

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ХАРЧОВИХ ТЕХНОЛОГІЙ

Інститут (факультет) Автоматизації і комп'ютерних систем
Кафедра Інформаційних технологій, штучного інтелекту і кібербезпеки

«До захисту в ЕК»

Директор інституту(декан факультету)
Андрій ФОРСЮК

(підпис) (ім'я та прізвище)

«13» грудня 2024р.

«До захисту допущено»

Завідувач кафедри

Сергій ГРИБКОВ

(підпис) (ім'я та прізвище)

«13» грудня 2024р.

КВАЛІФІКАЦІЙНА РОБОТА
НА ЗДОБУТТЯ ОСВІТНЬОГО СТУПЕНЯ МАГІСТРА

Зі спеціальності 122 «Комп'ютерні науки»

(код та назва спеціальності)

освітньо-професійної програми Управління інформацією та аналітика даних
на тему: Дослідження та створення аналітичної системи для аналізу продажів алкогольних напоїв

Виконав: здобувач 2 курсу, групи КН-2-4М

Дзюбан Іванна Іванівна

(прізвище, ім'я, по батькові повністю)

(підпис)

Керівник Харкянен Олена Валеріївна

(прізвище, ім'я та по батькові повністю)

(підпис)

Консультанти

(ім'я та прізвище)

(підпис)

(ім'я та прізвище)

(підпис)

Рецензент

(ім'я та прізвище)

(підпис)

Я як здобувачка Національного університету харчових технологій розумію і підтримую політику університету з академічної доброчесності. Я не надавала і не одержувала незгоєваної допомоги під час підготовки цієї роботи. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

Здобувач

(підпис)

Київ - 2024р.

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ХАРЧОВИХ ТЕХНОЛОГІЙ

Інститут (факультет) Автоматизації і комп'ютерних систем
 Кафедра Інформаційних технологій, штучного інтелекту і кібербезпеки
 Освітній ступінь магістр
 Спеціальність 122 «Комп'ютерні науки»
(код і назва)
 Освітньо-професійна програма Управління інформацією і аналітика даних
(назва)

ЗАТВЕРДЖУЮ

Завідувач
 кафедри Інформаційних технологій,
 штучного інтелекту і кібербезпеки

Грибков С.В.

“ 07 ” жовтня 2024 року

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ ЗДОБУВАЧА

Дзюбан Іванна Іванівна

(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження та створення аналітичної системи для аналізу продажів алкогольних напоїв
 керівник роботи Харкянен Олена Валеріївна доцент, кандидат технічних наук,
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)
 затверджені наказом закладу вищої освіти від 7 жовтня 2024 року №884-кс
2. Строк подання здобувачем роботи 6 грудня 2024 року
3. Вихідні дані до роботи Аналітичні дані про алкогольні напої та кореляції їх покупки клієнтами.
 Статистичні дані з платформи Kaggle.
4. Зміст пояснювальної записки (перелік питань, які потрібно розробити)
Розділ 1. Огляд літератури та існуючих підходів.
Розділ 2. Обґрунтування вибору напрямку досліджень і методів вирішення задач
Розділ 3. Реалізація основних завдань плану наукових досліджень.
5. Перелік графічного матеріалу
Аналітичні зрізи з масиву даних

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
1	Харкянен О.В., доцент	08.10.2024	19.10.2024
2	Харкянен О.В., доцент	19.10.2024	15.11.2024
3	Харкянен О.В., доцент	18.11.2024	02.12.2024

7. Дата видачі завдання 7 жовтня 2024 року

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів виконання кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
1	Аналіз предметної галузі алкогольних напоїв	08.10.2024-12.10.2024	Виконано
2	Створення аналітичної системи для аналізу продажів алкогольних напоїв	13.10.2024-27.11.2024	Виконано
3	Дослідження результатів	26.11.2024-06.12.2024	Виконано
4	Створення документу кваліфікаційної роботи	19.10.2024-06.12.2024	Виконано
5	Створення автореферату	01.12.2024-03.12.2024	Виконано
6	Створення презентації	04.12.2024-06.12.2024	Виконано

Здобувач


(підпис)

Дзюбан І.І.

(прізвище та ініціали)

Керівник роботи


(підпис)

Харкянен О.В.

(прізвище та ініціали)

АНОТАЦІЯ

Дзюбан Іванна Іванівна – «Дослідження та створення аналітичної системи для аналізу продажів алкогольних напоїв».

Кваліфікаційна робота: 74 сторінок, 25 рисунків, 10 таблиць, 2 додатки, 95 джерел.

Метою кваліфікаційної роботи є дослідження методів аналізу великих даних та створення системи для аналізу продажів алкогольних напоїв на основі масивів даних з відкритих інформаційних джерел.

Розроблена аналітична система, яка включає збір та обробку даних. Проведено статистичний аналіз продажів алкогольних напоїв. Виявлено ключові тренди, включаючи сезонність та регіональні відмінності.

Ключові слова: АНАЛІЗ ДАНИХ, ПРОДАЖІ АЛКОГОЛЬНИХ НАПОЇВ, АНАЛІТИЧНА СИСТЕМА, PYTHON, ПРОГНОЗУВАННЯ, ТРЕНДИ, МАСИВ ДАНИХ.

SUMMARY

Dziuban Ivanna Ivanivna – “Research and creation of an analytical system for analyzing sales of alcoholic beverages”.

Qualification work: 74 pages, 25 figures, 10 tables, 2 appendices, 95 sources.

The purpose of the qualification work is to study methods of big data analysis and create a system for analyzing sales of alcoholic beverages based on data sets from open information sources.

An analytical system has been developed, which includes data collection and processing. A statistical analysis of sales of alcoholic beverages has been conducted. Key trends have been identified, including seasonality and regional differences.

Keywords: DATA ANALYSIS, ALCOHOLIC BEVERAGE SALES, ANALYTICAL SYSTEM, PYTHON, FORECASTING, TRENDS, DATA ASSESSMENT.

ЗМІСТ

ВСТУП	8
РОЗДІЛ 1. ОГЛЯД ЛІТЕРАТУРИ ТА ІСНУЮЧИХ ПІДХОДІВ	11
1.1 Аналіз ринку алкогольних напоїв та тенденцій його розвитку.....	11
1.2 Огляд досліджень щодо продажу алкогольних напоїв в Україні	15
1.3 Аналіз сучасних методів та підходів до обробки інформації.....	17
1.4 Висновок до першого розділу.....	20
РОЗДІЛ 2. ОБҐРУНТУВАННЯ ВИБОРУ НАПРЯМКУ ДОСЛІДЖЕНЬ І МЕТОДІВ ВИРІШЕННЯ ЗАДАЧ.....	22
2.1 Етапи розробки інформаційної системи	23
2.2 Дослідження методів для вирішення задач	24
2.3 Обґрунтування вибору інструментів реалізації	26
2.4 Висновок до другого розділу	28
РОЗДІЛ 3. РЕАЛІЗАЦІЯ ОСНОВНИХ ЗАВДАНЬ АНАЛІЗУ ПРОДАЖУ АЛКОГОЛЬНИХ НАПОЇВ	30
3.1 Формування переліку задач дослідження.....	30
3.2 Збір масивів даних, необхідних для реалізації системи.....	32
3.3 Компоненти інформаційної системи та їх налаштування.....	33
3.4 Структура даних бази даних	39
3.5 Розробка візуального представлення даних	40
3.6 Висновок до третього розділу.....	58
ВИСНОВКИ.....	59
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	60
ДОДАТКИ.....	69

Перелік умовних позначень

1. BI - бізнес-аналітика.
2. K-means - K-середніх.
3. FP - GrowthFrequent Pattern Growth. Зростання частих шаблонів.
4. RFM-аналіз - Recency, Frequency, Monetary. Аналіз давності, частоти та грошової цінності.
5. LSTM - Long Short-Term Memory. Довга короткострокова пам'ять.
6. КПІ - Ключові показники ефективності.
7. Веб-скрапінг - Web Scraping. Збір інформації з веб-сайтів.

ВСТУП

Актуальність теми. Сегмент алкогольної продукції дуже важливий та актуальний тому що, це дає великий прибуток державі, а також ця галузь стрімко розвивається. На теперішній час алкогольна продукція є однією з популярних і прибуткових товарних категорій на світовому ринку. Аналіз продажів даних дозволяє виявити не тільки основні тенденції у споживанні алкогольних напоїв, а й спрогнозувати поведінку ринку. Аналіз даних є особливо суттєвим для виробників і роздрібних мереж. Ринок має занадто високу конкуренцію, та має певні обмеження та вимоги до якості напоїв. Всі фактори ускладнюють прогнозування попиту алкогольної продукції, маркетингових стратегій та управління якості алкогольних напоїв. Завдяки ефективному аналізу продажів бізнес може своєчасно реагувати на зміни в попиті і підлаштовуватися до ринкових умов.

Зв'язок роботи з науковими програмами, планами, темами кафедри, університету. Кваліфікаційна робота виконана у відповідності з тематикою наукових досліджень кафедри інформаційних технологій, штучного інтелекту і кібербезпеки НУХТ.

Мета дослідження. Метою кваліфікаційної роботи є дослідження та розробка аналітичної системи для аналізу продажів алкогольних напоїв. Впровадження системи позитивно вплине на обсяг продажів, цінову політику та прибутки харчових підприємств. Алкогольний ринок залежний від багатьох факторів. Саме тому важливо розробити ефективну аналітичну систему, що дозволить проаналізувати великі обсяги даних та прийняти обґрунтовані управлінські рішення.

Держави регулюють виготовлення алкоголю, тому підприємства-виробники мають дотримуватися багатьох обмежень та зобов'язані враховувати соціальні аспекти у своїй діяльності. Аналітична система призначена для покращення ефективності формування витрат на виготовлення продукції за рахунок більш точного планування закупівлі ресурсів для виготовлення алкогольних напоїв та покращення логістики.

Використання аналітичних зрізів даних допоможе ефективніше планувати маркетингові кампанії. Проведення аналізу продажів алкогольної продукції сприятиме визначенню які акції, сезон, знижки та пропозиції найбільше впливають на збільшення прибутку від продажів алкогольних напоїв.

Завдання дослідження.

1. Дослідити ринок продажів алкогольних напоїв на основі відкритих даних.
2. Дослідити ринок споживання алкогольних напоїв на основі відкритих даних.
3. Визначити перелік задач аналізу продажів алкогольних напоїв та методи їх реалізації.
4. Розробити аналітичну систему продажів та споживання алкогольних напоїв з використанням сучасних ІТ-засобів для візуалізації та моніторингу даних.

Об'єкт дослідження. Дані про продажі та динаміку попиту на алкогольні напої.

Предмет дослідження. Алгоритми прогнозування продажів та споживання на алкогольних напоїв. Програмні засоби для візуалізації даних та інтерпретації отриманих результатів для підтримки прийняття управлінських рішень, що збільшить прибуток від продажів алкогольної продукції.

Методи дослідження.

- Веб-скрапінг – збір інформації з різних сайтів та відкритих джерел.
- Статистичний аналіз – визначення кореляцій у даних, сезонності, дослідження трендів та сезонних коливань.
- Кластерний аналіз - поділ ринків або клієнтів за регіонами, за споживчими уподобаннями щодо алкогольної продукції.

Наукова новизна. Полягає у визначенні переліку задач для аналізу продажів алкогольних напоїв на основі дослідження масивів даних; адаптації

методів аналізу великих даних для визначеного переліку задач продажів алкогольних напоїв.

Особистий внесок здобувача.

- Розробка концепції аналітичної системи для аналізу продажів алкогольних напоїв.
- Адаптація та реалізація алгоритмів очищення даних, прогнозування та візуалізації даних.

Публікації.

1. Дзюбан І.І., Харкянен О.В. «Розроблення систем прогнозування на основі машинного навчання» // «Штучний інтелект та інформаційні технології» (АІТ-2024): наукові праці І Міжнародної науково-практичної конференції, м. Київ, Україна, 3–4 червня 2024 р. – К.: НУХТ, 2024. – С. 49-50.

2. Дзюбан І.І., Харкянен О.В. «Аналіз великих даних у реальному часі: технології та алгоритми» // «Штучний інтелект та інформаційні технології» (АІТ-2024): наукові праці І Міжнародної науково-практичної конференції, м. Київ, Україна, 3–4 червня 2024 р. – К.: НУХТ, 2024. – С. 121–122.

РОЗДІЛ 1. ОГЛЯД ЛІТЕРАТУРИ ТА ІСНУЮЧИХ ПІДХОДІВ

1.1 Аналіз ринку алкогольних напоїв та тенденцій його розвитку

Проведений аналіз ринку продажів алкогольної продукції виявив що загальний обсяг споживання алкоголю з кожним роком зростає. Найпомітніший приріст вживання алкоголю припадає на 2023 рік. На цей час, якраз припадає скасування карантину, який пов'язаний із пандемією COVID-19. Швидко розширюється та покращується сегмент міцного алкоголю, який розлитий до металевих бляшанок об'ємом 0,33 або 0,5 л. Такий об'єм людям комфортніше носити з собою. Це в свою чергу дає переваги, і напій зручно брати на відпочинок на природі або у подорож. Найактивнішими споживачами спиртного є міленіали. Міленіали - це перше покоління, що виросло в епоху інтернету [3]. На нього припадає майже 62% від загального продажу алкоголю. Водночас міленіали шукають альтернативу цукру, а також нові чи незвичайні смаки [1]. З кожним днем все більше зростає попит на безалкогольні напої. Все більше і більше асортимент алкогольних напоїв поповнюється новими смаками.

Безалкогольні та слабоалкогольні напої набувають популярності. Вони стають частиною правильного способу життя та соціальною нормою. В майбутньому споживання буде рости та підтримуватися збільшенням рівня доходів клієнтів та новими учасниками ринку, що особливо актуально для Азії та США. Вживання алкоголю є популярним серед заможних споживачів та є прибутковим для компаній.

Крім того, відбувається диверсифікація сегментів. Добавляються нові та статусні категорії. Це міцні спиртні напої на основі агави, ірландський, американський та японський віскі. Ріст продажів статусних алкогольних напоїв припадає на 2020 році. Споживачі все частіше шукають компанії з виробництва напоїв, що демонструють прихильність до стійких практик виробництва алкогольних напоїв [2].

За даними [7] пиво є одним із найбільш популярних алкогольних напоїв в Україні. Близько 43% українців споживають пиво, 37% і 34% відповідно –

вину та ігристому. У міцних напоях на першому місці горілка (33%), на другому місці коньяк (25%), на третьому - ром та віскі (10%), а також джин і текіла (6%). Також у рейтингу є сидр має (11%) споживання, а після сидру місця посідають слабоалкогольні напої (11%) [7].

Приблизно 82% українців вживають алкогольні напої, проте 64% вживає алкоголь декілька разів на місяць або рідше. Водночас регулярно або доволі часто це роблять лише 18% (рис.1.1). Для 16% респондентів вживання алкоголю є важливим способом розслабитись після роботи. Десь 28% опитаних кажуть, що вживання алкоголю після початку повномасштабної війни збільшилось. Приблизно 26% кажуть, що їх вживання алкоголю стало більше, ніж до війни [5]. Чоловіки вживають алкоголь набагато частіше ніж жінки. Серед чоловіків більше тих, хто вживає алкоголь регулярно або на регулярній основі. Люди від 40 років вживають алкоголь частіше, але не на регулярній основі. А ось регулярне вживання більш характерно для молоді від 18 років до 40 років [5].



Рисунок 1.1 – Частота вживання алкогольних напоїв

Вживання алкогольних напоїв у різних областях України демонструє деякі відмінності. Наприклад, у Києві спостерігається більша кількість людей,

які вживають алкоголь. Однак, у цілому, частота споживання алкоголю не має значних відмінностей між різними регіонами [6]. Споживання алкоголю майже не впливає місце проживання. Найвищий рівень споживання алкоголю фіксують на Західній Україні. Друге місце за споживання алкоголю – Київ. У столиці найбільше п'ють пиво, сидр, шампанське, а на останньому місці міцний алкоголь. Третє місце займає північний регіон, де найбільше віддають перевагу горілці та коньяку. Середній вік людей, які вживають алкоголь в Україні, складає 35 років [7].

У ході соціологічного опитування українців запитали про їхнє ставлення до алкоголю. З респондентів 54% вважали, що існують безпечні дози алкоголю, які не завдають шкоди здоров'ю, тоді як 45% висловили протилежну думку. Кожен третій опитаний вважав вино менш шкідливим алкогольним напоєм, кожен п'ятий — домашні вина. Горілку менш шкідливим напоєм вважали лише 10% респондентів, пиво — 8%, а коньяк і самогон — 6-7%. [11].

Згідно з даними США, алкогольні вподобання мають соціально-економічний аспект. Зокрема, американці з вищою освітою та високими доходами часто обирають вино. Натомість люди з нижчим рівнем освіти та доходів частіше віддають перевагу пиву, порівняно з більш освіченими категоріями населення [10].

Таблиця 1.1. Популярність напоїв в США [10].

	Лікер, %	Вино, %	Пиво, %
Стать			
Чоловік	26	15	55
Жінка	32	45	21
Вік			
Вік від 18 до 34 років	39	16	41

Продовження таблиці 1.1. Популярність напоїв в США [10].

