

Minería de Datos

Proyecto Final

Ivan Saavedra, Ph.D.

saavedrai@uninorte.edu.co

Universidad del Norte
División de Ingenierías
Dpto. Ingeniería de Sistemas



202110

1. Running Data Analysis

Metodología

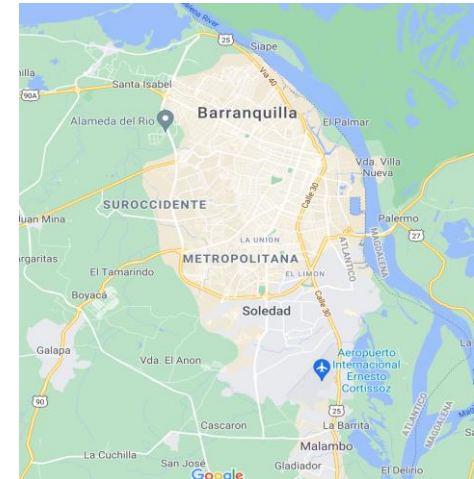
- **Búsqueda de datos históricos:** Datos históricos de una persona en particular han sido tomados de la aplicación wearable de <https://runkeeper.com/cms/>, los datos son suministrados en este ejercicio para su respectivo análisis.
- **Organización de los datos:** De ser necesario, usted deberá organizar los datos de manera apropiada para el desarrollo del análisis.
- **Análisis de los datos:** Usted necesitara desarrollar un análisis de probabilidad para responder a las preguntas objetivo.
 - Identifique que distribución de probabilidad se ajusta a la variable Distancia (km) ?
 - Cual es la probabilidad de correr una distancia de 10 km?
 - Cual es la probabilidad de correr una distancia mayor a de 15 km?
- **Visualización de los datos:** Usted necesitara crear las visualizaciones para dar soporte a su análisis.
 - Valor esperado vs valor observado
 - Diagrama de cajas
 - Otro?
- **Conclusiones:** Usted necesitara resaltar las conclusiones al final de sus análisis.

2. COVID-19 Mobil Testing/Vaccinations Sites

Usted ha sido asignado con la tarea de determinar cuales serian las ubicaciones optimas de nuevos puestos de vacunación móvil en la ciudad de Barranquilla para atender a la ciudadanía en barrios mas afectados por casos de COVID-19. Para ello, usted necesita encontrar los históricos del numero de casos nuevos/activos por barrio.

Metodología

- Encontrar la información necesaria
- Preparar el dataset a usar
- Realizar exploración estadística y visual analytics
- Seleccione la metodología y ejecute el desarrollo de la misma
- Sustente sus conclusiones y sugerencias
 - Muestre la precisión de su modelo
 - Realice map vizualizations



Source: <https://www.kpbs.org/news/2021/jan/12/san-diego-county-3524-new-covid-19-infections/>

3. Transit accidents severity

Como segunda asignación, usted debe crear un modelo que permita la categorización de la severidad de accidentes utilizando como base información de accidentabilidad de Estados Unidos para realizar una campaña de precauciones dependiendo de los atributos seleccionados.

(https://www.kaggle.com/sobhanmoosavi/us-accidents?select=US_Accidents_June20.csv)

Metodología

- Preparar el dataset a usar
- Seleccionar el conjunto de atributos a utilizar
- Realizar exploración estadística
- Seleccione la metodología y ejecute el desarrollo de la misma
 - Teniendo en cuenta los siguientes atributos: Temperatura, Humedad, Presión, Visibilidad, Precipitación, Condiciones climática
- Sustente sus conclusiones y sugerencias



Source: <https://www.pexels.com/photo/blur-car-caution-dash-163945/>

Entregables

- Archivos de Excel con los datos de entrada
- Un archivo de Jupyter Notebook con los desarrollo de los modelos
- Se sugiere que comente las secciones de manera adecuada para una mejor interpretación de su desarrollo
- Resumen ejecutivo de las tres asignaciones en PDF.
- La fecha de entrega es el día Miércoles 2 de Junio de 2021 vía catalogo web enlace de laboratorios.