**The Effect of Class Schedule on Traffic in the Hillsdale College Dining Hall:  
Reduction of High Density Arrivals Through Monte Carlo Simulation**

Philip Andrews, Michael Erickson, Mark Moreno, Sashabaw Neidbalski

MTH 380-02

Dr. Mark Panaggio

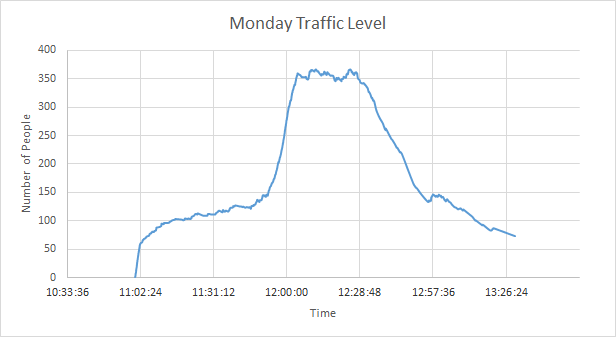
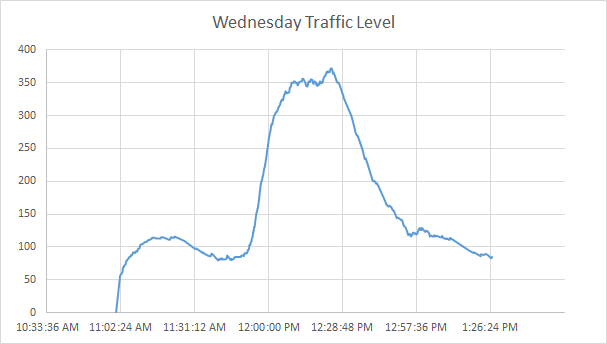
7 March, 2020

**Introduction**

Our group noticed large spikes in the number of students in the dining hall at Hillsdale that seem to correlate with when classes are released, bottlenecking the entrance. We wanted to show how the class schedule then affected the traffic flow in the dining hall during lunch hours: 11:00 - 13:30 Monday through Friday (we used a 24 hour system since it is easier to work with in Excel). Using this information, we wanted to simulate how a change in class schedule would reduce the traffic in the dining hall and attempt to minimize it. We contacted Bon Appetit's controller, Joann Alverez, who provided us with the arrival time of every student entry to the dining hall, accurate to the second, for the entire week of February 3rd, 2020. We used the Spring 2020 class schedule and Webadvisor to record exactly how many students are in class for every 5 minute interval from 10:00 - 13:30. With the data from Webadvisor and Bon Appetit, we formulated a model that simulates the number of students entering the dining hall every five minutes based on the class schedule, and then predicted how a change in schedule would reduce periods of heavy traffic.

**Model Formulation**

The data provided to us included the timestamp of each swipe into the dining hall for a full week’s worth of lunches. We started out by examining our data closely and visualizing it to get a better idea of what assumptions might need to be made and what our primary focus should be. For the sake understanding the flow of people in and out of the dining hall we plotted the number of people inside at any given moment by making the reasonable assumption that the time people stayed at lunch comes from a normal distribution with an average of about thirty minutes and a standard deviation of about ten minutes, examples of which can be seen below.



From these plots, we learned that Tuesday and Thursday did not have as significant issues with congestion, and therefore we decided to focus our efforts on M-W-F since these days have very similar class time slots. As expected, the largest congestion occurs between noon and one o’clock and the rest of this plot seemed to match up quite well with our real-world intuition. However, when formulating the model we wanted to avoid relying on assumptions about how long people stay at lunch and thus decided to focus purely on minimizing large numbers of arrivals. In general, the situation is tricky to model because of the dependency of arrivals on human habits, social interaction, and other intangible factors. Because of this, we realized that this problem would be suited to a Monte Carlo simulation based on a few key quantifiable dependencies.

The Poisson process is a fundamental counting process used to model arrivals of rare events. Common examples of applying this distribution include the rate of incoming emails, defects in a manufacturing process, or radioactive decay. It seems appropriate then that we begin our exploration of dining hall arrivals with the Poisson process. More specifically, the simulation should take a random sample from a Poisson with a mean arrival time parameter that is based on the dependencies we have identified. Using a random number generator inside an inverse poisson function can produce these samples. In Poisson\_Inv(x, λ), we generate a random number for x and the rest of the simulation comes from estimating the mean rate of arrivals λ based on chosen parameters.

The expected value parameter λ is fundamentally based on the parameters of the number of students who are in class in each time slot. Ultimately, these are the parameters we are interested in modifying, but for the sake of training our model, the actual values were gathered from WebAdvisor for each day of the week (see Example Monday Schedule).

When considering the reasons why someone would go to lunch, two primary effects are proposed to capture the largest influence on on the average arrival time λ:

1. a stable background rate of people who may not have other obligations and decide at a time whenever they are hungry to go eat
2. a larger more volatile rate of those who just got out of class who go directly to lunch because it would make less sense to wait ten minutes and then go to lunch

Thinking in terms of the class schedule is very difficult for productive modeling because class times overlap and there is no continuous pattern. Instead, we introduce the variables “lunchables” and “just released” which we propose are much more closely connected to modeling the situation but ultimately are based on our true variables, the class schedule.

**Lunchables (L):** the total number of people who are available to eat lunch at a given time

i.e. the total student body size minus the total number of people in class minus the number of people who have already eaten

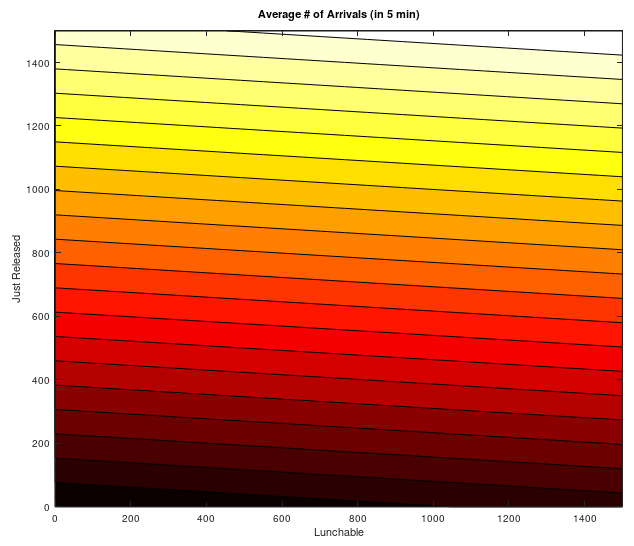
**Just Released (JR):** the total number of people who were just released from class at a given time

In the calculation of Lunchables, we assumed that the student body size is 1500. We were not able to record how many students are in classes that are scheduled only between the students and professors such as music lessons and very small classes, but their effect was assumed to be marginal. Since we would be working with a stochastic situation, we decided to bin the number of arrivals in 5 minute increments. Likewise, Lunchables and Just Released are calculated every five minutes which naturally makes sense because the number of students who are in class only changes on a resolution of five minutes.

Proposed model: λ(L,JR) = B\*L+J\*JR

B = base attendance rate (average % of people who are available who go to lunch within that 5 min bin)

J = just released attendance rate (average % of people who are just released from class who go to lunch within that 5 min bin)

Both of these parameters can easily be estimated from the given data for a particular day. On Monday, for example, B = 2.4% and J = 30%. A contour plot of this simple linear relationship the variables have on the average arrivals parameter λ can be seen below.

