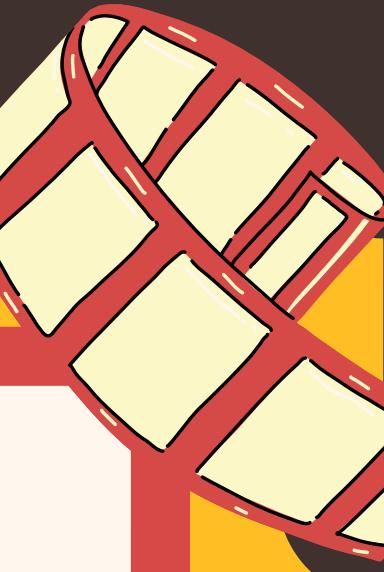


PHÂN TÍCH VÀ ĐỀ XUẤT CHO DOANH NGHIỆP KINH DOANH DỊCH VỤ CHIẾU PHIM



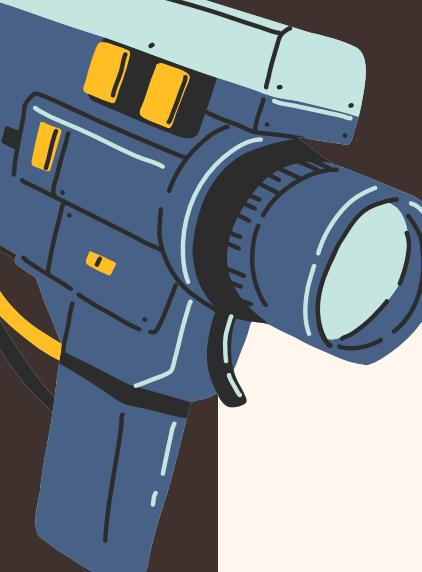
1

TỔNG QUAN ĐỀ TÀI

2

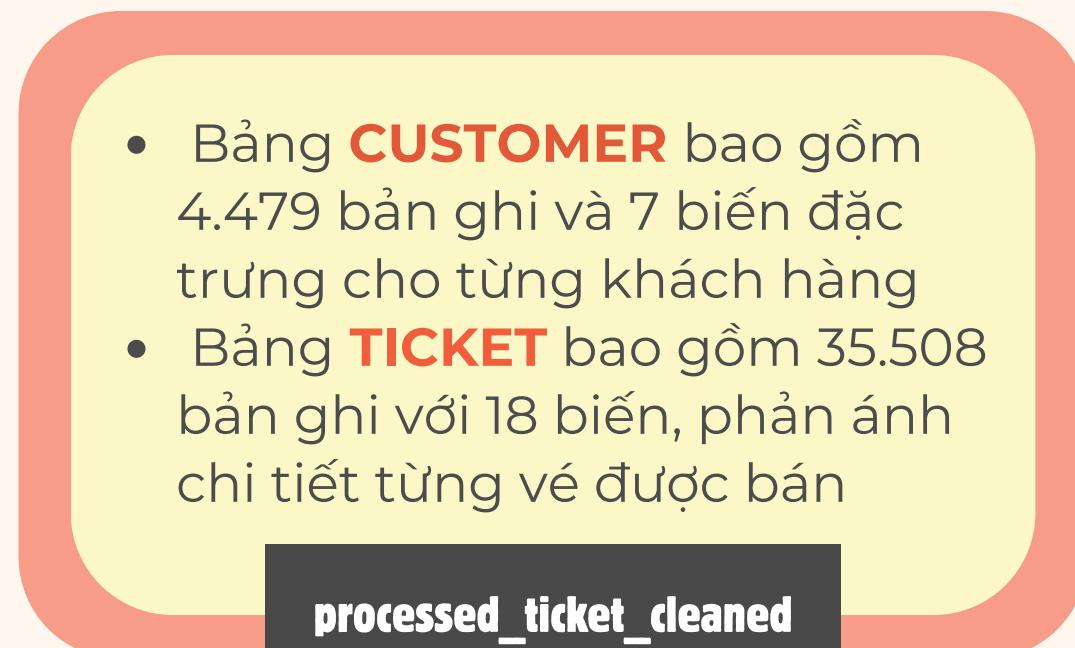
KHÁM PHÁ TẬP DỮ LIỆU



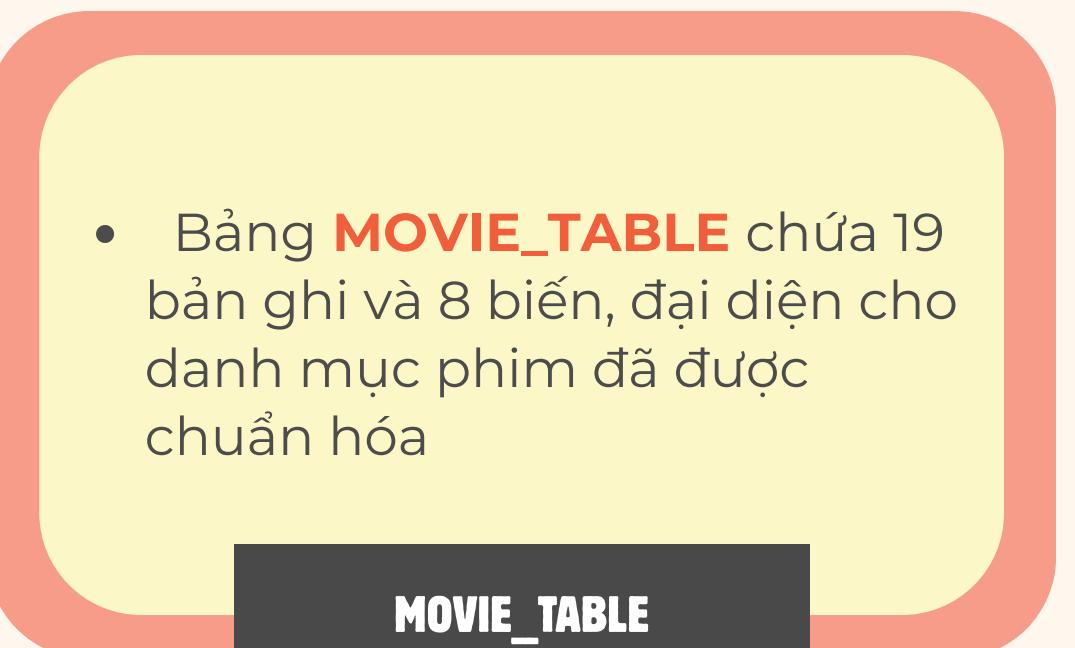


TẬP DỮ LIỆU

- Dữ liệu bán hàng thu thập trực tiếp từ hệ thống bán vé và quản lý phim tại 1 rạp phim ở **Đà Nẵng (T5/2019)**
- Gồm 2 tệp: **processed_ticket_cleaned** và **MOVIE_TABLE**

- 
- Bảng **CUSTOMER** bao gồm 4.479 bản ghi và 7 biến đặc trưng cho từng khách hàng
 - Bảng **TICKET** bao gồm 35.508 bản ghi với 18 biến, phản ánh chi tiết từng vé được bán

processed_ticket_cleaned

- 
- Bảng **MOVIE_TABLE** chứa 19 bản ghi và 8 biến, đại diện cho danh mục phim đã được chuẩn hóa

