



**UNIVERSIDAD
DEL PACÍFICO**

**Escuela de
Postgrado**

**LAS ECONOMÍAS DE AGLOMERACIÓN
EN LAS ÁREAS URBANAS DE PERÚ**

**Trabajo de Investigación presentado
para optar al Grado Académico de
Magíster en Economía**

Presentado por:

Anthony Pinedo Araujo

Paola Andrea Valer Cortez

Asesor: Alberto Elías Chong Lam

[0000-0002-5139-5386](tel:0000-0002-5139-5386)

Lima, febrero de 2024

REPORTE DE EVALUACIÓN DEL SISTEMA ANTIPLAGIO

A través del presente, Alberto Elías Chong Lam deja constancia que el trabajo de investigación titulado "LAS ECONOMÍAS DE AGLOMERACIÓN EN LAS ÁREAS URBANAS DE PERÚ" presentado por don Anthony Pinedo Araujo de acuerdo con el D.N.I. 44434487 y doña Paola Andrea Valer Cortez de acuerdo con el D.N.I. 72310853 para optar al Grado de Magister en Economía fue sometido al análisis del sistema antiplagio Turnitin el 03/03/2024 dando el siguiente resultado:

The screenshot shows a Turnitin report interface. The main content area displays the title of the thesis: "LAS ECONOMÍAS DE AGLOMERACIÓN EN LAS ÁREAS URBANAS DE PERÚ" and the author's name: "Trabajo de Investigación presentado para optar al Grado Académico de Magister en Economía". The report is from the Universidad del Pacífico, Escuela de Postgrado. On the right side, there is a sidebar titled "Resumen de coincidencias" (Summary of similarities) showing a total similarity score of 13%. Below this, a table lists 12 sources with their respective similarity percentages.

Rank	Source	Similarity %
1	WU World Bank	2 %
2	Investopedia	1 %
3	aprendiendo.com	1 %
4	WU World Bank	1 %
5	WU World Bank	<1 %
6	WU World Bank	<1 %
7	WU World Bank	<1 %
8	WU World Bank	<1 %
9	WU World Bank	<1 %
10	WU World Bank	<1 %
11	WU World Bank	<1 %
12	WU World Bank	<1 %

Resumen ejecutivo

Las ciudades son motores del crecimiento económico, concentrando la mayor parte de la actividad económica, la innovación y la productividad. Las aglomeraciones urbanas, la concentración de empresas e industrias en un mismo espacio, son clave para la mayor productividad urbana.

Este trabajo de investigación busca determinar el efecto de las aglomeraciones urbanas sobre la productividad de las empresas, analizando la relación entre la densidad poblacional en un área urbana y la productividad de las empresas allí ubicadas.

La investigación se basa en un análisis empírico de datos de: población por unidad de área, geolocalización y niveles de productividad de empresas sectorizadas, utilizando técnicas econométricas para estimar el efecto de la aglomeración urbana en tres niveles distintos y la productividad de las empresas.

Para ello se utilizó data de densidad poblacional por cada 1'x1' geográfico de área proporcionada por Meta. A partir de allí se delimitaron las áreas urbanas fijando un umbral de densidad mínima que debía cumplir. A partir de dichas áreas urbanas se definieron los centros urbanos y centros urbanos *expandidos* como los puntos y áreas de mayores concentraciones de densidad poblacional dentro del área urbana utilizando el modelo policéntrico. Por el lado de las empresas, se requirió los datos de SUNAT e INEI para geolocalizar a las compañías a través de la dirección registrada en el estado y medir sus niveles de productividad a través de la Encuesta Económica Anual.

Los resultados muestran una contribución negativa de las economías de aglomeración en la productividad para todas las empresas de -0.68, llegando a valores de -0.78 y -0.81 si las empresas pertenecen a un área urbana y un centro urbano expandido respectivamente.

Las razones detrás de este hallazgo pueden atribuirse al ciego desarrollo industrial o empresarial, al alto nivel de informalidad en el país o a las fricciones existentes en el mercado laboral.

Los resultados tienen importantes implicaciones para la política económica, permitiendo a los gobiernos locales diseñar políticas que fomenten la aglomeración urbana. Este trabajo contribuye a la literatura sobre economía urbana y al debate sobre el papel de las ciudades en el crecimiento económico.

Tabla de contenido

1. INTRODUCCIÓN	6
2. REVISIÓN DE LITERATURA	8
i. Economías de aglomeración	8
i. Economías de localización	10
ii. Economías de urbanización	11
ii. Delimitación de áreas urbanas	14
iii. Densidad poblacional y centros de negocios	16
3. MARCO METODOLÓGICO	19
i. Modelo	19
ii. Variables y bases de datos	22
i. Definición las áreas urbanas utilizando la data proveniente de los Mapas de Densidad de Alta Resolución (MDAR) de Meta	25
ii. Definición de los centros urbanos y centros urbanos <i>expandidos</i>	29
iii. Geolocalización de las empresas en base en el padrón de SUNAT	32
iv. <i>Scrapping</i> de empresas participantes de la EEA de 2016	33
v. Definición de índices de localización y de urbanización	34
4. ANÁLISIS DE LOS RESULTADOS	36
i. Estadísticas descriptivas	36
ii. Primera etapa	36
iii. Áreas urbanas	37
iv. Centros urbanos y centros urbanos <i>expandidos</i>	40
v. Segunda etapa	43
i. Geolocalización de empresas con base en el padrón de SUNAT	43
ii. <i>Scrapping</i> e identificación de las empresas participantes de la EEA de 2016	43
iii. Definición de índices de localización y urbanización	44
iv. Resultados	44
5. CONCLUSIONES	50
6. BIBLIOGRAFÍA	51
7. ANEXOS	55
Anexo N° 1	55
Anexo N° 2	59
Anexo N° 3	60
Anexo N° 4	64
Anexo N° 5	65

Índice de tablas

Tabla N° 1: Definición de las variables	23
Tabla N° 2: Número de observaciones por sector económico	24
Tabla N° 3: Estadísticas descriptivas de las variables de interés	36
Tabla N° 4: Áreas urbanas halladas	38
Tabla N° 5: Número de centros urbanos hallados por área urbana	41
Tabla N° 6: Número de pares región geográfica-sector por región geográfica	44
Tabla N° 7: Los efectos de las economías de localización y urbanización sobre la productividad de las empresas (a nivel de provincia).....	47
Tabla N° 8: Los efectos de las economías de localización y urbanización sobre la productividad de las empresas (a nivel de área urbana)	48
Tabla N° 9: Los efectos de las economías de localización y urbanización sobre la productividad de las empresas (a nivel de centro urbano <i>expandido</i>)	49
Tabla N° 10: Formatos por sector económico	59
Tabla N° 11: Estadísticas descriptivas de las variables de interés por sector.....	60
Tabla N° 12: Resultados de la estimación por sector.....	64
Tabla N° 13: Delimitación de áreas urbanas, centros urbanos y centros urbanos <i>expandidos</i>	65

Índice de figuras

Figura N° 1: Círculos concéntricos de Von Thünen	9
Figura N° 2: Cuadrículas de 1" × 1" proporcionadas por Meta para una sección de Lima.....	25
Figura N° 3: Comparación de los bordes obtenidos entre una cuadrícula de 1 km y de 1" geográfico	27
Figura N° 4: Posibles islas al pasar el algoritmo (caso de la ciudad de Trujillo)	28
Figura N° 5: Continuidad de cuadrículas.....	28
Figura N° 6: Poblaciones absolutas sobre cada cuadrícula para la ciudad de Trujillo	30
Figura N° 7: Poblaciones <i>extendidas</i> sobre cada cuadrícula para la ciudad de Trujillo	31
Figura N° 8: Posible superposición de <i>centros urbanos expandidos</i>	32
Figura N° 9: Visualización gráfica de las cinco ciudades más pobladas y sus áreas urbanas	39
Figura N° 10: Visualización gráfica de las cinco ciudades más pobladas y sus áreas urbanas, centros urbanos <i>expandidos</i> y centros urbanos.....	42

Índice de anexos

Anexo N° 1: Definiciones de índices de localización y urbanización	54
Anexo N° 2: Formatos por sector económico	58
Anexo N° 3: Estadísticas descriptivas de las variables de interés por sector.....	59
Anexo N° 4: Resultados de la estimación por sector	63
Anexo N° 5: Delimitación de áreas urbanas, centros urbanos y centros urbanos <i>expandidos</i>	64

1. INTRODUCCIÓN

Economías de aglomeración es un tema central en el ámbito de la economía urbana. En esta, se puede entender que la acumulación de empresas en una misma área geográfica impulsa la productividad de estas mismas mediante diversos mecanismos. Por otro lado, la delimitación de áreas urbanas en un país es crucial debido a la necesidad de planificar, gestionar y evaluar las ciudades de manera más eficiente.

Según la literatura existente, las economías de aglomeración se pueden clasificar en economías de localización y de urbanización. La primera se relaciona con los beneficios producto de la concentración de empresas de un mismo sector; mientras que la segunda está vinculada con las ventajas provenientes de la diversidad de sectores económicos en la ciudad donde la empresa se encuentre ubicada. En este sentido, el presente trabajo de investigación propone estimar las economías de aglomeración en cuanto a sus dos divisiones para 14 actividades económicas.

Para lograr este objetivo se considerará una estimación de dos etapas. En primer lugar, es esencial destacar que, a lo largo del documento, se asume que todos los sectores son competitivos y que las empresas dentro de un mismo sector económico poseen una función de producción homogénea y tecnologías de producción idénticas para todas las ciudades. En este marco, en la primera etapa, haciendo uso de una función de producción Cobb-Douglas y el residuo de Solow se estimará la productividad total de factores (PTF) por empresa perteneciente a cierto sector. Posteriormente, en la segunda etapa, se estimarán los impactos de interés de las economías de localización y urbanización.

Esta segunda etapa se realizará tomando en consideración que estas actividades se desarrollan dentro de una aglomeración; es decir, dentro de áreas urbanas que trascienden de una delimitación administrativa o dentro de lo que se definirá como centro urbano *expandido*. Por lo tanto, en primer lugar, se delimitarán estas áreas en función de las densidades poblacionales y de los centros urbanos, respectivamente. En consecuencia, el presente trabajo de investigación también propone la delimitación de tales áreas urbanas y centros urbanos *expandidos* para Perú debido a que no existen trabajos de esta naturaleza para el país.

En otras palabras, con el hallazgo de 65 áreas urbanas y 144 centros urbanos *expandidos* en todo el país es posible asociar una empresa con un área urbana y, posiblemente, con centro urbano *expandido*. Esta asociación, mediante la segunda etapa, se utilizará para determinar si estar en un área urbana o en un centro *expandido* afecta a la productividad positiva o negativamente. La hipótesis de aglomeración planteada para el presente trabajo sostiene que tanto las economías de localización como las de urbanización impactan positivamente en la productividad a nivel nacional.

En adición, la investigación busca contribuir al ámbito académico al proponer una modificación en la definición de ciudades basada en umbrales de densidades poblacionales que se ajustan mejor a grillas

más pequeñas. Por último, también propone un nuevo enfoque para verificar si las ciudades y sus centros conllevan mejoras en la productividad de las empresas.

Además de esta introducción, este documento de investigación comprende una revisión de literatura sobre economías de aglomeración, localización y urbanización, así como sobre las áreas y centros urbanos, en la segunda sección. En la tercera sección se presenta el marco metodológico. Los resultados preliminares son analizados en la cuarta sección, mientras que en la quinta sección se exponen las conclusiones, seguidas de la bibliografía y, por último, los anexos.

2. REVISIÓN DE LITERATURA

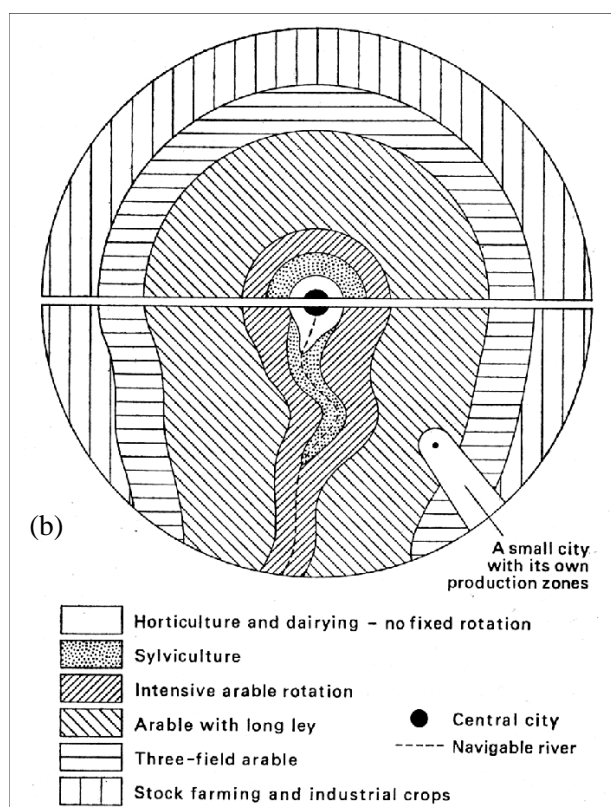
i. Economías de aglomeración

Las economías de aglomeración han sido objeto de estudio, desde hace varias décadas, por economistas, urbanistas y geógrafos. En particular, el concepto de “economías de aglomeración” fue introducido por Alfred Weber en 1909. Según él, es un elemento que, junto con los costos de transporte y los costos laborales, determina la localización óptima de las actividades empresariales.

Sin embargo, así como la aglomeración genera beneficios, también conlleva consecuencias, costos o *deseconomías de aglomeración* como el incremento del precio del suelo, el alargamiento de los desplazamientos (en tiempo y distancia) o los mayores niveles de contaminación. Sin embargo, es importante destacar que el continuo crecimiento que han experimentado los núcleos urbanos (Noticias ONU, 2018) sugiere que, históricamente, los beneficios de la aglomeración se han impuesto por sobre sus posibles costos, aunque debe existir un límite.

Adicionalmente es considerable destacar algunos precedentes de Weber como von Thünen (1826) y su “Teoría de la localización”. Los supuestos detrás del desarrollo de su teoría incluyen que 1) el espacio físico es homogéneo y uniforme (sin ríos ni montañas), 2) hay un único mercado localizado en el centro, 3) la dotación de factores es ubicua, 4) los costos de producción son independientes de la localización, 5) los agricultores quieren maximizar sus beneficios y 6) la distancia al mercado, factor clave, está asociada con los costos de transporte y depende de la localización geográfica de la actividad económica. De acuerdo con este modelo, la solución óptima para los productores agrarios era situarse lo más cerca posible del mercado. Además, von Thünen construyó un modelo del uso del suelo con varios círculos concéntricos alrededor del centro de la ciudad, donde los productos perecederos, voluminosos o pesados se producirían en las fajas más cercanas al centro; mientras que los productos de menor peso y volumen se especializarían en los cinturones más distantes, pero tendrían un precio más alto en el mercado debido a los costos de transporte relativamente más altos. En ese sentido, como puede verse en la parte (a) de la Figura N° 1, cada círculo se especializa en la producción de aquellos productos básicos agrícolas para los que se adapta mejor.

Figura N° 1: Círculos concéntricos de Von Thünen



Fuente: Grotewold (1959)

Finalmente, el autor incorporó dos factores modificadores a su modelo: un río navegable que aceleraba el transporte y una ciudad más pequeña que actuaba como mercado competitivo. En la Figura N° 1 (b) se puede notar que, a pesar de la inclusión de estas dos modificaciones, el patrón del uso del suelo es más complejo pero similar. Años más tarde, Alonso (1964) generalizó el trabajo de von Thünen al contexto urbano y reemplazó el “mercado” por el centro de empleo en el Distrito Central de Negocios (o *Central Business District*, CBD).

Launhardt (1882) es otro precedente de Weber con la primera aproximación referida a la localización de la actividad industrial. En su investigación analizó los patrones de ubicación de ciertas industrias manufactureras en términos de los costos de transporte, de las áreas de mercado y de las fuentes de materia prima, con el objetivo de encontrar el punto óptimo para minimizar los costos de transporte de la empresa.

Por otro lado, Marshall (1980) introdujo dos tipos de economías de aglomeración: economías internas y economías externas. Las economías internas son exclusivas de la empresa o se generan dentro de su gestión interna, como cuando una empresa reduce los costos y aumenta la producción. Las economías externas, por otro lado, se producen a partir de cambios más amplios fuera de la empresa como la industria o el sector. Estas últimas se dividen en una clasificación tradicional (Glaeser, Kallal, Scheinkman, & Shleifer, 1996): economías de localización y economías de urbanización. La primera está relacionada con los beneficios derivados de la concentración de varias empresas del mismo tipo de

industria o sector, ya sea por condiciones físicas o invitación de los gobernantes (Marshall, 1980). En particular, el autor observa que las industrias se especializan geográficamente, porque la proximidad favorece la transmisión de conocimiento intra-industrial (*knowledge spillovers*), reduce los costos de transporte de insumos y productos debido a los recursos compartidos (*input sharing*), y permite que las empresas se beneficien de un mercado laboral más eficiente (*labor market pooling*). Las economías de urbanización, por otro lado, están vinculadas a las ventajas provenientes del tamaño o la diversidad de la ciudad donde se encuentra ubicada la empresa (Jacobs, 1969), lo que incluye la proximidad al mercado, la disponibilidad de mano de obra, buenas comunicaciones y servicios financieros y comerciales.

Así, economías de localización se derivan de la ubicación o cercanía con respecto a otras empresas en la misma industria; mientras que, las economías de urbanización reflejan economías externas que funcionan entre empresas de diferentes industrias en una misma ciudad; es decir, las que están asociadas con la cercanía a la actividad económica general (Desmet & Fafchamps, 2005).

i. Economías de localización

El modelo Marshall-Arrow-Romer (MAR), formalizado por Glaeser *et al.* (1992), postula que la concentración de una industria en una región promueve la difusión del conocimiento entre las empresas (*knowledge spillovers*) y facilita la innovación en esa industria específica dentro de dicha región. Esta especialización fomenta la transmisión e intercambio de conocimientos entre empresas de la misma industria o similar, de ideas e información, de manera tácita o codificada, de productos y procesos a través de la imitación, las interacciones comerciales y la circulación interempresarial de trabajadores calificados, sin transacciones monetarias (Saxenian, 1994). De esta manera, una cierta configuración industrial local que incluya numerosas empresas pequeñas en industrias relacionadas o varios competidores yuxtapuestos en la misma industria tiende a mejorar la competitividad de la empresa al impulsar la innovación. En consonancia, van Oort y Stam (2006) destacan que la alta competencia actúa como un fuerte incentivo para el gasto en I+D, ya que las empresas se ven obligadas a innovar para sobrevivir. Esto no solo aumenta la probabilidad de adoptar una nueva tecnología o ser pionero, sino que también agiliza la difusión de información sobre la innovación y amplía el número potencial de adoptantes (Kelley & Helper, 1999). Es probable que estas externalidades de localización se intensifiquen a medida que la industria crece (Frenken, van Oort, Verburg, & Boschma, 2004)

Marshall identificó otros dos beneficios de la concentración geográfica: la agrupación del mercado laboral y el ahorro en los costos de transporte. En primer lugar, la concentración local de empresas en una misma industria aumenta las oportunidades de empleo para los trabajadores (*labor market pooling*). Con especial énfasis a la forma cómo se emparejan empleadores y trabajadores. En este caso particular, las economías de aglomeración pueden surgir porque la calidad de cada emparejamiento (o *match*) mejora cuando aumenta el número de agentes intentando emparejarse; también conocido como

“externalidad del emparejamiento” (Helsley & Strange, 1990). Además, la migración de estos trabajadores entre empresas contribuye a la difusión del conocimiento. En segundo lugar, las economías de escala derivadas del uso compartido de insumos (*input sharing*), como la mano de obra y la infraestructura, entre grandes concentraciones de empresas en la misma industria son otra fuente importante de economías de localización (Krugman, 1991). Esto está relacionado con los rendimientos crecientes que resultan de la provisión de ciertos bienes o servicios indivisibles, lo que permite una unidad adicional de este bien a un costo promedio considerablemente bajo. Asimismo, las empresas suelen ubicarse cerca de sus proveedores para reducir los costos de transporte y cerca de sus clientes para reducir los costos de distribución (Le Blanc, 2000).

ii. Economías de urbanización

La teoría de Jacobs (1969) enfatiza que la diversidad de industrias dentro de una región geográfica promueve las externalidades del conocimiento y, en última instancia, la actividad innovadora y el crecimiento económico. El intercambio de conocimientos complementarios entre diversas empresas y agentes económicos facilita la búsqueda y experimentación de la innovación. Una economía más diversa fomenta el intercambio de habilidades necesarias para el surgimiento de nuevos campos (Harrison, Kelley, & Gant, 1996).

Sin embargo, Combes (2000) especifica que esto presupone que los sectores son cercanos tecnológicamente, es decir, que las innovaciones de un sector pueden incorporarse a la producción de otra industria. Además, una infraestructura de transporte y comunicación que funcione bien, la proximidad a los mercados y un mejor acceso a servicios especializados son fuentes adicionales de externalidades de urbanización que facilitan el funcionamiento de las empresas y, por lo tanto, el crecimiento económico.

Aspectos empíricos

Ferragina y Mazzotta (2014) realizaron un análisis de los determinantes en la supervivencia empresarial basado en las economías de aglomeración y su impacto en la industria manufacturera, tanto nacional como multinacional. Utilizaron datos de panel compuestos por 884,232 observaciones de empresas, desglosados por sectores, provincias y empresas de interés en 110 provincias italianas entre los años 2002 y 2010. Definieron los índices de localización como el índice de especialización de Balassa¹, y el índice de urbanización como la densidad poblacional. Emplearon los modelos semiparamétricos de riesgo de Cox para explorar los determinantes de la duración de las empresas, encontrando que los efectos de los límites geográficos tienen poca importancia y que la diversificación es relevante para las empresas locales, pero no para las multinacionales.

¹ Dichos índices están definidos en el Anexo N° 1.

Qi Guo, Canfei He y Deyu Li (2016) exploraron los determinantes del emprendimiento manufacturero a nivel de prefecturas en China, destacando la influencia de las economías de localización y urbanización y la importancia de las relaciones tecnológicas y las pequeñas agrupaciones de empresas. Utilizaron datos de panel de la Encuesta Anual de Firms Industriales de los años 2001 a 2007, enfocándose en el sector manufacturero. Utilizaron el índice de especialización de Balassa¹ como índice de localización, mientras que la urbanización fue desglosada en varios factores. Los resultados mostraron que las economías de localización desempeñan un papel muy importante y positivo en el estímulo al emprendimiento, mientras que la urbanización no mostró un efecto significativo.

Feng Han, Rui Xie y Jiayu Fang (2018) estudiaron los efectos de la aglomeración de empresas en la eficiencia energética industrial utilizando datos de panel de 283 ciudades a nivel de prefectura en China entre los años 2003 y 2010, utilizando el modelo dinámico de Durbin. Definieron índices de localización o especialización de Balassa y de urbanización o diversificación¹ en función del número de trabajadores por sector-ciudad, encontrando que la localización y la diversificación no tuvieron un efecto significativo en la ciudad, pero redujeron significativamente las eficiencias energéticas de las ciudades vecinas.

Toshitaka Gokan, Ikuo Kuroiwa y Kentaro Nakajima (2019) examinaron los efectos de las economías de aglomeración en la productividad a nivel de empresa en Vietnam. Utilizaron la metodología propuesta por Mori y Smith (2014) para identificar la ubicación de los clusters para una industria y estimaron los efectos de las economías de localización y urbanización por sector-región, privado y extranjeros. Hicieron uso de datos del Censo de Establecimientos del año 2012 conducido por la Oficina General de Estadística. Los efectos de las economías de localización y urbanización fueron descompuestos en tres fuentes: transacciones entre industria, difusión de conocimiento entre empresas, y oportunidades de empleo. Encontraron que, las economías de aglomeración mejoran la productividad de la firma y que las economías de urbanización mejoran la productividad solo para empresas extranjeras. Sin embargo, estos efectos desaparecen para el caso de las empresas estatales.

Yin Huang, Tao Hong y Tao Ma (2020) analizaron el efecto de las externalidades de la red urbana sobre el crecimiento urbano y lo compararon con el de las economías de aglomeración desde la perspectiva de la teoría de las externalidades. Este estudio contrasta la forma tradicional de las teorías económicas y urbanas que enfatizan el papel de las economías de aglomeración en la promoción del crecimiento en una región versus las redes urbanas como dicho motor. Para ello, identificaron la red urbana nacional analizando su estructura basada en datos de frecuencia de trenes entre 273 distritos municipales en China. Construyeron el modelo urbano bajo el modelo espacial de Durbin examinando el impacto de las redes urbanas y lo compararon con las economías de aglomeración. Encontraron que las redes tienen un efecto significativo en la promoción del desarrollo económico urbano: las ciudades con mayor centralidad en las redes experimentaron un mayor crecimiento, a diferencia de las economías de aglomeración, que no dependen de la proximidad geográfica, sino de sus conexiones. En el análisis de

las aglomeraciones, utilizaron el mismo índice de localización de Balassa, pero para el de urbanización emplearon una variante que también dependía de la distribución de trabajadores por región¹.

Jiangjiang Kang, Wei Xu, Li Yu y Yueming Ning (2020) propusieron un marco teórico para comprender el cambio de la especialización manufacturera investigando los efectos de la localización, la urbanización y la globalización en el Delta del río Yangtze, China. Utilizaron datos de los años 2004, 2008 y 2013 del Censo Económico Empresarial de China. Modelaron los efectos de la localización, urbanización y la globalización utilizando una regresión de panel de efectos aleatorios con estimación GLS. Los hallazgos indicaron que la especialización manufacturera, definida por el KSI², mostraba una tendencia ascendente significativa entre 2004 y 2008, para luego ser reemplazada por una tendencia decreciente de 2008 a 2013. Desglosaron los índices de localización y urbanización en subfactores¹, encontrando que casi todos estos subfactores se correlacionaban de manera negativa con la especialización manufacturera. Además, encontraron que las áreas urbanas periféricas tenían un alto índice de especialización, mientras que el resto del conglomerado urbano estaba asociado con una baja especialización manufacturera en el Delta del río Yangtze.

Bird, Lebrand y Venables (2020) desarrollaron un modelo de equilibrio espacial de Asia Central y lo utilizaron para analizar posibles efectos de la “Iniciativa de la Franja y la Ruta”³ en la economía de la región. El modelo capturaba las unidades económicas internacionales y subnacionales y su conectividad entre ellas y con el resto del mundo. Debido a la naturaleza teórica del trabajo, definieron un índice de localización en función del número de firmas y la población¹. Encontraron que el ingreso real agregado variaba de una ganancia por el proyecto de menos del 2% del ingreso regional a un 3%, si se consideraban las economías de localización de escala y la movilidad laboral. También, encontraron variaciones dependiendo del tamaño de la región geográfica.

Yoo Ri Kim, Allan M. Williams, Sangwon Park y Jason Li Chen (2021) investigaron los efectos directos de las economías de aglomeración en la productividad del sector turismo, debido a la baja productividad asociada a dicho sector en países desarrollados. Utilizaron datos de panel de 2006 a 2016 del Reino Unido. Entre las variables consideradas estaba el coeficiente de localización de Balassa; sin embargo, los autores no consideraron los efectos de la urbanización. Encontraron un efecto significativo de las economías de aglomeración sobre la productividad en una región específica, siendo dicho efecto negativo para el caso del índice de localización.

Zidong Yu y Xintao Liu (2021) realizaron el estudio de las variaciones espaciales de las economías de aglomeración sectorial y sus relaciones con el ambiente construido y la sociodemografía en Hong Kong. Utilizando datos del censo poblacional del 2016, definieron el índice de localización como el índice de especialización de Balassa y, a partir de ello, midieron la distribución espacial de las economías de

² *Krugman Specialization Index*

³ Proyecto oficial de integración económica mundial del gobierno de la República Popular China.

aglomeración a nivel del Área de Circunscripción del Consejo de Distrito (DCCA). Luego, derivaron los factores del ambiente y los sociodemográficos y evaluaron sus asociaciones con las economías de aglomeración utilizando MCO y modelos GWR⁴. Encontraron heterogeneidad espacial en todos los sectores económicos, donde ciertos sectores tendían a concentrarse en pocas regiones particulares, teniendo dicha agrupación relaciones con diversos factores de ubicación.

ii. Delimitación de áreas urbanas

Duranton (2015), con el objetivo de delimitar las áreas urbanas, desarrolló un algoritmo sencillo basado en los desplazamientos de las personas por motivos laborales entre municipios. Este algoritmo se fundamenta en la idea de que, considerando tres distritos o municipios (C_1 , C_2 y C_3), se puede agregar⁵ a C_1 con C_2 si entre estos *intercambian* individuos por motivos de trabajo por encima de un umbral predefinido (asumido por el autor como el 10%). Además, existe un caso adicional en el que el tercer distrito puede no superar el umbral con los otros dos distritos por separado, pero sí con la unión de ambos. Es importante mencionar que el umbral es definido de manera arbitraria; sin embargo, a pesar de esto, los resultados obtenidos utilizando información del censo de 2010 de Colombia son robustos.

El problema con este método radica en que parte de las definiciones administrativas de cada distrito, lo que no permite discernir si es que estos están sobreestimados o subestimados. Por ejemplo, un distrito podría ser un área muy grande (o pequeña), pero con una población viviendo en un área más pequeña (o grande) dentro (o fuera) de esta división administrativa. Meta (anteriormente Facebook) ha utilizado este método para delimitar lo que ellos llaman “*commuting zones*” (o zonas de desplazamiento); es decir, áreas delimitadas de todos los países en función de en dónde viven y trabajan las personas, basándose en los servicios de geolocalización de los dispositivos móviles. Además, identifica los clústeres en donde la gente vive y trabaja en una misma región⁶. La ventaja de este método es que no se ve limitado por áreas administrativas; sin embargo, el problema con Meta es que no especifica de manera clara los parámetros utilizados, como, por ejemplo, el umbral.

Por otro lado, Uchida y Nelson (2009) desarrollaron un índice de aglomeración como una nueva forma de caracterizar las ciudades. Para crear este índice, utilizaron tres variables: la densidad poblacional, el tamaño de la población y los tiempos de viaje. Fijaron los umbrales para los valores mínimos de cada uno de estos tres criterios y, mediante el uso de un algoritmo de inteligencia artificial, determinaron los índices de aglomeración, comparándolos con datos de la población urbana medida por la Organización

⁴ *Geographically Weighted Regression*

⁵ Agregar es equivalente a formar una sola ciudad.

⁶ Los mapas de *commuting zones* pueden ser descargados desde:

<https://dataforgood.facebook.com/dfg/tools/commuting-zones>

de las Naciones Unidas (ONU). Para ello, crearon una grilla con cuadrículas de 1 km² (o el equivalente a 30" × 30" geográficos) para el año 2000, utilizando dos fuentes de datos: GRUMP⁷ y LandScan⁸.

Con el avance de la tecnología, surgieron nuevas formas de determinar las ciudades, como el uso de imágenes satelitales, ya sea en fotografías diurnas para calcular las densidades poblacionales como hizo Dijkstra y Poelman (2014), o el uso de imágenes nocturnas para analizar la luz emitida por las ciudades como hicieron Ch, Martin y Vargas (2014). En el caso de Dijkstra y Poelman (2014), propusieron un algoritmo basado en una nueva definición de ciudad: un área densamente poblada. Además, definieron otras dos categorías: área rural (bajamente poblada) y suburbios (medianamente poblados). Utilizando las bases de datos de LAU2⁹ y CORINE Land Cover¹⁰, trabajaron con grillas de 1 km² donde están geocodificadas la ubicación y la población. Dentro de su algoritmo, clusterizaron los mapas, completaron y corrigieron la clasificación, y tuvieron en cuenta los efectos de los bordes y cuadros no habitados, entre otros aspectos. Esta idea fue adoptada por Henderson y Kriticos (2019) para delimitar las ciudades con LandScan, aunque estos últimos proporcionaron más detalles sobre cómo se realizaron estas delimitaciones. En ambos casos, definieron umbrales para separar zonas urbanas (suburbios y ciudades con alta densidad) de zonas rurales. Para estos autores, las cuadrículas con densidades superiores a 1500 habitantes por km² y una población mínima de 50,000 habitantes son consideradas centros urbanos. Las cuadrículas con densidades superiores a 300 habitantes por km² y una población mínima de 5000 habitantes son consideradas clústeres urbanos. Aquellas que no están incluidas en ninguno de los dos grupos son consideradas zonas rurales. Debido a la existencia de cuadrículas vacías, utilizaron un *suavizamiento*, que consiste en agrupar varias cuadrículas que incluyan estos vacíos, promediando las densidades para obtener una densidad promedio para cada cuadrícula. Esta idea será adoptada en este trabajo para delimitar áreas urbanas en Perú: se buscará definir áreas urbanas como aquellas por encima de una determinada densidad, pero, a diferencia de Dijkstra y Poelman (2014), se utilizarán cuadrículas más pequeñas proporcionadas por Meta y se cambiará el suavizamiento de cuadrículas por una *densidad extendida*, que será definida más adelante. Por lo tanto, a diferencia del trabajo citado, las áreas urbanas en el presente estudio serán un polígono que encierre a todas aquellas cuadrículas que superen un umbral en la densidad extendida.

