



EMGVD (EYP3707)

Profesor: Pedro Luiz Ramos

Ayudante: Diego Bernal

Ayudantía 3

Primer Semestre - 2025

Enunciado I

El emprendimiento *Botillería aérea* esta interesado en colocar un puesto en la comuna de las condes con el fin de despachar la *alegría embotellada* a todos los hogares de la comuna. Para poder competir con las grandes compañías que dominan este rubro, se planteo la estrategia de estudiar el comportamiento de los competidores en el año 2024.

Su amiga Laura la cual tiene experiencias colocando darkstores en las comunas opulentas, le comparte una base de datos con todas las ventas por tipo de *bebida alcohólica* en la comuna, para este primer estudio usted se enfocara en las ventas de cerveza.

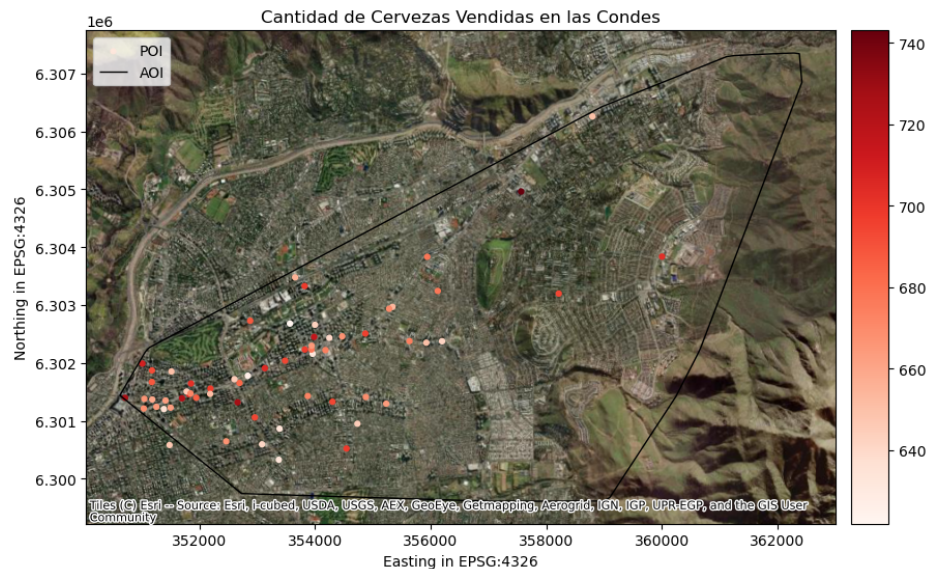


Figure 1: Mapa de ventas por tienda

Para esta gran tarea de identificar posibles zonas potenciales en la cuales instalar su negocio, antes de aplicar el modelo de optimización, debemos estudiar las relaciones entre nuestras variables.

Para ello usted aplica una grilla de 30^2mts con el fin de identificar cada zona potencial, después aplica una teselación de Voronoi para pasar los valores de las medias de las ventas para el 2024 por tienda a cada raster.

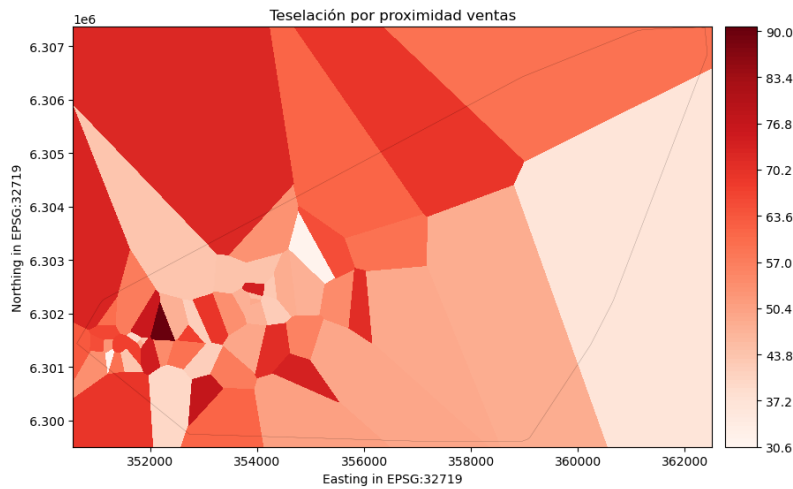


Figure 2: Teselación de Voronoi ventas promedio

1. Realice estadística descriptiva y comente sobre las relaciones entre los datos, serán lineales?.
2. Utilice la correlación de *Pearson* y plotee la matriz de correlación para observar si tenemos problemas de multicolinealidad.
3. Clasifiquemos la zona antes para ver si hay grupos utilice una metodología no supervisada para zonificar (knn).
4. Si las clases están desproporcionadas realice una estratificación, separe el data set en una proporción de 80/20, realice CV con 10 folds.
5. Pruebe correr el modelo completo con un modelo de regresión múltiple, tendremos un problema de estimadores robustos por la multicolinealidad, pruebelo.
6. Corra los modelos ahora con los algoritmos de subset-selection que están en el archivo *models.py*.
7. Ahora corra los modelos con los algoritmos *Ridge & Lasso*
8. Realice gráficos para estos dos últimos modelos en relación a sus parámetros con cada variable, y el RMSE or MSE.
9. plotee un gráfico con las teselaciones para cada modelo y los errores de cada modelo por raster como variable continua. Comente