**BIKE SALE ANALYSIS**

**Introduction**

This analysis was carried out to answer specific questions about bike buyers of different educational degree, age brackets income e.t.c in a specific region.

This analysis was done using the PYTHON PROGRAMMING language

**Objectives**

The primary goals of this analysis are

1. To understand the trends and patterns in bike sales dataset
2. To identify key factors driving sales such as product category, seasons or customer segment
3. To uncover insight into the most profitable product, customer behaviour and regional sales performance
4. To provide actionable insight through visualization that can guide business decisions and strategy

**Data Analysis Process**

**Data Collection**

1. This data was collected from an email sent by the class facilitator Joseph Elijah. The mail contained two files, which contained the Assessment spreadsheet and the group assessment question.
2. The assessment spreadsheet was opened using Microsoft excel, and several data columns and rows containing various data elements was observed.
3. The file containing the assessment questions was used to analyze the data.

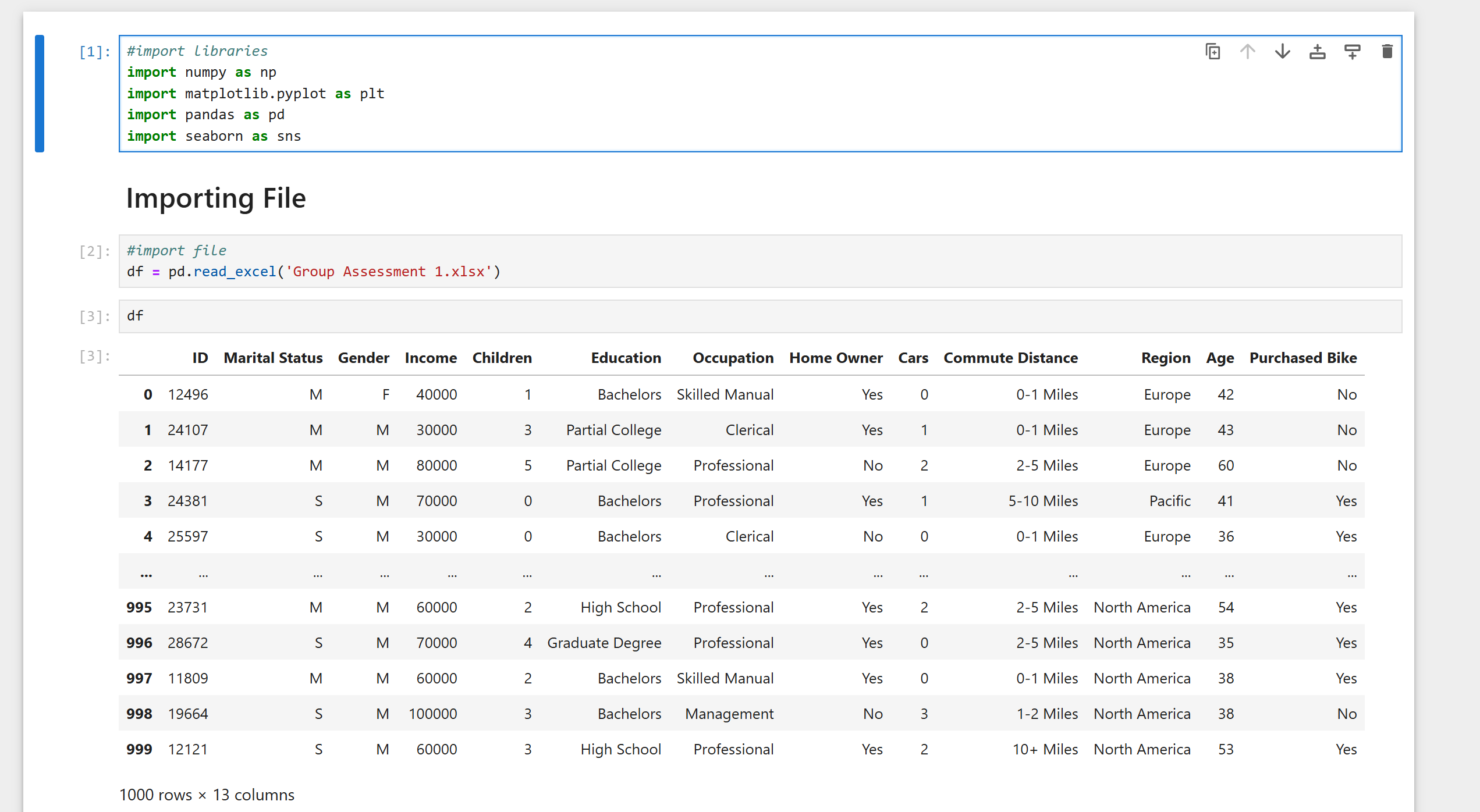
**Data Preprocessing**

The code for importing data libraries and file was imputed to prepare the python workspace

Numpy is imputed to work with numbers/integers, Pandas; to work on Series or Dataframe and also to alter data, Matplotlib and Seaborn are used for data visualization

The file was saved in a excel (.xlsx) format and was uploaded as such.

