**PROJECT REPORT**

**CSE-413**

**BIG DATA & IOT LAB**

**Submitted to:**

**Tahmina Sultana Priya**

**Lecturer,Department of CSE**

**Daffodil International University**

**Submitted by:**

**Md.Jannatul Ferdous (191-15-2497)**

**Mahin Ahmed (191-15-1029)**

**Saiful Islam Sagor (191-15-2678)**

**Md.Omer Faruk Tusher (192-15-13122)**

**Section: PC-A**

**Department of CSE**

**Daffodil International University**

**Submission Date: 06/12/2022**

**Abstract:**

The purpose of our project is to predict the populations growth-rate of different countries over the world. Global population growth is a challenging factor for the human race.

Distributed data processing platforms for cloud computing are important tools for large-scale data analytics. Apache Hadoop MapReduce has become the de facto standard in this space, though its programming interface is relatively low-level, requiring many implementation steps even for simple analysis tasks.

The main aim of this project is to analyze and predict the massive amount of data (world population), with the help of various types of tools such as apache spark which is used for real-time processing and analysis of large amounts of data.

**Introduction:**

The annual average rate of change of population size, for a given country, territory, or geographic area, during a specified period. It expresses the ratio between the annual increase in the population size and the total population for that year, usually multiplied by 100.

Understanding population growth is important for predicting, managing, monitoring, and eradicating pest and disease outbreaks.

**Objectives:**

The main objectives of this project are:

* To predict the number of population and growth rate of 2023.

