

**II Jornadas Internacionales de
Estadística Aplicada
5 y 6 de Diciembre de 2019**

**Two-step clustering: aplicación para la identificación de demandantes
potenciales de microcréditos en Salta**

Autora: Chocobar, Emilia Belén

Facultad de Ciencias Económicas Jurídicas y Sociales. Universidad Nacional de Salta. Salta.

Datos de contacto: chocobar_belen@hotmail.com. Tel. 3874864692

RESUMEN

Este trabajo plantea una metodología innovadora para la identificación de los demandantes potenciales de microcréditos. Se emplea la herramienta estadística *two-step clustering*, para la obtención de grupos con características similares, a partir de los datos de la Encuesta de Deuda Social Argentina del año 2010. El modelo obtenido ofrece una manera válida y fiable de gestionar los datos y establecer perfiles de los demandantes.

Los resultados indican que, en Salta, en el año 2010, existieron 20344 demandantes potenciales de microcréditos. De esta manera se busca proveer información útil para las instituciones micro financieras y al sector público.

Palabras claves: two-step clustering, demandantes potenciales, microcréditos.



INTRODUCCIÓN

El término “microcrédito” comienza a utilizarse desde los años 70, cuando organizaciones como Grameen Bank, ACCION International y Opportunity International empezaron a otorgar préstamos de montos pequeños a emprendedores, principalmente mujeres, con una garantía de grupo. Esto permitió superar la falta de garantías reales exigidas por la banca tradicional.

Con el tiempo, el microcrédito se ha transformado en un importante instrumento de política pública y privada, con el propósito de reducir la pobreza y ayudar a los más desfavorecidos. Su desarrollo, evolución y crecimiento es de gran importancia, primordialmente, por razones de equidad e inclusión social, porque muchos sectores de países en desarrollo o pobres no cuentan con fácil acceso al mercado de crédito formal por no contar con garantías, por poseer empleos precarios e inestables. El acceso al microcrédito, es una condición necesaria, pero no suficiente, para erradicar la pobreza y alcanzar metas de inclusión social (Morduch, 1998, Quibria, 2012).

El sector microfinanciero argentino, en comparación, con otros países de la región, como Perú y Bolivia, se encuentra en un estado incipiente, subdesarrollado y muy limitado (Navajas y Tejerina, 2006, Grandes, 2014). Tampoco existe amplia información respecto al tamaño del mercado, principalmente el conocimiento de la demanda de microcréditos. Grandes (2014), sostiene que un gran problema del mercado microfinanciero es la falta de conocimiento de la demanda. Si bien, existe una encuesta (Encuesta de Deuda Social Argentina) con un bloque de inclusión financiera que permite identificar a los demandantes, trabajos como los Grandes, et al. (2010), Grandes y Satorre (2012), Grandes y Carballo (2013) y Carballo (2016), emplean metodologías que consisten en un proceso de *filtering*. Esto quiere decir, que del total de la población se seleccionan personas con determinadas características, a saber, individuos bajo la línea de pobreza, características ocupacionales, grado de bancarización, etc. En el ámbito internacional, la metodología empleada tiene las mismas estructura, diferenciando únicamente por la base de datos de cada país o región.¹

En concordancia con Carballo (2016), se considera que el contar con una buena caracterización y estimación de la demanda potencial de microcréditos es un aporte invaluable para el desarrollo del sector microfinanciero. De esta forma, con una demanda aproximada, y caracterizada ayudará a tener una oferta más efectiva, promoviendo mayores y mejores oportunidades al desarrollo de este mercado.

Es por estos motivos, que el presente trabajo busca calcular el valor aproximado de la demanda potencial de microcréditos para el año 2010, con los datos de la Encuesta de Deuda Social Argentina, pero con una metodología no empleada con anterioridad en este campo: two step clustering. Es una herramienta de exploración, diseñada para descubrir agrupaciones naturales en un conjunto de datos que se diferencia de métodos tradicionales de conglomeración, por el hecho de incorporar variables tanto categóricas como continuas.

