El sistema consiste en una aplicación escrita en Python aplicando el framework SPARK para la manipulación, análisis y despliegue de datos. Todo esto utilizando el ambiente de desarrollo Jupyter lab y el formato de Jupyter Notebook. Además se hará uso de una base de datos en PostgreSQL, también para el almacenamiento y análisis de los datos en cuestión. En nuestro proyecto decidimos utilizar VS Code como IDE ya que soporta Jupyter Lab y como repositorio de datos y sistema de versionamiento, utilizamos Github.

En esta aplicación se analiza la posible correlación entre los datos demográficos, de empleo y participación económica; y los datos de actividad criminal en las distintas zonas del país delimitadas al nivel de distritos. Las dos fuentes de datos son los datos generados por el INEC en el censo nacional del 2011 y los datos recopilados por el OIJ para el mismo año.

El objetivo del proyecto consiste en obtener datos recopilados por dos fuentes distintas utilizando metodologías diferentes sobre los mismos grupos poblacionales. Esto supone que los datos son correspondientes en su agrupación, puesto que se trata de las misma división del mismo país, pero puede haber diferencias de forma en el cómo se escriben estos datos. Por tanto es necesario estandarizar, limpiar y empatar en los datos de las dos fuentes para poder analizar los datos correspondientes, en medida de lo posible con los datos que hay.

Así es que se desarrollan varias funciones en donde se limpian los datos de las dos fuentes y se estandarizan para poder compararlos y ver el grado de correspondencia entre los dos grupos de datos. Además se programan funciones para verificar el grado de correspondencia de la división de datos y otras más para hacer corresponder la mayor cantidad de datos posibles entre los dos conjuntos. Luego de esto, se guardan los datos depurados dentro de una base de datos PostgreSQL.

Después de este proceso, y dentro del mismo formato de Jupyter Notebook, se visualiza la información mediante gráficas en donde se comparan datos que pueden tener correlación. Como por ejemplo cantidad de delitos y tasa de ocupación, delitos por día de la semana entre otros.

Nosotros proponemos analizar también la relación que hay entre Género de la víctima, tasa de desempleo y día de la semana. Nuestra hipótesis la basamos en un ejercicio que se realizó en el 2015 por parte de Canal 7, la agencia de publicidad JWT y el INAMU en donde se pudo ver un incremento importante de las llamadas de reportes de violencia doméstica durante varios partidos de la selección nacional, frente al mismo espacio de tiempo cuando no hay partido (Fallas-Valenciano, E., & Treminio, C.). Queremos ver si los datos consultados corresponden con la premisa de este ejercicio y si el nivel de empleo afecta en este caso.

CITA APA: Fallas-Valenciano, E., & Treminio, C. (2016). Violencia de mujeres en Costa Rica: un problema para hacer conciencia. *Revista Hispanoamericana De Ciencias De La Salud*, *2*(2), 190–191. Recuperado a partir de https://www.uhsalud.com/index.php/revhispano/article/view/154