**COURSERA CAPSTONE**

**IBM Applied Data Science Capstone**

***Opening a new Hospital in Delhi, India***

**PART - 1**



**By: Yogesh Singhal**

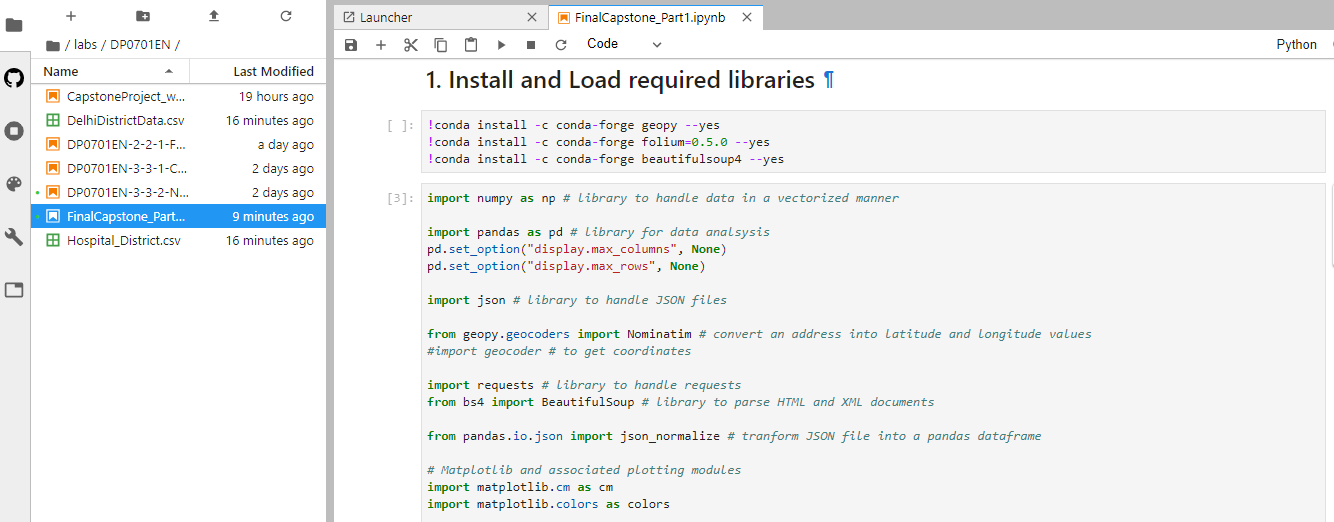
**April 2020**

**DATA DESCRIPTION**

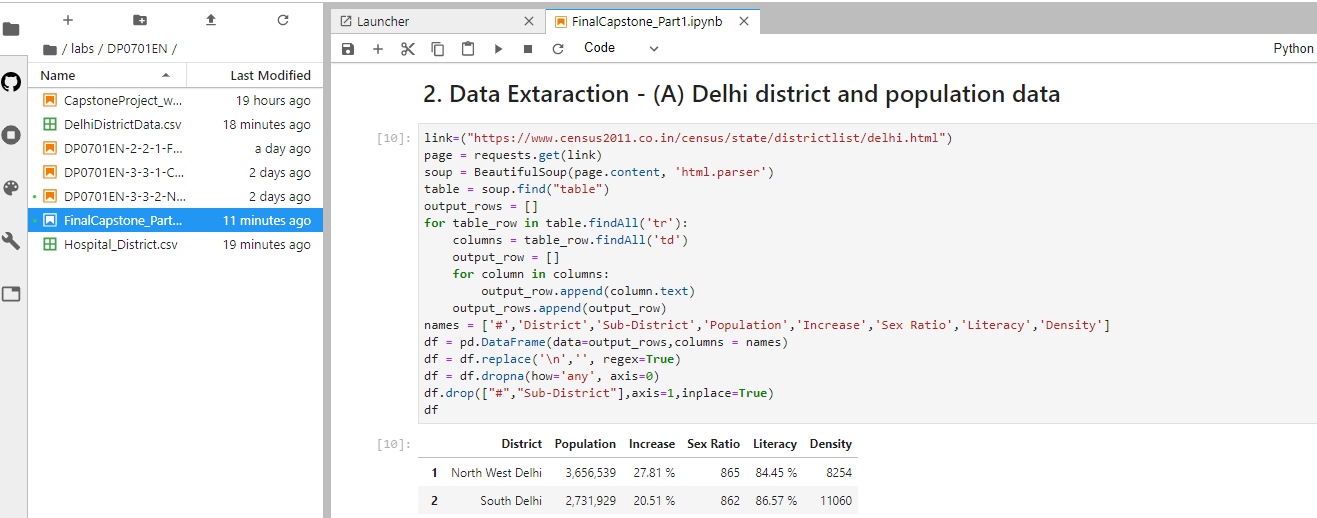
Data is prepared using:

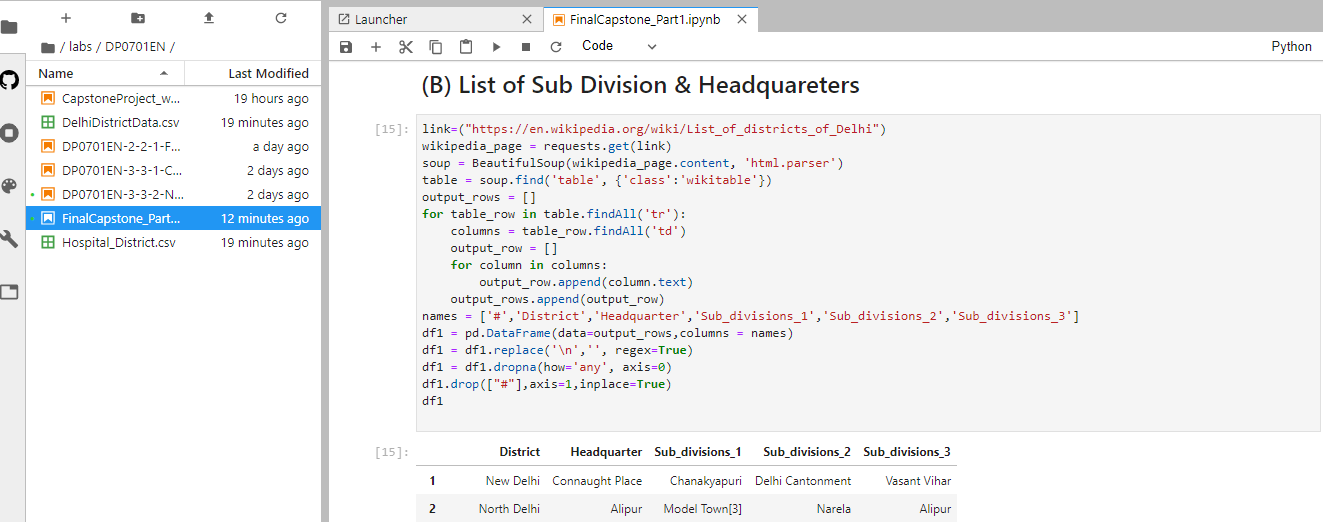
1. Data Extraction
2. Web Scraping from various sources.
3. Data Downloading from government website.
4. Data Preparation
5. Data Cleaning
6. Data merging
7. Data Definition
8. **Data Extraction**

