

UNIVERSIDAD NACIONAL AUTÓNOMA DE
MÉXICO

DIPLOMADO EN CIENCIA DE DATOS

Sector Asegurador

Primer Examen Módulo 1

Jessica Fernanda Rodríguez Mondragón

Septiembre 2020

Índice general

Resumen	II
Introducción	I
Calidad de datos	I
0.1. Diccionario de datos	I
0.2. Exploración de Datos	I
0.3. Análisis de valores ausentes	I
0.4. Identificación de Valores Atípicos	I
0.4.1. Método Percentiles	I
0.5. Ingeniería de Variables	I
0.6. Visualización de Datos	II
0.6.1. Asegurados y no trabajadores por Clave de Delegación	II
0.7. Conclusión y Propuesta	IX

Resumen

El sector asegurador necesita hacer proyecciones financieras con escenarios, según sus necesidades, para determinar la situación en la que se encuentra y las acciones a realizar. En este documento se presenta un análisis exploratorio y descriptivo de datos provenientes de la página oficial del IMSS, los cuales contienen información mensual del número de asegurados y salario base de cotización por desglose de dimensiones de interés.

Introducción

Desde la creación del Instituto Mexicano del Seguro Social (IMSS), en 1943, ha brindado seguridad social en el país. El IMSS, se destaca por brindar protección social a los trabajadores mexicanos y a sus familias, otorgando acceso y cobertura a sus asegurados en todo momento.

Desde 1943, el IMSS ha logrado un gran alcance y cobertura en todo el país, brindando atención médica a la población que no cuenta con acceso a la seguridad social, llegando a zonas rurales y comunidades indígenas del país, a través de programas que se llevan a cabo con el fin de beneficiar a estas comunidades.

El IMSS es el mayor prestador de servicios médicos en México, pero, no solo brinda servicios de salud, sino que también brinda apoyo del cuidado infantil, ofrece protección a riesgos económicos, brinda actividades culturales para los derechohabientes, deportivas, centros vacacionales, etc.

El IMSS, como la institución de seguridad social con el mayor número de derechohabientes, constantemente se enfrenta a grandes retos. En diciembre de 2018, el Instituto contaba con una población derechohabiente de 68.2 millones de personas, sin incluir a 13.1 millones que no cuentan con seguridad social y que reciben servicios médicos por parte del Instituto.

Anuado a esto, el sistema de pensiones de México, es decir, los esquemas de las pensiones y sus costos son uno de los principales problemas nacionales en México, son un problema social, que de acuerdo al Informe al Ejecutivo Federal y al Congreso de la Unión sobre la situación financiera y los riesgos del Instituto Mexicano del Seguro Social 2018-2019, se manifestará a partir de 2021, cuando la próxima generación a jubilarse bajo el último esquema de pensiones se dé cuenta que no posee una pensión, lo cual es un problema, por lo que se debe actuar desde ahora.

En este documento se presenta un análisis exploratorio y descriptivo sobre datos pertenecientes a los asegurados del IMSS, que contiene información desde 1997 con el fin de presentar una propuesta de mejora para las entidades federativas con un menor número de personas cotizantes al Seguro Social y con un salario base de cotización mínimo.

Calidad de datos

0.1. Diccionario de datos

Uno de los grandes retos al trabajar con datos oficiales provenientes del gobierno es entender los datos y las terminaciones que estos presentan, en esta sección, con el fin de entrar en contexto con los resultados, se presenta un diccionario de datos.

