

M03-Exercise-hr2ee

September 11, 2022

1 Mod 3 Excercise

Max Ryoo (hr2ee)

```
[1]: import re
import math
import pandas as pd
```

```
[2]: %matplotlib inline
```

1.0.1 Part 1

Use the Simple Polynesian alphabet (as found in the Schütz and Manning reading and printed below)

```
[3]: letters = pd.Series('p t k a i u'.split())
letters
```

```
[3]: 0    p
1    t
2    k
3    a
4    i
5    u
dtype: object
```

```
[4]: # weights = pd.Series([1/8,1/4,1/8,1/4,1/8,1/8])
weights = [1/8,1/4,1/8,1/4,1/8,1/8]
weights
```

```
[4]: [0.125, 0.25, 0.125, 0.25, 0.125, 0.125]
```

```
[5]: pd.DataFrame({'Letter':letters, 'Probability':weights})
```

```
[5]:   Letter  Probability
0      p         0.125
1      t         0.250
2      k         0.125
```

3	a	0.250
4	i	0.125
5	u	0.125

1.1 Part 2

Create a text by randomly selected seven letter words based on the weights (separated by spaces)

```
[6]: generated_words = []
for i in range(100):
    word = letters.sample(7, replace=True, weights=weights).str.cat()
    generated_words.append(word)
print(*generated_words)
```

pttkptu atiuatp pakptaa ttkuatp paaaauk pukuuua utaipia taautta ktaautt paaputa
 uaiaaia ttktiap kakiauu tpaptuk aituttt akaataa tutuaia iitaaik aautau ktuauta
 kutkttp tkauiaa kpkaipu aauktik ptiauaa atittap tutatti uaiapii ttattti attaaat
 tauiat tikuaiu tkpaipa uttitta ututtkt ttakpaa kitkkti ptatapa taptkpi piktata
 aaaattp ttpipta kpaitiu atptatt uipptuu ikpaiaa aaueaat utpiitu akkkutt akkukpa
 upatuka tkikppk puitkkp iatttui attakaa aiataau tiputka aattkkp aptapti iataata
 itpuiit pipaaap ttuauit aatkuua pttkppa kiitait uuitatk aakiitp tatpttt pkkaati
 ttptuta pttiiia iauakui uaukptt auapaat tatitti tiaptka kuataat tkiuaat atuaaup
 aakiptt iaakpat kkpptkp ttupatt ataakuu tutpipt akttiaa upappta ipapuak pttiapt
 aptattt katkkki ptttppp atttatt iutiitt tuakapa tpuktau kaaitut apaptpp aikuaaa

1.1.1 Part 3

Using the unigram term frequencies in that reading as weights, write a simple text generator from the Babel library you just created.

When you are done, submit your source code to the appropriate Assignment in Collab.

```
[7]: def get_message(m_len = 7, weights=[]):
    if len(weights) == 0:
        p = 1/letters.shape[0]
        weights = [p for i in range(letters.shape[0])]
    msg = letters.sample(m_len, replace=True, weights=weights).str.cat()
    return msg

def print_page(n_pages = 1, weights=[]):
    for i in range(n_pages):
        for i in range(40):
            print(get_message(80, weights))
        print()
        print('-'*80)
        print()
```

```
[8]: print_page(2, weights)
```

