

Київський національний університет імені Тараса Шевченка
Економічний факультет
Кафедра економічної кібернетики

КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

“ОЦІНКА ФАКТОРІВ УСПІХУ СТАРТАПІВ В ІТ-СФЕРІ”

студентки 4 курсу
спеціальності 051 «Економіка»
ОПП «Економічна кібернетика»
денної форми навчання
Сіриченко Владислави Анатоліївни

Науковий керівник:
Кандидат економічних наук, доцент
Шпирко Віктор Васильович

Засвідчую, що в цій роботі немає запозичень із
праць інших авторів без відповідних посилань
Студент _____

Роботу допущено до захисту перед ЕК
рішенням кафедри економічної кібернетики
від 05.06.2024 р., протокол № 15

Завідувач кафедри:
доктор економічних наук, професор
Ляшенко Олена Ігорівна

(підпис)

КИЇВ – 2024

РЕФЕРАТ

Кваліфікаційна робота бакалавра містить: 49 ст., 12 рис., 2 табл., 22 джерела, 2 додатки

Ключові слова: стартапи, інформаційні технології, венчурне фінансування, інновації, бізнес-моделі, зовнішнє середовище, ринкові умови, успіх стартапів, IT-стартапи, фактори успіху.

Об'єкт дослідження: Стартапи в сфері інформаційних технологій та фактори, що впливають на їх успішність.

Мета дослідження: Оцінити фактори, що впливають на успіх стартапів в IT-сфері, та розробити рекомендації для підвищення їхньої ефективності та конкурентоспроможності.

Методи дослідження: Кількісний аналіз даних (статистичний аналіз), теоретичні методи (аналіз літератури, систематизація та узагальнення).

Наукова новизна, теоретична значимість дослідження: Наукова новизна полягає у вперше проведеному комплексному аналізі факторів успішності стартапів у сфері інформаційних технологій, використанні сучасних методів класифікації через великі мовні моделі (LLM), та виявленні нових тенденцій у фінансуванні стартапів в Європі, зокрема значного інтересу до секторів охорони здоров'я та технологій. Теоретична значимість полягає у систематизації знань про ключові внутрішні та зовнішні фактори успішності стартапів, визначенні впливу венчурного капіталу та ринкових умов на їх розвиток, поглибленні розуміння ролі технологічних інновацій та бізнес-моделей, а також запропонованих методологічних підходах до дослідження успішності стартапів, які можуть бути використані для аналізу інших галузей та ринків.

Практична цінність: Розроблені рекомендації можуть бути використані засновниками стартапів для підвищення ефективності управління, інвесторами для оцінки перспективності стартапів, а також консультантами та аналітиками у сфері інформаційних технологій.

RESUME

Taras Shevchenko National University of Kyiv,

Faculty of Economics, Department of Economic Cybernetics

Keywords: startups, information technology, venture capital financing, innovation, business models, external environment, market conditions, startup success, IT startups, success factors.

The graduation research of student Vladyslava Sirychenko deals with the evaluation of success factors for startups in the IT sector.

The work is interesting for entrepreneurs, investors, consultants, and analysts in the field of information technology, as it provides comprehensive insights and practical recommendations for enhancing the effectiveness and competitiveness of IT startups.

Pages: 48, tables: 2, bibliog.: 22, append.: 2.

ЗМІСТ

ВСТУП.....	4
РОЗДІЛ 1: ЗАГАЛЬНИЙ ОГЛЯД ФАКТОРІВ УСПІШНОСТІ СТАРТАПІВ В ІТ ..	6
1.1 Вступ до ІТ-стартапів.....	7
1.3 Зовнішнє середовище стартапів.....	14
Висновки до розділу.....	18
РОЗДІЛ 2: ТЕОРЕТИЧНІ ОСНОВИ ФАКТОРІВ УСПІШНОСТІ СТАРТАПІВ В ІТ	
.....	19
2.1 Історичний огляд досліджень.....	19
2.2 Аналіз методологій досліджень	23
2.3 Аналіз даних по фінансуванню стартапів Індії	26
Висновки до розділу.....	28
РОЗДІЛ 3: АНАЛІЗ ДАНИХ ПРО ФІНАНСУВАННЯ СТАРТАПІВ У ЄВРОПІ...	29
3.1 Статистичний огляд показників.....	29
3.2 Класифікація стартапів через LLM та статистичний огляд результатів	33
3.3 Українські стартапи та фінансування.....	39
3.4 Американські стартапи та фінансування	40
Висновки до розділу та рекомендації для стартапів	42
ВИСНОВКИ	43
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	45
ДОДАТКИ	49

ВСТУП

Актуальність дослідження ІТ-стартапів зумовлена стрімким розвитком технологічного сектору та його значним впливом на глобальну економіку. На сучасному етапі розвитку інноваційного підприємництва важливим є дослідження факторів, що впливають на успішність ІТ-стартапів, а також оцінка внутрішніх і зовнішніх умов, що сприяють або перешкоджають їх розвитку.

Об'єктом дослідження є сам процес створення та розвитку ІТ-стартапів, а предметом виступає вивчення факторів успішності цих підприємств. Це дозволяє глибше зрозуміти механізми та умови, які сприяють інноваціям та виживанню стартапів в динамічному технологічному середовищі.

Метою даного дослідження є виявлення і аналіз ключових факторів, що визначають успіх ІТ-стартапів. Основними завданнями є вивчення наукових досліджень про фінансування, внутрішні та зовнішні умови діяльності стартапів, а також оцінка їх впливу на успішність бізнесу.

Для досягнення поставлених цілей використовуються методи критичного аналізу, статистичної обробки даних, класифікації та порівняльного аналізу.

Наукова новизна полягає у формуванні комплексного підходу до оцінки факторів успішності ІТ-стартапів, що дозволяє виявити взаємозв'язки між різними аспектами діяльності та їх вплив на ефективність інноваційного процесу. В ході дослідження було зібрано власний набір даних про стартапи в ЄС та використано сучасні методи класифікації по власному фреймворку через LLM, що дозволило оцінити та поглянути на статистичні дані, які були до цього недоступними.

Результати дослідження мають практичне значення для розробників політик, інвесторів та керівників стартапів, сприяючи формуванню стратегій для підвищення ефективності інноваційного розвитку в ІТ-секторі.

Інформаційна база дослідження охоплює наукові публікації, статистичні дані про фінансування стартапів в ЄС, а також результати дослідження публікацій про стан ринку стартапів у США та Україні.

Структура роботи включає три основні розділи, кожен з яких висвітлює окремий аспект дослідження, починаючи від загального огляду і закінчуючи детальним аналізом даних про фінансування. Висновки, наведені в кожному розділі, узагальнюються в останньому розділі, де формулюються рекомендації для подальшого розвитку ІТ-стартапів.

РОЗДІЛ 1: ЗАГАЛЬНИЙ ОГЛЯД ФАКТОРІВ УСПІШНОСТІ СТАРТАПІВ В ІТ

У вступному розділі даної дипломної роботи було поставлено за мету провести загальний огляд факторів успішності стартапів в ІТ-сфері. Розвиток технологій і швидке зростання індустрії інформаційних технологій створили сприятливі умови для заснування і розвитку стартапів. В той же час, висока конкуренція та специфіка ринку накладають певні виклики, успішний подолання яких можливий лише при правильному управлінні та стратегічному плануванні.

Важливим аспектом є аналіз внутрішніх та зовнішніх факторів, які впливають на успішність стартапу. До внутрішніх факторів належить якість продукту або послуги, командна робота, інноваційність та досвід засновників. Зовнішні фактори включають ринкову ситуацію, наявність конкурентів, правове регулювання індустрії та доступність фінансування. Розуміння та вміння управління цими факторами значно підвищує шанси на успіх.

Дослідження ролі інновацій та технологічних рішень у процесі заснування та розвитку стартапів в ІТ-сфері є критичним аспектом. Інновації не лише визначають унікальність продукту, але й сприяють його конкурентоспроможності на ринку. Якість і швидкість розробки продукту суттєво залежать від обраної технології та технічної бази.

Фінансування є ще одним вирішальним фактором. Обсяг і джерела фінансування можуть впливати на швидкість розвитку проекту та його можливості здійснити необхідні інвестиції в маркетинг, наукові дослідження та розробку.

На завершення, варто підкреслити, що успішність стартапу в ІТ-сфері є результатом складної взаємодії багатьох факторів. Управління стартапом, враховуючи специфіку ІТ-індустрії, вимагає глибокого розуміння технологій, ринку, потреб користувачів та вміння швидко адаптуватися до змін. Тому аналіз та

оцінка цих факторів є важливим етапом у підготовці та реалізації будь-якого стартап-проекту в сфері інформаційних технологій.

1.1 Вступ до IT-стартапів

У вступному розділі до теми "Оцінка факторів успіху стартапів в IT-сфері" розглядається основне поняття та означення стартапів у сфері інформаційних технологій (IT). В контексті глобалізованого світу, де інновації та технологічний розвиток виступають ключовими драйверами економічного зростання, IT-стартапи набувають особливого значення. Вони стають не просто елементом новітнього економічного порядку, але й інструментом соціального й культурного впливу, сприяючи формуванню нових тенденцій у суспільстві та бізнес-середовищі.

Специфіка IT-стартапів полягає у їх високій інноваційності, технологічній направленості та орієнтації на швидке зростання. Відмінною рисою таких стартапів є використання передових IT-розробок для створення унікальних продуктів чи послуг, які мають потенціал радикально змінити звичайні підходи в традиційних галузях або створити нові ринкові ніші. Важливою характеристикою IT-стартапів є їх гнучкість та здатність швидко адаптуватися до змін у технологічному середовищі та потребах користувачів.

В контексті дослідження процесу визначення та оцінки факторів успіху IT-стартапів, значущим є розуміння особливостей їх функціонування та розвитку. Стартапи в IT-сфері часто процвітають завдяки інноваційному підходу, залученню інвестицій, ефективному управлінню технологічними процесами, та створенню співпраці між різними учасниками технологічного ландшафту. Врахування цих аспектів є вирішальним для забезпечення довгострокового успіху та сталого росту високотехнологічних підприємств.

У підсумку, розуміння сутності та специфіки IT-стартапів є фундаментальним для оцінки факторів їх успіху. Визначення ключових параметрів, на які треба

звертати увагу при аналізі стартапів в ІТ-сфері, є критичним не лише для науковців та підприємців, але і для інвесторів, які прагнуть підтримувати інноваційний розвиток та реалізацію потенціалу новітніх технологій. Таким чином, перша глава надає необхідний теоретичний фундамент для подальшого дослідження ключових факторів, які сприяють успіху стартапів у сфері інформаційних технологій.

1.1.1 Визначення ІТ-стартапу

У вступі до розділу 1.1.1 "Визначення ІТ-стартапу" розглянемо концепцію стартапу у контексті інформаційних технологій. Як відомо, стартап може бути визначений як суб'єкт підприємницької діяльності на ранніх етапах свого розвитку, який має на меті розробити і запропонувати ринку інноваційний продукт чи послугу. Важливою особливістю стартапів у сфері інформаційних технологій є їхня орієнтація на швидке зростання та масштабування, часто з використанням сучасних цифрових платформ та інструментів.

Для детальнішого розуміння, варто виокремити ключові характеристики ІТ-стартапу, серед яких: інноваційність продукту або послуги, технологічна основа діяльності, високий потенціал зростання та орієнтація на глобальний ринок. Інноваційність визначається не лише новизною ідеї, а й тим, наскільки ефективно вона вирішує конкретні потреби користувачів. Технологічна основа передбачає, що продукти або послуги ґрунтуються на використанні новітніх досягнень у галузі програмування, штучного інтелекту, Blockchain, великих даних та інших перспективних технологій. Потенціал зростання і орієнтація на глобальний ринок вказують на те, що ІТ-стартап має амбіції виходу на широкі міжнародні ринки, а не обмежується локальним рівнем.

Додатково, стартапи в сфері інформаційних технологій характеризуються своєю тимчасовою організацією та пошуком ефективної бізнес-моделі, що дозволяє їм швидко адаптуватися до змін у ринковому середовищі і технологічних

тенденціях. Ця гнучкість та здатність до ітерацій є ключовими для успішної реалізації стартап-проектів.

На закінчення варто зазначити, що розуміння визначення та ключових характеристик ІТ-стартапу є фундаментальним для аналізу факторів успіху таких компаній. Інноваційні ідеї, здатність швидко адаптуватися і використання передових технологій створюють на підґрунті сучасної цифрової екосистеми ідеальні умови для розвитку новаторського підприємництва. Таким чином, глибоке розуміння елементів, що складають ІТ-стартап, значною мірою сприяє розкриттю секретів їх успішності та зростання.

1.1.2 Особливості ІТ-стартапів

У цьому розділі наша увага приділяється з'ясуванню ключових особливостей ІТ-стартапів, які є важливим елементом в екосистемі сучасних технологічних підприємств. ІТ-стартапи – це не просто малі новостворені компанії у сфері інформаційних технологій. Ці організації мають свої унікальні характеристики, які обумовлені особливостями технологічного ринку, швидкоплинністю технологій, високим рівнем невизначеності та потенціалом швидкого зростання. Розглянемо докладніше ключові особливості ІТ-стартапів, що відрізняють їх від традиційних підприємств та інших видів стартапів.

Перш за все, ІТ-стартапи характеризуються інноваційністю. Це не лише у впровадженні нових технологій або розробці нових продуктів і послуг, а й у способах ведення бізнесу, маркетингових стратегіях та моделях доходів. Іновації є ключовим драйвером, який забезпечує стартапам перевагу на ринку та можливість швидкого зростання.

Другою важливою особливістю є швидкість розвитку та масштабування. В умовах постійної технологічної зміни ІТ-стартапи мають бути гнучкими і здатними

швидко адаптуватися до змін на ринку. Це може включати розвиток нових функцій продуктів, оптимізацію процесів або швидке входження на нові ринки.

Третя особливість - високий рівень невизначеності та ризику. Засновники ІТ-стартапів часто працюють в умовах великої невизначеності щодо потенціалу ринку, реакції користувачів на продукт і можливості залучення інвестицій. Впоратися з цими викликами вимагає не тільки технічної експертизи, а й бізнес-номенклатури та гнучкості в прийнятті рішень.

Резюмуючи, ІТ-стартапи показують ряд унікальних характеристик, що робить їх важливою частиною глобальної технологічної екосистеми. Їх інноваційність, гнучкість і рівень адаптивності дають їм потенціал до швидкого зростання та розвитку, в той же час підвищуючи рівень ризику і невизначеності. Вивчення та оцінка факторів успіху цих стартапів можуть надати цінні вказівки як для засновників таких проєктів, так і для інвесторів і політиків в області малого і середнього підприємництва.

1.2 Ключові фактори успіху для ІТ-стартапів

Вступ до цього розділу присвячений визначенню та аналізу основних елементів, які впливають на успішність молодих компаній у сфері інформаційних технологій. У сьогоdnішньому нестабільному бізнес-середовищі стартапам необхідно враховувати безліч факторів, як внутрішніх, так і зовнішніх, аби забезпечити своє виживання та розвиток. Зростання конкуренції, швидкі зміни технологій, потреби споживачів та інвестиційний клімат - все це безпосередньо впливає на майбутнє молодих ІТ-компаній. Цей розділ покликаний ідентифікувати ключові фактори, які потенційно можуть підвищити шанси стартапів на успішну діяльність.

Одним із основних факторів успіху є інноваційність продукту або послуги, яку пропонує стартап. Новизна та унікальність рішень, здатність задовольнити

невеликі, але важливі потреби споживачів можуть відіграти вирішальну роль у виживанні компанії на ринку. При цьому, важливою є не лише сама ідея, а й її правильна реалізація та вибудовування ефективної бізнес-моделі, здатної забезпечити прибутковість.

Команда також грає надзвичайно важливу роль у розвитку стартапу. Злагожденість та професіоналізм команди, здатність швидко адаптуватися до змін та вирішувати складні задачі становить фундамент, на якому будується весь процес розвитку проекту. Особлива увага в цьому контексті приділяється лідерським якостям засновників та їхній здатності мотивувати та надихати свою команду.

Фінансування є ще одним вирішальним фактором. Залучення інвестицій, ефективне управління капіталом та вміння досягати до самоокупності без зайвих затримок є ключовими компонентами фінансової стійкості стартапу. Вміння представити свій проект інвесторам, переконати їх у перспективності бізнесу та використати отримані кошти з максимальною ефективністю є важливими навичками для засновників.

Технологічна база та інтелектуальна власність також мають значення. Захист інноваційних розробок, патентування ідей та забезпечення високого рівня технологічної підготовки продукту можуть стати решіткою, яка захистить стартап від конкуренції.

У підсумку, оцінка ключових факторів успіху для ІТ-стартапів свідчить про те, що в успішному стартапі важливі всі елементи: інноваційна ідея, сильна команда, сталий фінансування, технологічна ефективність та захист інтелектуальної власності. Комбінація цих чинників у правильному пропорцієні забезпечує не лише виживання на ринку, а й можливість для швидкого зростання та розвитку.

1.2.1 Технологічні інновації

У цьому розділі акцент зроблено на важливості інновацій у технологічному секторі як основного двигуна прогресу і успіху стартапів. Технологічні інновації визначаються як ключовий елемент, що впливає на здатність компаній не лише залишатися на плаву в умовах жорсткої конкуренції, але й рухатися вперед, здобувати нові ринкові ніші і відкривати необмежені можливості для росту.

В одному з джерел аналізується співвідношення між інноваційними рішеннями та комерційним успіхом стартапів, де зазначається, що впровадження новітніх технологічних рішень іде в ногу зі збільшенням доходів компанії та її впізнаваності на ринку. При цьому наголошується на тому, що інновації повинні базуватися на глибокому аналізі потреб кінцевих користувачів. [1]

У контексті іншого дослідження розкривається питання про різноманітні стратегії виходу на ринок для іноваційних технологічних продуктів, підкреслюючи велике значення виваженого підходу до розробки і запуску продукту. Особливу увагу приділяється методам тестування продуктів перед їх комерційним запуском, що допомагає виявити потенційні слабкі сторони й оптимізувати продукт для кінцевого споживача. [2]

Також освітлено аспект важливості залучення інвестицій для розвитку інноваційних проектів у технологічній сфері. За вказівкою деяких авторів, залучення фінансування є критичним кроком для стартапів, оскільки воно дозволяє не тільки розвивати існуючі продукти, але й проводити дослідження ринку та тестування нових ідей, що забезпечує компанії переваги перед конкурентами. [3]

Важливим аспектом, що висвітлюється у додатковому джерелі, є співпраця між стартапами та великими корпораціями в контексті розвитку та впровадження інновацій. Підкреслюється, що така співпраця може принести обопільну вигоду: стартапам - доступ до ресурсів та ринків, корпораціям - інноваційні рішення, які можуть бути інтегровані в їхні продукти або процеси. [4]

Загалом, цей розділ ретельно вивчає роль технологічних інновацій у успіху стартапів в ІТ-сфері, вказуючи на необхідність постійного пошуку нововведень і адаптації до змінюваних потреб ринку. Висвітлення різноманітних аспектів починаючи від розробки інноваційних продуктів до залучення інвестицій і співпраці з корпораціями, надає всебічне розуміння важливості інноваційного підходу для забезпечення довгострокового успіху стартапів в технологічній індустрії.

1.2.2 Бізнес-моделі та стратегії

Цей розділ присвячений аналізу бізнес-моделей та стратегій, які використовуються в стартапах в області інформаційних технологій. У цьому розділі детально досліджуються різні підходи до створення та масштабування продуктів в межах цього галузі, враховуючи особливості стартап-екосистеми та вимоги ринку.

