

ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ТЕЛЕКОМУНІКАЦІЙ
КАФЕДРА СИСТЕМНОГО АНАЛІЗУ

КВАЛІФІКАЦІЙНА РОБОТА

на тему: «Системний аналіз для оптимізації управління біг даними (Big Data) в організаціях»

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*

(підпис)

Єгор ПАНЬКОВ
Ім'я, ПРІЗВИЩЕ здобувача

Виконав: здобувач вищої освіт
групи САДМ-61

Єгор ПАНЬКОВ

Керівник: Сергій КОЗАЧЕНКО

Рецензент: _____

Київ 2023

**ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ТЕЛЕКОМУНІКАЦІЙ**

Кафедра Системного аналізу

Ступінь вищої освіти Магістр

Спеціальність 124 Системний аналіз

Освітньо-професійна програма Інтелектуальні системи управління

ЗАТВЕРДЖУЮ

Завідувач кафедру СА

_____ Ровіл НАФЄЄВ

«___» _____ 20__р.

**ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ**

Паньков Єгор Олександрович

Тема кваліфікаційної роботи: Системний аналіз для оптимізації управління біг даними (Big Data) в організаціях

керівник кваліфікаційної роботи Сергій КОЗАЧЕНКО к.е.н. ,

затверджені наказом Державного університету інформаційно-комунікаційних технологій

від «___» _____ 20__р. № ____

1. Строк подання кваліфікаційної роботи «___» _____ 20__р.

2. Вихідні дані до кваліфікаційної роботи: методи обробки великих даних

Інтернет-ресурси стосовно методів оптимізації управління біг даними

Науково-технічна література з питань, пов'язаних з алгоритмами машинного навчання

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

1. Основні сучасні методи обробки великих даних

2. Застосування машинного навчання для оптимізації управління

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	Підбір науково-технічної літератури		
2	Основні сучасні методи обробки великих даних		
3	Застосування машинного навчання для оптимізації		
4	Вступ, висновки, реферат		
5	Розробка обов'язкових матеріалів		
6	Попередній захист роботи		

Здобувач(ка) вищої освіти

(підпис)

Єгор ПАНЬКОВ

(Ім'я, ПРІЗВИЩЕ)

Керівник
кваліфікаційної роботи

(підпис)

Сергій КОЗАЧЕНКО

(Ім'я, ПРІЗВИЩЕ)

РЕФЕРАТ

Текстова частина кваліфікаційної роботи на здобуття освітнього ступеня магістра: 64 стор., 27 рис., 1 табл., 21 джерел.

Мета роботи – розробка та впровадження системного аналізу для оптимізації управління біг даними в організаціях. Ми прагнемо створити фреймворк, який дозволить організаціям ефективно використовувати біг дані для підтримки стратегічних рішень та покращення їхньої конкурентоспроможності.

Об'єкт дослідження – процеси управління біг даними в організаціях.

Предмет дослідження – методи та стратегії системного аналізу, спрямовані на оптимізацію управління біг даними.

Короткий зміст роботи: створено комплексний підхід до системного аналізу управління біг даними та розроблено конкретні стратегії оптимізації, які будуть враховувати поточний стан організації та тенденції розвитку галузі.

ABSTRACT

Text part of the master's qualification work: 64 pages, 27 pictures, 1 table, 21 sources.

The purpose of the work – development and implementation of system analysis to optimize data management in organizations. We aim to create a framework that will enable organizations to effectively use big data to support strategic decisions and improve their competitiveness.

Object of research – data management processes in organizations.

Subject of research – methods and strategies of system analysis aimed at optimizing the management of big data

Summary of the work: a comprehensive approach to the system analysis of big data management has been created and specific optimization strategies have been developed, which will take into account the current state of the organization and trends in the development of the industry

**ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ТЕЛЕКОМУНІКАЦІЙ**

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

Направляється здобувач Панькво Є.О до захисту кваліфікаційної роботи
(прізвище та ініціали)
за спеціальністю 124 Системний аналіз
освітньо-професійної програми Інтелектуальні системи управління
на тему: «Системний аналіз для оптимізації управління біг даними (Big Data) в організаціях»

Кваліфікаційна робота і рецензія додаються.

Директор ННІТ _____
(підпис)

Владислав КРАВЧЕНКО
(Ім'я, ПРІЗВИЩЕ)

Висновок керівника кваліфікаційної роботи

Здобувач Паньков Єгор Олександрович під час написання кваліфікаційної роботи показав відмінну теоретичну підготовку, володінням всіх необхідних аспектів і методів з системного аналізу, користуватися навчальною, довідковою і науково-технічною літературою. Працюючи над завданнями, які доручались керівником, проявив ініціативність, сумлінність та хист до інженерної роботи.

Все це дозволяє оцінити виконану кваліфікаційну роботу здобувача Панькова Єгора Олександровича оцінку «відмінно» та присвоїти йому кваліфікацію «Магістр з системного аналізу».

Керівник кваліфікаційної роботи _____
(підпис)

Сергій КОЗАЧЕНКО
(Ім'я, ПРІЗВИЩЕ)

«___» _____ 20__ року

Висновок кафедри про кваліфікаційну роботу

Кваліфікаційна робота розглянута. Здобувач Паньков Є.О. допускається до захисту даної роботи в Експертній комісії.

Завідувач кафедрою Системного аналізу
(назва)

(підпис)

Ровіл НАФЄЄВ
(Ім'я, ПРІЗВИЩЕ)

«___» _____ 20__ року

ВІДГУК РЕЦЕНЗЕНТА

на кваліфікаційну магістерську роботу

здобувача вищої освіти Панькова Єгора Олександровича
на тему «Системний аналіз для оптимізації управління біг даними (Big Data) в організаціях»

Актуальність.

Актуальність теми полягає в тому, що організації стають перед викликом ефективного управління цими масштабними обсягами даних. Справжнім викликом є не тільки обробка цих обсягів інформації, але і вміння здобувати з неї цінну інформацію для прийняття стратегічних рішень.

Позитивні сторони.

1. Автор роботи може висвітлює широкий спектр аспектів системного аналізу, пов'язаних з оптимізацією управління біг даними. Робота включає ретельне вивчення різних методів, моделей та стратегій для ефективного управління обсягами великих даних.
2. Автор врахував реальні вимоги і потреби організацій, його робота є практично цінною. Автор розглядає конкретні випадки впровадження системного аналізу в організаціях і надає чіткі практичні рекомендації.
3. Автор вносить нові підходи, методи, ідеї у область системного аналізу біг даних, це може визначити його роботу як науково значущу.

Недоліки.

1. В роботі мають місце деякі неточності в перекладі технічних термінів.
- Відзначені зауваження не впливають на загальну позитивну оцінку кваліфікаційної роботи

Висновок: кваліфікаційна робота на здобуття ступеня магістра заслуговує оцінку "відмінно", а здобувач Паньков Єгор Олександрович заслуговує присвоєння кваліфікації: «Магістр з системного аналізу»

Рецензент:

науковий ступінь, вчене звання

_____ *підпис*

_____ *Ім'я, ПРИЗВИЩЕ*

ЗМІСТ

ВСТУП.....	8
РОЗДІЛ 1.....	14
РОЗДІЛ 2.....	41
РОЗДІЛ 3.....	48
ВИСНОВКИ.....	71
ПЕРЕЛІК ПОСИЛАНЬ.....	72
ДЕМОНСТРАЦІЙНІ МАТЕРІАЛИ (Презентація)	73

Вступ

Актуальність теми. В сучасному світі, що швидко розвивається, обсяги інформації зростають експоненційно, зокрема завдяки постійному розширенню цифрового середовища та швидкому розвитку технологій. Біг дані (Big Data) визначаються як великі обсяги даних, що надходять з різних джерел і вимагають ефективного збору, обробки та аналізу.

Актуальність теми полягає в тому, що організації стають перед викликом ефективного управління цими масштабними обсягами даних. Справжнім викликом є не тільки обробка цих обсягів інформації, але і вміння здобувати з неї цінну інформацію для прийняття стратегічних рішень.

Big Data використовується в різних сферах та галузях і забезпечує обробку та аналіз великих обсягів даних, які раніше були важко або неможливо опрацьовувати традиційними методами. Ось детальний огляд того, як саме використовуються Big Data:

а) Бізнес та Маркетинг:

1) Аналітика клієнтів: Організації використовують дані для аналізу поведінки клієнтів та створення персоналізованих пропозицій та рекламних кампаній.

2) Прогнозування попиту: Big Data дозволяє прогнозувати попит на товари та послуги, що полегшує планування виробництва та постачання.

3) Аналіз ринку: Вивчення даних допомагає підприємствам зрозуміти тенденції ринку, дізнатися про конкурентні переваги та виявити нові можливості.

б) Фінанси та Банківська справа:

1) Кредитний аналіз: Аналіз Big Data дозволяє оцінювати кредитний ризик та приймати швидкі та точні рішення щодо видання кредитів.

2) Протидія шахрайству: Виявлення небезпечних транзакцій та інші методи аналізу даних допомагають виявляти та уникати фінансових шахрайств.

в) Охорона Здоров'я:

1) Персоналізована медицина: Аналіз геномних даних та інших медичних даних дозволяє створювати персоналізовані методи лікування та прогнозування захворювань.

2) Епідеміологічні дослідження: Вивчення великих обсягів даних дозволяє вченим слідкувати за поширенням захворювань та розробляти стратегії їхнього контролю.

г) Наука та Дослідження:

1) Астрономія: Обробка великих обсягів даних дозволяє виявляти нові об'єкти в космосі та аналізувати космічні явища.

2) Геноміка: Використання геномних даних для розуміння генетичних аспектів хвороб та розробки нових методів лікування.

г) Технології та Інтернет речей:

1) Управління мережею: Big Data використовується для моніторингу та управління великою кількістю підключених пристроїв.

2) Оптимізація енергоспоживання: Аналіз даних дозволяє ефективно використовувати енергію в інтернеті речей та інших системах.

Це лише декілька прикладів використання Big Data, і його роль продовжує зростати в різних галузях. Організації впроваджують стратегії обробки та аналізу великих обсягів даних для прийняття більш обґрунтованих стратегічних рішень та забезпечення конкурентоспроможності.

Значущість теми підкреслюється тим, що правильне управління біг даними стає стратегічно важливим елементом успішності будь-якої сучасної організації. Ефективність управління біг даними визначає конкурентоспроможність, здатність до інновацій та забезпечення сталого розвитку.

Окрім того, з поширенням цифрової трансформації та збільшенням кількості підключених пристроїв (Інтернет речей), управління біг даними стає необхідністю для забезпечення безпеки та конфіденційності інформації.

Таким чином, дослідження системного аналізу для оптимізації управління біг даними набуває великого значення в контексті сучасних викликів та можливостей, які принесла ера цифрової трансформації.

Мета і завдання дослідження. Метою нашого дослідження є розробка та впровадження системного аналізу для оптимізації управління біг даними в організаціях. Ми прагнемо створити фреймворк, який дозволить організаціям ефективно використовувати біг дані для підтримки стратегічних рішень та покращення їхньої конкурентоспроможності.

Завдання дослідження:

а) Проаналізувати поточний стан управління біг даними в організаціях:

1) Визначити основні виклики, з якими стикаються організації у сфері управління біг даними.

2) Оцінити рівень інтеграції та використання біг даних у різних галузях.

б) Визначити основні принципи системного аналізу та їх застосування до управління біг даними:

1) Розглянути концепції системного аналізу та визначити, як вони можуть бути адаптовані для вирішення проблем управління біг даними.

в) Розглянути та обґрунтувати методи оптимізації управління біг даними:

1) Дослідити різні стратегії розподіленої обробки даних та масштабованості.

2) Оцінити ефективність використання технологій хмарних обчислень.

г) Розробити модель системного аналізу для управління біг даними:

1) Створити концептуальну модель, яка враховує специфіку біг даних та інтеграцію з існуючими системами.

г) Визначити критерії ефективності та впровадження моделі:

1) Розробити критерії оцінки ефективності впровадження системного аналізу в управління біг даними.

2) Визначити етапи та план впровадження розробленої моделі.

д) Аналізувати витрати та користі від оптимізації управління біг даними:

1) Вивчити фінансові аспекти впровадження нової системи та порівняти їх із очікуваними користями.

е) Визначити та оцінити майбутні тенденції управління біг даними:

1) Прогнозувати можливі технологічні та стратегічні зміни у сфері управління біг даними.

ж) Висновки та рекомендації:

1) Підсумувати результати дослідження та надати конкретні рекомендації для організацій щодо оптимізації управління біг даними.

Ці завдання дозволять створити комплексний підхід до системного аналізу управління біг даними та розробити конкретні стратегії оптимізації, які будуть враховувати поточний стан організації та тенденції розвитку галузі.

Об'єкт дослідження. Об'єктом нашого дослідження є процеси управління біг даними в організаціях.

Предмет дослідження. Предметом дослідження є методи та стратегії системного аналізу, спрямовані на оптимізацію управління біг даними.

Джерела дослідження. Основні джерела інформації включатимуть літературні джерела, наукові статті, звіти організацій, що вже застосовують системний аналіз в управлінні біг даними, та інші відповідні ресурси.

Наукова новизна одержаних результатів. Наукова новизна полягатиме у визначенні конкретних аспектів системного аналізу, які можуть бути успішно впроваджені для оптимізації управління біг даними в сучасних умовах.

