



APLICACIÓN DE CIENCIA DE DATOS EN TURISMO NOA

Modalidad: Trabajo de aplicación de conceptos y técnicas de administración en situación laboral de revista o ambiente real.

Año: 2024

Autor: Miguel Alberto Esper

E-Mail: alberto.esper23@gmail.com





Índice

Resumen	3
Palabras Clave	4
Introducción	4
Situación Problemática	5
Preguntas de Investigación	5
Objetivo General	6
Objetivos Específicos	6
Marco Metodológico	7
Marco Teórico	7
Aplicación	15
Encuesta	16
Preparación de base de datos	18
Análisis Clúster	18
Análisis de Correspondencia Múltiple	28
Otros análisis	38
Recomendaciones	39
Conclusiones	40
Bibliografía	41
Anexo	42



Resumen

El turismo es la actividad que realiza una persona con el objetivo de viajar por placer, recreación, ocio o cualquier otro motivo que no esté relacionado directamente con el trabajo. Puede manifestarse de diversas formas, que van desde viajes cortos de fin de semana hasta vacaciones prolongadas, las cuales pueden incluir una amplia gama de actividades.

En el ámbito de la investigación turística, la combinación de enfoques cuantitativos, cualitativos y mixtos desempeñan un papel crucial para comprender la complejidad del comportamiento y preferencias del turista. En este contexto, según Hernández Sampieri (2018) destaca a la investigación como un proceso sistemático, crítico y empírico que se aplica al estudio de un fenómeno o problema con el resultado de ampliar su conocimiento.

El siguiente trabajo de investigación, coloca mayor énfasis en aquellas personas no residentes de la provincia de Tucumán, ya que son aquellos que pueden ofrecer una perspectiva única y objetiva sobre el destino en específico, debido a que no están influenciados por locales o percepciones arraigadas.

El desarrollo tiene como desafío dejar de lado el estudio univariado y aprovechar las múltiples variables aplicando herramientas de datos como ser Análisis de Correspondencia Múltiple y Análisis Clúster. El enfoque metodológico aplicado será el mixto de diseño concurrente y modelo dominante cuantitativo, en donde los datos obtenidos tendrán origen en una encuesta dirigida a una muestra total de 149 personas, con un total de 23 variables distribuidas en preguntas abiertas y cerradas.



Los resultados obtenidos a través de la investigación permitirán proporcionar a las entidades interesadas datos relevantes con el fin de poder inferir en el comportamiento de mercado turístico de la provincia de Tucumán.

Palabras Clave: turismo, ciencia de datos, correspondencia múltiple, clúster.

Introducción

Hoy en día la tecnología se ha convertido en un motor de cambio profundo y transformador en todos los aspectos sociales y empresariales de la actualidad. El desarrollo tecnológico supera los límites y la forma en que interactuamos, trabajamos e incluso la forma en que enfrentamos los desafíos. La tecnología, poniendo énfasis en la ciencia de datos y la inteligencia artificial han abierto nuevas perspectivas y oportunidades para comprender y mejorar el entorno cambiante en el cual que vivimos.

En los que respecta a la ciencia de datos, su función principal es transformar datos en conocimientos prácticos. Al analizar patrones y relaciones, permite a las empresas o entidades públicas tomar decisiones y cursos de acción para el cumplimiento de sus objetivos.

Cabe destacar, que la revolución provocada por la Inteligencia artificial permite que las máquinas aprendan, se adapten y mejoren en otorgar respuestas llevando a la automatización y análisis a niveles superiores. Debido a esto, Turismo Noa, una empresa a través de la cual se da origen a este trabajo de investigación, se destaca



por brindar servicios de excelencia en cuanto a la investigación y planificación de estrategias abocadas al turismo.

El presente trabajo de investigación determina el comportamiento del mercado turístico de la provincia de Tucumán en base a la aplicación de herramientas de recolección de datos, entre los cuales serán utilizados encuestas dirigidas a personas no residentes y herramientas de análisis de datos, como ser el Análisis de Correspondencia Múltiple (ACM) y Análisis de Clúster con el fin de analizar las relaciones entre las variables y establecer conclusiones.

Situación Problemática

En el ámbito del turismo, entender los elementos que impactan en el comportamiento del mercado es fundamental para alcanzar el éxito. Tradicionalmente, el análisis univariado ha sido utilizado para examinar variables de manera individual, proporcionando información limitada en relación a las interacciones entre variables. El desafío de este trabajo de investigación, es explorar las múltiples relaciones entre las variables mediante la aplicación de herramientas de datos como ser el análisis de correspondencia múltiple y análisis clúster para proporcionar una visión más holística y significativa del comportamiento del mercado turístico.

Preguntas de Investigación

1. ¿De qué manera se está aplicando actualmente la inteligencia artificial y la ciencia de datos en el ámbito empresarial?



2. ¿Cuáles son las herramientas más apropiadas para llevar a cabo el análisis y la estructuración de los datos, considerando su aplicación en el ámbito empresarial?
3. ¿Cómo puede las herramientas de análisis de datos brindar una ventaja competitiva para la empresa?

Objetivo General

El objetivo general, consiste en la utilización de las herramientas propias de la ciencia de datos e inteligencia artificial (IA), que permiten el procesamiento, análisis y estructuración de la información precisa para la toma de decisiones.

Objetivos Específicos

- Comparar los diferentes casos aplicación de inteligencia artificial y ciencia de datos en empresas del sector turístico.
- Evaluar herramientas y tecnologías disponibles para el análisis y estructuración de datos, identificando ventajas, limitaciones y aplicaciones específicas.
- Determinar como la implementación de herramientas de análisis de datos puede mejorar la eficiencia operativa, la toma de decisiones y la satisfacción del cliente, y como estas mejoras pueden traducirse en una ventaja competitiva para la empresa.



Marco Metodológico

El siguiente trabajo de investigación adopta un enfoque mixto, de diseño concurrente y modelo dominante cuantitativo. La muestra está compuesta por personas que van desde los 18 años hasta más de 55 años, ubicados en diversas provincias de la Argentina.

Desde el punto de vista cuantitativo, la información es extraída de cuestionarios proporcionados a las personas, y analizada con herramientas de datos con el fin de determinar las relaciones entre las variables.

Marco Teórico

Turismo

El turismo es un sector de gran importancia en la economía global, caracterizado por su dinamismo y su capacidad para generar empleo, ingresos y desarrollo en diferentes regiones del mundo. Comprender el comportamiento del mercado turístico es crucial para tomar decisiones estratégicas tanto a nivel corporativo como gubernamental. En este contexto, el uso de herramientas de análisis de datos parece ser una poderosa herramienta para analizar, predecir y optimizar las actividades turísticas.

Factores que afectan el comportamiento del mercado turístico

El comportamiento del mercado turístico está influenciado por una variedad de factores, desde aspectos económicos y políticos hasta aspectos socioculturales y



ambientales. Los principales factores que afectan las decisiones de los turistas y las ofertas turísticas incluyen: condiciones económicas globales y nacionales, cambios en los hábitos de viaje, tendencias demográficas, avances tecnológicos, competencia entre destinos, calidad de los servicios turísticos, seguridad y estabilidad política, y percepción del destino.

Investigación

Según Hernández Sampieri (2018), la investigación es un conjunto de procesos sistemáticos, críticos y empíricos que se aplican al estudio de un fenómeno o problema con el resultado de ampliar su conocimiento. Dentro de la investigación podemos encontrar tres rutas de investigación, cuantitativa, cualitativa y mixta. Los procesos a través del cual se van a analizar los siguientes enfoques, comprende:

1. Observación y evaluación de fenómenos.
2. Suposiciones como consecuencia de la observación y evaluación.
3. Demostración del fundamento de las suposiciones.
4. Proponer nuevas observaciones y evaluaciones para consolidar las suposiciones.

Investigación cuantitativa

Las rutas de investigación cuantitativa se presentan como una herramienta para analizar de manera objetiva diversos aspectos del mercado turístico. A través del uso de datos numéricos y técnicas estadísticas, permiten examinar una amplia gama de variables. El proceso de la ruta cuantitativa es el siguiente:



Figura 1. Ruta cuantitativa

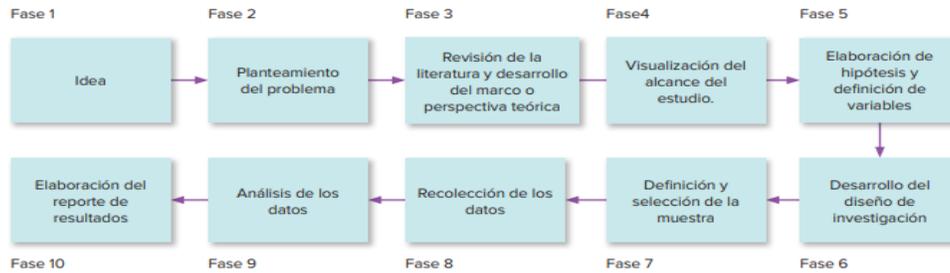


Figura 1.2. Proceso cuantitativo.

