

# CONCEPTOS DE DATA MINING APLICADOS A PROBLEMAS ADUANEROS

Lic. Cecilia Ruz

# Agenda

- ▶ **Clustering**
  - ▶ Clustering sobre las Aduanas del Interior
- ▶ Reglas de Asociación
  - ▶ Análisis de que productos se importan juntos

# Trabajo Realizado

## ▶ **Objetivo**

- ▶ Encontrar agrupaciones en las aduanas de acuerdo a su comportamiento en algunos grupos de variables. Todas las variables fueron medidas en el período 01/01/2004 – 31/12/2004

## ▶ **Descripción**

- ▶ Para caracterizar a las aduanas se decidió tomar dos grupos de variables distintas:
- ▶ Indicadores numéricos : que permiten caracterizar a las aduanas de acuerdo con su movimiento, ya sea de mercadería o de fondos.
- ▶ Clases de mercadería: en este punto se caracterizó a cada aduana por las 10 posiciones NCM que importaron/exportaron el mayor monto FOB y el mayor número de unidades en el período tratado. Como existe una gran variedad de NCM se tomaron sólo aquellas que aparecían repetidas en el top ten de por lo menos 2 aduanas.
- ▶ Para resolver este trabajo se utilizó el SPSS versión 12.0

# Indicadores Numéricos 1/2

- ▶ *Se tomaron diferentes indicadores y agrupó los mismos de acuerdo a su significado. Los indicadores numéricos se agruparon en las siguientes categorías:*
- ▶ *Declaraciones de Importación*
  - ▶ *Cantidad y monto FOB de las importaciones a consumo* [\[1\]](#) ( impo\_consumo.sql)
  - ▶ *Cantidad de manifiestos* [\[2\]](#) *marítimos de importación* ( sumarias.sql)
  - ▶ *Cantidad de manifiestos aéreos de importación* ( sumarias.sql)
  - ▶ *Cantidad de manifiestos terrestres de importación* ( sumarias.sql)
- ▶ *Declaraciones Detalladas Temporales*
  - ▶ *Cantidad y monto FOB de las importaciones temporales* [\[3\]](#)( impo\_temporales.sql)
  - ▶ *Cantidad y monto FOB de las exportaciones temporales* ( expo\_temporales.sql)
- ▶ *Declaraciones de Exportación*
  - ▶ *Cantidad y monto FOB de las exportaciones a consumo* ( impo\_consumo.sql)
  - ▶ *Cantidad de manifiestos marítimos de exportación* ( sumarias.sql)
  - ▶ *Cantidad de manifiestos aéreos de exportación* ( sumarias.sql)
- ▶ *Movimientos de mercadería ( dentro de una misma aduana y hacia otra)*
  - ▶ *Cantidad de transitos/traslados* [\[4\]](#) *sumarios por medio de transporte* ( sumarias.sql)
  - ▶ *Cantidad y monto de transitos/traslados detallados por medio de transporte* ( transitos\_traslados.sql)

# Indicadores Numéricos 2/2

- ▶ **Declaraciones Sumarias de Exportación**
  - ▶ *Cantidad de tránsitos de exportacion registrados por camion*
  - ▶ *Cantidad de tránsitos de exportación*
- ▶ **Unidades involucradas**
  - ▶ *Cantidad de contenedores en operaciones de importacion ( contenedores.sql)*
  - ▶ *Cantidad de contenedores en operaciones de exportacion ( contenedores.sql)*
  - ▶ *Unidades exportadas a granel por unidad de medida ( unidades.sql)*
  - ▶ *Unidades exportados en otros envases ( ni en contenedor ni a granel) ( unidades.sql)*
  - ▶ *Unidades importados a granel por unidad de medida ( unidades.sql)*
  - ▶ *Unidades importadas en otros envases ( ni en contenedor ni a granel) ( unidades.sql)*
- ▶ **Recaudación**
  - ▶ *Monto recaudado en impo ( recaudado\_aduana.sql)*
  - ▶ *Monto recaudado en expo ( recaudado\_aduana.sql)*
  - ▶ *Cantidad y monto de las garantías constituidas ( garantias.sql)*
- ▶ **Aclaración : Todos los datos se calcularan para el año 2004 y todos los montos están expresados en dolares estadounidense**

# Clase de Mercadería

- ▶ Para caracterizar a las aduanas se tomó el top ten de las mercaderías desde el punto de vista del monto FOB y desde el punto de vista de la cantidad de unidades estadísticas declaradas. En ambos casos se consideran los resultados tanto de importación como de exportación.
- ▶ **Preparación previa**
  - ▶ Una vez que se obtuvieron los top ten de las mercaderías para ambos casos se transformaron las mismas en variables dummy, con la única condición de que para que esto sucediera una determinada posición NCM ( caracterización internacional de la mercadería), tenía que encontrarse en por lo menos dos aduanas.

