

REPORTE ANALÍTICO DE INDICADORES

**INDICADORES DE  
TELECOMUNICACIONES  
E ÍNDICE DE BIENESTAR  
DIGITAL PARA ENTIDADES  
FEDERATIVAS DE MÉXICO**

**3ER CUATRIMESTRE DE 2022**

**CENTRO DE ESTUDIOS DEL IFT**

**MTRO. JOSÉ ALBERTO CANDELARIA BARRERA**

# Indicadores de telecomunicaciones e índice de bienestar digital para entidades federativas de México.

Mtro. José Alberto Candelaria Barrera

## Índice

<b>Indicadores de telecomunicaciones e índice de bienestar digital para entidades federativas de México.....</b>	<b>1</b>
Introducción.....	3
Información estadística y consideraciones metodológicas.....	5
1. Infraestructura de TIC´s.....	6
2. Adopción de TIC´s.....	6
3. Asequibilidad.....	7
4. Usos de TIC´s en la actividad económica.....	7
5. Usos de internet.....	8
Diagnóstico General.....	9
Determinación de agrupamientos para las entidades federativas.....	10
Ordenamiento de las Entidades Federativas y Grupos.....	12
Análisis de la información estatal.....	15
<b>Conclusiones.....</b>	<b>18</b>
<b>ANEXO METODOLÓGICO.....</b>	<b>20</b>
<b>Anexo 1.....</b>	<b>20</b>
Diagnóstico estatal de las variables de telecomunicaciones por entidad federativa.....	20
<b>Anexo 2.....</b>	<b>34</b>
Metodología y Análisis de <i>clústers</i> .....	34
Metodología aplicada al diseño del índice de desarrollo digital.....	38

## Introducción

Las telecomunicaciones tienen hoy en día una gran relevancia en la vida de las personas y en las actividades productivas de los países. De manera creciente su uso ha permitido ampliar y facilitar la comunicación entre los individuos, les ha dado acceso a servicios de educación, salud y financieros, así como al conocimiento y a la recreación. En lo económico, las telecomunicaciones son necesarias para promover la productividad, el crecimiento y la recuperación de la actividad productiva.

El progreso tecnológico de las telecomunicaciones es un fenómeno multifactorial que requiere de la consideración de distintos indicadores para ser capturado, por lo que su medición integral es compleja. La existencia de diferentes variables que señalan el avance de los distintos aspectos que conforman el desarrollo tecnológico dificulta establecer conclusiones y definir una estrategia de conectividad acorde con las necesidades del país. Asimismo, el progreso en materia de telecomunicaciones implica identificar la adopción efectiva de los servicios por parte de la población, ya que sólo a través de la apropiación de estos es que la digitalización se convierte en un factor multiplicador del bienestar y el crecimiento <sup>1</sup>.

El objetivo del trabajo es proveer un diagnóstico completo del avance estatal reportado, así como integrar información útil para identificar los rezagos en materia de telecomunicaciones, a partir de 21 indicadores básicos.

Cabe destacar que este ejercicio se elabora por tercer año consecutivo, permitiendo también, a través de los correspondientes reportes dar seguimiento al avance relativo de las entidades en materia de adopción de las TIC a través del tiempo (se ha realizado en los años 2020<sup>2</sup> y 2021<sup>3</sup>).

Los 21 indicadores que se usan para el análisis han sido clasificados en cinco dimensiones. Las primeras dos dimensiones son la infraestructura de telecomunicaciones y la adopción de tecnologías de la información y comunicaciones (TIC's). En ambos casos estas han sido utilizadas por la OCDE en algunas de sus publicaciones<sup>4</sup>. La tercera dimensión es la de asequibilidad de los

---

<sup>1</sup>Katz, R. & Koutroumpis, P. Measuring digitization: A growth and welfare multiplier. *Technovation* Volume 33, Issues 10-11, October–November 2013, Pages 314-319.

<sup>2</sup> <https://centrodeestudios.ift.org.mx/admin/files/indicadores/1627332327.pdf>

<sup>3</sup> <https://centrodeestudios.ift.org.mx/admin/files/indicadores/1639066089.pdf>

<sup>4</sup> Measuring the Digital Economy, A New Perspective. Diciembre, 2014. OCDE.

servicios de telecomunicaciones. Recientemente esta dimensión ha sido incorporada por Afshar Ali, Alam y Taylor<sup>5</sup> (2020) para actualizar el índice de desarrollo de telecomunicaciones de la Unión Internacional de Telecomunicaciones (UIT) llamándolo índice modificado del nivel de madurez de las TIC's, e incluyendo a su vez categorías como la eficiencia y la calidad. La siguiente categoría es la de usos de las TIC's en actividades económicas. La Comisión Europea en su publicación *The Digital Economy and Society Index (DESI) 2020* construye un índice compuesto que incluye como dimensión la integración de tecnología digital, la cual tiene como subdimensiones a la digitalización empresarial y el comercio electrónico. Por último, la categoría de los usos del internet, la cual es incorporada por Katz y Koutroumpis<sup>6</sup> en la construcción de un índice de digitalización.

Utilizando distintas técnicas estadísticas multivariadas se integran clústers o agrupamientos de las 32 entidades federativas con el objeto de identificar aquellos estados de la República Mexicana que presentan niveles similares de avance en conectividad, asequibilidad, en la adopción y uso de las tecnologías y servicios de telecomunicaciones. Adicionalmente, se construye un índice de desarrollo digital (IDD) para las entidades federativas que permite establecer un ordenamiento respecto al avance relativo de cada una. Esto es, el ejercicio identifica qué entidades reportan un desarrollo similar en cuanto a la dotación integral de los servicios de telecomunicaciones, así como en su adopción y usos, incluyendo 21 indicadores básicos.

A las dimensiones se les asigna una ponderación equitativa; en tanto que su elección busca reflejar la relevancia que tienen los ejes de política pública de telecomunicaciones y desarrollo digital establecidos por el gobierno federal, a través de la Secretaría de Infraestructura, Comunicaciones y Transportes<sup>7</sup>, así como, a partir del objetivo 3 del Programa Sectorial derivado del Plan Nacional de Desarrollo 2019-2024. En estos instrumentos de planeación se reconoce la relevancia de la promoción de infraestructuras de telecomunicaciones, del incremento de la cobertura y la disponibilidad de los servicios de telecomunicaciones y radiodifusión, así como el desarrollo de capacidades y habilidades digitales.

---

<sup>5</sup> Mohammad Afshar Ali, Khorshed Alam & Brad Taylor (2020) Incorporating affordability, efficiency, and quality in the ICT development index: Implications for index building and ICT policymaking, *The Information Society*, 36:2, 71-96, DOI: 10.1080/01972243.2019.1702601

<sup>6</sup> Katz, R. & Koutroumpis, P. *Using a digitization index to measure the economic and social impact of digital agendas*. [https://www.researchgate.net/publication/270806225\\_Using\\_a\\_digitization\\_index\\_to\\_measure\\_the\\_economic\\_and\\_social\\_impact\\_of\\_digital\\_agendas](https://www.researchgate.net/publication/270806225_Using_a_digitization_index_to_measure_the_economic_and_social_impact_of_digital_agendas)

<sup>7</sup> Página 9, [https://www.dof.gob.mx/nota\\_detalle.php?codigo=5596042&fecha=02/07/2020](https://www.dof.gob.mx/nota_detalle.php?codigo=5596042&fecha=02/07/2020)

En efecto, el desarrollo de capacidades y habilidades digitales es importante porque se vincula con la obtención de empleos mejor remunerados (Pirzada y Khan<sup>8</sup>), o con el acceso a más y mejor información (Van Deursen & Van Dijk<sup>9</sup>). No obstante, en México existen aún millones de personas que no tienen acceso a Internet; la mayor parte de ellas residen en zonas urbanas donde las carencias económicas frenan la conectividad<sup>10</sup>. Es así, que en función de los ejes de política pública y de los retos a los cuales se enfrenta actualmente el país, se opta por asignar una ponderación equitativa a las cinco dimensiones; es decir, de 20%.

Los resultados de este análisis multifactorial sobre el avance del sector de las telecomunicaciones a nivel estatal permiten establecer seis grupos, cada uno con niveles similares de infraestructura de telecomunicaciones, así como de su uso, asequibilidad y adopción. A partir del ordenamiento obtenido de los agrupamientos y del índice de desarrollo digital se concluye la Ciudad de México es la entidad federativa que muestra un mayor avance en el conjunto de los distintos servicios de telecomunicaciones con respecto al resto de entidades. En el otro extremo se encuentran Chiapas y Oaxaca.

El reporte corrobora que los estados con menor nivel de vida son también los que tienen menor equipamiento y adopción de los servicios de telecomunicaciones.

El trabajo busca proveer al Instituto Federal de Telecomunicaciones (IFT) con un insumo útil que permita identificar brechas entre las treinta y dos entidades federativas, a fin de generar información analítica útil para el diseño e implementación de políticas públicas que acorten esas diferencias.

## Información estadística y consideraciones metodológicas

Para el ejercicio que se presenta en este reporte, se utilizan veintiún indicadores del sector de las telecomunicaciones diferenciados en cinco categorías o dimensiones, a saber: Infraestructura de TIC´s; adopción de TIC´s; asequibilidad; usos de TIC´s en la actividad económica; usos del internet que realiza la población de cada entidad federativa. Estas categorías sirven para agrupar correctamente a las variables, y, a la vez, adjudicarles un peso ponderado el cual es de utilidad al

---

<sup>8</sup> Pirzada, K. & Khan F. Measuring Relationship between Digital Skills and Employability. European Journal of Business and Management [www.iiste.org](http://www.iiste.org) ISSN 2222-1905 (Paper) ISSN 2222-2839 (Online). Vol.5, No.24, 2013.

<sup>9</sup> Van Deursen, A. & Van Dijk J. Improving digital skills for the use of online public information and services. Government Information Quarterly. Volume 26, Issue 2, April 2009, Pages 333-340.

<sup>10</sup> Ver ENDUTIH 2019. <https://www.inegi.org.mx/programas/dutih/2019/>

momento de construir el índice de bienestar digital. A continuación, se explica a cada una de dichas categorías y los indicadores que las integran.

## 1. Infraestructura de TIC's.

Esta categoría incluye indicadores que miden la extensión con la cual la población tiene acceso a infraestructura de telecomunicaciones, ya sea que se trate de redes fijas, móviles o de TV restringida. En este sentido se utilizan dos mediciones principales: la penetración y la teledensidad de las TIC's. A esta categoría se le asigna una ponderación de 20% e incluye siete indicadores. La fuente de información es el Banco de Información en Telecomunicaciones (BIT) del Instituto Federal de Telecomunicaciones<sup>11</sup>; en el caso del porcentaje de población cubierta por tecnología de red 4G para servicio móvil la fuente es el Sistema de Análisis de Telecomunicaciones y Conectividad (SATYC)<sup>12</sup>.

- I. Penetración de telefonía fija (BIT)
- II. Penetración de banda ancha fija (BIT)
- III. Penetración de TV restringida (BIT)
- IV. Teledensidad de telefonía móvil (BIT)
- V. Teledensidad de internet móvil (BIT)
- VI. Penetración de banda ancha fija con acceso de fibra óptica (BIT)
- VII. Porcentaje de población cubierta por tecnología de red 4G para servicio móvil (SATYC)

## 2. Adopción de TIC's.

La segunda dimensión<sup>13</sup> es la adopción de TIC's, la cual refleja tanto el equipamiento con el que cuentan las personas en sus hogares, como la implementación exitosa de dichas tecnologías, lo que permite expresar su utilización. A esta dimensión se le asigna una ponderación del 20%. La fuente de

---

<sup>11</sup> <https://bit.ift.org.mx/BitWebApp/>

<sup>12</sup> <http://satyc.ift.org.mx:8010/connect/analyst/mobile/#/login>

<sup>13</sup> *Idem.*

información de todas las variables es la Encuesta Nacional sobre Disponibilidad y Uso de Tecnologías de la Información en los Hogares (ENDUTIH) 2021<sup>14</sup>, del INEGI.

- VIII. Porcentaje de usuarios que cuentan con teléfono celular.
- IX. Porcentaje de usuarios que cuentan con Smartphone.
- X. Porcentaje de usuarios de Internet que lo utilizan diariamente.
- XI. Porcentaje de usuarios de teléfono celular inteligente con conexión móvil a Internet.

### 3. Asequibilidad

La tercera dimensión es la de asequibilidad, y se refiere a la capacidad que tienen los usuarios de adquirir algún tipo de servicio (por ejemplo, un plan tarifario de pospago) o de gastar en algún servicio por encima de una cantidad mínima. La ONU recomienda un límite en el precio promedio del paquete de no más del 2% del ingreso del hogar. Para este reporte los indicadores se han definido de manera positiva, es decir, denotando que se encuentra en una situación favorable respecto al indicador en lo particular. Así para fines de integrar esta dimensión se consideró el porcentaje de usuarios de telefonía celular que gastan \$200 pesos, o más, al mes. Además, se incluye el porcentaje de usuarios de telefonía celular con plan tarifario pospago que integra a la población que dispone de capacidad crediticia y normalmente, con una mayor capacidad de consumo.

Así, esta dimensión únicamente incluye dos indicadores, y se le asigna una ponderación del 20%. La fuente de información es la ENDUTIH 2021, del INEGI.

- XII. Porcentaje de usuarios de telefonía celular con plan tarifario (pospago).
- XIII. Porcentaje de usuarios de telefonía celular que gastan \$200 pesos, o más, al mes.

