# Shared USE Path T-4 project dictionary

This folder contains all the code for the shared use path task-4 project. There are three sub-heading in this report which are described in detailed, the subheadings are:

1. Important files and folders
2. Data preparation code
3. Analysis code
4. Site selection tool

## Important files and folder

The “SUP” folder contains all the files and codes used for the Task-4 of the shared used path project. Inside the folder, a number of jupyter notebooks are present which is described in the next section. The “Presentation” folder, “Report” folder contains files as suggested by their name. The “Data” folder contains all the data used for the preparation of the final dataset. The “Site Selection Tool” contains information on the site selection tool preparation.

Inside the “Data” folder, the “Temp” folder contains all the csv, shape, and excel files prepared during the preparation of the dataset. Different files are saved during the calculation process which is used to create the final dataset. The excel file named “FullSUPIdentifiedData” inside the Data folder contains all information manually gathered related to SUP location and other features in the initial stage of the project. As of making of this report “FULL\_prepared\_data\_ampm\_v11.csv” is the final dataset after joining all the prepared data.

Another folder inside the “Data” folder is “Landuse” which contains the landuse data. The specific location of all land use data used for calculation is Data>Landuse>All landuse. The name of the county landuse data folders are coded, if the last word of the folder has “County” then the data is downloaded from TNRIS and if the last word of the folder has “C” then it is downloaded from tax appraisal office. If the last word of the folder is neither “C” nor “County” then it is downloaded from open city data.

## Data preparation code

This folder contains all the codes and data used to prepare the final dataset. Each Jupiter notebook containing the codes are explained in order of the process conducted while making the dataset. When working with spatial data, the maps are projected to NAD83/Texas Central (ft US). The European Petroleum Survey Group (EPSG) code for the projection is 2277.

### Count and ACS

1. data\_preparation\_count\_acs  
   This contains the code used to process the raw data obtained from Dr. Phil named “validation\_counts as of 2022122122558.csv” and exports a file named “raw\_data\_” inside Data/Temp which contains hourly count data. Count wrt day of week is also exported from here. This also calculates ACS data for each location and saves a file named FULL\_station\_and\_demographic\_025mile.csv.

### Roadway

1. data\_preparation\_roadway  
   This jupyter notebook contains codes used to calculate the roadway density, ADT, previous year ADT and max speed. All the lengths and area of buffers are in feet and sq. feet.

### Land use

1. data\_preparation\_landuse  
   After the local land use data is collected and prepared, the land use geospatial files are imported and joined. The joined file is saved as “Texas\_landuse\_raw.shp”. The land use code form this raw file is then converted into ‘residential’, ‘commercial’, ‘industrial’ and ‘vacant’ using a function named “replacelu” in the code. A comment “#s” which means self-assigned. Some codes for Whichita county were manually checked. “#g” was commented on land use codes which were checked using google maps and #g\_x was assigned to land use codes which were checked using google maps and the land use code used was not correct.  
   After converting all the land use codes, the rear of each type and the entropy is calculated for each station.  
   In addition, OSM data for Dallas is also checked for land use data and OSM land use visualization is prepared in this jupyter notebook.

### Replica land use

1. data\_preparation\_replicalanduse:  
   replica data is imported and the land use types around each count location for different buffer sizes are calculated. The land use codes are then converted into ‘residential’, ‘commercial’, ‘industrial’ and ‘vacant’. Entropy is also calculated for each count location.

### OSM intersection and Transit

1. data\_preparation\_osmintersection\_transit  
   The code contains, the calculation of number of intersections using OSM. It also has codes which was used extract SLD data for locations but not used due to too many missing values. The code also contains, transit related variable calculations.

### Job accessibility

1. data\_preparation\_access\_networkx  
   The code contains the job accessibility calculation. The codes are well commented. The travel speed is changed manually for walking and cycling, and the code is run separately for both. LODES data is used to calculate the number of jobs around each location.

### Weather and OSM proximity

1. data\_preparation\_weather\_osmproximity  
   Weather data is prepared using the different technique. First weather stations nearby each location are found out by calculating the distance between count station and weather station. Only stations within 10KM distance are selected for getting weather related data. If a station has missing value, then next nearest station is used to get the missing data. All the weather variables are joined to station for different years and saved.

### Joining all data

1. data\_preparation\_joining\_all  
   This Jupyter notebook joins all the calculated data into the final dataset. The first part of the dataset is “Combining all” which combines all data (except weather) to a final dataset. Then Features from BPCX is added to the dataset (like count method, surface type, streetlight,). After that the data set is formatted for modelling, which consists of extracting year information from columns and adding year data. Weather data is added after the dataset is modified and includes a separate column named “year”.   
   ADT variable is treated (all network type ADT is added within each buffer size to have a single ADT variable for each buffer sizes. AM peak and PM peak count data is calculated and added to the dataset. Missing data from weather variable is calculated. Finally, a new variable named ‘city\_area’ is added.

## Analysis code

There are two jupyter notebook which contains the preliminary analysis. The first one is “preliminary\_analysis\_buffer” which contains the buffer analysis and the second one is “premilinary\_analysis\_exploratory\_impfeatures” which contains the exploratory analysis, descriptive statistics table preparation code and Texas map with SUP counts visualization.

## Site selection tool

The folder “Site Selection Tool” inside “SUP” folder contains the information related to site selection tool. Excel file “AHP” contains the AHP weight calculation. The method used to create the site selection tool is mentioned below in detail. The following describes creation of weighted raster.

1. Create Proximity Maps:

a. Open the 'Geoprocessing' pane by clicking 'Analysis' > 'Tools' on the main toolbar.

b. In the 'Geoprocessing' pane, search for 'Euclidean Distance' and select it from the list.

c. For each input shapefile (parks, water body, university), follow these steps:

i. Choose the shapefile as the 'Input raster or feature source data'.

ii. Set an appropriate 'Output distance raster' name and location.

iii. Set a 'Maximum distance' if desired, or leave it blank to compute the entire extent.

iv. Click 'Run' to create a distance raster for each input shapefile.

1. Normalize the Rasters:

a. In the 'Geoprocessing' pane, search for 'Rescale by Function' and select it from the list.

b. For each distance raster created in the previous step, follow these steps:

i. Choose the distance raster as the 'Input raster'.

ii. Set an appropriate 'Output raster' name and location.

iii. Choose 'Linear' as the 'Transformation function'.

iv. Set the 'Output range min' and 'Output range max' values (e.g., 0 and 1).

v. Click 'Run' to normalize each distance raster.

1. Assign Weights using MCDM:

a. Determine the weights for each layer based on AHP.

i. For AHP mehod, make rows and columns with the variables in raster

ii. Create comparision matrix. Assign pairwise values from 1 to 9

iii. Beginning from first row, assign number according to importance. for example a value of 3 for A/B is A is 3 times imp than B and B/A will automatically be 1/3.

iv. Check consistency: Finally, you need to check the consistency of the pairwise comparisons by calculating the consistency ratio (CR). The CR is calculated by dividing the consistency index (CI) by the random index (RI). The CI is a measure of how consistent the pairwise comparisons are, and the RI is a value derived from the size of the matrix. If the CR is less than 0.1, then the pairwise comparisons are considered consistent.

b. In the 'Geoprocessing' pane, search for 'Raster Calculator' and select it from the list.

c. Create a weighted sum expression using the weights obtained in the previous step (e.g., "Weight1 \* Normalized\_Parks + Weight2 \* Normalized\_Water\_Bodies + Weight3 \* Normalized\_Universities").

d. Set an appropriate 'Output raster' name and location.

e. Click 'Run' to create the weighted raster.