Documentación Segmentación de Mercados

**Cliente xxxxx**

**xxxxxx/ Área Cumplimiento**

**Julio 2017**

Contenido

[I. Antecedentes 3](#_Toc490749895)

[II. Objetivo principal 3](#_Toc490749896)

[III. Conceptualización 3](#_Toc490749897)

[IV. Variables 6](#_Toc490749900)

[V. Consideraciones para ETL 6](#_Toc490749901)

[VI. Transformaciones aplicadas a la data 7](#_Toc490749902)

[VII. Resultados Modelamiento 7](#_Toc490749903)

[VIII. Descriptivo de segmentos 8](#_Toc490749904)

[IX. Metodología Conocimiento de Mercados 8](#_Toc490749907)

[X. Mejora continua 10](#_Toc490749912)

[XI. Anexos 10](#_Toc490749913)

# Antecedentes

La nueva resolución de la S.B.S N° 2660-2015 dentro del subcapítulo II, en el Artículo 34° estipula “Las empresas deben formar segmentos de mercado, estableciendo grupos que guardan una homogeneidad interna, pero heterogeneidad entre ellos, de acuerdo con una o varias variables.”

Producto de la regulación es que se necesita realizar una segmentación de mercados, utilizando técnicas estadísticas, dentro de un proyecto de data mining que asegure que los segmentos cumplan con las características deseadas de homogeneidad interna y heterogeneidad externa.

# Objetivo principal

Generar mediante clustering, segmentos de mercados para los clientes AFPP.

*Objetivo secundario*

Desarrollar el proyecto de tal manera que permita su ejecución de manera periódica.

# Conceptualización

Debido a la naturaleza de negocio de AFPP, sólo se tienen clientes del tipo persona natural.

* Los clientes serán divididos en 2 grupos:
  + Persona natural con aportes fijos (Definidos en base a número de operaciones).
  + Persona natural con aportes voluntarios (Definidos en base a número de operaciones).
* Se utilizarán para hacer las segmentaciones las variables Monto y Número de operaciones.
* La segmentación será realizada mediante la técnica de K-Medias. De existir algún grupo con una cantidad muy pequeña de datos, se utilizará alguna segmentación jerárquica.

# Clustering por K-medias

Clustering o segmentación por el algoritmo de K-medias, es una técnica no supervisada de clasificación, perteneciente a los métodos no jerárquicos, este método permite agrupar a los individuos u objetos homogéneos entre si y dividir a los mismos en grupos heterogéneos.

La constitución de los conglomerados se realiza mediante procedimientos de partición no jerárquicas y la asignación de los objetos a los conglomerados se realiza optimizando el criterio de selección.

