Syracuse University
School of Information Studies

Interactive Dashboard for Energy Consumption Analysis

Team 3
Esha Alva
Maithili Salankar
Atharva Gadiwan
Tejas Mistry

Introduction

- Energy consumption analysis involves examining how electricity or power is used by residential properties.
- Understanding energy usage patterns is crucial for several reasons.
- In the context of global warming and increasing energy demand, responsible energy management is essential.
- Data-driven insights play a pivotal role in optimizing energy resources and reducing environmental impact.

Problem Statement

- Energy Company: eSC is facing the challenge of increased demand for electricity.
- Cause: Global warming has led to hotter summers, resulting in higher energy consumption for cooling homes.
- Goal: eSC aims to reduce energy usage during hot summers to meet demand without building new energy production facilities.

Objective

- Objective 1: Identify key factors influencing energy usage, beyond temperature.
- Objective 2: Develop predictive models to forecast energy demand and anticipate peak periods.
- Objective 3: Design and implement targeted interventions to encourage customer energy conservation.
- Objective 4: Continuously monitor and measure the effectiveness of interventions, refining strategies as needed.
- Overall Goal: Ensure reliable energy supply for customers, mitigate environmental impact, and build a sustainable future.

Data Reading

1. Static House Data

Use of arrow package to reads the Parquet file and stored in new variable 'staticHouseData'

```
# 1. Load Static House Data
url_static <- "https://intro-datascience.s3.us-east-2.amazonaws.com/SC-data/static_house_info.parquet"
staticHouseData <- arrow::read_parquet(url_static)
```

2. Energy Usage and Weather Data

```
# 2. Load Energy Usage Data
url_base <- "https://intro-datascience.s3.us-east-2.amazonaws.com/SC-data/2023-houseData/"
# Initialize an empty data frame to store energy usage data
energyUsageData <- data.frame()

# Loop through each building ID and read corresponding energy consumption data
for (building_id in unique_building_id) {
  url <- paste0(url_base, building_id, ".parquet")
  # Use tryCatch to handle potential errors in reading data
  tryCatch({
    energy_data <- arrow::read_parquet(url)
    energyUsageData <- rbind(energyUsageData, energy_data)},
  error = function(e) {
    cat("Error for building_id:", building_id, "\n")})}
```

Data Manipulation

- Calculated total energy consumption
- Grouped by unique date-time and calculated hourly energy consumption
- Formatted date-time in Y-M-D-H format for merging and removed unnecessary variables
- Stored values in hourly_usage for energy usage data and in hourly_weather for weather data

```
# Calculate total energy consumption and format date-time
energy_columns <- grep("\\.energy_consumption$", names(energyUsageData), value = TRUE)
energyUsageData$total_energy_consumption <- rowSums(energyUsageData[energy_columns], na.rm = TRUE)
energyUsageData <- energyUsageData[, !(names(energyUsageData) %in% energy_columns)]
energyUsageData$time <- as.POSIXct(energyUsageData$time, format="%Y-%m-%d %H:%M:%S")
colnames(energyUsageData)[colnames(energyUsageData) == "time"] <- "date_time"

# Group by date-time and calculate hourly energy consumption
grouped <- energyUsageData %>%
  group_by(date_time) %>%
  summarise(hourly_energy_consumption = sum(total_energy_consumption))

hourly_usage <- as.data.frame(grouped)

# Format date-time for merging and remove unnecessary variables
hourly_usage$date_time <- format(hourly_usage$date_time, "%Y-%m-%d %H")
hourly_usage <- na.omit(hourly_usage)
```

Data Merging

- Merged Energy Usage and Weather Data
- Converted other Columns into numeric values

```r
# 4. Merge Energy Usage and Weather Data
merged_data <- merge(hourly_usage, hourly_weather, by = "date_time", all = TRUE)
merged_data <- na.omit(merged_data)

# Convert columns (excluding date_time) to numeric type
columns_to_convert <- setdiff(names(merged_data), "date_time")
merged_data[, columns_to_convert] <- lapply(merged_data[, columns_to_convert], as.numeric)
```

Modelling Techniques

- **XGBoost Model:**
- XGBoost is a machine learning algorithm known for its effectiveness in regression and classification tasks. Here's how it works:
  1. **Training Data:** Historical data on energy consumption, weather conditions, and other relevant features are used as input. The model learns from this data to identify patterns and relationships.
  2. **Prediction:** Once trained, the XGBoost model can make predictions on unseen data.
  3. **Evaluation:** The model's performance is assessed using Root Mean Squared Error (RMSE) metrics to measure how well its predictions align with actual energy consumption.