### 4. Usos de TIC's en la actividad económica.

La cuarta dimensión es la de usos de TIC's en la actividad económica. Lo que busca esta dimensión es reflejar el dominio que tienen los usuarios de la tecnología y el acceso a la red. Es decir, no basta simplemente con que las personas tengan una

<sup>14</sup> <https://www.inegi.org.mx/programas/dutih/2021/>

computadora o un teléfono inteligente conectados a internet para considerar a dicha persona como un incluido digitalmente, se precisa saber qué hacer con estas tecnologías. En este sentido es que se consideran indicadores como el porcentaje de usuarios que realizan pagos o compras por internet, o el porcentaje de usuarios que utilizan computadora para realizar sus actividades laborales. A esta categoría se le asigna una ponderación del 20%. La fuente de información de todas las variables es la ENDUTIH 2020, del INEGI.

- XIV. Porcentaje de usuarios que realizan compras por internet
- XV. Porcentaje de usuarios de Internet que realizan operaciones bancarias en línea.
- XVI. Porcentaje de usuarios que usan computadora para actividades laborales

## 5. Usos de internet.

La última categoría es la de los usos del internet, la cual es preciso incluir para definir el desarrollo digital con el que cuentan las entidades. En este sentido, se consideran variables como: el porcentaje de usuarios que lo utilizan para comunicación o para obtener información. A esta categoría se le asigna una ponderación del 20%. La fuente de información de todas las variables es la ENDUTIH 2020, del INEGI.

- XVII. Porcentaje de usuarios internet que lo utilizan para entretenimiento.
- XVIII. Porcentaje de usuarios de internet que lo utilizan para ver contenidos audiovisuales.
- XIX. Porcentaje de usuarios de internet que lo utilizan para comunicación.
- XX. Porcentaje de usuarios de internet que lo utilizan para obtener información.
- XXI. Porcentaje de usuarios que usan internet para realizar interacciones con el gobierno.

La muestra se integra por las treinta y dos entidades federativas de México. Los indicadores cuya fuente es el BIT se encuentran al cuarto trimestre del año 2021. Los indicadores cuya fuente es la ENDUTIH o el SACYT también se encuentran al cuarto trimestre del año 2021.

## Diagnóstico General

Los veintidós indicadores dan cuenta del desarrollo digital de las entidades federativas de México y son base del análisis de este estudio. El cuadro 1 siguiente muestra el promedio nacional y la desviación estándar que se registra en cada uno de estos indicadores. La dispersión más alta es la que presenta el indicador de penetración de telefonía fija (26), las penetraciones de banda ancha fija (18) y de TV restringida (16). Esto significa que la disparidad entre los valores de dichas variables es elevada.

**Cuadro 1.** Indicadores de desarrollo digital. Promedio y desviación estándar de los indicadores

N°	Dimensión	Nombre del indicador	Media	Desviación Estándar
1	1	Penetración telefonía fija	60	26
2	1	Penetración banda ancha fija	58	18
3	1	Penetración TV restringida	72	16
4	1	Teledensidad de telefonía móvil	100	10
5	1	Teledensidad de internet móvil	79	12
6	1	Penetración de banda ancha fija con accesos de fibra óptica	22	11
7	1	Porcentaje de población cubierta por tecnología de red 4G para servicio móvil	88	9
8	2	Porcentaje de usuarios que cuentan con teléfono celular	79	7
9	2	Porcentaje de usuarios que cuentan con smartphone	93	2
10	2	Porcentaje de usuarios de internet que lo utilizan diariamente	89	5
11	2	Porcentaje de usuarios de teléfono celular inteligente con conexión móvil a internet	87	4
12	3	Porcentaje de usuarios de telefonía celular con plan tarifario de pospago	15	5
13	3	Porcentaje de usuarios de telefonía celular que gastan \$200 pesos, o más, al mes	39	3
14	4	Porcentaje de usuarios de internet que realizan compras	31	6

15	4	Porcentaje de usuarios de internet que realizan operaciones bancarias en línea	24	5
16	4	Porcentaje de usuarios de usan computadora para actividades laborales	48	3
17	5	Porcentaje de usuarios de internet que lo utilizan para entretenimiento	89	3
18	5	Porcentaje de usuarios de internet que lo utilizan para ver contenidos audiovisuales	78	5
19	5	Porcentaje de usuarios de internet que lo utilizan para comunicarse	90	2
20	5	Porcentaje de usuarios de internet que lo utilizan para buscar información	90	2
21	5	Porcentaje de usuarios de internet que lo usan para realizar interacciones con el gobierno	41	7

En el Anexo 1 de este reporte se presenta una descripción detallada de las cifras estatales para cada indicador.

### Determinación de agrupamientos para las entidades federativas

Para integrar grupos de entidades federativas que se asemejan en cuanto a los veintiún indicadores de telecomunicaciones seleccionados y presentados en el Cuadro 1, se utiliza la técnica de partición conocida como método Ward<sup>15</sup>. La técnica permite identificar a los estados del país que son relativamente homogéneos entre sí, con base en la disponibilidad, adopción, asequibilidad y usos actuales de los servicios de telecomunicaciones de manera simultánea. Esto es, considerando los veintiún indicadores de manera simultánea se forman grupos

<sup>15</sup> El método de Ward es un procedimiento jerárquico en el cual, en cada etapa, se unen los dos clusters para los cuales se tenga el menor incremento en el valor total de la suma de los cuadrados de las diferencias, dentro de cada cluster, de cada individuo al centroide del cluster. Para una descripción detallada, véase: Peña, D., 2002. Análisis de Datos Multivariados, McGraw Hill.

integrados por entidades federativas con la mayor similitud entre sí, pero diferentes respecto a otros grupos definidos.

En el análisis se emplea una segunda técnica de agrupamiento conocida como K-medias<sup>16</sup>. El empleo de dos técnicas diferentes permite contrastar los resultados y tener mayor solidez en las conclusiones. Para una mayor descripción véase el Anexo 2 de este reporte.

A partir de la aplicación de este ejercicio estadístico, se definieron seis grupos o clústers como el número óptimo a considerar. Para llegar a esta conclusión se utiliza una gráfica conocida como Screeplot, (Gráfica A20 del Anexo Metodológico), que muestra el número óptimo de agrupamientos.

En función de los resultados que producen las dos técnicas, se consideró que el método de Ward utilizando la distancia Euclidiana es el más adecuado para el análisis, ya que otorgó resultados más robustos. El Cuadro 2 presenta los agrupamientos definidos.

Cuadro 2. Agrupamientos

Agrupamiento	Entidades Federativas
1	CDMX
2	Baja California, Baja California Sur, Colima, Querétaro
3	Nayarit, Sonora, Sinaloa, Tamaulipas
4	Aguascalientes, Chihuahua, Coahuila, Guanajuato, Jalisco, México, Morelos, Nuevo León, Quintana Roo, San Luis Potosí, Yucatán
5	Campeche, Durango, Guerrero, Hidalgo, Michoacán, Puebla, Tabasco, Tlaxcala, Veracruz, Zacatecas
6	Chiapas, Oaxaca

La metodología empleada permite establecer que los estados que integran cada grupo o clúster exhiben niveles similares en la penetración de los diferentes servicios de telecomunicaciones fijas y móviles, y en las variables de inclusión digital y de asequibilidad con base en los veintiún indicadores considerados simultáneamente. En este caso la Ciudad de México no comparte un nivel similar con ninguna otra entidad ya que queda aislada en su propio grupo. Por otra parte, Baja California,

<sup>16</sup> El método de K-medias tiene como objetivo generar una partición de un conjunto de  $n$  observaciones en  $k$  grupos. Cada grupo está representado por el promedio de los puntos que lo componen. El representante de cada grupo se denomina centroide. La cantidad de grupos a descubrir,  $k$ , es un parámetro que se debe fijar a priori. El método de clustering comienza con  $k$  centroides ubicados de forma aleatoria, y asigna cada observación al centroide más cercano. Después de asignarlos, los centroides se mueven a la ubicación promedio de todos los datos asignados a él, y se vuelven a reasignar los puntos de acuerdo a las nuevas posiciones de los centroides.

Baja California Sur, Colima y Querétaro quedan en un mismo grupo al considerar el agregado de los indicadores. En tanto que Chiapas y Oaxaca forman su propio grupo. Esto denota que este par de entidades no comparten un desarrollo en materia de TIC's que sea similar respecto a otras entidades.

Cabe destacar que las técnicas de agrupamiento permiten integrar en grupos a las entidades federativas con niveles similares de avance o rezago, pero no dan un orden en cuanto al mayor o menor desarrollo alcanzado.

A continuación, se estiman los índices que permiten determinar cuáles son los grupos o clústers, y las entidades federativas, que presentan un mayor avance respecto a su desarrollo digital.

## Ordenamiento de las Entidades Federativas y Grupos

Para ordenar los agrupamientos con base en los indicadores elegidos se estiman para cada entidad federativa dos métricas alternativas: un índice de variables estandarizadas (IVE) basado en la suma de las veintiún variables estandarizadas<sup>17</sup>; y un índice de desarrollo digital (IDD) para cada grupo.

El IVE se define como la suma de las veintiún variables estandarizadas; éste se emplea para estimar un índice para cada grupo. El IVE toma mayor valor en la medida que el nivel de equipamiento, asequibilidad, adopción y uso de los servicios de TIC's del agrupamiento es más elevado. Para facilitar la lectura, el índice se estima en una escala de valores que va de 1 a 100.<sup>18</sup>

De manera adicional, en este reporte se procede a la elaboración del IDD. El objetivo es contar con una herramienta que ofrezca una alternativa para clasificar a las entidades. Este se estima para cada grupo (Cuadro 2); y en la siguiente sección se presentan el IDD para cada entidad federativa (Cuadro 3). Para ello se utiliza la técnica de análisis de componentes principales<sup>19</sup>, la cual es de gran utilidad cuando se trabaja con bases de datos multivariadas, como lo es este caso. El análisis de los componentes principales permite observar en un solo índice la estructura de los 21 indicadores en su totalidad. La matemática subyacente al tema del análisis de los componentes principales es compleja, por lo tanto, no se aborda en el presente documento<sup>20</sup>, pero la esencia es la siguiente: se cuenta con un conjunto de datos con muchos indicadores o variables, es decir, multivariado,

---

<sup>17</sup> La estandarización consiste en una transformación lineal de las veintiuna variables con objeto de que tengan un promedio de cero y una desviación estándar igual a uno. Este proceso es necesario para poder sumar variables expresadas en unidades diferentes.

<sup>18</sup> Se hace obteniendo las diferencias entre el valor obtenido por el grupo 1 con respecto a cada uno de los cinco grupos restantes; posteriormente, a cada valor resultante se le resta a 100 para obtener el re escalamiento final.

<sup>19</sup> Peña, D., 2002. Análisis de Datos Multivariantes, McGraw Hill.

<sup>20</sup> *Ibidem*

el cual es simplificado convirtiendo a las variables originales en un número menor de “componentes principales”. Es así como los componentes principales representan la estructura subyacente de los datos<sup>21</sup>. En función de lo anterior, el diseño del IDD conlleva la aplicación de ponderadores a cada una de las cinco dimensiones definidas, para después construir el índice agregando los resultados. Esta metodología es la misma que aplica The Economist Intelligence Unit en su E-readiness rankings 2009<sup>22</sup>.

Contar con dos índices diferentes para el ordenamiento de las entidades federativas da solidez al ejercicio, además de que permite comparar los resultados que se obtienen con dos diferentes técnicas de ordenamiento, una ponderada y otra no ponderada.

El Cuadro 2 presenta los índices estimados, a partir de los cuales se ordenan los 6 grupos previamente definidos. Como puede constatarse, básicamente los dos índices ofrecen un mismo ordenamiento. Se desprende también, que tanto un índice no ponderado<sup>23</sup> (IVE) como uno ponderado (IDD) dan un resultado similar.

Cuadro 2. Entidades federativas según agrupamiento

Orden de agrupamiento	Estados	Índice de variables estandarizadas	Índice de Desarrollo Digital	IDH
1	CDMX	100	2.52	0.828
2	Baja California, Baja California Sur, Colima, Querétaro	89	1.36	0.732
3	Aguascalientes, Chihuahua, Coahuila, Guanajuato, Jalisco, México, Morelos, Nuevo León, Quintana Roo, San Luis Potosí, Yucatán	74	0.76	0.687
4	Nayarit, Sonora, Sinaloa, Tamaulipas	68	0.08	0.704

<sup>21</sup> Los componentes principales determinan la dirección o direcciones hacia la cual se encuentra la mayor varianza de los datos considerados. Por tanto, lo que se trata es de encontrar la línea recta que mejor se extienda a través de los datos al momento de ser proyectada. Este sería el primer componente principal. Para mayor explicación, véase: Peña, D., 2002. Análisis de Datos Multivariantes, McGraw Hill.

<sup>22</sup> E-readiness rankings 2009. A report from the Economist Intelligence Unit. The Economist. <http://graphics.eiu.com/pdf/e-readiness%20rankings.pdf>

<sup>23</sup> El índice de variables estandarizadas no usa explícitamente un ponderador ya que se define como un promedio aritmético de las variables estandarizadas. Por lo anterior, se da igual peso a cada una de las 19 variables.

