Total number of rows in raw data : 13647309

Null value count for each column:

|  |  |
| --- | --- |
| **Column Name** | **Number of Null Values** |
| fecha\_dato | 0 |
| ncodpers | 0 |
| ind\_empleado | 27734 |
| pais\_residencia | 27734 |
| sexo | 27804 |
| age | 0 |
| fecha\_alta | 27734 |
| ind\_nuevo | 0 |
| antiguedad | 0 |
| indrel | 0 |
| ult\_fec\_cli\_1t | 13622516 |
| indrel\_1mes | 149781 |
| tiprel\_1mes | 149781 |
| indresi | 27734 |
| indext | 27734 |
| conyuemp | 13645501 |
| canal\_entrada | 186126 |
| indfall | 27734 |
| tipodom | 0 |
| cod\_prov | 0 |
| nomprov | 93591 |
| ind\_actividad\_cliente | 0 |
| renta | 2794375 |
| segmento | 189368 |
| ind\_ahor\_fin\_ult1 | 0 |
| ind\_aval\_fin\_ult1 | 0 |
| ind\_cco\_fin\_ult1 | 0 |
| ind\_cder\_fin\_ult1 | 0 |
| ind\_cno\_fin\_ult1 | 0 |
| ind\_ctju\_fin\_ult1 | 0 |
| ind\_ctma\_fin\_ult1 | 0 |
| ind\_ctop\_fin\_ult1 | 0 |
| ind\_ctpp\_fin\_ult1 | 0 |
| ind\_deco\_fin\_ult1 | 0 |
| ind\_deme\_fin\_ult1 | 0 |
| ind\_dela\_fin\_ult1 | 0 |
| ind\_ecue\_fin\_ult1 | 0 |
| ind\_fond\_fin\_ult1 | 0 |
| ind\_hip\_fin\_ult1 | 0 |
| ind\_plan\_fin\_ult1 | 0 |
| ind\_pres\_fin\_ult1 | 0 |
| ind\_reca\_fin\_ult1 | 0 |
| ind\_tjcr\_fin\_ult1 | 0 |
| ind\_valo\_fin\_ult1 | 0 |
| ind\_viv\_fin\_ult1 | 0 |
| ind\_nomina\_ult1 | 0 |
| ind\_nom\_pens\_ult1 | 0 |
| ind\_recibo\_ult1 | 0 |

Datatype for each column variable:

|  |  |
| --- | --- |
| **Column Name** | **DataType** |
| fecha\_dato | timestamp (nullable = true) |
| ncodpers | double (nullable = true) |
| ind\_empleado | string (nullable = true) |
| pais\_residencia | string (nullable = true) |
| Sexo | string (nullable = true) |
| Age | string (nullable = true) |
| fecha\_alta | timestamp (nullable = true) |
| ind\_nuevo | string (nullable = true) |
| antiguedad | string (nullable = true) |
| Indrel | string (nullable = true) |
| ult\_fec\_cli\_1t | timestamp (nullable = true) |
| indrel\_1mes | string (nullable = true) |
| tiprel\_1mes | string (nullable = true) |
| Indresi | string (nullable = true) |
| Indext | string (nullable = true) |
| conyuemp | string (nullable = true) |
| canal\_entrada | string (nullable = true) |
| Indfall | string (nullable = true) |
| tipodom | string (nullable = true) |
| cod\_prov | string (nullable = true) |
| nomprov | string (nullable = true) |
| ind\_actividad\_cliente | string (nullable = true) |
| Renta | double (nullable = true) |
| segmento | string (nullable = true) |
| ind\_ahor\_fin\_ult1 | integer (nullable = true) |
| ind\_aval\_fin\_ult1 | integer (nullable = true) |
| ind\_cco\_fin\_ult1 | integer (nullable = true) |
| ind\_cder\_fin\_ult1 | integer (nullable = true) |
| ind\_cno\_fin\_ult1 | integer (nullable = true) |
| ind\_ctju\_fin\_ult1 | integer (nullable = true) |
| ind\_ctma\_fin\_ult1 | integer (nullable = true) |
| ind\_ctop\_fin\_ult1 | integer (nullable = true) |
| ind\_ctpp\_fin\_ult1 | integer (nullable = true) |
| ind\_deco\_fin\_ult1 | integer (nullable = true) |
| ind\_deme\_fin\_ult1 | integer (nullable = true) |
| ind\_dela\_fin\_ult1 | integer (nullable = true) |
| ind\_ecue\_fin\_ult1 | integer (nullable = true) |
| ind\_fond\_fin\_ult1 | integer (nullable = true) |
| ind\_hip\_fin\_ult1 | integer (nullable = true) |
| ind\_plan\_fin\_ult1 | integer (nullable = true) |
| ind\_pres\_fin\_ult1 | integer (nullable = true) |
| ind\_reca\_fin\_ult1 | integer (nullable = true) |
| ind\_tjcr\_fin\_ult1 | integer (nullable = true) |
| ind\_valo\_fin\_ult1 | integer (nullable = true) |
| ind\_viv\_fin\_ult1 | integer (nullable = true) |
| ind\_nomina\_ult1 | string (nullable = true) |
| ind\_nom\_pens\_ult1 | string (nullable = true) |
| ind\_recibo\_ult1 | integer (nullable = true) |

Exploratory Analysis:

1. Number of Female Customers (46%) are less than Number of Male customers (54%)
2. Last Date being Primary Customer the max value is 2016-05-30
3. Individual Customers (58%) are more than VIP (6%) and Students (36%) of total number of customers
4. Customer type has values of data type String, Double and Int
5. Average gross income of VIP customers is more than Individual and Student customers
6. Average gross income of female is lesser than male
7. Male with Individual customer types are maximum
8. Count of Guarantees is maximum for male individual type customers with average gross income approx. 139593
9. Number of savings account maximum for male individual type of customers with average gross income 139593
10. Number of derivada account maximum for male individual type of customers with average gross income 139593
11. Further analysis shows that all the products are consumed maximum by male individual type of customer