**Tools has been used:**

* **Colab**

**Language has been used:**

* **Python**

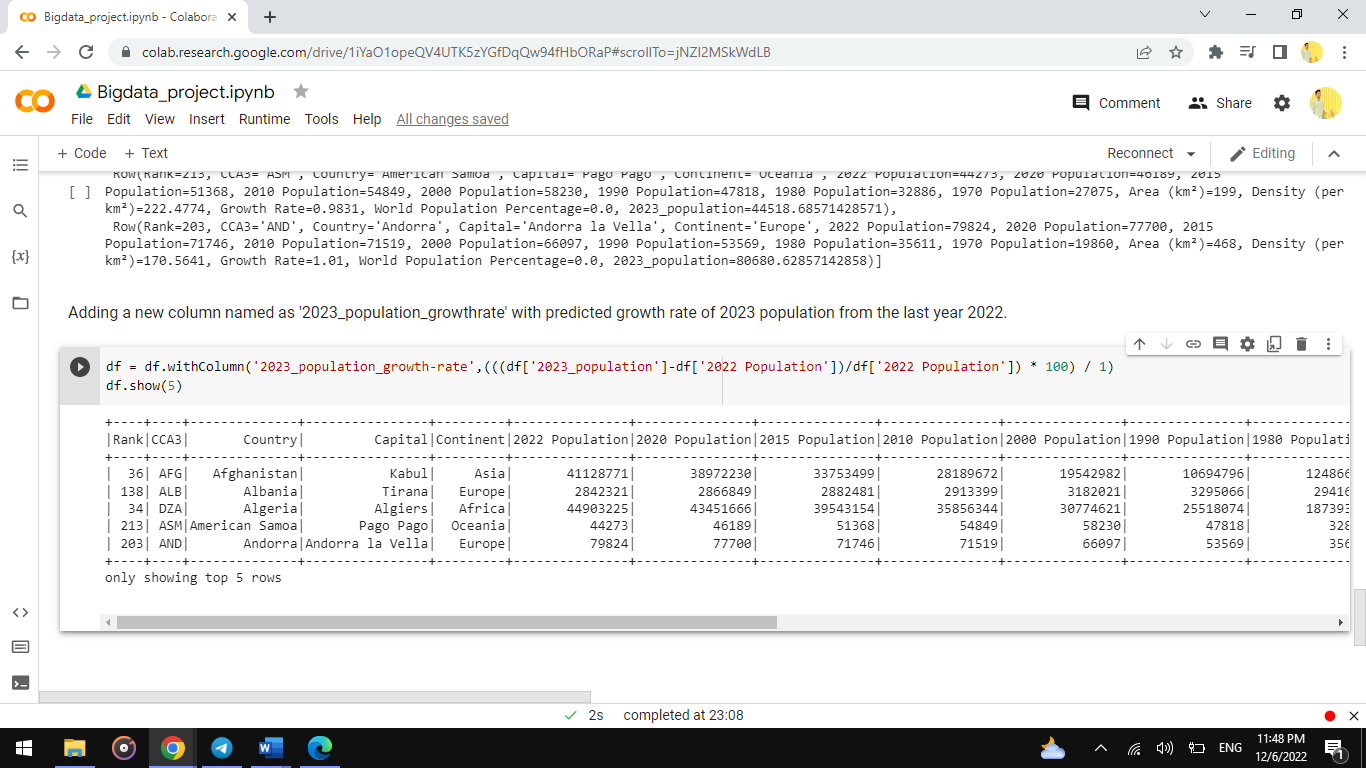
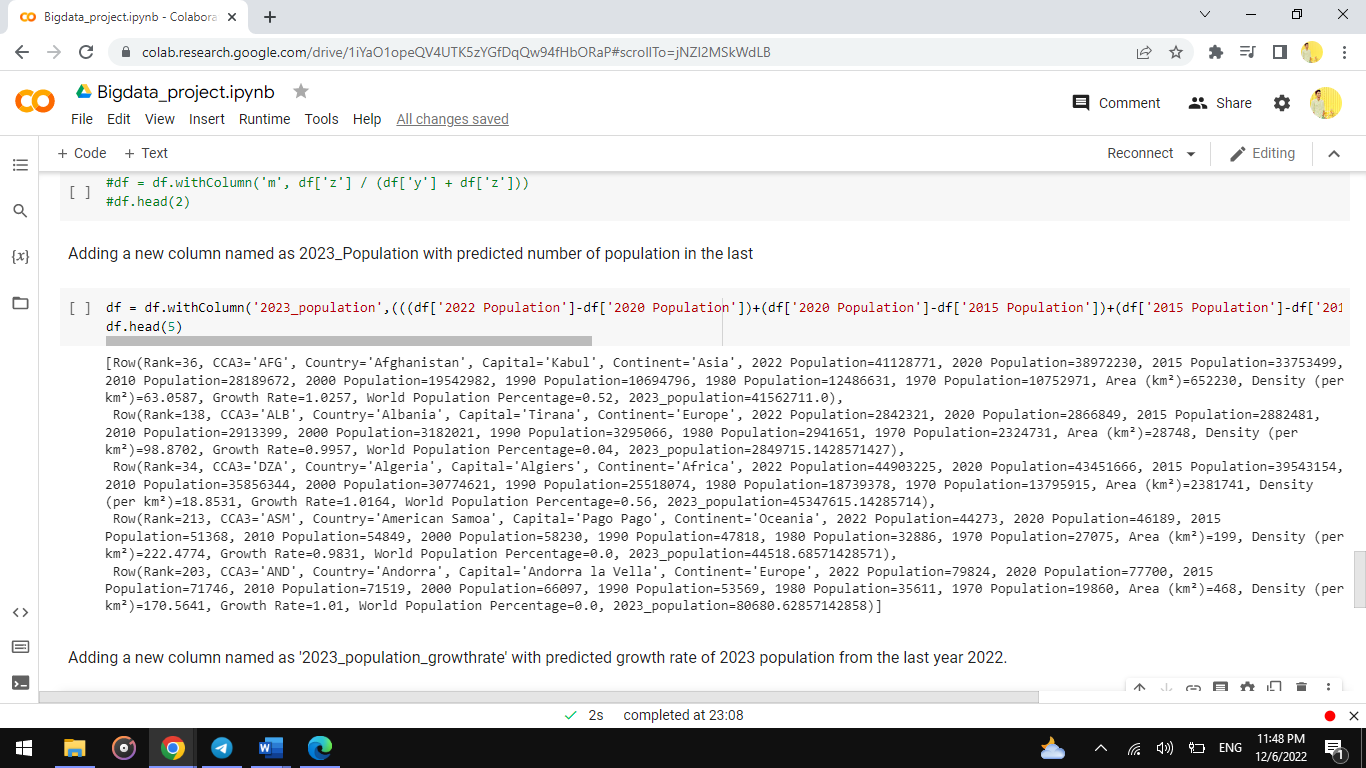