From this plot we see that decreasing the number of lunchable people and just released people will reduce λ at a given time. Of these two variables, reducing just released people will have a greater effect. To prevent jams of people, our solution should try to keep λ as small as possible which means reducing times of high just released people and lunchable people. This model proved to be a good starting point but students can not immediately make it to lunch when they are released from class so the model was updated to account for this wider spread across bins that just released students have.

Update 1 model: λ(L,JR)=(B)(L)+J/3[W1\*JR(-10)+JR(-5)+W2\*JR(0)]

In this model the arrivals in a given bin are now dependent on a weighted average of the number of students released 10 mins, 5 mins, and 0 mins before that 5 minute long bin. That is, for every time bin, our model prediction now takes into account the past 3 time bins to predict how many people would enter the dining hall. This now accounts for the fact that students take some time to move from class to lunch and the contribution of each five minute interval in the past can be described by W1 and W2.

W1 = weight of the contribution of students who were released 10 mins ago

W2 = weight of the contribution of students who were released 0 mins ago

The value of these weights is not immediately obvious, and is equivalent to asking how long most people take to move from class to lunch. We used the data from Monday 2/3/20 as a type of training data to make sure these weights seemed reasonable then used the rest of the days in the week as testing data. We assumed that students who are just released would have the largest contribution to the time bin 5 minutes ahead, followed by the bin 10 minutes ahead, and the least contribution would be to the immediate bin. The weights that seemed most reasonable were W1=0.7 and W2=0.4. These weights generalized well to Wednesday and Friday but could be subject to slight changes given more study. At this point examining a plot of the residuals indicated that there were no obvious trends, indicating that no major effects were being overlooked in this model.

The last effect our group noticed was that arrivals tended to be slightly higher during the middle of the lunch period likely because people tend to want to eat lunch with their friends. To capture this we decided to scale all of the predicted arrival models by a quadratic function F that simply gives slightly more weight to bins in the middle of the lunch period.

F = ((# bins/2)2-(# bins/2-current bin)2)+1

Final model: λ(L,JR)=F\*{(B)(L)+J/3[W1\*JR(-10)+JR(-5)+W2\*JR(0)]}

Our simulation predictions are generated by the following:

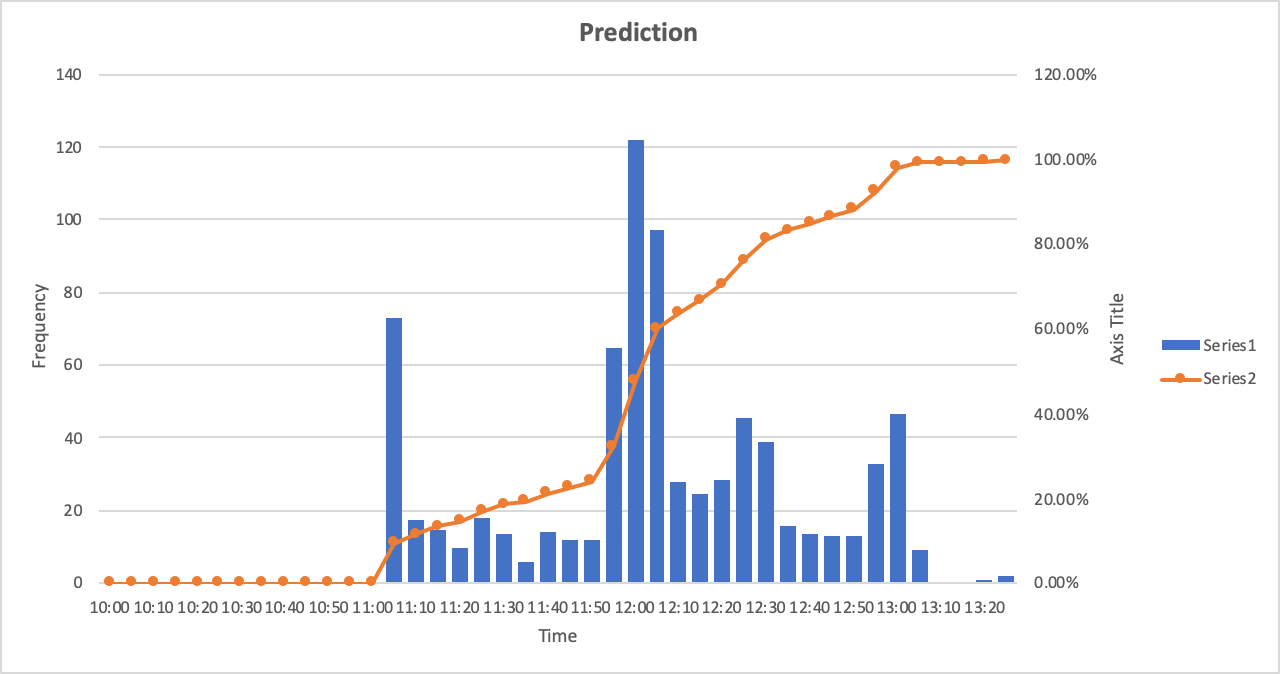
F\*Poisson\_Inv(Rand(),λ)

Running this simulation for many trials allowed us to define the set of possible outcomes and probabilistically quantify the effect that alterations to the class schedule can have on easing times of overcrowding.

**Analysis**

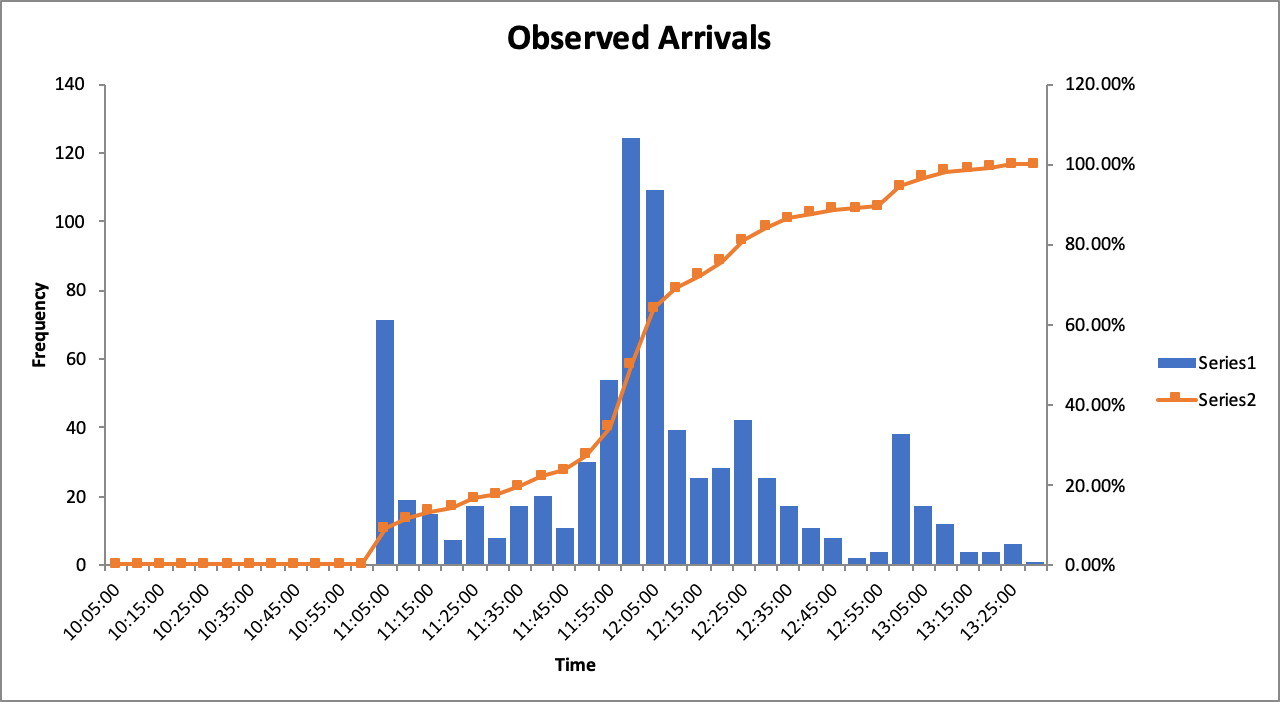
Once the form of the model was created, we began by applying it to the Monday data of people per class time slot (i.e. 10:00-11:00am etc.) to ensure a good approximation with low residuals and an accurate representation of the total number of swipes. Doing this yields a predicted number of people per five minute time slot that can be compared to the known swipe data from Bon Appetit. For Monday, the model returned the following plot of swipes per time block.

