**Made creatively by : Christopher Darren - 00000054804**

**IS 429 AL Big Data Analytics**

# LINK SUMBER DATASET:

[**https://www.kaggle.com/datasets/dsfelix/us-stores-sales**](https://www.kaggle.com/datasets/dsfelix/us-stores-sales)

**LINK LOGO PT KIMOCHIMART:**

[**https://www.canva.com/design/DAFlqveg3kI/Y4FCG5MKUaqiUgIu6x\_v\_Q/edit?utm\_content=DAFlqveg3kI&utm\_campaign=designshare&utm\_medium=link2&utm\_source=sharebutton**](https://www.canva.com/design/DAFlqveg3kI/Y4FCG5MKUaqiUgIu6x_v_Q/edit?utm_content=DAFlqveg3kI&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton)

**LINK FLOW PENELITIAN:**

[**https://drive.google.com/file/d/1Cb7JwUoOX62hBKopWHtrZtV19bgSfSqB/view?usp=sharing**](https://drive.google.com/file/d/1Cb7JwUoOX62hBKopWHtrZtV19bgSfSqB/view?usp=sharing)

# LINK SAS VIYA: <https://v4e038.vfe.sas.com/links/resources/report?uri=%2Freports%2Freports%2F54a750e1-1f76-4565-b926-0f9e300af4c9&page=vi6368>

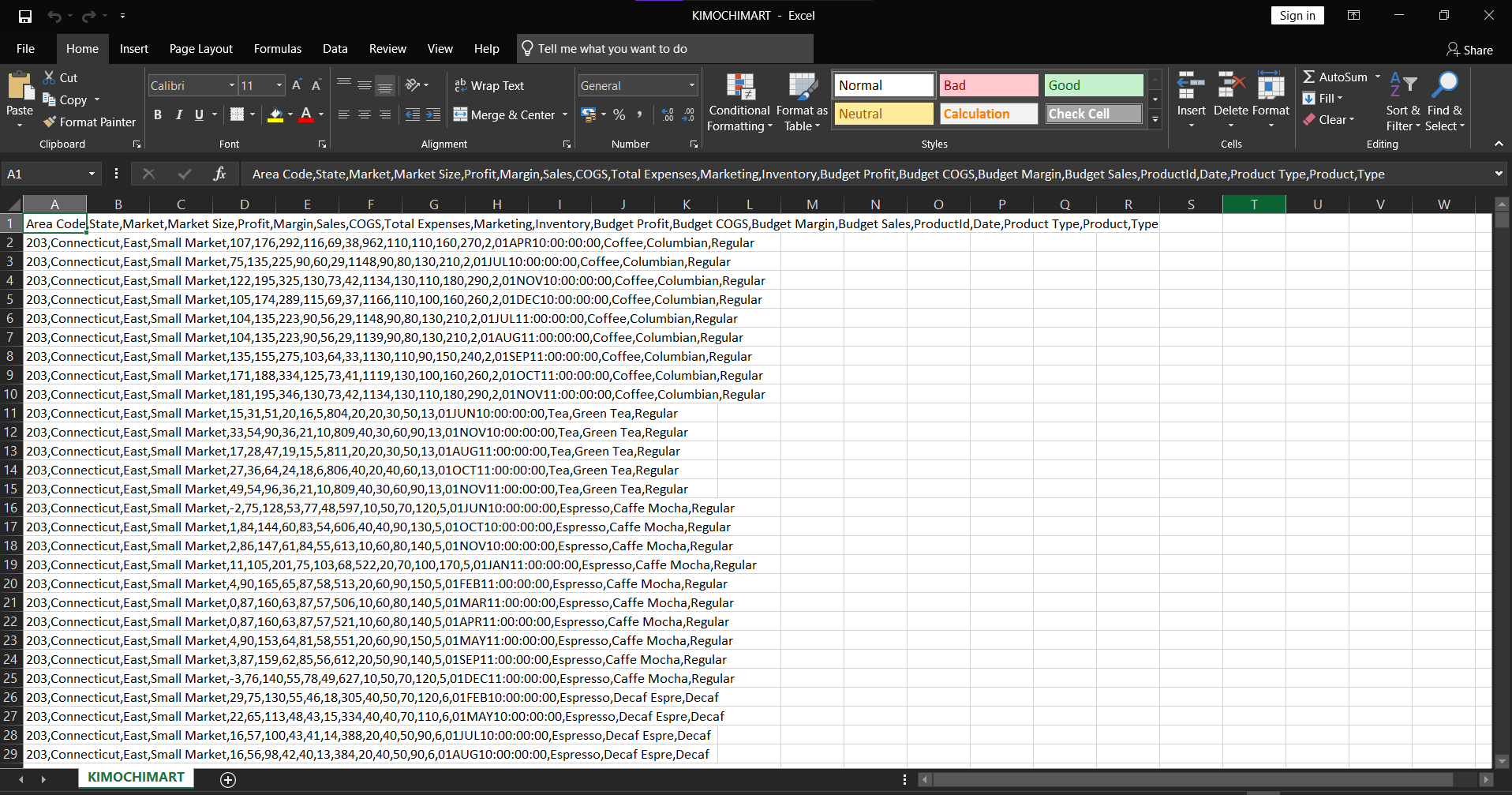


Peneliti mengambil bisnis KimochiMart. KimochiMart merupakan sebuah bisnis retail yang kegiatan sehari harinya adalah menjual produk andalan mereka berupa kopi dan teh yang terletak di United States, tepatnya di Oregon. KimochiMart sudah berdiri sejak tahun 2001 hingga sekarang. Tidak hanya itu KimochiMart juga memiliki banyak cabang sehingga menjadi salah satu bisnis ritel tersukses di United States pada masanya. Namun KimochiMart menghadapi kompetitor dimana pada tahun 2010 penjualan KimochiMart menurun, oleh karena itu diperlukannya teknik machine learning dalam menyelesaikan masalah yang sekarang sedang dihadapi oleh Kimochi Mart. Penjelasan lebih lengkapnya bisa lihat penjelasan dibawah ini.

# Question 1: CLO-3 Sub-CLO-6 and CLO-2 Sub-CLO-7 Weight (40 %)

* **Manage your Data Preparation and Exploration Process corresponding to the data curation process.**
* **The journal writing in section III contains an explanation of EDA (Exploratory Data Analysis).**
* **Create your EDA and Data preparation process with the SAS® Data Preparation Application.**

## Data Understanding

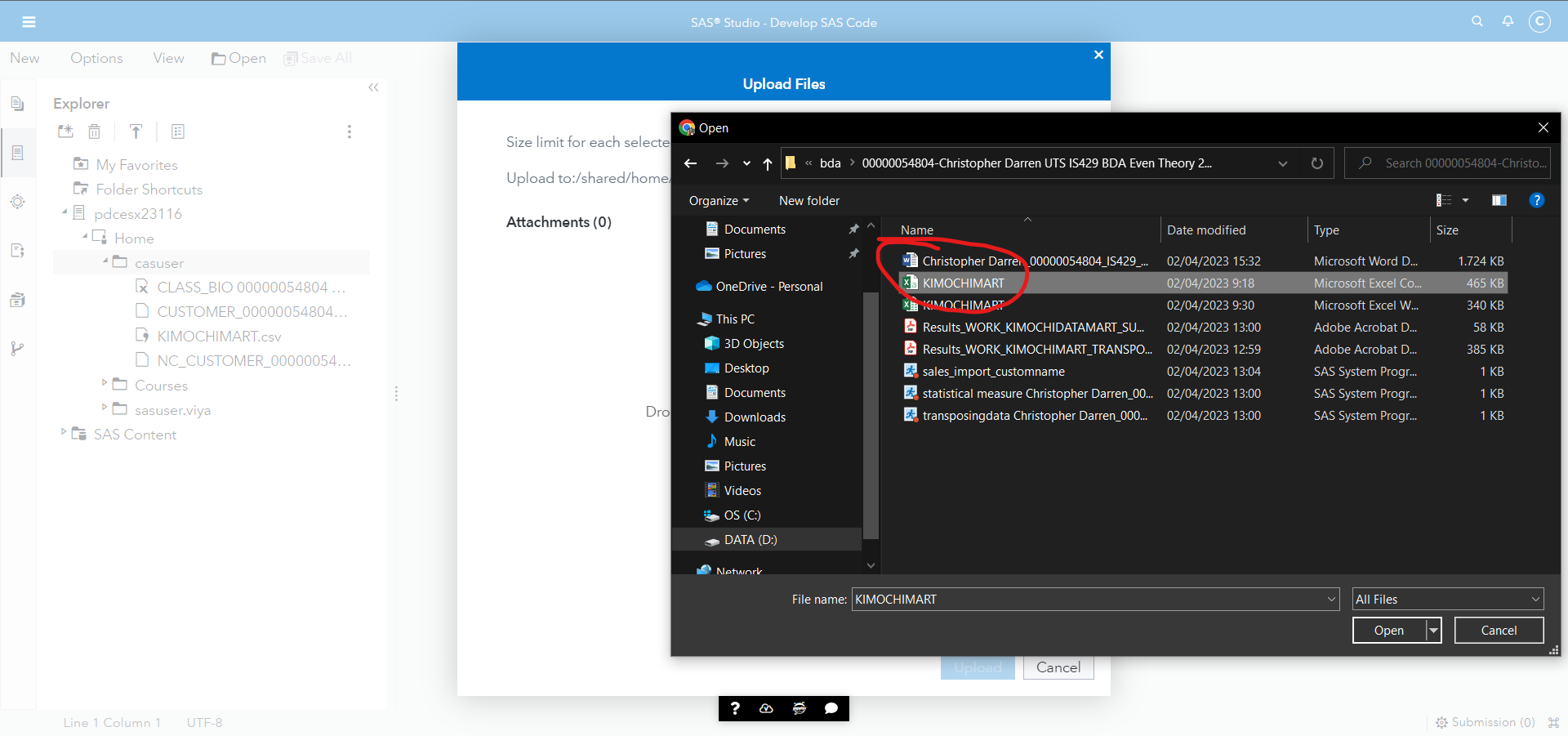


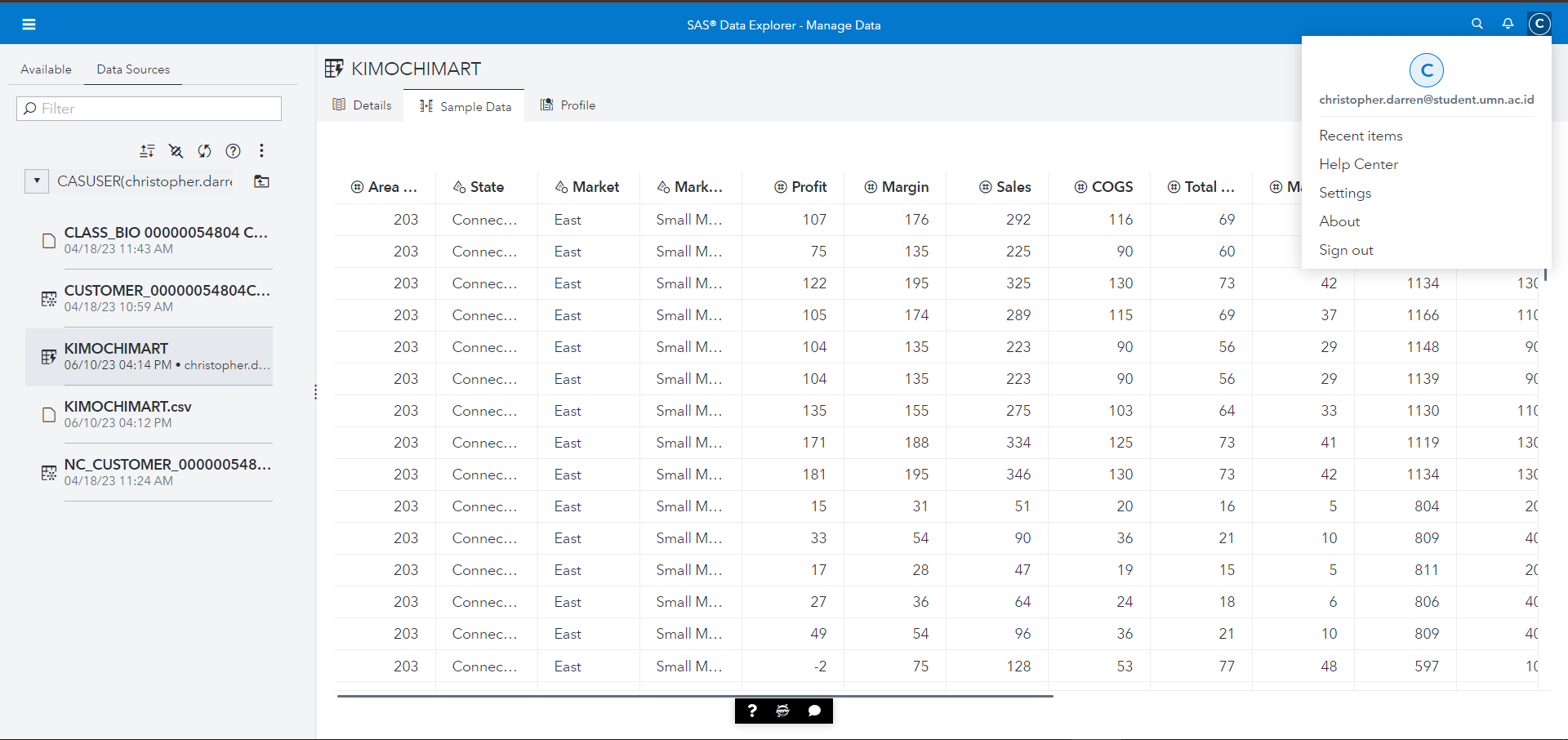
**Gambar 1. Data set mentah dalam bentuk .csv**

Berdasarkan gambar diatas, data KimochiMart diambil dari sumber kaggle.com, kaggle.com merupakan sebuah website yang digunakan untuk mengambil data. Data ini mempunyai rows ataupun record sebanyak 4000 rows. Bisa dibilang cukup karena datanya sudah termasuk banyak dan cocok untuk dipakai sebagai penelitian. Data ini berada di tahun 2010 - 2011. Untuk atributnya terdiri dari 20 atribut yang berbeda. Penjelasannya adalah sebagai berikut:

| **Atribut** | **Deskripsi** |
| --- | --- |
| Area code | kode dari store dan ini berbeda beda atau UNIQUE; |
| State | Lokasi *state* dari store tersebut; |
| Market | Lokasi region dari store tersebut, misal West, North, South, East dan lain lain; |
| Market Size | Ukuran dari *Store* tersebut apakah besar atau kecil. Diindikasikan dengan Small market, Big Market, dan lain-lain. |
| Profit | Keuntungan dalam Dollars ($); |
| Margin | Profit + Total Expenses ($) OR Sales - COGS ($); |
| Sales | nilai sales secara keseluruhan($)/ penjualan; |
| COGS | Cost of Goods Sold ($); |
| Total Expenses | Total pengeluaran untuk mendapatkan product agar bisa dijual dalam($); |
| Marketing | Pengeluaran dalam marketing dalam ($); |
| Inventory | Inventory Value of the Product in the Sale Moment ($); |
| Budget Profit | Expected Profit ($); |
| Budget COGS | Expected COGS ($); |
| Budget Margin | Expected Profit + Expected Total Expenses ($) OR Expected Sales - Expected COGS ($); |
| Budget Sales | Budget sebuah sales yang diharapkan akan keluar dalam ($) Expected Value Acquired in Sales ($); |
| ProductID | Merupakan ID product, dan ID ini unique artinya berbeda dari yang lain; |
| Date | Tanggal transaksi dilakukan |
| ProductType | Kategori produk ; |
| Product | Deskripsi product coffee secara detail; |
| Type | Tipe kopi ataupun tipe teh yang dibeli; |

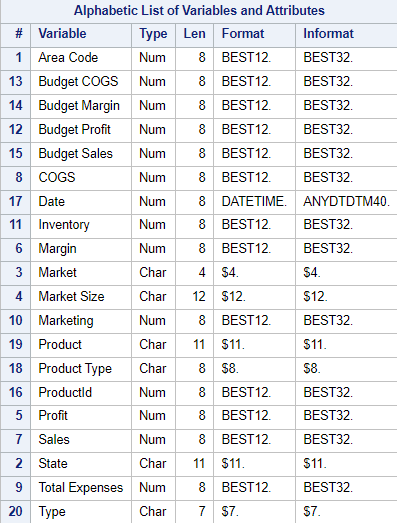
Untuk dataset yang sudah dijelaskan sebelumnya diperoleh dari KimochiMart adalah awalnya dengan format .csv , namun peneliti ingin mengimport data .csv tersebut kedalam SAS VIYA analytics supaya bisa lebih mudah dibaca dan tidak berantakan perbatasan garisnya. .Bisa dilihat pada gambar dibawah ini.

