CS210 Project Report

Project Title: Rythms of the heart: Personalized Heart Rate Analytics for Health Monitoring and forecasting

Name, Surname & ID:

Merve Bilgi 29117

Month, day, year: 19.01.2024

Motivation:

The motivation behind this heart rate analysis project stems from a desire for a deeper understanding of personal health and the factors influencing it. Heart rate data, a critical indicator of cardiovascular health, can reveal much about one's lifestyle and well-being. By meticulously analyzing this data, the intent is to uncover patterns and correlations with daily behaviors, stress levels, quality of sleep, and physical activity. This exploration seeks to leverage data-driven insights to foster a more informed approach to health, potentially leading to better health decisions, improved fitness strategies, and a more personalized health regimen. The overarching aim is to utilize the power of personal data to support a proactive stance on health maintenance and preemptive care.

Data Source:

The data source for this project is an extensive collection of heart rate metrics gathered through a personal health-tracking device. This advanced device is designed to record heart rate continuously throughout the day and night, capturing every heartbeat. The information is stored in a structured XML format, which includes precise timestamps for each recorded heart rate, ensuring data accuracy and traceability. The XML data encompasses a wide range of heart rate records, from resting periods to times of peak activity, providing a comprehensive dataset for analysis. This rich dataset forms the foundation for in-depth analysis, enabling a granular look at heart rate patterns in relation to various daily activities and external factors that might affect cardiovascular health. Through careful extraction and transformation of this XML data, valuable insights into personal health trends can be derived.

Data Analysis:

The analysis involved several stages:

1. Parsing the XML data to extract meaningful heart rate information.
2. Resampling the data to different time frames (hourly, daily, weekly) to observe trends.
3. Visualizing the data to identify patterns and outliers.
4. Conducting autocorrelation and partial autocorrelation analyses to understand the periodicity and lag effects.
5. Building an ARIMA model to forecast future heart rate trends based on historical data.

The data analysis component of this heart rate project was a multi-layered process that began with the careful extraction of heart rate measurements from an XML dataset. Initially, the XML data required parsing to filter out the pertinent information, namely the timestamps and heart rate values. Once extracted, the data was restructured into various timeframes to facilitate a detailed trend analysis, looking at hourly, daily, and weekly aggregates to discern any underlying patterns or anomalies.

Visual representations played a crucial role in making the data comprehensible. Charts and graphs were employed to provide a clear visual context for the heart rate behavior, highlighting any irregularities and painting a broader picture of the subject's cardiovascular activity.

The analysis delved deeper with autocorrelation and partial autocorrelation assessments, which helped in understanding the data's cyclical nature and the impact of previous heart rate readings on subsequent measurements. These analyses were instrumental in identifying the extent to which past and present heart rate data are interrelated.

Building upon these findings, an ARIMA model was constructed to project future heart rate values. This predictive model was trained on historical data, and through its iterative process, it provided forecasts that could guide anticipatory health decisions. This forecasting is particularly powerful as it encompasses both the immediate and distant future, harnessing the patterns decoded from the previous stages of analysis.

Such a comprehensive approach to data analysis not only illuminated the immediate state of cardiovascular health but also offered a predictive lens through which potential future states could be viewed and prepared for.

Findings:

The project's findings revealed insightful patterns in heart rate behavior. The data visualizations illuminated a rhythmic pattern, with specific hours demonstrating elevated averages, suggesting a link to daily physical activities or circadian rhythms. An intriguing variability in weekly data also surfaced, potentially reflecting the individual's unique weekly schedule or varying stress levels.

The ARIMA model's forecasts offer a glimpse into future heart rate trends, equipping the individual with a predictive tool to anticipate times of increased or decreased heart rate. This could prove invaluable for planning exercise, rest, and managing overall health.

This predictive capability, backed by a solid understanding of past and present heart rate data, paves the way for a proactive approach to health and fitness routines. The insights derived could inform personal health decisions, optimize workout times, and even aid in stress management by highlighting when relaxation techniques might be most beneficial.

