MathWorks Math Modeling Challenge 2021

Johns Creek High School

Team # 14482 Johns Creek Georgia

Coach: Julie Meert

Students: Jason Bao, Aditya Bora, Mehul Dhoot, Joseph Suharno, Austin Tsang

M3 Challenge THIRD PLACE—$12,000 Team Prize

JUDGES' COMMENTS

Specifically for Team # 14482 —Submitted at the Close of Triage Judging:

COMMENT 1: Very nice narrative here.

COMMENT 2: Your summary is well-written and does a good job of stating your motivations and conclusions. In general, try to avoid technical references, such as to data sets ("the data in D4") in an executive summary in favor of more general explanations of your work. Nice job presenting your first model. The model evaluation section was also appreciated, although more could have been said about the possible limitations of your assumption of using exponential functions, particularly with respect to long-term behavior. Good overall work on your second model: you give a through explanation of your methodology, present the conclusions clearly, and then discuss possible limitations. You present an excellent analysis in your third model of how the given regions influenced your model construction and how it may have changed if you needed to apply it to different population densities.

COMMENT 3: This was quite simply an excellent paper. You did an excellent job with your summary. It both summarizes your results and your approaches to each problem. With each problem you did an excellent job with your assumptions and then you also documented/supported those assumptions. Your models were very well developed and explained. Your graphs were crystal clear along with your conclusions. Excellent job with the sensitivity analysis and strengths/weakness. This is easily the best paper I read this year!

Defeating the Digital Divide

Executive Summary:

As epitomized by the current pandemic, high-speed internet has transcended its former role as a gateway to entertainment and now represents a vital factor for academic and financial success as well as physical and mental health. However, this growth in high-speed internet has unfortunately exacerbated wealth and geographical divides, presenting a new problem in determining how to best bridge this digital disparity. As such, our group seeks to explore the cost of bandwidth and each household's relative need for the internet to together combine this into establishing a node distribution plan that effectively meets the bandwidth needs of any region.

We first strove to determine the cost per unit of bandwidth in dollars per Mbps over the next 10 years for US and UK consumers. Thus, we developed two expressions for the price per Mbps of bandwidth for each country by using the average bandwidth download speed and the average plan price (wireless or wire internet), thus AS/PP. Then, by learning that the distribution of internet plans was 68.3% wireless and 31.7% wired, we computed a weighted average for the overall prices for both prices, eventually determining that by 2030, each Mbps of bandwidth costs $.10 and $.22 in the US and UK, respectively. This is a logical conclusion as in the past decade, internet plan prices have steadily decreased while innovations in broadband technology have created exponential increases in Mbps of bandwidth, causing greater transmissions of bandwidth for decreased prices and thus, decreases in cost per unit in the next decade.

Next, we worked to develop a model to predict a given household's need for the internet over the course of a year and tested this model against three given scenarios. Because we wanted to find the total amount of data each household uses, we discovered that this data amount usage equaled the bandwidth spent during a certain amount of time. As such, to understand how people utilize their time (respective data usage), we analyzed age and household income levels. To determine the respective weights for each factor, we performed linear regressions of the data in D4 to find the magnitude of lines of best fit which reflected upon the impact of each factor on hours spent on an activity. Knowing the time spent on certain activities and their respective bandwidth and data usage, we determined the minimum amount of required bandwidth for the three scenarios: (1) 3.090 Mbps for 90% and 3.604 Mbps for 99%, (2) 2.878 Mbps for 90% and 3.356 Mbps for 99%, and (3) 3.290 Mbps for 90% and 3.836 Mbps for 99%.

Finally, we established a model that optimizes the distribution of mobile broadband transmitting cellular nodes in a region. In order to do so, we determined the bandwidth demand of each subregion and created a model that calculates the minimum number of nodes outputting low, middle, or high band required to satisfy the bandwidth and area requirements. Whichever number of nodes is the lowest of the three bands is chosen for the particular subregion. Through this model, we determined that subregions 1, 4, 5 (A) require 1 node of high band while subregions 2, 3, 6 (A) require 2 nodes of high band, subregions 2, 6, 7 (B) require 3 nodes of high band while subregions 3, 4 (B) require 2 nodes of high band, subregion 1(B) needs 5 nodes of high band, and subregion 5 (B) needs 4 nodes of high band, and finally all subregions of C require 2 nodes of high band.

