

# Housing Market

## 1. Entendimiento de la situación actual:

El mercado inmobiliario es un sector que sufre subidas y bajadas en los precios de compra de propiedades, así como también la renta de alquiler; por lo cual es importante saber el impacto económico que puede llegar a tener esta industria. Esto puede variar año con año y presentarse situaciones económicas que pueden afectar el mercado como en el año 2008, donde se tuvo un período recesivo.

Este proyecto se enfocará en un estudio de mercado inmobiliario de Estados Unidos de América (EUA). Se realizará un análisis de qué zonas de los Estados Unidos conviene realizar inversiones inmobiliarias, independientemente de la situación económica que pueda llegar a estar sucediendo. Tomando en consideración factores como el tipo de departamento a comprar y en qué lugar, como también en qué momento del año es mejor realizar la inversión en un inmueble, clima, etc.

## 2. Objetivos:

Encontrar las mejores zonas para invertir en distintos tipo de inmuebles como:

- Departamentos
- Casas
- Condominios
- Edificios
- Complejos de casas

Buscar la relación que puede tener el precio de compra o alquiler de un inmueble tomando las siguientes características:

- Clima
- Delincuencia
- Tasa de empleo
- Población.

## 3. Alcances:

Se tomarán datasets de la página de Zillow (<https://www.zillow.com>), los cuales proporcionarán información sobre la compra, venta y renta de distintos tipo de viviendas, con los cuales se obtendrá un análisis del comportamiento inmobiliario de las distintas zonas de EUA y de esta forma determinar cuáles son las propiedades en las que es mejor invertir.

Se analizarán las condiciones que pueden afectar el precio de la propiedad tomando en cuenta características como la cantidad de población, nivel de delincuencia, poder adquisitivo de la zona, así como también el clima, entre otros. De los cuales se obtendrá información de datasets externos que puedan brindar una mayor información al respecto.

El proyecto se delimitará a las consideraciones anteriores sin tener en cuenta el panorama económico de EUA, así como también los períodos recesivos o de incertidumbre económica.

#### 4. Objetivos y KPIs asociados:

Para definir qué zonas son las más convenientes para invertir se tendrá en cuenta diferentes factores. Uno de ellos es el precio del inmueble, el de venta y el del alquiler. Relacionado a esto se utilizarán las siguientes KPIs:

- **Período de recuperación:** Este KPI determina el número de años que se necesitarán para recuperar la inversión inicial realizada en un inmueble.

$$\text{Período de recuperación} = \text{Coste del capital inicial para el proyecto} / \text{Ahorros anuales o beneficios del proyecto}$$

- **Rendimiento de las inversiones (ROI):** Es importante conocer el rendimiento de la inversión actual y pasada. Para esto se utiliza este indicador inmobiliario.

$$\text{ROI} = (\text{Beneficios netos} / \text{Inversión total}) * 100 \%$$

- **Precio medio del alquiler por inmueble:** Este indicador inmobiliario determina el precio medio del alquiler mensual para ayudar a comparar los cambios trimestrales o anuales.

$$\text{Precio medio del alquiler por inmueble} = \text{Total de ingresos mensuales} / \text{Total de inmuebles}$$

- **Compete Score:** Evalúa qué tan competitiva es un área en una escala de 0 a 100, donde 100 es la más competitiva. La competitividad de un mercado es un indicador de lo difícil que es vender una vivienda en una zona. El puntaje se calcula principalmente en función de cuatro entradas: número de ofertas competidoras, contingencias renunciadas, relación de venta a lista y número de días en el mercado.

- **Inmuebles disponibles vendidos.** Este indicador inmobiliario puede ayudar a identificar las condiciones del mercado actual de una zona concreta. Analiza cuántos inmuebles se han vendido y compara esa cantidad con el número de inmuebles disponibles en esa zona.

$$\text{Inmuebles disponibles vendidos} = \text{Total de inmuebles vendidos} / \text{Total de inmuebles disponibles}$$

También se analizarán otros factores como el clima y la seguridad del lugar.

