भारतीय प्रबंध संस्थान बेंगलूर
INDIAN INSTITUTE OF MANAGEMENT
BANGALORE

IMB 555

CUSTOMER ANALYTICS AT flipkart.com

NAVEEN BHANSALI, JITENDRA RUDRAVARAM, SHAILAJA GROVER AND U. DINESH KUMAR

Naveen Bhansali, Jitendra Rudravaram, Shailaja Grover and U Dinesh Kumar, Professor of Quantitative Methods and Information Systems, prepared this case for class discussion. This case is not intended to serve as an endorsement, source of primary data, or to show effective or inefficient handling of decision or business processes.

It was typical cloudy monsoon weather at Bangalore on July 28, 2015. In the Darwin room of Flipkart's Cessna Business Park office, Ravi Vijayaraghavan, the head of analytics and Pravin Shinde, senior manager Analytics were brainstorming various business problems that Flipkart as an e-commerce company was encountering. Flipkart had been putting in much effort and emphasis on the use of Analytics in every aspect of decision making. Forecasting demand for thousands of stock-keeping units (SKUs), predicting returns and cancellations of orders, predicting the reasons when customers contact the customer service centers, optimizing markdown pricing, identifying various types of frauds, optimizing vehicle routing, and enabling adherence to service-level agreements, were some of the typical problems that the analytics division of Flipkart was solving using state-of-the-art analytics techniques. In 2015, the team included about 100 data scientists mostly recruited from institutes such as the Indian Institute of Technology and Indian Institute of Management specifically for this purpose.

E-commerce in India had seen a compound annual growth rate (CAGR) of 34% since 2009 and was expected to exceed USD 22 billion by 2015[1] (**Exhibit 1**). Under the e-commerce head, e-travel in itself comprised 71% of the total e-commerce market, e-tailing which comprised online retail, and online marketplaces have been growing exponentially and well-poised to become the fastest growing segment, expected to reach USD 56 billion by 2023 (**Exhibit 2**).[2] The industry believed that growth was at an inflection point with the key drivers being broadband internet, rising standards of living, wider product range, and changing lifestyles of Indian consumers. Such high growth rate also created several business challenges to e-commerce companies as well as to their market place suppliers, among them profitability still remained a major challenge. The e-commerce companies in India incurred combined losses of around INR 10 billion[3] through heavy discounting to penetrate into the brick and mortar retail customer base.

Ravi Vijayaraghavan started the meeting by stating:

> We have been analyzing our data to gain insights, but, do we know the value of our customers? I think it is important for us to differentiate our customers through metrics such as customer lifetime value, which will help us to manage them effectively. For example, we can make our promotions effective if we know the customers with high customer lifetime value.

Customer lifetime value (CLV) is the net present value (NPV) of future cash flows (or profit). CLV is usually calculated at a customer segment level. The main challenge in calculating the lifetime value of customers of e-commerce companies such as Flipkart is that the exact life of the customer is unknown owing to data truncation; that is, the actual point in time of customer churn, may not be identified in e-commerce, since there would be no prior communication from the customer about the churn. Hence,

[1] Source: IndianOnlineseller.com - http://indianonlineseller.com/2014/08/delhi-biggest-online-shoppers-1-4-indian-buys-mobile/
[2] Source: Study of M&A scenario in Indian e-commerce market- http://www.novonous.com/case-studies/analysis-ma-scenario-indian-e-commerce-market
[3] Source: Live mint: http://www.livemint.com/Industry/5hz6UnSAB9gaAeZ4OobwMM/Online-retailers-losses-total-Rs1000-crore-so-far.html

traditional models of CLV calculation may not be appropriate for e-commerce companies such as Flipkart.

ABOUT FLIPKART

Flipkart, the poster child of Indian e-commerce, was an early entrant in the nascent Indian e-commerce market and quickly established itself as the leading company in this space. It was founded in 2007 by Sachin Bansal and Binny Bansal, both alumni of the Indian Institute Technology, Delhi. They pooled in INR 2,00,000 (approximately USD 3,150) each to start Flipkart in 2007. From a startup with an investment of just INR 4,00,00 (approximately USD 6,300), Flipkart had grown into an online retail giant, valued at over USD 15.2 billion as of 2015. Flipkart was running the marathon with ample support from private equity players such as Tiger Global, which invested over USD 1 billion as of 2015.[4] Flipkart sold over 30 million products from more than 50,000 sellers in 70+ categories and consisted of 30 exclusive brand associations, with in-a-day guarantee in 50 cities and same-day guarantee in 13 cities. Flipkart was 33,000 people strong and had over 50 million registered users with over 10 million daily visits and 8 million shipments per month. A burgeoning consumer class, coupled with a rising web-literate population and zealous venture capital funding propelled Flipkart to become India's answer to Alibaba and Amazon.

