

**ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ
ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
КАФЕДРА ІНФОРМАЦІЙНИХ СИСТЕМ ТА ТЕХНОЛОГІЙ**

КВАЛІФІКАЦІЙНА РОБОТА

на тему: «Методика управління Big Data в інформаційних системах на
основі системного аналізу»

на здобуття освітнього ступеня магістра
зі спеціальності 124 Системний аналіз
освітньо-професійної програми «Інтелектуальні системи управління»

*Кваліфікаційна робота містить результати власних досліджень.
Використання ідей, результатів і текстів інших авторів мають посилання
на відповідне джерело*

_____ Інна ЛОГОША
(підпис)

Виконала: здобувачка вищої освіти групи САДМ-61

_____ Інна ЛОГОША

Керівник: _____ Віталій СВАТКО
к.т.н., доцент

Рецензент: _____
науковий ступінь, Ім'я, ПРІЗВИЩЕ
вчене звання

Київ 2024

**ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ**
Навчально-науковий інститут інформаційних технологій

Кафедра інформаційних систем та мереж

Ступінь вищої освіти магістр

Спеціальність 124 Системний аналіз

Освітньо-професійна програма Інтелектуальні системи управління

ЗАТВЕРДЖУЮ

Завідувач кафедри ІСТ

_____ Каміла СТОРЧАК

«_____» ____20__р.

**ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ**

Логоша Інна Володимирівна

(прізвище, ім'я, по батькові здобувача)

1. Тема кваліфікаційної роботи: Методика управління Big Data в інформаційних системах на основі системного аналізу

керівник кваліфікаційної роботи Віталій СВЯТКО, к. т. н., доцент

(Ім'я, ПРИЗВИЩЕ, науковий ступінь, вчене звання)

затверджені наказом Державного університету інформаційно-комунікаційних технологій
від «_____» _____20__р. № _____

2. Строк подання кваліфікаційної роботи «_____» _____20__р.

3. Вихідні дані до кваліфікаційної роботи: демографічні дані, фінансові дані, дані про конкурентів, маркетингові показники, інфраструктурні показники, динамічні показники, аналітичні дані.

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити)

1. Проаналізувати вимоги до управління Big Data

2. Розробити методології системного аналізу для управління Big Data

3. Оцінити ефективності та перспективи управління Big Data

5. Перелік ілюстративного матеріалу: презентація

6. Дата видачі завдання «_____» _____20__р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
1	Перегляд наукової та технічної літератури на тему магістерської роботи		
2	Системний аналіз та організаційні структури управління бізнесом.		
3	Аналіз основних фінансово – економічних показників, структури управління, оцінка кадрового потенціалу досліджуваного підприємства		
4	Розробка рекомендації та пропозиції щодо системних підходів до вдосконалення організаційної структури управління бізнесом.		
5	Оформлення результатів дослідження.		
6	Проходження плагіату.		
7	Підготовка доповіді до захисту.		
8	Захист роботи		

Здобувачка вищої освіти

(підпис)

Інна ЛОГОША

(Ім'я, ПРІЗВИЩЕ)

Керівник
кваліфікаційної роботи

(підпис)

Віталій СВАТКО

(Ім'я, ПРІЗВИЩЕ)

РЕФЕРАТ

Текстова частина кваліфікаційної роботи на здобуття освітнього ступеня магістра:
_____ стор., _____ рис., _____ табл., _____ джерел.

Мета роботи – розробка та впровадження системного аналізу для оптимізації управління великими даними (Big Data) в організаціях.

Об'єкт дослідження – процеси управління Big Data в організаціях, зокрема їхні особливості, виклики та можливості оптимізації.

Предмет дослідження – методи та стратегії системного аналізу, які спрямовані на оптимізацію управління великими даними (Big Data), забезпечуючи ефективність і відповідність сучасним вимогам. Дослідження зосереджене на розробці універсального фреймворку для використання великих даних у підтримці стратегічних рішень, підвищенні продуктивності та конкурентоспроможності організацій.

Короткий зміст роботи: Дипломна робота присвячена аналізу та оптимізації управління великими даними (Big Data) за допомогою методів системного аналізу. У роботі розглядаються основні характеристики великих даних, такі як обсяг, швидкість, різноманітність і достовірність, а також методи їхньої обробки. Запропоновано концептуальну модель системного аналізу для управління великими даними, що враховує сучасні технології, такі як хмарні обчислення та NoSQL-бази. Проведено аналіз практичного застосування великих даних у бізнесі, фінансах, медицині та науці, а також досліджено вплив маркетингових факторів на трафік офлайн-магазинів. Використовуючи методи машинного навчання, виконано прогнозування трафіку на основі реальних даних, що підтверджує ефективність запропонованих підходів.

КЛЮЧОВІ СЛОВА: big data, системний аналіз, управління даними, аналіз даних, хмарні обчислення, NoSQL, NewSQL, моделювання, машинне навчання, алгоритми, медіа-аналітика, фінансовий аналіз, оптимізація, прогнозування, кластерний аналіз, лінійна регресія, тренди, сезонність, бізнес-аналітика

ABSTRACT

Text part of the master's qualification work:

_____ pages, _____ pictures, _____ table, _____ sources.

The purpose of the work is development and implementation of system analysis to optimize Big Data management in organizations.

Object of research is Big Data management processes in organizations, including their features, challenges, and opportunities for optimization.

Subject of research – methods and strategies of system analysis aimed at optimizing Big Data management, ensuring efficiency and compliance with modern requirements. The research focuses on the development of a universal framework for using big data to support strategic decisions, increase productivity and competitiveness of organizations.

Summary of the work: This thesis is dedicated to the analysis and optimization of Big Data management using system analysis methods. The paper discusses the main characteristics of big data, such as volume, speed, variety, and reliability, as well as methods of its processing. A conceptual model of system analysis for big data management is proposed, considering modern technologies such as cloud computing and NoSQL databases. The practical application of big data in business, finance, medicine and science is analyzed, and the influence of marketing factors on the traffic of offline stores is investigated. Using machine learning methods, traffic forecasting based on real data is performed, which confirms the effectiveness of the proposed approaches.

KEYWORDS: big data, system analysis, data management, data analysis, cloud computing, NoSQL, NewSQL, modeling, machine learning, algorithms, media analytics, financial analysis, optimization, forecasting, cluster analysis, linear regression, trends, seasonality, business intelligence

**ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ**

Навчально-науковий інститут інформаційних технологій

**ПОДАННЯ
ГОЛОВІ ЕКЗАМЕНАЦІЙНОЇ КОМІСІЇ
ЩОДО ЗАХИСТУ КВАЛІФІКАЦІЙНОЇ РОБОТИ
на здобуття освітнього ступеня магістра**

Направляється здобувач Логоша І. В. до захисту кваліфікаційної роботи
(прізвище та ініціали)
за спеціальністю 124 Системний аналіз
(код, найменування спеціальності)
освітньо-професійної програми Інтелектуальні системи управління
(назва)
на тему: «Методика управління Big Data в інформаційних системах на основі
системного аналізу».

Кваліфікаційна робота і рецензія додаються.

Директор ННІТ

(підпис)

Катерина НЕСТЕРЕНКО

(Ім'я, ПРІЗВИЩЕ)

Висновок керівника кваліфікаційної роботи

Здобувачка Інна ЛОГОША демонструє високий рівень підготовки та ґрунтовне опрацювання теми. Студентка добре розкрила актуальність теми, пояснивши значення технологій Big Data для сучасного управління інформацією. У роботі систематично аналізуються підходи до управління великими даними, включаючи методи системного аналізу та сучасні технології обробки інформації. Автор продемонструвала вміння працювати з теоретичним матеріалом, використовуючи наукові підходи та методи аналізу даних. Робота включає практичні дослідження, що свідчить про застосування аналітичних методів і вміння працювати з реальними даними.

Все це дозволяє оцінити виконану кваліфікаційну роботу здобувачки Інни ЛОГОШІ на оцінку «_____» та присвоїти їй кваліфікацію магістр з системного аналізу.

Керівник кваліфікаційної роботи _____

(підпис)

Віталій СВАТКО

(Ім'я, ПРІЗВИЩЕ)

«_____» _____ 20____ року

Висновок кафедри про кваліфікаційну роботу

Кваліфікаційна робота розглянута. Здобувачка Логоша І. В. допускається до захисту даної роботи в Екзаменаційній комісії.

Завідувач кафедру ІСТ

(назва)

(підпис)

Каміла СТОРЧАК

(Ім'я, ПРІЗВИЩЕ)

ВІДГУК РЕЦЕНЗЕНТА
на кваліфікаційну роботу
на здобуття освітнього ступеня магістра

здобувача вищої освіти Логоша Інна Володимирівна

(прізвище, ім'я, по батькові)

на тему «Методика управління Big Data в інформаційних системах на основі системного аналізу»

Актуальність.

Робота присвячена сучасній та актуальній проблемі управління великими даними, що є ключовим аспектом у цифровій економіці. Тема є надзвичайно важливою в умовах зростання обсягів інформації, які організації потребують аналізувати для прийняття обґрунтованих рішень. Застосування методів системного аналізу для оптимізації роботи з Big Data є перспективним напрямком, що може значно підвищити ефективність процесів у бізнесі, фінансах, науці, медицині та інших сферах

Позитивні сторони.

1. Автор розробив модель аналізу даних, яка включає методи прогнозування, обробки та оцінки ефективності впровадження Big Data.

2. В роботі присутній аналіз реальних даних, який демонструє застосування системного аналізу для бізнес-аналітики, управління трафіком клієнтів та оцінки ефективності маркетингових кампаній.

3. Використання таких методів, як регресійний аналіз, дисперсійний аналіз, XGBoost та машинне навчання для прогнозування даних робить дослідження цінним не тільки для теоретиків, а й для практиків.

Недоліки.

1. Деякі графіки та таблиці містять значний обсяг інформації без детальних пояснень, що може викликати труднощі у читача, не знайомого з методами аналізу даних.

2. Було б корисно представити порівняльний аналіз ефективності різних методів аналізу даних у конкретних сценаріях.

Відзначені зауваження не впливають на загальну позитивну оцінку кваліфікаційної роботи магістерської

Висновок: *кваліфікаційна робота на здобуття ступеня магістра заслуговує оцінку "_____", а здобувачка Інна ЛОГОША заслуговує присвоєння кваліфікації магістр з системного аналізу*

Рецензент:

науковий ступінь, вчене звання

підпис

Ім'я, ПРІЗВИЩЕ

ЗМІСТ

ВСТУП.....	9
РОЗДІЛ 1. ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ ЗАСАДИ СИСТЕМНОГО АНАЛІЗУ У ВИРІШЕННІ ПРОБЛЕМ УПРАВЛІННЯ ПІДПРИЄМСТВ	14
1.1 Визначення та суть системного аналізу	15
1.2 Роль системного аналізу в прийнятті управлінських рішень.....	16
1.3 Основні методи та підходи системного аналізу	16
1.4 Процес ухвалення рішень в управлінні.....	17
1.5 Використання системного аналізу для зменшення ризиків	17
РОЗДІЛ 2. ХАРАКТЕРИСТИКА ТА ПРИНЦИПИ РОБОТИ З ВЕЛИКИМИ ДАНИМИ (BIG DATA)	26
2.1 Визначення поняття "Big Data".....	26
2.2 Основні характеристики великих даних (Volume, Velocity, Variety, Veracity) ...	27
2.3 Розвиток технологій обробки великих даних	27
2.4 Використання великих даних у різних галузях	33
РОЗДІЛ 3. ЗАСТОСУВАННЯ СИСТЕМНОГО АНАЛІЗУ В УПРАВЛІННІ ВЕЛИКИМИ ДАНИМИ.....	41
3.1 Взаємозв'язок системного аналізу та управління великими даними	41
3.2 Основні підходи до інтеграції Big Data у стратегічне управління.....	43
3.3 Методи оцінки ефективності впровадження великих даних	44
РОЗДІЛ 4. ІНТЕГРАЦІЯ СИСТЕМНОГО АНАЛІЗУ В УПРАВЛІННЯ ВЕЛИКИМИ ДАНИМИ: ПІДХОДИ, ВИКЛИКИ ТА ПЕРСПЕКТИВИ	48
4.1 Використання великих даних у фінансовому секторі	48
4.2 Вплив Big Data на маркетинг та рекламу	69
ВИСНОВКИ	74
ПЕРЕЛІК ПОСИЛАНЬ	76
ДЕМОНСТРАЦІЙНІ МАТЕРІАЛИ (Презентація)	77

ВСТУП

Актуальність теми. Big Data – це усталена у світі концепція управління інформацією, яка активно використовується у різних сферах і галузях. Популярність терміну постійно зростає завдяки численним спробам оцінити реальний обсяг даних, що зберігаються або використовуються організаціями різного типу: діловими, соціальними чи технічними.

Протягом останніх років термін "Big Data" став невіддільною частиною сучасного інформаційного середовища. Великі дані можна розглядати як сукупність структурованої та неструктурованої інформації, яка характеризується великим обсягом, високою швидкістю надходження і обробки, а також різноманітністю. Ці характеристики роблять такі дані складними для обробки за допомогою традиційних управлінських процедур, баз даних та інструментів.

Простіше кажучи, великі дані – це дані, обсяг, швидкість та різноманітність яких перевищують можливості звичайних систем для їх ефективної обробки. Вони вимагають новаторських, економічно ефективних підходів до управління, що дозволяють не лише аналізувати інформацію, але й автоматизувати процеси, покращувати прийняття рішень та отримувати глибші інсайти.

Для опису великих даних найчастіше використовують модель "4V", яка включає чотири основні характеристики:

Обсяг (Volume) – величезна кількість даних, які постійно зростають.

Швидкість (Velocity) – швидке надходження даних у режимі реального часу.

Різноманітність (Variety) – різні типи даних, включаючи текст, зображення, відео, датчики тощо.

Достовірність (Veracity) – надійність та точність даних, які використовуються для аналізу.

Після появи концепції "4V" дослідники додали додаткові "V", такі як:

Візуалізація (Visualization) – методи представлення даних, які допомагають отримувати інсайти.

Цінність (Value) – практична користь, яку організація отримує від аналізу даних.

Попри розширення моделі, нові характеристики переважно деталізують або доповнюють основні "V", підтверджуючи ключову роль обсягу, швидкості, різноманітності та достовірності у роботі з великими даними. Саме ці особливості зробили **Big Data** основним інструментом сучасного цифрового світу, який активно впливає на розвиток бізнесу, науки та суспільства.

Управління великими даними ґрунтується на інтеграції сучасних технологій, таких як **хмарні обчислення**, а також спеціалізованих систем управління даними, наприклад **NoSQL** та **NewSQL**. Ці системи часто поєднуються зі структурованими підходами, що дозволяє забезпечити ефективну роботу з великими обсягами інформації. Завдяки цьому, такі технології створюють надійну основу для зберігання, обробки та аналізу даних.

Big Data використовується у різноманітних сферах, забезпечуючи обробку та аналіз великих обсягів даних, які раніше були недоступні для опрацювання традиційними методами. Нижче наведено детальний огляд застосування Big Data у ключових галузях:

1. Бізнес та маркетинг

Аналітика клієнтів: Використання великих даних для аналізу поведінки клієнтів дозволяє компаніям створювати персоналізовані пропозиції, підвищуючи ефективність рекламних кампаній.

Прогнозування попиту: Big Data допомагає передбачати потреби споживачів, що сприяє оптимізації виробництва та логістики.

Аналіз ринку: Глибоке вивчення даних дозволяє підприємствам виявляти ринкові тренди, оцінювати конкурентні переваги та відкривати нові бізнес-можливості.

2. Фінанси та банківська справа

Кредитний аналіз: Завдяки аналізу великих обсягів даних банки можуть оцінювати кредитоспроможність клієнтів швидше та точніше.

Протидія шахрайству: Використання Big Data сприяє виявленню підозрілих фінансових операцій, що допомагає попереджати шахрайство.

3. Охорона здоров'я

Персоналізована медицина: Аналіз геномних та медичних даних дає змогу розробляти індивідуальні методи лікування та прогнозувати ризики захворювань.

Епідеміологічні дослідження: Big Data дозволяє стежити за динамікою поширення хвороб і розробляти стратегії їхнього стримування.

4. Наука та дослідження

Астрономія: Великі дані використовуються для аналізу космічних явищ та відкриття нових небесних об'єктів.

Геноміка: Обробка геномних даних сприяє глибшому розумінню генетичних аспектів хвороб і створенню нових підходів до лікування.

5. Технології та Інтернет речей (IoT)

Управління мережею: Big Data забезпечує моніторинг і управління великою кількістю підключених пристроїв, оптимізуючи їхню роботу.

Оптимізація енергоспоживання: Аналіз даних дозволяє ефективно використовувати енергію в IoT-системах, сприяючи зменшенню витрат

Це лише окремі приклади застосування **Big Data**, але його роль продовжує стрімко зростати у різних галузях. Організації активно впроваджують стратегії обробки та аналізу великих обсягів даних для ухвалення більш обґрунтованих стратегічних рішень і забезпечення своєї конкурентоспроможності.

Значущість цієї теми підкреслюється тим, що правильне управління великими даними стає стратегічно важливим фактором успіху будь-якої сучасної організації. Ефективність роботи з Big Data безпосередньо впливає на здатність компаній до інновацій, зміцнення їхньої конкурентної позиції та забезпечення сталого розвитку.

Крім того, у зв'язку зі зростанням масштабів цифрової трансформації та збільшенням кількості підключених пристроїв (наприклад, через **Інтернет речей**), управління великими даними стає необхідною умовою для забезпечення безпеки, захисту конфіденційності та оптимізації використання інформаційних ресурсів.

Таким чином, дослідження та впровадження системного аналізу для оптимізації управління великими даними набувають критичного значення. Це

дозволяє організаціям ефективніше реагувати на сучасні виклики та використовувати можливості, які відкриває епоха цифрової трансформації.

Метою цього дослідження є розробка та впровадження системного аналізу для оптимізації управління великими даними (Big Data) в організаціях. Ми прагнемо створити універсальний фреймворк, який дозволить ефективно використовувати великі дані для підтримки стратегічних рішень, підвищення продуктивності та конкурентоспроможності організацій.

Завдання дослідження

1. Аналіз поточного стану управління великими даними в організаціях: Виявити ключові виклики, які постають перед організаціями у сфері управління великими даними. Оцінити рівень інтеграції та використання Big Data в різних галузях.
2. Визначення основних принципів системного аналізу та їх адаптація для управління великими даними: Розглянути концепції системного аналізу та визначити їхнє застосування у вирішенні проблем управління великими даними.
3. Розгляд і обґрунтування методів оптимізації управління великими даними: Дослідити стратегії розподіленої обробки даних та їх масштабованість. Оцінити ефективність використання технологій хмарних обчислень.
4. Розробка моделі системного аналізу для управління великими даними: Створити концептуальну модель, яка враховує специфіку великих даних та їх інтеграцію з існуючими інформаційними системами.
5. Визначення критеріїв ефективності впровадження моделі: Розробити критерії для оцінки ефективності застосування системного аналізу в управлінні великими даними. Сформулювати поетапний план впровадження розробленої моделі.
6. Аналіз витрат та вигод від оптимізації управління великими даними: Вивчити фінансові аспекти впровадження нової системи. Порівняти витрати з очікуваними вигодами для організацій.
7. Оцінка майбутніх тенденцій у сфері управління великими даними: Прогнозувати технологічні та стратегічні зміни, які впливатимуть на управління великими даними у майбутньому.

8. Формулювання висновків і рекомендацій: Підсумувати результати дослідження. Надати конкретні рекомендації організаціям щодо впровадження системного аналізу та оптимізації управління великими даними.

Запропоновані завдання забезпечать комплексний підхід до аналізу управління великими даними, дозволяючи розробити практичні стратегії оптимізації, які враховуватимуть специфіку кожної організації та сучасні тенденції галузі.

Об'єктом дослідження є процеси управління великими даними (Big Data) в організаціях, зокрема їхні особливості, виклики та можливості оптимізації.

Предметом дослідження виступають методи та стратегії системного аналізу, які спрямовані на оптимізацію управління великими даними, забезпечуючи ефективність і відповідність сучасним вимогам.