Практичне значення одержаних результатів. Результати дослідження будуть мати практичне значення для організацій, які прагнуть покращити своє управління біг даними, забезпечуючи ефективність, безпеку та конкурентоспроможність на ринку.

Розділ 1

Постановка проблеми у загальному вигляді та її зв'язок із важливими науковими і практичними завданнями. Складність управлінських завдань продовжує зростати, що вимагає постійного розвитку та удосконалення методологічних положень, методів і методичних рекомендацій. Наукова обґрунтованість рішень, їх оптимальність залежить, з одного боку, від ступеня досконалості методів, що використовуються в процесі ухвалення та реалізації рішень, з іншого — від рівня знань і володіння персоналом управління комплексом методів. У зв'язку з цим виникає необхідність уточнення теоретико-методологічних засад системного аналізу у вирішенні економічних завдань управління діяльністю підприємства.

Аналіз останніх досліджень і публікацій, в яких розпочато вирішення проблеми, що висвітлюється, і на які спирається автор. Проблемам теорії і практики системного аналізу в економічній сфері та управлінні присвячено роботи зарубіжних і вітчизняних дослідників, таких як В.С. Анфілатова, А.Л. Ємельянова, А.А. Кукушкіна [1], І. Блауберга [2], В. Лямця [3], С. Оптнера [4], О. І. Пономаренка [5], Ю. І. Черняка [6], Е. Юдіна [7], С. Янга [8] та ін. При цьому, на сьогодні, не існує єдиної точки зору на зміст поняття «системний аналіз» і етапи процесу прийняття управлінських рішень.

Виділення не вирішених раніше частин загальної проблеми, яким присвячується стаття. Залишаються невирішеними питання щодо практичного застосування системного аналізу під час ухвалення управлінських рішень на підприємствах і організаціях у всіх галузях економіки України.

Метою роботи є дослідження використання системного аналізу в управлінні та розробка практичних рекомендацій для забезпечення ухвалення оптимальних управлінських рішень. Розглянуті методичні підходи та методи до вирішення складних слабо структурованих проблем в управлінні. Виклад основного матеріалу дослідження. Від економічного обґрунтування

управлінських рішень залежить успіх діяльності будь-якої компанії. На сьогодні розроблено ряд наукових методів для підготовки рішень, щоб керівники могли ухвалювати їх обґрунтовано, вибираючи найкращий з можливих варіантів. Однак ці методи ще недостатньо активно використовуються в практиці управління. Значна частина керівників, покладаючись на свій досвід і на своє мистецтво управління, не приділяє належної уваги вивченню та впровадженню наукових методів управління.

Одночасне використання мистецтва управління та наукових методів і підходів дає високу результативність в управлінській діяльності. Від характеру проблеми, яку необхідно розв'язати, залежить процес ухвалення рішень і його суть. Управлінське рішення можна класифікувати залежно від числа можливих рішень, числа станів навколишнього середовища, зміни тих чи інших його параметрів. Універсальну систему ухвалення рішень створити неможливо, щоб вона розв'язувала всі типові управлінські задачі, а також враховувала усі випадки. У наукових публікаціях, підручниках з менеджменту наведені схеми процедури ухвалення рішень для типових управлінських задач, і в яких неодмінною складовою є системний аналіз [1].

Сучасне підприємство, яке діє в умовах ринкової економіки, представляє собою складну систему, що стикається з численними викликами та конкурентною боротьбою. Для досягнення конкурентних переваг та успіху компанії важливо розуміти його діяльність та взаємодії з навколишнім середовищем. В цьому контексті системний аналіз виступає як ефективний метод дослідження складних систем, що дозволяє аналізувати взаємозв'язки та взаємодії між їх складовими елементами.

Системний аналіз є сучасним науковим напрямком, який використовує методологію загальної теорії систем для дослідження об'єктів у різних галузях, таких як економіка, управління, техніка, виробництво та інші. Однією з його

основних мет є знаходження шляхів вирішення проблемних ситуацій, що виникають у предметних областях, які аналізуються.

У процесі системного аналізу використовуються принципи теорії систем, системний підхід, апарат дослідження операцій, кібернетики та системотехніки. Його завданням є розкриття змісту проблем перед керівниками для більш чіткого розуміння наслідків прийняття рішень. Це допомагає керівникам у більш обґрунтованому виборі оптимальних стратегій та врахуванні невизначеності та ризиків.

Використання системного аналізу в управлінні дозволяє встановлювати причинно-наслідкові зв'язки, аналізувати різні варіанти розв'язання проблем та надавати обґрунтовані рекомендації. Цей підхід також допомагає керівникам у виборі ефективних засобів досягнення поставлених цілей в умовах невизначеності та конкуренції на ринку.

Таким чином, системний аналіз є потужним інструментом управління, який допомагає розробляти стратегії, вибирати оптимальні рішення та досягати успіху в сучасному бізнес-середовищі.

Хто є споживачем даного товару? Відповіді на це питання можуть варіюватися. Зазвичай існують альтернативи, які важко порівнювати між собою, оскільки вони не мають однорідних одиниць вимірювань, або якщо вони мають, то розрахунки можуть бути умовними та відхилятися від реальності. Системний аналіз спрямований на порівняння варіантів рішень та вибір найоптимальнішого. Його метою є повне і всебічне порівняння різних можливостей вирішення проблеми з урахуванням якісних та кількісних показників витрачених ресурсів.

Використання сучасного системного аналізу в управлінні дозволяє визначати причинно-наслідкові зв'язки, аналізувати варіанти вирішення системних проблем з урахуванням обмежень, ризиків та невизначеностей

середовища. Він також надає обґрунтовані рекомендації для оптимального вибору або раціональної стратегії управління в складних ситуаціях. Методи моделювання використовуються для детального вивчення проблем.

Процес постановки задачі прийняття рішення може включати наступні кроки. Спочатку потрібно визначити межі системи, яка підлягає оптимізації. Далі обирається критерій ефективності для оцінки управлінських рішень. Внутрішні показники, що адекватно описують функціонування системи, обираються для аналізу. Завершальним етапом є побудова моделі, яка відображає взаємозв'язки між показниками та вплив незалежних факторів на результативність системи.

У сфері бізнесу ключовою потребою є обґрунтування управлінських рішень. Процес вибору рішення завжди розпочинається з виникнення проблемної ситуації та закінчується обранням одного з можливих варіантів. Обране рішення повинно ефективно вирішити проблему. Керівник стає важливим у тих випадках, коли в системі з'являється невизначеність, невідомість або ризик. Раціональність управлінських рішень значною мірою залежить від самого процесу їх підготовки та прийняття.

В теорії управління будь-який процес прийняття управлінських рішень складається із таких етапів:

- 1) підготовка і прийняття рішення.
- 2) реалізація прийнятого рішення.
- 3) оцінка наслідків реалізації результатів.

Розглядаючи методи системного аналізу, можна визначити, що для прийняття рішення потрібно пройти кілька етапів. Спочатку на першому етапі проводиться аналіз проблеми, визначення системи, розгляд структури системи,

визначення загальної мети та критерію системи, а також декомпозиція мети та визначення необхідних ресурсів для вирішення проблеми.

У початковий момент важливо визначити, чи є наявна економічна проблема реальною, а не уявною. Ключовою є точність формулювання проблеми, наприклад, «проблема нестачі програмістів для виконання проекту». Важливо структурувати проблему та вивчити її динаміку у часі. Також важливо визначити, чи є наша проблема пов'язаною з іншими, і якщо так, то як, а також розуміти, чи можна вирішити проблему в принципі чи ні. На цьому етапі необхідно визначити:

а) сутність проблеми, знайти причину наслідковий зв'язок. Потрібно з'ясувати такі питання: визначити особу, яка відповідальна за розв'язання проблеми; за яких обставин потрібно прийняти рішення; визначити термін розв'язання проблеми, чи потрібно взагалі приймати рішення;

б) це системна чи разова проблема;

в) кінцевий результат розв'язання проблеми;

г) критерії та обмеження щодо розв'язання проблеми. Необхідно визначити ознаки, на підставі яких буде проводитись оцінка розв'язання проблемної ситуації, упорядкувати ці ознаки за ступенем важливості.

На цьому етапі необхідно визначити деталі поставленого завдання, визначити об'єкти та їх компоненти для визначення межі проблеми, яку потрібно вирішити. Під час аналізу структури системи важливо побудувати ієрархію рівнів проблеми, розглянути її якщо вона має багатоаспектний характер, та провести специфікацію підсистем і процесів, якщо такі існують.

Під час аналізу структури системи необхідно побудувати ієрархію рівнів поставленої перед нами проблеми, якщо вона є багатогранною, зробити специфікацію підсистем і процесів проблеми, якщо такі є.

Потім визначається загальна мета та формуються критерії оцінки проблеми, а також їх декомпозиція за підсистемами. На цьому етапі проводиться ранжування цілей від верхнього рівня до нижнього, формулювання поточних та перспективних цілей, виявлення потреб у ресурсах. Оцінюються необхідні ресурси для досягнення мети, аналізуються наявні технології та ресурси, оцінюється стан проектів, які вже реалізовані, і тих, які заплановані.

На основі аналізу розробляються альтернативні варіанти розв'язання проблеми. Після цього проводиться оцінка всіх варіантів, і на її основі вибирається оптимальний варіант та приймається рішення.

У системному аналізі проблеми поділяються на три класи: добре структуровані (кількісно сформульовані), неструктуровані (якісно виражені), та слабо структуровані (змішані). Визначення критеріїв оцінки системи має важливе значення для управлінських рішень, оскільки вони часто не є очевидними і вимірюють комплексний ефект рішення на зацікавлені сторони.

Стейкхолдери впливають на управлінські рішення компанії і поділяються на "внутрішніх" (власники, співробітники) і "зовнішніх" (споживачі, постачальники, конкуренти, державні органи, некомерційні організації, ЗМІ). Розв'язання проблеми у системному аналізі включає в себе декомпозицію проблеми на завдання, які можуть бути новими або відомими. Вирішення економічних проблем вимагає отримання оцінок експертів та їх обробки.

Для неструктурованих проблем складним стає саме визначення проблеми. Системний аналіз в управлінні економікою дуже часто має справу зі слабо структурованими або зовсім неструктурованими проблемами. Саме тому отримання оцінок експертів, а також їх обробка є необхідними елементами системного аналізу більшості економічних проблем.

На етапі проектування формується комплекс заходів, моделей і проектів, необхідних для ефективного вирішення поставленої проблеми. Окрім цього, визначається порядок досягнення різних цілей, проводиться розподіл сфер діяльності та компетенцій. Усі ці кроки реалізуються в умовах тимчасових і ресурсних обмежень. На даному етапі важливо не лише визначити, як система буде функціонувати, а й врахувати її взаємодію з зовнішнім середовищем.

Побудова моделей є ключовим елементом цього процесу, де формалізований опис відображає зв'язок між цілями, засобами та очікуваними результатами. Застосування методів моделювання виявляється важливим етапом для вирішення практичних проблем. Це дозволяє отримати додаткові відомості про системи, проводити перевірку пропонованих рішень та виявляти нові варіанти розв'язання.

Після аналізу проблеми розробляються різні альтернативи для її вирішення. Цей процес включає у себе створення максимально можливої кількості варіантів. Це не лише допомагає уникнути шаблонних рішень, але й дозволяє провести всебічне порівняння та обрати оптимальний шлях вирішення проблеми.

Ключовим моментом прийняття управлінського рішення є операція вибору найкращої альтернативи з переліку розглянутих і аналізованих. Цей вибір базується на порівнянні різних варіантів за різними параметрами, встановленими під час моделювання. Оцінка включає показники, такі як ефективність рішення, час вирішення, витрати ресурсів, очікувані наслідки та ризики невиконання рішення.

Поміж визначеними етапами проектування важливо також враховувати зв'язок із соціальними та екологічними аспектами, а також передбачати можливі додаткові сценарії вирішення проблеми для більшої гнучкості у реагуванні на змінні умови.

Пошук та формування альтернативних рішень для вирішення поставленої проблеми представляє собою творчий процес, що вимагає розгляду різноманітних можливостей. Чим більше варіантів розв'язання розглядається, тим більше шансів знайти оптимальне рішення в умовах заданих обмежень. Однак проведення повного порівняння всіх альтернатив є складним завданням через значні витрати, тому зазвичай обмежуються аналізом лише кількох оптимальних варіантів.

Під час вибору найкращого рішення, його обговорення та прийняття, учасники процесу зобов'язані враховувати різні показники, такі як витрати ресурсів, часу, ризику та наслідки прийнятого рішення. Залежно від параметрів, одні варіанти можуть виявитися кращими за інші. Важливо зазначити, що раціональне рішення проблеми може вимагати значних витрат ресурсів і тривалого часу.

Умовність визначення оптимального варіанту рішення обумовлена тим, що єдиного критерію оптимальності рідко буває. Здебільшого при виборі за багатьма критеріями перевага надається прийнятному варіанту, який виглядає найкращим з певних позицій для осіб, що приймають рішення.

Після визначення оптимального варіанту, його реалізація може бути передана іншим підрозділам або проектним групам у компанії. На етапі реалізації широко використовуються методи планування, організації та контролю для ефективного виконання рішення.