Variable	Tipo de dato	Descripción
cve_delegacion	Int	Identificador de la delegación de adscripción al IMSS
cve_subdelegacion	Int	Identificador de la subdelegación de adscripción al IMSS
cve_entidad	Int	Identificador de la entidad federativa (catálogo IMSS)
cve_municipio	String	Identificador del municipio (catálogo IMSS)ó
sector_economico_1	Int	Identificador de sector económico a 1 posición
sector_economico_2	Int	Identificador de sector económico a 2 posiciones
sector_economico_4	Int	Identificador de sector económico a 4 posiciones
tamaño_patron	String	Identificador del rango de tamaño del registro patronal
sexo	Int	1: hombre 2: mujer
product_id	String	Id único del producto
rango_edad	String	Identificador de rango de edad
rango_salarial	String	Identificador de rango salarial
rango_uma	String	identificador del rango de la unidad de medida de actualización (UMA). A partir de febrero de 2017
asegurados	Int	Número de asegurados
no_trabajadores	Int	Asegurados sin un empleo asociado
ta	Int	Puestos de trabajo afiliados al IMSS (empleos asegurados o asegurados asociados a un empleo)

Conociendo los nombres de las variables y lo que significan se presenta la siguiente clasificación de variables:

- Continuas = 'asegurados', 'no_trabajadores', 'ta'
- Categóricas: 'cve_delegacion', 'cve_subdelegacion', 'cve_entidad', 'cve_municipio', 'sector_economico_1', 'sector_economico_2', 'sector_economico_4', 'tamaño_patron', 'sexo', 'rango_edad', 'rango_salarial', 'rango_uma'

0.2. Exploración de Datos

Dentro de este apartado, se muestran las características de nuestro conjunto de datos original, en primer instancia se cuenta con una dimension de 4,470,204, registros con 29 variables, sin embargo, para fines de este análisis, se reduce a trabajar con las variables descritas en el diccionario de datos, por lo que se reduce a 17 variables.

Para fines informativos, a continuación se muestran las categorías de las variables de acuerdo al diccionario de datos proporcionado por el IMSS.

■ Clave Delegación:

1. Aguascalientes
2. Baja California
3. Baja California Sur
4. Campeche
5. Coahuila
6. Colima
7. Chiapas
8. Chihuahua
9. Durango
10. Guanajuato
11. Guerrero
12. Hidalgo
13. Jalisco
14. Edo. México Oriente
15. Edo. México Poniente
16. Michoacán
17. Morelos
18. Nayarit
19. Nuevo León
20. Oaxaca
21. Puebla
22. Querétaro
23. Quintana Roo
24. San Luis Potosí
25. Sinaloa
26. Sonora
27. Tabasco
28. Tamaulipas
29. Tlaxcala
30. Veracruz Norte
31. Veracruz Sur
32. Yucatán
33. Zazatecas
34. Ciudad de México Norte
35. Ciudad de México Sur

■ Tamaño de registro patronal

- S1: con un puesto de trabajo
- S2: con dos y hasta cinco puestos de trabajo
- S3: con seis y hasta cincuenta puestos de trabajo
- S4: con cincuenta y uno y hasta doscientos puestos de trabajo
- S5: con doscientos cincuenta y uno hasta quinientos puestos de trabajo

- S6: con quinientos uno y hasta mil
- S7: con más de mil puestos de trabajo
- NA: no aplica

Para el caso de Rango de salario las dimensiones se encuentran como enlistadas desde W1 hasta W25, que significa hasta una vez el salario mínimo y mayor a 24 y hasta 25 veces el salario mínimo respectivamente.

A continuación se presenta el comportamiento de las diferentes variables continuas que se tienen, con el fin de observar los fenómenos que nos indican las variables.

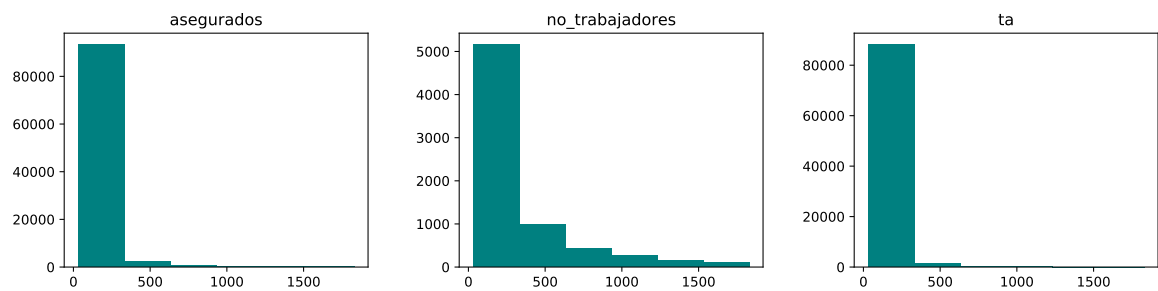


Figura 1:

Si bien, cabe mencionar que el total de asegurados del conjunto de datos, registrados en el IMSS es de 127,446,984. Mientras que los no trabajadores, son personas que están aseguradas ante el IMSS, pero no cuentan como trabajadoras al no estar registradas con un empleo asociado, las cuales podrían ser familiares de trabajadores cotizantes al IMSS, o bien, personas pensionadas, suman un total de 34,988,327, es decir, casi el .30 por ciento de las personas aseguradas.

Mientras que el comportamiento de la variable de Puestos afiliados al IMSS, es muy similar a la de asegurados, además que los puestos de trabajo suman el 19,495,952.

0.3. Análisis de valores ausentes

Dentro de este apartado se encuentra el tratamiento realizado a la tabla de datos original sobre los valores ausentes, la siguiente imagen permite identificar de una manera visual los valores ausentes dentro de la tabla original de datos:

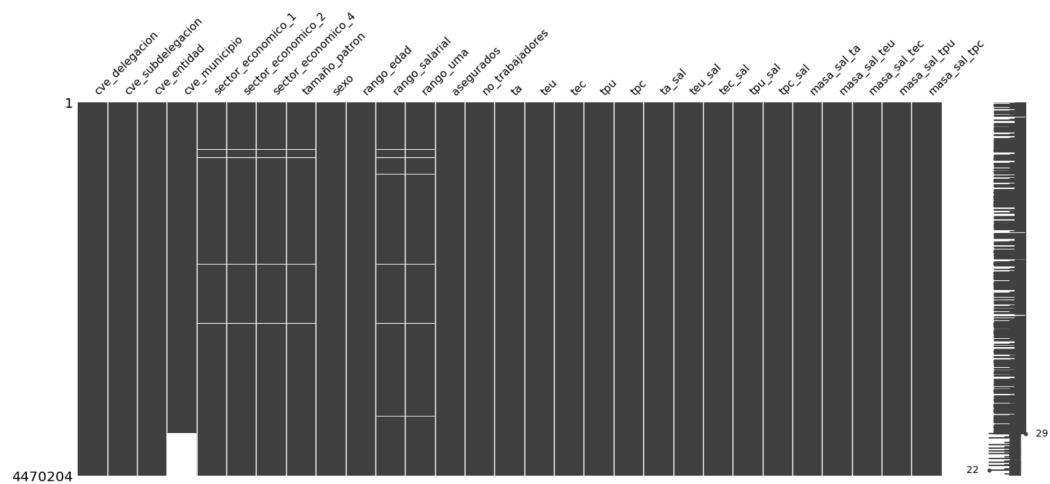


Figura 2:

Se observa que la clave de municipio cuenta con una gran cantidad de valores ausentes, sin embargo, esta cantidad está ausente ya que según el diccionario de datos proporcionado por el IMSS, cve_municipio toma el valor de NA cuando la entidad se refiere a Ciudad de México, ya que ciudad de México no tiene un municipio asociado, por lo que no se eliminará esta variable.

Para el caso de los sectores económicos y los rangos salariales los valores ausentes pertenecen al campo de 'No_trabajadores', es decir, no perciben un salario, por lo tanto, las variables que derivan de esta están como nulas, por lo que tampoco se procede a hacer un método de imputación de variables ausentes, ya que en este caso, el no tener información, proporciona información.

