

REPORTE ANALÍTICO DE INDICADORES

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

CENTRO DE ESTUDIOS DEL IFT

MTRO. JOSÉ ALBERTO CANDELARIA BARRERA

Las opiniones expresadas en este documento son de exclusiva responsabilidad del autor y no necesariamente representan la opinión del Instituto Federal de Telecomunicaciones.

ÍNDICE

Introducción	2
Información estadística y consideraciones metodológicas	4
Dimensión 1. Infraestructura de telecomunicaciones.	5
Dimensión 2. Adopción de TIC´s.	6
Dimensión 3. Inclusión digital para la actividad económica.	6
Dimensión 4. Usos de internet.	7
Diagnóstico General	7
Determinación de agrupamientos para las entidades federativas	9
Ordenamiento de las Entidades Federativas y Grupos	11
Análisis de la información estatal	14
Conclusiones	16
ANEXO METODOLÓGICO	18
Anexo 1	18
Diagnóstico estatal de las variables de telecomunicaciones por entidad federativa	18
Anexo 2	33
Metodología y Análisis de <i>clústers</i>	33
Metodología aplicada al diseño del índice de desarrollo digital	37

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 avance de las telecomunicaciones es un fenómeno multifactorial que requiere de la consideración de distintos aspectos que son medidos a través de diversos indicadores, 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 sectorial y la apropiación o uso de las mismas por parte de las personas dificulta establecer conclusiones generales y definir una estrategia de conectividad acorde con las necesidades específicas de cada entidad federativa. 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 los mismos es que la digitalización se convierte en un factor multiplicador del bienestar y el crecimiento¹.

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 19 indicadores básicos, clasificados en cuatro diferentes dimensiones: infraestructura de telecomunicaciones; adopción de tecnologías de la información y comunicaciones (TIC's), inclusión digital para la actividad económica de las personas, y el uso del internet.

Utilizando distintas técnicas estadísticas multivariadas se integran *clústers* o agrupamientos de las entidades federativas con el objeto de identificar aquellos estados de la República Mexicana que presentan niveles similares de avance en conectividad, en la adopción y uso de las tecnologías y servicios de telecomunicaciones.

Adicionalmente, se construye un índice de desarrollo digital 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

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

integral de los servicios de telecomunicaciones, así como en su adopción, incluyendo 19 indicadores básicos.

A las dimensiones definidas se les asigna una ponderación para reflejar la relevancia que tienen en función de los ejes de política pública de telecomunicaciones y desarrollo digital, establecidos por el gobierno federal, a través de la Secretaría de Comunicaciones y Transportes ², 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.

En efecto, el desarrollo de capacidades y habilidades digitales es importante porque se vincula con la obtención de empleos mejor remunerados (Pirzada y Khan³), o con el acceso a más y mejor información (Van Deursen & Van Dijk⁴). No obstante, en México existen aun 34 millones de personas que no tienen acceso a Internet; la mayor parte de ellas residen en zonas urbanas donde las carencias económicas frenan la conectividad⁵. 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 mayor ponderación a la dimensión de infraestructura de telecomunicaciones, la cual es de 40%; en tanto que a las tres dimensiones restantes se les asigna un ponderador de 20%, reflejando su relevancia.

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 y adopción. A partir del ordenamiento obtenido de los agrupamientos y del índice de desarrollo digital se concluye que Ciudad de México, Baja California, Baja California Sur, Nuevo León y Quintana Roo integran el grupo 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 Guerrero, Oaxaca y Chiapas. Así también, se detectan que existen mayores brechas entre estados en materia de la penetración de las líneas fijas y en la teledensidad de los servicios móviles. 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.

² Página 9, https://www.dof.gob.mx/nota_detalle.php?codigo=5596042&fecha=02/07/2020

³ Pirzada, K. & Khan F. *Measuring Relationship between Digital Skills and Employability*. European Journal of Business and Management www.iiste.org ISSN 2222-1905 (Paper) ISSN 2222-2839 (Online). Vol.5, No.24, 2013.

⁴ 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.

⁵ Ver ENDUTIH 2019. <https://www.inegi.org.mx/programas/dutih/2019/>

El trabajo busca proveer al personal del 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 19 indicadores relevantes del sector de las telecomunicaciones los cuales se presentan para cada una de las 32 entidades federativas. Los indicadores se clasifican en cuatro categorías o dimensiones según las características del fenómeno que miden, a saber: infraestructura de las telecomunicaciones, la adopción de TIC's, la inclusión digital para fines de la actividad económica, y los usos del internet. Estas 4 dimensiones muestran en el agregado los aspectos más relevantes asociados a la conectividad y la adopción de los servicios.

La clasificación de indicadores en diferentes dimensiones, ha sido explorada en otros estudios con fines similares, en los que también se estiman índices a partir de múltiples variables para otros países o conjuntos de países. Así por ejemplo, la OCDE⁶ construye un índice incluyendo tres dimensiones: infraestructura de telecomunicaciones, adopción de innovación y tecnología, y empleo y crecimiento. El Network Readiness Index (NRI)⁷ considera como uno de sus subpilares (categorías) al acceso a la tecnología, que incorpora indicadores de penetración y de teledensidad como los que se ocupan en el presente trabajo. En cuanto a la dimensión de inclusión digital para la actividad económica, en el Índice para la Economía Digital y Sociedad (*The Digital Economy and Society Index, DESI por sus siglas en inglés*) la Comisión Europea construye un índice compuesto que incluye entre otras, la dimensión de integración de tecnología digital, la cual incorpora la digitalización empresarial y el comercio electrónico. En el caso del presente reporte la dimensión de inclusión digital para la actividad económica incluye variables como el porcentaje de usuarios que realizan compras y pagos por internet, tratándose en esencia de comercio electrónico. Por último, Katz y Koutroumpis⁸ ocupan en su estudio la dimensión de los usos del internet, entre otras.

⁶ *Measuring the Digital Economy A New Perspective* de la OECD. Disponible en: <https://www.oecd.org/sti/measuring-the-digital-economy-9789264221796-en.htm>

⁷ <https://networkreadinessindex.org>

⁸ Katz, R. & Koutroumpis, P. *Using a digitization index to measure the economic and social impact of digital agendas*. Disponible en: https://www.researchgate.net/publication/270806225_Using_a_digitization_index_to_measure_the_economic_and_social_impact_of_digital_agendas

Las dimensiones son también útiles para adjudicar ponderaciones que reflejen la importancia relativa de cada categoría. Con ellas se define y estima en este reporte, un índice de desarrollo digital (IDD) que refleja el peso específico que la sociedad da a cada una de ellas. Para ese efecto, los ponderadores se asignaron, en función de los ejes de política pública establecidos por el gobierno federal, a través de la Secretaría de Comunicaciones y Transportes en su Programa Sectorial⁹. En este se reconocen como aspectos relevantes la promoción de infraestructuras de telecomunicaciones, el incremento de la cobertura y la disponibilidad de los servicios de telecomunicaciones y radiodifusión, así como el desarrollo de habilidades y capacidades digitales.

Los indicadores empleados en este reporte provienen de tres fuentes estadísticas: el Banco de Información en Telecomunicaciones (BIT) del Instituto Federal de Telecomunicaciones¹⁰, estas se refieren al cuarto trimestre de 2019; la ENDUTIH, con cifras para 2018; y el Sistema de Análisis de Telecomunicaciones y Conectividad¹¹ (SATyC) que presenta información de 2019. Esta diversidad de fuentes no afecta en ningún momento a la metodología utilizada.

A continuación, se explica cada una de las dimensiones; asimismo, se muestra la ponderación que se les asigna individualmente a cada una, así como los indicadores que las integran. Cabe resaltar que tanto las dimensiones como sus respectivas ponderaciones únicamente aplican en la construcción del IDD, no así en la generación de los agrupamientos o *clústers*.

Dimensión 1. Infraestructura de telecomunicaciones.

La categoría de infraestructura de telecomunicaciones 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 diferentes redes. Se incorporan también dos indicadores que reflejan la modernidad de las redes, estos son el porcentaje de población cubierta por tecnología de red 4G para servicio móvil y la penetración de banda ancha fija con acceso de fibra óptica

A esta categoría integrada por 7 indicadores se le asigna una ponderación de 40%. Su fuente de información es el BIT, salvo para el porcentaje de población cubierta por tecnología móvil de red 4G, cuya fuente es el Sistema de Análisis de Telecomunicaciones y Conectividad (SATyC)¹². El despliegue de redes requiere de fuertes inversiones. En particular, el costo puede ser elevado tratándose de la inversión en infraestructura para dotar de redes a las

⁹ Véase: Página 9, https://www.dof.gob.mx/nota_detalle.php?codigo=5596042&fecha=02/07/2020

¹⁰ Véase: <https://bit.ift.org.mx/BitWebApp/>

¹¹ Véase: <http://satyc.ift.org.mx:8010/connect/analyst/mobile/#/login>

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

localidades aisladas y en regiones montañosas, que son las que carecen de servicios. El ponderador asignado refleja el reto a cubrir a fin de proveer cobertura universal en materia de los servicios de telecomunicaciones. Los indicadores considerados son:

1. Penetración de telefonía fija
2. Penetración de banda ancha fija
3. Penetración de TV restringida
4. Teledensidad de telefonía móvil
5. Teledensidad de internet móvil
6. Penetración de banda ancha fija con acceso de fibra óptica
7. Porcentaje de población cubierta por tecnología de red 4G para servicio móvil

Dimensión 2. Adopción de TIC's.

En esta dimensión se refleja tanto el equipamiento TIC con el que cuentan las personas en sus hogares, como la implementación exitosa de dichas tecnologías la cual se ve reflejada en su utilización. A esta dimensión se le asigna una ponderación del 20%, al igual que a las restantes categorías, reflejando así la relevancia que tienen para lograr un desarrollo integral digital. En esta dimensión se incluyen indicadores: ¹³

1. Porcentaje de usuarios que se conectan a internet a través de TV inteligente
2. Porcentaje de usuarios que cuentan con teléfono celular
3. Porcentaje de usuarios que cuentan con Smartphone
4. Porcentaje de usuarios de internet que lo utilizan diariamente

Dimensión 3. Inclusión digital para la actividad económica.

La categoría de inclusión digital y actividad económica incluye indicadores sobre el uso que dan las personas con fines productivos y comerciales. No basta con que las personas tengan una computadora o un teléfono inteligente conectados a internet para considerar su inclusión digital, se precisa saber qué hacer con estas tecnologías. Es decir, la inclusión digital es también la rutina diaria para maximizar el tiempo y sus posibilidades, particularmente en la actividad productiva.

Esta categoría cuenta con una ponderación de 20% e incluye las siguientes variables:¹⁴

¹³ La fuente de información de todas las variables de esta categoría es la Encuesta Nacional sobre Disponibilidad y Uso de Tecnologías de la Información en los Hogares (ENDUTIH) 2018. INEGI. Véase: <https://www.inegi.org.mx/programas/dutih/2019/>

¹⁴ La fuente de información de las variables que integran esta categoría es la ENDUTIH 2018, del INEGI.

1. Porcentaje de usuarios que realizan compras por internet
2. Porcentaje de usuarios que realizan pagos por internet
3. Porcentaje de usuarios que usan computadora para actividades laborales
4. Porcentaje de usuarios que realizan interacciones con el gobierno

Dimensión 4. Usos de internet.

