

Результатом виконання роботи є веб-орієнтована інформаційна система готельного комплексу, яка дозволяє автоматизувати роботу готелю та зменшити зусилля користувача. Розроблена система є актуальною на сьогоднішній день і має практичну значимість. У результаті виконання проєкту вирішені завдання, поставлені на початку роботи.

#### Список літературних джерел

1. Застосування новітніх інформаційних систем управління готелем Електронний ресурс. Режим доступу : [https://tourlib.net/statti\\_ukr/ryabenka.htm](https://tourlib.net/statti_ukr/ryabenka.htm) Дата доступу: 02.10.2020
2. Информационные системы в гостиничном бизнесе. [Електронний ресурс]. Режим доступу: <https://maxi-booking.ru/blog/informatsionnaya-sistema/> Дата доступу: 22.09.2020
3. Мова програмування php [Електронний ресурс]. Режим доступу: <http://www.znannya.org/?view=PHP> Дата доступу: 08.10.2020
4. Разработка веб сайтов. Из чего она состоит? [Електронний ресурс]. Режим доступу: <http://www.simplemedia.ru/novosti/razrabotka-veb-sajtov-etapy-razrabotki/> Дата доступу: 02.10.2020
5. Работа с MySQL в PHP. [Електронний ресурс]. Режим доступу: <https://htmlacademy.ru/tutorial/php/mysql> Дата доступу: 29.10.2020

**УДК 004.82:004:85**

*Білич А. О., студент 1 курсу спеціальності  
122 «Комп'ютерні науки» СО Магістр  
Нескородева Т. В., канд. техн. наук, доцент,  
завідувач кафедри комп'ютерних наук та  
інформаційних технологій*

## **АНАЛІЗ ДАНИХ ПЕРЕГЛЯДУ ФІЛЬМІВ КОРИСТУВАЧАМИ ПЛАТФОРМИ IMDB**

*Донецький національний університет імені Василя Стуса, м. Вінниця, Україна*

Аналіз даних про перегляд фільмів дозволяє проаналізувати сучасні погляди на кіноіндустрію як вид людських інтересів, зробити статистику по жанрам, рокам, часу кінострічки, бюджет та оцінки користувачів, саме у цьому полягає оригінальність цього дослідження [1].

Актуальність полягає в тому, що платформа яку ми аналізуємо є унікальною базою, де користувачі можуть залишати оцінки про ті чи інші фільми. Загалом ця база кінофільмів не тільки дає можливість залишити оцінку але і дає змогу переглянути повну інформацію про кінострічку.

Розглянемо приклад такого аналізу даних на основі даних які представлені у дата-сеті Movies.csv (дані з фільмами ) та Credits.csv (Дані про бюджет фільму

та його прибуток чи втрати). Загалом ця база є частковою видимою інформацією яка платформа IMDB залишає публічною. IMDB складається з 2 наборів даних. Набір даних для фільму містить 4803 спостереження з 20 змінними, тоді як набір даних для кредитів містить 4803 спостереження з 4 змінними. Ми об'єднуємо обидва набори даних, щоб створити новий набір даних, який ми можемо використовувати для нашого аналізу надалі. Загалом для аналізу дата-сету пов'язаного з IMDB рейтингом ми використаємо перехресне з'єднання цих двох дата-сетів, щоб зробити певний аналіз даних. База була представлена у 2015 році. Нижче наводиться таблиця фільмів та основні поля які будемо аналізувати :

Загалом тут представлені основні такі елементи як :

