ANLY 482: ANALYTICS PRACTICUM

UNCOVERING CHINA MARKET-INSIGHTS FOR CHARLES & KEITH THROUGH THE USE OF DATA ANALYTICS

Proposal By:
Group 11
Toh Yan Ying, Edwin Tan Soon Hong, Audrey Jee Zhi Ying
10th January 2016

TABLE OF CONTENTS

1 OBJECTIVES

Through the application of data analytics, this project aims to discover market-insights such as to uncover associations between products purchased and to discover consumer buying trends within C&K's China market.

Based on the insights gathered, provide recommendations on how C&K may apply them to its business to positively affect consumer buying decisions and capture more market share in China.

The key objectives involves:
1. Visualise the overall generic consumer purchasing patterns and behaviour within China
2. Identify best-selling products and item set for different regions or tiers of cities
3. Highlight the differences between transactions of different locations in terms of product specifications or types

Some possible business areas where the project findings could be applied are as follow:
- Strategic placement of products within C&K's product catalogues
- In-store shelving decision
- Strategic marketing promotions to cross-sell and upsell, at a localized market level
- Suggest Market Basket Analysis algorithm solutions for C&K's e-commerce site

2 SECONDARY RESEARCH

In the paper Market Basket Analysis for a Supermarket based on Frequent Item set Matching (D, 2012), it showed an example of how market basket analysis can be used to better determine placement of products, design sales promotion for different segments of customers which will lead to increase in profit of the business.

This paper uses K-Apriori method as its main data mining algorithm. K-Apriori algorithm often demonstrates good performance with the sparse datasets which is similar to our dataset. The algorithm extracts a set of frequent item sets from the data and then pull out rules with the highest information content for different groups of customers by separating customers into different clusters. In our project, this algorithm can allow us to divide transaction data into different clusters initially and then find the frequent item sets and association rules for those categories separately. This is one way our group can approach the project.

3 DATA

Expected Sample Dataset

Transaction ID	Time of Transaction	Store Outlet	Cities	Item	Products' Category	Products ID	Specifications	Cost/RM
013052D	13th December 2014, 13:30:35	CKS144	Suzhou	Platform penny loafers	Shoes (Loafers)	310A	Red Felt Material	499
				Retro Aviators	Accessories (Bags)	422D	Black	359
014500D	13th December 2014, 13:41:12	CKS144	Suzhou	Collar Statement Necklace	Accessories (Necklace)	5001F	Gold	459

4 METHODOLOGY

4.1 DATA COMPILATION

We have identified the following sets of data to be useful in understanding C&K's local market in China:

- Demographic
  - Average age of country/region
  - Income
  - % of Female by Age Category
  - Population Size

- Economic
  - Gross Domestic Product (GDP)
  - GDP per capita

We will be obtaining these data sets for all the cities where C&K have a foothold in. These data will be extracted from China Statistical Yearbook Database and China Knowledge.

4.2 DATA PREPARATION

Data pre-processing is required to allow insights and knowledge discovery accurately. Data preparation operations such as reduction in number of attributes, outlier detection and discretization are performed to significantly increases the model's' accuracy that we will be using for analysis.

4.3 DIMENSIONALITY REDUCTION

Firstly, it is important for us to fully understand the dataset provided to us. As the dataset provided to us are of store transactions in C&K's China market, it is likely that the fields in the dataset are listed in Mandarin. Hence, we will first have to translate those fields into English before proceeding with the analysis. Thereafter, we will perform dimensionality reduction by keeping the data fields or attributes that are relevant and insightful towards our analysis and eliminate those that are not useful and unnecessary.

4.3.1 FILLING AND HANDLING

We will also eliminate the missing data before conducting our analysis. If necessary, these missing values should be filled in using an appropriate approach. Outliers and inaccurate values should be handled and removed from the dataset as well.

4.3.2 DISCRETIZATION/ BINNING

Continuous attributes should be encoded by discretizing the original values into a small number of value ranges as they provides more meaning to the analysis or bins. The main variables that we are focusing on is the price range.

After which, we will categorize the dataset based on the store location where the transactions were made. We will classify the stores based on the China City Tier System (Nexus Pacific, 2013), based on this classification method, i.e. Beijing will be classified as a Tier 1 City and all the stores located in the Tier 1 cities would be grouped together.

