1. **Агрегація даних**
2. **Мета роботи**

Навчитися агрегувати дані для наступної аналітики.

1. **Порядок виконання роботи і методичні вказівки з її виконання**
2. Завантажити користувальницькі дані  до реляційної БД. Дані представлені у форматі JSON з наступною схемою

[

        {

                ‘user\_id’: guid,

                ‘time’: date,

                ‘event\_id’: number,

                parameters: [

        ‘key1’: val1,

        ‘key2’: val2,

        ….

]

},

{....}]

Параметральна документація наведена нижче:

Таблиця 1.1 – Параметральна документація

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| event\_id | Подія | Параметер1 | Параметер2 | Параметер3 | Параметер4 |
| 1 | Запуск гри |  |  |  |  |
| 2 | Перший запуск гри | Стать (gender [string]) | Вік (age [int]) | Країна\* (country [string]) |  |
| 3 | Початок етапу | Етап  (stage [number]) |  |  |  |
| 4 | Завершення етапу | Етап  (stage [number]) | Перемога (win [bool]) | Час (секунд)  (time [number]) | Здобуто валюти\* (income [number]) |
| 5 | Покупка внутрішньоігрового предмета | Назва предмета  (item [string]) | Кошт (внутрішньоігровий)  (price [number]) |  |  |
| 6 | Покупка валюти | Назва паку  (name][string]) | Кошт (USD) (price [decimal]) | Здобуто валюти (income [number]) |  |

Параметри помічені зірочкою можуть бути відсутні.

1. При завантажені даних нормалізувати їх та забезпечити можливість шардингу.
2. Побудувати таблиці, що містять наступні загреговані значення на кожен день врахуйте, що дані є семпловані, тобто ми взяли 10% даних, що є рівномірно розподілені за користувачамі:

* DAU - кількість унікальних користувачів за добу
* New Users - кількість нових користувачей
* MAU - кількість унікальних користувачів за останні 30 діб
* Revenue - кількість USD, що зароблено за добу
* Курс внутрішньоігрової валюти за добу (прибуток за добу / кількість купленої валюти за добу)
* Поетапну статистку - кількість розпочатих етапів, кількість завершених етапів, кількість перемог, кількість отриманої внутрішньоігрової валюти (також перевести у USD)
* Попредметну статистику - кількість куплених речей, на яку суму внутрішньоігрової валюти, на яку суму USD

1. **Зміст звіту**
2. Схема БД, що буде побудована.
3. Програмний код ETL та агрегуючої систем
4. Загреговані таблиці
5. Висновки.
6. **Контрольні запитання і завдання**
7. Які основні проблеми у систем агрегації?
8. Опишіть базовий принцип агрегації даних.
9. Що таке ETL-система?
10. Навіщо використовуються ETL-системи?
11. Що відбувається на етапі вилучення даних?
12. Що відбувається на етапі переробки даних?
13. Що відбувається на етапі завантаження даних?
14. **Прогнозування даних**
15. **Мета роботи**

Навчитись прогнузувати дані використовуючи методи лінійної регресі

1. **Порядок виконання лабораторної роботи**
2. Спрогнозувати наступні параметри на наступні півроку

* DAU
* Кількість нових користувачей
* Щоденний прибуток
* Щоденну кількість проданих речей та їх вартість у USD

1. **Зміст звіту**
2. Розроблені математичні моделі.
3. Графіки з даними, що було спрогнозовано.
4. Висновки.
5. **Контрольні запитання і завдання**
6. Що таке ММ?
7. Які параметри ОД впливають на ММ?
8. Як виокремити ступінь впливу параметра ОД на ММ?
9. Навіщо потрібні регресійні моделі?
10. Як побудувати регресійну модель?
11. Що таке лінійна регресія?
12. **Сегментація користувачів**
13. **Мета роботи**

Розробити користувачів за групами, що мають схожі поведінкові ознаки.

1. **Порядок виконання лабораторної роботи**
2. Розбити користувачів за наступними сегментами

* Стать
* Вік
* Країна проживання
* Прибуток, що було принесено на останню добу, що міститься у статистиці (Tier 0 - не принесли прибутку у USD Tier 1-3 - розподілити користувачів використовуючи довільні методи кластеризації)
* Гравці, що порушували правила ігрової економіки (на якийсь час гравець витратив більше внутрішньоігрової валюти, ніж отримав)

1. Перерахувати загреговані значення за кожним сегментом
2. Зпрогнозувати параметри з лабораторної 2 по кожному сегменту
3. **Зміст звіту**
4. Математична модель, що використовувалась для кластеризації.
5. Отримані кластери.
6. Висновки.
7. **Контрольні запитання і завдання**
8. Що таке алгоритм кластеризації?
9. Що таке кластеризація?
10. Чим кластеризація відрізняється від класифікації?
11. Які проблеми мають алгоритми кластеризації?
12. Як працює алгоритм k-means?
13. **Візуалізація даних**
14. **Мета роботи**

Створити наочні уявлення аналітичних даних.

1. **Порядок виконання лабораторної роботи**
2. Побудувати графіки, що містять:

* Абсолютні та відносні показники кількості користувачів розподілених за когортами, що були визначені у лабораторній 3
* Графік DAU (наявного та передбаченого)
* Графік New Users (наявного та передбаченого)
* Графік прибутку (наявного та передбаченого)
* Розподіл прибутку за внутрішньоігровими предметами на довільну добу
* Графік курсу внутрішньоігрової валюти
* Графік складності етапів гри

1. **Зміст звіту**
2. Книга Power BI.
3. Зображення графіків, що було побудовано.
4. Висновки.
5. **Контрольні запитання і завдання**
6. Які приклади візуалізації ви знаєте?
7. Коли використовуються кільцеві діграми?
8. Коли використовуються воронки?
9. Коли використовуються мапи?
10. Які гарні приклади створення візуалізацій ви знаєте?
11. **Формування звітів**
12. **Мета роботи**

Розробити звіти, що можуть бути використані для ігрового аналізу.

1. **Порядок виконання лабораторної роботи**
2. Побудувати звіт у програмі Power BI. Кожна сторінка має містити таблицю та графік з даними. Забезпечити можливість фільтрації даних за довільними когортами та датами. Звіт має містити наступні сторінки:

* Users - розподіл користувачів за когортами, DAU, New Users
* Levels - Кількість розпочатих етапів, кількість завершених етапів, кількість перемог
* Revenue - Кількість прибутку (USD), кількість проданої валюти, передбачений прибуток
* Items - Кількість проданих речей, та отриманий прибуток за них

1. **Зміст звіту**
2. Зроблений звіт.
3. Скріншоти сторінок звіту.
4. Висновки.
5. **Контрольні запитання і завдання**
6. Які вимоги до макету звіту?
7. Навіщо вирівнювати звіт?
8. Як впливає переповнення звіту інформацією на користувача?
9. Як користувач сприймає звіт?
10. Які вимоги до кольорів звіту?