

Bogotá D.C., septiembre 10 de 2023

Entrega 1. Documento de propuesta de análisis

Integrante: Danilo Alfonso Hernández Montenegro

¿Cuál es la temática de mayor relevancia que debería tenerse en cuenta en Bogotá actualmente? ¿Por qué?

R/. Indudablemente, la temática que mayor relevancia debería tener actualmente en Bogotá es la “Movilidad sustentable”, porque es un modelo de traslado de bajo consumo de carbono que además de ser saludable, privilegia el elevar la calidad de vida urbana y el bienestar colectivo, así como la creación espacios públicos confortables que favorezcan la convivencia ciudadana.

2. ¿A qué ODS se vincula esta temática? ¿Cómo lo hace?

R/. Esta temática se vincula con la ODS 11. Ciudades y comunidades sostenibles, y lo hace de la siguiente manera:

2.1. En Bogotá D.C., El POT y las políticas de movilidad se han orientado progresivamente como una hoja de ruta para promover el ODS 11. Ciudades y comunidades sostenibles.

2.2. Bogotá por conducto del POT y las políticas del sector movilidad le apuesta cada vez más a la promoción de una movilidad activa (o no motorizada) como pieza fundamental del cumplimiento del ODS 11.

2.3. La promoción de los viajes a pie, del uso de la bicicleta (de pedaleo mecánico o asistido) y de vehículos de micro movilidad, hacen parte de la visión de ordenamiento territorial para la ciudad, mediante intervenciones en infraestructura, en mejorar la proximidad de la oferta de bienes y servicios y en la promoción cultural para reconocer que todos somos peatones y necesitamos una ciudad amigable y accesible para poder movernos de manera autónoma, segura y cómoda.

2.4. La ciudad viene avanzando en el desarrollo de obras de infraestructura e inversiones que permiten el cambio modal del transporte hacia fuentes de energía limpias y en la implementación de un ordenamiento del territorio que propenda por la reducción de las distancias espaciales entre la oferta y la demanda de bienes y servicios, para de esta forma reducir las barreras de localización garantizando que los ciudadanos accedan más

fácilmente a la oferta, reduciendo la presión sobre los medios de transporte público y privado convencionales.

2.5. La Primera Línea del Metro prevé la implementación a gran escala de infraestructura asociada al uso de la bicicleta mediante cicloparqueaderos de concentración que atenderán una alta demanda y estarán ubicados dentro de las estaciones, en el sótano o en el nivel de acceso, de acuerdo con la tipología de la estación.

3. Realice un análisis de la temática escogida con base con los datos abiertos explorados

Recopilación y exploración de datos:

A partir del acceso a la plataforma Datos Abiertos Bogotá (<https://datosabiertos.bogota.gov.co/>) se hizo una búsqueda exhaustiva de conjuntos de datos relacionados con la movilidad sustentable, se procedió a explorar los datos disponibles, identificando las variables relevantes y comprendiendo cómo se estructuran y relacionan entre sí.

Análisis de datos:

A continuación, se muestra el análisis realizado para uno de los Dataset (Intervención Malla Vial) utilizados y que cuenta con información de la Malla Vial, que se componen de variables del tipo discreto, continuo y categóricas y este conjunto de datos brinda 12.881 registros de Código de Identificación Vial-CIV y posee información de cada “CIV”, como se describe en las características del Dataset. La tarea principal consiste la implementación de algoritmos de aprendizaje supervisado- predecir respuesta si el CIV requiere intervención. La variable respuesta “Intervención” representa si un CIV requiere intervención. Si podemos usar la vida de uso de cada CIV para predecir otros CIV que requieran algún tipo de intervención, se pueden tomar mejores decisiones por parte de las entidades del Distrito Capital, responsables de realizar las intervenciones a la malla vial. Dentro de las variables importantes para la implementación de los modelos se logra identificar el tipo de superficie, la cantidad de carriles, la fecha de la última intervención, el tipo de calzada, la velocidad permitida y el estado superficial.