Time Series Model (ARIMA):

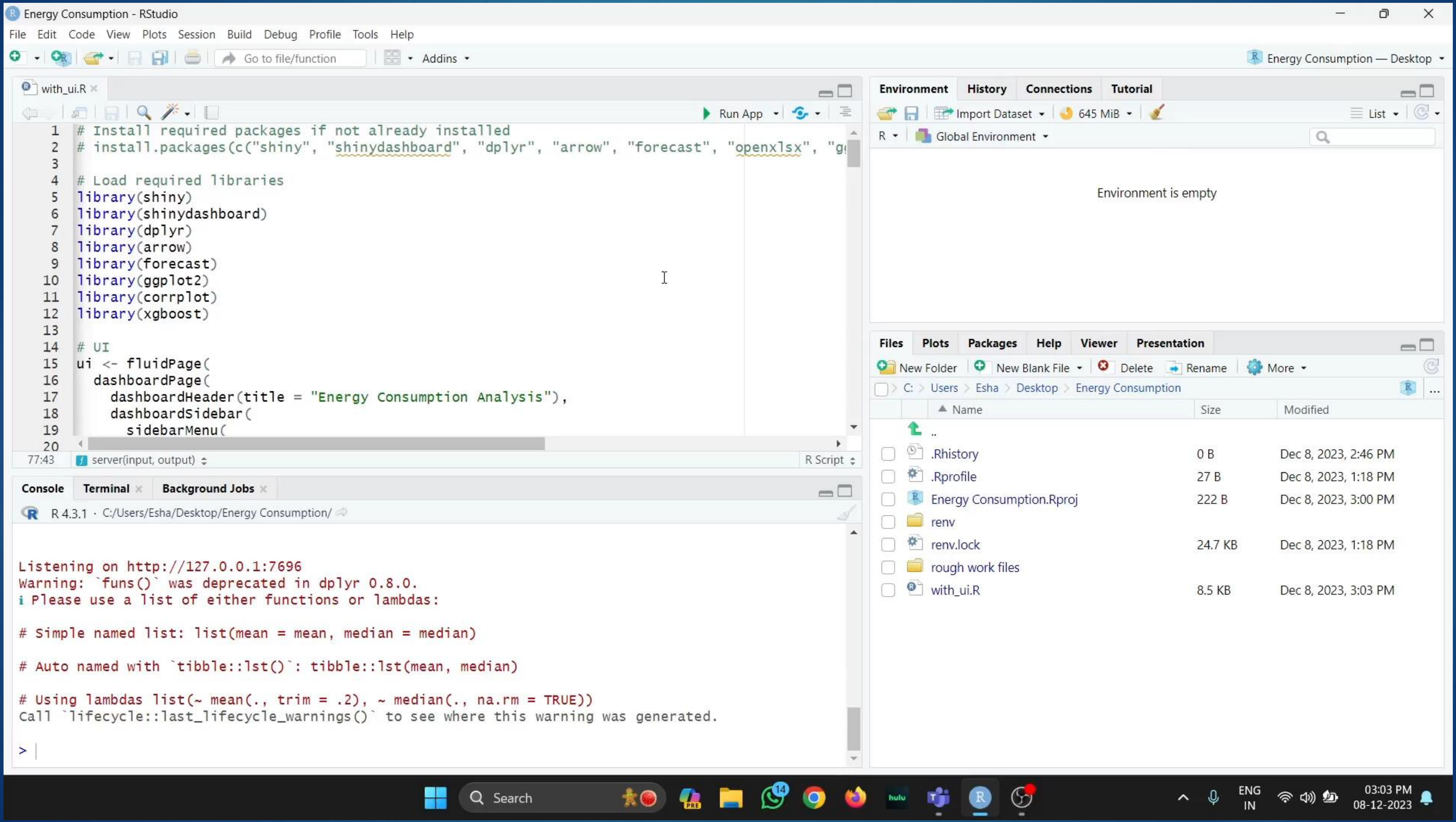
Time series models are specialized models used for forecasting time-dependent data, such as hourly energy consumption.