****



**Gambar 2. Data KimochiMart yang berhasil diimport menuju SAS VIYA**

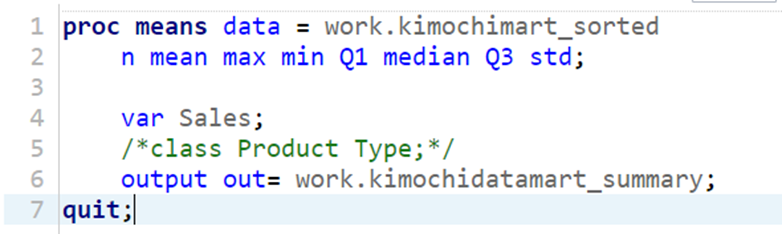
Berdasarkan tampilan tersebut, pada awalnya peneliti masuk ke bagian SAS Studio develop yang ada pada SAS VIYA kemudian peneliti mengimport data KimochiMart sesuai perintah dari SAS studio dengan cara *upload data*. Ketika sudah berhasil maka peneliti bisa langsung menuju halaman SAS VIYA - *manage data*  apakah datanya berhasil di import atau tidak. Terlihat jelas bahwa data KimochiMart berhasil di upload menuju SAS Viya sehingga bisa dilakukan *exploratory data analysis.* Selanjutnya peneliti akan menampilkan beberapa *data type* pada kolom KimochiMart yakni dengan cara menulis code PROC CONTENTS DATA=WORK.KimochiMart; RUN; pada aplikasi SAS Studio. Bila sudah di run maka hasilnya akan menjadi seperti ini.

****

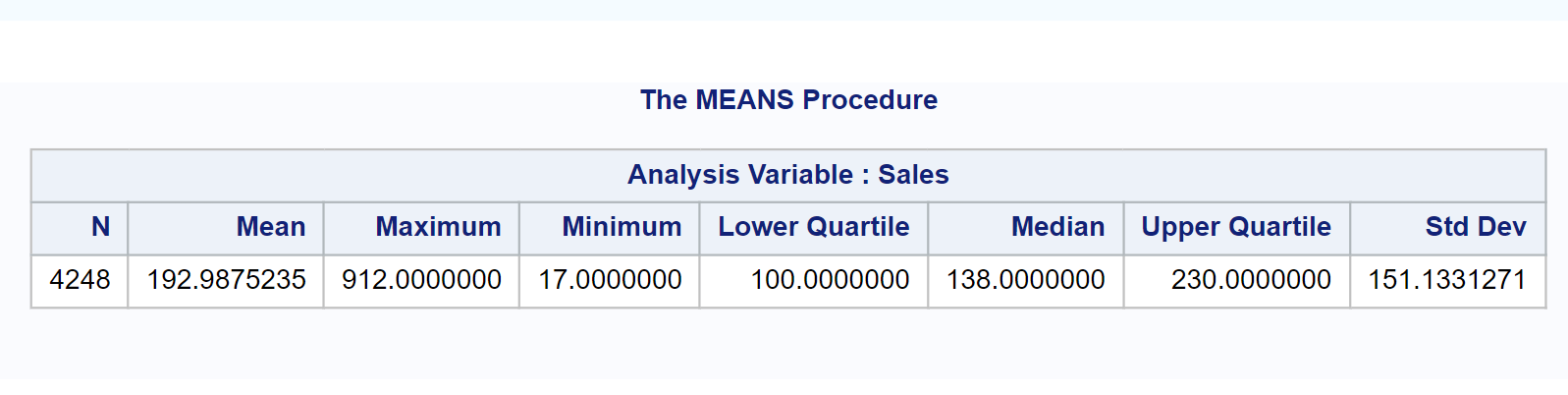
**Gambar 3. Melihat data type dari KimochiMart**

Berdasarkan gambar tersebut tipe data yang sering muncul adalah NUM, dan CHAR. Hal ini menunjukkan bahwa akan memudahkan peneliti dalam pembuatan model nantinya karena banyak numerik, begitu pula juga kategorikalnya yang bisa mendukung penelitian.

Kemudian peneliti juga melakukan perhitungan statistik untuk *Sales* pada KimochiMart dengan menggunakan SAS Studio. Caranya adalah dengan menulis kodingan di bawah ini



**Gambar 4. Perhitungan MEAN Sales dengan SAS PROC MEAN pada SAS STUDIO**

****

**Gambar 5. Hasil output MEANS PROC pada KimochiMart**

Berdasarkan hasil tersebut bisa dilihat bahwa Mean dari Sales KimochiMart berkisar di angka 192.9875235, kemudian nilai MAXIMUM di angka 912.0000000, nilai MINIMUM di angka 17.0000000, Median di angka 138.0000000, dan standar deviasi di angka 151.1331271. Dari data tersebut bisa disimpulkan bahwa datanya ini sangat banyak dan bervariatif.

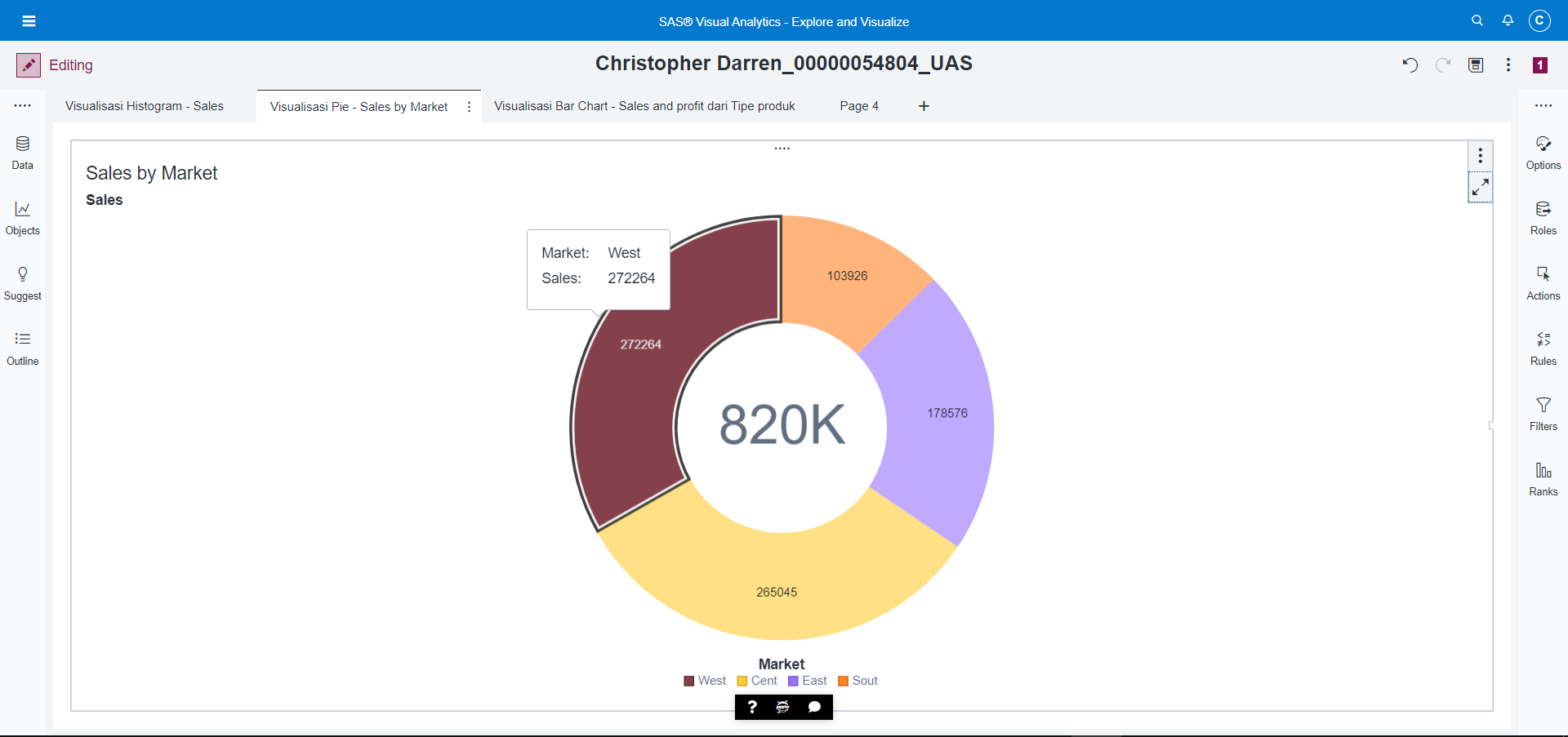
## EXPLORATORY DATA ANALYSIS

Peneliti juga tidak lupa melakukan exploratory data analysis dimana merupakan hal yang penting untuk dilakukan agar bisa mengetahui distribusi data dan memahami isi data dari KimochiMart ini terlihat seperti apa. Dengan melakukan EDA maka peneliti bisa mendapatkan sedikit insight dari data yang digunakan.

****

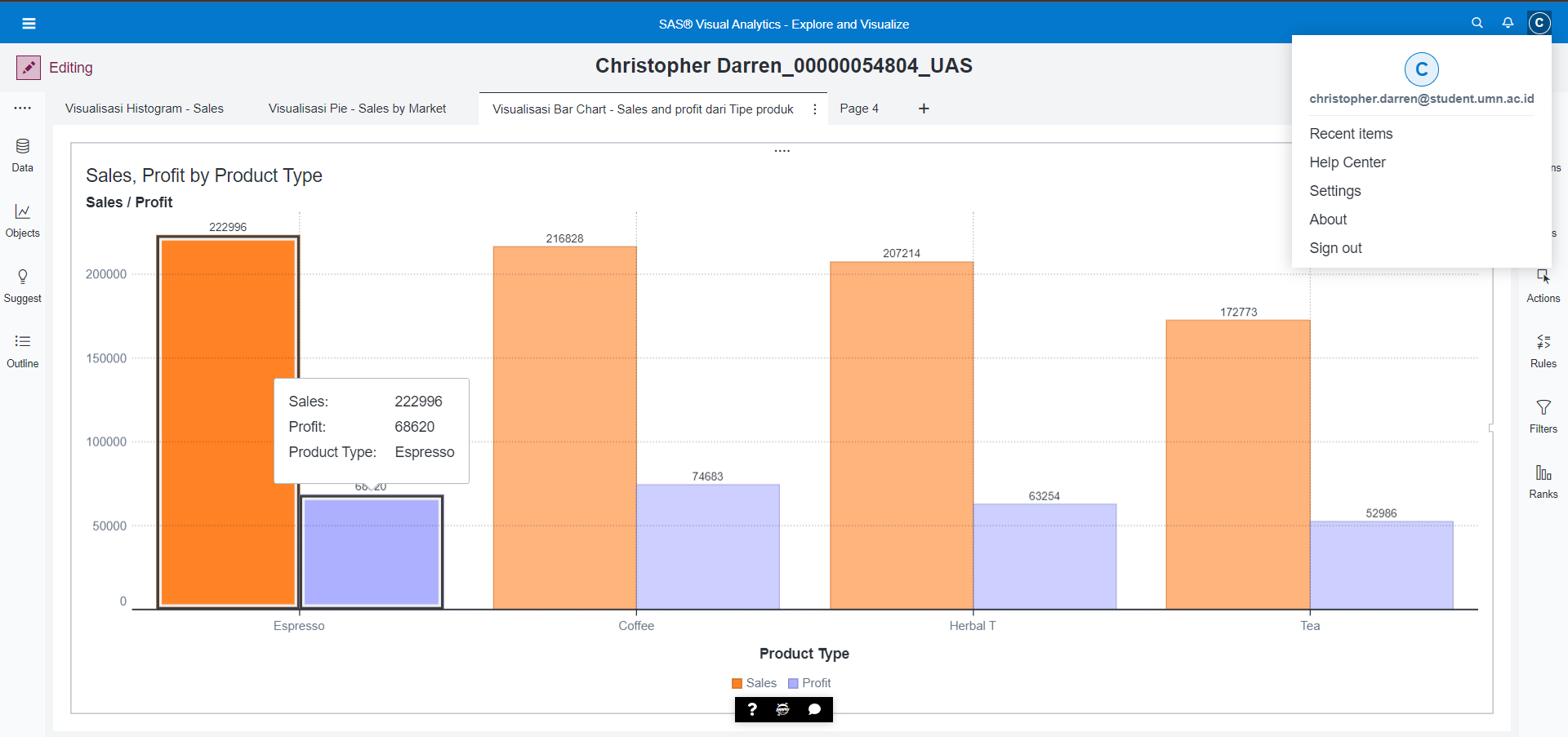
**Gambar 6. *Histogram* distribusi dari Sales pada KimochiMart**

Berdasarkan gambar tersebut terlihat bahwa *frequency* Sales dengan penjualan lower dan upper terbanyak frequency muncul sebanyak 921 kali, dalam artian baik *sales lower* dengan angka 80(dalam dollar) dan 120(dalam dollar)merupakan distribusi yang paling banyak dari antara semua sales yang terjadi pada KimochiMart.

