Simulation and Stocahstic Models - Group 2 Project

A Study of Server Capacity at Cafe Louis

Jacob French - Alif Nurrokhman - Jayden Hooper - Reece Pacia - James Boyle

Introduction

Members

Names	Emails	ORCID
Jacob French	frenchjaco@myvuw.ac.nz	0009-0004-7268-3581
Jayden Hooper	hooperjayd@myvuw.ac.nz	0009-0004-2644-5216
Reece Pacia	paciareec@myvuw.ac.nz	0009-0001-4748-604X
Alif Nurrokhman	nurrokmuha@myvuw.ac.nz	0009-0008-7814-9231
James Boyle	boylejame1@myvuw.ac.nz	0009-0002-1177-2922

Terminology

- Arrival Rate - λ
- Service Rate - μ
- Steady-State Attainable if - $\rho = \frac{\lambda}{\mu} < 1$
- Number of Servers - c
- Expected Waiting Time in System - W

Introduction

Queuing is traditionally defined as when people, materials or information must wait at certain times to receive a service (Teknomo, K 2012). This study will explain the basics of queuing theory and its relation to organisational performance and customer satisfaction in the hospitality industry. Queuing theory is frequently used to assess customer waiting time, server utilisation, earnings and costs. In hospitality, companies service's often cannot keep up with their arrivals, which results in long waiting times and loss of potential profits. In this study, we will study Cafe Louis's queuing system and prepare theoretical and simulation models that could be used to predict what effects changes in queue setup, such as changes in load, would have.

Problem Statement

Businesses have poor organisational performance and customer satisfaction due to service not keeping up with arrivals. If queuing theory was applied this could be maximised.

Objectives

- Examine whether a change in queue setup will increase organisational performance, to be measured using profit.
- Examine whether a change in queue setup will increase customer satisfaction, to be measured using waiting time.

Hypothesis

Based on the objectives of our study, the following hypothesis will be tested.

- Will increasing server capacity increase customer satisfaction and organisational performance?

Expectations

- Steady-State distribution is not attainable under current conditions. This will cause waiting time to tend to infinity.
- Organisation performance is not being maximised as service cannot keep up with arrivals, resulting in a loss of sales.
- Customer Satisfaction is not being maximised as service cannot keep up with arrivals, which results in long waiting times.
- Increasing Server capacity will fix all of the above.

Literature Review

For this study, we followed a handful of similar research papers done in this field in which they observed real-world systems and the effect that waiting time has on customer satisfaction. Queuing becomes an inconvenience and can be highly costly for organisations. Problematic queuing systems can significantly harm the customer's overall satisfaction (Ronald Anthony Nosek, Jr., & James P. Wilson, 2001).

A highly satisfied customer will likely provide repeat business, increasing revenues and profitability (Ronald Anthony Nosek, Jr., & James P. Wilson, 2001). Therefore we wanted the focus of this study to determine what factor could be optimised for the overall performance of an organisation to increase customer satisfaction. There are two approaches to improving customer satisfaction with waiting time: through decreasing actual waiting time as well as through enhancing customer waiting experience (Katz, Larson, and Larson, 1991; Davis and Heineke, 1994)

Because we cannot influence the waiting experience, a key metric we will use in this study is overall waiting time, as the amount of time customers must spend waiting can significantly affect their satisfaction (Davis and Vollman, 1990). Ronald Anthony Nosek, Jr., & James P. Wilson (2001) also considers that the measurement of customer satisfaction – is highly qualitative and subjective, and the relationship is generally inverse (i.e., in general, as waiting time decreases, satisfaction increases).

Teknomo, K (2012) states that a queuing system has two primary components: the customers and servers. Teknomo also says in their study on construction management that there is a direct relationship between the server and the customer. They stated that a system with fewer servers could incur customer costs (satisfaction), and adding servers to the system can incur higher server costs (organisation). As we think in terms of [the entire] system, the total cost of [the] queuing system must include both the server side and the customer side (Teknomo, K, 2012). Therefore, the optimum number of servers is the one that minimises the total cost (Teknomo, K, 2012).