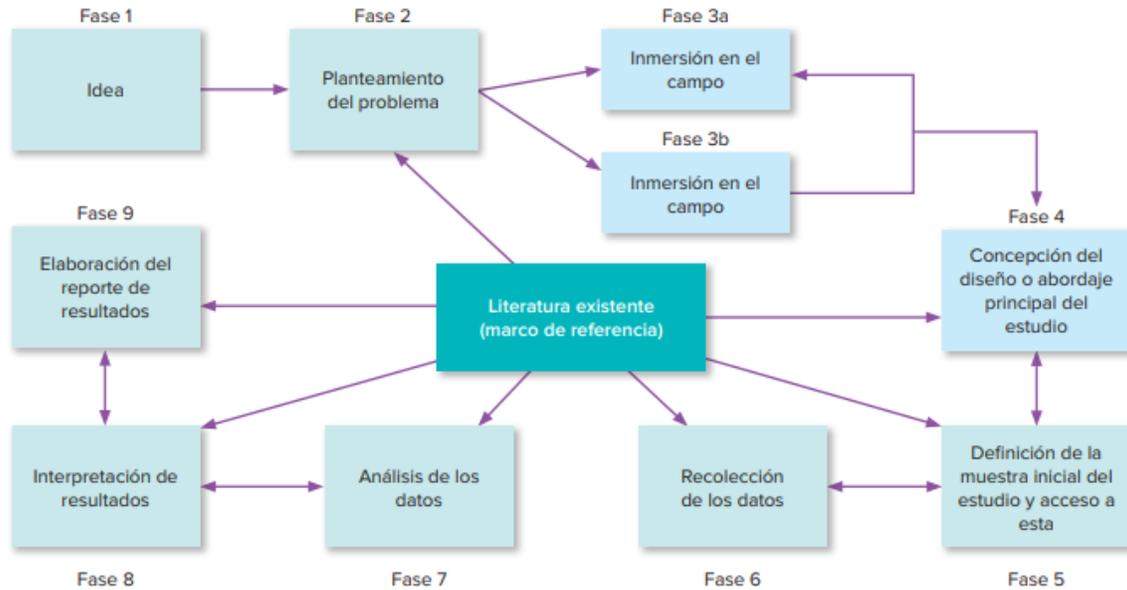
Fuente: Hernández Sampieri. (2018). Metodología de la investigación, Cap. 1, pág. 6

Investigación cualitativa

Según Hernández Sampieri (2018), la investigación cualitativa es un proceso que comienza examinando los hechos en sí y revisando los estudios previos, ambas acciones de manera simultánea, a fin de generar una teoría que sea consistente con lo que se está observando. Una de las principales fortalezas de las rutas cualitativas es la capacidad para capturar la complejidad de las interacciones humanas en el contexto bajo estudio, brindando la oportunidad de identificar tendencias emergentes y fenómenos no previstos que pueden ser cruciales para el diseño de estrategias. El proceso cualitativo abarca las siguientes etapas:



Figura 2. Rutas cualitativas



Fuente: Hernández Sampieri. (2018). Metodología de la investigación, Cap. 1, pág. 8

Investigación mixta

Comprende la combinación de datos cuantitativos y cualitativos, aprovecha la combinación de ambos para obtener una comprensión más completa y holística del mercado bajo estudio. Referido al mercado turístico, el enfoque mixto permitirá a los investigadores combinar métodos con el fin de optimizar la recopilación y análisis de los datos y maximizar la fiabilidad de los resultados obtenidos.

Inteligencia Artificial

Según Ray Kuzweil (2005), La inteligencia artificial es la ciencia de crear máquinas que puede pensar y actuar como los humanos. La IA ofrece diversas capacidades que pueden ser fundamentales en diversas áreas. Entre ellas:



1. Automatización de tareas: automatizar tareas repetitivas y rutinarias, liberando tiempo y recursos humanos para actividades más estratégicas y creativas.
2. Procesamiento avanzado de datos: procesar grandes volúmenes de datos de manera eficiente, identificando patrones y tendencias que podrían pasar desapercibidos mediante métodos tradicionales.
3. Aprendizaje automático: permite que los sistemas mejoren con la experiencia, adaptándose a cambios en los datos y mejorando su rendimiento con el tiempo.
4. Tomas de decisiones basada en datos: analizar datos complejos y proporcionar información valiosa para respaldar la toma de decisiones informada y estratégica.
5. Optimización de procesos: contribuye a la eficiencia y reducción de costos.
6. Innovación continua: abre nuevas posibilidades en áreas como la medicina, la investigación científica, la fabricación y otros sectores, impulsando el descubrimiento de soluciones creativas.

Es necesario reconocer la relación estrecha entre la Inteligencia artificial y la ciencia de datos, esta inferencia entre ambas ha transformado la manera en que se abordan los desafíos analíticos y las tomas de decisiones basadas en datos. No solo ha enriquecido las capacidades de la ciencia de datos, sino que también proporciona a la inteligencia artificial un terreno fértil para aplicar y mejorar sus algoritmos y modelos.

Ciencia de datos

La ciencia de datos, en la actualidad radica en la capacidad para extraer grandes cantidades de datos y transformarlo en conocimiento significativo, en base a



esto, DJ Patil y Jeff Hammerbacher (2011) definen a la ciencia de datos como el proceso de recopilar, limpiar, analizar y visualizar datos para obtener información y conocimiento. Es preciso mencionar dos aspectos fundamentales que componen a la ciencia de datos, aspectos técnicos el cual comprende habilidades y conocimientos necesarios para trabajar con datos, entre ellos la estadística, minería de datos y visualización. En segundo lugar, los aspectos creativos, en donde la importancia se centra en el pensamiento crítico y original para plantear preguntas relevantes y encontrar soluciones innovadoras. Hay que plantear cuatro pasos fundamentales que posee la ciencia de datos:

1. Entender el problema: El primer paso es comprender el problema que se quiere resolver. Partir de la identificación de los datos y comprender sus limitaciones.
2. Preparar los datos: los datos deber estar limpios y preparados para el análisis. Esto implica una depuración de dato incorrectos y transformarlos para que sean compatibles con los algoritmos de análisis.
3. Analizar los datos: aplicar técnicas estadísticas y de aprendizaje automático para extraer información de los datos.
4. Comunicar los resultados: los resultados del análisis deben ser comunicados de forma clara y concisa a los destinatarios.

Análisis univariado

El análisis univariado, es una técnica estadística que se encarga de analizar una única variable a la vez. Se encarga de describir y comprender la distribución, la



tendencia y la dispersión de una variable específica sin considerar alguna relación con otra variable.

Análisis Multivariado

Según Aldás y Uriel (2017), el análisis multivariado es una herramienta que permite estudiar conjunto de variables de forma simultánea. Dentro de este análisis podemos encontrar dos grupos principales.

- Grupos a priori: son aquellos que el investigador define previamente basándose en conocimientos o relaciones entre variables.
- Grupos a posteriori: son aquellos que se forman a partir de los datos analizados.

Las herramientas de análisis multivariado tienen aplicación en la exploración y descripción de datos, esto les permite a los investigadores poder identificar patrones, tendencias y relaciones ocultas entre una multiplicidad de variables. Desde el punto de vista del mercado turístico, tanto las empresas como las entidades públicas intentan comprender las complejas interacciones entre las variables que influyen en las decisiones de los viajeros y en la dinámica del mercado. Entre ellos:

- Segmentación del mercado turístico.
- Predicción de la demanda turística.
- Análisis de satisfacción y lealtad del cliente.
- Evaluación de la competitividad turística
- Análisis del comportamiento de compra.



Análisis clúster

Peña (2002) define al análisis clúster como una técnica multivariante que busca agrupar objetos en función de la similitud entre ellos. Se basa en la idea de que los objetos que son similares en un conjunto de variables también tenderán a ser similares a otras variables. Aldás y Uriel (2018) indican las siguientes recomendaciones a la hora de aplicar clúster.

- Definir claramente los objetivos de análisis.
- Seleccionar cuidadosamente las variables.
- Utilizar diferentes medidas de similitud.
- Validar los grupos identificados.
- Interpretar los grupos en el contexto del problema de investigación.

Análisis de correspondencia

Aldás y Uriel (2005), definen al análisis de correspondencia como una técnica.

- Multivariante: se basa en múltiples variables para analizar la relación entre dos conjuntos de categorías.
- Gráfica: Representa la relación entre las categorías en un espacio bidimensional.
- Descriptiva: no se realiza ninguna inferencia estadística.

Para poder aplicar el análisis de correspondencia es necesario.

1. Recopilar datos y organizarlos en una tabla de contingencias, conocida además como tabla de frecuencia cruzada.



2. Selección de variables y categorías: identificar las variables categóricas que se deseen analizar.
3. Cálculo de la frecuencia esperada: se calculan las frecuencias esperadas para cada celda de la tabla de contingencia bajo la hipótesis nula.
4. Análisis de las desviaciones: se comparan las frecuencias observadas en la tabla de contingencia con las frecuencias esperadas.
5. Representación gráfica: presentación de la información en gráficos de dispersión o diagrama de puntos.
6. Interpretación de los resultados: se interpreta la estructura de asociación revelada por el análisis de correspondencias. Se identifican patrones, relaciones y tendencias entre las categorías de las variables.

Aplicación

Desarrollo

Empresa: Turismo Noa

Servicios: La empresa tienen como funciones principales la planificación e investigación con el objetivo de mejorar las estrategias implementadas y brindar recomendaciones para el desarrollo del turismo, tanto a entidades del ámbito público como privado.



Encuesta

Para la obtención de datos en el trabajo de investigación, se llevó a cabo una encuesta la cual fue elaborada a través de la plataforma Google Form y compartida a los encuestados por medio de WhatsApp. La muestra estuvo compuesta por un total de 250 personas de las cuales respondieron 149. En cuanto a las características demográficas de los encuestados, este compuesto desde menores de 18 años hasta 55 años o más y, residentes de diversas provincias del territorio Argentino.

Tabla 1: Cantidad de personas encuestadas según residencia

Provincia	Cantidad
Buenos Aires	51
Catamarca	2
Chaco	1
Córdoba	14
Corrientes	3
Entre Ríos	2
Formosa	1
Jujuy	1
Mendoza	2
Misiones	4
Neuquén	2
Tucumán	3
Salta	8
San Juan	2
Santa Cruz	2
Santa Fe	11
Santiago del Estero	32
Otro	6
Total general	147

Fuente: Elaboración propia

Se definieron en total 18 variables:



- Variable X1: Edad
- Variable X2: Geográfica
- Variable X3: Experiencia de visita
- Variable X4: Propósito de visita
- Variable X5: Atracciones turísticas
- Variable X6: Calidad (Hotelería)
- Variable X7: Calidad (Gastronomía)
- Variable X8: Calidad (Turismo Cultural)
- Variable X9: Calidad (Turismo Aventura)
- Variable X10: Calidad (Turismo Religioso)
- Variable X11: Calidad (Turismo Eventos)
- Variable X12: Calidad (Servicios Públicos)
- Variable X13: Comparación oferta turística
- Variable X14: Método de descubrimiento
- Variable X15: Recomendación de Tucumán
- Variable X16: Actitud positiva/negativa
- Variable X17: Aspectos positivos
- Variable X18: Aspectos negativos
- Variable X19: Cantidad de Noches (Tucumán)
- Variable X20: Cantidad de Noches (Salta)
- Variable X21: Cantidad de Noches (Jujuy)
- Variable X22: Cantidad de Noches (Catamarca)



Variable X23: Cantidad de Noches (Santiago del Estero)

La encuesta se llevó a cabo en el período 2023, con un total de 23 preguntas entre las cuales existen preguntas abiertas, cerradas y de opción múltiple. La información obtenida del procesamiento de datos servirá para determinar el comportamiento del mercado turístico de Tucumán en relación con las demás provincias.

