**LAPORAN KUIS SPSS**

**ANALISIS PENJUALAN PROPERTI DI AMERIKA SERIKAT TAHUN 2001-2020**



**Disusun untuk memenuhi penugasan praktikum kuis pada mata kuliah Komputasi Statistik yang diampu oleh Yuliagnis Transver Wijaya, S.ST, M.Sc.**

Muhammad Ludvi Argorahayu

24

222313248

**KELAS 2KS3**

**PROGRAM STUDI D-IV KOMPUTASI STATISTIK**

**POLITEKNIK STATISTIKA STIS**

**JAKARTA**

**2025**

**PENDAHULUAN**

Diberikan sebuah dataset besar mengenai data penjualan properti yang ada di Amerika Serikat dalam kurun waktu 2001 sampai 2020. Dataset tersebut terlampir dalam tautan <https://s.stis.ac.id/data-properti-us> yang mana didalamnya terdapat raw data yang terdiri dari atribut serial number, list year, date recorded, town, address, asessed value, sale amount, sales ratio, property type, residential type, non use code, assessor remarks, dan location. Semantara itu jumlah recordnya sebanyak 997.214. Dalam praktikum kuis ini diminta untuk melakukan analisis dari awal untuk kemudian menghasilkan sebuah insight yang dapat berguna mengenai kasus penjualan properti di Amerika Serikat pada tahun 2001 sampai 2020 ini.

**PEMBAHASAN**

1. **Identifikasi Kebutuhan Data**
2. **Latar Belakang**

Dalam kurun waktu tahun 2001 sampai 2020 telah terjadi fluktuasi yang signifikan pada pasar properti di Amerika Serikat. Dalam proyek ini ingin dianalisis bagaimana tren harga properti yang ada di Amerika Serikat berubah dari tahun ke tahun dalam kurun waktu dua dekade ini, serta untuk mengetahui bagaimana persebaran wilayah di Amerika Serikat mengenai penjualan properti ini.

1. **Tujuan**

Analisis kasus penjualan properti di Amerika Serikat ini buertujuan untuk,

1. Mengidentifikasi tren harga jual properti di AS dari tahun 2001 sampai 2020.
2. Menggambarkan dinamika harga properti berdasarkan lokasi dan jenis properti.
3. Memberikan prediksi harga properti di masa depan berdasarkan pola historis.
4. **Kebutuhan Data**

Data yang digunakan berasal dari dataset penjualan properti yang mencakup 997.214 entri (record) dengan atribut serial number, list year, date recorded, town, address, asessed value, sale amount, sales ratio, property type, residential type, non use code, assessor remarks, dan location sesuai dengan yang terdapat dalam dataset tautan sebelumnya.

1. **Sumber Data**

Sumber data dari proyek ini tidak diidentifikasi karena dalam proyek praktikum kuis ini langsung diberikan datasetnya.

1. **Rancangan Pengolahan Data**

Proses melakukan analisis data dilakukan melalui beberapa tahapan agar menghasilkan insight yang berkualitas. Adapun rancangan pengolahan data sebagai berikut,

1. Preprocessing data

Tahapan preprocessing data menjadi tahapan yang paling banyak memkan waktu karena didalamnya memuat proses-proses perlakuan terhadap raw data sebalum dilakukan analisis statistik. Proses-proses yang dilakukan dalam preprocessing data yaitu,

1. Identifikasi tipe data setiap atribut.
2. Memperbaiki data missing.
3. Menangani pencilan.
4. Melakukan validasi data.
5. Melakukan pemilihan objek atau atribut data yang akan dianalisis.
6. Analisis eksploratif

Analisis eksploratif dilakukan untuk memahami pola umum dalam data, hubungan antar variabel, dan mendeteksi potensi insight awal. Beberapa kegiatan yang dilakukan dalam tahap ini adalah:

1. Melihat distribusi pada beberapa atribut yang dianalisis.
2. Menghitung ukuran statistik deskriptif.
3. Mengamati tran penjualan properti secara time series.
4. Mengidentifikasi penjualan properti berdasarkan tipe properti dan wilayah.
5. Analisis Prediktif

Analisis prediktif bertujuan untuk memodelkan pola dari data historis agar dapat digunakan dalam membuat prediksi atau estimasi. Dalam proyek ini akan dilakukan estimasi dengan menggunakan model ARIMA sesuai dengan arahan bahan materi praktikum.

1. **Pengambilan Data**

Proyek analisis data ini tidak meliputi proses pengambilan data karena dataset sudah disediakan sebagai bagan praktikum dalam tautan <https://s.stis.ac.id/data-properti-us>.

1. **Mengintegrasikan Data**

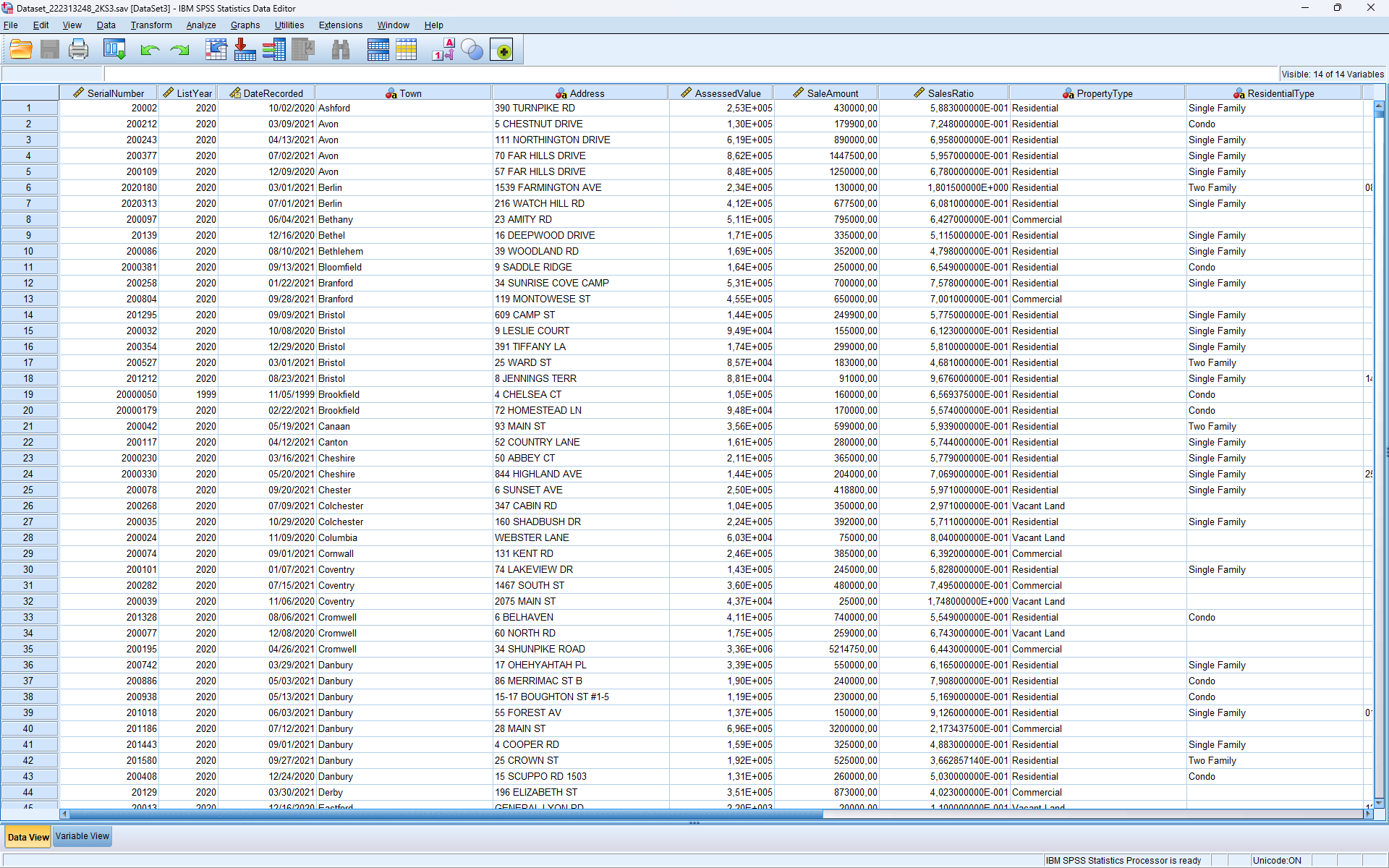
Dalam tahap integrasi data terdapat pemeriksaan kelayakan data untuk dilakuakan analisis yang meliputi pemeriksaan penanggung jawab dan pemeriksaan suatu data pakah bisa dipertanggungjawabkan. Akan tetapi karena dalam proyek ini dataset sudah langsung diberikan, maka diasumsikan bahwa dataset yang terkumpul terjamin kelayakannya dan dapat dipertanggungjawabkan.

