



ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
KHOA KHOA HỌC & KỸ THUẬT THÔNG TIN

XÂY DỰNG HỆ THỐNG DỰ ĐOÁN NĂNG LƯỢNG TIÊU THỤ REAL-TIME ỨNG DỤNG CÔNG NGHỆ DỮ LIỆU LỚN

SVTH: Team 10

- Phạm Đức Thể
- Trần Thành Luân
- Mai Đức Thuận

GVHD: TS. Đỗ Trọng Hợp

NỘI DUNG

1. Giới Thiệu

2. Bộ Dữ Liệu

3. Exploratory Data Analysis

4. Phương Pháp

5. Thực Nghiệm

6. Kiến Trúc Hệ Thống

7. Kết Luận

8. Demo

Giới Thiệu

- Trong vài năm qua, nhận dạng hoạt động trong các ngôi nhà thông minh đã trở thành một lĩnh vực nghiên cứu được quan tâm do có nhiều ứng dụng. IoT tập hợp mọi thứ trong ngôi nhà dưới một chiếc ô có khả năng giám sát và điều khiển từ xa như điều hòa không khí, hệ thống báo động, chiếu sáng, sưởi ấm, thông gió, hệ thống điện thoại, TV, Camera, v.v.
- Để nâng cao sự thoải mái và an ninh cho con người với mức tiêu thụ năng lượng thấp và quản lý năng lượng hiệu quả là một trong những trường hợp sử dụng IoT mà năng lượng được gửi đi hoặc tiêu thụ có thể được theo dõi.
- Hệ thống dự đoán mức tiêu thụ năng lượng trong tương lai với dữ liệu trong quá khứ được xây dựng.

Bộ Dữ Liệu

Bộ dữ liệu chúng tôi đã sử dụng là “**Smart Home Dataset with Weather Information**”:

- Source: [Kaggle](#).
- Bộ dữ liệu có 32 cột và hơn 500,000 điểm dữ liệu với khoảng thời gian cập nhật 1 phút/lần về năng lượng được sử dụng (tính bằng kW) bởi các thiết bị của ngôi nhà thông minh và điều kiện thời tiết của khu vực vào thời điểm đó.
- Bộ dữ liệu này ghi lại mức tiêu thụ năng lượng của từng phòng/thiết bị mỗi phút, từ ngày 1 tháng 1 năm 2016 (5:00 AM) đến ngày 16 tháng 12 năm 2016 (3:29 AM)

Bộ Dữ Liệu

Dữ liệu năng lượng

- Living Room, Kitchen, Home Office: Những cột này ghi mức sử dụng năng lượng của từng phòng trong nhà tính bằng KW.
- Fridge, Barn, Dish Washer, Microwave, Furnace, Wine Cellar, Garage Door, Well: Đây là số đọc của đồng hồ đo năng lượng của các thiết bị riêng lẻ được sử dụng trong nhà vào một ngày cụ thể.
- Generator, House overall: Các cột này chứa tổng mức sử dụng năng lượng của tất cả các thiết bị và tổng năng lượng do máy phát tạo ra vào một ngày cụ thể.

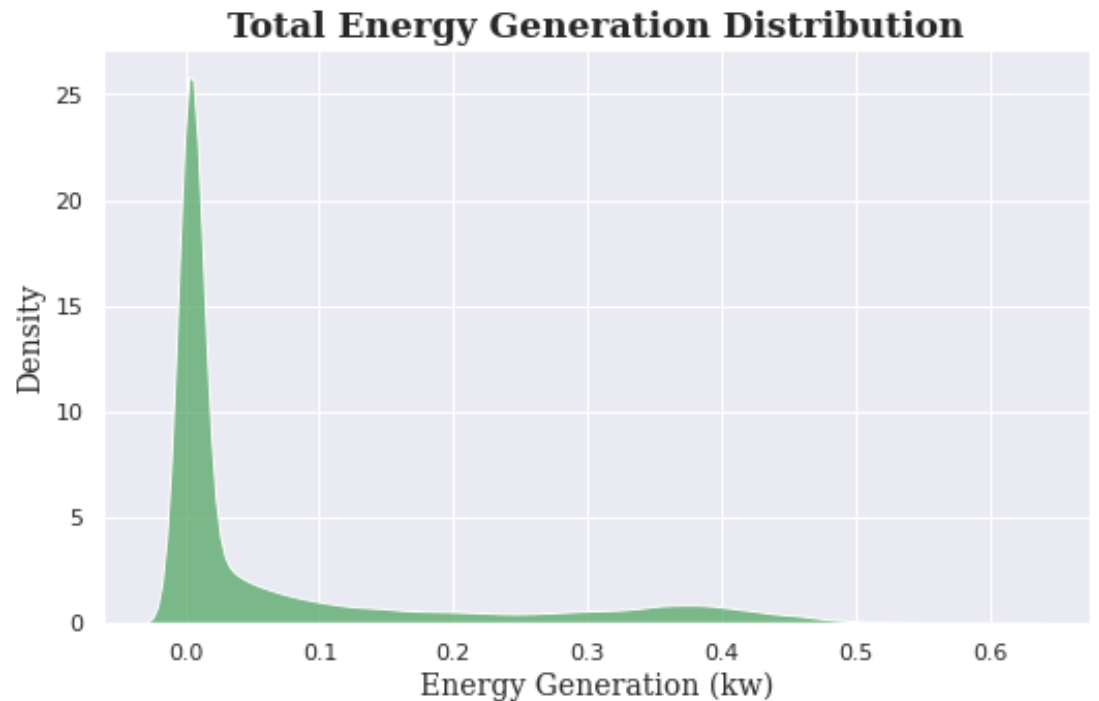
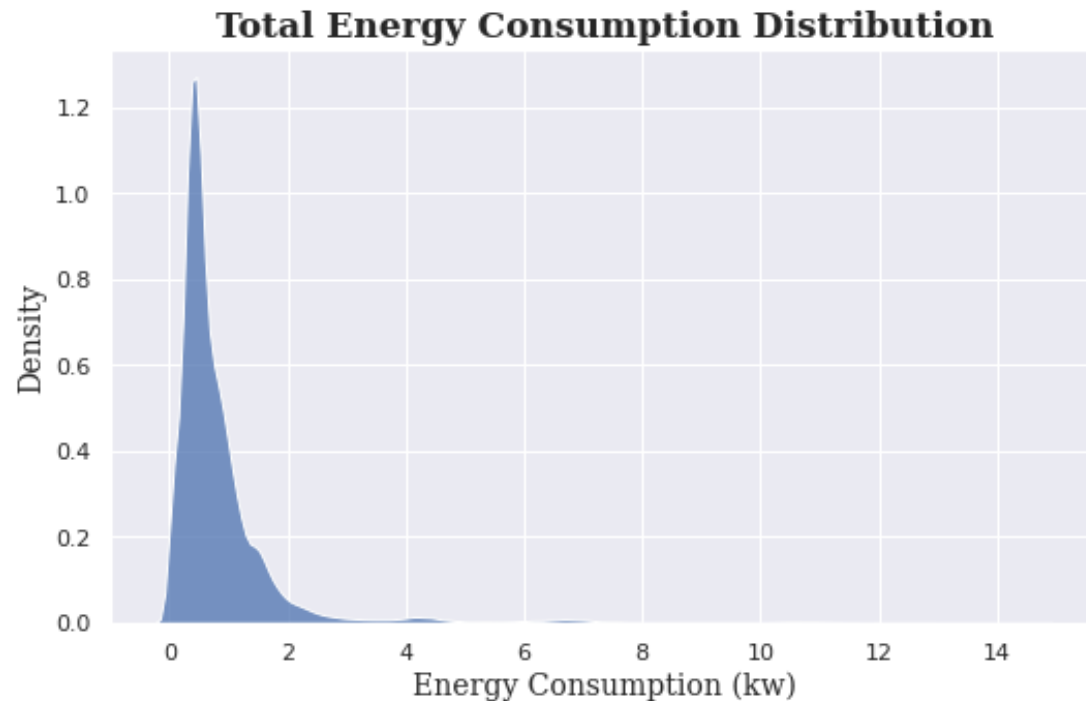
Bộ Dữ Liệu

Dữ liệu thời tiết

- Temperature, Humidity, Apparent Temperature: Đây là nhiệt độ và độ ẩm được ghi lại trong ngày tại địa điểm đó.
- Visibility, Dew Point, Wind Bearing, Wind Speed, Pressure: Phần này chứa tốc độ gió và các yếu tố khí hậu khác trên địa điểm đó ở các tỷ lệ khác nhau.