Table of Contents

1 Introduction

1.1 Restatement of the Problem

We are tasked with the following problem[1]:

1. Build a model to predict the cost per unit of bandwidth in dollars or pounds per Mbps over the next 10 years for consumers in the United States and the United Kingdom.
2. Create a flexible mathematical model to predict a given household's need for the internet over the course of a year. Apply your model to the example households listed below and determine the minimum amount of required bandwidth that would cover their total internet needs 90% of the time. What about 99% of the time?
    a. A couple in their early 30s (one is looking for work and the other is a teacher) with a 3-year-old child.
    b. A retired woman in her 70s who cares for two school-aged grandchildren twice a week.
    c. Three former M3 Challenge participants sharing an off-campus apartment while they complete their undergraduate degrees full-time and work part-time.
3. Develop a model that produces an optimal plan for distributing/placing cellular nodes in a region.

1.2 Global Assumptions

1. Bandwidth will be defined as download speed of data per second.
2. There will not be any major revolutionary advancement in broadband technology beyond the currently developed and used technologies (fiber optic, DSL, cable, etc.).
    a. **Justification**: Over the course of 10 years, there will solely be adaptations to current existing technology such as improving 5G networks as the development of major advancements in broadband technology require often spans across multiple decades as seen in the development of fiber-optic communication since 1880[2].
3. Inconsistencies in environmental conditions (such as severe weather, buildings, altitude differences) are negligible and can be ignored.
4. Sudden surges in population density (such as unexpected crowds) will be extremely uncommon. As such, their impact on internet infrastructure will be ignored.

2 Part I: The Cost of Connectivity

As the need to provide access to quick and stable internet grows due to our increasingly digital world, a situation that the coronavirus pandemic accelerated significantly[3], the cost of such an initiative remains highly unknown and a potential barrier toward this goal. As such, this section outlines the mathematical model for predicting the cost per unit of bandwidth in dollars per Mbps in both the United States and the United Kingdom and analyzes the cost in relation to wireless versus wired plans for customers across the next 10 years.

2.1 Local Assumptions

1. *The prices for 4G and 5G plans will be roughly equal and change at the same rates. That is, a price increase in 4G will result in an identical price increase in 5G, allowing us to calculate an overall price of broadband without having to calculate a weighted average stratified by 4G plans vs. 5G plans.*
   a. **Justification:** According to Ofcom, most internet providers are allowing consumers to switch from 4G to 5G plans at no additional cost[5]. We can assume this trend to continue as more consumers make the switch to 5G over the next decade.
2. *While broadband prices do vary due to differences in service providers (i.e., AT&T, Comcast, BT, etc.) or due to differences in materials (i.e., cable vs. DSL vs. fiber optic), we can assume that these differences are negligible when discussing overall broadband price increases or decreases.*
   a. **Justification:** There is no evidence to support that existing market shares will drastically increase since they have remained roughly constant since 2015 in both the United States and the United Kingdom[6][7]. For example, it is unlikely that Verizon will suddenly gain far greater control over the ISP market and therefore gain more influence over broadband prices.
3. *All consumers will utilize the same wired and wireless plans for the next ten years.*
   a. **Justification:** As stated above, all plans will have a negligible price difference, so we can assume all consumers pay the same prices for their wired or wireless plans.
4. *The costs of the wireless plans in the US are approximately 1.86 times the prices in the UK.*
   a. **Justification:** When comparing the prices of mobile internet between the US and UK, the prices in the US are 1.857 times greater than that of the UK, so this relationship can be applied to the prices of UK vs. US wireless plans[8]. This assumption was made because US wireless plan data was not available neither through the given data nor online.

2.2 Model Development

When developing our model, we first found an expression for the price per Mbps of bandwidth. If the average speed is *AS* and the average plan price is *PP*, then the average price per Mbps would be *AS/PP*. We then decided to differentiate between wired and wireless plans. Therefore, if the proportion of wired plans is $Prop_{wired}$ and the proportion of wireless plans is $Prop_{wireless}$, then the average price per Mbps for all plans would follow this equation:

$$Prop_{wired}\left(\frac{AS_{wired}}{PP_{wired}}\right) + Prop_{wireless}\left(\frac{AS_{wireless}}{PP_{wireless}}\right)$$

We used this equation to calculate for the respective prices of both countries. We then assumed that the equations for the speeds and prices of both wired and wireless plans were exponential functions. Using the data found in the provided data, we used regression to calculate the exponential trendlines to model these values over time and recorded the equations:

If x equals years from 2020:

US:

$$AS = AS_{wired} = AS_{wireless} = 3.896e^{0.193(x+11)}$$
$$PP_{wired} = 62.2e^{-0.009(x+8)}$$
$$PP_{wireless} = 44.267e^{-0.132(x+5)}$$

