**Alcances del fenómeno de la población habitante de calle desde una perspectiva social y de salud pública en el territorio de Medellín.**

**Resumen Descriptivo del Proyecto**

Debido al gran reto social y sanitario que representa la situación de los habitantes de calle en la ciudad de Medellín, este proyecto busca abordar esta realidad desde un enfoque investigativo con el fin de proporcionar información valiosa para afrontar las raíces de esta problemática. Mediante métodos de clasificación y segmentación que permitan predecir patrones de comportamiento y la probabilidad de que una persona termine en situación de calle con base en características y factores que determinan esta condición, y así contribuir a la toma de decisiones con respecto a las necesidades que se visibilizan dentro de este grupo poblacional.

**1. Problema.**

**a.** ¿Cuál es mi problema alrededor de mi proyecto?

La problemática de los habitantes de calle constituye un desafío tanto a nivel local como nacional y global, evidenciando una creciente preocupación a lo largo del tiempo. Este fenómeno se ve moldeado por una serie de factores estructurales y sociales que influyen en la presencia transitoria o permanente de individuos en las calles. Se evidencia una interrelación entre la falta de vivienda asequible, problemas de salud mental, abuso de sustancias, desempleo y violencia familiar, entre otros factores, todos ellos relacionados con esta problemática social y de salud pública. Es imperativo investigar y comprender las causas subyacentes de este fenómeno para proporcionar información y resultados impactantes que permitan abordar sus raíces desde una perspectiva de salud pública.

**b.** ¿Por qué se manifiesta el problema?

El problema se manifiesta por dinámicas de marginalidad, exclusión social, consumo de sustancias psicoactivas, limitado acceso a servicios de salud y disfunción familiar. Los habitantes de calle son el resultado de estas variantes, bien sea porque el sujeto decida elegir ese tipo de vida o porque la sociedad lo empuje a esa condición. En tal sentido, la exclusión social alienta la problemática de habitantes de calle, en cuanto son "los excluidos aquellos sujetos que han perdido su filiación con la población y no pueden integrarse y, por tanto, se sitúan fuera de las pautas de producción y consumo comúnmente admitidas por la sociedad"1.

**c.** ¿De quién es el problema?  
Es un problema que involucra a todos los ciudadanos, sin embargo es considerable el papel que deben desempeñar los entes territoriales para intervenir la actual situación, por lo tanto se consideraría el mayor responsable de esta problemática.

**2. Datos.**

**a.** Describa la fuente de datos (Sea más específico que en documento pasado)

Se tomarán los microdatos abiertos recolectados en el año 2019 disponibles en la plataforma del DANE (Departamento Administrativo Nacional de Estadística). El Censo de Habitantes de Calle, se realizó con el objetivo de contar con información actualizada y confiable sobre el volumen y las principales características socioeconómicas y demográficas de los habitantes de la calle ubicados en algunas de las principales ciudades del territorio nacional, para la formulación, gestión y evaluación de las políticas, planes y programas de las administraciones municipales, departamentales y nacionales, respecto a la atención integral, rehabilitación e inclusión social de dicha población 2.

**b.** ¿Cuál es la calidad que posee dicha fuente?

Los datos y la información en general que aparecen en el portal se han introducido siguiendo estrictos procedimientos de control de calidad, de acuerdo con el Decreto 262 de 2004, cuyo objetivo es garantizar la producción, la disponibilidad y la calidad de la información estadística estratégica. Para el caso del dataset de Habitantes de Calle 2019, el DANE cuenta con un diseño de métodos y mecanismos que buscan garantizar la calidad de la información respondiendo a los lineamientos de toda investigación: ¿Qué medir?, ¿Cómo medir?, ¿Dónde medir? y ¿Cuándo?. No obstante, el DANE no se responsabiliza por el uso e interpretación realizado por terceros 2.

**c.** ¿Es accesible y disponible?  
Si, en este caso particular del tema de Habitantes de Calle el acceso a los microdatos anonimizados es de uso público y de carácter gratuito. Se encuentra disponible en la página Web del DANE 2.

**3. Hipótesis.**

**a.** ¿Qué es lo que pretende probar?

Predecir patrones de comportamiento y la probabilidad de que una persona termine en situación de calle con base en características y factores que determinan esta condición.