1. **Menelaah Data**

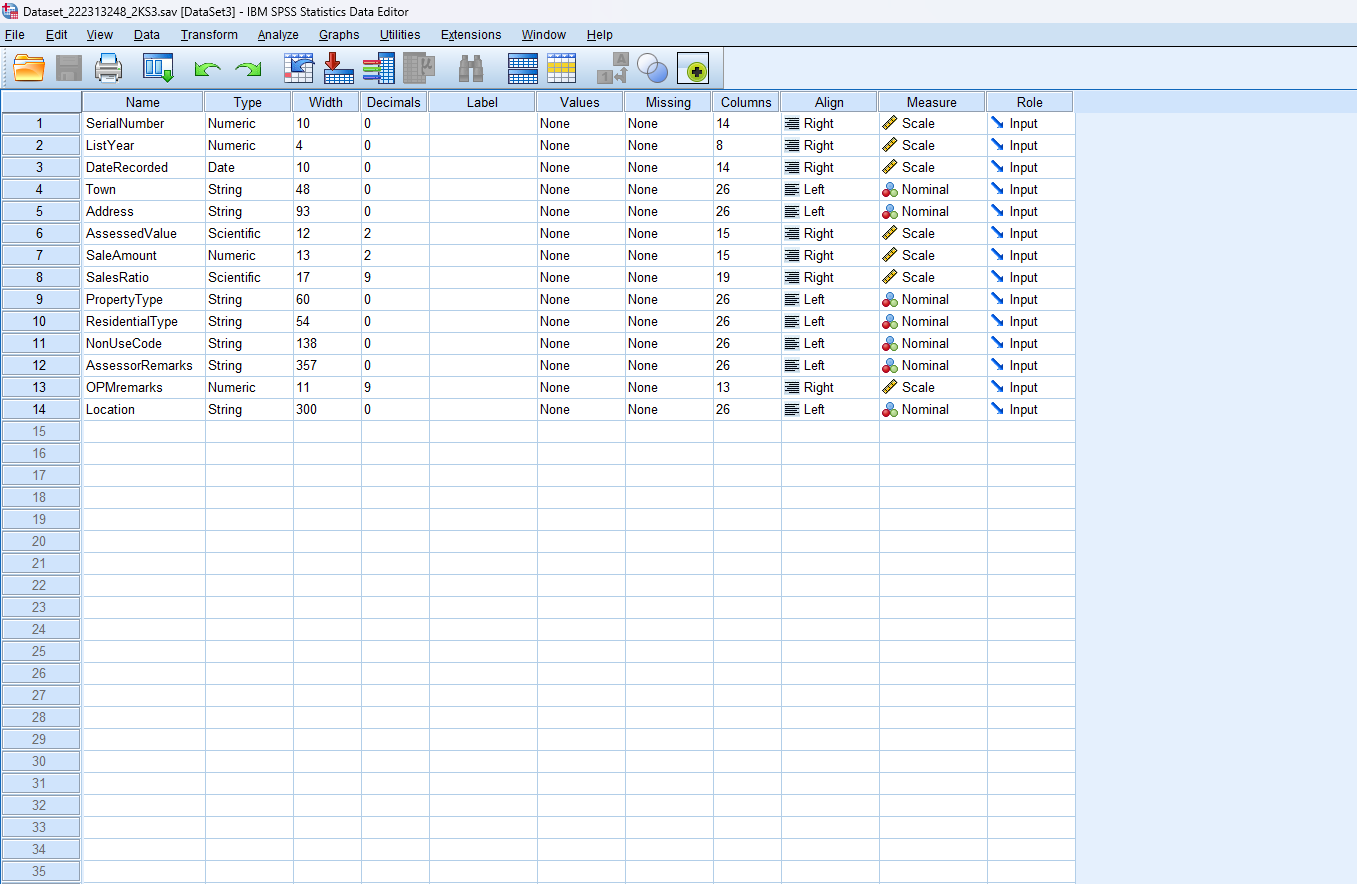
Menelaah data merupakan salah satu bagian dari preprocessing data bahkan sebagian bsear tahapan preprocessing data dilakukan pada tahap ini. Padsa tahap ini akan dilakukan identifikasi tipe data, Identifikasi datqa yang tidak relevan, pemeriksaan dan penanganan missing value, identifikasi masalah struktural data, dan penanganan terhadap data aneh dan pencilan.

1. **Entri Data**

Hal pertama yang harus dilakukan adalah memasukkan dataset yang dimiliki ke dalm software analisis dalam hal ini adalah SPSS. Dataset raw yang dimiliki memiliki ekstrensi csv sehingga dalam melakukan entri data ke SPSS diperlukan penyesuaian. Hasil entri ke dalam SPSS sebagai berikut, terdapat 14 field atau atribut dengan tipe data berbeda-beda dan jumlah record sebanyak 997.214 data.



1. **Identifikasi Tipe Data**

****

Atribut yang dimiliki oleh dataset sebanyak 14 atribut dengan tipe data setiap atribut berbeda-beda. Selanjutnya dengan memandang beberapa data di sheet data view, dilakukan beberapa penyesuaian terhadap properti yang dimiliki oleh atribut mulai dari type, width, decimals, skala pengukuran dan lain-lain. Adapun beberapa penyesuaian yang dilakukan sebagai berikut,

1. SerialNumber

Atribut SerialNumber memiliki tipe data numerik dan skala pengukuran nominal karena atribut ini digunakan sebagai id untuk properti, namun demikian atribut ini tidak unik jika dilihat secara keseluruhan akan tetapi akan menjadi unik jika dilihat dari segi tahun, kota, dan alamat.

1. ListYear

Atribut ListYear memiliki tipe data numerik dengan skala pengukuran nominal. Skala pengukuranj ini dipilih karena tahun meskipun berupa angka tidak memiliki arti jika dilakukan operasi aritmatika.

1. DateRecorded

Atribut DateRecorded memiliki tipe data date dan skala pengukuran nominal, di dalamnya memuat data tanggal, bulan, dan tahun pemeriksaaan properti.

1. Town

Atribut Town memiliki tipe data string karena terdiri dari kombinasi huruf, selain itu atribut ini memiliki skala pengukuran nominal.

1. Address

Atribut Address memiliki tipe data string karena terdiri dari kombinasi huruf dan angka, selain itu atribut ini memiliki skala pengukuran nominal.

1. AssessedValue

Atribut AssessedValue merupakan nilai yang diberikan assessor terhadap properti yang diperiksa. Atribut ini memiliki tipe data berupa numeric dan skala pengukuran rasio. Dalam SPSS skala pengukuran rasio akan diubah menjadi skala pengukuran scale.

1. SaleAmount

Atribut SaleAmount merupakan nilai jual yang dimiliki properti. Atribut ini memiliki tipe data berupa numeric dan skala pengukuran rasio. Dalam SPSS skala pengukuran rasio akan diubah menjadi skala pengukuran scale.

1. SalesRatio

Atribut SalesRatio memiliki tipe data berupa numeric dan skala pengukuran rasio. Dalam SPSS skala pengukuran rasio akan diubah menjadi skala pengukuran scale.

1. PropertyType

Atribut PropertyType memiliki tipe data string karena terdiri dari kombinasi huruf, selain itu atribut ini memiliki skala pengukuran nominal.

1. ResidentialType

Atribut ResidentialType memiliki tipe data string karena terdiri dari kombinasi huruf, selain itu atribut ini memiliki skala pengukuran nominal.

1. NonUseCode

Atribut NonUseCode memiliki tipe data string karena terdiri dari kombinasi angka dan huruf. Atribut ini juga memiliki skala pengukuran berupa nominal.

1. AssessorRemarks

Atribut AssessorRemarks merupakan catatan yang diberikan assessor terhadap properti yang diperiksa, atribut ini memiliki tipe data string karena terdiri dari kombinasi angka dan huruf. Atribut ini juga memiliki skala pengukuran berupa nominal.