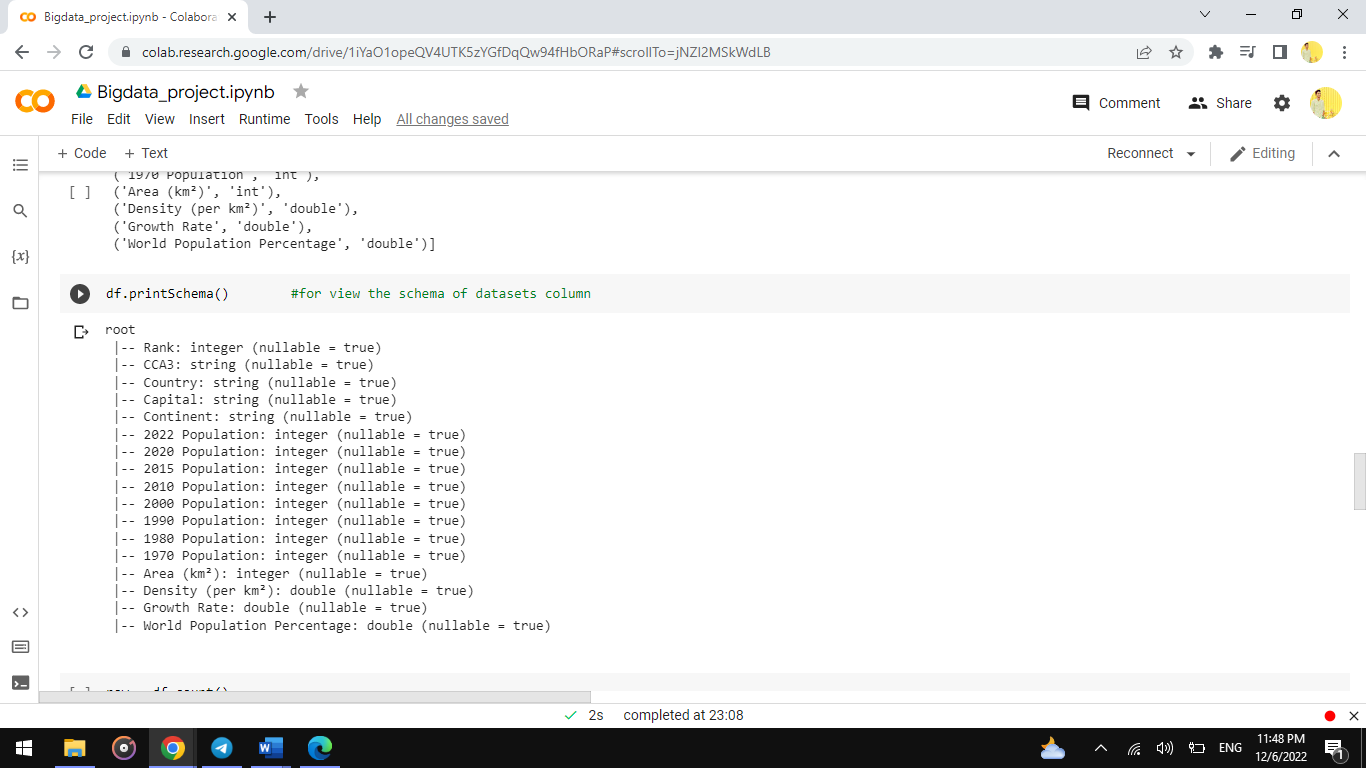
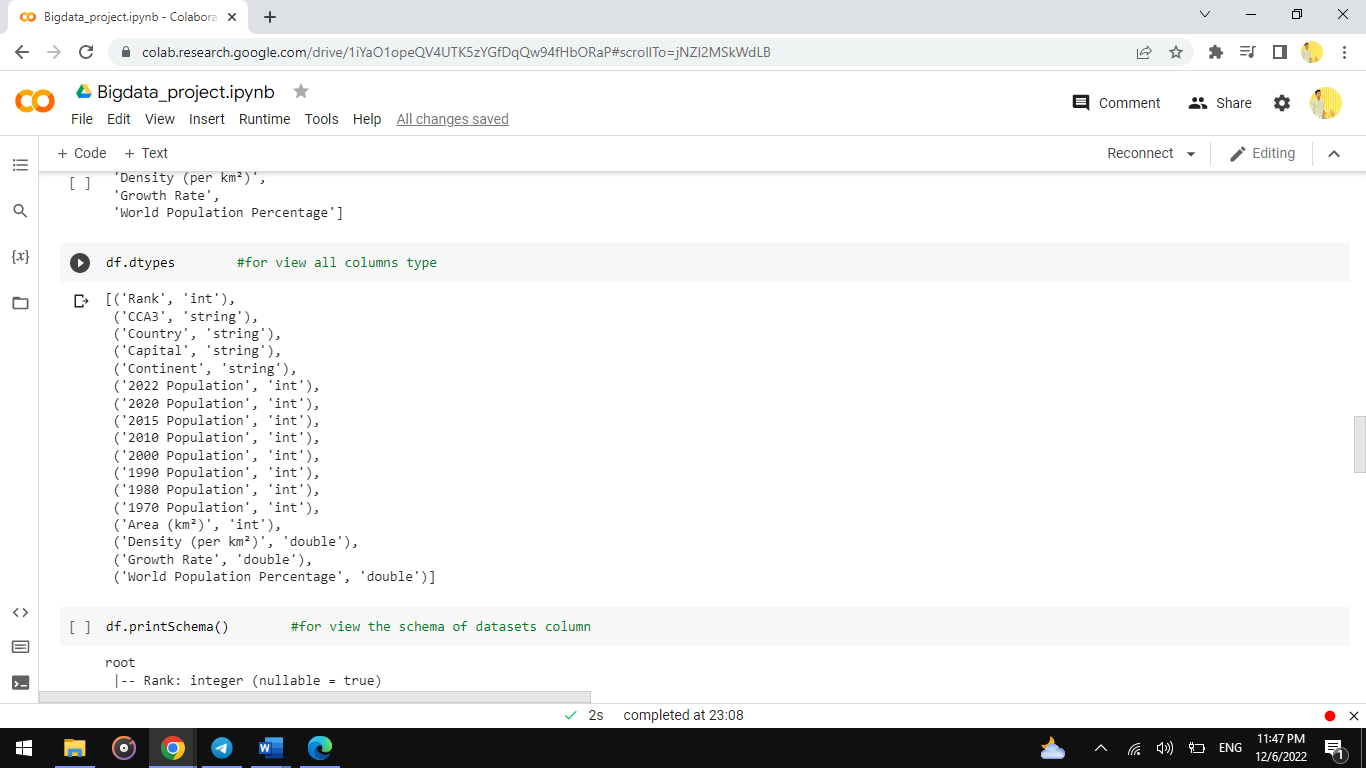
