Analyzing Machine and Sensor Data

Visualizations and Insights using R

By:
Anish Anand

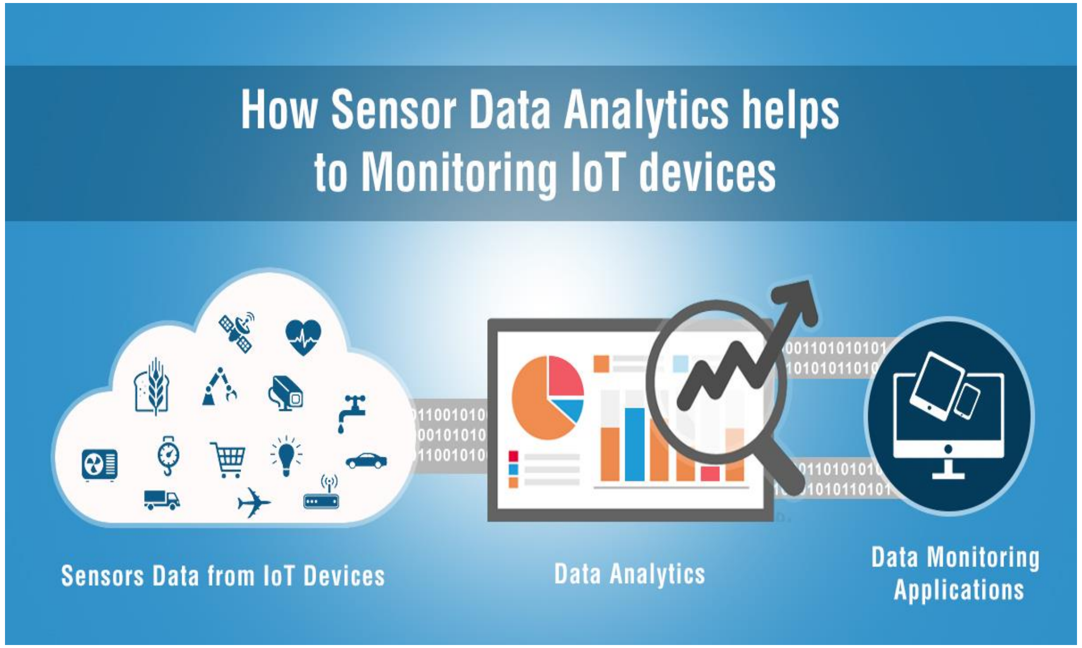
Analyzing Machine & Sensor Data

Sensor data

Sensor data is the output of a device that detects and responds to some type of input from the physical environment. The output may be used to provide information or input to another system or to guide a process. Sensor data is in integral component of the increasing reality of the Internet of Things (IoT) environment. In the IoT scenario, almost any entity imaginable can be outfitted with a unique identifier (UID) and the capacity to transfer data over a network. The huge volume of data produced and transmitted from sensing devices can provide a lot of information but is often considered the next big data challenge for businesses. To deal with that challenge, sensor data analytics is a growing field of endeavour.

Sensor Analytics

Sensor analytics is the statistical analysis of data that is created by wired or wireless sensors. A primary goal of sensor analytics is to detect anomalies. The insight that is gained by examining deviations from an established point of reference can have many uses, including predicting and proactively preventing equipment failure in a manufacturing plant, alerting a nurse in an electronic intensive care unit (eICU) when a patient's bloodpressure drops, or allowing a data centre administrator to make data-driven decisions about heating, ventilating and air conditioning (**HVAC**).

