UrbanSim Data Preparation Scripts

Hana Ševčíková

November 22, 2010

1 Unrolling Jobs from Establishments

1.1 Prerequisities

- Table of **businesses** that are assigned either to zones, or optionally to buildings or parcels.

  Mandatory columns: `zone_id`, `sector_id`, number of jobs (name configurable), building sqft (name configurable).

  Optional columns: `parcel_id`, `building_id`, `home_based` - maps to job's attribute `building_type` (should map to `home_based_status`), `join_flag`, `impute_building_sqft_flag`.

- Optionally, table of **zonal job control totals** (name configurable).

1.2 Procedure

1. User's settings:

   - `compute_sqft_per_job`: True or False
   - `unplace_jobs_with_non_existing_buildings`: True or False
   - `minimum_sqft` and `maximum_sqft`: numeric values. They default to 1 and 4000, respectively.

2. Initialize a job dataset with columns

   job_id
   sector_id
   building_id
   parcel_id
   zone_id

> **building_type** (corresponds to business attribute `home_based`)
>
> **sqft**
>
> **join_flag** (if `businesses.join_flag` exists)
>
> **impute_building_sqft_flag** (if `businesses.impute_building_sqft_flag` exists)

3. For each job in each business create an entry in the job dataset. Attribute `job_id` is an enumeration of the jobs.

4. If `compute_sqft_per_job` is True, the attribute `sqft` is computed (using 10% vacancy) as (building sqft - 0.1·building sqft)/number of jobs. The result is clipped between `minimum_sqft` and `maximum_sqft`. If `compute_sqft_per_job` is False, the `sqft` attribute is set to the building sqft of the business.

5. All missing values of `building_type` (i.e. smaller than 0) are set to 'non-home-based'.

6. If `unplace_jobs_with_non_existing_buildings` is True, all jobs that are assigned to a non-existing building are unplaced.

7. If a table of job control totals is not given go to 10., otherwise continue.

8. The table of control totals is expected to have columns `zone_id`, `sector_id`, and `jobs`.

9. For each zone and sector, check the actual number of jobs and compare it to the control total (CT). If there are less jobs than CT, continue. Otherwise compute the number of jobs to be deleted. Delete jobs randomly, starting from jobs that have neither `parcel_id` nor `building_id` assigned. If more jobs need to be deleted, continue by deleting jobs that have no `building_id` assigned, and only after that consider all jobs as candidate for being deleted.

10. Write the resulting table of jobs into cache (or other user-defined storage).

11. Table **building_sqft_per_job** is created by taking the median of the job `sqft` attribute over buildings of each building type and in each zone, clipped between min and max (defaults are 25 and 2000, respectively). Only jobs are considered that have `sqft` larger than 0.

2 Handling Residential Locations

1. In each zone, check the number of households and create new buildings if necessary to accommodate all households.

2. Run the household location choice model for each zone and assign buildings to households within their zone.

3 Handling Non-Residential Locations

3.1 Prerequisities

- Table of **jobs** (e.g. result of procedure described in Section 1).

  Mandatory columns: `job_id`, `building_id`, `parcel_id`, `building_type` (1=home based, 2=non-home based, 3=governmental), `sqft`

  Optional columns: `impute_building_sqft_flag`

  The table is assumed to have `building_id` assigned for parcels that have only one building.

- Table of **buildings** with columns `building_id`, `parcel_id`, `building_type_id`, `non_residential_sqft`

- Table of **building types**.

- Table **building_sqft_per_job**.

3.2 Procedure of placing jobs with known parcel id

1. Select all jobs that have `parcel_id` assigned but no `building_id` and group them by `building_type` into a governmental, non-home based and home-based group.

2. Place governmental jobs: For each parcel that have governmental jobs, sample randomly within governmental buildings in that parcel. Leave jobs unplaced if there are no governmental buildings.

3. Move jobs from the governmental group that were not placed in the step above into the non-home based group.

4. Determine the joint distribution of jobs over sectors and building types.

5. Place jobs from the non-home based group into buildings, depending on the available space, as follows.
   For each parcel that contains at least one job of this group:

   (a) Locate all buildings in this parcel. If there are no buildings go to the next parcel.
   (b) Determine the sqft capacity of each such building (i.e. `non_residential_sqft` minus occupied sqft). If there is no available capacity go to the next parcel.
   (c) Create a demand table with rows corresponding to buildings and columns corresponding to jobs which contains the values of jobs `sqft`. Zero values are replaced by corresponding values from the building_sqft_per_job table.

(d) Compute a table of supply-demand ratio where each cell is equal to the building's capacity divided by demand multiplied by the number of jobs. This value is multiplied by 0.9 in order to assure 10% vacancy. Thus each cell in this table contains a factor with the meaning 'If all jobs in this parcel would have this size, by how much would the sqft need to be decreased/increased to fit into this building, leaving 10% vacancy'.

(e) For buildings that have supply-demand ratio smaller than one in at least one cell, multiply the demand table by the supply-demand ratio.

(f) Using the job's distribution table from Step 4., create a probability table of the same shape as the demand table.

(g) Sort jobs by their sqft in decreasing order and sample a building to each job using the distribution table as sampling weights.

(h) Update job's `sqft` attribute by the corresponding value from the demand table, clipped between minimum and maximum set in Section 1.

6. Consider all non-home based jobs that were not placed in the previous step (excluding governmental jobs) and check if their parcel contains at least one residential building. In such a case, re-classify such a job to a home-based job.

7. For each non-home based job that was neither placed in Step 5. nor re-classify in Step 6., and has `impute_building_sqft_flag` equal to 1, sample a building using the distribution table from Step 4. and impute the corresponding sqft into the building. Update the attribute `non_residential_sqft` of such buildings and update the job's `sqft` attribute for such jobs by the corresponding value from a demand table created as in Step 5(c).

8. Place jobs from the home based group into buildings, as follows.
   For each parcel that contains at least one job of this group:

   (a) Locate all buildings in this parcel.

   (b) Compute the capacity for home-based jobs as total job space minus the number of home-based jobs. The total job space for a building is computed as follows: For a single family building, take the minimum of the number of persons and 2, sum-up over households in the building; For a multi-family building, take the minimum of the number of residential units and 50. If there is no available capacity in this parcel go to the next parcel.

   (c) Randomly sample residential buildings to jobs, taking into account the capacity.

9. Assign building type 1 (home-based) or 2 (non-home based) to jobs with missing values of the building type by sampling randomly with weights proportional to the actual home-based/non-home based distribution.

10. Store the updated table of jobs and buildings into cache (or other user-defined storage).

11. Update the building_sqft_per_job table.

3.3   Procedure of placing remaining jobs

1. Run the expected sales price model.

2. Run the development proposal choice model for each zone and generate proposals to accommodate all jobs.

3. Run the building construction model.

4. Run the employment location choice model for each zone and assign buildings to reamaining jobs within their zone.