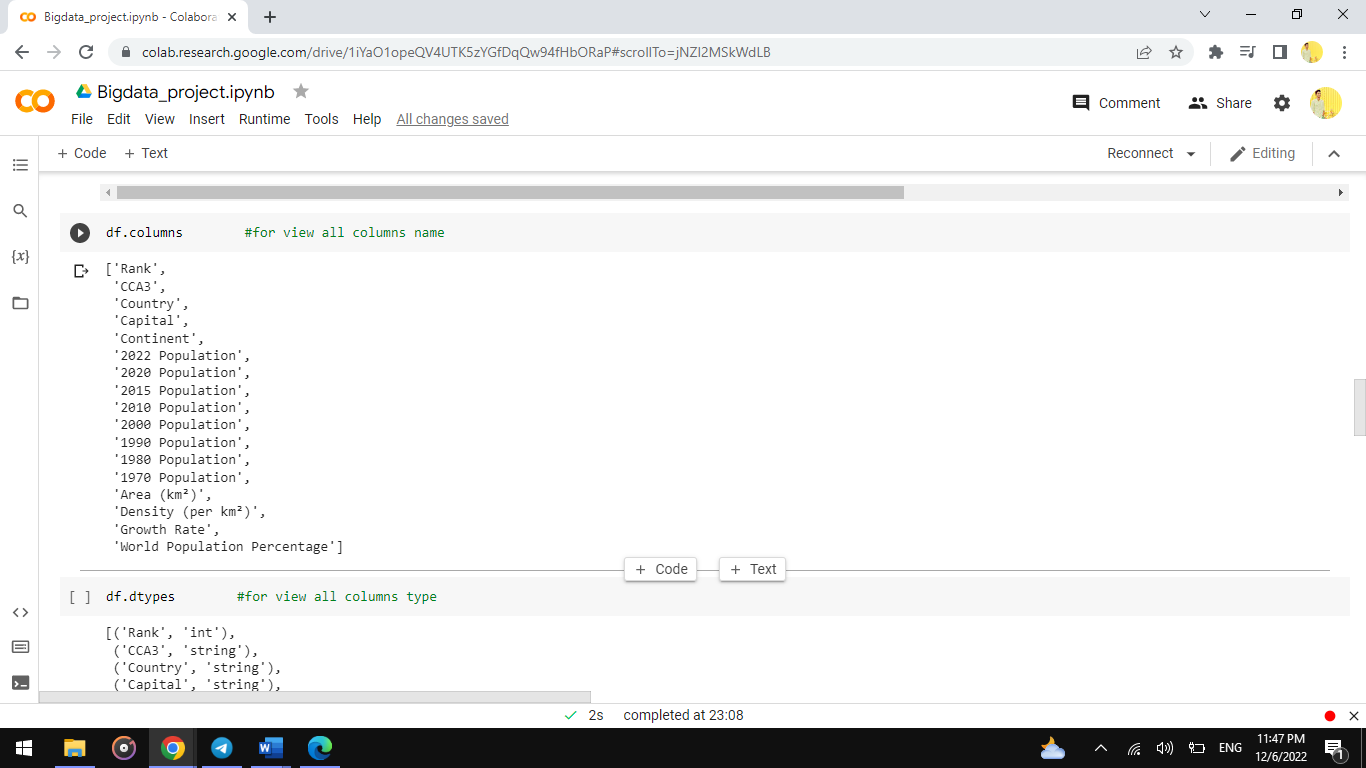
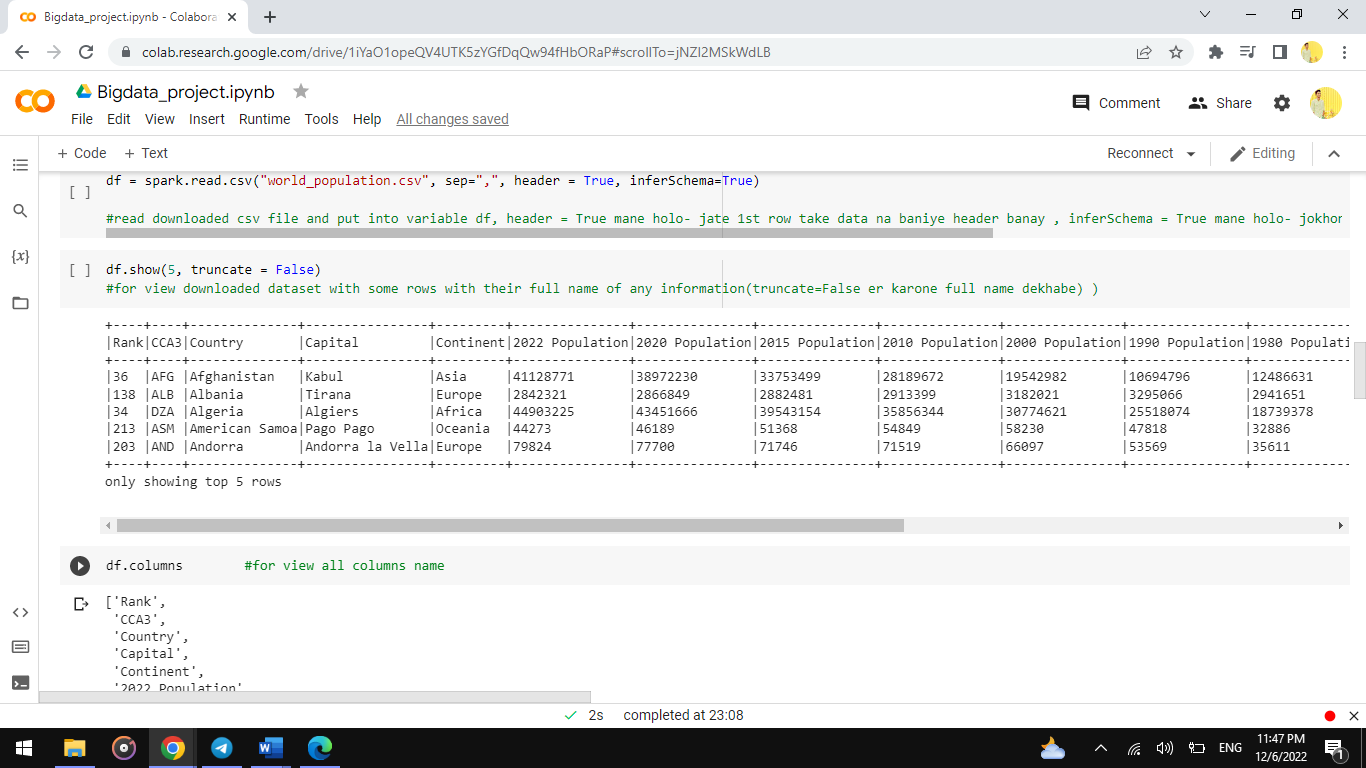
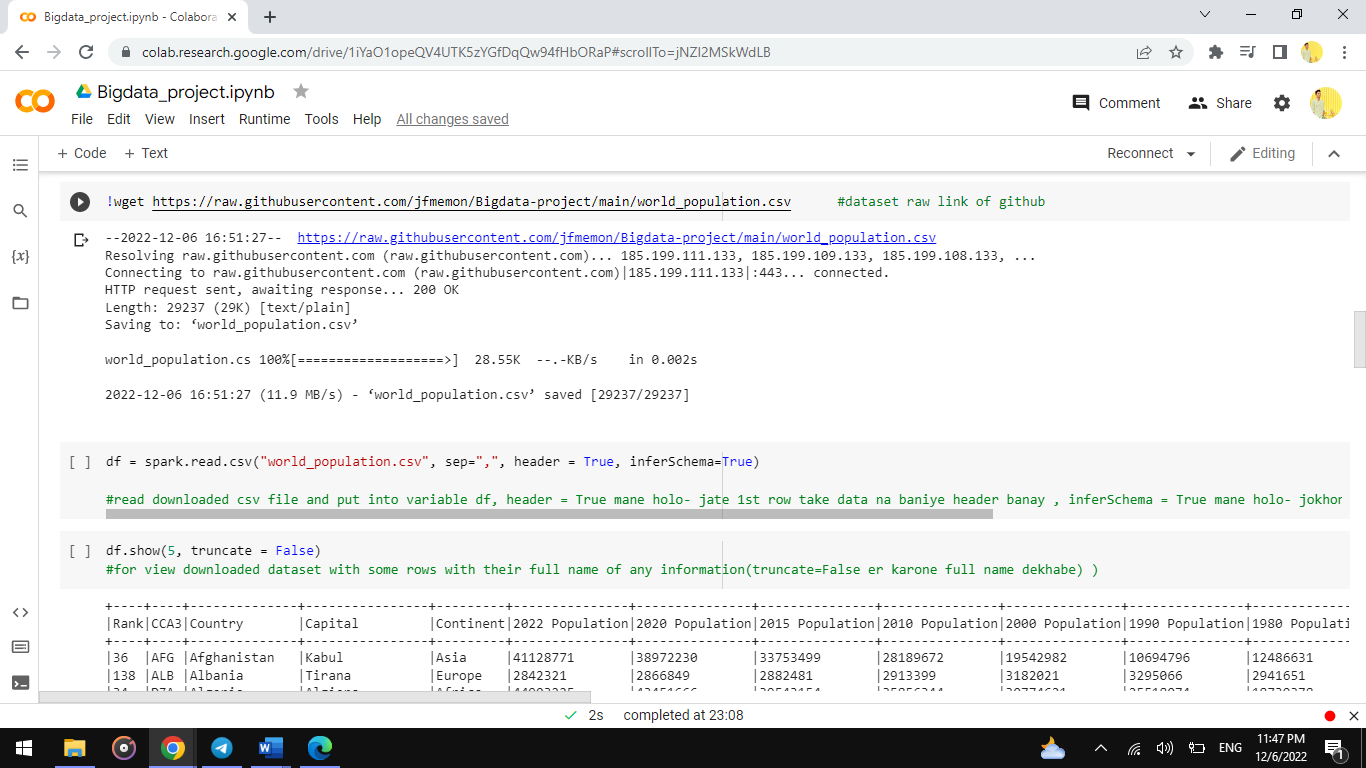