Вік від 35 до 54 років	32	24	41
Вік 55 років і старше	20	44	34
Місце проживання			
Схід	24	32	40
Середній Захід	27	23	49
Південь	34	30	32
Захід	29	33	36
Навчання			
Аспірантура	22	48	28
Тільки випускник коледжу	25	34	39
Навчався в коледжі	32	30	34
Не навчався в коледжі	33	16	47
Дохід			
100 000 \$ + сімейний дохід	25	35	36
\$40 000-100 000	31	27	40
Менше 40 000 доларів	33	23	43
Всього	29	30	38

Алкоголь є самим популярним напоєм у світі. Незважаючи на те, що пиво не є таким домінуючим, як це було чверть століття тому воно залишається найпопулярнішим із трьох основних видів алкоголю. Чоловіки часто віддають перевагу пиву і воно входить у топ-3 споживання алкоголю по всьому світу.

Приблизно семеро з 10 американців, які вживають алкоголь, кажуть, що останній раз випили протягом останнього тижня, 30% кажуть, що останній раз пили більше тижня тому [10].

Таким чином, можна констатувати що незважаючи на місце проживання, люди вживають алкоголь, а також мають подібні тенденції щодо смаків. Для кращого розуміння впливу культур і місця проживання проведемо аналіз статистичних даних про вживання алкогольних напоїв.

1.2 Огляд досліджень щодо продажу алкогольних напоїв в Україні

В Україні завдяки факторам зовнішнього і внутрішнього середовища ринок алкогольних напоїв є одним з пріоритетних, а вітчизняні підприємства нарощують його виробництво й виходять на міжнародні ринки [12].

Таблиця 1.2. Обсяги виробництва алкогольних напоїв в Україні [12].

Вид напою	Одиниця вимірювання	2017	2018	2019	2020	2021
Бренді	Млн. літрів	9,7	9,0	9,3	7,6	-
Горілка		54,4	50,9	47,7	50,7	51,3
Лікери		7,5	7,9	8,7	8,9	11,2
Слабоалкогольні напої		3,8	5,0	5,8	6,2	-

Продовження таблиці 1.2. Обсяги виробництва алкогольних напоїв в Україні [12].

Вино ігристе	Млн. літрів	3,7	3,4	2,7	2,3	2,2
Вино тихе		7,5	8,0	7,4	6,8	6,8
Напої ферментовані		2,8	3,1	3,7	4,8	5,0
Вина		1,6	1,3	1,4	0,5	0,5
Пиво		180	182	182	181	172

Аналізуючи дані, можна побачити, що у більшості видів алкогольних напоїв, до війни, спостерігались незначні коливання обсягів виробництва. Окрім виробництва ферментованих напоїв, слабоалкогольних напоїв та лікерів [15].

Таблиця 1.3. Частка продажу підприємствами роздрібною торгівлі алкогольних напоїв вітчизняного виробництва [15].

Алкогольні напої	2017	2018	2019	2020	2021
Алкогольні напої загалом	77.5	75.3	73.4	71.1	67.7
Горілка	71.4	67.9	65.8	63.3	57.9
Слабоалкогольні напої	93.8	94.1	94.3	95.0	91.4
Вина	62.1	57.6	52.1	47.8	46.1
Коньяк	79.9	77.1	65.6	66.2	63.2
Вина ігристі	74.7	69.1	60.3	56.9	55.4
Пиво	89.0	88.8	91.4	90.1	88.0

Широкий асортимент алкогольних напоїв сприяє задоволенню будь-якого уподобання споживача, але це і має певний негативний ефект. Із-за

великої кількості різноманіття це ускладнює вибір через відмінності класифікацій і групування за якістю та маркуванням алкогольних напоїв. Це в свою чергу може ускладнювати процес купівлі та призводити до плутанини, особливо для нових клієнтів.

Розглянемо переваги і недоліки виробництва і споживання алкоголю. До переваг можна віднести, що алкогольні напої мають стабільно високу частку на внутрішньому ринку і свідчить про лояльність споживачів до вітчизняних товарів. Це в свою чергу стимулює розвиток національного виробництва та покращення якості товарів. Український ринок алкогольних напоїв представлений великою кількістю категорій, що є конкурентною перевагою для місцевих виробників, допомагаючи їм виходити на міжнародні ринки. Українські підприємства активно нарощують виробництво для експорту, що може означати, що їх конкурентоспроможність та перспективи зростають на глобальному ринку.

Серед недоліків спостерігається поступове зниження продажів деяких видів алкогольних напоїв, а саме горілки, ігристих вин та коньяків. Дана тенденція свідчить про зростання конкуренції з імпортними товарами, зависока ціна або зміни у споживчих уподобаннях. Популярність українських алкогольних напоїв може страждати від коливань у світовій економіці, валютних ризиків і торгових бар'єрів. Це створює додаткові виклики для стабільного зростання експорту алкогольних напоїв українського виробництва.

Загалом, ринок алкогольних напоїв має перспективи для розвитку. Це в свою чергу збільшить частку експорту та завдяки високій якості продукції, різноманіттю товарів і лояльності споживачів зможе конкурувати на ринку алкогольних напоїв у всьому світі.

1.3 Аналіз сучасних методів та підходів до обробки інформації

Аналіз даних — це структурована послідовність кроків, які ведуть від необроблених даних до практичних ідей. Аналіз даних перевіряє, очищає,

перетворює та моделює дані, щоб витягнути інформацію та сприяти прийняттю рішень [8].

Аналіз відгуків споживачів здійснюється за допомогою методів обробки природної мови, які допомагають досліджувати відгуки клієнтів, коментарі в соціальних мережах або враження покупців. Такий підхід дає змогу виявляти позитивні та негативні аспекти, що сприяє покращенню продукції та якості обслуговування.

Тональний аналіз - визначає загальне ставлення до товару, що може бути корисним для визначення настроїв ринку та адаптації маркетингової стратегії [62].

Порівняємо методи обробки природної мови відгуків споживачів та тональний аналіз.

- аналіз відгуків — фокусується на оцінці продуктів або послуг споживачами;
- тональний аналіз — використовується для оцінки емоційного тону в текстах, наприклад, відгуків або повідомлень у соціальних мережах.

Таблиця 1.4. Переваги і недоліки аналіз відгуків споживачів та тональний аналіз.

Метод	Переваги	Недоліки
Аналіз відгуків споживачів	Дає можливість зрозуміти думки та настрої споживачів, а також легко впровадити з існуючими даними відгуків.	Не може точно інтерпретувати емоції, іронію, а також потребує очищення даних від шуму та пропусків.
Тональний аналіз	Виявлення емоцій та настроїв в текстах, а також підходить для обробки відгуків, соцмереж і форумів.	Складно працювати з сарказмом, а також залежить від точності алгоритмів і їх налаштування.

Автоматизоване ухвалення рішень здійснюється завдяки застосуванню алгоритмів штучного інтелекту та глибокого навчання, що дозволяє автоматично управляти акціями, встановлювати ціни та передбачати попит [64].

Роботизація маркетингових операцій - автоматизація маркетингових кампаній, яка ґрунтується на вивченні поведінки споживачів, аналізу їх уподобань і прогнозування реакцій на різні акційні пропозиції [65].

Порівняємо методи автоматизації процесів таких як, прийняття рішень та роботизація маркетингових операцій:

- автоматизоване прийняття рішень — дозволяє швидко приймати бізнес-рішення на основі алгоритмів;
- роботизація маркетингових операцій — автоматизація маркетингових задач, таких як обробка замовлень чи сегментація клієнтів.

Таблиця 1.5. Переваги і недоліки автоматизоване прийняття рішень та роботизація маркетингових операцій

Метод	Переваги	Недоліки
Автоматизоване прийняття рішень	Зменшує час прийняття рішень, а також підвищує ефективність.	Складний для налаштування, а також залежить від точності вхідних даних і алгоритмів.
Роботизація маркетингових операцій	Автоматизація рутинних завдань, а також знижує витрати та покращує точність виконання задач.	Дорогий та вимагає постійного моніторингу і налаштування.

Таким чином, перевагами сучасних методів обробки даних є:

- аналіз даних проводиться у реальному часі;
- методи машинного навчання та глибинного навчання підвищують точність прогнозування попиту;

- сегментація клієнтів дозволяє створювати персоналізовані маркетингові стратегії.

Недоліками та викликами сучасних методів обробки даних є:

- аналіз великих даних потребує потужної інфраструктури;
- робота з методами машинного навчання та великими даними потребує спеціалізованих знань;

- збір та аналіз персональних даних вимагає відповідності вимогам конфіденційності.

Використання сучасних методів дає можливість компаніям глибше розуміти потреби клієнтів. Це в свою чергу допомагає краще прогнозувати попит і приймати обґрунтовані бізнес-рішення. Це сприяє підвищенню конкурентоспроможності на ринку.

1.4 Висновок до першого розділу

Аналіз досліджень які відображають тенденції споживання, виробничі обсяги та особливості продажу алкогольних напоїв дозволили зробити наступні висновки, що зростання попиту на алкогольні напої у багатьох країнах свідчить про зміну споживацьких звичок, де особливу увагу привертають напої преміум-сегменту та інноваційні продукти, включаючи безалкогольні альтернативи та змішані коктейлі. Це свідчить про те, що сучасні споживачі шукають нові смаки і варіанти для свого дозвілля, зокрема, все більше уваги приділяється крафтовим та органічним напоям.

В Україні ситуація на ринку алкогольних напоїв формується під впливом кількох ключових факторів: економічної ситуації, змін у законодавстві, а також зовнішньоекономічних відносин. Дефіцит та обмеження імпорту часто призводять до зміщення акцентів у виробництві, сприяючи розвитку локальних виробників та зміцненню позицій національних брендів. У той же час важливо відзначити, що еволюція смакових уподобань споживачів також вимагає від виробників креативного підходу та адаптації до глобальних трендів. Наприклад, екологічна стійкість, використання біоресурсів та

впровадження інноваційних технологій у виробництві стають необхідними для того, щоб відповідати світовим стандартам.

Щодо особливостей продажу, важливо звернути увагу на зміни у форматі споживання: зростання онлайн-продажів і розвиток каналу доставки є важливими чинниками, що змінили традиційний процес купівлі. Пандемія COVID-19 лише підштовхнула цю тенденцію, оскільки споживачі все частіше звертаються до зручних форм покупки через інтернет. Це, у свою чергу, змінює і самі маркетингові стратегії компаній, які намагаються інтегрувати цифрові технології у свої бізнес-моделі.

Отже, ринок алкогольних напоїв, зокрема в Україні, демонструє активний розвиток, який має адаптуватися до глобальних змін і впроваджувати інновації, аби залишатися конкурентоспроможним. Важливо враховувати економічні, соціальні та культурні аспекти, які визначають переваги споживачів, та орієнтуватися на сталі практики й розвиток внутрішнього виробництва, що дозволить зміцнити позиції національних брендів та зробити ринок більш стійким у разі зовнішніх змін.

РОЗДІЛ 2. ОБҐРУНТУВАННЯ ВИБОРУ НАПРЯМКУ ДОСЛІДЖЕНЬ І МЕТОДІВ ВИРІШЕННЯ ЗАДАЧ

Для розробки системи, яка буде здатна ефективно обробляти великий обсяг даних і надавати аналітичні висновки щодо продажів алкогольних напоїв необхідно обрати відповідні методи дослідження і інструменти реалізації. Одним із найбільш динамічних сегментів споживчого ринку є ринок алкогольних напоїв. За даними світових статистичних агентств, продажі алкогольних напоїв у світі постійно зростають, зокрема в розвинутих країнах та нових економіках. Для ефективного управління ринком алкогольних напоїв потрібна точна і своєчасна аналітика. Це дозволить прогнозувати тренди, адаптувати маркетингові стратегії та оптимізувати канали збуту, а також забезпечить всебічний аналіз продажів алкогольних напоїв з метою оптимізації асортименту, ціноутворення та управління запасами.

Сучасні аналітичні інструменти, зокрема системи для збору, обробки та аналізу даних про продажі, є необхідними для отримання конкурентних переваг на ринку алкогольних напоїв. Однак наявні аналітичні системи часто не охоплюють усі аспекти: від збору даних до прогнозування, або їх інформація є локалізованою для окремих ринків, що обмежує можливості глобальної аналітики.

Ключовими завданнями цього дослідження є аналіз даних щодо продажів алкогольних напоїв, виявлення закономірностей і трендів, а також прогнозування майбутніх показників. Для багатьох компаній, що працюють на глобальному ринку алкогольних напоїв, відсутність єдиної системи, яка б надавала можливість аналізувати продажі в реальному часі, прогнозувати попит та зберігати всі необхідні дані в єдиному сховищі, є суттєвим бар'єром для оперативного прийняття рішень. Багато компаній використовують різноманітні інструменти для збору та обробки даних, але вони не інтегровані в єдину платформу, що створює труднощі для комплексного аналізу.

2.1 Етапи розробки інформаційної системи

Основні етапи дослідження в процесі створення інформаційної системи для аналізу продажів алкогольних напоїв охоплюють кілька ключових факторів, кожен з яких спрямований на досягнення глибокого розуміння даних і створення ефективних інструментів підтримки бізнес-рішень.

Початковим етапом є збір і інтеграція даних. Тут необхідно розробити системи для отримання інформації з різноманітних джерел — від касових апаратів і CRM-систем до звітів про роздрібні продажі та маркетингові кампанії. Усі дані повинні об'єднуватися у єдину базу з урахуванням достовірності та відповідності структури для подальшого використання.

Наступний важливий етап — очищення і попередня обробка. Він включає видалення дублікатів, виправлення помилок і заповнення пропусків, що забезпечить високу якість даних. Паралельно виконується стандартизація форматів і категорій, що полегшить подальший аналіз і підвищить точність результатів.

Ключовим завданням є визначення показників ефективності, які допоможуть оцінити рівень продажів і споживчих уподобань. Визначення середнього чеку, обороту та частоти покупок дозволить глибше зрозуміти динаміку продажів. Також важливо створити метрики для порівняння результатів у різних часових проміжках і сегментах ринку.

Аналіз закономірностей і трендів займає значну частину роботи і передбачає вивчення впливу сезонності, святкових періодів і зовнішніх обставин на продажі. Кластеризація клієнтів за купівельними звичками відкриває можливість сегментації і цільового маркетингу.

Прогнозування є важливим інструментом для планування майбутніх продажів. Побудова моделей, які враховують поточні тренди, попит і зміни на ринку, сприяє створенню точних прогнозів. Впровадження алгоритмів машинного навчання може допомогти в оцінці ризиків і виявленні невидимих можливостей для бізнесу.

Завершальним етапом є створення системи візуалізації даних та звітності. Розробка інтерактивних масивів даних дозволяє користувачам легко розуміти і використовувати результати аналізу для швидкого ухвалення рішень. Автоматизовані звіти спрощують процес відстеження результатів і контролю за ефективністю бізнес-процесів. Основні задачі дослідження в процесі створення аналітичної системи для аналізу продажів алкогольних напоїв охоплюють кілька ключових аспектів, кожен з яких спрямований на досягнення глибокого розуміння даних і створення ефективних інструментів підтримки бізнес-рішень.

2.2 Дослідження методів для вирішення задач

Для досягнення поставлених цілей у створенні аналітичної системи для дослідження продажів алкогольних напоїв пропонується використання низки методів. Серед них статистичний аналіз даних, що включає розрахунок середніх обсягів продажів, середніх цін та розподілу продажів за регіонами, часом і іншими важливими факторами і дозволяє отримати детальну картину поведінки споживачів та динаміки ринку.

Автоматичне усунення дублікатів і заповнення пропусків у даних стає можливим завдяки застосуванню спеціалізованих алгоритмів, що покращують якість інформації. Цей етап забезпечує точність подальших розрахунків і дозволяє уникнути спотворення результатів.

Аналіз часових рядів є важливим для дослідження сезонних коливань і виявлення довгострокових трендів у продажах. Це допомагає виявити повторювані шаблони у поведінці покупців, що є ключовим для планування стратегії бізнесу. Інтеграція штучного інтелекту, зокрема, створення рекомендаційних систем, надає можливість персоналізації маркетингових кампаній і покращення взаємодії з клієнтами.

Для виявлення взаємозв'язків між ціною алкогольних напоїв і обсягами продажів, а також для дослідження відношень між різними категоріями напоїв, застосовується кореляційний аналіз. Це дозволяє виявити, як зміни в ціновій

політиці впливають на обсяги продажів, і сприяє формуванню оптимальної цінової стратегії.

Методи машинного навчання, такі як лінійна регресія і дерева рішень, використовуються для прогнозування майбутніх продажів, дозволяють враховувати складні залежності і впливи, що виникають у процесі продажів, а також адаптувати прогнози до змін на ринку.

Для виконання аналітичних завдань пропонуються наступні інструменти:

1. ClickHouse - це високопродуктивна, орієнтована на колонки система управління базами даних SQL (СУБД) для онлайн-аналітичної обробки (OLAP). Вона доступна як програмне забезпечення з відкритим кодом, так і як хмарна пропозиція [82].

2. Grafana - програмне забезпечення з відкритим вихідним кодом дозволяє запитувати, візуалізувати, сповіщати та досліджувати показники, журнали та трасування, де б вони не зберігалися. Інструмент для візуалізації даних та моніторингу, який інтегрується з різними джерелами даних, включаючи ClickHouse [83].

3. VirtualBox є одним з найвідоміших міжплатформених інструментів для віртуалізації з відкритим вихідним кодом. Він дозволяє розробникам ефективніше розгортати код, запускаючи різні операційні системи на одному пристрої. IT-команди та постачальники рішень використовують VirtualBox для зниження витрат на операційну діяльність і скорочення часу, необхідного для безпечного розгортання додатків як локально, так і в хмарі [84].

4. Docker - це інструмент, який використовується для автоматизації розгортання програм у легких контейнерах для їх ефективної роботи у різних середовищах ізольовано [85].

5. Python - універсальна мова програмування високого рівня, яка зазвичай інтерпретується [86]. Вона спрямована на просування легкого для

читання, лаконічного стилю програмування та написання скриптів для очищення даних.

