

Product Affinity Associative Rules Mining (ARM)

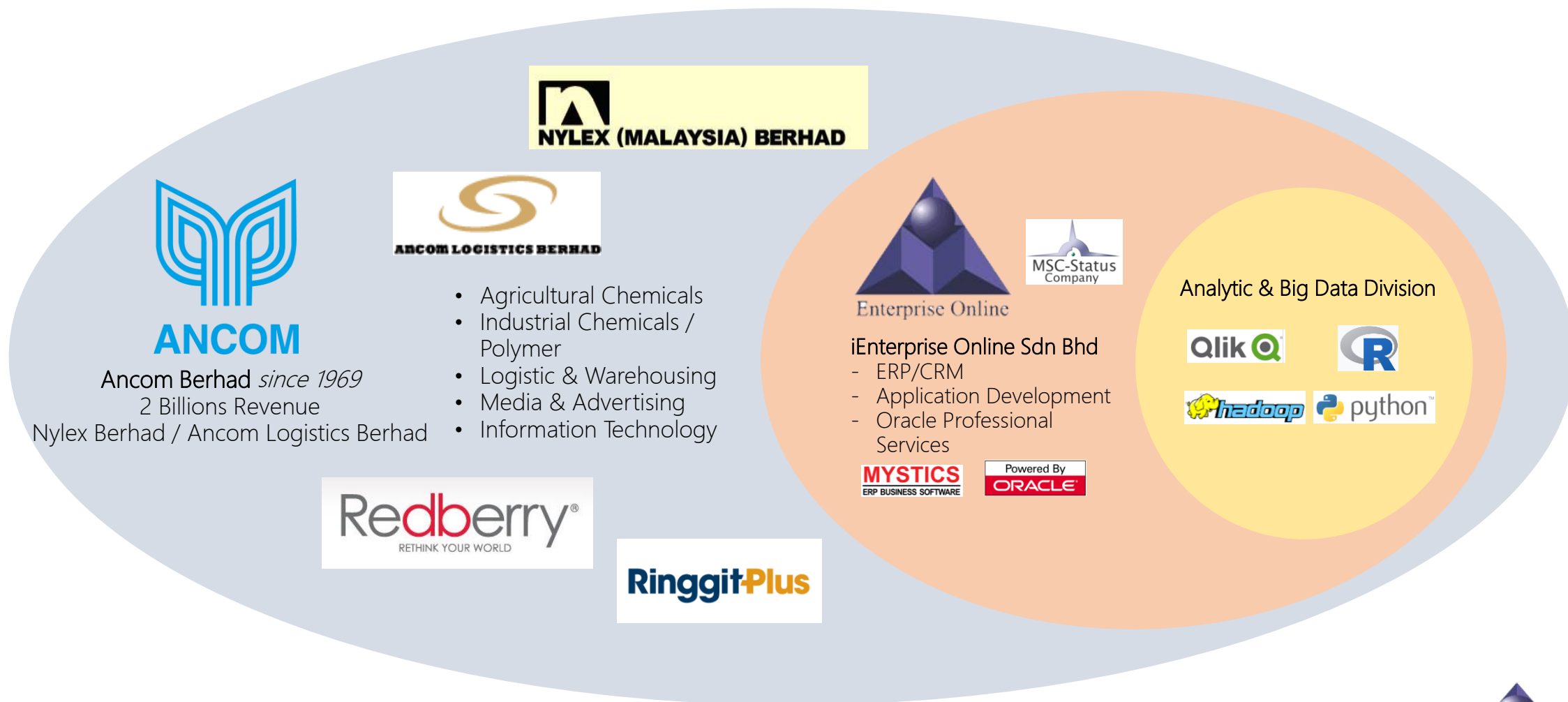
Presented by YK. Tan



Agenda

- Company Introduction
- Product Affinity
- Associative Rules
- Uses Case

Introduction: iEnterprise Online is a subsidiary of ANCOM Group



Introduction: Business Coverages



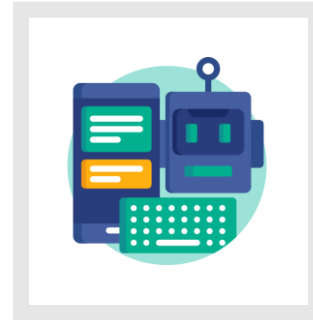
Visualization &
Analytics
Application



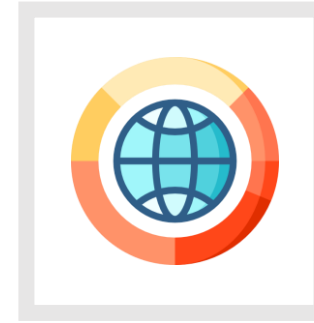
Embedded
Analytics
Application



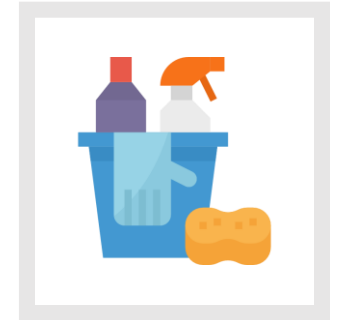
Enterprise
Reporting &
Delivery Application



Advance
& Predictive
Modelling



Big Data
Consulting &
Implementation



Data
Quality
Management

Product Affinity

Definition:

- Also known as market basket analysis
- Analysis that identifies the relationship between customers and attributes associated with them























Associative Rules Mining – Apriori Rules

Associative Rules Mining - Apriori Rules

- Rule based machine learning to find frequent patterns, correlations, associations between variables in large database.

