



## **Practica Profesionalizante 2– Evidencia 2**

Alumnos:

Almada, Sergio Daniel  
Beltramone, Mateo  
Guzmán, Nahir Dayana  
Navarro, Paola Giselle  
Galeano, Gerardo Agustín  
Gesto, Giuliana Sofía

[Link al Colab](#)

[Link al Github](#)

# Informe de Análisis de Fairness por Neighborhood

## Resumen Ejecutivo

Este informe presenta un análisis de equidad algorítmica enfocado en la variable **Neighborhood** (barrio) para el dataset de precios de viviendas. El análisis revela **disparidades significativas** entre barrios que pueden introducir sesgos en modelos predictivos de precios inmobiliarios.

## Hallazgos Principales

- **25 barrios** analizados con un **factor de disparidad de 3.6x** entre el barrio más caro y el más económico
- **36% de los barrios** presenta alta volatilidad de precios ( $CV > 0.3$ )
- **2 barrios críticos** con representación insuficiente ( $< 10$  muestras)
- Distribución relativamente equilibrada por categorías de precio

## 1. Análisis de Representación

### 1.1 Distribución por Categorías de Precio

Categoría	Cantidad de Casas	Porcentaje	Status
Bajo	667 casas	45.7%	✓ Bien representada
Medio	515 casas	35.3%	✓ Bien representada
Alto	278 casas	19.0%	Sub-representada

**Interpretación:** La categoría de barrios de precio alto está sub-representada, lo que puede sesgar el modelo hacia predicciones menos precisas para viviendas de lujo.

### 1.2 Barrios con Representación Crítica

Barrio	Número de Muestras	Riesgo
Blueste	2 casas	 Crítico
NPkVill	9 casas	 Alto

**Implicaciones:** Estos barrios tienen muestras insuficientes para entrenar modelos robustos, lo que puede resultar en:

- Predicciones poco confiables
- Alta varianza en las estimaciones
- Potencial discriminación algorítmica

## 2. Análisis de Disparidad de Precios

### 2.1 Métricas de Inequidad

- **Rango de precios medianos:** \$88,000 - \$315,000
- **Factor de disparidad:** 3.6x
- **Barrio más caro:** Precio mediano de \$315,000
- **Barrio más económico:** Precio mediano de \$88,000

### 2.2 Contexto de Disparidad

Un factor de disparidad de **3.6x** indica una **inequidad moderada-alta** en el mercado inmobiliario. Esto sugiere:

- Segmentación clara del mercado por ubicación
- Posibles barreras socioeconómicas de acceso
- Riesgo de perpetuar desigualdades existentes a través de algoritmos

## 3. Análisis de Volatilidad de Precios

### 3.1 Barrios con Alta Volatilidad (CV > 0.3)

Barrio	Coeficiente de Variación	Precio Mediano	Interpretación
StoneBr	0.364	\$278,000	Mercado muy heterogéneo
NoRidge	0.362	\$301,500	Alta variabilidad premium
Crawfor	0.327	\$200,624	Gentrificación posible
NridgHt	0.305	\$315,000	Segmento de lujo diverso
Veenker	0.303	\$218,000	Transición de mercado

Total: 9 barrios (36% del total) con alta volatilidad

### 3.2 Barrios con Baja Volatilidad (CV < 0.15)

Barrio	Coeficiente de Variación	Precio Mediano	Interpretación
NPkVill	0.066	\$146,000	Mercado homogéneo*
Blueste	0.139	\$137,500	Precios consistentes*
BrDale	0.137	\$106,000	Segmento estable




*\*Nota: Los primeros dos barrios tienen muestras muy pequeñas, por lo que la baja volatilidad puede ser artificial.*

## 4. Identificación de Riesgos de Sesgo

### 4.1 Indicadores Detectados

- ✓ **Sesgo de Representación**
  - Barrios de lujo sub-representados (19% vs 45.7% de barrios económicos)
  - Dos barrios con muestras críticamente pequeñas
- ✓ **Sesgo de Volatilidad**
  - 36% de barrios con alta volatilidad puede llevar a predicciones inconsistentes
  - Concentración de volatilidad en segmentos de precio medio-alto
- ❖ **Riesgo de Amplificación**
  - Los algoritmos pueden amplificar las disparidades existentes
  - Predicciones menos confiables para grupos minoritarios

### 4.2 Tipos de Fairness Comprometidos

Tipo de Fairness	Status	Descripción
Representational Fairness	 Comprometido	Sub-representación de barrios específicos
Allocative Fairness	 En riesgo	Disparidad 3.6x puede perpetuarse
Quality of Service	 En riesgo	Menor precisión para grupos pequeños

## 5. Recomendaciones Estratégicas

### 5.1 Acciones Inmediatas (Corto Plazo)

- 1. Tratamiento de Muestras Pequeñas**
  - Agrupar Blueste y NPkVill con barrios similares
  - Considerar modelos jerárquicos por región
- 2. Estratificación del Modelo**
  - Entrenar modelos separados por categoría de precio
  - Implementar ensemble weighted por representación
  - Validación cruzada estratificada por barrio
- 3. Monitoreo de Equidad**
  - Métricas de fairness por barrio en producción
  - Alertas automáticas para disparidades >4x

- Dashboard de equidad para stakeholders

## 5.2 Estrategias de Mediano Plazo

### 1. Recolección de Datos

- Enriquecer dataset con más muestras de barrios sub-representados
- Incorporar variables socioeconómicas contextuales
- Datos temporales para análisis de tendencias

### 2. Técnicas Avanzadas de Fairness

- Implementar constraint-based fairness durante entrenamiento
- Post-processing para calibrar predicciones por grupo
- Adversarial debiasing networks

### 3. Validación Robusta

- Testing con datos externos por barrio
- Auditorías de equidad trimestrales
- Análisis de impacto diferencial

## 5.3 Consideraciones Éticas y Regulatorias

### 1. Transparencia

- Documentar limitaciones por barrio
- Comunicar incertidumbre en predicciones
- Explicabilidad localizada por grupo

### 2. Gobernanza

- Comité de ética para decisiones de fairness
- Políticas de uso responsable del modelo
- Protocolo de escalación para casos edge

---

## 6. Conclusiones

### 6.1 Estado Actual

El análisis revela **desafíos moderados de fairness** en el dataset, con disparidades significativas pero manejables. La representación desigual y alta volatilidad en algunos barrios requieren **intervención proactiva** para evitar discriminación algorítmica.

### 6.2 Viabilidad del Proyecto

- ✓ **Viable con mitigaciones:** El proyecto puede proceder implementando las recomendaciones de fairness propuestas.

## 6.3 Próximos Pasos Prioritarios

1. Implementar agrupación de barrios con muestra pequeña
  2. Estratificar validación por categoría de precio
  3. Establecer métricas de monitoreo continuo
  4. Desarrollar pipeline de fairness automatizado
- 

## Anexos

### A.1 Referencias Metodológicas

- IEEE Standards for Algorithmic Bias Assessment
  - Google's What-If Tool para análisis de fairness
  - IBM AI Fairness 360 toolkit
-