paktuatttkaaatktiaukutttupuaaktipktptpttatpuuaatpuaiuikapauatpitkippputitpipiakaa
tuuutptttttaktaataiaiapuaaappkuappakaitatakpttkaaataakappaattputapattiuiakak
ktitpataaukappttutiaptkaktakaaauitptkapktpiiutktttattaiakukiutaiktppptiatitakia
atuapaaituitttiiikaaiuaiuauikautkiiuattaaauaaktttppappuiaaiikkuatiitptkikiiaaa
pptatakttitppipaatuapaupppappakitpkaiiappuiitputitapkuuapiuiiiktpatituakaitiaat
tuttkukkaauiiitktkikaikiaipaikiaapaipuaaituptiaaitauiaataatiuattkattkuatkaptuika
tpttuatiikauppauiiktutuuuikutptupaptkitiiapkaaiitakitpaaautitpuptaupkaaaatupppkp
auiaattaatkttaatauapakauaktkptkakppptatukaiitauiakpiupaakuittaitakauipatipatkta
itapkuatiuutkaiikaapkiktutkukuuaaukukatupptiitatapuupaukuuitiaiputiaauatkipttp
iiaipapikpattatuutkaaupiatkptutupakpaktituaatppiputtpaitaiatkttkkataakaaaititup
tatataaaaitaakiukapiaipkptapaakppikiataptkaakpkktautptiaaiaakpstkptpapakaitatka
utupikuaaittkutaupkpiutatkkakukuaiaitaitukauptttitkkpukkaiaauktpaakkukkpuaapuaki
pkptapikppptuuaapipttapappattkauuitupikattatatutakutuktukaiaaakukatttattutaiapt
aatupptkaupaiktpikaaaipaatautttkputatauaattaauttkutiiappiptkktuipattipatiipauiai
apattaatpppkauaaaaatttttaitatatkututputupaaiuitpkaaaaakaaupatktttuppakkkkt
uttpuauuiipaaaakkatkaakattauiuaaittipuaapaptattkkkuutkiiakpuitktuuiiaaaaapkttki
tikakiauuppkutiktupkaukuptppiitttatatuatuapatpttpauatkiauiuktktpakiaiktaaitaku
tppupaaittipttakuuaktiiaiktptaiutpitkaikiapaipapatkuatutitkatukuiaattatapapiiaua
puataaukktupapaakpaaptpakttiatuappttitkkkauaptuiiktikpktttuuaaiikapakpkpauat
iaitpaataatukutapaapaukkkautiiapptkutputkpatpputuitpiaiptptaiupipatkpaiakkikiau
ipatatpttpttuttpkiuiiatakptptktiukpittaututititkikaututaaaupttauaaaupkppukktta
atkpaauppipttttuiuuiatiapiaatupputpatkipttttaupkpttpttittituitaaapaktpkuitiutkt
iituaaauiuptpattipuuupptpuiipapkktptitutataaiittutkkiaattkakutpppkptiptaitaitata
ittpaakaaiiakipktpipitpatiuktptkuuaitiaapktitutkuaktiktaiiukipputuuutptakata
uaatuiiaapkpataatiippatikpatauuupkkptikkutitpittaitiakupitttiuktautkakktikakaap
puatatuattutataitpuauiputaapaatpatkatiptapakiaktkuuuuptpaauuaauikuapitpikppuii
uauiaaaaapaptatkikptkkikiauaptakkkaitaiitipiutaiatatakutukpkttttuuputkktpptipa
takktaaaakutatuiuaaattiptuaiktkauttatauaapttaaptitauttauitupakkaapuuttapaiaatkt
puttattkpiapatppkpitapauatappptpttikupitiiittkaitutppktatkppitkukttaatkpkkikaiuati
ittkptpttiitkukaatauaataturkpkkpiapakauitaptaktptkutuktaapaapiakttaituatptattat
aauiupputitattpttauptikukatitttkaatttatuaaakittauuataatpttattkttauiiaitptatittut
uputtaputuktukaiaiptiatatpapaaaattptttaukiiapapiuauattauukuukaaapikukpkitutuaakp
aupaupupauttkiktktipakaaautaturkpkkptatuptkiptaakukktkuiiikakkuaatupaaatautkput
autttitppuptptaaakpiitapttupuptkaaauiuktkatautkttkitttkpukkapktuautptpstkputkta
utkippaakakaptaaptatuaauukktutttataturktaiuttatapatutitktattptaktpapuktautauaaiikuttu
ptttapupuktaapukaaatakiptttaakppkapappatuaupiuptuitittitattutttaitpttaiputakktaka
ktttaaiupuiuaattiauaakitiapiuaaakupktkkpptptuututiktittitaatattaupatuptattaitaitpk
atttaakaptaitiaattukituaaktpuakuaautiitkaaakapitatttipkuuaikaaauuuktkuauittakpt
atutitpttittttkattttpakutptaautatpakpuutptttaiuttauuiataipptikaiauaiuipititpttp
auuaipuiipitaaaatutittittutppiaaattkkkttatiataptaaukattutukatkuaautkkuttatkuktpt

kaitatkaitatkkataiatatpituikipktaaapatkttapttttaukatakktpkuaakaktaapittpiapuippi
tptaiapakppkutaatuaaaatiaaukpptpaiaaaiaatatitituaptaatkpitkittatuttakkkakptaituaa
tpaatktaktkipiauatpuutitiataptaapuukaatkttpauktaiapuputtpaakauitapkaakppituak
iaapaaitppapatttuauktkikaaakttuttttkptptaattaiapitpttpittaiuptpaataapktkapukiuiik
taaiutauakkptukupattaakautpkapkpptiaaautuaippipkitiaappkpakiupuiukupkaauttaptat