Наукова новизна дослідження полягає у розробці та обґрунтуванні конкретних аспектів системного аналізу, які можуть бути ефективно впроваджені для оптимізації управління великими даними в умовах цифрової трансформації та швидкого розвитку технологій.

Практичне значення одержаних результатів

Результати дослідження матимуть вагоме практичне значення для організацій, що прагнуть:

- Оптимізувати свої процеси управління великими даними.
- Підвищити ефективність використання інформаційних ресурсів.
- Забезпечити високий рівень безпеки та конфіденційності даних.
- Зберігати та посилювати свою конкурентоспроможність на сучасному ринку.

РОЗДІЛ 1. ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ ЗАСАДИ СИСТЕМНОГО АНАЛІЗУ У ВИРІШЕННІ ПРОБЛЕМ УПРАВЛІННЯ ПІДПРИЄМСТВ

Питання, що порушуються в загальному вигляді та у зв'язку з фундаментальними науковими і практичними проблемами. Зростаюча складність управлінських проблем вимагає постійного розвитку та вдосконалення методологічних приписів, методів і рекомендацій. Наукова обґрунтованість та оптимальність рішень залежить, з одного боку, від досконалості методів, що використовуються в процесі прийняття та реалізації рішень, а з іншого - від рівня знань і навичок менеджерів у використанні комплексу методів. З цієї точки зору необхідно визначити теоретико-методологічні засади системного аналізу у вирішенні економічних проблем управління діяльністю підприємств.

Метою даної роботи є дослідження використання системного аналізу в управлінні та розробка практичних рекомендацій для забезпечення прийняття оптимальних управлінських рішень. У статті розглядаються методологічні підходи та методи вирішення складних і неструктурованих проблем в управлінні. Виклад основного матеріалу дослідження Успіх будь-якого підприємства залежить від економічного обґрунтування семи управлінських рішень. На сьогоднішній день розроблено низку наукових методів підготовки рішень для того, щоб менеджери могли раціонально приймати рішення та обирати найкращий з можливих варіантів. Однак ці методи ще не знайшли широкого застосування в управлінській практиці. Багато керівників покладаються на власний досвід і методи управління і не приділяють достатньої уваги вивченню і застосуванню наукових методів управління.

Одночасне застосування мистецтва управління та наукових методів значно підвищує ефективність управлінської діяльності. Процес ухвалення рішень і його зміст залежать від характеру проблеми, яку необхідно розв'язати. Управлінські рішення можуть бути класифіковані залежно від кількості можливих варіантів вирішення, числа станів навколишнього середовища, а також змін тих чи інших параметрів середовища.

Однак створення універсальної системи ухвалення рішень, яка б могла розв'язувати всі типові управлінські завдання і враховувати всі можливі випадки, є практично неможливим. Натомість у наукових публікаціях і підручниках з менеджменту пропонуються схеми процедур ухвалення рішень, орієнтовані на типові управлінські задачі.

Однією з ключових складових цих схем є системний аналіз, який забезпечує структурований підхід до розв'язання проблем, враховуючи всі аспекти середовища та його параметри. Системний аналіз дає змогу знизити ризики, підвищити обґрунтованість ухвалених рішень і забезпечити ефективність реалізації управлінських стратегій.

Сучасні підприємства, що працюють в умовах ринкової економіки, є складними системами, які стикаються з численними викликами та конкуренцією. Для досягнення конкурентних переваг та успіху важливо розуміти взаємодію між діяльністю компанії та її оточенням. У цьому контексті системний аналіз є ефективним методом дослідження складних систем, що дозволяє проаналізувати взаємозв'язки та взаємодію між елементами, які складають систему.

1.1 Визначення та суть системного аналізу

Системний аналіз - це сучасна наукова дисципліна, яка використовує методологію загальної теорії систем для дослідження проблем у різних галузях, таких як економіка, менеджмент, інженерія та виробництво. Однією з восьми основних цілей системного аналізу є пошук рішень проблемних ситуацій, що виникають у сфері аналізу.

Процес системного аналізу використовує принципи теорії систем, системні підходи, дослідження операцій, кібернетики та інструменти системної інженерії. Його мета - прояснити проблему, щоб менеджери могли чіткіше розуміти наслідки своїх рішень. Це допомагає менеджерам робити більш обґрунтований вибір найкращої стратегії та враховувати невизначеності і ризики.

Таким чином, системний аналіз є потужним управлінським інструментом для формулювання стратегій, вибору найбільш прийнятних рішень і досягнення успіху в сучасному бізнес-середовищі.

1.2 Роль системного аналізу в прийнятті управлінських рішень

Хто є споживачами цього продукту? Відповіді на це питання можуть бути різними. Часто існують альтернативи, які важко порівняти між собою, оскільки одиниці виміру не стандартизовані або, навіть якщо вони стандартизовані, розрахунки можуть бути умовними і відхилятися від реальності. Системний аналіз спрямований на порівняння варіантів рішень і вибір найбільш прийняттого з них. Метою є повне і всебічне порівняння різних можливостей вирішення проблеми з урахуванням якісних і кількісних показників витрачених ресурсів.

Застосування сучасного системного аналізу в управлінні дозволяє глибоко досліджувати причинно-наслідкові зв'язки, оцінювати альтернативні варіанти вирішення системних проблем з урахуванням обмежень, ризиків і невизначеностей середовища. Завдяки цьому підходу створюються обґрунтовані рекомендації, які допомагають у виборі оптимальної стратегії управління навіть у складних і динамічних умовах.

1.3 Основні методи та підходи системного аналізу

Методи моделювання є невід'ємною частиною системного аналізу. Вони дають змогу детально вивчати проблеми, прогнозувати наслідки рішень і розробляти ефективні сценарії дій. Завдяки моделюванню управлінці можуть краще оцінювати можливі ризики та підвищувати якість ухвалених рішень.

Процес постановки задачі прийняття рішень включає кілька послідовних етапів:

1. **Визначення меж системи**, яка потребує оптимізації. Це дозволяє чітко окреслити область дослідження та уникнути надмірної складності.
2. **Вибір критерію ефективності**, за допомогою якого буде оцінюватися якість управлінських рішень.

3. **Оцінка внутрішніх показників**, що адекватно описують функціонування системи та підлягають аналізу.

4. **Побудова моделі**, яка відображає взаємозв'язки між обраними показниками, а також вплив незалежних факторів на результативність системи.

1.4 Процес ухвалення рішень в управлінні

У сфері бізнесу особливо важливо мати обґрунтовані управлінські рішення. Процес вибору такого рішення починається із виявлення проблемної ситуації та завершується ухваленням оптимального варіанту з наявних альтернатив. Обране рішення має бути спрямоване на ефективне вирішення поставленої проблеми.

Роль керівника стає особливо важливою у ситуаціях, коли система стикається з невизначеністю, невідомістю чи ризиками. У таких випадках він повинен використовувати доступні інструменти, щоб знизити рівень невизначеності та підвищити впевненість у правильності вибору.

Рациональність управлінських рішень залежить не лише від результату, а й від самого процесу їхньої підготовки та прийняття. Правильна організація цього процесу сприяє зниженню ризиків, оптимізації ресурсів та досягненню поставлених цілей.

У теорії менеджменту будь-який процес прийняття управлінських рішень складається з наступних етапів:

- 1) Підготовка та прийняття рішення
- 2) Реалізація рішення
- 3) Оцінка результатів

1.5 Використання системного аналізу для зменшення ризиків

Розглядаючи підхід системного аналізу, можна визначити, що прийняття рішень вимагає декількох етапів. Перший етап передбачає аналіз проблеми, визначення системи, вивчення структури системи, визначення загальних цілей і критеріїв системи, декомпозицію цілей і визначення ресурсів, необхідних для вирішення проблеми.

У процесі вирішення економічної проблеми першим кроком є визначення, чи є вона реальною, а не уявною. Точність формулювання проблеми має вирішальне значення. Наприклад, проблему можна сформулювати як: «нестача програмістів для виконання проєкту».

Важливо структурувати проблему, дослідити її динаміку у часі та виявити можливі взаємозв'язки з іншими проблемами. Якщо такі зв'язки існують, необхідно зрозуміти їхній характер і визначити, чи можна проблему вирішити в принципі.

На цьому етапі слід виконати такі кроки:

1. Визначити сутність проблеми та її причинно-наслідкові зв'язки:
 - З'ясувати, хто відповідальний за розв'язання проблеми.
 - Встановити обставини, за яких необхідно ухвалити рішення.
 - Визначити термін розв'язання проблеми та з'ясувати, чи потребує вона вирішення взагалі.
2. Оцінити тип проблеми:
 - Зрозуміти, чи є проблема системною (тобто постійною) чи разовою.
3. Сформулювати кінцевий результат:
 - Чітко визначити, який результат очікується від розв'язання проблеми.
4. Встановити критерії та обмеження:
 - Визначити ознаки, за якими оцінюватиметься розв'язання проблемної ситуації.
 - Розставити ці ознаки за пріоритетністю, щоб забезпечити ефективну оцінку результатів.

Цей підхід дозволяє максимально чітко окреслити проблему, оцінити її значущість і знайти оптимальний шлях до її вирішення.

На цьому етапі необхідно визначити деталі завдання та ідентифікувати об'єкти та їхні складові, щоб визначити обсяг проблеми, яку потрібно вирішити. При аналізі структури системи важливо ієрархізувати рівні проблеми, врахувати проблеми багатовимірною характеру та визначити підсистеми і процеси, якщо такі є.

При аналізі структури системи, якщо проблема є багатогранною, необхідно встановити ієрархію рівнів і створити специфікацію підсистем і процесів проблеми, якщо такі є. Потім встановлюються загальні цілі, розробляються критерії оцінки проблеми та розподіляються на підсистеми. На цьому етапі цілі ранжуються зверху вниз, розробляються поточні та майбутні цілі і визначаються потреби в ресурсах. Оцінюються ресурси, необхідні для досягнення цілей, аналізуються існуючі технології та ресурси, а також оцінюється стан проектів, які вже реалізовані або заплановані.

На основі аналізу формуються альтернативні варіанти вирішення проблеми. Потім всі альтернативи оцінюються, і на основі цієї оцінки приймається рішення шляхом вибору найкращого варіанту. У системному аналізі проблеми поділяються на три категорії: структуровані (кількісно сформульовані), неструктуровані (якісно виражені) і неструктуровані (змішані). Визначення критеріїв оцінки системи є важливим для прийняття управлінських рішень. Це пов'язано з тим, що системи часто є нетривіальними і вимірюють комплексний вплив рішень на зацікавлені сторони.

Стейкхолдери - це люди, які впливають на бізнес-рішення компанії, їх можна поділити на «внутрішніх» (власники, працівники) та «зовнішніх» (споживачі, постачальники, конкуренти, державні органи, некомерційні організації, ЗМІ). Вирішення проблем у системному аналізі передбачає декомпозицію проблеми на нові або відомі завдання. Вирішення економічних проблем вимагає отримання та обробки експертних оцінок. У випадку неструктурованих проблем саме визначення проблеми стає складним. Системний аналіз в економічному управлінні часто має справу з неструктурованими або неструктурованими проблемами. Тому отримання та обробка експертних оцінок є важливою частиною системного аналізу більшості економічних проблем.

На етапі проектування формується комплекс заходів, моделей і проектів, необхідних для ефективного вирішення поставленої проблеми. Визначається порядок досягнення різних цілей, проводиться розподіл сфер діяльності та

компетенцій між учасниками процесу. Усі ці дії виконуються з урахуванням наявних тимчасових і ресурсних обмежень.

На цьому етапі важливо не лише визначити, як система буде функціонувати, але й врахувати її взаємодію із зовнішнім середовищем. Особливу роль відіграє побудова моделей, які дозволяють формалізувати зв'язок між цілями, засобами та очікуваними результатами. Методи моделювання є ключовим інструментом цього процесу, оскільки вони:

Дозволяють детально досліджувати систему.

Забезпечують перевірку пропонованих рішень.

Сприяють виявленню нових, потенційно ефективних варіантів розв'язання проблеми.

Після завершення аналізу проблеми розробляються різні альтернативи для її вирішення. Цей процес передбачає створення максимально можливої кількості варіантів. Такий підхід не лише зменшує ризик використання шаблонних рішень, але й забезпечує всебічне порівняння альтернатив. Це дозволяє обрати оптимальний шлях для вирішення проблеми, враховуючи її специфіку та наявні обмеження.

Важливою частиною процесу прийняття рішень в управлінні є завдання вибору найкращого серед оцінених та проаналізованих варіантів. Цей вибір ґрунтується на порівнянні різних альтернатив за різними параметрами, визначеними в процесі моделювання. Оцінка включає такі показники, як ефективність рішення, час, витрачений на прийняття рішення, витрати ресурсів, очікувані результати та ризик невдачі.

На етапі розробки проекту також важливо враховувати актуальність соціальних та екологічних міркувань і включати можливі додаткові сценарії рішень, щоб забезпечити гнучкість у реагуванні на мінливі умови.

Пошук та формування альтернативних рішень для розв'язання поставленої проблеми є творчим процесом, який потребує розгляду різноманітних можливостей. Чим більше альтернатив розглядається, тим вища ймовірність знайти оптимальне рішення в межах заданих обмежень.

Проте повне порівняння всіх можливих варіантів є складним завданням через значні витрати ресурсів і часу. Тому зазвичай аналіз обмежується розглядом кількох найоптимальніших альтернатив, відібраних на основі попередньої оцінки.

У процесі вибору найкращого рішення та його обговорення необхідно враховувати такі ключові показники:

- **Витрати ресурсів** (фінансові, людські, матеріальні).
- **Час**, необхідний для реалізації рішення.
- **Рівень ризиків**, пов'язаних із кожною альтернативою.
- **Очікувані наслідки** прийнятого рішення (як позитивні, так і потенційно негативні).

Залежно від цих параметрів певні альтернативи можуть бути кращими за інші. Важливо зазначити, що раціональне рішення проблеми може вимагати значних ресурсів і тривалого часу, однак його доцільність має бути підтверджена відповідним аналізом і прогнозом результатів.

Цей підхід дозволяє забезпечити зважений і обґрунтований вибір, який максимально відповідає умовам та обмеженням конкретної ситуації.

Умовність у визначенні оптимального рішення пов'язана з тим, що рідко існує єдиний критерій оптимальності. У більшості випадків при виборі серед багатьох критеріїв перевага надається тому прийнятному варіанту, який здається найкращим для особи, що приймає рішення, з певної точки зору.

Після того, як визначено найкращі варіанти, їх реалізацію можна делегувати іншим відділам або проектним групам всередині компанії. На етапі впровадження зазвичай використовуються методи планування, організації та управління, щоб забезпечити ефективну реалізацію рішень.

План імплементації включає визначення дій - що робити, з ким, як, де і коли. Ці дії повинні бути чітко задокументовані. На етапі управління впровадженням використовуються різні методи, включаючи управління проміжними і кінцевими результатами та дотримання графіка впровадження. Основна мета системи управління полягає в тому, щоб забезпечити виконання рішень і найвищу якість виконання.

На етапі аналізу наслідків рішення здійснюється оцінка його ефективності та результативності стосовно поставленої мети. Ефективні рішення дозволяють досягти цілей з мінімальними витратами ресурсів, забезпечуючи оптимальний баланс між витратами та результатами.

Цей етап передбачає врахування як кількісних, так і якісних наслідків реалізованого рішення:

- Кількісні показники включають витрати, економію ресурсів, продуктивність та інші числові метрики.
- Якісні аспекти стосуються таких чинників, як задоволеність зацікавлених сторін, соціальний вплив та довгострокові перспективи.
 - Ключовими діями на цьому етапі є:
- Здійснення зворотного зв'язку з усіма зацікавленими сторонами для отримання їхньої оцінки результатів.
- Обчислення показників ефективності для об'єктивного порівняння фактичних результатів із початковими цілями.

Процес ухвалення управлінських рішень є складним і багатогранним, оскільки на нього впливають численні фактори, такі як:

- Рівень ризику, пов'язаний із кожною альтернативою.
- Часові рамки, встановлені для прийняття та реалізації рішення.
- Політика організації, яка може обмежувати вибір певних рішень або встановлювати додаткові критерії.

Важливим аспектом є інтеграція аналізу наслідків у загальний процес управління. Це забезпечує не лише оцінку вже прийнятих рішень, але й формує підґрунтя для прийняття більш ефективних рішень у майбутньому.

Етап аналізу результатів рішення передбачає детальну оцінку його ефективності з урахуванням початкової постановки завдання. Визнано, що ефективні рішення - це ті, які досягають найбільших цілей з найменшими витратами ресурсів. Важливо враховувати якісні та кількісні наслідки рішень, залучати зацікавлені сторони до обговорення, забезпечувати зворотний зв'язок та розраховувати ключові показники ефективності. На процес прийняття рішень

впливає низка факторів, зокрема рівень ризику, час, доступний для прийняття рішення, та політика організації.

Успіх у бізнесі сьогодні обумовлений не тільки дотриманням обраного напрямку розвитку компанії та маркетингової стратегії, але й технічною підтримкою прийняття рішень, автоматизацією управління компанією та ефективною роботою інформаційних систем.

Різні бізнес-процеси потребують автоматизації для спрощення операцій. Однією з таких ключових технологій є використання систем управління взаємовідносинами з клієнтами (CRM), де CRM-менеджери вивчають взаємовідносини між клієнтом і компанією з метою забезпечення найвищого рівня обслуговування та лояльності клієнтів. CRM-система автоматизує взаємовідносини з клієнтами, створює і управляє єдиною базою даних клієнтів і бізнес-партнерів, ефективно контролює роботу відділу продажів, отримує аналітичні дані і надає можливості для поліпшення взаємодії з клієнтами.

CRM-фахівці використовують різні програмні інструменти, в тому числі CRM-системи, для досягнення ефективного CRM. Ці інструменти автоматизують бізнес-процеси і дають можливість усім підрозділам компанії взаємодіяти з клієнтами відповідно до концепції CRM. Ці системи не тільки вирішують завдання задоволення та утримання клієнтів, але й оптимізують діяльність компанії за рахунок скорочення витрат на пошук і обробку інформації, управління продажами та інші бізнес-процеси.

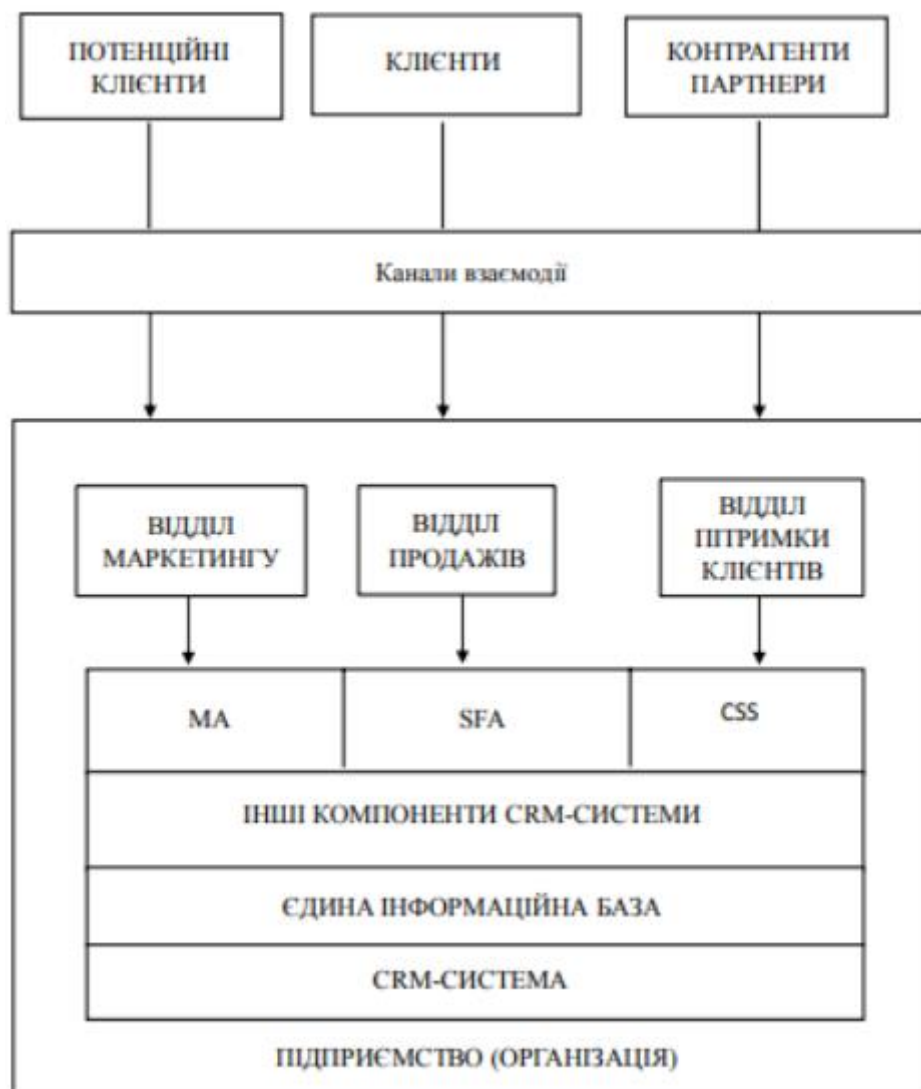


Рис. 1. CRM-система використовується для управління взаємодією між компанією (організацією) та її клієнтами.

