

ANÁLISIS DE CLÚSTERES Y SEGMENTACIÓN EN EL MERCADO DE CRUCEROS DE URUGUAY

ALICIA BELLANI

aliciabellani@hotmail.com

*Facultad de Ciencias Económicas y de Administración, Universidad de la República, Grupo de Investigación en Dinámica Económica,
Eduardo Acevedo 1139, Montevideo, Uruguay*

LEONARDO MORENO

mrleo@iesta.edu.uy

*Facultad de Ciencias Económicas y de Administración, Universidad de la República, Departamento de Métodos Cuantitativos,
Eduardo Acevedo 1139, Montevideo, Uruguay*

SANDRA ZAPATA AGUIRRE

szapata@ccee.edu.uy

*Facultad de Ciencias Económicas y de Administración, Universidad de la República, Departamento de Administración,
Eduardo Acevedo 1139, Montevideo, Uruguay*

Recibido (30/10/2018)

Revisado (26/12/2018)

Aceptado (30/12/2018)

RESUMEN: El objetivo del estudio es aportar información para la toma de decisiones en el marco de políticas turísticas aplicadas a la gestión de destinos. Diversos mecanismos de segmentación de mercado son aplicados de forma frecuente en el análisis de turismo y son utilizadas diferentes metodologías estadísticas según el fin que se persigue. En este trabajo se aplica una técnica de clasificación no supervisada que tiene por objetivo determinar grupos de variables. Se hace énfasis en un criterio de poda de variables que determina un grupo de ruido y en la comparación de dos metodologías diferentes en referencia al trato de las variables cualitativas. A partir de ellas se realiza una segmentación de los cruceristas. En esta última etapa es novedosa la aplicación de un criterio selección de variables que intenta emular los grupos obtenidos con las variables originales. Son analizados los resultados obtenidos en cada paso y se realiza una comparativa con clústeres determinados en temporadas anteriores.

Palabras Clave: Clústeres de variables, k -medias, selección de variables, turismo de cruceros

ABSTRACT: Diverse market segmentation mechanisms have been frequently applied in the tourism analysis and different statistical methodologies have been used according to the objective pursued. This study applies a segmentation technique that tries to determine groups of variables, based on their quadratic correlation for quantitative variables or based on the sum of correlation ratios for qualitative variables. In each of the clusters of variables, the subgroup of determinant variables are identified using a pruning technique which cancels the effect of unnecessary variables. From these set of determinant variables we conduct a cruise segmentation analysis using k -means. The results show the formation of four clusters in which the key variables are age, residence, travel party size and the port of visit. Then, we analyze the cluster expenditure type behavior, finding that shopping is the most important spending in all the groups. The study aims to provide information for decision making within the framework of tourism policies, in reference to each of the groups obtained.

Keywords: Cruise tourism, variable clusters, k -means, variable selection

1. Introducción

La industria de cruceros a nivel mundial ha experimentado un crecimiento sostenido en los últimos 15 años, comenzando la década del 90 con casi 4 millones de pasajeros hasta llegar a más de 23 millones en 2014, ver [CLIA, 2018] y [Brida et al., 2014a]. En los últimos cuatro años el sector ha experimentado un crecimiento promedio anual de casi el 9%. El sector de Crucerismo es el sector más dinámico del turismo y de las actividades recreativas en general y se prevé que este dinamismo continúe en el futuro. Según CLIA (Cruise Lines International Association) en los últimos años el dinamismo de la industria se debió al mayor número de cruceros y a la mayor diversidad de rutas en especial los cruceros que navegan por ríos, así como también a las nuevas generaciones que están optando por este tipo de turismo, ver [CLIA, 2018].

En este contexto, los países de América del Sur se han beneficiado también de este crecimiento y Uruguay no ha sido la excepción. Si bien el turismo de cruceros sigue siendo una pequeña porción del total de turistas que ingresan al país, este no es ajeno al dinamismo mundial que caracteriza al mercado (según datos obtenidos del anuario que publica el Ministerio de Turismo de Uruguay, en 2013 el total de cruceristas alcanzó un 13% del total de turistas que visitaron Uruguay, en 2014 el 12%, en 2015 el 9,8% y 8,4% en 2016).

En Uruguay los cruceros visitan dos puertos: la capital (Montevideo) y una ciudad costera con un gran desarrollo en la actividad de turismo de playa (Punta del Este). Montevideo se transformó hace muy poco en un puerto de embarque mientras que Punta del Este es un puerto de escala. Otra diferencia entre ambos puertos es que en Montevideo los pasajeros desembarcan directamente en el muelle, en Punta del Este el crucero ancla a cierta distancia y los pasajeros son trasladados en bote hasta el muelle [Brida et al., 2015a].

En la temporada 2013-2014 se registró un incremento de 6,3% en la llegada de cruceros con respecto a la temporada anterior, crecimiento que se explica en mayor medida por la afluencia de cruceros a Punta del Este la cual llegó a cifras récord, ver [Ministerio de Turismo, 2015]. El gasto también tuvo cifras récord en esa misma temporada llegando a 19 millones de dólares corrientes. En cuanto a la cantidad de personas que llegaron en la modalidad de crucero se registra un pico de cruceristas arribados en la temporada de 2012-2013 (411.937 cruceristas). Cabe notar que estas cifras disminuyeron de forma leve en las siguientes temporadas.

Por lo anteriormente planteado, resulta importante estudiar en detalle las características de la demanda de cruceristas que llega al país para que el mismo pueda aprovechar al máximo los beneficios asociados a esta actividad. En este sentido, el objetivo de este trabajo es encontrar y analizar grupos homogéneos de pasajeros que desembarcaron en los puertos uruguayos. Esto permite identificar segmentos de mercado que luego podrán ser analizados en detalle y de acuerdo con sus características, apoyar la toma de decisiones empresariales y políticas orientadas a mejorar la relación entre el destino y el turista, encontrar mercados objetivos y ofrecer la mayor cantidad de actividades que le sean atractivas a los cruceristas que llegan al país. Por otro lado también es de interés poder analizar la evolución temporal de este tipo de turismo en el país.

Las técnicas de segmentación de mercado se han utilizado de forma extensiva en la investigación académica, ver [Liao et al., 2012], [Romero and Ventura, 2013] y [Ngai et al., 2009]. Estas técnicas son útiles para conocer las preferencias de los consumidores y poder ajustar los productos y servicios ofrecidos por un destino para satisfacer las necesidades de los turistas [Pesonen, 2012] y ver por ejemplo los distintos flujos económicos que ellos determinan, ver [Brida et al., 2018]. Para lograr el objetivo planteado se realizan tres etapas de análisis en las cuales se adoptan técnicas de agrupamiento tanto para variables como para unidades. El primer paso consiste en la reducción del conjunto original de variables. Esta reducción del número de variables tiene el propósito de encontrar variables dependientes que podrían estar brindando información similar y al mismo tiempo encontrar e interpretar dimensiones comunes a dichas variables.

A diferencia de trabajos recientes sobre segmentación en [Brida et al., 2014a], además de utilizar un algoritmo de *k*-medias en lugar de un algoritmo jerárquico, se considera un grupo de variables de ruido que no presentan una correlación muy baja con cada clúster de variable. En particular son utilizados dos enfoques distintos para el tratamiento de las variables cualitativas y los resultados obtenidos son comparados.

Después de encontrar los grupos de variables el siguiente paso es segmentar los datos a través del algoritmo de agrupamiento de Ward para determinar grupos de personas. Lo innovador en este caso respecto a [Brida et al., 2014a] es la asignación de cierto nivel de importancia a cada variable en el momento de determinación de los clúster de individuos. En este sentido seguimos una metodología novedosa pero de escaso uso en las aplicaciones (desarrollada en [Fraiman et al., 2008]).

Por último, mediante el análisis CART se pretende encontrar las variables que impulsaron la formación de los conglomerados.