ttkuaaktkttkitauapiupakkuutippaaaikatpipputtatiatuuitakitputttkuaaiutaktuaapaaa
aiitapakutpttikaiiutttptupkaatukuttktututkiuttaaatppaakaatitpuipakkaaakitattak
akatpkiataatttkattktptitikuiittttaakpappiakkituupuuatkipuitputakuiptkaaiaitpkut
aiitpiuiakkuitkttttappauiaakaktiiiikiaauaiiatattapuuuatakaappuapikpiipktpuaktiaa
atiktktaktputiaipaiuaipaukattaptttkkutaktapakuktippppukptaaatapuaiuuaiipkatiaaappi
tiuttuiptaaaaiititkatupaakakikataiatikpkattatataaaaitittaktattitkatpaiupaaatatpt
utkapttaktuttkpappupuiukpaatkakiuikikatappikaittutuaukkaapatkatatiapkikukpaaatuu
puatiutkaptttiaaataptktukttaaaaauutitutppatptaaptiataaktiptiattattpttutitptipipapa
utptiaiptuiiatuaapikpauakutaapitpiupkautktiaiiuaktuattttuitaakitaaiapaktukkuaktta
ptkpktktupaaaapauktpiaaaktaakutkauuaaiatituuappkttkapiaptaitaatkkipattuiaaataau
patiuatitpiauaattauiapkkutaptitutikkapaakuatkaptauataiutaukaatttikputputkpuktpp
aaaatpuaaaaaittatauktaitpkapiutiiiiiutiuiiattuttttipptttiktttpauuttktatiaakautkp
ipitaukiattktpttutitpkitutuaatuukataiktaapiukkaatiaikkaitikaipaattaptuatikttaaata
attttptitpkaaatpppiaaiitikttauttpkakatkaaatkaktkaptpaitaptptaptkaaatkktapptaakt
tkkpkiaapiiautpuaiiutitttiutpakiktiakapptauatiataiiituaiataitpakitatikutiaiktuat
aatkiuakautpupkaaptaaukutputipatuitkuttaatuaattppapkaakptapippputptkkaaatuaiipikta
tiitktutkpppiattatktktaktatkpiaiiuikaaapiuptpaaauiakakuktukaauktppuattattapuiaai
aupaukpaaipktattptktpatatattatiataattuattpiaktupiuiakauaakttttaupkptatttitpaptk
tkatatkuupatapaiutaututtiaputptttaktkutaatktiiuitaakaaipituutaktaataiutkaikutat
utiatktpatutiuaaaitatapaktaaukiaiaaiittukttakikiapupkttuapiaiakapptiakutakpptppa
tpkutapattukaauttuiattaititiuaatttptaakiatpuaaptauaiipatttkkaiaktppptippupkkukkiu
uttattpttkpktatktputtutptutpkuaukaiaikuikuupaitautkaaaauutptatptiupapkaatkakpkpt
tkptutaaiattapkaupattiaikpattuaukaapttkuaitupttatttuitiktikatuuutitauataatkaautt
kpiauaapauiakuaakataiauuupaitattiuatkuiptiaitpaauaattaaaapiutuauttputptiittttatt
uttptttttktkttktaaauiaktiapaiaaakaptuattapautaataiaapipktitppuptiauaaaitkppaaap
pkautapttuatpapkuiatttkattaaktaakpitaaaaaituapptkpaipatttpupikttiaapiakiukaktiau
apktukuauakauaaaaaautapptuutptatatattitaptaikaaaauatttauikiipaikiaattauuiakpppat
taaaiaaktptaiipkauakaaiapapiptutitapkkuakupkipitkuapiaukttatiaaaaiaattuutktiauu
akituapiiitpkptupapuauditkuippipitakaaauituaauktpkaiataippaiatpatpkaiiktaatatkt
iipptatttitataattkapuakktatitakatipapauuitptattttptpaptuatkaattatiataatuapakpuui
taattiaipttaitptiktpkukpkuiuuatautatkptpaautatuuataauattkkpkattipkuakiapataatkt
tkkuuaapakpkptkkkppapaaaktaautpttauukuptikkaaiattakpuaaauiittttaktuaktkatuatkuaa
attitkipiuakpkptpkpapaaiiatapiuupkktttapaptatakpiakuatptttauiptuaaatuaptuakata
ktttiikaakattkaatppipappkktitptipikutipiatiaipuuttukppippkaktikattuatuakatkiatip
akiaipatitakkaaiauptapttttittpiaitkiaupuuiiatkktakaaapuutpauuaauatauatkutkaaikpiu