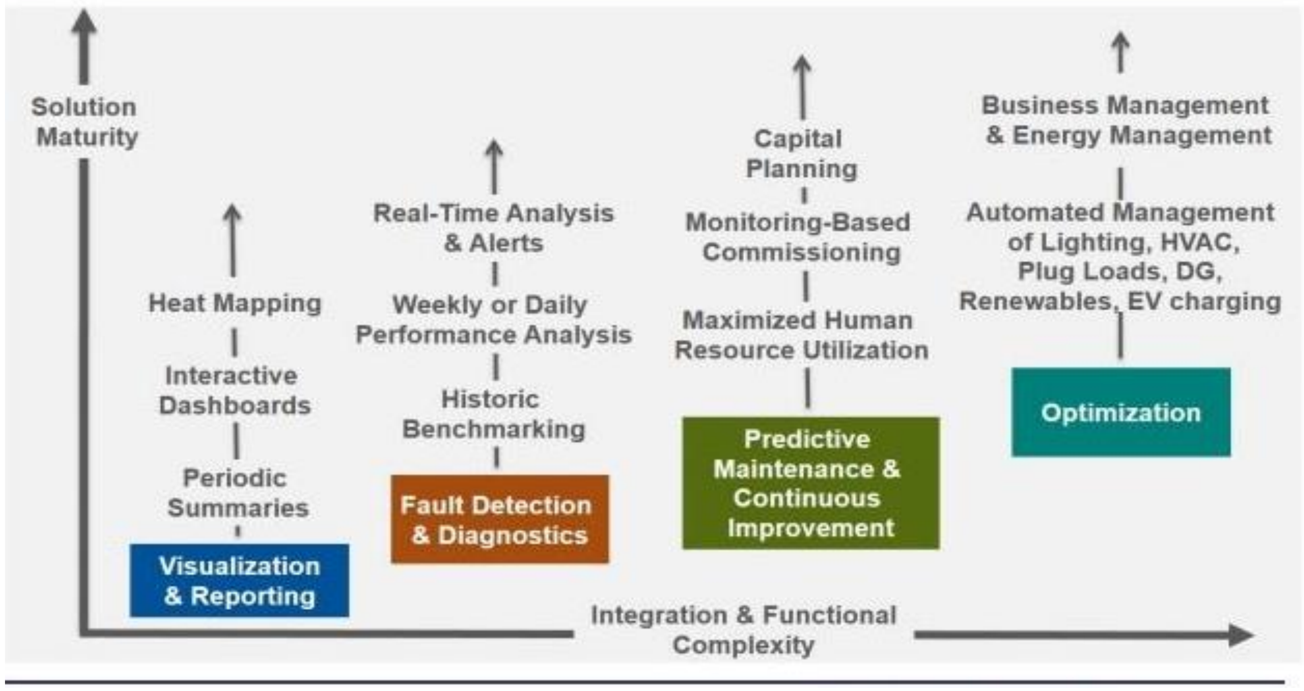


Analyzing Machine & Sensor Data

HVAC (heating, ventilating/ventilation, and air conditioning)

HVAC (heating, ventilating/ventilation, and air conditioning) is the technology of indoor and vehicular environmental comfort. Its goal is to provide thermal comfort and acceptable indoor air quality. HVAC system design is a sub discipline of mechanical engineering, based on the principles of thermodynamics, fluid mechanics, and heat transfer. Refrigeration is sometimes added to the field's abbreviation as HVAC&R or HVACR, or ventilating is dropped as in HACR (such as the designation of HACR-rated circuit breakers).

HVAC is an important part of residential structures such as single family homes, apartment buildings, hotels and senior living facilities, medium to large industrial and office buildi-ngs such as skyscrapers and hospitals, on-board vessels, and in marine environments, where safe and healthy building conditions are regulated with respect to temperature and humidity, using fresh air from outdoors.



Analyzing Machine & Sensor Data

BUSINESS OBJECTIVE

Sensor Data

A sensor is a device that measures a physical quantity and transforms it into a digital signal. Sensors are always on, capturing data at a low cost, and powering the “Internet of Things.”

Potential Uses of Sensor Data

Sensors can be used to collect data from many sources, such as:

- To monitor machines or infrastructure such as ventilation equipment, bridges, energy meters, or airplane engines. This data can be used for predictive analytics, to repair or replace these items before they break.
- To monitor natural phenomena such as meteorological patterns, underground pressure during oil extraction, or patient vital statistics during recovery from a medical procedure.

This project is based on about how to refine data from heating, ventilation, and air conditioning (HVAC) systems, and how to analyze the refined sensor data to maintain optimal building temperatures

Expected Output:

The company requires us to accomplish three goals with this data:

- Reduce heating and cooling expenses.

Analyzing Machine & Sensor Data

- Keep indoor temperatures in a comfortable range between 65-70 degrees.
- Identify which HVAC products are reliable, and replace unreliable equipment with those models.

These analysis will be very helpful for facilities department to initiate data-driven strategies to reduce energy expenditures and improve employee comfort.

Analysis need to be performed:

- Data visualization/analysis by mapping the buildings that are most frequently outside of the optimal temperature range. Calculate count of extremetemp (i.e. where the temperature was more than five degrees higher or lower than the target temperature) by each country and temprange
- Which country offices run hot (Hot offices can lead to employee complaints and reduced productivity) and which offices run cold (Cold offices cause elevated energy expenditures and employee discomfort). Calculate count of offices run in hot and count of office run in cold by country.
- Our data set includes information about the performance of five brands of HVAC equipment, distributed across many types of buildings in a wide variety of climates. We can use this data to assess the relative reliability of the different HVAC models(i.e We can see that the which model seems to regulate temperature most reliably and maintain the appropriate temperature range). Calculate count of extremetemp by hvacproduct .

Analyzing Machine & Sensor Data

DATA AVAILABILITY

The analysis had to be performed using the following file format given:

- ❑ **building.csv** :- A comma separated value file (.csv format) file of building id and its details was given which contained semi-structure format data. The file contains 5 fields namely
- BuildingID
- BuildingMgr
- BuildingAge
- HVACproduct and
- Country
- ❑ **HVAC.csv** :- A comma separated value file (.csv format) file of sensor log(HVAC log) and its details was given which contained semi-structure format data. The sensor contains the temperature reading sensor took at a particular time in particular building.

