MathWorks Math Modeling Challenge 2021

High Technology High School

Team # 14694 Lincroft, New Jersey

Coach: Raymond Eng

Students: Adithya Balachandran, Lasya Balachandran, David Chang, Alexander Postovskiy, Hazem Zaky

M3 Challenge RUNNER UP—$17,000 Team Prize

JUDGES' COMMENTS

Specifically for Team # 14694 —Submitted at the Close of Triage Judging:

COMMENT 1: Good that you provided some results in your summary. Summaries are the first thing read and set an indication of what is to come. Model 1 good, but present and discuss R^2. Need to include all 10 years, not just a few. Why a normal distribution for Model 2—did you analyze the data? The idea of SNR is excellent but you misused it. Overall better than average.

COMMENT 2: The choice of a logistic curve for modeling internet bandwidth in problem 1 is an interesting choice, but one that could benefit from closer scrutiny. Tables, such as 2.2.1, would benefit from labeling the units of numbers provided. The extrapolation to a logistic curve in Figure 2.2.1 is curious; it seems the data would better support a linear relationship which would be more conservative. The sensitivity analysis to provide 80% confidence intervals was original and fresh, for parts 1 and 3.

Defeating the Digital Divide: Internet Costs, Needs, and Optimal Planning

Executive Summary

In a world increasingly reliant on high-speed Internet, disadvantages faced by those who lack sufficient access have emerged as a major issue. Given the worldwide effects of the COVID-19 pandemic, the Internet is now even more essential for a variety of activities such as attending school remotely, safely accessing healthcare, working from home, civic participation, and access to information [1, 2, 3]. For the foreseeable future, there are several options for connecting to the Internet available for solving this crisis—from wired to mobile and satellite—but the viability of each remains uncertain. With the need to analyze Internet costs and bandwidth requirements for various communities, as well as the optimization of a potential connection plan, this paper provides mathematically founded insights on this issue.

Since one crucial factor for providing high-speed Internet to those in need is its cost, we first analyzed the cost per unit bandwidth (in Mbps, Megabits per second) for representative US and UK consumers. We focused on wired, mobile, and satellite Internet, both in the US and the UK. Within these categories, Internet bandwidth was modeled by fitting the given data [8] to a logistic curve. After multiplying the current cost of Internet access per unit bandwidth by the ratio of current bandwidth to predicted future bandwidth, our model was able to project the average monthly cost per unit bandwidth of Internet for the previously mentioned categories in 2025 and 2030.

We then developed a model to determine the average bandwidth required for a given household. This was calculated by determining how a household's income level and its denizens' ages affected the number of hours spent on different activities using the Internet. We characterized the average bandwidth using a normal distribution to determine the minimum bandwidth that would cover 90% and 99% Internet availability. The first household, containing a couple in their early 30's and a 3-year-old child, required 26.1 Mbps and 31.2 Mbps, respectively. The second household, containing an elderly woman in her 70's and two grandchildren, required 16.4 Mbps and 19.8 Mbps, respectively. The third household, containing three former M3 Challenge participants in college and working part-time, required 28.8 Mbps and 36.9 Mbps, respectively.

Finally, we optimized the number of mobile cellular towers for every subregion in the three regions provided [8]. After classifying cellular towers as low-, medium-, or high-band, we developed a set of linear inequalities that take into account both inherent properties of each tower type as well as demographics specific to individual subregions. We were able to calculate effective areas for the cellular towers using the COST Hata Model and bandwidth needs with the model from Part II. Optimization was performed using integer programming techniques, which successfully minimized the number of cellular towers for each subregion.

With the increasing need for high-speed Internet due to the rising popularity of social media and online education and work resulting from the COVID-19 pandemic, many people, particularly those in rural and low income areas, continue to struggle with connectivity issues. We believe the models outlined in our paper provide valuable insight that will help address the technical, logical, and economic challenges of providing adequate Internet access worldwide.

Contents

1 Introduction

This section delineates the components of the modeling problem and their objectives. Global assumptions applying to the entire modeling process are also listed.

1.1 Restatement of the Problem

The problem we are tasked with addressing is as follows:

1. Build a mathematical model that predicts the cost per unit of bandwidth in dollars or pounds per Megabit per second (Mbps) over the next 10 years for consumers in the United States and the United Kingdom.

2. Create a mathematical model that predicts the Internet access required over the course of a year for the following three households: a couple in their early 30's (one is looking for work and the other is a teacher) with a 3-year-old child; a retired woman in her 70's who cares for two school-aged grandchildren twice a week; and three former M3 Challenge participants sharing an off-campus apartment while they complete their undergraduate degrees full-time and work part-time. Apply this model to determine the minimum amount of required bandwidth for 90% and 99% of the households' Internet needs.

3. Develop a model that will create an optimal plan for placing cellular nodes providing 4G and 5G access in three given hypothetical regions, taking into account the population, demographic data, and bandwidth needed for the regions.

2 Part I: The Cost of Connectivity

High-speed Internet is necessary for attending school remotely, safely accessing healthcare, working from home, civic participation, and information access [1, 2, 3]. However, ensuring that everyone, specifically those in rural and low-income areas, has sufficient access can be an economic challenge [4]. This section outlines a mathematical model for predicting the cost per unit of bandwidth in dollars or pounds per Megabit per second (Mbps) over the next 10 years for consumers in the United States and the United Kingdom.

2.1 Assumptions

1. *No major disruptive technology will be introduced within the next 10 years.* Currently, the most recent technology implemented in the US and UK is 5G in 2019. 6G, which is estimated to be able to obtain speeds up to 1 terabyte per second, is expected to be commercially available only after 2030 [5]. For the scope of the next 10 years, we will not include this in our model. In addition, entirely novel Internet connection technologies cannot be predicted and are necessarily excluded.

2. *The only sources of Internet are wired—including DSL, cable, and fiber-optic—mobile, and satellite.* All other sources of Internet connection are either obsolete or negligible.

3. *Most wired Internet service providers essentially have a monopoly over their consumers.* For example, Comcast and Charter are the only available Internet service providers for more than 47 million people in the United States. For millions more, alternatives provide slower speeds at higher costs [6].

4. *The demand for high-speed Internet does not significantly affect the cost per Megabit per second.* During early 2020, Internet usage rose by 30-50% in parts of the United States, but prices were not significantly affected [7].

5. *The level of infrastructure and average Internet speed will not affect the cost of an Internet plan.* The change in the price of the Internet plan can be assumed to be negligible, as we are determining the cost per unit bandwidth. The change in Internet speed will be a much more significant effect, as is seen in the provided data [8].

6. *Satellite data speed will be roughly similar in the US and the UK.* We were unable to find data for satellite Internet speed in the UK; however, since geography should not significantly affect satellite coverage, this assumption is a reasonable simplification.