El trabajo se organiza de la siguiente manera. A continuación se realiza un análisis de la literatura, luego se describe la base de datos que se utilizará en este trabajo. Se realiza un breve introducción a metodología utilizada, se describen las etapas que se siguieron y posteriormente se enumeran los principales hallazgos. Por último se mencionan algunas implicancias en política y marketing que se derivan de los resultados y se mencionan aquellas diferencias significativas respecto a los clústeres encontrados en la temporada anterior.

2. Revisión de la literatura

Se han implementado varios algoritmos aplicados a la técnica de clústeres, por ejemplo el de *k*-medias, de particionamiento, distancias pequeñas entre los miembros del mismo, áreas densas del espacio de datos, intervalos o distribuciones estadísticas particulares, etc., que permiten categorizar los clústeres en Clúster Particionado y Clúster Jerárquico. El algoritmo apropiado depende del conjunto de datos (cualitativo y/o cuantitativo) que se analiza y el uso que se le dará a los resultados, ver [Joshi and Kaur, 2013] y [Sánchez-Hernández et al., 2013].

En lo que se refiere al turismo, esta técnica también se ha utilizado ampliamente con el objetivo de conocer las características de los turistas que llegan de visita. Es el caso del trabajo de [Bloom, 2004] que con el objetivo de solventar las limitaciones que imponen las técnicas de regresión lineal, realiza un ejercicio para encontrar relaciones no lineales mediante la segmentación del mercado, aplicadas a los turistas que visitan Ciudad del Cabo. Algo similar se realiza en el trabajo de [Dolnicar et al., 2012] en el cual se aplica una técnica particular de clústeres (biclústeres) y también en el trabajo de [Kim et al., 2003], que aplica este tipo de técnica al análisis de las decisiones turísticas que realizan las personas mayores que visitan la zona Oeste de Australia.

Sin embargo, la aplicación de estas técnicas no se ha implementado de forma tan extensa en el estudio y análisis de la experiencia de los cruceristas que visitan un puerto de escala. En trabajos realizados para Cartagena de Indias, [Brida et al., 2012] y [Brida et al., 2015b], encontraron tres clústeres alrededor de la variable “tiempo empleado en el destino”: pasajero en tránsito, estadía corta y estadía larga, concluyendo que el gasto realizado por el grupo de turistas que está poco tiempo en el destino es casi imperceptible.

En el trabajo del año 2015, se encontraron 6 segmentos de mercado cuyas variables centrales fueron, nacionalidad, satisfacción, percepción de seguridad y gasto. En este último trabajo el resultado fundamental al cual llegaron fue la necesidad de mejorar la percepción de seguridad y la posibilidad de trabajar sobre programas turísticos específicos para el grupo de jóvenes sudamericanos.

En [Andriotis and Agiomirgianakis, 2010] encuentran que los cruceristas que llegan a la ciudad de Heraklion (Creta, Grecia) no lo hacen motivados por el sol y la playa como sucede con otros

destinos del Caribe o del Mediterráneo, sino que están motivados en explorar los aspectos culturales del destino.

Por su parte, en [Thurau et al., 2007] implementan la técnica de segmentación de mercados a los cruceristas que llegan a Panamá, y encuentran que el ecoturismo tiene mayor nivel de preferencia al que se pensaba previo al estudio. Para Uruguay se encontraron dos trabajos que utilizaron la metodología de clústeres. En uno de ellos [Brida et al., 2014b], se segmenta el mercado de cruceristas en tres grupos según el país de residencia, las ciudades visitadas y visitas previas al país.

Se concluye que los turistas que visitan por primera vez Punta del Este, los jubilados que visitan Montevideo y los brasileños, constituyen las categorías de turistas que podrían intensificar en mayor medida, el impacto económico proveniente de la industria de cruceros.

En el segundo trabajo, se utiliza el algoritmo ClustOfVar y se encuentran cinco clústeres que demuestran que el agrado de los cruceristas con la gente local y características tales como la edad y el género son las variables más importantes para explicar el impacto económico del turismo de cruceros, ver [Brida et al., 2014a]. Estos últimos resultados son utilizados como insumo en la comparación con el presente trabajo.

3. Descriptiva de la base de datos

Para llevar a cabo este estudio se utilizan los datos de la encuesta realizada por el Ministerio de Turismo durante la temporada 2014-2015 y la temporada 2015-2016. La base de datos cuenta con 6207 casos y 45 variables.

Las encuestas se implementaron en los 2 puertos de desembarco (Montevideo y Punta del Este), por lo que se realizó una muestra independiente en cada puerto. Para obtener dicha muestra se llevan a cabo 2 etapas: por un lado se sortean los cruceros y por otro lado los “grupos de viaje” que descienden de los mismos.

En una primera fase, los cruceros fueron seleccionados mediante muestreo sistemático, a partir de una lista de buques que se espera lleguen durante la temporada. En la segunda etapa, se eligen con equiprobabilidad los grupos de viaje a ser encuestados. Las entrevistas se realizan cuando los cruceristas están subiendo al barco, o sea cuando su viaje ha terminado.

La Tabla 1 presenta la lista de variables que se utilizan para el análisis. Las mismas pueden agruparse en: socio-demográficas, relacionadas con el viaje y las psicográficas.

Se observa que el 58% de las personas descendidas son mujeres. La mayor parte de los cruceristas tienen entre 30 y 59 años (69%). Los países vecinos aportan un porcentaje elevado en este tipo de turismo (51% en Brasil y 21% en Argentina). Del total visitan nuestro país por primera vez el 71%. En cuanto a la ocupación el 32% fueron profesionales, técnicos o docentes, el 27% fueron jubilados y el 18% empresarios, comerciantes o patrón.

El principal motivo de agrado fue la gente, cordialidad y la atención (28%), seguido por la arquitectura, los edificios y monumentos (18%). De las cosas que mencionaron como desagradado destacan los precios (58%) y la higiene (18%). El gasto promedio ha sido de USD 34, siendo compras (60%) y alimentación (24%) los principales rubros de gasto.

Con respecto a temporadas anteriores (ver [Brida et al., 2014b] y [Brida et al., 2014a]) se encuentra que el país de origen prevaeciente de los pasajeros continúa siendo Argentina y Brasil, con un notorio aumento de la presencia de brasileños y un disminución de los pasajeros argentinos en las temporadas analizadas en este estudio. Asimismo, en promedio, en la mayoría de los pasajeros que visitan los dos destinos de estudio, se encontró un leve aumento de la cifra de los que por primera vez visita el país.

Tabla 1: Index on isotropic matrix with exponential parameter

Variables cualitativas			Variables cuantitativas		
Etiquetas	Descripción	Porcentaje	Etiquetas	Descripción	Porcentaje
Género			Puerto de vista		
Mujeres	Cantidad de mujeres en el grupo	0,58	Mdeo	Puerto de Montevideo	0,61
hombres	Cantidad de hombres en el grupo	0,42	Punta	Puerto de Punta del Este	0,39
Residencia			Gustos y desgustos		
Resid-arg	Residentes argentinos	0,21	GustaEdif	Agrado Arquitectura, edificios y monumentos	0,18
Resid-bra	Residentes brasileños	0,51	GustaPaisaje	Agrado Paisajes, plazas y espacios verdes	0,09
Resid-otrosAm	Residentes en otros países de América	0,02	GustaHigiene	Agrado Higiene	0,1
Resid-norte	Residentes en Norteamérica	0,13	GustaMtdo	Agrado Montevideo	0,04
Resid-euro	Residentes en Europa	0,1	GustaMercPuerto	Agrado Mercado del Puerto	0,08
Resid-Otros	Residentes en otros países	0,03	GustaGente	Agrado Gente, cordialidad, atención	0,04
Resid-Otros	Residentes en otros países	0,03	GustaGente	Agrado Gente, cordialidad, atención	0,04
Ocupación					
Jubilado	Jubilado o Pensionista	0,27	GustaTranq	Agrado Tranquilidad, seguridad	0,08
AmadeCasa	Ama de casa	0,03	GustaComida	Agrado Comida, gastronomía	0,04
Estudiante	Estudiante	0,02	DesagradNoTiendas	Desagrado Tiendas cerradas	0,61
Ocup-Otros	Otra ocupación	0,05	GustaPunta	Agrado Punta del Este	0,03
Desocupado	Desocupado	0	DesagradPrecios	Desagrado Precios	0,58
Empresario	Empresario, comerciante, Patrón	0,18	DesagradAtencion	Desagrado Mala atención en los servicios	0,10
Profesional	Técnico, Profesional, docente	0,32	DesagradHigiene	Desagrado Higiene	0,61
Empleado	Empleado público y privado, vendedor, cajero	0,13	GustaOrganizacion	Agrado organización	0,39
Repetición de la visita					
PrimeraVez	Visita por primera vez	0,71	GustaCasapueblo	Agrado Casapueblo	0,02

