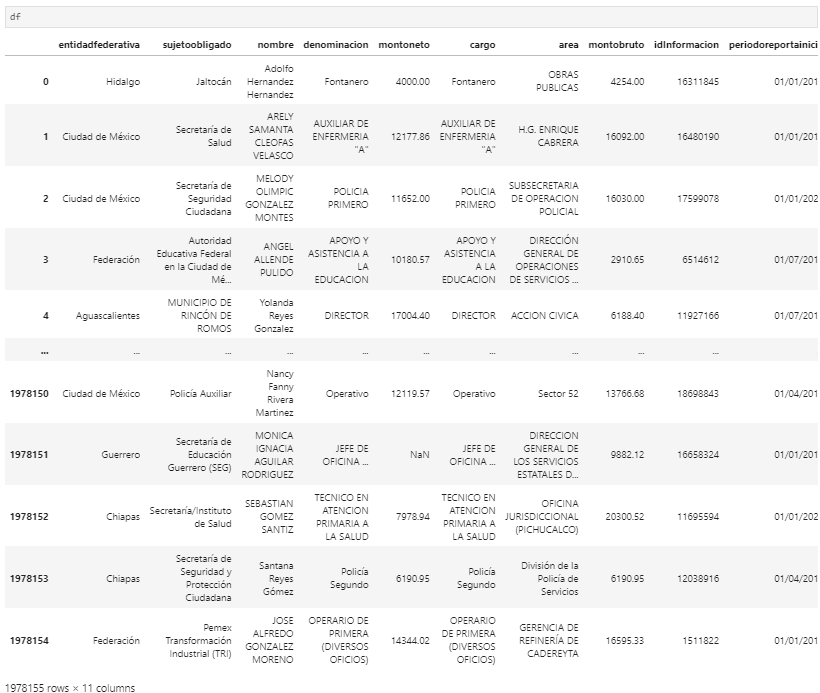
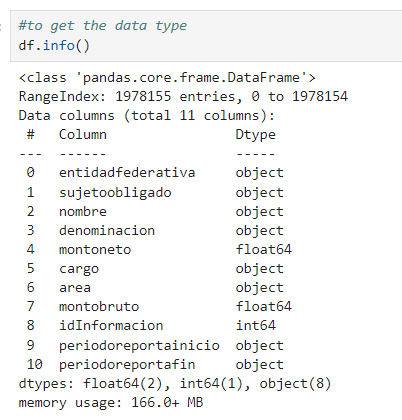
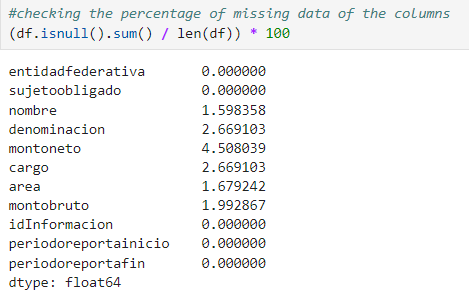
Data description

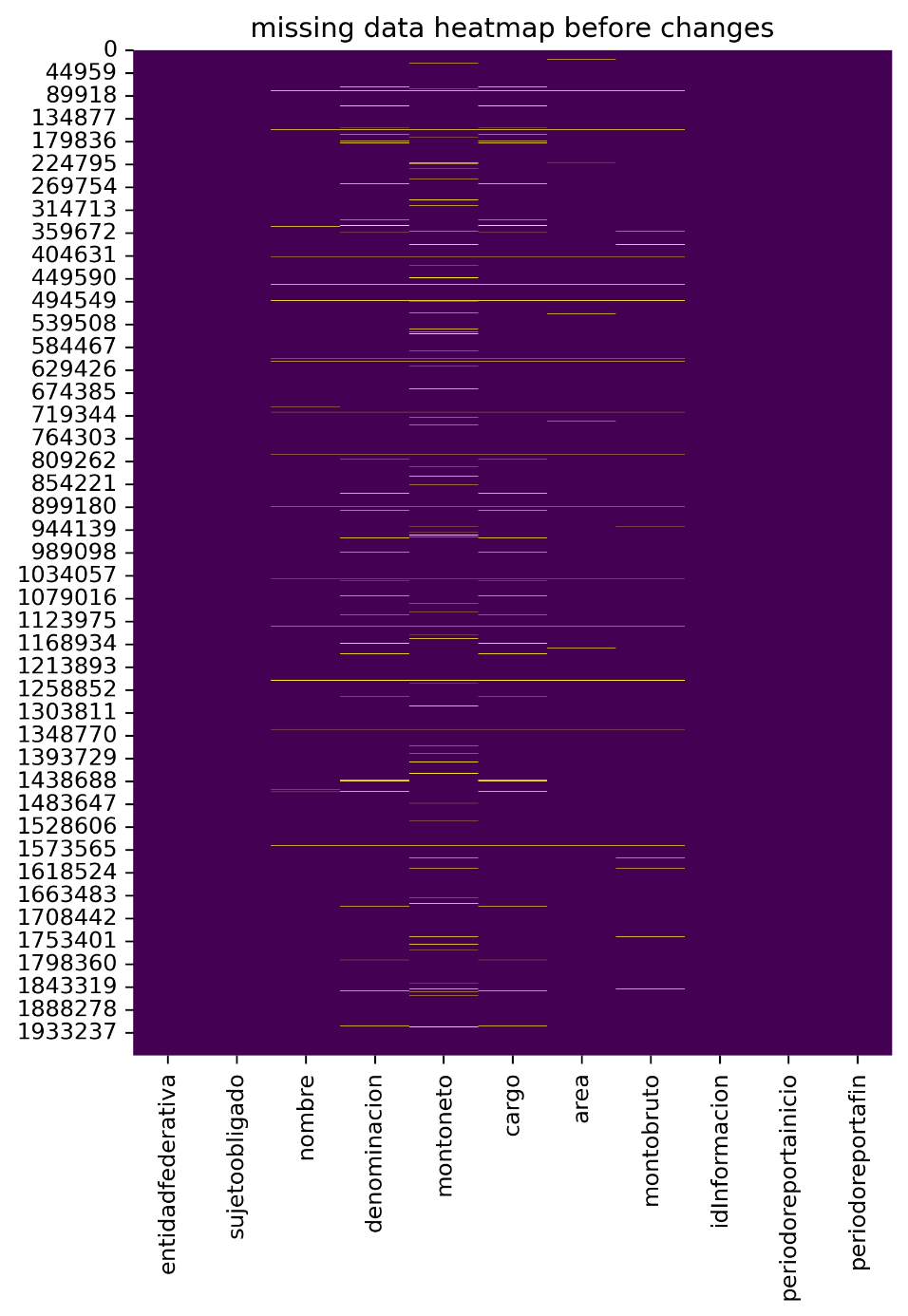
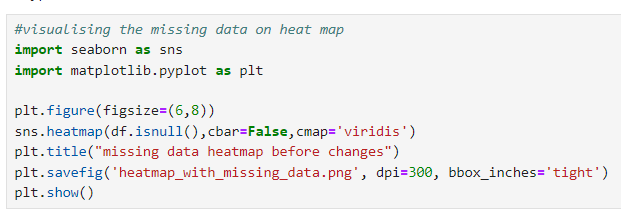
