Overview of the Restaurant Industry

Restaurants have been around for thousands of years, connecting families and communities all around the world. A survey performed by Oxford University found that when people eat with others, they are more likely to be fulfilled physically, mentally, and emotionally. Cultivating a sense of inclusivity inherently builds value and fosters belonging. Dining out strengthens human connections, thereby creating a stronger community. Beyond the important social role that restaurants play in the community, they also are a critical part of a successful economy (Massenkoff & Wilmers, 2023).

There are millions of restaurants around the world, employing millions of people around the world. These restaurants benefit the job market, as well as the local and national economy. According to the National Restaurant Association, in the United States, 12.33 million Americans work in the restaurant industry generating $94.8 billion in sales (*Economic,* n.d.). Spending money at local restaurants can generate four times more economic benefits than other local businesses in other industries. A culmination of local economic benefits supports the national economy as well (Unnevehr, 2017).

The restaurant industry is a key component of human culture and economic prosperity. However, with 30% of restaurants closing within the first year due to economic challenges and staff management issues, leveraging data analytics is crucial for maintaining a competitive edge.

Ethical Considerations

In beginning this project, I carefully considered the ethical implications. Conducting research ethically ensures voluntary and consensual participation, and it protects sensitive information through anonymization (text). The data from this study was obtained from Kaggle, a platform offering courses, discussion threads, code, competitions, and datasets. To access the data, I signed an agreement to use it solely for academic purposes. The data was already anonymized, using restaurant IDs rather than identifying information. Now, I will tell you more about the data itself.

Hypotheses

The purpose of this research project is to determine whether seasonality affects restaurant business and, if so, whether seasonal trends can be used to predict future visitor numbers. Additionally, the project investigates whether there is a significant difference in the number of visitors to a restaurant on holidays versus non-holidays.

The null hypothesis for the first part of this research is that there is no statistically significant trend or seasonality in the data. If this hypothesis is rejected, the alternative hypothesis is that there is a significant seasonal trend present in the data.

For the second part of the research, the null hypothesis is that there is no statistically significant difference in the average number of visitors on holidays compared to non-holidays. If this null hypothesis is rejected, the alternative hypothesis is that there is a statistically significant difference in the average number of visitors on holidays versus non-holidays.

Data

To test these hypotheses, the dataset used for this project includes reservation information from over 800 restaurants in Japan, covering the time between January of 2016 to April of 2017. While this dataset contains a multitude of variables, the focus will be on calendar dates, a binary indicator of whether a day is a holiday, and the number of visitors.

Tools and Techniques

To perform this research project, I used R in RStudio for data preparation and time series forecasting analysis. The methods I used were ARIMA modeling with Fourier terms and ANOVA. I first had to prepare the data for analysis.

I selected the relevant variables, aggregated the data across all restaurants for each day, and addressed outliers and missing data points through interpolation.

After this, I was able to begin performing time series analysis to test for seasonality. I used ARIMA modeling with Fourier terms. ARIMA, which combines auto-regression, integration, and moving average processes builds an accurate predictive model (Boomija et al., 2021). Given the limited data span of only a year and a half, traditional time series decomposition was not possible as it requires at least two periods to identify seasonal patterns accurately. Fourier terms were used to decompose the data into smaller components to identify trends (Andrei, 2023).

ANOVA was used to calculate and compare the mean values of two groups—in this case the average number of visitors on non-holidays and holidays. By applying ANOVA to the dataset, I determined the impact holidays have on the amount of restaurant business during these events.

ARIMA Results

Having built the time series forecasting model using no autoregressive terms, first-order differencing, and no moving average terms, the model appears to have captured some seasonality. As shown, the forecasted pattern closely follows the trend of the historical data. The S1-12 and C1-12 coefficients, with values of -63289.49 and 32225.87 respectively, suggest some seasonal effects. However, the large standard errors of 5253.24 and 47400.44 indicate that this seasonality is not statistically significant. This is further confirmed by the ACF1 value of -0.635, suggesting that the model is not capturing all patterns in the data. Therefore, I failed to reject the null hypothesis, indicating no statistically significant seasonality in the dataset. While using Fourier terms provides an approximation, it is less accurate and comprehensive compared to traditional forecasting methods.

ANOVA Results

The ANOVA analysis was conducted to compare the mean number of visitors on holidays versus non-holidays. The results calculated a p-value of 0.143, which is above the commonly accepted significance level of 0.05. This suggests that there is no statistically significant difference in the average number of visitors on holidays compared to non-holidays. Therefore, I failed to reject the null hypothesis, indicating that that holidays do not significantly impact visitor numbers in the dataset. These results highlight that holidays do not significantly influence restaurant business within the specified period.

RECOMMENDATIONS

Moving forward, I would recommend improving the predictive model to achieve more accurate and relevant forecasts. First, extending the data collection period is critical, having at least two cycles of historical data would better capture seasonal trends. However, new restaurants without two years of historical data would still benefit from predicting visitor numbers by incorporating additional variables, such as weather conditions, into the model. This could significantly improve forecasting accuracy, especially in datasets with limited historical data. External factors can provide a more nuanced understanding of visitor patterns and improve model performance.

Additionally, I suggest that each restaurant conduct its own analysis. Given that each establishment has unique characteristics and operational dynamics, building tailored models will account for specific nuances and provide insights customized to each business’s individual needs.

I would recommend pursuing further research into seasonality in the restaurant industry. By developing successful models that incorporate historical data with additional variables unique to each restaurant, more businesses can leverage data analysis for improved decision-making and operational efficiency.

\*\*\*However, using Fourier terms provides an approximation, less accurate than traditional methods\*\*\*

Andrei, F. (2023, December 7). *Modeling variable seasonal features with the Fourier transform*. Medium. https://towardsdatascience.com/modeling-variable-seasonal-features-with-the-fourier-transform-18c792102047

Boomija, G., Anandaraj, A., Nandhini, S., & Lavanya, S. (2018). Restaurant Visitor Time Series Forecasting Using Autoregressive Integrated Moving Average. *Journal of Computational and Theoretical Nanoscience*, *15*(5), 1590–1593. ResearchGate. https://doi.org/10.1166/jctn.2018.7345

*Economic Indicators*. (n.d.). NRA. https://restaurant.org/research-and-media/research/economists-notebook/economic-indicators/

Franks, B. (2019, August 10). *The Most Important Step Toward Ethical Analytics: Intentionality*. LinkedIn. https://www.linkedin.com/pulse/most-important-step-toward-ethical-analytics-bill-franks/

Massenkoff, M., & Wilmers, N. (2023, August). *Rubbing Shoulders: Class Segregation in Daily Activities* [Review of *Rubbing Shoulders: Class Segregation in Daily Activities*]. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=4516850#paper-citations-widget

Unnevehr, L. (2017, March). *Economic Contribution of the Food and Beverage Industry*. Committee for Economic Development; The Conference Board. https://www.ced.org/pdf/Economic\_Contribution\_of\_the\_Food\_and\_Beverage\_Industry.pdf

"C:\Users\Rachel\Documents\Grad School\MIS581\Portfolio Project\PortProj.R"