



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE ESTUDIOS SUPERIORES ACATLÁN

**Proyecto Final – Módulo I**  
**Análisis del Mercado Inmobiliario en México de 2013 a 2016**

PRESENTA

**Edgar David Cardoso Olvera**  
**314551885**

PROFESOR

**Carla Paola Malerva Resendiz**

DIPLOMADO

**Ciencia de Datos**

5 de noviembre de 2020

# Índice general

<b>1. Introducción</b>	<b>3</b>
1.1. Objetivo . . . . .	3
1.2. Metodología del Análisis . . . . .	3
1.3. Fuente de Datos . . . . .	3
<b>2. Análisis Exploratorio de Datos</b>	<b>4</b>
2.1. Diccionario de Datos . . . . .	4
2.2. Tipos de Variables . . . . .	5
2.3. Completitud . . . . .	5
2.4. Estadística Descriptiva . . . . .	6
2.5. Visualización de Datos . . . . .	6

# Capítulo 1

## Introducción

Una de las necesidades básicas y primordiales del ser humano es la vivienda, es por eso que el sector inmobiliario es tan importante para los países. Este mercado puede afectar factores tanto políticos como económicos, como se vio en la crisis inmobiliaria de 2008 que dañó a toda la economía mundial.

En México, siendo uno de los países con mayor población mundial, el mercado inmobiliario es uno de los sectores que mantiene a flote el Producto Interno Bruto. En el primer trimestre de 2020 el mercado de bienes raíces representó un 10 % del PIB en México, contribuyendo con 2.4 billones de pesos según datos de Inegi.

### 1.1. Objetivo

El objetivo principal de este proyecto es analizar y conocer el comportamiento del mercado inmobiliario mexicano a partir del año 2013 a 2016, esto con ayuda de la base de datos más grande del mundo, Internet. Se analizarán componentes primordiales en anuncios de venta y renta de inmuebles en internet, como lo son: fechas, precios, tipo de inmueble, ubicación, superficie, descripción, entre otras. Todo lo anterior con la finalidad de comprender uno de los mercados más grandes e importantes en México.

### 1.2. Metodología del Análisis

Al recabar la información del mercado inmobiliario en México se realizará un análisis con respecto a la calidad de los datos, esto con el fin de mejorar la toma de decisiones y facilitar el procesamiento de la información. Lo anterior tiene como objetivo conocer la falta de información en la base de datos principal y comprender qué características son importantes para el proyecto. De igual manera se le dará un formato homogéneo a la información para así poder trabajar de una manera más eficiente.

Después se realiza un análisis estadístico para conocer las distribuciones de los datos y su comportamiento, esto es necesario para identificar valores fuera de los rangos comunes dentro de la base. Este mismo análisis ayudará a la creación de visualizaciones, que facilitarán la comprensión de la información del mercado inmobiliario.

### 1.3. Fuente de Datos

El conjunto de datos se obtuvo de la página web <https://data.world/>. La fuente original de los datos contiene dos archivos de inmuebles publicados en México del año 2013 al 2016 divididos por el tipo de operación: Venta y Renta. Los archivos anteriores fueron unidos para crear una base de datos más completa y funcional con un total de 188,525 registros. Un agradecimiento al usuario @properati por haber proporcionado esta fuente de datos esencial para el análisis inmobiliario en México.

## Capítulo 2

# Análisis Exploratorio de Datos

El análisis exploratorio de la información tiene como objetivo reconocer patrones significativos en nuestros datos, encontrar irregularidades en la información y comprende de manera rápida y eficaz el conocimiento que estos transmiten. También brindará información relevante para crear visualizaciones para un mejor entendimiento.

### 2.1. Diccionario de Datos

En la siguiente tabla se da a conocer las características que se tienen de cada publicación dentro de la tabla de datos, esto es fundamental para conocer con herramientas contamos para el análisis de la información.

Nombre de la Variable	Tipo de Dato	Descripción
created_on	String	Fecha de Publicación
operation	String	Tipo de Operacion: Venta o Renta
property_type	String	Tipo de Propiedad: Casa, Departamento, Tienda, PH
place_name	String	Nombre de Municipio o Colonia
place_with_parent_names	String	Ubicacion con País, Estado, Municipio, Colonia
geonames_id	String	ID Geográfico
lat-lon	String	Latitud y Longitud Unidas
lat	Float64	Latitud con Decimales
lon	Float64	Longitud con Decimales
price	Float64	Precio de Inmueble
currency	String	Tipo de Moneda en la Publicación
price_aprox_local_currency	Float64	Precio Aproximado en Pesos Mexicanos
price_aprox_usd	Float64	Precio Aproximado en Dólares Americanos
surface_total_in_m2	Float64	Superficie total en m2
surface_covered_in_m2	Float64	Superficie construida en m2
price_usd_per_m2	Float64	Precio en Dólares Americanos por m2
price_per_m2	Float64	Precio por m2
floor	Float64	Piso de Departamento
rooms	Float64	Número de Cuartos
expenses	Float64	Gastos Adicionales
properati_url	String	Dirección Web
description	String	Descripción del Inmueble
title	String	Título del Anuncio
image_thumbnail	String	Dirección Imagen

