

### Tutorial 3

**NAME:** ANGIE CHONG SIN YI (S2113239)

1. What are the questions that a data scientist can ask on Covid-19 data? Think of some good questions and then categorize your questions as descriptive, exploratory, inferential and predictive.

**Descriptive:** How many new cases and deaths from Covid-19 based on demographics such as gender, age and location?

**Exploratory:** What is the relationship between patient's age and the number of new cases and deaths?

**Inferential:** To identify the potential risk factors in hospital field.

**Predictive:** What are the predictive durations for rise in number of new cases and death cases.

2. Web scraping with R

#### SCRIPT

```
library(xml2)
library(rvest)
library(stringr)
```

```
#read URL
```

```
url<-
'https://www.amazon.in/dp/B07WJV5KPL/ref=sspa_dk_hqp_detail_aax_0?psc=1&spLa=
ZW5jcnlwdGVkUXVhbGlmaWVyPUFQR0hRSVpFQTg4N0wmZW5jcnlwdGVkSWQ9
QTA1MjQ2MzczT1FXQlU5VlAyNzdUJmVuY3J5cHRlZEFkSWQ9QTAWNzU1OTgxT
FdET1hBMIBNOTM4JndpZGldE5hbWU9c3BfaHFwX3NoYXJlZCZhY3Rpb249Y2xp
Y2tSZWRpcmVjdCZkb05vdExvZ0NsaWNrPXRydWU='
webpage<-read_html(url)
```

```
#title
```

```
title_html<-html_nodes(webpage,'h1#title')
title<-html_text(title_html)
head(title)
str_replace_all(title,"[\r\n]", "")
```

```
#price of product
```

```
price_html <- html_nodes(webpage, 'span#priceblock_ourprice')
price <- html_text(price_html)
head(price)
```

```
#product description
```

```
desc_html <- html_nodes(webpage, 'div#feature-bullets')
```

```

desc <- html_text(desc_html)
head(desc)
str_replace_all(desc,"[\r\n]", "")

#rating
rate_html <- html_nodes(webpage,'span#acrPopover')
rate <- html_text(rate_html)
head(rate)
str_replace_all(rate,"[\r\n]", "")

#Size of the product
size_html <- html_nodes(webpage, 'div#variation_style_name')
size_html <- html_nodes(size_html, 'span.selection')
size <- html_text(size_html)
head(size)
str_replace_all(size,"[\r\n]", "")

#Color of the product
color_html <- html_nodes(webpage, 'div#variation_color_name')
color_html <- html_nodes(color_html, 'span.selection')
color <- html_text(color_html)
head(color)
str_replace_all(color , "[\r\n]", "")

#combine all the list
frameproduct_data <- data.frame(Title=title, Price=price,
Description=desc, Rating=rate, Size=size, Color=color)
str(product_data)

```

## **OUTPUT**

```

> library(xml2)
> library(rvest)
> library(stringr)
>
> #read URL
> url<-
'https://www.amazon.in/dp/B07WJV5KPL/ref=sspa_dk_hqp_detail_aax_0?psc=1&spLa=
ZW5jcnlwdGVkUXVhbGlmaWVyPUFQR0hRSVpFQTg4N0wmZW5jcnlwdGVkSWQ9
QTA1MjQ2MzczT1FXQlU5VlAyNzdUJmVuY3J5cHRlZEFkSWQ9QTAwNzU1OTgxT
FdET1hBMIBNOTM4JndpZGdlE5hbWU9c3BfaHFwX3NoYXJlZCZhY3Rpb249Y2xp
Y2tSZWRpcmVjdCZkb05vdExvZ0NsaWNrPXRydWU='
> webpage<-read_html(url)
>
> #title
> title_html<-html_nodes(webpage,'h1#title')
> title<-html_text(title_html)