And the dataset was displayed as shown below;

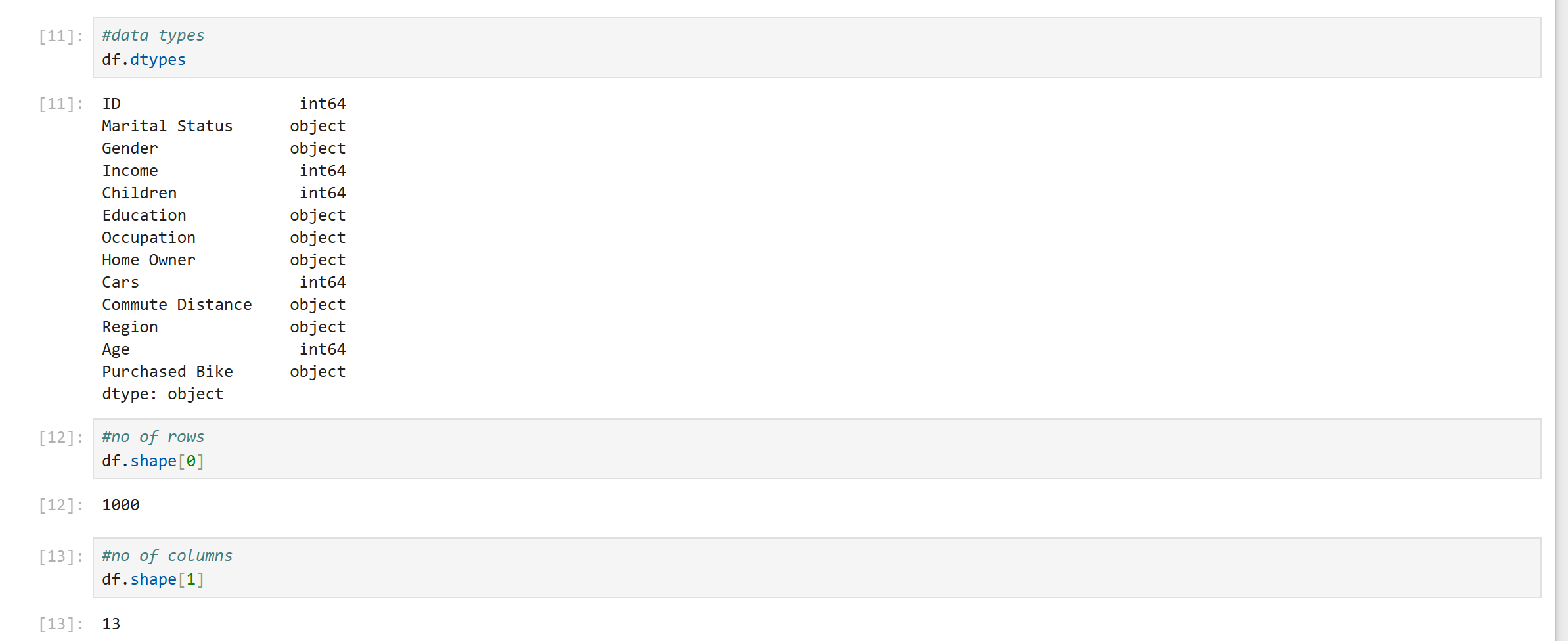


**DATA EXPLORATION**

We made use of Pivot tables to explore the formatted data set and answer the given questions, we also used charts e.g. bar and line charts to be able to perfectly visualize the answers.

Using the code df.shape, there are 1000 rows and 13 columns from the dataset presented above

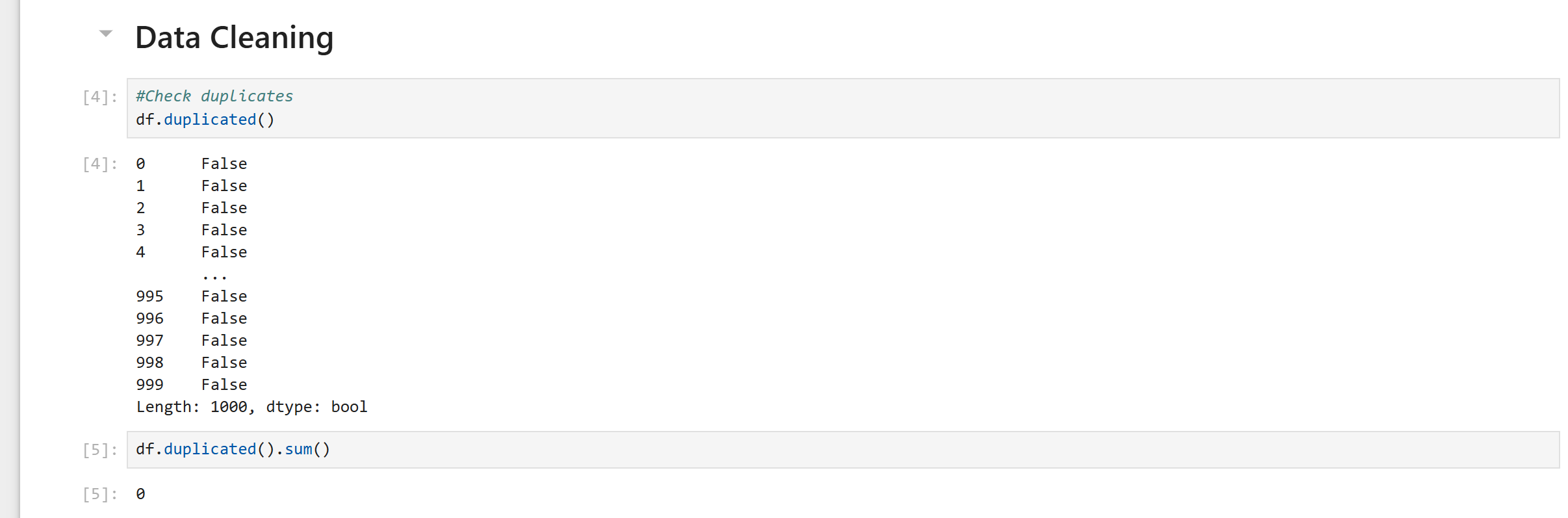
Using the “data.dtypes” code, all columns with string values are classified as ‘objects’ and columns with whole numbers are classified as ‘integer’ data types.



