Proyecto Inmobiliario Costa Atlántica Argentina

Marzo, 2023

# Introducción

## Problema

Después de la Pandemia, se ha observado que muchas personas han empezado a moverse de las grandes ciudades a hacia ciudades más pequeñas en búsqueda de sitios más tranquilos para vivir gracias a la posibilidad del trabajo virtual.

## Interesados

Un grupo de inversionistas quieren hacer un estudio en la Costa Atlántica Argentina para determinar la o las ciudades con mayor potencial para desarrollar un proyecto inmobiliario con los servicios de conexión de internet necesarios para el trabajo desde casa.

# Adquisición de Datos y Limpieza

## Fuentes de Datos

Para realizar esta investigación se emplearon datos geográficos del repositorio público de datos de la República Argentina, esta documentación corresponde a la API de normalización de datos geográficos, junto con los datos utilizados por el servicio. Los conjuntos de datos de provincias, departamentos, municipios, localidades y vías de circulación son elaborados por la Dirección Nacional de Datos e Información Pública en base a datos publicados por IGN, BAHRA e INDEC.

Disponible en el siguiente enlace: [https://datos.gob.ar/ar/dataset/jgm-servicio-normalizacion-datosgeograficos/archivo/jgm\_8.12](https://datos.gob.ar/ar/dataset/jgm-servicio-normalizacion-datos-geograficos/archivo/jgm_8.12)

También se empleó un detalle del acceso de Internet y velocidades promedio de Internet fijo por localidad proporcionadas por la ENACOM, ya que se busca localidades con buena infraestructura de servicios de Internet. Datos disponibles por el ente regulador de Telecomunicaciones de Argentina.

Disponible en el siguiente enlace: [https://datosabiertos.enacom.gob.ar/dashboards/20000/acceso-ainternet/](https://datosabiertos.enacom.gob.ar/dashboards/20000/acceso-a-internet/)

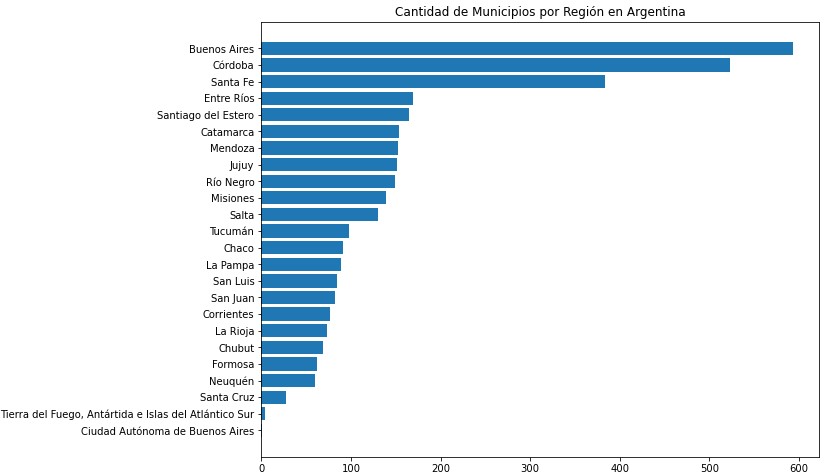
Y por último una lista de las ciudades balnearias más importantes de la Costa Atlántica, ya que son las ciudades más populares y son tentativamente la mejor opción para tener en cuenta en el lugar deseado para invertir. Los datos fueron obtenidos de Wikipedia.

Disponibles en el siguiente enlace:

<https://es.wikipedia.org/wiki/Localidades_balnearias_del_mar_Argentino>

## Limpieza y Exploración de los Datos

Se hizo una exploración de los datos geográficos del repositorio nacional para determinar sus características y la información contenida.



*Figura 1. Distribución de municipios por localidad*

A partir de la lista de las ciudades más populares de la costa, se cruzó con los datos del repositorio geográfico.



Quedando como final un dataset de localidades de las provincias que tienen ciudades con salida a la Costa Atlántica con su información de latitud y longitud. Donde el target es la variable localidad.

