



# Estimación de la Densidad de Licencias Comerciales en Nueva York



## Objetivo

Desarrollar un modelo predictivo para estimar la densidad de licencias comerciales por habitante, utilizando características demográficas por código postal en Nueva York.

## Fuentes de datos



### Datos demográficos

Estadísticas, por código postal, de personas que participan en programas del Depto. de Desarrollo Comunitario y Juvenil de Nueva York.

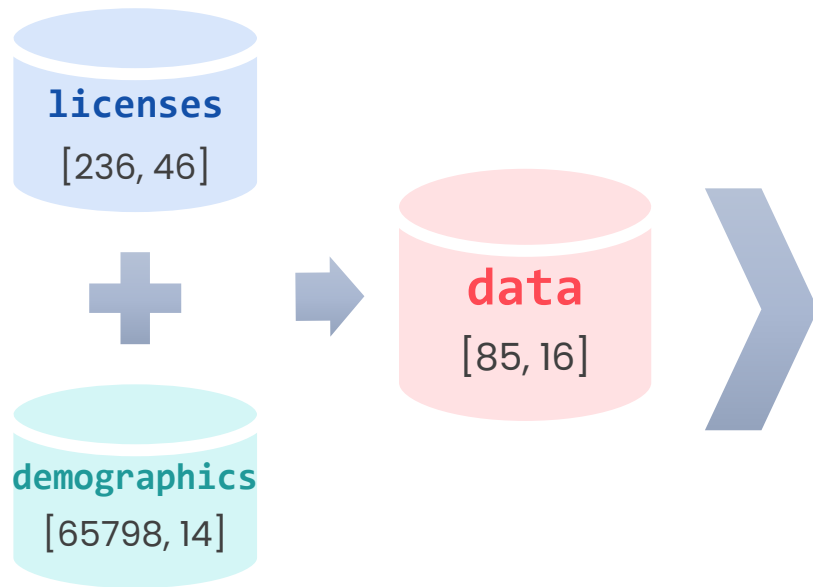


### Licencias

Emitidas por el Depto. de Protección al Consumidor y al Trabajador (DCWP) para empresas y personas naturales en Nueva York.