**b.** ¿Cuáles pueden ser las respuestas esperadas a lo que pretende probar?

* Encontrar públicos objetivos y puntuales que puedan ser identificados para plantear y/o proponer posibles planes de acción como: programas de sensibilización, atención médica oportuna, oportunidades de reinserción en sociedad, programas de psicología, entre otros.
* Revelar patrones de comportamiento de la población en estudio.
* Visibilizar las necesidades específicas de la población habitante de calle.
* Evidenciar las probabilidades que tiene una persona de terminar en situación de calle.
* Contribuir al planteamiento y formulación de las políticas públicas orientadas a estos grupos poblacionales.

**c.** ¿Qué acciones debo tomar para cada respuesta anterior?

* Identificar patrones subyacentes específicos para grupos con características y condiciones de vida similares, que permitan presentarle un programa según sus atributos particulares del grupo al que pertenece; siendo objetivos en las necesidades más urgentes: médicas, psicológicas, de acompañamiento, entre otras.
* Analizar la población de estudio para determinar tendencias en ella.
* Modelar la probabilidad de que un individuo termine en situación de calle.

**4. Solución.**

**a.** Tipo. (¿Qué tipo de algoritmo de ML es? ¿Qué técnica es?

Análisis multivariado

* Análisis de componentes principales (ACP), para reducir la dimensionalidad de las variables.
* Análisis de correspondencia múltiple (MCA) para identificar la relación entre las categorías de las variables cualitativas.

Modelo supervisado

* La información se va analizar mediante la metodología de Modelo de regresión logística para determinar la probabilidad de que un individuo pueda terminar en situación de calle.

Modelo no supervisado

* Modelo de clasificación no supervisado K-Means (k-means), que nos permita identificar grupos de individuos con comportamientos similares, para posteriormente proponer los planes de acción según las necesidades específicas de cada grupo.

**b.** ¿Cómo debe ser implementado?

* Análisis de componentes principales (ACP)

Preparar los datos: desde la limpieza hasta aplicar los formatos adecuados, verificar valores faltantes y eliminar columnas que no sean relevantes para el análisis de ACP.

Normalizar los datos: escalar los datos si es necesario con StandardScaler de scikit-learn

Aplicar ACP: se importa PCA de sklearn.decomposition para ajustar el modelo a los datos y transformarlos en el espacio de componentes principales.

Análisis de resultados: examinar los componentes principales resultantes para entender cuales son las principales características y como explican la varianza de los datos.

* Modelo de regresión logística

Preparar los datos: desde la limpieza hasta aplicar los formatos adecuados, verificar valores faltantes, codificar las variables categóricas y separar las características de la variable a predecir (objetivo).

Dividir los datos: Dividir los datos en conjuntos de entrenamiento y prueba para poder evaluar el rendimiento del modelo. Se utilizará la función train\_test\_split de scikit-learn.

Crear y entrenar el modelo: Importar la clase LogisticRegression de scikit-learn y crear una instancia del modelo. Luego, utilizar el método fit para entrenar el modelo de los datos de entrenamiento.

Evaluar el modelo: Una vez entrenado, utilizar los datos de prueba para evaluar el rendimiento del modelo. Usar métricas como precisión, recall, y el área bajo la curva ROC para evaluar la calidad de las predicciones.

* Análisis de correspondencia múltiple

Preparar los datos: desde la limpieza hasta aplicar los formatos adecuados, verificar valores faltantes y codificar las variables categóricas.

Se utiliza pip install prince de la biblioteca Prince

Crear y ajustar el modelo de análisis de correspondencia: Utilizar la clase MCA de la biblioteca Prince para crear una instancia del modelo y ajustarlo a los datos.

Visualizar los resultados: Explora los resultados del análisis de correspondencia para entender las relaciones entre las categorías de las variables cualitativas.

* Modelo de clasificación no supervisado K-Means

Preparar los datos: desde la limpieza hasta aplicar los formatos adecuados, verificar valores faltantes y codificar las variables categóricas.

Normalizar los datos: normalizar los datos antes de aplicar K-means para que todas las características tengan la misma escala. Se puede utilizar StandardScaler de scikit-learn.

