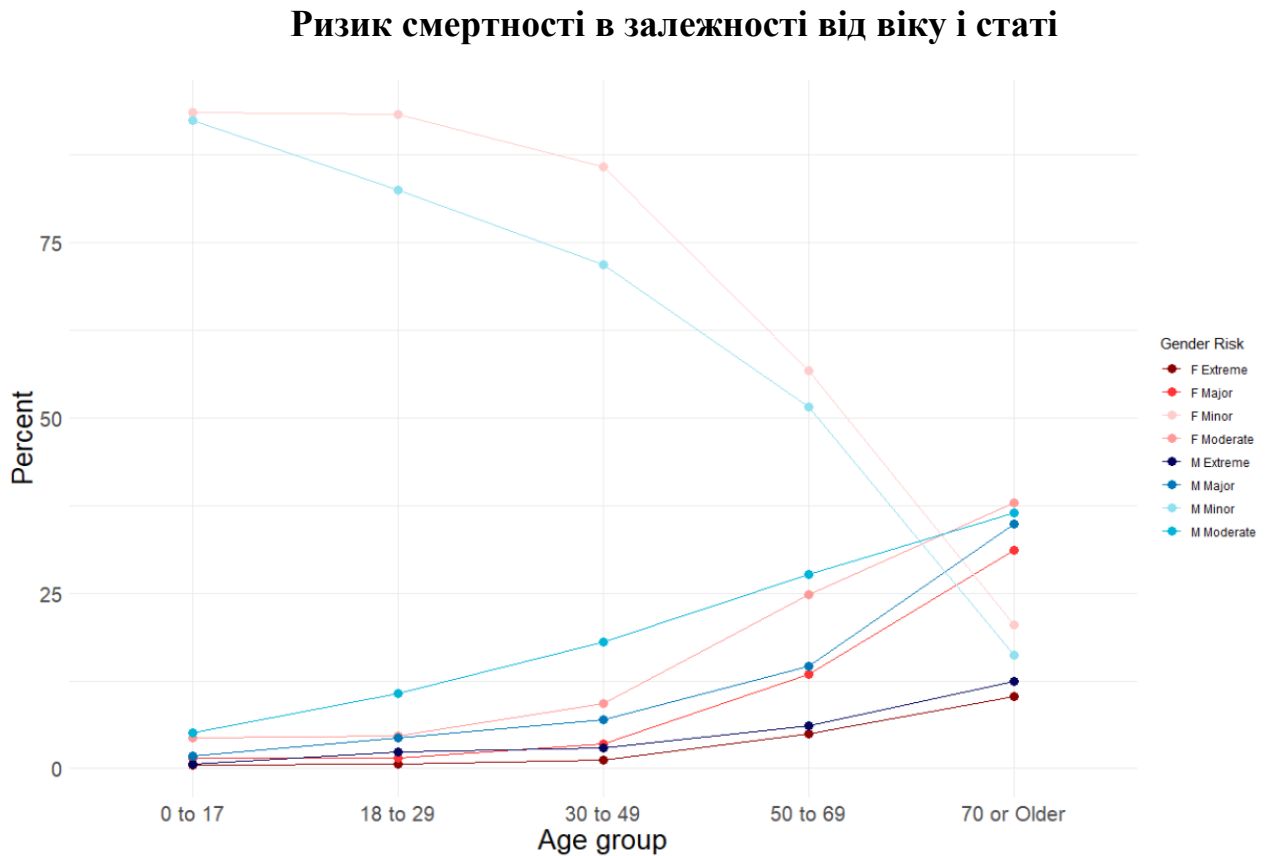


рис. 9 - Ризик смертності у відсотках в залежності від віку та статі

На цьому графіку зображено розподіл ризиків смертності за статтю та віком. Більш світлі кольори відповідають меншому ризику, чим темніша - тим більший. Із графіку можна винести такі спостереження:

- з віком збільшується відсоток людей, що мають більший ризик смертності. Можливо, саме вік сам по собі впливає на визначення цього ризику для конкретного випадку.
- відсоток чоловіків у середньому, високому і екстремальному станах більший за відповідний відсоток жінок для усіх вікових категорій, крім 0-17 років, і відповідно, менший за відсоток жінок для легкого стану. Це теж може бути наслідком теорії про ігнорування чоловіками дрібних симптомів чи бути пов'язаним з іншими факторами.