Associative measures:

- Support: Measure of item's occurrence within the whole dataset
- Confidence: Likelihood of occurrence of item y when item x being purchased.
- Lift: Indicator of the strength of association between items.

Transaction 1	   
Transaction 2	  
Transaction 3	 
Transaction 4	 
Transaction 5	   
Transaction 6	  
Transaction 7	 
Transaction 8	 

$$\text{Support} \{\text{apple}\} = \frac{4}{8}$$

$$\text{Confidence} \{\text{apple} \rightarrow \text{beer}\} = \frac{\text{Support} \{\text{apple}, \text{beer}\}}{\text{Support} \{\text{apple}\}}$$

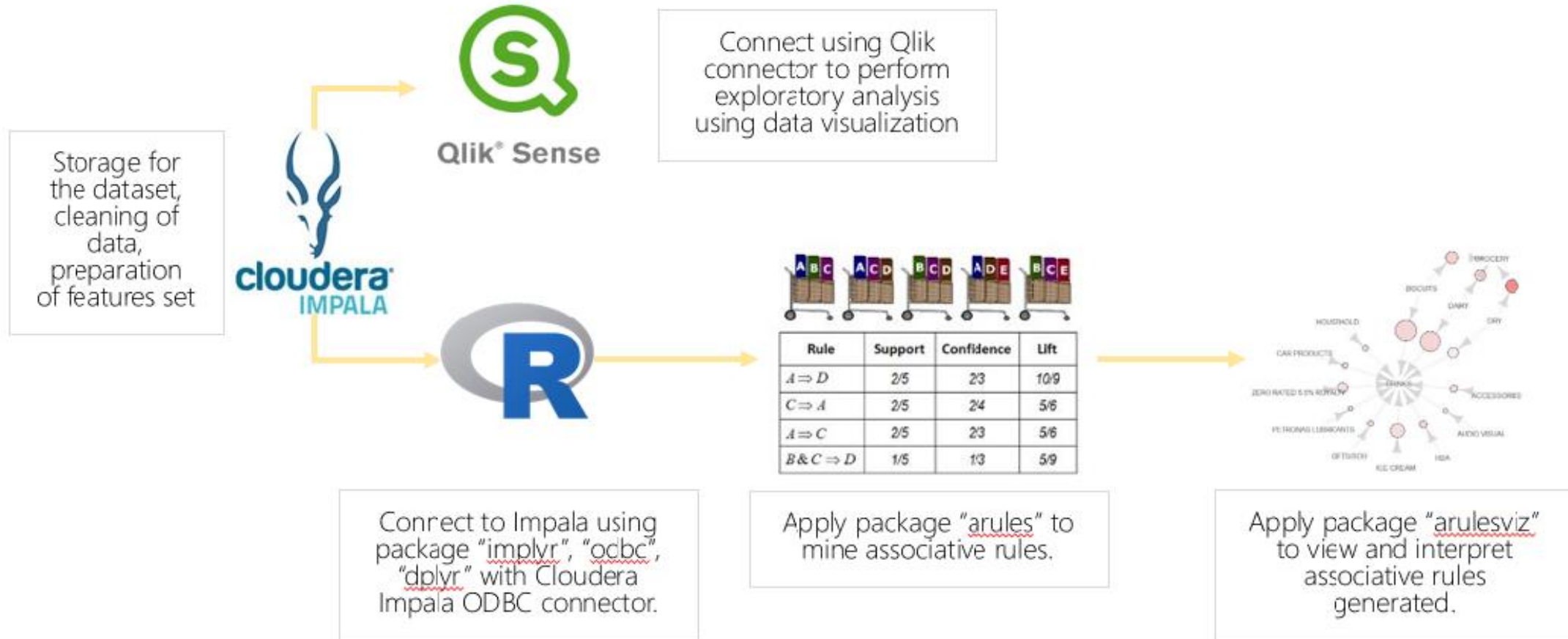
$$\text{Lift} \{\text{apple} \rightarrow \text{beer}\} = \frac{\text{Support} \{\text{apple}, \text{beer}\}}{\text{Support} \{\text{apple}\} \times \text{Support} \{\text{beer}\}}$$