La última categoría incluye otros usos del internet¹⁵. Se busca comparar el grado de adopción digital que reportan los usuarios en las distintas las entidades federativas, considerando los servicios de comunicación, búsqueda de información y el entretenimiento. A esta categoría se le asigna una ponderación del 20%.¹⁶

1. Porcentaje de usuarios de internet que lo utilizan para comunicación
2. Porcentaje de usuarios de internet que lo utilizan para obtener información
3. Porcentaje de usuarios de internet que lo utilizan para ver contenidos audiovisuales
4. Porcentaje de usuarios internet que lo utilizan para entretenimiento

Diagnóstico General

Los diecinueve 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 dispersión o disparidad (desviación estándar) que se registra en cada uno de los indicadores. En algunos casos se pueden observar fuertes desviaciones estándar, lo cual significa que existe mayor dispersión en los datos. Tal es el caso de la penetración de telefonía móvil con un promedio para las treinta y dos entidades de 50 y una desviación estándar de 26.3. Así, mientras la Ciudad de México presenta una penetración de 156 líneas de telefonía fija por cada 100 hogares, seguido de Nuevo León y Querétaro con 103 líneas y 81 líneas por cada 100 hogares, respectivamente, Tabasco, Oaxaca y Chiapas cuentan con 25, 24 y 14 líneas de telefonía fija por cada 100 hogares. Estas cifras corroboran la amplitud de la brecha digital regional: la Ciudad de México tiene más de 10 veces la penetración reportada en Chiapas.

¹⁵ 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

¹⁶ La fuente de información de las variables de esta categoría es la ENDUTIH 2018, del INEGI.

En la teledensidad de internet móvil la media nacional es de 77 suscripciones por cada 100 habitantes, y la desviación de 13.8. La mayor teledensidad la presenta Quintana Roo con 99 suscripciones a internet móvil por cada 100 habitantes, seguido de Sonora y CDMX con 98 y 96 suscripciones por cada 100 habitantes, respectivamente. Guerrero, Oaxaca y Chiapas registran el mayor rezago con cifras de 57, 56 y 44.

En los otros indicadores considerados no existe tanta dispersión. Tratándose del porcentaje de usuarios que utilizan internet para comunicación la desviación es de solo 2.1. El promedio nacional es de 90%.

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

N°	Dimensión	Nombre del Indicador	Media	Desviación Estándar
1	1	Penetración de telefonía fija	50	26.3
2	1	Penetración de banda ancha fija	52	18
3	1	Penetración de TV restringida	59	11.2
4	1	Teledensidad de telefonía móvil	96	11
5	1	Teledensidad de internet móvil	77	13.8
6	1	Penetración de banda ancha fija con acceso de fibra óptica	11	9.3
7	1	Porcentaje de población cubierta por tecnología de red 4G para servicio móvil	88	11.4
8	2	Porcentaje de usuarios que se conectan a internet a través de TV inteligente	11	4.3
9	2	Porcentaje de usuarios que cuentan con teléfono celular	74	7.3
10	2	Porcentaje de usuarios que cuentan con Smartphone	63	8.8
11	2	Porcentaje de usuarios de internet que lo utilizan diariamente	80	5.9
12	3	Porcentaje de usuarios que realizan compras por internet	14	4.9
13	3	Porcentaje de usuarios que realizan pagos por internet	11	3.8
14	3	Porcentaje de usuarios que usan computadora para actividades laborales	40	4.2
15	3	Porcentaje de usuarios que realizan interacciones con el gobierno	29	4.7
16	4	Porcentaje de usuarios de internet que lo utilizan para comunicación	90	2.1

17	4	Porcentaje de usuarios de internet que lo utilizan para obtener información	86	2.3
18	4	Porcentaje de usuarios de internet que lo utilizan para ver contenidos audiovisuales	77	4.6
19	4	Porcentaje de usuarios internet que lo utilizan para entretenimiento	90	2.1

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 diecinueve indicadores de telecomunicaciones seleccionados y presentados en el Cuadro 1, se utiliza la técnica de partición conocida como K-Medias¹⁷. Esta técnica utiliza la distancia entre los datos para saber si éstos son parecidos o diferentes entre sí, y es aceptada generalmente por la robustez de sus resultados. La técnica permite identificar a los estados del país que son relativamente homogéneos entre sí, con base en la disponibilidad y adopción actual de los servicios de telecomunicaciones de manera simultánea. Esto es, considerando los diecinueve indicadores de manera simultánea se forman grupos 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, el método Ward¹⁸. 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.

¹⁷ Para una descripción detallada, véase: Peña, D., 2002. Análisis de Datos Multivariados, McGraw Hill.

¹⁸ El segundo método se trata de una técnica jerárquica conocida como método Ward. La idea de ambas técnicas es que aquellas naciones que compartan características semejantes estarán juntas en un mismo grupo, y a su vez, separadas de los otros grupos definidos. En función de los resultados que generan ambas técnicas, se consideró que el método de K-Medias es el más apropiado para el análisis.

En función de los resultados que producen las dos técnicas, se consideró que el método de Ward (con distancia Manhattan) 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, BC, BCS, NL, QROO
2	AGS, COAH, COL, CHIH, JAL, SON, SIN, TAM, YUC
3	CAM, DGO, GTO, MICH, NAY, SLP, VER, ZAC
4	QRO, MEX, MOR
5	HGO, PUE, TAB, TLAX
6	CHIS, GRO, OAX

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, con base en los diecinueve indicadores considerados simultáneamente. Así, por ejemplo, la Ciudad de México comparte un nivel similar a entidades como Baja California, Baja California Sur, Nuevo León y Quintana Roo, considerando el agregado de los indicadores, por lo que quedan integradas en un mismo grupo. No así Oaxaca, Guerrero o Chiapas que presentan el mayor rezago. Estas tres entidades a su vez, son similares en cuanto al desarrollo sectorial y se incluyen en un grupo aparte.

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 diecinueve variables estandarizadas¹⁹; y un índice de desarrollo digital (IDD) para cada grupo.

El IVE se define como la suma de las diecinueve variables estandarizadas; este se emplea para estimar un índice para cada grupo. El IVE toma mayor valor en la medida que el nivel de equipamiento, adopción y uso de los servicios del agrupamiento es más elevado. Para facilitar la lectura, el índice se estima en una escala de *valores que va de 1 a 100*.²⁰

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 el promedio para cada grupo (Cuadro 2); así también, en la siguiente sección se presentan IDD para cada entidad federativa (Cuadro 3). Para ello se utiliza la técnica de análisis de componentes principales²¹, 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 19 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 documento²², pero la esencia es la siguiente: se cuenta con un conjunto de datos con muchos indicadores o variables, es decir, multivariado, el cual es simplificado convirtiendo a las variables originales en un número menor de “componentes principales”. Es así que los componentes principales son la estructura subyacente de los datos²³.

¹⁹ La estandarización consiste en una transformación lineal de las diecinueve 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.

²⁰ Esto 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 rescalamiento final.

²¹ Peña, D., 2002. Análisis de Datos Multivariantes, McGraw Hill.

²² *Ibidem*

²³ 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

En función de lo anterior, el diseño del IDD conlleva la aplicación de ponderadores a cada una de las cuatro 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 2009²⁴.

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 ponderado²⁵ (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, BC, BCS, NL, QROO	100.0	2.8	0.823
2	AGS, COAH, COL, CHIH, JAL, SON, SIN, TAM, YUC	86.9	0.9	0.793
3	QRO, MEX, MOR	86.5	0.9	0.782
4	HGO, PUE, TAB, TLAX	69.0	-1.1	0.762
5	CAM, DGO, GTO, MICH, NAY, SLP, VER, ZAC	69.0	-1.3	0.761
6	CHIS, GRO, OAX	51.0	-3.2	0.708

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.

²⁴ E-readiness rankings 2009. A report from the Economist Intelligence Unit. The Economist.

<http://graphics.eiu.com/pdf/e-readiness%20rankings.pdf>

²⁵ 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.

A partir del ordenamiento obtenido, se tiene que el *clúster* formado por Ciudad de México, Baja California, Baja California Sur, Nuevo León y Quintana Roo tiene los índices más altos, con un valor igual a 100 para el IVE y 2.8 tratándose del IDD. El segundo *clúster* conformado por Aguascalientes, Coahuila, Colima, Chihuahua, Jalisco, Sonora, Sinaloa, Tamaulipas y Yucatán tiene un IVE igual a 86.9 y un IDD igual a 0.9. En ese orden sigue el grupo de Querétaro, Estado de México y Morelos con un IVE de 86.5 y también un IDD igual a 0.9. El Mapa 1 resume los agrupamientos.

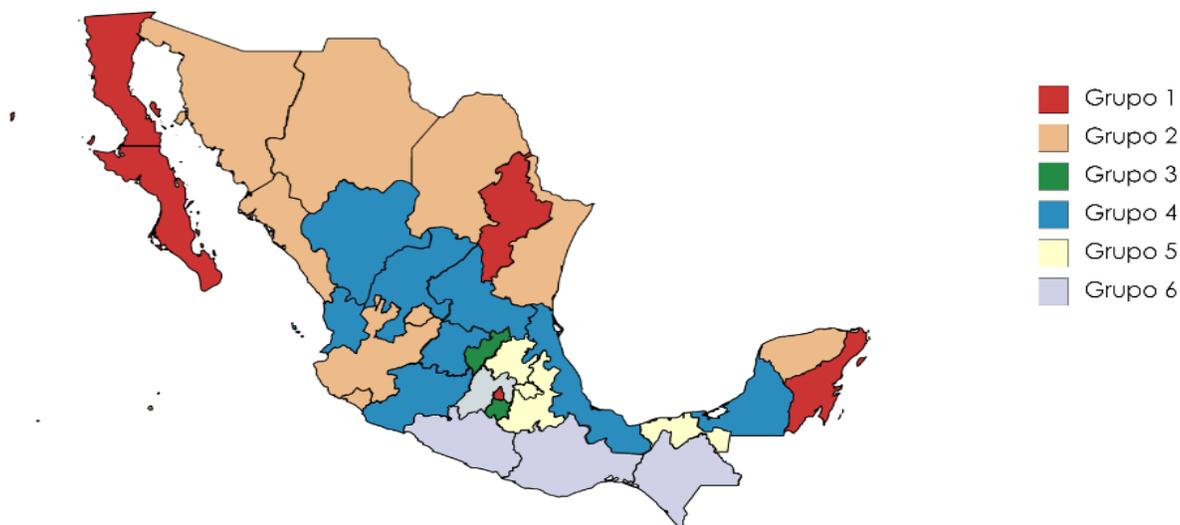
Con base en el IVE el grupo integrado por Hidalgo, Puebla, Tabasco y Tlaxcala presenta un nivel igual que el *clúster* que agrupa a Campeche, Durango, Guanajuato, Michoacán, Nayarit, San Luis Potosí, Veracruz y Zacatecas, ambos grupos con un valor de 69, por lo que podrían ubicarse en un mismo grupo. Sin embargo, de acuerdo con el IDD, las primeras entidades mencionadas tienen un mejor posicionamiento. No así con base en el IDD. Finalmente, el grupo de Chiapas, Guerrero y Oaxaca presentan el mayor rezago con un IVE de 51 y un IDD de -3.2.

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 Desarrollo²⁶ (IDH, por sus siglas) para México, el cual mide el nivel de desarrollo de cada país 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 por CDMX, Baja California, Baja California Sur, Nuevo León y Quintana Roo alcanza un IDH promedio igual a 0.823. Por su parte, el segundo *clúster* integrado por Aguascalientes, Coahuila, Colima, Chihuahua, Jalisco, Sonora, Sinaloa, Tamaulipas y Yucatán, tiene un IDH menor respecto al del primer grupo, con un valor de 0.793. Finalmente, el grupo conformado por Chiapas, Guerrero y Oaxaca, tiene el IDH promedio más bajo, con una cifra igual a 0.708. 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.

²⁶ https://es.wikipedia.org/wiki/Anexo:Entidades_federativas_de_México_por_IDH

Mapa 1. Entidades federativas según agrupamiento



Fuente: Elaboración propia a partir del Cuadro 2.

Análisis de la información estatal

Para completar el panorama presentado a nivel agrupamiento, se estimaron IDD para cada una de las entidades federativas. Estas cifras se presentan en el Cuadro 3, en el que también se presentan los valores para cada una de las 4 dimensiones que, ponderadas, integran el IDD. Tal y como puede observarse la Ciudad de México reporta el IDD más elevado con un índice de 3.6, esto a pesar de no liderar dos de las categorías, la de adopción de TIC's y la de usos de internet. En cuanto a la adopción de TIC's, la CDMX (2.81) reporta menor cifra que Nuevo León (3.25) y Sonora (3.62). Mientras que, en los usos de Internet, CDMX (0.95) está por debajo de entidades como Nuevo León (1.40), Baja California (3.10), Baja California Sur (1.25), entre otras.