2.2 Model Development

The primary effect on the cost of Internet per unit bandwidth in the near future will be the continuing increase in Internet speed. This is much more significant than a change in pricing of Internet plans, as Internet service providers largely have monopolies over their operating regions and can thus set costs as they desire. Speed increases of an order of magnitude can occur, but prices not change [8]. We model average Internet speed over time using a logistic curve.

$$S = L\left(\frac{1}{1+e^{-k(t-t_0)}}\right) + S_0. \quad (1)$$

L is the asymptotic maximum of the curve—taken at 1000 Mbps, the limit of current technology such as fiber-optic cables [13]. k is a constant associated with the magnitude of exponential growth. t_0 is the midpoint of the logistic curve—the point in time of greatest speed increase, which we project is in the near future. S_0 is an additive constant used to appropriately account for initial values. We first fit a logistic curve to the average Internet speeds recorded from 2009 to 2017 for US and UK wired and mobile Internet [8]. We obtained limited amounts of data on satellite speeds in the US from 1997 to 2009 [10], and added a data point for 2020, obtained by taking the average measured speeds over the last 12 months of the two largest satellite Internet companies in the US, HughesNet and Viasat [14, 15]. The values returned for k, t_0, and c are included in Tables 2.2.1 and 2.2.2, as well as the R^2 values for each of the fits.

Table 2.2.1: US Internet Speed Constants Determined from Curve Fitting

Variable	Wired	Mobile	Satellite
k	0.225	0.455	0.214
t_0	2034.9	2027.3	2040.5
S_0	1.014	0.982	0.133
R^2	0.997	0.878	0.999

Table 2.2.2: UK Internet Speed Constants Determined from Curve Fitting

Variable	Wired	Mobile	Satellite
k	0.184	0.030	0.214
t_0	2038.6	2070.7	2040.5
S_0	-1.249	-144.5	0.133
R^2	0.991	0.804	0.999

As can be seen, all of the R^2 values were at least 0.8, with the majority being at least 0.99. This indicates that our logistic growth model is a strong match for the data provided. One reason why the R^2 value for the UK mobile Internet data is a little lower than the rest of the values is that the provided UK data has a jump in 2015 when LTE-A was introduced in the UK.

Below are graphs of the regression for the average US and UK wired Internet speeds.

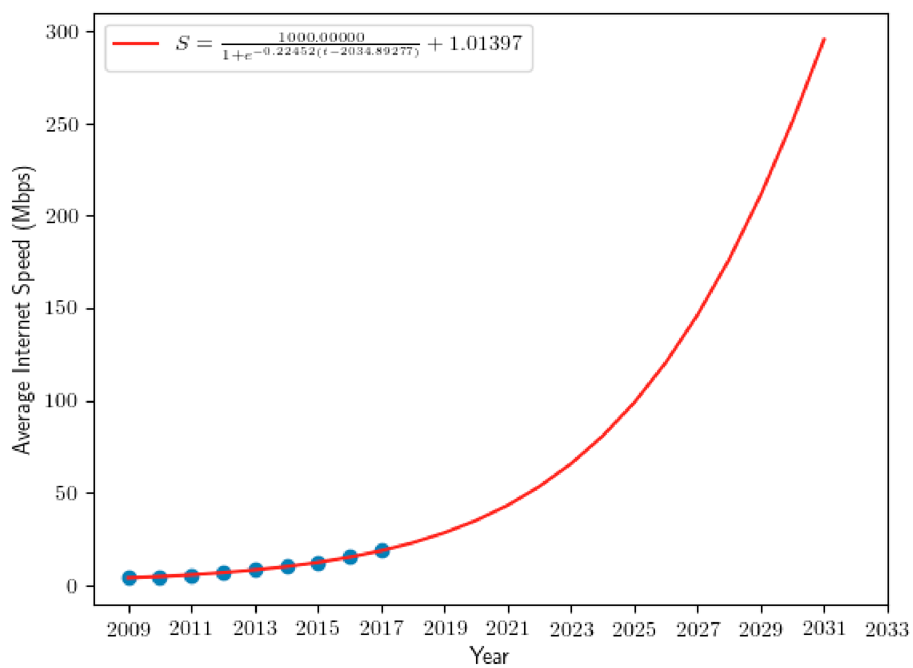

Figure 2.2.1: Regression of Average US Wired Internet Speeds

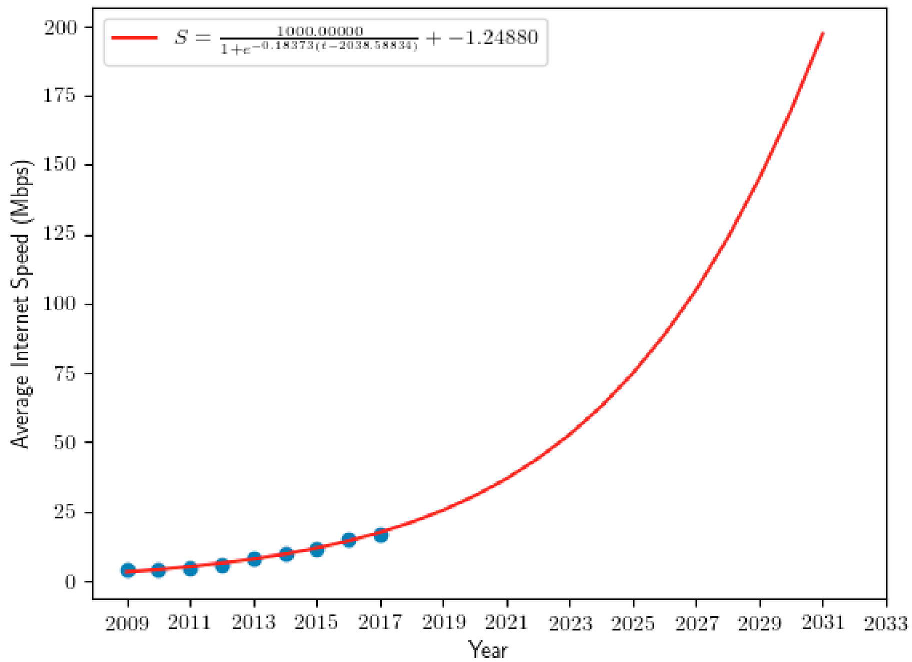

Figure 2.2.2: Regression of Average UK Wired Internet Speeds

For predicting the cost per unit bandwidth in the near future, we assume that the cost of each plan is approximately constant. Therefore,

$$C = \frac{A}{S}, \tag{2}$$

where C denotes the cost per bandwidth, S denotes the Internet speed recorded, and A is a constant determined by initial values. We use this equation for each of the Internet sources (wired, mobile, and satellite). After finding the current costs and predicting the Internet speeds over the next 10 years, we use Equation (2) to determine the cost per unit bandwidth for each of the Internet sources. Finally, we determine the average cost as the following:

$$C_{\text{avg}} = \frac{C_{\text{wired}} + C_{\text{mobile}} + C_{\text{satellite}}}{3}. \tag{3}$$

The current price per unit bandwidth can be multiplied by the ratio of current bandwidth to predicted future bandwidth to obtain the predicted future price per unit bandwidth. We can take an average of the three Internet types to obtain a final value. An unweighted average is used because with the great variability of regional costs, no single metric is able to be more than a signifier of a global trend. It is thus important to consider all forms of available Internet, as well as an aggregate metric.