В одному з джерел описуються загальні тенденції в розвитку стартапів, підкреслюючи важливість адаптивності бізнес-моделей до змінних умов ринку та потреб споживачів. Особлива увага приділяється механізмам залучення інвестицій та ролі венчурного капіталу в просуванні інноваційних проєктів у сфері ІТ. [5]

Як згадується в певному дослідженні, успіх ІТ-стартапів значною мірою залежить від здатності швидко адаптуватися до нових технологій та методологій розробки, включаючи застосування DevOps та мікросервісів. Це не лише сприяє ефективній розробці продукту, але й забезпечує високу швидкість внесення змін та оптимізацію процесів. [6]

Важливим аспектом, на який звертає увагу інше джерело, є структурування бізнес-процесів та формування команди, здатної ефективно втілювати обрану бізнес-модель у життя. Ключовими елементами успіху є залучення висококваліфікованих фахівців, встановлення ефективної взаємодії між різними підрозділами та створення мотивуючої системи винагород. [7]

Ще одне дослідження підкреслює значення інновацій та креативності в бізнес-моделях стартапів ІТ-галузі. Інтенсивна конкуренція та швидкі зміни технологічного середовища вимагають постійного пошуку нових ідей та підходів для забезпечення стійкого розвитку та прогресу проєктів. [8]

У підсумку, успіх стартапів в ІТ-сфері в значній мірі залежить від обраної бізнес-моделі та стратегії розвитку, здатності адаптуватися до змінних умов, впровадження інновацій та збудування ефективної команди. Враховуючи ці аспекти, ключовим для стартапів є розуміння потреб ринку та вибір гнучких підходів до розвитку продуктів і послуг.

1.3 Зовнішнє середовище стартапів

Зовнішнє середовище стартапів відіграє ключову роль у формуванні їхнього потенціалу до успіху в ІТ-сфері. Воно охоплює широкий спектр факторів, які безпосередньо або опосередковано впливають на діяльність молодих технологічних компаній. Важливість розуміння цього середовища полягає в можливості ідентифікації зовнішніх можливостей та загроз, які можуть впливати на стратегічне планування та прийняття рішень у межах стартапу.

Економічні умови є одним із найважливіших аспектів зовнішнього середовища, що включає такі показники, як рівень інфляції, валютні курси, відсоткові ставки, загальний економічний клімат країни, у якій діє стартап. Економічна стабільність позитивно впливає на інвестиційну привабливість технологічних проєктів, сприяє залученню зовнішнього фінансування.

Політичне середовище також має велике значення для стартапів в ІТ-сфері, оскільки включає в себе регуляторну політику, що визначає правові рамки для діяльності в сфері високих технологій, запатентовані технології, стандартизацію тощо. Політична стабільність та передбачуваність сприяють зростанню довіри інвесторів і, як наслідок, збільшенню інвестицій у сектор ІТ.

Соціокультурний аспект зовнішнього середовища охоплює цінності, уподобання споживачів, рівень освіти населення та його відкритість до нових технологій, що безпосередньо впливає на прийняття та впровадження нових ІТ-продуктів. Розуміння цих тенденцій є важливим для розробки продукту, який відповідатиме потребам ринку.

Технологічне середовище стартапів включає в себе доступність новітніх технологій, рівень технологічної конкуренції та швидкість розвитку технологічних інновацій. Воно безпосередньо впливає на можливість стартапу пропонувати нові або удосконалені продукти і послуги, які відповідають потребам ринку та випереджають конкурентів.

Врахування усіх цих аспектів зовнішнього середовища є критично важливим для успіху стартапів у ІТ-сфері. Воно дозволяє керівництву компаній ефективно адаптувати свою стратегічну поведінку, мінімізувати ризики та збільшити шанси на успішне входження на ринок. Тому детальний аналіз зовнішнього середовища повинен стати невід'ємною частиною стратегічного планування кожного стартапу в інформаційних технологіях.

1.3.1 Роль венчурного фінансування

Даний розділ присвячений аналізу значення венчурного капіталу для розвитку та успішної діяльності стартапів у ІТ-сфері. У контексті стрімкого розвитку цифрових технологій та зростаючої конкуренції на ринку, залучення венчурного фінансування набуває особливої ролі в обігу стартапів, надаючи їм можливість стати більш конкурентоспроможними та інноваційними.

Венчурне фінансування є ключовим фактором, що сприяє залученню додаткових ресурсів для розвитку інноваційних проєктів, наданню менторської підтримки молодим компаніям та доступу до кола професійних консультантів і потенційних клієнтів. Партнерство з венчурними фондами дозволяє стартапам в ІТ-

сфері отримати не тільки фінансування, але й цінний досвід управління та розвитку бізнесу.

Злагоджений механізм венчурного фінансування вимагає від стартапів готовності до тісної співпраці з інвесторами, постійного аналізу ринку та розроблення інноваційних продуктів, що задовольняють потреби користувачів. Важливим аспектом є також здатність стартапу демонструвати свої конкурентні переваги та потенціал для масштабування на глобальному ринку.

За останні роки, у зв'язку з технологічним прогресом, роль венчурного фінансування у розвитку стартапів значно посилилася. Наразі цей вид інвестицій відіграє незамінну роль у формуванні екосистеми інновацій, стимулюванні технологічного підприємництва та розвитку нових продуктів у ІТ-сфері.

У підсумку можна сказати, що роль венчурного фінансування у розвитку стартапів ІТ-сфери є критичною. Воно не лише забезпечує необхідні фінансові ресурси для росту та інновацій розвитку, але й відкриває доступ до важливих ринкових знань, мережі контактів і стратегічних партнерств. Таким чином, венчурне фінансування підтримує інноваційні процеси цифрової економіки, сприяє створенню новітніх технологій і формує основу для успішного майбутнього ІТ-стартапів.

1.3.2 Вплив ринкових умов

У цьому розділі ми розглядаємо вплив ринкових умов на успіх стартапів в ІТ-сфері, адже ринкові умови є критичним зовнішнім чинником, який безпосередньо визначає можливості для зростання та розвитку стартапів. Оцінка ринкової ситуації, аналіз конкурентоспроможності ідеї, визначення потенційної аудиторії і клієнтурі допомагає власникам стартапів формувати адекватні стратегії входу на ринок та подальшого просування.

Перш за все, суттєвим аспектом є наявність попиту на запропонований продукт чи послугу. Стартапи в ІТ-сфері повинні ґрунтуватися на детальному аналізі потреб ринку, щоб гарантувати, що їхня пропозиція зустрічатиметься з реальним інтересом. Це означає, що підприємці мають чітко ідентифікувати проблему, яку вони вирішують, і переконатися, що для цієї проблеми існує достатній обсяг зацікавленої аудиторії.

Другий важливий аспект – це розуміння конкурентного середовища. Необхідно провести глибокий аналіз присутніх на ринку гравців, їх сильних та слабких сторін, а також визначити потенційні бар'єри для входу. Ретельне дослідження конкурентів дозволяє новим стартапам розробити унікальну ціннісну пропозицію, яка допомагає виокремитися на ринку та залучити перших клієнтів, не дивлячись на високу конкуренцію.

Третім фактором є доступність ресурсів та інфраструктури. Для ІТ-стартапів критично важливо мати доступ до необхідних технологій, висококласних фахівців та розвинутої інформаційної інфраструктури. Крім того, на ранніх етапах розвитку важливою є підтримка у вигляді інвестицій, грантів або венчурного фінансування, які дозволяють стартапам налагоджувати процеси та швидко масштабуватися.

Отже, ринкові умови грають домінуючу роль у визначенні потенціалу стартапів у ІТ-сфері до успіху. Без глибокого аналізу попиту, конкурентного середовища та доступності ресурсів стартапам буде складно знайти своє місце на ринку, залучити клієнтів та досягти фінансової стійкості. Постійний моніторинг змін у ринкових умовах, адаптація до поточних вимог ринку та гнучкий підхід до розвитку продукту – це ключ до успішного масштабування та визнання стартапу на ринку.

Висновки до розділу

У розділі 1 було проведено загальний огляд факторів успішності стартапів у сфері інформаційних технологій. Розвиток технологій і швидке зростання ІТ-індустрії створили сприятливі умови для заснування стартапів, проте висока конкуренція та специфіка ринку вимагають грамотного управління та стратегічного планування. Важливими аспектами успішності є аналіз внутрішніх факторів, таких як якість продукту, командна робота, інноваційність та досвід засновників, а також зовнішніх факторів, включаючи ринкову ситуацію, конкуренцію, правове регулювання та доступність фінансування. Дослідження ролі інновацій і технологічних рішень показало, що вони є критичними для забезпечення конкурентоспроможності та успіху на ринку. Фінансування, зокрема венчурне, виявилось вирішальним фактором для розвитку стартапів, оскільки воно забезпечує необхідні ресурси для росту та інновацій. Ринкові умови також мають значний вплив, оскільки попит на продукт, конкурентне середовище та доступність ресурсів визначають потенціал успішності стартапу. Таким чином, комплексний аналіз та оцінка цих факторів є важливими для підготовки та реалізації будь-якого стартап-проекту в ІТ-сфері.

РОЗДІЛ 2: ТЕОРЕТИЧНІ ОСНОВИ ФАКТОРІВ УСПІШНОСТІ СТАРТАПІВ В ІТ

Стартапи в галузі інформаційних технологій стали рушійною силою інновацій та економічного зростання у XXI столітті. Проте, незважаючи на величезний потенціал, більшість ІТ-стартапів зазнають невдачі на ранніх стадіях свого розвитку. Це підштовхнуло дослідників до вивчення факторів, які сприяють успішності таких компаній.

Протягом останніх двох десятиліть було проведено чимало досліджень, спрямованих на виявлення ключових детермінант успіху ІТ-стартапів. Ці дослідження охоплювали широкий спектр аспектів, від бізнес-моделей та ринкових стратегій до організаційної структури, управління персоналом та інноваційних підходів.

У цьому розділі буде представлено історичний огляд досліджень факторів успішності ІТ-стартапів, аналіз використаних методологій та критичний огляд висновків попередніх студій. Такий ретельний аналіз минулих досліджень допоможе визначити прогалини у знаннях та виявити нові перспективні напрямки для подальших досліджень у цій галузі.

2.1 Історичний огляд досліджень

Інтерес до дослідження факторів успішності ІТ-стартапів зародився наприкінці 1990-х років, коли відбувся бум інтернет-компаній та зародження нової цифрової економіки. Перші дослідження в цій галузі були зосереджені на аналізі бізнес-моделей та стратегій виходу на ринок новостворених технологічних компаній.

На початку 2000-х років увага дослідників почала зміщуватися в бік вивчення ролі підприємницького досвіду, командної динаміки та соціальних мереж у

досягненні успіху стартапами. Ці дослідження виявили, що наявність досвідченої команди та сильних зв'язків з інвесторами та партнерами є критично важливими для розвитку молодих ІТ-компаній.

З середини 2000-х років дослідники почали приділяти більше уваги інноваційним процесам та стратегіям розробки продуктів у ІТ-стартапах. Значний внесок у цю галузь зробили такі дослідники, як Стів Бланк та Ерік Ріс, які запропонували нові методології, зокрема "Customer Development" та "Lean Startup".

Протягом останнього десятиліття дослідження факторів успішності ІТ-стартапів стали більш різноманітними та міждисциплінарними. Вони охоплювали такі аспекти, як вплив екосистеми стартапів, ролі корпоративної культури, стратегії фінансування та підходи до управління інтелектуальною власністю.

Таким чином, дослідження факторів успішності ІТ-стартапів пройшли шлях від вузького фокусу на бізнес-моделях до більш комплексного та багатогранного вивчення різноманітних організаційних, людських та екосистемних чинників, що визначають успіх або невдачу цих інноваційних компаній.

2.1.1 Розвиток наукового інтересу до ІТ-стартапів

У другому розділі розглядається розвиток наукового інтересу до ІТ-стартапів. Вступ до даного розділу підкреслює значущість дослідження феномена стартапів в контексті інформаційних технологій, враховуючи їх дедалі зростаючий вплив на сучасну економіку та інноваційний потенціал.

У одному з джерел розглядається новий підхід до теорії ігор середнього поля з мейджорними та мінорними гравцями, акцентуючи увагу на оптимізації поведінки представницьких мінорних гравців та впливі мейджорних гравців на умови гри. Автори аргументують, що їх підхід дозволяє домогтися наближеного рівноважного стану в іграх з великою кількістю учасників, вносячи вклад у розуміння динамічних систем в контексті розвитку ІТ-проектів. [9]

В іншому дослідженні аналізуються масштабовані лінійно-квадратичні ігри середнього поля з участю мейджорного гравця. Розроблена методика масштабування та визначення рішень характеризується наявністю через встановлення рішень Риккати, що є ключовим для забезпечення асимптотичної розв'язності. Такий підхід демонструє не тільки теоретичну значимість, але й практичну застосовність при аналізі стратегій мейджорних гравців в ІТ-сфері. [10]

Далі, згадується дослідження, яке ставить під сумнів однозначність взаємодій між мейджорними та мінорними гравцями в іграх середнього поля, порівнюючи різні методи розв'язку - від теорії Неша до майстер-рівнянь. Автори встановлюють еквівалентність цих підходів, що вказує на глибину і складність інтеракцій в ІТ-проектах, де участь великих і малих гравців створює динамічне поле активності. [11]

В останньому обговорюваному матеріалі вивчається важливість полевих досліджень у контексті аналізу даних, отриманих зі супутникових знімків. Це дослідження підкреслює потребу в розробці адаптивних методів збору даних і представляє значимість таких підходів для оцінки соціоекономічних показників, взаємозв'язаних з розвитком ІТ-стартапів. [12]

Отже, ці матеріали разом становлять ґрунтовну основу для аналізу наукового інтересу до ІТ-стартапів, надають нові перспективи для розуміння динаміки в іграх середнього поля та підкреслюють значення комплексного підходу до збору та аналізу даних. Таке різноманіття підходів і методик висвітлює різнобічність і складність досліджуваної теми, водночас вказуючи на потенціал для подальших наукових досліджень у цій сфері.

2.1.2 Значні дослідження у сфері

У розділі 2.1.2 "Значні дослідження у сфері" нашого дипломного дослідження, присвяченого оцінці факторів успіху стартапів у ІТ-сфері, ми розглядаємо ключові

наукові роботи, що формують теоретичну та практичну базу для аналізу вибраних факторів. Важливість комплексного розуміння теми не може бути переоцінена, а тому аналіз значущих досліджень є невід'ємною частиною вивчення.

У одному з джерел розглянуто стратегії навчання нейронних мережеских моделей мови з великим словниковим запасом. Автори стверджують, що, незважаючи на популярність нейронних мережеских моделей в таких застосунках, як розпізнавання голосу та машинний переклад, навчання таких моделей все ще є значно більш витратним з комп'ютерного погляду, порівняно з моделями на основі підрахунку, такими як Kneser-Neu. Дослідження зосереджується на систематичному порівнянні стратегій представлення та навчання великих словників, включаючи softmax, ієрархічний softmax, цілеспрямований вибірок, оцінку шумового контрасту та самонормалізацію. Окрім того, вводиться розширення самонормалізації для правильної оцінки подібності та представлено ефективний варіант softmax. Кожен метод оцінюється на трьох популярних бенчмарках, з оглядом на продуктивність у роботі з рідкісними словами, компроміс між швидкістю та точністю, а також доповнюваність до Kneser-Neu. [13]

У підсумку можна зазначити, що аналіз вказаного дослідження показує глибоке розуміння проблем і потенційних рішень у сфері навчання нейронних мережеских моделей мов з великими словниковими запасами. Знання, отримане з цього джерела, стане в нагоді при розробці та оптимізації алгоритмів для стартапів в ІТ-сфері, які прагнуть досягнути успіху в галузі машинного навчання та обробки природної мови.

2.2 Аналіз методологій досліджень

Цей розділ присвячений аналізу методологій досліджень стартапів в ІТ-сфері, є основою для глибокого розуміння та аналізу факторів, які впливають на успіх стартапів. В даному розділі проводиться огляд різноманітних підходів та

методологій, які використовуються в наукових дослідженнях для оцінки факторів успіху стартапів, особливо у сфері інформаційних технологій. Цей аналіз дозволяє ідентифікувати найбільш релевантні методології, що будуть використовуватися в наступних розділах дипломної роботи.

Перш за все, важливо розуміти, що однією з ключових характеристик досліджень в області стартапів є мультидисциплінарний підхід. Стартапи в ІТ-сфері вимагають унікального поєднання технічних знань, комерційної стратегії та вміння розпізнавати та досягати ринкового потенціалу. У зв'язку з цим, для оцінки факторів успіху стартапів використовуються різні дослідницькі методи, що охоплюють як кількісні, так і якісні аналізи.

Кількісні методи зазвичай включають статистичний аналіз даних, отриманих з опитувань, ринкових досліджень та фінансових звітів. Ці методи дозволяють виявити кореляції між кількісними вимірами успіху, такими як прибутковість та швидкість зростання, та відповідними факторами. Якісні методи, у свою чергу, зосереджуються на аналізі інтенсивних інтерв'ю, кейс-стаді та історій успіху випускників інкубаторів та акселераторів. Ці методи дозволяють глибше зрозуміти унікальні особливості та індивідуальні стратегії успіху стартапів.

Важливою частиною аналізу методологій досліджень є вивчення теоретичних рамок. Це включає теорії підприємництва, інновацій, бізнес-стратегії, технічних інновацій та управління проектами. Розуміння теоретичної бази є ключовим для глибокого аналізу та інтерпретації даних у контексті стартапів в ІТ-сфері.

У підсумку, даний розділ пропонує всебічний аналіз різних методологій, які можуть бути застосовані для дослідження факторів успіху стартапів в ІТ-сфері. Детальний огляд кількісних та якісних методів, а також теоретичних рамок, дає можливість вибрати найефективніші підходи для подальшого глибокого аналізу. Це, в свою чергу, підвищує наукову цінність дослідження та сприяє розробці надійних висновків про фактори успіху стартапів у динамічному середовищі ІТ-

сфери.

2.2.1 Кількісні методи

У дослідженнях факторів успішності ІТ-стартапів широко застосовувалися кількісні методи, які дозволяли збирати та аналізувати числові дані з різних джерел. Ці методи відігравали важливу роль у виявленні статистично значущих зв'язків між різними змінними та формулюванні узагальнень на основі емпіричних даних.

Одним з найпоширеніших кількісних методів є опитування, які проводяться серед засновників стартапів, інвесторів, співробітників та інших зацікавлених сторін. Опитування дозволяють отримати інформацію про характеристики команди, бізнес-моделі, стратегії фінансування, ринкові умови та інші фактори, які можуть впливати на успіх або невдачу стартапу.

Крім опитувань, дослідники також використовують статистичний аналіз наявних даних про стартапи, таких як фінансові показники, розмір команди, інвестиції та метрики продуктивності. Ці дані можуть бути отримані з відкритих джерел або спеціалізованих баз даних. Застосовуючи різноманітні статистичні методи, такі як регресійний аналіз, кластерний аналіз або моделювання структурними рівняннями, дослідники можуть визначати значущі фактори та встановлювати причинно-наслідкові зв'язки.

Крім того, у деяких дослідженнях використовуються експериментальні методи, де контрольовані змінні маніпулюються для вивчення їх впливу на результати стартапів. Наприклад, можуть проводитися експерименти з різними підходами до розробки продукту, маркетинговими стратегіями або методами управління командою.

Загалом, кількісні методи дозволяють дослідникам перевірити гіпотези, виявити статистично значущі зв'язки та сформулювати узагальнення на основі великих обсягів даних. Однак ці методи мають певні обмеження, оскільки не

завжди можуть повністю врахувати складність та контекстуальні фактори, що впливають на успіх ІТ-стартапів.