Exploratory Data Analysis

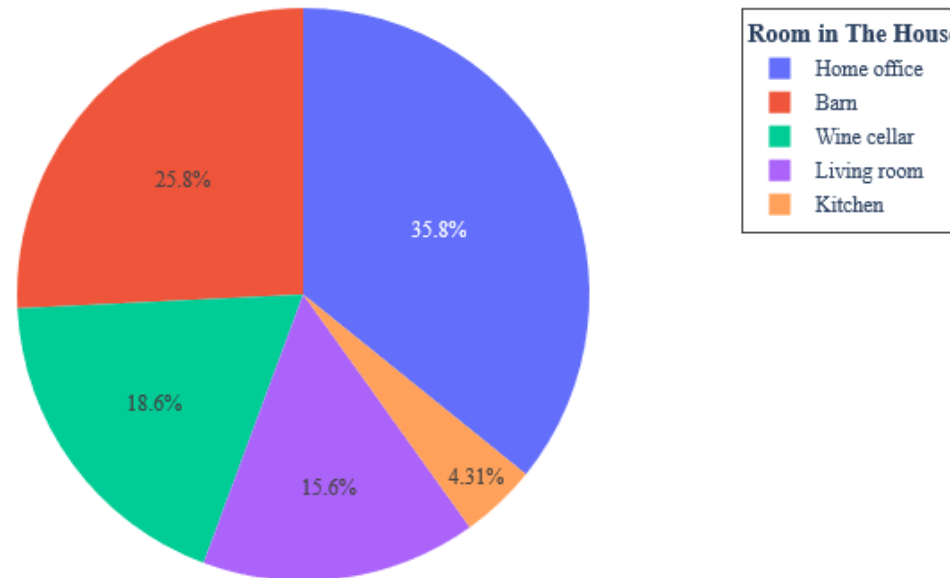
Basic Analysis - House Appliances



Exploratory Data Analysis

Basic Analysis - House Appliances

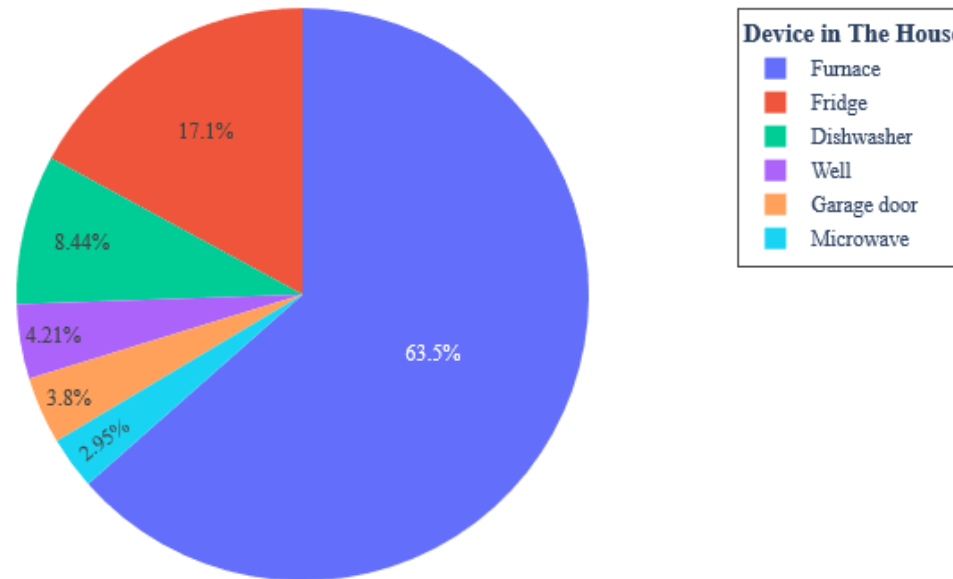
Energy Consumption in kW by Each Room in The House