El análisis se hace para 34 localidades que abarcan las 5 provincias seleccionadas.

*Figura 2. Provincias con localidades hacia la Costa Atlántica*

# Definición de Variables para el Modelamiento

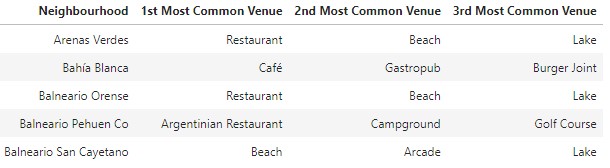
## Variables con información geográfica

Estas variables, corresponden a los sitios más populares en cada localidad con un radio de 500 metros. Para poder determinar las características de cada una y los servicios disponibles desde el punto de vista comercial.

Esta información se obtuvo por medio del servicio de FourSquare. <https://es.foursquare.com/developers/projects>

Donde se obtuvieron 224 lugares distribuidos en 75 categorías únicas para las 34 localidades en estudio.

A su vez se agruparon por los 3 sitios más populares por localidad:



*Figura 3. Ejemplo de la agrupación de lugares populares*

## Variables con información de Acceso a Internet

Otro tipo de información para determinar el potencial de la localidad es el tipo de acceso que tiene la localidad para para acceder al servicio de Internet fijo y móvil. Estos datos fueron proporcionados por la ENACOM que se actualizan de manera anual.



*Figura 4. Variables de acceso a los servicios de Internet* A su vez fueron agrupados en cuatro categorías:

* **Red Móvil:** Accesos de las Red móvil 2G y 3G
* **Red Banda Ancha:** Accesos ADSL, Cablemódem y Fibra
* **Red Banda Angosta:** Satelital, DialUp y Wireless
* **Teléfono Fijo:** Si tiene servicio de telefonía fija

# Normalización de Variables

La normalización es una técnica que se aplica a menudo como parte de la preparación de datos para el aprendizaje automático. El objetivo de la normalización es cambiar los valores de las columnas numéricas del conjunto de datos para usar una escala común, sin distorsionar las diferencias en los intervalos de valores ni perder información. La normalización también es necesaria para que algunos algoritmos modelen los datos correctamente.

Si bien existen varias técnicas para normalizar variables, en esta investigación se usaron variables categóricas que se pasan a Dummies (1 o 0).

*Figura*

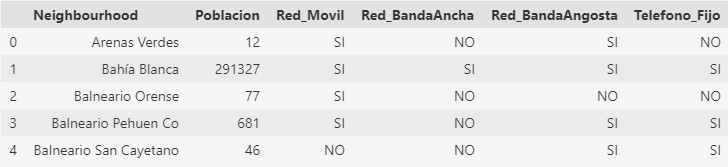
*5*

*. Ejemplo Variables geográficas*

*Figura*

*6*

*. Información descriptiva de las variables de Internet*



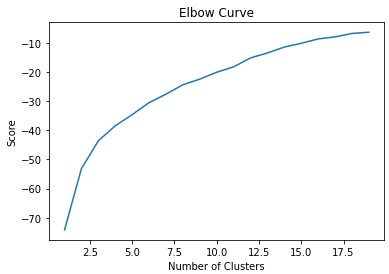
Las variables categóricas son las categorías por tipo de lugar y la información del tipo de conexión a los servicios de Internet.

# Modelamiento Clustering

El método K-means basa su funcionamiento en agrupar los datos de entrada en un total de k grupos definidos por un centroide, cuya distancia con los puntos que pertenecen a cada uno de los datos es la menor posible.

## Curva del Codo

Por medio de este método se va a determinar el valor óptimo de k centroides, en el cual la reducción en la suma total de la varianza deja de ser sustancial. Esto por medio de la optimización de la WCSS (Within Clusters Summed Squares).