****

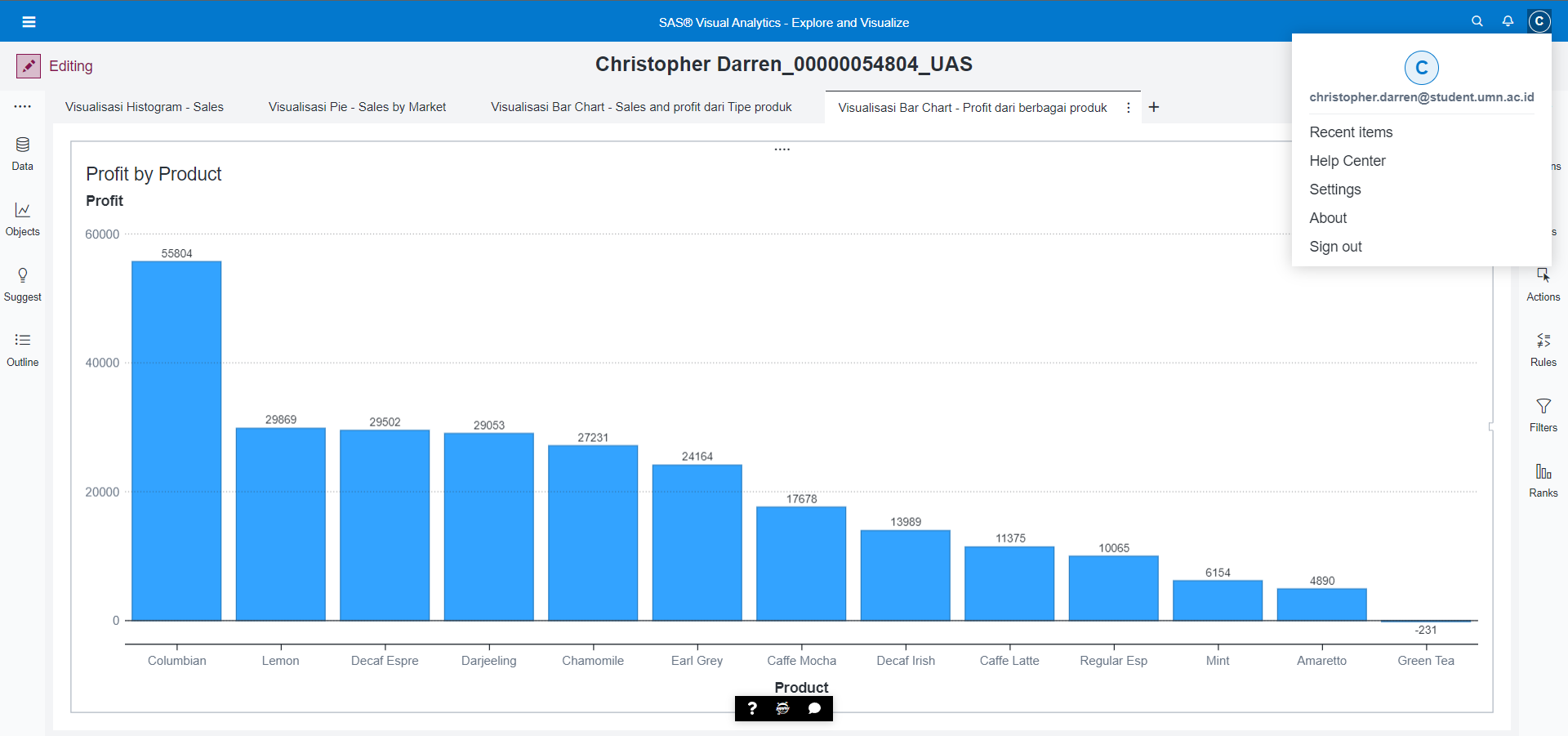
**Gambar 7. *Pie Chart* dari sales berdasarkan *region Market* pada KimochiMart**

Berdasarkan gambar tersebut total *Sales* pada market secara keseluruhan ada di angka 820.000, yang kemudian dipecah menjadi 4 region yakni West, Cent, East, dan South. Masing masing region tersebut mempunyai data sales tersendiri. Dari urutan Sales yang paling besar yakni dipegang oleh region *West* sebesar 272.264, diikuti urutan kedua oleh region *Cent* sebesar 265.045, urutan ketiga region *East* sebesar 178.576, dan urutan terakhir adalah region *South* sebesar 103.926.

****

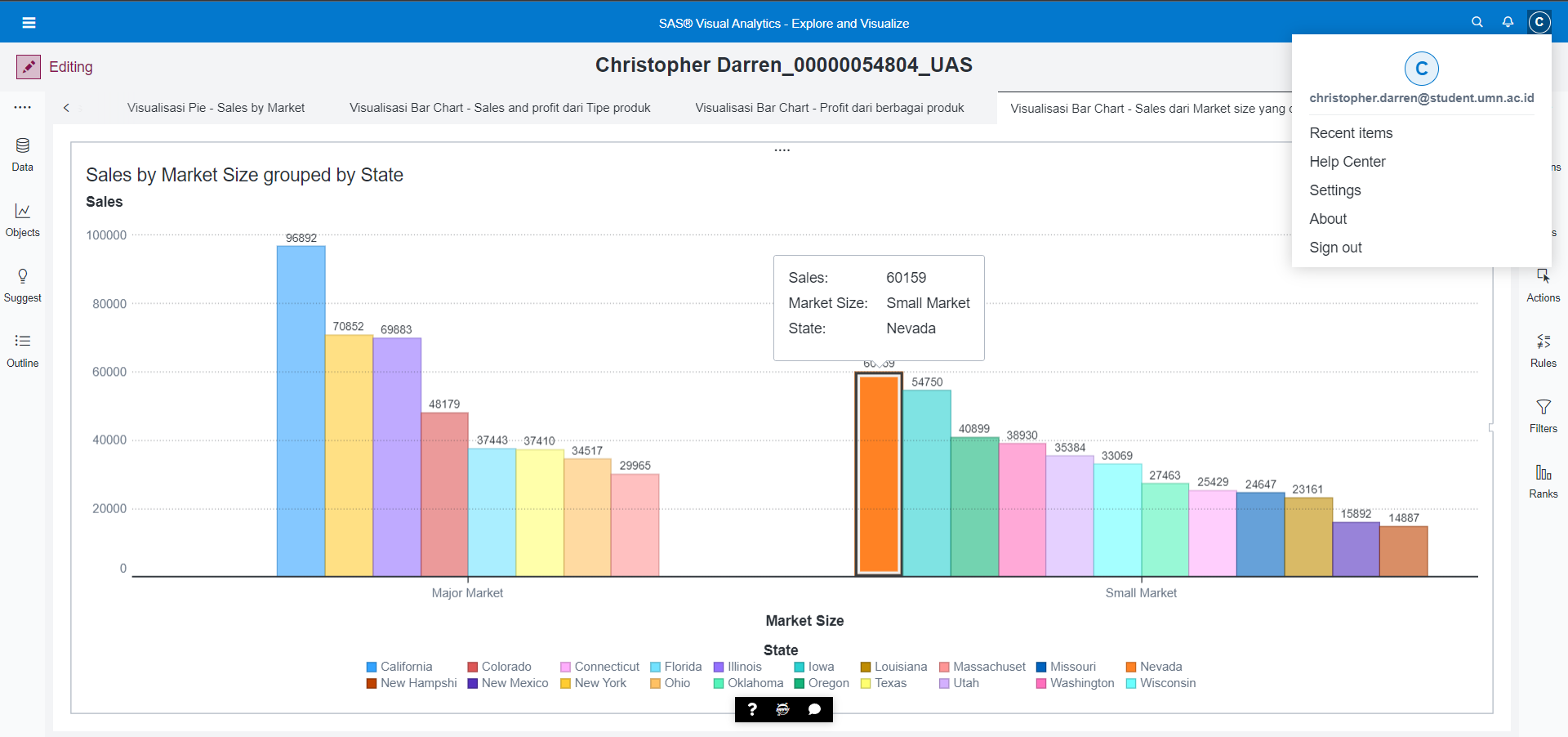
**Gambar 8. *Bar chart* dari sales dan profit berdasarkan tipe produk pada KimochiMart**

Berdasarkan gambar tersebut total Sales dari product type tertinggi diraih oleh tipe *Espresso* sebesar 222.996, urutan kedua diikuti oleh tipe *Coffee*, urutan ketiga yakni tipe *Herbal* *Tea,* dan urutan keempat adalah tipe *Tea.* Namun dalam sebuah bisnis ritel perlu melihat profit dari sales juga. Sales tinggi namun profit rendah sama saja dengan rugi maupun tidak sesuai ekspektasi. Pada gambar tersebut *profit* yang paling besar dari antara keempat tipe produk tersebut adalah tipe *Coffee* sebesar 74.683 meski Salesnya tidak sebanyak tipe *Espresso* dan yang paling rendah terdapat pada tipe *Tea* yang hanya sebesar 52.986.

****

**Gambar 9. *Bar Chart* dari profit berdasarkan produk yang terjual pada KimochiMart**

Berdasarkan gambar tersebut total profit dalam dollar, paling besar ada di produk *Columbian* sebesar 55.804(dalam dollar), diikuti urutan kedua oleh Lemon sebesar 29.869, urutan ketiga ada *Decaf Espresso* sebesar 29.053(dalam dollar). Kemudian ada 1 produk yang profitnya merugikan yakni *Green Tea* -231 yang berarti customer tidak suka dengan produk *Green Tea* dan alangkah baiknya difokuskan ke produk yang lebih membuat *profit*.

****

**Gambar 10. *Bar Chart* dari Sales Market size yang di grup berdasarkan lokasi state**

Berdasarkan gambar tersebut secara garis besar Market size *Major Market* mendominasi total sales dengan list state: California, New York, Illinois, Colorado, Florida, Texas, Ohio, dan Massachuset. Dengan total sales paling tinggi berada di state California, major market. Kemudian di sisi lain diikuti oleh Market size *Small Market*. Pada Small Market, total sales paling banyak ada di state Nevada sebesar 60.159, diikuti oleh Iowa sebesar 54.750. Adapun ada hal menarik dari gambar tersebut karena pada segmen *Small Market* ,beberapa mendekati total sales dari market *Major Market* contohnya Iowa yang bisa mengalahkan state Colorado, dan perbedaan juga lumayan meski beda ribuan dollar saja.

## Data Preparation

Pada tahap ini peneliti melakukan preprocessing data, pada data KimochiMart peneliti melakukan 2 cara yakni menangani missing values, dan mengecek duplikasi data.

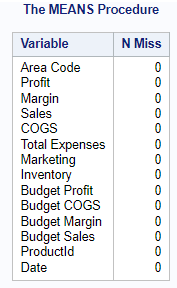
1. Menangani Missing Values

****

**Gambar 11. Code untuk menjalankan missing values**

Peneliti ingin tahu apakah data KimochiMart terdapat missing value ataupun tidak, supaya tidak mengganggu proses pada pembuatan model. Adapun caranya adalah digambar atas(gambar 11).

Bila di run maka hasil SAS studionya adalah sebagai berikut:

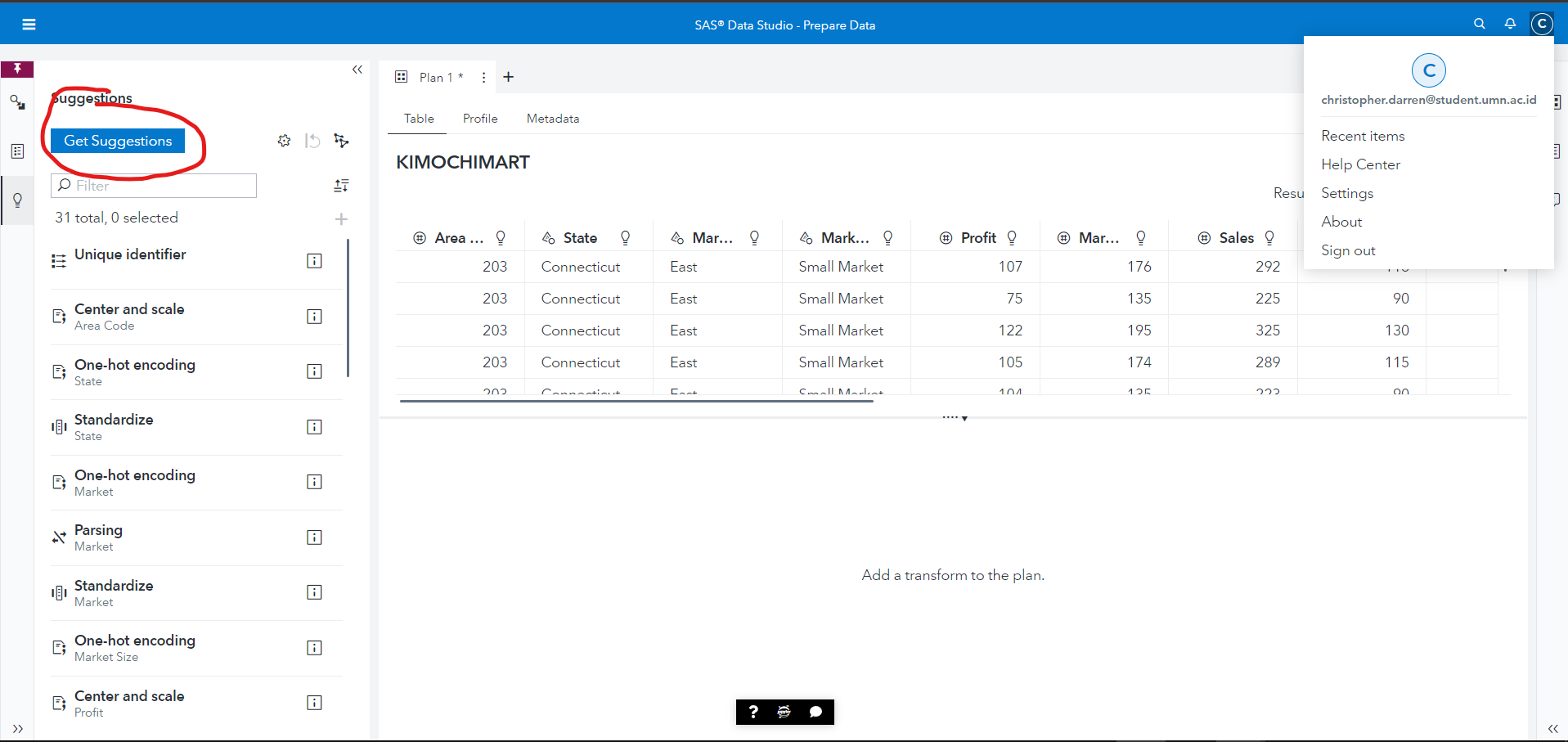
****

**Gambar 12. Hasil PROC MEANS NMiss untuk *missing value***

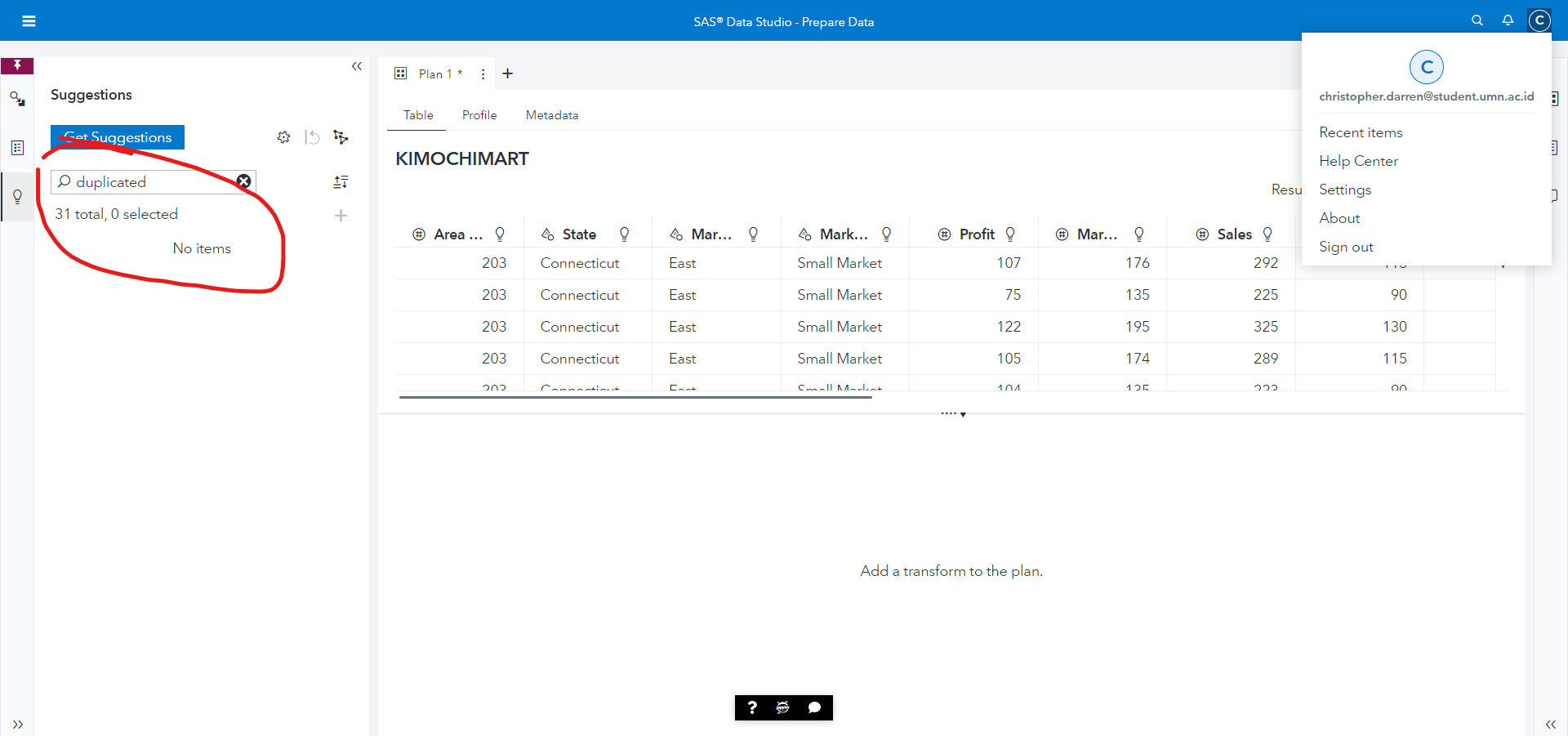
Hasil dari PROC means menggambarkan bahwa pada data KimochiMart tidak memiliki missing values atau N Miss sehingga data ini siap dipakai untuk keperluan modeling.

1. Menangani duplikasi data

Setelah peneliti tidak menemukan missing value pada tahap *finding missing values*, langkah selanjutnya adalah untuk memeriksa apakah ada *duplicate data.* Duplikasi data bisa terjadi karena kesalahan manusia dalam menginput data ataupun ada kesalahan dari sistem yang tidak bisa terhindarkan meski dalam menggunakan teknologi secanggih pun. Adapun peneliti ingin melihat apakah data Kimochi Mart memiliki duplikasi data atau tidak. Adapun caranya adalah sebagai berikut.

****

**Gambar 13. Get suggestions untuk mencari duplicate data**

****

**Gambar 14. Get suggestions tidak menemukan suggestion untuk remove duplicate data pada data KimochiMart.**

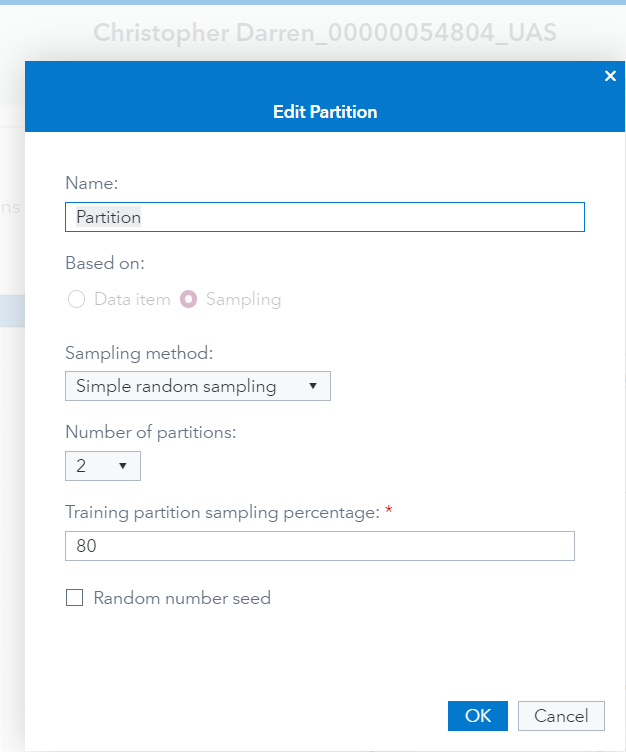
Namun setelah mengklik tombol *get suggestions* tidak terdeteksi adanya duplikasi data sehingga bisa dikatakan data ini sudah aman dan bisa dilanjutkan menuju proses modeling. Secara garis besar data KimochiMart tidak ada duplikasi data sehingga peneliti bisa lanjut ke proses berikutnya.

# 2. Question 2: CLO-3 Sub-CLO-11, CLO-2 Sub-CLO-12 and CLO-1 Sub-CLO-14 Weight (60 %)

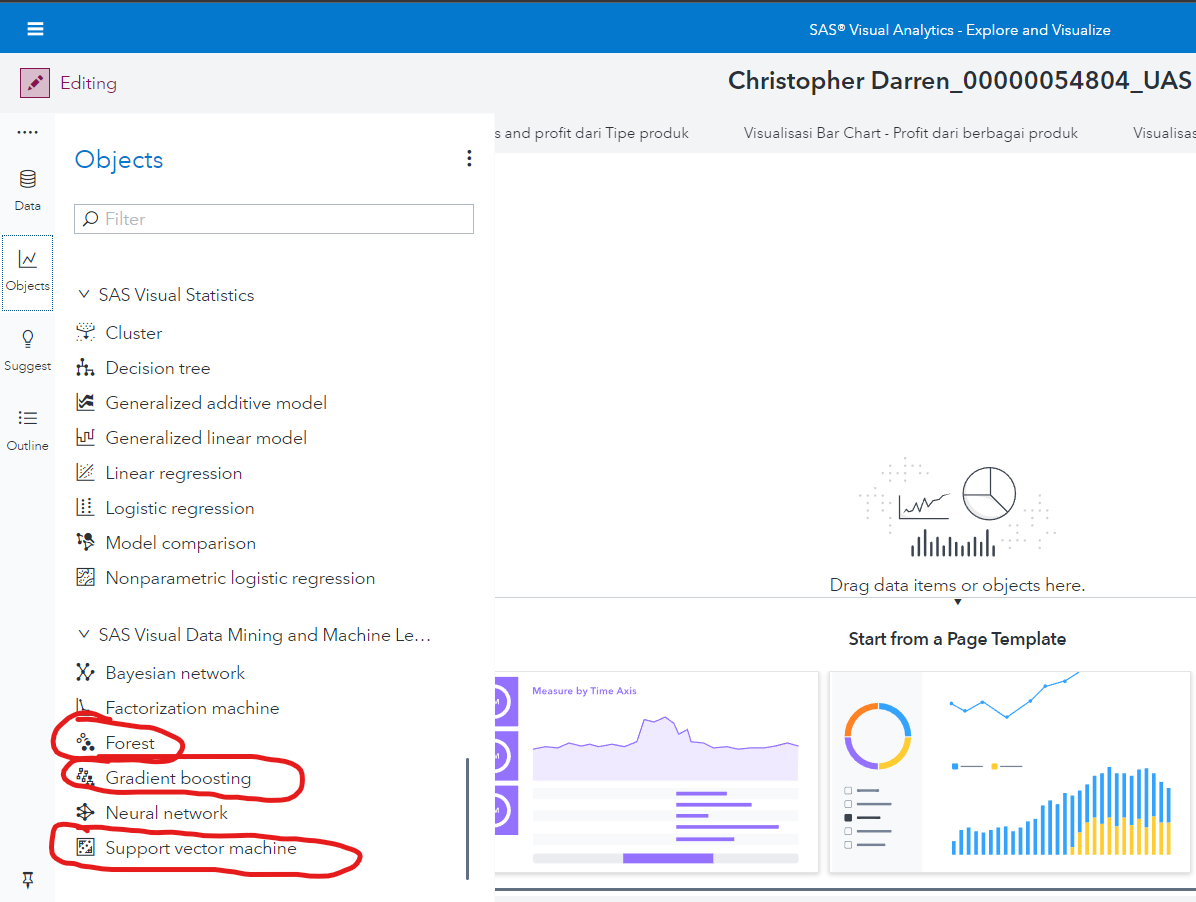
* **Build and manage your Analytics model using SAS® Software.**
* **The journal writing in section IV contains an explanation of Modeling your data .**
* **Create your Analytical Model by design and concept of SAS® VDMML with Forest or SVM algorithms with comparison of other algorithms.**

## Analytics Model

Peneliti kemudian menggunakan beberapa teknik *machine learning* untuk mengolah data KimochiMart seperti yang sudah dijelaskan pada latar belakang yakni dengan menggunakan *support vector machine, random forest ,* dan gradient boosting yang akan dibahas pada BAB IV HASIL DAN PEMBAHASAN menggunakan SAS VIYA analytics. Peneliti nantinya juga menggunakan variable standar *testing* *split data* yakni 80:20, terutama dalam membuat model prediksi nantinya. Untuk cara pembagian split data bisa lihat gambar dibawah ini.

****

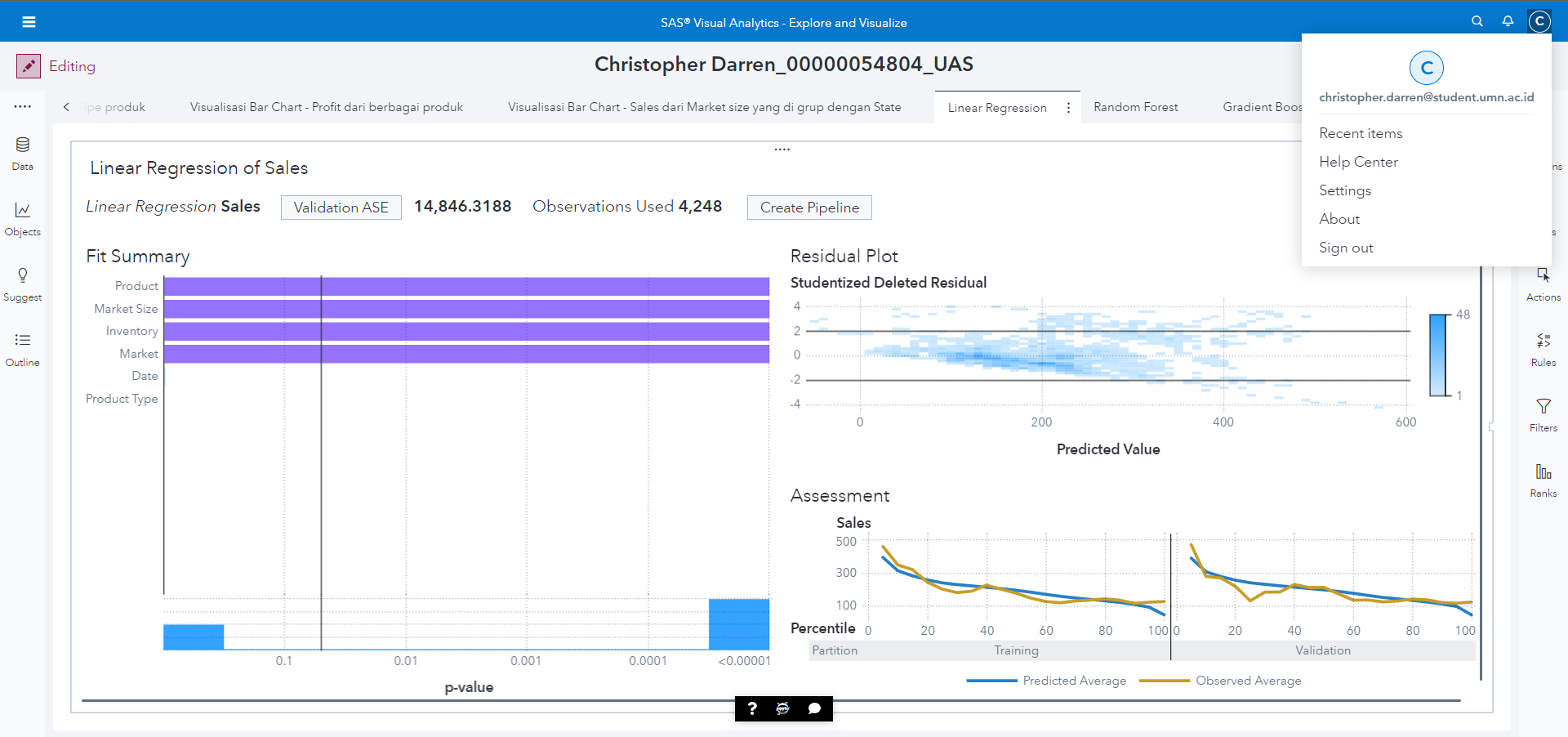
**Gambar 15. Proses splitting data 80:20**

****

**Gambar 16. Tampilan utama SAS VIYA untuk membuat modeling.**

Gambar diatas merupakan tampilan ketika peneliti ingin membuat modelling. Terlihat dengan jelas bahwapeneliti menggunakan 3 algoritma yakni forest, gradient boosting, dan support vector machine.