Por otro lado, Ch *et al.* (2014) desarrollaron un método para medir el tamaño y el crecimiento de las ciudades utilizando la luz nocturna captada por los satélites para diversas partes del mundo, proporcionadas por la Asociación Nacional Oceánica y Atmosférica (NOAA). Esta propuesta fue

⁷ Global Rural-Urban Mapping Project; las bases de datos pueden ser descargadas desde:

<https://sedac.ciesin.columbia.edu/data/collection/grump-v1/sets/browse>

⁸ Las bases de datos actualizadas pueden ser descargadas, previo registro, desde:

<https://landscan.ornl.gov/landscan-datasets>

⁹ Local Administrative Unit proporcionado por Eurostat, las bases de datos pueden ser descargadas desde:

<https://ec.europa.eu/eurostat/web/nuts/local-administrative-units>

¹⁰ Un repositorio de imágenes satelitales que funciona desde 1985. Las bases de datos pueden ser descargadas desde:

<https://land.copernicus.eu/pan-european/corine-land-cover>

inicialmente desarrollada por Imhoff, Lawrence, Stutzer y Elvidge (1997), Sutton (2003) y Small, Pozzi y Elvidge (2005), y ha sido mejorada por diferentes investigadores a lo largo de los años. Ch *et al.* (2014) utilizaron imágenes satelitales de luces nocturnas en alta resolución de los años 2000 y 2010 para medir áreas metropolitanas y cómo han crecido a partir de ellas. Debido a los problemas intrínsecos de las imágenes de la luminosidad nocturna, como la borrosidad, la variabilidad de las luces a lo largo del tiempo y los diferentes efectos captados por diferentes satélites, los autores presentaron metodologías propuestas por otros investigadores para abordar estos problemas y obtener imágenes en las que se pueda confiar, a partir de las cuales estimaron los tamaños de las ciudades, la población, entre otros aspectos.

Un estudio interesante, llevado a cabo por Bosker, Park y Roberts (2019), comparó los cuatro métodos presentados anteriormente para el caso de un país en vías de desarrollo como Indonesia¹¹. Los resultados obtenidos en cuanto a la delimitación de áreas urbanas difirieron entre sí. En particular, el método de Duranton (2015) generó un mayor número de áreas urbanas. Mientras que los resultados obtenidos utilizando el método de la clusterización de Dijkstra y Poelman (2014) y el índice de aglomeración de Uchida y Nelson (2009) arrojaron resultados poco confiables cuando los umbrales fueron muy bajos. Por otro lado, cuando los umbrales fueron muy grandes, los resultados obtenidos por el método de clusterización parecieron “razonables”, aunque diferían de los obtenidos a partir de la luz nocturna.

iii. Densidad poblacional y centros de negocios

El modelo monocéntrico fue formalmente introducido por Alonso (1964) y posteriormente fue ampliado por diferentes autores, quienes incluyeron nuevos aspectos como el transporte, la producción y los costos de vivienda, los cuales fueron unificados por Fujita (1989). Cuando se hace referencia a un modelo monocéntrico, se plantea que en una ciudad existe un único punto de referencia donde se desarrollan todas las actividades económicas productivas, asumiendo que todas las personas trabajan y viven alrededor de este lugar. A este lugar se le conoce como el distrito central de negocios (CBD, por sus siglas en inglés). Los modelos solían partir del supuesto de la existencia de dicho centro; sin embargo, fue gracias a Fujita (1986) que se lograron formalizar los elementos para la existencia endógena de los CBD.

Además del modelo monocéntrico, también existe el modelo policéntrico que fue propuesto por Fujita y Ogawa (1982). Este modelo surge de la idea de que a veces, cuando los costos de desplazamiento son relativamente altos, es posible la formación de varios distritos de negocios donde se permite el equilibrio múltiple.

Al combinar estos modelos con la densidad poblacional, obtenemos que el modelo básico de densidad de población derivado del modelo monocéntrico está dado por:

¹¹ El cual puede ser un buen referente para el caso de Perú

$$D(u) = A \cdot \exp\left(-\sum_{k=1}^{k=q} b_k u^k\right) \quad [1]$$

Donde,

$D(u)$ es la densidad de población a una distancia u del CBD y A , b_k son parámetros del modelo.

Siguiendo a Clark (1951), Mills (1970) y Newling (1966), tomaremos $q = 1$. La ecuación [1] debe ser calibrada bajo la condición (McDonald y Bowman, 1976):

$$\int_0^{\bar{u}} A \cdot \exp(-bu) f(u) du = \text{Población total} \quad [2]$$

Para este modelo, existen estudios que muestran los posibles problemas que podrían surgir dependiendo del método utilizado: OLS, MCNL o *Mill's Two Points Method* (Greene y Barnbrock, 1978; White, 1977; Straszheim, 1974). Los principales problemas están vinculados con la especificación del tipo de error asumido y el método escogido; sin embargo, existen pruebas que pueden ser utilizadas para validar la especificación de los errores.

Otro tipo de sesgo suele estar relacionado con las diferentes áreas de las regiones donde se analiza su densidad (Frankena, 1978). Es decir, dada una distancia u existe una región con un área A_u tal que, para diferentes distancias, podrían existir regiones con distintas áreas, esto es común cuando se analizan distancia entre distritos sobre un distrito central. Sin embargo, esto no supondrá un problema mayor en nuestro análisis, ya que todas las cuadrículas tienen las mismas dimensiones en área. Además, existen otros factores que influyen en la densidad, como la propia existencia de otros centros, tal como veremos en las siguientes líneas.

Cuando pasamos a un modelo policéntrico, la ecuación básica tomaría la siguiente forma:

$$D_i = \sum_{j=1}^{j=m} A_j \cdot \exp\left(-\sum_{k=1}^{k=q} b_{jk} d_{ij}^k\right) \quad [3]$$

Donde:

D_i es la densidad de la población en la ubicación i , d_{ij} es la distancia entre la ubicación i y el centro j y A_j , b_{jk} son parámetros asociados con el centro j .

Esta ecuación es simplemente una generalización de la ecuación para el caso monocéntrico, donde los parámetros A_j pueden interpretarse como la cantidad de “influencia” del j -ésimo centro sobre el área urbana.

Una nota importante es que la forma de medir la distancia no se puede reducir únicamente a la métrica Euclidiana (Griffith y Wong, 2006). En general, dado que la i -ésima región que tiene una densidad D_i cuyo centroide se encuentra en el punto con coordenadas (x_i, y_i) y está a una distancia d_{ij} del j -ésimo centro cuyas coordenadas son (X_j, Y_j) , se puede asumir una métrica de Minkowsky de forma general tal que la distancia d_{ij} sea definida como:

$$d_{ij} = (|x_i - X_j|^p + |y_i - Y_j|^p)^{1/p}$$

Esto es especialmente útil si queremos tener en cuenta los efectos de la naturaleza curvilínea de la tierra y la propia geometría interior de los caminos en las ciudades. Por ejemplo, en una ciudad con calles en formas cuadrículadas, quizás una métrica con $p = 1$ podría representar mejor las distancias.

3. MARCO METODOLÓGICO

Como se ha revisado en la sección anterior, las economías de aglomeración mejoran la productividad de las empresas a través de economías externas, ya sea bajo la forma de economías de localización o de economías de urbanización. Estas pueden manifestarse mediante la transmisión de conocimiento intraindustrial (*knowledge spillovers*), la reducción de los costos de transporte de insumos y productos por los recursos compartidos (*input sharing*) y los beneficios por un mercado laboral más eficiente (*labor market pooling*) en el caso de las economías de localización; y a través de la fertilización cruzada de diferentes industrias en un territorio en el caso de las economías de urbanización.

i. Modelo

Primera etapa

Para estimar la presencia de un patrón de economías externas de aglomeración por industria (o sector), se considerará la función de producción Cobb-Douglas, que incorpora seis factores y asume rendimientos constantes y neutralidad en el sentido Hicks:

$$Y_i = A_i F(L_i, K_i, T_i, N_i, M_i, R_i) = A_i K_i^\alpha L_i^\beta T_i^\gamma N_i^\rho M_i^\theta R_i^\mu \quad [4]$$

donde (a) Y_i es la producción total para la empresa i ; (b) A_i captura el componente tecnológico para la empresa i ; es decir, representa la productividad total de los factores (PTF); (c) K_i hace referencia al nivel de capital para la empresa i ; (d) L_i está vinculado con el empleo o mano de obra para la empresa i ; (e) T_i es el nivel de tierra que hace uso la empresa i ; (f) N_i es el nivel de intangibles que utiliza la empresa i ; (g) M_i está vinculado con el nivel de insumos utilizado por la empresa i ; y (h) R_i hace referencia a los servicios de terceros utilizados por la empresa i .

Al loglinealizar la ecuación para expresar la producción como una ecuación lineal generalizada y considerarla por sectores, la ecuación transforma en:

$$y_i^s = a_i^s + \alpha^s k_i^s + \beta^s l_i^s + \gamma^s t_i^s + \rho^s n_i^s + \theta^s m_i^s + \mu^s r_i^s + \epsilon_i^s \quad [5]$$

Asimismo, es importante mencionar que se adaptará la ecuación lineal generalizada a cada sector económico según la incidencia de los factores en su proceso productivo. Por ejemplo, para el caso del sector de agencias de viaje (AV), se utilizará la siguiente función:

$$y_i^{AV} = \alpha_i^{AV} + \alpha^{AV} k_i^{AV} + \beta^{AV} l_i^{AV} + \mu^{AV} r_i^{AV} + \epsilon_i^{AV} \quad [6]$$

La razón por la cual en la ecuación [6] no se considera la tierra, los insumos y los intangibles es porque el trabajo realizado por las agencias de viaje consiste en conectar a los turistas con los proveedores finales, que muchas veces no son ellas mismas. En ese sentido, la mayor parte del trabajo puede llevarse a cabo mediante coordinaciones remotas, sin la necesidad de contar con un espacio físico ni con

intangibles. Adicionalmente, al ser empresas *intermediarias*, no necesitan ningún tipo de insumo para proveer el servicio.

En ese sentido, se analizaron los 13 sectores productivos restantes y se obtuvieron las siguientes funciones de producción loglinealizadas:

1. Sector agroindustria¹² – AG

$$y_i^{AG} = a_i^{AG} + \alpha^{AG} k_i^{AG} + \beta^{AG} l_i^{AG} + \gamma^{AG} t_i^{AG} + \theta^{AG} m_i^{AG} + \mu^{AG} r_i^{AG} + \epsilon_i^{AG}$$

2. Sector centros educativos no estatales¹³ – CE

$$y_i^{CE} = a_i^{CE} + \alpha^{CE} k_i^{CE} + \beta^{CE} l_i^{CE} + \gamma^{CE} t_i^{CE} + \mu^{CE} r_i^{CE} + \epsilon_i^{CE}$$

3. Sector comercio y servicios¹⁴ - CS

$$y_i^{CS} = a_i^{CS} + \alpha^{CS} k_i^{CS} + \beta^{CS} l_i^{CS} + \mu^{CS} r_i^{CS} + \epsilon_i^{CS}$$

4. Sector construcción¹⁵ - C

$$y_i^C = a_i^C + \alpha^C k_i^C + \beta^C l_i^C + \theta^C m_i^C + \mu^C r_i^C + \epsilon_i^C$$

5. Sector establecimientos de hospedaje¹⁶ – HO

$$y_i^{HO} = a_i^{HO} + \alpha^{HO} k_i^{HO} + \beta^{HO} l_i^{HO} + \gamma^{HO} t_i^{HO} + \theta^{HO} m_i^{HO} + \mu^{HO} r_i^{HO} + \epsilon_i^{HO}$$

6. Sector hidrocarburos¹⁷ – HI

$$y_i^{HI} = a_i^{HI} + \alpha^{HI} k_i^{HI} + \beta^{HI} l_i^{HI} + \gamma^{HI} t_i^{HI} + \theta^{HI} m_i^{HI} + \mu^{HI} r_i^{HI} + \epsilon_i^{HI}$$

7. Sector pesca¹⁸ - P

$$y_i^P = a_i^P + \alpha^P k_i^P + \beta^P l_i^P + \theta^P m_i^P + \mu^P r_i^P + \epsilon_i^P$$

8. Sector manufactura¹⁹ - M

$$y_i^M = a_i^M + \alpha^M k_i^M + \beta^M l_i^M + \gamma^M t_i^M + \theta^M m_i^M + \mu^M r_i^M + \epsilon_i^M$$

9. Sector servicios eléctricos – SE

$$y_i^{SE} = a_i^{SE} + \alpha^{SE} k_i^{SE} + \beta^{SE} l_i^{SE} + \gamma^{SE} t_i^{SE} + \rho^{SE} n_i^{SE} + \theta^{SE} m_i^{SE} + \mu^{SE} r_i^{SE} + \epsilon_i^{SE}$$

10. Sector transportes y comunicaciones²⁰ - TC

$$y_i^{TC} = a_i^{TC} + \alpha^{TC} k_i^{TC} + \beta^{TC} l_i^{TC} + \mu^{TC} r_i^{TC} + \epsilon_i^{TC}$$

11. Sector universidades no estatales²¹ – U

$$y_i^U = a_i^U + \alpha^U k_i^U + \beta^U l_i^U + \gamma^U t_i^U + \rho^U n_i^U + \mu^U r_i^U + \epsilon_i^U$$

12. Sector restaurantes²² – R

$$y_i^R = a_i^R + \alpha^R k_i^R + \beta^R l_i^R + \gamma^R t_i^R + \theta^R m_i^R + \mu^R r_i^R + \epsilon_i^R$$

¹² No se está considerando el factor de producción: intangibles.

¹³ No se están considerando los factores de producción: intangibles e insumos.

¹⁴ No se están considerando los factores de producción: tierra, intangibles e insumos.

¹⁵ No se están considerando los factores de producción: tierra e intangibles.

¹⁶ No se está considerando el factor de producción: intangibles.

¹⁷ No se está considerando el factor de producción: intangibles.

¹⁸ No se están considerando los factores de producción: tierra e intangibles.

¹⁹ No se está considerando el factor de producción: intangibles.

²⁰ No se están considerando los factores de producción: tierra, intangibles e insumos.

²¹ No se está considerando el factor de producción: insumos.

²² No se está considerando el factor de producción: intangibles.

13. Sector servicios²³ – S

$$y_i^S = a_i^S + \alpha^S k_i^S + \beta^S l_i^S + \mu^S r_i^S + \epsilon_i^S$$

Por último, la PTF se calcula según el residuo de Solow, considerando la función de producción estimada previamente mediante la siguiente ecuación:

$$a_i^S = y_i^S - \hat{\alpha}^S k_i^S - \hat{\beta}^S l_i^S - \hat{\gamma}^S t_i^S - \hat{\rho}^S n_i^S - \hat{\theta}^S m_i^S - \hat{\mu}^S r_i^S \quad [7]$$

Segunda etapa

Como se ha mencionado anteriormente, los fenómenos de aglomeración influyen en la PTF; de tal manera que, en particular, las economías de locación y urbanización influyen sobre A . Así, modelamos el componente tecnológico de la siguiente manera:

$$a_i = \delta \text{loc}_{is}^g + \gamma \text{urb}_{is}^g + \mu_j \text{dpto}_{ij} + \rho \text{au}_i + \sigma \text{cue}_i + \theta_m \text{tam}_{im} + \beta(2015 - \text{inicio}_i) + u_i \quad [8]$$

donde loc es una medida de las economías de localización y urb es una medida de las economías de urbanización, ambas para el sector s , evaluadas para la empresa i . Estas dos medidas se estimarán en función de los niveles geográficos (g): provincia (prov), área urbana (au) y centro urbano *expandido* (cue). Asimismo, dpto es una *dummy* de acuerdo con el departamento donde la empresa i realiza sus actividades; esta variable representa la ventaja comparativa del lugar. La variable tam es una *dummy* que hace referencia al tamaño de la empresa²⁴; y, por último, inicio es el año de inicio de actividades de la empresa i y la diferencia $(2015 - \text{inicio})$ refleja la edad de la empresa o experiencia.

Nuestras variables de interés son loc_{is}^g y urb_{is}^g ; así, el objetivo es estimar los parámetros δ y γ de tal manera que hallemos en qué medida impactan las economías de localización y de urbanización sobre la productividad de las empresas.

La hipótesis de las economías de aglomeración sugiere que $\hat{\delta} > 0$ y $\hat{\gamma} > 0$, lo que significa que los sectores con un mayor potencial para las economías de aglomeración tienen una mayor productividad.

La estrategia empírica consiste en primero estimar la ecuación [5] de acuerdo con las especificaciones por sector económico, de donde se hallará a_i^S mediante el residuo de Solow (ecuación [7]). Luego, se procede a estimar la ecuación [8] haciendo uso del resultado anterior. Ambas ecuaciones se pueden estimar mediante regresiones simples de mínimos cuadrados ordinarios (MCO) siempre y cuando todas

²³ No se están considerando los factores de producción: tierra, intangibles e insumos.

²⁴ De acuerdo con el DS N° 013-2013-PRODUCE, las empresas pueden clasificarse en:

- Se denomina Microempresa a aquella cuyas ventas anuales no exceden de 150 UIT.
- Se denomina Pequeña Empresa a aquella cuyas ventas anuales sobrepasan las 150 UIT, pero no exceden de 1,700 UIT.
- Se denomina Mediana Empresa a aquella cuyas ventas anuales sobrepasan las 1,700 UIT, pero no exceden de 2,300 UIT.
- Se denomina Gran Empresa a aquella cuyas ventas anuales sobrepasan las 2,300 UIT.

las variables independientes sean observables y, al menos, débilmente exógenas. Sin embargo, existen problemas de estimación.

Problemas de estimación

Existen algunas limitaciones en el modelo propuesto que surgen al estimar las funciones de producción, en particular, la simultaneidad que implica que los factores de producción pueden estar correlacionados con shocks de productividad inobservables²⁵, y la heterogeneidad no observada. Desde mediados de la década de 1990, varios autores han desarrollado procedimientos de estimación de funciones de producción para hacer frente a estos problemas. Tal es el caso de Levinson y Petrin, quienes, en 2003, con el objetivo de resolver el problema de simultaneidad o variables omitidas²⁶, proponen utilizar la variable observable de insumos para controlar por la productividad inobservable.

Sin embargo, según Akerberg, Caves y Frazer (2015), la solución propuesta es parcial, ya que no aborda el problema de dependencia condicional entre los factores de producción, lo que hace que las estimaciones sean inconsistentes. En particular, para el método de Levinsohn y Petrin, todas las variables de decisión dependen de las mismas variables de estado, lo cual generaría la existencia de colinealidad entre los factores de producción.

ii. Variables y bases de datos

Ecuación [5]

Se utilizará de la Encuesta Económica Anual (EEA) del Instituto Nacional de Estadística e Informática (INEI) del ejercicio económico correspondiente al año 2015. Esta encuesta facilita la recolección de información económica y financiera de las empresas que operan en el territorio nacional y hayan realizado actividades económicas durante el período contable o año calendario (Instituto Nacional de Estadística e Informática, 2017). La encuesta se lleva a cabo de manera anual desde 1997, cuando se aplicó un único cuestionario para empresas en los sectores de manufactura, pesca extractiva y minería. En la actualidad, se recopila información de 14 actividades económicas que incluyen agencias de viajes, agroindustria, centros educativos no estatales, comercio, construcción, establecimientos de hospedaje, hidrocarburos, pesca (extractiva, transformativa y acuicultura), manufactura, servicios eléctricos, transporte y comunicaciones, universidades no estatales, restaurantes y servicios.

²⁵ El primer problema de estimación proviene del sesgo de simultaneidad. En ese caso, posiblemente k_{it}^s y l_{it}^s se encuentren correlacionados debido a que, como consecuencia de un shock económico puede que algunas empresas pueden cerrar y despedir empleados (en el caso el shock sea negativo) o algunas empresas pueden abrir nuevas sucursales y contratar más empleados (en el caso el shock sea positivo).

²⁶ Esta está vinculada con el hecho de que algunas variables no observadas por el econometrista pueden relacionarse con la productividad total de los factores como con las variables independientes.

Es importante destacar que en el *Capítulo 1 - Identificación y ubicación de la empresa*, a cada empresa se le asigna un sector y un formato específico²⁷.

Tabla N° 1: Definición de las variables

Variable	Nombre	Medición
Y_i	Producción total	Margen comercial + producción del ejercicio
K_i	Capital	Maquina y equipo de explotación + unidades de transporte + muebles y enseres + equipos diversos + herramientas y unidades de reemplazo
L_i	Mano de obra	Ejecutivos + empleados y obreros permanentes y eventuales + propietario + personal no remunerado + personal de recibos por honorarios + comisionistas
T_i	Tierra	Terrenos + edificaciones
N_i	Intangibles	Concesiones, licencias y otros derechos + patentes y propiedad intelectual + programas de computadora + costos de exploración y desarrollo + formulas, diseños y productos + reservas de recursos extraíbles + plusvalía mercantil + otros activos intangibles
M_i	Insumos	Compra de materias primas, auxiliares, suministros, repuestos, envases y embalajes
R_i	Servicios de terceros	Transporte, correos y gastos de viajes (alojamiento, alimentación y otros), asesoría y consultoría, producciones encargadas de terceros, mantenimiento y reparaciones, alquileres (terrenos, edificaciones, maquinaria y equipo explotación, equipo de transporte y equipos diversos) + servicios básicos (energía eléctrica, gas y agua + teléfono, internet, radio y cable) + publicidad, publicaciones, relaciones públicas + servicios de contratistas + otros servicios prestados por terceros

Fuente: Elaboración propia

En ese sentido, para obtener la **producción total, insumos y servicios de terceros** de cada empresa, se utilizará el *Capítulo 3 – Estado de resultados* o el *Capítulo 4M – Producción, consumo y resultado del ejercicio*.

Para estimar el nivel de **capital, tierra e intangibles**, se empleará el *Capítulo 5 – Resumen del movimiento de activo inmovilizado e intangible (neto)* o el *Capítulo 7M – Resumen del movimiento de activo inmovilizado e intangible (neto)*.

²⁷ Las empresas pueden clasificarse dentro de tres formatos:

- Formato M: para empresas cuyos ingresos netos sean hasta 6,885,000 soles; equivalente a empresas “medianas o pequeñas”
- Formatos D2 y F2: para empresas cuyos ingresos netos sean de 6,885,001 soles a más; equivalente a empresas “grandes”
- Formato A, A1, B1, P, U y R2: para todas las empresas de un mismo sector

En ese sentido, de acuerdo con el sector económico y las ventas registradas en SUNAT a cada empresa se le asigna una encuesta de un formato en particular. Por ejemplo, a una empresa manufacturera mediana se le entrega la encuesta tipo M o a una empresa de hidrocarburos se le asigna la encuesta tipo U. Ver anexo N° 2.

Para recabar el empleo o la **mano de obra** necesaria para la producción, se utilizará como *proxy* la cantidad de empleados, ya que existe una fuerte relación entre la productividad del trabajo y el número de contratados. Para esto, se utilizará el *Capítulo 11 – Personal ocupado* o el *Capítulo 06MB – Personal ocupado*.

Algunas observaciones en la muestra de la EEA contienen información faltante. Por lo tanto, se llevó a cabo un procedimiento de limpieza de datos, que implica eliminar las observaciones con las siguientes características: variables clave como la producción total o los factores de producción con valores negativos o faltantes. Este proceso reduce nuestras observaciones de 10,618 en todos los sectores a 8,952, cifra que representa más del 84% de la muestra original. En la Tabla N° 2 se muestra que la proporción de observaciones retenidas se mantiene en la gran mayoría de sectores.

Tabla N° 2: Número de observaciones por sector económico

(a)	(b)	(c)	(a)	(b)	(c)		
Agencias de viaje	108	106	98.1%	Agroindustria	273	218	79.9%
Centros educativos no estatales	235	157	66.8%	Comercio	3,244	3,033	93.5%
Construcción	756	658	87.0%	Establecimientos de hospedaje	145	87	60.0%
Hidrocarburos	76	47	61.8%	Pesca	198	183	92.4%
Manufactura	1,958	1,169	59.7%	Servicios eléctricos	84	38	45.2%
Transportes y comunicaciones	894	855	95.6%	Universidades no estatales	83	60	72.3%
Restaurantes	257	155	60.3%	Servicios	2,307	2,186	94.8%

Fuente: Elaboración propia

Nota: Donde en la columna (a) se tiene la cantidad de observaciones totales; en la columna (b) la cantidad de observaciones retenidas y en la columna (c) el porcentaje de observaciones retenidas.

Ecuación [8]

Como se mencionó anteriormente, se crearán las variables de localización y urbanización en dos niveles geográficos: áreas urbanas y centros urbanos *expandidos*. En particular, se construirán tomando como referencia las ubicaciones de los locales principales y anexos de las empresas ubicadas en las áreas urbanas. Este procedimiento constará de cinco pasos:

- i. Definición las áreas urbanas utilizando la data proveniente de los Mapas de Densidad de Alta Resolución (MDAR) de Meta

- ii. Definición de los centros urbanos y centros urbanos *expandidos*
- iii. Geolocalización de las empresas con base en el padrón de SUNAT
- iv. *Scrapping* de empresas participantes de la EEA de 2016
- v. Definición de los índices de localización y de urbanización

A continuación, se explicará cada uno de los pasos mencionados.

- i. Definición las áreas urbanas utilizando la data proveniente de los Mapas de Densidad de Alta Resolución (MDAR) de Meta

Para delimitar las áreas urbanas en Perú, adaptaremos la metodología propuesta por Dijkstra y Poelman (2014) ajustándola a grillas pequeñas y no continuas. En contraste con los autores mencionados, utilizaremos datos provenientes de los Mapas de Densidad de Alta Resolución (MDAR) proporcionados por Meta para el año 2015²⁸. Estos mapas, al igual que los provenientes de otras fuentes como LandScan y WorldPop, se fundamentan en datos censales, análisis de imágenes satelitales, luz nocturna, entre otros, combinados con un algoritmo de inteligencia artificial que proporciona densidades de población por cada cuadrado de un segundo geográfico de arista (aproximadamente 30 m x 30 m). Al descargar y geolocalizar los datos proporcionados por Meta, obtenemos la Figura N° 2, la cual muestra un nivel de detalle impresionante, permitiendo incluso distinguir la silueta del río Rímac pasando por el centro de la ciudad de Lima.

Figura N° 2: Cuadrículas de 1" x 1" proporcionadas por Meta para una sección de Lima



Fuente: Elaboración propia

El uso de estos MDAR podría plantear un problema significativo debido a la gran cantidad de puntos por analizar: más de 3.7 millones de cuadrículas habitadas solo para Perú. Por ese motivo, es necesario

²⁸ Los Mapas de Densidad de Alta Resolución pueden ser descargados desde:
<https://dataforgood.facebook.com/dfg/tools/highresolution-population-density-maps>

encontrar una forma de dividirlos en subgrupos para evitar un excesivo costo computacional. Los datos proporcionados por Meta están dados en grados geográficos sin ninguna delimitación administrativa. Por lo tanto, para agilizar el procesamiento de datos, se utilizarán las divisiones administrativas proporcionadas por el Instituto Nacional de Informática y Estadística (INEI) para ubicar los puntos provistos por Meta dentro de estas divisiones. Al realizar este primer emparejamiento entre puntos y divisiones, surgen ciertas dificultades, como que algunos puntos no se encuentran dentro de ninguna de estas divisiones administrativas. Esto se debe, principalmente, a la precisión de las mediciones o a los posibles puntos base que hayan sido utilizados por el INEI o Meta²⁹. Sin embargo, estos errores no son muy significativos, por lo que se procedió a desarrollar un algoritmo para asignar los puntos que no encajan en ninguna división administrativa a la división más cercana. Es importante señalar que este problema se presenta en los límites exteriores del país.

Dijkstra y Poelman (2014) definen las zonas urbanas como aquellas regiones donde en una cuadrícula de $1 \text{ km} \times 1 \text{ km}$ viven, al menos, 300 personas y que, en total, tenga una población mínima de 5000 personas. Sin embargo, dado que las cuadrículas provenientes de los MDAR son de $1'' \times 1''$ geográfico, se generalizará esta proposición:

Proposición 3.1: Sea x_i, y_i los centroides de las i -ésima cuadrícula, entonces este punto es candidato a ser parte de una zona urbana si existe alguna cuadrícula de $1 \text{ km} \times 1 \text{ km}$ que lo contenga tal que, la población dentro de esta nueva cuadrícula sea mayor a un umbral (L_1) definido por el usuario.

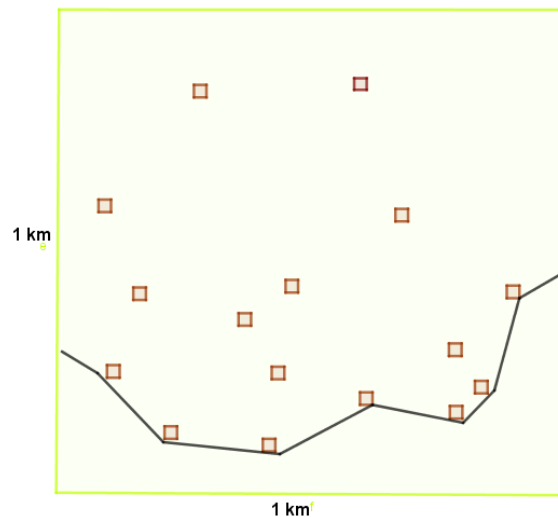
La proposición anterior podría haber sido lenta de ejecutar si no se hubieran ubicado de los puntos en sus divisiones administrativas. Para abordar este problema, se sugiere, según la realidad de cada país, generar grandes divisiones que aseguren que en las fronteras no haya una acumulación significativa de población. En el caso de Perú, las delimitaciones provinciales suelen cumplir con este requisito, con la excepción de Lima y Callao, que son dos provincias administrativamente separadas que podrían formar una única zona urbana.

La Proposición 3.1 permitirá tener una geometría más precisa en los bordes de las zonas urbanas, ya que basta con que el centroide de la cuadrícula pertenezca a alguna cuadrícula mayor de $1 \text{ km} \times 1 \text{ km}$ para que satisfaga el umbral establecido. Se puede afirmar que, como máximo, una cuadrícula de $1 \text{ km} \times 1 \text{ km}$ estará conformada por aproximadamente mil cuadrículas de $1''$ geográfico de arista. Sin embargo, como se puede observar en la Figura N° 3, estas cuadrículas rojas pueden no ser continuas. En este sentido, suponiendo que las cuadrículas marcadas satisfacen la Proposición 3.1 y que se encuentran hacia el límite inferior del área urbana, entonces para delimitarla ya no se utilizaría la

²⁹ En delimitaciones, debido a la que la tierra no es plana, y al trabajar con instrumentos planimétricos, siempre existirán factores que debemos agregar para corregir la curvatura terrestre.

cuadrícula de $1 \text{ km} \times 1 \text{ km}$, sino que se trazaría un polígono que encierre a esos puntos por la parte inferior, tal como se muestra en la línea negra de la misma figura.

Figura N° 3: Comparación de los bordes obtenidos entre una cuadrícula de 1 km y de $1''$ geográfico



Fuente: Elaboración propia

La Proposición 3.1, por sí sola, solo tiene en cuenta las densidades poblacionales; sin embargo, también se deben considerar los umbrales de población mínima de la zona urbana. Es decir:

Proposición 3.2: Se define un área urbana como el conjunto de puntos sobrevivientes a la Proposición 3.1 donde la población total de las aglomeraciones encontradas es mayor que un umbral definido (L_2) por el usuario.

Con el uso de ambas proposiciones, es posible asegurar que las zonas urbanas tengan un mínimo de población y un mínimo de área establecido por el investigador. Sin embargo, puede darse el caso de hallar clústeres aislados, lo suficientemente cercanos para caber dentro de una cuadrícula de 1 km^2 y que satisfagan el umbral establecido por el usuario, como puede observar en la Figura N° 4.

+

+ #

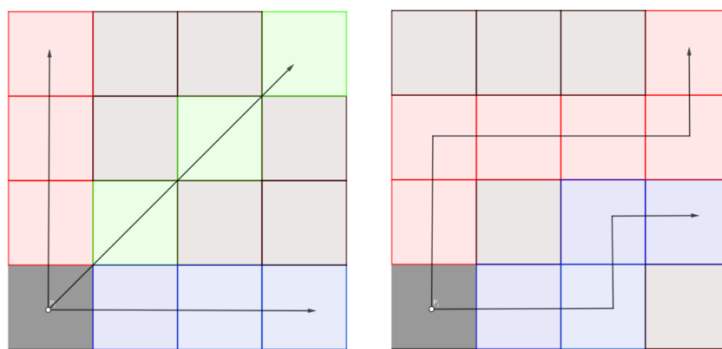
Figura N° 4: Posibles islas al pasar el algoritmo (caso de la ciudad de Trujillo)



Fuente: Elaboración propia

Se llamarán “islas” a estos clústeres aislados. En ese sentido, para garantizar que dichas islas sean excluidas del análisis, se incluirá un filtro adicional para asegurar la continuidad de las cuadrículas.