Preparación de base de datos

1. En primer lugar, se procedió a la revisión de la base de datos, con el fin de identificar posibles problemas relacionados a errores de entrada o inconsistencias.
2. Se continuó con la eliminación de datos sea incongruentes con el objeto bajo estudio.
3. Codificación de las preguntas abiertas y cerradas.
4. Categorización de las preguntas abiertas, la cual consiste en agrupar aquellas respuestas similares de los encuestados y definir las en categorías generales.
5. Por último, se estructuró la base de datos con la información ya organizada, con el fin de facilitar la manipulación y análisis de los datos.

Análisis Clúster

El primer paso para poder aplicar el análisis de clúster fue revisar la base de datos de encuestados y seleccionar aquellas variables cuantitativas que consideramos relevantes para el caso bajo estudio. En este caso, fueron la “recomendación de la provincia de Tucumán” y la “cantidad de noches de alojamiento”. Para poder trabajar



de forma correcto con la información, es necesario eliminar aquellos datos que consideramos atípicos y que entorpecerían el análisis, de esta forma se llevó a cabo el cálculo de la distancia de Mahalanobis.

Tabla 1. Base de datos para el análisis clúster.

<i>Recomendación</i>	<i>Noches de alojamiento</i>
4	1
5	0
7	4
5	0
3	1
6	2

Fuente: Elaboración propia.

Code R Studio

```
#Cálculo de la media
```

```
centro_datos <- colMeans(datos)
```

```
#Cálculo de covarianza
```

```
covarianza <- cov(datos)
```

```
dist_maha <- mahalanobis(datos, centro_datos, covarianza)
```



Tabla 2. Distancia de Mahalanobis

[1]	1.78035747	3.10726154	2.76281822	3.10726154	2.92828043	0.15369757
[7]	0.02714556	0.88718507	1.86424403	2.76281822	2.45283656	9.17739634
[13]	1.84922302	0.88718507	0.36815138	0.15369757	1.80947512	8.62531643
[19]	0.56743975	0.15369757	7.10885875	1.80947512	1.84922302	1.80947512
[25]	3.05411028	1.84922302	10.21797685	2.45283656	0.36815138	0.36815138
[31]	2.45283656	9.96888805	0.36815138	0.74780743	1.09999235	0.02714556
[37]	2.92828043	2.92828043	0.36815138	1.80947512	1.17671505	0.02714556
[43]	1.17671505	0.52714300	1.14193563	1.78035747	0.02714556	1.86424403
[49]	0.02714556	1.86424403	0.88718507	2.38717741	1.86424403	1.86424403
[55]	0.15369757	0.74780743	1.14193563	0.36815138	2.05070677	0.74780743
[61]	3.33870065	0.36815138	0.88718507	0.52714300	0.88718507	2.34121889
[67]	1.14193563	1.14193563	1.80947512	0.36815138	1.17671505	2.34121889
[73]	0.74780743	0.02714556	0.88718507	0.88718507	0.56743975	2.34121889
[79]	0.02714556	0.88718507	1.07529434	1.14193563	1.84922302	1.84922302

Fuente: Elaboración propia mediante Software estadístico R

Una vez determina la distancia de Mahalanobis, procedemos a introducir una sentencia en el Software R que indicará cuales son los valores atípicos que deben ser eliminados de la base de datos.

1. En primer lugar, debemos calcular el “**valor crítico**” = **5.9985658**, el cual servirá como guía para la eliminación de datos atípicos.

- El valor crítico está compuesto por “p”, me indica el número de columnas.

Sentencia: `p <- ncol(datos)`

- Asignación del valor “Alpha”. Sentencia: `Alpha <- 0.05`

- Cálculo de chi-cuadrado para determinar la asociación de las variables.

Sentencia: `qchisq(1 - alpha, df = p)`

2. En segundo lugar, mediante la sentencia:

```
observaciones_atipicas <- which(dist_maha > valor_critico)
```

```
print(observaciones_atipicas), nos permitirá visualizar aquellas observaciones
```

```
que se deben quitar para tener una base de datos escalada.
```



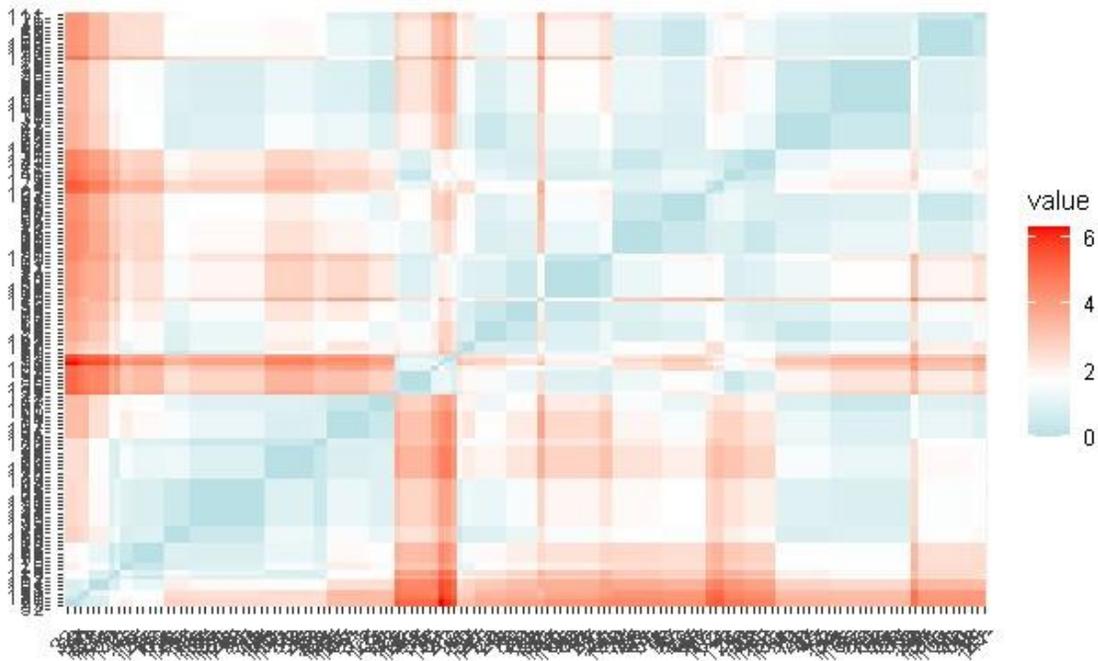
Luego, se procede a realizar el análisis clúster. En primer lugar, es necesario determinar la matriz de distancias. Se decide trabajar con la distancia euclídea.

Code R Studio

```
dist <- hclust(d = dist(x = datos, method = "euclidean"), method = "complete")
```

```
res.dist <- get_dist(datos, stand = TRUE, method = "euclidean")
```

Gráfico 1. Matriz de distancias



Fuente: Elaboración propia mediante Software estadístico R Studio

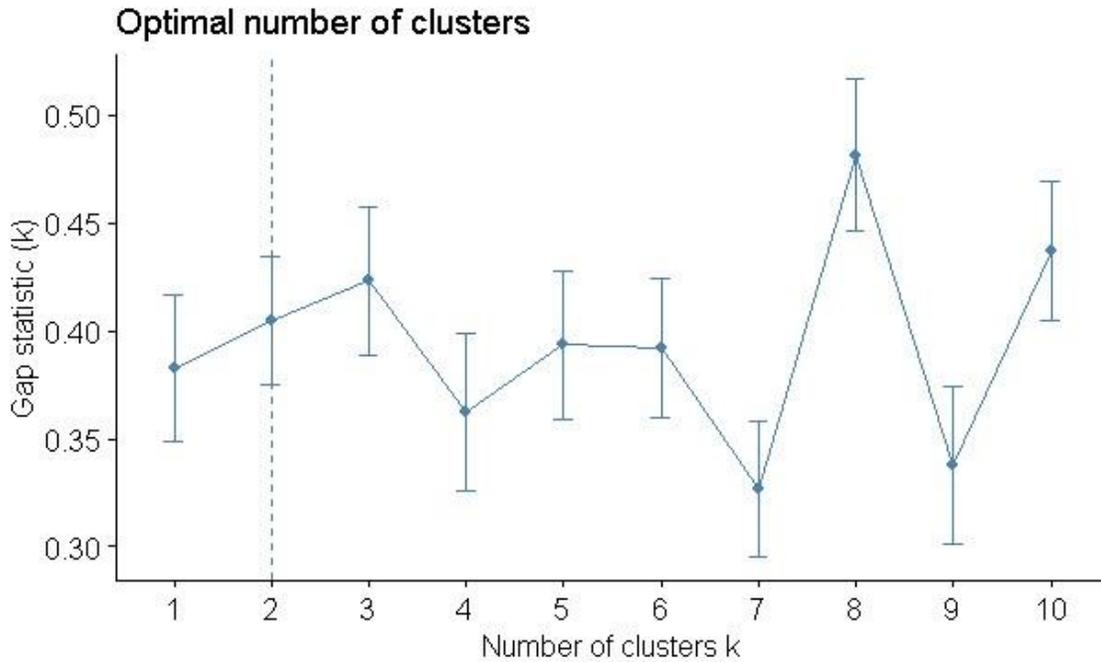
Continuando con el análisis, procedemos a calcular la cantidad óptima de clúster.