El trabajo se estructura de la siguiente manera, en el primer apartado se explica la metodología two step clustering, utilizada para el cálculo de los demandantes potenciales. A continuación, se exponen el desarrollo del trabajo en conjunto con los

¹ Ver Hassan et al. (2015), Navajas y Tejerina (2006), Gahanghro y Khan (2015), Jassim et al. (2019), etc.

principales resultados, cantidad de clusters encontrados, características de cada uno de ellos y selección de los demandantes. Finalmente, se plantea la conclusión, observaciones y recomendaciones.

METODOLOGÍA

Análisis de conglomerados (clúster)

El Análisis Clúster es una técnica estadística multivariante de clasificación automática de datos. A partir de una serie de datos, se busca ubicar cada observación en grupos homogéneos, clusters. La cantidad de conglomerados no necesariamente son conocidos desde el principio de la investigación, pero si sugeridos por la esencia de los datos. El objetivo es agrupar elementos con el mayor nivel de homogeneidad en cada grupo y la máxima diferencia entre ellos. De esta manera, las observaciones con similitudes son asignadas a un mismo grupo, mientras que aquellos disimiles se ubican en otro distinto. (Fernández, 2011).

Two step clustering

El análisis de conglomerados en dos fases, también conocido como bietapico, o two-step clustering, es una herramienta que permite descubrir agrupaciones naturales de conjuntos de datos (Pérez, 2011). Fue planteado originalmente por Chiu, Fang Chen Wang y Jeris el año 2001 en su artículo “*A Robust and Scalable Cluster ing Algorithm for MixedType Attributes in Large Database Environment*” (pág.263-268).

Este método de dos fases tiene las siguientes características únicas respecto a otros métodos de análisis de conglomerados tradicionales.

- Determinación automática del número óptimo de conglomerados.
- Posibilidad de utilizar variables categóricas y continuas para la creación de los grupos. Suponiendo una distribución normal multinomial conjunta y asumiendo que las variables son independientes.
- Analiza gran cantidad de datos.

El procedimiento consta de los siguientes pasos:

• *Paso 1: Pre-clusters de los casos.* En esta etapa se procesan todos los registros utilizando el método de agrupación secuencial (Theodoridis y Koutroumbas 1999), analizando cada una de las observaciones y decidiendo si ese dato, debe incluirse en los grupos previamente creados o debe formar un nuevo cluster. Si el número de clúster se vuelve muy elevado entonces el criterio de distancia se incrementa y los clústeres cuya distancia es ahora menor que la distancia modificada, son fusionados. Este proceso se realiza a través del algoritmo BIRCH (Zhang et al. 1996) con el cual se crea un *Árbol de Características*. Este árbol está compuesto por diferentes niveles de nodos, donde cada uno de ellos contiene un número de entradas determinadas. Una entrada para un *Nodo Hoja* representa un subcluster y donde las entradas en los *Nodo Rama* sirven para guiar un nuevo registro al *Nodo Hoja* correcto. Las medidas de distancias utilizadas son: log-verosimilitud (para las variables categóricas) y la euclidiana (para las variables continuas).

El proceso comienza colocando el primer dato en el Nodo Raíz para encontrar la entrada más cercana y así alcanzar a su correspondiente Nodo Rama, aquí encuentra

una nueva entrada (la más próxima) y viaja por ella hasta su Nodo Hoja correspondiente. El proceso continúa hasta que se completa el análisis para todos los datos.

- Paso 2: Clúster de los casos, se aplica un modelo basado en una técnica jerárquica, la medida de proximidad es la función de log-verosimilitud. Los datos de entrada son los subclusters del paso 1, y los agrupa en el número de clusters idóneo. Para determinar el número de clusters en forma automática cada una de estas soluciones de clúster se compara utilizando el Criterio Bayesiano (BIC) de Schwarz o el Criterio de Información de Akaike (AIC) como criterio de agrupamiento.