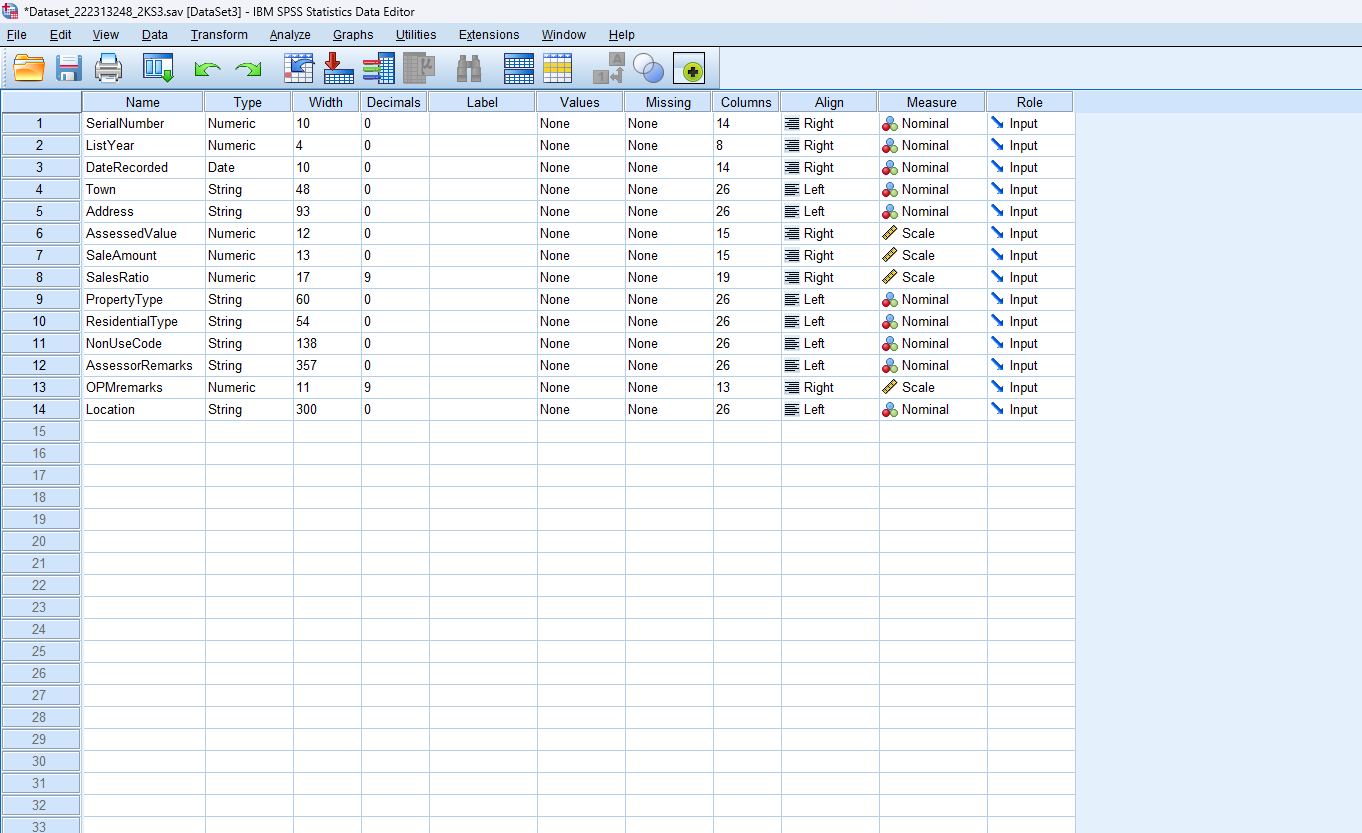
1. OPMremarks

Atribut OPMremarks memiliki tipe data numerik dengan sakala pengukuran Scale.

1. Location

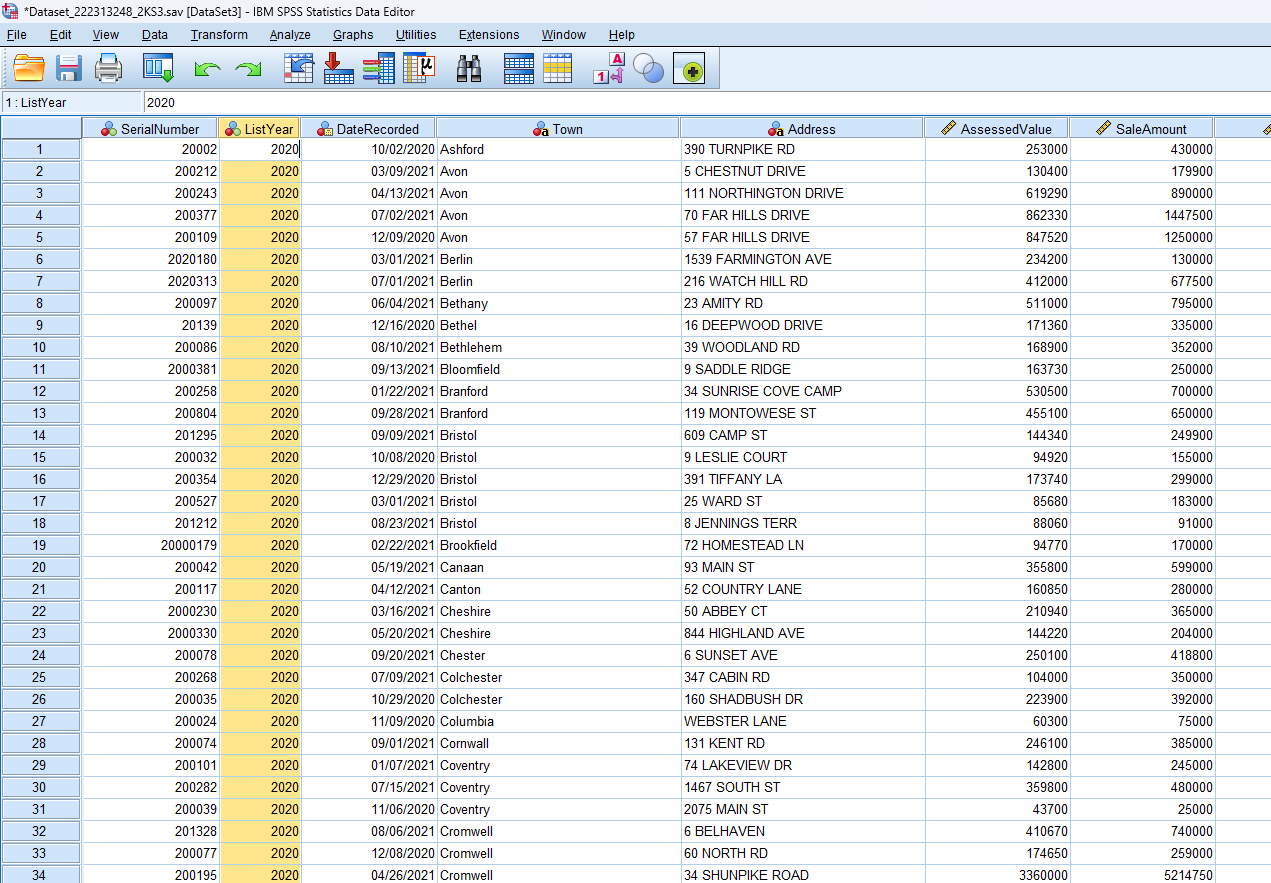
Atribut Location berisi informasi mengenai titik koordinat yang dimiliki oleh properti. Atribut ini memiliki tipe data string karena terdiri dari kombinasi angka, simbol, dan huruf. Selain itu, atribut ini memiliki skala pengukuran nominal. Atribut lokasi akan berguna ketika dilakukan analisis spasial yang mana akan memanfaatkan data titrik koordniat yang dimiliki.

Setelah dilakukan penyesuaian properti dari satiap atribut, diperoleh hasil sebagai berikut,