Складання плану реалізації включає в себе визначення дій: що, кому, з ким, як, де і коли робити. Ці дії повинні бути чітко документовані. На етапі контролю за реалізацією використовуються різні методи, такі як контроль за проміжними та кінцевими результатами, а також контроль за дотриманням графіка виконання.

Головна мета системи контролю - забезпечення гарантій виконання рішень та максимальної якості їх реалізації.

На етапі аналізу наслідків рішення проводиться оцінка ефективності та результативності в порівнянні із зазначеною метою. Ефективні рішення дозволяють досягти поставлених цілей при мінімальних витратах ресурсів. Важливо врахувати кількісні та якісні наслідки рішення, здійснювати зворотний зв'язок зі зацікавленими сторонами та обчислювати показники ефективності. Прийняття управлінських рішень під впливом різноманітних факторів, таких як ступінь ризику, відведений час та політика організації, є складним і багатограним процесом.

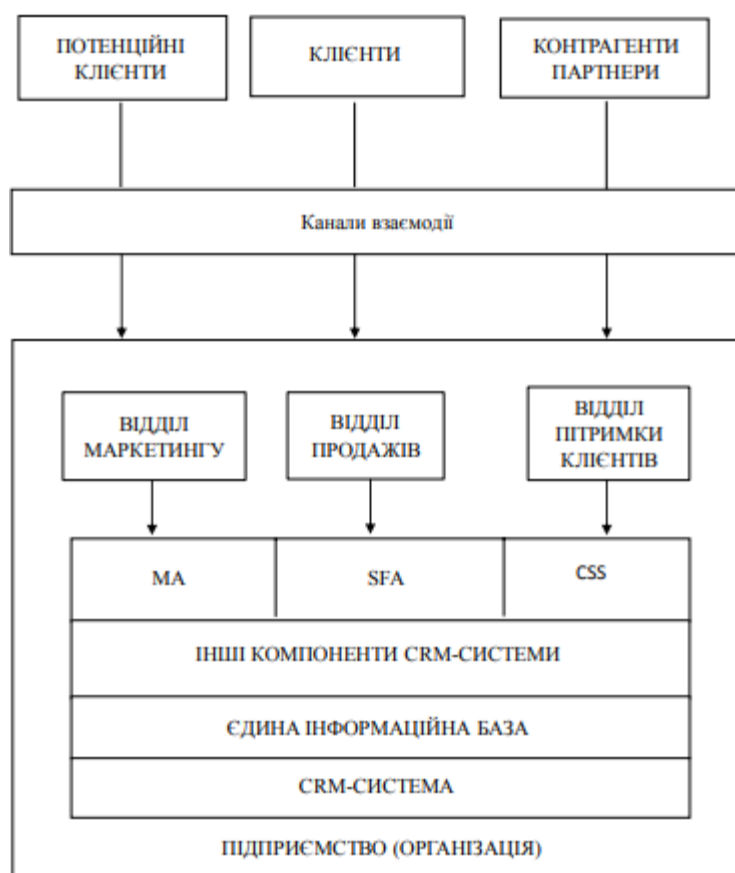
На етапі аналізу наслідків прийняття рішення проводиться докладна оцінка його ефективності, враховуючи початкові постановки завдання. Визнається, що ефективним є таке рішення, яке дозволяє досягти максимальних цілей при мінімальних витратах ресурсів. Важливо розглядати якісні та кількісні наслідки рішення, залучати зацікавлені сторони у дискусії, проводити зворотний зв'язок та розраховувати ключові показники ефективності. Процес ухвалення управлінських рішень піддається впливу низки факторів, таких як рівень ризику, виділений час для прийняття рішення та організаційна політика.

Успіх у сучасному бізнесі обумовлений не лише вірністю вибраним напрямкам розвитку компанії та стратегіям маркетингу, але також технічною підтримкою прийняття рішень, автоматизацією управління підприємством та ефективною функціональністю його інформаційної системи.

Для спрощення операцій в бізнесі необхідна автоматизація різноманітних бізнес-процесів. Однією з таких ключових технологій є використання систем управління відносинами з клієнтами (CRM). CRM-менеджери, які вивчають взаємодію фірми з клієнтами, спрямовані на забезпечення максимально високого

рівня обслуговування та лояльності клієнтів. Вони також активно працюють над залученням нових клієнтів та оптимізацією процесів. CRM-системи надають можливість автоматизувати відносини з клієнтами, створювати та підтримувати єдину базу клієнтів і контрагентів, ефективно контролювати роботу відділу продажів, отримувати аналітичні дані та поліпшувати взаємодію з клієнтами.

Для ефективної діяльності CRM-фахівці використовують різні програмні засоби, включаючи CRM-системи, які сприяють автоматизації бізнес-процесів,



забезпечують взаємодію всіх підрозділів підприємства з клієнтами відповідно до концепцій CRM. Ці системи не лише розв'язують завдання, спрямовані на задоволення та утримання клієнтів, але й оптимізують діяльність компанії, меншуючи витрати на пошук та обробку інформації, управління продажами та інші бізнес-процеси.

***Рис. 1. Схема управління взаємодією підприємства (організації)
з клієнтами за допомогою CRM-системи***

Способи взаємодії клієнтів із CRM-менеджерами розподілені між різними каналами, такими як електронна пошта, телефон, sms-повідомлення, факс, особисті зустрічі з менеджером, ICQ, корпоративний веб-сайт, точки продажу продукції та участь у соціальних мережах. Основові функціональні компоненти більшості CRM-систем включають модуль SFA (Sales Force Automation) - автоматизація роботи торгових представників; модуль MA (Marketing Automation) - автоматизація маркетингових процесів; модуль CSS (Customer Service & Support) - автоматизація обслуговування та підтримки клієнтів.

Після завершення будь-якої акції чи маркетингового заходу важливо провести аналіз їхньої ефективності, щоб належним чином зіставити витрати та досягнуті результати. Цей аналіз дозволяє виявити помилки та підготувати поліпшений план для майбутніх акцій.

Розробка універсальної формалізованої моделі управління взаємодією підприємства чи організації з клієнтами зустрічає значні труднощі через складність та низьку структурованість цього процесу. Незважаючи на це, підвищення ефективності цього процесу можливо шляхом впровадження та використання систем підтримки прийняття рішень.

Висновки. Системний аналіз має визначальне значення для ефективного управління підприємством, компанією чи організацією, оскільки він дозволяє виявляти проблеми та можливості для їх розв'язання. Системний аналіз об'єднує різні методи дослідження складних соціально-економічних систем, визначення проблем та планування дій для їх вирішення. Тому кожне управлінське рішення повинно ґрунтуватися на результатах системного аналізу. Знання менеджерами про етапи процесу управлінського рішення та методи забезпечення якості рішень,

які приймаються на кожному етапі, з урахуванням ситуаційних характеристик і чинників, що впливають на формування та реалізацію рішень, сприяє підвищенню ефективності управління організацією в цілому.

Великі дані (Big Data) представляють собою обсяги структурованих і неструктурованих даних величезних розмірів та різноманітності, які можуть бути ефективно оброблені програмними інструментами, що горизонтально масштабуються. Це поняття з'явилося наприкінці 2000-х років та стало альтернативою традиційним системам управління базами даних та рішенням у сфері бізнес-аналітики [1].

Згідно з аналітиками компанії IBS, глобальний обсяг даних демонструє наступні розміри [2]:

- 2003 г. – 5 ексабайт даних (1 ЕБ = 1 млрд гігабайтів)
- 2008 г. – 0,18 зетабайта (1 ЗБ = 1024 ексабайти)
- 2015 г. – більше 6,5 зетабайт
- 2021 г. – 44–48 зетабайта
- 2025 г. – цей об'єм збільшується ще у 10 разів.

Великі дані – це сукупність технологій, покликаних здійснювати три операції:

а) Обробляти більші, у порівнянні зі «стандартними» сценаріями, об'єми даних.

б) Уміти працювати з даними, що швидко надходять у дуже великих об'ємах. Тобто даних не просто багато, а їх постійно стає все більше й більше.

в) Вміти працювати зі структурованими і мало структурованими даними паралельно і у різних аспектах.

Вважається, що ці характеристики дозволяють виявляти приховані закономірності, які виходять за межі людського сприйняття, відкриваючи безпрецедентні можливості оптимізації різних сфер: державного управління, медицини, телекомунікацій, фінансів, транспорту, виробництва тощо.

Основні принципи роботи з великими даними визначаються характеристиками VVV (фізичний об'єм, швидкість приросту, різноманіття). Цей підхід був розроблений компанією Meta Group у 2001 році для підкреслення важливості управління даними в усіх аспектах.

Згодом з'явилися інші інтерпретації, такі як чотири V (додатково - достовірність), п'ять V (життєздатність та цінність), сім V (змінність та візуалізація). Наприклад, компанія IDC інтерпретує четверте V як цінність, підкреслюючи економічну доцільність обробки великих обсягів даних відповідно до умов [3]. За визначеннями, основні принципи взаємодії з великими даними включають:

Горизонтальна масштабованість. Це — базовий принцип обробки великих даних. Як вже було зазначено, великих даних з кожним днем стає все більше. Відповідно, необхідно збільшувати кількість обчислювальних вузлів, за якими розподіляються ці дані, при чому обробка має відбуватись без погіршення продуктивності.

Відмовостійкість. Цей принцип витікає з попереднього. Оскільки обчислювальних вузлів у кластері може бути багато (іноді десятки тисяч) та їх кількість, не виключено, буде збільшуватись, зростає ймовірність виходу машин з ладу. Методи роботи з великими даними мають враховувати ймовірність таких ситуацій і передбачати превентивні заходи.

Локальність даних. Оскільки дані розподілені по великій кількості обчислювальних вузлів, то, якщо вони фізично знаходяться на одному сервері, а

обробляються на іншому, витрати на передачу даних можуть бути невиправдано великими. Тому обробку даних бажано проводити на тій же машині, на якій вони зберігаються

Ці принципи відрізняються від тих, які характерні для традиційних, централізованих, вертикальних моделей зберігання добре структурованих даних. Власне, для роботи з великими даними розробляються підходи і технології.

Методи та прийоми аналізу великих даних представлені різними інструментами та підходами. Компанія McKinsey, що спеціалізується на стратегічному управлінні, виділяє 11 ключових методів і технік, які застосовуються до великих обсягів і різноманітності даних.

Серед цих методів важливе місце займає Data Mining, який включає в себе виявлення раніше невідомих знань у даних, таких як навчання асоціативним правилам, класифікація, кластерний та регресійний аналіз. Крім цього, використовується краудсорсінг, що дозволяє класифікувати та збагачувати дані силами великої кількості осіб, що не належать до трудових стосунків.

З метою глибинного аналізу, використовуються методи змішування та інтеграції даних, такі як цифрова обробка сигналів та обробка природної мови, включаючи тональний аналіз.

Машинне навчання, що включає навчання з учителем та без учителя, забезпечує створення моделей на основі статистичного аналізу для отримання прогнозів на основі базових моделей.

До інших методів входять штучні нейронні мережі, мережевий аналіз, оптимізація, включаючи генетичні алгоритми, які використовуються для

розв'язання задач оптимізації та моделювання шляхом випадкового підбору та комбінування параметрів.

Принципи обробки великих даних включають горизонтальну масштабованість, відмовостійкість та локальність даних. Система великих даних повинна бути розширюваною, відмовостійкою та враховувати локальність даних для зниження витрат на їх передачу.

Застосування великих даних активно використовується у фінансовому та медичному секторах, в інтернет-компаніях та в державному управлінні.

Сучасний ринок також свідчить про значний розвиток сервісів обробки великих даних. Наприклад, з 2012 року у бізнес-аналітиці успішно впроваджено хмарний сервіс Google BigQuery, що забезпечує аналіз великих даних у режимі реального часу. Інші провідні компанії також пропонують сервіси, побудовані на аналізі великих даних, використовуючи їх для фільтрації спаму, прогнозу пробок, управління безпілотними автомобілями та інші іноваційні застосування.

Великі обсяги даних знаходять своє активне застосування в різних галузях, особливо в фінансах, медицині, технологічних та інтернет-компаніях, а також в сфері державного управління.

Розвиток сервісів Big Data стає очевидним з початку 2010-х років. Наприклад, у 2012 році Google введе на ринок бізнес-аналітики свій хмарний сервіс Google BigQuery, який забезпечує аналіз великих даних у режимі реального часу. За рік цей сервіс успішно інтегрується з Google Analytics Premium - платною версією веб-лічильника. Недавно компанія Google представила Cloud Bigtable - масштабований хмарний сервіс баз даних, який є ще однією новітньою доповненням до їхнього портфелю продуктів.

Велика кількість сервісів провідних компаній, які займаються обробкою великих обсягів даних, базується на застосуванні різноманітних технологій

аналізу великих даних. Ці технології охоплюють різні аспекти, такі як пошукові алгоритми на основі нейромереж, машинний переклад, фільтрація спаму, таргетинг у контекстній рекламі, прогнозування пробок та погоди, розпізнавання мови і образів, а також системи управління безпілотними автомобілями та інші інноваційні застосування.