For the system we observed for this study, we state that the time it takes for a server to complete a request strongly influences the overall waiting time of a customer. We can optimise waiting time by optimising the server side of the system (hence our hypothesis). For an organisation to perform optimally, customer sanctification and operational costs are maximised. Because queuing theory is rich in optimisation (Teknomo K, 2012), we will utilise the tools that queuing theory provides to attain precise performance measures to assess real-world systems.

Methodology

Data Collection

Before collecting data, we obtained permission to do so via email. Once permission was granted, we began compiling our data between 11:00-14:00 on weekdays, as this window was peak busy hours for the cafe, and we could collect data much faster. We collected data on the system's metric arrival times, service times and waiting times. We ensured that our data-collecting methods were not invasive, including not collecting personal information or disrupting the queue itself.

The tools we decided to use for our data collection were:

- Notebooks, pens, and laptops for recording and capturing data (one per group member)
- Online stopwatch to time multiple people at once
- Google Sheets document with the data set, to be later converted into CSV as a machine-readable format

Data Analyse Methods

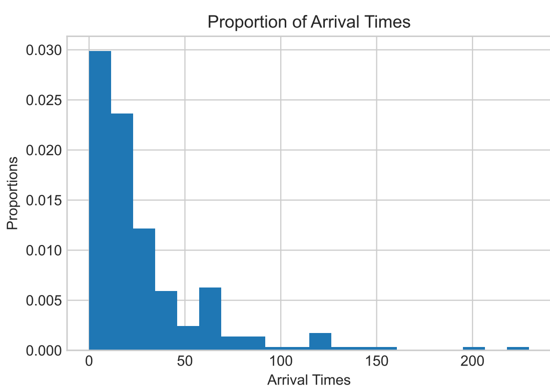
For analysing the data, we used Python for exploratory analysis and utilised the 'disfit' package to find and plot the "best-fit" distribution for our data. Additionally, we used a Python function provided by Dr Binh to perform chi-square goodness of fit tests. Other Python packages we used were pandas, numpy, matplotlib and scipy.

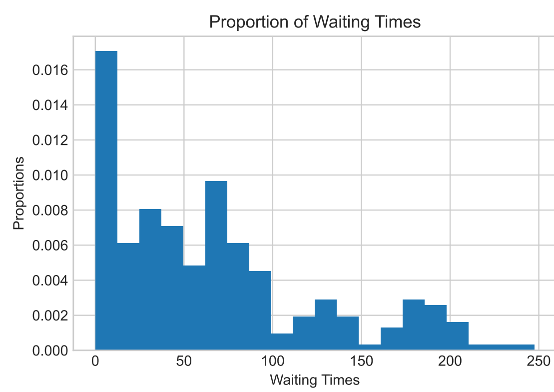
Reproducible Repository

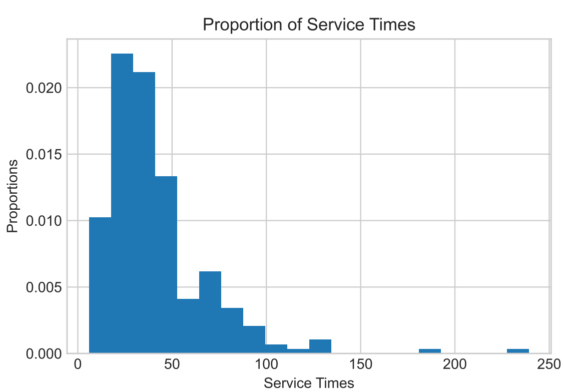
We set up a reproducible repository described by Frery, Gomez & Medeiros (2020) on GitHub so that this study may be reproduced. 312-Group-2-Project

