**DOCUMENTACIÓN CARPETA 02\_ELEMPLEO\_COVID**

Esta carpeta fue contiene un proyecto sobre la relación entre la llegada del Covid a Colombia y la evolución del mercado laboral colombiano. Para esto analizamos principalmente la evolución de las vacantes de empleo colombianas en Elempleo (observamos también Computrabajo, pero esta bolsa de empleo no tiene información antes de diciembre 2019 y las urls se mueven mucho). Adicionalmente, usamos también datos de la GEIH para (1) ver la relación con las vacantes y (2) caracterizar a la población ocupada colombiana.

1. **“blog”:** contiene diferentes versiones (incluyendo la final) en word de los tres blogs publicados en La silla vacía. Las entradas fueron:
   1. Evolución de las ofertas laborales, relación con el desempleo nacional y análisis por sector económico.
   2. Evolución de las vacantes por nivel educativo requerido y salario, además, usando la GEIH, se caracteriza a los ocupados por variables demográficas.
   3. Análisis de texto de las profesiones, habilidades y demás requisitos antes y luego de la pandemia.
2. **“data”**: contiene todas las bases utilizadas para el análisis, que en resumen son Elempleo, GEIH, Computrabajo y Elempleo Costa Rica. Acá los datos se dividen en raw (crudos), cleaned (procesados) y aggregated (datos que solicitaron de manera agregada, por ejemplo: total de ofertas laborales por día).
3. **“documents”:** Acá hay documentación de los datos usados. Por ejemplo, diccionario de la GEIH, anexos de empleo del DANE, clasificación de los sectores en Elempleo por código ciiu (para reagrupar) y conteos de las variables que existen en la GEIH luego de la pandemia (por cambios en los módulos).
4. **“graphs”:** Todos los gráficos usados para los blogs y el paper están acá. En la carpeta está todo lo de bolsas de empleo, sin embargo, hay dos subcarpetas con gráficos de candidatos de Elempleo y de la GEIH.
5. **“literature”:** Papers que consideramos relevante para el análisis. Adicionalmente, en Excel y en Xmind hay un resument de los papers relacionados (recomiendo abrir el de xmind).
6. **“old”:** como los datos se fueron actualizando semana a semana se dejó una copia de las carpetas, incluyendo gráficos, datos y scripts pasados, en caso de que algo saliera mal y como chequeo de que las series que se habían visto antes se mantuvieran bien (por ejemplo, que los gráficos del total de ofertas laborales de marzo se mantuviera igual al actualizar en abril)
7. **“scripts”:** acá está todo lo que obtiene, limpia y analiza los datos.
   1. **Scrapping\_computrabajo.py, scrapping\_elempleo\_cr.py y Scrapping\_web\_cluster.py**: Estos scripts recogen los microdatos crudos sobre vacantes de Computrabajo y Elempleo (Costa Rica y Colombia). El de Computrabajo saca una lista de Urls sobre las que luego se itera para obtener información (output\_computrabajo). Por otra parte, los de Elempleo iteran sobre seriales (que están al final de las urls de ofertas, ejemplo: <https://www.elempleo.com/co/ofertas-trabajo/digitador/1884553630>) y obtienen toda la información que usamos (output\_elempleo).
   2. **00\_Gen\_DataSet.R y 00\_Run.R:** acá se limpian los datos de Elempleo, se agregan los id de municipios, se limpian strings, dummies por algunos requisitos y se quitan duplicados por descripción de la oferta. Se guardan las bases en formato csv (output\_elempleo\_cleaned, en raw).
   3. **01\_scrapping\_computrabajo\_complete.R:** este script itera sobre todas las urls recuperadas en (a) (output\_computrabajo) y recupera toda la información disponible de las ofertas, como municipio, salario, y descripción. El resultado se guarda en .rds (output\_computrabajo\_cleaned, en raw).
   4. **02\_gen\_datasets.R:** acá se unen todas las bases procesadas en (b) y (c) (output\_elempleo\_cleaned y output\_computrabajo\_cleaned). Se codifican los factores, arreglan las fechas para usar, quitan duplicados por mes y descripción, agregan los sectores económicos de las empresas agrupados y se guardan las bases listas para analizar (elempleo.rds, elempleo\_cr y computrabajo.rds, en cleaned).
   5. **03\_analysis.R:** este script hace un análisis general del comportamiento de las ofertas laborales en Elempleo y Computrabajo. Se observa cómo ha cambiado el número de ofertas en el tiempo y se miran también cómo han cambiado las ofertas con características específicas (en Bogotá, de teletrabajo, en el sector salud, etc). Los outputs de este script se guardan en la carpeta graphs.
   6. **04\_analysis\_economic\_sector.R**: Acá se construyen los gráficos usados en el primer blog sobre los cambios en las ofertas laborales por sector económico. fun\_sector\_2019\_2020 es una función que saca el total de ofertas por sector económico diariamente y semanalmente, y las grafica. Adicionalmente, compara la diferencia del logaritmos del total de ofertas semanales respecto al inicio de febrero.
   7. **05\_geih\_unemployment.R:** utiliza microdatos de la GEIH para sacar bases pequeñas. La primera sección “UNEMPLOYMENT DATASETS” obtiene indicadores del mercado laboral, como tasa de desempleo, tasa de ocupación, etc, a nivel nacional, por ciudades y sector económico. Crea una base que se llama “geih.rds” en cleaned. La segunda sección “WAGE DATASET” se queda con los microdatos de la encuesta entre 2019 y 2020, y variables sobre ingresos, género, afiliación laboral, etc, para caracterizar a los ocupados. Crea una base que se llama “geih\_wages.rds” en cleaned.
   8. **06\_geih\_vs\_elempleo.R:**

**PASOS PARA ACTUALIZAR LOS DATOS Y GRÁFICOS DE BOLSAS DE EMPLEO**