El segundo puesto en el índice de bienestar digital es para Nuevo León y Baja California, ambos con 2.7. En contraste Guerrero, Oaxaca y Chiapas acusan el menor desarrollo digital, con índices de -2.6, -3 y -4.01, respectivamente.

Cuadro 3. Estimación estatal del índice de Desarrollo Digital

Entidad	Dimensión 1 Infraestructura telecom (40%)	Dimensión 2 Acceso a TIC (20%)	Dimensión 3 Inclusión Dgtl. en la actividad económica (20%)	Dimensión 4 Usos del Internet (20%)	IDD
CDMX	5.44	2.81	3.45	0.95	3.6
NL	3.31	3.25	2.15	1.40	2.7
BC	2.88	2.76	1.73	3.10	2.7
BCS	2.73	2.01	2.41	1.25	2.2
QROO	1.80	2.10	2.02	0.22	1.6
QRO	2.05	0.59	1.58	0.98	1.5
SON	1.51	3.62	0.91	-0.49	1.4
JAL	1.89	1.54	0.95	0.36	1.3
COL	0.90	1.95	1.51	1.23	1.3
AGS	1.16	1.18	1.46	0.61	1.1
YUC	-0.03	1.91	1.34	1.45	0.9
CHIH	0.70	0.90	0.98	0.94	0.8
MEX	0.63	-0.15	0.27	1.61	0.6
MOR	1.05	-0.09	-0.25	1.05	0.6
SIN	0.69	1.23	0.01	-0.05	0.5
COAH	0.17	0.94	0.51	0.54	0.5
TAM	0.47	0.85	0.29	-0.32	0.4
NAY	-0.36	-0.19	-0.81	0.35	-0.3

SLP	-0.78	-0.76	-0.35	0.47	-0.4
GTO	-0.68	-1.18	-1.03	-0.60	-0.8
MICH	-1.22	-1.45	-1.08	-0.63	-1.1
HGO	-1.42	-1.84	-1.19	-0.59	-1.3
VER	-1.73	-1.52	-1.36	-0.19	-1.3
DGO	-1.08	-0.48	-1.39	-2.58	-1.3
TAB	-1.74	-1.38	-1.47	-0.30	-1.3
PUEB	-1.35	-2.35	-1.77	0.16	-1.3
TLAX	-1.20	-1.12	-1.71	-1.50	-1.3
CAM	-2.41	-0.03	0.34	-2.94	-1.5
ZAC	-2.24	-1.71	-1.91	-0.08	-1.6
GRO	-2.65	-3.95	-2.36	-1.35	-2.6
OAX	-3.70	-3.24	-2.03	-2.40	-3.0
CHIS	-4.79	-4.89	-3.18	-2.64	-4.1

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 al avance alcanzado a partir de diecinueve indicadores de telecomunicaciones clasificados en cuatro dimensiones básicas: la infraestructura de telecomunicaciones; la adopción de TIC's; la inclusión digital para la actividad económica, y los usos del internet. A estas dimensiones se les asigna una ponderación en función de los ejes de política pública establecidos por el gobierno federal²⁷, la cual incluye en sus programas la promoción de infraestructuras, el incremento de la cobertura y la

²⁷ https://www.dof.gob.mx/nota_detalle.php?codigo=5596042&fecha=02/07/2020

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 entidades federativas como la Ciudad de México, Baja California, Baja California Sur, Nuevo León y Quintana Roo forman un grupo con avance similar, mediante el método Ward y la distancia Manhattan (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 Guerrero, 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 que con base en los resultados alcanzados en el presente estudio se 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 los grupos 6, 5 y 4.

ANEXO METODOLÓGICO

Anexo 1.

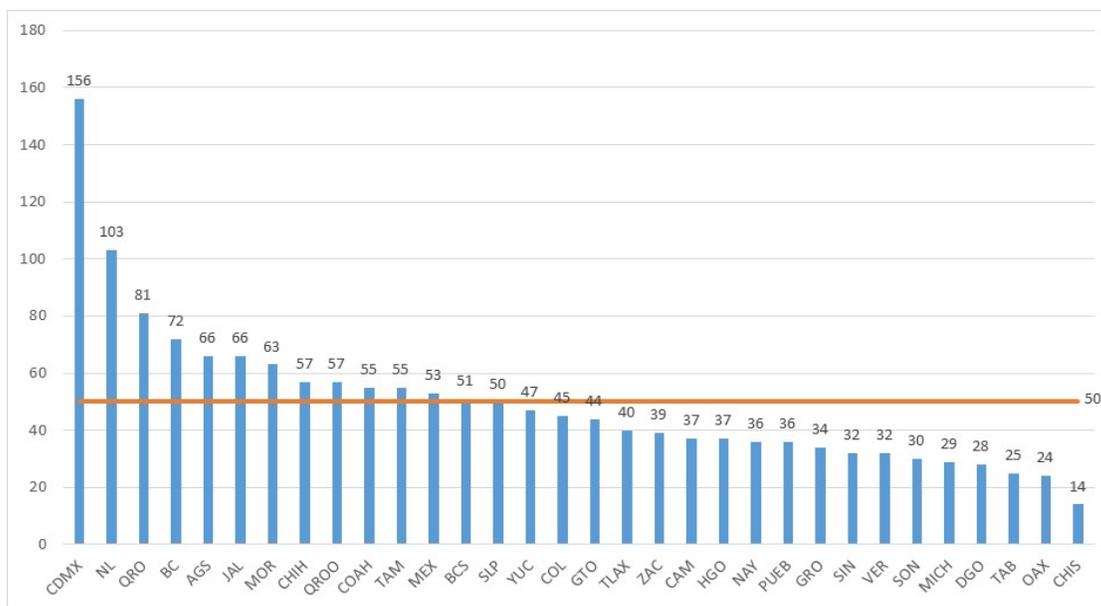
Diagnóstico estatal de las variables de telecomunicaciones por entidad federativa

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 inclusión digital o los usos del internet.

Los diecinueve indicadores de telecomunicaciones que se utilizan en el presente reporte se analizan de forma gráfica a continuación.

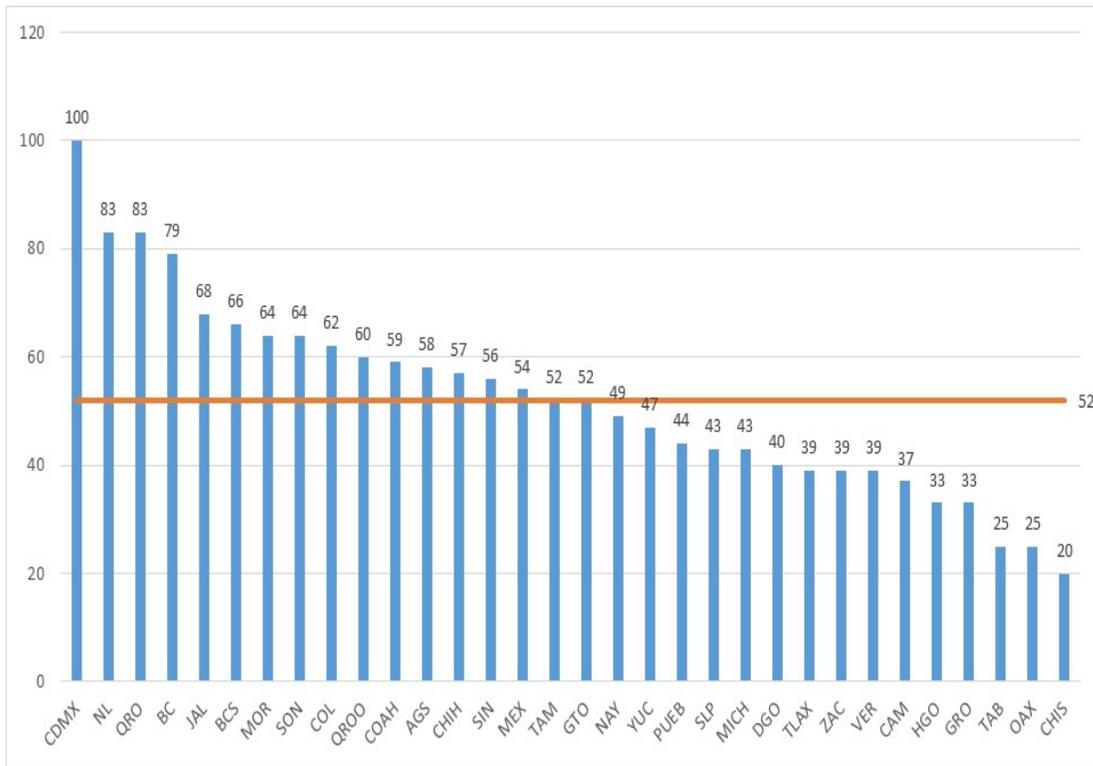
La Gráfica A1 muestra la penetración de telefonía fija por hogares para cada uno de las treinta y dos entidades federativas del país. En este caso la CDMX presenta una penetración de 156 líneas de telefonía fija por cada 100 hogares, seguido de Nuevo León y Querétaro con 103 líneas y 81 líneas por cada 100 hogares, respectivamente. Por otra parte, los estados de Tabasco, Oaxaca y Chiapas acusan el mayor rezago con 25, 24 y 14 líneas de telefonía fija por cada 100 hogares. Estas cifras corroboran la amplitud de la brecha digital regional: la CDMX tiene más de 10 veces la penetración reportada en Chiapas. La media de las 32 entidades federativas está representada mediante una línea horizontal, siendo igual a 50.

Gráfica A1. Penetración de telefonía fija



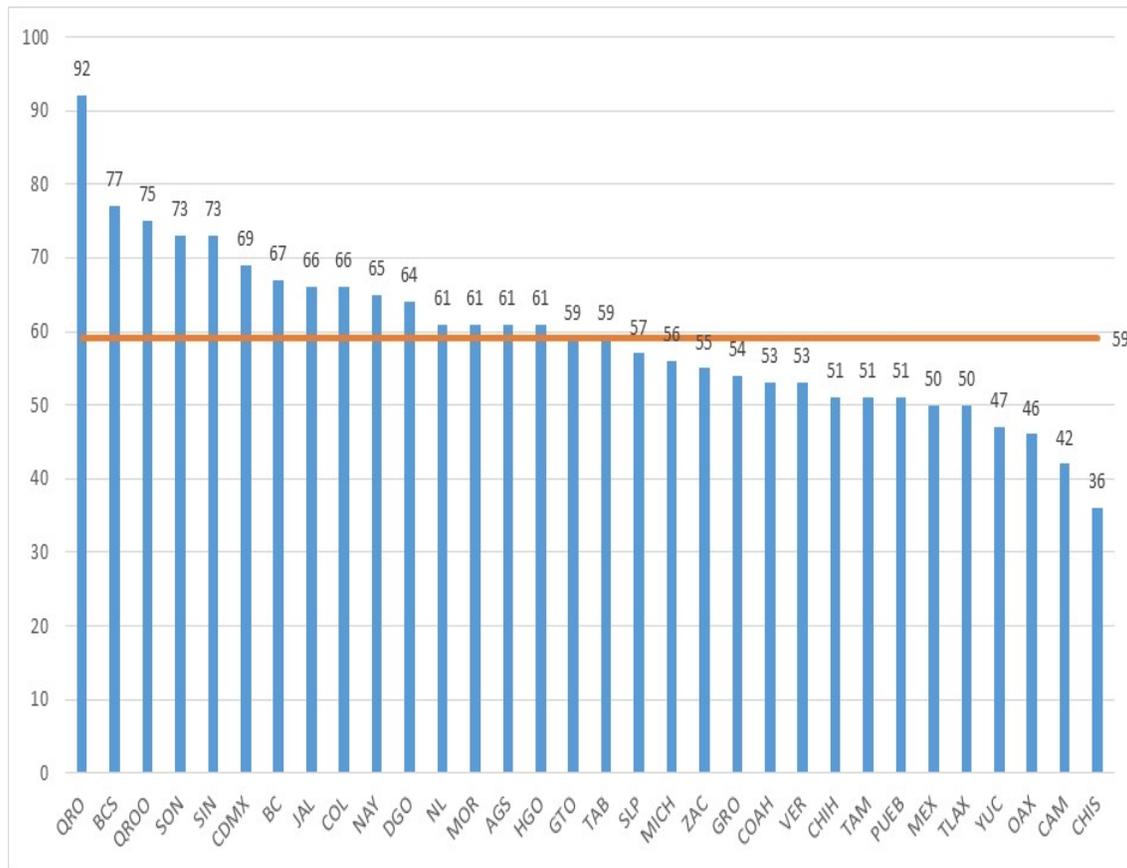
La Gráfica A2 muestra la variable de penetración de banda ancha fija por cada 100 hogares. Esta variable muestra un panorama similar a la variable anterior; es decir, la Ciudad de México se encuentra en el primer sitio con 100 accesos de banda ancha fija por cada 100 hogares, le siguen los estados de Nuevo León y Querétaro con 83 accesos por cada 100 hogares cada uno. Las últimas tres posiciones se encuentran ocupadas nuevamente por Tabasco, Oaxaca y Chiapas con 25, 25 y 20 accesos de banda ancha fija por cada 100 hogares, respectivamente. La media nacional es igual a 52 accesos de banda ancha fija por cada 100 hogares.

