

UNIVERSIDAD PRIVADA “FRANZ TAMAYO”

FACULTAD DE INGENIERÍA

CARRERA DE INGENIERÍA DE SISTEMAS



**“ANÁLISIS DE DATOS CENSALES PARA DETERMINAR LA
UBICACIÓN OPTIMA Y NECESARIA DE CENTROS EDUCATIVOS
EN BOLIVIA”**

ESTUDIANTE: Gabriela Micaela
Durán Villafán

ASIGNATURA: Big Data

DOCENTE: Ing. Enrique Laurel

LA PAZ – BOLIVIA

2025

Índice de Contenido

Resumen	3
1. Introducción	4
2. Planteamiento del Problema.....	5
3. Objetivos de Investigación.....	5
3.1. Objetivo General	5
3.2. Objetivos Específicos	6
4. Antecedentes	6
5. Alcances del Proyecto	7
6. Marco Teórico	8
7. Ingeniería del Proyecto.....	10
8. Límites.....	12
9. Metodología	13
9.1. Fuentes de Datos	13
9.2. Arquitectura Técnica y Proceso de Análisis.....	14
10. Resultados Preliminares y Hallazgos Clave	15
10. Conclusiones.....	19
REFERENCIAS BIBLIOGRÁFICAS.....	20
ANEXOS.....	22
Anexo 1	22
Anexo 2	22
Anexo 3	23
Anexo 4	23

Resumen

La distribución equitativa de la infraestructura educativa es un desafío fundamental para el desarrollo social en Bolivia. Este proyecto propone un modelo analítico basado en Big Data, diseñado para operar a nivel departamental. Su objetivo principal es optimizar la planificación educativa identificando, entre los nueve departamentos del país, aquellos con mayor necesidad prioritaria de nuevos centros educativos. Para ello, se procesan y analizan de manera integrada datos censales demográficos y registros administrativos de la oferta educativa actual, todos agregados y analizados por unidad departamental. La metodología elimina la necesidad de geocodificación a nivel de coordenadas precisas, centrándose en cambio en el análisis comparativo y agregado por departamento. Se calculan indicadores clave como la densidad de población en edad escolar, la relación alumno-escuela y la brecha entre oferta y demanda educativa para cada entidad territorial. Los resultados se sintetizan en un ranking o índice de priorización departamental, acompañado de visualizaciones claras (mapas temáticos y dashboards), que permiten a las autoridades nacionales y departamentales focalizar la inversión pública en las regiones con mayor déficit estructural. Este enfoque macro proporciona una herramienta de soporte decisional estratégica, basada en evidencia cuantitativa, para mejorar la equidad en el acceso a la educación en Bolivia.

Palabras clave: Big Data, planificación educativa, análisis departamental, optimización de recursos, políticas públicas, Bolivia, datos censales.

1. Introducción

La planificación urbana y la asignación eficiente de recursos públicos son pilares fundamentales para el desarrollo social y económico de cualquier país. En este contexto, la educación se erige como un derecho básico y un motor de movilidad social, cuya accesibilidad y calidad dependen en gran medida de una distribución territorial equitativa de los centros educativos. Sin embargo, la dinámica demográfica, caracterizada por constantes flujos migratorios y cambios en la densidad poblacional, genera desfases entre la oferta educativa existente y las necesidades reales de la población.

Tradicionalmente, las decisiones sobre la localización de nuevas escuelas o la ampliación de las existentes se han basado en métodos cualitativos, estudios de demanda parciales o criterios políticos, lo que a menudo resulta en una asignación ineficiente de los recursos. Esto puede traducirse en "desiertos educativos" en algunas zonas, donde los niños y jóvenes deben recorrer grandes distancias para acceder a la educación, y en una subutilización de la infraestructura en otras.

La era del Big Data ofrece una oportunidad sin precedentes para revolucionar este campo. Los censos de población y vivienda, junto con otras fuentes de datos administrativos, contienen un vasto y rico ecosistema de información sobre la composición demográfica, socioeconómica y geográfica de un territorio. El análisis de estos macrodatos permite pasar de la intuición a la evidencia, de la planificación estática a la dinámica.

Este proyecto propone el uso de técnicas de análisis de datos para procesar información censal, con el fin de identificar de manera precisa, objetiva y cuantificable las áreas

geográficas con mayor necesidad de nuevos centros educativos. El objetivo final es proporcionar a los planificadores y autoridades una herramienta de soporte decisional basada en datos que optimice la inversión pública y garantice el derecho a la educación.

