# ETL Steps in PowerBi

There was a total of 490578 rows and 22 columns. We have kept the ETL steps simple as follows:

* Removing unnecessary columns - FILE NUMBER
* Changing Data Types
  + OFFENSE DATE – Changed Data Type to US Locale (MM/DD/YYYY)
  + OFFENSE TIME – A custom formula was written to correct the values
* Fixing Errors
  + OFFENSE TIME – 17 Errors were present and were removed
  + DISTRICT – 660 Errors were present. District is recorded as a numerical value however the error was zero was inputted as the letter “O”. The same was fixed by replacing it.
  + SUBZONE - 660 Errors were present. Subzone is recorded as a numerical value however the letter “H” was included which caused the error. The same was fixed by replacing it with the number six.
* Dealing with Missing Values
  + STREET DIRECTION – Has more than 83% missing values, column removed
  + Columns with missing values as follows were removed:

|  |  |
| --- | --- |
| **COLUMN NAMES** | **MISSING VALUES** |
| FULL ADDRESS | 7313 |
| ADDRESS NUMBER | 8130 |
| STREET NAME | 7313 |
| STREET TYPE | 23913 |
| CITY | 7313 |
| STATE | 7313 |
| ZIP CODE | 3047 |
| DISTRICT | 23122 |
| ZONE | 23122 |
| SUBZONE | 23122 |
| COMPLETE DISTRICT | 23122 |
| COUNCIL DISTRICT | 3315 |
| CRIME PREVENTION DISTRICT | 3305 |
| GEOLOCATION | 7313 |