

# Modelo de Regresión para la predicción de la evolución del COVID-19 en la Ciudad de Buenos Aires a partir de factores ambientales y demográficos.

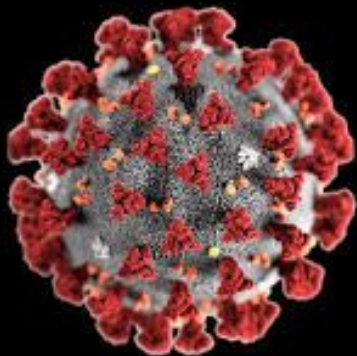
Cluster AI - Universidad Tecnológica Nacional  
Noviembre 2020

Demaestri, J. & Gutierrez, M.

clusterAI

# Introducción

clusterAI



El objetivo del siguiente estudio es el empleo de herramientas de Machine Learning para la predicción de la evolución del COVID-19 a partir de factores climáticos, demográficos y de movilidad urbana.



# Ciudad de Buenos Aires

clusterAI

**Buenos Aires**, oficialmente **Ciudad Autónoma de Buenos Aires (CABA)** o **Ciudad de Buenos Aires**, es la capital y ciudad principal (más poblada y de mayor tamaño) de la República Argentina.

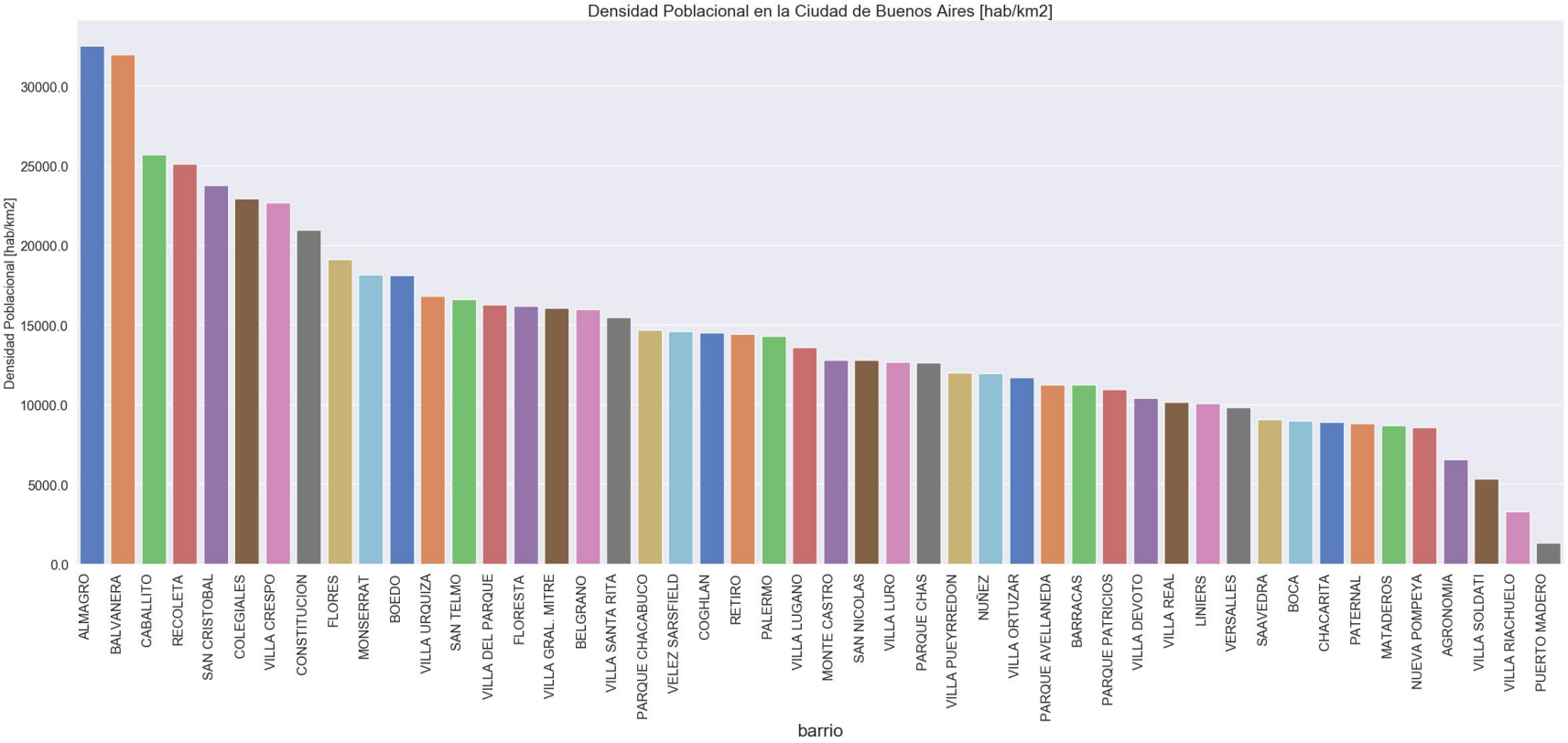
La población estimada de la ciudad es de 3 075 646 habitantes y la del Aglomerado 'Gran Buenos Aires' (conformado por la propia Ciudad Autónoma de Buenos Aires junto a ciertos partidos-municipios de la lindera Provincia de Buenos Aires) es de 17.541.141 habitantes, siendo este último el segundo aglomerado de Hispanoamérica, el segundo de América del Sur y una de las 20 mayores ciudades del mundo.