Exploratory Data Analysis

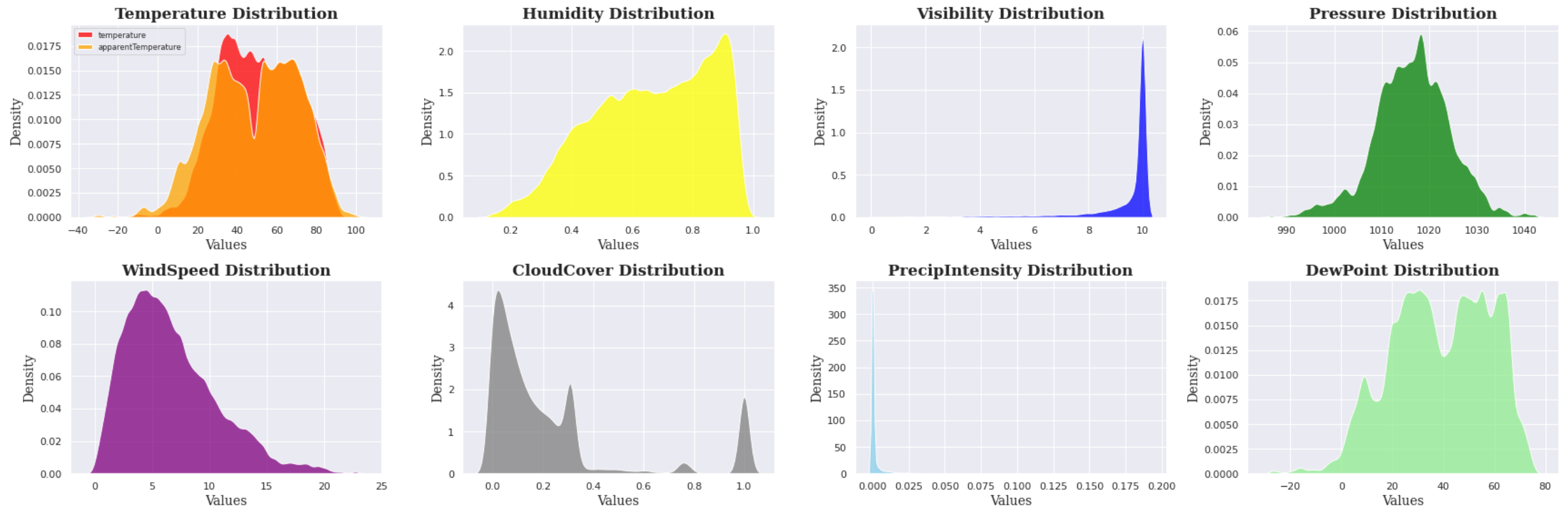
Basic Analysis - House Appliances

Energy Consumption in kW by Each Device in The House



Exploratory Data Analysis

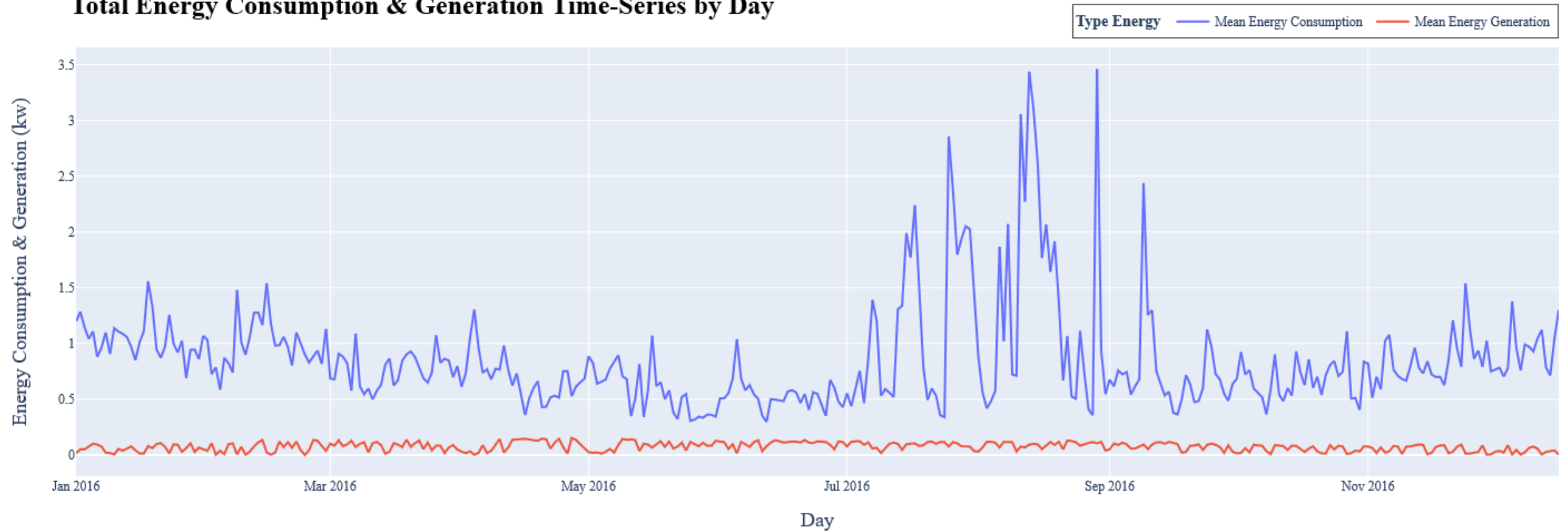
Basic Analysis - Weather Information



Exploratory Data Analysis

Time Series Analysis - House Appliances

Total Energy Consumption & Generation Time-Series by Day



Exploratory Data Analysis

Time Series Analysis - House Appliances