Code R Studio

```
fviz_nbclust(datos, kmeans, method = "gap_stat")
```



Gráfico 2. Número óptimo de clúster mediante el método Gap Static



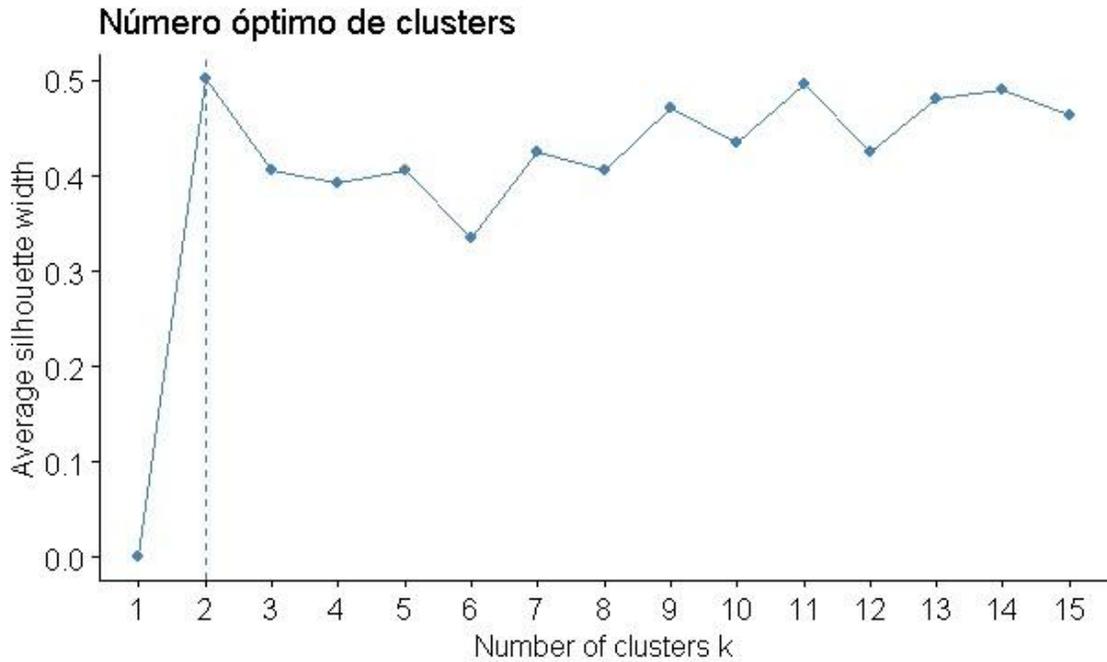
Fuente: Elaboración propia mediante Software estadístico R Studio

Code R Studio

```
fviz_nbclust(x =datos, FUNcluster = kmeans, method = "silhouette", k.max = 15) +  
labs(title = "Número óptimo de clusters")
```



Gráfico 3. Número óptimo de clúster mediante el método Silhouette



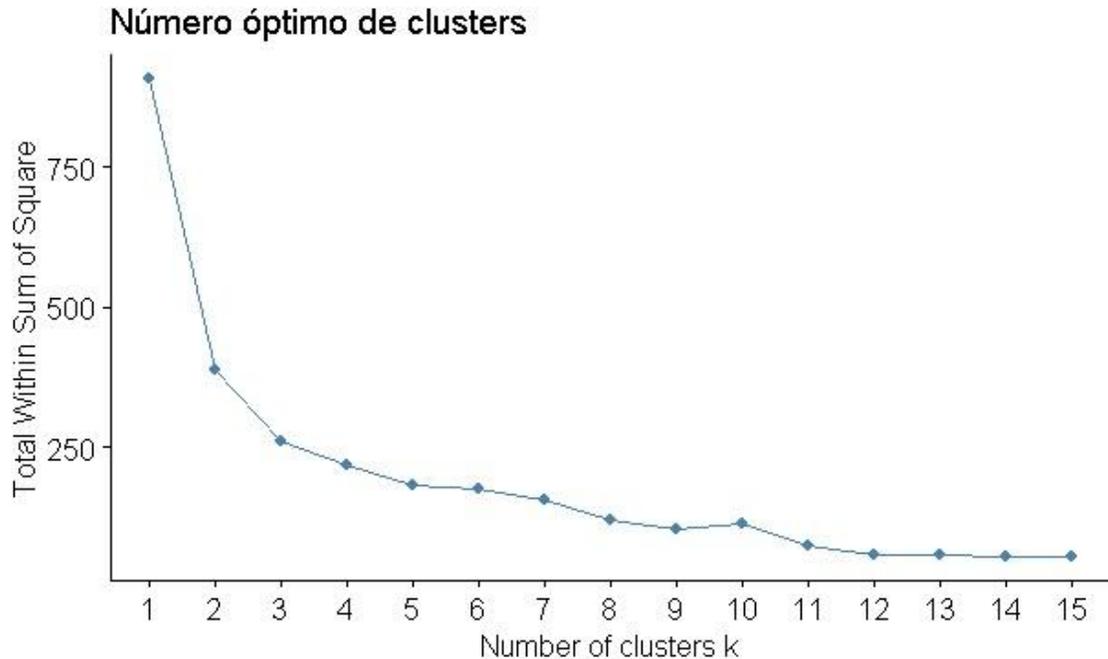
Fuente: Elaboración propia mediante Software estadístico R Studio

Code R Studio

```
fviz_nbclust(x = datos, FUNcluster = kmeans, method = "wss", k.max = 15) +  
  labs(title = "Número óptimo de clusters")
```



Gráfico 4. Número óptimo de clúster mediante el método Wss



Fuente: Elaboración propia mediante Software estadístico R Studio

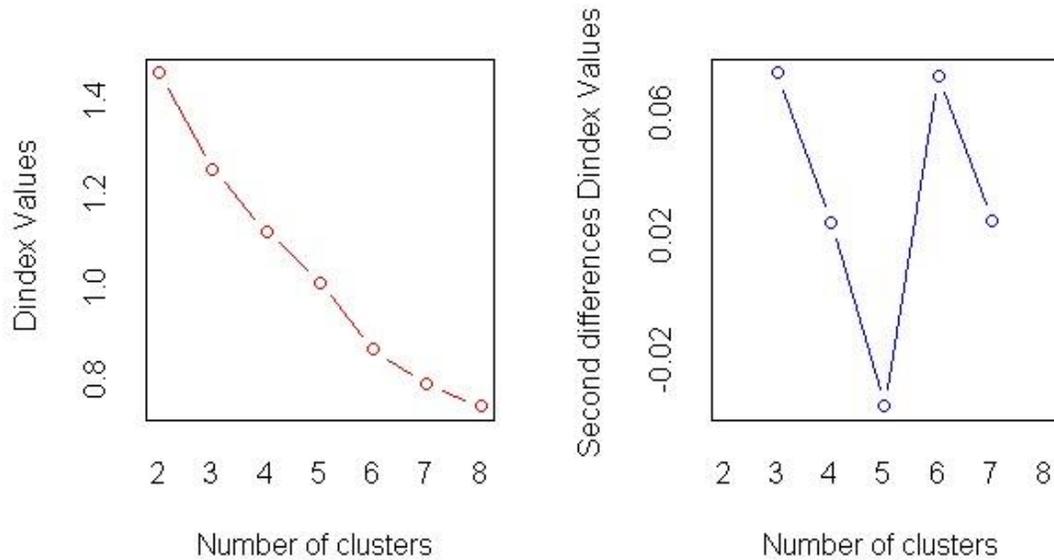
En base a lo observado, por cada uno de los métodos aplicados, la cantidad optima de clúster sería 2, para ellos realizamos una prueba que determinará cual es la prueba más congruente en cuanto a número optimo.

Code R Studio

```
res<-NbClust(datos, distance = "euclidean", min.nc=2, max.nc=8, method ="ward.D2",  
index = "alllong")
```



Gráfico 5. Número óptimo de clústeres



Fuente: Elaboración propia mediante Software estadístico R Studio

Gráfico 6. Número óptimo de clúster aplicando todos los métodos

```
*****  
* Among all indices:  
* 11 proposed 2 as the best number of clusters  
* 1 proposed 3 as the best number of clusters  
* 4 proposed 4 as the best number of clusters  
* 1 proposed 5 as the best number of clusters  
* 5 proposed 6 as the best number of clusters  
* 2 proposed 8 as the best number of clusters  
  
***** conclusion *****  
  
* According to the majority rule, the best number of clusters is 2
```

Fuente: Elaboración propia mediante Software estadístico R Studio

Según la información expuesta, se observa que definitivamente el número óptimo de clúster es 2.

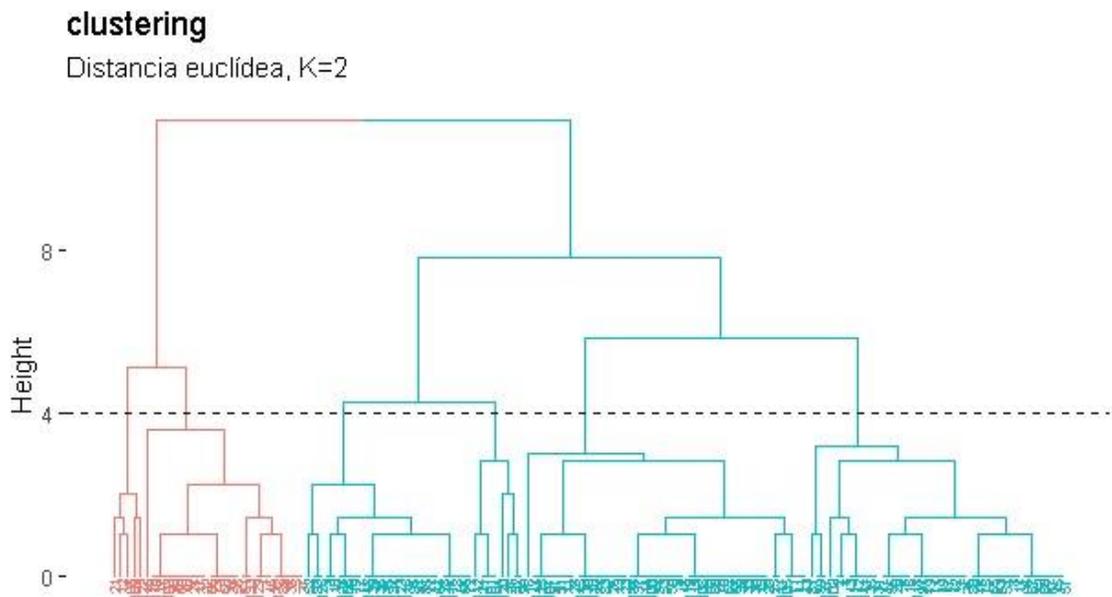


Para obtener una mejor visualización, se procede a la realización de un dendograma con la siguiente sentencia:

Code R Studio

```
fviz_dend(x = dist, k = 2, cex = 0.4) + geom_hline(yintercept = 4, linetype =  
"dashed") + labs(title = "clustering", subtitle = "Distancia euclídea, K=2")
```

Gráfico 7. Dendograma



Fuente: Elaboración propia mediante Software estadístico R Studio

Por último, se decide formar dos clústeres, aplicando el método k-means.