Ці теорії також потребують подальших досліджень.

Ризик смертності в залежності від раси

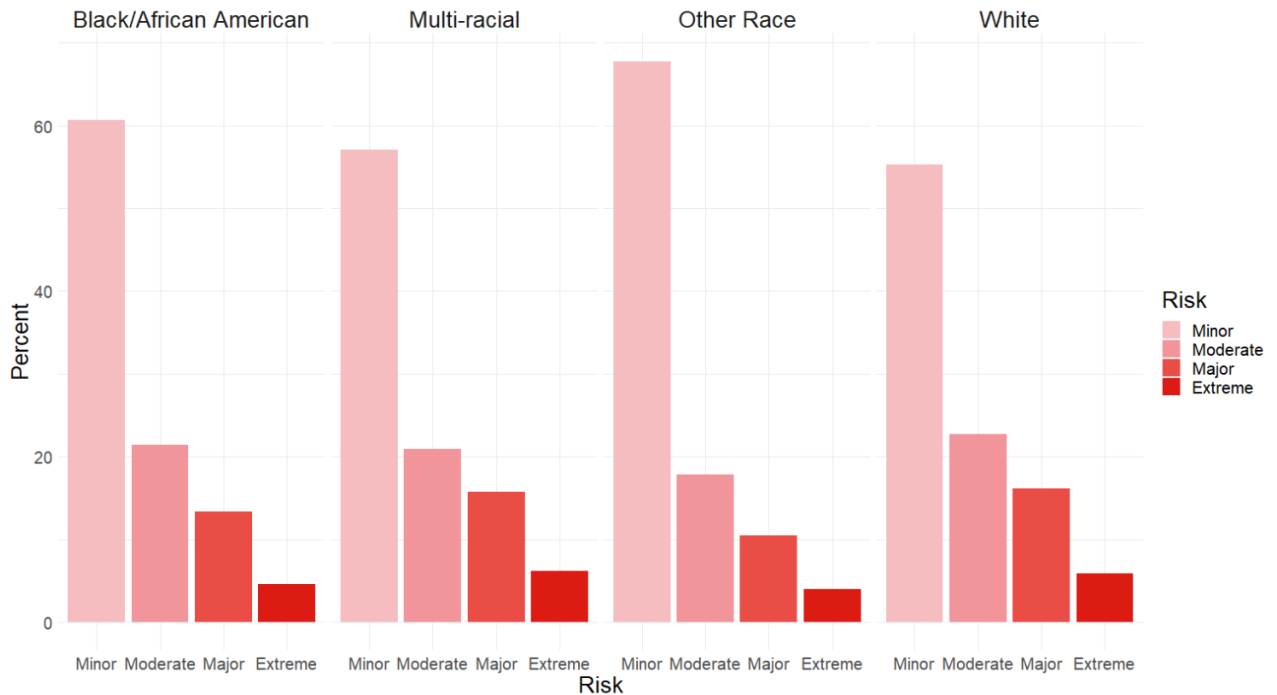


рис. 10 - Ризик смертності у відсотках в залежності від віку та статі

Цей графік показує розподіл залежності ризику смерті пацієнтів залежно від їх раси. Кількості випадків для станів з рівнем смертності Extreme приблизно однакові, хоч це необхідно підтвердити статистичними дослідженнями.