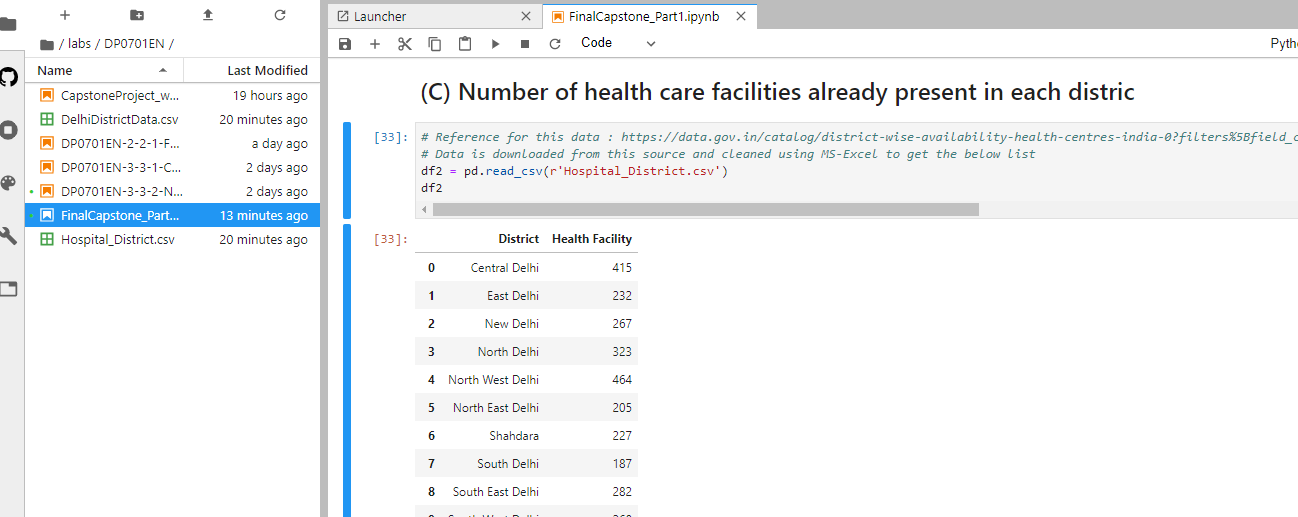
**Load the require libraries to start data extraction.**

****

**Data is extracted from various sources to get the district level information and number of hospitals in each district.**

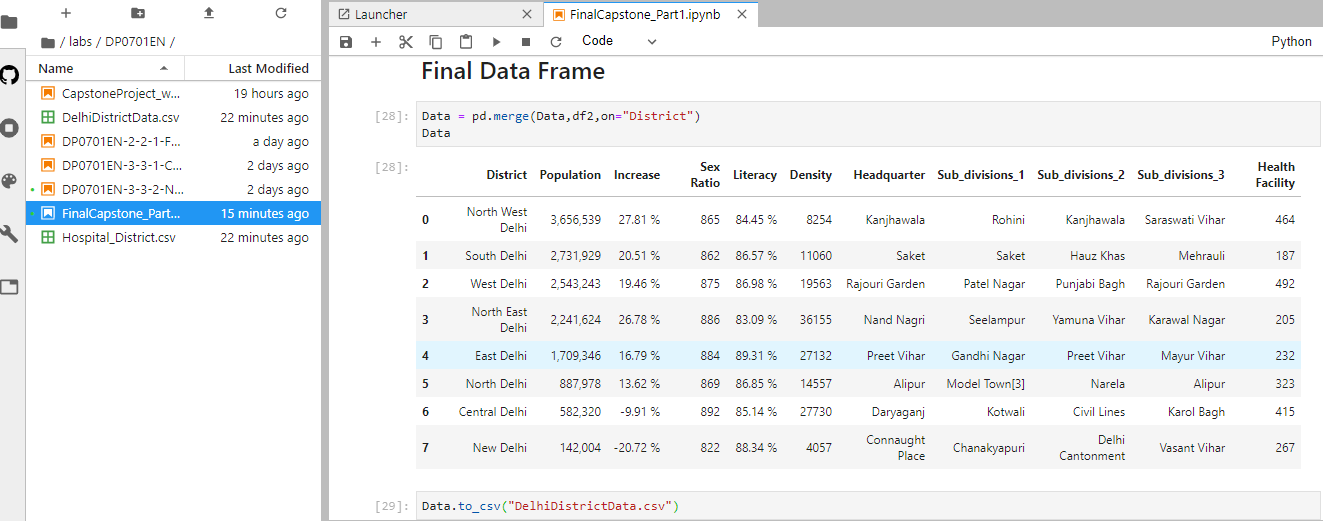
****

****

****

1. **Data Preparation**

**All the data is extracted now the a common data frame is formed to get the final required data which will be further processed to solve the business problem.**

****

1. **Data Definition**

The final Data frame contains 11 columns:

1. District: Delhi is divided into various districts; this column represents the districts in Delhi.
2. Population: Each district populations is represented in this column.
3. Increase: Increase in population since last decade.
4. Sex Ration: Female, Male ratio is represented in this column.
5. Literacy: represents literacy rate in the respective districts.
6. Density: People density in respective districts.
7. Headquarter: Headquarter of the respective districts lies in these areas.
8. Subdivision: Divisions under each district.
9. Health Facility: Number of hospitals under each district.