MathWorks Math Modeling Challenge 2021

Adlai E Stevenson High School
Team # 14485 Lincolnshire, Illinois
Coach: Paul Kim
Students: Spandan Goel, Andrew Liu, Joy Qu, Greycen Ren, Gabriel Visotsky

M3 Challenge FINALIST—$6,500 Team Prize

JUDGES' COMMENTS

Specifically for Team # 14485 —Submitted at the Close of Triage Judging:

COMMENT 1: This is an overall excellent paper. Your summary was one of the best I read. You did an excellent job of both giving an overview of your results and discussing your approach to each problem. The approaches to each problem were clear and explained very well in your narrative. A minor criticism is that the tables and graphs in problem 1 had some unclear labeling. I had to go back into the text to understand the units in a few cases. This did not seem to be an issue in problems 2 and 3. Your models were all well thought out, well supported, and well analyzed. This was an excellent submission!

COMMENT 2: Good job on preparing the model and presenting the paper. It is evident that the team has thought the entire modeling process through, paid attention to every broad aspect of the question, and presented their work as a neatly written work. A few thoughts for future improvement: please discuss long-term implications of your work. How likely is your prediction based on your model to be accurate, especially far out into the future? Try applying/modifying your model for different households with different demographics, and analyze how the predictions would change. Overall, very good work!

COMMENT 3: Good job with assumptions. You presented good critical and analytical thinking.

Defeating the Digital Divide

Team 14485

February 28, 2021

0   Executive Summary

We often see toddlers squealing on iPads playing Angry Birds while their parents sit a few feet away, dining on steak and potatoes. The sight of young children playing with technology is becoming more and more common in this digital age. Likewise, the information we consume from digital devices increases at an outstanding rate, especially with the integration of technology in schools and workplaces. However, this is all only possible with internet access. Those without sufficient access may have disadvantages in participating in remote school and work, receiving adequate healthcare, accessing information, etc. But internet access and mobile broadband can be costly for high speeds. To further analyze the future of internet, we created three models that consider internet costs, needs and optimal planning.

The cost for bandwidth has been decreasing since the past decade in both the US and the UK [1]. Using existing data on broadband subscriptions [29], we created a model that predicts the per-unit bandwidth cost over the next 10 years for US and UK consumers. Our model utilizes a differential equation relating per-unit bandwidth cost and time along with factors including supply and demand of internet access. In order to derive numerical calculations for the demand factor, we also use a logistic equation to approximate the number of people with internet access over time. We then integrate our function and compute the per-unit cost of bandwidth in 2031, to get \$0.38 and \$0.54 in the US and UK, respectively.

In order to create a model for the minimum bandwidth required for 90% and 99% of internet usage, we first looked to find what bandwidth is primarily used for. We determined that the main categories of internet usage are entertainment, bandwidth used for enjoyment purposes, work, and school. We then further specified the categories into activities for more nuance; for example social media would be an activity of entertainment. By multiplying the bandwidth rate required for a certain activity by the amount of time that activity takes, we found the total bandwidth required per person per activity. We then summed the this over all activities, categories, and finally over all people in a family. However, to create a more robust model and to account for realistic variation in the time an activity takes, we created a probability density function (PDF) to model the biggest bandwidth consumer and one of the most varied activities: online TV streaming. Taking all of this into account, we ran a Python simulation that gave us the standard deviation and the mean of bandwidth consumption for each of the three households, which gave the 90% and 99% levels. We found that the 90% and 99% levels are 4150.42 and 4152.5 GB per year for Family 1, 4548.26 and 4554.09 GB per year for Family 2, and 12188.52, and 12191.33 GB per year for Family 3.

We then turned our focus towards 5G mobile broadband infrastructure. Cellular nodes are placed in various locations around the world, allowing smartphone users to access 4G/5G data. The location of these cellular nodes are chosen after mathematical calculations to maximize the number of users that can connect to the towers in a certain geographic area. Using factors such as population density, percentage of smart phone users, income, wage and cost of living [1], we created a model that produces an optimal plan for placing cellular nodes in a given region. We applied the model in three hypothetical regions A, B, and C, ranking the subregions within each region from most to least important in terms of receiving a cellular node, creating an optimal distribution plan to the subregions. We find that out of all the subregions, C3, C4, and C5 should be distributed cellular towers first. Region by region, the first subregions should be A5, B3, and C3.

Contents

1 Part I: The Cost of Connectivity

1.1 Restatement of the Problem

We are tasked to do the following:

- Create a model to find the per-unit price of bandwidth.

- Use our model to predict values over the next 10 years.

Our model uses a differential equation relating per-unit bandwidth cost over time. Factors considered in this equation include supply and demand of broadband internet access, in which future supply is found through existing data predictions and future demand is calculated by a logistic model predicting the number of people with broadband internet access. For the internet demand logistic model, we use existing data on the increasing number of internet subscriptions starting from year 2000 to year 2018 to model growing access of internet broadband on a timeline. After determining supply and demand, we apply the final model that predicts per-unit bandwidth cost in the US and UK in 2030.

1.2 Assumptions

1. The number of fixed subscriptions for broadband internet is proportional to the number of people with access to broadband internet.

   - **Justification**: Broadband internet access is primarily geographical, where separated geographical areas uniquely provide broadband access. Therefore, everyone amongst those with subscriptions and access resides in similar locations. Hence, service for broadband internet is determined by the proportion of people in a geographical area— the number of people with access—and the rate at which they subscribe to service providers.

2. The change in population of the United States is negligible.

   - **Justification**: This is negligible in the final calculations

3. The growth of broadband access follows a logistic curve.

   - **Justification**: Initially, internet is an unknown technology and the growth of broadband internet access is slow. However, as the use of internet becomes more widespread, the growth of broadband access rapidly increases to accommodate increasing popularity. Eventually, as access to broadband increases to a carrying capacity, lack of access is sequestered amongst those that are hardest to provide access to. This relative model of growth is consistent with that of a logistic growth model.

4. An economic model of supply and demand can be used to determine the equilibrium price as the per-unit bandwidth cost.

   - **Justification**: The equilibrium price is the market price that best profits sellers. The best price for sellers of internet bandwidth is the cost that they will put in the market, which is the cost per unit bandwidth for consumers.

5. The impact of supply and demand in the US and UK will be the same.

   - **Justification**: Consumer habits amongst those in the US and UK are similar [21], meaning that they will have similar views of the market. Consequently, the forces of supply and demand will have similar effects.

6. The cost of bandwidth follows a logistic decay.

   - **Justification**: At first, when bandwidth technology first comes out, the decrease in price is slow as technology is underdeveloped. However, price drops rapidly as new technological developments occur until, eventually, the decrease in price slows down again as the commercial prices reaches closer to the raw price of production.

7. The cost of bandwidth in the UK is proportional to the cost in the US.

   - **Justification**: Given that consumer habits are similar, especially for internet services [21], it is likely that the US and UK maintain a proportional cost difference.

1.3 Model Development

1.3.1 Parameters

1. **Year** (t).

2. **Number of People with Broadband Access** ($N_a(t)$). The number of people in the United States that have access to broadband internet as a function of time.

3. **Number of People with Broadband Subscriptions** ($N_s(t)$). The number of people in the United States that have subscriptions to broadband internet as a function of time.

4. **US total population** (c). The total US population represents the carrying capacity in our logistic regression.

5. **Bandwidth cost** ($C(t)$). A model of the per-unit bandwidth cost as a function of time.

6. **Demand** (d). The demand for broadband internet access.

7. **Supply** (s). The supply of broadband internet access.