0.4. Identificación de Valores Atípicos

Un valor atípico, o también conocido como outlier, tiene un comportamiento especial, y único a diferencia de los demás, existen técnicas estadísticas de identificación de estos valores, por lo que para poder identificarlos en los datos de los asegurados usaremos el método de Percentiles.

0.4.1. Método Percentiles

Los valores atípicos se caracterizan por estar en los extremos de las distribuciones de las variables, por lo que se usa el método de percentiles, excluyendo todos aquellos menores al 1 % de los datos y al 99 %.

Al realizar este método en los datos del IMSS, la mayor parte de outliers provienen de la variable No Trabajadores, es decir no sería correcto eliminarlos, ya que para realizar un análisis más completo se tomará la variable de No Trabajadores.

0.5. Ingeniería de Variables

Como se muestra en el apartado de diccionario de Datos, se cuenta con un gran número de variables que son útiles para realizar un análisis e incluso un modelo, sin embargo, a partir de las variables originales se pueden derivar o crear nuevas variables, a este proceso se le conoce como Ingeniería de Variables.

El siguiente listado, muestra las variables que dieron origen a otras nuevas:

- rango_uma:
 1. cantidad_tope: Cantidad del tope de la UMA
 2. uma_valor: Conversión de cantidad a valor de la UMA (Tomando de referencia el valor de la UMA en 2020 \$ 86.88)
 3. uma_mensual: El valor mensual de la UMA se calcula multiplicando su valor diario por 30.4 veces)
- rango_salarial:
 1. cantidad_tope_salario: cantidad del tope del salario diario de cotización
 2. salario_diario: resultado de multiplicar la cantidad topada, con la cantidad de 185.56 mxn que es el salario minimo diario en 2020.
- asegurados y no_trabajadores:
 1. avsnt: Equivalencia de los no trabajadores entre los asegurados
 2. asegurados_trab: Asegurados trabajadores

Las variables anteriores, permiten dar el siguiente paso, realizar visualizaciones y análisis descriptivo.

0.6. Visualización de Datos

La visualización de datos permite conocer las características de los datos y presentar resultados de forma gráfica, por lo que en la siguiente sección se muestra un análisis desarrollado por clave de entidad, asegurados y rangos salariales.

0.6.1. Asegurados y no trabajadores por Clave de Delegación

Se cuentan con 35 claves de delegaciones, sin embargo, cabe aclarar que el nombre de la descripción de clave de delegación está dividida por estados de la República Mexicana, más allá de entidades.

Como primera pregunta a responder, es de interés saber cuál es el porcentaje del número de asegurados por 'Clave de Delegación', el cuál se muestra en la siguiente figura:

Porcentaje de Asegurados por clave de delegación

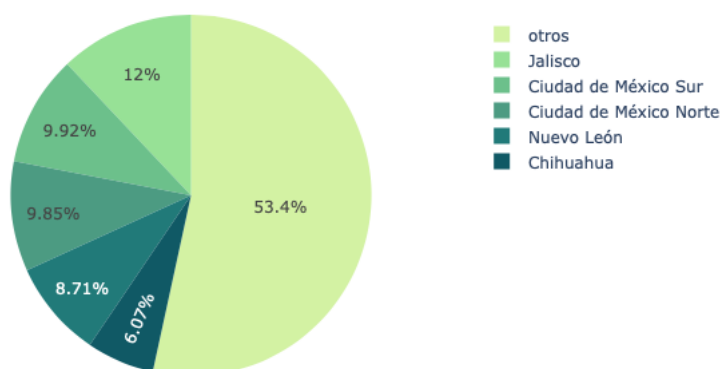


Figura 3:

Dado lo anterior, casi el 50 % de asegurados pertenecen Jalisco con un 12 % de asegurados, Ciudad de México Norte y Sur sumando un 19.177 % en total, y Nuevo León con 8.71 % de asegurados, mientras el resto de los estados suman el 53 % del total.

Siguiendo el mismo análisis se mostrará para las personas No Trabajadoras pero que sí están afiliadas al IMSS.