Взаємодія клієнтів із CRM-менеджерами здійснюється через різні канали зв'язку, серед яких електронна пошта, телефонні дзвінки, SMS-повідомлення, факс, особисті зустрічі, месенджери (наприклад, ICQ), корпоративні веб-сайти, точки продажу та соціальні мережі. Така багатоканальність забезпечує гнучкість у комунікаціях і дозволяє ефективніше задовольняти потреби клієнтів.

Основні функціональні компоненти сучасних CRM-систем включають:

1. Модуль SFA (Sales Force Automation) – автоматизація роботи торгових представників, що сприяє підвищенню продуктивності відділу продажів.

2. Модуль MA (Marketing Automation) – автоматизація маркетингових процесів, яка дозволяє ефективніше планувати та виконувати кампанії.

3. Модуль CSS (Customer Service & Support) – автоматизація обслуговування та підтримки клієнтів для забезпечення їхнього задоволення.

Після завершення будь-якої акції чи маркетингового заходу обов'язковим є проведення аналізу їхньої ефективності. Це дозволяє порівняти витрати з досягнутими результатами, виявити помилки та підготувати більш ефективні стратегії для майбутніх кампаній.

Розробка універсальної формалізованої моделі управління взаємодією з клієнтами стикається зі значними труднощами через складність і низьку структурованість цього процесу. Проте підвищення ефективності можливо завдяки впровадженню сучасних систем підтримки прийняття рішень (DSS). Ці системи сприяють обробці великих обсягів даних, покращують якість рішень і забезпечують інтеграцію багатоканальної комунікації з клієнтами.

Системний аналіз має вирішальне значення для ефективного управління підприємствами, компаніями та організаціями. Системний аналіз поєднує в собі різні методи дослідження складних соціально-економічних систем, виявлення проблем і планування дій для їх вирішення. Тому всі управлінські рішення повинні ґрунтуватися на результатах системного аналізу. Знання менеджерів про етапи процесу прийняття управлінських рішень і про те, як забезпечити якість рішень, що приймаються на кожному етапі, з урахуванням ситуаційних особливостей і факторів, що впливають на формування і реалізацію рішень, сприяє підвищенню загальної ефективності управління організацією.

РОЗДІЛ 2. ХАРАКТЕРИСТИКА ТА ПРИНЦИПИ РОБОТИ З ВЕЛИКИМИ ДАНИМИ (BIG DATA)

2.1 Визначення поняття "Big Data"

Великі дані (Big Data) — це величезні обсяги структурованої та неструктурованої інформації, яка характеризується різноманітністю та постійним зростанням. Для їхньої обробки використовуються спеціалізовані програмні інструменти, які здатні горизонтально масштабуватися, забезпечуючи ефективну роботу з такими даними. Це поняття виникло наприкінці 2000-х років і стало альтернативою традиційним системам управління базами даних та класичним рішенням у сфері бізнес-аналітики.

Розвиток глобального обсягу даних

Згідно з даними аналітиків компанії IBS, обсяги даних у світі демонструють стрімке зростання:

- 2003 р. — 5 ексабайт даних (1 ЕБ = 1 мільярд гігабайтів).
- 2008 р. — 0,18 зетабайт (1 ЗБ = 1024 ексабайти).
- 2015 р. — понад 6,5 зетабайт.
- 2021 р. — 44–48 зетабайт.
- 2025 р. — прогнозується зростання у 10 разів порівняно з обсягом 2021 року.

Основні характеристики та операції Big Data

Великі дані включають сукупність технологій, які покликані виконувати три ключові завдання:

1. Обробка великих обсягів даних

- Big Data дозволяє працювати з обсягами даних, які значно перевищують стандартні сценарії.

2. Робота з поточними даними

- Системи обробляють дані, що постійно надходять у великих обсягах, і можуть ефективно функціонувати в умовах швидкого зростання інформації.

3. Універсальність обробки даних

- Інструменти Big Data здатні одночасно працювати як зі структурованими, так і з неструктурованими даними, забезпечуючи їхній аналіз у різних аспектах.

Big Data сьогодні є невід'ємною частиною багатьох галузей, від бізнесу до науки, забезпечуючи нові можливості для аналізу, прогнозування та прийняття рішень у масштабах, недосяжних для традиційних технологій.

Вважається, що ці функції виявляють приховані закономірності за межами людського сприйняття і пропонують безпрецедентні можливості для оптимізації різних секторів, включаючи державне управління, охорону здоров'я, телекомунікації, фінанси, транспорт і виробництво.

2.2 Основні характеристики великих даних (Volume, Velocity, Variety, Veracity)

Основні принципи роботи з великими даними визначаються характеристиками VVV (фізична кількість, темпи зростання та різноманітність). Цей підхід був розроблений компанією Meta Group у 2001 році і підкреслює важливість усіх аспектів управління даними. З часом з'явилися й інші інтерпретації, такі як чотири V (адитивність, надійність), п'ять V (застосовність, цінність) і сім V (варіативність, видимість). Наприклад, IDC трактує четверту V як цінність і наголошує на економічній доцільності обробки великих обсягів даних у кожному конкретному випадку [3].

2.3 Розвиток технологій обробки великих даних

Основні принципи взаємодії з великими даними забезпечують ефективну та надійну роботу з інформацією у великих масштабах. До таких принципів належать:

1. Горизонтальна масштабованість

Це ключовий принцип обробки великих даних. Зі зростанням обсягів даних необхідно збільшувати кількість обчислювальних вузлів, між якими розподіляється обробка інформації. Горизонтальна масштабованість дозволяє додавати нові вузли до кластеру без погіршення продуктивності системи,

забезпечуючи її ефективну роботу навіть у випадках значного зростання обсягів даних.

2. Відмовостійкість

Відмовостійкість тісно пов'язана з принципом масштабованості. Оскільки великі обчислювальні кластери можуть містити тисячі вузлів, ймовірність виходу з ладу одного або декількох із них зростає. Системи для роботи з великими даними повинні враховувати ці ризики та забезпечувати механізми для автоматичного виявлення, ізоляції та відновлення роботи після таких збоїв. Це гарантує безперервність обробки даних і захищає від втрати інформації.

3. Локальність даних

Для зменшення витрат на передачу даних між вузлами обчислювального кластеру обробка повинна проводитися максимально близько до місця зберігання даних. Якщо дані фізично знаходяться на одному сервері, а обробляються на іншому, це може призвести до значних витрат на передачу інформації. Локальність даних дозволяє мінімізувати такі витрати, підвищуючи швидкість і ефективність обробки.

Ці принципи забезпечують оптимізацію роботи з великими даними, дозволяючи системам залишатися продуктивними, надійними та економічно ефективними навіть за умов стрімкого зростання обсягів інформації.

Методи і техніки аналізу великих даних представлені широким розмаїттям інструментів і підходів. Компанія McKinsey, що займається стратегічним управлінням, виділила 11 ключових методів і прийомів, які застосовуються до великих обсягів різноманітних даних. Важливою частиною цих методів є інтелектуальний аналіз даних, який передбачає виявлення раніше невідомої інформації в даних, наприклад, вивчення правил асоціацій, класифікацію, кластерний аналіз і регресійний аналіз. Крім того, краудсорсинг використовується для класифікації та збагачення даних великою кількістю людей, які не перебувають у трудових відносинах.

Для більш поглибленого аналізу використовуються методи злиття та інтеграції даних, такі як цифрова обробка сигналів і обробка природної мови.

Машинне навчання, яке включає як контрольоване, так і неконтрольоване навчання, будує моделі на основі статистичного аналізу і генерує прогнози на основі базової моделі. Інші методи включають штучні нейронні мережі, мережевий аналіз та оптимізацію за допомогою генетичних алгоритмів, які використовуються для вирішення проблем оптимізації та моделювання шляхом випадкового вибору та комбінування параметрів.

Принципи обробки великих даних включають горизонтальну масштабованість, відмовостійкість і локальність даних. Системи великих даних повинні бути масштабованими, відмовостійкими та враховувати локальність даних, щоб зменшити витрати на передачу даних. Додатки для роботи з великими даними активно використовуються у фінансовому та медичному секторах, інтернет-компаніях та державному управлінні.

На сучасному ринку також відбулися значні зміни у сфері послуг з обробки великих даних. Наприклад, з 2012 року Google BigQuery, хмарний сервіс, що пропонує аналіз великих даних у режимі реального часу, успішно використовується в бізнес-аналітиці. Інші провідні компанії також пропонують послуги, засновані на аналітиці великих даних, які вони використовують для фільтрації спаму, прогнозування трафіку, безпілотного керування транспортними засобами та інших інноваційних застосувань.

Великі дані активно використовуються в різних секторах, особливо у фінансах, охороні здоров'я, технологіях, інтернет-компаніях та уряді.

Розвиток сервісів Big Data став особливо помітним з початку 2010-х років. У 2012 році компанія Google вивела на ринок бізнес-аналітики хмарний сервіс Google BigQuery, який надає можливість аналізувати великі дані в режимі реального часу. Через рік цей сервіс був успішно інтегрований із Google Analytics Premium, платною версією популярного веб-аналітичного інструменту.

Нещодавно Google представила ще один масштабований хмарний сервіс баз даних — Cloud Bigtable, який розширив портфель продуктів компанії, орієнтованих на обробку великих обсягів даних.

Провідні технологічні компанії, які спеціалізуються на Big Data, розробляють численні сервіси, що базуються на застосуванні сучасних методів аналізу великих даних. Ці технології охоплюють широкий спектр застосувань, зокрема:

- Пошукові алгоритми на основі нейронних мереж.
- Машинний переклад для автоматизації перекладів між різними мовами.
- Фільтрація спаму для електронної пошти та інших цифрових платформ.
- Таргетинг у контекстній рекламі, який дозволяє компаніям точніше досягати своєї аудиторії.
- Прогнозування пробок і погодних умов для оптимізації транспортних і логістичних процесів.
- Розпізнавання мови та образів, що використовується у системах штучного інтелекту.
- Системи управління безпілотними автомобілями, які демонструють інноваційний підхід до транспорту.

Ці сервіси та технології не лише сприяють ефективній роботі з великими даними, але й відкривають нові можливості для інновацій та цифрової трансформації в різних галузях.

Аналіз переваг та обмежень використання технологій великих даних для бізнесу виявляє деякі важливі переваги, які слід враховувати при їх впровадженні:

1. Полегшення планування: великі дані спрощують процес планування і дозволяють компаніям більш ефективно визначати свої стратегії.
2. Швидкість запуску нових проектів: технологія великих даних прискорює запуск нових проектів завдяки швидкому аналізу великих обсягів даних.
3. Потенціал для збільшення попиту на проекти: Використовуючи великі дані, компанії можуть точніше прогнозувати, які проекти є популярними та затребуваними на ринку.

4. Оцінка задоволеності користувачів Аналіз даних дозволяє оцінити рівень задоволеності клієнтів і таким чином підвищити якість продуктів і послуг.

5. Спрощення пошуку та залучення цільової аудиторії: Застосування Big Data спрощує і поліпшує стратегії залучення та утримання цільової аудиторії

6. Спрощення пошуку та залучення аудиторії: використання великих даних спрощує та вдосконалює стратегії охоплення та утримання цільової аудиторії.

7. Оптимізація інтеграції ланцюгів поставок Великі дані допомагають оптимізувати та покращити інтеграцію ланцюгів поставок, забезпечуючи ефективну логістику та управління запасами.

8. Покращення якості обслуговування клієнтів та прискорення діалогу: аналітика великих даних може допомогти покращити якість обслуговування клієнтів та прискорити взаємодію з ними.

9. Підвищення лояльності існуючих клієнтів Великі дані підвищують лояльність, дозволяючи компаніям зрозуміти потреби та очікування своїх клієнтів.

Використання технологій великих даних у маркетингу відкриває значні можливості для бізнесу:

1. Розуміння вашого бізнесу в цифрах: великі дані можуть допомогти вам детально проаналізувати фінансові показники та визначити ефективні стратегії;

2. Дослідження конкурентів: аналіз даних може дати уявлення про поведінку конкурентів і дозволить вам адаптувати вашу стратегію;

3. Дослідження клієнтів: великі дані можуть забезпечити детальний аналіз поведінки та потреб клієнтів;

4. Підвищення рівня розуміння та аналізу в маркетингу: технології великих даних дозволяють маркетологам створювати більш точні та детальні профілі споживачів.

5. Точне уявлення про цільових споживачів: великі дані дозволяють маркетологам створювати більш точні та детальні профілі споживачів;

6. Прогнозування реакції споживачів на маркетингові повідомлення: великі дані дозволяють маркетологам створювати більш точні та детальні профілі

споживачів. Прогнозування реакції споживачів на маркетингові повідомлення: великі дані можна використовувати для прогнозування реакції споживачів на різні маркетингові повідомлення.

7. Прогнозування реакції споживачів на різні маркетингові повідомлення: прогножуючи реакцію споживачів на різні маркетингові повідомлення, компанії можуть ефективно адаптувати свої стратегії та досягти більшої ефективності комунікації

8. Максимальна персоналізація рекламних повідомлень: використовуючи великі дані, компанії можуть враховувати минулу історію покупок, вподобання та поведінку кожного клієнта, щоб створювати високо персоналізовані повідомлення

9. Збільшення перехресних продажів, повторних покупок і ремаркетингу: технології великих даних можна використовувати для ефективного впливу на клієнтську базу з метою збільшення перехресних продажів, стимулювання повторних покупок і підвищення успішності ремаркетингових кампаній, проведення успішних ремаркетингових кампаній.

10. Аналіз та виявлення причин популярності популярних продуктів та товарів: великі дані можуть бути використані для більш глибокого аналізу попиту та виявлення факторів, що впливають на популярність конкретних продуктів та товарів.

11. Вдосконалення продуктів і послуг та підвищення лояльності клієнтів: Використовуючи аналітику великих даних, компанії можуть постійно вдосконалювати свої продукти та послуги, враховуючи реальні потреби та вимоги своїх клієнтів, що призводить до підвищення лояльності;

12: Використання технологій великих даних покращує процеси обслуговування, роблячи їх більш ефективними та клієнтоорієнтованими.

13. Зниження витрат на взаємодію з постачальниками та клієнтами: Використання великих даних спрощує та оптимізує процес взаємодії з постачальниками та клієнтами, а також зменшує витрати на цих етапах. Економії витрат можна досягти на цих етапах.

Обсяг існуючої літератури з управління великими даними та системної аналітики дозволяє поглиблено вивчати різні аспекти цієї теми. Особливої уваги заслуговують дослідження, присвячені стратегіям оптимізації управління великими даними та використанню системної аналітики для вирішення цієї проблеми. Порівняльний аналіз результатів Результати аналізу літератури показали, що більшість досліджень зосереджені на технічних аспектах та аналізі великих даних. Однак обмежена кількість досліджень вказує на необхідність більш глибоких досліджень системного аналізу як стратегічного інструменту оптимізації управління великими даними в організаціях.

Аспекти, які ще не вивчені детально

1) Інтеграція системного аналізу в існуючі підходи до управління великими даними: не всі дослідження розглядають взаємодію системного аналізу з існуючими стратегіями управління великими даними;

2) Адаптація системного аналізу до реального середовища великих даних: деякі дослідження є дуже теоретичними;

3) Адаптація системного аналізу до реального середовища великих даних: деякі дослідження є занадто теоретичними. Бракує досліджень, що досліджують застосування концепцій системного аналізу в реальному виробничому середовищі;

4) Визначення найкращої стратегії впровадження системного аналізу: бракує комплексного підходу до визначення етапів впровадження системного аналізу в організації.

2.4 Використання великих даних у різних галузях

Інтеграція системного аналізу в сучасні підходи до управління великими даними означає використання концепцій і методів системного аналізу як фундаментального інструменту для вдосконалення процесу управління та аналізу великих даних в організації. Розглянемо цю тему трохи детальніше:

1. Системи та підсистеми: великі дані можна розглядати як складну систему, що складається з різних компонентів (підсистем), включаючи джерела даних, інструменти аналізу та людський фактор.

2. Взаємозв'язки та взаємодії: аналіз великих даних вимагає розуміння взаємозв'язків і взаємодій між різними частинами системи, чого можна досягти за допомогою методів системного аналізу.

3. Моделювання процесів: Системний аналіз дозволяє створювати моделі процесів управління біг даними, що полегшує їхнє розуміння та оптимізацію.

4. Виявлення взаємозв'язків: Системний аналіз можна використовувати для виявлення складних взаємозв'язків і залежностей між різними частинами системи управління великими даними.

5. Інтегрований підхід Системний аналіз дозволяє розглядати управління великими даними як інтегровану систему.

6. Ефективність і стратегічний підхід: системний аналіз може бути використаний для розробки більш ефективного і стратегічного підходу до управління великими даними.

Недосліджені аспекти. Адаптація системного аналізу до вимог великих даних: деякі аспекти, такі як адаптація традиційних методів системного аналізу до великих даних, ще не повністю вивчені. Роль системного аналізу в розробці стратегій роботи з великими даними: важливо розглядати системний аналіз як ключовий елемент у визначенні та формулюванні стратегій роботи з великими даними. Включення системного аналізу в управління великими даними уможливорює більш гнучкі та ефективні стратегії, які враховують складнощі та взаємодію в цьому динамічному середовищі. Для ефективного управління важливо проактивно аналізувати і виявляти всі невикористані можливості. У широкому розумінні управління - це вплив, який суб'єкт управління здійснює на об'єкт управління для досягнення поставлених цілей.

Науковою основою менеджменту є його принципи та методи. До принципів належать науковість, безперервність, гнучкість і цілеспрямованість. На практиці управління здійснюється за допомогою методів. До методів належать економічні, психологічні та адміністративні (розпорядчі) методи. Всі методи тісно взаємопов'язані і взаємозалежні. Роль економічних методів управління значно зросла з переходом економіки України до світового ринку і зростанням її

незалежності. Ці методи дають можливість суб'єктам господарювання робити власний вибір, мобілізувати творчу активність і підприємливість у межах існуючих форм і способів діяльності. Роль системи управління безпосередньо пов'язана з кінцевими результатами діяльності підприємства, тому слід використовувати всі можливості для досягнення позитивних результатів. На мою думку, саме систематичний та всебічний аналіз може вказати на вирішення проблем та прийняття правильних управлінських рішень задля власного добробуту та подальшого процвітання підприємства.

Комплексний економічний аналіз є системним дослідженням господарської діяльності, яке розглядає всі її аспекти у взаємозв'язку та взаємообумовленості. Він забезпечує достовірну оцінку діяльності підприємства та дозволяє виявити наявні резерви, враховуючи кінцеві результати.

