**Coding challenge**

**Python**

**Name: Rohan Vinayak Chaudhari Email:**[**chaudharirohan24@gmail.com**](mailto:chaudharirohan24@gmail.com)

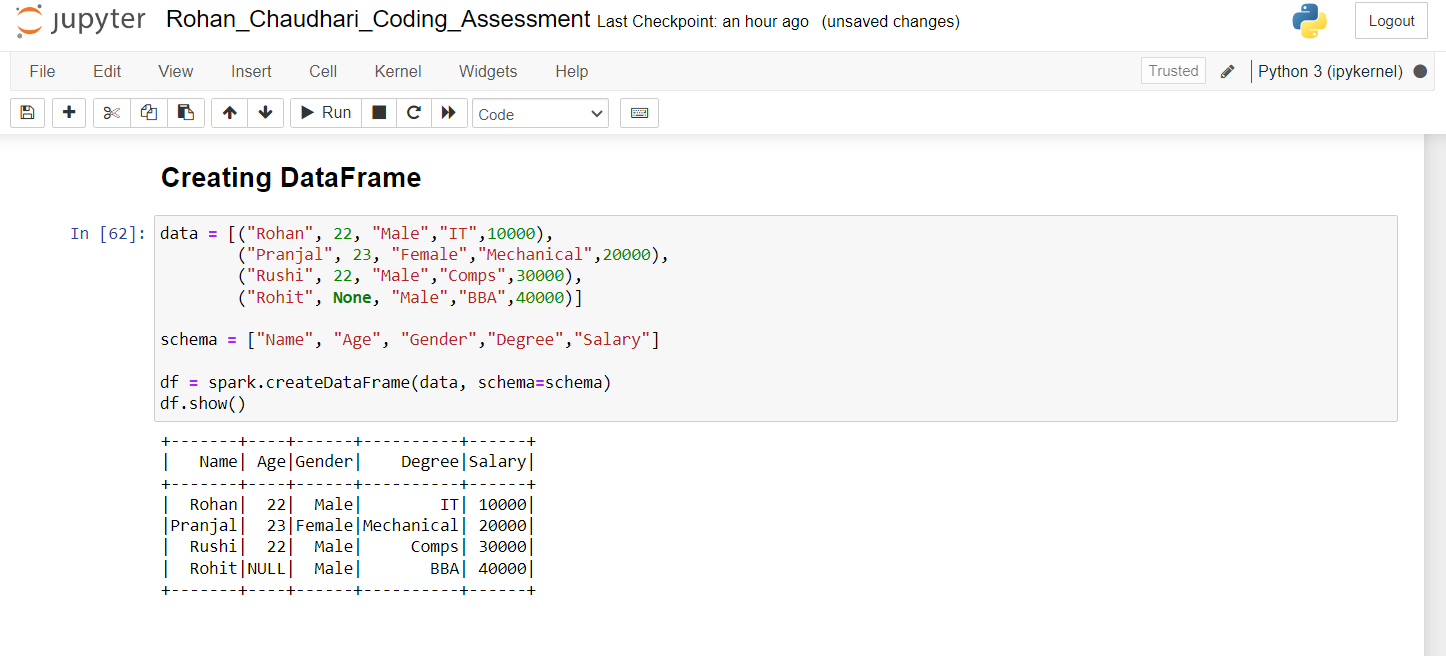