2. Planteamiento del Problema

Actualmente, la planificación de la infraestructura educativa en muchas regiones no responde de manera ágil y precisa a los cambios demográficos. Esto genera una distribución geográfica desequilibrada de la oferta educativa, manifestándose en dos problemas principales:

- Sobresaturación de escuelas en departamentos donde la demanda ha decrecido, llevando a una subutilización de recursos.
- Carencia de establecimientos en departamento de crecimiento poblacional reciente, forzando a los estudiantes a largos y costosos desplazamientos o limitando su acceso a la educación.
- Una insuficiente cobertura en departamentos con población dispersa o en áreas periféricas, obligando a los estudiantes a recorrer largas distancias o, en casos extremos, abandonando el sistema educativo.

3. Objetivos de Investigación

3.1. Objetivo General

Desarrollar un análisis basado en Big Data que, mediante el procesamiento integración y análisis de datos censales y registros administrativos agregados a nivel departamental,

identifique departamentos de Bolivia con mayor necesidad para la construcción de nuevos centros educativos.

3.2. Objetivos Específicos

- Recopilar, limpiar e integrar los conjuntos de datos censales y la ubicación geográfica de los centros educativos existentes.
- Identificar indicadores de demanda y oferta educativa por departamento y los centros existentes para su representación espacial.
- Realizar un análisis para identificar departamentos con alta concentración de población en edad escolar y baja oferta educativa.
- Aplicar algoritmos de optimización para localización de servicios que indiquen los departamentos ideales para nuevos centros.
- Visualizar los resultados mediante dashboards que presenten de forma clara los resultados obtenidos.

4. Antecedentes

La aplicación de técnicas cuantitativas y de análisis espacial a la planificación de servicios públicos tiene una larga tradición en la literatura académica (ReVelle & Swain, 1970). Sin embargo, su aplicación específica a la localización de escuelas utilizando grandes volúmenes de datos censales es un campo de desarrollo más reciente, potenciado por la revolución del Big Data y la mayor disponibilidad de información geoespacial.

A nivel internacional, estudios como el de Singleton et al. (2012) en el Reino Unido utilizaron datos del censo y SIG para evaluar la accesibilidad a escuelas primarias, identificando

desigualdades socioespaciales. En América Latina, se han realizado esfuerzos similares, aunque a escalas más locales. Por ejemplo, en Brasil, proyectos han integrado datos del censo y simulaciones de accesibilidad para planificar el transporte escolar (Páez et al., 2018).

En el contexto boliviano, si bien el Instituto Nacional de Estadística (INE) genera los datos de unidades educativas, existe una brecha evidente en la integración sistemática de estas fuentes para la planificación estratégica y prospectiva. Este proyecto se sitúa precisamente en esa brecha, proponiendo un enfoque metodológico moderno, reproducible y basado en evidencia que aún no ha sido implementado de manera integral en el país, salvo por iniciativas parciales o estudios académicos aislados.

5. Alcances del Proyecto

- **Alcance Geográfico y Unidad de Análisis:** El análisis se circunscribe exclusivamente al nivel departamental de Bolivia. No se realizarán análisis a nivel municipal, de manzana censal o coordenadas específicas, lo que permite un enfoque macro estratégico y de viabilidad inmediata para la planificación nacional.
- **Alcance de Datos:** Se utilizarán fuentes de datos oficiales y agregadas. La principal fuente demográfica será el Censo Nacional de Población y Vivienda más reciente, mientras que la oferta educativa se obtendrá del Registro Nacional oficial. No se utilizarán datos de encuestas de hogares para geocodificación, sino únicamente como referencia complementaria de contexto.
- **Alcance Metodológico:** Se desarrollará un modelo de priorización basado en indicadores comparables entre departamentos. No incluirá la selección de sitios específicos dentro de un departamento, el diseño arquitectónico, el cálculo de costos

de construcción ni el análisis de factibilidad técnica de terrenos. Su propósito es la priorización interdepartamental.

- **Alcance Poblacional:** Se enfoca en la población en edad escolar oficial (por ejemplo, rangos de 4 a 18 años), segmentada según los niveles educativos (inicial, primaria, secundaria).
- **Alcance Temporal:** El análisis será un diagnóstico basado en los datos más recientes disponibles al momento de su ejecución. No constituye un sistema de monitoreo en tiempo real, aunque la metodología es replicable para futuras actualizaciones con nuevos datos censales.