MOVIE_TABLE

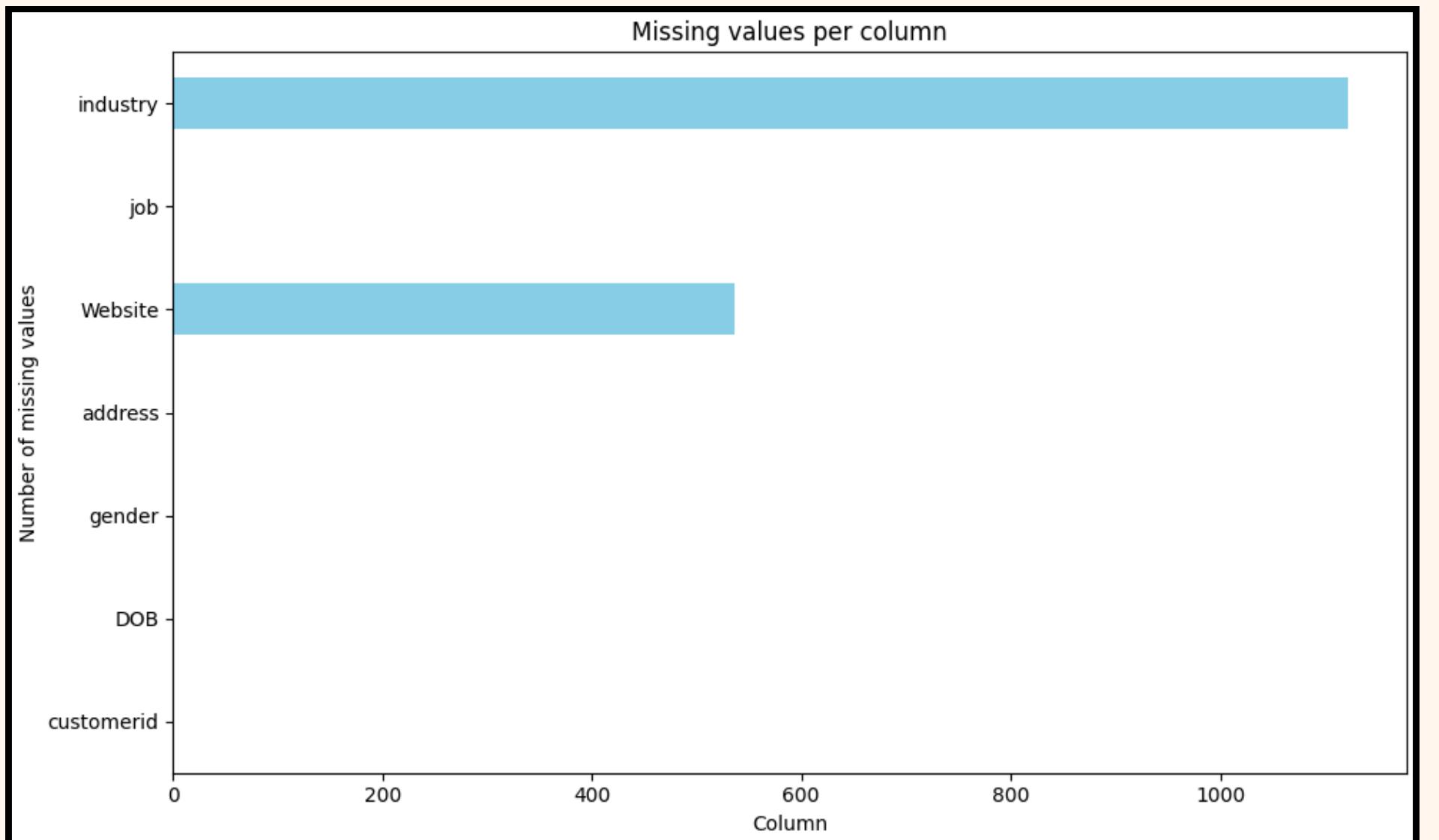


KHÁM PHÁ DỮ LIỆU - EDA

EDA CUSTOMER TABLE

KIỂM TRA MISSING VALUES

*Biểu đồ số lượng giá trị
thiếu theo cột*

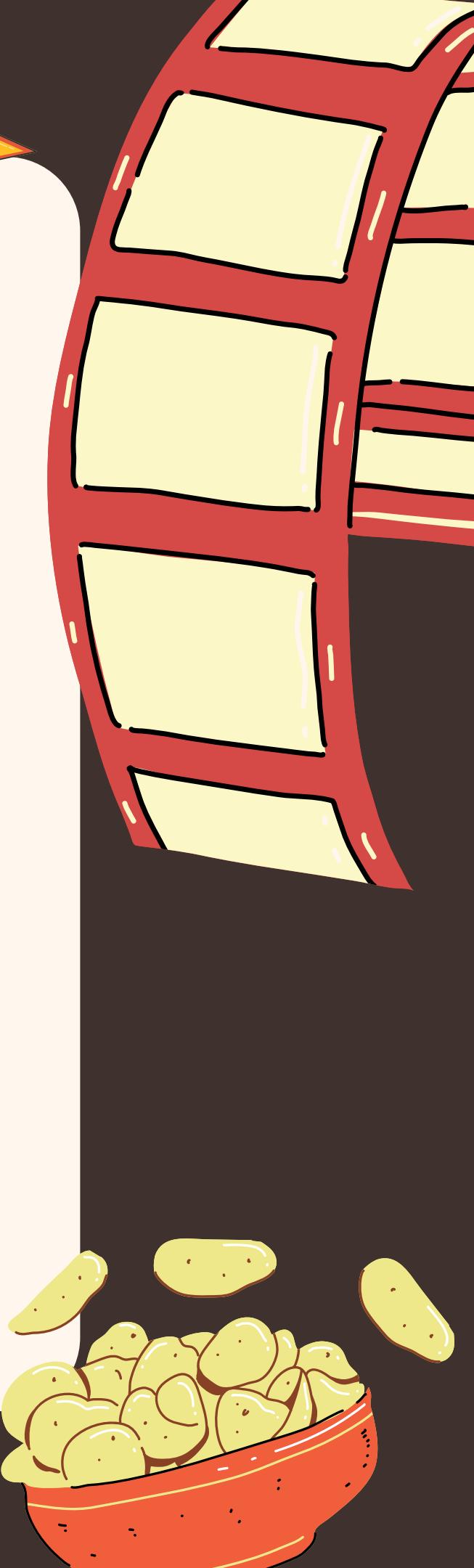
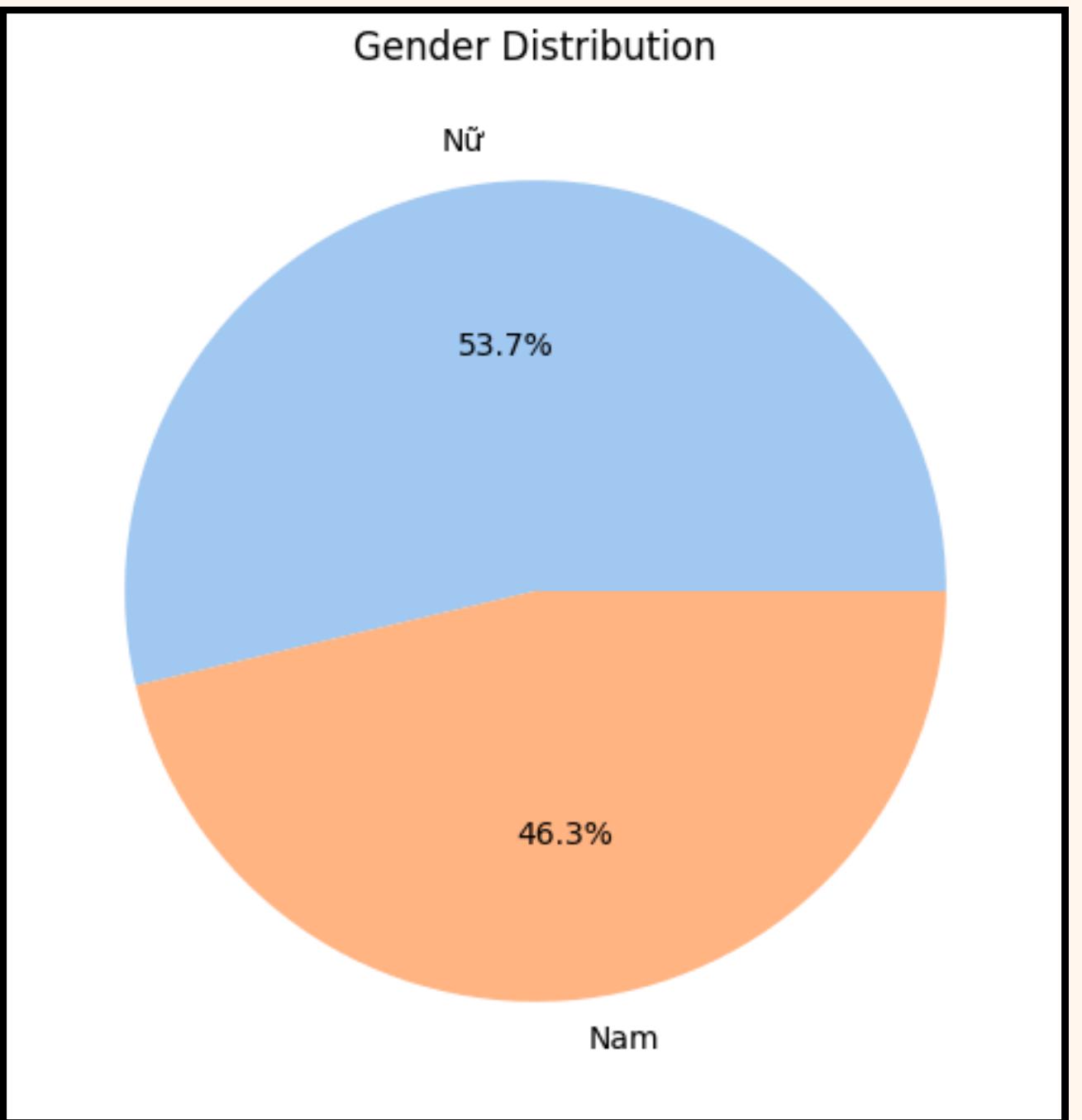