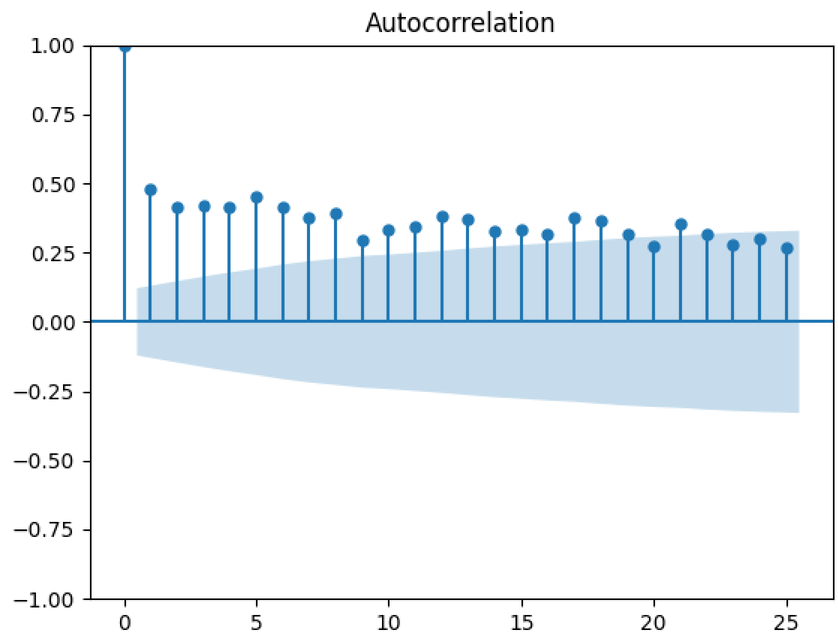
Limitations and Future Work:

The project encountered several limitations, primarily the reliability of wearable device data. These devices, while convenient, may not always provide medical-grade accuracy. For enhanced validity, future efforts could involve calibration against clinical-grade devices.

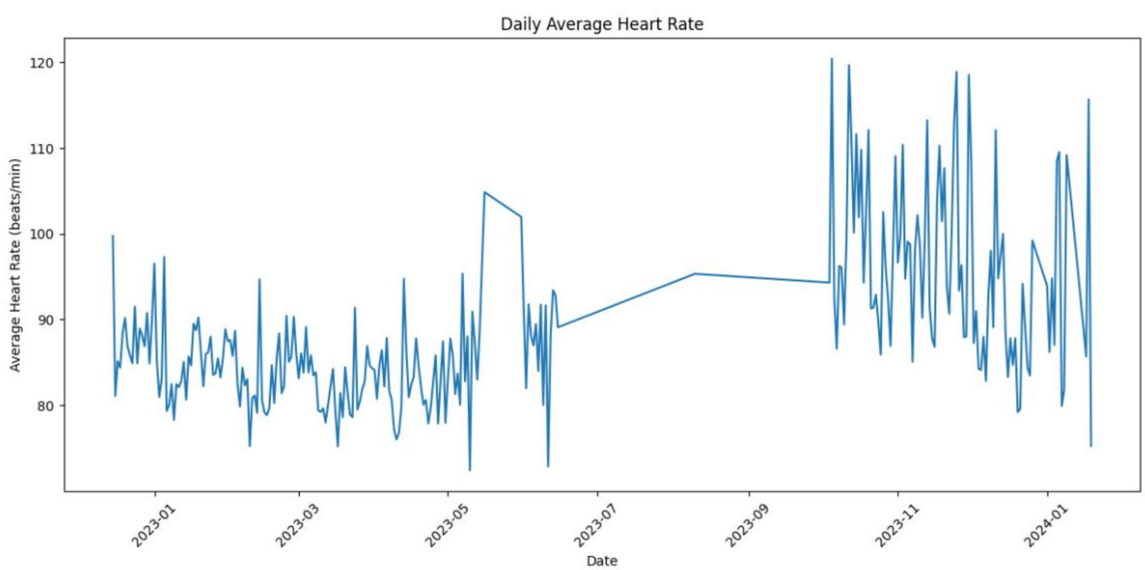
Looking ahead, there's substantial scope to enrich the analysis by incorporating more health indicators, such as blood pressure or sleep quality. Integrating multi-dimensional health data can offer a holistic view and potentially reveal deeper insights into cardiovascular health and overall well-being.

Further research could also explore machine learning models to identify more complex patterns and predict outcomes with greater precision. Expanding the dataset to include a diverse population would also improve the model's robustness and applicability to general health monitoring and personalized medicine.