5	Campeche, Durango, Guerrero, Hidalgo, Michoacán, Puebla, Tabasco, Tlaxcala, Veracruz, Zacatecas	53	-1.03	0.665
6	Chiapas, Oaxaca	34	-3.21	0.596

A partir del ordenamiento obtenido, se tiene que la Ciudad de México forma un *clúster* en solitario, y mantiene los índices más altos; con un valor igual a 100 para el IVE y 0.828 tratándose del IDD. El segundo *clúster* conformado por Baja California, Baja California Sur, Colima y Querétaro tiene un IVE igual a 89 y un IDD igual a 1.36. En ese orden sigue el grupo de Aguascalientes, Chihuahua, Coahuila, Guanajuato, Jalisco, México, Morelos, Nuevo León, Quintana Roo, San Luis Potosí y Yucatán con un IVE de 74 y un IDD igual a 0.76. El Mapa 1 resume los agrupamientos.

A continuación, el grupo conformado por Nayarit, Sonora, Sinaloa y Tamaulipas presenta un IVE igual a 68, y un IDD de 0.08. Por su parte, Campeche, Durango, Guerrero, Hidalgo, Michoacán, Puebla, Tabasco, Tlaxcala, Veracruz y Zacatecas con un IDD de -1.03 se encuentra arriba de Chiapas y Oaxaca que tienen un IDD de -3.21

Aunado a lo anterior, el análisis se enriquece si se toma en consideración un indicador del nivel de bienestar social de la población. Específicamente se emplea en este reporte el Índice de Desarrollo Humano del Programa de Naciones Unidas para el Desarrollo<sup>24</sup> (IDH, por sus siglas) para las entidades federativas de México, el cual mide el nivel de desarrollo de cada entidad federativa atendiendo a variables como la esperanza de vida, la educación o el ingreso per cápita.

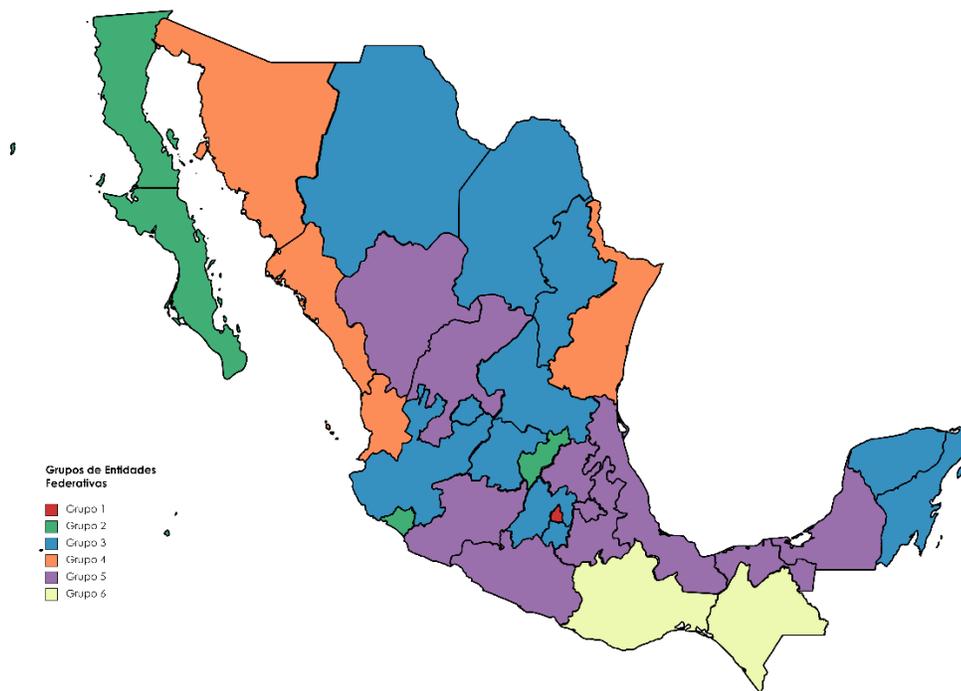
Para cada agrupamiento se obtiene el promedio del IDH de las entidades federativas que integran a cada *clúster* (ver Cuadro 2). El primer *clúster* conformado únicamente por la Ciudad de México alcanza un IDH más alto con 0.828. Por su parte, el segundo *clúster* integrado por Baja California, Baja California Sur, Colima y Querétaro tiene un IDH promedio igual 0.732. El grupo formado por Aguascalientes, Chihuahua, Coahuila, Guanajuato, Jalisco, México, Morelos, Nuevo León, Quintana Roo, San Luis Potosí y Yucatán tiene un IDH promedio igual a 0.687.

El grupo formado por Nayarit, Sonora, Sinaloa y Tamaulipas tiene un IDH promedio igual a 0.704. Le sigue el *clúster* que agrupa a los estados de Campeche, Durango,

<sup>24</sup> [https://es.wikipedia.org/wiki/Anexo:Entidades\\_federativas\\_de\\_M%C3%A9xico\\_por\\_IDH](https://es.wikipedia.org/wiki/Anexo:Entidades_federativas_de_M%C3%A9xico_por_IDH)

Guerrero, Hidalgo, Michoacán, Puebla, Tabasco, Tlaxcala, Veracruz y Zacatecas con un IDH promedio de 0.665. Por último, Chiapas y Oaxaca presentan un IDH promedio igual a 0.596.

Las cifras sugieren que el nivel de equipamiento de infraestructura de telecomunicaciones y la adopción de los servicios es mayor en los grupos de entidades que también presentan un mayor nivel de vida, de acuerdo con el IDH.



Mapa 1. Entidades Federativas de México por Grupos.

## Análisis de la información estatal

Para completar el panorama presentado a nivel agrupamiento, se estimó el índice de desarrollo digital o IDD para cada una de las entidades federativas. Estas cifras se presentan en el Cuadro 3, en el que también se incluyen los valores para cada una de las cinco dimensiones que ponderadas integran el IDD. Tal y como se puede

observar la Ciudad de México reporta el IDD más elevado con un índice de 2.52; sin embargo, la CDMX no lidera en todas y cada una de las cinco dimensiones. Por ejemplo, para la dimensión 4, Uso de TIC's en actividades económicas, Baja California Sur (2.72) y Colima (1.72) reportan mejores resultados que CDMX (1.68). En el índice global a CDMX le siguen Colima (1.46) y Nuevo León (1.46). Por su parte, las entidades peor posicionadas son Oaxaca (-3.05) y Oaxaca (-3.36).

Cuadro 3. Estimación estatal del índice de Desarrollo Digital

Posición	Entidad	Dimensión 1. Infraestructura de TIC's (20%)	Dimensión 2. Adopción de TIC's (20%)	Dimensión 3. Asequibilidad (20%)	Dimensión 4. Uso de TIC's en actividades económicas (20%)	Dimensión 5. Usos de Internet (20%)	Índice de Desarrollo Dígital (IDD)
1	CDMX	5.17	1.96	1.89	1.68	1.89	2.52
2	COL	1.25	1.34	1.95	1.72	1.03	1.46
3	NL	2.86	1.31	1.06	1.39	0.64	1.46
4	BCS	1.51	1.51	1.78	2.72	-0.66	1.38
5	QRO	2.47	0.72	0.38	1.34	1.81	1.35
6	BC	1.65	1.72	1.28	0.26	1.37	1.26
7	CHIH	1.07	0.65	1.21	2.26	0.80	1.20
8	JAL	1.88	1.55	0.38	0.55	0.61	1.00
9	YUC	0.26	0.48	0.73	1.00	2.48	1.00
10	COAH	0.98	0.80	0.67	0.70	1.79	0.99
11	QROO	0.50	1.20	0.53	2.32	-0.66	0.78
12	MEX	1.21	0.93	0.68	0.03	0.60	0.70
13	AGS	0.70	0.52	0.47	0.08	1.47	0.65
14	SON	1.56	1.26	0.33	-0.41	0.23	0.60
15	MOR	1.29	0.16	-0.00	0.78	0.63	0.57
16	TAM	0.79	1.05	-0.79	-0.11	0.56	0.30
17	SLP	-0.29	0.63	-0.57	0.28	0.74	0.16
18	SIN	1.38	0.81	-0.13	-0.51	-0.81	0.15
19	GTO	-0.04	-0.34	0.89	-0.71	-0.40	-0.12
20	CAM	-1.78	-0.704	-0.13	0.76	1.03	-0.17
21	PUE	-1.08	-0.89	0.12	-1.45	1.24	-0.41
22	HGO	-1.35	-0.79	-0.46	-2.07	2.04	-0.53
23	NAY	-0.08	0.87	-1.54	-1.44	-1.49	-0.74
24	TLAX	-0.38	-0.03	-1.62	-2.41	0.04	-0.88
25	DGO	-1.10	-1.14	-1.01	0.09	-2.13	-1.06
26	TAB	-1.94	-0.52	-1.40	-1.26	-0.47	-1.12

27	MICH	-1.70	-0.98	-0.59	0.30	-2.76	-1.15
28	ZAC	-2.07	-0.99	-0.68	-1.56	-1.57	-1.38
29	VER	-1.67	-1.52	-2.43	-0.75	-1.51	-1.58
30	GRO	-2.77	-2.47	-1.79	-1.95	-0.88	-1.98
31	OAX	-4.65	-4.74	-1.36	-1.65	-2.85	-3.05
32	CHIS	-5.65	-4.42	0.17	-2.03	-4.84	-3.36

## Conclusiones

El presente Reporte permite obtener una perspectiva del desarrollo en materia de telecomunicaciones de las 32 entidades federativas de México. Mediante la realización de un análisis de clústers y la construcción de un índice de desarrollo digital se clasifican las entidades de acuerdo con el avance alcanzado a partir de veintiún indicadores de telecomunicaciones clasificados en cinco dimensiones básicas: la infraestructura de telecomunicaciones; la adopción de TIC's; la asequibilidad, el uso de TIC's en actividades económicas, y los usos del internet. A estas dimensiones se les pondera de manera equitativa con un 20% a cada una. También se toman en consideración los ejes de política pública establecidos por el gobierno federal<sup>25</sup>, los cuales incluyen en sus programas la promoción de infraestructuras, el incremento de la cobertura y la disponibilidad de los servicios de telecomunicaciones y radiodifusión, así como el desarrollo de habilidades y capacidades digitales.

El análisis de clústers conlleva la aplicación de dos técnicas de agrupamiento: el método de K-medias y el método Ward. Se opta por seleccionar seis clústers o agrupamientos. Por otra parte, el diseño del índice de desarrollo digital se lleva a cabo mediante el análisis de componentes principales.

El reporte permite establecer que la Ciudad de México es la entidad con el desarrollo digital más elevado, seguida de los estados de Baja California, Baja California Sur, Colima y Querétaro los cuales conforman un grupo ya que cuentan con un desarrollo digital similar. Los resultados se estiman en base al método de K-medias (ver Cuadro 2). Este resultado se corrobora con la estimación del índice de desarrollo digital que resulta más elevado para las entidades antes señaladas (ver Cuadros 2 y 3).

Por otra parte, los estados de Oaxaca y Chiapas se encuentran a la zaga en el índice de desarrollo digital (ver Cuadros 2 y 3) y consistentemente conforman el grupo más rezagado de entre los seis que se clasifican.

Por último, se reconoce que la creciente importancia de las telecomunicaciones en las actividades económicas y en el nivel de vida de las personas hace imperativo el continuar por la senda del crecimiento y la inversión en el sector. Es así como con base en los resultados alcanzados en el presente estudio se

---

<sup>25</sup> [https://www.dof.gob.mx/nota\\_detalle.php?codigo=5596042&fecha=02/07/2020](https://www.dof.gob.mx/nota_detalle.php?codigo=5596042&fecha=02/07/2020)

recomienda continuar el diseño de políticas públicas y regulación para impulsar un mayor crecimiento en la penetración de los servicios de las telecomunicaciones.

De manera prioritaria se deberá atender el equipamiento y los esfuerzos de inclusión digital en las entidades que conforman el grupo 6.

# ANEXO METODOLÓGICO

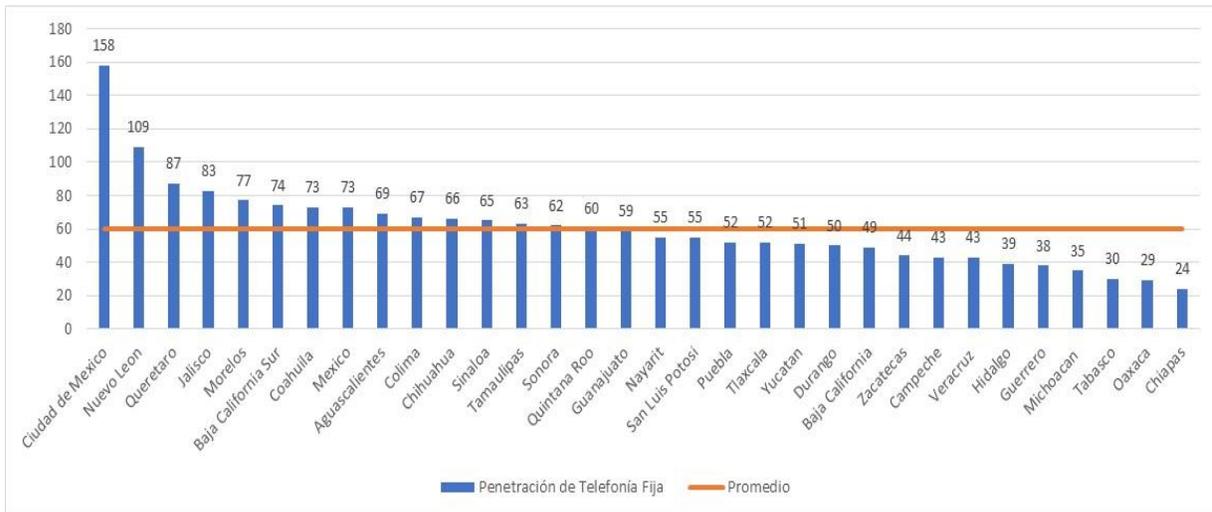
## Anexo 1.