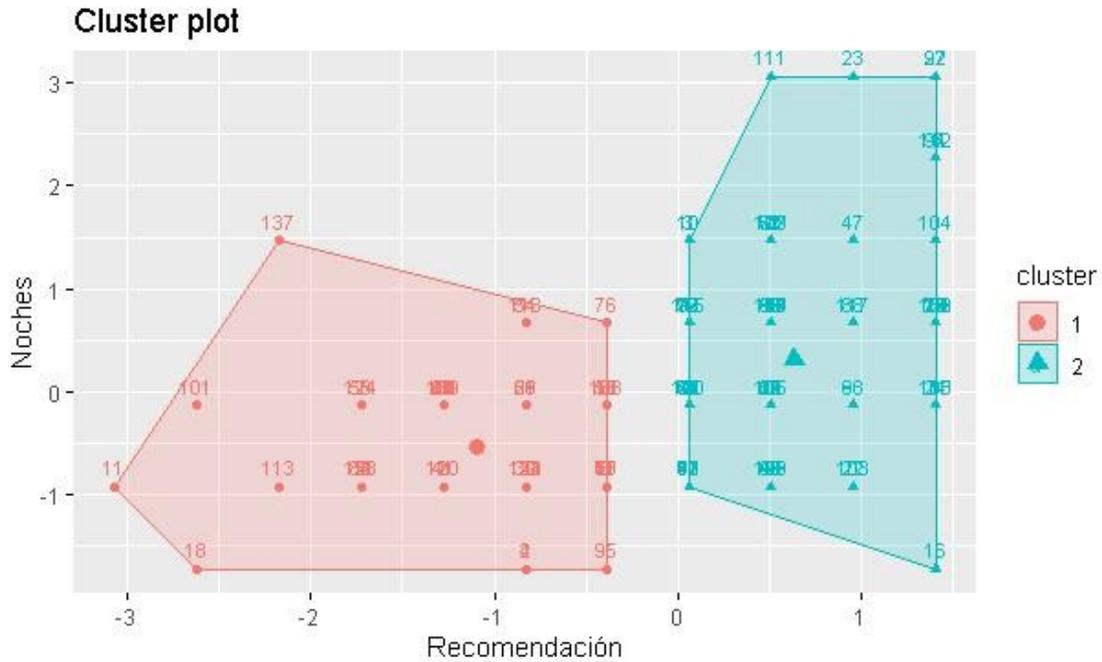
Code R Studio:

```
km.res <- kmeans(datos, 2)
```

```
fviz_cluster(km.res, data = datos, labelsize = 9)
```



Gráfico 8. Clústeres definidos



Fuente: Elaboración propia mediante Software estadístico R Studio

Para finalizar con el análisis clúster, conociendo las medias de cada uno de los grupos podemos concluir los siguientes resultados:

Code R Studio

```
A<-aggregate(datos, by = list(km.res$cluster), FUN = mean)
```

Tabla 3. Medias de cada uno de los clústeres mediante método k-means

Grupo	X15 = Recomendación	X19 = Noches
1	4.411765	1.490196
2	8.272727	2.545455

Fuente: Elaboración propia mediante Software estadístico R Studio



- Grupo 1: Las personas encuestadas que otorgan una recomendación de 4 puntos a la provincia de Tucumán deciden alojarse en promedio 1 noche.
- Grupo 2: Las personas de este grupo, que otorgan una puntuación de 8, deciden alojarse en promedio entre 2 a 3 noches.

Análisis de Correspondencia Múltiple

Para la aplicación del análisis de correspondencia múltiple, se utilizó la base de datos ya organizada y estructurada. Como primera medida, se establecieron diversas relaciones entre las variables cualitativas con el fin de obtener aquella que mejor se adapte para el trabajo bajo estudio.

Code R Studio:

```
library(readxl)
```

```
basa_datos <- read_excel("C:/Users/Vaio/Desktop/NUEVA PP/Libro1.xlsx")
```

```
View(basa_datos)
```

Tabla 3. Extracción de base de datos depurada para ACM

Edad	Comparación oferta turística
Adolescente	Inferior a otras provincias
Adolescente	Inferior a otras provincias
Adolescente	No sabe/No contesta
Adulto Joven	Idéntica a otras provincias
Adulto mediana edad	Inferior a otras provincias
Adolescente	Inferior a otras provincias
Adulto Mayor	Idéntica a otras provincias

Fuente: Elaboración propia



La variable “Edad” se refiere a los distintos grupos etarios de los encuestados, agrupado en un total de cinco categorías, entre ellas: “Menor”, “Adolescente”, “Adolescente Joven”, “Adulto mediana edad” y “Adulto Mayor”.

En relación a la variable “Comparación de ofertar turística”, se agrupó en cuatro categorías: “Inferior a otras provincias”, “idéntica a otras provincias”, “Superior a otras provincias” y “No sabe/No contesta”.

Una vez seleccionada la base de datos con sus respectivas variables, se procedió al análisis descripto de cada una de las variables con sus correspondientes categorías.

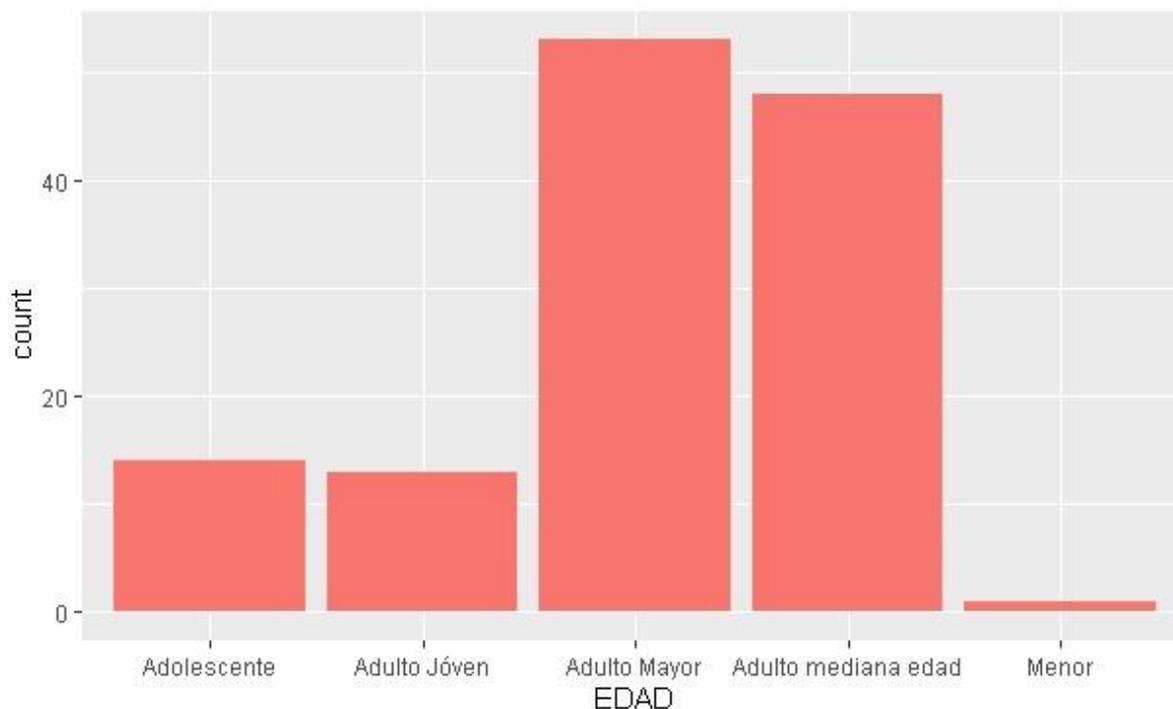
Podemos observar en el siguiente gráfico de “Edad”, que la categoría con mayor influencia es “Adultos Mayores”.

Code R Studio:

```
ggplot(data = base_datos )+ geom_bar(aes(x = EDAD,  
fill = "Red"))+  
theme(legend.position = "none")
```



Gráfico 8. Observaciones por categorías en Edad.