Monday Prediction:



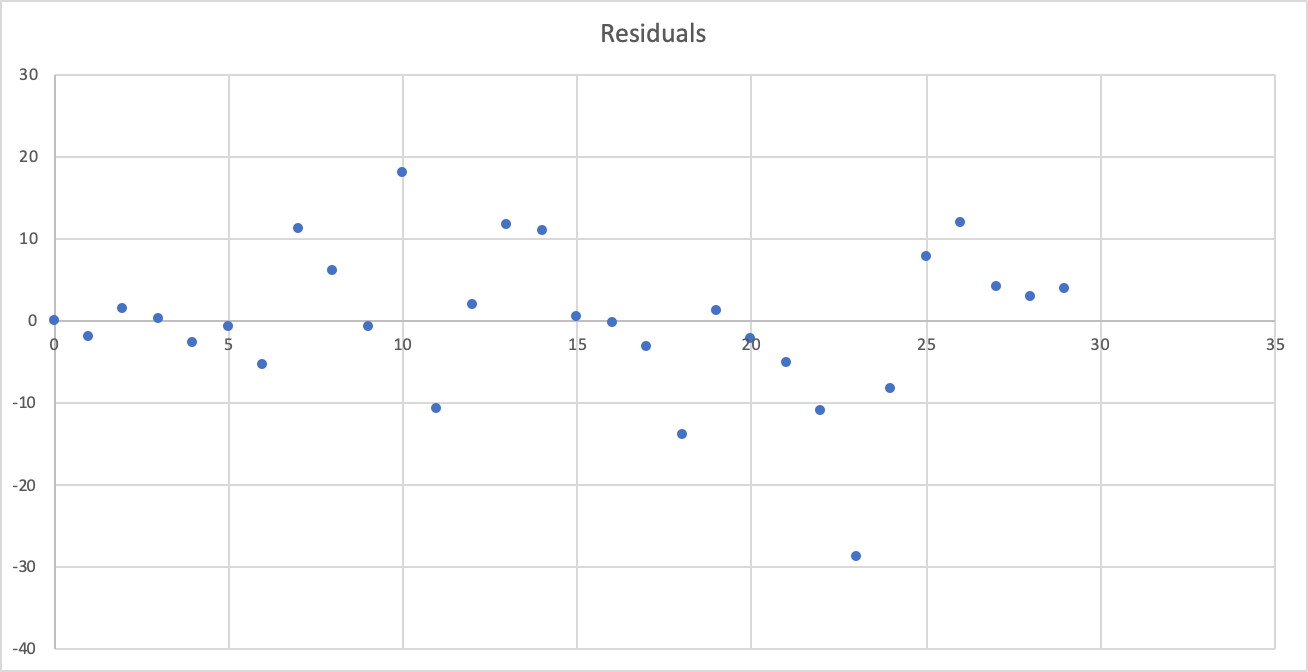
As the model is dependent on a random variable, this plot is only one of many, but due to how small the fluctuations in our predictions are, it is still a good representation of our model. Comparing this to the graph of observed arrivals, we see an apparent visual similarity in the shape of the distribution.

Monday Observed:



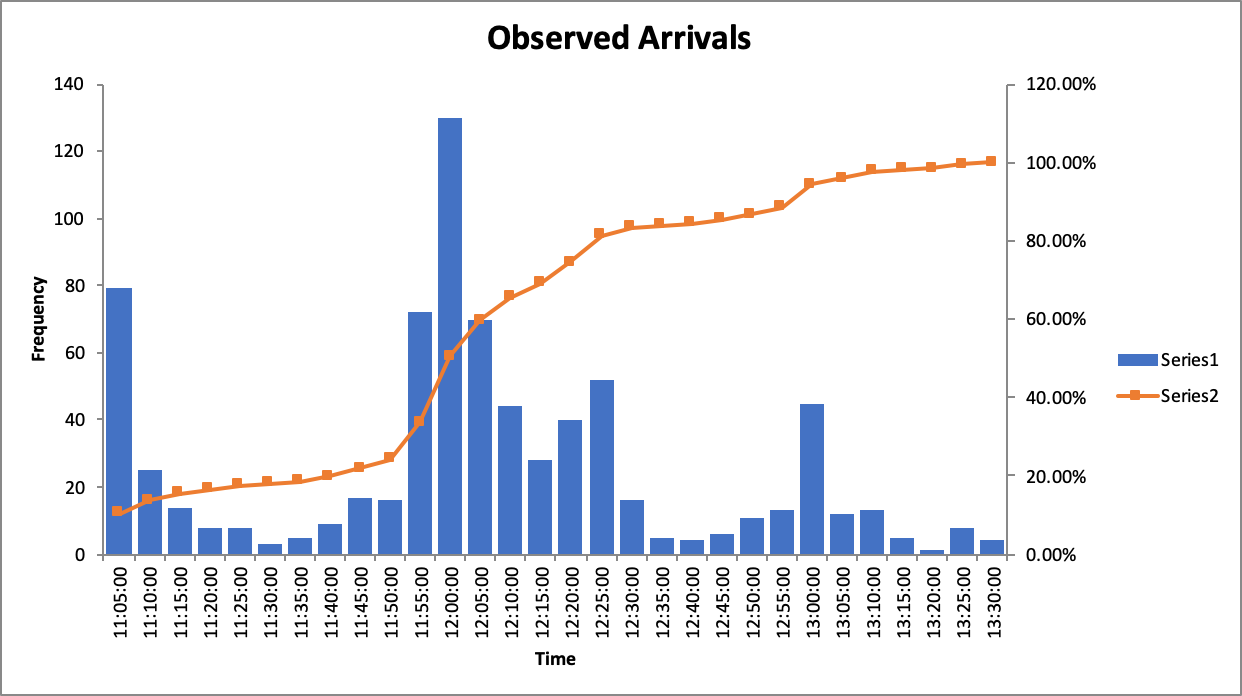
The accuracy of the model is seen better in the residual plot, the sum of the residuals being relatively small at just under 200, the R2 at about 3500, and the predicted total number of people to swipe in being approximately 780-787 as compared to the observed 785.