Аналіз переваг і обмежень використання технологій Big Data в бізнесі розкриває ряд ключових переваг, які варто враховувати при їхньому впровадженні:

- 1) Полегшення планування: Використання Big Data спрощує процес планування, дозволяючи бізнесу більш ефективно визначати свої стратегії.
- 2) Швидкість запуску нових проектів: Технології Big Data сприяють швидшому запуску нових проектів завдяки швидкому аналізу великих обсягів даних.
- 3) Збільшення шансів на затребуваність проекту: Використання Big Data дозволяє бізнесу більш точно прогнозувати, які проекти будуть популярними та затребуваними на ринку.
- 4) Оцінка задоволеності користувачів: Можливості оцінки задоволеності користувачів за допомогою аналізу даних дозволяють покращити якість продуктів і послуг.
- 5) Спрощення пошуку та залучення цільової аудиторії: Застосування Big Data спрощує і поліпшує стратегії залучення та утримання цільової аудиторії.
- 6) Прискорення взаємодії з клієнтами та контрагентами: Використання технологій Big Data дозволяє ефективно прискорити комунікацію та взаємодію з клієнтами та партнерами.

7) Оптимізація інтеграції в ланцюгу постачань: Big Data допомагає оптимізувати і покращити інтеграцію в ланцюгу постачань, забезпечуючи ефективну логістику та управління запасами.

8) Підвищення якості клієнтського сервісу та швидкості взаємодії: Аналіз великих даних сприяє підвищенню якості обслуговування клієнтів та прискоренню взаємодії з ними.

9) Підвищення лояльності поточних клієнтів: Застосування Big Data дозволяє бізнесу зрозуміти потреби та очікування своїх клієнтів, що призводить до підвищення рівня лояльності.

Застосування технологій Big Data в маркетингу відкриває широкі можливості для бізнесу:

1. Розуміння бізнесу в цифрах: Big Data допомагає здійснити глибокий аналіз фінансових показників та визначити ефективні стратегії.

2. Вивчення конкурентів: Аналіз даних дозволяє отримати інсайти щодо дій конкурентів та адаптувати власні стратегії.

3. Дослідження клієнтів: Використання Big Data дозволяє здійснювати глибокий аналіз клієнтської поведінки та потреб.

4. Підвищення рівня розуміння та аналітики в маркетингу: Технології Big Data розвивають нові можливості для зменшення витрат і збільшення обсягів продажів.

5. Створення точних портретів цільових споживачів: Big Data дозволяє маркетингологам створювати більш точні та деталізовані профілі своєї аудиторії.

6. Передбачення реакції споживачів на маркетингові повідомлення: Застосування Big Data дозволяє попередньо передбачати реакції клієнтів на різноманітні маркетинг

7. Прогнозування реакції споживачів на маркетингові повідомлення: Впереджаючи реакції клієнтів на різноманітні маркетингові повідомлення, компанії можуть ефективно адаптувати свої стратегії та забезпечити більшу ефективність комунікації.

8. Максимальна індивідуалізація рекламних повідомлень: Застосування Big Data дозволяє створювати на максимально індивідуальному рівні рекламні повідомлення, враховуючи особисті попередні покупки, вподобання та поведінку кожного клієнта.

9. Зростання крос-продажів, повторних продажів, ремаркетингу: Використання технологій Big Data дозволяє ефективно впливати на клієнтську базу для збільшення крос-продажів, стимулювання повторних покупок та успішного проведення ремаркетингових кампаній.10. Пошук і визначення причин популярності затребуваних товарів і продуктів.

11. Аналіз та визначення причин популярності високо популярних товарів та продуктів: Великі дані дозволяють проводити більш глибокий аналіз попиту, сприяючи виявленню факторів, що впливають на популярність конкретних товарів і виробів.

12. Вдосконалення продуктів і послуг, підвищення лояльності клієнтів: Використання аналітики Big Data дозволяє компаніям постійно вдосконалювати свою продукцію та послуги, враховуючи реальні потреби та вимоги клієнтів, що призводить до підвищення рівня лояльності.

13. Підвищення якості обслуговування: Застосування технологій Big Data допомагає удосконалити процеси обслуговування, роблячи їх більш ефективними та зорієнтованими на потреби клієнтів.

14. Зменшення витрат на взаємодію з постачальниками та клієнтами: Використання Big Data спрощує та оптимізує процеси взаємодії з

постачальниками та клієнтами, що дозволяє знижувати витрати компанії на цих етапах.

Сучасний обсяг літератури щодо управління біг даними та системного аналізу дозволяє глибоко вивчити різні аспекти цієї теми. Доцільно розглянути роботи, які зосереджуються на стратегіях оптимізації управління біг даними та використанні системного аналізу для розв'язання цього завдання.

1. "Big Data: A Revolution That Will Transform How We Live, Work, and Think" (Великі дані: революція, яка трансформує наше життя, роботу і мислення) - Віктор Майер-Шенбергер та Кеннет Кукуєр:

Ця книга надає комплексний огляд Big Data, зосереджуючись на його впливі на суспільство та бізнес. Виділено важливі аспекти збору та аналізу великих обсягів даних.

2. "Big Data Analytics: A Literature Review Paper" - Ваніш Майа та Інга Хаврлішек:

Ця робота надає систематичний огляд літератури з аналізу великих даних. Автори аналізують різні підходи до обробки та використання великих обсягів даних у різних галузях.

3. "System Analysis of Big Data Analytics in Healthcare: A Survey" - Асім Інам та Шакіл Ахмед:

Ця стаття фокусується на застосуванні системного аналізу до аналітики великих даних у сфері охорони здоров'я. Вона визначає ключові виклики та переваги використання системного аналізу для оптимізації управління біг даними в цій галузі.

4. "A Systematic Literature Review on Big Data Adoption: An Integrated View" - Елісавет Гречу та інші:

Робота виконує систематичний огляд літератури з проблеми впровадження великих даних. Вона надає інтегрований погляд на аспекти, що впливають на прийняття великих даних в організаціях.

Порівняльний аналіз результатів. За результатами аналізу літератури визначено, що багато досліджень фокусуються на технічні аспекти великих даних та їхній аналіз. Проте, обмежений обсяг робіт засвідчує потребу в детальному вивченні системного аналізу як стратегічного інструменту для оптимізації управління біг даними в організаціях.

Аспекти, які ще не дуже докладно досліджено:

1) Інтеграція системного аналізу в сучасні підходи до управління біг даними:

- Не всі дослідження розглядають взаємодію системного аналізу з існуючими стратегіями управління біг даними.

2) Адаптація системного аналізу до реальних умов великих обсягів даних:

- Деякі дослідження надто теоретичні; не вистачає робіт, що досліджують реалізацію концепцій системного аналізу в умовах реального виробництва.

3) Визначення оптимальних стратегій впровадження системного аналізу:

- Бракує комплексних підходів до визначення етапів впровадження системного аналізу в організаціях.

Інтеграція системного аналізу в сучасні підходи до управління біг даними передбачає використання концепцій та методів системного аналізу як основного інструменту для вдосконалення процесів управління та аналізу великих обсягів даних в організації. Давайте розглянемо це більш детально:

1) Системи та підсистеми: Великі дані можуть бути розглянуті як складна система, що складається з різних компонентів (підсистем), таких як джерела даних, інструменти аналізу, та людський фактор.

2) Взаємозв'язки та взаємодії: Аналіз великих даних потребує розуміння взаємозв'язків та взаємодій між різними частинами системи, що може бути досягнуто через методи системного аналізу.

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

4) Ідентифікація взаємозв'язків: Застосування системного аналізу допомагає виявляти складні взаємозв'язки та залежності між різними частинами системи управління біг даними.

5) Комплексний підхід: Системний аналіз дозволяє розглядати управління біг даними як комплексну систему, що сприяє глибшому розумінню та оптимізації всіх її аспектів.

6) Ефективність та стратегічність: Використання системного аналізу дозволяє розробляти більш ефективні та стратегічні підходи до управління біг даними, забезпечуючи врахування усіх факторів впливу.

Невивчені аспекти. Адаптація системного аналізу до вимог біг даних: Деякі аспекти, такі як адаптація традиційних методів системного аналізу до умов великих обсягів, ще не отримали належного дослідження Роль системного аналізу у розробці стратегій обробки великих даних: Важливо розглядати системний аналіз як ключовий елемент у визначенні та розробці стратегій обробки великих даних. Інтеграція системного аналізу в управління біг даними дозволить створити більш гнучкі та ефективні стратегії, що враховують складність та взаємодії у цьому динамічному середовищі.

Для якісного управління важливим є питання попереднього аналізу та виявлення всіх невикористаних можливостей. У широкому розумінні управління – це вплив суб'єкта управління на об'єкт управління для досягнення поставлених цілей. Наукову основу управління становлять його принципи й методи. До принципів відносять науковість, безперервність, гнучкість, цілеспрямованість тощо.

На практиці управління здійснюється за допомогою методів. До них відносять економічні, психологічні й адміністративні (розпорядницькі) методи. Всі методи перебувають у тісному взаємозв'язку й взаємозалежності. З переходом економіки України на світові ринки і посиленням самостійності суттєво зросла роль економічних методів управління. Вони дають суб'єкту управління можливість самостійного вибору в межах наявних обмежень форм і способів дій, мобілізуючи в такий спосіб творчу активність і заповзятливість. Роль системи управління має прямолінійну залежність з кінцевими результатами діяльності підприємства, а тому повинна використовувати всі можливості для досягнення позитивних результатів. На наш погляд, саме системний комплексний аналіз здатний вирішувати поставлені завдання та вказувати на прийняття правильних управлінських рішень для власного добробуту та подальшого процвітання підприємства.

Як відомо, комплексний економічний аналіз є системним дослідженням господарської діяльності в тісному взаємозв'язку та обумовленості всіх його розділів, дає достовірну оцінку та дає змогу виявляти наявні резерви з урахуванням кінцевих результатів. Тобто слід відзначити, що комплексний економічний аналіз вивчає всі аспекти інформаційної моделі виробничо-фінансової діяльності підприємств. Саме інформаційна модель повинна забезпечувати для економічного аналізу формування нормативних, облікових та звітних показників за зміною організаційно-технічного рівня виробництва;

використанням засобів та предметів праці; використанням трудових ресурсів; формуванням обсягів виробництва і реалізації продукції; виробничими затратами і собівартістю продукції; прибутком та рентабельністю продукції; відтворенням та обігом виробничих фондів; рентабельністю виробництва; фінансовим станом; економічним стимулюванням тощо.

Наступним кроком комплексного підходу є необхідність порівняння діяльності власного підприємства з діяльністю підприємств-аналогів. Це дає можливість виявити рівень конкурентоспроможності та скоригувати роботу власного підприємства. Саме в умовах інтеграції на Європейські ринки цей підхід стає особливо актуальним.

В сучасних же умовах необхідним стає використання саме методів системного аналізу, тобто тих проблем, які базуються на дослідженні як сукупності взаємопов'язаних елементів.

За використання комплексного підходу особливу увагу слід приділити саме системі показників та методів, за допомогою яких проводитиметься аналіз та вирішуватиметься та чи інша поставлена проблема. Адже головна проблема сучасного підприємства полягає у високій собівартості та низьких доходах, а інколи навіть у збитковості діяльності.

Отже, на наш погляд, головну увагу слід приділити саме побудові системи показників, які будуть мати ланцюговий зв'язок між собою і в результаті зможуть без зайвих затрат переорієнтувати підприємство на стандарти світових ринків та забезпечити довгострокову конкурентоспроможність власного бізнесу. За вдалої побудови системи показників можна виявити невикористані резерви та задіяти їх в роботу без зайвих затрат, а, навпаки, з певною вигодою. Це дасть можливість підприємствам поступово здійснити перебудову всієї системи планомірно та без значних фінансових втрат.

В ході проведення комплексного аналізу особливу увагу необхідно приділити якості та достовірності інформації. В сучасному потоці інформації, особливо інформації в мережі Інтернет, досить часто можна знайти неперевірену інформацію, використання якої може завести управління по хибному шляху.

Для якісної і компетентної роботи необхідна певна, кількісно обмежена інформація. Якщо ця умова не виконується, то виникає двостороння аномальна ситуація:

1) недостатня кількість інформації не дає змоги одержати повне уявлення про стан досліджуваного об'єкта, що приводить до інтуїтивного доповнення можливими варіантами;

2) надмірна кількість інформації знижує швидкість її оброблення й осмислення, вимагає додаткових витрат часу і праці; у такій ситуації не завжди вдається виявити головне, вловити закономірності і тенденції, встановити причинно-наслідкові зв'язки, в результаті чого можуть бути зроблені неправильні висновки, неправильно намічені шляхи подолання кризових ситуацій. В усіх випадках інформаційні потоки повинні бути якісними, достовірними, обґрунтованими та правдивими. Оцінюється якість інформації за чотирма ознаками: повнота за умови мінімального обсягу, що забезпечує її репрезентативність; достовірність, своєчасність надходження; можливість практичного застосування для потреб аналізу.

Вимоги сучасного менеджменту потребують відповідних знань стосовно нагромадження та зберігання інформації. Відібрані для використання джерела інформації підлягають формальній і суттєвій перевіркам. За формальної перевірки встановлюється правильність оформлення документів та звітів з точки зору наявності і заповнення реквізитів, таблиць, підписів, здійснюється перевірка сум, підсумків та погодженості взаємозалежних величин. За суттєвої перевірки

встановлюють доброякісність матеріалів, їх відповідність об'єктивній дійсності, можливість використання для аналізу. Тут потрібні знання теорії і практики економічного аналізу, належна кваліфікація і досвід роботи, а також знання технологічних дисциплін.