  The file contains fields namely
- **Date, Time**
- **TargetTemp, ActualTemp**
- **System,SystemAge and**
- **BuildingID**

building.csv


```
BuildingID,BuildingMgr,BuildingAge,HVACproduct,Country1,M1,25,AC1000,USA
2,M2,27,FN39TG,France3,M3,28,JDNS77,Brazil4,M4,17,GG1919,Finland5,M5,3,ACM
6,M6,9,AC1000,Singapore7,M7,13,FN39TG,South Africa8,M8,25,JDNS77,Australia
9,M9,11,GG1919,Mexico10,M10,23,ACMAX22,China11,M11,14,AC1000,Belgium
12,M12,26,FN39TG,Finland13,M13,25,JDNS77,Saudi Arabia14,M14,17,GG1919,Gern
15,M15,19,ACMAX22,Israel16,M16,23,AC1000,Turkey17,M17,11,FN39TG,Egypt
18,M18,25,JDNS77,Indonesia19,M19,14,GG1919,Canada20,M20,19,ACMAX22,Arger
```


HVAC.csv


```
Date,Time,TargetTemp,ActualTemp,System,SystemAge,BuildingID6/1/13,0:00:01,66,58,13,20,46/2/13,1:00:01,69,68,3,
6/5/13,4:00:01,68,74,16,9,36/6/13,5:00:01,67,56,13,28,46/7/13,6:00:01,70,58,12,24,26/8/13,7:00:01,70,73,20,26,1
6/11/13,10:00:01,67,70,10,17,156/12/13,11:00:01,69,62,2,11,76/13/13,12:00:01,69,73,14,2,156/14/13,13:00:01,65
6/17/13,16:00:01,67,57,15,7,66/18/13,17:00:01,66,57,12,5,136/19/13,18:00:01,69,58,8,22,46/20/13,19:00:01,67,5
6/23/13,22:00:01,67,65,6,18,206/24/13,23:00:01,67,61,15,13,66/25/13,0:13:19,70,71,19,14,186/26/13,1:13:19,67,
6/29/13,4:13:19,66,77,7,6,136/30/13,5:13:19,70,63,7,13,56/1/13,6:13:19,68,66,11,15,36/2/13,7:13:19,69,60,14,1,1
/6/13,11:13:19,67,72,16,8,196/7/13,12:13:19,67,73,17,30,26/8/13,13:13:19,65,75,4,13,106/9/13,14:13:19,67,68,2,
6/12/13,17:13:19,69,64,18,25,26/13/13,18:13:19,67,68,2,14,16/14/13,19:13:19,68,75,9,22,56/15/13,20:13:19,69,5
6/18/13,23:13:19,66,75,1,13,126/19/13,0:43:51,70,65,11,9,56/20/13,1:43:51,69,65,7,30,26/21/13,2:43:51,66,56,16
6/24/13,5:43:51,70,63,15,30,206/25/13,6:43:51,65,63,11,13,176/26/13,7:43:51,67,78,13,22,186/27/13,8:43:51,68,
6/30/13,11:43:51,68,67,8,17,66/1/13,12:43:51,69,79,6,15,76/2/13,13:43:51,65,72,20,26,126/3/13,14:43:51,65,57,3
6/6/13,17:43:51,67,79,1,9,106/7/13,18:43:51,67,72,13,29,156/8/13,19:43:51,69,76,11,25,116/9/13,20:43:51,65,67,
6/12/13,23:43:51,68,69,7,26,206/13/13,0:13:20,67,77,8,19,126/14/13,1:13:20,70,68,6,1,16/15/13,2:13:20,70,57,7,
6/18/13,5:13:20,70,58,6,17,196/19/13,6:13:20,69,75,9,21,66/20/13,7:13:20,69,75,13,14,166/21/13,8:13:20,69,55,1
6/24/13,11:13:20,65,65,19,24,26/25/13,12:13:20,68,55,14,5,206/26/13,13:13:20,68,67,18,28,96/27/13,14:13:20,66
6/30/13,17:13:20,65,57,17,9,126/1/13,18:13:20,68,65,7,21,126/2/13,19:13:20,65,72,4,1,136/3/13,20:13:20,66,67,1
6/6/13,23:13:20,70,70,13,3,56/7/13,0:33:07,68,72,18,8,16/8/13,1:33:07,70,80,16,24,96/9/13,2:33:07,67,76,16,23,9
```

Analyzing Machine & Sensor Data

APPROACH

In this problem, we will focus on sensor data from building operations. Specifically, we will refine and analyze the data from Heating, Ventilation, Air Conditioning (HVAC) systems in 20 large buildings around the world
The basic approach is to go through the dataset and see the factors which can lead to increase in temperature of offices beyond the NORMAL range. Those factors can be listed as :

- Geographical location
- Age of building
- HVAC product installed
- Age of system
- The difference between Actual Temperature and Target temperature can be used to predicted the extremities of excess temperature.

The above factors will be taken into consideration and basic merge operation is to be done using the building id as primary key on the data. After that subsequent feature engineering will be performed by making new variables corresponding to Temperature Difference, Temperature range and Extreme Temperature.

Further visualizations can be made by grouping the variables and plotting in R using ggplot2 or using power view of excel 2013.