Figura N° 5: Continuidad de cuadrículas



Fuente: Elaboración propia

Al analizar la Figura N° 5, es posible concluir que, para garantizar la continuidad de las cuadrículas, basta con verificar que, al aumentar la grilla de valor en 1 km de arista, el aumento de debe ser al menos igual al valor del umbral de densidad por kilómetro cuadrado establecido por el usuario. Esto se debe a que, al pasar de una grilla de 1 km a 2 km, como mínimo debe atravesar al menos una grilla habitada de 1 km²; de igual manera, al pasar de 2 km a 3 km y de 3 km a 4 km

Proposición 3.3: Sea x_i, y_i los centroides de las i -ésima cuadrícula que hayan sobrevivido a la Proposición 3.2; es decir, que pertenecen a una cuadrícula de 1 km de arista que cumplen con el umbral definido por el usuario, es posible decir que cumplen la condición de continuidad si

al extender dicha cuadrícula en un kilómetro cuadrado en cualquier dirección, la diferencia de población entre la nueva cuadrícula y la antigua debe ser mayor que el umbral definido inicialmente.

Para efectos del presente análisis, aplicaremos la Proposición 3.3 hasta una extensión de 4 km, asegurando la continuidad hasta dicha distancia. Y, por último, es importante volver a aplicar la Proposición 3.2 a los puntos que sobrevivan a la Proposición 3.3 para asegurar que sigan cumpliendo la condición de población mínima.

Con ello, es posible confirmar que las áreas urbanas cumplen con las tres condiciones:

- Proposición 3.1: Población mínima por km^2
- Proposición 3.2: Población total mínima
- Proposición 3.3: Que sea un área continua

Con estos pasos, es posible determinar las cuadrículas de $1'' \times 1''$ geográfico que pertenecen a un área urbana; sin embargo, para determinar el área del clúster urbano, se procede a generar un polígono cóncavo que encierre a todas estas cuadrículas. Para ello, seguiremos el enfoque de Park y Oh (2012), quienes proponen un algoritmo eficiente³⁰ para la creación de estos polígonos que encierran un conjunto de puntos.

Para el presente trabajo, se probarán diferentes umbrales L_1 y L_2 para corroborar cuáles reflejan mejor la realidad peruana.

ii. Definición de los centros urbanos y centros urbanos *expandidos*

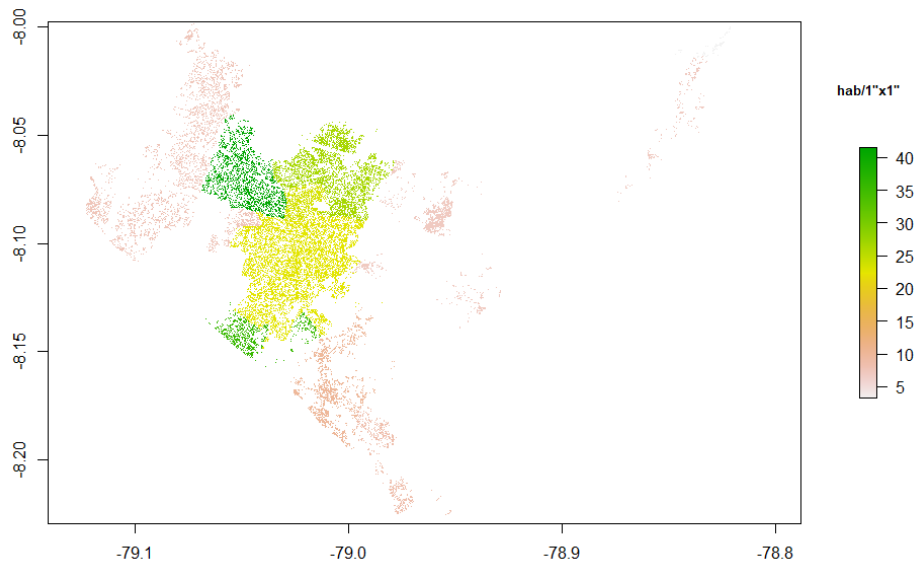
Para la determinación de los centros urbanos, tenemos dos opciones: trabajar con el modelo monocéntrico o policéntrico. De acuerdo con las definiciones dadas en las ecuaciones [1] y [3], tenemos que los centros urbanos son los puntos donde se tienen los niveles de densidad poblacional más elevados. Sin embargo, dependiendo del tamaño y distribución de las poblaciones, el modelo monocéntrico podría resultar poco común. En general, podemos considerar que el modelo monocéntrico como un caso particular del modelo policéntrico cuando $m = 1$ en la ecuación (3). Esto significa que podemos trabajar en forma general bajo un modelo con m centros en una determinada población.

Para determinar los centros, se utilizará el hecho de que un centro es un punto donde se acumula la mayor densidad poblacional. Por lo tanto, primero se deben determinar las densidades poblacionales. Para este caso, no se puede asumir la densidad de un punto como la densidad de una cuadrícula de $1''$ de arista que brinda Meta, ya que son grillas discontinuas y muy pequeñas. Si una grilla tiene baja

³⁰ El algoritmo mencionado fue mejorado por Vladimir Agafonkin liberando en el paquete *concaveman* para el software estadístico R. La documentación completa puede ser encontrada en: <https://github.com/joelgombin/concaveman>

densidad, no significa que la densidad alrededor de ese punto sea baja; podría ser baja en ese punto, pero alta en los que lo rodean.

Figura N° 6: Poblaciones absolutas sobre cada cuadrícula para la ciudad de Trujillo



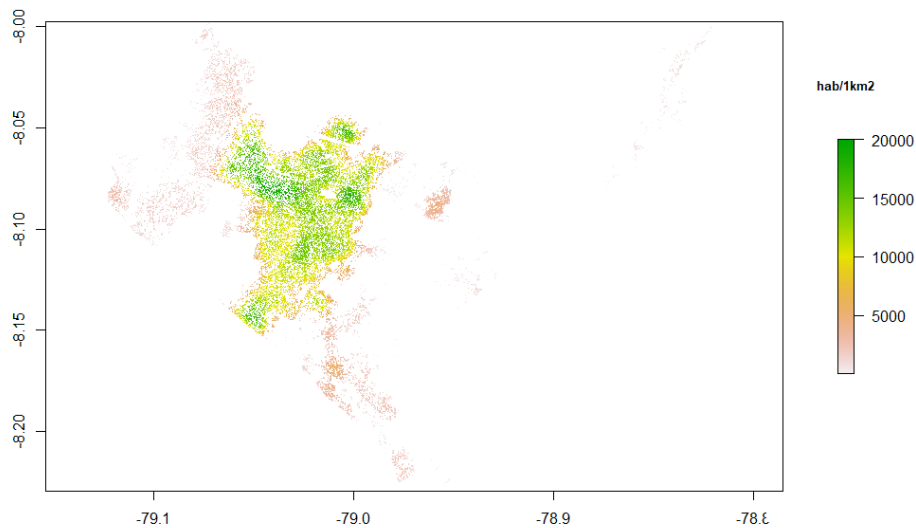
Fuente: Elaboración propia

En la Figura N° 6, se observa que, al graficar las grillas con sus respectivas poblaciones, no se dispone de información que permita calcular los centros, ya que las cuadrículas con densidades altas podrían estar más separadas (generando poca aglomeración), mientras que las cuadrículas con bajas densidades podrían estar más condensadas (generando alta aglomeración).

Por ese motivo, en primer lugar, se determinará lo que llamaremos *población extendida*. Esto significa que a cada punto se le asignará el equivalente a la población que existe en una cuadrícula de 1 km × 1 km con el centroide de la cuadrícula de 1'' de arista en el centro de esta cuadrícula. Esto permitirá tener una visión más clara del *peso* que tiene cada punto ya que altos valores de esta población extendida podrían indicar grandes acumulaciones de población alrededor de ese punto, como se muestra en la Figura N° 7.

Bajo esta nueva definición, para determinar los centros, podemos utilizar la idea de que el primer centro será el punto con mayor población extendida, y se entiende que, conforme nos alejamos de este punto, la densidad debería ir disminuyendo hasta el punto en el que otro centro podría influenciar y causar un aumento de la densidad poblacional. En ese momento, se procede a buscar el siguiente punto con mayor densidad extendida y se repite el proceso hasta encontrar todos los centros. Para evitar fluctuaciones propias de este tipo de datos, se evalúa la disminución de densidad cada R metros de radio. Es decir, si evaluamos la densidad poblacional en un radio R del primer centro, este debe ser mayor a la densidad poblacional de los siguientes R metros y así sucesivamente, hasta que se encuentre la influencia de otro(s) centro(s).

Figura N° 7: Poblaciones *extendidas* sobre cada cuadrícula para la ciudad de Trujillo



Fuente: Elaboración propia

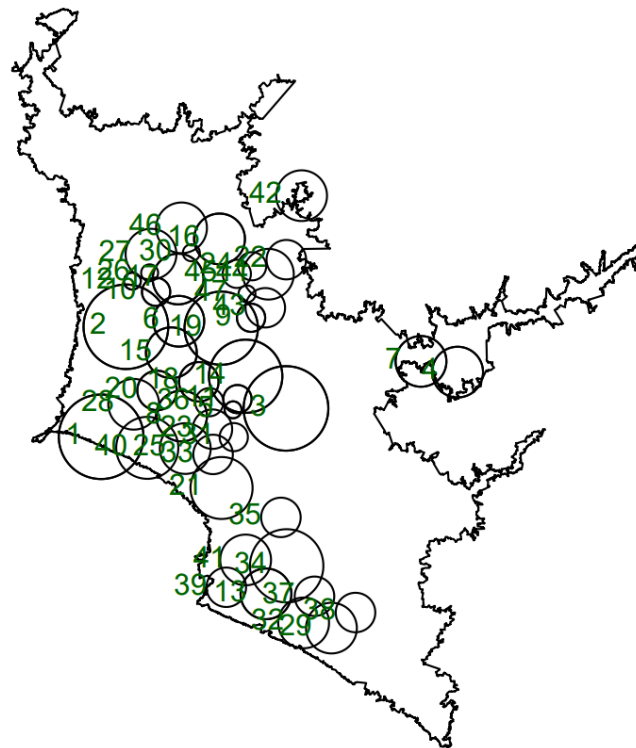
Implementar la idea propuesta líneas arriba supone anteriormente conlleva algunos problemas ya que en la práctica, especialmente si existen múltiples núcleos que afectan al centro inicial, ya que sería complicado determinar qué centro adicional en particular es el que está afectando a la reducción de la densidad. Para sistematizar este proceso, aprovecharemos el hecho de que, dado el centroide de un conjunto de cuadrículas, si tomamos el punto de mayor densidad extendida, entonces al tomar semicírculos concéntricos en dirección al centroide, la densidad poblacional debería ir disminuyendo hasta que, por influencia de los demás centros, esta densidad empiece a aumentar, y se tomará nota de ese radio. Con dicho radio, podemos eliminar los puntos que caigan dentro del círculo con centro en el punto de mayor densidad y el radio encontrado anteriormente. Con esa nueva región, repetiremos el proceso para hallar el siguiente centro, y así sucesivamente.

Los *puntos* encontrados anteriormente serán llamados centros urbanos, y el área sobre el radio en el cual la gradiente de densidad hace la variación mostrando la acción de los demás centros será llamado centro urbano *expandido*, que viene a ser la región de predominancia del centro específico, ya que fuera de ella, la influencia de los demás centros empieza a afectar el comportamiento decreciente de la densidad.

La desventaja de aplicar esta metodología es que podría suceder que tengamos dos centros urbanos *expandidos* superpuestos, como se muestra en la Figura N° 8. Esto puede ocurrir debido a que, en la metodología planteada, tengamos el caso en el que después de n centros, el punto de densidad máxima para determinar el centro $n + 1$ es menor que el punto exterior más cercano al primer centro (que se asume tiene la densidad máxima de la ciudad), por lo que el algoritmo tomará ese punto. Es por ello que es probable que se formen centros urbanos *expandidos* alrededor de los centros con mayor densidad. Para lidiar con este problema, se propone realizar un paso posterior para limpiar los

resultados: que dado que calculamos el $(i+1)$ -ésimo centro urbano y su radio de variación de gradiente, trazamos un círculo concéntrico con el radio encontrado. Luego, condicionamos a que, si dicho círculo cruza con alguno de los i centros urbanos *expandidos* encontrados anteriormente, dicho centro queda excluido. Con ello, aseguramos que no existan superposiciones y evitamos la ambigüedad del punto de densidad extendida máxima.

Figura N° 8: Posible superposición de *centros urbanos expandidos*
Caso de la ciudad de Lima



Fuente: Elaboración propia

iii. Geolocalización de las empresas en base en el padrón de SUNAT

Para realizar la geolocalización de cada empresa según con su dirección y ubigeo, se solicitó el padrón de empresas a SUNAT para el año 2015. Esto se llevó a cabo con el fin de cartografiar todas las empresas peruanas y comprender cómo se distribuyen dentro de las áreas urbanas. Se utilizó el API de Google, que permite, a través de una dirección en texto, obtener su ubicación geográfica. Sin embargo, este proceso presentó algunos problemas, como:

- Direcciones en blanco dentro del padrón: Esto es común para empresas del tipo personas naturales sin negocio, ya que muchas veces son trabajadores de quinta categoría que obtienen su RUC de forma automática por declaración de las empresas que tienen a estas personas en sus planillas. El mismo escenario se repite para personas naturales con negocio, donde la dirección no se brinda por motivos de privacidad.

- Direcciones incompletas y/o inválidas: Algunas direcciones no tenían el nombre completo, algunas poseían abreviaciones diferentes, avenidas que no existen en la base de datos de Google, o la ubicación por manzanas y lotes sin ninguna calle de referencia, entre otros.

Es el segundo error el causante de que las geolocalizaciones hayan salido erradas, manifestándose en ubicaciones fuera del territorio nacional, con ubigeos de la coordenada diferente del ubigeo registrado en su RUC, entre otros problemas.

iv. *Scrapping* de empresas participantes de la EEA de 2016

De acuerdo con el modelo propuesto, es necesario identificar a las empresas. Es importante mencionar que, debido a la confidencialidad de la información sensible, las entidades estatales no proporcionan esta información. Por lo tanto, se ha desarrollado un método para identificarlas.

Aunque no es posible vincular directamente a una empresa con una encuesta específica, sí es factible determinar si una empresa participó en la EEA del año contable 2015. Esto se puede lograr consultando en la página web del INEI³¹, donde se puede acceder a la sección “Historial de Empresas Participantes en las EEA’s anteriores”, ingresar el número de RUC de la empresa y obtener la información necesaria. Sin embargo, dado que hay más de 390,017³² empresas válidas brindadas por SUNAT, se programó un bot automático en R para que realice la búsqueda y captura de información por cada RUC, identificando cuáles participaron en la EEA para el año contable 2015.

Una vez identificado el conjunto de empresas que participaron en la encuesta, se diseñó otro código para realizar un *match* entre: sector, formato de encuesta, ubigeo y año de inicio de actividades. En caso de tener más de un posible *match*, se aproximó a aquella empresa cuyo número de trabajadores declarado a SUNAT sea lo más cercano posible al declarado en la encuesta. Sobre este punto, se pudo notar que los datos declarados por las empresas casi siempre eran muy cercanos a los trabajadores de quinta categoría, por lo que se esta aproximación se hizo hacia los trabajadores de quinta categoría declarados a SUNAT.

Esto conllevó a otros problemas como:

- a) Empresas que cambiaron su ubigeo después de 2015: Para abordar este problema, se realizó una búsqueda de las modificaciones de los ubigeos de las empresas a lo largo de las encuestas de los años siguientes, y se utilizó el último ubigeo declarado³³.

³¹ http://censos.inei.gob.pe/economicas/Portal_RucParticipacionEEA.asp

³² Es importante mencionar que, para hacer el *scrapping* no se puede restringir sólo a las empresas “válidas” ya que podría haber alguna empresa mal geoubicada por Google que pudo haber participado en la encuesta; por lo que, si alguna de estas aparecía como participante, se hace un ajuste manual posterior para no perder esa empresa.

³³ En algunos casos se hicieron búsquedas manuales históricas de varias fuentes para obtener el ubigeo lo más preciso para el año 2015.

- b) Empresas que cambiaron su fecha de inicio de actividades: Se observó que algunas empresas mencionan un año de inicio en 2015 y otro diferente en años posteriores. Para resolver esto, se realizó una búsqueda histórica para obtener la última declaración.
- c) No fue posible realizar ningún *match* con las empresas cuyos RUC comienzan con 10, que corresponden a personas naturales con empresas. Dado que la información sobre las direcciones no es accesible debido a la confidencialidad, este tipo de empresas no pudo ser incluido en el análisis.

v. Definición de índices de localización y de urbanización

Índice de localización

Se utilizará el siguiente índice de aglomeración de Balassa para la empresa en la región geográfica g perteneciente al sector s :

$$LOC_{g,s} = \frac{E_{g,s}/E_g}{E_s/E}$$

Donde

- $E_{g,s}$ es el empleo total del sector s en la región geográfica g
- E_g es la cantidad total de empleo en la región geográfica g
- E_s es la cantidad total de empleo nacional del sector s
- E es la cantidad total de empleo nacional

Si el índice de Balassa es superior a 1, implica especialización. Si es menor a 1, significa que no hay especialización.

La razón por la cual se optó por este indicador es porque las bases de datos obtenidas proveen la información necesaria para obtenerlo. Adicionalmente, como se puede observar en el Anexo N° 1, es uno de los índices más utilizado por los investigadores en los últimos años.

Índice de urbanización

Se utilizará el siguiente índice de urbanización para la empresa localizada en la región geográfica g y perteneciente al sector s :

$$URB_{g,s} = \frac{\frac{1}{\sum_{s'=1, s' \neq s}^n \left(\frac{E_{gs'}}{(E_g - E_{gs})} \right)^2}}{\frac{1}{\sum_{s'=1, s' \neq s}^n \left(\frac{E_{s'}}{(E - E_s)} \right)^2}}$$

Donde,

- E_{ig} representa la cantidad de empleo del sector s' excepto el sector s en la región geográfica g
- $E_{s'}$ es el empleo nacional del sector s' excepto el sector s

La razón por la cual se optó por este indicador es porque las bases de datos obtenidas proveen la información necesaria para obtenerlo. Además, a diferencia de otros indicadores, este se fundamenta en la mano de obra de las empresas en lugar de la población de las regiones.

Estos índices se calculan con la información de las empresas válidas de SUNAT. Así, habrá un coeficiente de LOC y URB por cada par <región geográfica-sector>.

Dado que estos índices dependen de la cantidad de trabajadores, para el caso de los locales anexos, se determinó el número de trabajadores en estos según el principio de proporcionalidad: los locales anexos en ciudades más pobladas tendrán una mayor cantidad de trabajadores. Con esto, se realiza la distribución de personal entre los locales principales y sus anexos.

Luego, se determinó a qué <región geográfica-sector> pertenece cada empresa participante de la EEA y se le asignaron los índices correspondientes a dicho par, calculados anteriormente.

4. ANÁLISIS DE LOS RESULTADOS

Después de describir los modelos y las bases de datos en los capítulos anteriores, en el presente capítulo, se reportan, analizan y evalúan los resultados estimados.

i. Estadísticas descriptivas

Las estadísticas descriptivas se muestran en la Tabla N° 3³⁴. Nuestros datos cubren 8,952 observaciones. La producción total promedio es de 39.6 millones de soles. El número medio de empleados es de 219 y el capital medio de 21.4 millones de soles. Como se puede observar, las desviaciones estándar son bastante altas en relación con las medias, lo que indica la existencia de distribuciones muy sesgadas.

Tabla N° 3: Estadísticas descriptivas de las variables de interés

Variable	Obs	Mean	Std. dev.	Min	Max
Producción total	8,952	39.58	238.65	0.000760	9,449.31
Capital	8,952	21.36	315.53	0.000001	21,761.32
Mano de obra	8,952	219.09	2,844.46	1.000000	261,274.00
Tierra	1,931	39.49	184.88	0.001017	3,863.42
Intangibles	98	11.13	48.42	0.000001	402.09
Insumos	3,330	25.28	217.75	0.000001	7,166.41
Servicios de terceros	8,952	10.31	72.06	0.000020	4,071.24

Fuente: Elaboración propia

Nota: La producción total y los factores de producción (salvo la mano de obra) están medidos en millones de soles y la mano de obra en cantidad de personas.

ii. Primera etapa

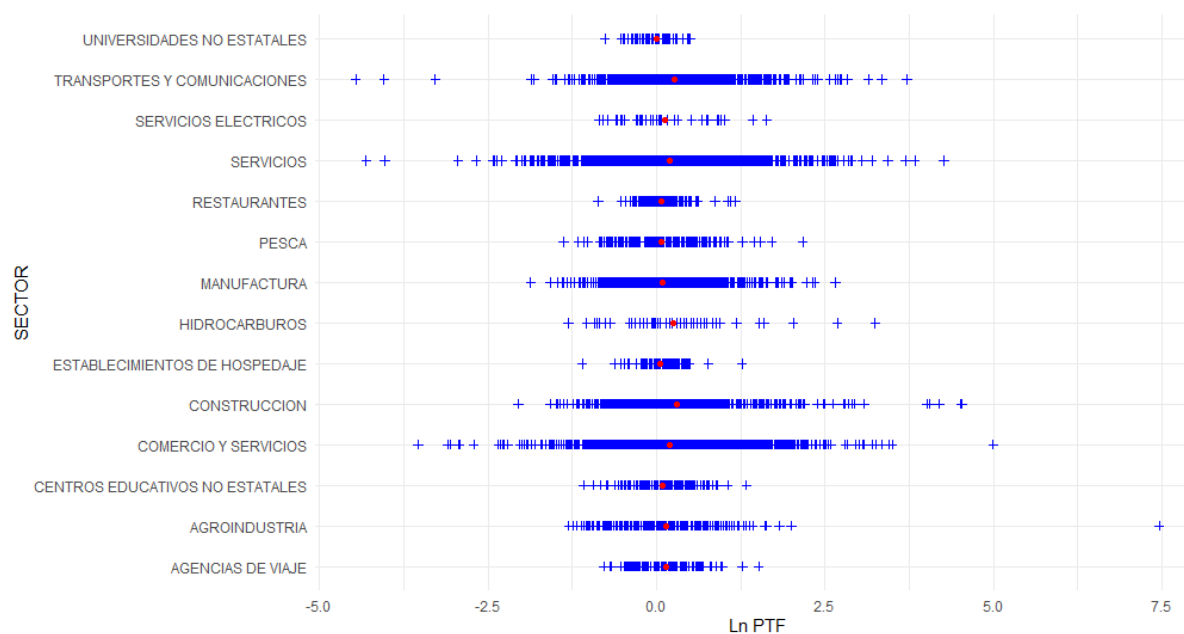
En el Gráfico N° 1 se pueden observar los resultados de la regresión [7] adaptada a cada uno de los sectores³⁵. Es decir, la estimación de la PTF por empresa (en logaritmo) para cada sector económico para 2015. Se puede observar la existencia de valores extremos como, por ejemplo, un valor de 7.45 para una empresa del sector agroindustria o un valor de -4.46 para una empresa perteneciente al sector transportes y comunicaciones. También se puede notar que el rango de la PTF es más amplio para los sectores comercio y servicio, construcción, transportes y comunicaciones y servicios. Asimismo, los sectores hidrocarburos y servicios eléctricos parecen tener una mayor dispersión de PTF.

Adicionalmente, en el Gráfico N° 1 se encuentra marcada en círculo rojo la PTF promedio por sector. Estos varían en el rango de -0.005 y 0.3; es decir, no varía mucho entre sectores.

³⁴ En el Anexo N° 3 se encuentran las estadísticas descriptivas de las variables de interés por sector económico.

³⁵ En el Anexo N° 4 se encuentran los resultados de la estimación de la regresión [5] personalizada para cada sector. Estos coeficientes estimados sirvieron de insumo para obtener los resultados de la regresión [7]; es decir, la PTF (en logaritmo) por empresa.

Gráfico N° 1: Rangos de la PTF (en logaritmos) por sector económico para 2015



Fuente: Elaboración propia

iii. Áreas urbanas

Implementada la metodología descrita anteriormente para el caso de Perú, se realizaron pruebas con distintos umbrales L_1 , desde 250 hasta 1,500 habitantes por kilómetro cuadrado. Se encontraron resultados creíbles utilizando uso de un L_1 equivalente a 500 habitantes por kilómetro cuadrado. En particular, con este resultado y usando un L_2 igual a 10,000 personas por área urbana, se identificaron 65 áreas urbanas, como se muestra en la Tabla N° 6. En la lista, Lima se encuentra en primer lugar con una población cercana a los diez millones de habitantes dentro de la zona urbana y un área de 1,208 km². En segundo lugar, está Arequipa, seguida inmediatamente por la ciudad de Trujillo. Es importante destacar que ambas ciudades tienen poblaciones muy similares; sin embargo, el área ocupada por el área urbana de Arequipa es más del doble que la de Trujillo. Una posible razón para esto podría ser que Trujillo se encuentra en una zona costera donde no hay accidentes geográficos pronunciados que limiten la expansión de la ciudad, como se observa en la Figura N° 9 (c); a diferencia de Arequipa, donde la presencia de ríos, lagos, montañas y zonas de cultivos dispersa la población, como se ve en la Figura N° 9 (b). Esto implica que la densidad promedio en Arequipa es menor que la de Trujillo.

Un escenario similar se observa en el caso de Huancayo, con una superficie muy parecida a la de Trujillo, pero con una población un poco mayor a la mitad de la de Trujillo. Por otro lado, en Cusco, donde la ciudad se ha concentrado en un área sin zonas de cultivo y con pocos accidentes geográficos, lo que le ha resultado en una alta densidad promedio. Además, al comparar dos ciudades costeras, se encuentran marcadas diferencias, como en el caso de Chiclayo, que tiene un área mayor que Trujillo, pero una población significativamente menor. Una posible explicación para esto podría ser que

Chiclayo se ha expandido a través de grandes áreas de cultivo, como se observa en la Figura N° 9 (d), mientras que, en el caso de Trujillo, las áreas de cultivo rodean la ciudad.

Tabla N° 4: Áreas urbanas halladas

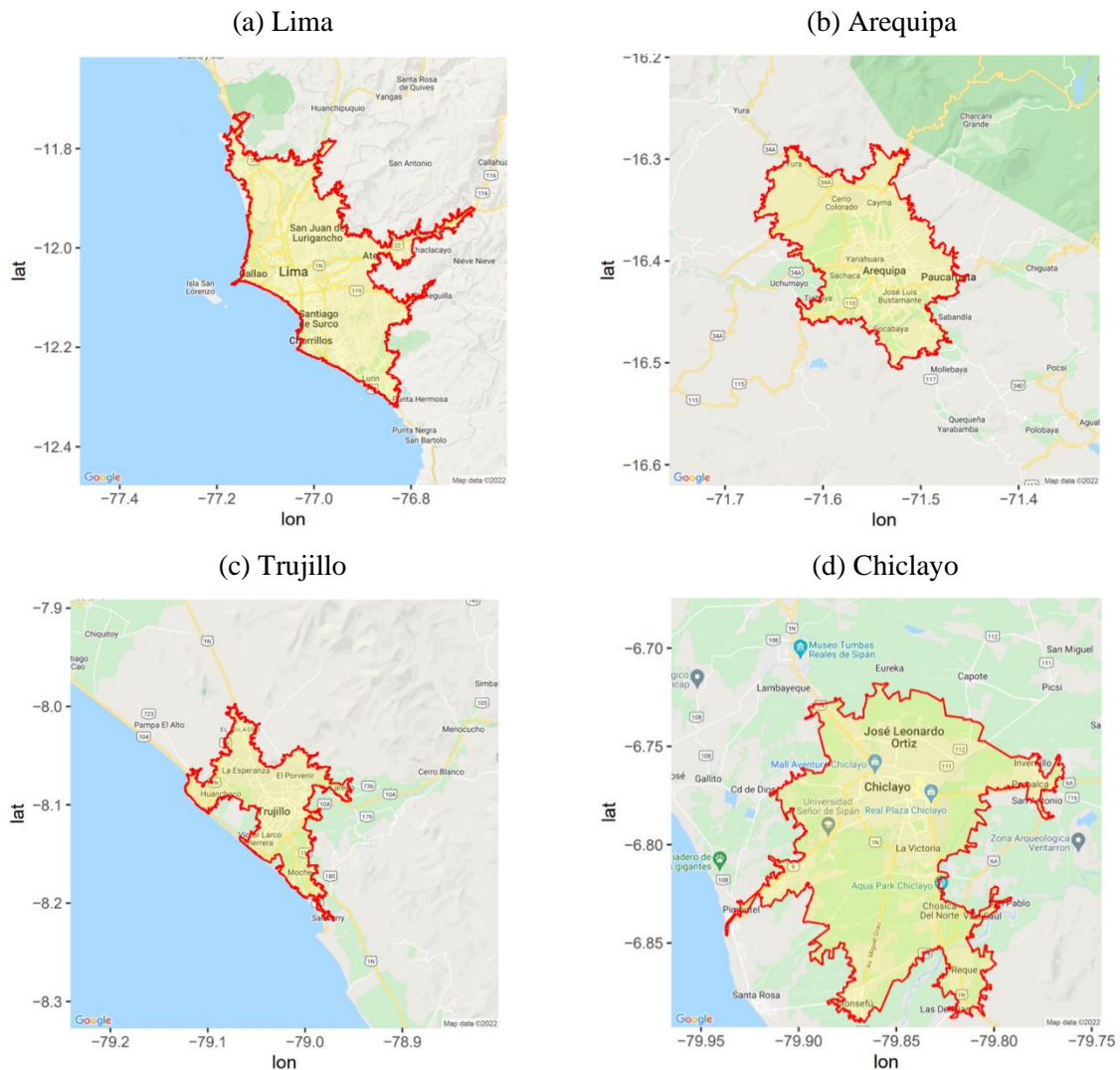
N°	Área urbana	Población	Área (km ²)	N°	Área urbana	Población	Área (km ²)
1	Lima	9,643,315	1,208	34	Mariscal Nieto	50,516	30
2	Arequipa	911,293	285	35	Abancay	49,453	22
3	Trujillo	910,656	141	36	Leoncio Prado	48,229	12
4	Chiclayo	641,197	162	37	Moyobamba	46,856	17
5	Huancayo	462,100	142	38	Andahuaylas	46,289	31
6	Piura	423,689	121	39	Huanta	40,494	23
7	Cusco	404,066	73	40	Canchis	34,889	19
8	Coronel Portillo	315,345	94	41	Pasco	34,108	5
9	Ica	296,421	123	42	Caylloma	33,591	29
10	Santa	283,658	60	43	Camaná	32,099	12
11	Tacna	259,299	92	44	Tarma	31,813	11
12	Cajamarca	234,597	153	45	Rioja	29,328	4
13	San Román	231,372	101	46	Lambayeque	25,206	26
14	Sullana	198,481	78	47	Huancavelica	24,575	5
15	Huamanga	185,565	50	48	Satipo	24,468	12
16	Chincha	166,861	53	49	Nasca	23,158	11
17	Huánuco	163,064	73	50	Chepén	21,453	4
18	San Martín	138,320	37	51	Hualgayoc	21,163	33
19	Puno	121,609	36	52	Zarumilla	19,811	4
20	Huaura	116,128	42	53	Alto Amazonas	19,478	5
21	Huaraz	104,584	63	54	La Convención	17,405	6
22	Maynas	102,732	57	55	Chupaca	16,963	14
23	Huaral	97,945	60	56	Casma	16,812	6
24	Tumbes	93,813	13	57	Utcubamba	15,565	5
25	Pisco	84,121	24	58	Virú	15,287	6
26	Paita	79,808	13	59	Sánchez Carrión	14,761	9
27	Talara	73,446	13	60	Celendín	14,597	10
28	Cañete	67,758	87	61	Urubamba	14,586	9
29	Ilo	64,500	18	62	Chachapoyas	14,117	4
30	Barranca	64,146	21	63	Mariscal Cáceres	13,386	3
31	Tambopata	57,918	22	64	Cajabamba	11,184	9
32	Chanchamayo	55,393	25	65	Huaylas	10,232	7
33	Jaén	54,451	14				

Fuente: Elaboración propia

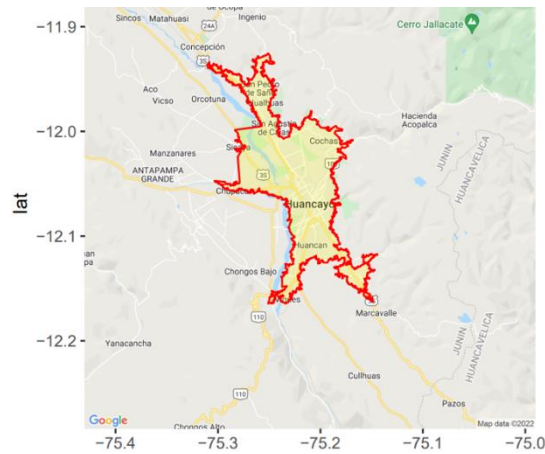
En la Figura N° 9 se presentan los mapas de las áreas urbanas de las seis ciudades más pobladas de Perú, de acuerdo con los resultados obtenidos previamente en la Tabla N° 6. Las imágenes están mostradas con diferentes niveles de zoom satelital para permitir una mejor apreciación de las

superficies. Específicamente, para Lima se utilizó un zoom de nivel 10; para Arequipa, Trujillo y Huancayo, se empleó un zoom de nivel 11; y para Chiclayo, se utilizó un zoom de nivel 12. Además, en el Anexo N° 5 se pueden encontrar los mapas correspondientes a las otras 59 áreas urbanas identificadas.