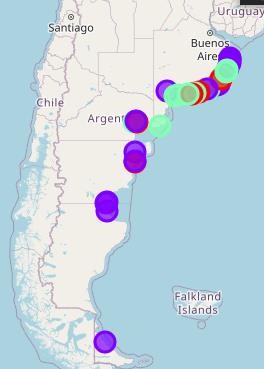


*Figura 7. Curva el Codo*

A partir de la curva, el valor que mantiene una buena relación entre score y cantidad de clusters es K=3. Este será el valor empleado para generar las agrupaciones de las variables.

# Resultados

Después de ejecutar el modelo, se obtuvieron 3 agrupaciones el cual tienen la siguiente distribución geográfica:



Clúster

1

Clúster

3

Clúster

2

*Figura 8. Distribución geográfica de los Cluster*

## Agrupación 1

Este grupo se caracteriza por no tener Red Móvil y el 40% no tiene servicios de teléfono fijo, por tal motivo queda descartado para una posible inversión.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Provincia** | **Localidad** | **Cluster** | **Poblacion** | **Red\_Movil** | **Red\_BandaAncha** | **Red\_BandaAngosta** | **Telefono\_Fijo** | **1st Most**  **Common Venue** | **2nd Most**  **Common Venue** | **3rd Most**  **Common Venue** |
| Buenos  Aires | Santa  Clara del Mar | 0 | 7713 | NO | SI | SI | NO | Art Gallery | New  American Restaurant | Beach |
| Buenos  Aires | Claromecó | 0 | 2081 | NO | SI | SI | SI | Restaurant | Bakery | Beach |
| Río Negro | Puerto  San  Antonio Este | 0 | 381 | NO | SI | SI | SI | Argentinian  Restaurant | Bistro | Restaurant |
| Chubut | Playa  Magagna | 0 | 76 | NO | NO | SI | NO | Beach | Arcade | Lake |
| Buenos Aires | Balneario  San Cayetano | 0 | 46 | NO | NO | SI | SI | Beach | Arcade | Lake |

*Figura 9. Resultados Agrupación 1*

## Agrupación 2

Este grupo se caracteriza por tener las localidades con mayor población, tienen la mayoría de los servicios disponibles para acceder a Internet y gran variedad de lugares populares además de la playa lo que puede indicar mayor actividad comercial.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Provincia** | **Localidad** | **Cluster** | **Poblacion** | **Red\_Movil** | **Red\_BandaAncha** | **Red\_BandaAngosta** | **Telefono\_Fijo** | **1st Most**  **Common**  **Venue** | **2nd Most**  **Common**  **Venue** | **3rd Most**  **Common**  **Venue** |
| Buenos  Aires | Mar del  Plata | 1 | 560913 | SI | SI | SI | SI | Café | Theater | Restaurant |
| Buenos  Aires | Bahía Blanca | 1 | 291327 | SI | SI | SI | SI | Café | Gastropub | Burger  Joint |
| Chubut | Comodoro  Rivadavia | 1 | 175196 | SI | SI | SI | SI | Gift Shop | Supermarket | Food &  Drink Shop |
| Buenos  Aires | Necochea -  Quequén | 1 | 84784 | SI | SI | SI | SI | Café | Italian  Restaurant | Burger  Joint |
| Chubut | Puerto  Madryn | 1 | 81315 | SI | SI | SI | SI | Gym Pool | Arcade | Lottery  Retailer |
| Tierra del  Fuego,  Antártida e Islas del  Atlánt... | Río Grande | 1 | 66475 | SI | SI | SI | SI | Gym | Furniture / Home Store | Gym /  Fitness Center |
| Santa  Cruz | Caleta Olivia | 1 | 51733 | SI | SI | SI | NO | Candy Store | Pub | Park |
| Buenos Aires | Miramar | 1 | 29433 | SI | SI | SI | SI | Café | Breakfast Spot | Italian Restaurant |
| Buenos  Aires | Mar de Ajó -  San Bernardo | 1 | 28466 | SI | SI | SI | SI | Café | Argentinian  Restaurant | Pizza  Place |
| Buenos  Aires | Pinamar | 1 | 25397 | SI | SI | SI | SI | Beach | Resort | Golf  Course |
| Buenos  Aires | Santa  Teresita -  Mar del Tuyú | 1 | 23581 | SI | SI | SI | SI | Nightclub | Coffee Shop | Café |
| Río Negro | San Antonio  Oeste | 1 | 16265 | SI | SI | SI | SI | Gym /  Fitness  Center | Bar | Pharmacy |
| Buenos Aires | San  Clemente del Tuyú | 1 | 12126 | SI | SI | SI | SI | Pizza  Place | Plaza | Bed & Breakfast |
| Chubut | Rada Tilly | 1 | 9098 | SI | SI | SI | SI | Clothing  Store | Arcade | Lottery  Retailer |
| Chubut | Playa Unión | 1 | 6775 | SI | SI | SI | SI | Ice  Cream Shop | Surf Spot | Beach |
| Buenos  Aires | Las Toninas | 1 | 5278 | SI | SI | SI | SI | City | Arcade | Lottery  Retailer |
| Buenos  Aires | Chapadmalal | 1 | 4112 | SI | SI | SI | SI | Resort | Beach | Italian  Restaurant |