1.3.2 Model Derivation

We begin by looking to quantify the rate of change of per-unit bandwidth cost over time as a differential equation. In a free market, the equilibrium price of a market is determined by the supply and demand for a good [14]. With consideration to *Assumption 4*, we will use the supply and demand as factors in determining the cost per unit bandwidth. Other factors such as cost of energy or cost of long distance services are not considered because these are all expenses. Due to the nature of the supply-demand model, the equilibrium point already takes into account any changes in expenses because it reflects highest profit, not revenue. Therefore, a possible increase in energy expenditures in the future is not considered in our model, as it is already accounted for when we consider profit.

We use historic data on internet access in the US to mirror general internet access in the UK, as the technological infrastructure for internet is comparable between the two countries [27]. In this case, over time, both supply and demand of internet access increase because, as the US and UK become even more digital, producers increase their supply as more consumers connect to the internet for long periods of time, especially with the impact of COVID-19 on increasing reliance on the internet. In terms of a simple economic supply-demand model, an increase in supply decreases the equilibrium price whereas an increase in demand increases the equilibrium price. Following *Assumption 6*, there is a minimum price for the cost of bandwidth. Intuitively, this would occur when the cost of bandwidth is equal to the cost of production. Given that there is a roughly 12.82% margin for bandwidth [24], with a cost of 43 cents in the US [5], the raw cost of production is around 38 cents. Thus, we can write that

$$\frac{dC}{dt} = -k_1 \cdot \frac{1}{s} \cdot d \cdot C \cdot (0.38 - C) \tag{1}$$

with constant k_1 and rates of change s, d. On the supply side, the expansion of the submarine cable market increases by a cumulative annual growth rate of 7.1% from 2020 to 2027 [15]. As submarine cables account for nearly 90% of all internet traffic [15] [26], this growth approximately accounts for the entire growth of the supply. For sake of brevity, we assume this to continue to 2030. Thus, for our purposes, we know the value of s to be 1.071.

Determining demand growth of broadband internet access is a more rigorous process. However, we can use internet access as a mirrored function for demand, as consumption by consumers, in this case an increased access to the internet, is the primary component of aggregate demand [20].

Since the growth of internet services in general cannot increase exponentially, instead, as per *Assumption 3*, we use a logistic model to regressively quantify a function for demand. Using data for broadband internet access, we can determine the rate of increase in internet access in general, which represents an increase in demand of internet.

We begin by looking for a way to quantify the growth of access to broadband internet services. Unfortunately, there is no existing data regarding historical access to broadband internet in the US. Therefore, we must make use of *Assumption 1* to create approximate historical data for access to broadband internet.

Our estimate takes the form of the following equation:

$$N_s(t) = \lambda_1 \cdot N_a(t) \tag{2}$$

Problematically, the only known value for $N_a(t)$ is for 2020, a year where no historical data is available for $N_s(t)$, which only spans from 2000 to 2018. However, using *Assumption 3* and the data from **Table 1.1**, we are able to estimate the value of $N_s(2020)$ with a logistic regression to determine a numeric value for λ_1. Our logistic takes the general form

$$N_s(t) = \frac{c}{1 + Ae^{-b(t-2000)} + d}$$

where A, b, d are constants and c is the population of the United States, a logical maximum number of people with subscriptions to broadband internet. Fitting this equation to the historical values of $N_s(t)$ from $t = 2000$ to $t = 2018$, we can numerically determine the values of our constants. Plugging in our constant values, we determine a quantitative equation to approximate $N_s(t)$:

$$N_s(t) = \frac{330.01}{1 + 0.581562e^{-0.147446(t-2000)}} - 207.183, \qquad r^2 = 0.9926 \tag{3}$$

Number of broadband subscriptions $N_s(y)$ over time	
Year	$N_s(y)$
2000	7.07
2001	12.79
2002	19.88
2003	27.74
2004	37.35
2005	51.16
2006	60.24
2007	71.70
2008	77.13
2009	79.99
2010	84.52
2011	88.32
2012	92.51
2013	96.03
2014	97.81
2015	102.21
2016	105.71
2017	108.19
2018	110.57

Table 1.1: Number of broadband subscriptions, in millions, from 2000-2018 [29]

Number of people with internet access $N_a(y)$ over time	
Year	$N_a(y)$
2000	19.44
2001	35.17
2002	54.66
2003	76.27
2004	102.69
2005	140.66
2006	166.07
2007	197.14
2008	212.06
2009	219.93
2010	232.38
2011	242.83
2012	254.35
2013	264.03
2014	268.92
2015	281.02
2016	290.64
2017	297.46
2018	304.01

Table 1.2: Number of people with internet access, in millions, from 2000-2018

Computing $N_s(2020)$, we determine that approximately 113.15 million people have subscriptions to broadband internet in 2020. Additionally, we know the number of people without broadband access [39] and the population of the United States in 2020 [7], allowing us to compute the value

of $N_a(2020)$, approximately 311.10 million people. Plugging in $N_a(2020)$ and $N_s(2020$ to (1), we determine λ_1 to be approximately 2.74945. Using (1) and our value of λ_1, we can calculate the value of $N_a(y)$ annually from 2000 to 2018 as seen in **Table 1.2**.

Once again, per *Assumption 3*, we can fit our data to a logistic curve, taking the form $N_a(t) = \frac{330.01}{1+Ae^{-b(t-2000)}}$ with $A = 6.08569$ and $b = 0.265082$. This gives us our numeric equation that relates $N_a(t)$ and t.

$$N_a(t) = \frac{330.01}{1 + 6.08569e^{-0.265082(t-2000)}}, \qquad r^2 = 0.9697 \tag{4}$$
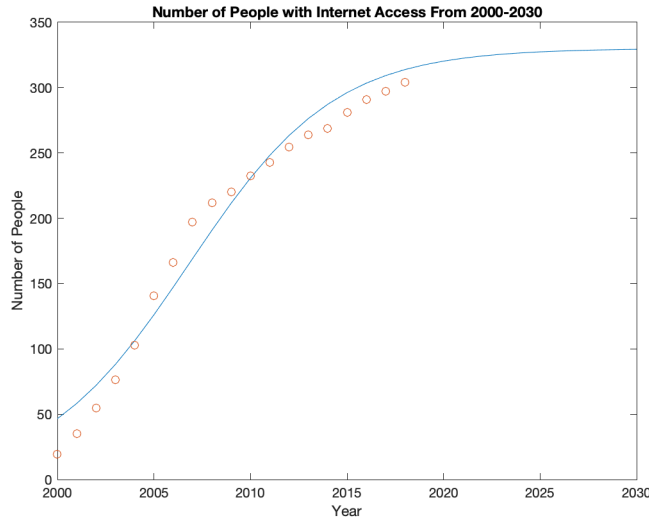

Fig 1.1

In a standard logistic equation following the form $\frac{dN}{dt} = rN(K - N)$, the intrinsic rate of growth, r is expressed in the solution as $rK = b$. In our equation, we can thus solve for r by dividing b by c, the "carrying capacity" for internet demand in the US. This yields $d = 1.000803$.

We can now solve for C. Integrating our differential equation for C, we find that

$$C(t) = \frac{0.38}{1 - Ae^{-0.38 \cdot k_1 \cdot \frac{d}{s} \cdot (t-2021)}} \tag{5}$$

We can now delineate our function $C(t)$ into $C_{\text{US}}(t)$ and $C_{\text{UK}}(t)$ for the cost of bandwidth in the US and UK respectively. We know that $C_{0\text{US}}$, the current price per Megabit per second in the US, is 43 cents. Further, in the UK, it costs 60 cents per Mb/s, giving us the value of $C_{0\text{UK}}$ [5].

We can then plug in $t = 2021$ and set the equation equal to C_0 to compute A, finding that $A_{\text{US}} = 0.1163$.