# Resultados Obtenidos

- ▶ Combinando ambos tipos de variables se obtuvo un agrupamiento en 13 clusters.
- ▶ Algunos ejemplos:
  - ▶ En el cluster 1 estaban las aduanas de SANTO TOME, BERNARDO DE YRIGOYEN e IGUAZU, que de alguna forma representan las “brasileras”, todas tienen límite con Brasil y son cercanas desde el punto de vista geográfico
  - ▶ En el cluster 3 estaban MENDOZA, CAMPANA, PASO DE LOS LIBRES podría denominarse “las grandes” son aduanas con mucho movimiento, fundamentalmente tránsitos terrestres.
  - ▶ En el cluster 4 estaban SAN LORENZO, ROSARIO, NECOCHEA, BAHIA BLANCA, las “graneleras”. Son aduanas con fuerte perfil exportador. Sus exportaciones varían entre U\$833 millones y U\$7.054, la cantidad de kilos brutos exportados varía entre 4.658 millones y 31.803 millones. Los productos que figuran entre las principales mercaderías de exportación para todas las aduanas son : maíz, trigo, porotos, aceite de girasol y los residuos de la extracción de aceite/grasas. Todas estas aduanas importan abonos, lo que refuerza su caracterización como aduanas graneleras.
  - ▶ En el cluster 10 VILLA REGINA, SAN ANTONIO OESTE, LA RIOJA, SAN JUAN, SAN RAFAEL, las aduanas “frutícolas”. Las exportaciones de este grupo varían entre U\$44 millones y U\$262 millones. Exportan peras, duraznos y ciruelas

# Agenda

- ▶ Clustering
  - ▶ Clustering sobre las Aduanas del Interior
- ▶ **Reglas de Asociación**
  - ▶ Análisis de que productos se importan juntos



# Modelo Temporal

- ▶ El modelo temporal contempla la posibilidad de que un itemset sea frecuente sólo en algunos subintervalos del período considerado.
- ▶ La regla temporal  $X \rightarrow Y$ :  $fl_{XUY}$  se mantiene en  $D$  con soporte temporal  $s_1, \dots, s_n$  y confianza  $c_1, \dots, c_n$  si  $s_1\%, \dots, s_n\%$  de las transacciones de  $D$  en  $fl_{XUY} = \{\text{subintervalo}_1, \dots, \text{subintervalo}_n\}$  contienen  $XUY$  y  $c_1\%, \dots, c_n\%$  de las transacciones que incluyen a  $X$  también incluyen a  $Y$ .

# Aplicación a un caso real

Los datos utilizados corresponden a todas las importaciones a consumo realizadas en la Argentina durante el período 1995-2003.

Para caracterizar las importaciones se tomaron en cuenta los siguientes datos:

- ▶ fecha en que se produjo la importación
- ▶ posiciones arancelarias involucradas en la misma

# Objetivo

Determinar que productos se importan juntos para permitir asignar un mayor control a aquellos casos donde se declare un producto y no el otro.

# Preparación previa

- ▶ Eliminación de posiciones referidas a “los demás”
- ▶ Eliminación de posiciones repetidas en una misma transacción
- ▶ Eliminación de destinaciones que quedaron sin detalle
- ▶ Conversión de las posiciones arancelarias en números consecutivos

# Resultados Obtenidos

- ▶ Como resultado del análisis se obtuvieron 17 reglas que son válidas en distintos intervalos de tiempo
- ▶ Algunos ejemplos de reglas son
  - ▶ 8471.60.53.000Q->8471.60.52.000H en el intervalo 22/04/2003-11/08/2003 con una confianza de 0.83
  - ▶ 8471.60.52.000H->8471.60.53.000Q en el intervalo 22/04/03-11/08/2003 con una confianza de 0.84
- ▶ La posición 8471.60.52.000H representa teclados para computadoras , mientras que la posición 8471.60.53.000Q indica que se trata de Indicadores o apuntadores (por ejemplo: "mouse" y "trackball").
- ▶ Como ambas reglas se originan en el mismo itemset tienen el mismo soporte, que es de 0,02014

# Preguntas



**Muchas Gracias!!!**