DESARROLLO

Para la identificación de los demandantes potenciales de microcréditos se emplearon los datos de la Encuesta de Deuda Social Argentina (EDSA) del año 2010. La EDSA es elaborada por el Observatorio de la Deuda Social Argentina y se realiza anualmente desde 2004. La encuesta se efectúa cada año, sobre una muestra probabilística representativa de hogares y personas que en el período de referencia residen en diecisiete aglomerados urbanos del país. El área cubierta es una muestra representativa de conglomerados tales como el Área Metropolitana y Gran Buenos Aires y Áreas Metropolitanas de las provincias (por ejemplo, Gran Salta, Gran Córdoba, Gran Mendoza, etc).

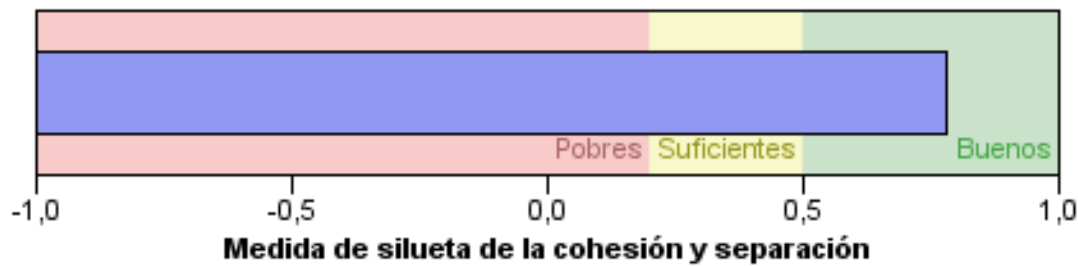
En el año 2010, esta encuesta, incluye el bloque de Inclusión Financiera en el capítulo correspondiente a “Trabajo y Seguridad Social”. En esta sección se incluyen preguntas que permiten distinguir si la persona desea tener un crédito, si ya solicitó uno, si le negaron el crédito, el origen institucional, etc, lo cual genera información idónea para el cálculo y la caracterización de la demanda potencial de microcréditos.

Las variables seleccionadas para la construcción de clusters son:

- Individuos cuyos ingresos no superan el valor del salario mínimo, vital y móvil (SMVYM) en ese año (\$2000),
- Si desea o no tener un crédito,
- Poseen o no cuenta bancaria,
- Poseen o no tarjeta de crédito de compra.

Al incorporar cada una de estas variables en el modelo, se obtuvieron 6 clusters, y como muestra la figura 1, la cantidad de conglomerados generados es buena ya que los datos evidencian de forma razonable y fuerte la estructura de dichas agrupaciones.