4. Metodología

Como ya se mencionó anteriormente, las técnicas de segmentación se han utilizado de forma extensiva, lo que ha generado una evolución importante en las mismas de manera de solventar algunas limitaciones, errores y malas interpretaciones que surgen de su aplicación en la investigación. Un problema que surge al momento de agrupar unidades (crear clústeres) es la redundancia en la información de las variables. Este problema se origina en la posibilidad de que las variables estén fuertemente correlacionadas, determinando así la duplicación de información y la clasificación de unidades es a veces engañosa. Otro problema es la inclusión en el análisis de variables que no presentan asociación con la variable a ser explicada, por tanto son meramente ruido que perturban el problema. En el trabajo (a diferencia de [Brida et al., 2014a]) se observa la presencia de ciertas variables no influyentes en el problema de segmentación, por tanto es fundamental poder detectar las mismas con el fin de obtener resultados más claros y precisos. La determinación de clústeres de variables es una alternativa válida para la reducción de la dimensionalidad de los datos, presentando la ventaja de una mayor interpretabilidad de los resultados frente a otras alternativas como, por ejemplo, el análisis factorial.

En esta sección describiremos los procedimientos que empleamos para determinar aquellos grupos (también denominados clústeres) de variables de mayor importancia para la delimitación de los diferentes comportamientos de los turistas cruceristas en Uruguay en la temporada 2014-2015 y 2015-2016.

4.1. Clúster de Variables

La primera etapa tiene por objetivo determinar grupos de variables asociadas al comportamiento de los turistas cruceristas. La metodología de [Vigneau and Qannari, 2003] nos permite identificar dichos clústeres, basados en el análisis de componentes principales (PCA) y mediante la construcción de variables latentes.

Se observa un resultado novedoso, ambos mecanismos proporcionan en nuestro caso clústeres similares de variables. Posterior a la determinación de estos grupos, en una segunda etapa, se

realiza una “poda” de estos. Es decir, se excluyen de los grupos aquellas variables que presentan poca correlación, en referencia a su varianza, respecto a la variable latente que determina el grupo, ver [Vigneau et al., 2016]. Estas variables descartadas pasan a formar parte de otro clúster que denominaremos clúster residuo. Posteriormente se determinan grupos de turistas cruceristas mediante algoritmo de k -medias. Se realiza una búsqueda de aquellas variables que son determinantes de estos clúster de variables. La metodología implementada para la selección de estas variables se basa en el trabajo de [Fraiman et al., 2008]. Puesto que el número de variables es elevado, para la implementación del algoritmo forward-backward, en el algoritmo de [Fraiman et al., 2008], las variables se van eligiendo secuencialmente de los clúster de variables y del grupo de las variables residuo obtenidas en la primera etapa (esto genera un nexo de dependencia entre ambas etapas).

El trabajo puede situarse como una extensión del estudio de [Brida et al., 2014a] pero con diferencias metodológicas sustanciales que nos permitirán analizar la evolución que ha seguido el mercado de los cruceristas en función de existencia o no de nuevos perfiles de pasajeros. En este trabajo se busca determinar clústeres de variables más robustos, podando aquellas variables que cumplen un papel de ruido en el modelo. A diferencia de [Brida et al., 2014a], en lugar de utilizar procedimientos jerárquicos, se determinan los grupos a partir de algoritmos basados en k -medias, tanto en la primera (clúster de variables) como en la segunda parte del trabajo (clúster de individuos). Además, se utiliza un paquete (ClustVarLV) que mide las variables como cuantitativas, a diferencia del paquete que utiliza el trabajo mencionado (ClustOfVar) que toma variables mixtas. En lo que se refiere a los clúster de individuos, ambos trabajos difieren en el método utilizado para identificar las variables más significativas; en el trabajo de [Brida et al., 2014a] se utiliza el método CART y en este trabajo se utiliza, como método novedoso, el algoritmo desarrollado en [Fraiman et al., 2008]. Por otro lado en el trabajo anterior la nacionalidad fue considerada como una variable explicativa para la determinación de los clúster. En este trabajo se busca observar como la tipología de los cruceristas encontrada aporta información sobre la procedencia de este.

Consideramos en particular que este tipo de encuestas pueden estar contaminadas por diversos motivos (por ejemplo falta de tiempo del crucerista o la “deseabilidad social” al contestar las preguntas). Por tanto es relevante podar determinada información que distorsiona el problema en cuestión.

En la primera etapa se busca una medida de similaridad entre las variables. En el caso del método CLV introducido por [Vigneau and Qannari, 2003] se consideran las variables agrado y desagrado como cuantitativas dicotómicas tomando valores 0 y 1. Llamamos x_j con $j = 1, \dots, p$ al vector determinado por la respuesta a la j -ésima variable, donde p es el número de variables. Si se fijan a priori K grupos se modelan las variables latentes c_k asociada al k -ésimo grupo de forma de maximizar la expresión,

$$S = \sum_{k=1}^K \sum_{j=1}^p \delta_{kj} \text{cov}(x_j, c_k), \quad (1)$$

donde la $\text{Var}(c_k) = 1$ y δ_{kj} es la función indicatriz que vale 1 si la variable j se encuentra en el grupo k . Como los autores demuestran, fijada la partición de variables, las variables latentes que optimizan la expresión S son

$$c_k = \frac{x'_k}{\sqrt{\text{Var}(x'_k)}} \quad (2)$$

Podemos encontrar dos algoritmos para el agrupamiento de las variables:

- (1) algoritmo de agrupación jerárquica,

(2) algoritmo de partición (k -medias).

Ambos algoritmos de agrupación tienen como objetivo maximizar el criterio de homogeneidad. Este criterio implica que las variables de un clúster estén fuertemente asociadas a una variable cuantitativa sintética central (variable de resumen).

Esta relación se mide, como ya se mencionó anteriormente, como la suma de los coeficientes de correlación o correlaciones cuadráticas según el tipo de variable. Adicionalmente, se realiza un bootstrap para medir la estabilidad de las particiones de las variables y se puede utilizar como criterio para determinar el número ajustado de clústeres a realizar.

A partir del algoritmo de k -medias, que es el que utilizamos en este trabajo, el algoritmo iterativo para determinar los clústeres es entonces,

(Algoritmo 1)

- Se fijan la cantidad de grupos y se selecciona una partición inicial de variables.
- Se determinan las variables latentes c_k .
- Se reasignan las variables al grupo k -ésimo donde maximiza la covarianza con c_k .
- Se reitera el procedimiento con la nueva partición.