Total Energy Consumption & Generation Time-Series by Weekday



Exploratory Data Analysis

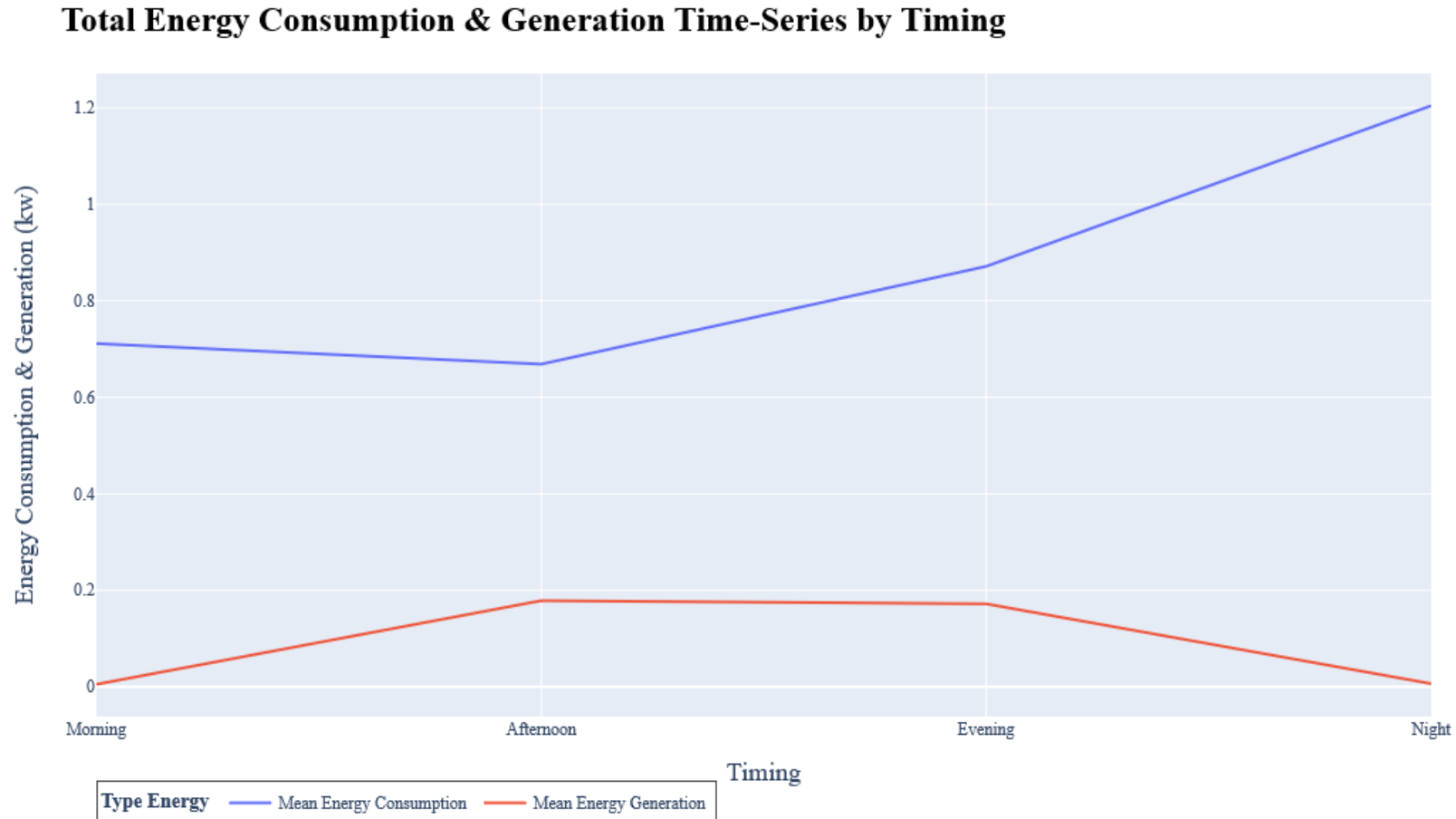
Time Series Analysis - House Appliances

Total Energy Consumption & Generation Time-Series by Month



Exploratory Data Analysis

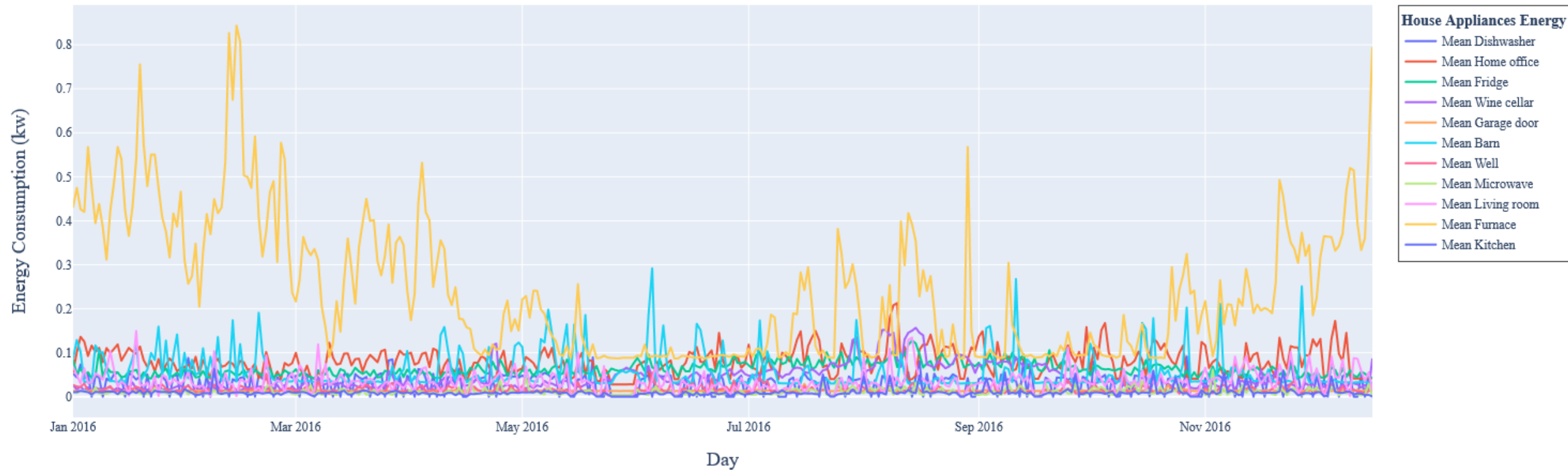
Time Series Analysis - House Appliances



Exploratory Data Analysis

Time Series Analysis - House Appliances

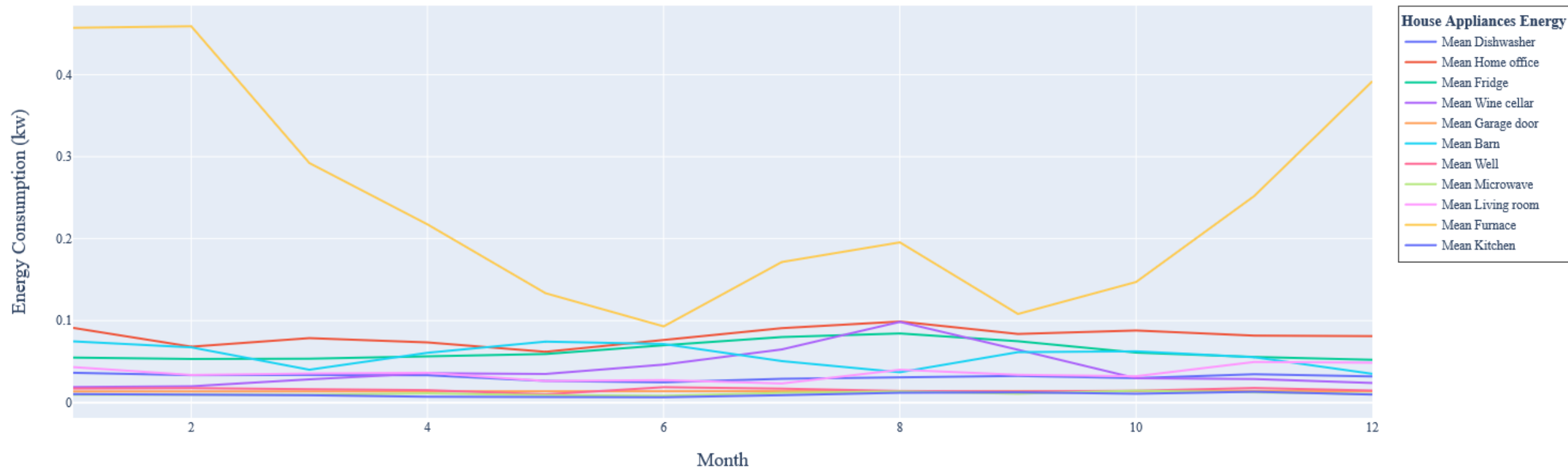