Our group believes that this method of classification is a very robust way of grouping C&K's stores in China as it takes into account the city's population size, Gross Domestic Product (GDP), Average economic growth, connectivity as a transpiration hub as well as the city's historical and cultural significance. Furthermore, based on the theories of Consumer Behaviour, customer buying preferences differs based on factors such as level of disposable income, lifestyles and economic environment which is largely characterised by which tier of city one lives in. For example, the buying attitudes of a consumer in an urban metropolis may be different from that of a consumer in a provincial capital. (Assumption: Customer that purchased from a store in Beijing is likely to live and work in Beijing) Hence, by slicing up the data into specific local categories will help provide our group a better understanding of consumer behaviour and trends.

Following, we will bin the data into different categories. Based on the product category list extracted from C&K's e-commerce site, we will bin the different SKU's into their respective product categories. Furthermore, based on the interquartile range of product prices of the different product categories, we will bin the products based on their price range.

4.4 DATA ANALYSIS

4.4.1 PRELIMINARY DATA ANALYSIS

To better understand C&K's local markets, based on the tiers of city, basic reporting would be performed to provide a consolidated overview of the local territories. Areas of reporting includes:

- % contribution of Classified Cities by Tiers to C&K's profits
  - E.g. Tier 1 Cities contribute 80% of C&K's sales revenue
- % of product category contribution of Classified Cities by Tiers to C&K's profits
  - E.g. In Tier 1 Cities, Heels contributes 45% of the tiered city sales revenue.
- Population Size of Classified Cities by Tiers
- Average age of Classified Cities by Tiers
- Average income of Classified Cities by Tiers

4.4.2 MARKET BASKET ANALYSIS (MBA)

It is a collection of undirected data mining methods for discovering customer purchasing patterns by finding associations between different items in customers' shopping carts. This project will focus on the Apriori Algorithm as a means to identify the actionable rules present in the in-store transaction data provided by C&K.

MBA would be conducted on different data levels in attempts to discover different customer purchasing patterns. The different levels this project hopes to explore are as follows:

- Product Category
- Tiers of City
- Product Materials
- Product Price Range

5 WORK PLAN

Project Timeline

Task	In Charge	Man day	W0	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13	W14	W15
PROJECT PREPARATION																		
Sponsors Sourcing	All	5	▓															
Project & Sponsors Finalisation	All	1		▓														
NDA Confirmation	Audrey & Yan Ying	3		▓														
Secondary Research	Edwin	2		▓														
INITIAL PROJECT SUBMISSION																		
Weekly Consultation	All	1		▓														
Initial Project Report	Audrey	5		▓														
Wiki Page Design Set-Up	Yan Ying	1		▓														
Wiki Page Content Update	Edwin	1		▓														
DATA PREPARATION																		
Data Cleanup	All	8			▓	▓												
Data Familiarisation and Understanding	All	2			▓													
Variables Selection	All	2			▓	▓												
DATA ANALYSIS																		
Model Implementation and Selection	Edwin	5				▓	▓											
Market Basket Analysis	Audrey	10					▓	▓										
Trends Identification	Yan Ying	8							▓	▓								
MID-TERM REVIEW																		
Interim Report	Audrey	8								▓	▓							
Presentation Slides	Yan Ying	2									▓							
Wiki Page Content Update	Edwin	2									▓							
Mid Term Presentation and Feedback	All	1										▓						
MODIFICATION & IMPROVEMENT																		
Further Analysis	Audrey	8										▓	▓	▓				
Changes based on Feedback	Edwin	5										▓	▓	▓				
Recommendation	Yan Ying	5												▓	▓			
FINAL PROJECT REVIEW																		
Final Report	Audrey	8													▓	▓		
Final Presentation Slides	Yan Ying	2														▓		
Final Wiki Page Update	Edwin	2														▓	▓	
Final Presentation	All	1																▓

6 REFERENCES

Charles & Keith Brand Profile. (n.d.). Retrieved from http://www.charleskeith.com/us/brand-profile/

China Knowledge. (n.d.). Retrieved from http://www.chinaknowledge.com/Newswires/Default.aspx

China Statistical Yearbook Database. (n.d.). Retrieved from
http://tongji.cnki.net/overseas/brief/result.aspx

D, L. C. (2012). Retrieved from http://ijcsi.org/papers/IJCSI-9-5-3-257-264.pdf

Nexus Pacific. (2013). Retrieved from http://nexus-pacific.com/blog/2013/7/9/china-city-tier-system-how-it-works-and-why-its-useful

NLambassade. (2014). Retrieved from
http://china.nlambassade.org/binaries/content/assets/postenweb/c/china/zaken-doen-in-china/sectoren/creatieve-industrie/china-fashion-industry-kansenrapport.pdf

SCMP. (2015). Retrieved from http://www.scmp.com/business/china-business/article/1820390/fast-fashion-brands-bring-new-look-chinas-retail-property