2.2.2 Якісні методи

В цьому розділі розглядаються різноманітні методології якісного аналізу, які можуть бути використані для оцінки факторів успіху стартапів в ІТ-сфері. Якісні методи дослідження характеризуються глибоким і всебічним підходом до аналізу зібраних даних, що дозволяє більше зосередитись на розумінні явищ, ніж на їх кількісному вимірі. Використання якісних методів важливе для з'ясування механізмів функціонування стартапів, оскільки дозволяє розкрити унікальні історії, досвід та перспективи засновників та команд.

Одним з основних якісних методів є глибинне інтерв'ю. Цей метод включає детальні бесіди з представниками стартап-компаній, інвесторами, експертами в ІТ-сфері. Глибинні інтерв'ю дозволяють зібрати унікальні дані про особистий досвід, мотивації, виклики та стратегії, які використовуються в процесі розвитку стартапів. Важливим є створення довірчого середовища під час інтерв'ю, щоб респонденти могли відверто поділитися своєю історією.

Іншим методом якісного аналізу є аналіз випадків (case study). Цей метод передбачає детальне вивчення одного або кількох прикладів стартапів з метою виявлення зразків, механізмів успіху чи причин невдач. Аналіз випадків допомагає глибше зрозуміти контекстуальні фактори, які могли впливати на успіх чи неуспіх стартапів, виявляючи при цьому специфіку ІТ-індустрії.

Фокус-групи також є значущим якісним методом, який може бути включений до стратегії оцінки успішності стартапів. Участь в фокус-групах дозволяє зібрати думки та оцінки відносно певних ідей чи проектів від групи людей. Це сприяє генерації нових ідей і розширення розуміння про проблеми та можливості, які можуть виникати під час роботи стартапу в ІТ-сфері.

Нарешті, етнографічні дослідження можуть бути застосовані для занурення в середовище стартапів з метою спостереження за діяльністю засновників та команд на щоденній основі. Етнографія дозволяє зібрати дані про внутрішні процеси, культуру комунікації, прийняття рішень та інші важливі аспекти роботи.

У підсумку, якісні методи дослідження є ключовими для оцінки факторів успіху стартапів в ІТ-сфері. Вони дозволяють не лише виявити та аналізувати такі фактори, але й глибше зрозуміти процеси, які відбуваються всередині стартапів, їхню культуру, цілі та виклики, з якими вони стикаються.

2.3 Аналіз даних по фінансуванню стартапів Індії

Аналіз даних по фінансуванню стартапів Індії є важливим елементом дослідження, що дозволяє глибше зрозуміти фінансові аспекти розвитку інноваційних компаній у сфері інформаційних технологій. Фінансування стартапів є ключовим чинником, що впливає на їхній успіх або невдачу, оскільки достатні фінансові ресурси забезпечують можливість реалізації бізнес-ідеї, розробки продукту, маркетингових кампаній та масштабування бізнесу.

Фінансування стартапів зазвичай здійснюється через різні джерела, такі як власні кошти засновників, інвестиції від венчурних капіталістів, бізнес-ангелів, корпоративні інвестори та державні гранти. Кожне з цих джерел має свої особливості, переваги та недоліки, які впливають на загальну стратегію розвитку стартапу.

Дані про фінансування стартапів Індії включають різноманітні показники, такі як загальний обсяг залучених інвестицій, кількість раундів фінансування, типи інвесторів, оцінка компаній на різних стадіях розвитку та структура капіталу. Аналіз цих даних дозволяє виявити тенденції та патерни, що сприяють успішному залученню фінансування, а також визначити критичні моменти, коли стартапи найбільше потребують фінансової підтримки.

Одним з важливих аспектів аналізу є вивчення зв'язку між типом інвесторів та успішністю стартапів. Наприклад, венчурні капіталісти часто надають не тільки фінансові ресурси, але й експертну підтримку, доступ до мережі контактів та стратегічні поради, що може значно підвищити шанси на успіх стартапу. З іншого боку, залучення коштів від бізнес-ангелів на ранніх стадіях розвитку може бути менш формальним та більш гнучким, що дозволяє засновникам зосередитися на розробці продукту та побудові команди.

Ще одним ключовим аспектом є аналіз динаміки фінансування на різних стадіях розвитку стартапів. На початкових етапах стартапи часто залучають менші суми інвестицій для проведення досліджень та розробки прототипів. На стадії зростання фінансування зазвичай збільшується для масштабування бізнесу, виходу на нові ринки та проведення маркетингових кампаній. На пізніших стадіях залучені кошти можуть бути спрямовані на придбання інших компаній, розширення продуктової лінійки та підготовку до IPO.

Аналіз даних про фінансування стартапів також включає вивчення успішних кейсів та визначення факторів, що сприяли їхньому фінансовому успіху та були ключовими для нього. Це може включати оцінку ринкових умов, стратегій залучення інвесторів, управління фінансовими ресурсами та побудови стосунків з інвесторами. Такий аналіз допомагає ідентифікувати найкращі практики та рекомендації для стартапів, які прагнуть залучити інвестиції та досягти успіху на ринку.

Таким чином, аналіз даних по фінансуванню стартапів Індії є критично важливим для розуміння фінансових механізмів, що сприяють успішному розвитку інноваційних компаній. Такий аналіз дозволяє виявити ключові фактори успіху, а також якісно визначити основні виклики та ризики, пов'язані з залученням фінансування та інвестицій на різних стадіях життєвого циклу стартапу.

Висновки до розділу

У другому розділі було розглянуто теоретичні основи факторів успішності стартапів в ІТ-сфері, які стали рушійною силою інновацій та економічного зростання у ХХІ столітті. Протягом останніх двох десятиліть було проведено численні дослідження, спрямовані на виявлення ключових детермінант успіху ІТ-стартапів, включаючи бізнес-моделі, ринкові стратегії, організаційну структуру, управління персоналом та інноваційні підходи. Історичний огляд показав еволюцію наукового інтересу до цієї теми, від аналізу бізнес-моделей та підприємницького досвіду до ролі інновацій та технологій. Аналіз методологій досліджень виявив важливість як кількісних, так і якісних методів для комплексного розуміння успішності стартапів. Кількісні методи, такі як статистичний аналіз та опитування, дозволяють виявити кореляції між різними факторами успіху, тоді як якісні методи, такі як глибинні інтерв'ю та аналіз випадків, надають глибоке розуміння унікальних особливостей та стратегій розвитку стартапів. Зокрема, було підкреслено важливість фінансування для розвитку стартапів, включаючи залучення інвестицій від венчурних капіталістів, бізнес-ангелів та інших джерел. Висновки другого розділу акцентують увагу на необхідності подальших досліджень для заповнення прогалин у знаннях та виявлення нових перспективних напрямків для підвищення успішності ІТ-стартапів.

РОЗДІЛ 3: АНАЛІЗ ДАНИХ ПРО ФІНАНСУВАННЯ СТАРТАПІВ У ЄВРОПІ

3.1 Статистичний огляд показників

Для цього розділу ми використовували власний датасет, що був зібраний із сайту www.eu-startups.com. В результаті збору даних вдалося отримати 13420 записів про стартапи, що включає такі дані:

Таблиця 3.1. Колонки та опис зібраних даних

Назва	Опис
Name	Назва стартапу
Category	Країни, в яких стартап базується
Description	Опис стартапу
Based in	Місто, в якому базується стартап
Tags	Теги, що описують діяльність стартапу
Total funding	Загальна сума фінансування, отриманого стартапом
Founded Year	Рік заснування стартапу
Website	Веб-сайт стартапу

Джерело: Зібраний датасет

Датасет включає детальну інформацію про фінансування стартапів в ЄС, що дозволяє здійснити всебічний статистичний огляд ключових показників. Далі наведено детальний опис графіків, що ілюструють різні аспекти фінансування стартапів.

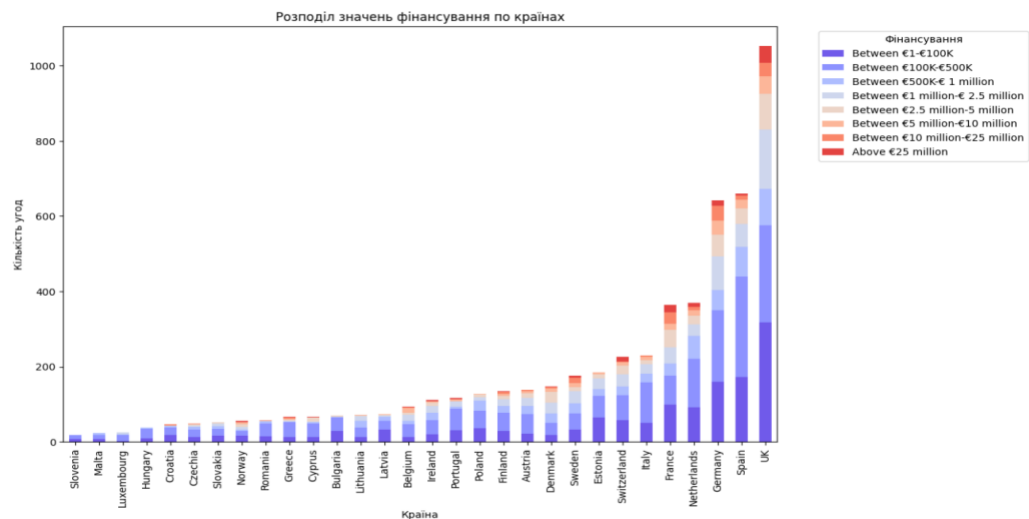


Рисунок 3.1. Розподіл значень фінансування по країнах

Джерело: візуалізація на основі зібраного датасету

На графіку (рис. 3.1) ми бачимо кількість стартапів по кожній країні. Найбільше угод було укладено у Великобританії, Німеччині та Іспанії. Це вказує на те, що ці країни є провідними в залученні фінансування.

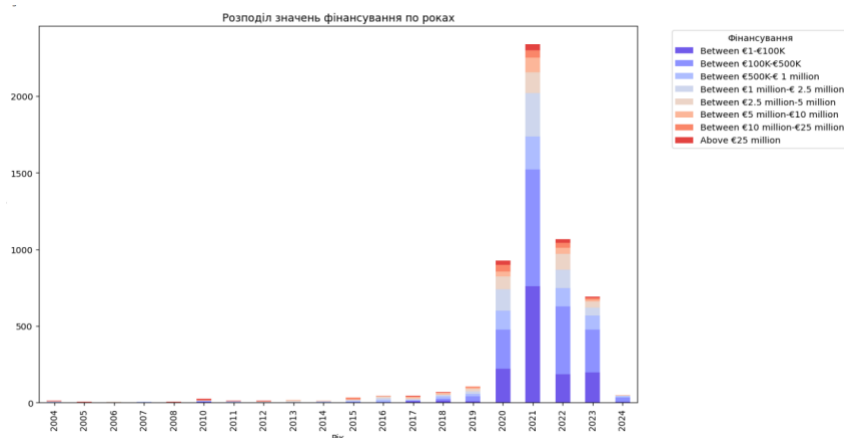


Рисунок 3.2. Розподіл значень фінансування по роках

Джерело: візуалізація на основі зібраного датасету

Графік на рисунку 3.2 показує розподіл фінансування по роках. Ми бачимо, що найбільше угод було укладено у 2021 році. Це може свідчити про значний ріст інвестицій в цей період.

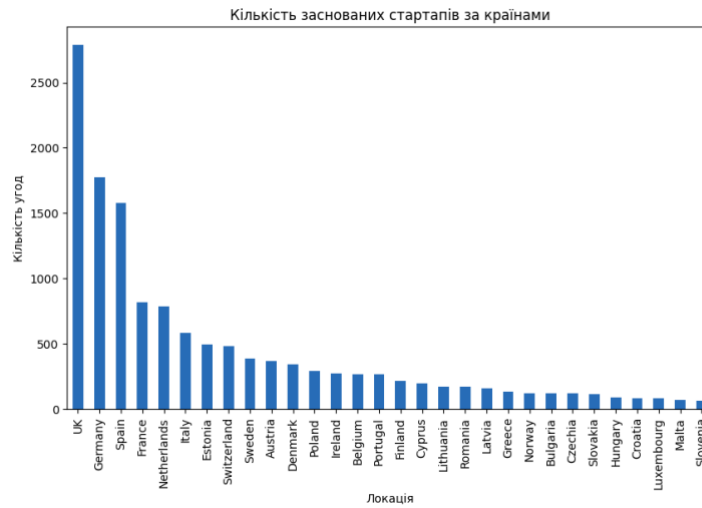


Рисунок 3.3. Кількість заснованих стартапів за країнами

Джерело: візуалізація на основі зібраного датасету

Графік на рисунку 3.3 відображає кількість стартапів по різних локаціям. Знову ж таки, найбільша кількість угод була укладена у Великобританії, Німеччині та Іспанії, що підтверджує попередній графік.

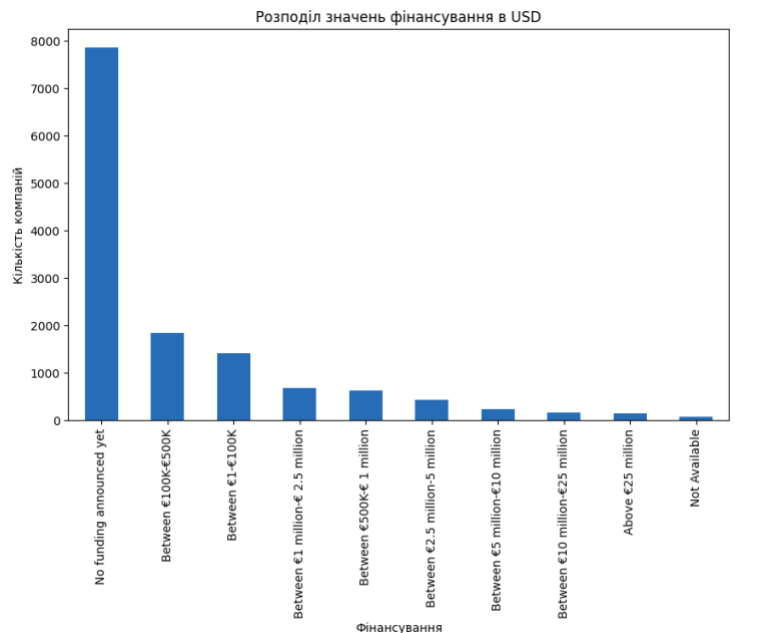


Рисунок 3.4. Розподіл значень фінансування в USD

Джерело: візуалізація на основі зібраного датасету

Графік на рисунку 3.4 показує розподіл угод за категоріями фінансування. Найбільша кількість угод стосується категорії "No funding announced yet", що може вказувати на відсутність даних про фінансування у великій кількості компаній чи недоступність таких даних.

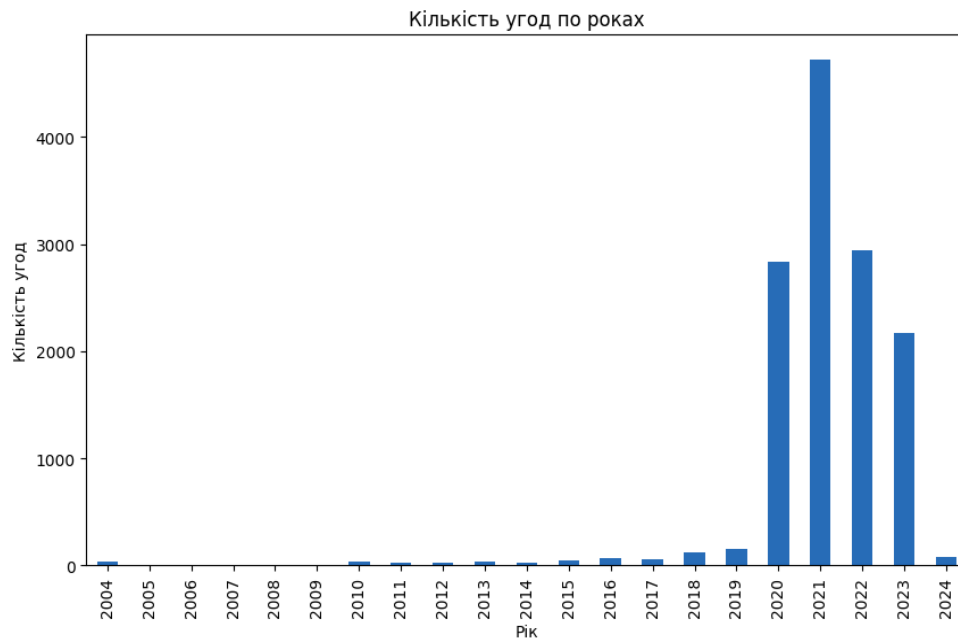


Рисунок 3.5. Кількість угод по роках

Джерело: візуалізація на основі зібраного датасету

Графік на рисунку 3.5 показує кількість угод по роках. Ми бачимо, що найбільше угод було укладено у 2021 році, що підтверджує дані попередніх графіків.

Отже, використовуючи дані з візуалізацій можна зробити такі висновки:

1. Лідери за кількістю стартапів - Великобританія, Німеччина та Іспанія є лідерами за кількістю компаній, що свідчить про їхню привабливість для інвесторів.

2. Пік інвестицій був у стартапи, які засновані в 2021 році, що може свідчити про зростання інвестиційної активності в цьому році, скоріш за все це пов'язано з розвитком штучного інтелекту.
3. Відсутність даних про фінансування - значна кількість угод не має даних про фінансування, що вказує на необхідність покращення звітності або доступу до даних.
4. Розподіл фінансування - найпоширеніші категорії фінансування включають відносно невеликі суми, що свідчить про те, що більшість компаній отримує фінансування на рівні до 1 мільйона євро.

Ці дані можуть бути корисними для аналізу інвестиційних тенденцій та трендів, а також для прийняття рішень щодо стратегій залучення фінансування в стартапи.

3.2 Класифікація стартапів через LLM та статистичний огляд результатів

У цьому розділі ми застосували метод класифікації стартапів за допомогою великих мовних моделей (LLM), зокрема GPT-3.5. Для цього використано датасет із Kaggle під назвою "Indian Startup Funding". Після попереднього очищення даних і заповнення пропущених значень ми приступили до класифікації стартапів за їх секторами та індустріями.

Процес класифікації включав розробку спеціальної класифікаційної рамки (Табл. 3.2), що охоплює різноманітні сектори, до яких належать стартапи, такі як технології, електронна комерція, охорона здоров'я, сільське господарство, виробництво, енергетика та стійкість, медіа та розваги, а також фінансові послуги. Кожен сектор включає кілька підкатегорій, що дозволяє більш детально та якісно класифікувати стартапи.