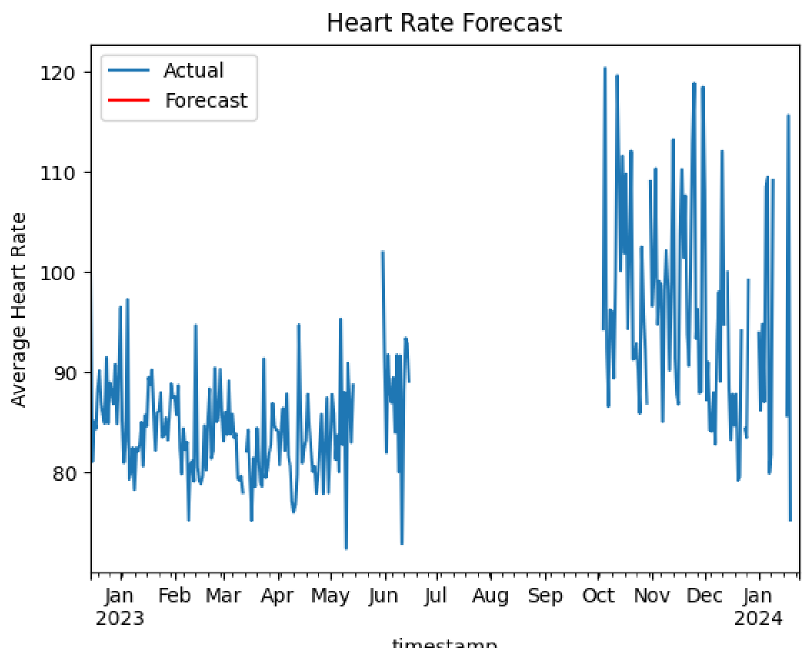
Visuals:



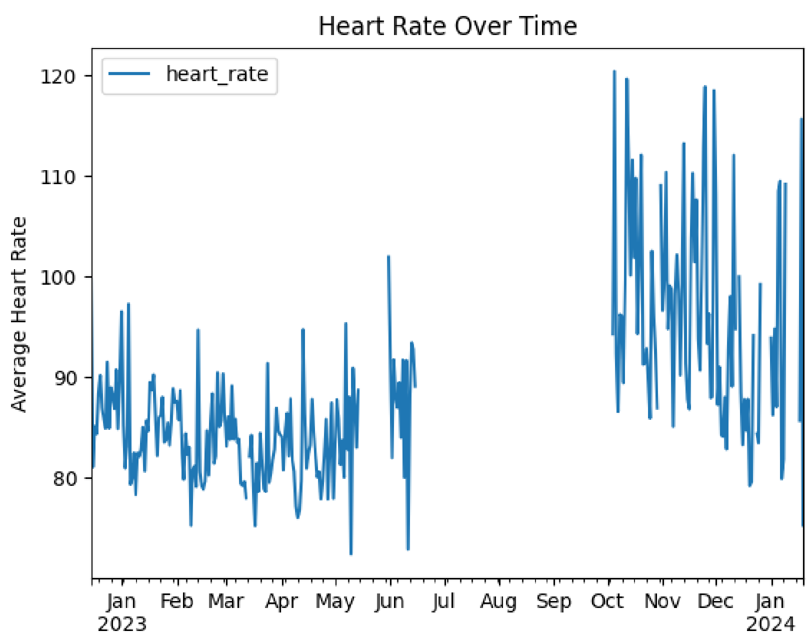
Autocorrelation Plot: This graph shows the correlation of the dataset with itself at different lags. The slow decline of the autocorrelation values as the lag increases suggests a possible long-term pattern or trend in your heart rate data. This could indicate a seasonal effect or other underlying patterns.