Також, для мультирасових, афроамериканських та білих пацієнтів кількість випадків рівня ризику Major приблизно однакова і, якщо дивитись на графік, приблизно на рівні 15%, в той час як пацієнти азійського походження мають цю характеристику на рівні приблизно 10%. Якщо ця різниця статистично значуща, то можна було б зробити висновок, що вони потрапляють в лікарню з цим ризиком смерті менше за інші раси. Схожа ситуація з рівнем Moderate - пацієнти азійського походження теж можуть мати менший відсоток за інших, але тут різниця між значеннями візуально ще менша.

До того ж, варта уваги категорія ризику смертності Minor - якщо дивитись виключно на дані графіку, то пацієнти азійського походження мають помітно більший відсоток звернень в лікарню з випадками, ризик смертності яких мінімальний.

Висновки

В ході роботи було досліджено набір даних пацієнтів лікарень штату Нью-Йорк за 2015 рік. Із проведеного дослідження можна винести такі гіпотези і можливі цікаві закономірності, які обов'язково потребують подальшої перевірки на достовірність:

1. Розподіл собівартостей та цін за обслуговування в лікарнях схожий на логарифмічно нормальний.
2. Розподіл тривалостей перебування в лікарні не схожий на якісь специфічні, а кількість випадків спадає за збільшення тривалості.
3. Щодо характеристик пацієнтів за віком:
 - a. Їх розподіл відносно кількості людей цієї ж категорії в населенні дає підґрунтя для спостереження, що люди до 50 років потрапляють в лікарню приблизно з однаковою частотою, для 50-69 років ця частота зростає, а найчастіше туди потрапляють люди віку 70+ років, якщо дані різниці є статистично значущими. Це відповідає попереднім припущенням про те, що в лікарню здебільшого потрапляють люди старшого віку.
 - b. Чим старша людина, тим більше в середньому час, який вона проводить в лікарні, до того ж, ця закономірність для різних вікових груп навіть віддалено схожа на лінійну. Це протирічить нашому попередньому припущенню, що для людей 40+ тривалість перебування в лікарні буде приблизно однаковою. Також, очікувана тривалість для дітей була меншою на нашу думку, ніж виявилась насправді у цьому датасеті.
 - c. У 70+ років вибіркова ймовірність потрапити у лікарню з ризиком смерті moderate та major вища, ніж з ризиком смерті minor. Можливо, так стається чи через більшу кількість серйозних станів, чи через вплив віку на оцінку ризику смерті. Це відповідає нашим попереднім уявленням.
4. Щодо розподілу за статтю:
 - a. Чоловіки менше потрапляють в лікарню, при чому, навіть якщо врахувати їх відсоток в населенні.
 - b. Чоловіки в середньому лежать в лікарні довше на 1 день, якщо ця різниця статистично значуща.
 - c. Імовірність чоловіків потрапити в лікарню з ризиком смерті, більшим за minor, трохи більша, ніж у жінок. Для minor навпаки - жінки частіше за чоловіків потрапляють в лікарню з цим ризиком.

Наші припущення стверджували, що характеристики за статтю будуть приблизно однакові і в чоловіків, і в жінок. Це припущення, здається, доведеться відкинути, але лише після подальшої перевірки.

5. Щодо розподілу за расою:

- a. Представники білої раси потрапляють в лікарні менше, ніж можна було б очікувати за розподілом населення, а представники інших рас (тобто люди азійського походження та корінні жителі Америки)- більше.
- b. Представники інших рас перебувають в лікарні менше за інші 3 категорії, а афроамериканці - більше.
- c. Представники інших рас мають більший відсоток випадків з ризиком смерті тіло, ніж інші 3 категорії, відповідно, відсоток важчих випадків менший.

Ці результати помітно відрізняються від попередніх гіпотез, і причини цього потребують подальшого дослідження.

Наша команда планує і надалі досліджувати цей датасет, аби довести чи спростувати вже знайдені закономірності, а також віднайти, можливо, нові, пов'язані з вже розглянутими змінними чи ще ні.

