

# Actividad 3

Gocobachi Lemus Jesus Abel

25 de enero de 2021

## Parte I

### Introduccion

Una nueva actividad, ahora con la ayuda de una nueva biblioteca, la cual es pandas, la cual entre otras cosas, nos será de gran utilidad para el tratamiento de datos, en nuestro caso vamos a reutilizar los mismos con los cuales trabajamos en una lejana actividad 1. Pero ahora con un mayor análisis de los mismos, esto será un gran aprendizaje para aprender a manejar una gran cantidad de datos, poder crear copias de los mismos y editarlos sin ningún problema de alterar los originales.

## Parte II

### Bibliotecas utilizadas

#### 1. Biblioteca de pandas

Pandas son las siglas de "Python Data Analysis Library". Según la página de Wikipedia sobre Pandas, "el nombre se deriva del término "datos de panel", un término econométrico para conjuntos de datos estructurados multidimensionales". ¡Pero creo que es solo un lindo nombre para una biblioteca de Python súper útil!

Pandas es todo un cambiador de juego cuando se cometo ES para el análisis de datos con Python y es una de las herramientas más preferidos y ampliamente utilizado en los datos munging / discusiones si no el más utilizado. Pandas es un código abierto, de uso gratuito (bajo una licencia BSD) y fue escrito originalmente por Wes McKinney (aquí hay un enlace a su página de GitHub ).

Lo bueno de Pandas es que toma datos (como un archivo CSV o TSV, o una base de datos SQL) y crea un objeto Python con filas y columnas llamado marco de datos que se parece mucho a la tabla en un software estadístico (piense en Excel o SPSS para ejemplo. Las personas que están familiarizadas con R también verían similitudes con R).

En Computación y Ciencia de datos, pandas es una biblioteca de software escrita como extensión de NumPy para manipulación y análisis de datos para el lenguaje de programación Python. En particular, ofrece estructuras de datos y operaciones para manipular tablas numéricas y series temporales. Es un software libre distribuido bajo la licencia BSD versión tres cláusulas.<sup>1</sup> El nombre deriva del término "datos de panel", término de econometría que designa datos que combinan una dimensión temporal con otra dimensión transversal.<sup>2</sup>



## Parte III

# Conclusiones

La actividad realizada en lo personal me gusto, fue una buena manera de ir manejando una gran cantidad de datos, los cuales se pueden llegar a complicar debido a la gran cantidad de los mismos. La carga de trabajo fue un poco more extensa que la actividad previa, but fue de gran ayuda. Una de las mayores dificultades que pude tener fue la utilización de la nueva biblioteca, pero de igual forma estuvo aceptable y se pudo realizar de la mejor manera. Como nivel le asignaría un valor intermedio