

# Análisis de la infraestructura turística de las principales ciudades del país

PROYECTO FINAL, ESTADÍSTICA MULTIVARIADA

---

**Autores: Edgar Baquero & Enrique Santibáñez**

27 de Mayo de 2021

Centro de Investigación en Matemáticas,  
Maestría en Cómputo Estadístico.



# Contenido

Importancia y Objetivo del proyecto

Metodología

Descripción de las fuentes de información

Oferta turística

Demanda turística

Resultados

Intepretación de resultados

Conclusiones

# Importancia y Objetivo del proyecto

---

# Importancia y Objetivo del proyecto

1. Visualizar la infraestructura ciudades principales de México

# Importancia y Objetivo del proyecto

1. Visualizar la infraestructura ciudades principales de México
2. Usar técnicas multivariadas

# Importancia y Objetivo del proyecto

1. Visualizar la infraestructura ciudades principales de México
2. Usar técnicas multivariadas
3. Escalar los datos factíblemente

# Importancia y Objetivo del proyecto

1. Visualizar la infraestructura ciudades principales de México
2. Usar técnicas multivariadas
3. Escalar los datos factíblemente
4. Analizar la demanda (Llegada de turistas y ocupación)

# Importancia y Objetivo del proyecto

1. Visualizar la infraestructura ciudades principales de México
2. Usar técnicas multivariadas
3. Escalar los datos factíblemente
4. Analizar la demanda (Llegada de turistas y ocupación)
5. Analizar la oferta (Infraestructura turística)



# Importancia y Objetivo del proyecto

1. Visualizar la infraestructura ciudades principales de México
2. Usar técnicas multivariadas
3. Escalar los datos factíblemente
4. Analizar la demanda (Llegada de turistas y ocupación)
5. Analizar la oferta (Infraestructura turística)
6. Definir una medida de similaridad apropiada entre ciudades

# Importancia y Objetivo del proyecto

1. Visualizar la infraestructura ciudades principales de México
2. Usar técnicas multivariadas
3. Escalar los datos factíblemente
4. Analizar la demanda (Llegada de turistas y ocupación)
5. Analizar la oferta (Infraestructura turística)
6. Definir una medida de similaridad apropiada entre ciudades
7. Reconocer grupos de ciudades

# Metodología

---

1. Queremos una representación en  $q \leq N - 1$  dimensiones

# Metodología, MDS

1. Queremos una representación en  $q \leq N - 1$  dimensiones
2. Hay  $M = \frac{N(N-1)}{2}$  similaridades

# Metodología, MDS

1. Queremos una representación en  $q \leq N - 1$  dimensiones
2. Hay  $M = \frac{N(N-1)}{2}$  similaridades
3. Suponiendo un orden:

$$s_{i_1 k_1} < s_{i_2 k_2} < \dots < s_{i_M k_M}$$

# Metodología, MDS

1. Queremos una representación en  $q \leq N - 1$  dimensiones
2. Hay  $M = \frac{N(N-1)}{2}$  similaridades
3. Suponiendo un orden:

$$s_{i_1 k_1} < s_{i_2 k_2} < \dots < s_{i_M k_M}$$

4. En la nueva representación queremos preservar:

$$d_{i_1 k_1}^{(q)} > d_{i_2 k_2}^{(q)} > \dots > d_{i_M k_M}^{(q)}$$

# Metodología, MDS

1. Queremos una representación en  $q \leq N - 1$  dimensiones
2. Hay  $M = \frac{N(N-1)}{2}$  similaridades
3. Suponiendo un orden:

$$s_{i_1 k_1} < s_{i_2 k_2} < \dots < s_{i_M k_M}$$

4. En la nueva representación queremos preservar:

$$d_{i_1 k_1}^{(q)} > d_{i_2 k_2}^{(q)} > \dots > d_{i_M k_M}^{(q)}$$

5. Evaluaremos a través de la medida de Stress (Kruskal, 1964):

$$\text{Stress}(q) = \left\{ \frac{\sum_{i < k} \left( d_{ik}^{(q)} - \hat{d}_{ik}^{(q)} \right)^2}{\sum_{i < k} \left[ d_{ik}^{(q)} \right]^2} \right\}^{1/2}$$



La evaluación (según Kruskal) del Stress se puede interpretar a través:

Stress	Ajuste
20 %	pobre
10 %	normal
5 %	bueno
2.5 %	Excelente
0 %	Perfecto

Usamos la solución por Coordenadas principales.

