

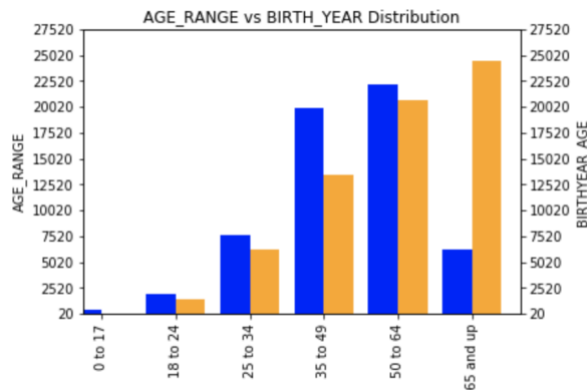
APP User File Summary

Data information:

- Number of distinct APP_USER_ID and email address: **664,632**
- 663,421 first names, 1211 null value, more than 94,000 shows 'Globe' as their first name.
- 662,267 last names
- Number of app users who voluntarily give their gender is 108,308 (16%)

Model performance (Using App users' first name to infer gender):

- Accuracy 81% - The prediction results matched 544076 (81%) of user's record and with the ratio of Male: Female = 64:36
- Google email domain takes account of more than 40%, followed by Hotmail domain with 10 %, the top 5 corporate domains are Rogers, Shaw, Sympatico, Telus and Bell.
- AGE and AGE_RANGE column have high discrepancy. Number of available AGE is 3578, number of available AGE_RANGE is 58283. AGE_RANGE is more reliable.
- AGE_RANGE column would be more accurate compared with BIRTH_YEAR distribution, 38% comes from 50 to 64 age group, followed by 35 to 49 age range (34%).



- One evident difference is that 65 and up age group takes account of 42% by BIRTH_YEAR, however, for AGE_RANGE is merely 10%.