This gives us the following equation:

$$C_{\text{US}}(t) = \frac{0.38}{1 - 0.1164e^{-0.38 \cdot k_1 \cdot \frac{d}{s} \cdot (t-2021)}} \tag{6}$$

We also must compute k_1. To compute a value for k_1, we use the provided data for 2012 [1]. Using median values, we can compute the median value of the cost per Mb/s of bandwidth for each city. Simply taking the average of these values gives a representative cost of \$2.37 per Mb/s. Plugging this into (6) gives $k_1 = 0.6959$

Per *Assumption 7*, we also know that we can write the cost of bandwidth in the UK as a proportion of that of the US with constant of proportionality $k_2 = \frac{60}{43}$. Plugging in times t from 2022 to 2031, we can then compute the following costs per megabit of bandwidth in the US and UK.

Price of bandwidth, $C(t)$, over time in the US and UK over time in USD		
Year	$C_{\text{US}}(t)$	$C_{\text{UK}}(t)$
2022	0.42	0.58
2023	0.41	0.57
2024	0.40	0.56
2025	0.40	0.55
2026	0.39	0.55
2027	0.39	0.54
2028	0.39	0.54
2029	0.39	0.54
2030	0.38	0.54
2031	0.38	0.54

Table 1.3: Price of bandwidth over time in the US and UK

1.4 Results and Discussion

We predicted that the cost of bandwidth per Mb/s in 2031 for the US and UK to be $0.38 and $0.54 respectively (note: UK values were also calculated in USD). These values are reasonable, especially given that internet service providers are offering 5G internet services without additional costs compounded from more established 4G services [31]. Additional competition would decrease costs while increasing technological advancements increase speeds, generally decreasing the per-unit cost of bandwidth.

1.5 Sensitivity Analysis

We seek to test our model's reliability by performing sensitivity analysis based on possible discrepancies due to the assumptions above. For example, despite *Assumption 1*, although unlikely, it's possible for someone to have broadband internet access without a fixed broadband subscription. The person who got broadband internet access without a fixed broadband subscription wouldn't be accounted for in the variable d, the demand of internet access, because d was derived based on fixed subscriptions only. As a result, $\frac{d}{s}$ would also be slightly decreased since d is in the numerator. We simulate $\frac{d}{s}$ as a changing variable since this value contains the most uncertainty due to discrepancies of real world supply and demand of internet access.

Sensitivity for $\frac{d}{s}$		
% change in variable	% change in $C_{US}(t)$ for year 2031	% change in $C_{\text{UK}}(t)$ for year 2031
-10 %	+ 0.242 %	+ 0.242 %
- 5%	+ 0.113 %	+ 0.113 %
+5 %	- 0.098 %	- 0.098 %
+10 %	- 0.194 %	- 0.194 %

Table 1.4: Sensitivity Analysis for $\frac{d}{s}$

To calculate the percentage change for $C_{US}(t)$ in the above table, we plugged in 2031 to equation (6) with 0.95 $\frac{d}{s}$ first while keeping all other parts of the equation the same. Then we adjusted to 1.05 $\frac{d}{s}$. In the same manner, we adjusted $\frac{d}{s}$ for +/- 10%. Finally, we repeated this process by calculating percentage change for $C_{UK}(t)$ with +/- 5% and + 10% $\frac{d}{s}$ in the UK. We noticed that the effect of changing $\frac{d}{s}$ on percentage change of bandwidth cost is very small, with each all changes being less than 0.25%. As a result, such small effects means that the actual cost would not change when rounding to the hundredths place, remaining at \$0.38 and \$0.54. Hence, the validity of our model is confirmed.

1.6 Strengths and Weaknesses

1.6.1 Strengths

Our model incorporates historical data to draw relevant conclusions about the future, giving statistical credence to our observations. We use economic implications to highlight the influence of various factors on the cost of bandwidth. Our use of cost data eliminates the need to consider various factors that also influence cost such as increasing efficiency or power as calculated through the Shannon-Hartley theorem [30][32], as those costs are already assumed.

1.6.2 Weaknesses

Our model overextrapolates trends from the US for the UK. We rely on trends to support our observations towards demand and supply. Our calculation for demand is also indirect, potentially resulting in errors due to intermediate simplification. Per the Shannon-Hartley theorem [30] [32], we fail to recognize the impact of additional Gaussian white noises from further development of cables, which could impact the speed of internet services. We overgeneralize our data as a uniform distribution across the geographical stratification of the country. Our model may be predictive in regions with higher infrastructure development but fails to consider the different environmental factors in rural and technologically underdeveloped regions [33]. Lastly, our model does not consider economic inflation or deflation. Economic impacts could change future CPI for internet services, and the value of the US dollar or UK pound changes over time. We also fail to recognize that the lower bound for bandwidth cost changes as new technologies allow for the cheaper and more efficient production of internet services by providers. This causes our model to likely overestimate the per-unit cost of bandwidth in 10 years time.

2 Part 2: Bit by Bit

2.1 Restatement of the Problem

We are tasked to do the following:

- Create a flexible mathematical model to predict a household's internet needs over a year

- Apply the model to the three provided example households

- Analyze the three households with our model to determine the required bandwidth for 90% and 99% annual bandwidth usage

Our model for Part II categorizes internet usage in an household into three primary categories. We consider the time spent per day on each of these categories through various activities that make up the majority of the internet usage in that category. Assume that any time we refer to a *Family*, we are referring to one of the three households given to us in the prompt.

2.2 Assumptions

1. We can use the normal model to predict the 90 and 99 percent confidence intervals for the true proportion of bandwidth required for each of the three listed "families" in each internet usage category.

   - **Justification:** Because of the random variability of factors, such as the amount of television and homework assigned per day, we can utilize the normal model and standard deviation calculations to calculate our confidence intervals. Also, because of the way that our modern world revolves around the internet, the large sample size of internet usage events occur in multitudes, thereby making the normal model more effective. Although we acknowledge that certain factors may not be independent, ultimately the percentage difference between differing values of the standard deviation as a result of smaller changes in the way we consider independence are negligible to the overall result of the model. Lastly, the bandwidth usage is randomized between different users. Hence, the assumption conditions for a normal model in this case are met.

2. We can approximate each person within the three listed families to be an average representative of each of the respective age groups that they are a part of.

   - **Justification**: Because this is something that is not specifically part of the prompt, and we believe it is the most reasonable assumption of the internet usage of each family, we assume it to be true. However, we ultimately acknowledge the shortcomings of presuming that M3 undergrads are not more productive than the average person in their age group.

3. Entertainment, schoolwork, and work encompass all categories of internet usage.

   - **Justification**: Students and kids primarily use the internet for school. Adults and college students use the internet primarily for their work. Outside of this, people use the internet primarily for entertainment purposes, including but not limited to gaming,

social media, and streaming services. Any other usages for the internet beyond the scope of these three categories we assume to be negligible based on data for the biggest broadband consumption activities [36].

4. Zoom is the primary video conferencing platform for all school- and job-related work.

   - **Justification**: Zoom and other platforms such as Google Meets, Skype, Hangouts, and Whatsapp have similar bandwidth usages [8]. Furthermore, Zoom is the most used conferencing platform, making it most representative of the service in general.

5. Bandwidth usage remains constant throughout the year.

   - **Justification**: Seasonal and weather effects on internet are negligible for the purpose of this model.

2.3 Model Development

We are tasked with creating an effective model to predict any given family's internet needs over the course of the year. Let $\mathcal{F}$ denote the set of families given to us in the prompt, where each family $F \in \mathcal{F}$ is itself a set consisting of each person described as part of the household.

Our model is designed to predict any family's internet needs based on a number of the major categories of activities that may take up their time during the day. Per *assumption 3*, let $\mathcal{C} =$ {Entertainment, School, Work} be the set of the major categories of activities that we consider with our model, where each $C \in \mathcal{C}$ is itself a set consisting of the activities we consider as part of each category.

