DATA-51100: Statistical Programming
Programming Assignment 5 – Data Preparations and Statistics

Introduction
The file `cps.csv` (attached) contains school profile information for Chicago Public Schools. Your program will derive some data from it and then generate some statistical information.

Requirements
You are to create a program in Python that performs the following:

1. Loads the cps.csv file (assume it's in the current directory) and create a DataFrame object from it.

2. Based on the data contained in the cps.csv file, generates a dataframe with the following information:
    a. School_ID
    b. Short_Name
    c. Is_High_School
    d. Zip
    e. Student_Count_Total
    f. College_Enrollment_Rate_School
    g. Lowest Grade Offered (derived from Grades_Offered_All column)
    h. Highest Grade Offered (derived from Grades_Offered_All column)
    i. Starting Hour (derived from School_Hours column)

    The values for a-f are based on existing columns in the data. For g-i, you will need to generate new columns which derives information from existing ones.

    Replace the missing numeric values with the mean for that column.
    Display the first 10 rows of this dataframe.

3. Displays the following information:
    a. Mean and standard deviation of College Enrollment Rate for High Schools
    b. Mean and standard deviation of Student_Count_Total for non-High Schools
    c. Distribution of starting hours for all schools
    d. Number of schools outside of the Loop Neighborhood (i.e., outside of zip codes 60601, 60602, 60603, 60604, 60605, 60606, 60607, and 60616)

Additional Requirements

1. The name of your source code file should be `DataStats.py`. All your code should be within a single file.
2. You need to use the pandas DataFrame object for storing data.
3. Your code should follow good coding practices, including good use of whitespace and use of both inline and block comments.
4. You need to use meaningful identifier names that conform to standard naming conventions.
5. At the top of each file, you need to put in a block comment with the following information: your name, date, course name, semester, and assignment name.

What to Turn In

You will turn in the **single DataStats.py file** as well as **a screenshot of your output(s)** using BlackBoard.

Sample Program Output

```
DATA 51100 - [Semester] [Year]
Name: [PUT YOUR NAME HERE]
PROGRAMMING ASSIGNMENT #5
```

	School_ID	Short_Name	Is_High_School	Zip	Student_Count_Total	College_Enrollment_Rate_School	Lowest_Grade	Highest_Grade	School_Start_Hour
0	609952	GREENE	False	60609	415	58.084302	PK	5	8
1	609869	LANGFORD	False	60636	241	58.084302	PK	8	8
2	609896	DRUMMOND	False	60622	346	58.084302	PK	8	8
3	610590	BRONZEVILLE CLASSICAL	False	60609	91	58.084302	K	2	7
4	610087	BLAIR	False	60638	248	58.084302	PK	2	7
5	610503	FRAZIER PROSPECTIVE	False	60624	198	58.084302	K	8	7
6	400164	INSTITUTO - LOZANO HS	True	60608	78	21.900000	9	12	8
7	610059	MAYER	False	60614	760	58.084302	PK	8	8
8	610206	TWAIN	False	60638	1094	58.084302	PK	8	8
9	609872	PEREZ	False	60608	318	58.084302	PK	8	8

```
College Enrollment  Rate for High Schools = 58.08 (sd=25.07)

Total Student Count for non-High Schools = 521.55 (sd=268.64)

Distribution of Starting Hours:
 8am:  415
 7am:  193
 9am:   40

Number of schools outside Loop: 634
```