Reporte Big\_Data\_TP2\_Grupo20.pdf

Punto 1) Se identifica a las personas desocupadas como, las personas que por el momento no están trabajando pero están en la búsqueda, es decir están dispuestos a trabajar.

Punto 2-b) Las variables que tienen más valores faltantes son GDECCFR (Numero del conjunto de aglomerados de 500.000 y más habitantes) y PDECCFR (número de decil de ingreso per cápita familiar del conjunto de aglomerados de menos de 500.000 habitantes). El año es 2024.

Punto 2-c) Se realizó una limpieza en las 15 variables de ambos años, lo que mostro que había más valores faltantes de los que habíamos observado en el paso anterior. Para corregir los valores sin sentido, buscamos en los diccionarios de EPH cuáles eran estos valores (9, 99, 999, 9999, -7, -8, 0), estos eran iguales en ambos años. Nos dio como resultado que el año que tenía mayor cantidad de valores faltantes era 2024, con 7647.

Punto 3) En el gráfico, podemos observar que la cantidad de hombres y mujeres es similar. Podemos ver que en 2024 hay una mayor cantidad de personas registradas, esto se puede deber a crecimiento poblacional.