Gráfica A2. Penetración de banda ancha fija



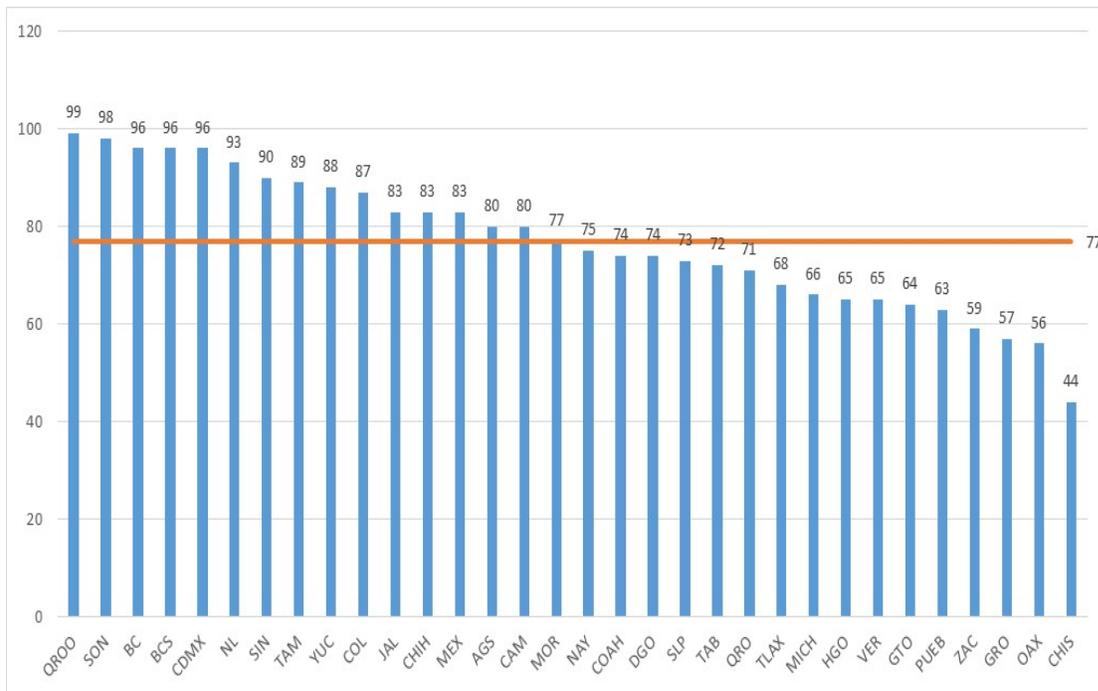
La gráfica A3 corresponde a la penetración de TV restringida por cada 100 hogares. La primera posición la ocupa el estado de Querétaro con 92 suscripciones a TV restringida por cada 100 hogares. A continuación, se encuentran Baja California Sur y Quintana Roo con 77 y 75 suscripciones por cada 100 hogares, respectivamente. Los menores niveles de penetración los acusan los estados de Oaxaca, Campeche y Chiapas con 46, 42 y 36 suscripciones por cada 100 hogares, respectivamente. Por otra parte, la media nacional corresponde a 59 suscripciones de TV restringida por cada 100 hogares.

Gráfica A3. Penetración de TV restringida



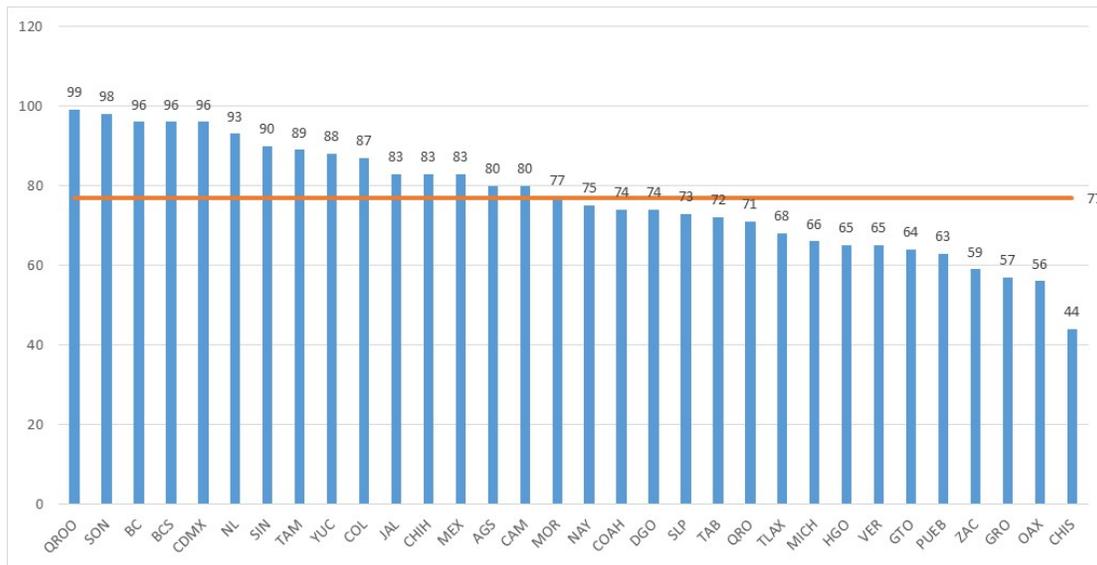
En la Gráfica A4 se observa la teledensidad de telefonía móvil por cada 100 habitantes. El primer sitio es para Sonora con 116 líneas de telefonía móvil por cada 100 habitantes; le siguen Baja California y Baja California Sur con 114 y 113 líneas de telefonía móvil por cada 100 habitantes. En el otro extremo se encuentra Oaxaca, Guerrero y Chiapas con 79, 78 y 72 líneas de telefonía móvil por cada 100 habitantes. La media nacional es igual a 96 líneas de telefonía de móvil por cada 100 habitantes.

Gráfica A4. Teledensidad de telefonía móvil



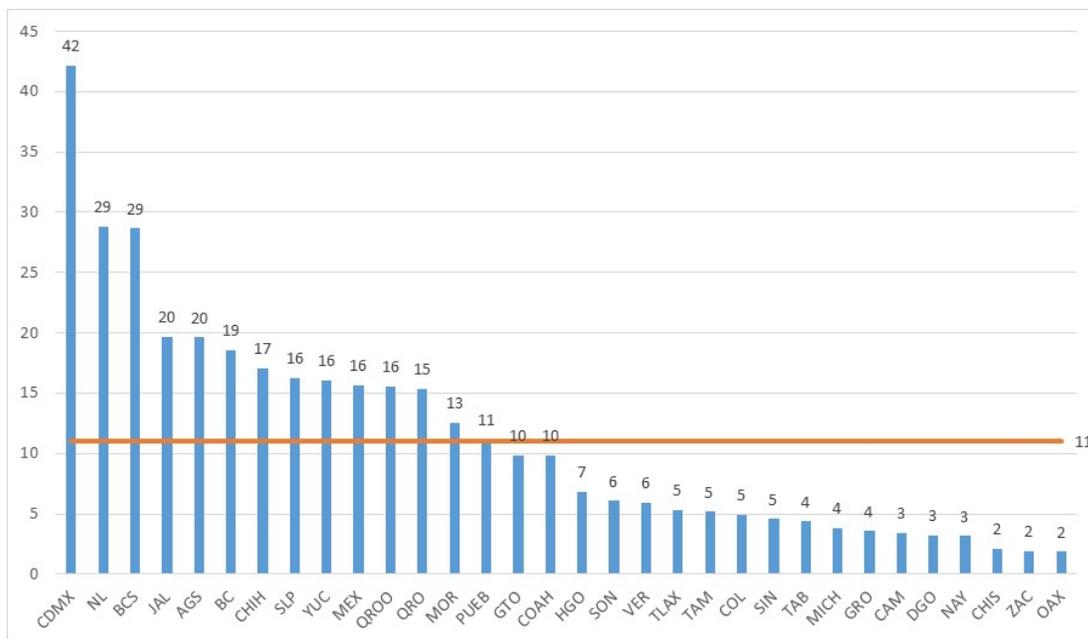
La gráfica A5 muestra la variable de teledensidad de internet móvil por cada 100 habitantes. El primer sitio lo ocupa Quintana Roo con 99 suscripciones a internet móvil por cada 100 habitantes, seguido de Sonora y CDMX con 98 y 96 suscripciones por cada 100 habitantes, respectivamente. Guerrero, Oaxaca y Chiapas son los estados con menores cifras de teledensidadde internet móvil con 57, 56 y 44 suscripciones por cada 100 habitantes, respectivamente. En tanto que el promedio nacional se encuentra en 77 suscripciones a internet móvil por cada 100 habitantes.

Gráfica A5. Teledensidad de internet móvil



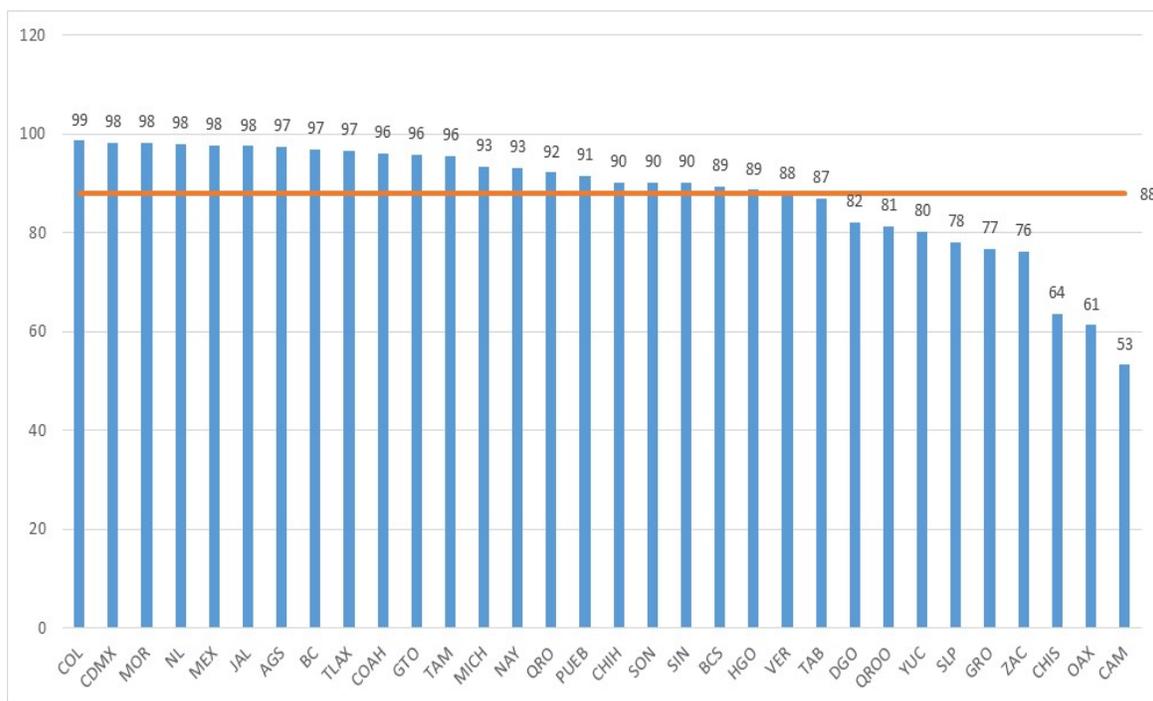
La Gráfica A6 corresponde a la variable de penetración de banda ancha fija con acceso de fibra óptica por cada 100 hogares. El primer lugar lo ocupa CDMX con 42 accesos por cada 100 hogares, seguido de Nuevo León y Baja California Sur, ambos con 29 accesos por cada 100 hogares. Chiapas, Zacatecas y Oaxaca presentan el mayor rezago, los tres con 2 accesos de banda ancha fija mediante fibra óptica por cada 100 hogares. Esta es sin lugar a dudas el indicador que muestra una mayor desigualdad. Por su parte el promedio nacional corresponde a 11 accesos por cada 100 hogares.