Porcentaje de no trabajadores por clave de delegación

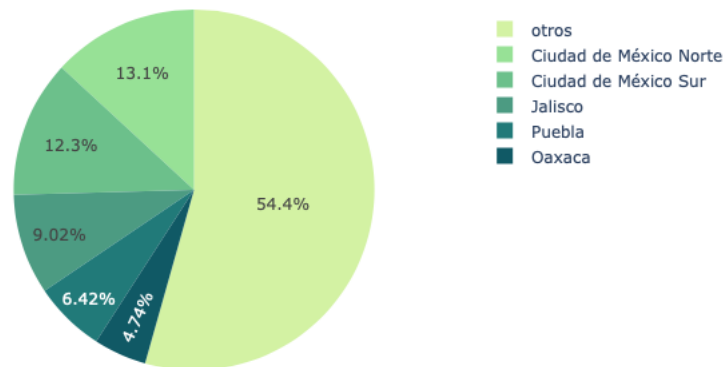


Figura 4:

Se muestra que La ciudad de México Norte, Sur, Jalisco, Puebla y Oaxaca, suman el 46 % de los no trabajadores, es decir, existe una mayor probabilidad de ser no trabajador si la zona de residencia está en Ciudad de México.

Por otra parte, es de interés saber cuál es el porcentaje de las personas no trabajadoras respecto a las aseguradas, como se muestra en la siguiente figura, se puede notar que respecto a la proporción, en Guerrero, la mayoría de los asegurados, no cuentan con un empleo asociado con el cual cotizar al Seguro Social.

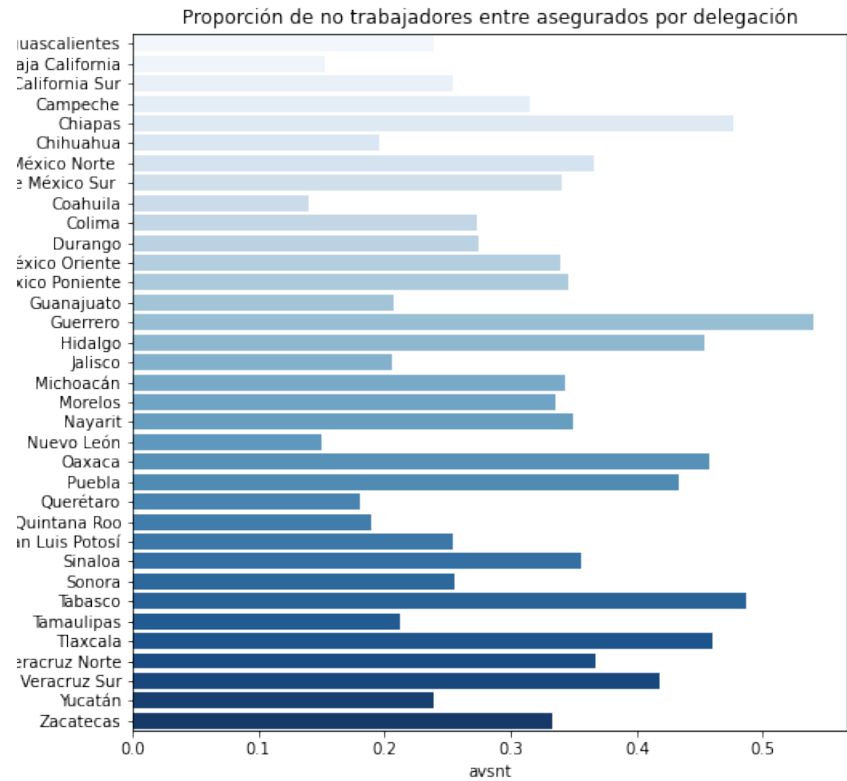


Figura 5:

Además de lo anterior, se sabe que el salario mínimo diario percibido por los trabajadores en México es de \$185.56 mxn, por lo que, en la siguiente gráfica se muestra la mediana del salario percibido en cada estado de la República Mexicana por los asegurados cotizantes al IMSS:

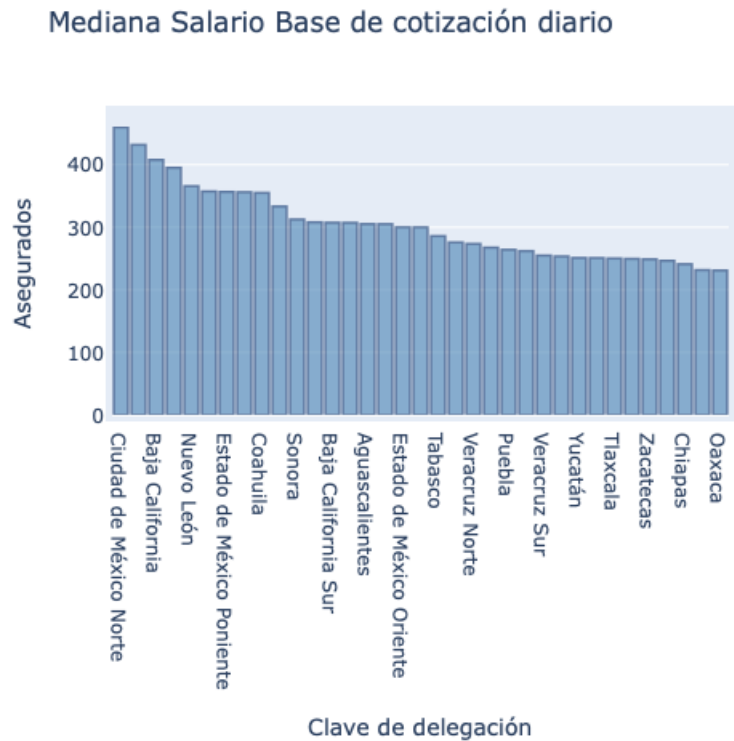


Figura 6:

Podemos notar que el la mediana del salario base de cotizacion por los diferentes Estados de la República se mantiene entre 200 MXN y 500 MXN, siendo Oaxaca y Ciudad de México Norte los extremos respectivamente.

Ahora bien, por otro lado se muestra la proporción de solamente aquellos Trabajadores Asegurados que cuentan con un salario base de cotización:

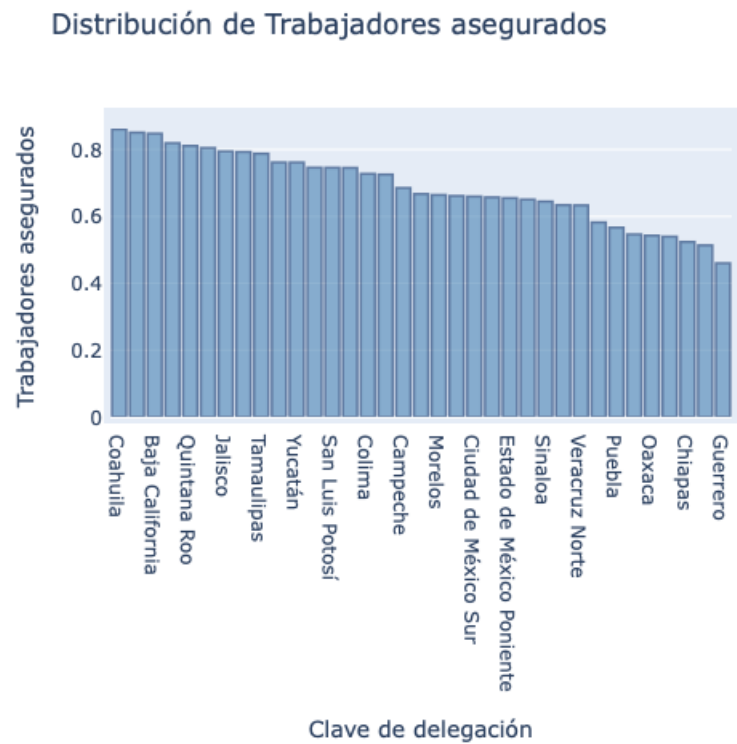


Figura 7:

De igual manera es importante conocer el comportamiento que tienen los puestos de trabajo que están afiliados al IMMS, en comparación con los asegurados, en donde puestos de trabajo corresponde a la gráfica en color azul.