Таблиця 3.2. Класифікаційна рамка

Категорія	Сектори
Technology	SaaS (Software as a Service), AI/ML (Artificial Intelligence & Machine Learning), FinTech (Financial Technology), Blockchain, EdTech (Educational Technology), HealthTech (Healthcare Technology), Cybersecurity
E-commerce	B2C (Business-to-Consumer), B2B (Business-to-Business), D2C (Direct-to-Consumer), Marketplaces, Logistics & Supply Chain, Hyperlocal Delivery
Healthcare	Telehealth, Diagnostics, Pharmaceuticals, Medical Devices, Biotechnology, Fitness & Wellness
Agriculture	AgriTech, Farm-to-Consumer Platforms, Agro-Equipment, Agri-Inputs, Precision Agriculture, Supply Chain
Manufacturing	Smart Manufacturing, Industrial Automation, 3D Printing, Materials Innovation, Supply Chain Optimization
Energy & Sustainability	Renewable Energy, Electric Vehicles, Carbon Capture & Storage, Waste Management, Water Conservation, Green Building
Media & Entertainment	OTT Platforms, Gaming, Content Production & Distribution, Music Streaming, Advertising & Marketing Technology, Influencer Platforms
Financial Services	Payments & Wallets, Lending & Credit, Investment Platforms, InsuranceTech, Wealth Management

Джерело: класифікаційна рамка розроблена автором

Для реалізації класифікації ми використовували API OpenAI GPT-3.5. Спочатку ми імпортували необхідні бібліотеки та налаштували середовище. Було створено клас GPTClassifier, який містить методи для побудови запитів до моделі, виконання класифікації та обробки даних. Основний метод `classify_dataframe` застосовує класифікацію до всіх записів у датасеті. Запит до LLM наведено на рисунку 3.11:

```
def build_prompt(self, description, single_label):
    if single_label:
        return (
            "Here's the classification framework for categorizing organizations: "
            + str(self.sectors)
            + ". Provide ONLY ONE CATEGORY from the list. Using this framework, "
            + "classify the organization based on its description by sectors. "
            + "Respond using a python-like list containing only ONE SECTOR. "
            + "Organization description:"
        )
    else:
        return (
            "Here's the classification framework for categorizing organizations: "
            + str(self.sectors)
            + ". Classification can be multi-labeled. Using this framework, "
            + "classify the organization based on its description by sectors. "
            + "Respond using a python-like list of sectors. Organization description:"
        )
```

Рисунок 3.6. Два варіанти запитів до LLM

Джерело: код написаний автором

Ключовим моментом було побудова запиту до моделі, який включав опис класифікаційної рамки та інструкції для моделі щодо того, як класифікувати кожен стартап на основі його опису. Класифікація могла бути як односекторною, так і багатосекторною, залежно від специфіки стартапу. Далі наведено аналіз даних на основі класифікацій, що були отримані в результаті виконання розробленого пз:

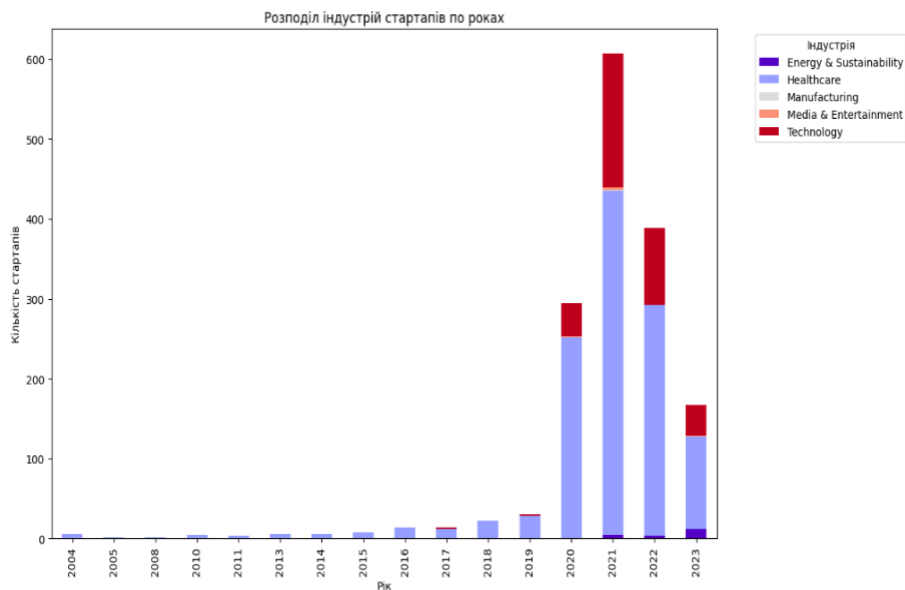


Рисунок 3.7. Розподіл індустрій стартапів по роках

Джерело: візуалізація на основі проведеної класифікації

Графік на рисунку 3.7 показує, як розвивалися різні індустрії стартапів по роках. Найбільший сплеск стартапів у сфері охорони здоров'я (Healthcare) спостерігається у 2021 році. Інші індустрії, такі як технології (Technology) та медіа і розваги (Media & Entertainment), також показують зростання у той же період.

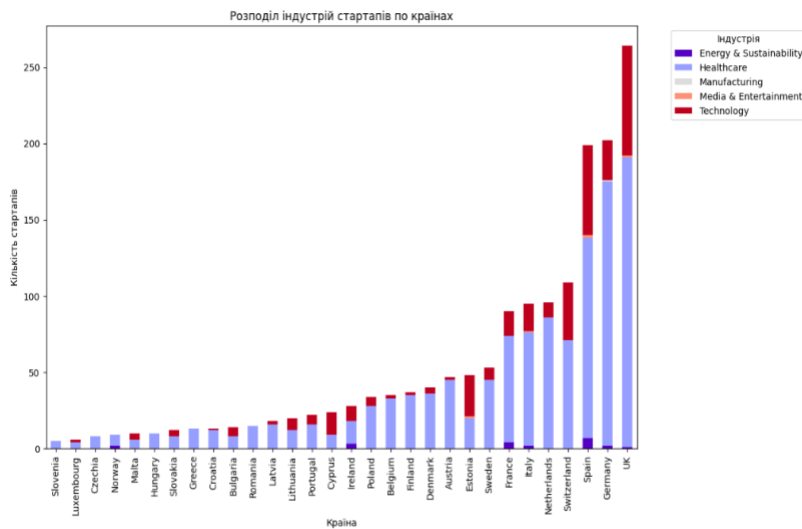


Рисунок 3.8. Розподіл індустрій стартапів по країнах

Джерело: візуалізація на основі проведеної класифікації

Графік на рисунку 3.8 відображає розподіл індустрій стартапів по різних країнах. Великобританія, Німеччина та Іспанія мають найбільшу кількість стартапів, переважно у сферах охорони здоров'я (Healthcare) та технологій (Technology). Це свідчить про їхню привабливість для стартапів та інвесторів.

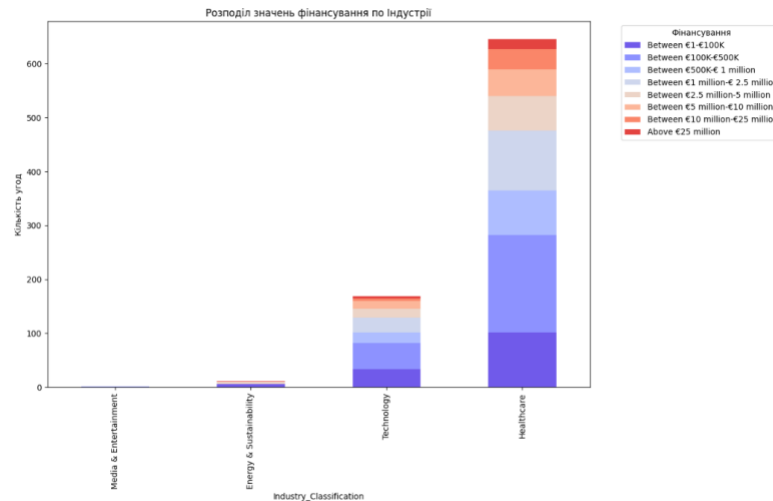


Рисунок 3.9. Розподіл значень фінансування по індустрії

Джерело: візуалізація на основі проведеної класифікації

Графік на рисунку 3.9 показує розподіл значень фінансування по різних індустріях. Найбільше фінансування отримують стартапи в сфері охорони здоров'я (Healthcare), що вказує на значний інтерес інвесторів до цієї галузі.

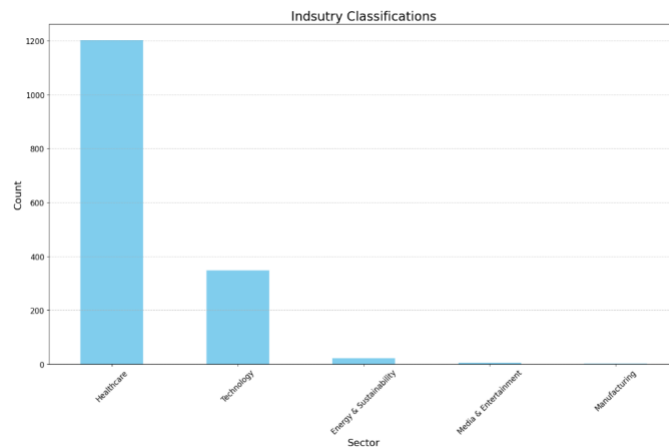


Рисунок 3.10. Кількість стартапів по індустріях

Джерело: візуалізація на основі проведеної класифікації

Графік на рисунку 3.10 показує кількість стартапів у різних індустріях. Охорона здоров'я (Healthcare) домінує, що підтверджує високий попит на стартапи у цій сфері.

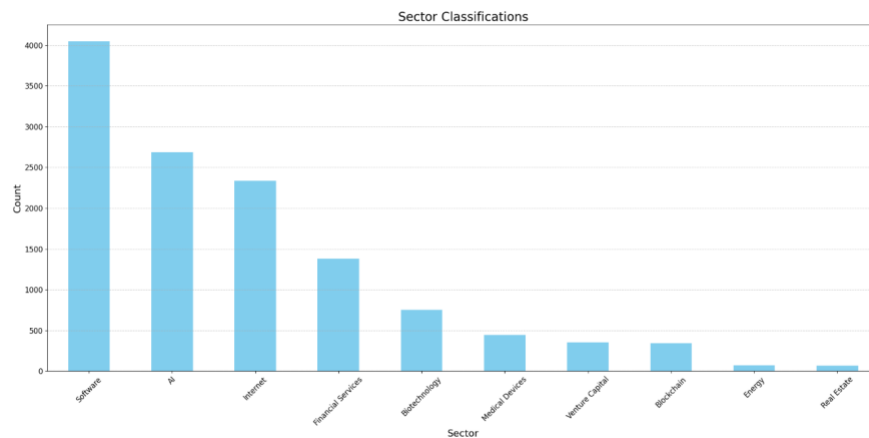


Рисунок 3.11. Кількість стартапів по секторах

Джерело: візуалізація на основі проведеної класифікації

Графік на рисунку 3.11 відображає кількість стартапів у різних секторах. Програмне забезпечення є провідною сферою, за ним йдуть AI та Інтернет, що свідчить про зростання інтересу до цих технологій.

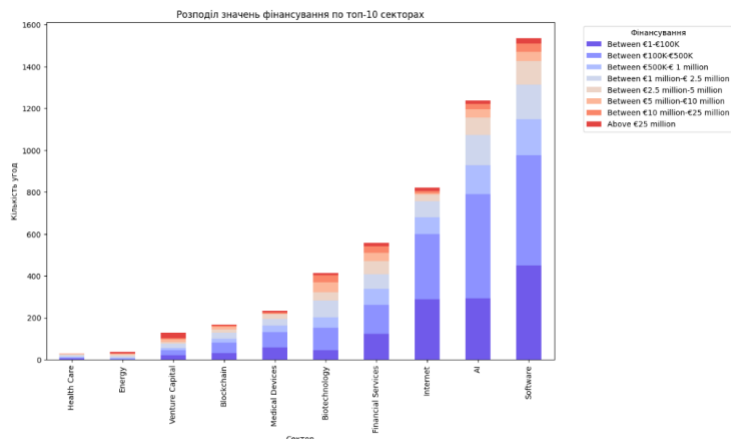


Рисунок 3.12. Розподіл значень фінансування по топ-10 секторах
Джерело: візуалізація на основі проведеної класифікації

Графік на рисунку 3.12 показує, як різні сектори отримують фінансування. Найбільше фінансування спрямоване на програмне забезпечення (Software), штучний інтелект (AI) та фінансові послуги (Financial Services). Це вказує на популярність і перспективність цих секторів серед інвесторів.

Отже, аналіз графіків підкреслює зростання сектору охорони здоров'я, що досягло піку у 2021 році, та продемонструвало значний інвестиційний інтерес. Стартапи у сферах технологій та охорони здоров'я зосереджені переважно у Великобританії, Німеччині та Іспанії, що вказує на їх високу привабливість для бізнесу та інвесторів. Розподіл фінансування підтверджує значний інтерес до секторів програмного забезпечення, штучного інтелекту та фінансових послуг, роблячи їх лідерами за обсягами інвестицій.

3.3 Українські стартапи та фінансування

Українські стартапи стикаються з численними викликами, коли йдеться про залучення іноземного фінансування. Однією з головних проблем є відсутність довіри з боку іноземних інвесторів до українського ринку, що зумовлює певну

упередженість до стартапів з України. Ця упередженість сприяє тому, що багато українських стартапів реєструються в інших країнах або видають себе за стартапи з інших країн для того, щоб отримати доступ до міжнародного фінансування.

Згідно з дослідженням Crunchbase, багато українських стартапів, незважаючи на те, що більшість їх співробітників знаходяться в Україні, реєструють свої штаб-квартири в таких містах, як Нью-Йорк, Лондон чи Сан-Франциско. Це явище не є новим, проте після початку повномасштабного вторгнення Росії, така практика стала ще більш поширеною. Основною причиною цього є бажання підвищити шанси на залучення інвестицій та уникнути ризиків, пов'язаних з політичною та економічною нестабільністю в Україні. [10] [11]

Дослідження [12] [13] також показує, що значна частина українських стартапів отримує фінансування від іноземних фондів лише після реєстрації в інших країнах. Наприклад, значна кількість українських стартапів переїхала до Польщі, де вони мали можливість отримати фінансування від місцевих та міжнародних інвесторів. Це підкреслює тенденцію до переміщення бізнесу за кордон для підвищення його інвестиційної привабливості.

Окрім цього, багато українських стартапів отримують фінансування через спеціальні програми підтримки від міжнародних організацій та компаній. Так, Європейська комісія запустила фонд у розмірі 20 мільйонів євро для підтримки українських стартапів, а компанії, такі як Google та AWS, пропонують гранти та безкоштовне використання своїх сервісів для українських підприємців. [14]

Ці заходи підтримки дозволяють українським стартапам продовжувати свою діяльність, незважаючи на складні умови. Однак, для повного розвитку та масштабування, багато стартапів все ж змушені шукати фінансування за кордоном, часто реєструючись у країнах з більш стабільною економічною та політичною ситуацією.

3.4 Американські стартапи та фінансування

Американські стартапи продовжують залишатися одним із ключових рушіїв інновацій і економічного зростання. Однією з головних особливостей стартап-сцени США є різноманітність джерел фінансування, доступних для підприємців на різних етапах розвитку бізнесу. Останніми роками значна частина інвестицій у США спрямовується на стартапи, що працюють у галузі штучного інтелекту (ШІ). У 2023 році понад 25% всіх інвестицій у стартапи в США були спрямовані на компанії, пов'язані з ШІ, що більш ніж удвічі перевищує рівень інвестицій у попередні роки. Це свідчить про значний інтерес інвесторів до технологій, які обіцяють трансформувати різні галузі економіки, від фінтеху до біотехнологій [15] [16].

Іншою важливою тенденцією є збільшення використання боргового фінансування серед стартапів. Хоча залучення боргового капіталу може бути привабливим через його відносну доступність, це також може становити ризик для компаній, які не досягають прогнозованих темпів зростання. Зокрема, такі великі компанії, як Uber та Airbnb, залучили значні суми боргового капіталу, але ці приклади також показують потенційні загрози, якщо борг не вдасться погасити вчасно [17].

Фінансова структура стартапів у США змінюється з часом. На початкових етапах підприємці зазвичай покладаються на власні заощадження, допомогу від родини та друзів, а також на особисті кредити. У міру зростання бізнесу та зменшення його непрозорості, значно зростає частка бізнесових кредитів та фінансування через торговий кредит. Стартапи з високою науково-дослідною активністю та інтелектуальною власністю найчастіше отримують фінансування від венчурних капіталістів та бізнесових кредитів [18].

Seed-фінансування в США залишилося відносно стабільним, незважаючи на загальне скорочення інвестицій у стартапи останніми роками. У 2022 році обсяг інвестицій на ранніх стадіях зріс на 10%, тоді як у 2023 році він знизився на 31%, що все ж було меншим зниженням, ніж на інших етапах фінансування [19]. Це свідчить про те, що навіть в умовах економічної нестабільності, інвестори продовжують підтримувати нові проекти, які мають великий потенціал для зростання.

Венчурні капіталісти та ангельські інвестори грають важливу роль у фінансуванні стартапів на різних етапах їх розвитку. Ангельські інвестори часто надають початковий капітал для стартапів на стадії seed-фінансування, тоді як венчурні капіталісти зазвичай інвестують на пізніших стадіях, таких як серія А, коли бізнес вже продемонстрував свою життєздатність і має певний ринковий успіх [20] [21].

Фінансування серії В у США також продовжує зростати. У 2021 році середній раунд серії В склав 45 мільйонів доларів, що на 50% більше порівняно з 2020 роком. Це свідчить про те, що стартапи на цьому етапі вже мають доведену технологію або ринкову тягу і готові до масштабування [22].

Таким чином, американські стартапи мають доступ до широкого спектру джерел фінансування, що дозволяє їм зростати та розвиватися. Однак це також вимагає від підприємців ретельного планування та управління фінансами, щоб уникнути потенційних ризиків.

Висновки до розділу та рекомендації для стартапів

На основі зібраних і аналізованих даних будл досліджено ключові тренди у фінансуванні стартапів у Європі, було виявлено лідерів за кількістю залучених інвестицій та ідентифікували галузі з найбільшим потенціалом зростання. Аналіз підтвердив значний інтерес до секторів охорони здоров'я та технологій, особливо у

таких країнах, як Великобританія, Німеччина та Іспанія. Стартапи, що працюють у сфері програмного забезпечення, штучного інтелекту, та фінансових послуг, продемонстрували найвищі рівні залучення капіталу.

Рекомендації для стартапів:

1. Фокус на галузі з високим потенціалом. Розглядайте можливість розвитку у секторах охорони здоров'я та технологій, оскільки ці галузі приваблюють значну увагу інвесторів і мають тенденцію до швидкого росту.
2. Розширення географії діяльності. Оскільки країни, такі як Великобританія, Німеччина та Іспанія, показали високу активність у фінансуванні стартапів, важливо розглядати ці ринки для розширення або реєстрації діяльності.
3. Покращення доступу до фінансових даних. Оскільки значна частина стартапів не має чіткої інформації про фінансування, важливо покращити звітність та доступ до фінансових даних, щоб підвищити прозорість для потенційних інвесторів.
4. Адаптація до нових реалій. Враховуючи пандемічні та геополітичні зміни, стартапи повинні бути гнучкими та готовими адаптувати свої стратегії залучення капіталу та ведення бізнесу залежно від змін у світовій економіці та інвестиційному кліматі.

Ці рекомендації можуть допомогти стартапам ефективніше залучати інвестиції та оптимізувати свої стратегії розвитку у швидко змінюваних умовах ринку.

ВИСНОВКИ

У даній роботі ми провели всебічний аналіз факторів успіху ІТ-стартапів, заснований на огляді наукових досліджень, детальному вивченні ключових показників фінансування, а також оцінці внутрішніх і зовнішніх умов, що впливають на діяльність стартапів у Європі та зокрема в Україні.

Технологічні інновації та ефективні бізнес-моделі виявились ключовими факторами, що визначають успіх ІТ-стартапів. Ці елементи дозволяють компаніям адаптуватися до швидко змінних ринкових умов і виходити на передові позиції в конкурентній боротьбі. Венчурне фінансування та ринкові умови також істотно впливають на успішність стартапів, підтримуючи їхній швидкий розвиток і масштабування.