The use of e-commerce to buy products and services was spreading at a fairly rapid pace in the psyche of the Indian consumer. In Indian cities such as Bangalore, lack of time, ease of shopping, and attractive pricing were major drivers for online shopping. On the other hand, accessibility to a variety of products encouraged customers from smaller towns and cities to opt for the online route.

CUSTOMER ANALYTICS AT FLIPKART

E-commerce companies such as Flipkart had access to huge amount of data, available for applying predictive, and prescriptive analytics to take data-driven decisions. Flipkart had a strong analytics team headed by Ravi Vijayaraghavan, which used statistical models and machine learning algorithms to generate crucial customer insights. Flipkart had more than 50 million registered users and the transaction data of these customers could be used in a far more meaningful way using analytics to predict online consumer behavior.

In 2015, the Indian e-commerce market space was facing immense competition owing to the entry of Jabong, HomeStop18, Infibeam, Indiaplaza, Snapdeal, and a plethora of other pure-play and multi-channel e-commerce companies. The continued growth of e-commerce and tough competition compelled Flipkart to seek a competitive advantage through more sophisticated analytics. Flipkart has been taking several analytics-enabled decisions, for instance, using web analytics to determine which landing pages encourage customers to make a purchase as well as which pay per click ad campaigns were most

[4] Source: http://www.business-standard.com/article/specials/tiger-global-flipkart-s-largest-investor-and-second-largest-stockholder-in-amazon-115111701126_1.html

effective. In the face of tremendous competition, more than ever before, the analytics team at Flipkart aimed to predict customer demand for the products, understand its customer's loyalty, assess the true impact of customer retention strategies (discounts, coupons, and extra services), and focus on customer segments with higher retention and spend potential.

In 2015, Flipkart wanted to understand its customers better and retain most of them through effective promotions, since customer retention is less expensive as compared to customer acquisition. Unlike the churn in the telecom sector, which was clearly defined and captured (in the instance of postpaid customers), churn for e-commerce companies was difficult to define and capture, as these events were unobserved. Across e-commerce companies, the customer churn may be very high owing to reasons such as need fulfilment, cessation of demand, competition, and so on. However, it was important to capture customer churn and identify which customers should be retained.

CHURN ANALYSIS AND LIFETIME VALUE

E-commerce companies faced a scenario of inconsistent customer purchase pattern wherein the gap between purchases could stretch far more than 6 months. Even though most of these buyers could come back after a gap of 5–6 months, Flipkart aimed to identify high value customers, and subsequently increase purchase traction among them.

The analytics team at Flipkart wanted to model customer purchase patterns, repeat buyer trends and calculate churn probabilities to help them identify the repeat customer segment to focus more on these customers for their marketing and promotional strategies. The final objectives of this exercise were to forecast the revenue generated from existing customers and calculate their lifetime value.

DATA DESCRIPTION

In order to carry out a detailed customer value assessment addressing customer churn issues, Flipkart collected sample transactional data spanning across 2 years: January 2013 to December 2014 such that all the 30,000 customers in the sample had made at least one purchase in January 2013. This was done to ensure new customers in the above said period were excluded from the study. Variable description of the data is provided in **Exhibit 3**.

DATA ANALYSIS

To understand customer churn and lifetime value, Pravin's team decided to use Discrete Time Markov Chains (DTMC). To build the churn model, the team first had to identify the period of inactivity (gap between transactions) to define churn. Gap was thus defined as the difference in months between two successive purchases or the difference between the current month (despite no purchase) and the last purchase month. The team put together the frequency distribution of all the purchases at different gaps, provided in **Exhibit 4.** This frequency table was used to define churn, that is, the absorbing state and built a DTMC as shown in **Exhibit 5.** The states of this Markov Chain were defined on the basis of recency of purchase as shown in **Exhibit 6**. The Transition Probability Matrix (TPM) for this DTMC was computed using the transaction data and is shown in **Exhibit 7**.