KHÁM PHÁ DỮ LIỆU - EDA

EDA CUSTOMER TABLE

PHÂN PHỐI BIẾN GENDER

Biểu đồ phân phối của biến Gender

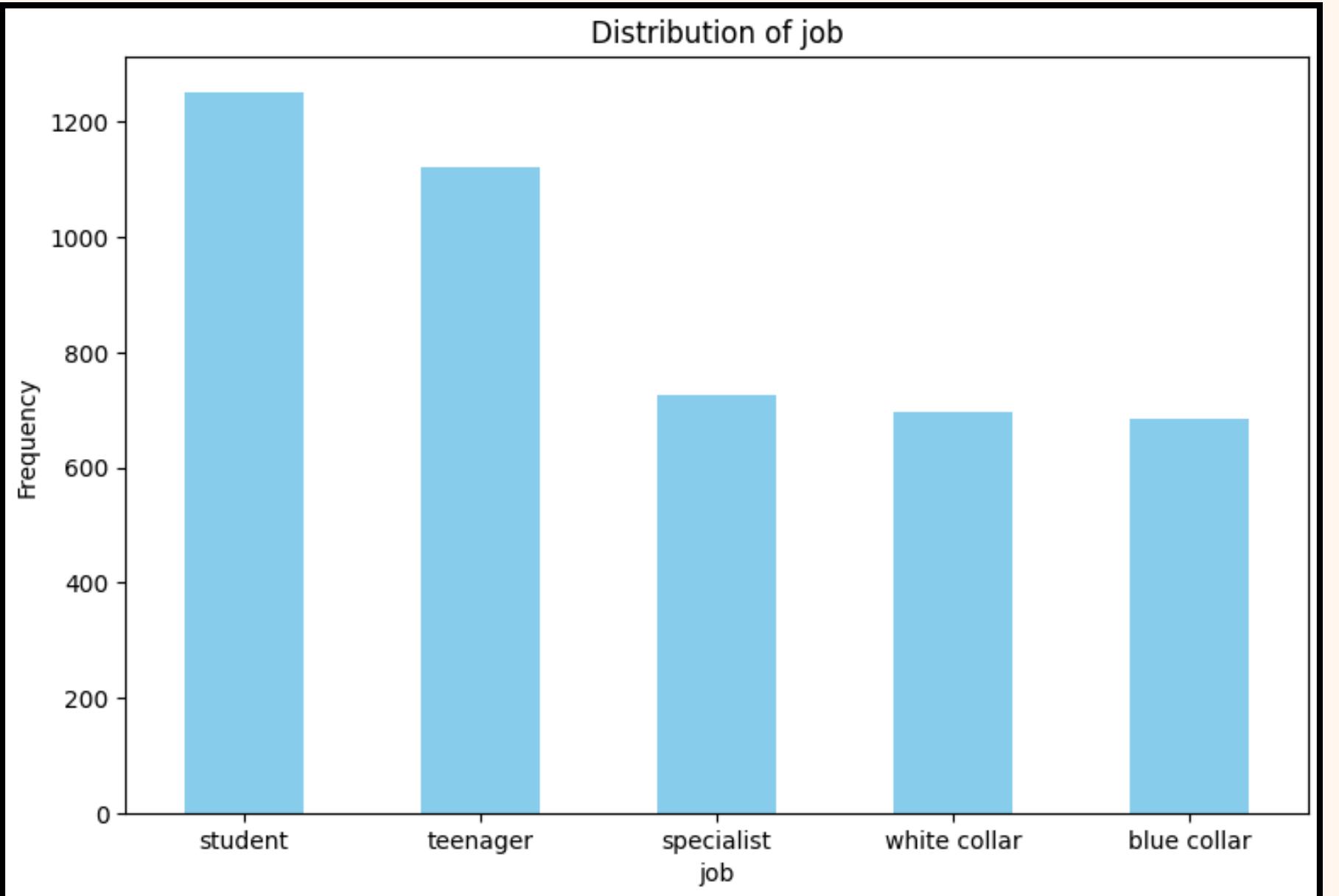


KHÁM PHÁ DỮ LIỆU - EDA

EDA CUSTOMER TABLE

PHÂN PHỐI BIẾN JOB

Biểu đồ phân phối của
biến Job



KHÁM PHÁ DỮ LIỆU - EDA

EDA TICKET TABLE

KIỂM TRA MISSING VALUE

Số lượng giá trị thiếu
của từng biến

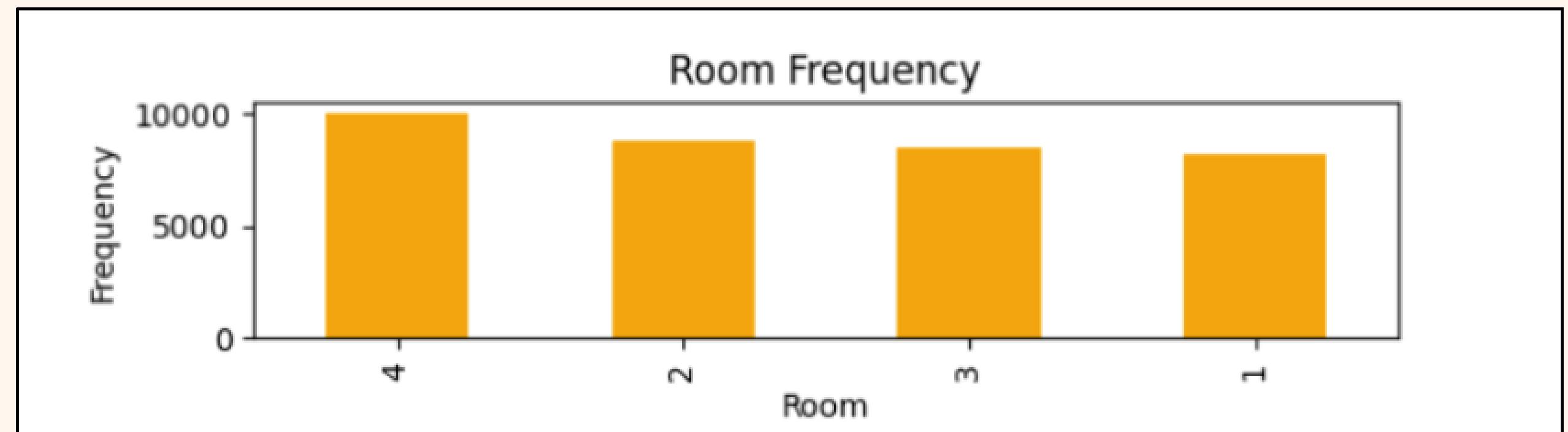
Missing Values:	
orderid	0
cashier	0
saledate	96
customerid	0
date	0
time	0
room	0
film	0
ticketcode	0
slot	0
ticket price	0
total	0
popcorn	96
slot type	0
ticket type	0
discount	0
have_discount	0
dtype: int64	



KHÁM PHÁ DỮ LIỆU - EDA

EDA TICKET TABLE

TẦN SUẤT PHÒNG CHIẾU



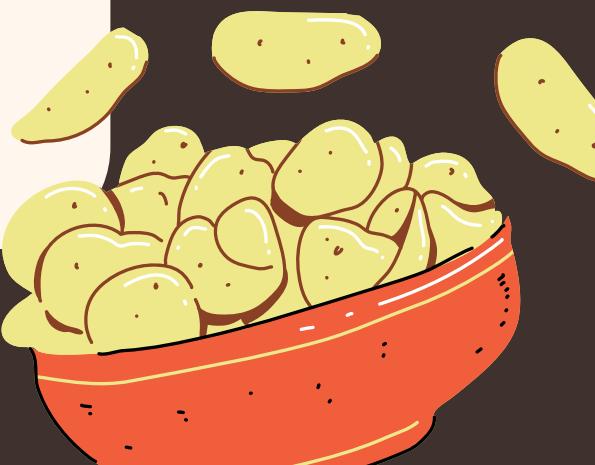
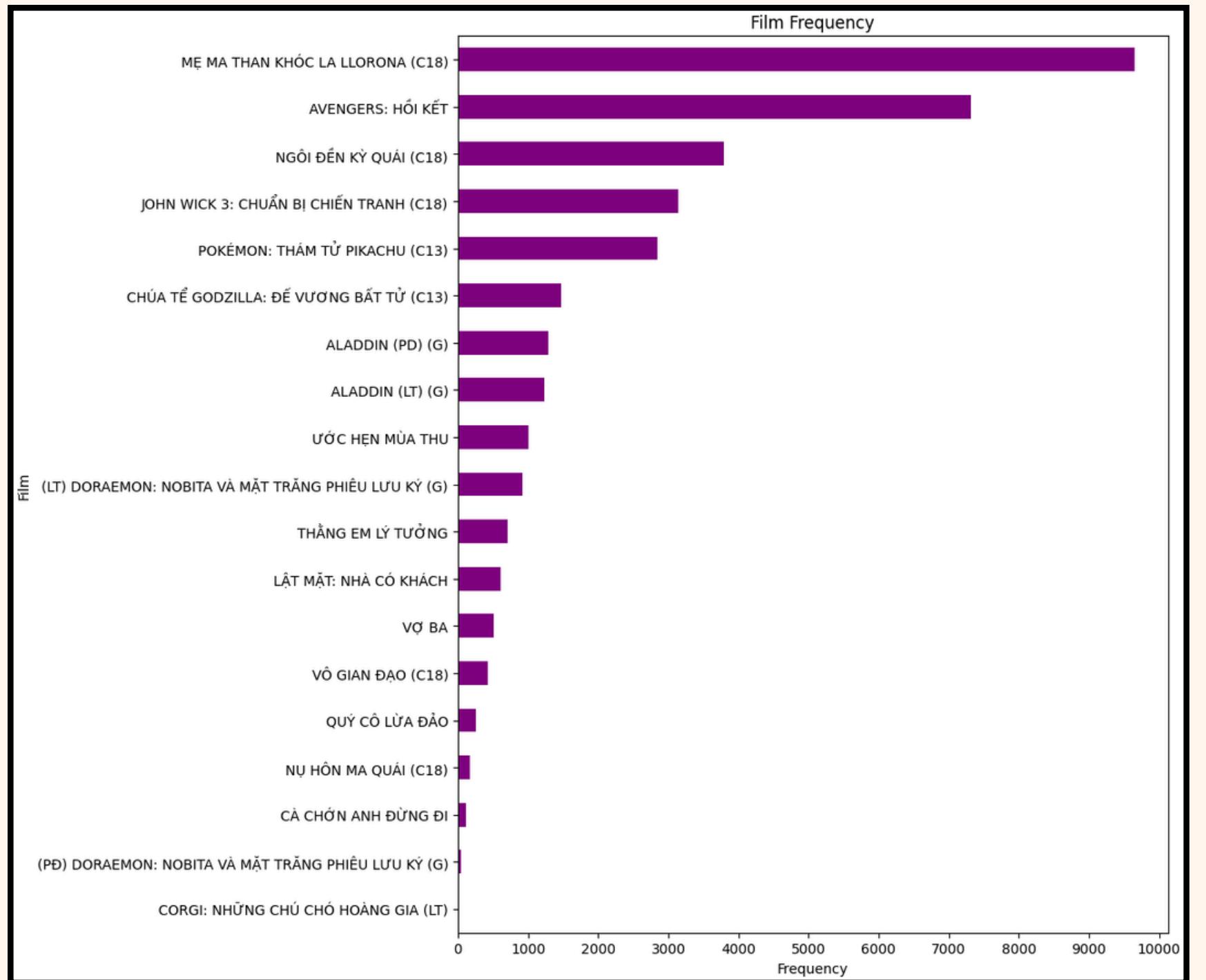
Biểu đồ tần suất sử dụng phòng chiếu