**DATA CLEANING AND FORMATTING**

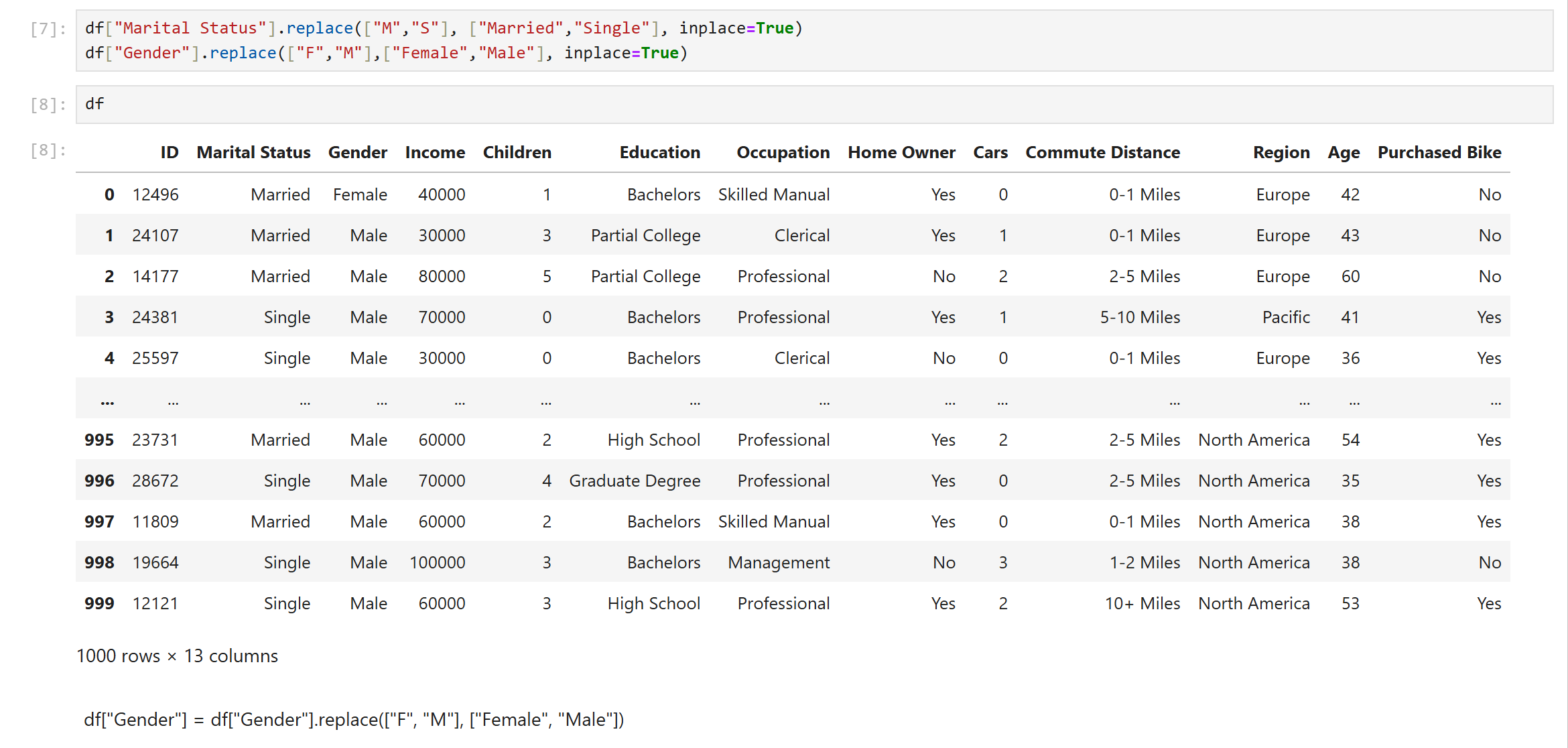
**Removing Duplicates**

The data was checked for duplicates using the code data.duplicated() and the code was run and there was no duplicate found in the data



