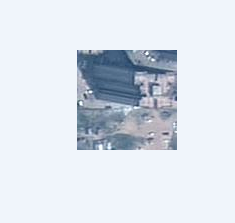
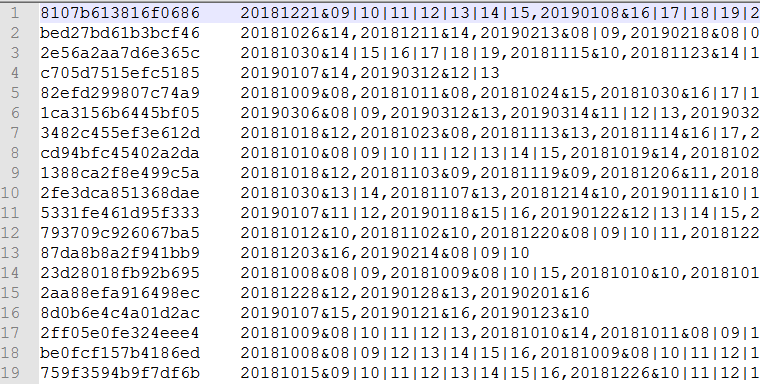
# Urban Region Function Classification

Website: <https://dianshi.baidu.com/competition/30/question>

## The goal:

Class the image (100\*100 pixels)/User visit time data to 9 classes.

|  |  |
| --- | --- |
| CategoryID | Functions of Areas |
| 001 | Residential area |
| 002 | School |
| 003 | Industrial park |
| 004 | Railway station |
| 005 | Airport |
| 006 | Park |
| 007 | Shopping area |
| 008 | Administrative district |
| 009 | Hospital |

### Initial analyses:

1. Neural network must be used (huge data, no hard rule to classify them).
2. Images can only offer limited information (textures are very similar, such as residential area, school, shopping area and others). The User visits information is very important.

### Things need to do:

1. How to extract the user visit information?
2. How to combine the image and visit information together?
3. Data augmentation?

## 5/4/19

Data preparing.

1. Extract the visit information.
2. Total visitor (the number of records) (170000 -1, how to normalize it?)
3. 24/12 hours, the number of visits.
4. ~~12 months, the number of visits. (September,2018 ----February, 2019, no enough data)~~
5. The return times of each visitor.
6. The average stay time of every day.
7. The size of data of the record.(9MB-1k)

Normalization: assume all data are following the normal distributing, then estimate the parameters, then normalize them to 0-1.