This document describes the outline for the homogeneous multiplex generator from given comma separated format of a dataset.

INPUT 1 (Data File): Structure of the Dataset with N records and K features

- The data file should be in a csv (comma separated) format
- **Line 1:** The first line of the file must specify the **feature names**
  - **Format: <featureName$_1$, featureName$_2$, …, featureName$_K$>**
- **Line 2:** The second line specifies the **feature type**
  - **Format: <featureType$_1$, featureType$_2$, …, featureType$_K$>**
  - What are the different types of features that must be handled?
    - **Identifier/Primary Key**
      - The unique identifier corresponding to the record
      - Example – *University ID*: asd3421
    - **Nominal/Categorical**
      - The attributes that have a fixed domain to which the values can belong
      - Example – *Political Belief of a person*: Democratic
    - **Numeric**
      - It can be integer or floating point
      - Example – *Speed of a Vehicle in mph*: 75
    - **Geographical Location**
      - **Latitude**
      - **Longitude**
    - **Time**
      - Format: 24 HRS
      - Example – *Time of Accident*: 1330
    - **Date**
      - Format: MM-DD-YYYY
      - Example – *Date of Movie Release*: 09-19-2018
    - **Set**
      - Format for a set with m values: {value$_1$, value$_2$, … value$_m$}
      - Example – *Genre of a Movie*: {ACTION, COMEDY, DRAMA}
    - **Text**
      - Format: [<text>]
      - Example1 - *Tweet*: [Congratulations, France #FIFA2018 #Winners]
      - Example2 – *IMDB Review*: [The movie was crap]
- **(N lines) Line 3 to Line N+2:** Each line corresponds to the comma separated data record

INPUT 2 (Configuration File)

- **Line 1: # Dataset Path**
- **Line 2:** The path to the dataset must be specified here
- **Line 3: # Number of Nodes**
- **Line 4:** Specify the number of nodes per layer i.e. the number of records in the file

- **Line 5: # Number of Features**
- **Line 6:** Specify the number of features present in the dataset
- **Line 7: # Number of Layers**
- **Line 8:** Specify the number of layers that need to be generated for this layer, say L
- **(L lines) Line 9 to Line L+8:** Each line from here specifies the configuration for the layer to be generated. The configuration will depend on feature type and similarity metric. If a layer has to be generated for the i^{th} *feature,* then depending on the feature type the configuration needs to be provided in one of the following formats
    - **Nominal**
        - **NOMINAL,**<*i*>**,METRIC_EQUALITY**
    - **Numeric**
        - **NUMERIC,**<*i*>**,**<measurement unit of feature *i*>**,METRIC_THRESHOLD,EUCLIDEAN,**<threshold value>
        - **NUMERIC,**<*i*>**,**<measurement unit of feature *i*>**,METRIC_FIXEDRANGE,**<number of ranges, say m>**,[**<r_1>,<r_2>**],[**<r_2>,<r_3>**], …,[**<r_{m-1}>,<r_m>**]**
    - **Geographical Location**
        - **LOCATION,**<latitude feature ID, say *i*>,<longitude feature ID, say *j*>,<**MILES**/**KM**>, **METRIC_THRESHOLD,**<**HAVERSINE**/**EUCLIDEAN**>**,**<threshold value>
        - **LOCATION,**<latitude feature ID, say *i*>,<longitude feature ID, say *j*>,<**MILES**/**KM**>,**METRIC_FIXEDRANGE,**<number of ranges, say m>**,[**<r_1>,<r_2>**],[**<r_2>,<r_3>**], …,[**<r_{m-1}>,<r_m>**]**
    - **Time**
        - **TIME,<***i*>,**METRIC_THRESHOLD,EUCLIDEAN,**<threshold value>
        - **TIME,<***i*>,**METRIC_FIXEDRANGE,**<number of ranges, say m>**,[**<r_1>,<r_2>**],[**<r_2>,<r_3>**], …,[**<r_{m-1}>,<r_m>**]**
        - **TIME,<***i*>,**METRIC_THRESHOLD_SEGMENT,EUCLIDEAN,**<number of equi-sized segments>,<threshold value>
        - **TIME,<***i*>,**METRIC_FIXEDRANGE_SEGMENT,** <number of equi-sized segments>,<number of ranges, say m>**,[**<r_1>,<r_2>**],[**<r_2>,<r_3>**], …,[**<r_{m-1}>,<r_m>**]**
    - **Date**
        - **DATE,**<*i*>,**METRIC_THRESHOLD,EUCLIDEAN,**<**DAY/WEEK/MONTH/YEAR**>,<threshold value>
        - **DATE,**<*i*>,**METRIC_EQUALITY,**<**DAY/WEEK/MONTH/YEAR**>
    - **Set**
        - **SET,**<*i*>**,METRIC_THRESHOLD,**<**COSINE/JACCARD**>,<threshold value>
    - **Text**
        - **TEXT,**<*i*>**, METRIC_THRESHOLD,**<**COSINE/JACCARD**>,<threshold value>