Analyzing Machine & Sensor Data

Steps

1. Install and load the relevant packages.
2. Getting the Data Into R
   the whole data for the sensor analysis problem is given by 2 separate csv-format files:

- **building**
- **HVAC**

We will import the above files in R using *read.csv()* and extract the relevant features from the omniture logs file by sub setting the data frame.

3. Making three variables (temp_diff, temprange, extremetemp) in Hvac table

Temp_didff = actual temperature – target temperature

temprangecolumn indicates whether the actual temperature was:

- NORMAL – within 5 degrees of the target temperature.
- COLD – more than five degrees colder than the target temperature.
- HOT – more than 5 degrees warmer than the target temperature.

Extrememetemp: If the temperature is outside of the normal range, extremetemp is assigned a value of 1; otherwise its value is 0.

4. Proper joining of all 2 data frames using relevant key(BuildingID) is done so as to create a full sensor data-frame of 8000 observations and 14 features.

Analyzing Machine & Sensor Data

5. We don’t have any missing values in the frame so we can proceed to change the date format in default R format using *as.Date()* function.
6. At the end, proper visualizations are made using *qplot()* and *ggplot()* function so as to find relevant insights on the data and find the target audience and area for selling our products.

Analyzing Machine & Sensor Data

Visualizations

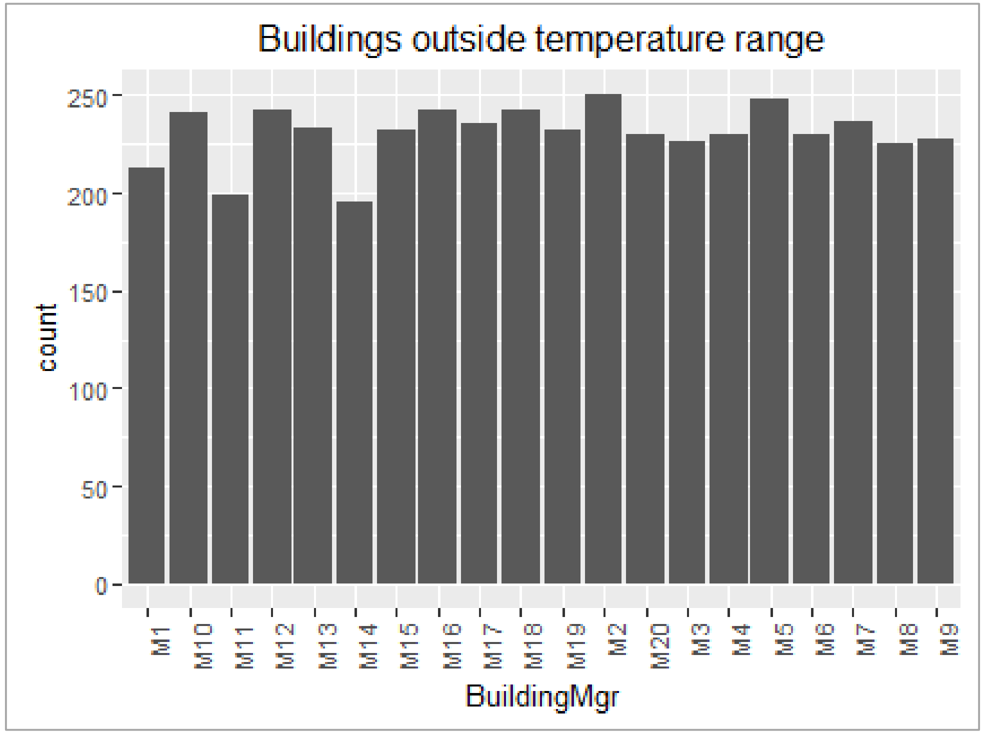
Hereby, I have included visualizations done on both R and Excel 2013(powerview). The code of visualizations in R are however shared in the code at the end.

Expected Outcomes

- ❑ **Data visualization/analysis by mapping the buildings that are most frequently outside of the optimal temperature range. Calculate count of extremetemp by each country and temprange**

The above problem can be divided into two parts

I. Data visualization of buildings outside of the optimal temperature range
II. Count of extreme temperature by each country