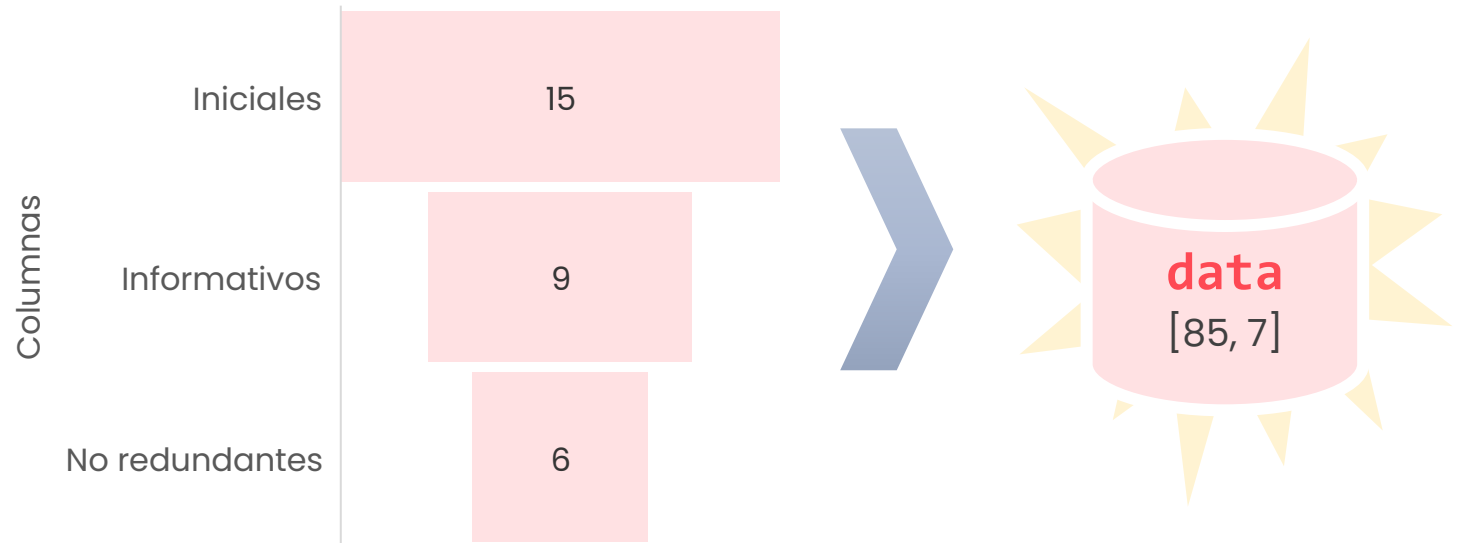
# Preprocesamiento y EDA

## Limpieza y unión de bases de datos



## Exploración de datos

Depuración de *predictores* de tabla **data**



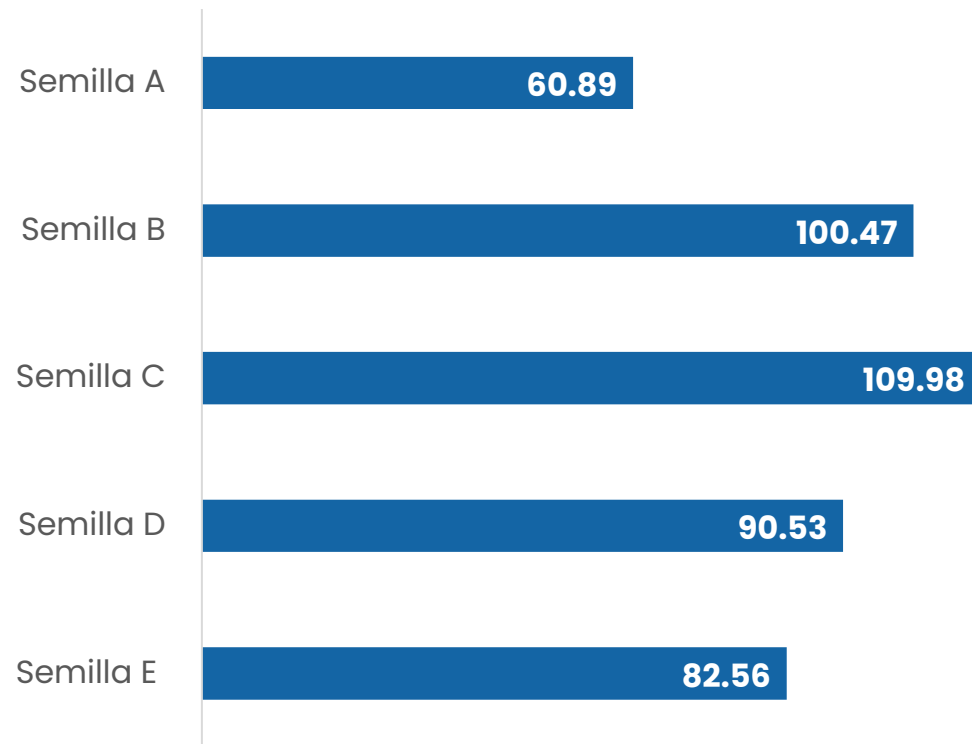
# Metodología



\*Se eligieron **modelos simples para evitar sobreajuste**, incluyendo técnicas de regularización y modelos capaces de capturar patrones no lineales para explorar posibles relaciones complejas en los datos.

# Análisis de resultados

Ej. RMSE de Lasso con Diferentes Semillas Aleatorias



## ► Modelos inestables

- De forma similar a como se ilustra en el gráfico, **todos los modelos** mostraron **métricas inconsistentes** con alta variabilidad.
- La alta variabilidad de las métricas sugiere modelos con **sobreajuste**.
- Los modelos tienen **baja capacidad de generalización**.

# Conclusiones



Los **modelos contruidos no son aptos** para predecir de forma fiable debido a su alta variabilidad y sobreajuste.

- La baja capacidad de generalización puede deberse al **tamaño limitado del conjunto de datos** y a la **debilidad de las relaciones** entre los predictores y la variable objetivo.
- Se recomienda ampliar el conjunto de datos y explorar variables adicionales – como demográficas, económicas o geográficas – para mejorar la robustez y capacidad predictiva de los modelos.