**Batch: Data Engineering 1**

# **Question1:** Execute Manipulating, Droping, Sorting, Aggregations, Joining, GroupeBy DataFrames

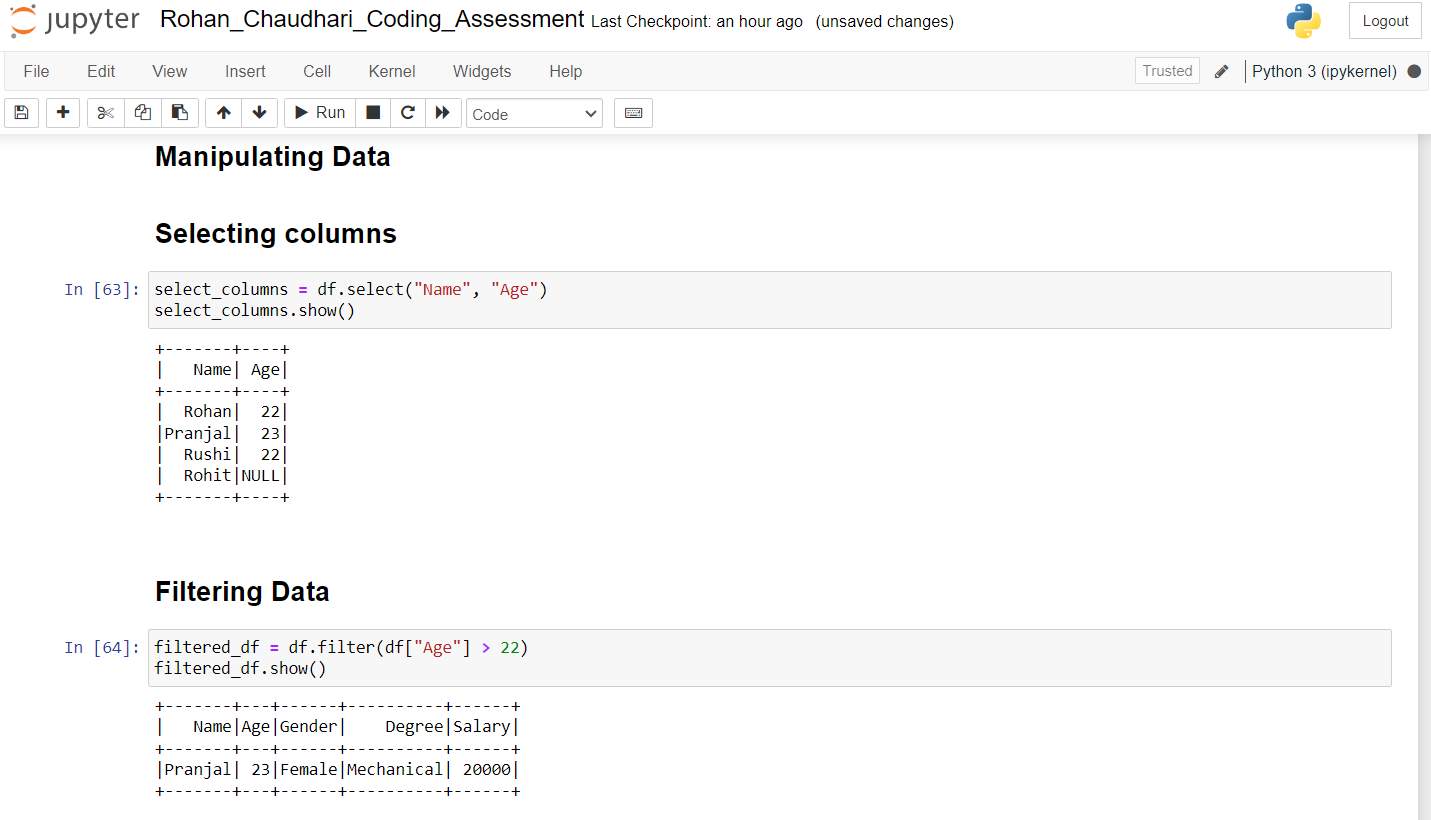
**Created Spark Session**

****