# Metodología, Selección de similaridad

Presentamos 3 tipos de similaridad calculables para los datos originales:

## 1. Similaridad Euclideana:

$$s_{ij} = \sqrt{\sum_{k=1}^N (c_{ik} - c_{jk})^2}$$

# Metodología, Selección de similaridad

Presentamos 3 tipos de similaridad calculables para los datos originales:

1. Similaridad Euclideana:

$$s_{ij} = \sqrt{\sum_{k=1}^N (c_{ik} - c_{jk})^2}$$

2. Similaridad Coseno:

$$s_{ij} = \cos \theta_{ij} = \frac{\vec{c}_i \cdot \vec{c}_j}{\|\vec{c}_i\| \|\vec{c}_j\|} = \frac{\sum_1^n c_{ik} c_{jk}}{\sqrt{\sum_1^n c_{ik}^2} \sqrt{\sum_1^n c_{jk}^2}}$$

# Metodología, Selección de similaridad

Presentamos 3 tipos de similaridad calculables para los datos originales:

1. Similaridad Euclideana:

$$s_{ij} = \sqrt{\sum_{k=1}^N (c_{ik} - c_{jk})^2}$$

2. Similaridad Coseno:

$$s_{ij} = \cos \theta_{ij} = \frac{\vec{c}_i \cdot \vec{c}_j}{\|\vec{c}_i\| \|\vec{c}_j\|} = \frac{\sum_1^n c_{ik} c_{jk}}{\sqrt{\sum_1^n c_{ik}^2} \sqrt{\sum_1^n c_{jk}^2}}$$

3. Similaridad de *Coxon* (Borg & Groenen, 2005):

$$s_{ij} = \sqrt{2(1 - r_{ij})} \quad (2.1)$$

Dado que hay  $N = 37$  ciudades. parece conveniente representarlas en jerarquías

1. Datos originales  $\implies$  Clústering Jerárquico

Dado que hay  $N = 37$  ciudades. parece conveniente representarlas en jerarquías

1. Datos originales  $\implies$  Clústering Jerárquico
2. Matriz de distancias  $\implies$  Clústering por  $K$ -Means

**Result:** conjunto de centroides

**while** *La asignación de clústeres cambie* **do**

Para alguna asignación de clúster,  $C$ , la varianza es minimizada respecto a  $\{m_1, \dots, m_K\}$ , llevando a la media del cluster asignado;

Dado el conjunto  $\{m_1, \dots, m_K\}$ , minimizamos

$\min_{C, \{m_k\}_1^K} \sum_{k=1}^K N_k \sum_{C(i)=k} \|x_i - m_k\|^2$ , asignando cada observación a su cluster más cercano::

$$C(i) = \operatorname{argmin}_{1 \leq k \leq K} \|x_i - m_k\|^2$$

**end**

**Algorithm 1:**  $K$ -means Clústering

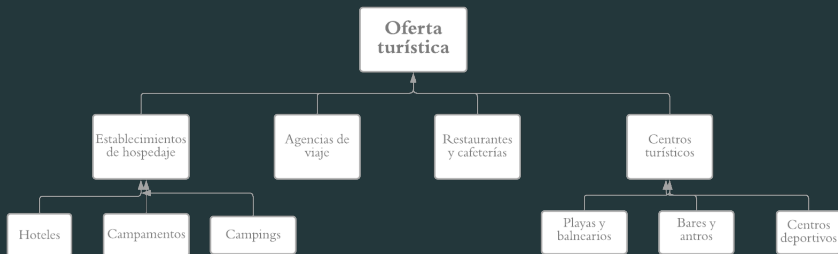
## Descripción de las fuentes de información

---



# Estructura de la oferta turística

Consideramos que la oferta turística de una ciudad se define por los servicios que cuenta cada ciudad para generar atracciones o comodidad a los turistas.