Figura N° 9: Visualización gráfica de las cinco ciudades más pobladas y sus áreas urbanas



(e) Huancayo



Fuente: Elaboración propia

iv. Centros urbanos y centros urbanos *expandidos*

Se han identificado los centros urbanos utilizando el criterio de variación de la gradiente y evitando la superposición de los mismos, tal como se describe en la metodología. En total, se han obtenido 144 centros en todo el país, cuyos centros por cada área urbana se detallan en el Anexo N° 5. Además, en la Tabla N° 5 se presenta la cantidad de centros por área urbana determinada en el paso anterior.

Tabla N° 5: Número de centros urbanos hallados por área urbana

N°	Área urbana	N° Centros	N°	Área urbana	N° Centros
1	Lima	18	34	Mariscal Nieto	2
2	Arequipa	4	35	Abancay	1
3	Trujillo	6	36	Leoncio Prado	2
4	Chiclayo	1	37	Moyobamba	1
5	Huancayo	2	38	Andahuaylas	2
6	Piura	2	39	Huanta	1
7	Cusco	2	40	Canchis	2
8	Coronel Portillo	2	41	Pasco	1
9	Ica	4	42	Caylloma	1
10	Santa	3	43	Camaná	2
11	Tacna	2	44	Tarma	1
12	Cajamarca	4	45	Rioja	1
13	San Román	3	46	Lambayeque	2
14	Sullana	5	47	Huancavelica	1
15	Huamanga	2	48	Satipo	3
16	Chincha	1	49	Nasca	1
17	Huánuco	4	50	Chepén	1
18	San Martín	1	51	Hualgayoc	3
19	Puno	2	52	Zarumilla	1
20	Huaura	1	53	Alto Amazonas	1
21	Huaraz	3	54	La Convención	1
22	Maynas	5	55	Chupaca	2
23	Huaral	3	56	Casma	1
24	Tumbes	2	57	Utcubamba	1
25	Pisco	3	58	Virú	2
26	Paíta	1	59	Sánchez Carrión	2
27	Talara	1	60	Celendín	1
28	Cañete	4	61	Urubamba	1
29	Ilo	2	62	Chachapoyas	1
30	Barranca	2	63	Mariscal Cáceres	1
31	Tambopata	1	64	Cajabamba	1
32	Chanchamayo	2	65	Huaylas	1
33	Jaén	1			

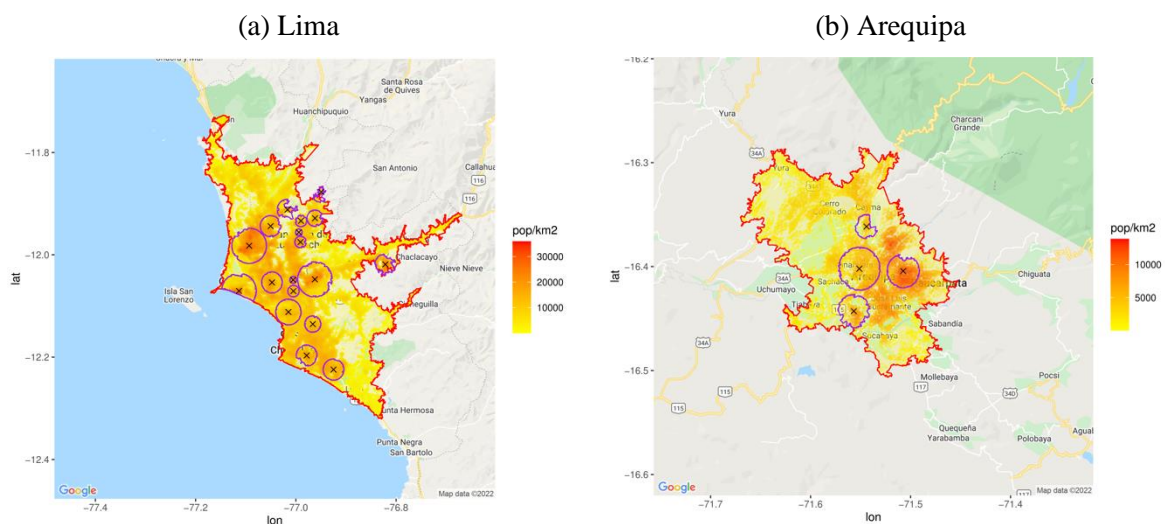
Fuente: Elaboración propia

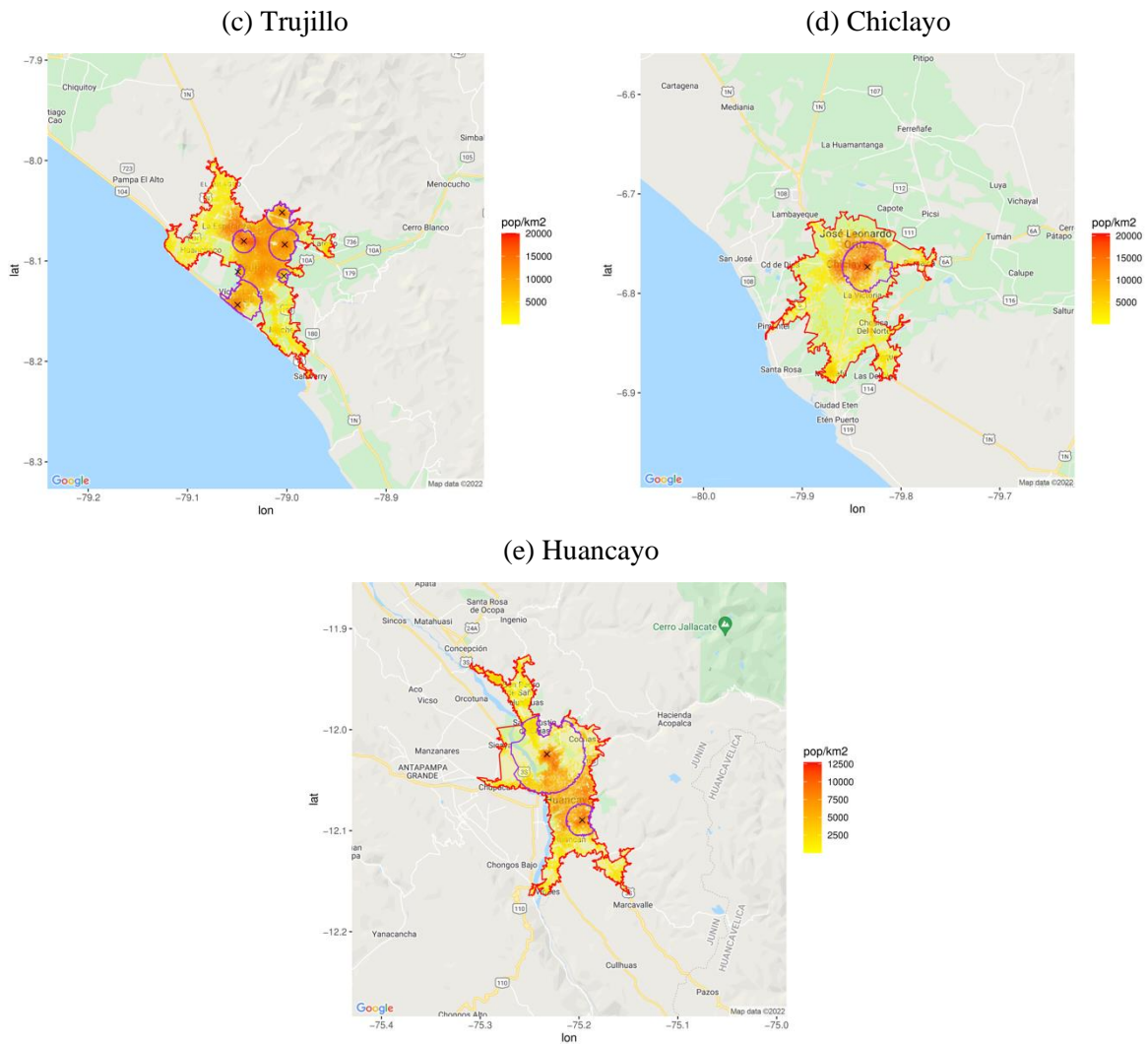
De todas las áreas urbanas, 31 poseen un solo centro, convirtiéndose en ciudades monocéntricas, lo que muestra que casi la mitad de las ciudades en Perú son de esta categoría. Por el otro lado, el resto de áreas urbanas son policéntricas, siendo el caso más especial el de Lima, que, como se puede observar en la Figura N° 10 (a), posee 18 centros a lo largo de toda la ciudad.

Asimismo, al determinar los radios sobre los cuales la gradiente de la población cae, y, por lo tanto, determinar lo que denominamos anteriormente como centros urbanos *expandidos*, se obtiene que estos son muy variados en tamaño y densidad de su punto central o centro urbano.

En la Figura N° 10 se muestran los centros urbanos y centros urbanos *expandidos* para las 5 ciudades más pobladas, y además se observan las escalas de lo que denominamos anteriormente población *extendida*. Se puede ver que los centros *expandidos* encierran a estos puntos con mayor densidad o población *extendida*. En particular, se puede notar el caso de la ciudad de Chiclayo, que, teniendo una población de más de seiscientas mil personas, se obtiene un único centro; a diferencia de Huancayo, que tiene una población significativamente menor, pero posee dos centros. Una posible explicación de este hecho es que, como se mencionó anteriormente, Chiclayo se ha expandido sobre espacios de cultivo, por lo que la densidad sobre los alrededores será muy baja respecto al centro de la ciudad, lo que impide la formación de centros adicionales. Un escenario distinto es el caso de la ciudad de Trujillo, que posee seis centros urbanos y cuya agricultura se ha desarrollado en los alrededores, lo que le ha permitido desarrollar concentraciones de población en distintos puntos de la ciudad. El caso de Arequipa, con una población parecida a la de Trujillo, pero con un área que la dobla, muestra cuatro centros urbanos, donde el primero de ellos es bastante pronunciado en comparación con los otros tres.

Figura N° 10: Visualización gráfica de las cinco ciudades más pobladas y sus áreas urbanas, centros urbanos *expandidos* y centros urbanos





Fuente: Elaboración propia

Nota: El área urbana es la delimitación en rojo, el o los centros urbanos *expandidos* están delimitados en morado y, por último, dentro de cada centro urbano *expandido* se puede hallar su centro urbano correspondiente en negro.

v. Segunda etapa

i. Geolocalización de empresas con base en el padrón de SUNAT

Con la metodología planteada y tras filtrar los resultados para eliminar las “anomalías” junto con las empresas aparentemente sin trabajadores, se obtuvo un total de 269,019 empresas válidas, con 181,604 locales anexos válidos, lo que suma un total de 450,623 empresas en todo el país. Con estos datos, se calcularán los índices de localización y urbanización para cada sector en cada región geográfica.

ii. *Scrapping* e identificación de las empresas participantes de la EEA de 2016

Con los datos obtenidos del *scrapping* y realizando el *match* con la data de SUNAT, se pudo identificar a 6,204 empresas participantes de la Encuesta Económica Anual del año contable 2015.

iii. Definición de índices de localización y urbanización

Como se mencionó inicialmente, los coeficientes de localización y urbanización fueron determinados a partir de la data de empresas de SUNAT. Es decir que, con las 450,623 empresas, para cada tipo de región geográfica, se obtuvo una cantidad de índices de localización (LOC) y urbanización (URB) que se emparejan con cada par <región geográfica-sector>:

Tabla N° 6: Número de pares región geográfica-sector por región geográfica

Provincia	Área urbana	Centro urbano <i>expandido</i>
2,379	928	1650

Fuente: Elaboración propia

Posteriormente, a cada una de las 6,204 empresas participantes identificadas se las asoció con un coeficiente de LOC y URB dependiendo de su ubicación y el sector en el que desempeñan sus actividades. Es importante mencionar que las empresas pertenecientes al mismo sector y ubicadas en la misma región geográfica tendrán los mismos coeficientes de LOC y URB.

iv. Resultados

Con base en los datos obtenidos de la Encuesta Económica Anual de 2016 a nivel nacional, que considera una muestra de empresas por cada uno de los 14 sectores de interés, este estudio identificó, como parte de la primera etapa, la productividad total de factores (en logaritmos) para cada empresa mediante la metodología del residuo de Solow.

Sobre esta base, esta sección describe los resultados del estudio empírico sobre el impacto de las economías de aglomeración y urbanización sobre la productividad de las empresas. La Tabla N° 7 presenta los resultados a nivel de provincia; la Tabla N° 8, a nivel de área urbana; y la Tabla N° 9, a nivel de centro urbano *expandido*. En las 3 tablas, la columna (1) presenta los resultados de la estimación MCO sin controles. En particular, para todos los casos, el impacto de ambos índices es significativo y afecta negativamente la productividad.

A partir de la columna (2) hasta la columna (12) -para las Tablas N° 7 y N° 8- o la columna (8) -para la Tabla N° 9- se incluyen una serie de controles, como si la empresa se encuentra en un área urbana (AU) o un centro urbano *expandido* (CUE), la cantidad de años de experiencia, el tamaño de la empresa y el departamento donde se encuentre ubicada.

Cabe destacar que, para la Tabla N° 8, se eliminó la posibilidad de considerar el control AU debido a que, como los resultados consideran a todas las empresas que pertenecen de por sí a un área urbana, esta *dummy* se vuelve automáticamente 1 para todos los casos. Esta decisión también se aplica para la Tabla N° 9; en esta se elimina la posibilidad de considerar como controles AU y CUE porque todas las empresas que pertenecían a un centro urbano *expandido* pertenecían también a un área urbana, lo cual significa que ambas *dummy* toman el valor de 1 para todos los casos.

Además, es importante notar que se incluyen controles por departamento de tal manera que midan la ventaja comparativa del lugar; sin embargo, los resultados son, en su gran mayoría, no significativos para los casos a nivel de provincia y área urbana. También, para los resultados a nivel de provincia, se observa que, para las regresiones (1) a (7), los impactos hallados no difieren en gran medida entre sí. Es decir, a pesar de incluir controles, las elasticidades no cambian y se encuentran en el rango de -0.0186 y -0.0165 para las economías de localización y entre -0.0661 y -0.592 para las economías de urbanización.

Para el caso a nivel de centro urbano *expandido*, los resultados para todas las regresiones consideradas son significativos y negativos. En particular, para las economías de urbanización, el efecto varía en función de qué controles se consideren. Si se considera la experiencia, el efecto sería de -0.785; si, además, se considera el tamaño de la empresa, el efecto pertenece al rango de -0.695 a -0.689; y si, por último, se considera adicionalmente el departamento donde se encuentra ubicada la empresa, el efecto pertenece al rango de -0.530 a -0.498.

Finalmente, es posible afirmar que, en promedio, la elasticidad de las economías de aglomeración sobre la productividad de una empresa es de -0.78 si esta se encuentra ubicada en un área urbana. Mientras que, la elasticidad desciende a -0.81 si la empresa forma parte de un centro urbano *expandido*. En la misma línea, pasar de estar fuera de un AU a pertenecer a uno hace que el efecto sobre la productividad de la empresa pase de -0.68 a -0.78 (una disminución de 0.10); y, pasar de estar fuera de un CUE a pertenecer a uno hace que el efecto pase de -0.78 a -0.81 (una disminución de 0.03).³⁶

La razón por la cual la elasticidad de las economías de localización y urbanización no produce el efecto esperado puede ser atribuible al ciego desarrollo industrial o empresarial en el país. Por lo tanto, para obtener los resultados determinados por la teoría el gobierno -en todos los niveles geográficos- debería desarrollar un proceso de planificación y un plan de expansión de los sectores y mejorar las características de los niveles de aglomeración específicos de cada sector para proporcionar el efecto positivo completo sobre la productividad.

Otra causa aparente de que los resultados hallados no sean los propuestos por la teoría puede ser atribuible al alto nivel de informalidad tanto de empresas como de trabajadores a nivel nacional. De acuerdo con una investigación económica (Loayza, 2008), Perú se ubica como uno de los países con una tasa de informalidad de las más altas del mundo. Es importante mencionar que pertenecer al sector informal supone estar al margen de las cargas tributarias y normas legales.

Por otro lado, el sector formal no está exento de fricciones. Los supuestos básicos de los que parte el modelo neoclásico del mercado de trabajo son:

1. Salarios y precios flexibles

³⁶ Los resultados considerados provienen de la suma de los hallazgos en negrita de cada tabla. Por ejemplo, para los resultados a nivel de provincia, se realizó la siguiente suma: $-0.0166 + -0.661 = -0.6776 \approx -0.68$.

2. No existen costos para los trabajadores en la búsqueda de trabajo ni para los empleadores en el aumento o reducción de la planilla
3. La fuerza de trabajo es homogénea entre sí

Sin embargo, en la vida real, estos supuestos no se cumplen; es decir, existen fricciones laborales. En primer lugar, los salarios no son flexibles debido a la existencia de un salario mínimo, lo cual puede introducir distorsiones en el mercado de trabajo si el valor determinado excede la remuneración de equilibrio del mercado. En segundo lugar, es evidente que la búsqueda de empleo por parte de los trabajadores genera costos para el mismo en términos de tiempo o recursos. Para el caso de las firmas, estas están sujetas a unas *reglas del juego* (leyes o normas) establecidas que no les permiten despedir a su fuerza laboral por la existencia de contratos; o la necesidad de realizar gastos adicionales para la contratación de empleados. Se puede afirmar que el empleo formal cuesta hasta un 55% más que el informal.

Por último, y como se ha revisado anteriormente, existen otros autores como Yoo Ri Kim, Allan M. Williams, Sangwon Park y Jason Li Chen (2021) o Jiangjiang Kang, Wei Xu, Li Yu y Yueming Ning (2020) que han hallado resultados similares: elasticidades negativas y significativas.

Tabla N° 7: Los efectos de las economías de localización y urbanización sobre la productividad de las empresas

(resultados a nivel de **provincia**)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Localización	-0.0178** (0.00725)	-0.0166** (0.00744)	-0.0177** (0.00726)	-0.0165** (0.00745)	-0.0186** (0.00772)	-0.0178** (0.00785)	-0.0181** (0.00784)	-0.0116 (0.00729)	-0.0117 (0.0073)	-0.0117 (0.00728)	-0.0142* (0.00785)	-0.0143* (0.00781)
Urbanización	-0.655*** (0.0793)	-0.661*** (0.0792)	-0.653*** (0.0795)	-0.660*** (0.0794)	-0.592*** (0.0779)	-0.597*** (0.0779)	-0.603*** (0.0781)	-0.0703 (0.121)	-0.0519 (0.123)	-0.0463 (0.123)	-0.16 (0.122)	-0.149 (0.122)
AU		Sí		Sí		Sí	Sí		Sí	Sí	Sí	Sí
CUE		Sí		Sí		Sí	Sí		Sí	Sí	Sí	Sí
Experiencia			Sí	Sí			Sí			Sí		Sí
Tamaño					Sí	Sí	Sí				Sí	Sí
Departamento								Sí	Sí	Sí	Sí	Sí
Constant	0.834*** (0.0769)	0.772*** (0.0834)	0.826*** (0.0789)	0.767*** (0.0848)	0.631*** (0.0819)	0.593*** (0.0878)	0.615*** (0.089)	0.344*** (0.12)	0.347*** (0.119)	0.349*** (0.119)	0.270** (0.119)	0.280** (0.119)
Observations	6,204	6,204	6,204	6,204	6,204	6,204	6,204	6,204	6,204	6,204	6,204	6,204
R-squared	0.012	0.013	0.012	0.013	0.041	0.041	0.042	0.029	0.03	0.03	0.052	0.053

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Fuente: Elaboración propia

Nota: AU hare referencia a “área urbana” y CUE a “centro urbano *expandido*”.

Tabla N° 8: Los efectos de las economías de localización y urbanización sobre la productividad de las empresas
(resultados a nivel de **área urbana**)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Localización	-0.0246** (0.0103)	-0.0248** (0.0103)	-0.0246** (0.0103)	-0.0248** (0.0103)	-0.0263** (0.0102)	-0.0265*** (0.0103)	-0.0267*** (0.0103)	-0.00673 (0.00989)	-0.00674 (0.00994)	-0.00656 (0.00993)	-0.0135 (0.0103)	-0.0133 (0.0103)
Urbanización	-0.755*** (0.0912)	-0.756*** (0.0913)	-0.752*** (0.0919)	-0.752*** (0.092)	-0.655*** (0.09)	-0.656*** (0.0901)	-0.666*** (0.0905)	0.142 (0.182)	0.151 (0.182)	0.154 (0.182)	-0.0211 (0.179)	-0.0209 (0.178)
CUE		Sí		Sí		Sí	Sí		Sí	Sí	Sí	Sí
Experiencia			Sí	Sí			Sí			Sí		Sí
Tamaño					Sí	Sí	Sí				Sí	Sí
Departamento								Sí	Sí	Sí	Sí	Sí
Constant	0.939*** (0.0877)	0.935*** (0.0878)	0.931*** (0.0905)	0.927*** (0.0905)	0.687*** (0.0924)	0.681*** (0.0926)	0.709*** (0.0948)	0.222* (0.134)	0.207 (0.131)	0.211 (0.131)	0.228 (0.14)	0.247* (0.139)
Observations	5,799	5,799	5,799	5,799	5,799	5,799	5,799	5,799	5,799	5,799	5,799	5,799
R-squared	0.013	0.013	0.013	0.013	0.041	0.041	0.042	0.031	0.032	0.032	0.053	0.054

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Fuente: Elaboración propia

Nota: CUE hace referencia a “centro urbano *expandido*”.

Tabla N° 9: Los efectos de las economías de localización y urbanización sobre la productividad de las empresas
(resultados a nivel de **centro urbano expandido**)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Localización	-0.0210** (0.0085)	-0.0210** (0.0085)	-0.0205** (0.00857)	-0.0204** (0.00859)	-0.0252*** (0.00772)	-0.0252*** (0.0077)	-0.0236*** (0.00801)	-0.0235*** (0.00801)
Urbanización	-0.786*** (0.0683)	-0.785*** (0.0689)	-0.689*** (0.0672)	-0.695*** (0.0676)	-0.530*** (0.107)	-0.529*** (0.107)	-0.498*** (0.103)	-0.495*** (0.103)
Experiencia		Sí		Sí		Sí		Sí
Tamaño			Sí	Sí			Sí	Sí
Departamento					Sí	Sí	Sí	Sí
Constant	0.896*** (0.0614)	0.893*** (0.0657)	0.687*** (0.0707)	0.715*** (0.0734)	0.689*** (0.0512)	0.693*** (0.0515)	0.720*** (0.0701)	0.735*** (0.0703)
Observations	3,258	3,258	3,258	3,258	3,258	3,258	3,258	3,258
R-squared	0.038	0.038	0.065	0.066	0.056	0.056	0.079	0.08

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Fuente: Elaboración propia

5. CONCLUSIONES

El presente documento de investigación fue realizado con el objetivo de estimar el efecto de las economías de aglomeración sobre la productividad de las empresas del país.

Las economías de aglomeración pueden dividirse en economías de localización y de urbanización. La primera de estas está asociada con los beneficios derivados de la concentración del mismo sector, mientras que la segunda está vinculada con las ventajas provenientes de la diversidad en cuanto a sectores económicos en la ciudad donde la empresa se encuentre ubicada.

La estrategia empírica consistió en dos etapas: en la primera se calculó la productividad de cada empresa mediante el residuo de Solow, partiendo de una función de producción Cobb-Douglas personalizada para cada uno de los 14 sectores económicos considerados; en la segunda etapa, se buscó identificar la elasticidad de las economías de localización y urbanización sobre la productividad hallada. Para lograrlo, se utilizó uso de información proveniente de la Encuesta Económica Anual de 2016 (para la realización de la primera etapa) y de SUNAT mediante dos solicitudes de acceso a la información pública (para la creación de los índices de localización y urbanización).

Estos índices se crearon para cada región geográfica considerada: provincia, área urbana y centro urbano *expandido*. Las áreas urbanas se definieron en función de las densidades poblacionales obtenidas a partir de Meta, siguiendo los criterios de umbral mínimo en la densidad poblacional extendida de 500 personas por kilómetro cuadrado, el umbral de población mínima de 10,000 personas por área urbana y el criterio de continuidad. Los centros urbanos *expandidos* se definieron alrededor de los puntos donde se acumula una mayor densidad poblacional extendida, siguiendo la disminución de la gradiente hasta el punto de cambio de signo que indicaba la acción del siguiente centro. Una vez finalizado este proceso, se volvió a iniciar sobre las regiones sobrantes de forma repetida hasta llegar a una población restante del 10% de la población total.

En cuanto a los hallazgos, se estimó que la contribución de las economías de aglomeración a la productividad en Perú es de -0.68 para todas las empresas; si, además, la empresa de interés pertenece a un AU, el efecto es de -0.78; y si, esta también forma parte de un CUE, la elasticidad es de -0.81.

Las razones detrás de este hallazgo pueden ser atribuirse al ciego desarrollo industrial o empresarial, al alto nivel de informalidad en el país o a las fricciones existentes en el mercado laboral.

6. BIBLIOGRAFÍA

- Akerberg, D. A., Caves, K., & Frazer, G. (Noviembre de 2015). Identification properties of recent production function estimators. *Econometrica*, 83(6), 2411-2451.
- Alonso, W. (1964). *Location and Land Use*. Cambridge: Harvard University Press.
- Bird, Julia, Lebrand, Mathilde y & Venables, Anthony J. (2020). The Belt and Road Initiative: Reshaping economic geography in Central Asia?. *Journal of Development Economics*, 144, 102441.
- Bosker, M., Park, J., & Roberts, M. (2019). Definition matters. metropolitan areas and agglomeration economies in a large developing country. World Bank Group.
- Ch, R., Martin, D., & Vargas, J. (2014). Measuring the size and growth of cities using nighttime light. Development Bank of Lantin America.
- Clark, C. (1951). Urban population densities. *J. Roy. Satist. Soc. Ser. A*, 114, 490-496. Dijkstra, L., & Poelman, H. (2014). A harmonised definition of cities and rural areas: the new degree of urbanisation. Washington DC, USA: European Commission Directorate - General for Regional and Urban Policy.
- Combes, P.-P. (Mayo de 2000). Economic Structure and Local Growth: France, 1984–1993. *Journal of Urban Economics*, 47(3), 329-355.
- Desmet, K., & Fafchamps, M. (Junio de 2005). Changes in the spatial concentration of employment across US counties: a sectoral analysis 1972–2000. 5(3), 261–284.
- Duranton, G. (2015). A proposal to delineate metropolitan areas in colombia. *Desarrollo y Sociedad*, 75, 223-264.
- Duranton, G., & Puga, D. (2004). Micro-foundations of urban agglomeration economies. En G. Duranton, J. Henderson, & W. Strange, *Handbook of Regional and Urban Economics* (Vol. 4, págs. 2063-2117).
- Ethier, W. J. (Junio de 1982). National and international returns to scale in the modern theory of international trade. *American Economic Review*, 72(3), 389–405.
- Feng Han, Rui Xie, & Jiayu Fang (2018). Urban agglomeration economies and industrial energy efficiency. *Energy*, 162, 45-59.
- Ferragina, Anna M., & Mazzotta, Fernanda (2014). Local agglomeration economies: what impact on multinational and national Italian firms' survival?. *Procedia - Social and Behavioral Sciences*, 100, 8-19.
- Frankena, M. W. (1978). A bias in estimating urban population density functions. *Journal of Urban Economics*, 5, 35-45.
- Frenken, K., van Oort, F., Verburg, T., & Boschma, R. A. (Diciembre de 2004). Variety and regional economic growth in the Netherlands. *Papers in Evolutionary Economic Geography*.
- Fujita, M. (1986). *Urban land use theory. in location theory*. London: Harwood Academic Publishers.
- Fujita, M. (1989). *Urban economic theory: Land use and city size*. Cambridge: Cambridge University Press.
- Fujita, M., & Ogawa, H. (1982). Multiple equilibria and structural transition of nonmonocentric urban configurations. *Regional Science and Urban Economics*, 12.

- Glaeser, E. L., D. Kallal, H. Scheinkman, J. A., & Shleifer, A. (Diciembre de 1992). Growth in Cities. *Journal of Political Economy*, 100(6), 1126-1152.
- Greene, D. L., & Barnbrock, J. (1978). A note on problems in estimating exponential urban density models. *Journal of Urban Economics*, 5, 285-290.
- Griffith, D., & Wong, D. (2006). Modeling population density across major us cities a polycentric spatial regression approach. *Journal of Geographical Systems*, 9, 53-75.
- Grotewold, A. (Octubre de 1959). Von Thunen in Retrospect. *Economic Geography*, 35(4), 346-355.
- Harrison, B., Kelley, M. R., & Gant, J. (Mayo de 1996). Specialization Versus Diversity in Local Economies: The Implications for Innovative Private-Sector Behavior. *Cityscape: A Journal of Policy Development and Research*, 2(2), 61-93.
- Helsley, R. W., & Strange, W. C. (Septiembre de 1990). Matching and Agglomeration Economies in a System of Cities. *Regional Science and Urban Economics*, 20(2), 189-212.
- Henderson, D. N., J. Vernon, & Kriticos, S. (2019). Learning by working in big cities. *Review of Economic Studies*, 103188.
- Imhoff, M. L., Lawrence, W., Stutzer, D. C., & Elvidge, C. (1997). A technique for using composite dmsp/ols city lights satellite data to map urban area. *Remote Sensing of Environment*, 61, 361-370.
- Jacobs, J. (1969). *The economy of cities*. New York: Random House.
- Jiangjiang Kang, Wei Xu, Li Yu, & Yueming Ning (2020). Localization, urbanization and globalization: Dynamic manufacturing specialization in the YRD mega-city conglomeration. *Cities*, 99, 102641.
- Kelley, M. R., & Helper, S. (1999). Firm Size And Capabilities, Regional Agglomeration, And The Adoption Of New Technology. *Economics of Innovation and New Technology*, 8(1-2), 79-103.
- Krugman, P. (1991). *Geography and Trade*. Cambridge: The MIT Press.
- Launhardt, W. (1882). Die Bestimmung des Zweckmässigsten Standortes einer Gewerblichen Anlage. *Zeitschrift des Vereines Deutscher Ingenieure*, 26, 106–115.
- Le Blanc, G. (2000). Regional specialization, local externalities and clustering in information technology industries. 40th Congress of the European Regional Science Association: "European Monetary Union and Regional Policy".
- Levinsohn, J., & Petrin, A. (2003). Estimating Production Functions Using Inputs to Control for Unobservables. *Review of Economic Studies*, 70(2), 317-341.
- Loayza, N. (2008). Causas y consecuencias de la informalidad en el Perú. *Revista de Estudios Económicos*(15), 43 - 64.
- Lucas, R. E. (Julio de 1988). On the mechanics of economic development. *Journal of Monetary Economics*, 22(1), 3-42.
- Marshall, A. (1980). *Principles of Economics*.
- Martin, P., Mayer, T., & Mayneris, F. (2011). Spatial concentration and plant-level productivity in France. *Journal of Urban Economics*, 69(2), 182-195.
- McDonald, J. F., & Bowman, H. W. (1976). Some tests of alternative urban population density functions. *Journal of Urban Economics*, 3, 242-252.