UK:

$$AS = AS_{wired} = AS_{wireless} = 3.274e^{0.299(x+11)}$$
$$PP_{wireless} = 27.214e^{-0.034(x+6)}$$
$$PP_{wireless} = 23.843e^{0.193(x+5)}$$

We then researched the distribution of the total plans to either wired or wireless plans. We found that in 2020, 68.3% of plans were wireless and 31.7% of plans were wired[9]. We assumed that these proportions were the same in the US and the UK. We also found that on average, the number of wireless plans increased by 3% per year and the number of wired plans increased by 5% per year. Using this data, we were able to create equations for $Prop_{wired}$ and $Prop_{wireless}$:

$$Prop_{wired} = \frac{0.317(1.05)^x}{0.683(1.03)^x + 0.317(1.05)^x}$$
$$Prop_{wireless} = \frac{0.683(1.03)^x}{0.683(1.03)^x + 0.317(1.05)^x}$$

We plugged these equations into the general price equation and used Python in order to get a model. In our model, we used the conversion rate of one pound is equal to 1.39 dollars to convert pounds to dollars and get a more accurate comparison between US prices and UK prices.

2.2.1 Defined Variables and Parameters in Model

Symbol	Term	Formula
$AS_{wired/wireless}$	Average Wired/Wireless Bandwidth Speed (Mbps)	US: $3.896e^{0.193(x+11)}$ UK: $3.274e^{0.299(x+11)}$
$Prop_{wired/wireless}$	Proportion of Wired/Wireless Plans	Wireless: $\frac{0.683(1.03)^x}{0.683(1.03)^x+0.317(1.05)^x}$ Wired: $\frac{0.317(1.05)^x}{0.683(1.03)^x+0.317(1.05)^x}$
PP_{wired}	Average Wired Plan Price ($ or £)	US: $62.2e^{-0.009(x+8)}$ UK: $27.214e^{-0.034(x+6)}$
$PP_{wireless}$	Average Wireless Plan Price ($ or £)	US: $44.267e^{-0.132(x+5)}$ UK: $23.843e^{0.193(x+5)}$

Table 2.2.1: Variables of Model of Connectivity Costs

2.2.2 Equations Utilized in Model

$$Prop_{wired}\left(\frac{AS_{wired}}{PP_{wired}}\right) + Prop_{wireless}\left(\frac{AS_{wireless}}{PP_{wireless}}\right)$$

<u>US:</u>

$$AS = AS_{wired} = AS_{wireless} = 3.896e^{0.193(x+11)}$$
$$PP_{wired} = 62.2e^{-0.009(x+8)}$$
$$PP_{wireless} = 44.267e^{-0.132(x+5)}$$

<u>UK:</u>

$$AS = AS_{wired} = AS_{wireless} = 3.274e^{0.299(x+11)}$$
$$PP_{wireless} = 27.214e^{-0.034(x+6)}$$
$$PP_{wireless} = 23.843e^{0.193(x+5)}$$
$$Prop_{wired} = \frac{0.317(1.05)^x}{0.683(1.03)^x + 0.317(1.05)^x}$$
$$Prop_{wireless} = \frac{0.683(1.03)^x}{0.683(1.03)^x + 0.317(1.05)^x}$$

2.3 Results

From the models our team created, we were able to predict the cost per Mbps of bandwidth in dollars from 2020 (our base year) to 2030, a span of 10 years. We wrote and ran a Python program that calculated the value of our equations above and used these values to model a price vs. time for the US and UK. Figure 2.3.1 shows the growth of US and UK bandwidth speeds while Figures 2.3.2–2.3.5 display the decrease in wired and wireless internet plan prices over time. Finally, Figure 2.3.6 and Table 2.3.1 present bandwidth prices per Mbps in dollars across the 10 years.