Observed Data

Arrival, Service and Waiting Times







Exploratory Statistics

Table 2: Observed Data

Metric	Inter_ArrivalTime	WaitingTime	ServiceTime
Count	251	251	251
mean	27.67	66.52	41.3
std	33.07	59.55	28.1
min	0	0	5.97
25%	6.52	18.27	23.73
50%	17.37	56.61	33.5
75%	34.33	90.4	49.35
max	229.5	247.6	239.3

Table 3: Expected Metrics based off Observed Data

Metric	Mean.s.	Rates.per.Hour
Expected Inter-Arrival Times	27.7	130
Expected Service Time	41.3	87
Waiting Time	66.5	–

Based on this table, we find that the rate of customers arriving is more significant than that of customers being served such that $\lambda > \mu$, therefore, steady-state is unattainable for the current system.

Models

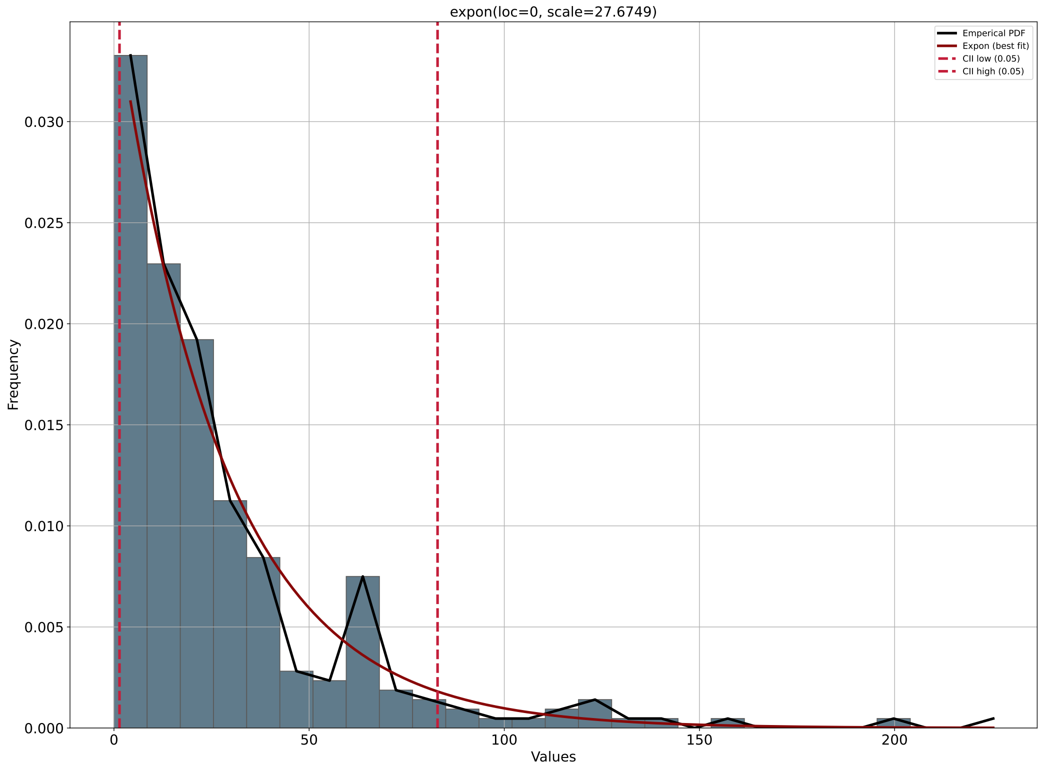
We built three models to represent our system. - Model 1 uses exponential distribution for arrival times and service times - Model 2 uses exponential for arrivals and Erlang for service - Model 3 uses exponential distribution for arrival times and uses empirical distribution for service times

Model 2

Best Fit Distributions

Inter-Arrival Times

Using the distfit package, we determined that the most suitable probability distribution for modelling inter-arrival times is the Exponential distribution, compared to the Erlang and Gamma distributions.



