

Analisis cluster pobreza laboral

Luis Bermeo

June 2021

1. Clusters

Obtén el crecimiento del empleo y del salario promedio de 2019 a 2020 para cada entidad federativa.
Obtén la pobreza laboral para el mismo periodo que antes
Realiza un análisis cluster

* con 4 cluster

* sin definir

Interpreta estadísticamente y económicamente los resultados.

Dado que queremos encontrar una relación entre las variables indicadas para cada estado de la republica nos gustaría hacer un método de clusterización no supervisado, de forma conveniente escogemos k-means pues nos permite dar una interpretación a partir de lo observado en las gráficas. Debido a que se tiene más de dos variables entonces se pueden seguir dos enfoques, uno donde se realicen cluster cada 2 variables o bien ocupar el método de componentes principales, en este trabajo se colocan ambos enfoques (el primero por la simplicidad del análisis y el segundo por ser un método más preciso). Cada entidad federativa se considera una observación y nos gustaría realizar una clasificación de acuerdo a las características observables. A continuación se muestra el método de clusterización por k-means usando 4 clasificadores.

1.1. cluster por separado k =4

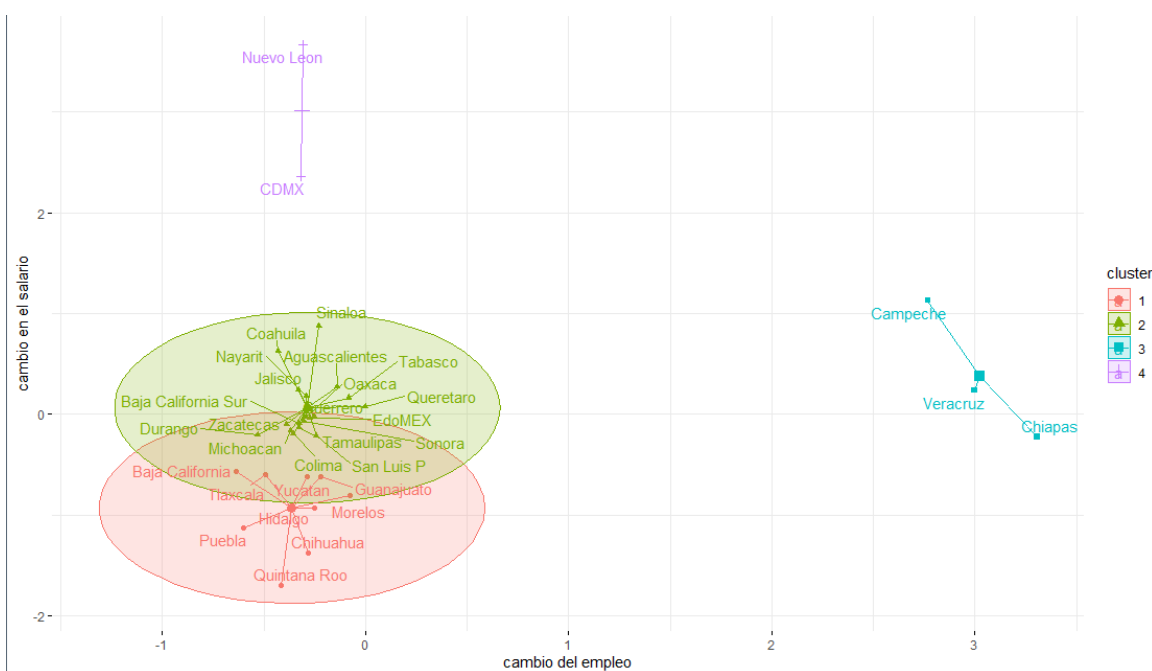


Figura 1: Cluster k = 4 Cambio del empleo y cambio de salarios

En la figura anterior se puede ver que la mayoría de los estados salvo los 3 que pertenecen al cluster 3 han experimentado un cambio negativo en el nivel de empleo esto se puede atribuir al efecto del confinamiento causado por la pandemia del