House Appliances Time-Series by Day



Exploratory Data Analysis

Time Series Analysis - House Appliances

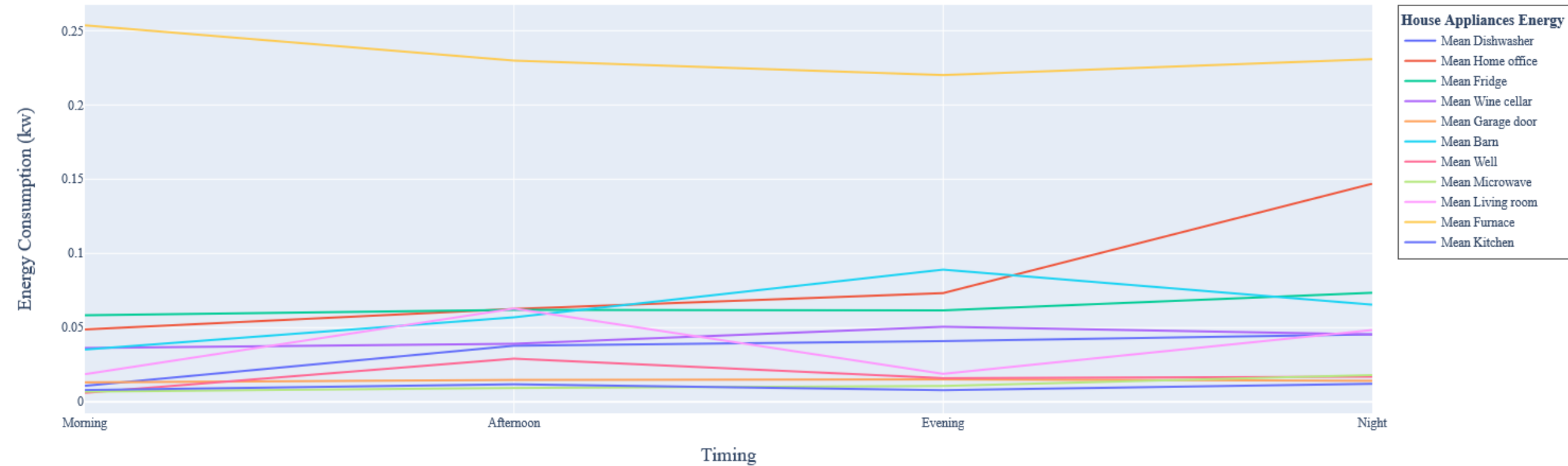
House Appliances Time-Series by Month



Exploratory Data Analysis

Time Series Analysis - House Appliances

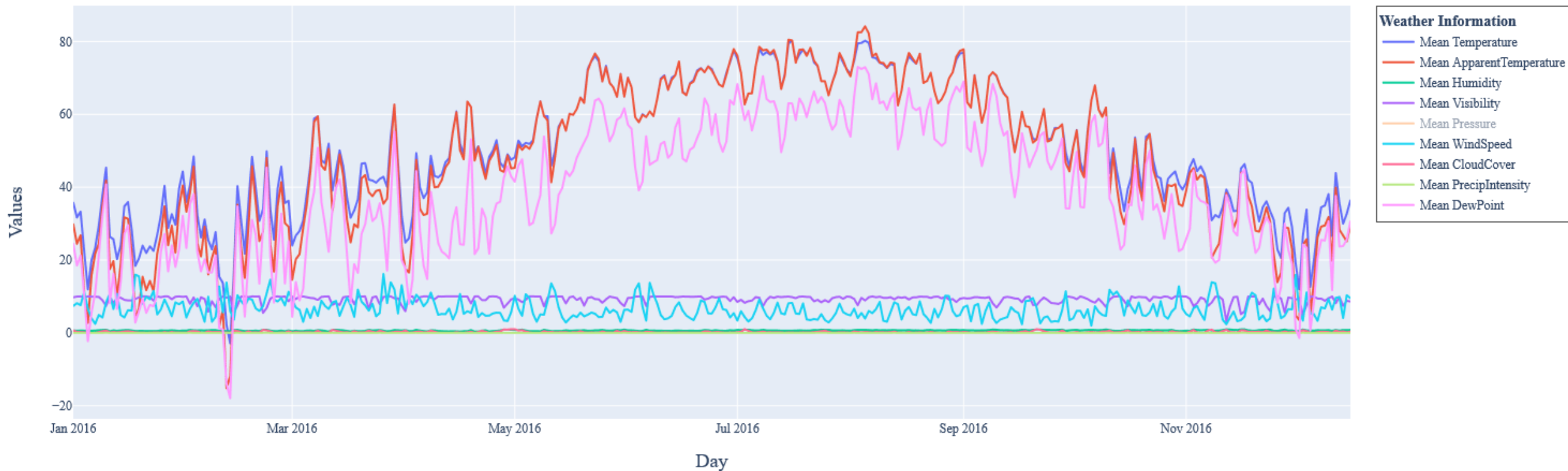
House Appliances Time-Series by Timing



Exploratory Data Analysis

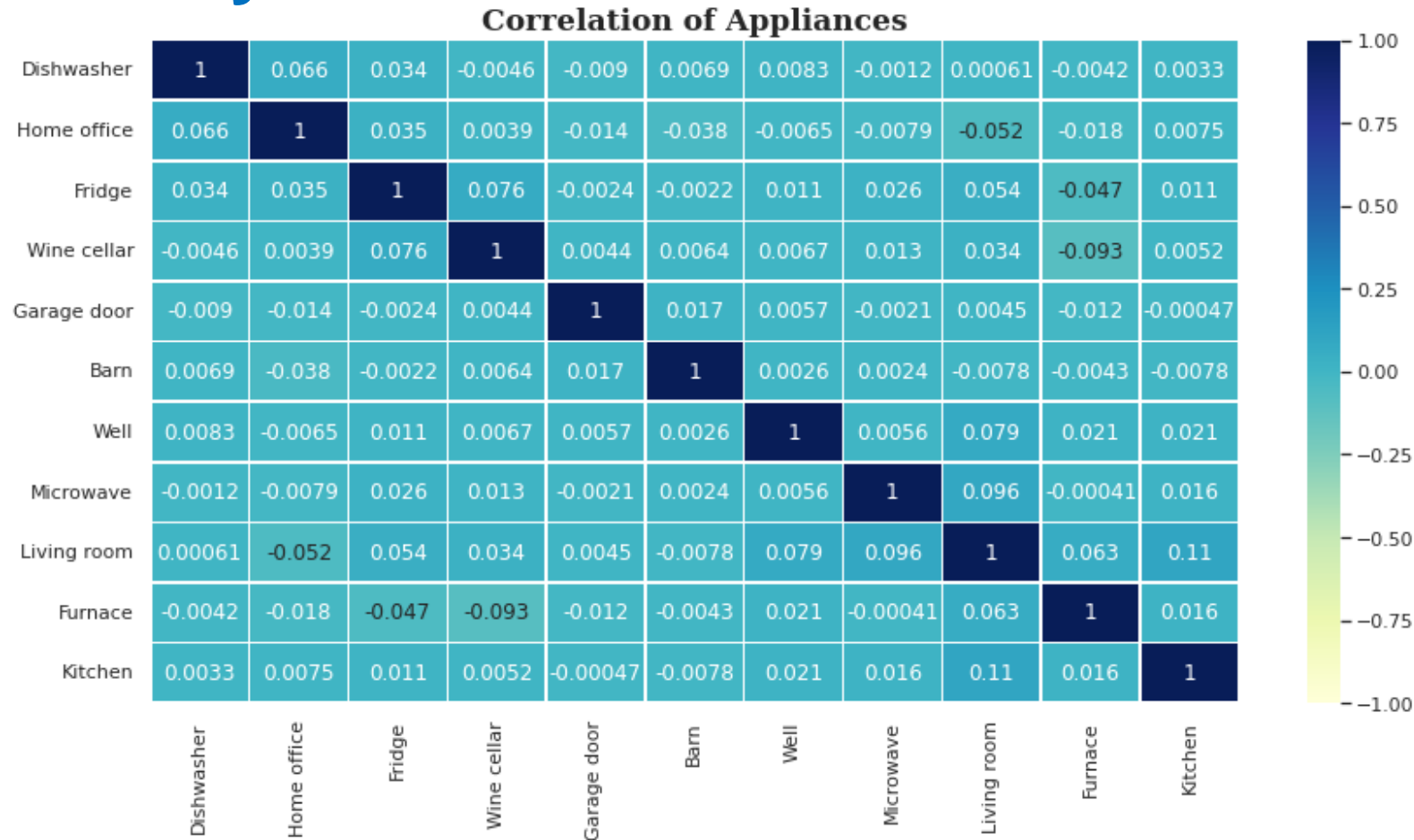
Time Series Analysis - Weather Information

