## Data-Driven Insights for the Kaiserliche Bohne: (Imperial Bean)



AI Generated Image

## Scenario

You are the lead of analytics consultant hired by the owner of " Kaiserliche Bohne," a traditional yet popular coffee shop located in Vienna, Austria. While charming and well-regarded, the owner feels increasing pressure from modern chains and wants to leverage data to optimize operations, understand customer behavior better, and increase revenue. They have recently implemented a new Point-of-Sale (POS) system and have provided you with transactional data spanning over two and a half years. The owner has a good intuition for the business but seeks data-backed evidence and actionable recommendations. Your task is to analyze this data, uncover key insights, and provide a forecast for popular products to aid in future planning.

## Objective

Your primary goal is to act as an analytics consultant, perform analytical work and a data narrative. You need to perform a thorough exploratory data analysis (EDA) to understand historical patterns and trends, followed by time series forecasting of popular products to predict future sales. Your final deliverable should provide actionable insights and recommendations for the Kaiserliche Bohne owner regarding product strategy, inventory management, and potentially promotional activities.

## Data

You are provided with a single CSV file (Kaffeehaus\_Kepler\_Transactions.csv) containing transactional data recorded from the POS system. **It will have data integrity issues which need to be investigated and addressed to** **complete the case.** This is because it is real data!

* Time Period: 10.01.2022 to 12.09.2024 (inclusive)
* Date Format: DD.MM.YYYY (Standard European Format)
* Time Format: HH:MM:SS (or similar, confirm during EDA)
* Currency: Euros (€)

Columns:

1. transactionID
2. receipt
3. Posting.date
4. Posting.time
5. Invoice.number
6. Payment.method
7. Product
8. Product.group
9. Quantity
10. Gross.unit.price
11. Net.unit.price
12. Gross.discount
13. Net.discount
14. Gross.price
15. Net.price
16. VAT.rate
17. VAT.amount

*Within the variables there are relationships such as Quantity \* Gross.unit.price equals Gross.price. Examine the data to determine appropriate values for your analysis. Consultants are often handed data with limited information and must actively identify the subset of variables that should be used. (Also, allow for potential minor rounding differences in the system).*

## Tasks & *Suggested* Guiding Questions

Please note the items below are suggestive and do not represent mandatory nor a complete set of analyses. You’re free to address some of these suggestions or add your own.

Part 1: Data Cleaning and Exploratory Data Analysis (EDA)

* Data Loading & Preparation:
  + Load the dataset. Pay close attention to parsing the Posting.date and Posting.time columns correctly, given the EU format.
  + Check for missing values in crucial columns (e.g., Invoice.number, Product, Quantity, pricing columns). Decide on and justify a strategy for handling any missing data.
  + Verify data types and perform necessary conversions (e.g., numeric columns, categorical columns).
  + Perform basic consistency checks (e.g., Quantity > 0, prices >= 0, VAT calculations). Document any anomalies found.
* Overall Business Performance:
  + What is the overall trend in total sales revenue (Gross.price or Net.price - justify your choice) over the entire period? Visualize this (e.g., monthly aggregation for example).
  + How many unique transactions (Invoice.number) occurred? What is the trend in the number of daily/weekly/quarterly/annual transactions? This can also be examined by product and product group.
  + What is the average transaction value (total Gross.price per Invoice.number)? How has this changed over time (if at all)?
* Product Analysis:
  + Which are the top 10 best-selling products by Quantity? By Gross.price?
  + Which are the bottom 10 performing products by Quantity? By Gross.price?
  + Analyze sales performance by Product.group. Which categories contribute most to revenue?
  + Are there products frequently purchased together? (Hint: Analyze items within the same Invoice.number).
* Inferred Customer Propensity:
  + What are the peak hours of operation based on transaction frequency or sales volume?
  + What are the busiest days of the week? Are there significant differences between weekdays and weekends?
  + How does Payment.method usage vary?