Use Case

- Objective:
- Propose a marketing campaign
- Data-set Background:
- -24-hour chain retail shop
- -Consists of columns as below

	transaction_number	trans_datetime	prod_cat
1	1467670	2018-01-29 18:38:48	GROCERY-BREAD/BUNS/CAKES
2	12441406	2018-01-19 17:31:47	CIGARETTE-BRITISH AMERICAN TOBACCO
3	12524330	2017-11-28 15:15:11	SNACKS-NUTS
4	4378629	2018-02-02 06:40:09	DRINKS-CHOCOLATE
5	13005078	2018-01-18 07:43:10	FOOD SERVICES-PREPARED MEALS
6	11799477	2018-01-19 13:09:23	SNACKS-POTATO

Workflow



Data Preparation

Dataset consists of:

- Duplicates of records
- Missing values

Standardize inconsistency of date format

Summary of data cleaning as below:

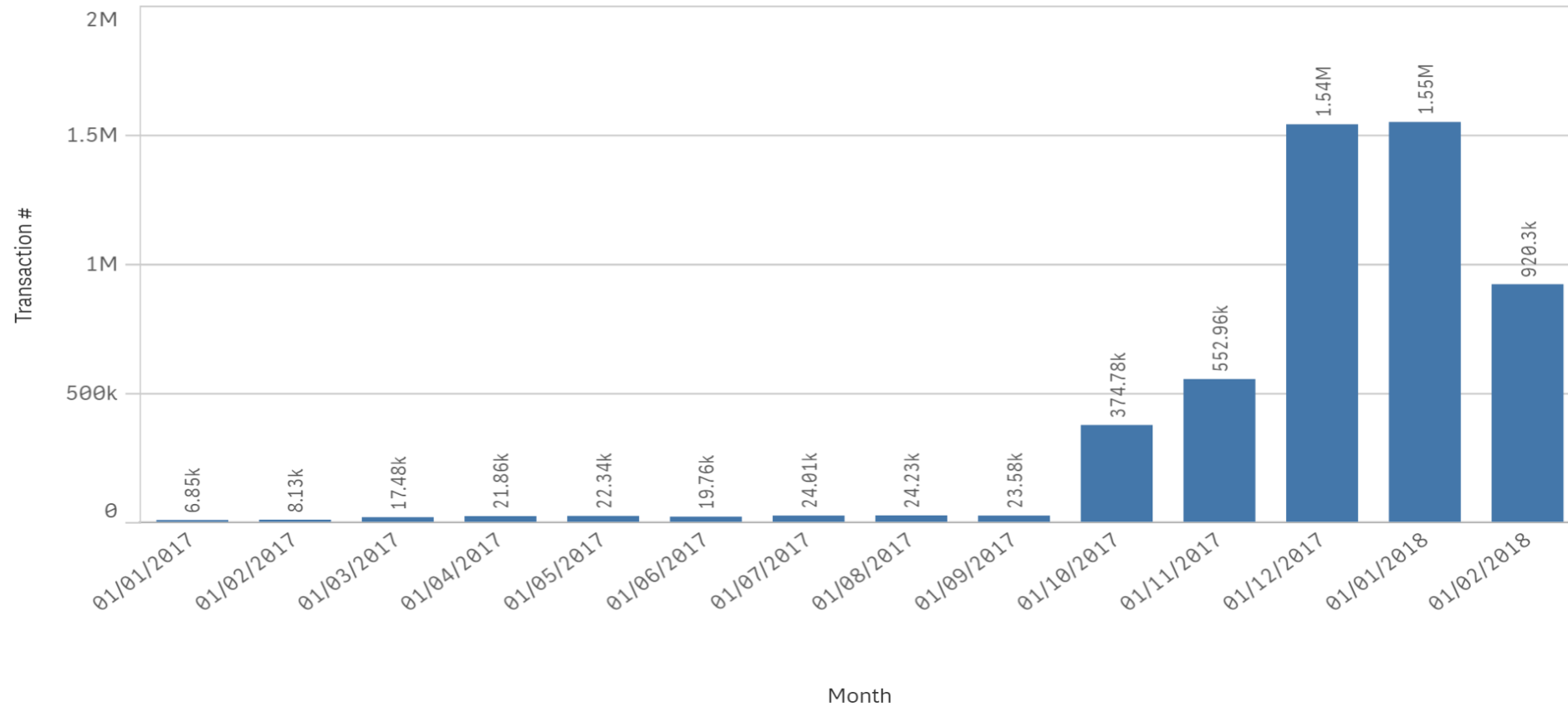
Total Rows	73,202,293
Duplicates	57,421,168
Missing Values in prod_cat	5,851,112
Final Rows #	9,930,013

05/01/2018 8:37:54
2018/01/11 9:50:11 AM
2018/01/11 9:50:11 PM
29-Dec-17 5:25:21 PM
29-Dec-17 5:25:21 AM
11/30/2017 6:47:57 PM

YYYY-MM-DD HH:mm:ss
2018-01-29 18:38:48

Transaction Distribution

Monthly Purchase Transaction





Use Case

GitHub Link: <https://github.com/ykai76/arulesretail>

Thank You