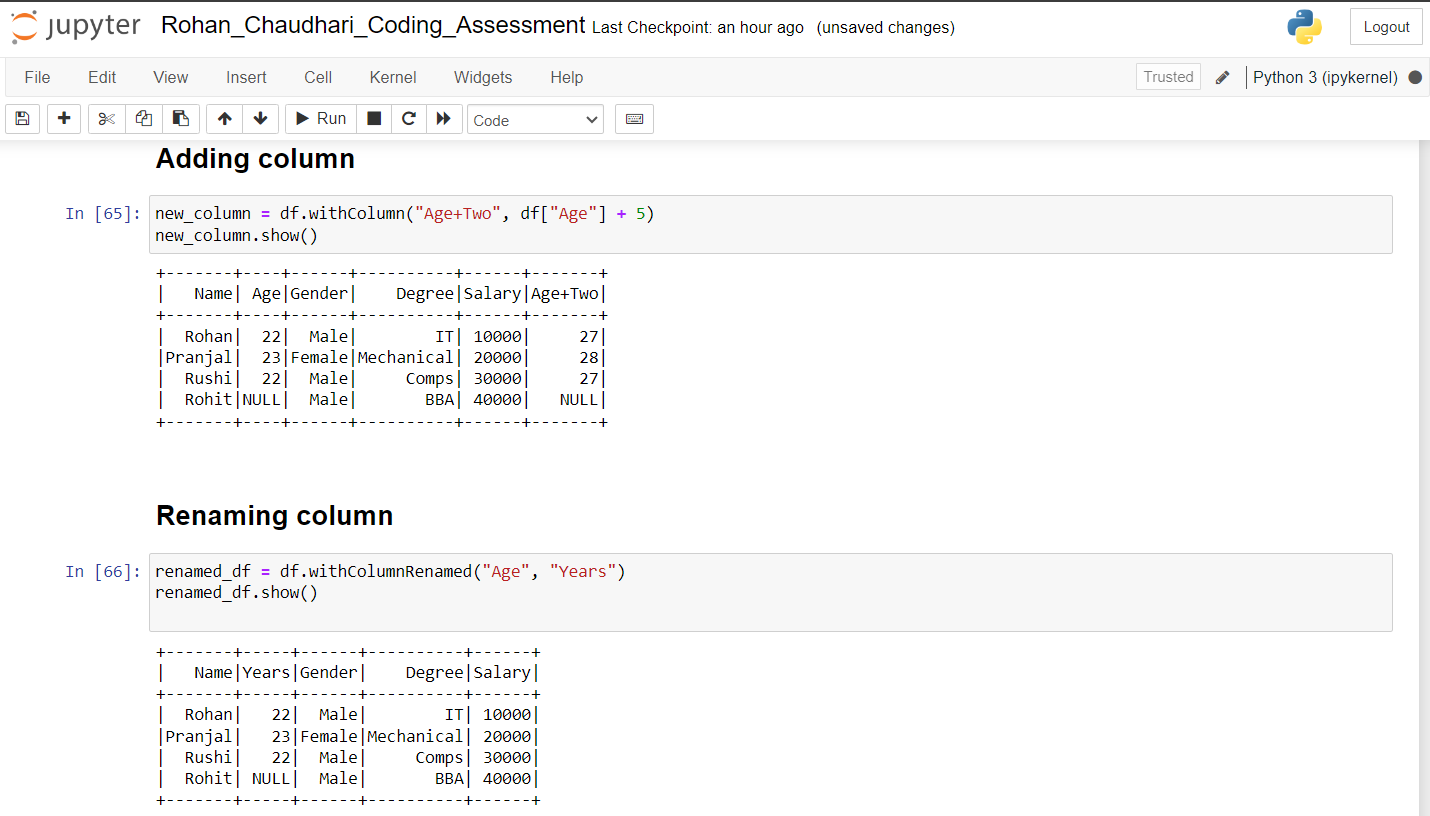
**Created DataFrame**

****

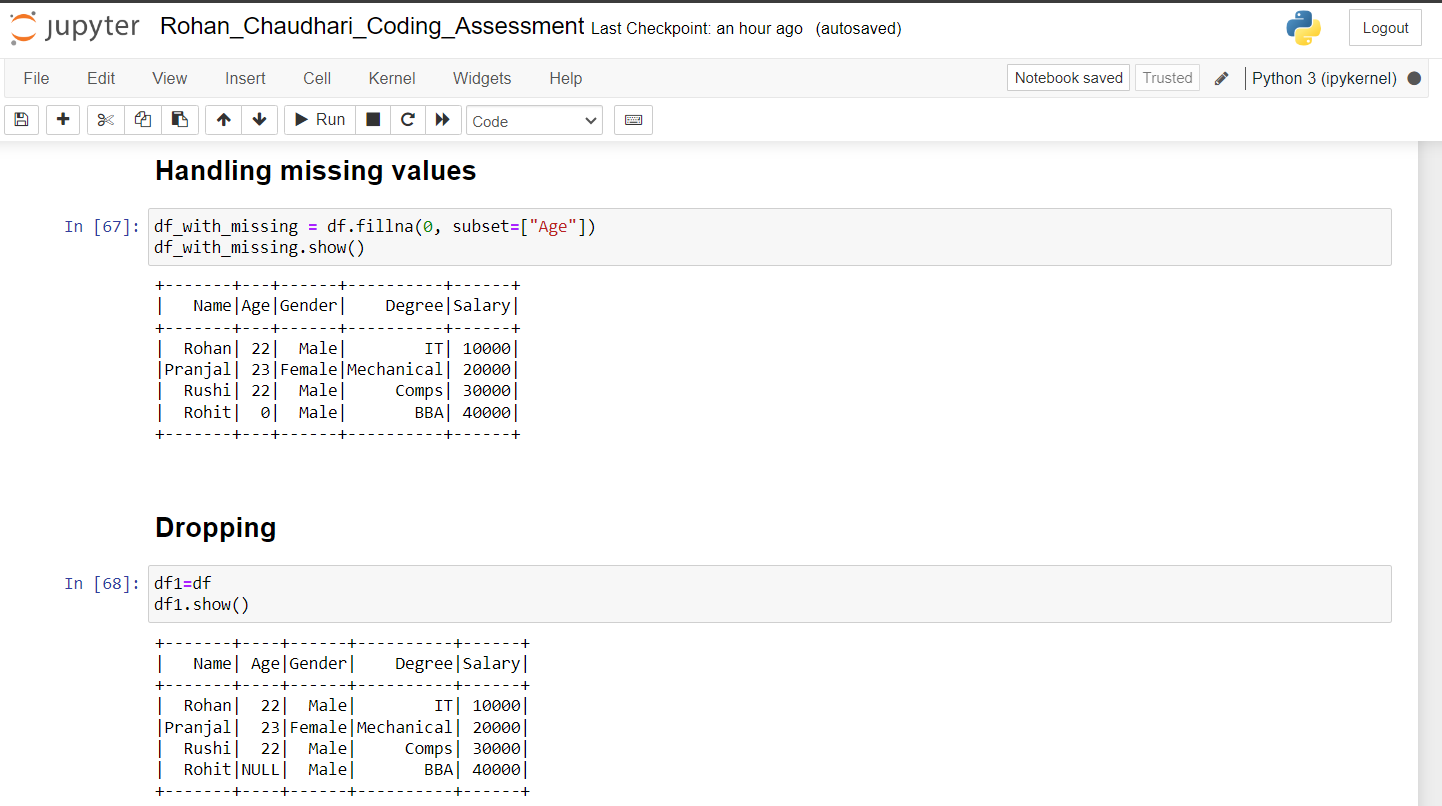
**Performed Operation as seletecting 2 columns and Filtering data based on age:**