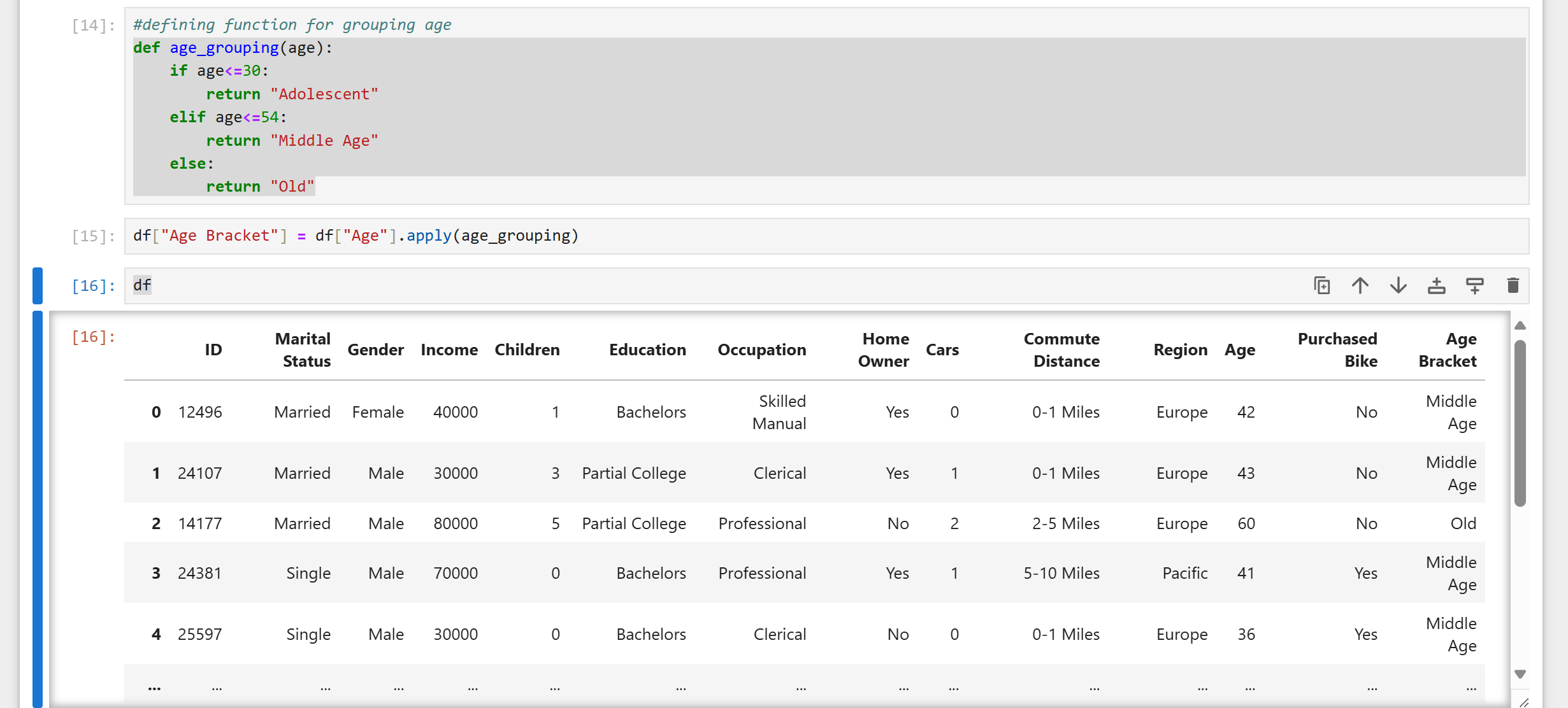
**Replacing columns**

The Marital status contained columns with the abbreviations M and S which was replaced to Married and Single and the Gender column had columns with the abbreviations M and F which was also replaced Male and Female.