Si las variables gustos y desgustos las consideramos como cualitativas, es decir, los grupos $\beta_1 = \{x_1, \dots, x_{p_1}\}$ y $\beta_2 = \{z_1, \dots, z_{p_2}\}$ son las variables cuantitativas y cualitativas respectivamente, la variable latente que identifica al grupo c_k esta dado por,

$$c_k = \operatorname{argmax}_u \left(\sum_{x_j \in C_k} r^2(x_j, u) + \sum_{z_j \in C_k} \tau^2(z_j, u) \right), \quad (3)$$

siendo r el coeficiente de correlación y τ la razón de correlación entre las variables, ver [Chavent et al., 2017]. El cálculo de c_k es la primer componente principal a través de algoritmo PCAMIX, ver [Kiers, 1991], aplicado a la matriz determinada por las columnas de las variables cuantitativas y la matriz determinada por las columnas de las variables cualitativas.

Al igual que en el algoritmo del método anterior se parte de una partición dada, se determinan las variables latentes y en cada paso se asocia cada variable a la variable latente con la cual presenta mayor asociación (coeficiente de correlación o razón de correlación según corresponda).

Podado de clústeres Se busca identificar un clúster de ruido, concepto introducido por [Dave, 1991]. Para identificar aquellas variables que quedan incluidas en los K clústeres del Algoritmo 1 pero serán llevadas a un nuevo grupo de ruido se utiliza la estrategia “ $K + 1$ ” como se denomina en [Vigneau et al., 2016]. En este caso la expresión a maximizar es

$$S = \sum_{k=1}^K \sum_{j=1}^p \delta_{kj} \operatorname{cov}(x_j, c_k) + \sum_{j=1}^p \left(1 - \sum_{k=1}^K \delta_{jk} \right) \rho \sqrt{\operatorname{Var}(x_j)}, \quad (4)$$

con $\operatorname{Var}(c_k) = 1$. Por tanto el Algoritmo 1 es modificado de la siguiente forma,

- Si, fijado j , $\operatorname{cov}(x_j, c_k) < \rho \sqrt{\operatorname{Var}(x_j)}$ entonces la variable x_j es asignada al grupo de ruido.
- Si $\operatorname{cov}(x_j, c_k) \geq \rho \sqrt{\operatorname{Var}(x_j)}$ entonces la variable x_j se asigna al grupo k si $\operatorname{cov}(x_j, c_k) < \operatorname{cov}(x_j, c_{k'})$ para todo $k \neq k'$

Es claro que el tamaño del grupo de ruido depende de la elección de ρ . La selección de este parámetro es discutida en [Vigneau et al., 2016] y [Cabilio and Masaro, 2001].

Finalmente, fijado el número de clúster en $K = 4$ (mediante el índice ajustado de Rand), se realiza el Algoritmo 1 a partir de la partición inicial obteniendo los clústeres direccionales de

variables. En la Tabla 4 (ver Anexo) se representan los clústeres de variables, la correlación de cada variable con la variable latente del grupo y su correlación con la variable latente del grupo con el que se encuentra correlacionada en segundo lugar.

Estos resultados fueron obtenidos bajo el supuesto de que las variables de gustos y desgustos son consideradas del tipo cuantitativo pero, si consideramos las variables gustos y desgustos como cualitativas, utilizando la librería *ClustofVar*, los clústeres de variables obtenidos son similares. También en la Tabla 4 figura con una cruz aquellas variables que el algoritmo asoció a un clúster diferente. En la misma Tabla pero en la columna denominada Poda se encuentran marcadas con una cruz las variables que pasaron a formar parte, luego de utilizada la estrategia “ $K + 1$ ”, del clúster de outliers. La proporción de poda fue de con $\rho = 0.3$.

La matriz de correlación entre las variables latentes que identifican los grupos es dada por la Tabla 2. Se observa aquí una asociación moderada en los tres primeros clústeres y un comportamiento no asociado de las variables del cuarto clúster de variables.

Tabla 2: Matriz de correlación entre las variables latentes.

	Comp1	Comp2	Comp3	Comp4
Comp1	1	0.22	0.35	-0.04
Comp2	0.22	1	-0.11	-0.02
Comp3	0.35	-0.11	1	-0.04
Comp4	-0.04	-0.02	-0.04	1

En esta segunda etapa se realizan clústeres de individuos basados en el algoritmo clásico de k -medias a partir del conjunto de todas las variables originales. En primera instancia se observa el comportamiento del gasto de los cruceristas en cada clúster. Por último, a partir del método introducido por [Fraiman et al., 2008] y los clústeres de variables determinados en la primera etapa se seleccionan las variables determinantes de estos grupos. Se observa el impacto causado en la distribución del gasto por rubro en cada grupo si sólo utilizamos las variables seleccionadas. A partir del criterio coeficiente de Silhouette se optó por trabajar con 5 grupos.

En la Figura 1 se observa las diferentes distribuciones de gasto total en cada grupo. Se marcan con una línea horizontal la mediana y los cuartiles en cada caso. En la Tabla 3 se observa el número de turistas en cada uno de los grupos, en donde la mayor parte se encuentran contenidos en grupo 1 y el grupo 3 que representan más del 70% de todos los individuos. En la Figura 2 se puede

Tabla 3: Número de turistas en cada uno de los 5 clústeres.

	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5
Turistas	2356	166	2150	1049	486
Porcentaje	38%	3%	35%	17%	8%

observar el diferente comportamiento del gasto por rubro en cada grupo en relación al gasto total.

A través del algoritmo forward-backward introducido en [Fraiman et al., 2008], se seleccionan qué variables son las de mayor influencia en la determinación de los clústeres. Los pasos del algoritmo son los siguientes.

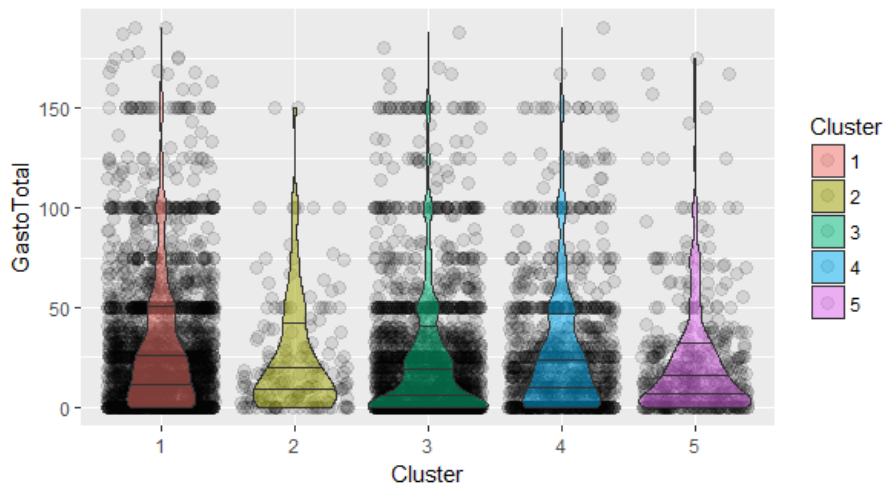


Figura 1: Distribución del gasto total en cada grupo.

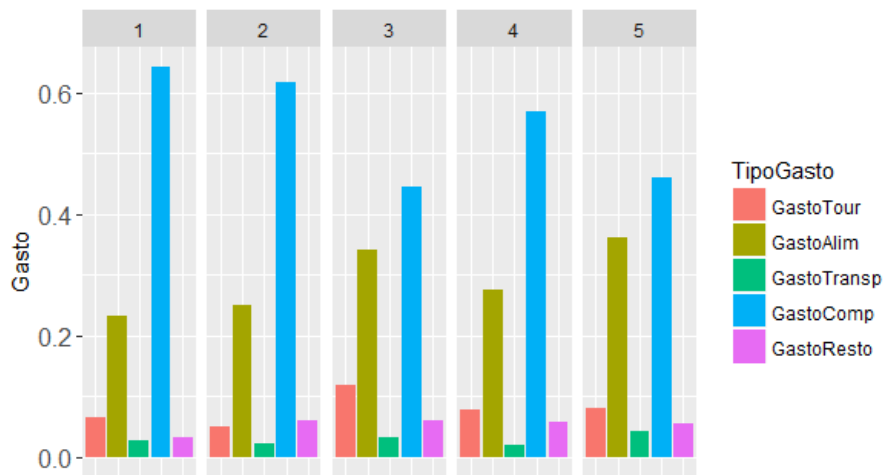


Figura 2: Proporción del gasto por rubro en cada clúster en relación al gasto total.