****

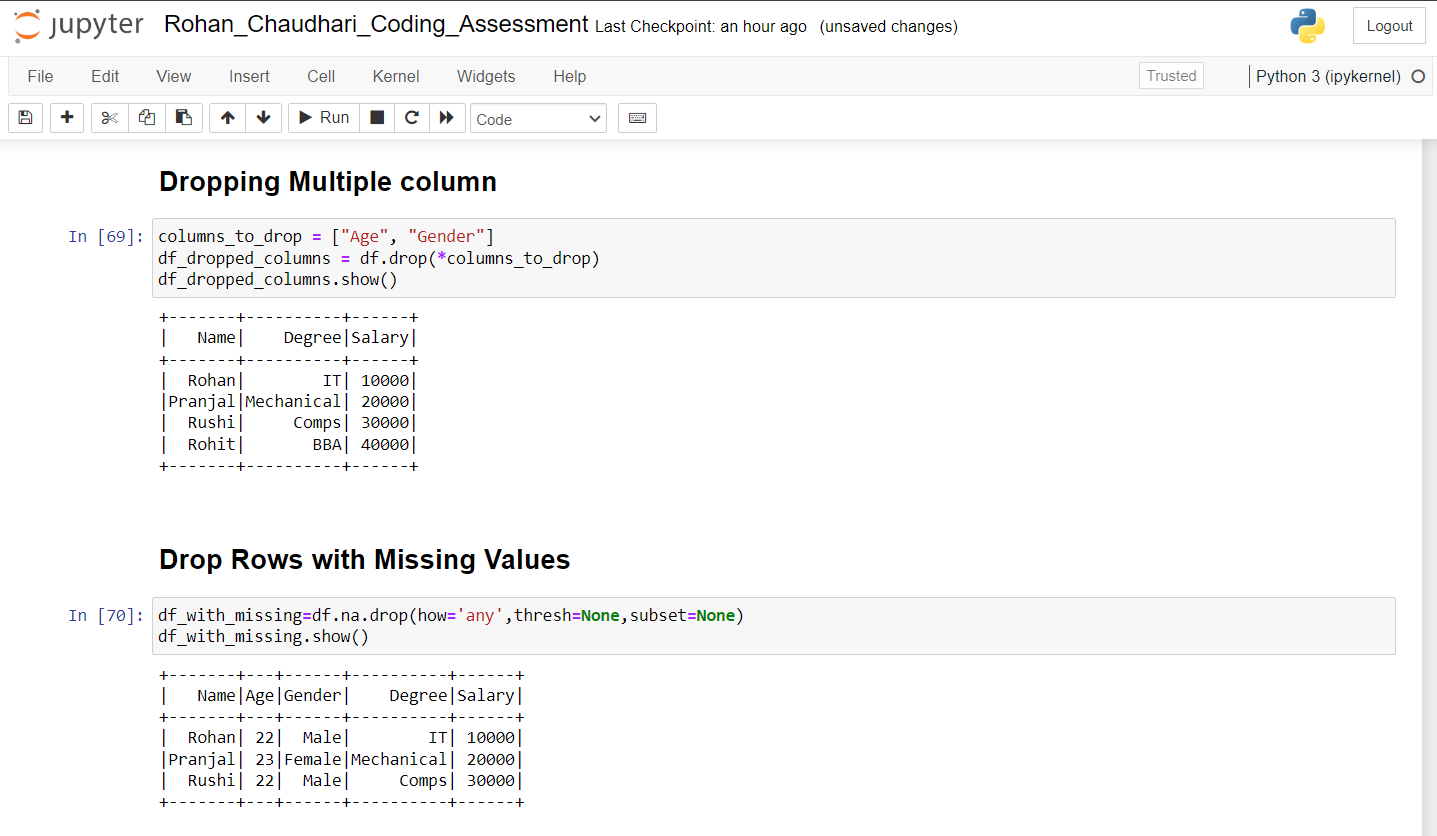
**Added a new column with 5 years ahed & Renamed the column:**

****

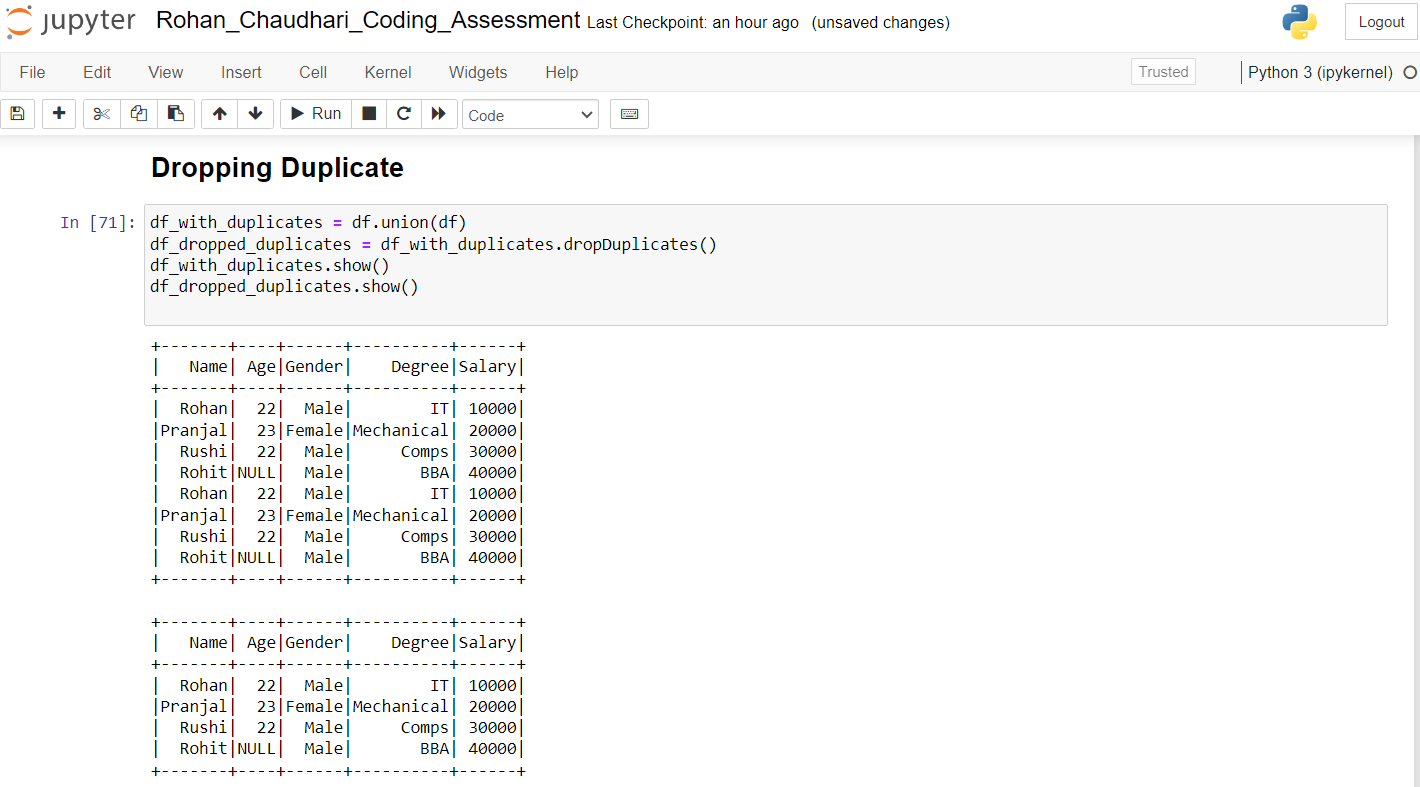
**Handled the missing value of age and filled it with 0:**