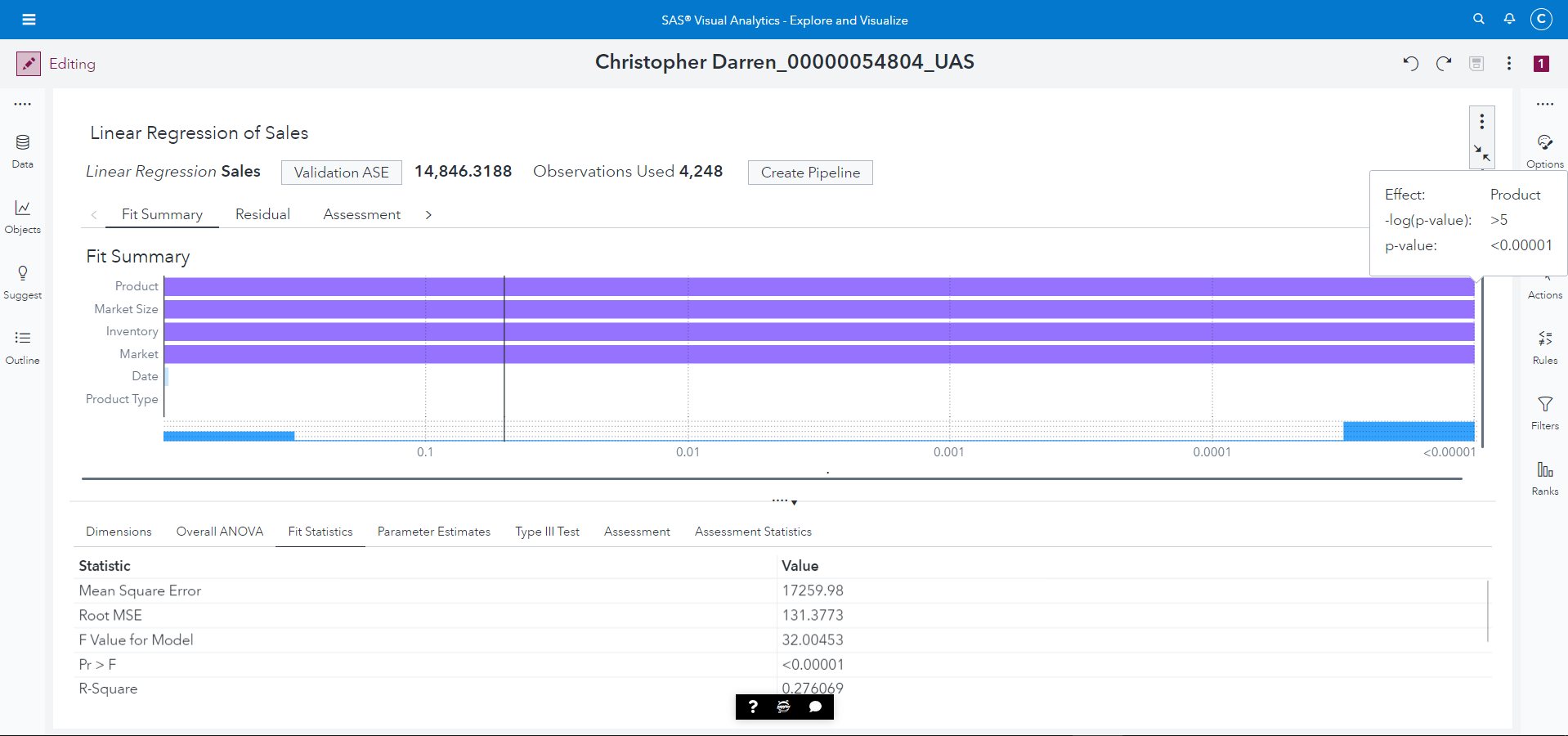
### Linear Regression

****

**Gambar 17. Visualisasi model Linear regression terhadap data KimochiMart**

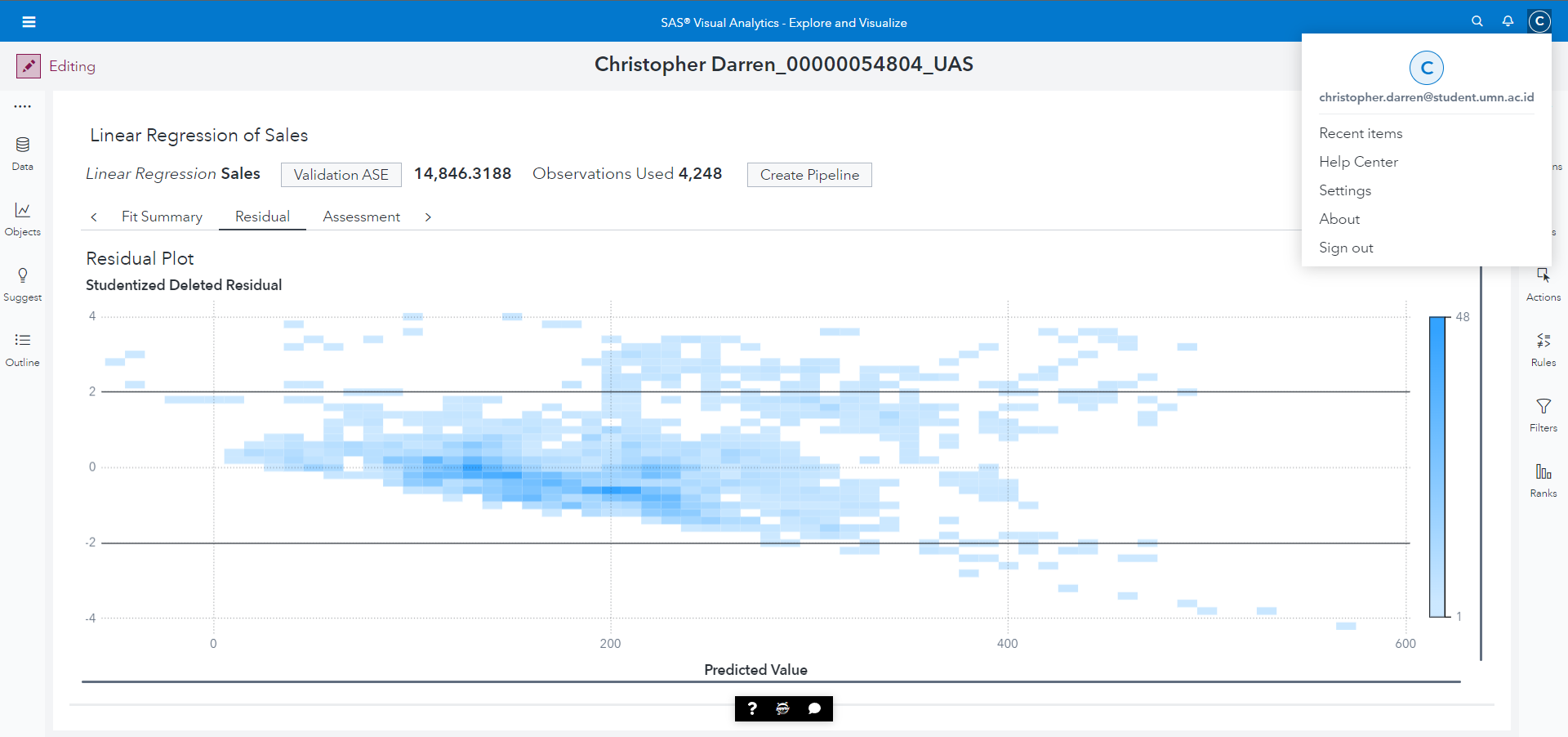
Berdasarkan hasil model linear regression pada gambar diatas dapat dilihat bahwa *total observation*, yakni data yang digunakan adalah sebanyak 4,248 records. Untuk *Response* yang digunakan adalah continuous yakni *Sales*, *Continuous Effects* yang digunakan adalah *Inventory*. Kemudian untuk classification yang digunakan adalah *Market Size, Product, Market, Date*, dan *Product Type*. Untuk hasil validation ASE nya sendiri adalah 14,846.3188.

Peneliti tidak lupa juga melakukan analisis mendalam dari 3 visualisasi yang berbeda.

****

**Gambar 18. Visualisasi dari fits summary linear regression terhadap data KimochiMart**

Berdasarkan gambar diatas, *fit summary* merupakan ringkasan dari model statistik yang sudah dibuat menggunakan linear regression. Terlihat bahwa hasil dari model tersebut terutama Pr > F adalah < 0.00001 yang berarti nilainya signifikan dan terdapat hubungan dari beberapa prediktor diatas yakni *Product, Inventory, Market Size,* dan *Market*. Tidak hanya itu jika diperhatikan pada bar yang berwarna ungu yang merepresentasikan prediktor memiliki p-value 0.00001 yang berarti mempunyai peran signifikan terhadap variabel target dalam hal ini target *response* nya adalah *Sales*.

****

**Gambar 19. Visualisasi residual plot dari linear regression terhadap data KimochiMart**

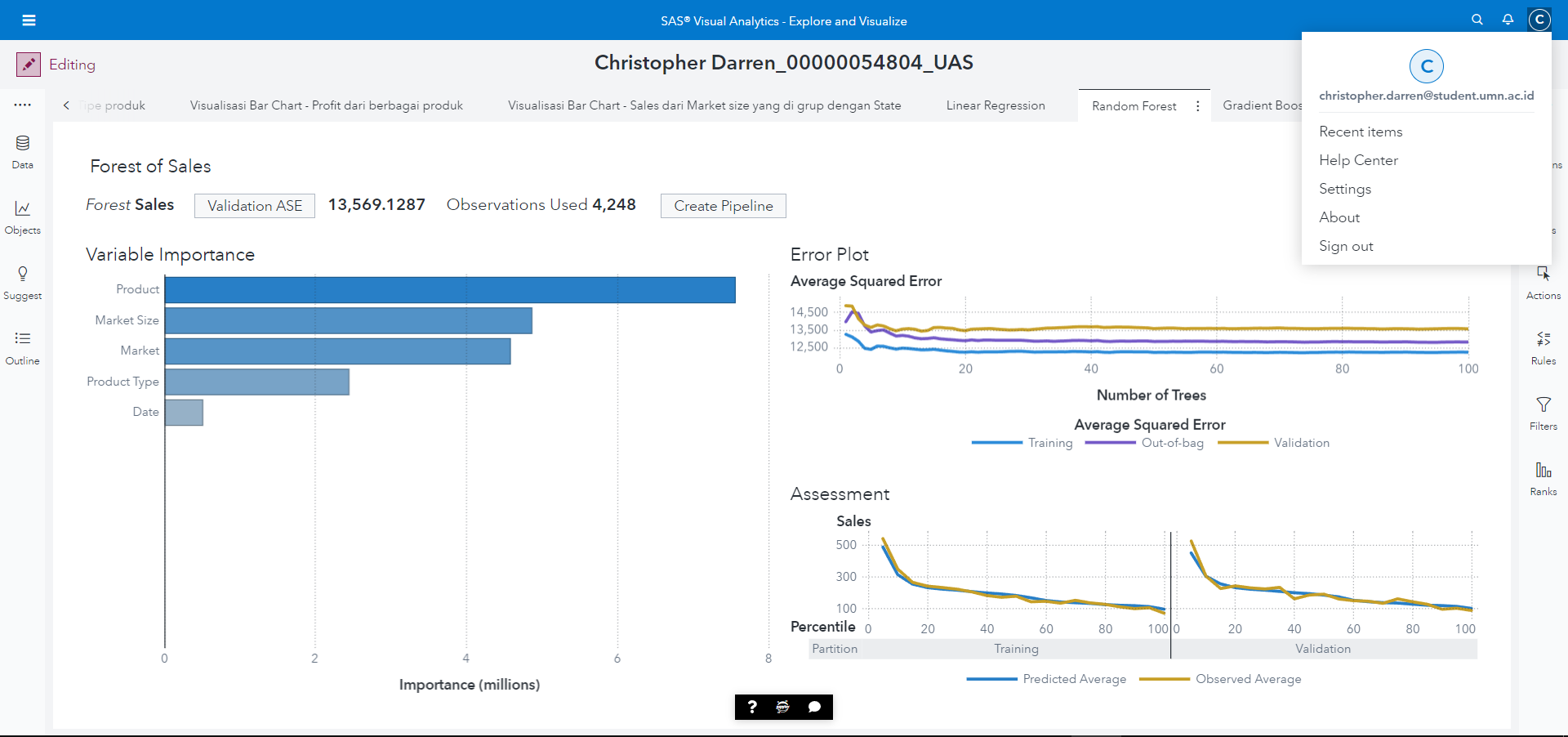
Berdasarkan gambar diatas, terlihat sebuah *residual plot.* *Residual plot* adalah sebuah nilai selisih antara value asli dan value prediksi dalam hal ini residual plot menggambarkan apakah model yang dibuat ini baik atau buruk. Data KimochiMart ketika dibuat model terlihat ada beberapa outlier yang berwarna biru muda pudar di bagian ujung. Apakah dikatakan buruk? Peneliti menjawab tidak begitu buruk karena data yang mengumpul juga masih banyak berada di *range* 0 axis y sampai 1. Untuk hasil prediksi dari hasil residual ini akan peneliti bahas pada gambar dibawah ini.

****

**Gambar 20. Visualisasi assessment plot untuk linear regression terhadap data KimochiMart**

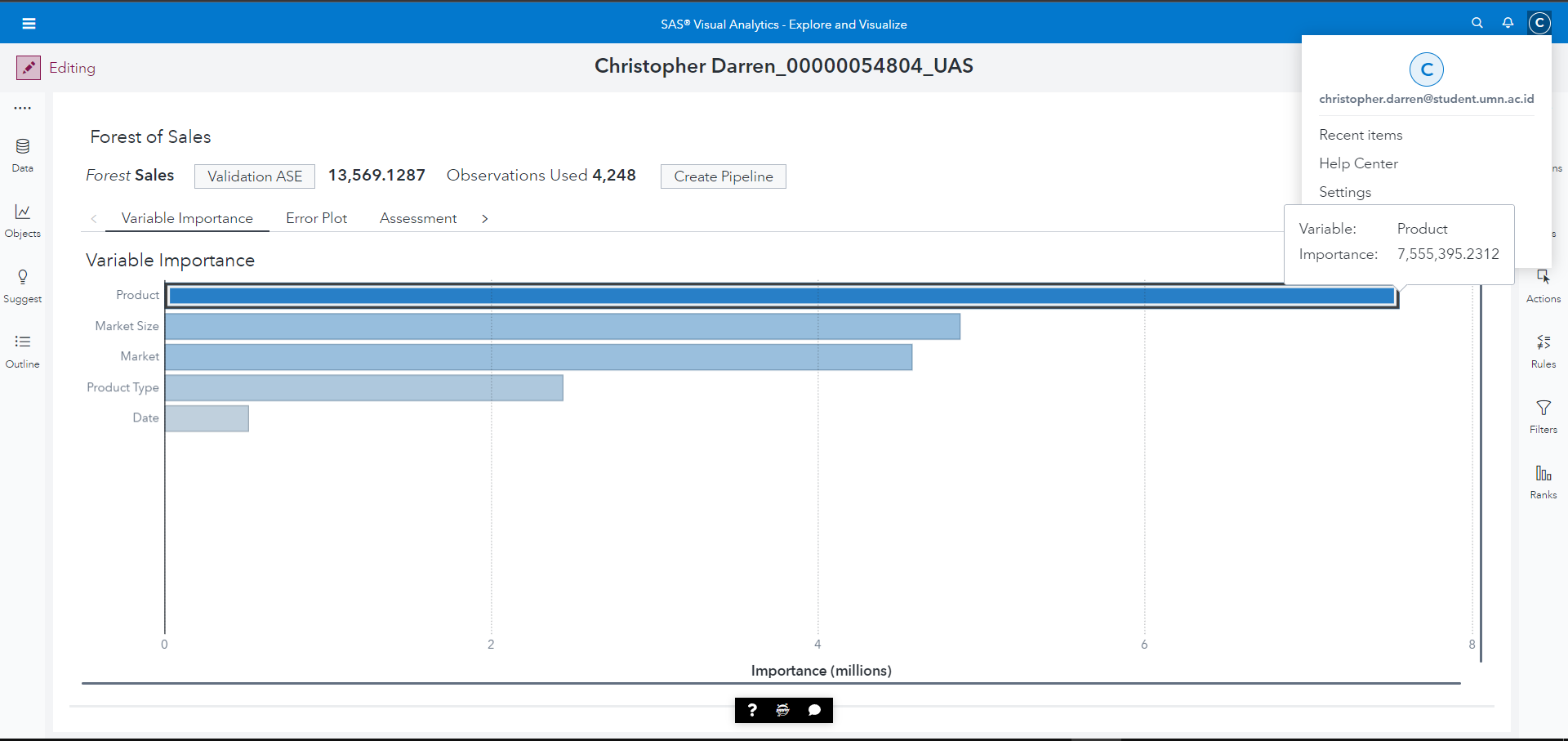
Berdasarkan gambar diatas terdapat 2 *assessment plot* yakni *training* dan *validation plot*. Peneliti melakukan hal demikian agar hanya tidak terpaku terhadap 1 model saja namun perlu pembuktian maka peneliti membuat *validation* juga setelah proses *training* pembuatan model selesai. Terlihat bahwa ada garis dengan legenda *predicted average* dan *observed average*. Pada plot training, memang prediksinya dimulai dengan kemiripan dengan observation plot namun semakin bergerak ke bawah ada beberapa perbedaan yakni observasinya yang tidak mengikuti prediksi pada plot *training* , bisa dibilang tidak terlalu buruk karena masih bisa mengikuti garis prediksi, namun ketika melihat dari sisi lain yakni *validation* model bisa dibilang bagus karena observed averagenya mengikuti garis predicted average, walaupun awal validation 20% rada sedikit jatuh untuk *observed average*nya, tapi seiring berjalannya model maka prediksi dan observed garisnya mendekati bahkan bertabrakan sehingga bisa dikatakan prediksinya berhasil dan kesalahan prediksi masih kecil.