```

```
> head(title)
[1] "\n\n\n\n\n\n\n\nniQOO Z5 5G (Mystic Space, 8GB RAM, 128GB Storage) | Snapdragon 778G 5G Processor | 5000mAh Battery | 44W FlashCharge\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n"
> str_replace_all(title,"[\r\n]",",")
[1] "iQOO Z5 5G (Mystic Space, 8GB RAM, 128GB Storage) | Snapdragon 778G 5G Processor | 5000mAh Battery | 44W FlashCharge"
>
> #price of product
> price_html <- html_nodes(webpage, 'span#priceblock_ourprice')
> price <- html_text(price_html)
> head(price)
[1] "₹23,990.00"
>
> #product description
> desc_html <- html_nodes(webpage, 'div#feature-bullets')
> desc <- html_text(desc_html)
> head(desc)
[1] "\n\n\n\n\n\n\nAbout this item\n\n\n\n\n\n\nQualcomm Snapdragon 778G 5G 6nm Octa-Core Processor\n\n\n\n\n\n\nSegment's 1st Smartphone with Enhanced LPDDR5 & UFS 3.1\n\n\n\n\n\n44W FlashCharge technology with massive 5000mAh battery with up to 50% in 23 mins*\n\n\n\n\n\n28837mm2 Liquid cooling system that reduces the CPU temperature by about 120C\n\n\n\n\n\n120Hz Refresh Rate | 240Hz Touch Sampling Rate\n\n\n\n\n\n64MP AF Main Camera, f/1.79 Aperture with GW3 sensor, packed with features like 4K video recording at 60FPS, Super Night Mode and much more. Along with 16MP Front Camera.\n\n\n\n\n\n2 Years of Android Updates & 3 Years of Security Updates\n\n\n\n\n\n\u009b\nSee more product details\n\n\n\n"
> str_replace_all(desc,"[\r\n]",",")
[1] "About this itemQualcomm Snapdragon 778G 5G 6nm Octa-Core ProcessorSegment's 1st Smartphone with Enhanced LPDDR5 & UFS 3.144W FlashCharge technology with massive 5000mAh battery with up to 50% in 23 mins*28837mm2 Liquid cooling system that reduces the CPU temperature by about 120C120Hz Refresh Rate | 240Hz Touch Sampling Rate64MP AF Main Camera, f/1.79 Aperture with GW3 sensor, packed with features like 4K video recording at 60FPS, Super Night Mode and much more. Along with 16MP Front Camera.2 Years of Android Updates & 3 Years of Security Updates\u009bSee more product details"
>
> #rating
> rate_html <- html_nodes(webpage,'span#acrPopover')
> rate <- html_text(rate_html)
> head(rate)
[1] "\n\n\n\n4.3 out of 5 stars\n\n\n\n\n\n"
[2] "\n\n\n\n4.3 out of 5 stars\n\n\n\n\n\n"
> str_replace_all(rate,"[\r\n]",",")
[1] "4.3 out of 5 stars" "4.3 out of 5 stars"
>
> #Size of the product
> size_html <- html_nodes(webpage, 'div#variation_style_name')
> size_html <- html_nodes(size_html, 'span.selection')
> size <- html_text(size_html)
```

```
> head(size)
character(0)
> str_replace_all(size,"[r\\n]", "")
character(0)
> 
> #Color of the product
> color_html <- html_nodes(webpage, 'div#variation_color_name')
> color_html <- html_nodes(color_html, 'span.selection')
> color <- html_text(color_html)
> head(color)
character(0)
> str_replace_all(color , "[r\\n]", "")
character(0)
> 
> #combine all the list
> frameproduct_data <- data.frame(Title=title, Price=price,
+ Description=desc, Rating=rate, Size=size, Color=color)
Error in data.frame(Title = title, Price = price, Description = desc, :
arguments imply differing number of rows: 1, 2, 0
> str(product_data)
'data.frame': 2 obs. of 6 variables:
 $ Title      : chr "\n\n\n\n\n\n\nniQOO Z5 5G (Mystic Space, 8GB RAM, 128GB Storage) | Snapdragon 778G 5G Processor | 5000mAh Ba"| __truncated__
"\n\n\n\n\n\n\nniQOO Z5 5G (Mystic Space, 8GB RAM, 128GB Storage) | Snapdragon 778G 5G Processor | 5000mAh Ba"| __truncated__
 $ Price       : chr "<U+20B9>23,990.00" "<U+20B9>23,990.00"
 $ Description: chr "\n\n\n\n\n\nAbout this item\n\n\n\n\n\nQualcomm Snapdragon 778G 5G 6nm Octa-Core Processor\n\n\n\n\n\nSegment's"| __truncated__
"\n\n\n\n\n\nAbout this item\n\n\n\n\n\nQualcomm Snapdragon 778G 5G 6nm Octa-Core Processor\n\n\n\n\n\nSegment's"| __truncated__
 $ Rating      : chr "\n\n\n\n4.3 out of 5 stars\n\n\n\n\n" "\n\n\n\n4.3 out of 5 stars\n\n\n\n\n"
 $ Size        : chr "\n\n8GB RAM +128GB Storage\n\n" "\n\n8GB RAM +128GB Storage\n\n"
 $ Color       : chr "Mystic Space" "Mystic Space"
```