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



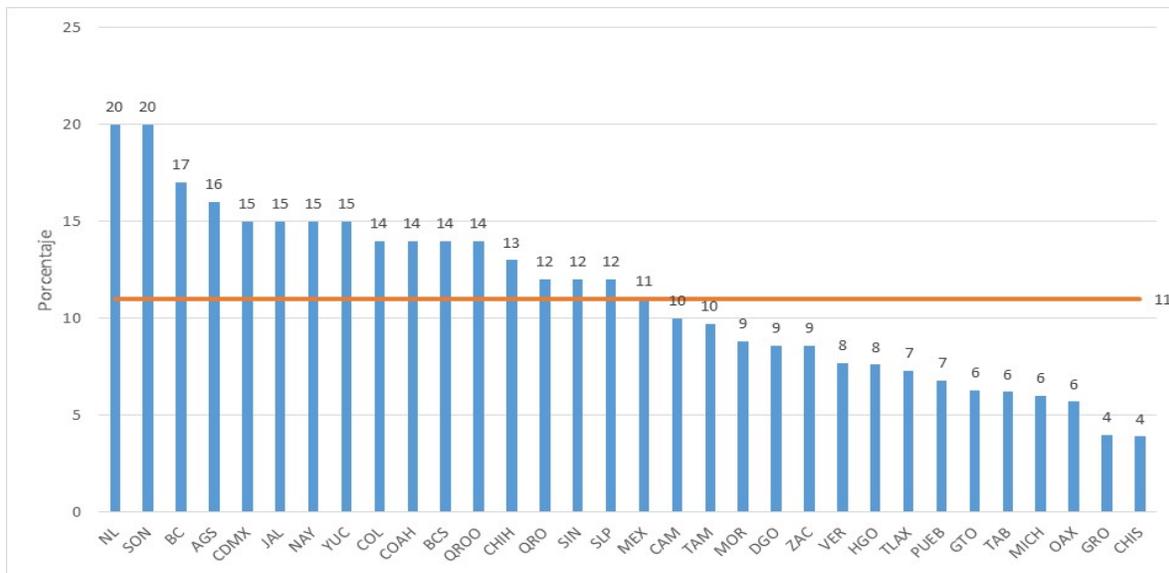
La Gráfica A7 muestra el porcentaje de población cubierta por tecnología de red 4G para servicio de telefonía móvil. El primer lugar es para Colima con 99 por ciento, seguida de CDMX y Morelos con 98%, respectivamente. La menor cobertura de 4G se registra en Chiapas, Oaxaca y Campeche con 64%, 61% y 53%, respectivamente. En cuanto a la media nacional ésta es igual a 88%.

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



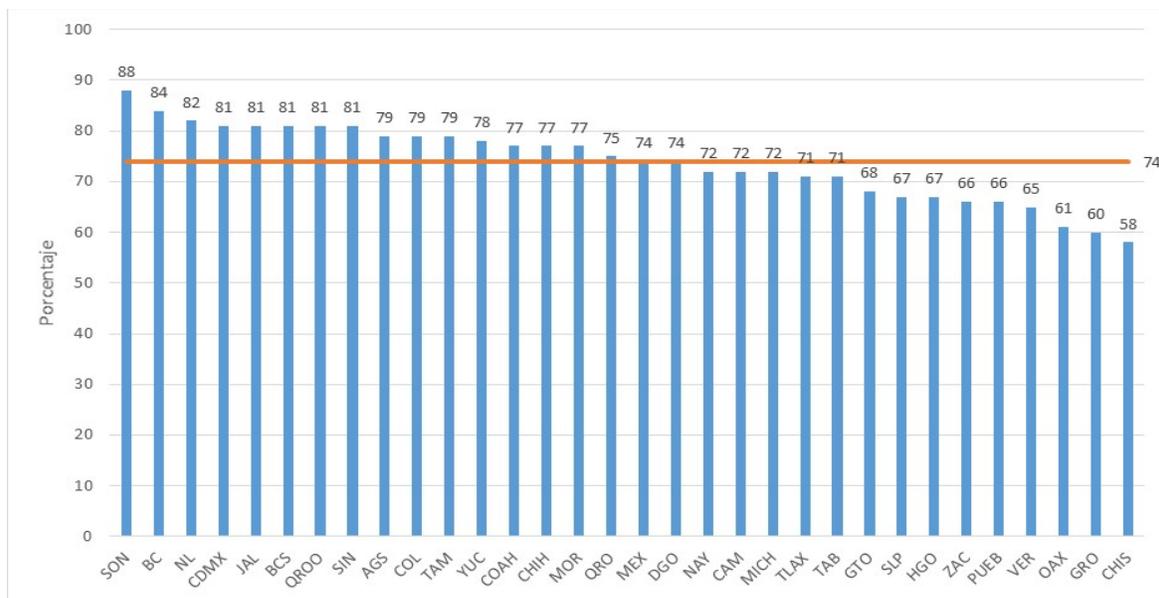
Considerando los indicadores de adopción de TICs, la Gráfica A8 muestra el porcentaje de usuarios que se conectan a internet a través de TV inteligente. Los estados de Sonora y Nuevo León ocupan la primera posición con el 20 por ciento, seguidos de Baja California con el 17 por ciento. Los puestos más rezagados son para las entidades federativas de Oaxaca, Guerrero y Chiapas con 6%, 4% y 4%, respectivamente. La media para las treinta y dos entidades es de 11%.

Gráfica A8. Porcentaje de usuarios que se conectan a internet a través de TV inteligente



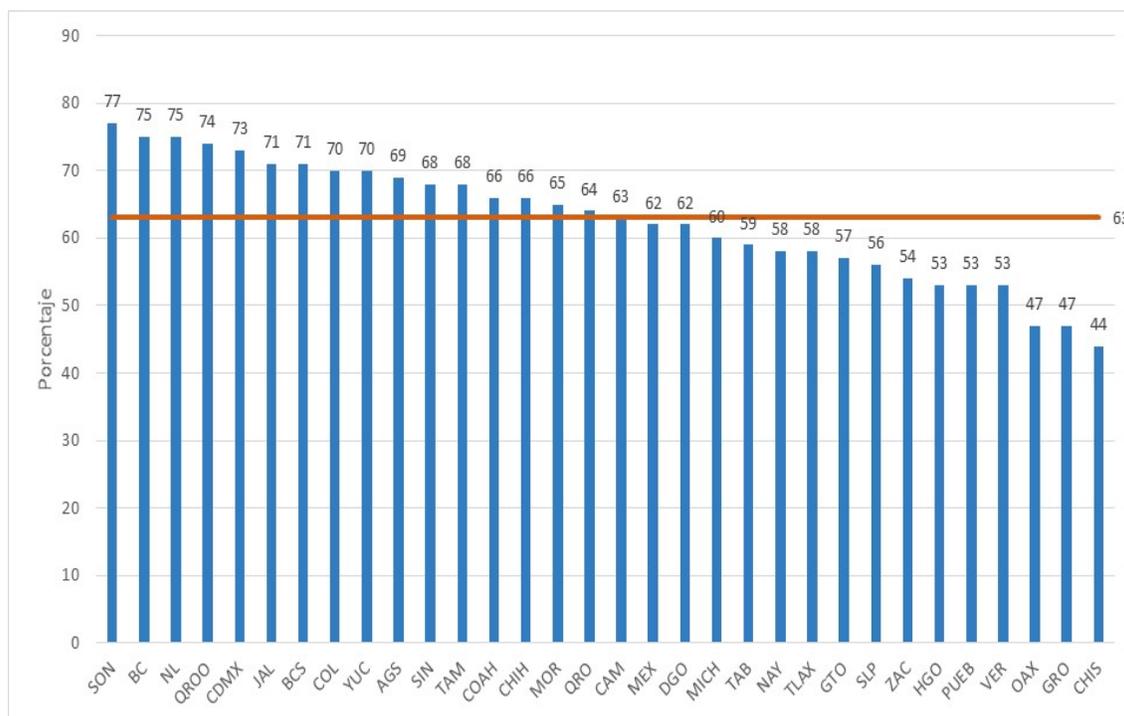
La Gráfica A9 muestra al porcentaje de usuarios que cuentan con teléfono celular siendo Sonora la entidad federativa que ocupa el primer lugar con un 88 por ciento, seguido de Baja California con un 84% y Nuevo León un 82%. En el otro extremo están Oaxaca, Guerrero y Chiapas con 61%, 60% y 58%, respectivamente. El promedio del porcentaje de usuarios a nivel nacional que cuentan con teléfono celular es del 74%.

Gráfica A9. Porcentaje de usuarios que cuentan con teléfono celular



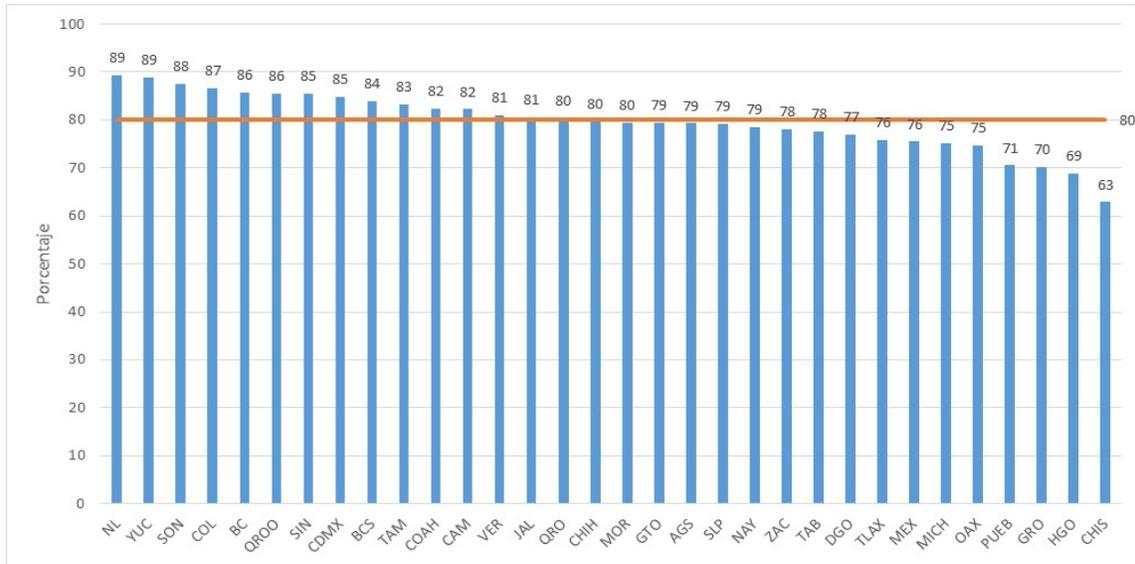
Por su parte, la gráfica A10 presenta el porcentaje de usuarios que cuentan con Smartphone o teléfono inteligente. Sonora ocupa nuevamente el primer lugar con 77 por ciento, seguido de Nuevo León y Baja California con 75 por ciento. Los puestos a la zaga los ocupan Oaxaca, Guerrero y Chiapas con 47%, 47% y 44%, respectivamente. A nivel nacional el promedio es de 63%.