Fuente: Elaboración propia mediante software estadístico R Studio

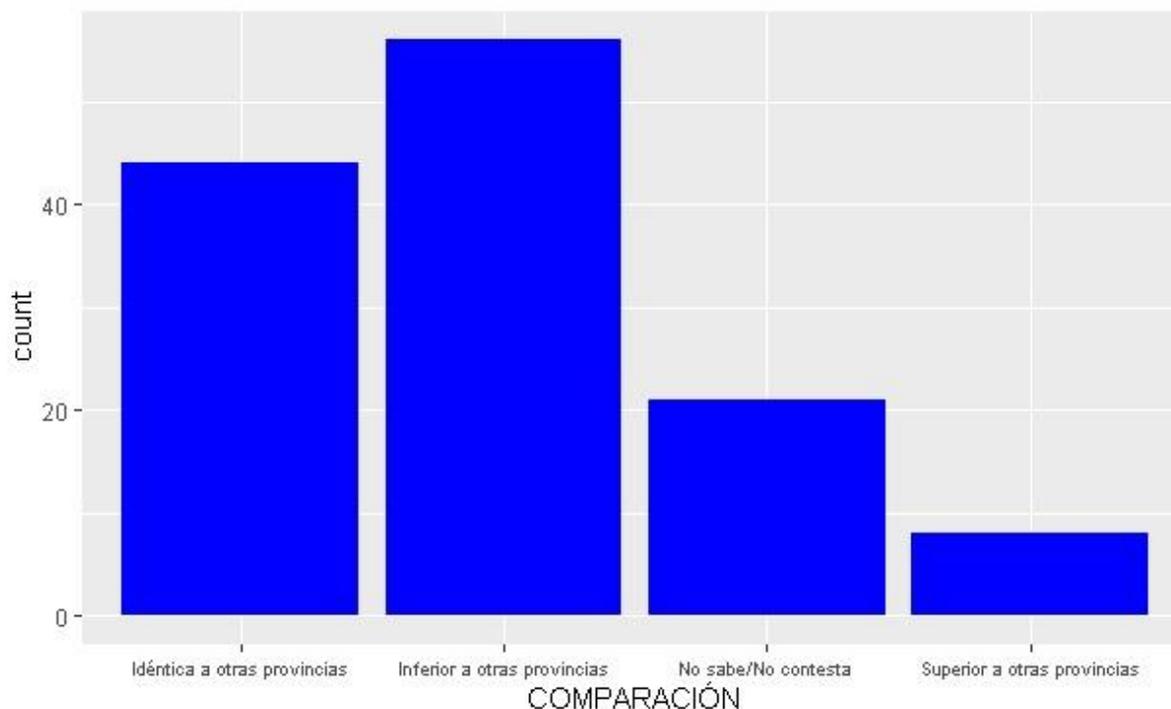
Observamos a continuación el gráfico de “Comparación de oferta turística” los encuestados hicieron hincapié en que la oferta turística de Tucumán es “Inferior a otras provincias”.

Code R Studio:

```
ggplot(data = base_datos )+  
geom_bar(aes(x = COMPARACIÓN), fill = "blue")+  
theme(legend.position = "none", axis.text.x = element_text(size = 7))
```



Gráfico 9. Observaciones por categoría Comparación de oferta turística



Fuente: Elaboración propia mediante Software estadístico R Studio

Continuando con el análisis, se llevó a cabo el cálculo del porcentaje de varianza explicado para cada una de las variables propuestas.

Code R Studio:

```
res.mca <- MCA(base_datos, graph = FALSE)
```

```
print(res.mca)
```

```
print(res.mca$var)
```

Tabla 4. Porcentaje de varianza explicada

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
EDAD	0.6816321	0.5457789	0.5287711	1.000000e+00	0.4712289
COMPARACIÓN	0.6816321	0.5457789	0.5287711	1.020018e-27	0.4712289

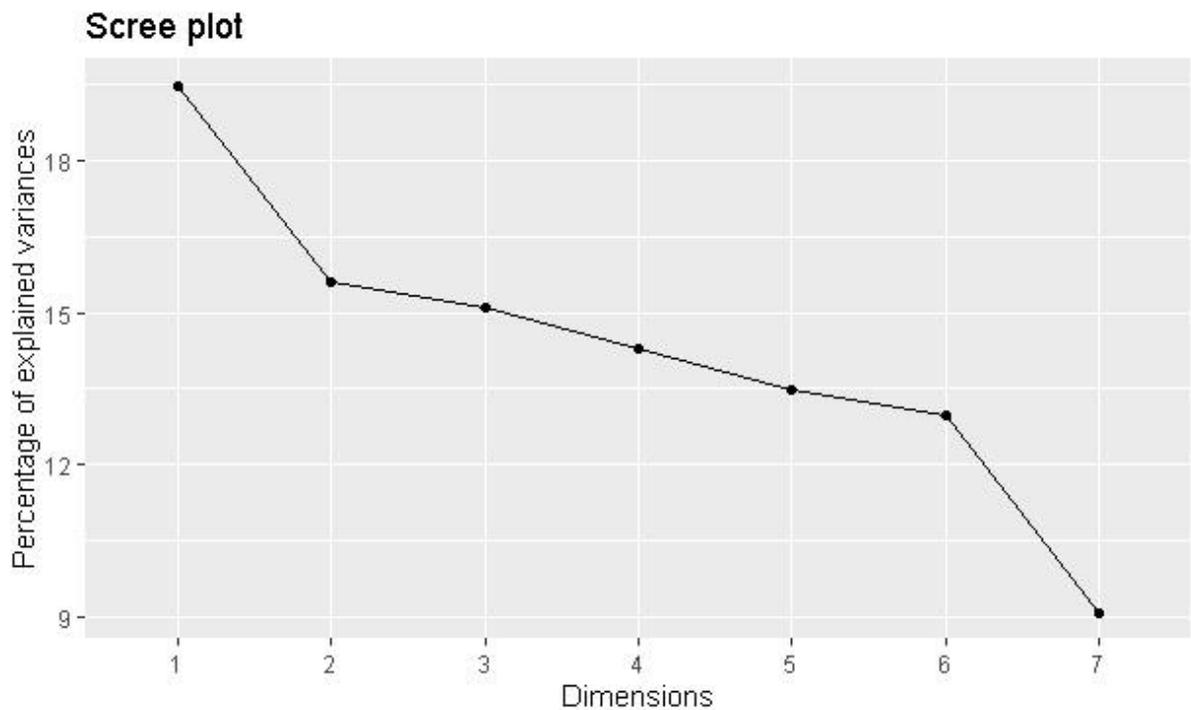
Fuente: Elaboración propia mediante Software estadístico R Studio



Code R Studio:

```
fviz_screplot(res.mca,geom="line")+  
theme_grey()
```

Gráfico 10. Porcentaje de varianza explicada según dimensiones



Fuente: Elaboración propia mediante Software estadístico R Studio

En base a los datos proporcionados gráficamente, podemos determinar que mayor porcentaje de varianza explicada lo encontramos en la dimensión 1 y 2 respectivamente.

Para reforzar el análisis se realizó dos gráficos que visualizan la contribución de las categorías de cada una de las variables para sus respectivas dimensiones.

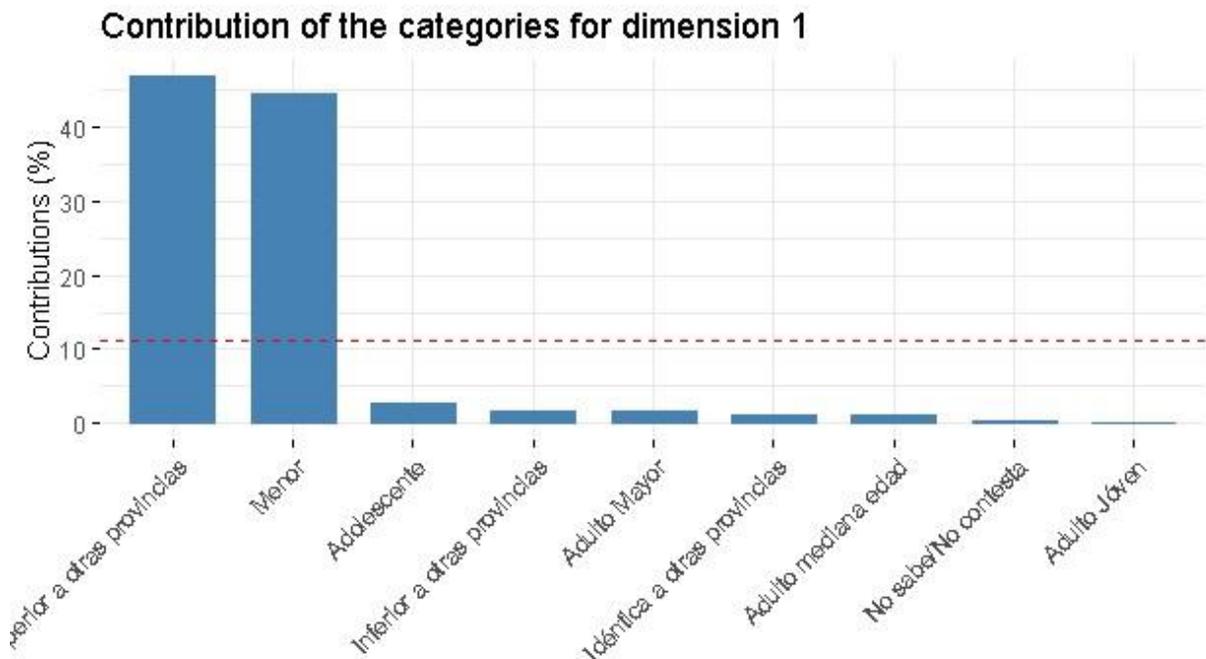


En el gráfico 11, podemos visualizar que la mayor contribución la encontramos en las categorías “Inferior a otras provincias” y “Menor”. Hay que tener en cuenta que la línea punteada de color rojo el valor promedio de contribuciones, si las mismas fuesen uniformes.