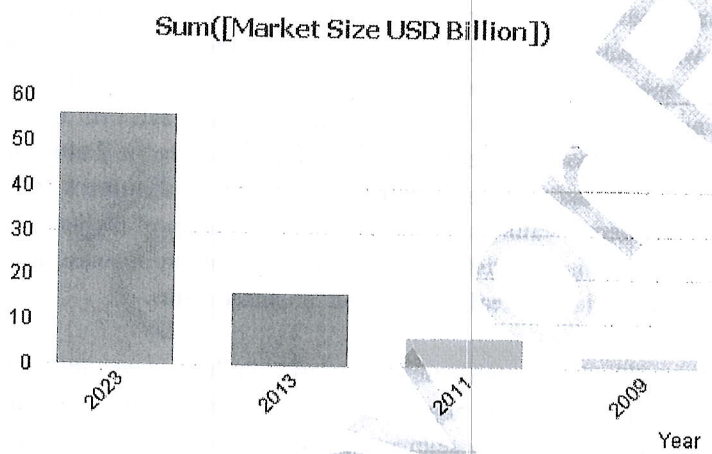
Pravin's team also wanted to forecast revenue from existing customers and therefore built a model using Recency (defined as when was the last purchase was made by the customer) and Monetary (defined as how much money was spent in the latest month which had a purchase). The team retained the recency state definitions and augmented the state space by adding monetary slabs for each recency level. These monetary slabs were identified from the frequency distribution provided in **Exhibit 8**. The extended state space (combination of recency and frequency) definition is shown in **Exhibit 9**. Whenever a customer is in an inactive state, the monetary values are retained from the month of the last purchase. The TPM for the Recency–Monetary DTMC is provided as an accompanying file to the case. **Exhibit 10** includes the average monetary value for each of the earlier identified monetary slabs.

The team was also keen to identify customer segments based on their current transactional attributes such as recency, monetary, and frequency. The objective was to provide a fairly accurate mechanism to study the purchase patterns of various customer segments and thereby enable effective promotions to increase customer spend and arrest customer churn. The team chose a quarterly transition time period to account for macro-level stochastic changes in the model. **Exhibit 11** shows the frequency distribution for recency, frequency (defined as the number of distinct days in the quarter when a purchase was made) and monetary at a quarterly level. The results from **Exhibit 11** were used to define a RFM-based DTMC state space, shown in **Exhibit 12**. Whenever a customer is in an inactive state, the frequency and monetary values are retained from the quarter of the last purchase. **Exhibit 13** shows the customer sub-segments in the active state, that is, Recency 1. **Exhibit 14** shows the truncated TPM with states from 1–8 (active states) and states from 9–33 (inactive states) clubbed together.

The transition probability matrices from each of the models were used to calculate the customer lifetime value for customers in each segment and subsequently build an effective campaign strategy to reduce churn and increase customer spend.

Exhibit 1

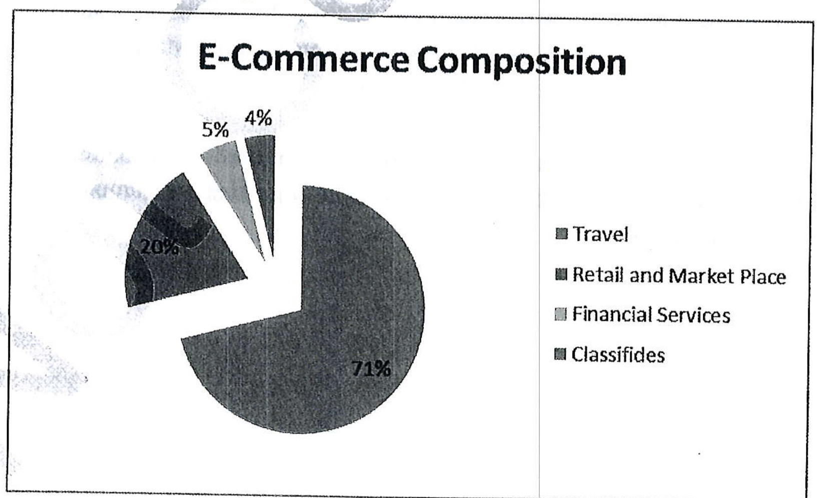
Growth of Indian e-commerce



Source: IndianOnlineseller.com - http://indianonlineseller.com/2014/08/delhi-biggest-online-shoppers-1-4-indian-buys-mobile/