**Figura 1:** Estructura de la oferta turística

# Información de la oferta turística

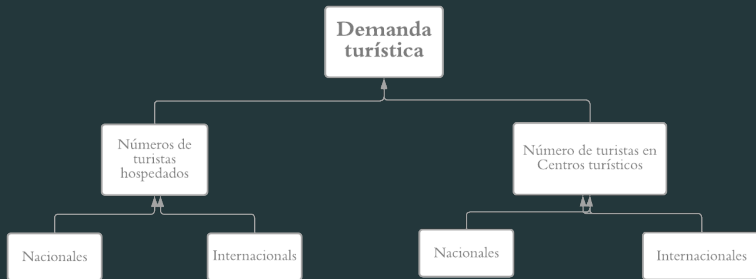
Los datos se han obtenido del Directorio Estadístico Nacional de Unidades Económicas 2019 (DE-NUE) que realiza el Instituto Nacional de Estadística y Geografía (INEGI) (INEGI, s.f.-a).



Figura 2: Map interactivo DENUE

# Estructura de la demanda turística

Para medir la demanda turística de cada ciudad, lo ideal sería tener el dato real de cuantos turistas existen por ciudad. Pero un *proxy* de este valor es considerar los registros hoteleros y de los centros turísticos.



**Figura 3:** Estructura de la demanda turística

## Información de la demanda turística.

El Monitoreo Hotelero Data-Tur reúne información estadística de las principales variables como flujo de viajeros internacionales, turismo doméstico, flujos aéreos y marítimos, actividades de alojamiento, las cuales en su conjunto ofrecen una perspectiva de la dinámica del sector turismo en México (INEGI, s.f.-b).