Film – Назва фільму

Genre – Жанр

Lead Studio – Назва студії виробника

Audience score – Оцінка аудиторії

	Film	Genre	Lead Studio	Audience score %	Profitability	Rotten Tomatoes %	Worldwide Gross	Year
2	Zack and Miri Make a Video	Romance	The Weinstein Company	70	1.747541667	64	\$41.94	2008
3	Youth in Revolt	Comedy	The Weinstein Company	52	1.09	68	\$19.62	2010
4	You Will Meet a Tall Dark Stranger	Comedy	Independent	35	1.211818182	43	\$26.66	2010
5	When in Rome	Comedy	Disney	44	0	15	\$43.04	2010
6	What Happens in Vegas	Comedy	Fox	72	6.267647029	28	\$219.37	2008
7	Water For Elephants	Drama	20th Century Fox	72	3.081421053	60	\$117.09	2011
8	WALL-E	Animation	Disney	89	2.896019067	96	\$521.28	2008
9	Waitress	Romance	Independent	67	11.0897415	89	\$22.18	2007
10	Waiting For Forever	Romance	Independent	53	0.005	6	\$0.03	2011
11	Valentine's Day	Comedy	Warner Bros.	54	4.184038462	17	\$217.57	2010
12	Tyler Perry's Why Did I get Married	Romance	Independent	47	3.7241924	46	\$55.86	2007
13	Twilight: Breaking Dawn	Romance	Independent	68	6.383363636	26	\$702.17	2011
14	Twilight	Romance	Summit	82	10.18002703	49	\$376.66	2008
15	The Ugly Truth	Comedy	Independent	68	5.402631579	14	\$205.30	2009
16	The Twilight Saga: New Moon	Drama	Summit	78	14.1964	27	\$709.82	2009
17	The Time Traveler's Wife	Drama	Paramount	65	2.598205128	38	\$101.33	2009
18	The Proposal	Comedy	Disney	74	7.8675	43	\$314.70	2009
19	The Invention of Lying	Comedy	Warner Bros.	47	1.751351351	56	\$32.40	2009
20	The Heartbreak Kid	Comedy	Paramount	41	2.129444167	30	\$127.77	2007
21	The Duchess	Drama	Paramount	68	3.207850222	60	\$43.31	2008
22	The Curious Case of Benjamin Button	Fantasy	Warner Bros.	81	1.78394375	73	\$285.43	2008
23	The Back-up Plan	Comedy	CBS	47	2.202571429	20	\$77.09	2010
24	Tangled	Animation	Disney	88	1.365692308	89	\$355.01	2010
25	Something Borrowed	Romance	Independent	48	1.719514286	15	\$60.18	2011
26	She's Out of My League	Comedy	Paramount	60	2.4405	57	\$48.81	2010
27	Sex and the City Two	Comedy	Warner Bros.	49	2.8835	15	\$288.35	2010
28	Sex and the City 2	Comedy	Warner Bros.	49	2.8835	15	\$288.35	2010
29	Sex and the City	Comedy	Warner Bros.	81	7.221795791	49	\$415.25	2008
30	Remember Me	Drama	Summit	70	3.49125	28	\$55.86	2010
31	Rachel Getting Married	Drama	Independent	61	1.384166667	85	\$16.61	2008
32	Penelope	Comedy	Summit	74	1.382799733	52	\$20.74	2008
33	P.S. I Love You	Romance	Independent	82	5.103116833	21	\$153.09	2007
34	Over Her Dead Body	Comedy	New Line	47	2.071	15	\$20.71	2008
35	Our Family Wedding	Comedy	Independent	49	0	14	\$21.37	2010
36	One Day	Romance	Independent	54	3.682733333	37	\$55.24	2011
37	Not Easily Broken	Drama	Independent	66	2.14	34	\$10.70	2009
38	No Reservations	Comedy	Warner Bros.	64	3.307180357	39	\$92.60	2007

Рисунок 1 – Таблиця даних Movies.csv

Об'єктами спостереження є фільми, що пропонувались за період від 1960 по 2015 роки. На першому етапі було створено описову статистику до кожного з показників та матрицю розсіювання для змінних з таблиці Movies.csv та Credits.csv.

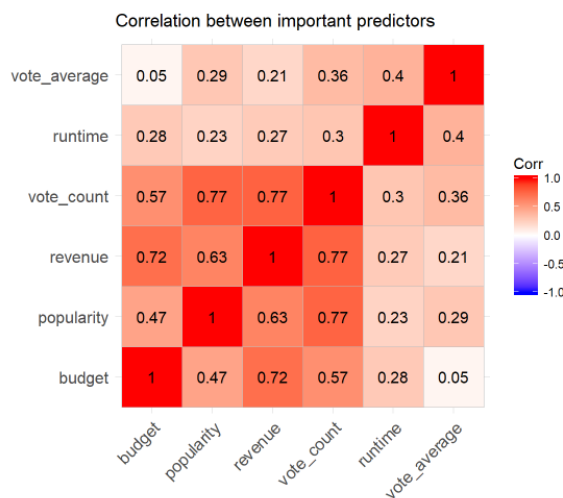


Рисунок 2 – Побудова матриць діаграм розсіювання

Також основним завданням є візуалізація даних, щоб можна було робити будь-які припущення. Для цього ми порівнюємо дані та робимо порівняння між важливими змінними. Наприклад порівнюємо бюджет фільму та інвестиції фільму

