**COURSERA CAPSTONE**

**IBM Applied Data Science Capstone**

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

**PART - 1**



**By: Tanvi Jain**

**February 2020**

**INTRODUCTION**

The detail analysis of Population Census 2011 published by Govt. of India for Delhi state reveal that population of Delhi has increased by 21.21% in this decade compared (2001-2011) to past decade (1991-2001). The density of Delhi state in the current decade is 29298 per sq mile.

* Delhi is a State of India with population of Approximate 1.68 Crores.
* The population of Delhi state is 16,787,941.
* The density of Delhi state is 11,320 per sq km.
* Delhi State is spread over 1,483 Sq Km.

Health sector infrastructure in Delhi comprises of 1,298 dispensaries, 1160 nursing homes, 230 maternity homes,178 polyclinics/special clinics, 88 hospitals and 17 medical colleges. Important vital indicators like Infant Mortality Rate, Neo-Natal Mortality Rate, Under Five Mortality Rate in respect of Delhi stand at lower levels i.e 18, 12 and 22 respectively in comparison to all-India levels at 34, 24 and 39 respectively. The total fertility rate of Delhi is 1.6 which is lowering among all states in India. Per capita expenditure on health in Delhi has increased from Rs 2116 in the year 2014-15 to Rs 2493 in 2017-18. The total expenditure in medical and public health sector of Delhi government has significantly increased from Rs 861.66 crore which accounts to 9.85 per cent of the total expenditure in 2007-08 to Rs 1912.42 crore which accounts to 13.28 per cent of the total outlay in 2017-18. To provide primary health care services at the doorstep of the citizens of Delhi, 189 Aam Aadmi Mohalla clinics have already been set up, it added.

The Delhi government has decided to remodel around 16 existing hospitals so as to enhance the number of existing beds as per Floor Area Ratio (FAR) norms. Around 7,000 new beds will be added as per planned remodelling of these existing 16 hospitals, it said.

**Business Problem**

The objective of this project is to take part of this remodelling process therefore, the proposal to make a new hospital having 500 beds. For this, the need is to analyse and select the best location in Delhi to open a new hospital. Using Data Science methodology and machine learning techniques as clustering, regression etc., this project aims to answer the business problem: In Delhi, India if government wants to open a hospital, where would the data science team will recommend building it?

**Target Audience of the Project**

This hospital will help to overcome the shortage of beds per 1000 population. Currently, the total number of hospital beds in the national capital increased from 48,096 in 2014-15 to 57,194 in the last fiscal, according to Delhi's Economic Survey report. This translated in a jump in beds per 1000 population from 2.68 to 2.99 in the corresponding periods. Delhi government need to further improve his number and provide the best health practices to its people and decline the death rate due to poor health care facilities in the country.

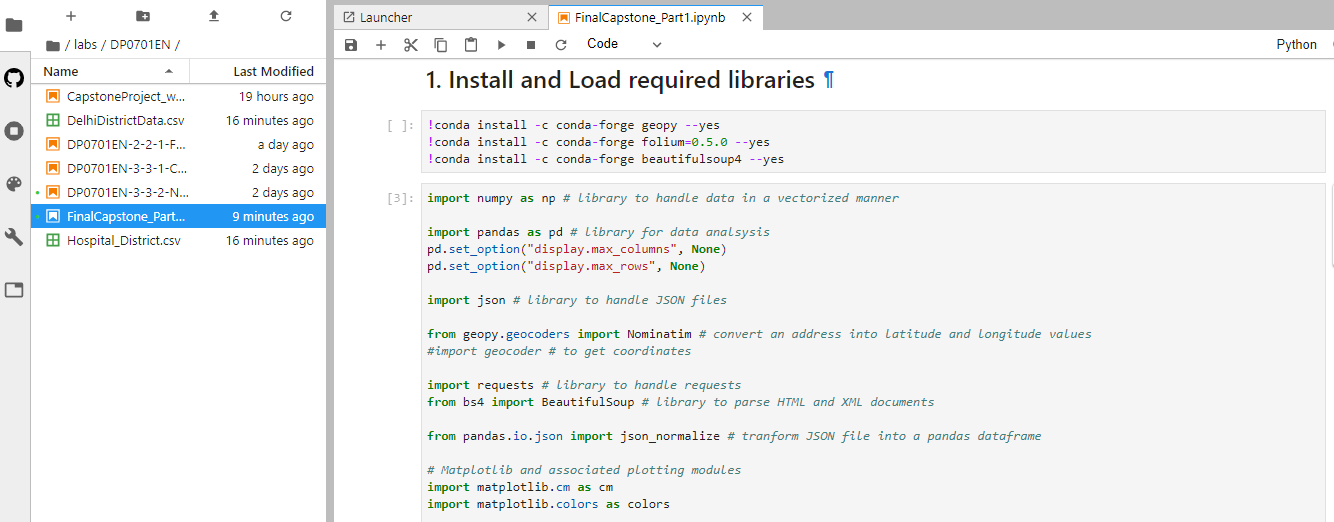
**DATA DESCRIPTION**

Data is prepared using:

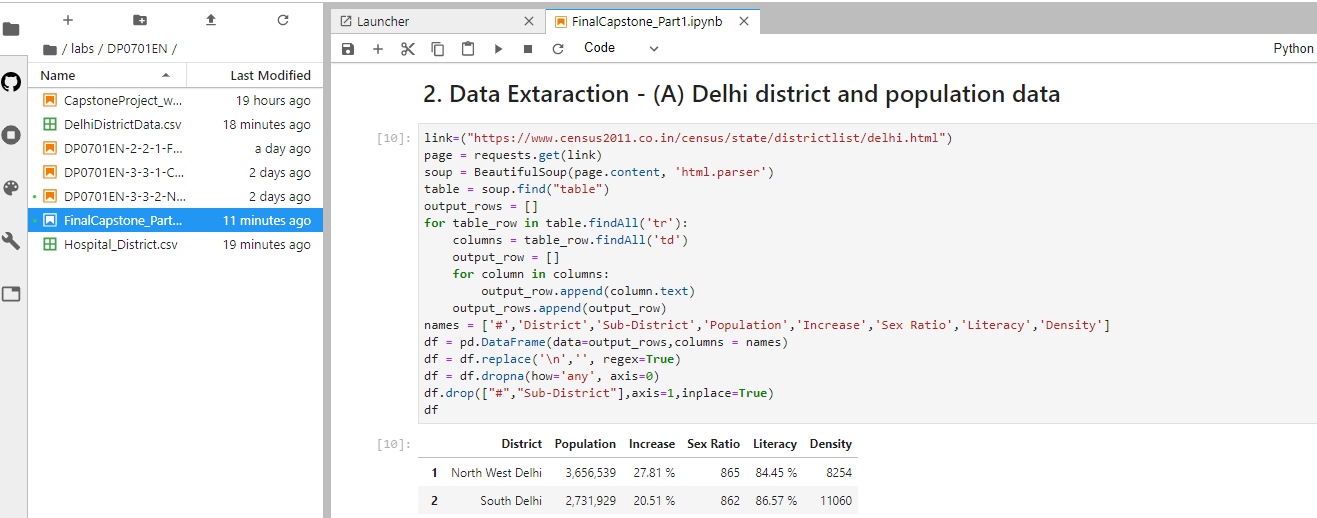
1. Data Extraction
2. Web Scraping from various sources.
3. Data Downloading from government website.
4. Data Preparation
5. Data Cleaning
6. Data merging
7. Data Definition