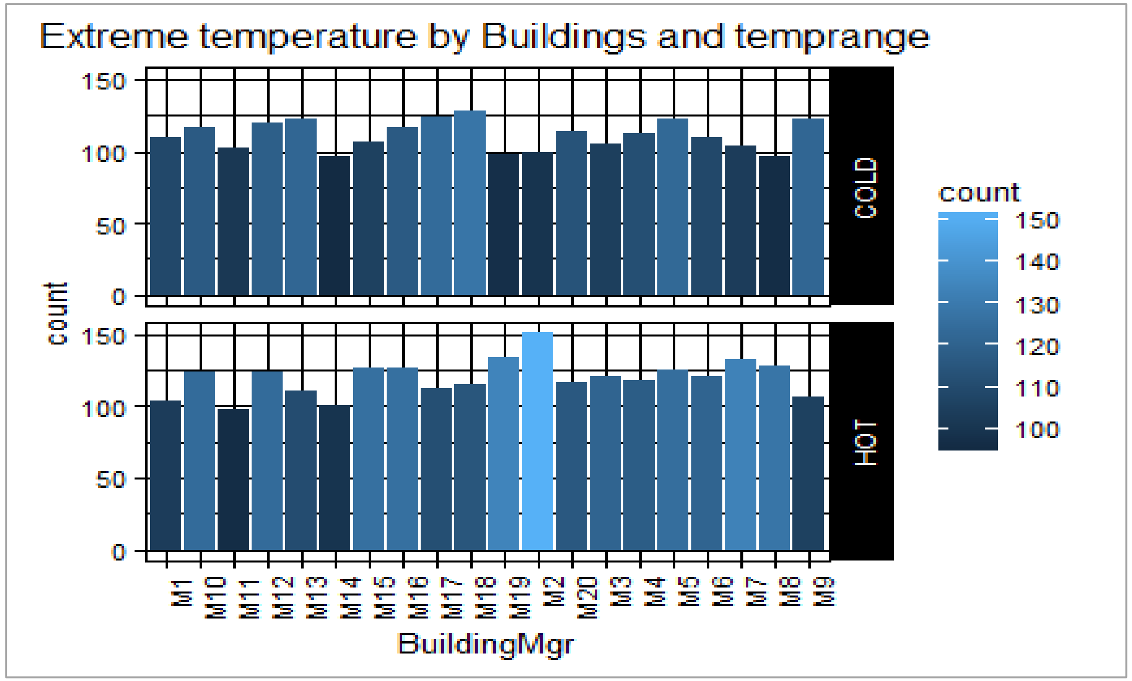


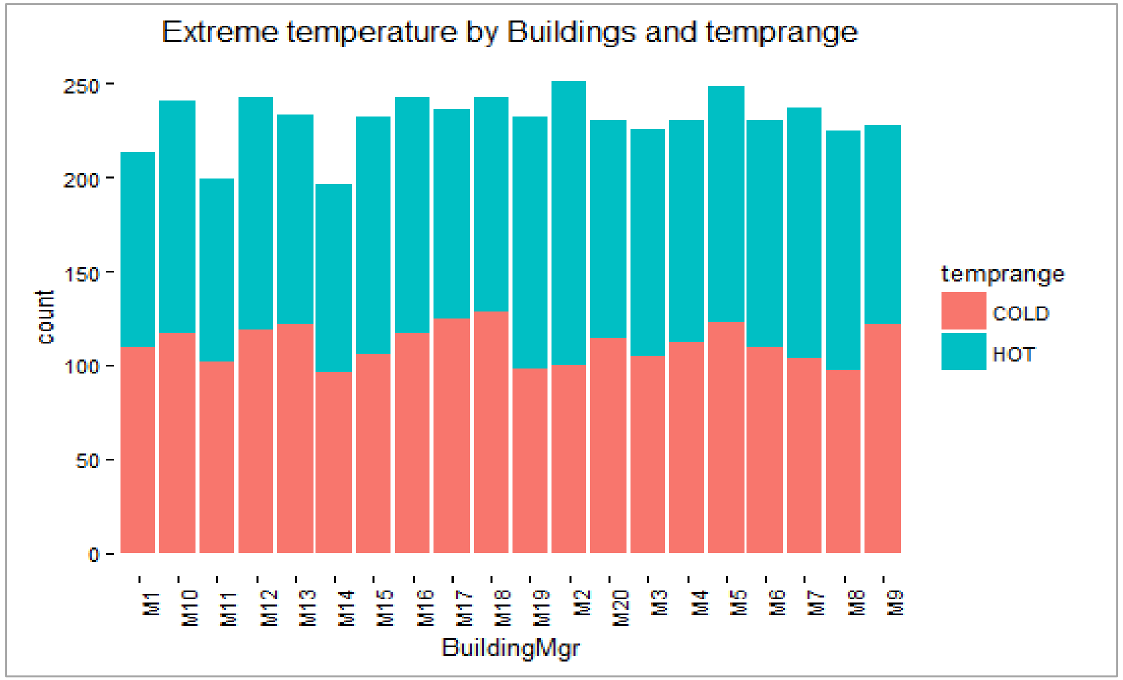
Analyzing Machine & Sensor Data

(I) The plot below corresponds to the buildings facing extreme temperatures. Perhaps due to geographic location of country in which it is situated or due to failure in the working of HVAC equipment installed in the building. The building outside normal temperature range due to faulty HVAC equipment can be seen as follows

- Building with ID 18,17,5,13,9 across the globe are not able to keep the Cold temperature under range
- Building with ID 2,19,7,8,16 across the globe are not able to keep the Hot temperature under range