2.3 Обґрунтування вибору інструментів реалізації

Для реалізації аналітичної інформаційної системи обрано наведені у таблицях 2.1 – 2.3 інструменти, оскільки вони найкраще відповідають поставленим завданням.

Розглянемо порівняння програм для контейнеризації Docker та Podman, наведене у таблиці 2.1.

Таблиця 2.1 Порівняння програм для контейнеризації Docker та Podman.

Параметр	Docker	Podman
Архітектура	Працює через демон, що керує контейнерами	Без демона, працює як окремі процеси
Інсталяція	Простий інстальатор із GUI-інструментами	Потребує налаштування
Сумісність із Docker	Використовує власний формат образів, але підтримує OCI	Повністю сумісний із Docker-образами
Оркестрація	Підтримує Docker Compose, інтеграція з Kubernetes	Оркестрація через Kubernetes

Програма для контейнеризації Docker було обрана завдяки тому що вона має:

- великий вибір готових образів на Docker Hub;
- детальнішу документацію та ширшу підтримку сторонніх інструментів;
- інтуїтивно зрозумілий інтерфейс.

Розглянемо порівняння програм для к для візуалізації даних Grafana та Power BI, наведене у таблиці 2.2.

Таблиця 2.2 Порівняння програм для візуалізації даних Grafana та Power BI.

Критерії	Grafana	Power BI
Призначення	Моніторинг та візуалізація метрик у реальному часі	Багатофункціональна аналітика
Інтеграції	Підтримка багатьох джерел даних (ClickHouse)	Інтеграція переважно з продуктами Microsoft
Продуктивність	Високошвидкісний моніторинг великих наборів даних	Може бути повільною для великих наборів даних
Вартість	Відкрите програмне забезпечення, безкоштовне	Платна підписка або ліцензія
Гнучкість	Просте налаштування кастомних дашбордів	Потребує більше зусиль для кастомізації

Програма для візуалізації даних Grafana була обрана завдяки тому що вона є:

- безкоштовною та підтримує багато різних джерел даних;
- більш адаптована для динамічного моніторингу метрик;
- забезпечує швидке підключення до систем для DevOps, ClickHouse.

Розглянемо порівняння програм для роботи з даними ClickHouse та PostgreSQL, наведене у таблиці 2.3.

Таблиця 2.3 Порівняння програм для роботи з даними ClickHouse та PostgreSQL.

Критерії	ClickHouse	PostgreSQL
Призначення	Аналітична база даних	Універсальна реляційна база даних
Продуктивність	Оптимізований для обробки великих обсягів даних	Краще підходить для транзакційних операцій

Продовження таблиці 2.3 Порівняння програм для роботи з даними
ClickHouse та PostgreSQL.

Швидкість	Надзвичайно швидке виконання запитів	Менш ефективний для великих аналітичних задач
Масштабованість	Легко обробляє петабайти даних	Обмежена масштабованість для великих даних
Вимоги до ресурсів	Мінімальні для аналітики	Вимагає більше ресурсів для складних задач

Програма для роботи з даними ClickHouse була обрана завдяки тому що вона:

- спрямована безпосередньо на роботу з аналітичними запитами;
- зменшує обсяг дискових операцій;
- призначена для обробки великих наборів даних, наприклад, аналізу логів чи продажів.

2.4 Висновок до другого розділу

Проведені у другому розділі дослідження демонструють важливість комплексного підходу до створення аналітичної системи для дослідження продажів алкогольних напоїв. Для забезпечення ефективної роботи системи необхідно враховувати етапи збору, підготовки та обробки даних, очищення та трансформацію даних для досягнення високої якості інформації. Застосування методів статистичного аналізу та побудова прогнозних моделей, зокрема за допомогою машинного навчання, допомагає отримати глибоке розуміння ринкових тенденцій і змінювань у споживчій поведінці.

Загалом інформаційна аналітична система призначена для виробників алкогольних напоїв, які можуть відстежувати та аналізувати ринкові тренди та тенденції, що сприятиме актуалізації програм виробництва алкогольної продукції та збільшенню прибутків.

В процесі розробки ІС важливим є використання сучасних мов програмування та інструментів, таких як Python для автоматизації процесів аналізу і очищення даних, СУБД ClickHouse для обробки великих даних, Grafana для візуалізації інформації. Використання цих інструментів дозволяє створити інтегровану платформу, яка не тільки обробляє дані, але й надає зручний інтерфейс для перегляду та інтерпретації результатів, що є основою для прийняття обґрунтованих бізнес-рішень.

Дослідження підкреслює необхідність адаптації обраних методів і технологій відповідно до конкретних умов і цілей, що забезпечує гнучкість та ефективність системи. Завдяки інтегрованому підходу, використанню сучасних методів обробки даних та прогнозуванню, харчові підприємства отримують можливість адаптувати свої стратегії, оптимізувати процеси і покращити взаємодію з клієнтами, що в цілому підвищує їх конкурентоспроможність на глобальному ринку алкогольних напоїв. Також системою можуть користуватися ритейли для аналізу популярності товарів, оцінки ефективності цінової політики та управління запасами магазину, що сприятиме врахуванням тенденцій сезонності асортименту для різних регіонів.

РОЗДІЛ 3. РЕАЛІЗАЦІЯ ОСНОВНИХ ЗАВДАНЬ АНАЛІЗУ ПРОДАЖУ АЛКОГОЛЬНИХ НАПОЇВ

Реалізація основних завдань передбачає виконання декількох етапів наукового дослідження, яке включає аналіз та обробку даних продажів алкогольних напоїв. Основна мета даного розділу полягає в детальному описі реалізації методів і процесів, що використовуються для досягнення цілей дослідження.

3.1 Формування переліку задач дослідження

На основі досліджень, наведених у попередніх розділах, та огляді джерел даних з інформацією про продажі алкогольної продукції було виділено перелік задач, що наведені у таблиці 3.1, які потребують візуалізації та аналізу.

Таблиця 3.1 Перелік задач для аналізу продажу алкогольної продукції

№	Задача	Засоби аналізу та представлення даних	Результат аналізу
1	Визначення найпопулярніших алкогольних напоїв за кількістю проданих пляшок	- SQL-запити та агрегація даних. - Візуалізація даних	Топ-10 найпопулярніших напоїв за кількістю проданих пляшок
2	Визначення мереж магазинів з найбільшим обсягом продажів алкогольних напоїв	- SQL-запити та агрегація даних. - Візуалізація даних.	Топ-10 мереж магазинів з найбільшими продажами алкоголю
3	Визначення місяців з найбільшим та найменшим прибутком від продажів	- SQL-запити та агрегація даних. - Аналіз часових рядів. - Візуалізація даних.	Статистика продажів продукції з найвищими та найнижчими прибутками в розрізі місяців. Дослідження сезонних коливань
4	Визначення обсягів продажів алкогольної продукції у святкові дні	- SQL-запити та агрегація даних. - Data Mining. - Аналіз часових рядів. - Візуалізація даних.	Статистика попиту на алкогольні напої у святкові дні

Продовження таблиці 3.1 Перелік задач для аналізу продажу алкогольної продукції

5	Визначення попиту на алкогольну продукцію	- SQL-запити та агрегація даних. - Візуалізація даних	Статистика продажу алкогольних напоїв
6	Визначення кількості проданих пляшок за одну транзакцію	- SQL-запити та агрегація даних. - Статистичний аналіз.	Статистика проданих пляшок за одну транзакцію
7	Визначення загальної суми продажів окремо для кожного дня тижня	- SQL-запити та агрегація даних. - Аналіз часових рядів. - Візуалізація даних.	Статистика загальної суми продажів окремо для кожного дня тижня
8	Визначення загального обсягу проданого алкоголю в літрах по роках	- SQL-запити та агрегація даних. - Аналіз часових рядів. - Візуалізація даних	Статистика продажів алкоголю в літрах по роках
9	Визначення категорій алкогольних напоїв з найвищою середньою роздрібною ціною пляшки	- SQL-запити та агрегація даних. - Статистичний аналіз. - Візуалізація даних	Статистика продажів категорій алкогольних напоїв
10	Визначення днів тижня з найбільшими продажами алкогольних напоїв	- SQL-запити та агрегація даних. - Аналіз часових рядів. - Візуалізація даних	Статистика продажів алкогольної продукції по днях тижня
11	Визначення середньої собівартості та роздрібною ціни пляшки алкогольного напою	- SQL-запити та агрегація даних. - Статистичний аналіз. - Візуалізація даних	Статистика середніх цін та собівартості пляшки алкогольного напою
12	Визначення продажу алкоголю по містах обраної країни	- SQL-запити та агрегація даних. - Візуалізація даних. - Геопросторовий аналіз.	Перелік міст кожної країни з найбільшим обсягом продажів

Продовження таблиці 3.1 Перелік задач для аналізу продажу алкогольної продукції

13	Визначення тенденцій продажів алкогольних напоїв по місяцях кожного року	- SQL-запити та агрегація даних. - Аналіз часових рядів. - Візуалізація даних	Тенденції і сезонність продажів по місяцях кожного року
14	Визначення прибутків від продажів по роках	- SQL-запити та агрегація даних. - Аналіз часових рядів. - Візуалізація даних.	Динаміка ринку продажів алкогольної продукції
15	Визначення рейтингу постачальників алкогольної продукції за обсягами продажів алкогольної продукції	- SQL-запити та агрегація даних. - Візуалізація даних.	Рейтинг постачальників алкогольної продукції

3.2 Збір масивів даних, необхідних для реалізації системи

Розглянемо етапи дослідження та розробки інформаційної системи:

- дані з набору продажів алкогольних напоїв, отримані сайту з системи Kaggle. Підготовка даних до аналізу потребує очистки. Очищення даних передбачало видалення дублікатів, обробку пропущених значень та перетворення форматів дат і числових значень для подальшого аналізу;
- статистичний аналіз даних виконаний з використанням бібліотеки Pandas. Методи аналізу включали побудову гістограм для дослідження розподілу продажів та часових рядів для виявлення сезонних тенденцій;
- модель прогнозування продажів будувалася на основі методів дослідження регресії. На даному етапі дослідження були випробувані кілька методів машинного навчання для визначення найкращого підходу до прогнозування.

Протягом реалізації проекту було зібрано великий обсяг матеріалів, які стали основою для магістерської роботи. Основні масиви даних, зібрані для аналізу, включають:

- обсяги продажів алкогольних напоїв за регіонами штату;
- показники продажів для різних категорій алкогольних напоїв;
- динаміка продажів у залежності від сезону, регіону.

Наведені матеріали були використані для побудови аналітичної системи, а також для візуалізації даних і представлення результатів у вигляді діаграм і графіків.

3.3 Компоненти інформаційної системи та їх налаштування

Система має багато компонентів. Вона запускається через VirtualBox, піднімається Docker контейнер після цього база даних зчитується до СУБД ClickHouse, а візуалізація здійснюється у засобі Grafana.

3.3.1 Налаштування віртуальної машини

Налаштування віртуальної машини для роботи з Docker і пов'язаними компонентами зберігається у Vagrant-файл.

Код Vagrant-файлу наведено нижче.

```
Vagrant.configure("2") do |config|
  config.vm.hostname = "certificate.vm"
  config.vm.define "certificate"
  config.vm.box = "ubuntu/jammy64"
  config.vm.provider "virtualbox" do |vb|
    vb.name = "certificate"
    vb.memory = 4096
    vb.cpus = 2
  end
  config.vm.network "forwarded_port", guest: 3000, host: 3000
```

```

config.vm.synced_folder ".", "/vagrant", owner: "vagrant", group: "vagrant"
config.vm.provision "shell", inline: <<-SHELL
sudo apt-get update && sudo apt-get upgrade -y
curl -fsSL https://get.docker.com -o get-docker.sh
sh get-docker.sh
sudo usermod -aG docker vagrant
sudo apt-get install -y docker-compose-plugin python3 python3-dev
python3-pip
pip install clickhouse_connect pandas
sudo systemctl enable docker
sudo systemctl start docker
cd /vagrant
docker compose down
docker compose build
docker compose up -d
SHELL
End

```

Наведений код автоматично підготує робоче середовище для роботи з Docker і мовою Python. Додає простий доступ до програми через порт 3000 та синхронізує файл між хостом і VM.

3.3.2 Створення бази даних і таблиць у ClickHouse

Для створення таблиць у базі даних потрібно виконати програмний код наведений нижче.

```

CREATE DATABASE IF NOT EXISTS certificate;
CREATE TABLE certificate.liquor_sales (
  `Invoice/Item Number` String,
  `Date` Date,
  `Store Number` Int32,
  `Store Name` String,

```

```

`Address` String,
`City` String,
`Zip Code` String,
`Store Location` String,
`County Number` Int32,
`County` String,
`Category` Int32,
`Category Name` String,
`Vendor Number` Int32,
`Vendor Name` String,
`Item Number` Int32,
`Item Description` String,
`Pack` Int32,
`Bottle Volume (ml)` Int32,
`State Bottle Cost` Float64,
`State Bottle Retail` Float64,
`Bottles Sold` Int32,
`Sale (Dollars)` Float64,
`Volume Sold (Liters)` Float64,
`Volume Sold (Gallons)` Float64
) ENGINE = MergeTree()
ORDER BY `Invoice/Item Number`;

```

Даний код перевіряє чи є база даних certificate, якщо її ще немає то вона створюється; додає таблицю liquor_sales для зберігання даних про продажі алкоголю; налаштовує таблицю на використання рушія MergeTree для ефективного зберігання та обробки даних.

3.3.3 Очищення даних мовою програмування Python

Для очищення даних було зроблено парсинг бази даних за допомогою мови Python в окремому файлі.

Код очищення бази даних наведено нижче.

```
import pandas as pd
from clickhouse_connect import get_client
file_path = './Iowa_Liquor_Sales.csv'
chunk_size = 1000
table_name = 'liquor_sales'
client = get_client(
    host='localhost',
    port='8123',
    username='vagrant',
    password='admin123',
    database='certificate'
)
for chunk in pd.read_csv(file_path, chunksize=chunk_size):
    float_columns = ['State Bottle Cost', 'State Bottle Retail', 'Sale (Dollars)',
'Volume Sold (Liters)', 'Volume Sold (Gallons)']
    for col in float_columns:
        chunk[col] = pd.to_numeric(chunk[col].replace('[,$,]', '', regex=True),
errors='coerce').fillna(0.0)
    int_columns = ['Pack', 'Bottle Volume (ml)', 'Bottles Sold', 'Category',
'Vendor Number', 'Item Number', 'Store Number', 'County Number']
    for col in int_columns:
        chunk[col] = pd.to_numeric(chunk[col], errors='coerce').fillna(0).astype(int)
    string_columns = ['Invoice/Item Number', 'Store Name', 'Address', 'City', 'Zip
Code', 'Store Location', 'County', 'Category Name', 'Vendor Name', 'Item
Description']
    for col in string_columns:
        chunk[col] = chunk[col].astype(str).fillna("")
    chunk['Date'] = pd.to_datetime(chunk['Date'], errors='coerce')
    print(chunk)
```

```
client.insert_df(table_name, chunk)
print("Дані успішно занесені.")
```

Код завантажує великі обсяги даних з CSV у базу даних, обробляючи дані для забезпечення коректності та консистентності. Після цього оптимізує роботу через розбиття даних на частини.

3.3.4 Налаштування бази даних у Grafana

Код для налаштування бази даних у СУБД ClickHouse для Grafana даних наведено нижче.

```
apiVersion: 1
datasources:
- name: ClickHouse Grafana Labs
  type: grafana-clickhouse-datasource
  access: proxy
  jsonData:
  server: clickhouse
  port: 9000
  username: vagrant
  secure: false
  tlsSkipVerify: true
  secureJsonData:
  password: admin123
  isDefault: true
```

Засіб Grafana підключено до ClickHouse через плагін grafana-clickhouse-datasource використовуючи сервер із хостом clickhouse і порт 9000. Це в свою чергу здійснює авторизацію з користувачем vagrant і паролем admin123.

3.3.5 Конфігурація Docker Compose

Наведений код автоматизує створення та запуск двох сервісів: ClickHouse та Grafana.

```
services:
```

clickhouse:

image: clickhouse/clickhouse-server:latest

container_name: clickhouse

environment:

- CLICKHOUSE_USER=vagrant
- CLICKHOUSE_PASSWORD=admin123
- CLICKHOUSE_DB=certificate

volumes:

- ./create-database.sql:/docker-entrypoint-initdb.d/create-database.sql

ports:

- "9000:9000"
- "8123:8123"

restart: unless-stopped

grafana:

image: grafana/grafana:latest

container_name: grafana

ports:

- "3000:3000"

depends_on:

- clickhouse

restart: unless-stopped

environment:

- GF_INSTALL_PLUGINS=grafana-clickhouse-datasource

volumes:

- ./provisioning:/etc/grafana/provisioning
- ./dashboards:/var/lib/grafana/dashboards

Спочатку запускається сервер бази даних ClickHouse, використовуючи облікові дані, передані через змінні середовища. Якщо все працює то автоматично створюється база даних certificate. Після цього виконуються SQL-запити з файлу create-database.sql.

В Grafana запускається сервер візуалізації, встановлюється плагін для інтеграції з ClickHouse. Grafana використовує локальні директорії для налаштувань і збереження дашбордів. Таким чином, Grafana налаштована працювати із ClickHouse (залежність `depends_on`), дозволяючи візуалізувати дані з бази.

3.4 Структура даних бази даних

База даних аналітичної системи складається з 12 мільйонів записів на основі інформації з джерела даних, що було завантажено з сайту Kaggle та має структуру, наведену у таблиці 3.2.