Особливу увагу під час створення інформаційних потоків необхідно приділити формуванню локальної мережі, в якій буде накопичуватися та зберігатися інформація. Сучасна локальна мережа повинна задовольняти вимоги системності, комплексності, оперативності, точності, прогресивності й динамічності. Тільки на основі виконання цих вимог забезпечуються об'єктивне пізнання стану досліджуваного об'єкта і тенденцій його розвитку, систематичне і цілеспрямоване підвищення ефективності господарської діяльності за результатами аналізу.

Під час створення інформаційної бази локальної мережі, по-перше, полегшується її використання; по-друге, чітко показується дійсна потреба у вихідних даних; по-третє, виключається дублювання, полегшується групування їх у блоки для одночасної обробки. Крім того, формалізація задач стає безпосередньо об'єктом математичного моделювання, що суттєво підвищує глибину досліджень. По-перше, для організації роботи в створенні та використанні інформаційних потоків характерно збереження цілості роботи за умови децентралізованої обробки інформації. Завдяки цьому досягаються об'єктивність роботи та її достовірність. В умовах децентралізованої обробки інформації цілість роботи не руйнується, не порушується єдність мети та завдань з точки зору його системних властивостей. По-друге, поставлені завдання вирішуються безпосередньо самим користувачем на своєму робочому місці. Спеціаліст веде особистий контроль над всіма стадіями процесу оброблення інформації, має можливість оцінити отримані результати, грамотно використати

їх для обґрунтування прийнятих рішень, задоволення різноманітних інформаційних потреб системи.

Створення інформаційних потоків системного аналізу вирішує такі завдання:

- зниження трудомісткості та вартості процесу;
- скорочення строків обробки даних, підвищення їх якості та достовірності;
- забезпечення встановлених строків виконання окремих операцій;
- підвищення гнучкості та оперативності керування процесом;
- удосконалення організації праці користувачів.

Це дасть змогу:

- своєчасно передавати та отримувати необхідну інформацію, що дасть можливість своєчасно виявляти та реагувати на зміну попиту і пропозиції на ринку;
- мати доступ до будь-яких джерел інформації, що найбільш ефективно буде працювати в сфері перерозподілу продукції;
- створити необхідну певну, кількісно обмежену інформацію, яка буде використовуватись для проведення аналізу.

На наш погляд, наведена методика створення інформаційних потоків дасть змогу сконцентрувати певні можливості на зростання економічної ефективності роботи управлінців, що буде сприяти підвищенню якості управління та забезпечувати безперебійну роботу перерозподілу матеріальних потоків.

На основі інформаційної моделі складається загальна блок-схема комплексного аналізу, в якій відображуються всі етапи виробництва та реалізації продукції. Спочатку дається попередня характеристика діяльності підприємства за системою найважливіших показників, потім детально аналізуються фактори та

причини, а завершальний етап складається з перевірки формування та використання всіх ресурсів, комплексного оцінювання діяльності підприємства; головним є те, що є всі можливості спрогнозувати подальший розвиток підприємства.

Розділ 2

Відповідно до першого розділу систематизуємо завдання дослідження в таблиці 1.2.

Таблиця 1.2 – 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?	<i>Джерело:</i> парсинг офіційних веб-сайтів компанії; <i>Метод:</i> побудова розподілу, розрахунок основних вибірових характеристик, Інтервальне оцінювання	Кожна компанія буде мати свій розподіл ціни на ноутбуки і візуально відображена у презентації та описані характеристики розподілу
2	Огляд бізнес показників ТМ Comfy	Яка існує тенденція офлайн трафіку в роздрібні магазини Comfy, чи існує сезонність трафіку	<i>Джерело:</i> внутрішня інформація компанії <i>Метод:</i> декомпозиція часового ряду офлайн трафіку	Графік загальної динаміки трафіку, окремий графік сезонності та тренду
		Чи залежить кількість трафіку від регіону	<i>Джерело:</i> внутрішня інформація компанії <i>Метод:</i> однофакторний дисперсійний аналіз	Стовбчикова діаграма кількості трафіку між найбільшими містами України, розподіли трафіку по містам
3	Аналіз впливу медіа-активності на офлайн трафік	Яка існує залежність між офлайн трафіком та факторами які можуть на нього впливати?	<i>Джерело:</i> інформація зібрана протягом дослідження, внутрішні дані <i>Метод:</i> 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 побудуємо гарфік, на якому буде зображено потенційна кількість трафіку якщо б не було карантинних обмежень та кількість яка була втрачена чемрез пандемію
		Який існує вплив не медійних факторів на трафік?	<i>Джерело:</i> інформація зібрана протягом дослідження, внутрішні дані <i>Метод:</i> лінійна регресія та її модифікації	
		Який існує вплив медійних факторів на трафік?	<i>Джерело:</i> інформація зібрана протягом дослідження, внутрішні дані <i>Метод:</i> лінійна регресія та її модифікації	
		Як вплинув карантин на офлайн трафік?	<i>Джерело:</i> інформація зібрана протягом дослідження, внутрішні дані <i>Метод:</i> лінійна регресія та її модифікації	
4	Прогноз офлайн	Який можна зробити прогноз офлайн трафіку на 2022 рік?	<i>Джерело:</i> інформація зібрана протягом дослідження, внутрішні дані	Графік динаміки офлайн трафіку на 2022 рік

	трафіку на 2022 рік		Метод: ансамблеві моделі	
--	---------------------	--	--------------------------	--

З таблиці 1.2 виділили завдання дослідження, пошукові питання, методи збору інформації та формат в якому інформація буде представлена на виході дослідження.

Основні теоретичні положення використаних методів:

а). Розрахунок основних вибірових характеристик:

середнє значення, біля якого групуються можливі значення випадкової величини; - ступінь розкиду цих значень щодо середнього; - асиметрію (або «скошеність») густини розподілу; - ексцес щільності розподілу [8].

б). Декомпозиція часового ряду, однофакторний дисперсійний аналіз:

Декомпозиція тимчасового ряду - це виділення тенденції, циклічних і випадкових коливань. Проведення декомпозиції дозволяє зрозуміти структуру ряду, побудувати його модель і екстраполювати рівні ряду на короткострокову перспективу [9].

Основною метою дисперсійного аналізу, фундаментальна концепція якого була запропонована Фішером у 1920 р., є дослідження значущості відмінності між середніми декількох груп даних або змінних [10]. У дисперсійному аналізі перевірка статистичної значущості відмінності між середніми декількох груп здійснюється на основі вибірових дисперсій. Ця перевірка проводиться за допомогою розбиття загальної дисперсії (варіації) на частини, одна з яких обумовлена випадковою помилкою (тобто внутрішньогруповою мінливістю), а друга пов'язана з відмінністю середніх значень. Якщо ця відмінність значуща, нульова гіпотеза щодо існування відмінності між середніми значеннями

відкидається на певному рівні значущості. Графічна інтерпретація сутності дисперсійного аналізу зображена на рисунку 2.1.

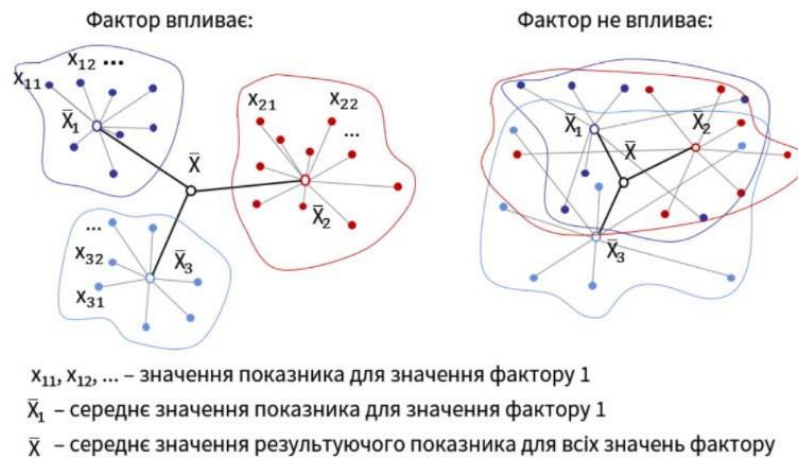


Рисунок 2.1 – Дисперсійний аналіз

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 об'єктів, що досліджуються за допомогою лінійної регресії та конкретних цілей дослідження можуть використовуватися різні методи оцінки невідомих параметрів. Найпопулярнішим є звичайний метод найменших квадратів. Він приймає за оцінку параметра значення, що мінімізують суму квадратів залишків по всіх спостереженнях:

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

У випадку класичної моделі лінійної регресії оцінка методу найменших квадратів є незміщеною, змістовною і найкращою лінійною незміщеною оцінкою (детальніше про ці статистичні властивості у статті метод найменших квадратів).

Серед інших методів оцінювання можна виділити наступні:

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

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

Цей метод є найкращим в сенсі максимальної вірогідності у випадку коли відхилення мають розподіл Лапласа. Метод найменших модулів є значно менш чутливим до викидів значень, ніж метод найменших квадратів, проте він може мати більш ніж один розв'язок і для нього не існує простої формули визначення оцінки.

- Метод максимальної вірогідності. Використовується коли відомі всі розподіли відхилень для всіх спостережень. При класичній і узагальненій моделях лінійної регресії з умовою нормальності відхилень приводить до того ж результату, що і метод найменших квадратів і 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 розподіли ціни ми тим самим порівнюємо пропозиції, які надають конкуренти, таким чином ми можемо визначити скільки ноутбуків пропонує на продаж кожен магазин, середня ціна на ноутбук в магазині, найчастішу ціну... В нашому випадку ми будемо екстраполювати отримані результати на всі товари в магазинах, тобто якщо ми визначимо, що середня ціна в магазині $A > B$, то магазин A є більш дорогим магазином (*Такий підхід є не*

зовсім доцільним у реальному житті, але у зв'язку з тим, що процес збору даних є достатньо довгим, для розрахункової роботи цього буде цілком достатньо).

Інформація у відповідь: Кожна компанія буде мати свій розподіл ціни на ноутбуки і візуально відображена у презентації та описані характеристики 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/ru/>
- 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: ми будемо в буквальному сенсі витягати дані з офіційних сайтів компаній. В області «Data science» подібну практику прийнято називати терміном веб-скрепінг (англ. Web Scraping).

Для проведення цієї процедури знадобляться знання html та python. Відповідно до джерела [12] HTML - це мова гіпертекстової розмітки, що виступає як будівельний матеріал сторінки. З його допомогою створюється основний контент - текстова частина, зображення, різні блоки та інше. Все це заповнюється за допомогою тегів спеціальних команд для браузера, які вводяться користувачем у спеціальному файлі з розширенням .html. Python - інтерпретована об'єктно-орієнтована мова програмування високого рівня зі строгою динамічною типізацією [13], проста у використанні, та водночас повноцінна мова програмування. Цю мову можна використовувати на універсальних і портативних програмах і використовувати його 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: для того щоб збільшити трафік, необхідно розміти його структуру, під цим я розумію виявлення загально тренду та сезонності. Відповідно до розуміння сезонності з'являється можливість приймати рішення щодо інструментів та підходів 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 3

Приступимо до виконання першого завдання дослідження. Маємо наступні 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/ru/>
- 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/>

Нам потрібно зібрати назву моделі ноутбуку HP та відповідну ціну по кожному магазину. Для цього будемо використовувати мову програмування Python.

В першу чергу імпортуємо необхідні бібліотеки (див. рис. 3.1):

```
# 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
```

Рисунок 3.1 – імпорт необхідних бібліотек

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

```
# Ініціалізуємо списки в які будемо складати назви та ціни ноутбуків
finaltotal = []
finalprice = []
headers = {'User-Agent': 'Mozilla/xxx (xxxxxxxxxxxxxxxxxxx) xxxxx/xxxxxxxxx
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")

```

Отримаємо наступну відповідь (див. рис. 3.2), 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

```

Рисунок 3.2 – Результат парсингу сайтів

Після того, як проведемо подібну процедуру над всіма сайтами, отримаємо загальну таблицю (див. рис. 3.3) 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 x 3 columns

Рисунок 3.3 – Набір даних, який отримали в результаті парсингу сайтів

Почнемо з того, що дізнаємося який обсяг ноутбуків пропонує кожен з магазинів (див. рис. 3.4, 3.5).