*Figura 10. Resultados Agrupación 2*

## Agrupación 3

Esta agrupación se caracteriza por tener localidades mas pequeñas, exceptuando las tres primeras son localidades con servicios básicos de acceso a Internet. Los lugares más populares son la playa, lagos/ríos o restaurantes de comida Argentina.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Provincia** | **Localidad** | **Cluster** | **Poblacion** | **Red\_Movil** | **Red\_BandaAncha** | **Red\_BandaAngosta** | **Telefono\_Fijo** | **1st Most**  **Common Venue** | **2nd Most**  **Common Venue** | **3rd Most**  **Common Venue** |
| Buenos  Aires | Villa  Gesell | 2 | 29593 | SI | SI | NO | SI | German  Restaurant | BBQ Joint | Supermarket |
| Buenos  Aires | Monte  Hermoso | 2 | 6351 | SI | SI | NO | SI | Argentinian  Restaurant | Bakery | Fast Food  Restaurant |
| Río  Negro | Las  Grutas | 2 | 4807 | SI | SI | SI | SI | Beach | Argentinian  Restaurant | Resort |
| Buenos  Aires | Mar Azul | 2 | 1797 | SI | NO | NO | SI | German  Restaurant | Bakery | Seafood  Restaurant |
| Buenos  Aires | Mar de  Cobo | 2 | 760 | NO | NO | SI | SI | Plaza | Market | Arcade |
| Río  Negro | El Cóndor | 2 | 746 | SI | NO | SI | SI | Argentinian  Restaurant | Lake | New  American Restaurant |
| Buenos  Aires | Balneario  Pehuen Co | 2 | 681 | SI | NO | SI | SI | Argentinian  Restaurant | Campground | Golf Course |
| Buenos Aires | Reta | 2 | 495 | SI | NO | SI | SI | Argentinian  Restaurant | Supermarket | Beach |
| Buenos  Aires | Mar  Chiquita | 2 | 487 | SI | NO | SI | SI | Restaurant | Diner | Lake |
| Buenos  Aires | Mar del  Sur | 2 | 453 | SI | NO | SI | SI | Breakfast  Spot | Hotel | Eastern  European Restaurant |
| Buenos Aires | Balneario Orense | 2 | 77 | SI | NO | NO | NO | Restaurant | Beach | Lake |
| Buenos  Aires | Arenas  Verdes | 2 | 12 | SI | NO | SI | NO | Restaurant | Beach | Lake |



Para acceder al análisis completo, visite el enlace para ver la Notebook completa[:](https://github.com/NCLIVIO/Coursera_Capstone/blob/main/Batalla%20de%20los%20Vecindarios/BatallaVecindarios_ProyectoFinal.ipynb) <https://github.com/whereismyminddd/Coursera_Capstone./blob/14ce72f27d202db464b37492220e4c9168a38780/BatallaVecindarios_Semana%20(2).ipynb>