

UNIVERSIDAD PRIVADA “FRANZ TAMAYO”

FACULTAD DE INGENIERÍA

CARRERA DE INGENIERÍA DE SISTEMAS



**“ANÁLISIS DE DATOS CENSALES PARA OPORTUNIDADES COMERCIALES
EN BOLIVIA”**

ESTUDIANTE: Gabriela Micaela Durán
Villafán

ASIGNATURA: Big Data

DOCENTE: Ing. Enrique Laurel

LA PAZ – BOLIVIA

2025

1. Introducción

En el contexto del desarrollo económico y empresarial en Bolivia, la toma de decisiones relacionada con la apertura de negocios y la elección del tipo de actividad comercial ha estado tradicionalmente fundamentada en la **intuición, la experiencia personal o la observación empírica del entorno**. Este enfoque, aunque prevalente entre emprendedores y pequeños comerciantes, adolece de una **falta de sustento técnico y analítico**, lo que con frecuencia deriva en **elevadas tasas de fracaso empresarial**. La carencia de información precisa y accesible sobre las condiciones socioeconómicas de las diferentes regiones del país limita la capacidad de los emprendedores para identificar áreas de oportunidad y minimizar los riesgos asociados a la inversión inicial.

En este sentido, el **Censo Nacional de Población y Vivienda**, desarrollado periódicamente por el Estado Plurinacional de Bolivia, constituye una fuente fundamental de información estructurada que permite comprender las **características demográficas, económicas, sociales y habitacionales** de la población. El censo realizado en el año 2012, y el próximo a ejecutarse en 2024, proporcionan datos de alta relevancia para la **formulación de políticas públicas**, pero también representan un **potencial insumo para el sector privado**, especialmente en lo que respecta a la **planificación estratégica y el desarrollo de emprendimientos** basados en evidencia empírica. La riqueza y amplitud de los datos censales ofrecen una oportunidad para transformar la manera en que se **analizan los mercados locales y se determinan las ubicaciones óptimas para diferentes tipos de negocios**.

El avance de las **tecnologías de inteligencia artificial (IA)** y el **análisis de grandes volúmenes de datos (Big Data)** posibilita actualmente el desarrollo de sistemas capaces de **procesar y extraer patrones significativos** a partir de conjuntos de información complejos, como los generados por los censos nacionales. En este contexto, se plantea la creación de un **modelo de inteligencia artificial orientado a la recomendación comercial**, que tenga la capacidad de analizar variables tales como la **ubicación geográfica, la densidad poblacional, los niveles de ingreso, la estructura etaria, la escolaridad y las condiciones de vivienda** de la población boliviana. Mediante el uso de **algoritmos de aprendizaje automático (machine learning)**, este sistema podría identificar correlaciones entre las características socioeconómicas de un área y el tipo de negocio con mayor probabilidad de éxito en la misma.

El propósito central de esta propuesta es **aprovechar los datos censales como un recurso estratégico para la toma de decisiones comerciales informadas**, reduciendo el grado de incertidumbre en el proceso de emprendimiento. A través del uso de la IA, se busca generar **recomendaciones personalizadas y basadas en evidencia**, que orienten a los emprendedores sobre las oportunidades de mercado más viables en función de la capacidad económica y las necesidades de consumo de la población en un determinado sector geográfico. Este enfoque permitirá **democratizar el acceso a información analítica**, tradicionalmente reservada a grandes empresas con capacidad de inversión en estudios de mercado, brindando así una herramienta innovadora y accesible para el **fortalecimiento del ecosistema emprendedor** en Bolivia.

Asimismo, la implementación de un sistema de este tipo podría contribuir a la **planificación territorial y al desarrollo económico local**, al proporcionar insumos para la **identificación de zonas con potencial comercial emergente** y la **distribución equitativa de la actividad económica**. La integración de los datos censales con la inteligencia artificial permitiría, además, **simular escenarios futuros** y proyectar el comportamiento de los mercados ante posibles cambios demográficos o migratorios, fortaleciendo la capacidad predictiva y estratégica de las decisiones empresariales.

En conclusión, el aprovechamiento de los datos del Censo Nacional de Población y Vivienda, complementado con herramientas avanzadas de inteligencia artificial, constituye una **oportunidad innovadora para transformar la toma de decisiones en el ámbito comercial y de emprendimiento en Bolivia**. Esta sinergia entre información estadística y tecnología emergente no solo busca optimizar la localización y tipo de negocios a establecer, sino también promover un modelo de desarrollo económico **más eficiente, inclusivo y sostenible**, donde la información y la ciencia de datos sean pilares para la generación de valor y el impulso de la competitividad nacional.

