

Мета: Вивчення технології та набуття практичних навичок аналізу Великих Даних та їх візуального відображення засобами **Tableau**.

Короткі теоретичні відомості

Tableau Software - це компанія, що займається програмним забезпеченням для аналізу та візуалізації даних, заснована в 2003 році Крістіаном Шабо, Петом Ханраханом та Крісом Столте, в Маунтін-В'ю, Каліфорнія. Протягом декількох останніх років Tableau утримує лідерські позиції на світовому ринку програмних продуктів BI і є прикладом якості програм для роботи з big data.

Алгоритм роботи з даними в Tableau дуже простий і складається з трьох основних етапів: вибір джерела даних, візуалізація даних, публікація результату.

Офіційний сайт Tableau Software:

<https://www.tableau.com/>

Інсталяція Tableau Desktop

Безкоштовна версія на 14 днів з повним функціоналом. Після закінчення пробного періоду, технічно можна почати новий пробний період. Для інсталяції продукту потрібно:

- Зайти на сайт Tableau в розділ Releases;
- Завантажити останню версію, клікнувши на кнопці Download Desktop;
- Запустити інсталяційний файл на ПК, далі діяти за інструкцією;
- При першому запуску буде запропоновано заповнення анкети. Заповніть обов'язкові поля та продукт буде готовий до використання.

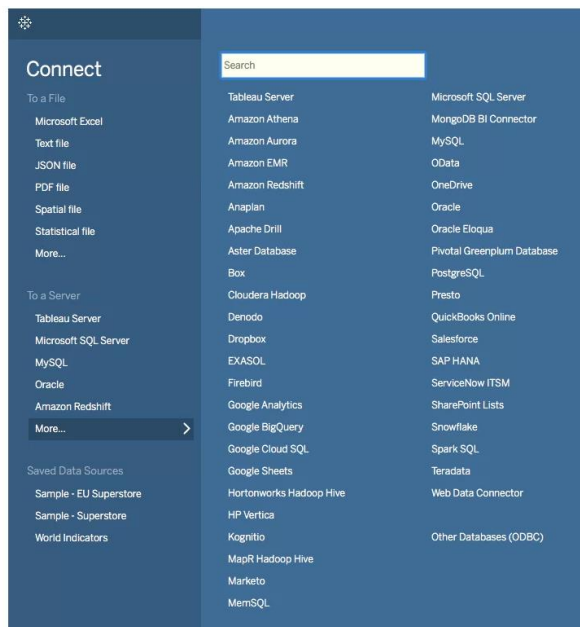
Інсталяція Tableau Public

Public - повністю безкоштовна версія Tableau Desktop, але з деякими обмеженнями. У публік не можна зберігати результати своєї роботи на ПК - тільки в свій профіль на сайті public.tableau.com. Цього буде достатньо для досягнення мети лабораторної роботи, але ви завжди можете завантажити версію Tableau Desktop, як більш зручну.

- Зайти на сайті Tableau Public - <https://public.tableau.com/s/>

- Введіть свій email у вікні на стартовій сторінці і натисніть кнопку Download the App;
- Потрібно створити обліковий запис на сайті Tableau Public через віконце в правому верхньому кутку: Sign In. У вікні в самому низу вибрати: Create one now for free, щоб зареєструватися.

Tableau Desktop підтримує безліч джерел даних: таблиці, статистичні файли, бази даних, багатовимірні куби і навіть онлайн-джерела, включаючи Google Analytics, Amazon Redshift і Salesforce. При запуску Tableau Desktop, список можливих джерел даних відображається на стартовому екрані. Файлові типи даних показуються на самому початку, потім перераховані сервери баз даних або останні джерела, до яких ви підключалися. Також можна безпосередньо вставляти табличні дані з буфера обміну. Залежно від джерела даних, можна вибрати пряме



підключення до даних або створення екстракту. При прямому підключенні, дані в Tableau будуть оновлюватися при кожній зміні джерела. Пряме підключення доступно не всім джерелам даних в Tableau і не завжди зручно у використанні. Пряме підключення збільшує навантаження на обчислювальні потужності і може уповільнювати процес роботи з даними в Tableau. Альтернативою є використання екстрактів.