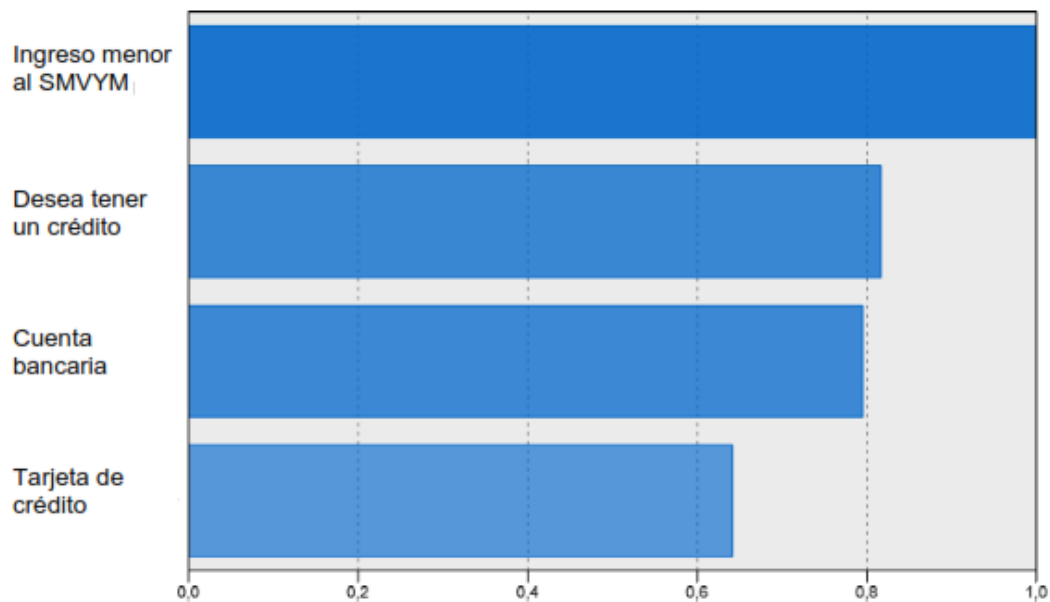
Figura 1. Calidad de clusters



Fuente: Elaboración propia.

La importancia de las variables predictoras está representada en el grafico 2, en donde la variable “ingreso menor al SMVYM” predomina como la más importante, con un 100% de importancia, para la clasificación y creación de los perfiles, seguida por el deseo de tener un crédito y el poseer una cuenta bancaria, con un nivel de importancia del 80%. Finalmente, con un poco más del 60% de importancia para el clustering, se encuentra la variable “tarjeta de crédito”.

Figura 2. Importancia del predictor



Fuente: Elaboración propia.

Selección del clúster

Para seleccionar el conglomerado correspondiente a los demandantes de microcréditos, se procedió a identificar rasgos distintivos de este grupo en cada uno de los clusters. Teniendo en cuenta la literatura antes analizada² y las características observadas de aquellos que poseen efectivamente un microcrédito en Salta (ver Anexo I) se consideran variables importantes a distinguir, personas con ingresos bajos, nivel educativo bajo (incluye secundario completo), situación ocupacional, no estar bancarizado y si tomaría un crédito.

² Ver PNUD (2005), Grandes et al. (2010)

Tabla 1. Resultados del modelo

		Clusters						
Características socioeconómicas		1	2	3	4	5	6	Total
N		26272	27368	30057	20344	45774	28733	178547
Porcentaje		0.15	0.15	0.17	0.11	0.26	0.16	100
Sexo	Varón	41.62	61.04	62.39	46.18	41.6	37.33	47.92
	Mujer	58.38	38.96	37.61	53.82	58.4	62.67	52.08
Edad	18-30	43.17	50.93	39.5	34.03	44.94	21.26	39.78
	31-40	11.93	17.83	33.91	19.96	11.69	31.29	20.48
	41-50	6.7	17.07	9.7	21.06	4.8	16.73	11.58
	50-60	12.33	6.43	9.81	14.55	20.21	5.47	12.15
	+60	25.86	7.73	7.08	10.4	18.36	25.25	16.03
Educación	Sin instrucción	0	0	0	0	4.32	0	1.11
	Primario incompleto	0	0	0	0	4.62	13.34	3.33
	Primario completo	0	1.65	11.63	38.37	16.67	24.65	14.83
	Secundario incompleto	23.69	23.16	14.63	39.33	16.76	0	18.28
	Secundario completo	21.19	12.24	34.9	16.61	36.53	43.41	29.11
	Terciario incompleto	0	11.96	17.11	0	8.1	7.67	8.02
	Terciario completo	15.14	22.48	11.89	5.68	7.68	6.46	11.33
	Universitario incompleto	23.51	14.74	4.92	0	2.8	4.47	7.99
	Universitario completo	16.47	13.77	4.92	0	2.53	0	6.01
Cobertura médica	Obra social	51.3	60.6	83.26	44.82	36.05	28.48	49.79
	Mutual	0	0	0	0	3.08	0	0.79
	Prepaga o Plan hospitalario privado	0	8.91	4.88	13.27	0	1.57	3.95
	PAMI	20.28	0	2.16	0	11.45	16.64	8.96