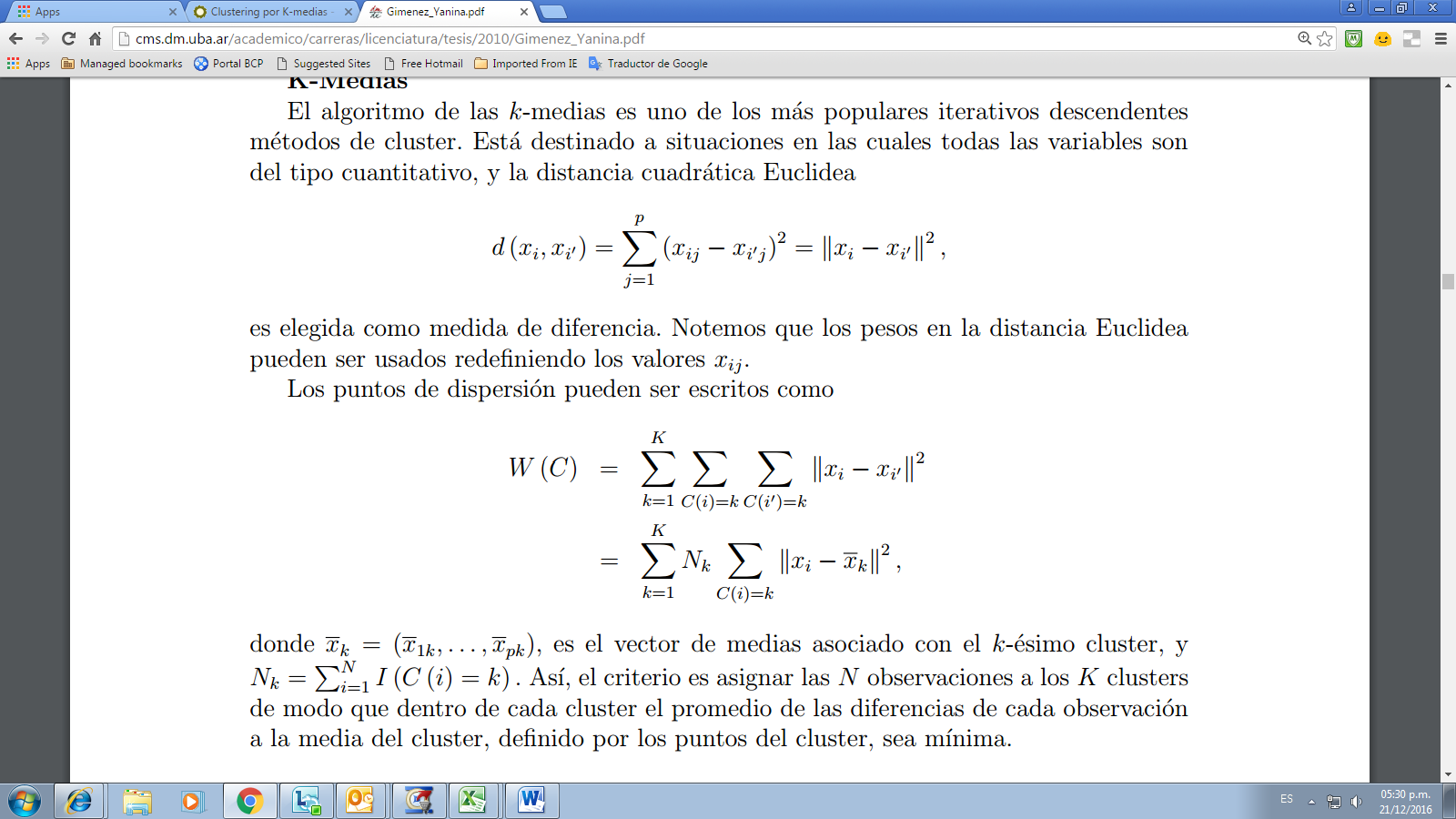
El procedimiento aproxima por etapas sucesivas un cierto número (prefijado) de clusters haciendo uso de los centroides de los puntos que deben representar.

El algoritmo se compone de los siguientes pasos:

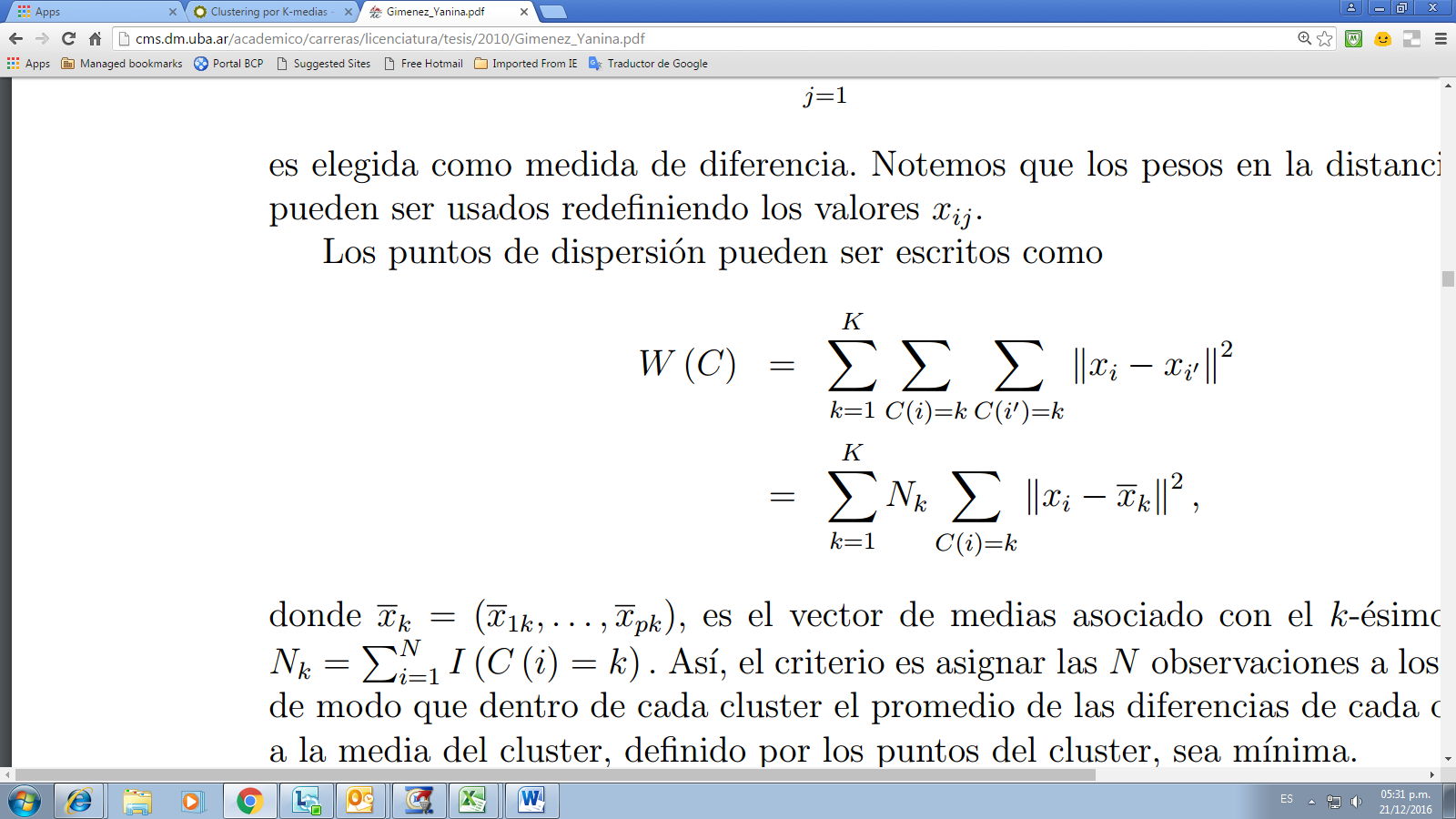
1. Sitúa K puntos en el espacio en el que "viven" los objetos que se quieren clasificar. Estos puntos representan los centroides iniciales de los grupos.
2. Asigna cada objeto al grupo que tiene el centroide más cercano.
3. Tras haber asignado todos los objetos, recalcula las posiciones de los K centroides.
4. Repite los pasos 2 y 3 hasta que los centroides se mantengan estables. Esto produce una clasificación de los objetos en grupos que permite dar una métrica entre ellos.

Aunque se puede probar que este algoritmo siempre termina, no siempre la distribución que se alcanza es la más óptima, ya que es muy sensible a las condiciones iniciales.

El algoritmo de las k-medias es uno de los más populares iterativos descendentes métodos de clúster. Está destinado a situaciones en las cuales todas las variables son del tipo cuantitativo, y la distancia cuadrática Euclídea



Es elegida como medida de diferencia. Notemos que los pesos en la distancia Euclidea pueden ser usados redefiniendo los valores xij. Los puntos de dispersión pueden ser escritos como



Donde xk = (x1k,. . ., xpk), es el vector de medias asociado con el k-ésimo clúster, y Nk = PN i=1 I (C (i) = k). Así, el criterio es asignar las N observaciones a los K clústers de modo que dentro de cada clúster el promedio de las diferencias de cada observación a la media del clúster, definido por los puntos del clúster, sea mínima.