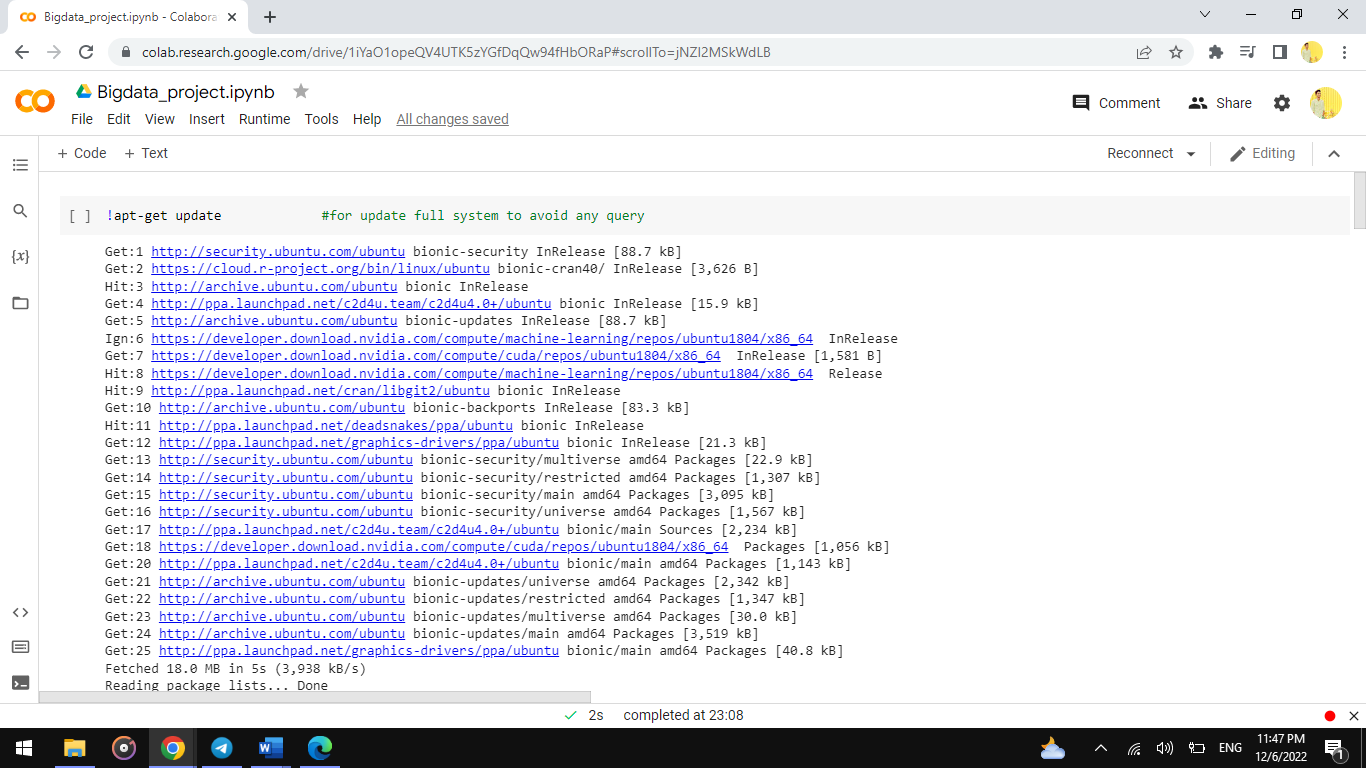
**Framework has been used:**