	No tiene	28.42	30.49	9.7	41.9	49.42	53.31	36.51
Situación ocupacional	Ocupado	53.17	62.91	77.8	43.68	72.95	72.9	65.97
	Desocupado (más de 6m)	0	2.64	0	0	4.06	0	1.45
	Ama de casa	11.2	14.79	8.12	38.64	2.53	0	10.33
	Jubilado	25.86	7.73	7.08	17.68	10.02	13.76	12.98
	Otros inactivos	9.77	11.92	7	0	10.44	13.34	9.27
Ocupación principal (solo ocupados)	Patrón	0	0	0	0	6.34	0	1.8
	En relación de dependencia	71.62	41.48	73.07	61	39.51	40.96	52.15
	Prof independiente	17.79	18.75	0	0	0	6.13	5.94
	Cuenta propia no profesional	10.59	33.05	20.6	25.98	44.16	40.03	31.77
	Empleado domestico	0	0	0	0	0	9.43	1.68
	Changas. Trabajos eventuales	0	6.72	0	0	4.43	3.45	2.85
	Otros	0	0	0	13.01	5.56	0	2.56
Ingreso total del hogar en el último mes	Media	5138.46	4441.67	3238.75	1264.29	3500	1144.74	
	Desvío	2296.56	2881.74	792.312	414.371	2641.02	541.319	
Rango de ingreso	<499	0	0	0	0	0	14.07	2.26
	500-999	0	0	0	11.35	0	20.3	4.56
	1000-1499	0	0	0	42.46	0	19.36	7.95
	1500-1999	0	0	0	46.19	0	46.27	12.71
	2000-2499	11.95	20.62	11.93	0	17.32	0	11.37
	2500-2999	0	7.69	9.41	0	17.67	0	7.29
	3000-3999	9.77	8.05	41.83	0	10.73	0	12.47
	4000-4999	6.7	8.01	6.68	0	0	0	3.34

	5000-7499	30.68	11.49	4.92	0	4.22	0	8.19
	>7500	40.9	44.13	25.23	0	50.06	0	29.86
Características crediticias								
Cuenta bancaria	Si	100	26.32	0	0	0	0	19.71
	No	0	73.68	100	100	100	100	80.29
Tarjeta de crédito	Si	78.95	47.37	100	78.57	0	0	44.78
	No	21.05	52.63	0	21.43	100	100	55.22
Solicitó un crédito	Si, para un emprendimiento	5.26	10.53	4.76	7.14	0	0	2.67
	Si, para vivienda	21.05	21.05	4.76	7.14	9.38	5.26	11.12
	No	73.68	68.42	90.48	85.71	90.63	89.47	85.48
	No responde	0	0	0	0	0	5.26	0.72
Institución que otorgó el crédito	Familiar o amigo	30.84	14.69	0.00	76.37	0.00	100.00	23.97
	Banco	69.16	85.31	100.00	0.00	100.00	0.00	74.20
	Asociación civil (ONG) de su barrio o cercanía	0.00	0.00	0.00	23.63	0.00	0.00	1.84
Monto	Valor promedio	2060	51100	8000	500	5333.3	3000	
	Desvío	1915.202	84547.62	2828.42	0	4163.3	0	
Tomaría un crédito	Si, con fin productivo	0	63.16	0	21.43	0	0	
	Si, para vivienda o consumo	0	36.84	0	28.57	0	0	
	No me interesa	100	0	100	50	100	100	

Fuente: Elaboración propia.

En base a los resultados descriptivos, se puede considerar que el clúster que representa a los demandantes potenciales de microcréditos en Salta es el número 4. Esto se debe a que es el conglomerado con mayor porcentaje de personas con nivel educativo hasta secundario completo; el promedio del ingreso, es el segundo más bajo de todos los clusters. Para cumplir con cierta capacidad de repago de la deuda, los ingresos de este grupo varían entre los \$500-\$2000. Acorde a PNUD (2005) se caracterizan a los potenciales demandantes como aquellos que están excluidos del sector bancario tradicional, por la falta de capacidades y garantías necesarias, en el clúster seleccionado se puede observar que ninguno de ellos posee una cuenta bancaria. En este grupo, solo el 7% solicitó un crédito con anterioridad, y el monto (en promedio) fue de \$500, el más bajo de los seis clusters. Adicionalmente, si se observa qué institución otorgó el préstamo, se puede notar que ninguno de ellos lo obtuvo en un banco, la mayoría lo recibió por un amigo o familiar, lo que reafirmaría el hecho de que son individuos excluidos de sistema bancario formal.