Додаток А.
Коди CSS [2]

- 1 Tuberculosis
- 2 Septicemia (except in labor)
- 3 Bacterial infection; unspecified site
- 4 Mycoses
- 5 HIV infection
- 6 Hepatitis
- 7 Viral infection
- 8 Other infections; including parasitic
- 9 Sexually transmitted infections (not HIV or hepatitis)
- 10 Immunizations and screening for infectious disease
- 11 Cancer of head and neck
- 12 Cancer of esophagus
- 13 Cancer of stomach
- 14 Cancer of colon
- 15 Cancer of rectum and anus
- 16 Cancer of liver and intrahepatic bile duct
- 17 Cancer of pancreas
- 18 Cancer of other GI organs; peritoneum
- 19 Cancer of bronchus; lung
- 20 Cancer; other respiratory and intrathoracic
- 21 Cancer of bone and connective tissue
- 22 Melanomas of skin
- 23 Other non-epithelial cancer of skin
- 24 Cancer of breast
- 25 Cancer of uterus
- 26 Cancer of cervix
- 27 Cancer of ovary
- 28 Cancer of other female genital organs
- 29 Cancer of prostate
- 30 Cancer of testis
- 31 Cancer of other male genital organs
- 32 Cancer of bladder
- 33 Cancer of kidney and renal pelvis
- 34 Cancer of other urinary organs
- 35 Cancer of brain and nervous system
- 36 Cancer of thyroid
- 37 Hodgkin`s disease
- 38 Non-Hodgkin`s lymphoma
- 39 Leukemias
- 40 Multiple myeloma
- 41 Cancer; other and unspecified primary
- 42 Secondary malignancies
- 43 Malignant neoplasm without specification of site

- 44 Neoplasms of unspecified nature or uncertain behavior
- 45 Maintenance chemotherapy; radiotherapy
- 46 Benign neoplasm of uterus
- 47 Other and unspecified benign neoplasm
- 48 Thyroid disorders
- 49 Diabetes mellitus without complication
- HCUP CCS 10/27/2011 19
- 50 Diabetes mellitus with complications
- 51 Other endocrine disorders
- 52 Nutritional deficiencies
- 53 Disorders of lipid metabolism
- 54 Gout and other crystal arthropathies
- 55 Fluid and electrolyte disorders
- 56 Cystic fibrosis
- 57 Immunity disorders
- 58 Other nutritional; endocrine; and metabolic disorders
- 59 Deficiency and other anemia
- 60 Acute posthemorrhagic anemia
- 61 Sickle cell anemia
- 62 Coagulation and hemorrhagic disorders
- 63 Diseases of white blood cells
- 64 Other hematologic conditions
- 650 MHSA: Adjustment disorders
- 651 MHSA: Anxiety disorders
- 652 MHSA: Attention-deficit, conduct, and disruptive behavior disorders
- 653 MHSA: Delirium, dementia, and amnestic and other cognitive disorders
- 654 MHSA: Developmental disorders
- 655 MHSA: Disorders usually diagnosed in infancy, childhood, or adolescence
- 656 MHSA: Impulse control disorders, NEC
- 657 MHSA: Mood disorders
- 658 MHSA: Personality disorders
- 659 MHSA: Schizophrenia and other psychotic disorders
- 660 MHSA: Alcohol-related disorders
- 661 MHSA: Substance-related disorders
- 662 MHSA: Suicide and intentional self-inflicted injury
- 663 MHSA: Screening and history of mental health and substance abuse codes
- 670 MHSA: Miscellaneous mental health disorders
- 76 Meningitis (except that caused by tuberculosis or sexually transmitted disease)
- 77 Encephalitis (except that caused by tuberculosis or sexually transmitted disease)
- 78 Other CNS infection and poliomyelitis
- 79 Parkinson`s disease
- 80 Multiple sclerosis
- 81 Other hereditary and degenerative nervous system conditions
- 82 Paralysis
- 83 Epilepsy; convulsions
- 84 Headache; including migraine

