Application Of Market Basket Analysis To Understand Students' Career Options: A Study On Management Under Graduate Students At IU, Mizoram

** Mrinmoy Bhattacharjee*
*** Pranjal Kalita*

INTRODUCTION

The Higher education sector has grown at a significant rate during the last decade in India. The days when higher education was a matter of national policy and government regulation are rapidly fading. Higher Education provisioning is now globalized, and in many ways, is now a commercialized affair. Since the late 1990s, the higher-education market has been growing by 7 per cent a year. The Economist Survey on higher education indicates that the annual fee income alone is estimated at $ 30 billion. While private profit seeking companies have entered the education business, even government-controlled universities are seeking independence from governmental authority. Hence, a University is no longer only a place where students apply to study. Universities are now actively pursuing students using a wide variety of strategies to market their courses and are providing them guidance and training towards attaining their career goals (icrier.org). Education being a service, customer satisfaction is the ultimate metric of success here (Joshi, 2011). The growth of higher education in India has been phenomenal. Starting with 1950-51, there were only 263,000 students in all disciplines in 750 colleges affiliated to 30 universities. This had grown by 2005 to 11 million students in 17,000 Degree colleges affiliated to 230 universities and non-affiliated university-level institutions. In addition, there are about 10 million students in over 6500 vocational institutions. The enrolment is growing at the rate of 5.1 per cent per annum. However, of the Degree students, only 5 per cent are enrolled into engineering courses, while an overall 20 per cent are enrolled in sciences. The demand for professional courses is higher (Kaul, 2006).

There is a sea of career options in front of the students after their graduation. Some of them opt for higher professional studies, while some prefer jobs, whereas, some would like to build themselves as entrepreneurs. The student population is heterogeneous in nature, and so are their choices and preferences regarding a career. There are lots of factors that govern the process of career selection by the students. Universities have been able to understand the needs of the students to a certain extent. However, there is a need for further research and analysis to understand the trends of career selection by the students pursuing professional graduate courses and provide them coaching and guidance in that direction. One way of effectively addressing this issue is through the analysis and presentation of data or data mining. Data mining enables organizations to use their current capabilities to uncover and understand hidden trends and patterns in vast databases (Luan, 2004). *'Market Basket Analysis'* is one of the data mining techniques, which can help to find out the patterns or trends of the students' career selection after their graduation (Hubinger, 2011).

CONCEPTUAL FRAMEWORK

Market Basket Analysis was formulated by Agrawal et al. in 1993 and is often referred to as the market-basket problem. In this problem, a set of items and their collection of transactions are given, which are subsets (referred to as baskets). The task is to find the relationships between the presences of various items within these baskets (Shalom, 1997). There are numerous applications of data mining which fit into this framework (Pujari, 2008). The problem has been named after its classic application in the supermarket. In the context of the supermarket, the problem is to analyze customers' buying habits by finding *associations* between the different items that customers place in their shopping baskets. The discovery of such association rules can help the retailer develop marketing strategies by gaining insights into matters

* *Senior Faculty Member (Management),* The ICFAI University, Mizoram, Salem Veng, Chaltlang, Aizawl, Mizoram - 796012.
E-mail: mrinmoy_mrin@yahoo.com
** *Faculty Member (IT),* The ICFAI University, Mizoram, Salem Veng, Chaltlang, Aizawl, Mizoram - 796012.
E-mail: pkalita1985@yahoo.com