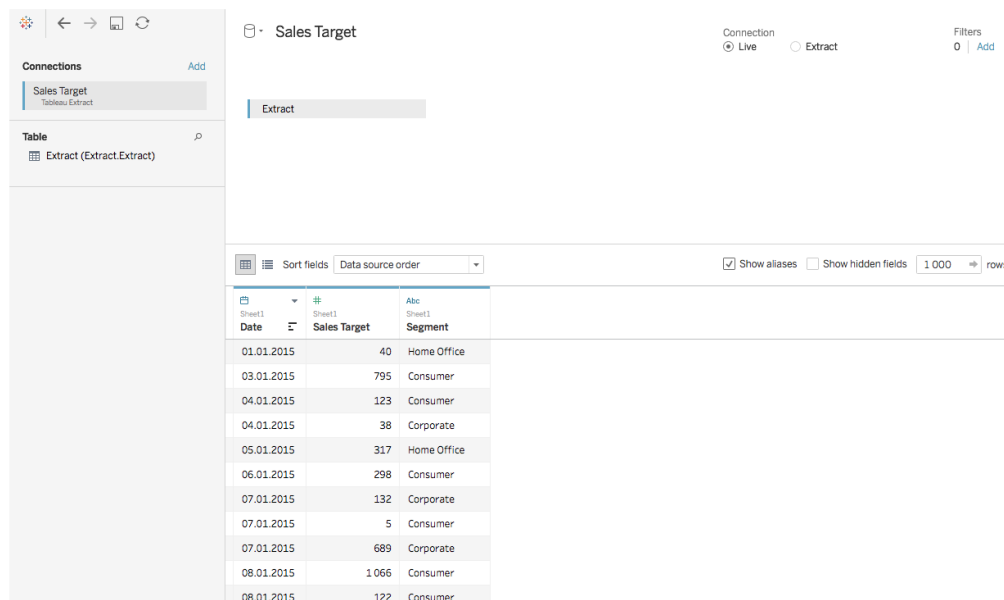
Екстракт - це одномоментний стислий знімок наявних даних в джерелі, який зберігається локально на комп'ютері і підвантажується в пам'ять програми. При створенні екстракту, його можна оптимізувати, застосувавши фільтри і агрегацію. Тим самим знизиться обсяг даних і підвищиться швидкість роботи. Оновлення екстракту може відбуватися кожні 15 хвилин. При цьому, можна виконувати як повне оновлення екстракту з нуля, так і інкрементальне оновлення, яке додає тільки нові рядки.

Для роботи з екстрактами використовується *технологія Data Engine* - це постовпчикова in-memory СУБД, використання якої суттєво прискорює роботу з аналітичними запитами для великих джерел даних. У версії 10.5 поточний Data Engine був замінений на ще більш ефективну in-memory технологію *Hyper*.

Оптимальний вибір способу підключення джерела даних між Live і Extract залежить від особливостей бази даних і конектора, частоти оновлення, швидкості роботи, складності запитів і обсягу даних.

Підключення джерела даних

Після підключення до джерела, відкривається екран налаштування джерела даних: вкладка ***DataSource***. Це проміжний етап перед початком роботи з візуалізацією даних.



Sheet1	#	Sheet1	Sheet1
Date		Sales Target	Segment
01.01.2015	40		Home Office
03.01.2015	795		Consumer
04.01.2015	123		Consumer
04.01.2015	38		Corporate
05.01.2015	317		Home Office
06.01.2015	298		Consumer
07.01.2015	132		Corporate
07.01.2015	5		Consumer
07.01.2015	689		Corporate
08.01.2015	1 066		Consumer
08.01.2015	122		Consumer

На цьому екрані можна:

- об'єднати кілька таблиць з даними (join);
- зробити union (зчеплення декількох таблиць);
- вибрати тип підключення Live / Extract;
- налаштувати типи полів;
- перейменувати поля;
- приховати непотрібні поля;
- додати фільтри на рівні джерела даних.

Як Tableau приєднується до джерел даних, можна подивитися на *video*:

https://www.youtube.com/watch?v=GzX2M6y0L_g

YouTube-канал **Tableau**:

<https://www.youtube.com/c/tableausoftware/featured>

Tableau Prep - новий продукт для підготовки даних.

При проведенні аналізу даних, 80% часу займає їх підготовка, і лише 20% - їх візуалізація і подальша аналітика. Найчастіше дані в «сирому» вигляді не придатні для завантаження в системи аналітики та подальшого аналізу: майже завжди вони вимагають процедур з очищення даних.

Цю задачу вирішують **ETL-інструменти** - від англ. Extract, Transform, Load - але, як правило, вони дорогі і вимагають хорошого рівня підготовки. Tableau вирішує зробити ETL-інструменти доступними і зрозумілими; може бути навіть орієнтованими на початківців користувачів, які хочуть швидко вирішувати свої щоденні завдання.

Назва нового ETL-продукту від Tableau - **Tableau Prep**. Інтерфейс Tableau Prep зовні схожий на Tableau Desktop. Тому процес адаптації до нового продукту у поточних користувачів можна назвати простим.

Ресурси для навчання:

Дуже ефективним є ознайомлення на самій ранній стадії роботи з Tableau за короткими відео, які підготовлені за тематиками самою компанією:

<https://www.tableau.com/learn/training/20203>

Також дуже добре допомагає робота з структурованою Базою Знань питань і відповідей:

<https://www.tableau.com/support/knowledgebase#desktop>

а також з набором онлайн-документації, яка дозволяє швидко знаходити відповіді на питання:

<https://www.tableau.com/support/help>

Збереження результатів

У Tableau працюють стандартні команди MS Windows; наприклад, Ctrl + S на клавіатурі збереже роботу. Можна переглядати місце розташування файлу, щоб зберегти робочу книгу, або перейти за замовчуванням, дозволяючи Tableau зберігати робочу книгу в папці Workbooks у сховищі My Tableau.