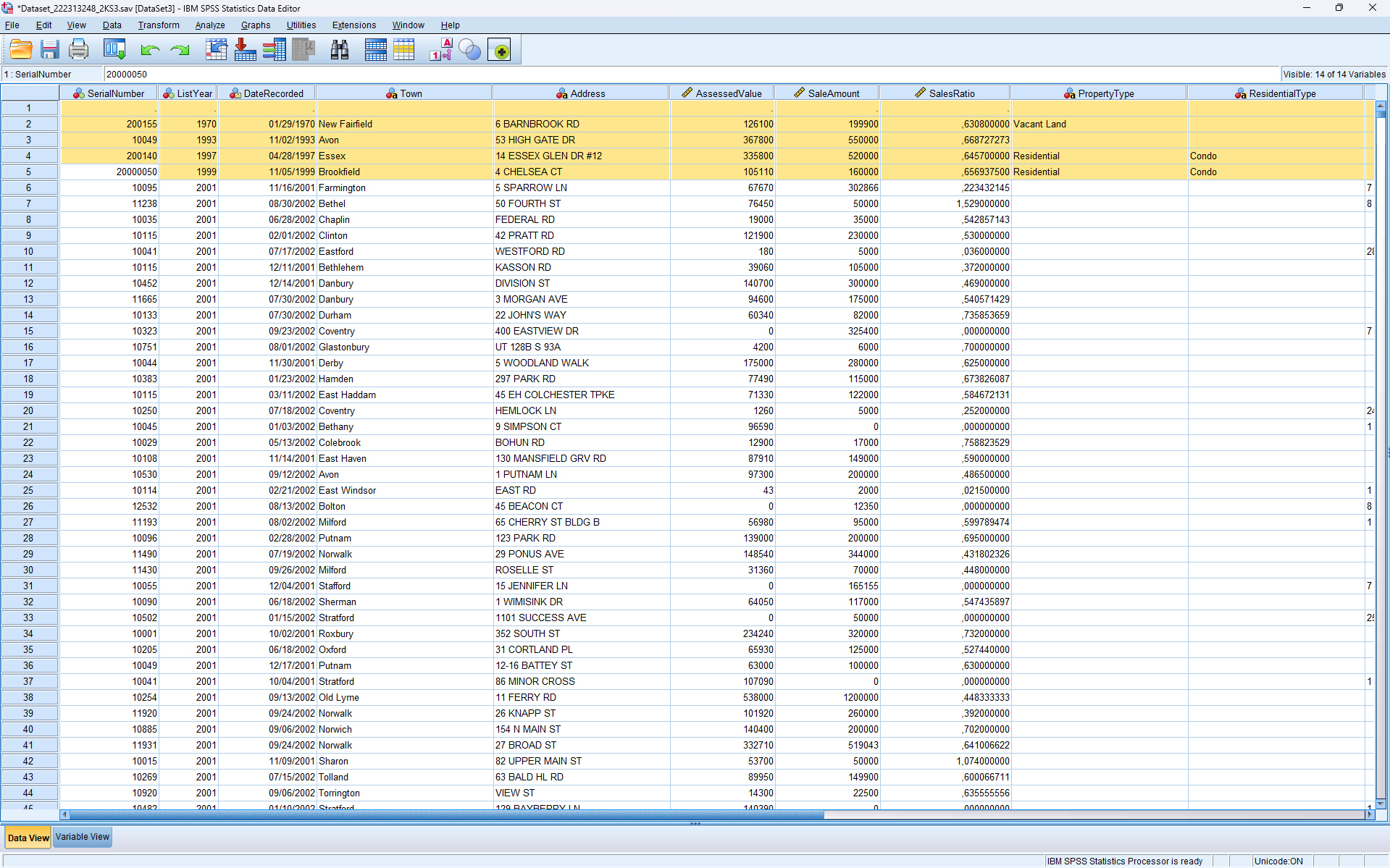


1. **Identifikasi Data Irrelevant**

Data yang akan dianalisis dalam proyek ini adalah data penjualan properti di Amerika Serikat dalam kurun waktu 2001 sampai 2020. Oleh karena itu akan dilakukan pengecekan apakah data yang terkumpul memuat data penjualan properti diluar kurun waktu tersebut. Batas atas tahun diperoleh 2020 sebagaimana tangkapan layar dibawah sehingga batas atas sudah sesuai.



Selanjutnya dilakukan pengecekan terhadap batas bawah tahun pemeriksaan, diperoleh hasil sebagai berikut,



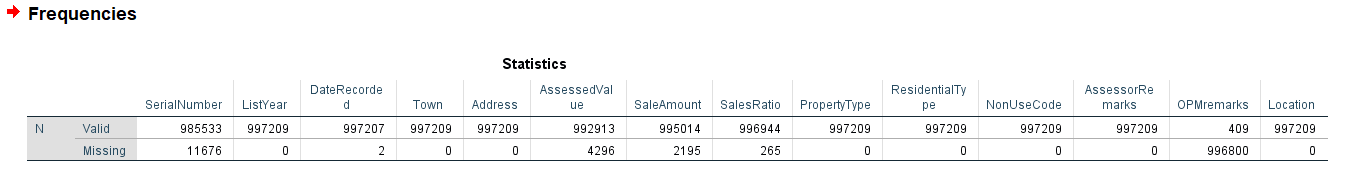
Terdapat empat record yang mana properti diperiksa diluar kurun waktu yang sedang dianalisis yaitu pada tahun 1970, 1993, 1997, dan 1999. Selain itu juga terdapat satu record yang mana untuk semua atributnya kosong sehingga tidak memiliki arti. Oleh karena itu terdapat 5 data yang tidak relevan untuk dianalisis dalam proyek ini sehingga bisa dilakukan removing data. Data record yang tersisa menjadi 997.209.

1. **Identifikasi Data Aneh**

Selanjutnya identifikasi kemunculan data aneh dilakukan. Pada identifikasi ini dimulai dengan mengidentifikasi dataaneh pada atribut PropertyType dan ResidentialType. Dengan menampilkan frekuensinya sebagai berikut,

1. **Identifikasi Missing Value**

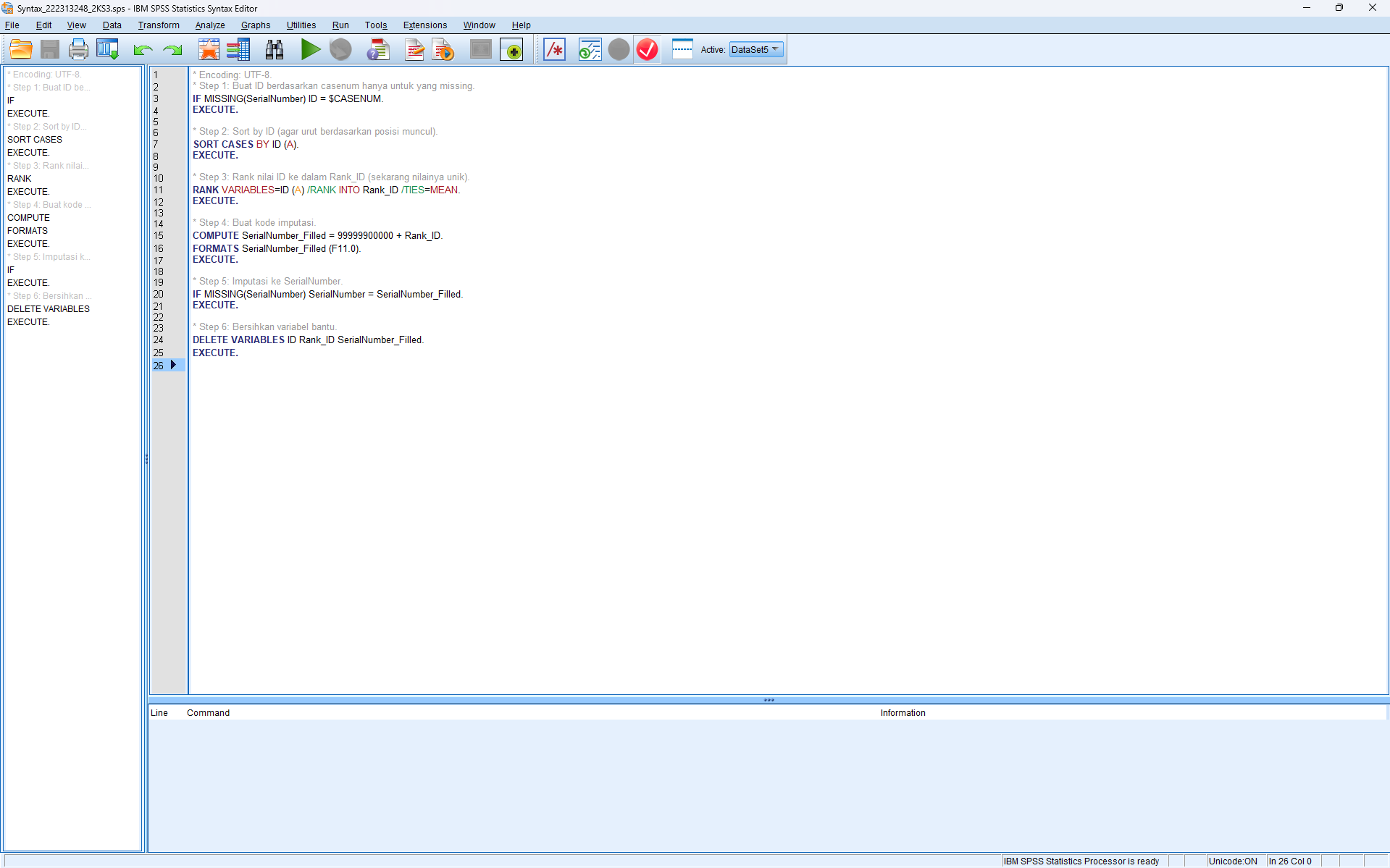
Sebelum melakukan analisis perlu diperiksa apakah data yang sudah dimiliki sudah lengkap semua atau terdapat beberapa data yang tidak lengkap sehingga perlu tindakan. Oleh karena itu dilakukan pengecekan missing value di SPSS untuk semua atribut. Pengecekan missing value dari data sebanyak 997.209 menghasilkan ringkasan sebagai berikut,



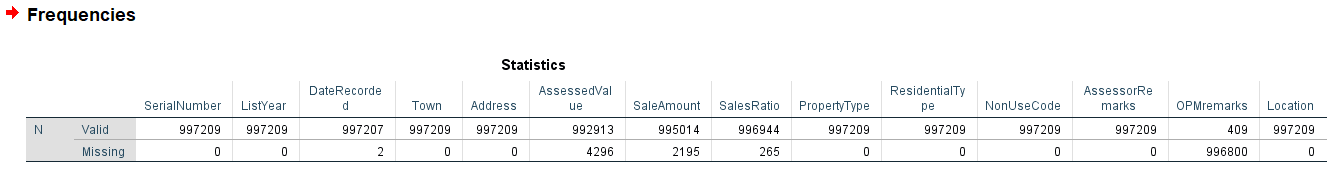
Terdapat 6 atribut yang memiliki missing value yaitu SerialNumbetr, DateRecorded, Assessedvalue, SaleAmount, SalesRatio, dan OPMremarks. Masing-masing atribut tersebut akan diberikan tindakan yang berbeda-beda.