covid-19, sin embargo se pueden identificar grupos que han experimentado un incremento en los salarios promedio y aquellos que han tenido una disminución, aquellos estados que han tenido un incremento en el nivel de salario (cluster 2 y 3) puede atribuirse a que la mayor parte de los despidos se realizaron en empleos con personal no cualificados, servicios (restaurantes, bares, peluquerías, etc) y puestos de salario con salario bajo asociado debido que muchas de estas actividades tuvieron que cerrar por completo y sus empleados no pudieron realizar home office (en contra al personal cualificado con salarios altos, puestos de cuello blanco cuyas actividades podían realizarse desde sus hogares), esta hipótesis se confirma con el cluster 4 (CDMX y Nuevo León) los cuales concentran una parte importante del personal cualificado (servicios financieros, administrativos, etc) y que al disminuir la masa de trabajadores con salario bajo se refleja en un incremento importante de los salarios. Por otro lado estados donde los salarios son mas homogéneos (cluster 2 y 3) se observa una disminución o un ligero aumento del salario. Para el Cluster 3 se puede explicar por que estos estados se han encontrado con bajos niveles de contagio (Campeche hasta ahora es el único estado en semáforo verde) lo cual puede no haber provocado la caída del empleo y actividades esenciales como la agricultura pueden reflejar el incremento del empleo (aunque en nivel salarial no aumentaron considerablemente).

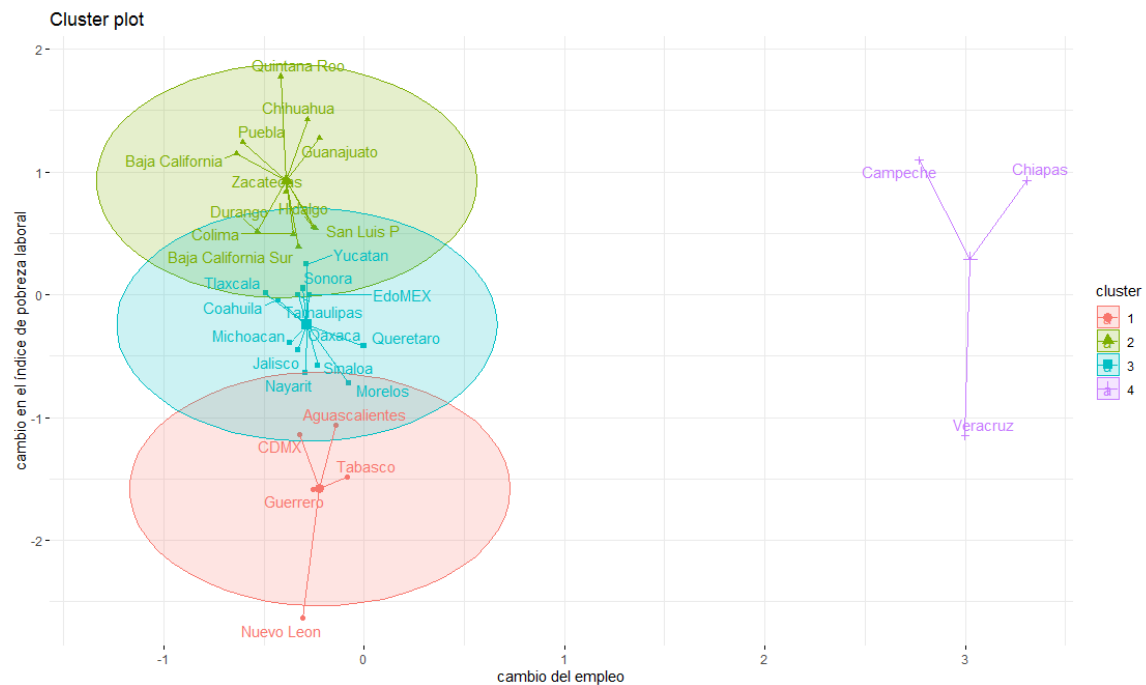


Figura 2: Cluster k = 4 Cambio del empleo e índice de pobreza laboral

Ahora se puede observar que algunos estados experimentaron una caída del empleo acompañada con una disminución de la pobreza laboral entre los que se encuentran CDMX y Nuevo León esto puede deberse como se comentó en el análisis anterior a que en estos estados los despidos ocurrieron principalmente en aquellos trabajadores con un salario bajo asociado (por debajo del umbral de pobreza) lo cual disminuye el índice de pobreza laboral en estos estados. También se observa que la mayoría de los estados (cluster 2 y 3) han tenido un incremento en el índice de pobreza laboral y una caída de los niveles de empleo, en general puede deberse debido a que la mayoría de la población gana mas homogéneamente en lugar de realizar despidos algunos empleadores optaron por disminuir jornadas y salarios por lo que una cantidad importante de los trabajadores cayó por debajo del umbral de pobreza laboral principalmente en los estados manufactureros del centro-bajo y Quintanarro cuya actividad económica depende exclusivamente del turismo que fue fuertemente golpeado por la pandemia.