1. **Training Data:** The historical energy consumption data is converted into a time series format. The ARIMA model considers the patterns and seasonality in the data.
2. **Forecasting:** The ARIMA model is capable of generating forecasts for future time points. In this case, it can predict energy usage for upcoming hours or days during hot summers.
3. **RMSE Evaluation:** Similar to the XGBoost model, the ARIMA model's performance is evaluated using RMSE metrics.

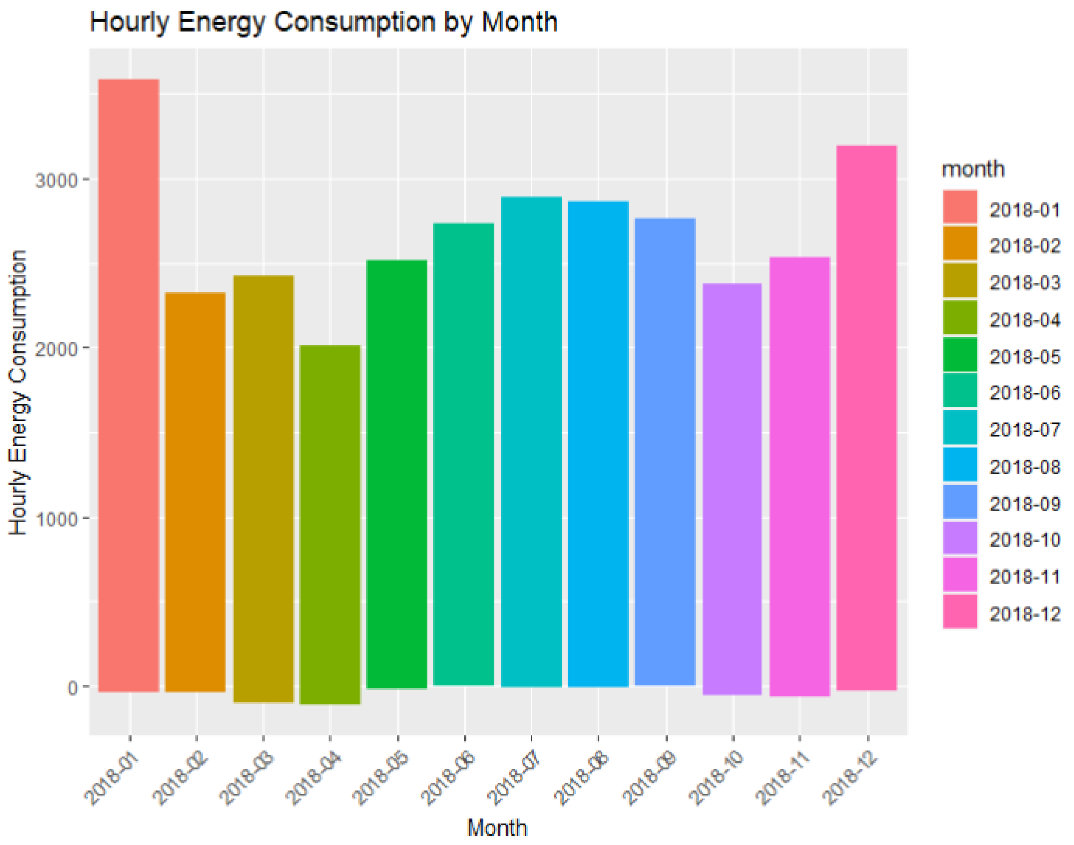
Data Visualization and Insights

Syracuse University
School of Information Studies

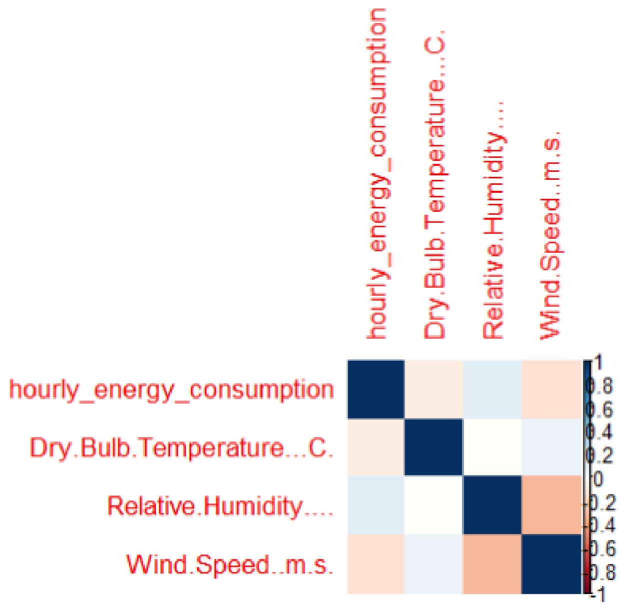
Live Demonstration



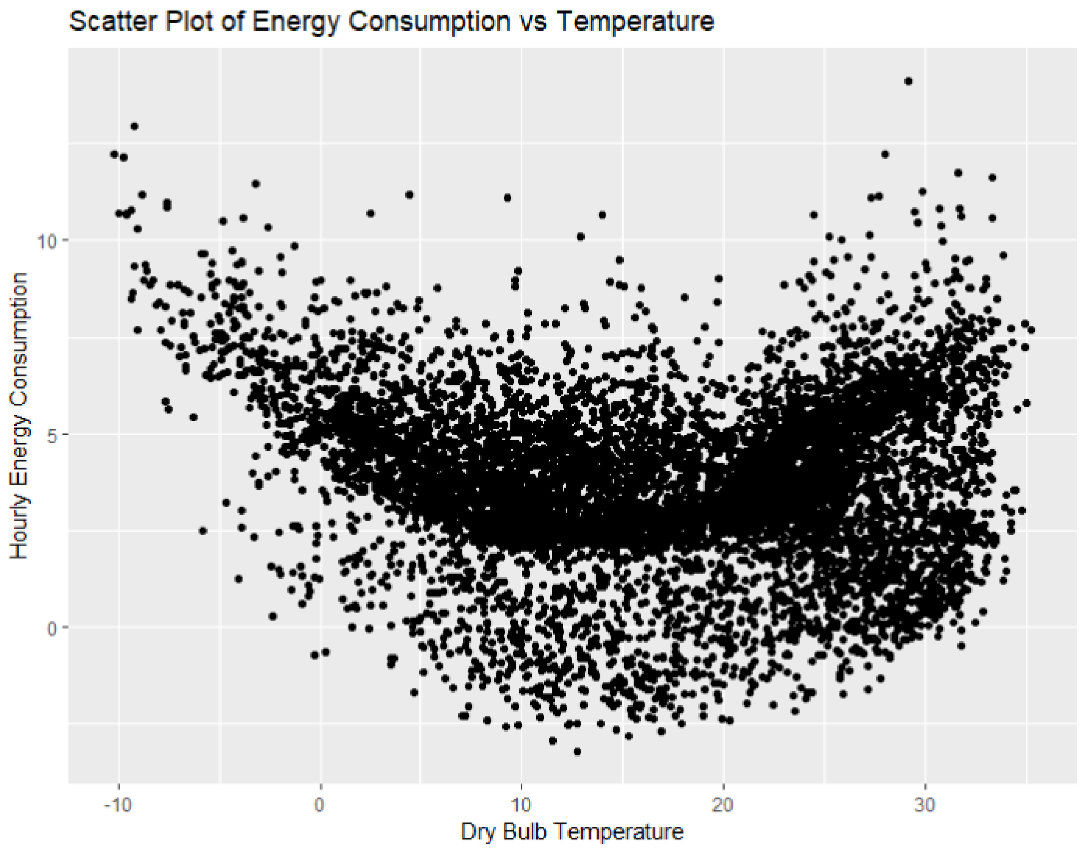
Monthly Energy Consumption



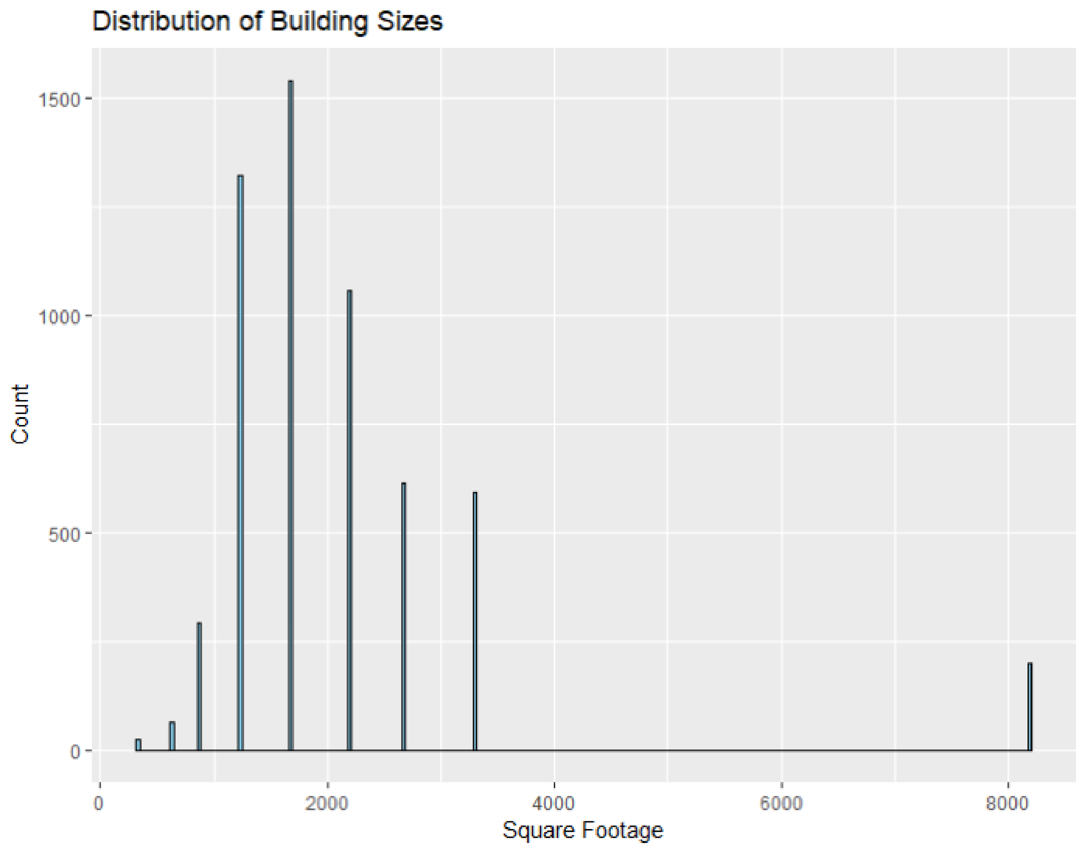
Correlation Plot



Scatter Plot



Building Size Distribution