# Clustering por método Ward

Él método Ward es parte de los método jerárquicos, estos se caracterizan porque en cada paso del algoritmo sólo un objeto cambia de grupo y los grupos están anidados en los de pasos anteriores. Si un objeto ha sido asignado a un grupo ya no cambia más de grupo. Recomendado para “n<=200”.

El Método Ward calcula la media de las variables de los conglomerados.

* Calcula la distancia euclídea al cuadrado entre cada objeto y la media de su conglomerado.
* Se suman todas las distancias.
* Se combinan aquellos conglomerados que tengan el menor incremento en la ST de las distancias al cuadrado dentro de cada conglomerado, es decir los que producen menor incremento en la varianza intragrupal.

# Variables

Para la segmentación se utilizarán las siguientes variables:

* CUSPP: Variable de identificación del cliente.
* Monto total: Promedio mensual de los movimientos totales del cliente durante el último año en dólares.
* Número total de operaciones: Promedio mensual de número total de operaciones del cliente durante el último año en dólares.

# Consideraciones para ETL

* Se ha trabajado con información correspondiente a operaciones de enero a diciembre del año 2015.
* Para identificar a los clientes que dejaron de serlo y que por lo tanto el último mes de registro tendrían cantidades muy grandes que inflarían desproporcionadamente los montos verdaderos se trabajó de la siguiente manera:
  1. Clientes con sólo una operación por mes: Se filtraron los clientes donde el último mes movieron un monto mayor a 2.5 veces el movimiento promedio.
  2. Clientes con a lo más 2 operaciones más que el número de meses que tiene registrado:
     + Si el último mes sólo tuvo una operación, se filtran los valores mayores a 3 veces el movimiento promedio.
     + Si el último mes tuvo 2 operaciones, se filtran los valores mayores a 5 veces el movimiento promedio.
     + Si el último mes tuvo 3 operaciones, se filtran los valores mayores a 7 veces el movimiento promedio.
  3. Clientes con número variable de operaciones por mes: Se filtraron los clientes donde el último mes movieron un monto mayor a 4 veces el movimiento promedio.
* Por último, mencionar que para el proceso de modelamiento, los 2 primeros grupos fueron agrupados debido a que no presentaba diferencia significativas, este nuevo grupo es el llamado de aporte fijo, mientras que el último grupo el de aporte voluntarios.

# Transformaciones aplicadas a la data

**Grupo 1 – aportes fijos**

* Se realizó un reemplazo de los límites superiores:
  + Monto: 5,000.
  + Numero: 1.

**Grupo 2 – aportes voluntarios**

* Se realizó un reemplazo de los límites superiores:
  + Monto: 1,000.
  + Numero: 5.

# Resultados Modelamiento

En total se realizaron 2 modelos con 5 clústeres.

**VII.1. Grupo 1 – aportes fijos**

|  |  |
| --- | --- |
| Modelo | N° Clústeres |
| K-Medias | 4 |

**VII.2. Grupo 2 – aportes voluntarios**

|  |  |
| --- | --- |
| Modelo | N° Clústeres |
| K-Medias | 3 |

Los resultados de los estadísticos descriptivos de la segmentación se encuentran en las tablas de SAS ***“*PARAM\_empresa1*”*** y ***“*PARAM\_empresa2*”*** ubicadas en la librería INPUT del SAS Grid.

