Time Series and Classification Project
The Report

Title: **Cinema Ticket Sales Forecasting: A Time Series Analysis for Strategic Growth in the Film Industry**

Amira
BOUDAOUD
Section 1 Group 2

Table Of Content :

I. Some general information and concerns about the project:

An HTML page that displays the executed version of the R notebook for this project, detailing each step taken, along with charts and their explanations and discussions of the different results . Included in the folder also a markdown file with the project's source code, and the dataset used within a CSV file. The HTML page covers all the detailed steps and visualizations, so there's no need to further mention them in this document.

II. Abstract : Definition of the project and its main goal.

This project explores the potential of predictive modeling to advance the cinema industry's efficiency and profitability by accurately forecasting ticket sales. Focused on enhancing operational decision-making, the study involves the analysis of historical sales and movie-specific data, enabling the optimization of screening schedules, targeted marketing, and pricing strategies. The ultimate goal is to provide actionable insights for resource allocation and investment planning, aiming to bolster the financial viability and strategic growth of cinema ventures. Through this endeavor, the project seeks to deliver a forecasting model that aids industry stakeholders in navigating the dynamic entertainment market.

III. Introduction: Definition of the dataset used.

About eight months sales history of different cinemas with detailed data of screening , during 2018 with encoded anonymized locations.

Link to it : https://www.kaggle.com/datasets/arashnic/cinema-ticket/data

Definition of the columns:

Film_code: Unique movie id

Cinema_code :Unique cinema id

Total_sales (our target column) :total sale per screening time

Tickets_sold: number of tickets sold

Tickets_out: Number of tickets canceled

Show_time: screening time in each day

Occu_perc: occupation percent of cinema by means of available capacity

Ticket_price: price of ticket at show time

Ticket_use: total number of tickets used

Capacity: capacity of the cinema

date , month, quarter , day: the date, month, quarter, and day on which each cinema screening occurred, categorizing each screening's timing into different temporal segments for analysis.

VI. General steps taken in the project:

We started by taking a close look at our data, understanding its structure which included dates broken down by day, month, quarter, and individual days. Through this preliminary examination, we decided that daily data was most relevant for our study (you can find a more detailed explanation in our R notebook/html page). We split our dataset into a training set for building the model and a testing set for evaluation and validation purposes.

Our initial analysis indicated that the variability in our data wasn't constant, so we used a Box-Cox transformation, specifically a log transform, to stabilize it. We then checked the time series for stationarity with an ACF plot and the ADF test. Our findings showed that differencing was necessary. After experimenting with first, second, and third differences, we found that differencing once (d=1) was sufficient to achieve stationarity in our time series.

Next, we moved on to specifying our model. We examined the ACF plot to gauge the number of MA terms (q), and the PACF plot for the number of AR terms (p), and used the armasubsets function to visualize the best models based on BIC. We also employed the auto.arima function, which selects the optimal model by minimizing BIC and AIC. We began with three potential models, but after assessing the residuals, we narrowed our choices down to two for forecasting. We forecasted ticket sales for the coming 30 days and then evaluated the forecast's accuracy using our test data. Sadly, the models did not yield accurate predictions due to the data's complexity.

To address the seasonal pattern we observed, we tried a SARIMA model, but it did not improve our results.

VII. Conclusion:

In the broader context of the project, the challenges in accurately forecasting ticket sales underscore the necessity of adopting more sophisticated, flexible modeling approaches that can accommodate the cinema industry's unique dynamics. This may involve exploring non-linear models, incorporating external data sources, and leveraging other algorithms that can adapt to complex patterns and relationships within the data.Thus,fitting simple models (trigonometric trend components,ARIMA, or SARIMA) will not show raisonable results**.**