Furthermore, to better approximate usage within these three categories, we looked to split them into smaller activities. Per Assumption 2, we assume that the activities that we did not consider as part of our model to contribute negligibly to overall broadband connectivity, and thus we assume that they are all encompassing with respect to internet usage. These actives are as follows: Entertainment = {streaming tv, surfing internet, streaming video, social media, gaming, music}; School = {homework, school-related Zoom}; and Work = {work-related Zoom}. These activities account for a large majority of the broadband usage of each category and thus can be used to calculate broadband usage.

2.3.1 Parameters

1. **Total bandwidth per family F over one year $(B(F))$.** This is the function taking in any family $F \in \mathcal{F}$ and returning the total bandwidth usage by F over the course of a year.

2. **Time per activity $(T(A, P))$.** This is the function taking in any activity A and returning the amount of time that it is engaged in, per day.

3. **Average rate of bandwidth consumption $(\mu(A))$.** The average rate of bandwidth consumption required for any activity A, in megabits per hour.

4. **Total bandwidth per person per category $(f(C, P))$.** This is the total bandwidth used by a specific person, from a certain category of usage per day.

2.3.2 Model Derivation

When looking to find a model for the total bandwidth in megabits used by a family over the course of a year, we must first split up the calculation into time and bandwidth used by every person in the household. Since each person in every family has different digital habits, it is important to account for such variation in bandwidth usage. However, it is also important to account for the differences in bandwidth usage not only between different categories but also between separate activities within those categories. Therefore, for family $F \in \mathcal{F}$, the amount of bandwidth used by each person $P \in F$, for each category $C \in \mathcal{C}$, the following equation can be written:

$$f(C, P) = \sum_{A \in C} T(A, P) \cdot \mu(A) \tag{7}$$

Therefore, in order to calculate the total bandwidth usage in megabits for F, we sum over all people and categories in the family:

$$B(F) = \sum_{P \in F} \sum_{C \in \mathcal{C}} f(C, P). \tag{8}$$

Per *Assumption 5*, we assume that $\mu(A)$ stays constant, and therefore we can directly plug known values into our equation. However, we cannot make this same assumption for $T(A, P)$, as the amount of time that accomplishing the same activity over many times in the same year may differ substantially. Therefore, we instead model the function for certain activities A as a normal model, owing to the random and unpredictable nature of the time that many of the most bandwidth-consuming activities may take. More specifically, because the activity "Streaming TV" takes up the most amount of bandwidth out of all of our activities (See Table 2.2), we model $T(\text{Streaming TV}, P)$ using a normal distribution, in order to account for the wild variability that results in $B(F)$ that results from smaller shifts in the value of the function. In our model, we also utilized known values for the standard deviation in the amount of homework students (undergrad and in high school) receive each day, to account for the relatively strong impact that the amount of homework has on the daily bandwidth usage for students.

In order to manually calculate a reliable normal distribution for the amount of television that different age demographics streamed per day, we opted to create a discrete probability density function (PDF) for $T(\text{Streaming TV}, P)$, based on data which gives the probability of an individual to use online video for a certain amount of time [28]. This allowed us to model the random chance that a randomly selected individual would watch a certain hours of streamed television per week, by dividing the percent of individuals in the survey who reported they watched those hours of streamed television into the categories they fell into. This is summarized in the table below:

PDF for time spent per week streaming TV (in hours)					
1-2 hours	3-4 hours	5-7 hours	8-10 hours	11-19 hours	20-24 hours
0.1395	0.098	0.05867	0.04567	0.01278	0.0194

Table 2.1: Probability Density Function for time spent per week streaming television (in hours)[1]

[28]

This was simulated for an entire year using a Python script (which can be found in Code

---

[1]Values under the columns represent the same value for each hour (e.g., the value under 1-2 hours represents the equal probability of 1 hour and 2 hours of streamed television)

Appendix II). This simulation was run over 3000 episodes (meaning 3000 different times), and the results were plotted with *matplotlib* in a histogram, which is shown below
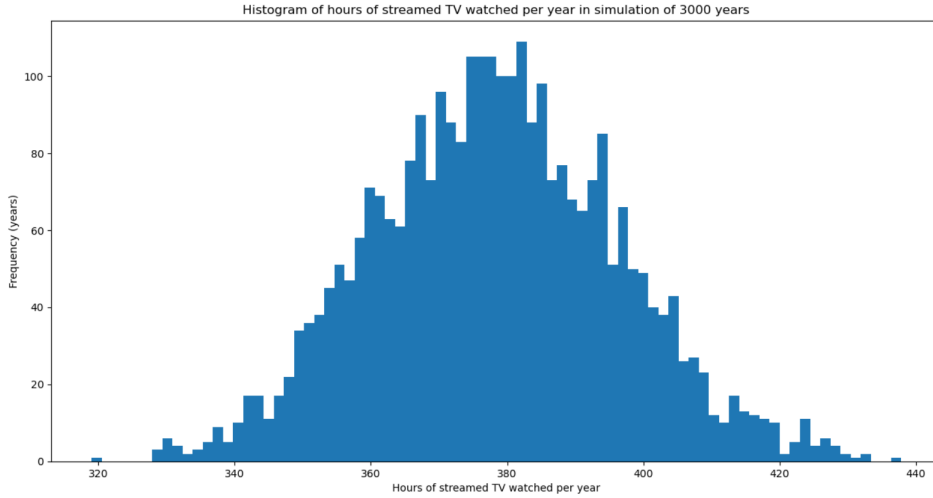

Fig 2.1: Histogram of Hours of Streamed TV Watched per Year

The script also outputted the exact numbers for the mean hours of streamed television per user over a year (378.26) and the standard deviation (17.99), which were used later in the model to calculate the data for the time spent per day on streaming television for the entire family. We then proportionally extrapolated this data to all age groups, as the study was centered around the 35-49 age group, using the provided data[1].

2.4   Results and Discussion

The following table details known values for $\mu(A)$ over all of our activities A:

Bandwidth Usage Rates by Activity (μ_0), in Megabits per hour									
Value	Streaming TV	Surfing	Streaming Videos	Social Media	Gaming	Music	Work-related Zoom	School-related Zoom	Homework
Bandwidth Usage	24012 [34]	119.88 [3]	18000 [16]	972 [38]	2160 [10]	576 [17]	13680 [9]	13680 [9]	119.88 [3]

Table 2.2: Data for Bandwidth Usage of Considered Internet Activities

For each person P, we collected known data values for $f(P, A)$. In order to sum these values in families, we either combined values into one or summed them separately, depending on whether or not they were independent (detailed below). For each independent addition, we added standard deviations in quadrature. Table 2.3 below details known values for $\sum_{P \in F} T(A, P)$ over all of our activities A, where F is Family 1, the couple in their 30s and their 3-year-old child.

Family 1 Data, in Hours per Day									
Value	Streaming TV	Surfing	Streaming Videos	Social Media	Gaming	Music	Work-related Zoom	School-related Zoom	Homework
Mean Time Per Day	.9976 $\pm$ 0.03378	6.724	1.633	2.586	2.346	0.4619	2.113	0	0

Table 2.3: Data for Bandwidth Usage of Family 1 [1][13][2][6][35]

For this table, we assumed that the couple in the family shared the same streaming television times, and that this television streaming rate was independent from that of their 3-year-old. We further assumed that, for every other activity, their data usage was independent. The next table details known values for $\sum_{P \in F} T(A, P)$ over all of our activities A, where F is Family 2, the retired woman in her 70s and her two school-aged grandchildren.

Family 2 Data, in Hours per Day									
Value	Streaming TV	Surfing	Streaming Videos	Social Media	Gaming	Music	Work-related Zoom	School-related Zoom	Homework
Mean Time Per Day	2.211 ± 0.0995	0.7378	0.9308	1.730	1.435	0.2635	0	1.755	5.570 ± 1.815

Table 2.4: Data for Bandwidth Usage of Family 2 [1][13][2][6][35]

For this table, similar to that of Table 2.3, we assumed that the woman's two grandchildren shared the same streaming television times, and that this television streaming rate was independent from that of their grandmother. We further assumed that, for every other activity, their data usage was independent. The next table details known values for $\sum_{P \in F} T(A, P)$ over all of our activities A, where F is Family 3, the three former $M3$ undergraduate students.

