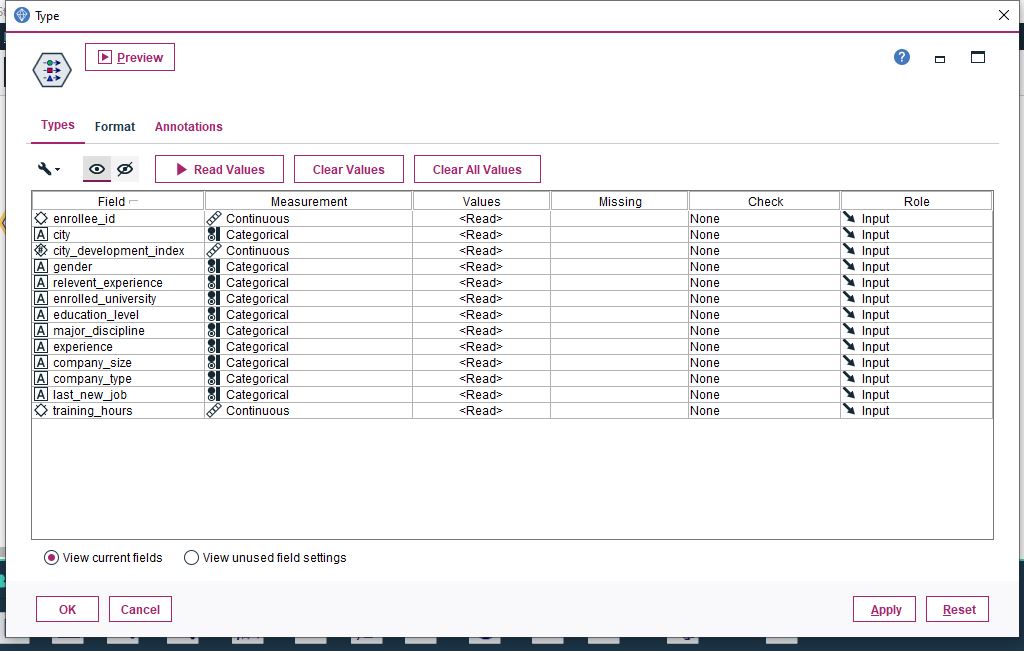
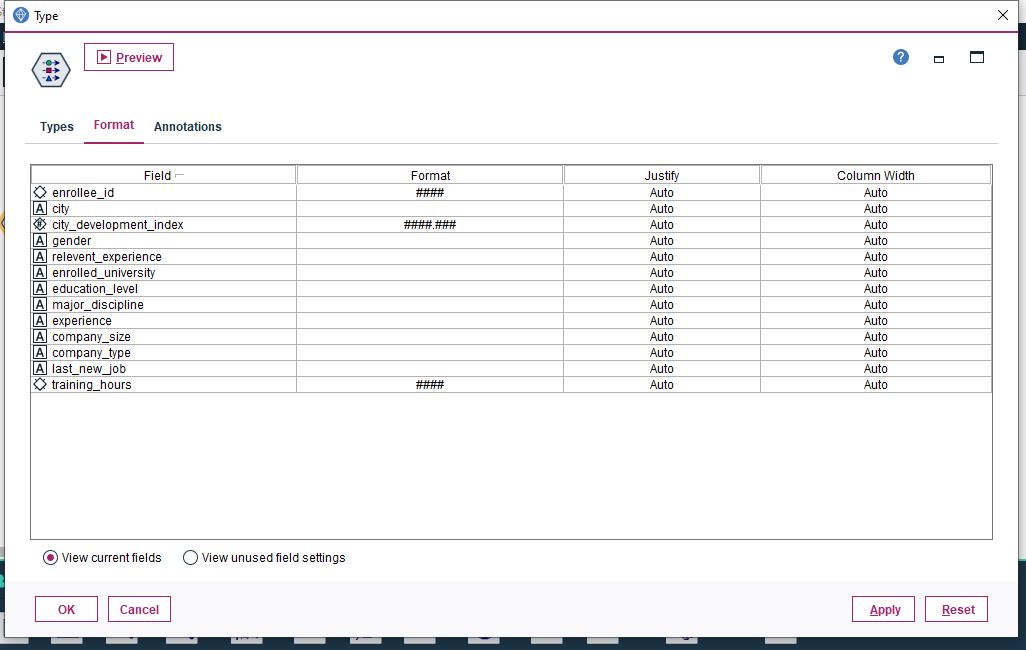
**PM**