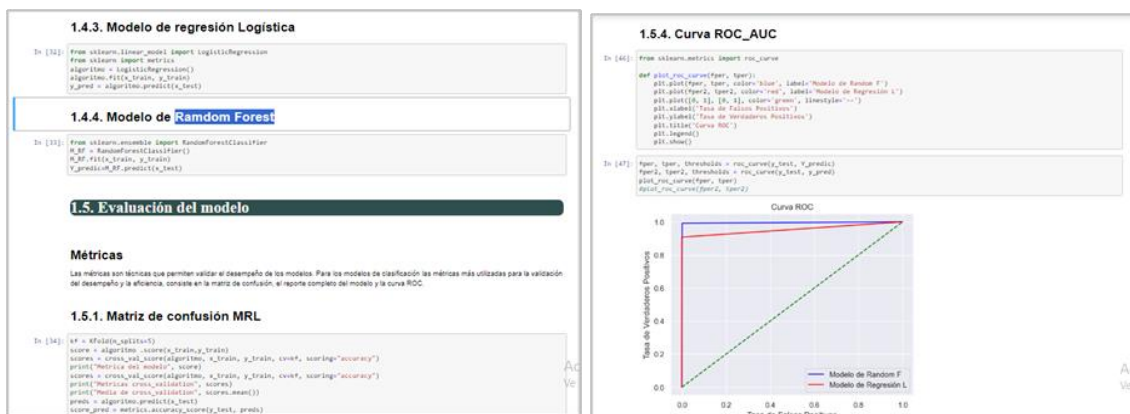
El dataset inicialmente pasa por varios procesos como se menciona a continuación: Data cleaning, Procesamiento, Escalamiento o normalización de las variables, División de los datos para entrenamiento y prueba de los modelos, Implementación del modelo Random Forest, Implementación del modelo Regresión logística, Evaluación de métricas de eficiencia. Las herramientas utilizadas para el análisis de datos son: Python, Jupyter Notebook y librerías (Pandas, Numpy, Matplotlib, SkLearn), para examinar y extraer información útil de los datos.

Análisis gráfico del dataset y análisis descriptivo: Comprensión de tendencias, patrones o relaciones existentes en los datos. El análisis gráfico de los datos permite validar el comportamiento del Dataset, y nos aporta valor y robustez a los análisis preliminares que se realizan para la gestión de los modelos. Una de las bondades de esta técnica de análisis estadístico, consiste en evidenciar la distribución de los datos, si se presentan sesgos, si nos enfrentamos a datos desbalanceados, datos atípicos, el comportamiento de la curtosis y el comportamiento de la variable respuesta. **Para este caso de estudio, y utilizando una gráfica tipo pie, podemos evidenciar que el 41% de la población de la muestra requiere intervención en el pavimento. Otro de los reportes más interesantes es el ProfileReport, este nos genera un reporte completo de todas las variables, comportamiento, correlación e interacción entre variables, consiste en una visión más holística del Dataset, proporcionando información valiosa para determinar si nuestro Dataset se encuentra lo suficientemente estructurado para continuar con la fase de implementación de los modelos.**



Modelos: Regresión logística, Random Forest. Evaluación de los modelos. **Para caso de estudio, la métrica en Test es 0.9615832363213038 = 96%.**

En la etapa final del ciclo de modelado, se utilizaron métricas de evaluación para medir el rendimiento del modelo. Las métricas comunes incluyen la precisión, el recall, la F1-score y la curva ROC. Estas métricas permiten comparar diferentes modelos y seleccionar el que tenga el mejor rendimiento para el problema en cuestión. **Si bien los dos modelos implementados cuentan con una alta eficiencia, luego de analizar la gráfica de la curva ROC, se logra identificar que el modelo de Random_Forest tiene una mayor área de cobertura para la clasificación de datos, lo que nos indica que su desempeño es mayor que el modelo de regresión logística.**



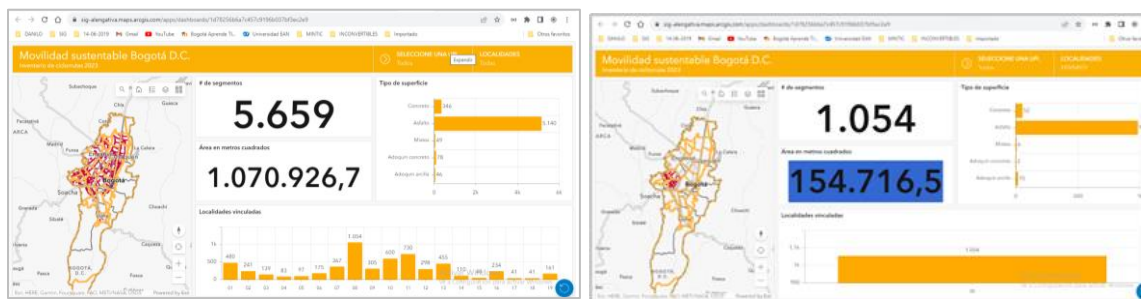
Análisis espacial

Se utilizaron las siguientes herramientas de análisis de datos espaciales: ArcGIS Pro, ArcGIS Online y herramientas de visualización (Dashboard ArcGIS, Power BI)