Комплексний економічний аналіз охоплює всі аспекти інформаційної моделі виробничо-фінансової діяльності підприємств. Ця інформаційна модель є ключовим інструментом для формування даних, необхідних для аналізу, зокрема:

- Нормативні, облікові та звітні показники відповідно до змін в організаційно-технічному рівні виробництва.
- Використання засобів і предметів праці, що впливають на ефективність виробничих процесів.
- Трудові ресурси, їхній розподіл і продуктивність.
- Обсяги виробництва та реалізації продукції, що відображають ринкову активність підприємства.
- Виробничі витрати та собівартість продукції, які є основою для оцінки рентабельності.
- Прибуток і рентабельність продукції, як показники фінансової ефективності.
- Відтворення та обіг виробничих фондів, що забезпечують стійкість виробничих процесів.
- Рентабельність виробництва, що характеризує загальну ефективність діяльності підприємства.

- Фінансовий стан, який відображає платоспроможність та ліквідність підприємства.
- Економічне стимулювання, спрямоване на підвищення продуктивності та ефективності діяльності.

Таким чином, інформаційна модель дозволяє отримувати повну та актуальну картину діяльності підприємства, забезпечуючи аналітичну базу для прийняття управлінських рішень та підвищення економічної ефективності.

Наступним кроком у комплексному підході є порівняння результатів діяльності вашої компанії з аналогічними компаніями. Це дозволить компанії визначити свій рівень конкурентоспроможності та скоригувати власний менеджмент. Такий підхід особливо важливий в контексті інтеграції в європейський ринок.

Сучасна ситуація вимагає використання методів системного аналізу, тобто проблем, заснованих на розгляді сукупності взаємопов'язаних елементів як сукупності взаємопов'язаних елементів.

При використанні комплексного підходу особливу увагу потрібно приділити системі показників і методів, які використовуються для аналізу та вирішення конкретних проблем. Адже основними проблемами сучасних підприємств є високі витрати та низька рентабельність, а іноді навіть збиткова діяльність. Тому найбільшу увагу, на нашу думку, слід приділяти побудові системи показників, які мають ланцюговий взаємозв'язок між собою і, як наслідок, здатні без зайвих витрат привести підприємства у відповідність до світових ринкових стандартів та забезпечити довгострокову конкурентоспроможність своїх підприємств. Успішне створення системи показників дозволить виявити невикористані резерви та використати їх таким чином, щоб вони не призводили до зайвих витрат, а, навпаки, приносили певну вигоду. Це дозволить підприємствам поступово перебудовувати всю систему планомірно і без значних фінансових втрат.

При проведенні due diligence особливу увагу слід приділяти якості та достовірності інформації. Сучасні інформаційні потоки, особливо в Інтернеті,

часто містять неперевірену інформацію, використання якої може спрямувати керівництво в хибному напрямку.

Для того, щоб виконати якісну, компетентну роботу, необхідний певний обсяг кількісно обмеженої інформації. Якщо ця умова не виконується, виникають дві аномальні ситуації:

1) недостатня кількість інформації не дає повного уявлення про досліджувану ситуацію і призводить до інтуїтивного добору можливих варіантів;

2) надмірна кількість інформації знижує швидкість обробки і розуміння, вимагаючи додаткових витрат часу і зусиль;

3) недостатня кількість інформації не дає повного уявлення про досліджувану ситуацію і призводить до інтуїтивного добору можливих варіантів, вимагаючи додаткових витрат часу і зусиль.

У таких ситуаціях не завжди вдається визначити ключові проблеми, виявити закономірності та тенденції, встановити причинно-наслідкові зв'язки, що може призвести до помилкових висновків або неправильного способу врегулювання кризової ситуації. У всіх випадках інформаційний потік має бути якісним, достовірним, перевіреним і правдивим. Якість інформації оцінюється за чотирма критеріями повнота (мінімальна кількість і репрезентативність), достовірність, своєчасність отримання і потенційна корисність для потреб аналізу.

Сучасний менеджмент вимагає відповідних знань щодо зберігання та захисту інформації. Відібрані для використання джерела інформації підлягають формальній та змістовній перевірці. Формальна перевірка визначає точність документів і звітів з точки зору наявності та повноти реквізитів, таблиць і підписів, а також перевіряє узгодженість сум, підсумків і взаємопов'язаних величин. Аудит по суті визначає якість матеріалу, його відповідність об'єктивній дійсності та потенціал для використання в аналізі. Це вимагає знання теорії та практики економічного аналізу, відповідної кваліфікації та практичного досвіду, а також технічних знань.

При створенні інформаційних потоків особливу увагу слід приділити організації локальної мережі, яка буде використовуватися для накопичення та

зберігання даних. Сучасна локальна мережа має відповідати низці вимог, які забезпечують її ефективність і надійність:

- Системність – інтеграція всіх елементів мережі в єдину систему, яка дозволяє комплексно обробляти дані.
- Комплексність – охоплення всіх аспектів діяльності підприємства.
- Оперативність – швидкий доступ до даних для забезпечення своєчасного аналізу.
- Точність – забезпечення коректності та достовірності інформації.
- Прогресивність – використання сучасних технологій для підвищення продуктивності.
- Динамічність – здатність адаптуватися до змін і зростаючих потреб підприємства.

Виконання цих вимог створює основу для:

1. Об'єктивного вивчення стану досліджуваного об'єкта та аналізу тенденцій його розвитку.
2. Систематичного підвищення ефективності господарської діяльності підприємства на основі результатів проведеного аналізу.

Така організація локальної мережі забезпечує ефективне управління інформаційними потоками, що є важливим елементом успішного функціонування сучасного підприємства.

При створенні бази знань для локальної мережі, по-перше, вона робиться простою у використанні. По-друге, чітко формулюється реальна потреба у вихідних даних. По-третє, усуваються повтори і полегшується групування в блоки для одночасної обробки. Крім того, формалізація завдання стає безпосередньо об'єктом математичного моделювання, що значно збільшує глибину роботи. По-перше, організація завдань зі створення та використання інформаційних потоків характеризується збереженням цілісності роботи в умовах децентралізованої обробки інформації. Це забезпечує об'єктивність і надійність роботи. При децентралізованій обробці інформації не порушується цілісність роботи і не порушується єдність цілей і завдань з точки зору її системного характеру. По-друге,

завдання вирішуються безпосередньо користувачем на його власному робочому місці. Фахівець має можливість особисто керувати всіма етапами процесу обробки інформації, оцінювати отримані результати, правильно використовувати їх для обґрунтування прийнятих рішень і задоволення різних інформаційних потреб системи.

Створення інформаційних потоків для системного аналізу вирішує наступні завдання:

- Знижує трудомісткість і вартість процесу
- Скорочує час обробки даних, підвищує їх якість і достовірність
- Забезпечує терміни виконання певних операцій
- Підвищує гнучкість і оперативність управління процесом
- Покращує організацію роботи користувача.

Це дає змогу:

- своєчасну передачу та отримання необхідної інформації, завдяки чому можна виявити зміни в ринковому попиті та пропозиції і вчасно відреагувати на них;
- доступ до всіх джерел інформації, які найбільш ефективно працюють у сфері перерозподілу продукції;
- специфічну, кількісно обмежену інформацію, що використовується в аналізі виробництва.

На мою думку, наведена вище методика формування інформаційних потоків дозволяє менеджерам зосередитися на конкретних можливостях підвищення економічної ефективності, підвищує якість управління та сприяє забезпеченню безперебійної роботи перерозподілу матеріальних потоків.

На основі інформаційної моделі складається загальна блок-схема комплексного аналізу, що відображає всі етапи виробництва і реалізації продукції. Спочатку за системою найважливіших показників дається попередня характеристика діяльності підприємства, потім проводиться детальний аналіз факторів і причин, а на завершальному етапі перевіряється формування і використання всіх ресурсів і дається комплексна оцінка діяльності підприємства.

РОЗДІЛ 3. ЗАСТОСУВАННЯ СИСТЕМНОГО АНАЛІЗУ В УПРАВЛІННІ ВЕЛИКИМИ ДАНИМИ

3.1 Взаємозв'язок системного аналізу та управління великими даними

Відповідно до Розділу 1, цілі дослідження систематизовані в таблиці 3.1 нижче.

Таблиця 3.1 – Завдання дослідження

№	Завдання дослідження	Пошукові питання	Метод отримання та джерело інформації	Формат отриманої інформації
1	Аналіз конкурентного середовища	Якою є щільність розподілу ціни на ноутбуки фірми HP в магазинах конкурентів?	Джерело: парсинг офіційних веб-сайтів компанії;	Кожна компанія буде мати свій розподіл ціни на ноутбуки і візуально відображена у презентації та описані характеристики розподілу
			Метод: побудова розподілу, розрахунок основних вибірових характеристик, Інтервальне оцінювання	
2	Огляд бізнес показників ТМ Comfy	Яка існує тенденція офлайн трафіку в роздрібні магазини Comfy, чи існує сезонність трафіку	Джерело: внутрішня інформація компанії	Графік загальної динаміки трафіку, окремий графік сезонності та тренду
			Метод: декомпозиція сезонності та тренду часового ряду офлайн трафіку	
		Чи залежить кількість трафіку від регіону	Джерело: внутрішня інформація компанії	Стовбчикова діаграма кількості трафіку між найбільшими містами України,
			Метод: однофакторний дисперсійний аналіз	

				розподіли трафіку по містах
3	Аналіз впливу медіа-активності на офлайн трафік	Яка існує залежність між офлайн трафіком та факторами які можуть на нього впливати?	Джерело: інформація зібрана протягом дослідження, внутрішні дані Метод: лінійна регресія та її модифікації	Графік впливу факторів на трафік, на якому буде виділено органічний рост мережі Comfy, сезонність, медіа-активність, вплив конкурентів. Для відображення впливу карантину буде побудовано гарфік, на якому буде зображено потенційна кількість трафіку якщо б не було карантинних обмежень та кількість яка була втрачена через пандемію
		Який існує вплив не медійних факторів на трафік?	Джерело: інформація зібрана протягом дослідження Метод: лінійна регресія та її модифікації	
		Який існує вплив медійних факторів на трафік?	Джерело: інформація зібрана протягом дослідження, внутрішні дані Метод: лінійна регресія та її модифікації	
		Як вплинув карантин на офлайн трафік?	Джерело: інформація зібрана протягом дослідження, внутрішні дані Метод: лінійна регресія та її модифікації	
4	Прогноз офлайн трафіку на 2022 рік	Який можна зробити прогноз офлайн трафіку на 2022 рік?	Джерело: інформація зібрана протягом дослідження, внутрішні дані	Графік динаміки офлайн трафіку на 2022 рік
			Метод: ансамблеві моделі	

У таблиці 1.2 наведені цілі опитування, запитання, методи збору інформації та формат, у якому інформація представлена наприкінці опитування.

Основні теоретичні положення методу, що використовується:

а). Розрахунок основних характеристик вибірки: - середнє значення, навколо якого групуються можливі значення випадкової величини; - ступінь

розсіювання цих значень відносно середнього значення; - асиметрія (або «асиметрія») щільності розподілу; - ексцес щільності розподілу [8].

б) Декомпозиція часових рядів, однофакторний дисперсійний аналіз: декомпозиція часового ряду - це виявлення трендів, циклічних коливань та випадкових коливань. Декомпозиція дозволяє зрозуміти структуру часових рядів, побудувати моделі часових рядів та спрогнозувати короткостроковий рівень часових рядів [9].

3.2 Основні підходи до інтеграції Big Data у стратегічне управління

Основна мета дисперсійного аналізу, базової концепції, запропонованої Фішером у 1920 році, полягає у вивченні значущості різниці між середніми значеннями різних груп даних або змінних [10]. В дисперсійному аналізі статистична значущість різниці між середніми значеннями декількох груп перевіряється на основі вибіркової дисперсії. Це виконується шляхом поділу загальної дисперсії (варіації) на частини, одна з яких є результатом випадкової помилки (тобто внутрішньої групової варіації), а інша - результатом різниці в середніх значеннях. Якщо ця різниця є значущою, то нульова гіпотеза про відсутність різниці між середніми відхиляється на певному рівні значущості. Графічна інтерпретація суті дисперсійного аналізу показана на рисунку 3.1

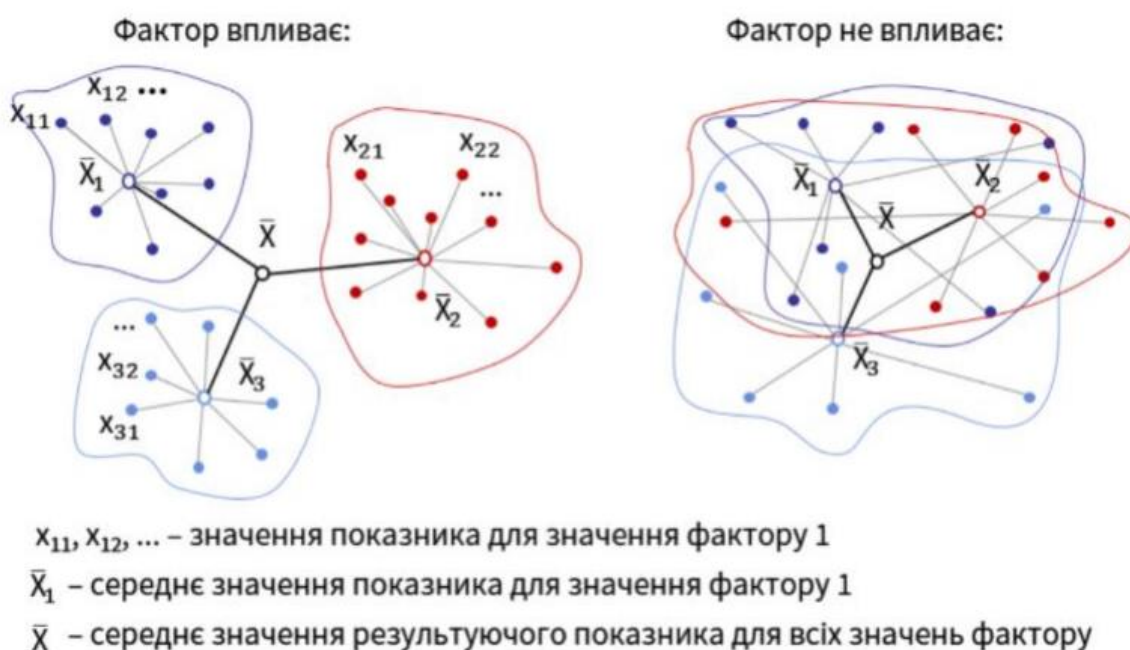


Рис. 3.1 – Дисперсійний аналіз

в) Лінійна регресія. Нехай маємо $\{y_i, x_{i1}, \dots, x_{ip}\}_{i=1}^n$ певний набір даних, де n – кількість елементів, y_i – залежна змінна, x_{ip} – фактори що впливають на залежну змінну, лінійна регресійна модель передбачає залежність між змінною Y та X є лінійною.

Загальна лінійна регресійна модель має наступний вигляд [11]:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_K x_K + u,$$

де u - випадкова похибка, розподіл якої в загальному випадку залежить від незалежних змінних але математичне сподівання якої дорівнює нулеві.

Залежно від теми, що вивчається в лінійній регресії, та конкретних цілей дослідження, для оцінювання невідомих параметрів можуть використовуватися різні методи. Найпоширенішим є звичайний метод найменших квадратів. Він бере за оцінку параметра значення, яке мінімізує суму квадратів залишків за всіма спостереженнями:

$$\beta = \arg \min_{\beta} \sum_{i=1}^n \left| y_i - \beta_0 - \sum_{j=1}^K X_{ij} \beta_j \right|^2 = \arg \min_{\beta} \|y - X\beta\|^2.$$

Для класичних лінійних регресійних моделей оцінка за методом найменших квадратів є незміщеною, значущою та найкращою лінійною незміщеною оцінкою (див. статтю про метод найменших квадратів для отримання додаткової інформації про ці статистичні властивості).

3.3 Методи оцінки ефективності впровадження великих даних

Серед інших методів оцінювання можна виділити наступні:

- Метод найменших модулів, що знаходить мінімум суми не квадратів відхилень, а їх абсолютних значень:

$$\beta = \arg \min_{\beta} \sum_{i=1}^n \left| y_i - \beta_0 - \sum_{j=1}^K X_{ij} \beta_j \right|$$

Цей метод є найкращим з точки зору максимальної правдоподібності, коли викиди підпорядковуються розподілу Лапласа. Він менш чутливий до пропусків,

ніж метод найменших квадратів, але може мати більше одного розв'язку і не має простої формули для визначення оцінки.

- Метод максимальної правдоподібності. Використовується, коли відомий розподіл всіх відхилень всіх спостережень. Для класичних та узагальнених лінійних регресійних моделей з умовою нормальності відхилень призводить до тих самих результатів, що й метод найменших квадратів та узагальнений метод найменших квадратів відповідно.

Обґрунтування пошукових питань:

1 Пошукове питання: Якою є щільність розподілу ціни на ноутбуки фірми HP в магазинах конкурентів?

Мета Порівнюючи розподіл цін, порівнюються пропозиції конкурентів. Таким чином, можна визначити кількість ноутбуків, проданих кожним магазином, середню ціну за ноутбук в магазині, найпоширенішу ціну ... У нашому випадку ми оцінюємо результати для всіх товарів у магазині. Іншими словами, якщо ми визначаємо, що середня ціна в магазині $A > B$, то магазин A є дорожчим магазином (такий підхід не зовсім доречний в реальності, але через досить тривалий процес збору даних такий розрахунок працює

Інформація у відповідь Кожна компанія має власний розподіл цін на ноутбуки, який наочно показано в презентації та пояснено особливості розподілу.

Джерела інформації: сайти найбільших компаній які займаються продажом побутової техніки та електроніки в Україні. Перелічимо їх:

- 1) Comfy. Офіційний сайт компанії: <https://Comfy.ua/>
- 2) Comfy. Офіційний сайт компанії: <https://Comfy.ua/>
- 3) Allo. Офіційний сайт компанії: <https://allo.ua/>
- 4) Foxtrot. Офіційний сайт компанії: <https://www.foxtrot.com.ua/>
- 5) Rozetka. Офіційний сайт компанії: <https://rozetka.com.ua/>
- 6) Moyo. Офіційний сайт компанії: <https://www.moyo.ua/>
- 7) Citrus. Офіційний сайт компанії: <https://www.citrus.ua/>

Методологія отримання інформації: у буквальному сенсі - вилучення даних з офіційного сайту компанії. У сфері науки про дані цю практику часто називають веб-скрепінгом. (Web Scraping)

Ця процедура вимагає знання HTML та Python. Згідно з джерелом [12], HTML - це мова гіпертекстової розмітки, яка слугує будівельним матеріалом для створення сторінки. Вона використовується для створення основного контенту, включаючи текст, зображення та різні блоки. Все це заповнюється за допомогою спеціальних командних тегів браузера, які користувач вводить у спеціальний файл з розширенням .html. Python - це високорівнева об'єктно-орієнтована мова програмування, що інтерпретується зі строгою динамічною типізацією [13], яка є простою у використанні і в той же час повноцінною, зі строгою динамічною типізацією [13]. мовою програмування. Мова може бути використана для універсальних та портативних додатків і може використовуватися на будь-якій операційній системі.

1 Етап: зробити запит на сервер

2 Етап: розпарсити html

3 Етап: відповідно до тегів витягти необхідні дані

4 Етап: подати у вигляді dataframe (таблиця з даними)

2 Пошукове питання: Яка існує тенденція офлайн трафіку в роздрібні магазини Comfy, чи існує сезонність трафіку?

Мета: Щоб збільшити трафік, необхідно розуміти його структуру. Розуміння сезонності може допомогти визначити інструменти та підходи для збільшення трафіку.

Інформація у відповідь: графік тренду трафіку та графік сезонності.

Джерела інформації: внутрішня інформація компанії

Методологія отримання інформації: в великих компаніях існують відділи які займаються збором даних, для отримання необхідної інформації мені потрібно лише звернутися до них.

3 Пошукове питання: Чи залежить кількість трафіку від регіону?

Мета: зрозуміти чи впливає регіон на відвідування магазинів

Інформація у відповідь: Відповідь так/ні і з якою ймовірністю

Джерела інформації: зібрані на минулих етапах дослідження

4 Пошукове питання: Чи існує лінійна залежність між трафіком та факторами?