Таблиця 3.2. Структура бази даних

Назва поля	Тип даних	Значення
Invoice/Item Number	String	Номер рахунка/товару
Date	DateTime	Дата
Store Number	Int32	Номер магазину
Store Name	String	Назва магазину
Address	String	Адреса
City	String	Місто
Zip Code	String	Поштовий індекс
Store Location	String	Розташування магазину
County Number	Int32	Номер округу
County	String	Округ
Category	Int32	Категорія
Category Name	String	Назва категорії
Vendor Number	Int32	Номер постачальника
Vendor Name	String	Назва постачальника

Продовження таблиці 3.2. Структура бази даних

Item Number	Int32	Номер товару
Item Description	String	Опис товару
Pack	Int32	Кількість у пакуванні
Bottle Volume (ml)	Int32	Об'єм пляшки (мл)
State Bottle Cost	Float32	Собівартість пляшки
State Bottle Retail	Float32	Роздрібна ціна пляшки
Bottles Sold	Int32	Продано пляшок
Sale (Dollars)	Float32	Продаж (долари)
Volume Sold (Liters)	Float32	Проданий об'єм (літри)
Volume Sold (Gallons)	Float32	Проданий об'єм (галони)

Таблиця 3.1 містить дані про продажі алкогольних напоїв, їх типи та структурований опис. Кожне поле містить певний тип даних і конкретне призначення. Масив даних містить зокрема інформацію про продажі, адреси, дати продажів, описи товару, категорії тощо. Всього використовується 24 таблиці з різними типами даних. Інформація, наведена в таблиці є основою для створення аналітичних звітів, побудови прогнозів і прийняття управлінських рішень.

3.5 Розробка візуального представлення даних

Розглянемо процес створення візуального представлення даних у відповідності зі списком задач, наведених у таблиці 3.1

Запит по найпопулярнішим напоям формує топ-10 напоїв по кількості проданих пляшок. Можна сказати, що самим популярним напоєм є Black Velvet (понад 2,5 млн пляшок), а між лідером та рештою товару є певний розрив по обсягах продажів. Це свідчить, що Black Velvet є абсолютним

лідером продажів. Також бачимо багато інших брендів алкоголю, що свідчить про широкий асортимент напоїв.

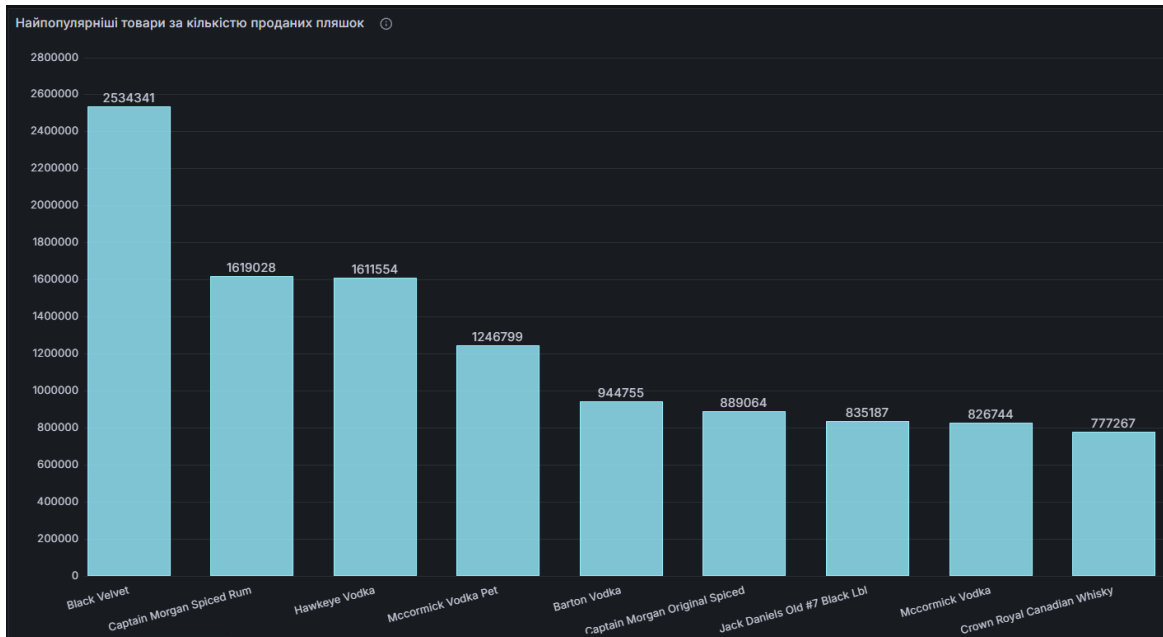


Рисунок 3.1 - Дані про найпопулярніші алкогольні напої

SQL код:

```
SELECT
  `Item Number`,
  `Item Description`,
  SUM(`Bottles Sold`) AS total_bottles_sold
FROM certificate.liquor_sales
GROUP BY `Item Number`, `Item Description`
ORDER BY total_bottles_sold DESC
LIMIT 10;
```

Наступний запит демонструє загальний розподіл доходу між топ 40 містами по сумі продажів у доларах. Найбільші продажі спостерігаються у місті Des Moines (понад 14 мільйонів доларів). Це може означати, що в ньому проживає багато споживачів алкогольної продукції, наявна велика кількість магазинів та споживачі платоспроможні. Отже, дане місто має ключове значення для бізнесу та логістики. Інші міста мають менший обсяг продажу в порівнянні з Des Moines, але є перспективними. Для збільшення продажів

доцільно провести маркетингову компанію та забезпечити доступ до найпопулярніших товарів.

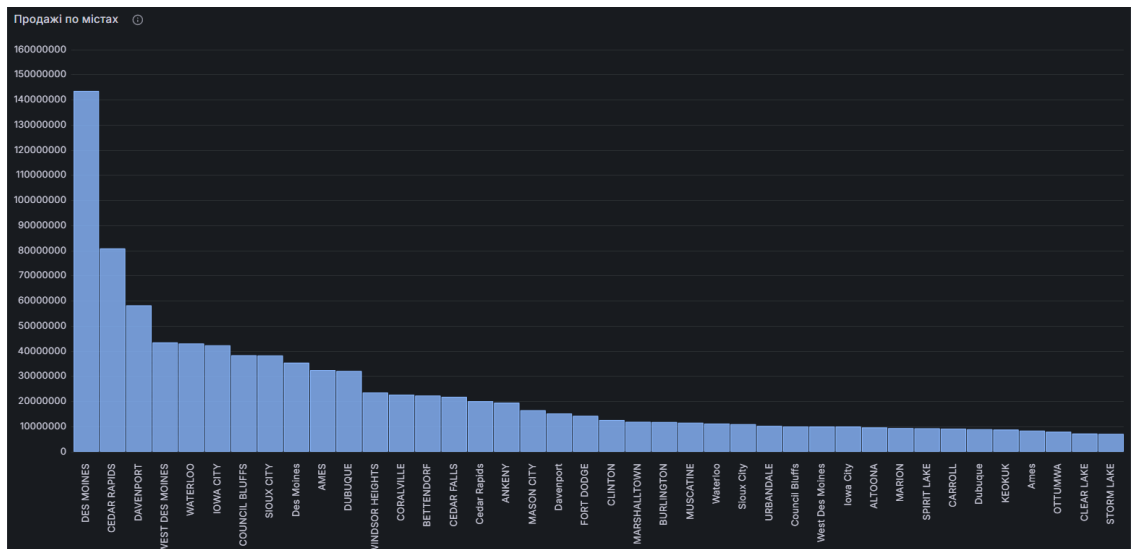


Рисунок 3.2 - Дані про продаж по містах

SQL код:

```
SELECT
```

```
  `City`,
```

```
  SUM(`Sale (Dollars)`) AS total_sales
```

```
FROM certificate.liquor_sales
```

```
GROUP BY `City`
```

```
ORDER BY total_sales DESC
```

```
LIMIT 40;
```

Наступний запит аналізує суму продажів по округах. Тобто підраховує загальний дохід по округам та сортує дані у спадаючому порядку. Округ Polk є лідером по продажам (понад 28 мільйонів доларів). З графіку, наведеному на рисунку 3.3 видно, що з 5-го по 10-тий місьць округи мають більш рівномірний розподіл продажів алкогольних напоїв. Основними ключовими округами за зростанням прибутку є Linn і Scott.

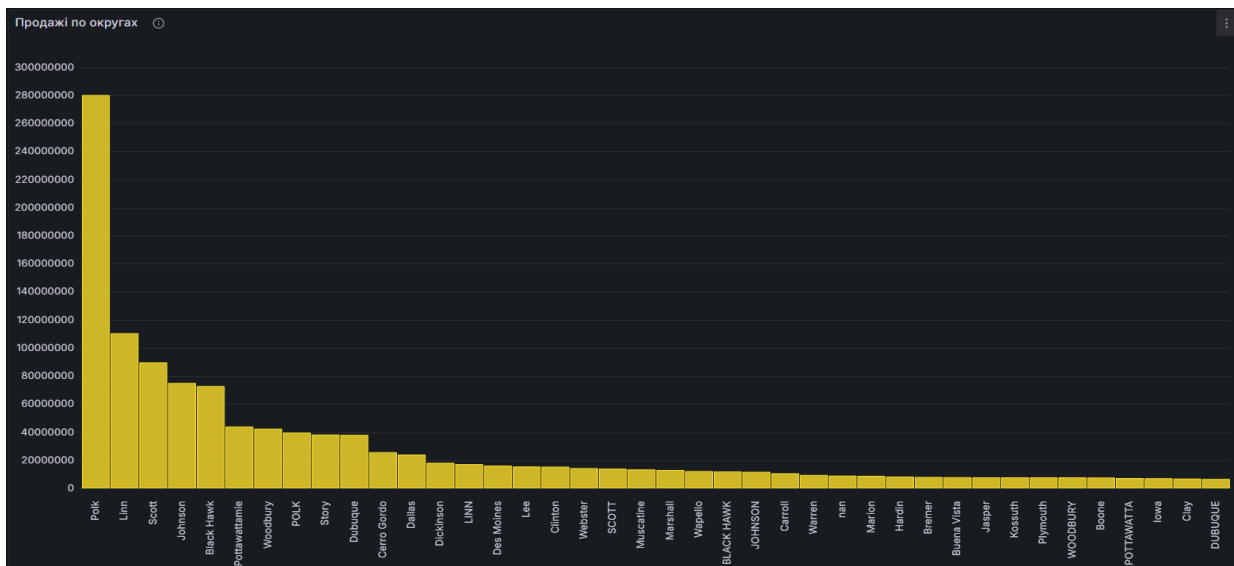


Рисунок 3.3 - Дані про продаж по округам

SQL код:

SELECT

`County`,

SUM(`Sale (Dollars)`) AS total_sales

FROM certificate.liquor_sales

GROUP BY `County`

ORDER BY total_sales DESC

limit40;

Графік, наведений на рисунку 3.4 визначає топ-10 магазинів з найбільшим обсягом продажів в магазинах. Магазин Ну-Vee #3 | Des Moines є топ по продажами алкогольних напоїв, які перевищують 4,7 мільйона доларів. Друге місце посідає Central City 2, але з меншими продажами чим в першому магазині 3,4 мільйона доларів. Магазины з 3-го місця мають менше прибутку та більш рівномірний розподіл доходів.

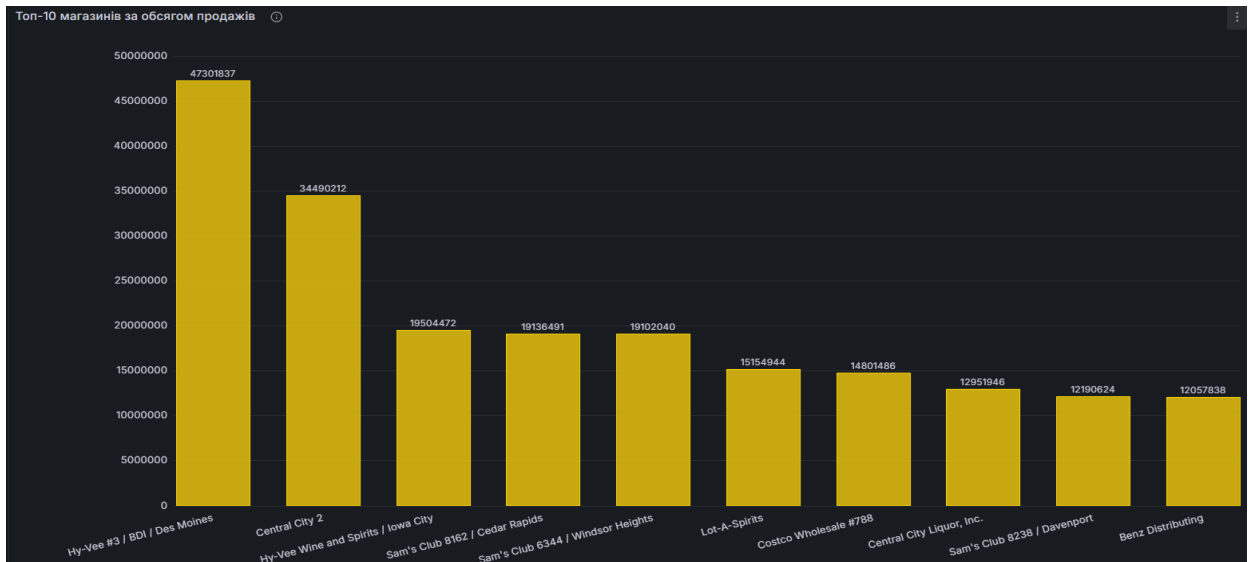


Рисунок 3.4 - Дані про топ-10 магазинів за доходами

SQL код:

SELECT

 `Store Number`,

 `Store Name`,

 SUM(`Sale (Dollars)`) AS total_sales

FROM certificate.liquor_sales

GROUP BY `Store Number`, `Store Name`

ORDER BY total_sales DESC

LIMIT

10;

Графіки, наведені на рисунках 3.5-3.7 мають червоні стовпчики це сума продажів у доларах за кожен місяць (*total_sales*), а помаранчеві - обсяг проданого алкоголю в літрах за кожен місяць. Можна спостерігати, що 10-й місяць (жовтень) має найбільший прибуток по всім параметрам. Це може бути пов'язано з сезонними факторами (тобто святами). Листопад (11-й місяць) та грудень (12-й) також мають високий прибуток. Можливо це зв'язано з підготовкою до зимових свят. Січень (1-й) має найменші продажі алкогольних напоїв, що може свідчити про зниження попиту після свят. Більшість місяців показують майже однакові показники по продажам.

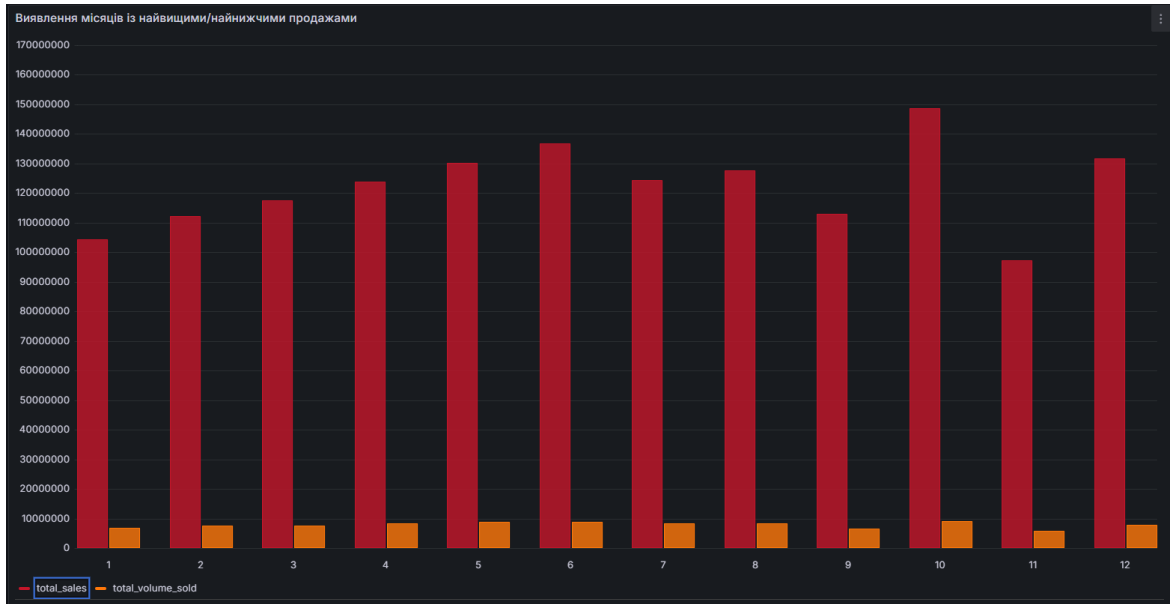


Рисунок 3.5 - Виявлення місяців із найвищими/найнижчими продажами

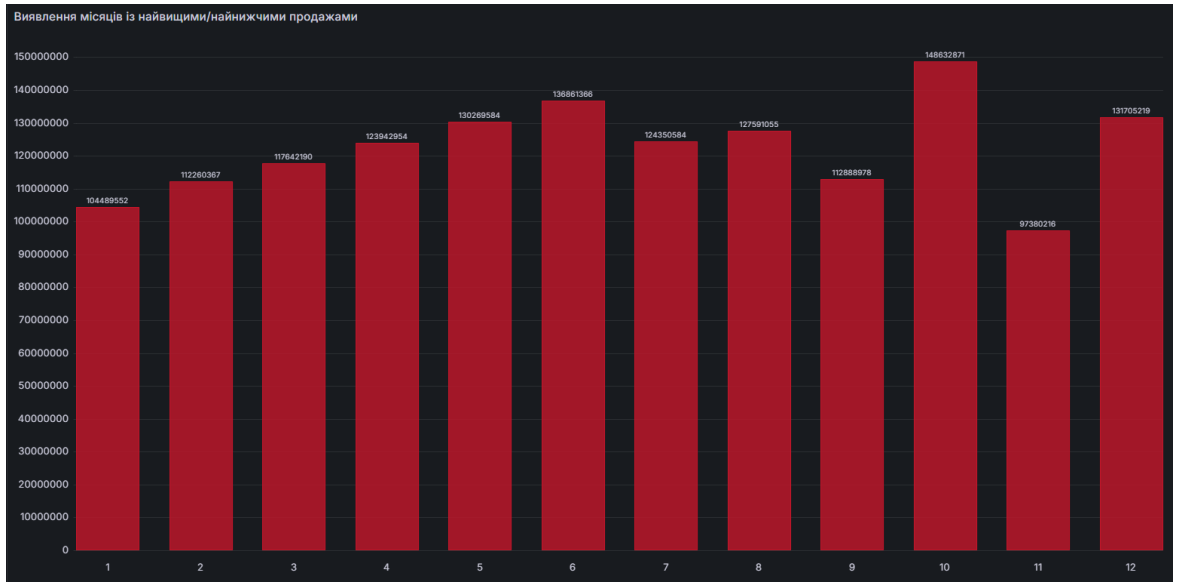


Рисунок 3.6 - Дані про суму продажів у доларах



Рисунок 3.7 - Дані про обсяг проданого алкоголю в літрах

SQL код:

```
SELECT
```

```
  toMonth(`Date`) AS month,
```

```
  SUM(`Sale (Dollars)`) AS total_sales,
```

```
  SUM(`Volume Sold (Liters)`) AS total_volume_sold
```

```
FROM `certificate`.`liquor_sales`
```

```
WHERE `Date` IS NOT NULL
```

```
GROUP BY month
```

```
ORDER BY month;
```

Графік, наведений на рисунку 3.8 обчислює середні продажі за кожним місяцем, щоб виявити сезонні тенденції. Інтенсивність кольору залежить від величини середніх продажів: світлі кольори - вищі продажі, а темні кольори - нижчі продажі. На графіку можна побачити, що в жовтні найвищі середні продажі, а січень має найнижчі. Протягом всього року спостерігаються рівномірні середні продажі.