Chi-Square GOF

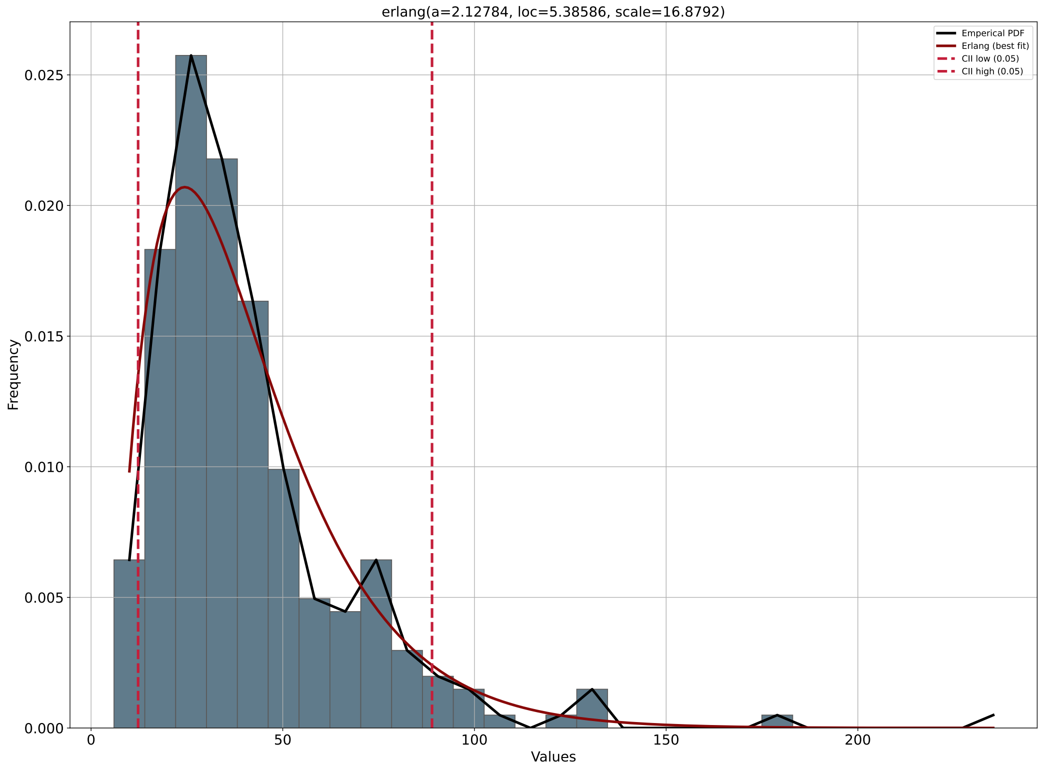
$$H_0 : \text{The sample is exponentially distributed}$$

$$H_1 : \text{The sample is not exponentially distributed}$$

However, we conducted a chi-squared goodness-of-fit test following the null hypothesis that the sample data follows an Exponential distribution and the alternative hypothesis that it does not. The test resulted in a chi-square statistic of 64.836 with a corresponding p-value of 2.184e-05. These results suggest insufficient evidence to support the null hypothesis that the sample is exponentially distributed. Therefore, we cannot conclude that the inter-arrival times are Exponentially distributed. Further investigation may be necessary to identify a more appropriate distribution for modelling the inter-arrival times.

Service Times

Using the distfit package, we identified the most suitable probability distribution for modelling service time as the Erlang distribution after comparing it to the Exponential and Gamma distributions.



Erlang.

$$H_0 : \text{The sample is Erlang distributed}$$

$$H_1 : \text{The sample is not Erlang distributed.}$$

We then conducted a chi-squared goodness-of-fit test with the null hypothesis that the sample data follows an Erlang distribution and the alternative hypothesis that it does not. The resulting chi-square statistic was 26.492 with a corresponding p-value of 0.546. Therefore we conclude that there is sufficient evidence to accept the null hypothesis and conclude that the service times are Erlang distributed. This implies that the Erlang distribution provides a good fit for modelling the service times in our study.