Зберігаючи роботу, можна вказати тип файлу. Параметри типу файлів наступні:

Таблиця робочих книг (.twb): зберігає всі аркуші та інформацію про їх з'єднання у файлі робочої книги - але дані не включені в робочу книгу, "упаковану" таблицею (.twbx): зберігає всі аркуші та інформацію про їх з'єднання та будь-які локальні ресурси (наприклад, локальні джерела даних файлів, фонові зображення, спеціальне геокодування тощо).

Tableau Gallery

Приклади і опису візуалізацій Tableau для різних сфер діяльності:

<https://www.tableau.com/solutions/gallery>

Tableau Public - Viz of the Day

Безкоштовний сервіс публікації візового і дашборда з більш ніж 1 млн робіт

<https://public.tableau.com/en-us/gallery/?tab=viz-of-the-day&type=viz-of-the-day>

Корисні YouTube-інтернет ресурси:

<https://www.youtube.com/watch?v=KsAnnCyLjgE>

<https://habr.com/ru/post/502378/>

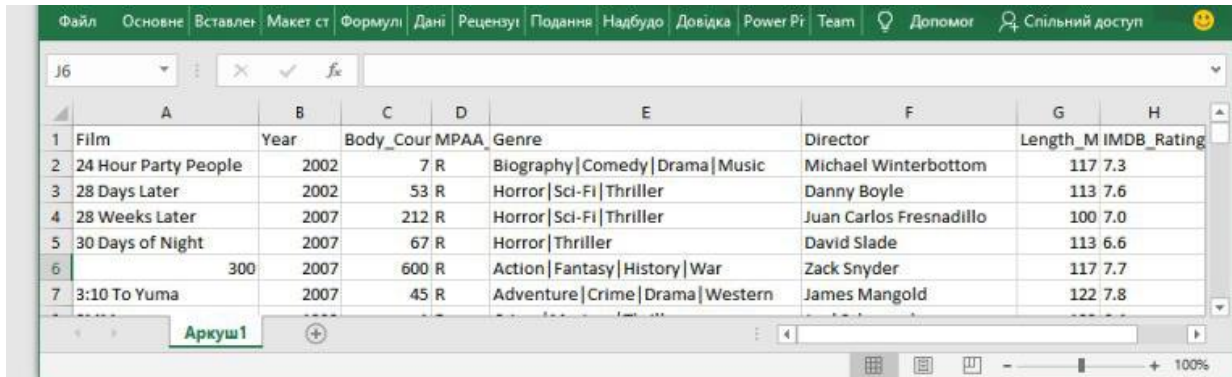
<https://habr.com/ru/post/428159/>

<https://www.youtube.com/watch?v=kV71JRUL8YY&t=1286s>

<https://www.youtube.com/watch?v=W-yxKNF-2q4>

Приклад створення Tableau dashboard

Формат файлу - це розділений комами файл (або CSV), частково показаний на наступному скріншоті:



	A	B	C	D	E	F	G	H
1	Film	Year	Body_Count	MPAA	Genre	Director	Length	IMDB_Rating
2	24 Hour Party People	2002	7	R	Biography Comedy Drama Music	Michael Winterbottom	117	7.3
3	28 Days Later	2002	53	R	Horror Sci-Fi Thriller	Danny Boyle	113	7.6
4	28 Weeks Later	2007	212	R	Horror Sci-Fi Thriller	Juan Carlos Fresnadillo	100	7.0
5	30 Days of Night	2007	67	R	Horror Thriller	David Slade	113	6.6
6	300	2007	600	R	Action Fantasy History War	Zack Snyder	117	7.7
7	3:10 To Yuma	2007	45	R	Adventure Crime Drama Western	James Mangold	122	7.8

Рис.1 Excel файл, який буде використовуватись в лабораторній для подальшого опрацювання

Починаємо з завантаження початкової сторінки Tableau (за замовчуванням Tableau відкриває стартову сторінку). Стартова сторінка Tableau - це центральне місце, з якого можна Connect (до даних), Open (існуючі робочі книги) та Discover (вміст, створений спільнотою Tableau). На стартовій сторінці (рис. 2) в розділі підключення натиснути Excel (на панелі Connect перелічено різні типи даних, до яких можна підключитися; наш файл був збережений у вигляді CSV-файлу, потім відкритий в MS Excel, переглянутий, і зберігається як робочий аркуш). Далі, у діалоговому вікні «Open» ми можемо перейти до нашого файлу та відкрити його.

(У файлі даних міститься 545 фільмів з 1949 по 2013 рік. Загиблими вважаються персонажі (люди, монстри, зомбі, прибульці), тіло яких відображається на екрані. Якщо це сцена з "натовпом" - як вибух Зірки смерті, то ці персонажі не враховуються).

