

ESTUDIOS ECONÓMICOS URBANOS_03

PAAE PATRONES DE AGLOMERACIÓN ESPACIAL

GEOGRAFÍA ECONÓMICA DEL PARTIDO DE GRAL SAN MARTÍN

DECANO

Marcelo Paz

Escuela de Economía y Negocios

DIRECTOR

Joaquín Ignacio Tomé

Centro de Estudios Económicos Urbanos

COORDINADOR

Roberto Arazo Silva

Centro de Estudios Económicos Urbanos

INVESTIGADORES

Roberto Arazo Silva

Bárbara Stern

María Emilia Saez Reynoso

Agustina Herrera

Centro de Estudios Económicos Urbanos

INVESTIGADOR ASOCIADO

Federico Catalano

Centro de Estudios Económicos Urbanos

EQUIPO DE INVESTIGACIÓN

Roberto Arazo Silva

Coordinación

Federico Catalano

Desarrollo Conceptual

Análisis Espacial y de Contenido

Bárbara Stern

Revisión General

Agustina Herrera

María Emilia Saez Reynoso

Diseño Gráfico y Editorial

EDITORIAL

Estudios Económicos Urbanos

N°03 - ISSN: 2683-975X

Escuela de Economía y Negocios (EEyN)

Universidad Nacional de San Martín (UNSAM)

Octubre 2019

Caseros 2241, San Martín, Prov. de Buenos Aires

inveeyn@unsam.edu.ar

tel. 4580 7250 int. 147

www.unsam.edu.ar/escuelas/economia/

ÍNDICE

00

INTRODUCCIÓN

Objetivos

3-4

3

01

METODOLOGÍA

Unidad de Análisis

La Base de Datos: Unidades Económicas Geolocalizadas

5-6

5

6

02

ANÁLISIS ESPACIAL DE NÚCLEOS DE ACTIVIDAD ECONÓMICA

Agrupamiento Espacial por Distancias

Agrupamiento Espacial por Atributos

7-20

7

10

03

REFLEXIONES FINALES

21-22



CENTRO DE ESTUDIOS ECONÓMICOS URBANOS

Motivada por su compromiso social y vocación de servicio frente a las problemáticas que aquejan a la Argentina, la Escuela de Economía y Negocios de la UNSAM crea el **Centro de Estudios Económicos Urbanos**. Este es un espacio dedicado a la investigación aplicada y al asesoramiento profesional, que aborda uno de los principales dilemas que enfrenta la sociedad contemporánea:

¿CÓMO HACER CIUDADES MÁS INCLUSIVAS Y SOSTENIBLES?



MISIÓN

El CEEU procura impulsar **proyectos de investigación aplicada** vinculados con la problemática urbana, y desarrollar productos, y servicios, para el asesoramiento de los distintos actores del sector público y privado, para posicionar en agenda **políticas públicas** a implementarse en materia de desarrollo urbano.

VISIÓN

Este espacio trabaja para lograr ciudades sostenibles e inclusivas, y busca transformarse en un centro en el que se **promueve el debate, la generación de datos abiertos**, la innovación y el conocimiento, para gestionar políticas urbanas sostenibles e inclusivas.

EJES

POLÍTICAS Y PROYECTOS URBANOS INTEGRALES

- planificación y gestión sostenible del suelo urbano
- inclusión socio-urbana y desarrollo comunitario
- proyectos urbanos integrales y participativos
- gestión ambiental, cambio climático y resiliencia
- estrategias en sistemas de transporte masivo en ciudades

GOBERNANZA Y GESTIÓN METROPOLITANA

- gobernanza metropolitana
- organización y fortalecimiento institucional
- agencias especializadas de desarrollo urbano
- herramientas de participación y gestión comunitaria
- financiamiento del desarrollo urbano

00_

INTRODUCCIÓN

A lo largo de la investigación, y a partir de diversos métodos de análisis espacial, se estudian los patrones de localización territorial de las distintas unidades económicas industriales, comerciales y de servicios del Partido de General San Martín. Además, considerando el reconocido perfil industrial del municipio, se indaga de manera particular en aquellas vinculadas a dicho sector, se identifican los principales núcleos de actividad y se verifica si estos siguen alguna lógica de localización específica.

Entre los principales resultados, se puede avanzar que si bien en la localización de las unidades económicas analizadas se reconocen áreas de mayor concentración, difícilmente se pueda afirmar que estas sigan una lógica regular en cuanto a sus atributos. En otras palabras, no se evidencian con claridad, zonas con mayor recurrencia de pequeñas empresas o algún tipo de vínculo entre distintos eslabones de una misma cadena productiva. Esto por mencionar sólo algunos ejemplos en relación a los hallazgos.

Las empresas del partido de General San Martín, obedecen a una lógica irregular y heterogénea a la hora de ubicarse en el territorio. El tejido industrial, por ejemplo, suele entre mezclarse con otras actividades. Si bien existen áreas con mayor concentración que otras, no se encuentra una referencia territorial en cuanto a cómo estas se geolocalizan, tanto por su cantidad de empleados o por los rubros de actividad a los que pertenecen.

OBJETIVOS Y MOTIVACIÓN

El análisis de la configuración espacial del entretejido económico en general, y del industrial en particular, en el municipio de General San Martín, es realizado por el Centro de Estudios Económicos Urbanos (CEEU) con la finalidad de complementar a partir de una perspectiva poco utilizada, las distintas investigaciones de carácter económico existentes sobre el partido. Esto es, reconocer cuáles son los resultados territoriales del proceso de localización de las unidades productivas en el distrito.

Para su concreción, es necesario plantear como objetivo principal, la identificación espacial de núcleos de actividad económica. Es decir, la ubicación de aquellas zonas que dentro del partido de General San Martín tienen un mayor dinamismo por aglomeración de distintos tipos de actividades.

Cabe destacar que en el 2014, San Martín es declarado capital nacional de la mediana y pequeña industria¹. Con esto, queda clara la relevancia que tienen las unidades económicas vinculadas a la producción industrial. Ahora bien, vale remarcar que esto no solamente es importante en términos de generación de empleo y en la conformación de un perfil económico específico del partido.

Esta tendencia tiene un correlato territorial que impacta directamente en la organización espacial y funcionamiento de distintos hechos urbanísticos del municipio: la accesibilidad y el transporte, la oferta de comercio y servicios, la distribución de

¹ Mediante la Ley 27.110 promulgada el 17 de diciembre de 2014 por el Senado y la Cámara de Diputados de la Nación.

viviendas, entre otros. La forma en que se localizan los diferentes eslabones productivos en el espacio, responden a lógicas escalares -económicas y urbanas-, tanto hacia dentro como hacia afuera de los límites administrativos del partido.

Razón por la cual, se considera relevante indagar en cómo se distribuyen las industrias dentro del Partido de General San Martín: ¿Siguen estas algún patrón específico? ¿Cuál es el alcance territorial de su localización? Esto es, ¿Se ubican en distintos sectores del municipio o están concentradas en zonas específicas?; Y, eventualmente, ¿Incide esto en el funcionamiento del sistema urbano del municipio?

Planteados estos interrogantes, también se busca entender si las lógicas de aglomeración espacial de las unidades económicas tenían algún correlato territorial para las ramas de actividad industrial; es decir, si las industrias del municipio también seguían alguna lógica de ubicación territorial específica.

Para llevar a cabo el análisis, se desarrolla un proceso de consolidación de información de unidades económicas del partido de General San Martín, a partir de un registro de información oficial y de un extenso trabajo de complementación de dicha base con atributos adicionales.

Posteriormente los registros son analizados con dos algoritmos de clustering espacial implementados en lenguaje de programación Python: DBSCAN y KMeans. Por medio de los mismos, se estudia si las unidades económicas previamente mencionadas siguen lógicas de aglomeración espacial tanto por distancia recíproca como por conjunción de un compendio específico de atributos.

01_

METODOLOGÍA

UNIDAD DE ANÁLISIS

La base de datos que se utiliza en el estudio, se construye a partir de una solicitud inicial de información a la Administración Nacional de la Seguridad Social (ANSES), con la finalidad de contabilizar las empresas pertenecientes al distrito de General San Martín. Dicha información es robustecida, con el propósito de entender algunos de los atributos básicos de estas unidades económicas, entre ellos, su perfil productivo y, fundamentalmente, su lógica de ubicación en el territorio.

Para tal propósito, se dispone de un conjunto de Claves Únicas de Identificación Tributaria (C.U.I.T.) radicadas en el partido de General San Martín. Esto incluye una serie de registros con el nombre de la razón social, su domicilio fiscal y la cantidad de empleados.