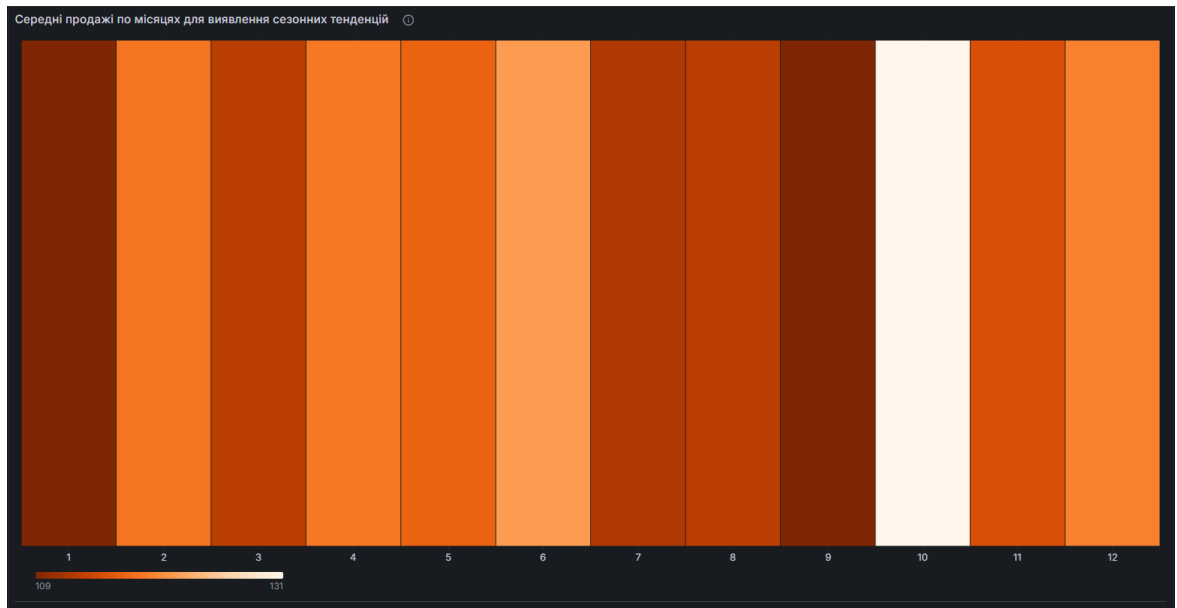


Рисунок 3.8 - Дані про середні продажі по місяцях

SQL код:

SELECT

toMonth(`Date`) AS month,

AVG(`Sale (Dollars)`) AS average_monthly_sales

FROM certificate.liquor_sales

GROUP BY month

ORDER BY month;

На наступному графіку 3.9 показано динаміку поливань продажів у святкові дні. Це означає продажі в доларах (`total_sales`) та в літрах (`total_volume_sold`). Помітні високий прибуток у великі свята (Новий рік, Різдво, День праці, День Конституції України). Код представлено в додатку А [A1].

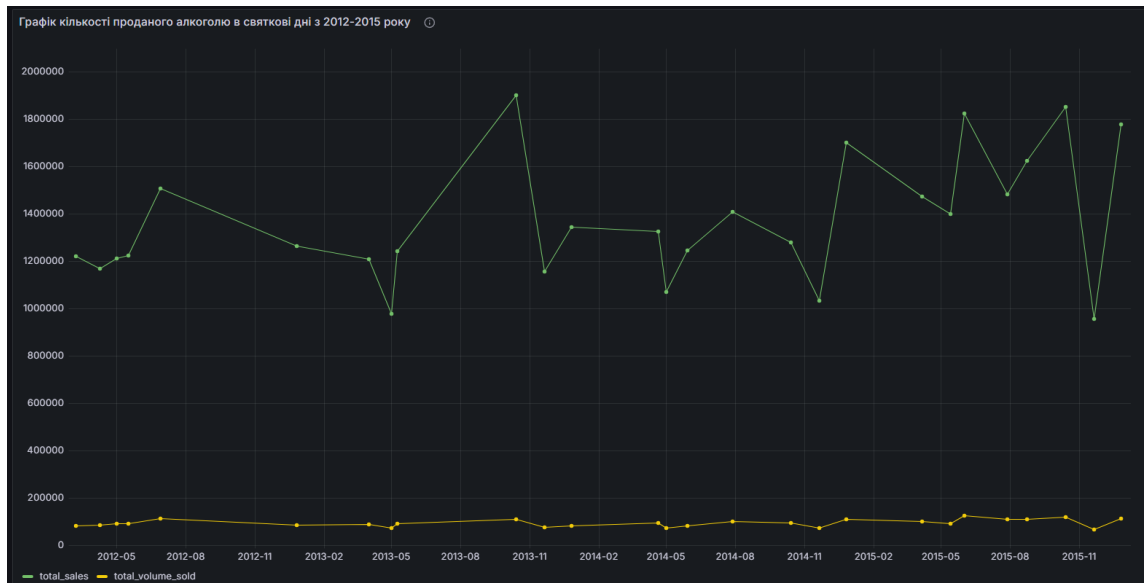


Рисунок 3.9 - Графік прибутків у свята

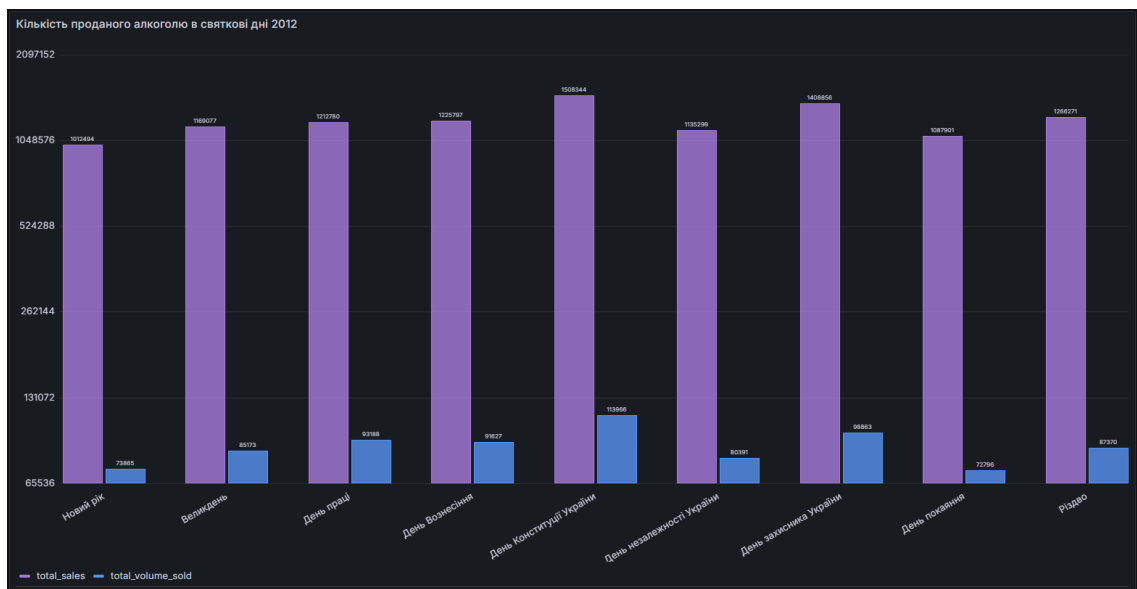


Рисунок 3.10 - Кількість проданого алкоголю в святкові дні з 2012

В 2012-му році можна побачити максимальні продажі на «День Конституції України», а також «День захисника України». Цей графік на відміну від попереднього має додаткове обмеження по рокам. Код відображено в додатку А [A2]. Інші графіки відображено в додатку Б, код подібний, як в 12-му році, просто поміняні роки на 13-15-ті роки відповідно.

У 2013 році високі продажі на «День Конституції України» та «День захисника України» залишаються високими, зрівнюючи з попереднім роком код наведено в. Додатку Б.

Починаючи з 2014-го року можна побачити, що більшість продажів пішло на Різдво код наведено в. Додатку Б.

У 2015-му році в топі по продажам алкоголю залишається день захисника України та Різдво. Код представлено код наведено в. Додатку Б.

Збільшення продажів алкогольних напоїв може бути пов'язане з сезонними факторами чи змінами в попиті на алкоголь. Тобто обсяг продажів при цьому зростає незначно.

Графік на рис.3.14 показує, що постачальник Diageo Americas є на першому місці по продажах, що перевищують 27 мільйонів доларів. Інші великі постачальники, такі як Jim Beam Brands та Pernod Ricard USA/Austin Nichols, мають менші продажі ніж у Diageo Americas і їх дохід є 11 та 9 мільйонів доларів відповідно.

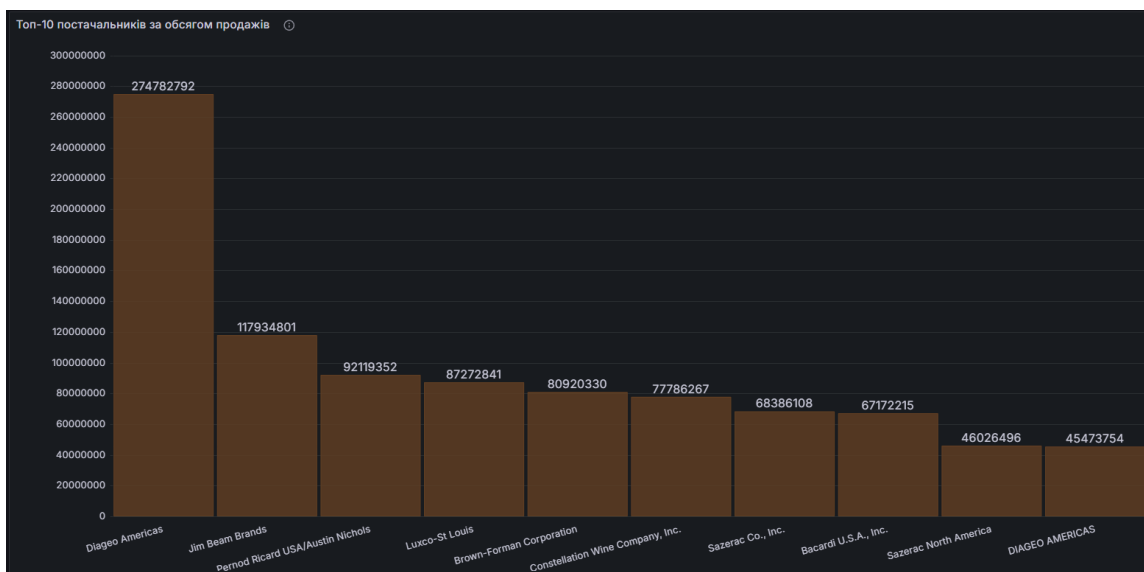


Рисунок 3.14 - Графік постачальників з найбільшим обсягом продажів

SQL код:

SELECT

`Vendor Name`,

SUM(`Sale (Dollars)`) AS total_sales

FROM certificate.liquor_sales

GROUP BY `Vendor Name`

```
ORDER BY total_sales DESC
LIMIT 10;
```

Лінійний графік, наведений на рисунку 3.15 показує стабільне зростання продажів з 2012 по 2015 рік. Зростання є поступовим, без особливих коливань. Зростання може бути обумовлене економічним розвитком або успішною маркетинговою стратегією від виробників алкогольної промисловості.

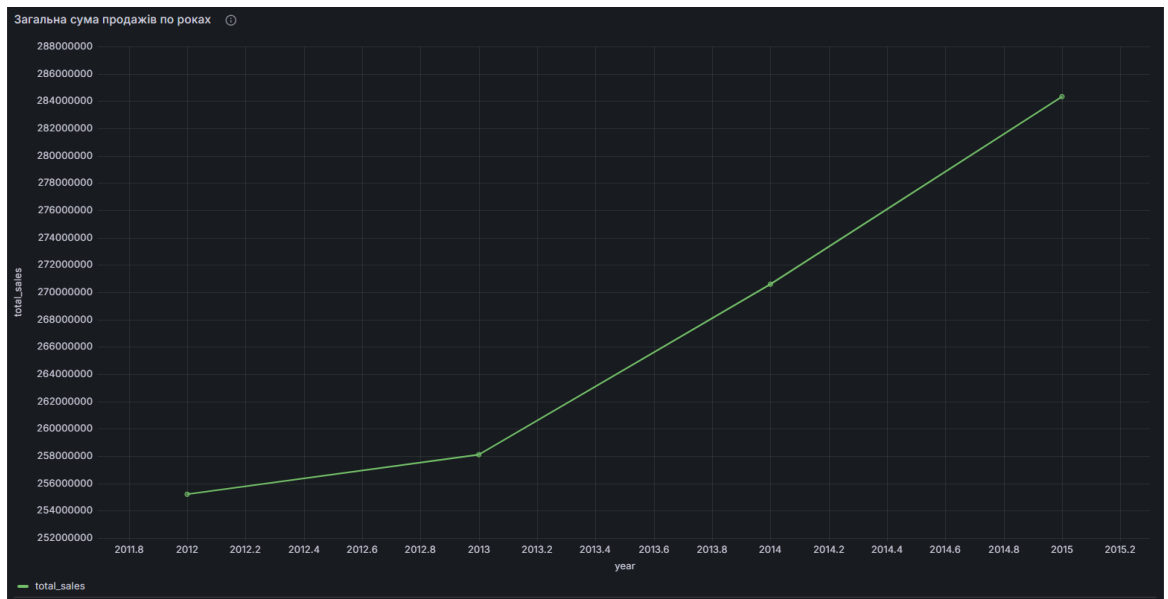


Рисунок 3.15 - Загальна сума продажів у доларах для кожного року

SQL код:

```
SELECT
  toYear(`Date`) AS year,
  SUM(`Sale (Dollars)`) AS total_sales
FROM certificate.liquor_sales
WHERE toYear(`Date`)
GROUP BY year
ORDER BY year;
```

Графік, наведений на рисунку 3.16 має найвищі обсяги продажів спостерігається у Vodka 80 Proof (понад 18 мільйонів пляшок). Це показує, що

є стабільний і високий попит на горілку стандартної міцності. Далі по графіку іде Cream Liqueurs (11 мільйонів пляшок) та Straight Bourbon Whiskeys (6,6 мільйонів пляшок). Вони мають високий попит, але не достатньо високий, як у лідера продажів на ринку. Клієнти дають значну перевагу універсальним напоям, бо горілка і є таким напоєм у більшості випадках.