like *"which items are most frequently purchased by customers"* (scribd.com). Category management, which is fully adopted with the grocery sector, now intends to move towards the fashion sector (Rashid, 2011). It also helps in inventory management, sales promotion strategies, etc. To have a transparent idea about market basket analysis, one needs to go through some of the common concepts and terminologies related to it. The foundation on which market basket analysis stands is the rule of association. It is widely accepted that the discovery of the association rule is solely dependent on the discovery of frequent sets. This rule tries to establish relations between items. The set of items are usually termed as *itemset* in this context. The second important concept in this analysis is support. A transaction t is said to support an item l_i, if l_i is present in t (Mahanta, 2005). t is said to support a subset of items $X \subseteq A$, if t supports each item l in X. An item set $X \subseteq A$ has a *support* s in T, if s % of the transactions in T supports X (rimtengg.com). The support can also be defined as fractional support, which means the proportion of transactions supporting X *in* T (Pujari, 2008). Another important concept in this regard is *confidence*. The confidence of a rule is defined *Conf* $(X \Rightarrow Y)$ = *Supp* $(X \cup Y)$/*Supp*(X). For example, the rule {milk, bread} $\Rightarrow$ {butter} has a confidence of 0.2 / 0.4 = 0.5 in the database, which means that for 50% of the transactions containing milk and bread, the rule is correct. Confidence can be interpreted as an estimate of the probability $P(Y \mid X)$, the probability of finding the RHS of the rule in transactions under the condition that these transactions also contain the *LHS* (Hipp et al., 2000). For a given transaction database T, an *association rule* is an expression of the form $X \Rightarrow Y$, where X and Y are subsets of A and $X \Rightarrow Y$ holds with confidence τ, if τ % of transactions in D that support X also supports Y. The rule $X \Rightarrow Y$ has a support σ in the transaction set T *if* σ % of transactions in T support $X \cup Y$. The intuitive meaning of such a rule is that a transaction of the database which contains X tends to contain Y (Pujari, 2008).

MARKET BASKET ANALYSIS AND STUDENT CAREER OPTIONS

The present study was conducted to find the inter relationships among various factors influencing the management under graduate students' career choice options after their graduation. The prime objective of this study is to ascertain the trends and patterns of students' career choices after their graduation, which would in turn help the institution/ university to develop effective career programs. The outcome of this study would help in scientifically understanding the career inclination of students after their graduation. The study is, however, confined to management undergraduate students at ICFAI University, Mizoram (IUM). For the purpose of study, three consecutive batches of passed out BBA students were considered. Further, the factors which were taken into account to develop tables for market basket analysis were, passing grades of the students (A and B grade), gender of the students, their choice of career after graduation (Job and Higher Studies). The association rule was applied, which includes determination of *support and confidence* for various relationships. The support and confidence are numeric values expressed in terms of percentage. The first batch (2005-08) consisted of 16 number of students and hence, 16 numbers of transactions were considered ($T_1, T_2, T_3 \ldots T_{16}$), the second batch (2006-09) consisted of 11 students and hence, 11 number of transactions were considered ($S_1, S_2, S_3 \ldots S_{11}$) and the third batch (2007-10) consisted of 22 students and thus, 22 number of transactions ($P_1, P_2, P_3 \ldots P_{22}$) were considered . There are six items in the *itemset {A, B, M, F, Job, HE}* considered for each batch where, *A* means students who have secured *'A grade'*, *B* means students who have secured *'B grade'*, *M* stands for *'Male'* students, *F* stands for *'Female'* students, *'Job'* represents those who had opted for the job after graduation and *'HE'* represents those who had opted for *'higher education'* after their graduation. In the tables, '0' represents the absence of a particular item in the transaction and '1' represents the presence of a particular item in the transaction.

Calculation of ***support*** and ***confidence*** for the transactions of 2005-08 student batch is as follows (see Table 1) :

Support (Job) = 68.75%

Support (HE) = 31.25%

A $\Rightarrow$ Job [Support = 25%, Confidence = 100%]

A $\Rightarrow$ HE [Support = 0%, Confidence = 0%]

B $\Rightarrow$ Job [Support = 43.75%, Confidence = 58.33%]

B $\Rightarrow$ HE [Support = 31.25%, Confidence = 41.67%]

M $\Rightarrow$ Job [Support = 25%, Confidence = 57.14%]

M $\Rightarrow$ HE [Support = 18.75%, Confidence = 42.86%]

F $\Rightarrow$ Job [Support = 43.75%, Confidence = 77.78%]

Table 1 : Table Representing The Items And Their Transactions In 2005-08 Batch						
Transaction ID	**A**	**B**	**M**	**F**	**Job**	**HE**
T_1	0	1	1	0	1	0
T_2	0	1	1	0	1	0
T_3	1	0	0	1	1	0
T_4	1	0	1	0	1	0
T_5	0	1	0	1	1	0
T_6	0	1	0	1	1	0
T_7	0	1	0	1	1	0
T_8	0	1	0	1	1	0
T_9	1	0	0	1	1	0
T_{10}	1	0	0	1	1	0
T_{11}	0	1	1	0	0	1
T_{12}	0	1	0	1	0	1
T_{13}	0	1	1	0	1	0
T_{14}	0	1	1	0	0	1
T_{15}	0	1	0	1	0	1
T_{16}	0	1	1	0	0	1

