T5 Boot camp

Data science

Project proposal

Estate's dataset

mohammad alajmi

mohammad almalki

Instructor:

Mejdal Al-Qahtani

Question/problem statement:

An Estate agent asked to make a system that predict the prices of estates to help the agent to distinguish between the overpriced estate and other way around, as the estate agent said that the data could be collected from different websites, , the estate agent specified some neighborhood is not desired so it should be disqualified from the model due to the low accuracy that it has, the agent also asked to study the data and visualize it to help him to have better understanding about estate market.

In this project we are willing to answer these following questions:

1. What is the best model to build a model that predict the estate market?

2. To study the market and to train the model predict the prices

3. To study the market and to test the model predict the prices and to check for it is accuracy

The Data set:

The used data set in this project is derived from different websites, the data collected using web scraping library called beautiful soap, it is about the prices of estates and other information about the estate, the data has about five thousand rows before the cleaning and about 7 columns.
The columns describe a Avg. Income of residents of the city house is in, Avg Age of Houses in same city, Avg Number of Rooms for Houses in same city.
The columns also describe Avg Number of Bedrooms for Houses in same city, Population of city house is in, price of the Estate and Address for the house.

The used tools:

I have used different tools to achieve this goal such as:

1- Python programming language
2- Numby library
3- Pandas' library
4- Matplotlib library
5- SeaBorn
6- Sklearn
7- BeautifulSoup
8- Request library

The MVP:

The main goal of this project is to study the Estates market and to train a model to predict the prices of the estates of and testing it using linear regression models.