### Random Forest

****

**Gambar 21. Visualisasi model random forest terhadap data KimochiMart**

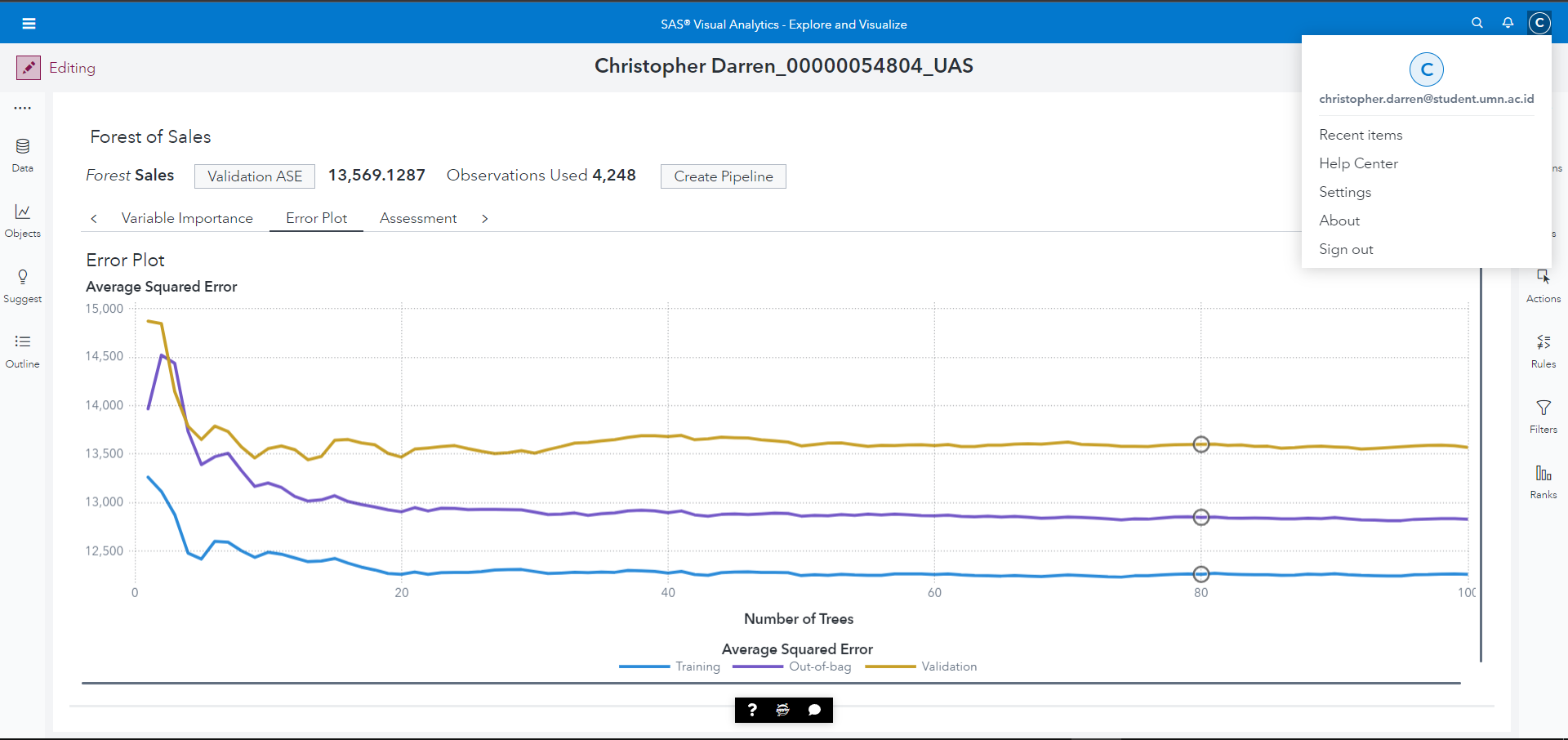
Berdasarkan hasil model random forest pada gambar diatas dapat dilihat bahwa total observation, yakni data yang digunakan adalah sebanyak 4,248 records. Untuk *Response* yang digunakan adalah continuous yakni *Sales*. Kemudian untuk classification yang digunakan adalah *Market Size, Product, Market, Date*, dan *Product Type*. Untuk hasil validation ASE nya sendiri adalah 13,569.1287. Sama seperti model sebelumnya, pada bagian random forest model, peneliti akan melakukan deep dive hasil model yang sudah dibuat pada gambar 21. Berikut penjelasan visualisasi masing masing dari random forest:

****

**Gambar 22. Visualisasi dari variable importance random forest terhadap data KimochiMart**

Berdasarkan gambar diatas, terlihat sebuah bar kumpulan *variable importance* dari data KimochiMart. *Variable importance* yang paling berperan besar dalam memainkan peran prediksi response *Sales* yakni *Product*. Urutan berikutnya yakni Market Size -> Market -> Product Type dan terakhir adalah Date. Dengan demikian hasil prediksi tergantung seberapa penting variable importance yang terpilih sebagai prediktor.

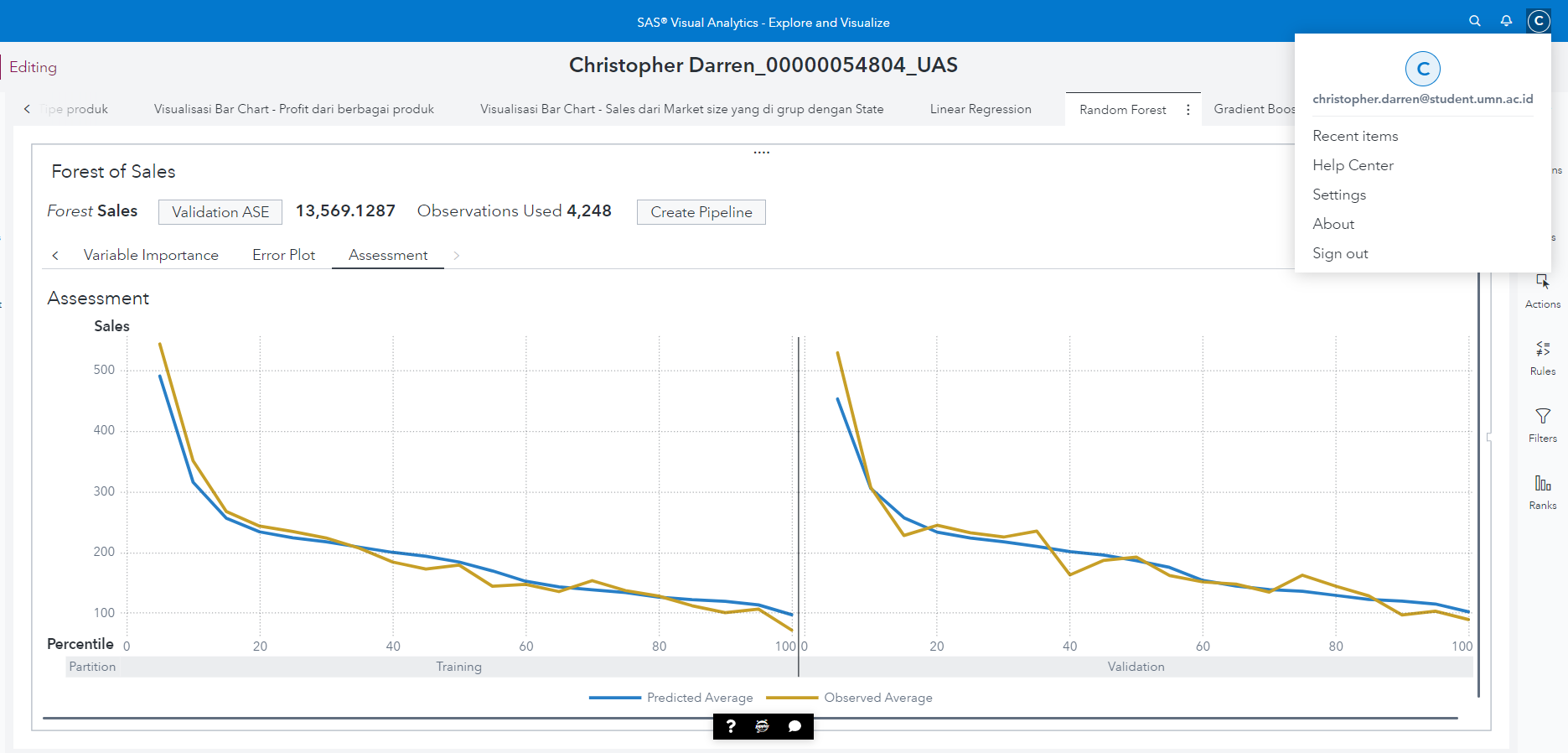
Kemudian peneliti juga membuat visualisasi sebagai berikut.

****

**Gambar 23. Visualisasi dari error plot random forest terhadap data KimochiMart**

Berdasarkan gambar diatas, terlihat sebuah *error plot*. Error plot singkatnya untuk melihat seberapa jauh error yang dihasilkan dari jumlah tree yang dihasilkan , terutama pada *training model* dan *validation model.* Pada data KimochiMart yang sudah dilakukan training dan validation terlihat bahwa pembuatan model awalnya sangat jelek dan garisnya bergerigi terutama dari garis *training*(garis biru), *garis out-of-bag*(garis ungu), dan *garis validation*(garis kuning) namun seiringnya model berjalan maka garisnya bisa dilihat lurus dan kesalahan sudah ditangani dengan baik karena sudah ada penambahan pohon dari ensemble model, sehingga kesalahan sudah hilang.

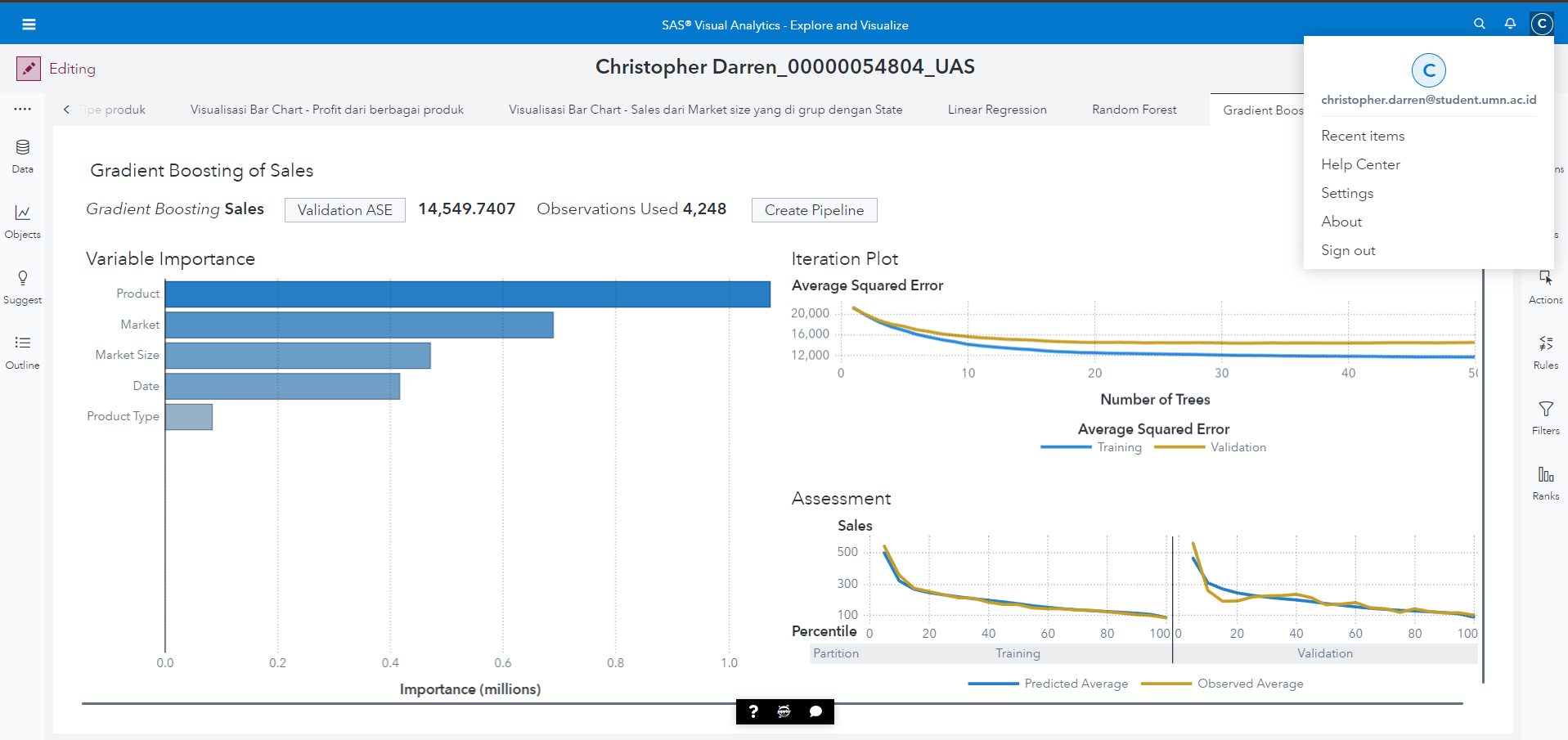
Berikutnya peneliti juga melihat dari assessment dibawah ini.