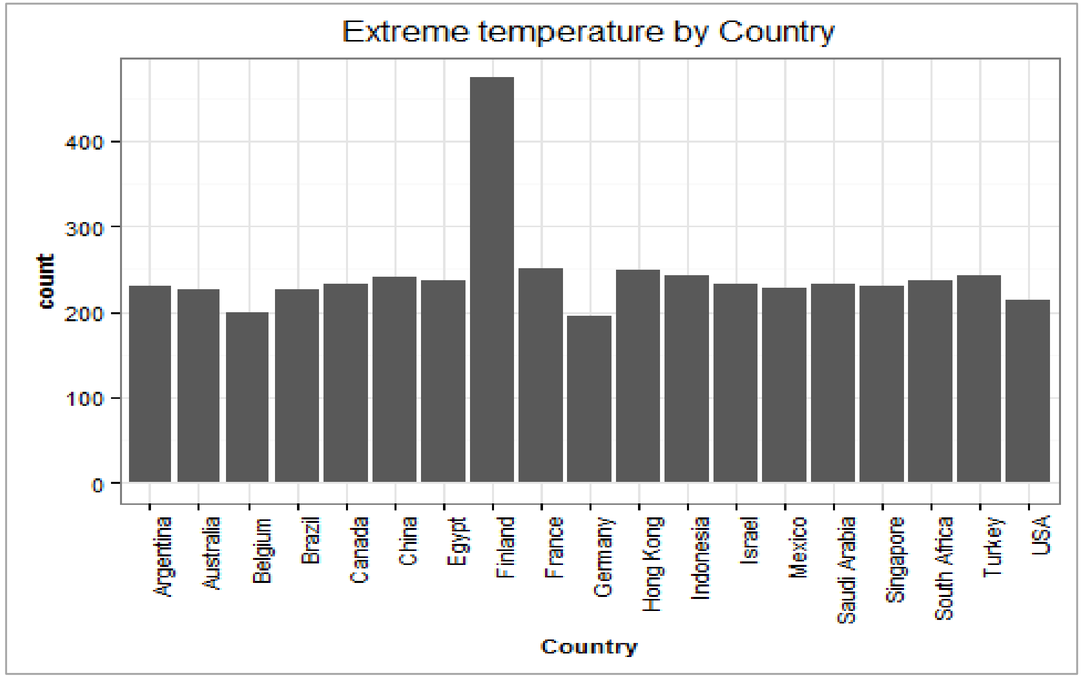
Moreover, the left plot shows the buildings with count as function of color and left plot corresponds to stacked bar chart

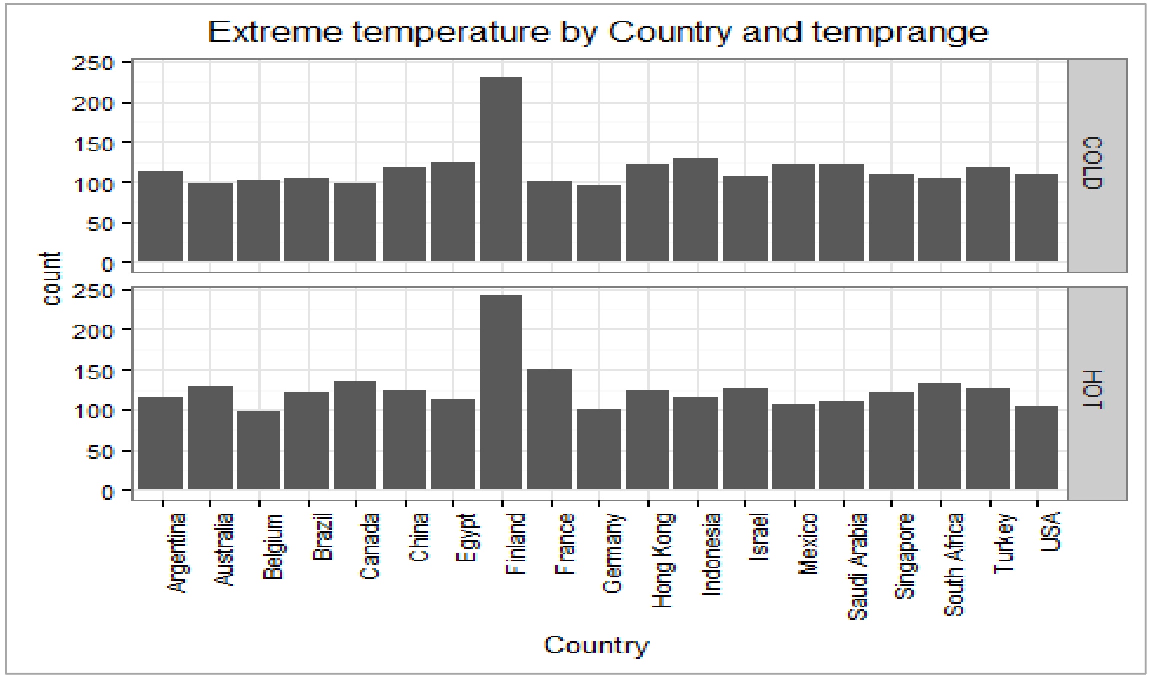




Analyzing Machine & Sensor Data

(II) Finland, France and Hong Kong top the list with most count of extremetemp (i.e where the temperature was more than five degrees higher or lower than the target temperature), which shows that the HVAC equipment of these countries are more sensitive to adverse climatic conditions and require further maintenance on them. However, the difference of count is not that significant, yet we can see that Finland tops the list and requires special concern in terms of HVAC equipment's reinstallation/ maintenance in buildings.