Результати аналізу показали, що європейські стартапи, особливо у галузях охорони здоров'я та технологій, продовжують приваблювати значні інвестиції. Це вказує на великий інтерес інвесторів і великий потенціал для подальшого розвитку. Однак українські стартапи стикаються з викликами на міжнародній арені, зокрема з проблемами довіри з боку іноземних інвесторів і політичною нестабільністю, що спонукає багатьох до реєстрації за кордоном для покращення своїх шансів на залучення фінансування.

На підставі проведеного дослідження, рекомендується стартапам активізувати пошук венчурного фінансування і максимально використовувати наявні ринкові можливості. Також важливим є зосередження на інноваціях та розробці унікальних бізнес-моделей, що допоможе їм виокремитися серед конкурентів і забезпечити стале зростання.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. T. Przedzinski. Methodology for development of scientific software and test frameworks in function of precision of the expected results. ArXiv. URL: <https://arxiv.org/abs/2203.11650> (date of access: 24.03.2024).
2. Carmine Giardino, Michael Unterkalmsteiner, Nicolò Paternoster, Tony Gorschek, Pekka Abrahamsson. What do we know about software development in startups?. ArXiv. URL: <https://arxiv.org/abs/2307.13707v1> (date of access: 24.03.2024).
3. Aurelio Patelli, Giulio Cimini, Emanuele Pugliese, Andrea Gabrielli. The scientific influence of nations on global scientific and technological development. ArXiv. URL: <https://arxiv.org/abs/1704.03768> (date of access: 24.03.2024).
4. Maximillien de Bayser, Vinicius Segura, Leonardo Guerreiro Azevedo, Leonardo P. Tizzei, Raphael Melo Thiago, Elton Soares, Renato Cerqueira. DevOps and Microservices in Scientific System development. ArXiv. URL: <https://arxiv.org/abs/2112.12049> (date of access: 24.03.2024).
5. Rene Carmona, Xiuneng Zhu. A Probabilistic Approach to Mean Field Games with Major and Minor Players. ArXiv. URL: <https://arxiv.org/abs/1409.7141> (date of access: 24.03.2024).
6. Yan Ma, Minyi Huang. Linear quadratic mean field games with a major player: The multi-scale approach. ArXiv. URL: <https://arxiv.org/abs/1903.08780v2> (date of access: 24.03.2024).
7. Minyi Huang. Linear-Quadratic Mean Field Games with a Major Player: Nash certainty equivalence versus master equations. ArXiv. URL: <https://arxiv.org/abs/2009.13050> (date of access: 24.03.2024).
8. Ian Bolliger, Tamma Carleton, Solomon Hsiang, Jonathan Kadish, Jonathan Proctor, Benjamin Recht, Esther Rolf, Vaishaal Shankar. Ground Control to Major

- Tom: the importance of field surveys in remotely sensed data analysis. ArXiv. URL: <https://arxiv.org/abs/1710.09342v1> (date of access: 24.03.2024).
9. Welin Chen, David Grangier, Michael Auli. Strategies for Training Large Vocabulary Neural Language Models. ArXiv. URL: <https://arxiv.org/abs/1512.04906> (date of access: 24.03.2024).
 10. Joanna Glasner. Capital Still Flows To Ukrainian-Staffed Startups, Though Founders Are Mostly Expats. March 18, 2024. Crunchbase. URL: <https://news.crunchbase.com/venture/ukraine-tech-startups-vc-funding-founders-amid-war/> (date of access: 24.03.2024).
 11. Nataly Veremeeva. The voice of Ukrainian start-ups: More than half of start-ups continue their operations exclusively from Ukraine. August 16, 2022. Tech Ukraine. URL: <https://techukraine.org/2022/08/16/the-voice-of-ukrainian-start-ups-report/> (date of access: 24.03.2024).
 12. Maia Yarova. Financing Ukrainian startups during the war: where to look for investments. 19 November, 2022. AIN.Capital. URL: <https://ain.capital/2022/11/19/strongfinancing-ukrainian-startups-during-the-war/> (date of access: 24.03.2024).
 13. Mariia Denysiuk. Ukrainian startups in Poland: challenges, state of funding, and plans — research. 01 February, 2024. AIN.Capital. URL: <https://ain.capital/2024/02/01/ukrainian-startups-in-poland-research/> (date of access: 24.03.2024).
 14. Patricia Allen. Ukrainian startups in Poland: challenges, state of funding, and plans — research. July 18, 2022. EU-Startups. URL: <https://www.eu-startups.com/2022/07/supporting-ukraines-startup-ecosystem-an-inside-look-into-the-eics-e20-million-fund/> (date of access: 24.05.2024).

15. Joanna Glasner. AI's Share Of US Startup Funding Doubled In 2023. August 29, 2023. Crunchbase News. URL: <https://news.crunchbase.com/ai-robotics/us-startup-funding-doubled-openai-anthropic-2023/> (date of access: 24.03.2024).
16. Alison Zeller. Top 10 Startup Trends (2024 & 2025). May 27, 2024. Exploding Topics. URL: <https://explodingtopics.com/blog/startup-trends> (date of access: 24.05.2024)
17. Mikey Tom. Startups increasingly turning to debt financing despite dangers. March 28, 2017. Pitch Book. URL: <https://pitchbook.com/news/articles/startups-increasingly-turning-to-debt-financing-despite-dangers> (date of access: 24.05.2024)
18. Carmen Cotei, Joseph Farhat. The Evolution of Financing Structure in U.S. Startups. February 17, 2017. Pepperdine Digital Commons. URL: <https://digitalcommons.pepperdine.edu/jef/vol19/iss1/4/> (date of access: 24.05.2024)
19. Gené Teare. US Seed Investment Actually Held Up Pretty Well For The Past 2 Years. Here's What That Means For 2024. January 24, 2024. Crunchbase News. URL: <https://news.crunchbase.com/seed/us-investment-held-up-forecast-2024/> (date of access: 24.05.2024)
20. Jason Gordon. Series or Rounds of Financing (Startups) - Explained. April 20th, 2024. The Business Professor. URL: https://thebusinessprofessor.com/en_US/business-transactions/series-or-rounds-of-financing-startups-definition (date of access: 24.05.2024)
21. Gené Teare. Lower Valuations, Higher Bar: What It's Like To Raise A Seed Round In 2024. January 25, 2024. Crunchbase News. URL:

<https://news.crunchbase.com/seed/us-startup-valuations-lower-2024/> (date of access: 24.05.2024)

22. Joanna Glasner. What Current Series B Trends Say About The US Startup Scene. March 10, 2022. Crunchbase News. URL: <https://news.crunchbase.com/business/series-b-startup-vc-fundraising-deal-trends/> (date of access: 24.05.2024)

ДОДАТКИ

ДОДАТОК А.

Збір, обробка та класифікація даних

```
# %% [code] {"execution":{"iopub.status.busy":"2024-05-14T18:59:21.892588Z","iopub.execute_input":"2024-05-14T18:59:21.892985Z","iopub.status.idle":"2024-05-14T18:59:21.922278Z","shell.execute_reply.started":"2024-05-14T18:59:21.892954Z","shell.execute_reply":"2024-05-14T18:59:21.921257Z"}}
```

```
import pandas as pd
```

```
df = pd.read_csv('/kaggle/input/indian-startup-funding/startup_funding.csv')
```

```
df
```

```
# %% [code] {"execution":{"iopub.status.busy":"2024-05-14T18:59:21.923778Z","iopub.execute_input":"2024-05-14T18:59:21.924094Z","iopub.status.idle":"2024-05-14T18:59:23.600017Z","shell.execute_reply.started":"2024-05-14T18:59:21.924057Z","shell.execute_reply":"2024-05-14T18:59:23.598976Z"}}
```

```
import pandas as pd
```

```
import matplotlib.pyplot as plt
```

```
import seaborn as sns
```

```
# 1. Understanding Structure
```

```
print(df.head()) # Check the first few rows
```

```
print(df.info()) # Get column data types and missing values
```

```
# 2. Descriptive Statistics
```

```
# Numeric columns
```

```

numeric_stats = df.describe()
print("Numeric columns summary:")
print(numeric_stats)

# Categorical columns
categorical_cols = df.select_dtypes(include=['object']).columns
for col in categorical_cols:
    print(f"Unique values in '{col}': {df[col].nunique()}")

# 3. Visual Analysis
# Missing values heatmap
plt.figure(figsize=(12, 8))
sns.heatmap(df.isnull(), cbar=False, cmap='viridis', yticklabels=False)
plt.title('Missing Values Heatmap')
plt.show()

# Investment Amount Distribution
df['Amount in USD'] = pd.to_numeric(df['Amount in USD'], errors='coerce')
plt.figure(figsize=(10, 6))
sns.histplot(df['Amount in USD'].dropna(), kde=True, bins=30)
plt.title('Investment Amount Distribution')
plt.xlabel('Amount in USD')
plt.ylabel('Frequency')
plt.show()

# Investment Type Distribution
plt.figure(figsize=(10, 6))
sns.countplot(y='InvestmentnType', data=df, order=df['InvestmentnType'].value_counts().index)
plt.title('Investment Type Distribution')
plt.xlabel('Count')

```

```
plt.ylabel('Investment Type')
```

```
plt.show()
```

```
# City-wise Investment Analysis
```

```
top_cities = df['City Location'].value_counts().head(10)
```

```
plt.figure(figsize=(10, 6))
```

```
sns.barplot(x=top_cities.values, y=top_cities.index)
```

```
plt.title('Top 10 Cities by Number of Investments')
```

```
plt.xlabel('Number of Investments')
```

```
plt.ylabel('City Location')
```

```
plt.show()
```

```
# %% [code] {"execution":{"iopub.status.busy":"2024-05-14T18:59:23.601638Z","iopub.execute_input":"2024-05-14T18:59:23.602015Z","iopub.status.idle":"2024-05-14T18:59:23.607239Z","shell.execute_reply.started":"2024-05-14T18:59:23.601983Z","shell.execute_reply":"2024-05-14T18:59:23.606140Z"}}

print(df.columns)
```

```
# %% [code] {"execution":{"iopub.status.busy":"2024-05-14T18:59:23.608685Z","iopub.execute_input":"2024-05-14T18:59:23.609422Z","iopub.status.idle":"2024-05-14T18:59:23.629855Z","shell.execute_reply.started":"2024-05-14T18:59:23.609396Z","shell.execute_reply":"2024-05-14T18:59:23.628931Z"}}

# Identify columns with categorical data types
```

```
categorical_cols = df.select_dtypes(include=['object']).columns
```

```
# Loop through each categorical column to print value counts
```

```
for col in categorical_cols:
```

```
    value_counts = df[col].value_counts(dropna=False)
```

```
    print(f"Value counts for '{col}':")
```

```
    print(value_counts)
```

```
print("\n")
```

```
# %% [markdown]
```

```
# # Classify
```

```
# %% [code] {"execution":{"iopub.status.busy":"2024-05-14T18:59:23.631916Z","iopub.execute_input":"2024-05-14T18:59:23.632222Z","iopub.status.idle":"2024-05-14T18:59:34.957330Z","shell.execute_reply.started":"2024-05-14T18:59:23.632198Z","shell.execute_reply":"2024-05-14T18:59:34.956282Z"}}
```

```
!pip install "openai<1.0.0"
```

```
# %% [code] {"execution":{"iopub.status.busy":"2024-05-14T18:59:34.958944Z","iopub.execute_input":"2024-05-14T18:59:34.959303Z","iopub.status.idle":"2024-05-14T18:59:46.346226Z","shell.execute_reply.started":"2024-05-14T18:59:34.959263Z","shell.execute_reply":"2024-05-14T18:59:46.344983Z"}}
```

```
pip install ratelimit
```

```
# %% [code] {"execution":{"iopub.status.busy":"2024-05-14T18:59:46.348343Z","iopub.execute_input":"2024-05-14T18:59:46.348688Z","iopub.status.idle":"2024-05-14T18:59:46.354235Z","shell.execute_reply.started":"2024-05-14T18:59:46.348657Z","shell.execute_reply":"2024-05-14T18:59:46.353188Z"}}
```

```
import pandas as pd
```

```
import numpy as np
```

```
#from tqdm.auto import tqdm
```

```
import ast
```

```
import json
```

```
import openai
```

```
from tqdm import tqdm
```

```
#           %%           [code]           {"execution":{"iopub.status.busy":"2024-05-14T18:59:46.355432Z","iopub.execute_input":"2024-05-14T18:59:46.355709Z","iopub.status.idle":"2024-05-14T18:59:46.365916Z","shell.execute_reply.started":"2024-05-14T18:59:46.355686Z","shell.execute_reply":"2024-05-14T18:59:46.364888Z"}}
```

```
framework = {  
  "Technology": [  
    "SaaS (Software as a Service)",  
    "AI/ML (Artificial Intelligence & Machine Learning)",  
    "FinTech (Financial Technology)",  
    "Blockchain",  
    "EdTech (Educational Technology)",  
    "HealthTech (Healthcare Technology)",  
    "Cybersecurity"  
  ],  
  "E-commerce": [  
    "B2C (Business-to-Consumer)",  
    "B2B (Business-to-Business)",  
    "D2C (Direct-to-Consumer)",  
    "Marketplaces",  
    "Logistics & Supply Chain",  
    "Hyperlocal Delivery"  
  ],  
  "Healthcare": [  
    "Telehealth",  
    "Diagnostics",  
    "Pharmaceuticals",  
    "Medical Devices",  
    "Biotechnology",  
    "Fitness & Wellness"  
  ],  
}
```

"Agriculture": [

- "AgriTech",
- "Farm-to-Consumer Platforms",
- "Agro-Equipment",
- "Agri-Inputs",
- "Precision Agriculture",
- "Supply Chain"

],

"Manufacturing": [

- "Smart Manufacturing",
- "Industrial Automation",
- "3D Printing",
- "Materials Innovation",
- "Supply Chain Optimization"

],

"Energy & Sustainability": [

- "Renewable Energy",
- "Electric Vehicles",
- "Carbon Capture & Storage",
- "Waste Management",
- "Water Conservation",
- "Green Building"

],

"Media & Entertainment": [

- "OTT Platforms",
- "Gaming",
- "Content Production & Distribution",
- "Music Streaming",
- "Advertising & Marketing Technology",
- "Influencer Platforms"

```
],  
  "Financial Services": [  
    "Payments & Wallets",  
    "Lending & Credit",  
    "Investment Platforms",  
    "InsuranceTech",  
    "Wealth Management"  
  ]  
}
```

```
#           %%           [code]           {"execution":{"iopub.status.busy":"2024-05-  
14T18:59:46.369211Z","iopub.execute_input":"2024-05-  
14T18:59:46.369497Z","iopub.status.idle":"2024-05-  
14T18:59:46.386144Z","shell.execute_reply.started":"2024-05-  
14T18:59:46.369466Z","shell.execute_reply":"2024-05-14T18:59:46.385069Z"}}
```

```
import requests
```

```
import json
```

```
import numpy as np
```

```
from tqdm import tqdm
```

```
class GPTClassifier:
```

```
    def __init__(self, framework, openai_api_key):
```

```
        self.framework = framework # format: {industry1: [sector1, sector2,...],...}
```

```
        self.sectors = [sector for sectors in self.framework.values() for sector in sectors]
```

```
        self.api_key = openai_api_key
```

```
        self.headers = {
```

```
            "Authorization": f"Bearer {self.api_key}",
```

```
            "Content-Type": "application/json"
```

```
        }
```

```
        self.api_url = "https://api.openai.com/v1/chat/completions"
```

```

def classify(self, description, single_label):
    prompt = self.build_prompt(description, single_label)
    payload = {
        "model": "gpt-3.5-turbo",
        "messages": [
            {"role": "system", "content": prompt},
            {"role": "user", "content": description}
        ],
        "temperature": 0.4
    }
    try:
        response = requests.post(self.api_url, headers=self.headers, json=payload)
        response.raise_for_status() # Raises an HTTPError for bad responses
        text_response = response.json()['choices'][0]['message']['content']
        return text_response
    except requests.RequestException as e:
        print(f"Error occurred: {str(e)}")
        print('description = ', description)
        return np.NaN

```

```

def build_prompt(self, description, single_label):
    if single_label:
        return (
            "Here's the classification framework for categorizing organizations: "
            + str(self.sectors)
            + ". Provide ONLY ONE CATEGORY from the list. Using this framework, "
            + "classify the organization based on its description by sectors. "
            + "Respond using a python-like list containing only ONE SECTOR. "
            + "Organization description:"

```

```

    )
else:
    return (
        "Here's the classification framework for categorizing organizations: "
        + str(self.sectors)
        + ". Classification can be multi-labeled. Using this framework, "
        + "classify the organization based on its description by sectors. "
        + "Respond using a python-like list of sectors. Organization description:"
    )

```

```

def classify_dataframe(self, df, output_dir='./', batch_size=1000, use_batches=False,
single_label=True):
    tqdm.pandas()
    df['Sector_Classification'] = np.NaN
    df['Industry_Classification'] = np.NaN
    if not use_batches:
        df['Sector_Classification'] = df['technical_description'].progress_apply(lambda x:
self.classify(str(x), single_label))
        df = self.map_sectors_to_industries(df)
    else:
        n_rows = df.shape[0]
        start_idx = 0
        while start_idx < n_rows:
            end_idx = start_idx + batch_size
            batch = df.iloc[start_idx:end_idx]
            batch['Sector_Classification'] = batch['technical_description'].progress_apply(lambda x:
self.classify(str(x), single_label))
            batch = self.map_sectors_to_industries(batch)
            df.iloc[start_idx:end_idx] = batch
            batch.to_csv(f"{output_dir}{start_idx}-{end_idx}_batch.csv")
            print('processed:', end_idx)

```

```

        start_idx += batch_size

    df.to_csv(f"{output_dir}classified_dataset.csv")

    return df

def map_sectors_to_industries(self, df):

    sector_to_industry = {sector: industry for industry, sector_list in self.framework.items() for
sector in sector_list}

    df['Industry_Classification'] = df['Sector_Classification'].apply(lambda x:
[sector_to_industry.get(i) for i in x if i in sector_to_industry])

    return df

# %% [code] {"execution":{"iopub.status.busy":"2024-05-
14T18:59:46.387242Z","iopub.execute_input":"2024-05-
14T18:59:46.387547Z","iopub.status.idle":"2024-05-
14T18:59:46.396852Z","shell.execute_reply.started":"2024-05-
14T18:59:46.387523Z","shell.execute_reply":"2024-05-14T18:59:46.396100Z"}}

import os

path = '/kaggle/input/indian-startup-funding/startup_funding.csv'

if os.path.exists(path):

    df = pd.read_csv(path)

else:

    print(f"File not found at path: {path}")

# description_technical - for classification, description - for frontend, use this block only if there is
no description_technical column

df['Combined_Info'] = df.apply(lambda row: f"{row['Startup Name']} - {row['Industry Vertical']} -
{row['SubVertical']}", axis=1)

df['technical_description'] = df['Combined_Info']

```

```

openai_api_key = 'API_KEY' # Load from a secure place

classifier = GPTClassifier(backend, openai_api_key)

classifier.classify_dataframe(df, batch_size=100, output_dir='/kaggle/working/classified_final_')

# %% [code] {"execution":{"iopub.status.busy":"2024-05-14T18:59:46.400545Z","iopub.execute_input":"2024-05-14T18:59:46.400800Z","iopub.status.idle":"2024-05-14T18:59:46.447829Z","shell.execute_reply.started":"2024-05-14T18:59:46.400779Z","shell.execute_reply":"2024-05-14T18:59:46.446916Z"}}
classified_df = pd.read_csv('/kaggle/input/indian-startups-classified/classified_final_classified_dataset.csv')

# Assuming classified_df is your DataFrame
classified_df['Sector_Classification'] = classified_df['Sector_Classification'].str[2:-2]

industry_mapping = {}

for industry, sectors in framework.items():
    for sector in sectors:
        industry_mapping[sector] = industry

classified_df['Industry_Classification'] = classified_df['Sector_Classification'].map(industry_mapping)

classified_df

# %% [code] {"execution":{"iopub.status.busy":"2024-05-14T18:59:46.448822Z","iopub.execute_input":"2024-05-14T18:59:46.449185Z","iopub.status.idle":"2024-05-14T18:59:46.449185Z","shell.execute_reply.started":"2024-05-14T18:59:46.449185Z","shell.execute_reply":"2024-05-14T18:59:46.449185Z"}}