- (1) Se comienza por ingresar aquella variable “más influyente”, es decir aquella en que la asignación en los clústeres es más afectada por su ausencia. Se congela una a una cada variable (las restantes se sustituyen por la estimación de su esperanza condicional) y se elige aquella que produce menos variación en la determinación de los clústeres (respecto al grupo con el conjunto de variables).
- (2) Este procedimiento se reitera secuencialmente hasta superar un porcentaje fijado de bien clasificados respecto a los grupos originales.
- (3) El paso tres es la realización de una búsqueda backward. Las variables son sustituidas una a una por su esperanza condicional y se van eliminando aquellas que producen menor cambio en la asignación de los individuos.

Como indican los autores, el algoritmo es computacionalmente costoso. En tal sentido, el agregado de variables se realiza de acuerdo a los grupos de variables obtenidos en la primera etapa del

trabajo. Se comienza introduciendo la variable más asociada a la variable latente de cada grupo de variables y así sucesivamente hasta alcanzar el umbral de coincidencias fijado. Para estimar la media condicional se realiza el promedio del valor de los 20 vecinos más cercanos. A partir de dicho algoritmo, con las variables “entre30-44”, “entre45-59”, “TotPersonas”, “Mdeo”, “Residbra”, “mujeres”, “Jubilado”, “GustaMercPuerto” y “DesagradPrecios”, se logra reconstruir los grupos originales con una eficiencia del 85%.

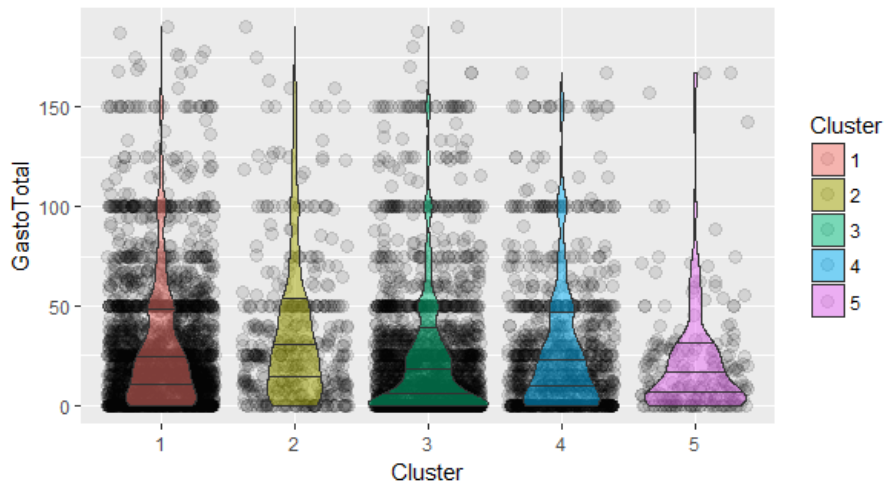


Figura 3: Distribución del gasto total en cada grupo (grupos determinados por k-medias sólo con las variables seleccionadas).

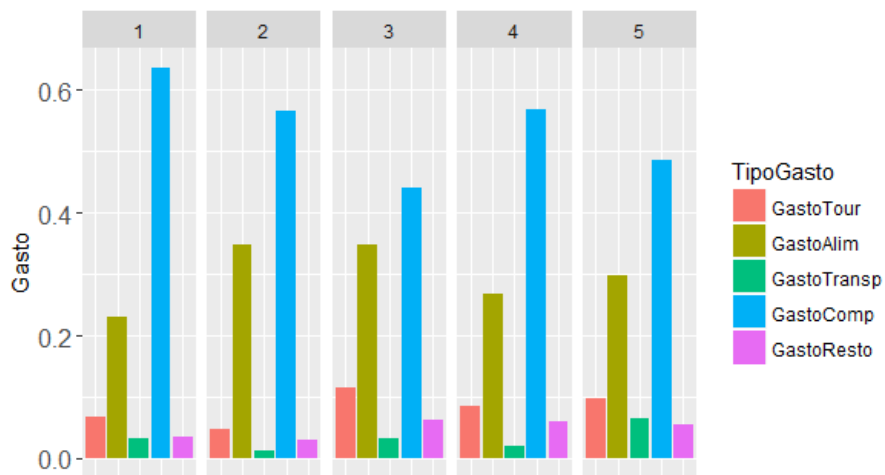


Figura 4: Proporción del gasto por rubro en cada clúster en relación al gasto total (grupos determinados por k-medias sólo con las variables seleccionadas).

5. Resultados

En esta sección se describen los principales hallazgos sobre la segmentación de los turistas cruceristas según la metodología utilizada y se comparación con la temporada 2012-2013.

5.1. Clúster de variables

Luego de podados los clústeres obtenemos 4 grupos . El grupo 1 está caracterizado por variables de edad y profesión. Las personas de mediana edad (entre 45 y 59) están relacionadas positivamente con la variable latente del grupo, mientras que las personas mayores (por encima de 60 años) están relacionadas negativamente. Este resultado se confirma con las variables de profesión, la relación positiva con las profesiones activas (empresarios y profesional) y la negativa con las profesiones inactivas (jubilados) y que además residen en países europeos.

El grupo 2 también tiene un componente fuerte relacionado con las variables de edad. Tiene una correlación directa con las personas de entre 30 y 45 años de nacionalidad brasilera que visitan por primera vez el país. La variable "Residente en América del norte" tiene una correlación negativa al igual que la variable que indica la repetición del destino.

El grupo 3 agrupa las variables de género y de tamaño del grupo. Todas las variables están correlacionadas positivamente. En este sentido, este grupo tiene un alto porcentaje de mujeres que viajan en grupos. Estos grupos de viajeros pueden estar representando a turistas que viajan en familia, con personas menores de edad y jóvenes de entre 18 y 29 años. La variable hombre también tiene una correlación positiva con el grupo, lo cual refuerza la interpretación de este grupo.

El grupo de variables más numeroso, el grupo 4, se puede identificar como el destino del crucero. En una dirección se encuentran aquellas variables relacionadas al destino Montevideo y en la dirección contraria están las variables relacionadas con el destino Punta del Este. De acuerdo con esto, los turistas que desembarcan y visitan Montevideo están satisfechos con el Mercado del Puerto, los Edificios, la Ciudad vieja y la comida y les desagrada la higiene y las pocas opciones de tiendas (que puede estar relacionado con la llegada de cruceros el fin de semana donde las tiendas cercanas a la zona portuaria están cerradas). Por otro lado, los turistas que desembarcan y visitan la ciudad de Punta del Este, están satisfechos con las variables Gusta Punta, Gusta Casa Pueblo, Gusta Higiene y Desagrada precios (concordante con la condición de Punta del Este de ser un destino caro). Adicionalmente, el desembarco en Punta del Este está asociado a los turistas residentes en Argentina.

En base a estos resultados se pudo identificar que las variables asociadas a la Edad, el Total de personas en el grupo de viaje, la Residencia y el Puerto de visita son las variables que determinan la formación de clústeres. La Edad está presente en tres de los cuatro grupos y en dos de ellos (clúster 1 y 2), es la variable con mayor correlación del clúster. Con las variables de Residencia sucede algo similar; está presente en los clústeres 1, 2 y 4. El total de personas en el grupo de viaje define el clúster 3 y el puerto de visita el clúster 4.