Table- 1 (Source: Authors)

(A= Secured 'A' Grade, B= Secured 'B' Grade, M= Male Students, F=Female Students, Job= Students opted for job, HE=Students opted for higher education)

$F \Rightarrow HE$ [Support = 12.5%, Confidence = 22.22%]
$\{A, M\} \Rightarrow Job$ [Support = 6%, Confidence = 100%]
$\{A, M\} \Rightarrow HE$ [Support = 0%, Confidence = 0%]
$\{A, F\} \Rightarrow Job$ [Support = 18.75%, Confidence = 100%]
$\{A, F\} \Rightarrow HE$ [Support = 0%, Confidence = 0%]
$\{B, M\} \Rightarrow Job$ [Support = 18.75%, Confidence = 50%]
$\{B, M\} \Rightarrow HE$ [Support = 18.75%, Confidence = 50%]
$\{B, F\} \Rightarrow Job$ [Support = 25%, Confidence = 66.67%]
$\{B, F\} \Rightarrow HE$ [Support = 12.5%, Confidence = 33.33%]

Table 1 reflects that 68.75% of the students opted for a job in the first batch, whereas students opting for higher education (HE) were 31.25%. Further, all (100%) the students who secured *'A'* grade in this batch opted for a job but out of all the students securing *'B'* grade, only 58.33% opted for a job and the rest 41.67% went for higher education. Again, if we consider gender wise career option, then 57.14% of the male students went for a job, whereas 77.78% of the female students opted for a job.

Further, considering grades and gender together, it was found that all male and female students securing *'A'* grade went for a job. However, out of all the male students securing *'B'* grade, 50% went for the job and the rest 50% opted for higher education, whereas, in case of female students, the percentage opting for a job (66.67%) was significantly higher than the percentage of female students opting for higher education (33.33%).

Thus, it is clear from the above analysis that most of the students opted for a job, irrespective of the grades and a very insignificant percentage of students opted for higher education, where the percentage of female students was again lower.

Calculation of ***support*** and ***confidence*** for the transactions of 2006-09 student batch is as follows (see Table 2) :

Support (Job) = 27.27%
Support (HE) = 72.72%

Table 2 : Table Representing The Items And Their Transactions In 2006-09 Batch						
Transaction ID	**A**	**B**	**M**	**F**	**Job**	**HE**
S_1	1	0	0	1	0	1
S_2	1	0	1	0	0	1
S_3	0	1	0	1	0	1
S_4	1	0	1	0	0	1
S_5	0	1	1	0	1	0
S_6	0	1	1	0	0	1
S_7	0	1	1	0	0	1
S_8	0	1	1	0	0	1
S_9	0	1	0	1	0	1
S_{10}	0	1	0	1	1	0
S_{11}	0	1	0	1	1	0

Table- 2 (Source: Authors)

(A= Secured 'A' Grade, B= Secured 'B' Grade, M= Male Students, F=Female Students, Job= Students opted for job, HE=Students opted for higher education)

$A \Rightarrow Job\ [Support = 0\%, Confidence = 0\%]$
$A \Rightarrow HE\ [Support = 27.27\%, Confidence = 100\%]$
$B \Rightarrow Job\ [Support = 27.27\%, Confidence = 37.5\%]$
$B \Rightarrow HE\ [Support = 45.45\%, Confidence = 62.5\%]$
$M \Rightarrow Job\ [Support = 9\%, Confidence = 16.67\%]$
$M \Rightarrow HE\ [Support = 45.45\%, Confidence = 83.33\%]$
$F \Rightarrow Job\ [Support = 18\%, Confidence = 40\%]$
$F \Rightarrow HE\ [Support = 27.27\%, Confidence = 60\%]$
$\{A, M\} \Rightarrow Job\ [Support = 0\%, Confidence = 0\%]$
$\{A, M\} \Rightarrow HE\ [Support = 18.18\%, Confidence = 100\%]$
$\{A, F\} \Rightarrow Job\ [Support = 0\%, Confidence = 0\%]$
$\{A, F\} \Rightarrow HE\ [Support = 9\%, Confidence = 100\%]$
$\{B, M\} \Rightarrow Job\ [Support = 9\%, Confidence = 25\%]$
$\{B, M\} \Rightarrow HE\ [Support = 27.27\%, Confidence = 75\%]$
$\{B, F\} \Rightarrow Job\ [Support = 18.18\%, Confidence = 50\%]$
$\{B, F\} \Rightarrow HE\ [Support = 18.18\%, Confidence = 50\%]$