****

**Gambar 24. Visualisasi dari assessment plot random forest terhadap data KimochiMart**

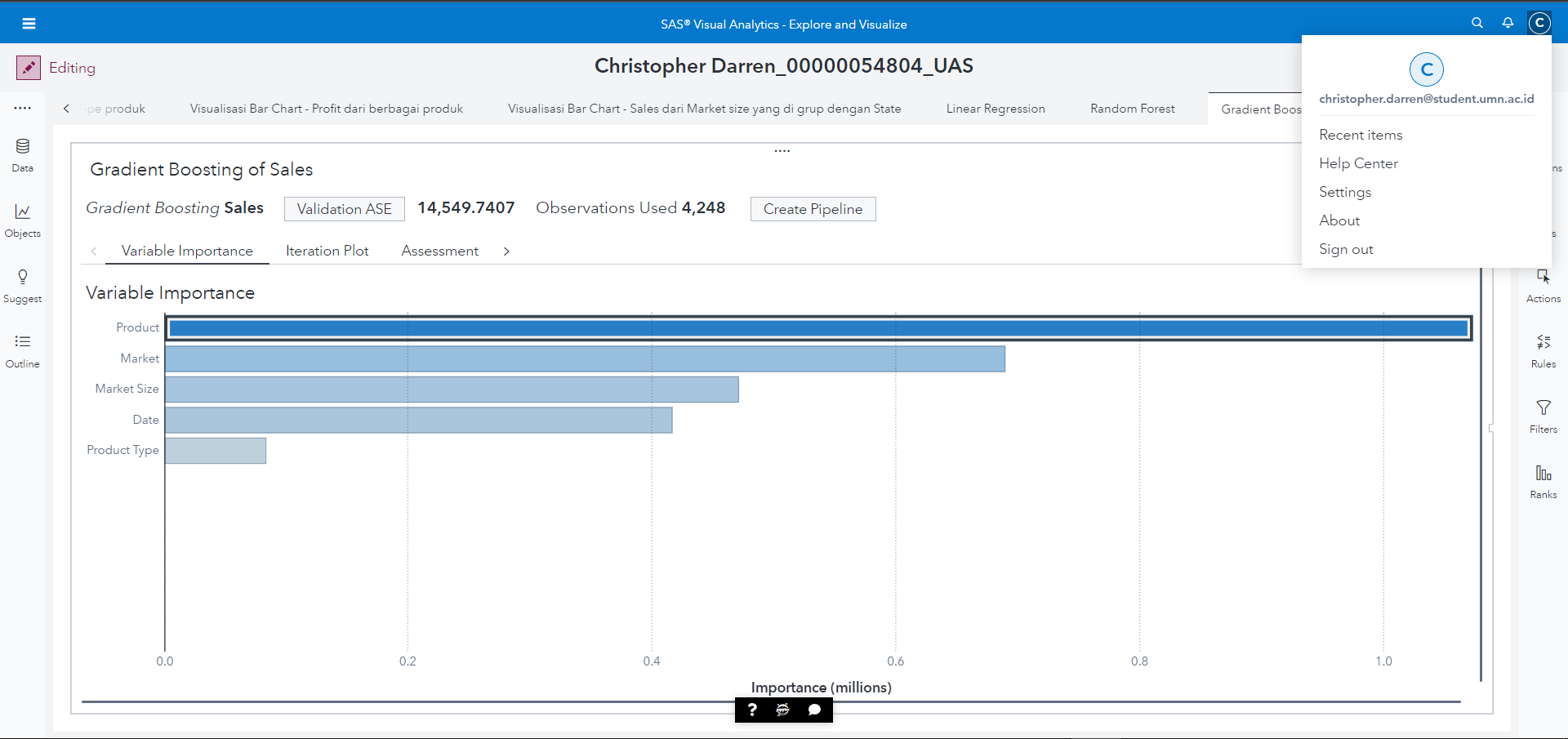
Berdasarkan gambar diatas, terlihat sebuah *assessment* plot yakni *training plot* dan *validation* *plot* dari model random forest yang sudah dibuat. Peneliti melakukan hal demikian sama seperti pada pembuatan *linear regression model* agar hanya tidak terpaku terhadap 1 model saja namun perlu pembuktian maka peneliti membuat *validation* juga setelah proses *training* pembuatan model selesai. Hasilnya adalah baik training dan validation perbedaannya tidak begitu jauh bahkan peneliti bisa bilang hasilnya mirip sehingga kesalahan dalam prediksinya jauh lebih kecil dari model sebelumnya(linear regression).

### Gradient Boosting

****

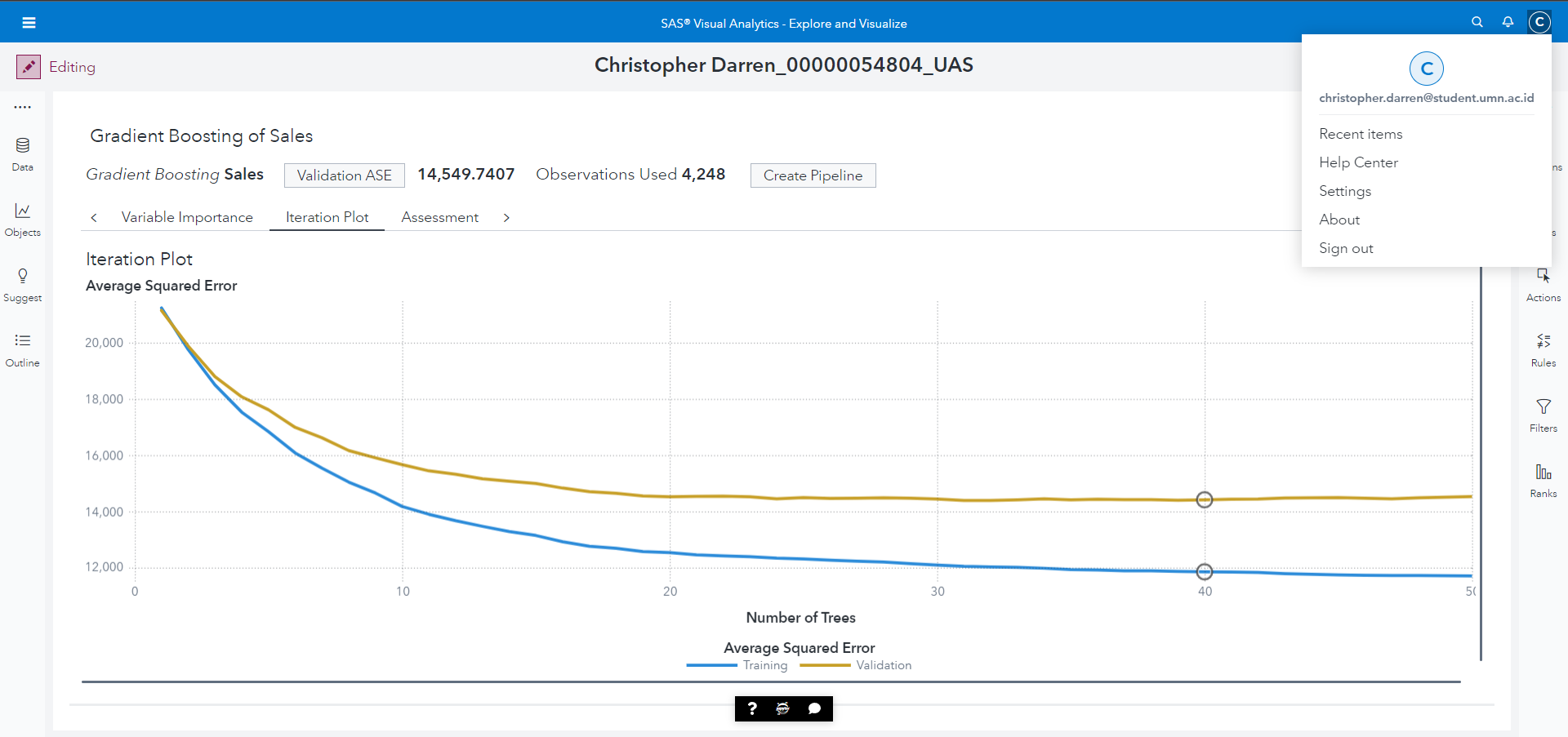
**Gambar 25. Visualisasi model gradient boosting terhadap data KimochiMart.**

Berdasarkan hasil model gradient boosting pada gambar diatas dapat dilihat bahwa total observation, yakni data yang digunakan adalah sebanyak 4,248 records. Untuk *Response* yang digunakan adalah continuous yakni *Sales*. Kemudian untuk classification yang digunakan adalah *Market Size, Product, Market, Date*, dan *Product Type*. Untuk hasil validation ASE nya sendiri adalah 14,549.7407. Sama seperti model sebelumnya, pada bagian random forest model, peneliti akan melakukan deep dive hasil model yang sudah dibuat pada gambar 25. Peneliti sama seperti sebelumnya melakukan analisa lebih dalam dari beberapa visualisasi yang sudah dibuat oleh gradient boosting.

****

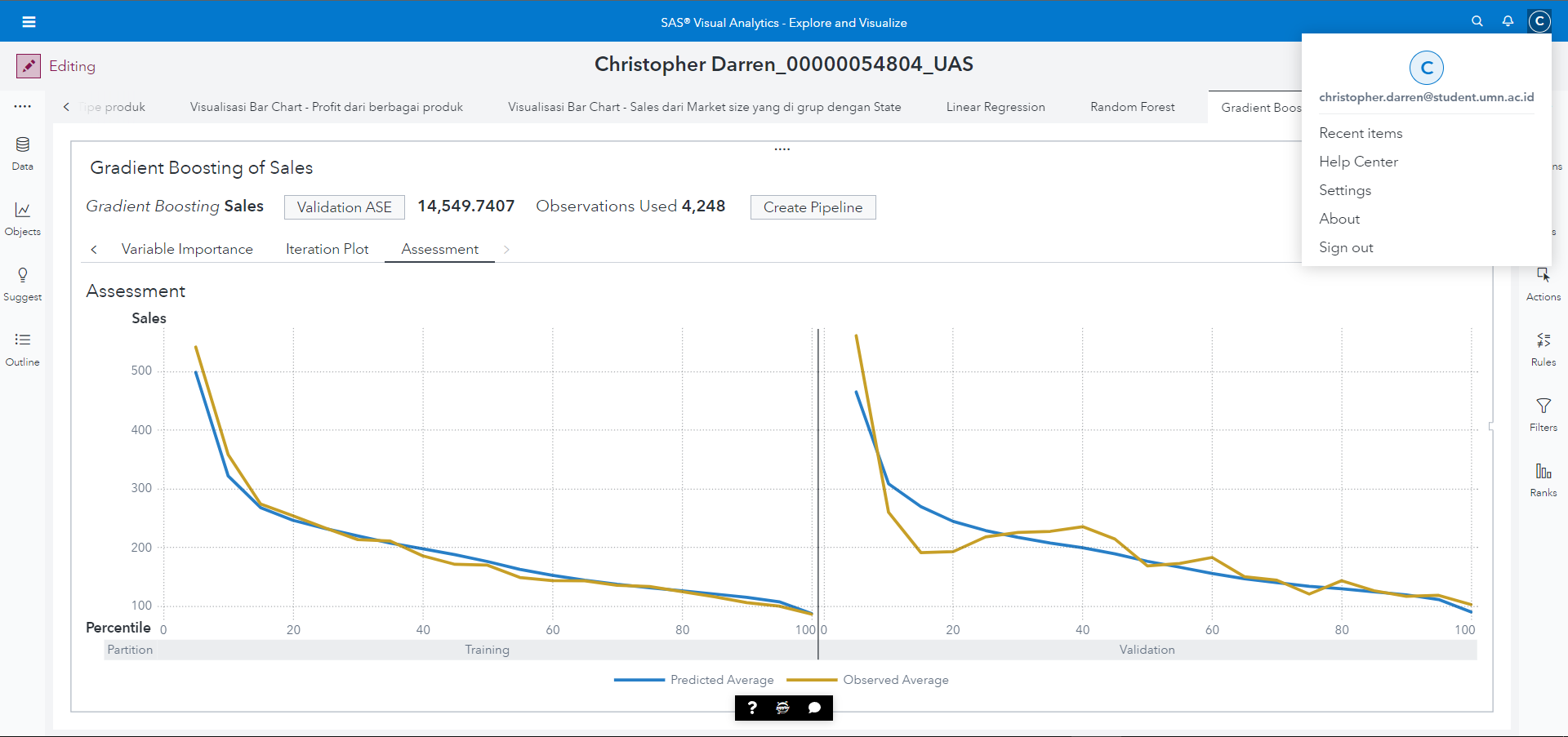
**Gambar 26. Visualisasi dari *variable importance* gradient boosting terhadap data KimochiMart**

Berdasarkan gambar diatas, terlihat sebuah bar kumpulan *variable importance* dari model gradient boosting data KimochiMart. *Variable importance* yang paling berperan besar dalam memainkan peran prediksi response *Sales* yakni *Product*. Urutan berikutnya yakni *Market* -> *Market Size* -> *Date* dan terakhir adalah *Product Type*. Dengan demikian hasil prediksi tergantung seberapa penting variable importance yang terpilih sebagai prediktor.

****

**Gambar 27. Visualisasi dari Iteration Plot gradient boosting terhadap data KimochiMart**

Berdasarkan gambar diatas, terlihat sebuah *Iteration plot.* *Iteration plot* singkatnya adalah meningkatkan iterasi agar bisa berjalan dengan optimal dan mencari model yang paling baik. Kesimpulan dari iteration plot data KimochiMart ini adalah model *training* dan validationnya dalam membuat iterasinya bagus sehingga membuat sebuah prediksinya tepat dan akurat.

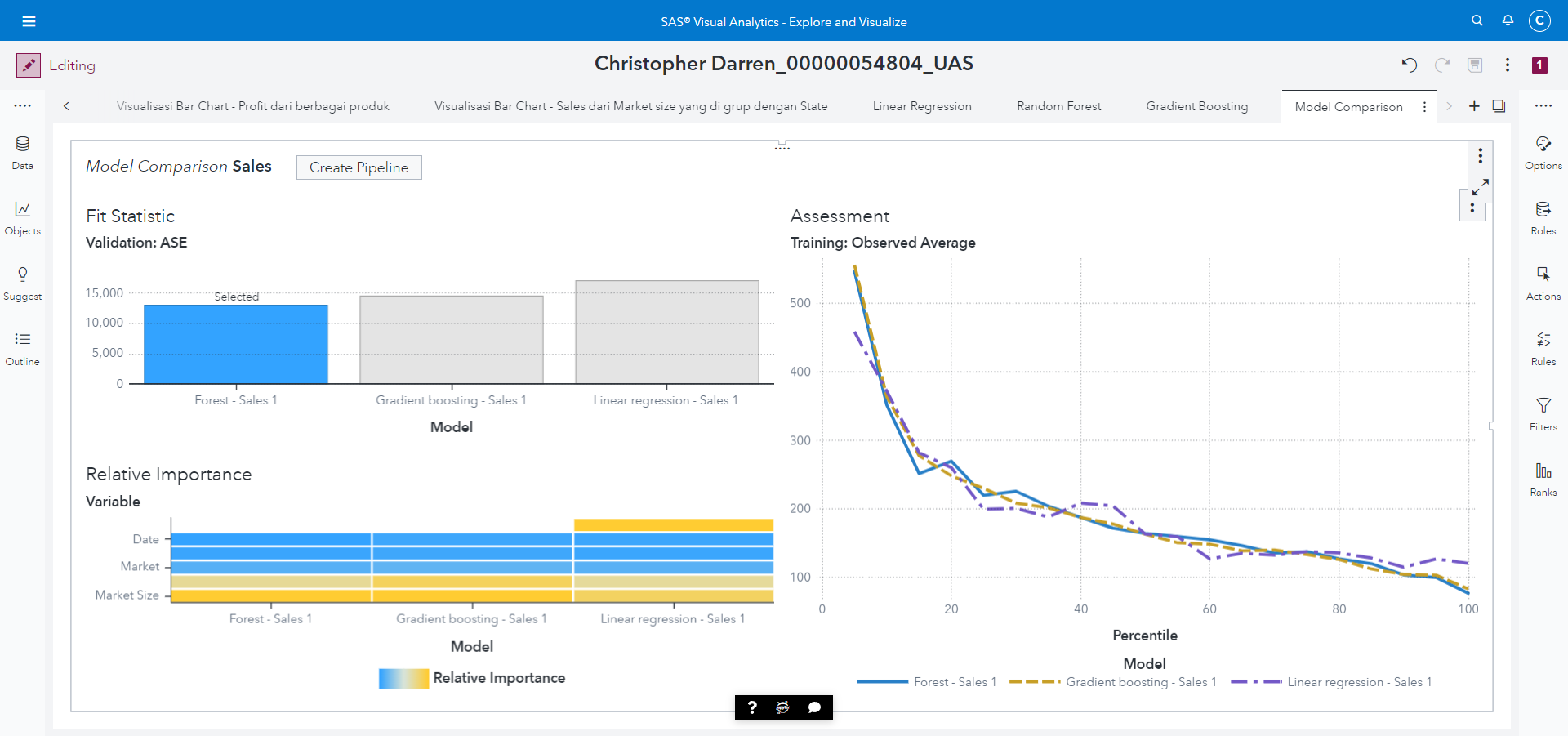
****

**Gambar 28. Visualisasi dari assessment plot gradient boosting terhadap data KimochiMart**

Berdasarkan gambar diatas, terlihat sebuah assessment plot yang memiliki 2 pengerjaan yakni melewati *training* dan *validation*. Garis biru yang merepresentasikan *predicted Average* dan garis kuning yang merepresentasikan *observed average*. Hasil dari training membuktikan bahwa hasil prediksi dan observation nya baik dan rapat satu sama lainnya dalam artian bahwa modelnya bagus. Di sisi lain *validation*  pada saat dijalankan seiring waktu hasilnya juga makin membaik dan bisa disimpulkan bahwa prediksinya berjalan sesuai dengan *observed average*.