Мета: зрозуміти чи можемо ми в подальшому використовувати лінійну регресію

Інформація у відповідь: Відсоток кореляції Пірсона

Джерела інформації: зібрані на минулих етапах дослідження

5, 6, 7 Пошукове питання: Який існує вплив не медійних факторів на трафік? Який існує вплив медійних факторів? Як вплинув карантин на офлайн трафік?

Мета: розуміти внесок певних факторів на офлайн трафік

Інформація у відповідь: Список факторів та їх внесок у модель

Джерела інформації: зібрані на минулих етапах дослідження

9 Пошукове питання: Який можна зробити прогноз офлайн трафіку на 2022 рік?

Мета: зробити прогноз

Інформація у відповідь: графік з прогнозом

Джерела інформації: зібрані на минулих етапах дослідження

РОЗДІЛ 4. ІНТЕГРАЦІЯ СИСТЕМНОГО АНАЛІЗУ В УПРАВЛІННЯ ВЕЛИКИМИ ДАНИМИ: ПІДХОДИ, ВИКЛИКИ ТА ПЕРСПЕКТИВИ

4.1 Використання великих даних у фінансовому секторі

Перейдемо до виконання першого завдання цього дослідження. Джерела інформації:

- 1) Comfy. Офіційний сайт компанії: <https://Comfy.ua/>
- 2) Comfy. Офіційний сайт компанії: <https://Comfy.ua/>
- 3) Allo. Офіційний сайт компанії: <https://allo.ua/>
- 4) Foxtrot. Офіційний сайт компанії: <https://www.foxtrot.com.ua/>
- 5) Rozetka. Офіційний сайт компанії: <https://rozetka.com.ua/>
- 6) Moyo. Офіційний сайт компанії: <https://www.moyo.ua/>
- 7) Citrus. Офіційний сайт компанії: <https://www.citrus.ua/>

Необхідно зібрати назви моделей ноутбуків HP у кожному магазині та відповідні ціни на них. Для цього використовується мова програмування Python.

В першу чергу імпортуємо необхідні бібліотеки (див. рис. 4.1),:

```
# Загрузка необхідних бібліотек
from bs4 import BeautifulSoup
import requests
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

Рис 4.1 – Імпорт необхідних бібліотек

Далі наведемо приклад коду на рис 4.2, за допомогою якого будемо витягувати дані:


```

# Ініціалізуємо списки в які будемо складати назви та ціни ноутбуків
finaltital = []
finalprice = []
headers = {'User-Agent': 'Mozilla/xxx (xxxxxxxxxxxxxxxxxxx) xxxxx/xxxxxxxxxx Firefox/23.0'}
# Основний код
# Робимо запит на сервер
for page in range(1,12):
    if page == 1:
        web_foxtrot =
        requests.get(f"https://www.foxtrot.com.ua/ru/shop/noutbuki_hp.html",
        headers=headers)
    else:
        web_foxtrot =
        requests.get(f"https://www.foxtrot.com.ua/ru/shop/noutbuki_hp.html?page={page}",
        headers=headers)
    if web_foxtrot.status_code == 200:
        print(f"Page {page} is Success")
    else:
        print("Something wrong")
    # Парсимо html
    sp = BeautifulSoup(web_foxtrot.content, "xxxxxxx")
    # Витягуємо необхідні теги
    title = sp.find_all("a", "card_title")
    price = sp.find_all("div", "card-price")
    titleloop = [titles.text for titles in title]
    priceloop = [prices.text for prices in price]
    # Чистимо дані
    for i in priceloop:
        x = int(i.replace(" ", "").replace("₴", "").replace("\n", ""))
        finalprice.append(x)
    for i in titleloop:
        finaltital.append(i)
    # Перевіряємо отримані дані
    if len(finalprice) == len(finaltital):
        print(f"Ready for df, the length of data is: {len(finalprice)}")
    else:
        print("Check length of data")

```

Рис. 4.2 Код запиту

Отримаємо наступну відповідь (див. рис. 4.3), яка свідчить, що всі сайти успішно загрузилися, та кількість елементів складає 287.

```

Page 1 is Success
Page 2 is Success
Page 3 is Success
Page 4 is Success
Page 5 is Success
Page 6 is Success
Page 7 is Success
Page 8 is Success
Page 9 is Success
Page 10 is Success
Page 11 is Success
Ready for df, the length of data is: 287

```

Рис. 4.3 – Результат парсингу сайтів

Після того, як проведемо подібну процедуру над всіма сайтами, отримаємо загальну таблицю (див. рис. 4.4) яка складає 2121 елемент та можемо приступати до аналізу.

	Title	Price	Brand
0	Ноутбук HP Envy x360 Convertible 15-es0005ua ...	39999	Eldorado
1	Ноутбук HP 250 G8 (2W9A9EA) Dark Ash Silver	19465	Eldorado
2	Ноутбук HP 240 G8 (2X7R5EA) Dark Ash Silver	31699	Eldorado
3	Ноутбук HP Laptop 15s-fq2014ua (437M3EA) Silver	21999	Eldorado
4	Ноутбук HP EliteBook 840 G8 (336D4EA) Silver	49999	Eldorado
...	...	...	...
2116	Ноутбук HP Pavilion 14-dv0026ur Gold (398N1EA)	22499	Foxtrot
2117	Ноутбук HP Spectre x360 14-ea0000ur Natura Sil...	35999	Foxtrot
2118	Ноутбук HP Spectre x360 13-aw2012ur Poseidon B...	56999	Foxtrot
2119	Ноутбук HP ZBook Create G7 Turbo Silver (2W982...	77555	Foxtrot
2120	Ноутбук HP ENVY 15-ep0029ur Silver (219Y2EA)	39999	Foxtrot
2121 rows × 3 columns			

Рис 4.4 – Набір даних, який отримали в результаті парсингу

Почнемо з того, що дізнаємося який обсяг ноутбуків пропонує кожен з магазинів (див. рис. 4.5, 4.6).

```

y = df["Brand"].value_counts()
plt.figure(figsize=(8,5))
sns.barplot(x = y.index, y = y.values, palette="Blues_d")
plt.title('Кількість ноутбуків в магазинах')
plt.ylabel('Кількість', fontsize=12)
plt.xlabel('Бренд', fontsize=12)

```

Рис. 4.5 – Розрахунок обсягу ноутбуків НР

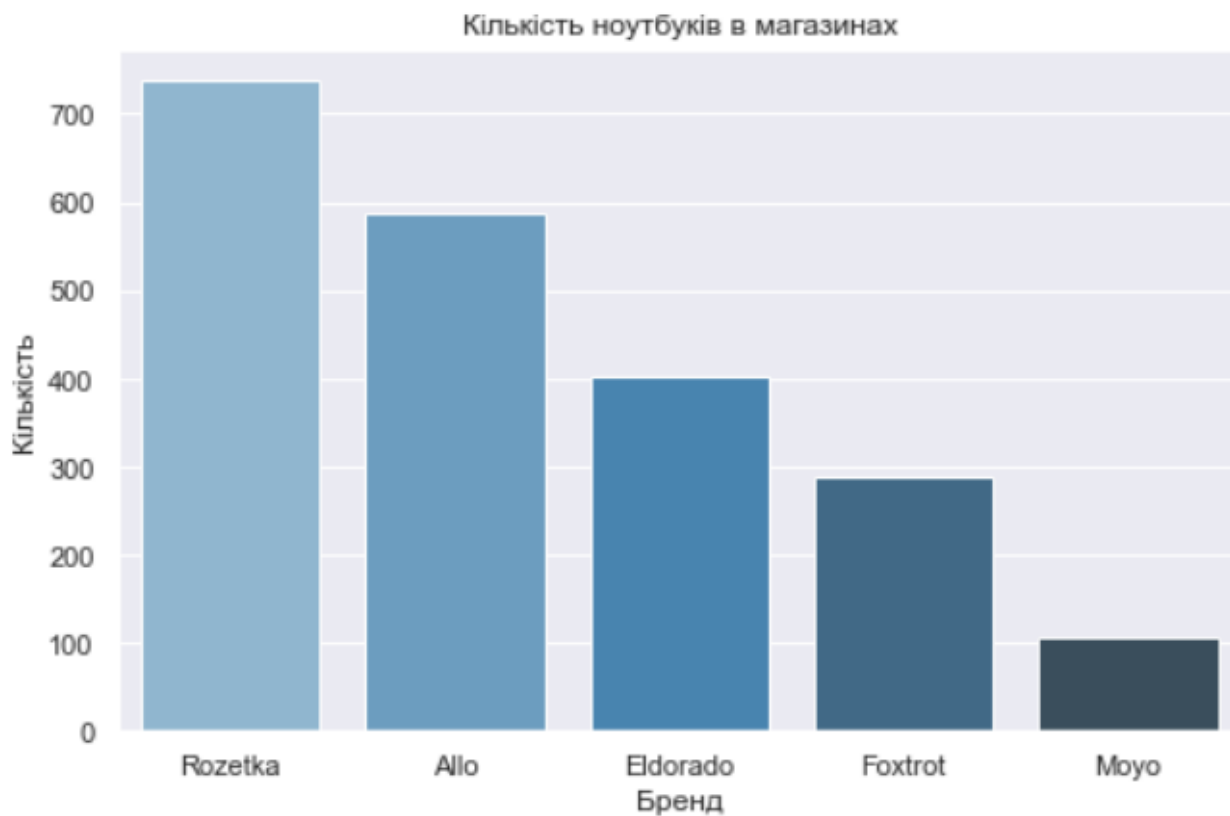


Рис. 4.6 – Обсяг ноутбуків НР

Це означає, що найбільше ноутбуків продається в магазинах Rozetka. Не забудьте екстраполювати цей результат на всі товари в кожному магазині. Таким чином, ми дійшли висновку, що Rozetka має найбільший асортимент товарів, тоді як MoYo має найменший асортимент.

Наступним етапом подивимося на максимальну та мінімальну ціну ноутбуку, та магазини які їх пропонують

Отримаємо наступну відповідь (див. рис. 4.7)

	Title	Price	Brand
1332	Ноутбук HP 250 G8 (2W8V6EA)	111299	Allo
1497	HP Probook 6570b (H5E72EA) "Refurbished"	5320	Allo
2099	Ноутбук HP ZBook Create G7 Turbo Silver (2H6U5...	111299	Foxtrot

Рис 4.7 – Мінімальна та максимальна ціна ноутбука

Allo - найдешевший і найдорожчий ноутбук на ринку. Це твердження можна проілюструвати за допомогою графіка, який називається «boxplot» (див. рис. 4.8).

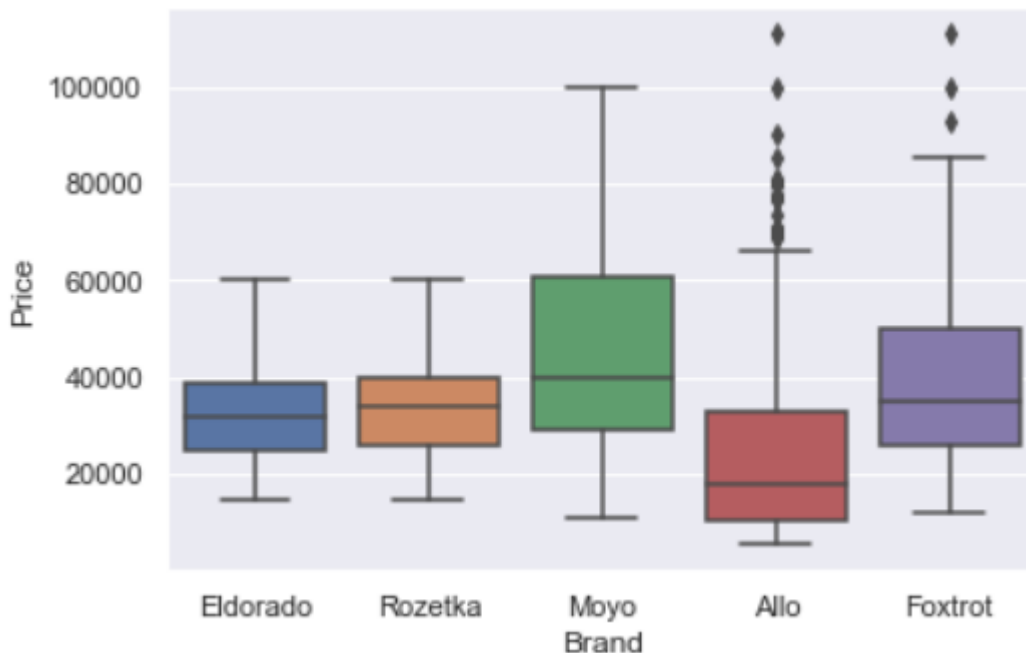


Рис. 4.8. Boxplot

На графіку також видно лінію з медіаною. Тепер розрахуємо центральну тенденцію (див. рис 4.9.).

```
roz = df[df["Brand"] == "Rozetka"]
el = df[df["Brand"] == "Comfy"]
allo = df[df["Brand"] == "Allo"]
fox = df[df["Brand"] == "Foxtrot"]
moyo = df[df["Brand"] == "Moyo"]

def MCT(x):
    mean = np.mean(x)
    mediana = np.median(x)
    moda = mode(x)
    return mean, mediana, moda

print("Rozetka:", MCT(roz["Price"]))
print("Comfy:", MCT(el["Price"]))
print("Allo:", MCT(allo["Price"]))
print("Foxtrot:", MCT(fox["Price"]))
print("Moyo:", MCT(moyo["Price"]))
```

Рис. 4.9. - Розрахунок показників мір ц.т

Отриманий результат буде виглядати наступним чином (див. рис. 4.10)

```
Rozetka: (33876.6078697422, 33999.0, 26999)
Eldorado: (32765.58064516129, 31699.0, 26999)
Allo: (23745.372448979593, 18055.0, 34999)
Foxtrot: (39382.24041811847, 34999.0, 34999)
Moyo: (45490.83018867925, 39999.0, 29999)
```

Рис. 4.10 – Результати розрахунку показників мір ц.т

Зобразимо отриманий результат графічно (див. рис. 4.11, 4.12)

```
shops = ["Rozetka", "Comfy", "Allo", "Foxtrot", "Moyo"]
mean = [MCT(roz["Price"])[0], MCT(el["Price"])[0], MCT(allo["Price"])[0],
MCT(fox["Price"])[0], MCT(moyo["Price"])[0]]
mediana = [MCT(roz["Price"])[1], MCT(el["Price"])[1], MCT(allo["Price"])[1],
MCT(fox["Price"])[1], MCT(moyo["Price"])[1]]
moda = [MCT(roz["Price"])[2], MCT(el["Price"])[2], MCT(allo["Price"])[2],
MCT(fox["Price"])[2], MCT(moyo["Price"])[2]]

df2 = pd.DataFrame({"Brand": shops, "Mean":mean, "Mediana":mediana, "Mode":moda})
df3 = df2.melt(id_vars=["Brand"], value_vars=df2.columns[1:])
```

Рис. 4.11 – Розрахунок міри центральної тенденції

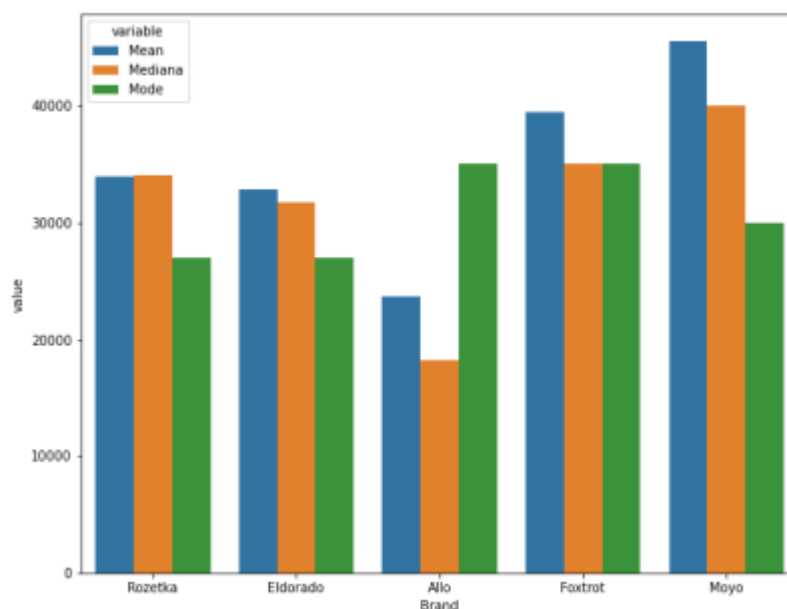


Рис. 4.12 – Міри центральної тенденції

Ми отримали дуже інформативний графік. Медіана і середнє значення показують, що ціни в магазині Мою вищі, ніж в інших магазинах. Найдешевшим є магазин Allo.

Після обчислення міри центральної тенденції настав час побудувати гістограму відносних кумулятивних частот.

Процедуру побудови емпіричного розподілу варто розглянути більш детально. Для цього використовується Excel.

- 1) Будуємо варіаційний ряд (див. рис. 4.13);

A	B	C	D	E	F	G	H
	Дані, ціна на ноутбук HP в Rozetka		Варіаційний ряд		Розмах вибірки	45 555	
	24 999		14 444		Оцінки повної вибірки		
	25 999		14 444		Вибіркове середнє	33 877	
	14 444		14 444		Вибіркова дисперсія	105 261 625	
	35 999		14 444		Мода	26 999	
	39 999		14 444		Медіана	33 999	
	37 999		14 444		25 процентіль	25 999	
	26 999		14 444		75 процентіль	39 999	
	26 999		14 444		Оцінки групованої вибірки		
	29 999		14 444		Мат.сподівання	24 865	
	20 599		14 444		Вибіркова дисперсія	135 557 154	
	35 999		14 444		Мода	37 114	
	49 999		16 399		Медіана	38 414	
	33 999		16 399				
	27 999		16 399				

Рис. 4.13 – Варіаційний ряд

- 2) Рахуємо точкові оцінки повної вибірки;
- 3) Будуємо груповану вибірку (див. рис. 4.14);

I	J	K	L	M	N	O	P	Q	R	S	T
№ інтервалу	Ліва межа інтервалу	Права межа інтервалу	Середина інтервалу	Частота інтервалу	Відносна частота інтервалу	Відносна накопичена частота інтервалу		Формула Стерджесса			
1	14444	18999,50	16721,75	64	8,7%	8,7%		n	10,53	10	
2	18999,50	23555,00	21277,25	52	7,1%	15,7%		h	4555,50		
3	23555,00	28110,50	25832,75	144	19,5%	35,3%					
4	28110,50	32666,00	30388,25	50	6,8%	42,1%					
5	32666,00	37221,50	34943,75	151	20,5%	62,6%					
6	37221,50	41777,00	39499,25	149	20,2%	82,8%					
7	41777,00	46332,50	44054,75	62	8,4%	91,2%					
8	46332,50	50888,00	48610,25	14	1,9%	93,1%					
9	50888,00	55443,50	53165,75	15	2,0%	95,1%					
10	55443,50	59999,00	57721,25	36	4,9%	100,0%					

Рис. 4.14 – Групована вибірка

- 3) Рахуємо оцінки групованої вибірки;
- 4) Будуємо гістограму частот (див. рис. 4.15);

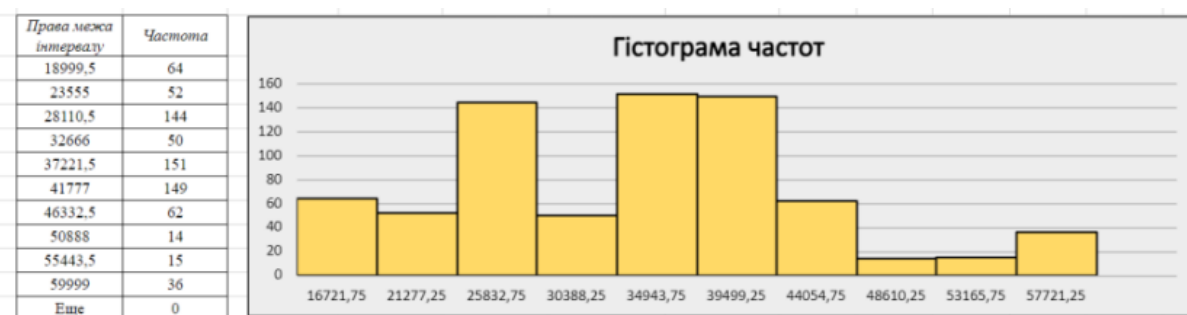


Рис. 4.15. – Гістограма частот

- 5) Будуємо кумулятивну криву (див. рис. 4.16);

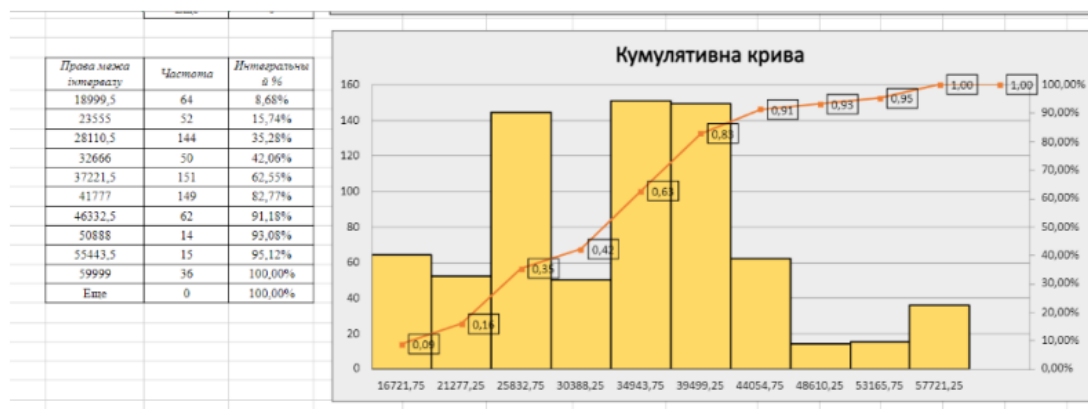


Рис. 4.16 – Кумулятивна крива

- 6) Робимо висновок у форматі «Питання-Відповідь»

1) Чому відрізняються оцінки по повній вибірці та по групованій?

При обчисленні оцінок для згрупованих вибірок невідомо, де більшість елементів потрапляє в інтервал. Однак істинне значення оцінки в генеральній сукупності може бути близьким як до повної вибірки, так і до згрупованої вибірки.

2) *Які висновки можна зробити, маючи результати про особливості розподілу генеральної сукупності?*

Розподіл має від'ємний ексцес (-0,04) та додатну асиметрію (0,44) через нахил вліво. Завдяки додатній асиметрії мода розподілу є меншою за середнє значення.

3) *Якою є гіпотеза про розподіл генеральної сукупності?*

Вибірка ще не має достатньої кількості елементів, але цілком ймовірно, що розподіл буде відповідати нормальному розподілу по мірі того, як з'являтиметься більше даних.

Ми можемо перевірити розподіл на нормальність за допомогою критерію згоди Шапіро-Уилка (див. рис. 4.17, 4.18)