****

**Dropped More than 1 column & rows with missing value:**

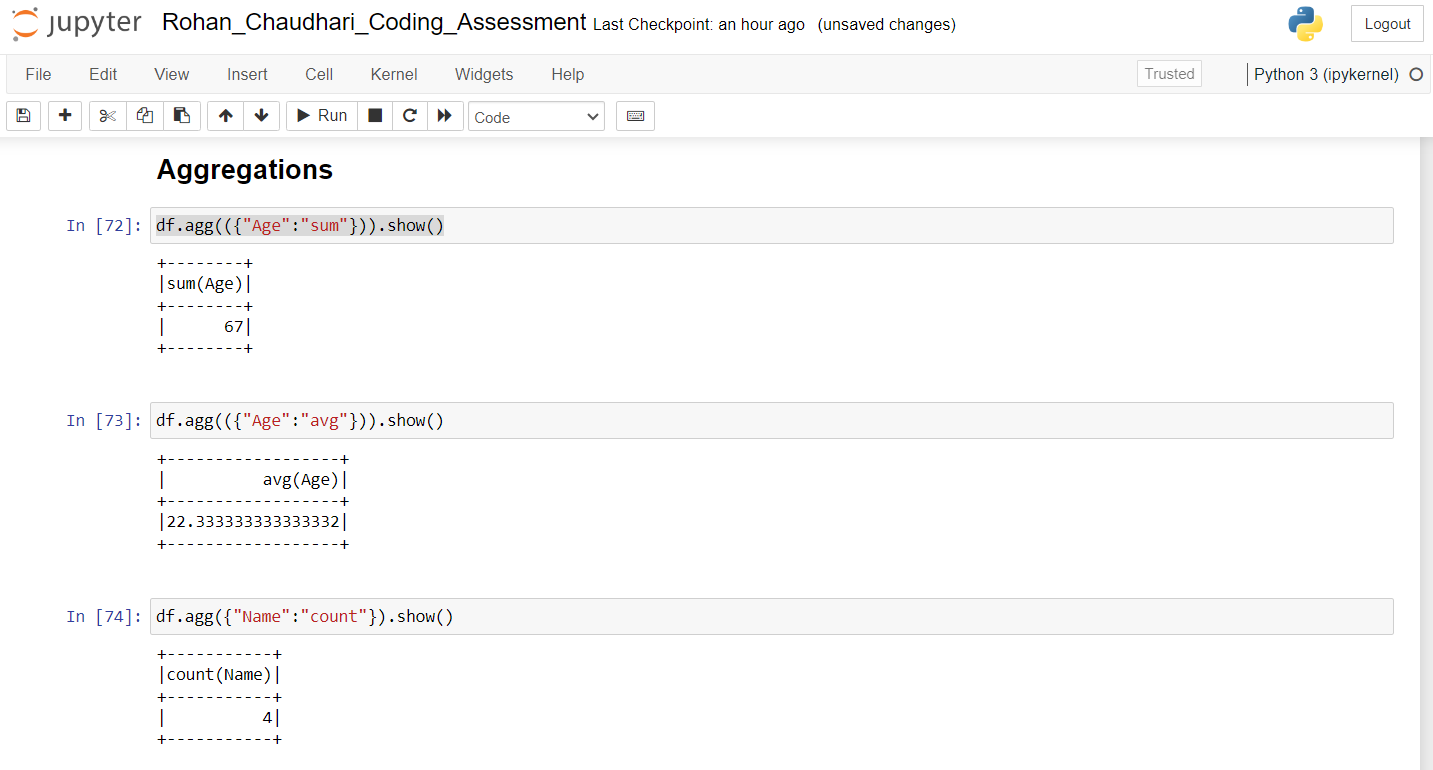
****

**Drop duplicate columns so for that first created duplicate rows with help of union:**

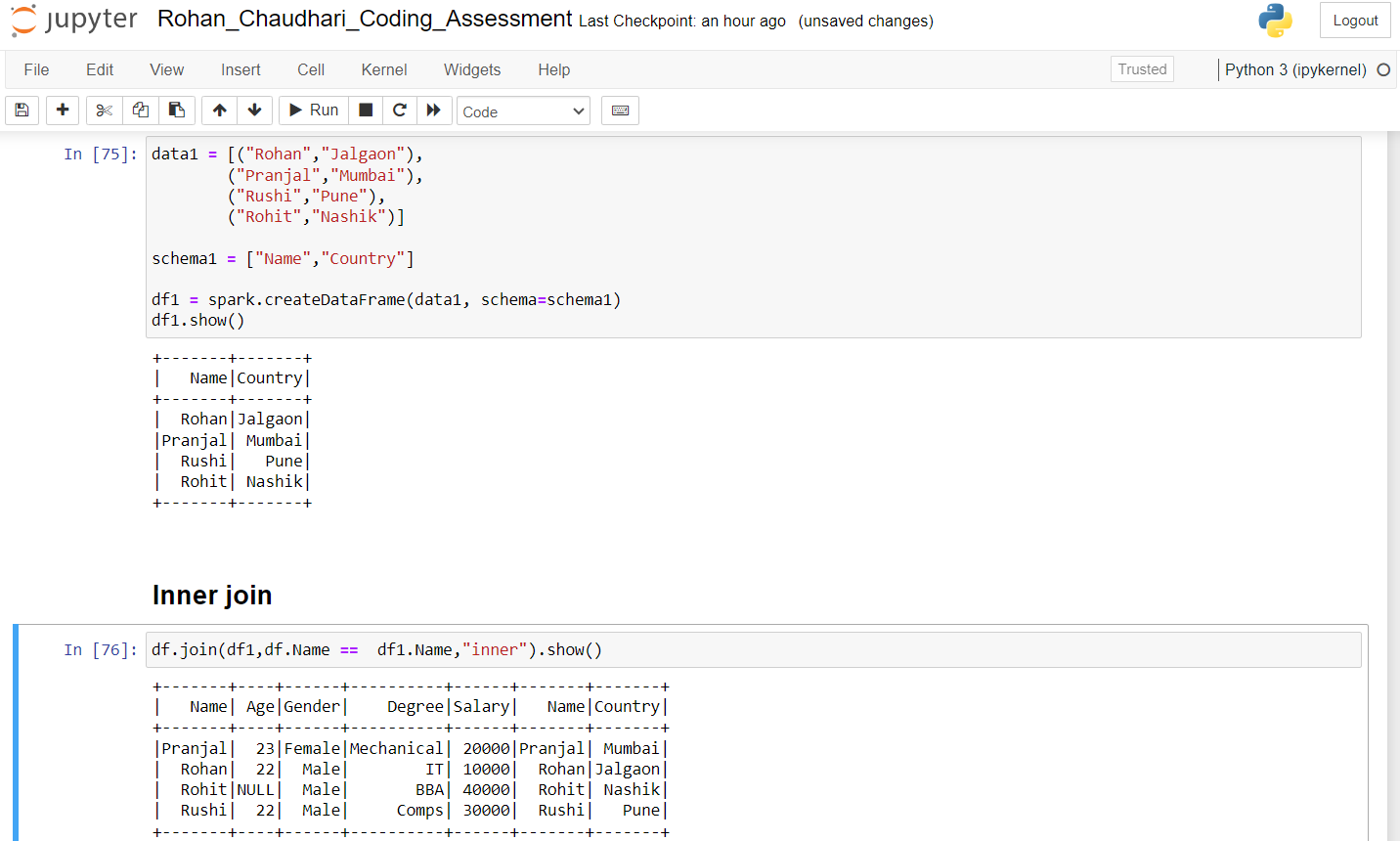