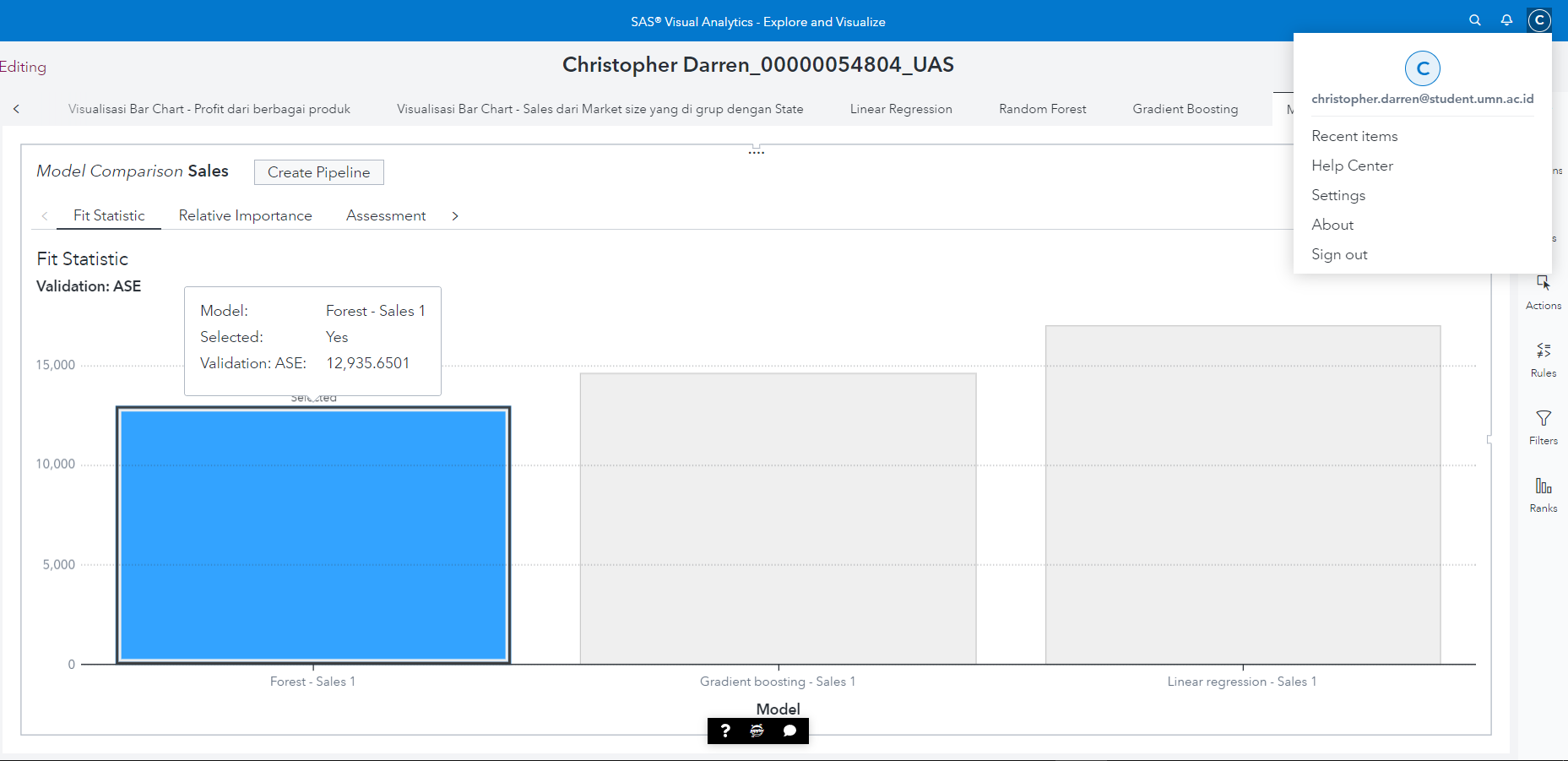
## Evaluation

Setelah peneliti membuat 3 model yang berbeda maka perlu dilakukan sebuah hasil evaluation guna memilih model paling baik Model Comparison. Pembuatan Model comparison bisa dilakukan dengan cara mendrag model comparison menuju canvas pekerjaan SAS VIYA, bila sudah di drag maka akan otomatis muncul sebuah pop untuk pilihan 3 algoritma, bila sudah di select maka secara otomatis akan membuat modelnya.



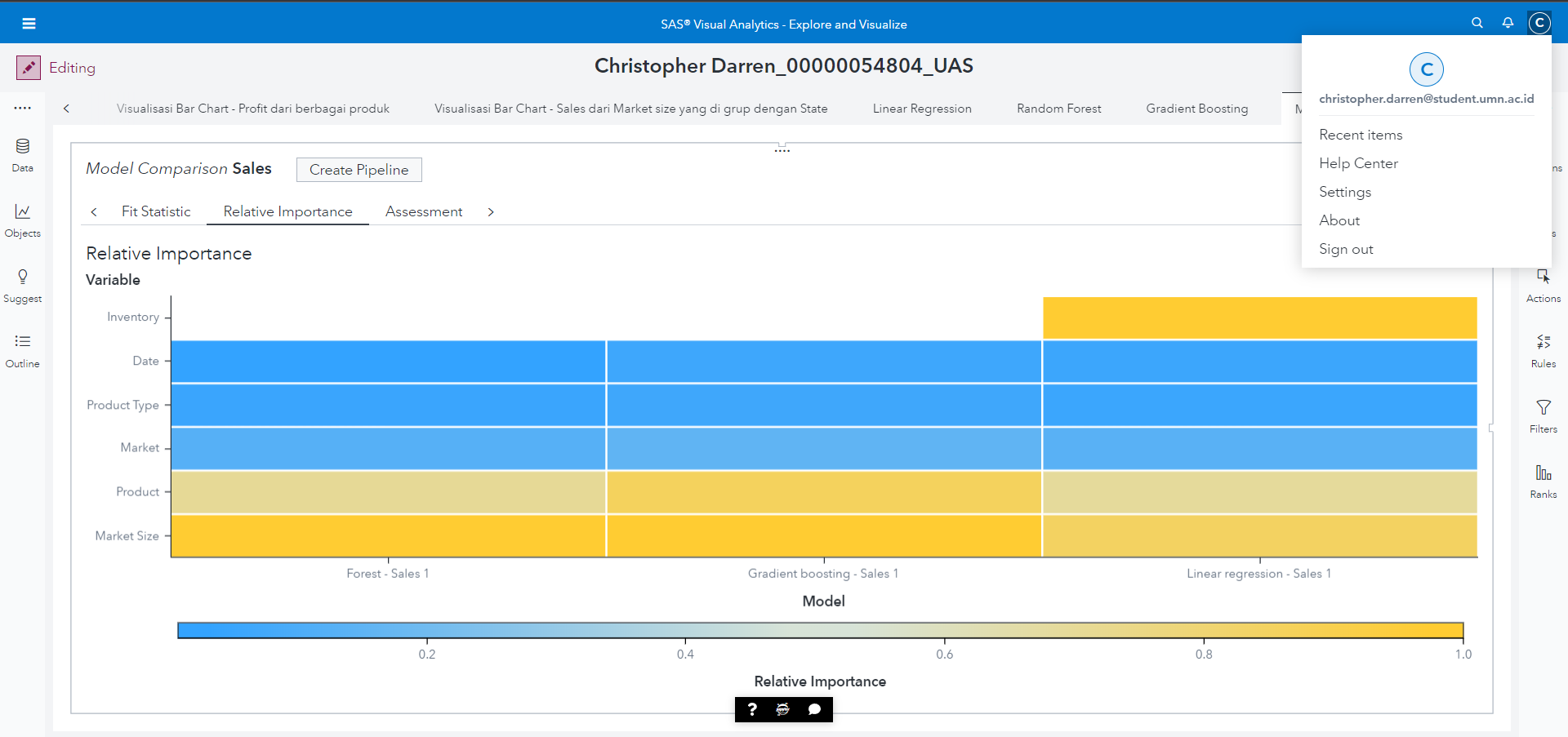
**Gambar 29. Visualisasi dari model comparison terhadap data KimochiMart**

Berdasarkan gambar diatas merupakan hasil secara keseluruhan dari ketiga model dari peneliti yang sudah buat, yakni linear regression, gradient boosting, dan random forest. Peneliti akan deep dive dari beberapa visualisasi yang sudah dibuat diatas.

****

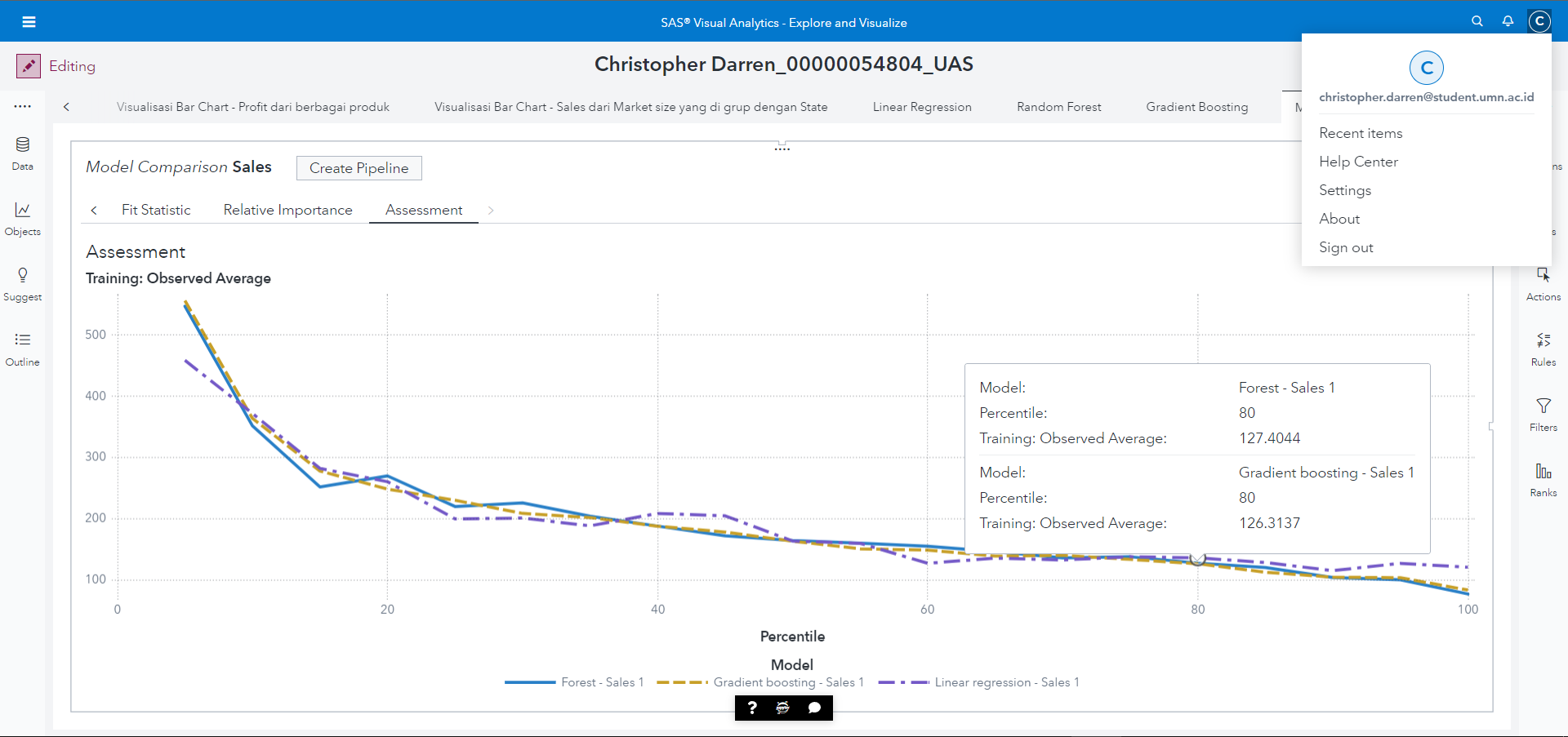
**Gambar 30. Visualisasi dari fit statistic data KimochiMart**

Berdasarkan perbandingan model diatas, model yang paling bagus adalah model *random forest*, karena validation ASE nya yang paling kecil yakni 12,935.6501. Dalam ilmu machine learning, hasil ASE yang kecil berarti perhitungan nya benar dan prediksi yang dilakukan itu yang paling baik sehingga random forest yang pemenangnya. Kemudian urutan keduanya ditempati gradient boosting, dan urutan ketiga ditempati oleh linear regression. Algoritma *machine* memilih prediksi yang paling tepat dari antara 3 tersebut dan secara default akan dihighlight.

****

**Gambar 31. Visualisasi dari *relative importance* terhadap data KimochiMart**

Berikutnya terdapat visualisasi *relative importance* dimana melihat pengaruh apa yang paling penting dalam melakukan prediksi. Terlihat bahwa semakin kuning relative *importance* nya maka semakin ada pengaruh terhadap prediksi, contohnya adalah pada *inventory*, khususnya model linear regression yang paling berpengaruh sedangkan model lain tidak. Pada kolom *Date*, *Product Type*, dan *Market* tidak ada pengaruh terhadap ketiga algoritma tersebut. Kemudian *relative importance* yang selanjutnya ada pengaruh yakni kolom *Product* dan *Market Size.*

****

**Gambar 32. Visualisasi dari assessment plot 3 model terhadap data KimochiMart.**

Berikutnya terdapat sebuah visualisasi mengenai assessment plot terhadap 3 model yang berbeda. Garis biru diwakili oleh random forest, garis kuning merupakan gradient boosting, dan garis ungu linear regression. Dari visualisasi tersebut bisa disimpulkan bahwa garis menurun yang paling bagus adalah random forest, walaupun pembuatan model awalnya tidak begitu bagus, tapi seperti penelitian sebelumnya seiring berjalannya waktu prediksinya semakin bagus dibandingkan model model lain karena pada random forest semua dikarenakan karena penambahan ensembel tree dari setiap berapa percentage sehingga jika garisnya menurun terus maka hasil prediksinya akan semakin baik.