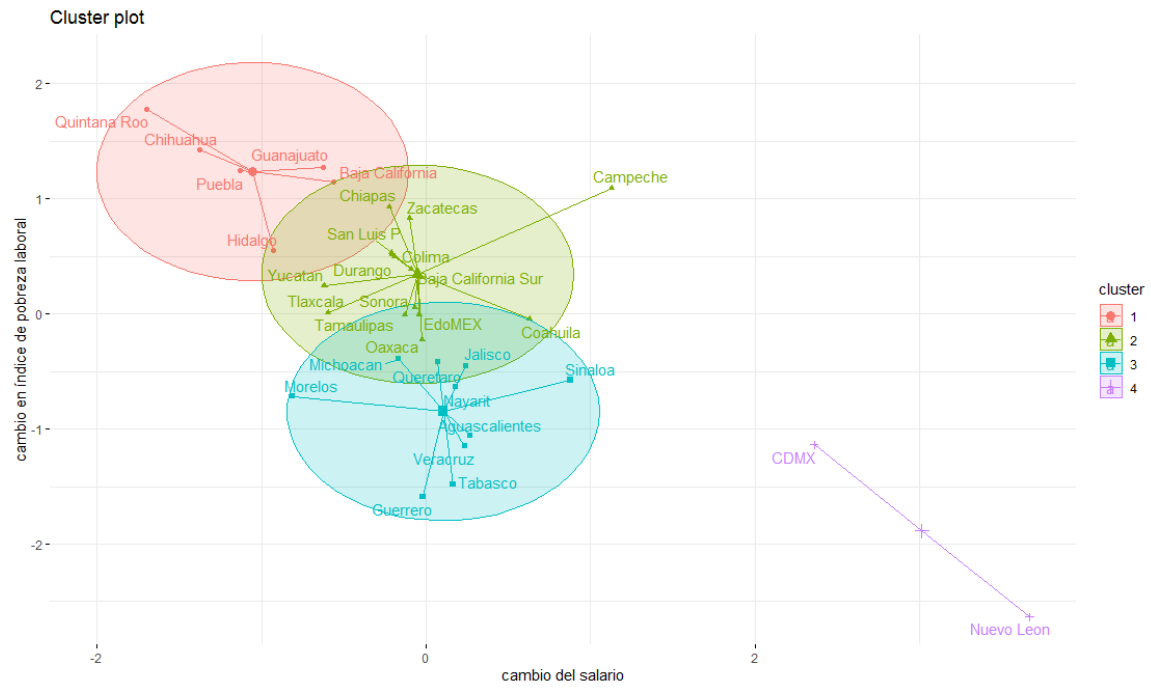


Figura 3: Cluster k = 4 Cambio de salarios e índice de pobreza laboral

Cuando realizamos cluster por índice de pobreza y salario los resultados son congruentes con los análisis anteriores, primero se observa que CDMX y Nuevo León son los estados que han experimentado un incremento del salario y una disminución del índice de pobreza laboral lo que confirma que en estos estados se disminuyó en nivel de trabajadores con un salario bajo asociado por lo que el promedio se incrementó (y no fue necesariamente por un incremento en la riqueza per capita promedio), por otro lado en los estados del cluster 1 y 2 (Centro y norte principalmente) el índice de pobreza aumentó mientras disminuyó el salario lo cual confirma que los trabajadores en promedio en estas entidades ganan en promedio menos y aumentó el número de ellos que ganan por debajo del umbral de pobreza posiblemente por una reducción de las jornadas laborales o un acuerdo entre trabajador y patrón por disminuir el salario a cambio de evitar despidos, en el cluster 1 se muestran los estados más afectados entre los que destaca Quintana Roo posiblemente por depender completamente del turismo. Para los estados del cluster 3 (sur principalmente) hay una disminución de la pobreza laboral y de los salarios esto puede deberse a que se realizaron despidos principalmente en trabajadores con un salario bajo (por debajo del umbral de pobreza) y para los trabajadores con salario más alto se pudo acordar una reducción de los salarios.