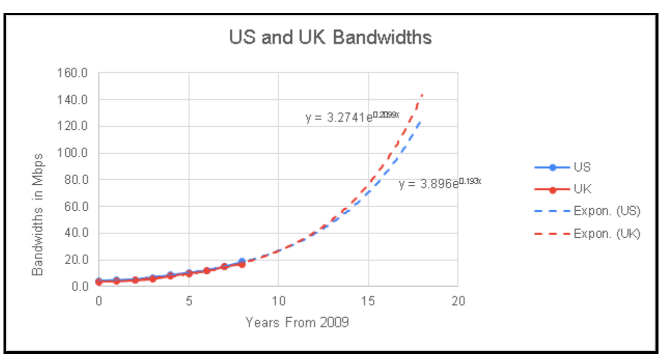

Figure 2.3.1: Graph of US and UK Bandwidth Speeds from 2009 to Projected 2030

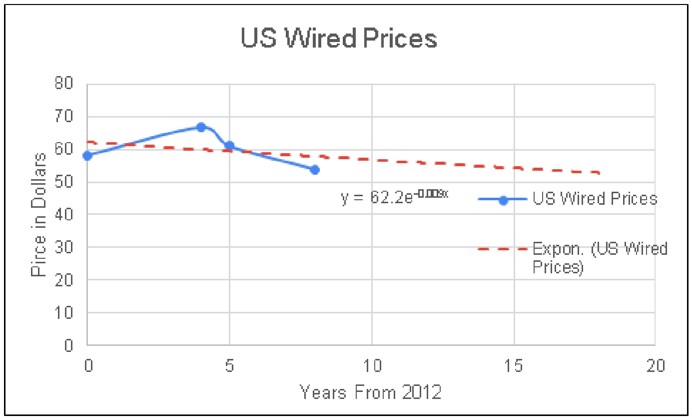

Figure 2.3.2: Graph of US Wired Internet Plan Prices from 2012 to Projected 2030

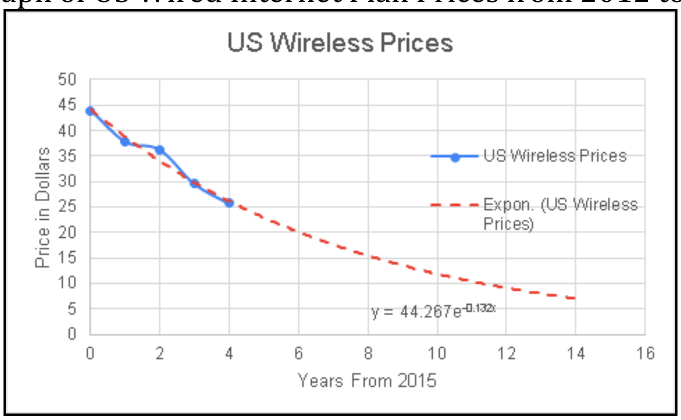

Figure 2.3.3: Graph of US Wireless Internet Plan Prices from 2015 to Projected 2030

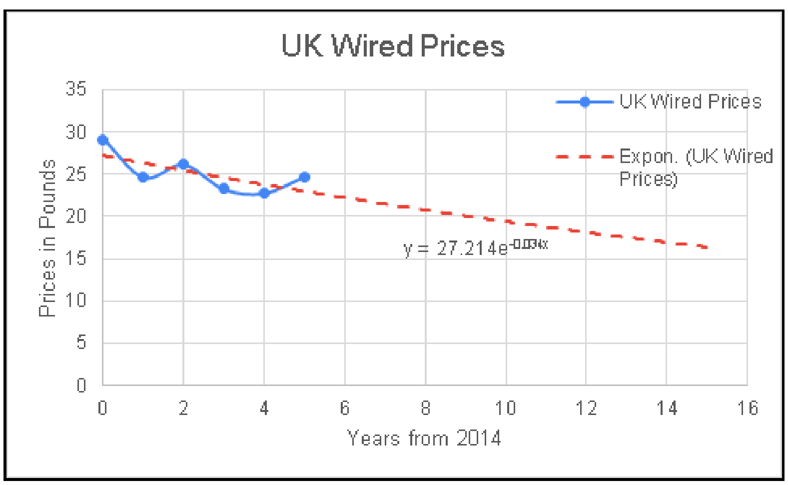

Figure 2.3.4: Graph of UK Wired Internet Plan Prices from 2015 to Projected 2030

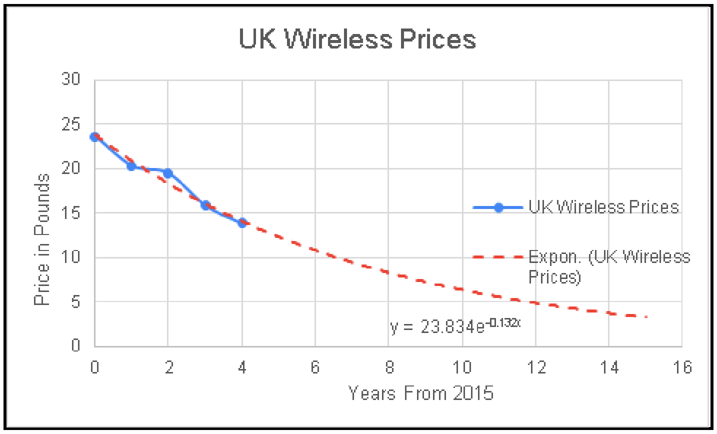

Figure 2.3.5: Graph of UK Wireless Internet Plan Prices from 2015 to Projected 2030

