Group 1

Individual Report

Tian Bai

300343541

1. the brief description of the system being studied, and what was our team done.

Firstly, we decided to simulate a M/M/1 system, because a single server which will be easily for collect data. For reach M/M/1 system simulation we need collect arrival time and service time for this single server and it should have a range of service time to avoid to collect data hardly and also have a not huge waiting buffer. During our consideration and learn from other previous project then we decide to collect data at Tommy Million Pizza, which do not like Maki Mono have a quick service time and a huge waiting line.

Secondly, when we get rough data then we read and trim it into two sets which contain interarrival time and service time and we do the analysis on these data for find which distribution is the most fitting these data.

Finally, we build three simulation models (M/M/1, M/G/1 and empirical) and using the data which we get in the second part for simulate on three models and conclude from the results.

2. A description of my activities in the investigation.

I take part in the whole part of this group project. For the first part collect data, I

take two hours on collecting data and I had collected 76/277 data. For the second part I write two scripts one for load and trim our data, another for analysis our data. For the third part that I build three simulation models (M/M/1, M/G/1 and empirical) run on the result that I get from second part.

3. What did I like/dislike during my work on the project?

On the one hand, I dislike collecting data, which is the most boring work in this project. However, it is the most important job in the whole project. On the other hand, I like write script whatever which type of script. Because of that I like programming on python, and I like using pithing to handle mathematical problems.

4. What difficulties did I have to overcome?

During my project, I have handle many problems in every stage.

When I was collecting data, there is one problem which will always happen in the real-world cases, however, we have not cover it during our meeting. They are when come to Pizza and they need decide what to order and during this period might new people come in and cut in the waiting line or they cancel to order. This case influence on our collect sheet, because when people cancel to order then we

cannot record him. For handle this problem is not difficulty that is we just need a service time which is made by end service time minus start service time. So, it does not matter when the service start.

When I load our data, it also has one problem and that is the data that my script read which is type of string (e.g. 12:45:06) and we just want a numerical type of time which might be float or integer. Because of this case, convert string to numerical type seems like a challenge. However, python support datetime type which helps me handle this problem by method strftime() which need I set a format of our data and use it for parsing. The Datetime type will also have methods that help get interval of two time and return by second or other units.

When I analyze which is the most fitted distribution for our interarrival time and service time. There is one problem that is for different distribution that have different encapsulated function which will return different fitted value of given data, However, the Scipy package do not have a clear document for illustrate return values. For example, for gamma distribution scipy.stats.gamma.fit(data, floc) will return alpha,loc and beta, compare with scipy.stat.expon.fit(data,floc) will return loc and lambda.

Furthermore, For the degree of freedom, first I algorithm that is df = k – q – 1 to find it, where k is number of our bins and q is the number of estimated parameters.

The p-value of some distributions test under this type of degree of freedom always get a wrong answer which is shown as 'nan'. The reason of this is the value is too small to show. Using adjusted degree freedom which is equal to q (number of estimated parameters) could handle this problem.

For the simulation model. I handle one problem that is when I realize that in Simpy it using minute for the time unit, However, I using second for time unit of my result of my load data analyzing part. I re-run the analyzing by using minutes as unit of my result then it works on my models.

5. What would I do differently if I were to start the project anew?

If I start the project anew, then I want to do two new things. First new thing is I want to collect a whole day's data rather than just collect data in an interval of a time during a day, the reason of this is I want to simulate a business more "truly". If the data could get more then the more accurate for its data analysis. Second is for the simulation model that I want to make it more complexity, I want to create a M/G/2 simulation model or a queuing network, because I want to treat menu as another server, when people go into this system and he first will look at menu it will takes an exponential service time then he could then he might just go away or continue to next server for service.

6. Any personal observations/thoughts on the setup and usefulness of the project.

A group should have a thoughtful plan. For example, for my group that we forget to talk about criteria that should be regard as end service. I using when a person starts ordering and when a person get his pizza then treat as he end of service. However, my groupmate treats that people finish their payment as end of service. It makes our data have different distribution so that we cannot using these data together.

In these project, that I think it is not difficult and not too many works, anyone could do the whole works by himself. I also think that we can use other tools such as R and SAS for cross validation.