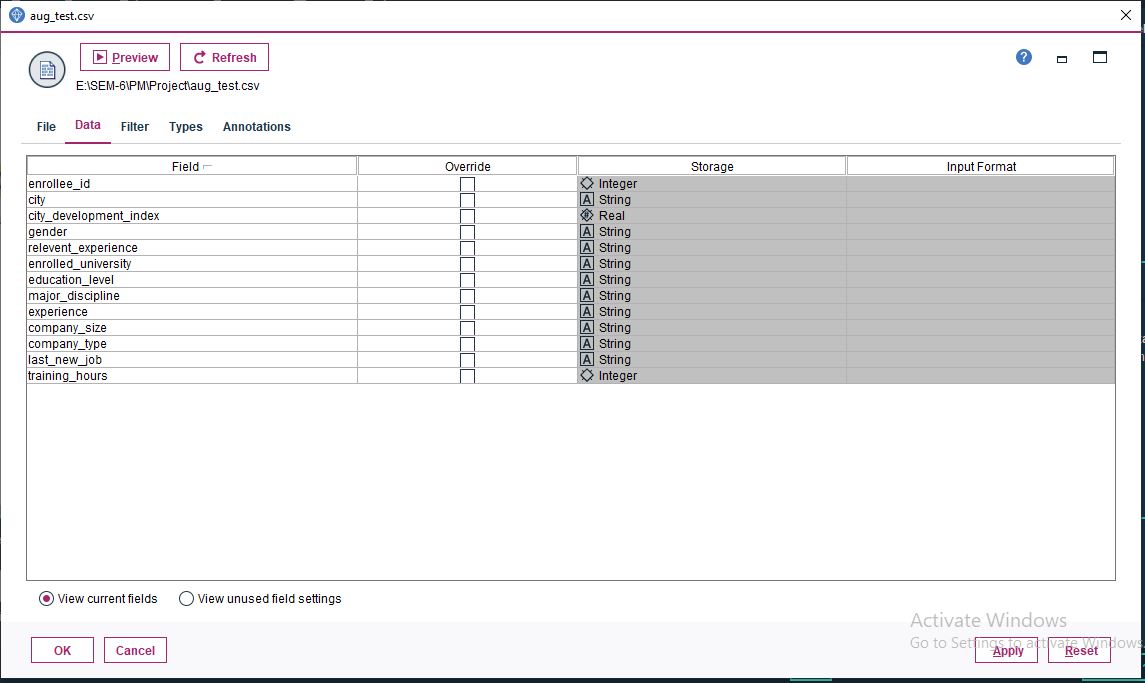
**Project Review-1**

**Data Description Report:-**

1. **Aug\_train.csv:-** This file contains 14 fields.

****

****

****

1. enrolle\_id:- unique id for candidates

* Measurement is continuous as it has continues values for id’s.
* Data type is integer as it accepts integer values for unique ids.
* Format = #### i.e. it accepts max 4 digit int value.
* Role = Input means this field contributes in modelling.

1. City:- this contains city code for each city

* Measurement is categorical as it has diff categories (ex-city\_41, city\_13 etc).
* Data type is string.
* Format = None as it accepts string.
* Role = Input i.e. it contributes to the model we will be creating.

1. City\_development\_index:- This indicates the development index of the city which is scaled

* Measurement is continuous indicating it has continuous range of index (ex-0.827,0.92 etc)..
* Data Type is Real means it accepts real numbers.
* Format = ####.### i.e. max 4 digit before decimal and max 3 digit after decimal.
* Role = Input i.e. it plays a role in modelling

1. Gender:- tells about gender of the candidate.

* Measurement level is categorical indicating it has diff categories (ex-Female,Male etc)
* Data type is string
* Format = None as the data type is string
* Role = input means it acts as an input in the modelling

1. Relevant\_experience :- tells about relevant experience of the candidate.

* Measurement level is categorical (ex-has relevant experience etc)
* Data type is String
* Format = None as data type is string
* Role = input

1. Enrolled\_university:- type of university course enrolled if any.

* Measurement is categorical (ex-Full time course, no\_enrollment etc)
* Data type is String
* Format = None as String is the data type
* Role = Input

1. Education\_level :- tells education level of candidate

* Measurement is categorical (ex- Graduate, High School etc)
* Data type is String
* Format = None as data type is string
* Role = input

1. Major\_disciplin :- indicates education major of candidate

* Measurement is categorical (ex-STEM, Other)
* Data type is String
* Format = None as data type is string
* Role = input

1. Experience :- tells experience of candidate in years

* Measurement is categorical (ex-9,<1 etc)
* Data type is String
* Format = None as data type is string
* Role = input

1. Company\_size :- tells no. of employees in current in employee’s company

* Measurement is categorical (ex-50-99 etc)
* Data type is String
* Format = None as data type is string
* Role = input

1. Company\_type:- tells type of company of employee

* Measurement is categorical (ex- Pvt Ltd)
* Data type is String
* Format = None as data type is string
* Role = input

1. Last\_new\_job:- no. of years between previous job and current job

* Measurement is categorical (ex- 1, never)
* Data type is String
* Format = None as data type is string
* Role = input

