MA615 Final Project – Job Analysis

Yinfeng Zhou

2020/12/11

In this project, my first goal is to do some text analysis of job description. By using API from `Adzuna` and `The Muse`, I fetched two datasets. The code for fetching `The Muse`, searching for those relating to `Data Science`:

```
musekey<-"f4d1dda85a1f43c8a1b0071c0b332b663be1a26c887a61bafb6174dcea54209f"
muselabel<-"MSSP BU"
museurl<-"https://www.themuse.com/api/public/jobs?category=Data%20Science&page=1&api_key=f4d1dd
a85a1f43c8a1b0071c0b332b663be1a26c887a61bafb6174dcea54209f"

musegenerator<-function(page){
  museurl<-paste("https://www.themuse.com/api/public/jobs?category=Data%20Science&page=",page,
"&api_key=f4d1dda85a1f43c8a1b0071c0b332b663be1a26c887a61bafb6174dcea54209f",sep="")
  return(museurl)
}

musedata<-fromJSON(museurl,flatten=TRUE)
muse<-musedata$results
musetemp<-muse
i=1
while(nrow(musetemp)==20&&nrow(muse)<1000){
  i=i+1
  museurl<-musegenerator(i)
  datatemp<-fromJSON(museurl,flatten=TRUE)
  musetemp<-datatemp$results
  muse%<>%rbind(musetemp)
}
```

The code for fetching the Adzuna job, searching for those having `data` in their titles:

```
appid<-"743becb0"
appkey<-"d9861c085b15df57a08a8d4c49c3ac8a"
npage<-20
apigenerator<-function(appid,appkey,page,npage){
  apiurl<-paste("https://api.adzuna.com/v1/api/jobs/gb/search/",page,"?app_id=",appid,"&app_key
=",appkey,"&results_per_page=",npage,"&title_only=data",sep="")
  return(apiurl)
}


apiurl2<-"https://api.adzuna.com/v1/api/jobs/gb/search/1?app_id=743becb0&app_key=d9861c085b15df
57a08a8d4c49c3ac8a&results_per_page=20&title_only=data"
data2<-fromJSON(apiurl2,flatten=TRUE)
result2<-data2$results
result2%<>%select(-contract_time)
resulttemp2<-result2
i=1
while(nrow(resulttemp2)==20&&nrow(result2)<1000){
  i=i+1
  tempurl<-apigenerator(appid,appkey,i,npage)
  datatemp<-fromJSON(tempurl,flatten=TRUE)
  resulttemp2<-datatemp$results
  result2%<>%inner_join(resulttemp2)
}
```

In the Adzuna dataset, there are salary data and the location. However, because of the setting from the website, developers can only fetch 500 bytes at maximum of the job description. Therefore, the Adzuna datset is not suitable for text analysis. However, we can use the location data and `Leaflet` package to map the jobs.

In the `The Muse` dataset, there are complete descriptions of the job. My interest will be, what is the difference of words used in the job descriptions between different companies ? By using wordcloud and histograms, we can easily have a glance on the style of these companies.
After cleaning the dataset below and use `wordcloud2` package, I successfully develop the shinyapp (https://yinfeng.shinyapps.io/JobAnalysis/). Due to the limit of time, the mapping function will be developed in the future with the `Adzuna` dataset.

```
muse%>%select(c("id","publication_date","name","contents","locations","refs.landing_page","comp
any.id","company.name"))->museset
museset$contents%<>%str_replace_all("\\<(.*?)\\>","")
museset%<>%separate(publication_date,c("publication_date","temp"),sep="T")
museset%<>%select(-temp)
museset$publication_date%<>%ymd()
museset%<>%unnest_longer(locations)
```


data
experience
business
team
will
work
development
people
learning
research
skills
machine
including
working
science
technical
knowledge
management
analytics
engineering
support
technology
information
years
build
solutions
product
new
opportunity
gender
status
analysis
teams
design
strong
ability
complex
global
degree
insights
statistics
communication
national
world
identity
disability
services
across
amp
you
develop
projects
required
requirements
problems
modeling
qualifications
orientation
process
statistical
related
environment
tools
position
help
using
clinical
large
looking
written
models
computer
performance
customer
religion
intelligence
understand
operations
software
systems
employment
products
analytical
preferred
industry
understanding
company
protected
opportunities
together
customers
technologies
employees
programming
sexual
quality
field
identify
equal
python
sql
drive
key
race
origin
veteran
committed
techniques
leadership
applicants
qualified
professional
applications
employer
reporting
language
partners
candidates
algorithms
developing
financial
medical
provide
project
various
manage
around
group
based
highly
within
service
multiple
applicable
health
diverse
processes
deep
community
bring
relevant
facebook
advanced
collaborate
mathematics
solving
program
quantitative
expertise
providing
scientific
responsible
training
security
leading
responsibilities
federal
culture
make
deliver
age
one
job
time
best
can
our
the
this
use
part
also
join
well
need
apply
ensure
building
needs
action
cloud
improve
candidate
innovation
processing
platform
internal
decision
digital
clients
level
high
receive
future
create
andor
color
ways
vision
like
role
insights
impact
mission
lead
state
please
strategic
without
healthcare
accommodations
different
regard
engineers
innovative
model
individual
following
effectively
excellent
expression
delivery
scientists
architecture
infrastructure
strategy
successful
operational
testing
equivalent
affirmative
authorization
connect
aws
risk
sex
areas
big
etc
gsk
plus
youll
career
power
growth
scale
focus
learn
results
benefits
diversity
decisions
creating
minimum
include
stakeholders
senior
metrics
critical
verbal
concepts
methods
closely
applied
communities
organization
application
marketing
communicate
implement
maintain
experiences
business
initiatives
industry
members
functional
capabilities
analyze
crossfunctional
companies