2.3 Results

Using the method described above and the speeds found in Table 2.2.1 and 2.2.2, we determine the following average monthly costs.

Table 2.3.1: Average Monthly Cost per Mbps of Internet Types in the US

Year	Wired	Mobile	Satellite	Average
2020	$1.96 [16]	$1.56 [17]	$8.05 [18]	$3.86
2025	$0.70	$0.21	$2.85	$1.25
2030	$0.27	$0.07	$1.04	$0.46

Table 2.3.2: Average Monthly Cost per Mbps of Internet Types in the UK

Year	Wired	Mobile	Satellite	Average
2019	$1.37 [8]	$0.63 [8]	$9.93	$3.98
2025	$0.46	$0.33	$2.85	$1.21
2030	$0.20	$0.23	$1.04	$0.49

Wired and mobile data for 2019 was obtained by averaging provided data points. Prices were converted from GBP to USD with a conversion factor of 1.39 dollars/pound, the current exchange rate.

Internet costs are decreasing overall, in both countries and across all three mechanisms of delivery, a welcome sign in an increasingly digital world. However, before one gets too heartened by five-fold cost reductions it is important to consider that the actual prices to the consumers are not likely to fall greatly. Increasing amounts of data are being transmitted; Internet service providers will increase bandwidth availability at a price tier but not necessarily lower prices. Nevertheless, these strong downward trends do indicate, however, that the infrastructures of the United States and the United Kingdom are prepared to enter into the third decade of the digital millennium.

2.4 Sensitivity Analysis

Table 2.4.1: Effect of Adjusting Variables on 2030 Costs

Variable	US +10%	US −10%	UK +10%	UK −10%
k wired	2.0%	−1.3%	2.2%	1.5%
k mobile	0.0%	−0.3%	5.3%	3.2%
k satellite	17.1%	−13.4%	16.4%	12.6%
t_0 wired	11.9%	−6.3%	8.6%	4.6%
t_0 mobile	1.6%	−0.5%	9.2%	4.3%
t_0 satellite	105.2%	−42.1%	98.7%	38.6%
S_0 wired	0.0%	0.0%	0.0%	0.0%
S_0 mobile	0.0%	0.0%	3.4%	2.2%
S_0 satellite	0.0%	0.0%	0.0%	0.0%

We conducted a sensitivity analysis on the variables k, t_0, and S_0 used in our model by changing each constant by +10% and −10%. We then calculated the resulting change in the average cost of Internet speed. As can be seen in Table 2.4.1, our model is resilient to changes in the variables k, t_0, and S_0, which were determined by curve-fitting. The major abnormality occurs in the variable t_0 for the satellite Internet. This is due to the remoteness of its normal value: at twenty years into the future it is far away from our data points, and thus more sensitive to changes. Satellite costs are also higher, leading to a greater share of the overall average. Furthermore, t_0 is the most affected variable because it denotes the center point of the logistic curve, which is where the rate of change is highest. Ultimately, the model is flexible, and one is able to look at its components to figure out the underlying changes.

2.5 Strengths and Weaknesses

Our logistic model fits the data well as can be seen from the R^2 values from Tables 2.3.1 and 2.3.1. Furthermore, our model accounts for the various types of Internet, including wired, mobile, and satellite. We account for the expected increase in satellite Internet speeds. This is particularly useful because satellite Internet can provide additional access to users in rural communities who do not have other low-cost and high-speed options. Finally, our model is robust and resilient to changes in the determining variables as can be seen from our sensitivity analysis.

A weakness is that our model does not appear to be effective at predicting the speeds more than 10 years into the future. This is because new technology in the form of 6G and further innovations in existing technologies (fiber-optic, satellite) will be expected after 2030 [5].

3 Part II: Bit by Bit

In addition to being an economic burden, sufficient high-speed Internet can be a technical and logistical challenge. In this section, we formulate a model that predicts a given household's Internet need over the course of a year and apply the model to the following five households: a couple in their early 30's (one is looking for work and the other is a teacher) with a 3-year-old child; a retired woman in her 70's who cares for two school-aged grandchildren twice a week; and three former M3 Challenge participants sharing an off-campus apartment while they complete their undergraduate degrees full-time and work part-time. Using the model, we determined the minimum amount of required bandwidth that would cover 90% and 99% of the households' Internet needs.

3.1 Assumptions

1. *Bandwidth follows a normal distribution.* Following the methodology used for dimensioning network links by Pras et al., it can be assumed that the bandwidth follows a normal distribution [19].

2. *Using a TV connected game console is considered online gaming.* A TV connected game console is used to play video games. Video games likely use similar bandwidth to online games.

3. *Watching traditional television, using a TV connected Internet device, and using video on a computer are considered video streaming.* Each of these activities involves watching video on a device which is considered streaming. They would likely use similar

bandwidth depending on whether the streaming is done in high definition or standard definition.

4. *The expected bandwidth for video streaming a given activity is based on the proportion used in high definition (HD) and the proportion used in standard definition (SD).* Video streaming can be done in either HD or SD. The proportion of each shows the probability that a given definition would be used in the household.

5. *Average income for a given age will remain constant.* For 2006 to 2019, the percent distribution of average incomes for households across the United States remained approximately the same [20].

6. *Average income will not affect use of electronics if the consumer is retired.* Those considered to be retired will not be affected by the normal work hours associated with annual income. Therefore the amount of hours spent using electronics will not be affected by income.

7. *College students will attend online classes during the next year.* Due to the pandemic, it is highly unlikely that colleges will allow students to return to campus during the upcoming school year.

8. *All jobs within the next year will be remote.* Since introducing new members of the workforce to the workplace can increase chances of spreading COVID-19, it has been assumed that all jobs will be conducted from home.

3.2 Model Development

The model uses an expected income range based on the ages of the people in the household to determine the number of hours the household spends on several different categories. This expected number of hours is used to calculate the Megabits necessary for the household over the period of a week, which is then divided by the number of hours the household would be awake to calculate the necessary bandwidth.

3.2.1 Determining the Income Range

The expected income of a household is the sum of the average incomes for the ages of the people in the household [21]. This income will be mapped to one of the following four ranges: less than $25k, between $25k and $50k, between $50k and $75k, and more than $75k.

3.2.2 Determining the Hours Spent on the Internet

The hours spent on the Internet will be split into 7 categories: watching traditional television, using a TV connected game console, using a TV connected Internet device, Internet on a computer, total app/web on a smartphone, total app/web on a tablet, and total time spent online for school/work. We represent the hours per week for each of the categories by the following equation:

$$h = a \cdot \left(\frac{i}{e}\right), \tag{4}$$

where h is the expected number of hours per week for that category for the household, a is the expected number of hours per week based on the ages, i is the average number of hours per week based on the income range, and e is the expected number of hours per week across all of the income ranges. For each of these categories, using the average number of hours per week spent for each age group from the Nielsen Corporation, we determine a for the household [8]. Since the actual number of hours can vary depending on the income, we multiply by $\frac{i}{e}$, which accounts for higher or lower hours spent depending on the income level. Below is a table of the $\frac{i}{e}$ factors that multiply a.