Exhibit 2

Composition of Indian e-commerce sector



Source: Study of M&A scenario in Indian e-commerce market- http://www.novonous.com/case-studies/analysis-ma-scenario-indian-e-commerce-market

Exhibit 3

Unique Identifier: account_id (Account id of the customer)

Variable	Description
item_selling_price	Selling price at the time of purchase
order_creation_date	Date when order was placed
Pincode	Address pin code
product_id	Product id of the product being purchased
unit_quantity	Quantity ordered
unit_status	Order delivery status

Source: Primary Data from Flipkart

Exhibit 4

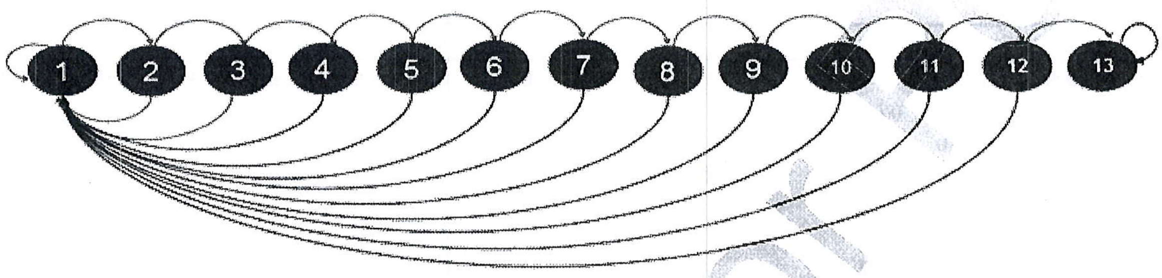
Gaps between transactions (measured in months) in purchase by customers

Gaps	Count of Purchases	%	Cumulative %
0	91806	52.80	52.80
1	31969	18.38	71.18
2	16423	9.44	80.63
3	9649	5.55	86.17
4	6263	3.60	89.78
5	4449	2.56	92.33
6	3165	1.82	94.15
7	2323	1.34	95.49
8	1638	0.94	96.43
9	1319	0.76	97.19
10	973	0.56	97.75
11	797	0.46	98.21
12	585	0.34	98.55
13	492	0.28	98.83
14	394	0.23	99.06
15	360	0.21	99.26
16	283	0.16	99.42
17	253	0.15	99.57
18	198	0.11	99.68
19	148	0.09	99.77
20	152	0.09	99.86
21	129	0.07	99.93
22	120	0.07	100.00
Total	**173888**		

Source: Based on the data provided by Flipkart

Exhibit 5

Transition diagram between recency states



Source: Based on the data provided by Flipkart

Exhibit 6: **Recency states**

State	Recency Level	Explanation
1	1	Purchase made this month
2	2	Purchase made one month ago
3	3	Purchase made two months ago
4	4	Purchase made three months ago
5	5	Purchase made four months ago
6	6	Purchase made five months ago
7	7	Purchase made six months ago
8	8	Purchase made seven months ago
9	9	Purchase made eight months ago
10	10	Purchase made nine months ago
11	11	Purchase made ten months ago
12	12	Purchase made eleven months ago
13	**13**	**Purchase made twelve months ago, hence churned**

Source: Based on the data provided by Flipkart

Exhibit 7

One-step transition probability matrix (recency states)

States	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0.511	0.489	0	0	0	0	0	0	0	0	0	0	0
2	0.365	0	0.635	0	0	0	0	0	0	0	0	0	0
3	0.300	0	0	0.700	0	0	0	0	0	0	0	0	0
4	0.244	0	0	0	0.756	0	0	0	0	0	0	0	0
5	0.205	0	0	0	0	0.795	0	0	0	0	0	0	0
6	0.180	0	0	0	0	0	0.820	0	0	0	0	0	0
7	0.153	0	0	0	0	0	0	0.847	0	0	0	0	0
8	0.137	0	0	0	0	0	0	0	0.863	0	0	0	0
9	0.105	0	0	0	0	0	0	0	0	0.895	0	0	0
10	0.103	0	0	0	0	0	0	0	0	0	0.897	0	0
11	0.091	0	0	0	0	0	0	0	0	0	0	0.909	0
12	0.079	0	0	0	0	0	0	0	0	0	0	0	0.921
13	0	0	0	0	0	0	0	0	0	0	0	0	1

