**HOMEWORK 1**

| Name | *Huỳnh Quốc Việt* |
| --- | --- |
| Student ID | *SE194225* |

**Note:**

* To submit, name the file consisting your answers in the following format:

[YourName]\_[Your StudentID]\_Homework1

Ex: NguyenVanA\_123456\_Homework1

In this homework, your task is to load Ames Housing Data given by your instructor and answer following question

**Question 1 (2pt)**

What is the min and max sale price? Show evidence (code and results from notebook or console)

|  |
| --- |

**Question 2: (2pt)**

What are the total missing value given by the dataset? Show evidence (code and results from notebook or console)

|  |
| --- |

**Question 3: 2(pt)**

* To address potential outliers, filter the dataset to include only entries where the 'Gr Liv Area' is less than or equal to 4000.
* Next, generate polynomial features (at the second degree) and interaction terms for the 'Lot Area' and 'Overall Qual' columns using the PolynomialFeatures function as demonstrated below:
  + pf = PolynomialFeatures(degree=2)

**Your task is to**:

1. Display images of top 5 elements and last 5 elements of resulting dataframe (dataframe after transform using PolynomialFeatures function)
2. What are the "Lot Area Overall Qual" values for IDs 0 and 2924?

|  |
| --- |

**Question 4: (3pt)**

Read load the dataset. Don’t use threshold in question 3 anymore.

To know number of outlier of dataset, you must do as follow:

1. Find the first quartile (Q1), which is the value below which 25% of the data falls.
2. Find the third quartile (Q3), which is the value below which 75% of the data falls.
3. Calculate the IQR = Q3 – Q1
4. Find lower bound and upper bound based on the following formula

Lower = Q1 – 1.5\* IQR

Upper = Q3 +1.5\*IQR

**Your task is to answer the following question:**

* What is the value of *lower bound* and *upper bound* of Column “Gr Liv Area” of the given dataset. Show evidence (code and results from notebook or console)

|  |
| --- |

**Question 4: (1pt)**

How many number of outliers according to lower bound and upper bound value you detected in Question 4

|  |
| --- |