Relevancia del Análisis de Datos

La analítica de datos aplicada al censo permite:

- **Reducir el riesgo** en inversiones comerciales
- **Promover el desarrollo económico local** mediante la conexión entre oferta y demanda

2. Justificación

Importancia desde la Perspectiva de Big Data

Este problema requiere un enfoque de Big Data porque involucra el procesamiento y análisis de millones de registros censales con múltiples variables interdependientes. Las técnicas tradicionales de análisis resultan insuficientes para:

- Procesar eficientemente el volumen de datos censales
- Identificar patrones complejos entre múltiples variables
- Generar visualizaciones interactivas que permitan exploración intuitiva

Justificación Técnica:

- Demostración práctica de aplicaciones de Big Data para problemas reales
- Desarrollo de habilidades en procesamiento de datos masivos
- Creación de metodologías replicables para análisis comerciales

Justificación social:

- Empoderamiento de emprendedores con información accesible
- Reducción de la tasa de fracaso de pequeños negocios
- Fomento del desarrollo económico local

Justificación Económica:

- Optimización de inversiones comerciales
- Identificación de oportunidades de mercado no explotadas
- Mejora en la competitividad del sector comercial

3. Planteamiento del Problema

Descripción del Problema

Actualmente, emprendedores, pequeñas empresas y inversionistas en Bolivia carecen de herramientas accesibles y basadas en datos para identificar:

- Dónde ubicar sus establecimientos comerciales
- Qué tipo de negocios tienen mayor potencial en diferentes zonas
- Qué características debe tener su oferta comercial para satisfacer la demanda local

Qué se Desconoce

- Patrones de compatibilidad entre características poblacionales y tipos de negocios exitosos
- Distribución geográfica de segmentos de consumidores específicos
- "Huecos de mercado" en diferentes regiones urbanas y rurales
- Correlación entre indicadores socioeconómicos y hábitos de consumo

Decisiones que se Beneficiarán

- **Emprendedores:** Elección de ubicación y tipo de negocio
- **Empresas establecidas:** Decisiones de expansión y apertura de nuevas sucursales
- **Gobiernos municipales:** Planificación de desarrollo comercial zonificado
- **Inversionistas:** Identificación de sectores y regiones con potencial de crecimiento

4. Objetivos de Investigación

Objetivo General

Desarrollar un sistema de análisis y visualización de datos censales que identifique oportunidades comerciales y ubicaciones óptimas para diferentes tipos de negocios en Bolivia.

Objetivos Específicos

- Recopilar datos poblacionales para el uso de estos en análisis.
- Analizar los datos recopilados para generar graficas y realizar un estudio de los mejores sectores en los que se puede iniciar un negocio
- Generar un modelo de inteligencia artificial que permita dar recomendaciones de acuerdo al sector en la ciudad.

5. Fuentes de Datos

Datos Primarios

- **Fuente:** Instituto Nacional de Estadística (INE) - Censo Nacional de Población y Vivienda 2024
- **Tipo:** Datos estructurados en formato tabular
- **Volumen estimado:** 10 millones de registros (población boliviana) × 80+ variables
- **Características:** Datos anonimizados a nivel individual, agregables por manzana, barrio, distrito, municipio

Variables Principales a Utilizar

- **Demográficas:** _AGE, _SEX, _URBAN
- **Socioeconómicas:** _EDATTAIN, _WALL, _ROOF, _FLOOR
- **Tecnología y Bienes:** _COMPUTER, _INTERNET, _AUTO, _MOTORCYC
- **Servicios:** _WATSRG, _TOILET, _SEWAGE, _ELECTRIC
- **Movilidad:** _TRANS
- **Hogar:** _ROOMS, _BEDROOMS, _OWNERSHIP

6. Cronograma Estimado

Fase 1: Preparación de Datos (3-4 semanas)

- Descarga y carga de datos censales
- Limpieza y estandarización de variables
- Geocodificación y preparación de datos espaciales
- Creación de índices compuestos (poder adquisitivo, nivel educativo)

Fase 2: Análisis Exploratorio (2-3 semanas)

- Estadísticas descriptivas por región
- Análisis de correlaciones entre variables
- Identificación de patrones demográficos
- Segmentación inicial de mercados

Fase 3: Desarrollo de Modelos (3-4 semanas)

- Creación de perfiles de compatibilidad negocio-población
- Desarrollo de algoritmos de scoring por tipo de negocio
- Validación cruzada con datos existentes
- Optimización de parámetros

Fase 4: Visualización y Dashboard (2-3 semanas)

- Diseño de interfaz de usuario
- Desarrollo de mapas interactivos
- Implementación de filtros y búsquedas
- Pruebas de usabilidad

Fase 5: Documentación y Validación (1-2 semanas)

- Documentación técnica y de usuario
- Casos de estudio de validación
- Preparación de informe final
- Presentación de resultados

7. Relación con las 5V del Big Data

Volumen

- **Datos:** ~5 millones de registros × 80+ variables
- **Procesamiento:** Requiere técnicas de procesamiento distribuido
- **Almacenamiento:** Estimado 2-5 GB en formato comprimido

Velocidad

- **Actualización:** Datos estáticos (censo decenal)
- **Procesamiento:** Análisis por lotes
- **Respuesta:** Visualización de diagramas una vez procesados los datos

Variedad

- **Estructurados:** Datos tabulares del censo (95%)
- **Semi-estructurados:** Metadatos y diccionarios de variables
- **Geográficos:** Coordenadas y delimitaciones territoriales
- **Textuales:** Descripciones de variables y categorías

Veracidad

- **Origen:** Datos oficiales del INE, máxima confiabilidad institucional
- **Calidad:** Alta cobertura y estandarización metodológica
- **Limitations:** Posibles sesgos de no respuesta, requiere validación cruzada

Valor

- **Directo:** Herramienta para toma de decisiones comerciales informadas
- **Estratégico:** Reducción de riesgo en inversiones
- **Social:** Promoción del desarrollo económico local
- **Académico:** Metodología replicable para análisis similares