- 85 Coma; stupor; and brain damage
- 86 Cataract
- 87 Retinal detachments; defects; vascular occlusion; and retinopathy
- 88 Glaucoma
- 89 Blindness and vision defects
- 90 Inflammation; infection of eye (except that caused by tuberculosis or sexually transmitted disease)
- 91 Other eye disorders
- 92 Otitis media and related conditions
- 93 Conditions associated with dizziness or vertigo
- 94 Other ear and sense organ disorders
- 95 Other nervous system disorders
- 96 Heart valve disorders
- 97 Peri-; endo-; and myocarditis; cardiomyopathy (except that caused by tuberculosis or sexually transmitted disease)
- 98 Essential hypertension
- 99 Hypertension with complications and secondary hypertension
- HCUP CCS 10/27/2011 20
- 100 Acute myocardial infarction
- 101 Coronary atherosclerosis and other heart disease
- 102 Nonspecific chest pain
- 103 Pulmonary heart disease
- 104 Other and ill-defined heart disease
- 105 Conduction disorders
- 106 Cardiac dysrhythmias
- 107 Cardiac arrest and ventricular fibrillation
- 108 Congestive heart failure; nonhypertensive
- 109 Acute cerebrovascular disease
- 110 Occlusion or stenosis of precerebral arteries
- 111 Other and ill-defined cerebrovascular disease
- 112 Transient cerebral ischemia
- 113 Late effects of cerebrovascular disease
- 114 Peripheral and visceral atherosclerosis
- 115 Aortic; peripheral; and visceral artery aneurysms
- 116 Aortic and peripheral arterial embolism or thrombosis
- 117 Other circulatory disease
- 118 Phlebitis; thrombophlebitis and thromboembolism
- 119 Varicose veins of lower extremity
- 120 Hemorrhoids
- 121 Other diseases of veins and lymphatics
- 122 Pneumonia (except that caused by tuberculosis or sexually transmitted disease)
- 123 Influenza
- 124 Acute and chronic tonsillitis
- 125 Acute bronchitis
- 126 Other upper respiratory infections
- 127 Chronic obstructive pulmonary disease and bronchiectasis

- 128 Asthma
- 129 Aspiration pneumonitis; food/vomitus
- 130 Pleurisy; pneumothorax; pulmonary collapse
- 131 Respiratory failure; insufficiency; arrest (adult)
- 132 Lung disease due to external agents
- 133 Other lower respiratory disease
- 134 Other upper respiratory disease
- 135 Intestinal infection
- 136 Disorders of teeth and jaw
- 137 Diseases of mouth; excluding dental
- 138 Esophageal disorders
- 139 Gastroduodenal ulcer (except hemorrhage)
- 140 Gastritis and duodenitis
- 141 Other disorders of stomach and duodenum
- 142 Appendicitis and other appendiceal conditions
- 143 Abdominal hernia
- 144 Regional enteritis and ulcerative colitis
- 145 Intestinal obstruction without hernia
- 146 Diverticulosis and diverticulitis
- 147 Anal and rectal conditions
- 148 Peritonitis and intestinal abscess
- 149 Biliary tract disease
- 150 Liver disease; alcohol-related
- 151 Other liver diseases
- 152 Pancreatic disorders (not diabetes)
- 153 Gastrointestinal hemorrhage
- 154 Noninfectious gastroenteritis
- 155 Other gastrointestinal disorders
- HCUP CCS 10/27/2011 21
- 156 Nephritis; nephrosis; renal sclerosis
- 157 Acute and unspecified renal failure
- 158 Chronic kidney disease
- 159 Urinary tract infections
- 160 Calculus of urinary tract
- 161 Other diseases of kidney and ureters
- 162 Other diseases of bladder and urethra
- 163 Genitourinary symptoms and ill-defined conditions
- 164 Hyperplasia of prostate
- 165 Inflammatory conditions of male genital organs
- 166 Other male genital disorders
- 167 Nonmalignant breast conditions
- 168 Inflammatory diseases of female pelvic organs
- 169 Endometriosis
- 170 Prolapse of female genital organs
- 171 Menstrual disorders
- 172 Ovarian cyst