Part 2: Time Series Forecasting

* Data Aggregation:
  + Aggregate the transactional data into a suitable time series format for forecasting for “n” popular products. Common periodicity choices include daily total gross sales or weekly total gross sales. Justify your choice of aggregation; which product(s) or group(s), what periodicity (daily vs. weekly vs monthly etc) and the variable (Gross.price vs. Net.price vs. Quantity).
  + For added accuracy, you can explore the impact missing days/weeks (i.e. imputation), holidays in Austria, discounts or other events.
* Time Series Exploration & Decomposition:
  + Visualize the aggregated time series data.
  + Identify and discuss any observable patterns:
  + Decompose the time series into its trend, seasonal, and residual components. Visualize the decomposition. Does the decomposition confirm your visual observations?
* Algorithmic Forecasting:
  + Split your aggregated time series data into a training set and a testing set (For example, use data up to a point in 2024 for training, and the rest for testing).
  + Implement the Holt-Winters exponential smoothing method or apply other more sophisticated methods based on the book, or other research.
    - Critically, you must decide whether the trend and seasonal components are additive or multiplicative. Justify your choice based on the EDA and visualization of the time series (e.g., does the seasonal variation amplitude increase over time?).
    - Specify the seasonal period (e.g., 7 for weekly seasonality if using daily data, 52 for annual seasonality if using weekly data).
  + Train the model.
* Model Evaluation & Forecasting:
  + Generate forecasts on the testing set period.
  + Evaluate the model's performance using appropriate metrics (e.g. Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE)). Interpret these metrics in the context of the coffee shop's sales figures for the chosen products.
  + Visualize the actual values from the test set against the forecasted values and confidence intervals. How well did the model capture the patterns?
  + Once satisfied with the model (or after documenting its limitations), retrain the model on the *entire* dataset of time for the chosen product(s).
  + Generate a forecast for a relevant future period (e.g., the next 4-12 weeks). Visualize this forecast, including prediction intervals.
* Interpretation & Recommendations:
  + Based on the forecast, what recommendations can you provide to the Kaiserliche Bohne owner regarding:
    - Staffing: Aligning staff levels with predicted busy/slow periods?
    - Inventory: Ordering products based on anticipated demand?
    - Marketing/Promotions: Planning promotions during forecasted lulls?
    - Revenue: Given the products selected what is the expected revenue, and revenue change the business owner could expect in the forecasted period?
  + Discuss the limitations of your forecast (e.g., assumption that past patterns continue, inability to predict external shocks not present in historical data).

## Deliverables

Submit the following:

1. Code: A well-commented R Script or R Markdown file containing all code used for data loading, cleaning, EDA, time series analysis, modeling, and visualization.
2. Executive Summary / Report: A concise report (max 5 pages)
3. A recorded narration in the presentation, or a link to a recording wherein you’re presenting your data efforts & insights
4. Presentation slides aimed at the the Kaiserliche Bohne owner owner (*a non-technical stakeholder* requiring robust explanations). It should include:
   * An agenda
   * A brief overview of the project objective.
   * A data set overview
   * Key findings from the EDA (use clear visualizations & summary statistics).
   * A summary of the forecasting results (including a clear visualization of the forecast).
   * Actionable recommendations based on your analysis and forecast.
   * A brief note on limitations.

## Hints and Considerations

* Pay careful attention to data types, especially dates and numeric fields. The EU date format (DD.MM.YYYY) is critical.
* Consider the impact of holidays or special events in Vienna if you can identify them or infer potential effects (though external data isn't provided, look for anomalies).
* Seasonality might exist at multiple levels (time of day, day of week, season of year). Your time series aggregation will influence which seasonality is easiest to model. For example it is likely there is intraday (hourly) ordering patterns such as coffee in the morning and also daily, weekly and monthly patterns such as more coffee being sold on Monday vs another weekday.
* Justify your choices: Why did you choose a specific aggregation level? Why additive/multiplicative seasonality? Why Holt-Winters or some other chosen approach?
* Focus on clear visualizations and interpreting results in the business context. The owner cares about coffee and customers, not just model parameters.

## Other Notes

Submit R code, a narration of a live business presentation (can be a standalone video file, or inserted within the powerpoint) , a set of slides, and a written supplemental. If you submit documents with links to youtube videos or cloud drive files, you must ensure all links are accessible. Links set to private which are not viewable or downloadable will automatically result in 0 for that section of the rubric.

You are not allowed to use an ai avatar, or speech creation for narration. The purpose of the business case presentation is to improve *your* presentation skills. In a business setting you will still be expected to articulate your findings and not send an avatar. As a result services like <https://elevenlabs.io/> or <https://www.heygen.com/> are not permitted.

Submit a **formal business report** to represent the entirety of your work including the data overview, process, findings, and implications. Anything less than a professionally written and organized report will be considered sub-optimal. Do not submit a series of bullet points, code or markdown. The use of external and verifiable sources is expected to add context and support any component of the paper. **The minimum is 2 pages maximum is 5**. **Double spaced and 12 point font.**