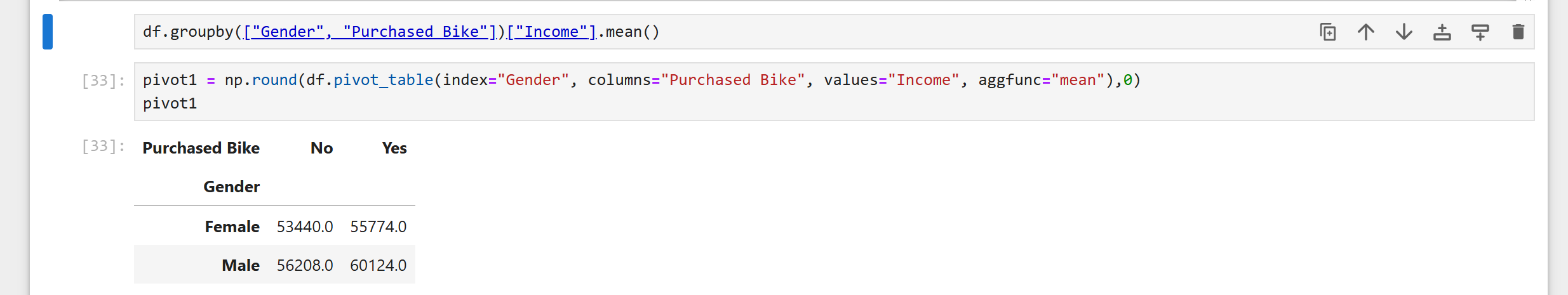
Age classification

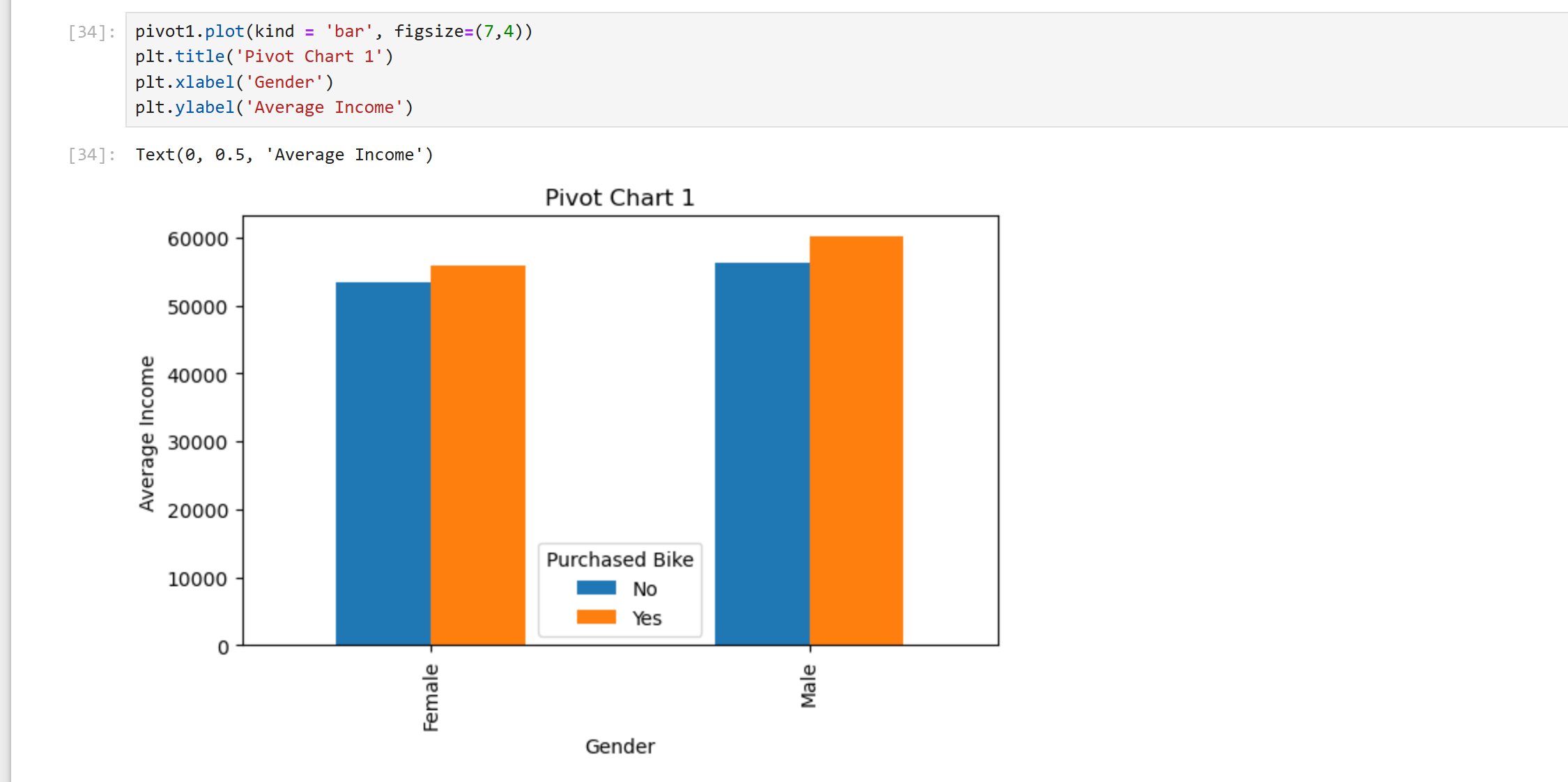
The age column was classified into age brackets in a different column with adolescents (0-30), middle age (31-54), and old (55+) using the code



**Data Analysis/Visualization**

1. The average income of each gender and their bike purchase history was determined by first grouping and then finding their average using the code then the chart was plotted;