A su vez, la base es complementada con un arduo trabajo de data wrangling que se realiza a partir del scrapeo del sitio web www.cuitonline.com. La clave de identificación tributaria se utiliza para acceder a la portada de cada razón social, de la cual se extrae el número de actividades declaradas, con fecha de inicio y descripción. A partir del uso de expresiones regulares y de la librería de Python BeautifulSoup², se identifican los <div> que contienen la información detallada, y se consolida la base de datos con mayor información.

El proceso previamente descrito es fundamental para saber si, además de existir algún patrón específico de aglomeración territorial de las unidades productivas, la mismas se encuentran localizadas en función de la cercanía entre ramas de actividad. Para lo cual es esencial el reconocimiento del rubro al que pertenecen.

Por otro lado, en el domicilio fiscal registrado en la base para cada una de las firmas, se identifica el inconveniente de no seguir ningún patrón de notación estandarizada, sumado al hecho de no incluir la ubicación precisa en términos de latitud y longitud. Razón la cual se recurre al uso de la API de Geocoding de la plataforma Google Maps. De esta manera, no sólo se estandariza el conjunto de direcciones físicas, sino que también se utiliza la latitud y longitud de cada registro para poder geolocalizarlo en el mapa del partido de General San Martín.

El resultado de todos los pasos anotados, y en función de la finalidad del estudio, es la consolidación de la base de datos con los atributos principales de las unidades económicas de análisis: la localización espacial, la rama de actividad a la que pertenece y algún valor que permite denotar su tamaño o escala productiva. Para este último caso, la cantidad de empleados y de actividades declaradas para cada CUIT.

La construcción y análisis de la base datos previamente descrita, fue realizada en su totalidad en cuadernos de Jupyter con Python. Se prevé poder disponibilizar el código con el que se realiza el trabajo, una vez anonimizados los registros de la solicitud efectuada. Esto con el objetivo principal de compartir el código con el que se consolidan los registros.

² Para mayor información se puede consultar el siguiente vínculo: [Beautiful Soup](#)

³ Para profundizar sobre la herramienta se puede consultar el siguiente vínculo: [API de Geocoding](#)

LA BASE DE DATOS: UNIDADES ECONÓMICAS GEOLOCALIZADAS

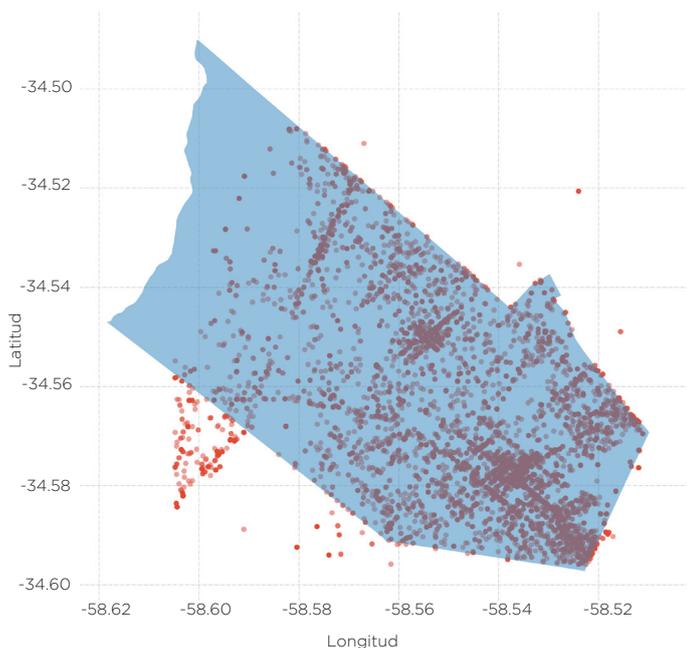
El resultado de las distintas etapas del proceso de data wrangling, se muestra a continuación, en los mapas con las unidades económicas geolocalizadas en el partido de General San Martín.

Vale aclarar que, en el mapa de la izquierda, los puntos rojos representan tanto el grupo de registros que fueron correctamente geolocalizados por la API de Google, así como aquellos por fuera de los límites administrativos del municipio. Con esto, también es importante mencionar que la base inicial contaba con 5.074 CUIT únicas. Cantidad

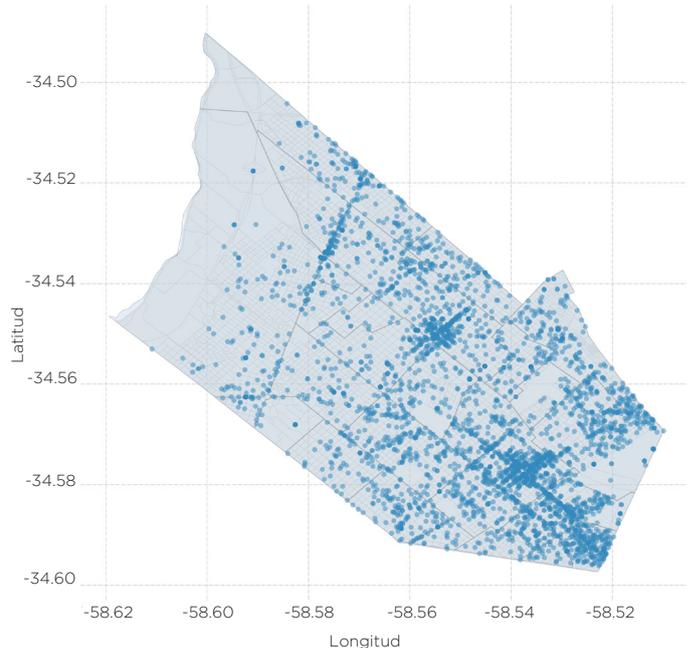
de observaciones a partir de la cual se establece el grupo de datos que tiene un margen de error demasiado alto en la geolocalización. Por lo cual, son descartados de la base de registros. Asimismo, se excluyen de la base las unidades económicas ubicadas, en su mayoría, en la localidad de Loma Hermosa (en las adyacencias de las extensiones de las rutas 4 y 8); si bien están correctamente especializadas, tampoco se encuentran en los límites administrativos de San Martín.

De esta forma, la base de datos se conforma con un total de 4.096 registros. Donde aproximadamente el 35% (1.395 registros) del universo total de la muestra, obedece a la actividad industrial; sector y observaciones en los que en el estudio se enfatiza.

M1 | TOTAL UNIDADES ECONÓMICAS GEOLOCALIZADAS



M2 | BASE FINAL UNIDADES ECONÓMICAS A ANALIZAR



Fuente: Elaboración propia.

02_

ANÁLISIS ESPACIAL DE NÚCLEOS DE ACTIVIDAD ECONÓMICA

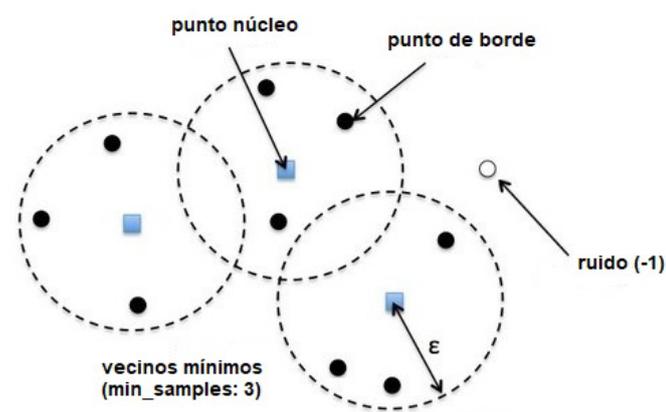
AGRUPAMIENTO ESPACIAL POR DISTANCIAS

El primer método de análisis aplicado para identificar zonas con mayor densidad de actividad económica, es un algoritmo comúnmente utilizado en el campo del “clustering”⁴ espacial. De esta manera, la base de datos descrita en las secciones previas, es estudiada con “Density-Based Spatial

Clustering of Applications with Noise” (DBSCAN)⁵; una herramienta de machine learning desarrollada para detectar patrones de agrupación en un espacio geográfico determinado.