**Superficie:** 203 km<sup>2</sup>

**Área metropolitana:** 4,758 km<sup>2</sup>

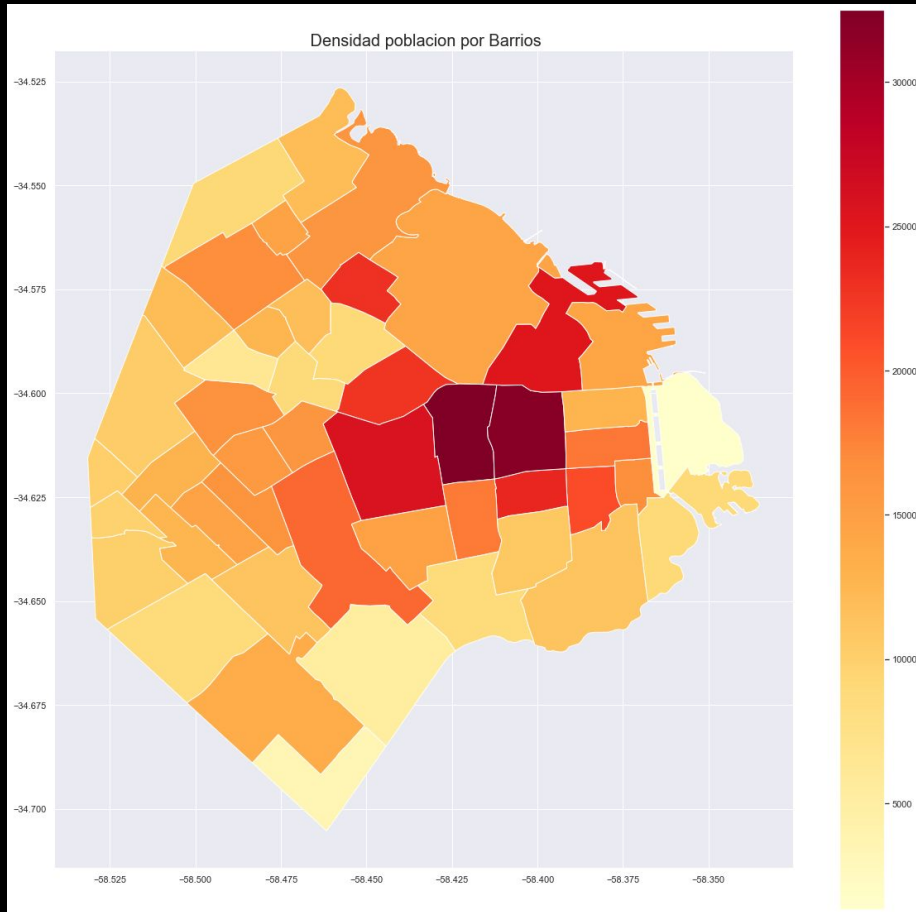
**Coordenadas :** 34°35'59"S 58°22'55"O

# Demografía de la Ciudad de Buenos Aires



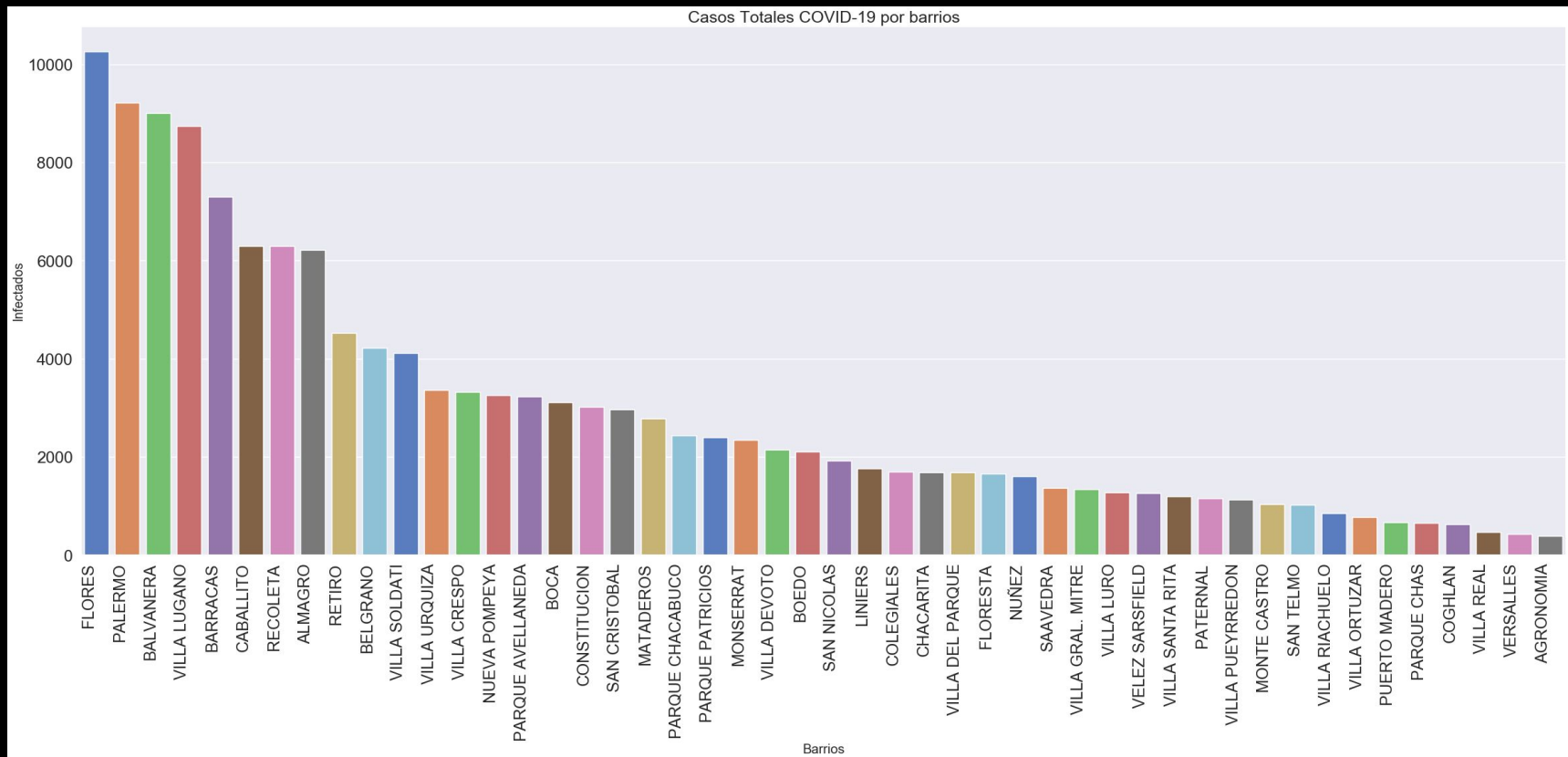
# Demografía de la Ciudad de Buenos Aires