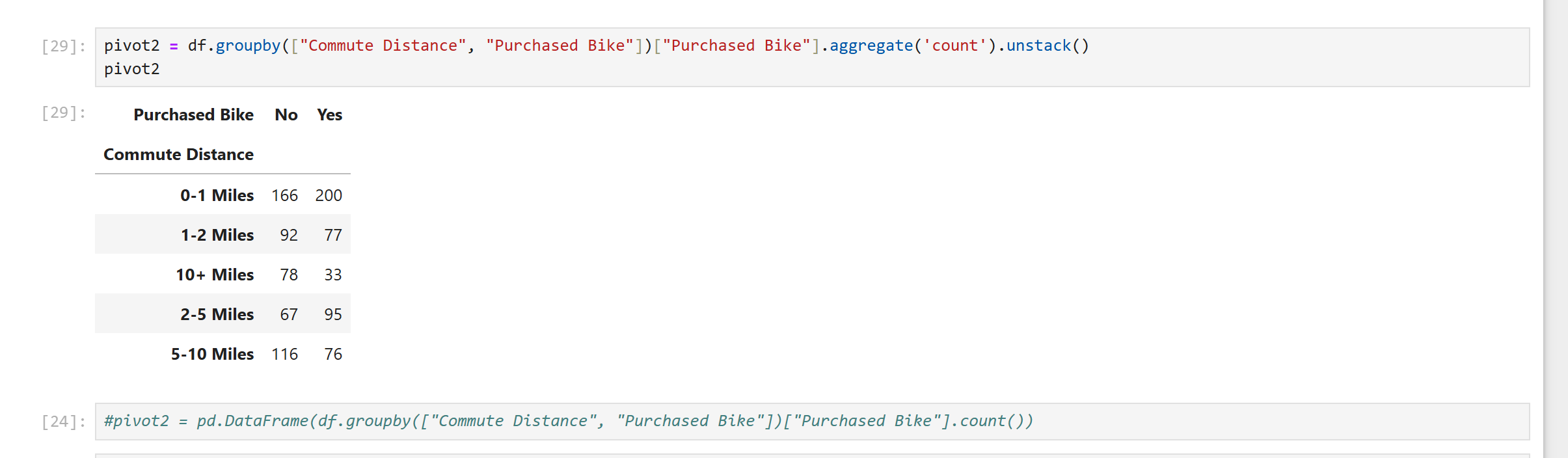


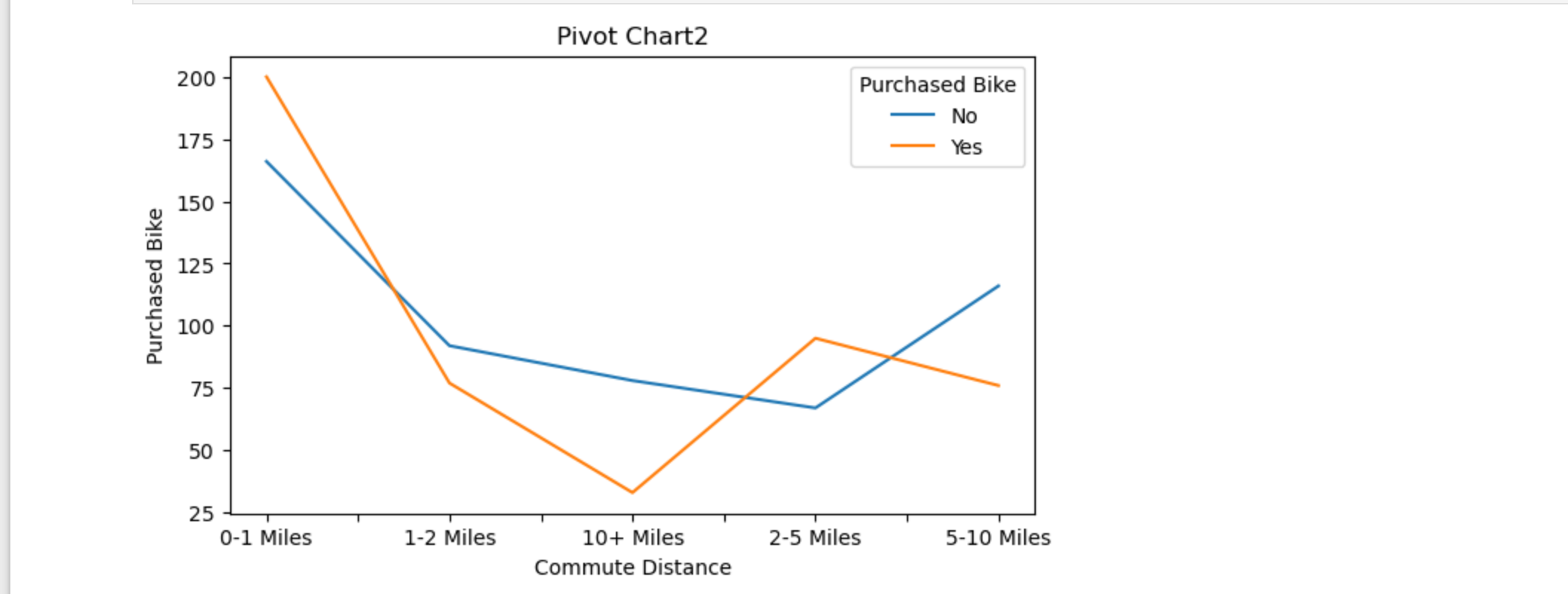


It showed that in the female category, those who purchase bike earn more and this is also seen in the male category