6. Marco Teórico

Para fundamentar este proyecto, es esencial definir los conceptos clave que sustentan el análisis propuesto.

Equidad Territorial en la Provisión de Servicios Públicos: Este principio postula que la distribución de recursos e infraestructura pública debe compensar las desventajas históricas y geográficas para lograr igualdad de oportunidades (Soja, 2010). En educación, la equidad territorial implica que la calidad y el acceso no deben depender del departamento o región donde se nace. La falta de equidad se manifiesta en "desiertos educativos" a nivel regional, donde la oferta es insuficiente para la demanda existente (Talen, 1998).

Big Data

El término "Big Data" se refiere a conjuntos de datos de gran volumen, velocidad y variedad que requieren tecnologías avanzadas para su procesamiento y análisis (Kitchin,

2014). Big data permite transformar estos datos en "insights" accionables, mejorando la transparencia, eficiencia y efectividad de las decisiones gubernamentales (Mergel et al., 2016).

Análisis Espacial

El análisis espacial es un conjunto de técnicas para el estudio de la localización, distribución y patrones de fenómenos geográficos (O'Sullivan & Unwin, 2010). Los SIG son plataformas computacionales que integran, almacenan, analizan y visualizan datos georreferenciados, siendo fundamentales para identificar patrones de demanda y oferta de servicios (Longley et al., 2015).

Algoritmos de Optimización para Localización de Servicios

Son modelos matemáticos diseñados para determinar la ubicación óptima de uno o más servicios (p. ej., escuelas, hospitales) para servir a una población distribuida en el espacio.

Dos modelos relevantes son:

- **Modelo p-mediana:** Busca localizar p instalaciones (escuelas) de manera que se minimice la distancia total o promedio que la población debe recorrer para acceder al servicio más cercano (ReVelle & Swain, 1970).
- **Análisis de Densidad Kernel:** Técnica no paramétrica para estimar la densidad de puntos (p. ej., población en edad escolar) en un área continua, permitiendo identificar "puntos calientes" (hotspots) de alta concentración (Silverman, 1986).

7. Ingeniería del Proyecto

La arquitectura técnica del proyecto sigue un flujo de datos estructurado en cuatro etapas, implementado con herramientas de código abierto para garantizar reproducibilidad y bajo costo:

1. Adquisición y Almacenamiento de Datos Agregados:

- **Proceso:** Descarga automatizada o manual de los conjuntos de datos oficiales desde los portales web del INE (Instituto Nacional de Estadística) y el Ministerio de Educación. Los datos relevantes ya se encuentran típicamente en tablas agregadas por departamento o pueden ser sumados fácilmente a ese nivel sin necesidad de datos a nivel individual.
- **Herramientas:** Scripts en Python utilizando librerías como pandas para la manipulación de datos tabulares.
- **Almacenamiento:** Los datos limpios y consolidados se almacenan en archivos estructurados (CSV, Excel) dada la escala manejable a nivel departamental.

2. Transformación y Limpieza (ETL - Extract, Transform, Load):

- **Limpieza:** Se depuran inconsistencias, se manejan valores nulos y se estandarizan los nombres de los departamentos para garantizar la correspondencia entre las diferentes fuentes de datos.

3. Análisis y Modelado de Priorización:

- **Análisis Exploratorio de Datos (AED):** Se generan estadísticas descriptivas, gráficos de barras comparativos para visualizar la distribución de cada indicador entre departamentos. Herramientas: matplotlib en Python.
- **Construcción del Índice de Priorización:** Se selecciona un conjunto de indicadores normalizados y se agregan en un índice compuesto. Un método simple puede ser la suma de percentiles o un método multicriterio básico. El resultado es un score o ranking para cada departamento.

4. Visualización y Despliegue:

- **Desarrollo del Dashboard:** Se crea una aplicación web interactiva. Esta aplicación leerá los datos del DataMart departamental.
- **Funcionalidades:** La interfaz permitirá:
 - a) Visualizar graficos mostrando el índice de priorización o por cualquier indicador individual.
 - b) Consultar una tabla ordenable con el ranking y los valores de todos los indicadores por departamento.
 - c) Filtrar la información por nivel educativo (inicial, primaria, secundaria) si los datos lo permiten.
- **Producto Final:** Un prototipo de herramienta web accesible, que presenta un dashboard argumentado con datos.