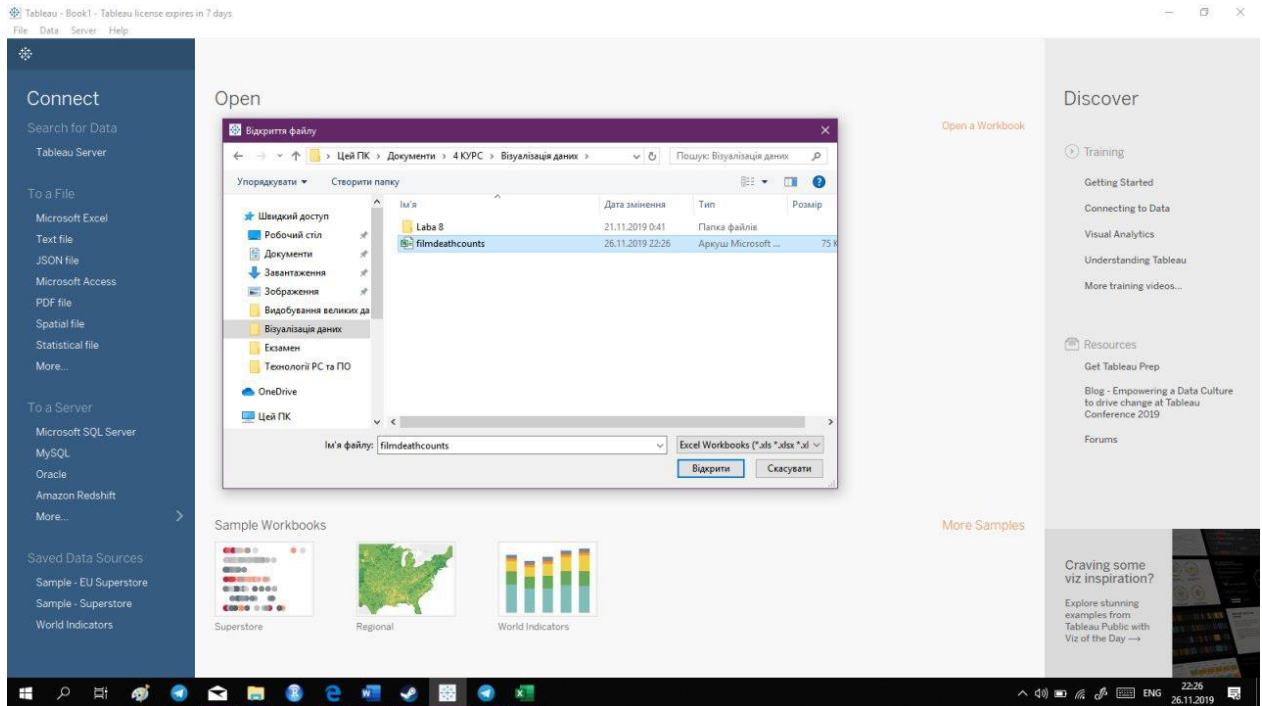


Рис. 2. Додавання файлу

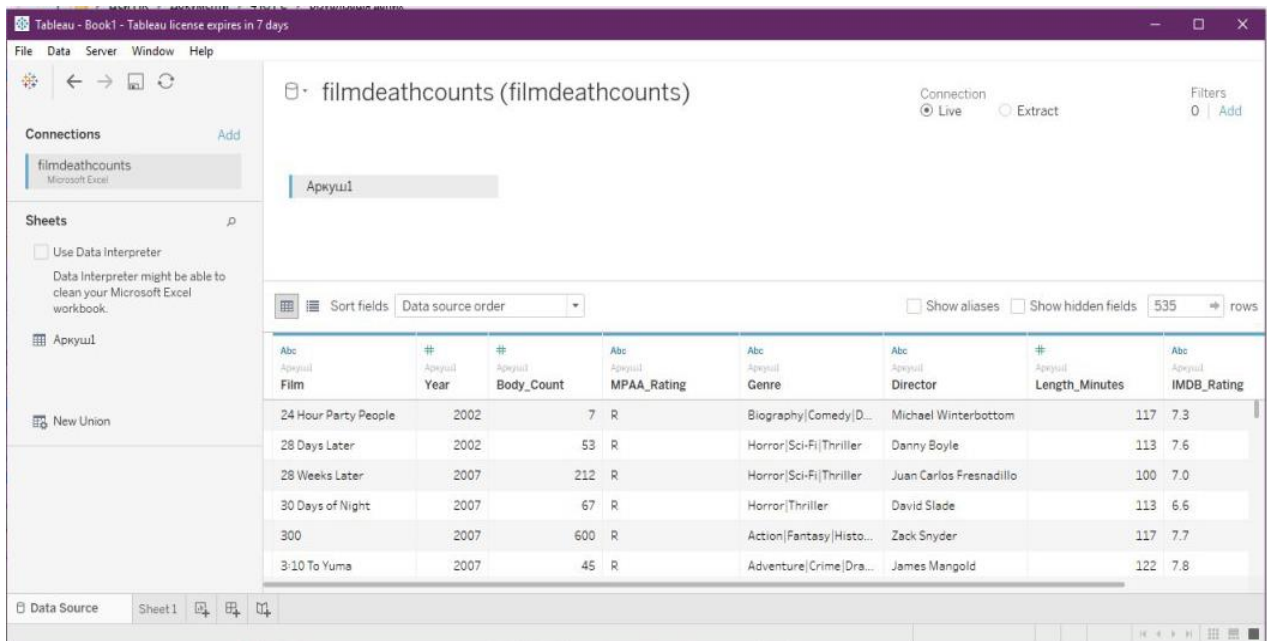


Рис. 3. Завантаження файлу

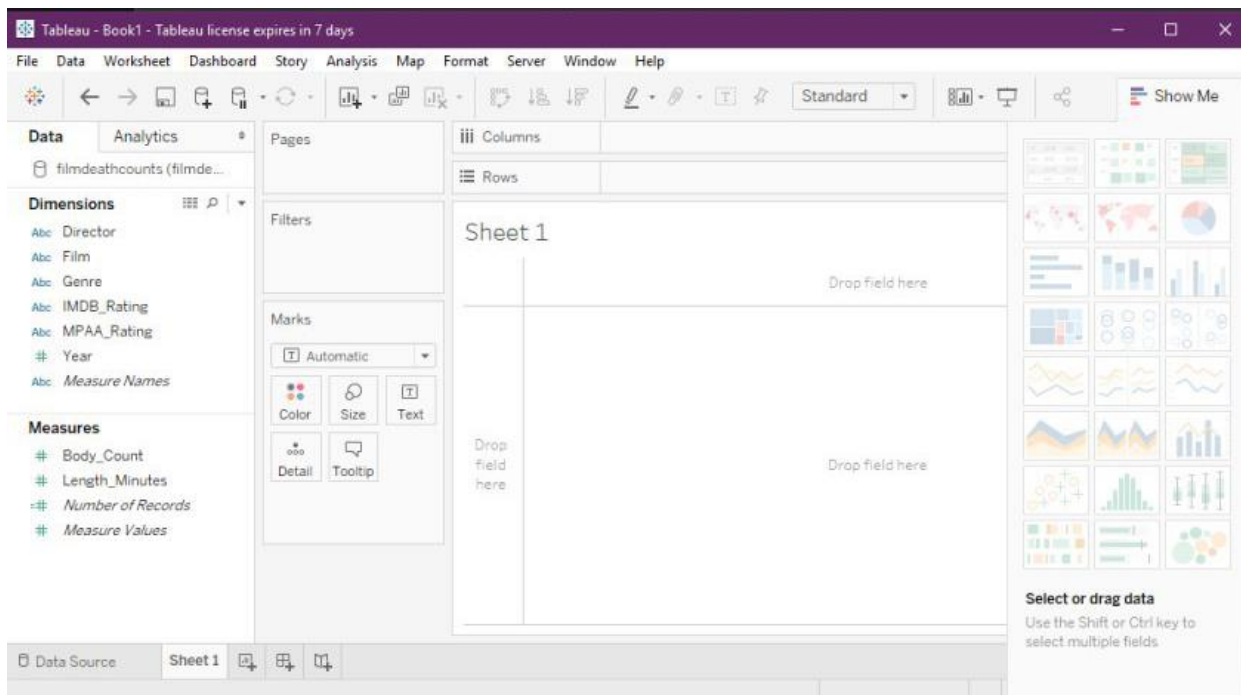


Рис. 4. Аркуш для побудови візуалізації

У лівій частині робочого аркуша знаходиться панель даних (показано на наступному скріншоті). Панель даних Tableau містить безліч різних видів полів, включаючи параметри та заходи.