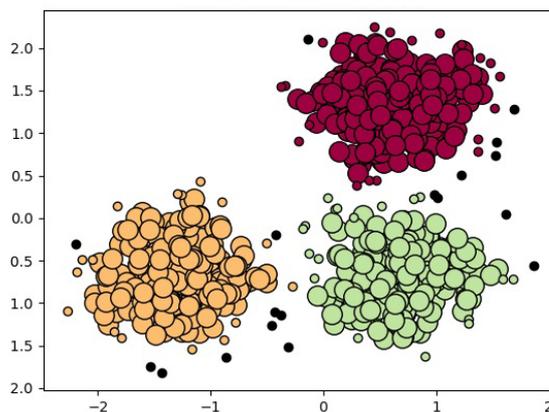
Para ello se utiliza Scikit Learn, una biblioteca de aprendizaje automático implementada en el lenguaje de programación Python. En este paquete, DBSCAN funciona esencialmente a partir de dos parámetros principales: epsilon y min_samples⁶, lo que determina el radio de proximidad y el mínimo de vecinos a considerar.

I1 | ESQUEMA DEL MÉTODO DBSCAN



Fuente: [Simon & Zebin](#).

I2 | EJEMPLIFICACIÓN DEL ALGORITMO DE CLUSTERING DBSCAN



Fuente: [Scikit Learn](#).

⁴ Para profundizar sobre el método de análisis se puede seguir el vínculo: [Clustering](#)

⁵ Para mayor información sobre la herramienta se puede consultar el siguiente vínculo: [DBSCAN](#)

⁶ Para profundizar en los parámetros del algoritmo se sugiere revisar la documentación oficial en: [Parámetros Cluster DBSCAN](#)

Tal como se explica en la documentación oficial, el aspecto más relevante de DBSCAN es el concepto de muestras núcleo. Es decir, aquellas que se encuentran en áreas de alta densidad. Así, un clúster puede ser definido como tal, a partir de un conjunto de puntos que por medio de un atributo de distancia⁷, se consideran cercanos o próximos entre sí. En este marco, tanto `min_samples` como `eps`, son los parámetros con los que dispone el algoritmo para definir la noción de densidad. Con valores de `min_samples` más altos o de `eps` más bajos, se puede establecer un nivel de densidad mayor para formar un clúster.

De esta forma, existiendo un número mínimo de muestras, muchos puntos que se encuentren a una distancia `eps` de otro punto dado, conformarán clusters considerando a este último como punto núcleo. Los puntos que se encuentren en el borde o cercanos a la frontera del radio considerado se denominan “de borde”. Los que no entren dentro de la agrupación, se consideran outliers o ruido (-1).

Este algoritmo, también cuenta con una métrica que permite evaluar la calidad del agrupamiento dentro del cluster conformado. Este se conoce como *silhouette coefficient*⁸ o coeficiente de silueta, y evalúa qué tan cerca está un punto de sus compañeros en comparación con qué tan cerca está de los puntos en el grupo más cercano. Un alto coeficiente de silueta indica clusters bien definidos, y un valor bajo indica un nivel de agrupamiento débil o atípico (Geoff Boeing, 2016).

Este coeficiente se compone de dos métricas. La distancia media entre un núcleo y todos los demás puntos de la misma clase (a), y la distancia media entre dicho núcleo y todos los demás puntos en el siguiente grupo más cercano (b). Como se puede apreciar en la fórmula del coeficiente, la diferencia entre ambas es normalizada por el máximo entre a y b. En el divisor se evalúa el máximo siendo {(a si a>=b) o (b si a<b)}. En términos más sencillos, la diferencia entre ambas medias se expresa en función de la mayor de ellas:

$$s = \frac{b - a}{\max(a, b)}$$

Como se menciona anteriormente, DBSCAN es un algoritmo comúnmente utilizado para la detección de centros de actividad. En líneas generales, el mismo trabaja a partir de la proximidad entre puntos de interés teniendo en cuenta un radio de distancia y una cantidad mínima de vecinos. Con esto, es importante aclarar que no hay una combinación ideal de parámetros para la identificación de centros de actividad o clusters. Sino antes, una correcta interpretación de la herramienta ajustada a la naturaleza del problema que se quiere abordar. Sólo por mencionar un aspecto, un radio de 500m o 10km, puede ser mucho o poco, en función del área que se está estudiando y del tipo de puntos que se quiere evaluar.

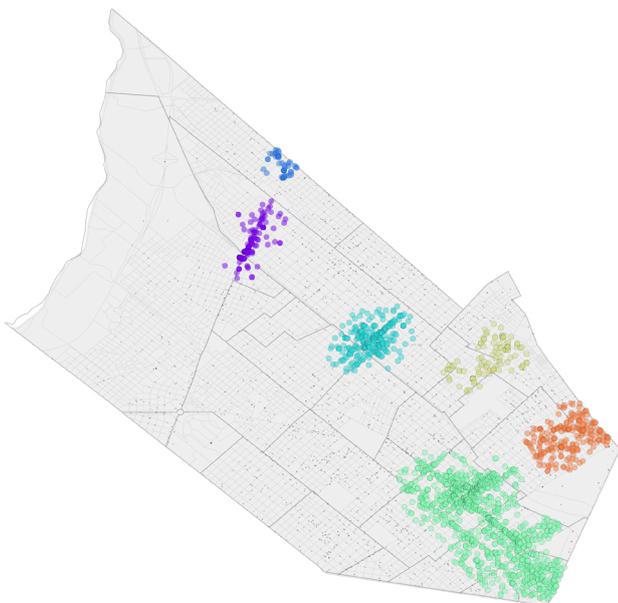
⁷ Se dice atributo porque para computar distancia el dataset debe contar con una columna de latitud y otra de longitud.

⁸ Para mayor información se puede consultar el siguiente vínculo: [Silhouette Coefficient](#)

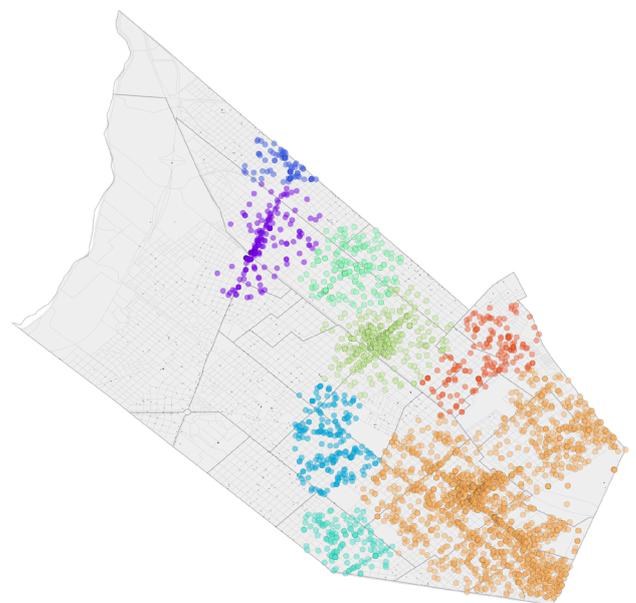
Así, es factible apreciar que con un “epsilon” de 300 metros y un “min_samples” de 50, el algoritmo arroja un total de 6 clusters con un silhouette coefficient de 0.155. Ahora bien, esto ¿es mucho o poco? Como se establece inicialmente, no existe una fórmula precisa, más bien distintos niveles de exigencia con los que opera el algoritmo. De esta manera, y considerando como parámetros un mínimo de 50 empresas registradas en un radio de hasta 300 metros, el mapa 3 evidencia que los principales centros de actividad económica se encuentran en Villa Lynch, San Martín centro, Villa Ballester y la Ruta 4 (en las inmediaciones de José León Suárez).

Al ajustar nuevamente los parámetros del algoritmo, ahora con un “epsilon” de 500 metros y un “min_samples” de 70 firmas vecinas, el silhouette coefficient asciende a 0.284. Un resultado un poco más elevado en comparación al caso anterior. Ahora, si bien el escenario mejora y el número de clusters aumenta, el patrón general de aglomeración espacial sigue siendo el mismo. Los principales centros de actividad económica no cambian, sino que se extienden los grupos o categorías en relación al primer análisis. Siempre, con un protagonismo marcado de las localidades de Villa Lynch, San Martín centro, Villa Ballester y José León Suárez.

M3 | CENTROS DE ACTIVIDAD ECONÓMICA EN 6 CLUSTERS



M4 | CENTROS DE ACTIVIDAD ECONÓMICA EN 8 CLUSTERS



AGRUPAMIENTO ESPACIAL POR ATRIBUTOS

En segunda instancia se implementa el método de análisis KMeans⁹, que a diferencia de DBSCAN, es un algoritmo de clasificación que trabaja resumiendo estadísticamente las características principales del fenómeno a clasificar. Tal como lo define Dani Arribas:

“La idea central de la agrupación estadística es sintetizar la información contenida en varias variables creando un número relativamente pequeño de categorías. Cada observación en el conjunto de datos se asigna a una, y solo a una categoría, dependiendo de sus valores para las variables originalmente consideradas en la clasificación.” (Arribas, 2016)

Así, se busca reducir la complejidad de un problema multidimensional. Es decir, identificar variables significativas y evaluar su comportamiento para regiones o divisiones administrativas específicas.