8. Límites

- **Nivel de Agregación:** El mayor límite es la escala departamental. Al no trabajar con datos a nivel municipal o comunitario, el modelo no puede identificar necesidades específicas dentro de un departamento, donde pueden coexistir áreas saturadas y áreas con déficit. Su utilidad es para la planificación macro y la asignación de recursos entre departamentos, no para la micro-localización de escuelas.
- **Indicadores de Proximidad y Accesibilidad:** Al no usar coordenadas, no se pueden calcular distancias reales de viaje de los estudiantes a las escuelas. Los indicadores de "accesibilidad" se limitan a medidas de densidad y ratio, que son proxies menos precisas de la facilidad real de acceso.
- **Factores Cualitativos y de Contexto:** El modelo se basa exclusivamente en datos cuantitativos disponibles. No incorpora factores cruciales como la calidad de la infraestructura existente, el estado de los caminos, las barreras culturales o lingüísticas, o las prioridades políticas establecidas en planes de desarrollo específicos, los cuales pueden modificar la priorización.
- **Actualidad de los Datos Censales:** Los censos de población se realizan cada década aproximadamente. El análisis basado en el último censo puede no reflejar cambios demográficos muy recientes, como migraciones masivas o crecimiento urbano acelerado posterior al censo.
- **Capacidad vs. Matrícula Real:** El modelo utiliza capacidad estimada (ej. número de escuelas) o población en edad escolar como proxy de demanda. No siempre se dispone de datos de matrícula real por departamento, la cual puede ser menor debido a la deserción escolar, introduciendo un sesgo en el cálculo de la brecha.

9. Metodología

9.1. Fuentes de Datos

El modelo se sustentará en la integración de fuentes de datos oficiales y públicas, las cuales serán procesadas y analizadas a un nivel de agregación departamental. Las principales fuentes son:

- **Censo Nacional de Población y Vivienda (CPV):** Constituye la fuente primaria de datos demográficos. Proporciona información estructurada sobre el volumen, composición por edad, distribución espacial y características socioeconómicas de la población, desagregada por departamento. Para este proyecto, se extraerán los datos correspondientes a la población en edad escolar para cada uno de los nueve departamentos de Bolivia (INE, 2012).
- **Registro Nacional de Instituciones Educativas:** Administrado por el Ministerio de Educación, este registro contiene el inventario oficial de todas las unidades educativas del sistema público. De él se obtendrán datos clave agregables a nivel departamental, como el número total de establecimientos, su ubicación por municipio y departamento, y métricas de capacidad asociada (número de aulas, docentes, si están disponibles). Este conjunto de datos representa la variable de oferta educativa en el modelo (Ministerio de Educación, 2024).
- **Órgano electoral plurinacional (OEP):** Este registro contiene el inventario total de recintos registrados en su mayoría colegios a nivel Bolivia actualizado según las ultimas elecciones presidenciales del 2025, dicho conjunto de datos nos ayuda a realizar el conteo de establecimientos por departamento (OEP, 2025)

9.2. Arquitectura Técnica y Proceso de Análisis

El flujo de trabajo se ha diseñado como una etapa secuencial y reproducible, estructurado en cuatro etapas principales que transforman datos brutos en recomendaciones estratégicas:

1. **Adquisición y Preprocesamiento:** En esta etapa inicial, se recopilarán todos los conjuntos de datos mencionados desde sus portales oficiales en línea. Los datos tabulares (CSV, Excel) se cargarán en un entorno de trabajo controlado. El preprocesamiento consistirá en limpieza de datos: identificación y tratamiento de valores nulos, eliminación de duplicados, corrección de inconsistencias en los nombres de departamentos y estandarización de formatos de fechas y códigos. El objetivo es crear un conjunto de datos base limpio y consistente para cada fuente.
2. **Transformación, Integración y Cálculo de Indicadores:** Aquí, los datos de las distintas fuentes se unificarán. Se cruzarán las tablas demográficas (Censo) y de oferta educativa (Registro) utilizando el departamento como clave de unión. A partir de esta integración, se procederá al cálculo sistemático de los indicadores departamentales clave que alimentarán el modelo de priorización. Estos incluyen:
 - Población en edad escolar por departamento.
 - Número total de unidades educativas por departamento.
 - Población en edad escolar dividida entre el número de escuelas en el departamento.
 - Población en edad escolar por kilómetro cuadrado de superficie departamental.
 - Diferencia entre las variables alumno/escuela de un departamento y el promedio nacional.