2. Distance travelled by commuters and bike purchase

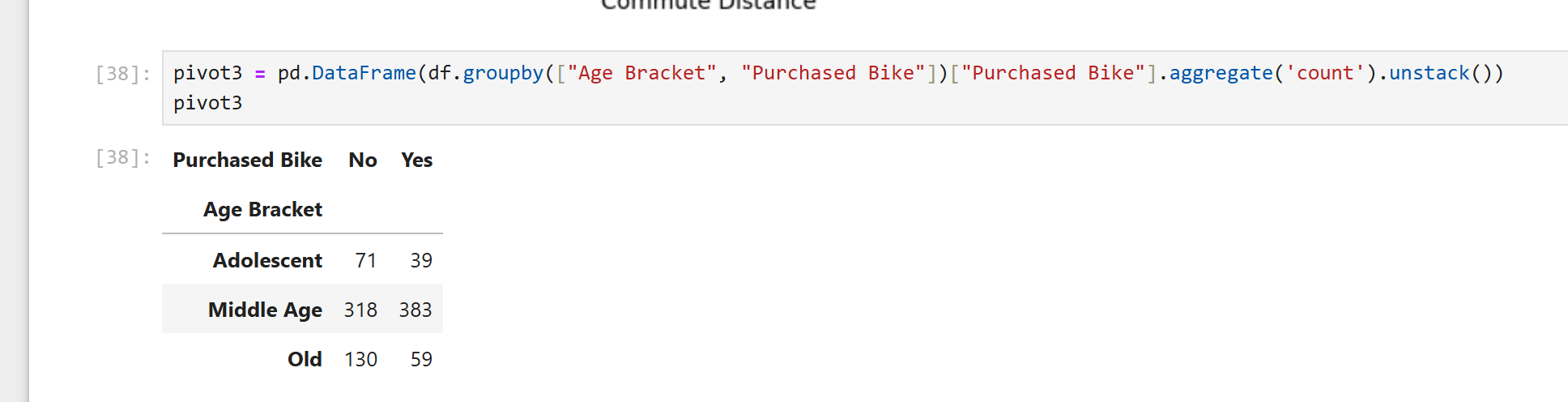
This was determined by grouping the Commute Distance and the Purchased Bike column using the code “df.groupby” and the chart was plotted;

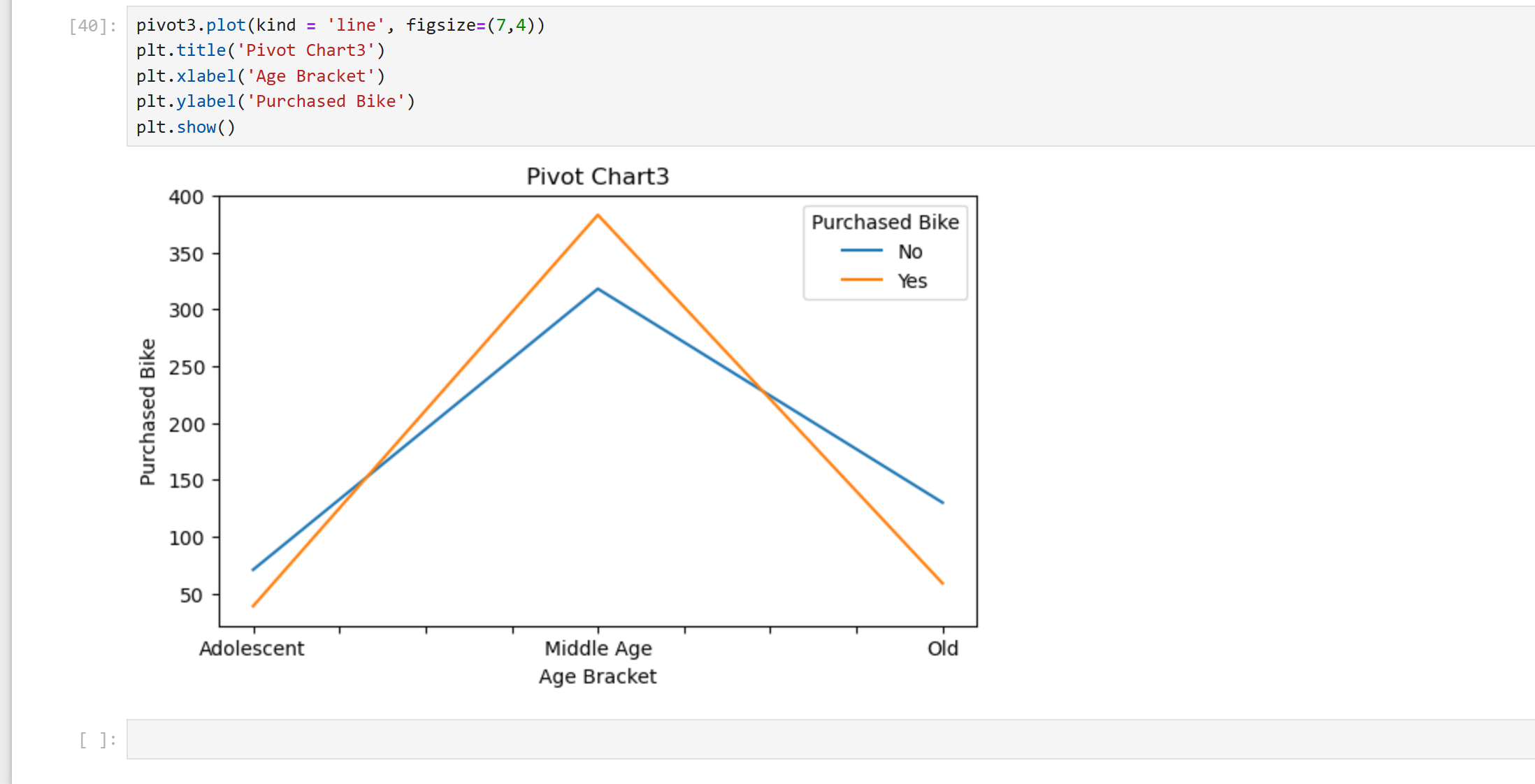




1. Age bracket and bike purchase.

The columns of Age bracket and bike purchase was grouped using the “df.groupby” and the chart was plotted as seen below