### Diagnóstico estatal de las variables de telecomunicaciones por entidad federativa.

El objetivo del presente trabajo es identificar las similitudes o disimilitudes que guardan las treinta y dos entidades federativas del país, con respecto al acceso y adopción por parte de su población de las tecnologías de la información y comunicaciones; así como de la asequibilidad, la inclusión digital o los usos del internet. Los veintiún indicadores de telecomunicaciones que se utilizan en el presente reporte se analizan de forma gráfica a continuación.

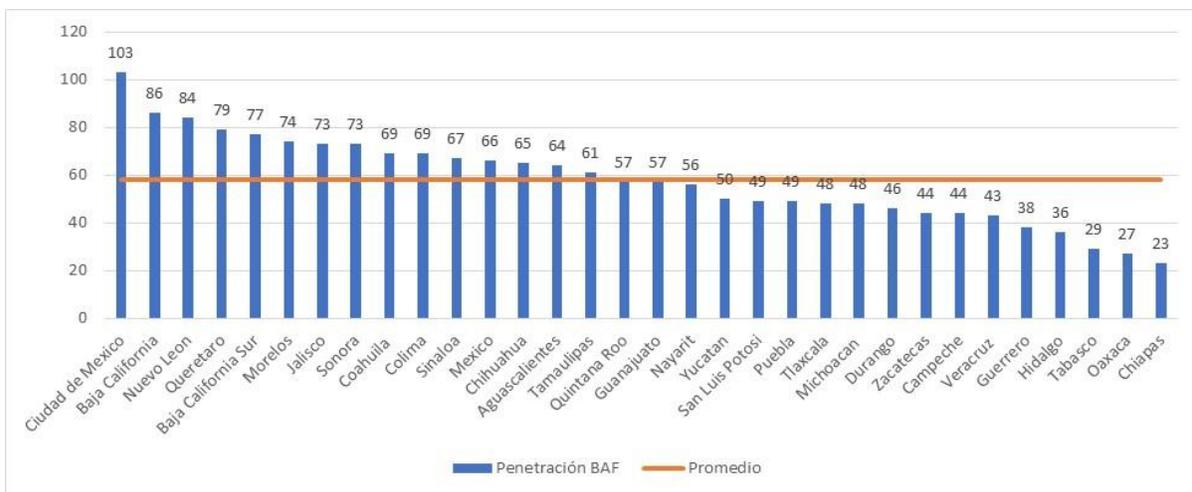
La Gráfica A1 muestra la penetración de telefonía fija por hogares para cada una de las treinta y dos entidades federativas del país. La CDMX presenta una penetración de 158 líneas telefónicas por cada 100 hogares, seguido de Nuevo León con 109 y Querétaro con 87. Por otra parte, los estados de Tabasco, Oaxaca y Chiapas acusan el mayor rezago con 30, 29 y 24 líneas de telefonía fija por cada 100 hogares. Estas cifras corroboran la amplitud de la brecha digital regional: la CDMX tiene casi ocho veces la penetración reportada en Chiapas. La media de las 32 entidades federativas está representada mediante una línea horizontal, siendo igual a 60.

Gráfica A1. Penetración de telefonía fija



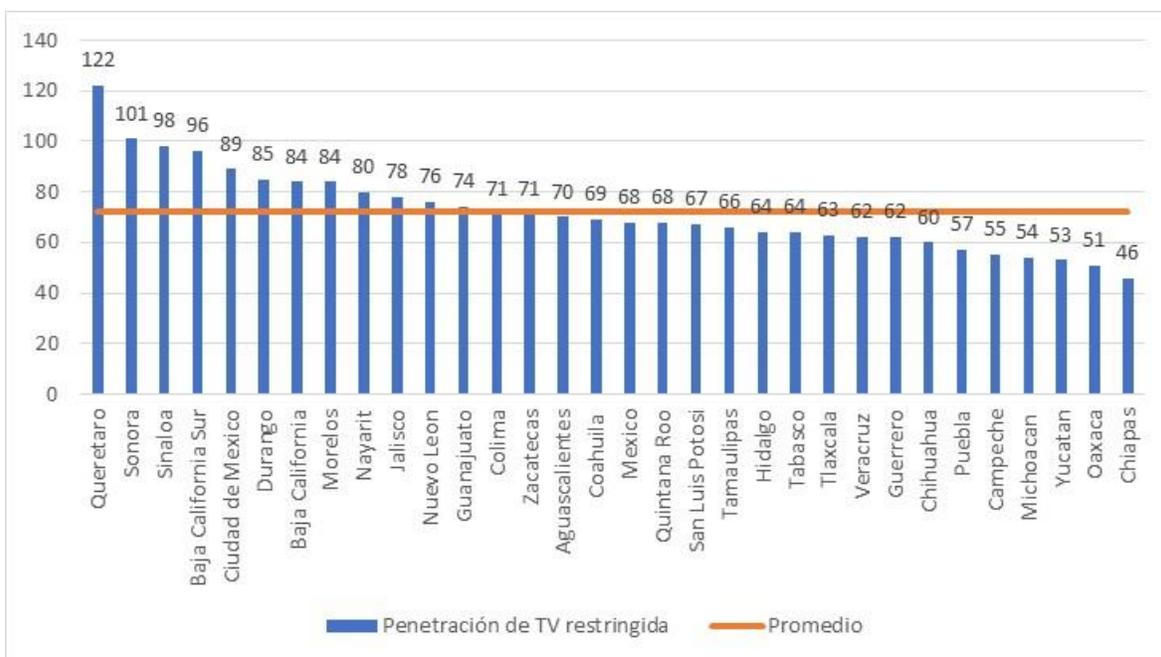
Por su parte, Gráfica A2 muestra la variable de penetración de banda ancha fija por cada 100 hogares. Esta variable muestra un panorama similar a la variable anterior; es decir, la Ciudad de México se encuentra en el primer sitio con 103 accesos de banda ancha fija por cada 100 hogares, le siguen los estados de Baja California y Nuevo León con 86 y 84 accesos por cada 100 hogares, respectivamente. Las últimas tres posiciones se encuentran ocupadas por Tabasco, Oaxaca y Chiapas con 29, 27 y 23 accesos de banda ancha fija por cada 100 hogares, respectivamente. La media nacional es igual a 58 accesos de banda ancha fija por cada 100 hogares.

Gráfica A2. Penetración de banda ancha fija



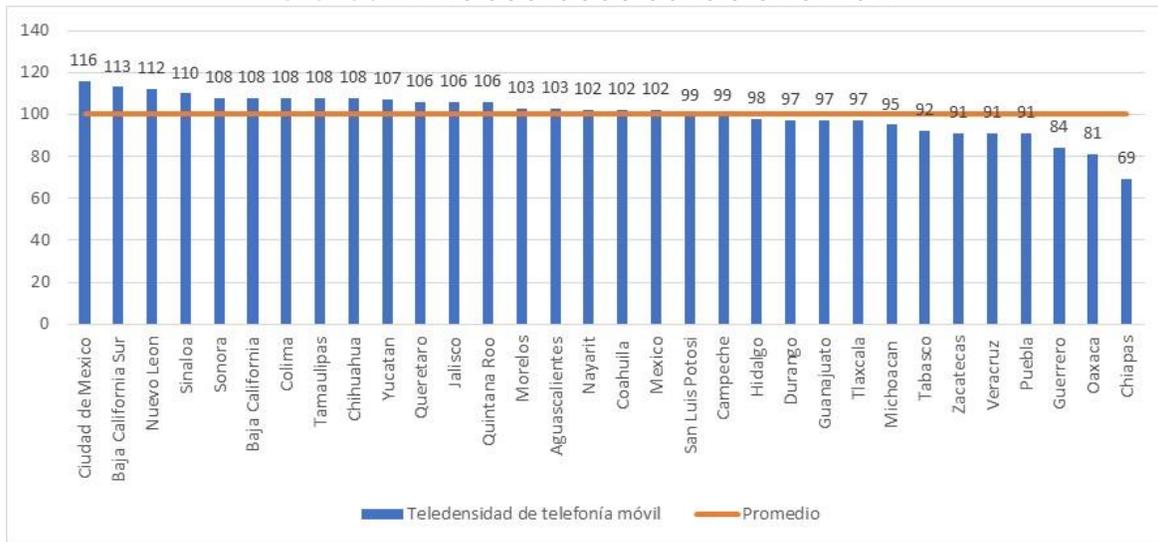
La gráfica A3 corresponde a la penetración de TV restringida por cada 100 hogares. La primera posición la ocupa el estado de Querétaro con 122 suscripciones a TV restringida por cada 100 hogares. A continuación, se encuentran los estados de Sonora y Sinaloa con 101 y 98 suscripciones por cada 100 hogares, respectivamente. Los menores niveles de penetración los acusan los estados de Yucatán, Oaxaca y Chiapas con 53, 51 y 46 suscripciones por cada 100 hogares, respectivamente. Por otra parte, la media nacional corresponde a 72 suscripciones de TV restringida por cada 100 hogares.

Gráfica A3. Penetración de TV restringida



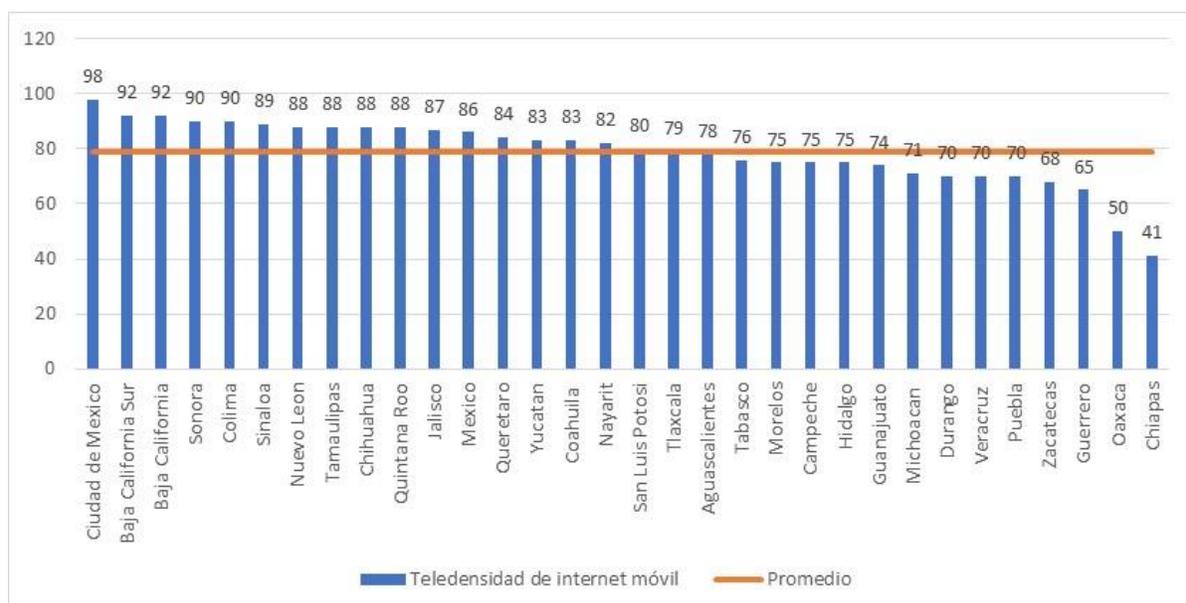
En la Gráfica A4 se observa la teledensidad de telefonía móvil por cada 100 habitantes. El primer sitio es para CDMX con 116 líneas de telefonía móvil por cada 100 habitantes; le siguen Baja California Sur y Nuevo León con 113 y 112 líneas de telefonía móvil por cada 100 habitantes. En el otro extremo se encuentra Guerrero, Oaxaca y Chiapas con 84, 81 y 69 líneas de telefonía móvil por cada 100 habitantes. La media nacional es igual a 100 líneas de telefonía de móvil por cada 100 habitantes; es decir, una por cada habitante del país.

Gráfica A4. Teledensidad de telefonía móvil



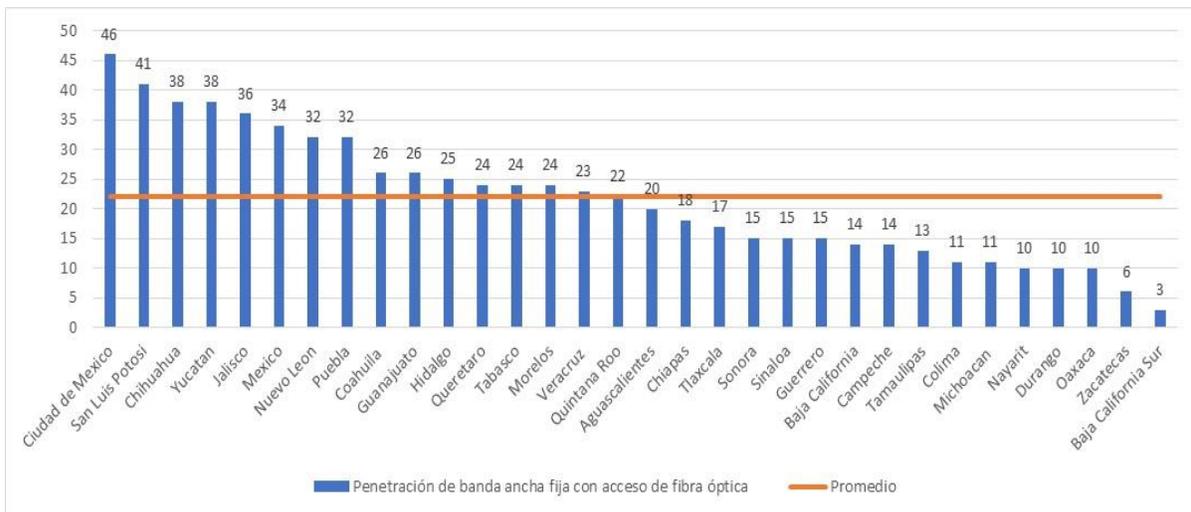
La gráfica A5 muestra la variable de teledensidad de internet móvil por cada 100 habitantes. El primer sitio lo ocupa CDMX con 98 suscripciones a internet móvil por cada 100 habitantes, seguido de Baja California Sur y Baja California con 92 suscripciones por cada 100 habitantes cada uno. Guerrero, Oaxaca y Chiapas son los estados con menores cifras de teledensidad de internet móvil con 65, 50 y 41 suscripciones por cada 100 habitantes, respectivamente. En tanto que el promedio nacional se encuentra en 79 suscripciones a internet móvil por cada 100 habitantes.