clusterAI



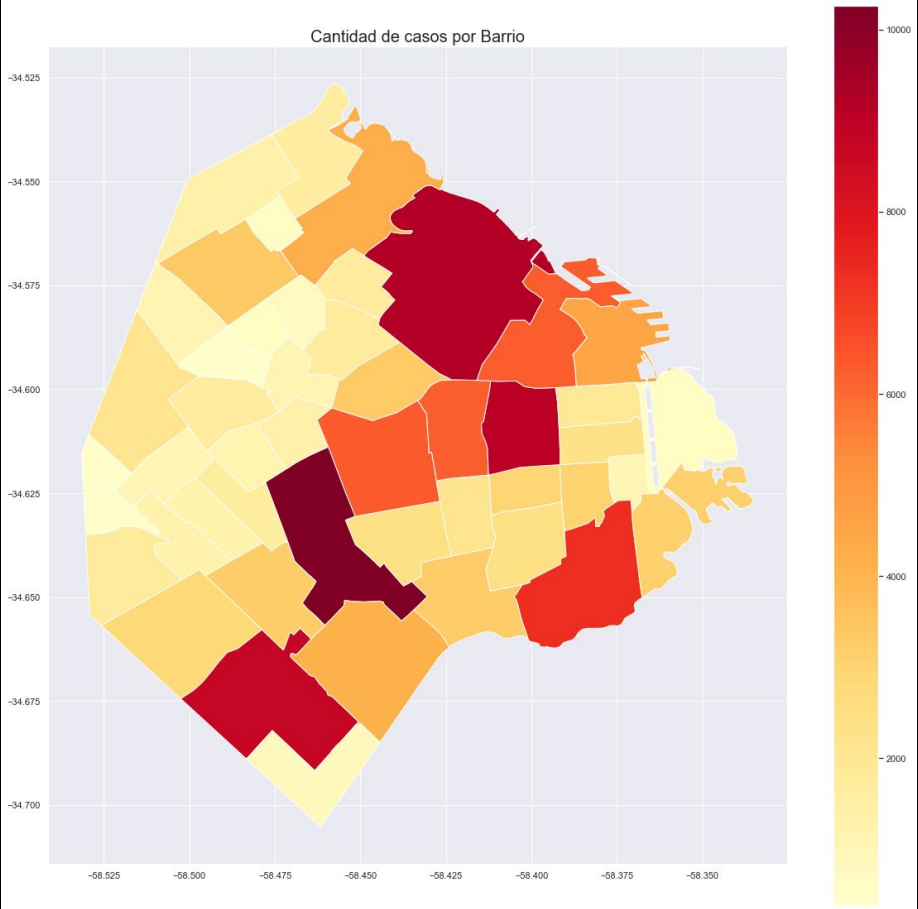
# Casos de COVID-19 según cada Barrio\*

clusterAI



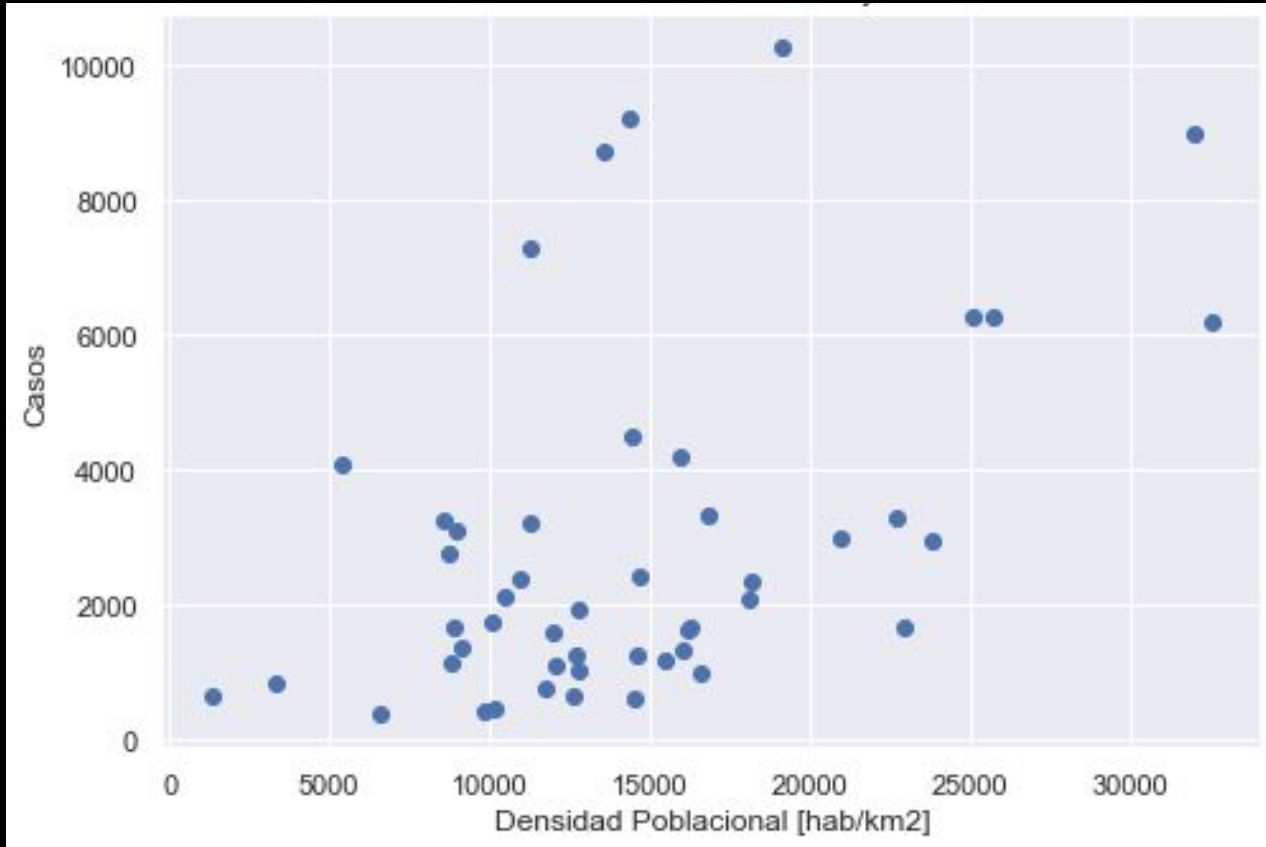
\*Información actualizada: Octubre 2020

# Casos de COVID-19 según cada Barrio



# Relación entre Densidad y Casos

clusterAI

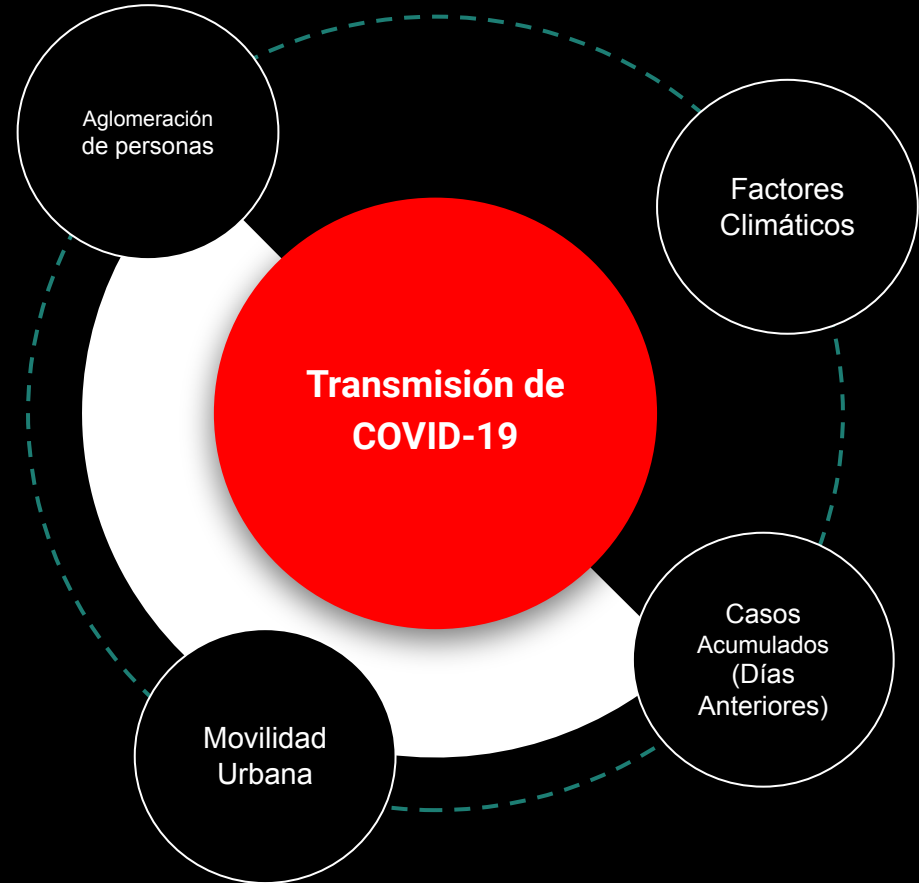




# Hipótesis

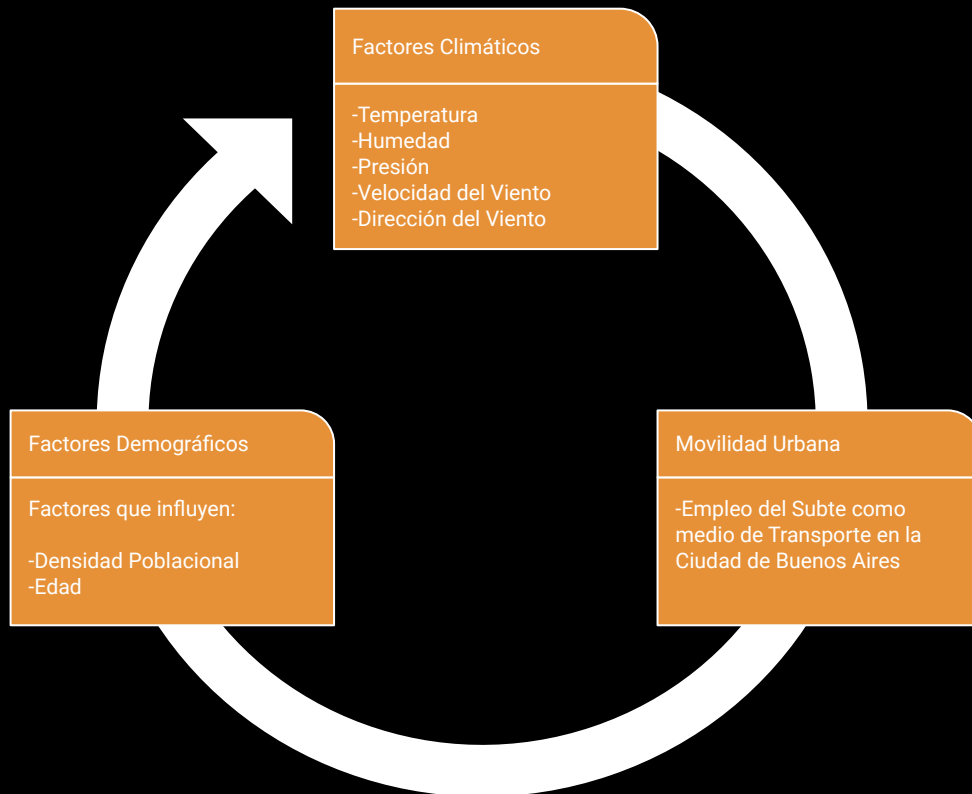
La transmisión del SARS-CoV-2 se produce mediante pequeñas gotas expulsadas al hablar, estornudar, o toser, que al ser despedidas por un portador pasan directamente a otra persona mediante la inhalación, o quedan sobre los objetos y superficies que rodean al emisor, y luego, a través de las manos, que lo recogen del ambiente contaminado, toman contacto con las membranas mucosas orales, nasales y oculares, al tocarse la boca, la nariz o los ojos (Xian Peng, Xin Xu, Yuqing Li, Lei Cheng, Xuedong Zhou, & Biao Ren, 2020). Esta última es la principal vía de propagación, ya que el virus puede permanecer viable hasta por días en todo tipo de lugares.

Debido a la alta contagiosidad de este virus y su rápida propagación en las grandes urbes del Mundo se ha establecido la hipótesis de que los factores demográficos y sociales, influidos también por los factores climatológicos, tienen alta incidencia en la aceleración de contagios.



