	<b>PRUEBA DE CONOCIMIENTOS CIENTÍFICO DE DATOS</b>		
	<b>VICEPRESIDENCIA ANALÍTICA DE NEGOCIO - DEPTO. ANALÍTICA DE DATOS NO ESTRUCTURADOS</b>		
	<b>Depto. Datos No Estructurados</b>	<b>Elaborado</b> <b>Juan Felipe Peña</b>	<b>Versión:2</b>

## Prueba de Conocimiento Científico de Datos: Análisis de precios de Vivienda

**Entrega:** [juan.pena@davivienda.com](mailto:juan.pena@davivienda.com), [javier.herrera@davivienda.com](mailto:javier.herrera@davivienda.com)

En el equipo de Analítica buscamos personas con excelentes capacidades técnicas a las que les guste explorar los últimos avances en IA para asumir retos de especial dificultad. Aunque la siguiente prueba busca, en primera instancia, corroborar su idoneidad técnica, para nosotros es muy importante que esta destreza venga acompañada de dos elementos adicionales:

1. Creatividad para encarar desafíos técnicos en procesamiento de datos..
2. Habilidad para comunicar el trabajo hecho y sus resultados a un público general o experto.


Por lo tanto, el objetivo de la prueba es tener una visión completa de cómo es su compromiso, capacidad analítica, creatividad técnica y comunicación en la solución de un problema puntual de analítica, la cual, en la práctica, se verá reflejada en una presentación a los jurados.

Para el desarrollo de esta prueba cuentan con una semana, tiempo en el que deben desarrollar la solución del problema mediante código debidamente documentado, depositado en un cuaderno de Python (Google Colab), así como una presentación gerencial de su trabajo. De parte de los jurados se evaluarán estos dos elementos y se comprometen a respetar la propiedad intelectual de los desarrollos hechos por ustedes donde haya lugar (modelos personalizados, APIs propias, técnicas especiales, etc). Si lo desea, puede hacerlo explícito mediante licencia Creative Commons.

Puede hacer los supuestos que considere necesarios. Esto depende de la forma en que usted aborde el problema. Aunque los cuadernos de Python suministrados sugieren caminos, no hay una solución única. Fundamentalmente queremos ver soluciones ordenadas, creativas (si consideran que esta cualidad es pertinente en la resolución del problema) y bien comunicadas.

### Introducción

La estimación de los precios de la vivienda es esencial tanto para los propietarios como para los inversores, ya que ambos necesitan entender el valor de este activo inmobiliario. Para

	<b>PRUEBA DE CONOCIMIENTOS CIENTÍFICO DE DATOS</b>		
	<b>VICEPRESIDENCIA ANALÍTICA DE NEGOCIO - DEPTO. ANALÍTICA DE DATOS NO ESTRUCTURADOS</b>		
	<b>Depto. Datos No Estructurados</b>	<b>Elaborado</b> <b>Juan Felipe Peña</b>	<b>Versión:2</b>

muchas personas, comprar una propiedad es una de las decisiones y compras más importantes en la vida.

Además de la asequibilidad de una vivienda, otros factores, como la conveniencia del lugar y las perspectivas de inversión a largo plazo, también afectan el proceso de toma de decisiones. El mercado inmobiliario está expuesto a muchas fluctuaciones en los precios debido a las correlaciones existentes con muchas variables, algunas de las cuales no se pueden controlar o incluso pueden ser desconocidas. Los precios de las viviendas pueden aumentar rápidamente (o en algunos casos, también bajan muy rápido).


Algunas aplicaciones para un banco son:

1. Origenación: Establecer el valor comercial del inmueble a financiar utilizado para la aprobación final.
2. Retanqueo: Actualizar el valor comercial de garantía ya existente para aprobación de nuevos cupos de crédito.
3. Monitoreo portafolio garantías: Valoración del portafolio de las garantías para cumplimiento normativo y para el análisis de riesgos del colateral.
4. Normalización de cartera: Evaluar los préstamos existentes, evaluar los acuerdos de refinanciamiento y daciones en pago.

**El objetivo es que usando las características de las viviendas de la base de datos reconozca patrones y prediga el precio de tal manera que pueda ser usado en otras viviendas donde no exista aún un avalúo.**

## Instrucciones Importantes

- Recuerde que la variable que debe estimar es la de **valor\_total\_avaluo**
- **train\_precios\_vivienda.csv**: Este conjunto de datos tiene características de los inmuebles y su precio total (incluye áreas privadas, públicas, parqueaderos y cuarto útil en caso de tener). La idea es que cree el modelo teniendo en cuenta esta información.
- Entregar **predict\_prices.csv** con dos columnas: **id** y **valor\_total\_avaluo\_predict** (Valor predicho de valor\_total\_avaluo) para la base de **test\_precios\_vivienda.csv**
- El archivo **metadata.pdf** tiene información de las variables.
- El archivo **Puntosdeinteres.csv** tiene información complementaria.
- La evaluación del modelo tendrá en cuenta la métrica **MAPE**