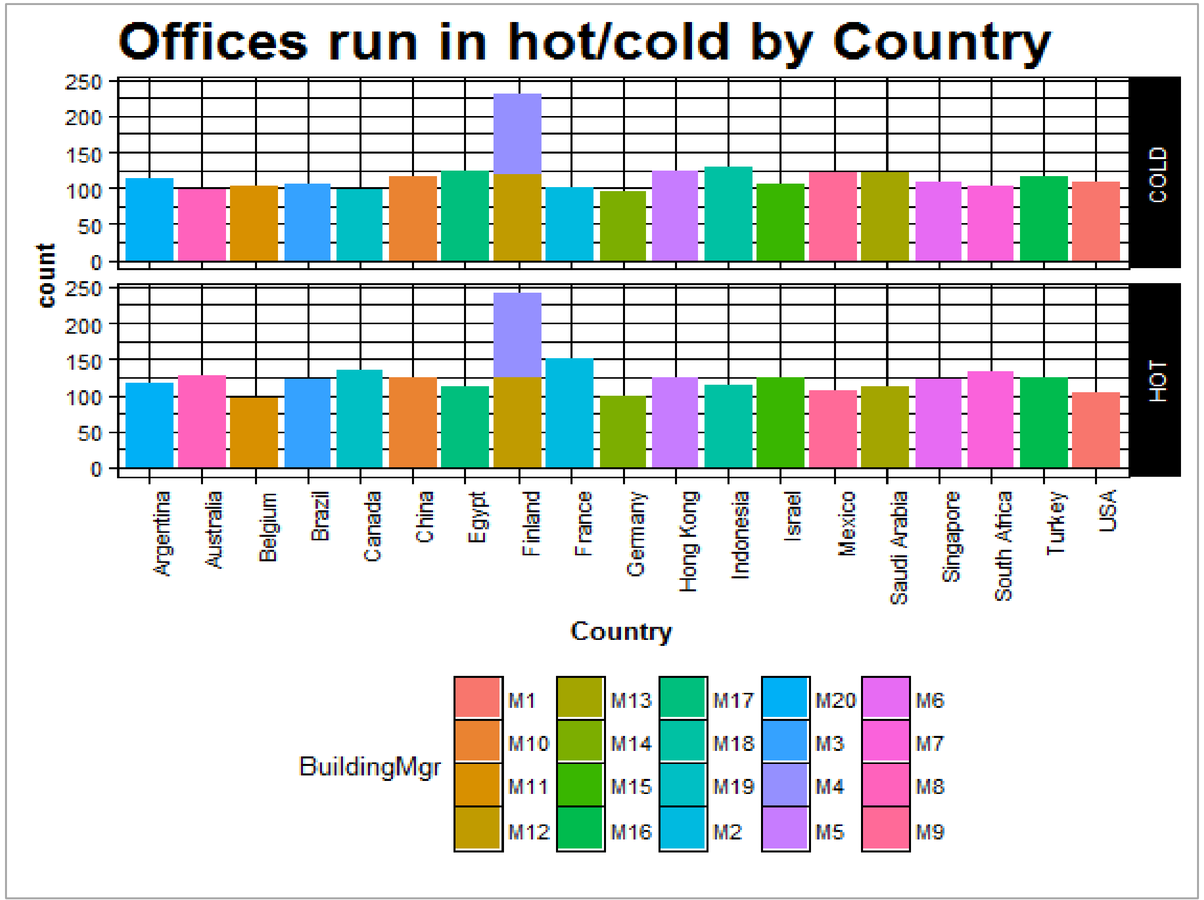
Analyzing Machine & Sensor Data

- ❑ **Which country offices run hot (Hot offices can lead to employee complaints and reduced productivity) and which offices run cold (Cold offices cause elevated energy expenditures and employee discomfort). Calculate count of offices run in hot and count of office run in cold by country.**

Among all the offices, the offices situated in Finland runs at extreme hot and cold temperature. In the adjacent plot, we can see that there are 2 offices,M12 and M4 which contribute to overall count of extreme temperature in Finland. By seeing the building age from the data, we also found out that M12 buildings were beyond 26 years of age,hence improper maintenance of HVAC equipment could have contributed to the same.