```

```
14T18:59:46.458739Z", "shell.execute_reply.started": "2024-05-14T18:59:46.449162Z", "shell.execute_reply": "2024-05-14T18:59:46.457337Z"}}}
```

```
import pandas as pd
```

```
# Assuming classified_df is your DataFrame and framework is the provided framework dictionary
```

```
def filter_and_map_sector(sector):
```

```
    for industry, sectors in framework.items():
```

```
        if sector in sectors:
```

```
            return sector # Return the first matching sector
```

```
    return None # Return None if no matching industry is found
```

```
classified_df['Sector_Classification'] = classified_df['Sector_Classification'].apply(filter_and_map_sector)
```

```
    #           %%           [code]           {"execution":{"iopub.status.busy":"2024-05-14T18:59:46.460416Z", "iopub.execute_input": "2024-05-14T18:59:46.461226Z", "iopub.status.idle": "2024-05-14T18:59:46.471494Z", "shell.execute_reply.started": "2024-05-14T18:59:46.461191Z", "shell.execute_reply": "2024-05-14T18:59:46.470349Z"}}}
```

```
# Dictionary mapping long names to short names
```

```
short_names_mapping = {
```

```
    'Sector_Classification': {
```

```
        'Marketplaces': 'Marketplaces',
```

```
        'B2C (Business-to-Consumer)': 'B2C',
```

```
        'SaaS (Software as a Service)': 'SaaS',
```

```
        'EdTech (Educational Technology)': 'EdTech',
```

```
        'HealthTech (Healthcare Technology)': 'HealthTech',
```

```
        'D2C (Direct-to-Consumer)': 'D2C',
```

```
        'Hyperlocal Delivery': 'Hyperlocal',
```

```
        'Logistics & Supply Chain': 'Logistics',
```

```
        'AI/ML (Artificial Intelligence & Machine Learning)': 'AI/ML',
```

'Lending & Credit': 'Lending',
'Payments & Wallets': 'Payments',
'Advertising & Marketing Technology': 'AdTech',
'FinTech (Financial Technology)': 'FinTech',
'B2B (Business-to-Business)': 'B2B',
'AgriTech': 'AgriTech',
'Content Production & Distribution': 'Content',
'OTT Platforms': 'OTT',
'Fitness & Wellness': 'Fitness',
'Telehealth': 'Telehealth',
'Gaming': 'Gaming',
'InsuranceTech': 'InsuranceTech',
'Cybersecurity': 'Cybersecurity',
'Investment Platforms': 'Investment',
'Renewable Energy': 'Renewable',
'Wealth Management': 'Wealth',
'Electric Vehicles': 'EV',
'Medical Devices': 'MedDevices',
'Diagnostics': 'Diagnostics',
'Smart Manufacturing': 'SmartManuf',
'Waste Management': 'Waste',
'Industrial Automation': 'Automation',
'Music Streaming': 'Music',
'Pharmaceuticals': 'Pharma',
'Materials Innovation': 'Materials',
'Green Building': 'GreenBuilding',
'Blockchain': 'Blockchain',
'3D Printing': '3DPrinting',
'Supply Chain': 'SupplyChain',
'Biotechnology': 'Biotech',

```
'Water Conservation': 'Water',
'Agro-Equipment': 'AgroEquip',
'Agri-Inputs': 'AgriInputs'
}
}
```

```
# Create a new column with shorter names
```

```
classified_df['Short_Sector_Name']
classified_df['Sector_Classification'].map(short_names_mapping['Sector_Classification'])
```

```
#           %%           [code]           {"execution":{"iopub.status.busy":"2024-05-
14T18:59:46.473541Z","iopub.execute_input":"2024-05-
14T18:59:46.473813Z","iopub.status.idle":"2024-05-
14T18:59:47.166533Z","shell.execute_reply.started":"2024-05-
14T18:59:46.473791Z","shell.execute_reply":"2024-05-14T18:59:47.165463Z"}}}
```

```
import matplotlib.pyplot as plt
```

```
vc = classified_df['Short_Sector_Name'].value_counts()
```

```
plt.figure(figsize=(20, 10))
```

```
vc.plot(kind='bar', color='skyblue')
```

```
plt.title('Sector Classifications', fontsize=20)
```

```
plt.xlabel('Sector', fontsize=16)
```

```
plt.ylabel('Count', fontsize=16)
```

```
plt.xticks(rotation=45, fontsize=12)
```

```
plt.yticks(fontsize=12)
```

```
plt.grid(axis='y', linestyle='--', alpha=0.7)
```

```
plt.tight_layout()
```

```
plt.show()
```

```
# %% [code] {"execution":{"iopub.status.busy":"2024-05-14T18:59:47.167791Z","iopub.execute_input":"2024-05-14T18:59:47.168131Z","iopub.status.idle":"2024-05-14T18:59:47.566132Z","shell.execute_reply.started":"2024-05-14T18:59:47.168097Z","shell.execute_reply":"2024-05-14T18:59:47.565029Z"}}
```

```
import matplotlib.pyplot as plt
```

```
vc = classified_df['Industry_Classification'].value_counts()
```

```
plt.figure(figsize=(15, 10))
```

```
vc.plot(kind='bar', color='skyblue')
```

```
plt.title('Industry Classifications', fontsize=20)
```

```
plt.xlabel('Sector', fontsize=16)
```

```
plt.ylabel('Count', fontsize=16)
```

```
plt.xticks(rotation=45, fontsize=12)
```

```
plt.yticks(fontsize=12)
```

```
plt.grid(axis='y', linestyle='--', alpha=0.7)
```

```
plt.tight_layout()
```

```
plt.show()
```

```
# %% [code]
```

```
import pandas as pd
```

```
import seaborn as sns
```

```
import matplotlib.pyplot as plt
```

```
# Видаляємо рядки з некоректними значеннями у стовпці "Amount in USD"
```

```
classified_df = classified_df.dropna(subset=['Amount in USD'])
```

```
classified_df = classified_df[classified_df['Amount in USD'].str.replace(',', '').str.isnumeric()]
```

```

# Перетворюємо стовпець "Amount in USD" на числовий тип
classified_df['Amount in USD'] = classified_df['Amount in USD'].replace(',', '',
regex=True).astype(float)

# Групуємо дані за індустрією та обчислюємо загальний та середній рівень фінансування
industry_funding_total = classified_df.groupby('Industry_Classification')['Amount in
USD'].sum().reset_index()

industry_funding_avg = classified_df.groupby('Industry_Classification')['Amount in
USD'].mean().reset_index()

# Групуємо дані за сектором та обчислюємо загальний та середній рівень фінансування
sector_funding_total = classified_df.groupby('Sector_Classification')['Amount in
USD'].sum().reset_index()

sector_funding_avg = classified_df.groupby('Sector_Classification')['Amount in
USD'].mean().reset_index()

# Візуалізація загального рівня фінансування за індустрією
plt.figure(figsize=(12, 8))

sns.barplot(x='Amount in USD', y='Industry_Classification',
data=industry_funding_total.sort_values(by='Amount in USD', ascending=False), color='skyblue')

plt.title('Total Funding by Industry')
plt.xlabel('Total Funding (in USD)')
plt.ylabel('Industry')
plt.show()

# Візуалізація середнього рівня фінансування за індустрією
plt.figure(figsize=(12, 8))

sns.barplot(x='Amount in USD', y='Industry_Classification',
data=industry_funding_avg.sort_values(by='Amount in USD', ascending=False), color='skyblue')

plt.title('Average Funding by Industry')
plt.xlabel('Average Funding (in USD)')
plt.ylabel('Industry')
plt.show()

```

```

# Візуалізація загального рівня фінансування за сектором

plt.figure(figsize=(14, 10))

sns.barplot(x='Amount in USD', y='Sector_Classification',
data=sector_funding_total.sort_values(by='Amount in USD', ascending=False), color='skyblue')

plt.title('Total Funding by Sector')
plt.xlabel('Total Funding (in USD)')
plt.ylabel('Sector')
plt.show()

```

```

# Візуалізація середнього рівня фінансування за сектором

plt.figure(figsize=(14, 10))

sns.barplot(x='Amount in USD', y='Sector_Classification',
data=sector_funding_avg.sort_values(by='Amount in USD', ascending=False), color='skyblue')

plt.title('Average Funding by Sector')
plt.xlabel('Average Funding (in USD)')
plt.ylabel('Sector')
plt.show()

```

```

# %% [code] {"execution":{"iopub.status.busy":"2024-05-
13T15:08:30.920241Z","iopub.execute_input":"2024-05-
13T15:08:30.920922Z","iopub.status.idle":"2024-05-
13T15:08:30.947058Z","shell.execute_reply.started":"2024-05-
13T15:08:30.920884Z","shell.execute_reply":"2024-05-13T15:08:30.945905Z"}}

```

```

classified_df = classified_df.dropna(subset=['Industry Vertical', 'SubVertical', 'City Location'])
classified_df

```

```

# %% [markdown]

```

```

# # Predict

```

```

# %% [code]

```

```
classified_df = pd.read_csv('/kaggle/input/indian-startups-  
classified/classified_final_classified_dataset.csv')
```

```
# Assuming classified_df is your DataFrame
```

```
classified_df['Sector_Classification'] = classified_df['Sector_Classification'].str[2:-2]
```

```
industry_mapping = {}
```

```
for industry, sectors in framework.items():
```

```
    for sector in sectors:
```

```
        industry_mapping[sector] = industry
```

```
classified_df['Industry_Classification'] =  
classified_df['Sector_Classification'].map(industry_mapping)
```

```
classified_df
```

```
import pandas as pd
```

```
# Assuming classified_df is your DataFrame and framework is the provided framework dictionary
```

```
def filter_and_map_sector(sector):
```

```
    for industry, sectors in framework.items():
```

```
        if sector in sectors:
```

```
            return sector # Return the first matching sector
```

```
    return None # Return None if no matching industry is found
```

```
classified_df['Sector_Classification'] =  
classified_df['Sector_Classification'].apply(filter_and_map_sector)
```

Dictionary mapping long names to short names

short_names_mapping = {

 'Sector_Classification': {

 'Marketplaces': 'Marketplaces',

 'B2C (Business-to-Consumer)': 'B2C',

 'SaaS (Software as a Service)': 'SaaS',

 'EdTech (Educational Technology)': 'EdTech',

 'HealthTech (Healthcare Technology)': 'HealthTech',

 'D2C (Direct-to-Consumer)': 'D2C',

 'Hyperlocal Delivery': 'Hyperlocal',

 'Logistics & Supply Chain': 'Logistics',

 'AI/ML (Artificial Intelligence & Machine Learning)': 'AI/ML',

 'Lending & Credit': 'Lending',

 'Payments & Wallets': 'Payments',

 'Advertising & Marketing Technology': 'AdTech',

 'FinTech (Financial Technology)': 'FinTech',

 'B2B (Business-to-Business)': 'B2B',

 'AgriTech': 'AgriTech',

 'Content Production & Distribution': 'Content',

 'OTT Platforms': 'OTT',

 'Fitness & Wellness': 'Fitness',

 'Telehealth': 'Telehealth',

 'Gaming': 'Gaming',

 'InsuranceTech': 'InsuranceTech',

 'Cybersecurity': 'Cybersecurity',

 'Investment Platforms': 'Investment',

 'Renewable Energy': 'Renewable',

 'Wealth Management': 'Wealth',

 'Electric Vehicles': 'EV',

 'Medical Devices': 'MedDevices',

```
'Diagnostics': 'Diagnostics',
'Smart Manufacturing': 'SmartManuf',
'Waste Management': 'Waste',
'Industrial Automation': 'Automation',
'Music Streaming': 'Music',
'Pharmaceuticals': 'Pharma',
'Materials Innovation': 'Materials',
'Green Building': 'GreenBuilding',
'Blockchain': 'Blockchain',
'3D Printing': '3DPrinting',
'Supply Chain': 'SupplyChain',
'Biotechnology': 'Biotech',
'Water Conservation': 'Water',
'Agro-Equipment': 'AgroEquip',
'Agri-Inputs': 'AgrilInputs'
}
}
```

```
# Create a new column with shorter names
classified_df['Short_Sector_Name']
classified_df['Sector_Classification'].map(short_names_mapping['Sector_Classification'])
```

```
classified_df = classified_df.dropna(subset=['Industry Vertical', 'SubVertical', 'City Location'])
```

```
# %% [code]
```

```
import pandas as pd
```

```
import numpy as np
```

```
from sklearn.model_selection import train_test_split
```

```
from sklearn.pipeline import Pipeline
```

```
from sklearn.compose import ColumnTransformer
```

```

from sklearn.preprocessing import StandardScaler, OneHotEncoder
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
import matplotlib.pyplot as plt

from sklearn.linear_model import LinearRegression
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestRegressor, GradientBoostingRegressor,
AdaBoostRegressor
from sklearn.neighbors import KNeighborsRegressor
from sklearn.svm import SVR
from sklearn.neural_network import MLPRegressor
from xgboost import XGBRegressor
from lightgbm import LGBMRegressor

# %% [code]

# Convert 'Date dd/mm/yyyy' to datetime, coercing errors
classified_df['Date dd/mm/yyyy'] = pd.to_datetime(classified_df['Date dd/mm/yyyy'],
format='%d/%m/%Y', errors='coerce')

# Drop rows where date conversion failed
classified_df = classified_df.dropna(subset=['Date dd/mm/yyyy'])

# Extract date features
classified_df['Week'] = classified_df['Date dd/mm/yyyy'].dt.isocalendar().week
classified_df['Month'] = classified_df['Date dd/mm/yyyy'].dt.month
classified_df['Year'] = classified_df['Date dd/mm/yyyy'].dt.year
classified_df['Day_of_Week'] = classified_df['Date dd/mm/yyyy'].dt.dayofweek # Monday=0,
Sunday=6

# Replace currency symbols and commas, then convert to numeric, setting errors to NaN

```

```
classified_df['Amount in USD'] = classified_df['Amount in USD'].replace(['\$'], "",
regex=True).replace('undisclosed', np.nan).replace('unknown', np.nan)
```

```
classified_df['Amount in USD'] = pd.to_numeric(classified_df['Amount in USD'].str.replace(',', ""),
errors='coerce')
```

```
# Drop rows with NaN in 'Amount in USD' after conversion
```

```
classified_df = classified_df.dropna(subset=['Amount in USD'])
```

```
# Define features and target
```

```
X = classified_df.drop(['Amount in USD', 'Date dd/mm/yyyy'], axis=1)
```

```
y = classified_df['Amount in USD']
```

```
# Split the data into training and testing sets
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
# Define preprocessing for numerical and categorical data
```

```
numerical_features = ['Week', 'Month', 'Year', 'Day_of_Week']
```

```
categorical_features = ['City Location', 'InvestmentType', 'Sector_Classification',
'Industry_Classification']
```

```
numerical_transformer = StandardScaler()
```

```
categorical_transformer = OneHotEncoder(handle_unknown='ignore')
```

```
preprocessor = ColumnTransformer(
```

```
    transformers=[
```

```
        ('num', numerical_transformer, numerical_features),
```

```
        ('cat', categorical_transformer, categorical_features)
```

```
    ])
```

```
# Create the final pipeline
```

```
pipeline = Pipeline(steps=[('preprocessor', preprocessor)])
```

```
# Fit and transform the training data
X_train_transformed = pipeline.fit_transform(X_train)
X_test_transformed = pipeline.transform(X_test)

# List of models to evaluate
models = {
    "Linear Regression": LinearRegression(),
    "Decision Tree": DecisionTreeRegressor(),
    "Random Forest": RandomForestRegressor(),
    "Gradient Boosting": GradientBoostingRegressor(),
    "AdaBoost": AdaBoostRegressor(),
    "KNN": KNeighborsRegressor(),
    "SVR": SVR(),
    "MLP": MLPRegressor(),
    "XGBoost": XGBRegressor(),
    "LightGBM": LGBMRegressor()
}

# Dictionary to store the results
results = {}

# Train and evaluate each model
for name, model in models.items():
    model.fit(X_train_transformed, y_train)
    y_pred = model.predict(X_test_transformed)
    mae = mean_absolute_error(y_test, y_pred)
    mse = mean_squared_error(y_test, y_pred)
    r2 = r2_score(y_test, y_pred)
    results[name] = {"MAE": mae, "MSE": mse, "R2": r2}
```

```
# Convert the results to a DataFrame for easier visualization
results_df = pd.DataFrame(results).T

# Plot the results
fig, axes = plt.subplots(1, 3, figsize=(18, 5)) # Corrected line
results_df[['MAE']].plot(kind='bar', ax=axes[0], legend=False, title='Mean Absolute Error (MAE)')
results_df[['MSE']].plot(kind='bar', ax=axes[1], legend=False, title='Mean Squared Error (MSE)')
results_df[['R2']].plot(kind='bar', ax=axes[2], legend=False, title='R2 Score')

# Display the plots
plt.tight_layout()
plt.show()

# %% [code]
results_df

# %% [code]
# Get the top 10 features based on absolute coefficient values
top_coefficients_df = coefficients_df.head(20)

# Plot the top 10 coefficients
plt.figure(figsize=(10, 6))
sns.barplot(data=top_coefficients_df, x='Coefficient', y='Feature', orient='h')
plt.title('Top 10 Linear Regression Coefficients')
plt.xlabel('Coefficient')
plt.ylabel('Feature')
plt.show()
```


ДОДАТОК Б.