Table 3.2.1: Hours per Week for Each Income Level Divided by the Expected Hours per Week for the Activity

Activity	< \$25k	\$25k–\$50k	\$50k–\$75k	> \$75k
Traditional Television	1.39 [8]	1.14 [8]	0.95 [8]	0.75 [8]
TV Connected Game Console	1.57 [8]	1.24 [8]	0.846 [8]	0.664 [8]
TV Connected Internet Device	1.42 [8]	1.23 [8]	0.965 [8]	0.668 [8]
Internet on Computer	1.29 [8]	1.05 [8]	0.982 [8]	0.845 [8]
Total App/Web on a Smartphone	1.13 [8]	1.04 [8]	1.04 [8]	0.896 [8]
Total App/Web on a Tablet	1.01 [8]	1.05 [8]	1.03 [8]	0.947 [8]
School/Work	0.784 [22]	0.987 [22]	1.15 [22]	1.16 [22]

3.2.3 Determining the Bandwidth (Mbps)

$$T = \sum h_k b_k, \tag{5}$$

where T is the total Megabits necessary and h_k and b_k are the average hours per week and expected bandwidth, respectively, for category k. The hours per week must be converted to seconds by multiplying by 3600 (number of seconds in an hour). Figure 3.2.1 and Table 3.2.3 below show the total Megabits for the three households. Household 1 refers to the couple in their early 30's with a 3-year-old child. Household 2 refers to the retired woman in her 70's who cares for two school-aged grandchildren twice a week. Household 3 refers to the three former M3 Challenge participants sharing an off-campus apartment while they complete their undergraduate degrees full-time and work part-time.

Table 3.2.2: Expected Bandwidth and Standard Deviation of Bandwidth

Activity	Bandwidth (Mbps)	Standard Deviation
Traditional Television	6.05 [8]	1.07
TV Connected Game Console	2 [8]	1
TV Connected Internet Device	1 [8]	0.5
Internet on Computer	2 [22]	1
Total App/Web on a Smartphone	1 [8]	0.5
Total App/Web on a Tablet	5 [8]	1.5
School/Work	2.5 [8]	1.5

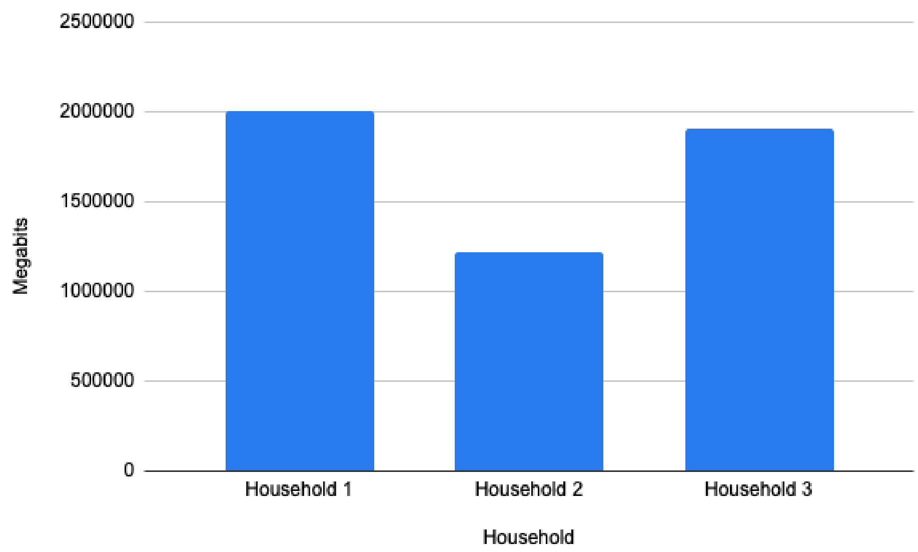

Figure 3.2.1: Bar Graph of Total Megabits for the Three Households

Table 3.2.3: Total Megabits for Each Household

Household	Total Megabits
Couple in 30's, 3-year-old child	2004688
Woman in 70's, two grandchildren	1216215
Former M3 Challenge participants	1907057

Humans spend about 720 minutes a day using technology [23]. Since the information provided was for a week, we can divide the total Megabits by 5040 minutes (302400 seconds). This value is the average bandwidth necessary for each person, so we must multiply by the total number of people in the household to obtain the average bandwidth μ.

3.2.4 Minimum Amount of Required Bandwidth

Bandwidth follows a normal distribution. The equation for a normal distribution is

$$d = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(R-\mu)^2}{2\sigma^2}}, \tag{6}$$

where d represents the bandwidth distribution, R is the minimum bandwidth for a given Internet availability, μ is the average bandwidth, and σ is the standard deviation. The bandwidth requirements for a given percentage of Internet availability (i.e., a fraction of time over which Internet needs must be met) are then given by

$$\alpha = \phi\left(\frac{R-\mu}{\sigma}\right) \implies \phi^{-1}(\alpha) = \frac{R-\mu}{\sigma}, \tag{7}$$

where α is the Internet availability and ϕ is the cumulative distribution function for a standard normal distribution. The standard deviation σ for each Internet activity is obtained by finding the average distance between the mean bandwidth and the upper and lower bounds of recommended bandwidth required per activity. Since variances add, the total standard deviation, σ_t, can be calculated using the equation

$$\sigma_t = \sqrt{\sum (d_k \cdot \sigma_k)^2}, \tag{8}$$

where d_k is the average bandwidth (total Megabits for the category divided by 302400 seconds) and σ_k is the standard deviation for each category k.

3.3 Results

Using the method described above, we calculated the average bandwidth for each household and the minimum bandwidth necessary for 90% and 99% Internet availability. The results are shown in Table 3.3.1.

Table 3.3.1: Bandwidth Needed for Each Household

Household	Average Bandwidth (Mbps)	Bandwidth for 90% Internet (Mbps)	Bandwidth for 99% Internet (Mbps)
Couple in 30's, 3-year-old child	19.9	26.1	31.2
Woman in 70's, two grandchildren	12.1	16.4	19.8
Former M3 Challenge participants	18.9	28.8	36.9

Below are bar graphs for the average bandwidth and minimum bandwidth for 99% Internet availability. Household 1 refers to the couple in their early 30's with a 3-year-old child. Household 2 refers to the woman in her 70's who cares for two school-aged grandchildren twice a week. Household 3 refers to the three former M3 Challenge participants.