```
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)
```

Рисунок 3.4 – Розрахунок обсягу ноутбуків HP

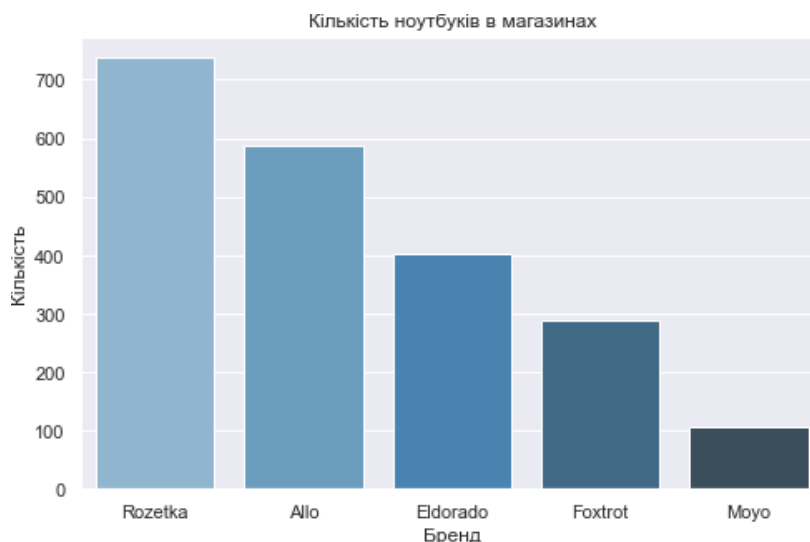


Рисунок 3.5 – Обсяг ноутбуків HP

Отже, маємо, що магазин Rozetka має найбільшу кількість ноутбуків на продаж. Пам'ятаємо, що отримані висновки ми будемо екстраполювати на всі товарні позиції по магазинам. Тому, робимо висновок, що найбільший асортимент товарів пропонує магазин Rozetka, а найменший магазин Мoyo.

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

```
df[df["Price"]== df["Price"].max()]
df[df["Price"]== df["Price"].min()]

# Краще так:
df[(df["Price"] == df["Price"].max()) | (df["Price"]== df["Price"].min())]
```

Отримаємо наступну відповідь (див. рис. 3.4).

	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

Рисунок 3.4 – Мінімальна та максимальна ціна ноутбука

Бачимо, що магазин Allo має ноутбук по найменшій ціні на ринку та найбільшій ціні, що свідчить про найбільший розмах. Це твердження можемо проілюструвати за допомогою графіку який називається 'boxplot' (див. рис. 3.5).

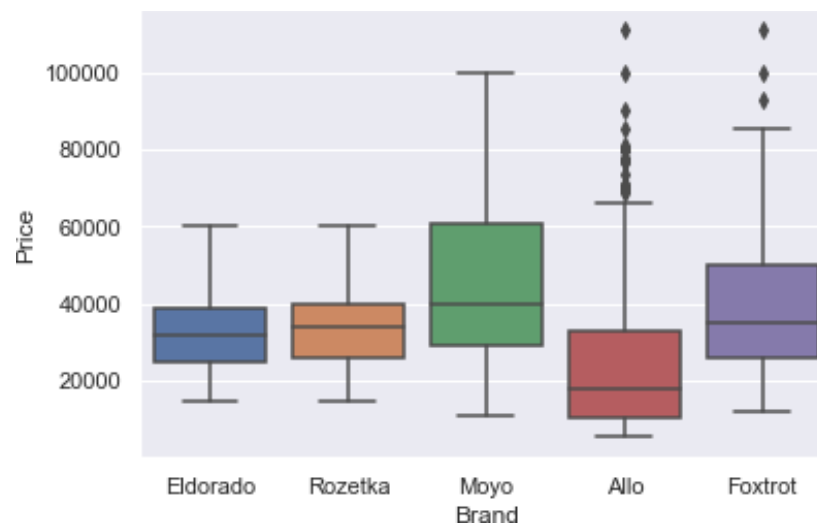


Рисунок 3.5 – Boxplot

Також на графіку можна побачити лінію, яка є медіаною. Розрахуємо показники центральної міри тенденції (див. рис. 3.6)

```
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"]))
```

Рисунок 3.6 – Розрахунок показників мір ц.т

Отриманий результат буде виглядати наступним чином (див. рис. 3.7)

```
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)
```

Рисунок 3.7 – Результати розрахунку показників мір ц.т

Зобразимо отриманий результат графічно (див. рис. 3.8, 3.9)

```
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:])
```

Рисунок 3.7 – Розрахунок міри центральної тенденції

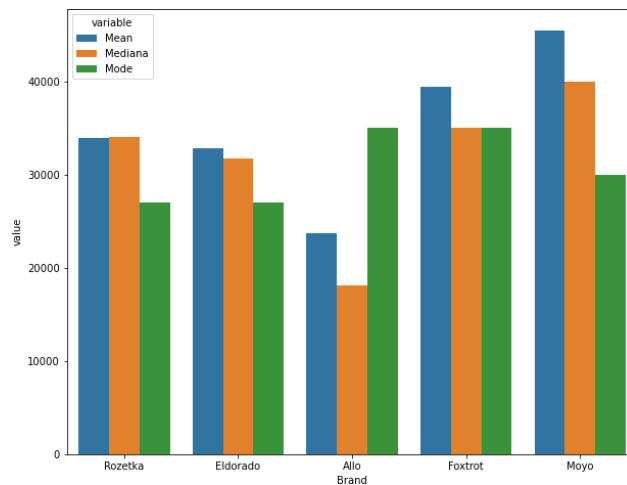


Рисунок 3.8 – Міри центральної тенденції

Отримали дуже інформативний графік. Він показує що з точки зору медіани та середнього значення, в магазині Моуо ціни дорожче ніж в інших магазинах. Найдешевшим є магазин Allo.

Ми вже розрахували показники міри центральної тенденції, тепер настав час, щоб побудувати гістограму відносних накопичених частот.

Пропоную розглянути процедуру побудови емпіричного розподілу більш детально. Для цього будемо використовувати Excel.

1) Будуємо варіаційний ряд (див. рис. 3.9);

A	B	C	D	E	F	G	H
	Дані, ціни на ноутбуки 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		Максимальна	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				

Рисунок 3.9 – Варіаційний ряд

2) Рахуємо точкові оцінки повної вибірки;

3) Будуємо групувану вибірку (див. рис. 3.10);

I	J	K	L	M	N	O	P	Q	R	S	T
№ інтервалу	Ліва межа інтервалу	Право межа інтервалу	Середня 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%	15.8%		h	4555.50		
3	23555.00	28110.50	25832.75	144	18.9%	34.7%					
4	28110.50	32666.00	30388.25	50	6.8%	41.5%					
5	32666.00	37221.50	34943.75	151	20.5%	62.0%					
6	37221.50	41777.00	39499.25	149	20.2%	82.2%					
7	41777.00	46332.50	44054.75	62	8.4%	90.6%					
8	46332.50	50888.00	48610.25	14	1.9%	92.5%					
9	50888.00	55443.50	53165.75	15	2.0%	94.5%					
10	55443.50	59999.00	57721.25	36	4.9%	100.0%					

Рисунок 3.10 – Групована вибірка

- 3) Рахуємо оцінки групованої вибірки;
- 4) Будуємо гістограму частот (див. рис. 3.11);

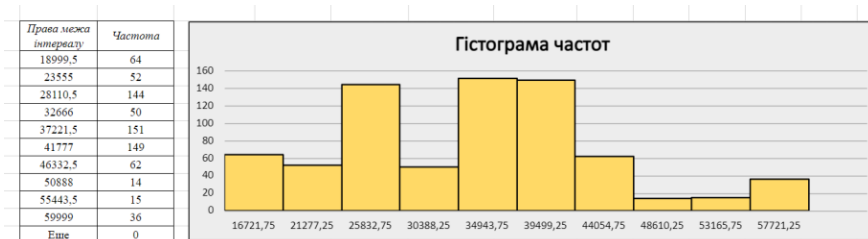


Рисунок 3.10 – Гістограма частот

- 5) Будуємо кумулятивну криву (див. рис. 3.12);



Рисунок 3.12 – Кумулятивна крива

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

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

Коли ми розраховуємо оцінки групованої вибірки, ми не знаємо де знаходиться більшість елементів в межах інтервалу. Однак істинне значення оцінок в генеральній сукупності можуть бути близькими і до повної вибірки і до групованої.

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 скошеність в ліву сторону. За рахунок додатної асиметрії мода розподілу менша за середнє значення.

3) Якою є гіпотеза про розподіл генеральної сукупності?

Поки що в вибірці немає достатньо кількості елементів, але скоріше за все, розподіл буде підпорядковуватися нормальному розподілу зі збільшенням даних.

Ми можемо перевірити розподіл на нормальність за допомогою критерію згоди Шапіро-Уилка (див. рис. 3.13, 3.14)

```
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"]))
```

Рисунок 3.13 – Перевірка на нормальність (1)

Результат покаже, що p value дорівнює 0,17. Це більше ніж 0,05 тому ми не можемо відхилити основну гіпотезу, яка свідчить що розподіл є нормальним.

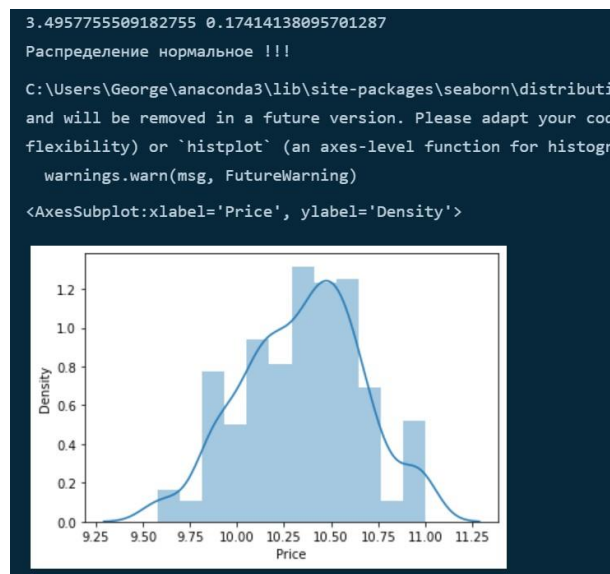


Рисунок 3.14 – Перевірка на нормальність (2)

Ну і на останок порівняємо вибіркові характеристики магазинів між собою

```
def comp_param(x):
    mean = np.mean(x)
    mediana = np.median(x)
    moda = mode(x)
    std = np.std(x)
```

```

var = std ** 2
min1 = min(x)
max1 = max(x)
range1 = max1 - min1
skewness1 = skew(x)
kurtosis1 = kurtosis(x)
return mean, mediana, moda, std, var, min1, max1, range1, skewness1, kurtosis1
shops = ["Rozetka", "Eldorado", "Allo", "Foxtrot", "Moyo"]
mean = [comp_param(roz["Price"])[0], comp_param(el["Price"])[0],
comp_param(allo["Price"])[0], comp_param(fox["Price"])[0],
comp_param(moyo["Price"])[0]]

mediana = [comp_param(roz["Price"])[1], comp_param(el["Price"])[1],
comp_param(allo["Price"])[1], comp_param(fox["Price"])[1],
comp_param(moyo["Price"])[1]]

moda = [comp_param(roz["Price"])[2], comp_param(el["Price"])[2],
comp_param(allo["Price"])[2], comp_param(fox["Price"])[2],
comp_param(moyo["Price"])[2]]

std = [comp_param(roz["Price"])[3], comp_param(el["Price"])[3],
comp_param(allo["Price"])[3], comp_param(fox["Price"])[3],
comp_param(moyo["Price"])[3]]

var = [comp_param(roz["Price"])[4], comp_param(el["Price"])[4],
comp_param(allo["Price"])[4], comp_param(fox["Price"])[4],
comp_param(moyo["Price"])[4]]

min1 = [comp_param(roz["Price"])[5], comp_param(el["Price"])[5],
comp_param(allo["Price"])[5], comp_param(fox["Price"])[5],
comp_param(moyo["Price"])[5]]

max1 = [comp_param(roz["Price"])[6], comp_param(el["Price"])[6],
comp_param(allo["Price"])[6], comp_param(fox["Price"])[6],
comp_param(moyo["Price"])[6]]

range1 = [comp_param(roz["Price"])[7], comp_param(el["Price"])[7],
comp_param(allo["Price"])[7], comp_param(fox["Price"])[7],
comp_param(moyo["Price"])[7]]

skewn = [comp_param(roz["Price"])[8], comp_param(el["Price"])[8],
comp_param(allo["Price"])[8], comp_param(fox["Price"])[8],
comp_param(moyo["Price"])[8]]

kurtos = [comp_param(roz["Price"])[9], comp_param(el["Price"])[9],
comp_param(allo["Price"])[9], comp_param(fox["Price"])[8],
comp_param(moyo["Price"])[9]]

```