Gráfica A10. Porcentaje usuarios que cuentan con Smartphone



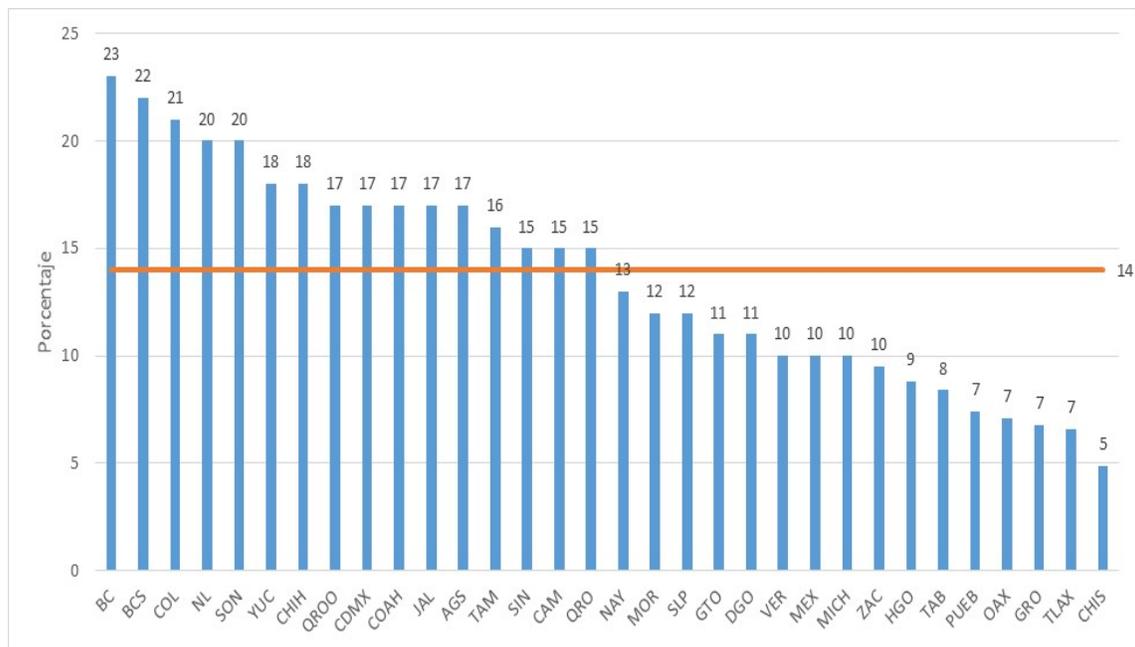
La Gráfica A11 presenta al porcentaje de usuarios de internet que lo utilizan diariamente. Nuevo León y Yucatán ocupan el primer sitio con 89 por ciento, seguidos por Sonora con 88 por ciento. En las posiciones más rezagadas se encuentran Guerrero, Hidalgo y Chiapas con 70%, 69% y 63%, respectivamente. El promedio de los treinta y dos estados es de 80%.

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



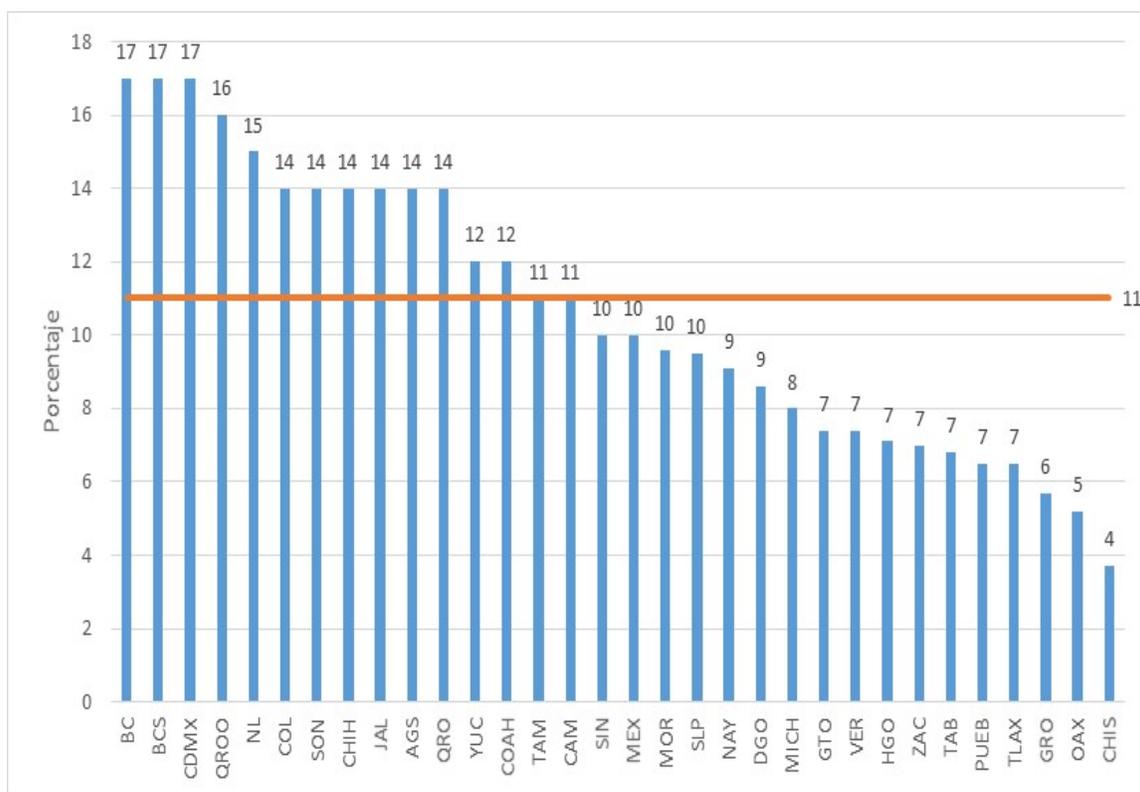
El diagnóstico sobre la inclusión digital para la actividad económica inicia con el porcentaje de usuarios que realizan compras por internet (Gráfica A12). El primer lugar es para Baja California con 23%, seguido de Baja California Sur y Colima con 22 y 21 por ciento, respectivamente. Por su parte Guerrero, Tlaxcala y Chiapas reportan los menores porcentajes con 7%, 7% y 5%, respectivamente. La media nacional es de 14%.

Gráfica A12. Porcentaje de usuarios que realizan compras por internet



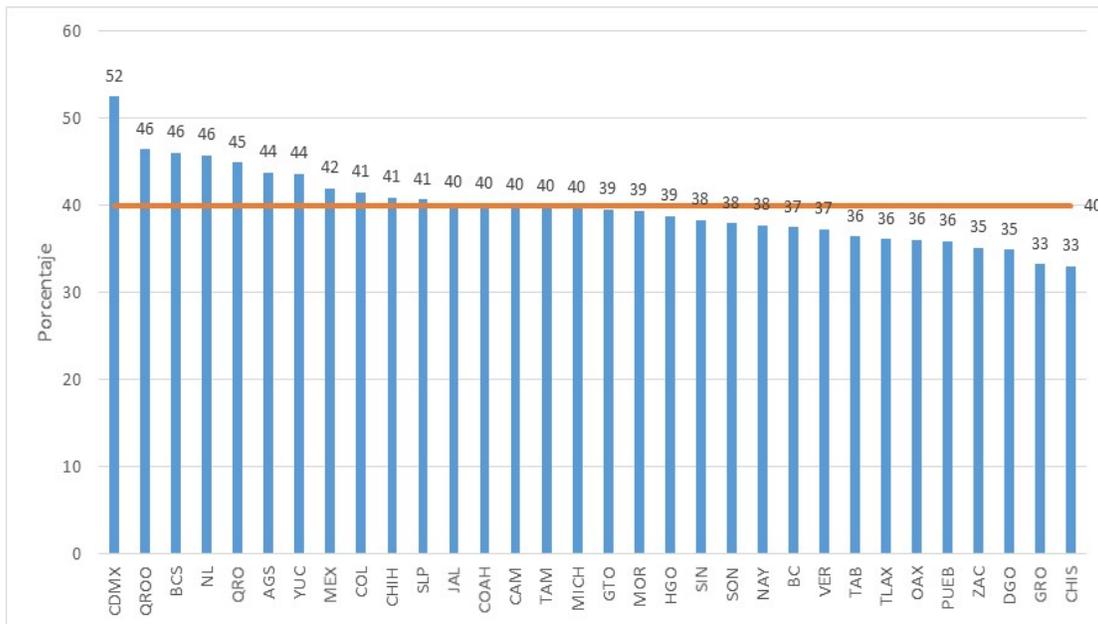
Por su parte, la Gráfica A13 muestra al porcentaje de usuarios que realizan pagos por internet. En este caso los estados de Baja California, Baja California Sur y CDMX reportan el mayor porcentaje, las tres con el 17 por ciento. Los que tienen menor aplicación de pagos en línea son Guerrero, Oaxaca y Chiapas con 6%, 5% y 4%, respectivamente. En este caso se constata que el uso del internet para la realización de pagos es aun baja a nivel nacional (11%). No obstante, no se aprecia una disparidad tan elevada como en otros indicadores analizados.

Gráfica A13. Porcentaje de usuarios que realizan pagos por internet



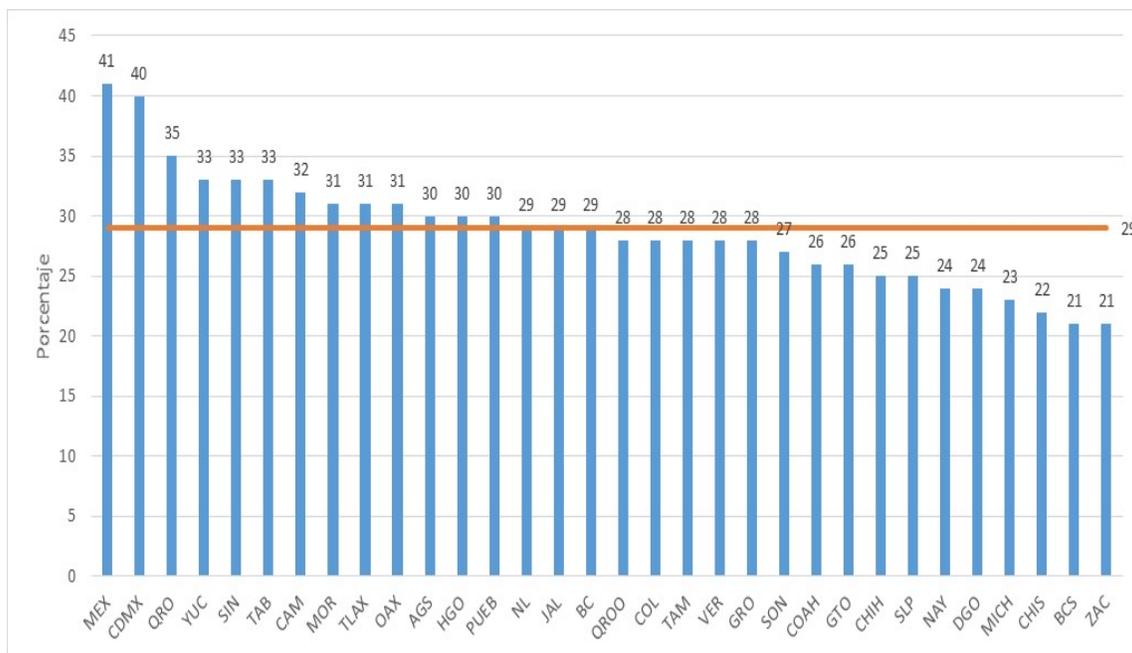
En la Gráfica A14 se observa al porcentaje de usuarios que utilizan computadora para realizar actividades laborales. El primer sitio es para CDMX con el 52%, seguido de Quintana Roo y Baja California Sur con 46%. Las posiciones más rezagadas las ocupan Durango, Guerrero y Chiapas con 35%, 33% y 33%, respectivamente. Por su parte la media nacional se encuentra al 40%.

