

## Homework #2 – Fill Missing Fields

- **FULLVAL, AVLAND and AVTOT**

- The variables FULLVAL, AVLAND and AVTOT denote the market value of the property. They can be determined by the variables Tax Class (TAXCLASS) and the Zip code.
- First, group by Tax class, and then by Zip code, and take the median for filling empty values.
- Unreasonable values: Replace 0's in this variable by the same method.
- If there are fewer than 5 values in each group, group by a broader category (Just Tax Class).

- **LTFRONT and LTDEPTH**

- The variables LTFRONT and LTDEPTH can be determined by the variables - Borough (B) and Tax Class (TAXCLASS) and Block, since the lot front length/breadth varies for properties across, for example, Manhattan and Queens.
- First, group by Borough and then Tax class, and take the median for filling empty values.
- Unreasonable values: Replace 0's in this variable by the same method.
- If there are fewer than 5 values in each group, group by a broader category (Just Borough).

- **BLDFRONT and BLDDEPTH**

- The variables BLDFRONT and BLDDEPTH can be determined by the variables - Borough (B) and Tax Class (TAXCLASS) and Block, since the lot depth/breadth varies across, for example, Manhattan and Queens.
- First, group by the Borough and then the Tax class, and take the median for filling empty values.
- If there are fewer than 5 values in each group, group by a broader category (Just Borough).

- **ZIP**

- The Zip value of the property can be determined grouping by the variables Borough and Block, and then taking the mode.
- If there are very few values in the group, a wider category (Borough) can be used.
- If the zip codes are not of length 5 or 9, then they are unusual values and can be replaced by the above method.

- **STORIES**

- The number of stories in a property can be determined by grouping by Borough, Tax class and Block, and taking the mode.
- If there are very few values in a group, take a broader category (Borough and Tax Class).