1. This dataset chosen is “Mexican Federal Government Salaries” from the Kaggle website: <https://www.kaggle.com/datasets/ivansabik/mexican-federal-government-salaries>
2. This dataset has 1978155 rows × 11 columns
3. Data types of all the columns are below

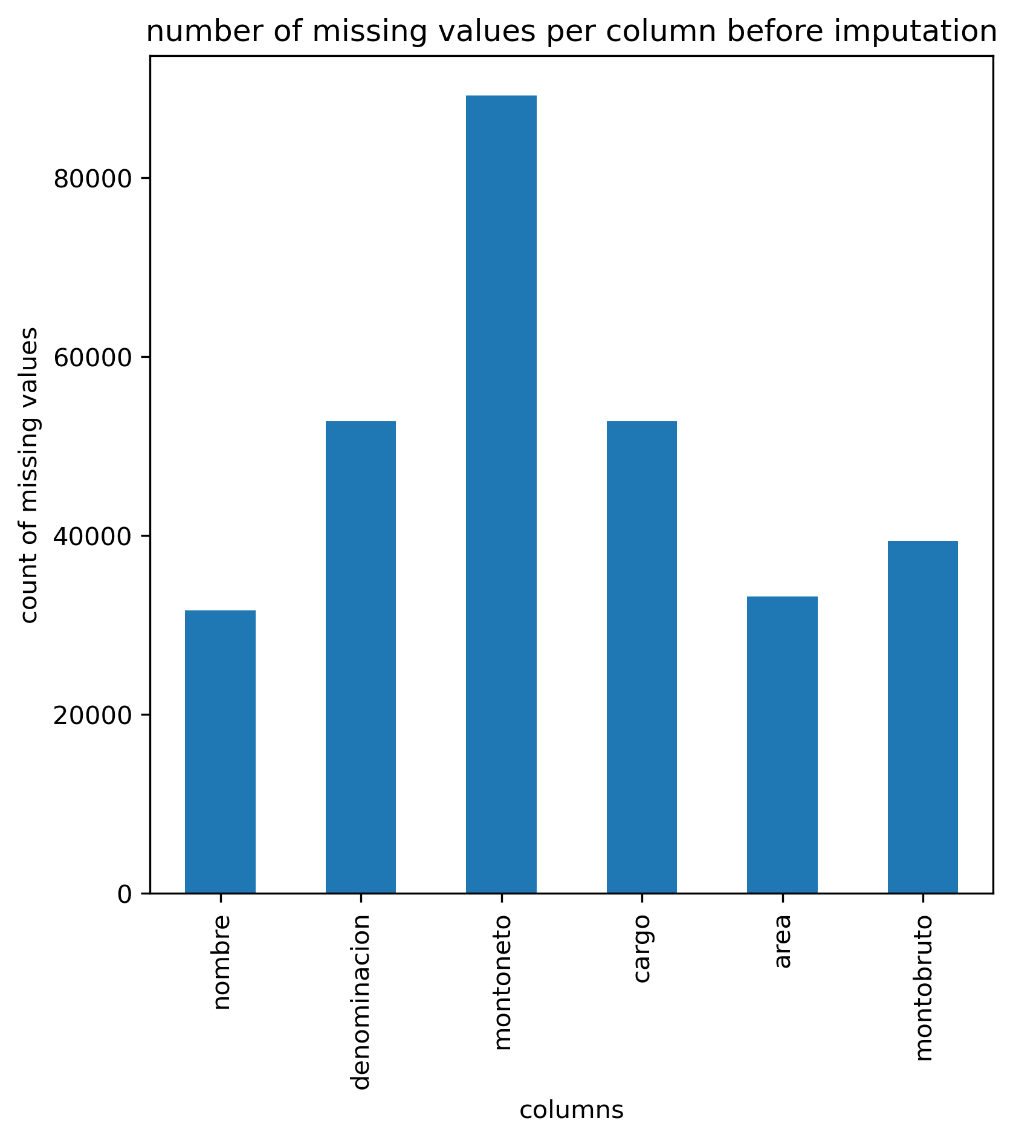
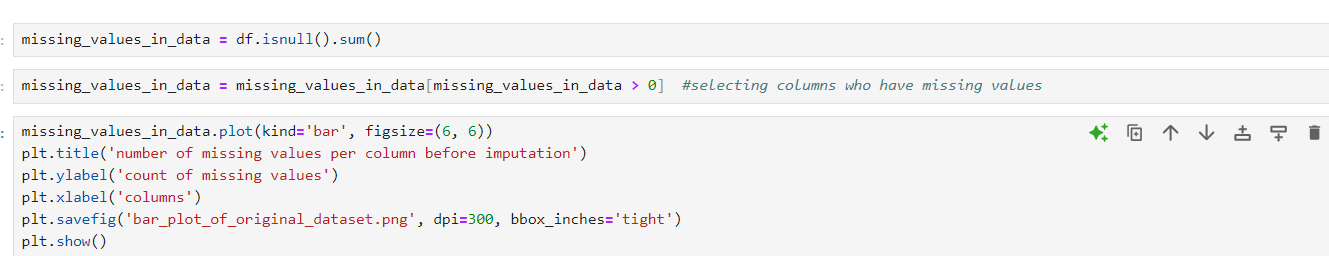
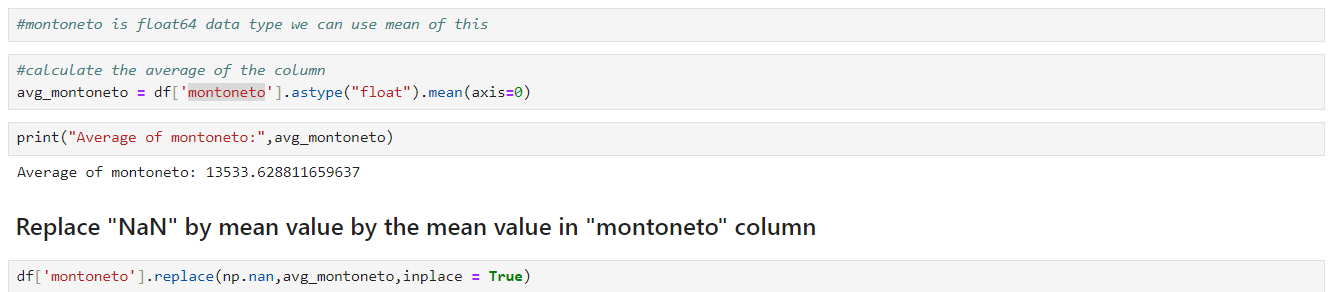


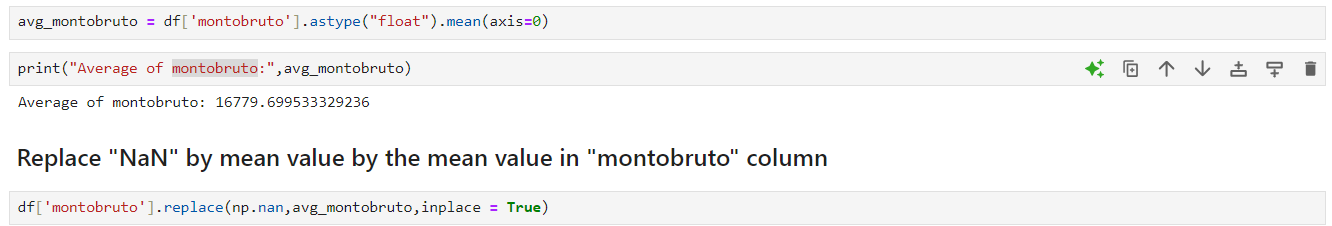
1. Checked the missing value in the dataset