Residual Plot Monday

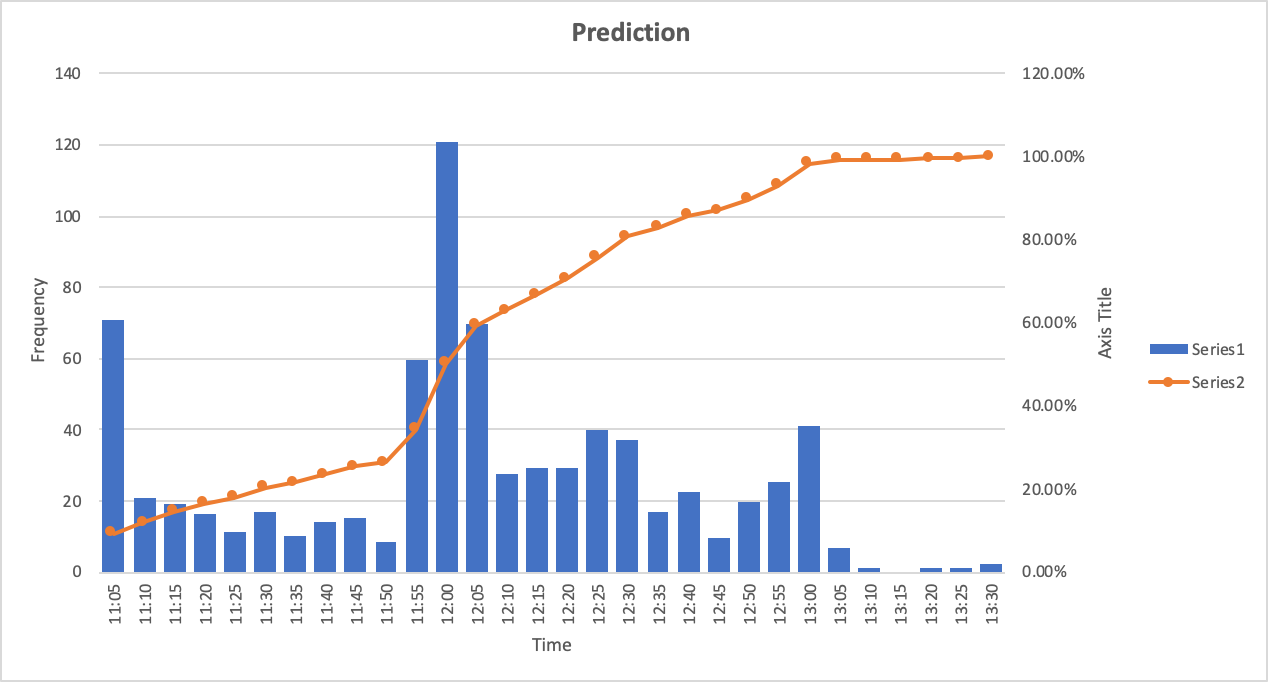


That the residuals were not fanning, linear, or curvilinear gave indication that the model properly fit the trend given in the observed data. It is important to note that the model was generated based on the Monday data, so to check for overfitting of that particular set, the model was also checked on Wednesday and Friday data in the same manner. Wednesday produced the following plots for observed data, model prediction, and residuals.