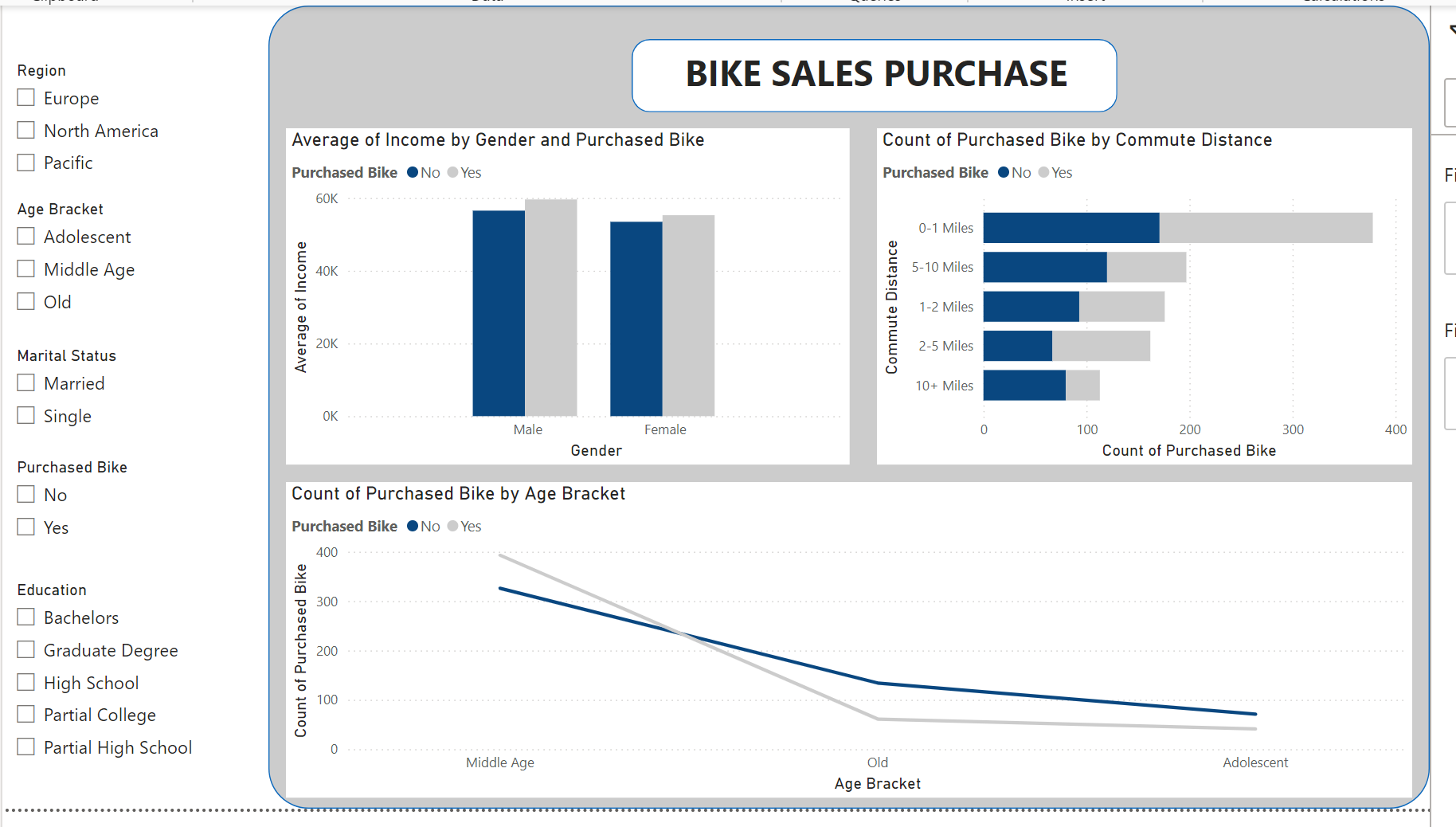




The chart reveals that there the middle aged group both have the highest number of persons who purchased bike and who also did not purchase bike.

**Reporting**

The report was done using the Power BI visualization tool and the questions were answered thus;



**Conclusion**

Based on this analysis, we’ve been able to answer some basic business question, which include:

* Which age bracket should be given less attention by bike manufacturers
* How European Middle Age Singles on bike purchase behave
* Which academic qualification earns more

**Recommendation**

* Less attention should be given to the Middle Age group by the Bike Manufacturers because it is discovered, from the chart, that they usually buy more. Hence, more marketing focus should instead be on the adolescent and old age bracket because their count of bike purchase is relatively low
* For European middle age singles on bike purchase, these people are observed to travel lesser miles, even with a purchase bike count of about 70. Hence, they’re less commuters and would likely spend time at home. There’s possibility of them working from home and only need bikes to get home items nearby
* The Graduate degree has the highest average income of $70,000. Bike Manufacturers should target them more, especially singles from the Pacific Region with the highest average income.