# Descriptivo de segmentos

# Grupo 1 – aportes fijos

Descriptivo de segmentos para el grupo de aportes fijos.

|  |  |  |  |
| --- | --- | --- | --- |
| Clusters | Promedio ($) | Mínimo ($) | Máximo ($) |
| 1 | 113.01 | 0.09 | 365.83 |
| 2 | 4,692.79 | 3,362.33 | 5,000.00 |
| 3 | 602.76 | 365.84 | 1,312.61 |
| 4 | 1,956.45 | 1,312.98 | 3,359.18 |

# Grupo 2 – aportes voluntarios

Descriptivo de segmentos para el grupo de aportes voluntarios

|  |  |  |  |
| --- | --- | --- | --- |
| Clusters | Promedio($) | Mínimo ($) | Máximo ($) |
| 1 | 2.06 | 1.25 | 3.14 |
| 2 | 4.13 | 2.75 | 5.00 |
| 3 | 2.64 | 1.25 | 5.00 |

# Metodología Conocimiento de Mercados

# Introducción

Se ha establecido que la variable que permite conocer las características del mercado en las que opera el EMPRESA es la descripción estadística del punto anterior con el fin de perfilarlos, es decir:

* Para el grupo I de aportes fijos, los segmentos son caracterizados de acuerdo a los montos promedio, mínimo y máximo en dólares de su movimiento promedio.
* Para el grupo II de aportes voluntarios, los segmentos son caracterizados de acuerdo a los montos promedio, mínimo y máximo en dólares de su movimiento promedio.

# Escenario de alertas

Se tiene implementado un query desarrollado en MS Access que se ejecuta para obtener las alertas relacionadas a conocimiento de mercado en empresa. La lógica utiliza la información del Registro de Operaciones (RO) identificando a los clientes donde su promedio de movimientos del último mes supere en 3 desviaciones estándar del segmento al que pertenece. Estas alertas aplican tanto para el grupo de clientes correspondiente al de aportes fijos como aportes voluntarios. El detalle de la lógica se encuentra en el documento de Ficha Técnica correspondiente a Conocimiento de Mercado.

# Ejecución y periodicidad

La alerta del escenario para conocimiento de mercado es ejecutada de manera manual por el personal de Cumplimiento PLAFT empresa con periodicidad mensual para la alerta descrita. Las alertas resultantes de manera mensual al aplicativo de seguimiento para su respectiva revisión.

# Evaluación e investigación

El proceso de evaluación e investigación para las alertas generadas es realizado por el equipo de Cumplimiento PLAFT de empresa.

El procedimiento para la revisión de las alertas está descrito en la Guía de Procedimientos de Evaluaciones e Investigaciones, donde se indica que la revisión de las alertas se hace en 2 etapas:

* Evaluación de alertas: 30 días.
* Investigación de alertas: 40 días.

En ambos casos, el objetivo es la “Identificar tipologías de LAFT, validar e identificar señales de alerta contrastándolo contra el perfil del cliente”. Dentro de las alertas por conocimiento de mercado, los analistas utilizan los perfilamientos realizados a los segmentos de tal manera que pueden validar el perfil del cliente con el segmento de mercado al que pertenece.

# Mejora continua

Este informe será revisado y actualizado cada dos años, de modo tal que se puedan identificar oportunidades de mejora en la segmentación realizada.

