

PROYECTO DE DATA SCIENCE

PREDICCION DE PRECIO DE ARRIENDOS EN LA REGION METROPOLITANA

- Alexis Lema
- Andrés Madariaga
- Daniel Shinya
- Gonzalo Gaete
- Philip Magna

INDICE



Introducción



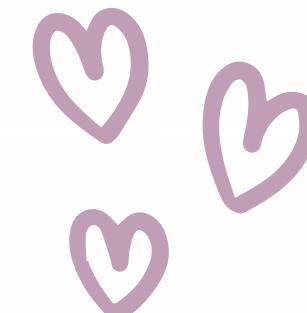
Metodología



Análisis descriptivo



Primer Modelo



INTRODUCCIÓN AL PROYECTO

Este proyecto tiene como **objetivo principal** predecir los valores de arriendo de domicilios en la Región Metropolitana (RM) de Santiago, Chile, considerando diversos parámetros y patrones que afectan el mercado inmobiliario.

A través de la metodología CRISP-DM, se utilizará web scraping y minería de datos para centralizar la información del mercado y proporcionar insights relevantes, tales como el análisis del perfil socioeconómico de las comunas, la viabilidad de arrendar según ingresos, y la proyección del valor futuro de propiedades.

OBJETIVOS

01

Objetivo 01

Predecir el precio mensual de arriendo y el precio UF/m^2 de departamentos en la Región Metropolitana.

02

Objetivo 02

Clasificar las zonas según diferentes factores como precio, perfil socioeconómico, etc.

03

Objetivo 03

Identificar tendencias en la oferta y demanda del mercado inmobiliario.

MOTIVACIONES

Decisión

Toma de decisiones informadas para arrendar según presupuesto y preferencias de ubicación.

Transparencia

Comparar zonas y detectar diferencias de precio por m².

Planificación

Estimar si un arriendo objetivo es alcanzable bajo restricciones.

METODOLOGÍA

Para el desarrollo de este proyecto se eligió el framework CRISP-DM, este framework fue electo debido a sus fases de entendimiento del negocio y deployment que serán importantes para nuestro proyecto

Para la recolección de datos se desarrolló un scraper que busca datos de propiedades en la región metropolitana, posteriormente se sigue con el flujo del framework y se prepararán los datos

Se generan variaciones de precios por cercanía al metro, y se integran datos abiertos para poder crear features con trazabilidad, en base a ello.



FRAMEWORK

CRISP-DM



Business
Understanding



Data
Understanding



Data Preparation



Modeling



Evaluation



Deployment

FRAMEWORK

Business Understanding



PROBLEMA

No existe una estimación directa y comparable del precio de arriendo por zona (RM) según los deseos del usuario.

SOLUCION

Modelo que predice precio (CLP y UF/m²) en zonas según presupuesto, características de vivienda, ubicación (aun no implementado), etc.

FRAMEWORK

Data Understanding



ANALISIS DE LOS DATOS

link	titulo	precio	moneda	superficie_util	superficie_total	dormitorios	banos
https://portalinmobiliario.com/MLC-161	Departamento Vecinal Id: 141768	380000	\$	40 m ²	46 m ²	2	1
https://portalinmobiliario.com/MLC-162	Departamento Santos Dumont Id: 683	360000	\$	35 m ²	40 m ²	1	1
https://portalinmobiliario.com/MLC-162	Departamento San Petersburgo Id: 15	330000	\$	50 m ²	50 m ²	3	1
https://portalinmobiliario.com/MLC-162	Departamento Avenida El Parrón Id: 15	320000	\$	38 m ²	38 m ²	2	1
https://portalinmobiliario.com/MLC-162	Departamento Avenida El Parrón Id: 15	220000	\$	19 m ²	19 m ²	1	1
https://portalinmobiliario.com/MLC-162	Departamento Quinta Avenida Id: 154	225000	\$	48 m ²	52 m ²	2	1
https://portalinmobiliario.com/MLC-163	Nuevo Y Equipado! Briones Luco 1d1l	150000	\$	30 m ²	30 m ²	1	1
https://portalinmobiliario.com/MLC-163	Departamento Abtao Id: 41396	150000	\$	29 m ²	32 m ²	1	1
https://portalinmobiliario.com/MLC-164	Amplio Estilo Mariposa 2d2b Vista Al F	480000	\$	50 m ²	60 m ²	2	2
https://portalinmobiliario.com/MLC-164	Penthouse En Arriendo, Amoblado.	75	UF	160 m ²	200 m ²	3	3
https://portalinmobiliario.com/MLC-164	Arriendo Dpto 3 Dormitorios Con Esta	580000	\$	70 m ²	70 m ²	3	3
https://portalinmobiliario.com/MLC-164	Departamento Con Buenos Accesos Y	49	UF	122 m ²	134 m ²	3	3

DTPM
Directorio de
Transporte Público
Metropolitano



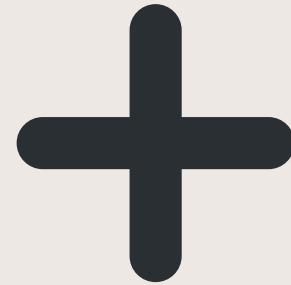
METRO
DE SANTIAGO

FRAMEWORK

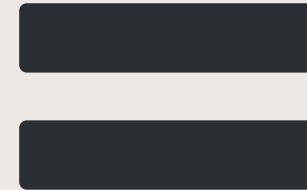
Data Preparation



Listado_arriendos.csv



Detalle_arriendos.csv



Deptos_out.csv

***LINK, TITULO Y MONEDA SOLO TIENEN RELEVANCIA EN EL DATA PREPARATION**

FRAMEWORK

Data Preparation



1

Se eliminan las columnas: link y titulo, ya que solo tienen relevancia al realizar el merge entre los dos dataset entregados por el scraper

2

Se trabaja solo con datos de peso chileno, (solo para esta presentación) de esta forma obtenemos la estimación final en pesos chilenos.

3

Se transforman a variables tipo float las superficies (superficie_util y superficie_total) para trabajar en escala numérica

	precio	moneda	superficie_util	superficie_total	dormitorios
0	380000	\$	40 m ²	46 m ²	
1	360000	\$	35 m ²	40 m ²	
2	330000	\$	50 m ²	50 m ²	
3	320000	\$	38 m ²	38 m ²	
4	220000	\$	19 m ²	19 m ²	
5	225000	\$	48 m ²	52 m ²	
6	150000	\$	30 m ²	30 m ²	
7			29 m ²	32 m ²	
8			50 m ²	60 m ²	
9			60 m ²	70 m ²	
10					

10

FRAMEWORK

Modeling



PRIMER MODELO

- **Modelo:** RandomForestRegressor.
- **Flujo:** imputación simple de features
 - Train/valid (75/25)
 - Métricas en CLP (MAE, RMSE, R²).
- **Inferencia:**
 - “Precio estimado de arriendo para 4 dormitorios, 2 baños, 140 m² útiles”, el modelo arma el vector de entrada (estima superficie total según la relación típica del dataset) y devuelve el precio en CLP.

Métricas validación (\$): MAE=112,588 | RMSE=190,287 | R²=0.803

Consulta (arriendo): 4 dormitorios, 2 baños, 140 m² útil
Precio estimado: ~\$ 1,604,206 CLP

¡MUCHAS GRACIAS!
¿PREGUNTAS?