****

**Performed Aggregation Functions:**

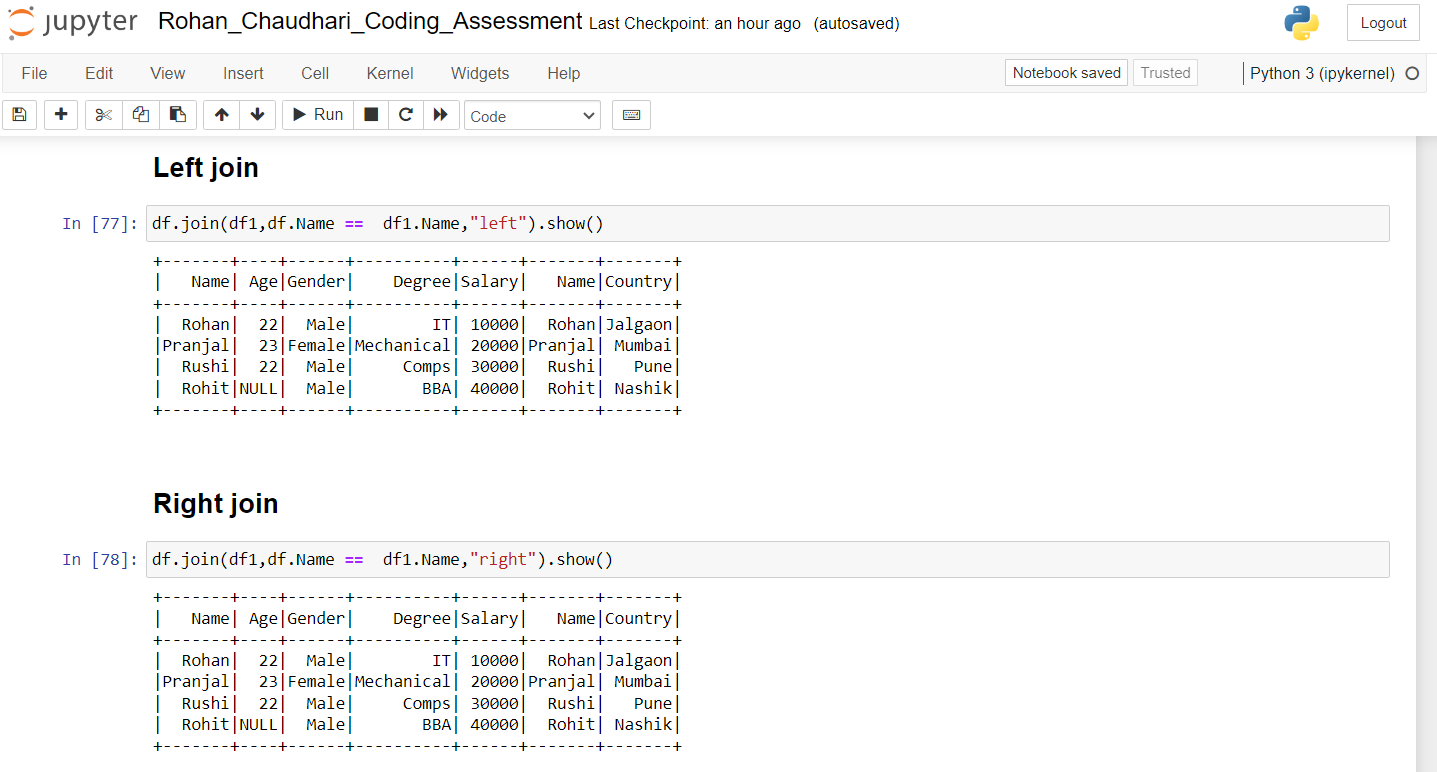
**Sum,Average,Count:**

****

**Created another dataframe for performing joins operation here same column kept is of