Family 3 Data (all numbers are in hours per day)									
Value	Streaming TV	Surfing	Streaming Videos	Social Media	Gaming	Music	Work-related Zoom	School-related Zoom	Homework
Mean Time Per Day	1.671 ± 0.0459	1.514	2.45	3.879	3.519	0.6929	5.143	7.286	7.286 ± 2.374

Table 2.5: Data for Bandwidth Usage of Family 3 [1][13][2][6][35]

For this table, as in the last two tables, we assumed that the three students shared television screen times. We further assumed that, for every other activity, their data usage was independent.

Using the results from these three tables, we then ran a Python simulation (which can be found in Code Appendix II) which simulated the internet activities conducted by the entire family. For each iteration of the simulation, we picked random values based on normal distributions for each of the activities in which we had standard deviations (streaming TV and homework), and, using (8), generated a single possibility for the total bandwidth usage for each family. By running our script over 1000 iterations, we were able to generate enough data points to create a rough mean and standard deviation for the total bandwidth requirements for each of the three families, depicted in Table 2.6:

Total Bandwidth Consumption By Family Over One Year (in megabits)		
Family (F)	Mean Usage (B(F))	Standard Deviation of Usage ($\sigma_{B(F)}$)
Family 1	33182955.822	15929.872
Family 2	36328818.447	44660.894
Family 3	97480617.470	21507.867

Table 2.6: Total Bandwidth Consumption By Family Over One Year of Internet Usage (in Megabits)

To see what covers their internet needs 90 and 99 percent of the time in the worst case scenarios, we can look at each Family's normal distribution and find the largest value of bandwidth usage. 90 percent of the time refers to the point which covers 90 percent of the distribution before it, which covers 40 percent of the data right above the mean in the normal distribution, and 50 percent of the data below the mean. This can be found with the *z-score* for an 80 percent confidence interval, which we know is 1.282 [12]. We also can apply this to the largest bandwidth usage 99 percent of the time, which again contains the bottom 50 percent of the normal distribution and the 49 percent of the values that are right above the mean, meaning that we can use the *z-score* for a 98 percent confidence interval, which we know is 2.326. From here, we can plug in our z-scores to get final values, which are summarized in a table

Total Bandwidth Required For a Family over a year (in Megabits)		
Family (F)	90 Percent Coverage ($1.282\sigma_{B(F)}$)	99 Percent Coverage ($2.326\sigma_{B(F)}$)
Family 1	4150.42	4152.50
Family 2	4548.26	4554.09
Family 3	12188.52	12191.33

Table 2.7: Total Bandwidth Required By Family for One Year of Internet Usage in Megabits for 90 and 99 % Coverage

These final numbers conclude the model and answer the question at heart, which is to determine the minimum amount of bandwidth required to ensure that the internet needs of all 3 Families are fulfilled.

2.5   Strengths and Weaknesses

2.5.1   Strengths

In our model, we made sure to account for the largest factors in determining the amount of bandwidth used. We also were able to split the model into categories and activities, which allowed for an in-depth calculation of the figures. Our final results are also very in-line with the current average amount of GBs of data used for a household (869.9 GB)[11]. Our model was also very cognizant of the current situation in which we live by factoring in both school-related Zoom and work-related Zoom, allowing for a more accurate results when calculating broadband usage during the COVID-19 pandemic.

2.5.2 Weaknesses

We did not consider the relationship of time between activities—for instance, if the undergraduate students got a lot of homework on one day, then the time they spend gaming or streaming television may decrease, and vice versa. We acknowledge our failure to consider the potential dependencies between each of our activities and that potential to negatively impact the robustness of our standard deviations. Another weakness to this model is that it does not account for seasonal variation, such as how students do not go to school (which is on Zoom) in the summer or during winter break.

3 Part 3: Mobilizing Mobile

3.1 Restatement of the Problem

We are tasked to do the following:

- Develop a model to predict the optimal distribution strategy of 5G cellular nodes.

- Apply our model to the three provided regions to demonstrate its flexibility.

We devise a system of scoring subregions within a given region to determine an optimal plan of cell tower distribution. We then use the data provided in the problem data set to consider the distribution of cellular nodes amongst the hypothetical subregions and regions.

3.2 Assumptions

1. The proportion of a region at each age is constant through its subregions

   - **Justification**: We don't have data on the age stratification of each subregion. However, in a defined geographical region, it is reasonable to presume this to be true.

2. Cell towers all have the same power and elevation

   - **Justification**: Helps to standardize the range of each cell tower, as elevation and power can impact the range [19] [40].

3. Other objects that disturb the transmission of cell towers are negligible

   - **Justification**: Assumed for the sake of brevity [40].

4. The cost of living is uniform in all the subregions of a region

   - **Justification**: The information for cost of living in each subregion is not provided. This is a reasonable estimate based on the mean cost of living index of the actual region in order to rank subregions.

5. The smart phone distribution is uniform amongst households in each subregion of a region.

   - **Justification**: The information for households with smart phone access in each subregion is not provided. This is a reasonable estimate based on the smart phone access proportion of the actual region in order to rank subregions.

6. All factors are equally weighted in ranking the subregions

   - **Justification**: There is no information about what factors are most important in the implementation of cell phone towers. Hence, we assume all factors to be equally weighted.

3.3   Model Development

We consider various factors that influence a region's usage of mobile broadband. For each subregion, each factor is assigned an equal weight and scaled from 1–100. These scales are in reference to the highest score in a category. The subregion with the highest score is the one prioritized for placing a cellular tower. Hence, with this method, we can rank each subregion from the three hypothetical regions in order from most to least prioritized for a cellular tower.

3.3.1   Parameters

1. **Cost of Living** (C_L). The index of the cost of living in a region.

2. **Median Wage** (W_M). The median wage of a region.

3. **Area** (A). The area of each subregion.

4. **Population** (P). The population of each subregion.

5. **Smart Phone Percentage** (S). The proportion of each population that has access to a smart phone.

6. **Importance** (I_r). The score assigned to each subregion for cell tower distribution.

7. **Adjusted Income** (W_a).

8. **Population Density** (σ).

9. **Number of Towers** (N).

10. **Proportion of Population between 15 and 64** (R)

3.3.2   Model Derivation

We consider five primary factors in the ranking of the distribution of cell phone towers: *adjusted income, population density, number of towers needed, proportion of population between 15 and 64,* and *smart phone percentage.* Per *Assumption 7,* we weight these factors equally, creating a generalized score for distribution.

First, we need to quantify our factors. To find adjusted income, we simply need to divide the median wage by the cost of living to find the effective wage in that region:

$$W_a = \frac{W_M}{C_L} \tag{9}$$

The adjusted wage is important because 5G technology is not cheap [25]. Those with more money are more likely to be financially capable and willing to back the establishment of 5G infrastructure. Next, we look at population density. We compute this by

$$\sigma = \frac{P}{A} \tag{10}$$

This is an important factor, as the more dense the population, the more people each individual tower can cover and, thus, the more effective each individual cell tower is. Indeed, cell towers are

optimally configured in regions with high population density [23]. Likewise, the number of towers that are needed to cover a subregion is important in deciding the distribution of towers [25][4]. We find that

$$N = \frac{A}{A_t} \tag{11}$$

where A_t represents the area covered by each tower, which is constant per *Assumption 2 and 3*. Hence, in relative scaling, we don't actually need to consider the value of A_t as we compare the number of towers amongst the subregions. For this factor, however, the lower the number of towers needed, the better. Hence, we take $\frac{1}{N}$ as the factor to be considered in our model.