СТАТИСТИЧНЕ ТА ВІЗУАЛЬНЕ ДОСЛІДЖЕННЯ ДАНИХ

```
# %% [code] {"_kg_hide-input":true}
import os
import string
import datetime
import numpy as np
import pandas as pd

import matplotlib.pyplot as plt
# import seaborn as sns
# color = sns.color_palette()

%matplotlib inline

from plotly import tools
import plotly.offline as py
py.init_notebook_mode(connected=True)
import plotly.graph_objs as go
import plotly.express as px

pd.options.mode.chained_assignment = None
pd.options.display.max_columns = 999

# %% [markdown]
# Let us read the dataset and look at the top few rows to start with.

# %% [code] {"_kg_hide-input":true}
df = pd.read_csv("../input/startup_funding.csv")
df.columns = ["SNo", "Date", "StartupName", "IndustryVertical", "SubVertical", "City",
"InvestorsName", "InvestmentType", "AmountInUSD", "Remarks"]

def clean_string(x):
    return str(x).replace("\\xc2\\xa0","").replace("\\\\xc2\\\\\\xa0", "")

for col in ["StartupName", "IndustryVertical", "SubVertical", "City", "InvestorsName",
"InvestmentType", "AmountInUSD", "Remarks"]:
    df[col] = df[col].apply(lambda x: clean_string(x))

df.head()

# %% [code] {"_kg_hide-input":true}
### Some more fixes in the data format. Will try to fix in the input file in next version #
df['Date'][df['Date']=='01/07/015'] = '01/07/2015'
df['Date'][df['Date']=='\\\\xc2\\\\\\xa010/7/2015'] = '10/07/2015'
df['Date'][df['Date']=='12/05.2015'] = '12/05/2015'
df['Date'][df['Date']=='13/04.2015'] = '13/04/2015'
df['Date'][df['Date']=='15/01.2015'] = '15/01/2015'
```

```

df['Date'][df['Date']=='22/01//2015'] = '22/01/2015'
df['Date'][df['Date']=='05/072018'] = '05/07/2018'

df["year"] = pd.to_datetime(df['Date'],format='%d/%m/%Y').dt.year
df["yearmonth"] = (pd.to_datetime(df['Date'],format='%d/%m/%Y').dt.year*100)+(pd.to_datetime(df['Date'],format='%d/%m/%Y').dt.month)

cnt_srs = df["year"].value_counts()
cnt_srs = cnt_srs.sort_index()
trace = go.Bar(
    x=cnt_srs.index,
    y=cnt_srs.values,
    marker=dict(
        color="#1E90FF",
    ),
)

layout = go.Layout(
    title=go.layout.Title(
        text="Number of funding deals over years",
        x=0.5
    ),
    font=dict(size=14),
    width=800,
    height=500,
)

data = [trace]
fig = go.Figure(data=data, layout=layout)
py.iplot(fig, filename="funding")

# %% [code] {"_kg_hide-input":true}
import datetime
df["yearmonth"] = pd.to_datetime(df['Date'],format='%d/%m/%Y').apply(lambda x:
datetime.date(x.year,x.month,1))

cnt_srs = df["yearmonth"].value_counts()
cnt_srs = cnt_srs.sort_index()
trace = go.Scatter(
    x=cnt_srs.index.astype(str),
    y=cnt_srs.values,
    marker=dict(
        color="#1E90FF",
    ),
)

layout = go.Layout(
    title=go.layout.Title(
        text="Number of funding deals - month on month",
        x=0.5
    ),
)

```

```

    font=dict(size=14),
    width=800,
    height=500,
)

data = [trace]
fig = go.Figure(data=data, layout=layout)
py.iplot(fig, filename="funding")

# %% [code] {"_kg_hide-input":true}
def clean_amount(x):
    #x = ".join([c for c in str(x) if c in ['0', '1', '2', '3', '4', '5', '6', '7', '8', '9']])
    x = str(x).replace(",","").replace("+","")
    x = str(x).lower().replace("undisclosed","")
    x = str(x).lower().replace("n/a","")
    if x == "":
        x = '-999'
    return x

df["CleanedAmount"] = df["AmountInUSD"].apply(lambda x: float(clean_amount(x)))
AmountInUSD = df["CleanedAmount"]
AmountInUSD = AmountInUSD[~np.isnan(AmountInUSD)]
AmountInUSD = AmountInUSD[AmountInUSD!=-999]
plt.figure(figsize=(8,6))
plt.scatter(range(len(AmountInUSD)), np.sort(AmountInUSD.values), color="#1E90FF")
plt.xlabel('index', fontsize=12)
plt.ylabel('Funding value in USD', fontsize=12)
plt.title("Distribution of funding value in USD")
plt.show()

# %% [markdown]
# There are some extreme values at the right. Let us see who are these very well funded startups.

# %% [code]
df.iloc[df["CleanedAmount"].sort_values(ascending=False).index[:5]]

# %% [code] {"_kg_hide-input":true}
df["CleanedAmount"][df["CleanedAmount"]==3900000000] = 50000000

cnt_df = df.groupby('year')['CleanedAmount'].agg(["sum", "mean"])
cnt_srs = cnt_df["sum"]
cnt_srs = cnt_srs.sort_index()
trace = go.Bar(
    x=cnt_srs.index,
    y=cnt_srs.values,
    marker=dict(
        color="#1E90FF",
    ),
)

layout = go.Layout(
    title=go.layout.Title(

```

```

        text="Sum of funding value in each year",
        x=0.5
    ),
    font=dict(size=14),
    width=800,
    height=500,
)

data = [trace]
fig = go.Figure(data=data, layout=layout)
py.iplot(fig, filename="funding")

# %% [code] {"_kg_hide-input":true}
cnt_srs = cnt_df["mean"]
cnt_srs = cnt_srs.sort_index()
trace = go.Bar(
    x=cnt_srs.index,
    y=cnt_srs.values,
    marker=dict(
        color="#1E90FF",
    ),
)

layout = go.Layout(
    title=go.layout.Title(
        text="Mean of funding value in each year",
        x=0.5
    ),
    font=dict(size=14),
    width=800,
    height=500,
)

data = [trace]
fig = go.Figure(data=data, layout=layout)
py.iplot(fig, filename="funding")

# %% [code] {"_kg_hide-input":true}
import re
import string
PUNCT_TO_REMOVE = string.punctuation
def remove_punctuation(text):
    """custom function to remove the punctuation"""
    return text.translate(str.maketrans("", "", PUNCT_TO_REMOVE))
df['InvestmentType'] = df['InvestmentType'].apply(lambda x: remove_punctuation(str(x)))

funding_map = {
    "SeedAngel Funding": "Seed Angel Funding",
    "SeedFunding": "Seed Funding",
    "PrivateEquity": "Private Equity",
    "Crowd funding": "Crowd Funding",
    "Angel Seed Funding": "Seed Angel Funding",

```

```

    "Seed Angel Funding": "Seed Angel Funding",
    "Seed Angle Funding": "Seed Angel Funding",
    "Seed Angle Funding": "Seed Angel Funding",
    "SeednFunding": "Seed Funding",
    "Seed funding": "Seed Funding",
    "Seed Round": "Seed Funding",
    "preSeries A": "PreSeries A",
    "preseries A": "PreSeries A",
    "Pre Series A": "PreSeries A"
}

for i, v in funding_map.items():
    df['InvestmentType'][df['InvestmentType']==i] = v

def horizontal_bar_chart(srs, color):
    trace = go.Bar(
        x=srs.values[::-1],
        y=srs.index[::-1],
        showlegend=False,
        orientation = 'h',
        marker=dict(
            color=color,
        ),
    )
    return trace

cnt_srs = df['InvestmentType'].value_counts()[:10]
layout = go.Layout(
    title=go.layout.Title(
        text="Count of Investment type",
        x=0.5
    ),
    font=dict(size=14),
    width=800,
    height=600,
)

data = [horizontal_bar_chart(cnt_srs, "#1E90FF")]
fig = go.Figure(data=data, layout=layout)
py.iplot(fig, filename="funding")

# %% [code] {"_kg_hide-input":true}
cnt_df = df.groupby('InvestmentType')['CleanedAmount'].agg(['size', 'sum', 'mean'])
cnt_df = cnt_df.sort_values(by="size", ascending=False)
cnt_df = cnt_df.iloc[:10,:]
cnt_srs = cnt_df["sum"]
cnt_srs = cnt_srs[["Seed Funding", "Seed Angel Funding", "PreSeries A", "Series A", "Series B",
"Series C", "Series D", "Private Equity"]]
layout = go.Layout(
    title=go.layout.Title(
        text="Sum of investment amount by Investment Type",
        x=0.5

```

```

    ),
    font=dict(size=14),
    width=800,
    height=600,
)

data = [horizontal_bar_chart(cnt_srs, "#1E90FF")]
fig = go.Figure(data=data, layout=layout)
py.iplot(fig, filename="funding")

# %% [code] {"_kg_hide-input":true}
#cnt_df = cnt_df.iloc[:10,:]
cnt_df = df.groupby('InvestmentType')['CleanedAmount'].agg(['size', 'sum', 'mean'])
cnt_df = cnt_df.sort_values(by="size", ascending=False)
cnt_srs = cnt_df["mean"]
cnt_srs = cnt_srs[["Seed Funding", "Seed Angel Funding", "PreSeries A", "Series A", "Series B",
"Series C", "Series D", "Private Equity"]]
layout = go.Layout(
    title=go.layout.Title(
        text="Mean Investment amount by Funding Round",
        x=0.5
    ),
    font=dict(size=14),
    width=800,
    height=600,
)

data = [horizontal_bar_chart(cnt_srs, "#1E90FF")]
fig = go.Figure(data=data, layout=layout)
py.iplot(fig, filename="funding")

# %% [code] {"_kg_hide-input":true}
location_map = {
    "Bengaluru": "Bangalore",
    "Delhi": "NCR",
    "New Delhi": "NCR",
    "Gurugram": "NCR",
    "Gurgaon": "NCR",
    "Noida": "NCR"
}
for i, v in location_map.items():
    df['City'][df['City']==i] = v

def horizontal_bar_chart(srs, color):
    trace = go.Bar(
        x=srs.values[::-1],
        y=srs.index[::-1],
        showlegend=False,
        orientation = 'h',
        marker=dict(
            color=color,
        ),
    ),

```

```

    )
    return trace

cnt_srs = df['City'].value_counts()[:14]
layout = go.Layout(
    title=go.layout.Title(
        text="Number of funding deals in each location",
        x=0.5
    ),
    font=dict(size=14),
    width=800,
    height=800,
)

data = [horizontal_bar_chart(cnt_srs, "#1E90FF")]
fig = go.Figure(data=data, layout=layout)
py.iplot(fig, filename="funding")

# %% [code] {"_kg_hide-input":true}
cnt_df = df.groupby('City')['CleanedAmount'].agg(["size", "sum", "mean"])
cnt_df = cnt_df.sort_values(by="size", ascending=False)
cnt_df = cnt_df.iloc[:14,:]

cnt_srs = cnt_df['sum']
cnt_srs = cnt_srs.sort_values(ascending=False)
layout = go.Layout(
    title=go.layout.Title(
        text="Total funding value by location",
        x=0.5
    ),
    font=dict(size=14),
    width=800,
    height=800,
)

data = [horizontal_bar_chart(cnt_srs, "#1E90FF")]
fig = go.Figure(data=data, layout=layout)
py.iplot(fig, filename="funding")

# %% [code] {"_kg_hide-input":true}
cnt_srs = cnt_df['mean']
cnt_srs = cnt_srs.sort_values(ascending=False)
layout = go.Layout(
    title=go.layout.Title(
        text="Mean funding value by location",
        x=0.5
    ),
    font=dict(size=14),
    width=800,
    height=800,
)

```

```

data = [horizontal_bar_chart(cnt_srs, "#1E90FF")]
fig = go.Figure(data=data, layout=layout)
py.iplot(fig, filename="funding")

# %% [code] {"_kg_hide-input":true}
temp_df = df.groupby(["City", "year"])["CleanedAmount"].agg(["size", "mean"]).reset_index()
cities_to_use = ["Bangalore", "NCR", "Mumbai", "Chennai", "Pune", "Hyderabad"]
temp_df = temp_df[temp_df["City"].isin(cities_to_use)]
temp_df.columns = ["City", "Year", "Count", "Mean"]

fig = px.scatter(temp_df, x="Year", y="City", color="City", size="Count")
layout = go.Layout(
    title=go.layout.Title(
        text="Number of funding deals by location over time",
        x=0.5
    ),
    font=dict(size=14),
    width=800,
    height=600,
    showlegend=False
)
fig.update_layout(layout)
fig.show()

# %% [code] {"_kg_hide-input":true}
fig = px.scatter(temp_df, x="Year", y="City", color="City", size="Mean")
layout = go.Layout(
    title=go.layout.Title(
        text="Mean funding value by location over time",
        x=0.5
    ),
    font=dict(size=14),
    width=800,
    height=600,
    showlegend=False
)
fig.update_layout(layout)
fig.show()

# %% [code] {"_kg_hide-input":true}
location_map = {
    "eCommerce": "E-Commerce",
    "ECommerce": "E-Commerce"
}
for i, v in location_map.items():
    df["IndustryVertical"][df["IndustryVertical"]==i] = v

cnt_srs = df["IndustryVertical"].value_counts()[:10]
layout = go.Layout(
    title=go.layout.Title(
        text="Number of funding deals by Industry Vertical",
        x=0.5
    )
)

```

```

    ),
    font=dict(size=14),
    width=800,
    height=600,
)

data = [horizontal_bar_chart(cnt_srs, "#1E90FF")]
fig = go.Figure(data=data, layout=layout)
py.iplot(fig, filename="funding")

# %% [code] {"_kg_hide-input":true}
temp_df = df.groupby(["IndustryVertical", "year"])["CleanedAmount"].agg(["size",
"mean"]).reset_index()
values_to_use = ["Consumer Internet", "Technology", "E-Commerce", "Healthcare", "Finance",
"Logistics"]
temp_df = temp_df[temp_df["IndustryVertical"].isin(values_to_use)]
temp_df = temp_df[temp_df["year"]!=2015]
temp_df.columns = ["IndustryVertical", "Year", "Count", "Mean"]

fig = px.scatter(temp_df, x="Year", y="IndustryVertical", color='IndustryVertical', size="Count")
layout = go.Layout(
    title=go.layout.Title(
        text="Number of funding deals by industry over time",
        x=0.5
    ),
    font=dict(size=14),
    width=600,
    height=600,
    showlegend=False
)
fig.update_layout(layout)
fig.show()

# %% [code] {"_kg_hide-input":true}
fig = px.scatter(temp_df, x="Year", y="IndustryVertical", color='IndustryVertical', size="Mean")
layout = go.Layout(
    title=go.layout.Title(
        text="Mean funding value by industry over time",
        x=0.5
    ),
    font=dict(size=14),
    width=600,
    height=600,
    showlegend=False
)
fig.update_layout(layout)
fig.show()

# %% [code] {"_kg_hide-input":true}
from wordcloud import WordCloud, STOPWORDS
import random
random.seed(123)

```

```

inv_names = []
for invs in df['InvestorsName']:
    for inv in str(invs).split(","):
        if inv != "":
            inv_names.append(inv.strip().lower().replace(" ", "_").replace("'", ""))

def grey_color_func(word, font_size, position, orientation, random_state=None,
                    **kwargs):
    #return "hsl(0, 0%%, %d%%)" % (font_size*3)
    return (100, 100, font_size*3)

def plot_wordcloud(text, mask=None, max_words=40, max_font_size=80, figure_size=(24.0,16.0),
                  title = None, title_size=40, image_color=grey_color_func):
    stopwords = set(STOPWORDS)
    more_stopwords = {'one', 'br', 'Po', 'th', 'sayi', 'fo', 'Unknown'}
    stopwords = stopwords.union(more_stopwords)

    wordcloud = WordCloud(background_color='white',
                          stopwords = stopwords,
                          max_words = max_words,
                          prefer_horizontal = 1.0,
                          max_font_size = max_font_size,
                          min_font_size = 10,
                          random_state = 42,
                          #color_func = lambda *args, **kwargs: (140,0,0),
                          #color_func = color_map(),
                          colormap="Blues",
                          width=1200,
                          height=600,
                          mask = mask)
    wordcloud.generate(str(text))

    plt.figure(figsize=figure_size)
    if image_color:
        #image_colors = ImageColorGenerator(mask);
        plt.imshow(wordcloud.recolor(color_func=image_color, interpolation="bilinear");
        plt.title(title, fontdict={'size': title_size, 'color': 'blue',
                                   'verticalalignment': 'bottom'})
    else:
        plt.imshow(wordcloud);
        plt.title(title, fontdict={'size': title_size, 'color': 'blue',
                                   'verticalalignment': 'bottom'})

    plt.axis('off');
    plt.tight_layout()

plot_wordcloud(' '.join(inv_names), title="Investors with most number of funding deals")

# %% [code] {"_kg_hide-input":true}
cnt_srs = pd.Series(inv_names).value_counts()[:10]
layout = go.Layout(
    title=go.layout.Title(

```

```

        text="Number of funding deals by Investor Name",
        x=0.5
    ),
    font=dict(size=14),
    width=800,
    height=600,
)

data = [horizontal_bar_chart(cnt_srs, "#1E90FF")]
fig = go.Figure(data=data, layout=layout)
py.iplot(fig, filename="funding")

# %% [code] {"_kg_hide-input":true}
inv_names = []
for invs in df['SubVertical']:
    for inv in str(invs).split():
        if inv != "":
            inv_names.append(inv.strip().lower().replace("'", ""))

def plot_wordcloud(text, mask=None, max_words=40, max_font_size=80, figure_size=(24.0,16.0),
                  title = None, title_size=40, image_color=False):
    stopwords = set(STOPWORDS)
    more_stopwords = {'one', 'br', 'Po', 'th', 'sayi', 'fo', 'Unknown', 'nan', ' nan'}
    stopwords = stopwords.union(more_stopwords)

    wordcloud = WordCloud(background_color='white',
                          stopwords = stopwords,
                          max_words = max_words,
                          prefer_horizontal = 1.0,
                          max_font_size = max_font_size,
                          min_font_size = 10,
                          random_state = 42,
                          #color_func = lambda *args, **kwargs: (140,0,0),
                          #color_func = color_map(),
                          colormap="Blues",
                          width=600,
                          height=300,
                          mask = mask)
    wordcloud.generate(str(text))

    plt.figure(figsize=figure_size)
    if image_color:
        #image_colors = ImageColorGenerator(mask);
        plt.imshow(wordcloud.recolor(color_func=image_color), interpolation="bilinear");
        plt.title(title, fontdict={'size': title_size, 'color': 'blue',
                                   'verticalalignment': 'bottom'})
    else:
        plt.imshow(wordcloud);
        plt.title(title, fontdict={'size': title_size, 'color': 'blue',
                                   'verticalalignment': 'bottom'})

    plt.axis('off');
    plt.tight_layout()

```

```

plot_wordcloud(' '.join(inv_names), title="Subvertical with most number of funding deals")
# %% [code] {"_kg_hide-input":true}
import os
import string
import datetime
import numpy as np
import pandas as pd

import matplotlib.pyplot as plt
# import seaborn as sns
# color = sns.color_palette()

%matplotlib inline

from plotly import tools
import plotly.offline as py
py.init_notebook_mode(connected=True)
import plotly.graph_objs as go
import plotly.express as px

pd.options.mode.chained_assignment = None
pd.options.display.max_columns = 999

# %% [markdown]
# Let us read the dataset and look at the top few rows to start with.

# %% [code] {"_kg_hide-input":true}
df = pd.read_csv("../input/startup_funding.csv")
df.columns = ["SNo", "Date", "StartupName", "IndustryVertical", "SubVertical", "City",
"InvestorsName", "InvestmentType", "AmountInUSD", "Remarks"]

def clean_string(x):
    return str(x).replace("\\xc2\\xa0", "").replace("\\\\xc2\\\\\\xa0", "")

for col in ["StartupName", "IndustryVertical", "SubVertical", "City", "InvestorsName",
"InvestmentType", "AmountInUSD", "Remarks"]:
    df[col] = df[col].apply(lambda x: clean_string(x))

df.head()

# %% [code] {"_kg_hide-input":true}
### Some more fixes in the data format. Will try to fix in the input file in next version #
df['Date'][df['Date']=='01/07/015'] = '01/07/2015'
df['Date'][df['Date']=='\\\\\\xc2\\\\\\xa010/7/2015'] = '10/07/2015'
df['Date'][df['Date']=='12/05.2015'] = '12/05/2015'
df['Date'][df['Date']=='13/04.2015'] = '13/04/2015'
df['Date'][df['Date']=='15/01.2015'] = '15/01/2015'
df['Date'][df['Date']=='22/01//2015'] = '22/01/2015'
df['Date'][df['Date']=='05/072018'] = '05/07/2018'

df["year"] = pd.to_datetime(df['Date'],format='%d/%m/%Y').dt.year