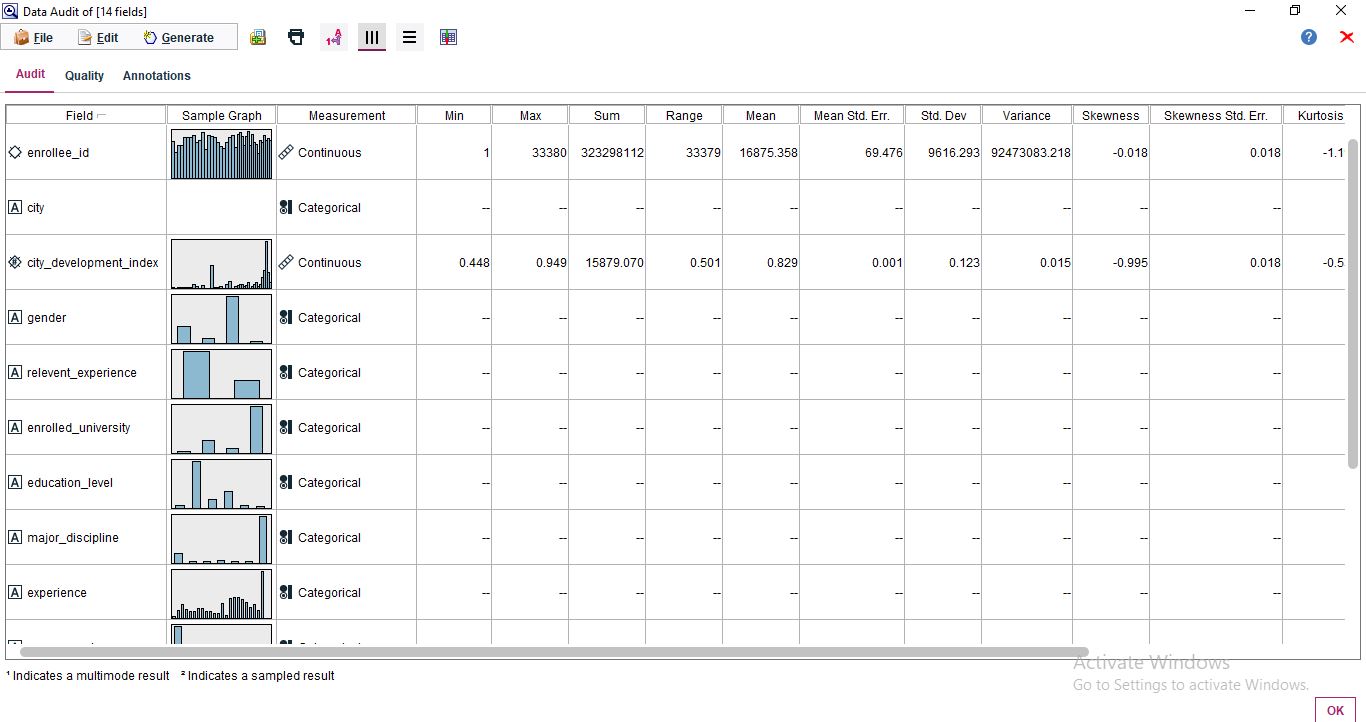
1. Training\_hours:- indicates training hours completed

* Measurement is continuous (ex- 21,10 etc)
* Data type is integer
* Format = #### i.e. max 4 digit int
* Role=input

1. Target:- indicates whether candidate will change job or not

* Measurement is continuous
* Data type is real
* Format = ####.### i.e. max 4 digit before decimal and max 3 digit after decimal.
* Role=input

**Data Exploration Report:-**

****

1. **Aug\_train.csv:-** This file contains 14 fields.
2. enrolle\_id:

* min value=1
* max value=33380
* sum= 323298112
* Range from 1 to 33380 (i.e. 33379)
* Mean= 16875.358
* Mean Standard Error= 69.476
* Standard Deviation= 9616.293
* Variance= 92473083.218
* Skewness= -0.018 means the data is skewed
* Skewness std. error= 0.018
* Unique=-
* Valid values= 19158

1. City:

* Unique=123
* Valid= 19158

1. City\_development\_index:

* min value= 0.448
* max value= 0.949
* sum= 15879.070
* Range = 0.501
* Mean= 0.829
* Mean Standard Error= 0.001
* Standard Deviation= 0.123
* Variance= 0.015
* Skewness= -0.995
* Skewness std. error= 0.018
* Unique=-
* Valid values= 19158

1. Gender:

* Unique=-
* Valid=14650

1. Relevant\_experience:

* Unique=2
* Valid= 19158

1. Enrolled\_university:

* Unique=4
* Valid= 18772

1. Education\_level:

* Unique=6
* Valid= 18698

1. Major\_discipline:

* Unique=7
* Valid= 16345

1. Experience:

* Unique=23
* Valid= 19093

1. Company\_size:

* Unique=9
* Valid= 13220

1. Company\_type:

* Unique=7
* Valid= 13018

1. Last\_new\_job:

* Unique=7
* Valid= 18735

1. Training\_hours:

* min value= 1
* max value= 336
* sum= 1252299
* Range = 335
* Mean= 65.367
* Mean Standard Error= 0.434
* Standard Deviation= 60.058
* Variance= 3607.019
* Skewness= 1.819
* Skewness std. error= 0.018
* Unique=-
* Valid values= 19158

1. Target:

* min value= 0.000
* max value= 1.000
* sum= 4777.000
* Range = 1.000
* Mean= 0.249
* Mean Standard Error= 0.003
* Standard Deviation= 0.433
* Variance= 0.187
* Skewness= 1.159
* Skewness std. error= 0.018
* Unique=-
* Valid values= 19158