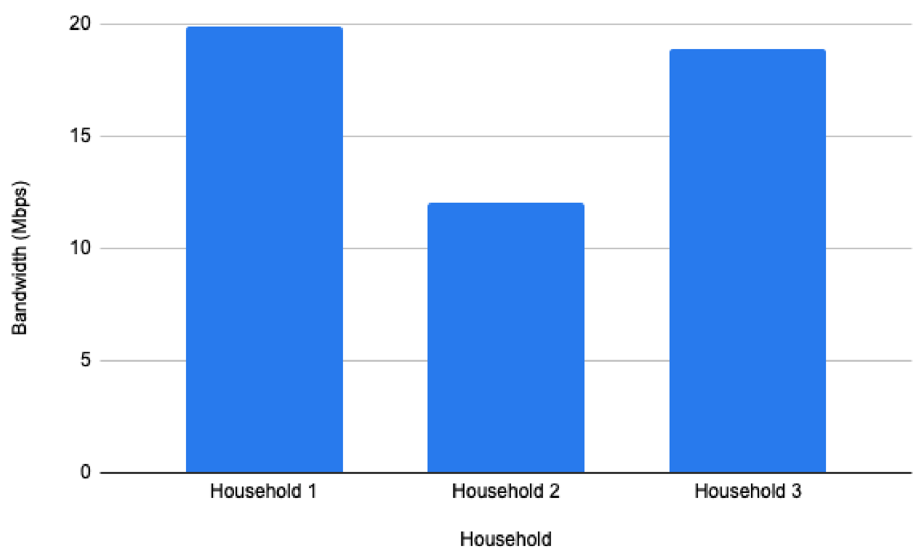

Figure 3.3.1: Bar Graph of Average Bandwidth for the Three Households

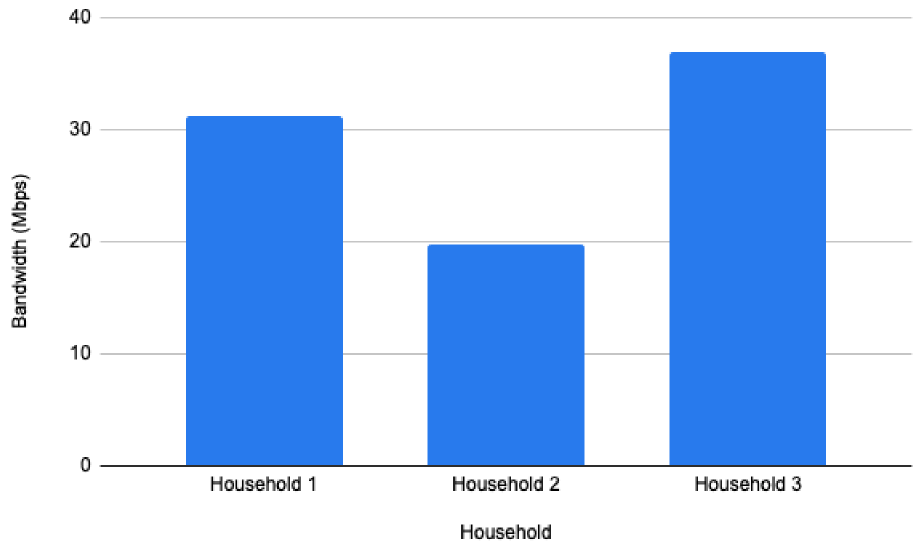

Figure 3.3.2: Bar Graph of Minimum Bandwidth for 99% Internet Availability

Our model concludes that for 90% and 99% Internet availability, the households require 26.1 Mbps and 31.2 Mbps, 16.4 Mbps and 19.8 Mbps, and 28.8 Mbps and 36.9 Mbps, respectively.

3.4 Sensitivity Analysis

Table 3.4.1 shows the sensitivity analysis for our residential broadband access bandwidth model. The variables present in our model are the hours spent on different activities requiring Internet access. For the above calculation in Table 3.3.1, the expected values of bandwidth required per activity and hours of usage are used. For the sensitivity analysis, we increased and decreased the value of $\frac{i}{e}$ by 10%.

Table 3.4.1: Effects of Changes in $\frac{i}{e}$ on Average Bandwidth

Household	Average Bandwidth (+10% for $\frac{i}{e}$)	Average Bandwidth (-10% for $\frac{i}{e}$)
Couple in 30's, 3-year-old child	21.9	17.9
Woman in 70's, two grandchildren	13.3	10.9
Former M3 Challenge participants	20.8	17.0

As expected, positive changes in the $\frac{i}{e}$ factor resulted in positive changes in the average bandwidth necessary at each of the households, and negative changes in the severity factor resulted in negative changes in the average bandwidth. It is also seen that the model is resilient to changes in the hours per week for each activity based on income.

3.5 Strengths and Weaknesses

Our model's primary strength is that it takes into account different income levels and the hours spent on each activity for different age groups. By taking into account the income levels, the model considers economic gaps that could affect technology used by the different groups. Also, by considering the many age groups, the model can account for varied uses of technology in households of different ages. Furthermore, the model ensures a minimum calculated bandwidth meets the Internet needs given a certain Internet availability by using a normal distribution, accounting for all the different possible bandwidths necessary. Finally, as shown in the sensitivity analysis, our model is resilient to changes in variables.

A weakness is that this model assumed that all jobs and college classes will be online for the next year due to COVID-19, as COVID-19 is still present, and there is not much data describing the trend of online classes post-COVID-19. For this reason, there is uncertainty regarding post-pandemic online behavior regarding education and work.

4 Part III: Mobilizing Mobile

Mobile broadband is transmitted from towers or nodes, so to minimize the cost-effectiveness of high-speed Internet, we must optimize the distribution and placement of cellular nodes [24]. This section outlines a model that determines the optimal distribution of low-band, mid-band, and high-band cellular towers in a region.

4.1 Assumptions

1. *There is a uniform population density with uniform demographics within each region.* This is reasonable to assume since each subregion may contain similar neighborhoods that contain similar demographics. It is unreasonable to assume a non-uniform population density as this information is not provided in [8].

2. *All three types of cellular towers have equal cost.* This is reasonable to assume since the majority of the cost of a cellular tower is not due to which type of band it is running, but due to physical requirements of construction [25].

3. *Each individual within a subregion will only use Internet provided by cellular towers within that subregion.* This is reasonable to assume because a cellular tower in another subregion would not provide optimal Internet service for the user since it is further away.

4. *The optimal number of cellular towers in a certain subregion is independent of the cellular towers in other regions.* This is clear if users are only using cellular towers within their subregion, and it allows us to consider each subregion separately to determine the optimal setup of cellular nodes.

5. *The given regions A, B, and C are all suburban areas.* From [8], one can verify that the population densities are below the threshold for urban areas.

6. *The number and type of cellular towers needed are dependent on the area of the subregion and the specific demographics of the subregion. In particular, we do not consider the shape of each subregion.* The specific dimensions of each region are not given, and considering the shape of each subregion is out of the scope of our model.

7. *The effective area covered by a cellular tower is the region such that the Internet speed recorded is at least midway between the minimum and maximum speeds for the cellular tower.* This is a reasonable assumption as areas closer to the cellular tower will have faster speeds, and areas farther from the cellular tower but within its reach will have slower speeds.

8. *The transmitted frequency from each cellular tower is the frequency halfway between the maximum and minimum frequencies for the range.* This is reasonable because the average frequency recorded is expected to be halfway between minimum and maximum frequencies.