- 173 Menopausal disorders
- 174 Female infertility
- 175 Other female genital disorders
- 176 Contraceptive and procreative management
- 177 Spontaneous abortion
- 178 Induced abortion
- 179 Postabortion complications
- 180 Ectopic pregnancy
- 181 Other complications of pregnancy
- 182 Hemorrhage during pregnancy; abruptio placenta; placenta previa
- 183 Hypertension complicating pregnancy; childbirth and the puerperium
- 184 Early or threatened labor
- 185 Prolonged pregnancy
- 186 Diabetes or abnormal glucose tolerance complicating pregnancy; childbirth; or the puerperium
- 187 Malposition; malpresentation
- 188 Fetopelvic disproportion; obstruction
- 189 Previous C-section
- 190 Fetal distress and abnormal forces of labor
- 191 Polyhydramnios and other problems of amniotic cavity
- 192 Umbilical cord complication
- 193 OB-related trauma to perineum and vulva
- 194 Forceps delivery
- 195 Other complications of birth; puerperium affecting management of mother
- 196 Other pregnancy and delivery including normal
- 197 Skin and subcutaneous tissue infections
- 198 Other inflammatory condition of skin
- 199 Chronic ulcer of skin
- 200 Other skin disorders
- 201 Infective arthritis and osteomyelitis (except that caused by tuberculosis or sexually transmitted disease)
- 202 Rheumatoid arthritis and related disease
- 203 Osteoarthritis
- 204 Other non-traumatic joint disorders
- 205 Spondylosis; intervertebral disc disorders; other back problems
- 206 Osteoporosis
- 207 Pathological fracture
- 208 Acquired foot deformities
- 209 Other acquired deformities
- HCUP CCS 10/27/2011 22
- 210 Systemic lupus erythematosus and connective tissue disorders
- 211 Other connective tissue disease
- 212 Other bone disease and musculoskeletal deformities
- 213 Cardiac and circulatory congenital anomalies
- 214 Digestive congenital anomalies
- 215 Genitourinary congenital anomalies

216 Nervous system congenital anomalies
217 Other congenital anomalies
218 Liveborn
219 Short gestation; low birth weight; and fetal growth retardation
220 Intrauterine hypoxia and birth asphyxia
221 Respiratory distress syndrome
222 Hemolytic jaundice and perinatal jaundice
223 Birth trauma
224 Other perinatal conditions
225 Joint disorders and dislocations; trauma-related
226 Fracture of neck of femur (hip)
227 Spinal cord injury
228 Skull and face fractures
229 Fracture of upper limb
230 Fracture of lower limb
231 Other fractures
232 Sprains and strains
233 Intracranial injury
234 Crushing injury or internal injury
235 Open wounds of head; neck; and trunk
236 Open wounds of extremities
237 Complication of device; implant or graft
238 Complications of surgical procedures or medical care
239 Superficial injury; contusion
240 Burns
241 Poisoning by psychotropic agents
242 Poisoning by other medications and drugs
243 Poisoning by nonmedicinal substances
244 Other injuries and conditions due to external causes
245 Syncope
246 Fever of unknown origin
247 Lymphadenitis
248 Gangrene
249 Shock
250 Nausea and vomiting
251 Abdominal pain
252 Malaise and fatigue
253 Allergic reactions
254 Rehabilitation care; fitting of prostheses; and adjustment of devices
255 Administrative/social admission
256 Medical examination/evaluation
257 Other aftercare
258 Other screening for suspected conditions (not mental disorders or infectious disease)
259 Residual codes; unclassified
260 E Codes: All (external causes of injury and poisoning)
2601 E Codes: Cut/pierceb