- Mills, E. (1970). Urban density functions. *Urban Studies*, 7, 5-20.
- Mori, T., & Smith, T. (2014). A probabilistic modeling approach to the detection of industrial agglomerations. *Journal of Economic Geography*, 14, 547–588.
- Zidong Yu, & Xintao Liu (2021). Urban agglomeration economies and their relationships to built environment and socio-demographic characteristics in Hong Kong. *Habitat International*, 117, 102417.
- Newling, B. (1966). Urban growth and spatial structure: Mathematical models and empirical evidence. *Geograph Rev.*, 56, 213-225.
- Noticias ONU. (16 de Mayo de 2018). Naciones Unidas. Recuperado el Marzo de 2022, de Departamentode Asuntos Económicos y Sociales: <https://www.un.org/development/desa/es/news/population/2018-world-urbanization-prospects.html>
- Olley, G. S., & Pakes, A. (Noviembre de 1996). The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica*, 64(6), 1263-1297.
- Park, J.-S., & Oh, S.-J. (2012). A new concave hull algorithm and concaveness measure for n-dimensional datasets. *Regional of Information Science and Engineering*, 28, 587-600.
- Qi Guo, Canfei He, & Deyu Li (2016). Entrepreneurship in China: The role of localisation and urbanisation economies. *Urban Studies*, 53, 2584-2606.
- Saxenian, A. (1994). Regional Advantage: Culture and Competition in Silicon Valley and Route 128. *The Business History Review*, 68(3), 429-431.
- Small, C., Pozzi, F., & Elvidge, C. D. (2005). Spatial analysis of global urban extent from DMSP-OLS night lights. *Remote Sensing of Environment*, 96, 277-291.
- Straszheim, M. R. (1974). Econometric issues in interpreting mills estimates of urban density gradients. *Journal of Urban Economics*, 1, 445-448.
- Sutton, P. C. (2003). A scale-adjusted measure of urban sprawl using nighttime satellite imagery. *Remote Sensing of Environment*, 86, 353-369.
- Toshitaka Gokan, Ikuo Kuroiwa, & Kentaro Nakajima (2019). Agglomeration economies in Vietnam: A firm-level analysis. *Journal of Asian Economics*, 62, 52-64.
- Uchida, H., & Nelson, A. (2009). Agglomeration index: Towards a new measure of urban concentration. Washington DC, USA: World Bank.
- van Oort, F., & Stam, E. (29 de Marzo de 2006). Agglomeration Economies and Entrepreneurship in the ICT Industry. (Report Series ERS-2006- 016-ORG).
- von Thünen, J. (1826). *The Isolated State*. Hamburg.
- Weber, A. (1909). *Theory of the Location of Industries*. Chicago: University of Chicago Press.
- White, L. J. (1977). How good are two-point estimates of urban density gradients and central densities. *Journal of urban Economics*, 4, 292-309.
- Yin Huang, Tao Hong, & Tao Ma (2020). Urban network externalities, agglomeration economies and urban economic growth. *Cities*, 107, 102882.
- Yoo Ri Kim, & Williams, Allan (2021). Spatial spillovers of agglomeration economies and productivity in the tourism industry: The case of the UK. *Tourism Management*, 82, 104201.

Zidong Yu, & Xintao Liu (2021). Urban agglomeration economies and their relationships to built environment and socio-demographic characteristics in Hong Kong. *Habitat International*, 117, 102417.

7. ANEXOS

Anexo N° 1: Definiciones de índices de localización y urbanización

Autores	Título de la investigación	Economías de localización	Economías de urbanización
<p>Anna Maria Ferragina, Fernanda Mazzotta (2014)</p>	<p>Local agglomeration economies: what impact on multinational and national Italian firms' survival?</p>	<p>Índice de especialización de Balassa:</p> $IS_{i,s} = \frac{L_{i,s}}{\frac{\sum_s L_{i,s}}{\sum_{is} L_{i,s}}}$	<p>1. Densidad de población</p> <p>2. Grado de distritalización de las provincias medido por el número de trabajadores en los sistemas del mercado laboral local.</p> <p>-</p> <p>Externalidades de Jacob</p> <p>1. Variedad relacionada</p> <p>2. Variedad intraindustrial</p> <p>3. Diversificación o variedad no relacionada</p>
<p>Qi Guo, Canfei He, Deyu Li (2016)</p>	<p>Entrepreneurship in China: The role of localisation and urbanisation economies</p>	$LOC = \frac{e_i/e}{E_i/E}$ <p>donde e_i es el empleo local en el sector i; e es el empleo manufacturero local total; E_i es el empleo nacional en el sector i y E es el empleo manufacturero nacional total.</p> <p>Descomponemos e_i en empleos de empresas: pequeñas, medianas y grandes.</p>	<p>-</p>

<p>Feng Han, Rui Xie, Jiayu Fang (2018)</p>	<p>Urban agglomeration economies and industrial energy efficiency</p>	<p><i>Agglomeration de especialización</i> (adoptado de Combes):</p> $S_i = \frac{E_{i,m}/E_i}{E_m/E} \quad S_{i,s} = \frac{E_{i,s}/E_i}{E_s/E}$ <p>donde S_i es el nivel general de especialización de la aglomeración en la ciudad i; mientras que, $S_{i,s}$ es la aglomeración de especialización de la subdivisión s de la industria de manufactura en la ciudad i. $E_{i,m}$ es el empleo total de la industria manufacturera en la ciudad i; E_i es la cantidad total de empleo en la ciudad i; E_m es la cantidad total de empleo en la manufactura nacional; E es la cantidad total de empleo nacional; $E_{i,s}$ representa la cantidad de empleo del sector manufacturero s en la ciudad i; y, por último, E_s es el empleo del sector s en el país.</p>	<p><i>Agglomeration de diversificación</i> (adaptado de Combes):</p> $D_i = \sum_s \frac{E_{is}}{E_i} \left[\frac{\frac{1}{\sum_{s'=1, s' \neq s}^n \left(\frac{E_{is'}}{(E_i - E_{is'})} \right)^2}}{\frac{1}{\sum_{s'=1, s' \neq s}^n \left(\frac{E_{s'}}{(E - E_s)} \right)^2}} \right] D_{i,s}$ $= \frac{1 / \sum_{s'=1, s' \neq s}^n (E_{is'} / (E_i - E_{is'}))^2}{1 / \sum_{s'=1, s' \neq s}^n (E_{s'} / (E - E_s))^2}$ <p>donde D_i es la aglomeración de diversificación de la manufactura en la ciudad i; mientras que, $D_{i,s}$ es la aglomeración de diversificación de la industria manufacturera de subdivisión s en la ciudad i. Cuanto mayor es el valor, mayor es el nivel de aglomeración de diversificación.</p> <p>Por otro lado, $E_{is'}$ representa la cantidad de empleo del sector s' excepto el sector s en la ciudad i; y, por último, $E_{s'}$ es el empleo del sector s' excepto el sector s en el país.</p>
<p>Toshitaka Gokan, Ikuo Kuroiwa, Kentaro Nakajima (2019)</p>	<p>Agglomeration economies in Vietnam: A firm-level analysis</p>	$AGG_{e(i)r}^t = \sum_j d_{jr} w_{ij}^t$ <p>donde d_{jr} es una <i>dummy</i> de conglomerado que equivale a 1 si el distrito r se detecta como un conglomerado de la industria j y w_{ij}^t es la fuerza de las relaciones interindustriales de tipo t entre las industrias i y j.</p>	<p>-</p>

<p>Yin Huang, Tao Hong, Tao Ma (2020)</p>	<p>Urban network externalities, agglomeration economies and urban economic growth</p>	<p><i>Agglomeration de especialización</i> (Beaudry and Schiffauerova, 2009):</p> $LQ_{ij} = \frac{E_{ij}/E_i}{E_j/E}$ <p>Nivel general de especialización de la aglomeración en la ciudad i:</p> $S_i = \max(S_{i,s})$	<p><i>Agglomeration de diversificación</i> (Jacobs' externalities):</p> $diversity_i = 1 / \left(\sum_s (E_{i,s}/E_i)^2 \right)$
<p>Jiangjiang Kang, Wei Xu, Li Yu, Yueming Ning (2020)</p>	<p>Localization, urbanization and globalization: Dynamic manufacturing specialization in the YRD mega-city conglomeration</p>	<p>Emplean cuatro variables:</p> <ol style="list-style-type: none"> 1. Tamaño de mercado (Ricardo, 1891) - ventas minoristas totales de bienes de consumo 2. Salario laboral promedio: para medir el efecto del costo de la mano de obra local en la especialización manufacturera (Traistaru, Nijkamp y Longhi, 2002) - salario promedio de los empleados en el trabajo 3. Precio de la tierra: refleja el costo de la tierra que influye en las actividades manufactureras (Malmberg & Maskell, 1997) - el precio promedio de la tierra sujeta a licitación, subasta o listado para la venta 4. Zonas de desarrollo: para reflejar el papel de los gobiernos locales en el desarrollo industrial (Storper, 2010) - número de zonas de desarrollo en un área determinado 	<p>Emplean tres variables (Su, 2018):</p> <ol style="list-style-type: none"> 1. Tamaño de la ciudad: se calculan a partir de la población empadronada de una ciudad - la población local total 2. Tasa de urbanización: se calculan a partir de la población empadronada de una ciudad 3. Tasa de crecimiento de la migración: se estima utilizando datos del censo de población. <p>También emplean dos <i>dummy</i> para investigar el efecto de los tipos urbanos diferenciados en la manufactura:</p> <ol style="list-style-type: none"> 3.1 Áreas urbanas primarias - <i>dummy</i> = 1 si un condado se encuentra en áreas primarias del núcleo urbano, o = 0 3.2 Áreas urbanas secundarias - <i>dummy</i> = 1 si un condado se encuentra en áreas secundarias del núcleo urbano, o = 0 3.3 Áreas urbanas periféricas - <i>dummy</i> = 1 si un condado se encuentra en áreas periféricas del núcleo urbano, o = 0

<p>Julia Bird, Mathilde Lebrand, Anthony J. Venables (2020)</p>	<p>The Belt and Road Initiative: Reshaping economic geography in Central Asia?</p>	$z_{2i} = \bar{z}_{2i} (1 + \alpha(n_{2i}/\bar{L}_i))^p$ <p>donde la productividad base z_{2i} se calibra en los datos. Por el lado derecho, \bar{z}_{2i} se mantiene constante en la simulación y se adiciona un componente que representa rendimientos crecientes en función de la densidad de empresas en la celda del sector; es decir, el número de empresas dividido por la población inicial (n_{2i}/\bar{L}_i).</p>	<p>-</p>
<p>Yoo Ri Kim, Allan M. Williams, Sangwon Park, Jason Li Chen (2021)</p>	<p>Spatial spillovers of agglomeration economies and productivity in the tourism industry: The case of the UK</p>	<p>Ratio entre el empleo turístico y el empleo total en una unidad espacial frente a la relación entre el empleo turístico total y el empleo total</p>	<p>-</p>
<p>Zidong Yu, Xintao Liu (2021)</p>	<p>Urban agglomeration economies and their relationships to built environment and socio-demographic characteristics in Hong Kong</p>	$LQ_{ij} = \frac{e_{ij}/e_i}{E_j/E}$ <p>donde e_{ij} y e_i denotan el empleo del sector j y el empleo total para i, respectivamente. Mientras que E_j y E son el empleo total de toda el área de estudio en el sector j y en todos los sectores.</p>	<p>-</p>

Fuente: Elaboración propia

Anexo N° 2: Formatos por sector económico

Tabla N° 10: Formatos por sector económico

Código del sector	Sector económico	Formato
1	Agencias de viaje	Único
2	Agroindustria	Grande Mediana o pequeña
3	Centros educativos no estatales	Único
4	Comercio y servicios	Grande Mediana o pequeña
5	Construcción	Grande Mediana o pequeña
6	Establecimientos de hospedaje	Único
7	Hidrocarburos	Único
8	Pesca	Grande Mediana o pequeña
9	Manufactura	Grande Mediana o pequeña
10	Servicios eléctricos	Único
11	Transportes y comunicaciones	Grande Mediana o pequeña
12	Universidades no estatales	Único
13	Restaurantes	Único
14	Servicios	Grande Mediana o pequeña

Fuente: Elaboración propia con base en la disposición técnica para la interpretación de datos de la Encuesta Económica Anual del INEI.

Anexo N° 3: Estadísticas descriptivas de las variables de interés por sector

Tabla N° 11: Estadísticas descriptivas de las variables de interés por sector

Sector: Agencias de viaje

Variable	Obs	Mean	Std. dev.	Min	Max
Producción total	106	10.45	14.15	0.23	80.07
Capital	106	0.87	1.92	0.00	17.82
Mano de obra	106	60.56	109.30	1.00	877.00
Servicios de terceros	106	6.54	9.95	0.02	62.62

Sector: Agroindustria

Variable	Obs	Mean	Std. dev.	Min	Max
Producción total	218	208.86	728.60	0.05	6,229.33
Capital	218	55.65	209.20	0.02	2,509.44
Mano de obra	218	695.58	1,588.87	1.00	13,211.00
Tierra	218	63.91	172.33	0.00	1,264.01
Insumos	218	86.52	373.93	0.00	4,775.87
Servicios de terceros	218	28.04	103.51	0.01	1,126.81

Sector: Centros educativos no estatales

Variable	Obs	Mean	Std. dev.	Min	Max
Producción total	157	18.47	33.81	0.32	319.07
Capital	157	5.53	18.35	0.01	202.01
Mano de obra	157	257.19	450.39	6.00	4,160.00
Tierra	157	19.84	48.63	0.00	309.32
Servicios de terceros	157	4.28	7.21	0.00	52.55

Sector: Comercio

Variable	Obs	Mean	Std. dev.	Min	Max
Producción total	3,033	13.98	51.49	0.00	933.41
Capital	3,033	3.80	34.01	0.00	1,452.71
Mano de obra	3,033	107.52	595.08	1.00	16,678.00
Servicios de terceros	3,033	4.93	20.92	0.00	400.63

Sector: Construcción

Variable	Obs	Mean	Std. dev.	Min	Max
Producción total	658	41.50	141.00	0.01	1,843.74
Capital	658	7.43	25.44	0.00	288.13
Mano de obra	658	184.52	476.53	1.00	4,965.00
Insumos	658	8.07	22.77	0.00	257.39

Servicios de terceros	658	16.54	90.01	0.00	1,851.61
-----------------------	-----	-------	-------	------	----------

Sector: Establecimientos de hospedaje

Variable	Obs	Mean	Std. dev.	Min	Max
Producción total	87	17.93	32.55	0.62	219.49
Capital	87	7.98	17.77	0.04	123.49
Mano de obra	87	161.64	264.71	3.00	1,539.00
Tierra	87	41.91	113.24	0.00	906.00
Insumos	87	2.22	4.24	0.00	26.92
Servicios de terceros	87	5.86	10.48	0.02	55.62

Sector: Hidrocarburos

Variable	Obs	Mean	Std. dev.	Min	Max
Producción total	47	651.96	1,630.04	2.32	8,342.86
Capital	47	742.56	2,351.45	0.03	12,895.46
Mano de obra	47	285.30	661.90	1.00	4,282.00
Tierra	47	170.97	461.75	0.01	2,189.19
Insumos	47	322.01	1,213.26	0.04	7,166.41
Servicios de terceros	47	103.18	211.45	0.01	1,058.62

Sector: Pesca

Variable	Obs	Mean	Std. dev.	Min	Max
Producción total	183	51.23	155.11	0.43	1,225.47
Capital	183	57.48	243.90	0.00	2,293.64
Mano de obra	183	200.16	472.17	1.00	4,608.00
Insumos	183	18.30	49.06	0.01	330.04
Servicios de terceros	183	9.23	27.34	0.00	237.56

Sector: Manufactura

Variable	Obs	Mean	Std. dev.	Min	Max
Producción total	1,169	65.66	217.28	0.17	3,840.88
Capital	1,169	26.04	124.99	0.00	2,205.45
Mano de obra	1,169	203.28	374.76	1.00	4,624.00
Tierra	1,169	18.73	68.32	0.00	1,185.48
Insumos	1,169	27.14	89.32	0.00	1,545.98
Servicios de terceros	1,169	11.84	44.12	0.00	784.53

Sector: Servicios eléctricos

Variable	Obs	Mean	Std. dev.	Min	Max
Producción total	38	500.67	756.97	1.49	2,937.23
Capital	38	659.38	977.91	0.37	3,839.93
Mano de obra	38	217.05	276.11	1.00	1,331.00
Tierra	38	382.65	818.44	0.21	3,863.42
Intangibles	38	24.84	75.94	0.00	402.09
Insumos	38	202.95	420.58	0.04	1,867.31
Servicios de terceros	38	33.79	39.14	0.22	165.21

Sector: Transporte y comunicaciones

Variable	Obs	Mean	Std. dev.	Min	Max
Producción total	855	52.13	380.31	0.00	9,449.31
Capital	855	51.48	774.35	0.00	21,761.32
Mano de obra	855	478.80	8,946.82	1.00	261,274.00
Servicios de terceros	855	24.86	188.51	0.00	4,071.24
Producción total	855	52.13	380.31	0.00	9,449.31

Sector: Universidades no estatales

Variable	Obs	Mean	Std. dev.	Min	Max
Producción total	60	114.89	168.79	1.04	750.50
Capital	60	31.86	48.06	0.11	237.40
Mano de obra	60	1,905.95	2,853.91	60.00	16,251.00
Tierra	60	173.81	426.50	0.17	2,902.05
Intangibles	60	2.45	6.34	0.00	42.59
Servicios de terceros	60	24.88	38.57	0.33	176.92

Sector: Restaurantes

Variable	Obs	Mean	Std. dev.	Min	Max
Producción total	155	19.72	47.27	0.19	471.60
Capital	155	3.09	8.98	0.01	70.06
Mano de obra	155	270.10	640.88	2.00	5,214.00
Tierra	155	4.32	11.83	0.00	100.14
Insumos	155	7.93	19.92	0.02	181.76
Servicios de terceros	155	4.72	11.57	0.00	90.56

Sector: Servicios

Variable	Obs	Mean	Std. dev.	Min	Max
Producción total	2,186	18.72	60.71	0.00	1,857.32
Capital	2,186	6.27	41.14	0.00	1,413.34
Mano de obra	2,186	201.18	728.90	1.00	15,000.00
Servicios de terceros	2,186	5.80	26.90	0.00	954.97

Fuente: Elaboración propia

Nota: La producción total y los factores de producción (salvo la mano de obra) están medidos en millones de soles y la mano de obra en cantidad de personas.

Anexo N° 4: Resultados de la estimación por sector

Tabla N° 12: Resultados de la estimación por sector

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Capital	0.0782* (0.0422)	-0.113 (0.0955)	0.167*** (0.0402)	0.136*** (0.015)	-0.0415 (0.0427)	0.221*** (0.0481)	0.152* (0.0884)	-0.00837 (0.0251)	0.0462** (0.0215)	-0.0235 (0.0859)	0.162*** (0.0237)	0.337*** (0.0457)	0.0944*** (0.0352)	0.0476* (0.0268)
Mano de obra	0.293*** (0.032)	0.451*** (0.0462)	0.331*** (0.00795)	0.355*** (0.0126)	0.444*** (0.0166)	0.348*** (0.0161)	0.414*** (0.086)	0.499*** (0.0272)	0.454*** (0.0151)	0.574*** (0.121)	0.386*** (0.0159)	0.276*** (0.0205)	0.325*** (0.0177)	0.357*** (0.00945)
Tierra		0.0474 (0.0471)	0.125*** (0.0259)			0.0291 (0.0314)	0.0454 (0.0723)		0.0226 (0.0148)	0.261*** (0.0648)		0.0461 (0.04)	-0.00946 (0.0216)	
Intangibles										0.0185 (0.0473)		0.0302* (0.0152)		
Insumos		0.301*** (0.103)			0.185*** (0.0441)	0.0609* (0.0349)	0.387*** (0.0801)	0.314*** (0.0728)	0.217*** (0.0493)	0.0269 (0.0668)			0.428*** (0.0678)	
Servicios de terceros	0.566*** (0.0638)	0.368*** (0.119)	0.309*** (0.0566)	0.426*** (0.0167)	0.257*** (0.0342)	0.299*** (0.0378)	0.293** (0.118)	0.172*** (0.0432)	0.290*** (0.0331)	0.483*** (0.121)	0.308*** (0.0309)	0.400*** (0.0556)	0.0926*** (0.0252)	0.445*** (0.0231)
Observations	106	218	157	3,028	657	87	46	183	1,169	37	853	60	155	2,177
R-squared	0.900	0.941	0.93	0.836	0.747	0.948	0.97	0.926	0.943	0.987	0.674	0.996	0.956	0.751

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

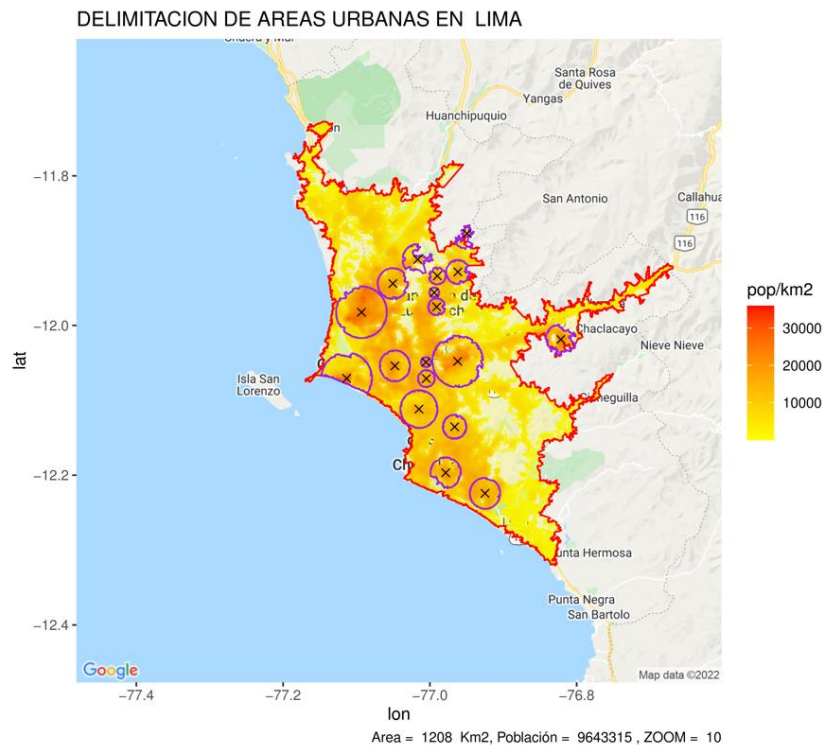
Fuente: Elaboración propia

Nota: Los sectores de cada regresión se pueden hallar codificados en el Anexo N° 2. Asimismo, las variables han sido log-linealizadas de su versión en niveles.

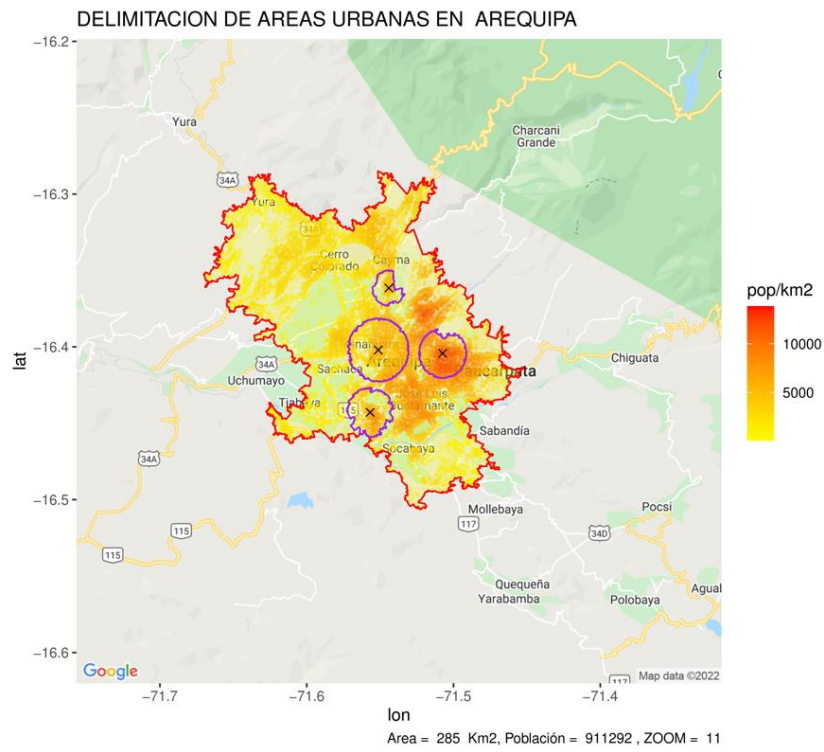
Anexo N° 5: Delimitación de áreas urbanas, centros urbanos y centros urbanos *expandidos*

Tabla N° 13: Delimitación de áreas urbanas, centros urbanos y centros urbanos *expandidos*