Weather Information Time-Series by Day



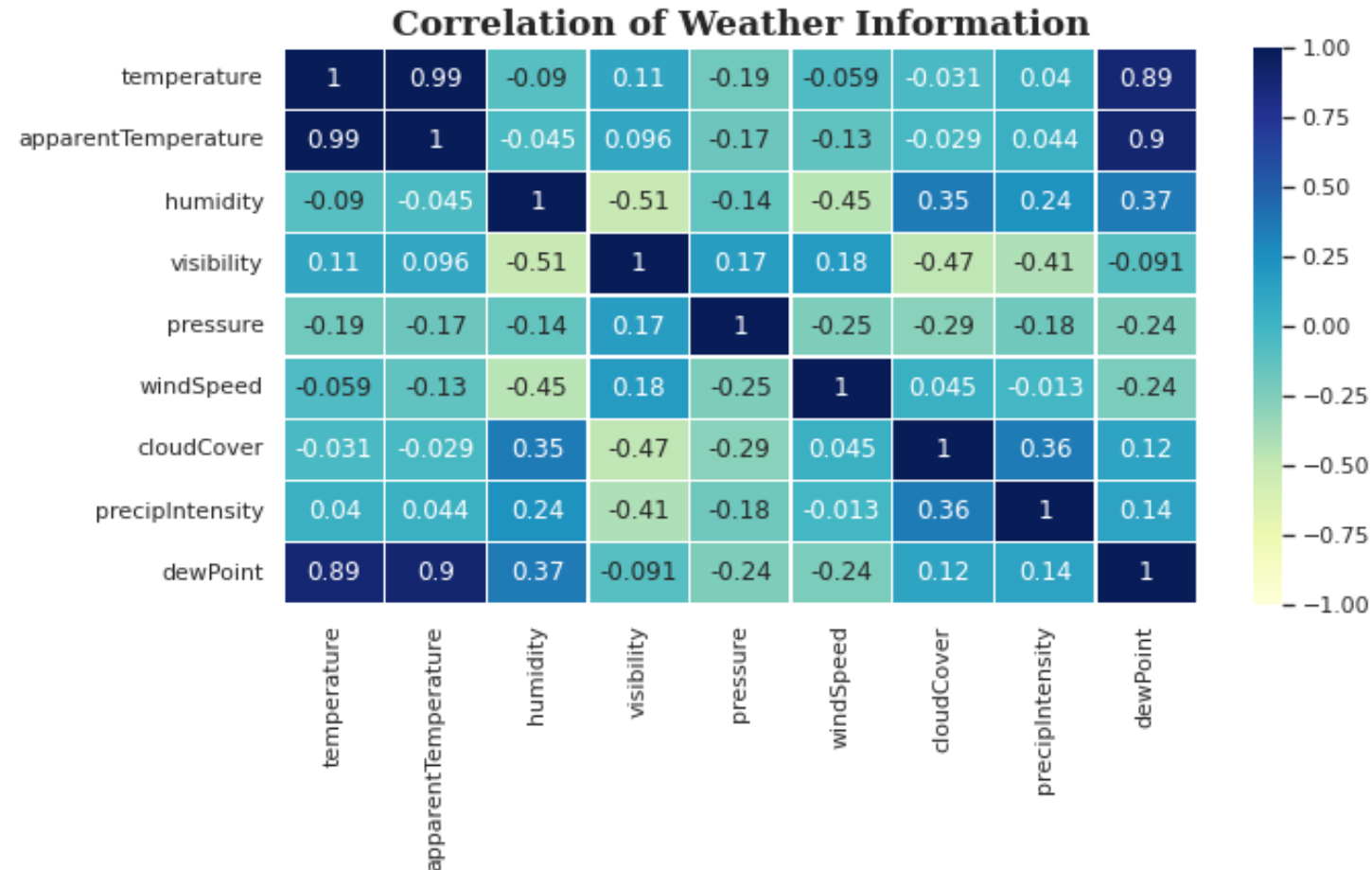
Exploratory Data Analysis

Correlation Analysis



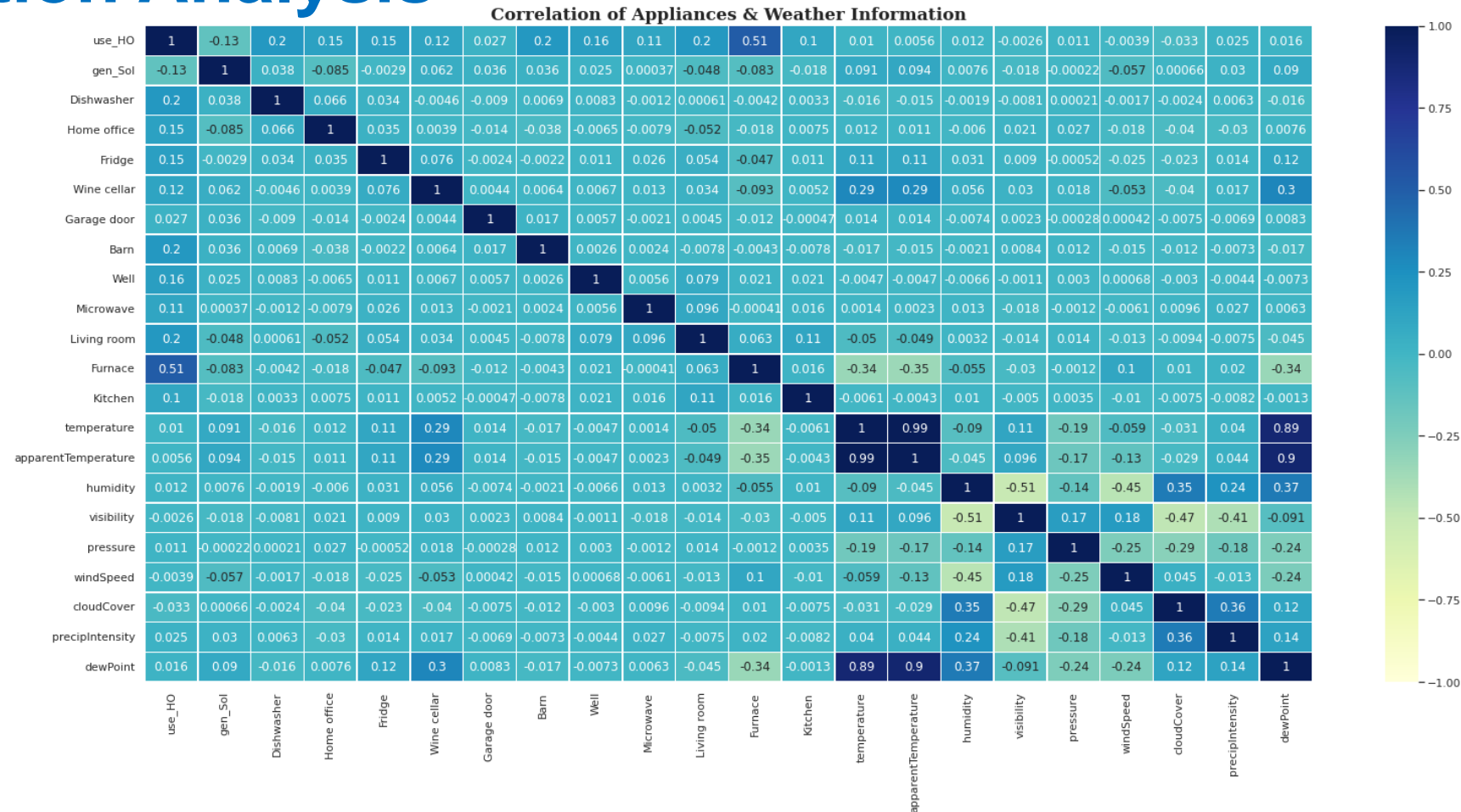
Exploratory Data Analysis

Correlation Analysis



Exploratory Data Analysis

Correlation Analysis



Phương Pháp

Univariate & Multivariate Time Series Forecasting

- TCNForecaster
 - LSTMForecaster
 - Seq2SeqForecaster
 - CNN
 - GRU
 - LSTM
- } BigDL - Chronos Forecaster

Phương Pháp

Univariate Time Series Forecasting

- Là một bài toán dự báo dữ liệu chuỗi thời gian mà chỉ có một chỉ số quan trọng được xem xét (chỉ có một biến độc lập được sử dụng để dự báo giá trị tiếp theo).
- Có nhiều model được sử dụng cho Univariate Time Series Forecasting bao gồm: ARIMA, Prophet, Neural Network Models (Feedforward, LSTM, GRU,...), BigDL-Chronos Models (Seq2SeqForecaster, TCNForecaster, LSTMForecaster,...).

Phương Pháp

Multivariate Time Series Forecasting

- Là một bài toán dự báo dữ liệu chuỗi thời gian mà nhiều chỉ số quan trọng được xem xét (nhiều biến độc lập được sử dụng để dự báo giá trị tiếp theo).
- Có nhiều model có thể sử dụng cho Multivariate Time Series Forecasting bao gồm: VAR, Prophet, Neural Network Models (Feedforward, LSTM, GRU,...), BigDL-Chronos Models (Seq2SeqForecaster, TCNForecaster, LSTMForecaster,...).

Phương Pháp

BigDL-Chronos Forecaster

- Chronos Forecaster là một thư viện trong BigDL, nó cho phép người dùng áp dụng các thuật toán Deep Learning vào dữ liệu chuỗi thời gian cho mục đích dự báo.
- Mục tiêu của Chronos Forecaster là cung cấp một giải pháp dễ sử dụng và có thể mở rộng để dự báo chuỗi thời gian, tận dụng sức mạnh của học sâu và nền tảng Apache Spark.

Phương Pháp

TCNForecaster (BigDL-Chronos Forecaster)

- TCNForecaster là một mô hình dự báo chuỗi thời gian được xây dựng trên nền tảng BigDL. Nó sử dụng mạng neural Temporal Convolutional Network (TCN) để dự báo giá trị trong tương lai.
- TCN là một neural network sử dụng kiến trúc mạng tích chập thay vì mạng hồi quy. Nó hỗ trợ các trường hợp multi-step và multi-variant.
- Causal Convolutions cho phép tính toán song song quy mô lớn khiến TCN có thời gian suy luận ít hơn so với mô hình dựa trên RNN như LSTM.

Phương Pháp

Seq2SeqForecaster (BigDL-Chronos Forecaster)

- Seq2SeqForecaster trong BigDL là một mô hình dự báo chuỗi thời gian dựa trên mô hình seq2seq (sequence to sequence). Đây là một trong những kiến trúc mạnh mẽ để xử lý dữ liệu chuỗi thời gian sử dụng Deep Learning.
- Mô hình Seq2SeqForecaster được thiết kế để dự báo các giá trị liên tục trong tương lai dựa trên các giá trị đầu vào trong quá khứ.

Phương Pháp

LSTMForecaster (BigDL-Chronos Forecaster)

- Là một trong những mô hình sử dụng cho bài toán dự báo chuỗi thời gian được định nghĩa trong thư viện BigDL và có thể chạy trên nền tảng Apache Spark. Nó sử dụng một kiến trúc mạng LSTM để dự báo giá trị tại thời điểm trong tương lai bằng cách học từ các giá trị đầu vào trong quá khứ.
- LSTM là một kiến trúc mạnh mẽ và linh hoạt cho việc xử lý dữ liệu chuỗi thời gian, vì nó có khả năng giữ lại thông tin trong khoảng thời gian dài và cho phép mô hình phân tích các mối quan hệ giữa các giá trị trong quá khứ và tương lai.

Phương Pháp

Convolutional Neural Network (CNN)

- Là một kiểu mạng neural Deep Learning được sử dụng để phân loại hoặc dự báo dữ liệu hình ảnh. Nó sử dụng một quá trình convolution để tìm các đặc trưng cụ thể trong dữ liệu và đặt nền tảng cho việc phân loại hoặc dự báo.
- Dữ liệu time-series có thể được chuyển đổi thành một ma trận và được sử dụng như một hình ảnh đầu vào cho CNN. Nó sẽ sử dụng các lớp convolution để tìm các đặc trưng quan trọng trong dữ liệu và sau đó sử dụng các lớp fully connected để tạo ra một dự báo.

Phương Pháp

Gated Recurrent Unit (GRU)

- Là một loại neural network hồi quy dựa trên việc sử dụng cùng một hàm tính toán để xử lý các đầu vào liên tục. Trong RNN, trạng thái hiện tại được sử dụng để dự báo trạng thái tiếp theo và quá trình này được lặp đi lặp lại cho mỗi đầu vào trong chuỗi.
- GRU có thể hoạt động tốt với dữ liệu time-series bằng cách giữ nguyên những thông tin quan trọng và loại bỏ những thông tin không cần thiết. Vì vậy, GRU có thể dự báo các sự kiện trong tương lai dựa trên dữ liệu quá khứ và hiện tại.

