Dulux/Yates running sheet

Automated Jobs for Yates 1

live chat 7 days summary 1

live chat 7 days summary by state 2

Alert email 4

Pull live chat from API 4

Live chat parser 5

Source file, code, and output 8

Live chat 8

Weather impact 10

The weather data 11

The weather station and Bunning stores 11

Live chat and count of chat per tag1 disease/pest 12

the final datasets for the modelling 12

Pest with Weather 13

the impact and output 13

Dulux one-off campaign analysis 14

# Automated Jobs for Yates

Below is a list of jobs running on the serve. Emails are sent to Yates team based on the time and threshold trigger.

## live chat 7 days summary

55 20 \* \* 1 /data/duluxftp/Yates/code/chat/dailyemail.sh

echo "Hi Kylie, please see attached the Yates live chat 7 days summary. Kind Regards, Yates AI robot" | mail -s "Yates live chat 7 days summary" -r "AI Robot <winston.lin@zetaris.com>" -c winston.lin@zetaris.com -a /tmp/Email\_Yates\_livechat\_dailytop10.csv Kylie.Grigg@yates.com.au tammy.huynh@yates.com.au [James.Jones1@duluxgroup.com.au](mailto:James.Jones1@duluxgroup.com.au)

select a.tag1,a.chat\_count\_7days as tag1\_count, tag2,chat\_count\_sub as tag2\_count from

(select tag1,count(distinct chatid) as chat\_count\_7days from livechat\_qlik

where (now()::date-to\_timestamp(chatdate, 'Dy, MM/DD/YY HH:MI:SS am' )::date) <= 7

and tag1 like '%1%'

group by tag1

order by count(distinct chatid) desc

limit 10) a

join

(select tag1,tag2,count(distinct chatid) as chat\_count\_sub from livechat\_qlik

where (now()::date-to\_timestamp(chatdate, 'Dy, MM/DD/YY HH:MI:SS am' )::date) <= 7

and tag1 like '%1%'

group by tag1,tag2

order by count(distinct chatid) desc) b

on a.tag1 = b.tag1

order by tag1\_count desc,tag1,tag2\_count desc)

## live chat 7 days summary by state

56 20 \* \* 1 /data/duluxftp/Yates/code/chat/chatbystate.sh

/usr/local/pgsql/bin/psql -f /data/duluxftp/Yates/code/chat/email.sql Yates -U zetaris -p 20004

echo "Hi Kylie, please see attached the Yates live chat 7 days summary by state. Kind Regards, Yates AI robot" | mail -s "Yates live chat summary by state" -r "AI Robot <winston.lin@zetaris.com>" -c winston.lin@zetaris.com -a /tmp/Email\_Yates\_livechat\_state.csv Kylie.Grigg@yates.com.au tammy.huynh@yates.com.au James.Jones1@duluxgroup.com.au

select a.state,a.tag1,a.topn as tag1\_rank,a.chat\_count\_7days as tag1\_count, tag2,chat\_count\_sub as tag2\_count from

(

select \*,rank() over (partition by state order by state,chat\_count\_7days desc) as topn from

(

select case when postcode::int between 6000 and 6999 then 'WA'

when postcode::int between 4000 and 4999 then 'QLD'

when postcode::int between 5000 and 5999 then 'SA'

when postcode::int between 6000 and 6999 then 'WA'

when postcode::int between 3000 and 3999 then 'VIC'

when (postcode::int between 2000 and 2599) OR (postcode::int between 2620 and 2899) OR ((postcode::int between 2921 and 2999)) then 'NSW'

when (postcode::int between 2600 and 2619) OR (postcode::int between 2900 and 2920) then 'ACT'

else 'others' end as state,tag1,count(distinct chatid) as chat\_count\_7days

from livechat\_qlik

where (now()::date-to\_timestamp(chatdate, 'Dy, MM/DD/YY HH:MI:SS am' )::date) <= 7

and tag1 like '%1%'

and length(postcode)>2

group by state,tag1

) pa

) a

join

(

select

case when postcode::int between 6000 and 6999 then 'WA'

when postcode::int between 4000 and 4999 then 'QLD'

when postcode::int between 5000 and 5999 then 'SA'

when postcode::int between 6000 and 6999 then 'WA'

when postcode::int between 3000 and 3999 then 'VIC'

when (postcode::int between 2000 and 2599) OR (postcode::int between 2620 and 2899) OR ((postcode::int between 2921 and 2999)) then 'NSW'

when (postcode::int between 2600 and 2619) OR (postcode::int between 2900 and 2920) then 'ACT'

else 'others' end as state,tag1,tag2,count(distinct chatid) as chat\_count\_sub from livechat\_qlik

where (now()::date-to\_timestamp(chatdate, 'Dy, MM/DD/YY HH:MI:SS am' )::date) <= 7

and tag1 like '%1%'

and length(postcode)>2

group by postcode,tag1,tag2

order by count(distinct chatid) desc

) b

on a.tag1 = b.tag1

and a.state = b.state

where a.topn <11

and a.state not like 'others'

and length(tag2) >1

order by state,topn

## Alert email

59 20 \* \* \* /data/duluxftp/Yates/code/chat/alertemail.sh

This is an alert from live chat as we are getting 7.5% or 75 count of chats since yesterday to a specific 1p pest/disease