# Objetivo

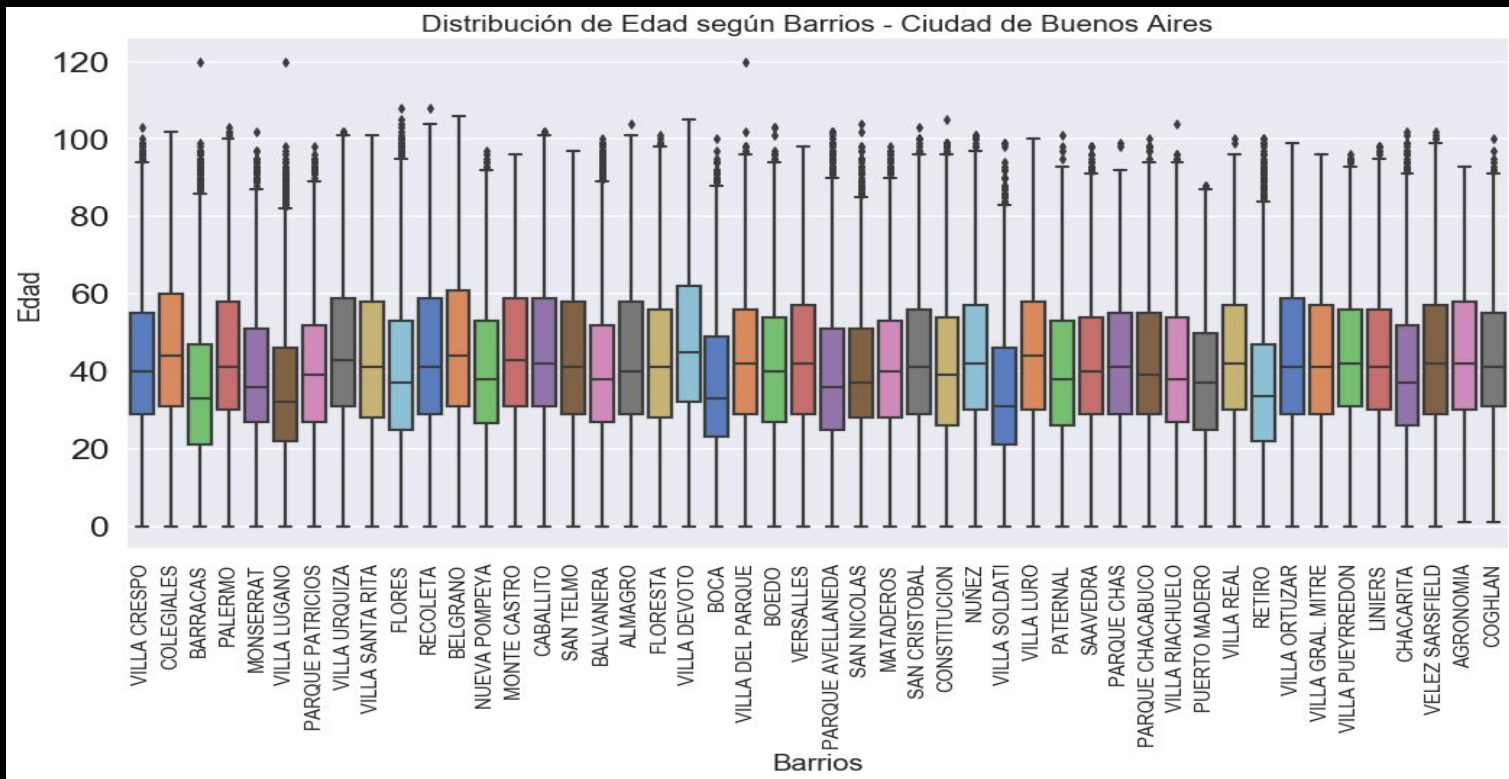
El objetivo de este estudio es plantear un modelo de Regresión, empleando herramientas de Machine Learning, para intentar predecir la cantidad de casos tomando como