Arrival pattern of customers

In order to understand why queues form operations managers have done extensive studies in the quest of solving long frustrating lines that send customers away. According to Maister (1985), queues form when the demand for a particular product or service exceeds the capacity of the firm providing the service/product. Reasons for this may be a shortage of available products in stock or servers, or there may be limitations to the available space where the service or product is provided. In order to appreciate the challenges experienced managers must know how much service or products should be made available. It should take into considerations the average length of waiting for a service, number of people in a waiting line, and the duration of providing the service rate. All these factors can be taken into consideration with the use of the queuing theory, which will be discussed in the following section (Kostecki M., 1996).

2. 1. 2. Queuing Theory

Every queue takes place within a system that has processes where customers come into the firm for a service or product. When the server is not immediately available, he will join the line at the end of the queue and eventually leaves the system once he has been served (Maister, 1985). The word ' customers' does not necessarily infer a true human being, but rather, it is used in a general sense, and it could represent a car in a traffic light, a call waiting to be routed or airplane waiting to take off. The queuing theory is the study of a queue through the use of mathematical modeling to evaluate the efficiency of queues. It is the foundation to resolving and establishing an effective queue management.

According to Schlechter amd Kira (2009), Queuing theory is the statistical research of queues. How long the queue is and how long it takes to wait in the line is predicted using the queuing theory. As a branch of operations research queuing theory is helps in making business decisions about the resources needed to provide effective service. Since Agner Krarup Erlang first described the Copenhagen telephone exchange, queuing theory has since been used in waiting lines management (Sundrapandian, 2009)

2. 1. 3. The Queue Management Concept

There are many types of models in the queue management system and different authors have grouped them differently.

Shortest Processed First

The most commonly used in restaurants and cafes is the shortest processed first system (SPF). Averill, (1993) accurately described it as the system that prioritizes services that are expected to take a short length are dealt with sooner. This system can work well but is problematic if the customers do not perceive the right degree of fairness from the system. This necessitates an explanation to customers why these customers are being served in an order other than the simple first-in-first-out and ensure that they understand and see the sense behind this approach.

The first-in-first-out (FIFO)

In this system, the first customer to arrive at the queue is served first and exits to give way to the next person. The first-in-first-out is seen as the most fair because customers are served in the order in which they came (chronologically) and the customers are usually aware of their positions and

those ahead or behind them. Where customers would like to be treated equally, this model is most suited (Nasar, 1988).

Single Queue

The single queue is the common snake or corral queue format. Each person waiting is served in turn and the format of the queue discourages pushing in. Customers are also comforted by the idea that they will be served fairly and that they can also see the queue moving forward.

Multiple Queues

In this form multiple servers serve multiple lines or single lines of customers. In multiple queues several customers can be channeled into one single line be served or alternatively, customers can select the line which to wait and the server from which to be served (Hall, 1991).

Head of Queue

According to Larson (1987), the Head of Queue is the system where the next person to be served waits in a single queue area. It is important that customers can see along the line of service position to avoid challenges caused by slack in service provision. Where there are more than 5 servers, this factor becomes increasingly challenging.

2. 1. 4. Queue Discipline

The queue discipline indicates the order in which members of the queue are selected for service. It is most frequently assumed that the customers are served on first come first serve basis. This is commonly referred to as FIFO (first in, first out) system. Occasionally a certain group of customers receive priority over the others even if they arrive late. This is commonly referred to

as Priority Queue. The queue discipline does not always take into account the order of arrival. The server chooses one of the customers to offer service at random. Such a service is known as Service in Random Order (SIRO). While allotting an item with high demand and limited supply such as a test match ticket or share of a public limited company, SIRO system is the only way of offering service when it is not possible to identify the order of arrival. It may also be so that every customer gets a time slice. If her service is not completed within that slice of time, he/she will re-enter the queue. This is known as Round Robin. State of Queuing System The transient state of a Queuing System is the state where the probability of the number of customers in the system depends upon time. The steady state of a Queuing System is the state where the probability of the number of customers in the system is independent of time.