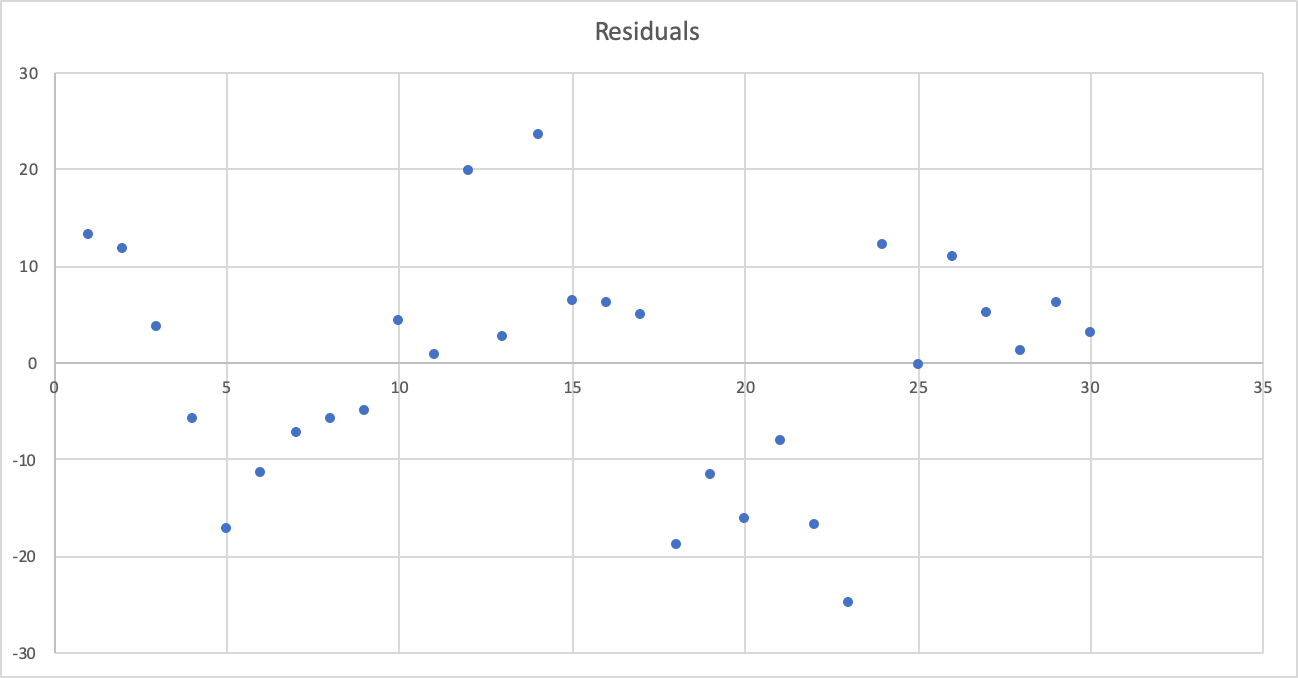
Wednesday Observation



Wednesday Prediction

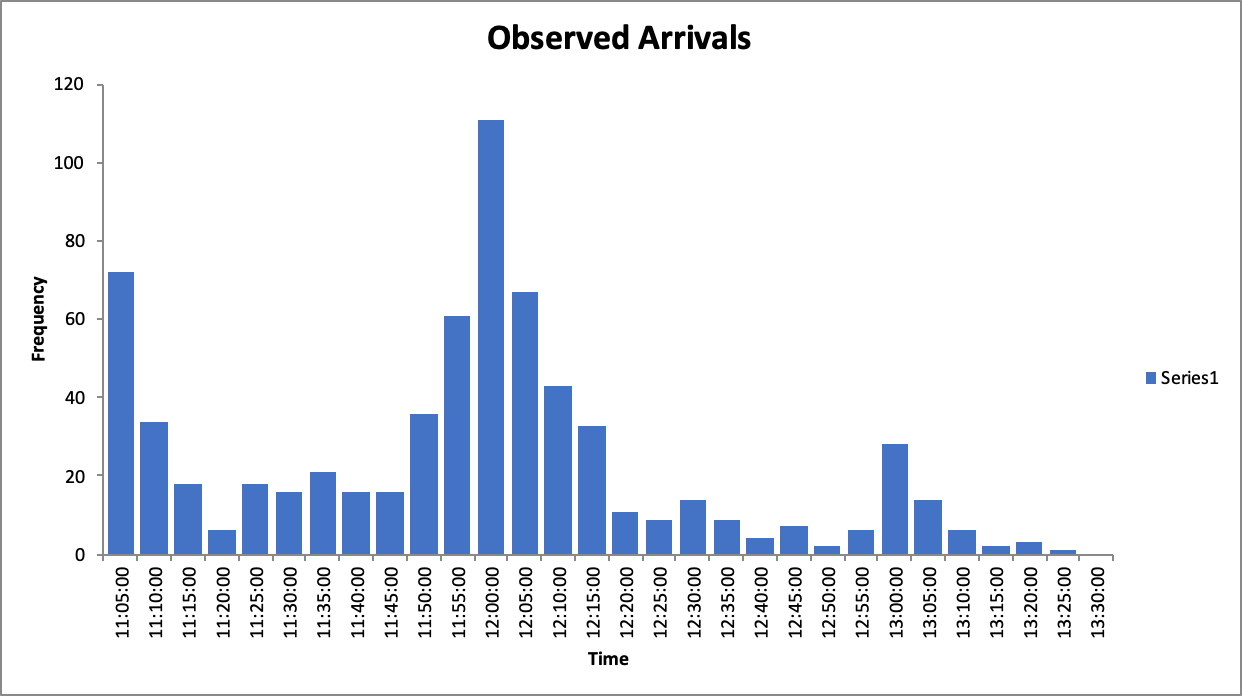


Wednesday Residuals

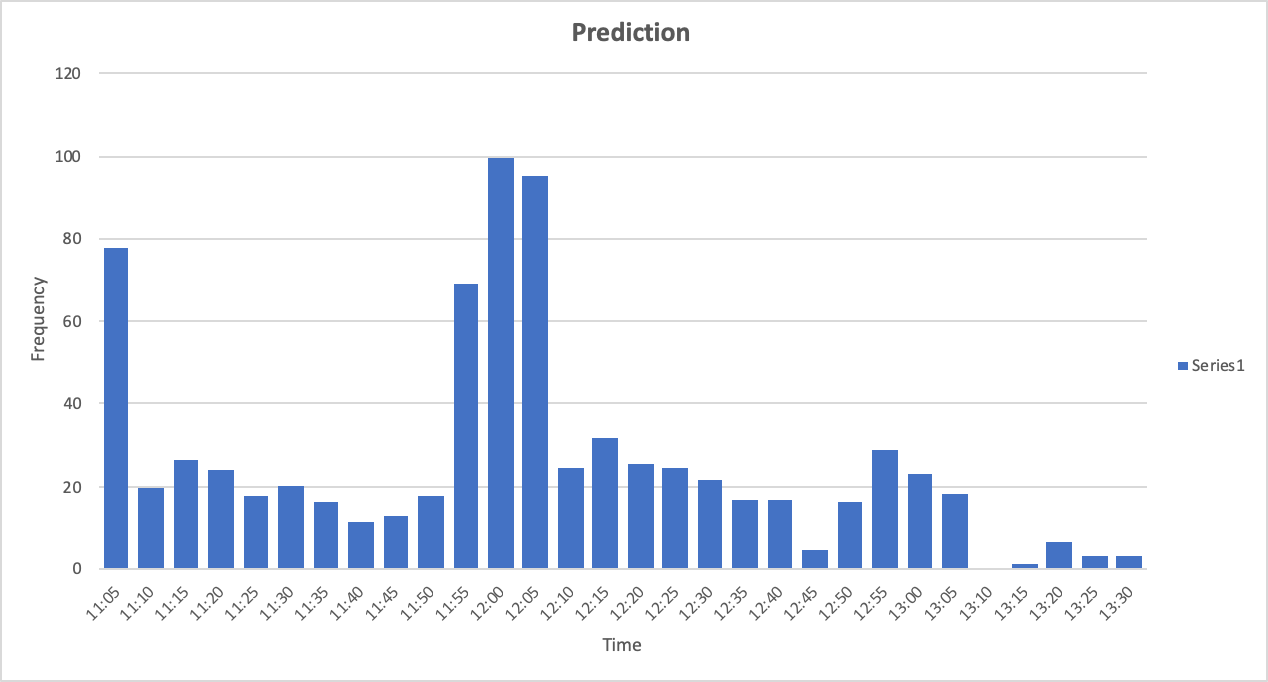


Here the sum of the residuals is slightly greater at 230, but the R2 value is significantly lower at around 2700, and the simulated number of total swipes is 760 on average with an observed value of 763. Additionally, the residual plot remains not as a fan, linear, or curvilinear plot, which demonstrates that the model still encapsulates the general trend of the Wednesday data that it had been untrained on. Moving on to the Friday data, the model produced the following plots.

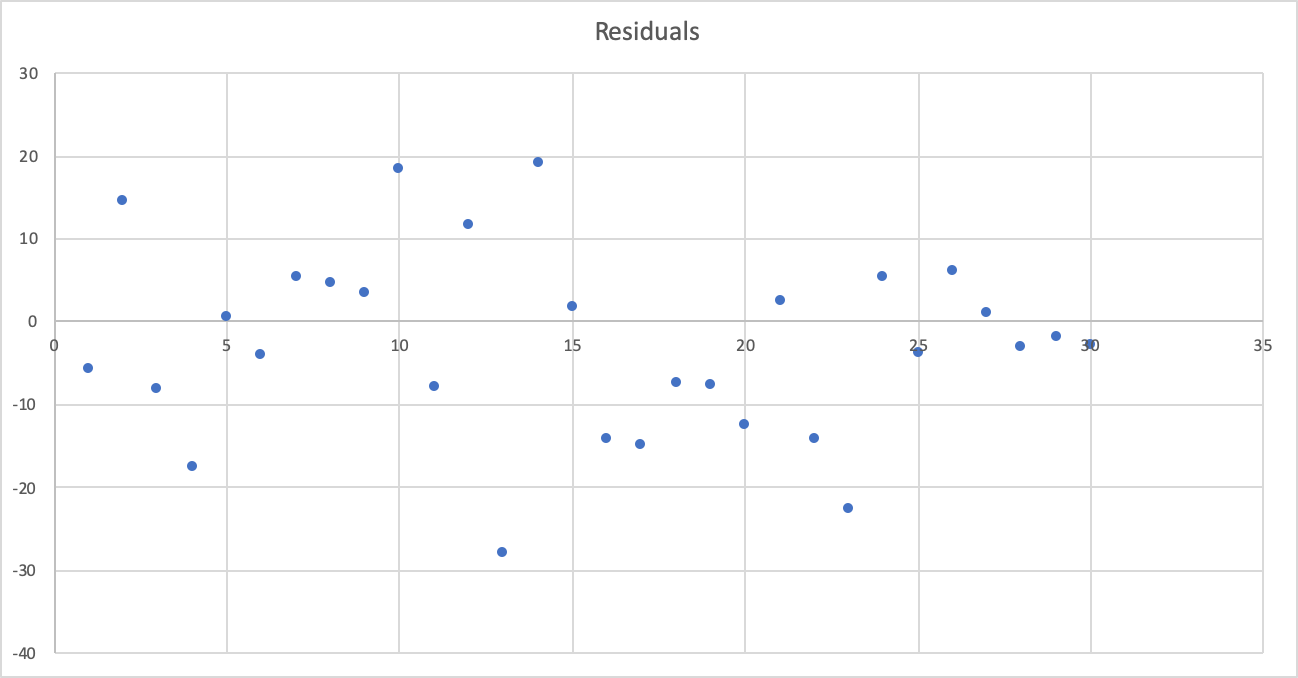
Friday Observation



Friday Prediction



Friday Residuals



Here the sum of the residuals is at 250, with the R2 value at around 4000. Additionally, the simulated number of total swipes is 770 on average with an observed value of 684. The predicted values for Friday are the most deviational from the observed data of all three days. However, on inspection, this seems to be rooted in that far fewer people appear to attend Friday lunch than Monday or Wednesday based on the observed data even though there are roughly the same number of people attending class over the entire time period. There seems to be another potential hidden variable of human behavior here that lies outside the current focus of our model. However, although there seems to be uncertainty surrounding the number of swipes predicted, that the general trend is still being fit can be seen in the residual plot. Once again, it is centered around the y-axis and displays no fanning, linear, or curvilinear traits. This constitutes the process we used to determine the feasibility of our model to accurately predict the number of swipes per five minute interval based only on class schedule.