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



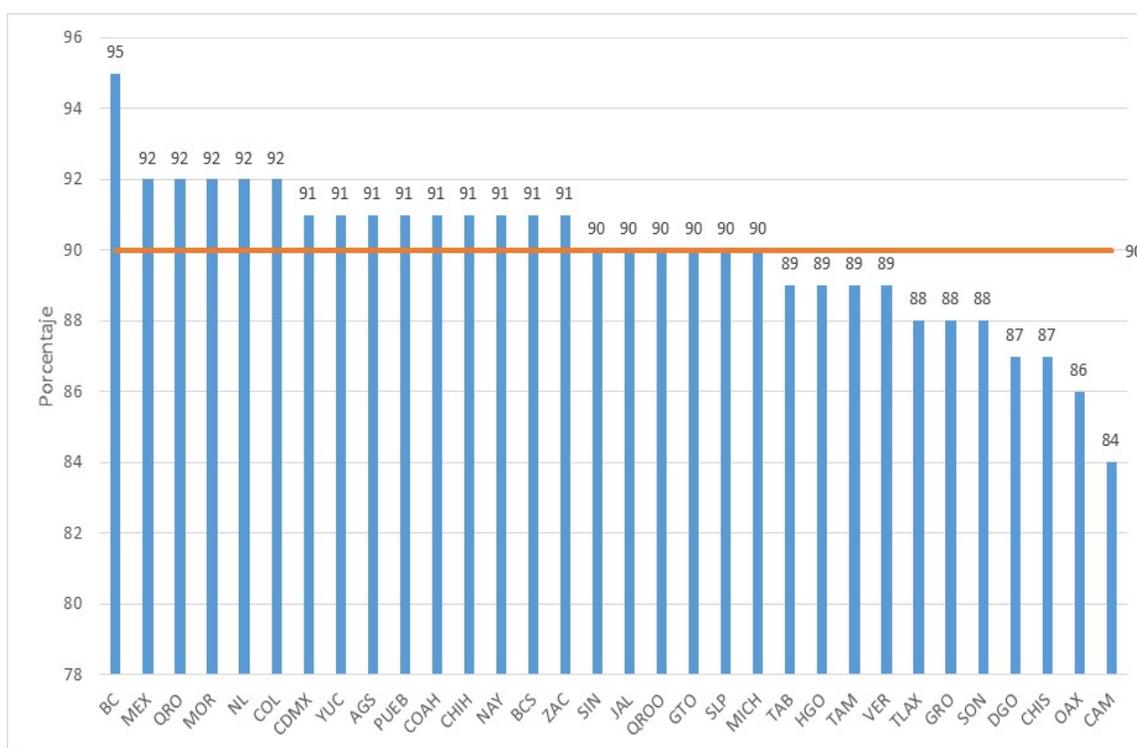
La Gráfica A15 presenta al porcentaje de usuarios de internet que realizan interacciones con el gobierno. El primer sitio es para el Estado de México con 41 por ciento, seguido por CDMX y Querétaro con 40 por ciento y 35 por ciento, respectivamente. En los lugares más rezagados se encuentran Chiapas, Baja California Sur y Zacatecas con 22%, 21% y 21%, respectivamente. A nivel nacional el promedio alcanza el 29%.

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



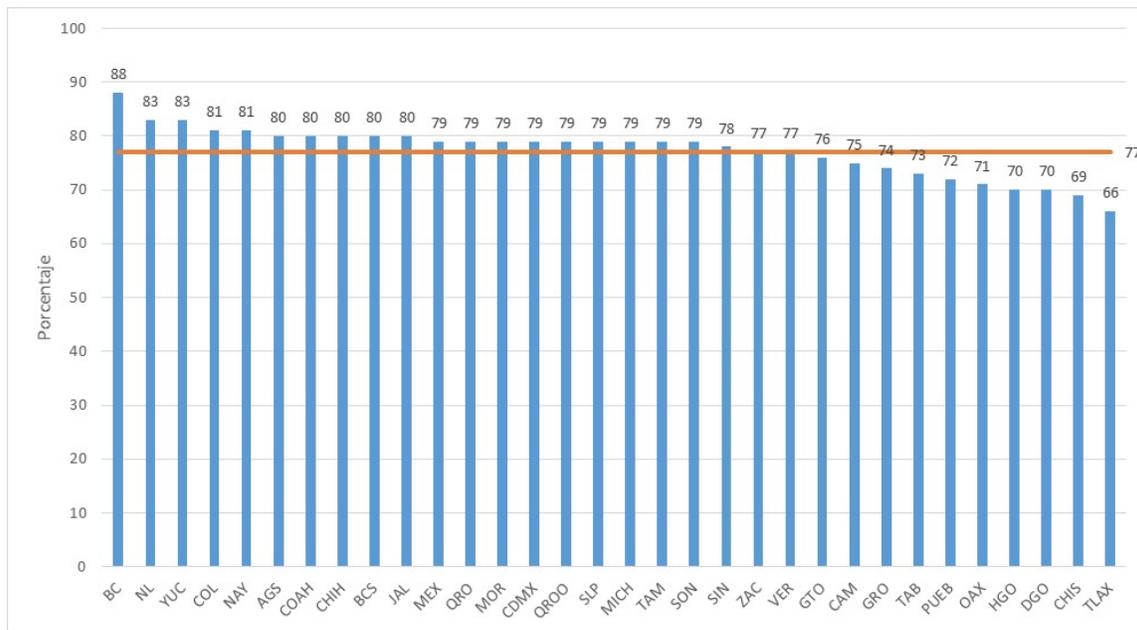
En la Gráfica A16 se observa al porcentaje de usuarios de internet que lo utilizan para entretenimiento. Esta es una modalidad de uso que cuenta ya con una elevada aceptación entre los usuarios en todas las entidades federativas. El mayor uso lo reporta Baja California con 95 por ciento, seguido del Estado de México y Querétaro con 92% y 92%, respectivamente. El menor uso se reporta en Chiapas, Oaxaca y Campeche con 87%, 86% y 84%, respectivamente. La media obtenida entre las treinta y dos entidades federativas es de 90%.

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



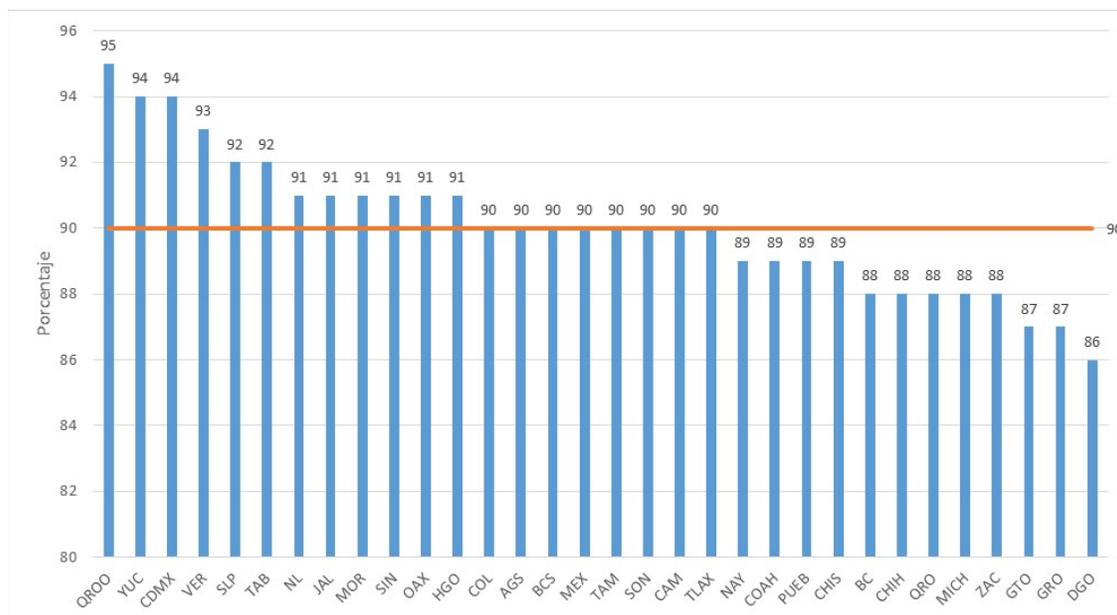
En la Gráfica A17 se observa al porcentaje de usuarios de internet que lo utilizan para ver contenidos audiovisuales. Baja California se encuentra a la cabeza con un 88%, seguido de Nuevo León y Yucatán con 83%. En el otro extremo se encuentran Durango, Chiapas y Tlaxcala con 70%, 69% y 66%, respectivamente. El promedio nacional es de un 77%.

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



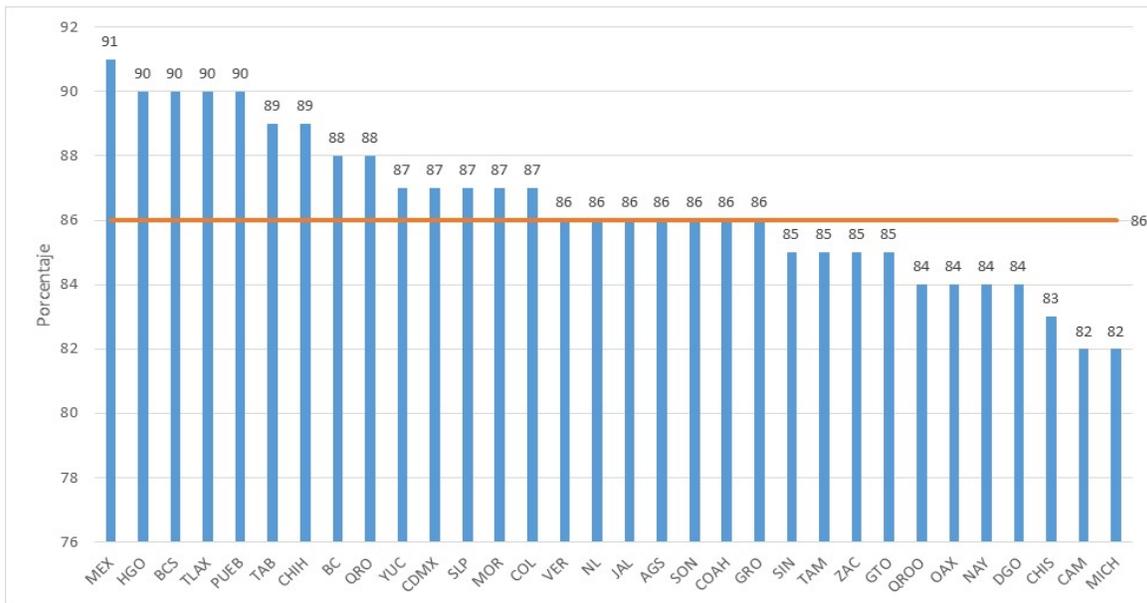
La Gráfica A18 presenta el porcentaje de usuarios de internet que lo utilizan para comunicarse. En el primer sitio se encuentra Quintana Roo con 95%, seguido de Yucatán y CDMX con 94 %. En los sitios más rezagados se encuentran los estados de Guanajuato, Guerrero y Durango con 87%, 87% y 86%, respectivamente. La media nacional corresponde a un 90%.

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



En la Gráfica A19 se muestra el porcentaje de usuarios de internet que lo utilizan para obtener información. El primer sitio es para el Estado de México con 91%, seguido de Hidalgo y Baja California Sur con 90%. Chiapas, Campeche y Michoacán reportan las menores cifras con 83%, 82% y 82%, respectivamente. En cuanto al promedio nacional, éste es igual a un 86%.