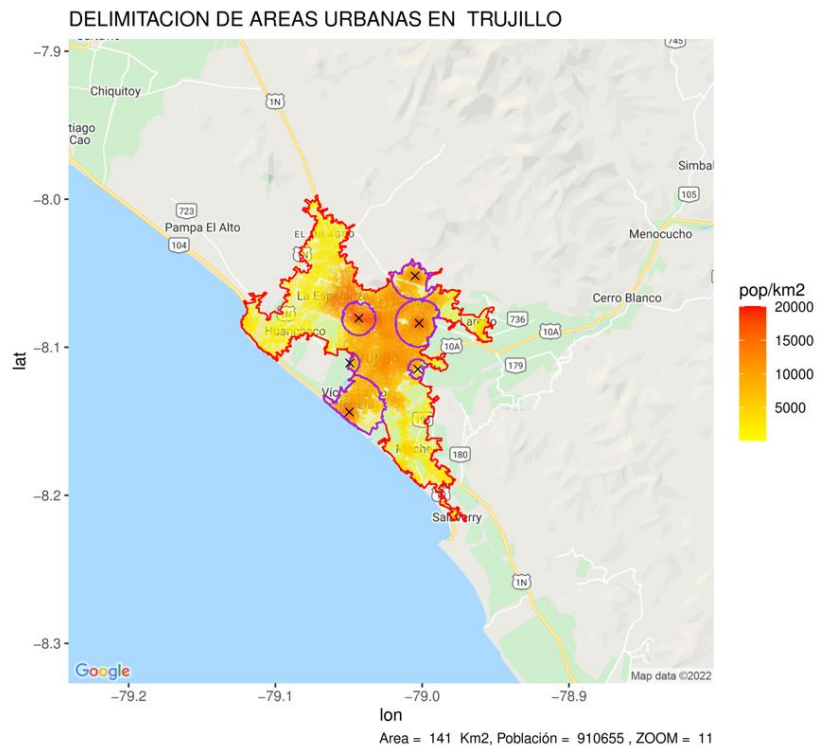
1. Lima



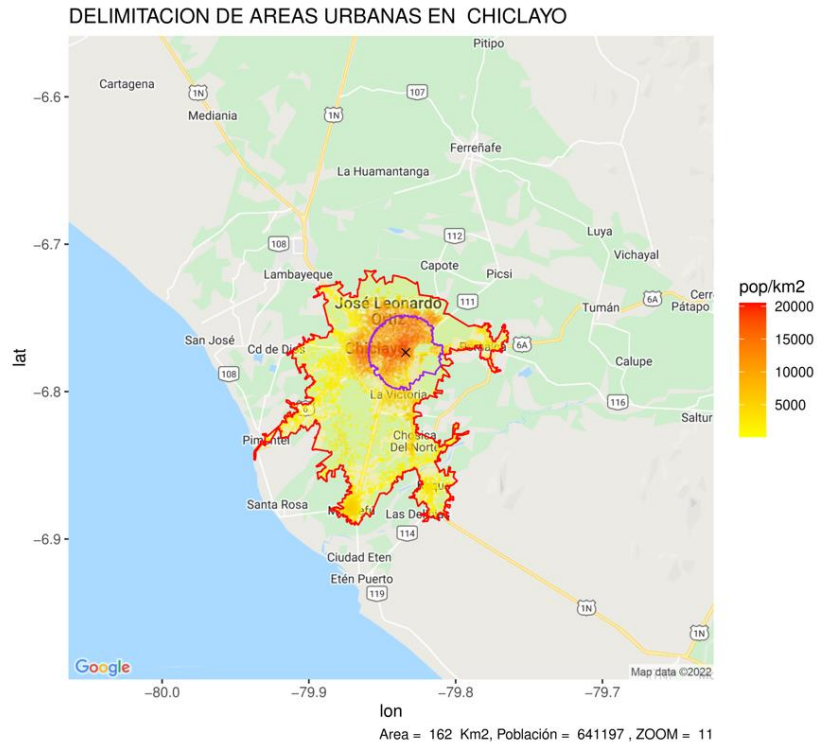
2. Arequipa



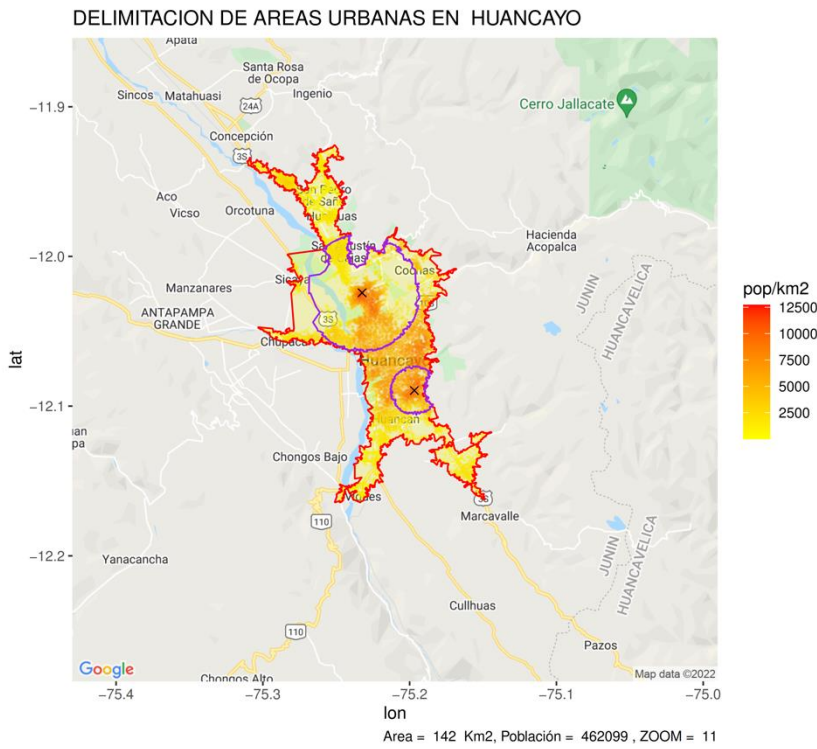
3. Trujillo



4. Chiclayo

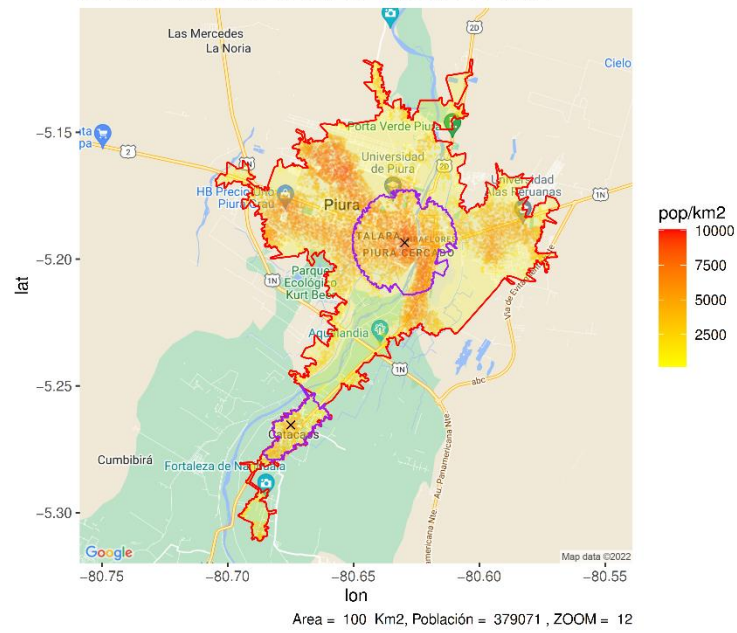


5. Huancayo



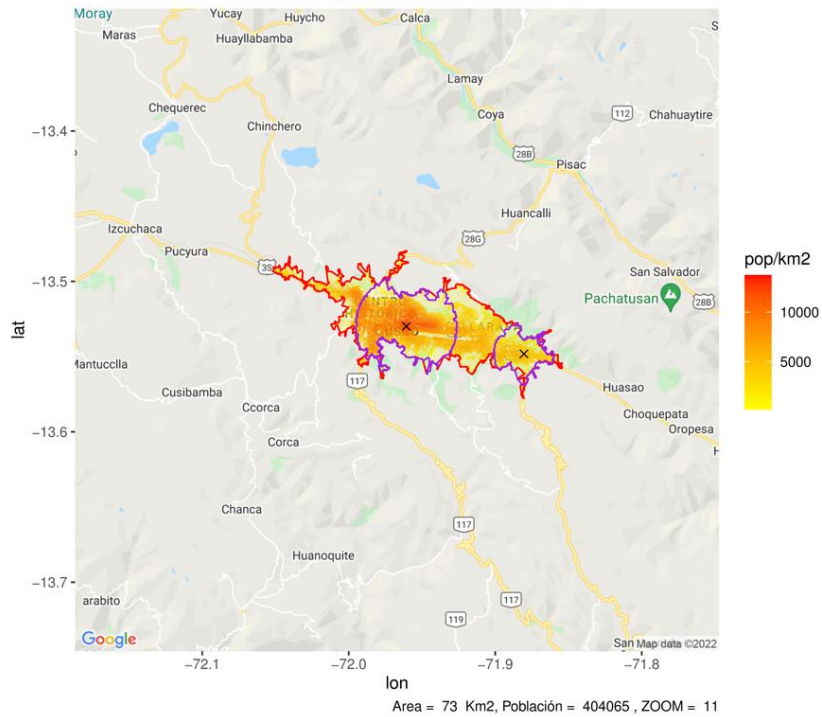
6. Piura

DELIMITACION DE AREAS URBANAS EN PIURA

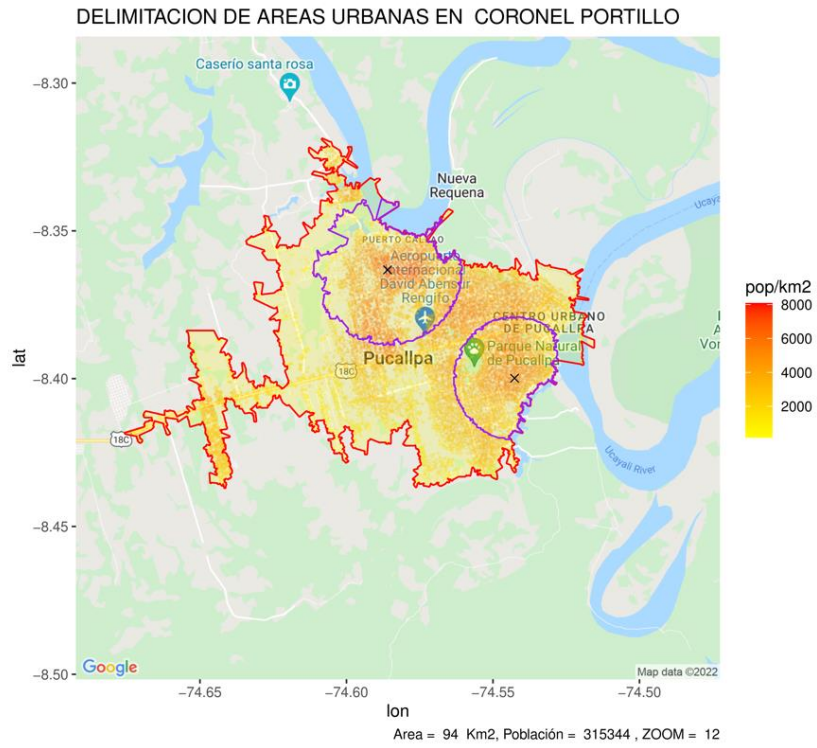


7. Cusco

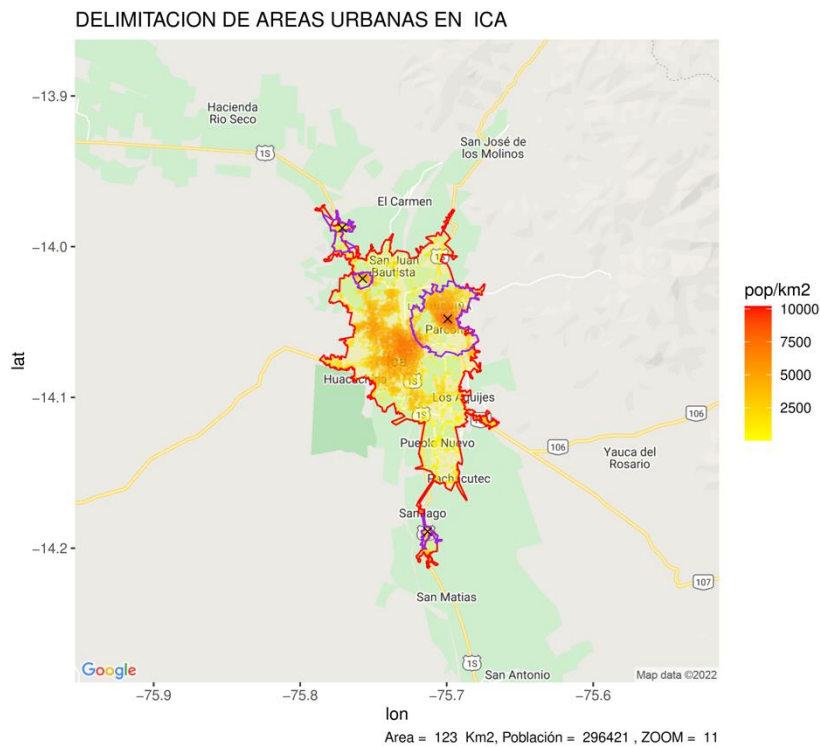
DELIMITACION DE AREAS URBANAS EN CUSCO



8. Coronel Portillo

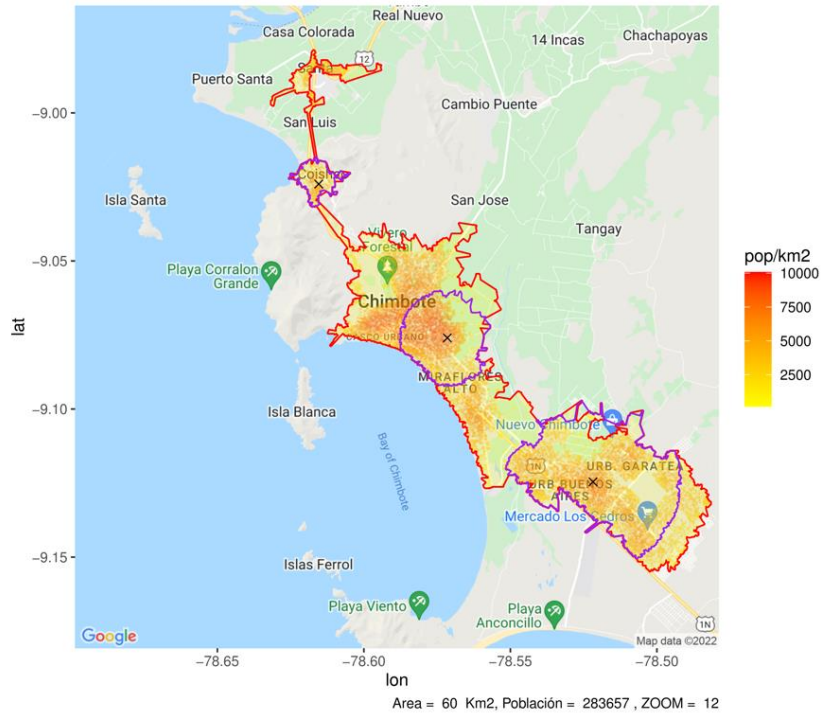


9. Ica



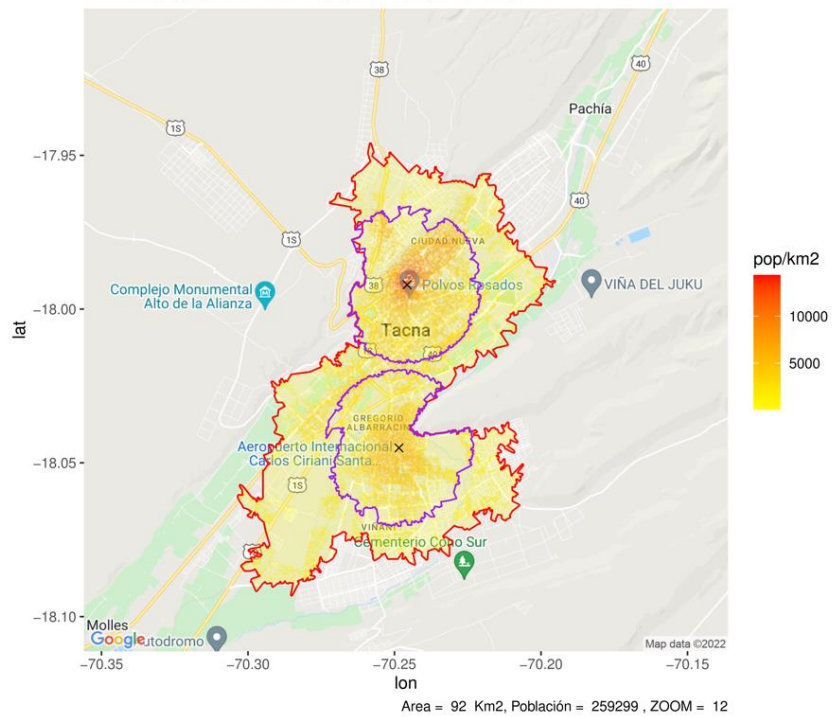
10. Santa

DELIMITACION DE AREAS URBANAS EN SANTA



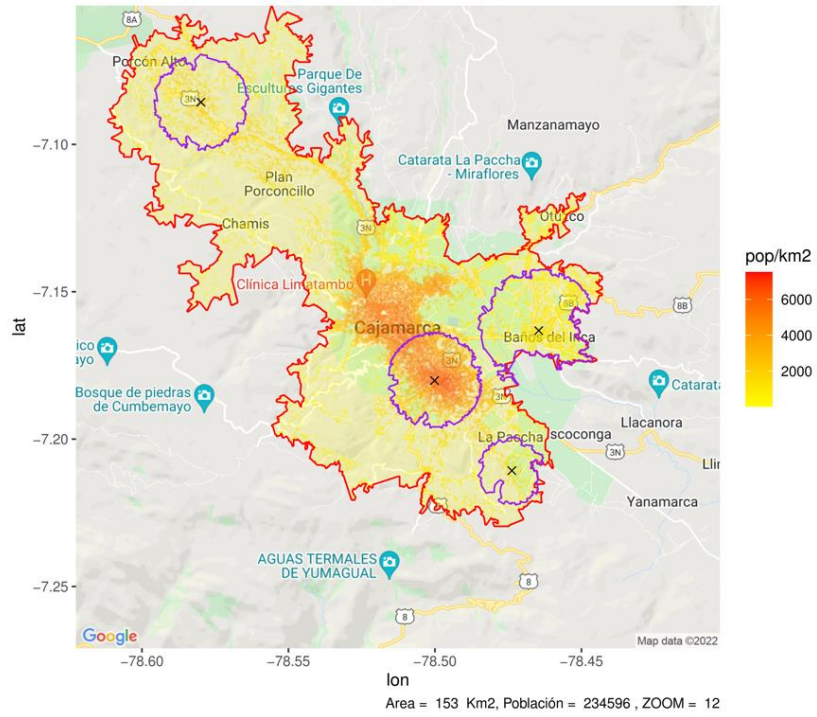
11. Tacna

DELIMITACION DE AREAS URBANAS EN TACNA



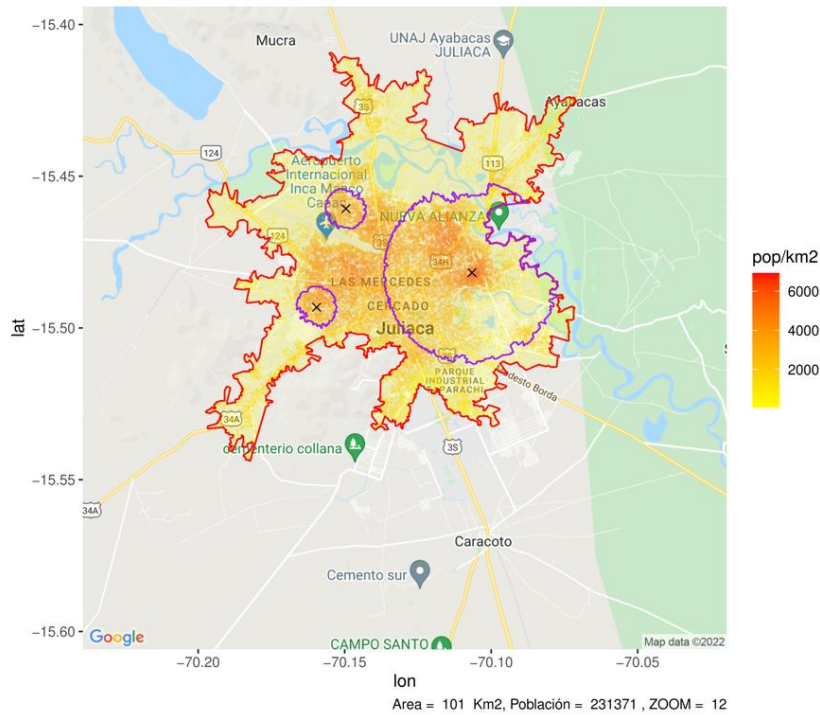
12. Cajamarca

DELIMITACION DE AREAS URBANAS EN CAJAMARCA



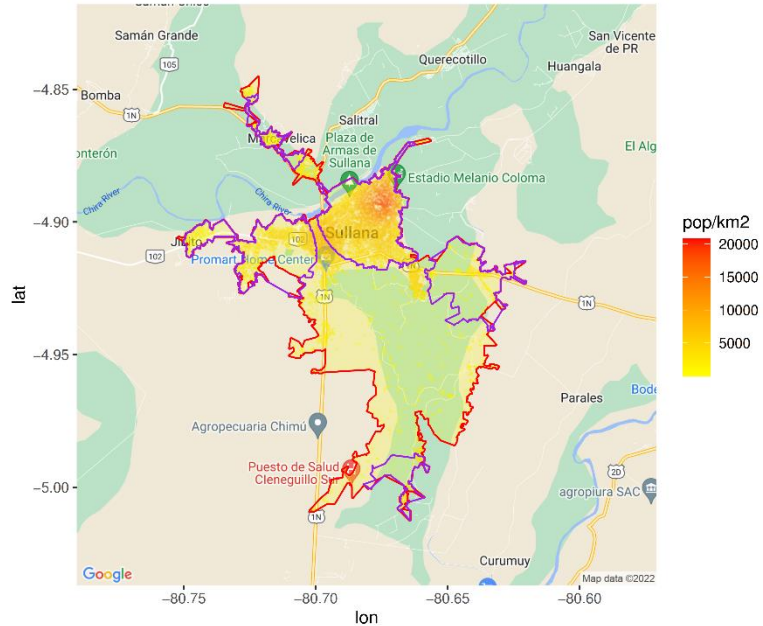
13. San Román

DELIMITACION DE AREAS URBANAS EN SAN ROMAN



14. Sullana

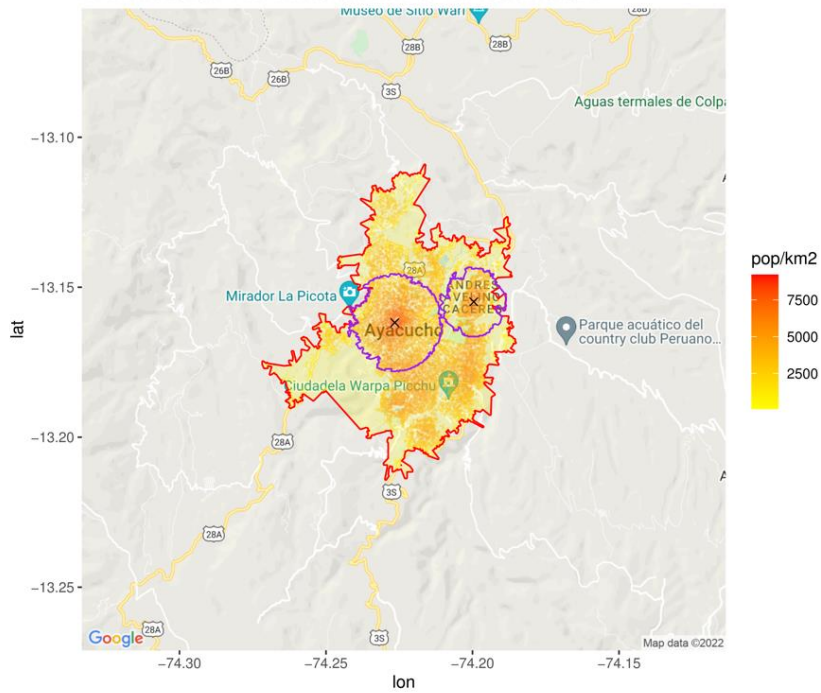
DELIMITACION DE AREAS URBANAS EN SULLANA



