

## Checkpoint 2 - Grupo XX

### Análisis Exploratorio

Realizar una breve descripción del dataset: cantidad de registros y columnas, etc.  
Comentar los features más destacables: tipo de dato, qué representa y por qué se destacan. Listar hipótesis o supuestos que tomaron.

### Preprocesamiento de Datos

Detallar las tareas más importantes que realizaron sobre el dataset, les dejamos algunas preguntas cómo guía:

- ¿Se eliminaron columnas (Nombre de la columna y motivo de eliminación)?
- ¿Detectaron correlaciones interesantes (entre qué variables y qué coeficiente)?
- ¿Generaron nuevos features?
- ¿Encontraron valores atípicos? ¿Cuáles? ¿Qué técnicas utilizaron y qué decisiones tomaron?
- ¿Qué columnas tenían datos faltantes?  
¿En qué proporción? ¿Qué se hizo con estos registros?

### Visualizaciones

Mostrar dos gráficos realizados: dispersión entre variables, histogramas, heatmaps, etc. que sean descriptivos del problema. Seleccionar aquellos que permitan entender cómo se distribuyen los datos, cómo se relacionan la variables etc.  
Comentar brevemente qué se está visualizando en cada caso y por qué los eligieron.

### Clustering

Mencionar la tendencia al clustering y la cantidad apropiada de grupos que se deben formar. Mostrar cómo llegaron a esa conclusión.  
Mostrar un gráfico sobre un mapa de CABA con los avisos coloreados según el grupo al que pertenecen. Describir en función de qué características se formaron dichos grupos.

## Clasificación

Mencionar cuál es la alternativa seleccionada para construir la variable “tipo\_precio” justificando su elección. Mostrar en un mapa de CABA los avisos coloreados por tipo\_precio ¿Qué diferencias o similitudes encuentran con el agrupamiento realizado por K-Means con 3 grupos?

Si llegaron a entrenar alguno de los modelos, mencionar cuáles y qué métricas obtuvieron en test y si realizaron nuevas transformaciones sobre los datos (encoding, normalización, etc) completando los ítems a y b:

### a. Construcción del modelo

#### Arbol de Decision

- ¿Optimizaron hiperparámetros? ¿Cuáles?
- ¿Utilizaron K-fold Cross Validation?¿Cuántos folds utilizaron?
- ¿Qué métrica utilizaron para buscar los hiperparámetros?
- Añadir imagen del árbol generado e incluir descripciones que consideren adecuadas para entender el mismo. Si es muy extenso mostrar una porción representativa.

#### Random Forest

- ¿Optimizaron hiperparámetros? ¿Cuáles?
- ¿Utilizaron K-fold Cross Validation?¿Cuántos folds utilizaron?
- ¿Qué métrica utilizaron para buscar los hiperparámetros?
- Mostrar la conformación final de uno de los árboles generados. Si es muy extenso mostrar una porción representativa y explicar las primeras reglas.

#### Modelo a Elección

- ¿Optimizaron hiperparámetros? ¿Cuáles?
- ¿Utilizaron K-fold Cross Validation?¿Cuántos folds utilizaron?
- ¿Qué métrica utilizaron para buscar los hiperparámetros?
- Mostrar la conformación final de uno de los árboles generados. Si es muy extenso mostrar una porción representativa y explicar las primeras reglas.

### b. Cuadro de Resultados

Realizar un cuadro de resultados comparando los modelos que entrenaron (entre ellos debe figurar cuál es el que seleccionaron como mejor predictor).

Medidas de rendimiento en el conjunto de TEST:

- F1
- Precision
- Recall
- Accuracy
- XXX: si seleccionaron alguna métrica adicional...

Confeccionar el siguiente cuadro con esta información:

Modelo	F1-Test	Precision Test	Recall Test	Accuracy Test	XXX
Arbol de Decision					
Random Forest					
...					

En cada caso ¿Cómo resultó la performance respecto al set de entrenamiento?