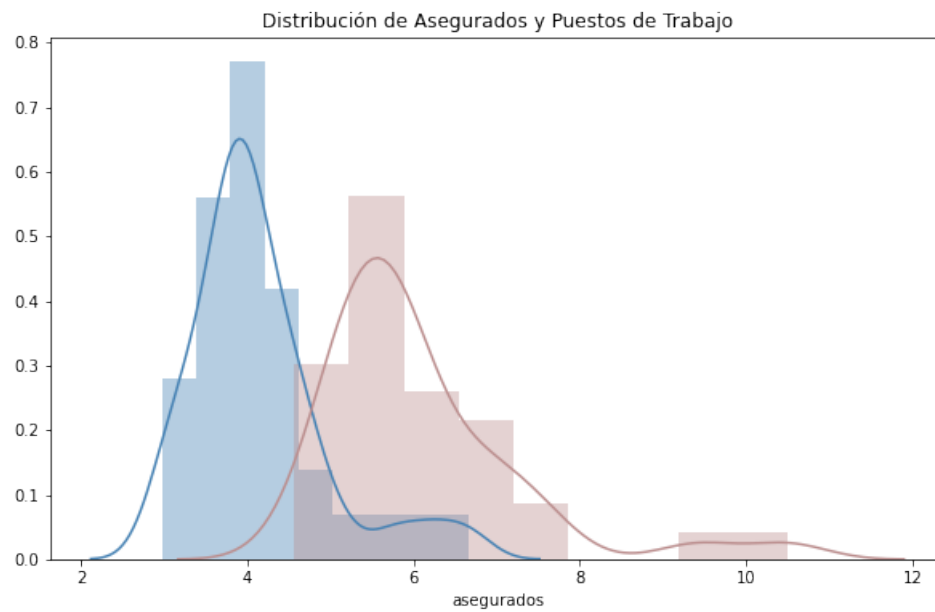


Figura 8:

Por otra parte, la mayoría de los asegurados se encuentran en un rango salarial W2, es decir, perciben de uno hasta 2 salarios mínimos diarios.

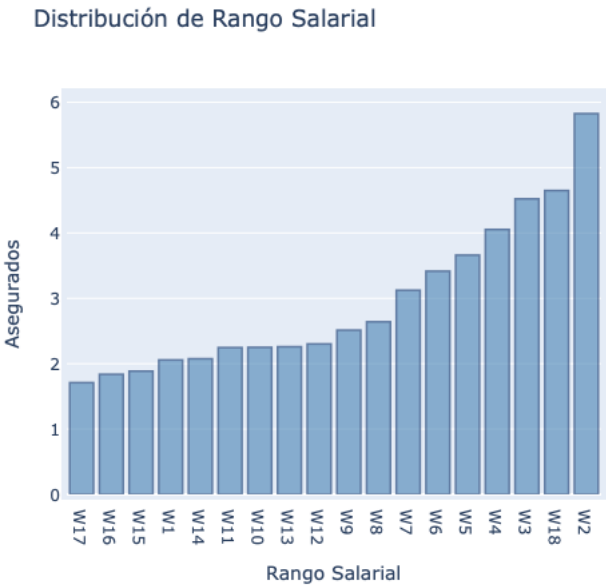


Figura 9:

De igual manera, la mayor parte de los asegurados pertenecen al Sector económico de Servicios para empresas, personas y el hogar.