Los datos seleccionados contienen información geográfica, como la ubicación de estaciones de bicicletas, paradas de transporte público, Localidades, Unidades de Planeamiento Local-UPL, ciclorrutas, cicloparqueaderos, cámaras salvavidas, Rutas zonales SITP, Ciclo alameda POT Bogotá D.C., Paraderos Zonales del SITP, Malla vial intervenida, estado de la malla vial, estaciones calidad del aire. Bogotá D.C., entre otros.

A continuación, se muestra el análisis realizado para uno de los Dataset (Ciclorrutas) utilizados y que cuenta con información del inventario de la infraestructura de ciclorrutas y este conjunto de datos brinda 5659 registros de Código de Identificación Vial-CIV

(segmentos de ciclorruta) y posee información de cada “CIV”, como se describe en las características del Dataset. La tarea principal consiste en analizar cómo se distribuyen las ciclorrutas (# de segmentos y área en metros cuadrados) por Localidades, por UPL, (opcional por UPZ), y cuales son los tipos de superficie. Para el correspondiente análisis, inicialmente esta capa geográfica solo tenía dos variables (Tipo de superficie, Código de Identificación Vial-CIV), por lo que se hizo necesario agregar nuevas variables como lo son: Localidad, UPL, UPZ, área en metros cuadrados. Estas variables se obtuvieron al realizar el geoprocresamiento en el software ArcGIS Pro.



La Localidad de Kennedy, es la localidad con mayor número de segmentos de ciclorruta (1.054) equivalente al 18,6%, con un área de 154.716,5 metros cuadrados, equivalente al 14,5%, y principalmente con tipo de superficie en asfalto que corresponde al 92,9 %. Contrariamente, las localidades La Candelaria y Rafael Uribe Uribe, son las que menor número de segmentos de ciclorruta tienen , 41 para cada una, equivalente al 0,7%, con área en metros cuadrados de 8.077,9 y 11.642,3 equivalente a un 0,7% y 1% respectivamente.

4. ¿Qué propuesta de acciones de mejoramiento frente a la misma se pueden plantear?

Con base en los resultados de los modelos, puede identificarse la malla vial que requiere priorizarse para que sea intervenida. Respecto a los análisis de datos espaciales se ha logrado identificar que localidades ameritan mayor infraestructura de ciclorrutas.

El análisis muestra que la infraestructura de ciclorrutas es insuficiente, por lo que se propone aumentar el área disponible para la mayoría de las localidades, aumentando la cantidad de carriles exclusivos para bicicletas o mejorar la conectividad entre las rutas existentes.

Referencias:

[1]. Exploradatos 3.0 - Google drive. (s/f). Google.com. Recuperado el 13 de agosto de 2023, de

<https://drive.google.com/drive/folders/1U5xWJrkkgB0ejA5DSCsYS7TWLrkWdPbb>

[2]. de la Megalópolis, C. A. (s/f). ¿Qué es la movilidad sustentable? gob.mx. Recuperado el 13 de agosto de 2023, de <https://www.gob.mx/comisionambiental/articulos/que-es-la-movilidad-sustentable?idiom>

[3]. Infraestructura de Datos Espaciales de Bogotá. (s/f). Ideca. Recuperado el 13 de agosto de 2023, de <https://www.ideca.gov.co/recursos/mapas/mapa-de-referencia-para-bogota-dc>

[4]. IDECA. (s/f). Mapas Bogotá. Gov.co. Recuperado el 13 de agosto de 2023, de <https://mapas.bogota.gov.co/>

[5]. (S/f). Gov.co. Recuperado el 13 de agosto de 2023, de <https://observatorio.movilidadbogota.gov.co/>

Pantallazo re social Facebook hashtag #ExploraDatosBogotá