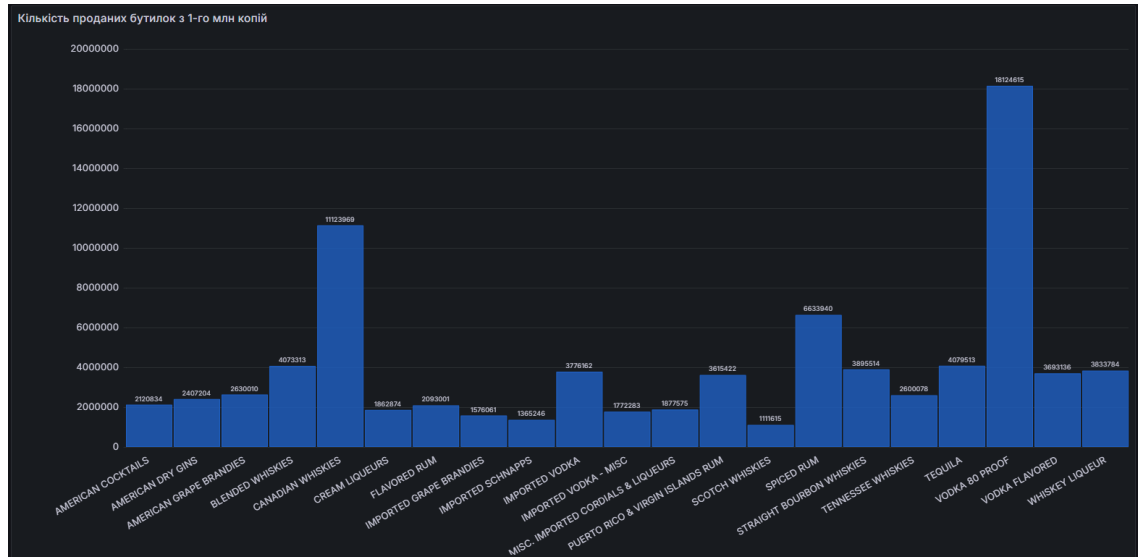


Рисунок 3.16 - Графік продажу пляшок алкоголю з найбільшим потенціалом

SQL код:

```
SELECT
```

```
  `Category Name` AS category,
```

```
  SUM(`Bottles Sold`) AS total_bottles_sold
```

```
FROM `certificate`.`liquor_sales`
```

```
WHERE `Date` IS NOT NULL
```

```
GROUP BY `Category Name` HAVING total_bottles_sold >1000000
```

```
ORDER BY `Category Name`;
```

Графік, наведений на рисунку 3.17 має дані, що Canadian Whiskies і Vodka 80 Proof є двома найпопулярнішими лідерами за всім обсягом продажів. Перші дві позиції мають майже половину від загальних продажів у топ-10 напоїв.

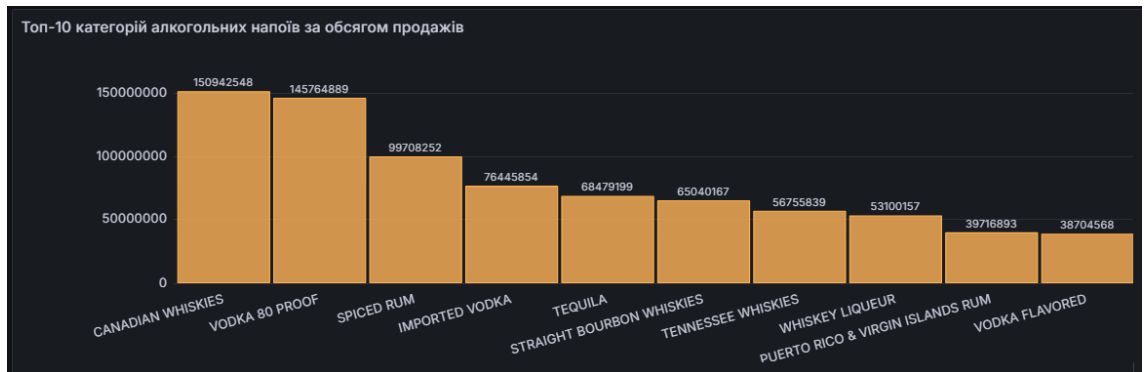


Рисунок 3.17 - Графік топ-10 категорій алкогольних напоїв за обсягом продажів у доларах

SQL код:

```
SELECT
  `Category Name`,
  SUM(`Sale (Dollars)`) AS total_sales
FROM certificate.liquor_sales
GROUP BY `Category Name`
ORDER BY total_sales DESC
LIMIT 10;
```

Графік, наведений на рисунку 3.18 показує середню кількість проданих пляшок на транзакцію. Тобто на один чек включає приблизно 8 пляшок. Це означає, що клієнти часто купляють оптом для святкувань чи тощо.

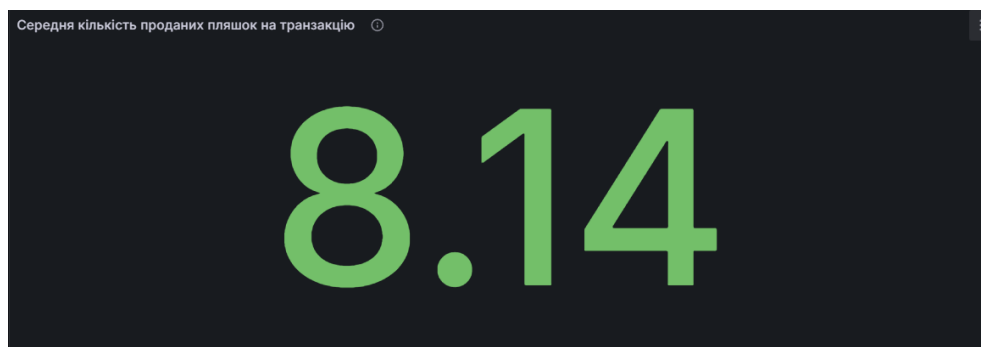


Рисунок 3.18 - Дані про середню кількість пляшок, проданих за одну транзакцію

SQL код:

```
SELECT
```

```
AVG(`Bottles Sold`) AS average_bottles_per_transaction
FROM certificate.liquor_sales;
```

Графік, наведений на рисунку 3.19 візуалізує дні з найбільшим продажем. Таким днем є понеділок. Це може вказувати на оптові закупівлі великої кількості напоїв або початок робочого тижня для закладів, де алкогольні напої були потрібні. Середа та четвер мають також мають достатньо високі продажі. В свою чергу в суботу є самі найменші продажі.

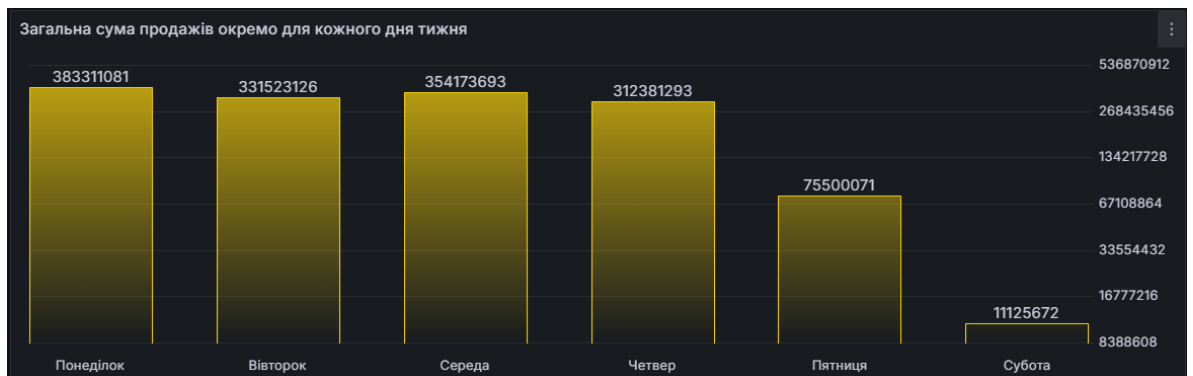


Рисунок 3.19 - Дані про суму продажів окремо для кожного дня тижня

SQL код:

```
SELECT
  toDayOfWeek(`Date`) AS day_of_week_number,
  CASE toDayOfWeek(`Date`)
    WHEN 1 THEN 'Понеділок'
    WHEN 2 THEN 'Вівторок'
    WHEN 3 THEN 'Середа'
    WHEN 4 THEN 'Четвер'
    WHEN 5 THEN 'П'ятниця'
    WHEN 6 THEN 'Субота'
  END AS day_of_week_name,
  SUM(`Sale (Dollars)`) AS total_sales
FROM certificate.liquor_sales
GROUP BY day_of_week_number, day_of_week_name
ORDER BY day_of_week_number;
```

Графік, наведений на рисунку 3.20 має загальний обсяг продажу алкоголю. На цьому графіку видно, що різниця між роками поступово. Це означає, що ринок алкоголю розвивався поступово.



Рисунок 3.20 - Дані про обсяг проданого алкоголю в літрах

SQL код:

```
SELECT
  toYear(`Date`) AS year,
  SUM(`Volume Sold (Liters)`) AS total_volume_liters
FROM certificate.liquor_sales
WHERE toYear(`Date`) NOT IN (2016, 2017)
GROUP BY year
ORDER BY year;
```

Графік, наведений на рисунку 3.21 показує, що HIGH PROOF BEER - AMERICAN має високу середню ціну (122 долари). Це може значить, що продукція є рідкісною або є преміальним брендом. Інші категорії, такі як JAPANESE WHISKY має середню ціну (78,7 долара), а це суттєво перевищує мінімальну середню ціну в (18,7 долара). Це значить, що широкий асортимент товарів з різним рівнем преміальності алкогольними напоями. Інші категорії мають збалансовану різницю між середньою та мінімальною ціною.

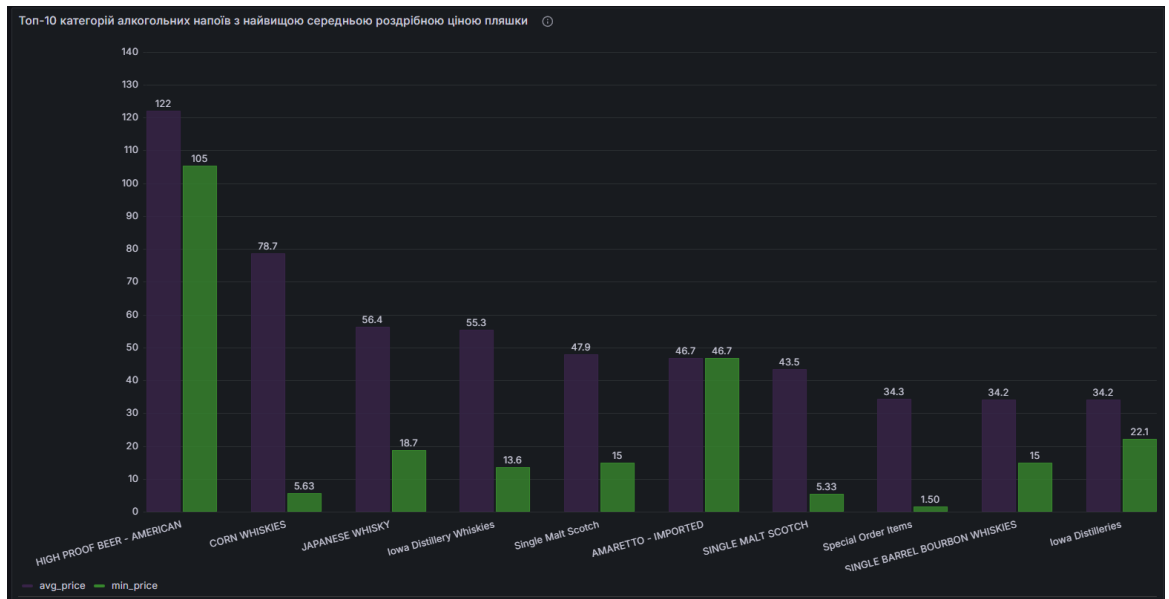


Рисунок 3.21 - Дані про категорії алкоголю з середньої ціною за пляшку

SQL код:

SELECT

`Category Name` AS category,

AVG(`State Bottle Retail`) AS avg_price,

MIN(`State Bottle Retail`) AS min_price

FROM certificate.liquor_sales

WHERE `Category Name` != "" AND isFinite(`State Bottle Retail`)

GROUP BY `Category Name`

ORDER BY avg_price DESC

LIMIT 10;

На графіку, наведеного на рисунку 3.22 можна спостерігати сезонний характер продажів. Високий попит в листопаді-грудні та низький у весняні та літні місяці. Динаміка має схожість з кожним роком. Це показує, що є кореляція даних у певні сезони.

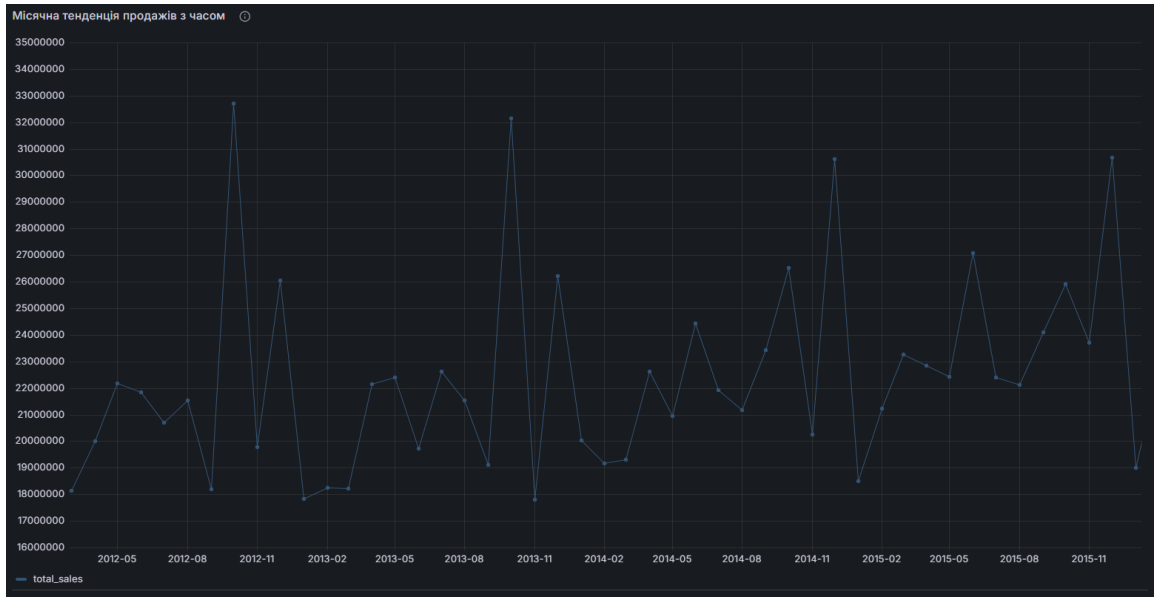


Рисунок 3.22 - Дані про зміни по продажам в місяці

SQL код:

SELECT

toStartOfMonth(`Date`) AS month,

SUM(`Sale (Dollars)`) AS total_sales

FROM certificate.liquor_sales

GROUP BY month

ORDER BY month;

Графік, наведений на рисунку 3.23 визначає загальну суму продажів у 10 днів з найвищими продажами алкоголю по різним рокам. Тут зберігається кореляція з великими святами. Найбільші продажі припадають на Новий рік, перед Різдвом та на інші великі свята.

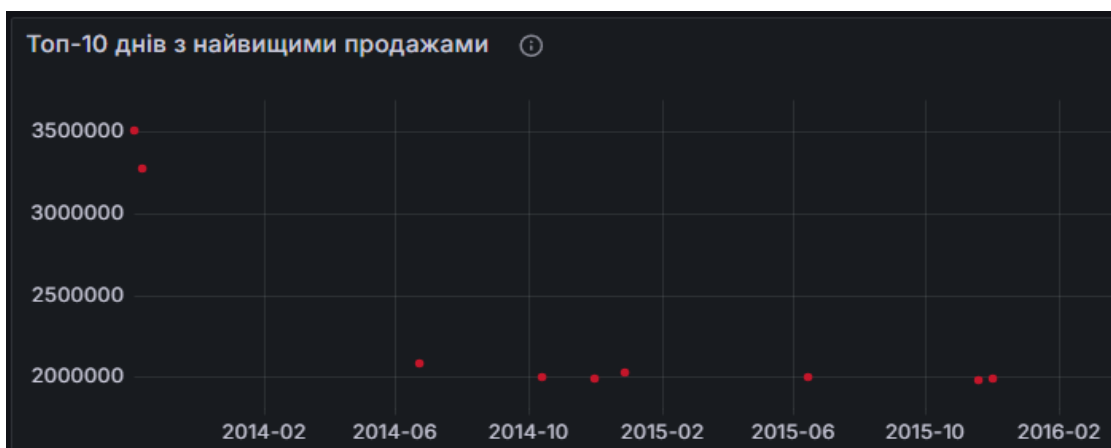


Рисунок 3.23 - Дані про топ-10 днів з найбільшим прибутком

```

SQL код:
SELECT
  `Date`,
  SUM(`Sale (Dollars)`) AS total_sales
FROM certificate.liquor_sales
GROUP BY `Date`
ORDER BY total_sales DESC
LIMIT 10;

```

Графік, наведений на рисунку 3.24 визначає середню ціну собівартості алкогольних напоїв. Зважаючи на дані, можна побачити, що жовта частина відповідає середній роздрібній ціні, зелена частина - середню собівартість пляшки, а синя частина - середній прибуток із продажу пляшки. Отже, собівартість значно менша, ніж роздрібна ціна. Основну частку ціни пляшки становить собівартість і прибуткова маржа. Роздрібна націнка на виріб демонструє стабільну рентабельність напою, що є позитивним показником для бізнесу та для майбутніх продажів.