* **PySpark**

**Dataset link:** [**Click\_here**](https://github.com/jfmemon/Bigdata-project/blob/main/world_population.csv)

**Project github link:** [**Click\_here**](https://github.com/jfmemon/Bigdata-project)

**Code & Output:**



**REFERENCES**

* <https://stackoverflow.com/questions/40728017/how-to-do-mathematical-operation-with-two-column-in-dataframe-using-pyspark>
* <https://pages.uoregon.edu/rgp/PPPM613/class8a.htm>
* <https://www.google.com/search?q=what+is+population+growth+rate&oq=what+is+population+growth+rate&aqs=chrome.0.0i20i263i512j0i512l8j0i390.8918j1j15&sourceid=chrome&ie=UTF-8>
* <https://research.google.com/colaboratory/faq.html#:~:text=Colaboratory%2C%20or%20%E2%80%9CColab%E2%80%9D%20for,learning%2C%20data%20analysis%20and%20education>.
* <https://en.wikipedia.org/wiki/Population_growth>
* <https://sparkbyexamples.com/pyspark/pyspark-lit-add-literal-constant/#:~:text=PySpark%20SQL%20functions%20lit(),by%20importing%20pyspark.sql.functions>
* <https://www.youtube.com/watch?v=uZqS6pJnosU&t=470s>