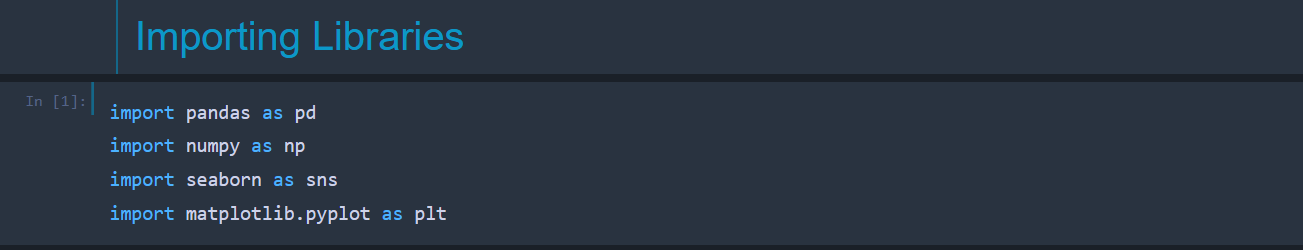
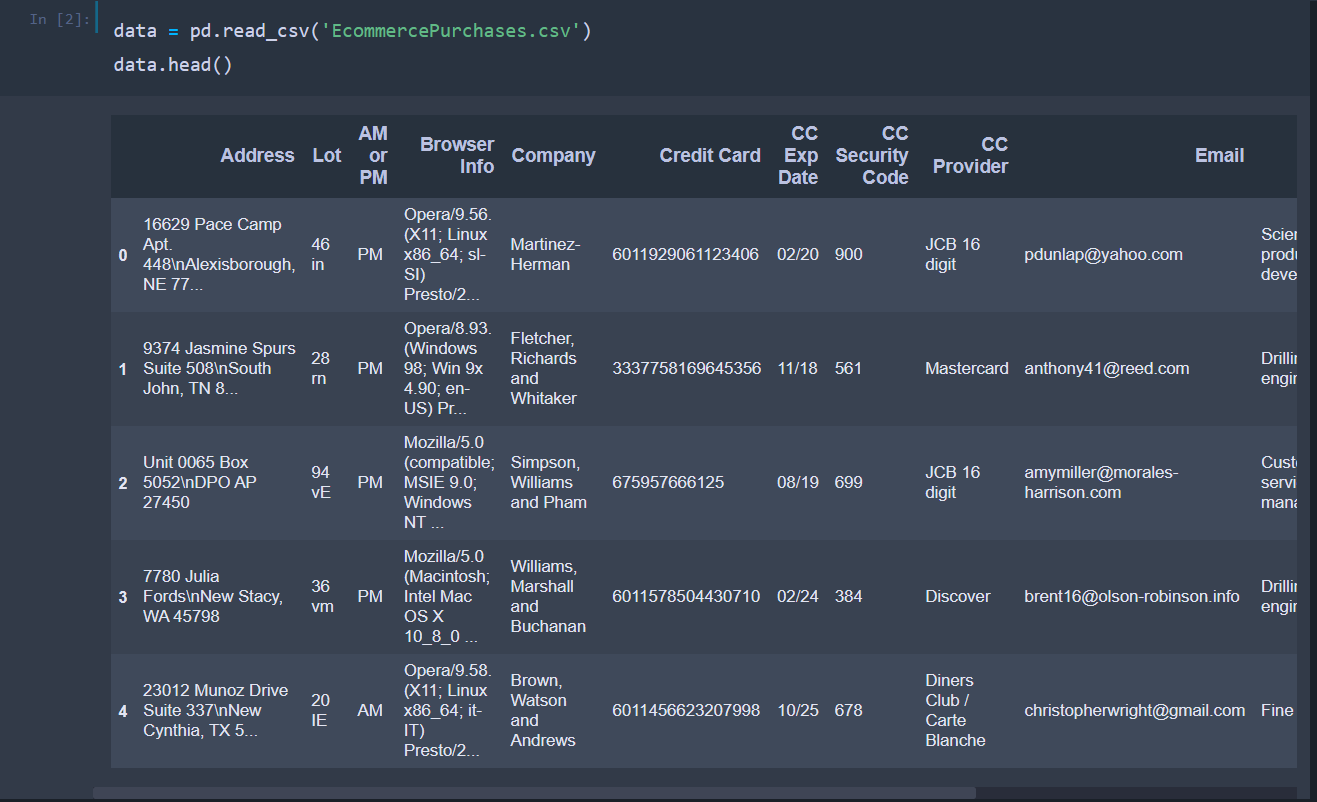
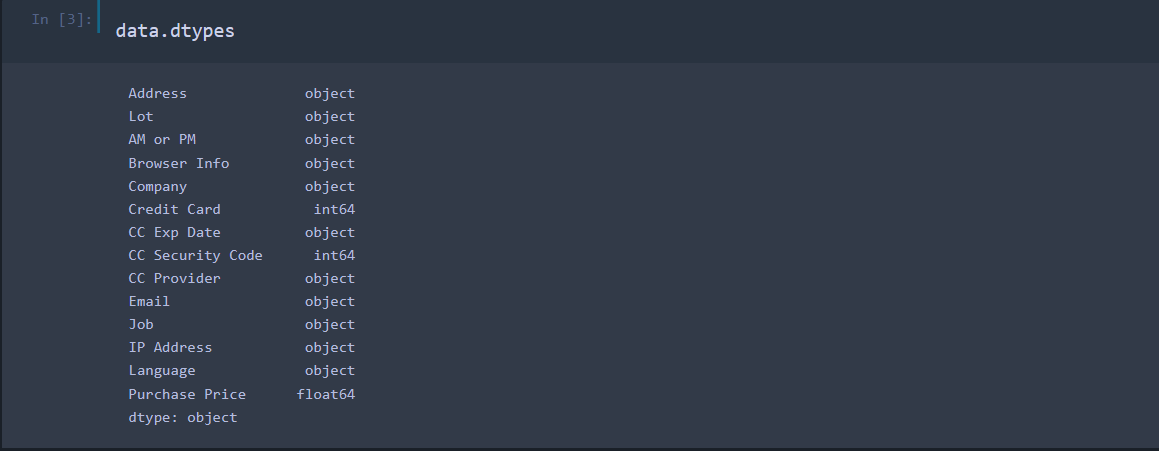
Ecommerce Data Analysis