Para avanzar con el análisis se observaron qué tipos de trabajos realizaban los individuos en cada grupo. En el clúster 4 se identificaron las siguientes actividades: ama de casa, ayudante de pintor de autos, guardia cárcel, vendedor de áridos, pasante en el poder judicial, vendedor de libros y trabajadores independientes.

CONCLUSIÓN

Actualmente el sector microfinanciero en Argentina se encuentra subdesarrollado y como afirma Carballo (2016), contar con una buena estimación y caracterización de los individuos potenciales demandantes de microcréditos es una gran oportunidad de crecimiento y desarrollo de este sector.

La mayoría de las metodologías empleadas, hasta el momento, tanto en el ámbito nacional como en el internacional, se basan en procesos de *filtering*, del total de la población encuestadas se filtran los individuos que reúnen características básicas de quienes son la población objetivo de microcréditos. Sin embargo, en este trabajo se plantea una metodología totalmente diferente e innovadora, basada en una técnica estadística de análisis de cluster: *two-step clustering*.

A partir de los datos proporcionados por la Encuesta de Deuda Social Argentina del año 2010, se pudieron identificar grupos con rasgos particulares en base a las variables seleccionadas. A su vez, dentro de esta clasificación, se calcularon los estadísticos descriptivos de las características socioeconómicas y crediticias de los individuos de cada uno de los conglomerados. Esto permitió distinguir que el clúster número 4 es el que corresponde a los demandantes potenciales de microcréditos, ya que cuentan con los rasgos propios de los demandantes efectivos analizados en Anexo I.

Adicionalmente, de acuerdo a los resultados obtenidos, se puede considerar que el modelo seleccionado es el correcto, ya que la calidad de la cantidad de los clústers es buena, y la importancia de todas las variables predictoras es mayor al 0.6.

Esto da lugar a afirmar que esta nueva metodología, basada en una herramienta estadística de proceso de clustering, es efectiva y permite identificar correctamente al grupo objetivo. En este caso, para Salta en el año 2010, los demandantes potenciales son 20344, el 11% de la muestra representativa.