Gráfica A5. Teledensidad de internet móvil



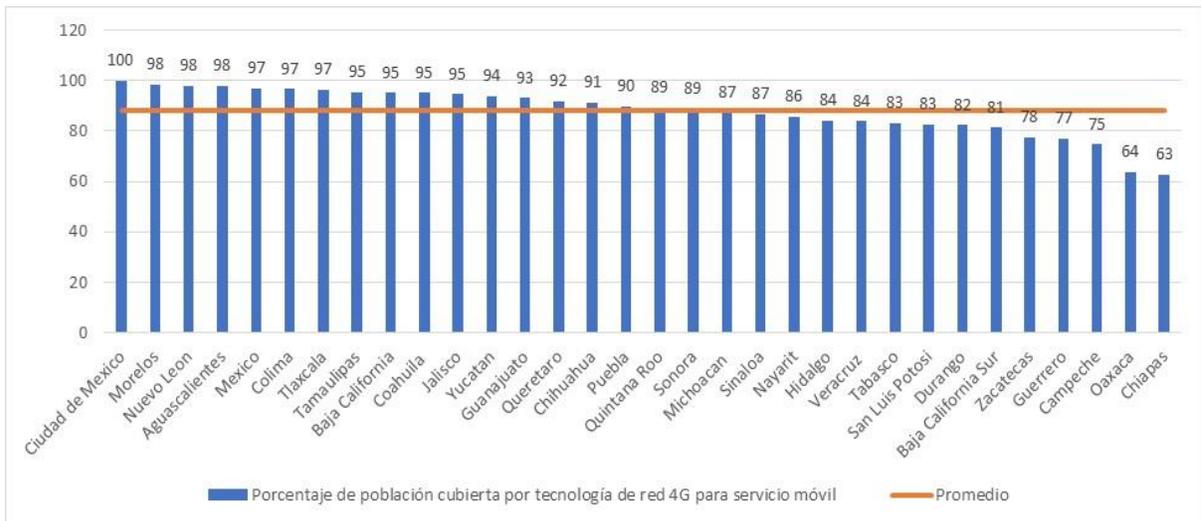
La Gráfica A6 corresponde a la variable de penetración de banda ancha fija con acceso de fibra óptica por cada 100 hogares. El primer lugar lo ocupa CDMX con 46 accesos por cada 100 hogares, seguido de San Luis Potosí con 41 accesos por cada 100 hogares y Chihuahua y Yucatán con la misma cifra de 38 accesos por cada 100 hogares. Oaxaca, Zacatecas y Baja California Sur presentan el mayor rezago. Por su parte el promedio nacional corresponde a 22 accesos por cada 100 hogares.

Gráfica A6. Penetración de banda ancha fija con acceso de fibra óptica



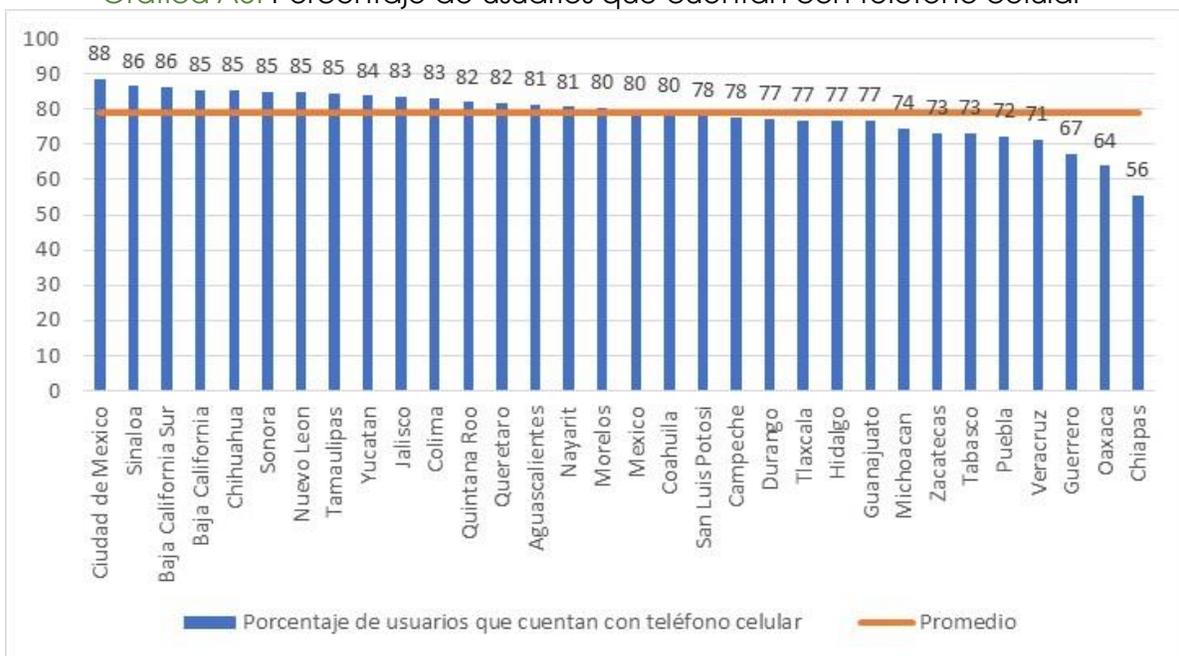
La Gráfica A7 muestra el porcentaje de población cubierta por tecnología de red 4G para servicio de telefonía móvil. El primer lugar es para CDMX con el 100 por ciento. Le siguen Morelos, Nuevo León y Aguascalientes con 98 por ciento. La menor cobertura de 4G se registra en Oaxaca y Chiapas con 64 por ciento y 63 por ciento, respectivamente. En cuanto a la media nacional ésta es igual a 88 por ciento.

Gráfica A7. Porcentaje de población cubierta por tecnología de red 4G para servicio móvil



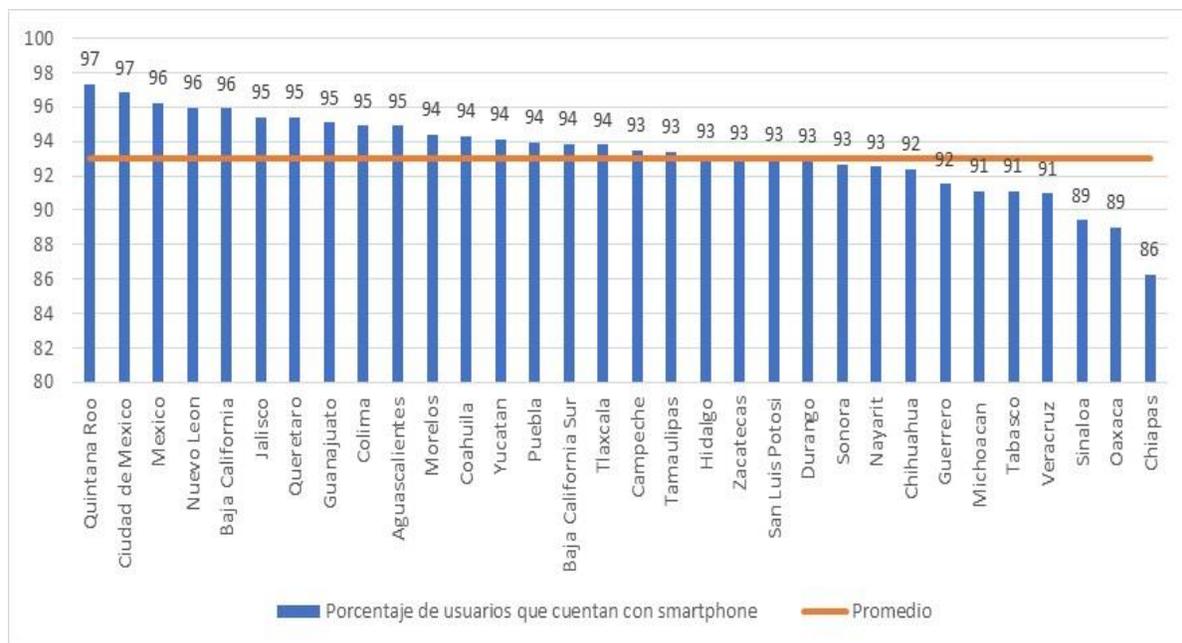
La Gráfica A8 muestra al porcentaje de usuarios que cuentan con teléfono celular siendo CDMX la entidad federativa que ocupa el primer lugar con un 88 por ciento, seguido de Sinaloa y Baja California Sur con un 86 por ciento. En el otro extremo se encuentran Guerrero, Oaxaca y Chiapas con 67 por ciento, 64 por ciento y 56 por ciento, respectivamente. El promedio del porcentaje de usuarios a nivel nacional que cuentan con teléfono celular es del 79 por ciento.

Gráfica A8. Porcentaje de usuarios que cuentan con teléfono celular



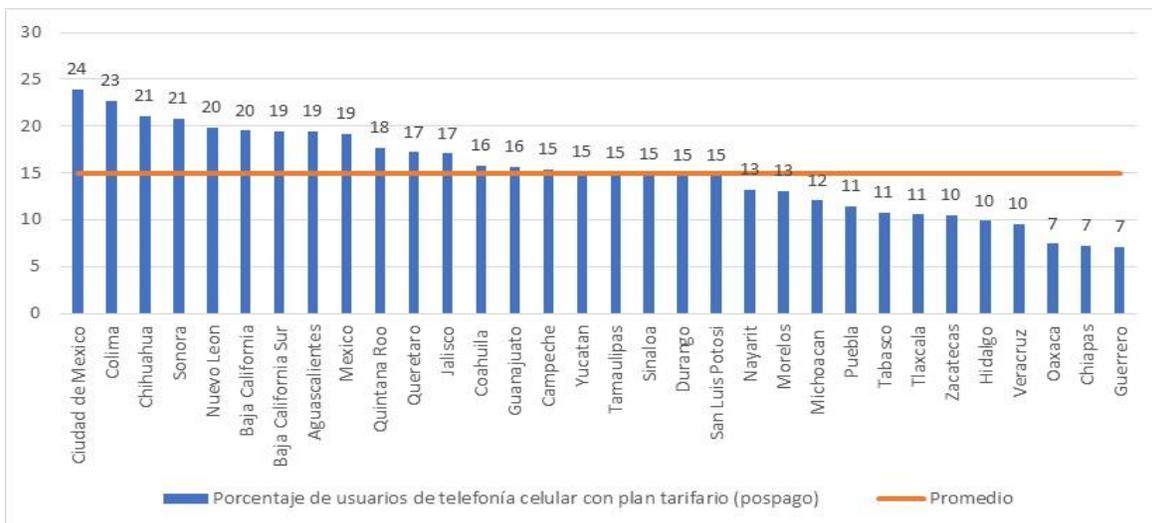
Por otra parte, la gráfica A9 presenta el porcentaje de usuarios que cuentan con Smartphone o teléfono inteligente. Quintana Roo y CDMX ocupan el primer lugar con 97 por ciento. Los puestos a la zaga los ocupan Sinaloa y Oaxaca con 89 por ciento, y Chiapas con 86 por ciento. A nivel nacional el promedio es de 93 por ciento.

Gráfica A9. Porcentaje de usuarios que cuentan con Smartphone



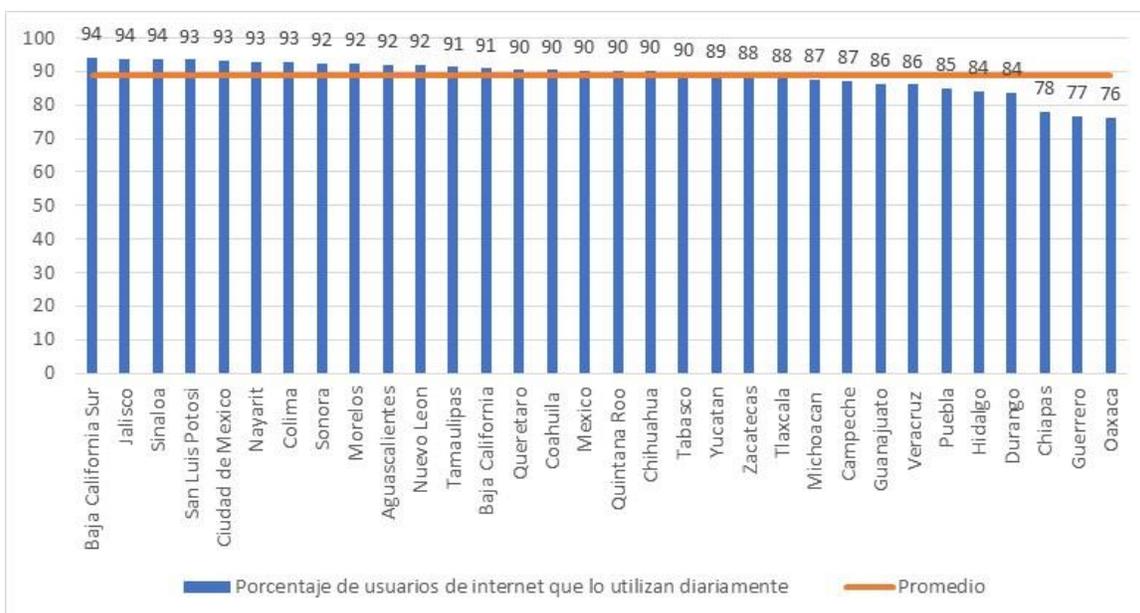
La Gráfica A10 muestra al porcentaje de usuarios de telefonía celular con plan tarifario (postpago) siendo CDMX la entidad federativa que ocupa el primer lugar con un 24 por ciento, seguido de Coahuila con 23 por ciento y Chihuahua y Sonora con 21 por ciento. En el otro extremo se encuentran Oaxaca, Chiapas y Guerrero con el 7 por ciento. El promedio a nivel nacional es del 15 por ciento.