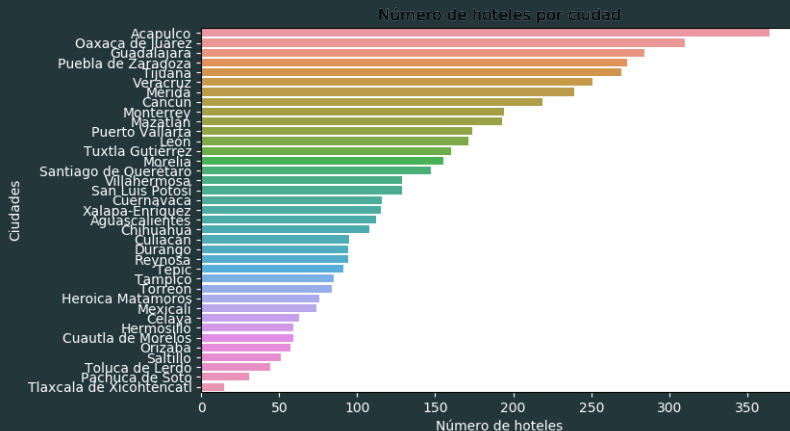


**Figura 4:** Logo Datatur

# Resultados

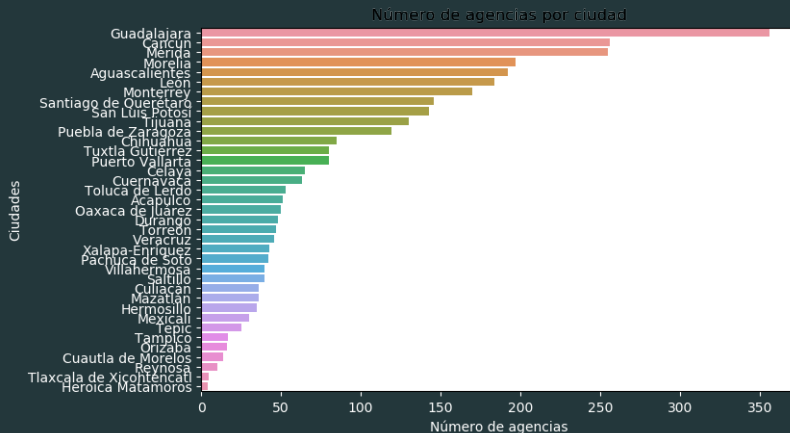
---

# Oferta hotelera



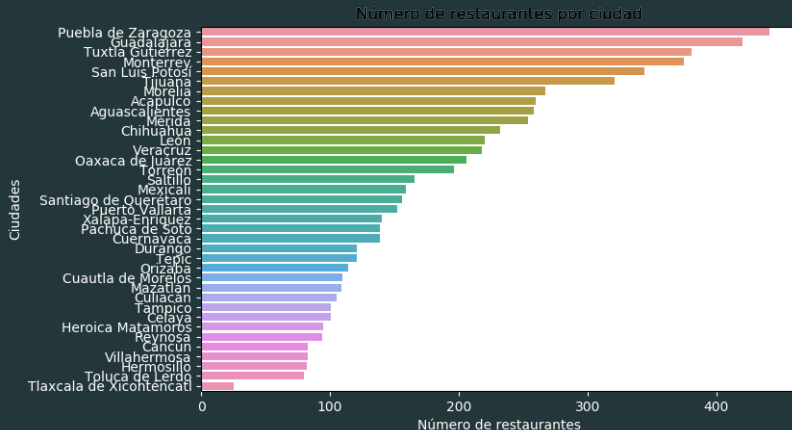
**Figura 5:** Distribución de la cantidad de hoteles por ciudad

# Agencias de viaje



**Figura 6:** Distribución de la cantidad de agencias de viajes por ciudad

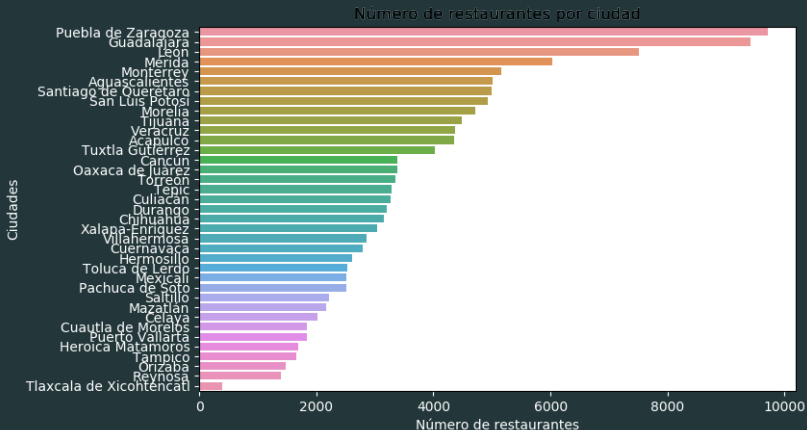
# Restaurantes, cafetería y fondas.



