**Task:**

1- Come up with a list of questions that Isaac would like to answer using the smart home dataset. These questions should be related to activity recognition, anomaly detection, and privacy. No need for formal research questions. The objective of this task is to see how he can structure of formalizing an analytic data problem

2- Exploring the smart home datasets using data visualization techniques. Mainly using plotting tools like Matplotlib and graph visualization tool like NetworkX. Using data visualization, he should be able to visualize the dataset to show patterns, data characteristics, and provide a visual summarization of the data.

**MIT Activity Dataset : Activity Recognition in the Home Setting Using Simple and Ubiquitous Sensors**Link: http://courses.media.mit.edu/2004fall/mas622j/04.projects/home/

**Summary**  
This dataset is about human activities in smart home. Two different smart homes are used to collect data in two weeks. The size of the dataset in both home is about 171.9 KB and the data format is in csv. There are 161 sensors in two different homes. More than one sensors are attached to one subject. The types of IoT devices are electric appliances, furniture and things you can find in a home such as light switch, door, oven, jewellery box etc. Importantly, the data format, features and the type of the features of the dataset are as follows :   
  
Data Format and Type : SensorID (Integer), Location (String), Object (String)

The type of features : SensorID (Ordinal), Location (Nominal), Object (Nominal)  
  
Data Format and Type : Heading(String), Category(String), Subcategory(String), Code(Integer)  
The type of features : Heading(Nominal), Category(Nominal), Subcategory(Nominal), Code(Ordinal)  
  
Assume one sensor is detected for one activity (Could be more)  
Data Format and Type : ACTIVITY\_LABEL(String), DATE(TimeStamp), START\_TIME(TimeStamp), END\_TIME(TimeStamp)   
The type of features : ACTIVITY\_LABEL(Nominal), DATE(Continuous), START\_TIME(Continuous), END\_TIME(Continuous)  
  
Data Format and Type : SENSOR1\_ID(Integer)  
The type of features : SENSOR1\_ID(Ordinal)  
  
Data Format and Type : SENSOR1\_OBJECT(String)  
The type of features : SENSOR1\_OBJECT(Nominal)  
  
Data Format and Type : SENSOR1\_ACTIVATION\_TIME(TimeStamp)  
The type of features : SENSOR1\_ACTIVATION\_TIME(Continuous)  
  
Data Format and Type : SENSOR1\_DEACTIVATION\_TIME(TimeStamp)  
The type of features : SENSOR1\_DEACTIVATION\_TIME(Continuous)  
  
**Number of Human subjects**2: The first subject was a professional 30-year-old woman who spent free time at home, and the second was an 80-year-old woman who spent most of her time at home.  
  
**Questions related to activity recognition, anomaly detection, and privacy**

How could anomaly activities be recognized in this dataset? (eg. outlier on a 2D/3D graph?)  
What are the patterns that can be studied to predict what the subject is going to next?   
In terms of privacy, what are the time that subjects not at home?  
How many subjects are in the home by checking the activation of sensors?  
Are there any mistakes when recognizing activities? (eg. activity is labelled as bathing but freezer is used during the time)  
Could the data tell persons' gender and approximate age?(eg. jewellery box is used)

**ARAS Datasets**

link: https://www.cmpe.boun.edu.tr/aras/  
  
**Summary**  
This dataset is about collecting real world activities from real people in real homes. Two houses are used to collect data for a full month. The size of the dataset in the lab is about 225 MB and the data format is in txt. There are 20 sensors in each house. Each sensor is located in different places such as wardrobe, TV receiver, chair and couch etc. The types of IoT devices are the sensors. Importantly, the data format, features and the type of the features of the dataset are as follows :  
  
Data Format :   
  
Sensor1(Integer), Sensor2(Integer), Sensor3(Integer), Sensor4(Integer), Sensor5(Integer), Sensor6(Integer), Sensor7(Integer), Sensor8(Integer), Sensor9(Integer), Sensor10(Integer), Sensor11(Integer), Sensor12(Integer), Sensor13(Integer), Sensor14(Integer), Sensor15(Integer), Sensor16(Integer), Sensor17(Integer), Sensor18(Integer), Sensor19(Integer), Sensor20(Integer), ActivityLabel\_Resident1(Integer), ActivityLabel\_Resident2(Integer)  
  
Type :  
  
Sensor1(Nominal), Sensor2(Nominal), Sensor3(Nominal), Sensor4(Nominal), Sensor5(Nominal), Sensor6(Nominal), Sensor7(Nominal), Sensor8(Nominal), Sensor9(Nominal), Sensor10(Nominal), Sensor11(Nominal), Sensor12(Nominal), Sensor13(Nominal), Sensor14(Nominal), Sensor15(Nominal), Sensor16(Nominal), Sensor17(Nominal), Sensor18(Nominal), Sensor19(Nominal), Sensor20(Nominal), ActivityLabel\_Resident1 (Nominal), ActivityLabel\_Resident2 (Nominal) **Number of Human subjects**

2: Resident 1 and Resident 2

**Questions related to activity recognition, anomaly detection, and privacy**  
  
How to know if the residents are sleeping or they are just lying on the bed?  
How can we tell the activities are performed by resident 1 or resident 2?  
What are the sequential activities of resident 1 and resident 2 (eg. Sleeping -> shaving -> toileting -> Changing clothes)?  
Could concurrent activities not be recognized because the data only showed one activity per resident?  
What is regular time which resident 1 and resident 2 are both not home?

**Tim van Kasteren activity recognition dataset**link: http://ailab.eecs.wsu.edu/casas/datasets.html  
  
**Summary**:   
This dataset is about collecting data for activity recognition. Two houses are used to collect data in two months. The size of the dataset in the lab is about 158 KB and the data format is in m(matlab) or txt. There are 24 sensors. Each sensor is also located in different places such as microwave, front door, Freezer and washing machine etc. The types of IoT devices are the sensors. Importantly, the data format, features and the type of the features of the dataset are as follows :  
  
Data Format and Type : Start Time (TimeStamp), End Time (TimeStamp), SensorID (Integer), Value (Integer)

The type of features : Start Time (Continuous), End Time (Continuous), SensorID (Ordinal), Value (Ordinal)  
  
Data Format and Type : StartTime(TimeStamp), End Time(TimeStamp), ActivityID(Integer)  
The type of features : StartTime(Continuous), End Time(Continuous), ActivityID(Ordinal)

**Number of Human subjects**Unknown because link to the page is not working.

**Questions related to activity recognition, anomaly detection, and privacy**

What are the sensors that are not activated frequently? (Anomaly activities)  
How to know if the residents are sleeping or they are just lying on the bed?  
What are the regular days that subjects are not home?  
What are the sequential activities in this data set and why is it dangerous to know? (break in to the house when resident is not home)  
 **Health Smart Home (HIS) datasets**

link: https://sites.google.com/site/thonyfleury/health-smart-home-his-datasets  
  
**Summary**   
This dataset is about collecting smart home data in TIMC-IMAG lab. The home is equipped and contains all the materials that can be used for different activities. The size of the dataset in both home is large because it includes video, audio and text files . In the dataset, the types of IoT devices are microphones, presence infra-red sensors, and webcams. Also, temperature and hygrometry sensors are installed inside the bathroom, and the dweller is equipped with an home-made accelerometer and magnetometer circuit that is placed under the left armpit. Importantly, the data format, features and the type of the features of the dataset are as follows :

**Number of Human subjects**10 different subjects  
  
  
  
**Questions related to activity recognition, anomaly detection, and privacy**