Due to the nature of our data, model, and means of applying it in a Monte Carlo approach, it became obvious that computational tools would be largely utilized in exploring the model and analysis. Although an analytic approach to implementing the model was not possible, a better understanding of our two main parameters, Just Released and Lunchables, could be made through an analytic display of their relationship. This has already been done in the prior Model Formulation section of the report, which displayed a contour plot of the average number of arrivals as dependent on those variables as well as a detailed explanation of how these variables influence the prediction in the full model.

The final step of the analysis is then to apply the model to each day with a different class schedule to achieve our goal of minimizing traffic flow. As our model is probabilistic in nature, we utilized the Monte Carlo approach of taking a large number of predictions and determining our minimization on the mean values generated from that set. For this project we decided to generate 200 models per day. At first we were interested in determining the mean number of swipes per time slot, and minimizing the residuals of the generated swipes. This would, in theory, return the class schedule needed to produce the most constant level of traffic possible. However, due to the nature of this problem as largely computation, attempting to minimize the mean value of 200 models for ten parameters (the class schedule time) is next to impossible. Any type of solver or minimization tool quickly finds a local minimum at almost the exact same values (often off from the originals by a factor of 0.001). Instead, we chose to simply take the maximum number of swipes per time slot for each day, determine 75% of that, and round it down to the nearest tens spot. The traffic will be sufficiently minimized if we can demonstrate that any particular 5 min window (of all time slots of the 200 models that were generated) has a 0.5% chance of having more predicted swipes than that value. In order to make this problem dependent on the number of classes per class time slot, we took the average number of students per class at those times (15, as determined from Web Advisor) and used that as an intermediary between the number of students that the model depends on and the number of classes that we were changing. Additionally, to account for the difficulty of being able to change classes around, and teacher preference for teaching specific times, we restrained ourselves to only being able to change 30 classes from one time slot to another for each day.

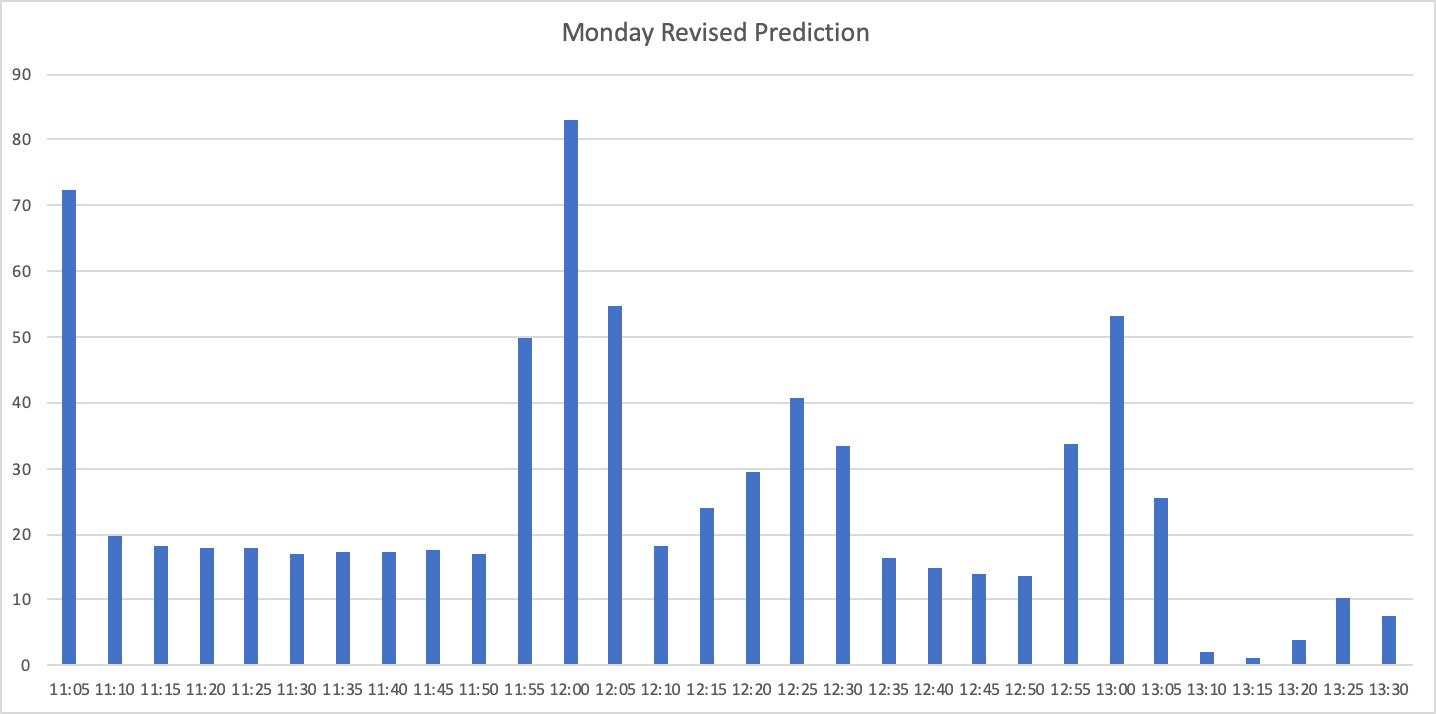
Beginning with Monday, the maximum number of swipes was 124 in a five minute block. Applying the 25% reduction rate and consequent rounding down led to needing to get 99.5% of all swipes down to under 90 per time block. We started with the original number of classes per class time schedule as shown below.

| Monday Original Schedule | |
| --- | --- |
| Class Slots | Number of Classes |
| 10-10:50 | 55 |
| 11-11:50 | 46 |
| 12-12:50 | 10 |
| 1-1:50 | 40 |
| 10:00-11:20 | 0 |
| 11-12:05 | 2 |
| 11-12:15 | 9 |
| 12-1:15 | 0 |
| 12-2:00 | 1 |
| 12:30/45-1:45/50 | 4 |

We ended up changing the 30 maximum class times to yield a percentage of 99.6% of time slots having less than 90 swipes. We focused on moving classes away from ending at 10:50am and 11:50pm to avoid the large spikes in available swipes and towards those ending partially through the hour, such as 11:20, 12:05, and 12:15. This yielded the following revised class schedule and plot of the average swipes per time slot for all 200 models.

| Monday Revised Schedule | |
| --- | --- |
| Class Slots | Number of Classes |
| 10-10:50 | 51 |
| 11-11:50 | 20 |
| 12-12:50 | 19 |
| 1-1:50 | 40 |
| 10:00-11:20 | 10 |
| 11-12:05 | 5 |
| 11-12:15 | 12 |
| 12-1:15 | 5 |
| 12-2:00 | 1 |
| 12:30/45-1:45/50 | 4 |