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



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*²⁸ (Gráfica A20) el cual es de utilidad para determinar el número de grupos a utilizar. La forma de interpretar la gráfica *Screeplot*²⁹ 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 van aumentando la cantidad de *clústers*. Es así que la gráfica sugiere trabajar con seis *clústers*, tal y como lo señala el círculo de color rojo, debido a que un séptimo grupo no mejora el modelo.

Una vez determinado el número de *clústers* con el cual se va a trabajar se inicia con la aplicación de un método jerárquico conocido como el método Ward³⁰. Este método construye una estructura conocida como dendograma, en 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 para cada *clúster* 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 rectángulo³¹.

Al usar seis *clústers* (Gráficas A21 y A22) se observa que la Ciudad de México queda aislada cuando se utiliza la distancia Euclidiana; en tanto que bajo la distancia Manhattan se agrupa junto a las entidades federativas de Baja California, Baja California Sur, Quintana Roo y Nuevo León. El resto de *clústers* varían para ambas distancias, salvo para el caso del que conforman los estados de Chiapas, Guerrero y Oaxaca. Otros dos grupos son muy similares, y sólo se

²⁸ *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.

²⁹ <https://campus.datacamp.com/courses/unsupervised-learning-in-r/unsupervised-learning-in-r?ex=9>

³⁰ Peña, D., 2002. *Análisis de Datos Multivariantes*, McGraw Hill

³¹ <https://xlinux.nist.gov/dads/HTML/manhattanDistance.html>

diferencian porque se intercambia el estado de Veracruz. Estos grupos son los que conforman bajo la distancia Euclidiana, por una parte: Campeche, Durango, Guanajuato, Michoacán, Nayarit, San Luis Potosí y Zacatecas; y, por otra, Hidalgo, Puebla, Tabasco, Tlaxcala y Veracruz.

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 A23). 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:

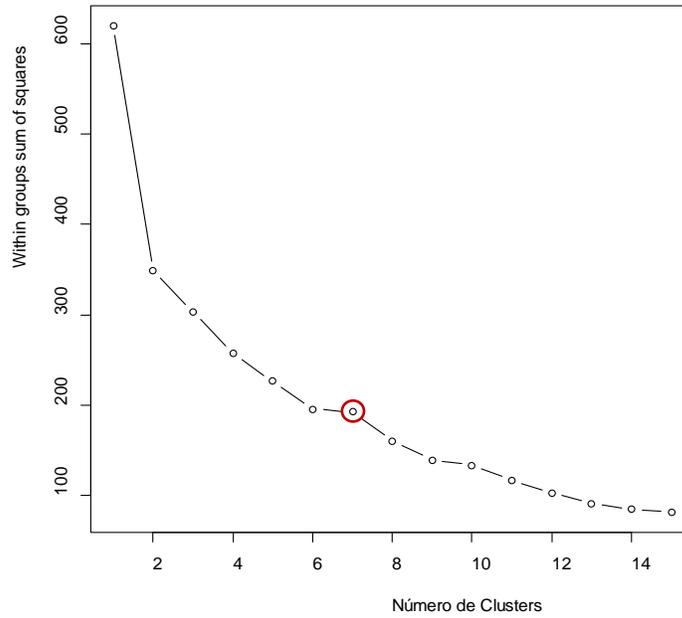
- i. Se especifica el número de *clústers* (k) que serán utilizados.
- ii. A partir de los datos se seleccionan de forma aleatoria k puntos como los centros iniciales (centroide³²) o medias.
- iii. Cada observación es asignada al centroide más cercano, basado en la distancia Euclidiana, o Manhattan, 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.

Aplicando este método para definir seis *clústers* (Gráfica A23) se observa que el primer grupo está conformado exclusivamente por la Ciudad de México. En el segundo grupo se encuentran Baja California, Baja California Sur, Quintana Roo, Nuevo León y Sonora. En el tercer grupo se encuentra únicamente Campeche. En tanto que el cuarto grupo tiene a Aguascalientes, Coahuila, Colima, Chihuahua, Jalisco, Estado de México, Morelos, Querétaro, Sinaloa, Tamaulipas y Yucatán. El quinto *clúster* tiene a los estados de Durango, Guanajuato, Michoacán, Nayarit, San Luis Potosí, Veracruz y Zacatecas. Por último, en el sexto grupo están Chiapas, Guerrero, Hidalgo, Oaxaca, Puebla, Tabasco y Tlaxcala.

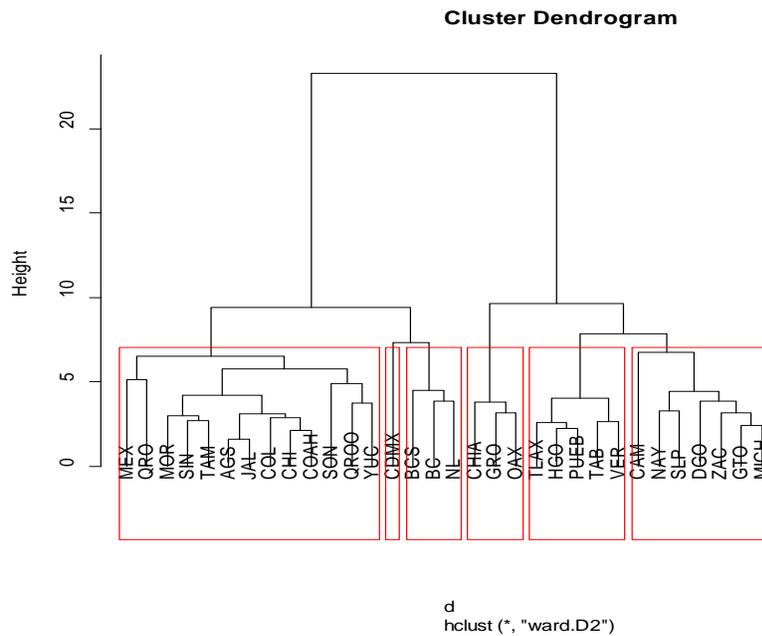
Los resultados se resumen en el Cuadro A1. El ejercicio estadístico completo considera que seis es el número óptimo de *clústers* a considerar.

³² 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.

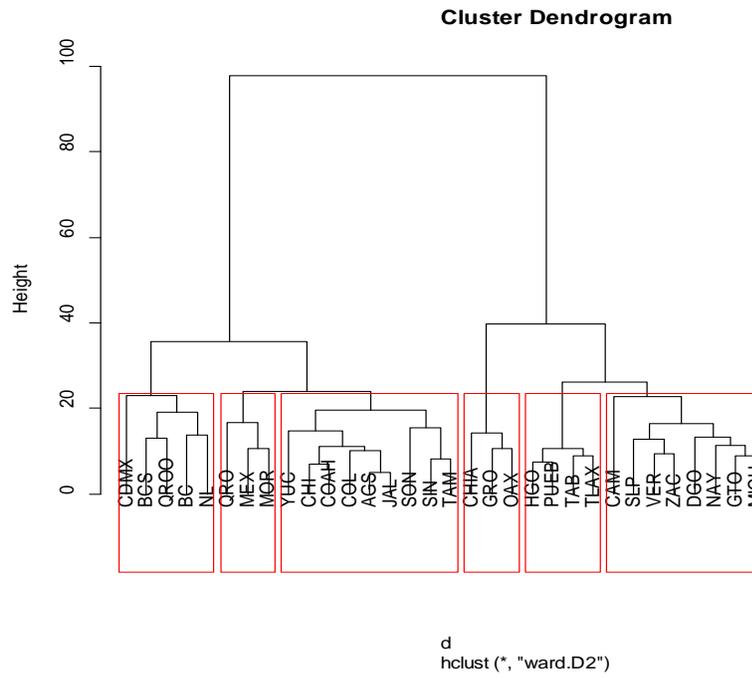
Gráfica A20. Screeplot



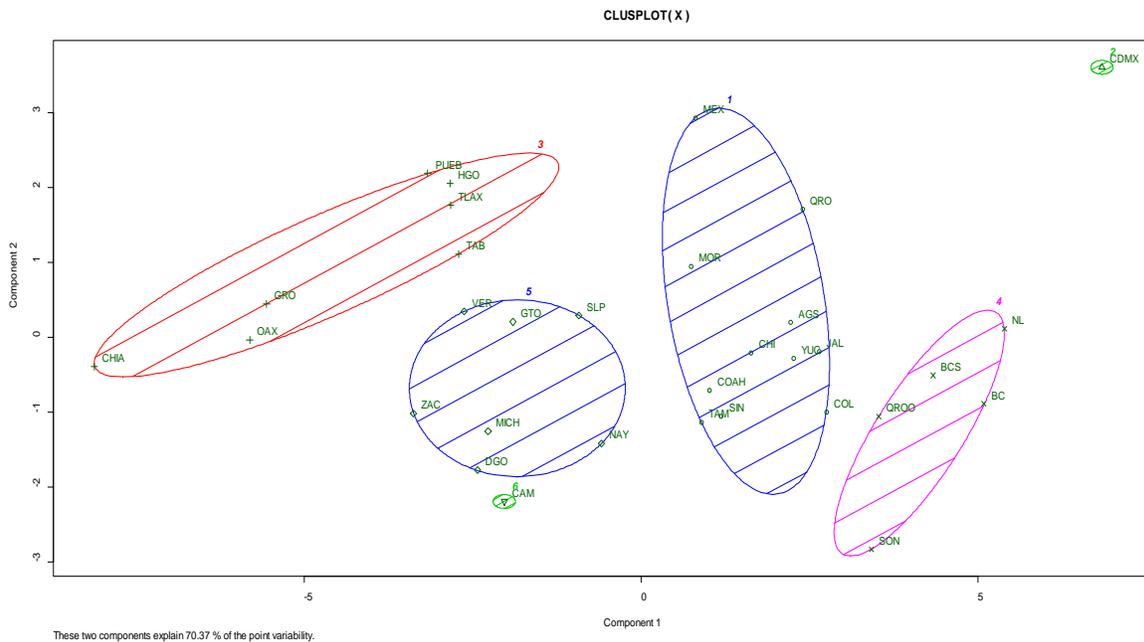
Gráfica A21. Método Ward. Dendograma-Distancia Euclidiana con seis clústers



Gráfica A22. Método Ward. Dendrograma-Distancia Manhattan con seis clústers



Gráfica A23. Método K-Medias con seis clústers



Cuadro A1. Resultados para seis agrupamientos según Metodología de Clústers

Método Ward		K-medias
Distancia Euclidiana	Distancia Manhattan	
CDMX	CDMX, BC, BCS, NL, QROO	CDMX
BC, BCS, NL	AGS, COAH, COL, CHIH, JAL, SON, SIN, TAM, YUC,	BC, BCS, NL, QROO, SON
CAM, DGO, GTO, MICH, NAY, SLP, ZAC	CAM, DGO, GTO, MICH, NAY, SLP, VER, ZAC,	CAM
AGS, COAH, COL, CHIH, JAL, MEX, MOR, QRO, QROO, SON, SIN, TAM, YUC	QRO, MEX, MOR	AGS, COAH, COL, CHIH, JAL, MEX, MOR, QRO, SIN, TAM, YUC
HGO, PUE, TAB, TLAX, VER	HGO, PUE, TAB, TLAX	DGO, GTO, MICH, NAY, SLP, VER, ZAC
CHIA,GRO OAX	CHIA, GRO,OAX	CHIA, GRO, HGO, OAX, PUE, TAB, TLAX

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 principales³³, 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

³³ Peña, D., 2002. Análisis de Datos Multivariantes, McGraw Hill.

(en este caso las treinta y dos entidades federativas) en un set de p componentes principales no correlacionados sobre los mismos sujetos.

En virtud de que el set de datos cuenta con 19 indicadores se puede decir que éste tiene 19 dimensiones. Por lo tanto, el número de eigenvectores y sus eigenvalores asociados será también de 19 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 eigen valor 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 cuatro dimensiones y su suma para obtener el resultado final, el cual se presentó en el Cuadro 3 del reporte.