2602 E Codes: Drowning/submersion

2603 E Codes: Fall

2604 E Codes: Fire/burn

2605 E Codes: Firearm

HCUP CCS 10/27/2011 23

2606 E Codes: Machinery

2607 E Codes: Motor vehicle traffic (MVT)

2608 E Codes: Pedal cyclist; not MVT

2609 E Codes: Pedestrian; not MVT

2610 E Codes: Transport; not MVT

2611 E Codes: Natural/environment

2612 E Codes: Overexertion

2613 E Codes: Poisoning

2614 E Codes: Struck by; against

2615 E Codes: Suffocation

2616 E Codes: Adverse effects of medical care

2617 E Codes: Adverse effects of medical drugs

2618 E Codes: Other specified and classifiable

2619 E Codes: Other specified; NEC

2620 E Codes: Unspecified

2621 E Codes: Place of occurrence

Додаток Б.
Коди MDC [6]

MDC	Description	<u>MS-DRG^{[1][2]}</u>
0	Pre-MDC	001 - 017
1	Diseases and Disorders of the Nervous System	020 - 103
2	Diseases and Disorders of the Eye	113 - 125
3	Diseases and Disorders of the Ear, Nose, Mouth And Throat	129 - 159
4	Diseases and Disorders of the Respiratory System	163 - 208
5	Diseases and Disorders of the Circulatory System	215 - 316
6	Diseases and Disorders of the Digestive System	326 - 395
7	Diseases and Disorders of the Hepatobiliary System And Pancreas	405 - 446
8	Diseases and Disorders of the Musculoskeletal System And Connective Tissue	453 - 566
9	Diseases and Disorders of the Skin, Subcutaneous Tissue And Breast	573 - 607
10	Diseases and Disorders of the Endocrine, Nutritional And Metabolic System	614 - 645

11	Diseases and Disorders of the Kidney And Urinary Tract	652 - 700
12	Diseases and Disorders of the Male Reproductive System	707 - 730
13	Diseases and Disorders of the Female Reproductive System	734 - 761
14	Pregnancy, Childbirth And Puerperium	765 - 782
15	Newborn And Other Neonates (Perinatal Period)	789 - 795
16	Diseases and Disorders of the Blood and Blood Forming Organs and Immunological Disorders	799 - 816
17	Myeloproliferative DDs (Poorly Differentiated Neoplasms)	820 - 849
18	Infectious and Parasitic DDs (Systemic or unspecified sites)	853 - 872
19	Mental Diseases and Disorders	876 - 887
20	Alcohol/Drug Use or Induced Mental Disorders	894 - 897
21	Injuries, Poison And Toxic Effect of Drugs	901 - 923
22	Burns	927 - 935
23	Factors Influencing Health Status and Other Contacts with Health Services	939 - 951

24	Multiple Significant Trauma	955 - 965
25	Human Immunodeficiency Virus Infection	969 - 977
	MDC Category Missing	981 - 989; 998,999

Джерела

1. <https://health.data.ny.gov/Health/Hospital-Inpatient-Discharges-SPARCS-De-Identified/82xm-y6g8>
2. <https://hcup-us.ahrq.gov/db/nation/nis/APR-DRGsV20MethodologyOverviewandBibliography.pdf>
3. <https://hcup-us.ahrq.gov/toolssoftware/ccs/CCSUsersGuide.pdf>
4. https://en.wikipedia.org/wiki/Major_Diagnostic_Category
5. https://en.wikipedia.org/wiki/Diagnosis-related_group
6. <https://hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp>
7. https://www.health.ny.gov/statistics/vital_statistics/2015/table01.htm
8. <https://usafacts.org/data/topics/people-society/population-and-demographics/our-changing-population/state/new-york?endDate=2021-01-01&startDate=2015-01-01>