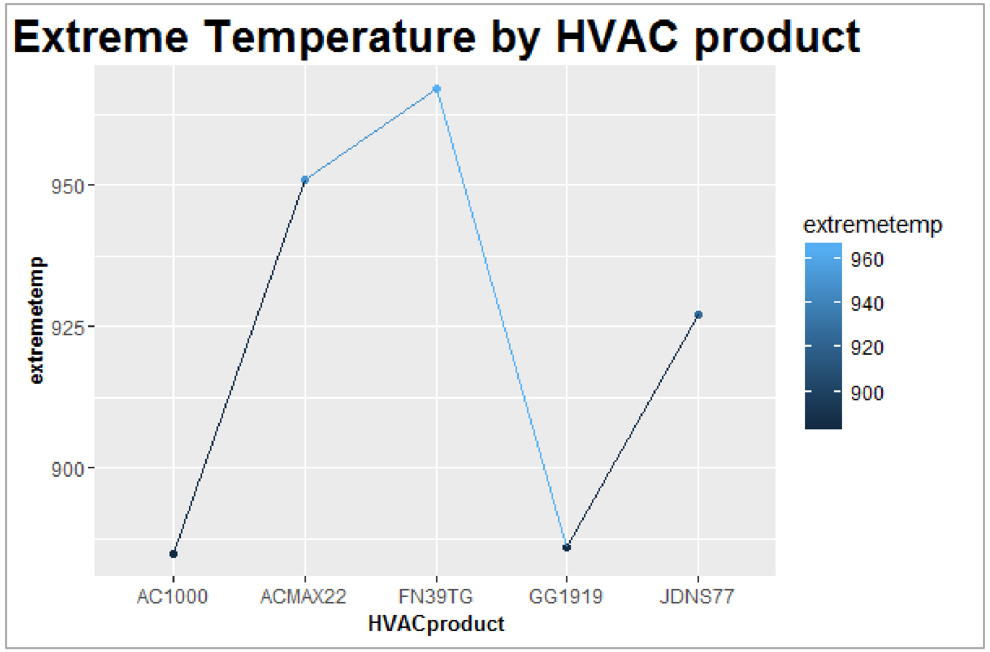
Moreover, we can see that the geographical locations of the countries also contributes in the extreme temperature faced at offices.

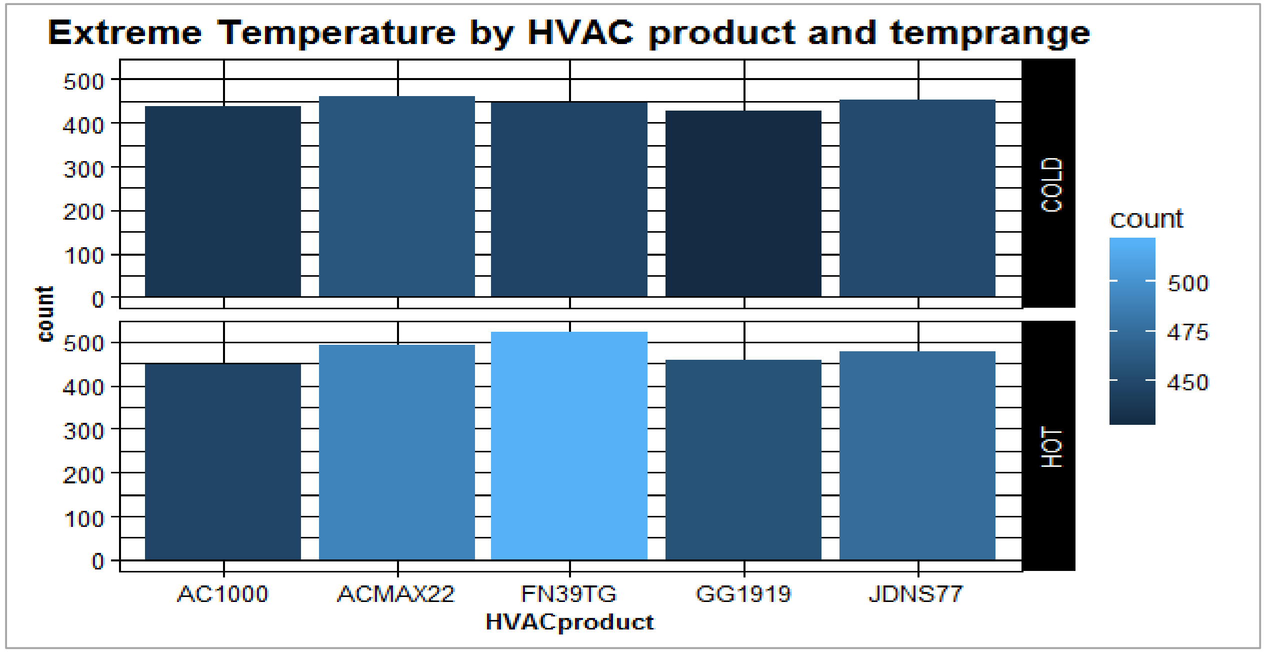
- Offices with ID M18,M17,M5,M13,M9 across the globe are not able to keep the Cold temperature under range
- Building with ID M2,M19,M7,M8,M16 across the globe are not able to keep the Hot temperature under range



Analyzing Machine & Sensor Data

- ❑ Our data set includes information about the performance of five brands of HVAC equipment, distributed across many types of buildings in a wide variety of climates. We can use this data to assess the relative reliability of the different HVAC models. Calculate count of extremetemp by hvacproduct





Performance of the HVAC product FN39TG is the poorest among all, in terms of regulating the temperature of the office. Moreover, the product ACMAX22 is the worst when it comes to regulating office temp. during cold temperature and FN39TG is worst when it comes to dealing with hot temperature. The buildings having HVAC equipment AC1000 and GG1919 have best temperature control as compared to others.