Source: Based on the data provided by Flipkart

Exhibit 8

Monetary value definition

Monetary Amount	**Frequency**	**Percentage**	**Cumulative Percentage**
499	74787	36.75	36.75
999	45780	22.50	59.25
1999	36828	18.10	77.35
4999	28090	13.81	91.16
9999	9917	4.87	96.03
229093	8075	3.97	100.00
Total	203477	100	-

Exhibit 9

State definition using recency and monetary

State	Recency Level	Monetary Level	Explanation
1	1	>9999	Purchase made this month for value higher than Rs. 9999
2	1	4999-9999	Purchase made this month for value between Rs. 4999-9999
3	1	1999-4999	Purchase made this month for value between Rs. 1999-4999
4	1	999-1999	Purchase made this month for value between Rs. 999-1999
5	1	499-999	Purchase made this month for value between Rs. 499-999
6	1	<499	Purchase made this month for value lower than Rs. 499
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
67	12	>9999	Purchase made twelve months ago for value higher than Rs. 9999
68	12	4999-9999	Purchase made twelve months ago for value between Rs. 4999-9999
69	12	1999-4999	Purchase made twelve months ago for value between Rs. 1999-4999
70	12	999-1999	Purchase made twelve months ago for value between Rs. 999-1999
71	12	499-999	Purchase made twelve months ago for value between Rs. 499-999
72	12	<499	Purchase made twelve months ago for value lower than Rs. 499
73	13	-	Purchase more than 12 months back hence churned

Source: Based on the data provided by Flipkart

Exhibit 10

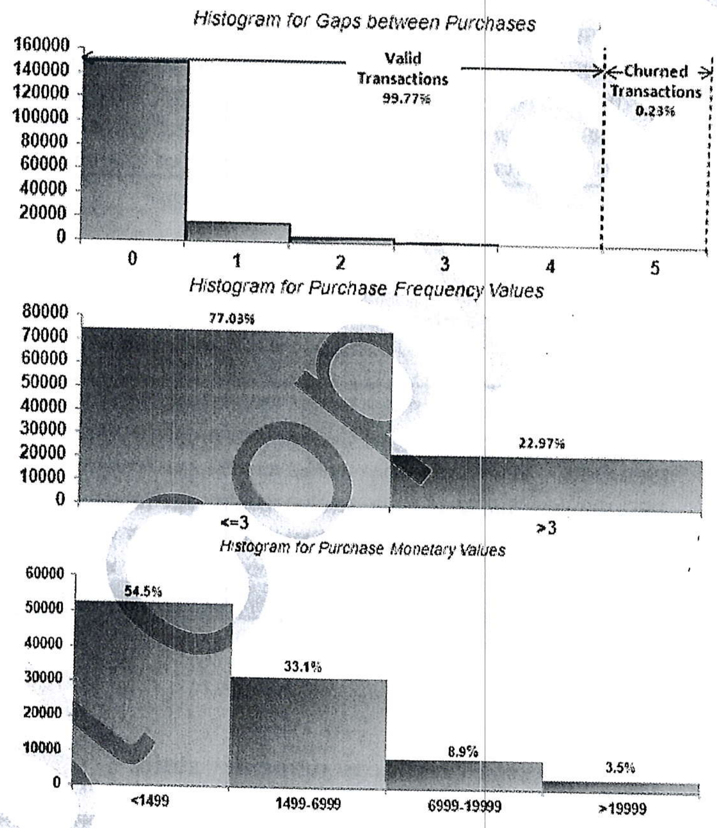
Average revenue in monetary states

State	State 1	State 2	State 3	State 4	State 5	State 6
Average Revenue / Level	22032	6977	3114	1423	720	304

