



Clustering Analysis

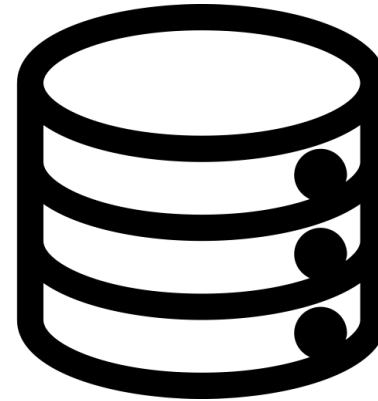


Context

1. Data preparation and cleaning steps
2. Is the data suitable for clustering?
3. Clustering
4. Future work

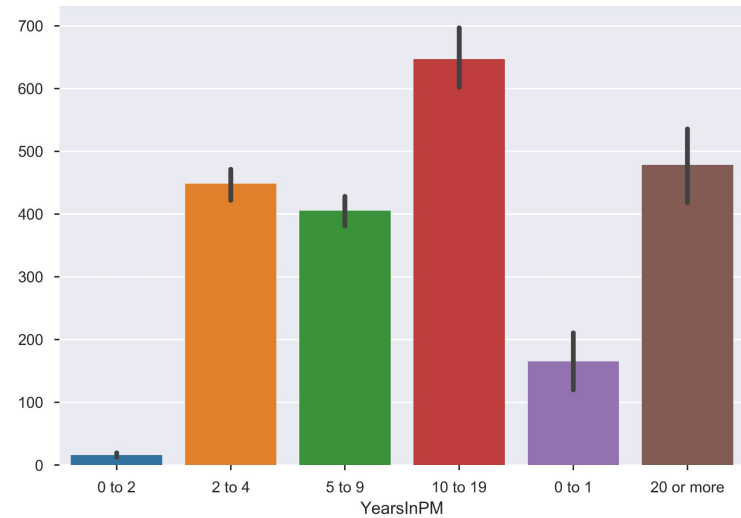
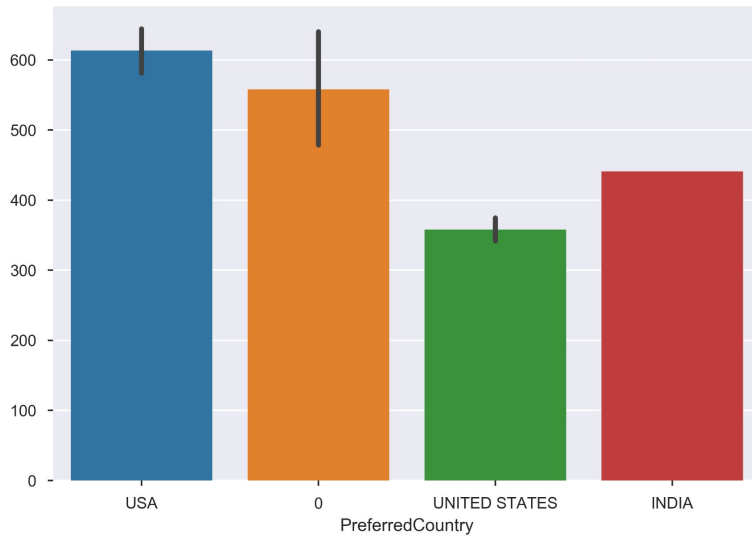
Data Prep and cleaning step

```
IsRegisteredUser      0
YearsInPM              865
PreferredCountry       0
Male                   0
Female                 0
RegisteredUserCreateDate 0
PMCOMUserFirstVisit    0
PMCOMUserLastVisit     0
PreferredRegion        334
PreferredStateProvince 108
HomeChapterName        482
PMCOMUserLastUpdate    0
PMCOMUserCompanySize   512
IsRegisteredUser      0
dtype: int64
```