Gráfica A10. Porcentaje de usuarios de telefonía celular con plan tarifario (postpago)



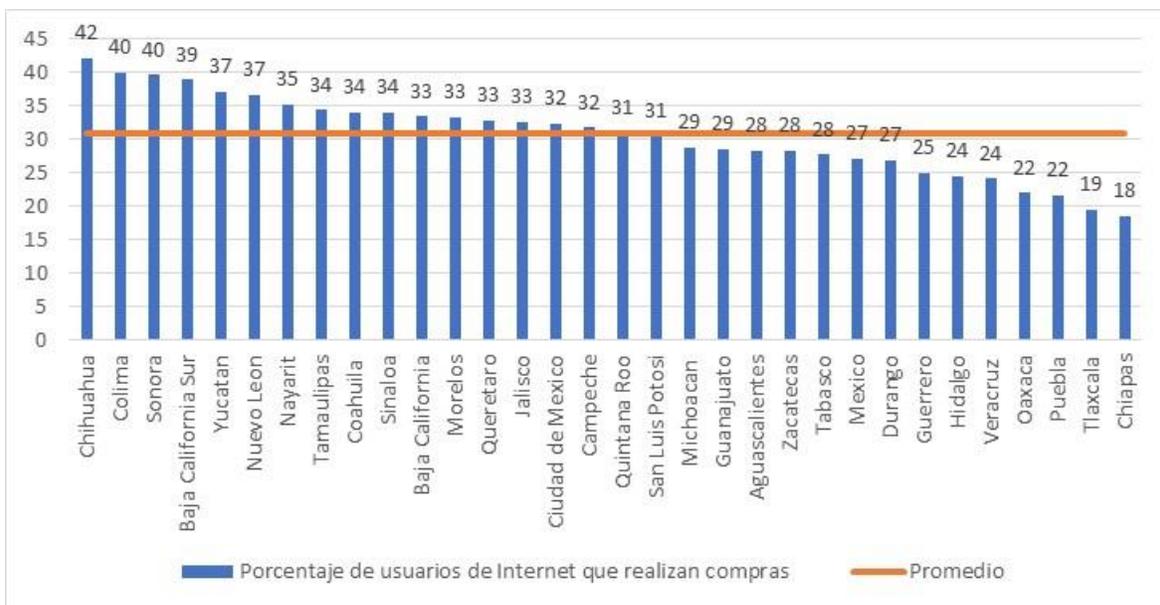
La Gráfica A11 presenta al porcentaje de usuarios de internet que lo utilizan diariamente. Baja California Sur, Jalisco y Sinaloa ocupan el primer sitio con 94 por ciento; seguidos de San Luis Potosí, Ciudad de México, Nayarit y Colima con un 93 por ciento. En las posiciones más rezagadas se encuentran Chiapas, Guerrero y Oaxaca con 78%, 77% y 76%, respectivamente. El promedio de los treinta y dos estados es de 89%.

Gráfica A11. Porcentaje de usuarios de internet que lo utilizan diariamente



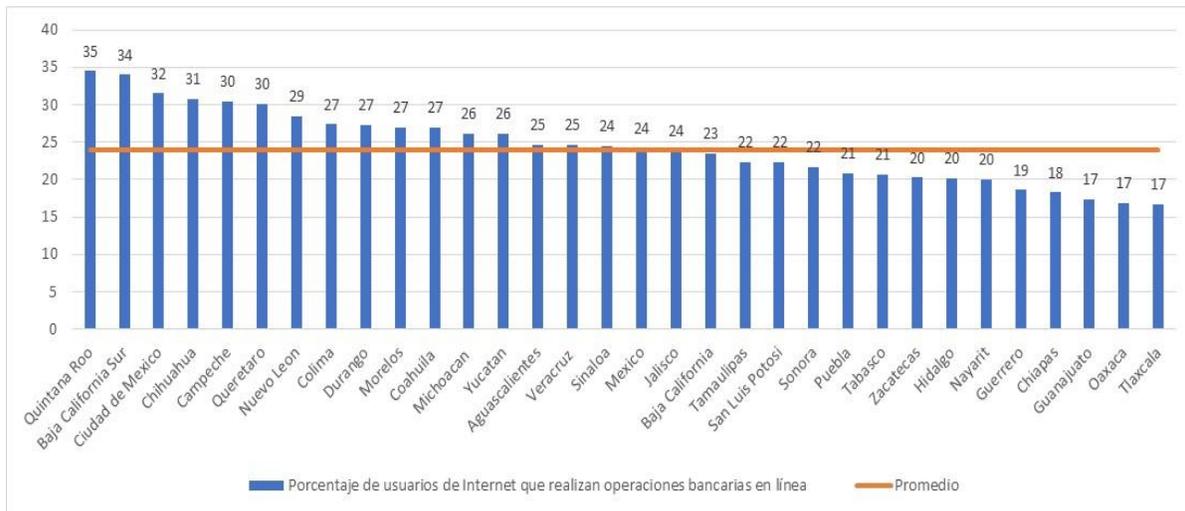
Los usos de las TIC's en actividades económicas se analizan con indicadores como el porcentaje de usuarios de Internet que realizan compras por ese medio (Gráfica A12). El primer lugar es para Chihuahua con 42 por ciento, seguido de Colima y Sonora con 40 por ciento cada una. Por su parte Puebla, Tlaxcala y Chiapas reportan los menores porcentajes con 22%, 19% y 18% respectivamente. La media nacional es de 31 por ciento.

Gráfica A12. Porcentaje de usuarios de Internet que realizan compras



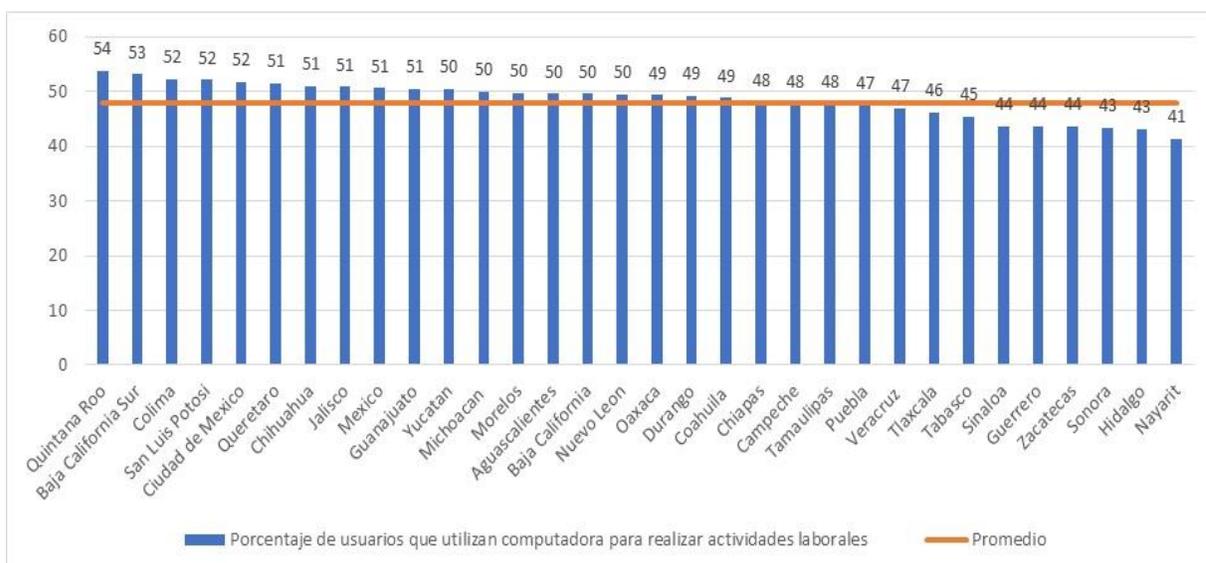
El porcentaje de usuarios de Internet que realizan operaciones bancarias en línea se muestra en Gráfica A13. En este caso Quintana Roo reporta el mayor avance con un porcentaje del 35 por ciento de los usuarios de Internet del estado, seguido por Baja California Sur con el 34 por ciento. Las entidades federativas con menor porcentaje de operaciones bancarias en línea son Guanajuato, Oaxaca y Tlaxcala, cada una con el 17 por ciento. A nivel nacional el promedio es de 24 por ciento.

Gráfica A13. Porcentaje de usuarios Internet que realizan operaciones bancarias en línea.



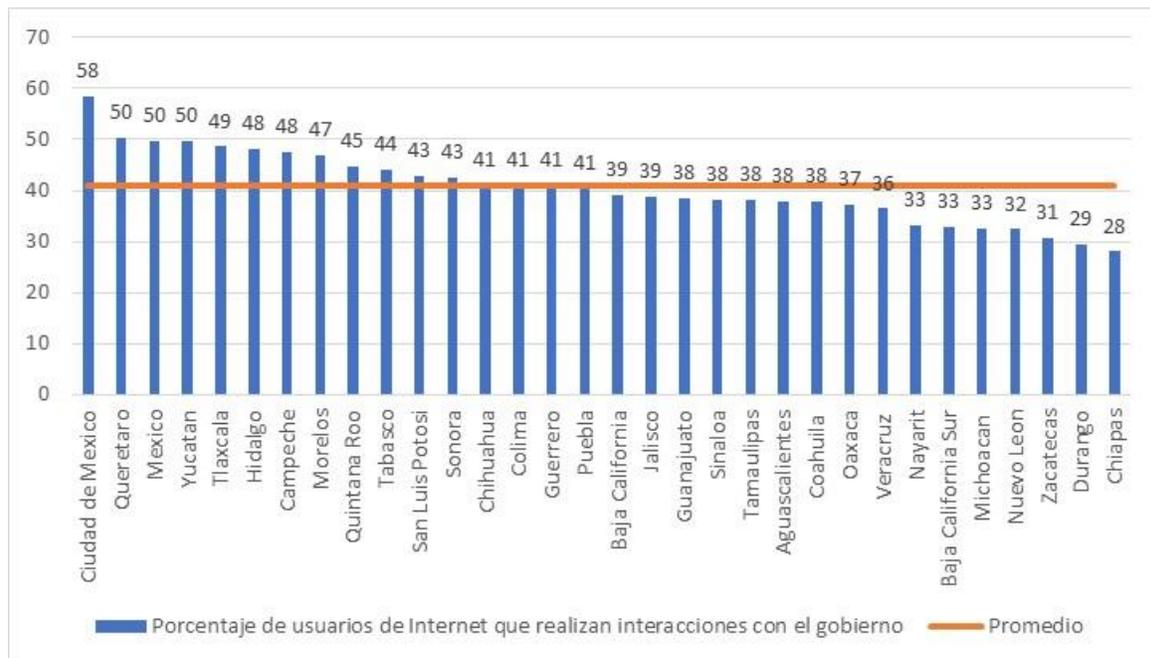
En la Gráfica A14 se observa al porcentaje de usuarios que utilizan computadora para realizar actividades laborales. El primer sitio es para Quintana Roo con 54 por ciento, seguido de Baja California Sur con 53 por ciento. Las posiciones más rezagadas las ocupan Sonora e Hidalgo con 43 por ciento, y Nayarit con el 41 por ciento. Por su parte la media nacional se encuentra al 48 por ciento.

Gráfica A14. Porcentaje de usuarios que usan computadora para actividades laborales.