3. **Análisis, Modelado y Priorización:**

- **Análisis Exploratorio de Datos (AED):** Se realizará un análisis estadístico descriptivo (media) y visual de los indicadores calculados. Se generarán gráficos para visualizar la distribución geográfica de cada indicador, permitiendo una identificación inicial de disparidades interdepartamentales.
- **Validación:** Los resultados del análisis se contrastarán con fuentes de validación externas, como informes de diagnóstico del Ministerio de Educación, planes de desarrollo departamental o estudios académicos previos, para evaluar su consistencia y validez aparente.

4. **Visualización y Comunicación:** Los hallazgos se transformarán en productos de conocimiento accesibles. Se desarrollará un dashboard interactivo que permita visualizar el mapa de Bolivia coloreado según el índice de priorización, explorar una tabla con el ranking y los valores de todos los indicadores, y filtrar la información. Adicionalmente, se generarán reportes ejecutivos estáticos con gráficos y mapas clave que resuman la evidencia para la toma de decisiones.

10. **Resultados Preliminares y Hallazgos Clave**

Aunque el proyecto se encuentra en fase de diseño metodológico, el enfoque propuesto permite proyectar el tipo de hallazgos y resultados que se esperan obtener al aplicar el modelo a los datos reales:

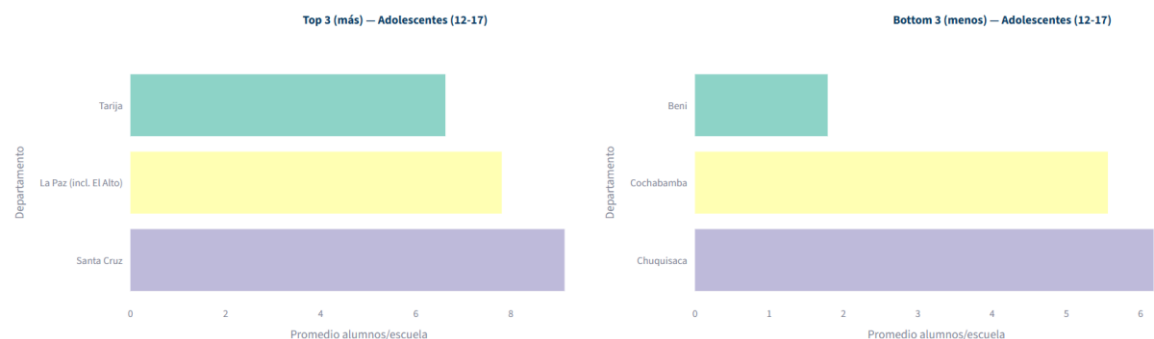
- **Identificación de Disparidades Interdepartamentales en Oferta Educativa:** Se anticipa que el análisis revelará diferencias significativas en el indicador ratio

alumno/escuela entre departamentos. Es probable que departamentos con alta densidad poblacional urbana (ej., La Paz, Cochabamba) o con alta dispersión poblacional (ej., Pando, Beni) presenten los ratios más altos, indicando una mayor presión sobre la infraestructura existente. Por el contrario, departamentos con una red escolar más densa en relación a su población mostrarán ratios más bajos.

Departamentos con MÁS / MENOS estudiantes por unidad (por grupo etario)