KHÁM PHÁ DỮ LIỆU - EDA

EDA TICKET TABLE

TẦN SUẤT PHIM

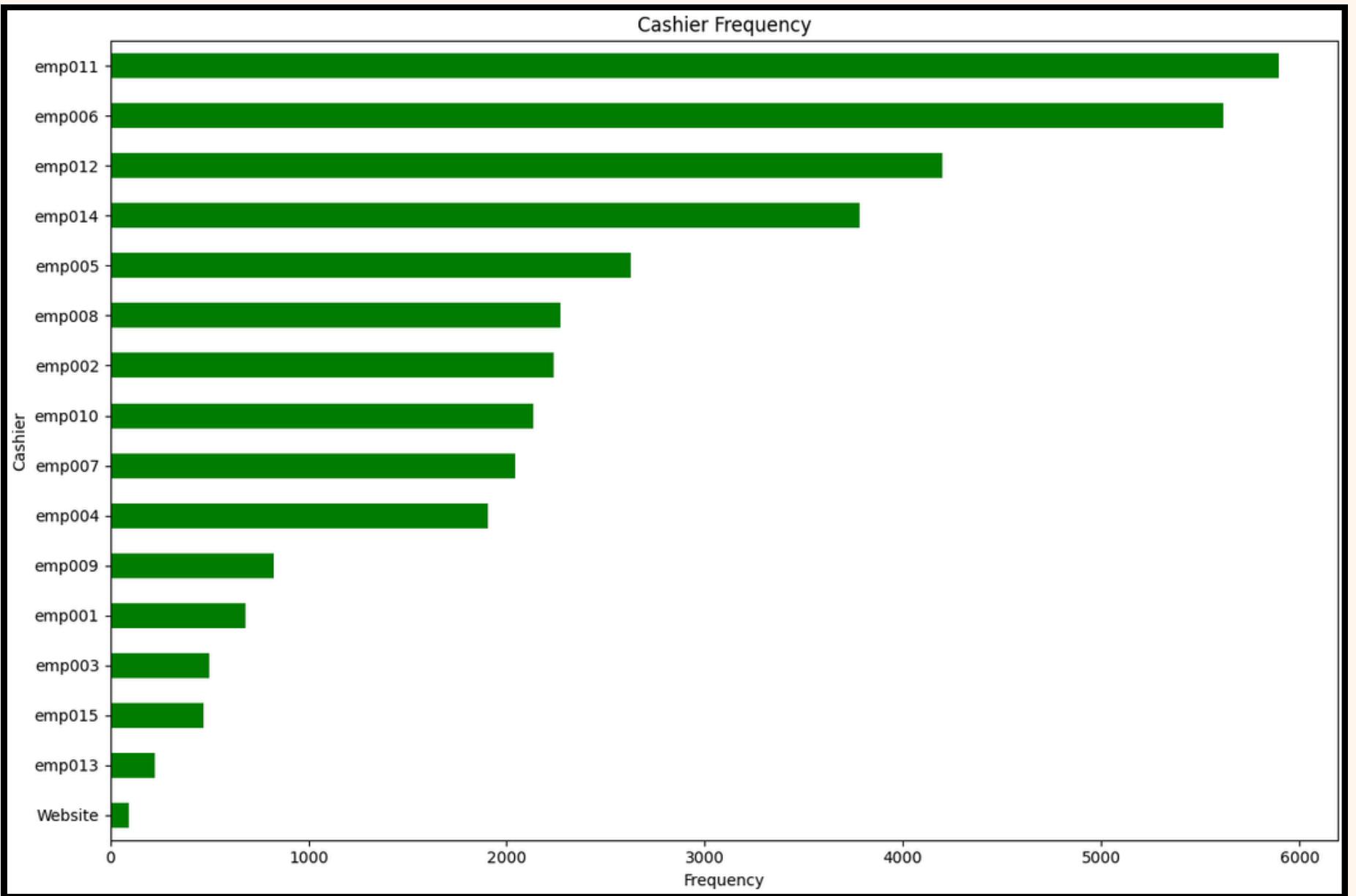
Biểu đồ phân phối tần suất chiếu phim



KHÁM PHÁ DỮ LIỆU - EDA

EDA TICKET TABLE

TẦN SUẤT NGƯỜI BÁN VÉ



Phân phối tần suất hoạt động của nhân
viên thu ngân

TIỀN XỬ LÝ DỮ LIỆU

VẤN ĐỀ

Một số đơn hàng có tổng giá vé (ticket_price) không khớp với tổng tiền (total) trong đơn hàng (orderid)

NGUYÊN NHÂN

- Lỗi nhập liệu
- Sai sót trong tính toán
- Dữ liệu bị cập nhật thiếu đồng bộ

XỬ LÝ BẰNG TICKET

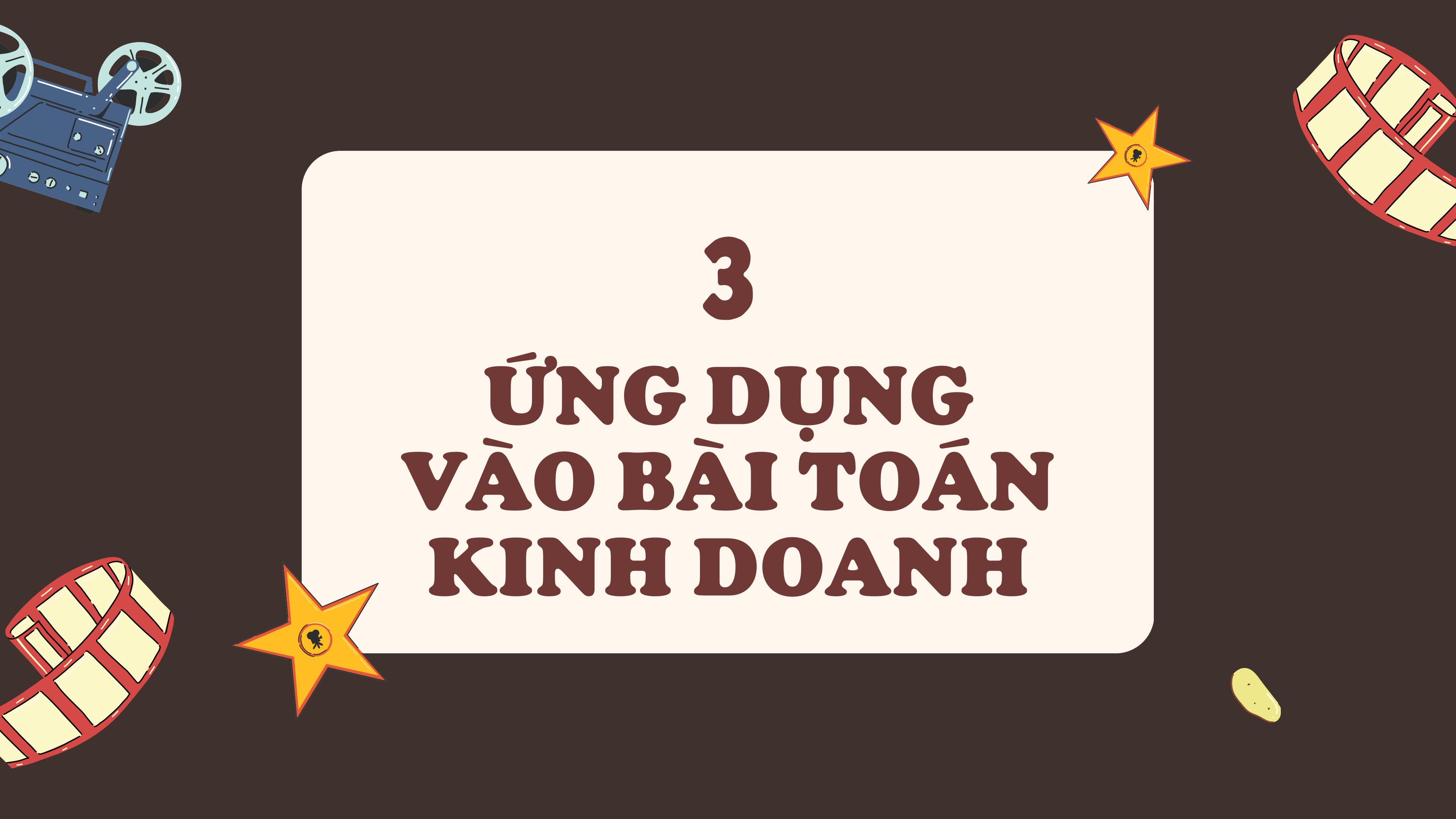
GIẢI PHÁP

- Dùng Python + pandas để kiểm tra và xử lý dữ liệu.
- Vé được nhóm theo **orderid**, tính lại tổng **ticket_price**, so sánh với **total**.
- Bổ sung dòng bị thiếu để đảm bảo tổng giá vé khớp với total.