Otra diferencia relevante respecto de DBSCAN es que, KMeans se implementa definiendo como un parámetro previo, la cantidad de clusters en los que se desea resumir la información. Esto es, en lugar de agrupar entidades geográficas por proximidad en un radio de distancia determinada, ahora se evalúa cómo se comportan las características de las empresas para dos divisiones administrativas de tamaño diferente.

De tal manera, se verifica si a valores similares de las variables le corresponde algún tipo de lógica de agrupamiento espacial. Para ello, se plantea trabajar con dos escalas o niveles de granularidad. Uno más agregado (fracciones censales) y otro más desagregado (radios censales). Esta comparación, se lleva a cabo con la finalidad de ver cómo el algoritmo construye categorías para distintos niveles de agregación.

Tal comparación permite evaluar cuán significativos son entre sí los clusters que arroja K-Means. Es decir, comparando cómo se compone cada categoría de los agrupamientos devueltos por el algoritmo, se puede ver si existe alguna coincidencia geográfica entre ambos niveles de agregación. Si se asemejan entre sí, seguramente lo que suceda a nivel de radio censal tenga un comportamiento similar al de las fracciones correspondientes. De lo contrario, el nivel de dispersión para los valores de las variables seleccionadas en el clustering podría ser muy alto y, seguramente, el agrupamiento resultante muy poco claro o definido. Esto es, una baja capacidad explicativa de las variables seleccionadas para evaluar algún tipo de lógica de agrupación geográfica.

Para llevar a cabo la clasificación, se eligen como variables principales la cantidad promedio de empleados y actividades declaradas por empresa en cada unidad administrativa (radios y fracciones censales), asimismo, el porcentaje de empresas por rama de actividad y tamaño.

En primer lugar, se evalúa cómo el algoritmo (KMeans) agrupa las mencionadas variables a nivel de fracción censal. Tal como se puede apreciar, la mayor cantidad de fracciones se agrupan en la categoría 2. Además, en su distribución territorial se observa como en gran medida coinciden con las zonas para usos industriales aprobadas en el código de ordenamiento urbano del partido. A su vez, como dicha categoría rodea (a manera de envolvente) a la segunda categoría con mayor frecuencia de casos (la número 3).

⁹ Para profundizar sobre el algoritmo se puede consultar el siguiente vínculo: [KMeans](#)

Analizando este patrón con mayor detenimiento, es factible suponer que el núcleo conformado por la categoría 3, refleja la condensación de actividades en tres de los sectores más vitales del municipio: José León Suárez, Villa Ballester y San Martín centro. Esto es, una dinámica vista previamente en los resultados arrojados por DBSCAN.

Ahora cabe profundizar en, ¿qué representan cada una de estas categorías o, en otras palabras, en qué valores de las mencionadas variables estos clusters se diferencian?, ¿en la cantidad de actividades, en su tamaño y/o en la rama a la que pertenecen?

Para analizar la composición de los clusters, se aborda en primera instancia el qué sucede con el promedio de empleados y actividades declaradas por empresa o razón social.

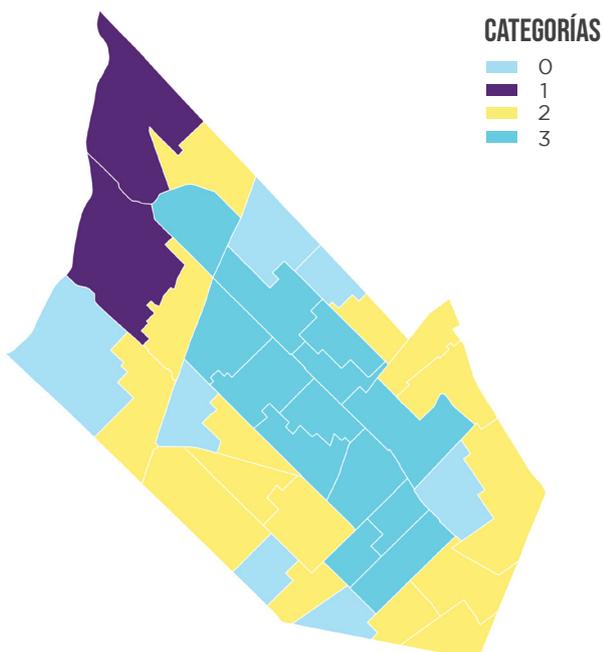
Como se puede ver, no se reconocen grandes diferencias entre los valores de cada cluster. Es

decir que, por ejemplo, si bien el cluster número 2 es el que más empleados en promedio tiene, también es claro que la distancia respecto del valor más bajo no es muy grande (tan sólo 4 teniendo en cuenta el cluster número 0). Por su parte, en relación al número de actividades, se concentran entre uno y dos, a partir de lo cual es viable pensar que son las cantidades más frecuentes declaradas por las empresas la muestra.

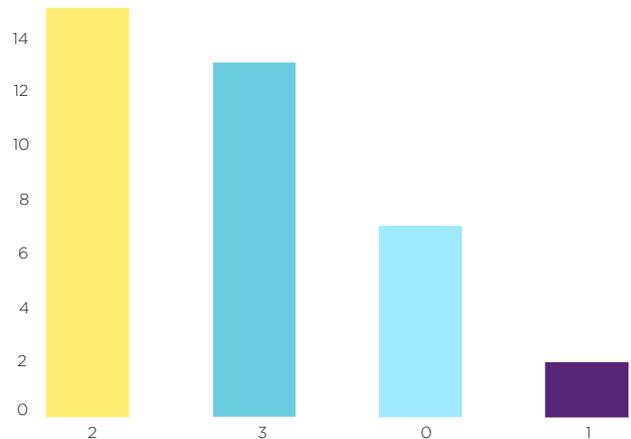
Con esto, se puede estimar que la muestra analizada probablemente se encuentre poblada de casos muy heterogéneos. Principalmente al tener en cuenta que la misma concentra industrias, comercios y servicios de distintos tipos y escalas. Aspecto que también se refleja en la composición de los clusters por tamaño y rubro de las razones sociales.

En cuanto a la variable tamaño de la firmas, es necesario anotar que para el desarrollo del estudio,

M5 | CLUSTERIZACIÓN ESPACIAL EN 4 CATEGORÍAS A NIVEL DE FRACCIÓN CENSAL



G1 | CANTIDAD DE FRACCIONES CENSALES POR CLUSTER



las empresas denominadas grandes (aquellas con más de 250 empleados), no se contemplan en la muestra¹⁰. De este modo, se observa que la mayor cantidad de firmas son pequeñas o microempresas. Hecho que refuerza la imagen de una muestra bastante heterogénea y, pone en evidencia, la dificultad de clasificar unidades económicas a partir de valores numéricos que reflejen atributos.

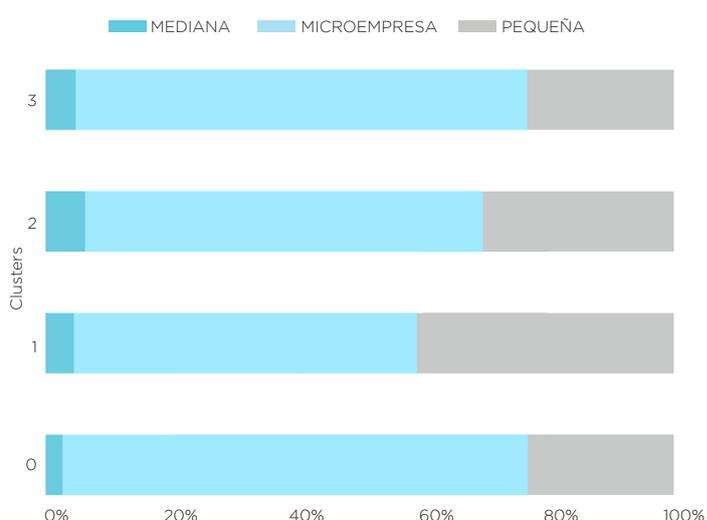
Respecto a los rubros, se identifica que en el conjunto de clusters existe cierta recurrencia de la rama “industria manufacturera”. Esto con excepción del número uno, donde la mayor proporción de casos corresponden a la rama “construcción”. También se reconoce la importancia de la rama “reparación de vehículos y automotores”. Se puede considerar que principalmente en estos tres rubros, oscilan las ramas de actividad entre los cuatro clusters. Con mayores y menores proporciones, pero con una constante presencia relativa superior.