Esto dispara diferentes implicaciones a nivel de política económica y de marketing estratégico orientadas a satisfacer necesidades de grupos específicos de cruceristas y atraer a otros. Los resultados sugieren que sería rentable para el destino organizar actividades acordes a los diferentes grupos etarios y principalmente orientados a las familias. Con esto se busca que la familia o el grupo bajen del barco, conozca el destino y que todos se vean atraídos con lo que puede ofrecer el puerto visitado.

También hay claros indicios de que el turista se comporta diferente según su residencia. De modo que, los turistas europeos y jubilados podrían estar buscando actividades guiadas orientadas a conocer el acervo cultural, gastronómico, arquitectónico, etc., mientras que las personas de mediana edad buscarían actividades menos estructuradas.

Otros resultados referentes a la Residencia estarían indicando que el brasilero viene por primera vez y el argentino visita Punta del Este. Este resultado se podría interpretar de dos maneras, una es que el brasilero no es leal al destino y la otra es que tal vez vuelve como turista de estancia, lo cual conlleva a estrategias bien diferentes.

En cuanto a los argentinos, es probable que el no visitar Montevideo esté reflejando la cercanía entre ambos lugares. Estos resultados también refuerzan la idea de crear estrategias específicas para cada perfil de crucerista. Por último, el clúster 4 también refleja información importante. Es claro que las variables de higiene y precios deben ser revisadas por todos los actores involucrados.

5.2. Clúster de individuos y selección de variables

De acuerdo con los resultados en esta etapa, se pueden obtener algunas conclusiones interesantes. Los grupos 1, 3 y 4 son los grupos más parecidos en términos de gasto total y el 1 y el 3 son los que acumulan mayor cantidad de observaciones (2356 y 2150 respectivamente). Se puede observar que los valores de la mediana, cuartiles y máximos son similares. Adicionalmente, estos 3 grupos son los que tienen valores más dispersos en términos de la distancia entre el valor mínimo (cero) y el máximo. El grupo 3 y el grupo 5 son los que tienen mayor acumulación de valores cercanos a cero, lo que significa que los individuos en estos grupos son los que menos gastos realizaron. Además, el grupo 3 es uno de los que acumula mayor cantidad de observaciones indicando que en este grupo muchos individuos gastaron poco o nada. El grupo 2 tiene la menor cantidad de observaciones y es el más homogéneo de todos. Se puede observar también, sobre todo en los grupos 1, 3 y 4, que hay observaciones acumuladas en valores redondos. Este resultado es bastante intuitivo, ya que cuando se le pregunta al crucerista cuánto gastó, es probable que haga una estimación redonda de su gasto y no responda un resultado exacto.

En la figura 2 se puede observar que todos los grupos realizaron su mayor gasto en Compras, pero fue el grupo 1 el que mayor gasto hizo en este rubro (el 38% de los individuos destinaron más al 60% de su presupuesto en Compras). En el rubro Alimentos, fue el grupo 5 el que se destacó en comparación con el resto de los grupos, y el grupo 3 lo hizo en el rubro tour. El gasto en transporte fue bajo en todos los grupos. Esto puede ser debido a que, quienes usaron transporte lo hicieron como parte de un tour o también por el hecho de que las ciudades de visita son chicas. Pero, por otro lado, puede estar reflejando que el transporte no es amigable para el turista y optaron por realizar visitas en tour o cercanas al puerto. Este es un aspecto muy interesante para investigar ya que podría ser un punto a mejorar.

Para una mejor y más acertada interpretación de los resultados se optó por trabajar con el conjunto de todas las variables y no determinar los clústeres a partir de las variables latentes. Sin embargo, los clústeres de variables obtenidos en la segunda etapa de la sección metodológica, son también de vital importancia en el momento de la selección de variables, pues la correlación con la latente indica el orden de entrada en el algoritmo forward-backward de la siguiente sección.

Para la selección de variables, cuando se toma en cuenta las variables más influyentes de cada clúster, los resultados no varían demasiado. En los grupos 1, 3 y 5 la distribución del gasto total parecería ser la misma que cuando se toman en cuenta todas las variables del clúster. Mientras que, en los grupos 4 y 2 se puede observar que, en el primero existían observaciones atípicas relacionadas con variables que no eran extremadamente influyentes para armar el grupo. Una vez que esas variables se extrajeron del grupo, disminuyó el rango entre el valor mínimo y máximo. En el grupo 2 se puede observar que la gráfica se afina y se alarga reflejando un aumento en el rango mín-máx.

La caracterización de los grupos se puede visualizar en la Figura 5 a través de un árbol de decisión. Dentro de las variables consideradas en este trabajo, solamente el puerto de desembarco, la edad y la ocupación de Jubilado discriminan los diferentes grupos.

La primera variable y más importante que determina la primera partición es el puerto de

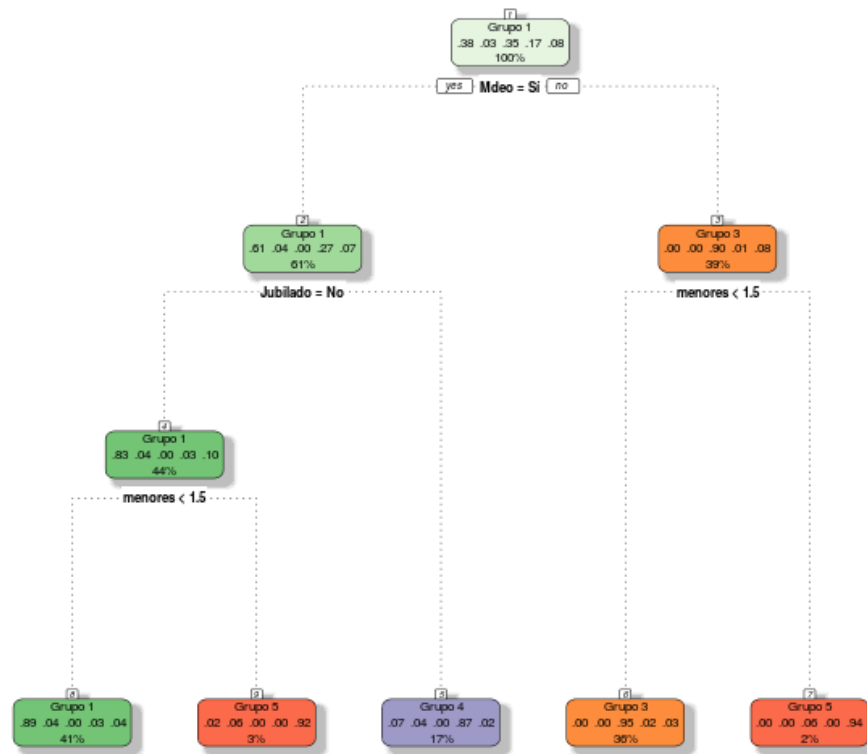


Figura 5: Proporción del gasto por rubro en cada clúster en relación al gasto total (grupos determinados por k-medias sólo con las variables seleccionadas).

desembarco. En la rama de la derecha, se encuentran aquellos que no desembarcaron en Montevideo (es decir, desembarcaron en Punta del Este) y se particiona nuevamente según la edad entre aquellos grupos de cruceristas que incluyen a lo sumo un menor de edad y el resto. Los mayores de edad son parte de casi la totalidad del Grupo 5 (94%) y el Grupo 3 esta conformado por los que tienen como máximo un menor (95%).

En la rama de la izquierda, están aquellos individuos que desembarcaron en Montevideo. Estos se particionan entre aquellos que son Jubilados y aquellos que no. La sub-rama de la derecha se caracteriza por los que son Jubilados que forma parte del 87% del Grupo 4 y los no jubilados, a su vez, se diferencian entre aquellos que son menores y los que no lo son. Los no jubilados y menores de edad conforman el 89% del Grupo 1 y los no jubilados mayores de edad conforman el 92% del Grupo 5.