Code R Studio:

```
fviz_contrib(res.mca, choice = "var", axes = 1, top = 15)+labs(title = "Contribution of the categories for dimension 1")
```

Gráfico 11. Contribución para categorías en la dimensión 1



Fuente: Elaboración propia mediante Software estadístico R Studio

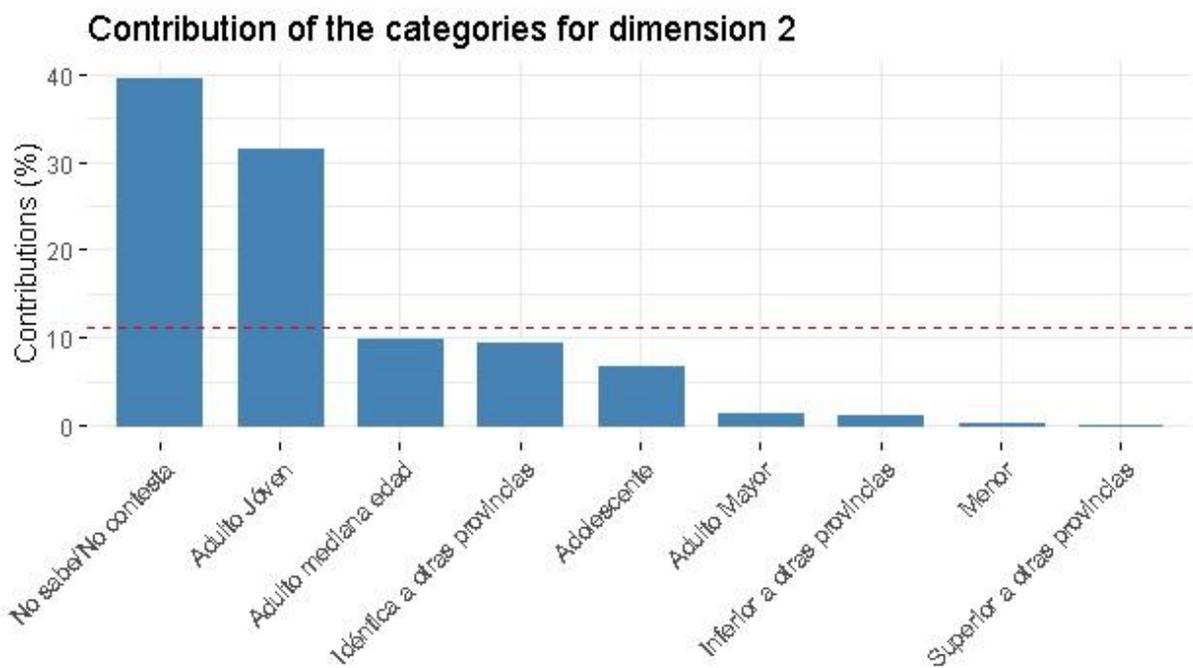
En el gráfico 12 podemos determinar que la mayor contribución la encontramos en las categorías “No sabe/No contesta” y “Adulto Joven”.



Code R Studio:

```
fviz_contrib(res.mca, choice = "var", axes = 2, top = 15)+labs(title = "Contribution of the categories for dimension 2")
```

Gráfico 12. Contribuciones para categorías en la dimensión 2



Fuente: Elaboración propia mediante Software R Studio

En el gráfico 13, se visualiza la representación de la contribución para ambos ejes para de esta manera determinar los patrones de asociación entre las categorías en un espacio menor de dimensión.

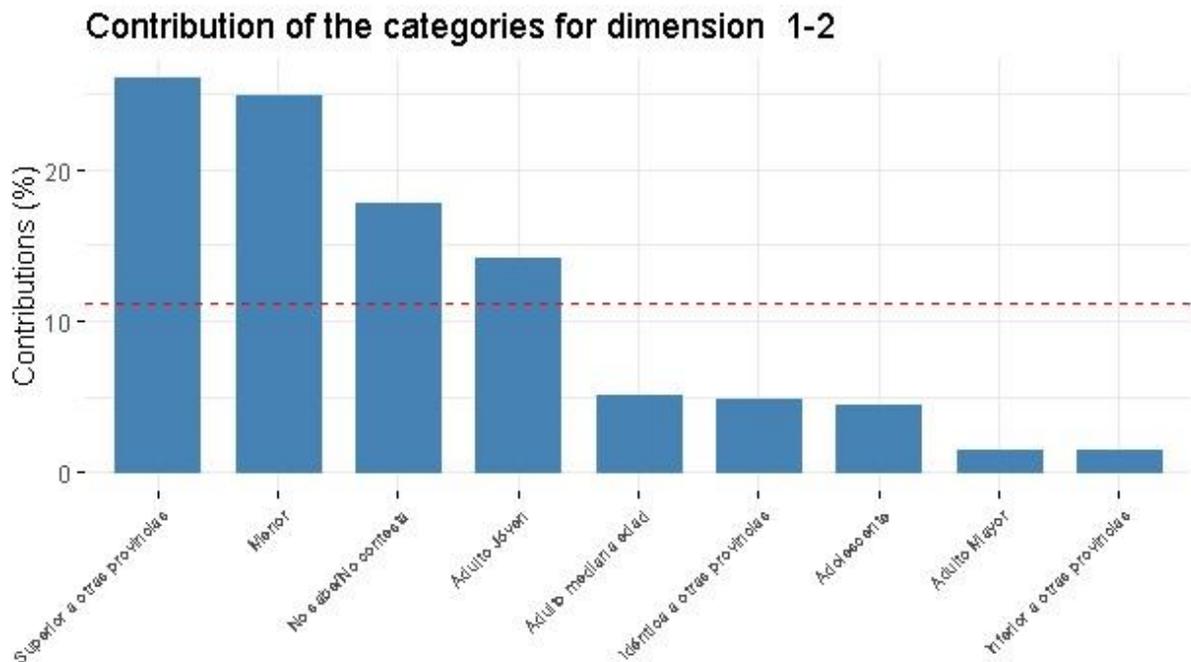
Code R Studio:

```
fviz_contrib(res.mca, choice = "var", axes = 1:2, top = 15)+labs(title = "
```



Contribution of the categories for dimension 1-2")+ theme(axis.text.x = element_text(size = 7))

Gráfico 13 . Contribuciones para categorías en ambas dimensiones 1 - 2



Fuente: Elaboración propia mediante Software estadístico R Studio

Para ir finalizando con el análisis de correspondencia, se realiza un gráfico que visualiza las correspondencias entre las categorías de cada una de las variables, además de un gráfico de co2 para determinar la calidad de que cada una de las representaciones.

Code R Studio

```
fviz_mca_var(res.mca, col.var = "Red", addEllipses = FALSE, repel = TRUE) +  
theme_minimal()
```



Gráfico 14. Contribuciones



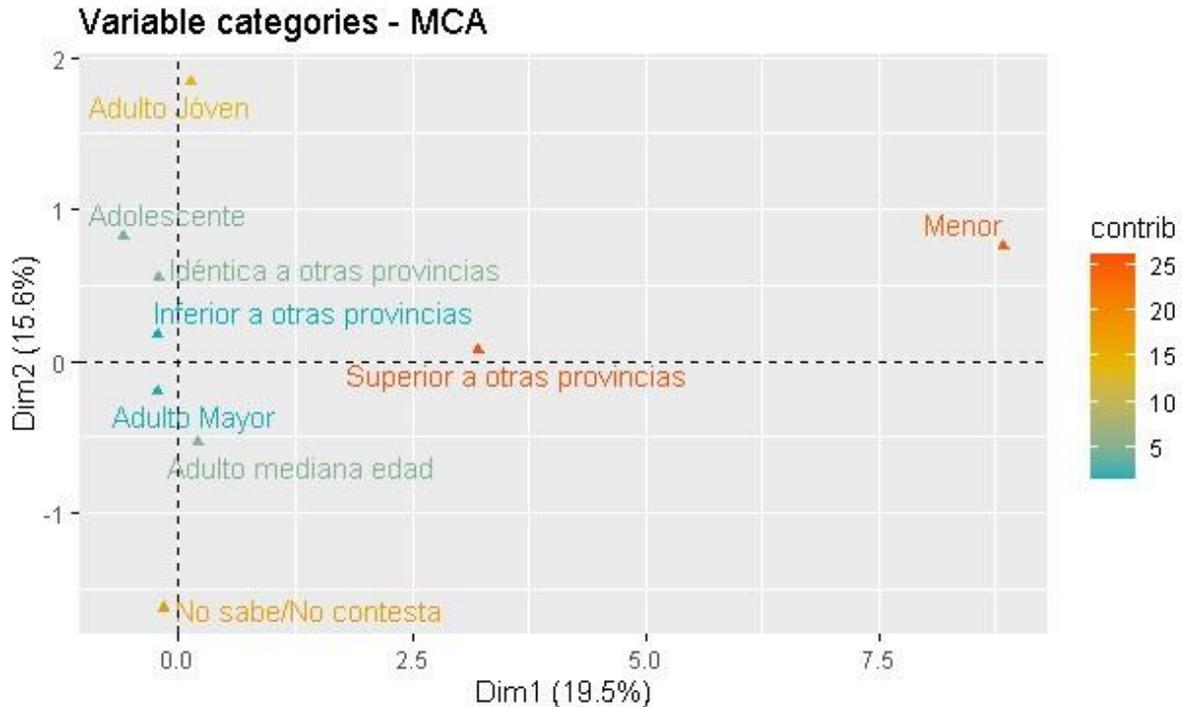
Fuente: Elaboración propia mediante Software estadístico R Studio.

Code R Studio:

```
fviz_mca_var(res.mca, col.var = "contrib",  
gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),  
ggtheme = theme_grey(), repel = TRUE)
```



Gráfico 15. Contribuciones según co2



Fuente: Elaboración propia mediante Software estadístico R

Relaciones observadas en los gráficos propuestos:

- Se observa una alta contribución por parte de las personas de edades “Menores a 18 años” y las que indicaron que la provincia de Tucumán tiene una oferta “Superior a la de otras provincias”. Sin embargo, no podemos establecer una asociación clara debido a su lejanía.
- Por otra parte, la asociación de “Adulto de mediana edad” y “No sabe/No contesta”, la cual establece que las personas comprendidas entre esas edades no tienen conocimiento sobre las ofertas turísticas de provincia de Tucumán en comparación a otras provincias.



- Por último, se puede establecer una asociación entre las personas de “Adulto Mayor” y la categoría “Inferior a otras provincias”. Pero hay que tener en cuenta que la contribución establecida por CO₂ es inferior a 10.

Otros análisis

Por último, con el fin de enriquecer los análisis y resaltar aquellos seis aspectos positivos e importantes que las personas encuestadas no residentes destacaron de la provincia de Tucumán, se estableció un ranking de importancia en base a la frecuencia calculada para cada una de las categorías comprendidas en la variable “aspectos positivos”. Se considera 1 el más importante hasta 6 el menos importante.