# Anexos

**Anexo 1.**

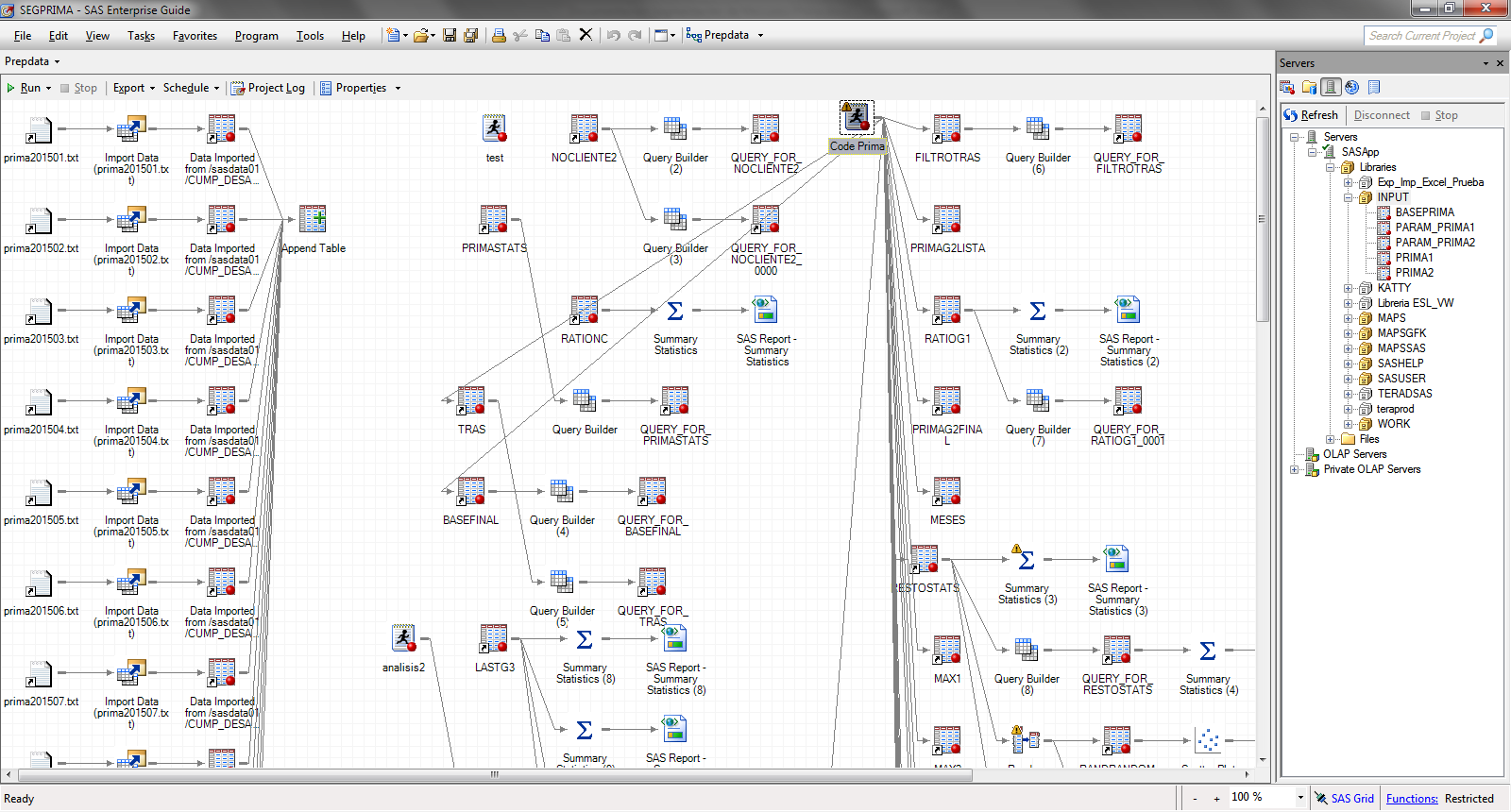
La segmentación se realizó en el siguiente proyecto que tiene la ruta:

*Z:\\_Gerencia\_Gestión\_de\_Cumplimiento\Inteligencia\_de\_Cumplimiento\(06)GestiónModelamiento\7\_Segmentación\Segmentación\_EMPRESAAFP\3. Modelamiento*