/usr/local/pgsql/bin/psql -f /data/duluxftp/Yates/code/chat/email.sql Yates -U zetaris -p 20004

if (( $(wc -c < /tmp/Email\_Yates\_livechat\_alert.csv)>1)); then

echo "Hi Kylie, This is an alert from live chat as we are getting 7.5% or 75 count of chats since yesterday to a specific 1p pest/disease. Kind Regards, Yates AI robot" | mail -s "Yates live chat disease and pest alert" -r "AI Robot <winston.lin@zetaris.com>" -c winston.lin@zetaris.com -a /tmp/Email\_Yates\_livechat\_alert.csv Kylie.Grigg@yates.com.au tammy.huynh@yates.com.au James.Jones1@duluxgroup.com.au

fi

select \* from

(

select tag1, chat\_count,round(chat\_count/(sum(chat\_count) over ()),2) as pct from

(select tag1,count(distinct chatid) as chat\_count from livechat\_qlik

where (now()::date-to\_timestamp(chatdate, 'Dy, MM/DD/YY HH:MI:SS am' )::date) <= 1

and (tag1 like '%1 Dise%' or tag1 like '%1 Pest%')

group by tag1

order by count(distinct chatid) desc) a

) b

where chat\_count >=75 or pct >= 0.075

## Pull live chat from API

10 3 \* \* \* /data/duluxftp/Yates/code/chat/chatdailypull.sh

#!/bin/sh

s=$(date -I -d " - 1 day")

curl "https://api.livechatinc.com/v2/chats?date\_from=$s&date\_to=$s&page=1" -u tammy.huynh1@yates.com.au:704e6acd8859ae43b170e68420c20465 -H X-API-Version:2 > /data/duluxftp/Yates/Data\_Inputs/LiveChat/daily/livechat${s}1

curl "https://api.livechatinc.com/v2/chats?date\_from=$s&date\_to=$s&page=2" -u tammy.huynh1@yates.com.au:704e6acd8859ae43b170e68420c20465 -H X-API-Version:2 > /data/duluxftp/Yates/Data\_Inputs/LiveChat/daily/livechat${s}2

curl "https://api.livechatinc.com/v2/chats?date\_from=$s&date\_to=$s&page=3" -u tammy.huynh1@yates.com.au:704e6acd8859ae43b170e68420c20465 -H X-API-Version:2 > /data/duluxftp/Yates/Data\_Inputs/LiveChat/daily/livechat${s}3

curl "https://api.livechatinc.com/v2/chats?date\_from=$s&date\_to=$s&page=4" -u tammy.huynh1@yates.com.au:704e6acd8859ae43b170e68420c20465 -H X-API-Version:2 > /data/duluxftp/Yates/Data\_Inputs/LiveChat/daily/livechat${s}4

curl "https://api.livechatinc.com/v2/chats?date\_from=$s&date\_to=$s&page=5" -u tammy.huynh1@yates.com.au:704e6acd8859ae43b170e68420c20465 -H X-API-Version:2 > /data/duluxftp/Yates/Data\_Inputs/LiveChat/daily/livechat${s}5

curl "https://api.livechatinc.com/v2/chats?date\_from=$s&date\_to=$s&page=6" -u tammy.huynh1@yates.com.au:704e6acd8859ae43b170e68420c20465 -H X-API-Version:2 > /data/duluxftp/Yates/Data\_Inputs/LiveChat/daily/livechat${s}6

curl "https://api.livechatinc.com/v2/chats?date\_from=$s&date\_to=$s&page=7" -u tammy.huynh1@yates.com.au:704e6acd8859ae43b170e68420c20465 -H X-API-Version:2 > /data/duluxftp/Yates/Data\_Inputs/LiveChat/daily/livechat${s}7

curl "https://api.livechatinc.com/v2/chats?date\_from=$s&date\_to=$s&page=8" -u tammy.huynh1@yates.com.au:704e6acd8859ae43b170e68420c20465 -H X-API-Version:2 > /data/duluxftp/Yates/Data\_Inputs/LiveChat/daily/livechat${s}8

curl "https://api.livechatinc.com/v2/chats?date\_from=$s&date\_to=$s&page=9" -u tammy.huynh1@yates.com.au:704e6acd8859ae43b170e68420c20465 -H X-API-Version:2 > /data/duluxftp/Yates/Data\_Inputs/LiveChat/daily/livechat${s}9

curl "https://api.livechatinc.com/v2/chats?date\_from=$s&date\_to=$s&page=10" -u tammy.huynh1@yates.com.au:704e6acd8859ae43b170e68420c20465 -H X-API-Version:2 > /data/duluxftp/Yates/Data\_Inputs/LiveChat/daily/livechat${s}10

curl "https://api.livechatinc.com/v2/chats?date\_from=$s&date\_to=$s&page=11" -u tammy.huynh1@yates.com.au:704e6acd8859ae43b170e68420c20465 -H X-API-Version:2 > /data/duluxftp/Yates/Data\_Inputs/LiveChat/daily/livechat${s}11

## Live chat parser

30 3 \* \* \* java -jar /data/duluxftp/Yates/code/chat/YatesLiveChatParser1.0.jar /data/duluxftp/Yates/Data\_Inputs/LiveChat/daily /data/duluxftp/Yates/Data\_Inputs/LiveChat/dailyparsed/chatfrom20160923.txt