**Figura 7:** Distribución de la cantidad de servicios de comida por ciudad



# Lugares de ocio y recreativos



**Figura 8:** Distribución de la cantidad de lugares de ocio y recreativos por ciudad

# Turistas extranjeros

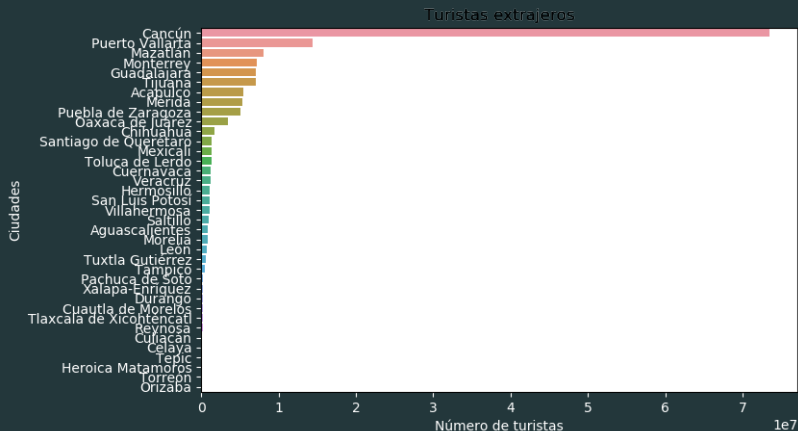


Figura 9: Turistas extranjeros

# Turistas nacionales

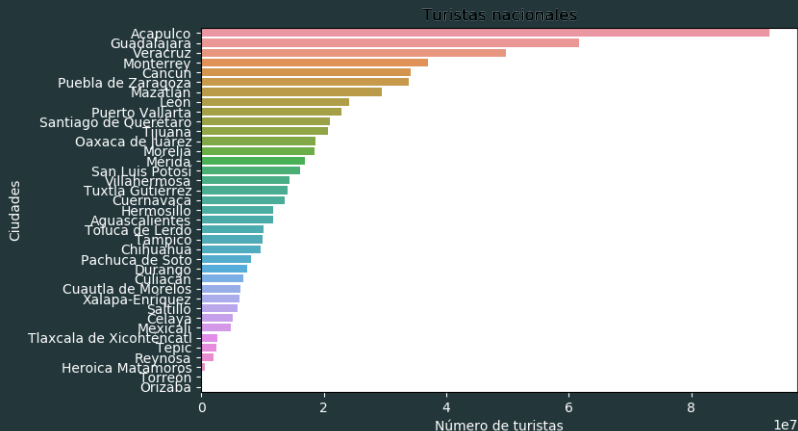


Figura 10: Turistas nacionales

- Para determinar las proximidades que existen entre las ciudades, consideremos la distancia euclidiana como primer enfoque. Pero esta medida de distancia no arrojaba buenos resultados.

- Para determinar las proximidades que existen entre las ciudades, consideremos la distancia euclidiana como primer enfoque. Pero esta medida de distancia no arrojaba buenos resultados.
- En un estudio similar María, 2012, utilizaron una medida de disimilaridad para cada par de observaciones mediante el coeficiente de correlación.

- Para determinar las proximidades que existen entre las ciudades, consideremos la distancia euclidiana como primer enfoque. Pero esta medida de distancia no arrojaba buenos resultados.
- En un estudio similar María, 2012, utilizaron una medida de disimilaridad para cada par de observaciones mediante el coeficiente de correlación.
- Considerando la transformación de Coxon (1) a partir de los datos obtenidos del DENUE hemos aplicado MDS, obteniendo la configuración de la Figura 11.

# Configuración MDS

El STRESS obtenido es de 0.1034, lo cual nos indica que el ajuste de los datos es regular.



Figura 11: Configuración MDS

- El segundo objetivo de este trabajo es determinar la demanda turística de cada ciudad. Para ello, con ayuda de los datos recabados de Datatur procedimos a realizar el análogo a lo que se hizo en oferta turística.



- El segundo objetivo de este trabajo es determinar la demanda turística de cada ciudad. Para ello, con ayuda de los datos recabados de Datatur procedimos a realizar el análogo a lo que se hizo en oferta turística.
- En este caso, la configuración obtenida con MDS tuvo mejores resultados (Ver Figura 8). Obtuvimos un STRESS de 0.01, lo que nos indica que el ajuste es muy bueno.

# Configuración MDS, demanda turística



Figura 12: Configuración MDS, demanda turística

# Intepretación de resultados

---

## Comparación de clustering y las configuración MDS

La última parte de este trabajo es comparar los métodos de clustering KMeans y Métodos jerárquicos, con la configuración obtenida en MDS. Para ello,

- ocupamos métodos conglomerados a los datos originales de oferta y demanda.
- ocupamos KMeans a la matriz de distancias de la transformación Coxon de la matriz de correlación entre las ciudades.