Рис. 5. Дані, які будуть опрацьовуватись

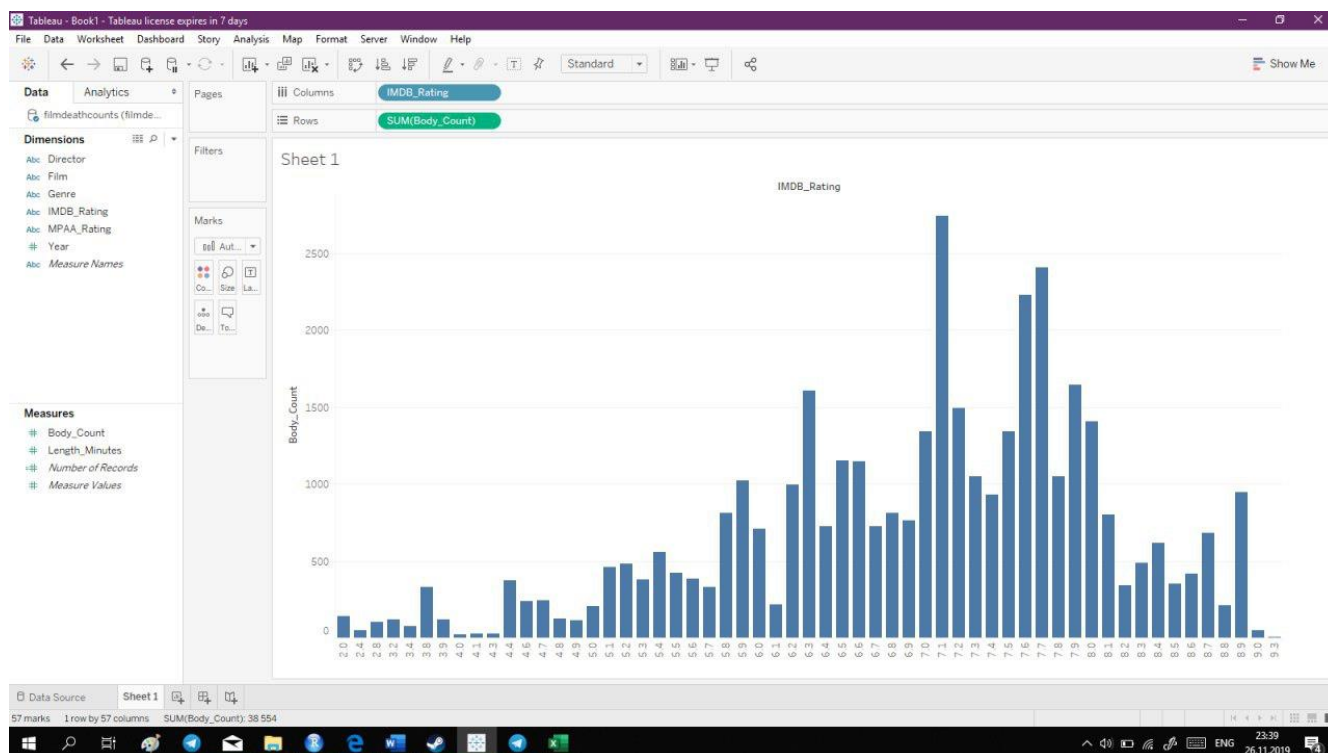


Рис. 5. Графік залежності рейтингу від кількості вбивств в фільмах

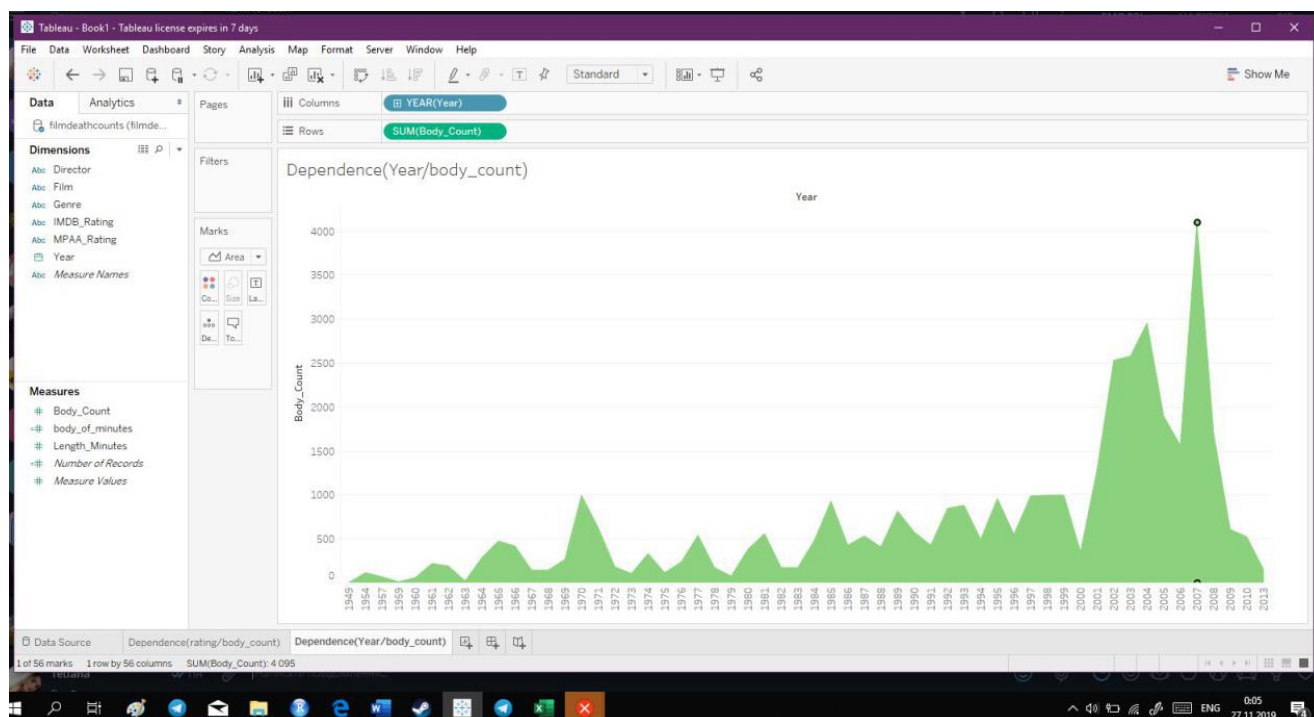


Рис. 6. Графік , на якому зображена залежність кількості загиблих в фільмах до їх року випуску

Ще одна корисна особливість - це можливість додавати обчислення безпосередньо до візуалізації. Якщо натиснути на меню з написом "Аналіз", а

потім створити "Розраховане поле", Tableau представляє діалогове вікно, де можна написати імпровізовані формули для додавання до своїх візуалізацій (без необхідності додавати їх у вихідний файл даних):