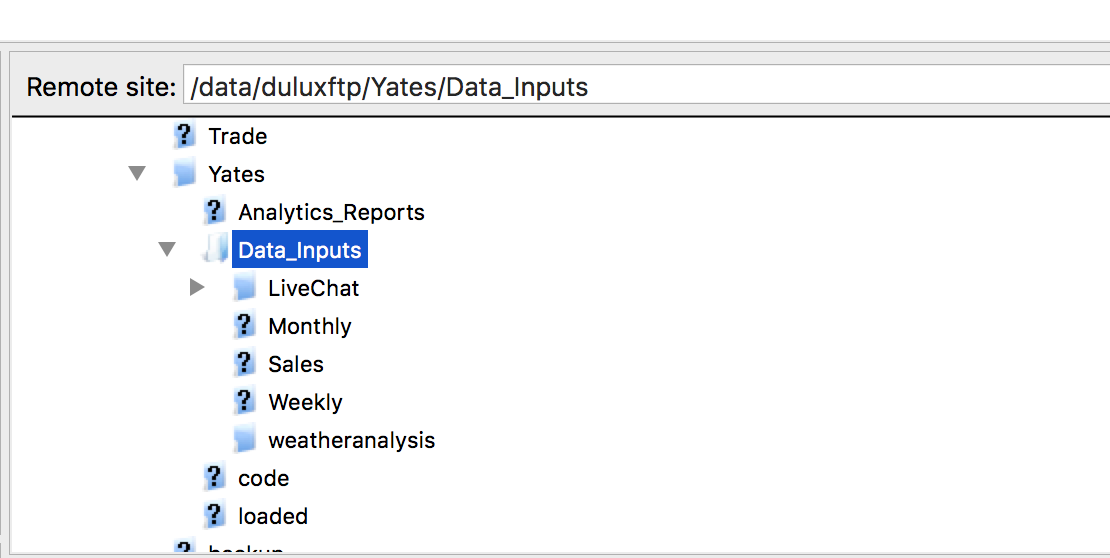
**import** scala.io.{Codec, Source}  
**import** java.io.{File, FileOutputStream, FileWriter, PrintWriter}  
**import** net.liftweb.json.JsonAST.JArray  
**import** net.liftweb.json.\_  
**import** chat.ChatUtils.\_  
**import** sys.process.\_  
  
  
*/\*\*  
 \* Created by winston on 21/09/2016.  
 \* to parse Yates livechat json files into csv format  
 \*/***object** ChatParser **extends** App{  
  
 **val** *usage* =  
 **"""  
 |Usage: ChatParser sourceFolder outputfile  
 |example: java -jar target/scala-2.11/YatesLiveChatParser.jar /Users/winston/data/dulux/livechat/dailyinput /Users/winston/data/dulux/livechat/dailyoutput/livechatoutput.txt  
 |  
 """**.stripMargin  
  
 **if** (args.length <2) *println*(*usage*);  
 **val** *in* = args(0)  
 **val** *outputfile* = args(1)  
  
 *// val in = "/Users/winston/data/dulux/livechat/dailyinput"  
 //val outputfile = "/Users/winston/data/dulux/livechat/dailyoutput/testchat.txt"* **new** File(*outputfile*)  
 **new** FileOutputStream(**new** File(*outputfile*))  
  
 **val** *src* = **new** File(*in*).listFiles().filter(\_.getName.endsWith(**""**))  
 *src*.foreach(*println*)  
  
 **val** *write* = **new** PrintWriter(**new** FileOutputStream(**new** File(*outputfile*),**true**))  
 *//write.write("postcode|chatid|visitor\_id|tag1|tag2|tag3|tag4|guestname|chatdate|chatyear|chatmonth|chatday|chathour|email|chattext \r\n")* **def** file2json(f:File):Array[String] ={  
 Source.*fromFile*(f)(Codec.*ISO8859*).getLines().filter(\_.contains(**"chats"**))  
 .mkString.replaceAll(**"\\{\"chats\":\\[\\{\"type\""**,**"\n\\{\"chats\":\\[\\{\"type\""**)  
 .split(**"\n"**)  
 }  
  
  
  
 **def** json2CSV(js:String) = {  
  
 **for** {  
 *JArray*(chats) <- *parse*(js) \\ **"chats"** *JObject*(chat) <- chats  
 *JField*(**"prechat\_survey"**, *JArray*(prechat\_survey)) <- chat  
 *JField*(**"id"**, *JString*(id)) <- chat  
 *JField*(**"visitor\_id"**, *JString*(visitor\_id)) <- chat  
 *JField*(**"tags"**, *JArray*(tags)) <- chat  
 *JField*(**"messages"**, *JArray*(messages)) <- chat  
  
  
  
 msglist = **for** {  
 *JObject*(prechat) <- prechat\_survey.tail  
 *JObject*(imsg) <- messages  
 *JField*(**"author\_name"**, *JString*(author\_name)) <- imsg  
 *JField*(**"text"**, *JString*(text)) <- imsg  
 *JField*(**"date"**, *JString*(chatdate)) <- imsg  
 *JField*(**"agent\_id"**, *JString*(agent\_id)) <- imsg  
 } *write*.write(prechat.mkString.postcode() + **"|"** + id + **"|"** + visitor\_id + **"|"** + *tagfunc*(tags) + **"|"** + author\_name + **"|"** + chatdate + **"|"** + chatdate.ymdhcols() + **"|"** + agent\_id + **"|"** + text.replace(**"/"**,**""**).replace(**"\\"**,**""**).replace(**"|"**,**""**).replaceAll(**"\n"**,**""**).replaceAll(**"\r"**,**""**) + **"\r\n"**)  
 *// println prechat.("postcode")* } **yield** id  
  
 }  
  
  
 **for** {  
 file <- **new** File(*in*).listFiles.filter(\_.getName.endsWith(**""**)).toIterator **if** file.isFile  
 line <- *file2json*(file).filter(\_.size >0)  
 } { *json2CSV*(line) }  
  
  
 *Seq*(**"sed"**,**"-i"**,**"$ d"**,*outputfile*)!  
  