4.2 Model Development

The primary considerations for the construction of cellular towers in a subregion is that the entire area is covered and that there is sufficient bandwidth to meet the needs of every person. In this section, we develop a flexible model to determine the quantity of low-band, mid-band, and high-band towers necessary to meet the demand for Internet. Low-band towers have the greatest range, but are the least powerful, operating on a frequency of 600 MHz to 700 MHz. Mid-band towers operate on a frequency of 2.5 GHz to 3.5 GHz. High-band towers have the shortest range, but operate on a frequency of 24 GHz to 39 GHz [8].

$$A_{\text{low}} \cdot n_{\text{low}} + A_{\text{mid}} \cdot n_{\text{mid}} + A_{\text{high}} \cdot n_{\text{high}} \geq A, \tag{9}$$

$$B_{\text{low}} \cdot n_{\text{low}} + B_{\text{mid}} \cdot n_{\text{mid}} + B_{\text{high}} \cdot n_{\text{high}} \geq B. \tag{10}$$

On a high level, our model is described by Equations (9) and (10). For each type of tower, we determine an effective area of coverage per tower. We then multiply this by the number of towers in that band and sum across all three bands to obtain the total area of coverage, which must be greater than or equal to A, the area of the subregion. A similar process is used for bandwidth: the total bandwidth for one tower is multiplied by the amount of that type of tower and this is summed across the three bands, resulting in a value greater than or equal to the total need for bandwidth in the subregion. This additionally allows for the easy adapting of this model to a specific model of tower, which has a definite value for its total bandwidth. The optimal distribution of towers will follow Equations (9) and (10) and minimize $x + y + z$ over the integers:

$$A_{\text{low}} \cdot B_{\text{low}} \cdot n_{\text{low}} + A_{\text{mid}} \cdot B_{\text{mid}} \cdot n_{\text{mid}} + A_{\text{high}} \cdot B_{\text{high}} \cdot n_{\text{high}} \geq A \cdot B. \tag{11}$$

It is also required for the necessary bandwidth to be present in all parts of the subregion, leading to Equation (11). For each band of tower, the effective area multiplied by the bandwidth multiplied by the number of towers is summed; the resulting quantity must be greater than or equal to the total area of coverage multiplied by the total requisite bandwidth. With all three of these conditions, the resulting solution will always be valid.

4.2.1 Calculating Effective Areas

Table 4.2.1: Variables Used in Model for Effective Area

Variable	Meaning
R	rate achievable (Mbps)
B	frequency bandwidth (Hz)
L_B	median path loss (dB)
f	frequency of band
h_B	effective height of the base station (m)
h_R	effective height of the receiver (m)
d	distance from base station (km)
$a(h_R, f)$	factor that depends on the environment (urban, suburban, etc.)
SNR	received signal to noise ratio
P	power transmitted from base station (dBm)
N	noise (dBm)

By the Shannon-Hartley Theorem [25], the Internet speed achievable R is

$$R = B \log_2 (1 + \text{SNR}) . \tag{12}$$

From Equation (12), we can determine the received signal to noise ratio from the rate. We use the rate R equal to the average of the minimum and maximum rates for each cellular tower according to Assumption 7 in section 4.1.

Then, as in Assumption 6, we assume the frequency transmitted f is the middle of each frequency band from [8]. To relate the frequency f and Internet speed R to the distance, we use the COST Hata Model. The COST Hata model is used to find the median path loss L_b of a wireless signal from the distance [26]. It is based on the Okumura model, which can be used to find path loss through urban environments. The model works best for low-band and mid-band towers, but it can also be utilized for high-band towers as it is dependent on the frequency. We have the following equation [26]:

$$\begin{aligned} L_B = {} & 46.3 + 33.3 \log_{10} \left(\frac{f}{1\,\text{MHz}} \right) - 13.82 \log_{10} \frac{h_B}{1\,\text{m}} - a(h_R, f) \\ & + \left(44.9 - 6.55 \log_{10} \frac{h_B}{1\,\text{m}} \right) \log_{10} \frac{d}{1\,\text{km}} \end{aligned} \tag{13}$$

where the variables are defined as in Table 4.2.1. Here, $a(h_R, f)$ is defined for suburban or rural environments to be

$$a(h_R, f) = \left(1.1 \log_{10} \frac{f}{\text{MHz}} - 0.7 \right) \frac{h_R}{1\,\text{m}} - \left(1.56 \log_{10} \frac{f}{\text{MHz}} - 0.8 \right) . \tag{14}$$

The path loss or attenuation can also be found through the following equation [32]:

$$\text{SNR} = \frac{\gamma P}{N} , \tag{15}$$

where γ is the attenuation. This can also be converted to dB:

$$\text{SNR} = P - N - L_b \ \ (\text{in dB}). \tag{16}$$

The noise can be calculated based on the value of the frequency bandwidth B for each type of cellular tower [27]. Furthermore, the value of P and B is known for each type of cellular tower given [28]. In addition, it is known that the average height of a house is 6 meters, so we use $h_R = 6$ m [30]. We also use the average height of a cellular tower as 60 m [31].

For each type of tower, we determine the distance r such that the rate achievable is midway between the lower and upper bounds of the cellular tower in [8]. The effective area A covered by the cellular tower is then given by the area of a circle with radius r:

$$A = \pi r^2. \tag{17}$$

The obtained values for r and A are shown in Table 4.2.2.

Table 4.2.2: Tower Types and Effective Area Covered

Band	r (miles)	A (square miles)
Low	2.735	23.49
Mid	0.465	0.679
High	0.0323	0.00328

As can be seen in Table 4.2.2, the effective area covered by the high-band tower is rather small. This is expected because the high-band tower has higher frequencies and covers much smaller areas.

4.2.2 Calculating Bandwidth Needs

We lack detailed information on the demographics of each subregion and thus used the median ages and incomes to calculate bandwidth needs as outlined in the previous section of paper (though limiting the activities to only smartphone and tablet use). We multiplied the bandwidth need of the median person in each subregion by the population of that subregion to arrive at the total mean bandwidth usage in Mbps. In computing the effective areas, we implicitly included extra capacity in the area where the download speed is less than half of the advertised speed, so peak loads are able to be dealt with.

4.3 Results

Using integer programming as shown in min-cell-towers.py, we minimize the number of cellular towers $n_{\text{low}} + n_{\text{mid}} + n_{\text{high}}$ given the constraints described in Equations (9) and (10). In tables 4.3.1, 4.3.2, and 4.3.3, optimizations for the number of cellular towers for each subregion in regions A, B, and C, respectively, are displayed.

Table 4.3.1: Number of Cellular Towers for Each Subregion of Region A

Subregion	Low-band Towers	Mid-band Towers	High-band Towers
1	1	1	0
2	0	2	0
3	0	2	0
4	1	1	0
5	0	1	1
6	1	0	1

Table 4.3.2: Number of Cellular Towers for Each Subregion of Region B

Subregion	Low-band Towers	Mid-band Towers	High-band Towers
1	3	0	1
2	2	0	1
3	1	0	1
4	1	0	1
5	4	0	1
6	2	0	1
7	3	0	1

Table 4.3.3: Number of Cellular Towers for Each Subregion of Region C

Subregion	Low-band Towers	Mid-band Towers	High-band Towers
1	0	1	1
2	0	1	1
3	0	1	1
4	0	1	1
5	0	1	1
6	0	1	1
7	1	0	1

In region A, it is seen that it is usually more optimal to use low-band and mid-band towers to cover the entire region. This is because the regions in A had moderate area while a generally lower required bandwidth. In region B, it was most optimal to use 1 high-band tower and some number of low-band towers to cover the region. This makes sense as the subregions in B were much larger, and we need low-band towers to provide optimal coverage. In region C, it is most optimal to use 1 mid-band tower and 1-high band tower in most of the regions. This is clear because most of the subregions of C had small area while requiring a moderate amount of bandwidth.