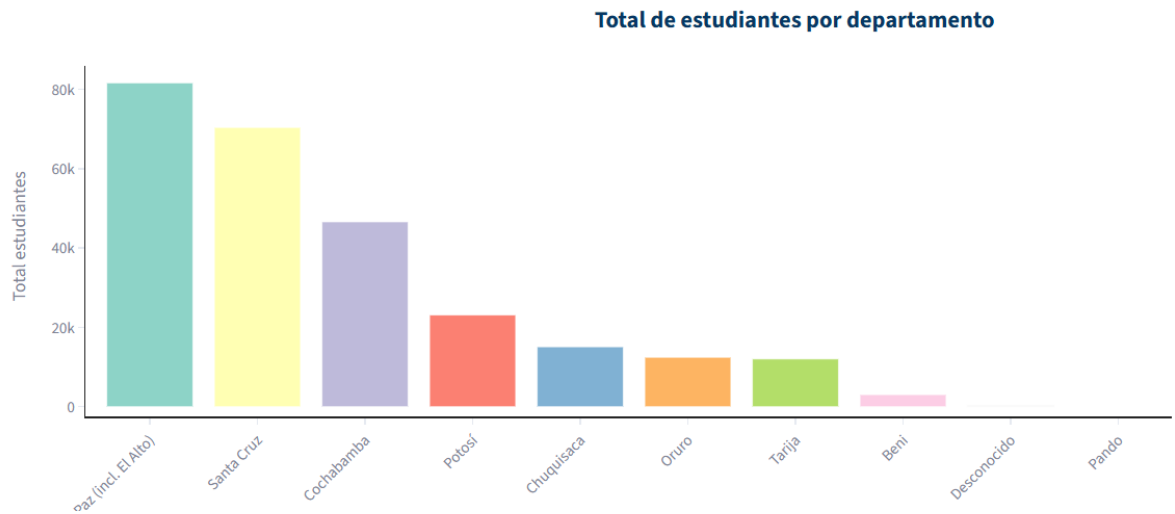
Mostramos, para cada grupo de edad, los 3 departamentos con mayor promedio de estudiantes por unidad educativa y los 3 con menor promedio (ignorando departamentos sin escuelas asignadas).

◆ Adolescentes (12-17) ↻



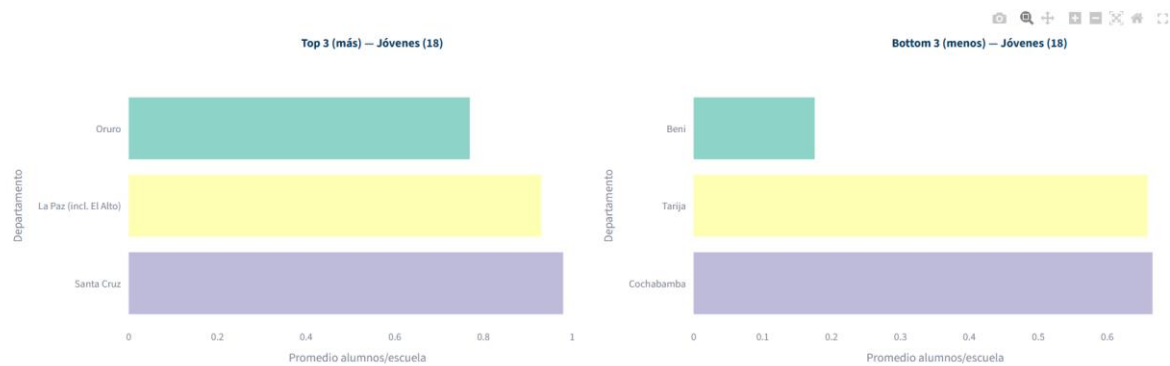
- **Detección de "Déficits Relativos" y "Excesos de Cobertura":** Al cruzar el indicador de densidad de demanda con la distribución geográfica de las escuelas, se podrán identificar departamentos que, a pesar de tener una población escolar absoluta no tan alta, presentan una baja densidad de escuelas por área geográfica, sugiriendo posibles "desiertos educativos" o problemas de accesibilidad por distancia. Asimismo, se podrían señalar departamentos donde la densidad de escuelas parece ser adecuada o incluso exceder la demanda potencial relativa.