Crear y ajustar el modelo K-means: Importar la clase KMeans de scikit-learn y crear una instancia del modelo. Luego, utilizar el método fit para ajustar el modelo a los datos.

Asignar clusters: Después de ajustar el modelo, utilizar el método predict para asignar cada instancia de los datos a un cluster.

Explorar y analizar los clusters: Analizar los clusters resultantes para entender las características de cada grupo y cómo se relacionan con los datos originales.

**c.** ¿Qué resultados se esperan?

* Probabilidades de que una persona termine en situación de calle con base en características demográficas, socioeconómicas, de salud mental, historial familiar, y demás factores asociados.
* Patrones de comportamiento de la población en estudio.
* Necesidades futuras de los habitantes de calle identificadas.
* Finalmente se espera que los resultados de esta investigación sean tenidos en cuenta para contribuir en el planteamiento y formulación de las políticas públicas orientadas a la población en situación de calle.

**5. KPI’s**

**a.** ¿Cómo espera evaluar el modelo?

* Para el análisis de componentes principales, se determina la reducción de variables mediante la sumatoria de la variabilidad explicada por estas variables en las principales dimensiones que recomienda el análisis.
* Para la regresión logística, se evaluará mediante el uso de metodologías de Matriz de confusión y curva ROC, las cuales dan un alcance de precisión, sensibilidad y determinan la efectividad de esta modelación.
* Para el análisis de correspondencia: se evalúa por la sumatoria de la variabilidad explicada de las dimensiones.
* Para el modelo de clasificación no supervisado K-Means, se evalúa mediante la elección de número grupo bajo la metodología del Codo, quien nos dice gráficamente cuál es el número óptimo de grupos en el que se pueden desagregar la población de estudio.

**b.** ¿Qué métricas debe usar?

* Análisis de componentes principales:
  + la varianza explicada acumulativa, lo ideal es que represente una proporción alta (>70%) con un número mínimo de componentes.
* Regresión logística:
  + la precisión (accuracy)
  + la sensibilidad (recall)
  + la especificidad
  + el valor F1
  + el área bajo la curva ROC (AUC)
* Análisis de correspondencia:
  + la proporción de inercia explicada
* Modelo de clasificación no supervisado K-Means:
  + la inercia intra-cluster, que representa la dispersión de los datos dentro de cada grupo
  + el punto de inflexión en la gráfica de la inercia intra-cluster (método del codo)

**6. Actores.**

**a.** ¿Quién es su cliente?

Gobierno local (Alcaldía municipal en el estudio)

**b.** ¿Quiénes son los interesados (Stakeholders)?

Lo entes territoriales interesados en transformar realidades sociales

**c.** ¿Quién usará la solución?

Se espera presentar a organismos gubernamentales y grupos comunitarios dedicados al bienestar de la población objeto de estudio. En la que se puedan integrar planes de acción y programas de intervención, según sean las políticas internas de cada organización. De igual manera se espera que el estudio esté publicado y disponible al público para personas que buscan más información sobre los individuos en situación de calle, permitiendo así tener una mayor difusión de los análisis que ayuden a entender un poco más esta población.

**d.** ¿A quién impactará?

El impacto de los análisis realizados en este proyecto, va más allá del grupo de riesgo descrito, va dirigido a mejorar las condiciones de nuestra comunidad, al proponer mejoras y planes que impacten a grupos muy específicos de la población de estudio, ayudando notablemente a que los múltiples esfuerzos, se concentren en cada subgrupo de la población objeto.

**REFERENCIAS**

1. Alfonso, Ó., Barrera, R., Bernal, P., Camargo, D., & Garzón, L. (2019, mayo). Vista de El ciclo mortal de los habitantes de calle en Bogotá. Teorías, olvidos, políticas y desenlaces fatales. Revistas Universidad Externado de Colombia, https://revistas.uexternado.edu.co/index.php/ecoins/article/view/5970/7933

2. Colombia - Censo de Habitantes de Calle - CHC- 2019. (2020, marzo). Microdatos - DANE. Recuperado 25 de abril de 2024, de https://microdatos.dane.gov.co/index.php/catalog/663/study-description