Area = 78 Km2, Población = 198481 , ZOOM = 12

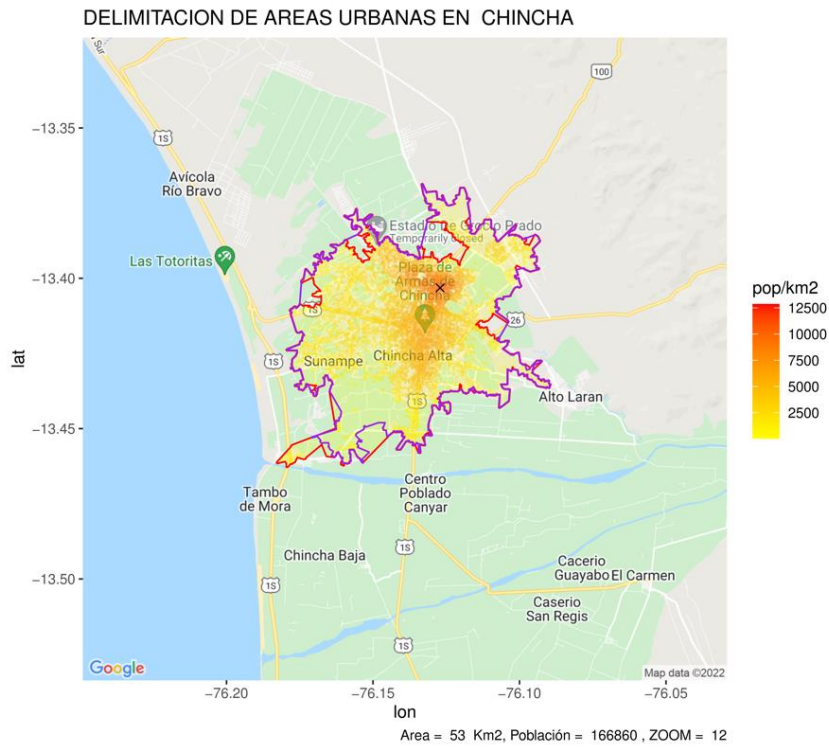
15. Huamanga

DELIMITACION DE AREAS URBANAS EN HUAMANGA

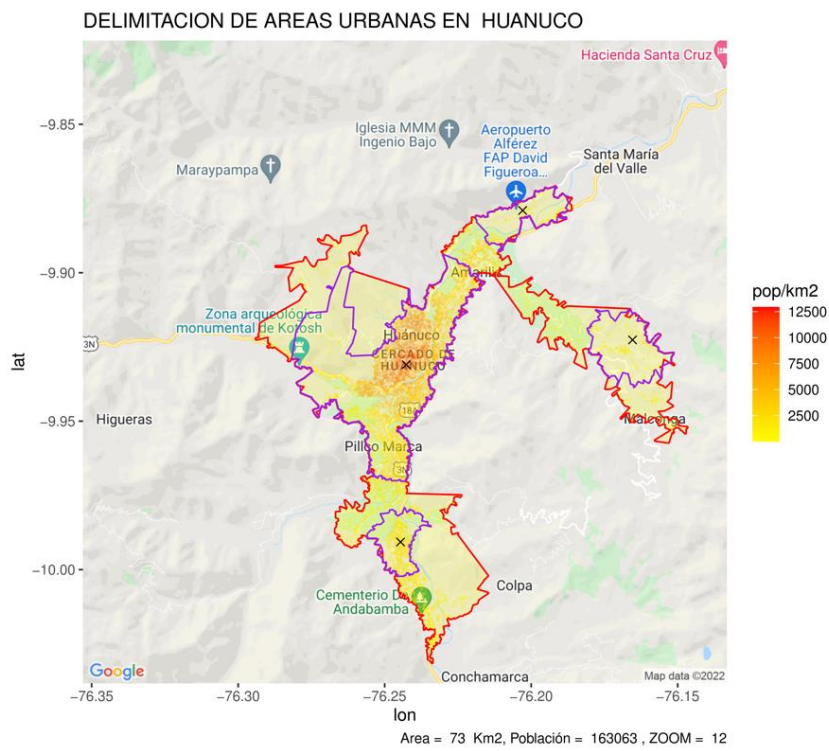


Area = 50 Km2, Población = 185565 , ZOOM = 12

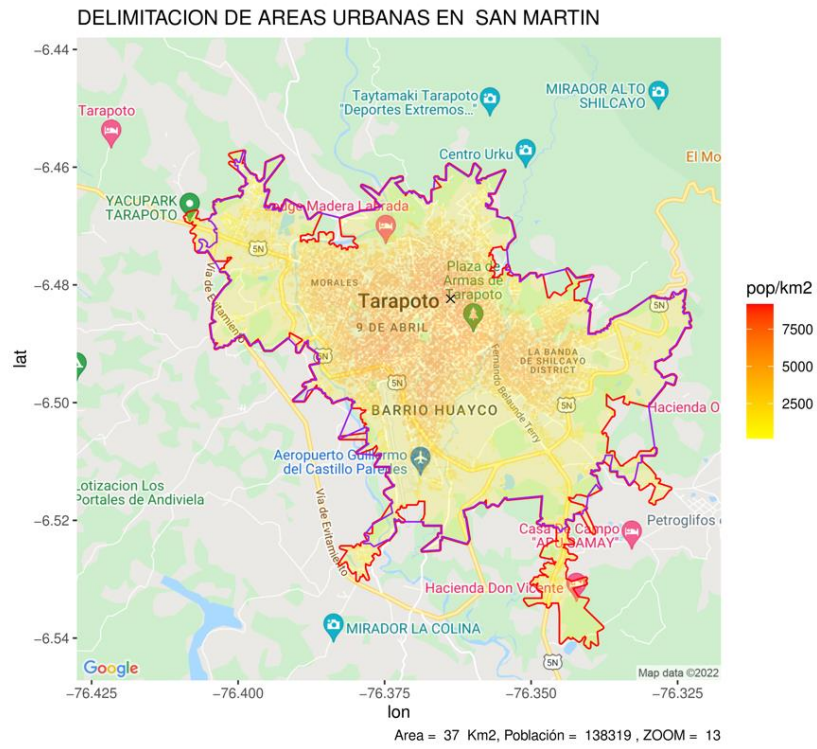
16. Chincha



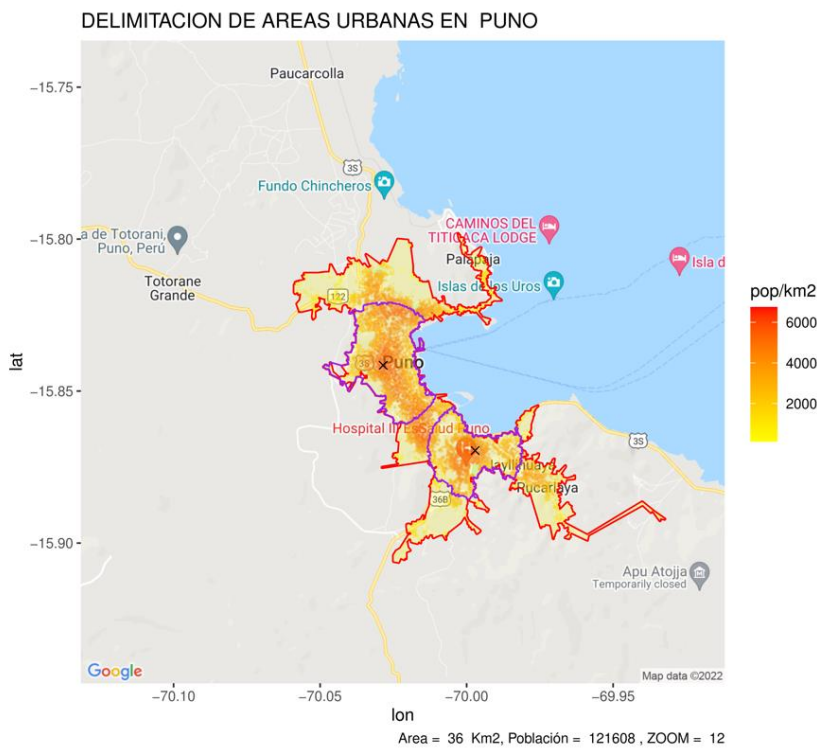
17. Huánuco



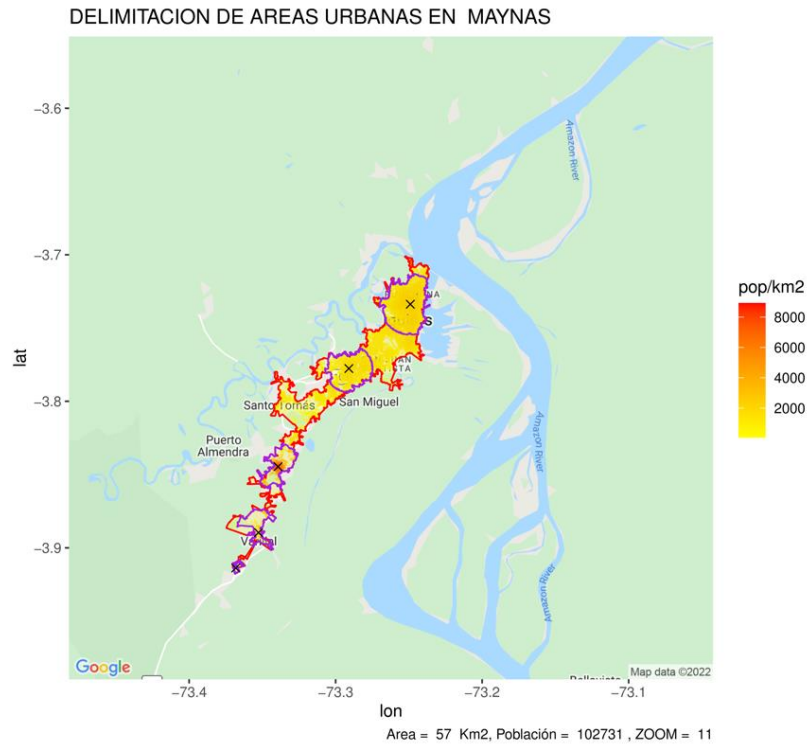
18. San Martín



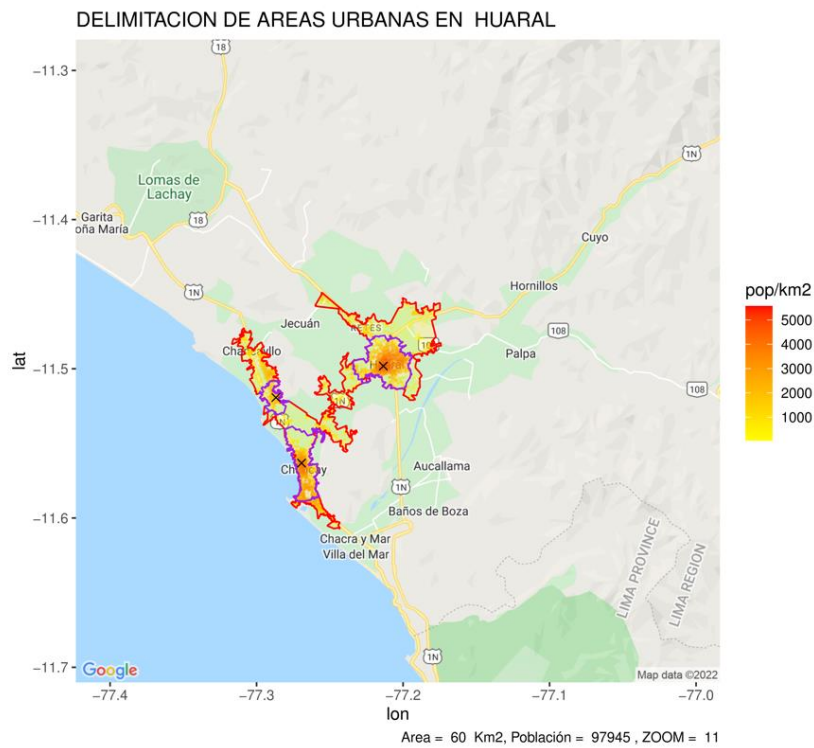
19. Puno



20. Maynas

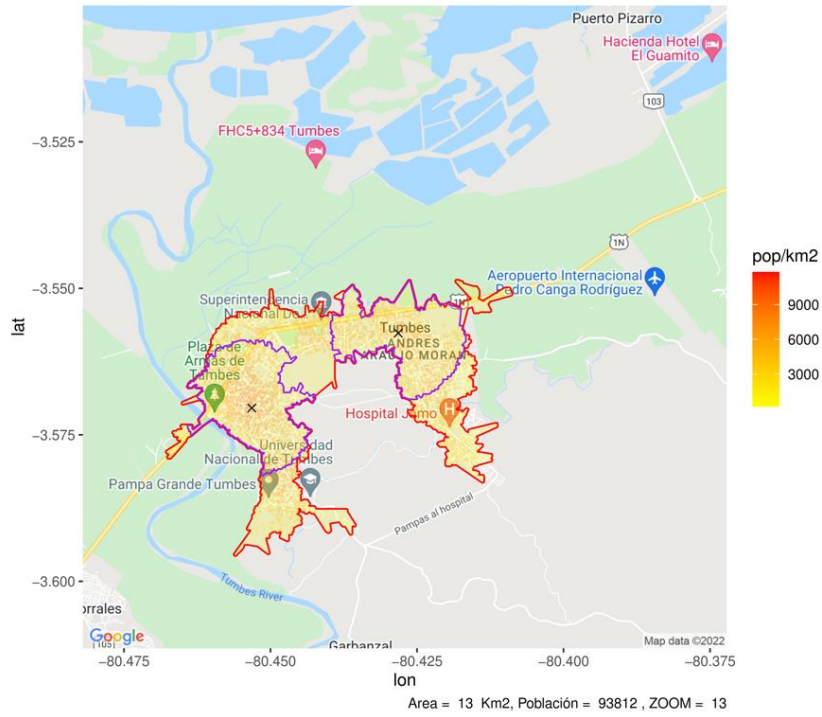


21. Huaral



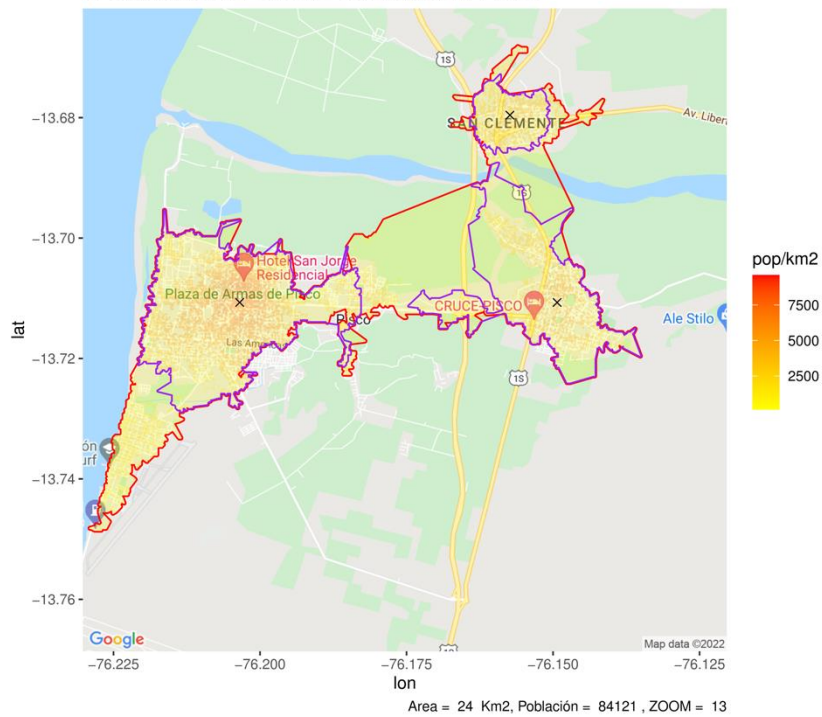
22. Tumbes

DELIMITACION DE AREAS URBANAS EN TUMBES

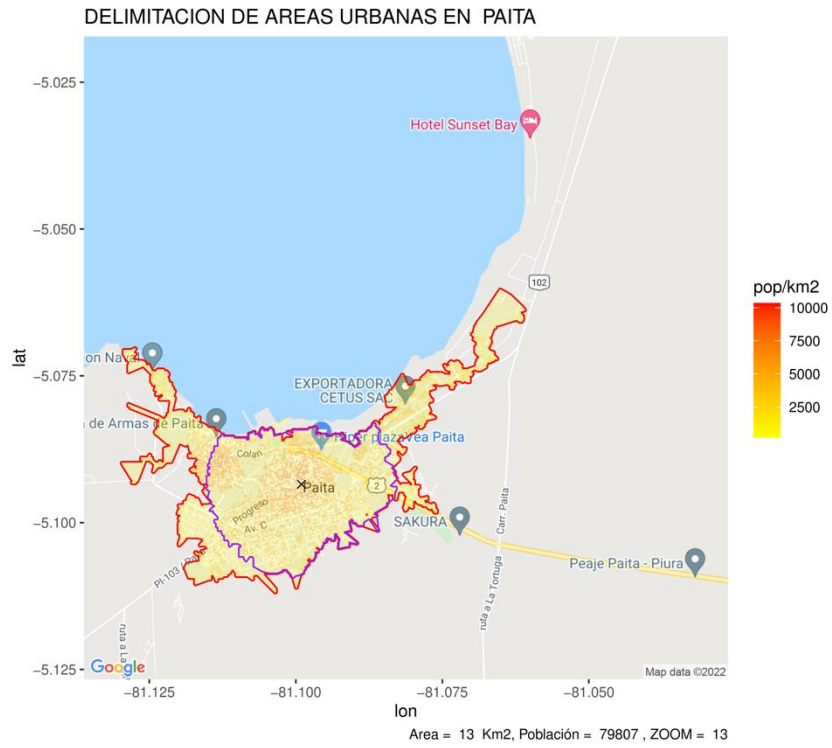


23. Pisco

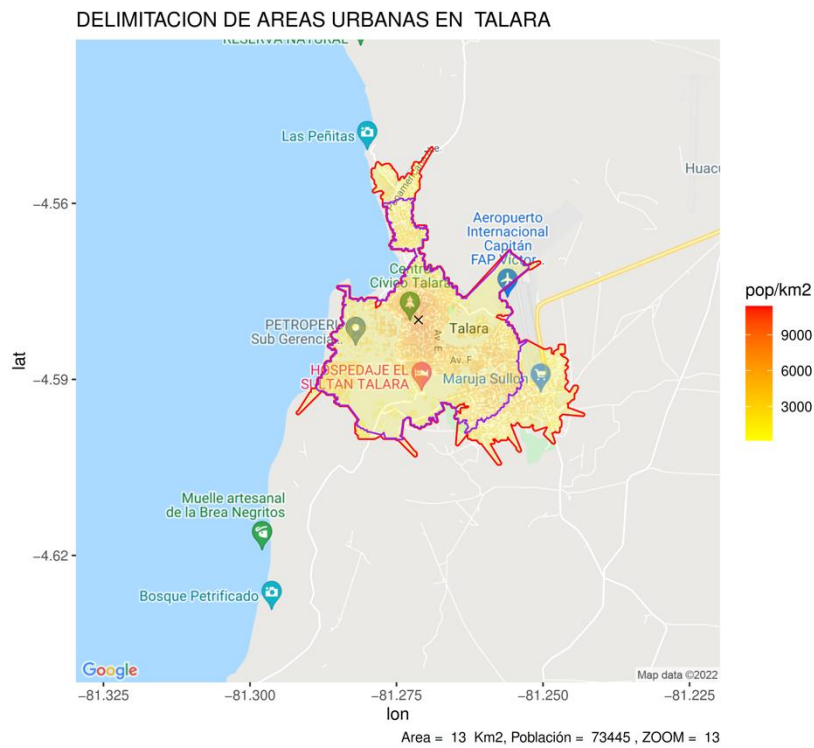
DELIMITACION DE AREAS URBANAS EN PISCO



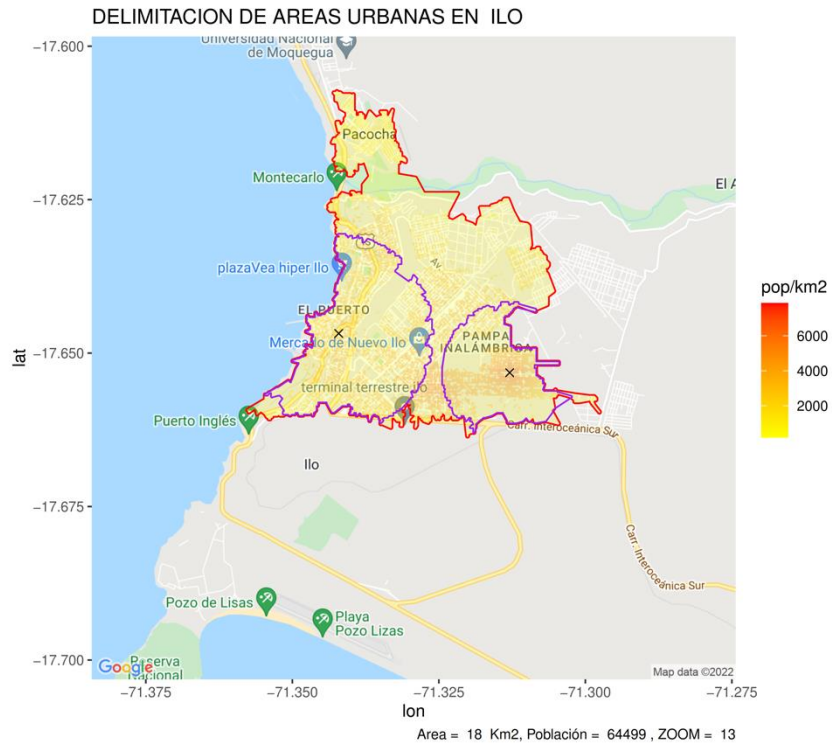
24. Paita



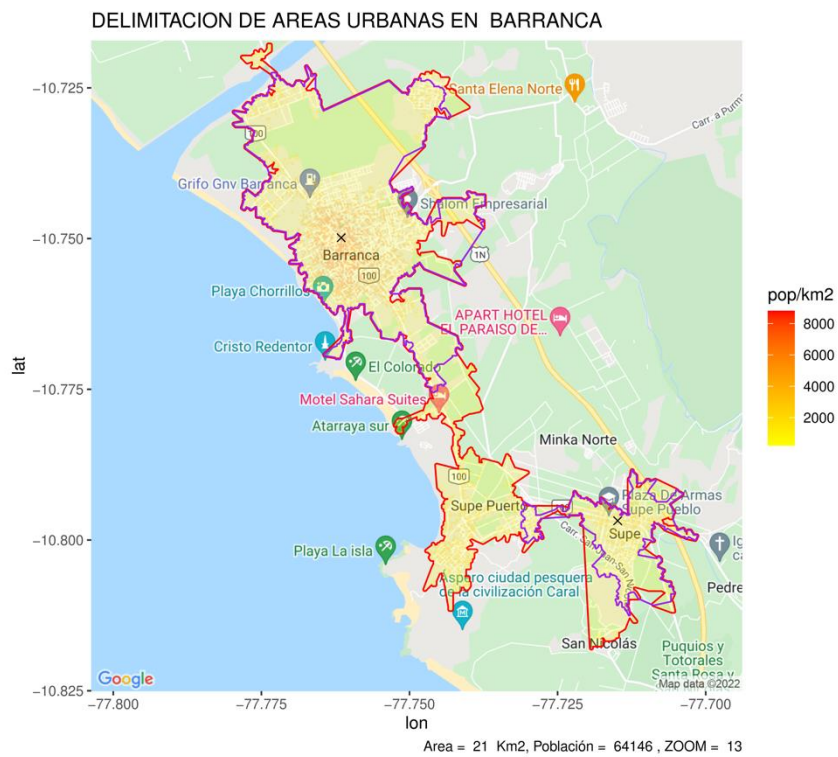
25. Talara



26. Ilo

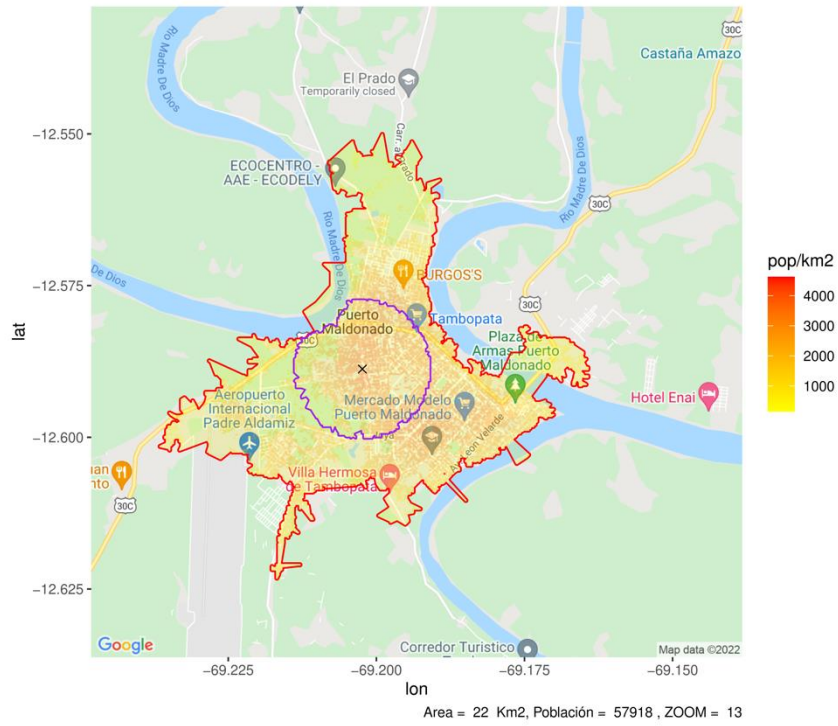


27. Barranca



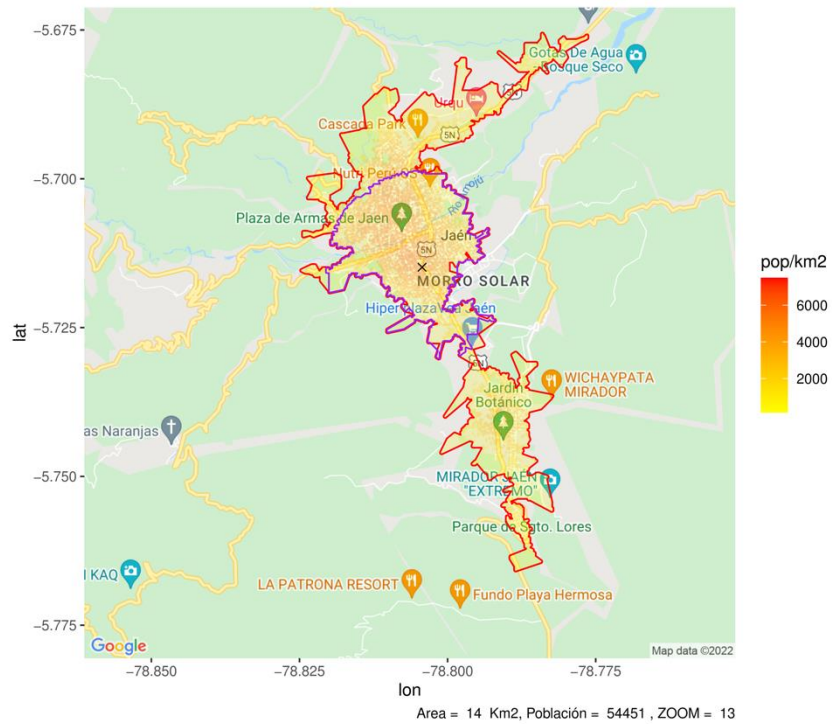
28. Tambopata

DELIMITACION DE AREAS URBANAS EN TAMBOPATA



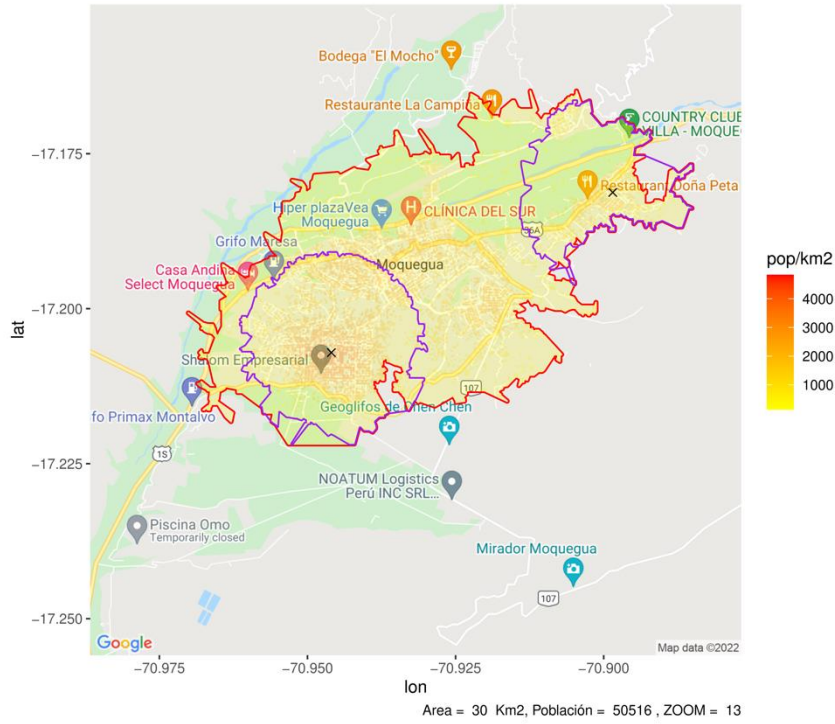
29. Jaén

DELIMITACION DE AREAS URBANAS EN JAÉN



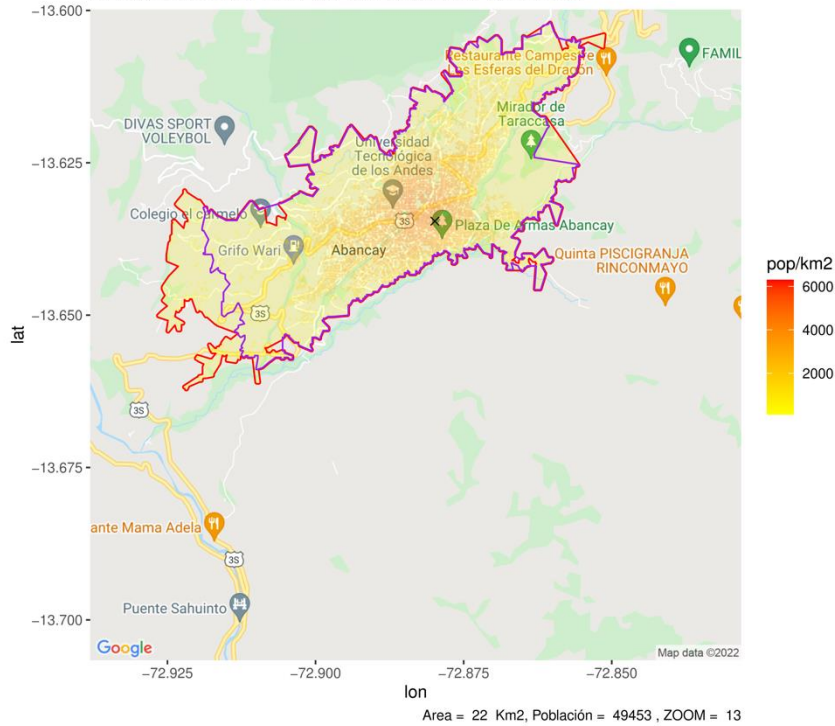
30. Mariscal Nieto

DELIMITACION DE AREAS URBANAS EN MARISCAL NIETO

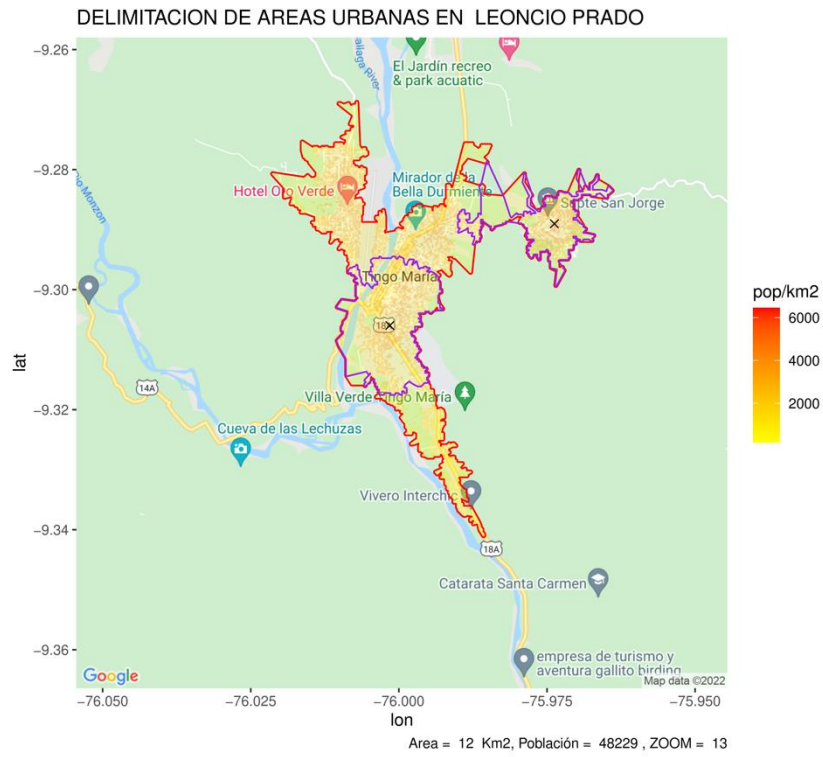


31. Abancay

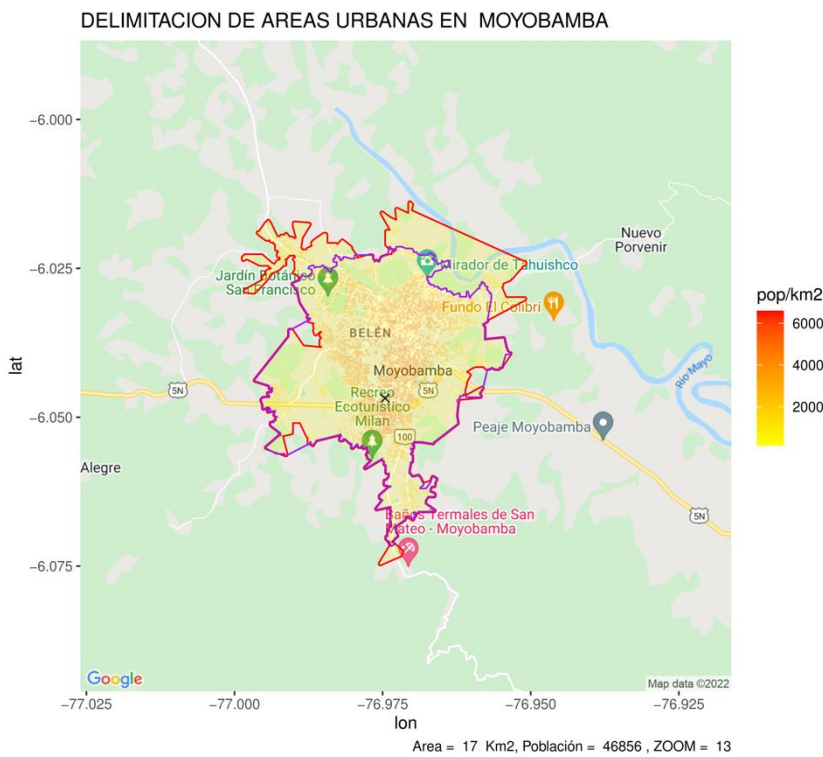
DELIMITACION DE AREAS URBANAS EN ABANCAY