Tabla 6. Ranking de aspectos positivos

Aspecto positivo	Ranking
Recursos naturales y paisajes	1
Historia y patrimonio cultural	2
Gastronomía	3
Infraestructura turística	4
Oferta artística y cultural	5
Accesibilidad y distancia	6

Fuente: Elaboración propia



Recomendaciones

En base a los resultados obtenidos, tener en cuenta.

- Teniendo en cuenta los resultados obtenidos a partir del análisis clúster en el estudio de las variables “recomendación” y “noches de alojamiento”, poner énfasis en la infraestructura hotelera de la provincia de Tucumán, evaluando la capacidad y calidad de los servicios hoteleros y demás lugares de alojamiento ofrecidos.
- Según la información obtenida del estudio de la variable “edad” en el ACM, se recomienda que para aquellas personas de rangos etarios que consideraron que la oferta turística de la provincia de Tucumán es inferior a otras provincias, desarrollar campañas de marketing dirigidas específicamente a este grupo demográfico, resaltando actividades y recursos que la provincia pueda ofrecer.
- Se recomienda continuar con las investigaciones y recopilación de información sobre el mercado turístico de Tucumán, ya que el comportamiento y las preferencias de los turistas pueden cambiar con el tiempo, por lo que es crucial mantenerse actualizado con el fin de poder aplicar las estrategias correspondientes.
- Por último, tener en consideración el ranking elaborado con aspectos positivos que los encuestados destacan en la provincia de Tucumán.



Conclusiones

Para concluir con el trabajo de investigación, resulta oportuno enfatizar en la aplicación de herramientas de análisis de datos para ahondar en las relaciones de variables existentes y obtener información de calidad, ya que las empresas atraviesan una complicación debido a la multiplicidad de variables que se les presenta.

Destacar que la utilización de Análisis de Correspondencia Múltiple y Análisis Clúster en combinación, nos permitió tener una perspectiva integral y enriquecedora sobre los hallazgos obtenidos. De esta forma consideramos a estas técnicas como fundamentales para una investigación analítica.

Por otro lado, la aplicación del Software estadístico R Studio para el correcto procesamiento y visualización de los datos, con el fin de ahorrar tiempos y evitar inconsistencias en el análisis manual.

Por último, en base a la investigación realizada se a podido determinar las relaciones existentes entre los rangos etarios y su determinación en cuanto al comparativo de la oferta turística de Tucumán, en base a su pensamiento crítico. Y por medio de clúster, encontrar el tiempo óptimo de noches de alojamiento y puntuación de la provincia.



Bibliografía

- ❖ Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2018). Metodología de la Investigación (6ta ed.). McGraw-Hill Education.
- ❖ Hernández Sampieri, R. (2018). Metodología de la investigación las rutas cuantitativa, cualitativa y mixta. México. Editorial Mc Graw Hill Education.
- ❖ Osterwalder, A., Pigneur, Y., Smith, A., & Bernarda, G. (2015). Diseñando la propuesta de valor. Bogotá, Colombia: Editorial Planeta Colombiana.
- ❖ Kaplan, R. S., & Norton, D. P. (2000). The balanced scorecard: Translating strategy into action: Second edition. Boston, MA: Harvard Business School Press.
- ❖ Pardinas, Felipe: (1991). Metodología y Técnicas de Investigación en Ciencias Sociales. 32a. Edición. Editorial Siglo XXI, Bogotá.
- ❖ DJ Patil y Jeff Hammerbacher. (2011). The Art of Data Science. First edition. O'Reilley Media.
- ❖ Kurzweil, R. (2005). La singularidad está cerca: Cuando los humanos transcendamos la biología. Berlín: Lola Books. Ballvé A. (2008).
- ❖ Peña, Daniel (2002): Análisis de Datos Multivariantes. México DF, México: Mc Graw-Hil.
- ❖ Aldás Joaquín, Uriel Ezequiel (2017). Análisis Multivariante aplicado con R. 2ª Edición. Editorial Paraninfo, S.A.



Anexo

1. ¿Cuál es su edad?
 - a. Menos de 18 años
 - b. 18 a 24 años
 - c. 25 a 34 años
 - d. 35 a 44 años
 - e. 45 a 54 años
 - f. 55 años o más
2. ¿En qué provincia reside actualmente?
 - a. Respuesta abierta
3. ¿Ha visitado alguna vez la provincia de Tucumán?
 - a. Si
 - b. No
4. ¿Cuál es o sería el propósito principal de su visita a la provincia de Tucumán?
(Puede seleccionar más de una opción)
 - a. Explorar la naturaleza y los paisajes
 - b. Conocer la gastronomía local
 - c. Participar de evento o recitals
 - d. Participar en actividades deportivas o de Aventura
 - e. Visitar familiares o amigos
 - f. Otro (respuesta abierta)



5. ¿Qué atracción turística ha visitado en la provincia de Tucumán? (Puede seleccionar más de una opción) Contestar UNICAMENTE en caso de haber visitado la provincia.
- a. Casa Histórica
 - b. Dique el Cadillal
 - c. Festival del limón
 - d. Fiesta Nacional de la Pachamama
 - e. Amaicha del Valle
 - f. Otro (respuesta abierta)
6. ¿Cómo calificaría la calidad de los servicios turísticos en la provincia de Tucumán? Contestar UNICAMENTE en caso de haber visitado la provincia.
- [Hotelería]
- a. Malo
 - b. Regular
 - c. Bueno
 - d. Excelente
 - e. No sabe/No contesta
7. ¿Cómo calificaría la calidad de los servicios turísticos en la provincia de Tucumán? Contestar UNICAMENTE en caso de haber visitado la provincia.
- [Gastronomía].
- a. Malo
 - b. Regular



- c. Bueno
 - d. Excelente
 - e. No sabe/No contesta
8. ¿Cómo calificaría la calidad de los servicios turísticos en la provincia de Tucumán? Contestar UNICAMENTE en caso de haber visitado la provincia.
[Turismo Cultural].
- a. Malo
 - b. Regular
 - c. Bueno
 - d. Excelente
 - e. No sabe/No contesta
9. ¿Cómo calificaría la calidad de los servicios turísticos en la provincia de Tucumán? Contestar UNICAMENTE en caso de haber visitado la provincia.
[Turismo Aventura].
- a. Malo
 - b. Regular
 - c. Bueno
 - d. Excelente
 - e. No sabe/No contesta
10. ¿Cómo calificaría la calidad de los servicios turísticos en la provincia de Tucumán? Contestar UNICAMENTE en caso de haber visitado la provincia.
[Turismo Religioso].



- a. Malo
- b. Regular
- c. Bueno
- d. Excelente
- e. No sabe/No contesta

11. ¿Cómo calificaría la calidad de los servicios turísticos en la provincia de Tucumán? Contestar UNICAMENTE en caso de haber visitado la provincia.
[Turismo de Eventos].

- a. Malo
- b. Regular
- c. Bueno
- d. Excelente
- e. No sabe/No contesta

12. ¿Cómo calificaría la calidad de los servicios turísticos en la provincia de Tucumán? Contestar UNICAMENTE en caso de haber visitado la provincia.
[Servicios públicos pasivos].

- a. Malo
- b. Regular
- c. Bueno
- d. Excelente
- e. No sabe/No contesta



13. ¿Cómo calificaría la oferta turística de la provincia de Tucumán en comparación con otras provincias del NOA (Jujuy, Salta, Tucumán, Santiago del Estero, Catamarca y La Rioja)? Contestar UNICAMENTE en caso de haber visitado la provincia.
- Inferior a otras provincias
 - Idéntica a otras provincias
 - Superior a otras provincias
 - No sabe/No contesta
14. ¿Cómo descubrieron los lugares turísticos de la provincia de Tucumán?
- Respuesta abierta
15. En una escala de 0 al 10 ¿Qué tan probable es que recomiende la provincia de Tucumán como destino turístico? Siendo 0 poco probable y 10 muy probable.
- 0
 - 1
 - 2
 - 3
 - 4
 - 5
 - 6
 - 7
 - 8
 - 9



k. 10

16. ¿Volvería a Tucumán por turismo? ¿Cuáles son las razones que lo llevan a tomar esa decisión? Contestar UNICAMENTE en caso de haber visitado la provincia.

a. Respuesta abierta

17. ¿Cuáles consideras que son los aspectos positivos que más se destacan en la oferta turística de la provincia de Tucumán? Contestar UNICAMENTE en caso de haber visitado la provincia.

a. Respuesta abierta

18. ¿Cuáles consideras que son los aspectos negativos que más se destacan en la oferta turística de la provincia de Tucumán? Contestar UNICAMENTE en caso de haber visitado la provincia.

a. Respuesta abierta

19. Por último, si tuvieras que organizar un viaje al Noroeste Argentino, ¿qué cantidad de noches te alojarías en cada una de las provincias? [Tucumán]

a. 0 noches

b. 1 noche

c. 2 noches

d. 3 noches

e. 4 noches

f. 5 noches

g. 6 noches o más



20. Por último, si tuvieras que organizar un viaje al Noroeste Argentino, ¿que cantidad de noches te alojarías en cada una de las provincias? [Salta]
- a. 0 noches
 - b. 1 noche
 - c. 2 noches
 - d. 3 noches
 - e. 4 noches
 - f. 5 noches
 - g. 6 noches o más
21. Por último, si tuvieras que organizar un viaje al Noroeste Argentino, ¿que cantidad de noches te alojarías en cada una de las provincias? [Jujuy]
- a. 0 noches
 - b. 1 noche
 - c. 2 noches
 - d. 3 noches
 - e. 4 noches
 - f. 5 noches
 - g. 6 noches o más
22. Por último, si tuvieras que organizar un viaje al Noroeste Argentino, ¿que cantidad de noches te alojarías en cada una de las provincias? [Catamarca]
- a. 1 noche
 - b. 2 noches



- c. 3 noches
- d. 4 noches
- e. 5 noches
- f. 6 noches o más

23. Por último, si tuvieras que organizar un viaje al Noroeste Argentino, ¿que cantidad de noches te alojarías en cada una de las provincias? [Santiago del Estero]

- a. 1 noche
- b. 2 noches
- c. 3 noches
- d. 4 noches
- e. 5 noches
- f. 6 noches o más