4.4 Sensitivity Analysis

We conducted a sensitivity analysis by raising and lowering each of the constants A_{low}, A_{mid}, and A_{high} by 10%. By running min-cell-towers.py again, we found that the number of cellular towers of each type was not affected in any of the 20 subregions from regions A, B, and C. This shows that our model is robust and resilient to changes in the effective area of each cellular tower.

4.5 Strengths and Weaknesses

Our model's greatest strength is its robust flexibility that allows it to be applied to a variety of situations, as the technical specifications of the cellular towers—in bandwidth, radio frequency, and dimensions—are all parameterized in various locations. Additionally, more precise results can be obtained in real life; the hypothetical regions presented to us did not have detailed demographic information for each of the subregions. With this added data, the model will produce better results. It can also be adapted to urban areas, as the COST Hata model of path loss is specifically designed to apply to a variety of realistic environments.

Unfortunately, we were unable to consider cellular towers that overlap between subregions due to a lack of data. We were not given detailed shapes or dimensioned maps of the hypothetical regions. It may be possible to create a more cost-effective plan for building cellular towers than we have created. However, excess capacity is not detrimental, as Internet loads vary.

5 Conclusion

5.1 Further Studies

Our first model does not currently account for technology in the form of 6G and future innovations in current technology which would likely impact the cost of bandwidth in the future. Taking this into account would strengthen our model, as it would be able to predict the cost more than 10 years from now. Our second model focuses mainly on the immediate years following COVID-19, so as more data becomes available for the long-term trends as a result of COVID-19, we will be able to better predict the necessary bandwidth for a given household. Finally, our third model does not consider the shape of each subregion and disregards the overlap of a cellular tower across multiple regions. With more detailed information and a thorough analysis about these factors, the accuracy of our model could be improved.

5.2 Conclusion

In Part I, we predicted the average cost per bandwidth (Megabits per second) for Internet in the United States and United Kingdom in 2030. We developed a logistic model for Internet speed in three different technologies: wired, mobile, and satellite. Our robust model estimated future Internet speeds and costs for each technology, which were averaged to produce a global indicator. In the United States, the cost of Internet per unit bandwidth was predicted to go down from \$3.86 in 2020 to \$0.46 in 2030. In the United Kingdom, the cost of Internet per unit bandwidth was predicted to go down from \$3.98 in 2019 to \$0.49 in 2030.

In Part II, we determined the minimum amount of bandwidth required that would cover 90% and 99% of total Internet needs for each of the three households. This was calculated by determining how each household's income level and ages for the individuals in the household affected the number of hours spent on the Internet. We mapped necessary bandwidth to a normal distribution to determine the minimum bandwidth that would cover a certain Internet availability. The first household was determined to require 26.1 Megabits per second to meet 90% of their Internet needs and 31.2 Megabits per second to meet 99% of their Internet needs. The second household was determined to require 16.4 Megabits per second to meet 90% of their Internet needs and 19.8 Megabits per second to meet 99% of their Internet needs. the third household was determined to require 28.8 Megabits per second to meet 90% of their Internet needs and 36.9 Megabits per second to meet 99% of their Internet needs.

In Part III, we minimized the total quantity of cellular towers necessary to provide coverage to each of the three hypothetical regions by constructing a model that determined how many low-band, mid-band, and high-band towers were needed. We split up each region into its constituent subregions, for each of which we determined the mean demand for bandwidth based on its demographics and our model from Part II. Additionally, we used models of radio wave propagation to determine the area on which a cellular tower of a specific band would be effective and utilized this information to ensure that the full area of each region retained high-speed coverage. The model we created is robust and can be applied to any set of region data, and would improve in precision with more accurate and comprehensive data about a region.

6 References

1. "The children left behind in a world of distance learning." (n.d.). Retrieved February 28, 2021, from https://parentzone.org.uk/article/children-left-behind-world-distance-learning.

2. Peyton, N. (2020, August 28). "'Who is standing up FOR US?'- Black, rural students left behind as U.S. schools go online." Retrieved February 28, 2021, from https://www.reuters.com/article/us-health-coronavirus-usa-education-feat-idUSKBN25O1XR.

3. Webster, P. (2020, April 11). "Virtual health care in the era of COVID-19." Retrieved February 28, 2021, from https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)30818-7/fulltext.

4. Willcox, J. K. (n.d.). "COVID-19 Relief Package Will Help Families in Need of Internet Access." Retrieved February 28, 2021, from https://www.consumerreports.org/broadband-access/covid-19-relief-package-will-help-families-in-need-of-internet-access/.

5. Christensen, G. (2020, August 31). "What is 6G and when will it come out?" Retrieved February 28, 2021, from https://searchnetworking.techtarget.com/definition/6G.

6. Britt, P. (2020, August 12). "Broadband Monopoly Report: At Least 49.7 Million Americans Served by Only One Provider." Retrieved February 28, 2021, from https://www. telecompetitor.com/broadband-monopoly-report-at-least-49-7-million-americans-served-by-only-one-provider/.

7. Bergman, A. and Iyengar, J. (2020, April 8). "How COVID-19 is affecting internet performance." Retrieved February 28, 2021, from https://www.fastly.com/blog/how-covid-19-is-affecting-internet-performance.

8. Defeating the Digital Divide Data, MathWorks Math Modeling Challenge 2021. https://m3challenge.siam.org/node/523.

9. Kucharavy, D. and De Guio, R. (2011, May 7). "Logistic substitution model and technological forecasting." Retrieved February 28, 2021, from https://www.sciencedirect.com/science/article/pii/S1877705811001469.

10. "Broadband Performance." (n.d.). Retrieved February 28, 2021, from https://transition.fcc.gov/national-broadband-plan/broadband-performance-paper.pdf.

11. "Connect America Fund Phase II Auction (Auction 903)." (2020, November 12). Retrieved February 28, 2021, from https://www.fcc.gov/auction/903.

12. "The satellite broadband industry is moving at hyperspeed." (2020, February 19). Retrieved February 28, 2021, from https://www2.deloitte.com/us/en/insights/industry/technology/future-of-satellite-internet.html.

13. enTouch. (2018, March 02). "What Is 1 Gig Internet?" Retrieved February 28, 2021, from https://www.entouch.net/what-is-1-gig-internet/.

14. 2021 “Viasat Speed Test & Statistics.” (2021, January 22). Retrieved February 28, 2021, from https://broadbandnow.com/Exede-Internet-speed-test.

15. 2021 “HughesNet Speed Test & Statistics.” (2021, January 25). Retrieved February 28, 2021, from https://broadbandnow.com/HughesNet-speed-test.

16. “The Cost of Connectivity 2020.” (n.d.). Retrieved February 28, 2021, from https://www.newamerica.org/oti/reports/cost-connectivity-2020/global-findings/.