Realizar los cluster por separado para su análisis resulta conveniente cuando el número de variables es pequeño sin embargo este no refleja del todo la relación entre la correlación que existe entre las variables y podría no ser la mejor técnica para poder clasificar las observaciones por lo que usaremos el método de componentes principales para producir 4 cluster mediante componentes principales.

1.2. 4 cluster usando PCA k=4

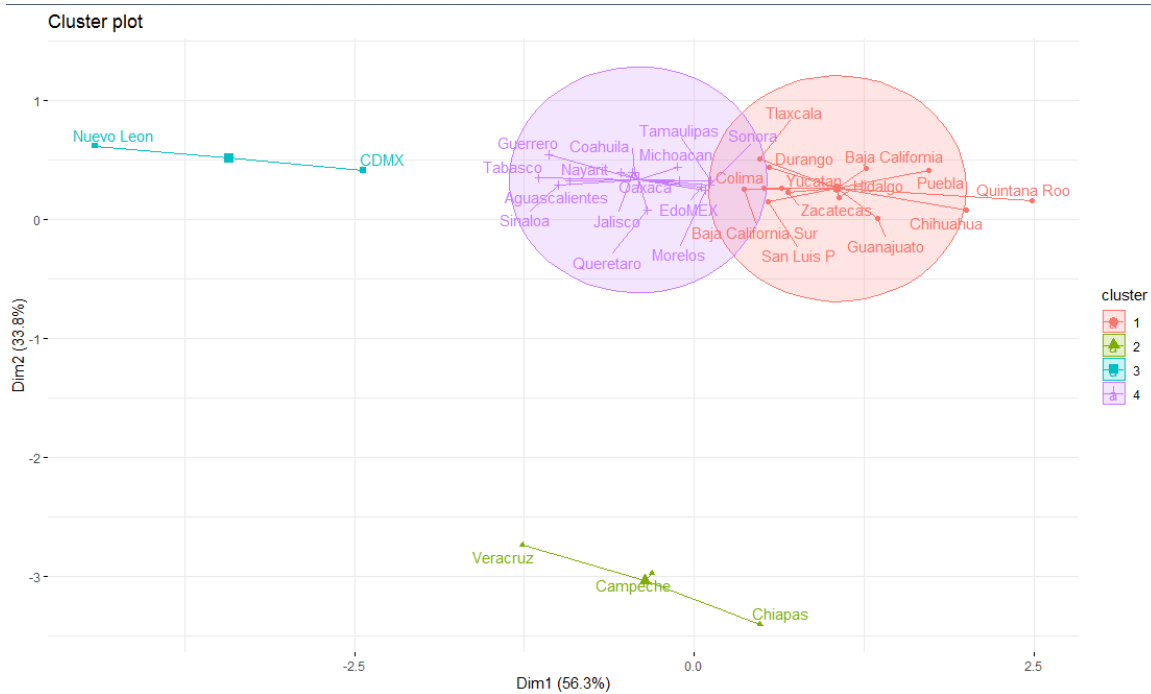


Figura 4: Cluster k = 4 usando PCA

Lo primero que se observa y es sumamente interesante es que los cluster 3 (CDMX-Nuevo León) y 2 (Veracruz Campeche y Chiapas) ya los habíamos obtenido con los análisis anteriores lo cual indica que entre estos grupos exhiben características muy parecidas sobre las tres variables. El porcentaje de explicación es alrededor del 70 %

A continuación se muestra la contribución de las variables a los componentes principales a la izquierda se muestra que cambio de salario e índice de pobreza explican el componente principal 1 y a la derecha que cambio en el empleo explica el segundo componente principal.