Table 2 reflects a contrary scenario than Table1. In the 2006 -09 batch, only 27.27% of the students went for a job, whereas, 72.72% of the students went for higher education. No student securing *'A'* grade opted for a job, i.e. all of them went for higher education. The trend is more or less similar in case of students securing *'B'* grade also i.e. out of all the students securing *'B'* grade, only 37.5% went for a job, and the rest opted for higher education.

Even gender wise analysis depicts that a high percentage of male and female students opted for higher education than a job. Again, joint consideration of grades and gender reveals that both male and female students securing *'A'* grade opted for higher education unlike the previous batch, where all the students securing *'A'* grade went for a job. In the case of students securing *'B'* grade also, 75% and 50% students went for higher education among males and females respectively.

Though there was a high rise in the overall percentage of students opting for higher education in the second batch of students, still, the percentage of female students among them was lower in comparison to male students in both the years.

Table 3 : Table Representing The Items And Their Transactions In 2007-10 Batch

Transaction ID	A	B	M	F	Job	HE
P_1	1	0	1	0	0	1
P_2	1	0	1	0	0	1
P_3	1	0	1	0	0	1
P_4	1	0	0	1	0	1
P_5	1	0	1	0	0	1
P_6	1	0	1	0	0	1
P_7	0	1	1	0	0	1
P_8	1	0	0	1	0	1
P_9	1	0	0	1	0	1
P_{10}	0	1	1	0	0	1
P_{11}	1	0	1	0	0	1
P_{12}	0	1	1	0	0	1
P_{13}	1	0	1	0	0	1
P_{14}	0	1	0	1	0	1
P_{15}	0	1	0	1	0	1
P_{16}	0	1	0	1	0	1
P_{17}	0	1	0	1	0	1
P_{18}	0	1	0	1	1	0
P_{19}	0	1	1	0	1	0
P_{20}	0	1	1	0	0	1
P_{21}	0	1	1	0	0	1
P_{22}	0	1	1	0	0	1

Table- 3 (Source: Authors)

(A= Secured 'A' Grade, B= Secured 'B' Grade, M= Male Students, F=Female Students, Job= Students opted for job, HE=Students opted for higher education)

Calculation of ***support*** and ***confidence*** for the above transactions of 2007-10 students batch is as follows (see Table 3) :

Support (Job) = 9%

Support (HE) = 91%

$A \Rightarrow$ Job [Support = 0%, Confidence = 0%]

$A \Rightarrow$ HE [Support = 45.45%, Confidence = 100%]

$B \Rightarrow$ Job [Support = 9%, Confidence = 16.67%]

$B \Rightarrow$ HE [Support = 45.45%, Confidence = 83.33%]

$M \Rightarrow$ Job [Support = 4.5%, Confidence = 7.15%]

$M \Rightarrow$ HE [Support = 59%, Confidence = 92.85%]

$F \Rightarrow$ Job [Support = 4.5%, Confidence = 12.5%]

$F \Rightarrow$ HE [Support = 31.8%, Confidence = 87.5%]

$\{A, M\} \Rightarrow$ Job [Support = 0%, Confidence = 0%]

$\{A, M\} \Rightarrow$ HE [Support = 31.81%, Confidence = 100%]

$\{A, F\} \Rightarrow$ Job [Support = 0%, Confidence = 0%]

$\{A, F\} \Rightarrow$ HE [Support = 13.63%, Confidence = 100%]