📌 Estudiantes por departamento

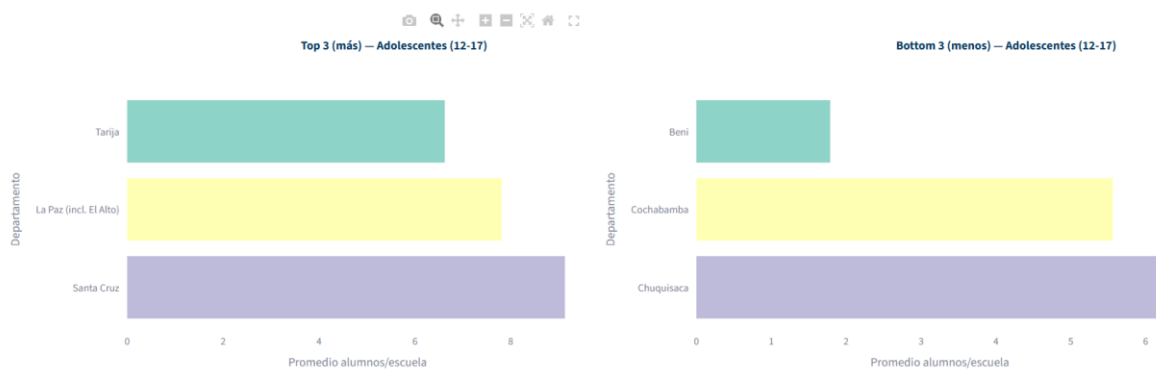


- **Ranking de Priorización Departamental:** El principal resultado concreto será un ordenamiento claro y cuantificado de los nueve departamentos según su necesidad relativa de inversión en nueva infraestructura escolar. Este ranking, respaldado por los indicadores objetivos, proporcionará una herramienta fundamental para la asignación estratégica de recursos del gobierno central. Por ejemplo, el modelo podría identificar que los departamentos La Paz, Santa Cruz y Cochabamba ocupan los primeros puestos en el índice de prioridad, recomendando que las próximas inversiones nacionales en construcción de escuelas se focalicen de manera preferente en esas regiones.

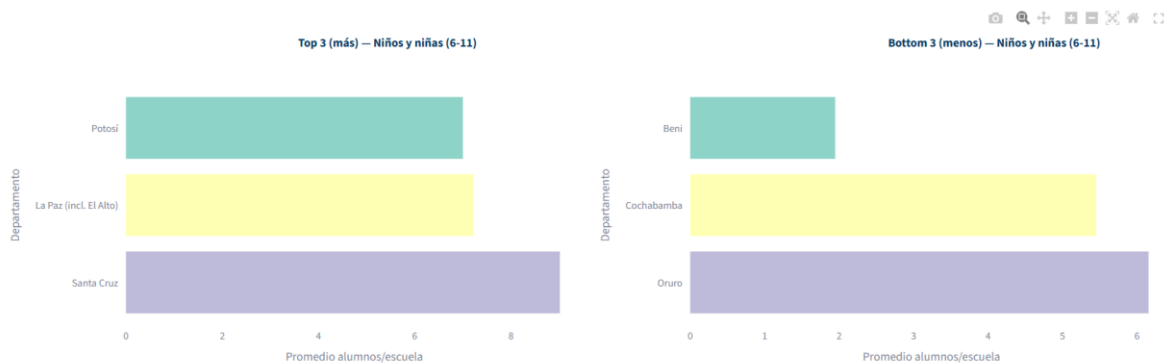
◆ Jóvenes (18)



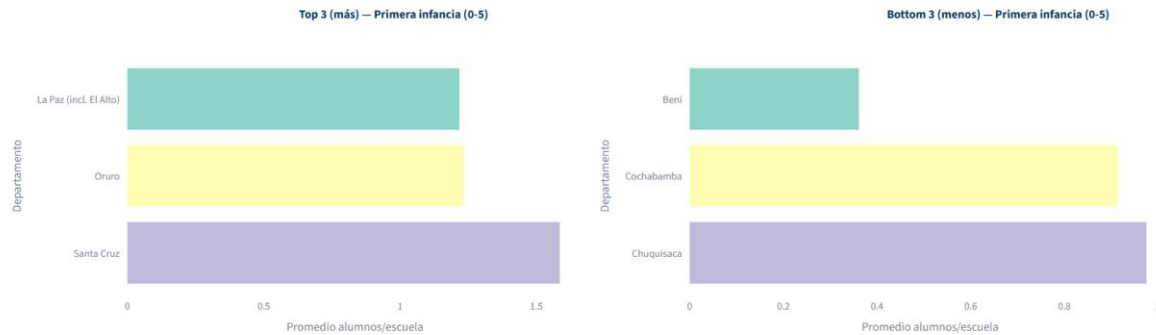
◆ Adolescentes (12-17)



◆ Niños y niñas (6-11)



◆ Primera infancia (0-5)



- **Visualización Clara para la Toma de Decisiones:** El dashboard final permitirá visualizar de inmediato la situación comparativa, responder preguntas como "¿Cómo se compara mi departamento con los demás en términos de presión sobre la infraestructura escolar?" y fundamentar sus solicitudes de inversión o sus planes de expansión con datos sólidos y fácilmente comunicables.