```
from scipy import stats
statistics, p_value = stats.normaltest(x)
print(statistics, p_value)

if p_value > 0.05:
    print("Распределение нормальное !!!")
else:
    print("Распределение не нормальное")

sns.distplot(np.log(el["Price"]))
```

Рис. 4.17 – Перевірка на нормальність (1)

Це призвело до р-значення 0,17. Оскільки це значення більше за 0,05, основну гіпотезу про нормальний розподіл не можна відкинути.

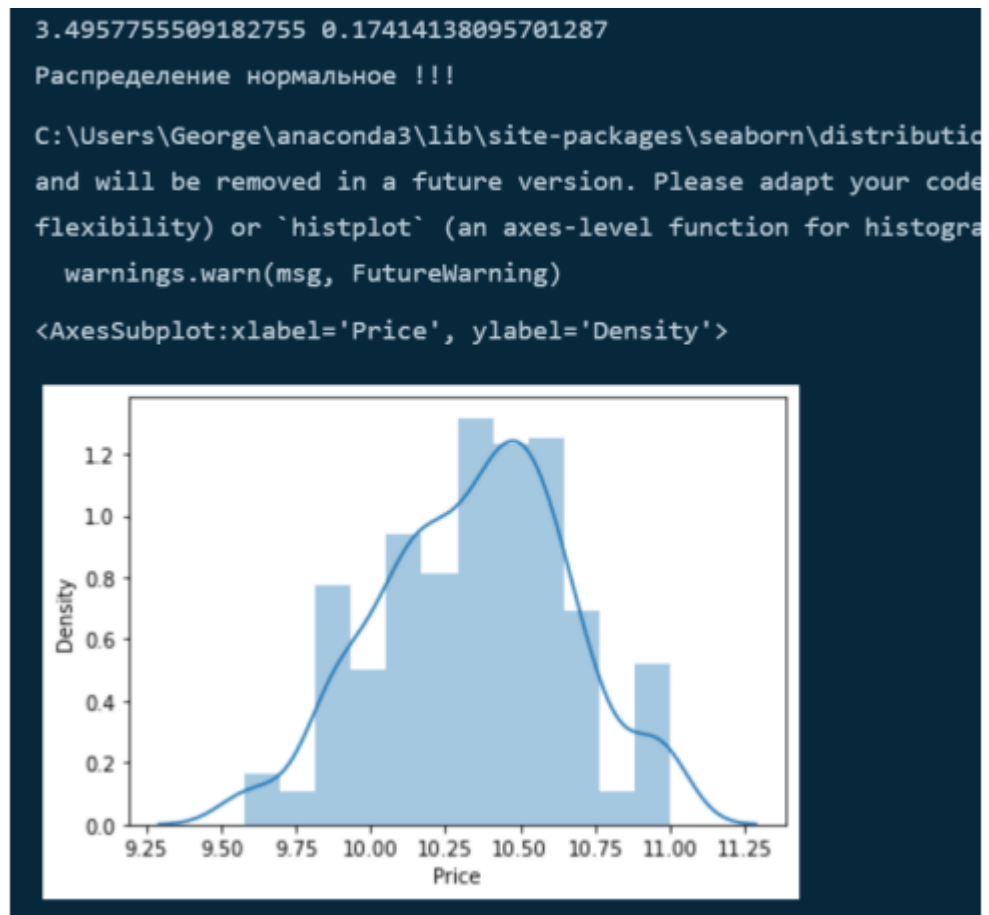


Рис. 4.18 – Перевірка на нормальність (2)

Нарешті, порівняйте вибіркові характеристики кожного магазину (рис. 4.19)

```

1  def comp_param(x):
2  mean = np.mean(x)
3  mediana = np.median(x)
4  moda = mode(x)
5  std = np.std(x)
6  48
7  var = std ** 2
8  min1 = min(x)
9  max1 = max(x)
10 range1 = max1 - min1
11 skewness1 = skew(x)
12 kurtosis1 = kurtosis(x)
13 return mean, mediana, moda, std, var, min1, max1, range1, skewness1, kurtosis1
14 shops = ["Rozetka", "Eldorado", "Allo", "Foxtrot", "Moyo"]
15 mean = [comp_param(roz["Price"])[0], comp_param(el["Price"])[0],
16 comp_param(allo["Price"])[0], comp_param(fox["Price"])[0],
17 comp_param(moyo["Price"])[0]]
18 mediana = [comp_param(roz["Price"])[1], comp_param(el["Price"])[1],
19 comp_param(allo["Price"])[1], comp_param(fox["Price"])[1],
20 comp_param(moyo["Price"])[1]]
21 moda = [comp_param(roz["Price"])[2], comp_param(el["Price"])[2],
22 comp_param(allo["Price"])[2], comp_param(fox["Price"])[2],
23 comp_param(moyo["Price"])[2]]
24 std = [comp_param(roz["Price"])[3], comp_param(el["Price"])[3],
25 comp_param(allo["Price"])[3], comp_param(fox["Price"])[3],
26 comp_param(moyo["Price"])[3]]
27 var = [comp_param(roz["Price"])[4], comp_param(el["Price"])[4],
28 comp_param(allo["Price"])[4], comp_param(fox["Price"])[4],
29 comp_param(moyo["Price"])[4]]
30 min1 = [comp_param(roz["Price"])[5], comp_param(el["Price"])[5],
31 comp_param(allo["Price"])[5], comp_param(fox["Price"])[5],
32 comp_param(moyo["Price"])[5]]
33 max1 = [comp_param(roz["Price"])[6], comp_param(el["Price"])[6],
34 comp_param(allo["Price"])[6], comp_param(fox["Price"])[6],
35 comp_param(moyo["Price"])[6]]
36 range1 = [comp_param(roz["Price"])[7], comp_param(el["Price"])[7],
37 comp_param(allo["Price"])[7], comp_param(fox["Price"])[7],
38 comp_param(moyo["Price"])[7]]
39 skewn = [comp_param(roz["Price"])[8], comp_param(el["Price"])[8],
40 comp_param(allo["Price"])[8], comp_param(fox["Price"])[8],
41 comp_param(moyo["Price"])[8]]
42 kurtos = [comp_param(roz["Price"])[9], comp_param(el["Price"])[9],
43 comp_param(allo["Price"])[9], comp_param(fox["Price"])[8],
44 comp_param(moyo["Price"])[9]]
45 49
46 df2 = pd.DataFrame({"Brand": shops, "Mean":mean, "Mediana":mediana, "Mode":moda,
47 "std": std, "var": var, "min": min1, "max": max1, "range":range1, "skew":skewn,
48 "kurt":kurtos})
49 df3 = df2.melt(id_vars=["Brand"], value_vars=df2.columns[1:])
50 df3.to_excel("C:/Users/xxxxx/xxxx/PF СТАТ/xxx.xlsx")
51

```

Рис. 4.19. - Код запиту

Отримаємо наступну таблицю 4.1

Таблиця 4.1 - Вибіркові характеристики

	Allo	Eldorado	Foxtrot	Moyo	Rpzetka
Ексцес	2,6	0,2	1,1	-0,5	-0,1
Максимальне значення	111 299	59 999	111 299	99 999	59 999
Середнє значення	23 745	32 766	39 382	45 491	33 877
Медіана	18 055	31 699	34 999	39 999	33 999
Мінімальне значення	5 320	14 444	11 999	10 999	14 444
Мода	34 999	26 999	34 999	29 999	26 999
Розмах	105 979	45 555	99 300	89 000	45 555
Асиметрія	1,6	0,6	1,1	0,7	0,4
Стандартне відхилення	17 716	10 283	17 918	21 776	10 253
Дисперсія	313 858 226	105 740 139	321 068 737	474 183 975	105 118 800

Нехай Allo – A, Eldorado – E, Foxtrot – F, Moyo – M, Rozetka – R, тоді з таблиці 4.1 маємо наступні нерівності:

Ексцес: $A > F > E > R > M$; Максимальне значення: $A = F > M > E = R$;
Середнє значення: $M > F > R > E > A$; Медіана: $M > F > R > E > A$; Мінімальне значення: $E = R > F > M > A$; Мода: $A = F > E = M = R$; Розмах: $A > F > M > R = E$;
Асиметрія: $A > F > M > E > R$; Стандартне відхилення: $M > F > A > E > R$;
Дисперсія: $M > F > A > E > R$.

Тепер перейдемо до 2 завдання дослідження - огляду результатів діяльності Comfy. Значущість: Використані дані є штучними і слугують лише прикладами статистичних методів та способів обробки даних. Почнемо з огляду офлайн-трафіку магазинів Comfy, щоб виявити поточні тенденції та сезонність. Перш за все, необхідно зібрати дані. Припустимо, що в компанії є відділ, який займається збором даних і має доступ до бази даних SQL, що містить дані про трафік. Наступний код (див. рис. 4.20) використовується для вилучення цього набору даних на Python.

```
import pandas as pd
from sqlalchemy import create_engine
import pymysql

db_connection_str = 'mysql+pymysql://xxxxxxxxxx@xxxxxxxx/xxxxxxxx'
db_connection = create_engine(db_connection_str)
df = pd.read_sql('SELECT * FROM trafficCity', con=db_connection)
```

Рис. 4.20 – Дані з трафіком(1)

Отримаємо наступний набір даних (див. рис. 4.21)

	Year	Week	City	Traffic
0	2018	1	Александрия	1937
1	2018	1	Белая Церковь	2954
2	2018	1	Белогород Днестровский	1162
3	2018	1	Бердичев	1679
4	2018	1	Бердянск	5083
...	...	...	...	...
12068	2021	53	Херсон	3660
12069	2021	53	Хмельницкий	4442
12070	2021	53	Черкасы	2186
12071	2021	53	Чернигов	5265
12072	2021	53	Черновцы	2245

12073 rows x 4 columns

Рис. 4.21 – Дані з трафіком (2)

В нашому наборі даних є місто та офлайн трафік протягом 2018-2021 року. Побудуємо графік загальної динаміки в місті Києві (див. рис. 4.22).

Динаміка офлайн трафіку



Рис. 4.22 – Динаміка трафіку в місті Києві

Графік показує чітку сезонність з циклічними піками навколо 46-ї та 51-52-ї неділі кожного року. Можна припустити, що саме на ці неділі припадають Чорна п'ятниця та Новий рік. Тому ми припускаємо, що ці свята впливають на динаміку трафіку і спробуємо визначити, який саме приріст трафіку викликаний наявністю цих свят.

Далі, проведемо декомпозицію часового ряду (в нашому випадку це трафік). Напишемо код, для декомпозиції ряду і отримаємо окремий вклад тренду сезонності та випадкової помилки (див. рис. 4.23).

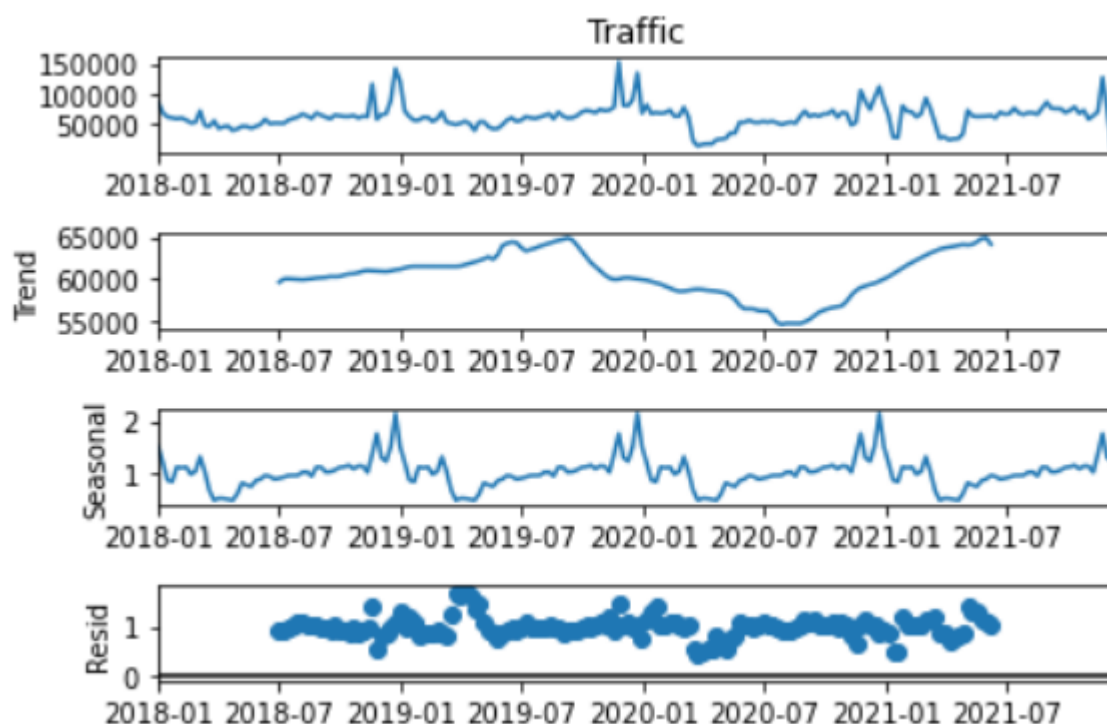


Рис. 4.23 – Декомпозиція часового ряду

При побудові моделі необхідна чітка сезонність, і ми також бачимо, що з 2018 по 2020 рік спостерігається позитивний тренд; з початку 2020 року по серпень ми бачимо спадний тренд через пандемію, а оскільки ми розглядаємо офлайн-трафік, то це мало б сенс лише в тому випадку, якби магазини були зачинені через карантинні обмеження.

Потім було перевірено гіпотезу про те, що кількість трафіку залежить від регіону. Таким чином, основна гіпотеза полягає в тому, що трафік не залежить від регіону, тоді як нульова гіпотеза полягає в тому, що трафік змінюється від регіону до регіону.

Спочатку зобразимо стовпчикову діаграму, та розподіли по кількості трафіку в найбільших містах (Київ, Дніпро, Харків, Одеса, Львів)(див. рис. 4.24).