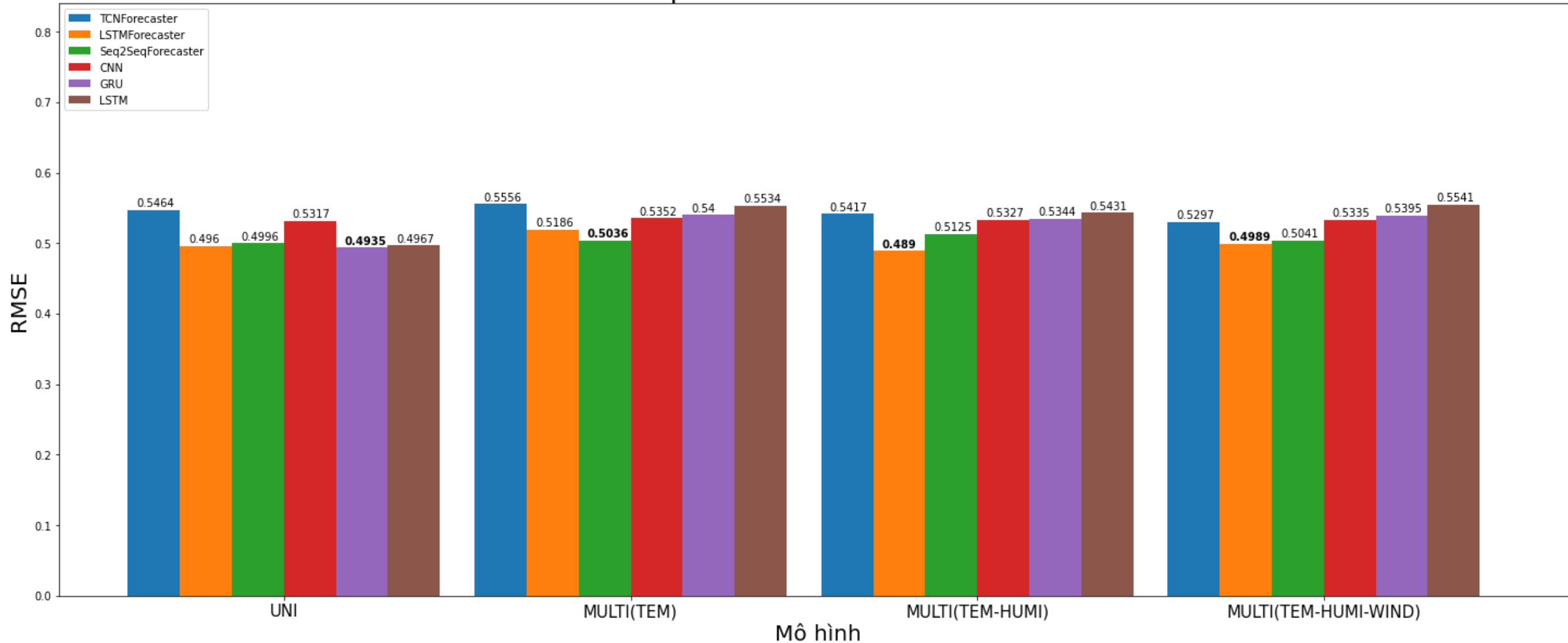
Phương Pháp

Long Short-Term Memory (LSTM)

- Là một loại mô hình RNN được thiết kế để xử lý dữ liệu chuỗi thời gian, trong đó một số thông tin của trạng thái trước được giữ lại để dùng cho các bước tiếp theo.
- So với GRU, LSTM có cấu trúc phức tạp hơn và yêu cầu nhiều tài nguyên hơn để huấn luyện, nhưng nó có khả năng xử lý tốt hơn một số bài toán NLP với dữ liệu chuỗi thời gian.