Source: Based on the data provided by Flipkart

Exhibit 11

Histogram of gaps (measured in quarters) between purchases



Source: Based on the data provided by Flipkart

Exhibit 12

RFM Based States

State	*R*	*F*	*M*
1	1-3	>3	>19999
2	1-3	>3	6999-19999
3	1-3	>3	1499-6999
4	1-3	>3	<1499
5	1-3	≤3	>19999
6	1-3	≤3	6999-19999
7	1-3	≤3	1499-6999
8	1-3	≤3	<1499
9	4-6	>3	>19999
10	4-6	>3	6999-19999
11	4-6	>3	1499-6999
12	4-6	>3	<1499
13	4-6	≤3	>19999
14	4-6	≤3	6999-19999
15	4-6	≤3	1499-6999
16	4-6	≤3	<1499
17	7-9	>3	>19999
18	7-9	>3	6999-19999
19	7-9	>3	1499-6999
20	7-9	>3	<1499
21	7-9	≤3	>19999
22	7-9	≤3	6999-19999
23	7-9	≤3	1499-6999
24	7-9	≤3	<1499
25	10-12	>3	>19999
26	10-12	>3	6999-19999
27	10-12	>3	1499-6999
28	10-12	>3	<1499
29	10-12	≤3	>19999
30	10-12	≤3	6999-19999
31	10-12	≤3	1499-6999
32	10-12	≤3	<1499
33	13-15	All	All

Source: Based on the data provided by Flipkart

Exhibit 13

Customer segments based on recency and frequency

Segments	Crème de la crème		High Spenders		Medium Spenders		Budget Buyers	
States	1	5	2	6	3	7	4	8
Frequency	>3	≤3	>3	≤3	>3	≤3	>3	≤3
Monetary Spend Value	>19999		6999-19999		1499-6999		<1499	
No of Customers*	458		1338		4439		6378	
Average Revenue	39480	32273	11347	11326	3636	2984	1116	631

*As per Q1 in the sample dataset
Source: Based on the data provided by Flipkart

Exhibit 14: Transition probability matrix with RM states

	Crème de la crème		High Spenders		Medium Spenders		Budget Buyers		Inactive
States	1	5	2	6	3	7	4	8	9 to 33
1	0.25	0.03	0.23	0.04	0.14	0.13	0.01	0.1	0.07
2	0.09	0.01	0.24	0.03	0.21	0.16	0.01	0.16	0.08
3	0.02	0.01	0.1	0.02	0.28	0.17	0.03	0.28	0.1
4	0.01	0	0.04	0.01	0.2	0.1	0.07	0.41	0.15
5	0.05	0.05	0.03	0.06	0.07	0.2	0.01	0.25	0.29
6	0.03	0.02	0.05	0.07	0.07	0.22	0	0.24	0.3
7	0.01	0.01	0.04	0.04	0.09	0.21	0.01	0.29	0.3
8	0.01	0.01	0.02	0.02	0.06	0.13	0.01	0.36	0.38
9	0.03	0.01	0.03	0.14	0.01	0.16	0.01	0.35	0.26
10	0.05	0.01	0.04	0.05	0.06	0.15	0.01	0.24	0.4
11	0.01	0.01	0.03	0.02	0.09	0.17	0.01	0.31	0.36
12	0	0	0	0.01	0.07	0.11	0	0.4	0.41
13	0.02	0.03	0.01	0.08	0.02	0.1	0	0.19	0.54
14	0.01	0.02	0.01	0.06	0.03	0.12	0	0.21	0.54
15	0.01	0.01	0.01	0.03	0.03	0.15	0	0.24	0.52
16	0	0.01	0.01	0.01	0.02	0.08	0	0.28	0.59
17	0	0.11	0.03	0.03	0	0.07	0	0.18	0.58
18	0	0	0.01	0.04	0.03	0.16	0	0.1	0.65
19	0.01	0	0.01	0.01	0.02	0.1	0.02	0.25	0.57
20	0	0	0	0.02	0	0.08	0.04	0.26	0.61
21	0	0.03	0	0.04	0.01	0.1	0	0.17	0.65
22	0	0.01	0	0.04	0.03	0.09	0	0.16	0.66
23	0	0.01	0	0.02	0.01	0.1	0	0.17	0.68
24	0	0	0	0.01	0.01	0.05	0	0.21	0.71
25	0	0	0	0	0	0	0	0	1
26	0	0.01	0.01	0.05	0.01	0.01	0	0.05	0.85
27	0.01	0	0.01	0	0.01	0.06	0	0.16	0.75
28	0	0	0	0	0	0.07	0	0.2	0.73
29	0	0.04	0	0.06	0	0.12	0	0.14	0.63
30	0	0.01	0.01	0.03	0	0.05	0.01	0.09	0.8
31	0	0	0.01	0.01	0.01	0.06	0	0.13	0.78
32	0	0	0	0.01	0	0.04	0	0.15	0.79
33	0	0	0	0	0	0	0	0	1

Source: Based on the data provided by Flipkart