Name:**

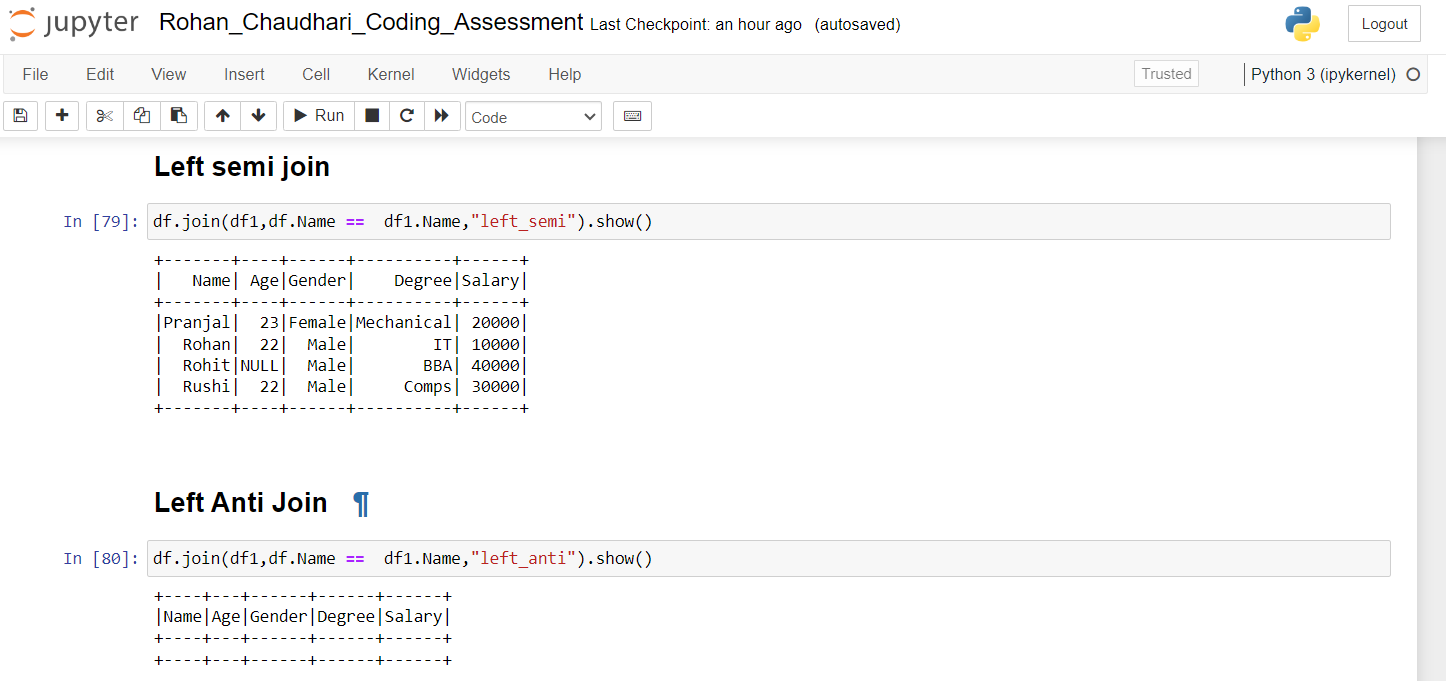
****

**Performed Left & Right Join operation with common column as Name:**

****

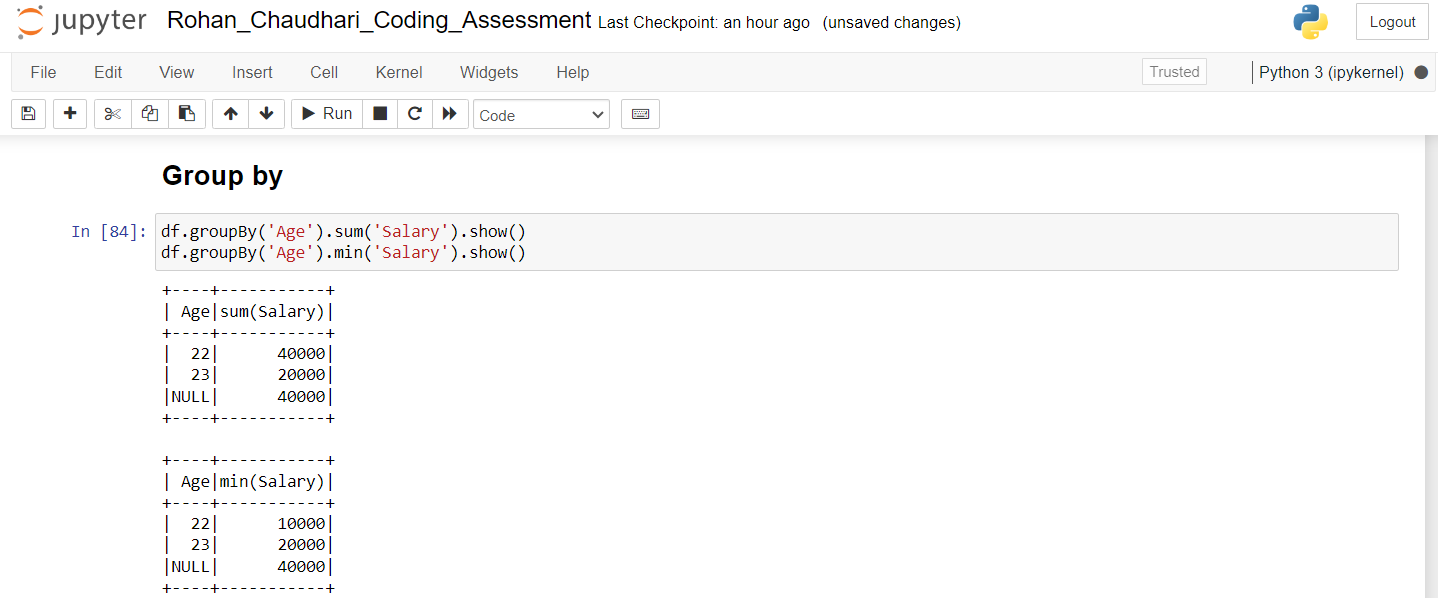
**Performed Left Semi & Anti Join operations:**