T1 | PROMEDIO DE EMPLEADOS Y ACTIVIDADES POR CLUSTER (FRACCIONES)

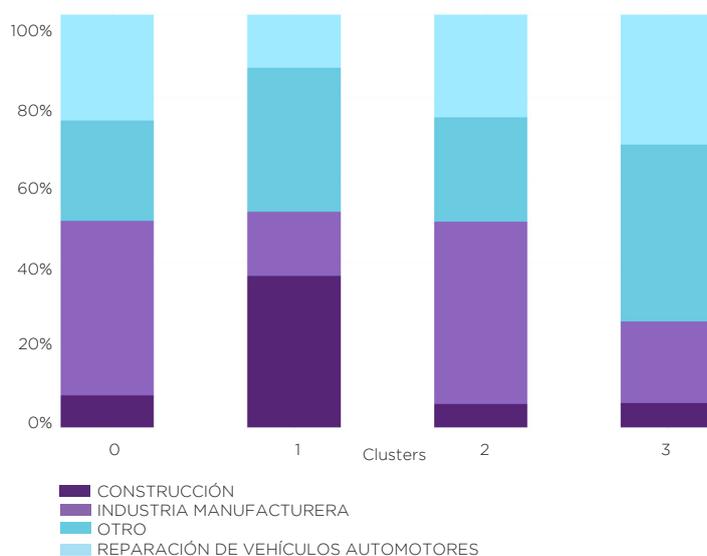
CLUSTER	EMPLEADOS	ACTIVIDADES
0	10	1
1	13	2
2	14	1
3	12	2

¹⁰ El enfoque territorial del trabajo demanda desestimar empresas con un número demasiado alto de empleados. Esto, al entender que al CUIT de cada firma se le asigna una dirección física fiscal, la cual comúnmente corresponde a una sucursal principal. A partir de esto, se supone que todos los empleados registrados para la firma se atribuyen de igual manera a esta geolocalización, la cual no necesariamente representa a todo un conjunto de sucursales. En otras palabras, la información de toda la firma se asocia en un solo punto en el espacio, mientras que es muy probable que en el territorio esto no funcione de la misma forma.

G2 | PORCENTAJE DE EMPRESAS POR TAMAÑO SEGUN CLUSTER (FRACCIÓN CENSAL)



G3 | PORCENTAJE DE EMPRESAS POR RAMA SEGUN CLUSTER (FRACCIÓN CENSAL)



Con base en lo anterior, cabe preguntarse si el comportamiento de tales valores corresponden a circunstancias arbitrarias, producto de la división administrativa, o bien a patrones geográficos específicos. Por lo pronto, es oportuno saber si el comportamiento se mantiene al analizar una división administrativa de mayor desagregación o menor escala, como los radios censales.

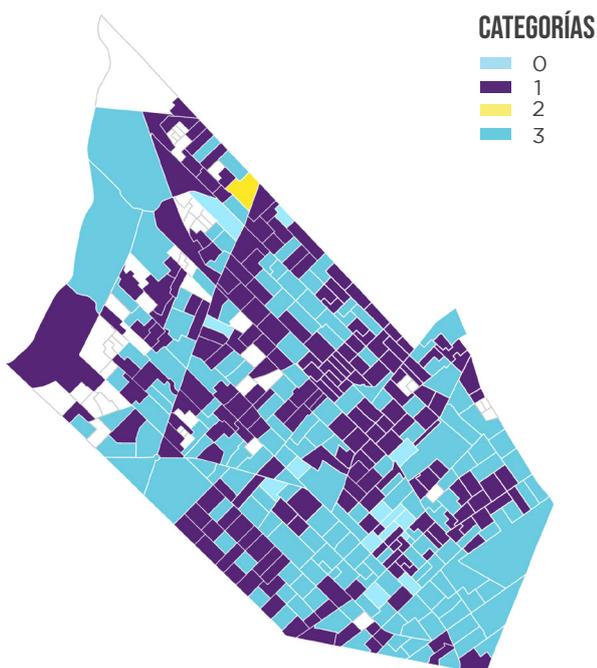
Esto implica entonces que los cluster se conforman a partir de una mayor cantidad de unidades territoriales (tantas como radios haya), las cuales a su vez, están compuestas por un menor número de casos cada una (en comparación con las fracciones censales). De tal forma, se evidencia que no existe un patrón espacial claro, sino más bien islas donde las variables de target se suponen más recurrentes. Esto, con una prevalencia general de los clusters 1 y 3.

En líneas generales, se ve que el comportamiento es bastante similar al identificado en las fracciones censales. Sobre todo en lo que respecta a la cantidad de actividades. Es decir, una mayor recurrencia de firmas con entre uno y dos actividades declaradas. Lo que nuevamente pone en evidencia una muestra compuesta por establecimientos pequeños y, por ende, heterogéneos entre sí.

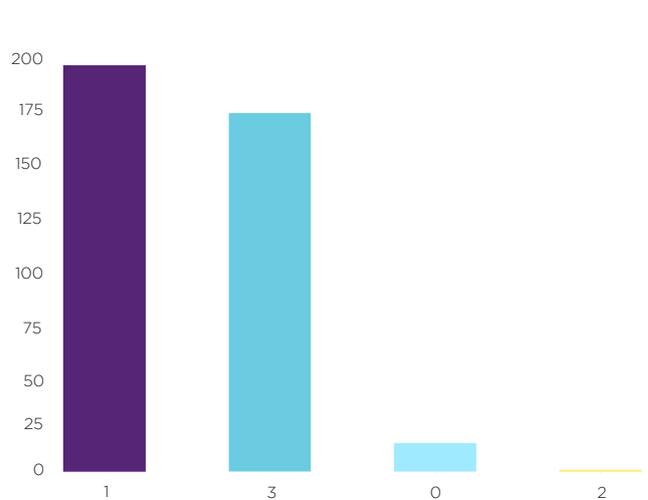
No obstante, si se encuentra una diferencia en los resultados por el cambio de escala territorial de análisis, la cual se refleja en la distribución de la cantidad de empleados. Al minimizar la unidades administrativas de corte, los promedios de empleados varían un poco más entre sí (siendo 18 el valor medio más alto y 6 el más bajo).

En relación al tamaño (según empleados) y rama de las empresas, la dinámica no dista demasiado de la expuesta con fracciones censales. Esto es

M6 | CLUSTERIZACIÓN ESPACIAL EN 4 CATEGORÍAS POR RADIO CENSAL



G4 | CANTIDAD DE RADIOS CENSALES POR CATEGORÍA

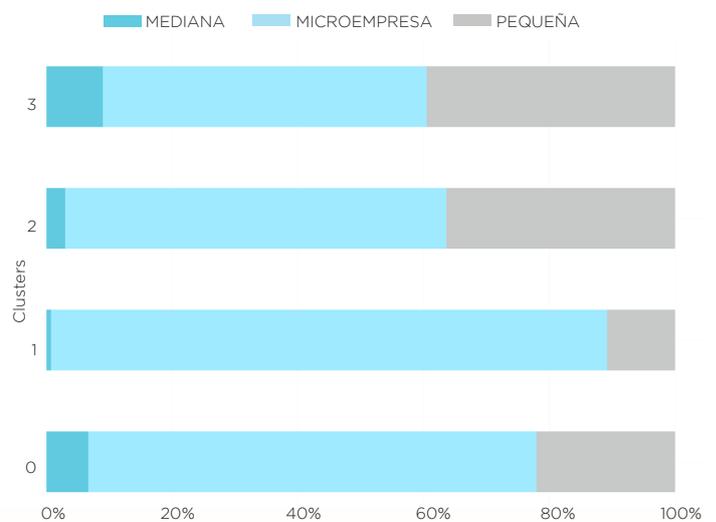


prevalencia de pequeñas y microempresas por un lado, y de ciertas ramas de actividad por el otro (las mismas que en el caso anterior).

Al tomar como referencia los dos niveles territoriales de análisis y los resultados hallados hasta ahora, se reconoce cierta dificultad para la obtención de diferencias entre patrones de localización espacial con valores tan heterogéneos. Aunque también es factible interpretar que puede ser un argumento suficiente para tratar de identificar la existencia de algún comportamiento específico en universos más reducidos.