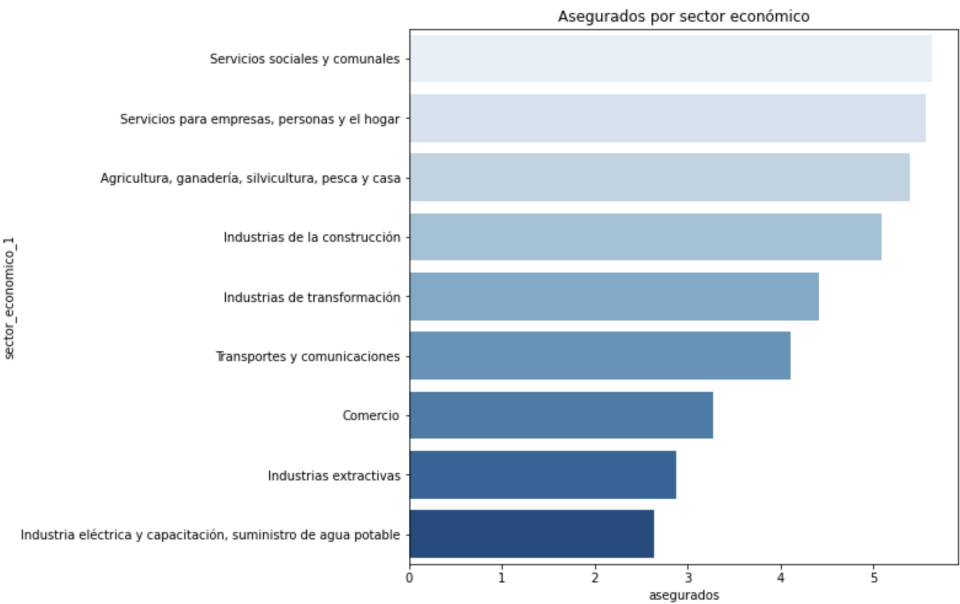


Figura 10:

0.7. Conclusión y Propuesta

Como primeras conclusiones obtenidas del análisis anterior, se busca saber cuáles son las delegaciones, o en este caso, los Estados de la República Mexicana, que tienen un performance menor a todos los demás respecto al salario base de cotización, a la cantidad de trabajadores asegurados que están afiliados al IMSS que cuentan con un Rango Salarial entre uno y hasta dos veces el salario mínimo ya que, especialmente estas personas y estos estados de la República, están cotizando con una menor cantidad ante el Seguro Social, lo cuál se verá afectado al momento del cálculo de una pensión. Si bien, un análisis de estadística descriptiva no es suficiente, este documento está planeado como base de una investigación y desarrollo más profunda.

Como segunda conclusión, el tema de las pensiones en México se ha visto modificado de acuerdo a la reforma al sistema de pensiones, la cual busca aumentar la pensión que ofrece el Sistema de Ahorro para el Retiro (SAR) a los trabajadores afiliados al Instituto Mexicano del Seguro Social (IMSS) y evitar una crisis futura.

Un punto importante es que los trabajadores promedio podrían aumentar hasta un 40 % su pensión, sin embargo, para los Estados de la República con menores ingresos y con menor salario base de cotización, no se verán muy beneficiados a comparación de los otros, por lo que se propone hacer un modelo de contraste de poblaciones, es decir, se tiene la información agrupada por diferentes dimensiones (poblaciones), las cuales poseen diferentes comportamientos, por lo que dependiendo de las características de cada población, por medio de la dimensión no supervisada, generar una clusterización que arroje los grupos de entidades y de asegurados cotizantes al IMSS con características similares. Los grupos de entidades que resulten más afectados, es decir, aquellos grupos de asegurados con un salario base de cotización bajo, serán aquellos a los que se les podría diseñar un apoyo o un programa por parte del gobierno que ayude a mejorar la expectativa de las pensiones de este grupo de asegurados, el cual esté diseñado desde educación económica sobre las pensiones, métodos de ahorro, hasta apoyos económicos o facilidades de abrir una cuenta de ahorro.