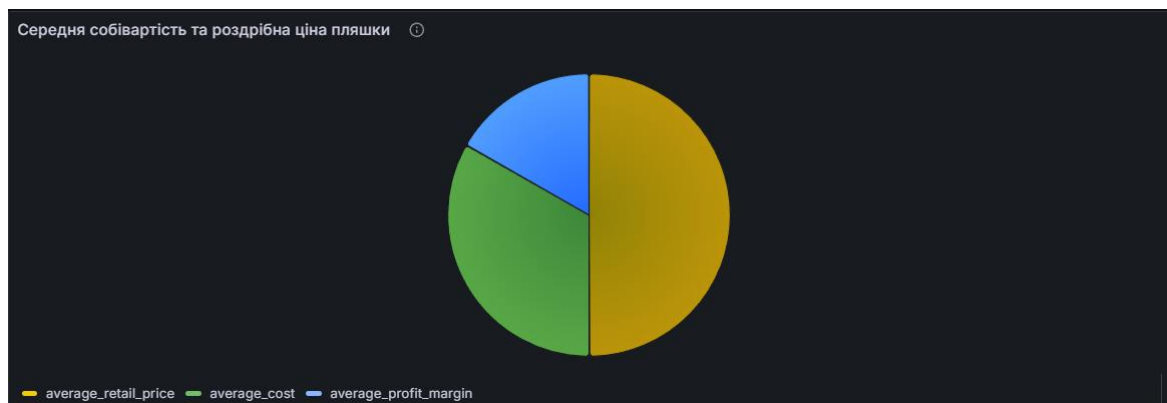


Рисунок 3.24 - Дані про середню собівартість виробів

```

SQL код:
SELECT
  AVG(`State Bottle Cost`) AS average_cost,
  AVG(`State Bottle Retail`) AS average_retail_price,
  AVG(`State Bottle Retail` - `State Bottle Cost`) AS average_profit_margin

```

FROM certificate.liquor_sales;

3.6 Висновок до третього розділу

У третьому розділі здійснено аналіз інструментів, які використовувалися для розробки аналітичної системи для дослідження продажів алкогольних напоїв.

Сформовано перелік основних задач для аналізу продажів алкогольних напоїв, способи їх реалізації, графічне відображення та результати аналізу.

Розглянуто етапи створення системи, а також її компоненти. Проведено аналіз компонентів системи та їх порівняння. Розроблено базу даних з таблицями, які містять інформацію про продажі алкогольних напоїв.

Наведено код програми налаштування віртуальної машини та конфігурацію Docker Compose Проведено очищення даних за допомогою мовою програмування Python. Здійснено візуалізацію даних та пояснення отриманих результатів. Були виявлені- кореляції у даних, зокрема на періоди свят, визначено найбільш популярні бренди напоїв, найприбутковіші місяці та дні тижня.

ВИСНОВКИ

Кваліфікаційна робота на тему «Дослідження та створення аналітичної системи для аналізу продажів алкогольних напоїв» дозволила дослідити великі обсяги даних, запропонувати способи дослідження актуальних питань щодо продажів алкогольних напоїв. Вона може бути використана для аналізу ринкових трендів і прогнозування майбутніх продажів алкогольних напоїв.

У ході роботи були виконані такі завдання:

1. Досліджено ринок продажів алкогольних напоїв на основі відкритих даних;
2. Досліджено ринок споживання алкогольних напоїв на основі відкритих даних;
3. Визначено перелік задач аналізу продажів алкогольних напоїв та методи їх реалізації;
4. Розроблено аналітичну систему продажів та споживання алкогольних напоїв з використанням сучасних ІТ-засобів для візуалізації та моніторингу даних.

У результаті виконання кваліфікаційної роботи досягнуто мети, яка полягала у дослідженні та створенні аналітичної системи для аналізу продажів алкогольних напоїв. Система позитивно вплине на обсяг продажів, цінову політику напоїв та прибутки харчових підприємств. Алкогольний ринок залежний від багатьох факторів. Саме тому важливим є розробка ефективної аналітичної системи, що дозволить проаналізувати великі обсяги даних та надати інформацію для прийняття обґрунтованих управлінських рішень.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Rivnepost. Тенденції та інновації в індустрії алкогольних напоїв. URL: <https://rivnepost.rv.ua/news/tendentsii-ta-innovatsii-v-industrii-alkoholnikh-napoiv> (дата звернення: 10.11.2024).
2. Drinks. Алкогольні напої у 2022: тренди і тенденції. URL: <https://drinks.ua/news/alkogolni-napoi-u-2022-trendi-tendencii/> (дата звернення: 12.11.2024).
3. Wikipedia. Міленіали. URL: <https://uk.wikipedia.org/wiki/%D0%9C%D1%96%D0%BB%D0%B5%D0%BD%D1%96%D0%B0%D0%BB%D0%B8> (дата звернення: 14.11.2024).
4. Аналіз регуляторного впливу. Встановлення заборони продажу алкогольних напоїв. URL: <https://kr-rada.gov.ua/site/uploads/files/Publicna-informacia/Regulyanorna-politika/analiz-vplivu-vstanovl-zaboroni-alcohol-napoiv.pdf> (дата звернення: 16.11.2024).
5. Suspilne.media. Тенденції споживання алкоголю в Україні. URL: <http://surl.li/rmuatj> (дата звернення: 18.11.2024).
6. Український інститут майбутнього. Як змінився психоемоційний стан українців. URL: <https://news.uifuture.org/yak-zminivsia-psikhoemosiyniy-stand-ukra/> (дата звернення: 20.11.2024).
7. Barout. Аналітика ринку алкоголю: основні тренди. URL: <https://barout.media/news/ukraine-alcohol-preferences-map-media/#:~:text=Пиво%20залишається%20найбільш%20популярним%20алкого%25> (дата звернення: 22.11.2024).
8. Simplilearn. Data Analysis Methods: Process & Types. URL: <https://www.simplilearn.com/data-analysis-methods-process-types-article> (дата звернення: 05.12.2024).
9. Barout. Відповідальне споживання: рейтинг Україна. URL: <https://barout.media/news/vidpovidalne-spozhyvannya-rating-ukraine/> (дата звернення: 25.11.2024).

10. News.gallup. Liquor Ties Wine as Second Favorite Adult Beverage. URL: <https://news.gallup.com/poll/264335/liquor-ties-wine-second-favorite-adult-beverage.aspx> (дата звернення: 26.11.2024).
11. UCDA. Аналіз споживчих трендів у ритейли. URL: <https://ukrainiancraftdistillers.com/yaki-micni-alkogolni-napoyi-vzhyvayut-ukrayinci#:~:text=Опитані%20групою%20«Рейтинг»%20серед%20найпопулярні%25> (дата звернення: 27.11.2024).
12. Товари і ринки. Алкогольні напої в Україні: проблеми та перспективи. URL: <https://journals.knute.edu.ua/commodities-and-markets/article/view/2105/1916> (дата звернення: 28.11.2024).
13. Barout. Рейтинг алкогольних напоїв у США: галуп-опитування. URL: <https://barout.media/news/gallup-poll-usa-alcohol-drinks-preference-wine-spirits-beer/> (дата звернення: 29.11.2024).
14. Barout. Алкогольні вподобання в Україні: карта медіа. URL: <https://barout.media/news/ukraine-alcohol-preferences-map-media/> (дата звернення: 30.11.2024).
15. Journals. Алкогольні ринки: аналіз і дослідження. URL: <https://journals.knute.edu.ua/commodities-and-markets/article/view/2105/1916> (дата звернення: 15.11.2024).
16. Barout. Алкогольні вподобання в Україні: карта медіа. URL: <https://barout.media/news/ukraine-alcohol-preferences-map-media/> (дата звернення: 20.11.2024).
17. Wikipedia. Long short-term memory. URL: [https://en.wikipedia.org/wiki/Long_short_term_memory#:~:text=Long%20short%2Dterm%20memory%20\(LSTM,commonly%20encountered%20by%20traditional%20RNNs](https://en.wikipedia.org/wiki/Long_short_term_memory#:~:text=Long%20short%2Dterm%20memory%20(LSTM,commonly%20encountered%20by%20traditional%20RNNs) (дата звернення: 12.11.2024).
18. Bayadera. Головні алкогольні тренди 2024 року. URL: <https://bayadera.ua/posts/glavnye-alkogolnye-trendy-2024> (дата звернення: 18.11.2024).

19. IBM. Моделі ARIMA: аналіз даних. URL: <https://www.ibm.com/topics/arima-model> (дата звернення: 12.11.2024).
20. Wikipedia. Автокореґресивно інтегрована модель ковзного середнього. URL: https://en.wikipedia.org/wiki/Autoregressive_integrated_moving_average (дата звернення: 13.11.2024).
21. Drinks-intel. Why the Alcohol Industry is Having a Sobering 2024 in the US Market. URL: <https://drinks-intel.com/intel/why-the-alcohol-industry-is-having-a-sobering-2024-in-the-us-market-intel/> (дата звернення: 15.11.2024).
22. Geeksforgeeks. SARIMA: Seasonal Autoregressive Integrated Moving Average. URL: <https://www.geeksforgeeks.org/sarima-seasonal-autoregressive-integrated-moving-average/> (дата звернення: 16.11.2024).
23. Gallup. Alcohol Consumption Increasingly Viewed as Unhealthy. URL: <https://news.gallup.com/poll/648413/alcohol-consumption-increasingly-viewed-unhealthy.aspx> (дата звернення: 17.11.2024).
24. Tfah. Pain in the Nation 2024. URL: <https://www.tfah.org/report-details/pain-in-the-nation-2024/> (дата звернення: 18.11.2024).
25. Unn. Новини України: аналітика споживання. URL: <https://unn.ua/> (дата звернення: 19.11.2024).
26. Wikipedia. Apache Hadoop. URL: https://uk.wikipedia.org/wiki/Apache_Hadoop (дата звернення: 20.11.2024).
27. Techdrinks. В Україні підвищується попит на міцні напої: результати дослідження ринку. URL: <https://techdrinks.info/v-ukrayini-pidvyshhuyetsya-popyt-na-mitsni-napoyi-rezultaty-doslidzhennya-rynku/> (дата звернення: 21.11.2024).
28. Wikipedia. Apache Spark. URL: https://en.wikipedia.org/wiki/Apache_Spark (дата звернення: 22.11.2024).
29. Wikipedia. Data Lake. URL: https://en.wikipedia.org/wiki/Data_lake (дата звернення: 23.11.2024).

- 30.Greenpost. Алкогольні тренди 2024. URL: <https://greenpost.ua/news/alkogolni-trendy-2024-i67554> (дата звернення: 24.11.2024).
- 31.Inventure. Аналіз ринку вина в Україні. URL: <https://inventure.com.ua/uk/analytics/investments/analiz-rinku-vina-v-ukrayini> (дата звернення: 25.11.2024).
- 32.Prophet. Facebook Prophet: Time Series Forecasting. URL: <https://facebook.github.io/prophet/> (дата звернення: 26.11.2024).
- 33.Sendpulse. Guide to Building an RFM Analysis. URL: <https://sendpulse.ua/blog/guide-to-building-an-rfm-analysis> (дата звернення: 27.11.2024).
- 34.Wikipedia. RFM-аналіз. URL: <https://uk.wikipedia.org/wiki/RFM-%D0%B0%D0%BD%D0%B0%D0%BB%D1%96%D0%B7> (дата звернення: 28.11.2024).
- 35.Datawiz. RFM-аналіз у роздрібній торгівлі. URL: <https://datawiz.io/uk/blog/rfm-analysis-in-retail> (дата звернення: 29.11.2024).
- 36.Researchgate. Trends in Alcoholic Beverage Consumption. URL: <http://surl.li/iuijgu> (дата звернення: 30.11.2024).
- 37.Wikipedia. Алгоритм Apriori. URL: <http://surl.li/rzavia> (дата звернення: 01.12.2024).
- 38.Wikipedia. Чисельні методи аналізу даних. URL: <http://surl.li/csenrt> (дата звернення: 10.11.2024).
- 39.Bayadera. Основні алкогольні тренди 2024 року. URL: <https://bayadera.ua/rus/posts/glavnye-alkogolnye-trendy-2024> (дата звернення: 12.11.2024).
- 40.Knu. Проблеми інформаційних технологій. URL: https://csc.knu.ua/media/study/asp/mod_probl_inf_tech_sys_analysis_ivohin/lecture/lec11.pdf (дата звернення: 13.11.2024).

41. Extension. Alcoholic Beverage Consumption and Purchasing Trends 2024. URL: <https://extension.psu.edu/alcoholic-beverage-consumption-and-purchasing-trends-2024> (дата звернення: 14.11.2024).
42. Wikipedia. Алгоритм Apriori: застосування у великих даних. URL: https://en.wikipedia.org/wiki/Apriori_algorithm (дата звернення: 15.11.2024).
43. DataScientist. Огляд алгоритму Apriori. URL: <http://datascientist.one/apriori-algorithm/> (дата звернення: 16.11.2024).
44. Gallup. Шість із десяти американців споживають алкоголь. URL: <https://news.gallup.com/poll/509501/six-americans-drink-alcohol.aspx> (дата звернення: 17.11.2024).
45. Theiwsr. Global Beverage Alcohol Market Set for Moderate Recovery in 2025. URL: <https://www.theiwsr.com/global-beverage-alcohol-market-set-for-moderate-recovery-in-2025-while-challenges-persist-in-2024/> (дата звернення: 18.11.2024).
46. DM. Часті шаблони та алгоритми FPGrowth. URL: <https://dm.cs.tu-dortmund.de/mlbits/frequent-pattern-fpgrowth/> (дата звернення: 19.11.2024).
47. Seventyfive. Тренди алкогольної індустрії. URL: <https://go.seventyfive.com/> (дата звернення: 20.11.2024).
48. LeadIQ. Аналітика Bevspot: стан ринку напоїв. URL: <https://leadIQ.com/c/bevspot/5a1d87d9240000240061be93> (дата звернення: 21.11.2024).
49. IT. Аналітичні методи в дослідженні споживання алкоголю. URL: <http://surl.li/zmgnhu> (дата звернення: 22.11.2024).
50. Wikipedia. Docker: ізоляція додатків. URL: <http://surl.li/unubiz> (дата звернення: 23.11.2024).
51. Wikipedia. Python як мова програмування. URL: <http://surl.li/uyaupw> (дата звернення: 24.11.2024).

- 52.Znu. Нейронні мережі в аналізі даних. URL: https://moodle.znu.edu.ua/pluginfile.php/675805/mod_resource/content/1/Su_bbotin_Neural-88-92.pdf (дата звернення: 25.11.2024).
- 53.Kaggle. Iowa Liquor Sales Dataset. URL: <https://www.kaggle.com/datasets/residentmario/iowa-liquor-sales/data> (дата звернення: 26.11.2024).
- 54.Wikipedia. Python (programming language). URL: [https://en.wikipedia.org/wiki/Python_\(programming_language\)](https://en.wikipedia.org/wiki/Python_(programming_language)) (дата звернення: 27.11.2024).
- 55.Microsoft. Power BI: Інструмент для аналітики. URL: <https://www.microsoft.com/de-de/power-platform/products/power-bi?market=de> (дата звернення: 28.11.2024).
- 56.Wikipedia. Power BI. URL: https://de.wikipedia.org/wiki/Power_BI (дата звернення: 12.11.2024).
- 57.Python.org. Офіційний сайт Python. URL: <https://www.python.org/> (дата звернення: 15.11.2024).
- 58.Wikipedia. Tableau Software. URL: https://es.wikipedia.org/wiki/Tableau_Software (дата звернення: 18.11.2024).
- 59.Tableau. Офіційний сайт Tableau. URL: <https://www.tableau.com/> (дата звернення: 21.11.2024).
- 60.Looker-bi. Інструмент бізнес-аналітики. URL: <https://cloud.google.com/looker-bi> (дата звернення: 11.12.2024).
- 61.Metabase. Платформа для створення дашбордів. URL: <https://www.metabase.com/> (дата звернення: 17.11.2024).
- 62.Wikipedia. Колаборативна робоча середа Colab. URL: <http://surl.li/izfdim> (дата звернення: 22.11.2024).
- 63.Colab. Google Colab: Інтерактивний середовище для програмування на Python. URL: <https://colab.research.google.com/> (дата звернення: 29.11.2024).

64. Wikipedia. Інформація про Docker. URL: <http://surl.li/kbuyqj> (дата звернення: 25.11.2024).
65. Stripo. Як маркетингова автоматизація ефективно залучає клієнтів. URL: <https://stripo.email/ua/blog/how-marketing-automation-effectively-drives-leads-to-your-saas-business/> (дата звернення: 28.11.2024).
66. Google Colab. Офіційний сайт. URL: <https://colab.google/> (дата звернення: 13.11.2024).
67. MariaDB. Офіційний сайт MariaDB. URL: <https://mariadb.com/> (дата звернення: 20.11.2024).
68. MariaDB Foundation. Інформація про MariaDB Foundation. URL: <https://mariadb.org/> (дата звернення: 30.11.2024).
69. Clickhouse. Офіційний сайт ClickHouse. URL: <https://clickhouse.com/> (дата звернення: 24.11.2024).
70. GitHub. Репозиторій ClickHouse. URL: <https://github.com/ClickHouse/ClickHouse> (дата звернення: 27.11.2024).
71. Pro-consulting. Консультаційні послуги. URL: <https://pro-consulting.ua/ua/> (дата звернення: 23.11.2024).
72. Drinks. Стан алкогольної індустрії у США. URL: <https://drinks.ua/news/sostojanie-otrasli-vodka-v-ssha/> (дата звернення: 10.11.2024).
73. Drinks. Аналіз ринку України. URL: https://pro-consulting.ua/ua/base/analiz-rynka-ukrainy?level1=alko&level2=Obshchei_rynok_AP&stat=1 (дата звернення: 18.12.2024).
74. SAP. Платформа для аналітики в хмарі. URL: <https://www.sap.com/products/technology-platform/cloud-analytics.html> (дата звернення: 26.11.2024).
75. Wikipedia. Oracle Cloud Enterprise Resource Planning. URL: https://en.wikipedia.org/wiki/Oracle_Cloud_Enterprise_Resource_Planning (дата звернення: 11.12.2024).