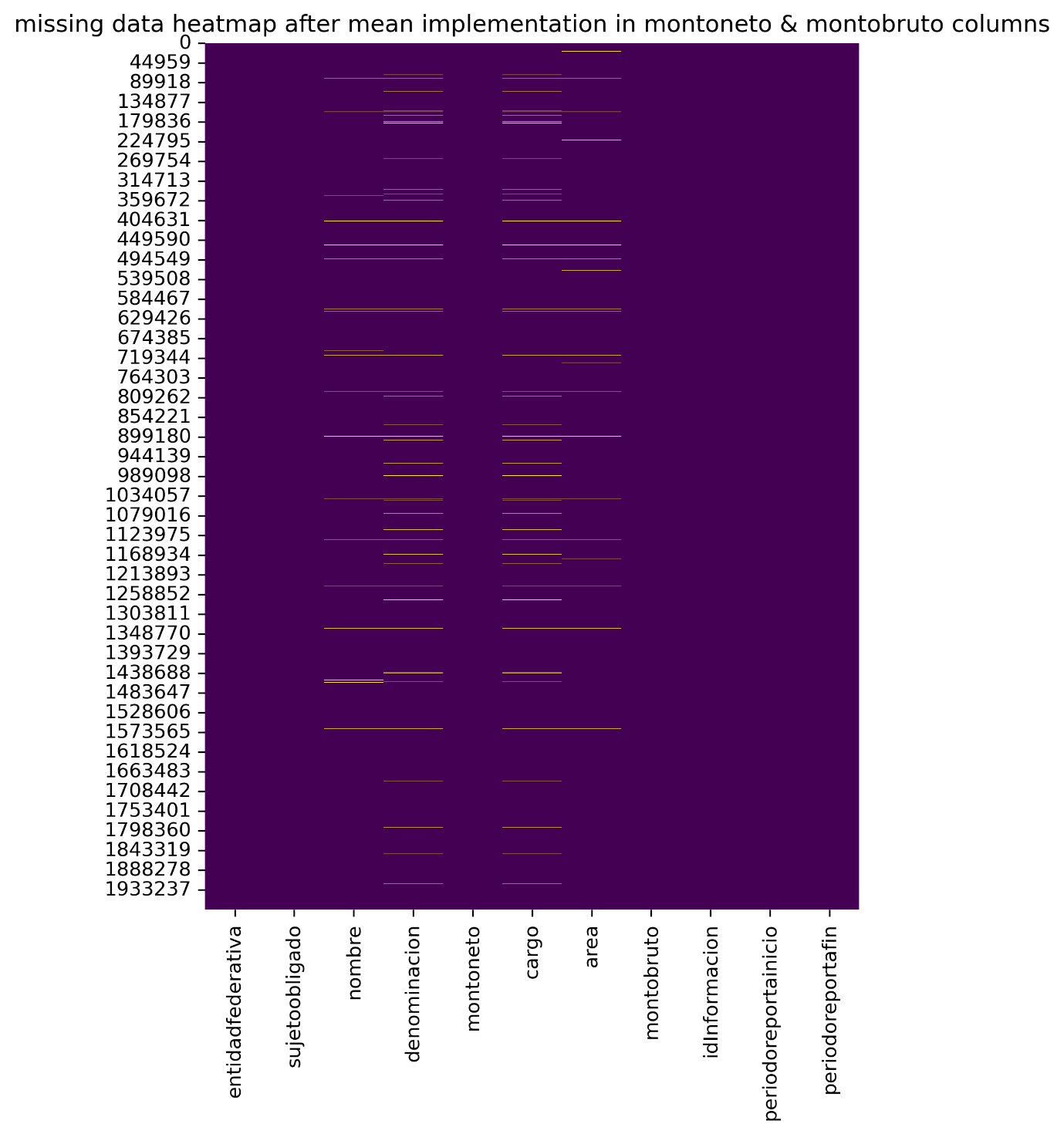
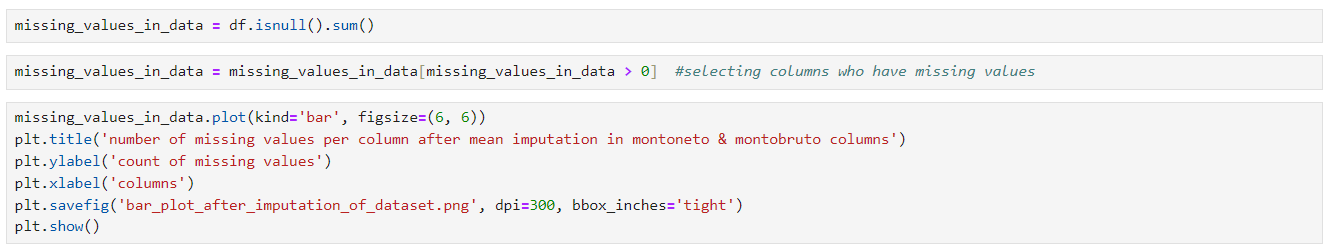
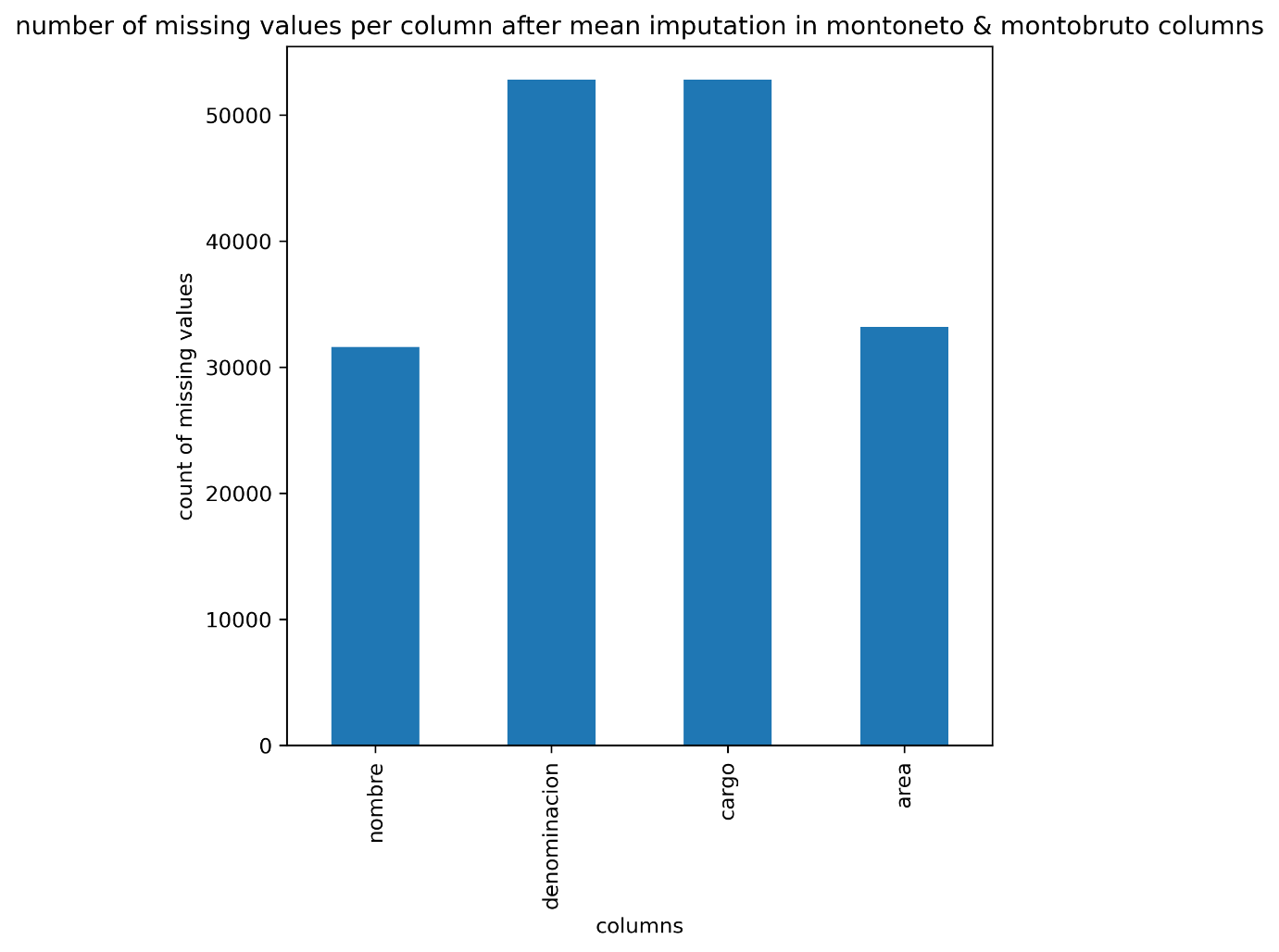
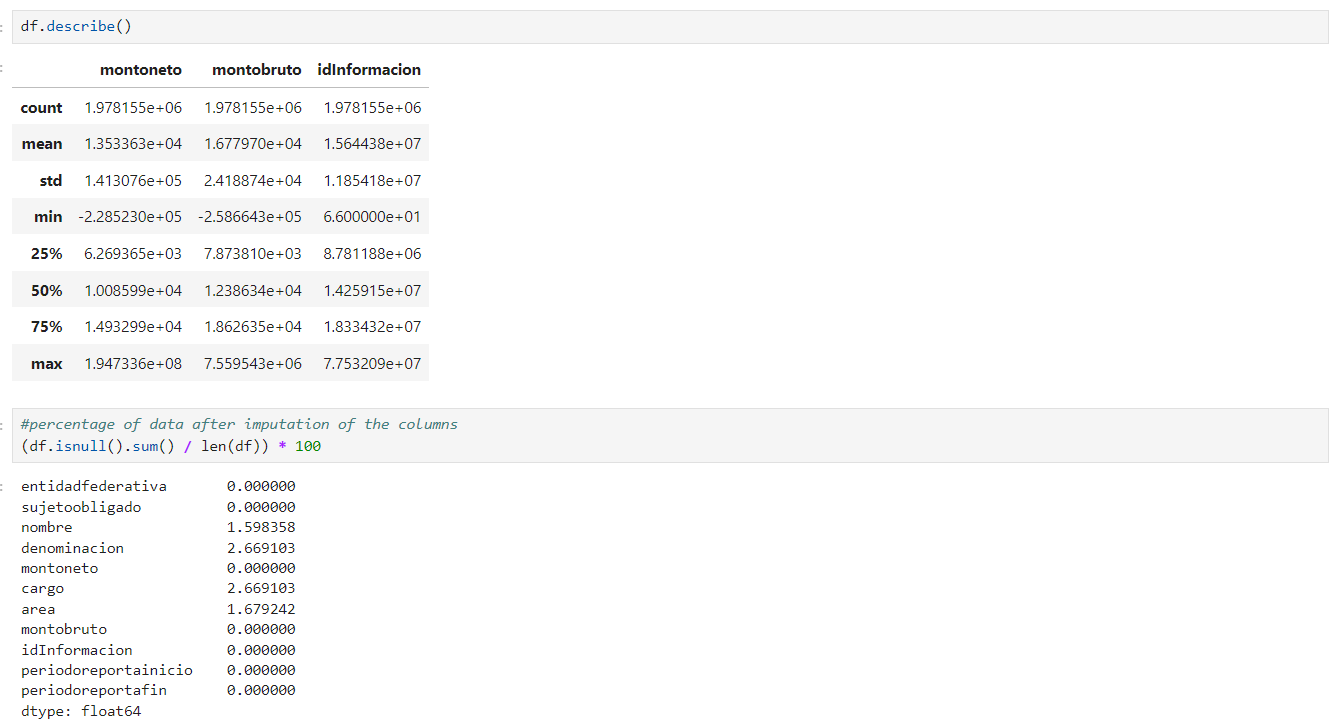


1. Percentage of missing data in the dataset 
2. Heatmap of missing data of original dataset



1. Bar plot of missing data of original data set 
2. As per analysis there’s missing data in six columns however only columns **montoneto** and **montobruto** have data type float64 where imputation can be done for data wrangling
3. Imputation in montoneto column
4. Imputation in **montobruto** column



1. Heat map after data imputataion  
2. Bar Plot after imputation of data set  
3. Dataset description after Data Wrangling 
4. Why mean imputation? This amount of dataset preserves central tendency that’s why mean represents the best option for missing numeric values.
5. For categorical data with data type “object” we can change the missing values to “unknown” to make similarities for estimation. 