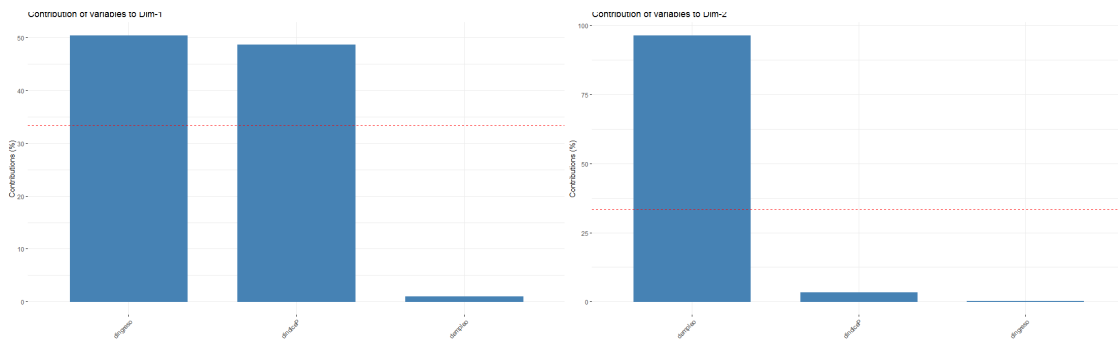


Figura 5: Contribución a los componentes principales

por lo tanto se puede ver que el salario y el índice de pobreza están altamente correlacionados (como era de esperarse). Entonces de este modo el cluster por componentes principales resume de mejor forma los resultados obtenidos por un análisis por pasos, para el cluster 1 el primer componente es bajo y el segundo componente es alto pues en estos estados los salarios aumentaron y se disminuyó el índice de pobreza laboral como consecuencia de que los despidos se realizaron en trabajadores con salarios bajos. Para el cluster 2 y 3 dado que los salarios son más homogéneos se observa que en el cluster 4 los despidos se realizaron de forma homogénea por lo que el componente principal 1 está cercano a cero a diferencia del cluster 1 donde seguramente los salarios se disminuyeron a la vez que se realizaron recortes con lo que se incrementó el índice de pobreza laboral y bajaron los salarios. El cluster 2 corresponde a los estados donde la pandemia afectó poco la actividad económica y por ello se observa incremento del empleo (componente 2) y centrado al cero para el componente 1.

1.3. Cluster sin definir

Para realizar la selección del número de cluster usamos cross-validation mediante la segmentación de la muestra en una de entrenamiento y una de test y utilizando el número de cluster que minimice el WSS (within sum square). En la siguiente figura se muestra una gráfica del WSS vs diferente número de cluster usando kmeans.

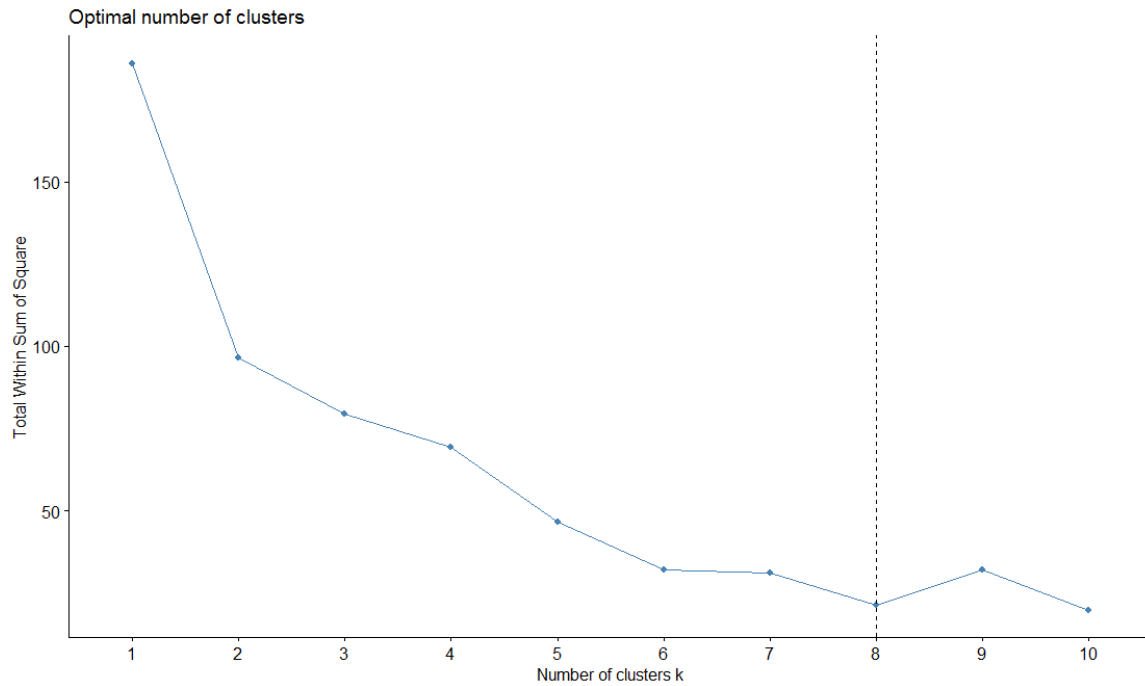


Figura 6: Selección del número de cluster optimos

Como se muestra en la figura anterior el número de cluster óptimos es 8, por lo que realizamos clusterización usando PCA como se muestra en la siguiente figura

