Montpellier Business School

Team project task

Note: Please answer all questions and hand in source codes and the results

(10 pts) **Q1. Descriptive analysis of data is essential for business analytics**

In France, metropole is composed of a city and its neighboring towns. Please work on the dataset, Bordeaux_Metropole.csv.

Bordeaux_Metropole.csv contains the following columns:

- date_mutation: date of transaction
- valeure_fonciere: property value
- code_postal: postal code
- section: section
- type_local: type of property (apartment or house)
- surface_relle_bati: size of property (in m^2)
- number_pieces_principales: number of rooms
- surface_terrain: size of land (in m^2)

(A) If a client is interested in apartments in postal codes, 33000, 33110 and 33440, and he wants to know historical transaction price for these three areas. Please export descriptive statistics of price and size of **apartments** (i.e., mean, std, min, max, median, Q25, Q75) for these three postal codes as a csv file (e.g., 33000.csv, 33110.csv and 33440.csv) in Bordeaux Metropole.

(10 pts) **Q2: Data analysis and multivariate linear regression model.**

Work with the real estate dataset, realestate.xlsx or realestate.csv

- Variables – numerical attributes
  - SalePrice: sale price (in thousand)
  - Size: size of the real estate (in square feet)
  - Beds: number of bedrooms
  - Baths: number of bathrooms
  - Num_Garage: number of garages
  - Year: when the real estate was built
- Variables: binary attributes
  - Highway: accessible to the highway in 10 minutes, yes or no
  - Aircondition: with or without air conditioner
  - Swimmingpool: with or without swimming pool

For question A to B, please hand in all source codes and the result.

A. (2 pts) You are asked to predict **Sale Price**, what are your reasonable choices from the dataset? Please report the total amounts of reasonable choices.

B. (8 pts) Continued with (A). Please use python code to list express all your reasonable choices in regression formula.