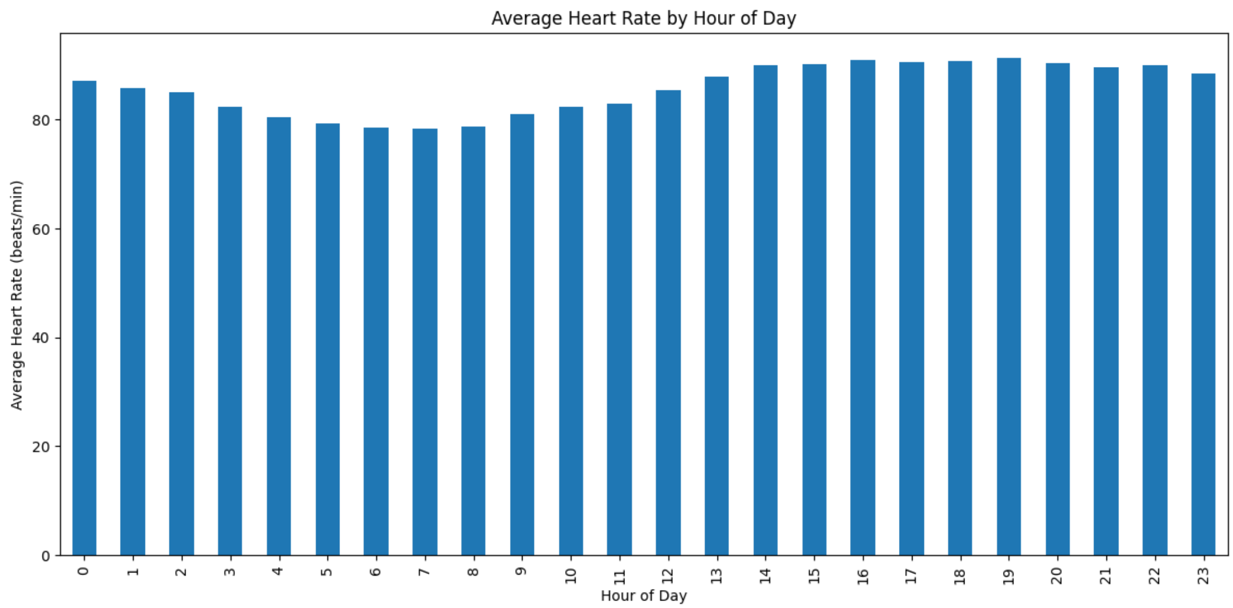
Daily Average Heart Rate: The line graph displaying daily average heart rate over time shows fluctuations that may correspond to daily activities or rest periods. There are notable peaks and troughs which may align with days of heightened activity or rest.



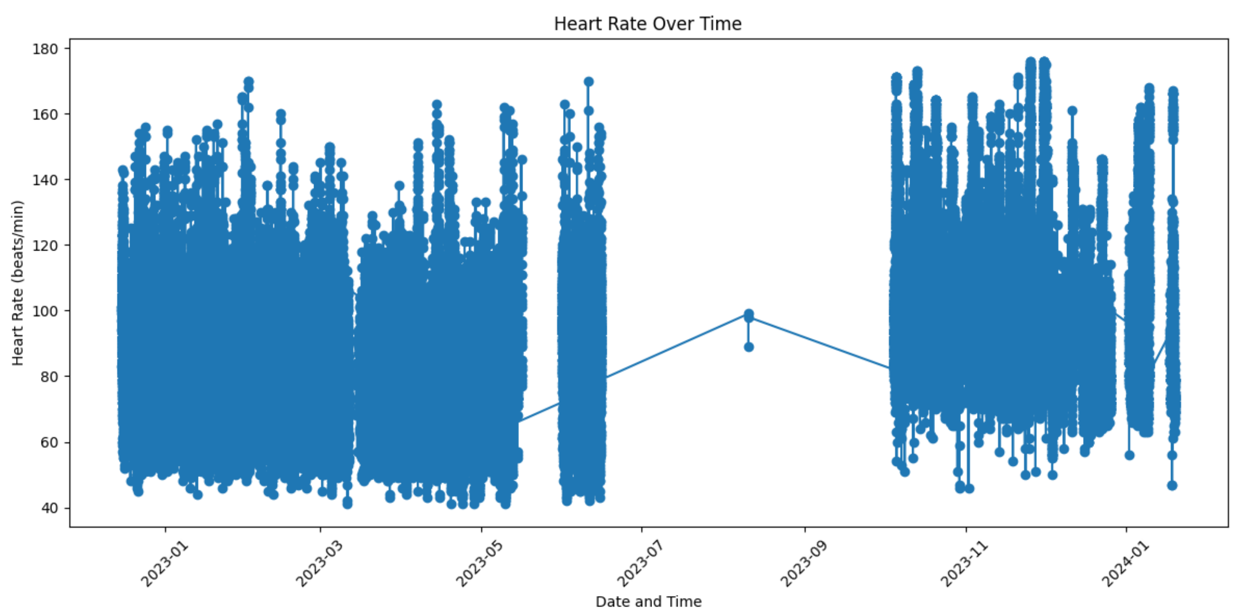
Heart Rate Forecast: The forecast plot indicates predictions of future heart rate based on past data. It shows the actual data in blue and the forecasted data in red. The model seems to predict a stable continuation of the most recent trends