Hasta el momento, sólo se ha trabajado la noción de aglomeración espacial en términos generales por unidades territoriales, para tratar de identificar clusters de actividad económica. Ahora bien, algo que también aporta al estudio de los patrones, es si las firmas responden a alguna lógica de agrupación

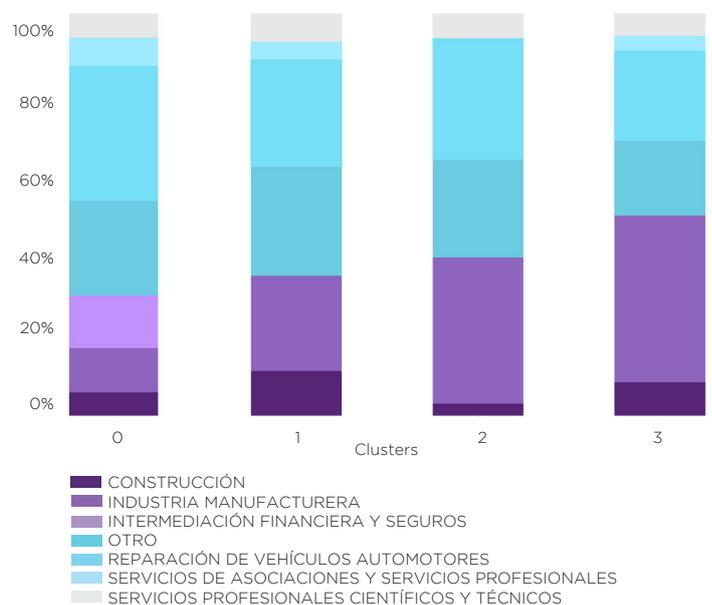
G5 | PORCENTAJE DE EMPRESAS POR TAMAÑO SEGÚN CLUSTER (RADIO CENSAL)



T2 | PROMEDIO DE EMPLEADOS Y ACTIVIDADES POR CLUSTER (RADIOS)

CLUSTER	EMPLEADOS	ACTIVIDADES
0	13	2
1	6	1
2	14	2
3	18	1

G6 | PORCENTAJE DE EMPRESAS POR RAMA SEGÚN CLUSTER (RADIO CENSAL)



por sector de actividad económica. Para ello se evalúa si aquellas razones sociales vinculadas a actividades netamente industriales, respetan alguna tendencia de ubicación dentro de los límites geográficos del municipio.

De las dos herramientas que se han utilizado, solo una analiza la noción de cercanía. La otra, trabaja a partir de valores similares de un grupo de atributos. Por lo tanto, se puede esperar que al analizar un grupo más homogéneo de observaciones o empresas, el comportamiento espacial sea un poco más claro al interior del partido.

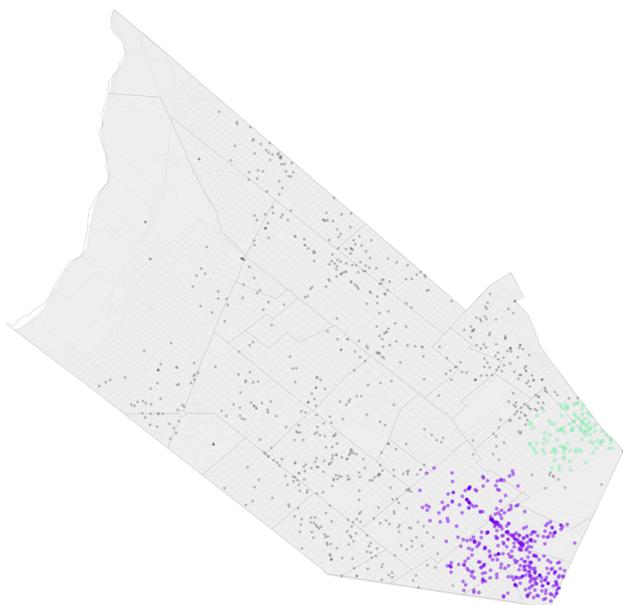
Para tal fin, y al utilizar un “epsilon” de 500 metros y un “min_samples” de 70 vecinos, dbscan configura un total de 2 clusters con un Silhouette coefficient de 0.286. Esto es, levemente superior al último resultado obtenido para centros de actividad económica. Si bien el resultado pareciera no ser demasiado significativo, es cierto que esto cobra un mayor sentido si se tiene en cuenta que el patrón de

aglomeración es mucho más concentrado que en el análisis previo. Es decir, las industrias del partido de General San Martín se agrupan mayormente en la localidad homónima, así como en Villa Maipú y Villa Lynch.

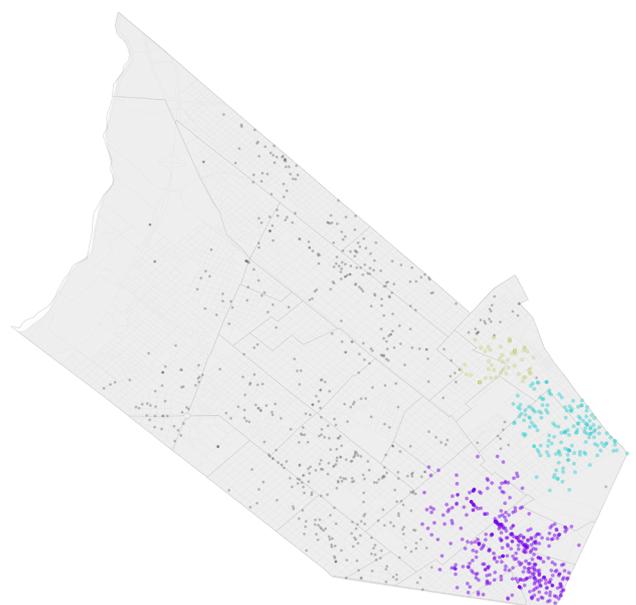
Si se ajustan levemente los parámetros del algoritmo, el resultado mejora un poco más. Al reducir la cantidad mínima de vecinos de 70 a 50, y manteniendo el mismo radio de distancia, el Silhouette coefficient aumenta a 0.318. Resultado que podría entenderse como contundente, si se aprecia que no hace más que reafirmar el patrón de aglomeración espacial anterior. Una vez más, una alta concentración de industrias en el límite del partido con la Ciudad de Buenos Aires, y al norte del polígono conformado por las Avenidas General Paz, 25 de Mayo, Gral. Manuel Savio y de los Constituyentes.

Más allá de estos resultados, es decir, de cuánto mejora el Silhouette Coefficient, es importante

M7 | CENTROS DE ACTIVIDAD INDUSTRIAL EN 2 CLUSTER



M8 | CENTROS DE ACTIVIDAD INDUSTRIAL EN 3 CLUSTER



remarcar que en ambos casos las industrias se concentran en un único sector del partido.

Lo que se traduce en un patrón de aglomeración distinto al reconocido en la sección inicial, donde se identifican núcleos de actividad en distintas zonas o localidades del partido. Ahora en cambio, se observa que la actividad industrial se aloja en un área más restringida.

Para profundizar el análisis, es conveniente evaluar la capacidad explicativa que tiene la interpretación de estas aglomeraciones en dos o tres categorías. Esto es, cuanto más se diferencian los grupos entre sí, conforme se agrupan en dos o en tres clusters.

Para ello, se implementa nuevamente K-Means sobre el dataset inicial (como en el paso anterior), pero con el filtro de aquellas CUIT en cuya rama de actividad se reconocen tanto un código como una descripción asociada al sector industrial (fabricación de metales, maquinaria, productos químicos, alimenticios, manufacturas en general, entre otros)¹¹. Con esto se reemplaza la rama de actividad por la subrama.

De esta manera se reduce el espectro de posibilidades, con la finalidad de trabajar sobre un grupo de datos que debería ser más homogéneo¹².

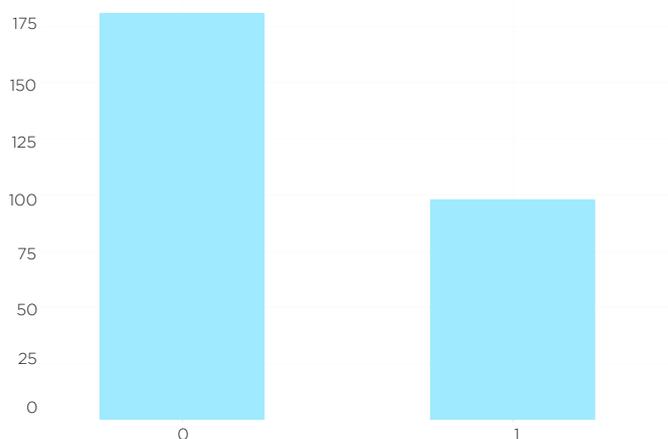
En este escenario del estudio, se elige analizar únicamente el nivel de agregación geográfica de menor escala. Esto, considerando que, si las empresas se aglomeran en una única área del partido, entonces debería haber una mayor cantidad de casos en unidades administrativas inferiores. Además, si se consideran las fracciones como unidad geográfica, el algoritmo hubiese contado

con pocos casos para asignar entre clusters (ya no todas las fracciones del partido sino un grupo más reducido).