```
import matplotlib.pyplot as plt
city = ["Київ", "Одеса", "Харків", "Львів", "Дніпро"]
a = df[df["City"].isin(city)].groupby("City")["Traffic"].sum()
plt.figure(figsize=(10, 8))
sns.barpplot(x = a.index, y = a, order=a.sort_values(ascending = False).index)
```

Рис 4.24 – Кількість трафіку в найбільших містах (1)

Отримаємо наступний графік (див. рис. 4,24), 4.25):

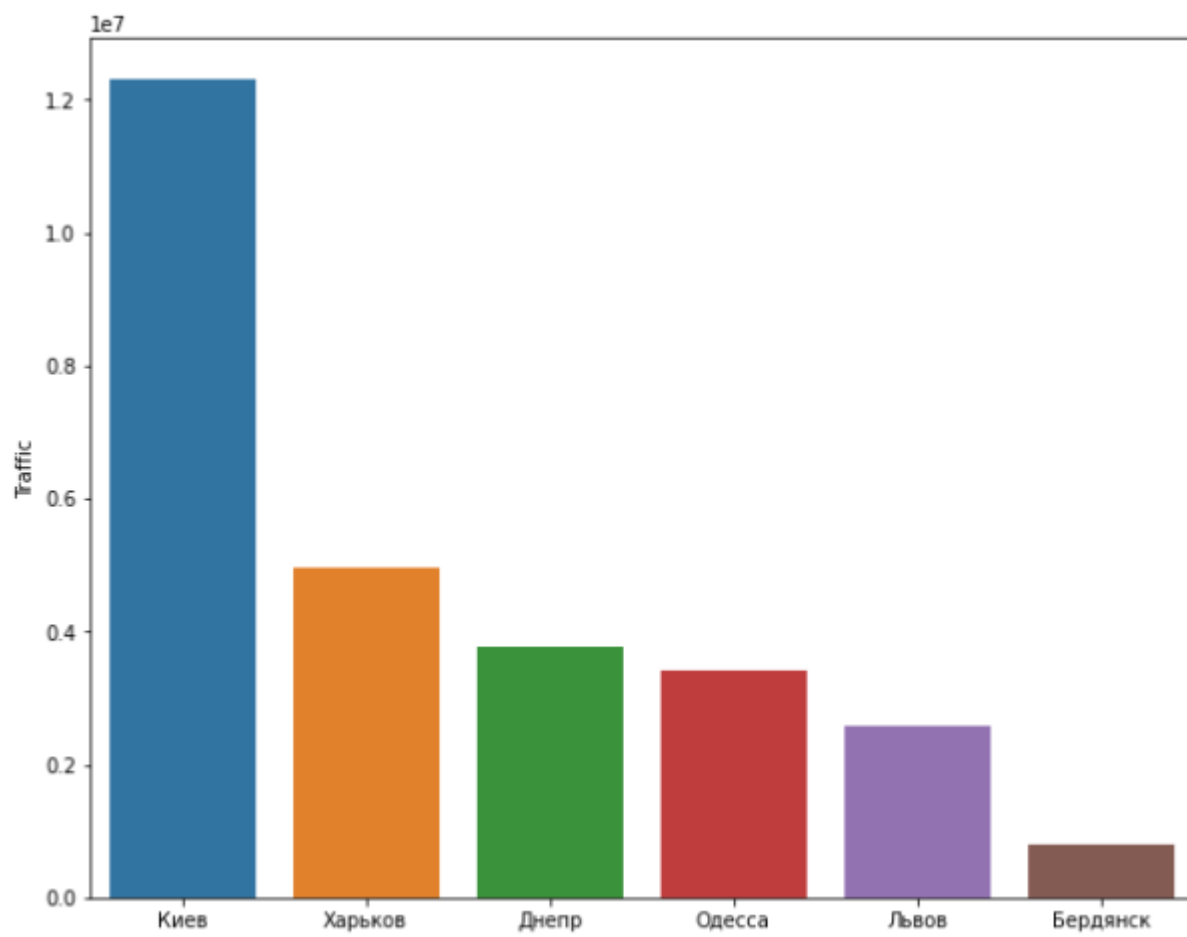


Рис. 4.24 – Кількість трафіку в найбільших містах (2)

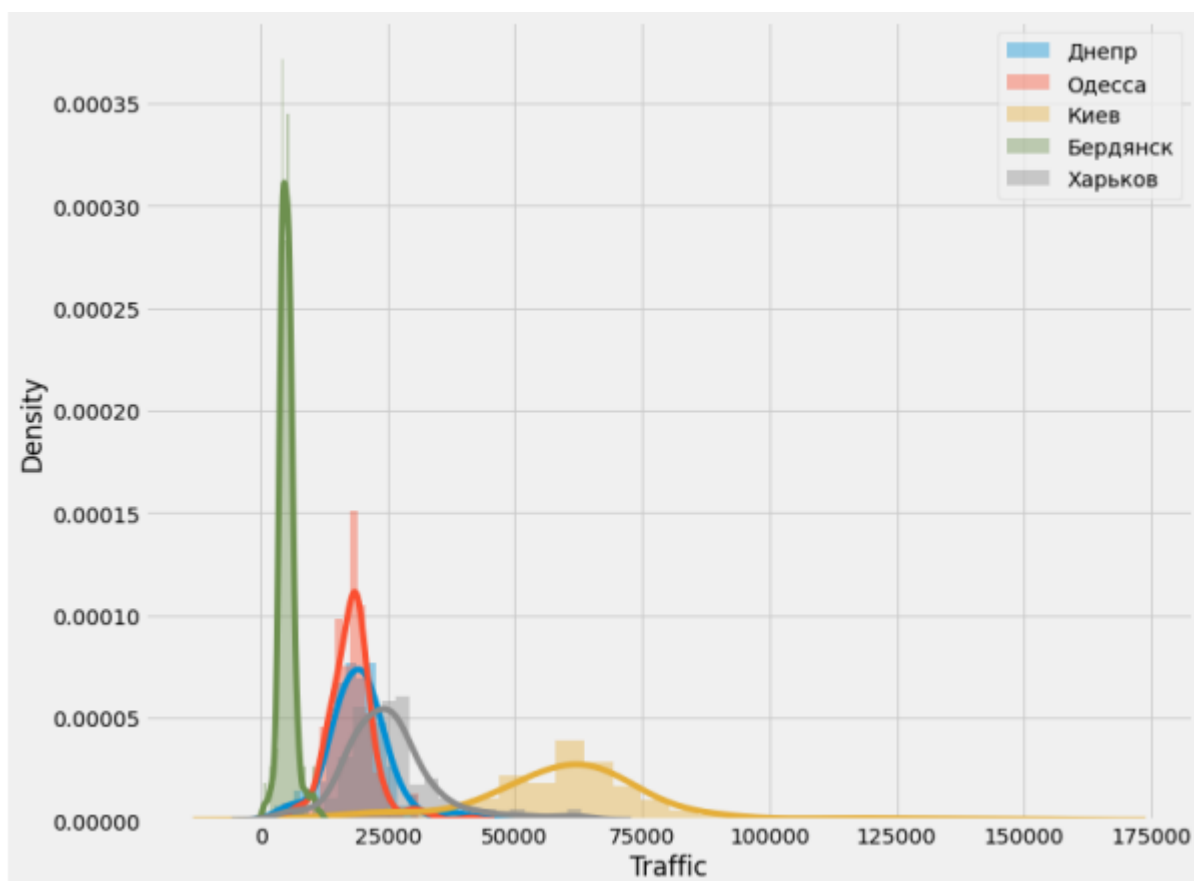


Рис. 4.25 – Кількість трафіку в найбільших містах

Графіки показують, що існує велика різниця в трафіку, але як справжній математик-аналітик, ви не повинні довіряти своїм очам і перевірити це математично, використовуючи однофакторний дисперсійний аналіз.

```
model = ols('Traffic' + '~' + "City", data = df).fit()
result_anova = sm.stats.anova_lm(model)
print(result_anova)
```

	df	sum_sq	mean_sq	F	PR(>F)
City	62.0	9.878437e+11	1.593296e+10	940.26259	0.0
Residual	12010.0	2.035122e+11	1.694523e+07	NaN	NaN

Орієнтуючись на F-статистику та р-значення, ми робимо висновок, що трафік варіюється в залежності від регіону, і приймаємо альтернативну гіпотезу майже зі 100% впевненістю. Перейдемо до третього завдання дослідження - аналізу впливу медіаактивності на офлайн-трафік магазинів Comfy.

Ця проблема вирішується шляхом побудови моделей, які наближають істинні значення. Апроксимація - це метод вибору наближення «складної» функції на інтервалі з сімейства «простих» функцій, що утримує похибку в певних межах. Апроксимація використовується, коли необхідно отримати функцію, подібну до даної, але простішу для обчислення або маніпулювання (похідна, інтеграл тощо) [15].

Прочитаємо вже створений набір даних. Оскільки даних (факторів) для побудови моделі занадто багато, не рекомендується занурюватися в дані і записувати кожен крок формування набору даних.

```
def read_file():
    df = pd.read_excel("C:/Users/George/Desktop/PP стат/Model.xlsx",
                      sheet_name="forModel")
    return df
df = read_file()
df.head()
```

Рис. 4.26 – Набір даних для побудови моделі

Наш набір даних для побудови моделі (див. рис. 4.26) має наступний вигляд

Year	Week	Traffic	Eldorado TV	Competitors TV	Eldorado digital	Competitors digital	Competitor's radio	Eldorado radio	Eldorado OOH	...
2014	1	516830	912.60	6154.23	327544.0	242973643	333	0	111128.000000	...
2014	2	470645	0.00	3144.38	NaN	241158044	93	0	111128.000000	...
2014	3	400100	747.94	2195.65	228663.0	224609929	205	0	111128.000000	...
2014	4	384467	695.55	2897.40	2781741.0	228308187	521	0	111128.000000	...
2014	5	382831	569.30	2628.87	4744420.0	176177652	239	0	109296.857143	...
...	...	...	...	...	...	...	...	...	...	...
2021	39	435642	509.58	3529.18	5087106.0	180481537	169	0	115689.000000	...
2021	40	469145	388.91	3535.31	5525681.0	166608333	75	0	133440.000000	...
2021	41	489419	439.40	3810.58	4648381.0	149422224	99	0	138510.000000	...
2021	42	399089	352.88	3309.22	6741722.0	170424814	0	0	138510.000000	...
2021	43	398311	349.03	2753.16	5822715.0	148889485	0	0	138510.000000	...

Рис 4.26 – Набір даних для побудови моделі

В нас є 408 строки та 24 колони (23 фактори, 1 цільова змінна). Перевіримо набір даних на пропущені елементи

Було написано функцію (див. рис 4.27) для пошуку відсутніх значень і, якщо їх знайдено, повернення користувачеві назви змінної та кількості відсутніх елементів, а також перетворення їх у середнє значення змінної.

```
def missing_value_info(df):
    for i in df.columns:
        if df[i].isnull().sum() > 0:
            print(f"Columns '{i}' has {df[i].isnull().sum()} missing value(s)")
            df[i] = df[i].fillna(np.mean(df[i]))
missing_value_info(df)
```

Рис. 4.27 – функцію, яка шукає пропущені значення

Отримали наступну відповідь, яка значить що було знайдено лише одне пропущене значення в змінній «Comfy digital» і замінено на середнє по змінній.

Перевіримо набір даних на наявність неметричних даних

```
def check_nonNumericData(df):
    for i in df.columns:
        if df[i].dtypes == object:
            print(f"Columns '{i}' has not numerical value")
check_nonNumericData(df)
```

Рис. 4.28 – функцію, яка перевіряє наявність неметричних даних

Написали функцію, яка шукає неметричні дані і повертає користувачеві назву змінної. Виявляється, змінна «Кількість місяців» містить неметричні дані.

Давайте змінимо це за допомогою методу під назвою «Одне гаряче кодування».
(див. рис. 4 29)

```
dummies = pd.get_dummies(df["Номер місяця"])
dummies.head()
```

	Август	Апрель	Декабрь	Июль	Май	Март	Ноябрь	Октябрь	Сентябрь	Февраль	Январь
0	0	0	0	0	0	0	0	0	0	0	1
1	0	0	0	0	0	0	0	0	0	0	1
2	0	0	0	0	0	0	0	0	0	0	1
3	0	0	0	0	0	0	0	0	0	0	1
4	0	0	0	0	0	0	0	0	0	0	1

Рис 4.29 – Кодування неметричних даних

Це означає, що з одного фактора було створено 12 факторів і розмірність набору даних збільшилася.

Для фактору медіа в модель також включено лагові зміни. Це пов'язано з тим, що, як відомо, ефект від реклами проявляється не одразу, а лише з часом.

```
col_list = set(df.columns.tolist()) - set(["Traffic", "Year", "Week", "Номер
месяца"])
X = pd.concat([df[col_list], dummies], 1)
X["lag_TV_el"] = X["Comfy TV"].shift()
X["lag_TV_comp"] = X["Competitors TV"].shift()
X["lag_dig_el"] = X["Comfy digital"].shift()
X["lag_dig_comp"] = X["Competitors digital"].shift()
X["lag_ooh_el"] = X["Competitors TV"].shift()
X["lag_ooh_comp"] = X["Competitor's OOH"].shift()
X = X.fillna(0)
y = df["Traffic"]
```

Рисунок 4.30 – Лагові зміни для медіа факторів

Так як дані мають різну розмірність, їх треба стандартизувати, аби модель могла коректно працювати (див. рис. 4.31)

```
names = X.columns.tolist()
scaler = StandardScaler().fit(X)
X = scaler.transform(X)
X = pd.DataFrame(data=X, columns=names)
```

Рис. 4.31 – Стандартизація моделі

Далі, необхідно перевірити цільову змінну на нормалість розподілу (одна з умов МНК) (див. рис. 4.32).

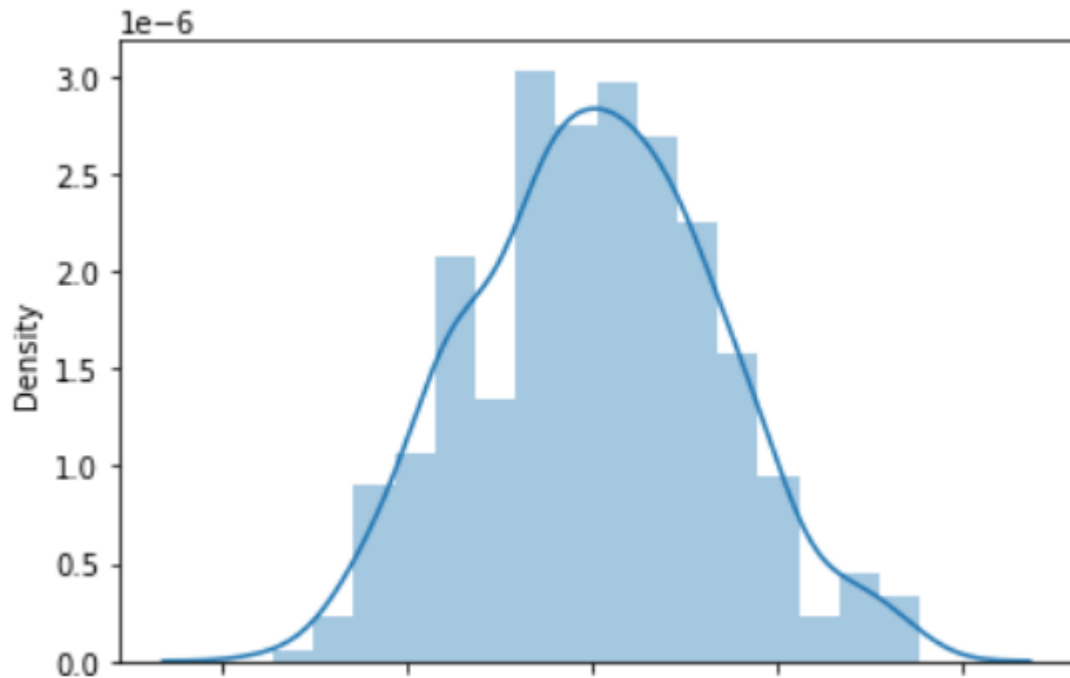


Рис. 4.32 – Розподіл цільової змінної (трафіку)

Графічно ніби як нормальний розподіл, перевіримо за допомогою тесту (див. рис. 4.33)

```
statistics, p_value = stats.normaltest(y_norm)
print(statistics, p_value)
#Відповідь
#0.4476 0.7994
```

Рисунок 4.33 – Розподіл цільової змінної (трафіку) (2)

$P\text{ value} > 0.05$, отже ми не можемо відхилити основну гіпотезу про нормальність розподілу. Наступним кроком побудуємо регресійну модель (див. рис. 4.34).

```
X = sm.add_constant(X)
model = sm.OLS(y, X).fit()
print(model.summary())
```

Рис 4.34 – Побудова регресійної моделі

Отримаємо наступні результати якості моделі (див. рис. 4.35).

OLS Regression Results						
=====						
Dep. Variable:	Traffic	R-squared:	0.876			
Model:	OLS	Adj. R-squared:	0.864			
Method:	Least Squares	F-statistic:	72.82			
Date:	Wed, 15 Dec 2021	Prob (F-statistic):	1.59e-145			
Time:	14:54:35	Log-Likelihood:	-4940.2			
No. Observations:	408	AIC:	9954.			
Df Residuals:	371	BIC:	1.010e+04			
Df Model:	36					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	4.008e+05	2278.508	175.920	0.000	3.96e+05	4.05e+05
Competitor's radio	1.09e+04	3361.471	3.241	0.001	4286.127	1.75e+04
Competitors digital	-4715.1821	4929.827	-0.956	0.339	-1.44e+04	4978.724
Viber	1.871e+04	3526.082	5.305	0.000	1.18e+04	2.56e+04
Competitor's OOH	-7.681e+04	1.28e+04	-6.005	0.000	-1.02e+05	-5.17e+04
К-во городов присутствия	-566.8875	4878.823	-0.116	0.908	-1.02e+04	9026.726
Eldorado TV	2.916e+04	3401.723	8.573	0.000	2.25e+04	3.59e+04
8 марта	6287.5593	2885.380	2.179	0.030	613.810	1.2e+04
1 января	2.571e+04	3248.253	7.916	0.000	1.93e+04	3.21e+04
Рождество	-5195.3216	3149.963	-1.649	0.100	-1.14e+04	998.700
ЧП	1.759e+04	3038.210	5.789	0.000	1.16e+04	2.36e+04
...						

Рис 4.35 – Модель лінійної регресії (методом МНК)

Маємо наступний розподіл залишків (див. рис. 4.36).

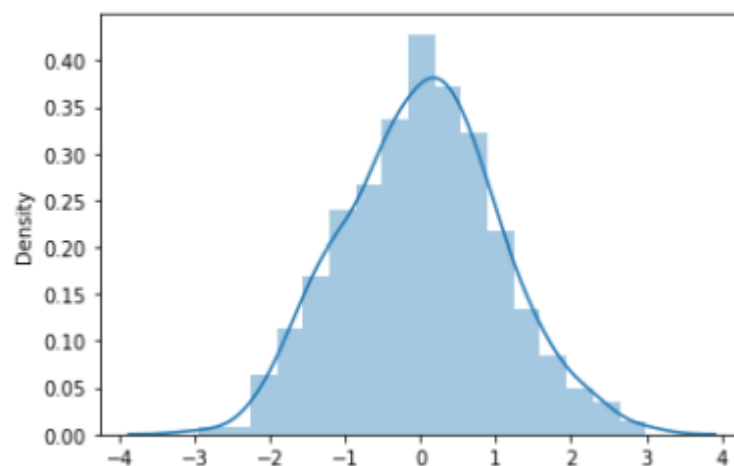


Рис. 4.36 – Розподіл залишків

Умови MLM виконуються, а залишки мають нормальний розподіл. З побудованої моделі можна виділити ознаки (фактори), які дійсно впливають на трафік. Для цього потрібно перевірити значущість коефіцієнтів, і в нашому випадку, якщо $p\text{-value} < 0,05$, то фактор є значущим.

4.2 Вплив Big Data на маркетинг та рекламу

Далі розрахуємо вплив кожного з факторів. Спочатку перерахуємо основні фактори: - органічний приріст трафіку (вказує на кількість трафіку, що не залежить від зовнішніх факторів) - сезонність (вказує на вплив свят) - медійна активність (вказує на кількість трафіку через рекламу) - вплив конкурентів (вказує на те, скільки потенційного трафіку було забрано конкурентами). Згідно з нашою моделлю, внесок факторів виглядає наступним чином (див. рис. 4.37).

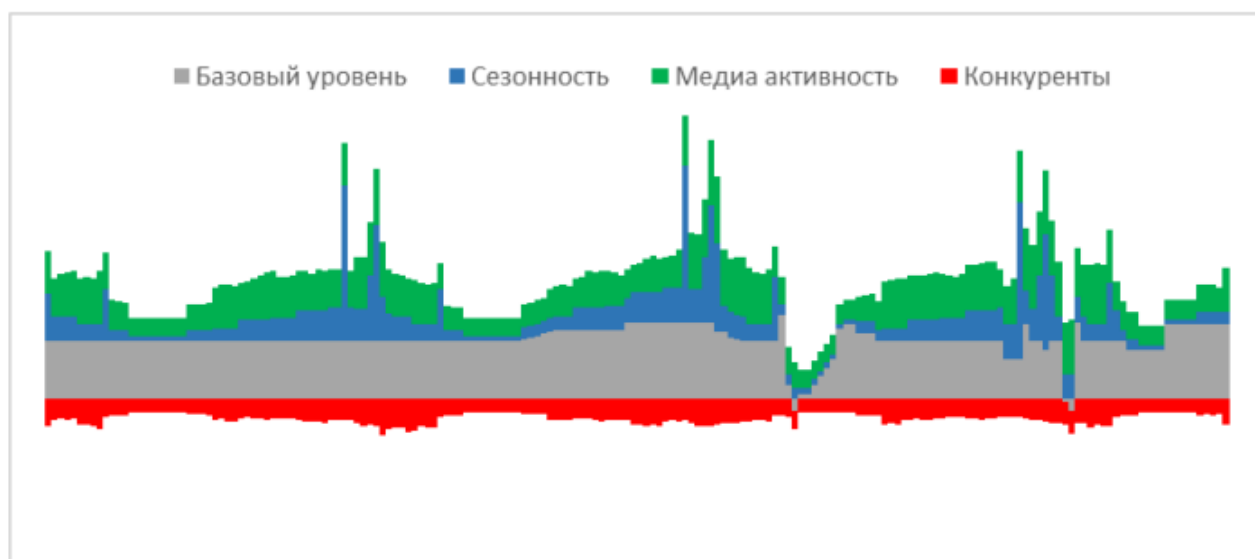


Рис. 4.37 – Вплив факторів

Проаналізуємо вплив кожного фактору більш детально. Базовий рівень та органічне розширення мережі забезпечують 47% трафіку до офлайн-магазинів. Базовий рівень включає:

- розмір роздрібної мережі та її розширення;
- розмір клієнтської бази та її розширення;
- зв'язок з клієнтською базою через розсилки;
- знання ТМ Comfy;
- імідж та сприйняття бренду Comfy;
- попередній досвід взаємодії з Comfy (якість обслуговування, сервісу, продукції);
- інші фактори (якість обслуговування, сервісу, продукції);
- інші фактори, які не піддаються кількісній або цифровій оцінці.

Сезонність включає сезонні коливання та вплив державних свят, на які припадає 29% трафіку. Існують певні закономірності в сезонних коливаннях:

- Серпень – грудень та січень – березень – ключові бізнес-сезони;
- У період із квітня по червень трафік суттєво просідає

За рахунок медіа конкурентів, трафік Comfy втрачає 12%.

Згідно з моделлю, Comfy залучає 36% свого трафіку через медіа-активності.

На основі цієї моделі отримано відповідь на питання дослідження про вплив карантинних обмежень. Вона показує, що карантинні обмеження призвели до значного зниження трафіку після 17 березня, але в травні споживча активність поступово відновилася, частково компенсуючи втрати в березні-квітні. Моделі показують, що під час карантинних обмежень було втрачено приблизно 1 мільйон потенційного трафіку (див. рис. 4.38).

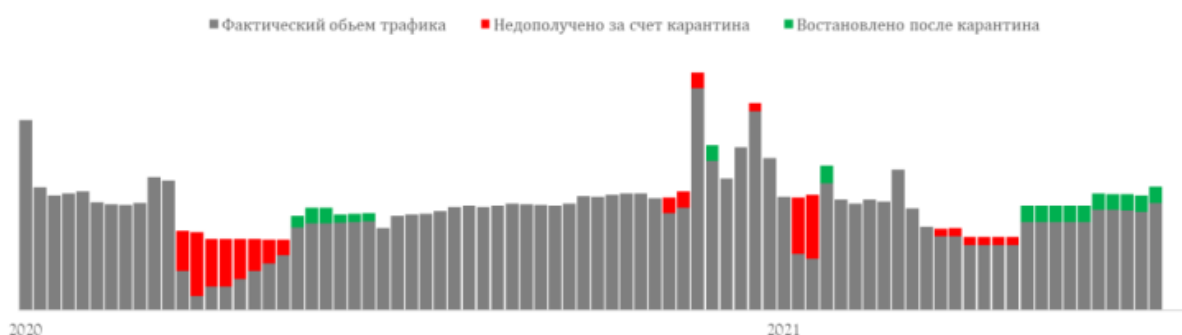


Рис. 4.38 – Вплив карантинних обмежень

Перейдемо до останнього завдання дослідження - прогнозу на 2022 рік. У попередньому розділі для отримання наближених значень змінних, що нас цікавлять, було використано метод лінійної регресії. Як вже зазначалося, цей метод передбачає лінійний зв'язок між факторами та трафіком, що дозволило нам визначити вплив кожного фактору. Ми зробили це припущення (наявність лінійного зв'язку) і отримали досить високу оцінку ($R^2 = 89\%$) якості моделі, але в реальності, в більшості випадків, зв'язок між цільовою змінною і факторами, що її описують, не є строго лінійним і слід використовувати більш складні методи. Тому для цієї регресійної задачі (прогнозування дійсних чисел) ми використовуємо метод XGBoost, який є дуже універсальним і популярним у сфері машинного навчання.