## 5. Repositorio:

Link del repositorio:

[https://github.com/julicar31/GRUPO\\_07-PF\\_DS4.git](https://github.com/julicar31/GRUPO_07-PF_DS4.git)

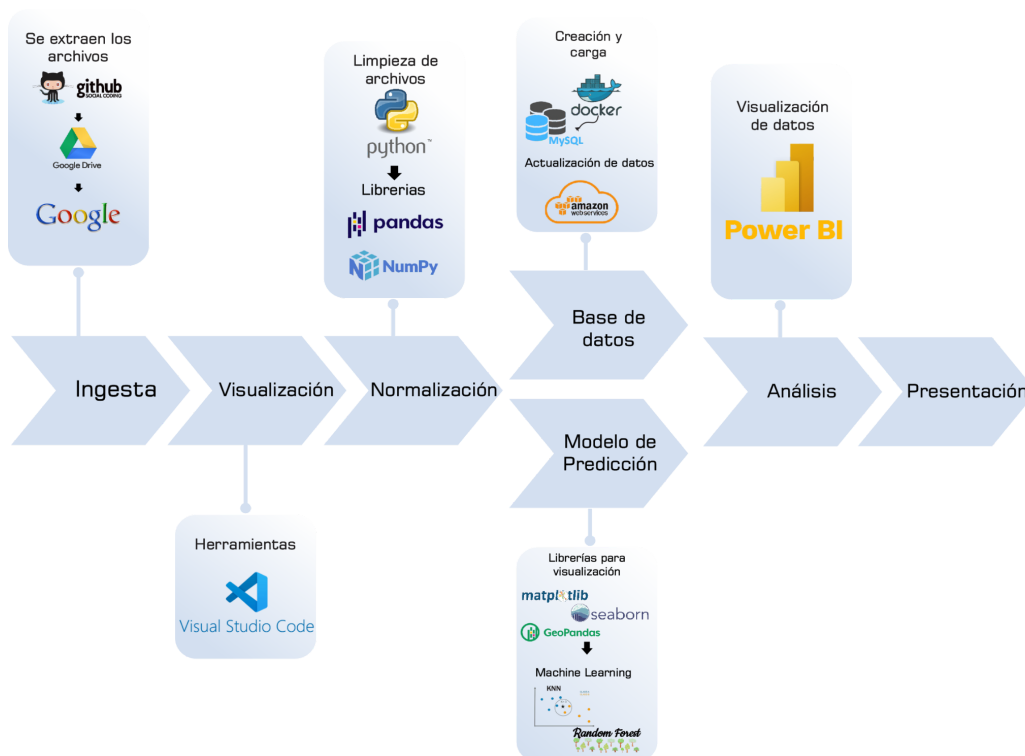
En este se encontraran los archivos que se estarán trabajando durante el proyecto:

Datasets consultados  
Archivos .py o .ipnyb  
Dashboard  
Informe (calidad, \*resultados)

## 6. Solución propuesta:

- Stack tecnológico: Se utilizará *Python* para la limpieza de datos y el código para la carga incremental de la información a una base de datos. Como base de datos se utilizará una imagen de *MySQL* en *Docker* y se conectará con el servicio RDS de AWS. Para el reporte y dashboard se hará uso de PowerBI.