Volviendo a la consigna anterior, la cantidad de clusters que maximice las diferencias entre grupos sigue un criterio específico. Este es, tratar de que cada cluster cuente con una cantidad similar de casos. Esto, para evitar que en el resultado se contabilice como un agrupamiento específico, aquellos radios que al no pertenecer a ningún grupo mayoritario, se geolocalizan de forma aleatoria en la superficie del territorio. Es decir, en aquellas situaciones en las que se reconoce como grupo un conjunto de radios separados por una distancia demasiado amplia.

Por eso, luego de evaluar el agrupamiento en dos o en tres categorías, se opta por descartar al último de ellos. Esto, basándose en la distribución de frecuencias de ambos escenarios.

67 | CANTIDAD DE RADIOS CENSALES POR CLUSTER EN 2 CATEGORÍAS



¹¹ En la sección metodológica del presente estudio, se especifican los códigos que se utilizan para la identificación de las ramas de actividad. Estos son extraídos del sitio web escrapearo. En lo que respecta a la rama de actividad, las mismas son parseadas con expresiones regulares en python para mejorar su lectura y poder separar los dos niveles que se reconocen en la descripción de actividades.

¹² El código y la descripción que se extraen de la web de CUIT online reconoce dos niveles. En primer lugar, una agrupación general, de donde se diferencian actividades del sector industrial. A partir del segundo nivel de descripción se construyen las categorías de subrama.

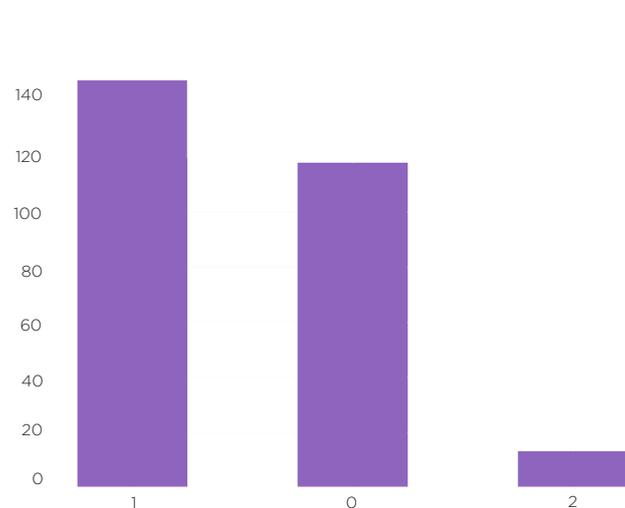
Como se puede observar para el agrupamiento en tres categorías, uno de los grupos cuenta con una cantidad mucho menor de radios que los demás. Así, se reconocen dos opciones. Una, considerando a este grupo minoritario como un cluster en sí mismo, o bien dividiendo el total de industrias en dos grupos.

Si bien es cierto que dichos radios van a contar siempre con la misma ubicación geográfica, también lo es que al repartirlos en dos grupos distintos los valores de los indicadores que se consideran para conformarlos (promedio de empleados, actividades declaradas y distribución por tamaño y subrama) deberían ser más robustos por contemplar una mayor cantidad de casos. Estos radios van a seguir estando aislados en el espacio pero, ahora pueden ser asimilados como el margen o periferia de los clusters principales.

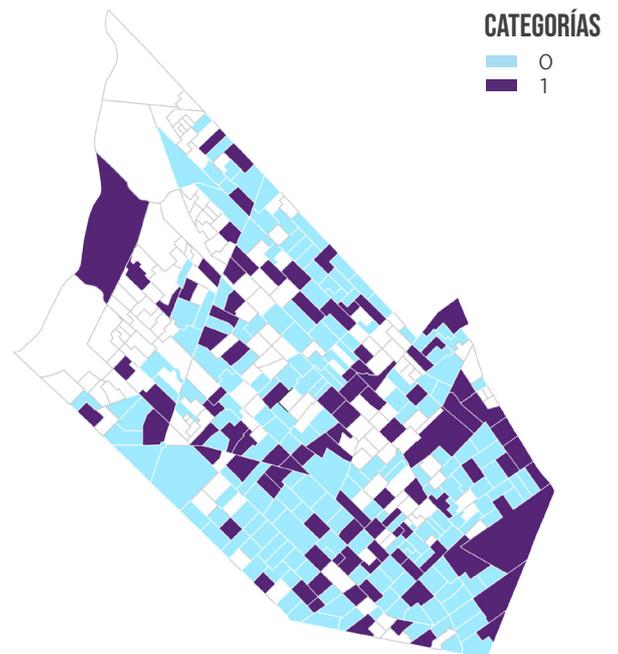
Un primer elemento que se puede reconocer, es que a diferencia del análisis previamente realizado para la misma división administrativa, en esta oportunidad la cantidad de radios censales con valores asignados (que pertenecen a un cluster) es mucho menor¹³.

De la misma forma, no podríamos decir que ambos grupos tienen una diferenciación clara en el territorio. Si bien es cierto que algunas zonas del partido se distinguen claramente por el cluster al que pertenecen, también lo es que en otras el patrón no es continuo. Es decir, que en el medio del área de cobertura de un cluster aparecen colores pertenecientes a otro. Ahora, también se puede observar que en algunas regiones del partido varios radios censales se agrupan formando un solo cluster.

68 | CANTIDAD DE RADIOS CENSALES POR CLUSTER EN 3 CATEGORÍAS



M9 | CLUSTERIZACIÓN ESPACIAL EN 2 CATEGORÍAS POR RADIO CENSAL



Fuente: Elaboración propia.

¹³ Esto se puede ver con claridad en la cantidad de radios en blanco. Los que no tienen valor o categoría, no son imputados en el algoritmo por no contar con casos a ser evaluados.

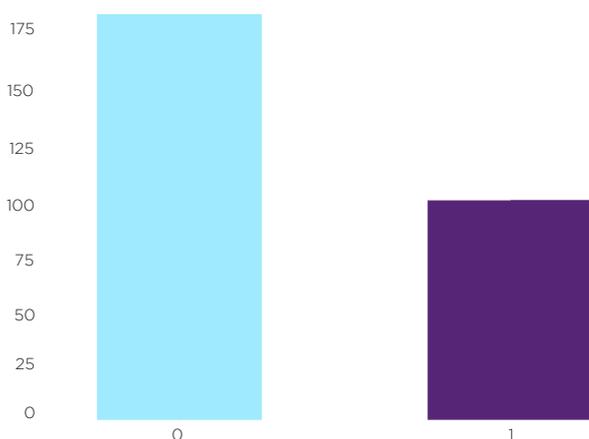
Por decirlo de otra manera, que su tamaño es lo territorialmente representativo como para que se identifiquen regiones de atributos similares (o de un mismo color). Entre ellas, la que se acotaron en análisis previos, como la zona adyacente a la Avenida General Paz. Dicho polígono sigue asumiendo un color único. Es decir, que parece que se está frente a un grupo de similares características.

Entre estas, se destaca cómo el cluster “0” cuenta con una mayor proporción de microempresas. El promedio de empleados para este agrupamiento es de 8 mientras que el número “1” cuenta con un nivel de 25 empleados promedio. Aquí tampoco se reconoce una distribución por subrama muy diferente. Para ambos clusters, las industrias vinculadas al metal, el caucho y el plástico, los productos alimenticios y la reparación de maquinaria, tienden a ser las más importantes en ambos agrupamientos.