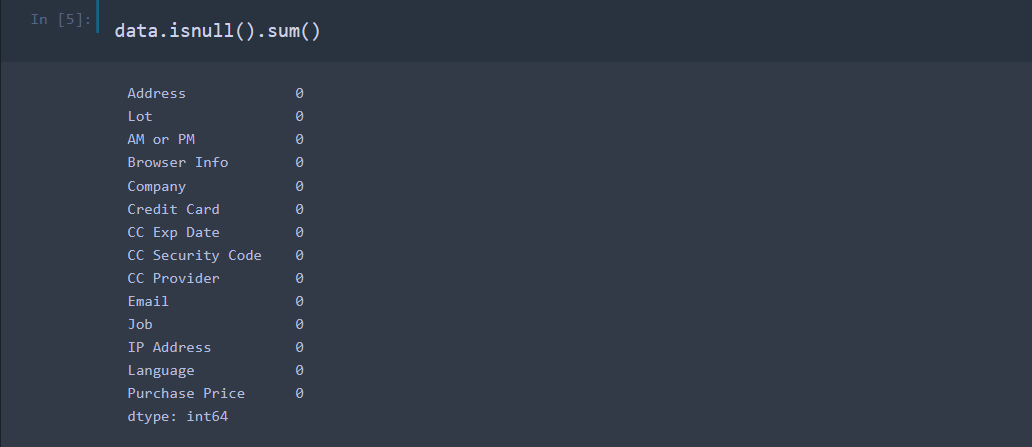
We all would have used **ecommerce sites** like **Amazon, Flipkart, Myntra, Snapdeal** etc at one point or the other in order to buy something. In this section we will be having a look at the dataset which contains information regarding customers who have done shopping on an **ecommerce site.** We will not be using any models in this section since we will be just analysing the data, in-order to draw insights from it and have a clear understanding of the available data.

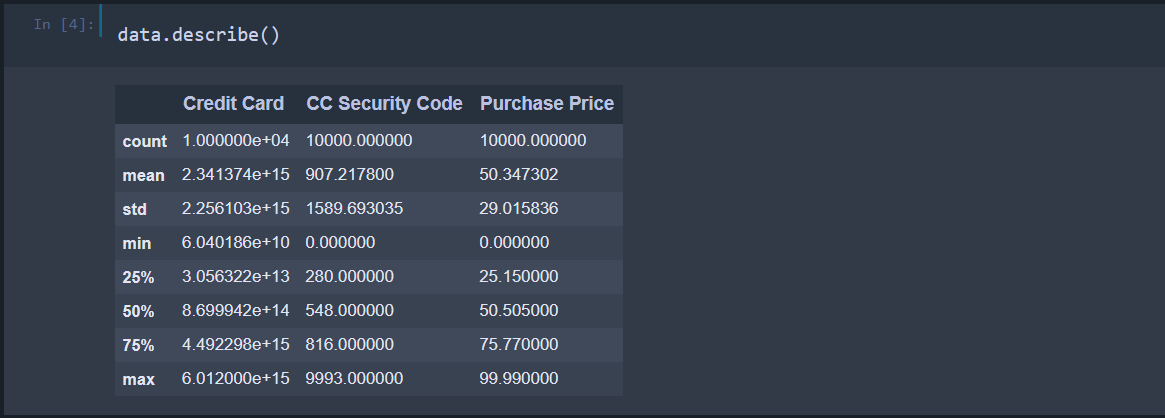
The first step remains same in almost all the problem statements i.e., to import the necessary **libraries** and **modules**. We import the basic libraries like NumPy, Pandas, Seaborn and Matplotlib.