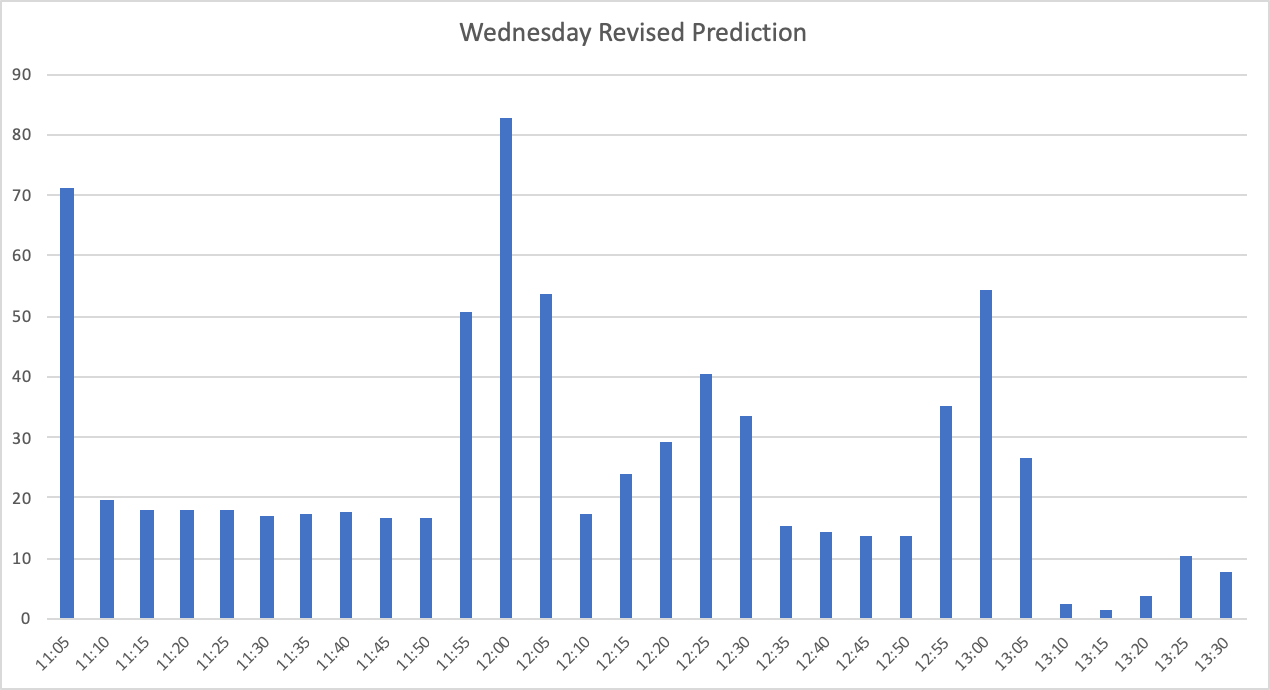
Monday Revised Swipe Prediction



Wednesday’s maximum number of swipes was 130, and so we once again simply needed to get beneath the 90 swipe threshold for 99.5% of all time slots. Below is the original number of classes per class time schedule. We once again changed 30 class times, this time with a percentage of 99.7% of time slots having less than 90 swipes. We continued focusing on moving classes away from ending at 10:50am and 11:50pm and towards those ending partially through the hour. This yielded the following revised class schedule and plot of the average swipes.

| Wednesday Original Schedule | | Wednesday Revised Schedule | |
| --- | --- | --- | --- |
| Class Slots | Number of Classes | Class Slots | Number of Classes |
| 10-10:50 | 55 | 10-10:50 | 50 |
| 11-11:50 | 45 | 11-11:50 | 20 |
| 12-12:50 | 11 | 12-12:50 | 20 |
| 1-1:50 | 40 | 1-1:50 | 40 |
| 10:00-11:20 | 0 | 10:00-11:20 | 10 |
| 11-12:05 | 2 | 11-12:05 | 5 |
| 11-12:15 | 9 | 11-12:15 | 12 |
| 12-1:15 | 0 | 12-1:15 | 5 |
| 12-2:00 | 1 | 12-2:00 | 1 |
| 12:30/45-1:45/50 | 4 | 12:30/45-1:45/50 | 4 |

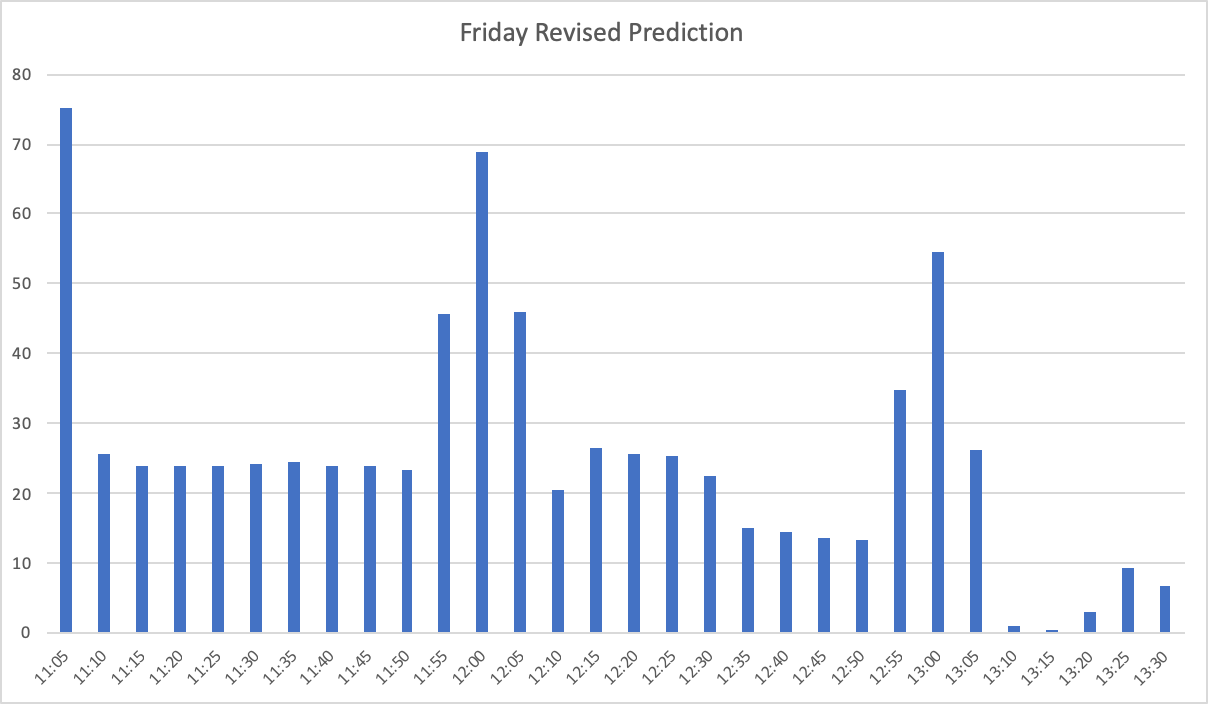
Wednesday Revised Swipe Prediction



Friday’s maximum number of swipes was 111, making the swipe threshold 80 for 99.5% of all time slots. The original number of classes per class time schedule is shown below. We once again changed 30 class times, this time with a percentage of exactly 99.5%, away from ending at 10:50am and 11:50pm and towards those ending partially through the hour. This yielded the following revised class schedule and swipes plot.

