**Yates/Dulux sales forecast model using weather, sales data and live chat: Phase 1**

(Zetaris) Hagen Schulte In den Bäumen / Stephen Anthony / Jessie Wang / Craig Dunn 

* Craig - Project Manager
* Stephen – Lead Data Scientist
* Hagen – Data Scientist
* Jessie – Data Analyst

(Dulux/Yates) Kylie Grigg / James Jones  

**Objectives:**

**-** Can we correlate key weather data points with product sales in order to predict potential sales opportunities moving forward?

**-** Can we correlate key weather data points with pest/disease outbreaks (data source Live Chat conversations) to predict potential sales opportunities moving forward?

**Data sources:**1.   Bureau of Metrology (BOM) weather information for the past 8 years across 3 states (WA / VIC / QLD) from about 160 weather stations  
2.   Bunnings sales & store information – 5 years across 3 location states (WA / VIC / QLD)   
3.   Yates target products – Stars and Robins (approximately 20 sku’s)   
4.   Yates Pest & Disease tagging - Through Live Chat data that is imported daily to Qlik platform

**Data correlation:**  
1.   Map BOM weather details to store locations   
2.   Map Yates target products to stores sales

**Data analysis:**3.  Use weekly sales data to analyse the weather impact on sales. Is there a repeat of weather trends where we see these spikes re-occur?   
4.  Use weekly sales data to analyse the weather impact on pest and disease outbreaks and uplift in sales. Is there a repeat of weather trends where we see these spikes re-occur?

Phasing – James sees a Phase 1 delivery of model results and insights by the end of August for presentation to internal teams.  Phase II work including the automation of reporting on a monthly basis to follow.

* Weather data:

The weather data used for the sales model can be downloaded from a ftp server of the BOM:

http://www.bom.gov.au/catalogue/anon-ftp.shtml

* Sales, Live Chat data is on the Dulux server:

IP:  43.243.201.28   port:22

Username: dulux

Password:z3t@r!s2015

psql Yates -U zetaris -p 20004

Sales data used under */data/duluxftp/Yates/Data\_Inputs/sales\_hagen.csv*  (these are all the sales data in the database, from 2012 Nov,  only 2015-2017 have been used for dashboard purposes by Jessie)

Live chat data under */data/duluxftp/Yates/Data\_Inputs/LiveChat/dailyparsed* (**all the files** in the folder.  The files are **all delimiter by pipe**)

Export csv:

Copy

(select start\_date,material,customer,sales,qty,cost,store,location,state,postcode,longitude,latitude,territory,area,product,bunnings\_sub\_class,cat2,cat3,cat4,brand,ph4

from sales\_bk

) TO '/tmp/sales\_hagen.csv' (FORMAT CSV, HEADER)  ;

Transfer file to other folder ( linux) : cp /tmp/sales\_hagen.csv  /data/duluxftp/Yates/Data\_Inputs/

* I have reduced the sales data to 3 states and 41 key materials (products) over a period of about 3.5 years. These are all items from the file attached which I got from Jessie. This leads to 176 sales locations  (e.g. Albany) and 343 stores (e.g.Bunnings Albany 2272) in Victoria, Queensland and WA.
* Customer and store are different:

|  |  |
| --- | --- |
| **Row Labels** | **Count of Customer** |
| Aust Capital Terr | 4 |
| New South Wales | 95 |
| Northern Territory | 3 |
| Queensland | 62 |
| South Australia | 20 |
| Tasmania | 8 |
| Victoria | 79 |
| Western Australia | 43 |
| **Grand Total** | **314** |

The address for dulux qliksense app is as follows:

http://43.243.201.153/sense/app/30a2351b-ea1d-4a30-a36d-804daef05ce8?qlikTicket=2imhOwnfxPpe.qud

you can use my user name and password:

user name: qlick1\jessiew

password: ch4ng3m3$jw​

**dulux ftp yates data inputs weatheranalysis: sales\_hagen\_material\_states**

**and weather\_weekly**

Sales\_hagen\_material\_states and weather weekly are newer intermediate version of sales\_version29August and weatherdata\_all.

If nearest weather station is more than 20 kms away from store the store is excluded.

sales\_hagen.csv (8 GB size holds all data from 2012 to 2017 including sales = 0)

from there: Sales\_hagen\_material\_states -> sales13 alas sales\_station.csv

sales13 has locality which is weather station name

lat, long of weather station which is nearest to the store.

Sales 13 holds around 160 stores and 100 weather stations.

Tables hold all weather data for Australian stations (around 700).

sales\_station\_weather (around 1 million rows, all stores and weather and sales data on a weekly basis from 2012 to 2017).

sales\_station\_weather3 (around 0.3 million rows, all stores and weather and sales data on a monthly basis from 2012 to 2017 with historical features).

Sales\_top\_states\_bunnings19 alas sales\_version29August(2017) holds sales data from around 100 stores and 60 stations. This is the version which has been used to create sales per area for Brisbane , Melbourne and Perth results.

Sales\_area 5,8,10 for Brisbane, Melbourne, Perth.

The R-code is on github on the Nielsen project (there is a dulux/yates repository under zetaris according to Stephen A. but I couldn’t access it).

https://github.com/zetaris/proj-nielsen-rtp/R:

* dulux.R
* makes sales data ready.R
* yales model script.R
* make weather data ready.R
* sample distant matrix.R
* live chat.R