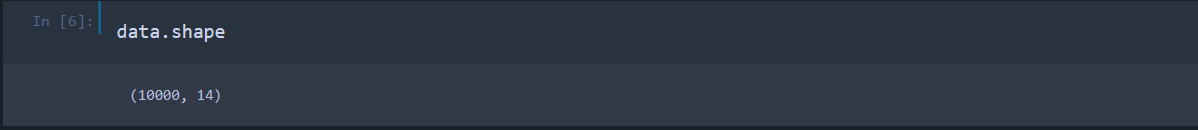
****After having the necessary modules and libraries in our project, we move ahead and import the dataset which we will be using in this project. The dataset contains many different features like **Address, Browser Info, Credit Card, Expiry date of credit card** etc. If we have a look at the data types of the values stored in the columns, we can see that most of the columns contains values of the type **object** and some columns have **integer** and **float** values.



Let’s analyse the dataset to see if we have any **Null** values present in the dataset which might affect the process of data analysis. We can use the **isnull()** method for this purpose.

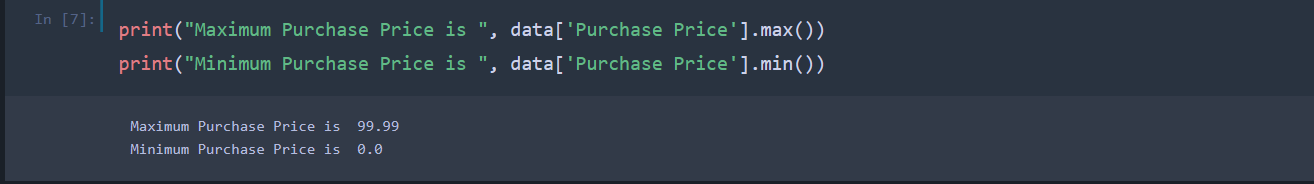


In order to have a better understanding about the numeric data present in the data, we can use the **describe** method on the dataset. This gives us insights about various properties like **mean, max, min** etc of various columns present in the dataset.

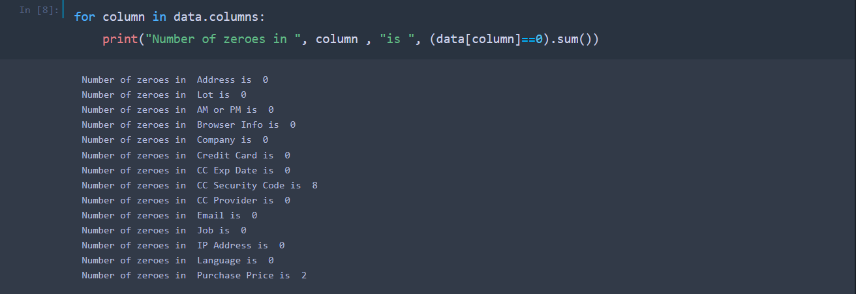


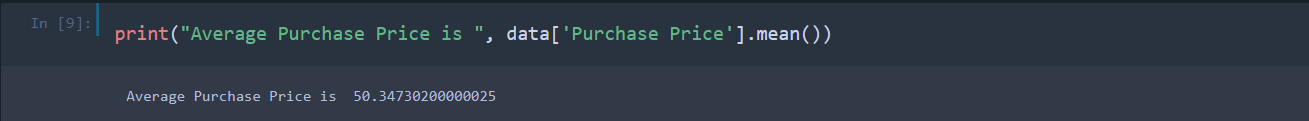
The output of the above code indicates the number of **rows** and **columns** present in the dataset. We can see that there are 10,000 rows and 14 columns (features).

Let’s move ahead and find out some interesting facts about the dataset. To start with, let’s look at the **maximum** and **minimum** purchaseprice.

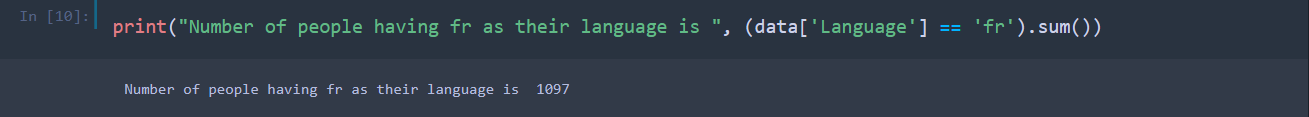


The **minimum** shopping price is shown as 0, which may sound absurd since shopping price cannot be 0. To get a clear understanding of this, we can have a look at the number of **zeroes** present in each column. The zeroes can be replaced with some other value like **mean** of the values present in the column where the 0 is present.