Con estos resultados y los detallados en el punto anterior podemos caracterizar mejor a los grupos. En este sentido, podemos identificar que el Grupo 1 que gastó más del 60% de su presupuesto en Compras, son individuos que desembarcaron en Punta del Este y cuya ocupación no es Jubilado (es decir, es un individuo económicamente activo) y por tanto no son menores de edad.

El Grupo 3, que gastó en tours más que el resto de los grupos, está caracterizado por cruceristas con a lo sumo un menor de edad que desembarcaron en Punta del Este. También es interesante caracterizar el Grupo 4 y 5 que, si bien no incluye a la mayor parte de la muestra (25%

del total), arrojan resultados interesantes. El Grupo 4, que también realizó un gasto importante en compras, está compuesto por Jubilados que desembarcan en Punta del este y el Grupo 5, que fue el que realizó el gasto más alto en alimentos, se caracteriza por individuos mayores de edad, en edad activa que desembarcaron en Punta del Este y Menores que desembarcaron en Montevideo.

El grupo 2 son cruceristas que desembarcan en Montevideo pero no tienen alguna caracterización fuerte que los identifique entre ellos, podríamos denominarlo como un grupo de ruido que por su tamaño no es relevante para los objetivos del trabajo.

6. Conclusiones y trabajo a futuro

Este trabajo se enmarca dentro de la literatura sobre el turismo de cruceros, utilizando una metodología poco extendida en este rubro como lo es el análisis de clústeres. En este sentido, este estudio llegó a resultados que contribuyen a la literatura general sobre el turismo y particularmente a la literatura de la industria de cruceros. Adicionalmente, se encontraron resultados que permiten obtener implicancias a nivel de varios actores involucrados en esta actividad y que refuerzan conclusiones obtenidas en otros estudios analizados en la literatura. A continuación, se enumeran los principales hallazgos de la investigación:

- A partir de un procedimiento robusto de clúster de variables se pudo identificar aquellos grupos de variables con mayor asociación entre ellas, y por otro lado un grupo de variables de ruido.
- Ambas metodologías utilizadas para el clúster de variables nos proporcionan clústeres de variables similares, e incluso aquellas variables que pasan a formar el clúster de ruido son las menos asociadas a la variable latente de cada grupo en *clustofVar*.
- Los clústeres de variables determinan una fuerte dependencia entre la edad de los cruceristas, el tipo de grupo en que viajan y la ocupación de sus integrantes.

Los clústeres anteriores nos permiten demostrar que sería redituable para el destino armar estrategias específicas para cada grupo objetivo de cruceristas que llegan al país. Esto permitiría mejorar la relación entre el destino y los cruceristas explotando al máximo el tiempo que le dedican a las ciudades que visitan. De acuerdo con esto, los resultados demuestran que la existencia de familias en el crucero exige que el destino les ofrezca actividades amigables a todos los miembros.

Se encontró también que los turistas europeos son en su mayoría personas mayores a 60 años que podrían estar buscando realizar tours guiados orientados a conocer el acervo cultural, gastronómico, arquitectónico, etc.

Con respecto a la residencia, variable clave en la formación de clústeres, se encontró que el residente en Brasil llega a los puertos uruguayos por primera vez y el argentino está atraído mayormente por Punta del Este.

Las estrategias tendientes a promocionar el turismo de cruceros en los países vecinos, deberían orientarse al grupo de turistas brasileños que aún no han venido al país así como también lograr mayor lealtad en aquellos que ya lo han visitado.

Otros resultados importantes y en los cuales se debe poner foco es en los puntos débiles de los destinos, como higiene, precios y la falta de tiendas en Montevideo y en los puntos fuertes que fueron los atractivos de ambas ciudades. Estos resultados refuerzan la idea de crear actividades específicas según el destino y el perfil del turista que desembarca del crucero y la necesidad de trabajar sobre la sinergia de todos los involucrados para lograr el mayor beneficio posible.

Si comparamos con la temporada de 2011-2012 analizada por [Brida et al., 2014a] encontramos algunas diferencias en la conformación de los perfiles de turistas que llegaron en esa temporada con respecto a la analizada en este trabajo. Las diferencias se vinculan básicamente con dos variables: Precio y lealtad al destino. En cuanto a los precios, se puede identificar que en la temporada 2011-2012 los turistas no se sintieron insatisfechos con los mismos, mientras que en las temporadas entre

2014 y 2016 los turistas encontraron que los destinos tenían precios altos en especial Punta del este (concordante con el encarecimiento del país frente a otros países de la región). En el trabajo de [Brida, et al., 2014c] se encontró que los brasileños tenían lealtad con Punta del este mientras que en este trabajo se encontró que los brasileños llegaban por primera vez. Este resultado podría estar indicando que los brasileños dejaron de repetir el destino y una de las causas podría ser el encarecimiento relativo. En cuanto a la lealtad del destino, los argentinos siguen repitiendo el destino, en especial Punta del este

En lo que se refiere a los resultados económicos, se pudo identificar que:

- Gasto en Compras: todos los grupos encontraron oportunidades para gastar en el rubro compras por sobre cualquier otro rubro, a pesar de que algunos turistas se fueron insatisfechos por la falta de tiendas (particularmente en Montevideo). El perfil de crucerista que gastó más en este rubro fue el Grupo 1 caracterizado por turistas no jubilados (económicamente activos) que bajaron en Punta del este. Es de destacar que, si bien
- Punta del este está asociado con precios altos, los turistas encuentran opciones de gasto más interesantes que en Montevideo. Sería interesante estudiar en detalle la incidencia del precio en la decisión de compra de los turistas porque parecería no ser una variable que reduce la demanda de bienes y servicio y esto podría definir el producto que se le ofrece al turista.
- Gasto en Tours: fue muy chico, aspecto que se debería estudiar con mayor detalle ya que podría estar respondiendo a la insatisfacción de los turistas con los tours ofrecidos por los destinos. El gasto en tours está asociado con un perfil de crucerista que viaja en familia y desciende en Punta del este.
- Gasto en Transporte: también fue muy chico lo que podría estar indicando insatisfacción por las posibilidades de transporte con las que cuentan.
- Gasto en Alimentos: el grupo que gastó más en alimentos fue el 5 caracterizado por turistas jubilados que desembarcaron en Punta del este.

Una salvedad que se debe tener en cuenta con respecto al menor gasto encontrado en los rubros tours y transporte es que podría estar respondiendo a la incapacidad de la encuesta de captar este gasto ya que es usual que el crucerista contrate el tour dentro de las opciones que le ofrece el crucero.

A diferencia del trabajo de [Brida et al., 2014a], en este trabajo se segmentó a los individuos por rubro de gasto. De todos modos, a nivel de gasto total se encontró una gran diferencia en el perfil de crucerista con mayor nivel de gasto. En la temporada 2011-2012, el grupo de turistas que gastó más fue un brasileño o norteamericano, Jubilado y satisfecho con el destino que desembarcó en Montevideo. Mientras que en las temporadas entre 2014 y 2016, el perfil de ese crucerista que gastó más fue un turista con profesión económicamente activa que desembarcó en Punta del este.

Como trabajo a futuro, sería interesante poder integrar a la encuesta de cruceristas otras variables que ayudarían a complementar estos resultados, tal como, el nivel de ingresos, el nivel educativo, el tiempo en destino, desagregar las preguntas sobre tours y compras, actividades que le gustaría realizar y no tuvo oportunidad, si volvería como turista de estancia, etc. Esto permitiría conocer el perfil de turista con mayor nivel de detalle y como consecuencia se mejorarían los productos y servicios que se le pueden ofrecer.