base los factores antes mencionados.



# Casos de COVID-19 en la Ciudad de Buenos Aires

clusterAI

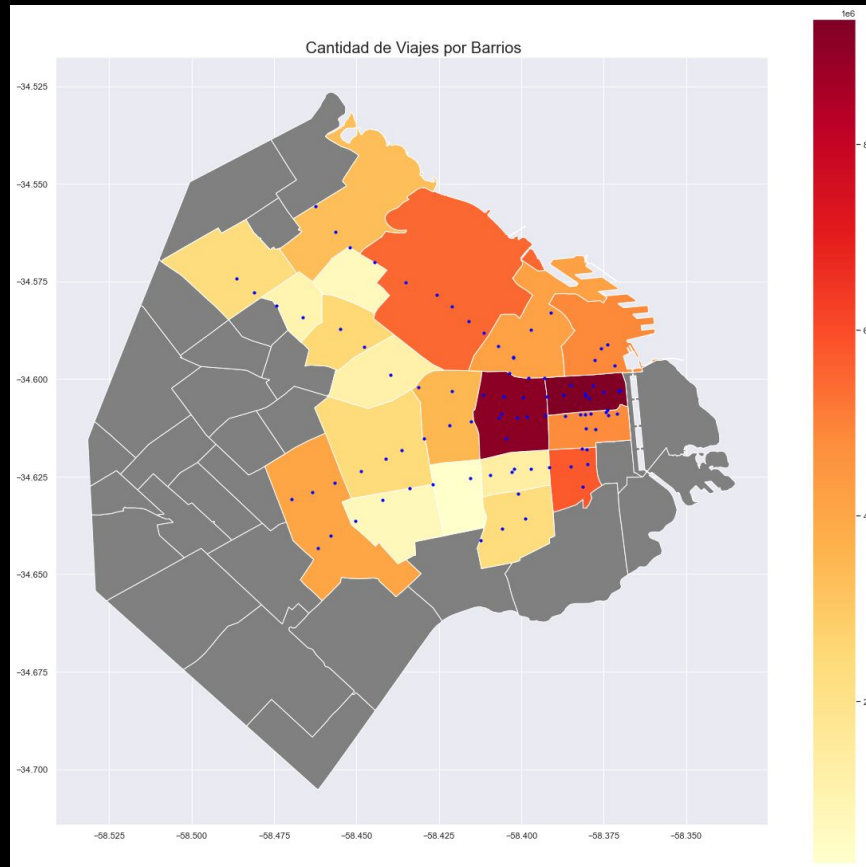
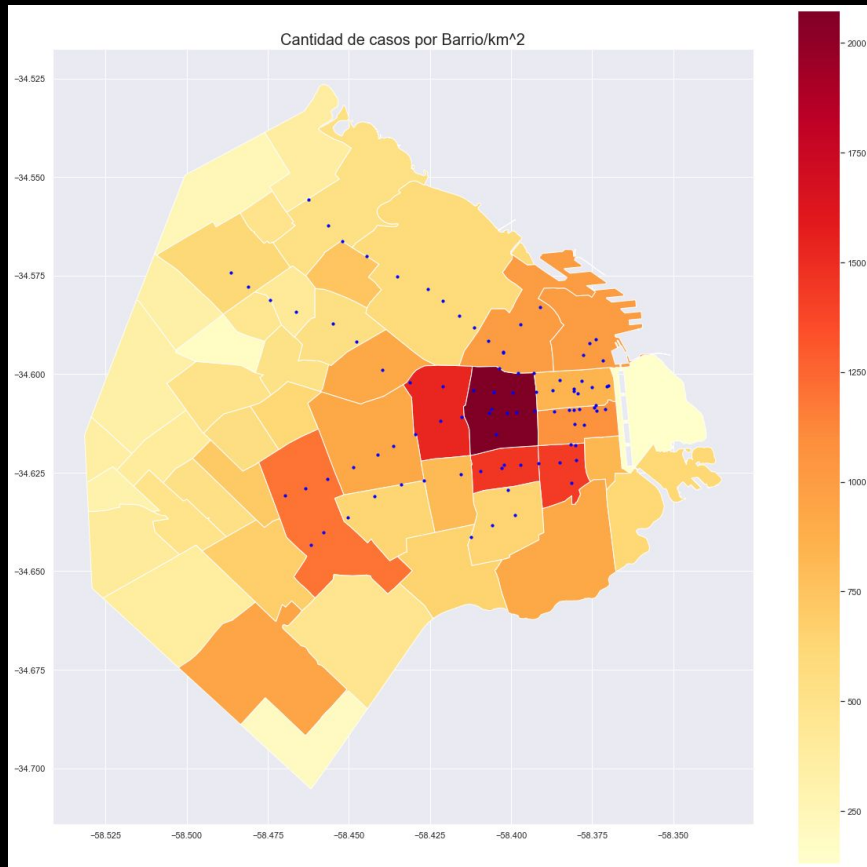
## Distribución según Edad\*