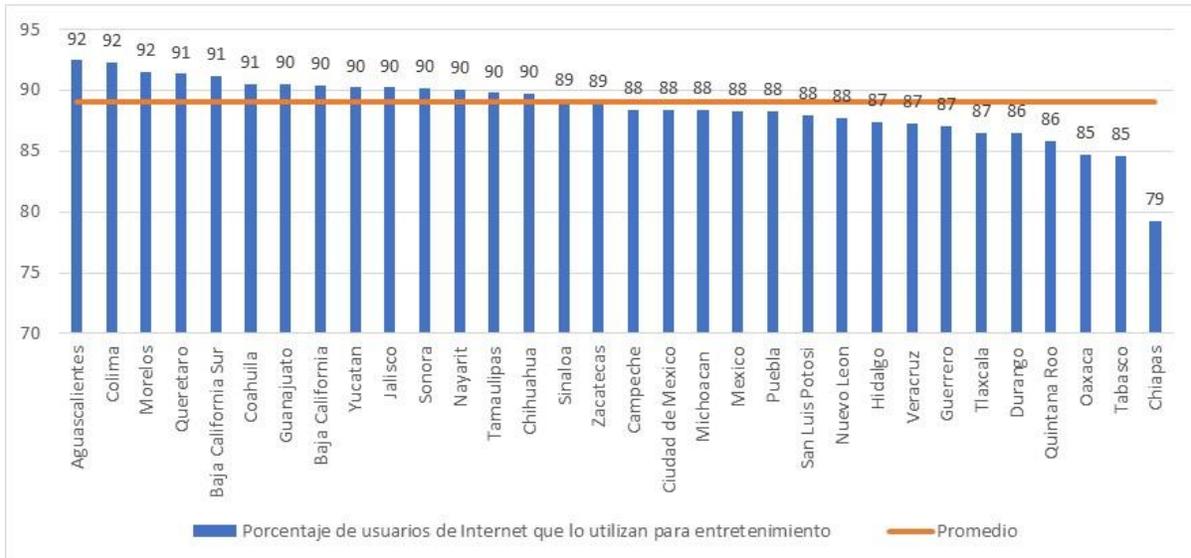
La Gráfica A15 presenta al porcentaje de usuarios de internet que realizan interacciones con el gobierno. El primer sitio es para CDMX con 58 por ciento, seguido por Querétaro, Estado de México y Yucatán con 50 por ciento. En los lugares más rezagados se encuentran Durango y Chiapas con 29 por ciento y 28 por ciento, respectivamente. A nivel nacional el promedio alcanza el 41 por ciento.

Gráfica A15. Porcentaje de usuarios que realizan interacciones con el gobierno



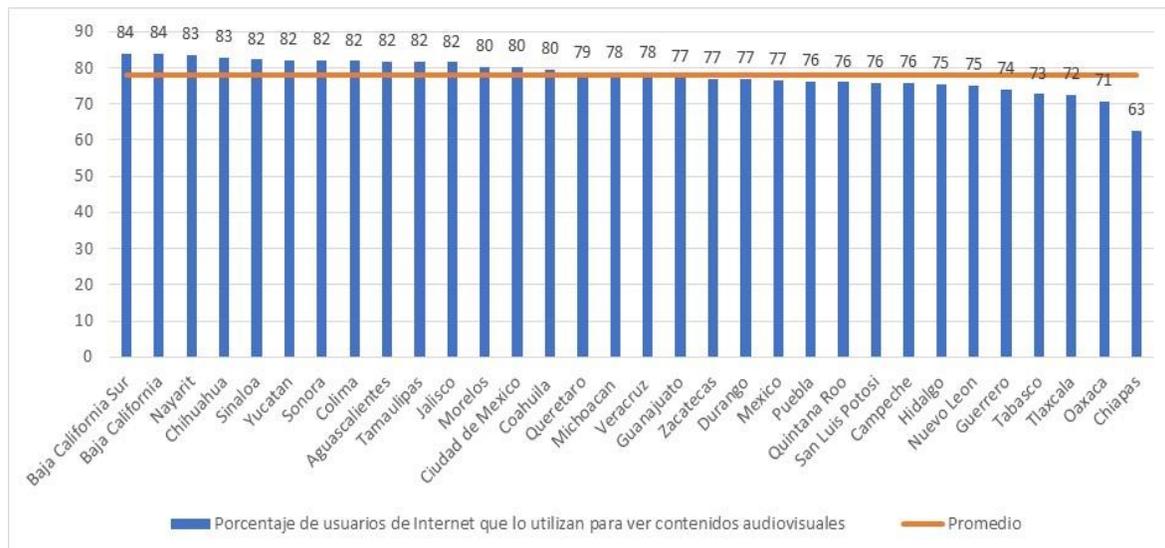
En la Gráfica A16 se observa al porcentaje de usuarios de internet que lo utilizan para entretenimiento. Esta es una modalidad de uso que cuenta ya con una elevada aceptación entre los usuarios en todas las entidades federativas. El mayor uso lo reportan Aguascalientes, Colima y Morelos con 92 por ciento. El menor uso se reporta en Tabasco y Chiapas con 85 por ciento y 79 por ciento, respectivamente. La media obtenida entre las treinta y dos entidades federativas es de 89 por ciento.

Gráfica A16. Porcentaje de usuarios internet que lo utilizan para entretenimiento



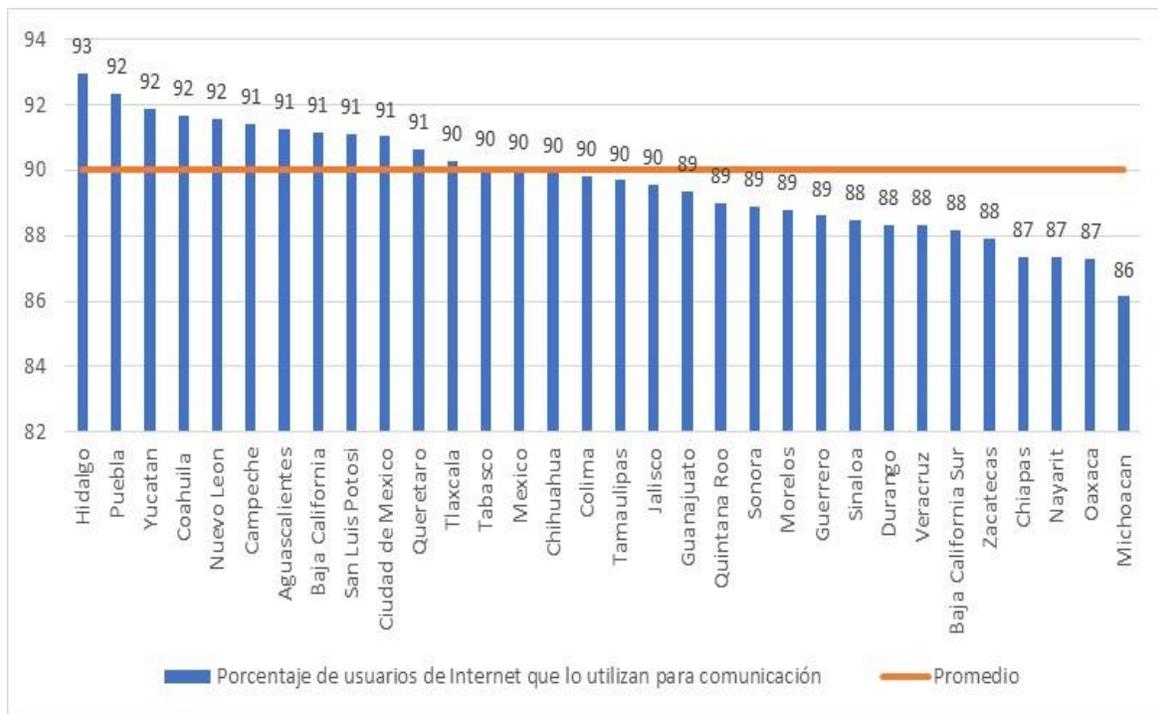
En la Gráfica A17 se observa al porcentaje de usuarios de internet que lo utilizan para ver contenidos audiovisuales. Baja California Sur y Baja California se encuentran a la cabeza con un 84 por ciento. En el otro extremo se encuentran Oaxaca y Chiapas con 71 por ciento y 63 por ciento, respectivamente. El promedio nacional es de un 78 por ciento

Gráfica A17. Porcentaje de usuarios de internet que lo utilizan para ver contenidos audiovisuales



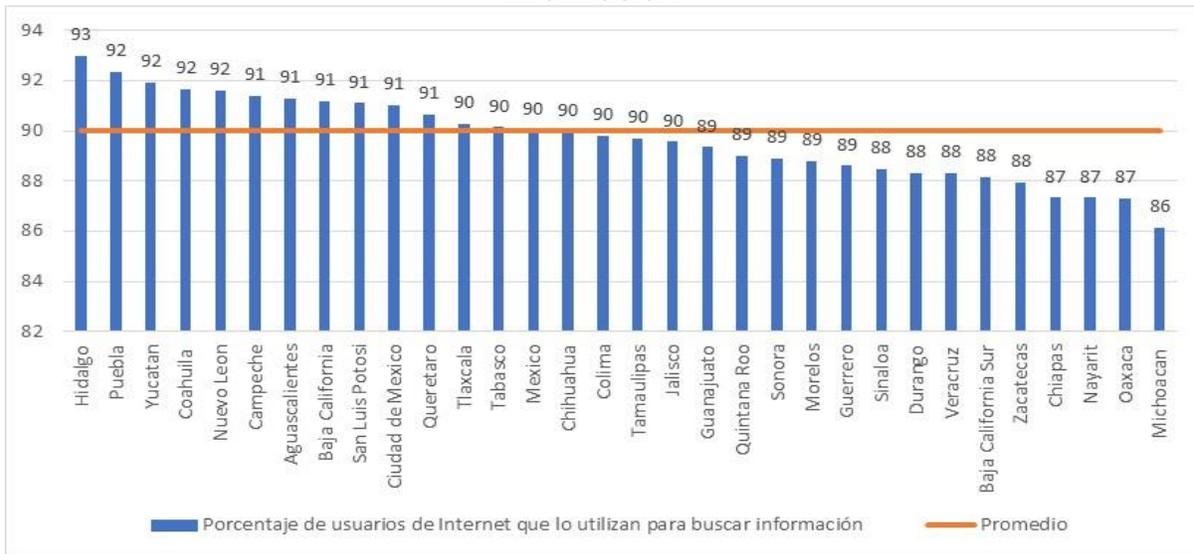
La Gráfica A18 presenta el porcentaje de usuarios de internet que lo utilizan para comunicarse. En el primer sitio se encuentra Hidalgo con 93 por ciento. En los sitios más rezagados se encuentran los estados de Oaxaca y Michoacán con 87 por ciento y 86 por ciento, respectivamente. La media nacional corresponde a un 90 por ciento.

Gráfica A18. Porcentaje de usuarios de internet que lo utilizan para comunicarse



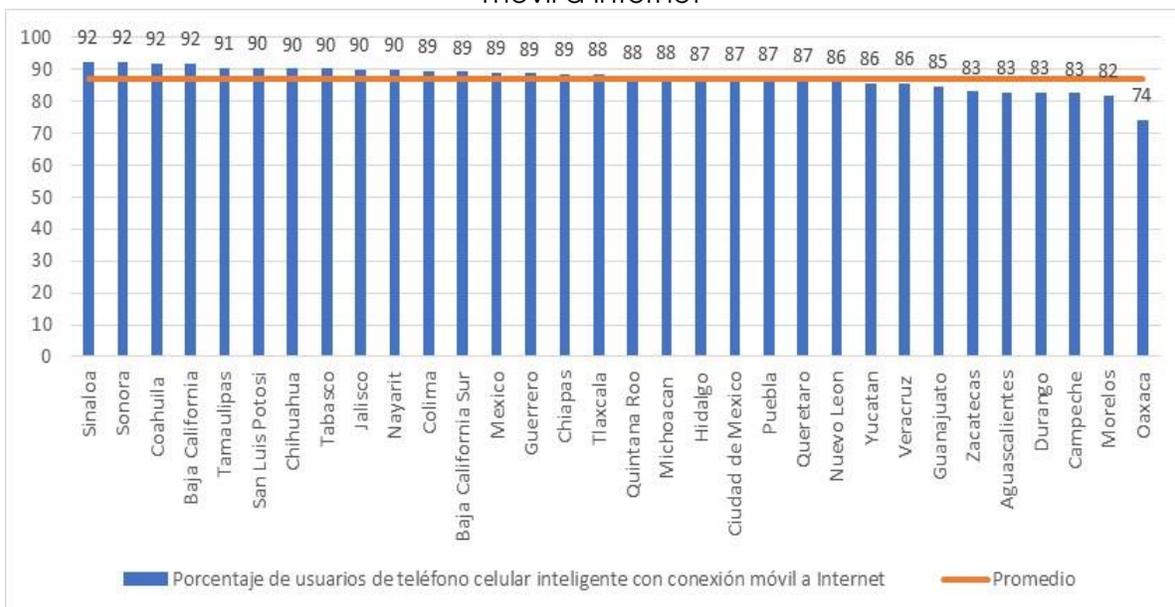
En la Gráfica A19 se muestra el porcentaje de usuarios de internet que lo utilizan para obtener información. El primer sitio es para Hidalgo con 93 por ciento. Oaxaca y Michoacán reportan las menores cifras con 87 por ciento y 86 por ciento, respectivamente. En cuanto al promedio nacional, éste es igual a un 90 por ciento.