10. Conclusiones

Este proyecto demuestra la viabilidad y el valor estratégico de aplicar un enfoque a un problema crítico de política pública en Bolivia: la planificación macro de la infraestructura educativa. Al optar por un análisis a nivel departamental, se elige un camino pragmático que equilibra la ambición analítica con la factibilidad operativa y la utilidad inmediata para los tomadores de decisiones nacionales. La metodología propuesta permite transformar los extensos y a menudo subutilizados datos censales y administrativos en un producto concreto con un ranking de priorización departamental que ofrece una base objetiva y transparente para debatir y dirigir la inversión pública.

Si bien el modelo no reemplaza los estudios de detalle a nivel local ni considera todos los matices del territorio, cumple una función esencial de diagnóstico comparativo y focalización estratégica. Proporciona una respuesta basada en evidencia a la pregunta: "¿Dónde debe priorizarse la inversión en nuevas escuelas a nivel nacional?" Al hacer visible y cuantificable la desigualdad interdepartamental en la oferta educativa, este proyecto contribuye a promover una gestión pública más racional, equitativa y orientada a resultados. Futuras extensiones de este trabajo podrían escalar el análisis al nivel municipal, integrar datos de movilidad o combinar el índice cuantitativo con evaluaciones cualitativas participativas, construyendo así un sistema de planificación educativa cada vez más robusto e integral para Bolivia.

REFERENCIAS BIBLIOGRÁFICAS

- Anselin, L. (2019). A Local Spatial Statistics Primer. En M. M. Fischer & P. Nijkamp (Eds.), Handbook of Regional Science (pp. 1-24). Springer.
https://doi.org/10.1007/978-3-642-36203-3_7-1
- Batty, M. (2013). The New Science of Cities. MIT Press.
- Chakraborty, A., McMillan, A., & Guhathakurta, S. (2020). Evaluating the role of urban spatial structure in mitigating the education equity gap. Journal of Planning Education and Research, *40*(3), 289-303.
- Goldberg, D. W. (2011). A Geocoding Best Practices Guide. North American Association of Central Cancer Registries.

- Kharrazi, A., Qin, H., & Zhang, Y. (2016). Urban big data and sustainable development goals: Challenges and opportunities. *Sustainability*, *8*(12), 1293. <https://doi.org/10.3390/su8121293>
- Kitchen, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. SAGE Publications.
- Levy, J. M. (2016). *Contemporary Urban Planning* (11th ed.). Routledge.
- Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2015). *Geographic Information Science and Systems* (4th ed.). Wiley.
- Mergel, I., Rethemeyer, R. K., & Isett, K. (2016). Big data in public affairs. *Public Administration Review*, *76*(6), 928-937.
- O'Sullivan, D., & Unwin, D. J. (2010). *Geographic Information Analysis* (2nd ed.). Wiley.
- ReVelle, C. S., & Swain, R. W. (1970). Central facilities location. *Geographical Analysis*, *2*(1), 30-42. <https://doi.org/10.1111/j.1538-4632.1970.tb00142.x>
- Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.
- Talen, E. (1998). Visualizing fairness: Equity maps for planners. *Journal of the American Planning Association*, *64*(1), 22-38.

ANEXOS

Anexo 1

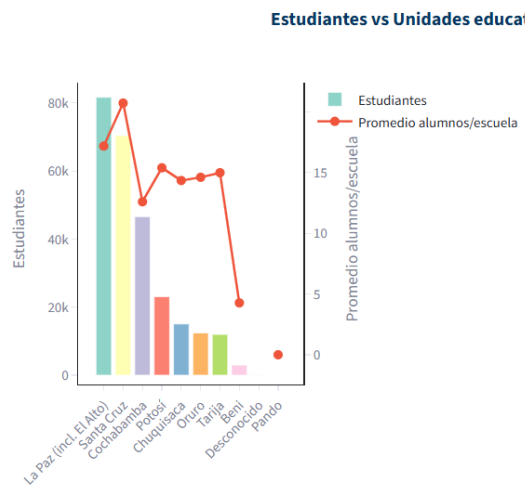
Visualización de datos totales manejados en el análisis



Anexo 3

Visualización de cantidad de estudiantes centros educativos por departamento

Estudiantes vs Unidades educativas



Anexo 4

Visualización de distribución por grupo etario

Distribución por grupo de edad