We also consider S and R, the proportion of people with smart phones and between the ages of 15 and 64 in a region, respectively, as only these people can use 5G infrastructure. 5G is cellular [37], meaning that only those with smart phones can actually benefit from cellular tower infrastructure. We only look at those between 15 and 64 because this age range represents the proportion of people that benefit most from cellular tower distribution; teenagers on average get their phones from 12-13 years old, making those older than 15 likely to own phones so that they are be able to use 5G networks [18]. Likewise, the average age of retirement is around 64 years, after which most people are less reliant on cellular data networks, instead being able to stay home throughout the day [22].

We normalize each factor such that the maximum value amongst the subregions is set to 100, allowing us to assign an importance score to each subregion between 0 and 500.

3.4   Results and Discussion

We determine the following values for each factor for each subregion in **Table 3.1**:

Rankings by subregion						
Subregion	W_a	σ	N	S	R	Total
A1	18	4	8	79	84	193
A2	20	14	13	79	84	210
A3	30	15	15	79	84	223
A4	22	1	6	79	84	192
A5	20	27	28	79	84	238
A6	30	5	5	79	84	203
B1	58	9	3	99	77	246
B2	78	4	2	99	77	260
B3	100	2	2	99	77	280
B4	65	4	4	99	77	249
B5	62	3	1	99	77	242
B6	85	5	3	99	77	269
B7	76	2	1	99	77	255
C1	89	7	26	100	100	322
C2	40	33	71	100	100	344
C3	79	56	100	100	100	435
C4	58	100	42	100	100	400
C5	63	42	77	100	100	382
C6	86	60	32	100	100	378
C7	63	45	29	100	100	337

Table 3.1: Rankings of subregions within hypothetical regions A, B, and C [1]

Hence, for region A, subregions 5, 3, 2, 6, 1, and 4 should receive a cellular node in that order. For region B, subregions 3, 6, 2, 7, 4, 1, and 5 should receive a cellular node in that order. For region C, subregions 3, 4, 5, 6, 2, 7, and 1 should receive a cellular node in that order.

3.5 Strengths and Weaknesses

3.5.1 Strengths

Our model is well-rounded in considering various factors that influence usage of mobile broadband. This increases accuracy of the rankings and reflects real behaviors of populations despite being hypothetical regions. We are able to recognize both bandwidth usage and economic sustainability in our model to best determine where to implement cell towers. We are able to avoid the complexity in determining the optimal distribution pattern geographically by comparing relative area covered, whereupon the area covered by each cellular node is no longer important.

3.5.2 Weaknesses

Our model is limited to only ranking subregion importance in relevancy to one another. Since in these hypothetical regions we don't know how the population is spread out within each subregion,

we can not determine where to place cellular towers within a subregion. We also assume that people younger than age 15 and older than age 64 don't use 5G, when there could be some people in these age groups who do use mobile broadband. Thus, actual usage of mobile broadband in some places may increase, so the rankings of some subregions could also increase. Further, in the actual creation of cell tower hexagonal plots [4][41], various other factors contribute, including the geographic topography as well as tall buildings [40] [23], limiting the effectiveness of 5G networks in urban areas, something that our model fails to consider. We were limited in our data for smart phone access and age stratification in subregions, leading to less robustness in the ranking of subregions within the same region with regards to those two factors.

References

[1] MathWorks Math Modeling Challenge 2021. Defeating the Digital Divide data. https://m3challenge.siam.org/node/523.

[2] K. Anderton. Week 6 household pulse survey: June 4-June 9. https://www.forbes.com/sites/kevinanderton/2019/03/21/research-report-shows-how-much-time-we-spend-gaming-infographic/?sh=54549ecb3e07.

[3] AT&T. Internet data calculator. https://www.att.com/support/data-calculator/.

[4] C. Beckmen. The evolution of base station antennas for mobile communications. https://www.researchgate.net/figure/Hexagonal-cells-in-a-GSM-system-created-by-placing-the-base-stations-at-the-intersection_fig4_4293285.

[5] Broadbandsearch.net. How do US internet prices compare to the rest of the world? https://www.broadbandsearch.net/blog/internet-costs-compared-worldwide#post-navigation-3.

[6] U.S. Census Bureau. Research report shows how much time we spend gaming. https://www.census.gov/data/tables/2020/demo/hhp/hhp6.html#techdoc.

[7] U.S. Census Bureau. U.S. and world population clock*. https://www.census.gov/popclock/.

[8] J. Caboz. Compared: SA video calls on Zoom, Skype, Teams, Hangouts and Whatsapp. https://www.businessinsider.co.za/heres-why-your-video-conference-app-keeps-acting-up-recommended-bandwidth-speeds-of-zoom-microsoft-teams-skype-google-hangouts-met-and-whatsapp-2020-4.

[9] Zoom Help Center. System requirements for Windows, macOS, and Linux. https://support.zoom.us/hc/en-us/articles/201362023-System-requirements-for-Windows-macOS-and-Linux.

[10] A. Choros. How much data do you need for gaming? https://www.whistleout.com.au/Broadband/Guides/how-much-data-do-I-need-for-online-gaming.

[11] CISCO. VNI complete forecast highlights. https://www.cisco.com/c/dam/m/en_us/solutions/service-provider/vni-forecast-highlights/pdf/United_States_2021_Forecast_Highlights.pdf.

[12] Richland Community College. Stats: Introduction to estimation. https://people.richland.edu/james/lecture/m170/ch08-int.html.

[13] G. Enoch. Nielsen - the total audience report — q3 2015. https://www.nielsen.com/wp-content/uploads/sites/3/2019/04/total-audience-report-q3-2015.pdf.

[14] J. Fernando. Law of supply and demand. https://www.investopedia.com/terms/l/law-of-supply-demand.asp.

[15] Market Analysis Forecast. Submarine cables market size, share & trends analysis report by application, by voltage, by end user, by component, by offerings, by region, and segment forecasts, 2020 - 2027. https://www.grandviewresearch.com/industry-analysis/submarine-cables-market.

[16] Youtube Help. Youtube system requirements.
https://support.google.com/youtube/answer/78358?hl=en.

[17] J. Hildenbrand. How much mobile data does streaming media use?
https://www.androidcentral.com/how-much-data-does-streaming-media-use.

[18] L. Hutton. 6 questions to ask before getting your child a cell phone.
https://www.familyeducation.com/life/kids-cell-phones/6-questions-ask-getting-your-child-cell-phone.

[19] H. Kanchwala. What are cell towers and how do they work?
https://www.scienceabc.com/innovation/cell-tower-work.html#the-range-of-a-cell-tower.

[20] W. Kenton. Aggregate demand. https://www.investopedia.com/terms/a/aggregatedemand.asp.

[21] S. Kraus. US & UK ecommerce markets: Similarities and differences.
https://www.linkedin.com/pulse/us-uk-ecommerce-markets-similarities-differences-stephen-kraus.

[22] D. Kurt. When to retire: The pros and cons of different ages.
https://www.investopedia.com/retirement/when-should-i-retire-pros-and-cons-different-ages/.

[23] Millman Land. What is a cell tower and how does a cell tower work?
https://millmanland.com/company-news/what-is-a-cell-tower-and-how-does-a-cell-tower-work/.

[24] Macrotrends.net. Bandwidth profit margin 2016-2020.
https://www.macrotrends.net/stocks/charts/BAND/bandwidth/profit-margins.

[25] A. Villas-Boas, M. Yan Huang. Why you shouldn't get too excited about 5G yet. https://www.businessinsider.com/5g-high-speed-internet-cellular-network-issues-switch-2019-4.

[26] J. Miller. And just when you think bandwidth prices can't go lower...
https://blog.telegeography.com/and-just-when-you-think-bandwidth-prices-cant-go-lower.