In order to maintain fairness in the queuing system, there are rules that have been set for the Queue Management System concept:-

Queue must be seen to be fair.

Queues must be managed efficiently and not allowed to deteriorate into a ' free-for all'.

The process must take into consideration positive feedback of progress of the movement of the line.

The process must be clearly defined; the beginning and the end points must be clearly seen.

Perception of queuing time should be effective and to the advantage of both the firm and the customer.

2. 1. 5. Psychology of Queues

There are a few key behavioral responses or reactions to queues, or waiting. Below are the propositions of Maister, (1985):

Time that is not filled with occupation seems longer than occupied time.

When customers are anxious, the waits seem longer.

When waiting time is not known, they seem longer than known, finite ones.

Where there is no fair play in the wait, they are longer than equitable waits.

When the service being sought is perceived as more valuable, the customer is will to wait longer.

When a customer is waiting on his own without company, the wait will be perceived as longer than in a group.

2. 2 Characteristics of Queuing Processes:

According to Deiner (1993) the queuing theory considers mainly 6 general characteristics of any queuing processes: Four of these will be discussed here; Arrival pattern of customers, Service pattern, System capacity and Queue disciplines.

Arrival Pattern of Customers

Customer arrivals are the input processes of a queuing system that is defined by:

Number of customers arriving at a given time period.

Statistical distribution of customers' inter-arrival times: This is the pattern that infers the times between customer arrivals: The Poisson process is the most commonly used in the distribution of the time until the next customer arrival is independent of the last arrival.

Other unusual aspects: Stationary (time-dependent) vs. non-stationary arrival patterns (non time-dependent).

(2) Service Pattern

The service pattern is the output aspect that and relies on the following:

The distribution of the service time: This refers to the time that has been spent during the service process from the beginning of service to the end, and the assumption being that it is not influenced by the number of customers present. The exponential distribution is most commonly used in this aspect.

Number of servers: This number influences the arrangement of a waiting line, which could be either in parallel or in series.

(3) Service Capacity

The service capacity is the available space for queuing is divided into the following classifications:

Finite service capacity: When a customer realizes that the waiting area is full when he arrives, and he is not able to join the queue, he reneges (leaves) and never returns.

Infinite: When the maximum allowable number of customers of the waiting area is evidently large, it is presumed to be infinite. This is a standardized assumption for most queuing systems.

(4) Queuing Disciplines

Order of Service

This specifies the order in which customers were chosen for service within a queue and represents the way the queue is organised (rules of inserting and removing customers to/from the queue). There are these ways which have been discussed in detail in 2. 1. 4

FCFS: First Come, First Served. This is the most commonly used discipline applied in the real world situations, such as check-in counters at the airport

LCFS: Last Come, First Served. This illustrates a reverse order service given to customer versus their arrival.

SIRO: Service in Random Order.

PD: Priority Discipline. Under this discipline, customers will be classified into categories of different priorities.

Structure of the Queue

According to Maister (1985) queues can be divided in to two categories:

Unstructured Queues

In the unstructured queues, customers make waiting lines unpredictable and different places and directions. These type of waiting lines are not usually

easy to manage and can be chaotic if the management concerned do not intervene at the right time. Such type of queues can be seen in ATMs and taxi queues.

Structured Queues

In this form, waiting lines have customers in fixed, predictable positions. These positions can either be defined by the arrival time of the customers as in first come, first serve and wait in line; or use of tickets with numbers and customers sit while they wait for their turns to be called; or making appointments through telephones and text messages. It combines two main factors: the number of servers and lines available. Among the disciplines under this category:

Single line and single server

This is one of the most common forms of waiting lines, which customers join a queue from the end depending on his arrival and be served by one server. This form of system encourages a high level of social justice and at the same time it allows social comparisons where queues are visible to the customers, such as queuing at a bus stop.