**Nota: indicar brevemente en qué consiste cada modelo de la tabla, por ejemplo**

Arbol de decision: {'max\_depth': 7, 'criterion': 'entropy', 'ccp\_alpha': 0.02}

Random Forest: {'criterion': 'entropy', 'min\_samples\_leaf': 1, 'min\_samples\_split': 16, 'n\_estimators': 50}

## Regresión

Si llegaron a entrenar alguno de los modelos, mencionar cuáles y qué métricas obtuvieron en test y si realizaron nuevas transformaciones sobre los datos (encoding, normalización, etc) completando los ítems a y b:

### a. Construcción del modelo

#### KNN

- ¿Utilizaron K-fold Cross Validation? ¿Cuántos folds utilizaron?
- ¿Qué métrica utilizaron para buscar los hiperparámetros?

- Evaluar la performance del modelo en el conjunto de evaluación, explicar todas las métricas. Comparar con la performance de entrenamiento.

#### XGBoost

- ¿Utilizaron K-fold Cross Validation? ¿Cuántos folds utilizaron?
- ¿Qué métrica utilizaron para buscar los hiperparámetros?
- Evaluar la performance del modelo en el conjunto de evaluación, explicar todas las métricas. Comparar con la performance de entrenamiento.

#### Modelo a elección:

- ¿Utilizaron K-fold Cross Validation? ¿Cuántos folds utilizaron?
- ¿Qué métrica utilizaron para buscar los hiperparámetros?
- Evaluar la performance del modelo en el conjunto de evaluación, explicar todas las métricas. Comparar con la performance de entrenamiento.

### **b. Cuadro de Resultados**

Realizar un cuadro de resultados comparando los modelos que entrenaron (entre ellos debe figurar cuál es el que seleccionaron como mejor predictor).

Medidas de rendimiento en el conjunto de TEST:

- MSE
- RMSE
- XXX: si seleccionaron alguna métrica adicional...

Confeccionar el siguiente cuadro con esta información:

Modelo	MSE	RMSE	XXX
KNN			
XGBoost			
...			

En cada caso ¿Cómo resultó la performance respecto al set de entrenamiento?

**Nota: indicar brevemente en qué consiste cada modelo de la tabla**

## Estado de Avance

### 1. Análisis Exploratorio y Preprocesamiento de Datos

**Porcentaje de Avance:** XX%/100%

**Tareas en curso:** detallar en qué puntos están trabajando.

**Tareas planificadas:** detallar con qué tareas tienen pensado continuar.

**Impedimentos:** mencionar temas que sean bloqueantes para ustedes.

- a) Exploración Inicial: detallar tareas pendientes si las hubiera.
- b) Visualización de los datos: detallar tareas pendientes si las hubiera.
- c) Datos Faltantes: detallar tareas pendientes si las hubiera.
- d) Valores atípicos: detallar tareas pendientes si las hubiera.
- e) Opcional: si están trabajando en alguna tarea adicional de preprocesamiento detallar aquí

### 2. Agrupamiento

**Porcentaje de Avance:** XX%/100%

**Tareas en curso:** detallar en qué puntos están trabajando.

**Tareas planificadas:** detallar con qué tareas tienen pensado continuar.

**Impedimentos:** mencionar temas que sean bloqueantes para ustedes.

### 3. Clasificación

**Porcentaje de Avance:** XX%/100%

**Tareas en curso:** detallar en qué puntos están trabajando.

**Tareas planificadas:** detallar con qué tareas tienen pensado continuar.

**Impedimentos:** mencionar temas que sean bloqueantes para ustedes.

#### 4. Regresión

**Porcentaje de Avance:** XX%/100%

**Tareas en curso:** detallar en qué puntos están trabajando.

**Tareas planificadas:** detallar con qué tareas tienen pensado continuar.

**Impedimentos:** mencionar temas que sean bloqueantes para ustedes.

#### Tiempo dedicado

Indicar brevemente en qué tarea trabajó cada integrante del equipo durante estas semanas. Si trabajaron en las mismas tareas lo detallan en cada caso (como en el ejemplo el armado de reporte). Deben indicar el promedio de horas semanales que dedicaron al trabajo práctico. En esta tabla solo deben incluir las tareas que realizaron luego de entregar el CHP1.

Integrante	Tarea	Prom. Hs Semana
Azul Villanueva	Detección de Outliers Armado de Reporte	4
Mateo Suster	Análisis de Correlaciones Armado de Reporte	3