```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

```
#Нэр үг
import csv
dataset_noun = []
with open('/content/drive/MyDrive/Машин сургалт/lab_03/mn_noun(нэр үгийн хувилал).csv', 'r')
  reader = csv.reader(file)
  for row in reader:
    l=[]
    text = row[0].split("\t")
    for i in range(7):
      if i != 1 and i != 4 and i!=3:
        l.append(text[i])
    t = l[1]
    l.remove(t);
    l.insert(0,t);
    dataset_noun.append(l);
pd.DataFrame(dataset_noun)
```

	0	1	2	3
0	нэрийн	нэр	ийн	-
1	нэрд	нэр	д	-
2	нэрийг	нэр	ийг	-
3	нэрээс	нэр	ээс	-
4	нэрээр	нэр	ээр	-
...	...	...	...	...
14587	залгааг	залгаа	г	-
14588	залгаанаас	залгаа	аас	н
14589	залгаагаар	залгаа	аар	г
14590	залгаатай	залгаа	тай	-
14591	залгаагаа	залгаа	аа	г

14592 rows × 4 columns

```
#үйл үг
import csv
dataset_verb = []
with open('/content/drive/MyDrive/Машин сургалт/lab_03/mn_verb(үйл үгийн хувилал).csv', 'r')
   reader = csv.reader(file)
```

```
    for row in reader:
     l=[]
     text = row[0].split("\t")
     for i in range(5):
       if i != 2 and i != 4:
         l.append(text[i])
       if i == 4:
         l.append("-");
     t = l[1]
     l.remove(t);
     l.insert(0,t);
     dataset_verb.append(l);

pd.DataFrame(dataset_verb)
```

	0	1	2	3
0	түгдэг	түгэх	дэг	-
1	түгэв	түгэх	в	-
2	түгжээ	түгэх	жээ	-
3	түгмээр	түгэх	мээр	-
4	түгээсэй	түгэх	ээсэй	-
...	...	...	...	...
15546	хэлмэгдтэл	хэлмэгдэх	тэл	-
15547	хэлмэгдмэгц	хэлмэгдэх	мэгц	-
15548	хэлмэгдвэл	хэлмэгдэх	вэл	-
15549	хэлмэгдэнгээ	хэлмэгдэх	нгээ	-
15550	хэлмэгдэлгүй	хэлмэгдэх	лгүй	-

15551 rows × 4 columns

```
#монгол үгнүүд
import pandas as pd
file = pd.ExcelFile('/content/drive/MyDrive/Машин сургалт/lab_03/mon-words.xlsx');
tmp_all = file.parse(file.sheet_names[0]).to_dict()
dataset_all = [];
tmp_all.pop('Category', None)
tmp_all.pop('POS', None)
tmp_all.pop('word_id', None)
for i in range(5000):
  tmp = dict.fromkeys(tmp_all,0);
  for key in tmp_all:
    tmp[key] = tmp_all[key][i]
  dataset_all.append(tmp);

pd.DataFrame(dataset_all)
```

```
pd.DataFrame(dataset_all)
```

	lemma	Derivation	Source word	Derived POS	Source POS
0	нэр	NaN	NaN	NaN	NaN
1	отряд	NaN	NaN	NaN	NaN
2	цутгах	NaN	NaN	NaN	NaN
3	хуйлрах	-рах	хуйлах	verb	verb
4	халих	NaN	NaN	NaN	NaN
...	...	...	...	...	...
4995	чимэглэл	NaN	NaN	NaN	NaN
4996	бүтэлгүй	NaN	NaN	NaN	NaN
4997	мөрөгцөг	NaN	NaN	NaN	NaN
4998	сэдэл	NaN	NaN	NaN	NaN
4999	залгаа	NaN	NaN	NaN	NaN

5000 rows × 5 columns

```
mgl_words = []
for dict in dataset_all:
  tmp = []
  tmp.append(dict['lemma']);
  tmp.append(dict['Source word']);
  tmp.append(dict['Derivation']);
  tmp.append('-')
  mgl_words.append(tmp);
mgl_words


#Бүх үгээ нэгтгэсэн нь

final_dataset = dataset_noun + dataset_verb + mgl_words
pd.DataFrame(final_dataset)
```

	0	1	2	3
0	нэрийн	нэр	ийн	-
1	нэрд	нэр	д	-
2	нэрийг	нэр	ийг	-
3	нэрээс	нэр	ээс	-
4	нэрээр	нэр	ээр	-
...	...	...	...	...
35138	чимэглэл	NaN	NaN	-
35139	бүтэлгүй	NaN	NaN	-
35140	мөрөгцөг	NaN	NaN	-
35141	сэдэл	NaN	NaN	-
35142	залгаа	NaN	NaN	-

35143 rows × 4 columns