1. Missing value SerialNumber

SerialNumber dalam data ini berperan selayaknya id, akan tetapi tidak unik jika dianalisis secara keseluruhan. Oleh karena itu SerialNumber tidak terlalu terlibat dalam analisis kuantitatif dalam tahap akhir nanti, sehingga untuk menangani missing value pada atribut ini akan dilakukan imputasi berupa kode khusus yang menandakan imputan missing. Karena tipe data dari atribut ini adalah numerik maka akan dilakukan inputan berupa data numerik yang khusus yaitu 999999\_. Tanda garis bawah akan diisi iteratif mulai dari 1 sampai 11676. Contoh 99999935 artinya adalah inputan missing value untuk SerialNumber yang ke-35. Dengan bantuan Artificial Intelligence akan dilakukan imputan tersebut melalui sintaks SPSS.



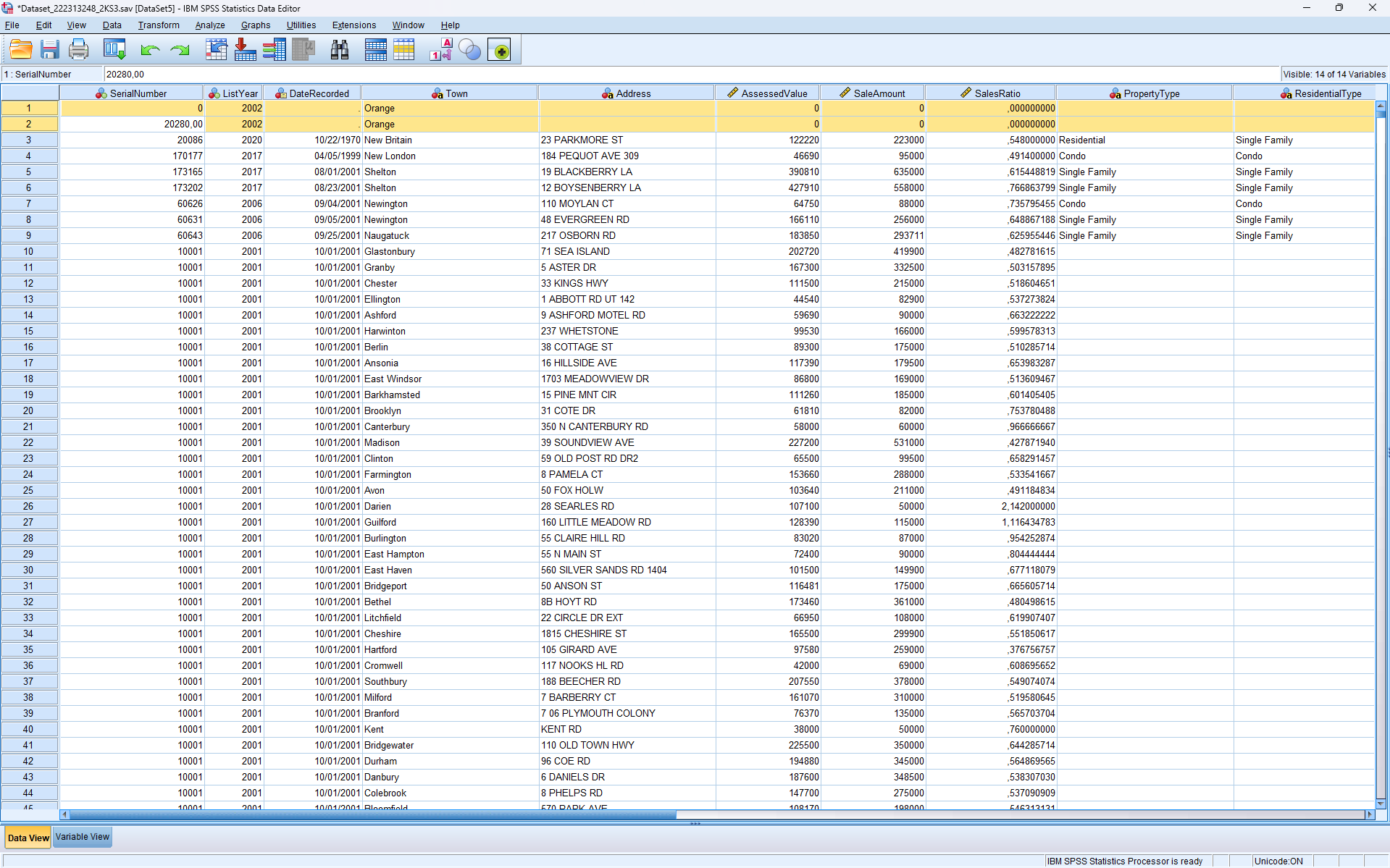
Selanjutnya untuk memeriksa hasilnya dilakukan pengecekan missing value lagi dengan hasil sebagai berikut,



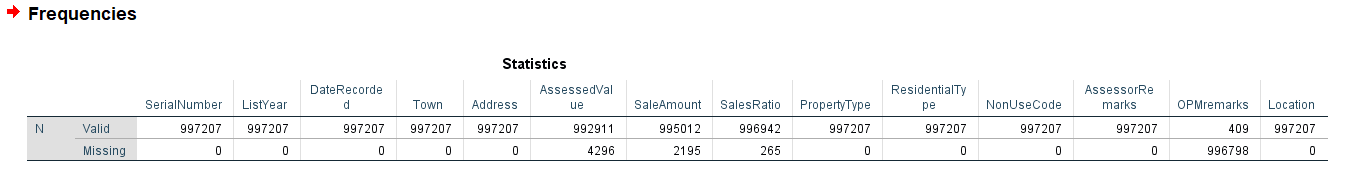
Terlihat bahwa missing value untuk atribut SerialNumber sudah nol tanpa mengurangi jumlah record.

1. Missing value DateRecorded

Identifikasi terhadap missing value yang dimiliki oleh atribut dateRecorded dapat dilihat dengan mengurutkan secara ascending atribut DateRecorded. Diperoleh hasil sebagai berikut,

