1. Download Data Facebook_comments.csv (zip file and unzip from LMS (module : linear models)
2. Read data to R using read.csv
3. Use function Create dummies to create dummy vars for column page_category with frequency cutoff 200
4. For columns 'Post Published Weekday'and'Base Date Time Weekday' replace ['Sunday','Monday'. . . . .] with [1,2, ...]
5. Instead of creating dummies for datetimetype columns its better to represent them with values which are cyclic in nature themselves . Create sin and cos columns for both the columns mentioned in (6) as follows :

```
df$col_sin=sin(2*pi*df$col/7)
df$col_cos=cos(2*pi*df$col/7)
```

6. Remove columns 'Post Published Weekday'and'Base Date Time Weekday' from the data
7. Break data into two parts(80/20) randomly
8. Run lm for your 80% data, and remove vars on the basis of vif with cutoff 5 , one by one
9. Run lm again and drop vars on the basis of p-values with cutoff 0.05 (see if you can use step function to reduce manual process here)
10. Check the performance of your model on 20% of the data (rmse)