Single line and multiple servers

In this form, customers join line up in a single queue as they arrive and multiple servers available where each of the customers is served by the next available server among all. Examples of such forms of lines can be found in airline check-in areas, or some fast-food restaurants, where customers follow a single line waiting to check in luggage or purchase food respectively, and the customer at the front of the line is called by the next available server.

Multiple line and single server

This form involves only one server attending to multiple waiting lines. An example to this would be car washing service, where only one washing area is available for cars lining in two queues.

Multiple lines and multiple servers

This is the form that has multiple servers and several waiting lines, where each line is served by a one server. This situation can commonly be found in Supermarket queues or Immigration Custom queues, where customers arrive at the queuing area, join in one of the waiting lines and wait until they move towards the front of their chosen line.

2. 3 Queues as Marketing Issue

In the context of operations research, the queuing theory is an effective tool to understand queues in a quantitative way. Though it does not work perfectly with queues that do not involve human beings, the reason for this is because it does not take human factors into account. In marketing, " selling a service is selling a relationship", the success of service delivery is determined not only by pure quantitative measures such as service efficiencies, but more importantly, by customers' satisfaction, which directly reflects the relationship between the company and the customers. As a result, customers' satisfaction has become an increasingly important issue to service operation management. Proper influencing of customers' perception of service quality would serve to contribute greatly in maximizing customer satisfaction, and this could not be done without thoroughly appreciating the psychology of queues. The previous sections describe how queues are

studied in a general context, and the following sections will cover queue in the psychosocial context.

2. 3. 1. Behavioral Aspects of Queuing

For a long time, many researchers have studied the behavioral issues of waiting. Four important issues are raised from the existing literature, and they are social justice, social comparisons, perceptions of waiting, sense of progress, and their effect on customers' decisions making to queuing and overall satisfaction of the service of a firm.

Social Justice

Earlier research has suggested that queues form a social system. Hall (1991) revealed that a waiting line is made up of a social system with norms, roles, and obligations for queuers. Later, Larson related waiting to social injustice. He defined social injustice as the violation of the principle of first in and first out. Queuers tend to think negatively of those who arrive later than they do but get served before them because they have sensed an unfair treatment. Both studies by Zhou and Soman (2003) suggested that when social injustice occurs in waiting, after the intrusion point the queuers are more likely to respond negatively to the cutting behavior than queuers ahead of the cutting point. And a customer's perception of a wait to be unfair is likely to lead to dissatisfaction with the wait and possibly with the entire service experience.

Social Comparison

Zhou and Soman also suggested an important social factor that customers in a waiting line tend to make social comparisons of relative waiting times with others in the queue.

These comparisons could be upward and downward, meaning comparisons with people ahead and behind respectively. From their study, it was shown that customer satisfaction while waiting and their decisions of leaving the queue are more likely to be influenced by downward comparisons. A queuer will feel more comfortable if he sees more people behind and is more likely to stay in line. However, it has been shown in their other research that social comparisons would not be an influential issue unless when the first-in-first-out principle is being violated.

Perception of Time

According to Maister (1985) customer perceive service satisfaction rather than measure actual service waiting time. For example, Katz, Larson and Larson (1991) researched perceptions of waiting under different situations in a bank. The results of the study deduced that clients tend to exaggerate the amount of waiting time by an average of 1 minute. This study posits that customer satisfaction seems to reduce as perceptions of waiting time increases. The total cost of receiving the services that customers perceived also affect their perception of waiting time. According to respondents, the total cost involved not only the cost of services, but also the cost of waiting which theoretically increases linearly with waiting time. The study by Baker and Cameron, (1996) indicate that waiting time acceptability determines a customer's mood and satisfaction of the service quality. The authors also

suggested that when people were not engaged, the waiting time would be perceived as long.