]

*/\*\*  
 \* Created by winston on 21/08/2016.  
 \*/***package** chat  
**import** java.time.LocalDateTime  
  
**object** ChatUtils {  
  
  
  
 **implicit class** tvalue(s:String){  
  
 **def** t():String= {  
 s.replace(**"JString("**,**""**).replace(**")"**,**""**)  
 }  
  
 **def** ymdhcols():String = {  
 **val** format = **new** java.text.SimpleDateFormat(**"EEE, MM/dd/yy HH:mm:ss"**)  
 *// val stime = "Mon, 08/01/16 08:09:50"* **val** newdate = format.parse(s)  
 **val** y = newdate.getYear-100  
 **val** m = newdate.getMonth+1  
 **val** d = newdate.getDate  
 **val** h = newdate.getHours  
 y+**"|"**+m+**"|"**+d+**"|"**+h  
 }  
  
  
 **def** postcode():String= {  
 s.substring(s.indexOf(**"Postcode"**)+32,s.indexOf(**"Postcode"**)+36).replaceAll(**"[^0-9]"**,**""**)  
  
 }  
  
 *// 08/01/16 08:09:50 am* }  
  
  
 **def** tagfunc[A](list:List[A]): String = list.size **match** {  
 **case** 0 => **"|||"  
 case** 1 => list.head.toString().t + **"|||"  
 case** 2 => list.head.toString().t + **"|"**+ list(1).toString().t + **"||"  
 case** 3 => list.head.toString().t + **"|"**+ list(1).toString().t + **"|"** + list(2).toString().t + **"|"  
 case** 4 => list.head.toString().t + **"|"**+ list(1).toString().t + **"|"** + list(2).toString().t + **"|"** + list(3).toString().t  
 **case** s **if** s>4 => list.head.toString().t + **"|"**+ list(1).toString().t + **"|"** + list(2).toString().t + **"|"** + list(3).toString().t  
 **case** \_ => **"|||"** }  
  
  
}

## Source file, code, and output

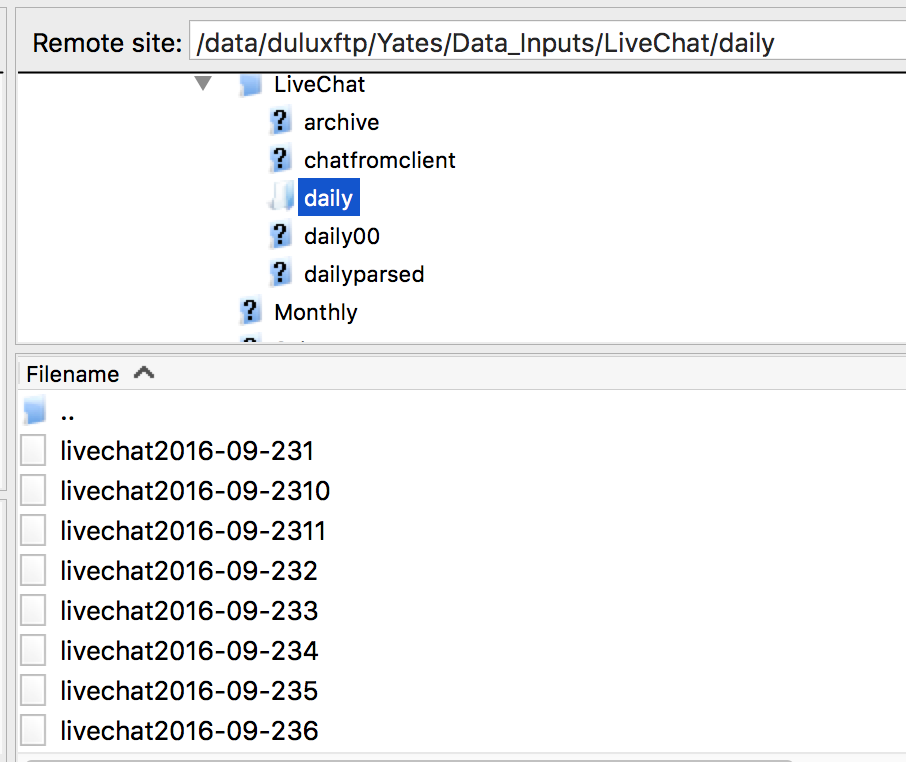
See below the ftp for all source files, code and output file. The temp working directory is /tmp