\*Información actualizada: Octubre 2020

# Casos / Superficie - Influencia del Subte BA

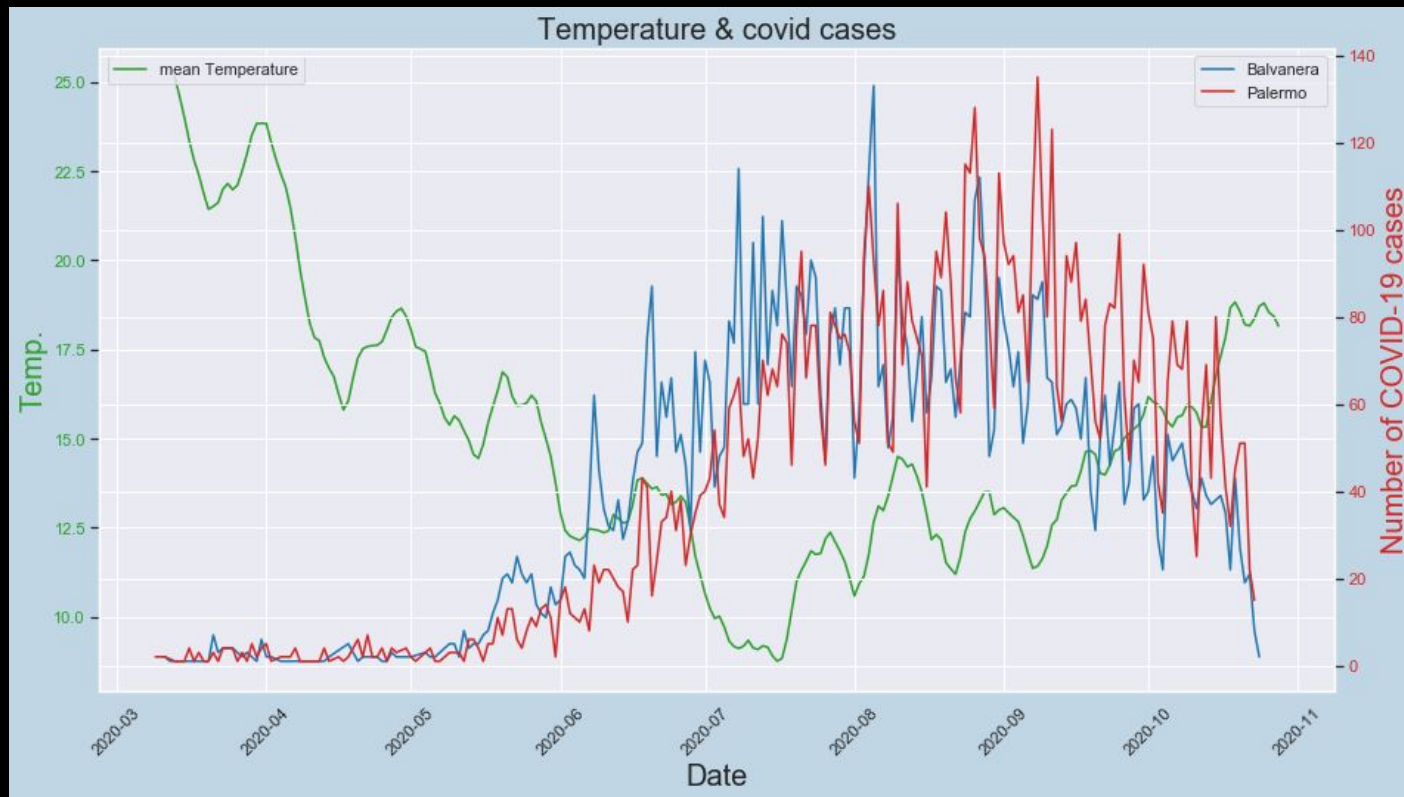
clusterAI



# Casos de COVID-19 en la Ciudad de Buenos Aires

## Incidencia de las Temperaturas

clusterAI



\*Información actualizada: Octubre 2020

# Modelos de Regresión para la Predicción de Evolución de COVID-19

clusterAI

## Modelos

Se emplearán para el armado del modelo de Regresión los siguientes algoritmos:

- Support Vector Regression
- KNN (K-Nearest Neighbors)
- Random Forest Regression

## Datos\*

El set de datos a emplear estará conformado por los siguientes:

- Datos de Subterráneo de la Ciudad de Buenos Aires, año 2020.
- Datos de Clima de la Ciudad de Buenos Aires, año 2020.
- Datos de COVID-19 Argentina y Ciudad de Buenos Aires, año 2020.

## Medición de los resultados

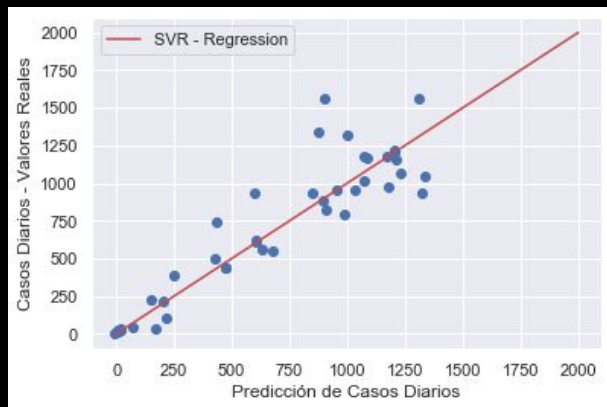
Para medir el rendimiento de los modelos de regresión se emplearán las siguientes métricas :

- $R^2$ : coeficiente de determinación. explica la proporción de la varianza de "y" que explica el modelo de regresión.
- MSE: Mean Squared Error (error cuadrático medio).