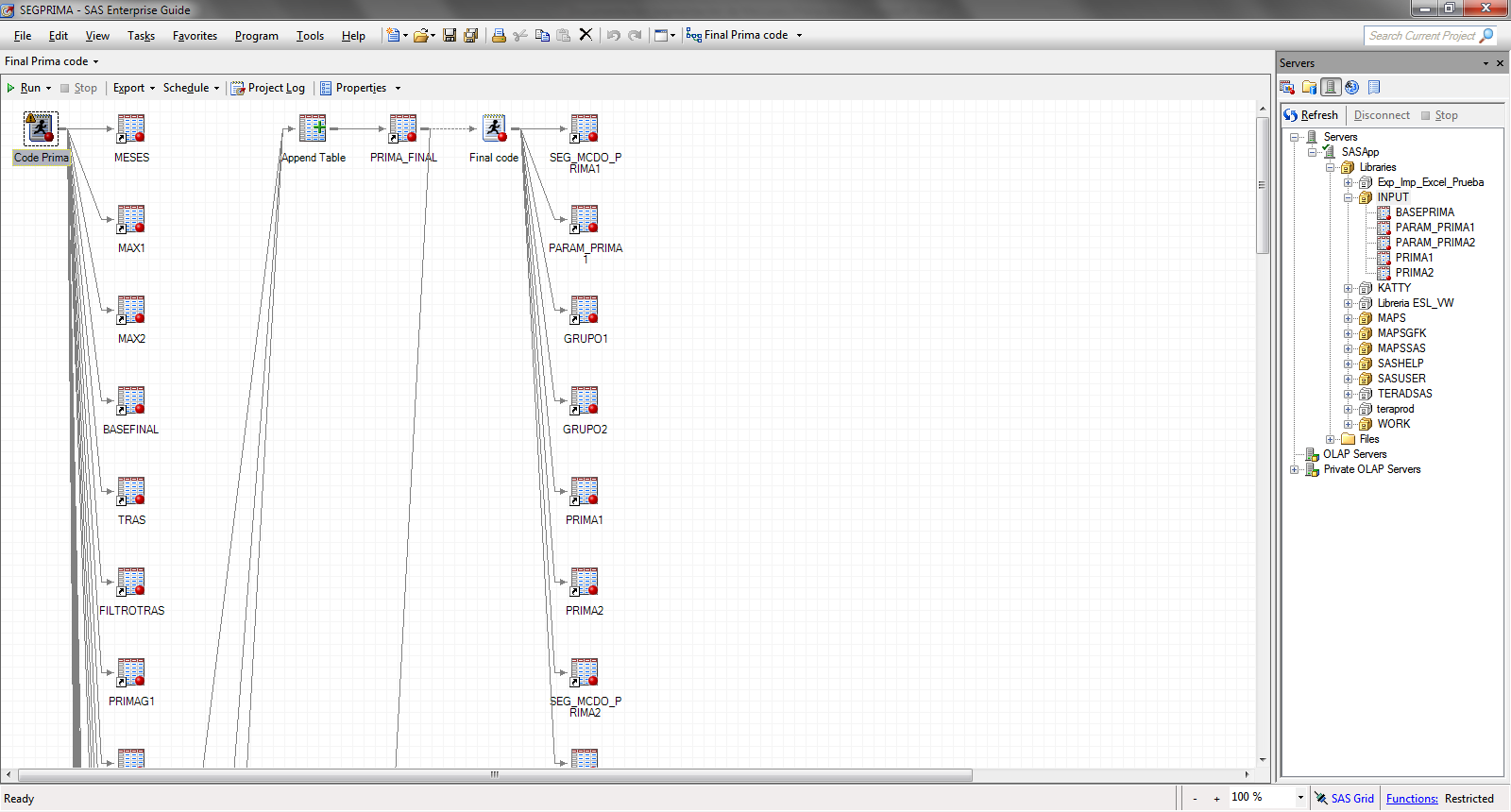


Este proyecto tiene dos “Process Flow”: PrepData y Final eMPRESA Code. En el Process Flow “PrepData” se preparó la información para el modelamiento, es decir, las transformaciones mencionadas en el punto de consideraciones de ETL. En el process flow “Final Empresa Code” se almacenan los códigos de la segmentación tanto para persona natural y jurídica obtenidos en el modelamiento en SAS Entrerprise Miner.

PrepData:



Final empresa Code:



* En SAS Enterprise Miner, el proyecto de segmentación de mercado PLAFT empresa se llama **“*SEG\_MCDO\_EMPRESA”*** ubicada en la siguiente ruta: SAS Foders/Proyecto\_Miner\_CUMPLIMIENTO y para el almacenamiento de la información se creó la siguiente librería: ***Libname DATA\_EPV '/sasdata15/CUMP\_DESA/SEG\_MCDO\_EMPRESA.***