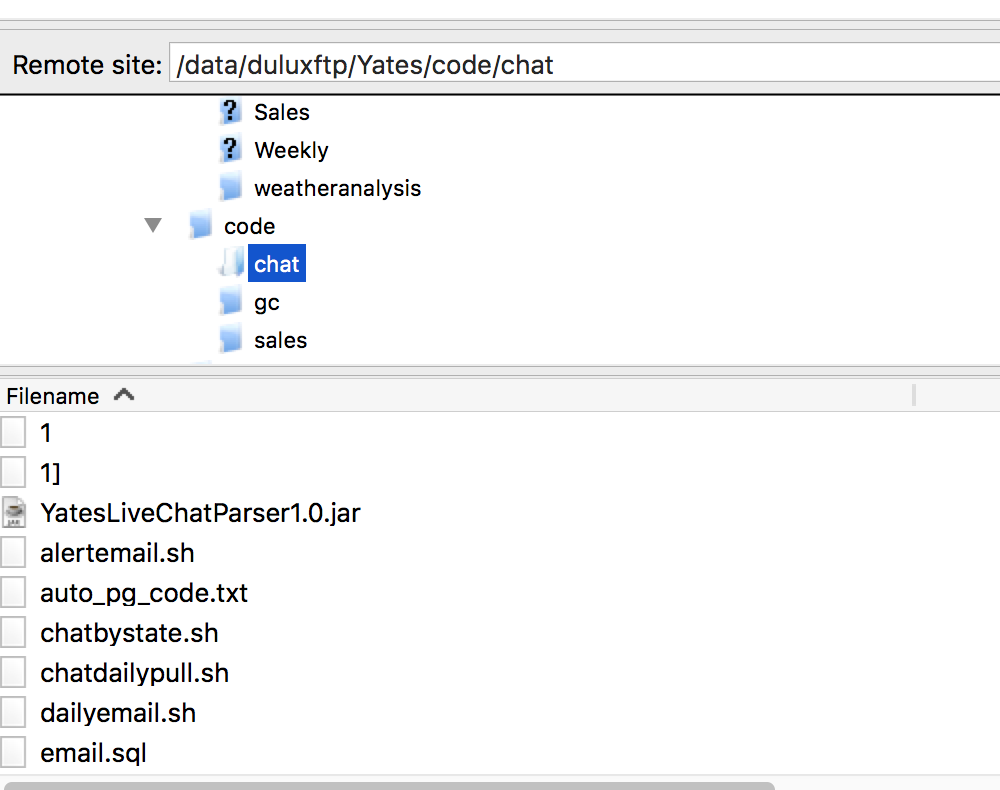
Refer to the infrastructures doc he ftp access details.



## Live chat

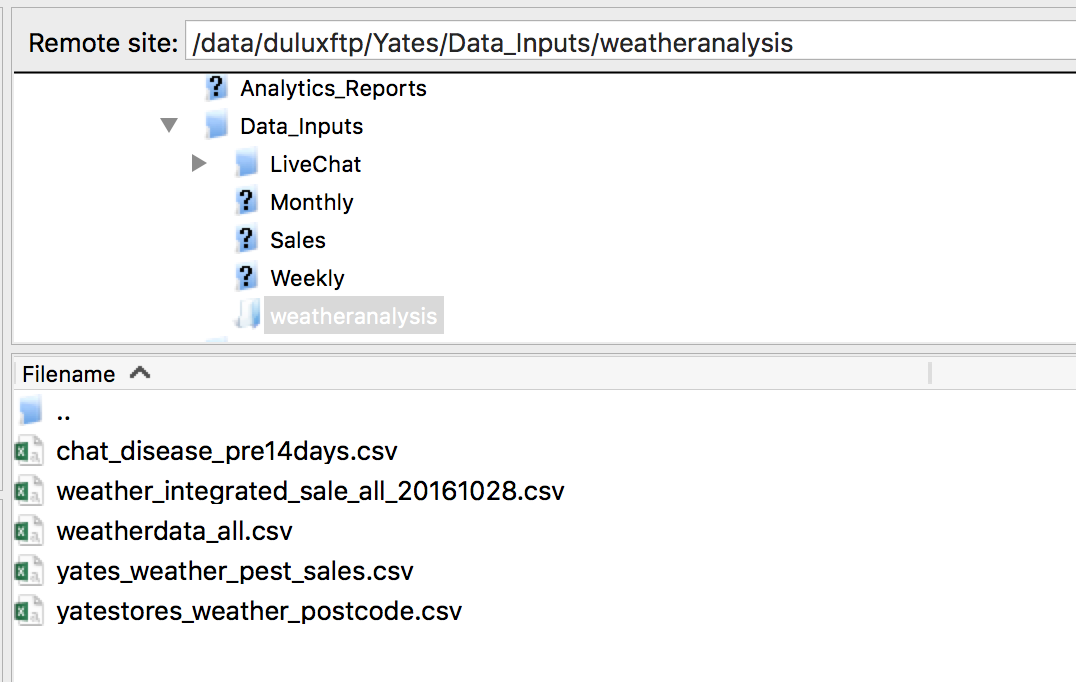
The live chat is pulled from the live chat website and parsed with the code in the code folder.





# Weather impact

The following datasets and processes were used for the one-off weather impact analysis. We manually ran a regression model to find the relationship between the sales with weather and pest disease.



### The weather data

weatherdata\_all.csv in ftp

stationname,date,evapo,rain,pan,maxtemperature,mintemperature,maxhumidity,minhumidity,windspeed,solarradiation

ALBURY AIRPORT,01/01/2009,7.8,0.0, ,26.5,11.3,83,23,5.72,34.69

ALBURY AIRPORT,02/01/2009,7.7,0.0, ,23.9,9.6,59,19,5.10,35.85

ALBURY AIRPORT,03/01/2009,7.1,0.0, ,28.8,10.5,62,18,2.20,35.37

ALBURY AIRPORT,04/01/2009,7.9,0.0, ,34.6,12.3,71,9,2.42,33.35

ALBURY AIRPORT,05/01/2009,8.5,0.0, ,35.8,12.9,55,7,2.44,35.14

ALBURY AIRPORT,06/01/2009,9.9,0.0, ,37.9,13.7,55,7,3.39,34.95

ALBURY AIRPORT,07/01/2009,10.5,0.0, ,38.9,16.1,57,10,4.31,32.70

ALBURY AIRPORT,08/01/2009,9.1,0.0, ,28.3,14.0,61,13,5.42,35.40

### The weather station and Bunning stores

yatestores\_weather\_postcode.csv

stationid,state,id,name,longid,lat,lon,store,latitude,longtitude,distance,mindist,storename,postcode,store\_s

