CINEMA TICKETS PREDICTIO

Prepared by:

Mohammed Saeed Al shehri

Munira Al Zahrani

Abdulelah Al dossari

We will do the following in this project:

1- **Cleaning Dataset**
2- **EDA**
3- **Modeling**

1.Problem Statement:

Cinema industry is not excluded from getting advantage of predictive modeling. Like other industry e.g., retail, banking, and restaurants, sale forecasts can help cinemas for cost reduction and better ROI. By forecasting sales, screening in different locations could be optimized as well as effective market targeting and pricing.

2. Information about Linear Regression:

Linear regression is one of the statistical techniques used in predictive analysis where can be used to find the relationship between a dependent variable (target) and one or more independent variables (predictors). The regression equation form with one dependent and one independent variable is defined by the formula:

$y = \theta_0 + \theta_1 X_1$, where

y = estimated dependent variable score, θ_0 = constant, θ_1 = regression coefficient, and X = score on the independent variable.

We are creating this project to predict ticket price with other columns and to answer   some questions, which are:

- What is the most ticket sold?

- Is the total_sales affect the week_name?

- Which film_code has the most tickets_sold?

- Which film_code has most occupation percentage?

3. Dataset:

After a long search for the appropriate dataset on several websites, we chose to use a dataset called "Cinema Tickets", which was extracted from Kaggle, which contains about 142524 and observations 14 Features. which shows cinema tickets based on 14 Features due to its features that could be very useful in our analysis, which are:

film_code, **cinema_code**, **total_sales**, **tickets_sold**, **tickets_out**, **show_time**, **occu_perc**, **ticket_price**, **ticket_use**, **capacity**, **date**, **month**, **quarter**, and **day**.

The dataset can be found from this link

4. Tools:

We used many libraries related to data analysis and machine learning, which are:

- NumPy.
- Pandas.
- Matplotlib.
- Seaborn.
- Datetime.
- Scikit-learn.

For execute: Jupyter Notebook.

5. MVP:

We will answer all questions that we have or at least some of them and also to do the regression, and also find the linear relationship between dependent variable and one or more independent variables.