

Developing and deploying an R 'app' on AWS

Roman Popat - Data Scientist
roman.popat@thedatalab.com
twitter: @rmnppt
web: rmnppt.com
Github: rmnppt

Things to cover:

1. Accessing an REST API
2. Publishing an R package on github
3. EC2 instance and R/Rstudio
4. RDS storage (MySQL and R)
5. Flexdashboard (Rmarkdown)
6. Scheduling linux tasks (crontab)

REST API examples



A case study



Login / Register



The banner features a background image of a city skyline. On the left, a large blue circle contains a white square with a green bar chart icon, a silhouette of a person, and the text '£31,667' and 'DARREN'S CV VALUE'. To the right of this circle, the text 'ValueMyCV' is displayed in blue, followed by 'WHAT ARE YOU WORTH?' in green. Below this text is a green button with a white upload icon and the text 'Upload your CV today'. A small white circle with a downward arrow is centered below the main text.

SEARCH JOBS

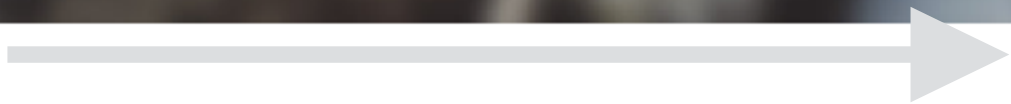
What?

e.g. job, company, title



Where?

e.g. city, county or postcode



What does an API call look like?

api.adzuna.com:80/v1/api/jobs/gb/search/
1?

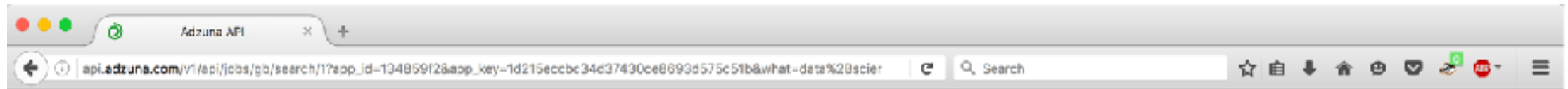
app_id=134859f2&app_key=1d215eccbbc34d374
30ce8693d575c51b&what=data%2Bscience



Data returned from API



JSON



```

__CLASS__
  Adzuna::API::Response::JobSearchResults
count
  1452
mean
  56323.02
results
  0. __CLASS__
    Adzuna::API::Response::Job
  adref
    eyJhbGciOiJIUzI1NiJ9.eyJpIjojNDE4NTI5MDg4Iiwic2l6IjBGU0F2QkRlVGhxTXFoaWFVZ1B0RmciQ.gX0vUPgbRUd_WX9mxpw_HbXHD8Bc_iXx2FVfmoQiyw
  category
    __CLASS__
      Adzuna::API::Response::Category
    label
      Accounting & Finance Jobs
    tag
      accounting-finance-jobs

  company
    __CLASS__
      Adzuna::API::Response::Company
    canonical_name
      Endsleigh Insurance
    display_name
      Endsleigh Insurance

  contract_type
    permanent
  created

```

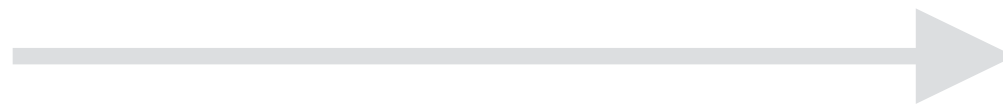
The R function

```
get_country_page <- function(  
  keyword,  
  country,  
  app_id,  
  app_key,  
  page  
) {  
  
  this_url <- paste0(  
    "http://api.adzuna.com:80/v1/api/jobs/",  
    country,  
    "/search/",  
    page, "?",  
    "app_id=", app_id,  
    "&app_key=", app_key,  
    "&results_per_page=50",  
    "&what=", keyword  
  )  
  
  dat <- fromJSON(this_url)  
  
  return(dat)  
}
```



Using the function

```
get_country_page(  
  keyword = "data science",  
  country = "gb",  
  app_id = id,  
  app_key = key,  
  page = 1  
)
```



Sharing Code

R session

```
create("adzunar")
```



Terminal

```
git init  
git add  
git commit  
git push
```



Sharing Code

```
devtools::install_github("rmnppt/adzunar")
```

```
library(adzunar)
```


```
get_country_page()
```




Things to cover:

1. Accessing an REST API
2. Publishing an R package on github
3. EC2 instance and R/Rstudio
4. RDS storage (MySQL and R)
5. Flexdashboard (Rmarkdown)
6. Scheduling linux tasks (crontab)

R/Rstudio on AWS - EC2

 Menu



<https://aws.amazon.com/blogs/big-data/running-r-on-aws/>

Things to cover:

1. Accessing an REST API
2. Publishing an R package on github
3. EC2 instance and R/Rstudio
4. RDS storage (MySQL and R)
5. Flexdashboard (Rmarkdown)
6. Scheduling linux tasks (crontab)

Things to cover:

1. Accessing an REST API
2. Publishing an R package on github
3. EC2 instance and R/Rstudio
4. RDS storage (MySQL and R)
5. Flexdashboard (Rmarkdown)
6. Scheduling linux tasks (crontab)

OAA - Obscure Amazon Acronyms

AWS - Amazon Web Services

EC2 - Elastic Cloud Compute

S3 - Simple Storage Service

RDS - Relational Database Service

finished