61078,NSW,61,WILLIAMTOWN RAAF,19420101..,-32.7932,151.8359,Bunnings Port Stephens 7309,-32.74882,152.07008,13.94632092,13.94632092, Port Stephens 7309,, Port Stephens

41525,QLD,41,WARWICK,19940131..,-28.2061,152.1003,Bunnings Warwick 8125,-28.2125,152.03344,4.094848561,4.094848561, Warwick 8125,, Warwick

90186,VIC,90,WARRNAMBOOL AIRPORT NDB,19980902..,-38.2867,142.4522,Bunnings Warrnambool 6225,-38.38362,142.52392,7.74287808,7.74287808, Warrnambool 6225,, Warrnambool

82138,VIC,82,WANGARATTA AERO,19870513..,-36.4206,146.305

### Live chat and count of chat per tag1 disease/pest

chat\_disease\_pre14days.csv

tag1,cyear,cweek,pre14chat

"1 Disease Anthracnose",2015.0,36.0,2

"1 Disease Anthracnose",2015.0,38.0,3

"1 Disease Anthracnose",2015.0,39.0,2

"1 Disease Anthracnose",2015.0,42.0,2

"1 Disease Anthracnose",2015.0,43.0,2

"1 Disease Anthracnose",2015.0,44.0,3

"1 Disease Anthracnose",2015.0,50.0,3

"1 Disease Anthracnose",2016.0,14.0,1

"1 Disease Anthracnose",2016.0,19.0,2

### the final datasets for the modelling

weather\_integrated\_sale\_all\_20161028.csv

store,product,saleyear,season,saleweek,qty,totalrain,maxtemp,maxhumidity,mintemp,minhumidity,maxwindspeed,rownumber,pre14rain

Bunnings Airlie Beach 8467,ZZZYts Rose Shield Insect&Disease500ml##,2013,1,8,1,68.80000000000001,32.6,97,,55,3.67,3,263.8

Bunnings Airlie Beach 8467,ZZZYts Rose Shield Insect&Disease500ml##,2013,2,15,2,124.8,28.0,99,18.4,64,6.20,4,324.4

Bunnings Airlie Beach 8467,ZZZYts Rose Shield Insect&Disease500ml##,2013,2,16,2,0.2,30.3,99,17.8,50,3.58,5,193.60000000000002

Bunnings Airlie Beach 8467,ZZZYts Rose Shield Insect&Disease500ml##,2013,2,19,1,31.800000000000004,27.5,98,14.3,51,6.33,6,125.0

### Pest with Weather

yates\_weather\_pest\_sales.csv

"saleyear","saleweek","store","product","season","qty","totalrain","maxtemp","maxhumidity","mintemp","minhumidity","maxwindspeed","rownumber","pre14rain",".PlantAdvice",".PlantHealth",".PlantProtection","DiseaseAnthracnose","DiseaseBlackspot","DiseaseBlossomEndRot","DiseaseCollarRot","DiseaseLeafCurl","DiseaseLeafSpot","DiseaseMyrtleRust","DiseasePowderyMildew","DiseaseRust","DiseaseSootyMould","DiseaseSunburn","HibiscusFlowerBeetle","PestAfricanBlackBeetle","PestAnt","PestAphid","PestArmyWorm","PestAzaleaLaceBug","PestBeetle","PestCaterpillar","PestCurlGrub","PestElmLeafBeetle","PestScale","PestSnailsSlugs","PestWoollyAphid","PhysicalDamageSunburn","PPlantProtection"

"1",2015,33,"Bunnings Rosebud 6075","Yts Mavrik Gun 750ml",3,2,6.8,15.7,95,2.1,48,4.12,91,39.8,0,0,0,0,0,0,0,1,0,0,1,1,0,0,0,0,0,0,0,0,0,1,0,0,3,0,0,0,4

"2",2015,33,"Bunnings Moorabbin 6363","Yts Pestoil 500ml",3,4,4.8,16.3,98,3.8,47,8.06,117,22.4,0,0,0,0,0,0,0,1,0,0,1,1,0,0,0,0,0,0,0,0,0,1,0,0,3,0,0,0,4

"3",2015,33,"Bunnings Box Hill 6035","Yts Pestoil RTU 750ml 6pk",3,2,3.6,16.9,99,1.6,47,4.77,130,28,0,0,0,0,0,0,0,1,0,0,1,1,0,0,0,0,0,0,0,0,0,1,0,0,3,0,0,0,4

"4",2015,33,"Bunnings Modbury 5210","Yts Pyrethrum RTU 3L",3,1,12,17.9,98,10.8,38,6.78,65,9.6,0,0,0,0,0,0,0,1,0,0,1,1,0,0,0,0,0,0,0,0,0,1,0,0,3,0,0,0,4

"5",2015,33,"Bunnings Northland 6308","Yts Liquid Copper Fungicide 200ml",3,2,3.6,16.9,99,1.6,47,4.77,120,19,0,0,0,0,0,0,0,1,0,0,1,1,0,0,0,0,0,0,0,0,0,1,0,0,3,0,0,0,4

### the impact and output

We’ve presented the sample output of regression. Here is the main code ran in R.

lm <-lm(qty ~ season + totalrain + maxtemp + maxhumidity + mintemp +

minhumidity + mintemp + minhumidity + maxwindspeed + pre14rain+ DiseaseAnthracnose+DiseaseBlackspot+DiseaseBlossomEndRot+DiseaseCollarRot+DiseaseLeafCurl+DiseaseLeafSpot+DiseaseMyrtleRust+DiseasePowderyMildew+DiseaseRust+DiseaseSootyMould+DiseaseSunburn+HibiscusFlowerBeetle+PPlantProtection+PestAfricanBlackBeetle+PestAnt+PestAphid+PestArmyWorm+PestAzaleaLaceBug+PestBeetle+PestCaterpillar+PestCurlGrub+PestElmLeafBeetle+PestScale+PestSnailsSlugs+PestWoollyAphid+PhysicalDamageSunburn+.PlantAdvice+.PlantHealth+.PlantProtection,

data=pestsales)

Here is an example to generate the visualization in R.