## Metodología de trabajo:



**Entregables:**

Semana 1. Propuesta del proyecto

Semana 2. Documentación  
Stack elegido y fundamentación  
Flujo de trabajo

Semana 3. Dashboard  
Modelo de Machine Learning  
Reporte

Semana 4. Repositorio  
Documentación

**Cronograma:**

IDENTIFICADOR	TAREA TÍTULO	MIEMBRO ENCARGADO	EMPEZAR FECHA	FIN ESPECULADO	DURACIÓN EN DÍAS	PCT DE LA TAREA ÍNTEGRO
1	Phase 1. Project Scope					
1.1	Planteamiento y alcance	Equipo	14/11/2022	15/11/2022	1	100%
1.2	Solución y cronograma	MV - KC	15/11/2022	16/11/2022	1	100%
1.3	Establecer KPI's	JC - EH	15/11/2022	16/11/2022	1	100%
1.4	Definir datasets	Equipo	16/11/2022	17/11/2022	1	100%
1.5	EDA preliminar	Equipo	16/11/2022	18/11/2022	2	90%
1.6	Presentación de avances	Equipo	18/11/2022	18/11/2022		100%
2	Phase 2. Data Process					
2.1	Validación de datos	Equipo	21/11/2022	22/11/2022	1	0%
2.2	Creación Data Warehouse	JC	22/11/2022	24/11/2022	2	0%
2.3	Definición del modelo ER	EH	22/11/2022	24/11/2022	2	0%
2.4	Pipelines	MV	22/11/2022	24/11/2022	2	0%
2.5	Automatización del DW	KC	22/11/2022	24/11/2022	2	0%
2.6	Presentación de avances y reportes	Equipo	25/11/2022	25/11/2022		0%
3	Phase 3. Data Analytics					
3.1	Dashboard	EH	28/11/2022	30/11/2022	2	0%
3.2	Implementación de modelos de ML	MV	28/11/2022	30/11/2022	2	0%
3.3	Ajustes de modelos de ML	KC	28/11/2022	30/11/2022	2	0%
3.4	Ajustes de automatización	JC	28/11/2022	30/11/2022	2	0%
3.5	Presentación de avances y reportes	Equipo	02/12/2022	02/12/2022		0%
4	Phase 4. Final Demo					
4.1	Preparación de la demo final	MV-EH	05/12/2022	06/12/2022	1	0%
4.2	Entregable final	JC -KC	05/12/2022	06/12/2022	1	0%
4.3	Presentación de la demo	Equipo	06/12/2022	06/12/2022		0%

- **Análisis preliminar de calidad de datos:**

- **Unique\_City\_ID:** *String*, identificador único para cada ciudad de Estados Unidos. Es un nombre compuesto por la ciudad, el condado y el estado.
- **City:** *String*, ciudad de Estados Unidos.
- **County:** *String*, condado de Estados Unidos.
- **State:** *String*, estado de Estados Unidos.
- **Date:** *String*, fecha en que ocurre un evento.
- **MedianListingPrice\_X:** *Float*, la mediana del precio de venta de un inmueble de tipo X.
- **MedianRentalPrice\_X:** *Float*, la mediana del precio de renta de un inmueble de tipo X.
- **House\_Price\_Index:** *Float*, índice de precio de inmuebles.
- **Stock\_Price\_Index:** *Float*, índice de precio de las acciones.
- **Consumer\_Price\_Index:** *Float*, índice de precios al consumidor.
- **Unemployment\_Rate:** *Float*, tasa de desempleo.
- **Real\_GDP:** *Float*, PBI real.
- **Mortgage\_Rate:** *Float*, tasa de interés hipotecario.
- **Real\_Disposable\_Income:** *Float*, ingresos reales desechables.
- **StateId:** *String*, abreviación de estado de Estados Unidos.
- **Type:** *String*, tipo de evento meteorológico.
- **Precipitation(in):** *Float*, cantidad en pulgadas de la precipitación en un período de tiempo.
- **Hours:** *Float*, cantidad de horas en que ocurrió el evento meteorológico.
- **Population:** *Integer*, población de un estado, condado o ciudad.
- **CrimePropertyRate:** *Float*, cantidad de crímenes de propiedad reportados cada 100.000 personas.
- **CrimeViolentRate:** *Float*, cantidad de crímenes violentos reportados cada 100.000 personas.