17. Allconnect. (2020, December 02). “Average cellphone plan cost is $113: Cost factors.” Retrieved February 28, 2021, from https://www.allconnect.com/blog/the-average-cost-of-a-cellphone-plan-in-the-us-is-113.

18. Schafer, D. (2021, January 26). “How Much Does Satellite Internet Cost?” Retrieved February 28, 2021, from https://www.satelliteinternet.com/resources/how-much-does-satellite-internet-cost/.

19. A. Pras, L. Nieuwenhuis, R. van de Meent, and M. Mandjes. “Dimensioning network links: A new look at equivalent bandwidth.” IEEE Network, 2009. Retrieved February 28, 2021, from https://ieeexplore.ieee.org/document/4804330.

20. Statista Research Department. (2021, January 20). “Distribution of household income in the U.S. 2019.” Retrieved February 28, 2021, from https://www.statista.com/statistics/758502/percentage-distribution-of-household-income-in-the-us/.

21. Schrager, A. (2016, January 13). “Americans are working fewer hours—and it’s making income inequality worse.” Retrieved February 28, 2021, from https://qz.com/574693/americans-working-less-than-ever-before/amp/.

22. “Income Percentile by Age Calculator [2020].” (2020, December 28). Retrieved February 28, 2021, from https://dqydj.com/income-percentile-by-age-calculator/.

23. Statista Research Department. (2014, August 06). “Awake time per day in the United Kingdom (UK) 2014, by age and gender.” Retrieved February 28, 2021, from https://www.statista.com/statistics/326907/awake-time-per-day-uk-by-age-and-gender/.

24. Verizon. (2020, June 01). “5G nodes: What they are and how they work.” Retrieved February 28, 2021, from https://www.verizon.com/about/news/towers-what-they-are-how-they-work.

25. “Catalog of Potential Expenses and Estimated Costs.” (n.d.). Retrieved February 28, 2021, from https://docs.fcc.gov/public/attachments/DA-17-154A2.pdf.

26. Singh, Y. (2012, December). “Comparison of Okumura, Hata and COST-231 Models on the Basis of Path Loss and Signal Strength.” Retrieved February 28, 2021, from https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.303.4057&rep=rep1&type=pdf.

27. “CS3282: Digital Communications ’05-’06, Section 7: The Shannon-Hartley Theorem.” (n.d.). Retrieved February 28, 2021, from http://www.cs.man.ac.uk/~barry/mydocs/CS3282/Notes/DC06_7.pdf.

28. “Physical Sources of Noise.” (n.d.). Retrieved February 28, 2021, from https://ieeexplore.ieee.org/document/4052064.

29. “Frequency Bands and Applications” — blog. (2020, October 05). Retrieved February 28, 2021, from https://www.jemengineering.com/blog-frequency-bands-and-applications/.

30. George. (2021, February 18). “How Tall Is a 2 Story House: An In-depth Analysis.” Retrieved March 01, 2021, from https://ladderhunt.com/how-tall-is-a-2-story-house.

31. Kanchwala, H. (2021, January 7). “What Are Cell Towers and How Do They Work?” Retrieved February 28, 2021, from https://www.scienceabc.com/innovation/cell-tower-work.html.

32. “Signal-to-Noise Ratio.” (n.d.). Retrieved February 28, 2021, from https://www.sciencedirect.com/topics/engineering/signal-to-noise-ratio.

7 Appendix

7.1 average-speed.py

```
import matplotlib.pyplot as plt
import numpy as np
from scipy.optimize import curve_fit

# Data for average wired and mobiles speeds
#wired
usWSpeed = [4.2,4.7,5.3,6.7,8.6,10.5,11.9,15.3,18.7]
ukWSpeed = [3.7,3.8,4.6,5.6,7.9,9.9,11.6,14.9,16.9]
#mobile
usMSpeed = [0.9,1.3,1.6,2.8,5.5,4.0,5.1,10.7]
ukMSpeed = [2.3,2.7,3.2,3.1,5.6,20.4,27.9,26.0]

#Included below is sample code for US/UK Wired Speed
#Small modifications are needed for Mobile speeds

#x-axis for data points
time1 = [t for t in range(9)]

#logistic function for model
def logistic(t, L, k,t0,c):
    return L/(1 + np.exp(-k*(t-t0)))+c

#introduce parameter bounds to refine model
param_bounds=([1000,-np.inf,0,-np.inf],[np.inf,np.inf,np.inf, np.inf])
#curve fit
popt, _ = curve_fit(logistic, time, ukSpeed, bounds=param_bounds)
L,k,t0,c = popt


# Sample calculation to compute R^2 values
residuals = [ukSpeed[i] - logistic(time[i],L,k,t0,c) for i in range(9)]
residualSquares = [residual**2 for residual in residuals]
residualsSum = sum(residualSquares)
totalSquares = sum([(ukSpeed[i]-sum(ukSpeed)/9)**2 for i in range(9)])
rSquared = 1 - (residualsSum)/totalSquares
print(rSquared)


#create plot; remember to shift t by 2009 to convert to years
x_line = np.arange(min(time), max(time)+15, 1)
y_line = logistic(x_line, L,k,t0,c)
years = [t+2009 for t in time]
x_line = np.arange(min(time)+2009, max(time)+15+2009, 1)
plt.rcParams['text.usetex'] = True
plt.scatter(years,ukSpeed)
bestFit = plt.plot(x_line, y_line, color='red')
plt.xlabel("Year")
plt.ylabel(r'Average Internet Speed (Mbps)')
plt.legend([r'$S = \frac{%.5f}{1+e^{-%.5f(t-%.5f)}}+%.5f$' %(L,k,t0+2009,c
```

```
    )])
plt.xticks([2009+2*i for i in range(13)])
plt.show()
```

7.2 min-cell-towers.py

```
from pulp import LpMinimize, LpProblem, LpVariable

# Bandwidths of low-, medium-, high- band cellular towers in Mbps
B_low = 140
B_mid = 500
B_high = 2000

# Effective areas of low-, medium-, high- band cellular towers in square miles
A_low = 23.49325202
A_mid = 0.6789549833
A_high = 0.003281288918

# Total bandwith and total area (varies by subregion)
B = 399.2085357 #in Mbps
A = 1.21 #in square miles

model = LpProblem(name="min-cell-towers", sense=LpMinimize)

# Define the decision variables
n_low = LpVariable(name="n_low", lowBound=0, cat='Integer')
n_mid = LpVariable(name="n_mid", lowBound=0, cat='Integer')
n_high = LpVariable(name="n_high", lowBound=0, cat='Integer')

# Add constraints
model += (A_low*n_low + A_mid*n_mid + A_high*n_high >= A)
model += (B_low*n_low + B_mid*n_mid + B_high*n_high >= B)
model += (A_low*B_low*n_low + A_mid*B_mid*n_mid + A_high*B_high*n_high >= A*B)

# Set objective
obj_func = n_low + n_mid + n_high
model += obj_func

# Solve the optimization problem
status = model.solve()

# Print results
for var in model.variables():
    print(f"{var.name}: {int(var.value())}")
```