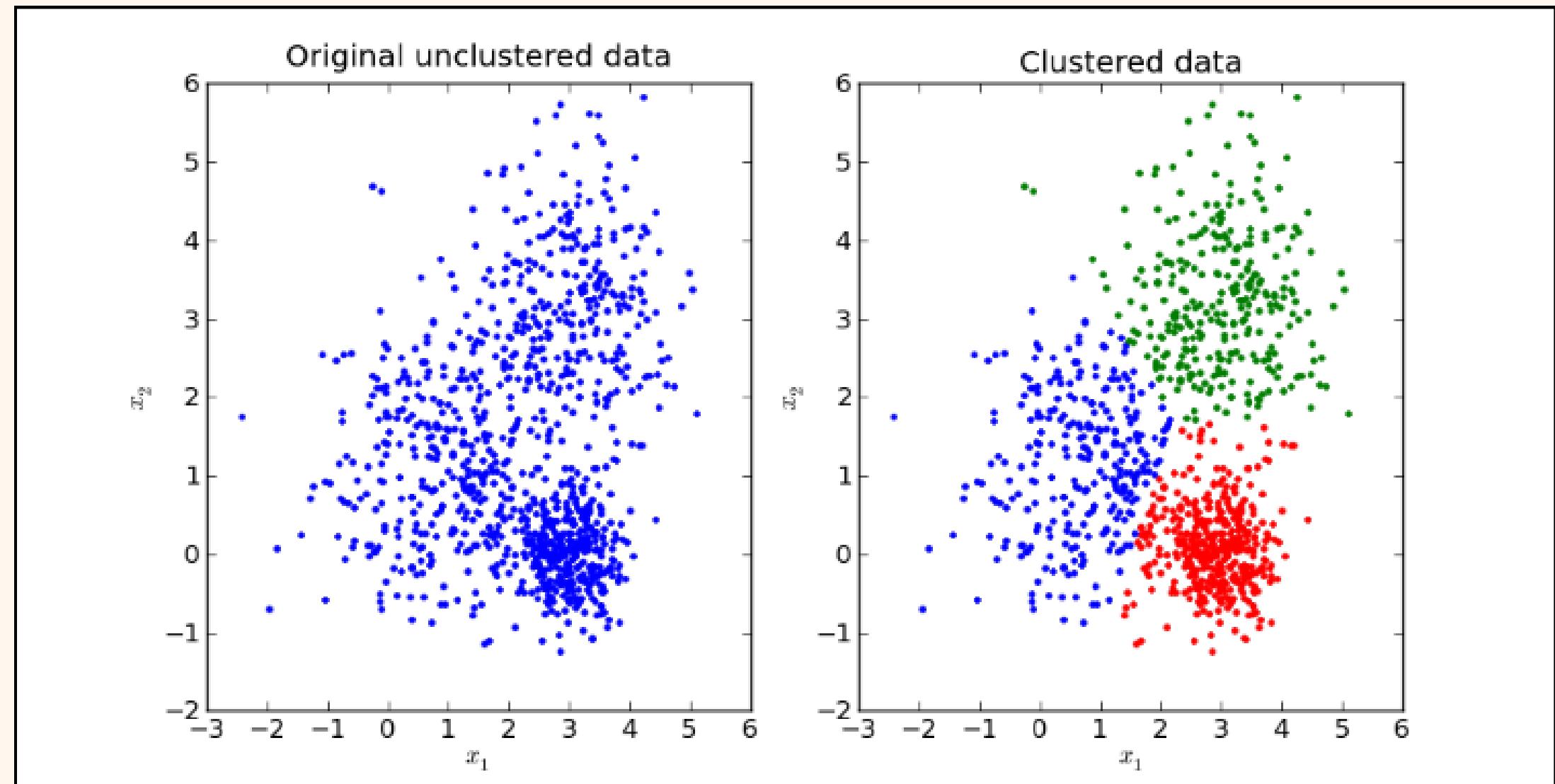
3

ỨNG DỤNG VÀO BÀI TOÁN KINH DOANH



LỰA CHỌN PHƯƠNG PHÁP

- Sử dụng 2 thuật toán phân cụm: **K-Means** và **DBSCAN**
- Sau đó, tiến hành phân tích và so sánh



Minh họa cho mô hình
phân cụm
(Nguồn: Mubaris)

QUY TRÌNH HUẤN LUYỆN MÔ HÌNH

XỬ LÝ DỮ LIỆU

TÍNH CHỈ SỐ
RFM

K-MEANS &
DBSCAN

- Lọc cột không liên quan, xử lý giá trị thiếu và bất thường.
- Chuyển ngày sinh → tuổi, mã hóa biến phân loại
- Chuẩn hóa các biến liên tục bằng StandardScaler.

- **Recency:** Ngày hiện tại – ngày giao dịch gần nhất.
- **Frequency:** Đếm số vé đã mua của từng khách.
- **Monetary:** Tổng số tiền đã chi tiêu.

K-MEANS

- Xác định số cụm tối ưu bằng Elbow Method
- Phân nhóm KH dựa trên RFM và đặc trưng cá nhân.

DBSCAN

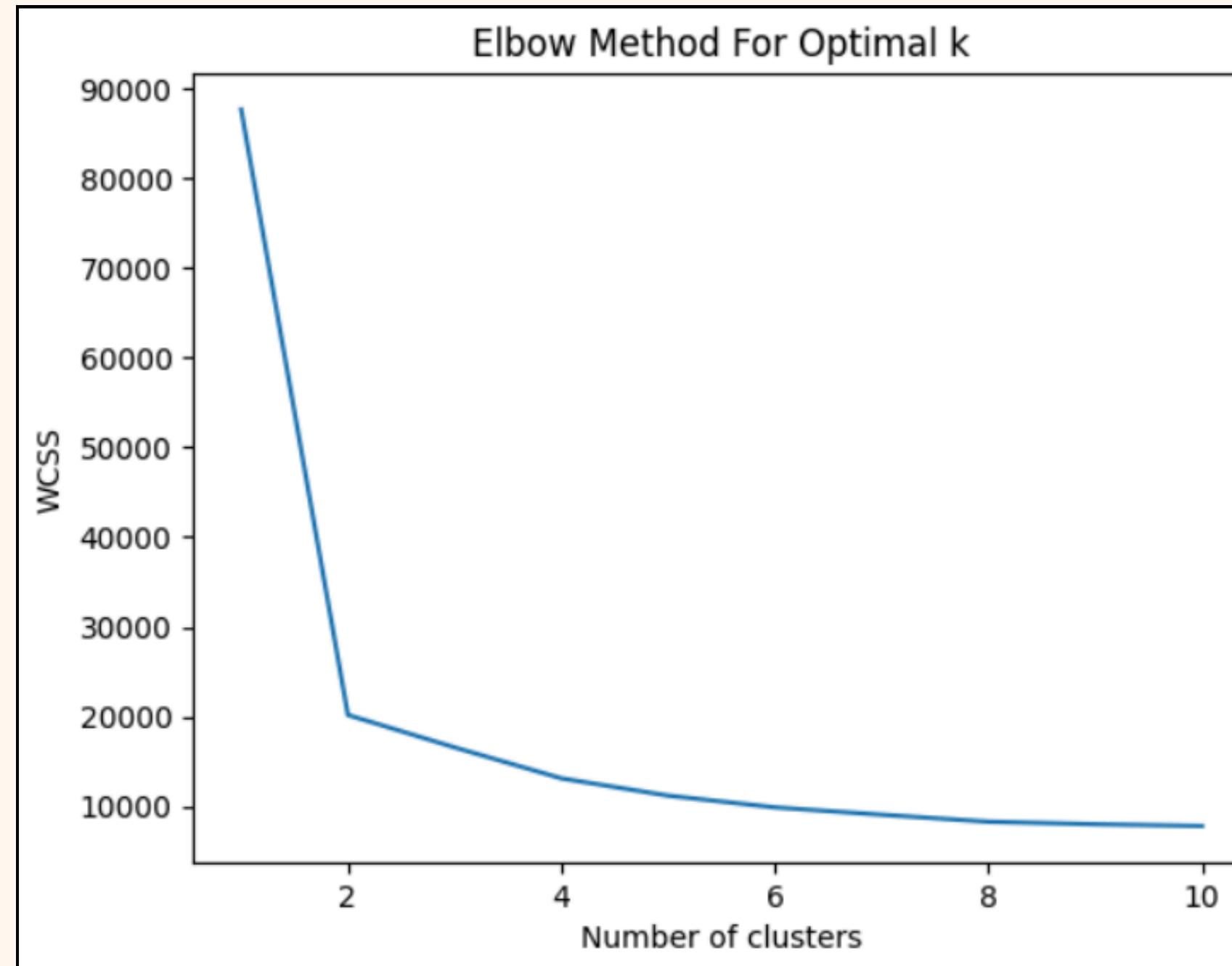
- Hiệu chỉnh siêu tham số
- Đánh giá chất lượng cụm bằng Silhouette Score.
- Loại trừ và ghi nhận kết quả

4

KẾT QUẢ MÔ HÌNH

KẾT QUẢ MÔ HÌNH

Biểu đồ Elbow – Chọn số cụm tối ưu



- Sử dụng phương pháp Elbow với số cụm từ 1 đến 10.
- Tính WCSS (Within-Cluster Sum of Squares) cho từng k.
- Elbow cho thấy k = 2 là tối ưu về kỹ thuật, nhưng không phù hợp với mục tiêu bài toán.

**XÁC ĐỊNH SỐ CỤM
TỐI ƯU**

- Chọn **k = 4**
- Kết quả phân cụm được gán vào cột cluster trong dữ liệu
- Gộp kết quả phân cụm với dữ liệu gốc để
 - + Phân tích sâu từng nhóm
 - + Thống kê từng cụm.

**ỨNG DỤNG K-
MEANS**

KẾT QUẢ MÔ HÌNH

--- THỐNG KÊ CHO CLUSTER 0 ---						--- THỐNG KÊ CHO CLUSTER 1 ---					
	Tổng	Trung bình	Min	Max	Mode		Tổng	Trung bình	Min	Max	Mode
Recency	48272	21.646637	11	30	26	Recency	0	0.0	0	0	0
Frequency	3095	1.387892	1	18	1	Frequency	4043	4043.0	4043	4043	4043
Monetary	356075000	159674.887892	45000	4530000	90000	Monetary	490635000	490635000.0	490635000	490635000	490635000
action	341	0.152915	0	12	0	action	885	885.0	885	885	885
action-adventure	3728	1.671749	0	57	0	action-adventure	4007	4007.0	4007	4007	4007
animation	0	0.000000	0	0	0	animation	244	244.0	244	244	244
comedy	1447	0.648879	0	27	0	comedy	2911	2911.0	2911	2911	2911
crime	418	0.187444	0	27	0	crime	1074	1074.0	1074	1074	1074
drama	262	0.117489	0	12	0	drama	401	401.0	401	401	401
family	638	0.286099	0	13	0	family	1694	1694.0	1694	1694	1694
fantasy	3090	1.385650	0	57	0	fantasy	3085	3085.0	3085	3085	3085
horror	2803	1.256951	0	32	0	horror	4329	4329.0	4329	4329	4329
melodrama	34	0.016247	0	4	0	melodrama	149	149.0	149	149	149
mystery	3186	1.428700	0	32	0	mystery	4235	4235.0	4235	4235	4235
romance	249	0.111659	0	12	0	romance	429	429.0	429	429	429
sci-fi	3728	1.671749	0	57	0	sci-fi	3192	3192.0	3192	3192	3192
thriller	2923	1.310762	0	32	0	thriller	4481	4481.0	4481	4481	4481
age	61410	27.538117	6	60	25	age	6	6.0	6	6	6
cluster	0	0.000000	0	0	0	cluster	1	1.0	1	1	1