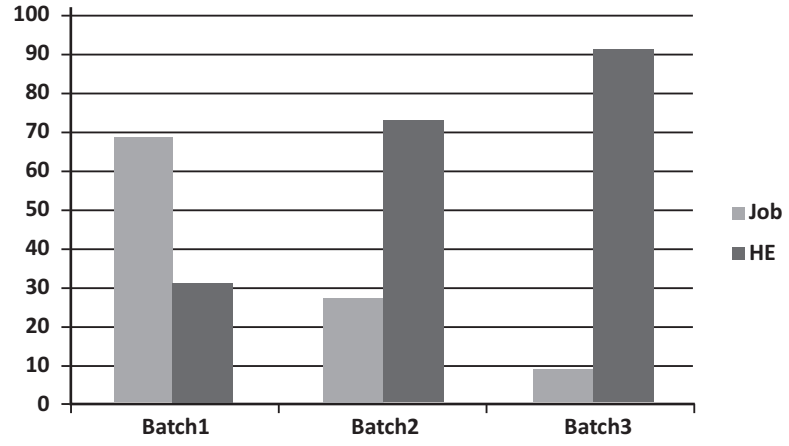
$\{B, M\} \Rightarrow$ Job [Support = 4.5%, Confidence = 14.29%]

{B, M} ⇒ HE [Support = 27.27%, Confidence = 85.71%]
{B, F} ⇒ Job [Support = 4.5%, Confidence = 20%]
{B, F} ⇒ HE [Support = 18.18%, Confidence = 80%]

Table 3 imitates the same trend of Table 2. In this batch also, there was a consecutive rise in percentage of students opting for higher education, where 91% of the students opted for higher studies in the field of management. Like the previous batch, all the students securing *'A'* grade, irrespective of gender, had opted for higher education. In case of students securing *'B'* grade, 85.71% of the males and 80% of the females selected higher education in their next venture. Percentage of female students continued to be lower than the percentage of male students in higher education, though the difference was marginal for this (3rd) batch.

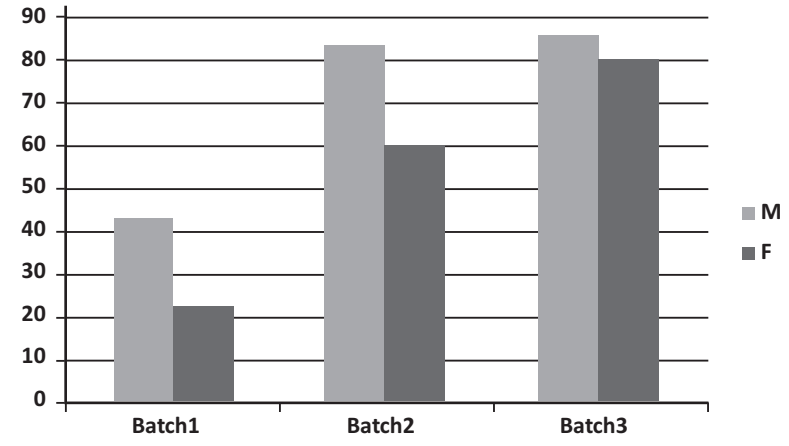
The batch wise analysis has been combined and is represented in the form of graphs for easy interpretation and for drawing conclusion (see Figures 1, 2 and 3).

Figure 1: Job Vs. Higher Education - Batchwise

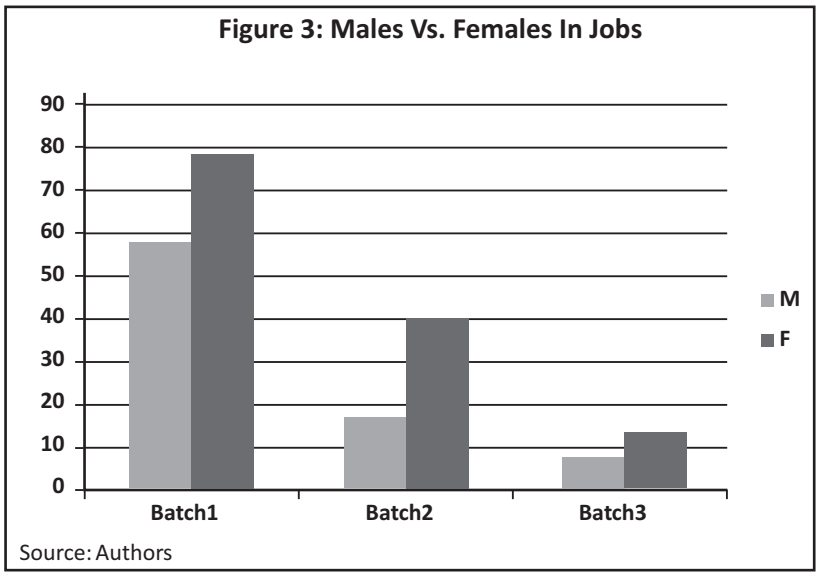


Source: Authors

Figure 2 : Males Vs. Females In Higher Education



Source: Authors



The first graph (Figure-1) clearly shows that there is a constant rise in the preference of students opting for higher education and there was a fall in preference for jobs. The second graph (Figure-2) reflects the decline in difference between male and female students opting for higher education, and the third graph (Figure-3) depicts the declining preference for jobs (gender wise) among the graduate students .