- MAE: Mean Average Error (Media del error)

\*Todos los datos fueron extraídos de portales gubernamentales. [datos.gob.ar](https://datos.gob.ar) y el portal de datos de la ciudad Autónoma de Buenos Aires.

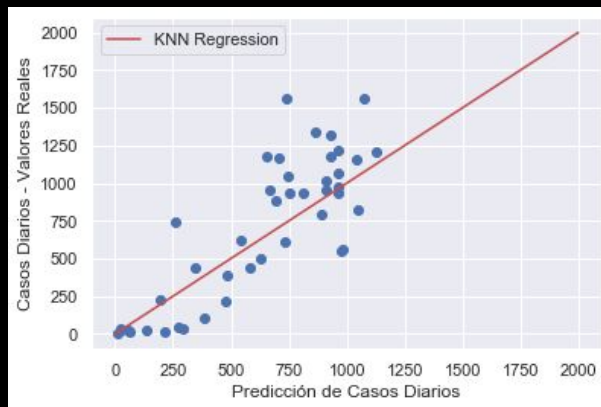
# Modelos de Machine Learning

clusterAI



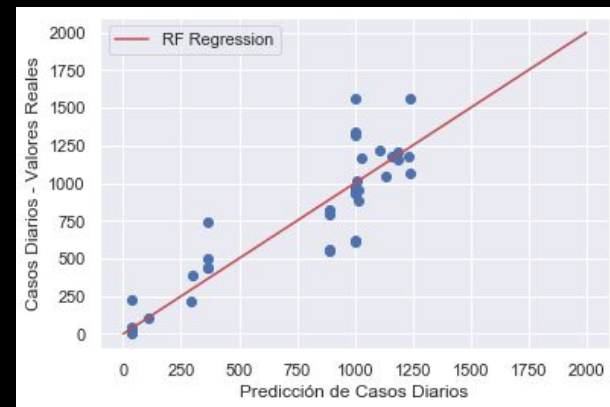
Support Vector Regression

R2 score: 0.846155  
MSE: 34788.083885  
MAE: 118.891602



KNN Regression

R2 score: 0.665938  
MAE: 209.630814  
MSE: 75539.543968



Random Forest Regression

R2 score: 0.847  
MAE: 124.590  
MSE: 34639.112

# Resultados y Conclusiones

	Model	R2	MSE	MAE
1	SVR	0.846	34788.084	118.892
2	KNN	0.666	75539.544	209.631
3	Random Forest	0.847	34639.112	124.590

Los modelos de Support Vector Regression y Random Forest Regression tuvieron un rendimiento aceptable, alcanzando una precisión del 84%.

El modelo KNN no presentó resultados aceptables hasta el momento.

Empleando estos modelos, podría predecirse con una considerable precisión la evolución del COVID-19 en la ciudad de Buenos Aires, a partir de los factores antes mencionados.



MUCHAS GRACIAS