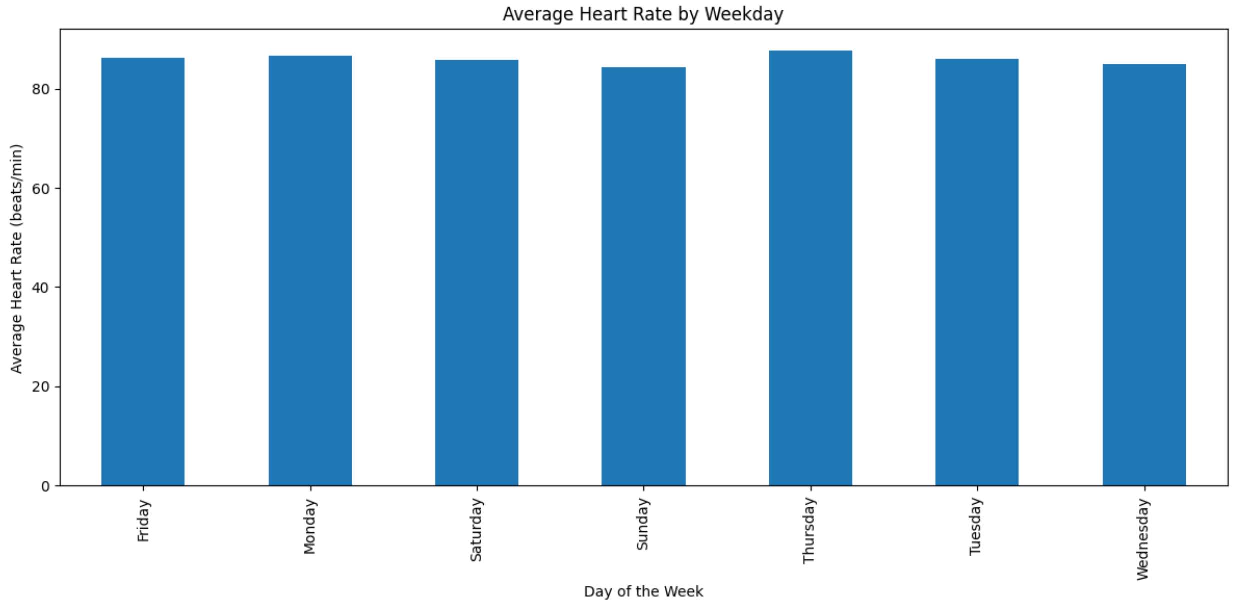
Heart Rate Over Time by Month: This visualization helps identify any longer-term trends or cycles over several months. Seasonal changes or changes in routine could explain any visible patterns.



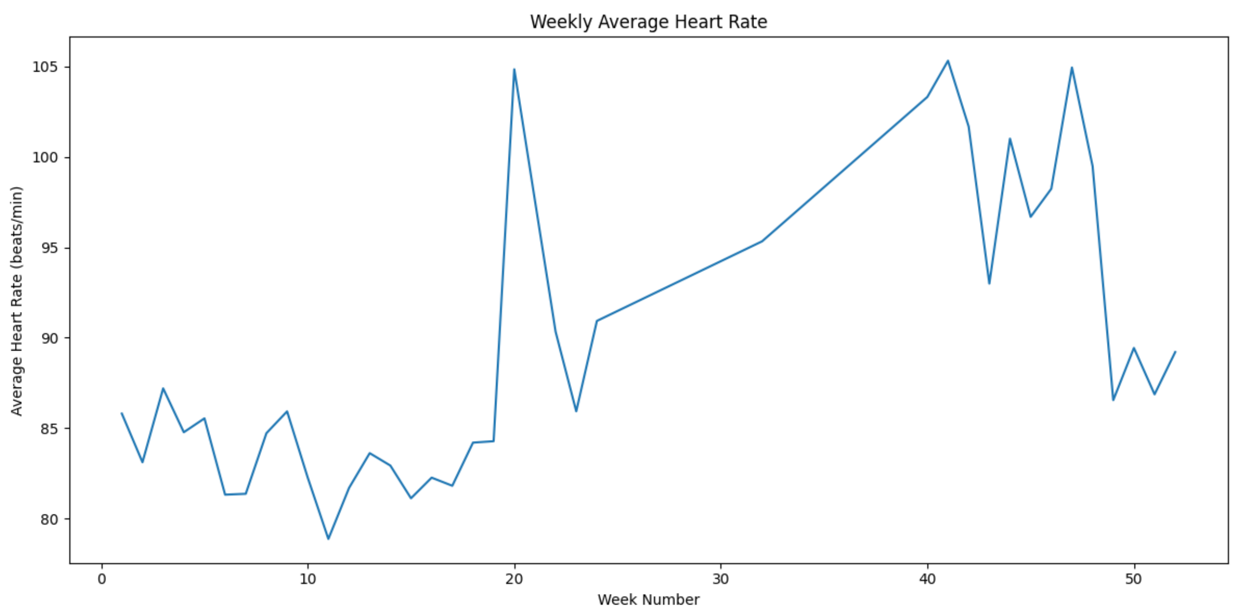
Average Heart Rate by Hour of Day: This bar chart is particularly useful for identifying times of day when heart rate is highest or lowest, which can be linked to daily routines, such as exercise or sleep.