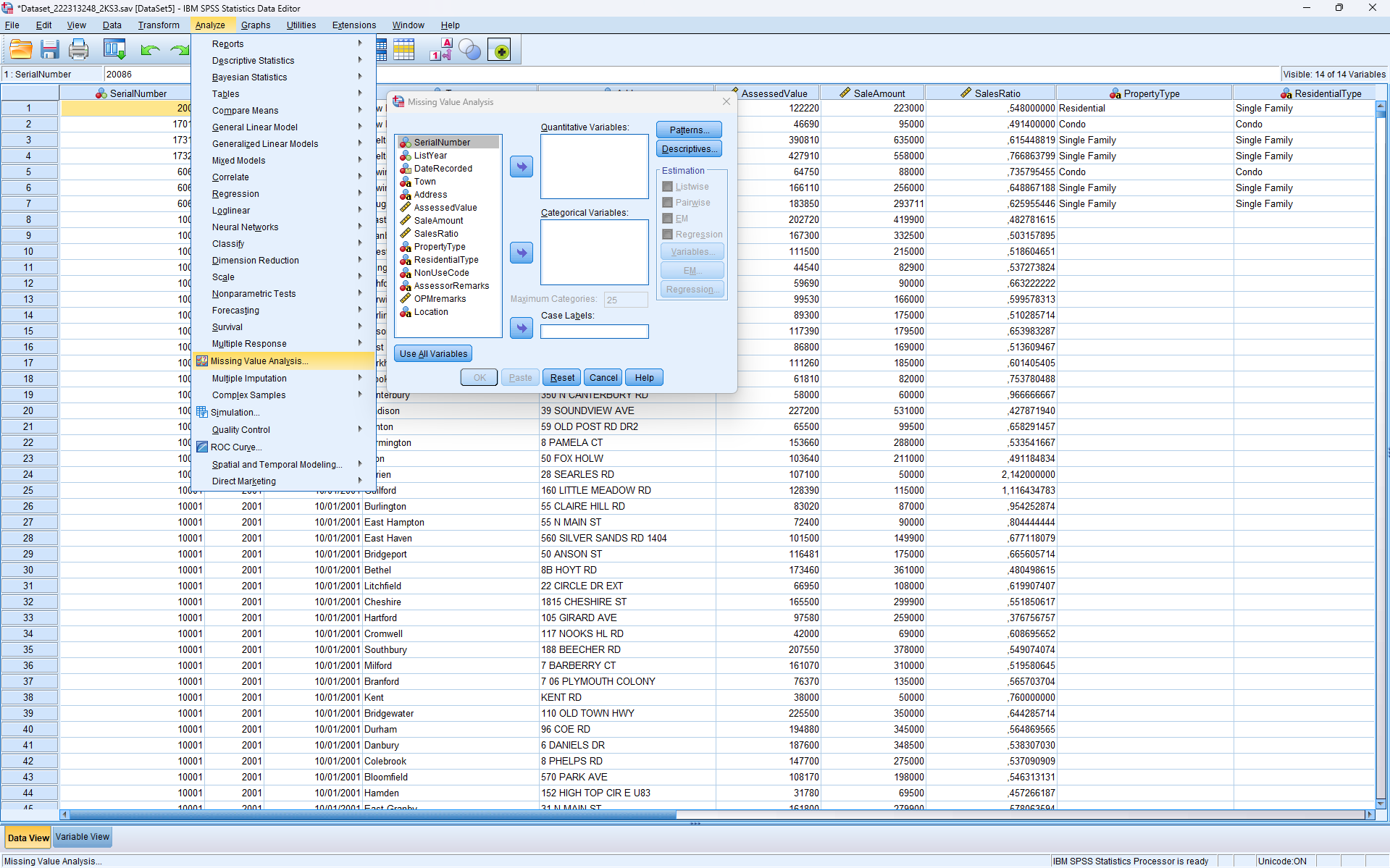


Terlihat bahwa data missing pada atribut ini tidak valid karena pada atribut yang lain juga kosong. Oleh karena itu untuk menangani missing value pada atribut DateRecorded ini dilakukan removing data sehingga sisa record data masih 997.207. Hasil pengecekan dapat dilihat pada tangkapan layar di bawah ini.

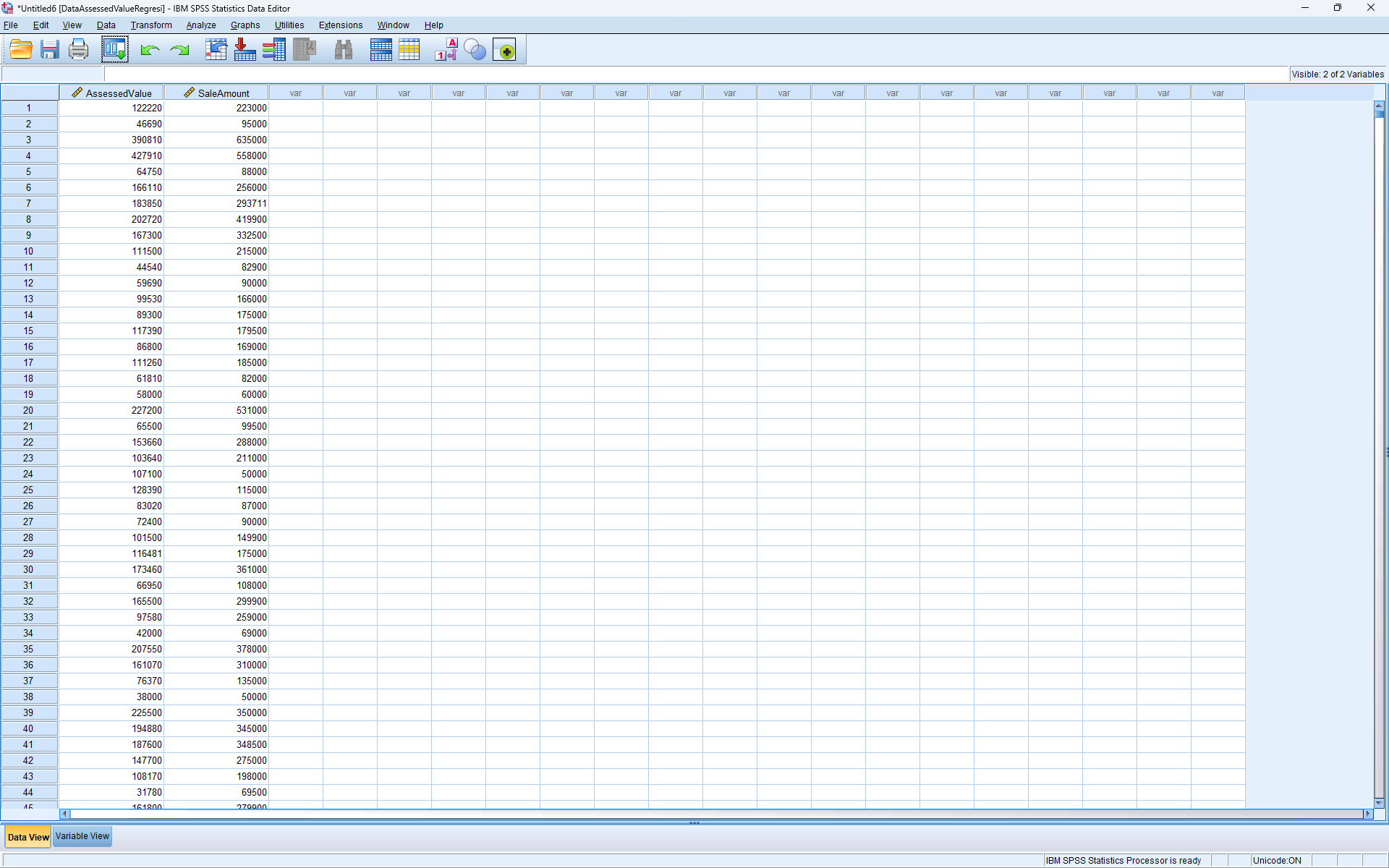


1. Missing value AssessedValue

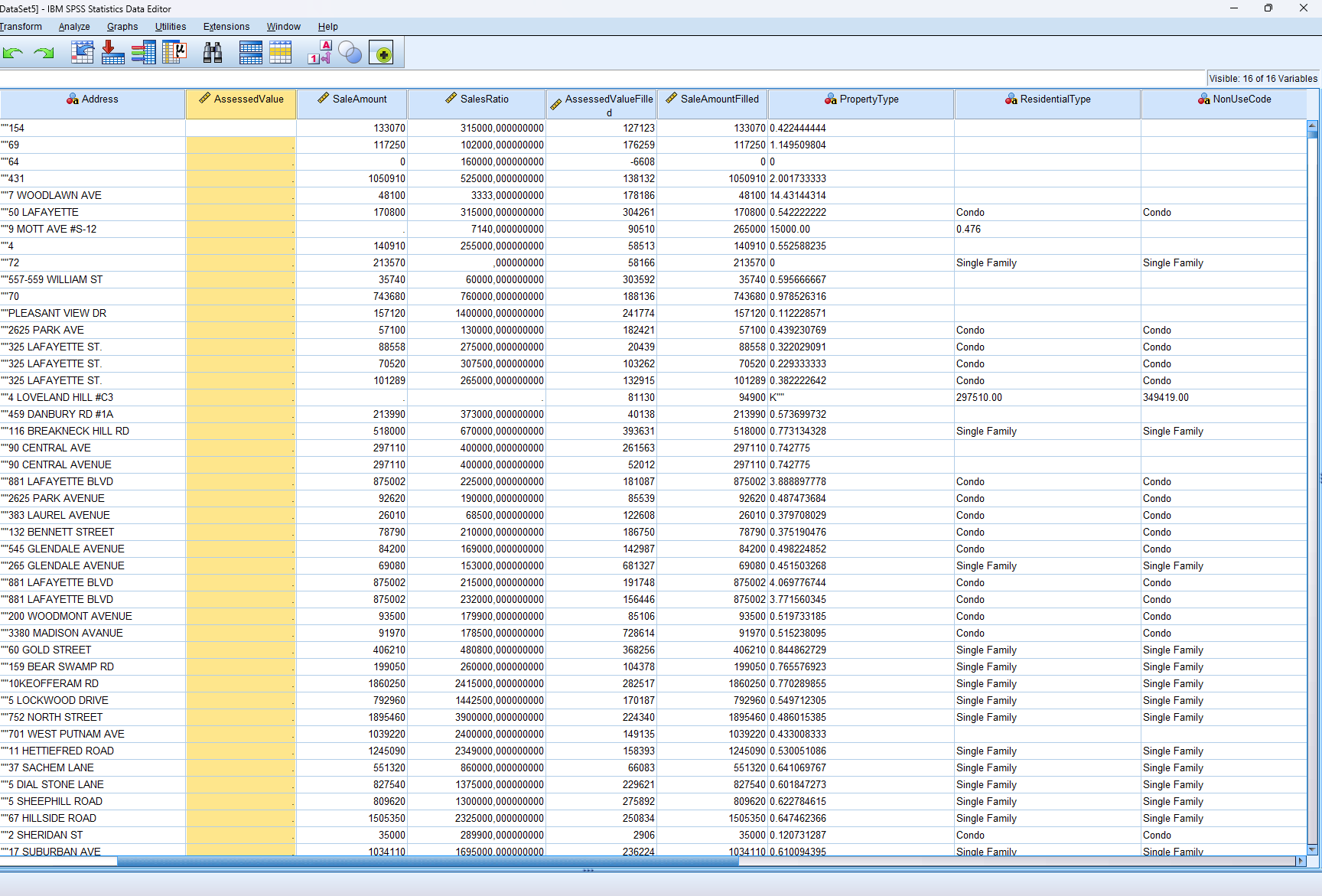
Atribut AssessedValue berisi data penilaian properti yang diberikan oleh assessor. Oleh karena itu data atribut ini menjadi sangat penting untuk dianalisis. Mengingat urgensi atribut ini, maka tidak langsung dilakukan remove data pada missing valueenya, akan tetapi akan dilakukan imputasi dengan menggunakan fitur yang ada di SPSS yaitu Missing Value Analysis yang mana akan menggunakan estimasi regresi dengan atribut lain yaitu SaleAmount. Hal ini dilakukan karena hubungan antara penilaian assessor dengan jumlah penjualan saling berhubungan secara logika. Adapun langkahnya adalah