Mô tả kết quả
phân cụm
KMeans

KẾT QUẢ MÔ HÌNH

--- THỐNG KÊ CHO CLUSTER 2 ---						--- THỐNG KÊ CHO CLUSTER 3 ---					
	Tổng	Trung bình	Min	Max	Mode		Tổng	Trung bình	Min	Max	Mode
Recency	0	0.0	0	0	0	Recency	13491	6.004005	0	20	0
Frequency	2854	2854.0	2854	2854	2854	Frequency	4649	2.068981	1	206	1
Monetary	330290000	330290000.0	330290000	330290000	330290000	Monetary	509815000	226886.960392	45000	23310000	90000
action	469	469.0	469	469	469	action	1442	0.641745	0	40	0
action-adventure	3037	3037.0	3037	3037	3037	action-adventure	4352	1.936805	0	201	0
animation	361	361.0	361	361	361	animation	370	0.164664	0	33	0
comedy	2496	2496.0	2496	2496	2496	comedy	3881	1.727192	0	221	0
crime	616	616.0	616	616	616	crime	1710	0.761015	0	42	0
drama	184	184.0	184	184	184	drama	541	0.240765	0	16	0
family	1437	1437.0	1437	1437	1437	family	2564	1.141077	0	120	0
fantasy	2622	2622.0	2622	2622	2622	fantasy	3483	1.550067	0	161	0
horror	3079	3079.0	3079	3079	3079	horror	4018	1.768162	0	249	0
melodrama	73	73.0	73	73	73	melodrama	251	0.111704	0	7	0
mystery	2172	2172.0	2172	2172	2172	mystery	2896	1.288830	0	166	0
romance	281	281.0	281	281	281	romance	336	0.149533	0	11	0
sci-fi	2015	2015.0	2015	2015	2015	sci-fi	2196	0.978193	0	134	0
thriller	2421	2421.0	2421	2421	2421	thriller	3893	1.732532	0	158	0
age	24	24.0	24	24	24	age	64983	28.919693	5	61	28
cluster	2	2.0	2	2	2	cluster	6741	3.000000	3	3	3

Mô tả kết quả
phân cụm
KMeans

KẾT QUẢ MÔ HÌNH

--- THỐNG KÊ CHO CLUSTER -1 ---						--- THỐNG KÊ CHO CLUSTER 0 ---					
	Tổng	Trung bình	Min	Max	Mode		Tổng	Trung bình	Min	Max	Mode
Recency	502	9.843137e+00	0	30	0	Recency	45985	13.892749	0	30	26
Frequency	7209	1.413529e+02	1	4043	1	Frequency	5498	1.661027	1	14	1
Monetary	863885000	1.693892e+07	45000	490635000	135000	Monetary	585710000	176951.661631	45000	1750000	90000
action	1489	2.880392e+01	0	885	0	action	1304	0.393968	0	11	0
action-adventure	7446	1.460000e+02	0	4007	0	action-adventure	5831	1.761631	0	22	0
animation	651	1.276471e+01	0	361	0	animation	229	0.069184	0	10	0
comedy	5685	1.114706e+02	0	2911	0	comedy	3512	1.061027	0	19	0
crime	1822	3.572549e+01	0	1074	0	crime	1538	0.464653	0	11	0
drama	644	1.262745e+01	0	401	0	drama	487	0.147130	0	7	0
family	3332	6.533333e+01	0	1694	0	family	2178	0.658006	0	14	0
fantasy	6059	1.188039e+02	0	3085	0	fantasy	4778	1.443505	0	22	0
horror	7690	1.507843e+02	0	4329	0	horror	4191	1.266163	0	32	0
melodrama	267	5.235294e+00	0	149	0	melodrama	171	0.051662	0	4	0
mystery	6609	1.295882e+02	0	4235	0	mystery	3764	1.137160	0	32	0
romance	731	1.433333e+01	0	429	0	romance	362	0.109366	0	8	0
sci-fi	5461	1.070784e+02	0	3192	0	sci-fi	4296	1.297885	0	22	0
thriller	7217	1.415098e+02	0	4481	0	thriller	4320	1.305136	0	32	0
age	2314	4.537255e+01	6	61	53	age	98812	29.852568	25	50	27
cluster	-51	-1.000000e+00	-1	-1	-1	cluster	0	0.000000	0	0	0

Mô tả kết quả
phân cụm
DBSCAN

KẾT QUẢ MÔ HÌNH

--- THỐNG KÊ CHO CLUSTER 1 ---						--- THỐNG KÊ CHO CLUSTER 2 ---					
	Tổng	Trung bình	Min	Max	Mode		Tổng	Trung bình	Min	Max	Mode
Recency	14975	13.725940	0	30	26	Recency	301	11.148148	0	29	0.0
Frequency	1883	1.725940	1	19	1	Frequency	51	1.888889	1	8	1.0
Monetary	231505000	212195.233731	45000	2340000	90000	Monetary	57150000	211666.666667	45000	1125000	90000.0
action	351	0.321723	0	18	0	action	13	0.481481	0	3	0.0
action-adventure	1782	1.633364	0	31	0	action-adventure	65	2.407407	0	19	0.0
animation	95	0.087076	0	8	0	animation	0	0.000000	0	0	0.0
comedy	1505	1.379468	0	27	0	comedy	33	1.222222	0	8	0.0
crime	445	0.407883	0	27	0	crime	13	0.481481	0	3	0.0
drama	252	0.230981	0	16	0	drama	5	0.185185	0	2	0.0
family	793	0.726856	0	20	0	family	30	1.111111	0	8	0.0
fantasy	1386	1.270394	0	23	0	fantasy	57	2.111111	0	17	0.0
horror	2313	2.120073	0	22	0	horror	35	1.296296	0	12	0.0
melodrama	68	0.062328	0	6	0	melodrama	1	0.037037	0	1	0.0
mystery	2094	1.919340	0	25	0	mystery	22	0.814815	0	6	0.0
romance	196	0.179652	0	12	0	romance	6	0.222222	0	2	0.0
sci-fi	1338	1.226398	0	31	0	sci-fi	38	1.407407	0	13	0.0
thriller	2153	1.973419	0	24	0	thriller	28	1.037037	0	6	0.0
age	25100	23.006416	16	29	24	age	197	7.296296	5	9	7.0
cluster	1091	1.000000	1	1	1	cluster	54	2.000000	2	2	2.0