[27] I. Morris. UK vs US: Who has the best tech? https://www.cnet.com/news/uk-vs-us-who-has-the-best-tech/.

[28] Limelight Networks. Market research: The state of online video 2019.
https://www.limelight.com/resources/white-paper/state-of-online-video-2019/.

[29] S. O'Dea. number of fixed broadband subscriptions in the US from 2000-2018 in millions.
https://www.statista.com/statistics/183614/us-households-with-broadband-internet-access-since-2009/.

[30] University of Cape Town. Channel capacity.
https://www.dip.ee.uct.ac.za/ñicolls/lectures/eee482f/04_chancap_2up.pdf.

[31] Office of Communications. Pricing trends for communications services in the UK.
https://www.ofcom.org.uk/_data/assets/pdf_file/0022/189112/pricing-trends-
communication-services-report.pdf.

[32] U. of Manchester. Cs3282: Digital communications '05-'06. section 7: The Shannon-Hartley
theorem. http://www.cs.man.ac.uk/b̃arry/mydocs/CS3282/Notes/DC06_7.pdf.

[33] C. Park. The cost of connectivity in West Virginia.
https://www.newamerica.org/oti/reports/cost-connectivity-west-virginia/.

[34] N. Perry. How much data does Netflix use? https://www.digitaltrends.com/movies/how-
much-data-does-netflix-use/.

[35] V. Rideout, M. Robb. The common sense census: Media use by tweens and teens.
https://www.commonsensemedia.org/sites/default/files/uploads/research/2019-census-8-to-
18-full-report-updated.pdf.

[36] M. Rosoff. Netflix and Youtube are America's biggest bandwidth hogs.
https://www.businessinsider.com/which-services-use-the-most-bandwidth-2015-12.

[37] S. Segan. What is 5G? https://www.pcmag.com/news/what-is-5g.

[38] J. A. Thales. How much data does social media and streaming apps use when roam-
ing? https://justaskthales.com/en/how-much-data-does-social-media-and-streaming-apps-
use-when-roaming/.

[39] T. Wheeler. 5 steps to get the internet to all Americans.
https://www.brookings.edu/research/5-steps-to-get-the-internet-to-all-americans/.

[40] WilsonPro. How building materials affect cell phone signal booster performance.
https://www.wilsonpro.com/blog/how-building-materials-affect-cell-phone-signal-booster-
performance.

[41] X. Yang. Performance analysis of hexagonal cellular networks in fading channels.
https://onlinelibrary.wiley.com/doi/full/10.1002/wcm.2573.

4 Code Appendix I: MATLAB

```
1 x = 2000:2030
2 y = 330.01./(1+6.08569*exp(-0.265072 * (x-2000)))
3 x1 = 2000:2018
4 y1 = [19.44 35.17 54.66 76.27 102.69 140.66 166.07 197.14 212.06 219.93 232.38
      242.83 254.35 264.03 268.92 281.02 290.64 297.46 304.01]
5 plot(x, y)
6 hold on
7 plot (x1, y1, 'r')
8 xlabel 'Year'
9 ylabel 'Number of People'
10 title 'Number of People with Internet Access From 2000-2030'
```

Listing 1: MATLAB code used to create Figure 1.1 in Part I: The Cost of Connectivity