| Friday Original Schedule | | Friday Revised Schedule | |
| --- | --- | --- | --- |
| Class Slots | Number of Classes | Class Slots | Number of Classes |
| 10-10:50 | 53 | 10-10:50 | 48 |
| 11-11:50 | 44 | 11-11:50 | 19 |
| 12-12:50 | 9 | 12-12:50 | 20 |
| 1-1:50 | 37 | 1-1:50 | 40 |
| 10:00-11:20 | 0 | 10:00-11:20 | 2 |
| 11-12:05 | 2 | 11-12:05 | 5 |
| 11-12:15 | 0 | 11-12:15 | 4 |
| 12-1:15 | 0 | 12-1:15 | 5 |
| 12-2:00 | 0 | 12-2:00 | 1 |
| 12:30/45-1:45/50 | 3 | 12:30/45-1:45/50 | 4 |

Friday Revised Swipe Prediction



**Interpretation and Refinement**

The low residual sum and R2 totals for our predictions for Monday, Wednesday, and Friday with the original class schedules, and the absence of any fanning, linear, or curvilinear trends in the residuals of any of the days, indicate that the model does a reasonably good job of simulating the actual flow of traffic into the dining hall. And our predictions for Monday, Wednesday, and Friday with the new class schedules, while they cannot be tested on any actual data, appear to account, as we would expect them to, for the changes we made to class schedules, with the exception of the fact that adding ten 10-11:20 classes on Monday and Wednesday did not produce any significant spike in traffic flow between 11:20 and 11:30 to correspond with the fair increase in the Just Released pool at this time. Instead, the shapes of the graphs for Monday and Wednesday are roughly the same in the 11-12 hour as Friday, even though we added only two 10-11:20 classes on Friday. Our model does capture the fact that there are fewer people available on Monday and Wednesday between 11 and 11:20 than on Friday by producing lower arrival rates over this time interval for these days than for Friday; however, the arrival rates for Monday and Wednesday remain lower than Friday through the rest of the 11-12 hour. Since we cannot really be sure how our model would actually fare were it to be tested, it cannot be said that one seemingly absent effect alone is enough to bring the validity of the model into question, especially since it models all the other changes to the class schedules almost exactly as would be expected for each of the three days. It could be that our model, which is dependent on only the Lunchables and Just Released parameters and the quadratic function F that gives more weight to the bins in the middle of the lunch period, is not quite complex enough and needs to take into account other factors that might come into play as we change class schedules. It is hard to say, however, what these factors might be, so unless the class schedules actually are significantly changed from their current structure (they tend to be about the same every semester), we cannot really hope to do any better than rely on the model we have, while perhaps making very slight adjustments to our three parameters in light of more data or information.

We could, perhaps learn more about the problem not only by testing and refining our model on different weeks (which may produce effects based on factors we weren’t aware of when we first constructed the model, allowing us to refine it based on this new data or information), but also by creating a new (although probably fairly similar) model for Tuesday and Thursday, and even maybe a third one for Saturday and Sunday. Studying these two pairs of days might provide insight into other factors that affect traffic flow that either weren’t significant enough when we restricted ourselves to our one week and three days, or were actually significant but were simply overlooked, and could lead us to consider not only adjusting our three parameters but also perhaps adding new parameters we simply weren’t able to think of when looking at only the one week and three days we started with.

Moreover, we could even consider modeling a function with entirely new variables other than the one we looked at in this model, rates of arrival. In the beginning stages of the modeling process, we were considering both arrival times (for which we had hard data) and departure times (for which we had no data and had to guess at intuitively) to create graphs of overall traffic levels throughout the lunch period, which we produced two examples of in the beginning of the “Model Formulation” step of this report. At the end of the day, however, we decided to settle on studying exclusively arrival times in our model, which was a reasonable decision considering that the problem of overcrowding in the dining hall typically happens at around the same time every day (as evidenced by the graphs) and is a problem not only of overall traffic levels but also of bottlenecking in arrival rates (there is always a long line and a crowd of people getting their food at around noon every Monday/Wednesday/Friday). Traffic levels, additionally, would be more difficult to model than arrival rates if we took into account the factor that people would presumably be more likely to leave if they have to go to class soon (so people who arrive at, say, 12:30 would probably stay shorter on average than people who arrive at 12:00, since a lot of them will have to get to class at 1:00.) While this would be a more challenging problem than simply looking at arrival rates, it could probably be done in similar fashion if we were able to get data on how long people who arrive at various times stay, which we could probably do by taking enough samples from different arrival times (probably through a random poll of students) and using this data to make generalizations on how long people on average people stay (and with what standard deviation) depending on the particular time they arrive.

One other thing we could consider would be to take a different approach to solving the general problem we settled on: reducing traffic flow into the dining hall. At the beginning of the modeling process we were unsure whether to do this by changing class schedules, as we ended up doing, or by extending the time the dining hall is open, probably at the beginning (since there is very little traffic by the end.) While this might have also proven effective in bringing about a 25% or so reduction in the maximum 5-minute bin over the lunch period, we weren’t sure it would be, especially once we added in the quadratic F that gave more weight to the middle of the lunch period. We felt like we had a good enough understanding of the effects produced by class schedules--the Lunchables and Just Released parameters--that it would be better to focus our model more on these effects, although we still had to consider to some extent the fact that students are more likely to want to go at the time where more people are there because they want to eat with friends. Trying to model opening the dining hall earlier would have made the problem mainly about this quadratic F, which was itself an estimate, so we were more confident in modeling changing class schedules since these are more or less hard data.

There were still points in time, however, where we had to exercise creativity, guesswork, and judgment. As we were trying to figure out how we were going to reduce traffic flow without trying to change more than is practically possible with class schedules, we initially thought about doing this by minimizing the function:

v(Yi - Ӯ) + wck

where Yi is the prediction for each 5-minute bin, Ӯ is the mean over the entire lunch period (all thirty 5-minute bins), c is the variable of the function--the number of changes made to class schedules--and v, w, and k are parameters we can control, with v and w representing weights for the two components of the function. Our plan was to make v significantly larger than w--say, v=.8 and w=.2--but to make k fairly large, say, k=2. This would ensure that for low values of c--which, presumably, would leave higher variance in the model--the variance would dominate and be too big (so we’d be forced to make enough changes to make a significant impact), but once c got high enough, raising it to an appropriate power would cause this factor to dominate and be too big (so we wouldn’t be allowed to make more changes than would be practically possible, per se.) The problem with approach, however, ended up being that we were unable to set up the minimization in a practical manner for so many different possible changes to the class schedule. So instead of trying to obtain a “precise” solution that in theory perfectly balanced the two objectives of reducing traffic flow and limiting the number of changes to the class schedule, we settled instead on minimizing traffic flow alone while subjecting it to the constraint that no more than 30 changes be made to the class schedule; and then, after coming up with a reasonable set of 30 changes to make to each of the three days, we simply checked to make sure that, 99.5 % of the time we ran the model, its maximum bin, rounded down to the nearest multiple of 10, was at least 25 % lower than the maximum bin for the original class schedule for that day. These somewhat arbitrary standards both for the maximum number of changes to the class schedule and for the threshold for satisfactory reduction in traffic flow reflect the reality that we were faced with on several occasions in the modeling process: while we would have ideally liked to add more complexity to the model, we often had to settle for a simpler model to make sure that we were able to obtain a satisfactory solution.