Problem Description

An architecture company has recently completed an initiative to digitalize their building layout designs. As part of this exercise, they have converted the plan view of buildings architected by them into bitmaps of size 640 x 480 pixels, as shown by the sample images below.

A repository of 1183 such views (i.e. buildings) has been created in the first phase. These images are available in the file **layouts.zip** which will be sent.

They are now ready to launch initiatives to use this data to improve their productivity, and facilitate robust designs. They are also very keen to mine the information potentially hidden in these views, and establish new and innovative processes to respond rapidly to customer requirements and demands. They are convinced that methods of Data Science in general, and Machine Learning in particular, can be effectively used to achieve these goals. Following are some of the requirements / questions they have posed:

1. *We have a hunch that our designs can be grouped into families, based on their shapes. Knowledge of these families will not only improve our insights into our own designs, but also help in standardizing them by creating design templates. We have been informed that multiple approaches can be used to do this, and we would like to see the results of at least two approaches.*
2. *Further, we would like to classify the complexity of layouts as **Low Complexity**, **Medium Complexity**, **High Complex**. Based on a formal analysis of the layouts we would like to create the criteria to decide their complexity. Can you carry out an appropriate analysis, establish the criteria, and classify the layouts?*
3. *We are very keen to speed up our layout design process by retrieving relevant prior layouts based on a set of gross parameters. For example, the architect usually knows the dimensions (length, width) of the tight-fitting box (see Appendix 1), the layout area, and the permissible layout complexity. When the architect specifies these parameters, the **design family / families** likely to provide the closest designs should be predicted. Layouts from such families can be retrieved for further detailing. This will save a lot of time and effort, and ensure that designs are based on past, successful designs!*
4. *We are open to other suggestions, about what else can be done using the image data and information mined from it. Please suggest more possibilities and ideas*

HINTS:

APPENDIX – 1: Tight Fitting Bounding Box

A tight-fitting bounding box around a contour is the one with minimum area. Such a box is usually not aligned with the horizontal and vertical axes of the image containing the contour – as shown in the image below.

- Outer blue box: Total extents of the image
- Inner dark blue polygon: Contour / layout under consideration
- Green box: The tight-fitting box covering the contour under consideration