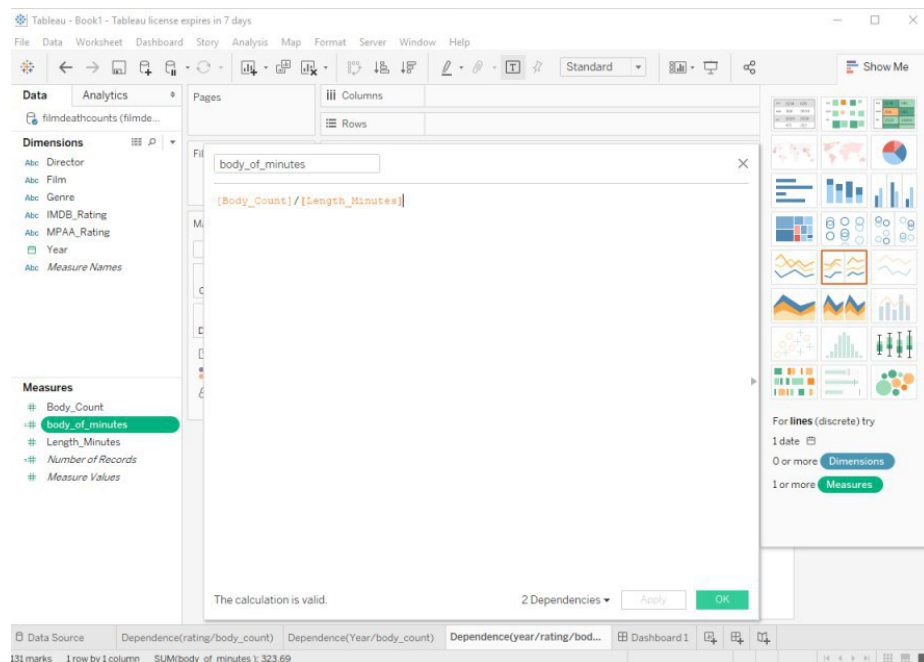


Рис. 7. Створюємо стовпець Body_of_minutes

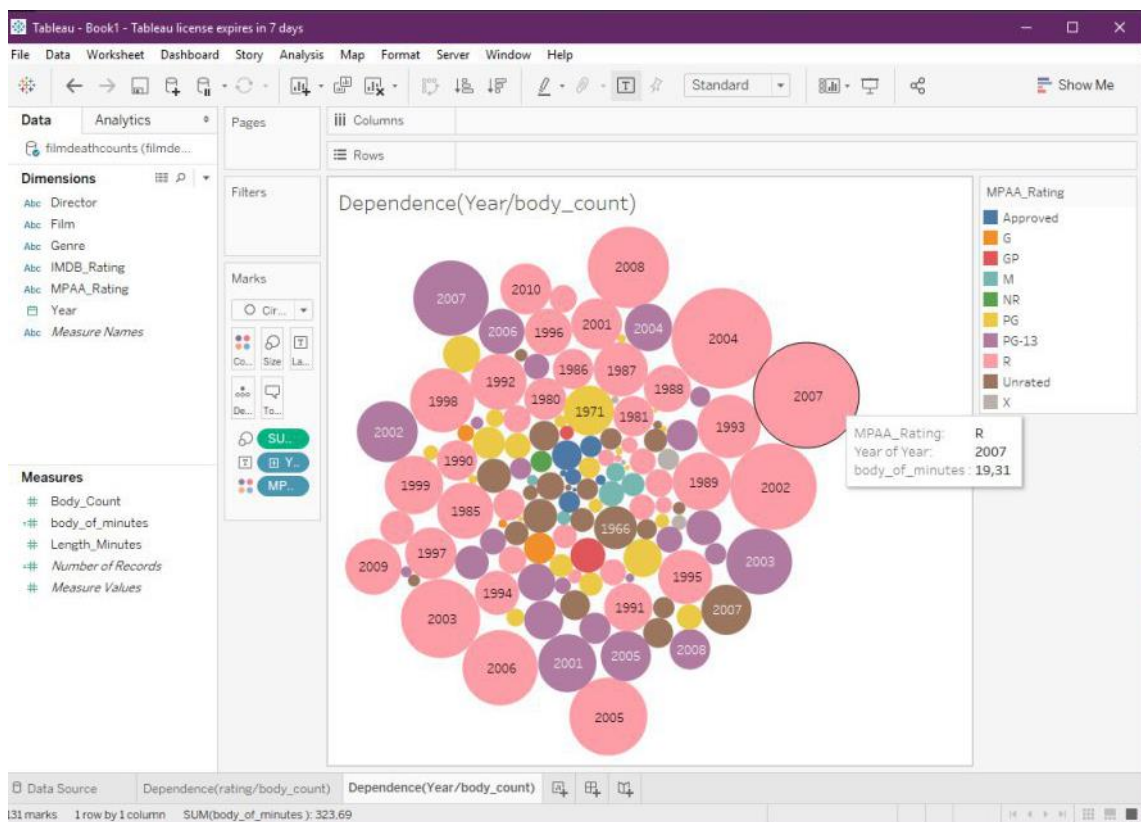


Рис.8. Бульбашкова діаграма на якій зображені рейтинг, рік та кількість смертей за хвилину (чим більша кулька тим більше смертей)