Model Results

Simulating the models for three hours produced the following results:

Table 4: Model Results

Model	Mean.Waiting.Time.s.
Model 1:M/M/1	1872.87
Model 2:Best-fit Distribution	1275.37
Model 3:Empirical Distribution	1704.03

Because a steady state is not attainable, all these results will approach infinite. Once we add more server capacity, steady-state should be achievable, and we expect that model 2 will be the best representation of the system.

The reason why we would not expect model 1 to be the best representation is possibly underfitting, as the distribution used for this model may not be the best representation.

We would not expect model 3 to be the representation because the model may be overfitting when using the empirical distribution.

Model 2 produces the shortest waiting time, and because it is fitted with the 'best fit' distribution, it should be the most accurate representation of our model. We will continue to analyse Model 2.

Cost Analysis

Assumptions

For this section, we make the following assumptions:

- Servers are paid a minimum wage of $22.70 per hour.
- Server cost per hour = c * 22.70
- The average spend = $5 (estimated as the item's average cost on men. u)Earnings per hour = average spend * service rate per hour. Whilst serving rate < arrival rate. Otherwise, service rate per hour = arrival rate.
- Earnings lost per hour = (service rate per hour - arrival rate per hour) * average spend, so does this. Whilst Arrival rate >= Service rate. Otherwise, earnings lost per hour = 0.
- Profit = Earnings per hour - server cost per hour -Assume all other costs are covered.

Results

Changing server capacity produces the following results:

Table 5: Change in Capacity Results (continued below)

Server.Capacity..c.	Expected.Waiting.Time	Arrival.Rate	Service.Rate
1	NA	130	87
2	77.162	130	130
3	45.96	130	130
4	42.22	130	130

Table 6: Table continues below

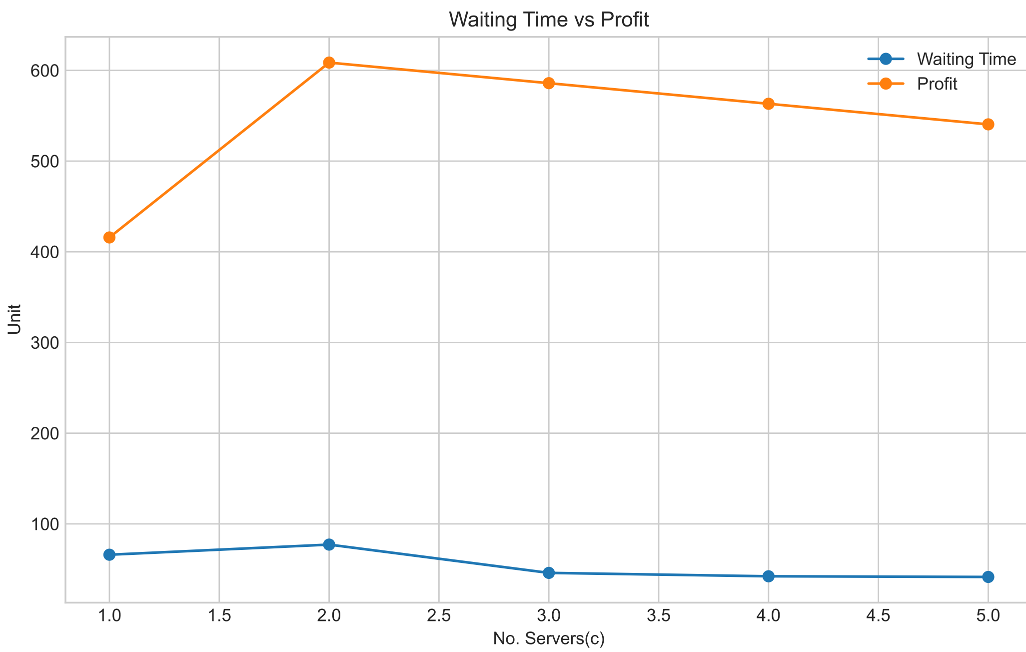
Utilisation.Factor	Earnings.per.hour...NZD.	Earnings.Lost.per.hour...NZD.
1.49	435.81	218.48
0.86	654.3	0
0.79	654.3	0
0.78	654.3	0

Profit.per.hour...NZD.
413.11
608.9
586.2

Profit.per.hour...NZD.
564.3

Results indicate that profit is maximised when the system has two servers; if the system has two or more servers, gain follows a linear trend 654.3 -wages * c. Results also indicate that waiting time shows no significant decrease after the system has three servers. This can be observed graphically.