```

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=123)
model = XGBRegressor(n_estimators=1000, max_depth=7, eta=0.1, subsample=0.7,
colsample_bytree=0.8)
model.fit(X_train, y_train)
y_pred = model.predict(X_test)

def mse(y1, y2):
    return np.mean((y1 - y2)**2)

def r2(y1, y2):
    return 1 - (sum((y1 - y2)**2)/sum((y1 - np.mean(y1))**2))

print(f"XGBoost: {mse(y_test, y_pred)}")

reg = linearRegression()
reg = reg.fit(X_train, y_train)
y_pred_lin = reg.predict(X_test)

print(f"Linear Regression: {mse(y_test, y_pred_lin)}")
print(f"Diference is {mse(y_test, y_pred) - mse(y_test, y_pred_lin)}")
print(r2(y_test, y_pred), r2(y_test, y_pred_lin))

#!!!!!!!!!!!!!!!!!!!!Результат!!!!!!!!!!!!!!!!!!!!!!!!!!!!
#XGBoost: 872658646.6944878
#Лінійна регресія: 2440205519.725219
#Різниця: -1567546873.030731
#0.9489575646051334 0.8572706142744758

```

Рисунок 4.39 – XGBoost

Побудовано модель XGBoost та порівняно квадратичні відхилення. Щоб побудувати модель та оцінити її ефективність, ми спочатку розділили дані на навчальний та тестовий набори, показавши, що метод XGBoost робить кращі прогнози, ніж лінійна регресія. Покажемо графік відхилень для тестового набору (див. рис. 4.40)

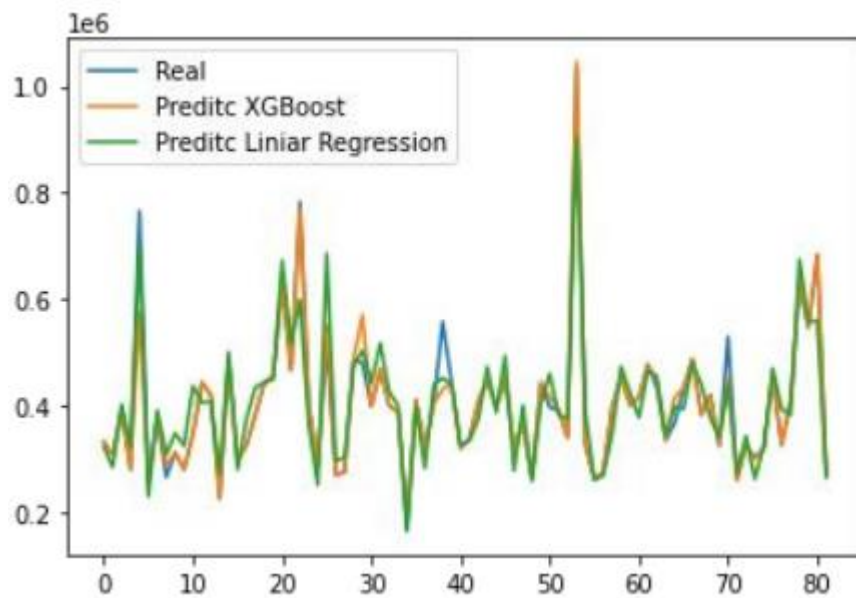


Рис 4.41 - Графік відхилень на тестовій вибірці

Виведемо графік відхилень на тренувальній вибірці (див. рис. 4.42).

```
plt.plot(range(len(y_train)), y_train, label = "Real value")
plt.plot(range(len(y_train)), model.predict(X_train), label = "Predictc XGBoost")
plt.plot(range(len(y_train)), reg.predict(X_train), label = "Predictc Liniar
Regression")
plt.legend()
plt.show()
```

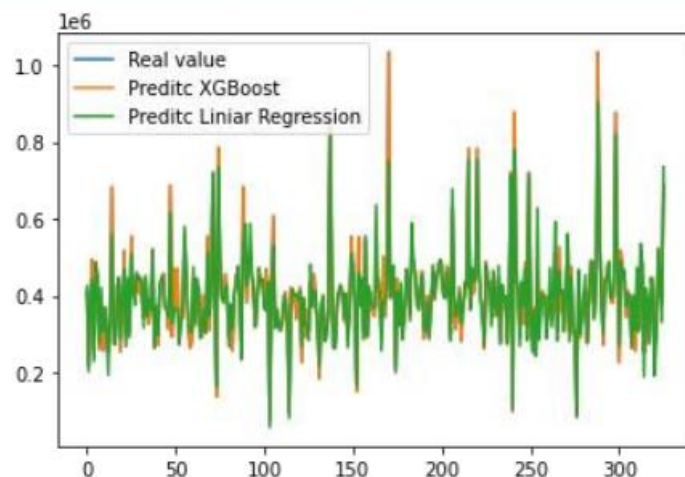


Рис 4.42 - Графік відхилень на тренувальній вибірці

З графіка видно, що метод XGBoost дуже добре описує цільову змінну. Тепер припустимо, що у нас є медіа-план на 2022 рік, і ми використовуємо цю модель для прогнозування трафіку. Ми отримаємо наступні прогнози (див. рис. 4.43).

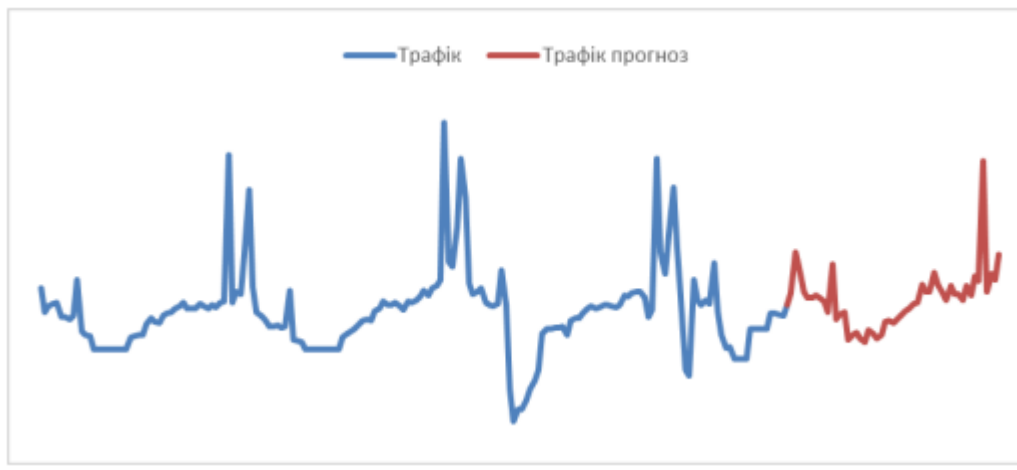


Рис. 4.43 – Прогноз трафіку на 2022 рік

Ми розрахували точкову оцінку прогнозу, але прогноз можна було б покращити, додавши інтервальний діапазон, в якому існує 95% ймовірність того, що трафік зменшиться. Це дало б краще розуміння бізнес-ситуації та допомогло б приймати ефективні рішення, коли це необхідно.

ВИСНОВКИ

Проведене дослідження підтверджує ключову роль технологій обробки великих даних (Big Data) у підвищенні ефективності управління сучасними організаціями. У сучасному середовищі, де обсяги інформації зростають експоненційно, традиційні підходи до аналізу та обробки даних виявляються недостатніми. Інтеграція Big Data з методами системного аналізу відкриває нові можливості для підприємств у всіх сферах їхньої діяльності. Ця синергія дозволяє ефективно вирішувати комплексні завдання, оптимізувати управлінські процеси та приймати обґрунтовані рішення навіть у складних і динамічних умовах.

Системний аналіз забезпечує комплексний підхід до вивчення господарської діяльності підприємств. Його застосування дозволяє охопити всі аспекти функціонування підприємства, виявити слабкі місця, оцінити потенційні резерви та сформулювати практичні рекомендації для досягнення стратегічних і тактичних цілей. Інформаційні моделі, побудовані на основі принципів системного аналізу, забезпечують не лише глибоке розуміння внутрішніх процесів підприємства, але й враховують вплив зовнішнього середовища. Це особливо важливо для підприємств, що прагнуть адаптуватися до змінних умов ринку та підвищувати свою конкурентоспроможність.

Особливе значення має адаптація традиційних методів системного аналізу до специфіки великих даних. В умовах, коли обсяги інформації постійно зростають, а її структура стає дедалі складнішою, необхідно вдосконалювати інструменти аналізу та розробляти нові підходи до обробки даних. Сучасні технології, такі як хмарні обчислення, NoSQL, NewSQL, а також методи моделювання, створюють основу для ефективної роботи з великими обсягами інформації. Ці технології дозволяють не лише зберігати та аналізувати дані, але й отримувати корисну інформацію для прийняття рішень у реальному часі.

Організація інформаційних потоків, зокрема через створення локальних мереж, що відповідають критеріям системності, оперативності, точності та динамічності, є важливим етапом у впровадженні Big Data. Такі мережі забезпечують надійність і швидкість обробки даних, мінімізують витрати на

передачу інформації та сприяють підвищенню ефективності управлінських рішень. Використання моделювання як інструменту для оцінки впливу управлінських рішень на систему дозволяє уникати помилок, аналізувати альтернативи та обирати найбільш ефективні стратегії.

Результати дослідження мають високу практичну цінність. Вони стануть у пригоді підприємствам, які прагнуть мінімізувати витрати, підвищити рентабельність, забезпечити високий рівень безпеки даних і покращити якість управління. Застосування системного аналізу в управлінні великими даними створює платформу для довгострокового розвитку підприємств, стимулює їхню продуктивність і сприяє досягненню стратегічних цілей. Завдяки поєднанню методів системного аналізу з технологіями Big Data підприємства отримують унікальні можливості для трансформації своєї діяльності, підвищення ефективності та впевненого виходу на нові ринки.

Таким чином, системний аналіз у поєднанні з технологіями Big Data не лише забезпечує якісне управління інформацією, але й стимулює інноваційний розвиток підприємств. Його впровадження дозволяє не лише вирішувати існуючі проблеми, але й запобігати можливим викликам у майбутньому, формуючи основу для стабільного зростання та процвітання.

ПЕРЕЛІК ПОСИЛАНЬ

1. S Maheshwari, P Gautam, CK Jaggi - International Journal of ..., 2021 - Taylor & Francis. Role of Big Data Analytics in supply chain management: current trends and future perspectives. [academia.edu](https://www.academia.edu)
2. Sahoo, S. (2021). Big data analytics in manufacturing: a bibliometric analysis of research in the field of business management. International Journal of Production Research, 60(22), 6793–6821. <https://doi.org/10.1080/00207543.2021.1919333>
3. H Lv, S Shi, D Gursoy - ... of Hospitality Marketing & Management, 2022 - Taylor & Francis. A look back and a leap forward: a review and synthesis of big data and artificial intelligence literature in hospitality and tourism. [researchgate.net](https://www.researchgate.net)
4. M Mariani, R Baggio - ... of Contemporary Hospitality Management, 2022 - emerald.com. Big data and analytics in hospitality and tourism: a systematic literature review. [unibo.it](https://www.unibo.it)
5. S Kaffash, AT Nguyen, J Zhu - International journal of production economics, 2021 - Elsevier. Big data algorithms and applications in intelligent transportation system: A review and bibliometric analysis. [deafrontier.net](https://www.deafrontier.net)
6. MM Rathore, SA Shah, D Shukla, E Bentafat... - IEEE ..., 2021 - ieeexplore.ieee.org. The role of ai, machine learning, and big data in digital twinning: A systematic literature review, challenges, and opportunities. [ieee.org](https://www.ieee.org)

ДЕМОНСТРАЦІЙНІ МАТЕРІАЛИ (Презентація)

ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
КАФЕДРА ІНФОРМАЦІЙНИХ СИСТЕМ ТА ТЕХНОЛОГІЙ

КВАЛІФІКАЦІЙНА РОБОТА

на тему: «Методика управління Big Data в інформаційних
системах на основі системного аналізу»

Виконала роботу студентка гр. САДМ-61

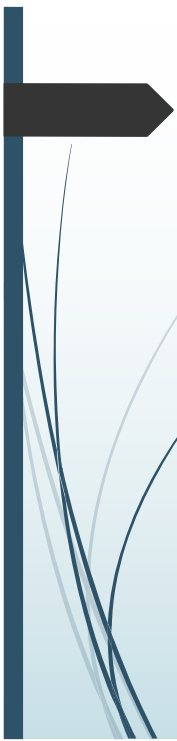
Інна ЛОГОША

Керівник роботи к. т. н. доцент

Віталій СВАТКО

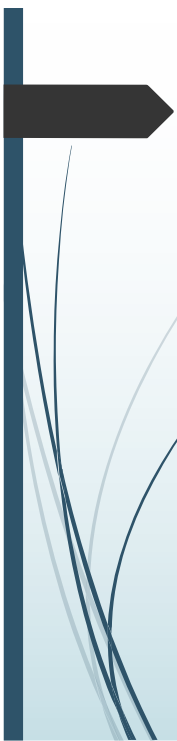
Вступ

- Актуальність:
 - - Великі дані (Big Data) – ключова технологія цифрової економіки.
 - - Необхідність системного аналізу для ефективного управління даними.
- Мета:
 - - Використання системного аналізу для оптимізації управління Big Data.



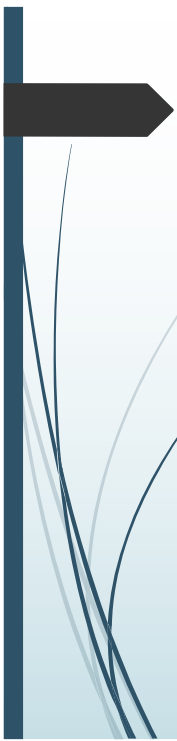
Основні характеристики Big Data

- Модель '4V':
 - - Обсяг (Volume)
 - - Швидкість (Velocity)
 - - Різноманітність (Variety)
 - - Достовірність (Veracity)
- Додаткові характеристики:
 - - Візуалізація (Visualization)
 - - Цінність (Value)



Основні характеристики Big Data

- Модель '4V':
 - - Обсяг (Volume)
 - - Швидкість (Velocity)
 - - Різноманітність (Variety)
 - - Достовірність (Veracity)
- Додаткові характеристики:
 - - Візуалізація (Visualization)
 - - Цінність (Value)



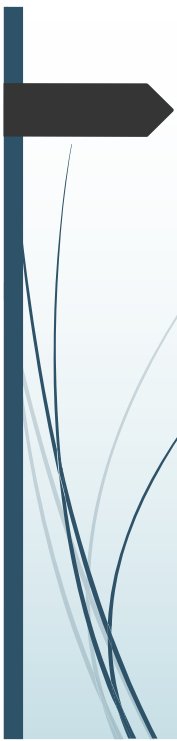
Використання Big Data в різних сферах

- 1. Бізнес: аналіз клієнтів, прогнозування попиту.
- 2. Фінанси: кредитний аналіз, виявлення шахрайства.
- 3. Охорона здоров'я: персоналізована медицина, контроль епідемій.
- 4. Наука: астрономія, геноміка.
- 5. Технології: IoT, управління мережею.



Системний аналіз у вирішенні проблем управління

- Системний аналіз дозволяє:
- - Визначати проблеми управління.
- - Будувати моделі для прогнозування.
- - Обирати найкращі стратегії вирішення.



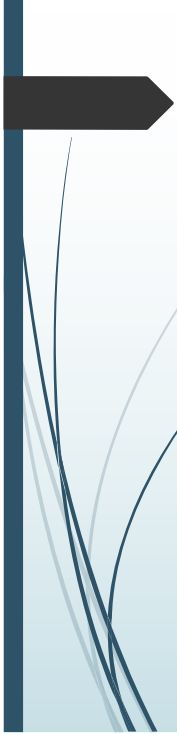
Основні етапи системного аналізу

- 1. Визначення меж системи.
- 2. Формування критеріїв ефективності.
- 3. Аналіз внутрішніх і зовнішніх факторів.
- 4. Побудова моделей управління даними.
- 5. Розробка альтернативних рішень.
- 6. Оцінка ризиків і варіантів оптимізації.
- 7. Запровадження рішень та їх моніторинг.



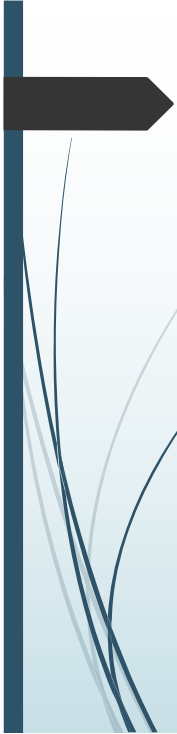
Технології для управління великими даними

- - Хмарні обчислення (Cloud Computing)
- - NoSQL та NewSQL бази даних
- - Машинне навчання та AI
- - Системи підтримки прийняття рішень (DSS)



Переваги впровадження системного аналізу

- - Оптимізація управління інформацією.
- - Зниження витрат на обробку даних.
- - Покращення точності ухвалення рішень.
- - Підвищення конкурентоспроможності.



Висновки та перспективи

- - Big Data – ключовий елемент цифрової трансформації.
- - Системний аналіз допомагає ефективно управляти даними.
- - Подальші дослідження необхідні для інтеграції Big Data у всі аспекти бізнесу.