```
df2 = pd.DataFrame({"Brand": shops, "Mean":mean, "Mediana":mediana, "Mode":moda,
"std": std, "var": var, "min": min1, "max": max1, "range":range1, "skew":skewn,
"kurt":kurtos})
df3 = df2.melt(id_vars=["Brand"], value_vars=df2.columns[1:])
df3.to_excel("C:/Users/xxxxx/xxxx/PP стат/xxx.xlsx")
```

Отримаємо наступну таблицю 3.1

Таблиця 3.1 - Вибіркові характеристики

	Allo	Eldorado	Foxtrot	Moyo	Rozetka
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 3.1 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: $M > F > A > E > R$.

Перейдемо к 2 завданню дослідження, огляду бізнес показників компанії Comfy. **Важливо: дані, які будуть використовуватися є штучними і слугують лише для прикладу статистичних методів та методології обробки даних.**

Почнемо з огляду офлайн трафіку в магазини Comfy, та визначимо існуючу тенденцію та сезонність. Для початку треба зібрати дані. Нехай в компанії є

відділ, який займається збором даних, а в нас є доступ до бази даних SQL, в якій є дані з трафіком. За допомогою наступного коду (див. рис. 3.14) ми витягнемо цей набір даних в Python.

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

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

Рисунок 3.15 – Дані з трафіком (1)

Отримаємо наступний набір даних (див. рис. 3.15)

	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 × 4 columns

Рисунок 3.15 – 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. 3.16).



Рисунок 3.16 – Динаміка трафіку в місті Києві

З графіку можемо побачити періодичні сплески, які виникають приблизно на 46 неділі та 51-52 кожного року і свідчать про наявність чітко виразної сезонності. Можемо здогадатися, що саме в цих неділях святкують Чорну П'ятницю та Новорічні свята. Отже робимо гіпотезу, що ці свята впливають на динаміку трафіку, та постараємося визначити який приріст в трафік робить наявність подібних 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. 3.15).

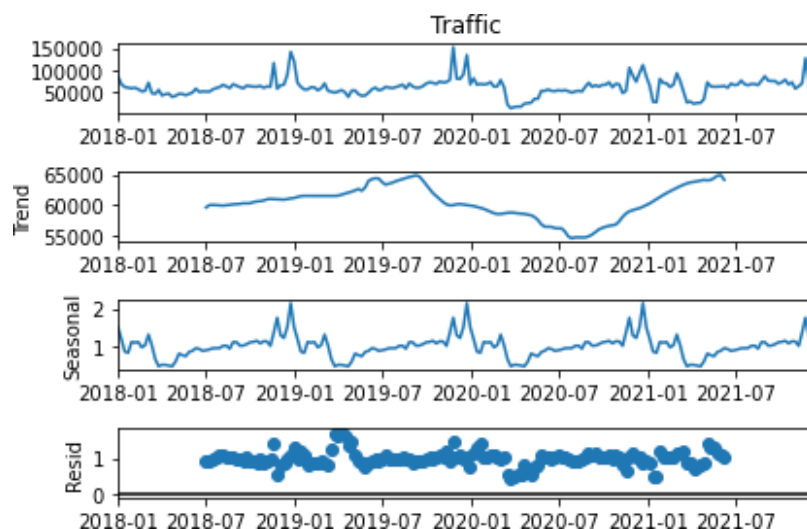


Рисунок 3.15 – Декомпозиція часового ряду

Висновок, маємо чітку сезонність, яка знадобиться нам при побудові моделі, також бачимо, що з 2018-2020 рік був позитивний тренд. З початку 2020 року по серпень місяць спостерігаємо спадаючий тренд, що був спричинений пандемією, а так як ми розглядаємо саме офлайн трафік, то логічним буде твердження, що через карантинні обмеження магазини просто були закриті.

Далі перевіримо гіпотезу, чи залежить кількість трафіку від регіону. Відповідно, основна гіпотеза – трафік не залежить від регіону, альтернативна – в залежності від регіону трафік відрізняється.

Спочатку зобразимо стовпчикову діаграму, та розподіли по кількості трафіку в найбільших містах (Київ, Дніпро, Харків, Одеса, Львів)(див. рис. 3.17).

```
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.barplot(x = a.index, y = a, order=a.sort_values(ascending = False).index)
```

Рисунок 3.17 – Кількість трафіку в найбільших містах (1)

Отримаємо наступний графік (див. рис. 3.16), (див. рис. 3.17):

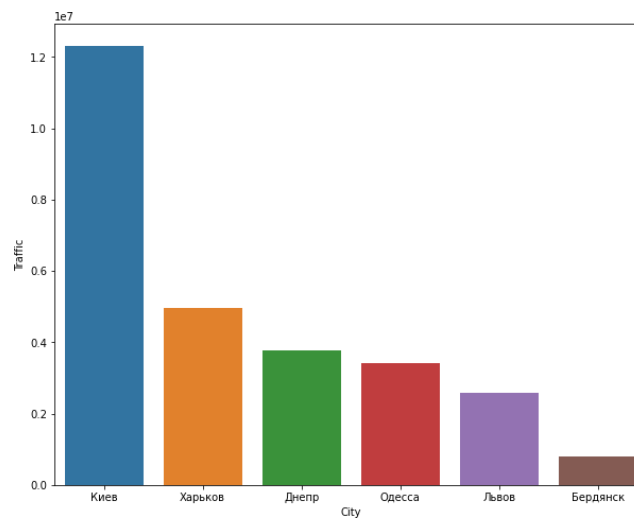


Рисунок 3.16 – Кількість трафіку в найбільших містах (2)

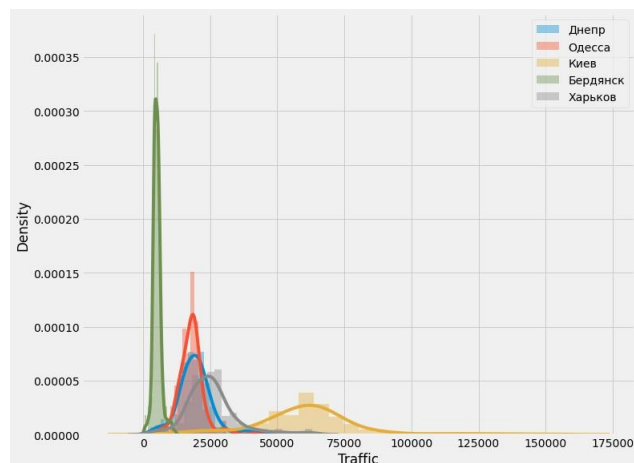


Рисунок 3.17 – Кількість трафіку в найбільших містах

Вже з графіку можна визначити суттєву різницю у трафіку, але ми як справжні математики-аналітики не довіряємо очам і повинні це перевірити математично за допомогою однофакторного дисперсійного аналізу.

```
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

Звертаємо увагу на статистику F та значення p value та робимо висновок, що в залежності від регіону трафік відрізняється, приймаємо альтернативну гіпотезу з майже 100% впевненості.

Переходимо до вирішення 3 завдання дослідження, аналізу впливу медіа-активності на офлайн трафік у магазини Comfy.

Вирішувати це завдання будемо за рахунок побудови моделі, яка буде апроксимувати реальні значення. Апроксимація - спосіб вибрати з сімейства "простих" функцій наближення для "складної" функції на відрізку, при цьому помилка не повинна перевищувати певної межі. Апроксимацію використовують, коли потрібно отримати функцію, схожу на цю, але більш зручну для обчислень та маніпуляцій (диференціювання, інтегрування тощо)[15].

Завантажимо вже сформований датасет, який необхідно лише почистити. Так як даних (факторів) для побудови моделі буде багато, це не доцільно занурюватися в дані та розписувати кожен крок формування датасету, так як процес цей механічний і не вартий уваги.

```
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()
```

Рисунок 3.18 – Набір даних для побудови моделі

Наш набір даних для побудови моделі (див. рис. 3.18) має наступний вигляд

Year	Week	Traffic	Eldorado TV	Competitors TV	Eldorado digital	Competitors digital	Competitor's radio	Eldorado radio	Eldorado ООН	...
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	...

Рисунок 3.18 – Набір даних для побудови моделі

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

Написали функцію (див. рис. 3.19), яка шукає пропущені значення та у разі знаходження вертає назву змінної та кількість пропущених елементів користувачу і змінює їх на середнє значення по змінній.

```
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)
```

Рисунок 3.19 – 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)
```

Рисунок 3.20 – функцію, яка перевіряє наявність неметричних даних

Написали функцію, яка шукає неметричні данні та вертає користувачу назву змінної. Отримали, що змінна «Номер місяця» має неметричні данні. Давайте замінімо їх за допомогою методу, який називається “One hot encoding”, проше кажучи він кодує дані у бінарний формат (див. рис. 3.19).

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

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

Рисунок 3.19 – Кодування неметричних даних

Тобто, ми з одного фактора зробили 12, тим самим збільшили розмірність нашого набору даних.

Додомо до моделі лагові зміні для медіа факторів, тому що як ми знаємо вплив реклами може бути не зразу, а через деякий час.

```
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 3.20 – Лагові зміні для медіа факторів

Так як дані мають різну розмірність, їх треба стандартизувати, аби модель могла коректно працювати (див. рис. 3.21)

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

Рисунок 3.21 – Стандартизація моделі

Далі, необхідно перевірити цільову змінну на нормальність розподілу (одна з умов МНК) (див. рис. 3.20).

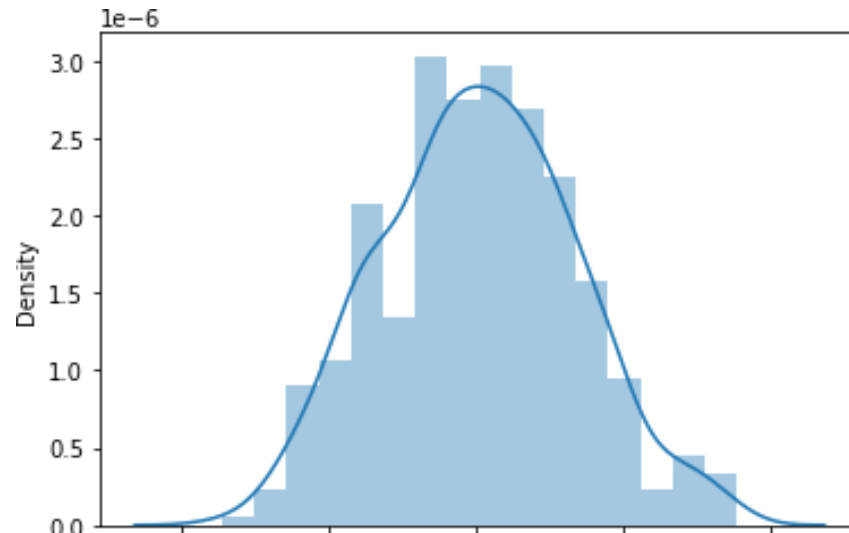


Рисунок 3.20 – Розподіл цільової змінної (трафіку)

Графічно ніби як нормальний розподіл, перевіримо за допомогою тесту (див. рис. 3.21)

```
statistics, p_value = stats.normaltest(y_norm)
print(statistics, p_value)

#Відповідь
#0.4476 0.7994
```

Рисунок 3.21 – 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. 3.22).

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

Рисунок 3.22 – Побудова регресійної моделі

Отримаємо наступні результати якості моделі (див. рис. 3.21).

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
K-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
...						

Рисунок 3.21 – Модель лінійної регресії (методом МНК)

Маємо наступний розподіл залишків (див. рис. 3.22)

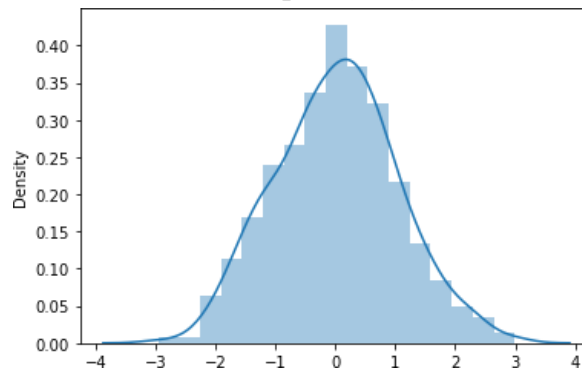


Рисунок 3.22 – Розподіл залишків

Необхідні умови мнк виконуються, залишки мають нормальний розподіл. З побудованої моделі можемо відібрати ознаки (фактори) які дійсно впливають на трафік. Для цього треба перевірити показник значимості коефіцієнта, в нашому випадку фактор є значимим, якщо $p \text{ value} < 0.05$.

Далі, розрахуємо вплив факторів. Спочатку виділимо основні фактори:

- Органічний ріст трафіку (показує кількість трафіку без участі зовнішніх факторів)

- Сезонність (показує вплив празників)

- Медіа-активність (показує скільки трафіку прийшло за допомогою реклами)

- Вплив конкурентів (показує скільки потенційного трафіку забрали конкуренти)

Відповідно до нашої моделі, маємо наступний вклад факторів (див. рис. 3.23)

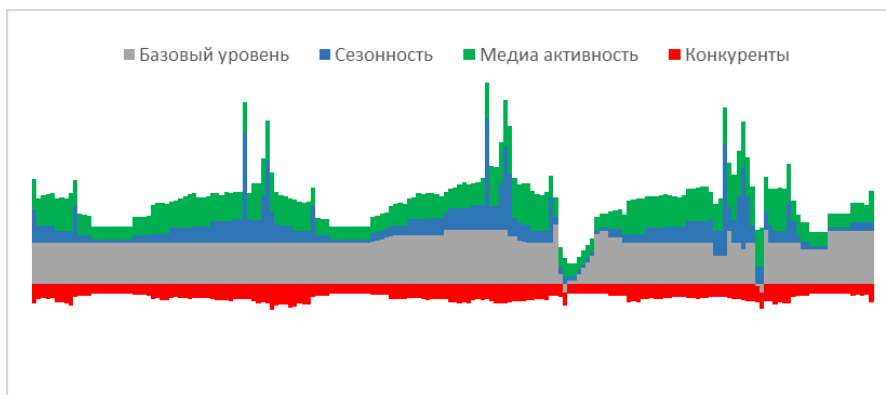


Рисунок 3.23 – Вплив факторів

Проаналізуємо вплив кожного з факторів детальніше. Базовий рівень та органічний розширення мережі займає 47% трафіку в офлайн магазини. Базовий рівень включає:

- Розмір торгової мережі та її розширення;
- Обсяг клієнтської бази та її розширення;
- Робота з клієнтською базою шляхом розсилок;
- Знання ТМ Comfy;
- Імідж та сприйняття бренду Comfy;
- Попередній досвід взаємодії з Comfy (якість обслуговування, сервіс, якість продукції);
- Інші фактори, які неможливо кількісно виміряти та оцифрувати.