Mô tả kết quả
phân cụm
DBSCAN

DASHBOARD

OVERVIEW DASHBOARD

MOVIE THEATRE OVERVIEW DASHBOARD



Total Revenue

1688M

Total Tickets

35.51K

Website Rate

0.27%

Popcorn Rate

0.35%

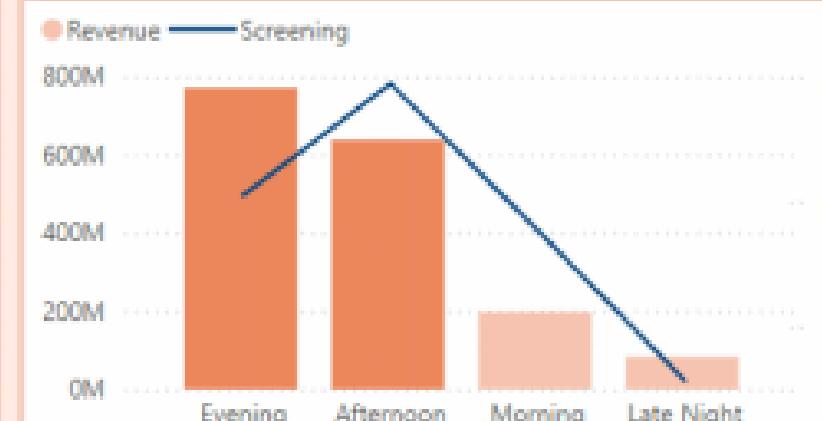
Screenings

733

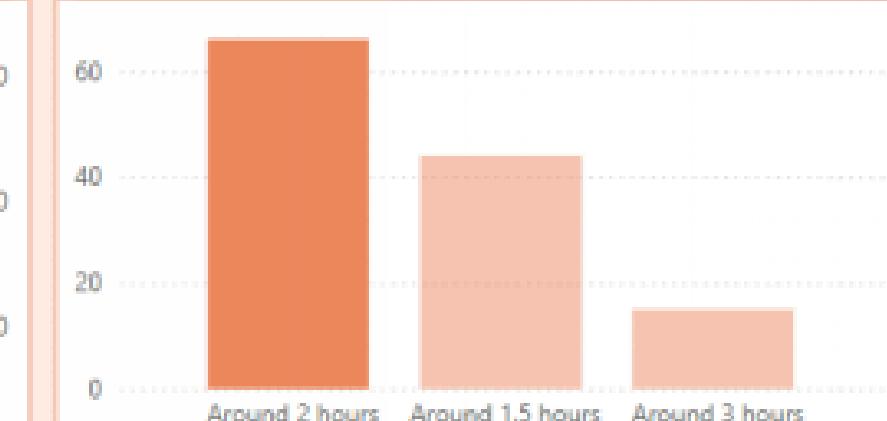
REVENUE AND SCREENINGS BY DATE OF MONTH



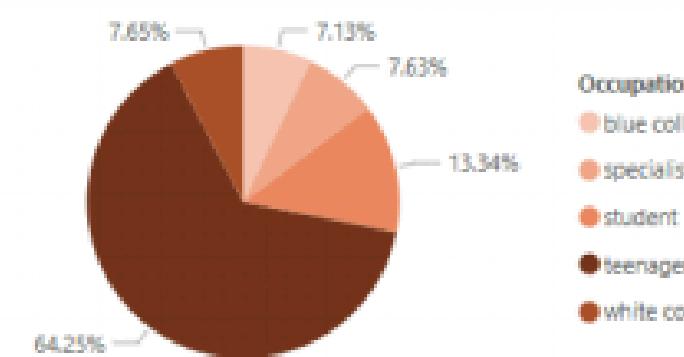
REVENUE AND SCREENINGS BY TIME OF DAY



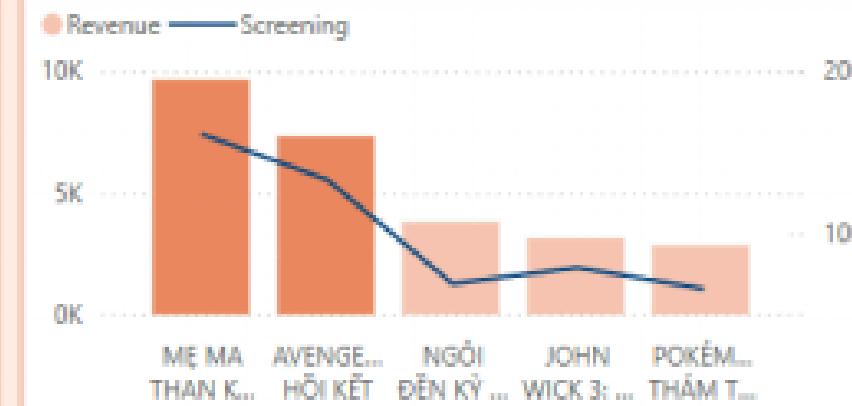
POPCORN SALES BY DURATION TYPE



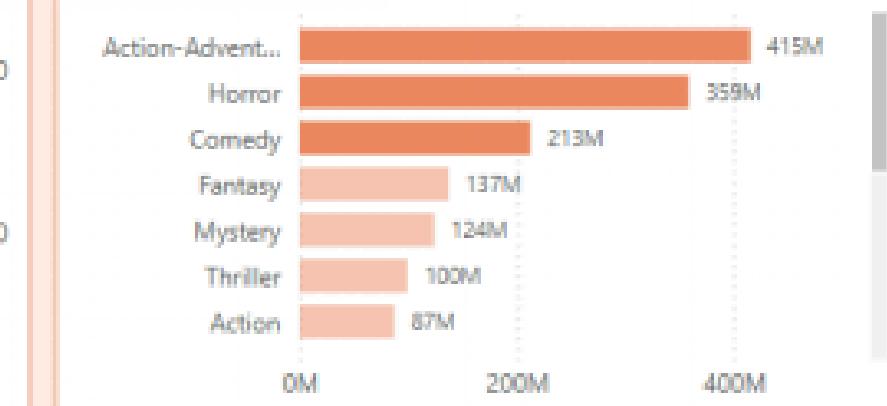
TICKET REVENUE DISTRIBUTION BY OCCUPATION



TOP 5 MOVIES BY TICKET REVENUE & SCREENING

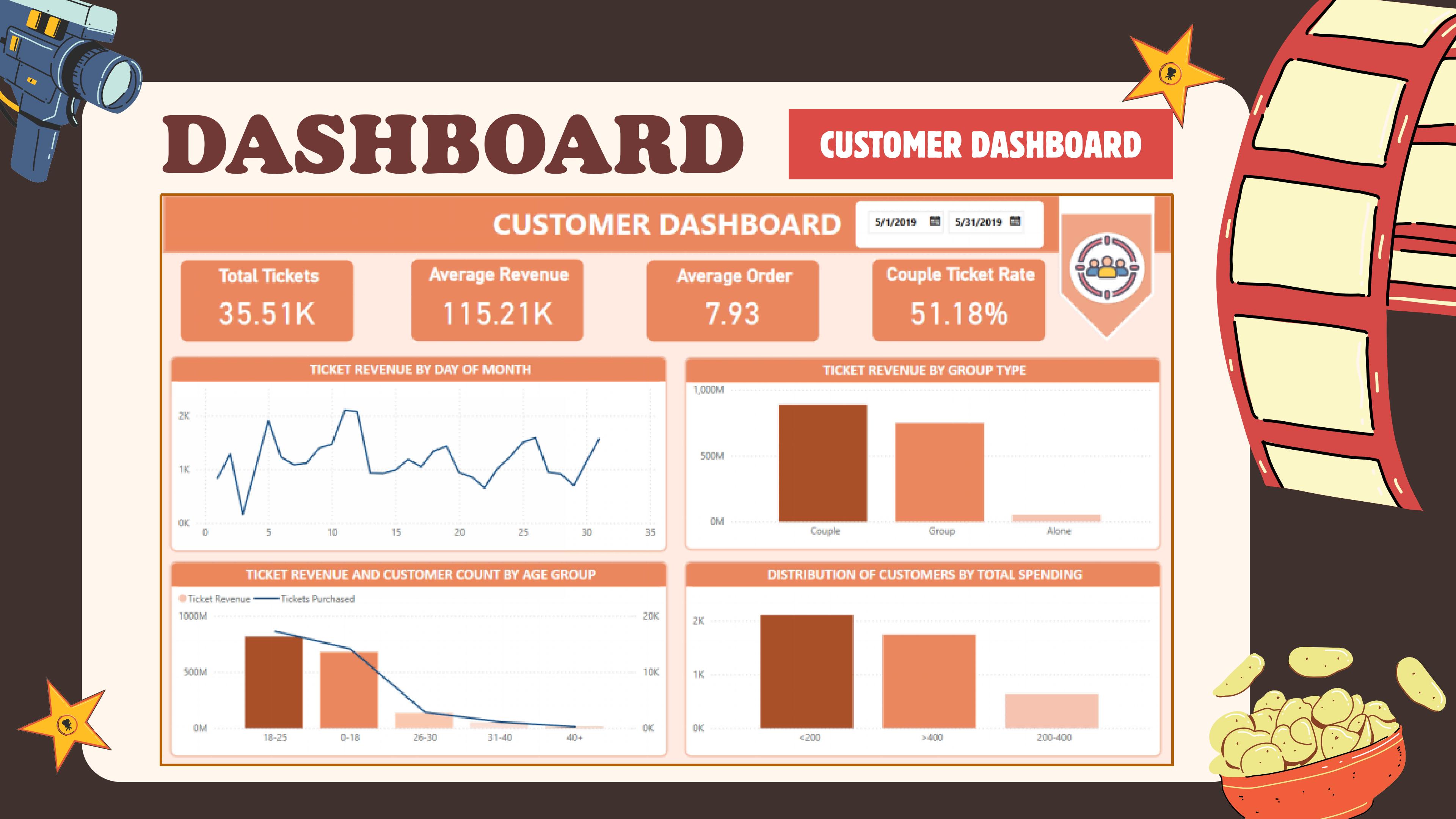
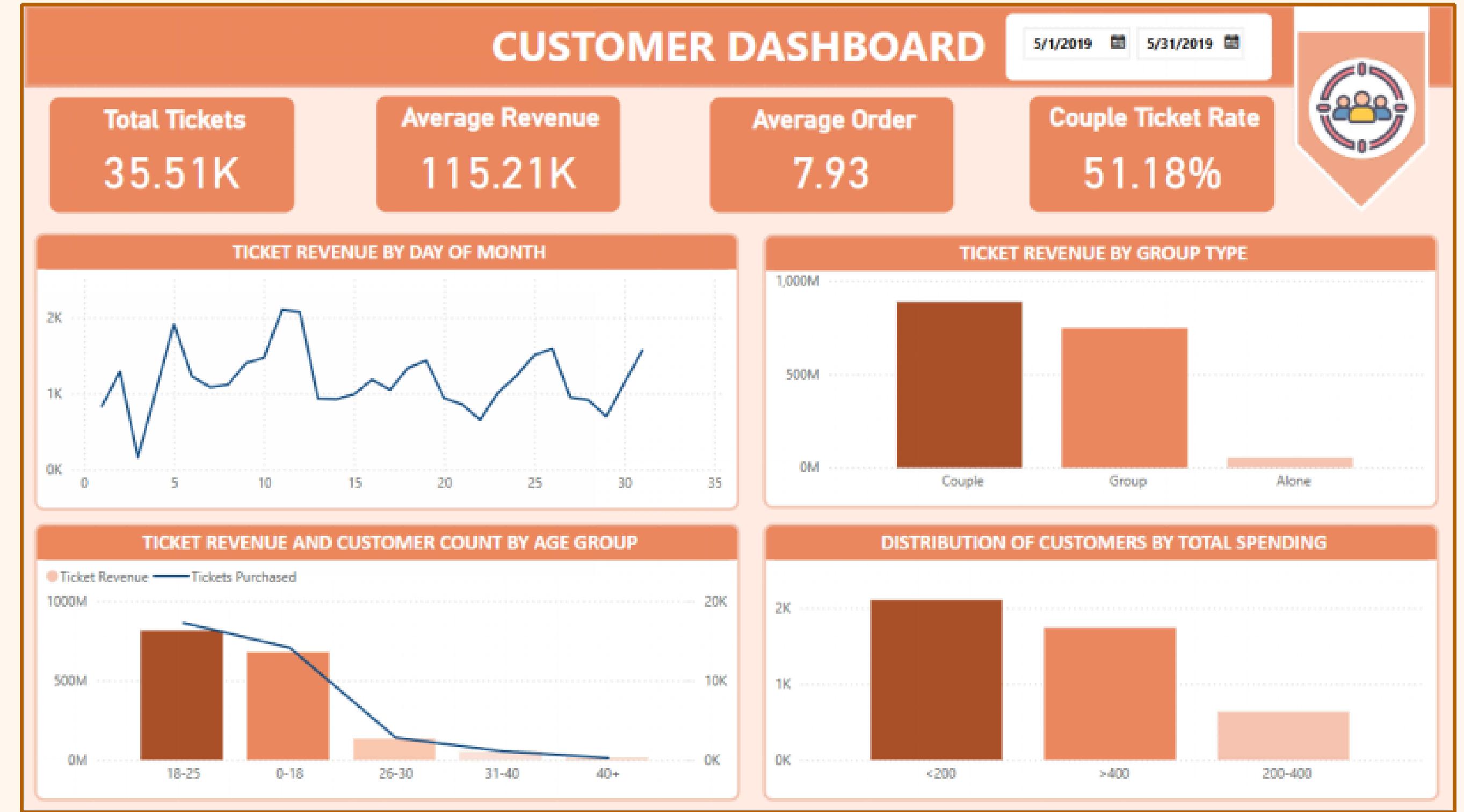