Profit & Waiting Time Plot



Profit Waiting Time Trade-off

As the optimal number of metrics servers disagree, we have a trade-off. Hence we suggest a new metric profit-waiting time to find the number of servers that balances profit and waiting time. This can be observed graphically.

The graph shows that profit - waiting time will be maximised if the system has three servers.

Conclusion & Discussion

Recommendations

In conclusion, our models indicate Cafe Louis does the following between 11-2 pm.

- If Cafe Louis values organisation performance over customer satisfaction, they should hire one additional staff member as this will maximise their profit.
- If Cafe Louis values customer satisfaction over organisational performance, they should hire two or more staff members, as this will decrease their waiting times significantly.
- If Cafe Louis wants to find a balance between the two, they should hire two staff members as this maximises the difference between profit and waiting times.

Discussion

We believe the waiting times calculated by our models were inaccurate. This is because Cafe Louis had busier periods between lectures(roughly). During these busier periods, the majority of our data points were collected. We believe this directly affected our waiting times and arrival rates; hence, we didn't have constant arrivals (an exponential distribution assumption). We believe the simulation models we built better represent this period rather than 11-2 pm. Suppose data was only observed and simulated for this period. In that case, we believe it would fix problems such as small p-values for the chi-square goodness of fit tests for exponentially distributed arrival rates, and we would see an increase in waiting time when adding additional servers. If a study was carried out on this, we believe it would produce significantly better results.

Evidence

Simulating our models for ten minutes rather than three hours results in much closer waiting time estimates to our observed mean. This is evidence that our models represent this period better.

Table 8: Model Results (continued below)

Model	Mean_Waiting_Time.s._3hours
Model 1:M/M/1	1872.87
Model 2:Best-fit Distribution	1275.37
Model 3:Empirical Distribution	1704.03
Observed	66

Mean_Waiting_Time.s._10mins
186.3
11.75
205.72
66

Conclusion

In conclusion, increasing server capacity will improve customer satisfaction and organisational performance.

References

- Aina, M. O. Y. P. N. B., & Kazeem, O. Queuing Theory and Customer Satisfaction: A Review of Performance, Trends and Application in Banking Practice (A Study of First Bank Plc Gwagwalada, Abuja Branch).
- Davis, M.M. and Heineke, J. (1994), "Understanding the Roles of the Customer and the Operation for Better Queue Management", *International Journal of Operations & Production Management*, Vol. 14 No. 5, pp. 21-34. https://doi.org/10.1108/01443579410056777
- Frery, A. C., Gomez, L., & Medeiros, A. C. (2020). A badging system for reproducibility and replicability in remote sensing research. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, 4988-4995.
- Katk. K, Larson. K, Larson. R. (1991). Prescription for Waiting Line Blues: Entertained, Enlighten and Engage, Sloan Management Review, https://www.researchgate.net/publication/304582002_Prescription_for_the_Waiting_in_Line_Blues_Entertain_Enlighten_Engage
- Teknomo, K. (2012). Queuing rule of thumb based on M/M/s queuing theory with applications in construction management. Civil Engineering Dimension, 14(3), 139-146.
- Ronald Anthony Nosek, Jr., & James P. Wilson. (2001). Queuing Theory and Customer Satisfaction: A Review of Terminology, Trends, and Applications to Pharmacy Practice.
- https://www.math.tamu.edu/~phoward/m442/queue.pdf