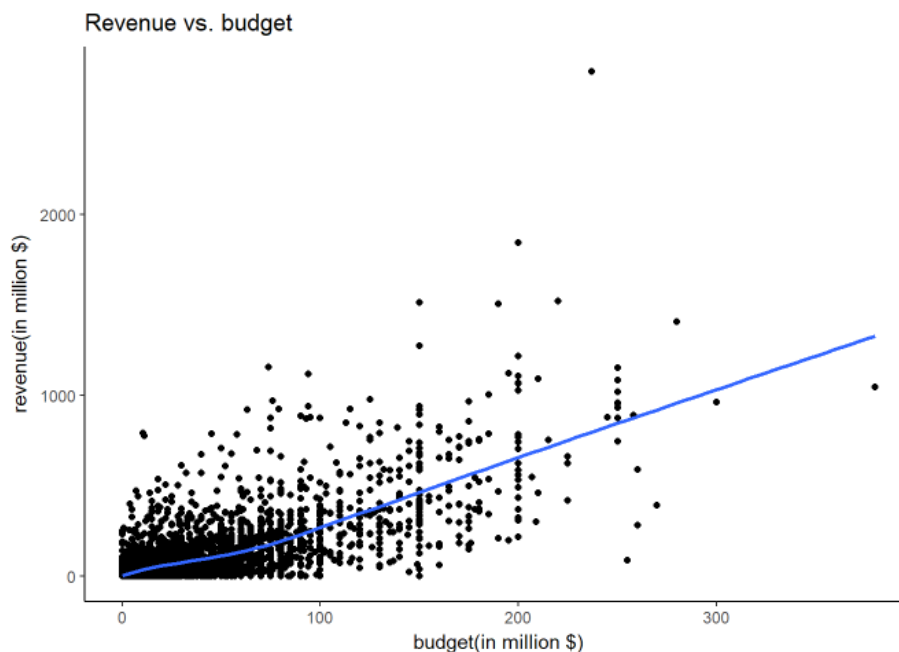


Рисунок 3 – Відношення інвестицій до бюджету

Можна сказати, що інвестиції, здається, позитивно впливають на прибутковість фільму.

Також якщо будемо говорити про візуалізацію будь-яких даних то можемо створити це на прикладі жанрів фільму

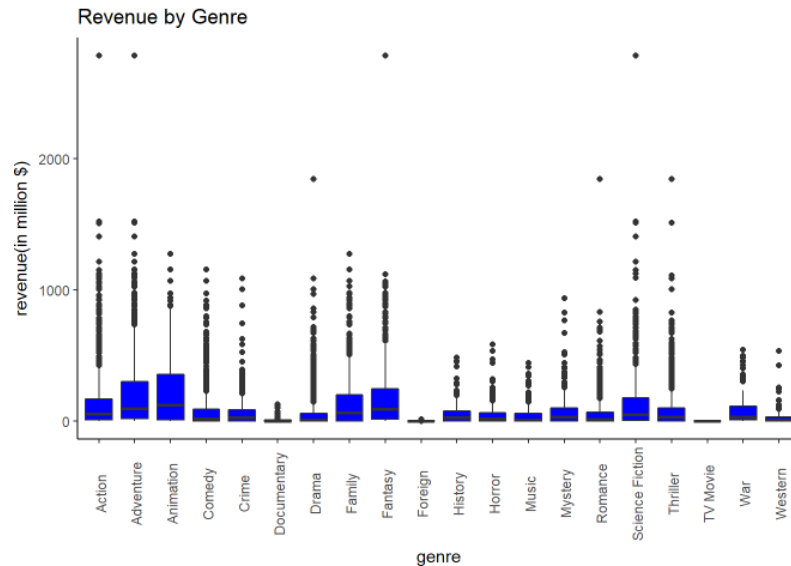


Рисунок 4 – Візуалізація даних щодо жанрів

Ми бачимо, що середні доходи жанрів анімації, пригод та фентезі є найвищими. Документальні та іноземні фільми, як правило, заробляють найменше.

Висновок : Отже використання аналізу даних дає можливість візуалізувати дані та створити діаграми для більш наглядного зберігання даних. Даний дата-сет дає нам можливість створення статистики і розуміння того які фільми є прибутковими, які жанри подобаються людям, скільки потрібно часу для фільму аби він не набридав глядачеві тощо.

#### Список літературних джерел

1. Сніжинський М.В., Нескородева Т.В. Аналіз даних про рейтинг фільмів на платформі оцінювання IMDB засобами мови R. Матеріали всеукраїнської науково-практичної конференції для студентів, аспірантів та молодих вчених "Прикладні інформаційні технології" (29 квітня 2020 року) - Вінниця: ДонНУ імені Василя Стуса. С.57- 58.
2. «Movies» статистика – [Електронний ресурс]. Режим доступу: <https://www.imdb.com/interfaces/>
3. Джеймс Г., Уиттон Д., Хасті Т., Тибишани Р. Введение в статистическое обучение с примерами на языке R. Пер. с англ. С. Э. Мاستицкого - М.: ДМК Пресс, 2017. – 456 с.