On the basis of overall findings, it is suggested that the university/institution needs to work towards satisfying the requirements of the students through providing guidance and mentoring in that direction. Most of the professional institutions and universities focus more on placement facilities after graduation for showing their supremacy and for luring students than providing them with active guidance towards their field of interest. In this case, it has been found that graduate students tend to pursue higher education more than choosing a job as a career option after graduation. Further, it is also understandable that students select those institutions or universities for their higher education which provide them with better facilities, ambience and opportunities. So, the universities/institutions need to improvise and develop their higher education facilities by developing infrastructure, both physical and intellectual, in order to retain the students after graduation. The migration of students to other institutes/universities after graduation can be gradually reduced by providing them what they want in the existing institution. Education is a service and thus, its search quality[1] cannot be judged like physical products. Over a period of time, the students have to realize and understand the credence quality[2] of it, which is only possible by providing superior facilities.

CONCLUSION

The higher education industry has become highly competitive and challenging in today's context. The dynamics of students' career choices have become highly complex too. Universities and professional institutions are no longer a place for mere knowledge attaining, but a center to actively guide and mentor students for achieving their career goal. In this race, these professional institutions and universities are indulging more into number game of placements, rather than providing guidance and facilitating the process to let the students achieve their personally desired goal. There is a greater need for understanding and realizing the students' interest by these institutions and universities. In doing so, the universities and institutions can benefit themselves by gaining the trust and confidence of the students and thereby, retaining them for further education.

REFERENCES

1. A. Kakoti Mahanta (2005) . *"Finding Locally And Periodically Frequent Sets And Periodic Association Rules"*, Lecture Notes in Computer Science, 2005.

[1] The quality of a product which can be judged by the customers before buying it.

[2] The quality of a product which can be judged only after using it.

2. B. Hubinger (2011) . "Influence Of Initial Heterogeneities And Recharge Limitations On The Evolution Of Aperture Distributions In Carbonate Aquifers", *Hydrology and Earth System Sciences Discussions*, 06/14/2011.

3. http://cran.fhcrc.org/web/packages/arules/vignettes/arules.pdf accessed on February 12, 2011.

4. http://www.icrier.org/pdf/WP_179.pdf accessed on January 10, 2011.

5. http://www.rimtengg.com/coit2007/proceedings/pdfs/42.pdf accessed on January 2, 2011.

6. http://www.scribd.com/doc/11573314/p139-Data-Mining-Mafia accessed on January 10, 2011.

7. Jochen Hipp, Ulrich Güntzer, and Gholamreza Nakhaeizadeh (2000). 'Algorithms For Association Rule Mining - A General Survey And Comparison'. *SIGKDD Explorations*, Volume 2, Number 2, pp. 1-58.

8. Joshi Sandhya (2011) . 'A Study Of Service Quality And Customer Satisfaction Across Various Service Providers In The Telecom Sector', *Indian Journal of Marketing,* Vol. 41, No. 9, p.55.

9. Kaul Sanat (2006). 'Higher Education In India: Seizing the Opportunity', May, *ICRIER*, pp. 14-15.

10. Luan Jing (2004). *'Data Mining Application in Higher Education'.* Executive report, SPSS Inc. pp. 22-24

11. Pujari Arun K. (2008). *'Data Mining Techniques'.* Universities Press (India), 10th Edition, pp. 69-74.

12. Rashid Sanuwar Md. (2011) . 'The Structured Frame Of Category Management For Optimization Of Retail Business', *Indian Journal of Marketing*, Vol. 41, No. 9, p. 45.

13. Shalom Tsur (1997). *"Dynamic Itemset Counting And Implication Rules For Market Basket Data"*, Proceedings of the 1997 ACM SIGMOD international conference on Management of data -SIGMOD 97, 1997.