**-- Creating Table**

**-- Creating HVAC and Building Table**

create table hvac

( H\_date string,

time string,

target\_temp int,

actual\_temp int,

system int,

system\_age int,

building\_id string)

ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde'

WITH SERDEPROPERTIES

(

"separatorChar" = ",",

"quoteChar" = "\""

)

tblproperties('serialization.null.format'='','skip.header.line.count'='1');

**-- create table building**

(building\_id string,

building\_mgr string,

building\_age int,

hvac\_product string,

country string)

ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde'

WITH SERDEPROPERTIES

(

"separatorChar" = ",",

"quoteChar" = "\""

)

tblproperties('serialization.null.format'='','skip.header.line.count'='1');

**-- Creating HVAC temperature Variables and merging every table**

CREATE TABLE hvac\_temperature as

select

\*,

actual\_temp - target\_temp as temp\_diff,

case

when (actual\_temp - target\_temp) > 5 THEN 'HOT'

when (actual\_temp - target\_temp) < -5 THEN 'COLD'

else 'NORMAL'end

as temp\_range,

case

when abs(actual\_temp - target\_temp) <5 THEN '0'

else '1' end

AS extreme\_temp

from hvac;

create table hvac\_building as

select

ht.\*,

b.building\_mgr,

b.building\_age,

b.hvac\_product,

b.country

from hvac\_temperature ht inner join building b on ht.building\_id = b.building\_id;

**-- Analysis**

**-- Q1. 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**

**-- or lower than the target temperature) by each country and temp range.**

select

building\_id,

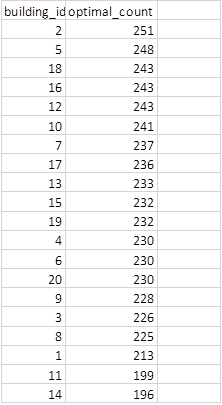
count(temp\_range) as optimal\_count

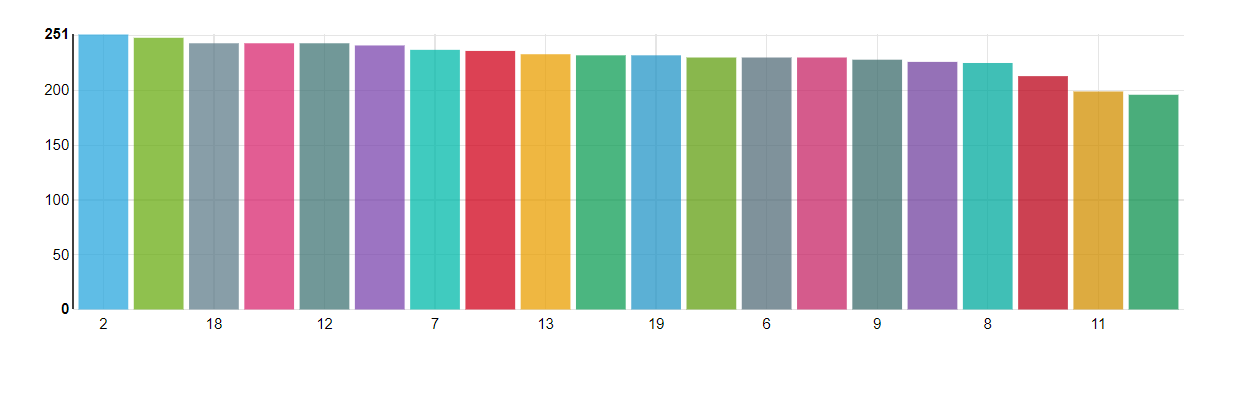
from hvac\_building

where temp\_range !='NORMAL'

group by building\_id

order by optimal\_count desc;