**Here Anti join output is null because every data in both table is matched:**

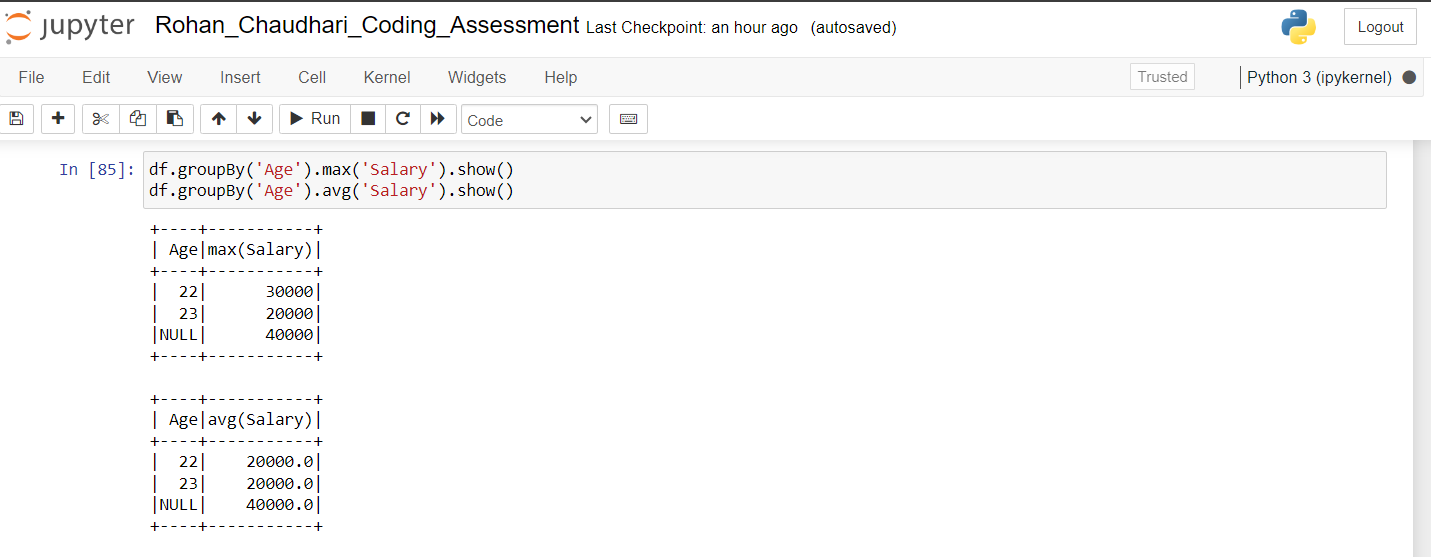
****

**GroupBY Operations:**

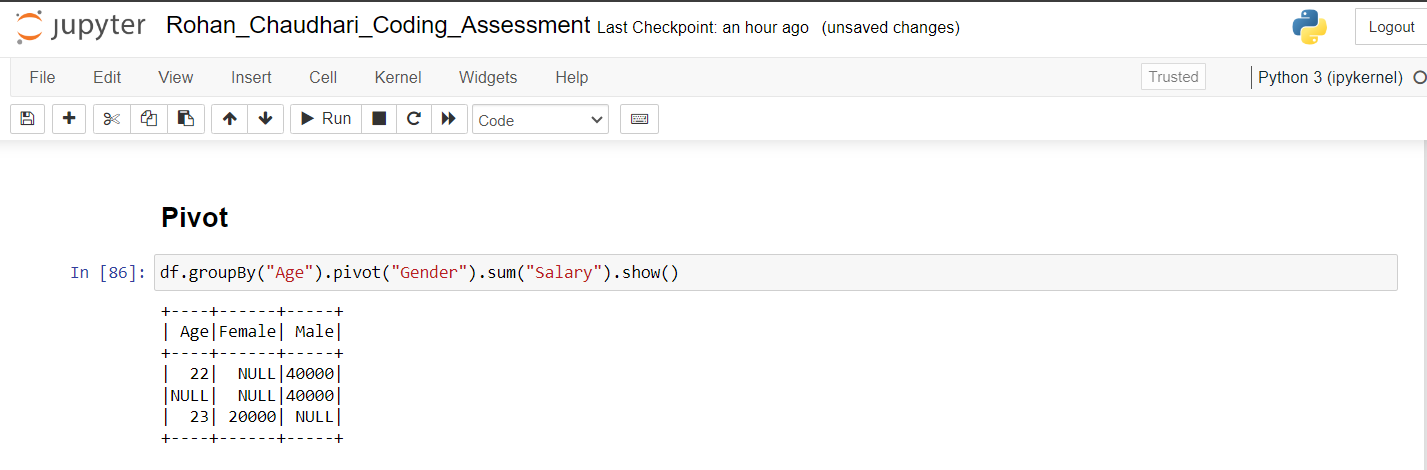
**Groupby with Sum & Min:**

****

**Groupby with Max& Avg:**

****

**Performed Pivot in groupBy with PIVOT kept as Gender:**

****