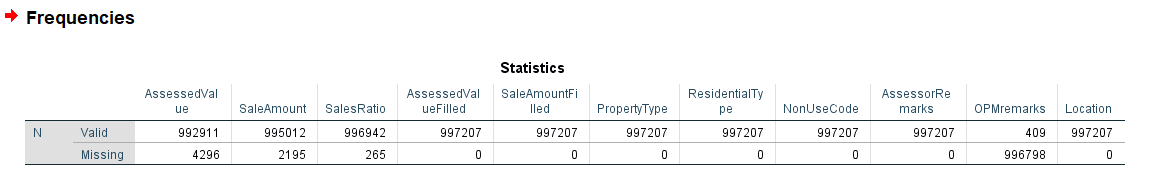
Lalu memasukkan variabel AssessedValue dan SaleAmount ke Quantitative Variables dan memilih estimasi Regression. Diperoleh data baru sebagai berikut,



Selanjutnya data tersebut disalin ke dataset utama dengan tetap memperhatikan urutannya. Di dataset utama akan disalin kedalam atribut baru yaitu AssessedvalueFilled dan SaleAmountFilled. Berikut adalah contoh hasil imputasinya,



Selanjutnya akan diperiksa apakah masih ada missing value untuk atribut AssessedValue. Jika masih ada maka akan dilihat proporsinya terlebih dahulu terhadap keseluruhan data sebelum dilakukan removing data.



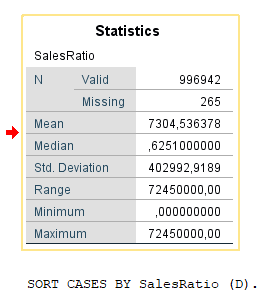
Diperoleh hasil bahwa variabel atau atribut AssessedvalueFilled dan SaleAmountFilled sudah tidak memiliki missing value. Terlepas apakah data itu wajar atau tidak akan dianalisis di tahap selanjutnya. Sehingga analisis data dapat terus dilanjutkan dengan memakai atribut yang telah diimputasi ini.

1. Missing value SaleAmount

Penanganan missing value dengan imputasi atribut SaleAmount sama persis dengan imputasi atribut Assessedvalue.

1. Missing value SalesRatio

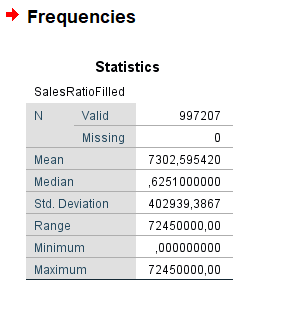
Missing value dari atribut SalesRatio sebanyak 265 sebagaimana yang ditampilkan pada tangkapan layar dibawah ini.



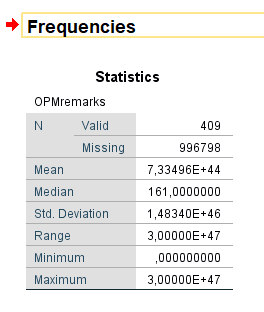
Upaya yang dilakukan untuk menangani masalah missing value pada atribut ini yaitu dengan melakukan imputasi nilai median ke missing value. Dengan menggunakan sintaks SPSS berikut,



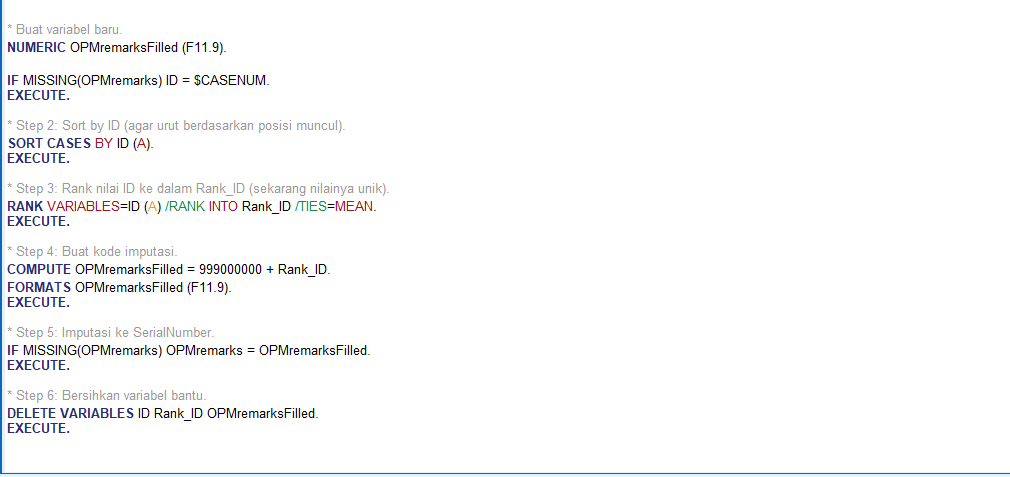
Diperoleh hasil bahwa missing value untuk atribut SalesRatio sudah nol.



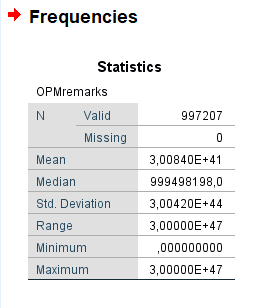
1. Missing value OPMremarks



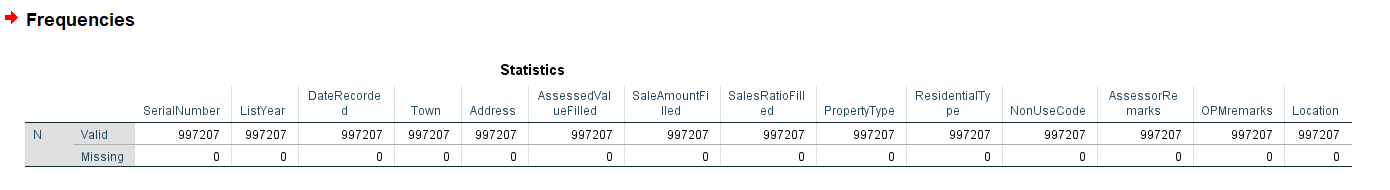
Atribut OPMremaks memiliki missing value sebanyak 996.798, dengan jumlah missing value sebanyak itu tidak wajar jika dilakukan penghapusan data. Mengingat data ini tidak terlalu berpengaruh pada tahap analisis statistik maka cukup dilakukan imputasi kode missing saja sperti hanya pada atribut SerialNumber. Dilakukan imputasi kode 999 dengan menggunakan sintaks berikut,



Setelah dilakuakn imputasi maka diperoleh hasil bahwa missing value untuk atribut OPMremarks sudah nol tanpa melakuakan removing data.

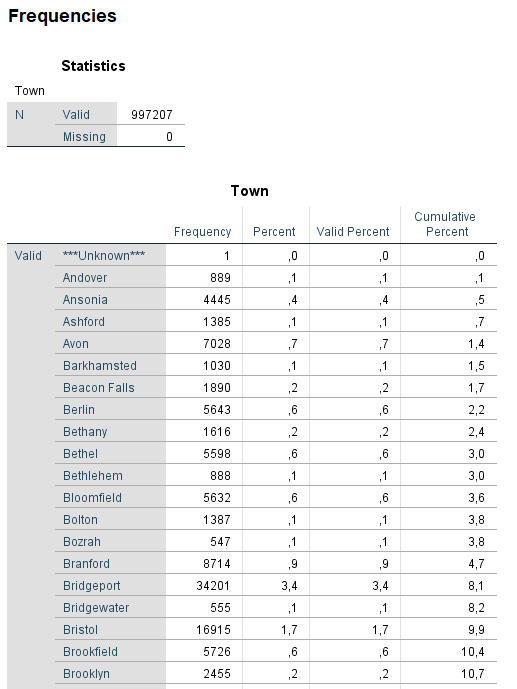


Setelah dilakukan penanganan terhadap missing value untuk setiap atribut, maka sudah dipastikan semua atribut yang ada dalam dataset sudah valid semua tidak ada yang missing.

