

 question.md

## Assignment 2

### Question

Find correlations in the given data:

1. To see the relation between literacy rate among females and GFR (GFR : General Fertility Rate)

### Data

- *fertility-rate.csv* : GFR values(total, rural, urban) for some states and UTs
- *literacy-rate.csv* : Literacy rate for states and UTs, given by region(rural/urban/general) and gender(male/female/total)

### Code Snippets

This exercise can be done in many different ways (ex. using modules like numpy, scipy, or even simple lists). The following code snippets are for the dataframe approach, i.e., data from csv files is loaded into dataframes, merged, then correlation is calculated.

- To load a CSV to a dataframe

```
import pandas as pd
df = pd.read_csv('filename.csv')
```

- To merge two dataframes by a column

```
merged_df = pd.merge(df1, df2, on=['column_name'], how='inner')
```

- To find correlation between two columns of a dataframe

```
column1 = df["column1_name"]
column2 = df["column2_name"]
correlation = column1.corr(column2)
```

### Helpful Links

- [How to find the correlation between two Pandas DataFrame columns in Python](#)
- [How do I read CSV data into a record array in NumPy?](#)