# Clustering oferta usando KNN

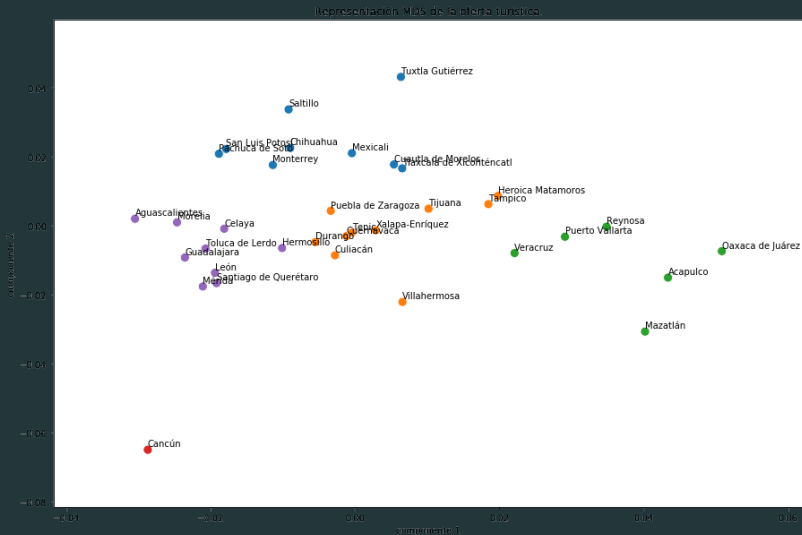


Figura 13: KNN a la matriz de distancias



# Clustering demanda usando KNN



Figura 15: KNN a la matriz de distancias

# Clustering demanda usando conglomerados

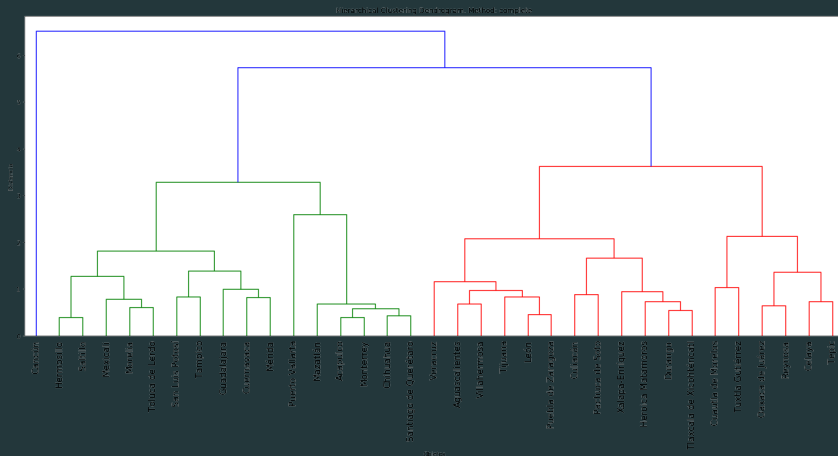


Figura 16: Análisis de Conglomerados a la matriz de datos.



# Conclusiones

---

- En conclusión podemos decir que ambas configuraciones de la demanda y la oferta turística fueron buenas, pero la demanda tuvo mejores resultados poder tener una interpretación de los componentes que los de la oferta.
- Además podemos concluir que las configuraciones obtenidas presentan muy buenos resultados comparados con los métodos de clustering no supervisado.

Gracias 😊

---

## Referencias

---



Borg, I. & Groenen, P. (2005). *Modern Multidimensional Scaling: Theory and Applications*. Springer.



INEGI. (s.f.-a). *Directorio Estadístico Nacional de Unidades Económicas*.

<https://www.inegi.org.mx/app/mapa/denue/default.aspx>



INEGI. (s.f.-b). *Resultados de la Actividad Turística 2019*. [https://datatur.sectur.gob.mx/RAT/RAT-2019-12\(ES\).pdf](https://datatur.sectur.gob.mx/RAT/RAT-2019-12(ES).pdf)



María, G. C. (2012). EL ANÁLISIS DE ESCALAMIENTO MULTIDIMENSIONAL: UNA ALTERNATIVA Y UN COMPLEMENTO A OTRAS TÉCNICAS MULTIVARIANTES.. *La Sociología en sus Escenarios*, (25). <https://revistas.udea.edu.co/index.php/ceo/article/view/11450>