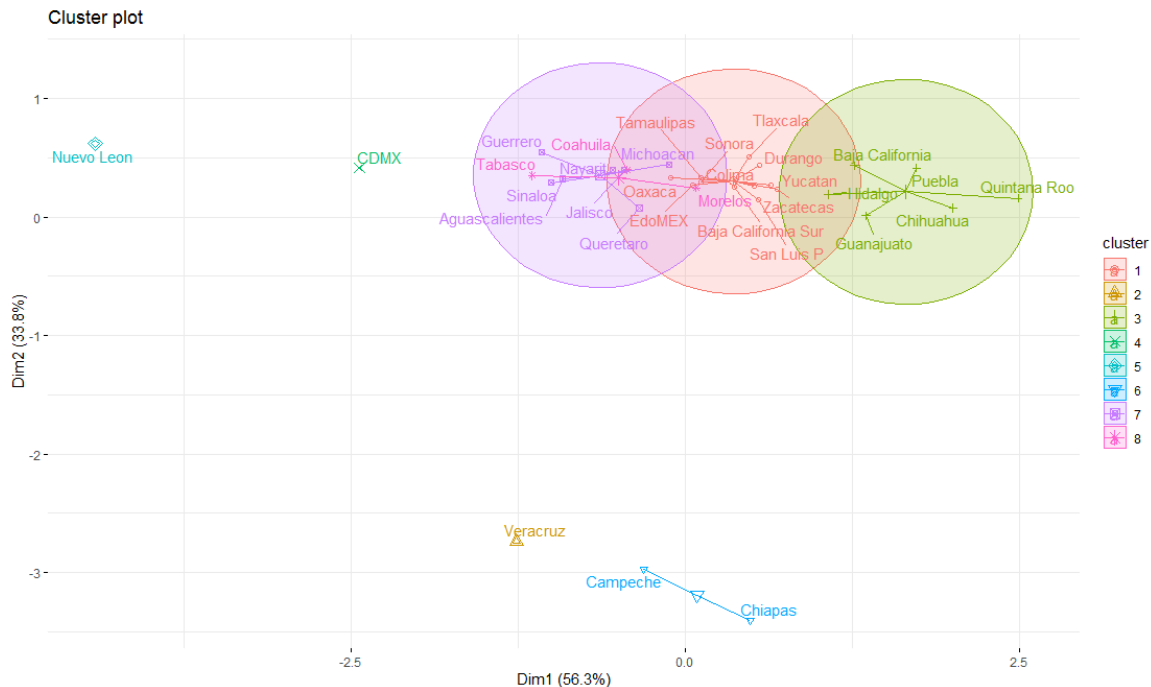


Figura 7: Cluster con k=8 usando PCA

Como se observa los cluster CDMX-Nuevo Leon y los poco afectados se descompusieron en clases de una y dos observaciones esto puede darse debido a la diferencia en los niveles de pobreza laboral-salario entre Nuevo Leon y CDMX por lo que resulta mejor clasificarlos por separado. Asimismo campeche-chiapas cuyas características socioeconomicas son mas parecidas parecen

mejor agruparse por separado de Veracruz esto es evidencia de los sectores economicos que predominan en estas regiones lo que provoca que el cambio en el empleo, salario e indice de pobreza sea significativo para consideralos por separado. Asimismo para el cluster 3 se observa que los estados de este grupo corresponden a los mismos del cluster 1 de salario y pobreza laboral que comparten características económicas que resultan mejor clasificarlas por separado de los otros estados, aquellos estados en el cluster 7 que son los que resultan menos afectados a nivel de salario y pobreza laboral corresponden a los estados donde los salarios ya eran bajos como para optar por reducirlos, así como estados con zonas turísticas importantes y de servicios donde los empleadores optaron por despedir a sus empleados. En cuanto la zona 1 corresponde principalmente al noreste donde la actividad industrial predomina y algunos empleadores optaron por despido acompañado de la reducción de salario y/o jornada laboral.