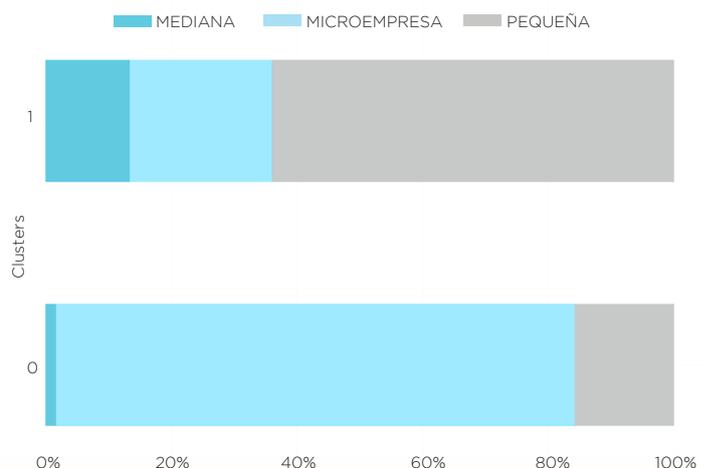
Así las cosas, parece quedar claro que los dos grupos, antes que en la subrama de actividad, parecen diferenciarse en el tamaño de las unidades económicas o productivas. Es decir, en su capacidad medida en cantidad de empleados. Por lo tanto, si el objetivo es identificar el principal patrón de aglomeración espacial de industrias dentro del partido, es viable establecer como respuesta, que este no responde a una conjunción de atributos muy clara.

Lo que significa que las industrias no se ubican dentro del territorio, por ejemplo, para formar algún tipo de economía de escala externa o por aglomeración, que les induzca a estar cerca de ramas de actividad afines o conexas. Lo que más las distingue dentro del territorio, en lo que respecta

G9 | CANTIDAD DE RADIOS CENSALES POR CATEGORÍA

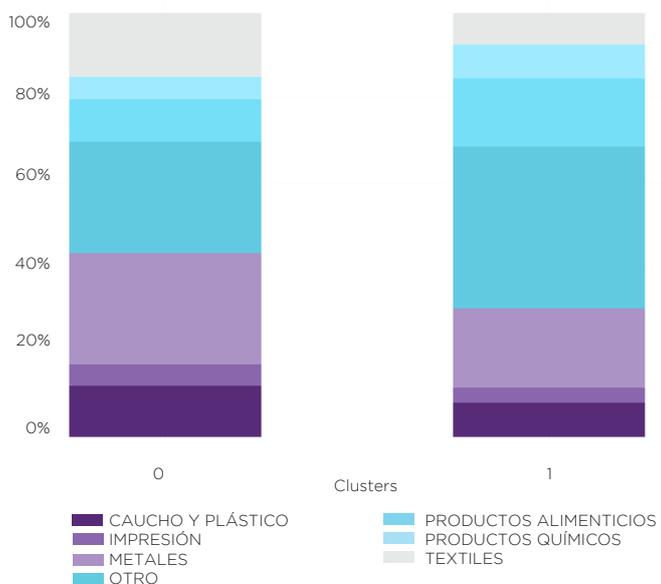


G10 | PORCENTAJE DE INDUSTRIAS POR TAMAÑO SEGÚN CLUSTER (RADIO CENSAL)



a su agrupamiento o aglomeración, es la cantidad de empleados y, principalmente, su proximidad a la Ciudad Autónoma de Buenos Aires. El grueso del universo de las unidades económicas asociadas a prácticas o actividades industriales se encuentran cercanas a la Avenida General Paz.

611 | PORCENTAJE DE INDUSTRIAS POR SUBRAMA SEGÚN CLUSTER (RADIO CENSAL)



Fuente: Elaboración propia.

03_

REFLEXIONES FINALES

Como se pudo evidenciar a lo largo del estudio, en el tejido industrial del partido de San Martín no se reconocen patrones claros de aglomeración espacial. Esto, en lo que respecta a la conformación de cadenas de valor agregado, cuyos eslabones se ubiquen cerca de ramas de actividad que pudieran resultar complementarias en sus procesos productivos. En otras palabras, no se distinguen economías de localización, donde las unidades de actividad industrial, comercial y de servicios, se ubiquen en el territorio siguiendo algún criterio de complementariedad o beneficio mutuo.

Lo que sí se nota en el comportamiento espacial de las firmas, es la tendencia a conformar núcleos o zonas de mayor concentración de actividades económicas. Sin embargo, en estas y como se acotó inicialmente, no se observa una estructura u orden que obedezca a la interacción por localización. No se distinguen valores significativos con respecto a la cantidad de empresas según tamaño y rama de actividad similar. Ahora bien, este es en definitiva, el patrón general que (a grandes rasgos), suele repetirse en los núcleos económicos reconocidos.

Los mayores niveles de actividad general (independiente al tipo de rama), suelen concentrarse en las zonas de Villa Lynch, General San Martín, Villa Ballester y José León Suárez. Áreas que aunque en gran medida presentan características urbanísticas y socioeconómicas diferentes entre sí, tienen como aspectos en común: facilidades de acceso (Ruta 8, Av. 25 de Mayo, la Línea Mitre Ramal Retiro José León Suárez, Ruta 4); y valores relativamente altos en cuanto a densidad habitacional y flujos de población flotante.

Al momento de analizar particularmente al sector industrial, los resultados arrojan que los núcleos perceptibles en el municipio se localizan en las adyacencias a la Avenida General Paz, en las localidades de Villa Lynch y Villa Maipú. Con énfasis en aquellos puntos que permiten mejores posibilidades de acceso, desde o hacia esta, como alrededor de las intersecciones de la Ruta 8 y la Av. Constituyentes con la Avenida General Paz.

La importancia que tiene la accesibilidad en la localización de las firmas industriales, está directamente vinculada con la incidencia del transporte en la viabilidad económica del sector. Tanto en la adquisición y costo de los insumos, como en la rentabilidad y comercialización de los productos. La Avenida General Paz conecta al partido de General San Martín con puertos aéreos y fluviales, así como con múltiples destinos a nivel metropolitano y regional

Es así que si bien no se identifican patrones de aglomeración espacial por afinidad en sub ramas de actividad, si se reconoce un patrón de localización industrial en función de la accesibilidad y el transporte, y las economías de escala interna de las empresas.

ILUSTRACIONES, GRÁFICOS, MAPAS Y TABLAS

ILUSTRACIONES

- I1: Esquema del Método DBSCAN
- I2: Ejemplificación del Algoritmo de Clustering DBSCAN

GRÁFICOS

- G1: Cantidad de fracciones censales por cluster
- G2: Porcentaje de empresas por tamaño según cluster (fracción censal)
- G3: Porcentaje de empresas por rama según cluster (fracción censal)
- G4: Cantidad de radios censales por categoría
- G5: Porcentaje de empresas por tamaño según cluster (radio censal)
- G6: Porcentaje de empresas por rama según cluster (radio censal)
- G7: Cantidad de radios censales por cluster en 2 categorías
- G8: Cantidad de radios censales por cluster en 3 categorías
- G9: Cantidad de radios censales por categoría
- G10: Porcentaje de industrias por tamaño según cluster (radio censal)
- G11: Porcentaje de industrias por subrama según cluster (radio censal)

TABLAS

- T1: Promedio de empleados y actividades por cluster (fracciones)
- T2: Promedio de empleados y actividades por cluster (radios)

MAPAS

- M1: Total de unidades económicas geolocalizadas
- M2: Base final unidades económicas a analizar
- M3: Centros de actividad económica en 6 clusters
- M4: Centros de actividad económica en 8 clusters
- M5: Clusterización espacial en 4 categorías a nivel de fracción censal
- M6: Clusterización espacial en 4 categorías por radio censal
- M7: Centros de actividad industrial en 2 cluster
- M8: Centros de actividad industrial en 3 cluster
- M9: Clusterización espacial en 2 categorías por radio censal

BIBLIOGRAFÍA

Arribas, D. Geographic Data Science with PySAL and the pydata stack. Spatial Clustering. Disponible en: http://darribas.org/gds_scipy16/ipynb_md/07_spatial_clustering.html

Boeing G., Urban Data Science. Cluster Analysis. Disponible en: <https://github.com/gboeing/urban-data-science/blob/master/15-Spatial-Cluster-Analysis/cluster-analysis.ipynb>

Pierobon, G. Medium. DBSCAN clustering for data shapes k-means can't handle well (in Python). Disponible en: <https://towardsdatascience.com/dbscan-clustering-for-data-shapes-k-means-cant-handle-well-in-python-6be89af4e6ea>

Vázquez Brust, A. Bit & Bricks. DBSCAN: Machine Learning para detectar centros de actividad urbana. Disponible en: <https://bitsandbricks.github.io/post/dbscan-machine-learning-para-detectar-centros-de-actividad-urbana/>

Centro de Estudios Económicos Urbanos (CEEU)

Escuela de Economía y Negocios (EEyN)
Universidad Nacional de San Martín (UNSAM)
Septiembre 2019
Caseros 2241, San Martín, Prov. de Buenos Aires
www.unsam.edu.ar/escuelas/economia/ceeu/
e-mail: ceeu.eeyn@unsam.edu.ar
tel. 4580 7250 int. 147