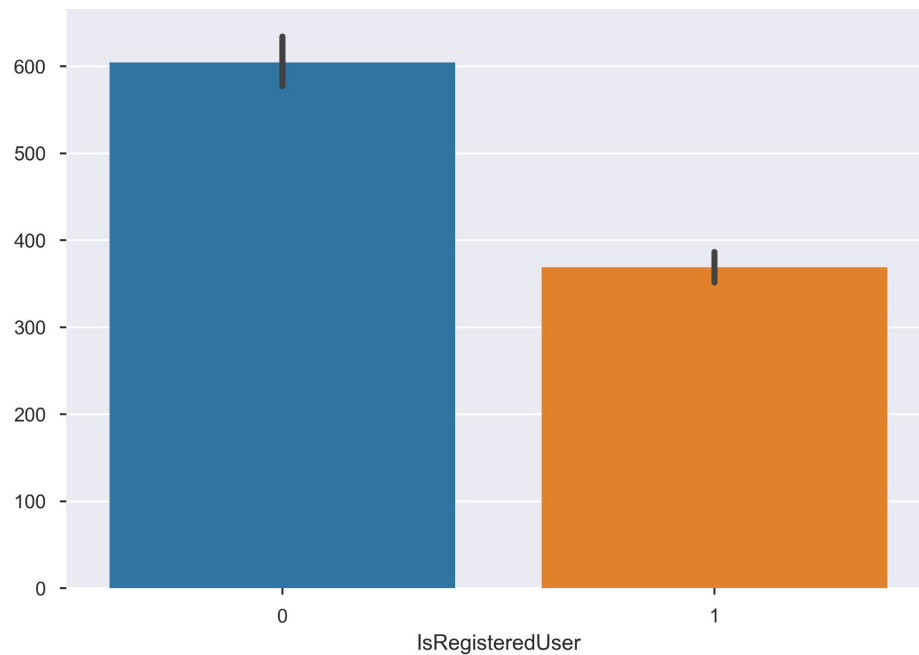
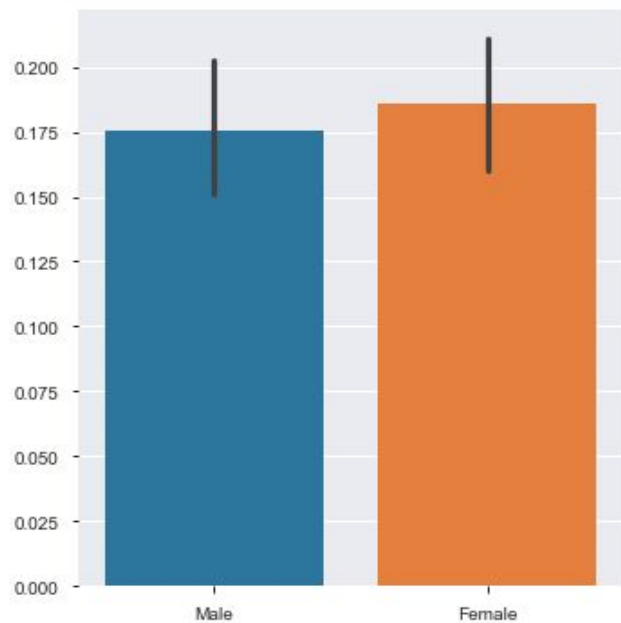




Analysis



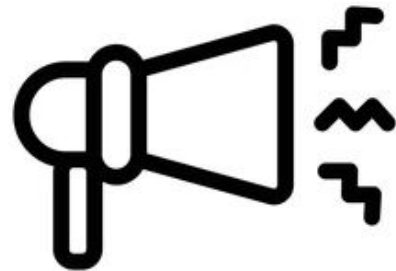
Analysis



Is the data suitable for clustering?

No, i wouldn't use all the data for clustering because some of the features are noise and contain no vital information.

I will drop features that don't contribute to the behaviour of a user



Clustering

I used Elbow Method For Optimal number of K and the optimal number is 5.