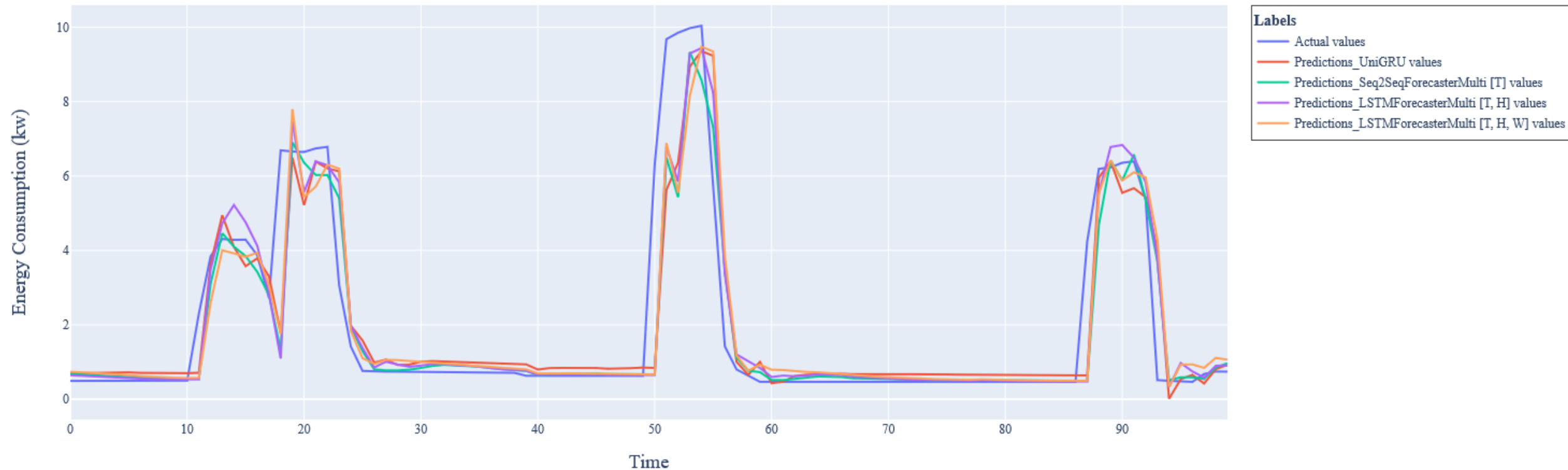
Thực Nghiệm

Kết quả RMSE của các mô hình



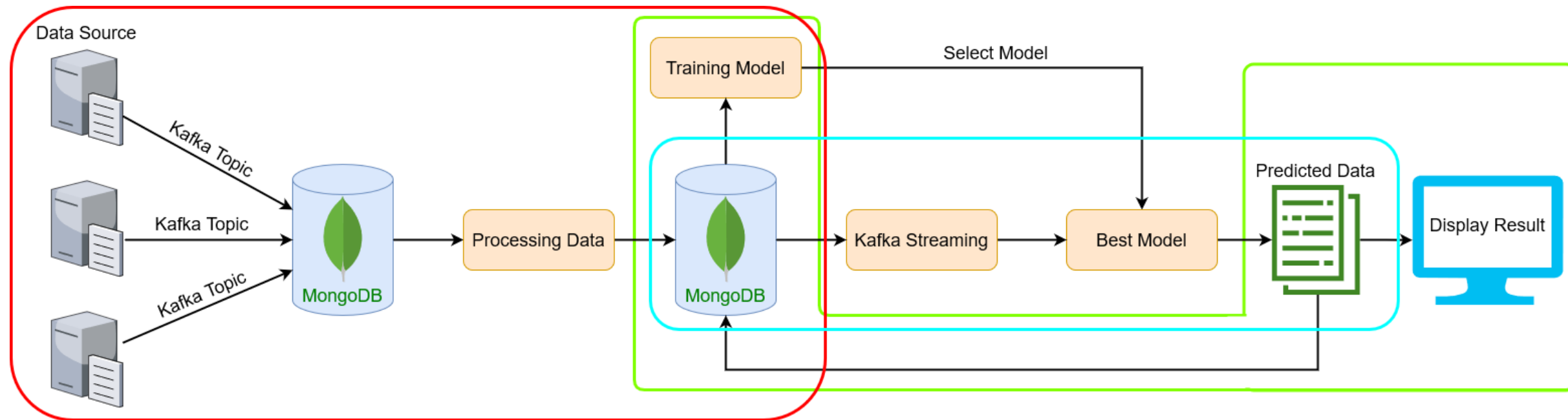
Thực Nghiệm

Result of Models on Test Set



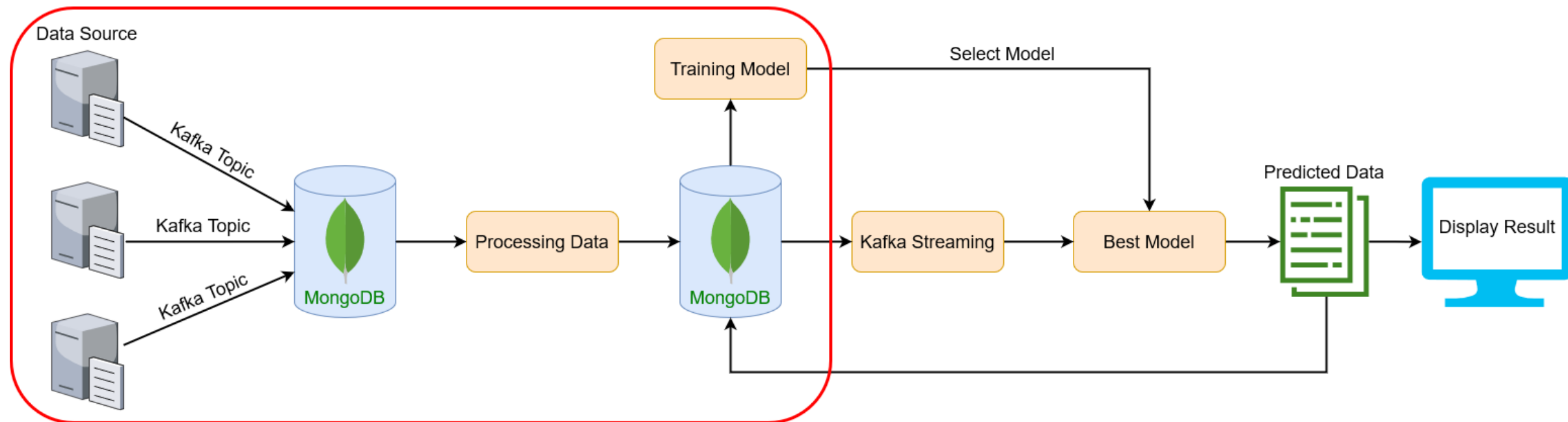
Kiến Trúc Hệ Thống

Tổng quan hệ thống



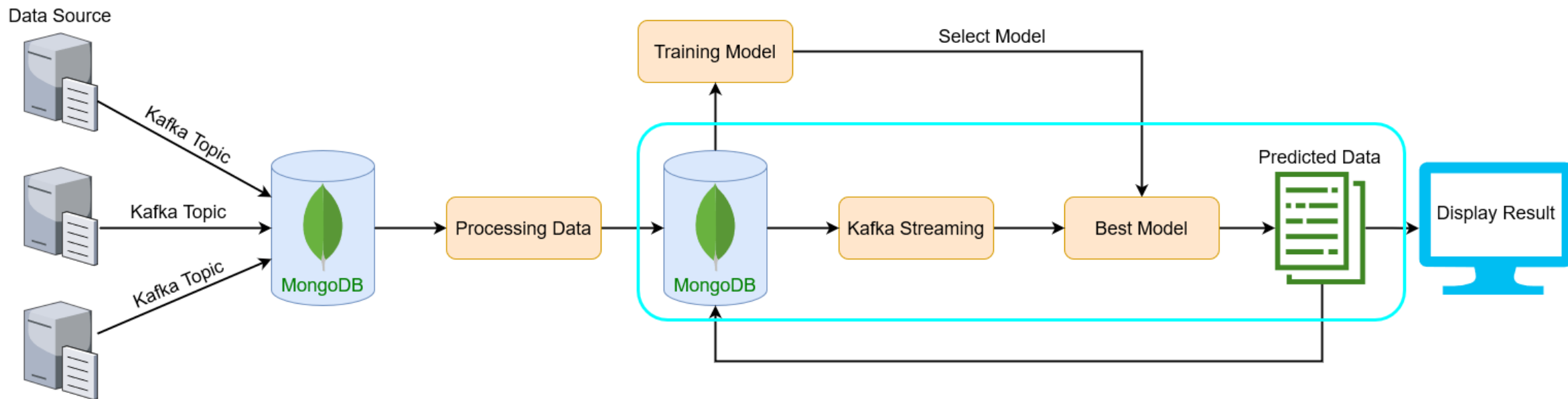
Kiến Trúc Hệ Thống

Giai đoạn 1: Xử lý dữ liệu & training model



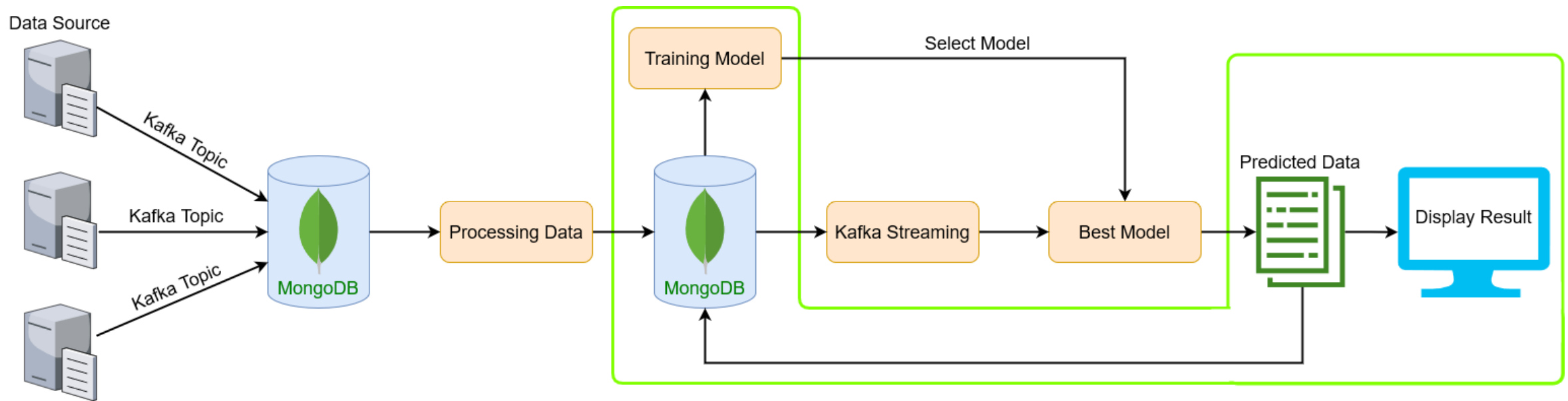
Kiến Trúc Hệ Thống

Giai đoạn 2: Chọn model & streaming predict data



Kiến Trúc Hệ Thống

Giai đoạn 3: Lưu lại dữ liệu streaming vào MongoDB



Kết Luận

- Chúng tôi đã tiến hành xử lý dữ liệu, lựa chọn phương pháp và tiến hành thực nghiệm để tìm ra kết quả tốt nhất. Từ đó xây dựng nên hệ thống dự đoán năng lượng tiêu thụ real-time để nâng cao sự thoải mái và an ninh cho con người với mức tiêu thụ năng lượng thấp và quản lý năng lượng hiệu quả.
- Hạn chế:
 - Kết quả dự đoán mang lại kết quả chưa thực sự cao.
 - Hệ thống vận hành chưa tốt.
- Hướng phát triển trong tương lai:
 - Áp dụng thêm nhiều mô hình: AutoformerForecaster, NBeatsForecaster, TCMFForecaster, MTNetForecaster, ARIMAForecaster, ProphetForecaster...
 - Kết hợp nhiều thuộc tính khác nhau trong bộ dữ liệu để có thể đưa ra kết quả dự đoán chính xác hơn.

THANKS FOR YOUR ATTENTION!



DEMO

https://youtu.be/2pS_oTZ_wEI