select

country,

temp\_range,

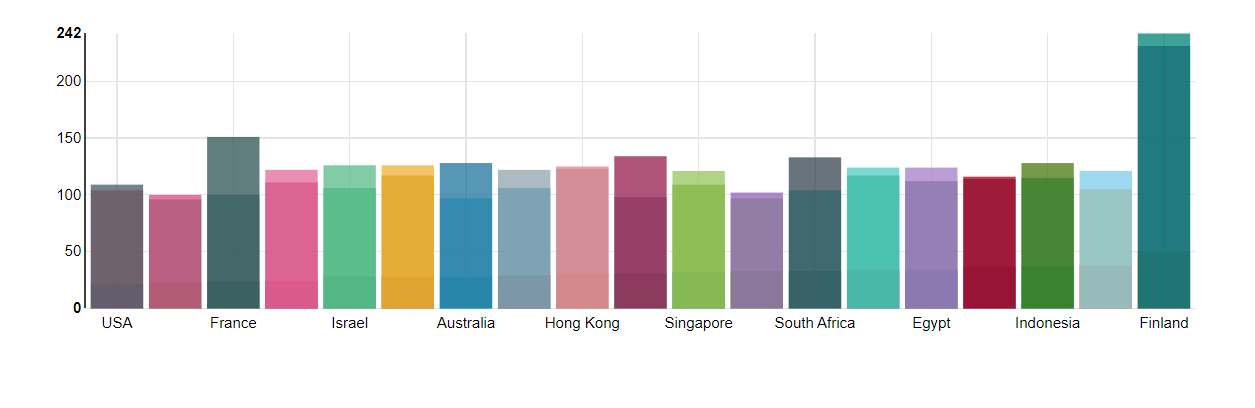
sum(cast(extreme\_temp as int)) as extreme\_temperature

from hvac\_building

group by country, temp\_range

order by extreme\_temperature;





**-- Q2. 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.**

select

country,

temp\_range,

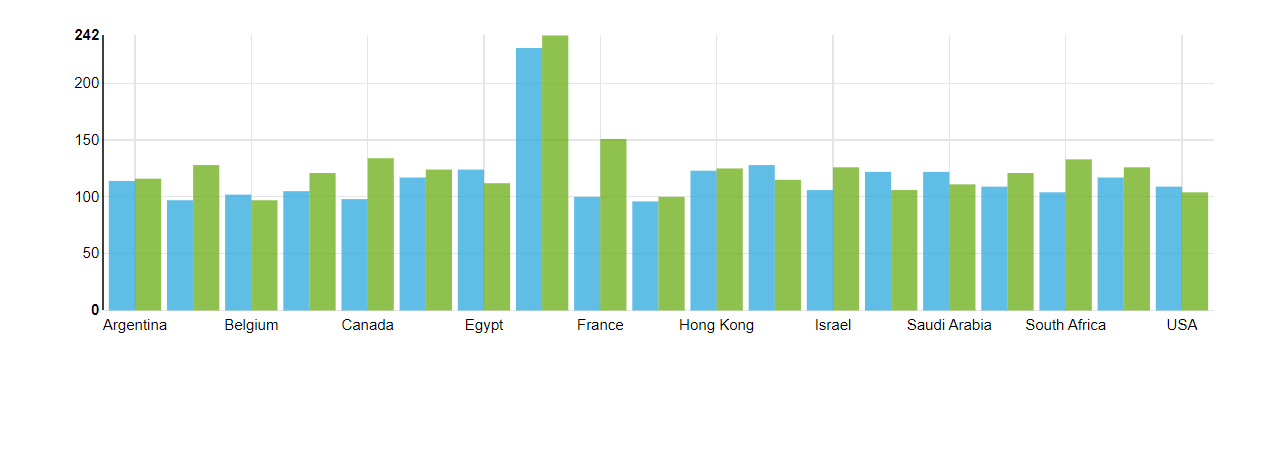
count(temp\_range) Office\_Count

from hvac\_building

where temp\_range != 'NORMAL'

group by country, temp\_range;





**-- Q3. 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**

**-- extreamtemp by hvacproduct**

select

hvac\_product,

sum(cast(extreme\_temp as int)) as extreme\_temperature

from hvac\_building

group by hvac\_product;