Solutions

- Promote energy-saving behavior: Based on the correlation plot, dry bulb temperature is the strongest factor influencing energy consumption. Educational campaigns and personalized recommendations could encourage customers to adjust their thermostats during hot periods, reducing reliance on air conditioning.
- Invest in smart grid technologies: The time series model suggests predictable patterns in energy consumption. Implementing smart grid technologies could optimize resource allocation and manage peak demand more effectively.
- Offer incentives for energy efficiency: The building size distribution indicates various building types within eSC's customer base. Providing rebates or other incentives for energy-efficient upgrades and retrofits could target specific building sizes and achieve greater impact.
- Partner with local communities: Collaborating with local organizations and initiatives can amplify awareness and encourage community-wide energy conservation efforts.

Conclusion

- The analysis reveals the significant influence of dry bulb temperature on energy consumption.
- Building size is also a relevant factor, indicating the need for targeted interventions.
- Data-driven insights, including predictive models and correlation analysis, offer valuable tools for managing peak demand and promoting energy conservation.
- Implementing the proposed solutions can help eSC:
    - Ensure reliable energy supply for customers.
    - Reduce peak demand and grid overload risks.
    - Enhance environmental sustainability.
    - Optimize resource allocation and cost savings.
    - Increase customer satisfaction and engagement.