Презентація роботи

Інформаційна панель (**dashboard**) - це сукупність (робочих аркушів та супровідної інформації), що відображається в одному місці, тому можна порівнювати та контролювати одночасно, а не потрібно клацати по окремих робочих аркушах. Як і у робочих аркушах, інформаційні панелі оновлюються останніми даними з їх джерела даних. Це означає, що коли змінюємо робочий аркуш, інформаційна панель, до якої він належить, оновлюється, а коли змінюємо подання на інформаційній панелі, робочий лист оновлюється.

Створюємо інформаційну панель аналогічно тому, як ми створили наш робочий аркуш. Після натискання на піктограму «Нова панель» інформаційної панелі (внизу робочої книги) можна натиснути на створені нами представлення (перелічені під Таблицями зліва) та перетягнути їх на аркуш панелі праворуч.

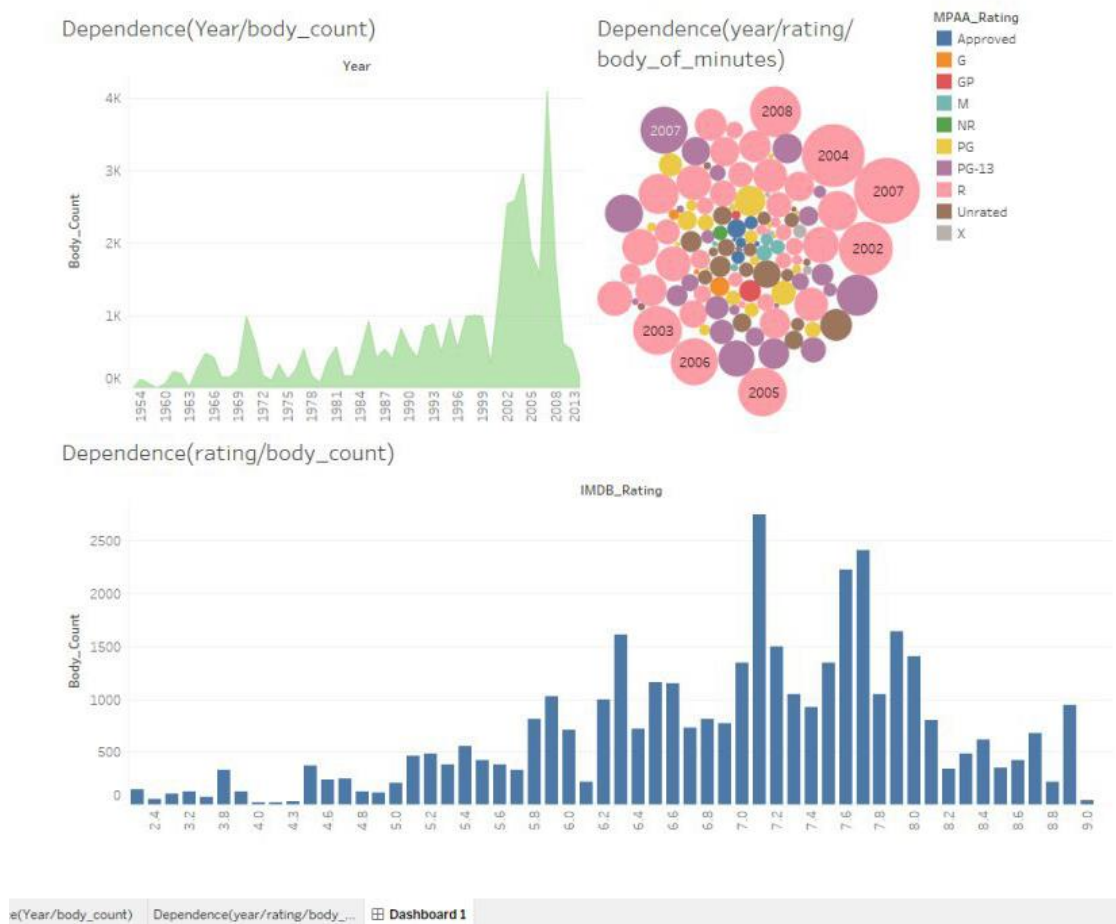


Рис.9. Презентація роботи

The result of the experiment



Рис. 10. Dashboard

Завдання для виконання:

1. Завантажити та встановити середовище Tableau.
2. Ознайомитись з порядком застосування Tableau та його функціональними можливостями.
3. Для виконання роботи обрати тестовий набір даних для завантаження: виробництво, аналіз продажів, банківська діяльність, фінансовий сектор, медицина, тощо. Навести опис вибраного набору даних.
4. Виконати аналіз та візуалізацію результатів засобами Tableau за зразками наведеними у вказівках до лабораторної роботи.
5. Зробити висновки та оформити звіт про виконання лабораторної роботи.