Gráfica A19. Porcentaje de usuarios de internet que lo utilizan para obtener información



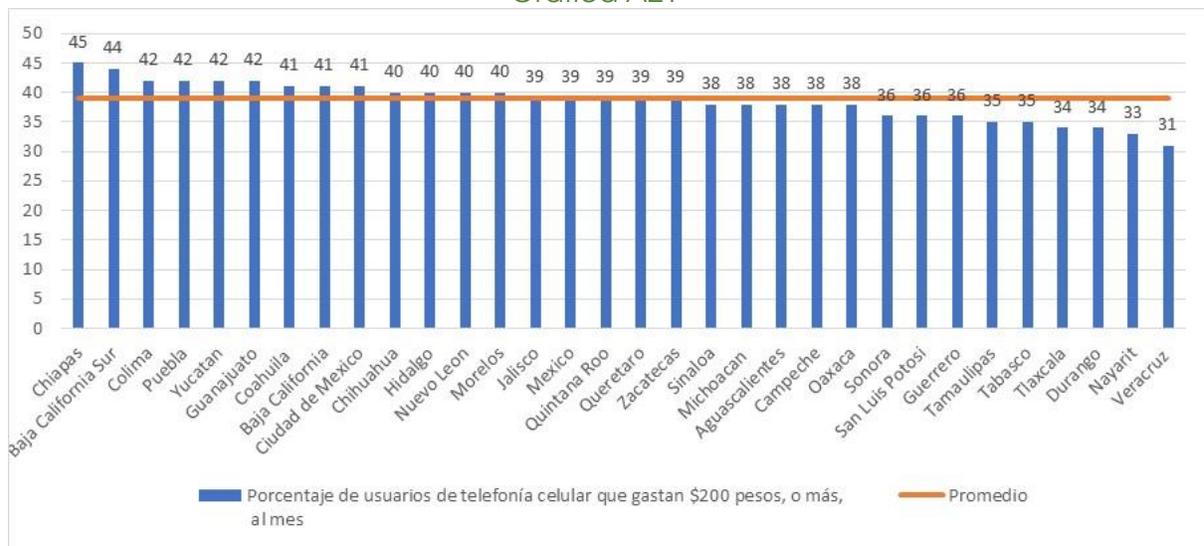
La Gráfica A20 muestra el porcentaje de usuarios de teléfono celular inteligente con conexión móvil a Internet. Las entidades mejor posicionadas son Sinaloa, Sonora, Coahuila y Baja California con un porcentaje de 92 por ciento. Las entidades con mayor rezago son Morelos con 82 por ciento y Oaxaca con 74 por ciento. El promedio nacional es igual a 87 por ciento.

Gráfica A20. Porcentaje de usuarios de teléfono celular inteligente con conexión móvil a Internet



La Gráfica A21 es referente al porcentaje de usuarios de telefonía celular que gasta 200 pesos o más al mes. Chiapas con 45 por ciento encabeza la gráfica, seguida Baja California Sur con 44 por ciento. Las entidades más rezagadas con Nayarit con 33 por ciento y Veracruz con 31 por ciento. El promedio de las 32 entidades federativas es 39 por ciento.

Gráfica A21



## Anexo 2

### Metodología y Análisis de *clústers*.

En esta sección se presenta el análisis de clúster que es utilizado en este reporte analítico, iniciando con un gráfico Screeplot<sup>26</sup> (Gráfica A22) el cual es de utilidad para determinar el número de grupos a utilizar. La forma de interpretar la gráfica Screeplot<sup>27</sup> consiste en observar en qué parte se rompe la curva formándose una especie de “codo”, lo cual es indicativo de que la calidad del modelo no se incrementa conforme va aumentando el número de clústers. Se sugiere trabajar con seis clústers debido a que un séptimo grupo no mejora el modelo.

<sup>26</sup> Screeplot. Diagrama en el cual se grafican los eigenvalores de los factores o de los componentes principales. Se utiliza para determinar el número de factores o componentes principales a mantener. Tan, P.-N., Steinbach, M., and Kumar, V. Introduction to Data Mining. Addison-Wesley, 2006.

<sup>27</sup> <https://campus.datacamp.com/courses/unsupervised-learning-in-r/unsupervised-learning-in-r?ex=9>

Una vez determinado el número de clústers con el cual se va a trabajar se inicia con la aplicación del método jerárquico conocido como el método Ward<sup>28</sup>. Este método construye una estructura conocida como dendograma la cual muestra en qué orden se han unido los clústers y cuál es su grado de proximidad. En este sentido se obtienen dos dendogramas en virtud de que utilizamos dos tipos de distancias en el algoritmo para darle robustez a los resultados: la Euclidiana y la Manhattan. Por una parte, la distancia Euclidiana se refiere a que la distancia más corta que une a dos puntos es una línea recta; en tanto que la distancia Manhattan consiste en que la distancia más corta viene dada por la suma de los dos catetos de un triángulo rectángulo<sup>29</sup>.

Al usar seis clústers (Gráficas A23 y A24) se observa que la Ciudad de México queda aislada cuando se utiliza cualquiera de las dos distancias. Los estados de Chiapas y Oaxaca forman un grupo el cual se repite en ambos casos.

Para dar robustez al análisis técnico, se sustituye la técnica jerárquica por la técnica de partición mediante el algoritmo de K-Medias repitiendo la estimación para seis clústers (Gráfica A25). La idea básica detrás del algoritmo de K-medias consiste en definir los agrupamientos de tal manera que la variación intra-grupo sea minimizada. El algoritmo de K-medias se puede resumir de la siguiente forma:

- I. Se especifica el número de *clústers* ( $k$ ) que serán utilizados.
- II. A partir de los datos se seleccionan de forma aleatoria  $k$  puntos como los centros iniciales (centroide<sup>30</sup>) o medias.
- III. Cada observación es asignada al centroide más cercano, basado en la distancia Euclidiana, o Manhattan, entre el punto y el centroide.
- IV. Para cada uno de los  $k$  *clústers* se actualiza el centroide del *clúster* por medio del cálculo del nuevo promedio para todos los puntos en el *clúster*. El centroide de un clúster  $K$  es un vector de longitud  $p$  que contiene las medias de todas las variables para las observaciones del  $K$ -ésimo clúster,  $p$  se refiere al número de variables.
- V. Se minimiza de forma iterativa el total de la suma de cuadrados. Esto es, se iteran los pasos iii y iv hasta que el *clúster* deje de cambiar o se alcance el número máximo de iteraciones.

---

28Peña, D., 2002. Análisis de Datos Multivariantes, McGraw Hill

29 <https://xlinux.nist.gov/dads/HTML/manhattanDistance.html>

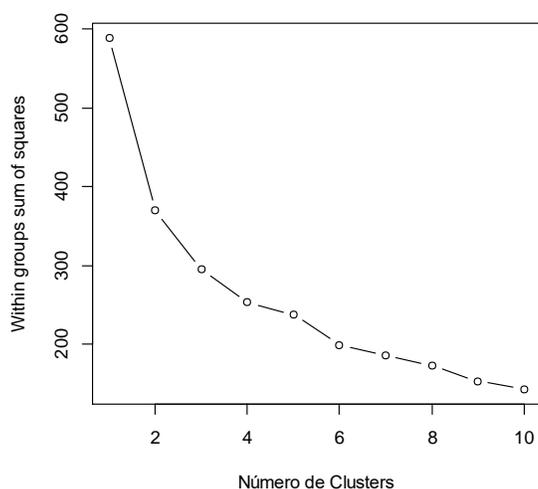
30 El centroide de un clúster se define como el punto equidistante de los objetos pertenecientes a dicho clúster. Tan, P.-N., Steinbach, M., and Kumar, V. *Introduction to Data Mining*. Addison-Wesley, 2006.

Aplicando este método para definir seis clústers (Gráfica A25) se observa que el primer grupo está conformado por la Ciudad de México.

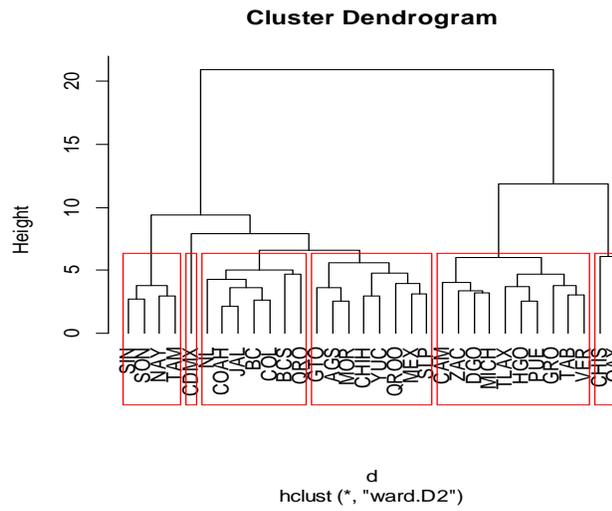
El segundo clúster lo integran Baja California, Baja California Sur, Colima y Querétaro. Por su parte, el tercer clúster se encuentra formado por las entidades federativas de Aguascalientes, Chihuahua, Coahuila, Guanajuato, Jalisco, México, Morelos, Nuevo León, Quintana Roo, San Luis Potosí y Yucatán. En el cuarto grupo se encuentra Nayarit, Sonora, Sinaloa y Tamaulipas; en el quinto están Campeche, Durango, Guerrero, Hidalgo, Michoacán, Puebla, Tabasco, Tlaxcala, Veracruz y Zacatecas ; y en el sexto, Chiapas y Oaxaca.

Los resultados se resumen en el Cuadro A1. El ejercicio estadístico completo considera que seis es el número óptimo de clústers a considerar.

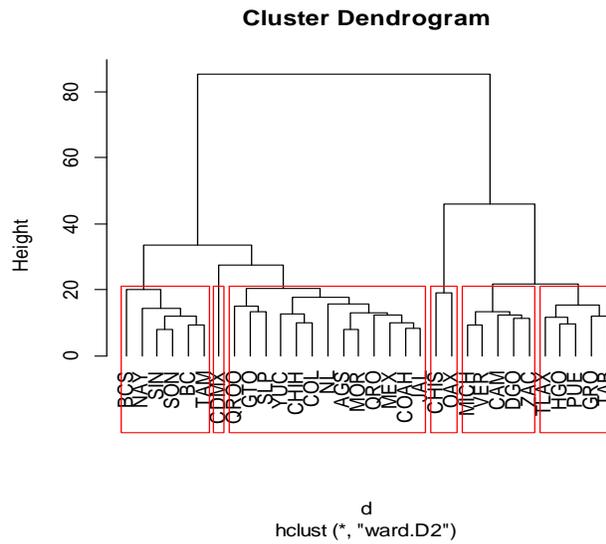
Gráfica A22. Screeplot



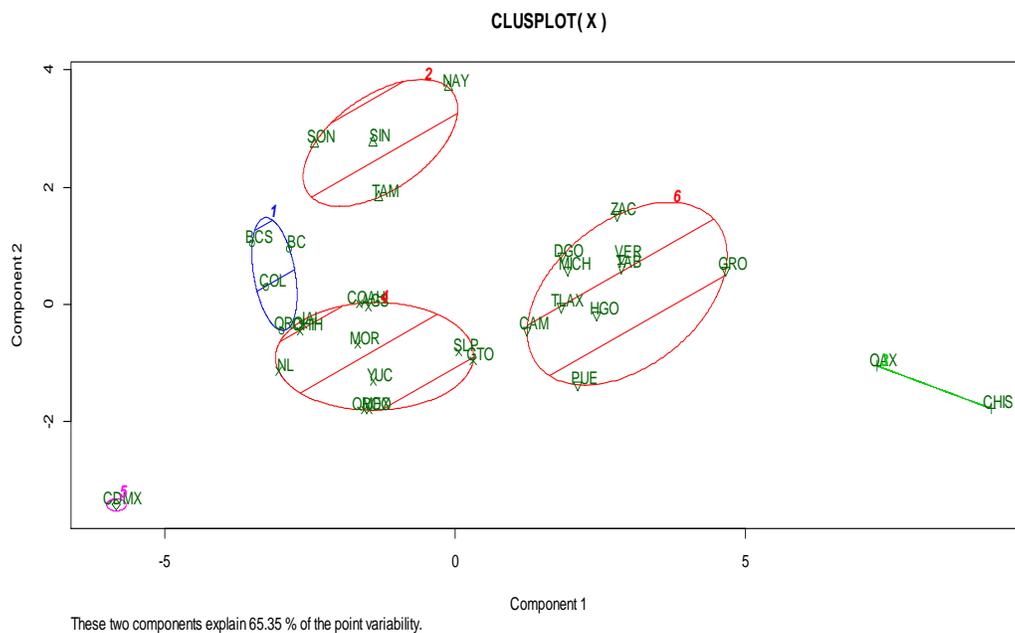
Gráfica A23. Método Ward-Distancia Euclidiana



Gráfica A24. Método Ward-Distancia Manhattan



Gráfica A25. Método K-medias.



## Metodología aplicada al diseño del índice de desarrollo digital.

El diseño del índice de bienestar digital conlleva la aplicación de la técnica de los componentes principales<sup>31</sup>, la cual es de gran utilidad cuando se trabaja con bases de datos multivariadas, como lo es en este caso. El análisis de los componentes principales va a permitir observar la “forma” de los datos en su totalidad.

En suma, los componentes principales son un tipo de transformación lineal de un set de datos, el cual ajusta dichos datos a un nuevo sistema de coordenadas, de tal forma que la parte más significativa de la varianza se encuentra en la primera coordenada, y cada coordenada subsecuente es ortogonal a la anterior y tiene una menor varianza. De esta forma, lo que se hace es transformar a un set de  $x$  variables correlacionadas sobre y sujetos (en este caso las treinta y dos entidades

31 Peña, D., 2002. Análisis de Datos Multivariantes, McGraw Hill.

federativas) en un set de  $p$  componentes principales no correlacionados sobre los mismos sujetos.

En virtud de que el set de datos cuenta con 21 indicadores se puede decir que éste tiene 21 dimensiones. Por lo tanto, el número de eigenvectores y sus eigenvalores asociados será también de 21 ya que siempre es igual al número de dimensiones del set de datos. Ahora, un eigenvector es algo similar a una dirección, en tanto que un eigenvalor es un número que dice cuánta varianza existe en los datos en dicha dirección. Es así que el eigenvector con el eigenvalor más alto será el primer componente principal.

En función de lo anterior es que el diseño del índice de desarrollo digital implica la ponderación de cada una de las cinco dimensiones y su suma para obtener el resultado final, el cual se presentó en el Cuadro 3 del reporte.