5   Code Appendix II: Python

```python
import numpy as np
import matplotlib.pyplot as plt
from tqdm import tqdm

# Each activity has a certain amount of megabit/hour bandwidth it uses, which is a
    constant listed below
streamingTVBandwidth = 24012
streamingVidBandwidth = 18000
workZoomBandwidth = 13680
schoolZoomBandwidth = workZoomBandwidth # the bandwidth used for school and work
    zoom can reasonably be assumed
                                        # to be the same as per TODO insert
    assumption
surfingBandwidth = 119.88
homeworkBandwidth = surfingBandwidth # the bandwidth used for surfing can
    reasonably be assumed to be the same
                                        # as the bandwidth used for homework as per
    TODO insert assumption
socialMediaBandwidth = 972
gamingBandwidth = 2160
musicBandwidth = 576

# we now proceed to define a class, Family, to assist in our code's execution


class Family:
    def __init__(self, name):
        self.identity = name
        # Name of the family is kept as a variable for cleanliness and
    identification.

        self.streamingTVTA = 0
        self.homeworkTA = 0
        # Streaming TV and homework are variable T(A) values, so they are set as
    separate variables which are to
        # be computed by numpy later. They are initialized to 0 in the class
    construction method.
        self.streamingTVTAMean = 0
        self.streamingTVTASigma = 0
        self.homeworkTAMean = 0
        self.homeworkTASigma = 0


        self.surfingTA = 0
        self.streamingVidTA = 0
        self.socialMediaTA = 0
        self.gamingTA = 0
        self.musicTA = 0
        self.workZoomTA = 0
        self.schoolZoomTA = 0
        # Surfing the Internet, Streaming Video, Social Media, Gaming, Music, Work
```

```
      -related Zoom , and School -related
44        # Zoom are all activities that are constant numbers for a person , yet they
       are defined as variables for the
45        # simplicity and cleanliness of the code.
46
47        self.entertainmentActivitiesTA = [self.streamingTVTA , self.surfingTA , self
      .streamingVidTA ,
48                                          self.socialMediaTA , self.gamingTA , self.
      musicTA]
49        # Streaming TV , Surfing the Internet , Streaming Video , Social Media ,
      Gaming , and Music are all activities
50        # that fall in the entertainment category.
51
52        self.workActivitiesTA = [self.workZoomTA]
53        # Work -related Zoom is the only activity that falls into the work category
      .
54
55        self.schoolActivitiesTA = [self.schoolZoomTA , self.homeworkTA]
56        # School -related Zoom and homework are the only activities that fall into
      the school category.
57
58    def generateChangingVariables(self):
59        # Streaming TV and homework are the only activities that need to be
      periodically regenerated due to their
60        # high variability , as mentioned earlier in the code.
61
62        self.streamingTVTA = np.random.normal(loc=self.streamingTVTAMean , scale=
      self.streamingTVTASigma)
63        # generate the T(A) value for streaming TV with np.random.normal to get a
      random value from the
64        # normal distribution representing the T(A) values the "loc" argument
      represents the mean and the
65        # "scale" argument represents the standard deviation.
66
67        self.homeworkTA = np.random.normal(loc=self.homeworkTAMean , scale=self.
      homeworkTASigma)
68        # generate the T(A) value for homework with np.random.normal to get a
      random value from the
69        # normal distribution representing the T(A) values the "loc" argument
      represents the mean and the
70        # "scale" argument represents the standard deviation.
71
72    def generateChangingParameters(self , streamingTVTAMean , streamingTVTASigma ,
      homeworkTAMean , homeworkTASigma):
73        self.streamingTVTAMean = streamingTVTAMean
74        self.streamingTVTASigma = streamingTVTASigma
75        self.homeworkTAMean = homeworkTAMean
76        self.homeworkTASigma = homeworkTASigma
77
78    def generateConstantVariables(self , surfingTA , streamingVidTA , socialMediaTA ,
      gamingTA , musicTA , workZoomTA ,
79                                  schoolZoomTA):
80        self.surfingTA = surfingTA
```

```
81              self.streamingVidTA = streamingVidTA
82              self.socialMediaTA = socialMediaTA
83              self.gamingTA = gamingTA
84              self.musicTA = musicTA
85              self.workZoomTA = workZoomTA
86              self.schoolZoomTA = schoolZoomTA
87
88          def generateFamilyOutput(self):
89              totalBandwidthUsage = []
90              # totalHomeworkBandwidth = []
91              for day in range(0, 365):  # iterating over all the days as the T(A)
        values are for each day
92                  thisDayBandwidthUsage = 0
93                  self.generateChangingVariables()
94                  # we first generate the variables that need to be regenerated on each
        day
95
96                  # now we can proceed to add to the bandwidth usage of each day the
        constant variables (T(A) does not
97                  # vary day by day)
98                  thisDayBandwidthUsage += self.surfingTA * surfingBandwidth
99                  thisDayBandwidthUsage += self.streamingVidTA * streamingVidBandwidth
100                 thisDayBandwidthUsage += self.socialMediaTA * socialMediaBandwidth
101                 thisDayBandwidthUsage += self.gamingTA * gamingBandwidth
102                 thisDayBandwidthUsage += self.musicTA * musicBandwidth
103                 thisDayBandwidthUsage += self.workZoomTA * workZoomBandwidth
104                 thisDayBandwidthUsage += self.schoolZoomTA * schoolZoomBandwidth
105
106                 # now we can proceed to add to the bandwidth usage of each day the
        changing variables
107                 thisDayBandwidthUsage += self.streamingTVTA * streamingTVBandwidth
108                 thisDayBandwidthUsage += self.homeworkTA * homeworkBandwidth
109
110                 # finally, we can add today's bandwidth usage to a list that maintains
         this value over the entire year
111                 totalBandwidthUsage.append(thisDayBandwidthUsage)
112
113             return totalBandwidthUsage
114
115
116 # Scenario 1
117 # Couple in their 30s
118 familyTeacher = Family("familyTeacher")
119 familyTeacher.generateChangingParameters(.9976437104, 0.0337808, 0, 0)  # this is
        the data for the streaming TV and homework
120                                                                     # activities
        for the first family
121 familyTeacher.generateConstantVariables(6.723804714, 1.633342857, 2.585714286,
        2.345714286, 0.4619142857,
122                                         2.11285, 0)
123 familyTeacherAverages = []
124 familyTeacherSums = []
125 for episode in tqdm(range(0, 1000)):  # we run 1000 episodes (years) for each
```

```
              family to get a reasonable mean and
126                                              # standard deviation. tqdm is used to track
           progress in the episodes
127         familyDailyOutputList = familyTeacher.generateFamilyOutput()  # generate the
           output of the family
128         familyTeacherAverages.append(np.average(familyDailyOutputList))  # add this to
            an array maintaining the averages
129         familyTeacherSums.append(np.sum(familyDailyOutputList))  # add this to an
           array maintaining the yearly sums
130
131 print(np.mean(familyTeacherAverages))  # np.mean gets the mean of the average
       daily bandwidth usage in a year
132 print(np.std(familyTeacherAverages))  # np.std gets the standard deviation
133 print()
134 print(np.mean(familyTeacherSums))  # np.mean gets the mean of the total yearly
       bandwidth usage
135 print(np.std(familyTeacherSums))  # np.std gets the standard deviation
136
137
138 # Scenario 2
139 # Old grandma with two grandchildren
140 familyGrandma = Family("familyGrandma")
141 familyGrandma.generateChangingParameters(2.211, 0.0995, 5.57, 1.815)  # this is
       the data for the streaming TV and homework
142                                                      # activities
       for the second family
143 familyGrandma.generateConstantVariables(0.7377530612, 0.9308435843, 1.729932352,
144                                          1.435102041, 0.2635065122, 0, 1.755102041)
145 familyGrandmaAverages = []
146 familyGrandmaSums = []
147 for episode in tqdm(range(0, 1000)):  # we run 1000 episodes (years) for each
       family to get a reasonable mean and
148                                              # standard deviation. tqdm is used to track
           progress in the episodes
149         familyDailyOutputList = familyGrandma.generateFamilyOutput()  # generate the
           output of the family
150         familyGrandmaAverages.append(np.average(familyDailyOutputList))  # add this to
            an array maintaining the averages
151         familyGrandmaSums.append(np.sum(familyDailyOutputList))  # add this to an
           array maintaining the yearly sums
152
153 print(np.mean(familyGrandmaAverages))  # np.mean gets the mean of the average
       daily bandwidth usage in a year
154 print(np.std(familyGrandmaAverages))  # np.std gets the standard deviation
155 print()
156 print(np.mean(familyGrandmaSums))  # np.mean gets the mean of the total yearly
       bandwidth usage
157 print(np.std(familyGrandmaSums))  # np.std gets the standard deviation
158
159 # Scenario 3
160 # Undergrads with part time job
161 familyUndergrads = Family("familyUndergrads")
162 familyUndergrads.generateChangingParameters(1.671, 0.04588748, 7.285714286,
```

```
      2.374070275)   # this is the data for the streaming TV and homework
163                                                                       # activities
      for the third family
164 familyUndergrads.generateConstantVariables(1.514271429, 2.450014286, 3.8788571429,
165                                              3.518571429, 0.6928714286, 5.142857143,
       7.285714286)
166 familyUndergradsAverages = []
167 familyUndergradsSums = []
168 for episode in tqdm(range(0, 1000)):  # we run 1000 episodes (years) for each
      family to get a reasonable mean and
169                                         # standard deviation. tqdm is used to track
      progress in the episodes
170     familyDailyOutputList = familyUndergrads.generateFamilyOutput()  # generate
      the output of the family
171     familyUndergradsAverages.append(np.average(familyDailyOutputList))  # add this
       to an array maintaining the averages
172     familyUndergradsSums.append(np.sum(familyDailyOutputList))  # add this to an
      array maintaining the yearly sums
173
174 print(np.mean(familyUndergradsAverages))  # np.mean gets the mean of the average
      daily bandwidth usage in a year
175 print(np.std(familyUndergradsAverages))  # np.std gets the standard deviation
176 print()
177 print(np.mean(familyUndergradsSums))  # np.mean gets the mean of the total yearly
      bandwidth usage
178 print(np.std(familyUndergradsSums))  # np.std gets the standard deviation
```

Listing 2: Python 3 code used to determine mean and standard deviation of yearly bandwidth usage by family for Part II: Bit by Bit

```python
1  import numpy as np
2  import matplotlib.pyplot as plt
3  import time
4
5  streamingTVPDFValues = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17,
       18, 19, 20, 21, 22, 23, 24]
6  # values for the probability density function of TV streaming
7
8  for val in range(0, len(streamingTVPDFValues)):
9      streamingTVPDFValues[val] = streamingTVPDFValues[val]/7
10
11
12 streamingTVPDFProbs = [0.1395, 0.1395, 0.098, 0.098, 0.176/3, 0.176/3, 0.176/3,
      0.137/3, 0.137/3, 0.137/3, 0.115/9,
13                          0.115/9, 0.115/9, 0.115/9, 0.115/9, 0.115/9, 0.115/9,
      0.115/9, 0.115/9, 0.0194, 0.0194,
14                          0.0194, 0.0194, 0.0194]
15 # probabilities of each value in the pdf of TV streaming
16
17 print(np.sum(streamingTVPDFProbs))
18 streamTimeEpisodeData = []
19
20 for episode in range(0, 3000):
21     totalStreamTime = 0
22     for day in range(0, 365):
23         streamTime = np.random.choice(a=streamingTVPDFValues, size=None, replace=
      True, p=streamingTVPDFProbs)
24         # print(f"Day {day+1}:\nStreamed: {streamTime}\nSchool: {schoolWorkTime}\
      nWork: {workTime}\n")
25         totalStreamTime += streamTime
26     streamTimeEpisodeData.append(totalStreamTime)
27
28 print(np.mean(streamTimeEpisodeData))
29 print(np.std(streamTimeEpisodeData))
30
31 plt.hist(streamTimeEpisodeData, bins=80)
32 plt.title('Histogram of hours of streamed TV watched per year in simulation of
      3000 years')
33 plt.xlabel('Hours of streamed TV watched per year')
34 plt.ylabel('Frequency (years)')
35
36 plt.show()
```

Listing 3: Python 3 code used to calculate standard deviations of time spent per day streaming TV through the internet in Part II: Bit by Bit