

# Minería de Datos

## Lab. #3

Ivan Saavedra, Ph.D.

[saavedrai@uninorte.edu.co](mailto:saavedrai@uninorte.edu.co)

Universidad del Norte  
División de Ingenierías  
Dpto. Ingeniería de Sistemas



202030

# Análisis de Distribuciones de Probabilidad

**Usted ha sido asignado para desarrollar un análisis exploratorio de probabilidad en dos conjuntos de datos históricos correspondientes a:**

- **la accidentalidad vehicular en la ciudad de Barranquilla, y**
- **al historial de los entrenamientos de un corredor.**

Se espera que usted ponga en practica los temas distribución de probabilidad tratados en clase para responder a las preguntas objetivo.

# 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. Vehicle Accident Analysis

### Metodología

- **Búsqueda de datos históricos:** Datos históricos de la accidentalidad en la ciudad de Barranquilla han sido descargados de la siguiente fuente :<https://www.datos.gov.co/en/Transporte/Accidentalidad-en-Barranquilla/yb9r-2dsi>. 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.
  - Realice la organización de los datos dentro de su notebook.
- **Análisis de los datos:** Usted necesitara desarrollar un análisis de probabilidad para responder a las preguntas objetivo.
  1.
    - Identifique que distribución de probabilidad se ajusta a la variable Cantidad de Accidentes Diarios?
    - Cual es la probabilidad de que se den menos de 10 Accidentes?
    - Cual es la probabilidad de que se presenten mas 15 accidentes?
  - 2
    - Analice que distribución de probabilidad se ajustaría a la variable Cantidad de Accidentes En Un Intervalo De Dos Horas Dentro del Horario de 6am a 6pm.
- **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.

# Análisis de Distribuciones de Probabilidad

## Entregable:

- Un archivo de Jupyter Notebook con el desarrollo del análisis
- Las conclusiones y respuestas al objetivo del análisis deben ser contestadas en el mismo notebook.
- Se sugiere que comente las secciones de manera adecuada para una mejor interpretación de su análisis.
- La fecha de entrega es el Viernes 04 de Septiembre del 2020 vía catalogo web enlace de laboratorios.