	<b>PRUEBA DE CONOCIMIENTOS CIENTÍFICO DE DATOS</b>		
	<b>VICEPRESIDENCIA ANALÍTICA DE NEGOCIO - DEPTO. ANALÍTICA DE DATOS NO ESTRUCTURADOS</b>		
	<b>Depto. Datos No Estructurados</b>	<b>Elaborado</b> <b>Juan Felipe Peña</b>	<b>Versión:2</b>

- Existe un identificador único por inmueble este es id. Todos los inmuebles son únicos, por lo cual no se tiene medido el valor de un mismo inmueble en diferentes puntos del tiempo.
- En el archivo metaData.pdf hay algunas descripciones de unas pocas variables. Se deja a interpretación la naturaleza de cada campo.
- En PuntosInteres.csv hallará puntos de interés geográfico con su respectiva longitud y latitud. La tercera columna es la categoría del punto de interés y la cuarta columna podemos encontrar el nombre si se tiene. (Utilicela) y describa en su informe cómo la utilizo, porque y los beneficios qué obtuvo de tenerla.

## Objetivo

El propósito de la prueba es medir sus capacidades de análisis, tratamiento, preparación y limpieza de datos. De igual manera se busca retar sus capacidades para la formulación de un producto innovador a partir de los datos disponibles y el análisis desarrollado.


Hagamos saber qué otros datos o atributos adicionará idealmente al conjunto de datos, para un producto más efectivo. Aquí, tenga en cuenta la factibilidad y el costo de obtener esos datos. Con esta información y sus propios análisis que utilidad extraería. Especifique para quién y de qué modo y el porqué es útil.

Adicionalmente con el uso anterior en mente, diseñe un sistema de manera teórica que bosqueje una solución que permita hacer disponible los resultados de su producto y que sean fácilmente consumibles por servicios externos, páginas web, servicio móvil, etc. No se tiene que desarrollar. Desarrollar una aplicación o sistema de información no da ningún punto extra y no será tomada en cuenta para la calificación total por lo que recomendamos no desarrollarla sino únicamente elaborar el bosquejo y describir la estrategia.(No se le olvide)

## Entregables

**Presentación Gerencial:** Una presentación (en formato PDF) donde se encuentre:

- Limpieza y manipulación de datos. (Transformaciones, tratamiento de variables, etc)
- Exploración de datos.
- Técnicas usadas o exploradas.
- Métricas resultantes de los modelos usados
- Dificultades encontradas.
- Aspectos importantes.
- Demás detalles que considere relevantes para la presentación de su proceso.

	<b>PRUEBA DE CONOCIMIENTOS CIENTÍFICO DE DATOS</b>		
	<b>VICEPRESIDENCIA ANALÍTICA DE NEGOCIO - DEPTO. ANALÍTICA DE DATOS NO ESTRUCTURADOS</b>		
	<b>Depto. Datos No Estructurados</b>	<b>Elaborado</b> <b>Juan Felipe Peña</b>	<b>Versión:2</b>

**Códigos:** También nos debe entregar la implementación de su análisis en el cuaderno, tenga en cuenta que el orden y la documentación del código será evaluada.

**Repositorio de Github** que incluya Cuaderno de Google Colab suministrado con el desarrollo hecho, debidamente documentado (o su URL). Los resultados de la ejecución de cada celda deben ser visibles sin necesidad de correr las celdas. Tenga en cuenta que no vamos a modificar rutas de ubicación de input/output, archivos auxiliares, etc. por lo que todo el contenido debe ser completo y consistente dentro del archivo zip entregado.

**(Opcional)** - Si su desarrollo necesita de archivos auxiliares para funcionar, por favor inclúyelos dentro del archivo zip de manera que sea consistente con la ejecución de los códigos hechos.

**Nota:** Los resultados deben entregarse por medio de correo electrónico a [juan.pena@davivienda.com](mailto:juan.pena@davivienda.com)

- Con el siguiente asunto: Prueba de Conocimiento Científico de Datos [nombre participante]
- Con la siguiente nomenclatura: 01\_nombrep participante.zip

## Evaluación

Se evaluarán sus capacidades de análisis, manejo de datos, programación y comunicación. Esto se realizará bajo la siguiente metodología:

- Desarrollo de los códigos: 30%
  - Orden (5%)
  - Documentación (5%)
  - Técnica (20%)
- Presentación con los resultados: 70%
  - Orden en el relato del proceso (20%)
  - Conocimientos técnicos (20%)
  - Completitud (10%)
  - Claridad (20%)