

## **Why predict Broadway sales?**

Theater-owning conglomerates (our main stakeholders) have an interest in accurate predictions of not just their annual box office gross for the coming year, but also how it is likely to fluctuate week-to-week (our main KPI), so they can budget accordingly. For example, if they can anticipate high sales in a coming week, they can make arrangements for how to spend that money so it is put to good use as quickly as possible (e.g. paying off a loan). While the models we build here predict gross sales, the Playbill data includes other potential outcome variables of interest that could be easily predicted using similar models. Reliable forecasts of the percentage of theater seats sold on a weekly basis can serve as valuable information for theater owners. This data can aid in planning activities such as renovations or facility maintenance, timed strategically during weeks with lower occupancy. Moreover, this knowledge can contribute to efficient workforce management, enabling the reduction of staff during less busy weeks, thus leading to cost savings.

## **Our Approach**

While most Broadway theaters are owned by conglomerates, it is probable that they maintain largely independent budgeting processes. Theaters also vary in their appeal to customers and how their owners use them which means that each theater likely has a slightly different pattern in its sales data. Thus, it is appropriate for us to build separate models to predict each theater individually, and for this, we used time series methods. We used the years 2013 to 2018 to design models and used them to test data in 2019 to assess their performance. For practical purposes, we chose to focus on the Gershwin and Majestic Theatres because their data contained the least numbers of missing values, which makes them easier to work with for a variety of reasons discussed in more detail in the README file of the GitHub repository. Further, the show playing at the theaters examined did not change during the period of interest, which otherwise would have caused a shock that we would have needed to deal with in a way that was beyond the ambitions of this project.

## **Results**

In our comprehensive time series analyses on weekly Broadway sales grosses, we employed the Root Mean Square Error (RMSE) as the yardstick for evaluating the performance of our models. The RMSE provides a measure of the discrepancy between our model's predictions and the actual sales grosses. A lower RMSE implies a model that is able to more accurately predict Broadway sales on a week-by-week basis. We trained and tested three distinct models: Autoregressive Model (AR), Seasonal Autoregressive Integrated Moving Average (SARIMA), and the Prophet Model. Their respective RMSE results were 382,071, 173,964, and 131,508, respectively, for the Gershwin Theatre. Based on these metrics, we selected the Prophet Model as our final model, as it achieved the lowest RMSE and an MAPE of 5.57%, which is plausibly accounted for largely by noise. Upon visualizing the forecasted time series on the test data, it was evident that the Prophet model's predictions closely adhered to the observed behavior. We strongly recommend the Prophet Model to our stakeholders. Not only does it adeptly handle seasonality and trends, but it also incorporates irregularities, such as holiday spikes, which none of the other models do. This advanced level of precision offers a significant advantage for theater owners in forecasting their revenue, as theater attendance tends to increase around holidays. A reliable revenue estimate aids in efficient financial planning and expense management. Additionally, an accurate revenue prediction model can result in increased profitability by reducing the risk of over- or under-budgeting. This leads to optimized operations and increased profitability.