Referencias bibliográficas

- Andriotis and Agiomirgianakis, 2010. Andriotis, K. and Agiomirgianakis, G. (2010). Cruise visitors' experience in a mediterranean port of call. *International Journal of Tourism Research*, 12(4):390–404.
- Bloom, 2004. Bloom, J. Z. (2004). Tourist market segmentation with linear and non-linear techniques. *Tourism Management*, 25(6):723–733.

- Brida et al., 2015a. Brida, J. G., Bukstein, D., and Tealde, E. (2015a). Exploring cruise ship passenger spending patterns in two uruguayan ports of call. *Current Issues in Tourism*, 18(7):684–700.
- Brida et al., 2015b. Brida, J. G., Bukstein, D., and Tealde, E. (2015b). Exploring cruise ship passenger spending patterns in two uruguayan ports of call. *Current Issues in Tourism*, 18(7):684–700.
- Brida et al., 2014a. Brida, J. G., Fasone, V., Scuderi, R., and Zapata-Aguirre, S. (2014a). Clustofvar and the segmentation of cruise passengers from mixed data: Some managerial implications. *Knowledge-Based Systems*, 70:128–136.
- Brida et al., 2018. Brida, J. G., Lanzilotta, B., Moreno, L., and Santiñaque, F. (2018). A non-linear approximation to the distribution of total expenditure distribution of cruise tourists in uruguay. *Tourism Management*, 69:62–68.
- Brida et al., 2012. Brida, J. G., Pulina, M., Riaño, E., and Zapata-Aguirre, S. (2012). Cruise passengers experience embarking in a caribbean home port. the case study of cartagena de indias. *Ocean & Coastal Management*, 55:135–145.
- Brida et al., 2014b. Brida, J. G., Scuderi, R., and Seijas, M. N. (2014b). Segmenting cruise passengers visiting uruguay: A factor–cluster analysis. *International Journal of Tourism Research*, 16(3):209–222.
- Cabilio and Masaro, 2001. Cabilio, P. and Masaro, J. (2001). Basic statistical procedures and tables.
- Chavent et al., 2017. Chavent, M., Kuentz, V., Liquef, B., Saracco, J., Chavent, M. M., and PCAmixdata, I. (2017). Package clustofvar.
- CLIA, 2018. CLIA (2018). 2018 cruise industry outlook.
- Dave, 1991. Dave, R. N. (1991). Characterization and detection of noise in clustering. *Pattern Recognition Letters*, 12(11):657–664.
- Dolnicar et al., 2012. Dolnicar, S., Kaiser, S., Lazarevski, K., and Leisch, F. (2012). Biclustering: Overcoming data dimensionality problems in market segmentation. *Journal of Travel Research*, 51(1):41–49.
- Fraiman et al., 2008. Fraiman, R., Justel, A., and Svarc, M. (2008). Selection of variables for cluster analysis and classification rules. *Journal of the American Statistical Association*, 103(483):1294–1303.
- Joshi and Kaur, 2013. Joshi, A. and Kaur, R. (2013). A review: Comparative study of various clustering techniques in data mining. *International Journal of Advanced Research in Computer Science and Software Engineering*, 3(3).
- Kiers, 1991. Kiers, H. A. (1991). Simple structure in component analysis techniques for mixtures of qualitative and quantitative variables. *Psychometrika*, 56(2):197–212.
- Kim et al., 2003. Kim, J., Wei, S., and Ruys, H. (2003). Segmenting the market of west australian senior tourists using an artificial neural network. *Tourism Management*, 24(1):25–34.
- Liao et al., 2012. Liao, S.-H., Chu, P.-H., and Hsiao, P.-Y. (2012). Data mining techniques and applications—a decade review from 2000 to 2011. *Expert systems with applications*, 39(12):11303–11311.
- Ministerio de Turismo, 2015. Ministerio de Turismo, U. (2015). Anuario 2015.
- Ngai et al., 2009. Ngai, E. W., Xiu, L., and Chau, D. C. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert systems with applications*, 36(2):2592–2602.
- Pesonen, 2012. Pesonen, J. A. (2012). Segmentation of rural tourists: Combining push and pull motivations. *Tourism and Hospitality Management*, 18(1):69–82.
- Romero and Ventura, 2013. Romero, C. and Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(1):12–27.
- Sánchez-Hernández et al., 2013. Sánchez-Hernández, G., Chiclana, F., Agell, N., and Aguado, J. C. (2013). Ranking and selection of unsupervised learning marketing segmentation. *Knowledge-based systems*, 44:20–33.
- Thurau et al., 2007. Thurau, B. B., Carver, A. D., Mangun, J. C., Basman, C. M., and Bauer, G. (2007). A market segmentation analysis of cruise ship tourists visiting the panama canal watershed: Opportunities for ecotourism development. *Journal of Ecotourism*, 6(1):1–18.
- Vigneau and Qannari, 2003. Vigneau, E. and Qannari, E. (2003). Clustering of variables around latent components. *Communications in Statistics-Simulation and Computation*, 32(4):1131–1150.
- Vigneau et al., 2016. Vigneau, E., Qannari, E., Navez, B., and Cottet, V. (2016). Segmentation of consumers in preference studies while setting aside atypical or irrelevant consumers. *Food quality and preference*, 47:54–63.

Anexo A. Anexo

Tabla 4: En la tabla siguiente se muestran las correlaciones de cada variable con la variable latente que determina el clúster y con la variable latente del clúster con la que presenta mayor correlación (sin contar el clúster del cuál forma parte). Por otro lado se marcan con una cruz aquellas variables que por el algoritmo de ClustofVar fueron asignadas a un clúster distinto y aquellas variables que posteriormente fueron podadas.

	Correlación	Corr con siguiente	ClustofVAR	Poda
Grupo 1				
entre45_59	0.91	0.33	-	-
mayores	-0.75	0.25	-	-
Jubilado	-0.51	0.28	-	-
Profesional	0.20	0.14	-	-
Empresario	0.18	0.04	-	-
Resid_euro	-0.17	0.13	-	-
Empleado	0.12	0.11	x	X
Ocup_Otros	0.07	0.04	-	X
AmadeCasa	0.04	0.03	x	X
Grupo 2				
entre30_44	0.99	0.35	-	-
Resid_bra	0.18	0.12	-	-
Resid_norte	-0.14	0.11	x	-
PrimeraVez	0.06	0.04	-	-
MasVeces	-0.06	0.04	-	-
Desocupado	0.04	0.01	x	X
DesagradAtencion	0.04	0.03	x	-
Grupo 3				
TotPersonas	0.99	0.36	-	-
mujeres	0.80	0.27	-	-
hombres	0.66	0.26	-	-
menores	0.53	0.16	-	-
entre18_29	0.27	0.05	x	-
Estudiante	0.06	0.03	x	x
Grupo 4				
Mdeo	0.98	0.03	-	-
Punta	-0.98	0.03	-	-
VisitMtdo	0.95	0.04	-	-
Visitpta	-0.92	0.05	-	-
GustaMercPuerto	0.22	0.02	-	-
GustaPunta	-0.18	0.02	-	-
GustaEdif	0.16	0.03	-	-
GustaCasapueblo	-0.16	0.02	-	-
DesagradPrecios	-0.16	0.05	-	-
GustaHigiene	-0.14	0.01	-	-
GustaMtdo	0.14	0.02	-	-
GustaGente	0.14	0.05	x	-
DesagradHigiene	0.14	0.01	-	-
Resid_arg	-0.13	0.09	x	-
GustaCiudadVieja	0.13	0.02	-	-
DesagradNoTiendas	0.11	0.01	-	-
Resid_Otros	0.08	0.07	x	-
GustaComida	0.08	0.04	-	-
GustaPaisaje	0.06	0.02	-	x
Resid_otrosAm	0.05	0.04	-	X
GustaOrganizacion	-0.04	0.02	X	X
GustaTranq	-0.03	0.01	-	X