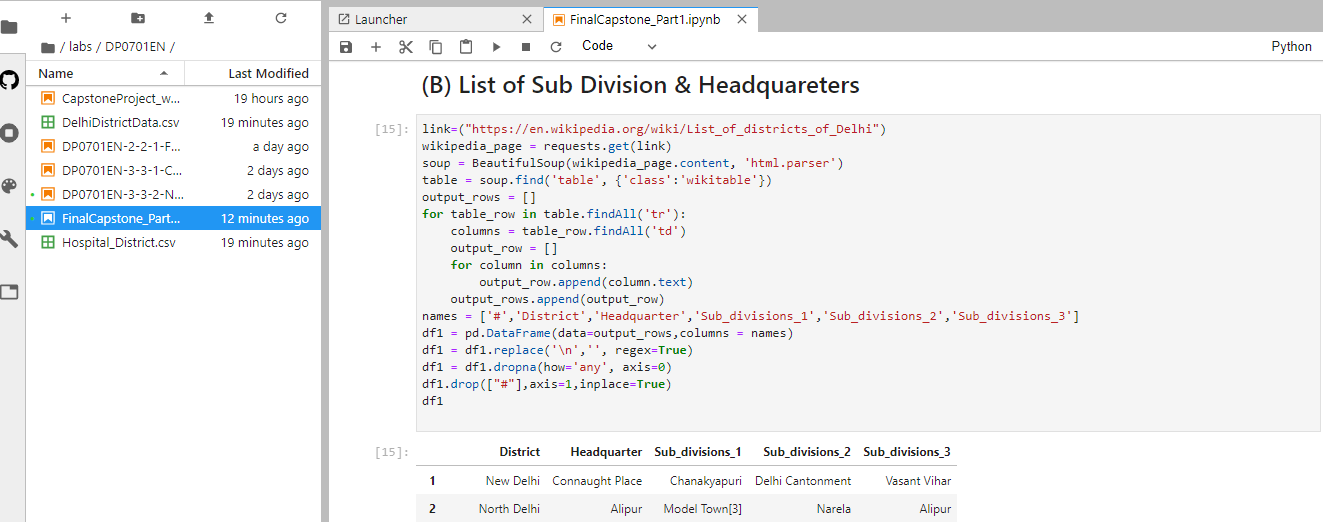
1. **Data Extraction**

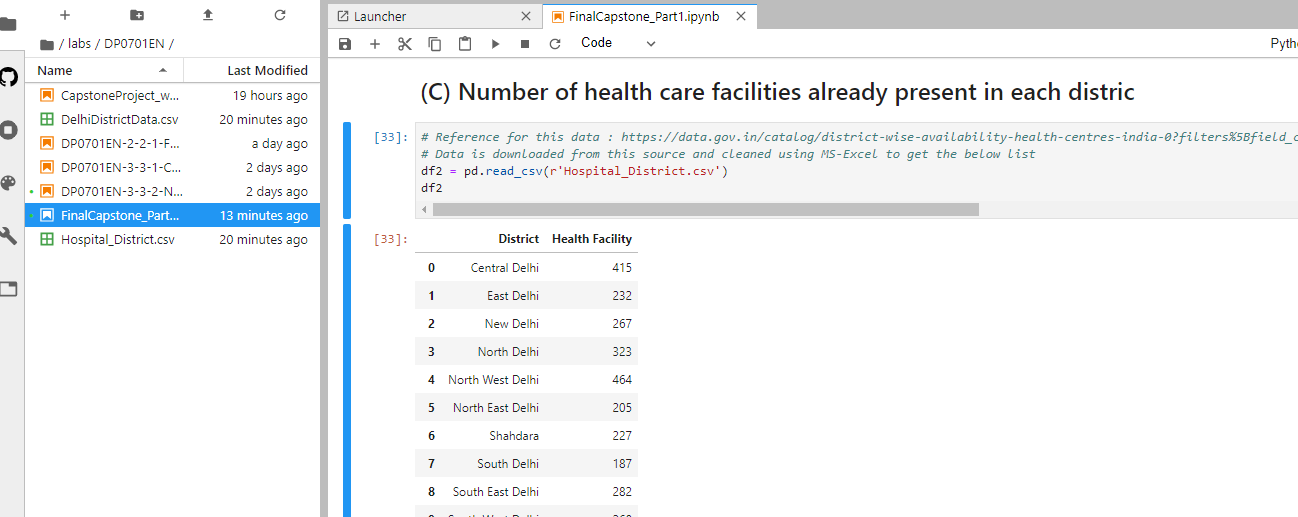
**Load the require libraries to start data extraction.**

****

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

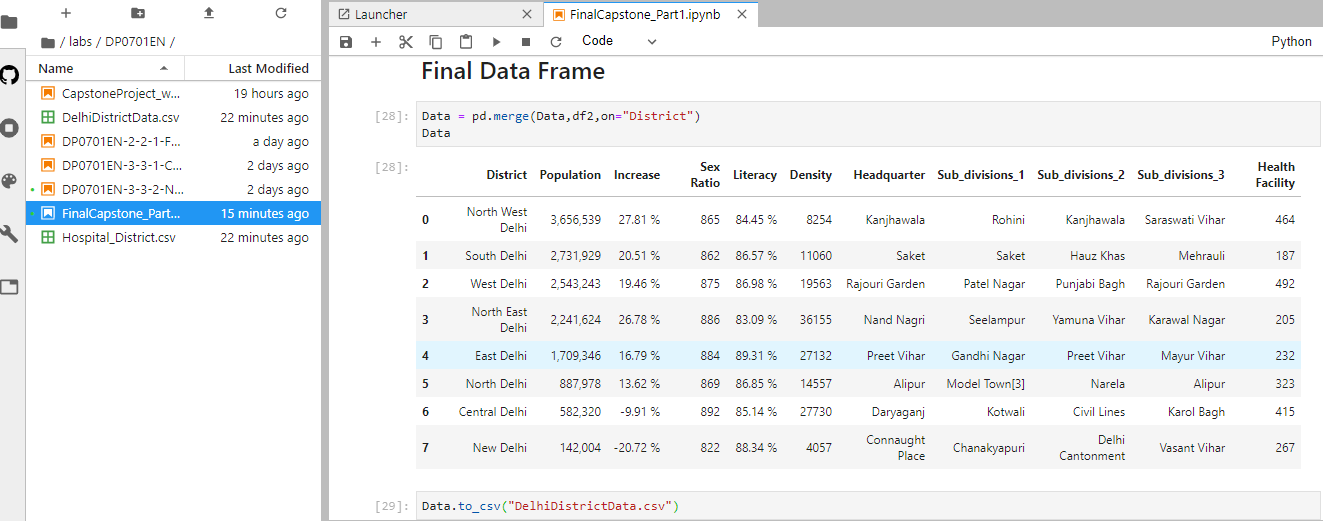
****

****

****

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.