SIT 720 - Machine Learning

Lecturer: Chandan Karmakar | karmakar@deakin.edu.au

School of Information Technology,
Deakin University, VIC 3125, Australia.

▾ Assessment Task 1 (20 marks)

Submission Instruction

1. Student should insert Python code or text responses into the cell followed by the question.

2. For answers regarding discussion or explanation, **maximum five sentences are suggested**.

3. Rename this notebook file appending your student ID. For example, for student ID 1234, the submitted file name should be A0_1234.ipynb.

4. Insert your student ID and name in the following cell.


```
# Student ID:

# Student name:
```


Background

Selection of housing is always difficult for someone seeking for a suitable one as it includes various factors and preferences. People prefer to buy a house considering many criteria like-population, quality of life, financial capability, as well as social and natural environments around the housing block. In this assignment you will be helping people choosing suitable housing for them according to their needs.

Dataset

Dataset file name: housing_dataset.csv

Dataset description: Dataset contains total 10 features (columns). It contains the location, housing age, population, number of families in a housing (block), number of rooms, average income of the families in that housing, ocean proximity and other informaiton. Each row indicates a record of a housing block containing the features mentioned earlier.

Features:

1. latitude (float): Latitude of the location of a housing in conventional geospace

2. longitude (float): Longitude of the location of a housing in conventional geospace
3. housing_age (int): Age of the housing in year, the higher number indicates the older housing
4. total_rooms (int): Total number of rooms in a housing
5. total_bedrooms (int): Total number of bedrooms in a housing
6. population (int): Total population of a housing
7. families (int): Total number of families living in a housing
8. average_income (float): Average income of the member of a housing in a scale of Tousand Dollar Per Month
9. ocean_proximity (string): Describing how close the housing is to the ocean
10. house_value (int): Average individual house price of a housing in Dollers

▾ Part-1: Basic Calculations: *(8 marks: 8 questions x 1 marks each)*

1. Find the distances of the farthest and nearest housing blocks from the house block described in the first row of the dataset.

```
# INSERT your code here.
```

2. Calculate the average age of the house blocks near the ocean.

```
# INSERT your code here.
```

3. Find the income of the housing block with the most and least population density (per family).

```
# INSERT your code here.
```

4. Calculate the price difference between the latest and oldest housing block from the dataset.

```
# INSERT your code here.
```

5. Calculate the cheapest price per room from the dataset.

```
# INSERT your code here.
```

6. Calculate the population density (per family) for the most and least wealthy housing

```
# INSERT your code here.
```

7. Calculate and print the total housing blocks located in the same place.

```
# INSERT your code here.
```

8. Calculate the price of expensive room grouped by ocean proximity.

```
# INSERT your code here.
```

▾ Part-2: Visualization: *(6 marks: 3 question x 2 marks each)*

1. Draw the population scatter plot against housing age and another against ocean proximity. From the graph conclude an assumption.

```
# INSERT your code here.
```

2. Draw a bar diagram of average values of all suitable columns. (excluding latitude, longitude and ocean proximity of course).

```
# INSERT your code here.
```

3. Visualize the differences in housing prices from the average price of housing using a bar diagram.

```
# INSERT your code here.
```

▾ Part-3: File Management: *(6 marks: 2 question x 3 marks each)*

1. Save the details of all housing blocks in a csv file having houses near oceans and lower than the average of the housing value.

```
# INSERT your code here.
```

2. Create a new housing dataset (a csv file) having only the location, total rooms and housing price information.

```python
# INSERT your code here.
```