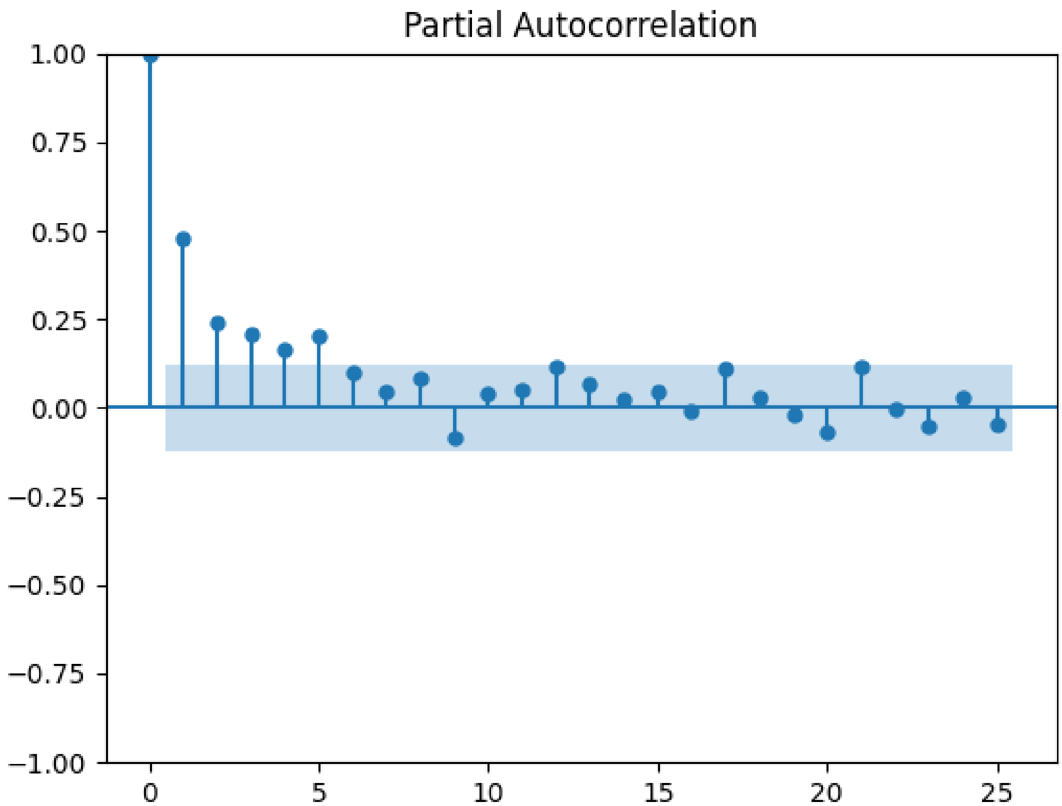
Heart Rate Over Time with Detailed Timestamps: A more granular look at heart rate variations throughout the day, this scatter plot may be used to pinpoint specific events or activities that cause heart rate spikes.



Average Heart Rate by Weekday: This chart suggests that there might be variations in heart rate averages based on the day of the week. It could reveal how weekly routines impact heart rate.



Weekly Average Heart Rate: By averaging heart rate data by week, it's possible to observe broader patterns and eliminate daily noise. Any sustained changes could be associated with lifestyle changes or stress levels.



Partial Autocorrelation Plot: This plot helps identify the direct effect of past data points on future data, independent of intermediate values. The significant initial lags may suggest that only the most recent past values have a direct impact on future heart rate, with the effect tapering off as we go further back in time.

For detailed visual, all the data processing and modeling codes, please refer to the GitHub repository. https://github.com/xmrvx/CS210-PROJECT-bmerve.git