



Afi

Escuela
de Finanzas

Análisis Multivariante

Máster en Data Science y Big Data en Finanzas

Javier Nogales – PhD Matemáticas

Profesor Titular, Estadística UC3M

www.est.uc3m.es/nogales @fjnogales 2020

Ejercicio Final

- Se proporciona un dataset: HomeSalesData.csv
- Contiene datos de unas 20,000 viviendas, cada una con 21 variables:
 - **id**: identificador para cada casa.
 - **date**: fecha de venta de la casa.
 - **price**: precio de venta de la casa.
 - **bedrooms**: número de habitaciones.
 - **bathrooms**: número de baños (puede que haya .5 baños que representarían baños con váter pero no ducha).
 - **sqft_living**: pies cuadrados del interior de una casa.
 - **sqft_lot**: pies cuadrados de terreno de una casa.
 - **floors**: número de plantas de una casa.
 - **waterfront**: variable dummy que vale 1 si la casa está enfrente del mar.
 - **view**: índice de 0 a 4 indicando como de buenas son las vistas.
 - **condition**: índice de 1 a 5 indicando como de buenas son las condiciones de las casas.
 - **grade**: índice de 1 a 13 indicando el grado de construcción y diseño de la casa (de 1-3 mala construcción y diseño, 7 nivel medio de construcción y diseño y de 11-13 muy buena construcción y diseño).
 - **sqft_above**: los pies cuadrados del espacio de la vivienda interior que está por encima del nivel del suelo.
 - **sqft_basement**: pies cuadrados del sótano de una casa.
 - **yr_built**: año de construcción de la casa.
 - **yr_renovated**: año de renovación de la casa.
 - **zipcode**: código postal de la casa.
 - **lat**: latitud.
 - **long**: longitud.
 - **sqft_living15**: pies cuadrados del interior de las casas de los 15 vecinos más próximos.
 - **sqft_lot15**: pies cuadrados del terreno de las casas de los 15 vecinos más próximos.

Ejercicio Final

- **Objetivo principal:** aplicar herramientas multivariantes: descriptivas, PCA y FA
- Tareas concretas:
 - ➊ Eliminar las variables id y date
 - ➋ Eliminar la variable price, aunque se puede conservar en memoria para interpretar mejor el output de PCA y FA
 - ➌ Preprocesado y feature engineering: 1 punto
 - ➍ Añadir análisis descriptivo multivariante (numérico y gráfico) previo a las herramientas: 2 puntos
 - ➎ Aplicar herramienta PCA e interpretar: 2 puntos
 - ➏ Aplicar herramienta FA e interpretar: 2 puntos
 - ➐ Conclusiones (explicar conocimiento ganado, diferencias entre PCA y FA, etc.): 2 puntos
 - ➑ Report (si es markdown mejor): 1 punto
- Subir al Campus Virtual el código y un pequeño informe (pdf) de no más de 10 páginas



Afi

Escuela
de Finanzas

© 2015 Afi Escuela de Finanzas. Todos los derechos reservados.