32. Leoncio Prado

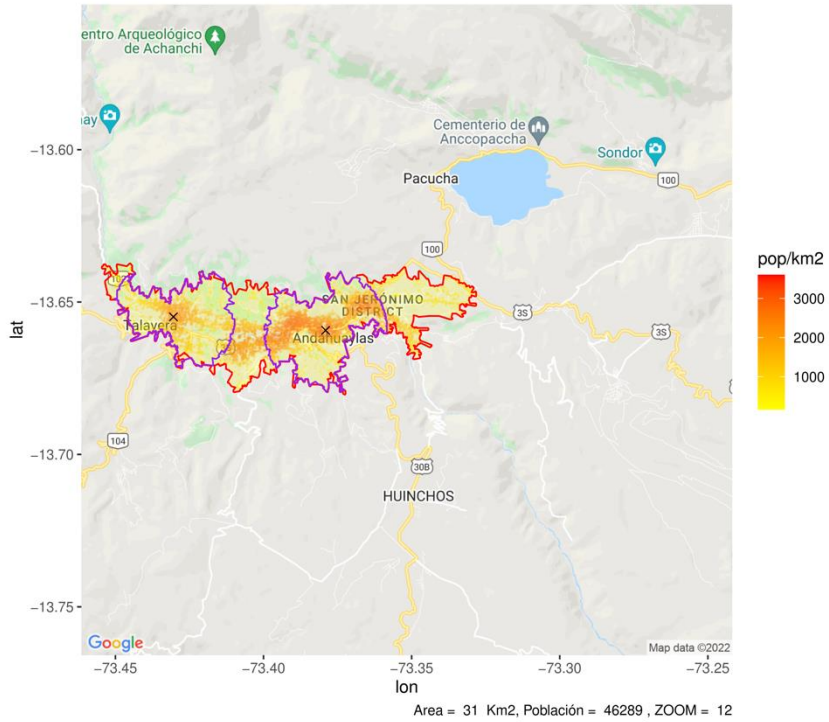


33. Moyobamba



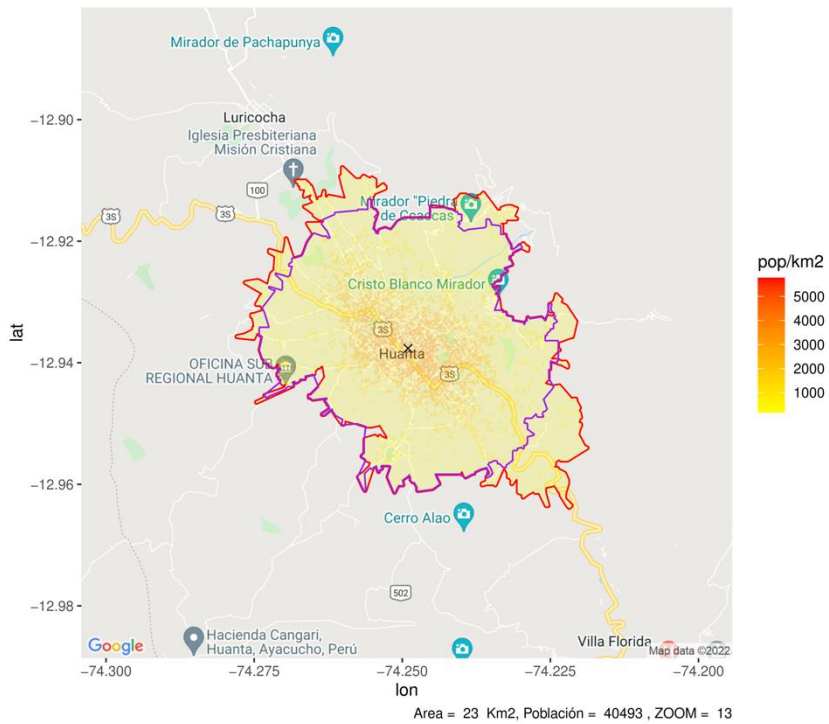
34. Andahuaylas

DELIMITACION DE AREAS URBANAS EN ANDAHUAYLAS



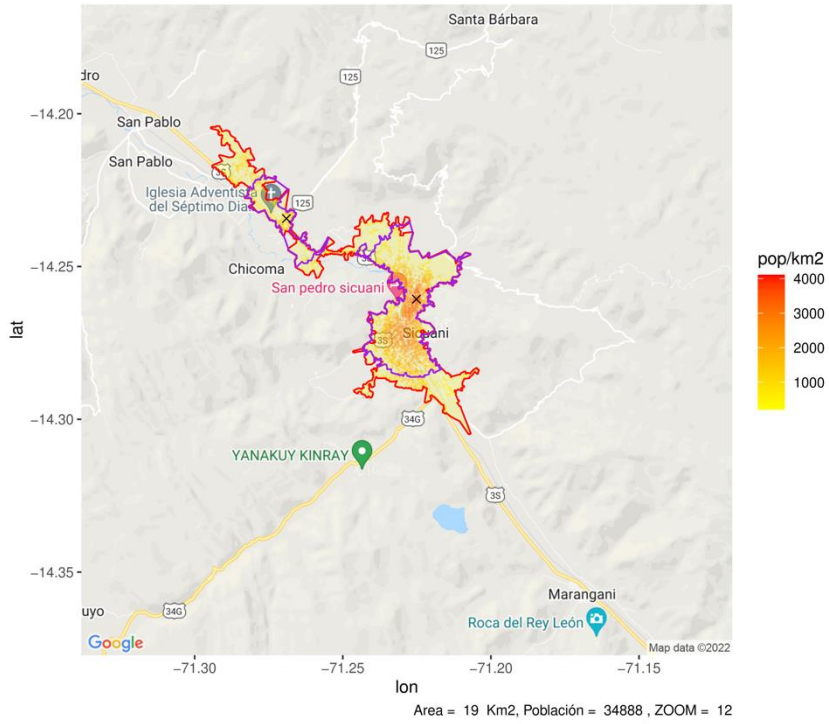
35. Huanta

DELIMITACION DE AREAS URBANAS EN HUANTA



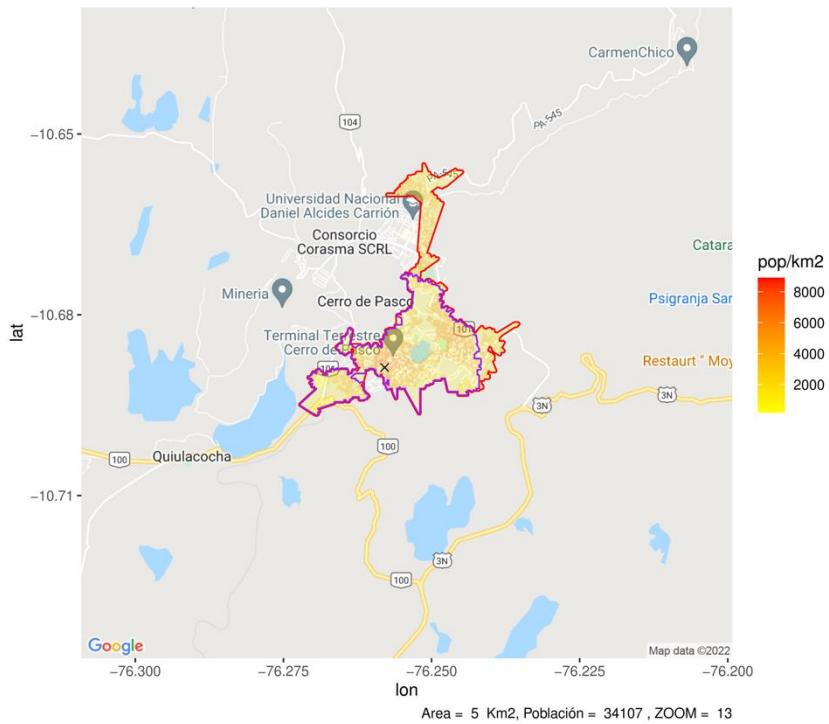
36. Canchis

DELIMITACION DE AREAS URBANAS EN CANCHIS

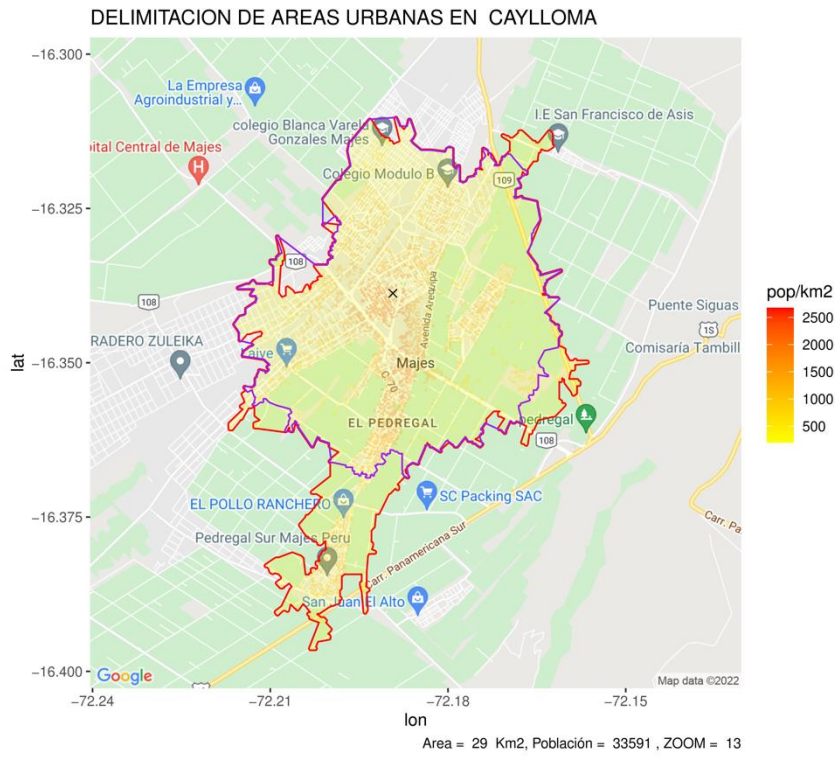


37. Pasco

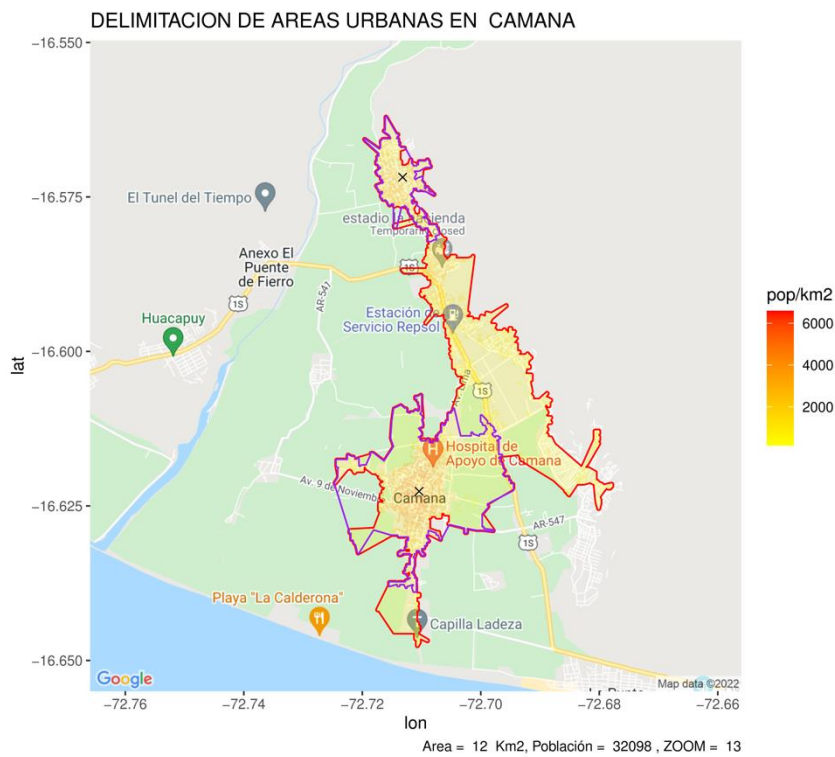
DELIMITACION DE AREAS URBANAS EN PASCO



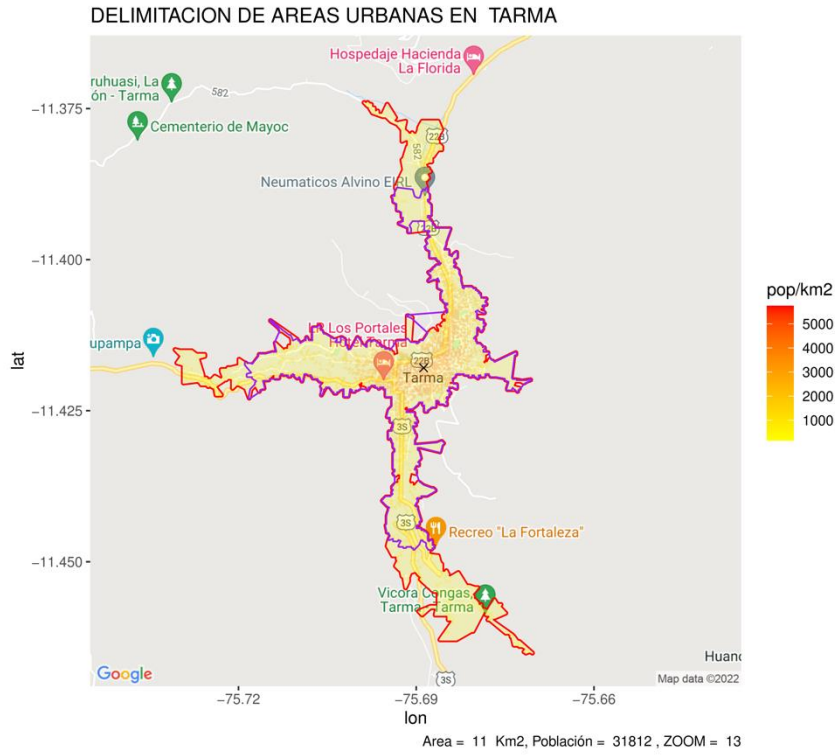
38. Caylloma



39. Camaná



40. Tarma

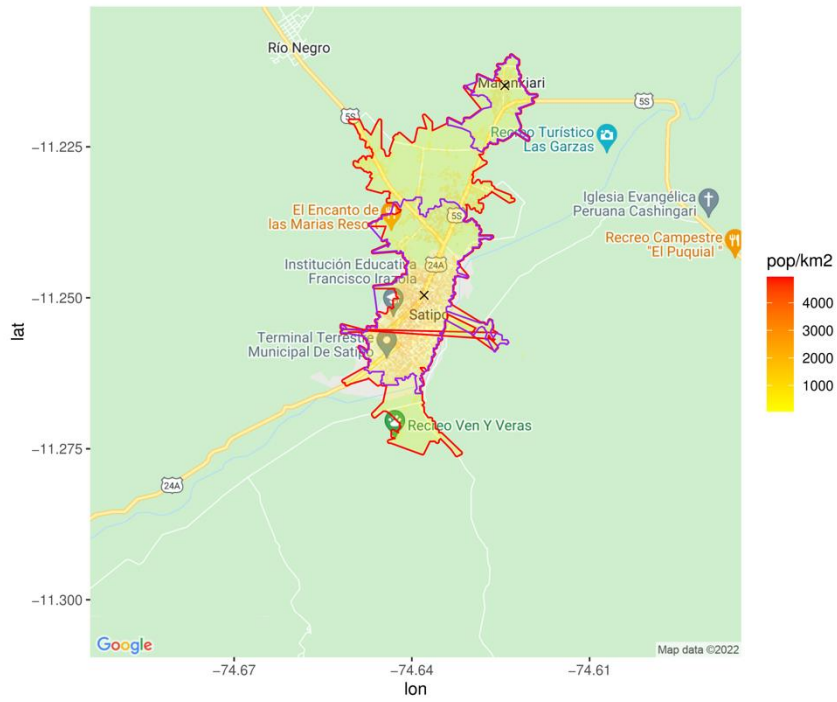


41. Rioja



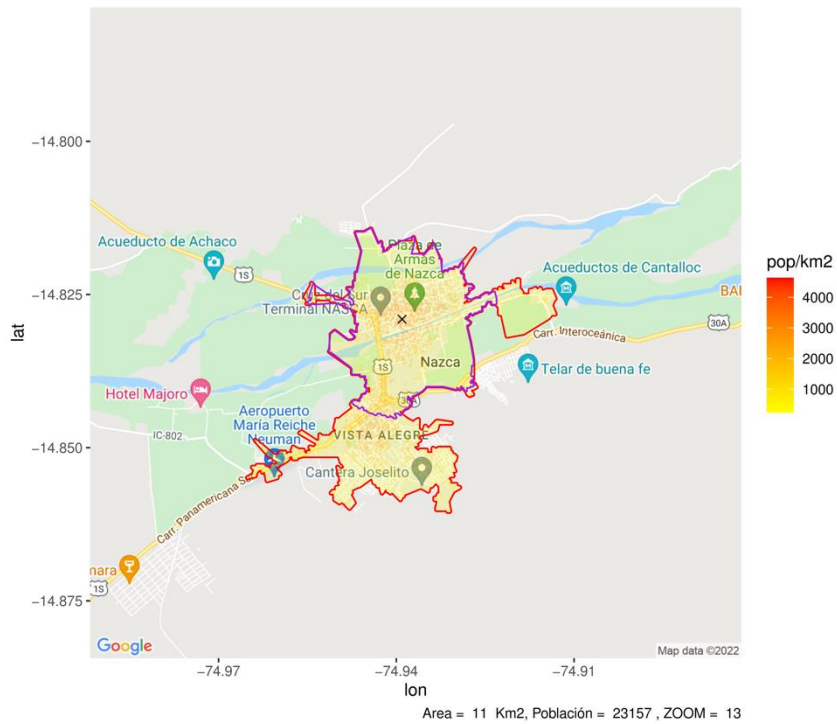
42. Satipo

DELIMITACION DE AREAS URBANAS EN SATIPO

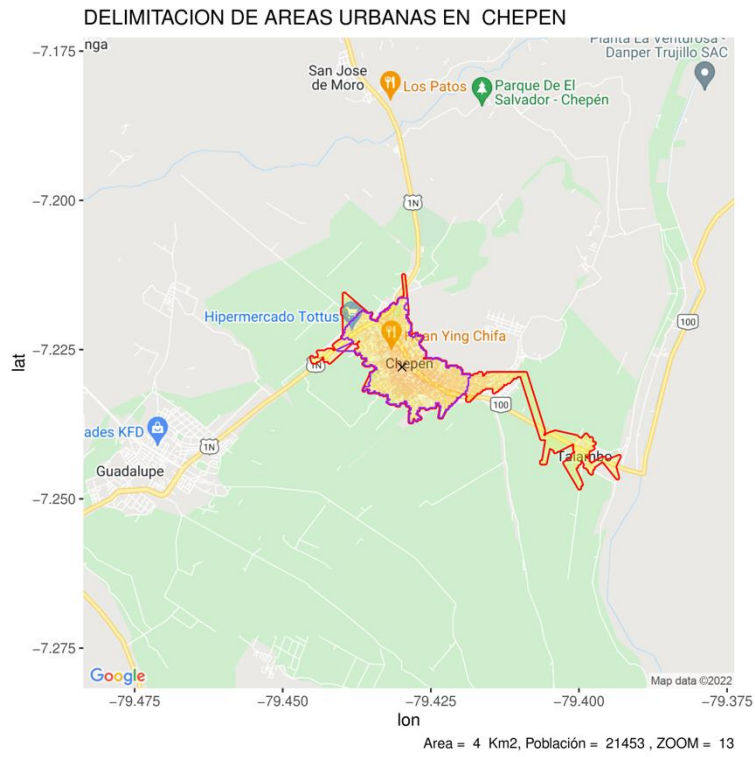


43. Nasca

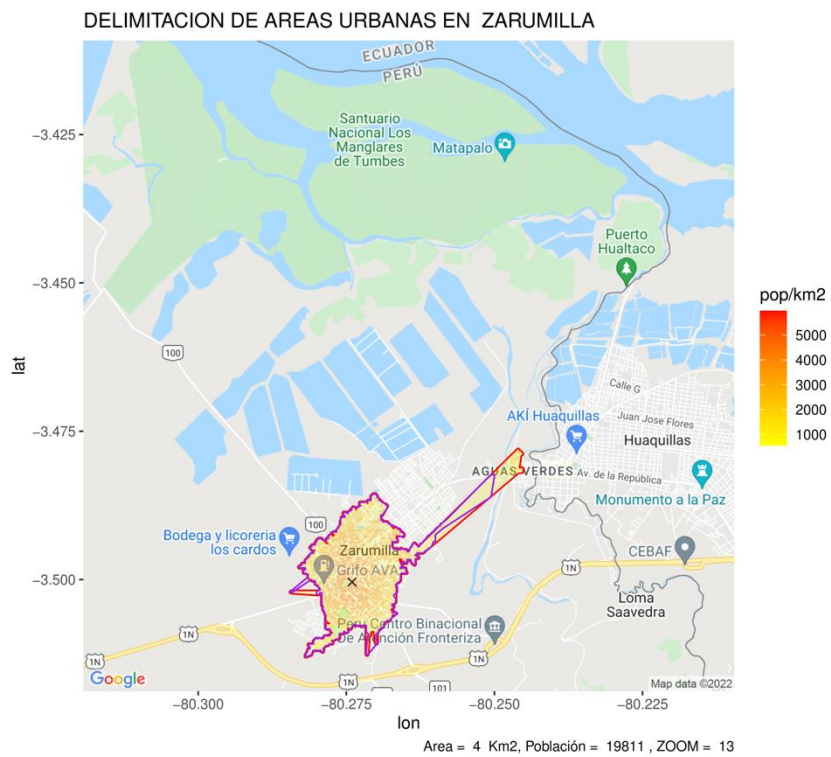
DELIMITACION DE AREAS URBANAS EN NASCA



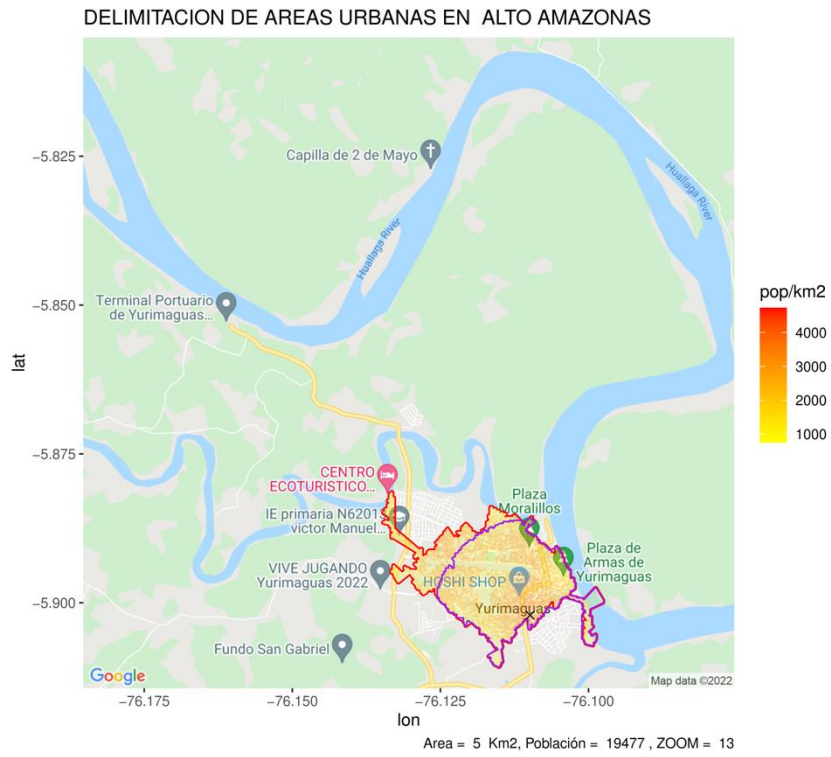
44. Chepén



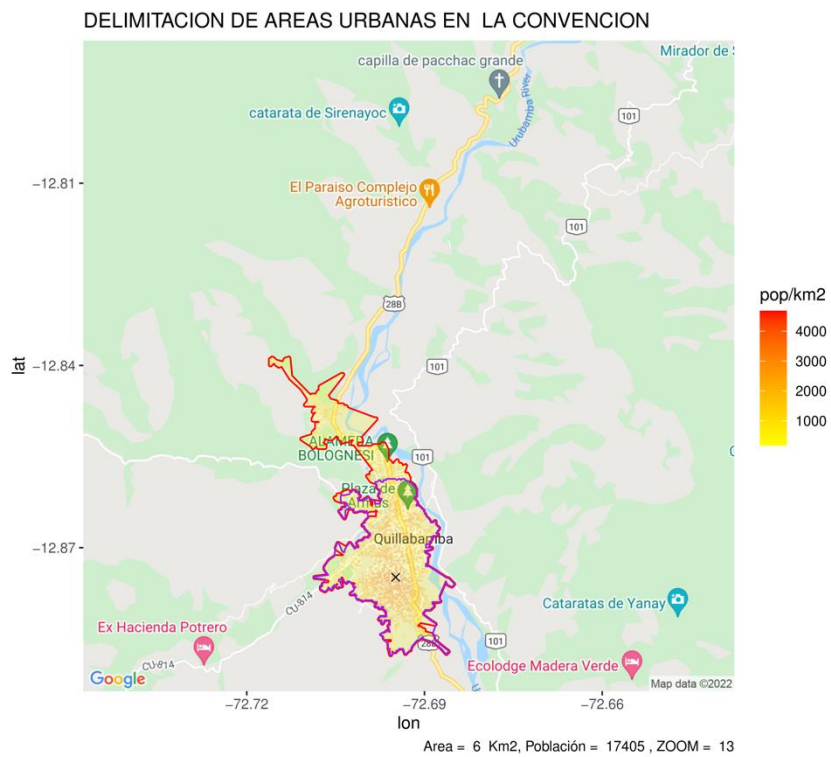
45. Zarumilla



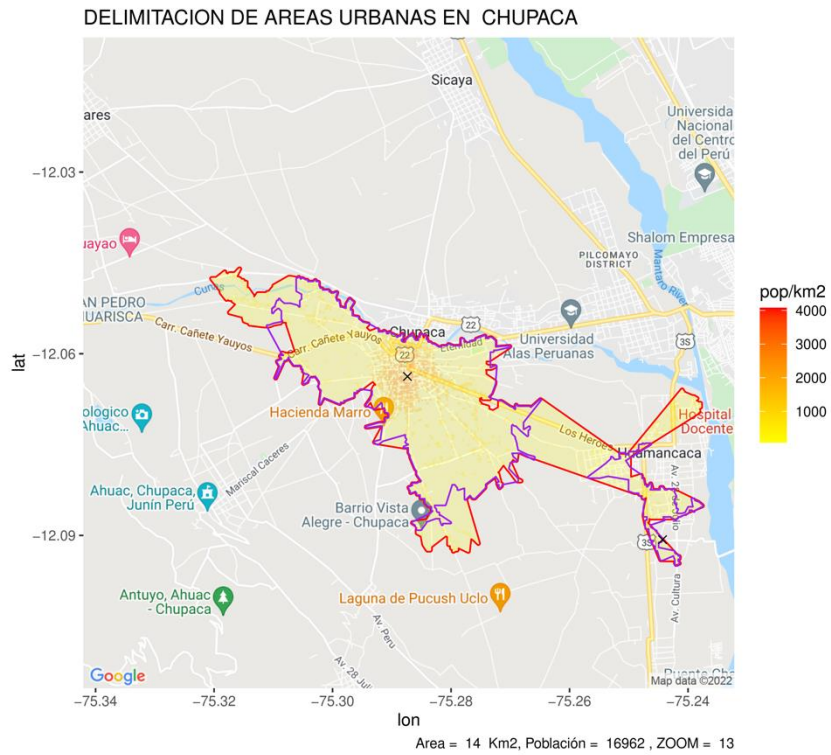
46. Alto Amazonas



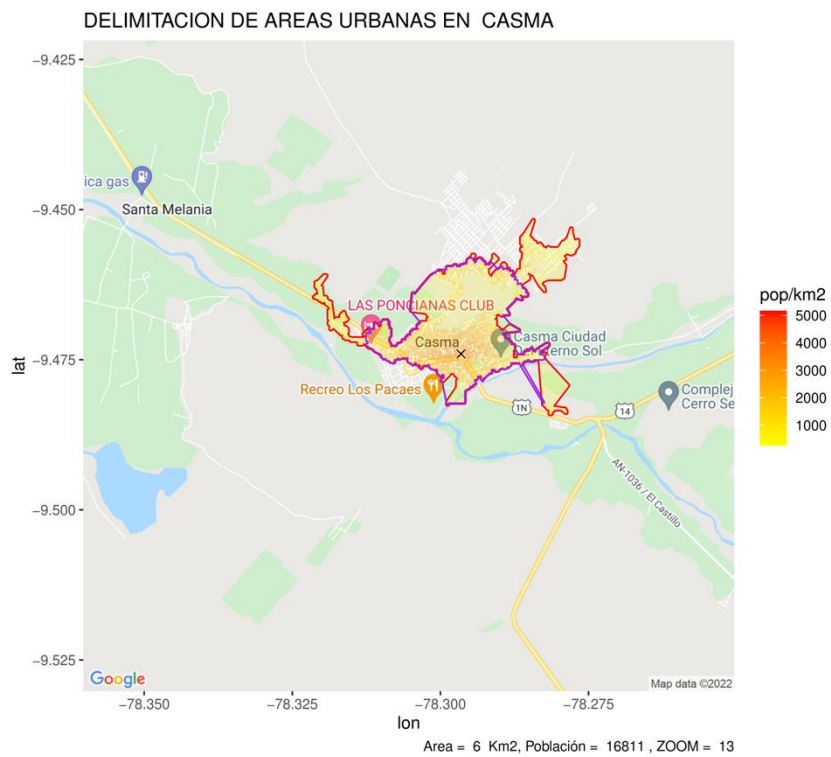
47. La Convención



48. Chupaca

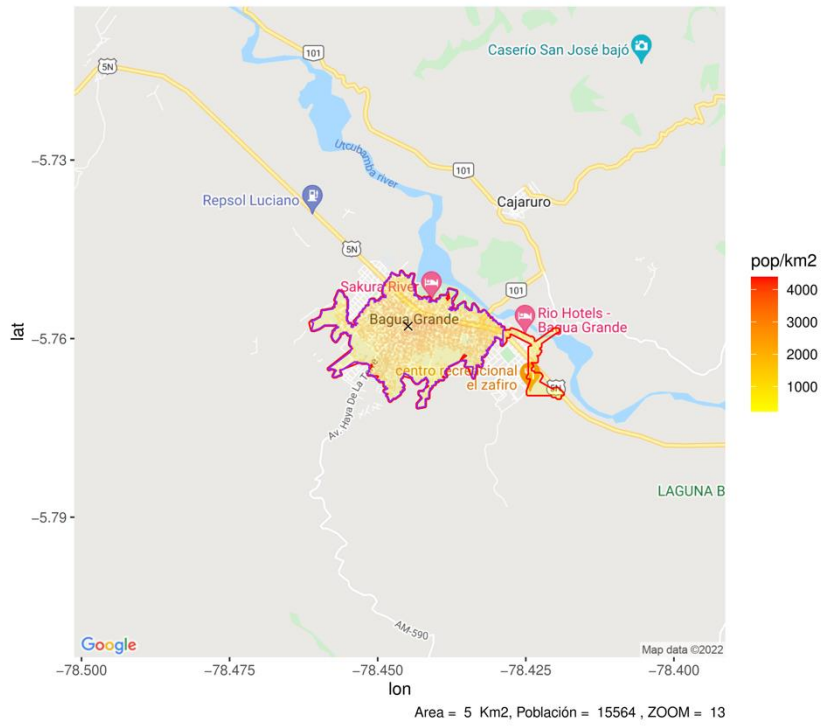


49. Casma



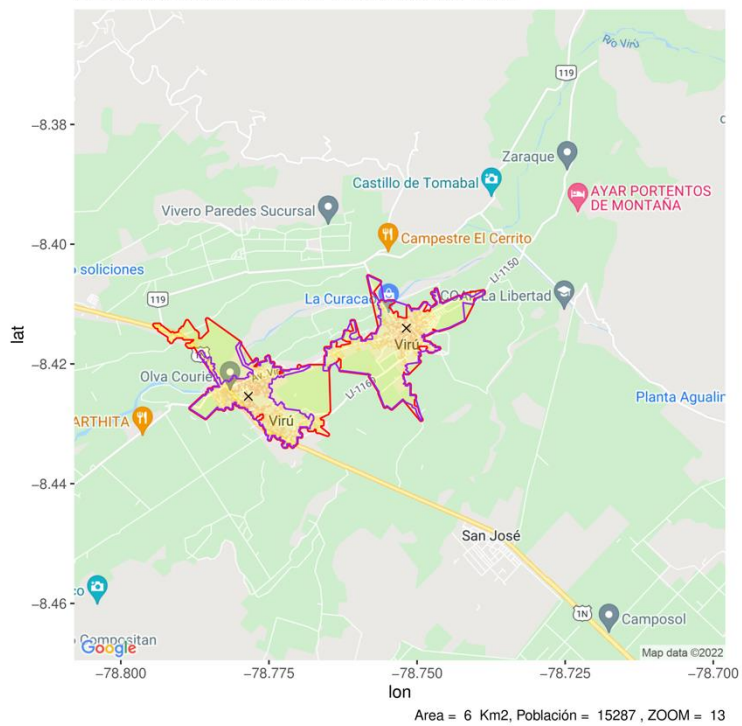
50. Utcubamba

DELIMITACION DE AREAS URBANAS EN UTCUBAMBA

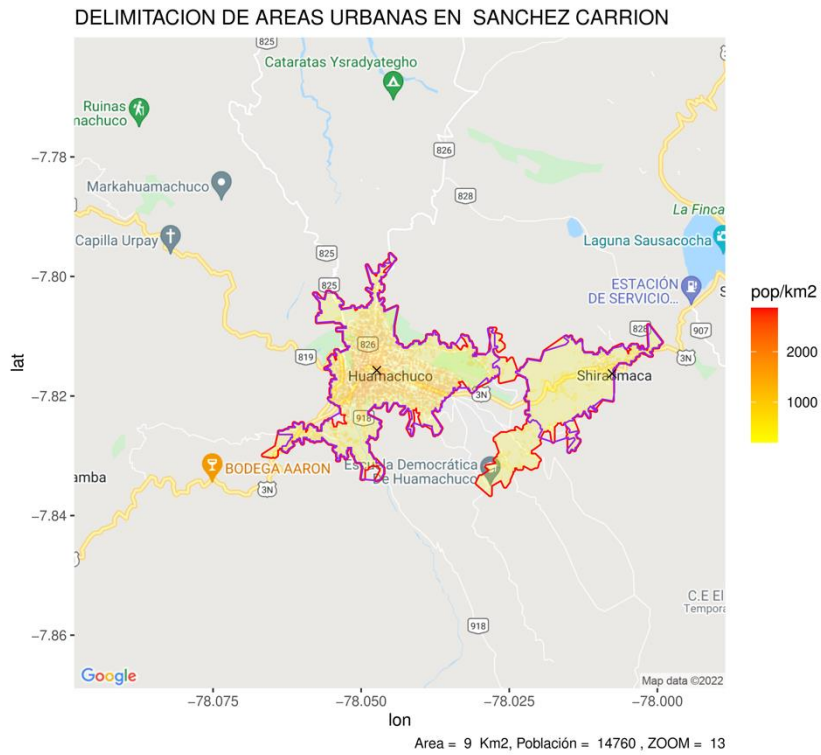


51. Virú

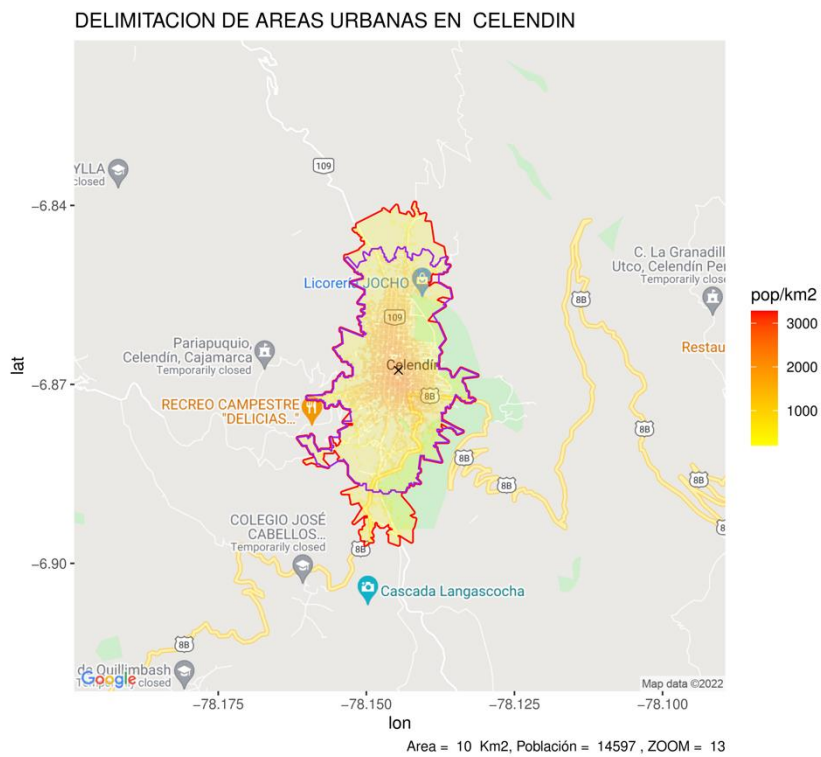
DELIMITACION DE AREAS URBANAS EN VIRU



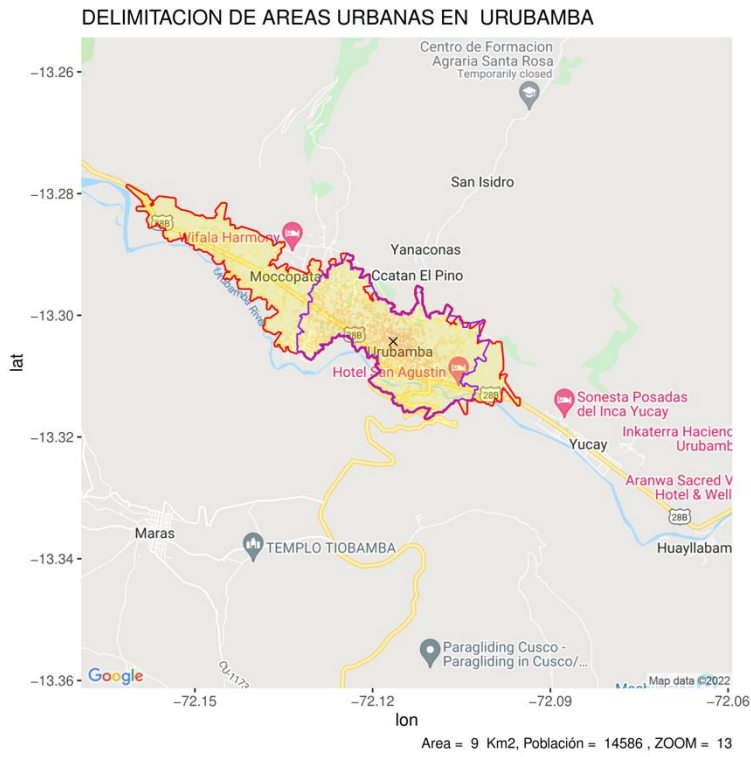
52. Sánchez Carrión



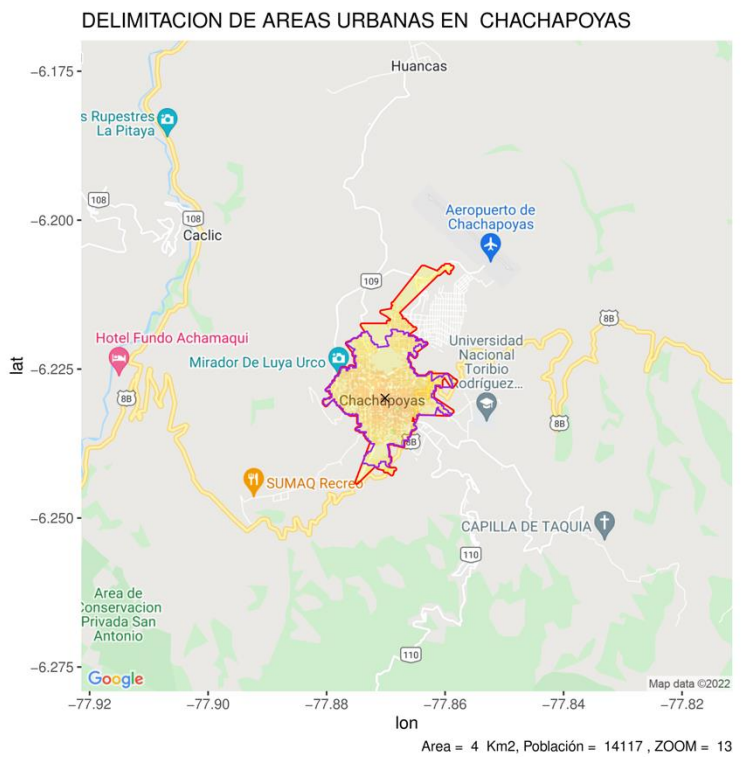
53. Celendín



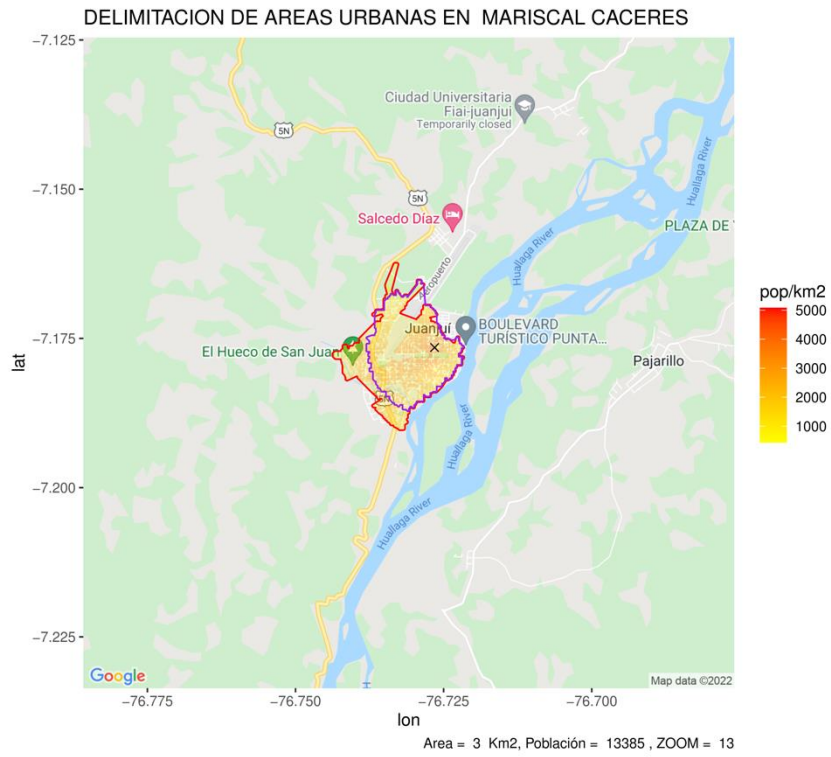
54. Urubamba



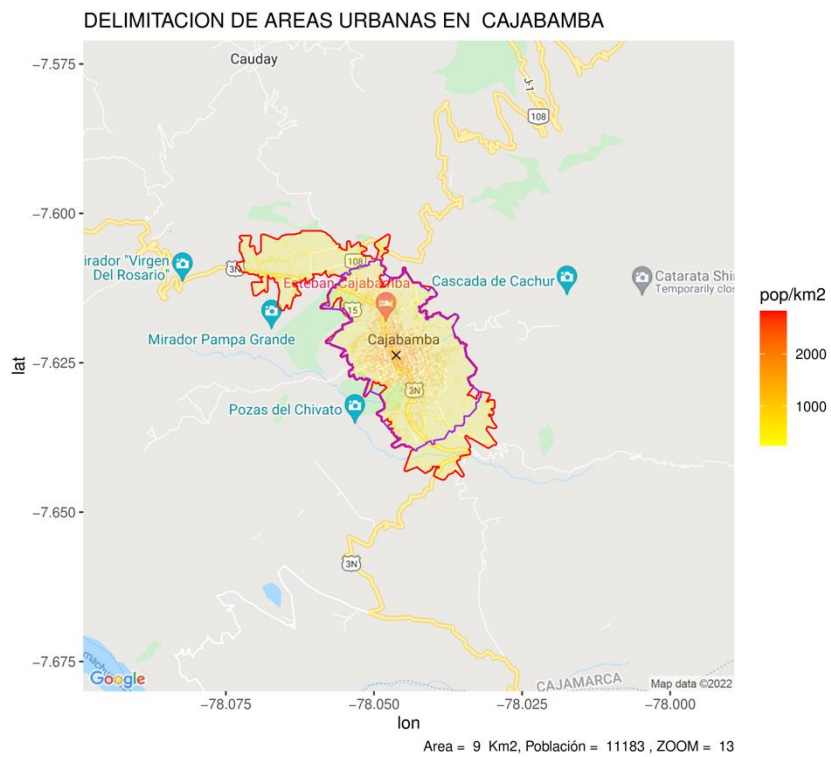
55. Chachapoyas



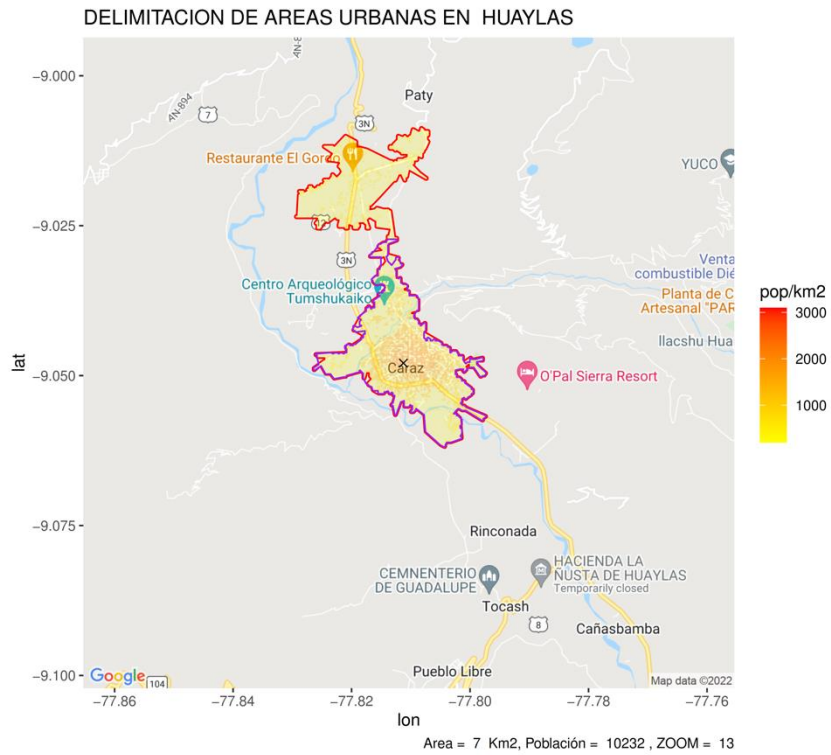
56. Mariscal Cáceres



57. Cajabamba



58. Huaylas



Fuente: Elaboración propia

Nota: El área urbana es la delimitación en rojo, el o los centros urbanos *expandidos* están delimitados en morado y, por último, dentro de cada centro urbano *expandido* se puede hallar su centro urbano correspondiente en negro.