Сезонність включає сезонні коливання та вплив празників і складають 29% трафіку. Сезонні коливання мають певну закономірність:

- 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%.

Медіа-активність Comfy включає наступні канали:

- Тв; Діджитал; Зовнішня реклама; Радіо.

Відповідно до побудованої моделі, за допомогою медіа-активності Comfy залучає 36% свого трафіку.

Спираючись на побудовану модель дамо відповідь на пошукове питання, яке стосується впливу карантинних обмежень. Відповідно маємо, що карантинні обмеження призвели до значного скорочення трафіку з 17 березня, але з початку травня активність споживачів поступово відновлюється, частково компенсуючи втрати березня-квітня. За моделлю, під час карантинних обмежень було втрачено близько 1 000 000 осіб потенційного трафіку (див. рис. 3.24).

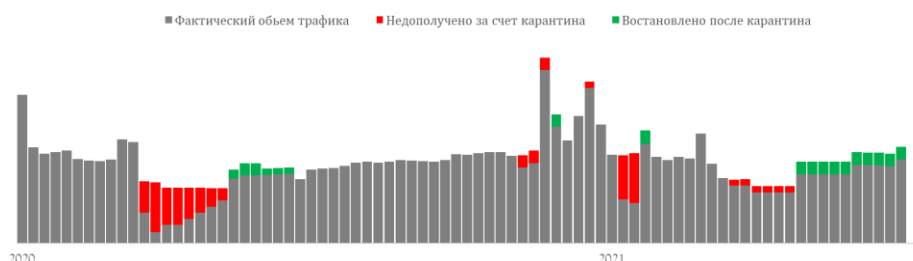


Рисунок 3.24 – Вплив карантинних обмежень

Перейдемо до вирішення останнього завдання дослідження, зробимо прогноз на 2022 рік. В минулий раз ми використовували метод лінійної регресії, для вирішення апроксимації цільової змінної. Як вже було сказано, цей метод передбачає лінійний зв'язок між факторами та трафіком, за допомогою якого ми могли визначити вплив кожного фактора. Ми зробили припущення (наявність лінійного зв'язку) та отримали досить високу оцінку якості моделі ($R^2 = 89\%$), але в дійсності скоріше за все взаємозв'язок між цільовою змінною та факторами які її описують не є строго лінійною, тому потрібно використовувати більш складні методи. Тому для цієї задачі регресії

(прогнозу дійсного числа) ми будемо використовувати метод XGBoost, який є досить універсальний і популярний в області 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"Diferense 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 3.25 – XGBoost

Ми побудували модель XGBoost та порівняли квадрати відхилень. Для того щоб побудувати модель та оцінити її якісь, спочатку данні ділять на тренувальну вибірку та тестову, це робиться для того, щоб перевірити як модель прогнозує дані, які вона ще не

бачила. Можемо побачити, що метод XGBoost робить дуже якісний прогноз, краще ніж лінійна регресія. Виведемо графік відхилень на тестовій вибірці (див. рис. 3.25).

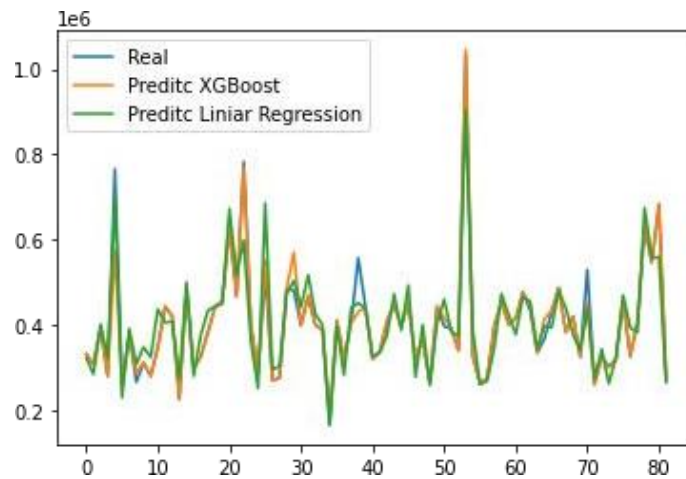


Рисунок 3.25 - Графік відхилень на тестовій вибірці

Виведемо графік відхилень на тренувальній вибірці (див. рис. 3.26).

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

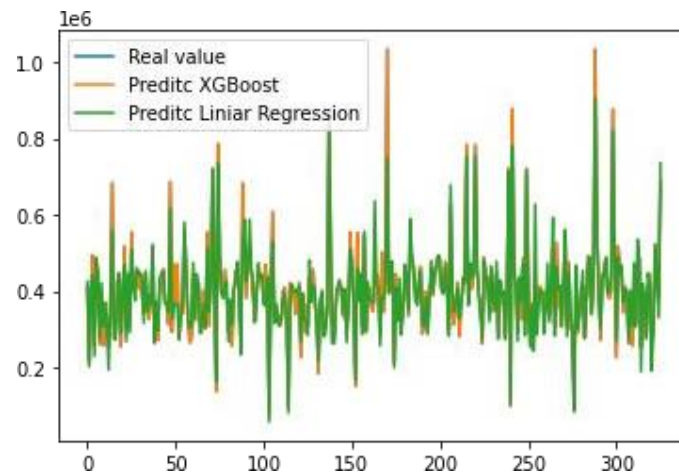


Рисунок 3.26 - Графік відхилень на тренувальній вибірці

З побудованих графіків бачимо, що метод XGBoost дуже добре описує цільову змінну. Тепер припустимо, що ми маємо медіа-план на 2022 рік, тоді за допомогою

побудованої моделі ми можемо побудувати прогноз трафіку. Отримаємо наступний прогноз (див. рис. 3.27).



Рисунок 3.27 – Прогноз трафіку на 2022 рік

Ми розрахували точкові оцінки прогнозу, але ми можемо вдосконалити наш прогноз тим, що додаймо інтервальний діапазон в якому опиниться трафік з ймовірністю 95%, що дасть змогу краще сприймати бізнес ситуацію і стануть у нагоді при необхідності у прийнятті ефективного рішення.

ВИСНОВКИ

Таким чином, проведені спостереження доводять, що саме комплексний системний аналіз спроможний технічно вирішувати питання, пов'язані з якісним управлінням. Комплексний аналіз дає змогу охопити всі структурні сторони підприємства, а системний підхід дає можливість побудувати таку економічну модель, яка дасть змогу управлінцю знайти та прийняти правильне рішення.

Для ефективного управління підприємством, компанією, організацією системний аналіз має першорядне значення, тому що з його допомогою виявляють проблеми та можливості їх розв'язання. Системний аналіз є сукупністю методів дослідження складних соціально-економічних систем, виявлення проблем, планування і реалізація дій, що спрямовані на їх ліквідацію. Тому будь-яке рішення повинно спиратися на результати системного аналізу. Знання менеджерів змісту етапів процесу вироблення управлінського рішення, методів, що забезпечують якість рішення, що набувається, а кожному етапі з урахуванням ситуаційних характеристик, чинників, що впливають на формування і реалізацію рішення, сприяє підвищенню ефективності управління організації в цілому.

ПЕРЕЛІК ПОСИЛАНЬ

1. Анфилатов В. С. Системный анализ в управлении: учебное пособие / Анфилатов В. С., Емельянов А. Л., Кукушкин А. А. ; под. ред. А. Л. Емельянова. — М.: Финансы и статистика, 2009. — 368 с
2. Блауберг И. В. Становление и сущность системного подхода / И. В. Блауберг. — М.: Мир, 2007. — 361 с
3. Лямец В. И. Системный анализ / В. И. Лямец. — Х. : ХТУРЭ, 2012.— 252 с.
4. Оптнер С. Системный анализ для решения деловых и промышленных проблем / С. Оптнер. — М.: Советское радио, 1989. — 416 с
5. Пономаренко О. І. Системні методи в економіці, менеджменті та бізнесі / О. І. Пономаренко. — К. : Либідь, 2015. — 240 с
6. Черняк Ю. И. Системный анализ в управлении экономикой / Ю. И. Черняк. — М. : Экономика, 1975. — 193 с
7. Янг С. Системное управление организацией / С. Янг; пер. с англ. — М.: Советское радио, 1982. — 456 с
8. Практика медіа-просування URL: http://www.economy.nayka.com.ua/pdf/6_2021/204.pdf
9. Практика медіа-просування URL: http://www.economy.nayka.com.ua/pdf/6_2021/204.pdf
10. Опис діяльності Comfy URL: <http://ir.duan.edu.ua/bitstream/.....pdf>
11. Фактори маркетингового середовища URL: <http://ir.duan.edu.ua/bitstream/.....pdf>
12. Маркетингові комунікації URL: <https://vlada-rykova.com/ua/chto-takoye-mediaplan/>
13. Основи медіа планування URL: <https://studme.com.ua/1566021211480/marketing/mediaplanirovanie.htm>
14. Медіа — активність URL: <https://faqukr.ru/marketing/94221-mediaplanuvannja-ce-mediaplanuvannja-v-reklami.html>
15. Описова статистика URL: <https://uk.wikipedia.org/wiki/.....>
16. Дослідження часових рядів URL: <https://uk.wikipedia.org/wiki/.....>
17. Однофакторний дисперсійний аналіз URL: <https://uk.wikipedia.org/wiki/.....>
18. Теоретичні основи лінійної регресії URL: <https://vseosvita.ua/library/linijna-regresia-43816.html>
19. Html URL: <https://www.parsehub.com/blog/what-is-web-scraping/>
20. Python URL: <https://ru.wikipedia.org/wiki/Python>

21. Апроксимація URL: <https://habr.com/ru/post/514818/>

СИСТЕМНИЙ АНАЛІЗ ДЛЯ ОПТИМІЗАЦІЇ УПРАВЛІННЯ БІГ ДАНИМИ (BIG DATA) В ОРГАНІЗАЦІЯХ

ПАНЬКОВ ЄГОР САДМ-51

Слайд 1

МЕТА/ОБ'ЄКТ/ПРЕДМЕТ ДОСЛІДЖЕННЯ

- **Мета роботи** – розробка та впровадження системного аналізу для оптимізації управління біг даними в організаціях. Ми прагнемо створити фреймворк, який дозволить організаціям ефективно використовувати біг дані для підтримки стратегічних рішень та покращення їхньої конкурентоспроможності.
- **Об'єкт дослідження** – процеси управління біг даними в організаціях.
- **Предмет дослідження** – методи та стратегії системного аналізу, спрямовані на оптимізацію управління біг даними.
- **Короткий зміст роботи:** створено комплексний підхід до системного аналізу управління біг даними та розроблено конкретні стратегії оптимізації, які будуть враховувати поточний стан організації та тенденції розвитку галузі.

Слайд 2

ЗАВДАННЯ ДОСЛІДЖЕННЯ

- Проаналізувати поточний стан управління біг даними в організаціях:
 - Визначити основні виклики, з якими стикаються організації у сфері управління біг даними.
 - Оцінити рівень інтеграції та використання біг даних у різних галузях.
- Визначити основні принципи системного аналізу та їх застосування до управління біг даними:
 - Розглянути концепції системного аналізу та визначити, як вони можуть бути адаптовані для вирішення проблем управління біг даними.
- Розглянути та обґрунтувати методи оптимізації управління біг даними:
 - Дослідити різні стратегії розподіленої обробки даних та масштабованості.
 - Оцінити ефективність використання технологій хмарних обчислень.
- Визначити критерії ефективності та впровадження моделі:
 - Розробити критерії оцінки ефективності впровадження системного аналізу в управління біг даними.
 - Визначити етапи та план впровадження розробленої моделі.

Слайд 3

АКТУАЛЬНІСТЬ ТЕМИ

- Актуальність теми полягає в тому, що організації стають перед викликом ефективного управління цими масштабними обсягами даних. Справжнім викликом є не тільки обробка цих обсягів інформації, але і вміння здобувати з неї цінну інформацію для прийняття стратегічних рішень.

Слайд 4

ВИСНОВКИ

Таким чином, проведені спостереження доводять, що саме комплексний системний аналіз спроможний технічно вирішувати питання, пов'язані з якісним управлінням. Комплексний аналіз дає змогу охопити всі структурні сторони підприємства, а системний підхід дає можливість побудувати таку економічну модель, яка дасть змогу управлінцю знайти та прийняти правильне рішення.

Для ефективного управління підприємством, компанією, організацією системний аналіз має першорядне значення, тому що з його допомогою виявляють проблеми та можливості їх розв'язання. Системний аналіз є сукупністю методів дослідження складних соціально-економічних систем, виявлення проблем, планування і реалізація дій, що спрямовані на їх ліквідацію. Тому будь-яке рішення повинно спиратися на результати системного аналізу. Знання менеджерів змісту етапів процесу вироблення управлінського рішення, методів, що забезпечують якість рішення, що набувається, а кожному етапі з урахуванням ситуаційних характеристик, чинників, що впливають на формування і реалізацію рішення, сприяє підвищенню ефективності управління організації в цілому.