Sense of Progress

It is the feeling that the effort that one is putting into an activity is being achieved to some degree. Zhou and Soman (2003) suggested that queuers' sense of progress of the wait is also influential towards enhancing customer satisfaction and acceptability of wait. The sense of advancement is initiated by different signals that tell the speed of the line, such as customer's physical displacement over time, their distance from the goal, changes in expectations about how much longer they need to Waiting Time.

Open-ended waiting: In open-ended waiting, we expect steady increase of subjective distress.

Hence when customers have no clear information about the remaining waiting period they become increasingly frustrated over time. This idea is consistent with Osuna's (1985) analytical model of waiting which suggests that the aversiveness of waiting increases over time, as a result of what he names the psychological cost of waiting. (Ziv Carmon and Daniel Kahneman, 1995),

2. 3. 2 Customers' Waiting Time versus Cost

Cost of Waiting is the value of wait in relation to customers reaching their goal. It is the cost associated with customers having to wait for a service. The progress of line advancement can be indicated in the information of expected waiting time given by service providers. Scientific researchers have

shown that the shape of waiting line, its visibility and speed of advancement influence the client's perception of the waiting time.

In addition, Osuna, (1985) demonstrated that the amount of stress rises as waiting time increases during the process causing an rise of the psychological cost of waiting and a decreasing customer satisfaction. A customer's stress level rises when faced by unclear expectations because most people seem to have problems coping with them. When waiting period is not provided, customers may have wrong perception of time. Also, if customers know they are getting closer to achieving their objectives, their stress level declines. Dubé, Leclerc, and Schmitt (1991b) have come up with a theoretical framework using the field theory to explain customers' affective response to waiting. The field theory suggested that stress is relieved as long as the customer is inside the area of the goal, and that anxiety remains higher when he is off the area. The degree to which of the above behavioral aspects influences a queuing system relies on the type of queuing system in operation. In studying the queuing discipline and the number of ways in which each discipline can be applied will help to determine the nature and function of a queuing setup.

2. 3. 3. Comparison of queue disciplines

The comparison of queue disciplines was made on the following categories as suggested by Hall, (1991):

Decision Making

When customers enter an area to be served, they will have to make two main decisions.

The first is whether or not to join a queue, and whether or not to continue waiting. In single waiting line, the customer can make his in a simpler context because they simply have to decide on the above two basic decisions. However, in a system of multiple lines, decision making becomes a challenge because it also involves other aspects such as social justice and comparisons.

Social Justice

As discussed earlier, social justice is high in any single-line queuing system because everyone in the line is guaranteed to be served in the order in which they arrived.

Sometimes, there might be the likelihood for injustice but at a higher level. A case scenario would be where three customers arrive at three varied times, though the servicing time for the three customers are the same, the customer who came in last might feel unfairly treated because he will have to wait longer than the first two. On the other hand, a system with multiple lines seems to encourage lower social justice. Depending on the luck of the customer, he might choose a waiting line in which is served by a faster server. This would in turn encourage customers from the other lines to switch to the faster line. In the case where switching line is not possible and the speed of each waiting line is visible, customers in slower queues are more likely to feel frustrated.

Comparisons

In single waiting line, customers can only make upward or downward comparisons, whereas, in the case of multiple lines, customers can also make lateral comparisons. For example, when two customers are in a system of multiple queues, they first started waiting side by side, but as the line moves forward, they keep track with each other's progress in the line to see which line is moving faster. This lateral comparison is served to help customers to decide on joining the better waiting line.

2. 4 ADVANTAGES OF QUEUE MANAGEMENT SYSTEM

The Queue Management System provides many advantages to the customer service provider and the customer himself. The advantages are:-

Staffs dealing with customers who have been well-treated do not get stressed and this increases job satisfaction.

Staff can serve more customers per hour if those customers are fed to each counter efficiently.

It can help the management by producing statistical reports on information such as arrival rates and patterns, waiting and service times, and default and reneging cases.

Enabling managers to monitor and set performance thresholds.

Customer will be treated with fairness and in more relaxed environment.

The queue flow will be smoother and thereby increasing the efficiency of queue management.