BIBLIOGRAFÍA

- Carballo, I. (2016). El federalismo en la demanda potencial de microcréditos en Argentina.
- Chiu, Tom & Fang, Dongping & Chen, John & Wang, Yao & Jeris, Christopher. (2001). *A Robust and Scalable Clustering Algorithm for Mixed Type Attributes in Large Database Environment*. Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 263-268. 10.1145/502512.502549.
- Fernández, S. (2011). *Análisis conglomerados*. Facultad de ciencias económicas y empresariales. Universidad Autónoma de Madrid. 2011
- Ghanghro, A. y Khan N. (2015). Estimating potential market size for microcredit in Pakistan.
- Grandes M. y Carballo I. (2013). *Nuevas estimaciones de la demanda potencial de microcréditos Argentina en 2011*. Universidad Católica Argentina. Facultad de Ciencias Económicas. Escuela de Negocios. Documento de Trabajo nº 9, 2013
- Grandes, M., Martiarena, A., Fariña, F. (2010). *¿Y dónde está la demanda? Una nueva metodología para cuantificar y caracterizar la demanda potencial por microcréditos en Argentina*. Documento de Trabajo de la Escuela de Negocios de la UCA 5.
- Grandes, M., Satorre M. (2011). *“¿Cuál es y cómo se caracteriza la demanda potencial de microcréditos en la Argentina? Nuevas estimaciones”* Argentina. Documento de Trabajo de la Escuela de Negocios de la UCA 7.
- Hassan, M.S., Ibrahim, K. (2015). *Sustaining small entrepreneurs through a microcredit program in Penang, Malaysia: A case study*. Journal of Human Behaviour in the Social Environment, 25(3), 182-191.
- Jassim J., Imtiaz, A. y Quereshi, A. (2019). *Estimating the potential market for microfinance in Pakistan including credit, savings, payments and insurance*.
- Molouny, L., Grandes, M. (2014). *Los determinantes de la demanda de microcréditos en Argentina*. Documento de trabajo No. 11. Escuela de Negocios. Facultad de Ciencias Económicas de la Universidad Católica Argentina. Disponible en: <http://bibliotecadigital.uca.edu.ar/repositorio/investigacion/determinantes-demanda>
- Morduch, J., (1998). *Does Microfinance Really Help the Poor? New Evidence from Flagship Programmes in Bangladesh*. Mimeo. New York University, Department of Economics
- Navajas, S. y Tejerina, L., (2006). *Microfinance in Latin America and the Caribbean: Connecting Supply and Demand*. Banco Interamericano de Desarrollo (BID), Washington DC.
- Pérez, C. (2011). *Técnicas de segmentación. Conceptos, herramientas y aplicaciones*. Madrid: Gaceta Grupo Editorial.
- Programa de las Naciones Unidas para el Desarrollo (2005). *Microfinanzas en la Argentina*, 1º edición, Buenos Aires: PNUD.
- Quibria M.G. (2012) *Microcredit and Poverty Alleviation: Can Microcredit Close the Deal?* Working Paper No. 2012/78. United Nations University-WIDER.
- Theodoridis, S. y Koutroumbas, K. (1999). *Pattern recognition*. Academic Press, New York.
- Zhang, T., Ramakrishnon, R. y Livny, M. (1996). *BIRCH: An efficient data clustering method for very large databases*. Proceedings of the ACM SIGMOD Conference on Management of Data. p. 103–114, Montreal, Canada

ANEXO I

En este apartado se realiza un análisis descriptivo de aquellas personas que poseen efectivamente microcréditos. Para ello se utilizan datos de encuestas realizadas a los beneficiarios salteños de las instituciones de CONAMI en el año 2018 y a las beneficiarias de PROMUJER 2017.

CARACTERISTICAS	POBLACION GENERAL	CONAMI		PROMUJER
		n=50		n=344
		Hombre	Mujer	
CARACTERISTICAS SOCIOECONOMICAS				
Sexo		10	90	
Edad				
20-30	8,10	0,00	13,50	9,10
31-40	21,00	50,00	16,20	20,40
41-50	31,20	0,00	24,30	23,40
51-60	21,80	50,00	18,90	31,00
60+	18,00	0,00	27,00	16,10
Nivel Educativo				
Primario		40,00	14,30	El 50 cuentan con estudios secundarios completos
Secundario		60,00	66,70	
Terciario		0,00	16,70	
Universitario		0,00	2,40	
Cobertura médica				
Obra social	39,90	20,00	64,30	38,20
Mutual	8,60	80,00	3,60	7,90
Planes y seguros	7,00	0,00	0,00	7,60
No paga ni le descuentan	42,90	0,00	10,70	46,20
Ns/Nc	1,60	0,00	21,40	0,00
Actividad				
Venta de comidas	28,30	0,00	36,10	27,60
Despensa, verdulería y pollería	15.9	0.0	13,90	15,60
Venta de ropa	14,50	0,00	22,20	13,60
Librería, peluquería y cosméticos			5,60	14,30
	13,30	0,00		
Artesanías	11,80	25	19,40	12,00
Otros	16,20	75	2,80	16,90
Actividad registrada	17,00	0,00	28,90	15,70
Ingreso total promedio				13592

CARACTERISTICAS CREDITICIAS				
No tienen cuenta bancaria	83,00	80,00	89,00	83,00
Disposición a acceder un préstamo				38,00
Disposición a solicitar un nuevo microcrédito	87,30	60,00	64,00	90,00
Monto promedio				\$10.000
Monto deseado respecto al crédito anterior				
Mayor				49,00
Igual				49,00
Menor				1,00