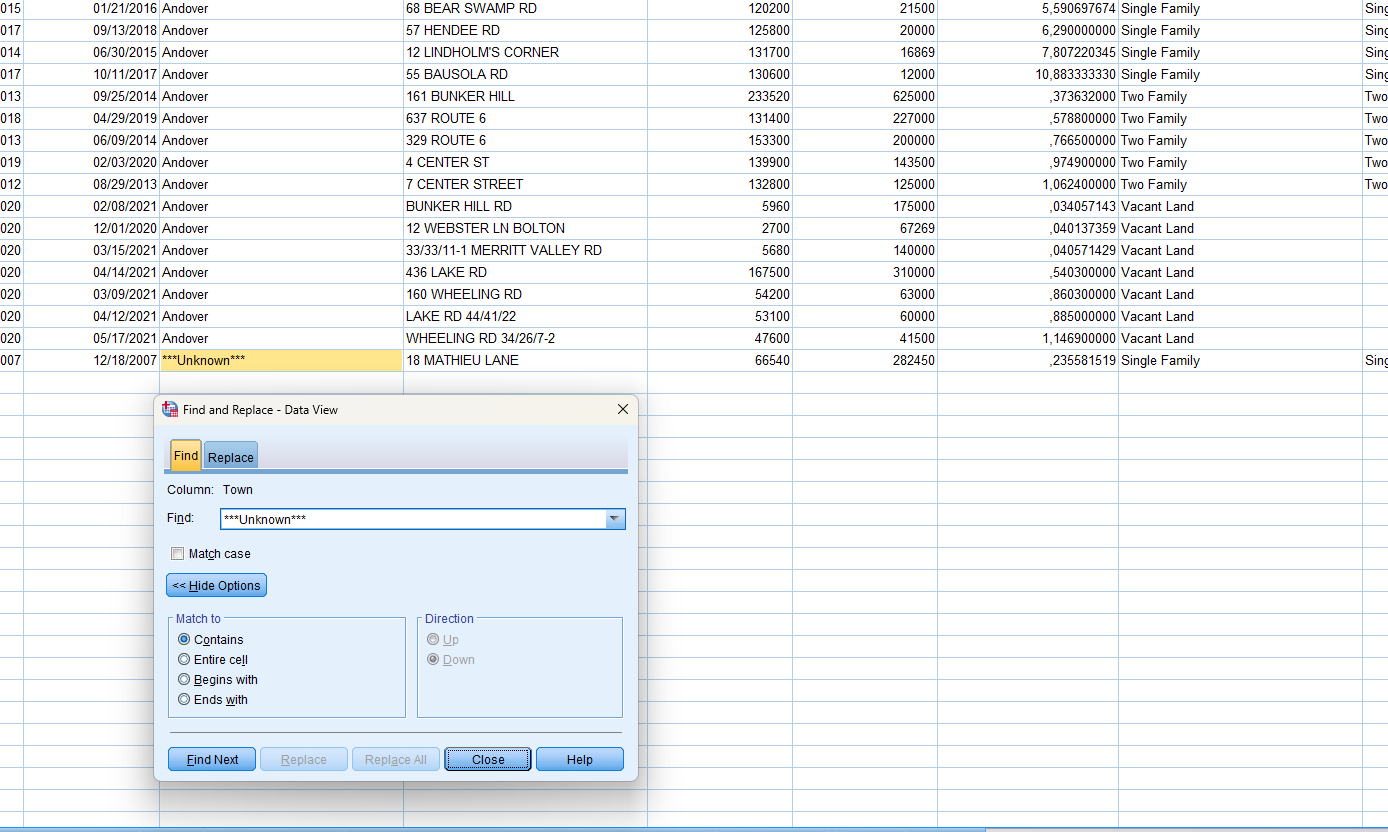


1. **Identifikasi Masalah Struktural**

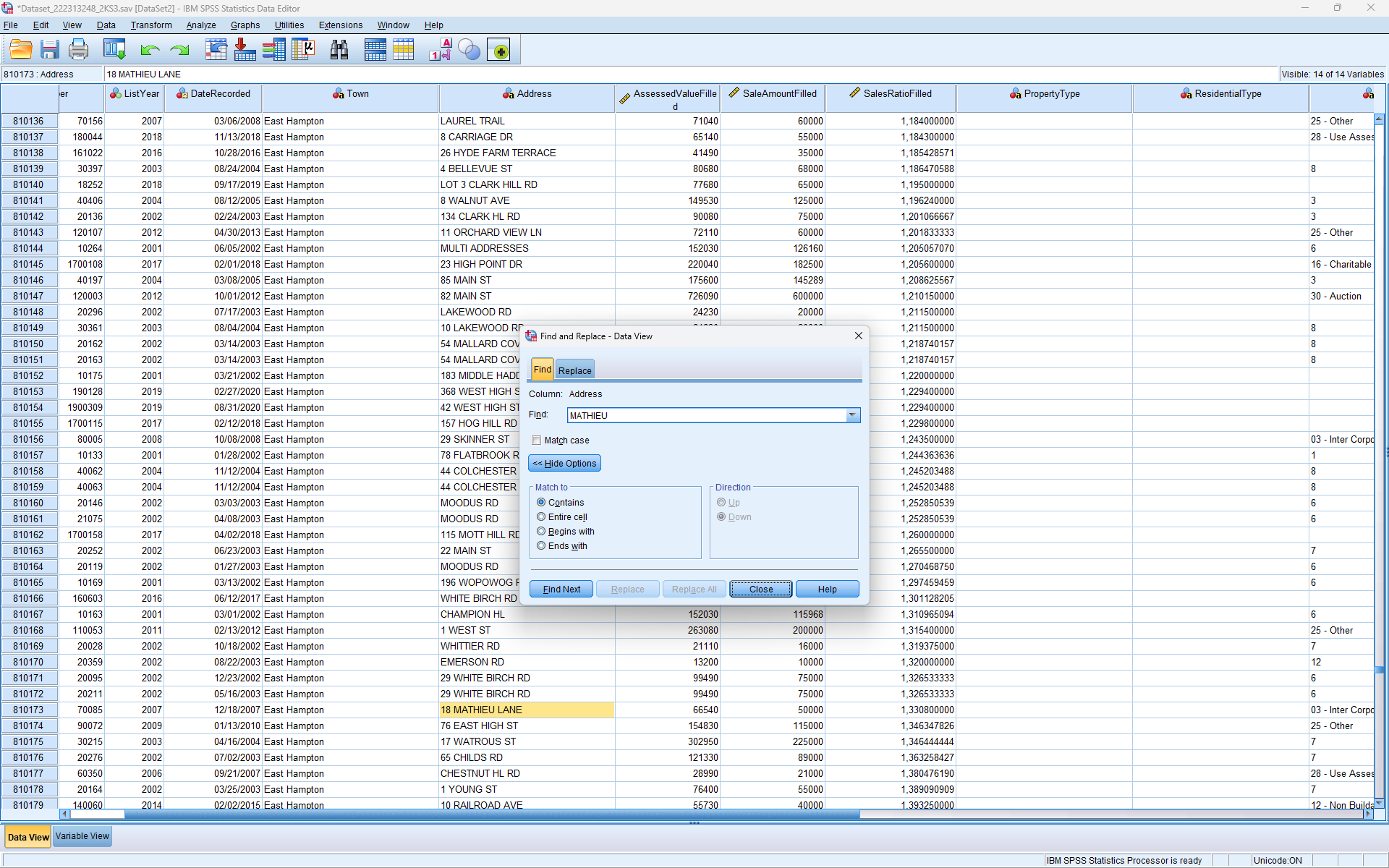
Atribut yang akan diidentifikasi masalah struktural datanya adalah Town.Pada atribut Town secara keseluruhan tidak terdapat masalah struktuiral yang berarti seperti typo yang banyak, akan tetapi terdapat satu masalah dimana ada kesalahan input yaitu terdapat kota \*\*\*Unknown\*\*\* sebagaimana tangkapan layar di bawah ini,



Selanjutnya dilakukan pencarian data nama kota \*\*\*Unknown\*\*\* dan diperoleh bahwa data ini memiliki Address MATHEU LANE.

\

Oleh karena itu untuk menangani kasus ini selanjutnya dilakukan pencarian data dengan Address MATHEU LANE dan diperoleh bahwa properti yang memiliki alamat MATHEU LANE berada di Kota East Hampton. Sehingga data \*\*\*Unknown\*\*\* tadi diperbarui menjadi East Hampton.

****

1. **Identifikasi Pencilan**
2. **Memvalidasi Data**
3. **Menentukan Objek Data**
4. **Membuat Business Intelligence**

**KESIMPULAN**