# Conditional Inference Tree

fit2 <- ctree(qty ~ season + totalrain + maxtemp + maxhumidity + mintemp +

minhumidity + mintemp + minhumidity + maxwindspeed + pre14rain ,

data=pyrethrum750)

plot(fit2, main="Conditional Inference Tree for pyrethrum750",

gp = gpar(fontsize = 6), # font size changed to 6

inner\_panel=node\_inner,

ip\_args=list(

abbreviate = FALSE,

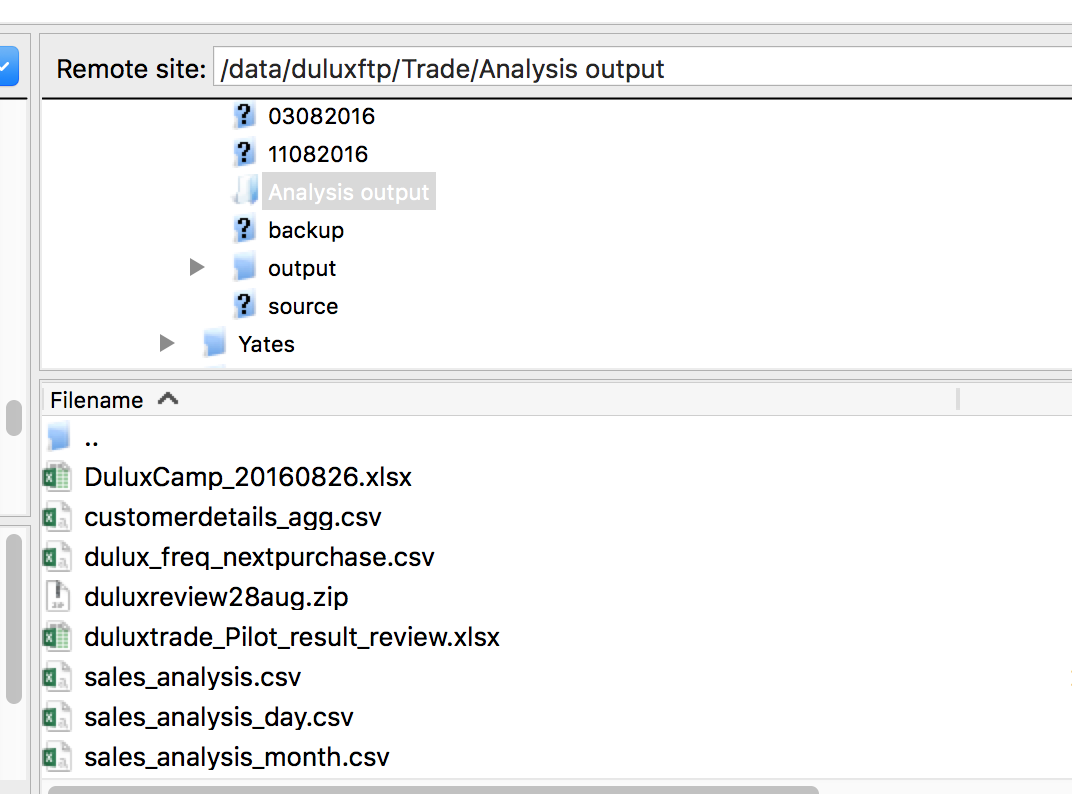
id = FALSE))

# Dulux one-off campaign analysis

The requirements from Dulux trade was to produce some aggregated summary for customers who are buying brush and paint. Also to understand if there are buying below/above median of each quantiles.

The output has been sent to client and reviewed in several workshops. There is also a copy in dropbox.

Due to the limited time we had, most of the summary and calc have been done in excel. However, there are some sql queries to get the basic datasets before put into Excel for calc. see the code below.



drop table if exists cp\_paintbrush;

create table cp\_paintbrush as

select customer\_id,billdocnum,material,material\_desc,prodhier3\_desc,category, prodhier2,region,fy\_year,calendardate,qty,volumn as volume,netvalue,region\_desc,custhier3,custhier2,plant\_desc, insp\_store

from core

where (category = 'Paint' or prodhier2 = 'BRUSH')

;

select distinct plant\_desc, insp\_store from cp\_paintbrush;

select count(distinct customer\_id),insp\_store from cp\_paintbrush where fy\_year =2015 group by insp\_store ;

select fy\_year ,count (distinct customer\_id),sum(netvalue) from cp\_paintbrush where netvalue >0 and custhier3 like 'INSP%' group by fy\_year;

create index idx\_cust on cp\_paintbrush(customer\_id);

create index idx\_fy\_year on cp\_paintbrush(fy\_year);