76. Microsoft Dynamics 365. Офіційний сайт. URL: <https://www.microsoft.com/de-de/dynamics-365> (дата звернення: 12.11.2024).
77. Google BigQuery. Сервіс аналітики даних. URL: <https://cloud.google.com/bigquery> (дата звернення: 19.11.2024).
78. AWS. Хмарні послуги від Amazon. URL: https://aws.amazon.com/de/?nc1=h_ls (дата звернення: 21.11.2024).
79. NielsenIQ. Інформація про споживчі дослідження. URL: <https://nielseniq.com/global/en> (дата звернення: 29.11.2024).
80. OTexts. Статистика та прогнозування. URL: <https://otexts.com/fpp2/stationarity.html> (дата звернення: 14.11.2024).
81. ACM Digital Library. Статті про статистику та аналіз даних. URL: <https://dl.acm.org/doi/10.1145/1133905.1133907> (дата звернення: 30.11.2024).
82. ClickHouse. Офіційна документація ClickHouse. URL: <https://clickhouse.com/docs/en/intro> (дата звернення: 10.12.2024).
83. Grafana. Документація Grafana. URL: <https://grafana.com/docs/grafana/latest/> (дата звернення: 13.12.2024).
84. Oracle VirtualBox. Інструмент для віртуалізації. URL: <https://www.oracle.com/virtualization/virtualbox/> (дата звернення: 22.12.2024).
85. Docker. Офіційний сайт Docker. URL: [https://en.wikipedia.org/wiki/Docker_\(software\)](https://en.wikipedia.org/wiki/Docker_(software)) (дата звернення: 15.11.2024).
86. Wikipedia. Python (programming language). URL: [https://de.wikipedia.org/wiki/Python_\(Programmiersprache\)](https://de.wikipedia.org/wiki/Python_(Programmiersprache)) (дата звернення: 17.12.2024).
87. Guru99. Посібник по ETL-тестуванню. URL: <https://www.guru99.com/uk/ultimate-guide-etl-datawarehouse-testing.html> (дата звернення: 12.12.2024).

- 88.Guru99. Процес ETL (Extract, Transform, Load). URL: <https://www.guru99.com/uk/etl-extract-load-process.html> (дата звернення: 28.11.2024).
- 89.ZTU. Наукові роботи з інформатики. URL: <https://conf.ztu.edu.ua/wp-content/uploads/2023/02/14.pdf> (дата звернення: 24.11.2024).
- 90.Mind the Graph. Геопросторовий аналіз. URL: <https://mindthegraph.com/blog/uk/%D1%89%D0%BE-%D1%82%D0%B0%D0%BA%D0%B5-%D0%B3%D0%B5%D0%BE%D0%BF%D1%80%D0%BE%D1%81%D1%82%D0%BE%D1%80%D0%BE%D0%B2%D0%B8%D0%B9-%D0%B0%D0%BD%D0%B0%D0%BB%D1%96%D0%B7/> (дата звернення: 25.11.2024).
- 91.VNTU. Дослідження в галузі інформатики. URL: <https://ir.lib.vntu.edu.ua/bitstream/handle/123456789/38995/138849.pdf?sequence=2> (дата звернення: 11.12.2024).
- 92.Pidru4niki. Статистичні методи в аналізі даних. URL: https://pidru4niki.com/1942111747789/informatika/statistichni_metodi_data_mining (дата звернення: 18.11.2024).
- 93.Data-B-I. Агрегатні функції в SQL. URL: <https://data-b-i.com/uk/article/agregatny-funktstii-v-sql.html> (дата звернення: 20.11.2024).
- 94.Методичні рекомендації до виконання кваліфікаційної роботи на здобуття освітнього ступеня «Магістр» спеціальності 122 «Комп'ютерні науки» освітньо-професійної програми «Управління інформацією та аналітика даних» денної форми здобуття освіти [Електрон. ресурс] / уклад. М. П. Костіков, С. В. Грибков, Н. В. Ліманська. К.: НУХТ, 2024. – 29 с.
- 95.ДСТУ 3008-2015. Інформація та документація. Звіти у сфері науки і техніки. Структура та правила оформлювання. К.: ДП «УкрНДНЦ», 2015.32 с.

ДОДАТКИ

Додаток А

Код запиту для формування графіку обсягу проданого алкоголю в святкові дні з 2012-2015 року [A1].

```
SELECT
  `Date`,
CASE
  WHEN `Date` = '2012-05-01' OR `Date` = '2013-05-01' OR `Date` = '2014-05-01' OR `Date` = '2015-05-01' THEN 'День праці'
  WHEN `Date` = '2012-03-08' OR `Date` = '2013-03-08' OR `Date` = '2014-03-08' OR `Date` = '2015-03-08' THEN 'Міжнародний жіночий день'
  WHEN `Date` = '2012-06-04' OR `Date` = '2013-06-01' OR `Date` = '2014-06-01' OR `Date` = '2015-06-01' THEN 'День захисту дітей'
  WHEN `Date` = '2012-06-28' OR `Date` = '2013-06-28' OR `Date` = '2014-06-28' OR `Date` = '2015-06-28' THEN 'День Конституції України'
  WHEN `Date` = '2012-08-24' OR `Date` = '2013-08-24' OR `Date` = '2014-08-24' OR `Date` = '2015-08-24' THEN 'День незалежності України'
  WHEN `Date` = '2012-10-14' OR `Date` = '2013-10-14' OR `Date` = '2014-10-14' OR `Date` = '2015-10-14' THEN 'День захисника України'
  WHEN `Date` = '2012-12-26' OR `Date` = '2013-12-26' OR `Date` = '2014-12-26' OR `Date` = '2015-12-26' THEN 'Другий день Різдва'
  WHEN `Date` = '2012-12-25' OR `Date` = '2013-12-25' OR `Date` = '2014-12-25' OR `Date` = '2015-12-25' THEN 'Різдво'
  WHEN `Date` = '2012-01-01' OR `Date` = '2013-01-01' OR `Date` = '2014-01-01' OR `Date` = '2015-01-01' THEN 'Новий рік'
  WHEN `Date` = '2012-04-08' OR `Date` = '2013-03-31' OR `Date` = '2014-04-20' OR `Date` = '2015-04-05' THEN 'Великдень'
  WHEN `Date` = '2012-04-06' OR `Date` = '2013-03-29' OR `Date` = '2014-04-18' OR `Date` = '2015-04-03' THEN 'Страшна п'ятниця'
```

```

WHEN `Date` = '2012-04-09' OR `Date` = '2013-04-01' OR `Date` = '2014-
04-21' OR `Date` = '2015-04-06' THEN 'Великодня понеділок'
WHEN `Date` = '2012-05-27' OR `Date` = '2013-05-19' OR `Date` = '2014-
06-08' OR `Date` = '2015-05-24' THEN 'Трійця'
WHEN `Date` = '2012-05-17' OR `Date` = '2013-05-09' OR `Date` = '2014-
05-29' OR `Date` = '2015-05-14' THEN 'День Вознесіння'
WHEN `Date` = '2012-11-21' OR `Date` = '2013-11-21' OR `Date` = '2014-
11-21' OR `Date` = '2015-11-21' THEN 'День покаяння'
WHEN `Date` = '2012-07-28' OR `Date` = '2013-07-28' OR `Date` = '2014-
07-28' OR `Date` = '2015-07-28' THEN 'День Української Державності'
ELSE 'Інше'
END AS holiday,
SUM(`Sale (Dollars)` ) AS total_sales,
SUM(`Volume Sold (Liters)` ) AS total_volume_sold
FROM `certificate`.`liquor_sales`
WHERE `Date` IN (
'2012-05-01', '2013-05-01', '2014-05-01', '2015-05-02',
'2012-03-08', '2013-03-08', '2014-03-08', '2015-03-08',
'2012-06-02', '2013-06-01', '2014-06-01', '2015-06-01',
'2012-06-28', '2013-06-29', '2014-06-28', '2015-06-28',
'2012-08-24', '2013-08-24', '2014-08-24', '2015-08-24',
'2012-10-14', '2013-10-14', '2014-10-14', '2015-10-14',
'2012-12-26', '2013-12-26', '2014-12-26', '2015-12-26',
'2012-12-25', '2013-12-25', '2014-12-25', '2015-12-25',
'2012-01-01', '2013-01-01', '2014-01-01', '2015-01-01',
'2012-04-08', '2013-03-31', '2014-04-20', '2015-04-05',
'2012-04-06', '2013-03-29', '2014-04-18', '2015-04-03',
'2012-04-09', '2013-04-01', '2014-04-21', '2015-04-06',
'2012-05-27', '2013-05-19', '2014-06-08', '2015-05-24',
'2012-05-17', '2013-05-09', '2014-05-29', '2015-05-14',

```

```

'2012-11-22', '2013-11-21', '2014-11-21', '2015-11-21',
'2012-07-28', '2013-07-28', '2014-07-28', '2015-07-28'
)
GROUP BY `Date`, holiday
ORDER BY `Date`;

```

Код запиту для формування обсягів проданого алкоголю у святкові дні 2014 р.

```

SELECT
`Date`,
CASE
WHEN `Date` = '2014-05-01' THEN 'День праці'
WHEN `Date` = '2014-06-27' THEN 'День Конституції України'
WHEN `Date` = '2014-08-21' THEN 'День незалежності України'
WHEN `Date` = '2014-10-14' THEN 'День захисника України'
WHEN `Date` = '2014-12-26' THEN 'Різдво'
WHEN `Date` = '2014-01-01' THEN 'Новий рік'
WHEN `Date` = '2014-04-21' THEN 'Великдень'
WHEN `Date` = '2014-05-29' THEN 'День Вознесіння'
WHEN `Date` = '2014-11-21' THEN 'День покаяння'
ELSE 'Інше'
END AS holiday,
SUM(`Sale (Dollars)`) AS total_sales,
SUM(`Volume Sold (Liters)`) AS total_volume_sold
FROM `certificate`.`liquor_sales`
WHERE `Date` IN (
'2014-05-01', -- День праці
'2014-06-27', -- День Конституції України
'2014-08-21', -- День незалежності України
'2014-10-14', -- День захисника України
'2014-12-26', -- Різдво

```

```
'2014-01-01', -- Новий рік  
'2014-04-21', -- Великдень  
'2014-05-29', -- День Вознесіння  
'2014-11-21' -- День покаяння  
)  
GROUP BY `Date`, holiday  
ORDER BY `Date`;
```

Додаток Б. Візуалізація даних

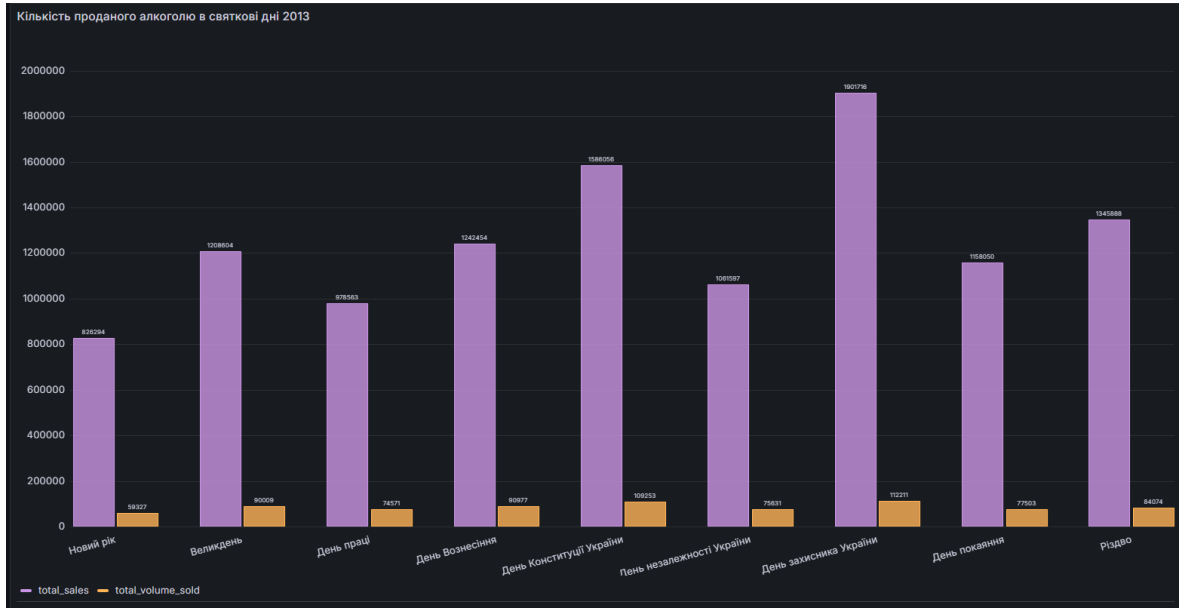


Рисунок Б.1 – Обсяги проданого алкоголю в святкові дні 2013 р

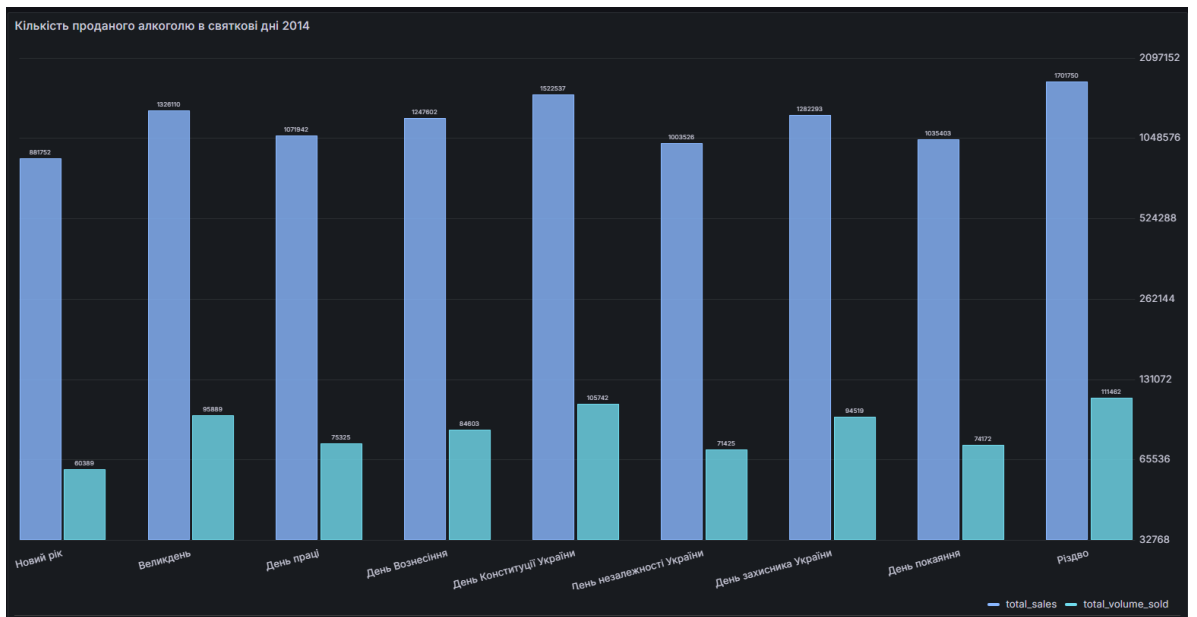


Рисунок Б.2 – Обсяги проданого алкоголю в святкові дні 2014 р

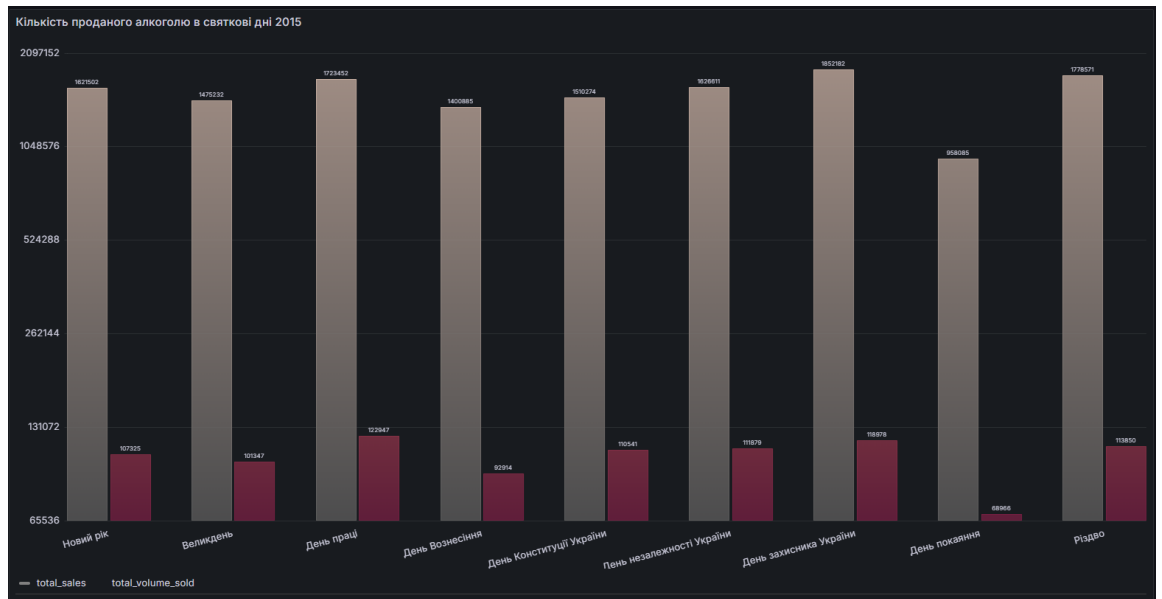


Рисунок Б.3 – Обсяги проданого алкоголю в святкові дні 2015 р