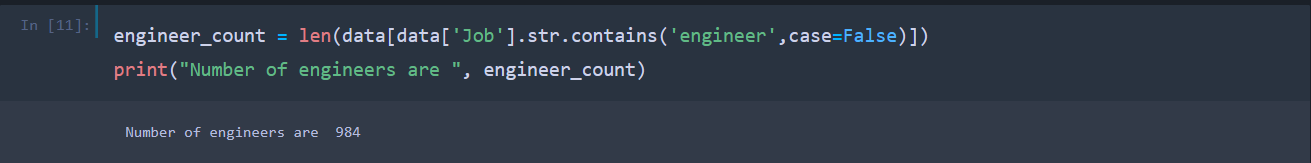
We can even look at the **average** purchase price using the **mean()** method on the **Purchase Price** column.

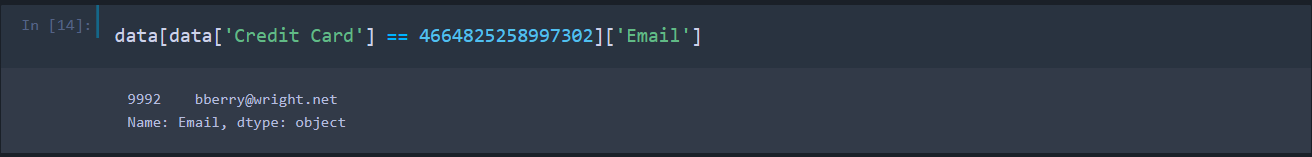


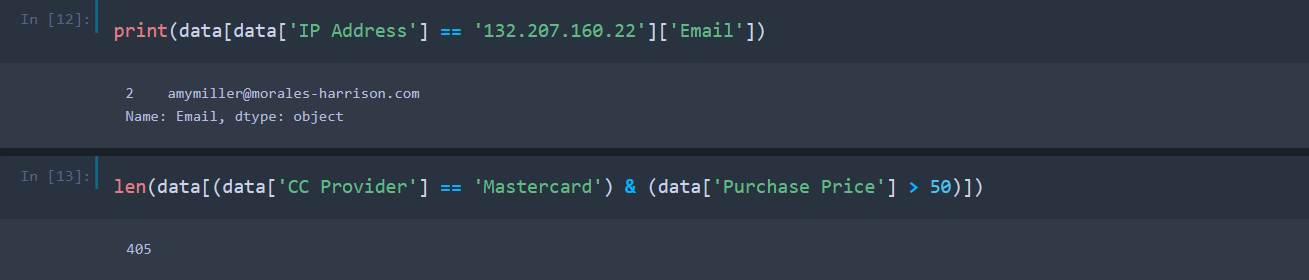
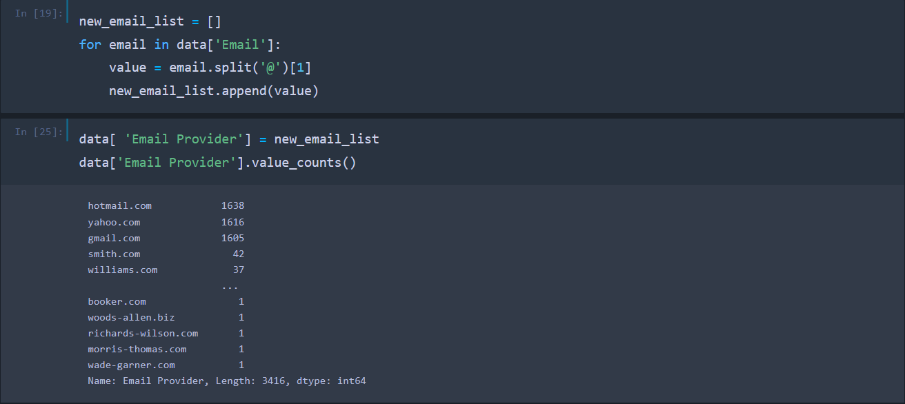
Next, let’s have a look at the **number of people** who speak French i.e., number of people who have **fr** in their language. **Fr** indicates **French**.



We can even get the count of number of engineers in the dataset by analysing the values present in the **Job** column.



We can also get the details of an individual customer based on certain feature like **IP Address**, **Credit Card Number**, **Credit Card Provider** etc.

Let’s look at the 5 most popular email providers which are used by the customers. The first 5 email providers have the highest number of customers.