

Resumen técnico del análisis de género en nomenclátor urbano

Este proyecto consistió en el análisis y visualización de la representación de género en los nombres de calles de la Ciudad, basado en datos geoespaciales extraídos de OpenStreetMap.

1. Extracción y preparación de datos:

Se realizó un raspado sistemático de la base de datos de OpenStreetMap para obtener los nombres de las calles junto con sus respectivas geometrías en formato WKT. Los datos fueron almacenados en archivos CSV para su posterior procesamiento.

2. Limpieza y extracción de nombres relevantes:

Se implementó una función para filtrar y extraer posibles nombres propios dentro del texto del nombre de la calle, descartando términos genéricos y geográficos que no aportan a la clasificación (e.g., "calle", "avenida", "plaza").

3. Clasificación de género mediante aprendizaje automático:

Se desarrolló un modelo supervisado de clasificación utilizando el algoritmo Multinomial Naive Bayes, entrenado con una base extensa de nombres personales (en español y Kichwa) etiquetados con su género correspondiente. El modelo fue aplicado a los nombres extraídos de las calles para predecir el género asociado, utilizando técnicas de procesamiento de texto basadas en n-gramas de caracteres para capturar patrones característicos de los nombres.

4. Conversión y estructuración geoespacial:

Los resultados, que incluyen el nombre de la calle, su geometría y la clasificación de género, fueron convertidos a formato GeoJSON para facilitar su integración en sistemas de información geográfica y plataformas de análisis.

5. Visualización y análisis en plataformas especializadas:

Los datos en formato GeoJSON fueron cargados en servicios de mapas de Azure y conectados a Power BI para la creación de dashboards interactivos. Esto permitió analizar la distribución espacial y la representatividad de género en el nomenclátor urbano, facilitando la toma de decisiones informadas y promoviendo la equidad en el espacio público.