--truncate table w\_cust\_insp\_2016;

create table w\_cust\_insp\_2015 as

select distinct customer\_id from cp\_paintbrush where netvalue >0 and insp\_store like 'Yes%' and fy\_year=2015;

create table w\_cust\_duluxstore\_2015 as

select distinct customer\_id from cp\_paintbrush where netvalue >0 and insp\_store not like 'Yes%' and fy\_year=2015;

create table w\_cust\_crossstore\_2015 as

select distinct customer\_id from w\_cust\_duluxstore\_2015 a where a. customer\_id not in (select customer\_id from w\_cust\_insp\_2015 );

select count(\*) from w\_cust\_insp\_2015;

select count(\*) from w\_cust\_duluxstore\_2015;

select count(\*) from w\_cust\_crossstore\_2015;

select distinct material from cp\_paintbrush where prodhier2 = 'BRUSH';

select \* from cp\_paintbrush where material NOT like 'N10%' AND PRODHIER2 = 'BRUSH';

copy cp\_paintbrush to '/tmp/cp\_paintbrush' csv header delimiter ',';

SELECT customer\_id, id, amount, ea\_year, circle\_id

, sum(amount) OVER (PARTITION BY circle\_id ORDER BY month) AS cum\_amt

FROM tbl

ORDER BY circle\_id, month;

/\*

where

--region = 'TFF'

--and

netvalue >0

and fy\_year =2015;

\*\*/

select distinct terrtry\_desc from core;

select distinct category, prodhier2 from core;

-- paint brush

-- average method

-- customers buying brush

drop table if exists d\_paintwithbrush;

create table d\_paintwithbrush as

SELECT

fy\_year, inspiration, paintvolume, brushqty, duluxbrushqty, nonduluxbrushqty,

quartile, customer\_id, txncount, lasttxndate, firsttxndate, "awayDays",

avgdays, triggerdate, nextpurchasedate, frequency, rewardmember,

avg(paintvolume) over (partition by fy\_year, inspiration,quartile order by fy\_year, inspiration,quartile) as avgpaint,

avg(brushqty) over (partition by fy\_year, inspiration,quartile order by fy\_year, inspiration,quartile)as avgbrush,

avg(paintvolume) over (partition by fy\_year, inspiration,quartile order by fy\_year, inspiration,quartile)/(avg(brushqty) over (partition by fy\_year, inspiration,quartile order by fy\_year, inspiration,quartile)) as conversionratio

, case

when (paintvolume/(avg(paintvolume) over (partition by fy\_year, inspiration,quartile order by fy\_year, inspiration,quartile)/(avg(brushqty) over (partition by fy\_year, inspiration,quartile order by fy\_year, inspiration,quartile)))) > brushqty then 'below'

else 'above'

end as target

FROM c\_customerdetailsperyear

where brushqty >0;

select count(\*) from d\_paintwithbrush;

//create table d\_quart as

select fy\_year, inspiration,quartile,count(distinct customer\_id), avgpaint,avgbrush,conversionratio from d\_paintwithbrush

group by fy\_year, inspiration,quartile, avgpaint,avgbrush,conversionratio

order by fy\_year, inspiration,quartile, avgpaint,avgbrush,conversionratio;

select \* from d\_quart;

// drop table if exists d\_nobrush;

//create table d\_nobrush as

select a.fy\_year, a.inspiration, paintvolume, brushqty, duluxbrushqty, nonduluxbrushqty,

a.quartile, customer\_id, txncount, lasttxndate, firsttxndate, "awayDays",

avgdays, triggerdate, nextpurchasedate, frequency, rewardmember

,b.conversionratio, paintvolume/b.conversionratio as brushopportunity

from c\_customerdetailsperyear a

left join d\_quart b

on a.fy\_year = b.fy\_year and a.inspiration = b.inspiration and a.quartile = b.quartile

where brushqty = 0;

create table d\_nobrush\_summary as

select fy\_year, inspiration, quartile,count(distinct customer\_id), sum(paintvolume) as paintvolume, sum(brushopportunity) as brushopportunity from d\_nobrush

group by fy\_year, inspiration, quartile

order by fy\_year, inspiration, quartile

;

select fy\_year, inspiration, quartile,frequency,count(distinct customer\_id), sum(paintvolume) as paintvolume, sum(brushopportunity) as brushopportunity from d\_nobrush

where fy\_year = 2016

group by fy\_year, inspiration, quartile,frequency

order by fy\_year, inspiration, quartile, frequency;

select \* from d\_nobrush\_summary;

-- below avg

create table d\_belowavg as

select \*, (paintvolume/conversionratio-brushqty) as brushopportunity from d\_paintwithbrush

where target like 'below'

;

-- below avg summary

create table d\_belowavg\_summary as

select fy\_year, inspiration, quartile,count(distinct customer\_id), sum(paintvolume) as paintvolume,sum(brushqty) as brushqty, sum(brushopportunity) as brushopportunity

from d\_belowavg

group by fy\_year, inspiration, quartile

order by fy\_year, inspiration, quartile;

select fy\_year, inspiration, quartile,frequency,count(distinct customer\_id), sum(paintvolume) as paintvolume,sum(brushqty) as brushqty, sum(brushopportunity) as brushopportunity

from d\_belowavg

where fy\_year =2016

group by fy\_year, inspiration, quartile, frequency

order by fy\_year, inspiration, quartile,frequency;

-- dulux /non dulux

drop table if exists d\_nondulux;

create table d\_nondulux as

select fy\_year, inspiration, quartile,count(distinct customer\_id), sum(paintvolume) as paintvolume, sum(duluxbrushqty) as duluxbrushqty,

sum(nonduluxbrushqty) as nonduluxbrushqty from d\_paintwithbrush

where nonduluxbrushqty >0

group by fy\_year, inspiration, quartile

order by fy\_year, inspiration, quartile

;

select \* from d\_nondulux;