REVENUE BY MOVIES GENRES



DASHBOARD

CUSTOMER DASHBOARD



DASHBOARD

CUSTOMER DASHBOARD

SCHEDULING DASHBOARD

TICKET REVENUE & VIEWING HABITS BY OCCUPATION

TICKET REVENUE BY DAY OF THE WEEK

Occupation	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
blue collar	0.83%	0.65%	0.81%	1.19%	0.92%	0.92%	1.79%
specialist	0.87%	0.76%	0.93%	1.22%	1.06%	1.01%	1.73%
student	1.75%	1.58%	1.39%	2.36%	2.02%	1.53%	2.67%
teenager	7.15%	6.82%	8.13%	10.94%	10.11%	9.47%	11.78%
white collar	0.77%	0.82%	0.81%	1.23%	1.29%	0.98%	1.74%

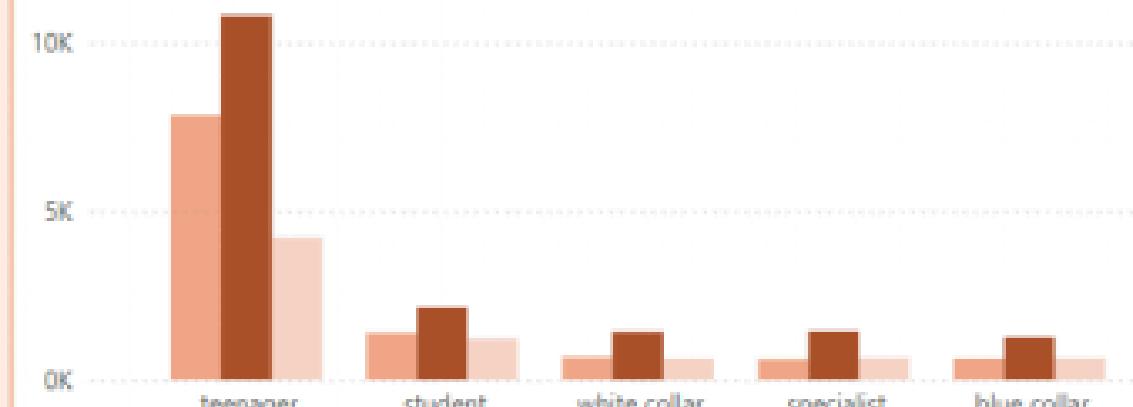
TICKET REVENUE BREAKDOWN BY OCCUPATION & TIME

TICKET REVENUE BY INDUSTRY AND TIME FRAME

Occupation	Morning	Afternoon	Evening	Late Night
blue collar	0.83%	2.21%	3.69%	0.38%
specialist	0.92%	2.20%	4.10%	0.35%
student	1.36%	4.24%	6.83%	0.87%
teenager	7.37%	26.80%	27.41%	2.82%
white collar	1.02%	2.39%	3.92%	0.30%

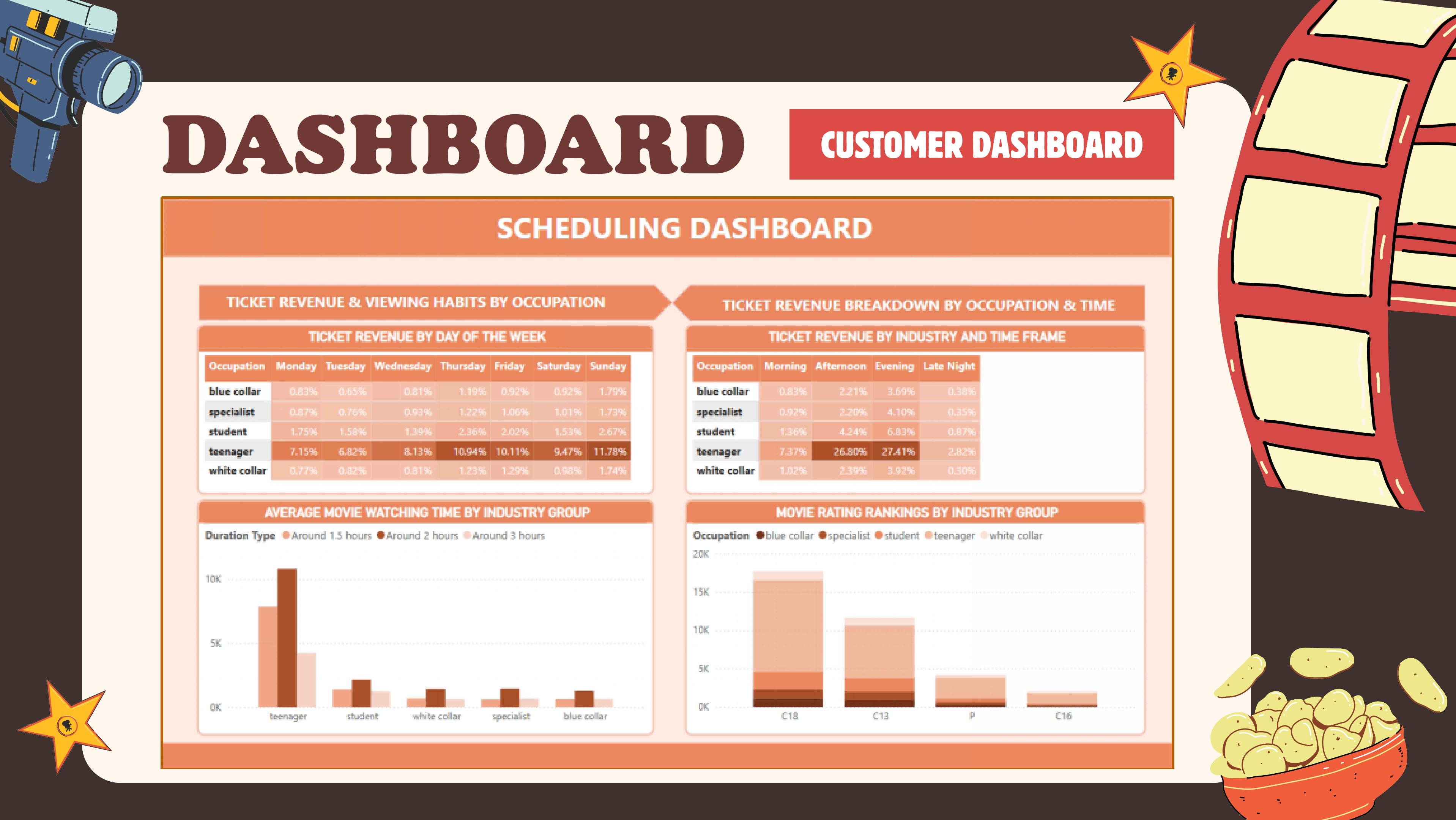
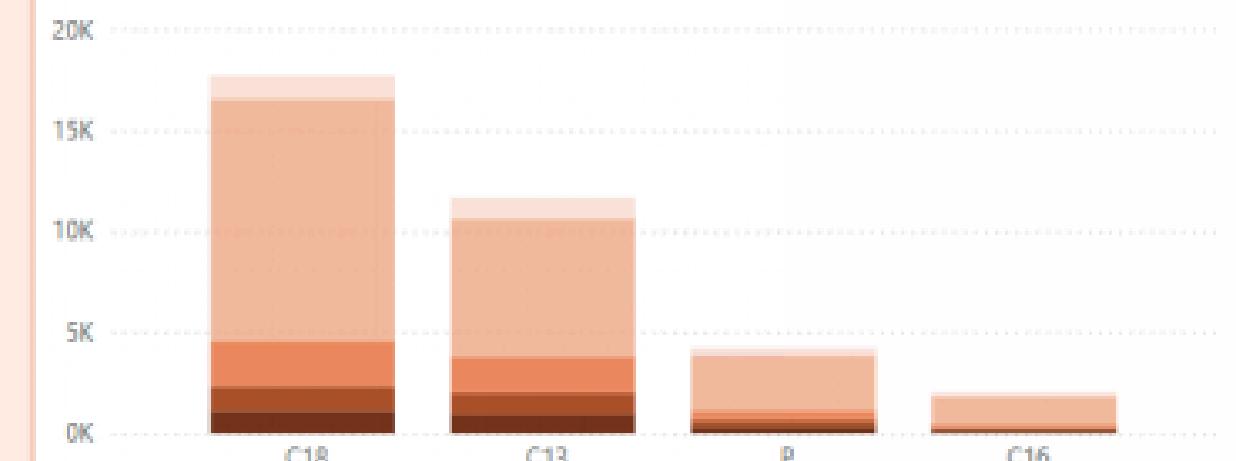
AVERAGE MOVIE WATCHING TIME BY INDUSTRY GROUP

Duration Type: ● Around 1.5 hours ● Around 2 hours ● Around 3 hours



MOVIE RATING RANKINGS BY INDUSTRY GROUP

Occupation: ● blue collar ● specialist ● student ● teenager ● white collar



ĐỀ XUẤT TỐI ƯU HÓA

LỊCH CHIẾU

Theo thời gian

- Tăng suất chiếu cho nhóm khách teenager và white collar.
- Bổ sung suất buổi chiều và tối trong tuần cho student.
- Cân nhắc thêm suất sáng và chiều trong tuần cho blue collar.

Theo thời lượng phim

- Phim dài (2 tiếng) nên chiếu cho teenager vào tối cuối tuần.
- Phim ngắn ưu tiên cho student và white collar vào các ngày giữa tuần.

Theo ngày trong tuần

- Cuối tuần nên dành cho các suất chiếu “hot” với phim có rating cao như C18 và C13.
- Ngày thường có thể giảm số suất chiếu, hoặc chọn chiếu các phim phổ thông, phim ngắn – để tiết kiệm chi phí vận hành mà vẫn giữ được lượt khách.



THANK YOU

For Your Attention