```

```

df["yearmonth"]
(pd.to_datetime(df['Date'],format='%d/%m/%Y').dt.year*100)+(pd.to_datetime(df['Date'],format='%d/%
m/%Y').dt.month)

cnt_srs = df["year"].value_counts()
cnt_srs = cnt_srs.sort_index()
trace = go.Bar(
    x=cnt_srs.index,
    y=cnt_srs.values,
    marker=dict(
        color="#1E90FF",
    ),
)

layout = go.Layout(
    title=go.layout.Title(
        text="Number of funding deals over years",
        x=0.5
    ),
    font=dict(size=14),
    width=800,
    height=500,
)

data = [trace]
fig = go.Figure(data=data, layout=layout)
py.iplot(fig, filename="funding")

# %% [code] {"_kg_hide-input":true}
import datetime
df["yearmonth"] = pd.to_datetime(df['Date'],format='%d/%m/%Y').apply(lambda x:
datetime.date(x.year,x.month,1))

cnt_srs = df["yearmonth"].value_counts()
cnt_srs = cnt_srs.sort_index()
trace = go.Scatter(
    x=cnt_srs.index.astype(str),
    y=cnt_srs.values,
    marker=dict(
        color="#1E90FF",
    ),
)

layout = go.Layout(
    title=go.layout.Title(
        text="Number of funding deals - month on month",
        x=0.5
    ),
    font=dict(size=14),
    width=800,
    height=500,
)

```

```

data = [trace]
fig = go.Figure(data=data, layout=layout)
py.iplot(fig, filename="funding")

# %% [code] {"_kg_hide-input":true}
def clean_amount(x):
    #x = ".join([c for c in str(x) if c in ['0', '1', '2', '3', '4', '5', '6', '7', '8', '9']])
    x = str(x).replace(",","").replace("+","")
    x = str(x).lower().replace("undisclosed","")
    x = str(x).lower().replace("n/a","")
    if x == "":
        x = '-999'
    return x

df["CleanedAmount"] = df["AmountInUSD"].apply(lambda x: float(clean_amount(x)))
AmountInUSD = df["CleanedAmount"]
AmountInUSD = AmountInUSD[~np.isnan(AmountInUSD)]
AmountInUSD = AmountInUSD[AmountInUSD!=-999]
plt.figure(figsize=(8,6))
plt.scatter(range(len(AmountInUSD)), np.sort(AmountInUSD.values), color="#1E90FF")
plt.xlabel('index', fontsize=12)
plt.ylabel('Funding value in USD', fontsize=12)
plt.title("Distribution of funding value in USD")
plt.show()

# %% [markdown]
# There are some extreme values at the right. Let us see who are these very well funded startups.

# %% [code]
df.iloc[df["CleanedAmount"].sort_values(ascending=False).index[:5]]

# %% [code] {"_kg_hide-input":true}
df["CleanedAmount"][df["CleanedAmount"]==3900000000] = 50000000

cnt_df = df.groupby('year')['CleanedAmount'].agg(["sum", "mean"])
cnt_srs = cnt_df["sum"]
cnt_srs = cnt_srs.sort_index()
trace = go.Bar(
    x=cnt_srs.index,
    y=cnt_srs.values,
    marker=dict(
        color="#1E90FF",
    ),
)

layout = go.Layout(
    title=go.layout.Title(
        text="Sum of funding value in each year",
        x=0.5
    ),
    font=dict(size=14),

```

```

        width=800,
        height=500,
    )

    data = [trace]
    fig = go.Figure(data=data, layout=layout)
    py.iplot(fig, filename="funding")

# %% [code] {"_kg_hide-input":true}
cnt_srs = cnt_df["mean"]
cnt_srs = cnt_srs.sort_index()
trace = go.Bar(
    x=cnt_srs.index,
    y=cnt_srs.values,
    marker=dict(
        color="#1E90FF",
    ),
)

layout = go.Layout(
    title=go.layout.Title(
        text="Mean of funding value in each year",
        x=0.5
    ),
    font=dict(size=14),
    width=800,
    height=500,
)

data = [trace]
fig = go.Figure(data=data, layout=layout)
py.iplot(fig, filename="funding")

# %% [code] {"_kg_hide-input":true}
import re
import string
PUNCT_TO_REMOVE = string.punctuation
def remove_punctuation(text):
    """custom function to remove the punctuation"""
    return text.translate(str.maketrans("", "", PUNCT_TO_REMOVE))
df['InvestmentType'] = df['InvestmentType'].apply(lambda x: remove_punctuation(str(x)))

funding_map = {
    "SeedAngel Funding": "Seed Angel Funding",
    "SeedFunding": "Seed Funding",
    "PrivateEquity": "Private Equity",
    "Crowd funding": "Crowd Funding",
    "Angel Seed Funding": "Seed Angel Funding",
    "Seed Angel Funding": "Seed Angel Funding",
    "Seed Angle Funding": "Seed Angel Funding",
    "Seed Angle Funding": "Seed Angel Funding",
    "SeednFunding": "Seed Funding",

```

```

    "Seed funding": "Seed Funding",
    "Seed Round": "Seed Funding",
    "preSeries A": "PreSeries A",
    "preseries A": "PreSeries A",
    "Pre Series A": "PreSeries A"
}

for i, v in funding_map.items():
    df['InvestmentType'][df['InvestmentType']==i] = v

def horizontal_bar_chart(srs, color):
    trace = go.Bar(
        x=srs.values[::-1],
        y=srs.index[::-1],
        showlegend=False,
        orientation = 'h',
        marker=dict(
            color=color,
        ),
    )
    return trace

cnt_srs = df['InvestmentType'].value_counts()[:10]
layout = go.Layout(
    title=go.layout.Title(
        text="Count of Investment type",
        x=0.5
    ),
    font=dict(size=14),
    width=800,
    height=600,
)

data = [horizontal_bar_chart(cnt_srs, "#1E90FF")]
fig = go.Figure(data=data, layout=layout)
py.iplot(fig, filename="funding")

# %% [code] {"_kg_hide-input":true}
cnt_df = df.groupby('InvestmentType')['CleanedAmount'].agg(['size', 'sum', 'mean'])
cnt_df = cnt_df.sort_values(by="size", ascending=False)
cnt_df = cnt_df.iloc[:10,:]
cnt_srs = cnt_df["sum"]
cnt_srs = cnt_srs[["Seed Funding", "Seed Angel Funding", "PreSeries A", "Series A", "Series B",
"Series C", "Series D", "Private Equity"]]
layout = go.Layout(
    title=go.layout.Title(
        text="Sum of investment amount by Investment Type",
        x=0.5
    ),
    font=dict(size=14),
    width=800,
    height=600,
)

```

```

)

data = [horizontal_bar_chart(cnt_srs, "#1E90FF")]
fig = go.Figure(data=data, layout=layout)
py.iplot(fig, filename="funding")

# %% [code] {"_kg_hide-input":true}
#cnt_df = cnt_df.iloc[:10,:]
cnt_df = df.groupby('InvestmentType')['CleanedAmount'].agg(['size', 'sum', 'mean'])
cnt_df = cnt_df.sort_values(by="size", ascending=False)
cnt_srs = cnt_df["mean"]
cnt_srs = cnt_srs[["Seed Funding", "Seed Angel Funding", "PreSeries A", "Series A", "Series B",
"Series C", "Series D", "Private Equity"]]
layout = go.Layout(
    title=go.layout.Title(
        text="Mean Investment amount by Funding Round",
        x=0.5
    ),
    font=dict(size=14),
    width=800,
    height=600,
)

data = [horizontal_bar_chart(cnt_srs, "#1E90FF")]
fig = go.Figure(data=data, layout=layout)
py.iplot(fig, filename="funding")

# %% [code] {"_kg_hide-input":true}
location_map = {
    "Bengaluru": "Bangalore",
    "Delhi": "NCR",
    "New Delhi": "NCR",
    "Gurugram": "NCR",
    "Gurgaon": "NCR",
    "Noida": "NCR"
}
for i, v in location_map.items():
    df['City'][df['City']==i] = v

def horizontal_bar_chart(srs, color):
    trace = go.Bar(
        x=srs.values[::-1],
        y=srs.index[::-1],
        showlegend=False,
        orientation = 'h',
        marker=dict(
            color=color,
        ),
    )
    return trace

cnt_srs = df['City'].value_counts()[:14]

```

```

layout = go.Layout(
    title=go.layout.Title(
        text="Number of funding deals in each location",
        x=0.5
    ),
    font=dict(size=14),
    width=800,
    height=800,
)

data = [horizontal_bar_chart(cnt_srs, "#1E90FF")]
fig = go.Figure(data=data, layout=layout)
py.iplot(fig, filename="funding")

# %% [code] {"_kg_hide-input":true}
cnt_df = df.groupby('City')['CleanedAmount'].agg(["size", "sum", "mean"])
cnt_df = cnt_df.sort_values(by="size", ascending=False)
cnt_df = cnt_df.iloc[:14,:]

cnt_srs = cnt_df['sum']
cnt_srs = cnt_srs.sort_values(ascending=False)
layout = go.Layout(
    title=go.layout.Title(
        text="Total funding value by location",
        x=0.5
    ),
    font=dict(size=14),
    width=800,
    height=800,
)

data = [horizontal_bar_chart(cnt_srs, "#1E90FF")]
fig = go.Figure(data=data, layout=layout)
py.iplot(fig, filename="funding")

# %% [code] {"_kg_hide-input":true}
cnt_srs = cnt_df['mean']
cnt_srs = cnt_srs.sort_values(ascending=False)
layout = go.Layout(
    title=go.layout.Title(
        text="Mean funding value by location",
        x=0.5
    ),
    font=dict(size=14),
    width=800,
    height=800,
)

data = [horizontal_bar_chart(cnt_srs, "#1E90FF")]
fig = go.Figure(data=data, layout=layout)
py.iplot(fig, filename="funding")

```

```
# %% [code] {"_kg_hide-input":true}
temp_df = df.groupby(["City","year"])["CleanedAmount"].agg(["size", "mean"]).reset_index()
cities_to_use = ["Bangalore", "NCR", "Mumbai", "Chennai", "Pune", "Hyderabad"]
temp_df = temp_df[temp_df["City"].isin(cities_to_use)]
temp_df.columns = ["City", "Year", "Count", "Mean"]
```

```
fig = px.scatter(temp_df, x="Year", y="City", color="City", size="Count")
layout = go.Layout(
    title=go.layout.Title(
        text="Number of funding deals by location over time",
        x=0.5
    ),
    font=dict(size=14),
    width=800,
    height=600,
    showlegend=False
)
fig.update_layout(layout)
fig.show()
```

```
# %% [code] {"_kg_hide-input":true}
fig = px.scatter(temp_df, x="Year", y="City", color="City", size="Mean")
layout = go.Layout(
    title=go.layout.Title(
        text="Mean funding value by location over time",
        x=0.5
    ),
    font=dict(size=14),
    width=800,
    height=600,
    showlegend=False
)
fig.update_layout(layout)
fig.show()
```

```
# %% [code] {"_kg_hide-input":true}
location_map = {
    "eCommerce": "E-Commerce",
    "ECommerce": "E-Commerce"
}
for i, v in location_map.items():
    df["IndustryVertical"][df["IndustryVertical"]==i] = v
```

```
cnt_srs = df["IndustryVertical"].value_counts()[:10]
layout = go.Layout(
    title=go.layout.Title(
        text="Number of funding deals by Industry Vertical",
        x=0.5
    ),
    font=dict(size=14),
    width=800,
    height=600,
```

```

)

data = [horizontal_bar_chart(cnt_srs, "#1E90FF")]
fig = go.Figure(data=data, layout=layout)
py.iplot(fig, filename="funding")

# %% [code] {"_kg_hide-input":true}
temp_df = df.groupby(["IndustryVertical", "year"])["CleanedAmount"].agg(["size",
"mean"]).reset_index()
values_to_use = ["Consumer Internet", "Technology", "E-Commerce", "Healthcare", "Finance",
"Logistics"]
temp_df = temp_df[temp_df["IndustryVertical"].isin(values_to_use)]
temp_df = temp_df[temp_df["year"]!=2015]
temp_df.columns = ["IndustryVertical", "Year", "Count", "Mean"]

fig = px.scatter(temp_df, x="Year", y="IndustryVertical", color='IndustryVertical', size="Count")
layout = go.Layout(
    title=go.layout.Title(
        text="Number of funding deals by industry over time",
        x=0.5
    ),
    font=dict(size=14),
    width=600,
    height=600,
    showlegend=False
)
fig.update_layout(layout)
fig.show()

# %% [code] {"_kg_hide-input":true}
fig = px.scatter(temp_df, x="Year", y="IndustryVertical", color='IndustryVertical', size="Mean")
layout = go.Layout(
    title=go.layout.Title(
        text="Mean funding value by industry over time",
        x=0.5
    ),
    font=dict(size=14),
    width=600,
    height=600,
    showlegend=False
)
fig.update_layout(layout)
fig.show()

# %% [code] {"_kg_hide-input":true}
from wordcloud import WordCloud, STOPWORDS
import random
random.seed(123)

inv_names = []
for invs in df['InvestorsName']:
    for inv in str(invs).split(","):

```

```

    if inv != "":
        inv_names.append(inv.strip().lower().replace(" ", "_").replace("'", ""))

def grey_color_func(word, font_size, position, orientation, random_state=None,
                    **kwargs):
    #return "hsl(0, 0%%, %d%%)" % (font_size*3)
    return (100, 100, font_size*3)

def plot_wordcloud(text, mask=None, max_words=40, max_font_size=80, figure_size=(24.0,16.0),
                  title = None, title_size=40, image_color=grey_color_func):
    stopwords = set(STOPWORDS)
    more_stopwords = {'one', 'br', 'Po', 'th', 'sayi', 'fo', 'Unknown'}
    stopwords = stopwords.union(more_stopwords)

    wordcloud = WordCloud(background_color='white',
                          stopwords = stopwords,
                          max_words = max_words,
                          prefer_horizontal = 1.0,
                          max_font_size = max_font_size,
                          min_font_size = 10,
                          random_state = 42,
                          #color_func = lambda *args, **kwargs: (140,0,0),
                          #color_func = color_map(),
                          colormap="Blues",
                          width=1200,
                          height=600,
                          mask = mask)
    wordcloud.generate(str(text))

    plt.figure(figsize=figure_size)
    if image_color:
        #image_colors = ImageColorGenerator(mask);
        plt.imshow(wordcloud.recolor(color_func=image_color), interpolation="bilinear");
        plt.title(title, fontdict={'size': title_size, 'color': 'blue',
                                   'verticalalignment': 'bottom'})
    else:
        plt.imshow(wordcloud);
        plt.title(title, fontdict={'size': title_size, 'color': 'blue',
                                   'verticalalignment': 'bottom'})

    plt.axis('off');
    plt.tight_layout()

plot_wordcloud(' '.join(inv_names), title="Investors with most number of funding deals")

# %% [code] {"_kg_hide-input":true}
cnt_srs = pd.Series(inv_names).value_counts()[1:10]
layout = go.Layout(
    title=go.layout.Title(
        text="Number of funding deals by Investor Name",
        x=0.5
    ),
    font=dict(size=14),

```

```

        width=800,
        height=600,
    )

    data = [horizontal_bar_chart(cnt_srs, "#1E90FF")]
    fig = go.Figure(data=data, layout=layout)
    py.iplot(fig, filename="funding")

    # %% [code] {"_kg_hide-input":true}
    inv_names = []
    for invs in df['SubVertical']:
        for inv in str(invs).split():
            if inv != "":
                inv_names.append(inv.strip().lower().replace(" ", ""))

    def plot_wordcloud(text, mask=None, max_words=40, max_font_size=80, figure_size=(24.0,16.0),
                      title = None, title_size=40, image_color=False):
        stopwords = set(STOPWORDS)
        more_stopwords = {'one', 'br', 'Po', 'th', 'sayi', 'fo', 'Unknown', 'nan', ' nan'}
        stopwords = stopwords.union(more_stopwords)

        wordcloud = WordCloud(background_color='white',
                              stopwords = stopwords,
                              max_words = max_words,
                              prefer_horizontal = 1.0,
                              max_font_size = max_font_size,
                              min_font_size = 10,
                              random_state = 42,
                              #color_func = lambda *args, **kwargs: (140,0,0),
                              #color_func = color_map(),
                              colormap="Blues",
                              width=600,
                              height=300,
                              mask = mask)
        wordcloud.generate(str(text))

        plt.figure(figsize=figure_size)
        if image_color:
            #image_colors = ImageColorGenerator(mask);
            plt.imshow(wordcloud.recolor(color_func=image_color), interpolation="bilinear");
            plt.title(title, fontdict={'size': title_size, 'color': 'blue',
                                      'verticalalignment': 'bottom'})
        else:
            plt.imshow(wordcloud);
            plt.title(title, fontdict={'size': title_size, 'color': 'blue',
                                      'verticalalignment': 'bottom'})
        plt.axis('off');
        plt.tight_layout()

    plot_wordcloud(' '.join(inv_names), title="Subvertical with most number of funding deals")

```

Київський національний університет імені Тараса Шевченка

Економічний факультет

Кафедра економічної кібернетики

ЗАВДАННЯ

на кваліфікаційну роботу бакалавра

студентки 4 курсу спеціальності 051 «Економіка», ОПП «Економічна
кібернетика»

Сіриченко Владислави Анатоліївни

1. Тема роботи: Оцінка факторів успіху стартапів в ІТ-сфері.
2. Термін завершення роботи: 02.06.2024 року.
3. Попередній захист роботи: 03.06.2024 року.
4. Об'єкт дослідження: стартапи в сфері інформаційних технологій та фактори, що впливають на їх успішність.
5. Предмет дослідження: фактори, що впливають на успішність стартапів в ІТ-сфері, зокрема зовнішні умови, бізнес-моделі, інноваційні технології, венчурне фінансування та ринкові умови.
6. Мета дослідження: оцінити фактори, що впливають на успіх стартапів в ІТ-сфері, та розробити рекомендації для підвищення їх ефективності та конкурентоспроможності.
7. Завдання дослідження:
 - 7.1 проаналізувати теоретичні основи стартапів у сфері ІТ;
 - 7.2 дослідити внутрішні та зовнішні фактори впливу на успішність стартапів;
 - 7.3 провести статистичний аналіз даних про фінансування стартапів та їх успішність;
 - 7.4 використати сучасні методи класифікації, зокрема великі мовні моделі (LLM), для аналізу факторів успіху;
 - 7.5 розробити рекомендації для засновників стартапів та інвесторів щодо підвищення ефективності управління та оцінки перспективності стартапів.

Науковий керівник: кандидат економічних наук, доцент Шпирко Віктор
Васильович

Студент:

(підпис)

Затверджено на засіданні кафедри економічної кібернетики
протокол № 4 від 22.11.2023 р.

Календарний план виконання кваліфікаційної роботи бакалавра

№	Етапи роботи	Терміни виконання	Відмітка керівника про виконання
1	Вибір теми кваліфікаційної роботи бакалавра	11.09.2023-07.10.2023	
2	Розробка та затвердження завдання кваліфікаційної роботи бакалавра	08.10.2023-22.11.2023	
3	Вивчення наукових робіт за темою дипломної роботи	01.12.2023-06.01.2024	
4	Аналіз проведених досліджень та методів оцінки факторів успішності стартапів	06.01.2024-28.01.2024	
5	Збір та попередня обробка даних про стартапи у сфері інформаційних технологій	29.01.2024-14.02.2024	
6	Використання статистичних методів для аналізу даних про фінансування та успішність стартапів	15.02.2024-09.03.2024	
7	Використання методів класифікації, зокрема великих мовних моделей (LLM), для аналізу факторів успіху	10.03.2024-08.04.2024	
8	Порівняння ефективності різних методів аналізу даних та їх результатів	09.04.2024-28.04.2024	

9	Розробка рекомендацій для засновників стартапів та інвесторів	29.04.2024-14.05.2024	
10	Написання висновків, коригування та оформлення остаточного варіанту роботи	15.05.2024-01.06.2024	

Науковий керівник: Шпирко Віктор Васильович

Студент: Сіриченко Владислава Анатоліївна