---

## 2.2. Tipos de Variables

- Variables Continuas:  
lat, lon, price, price\_aprox\_local\_currency, price\_aprox\_usd, surface\_total\_in\_m2, surface\_covered\_in\_m2, price\_per\_m2, price\_usd\_per\_m2, floor, rooms, expense
- Variables Discretas:  
operation, property\_type, currency, estado, municipio, colonia, geonames\_id
- Variables de Fechas:  
created\_on
- Variables de Texto:  
properati\_url, description, title, image\_thumbnail

## 2.3. Completitud

Nombre de la Variable	% De Nulos
created_on	0.00 %
operation	0.00 %
property_type	0.00 %
place_name	0.00 %
place_with_parent_names	0.00 %
geonames_id	100.00 %
lat-lon	16.94 %
lat	16.94 %
lon	16.94 %
price	1.39 %
currency	1.39 %
price_aprox_local_currency	1.39 %
price_aprox_usd	1.39 %
surface_total_in_m2	48.21 %
surface_covered_in_m2	3.95 %
price_usd_per_m2	58.79 %
price_per_m2	11.23 %
floor	84.77 %
rooms	96.78 %
expenses	99.89 %
properati_url	0.00 %
description	0.00 %
title	0.00 %
image_thumbnail	2.93 %

Algunas variables como geonames\_id que contiene un sólo dato o la variable expense con 207 datos serán eliminados en los siguientes pasos ya que contienen mas del 90 % de información nula, lo cuál perjudica el análisis de la información.

---

## 2.4. Estadística Descriptiva

	c_lat	c_lon	c_price	c_price_aprox_local_currency
count	156134	156134	185454	185454
mean	21.034041540242	-99.0133530053932	2559721.73365363	3183243.06557337
std	2.97352809912203	5.17021181128755	6038333.16636718	7963734.31420877
min	14.843818	-117.228632	50	940.42
10 %	18.917542	-103.438171	13500	14864.16
20 %	19.16150848	-101.00246024	155000	252693.13
30 %	19.367960277	-100.356804	580000	672019.707
40 %	19.454332	-99.6030794	950000	1090307.49
50 %	20.1242815	-99.2277873	1400000	1565706.81
60 %	20.6965424	-99.168088	1880000	2105776.52
70 %	21.039522	-98.979619523	2570554	2873765.76
80 %	22.272658848	-97.837899572	3567479.99999999	4063873.78
90 %	25.643093	-89.636932	5900000	6936675.98
max	41.577487	99.206936	945000000	936451275.4

	c_price_aprox_usd	c_surface_total_in_m2	c_surface_covered_in_m2	c_price_per_m2
count	185454	97630	180635	166908
mean	169244.919603299	448.098811840623	1993.31448501121	16247.7332532109
std	423411.453007493	3017.1541337506	563709.794917386	531832.067114607
min	50	-396	-324	0.008
10 %	790.29	0	46	99.7603322270001
20 %	13435.05	35	70	1813.6720144
30 %	35729.575	90	90	6428.571429
40 %	57968.87	120	120	8333.333333
50 %	83244.64	160	150	9941.0024365
60 %	111958.77	200	190	11410.25641
70 %	152790.8	282	235	13279.209701
80 %	216065.81	400	300	16190.47619
90 %	368805.38	674	431	23500
max	49788727.19	200000	230303030	199500000

Al observar las variables de coordenadas vemos que los valores máximos y mínimos de latitud y longitud difieren en las coordenadas máximas y mínimas del territorio mexicano, al crear los distintos gráficos se analizarán estos casos particulares.

En el caso del precio de propiedades se analizará el valor máximo de esta variable ya que existen propiedades con un precio superior a los 900 millones de pesos mexicanos.

La variable superficie total será analizada detalladamente, ya que en el primer percentil la longitud total es de 0 metros cuadrados.

## 2.5. Visualización de Datos

Gráficas