Write a program in Python according to your variant by using “brooklyn_sales_map.csv” dataset. Write a small report that should contain title page, task, program code and results. To complete the task you can use books “Learning Spark. Lightning fast data analysis”, “PySpark SQL Recipes” and of course internet.

Variant 1.

- Find the average price of housing (sale_price) and output a new table, which will contain two columns – sale price and deviation from average value.
- Find average price of housing (sale_price) for every borough.
- Output the average housing area (gross_sqft) for every combination of tax class and year of sale.
- Output a table that contains the number of null values for every column.

Variant 2.

- Find the average area of housing (gross_sqft) and output a table, which contains gross_sqft and deviation from average value.
- Find the average area of housing for each year it was built (year_built).
- Find the average sale price for every combination of neighborhood and building class category.
- In the initial dataframe delete the rows with records of housing that was built after year 2000, and also those which only contain null values.

Variant 3.

- Find the average price of housing sale and output a new table that contains two columns – sale price and percent of its deviation from average value.
- Output a table, which contains all building class categories and the number of records belonging to those categories.
- Output a table, which contains average values for every column in dataframe.
- In the initial dataframe fill the null values with average by column.

Variant 4.

- Add a new column to the initial dataset, which will contain the “age” of building.
- Output a table that contains average sales date for every combination of zip code and tax class.
- Output a table which contains sum of sale prices over every combination of tax class and zip_code.
- Create a new table that has 10 columns of the initial dataframe. Columns may contain null values, but null values should not be predominant. Afterwards, delete all rows that contain only null values.

Variant 5.

- Find the average year when housing were built and output a new table which contains housing year built and deviation from average value.
- Sort dataset by sale price (ascending) and zip code (descending) at the same time.
- Output a table with the highest sale prices for every combination of neighborhood and building class category.