

1. Gender column has two categories M and F so encoding it
2. The age is in the format of xx-xx, so the outcomes can be
   1. avg
   2. Segment into categories (teen, working, adults etc) and then encode it
3. Ocupation column is from 0-20 making it an array of 21
   1. Group them and encode
   2. Depending on the
4. Drop userid and prodcut id
5. Gender ko ohe male(1) female (0)
6. Age class ko make separate columns and drop first
7. Categies ko ohe and drop first
8. Stay use regex and only the first digit
9. Merging train and test data
   1. Df= df\_train.append(df\_test)
10. df.info, df.describe, user\_id can be deleted
11. Gender =ohe, dropfirst
12. Age gets ordinal training bcus 0-17 don’t shop much (label encoding can be done ig)
13. City categories ohe, dropfirst
14. Replace prod\_cat null val with mode
15. Stay in current years
16. Pairplot