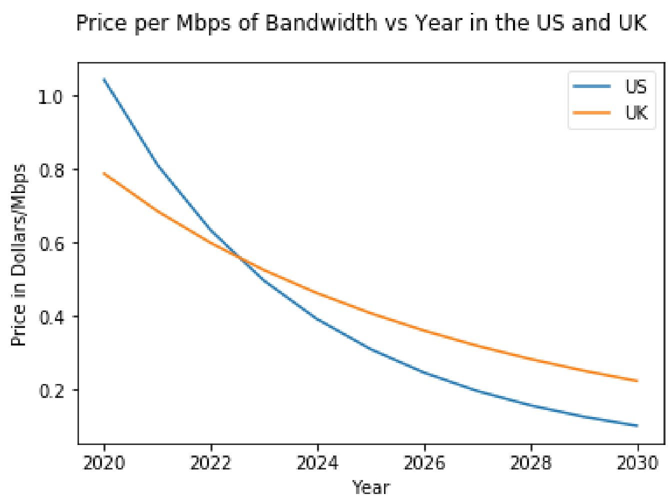

Figure 2.3.6: Graph of US and UK Price per Mbps of Bandwidth in Dollars from 2020 to 2030

Year	US Price in Dollars per Mbps	UK Price in Dollars per Mbps
2020	1.044	0.788
2021	0.811	0.686
2022	0.634	0.599
2023	0.497	0.526
2024	0.392	0.463
2025	0.311	0.408
2026	0.247	0.361

2027	0.197	0.320
2028	0.158	0.284
2029	0.127	0.252
2030	0.102	0.224

Table 2.3.1: Distribution of US and UK Bandwidth Prices per Mbps in Dollars (2020-2030)

Thus, our model concludes that across the next 10 years, the cost per unit of bandwidth in dollars per Mbps in both the US and UK will exponentially decrease, with a sharper decrease in the United States. By 2030, the US price and UK price will be about $0.10 and $0.34 per unit, respectively.

2.4 Model Evaluation

2.4.1 Validation

The results of our model are reasonable in the context of the problem. As the prices in wired and wireless plans decrease over time, as seen in the trends from the early 21st century to 2020, this trend coincides with the exponential increase in broadband speeds[10], creating a drastic decrease in price per Mbps of bandwidth as reflected in the table.

2.4.2 Sensitivity Analysis

Constant	% Change in Constant	US % Change	UK % Change
0.683 ($Prop_{wireless}$ in 2020)	+5 %	-2.04%	+1.65%
0.683 ($Prop_{wireless}$ in 2020)	-5 %	+2.49%	-3.34%
1.03 ($Prop_{wireless}$ Yearly Multiplier)	+5 %	-21.0%	+15.2%
1.03 ($Prop_{wireless}$ Yearly Multiplier)	-5 %	+25.4%	-18.4%
0.317 ($Prop_{wired}$ in 2020)	+5 %	+2.35%	-1.67%
0.317 ($Prop_{wired}$ in 2020)	-5 %	-2.35%	+1.82%
1.05 ($Prop_{wired}$ Yearly Multiplier)	+5%	+24.1%	-17.5%
1.05 ($Prop_{wired}$ Yearly Multiplier)	-5%	-22.8%	-16.0%

Table 2.4.1: Sensitivity Analysis for Part I

As seen in the table, the change in the US's price was inversely related to a change in the proportion of wireless plan customers but directly proportional to the proportion of wire plan customers while the opposite occurred in the UK.

2.5 Strengths and Weaknesses

Overall, one of the strengths of our model is that it bases the changes in price on the changes in price plans rather than individual variables like materials, labor, etc. which contributes to the simplicity and conciseness of the model. Additionally, our model is fundamentally centered around historical trends (real data) rather than future values, which provides concrete support for the models that we developed. Finally, our model is accurate in the short-term because changes in the exponent coefficients cause minimal changes (minimal error) in the country prices in the short run.

However, our model used a ratio factor to determine the US wireless prices because online data for US wireless plan prices was not available, and we instead chose to compare the ratio between mobile internet prices between the US and UK and apply this to wireless prices. The actual US wireless prices could deviate from this ratio factor, which would distort actual change in US cost per unit of bandwidth. Additionally, as seen in the sensitivity table, our model is not resistant to changes, with changes in the constants by 5% causing the results to change by nearly 25%. Moreover, because our model is an exponential function, changes in the power coefficient causes bigger changes in the variable as time passes, signifying a weakness in our model.

3 Part II: Bit by Bit

In order to better gauge the widespread necessity of high-speed internet, the respective needs of a variety of households must be quantitatively determined. However, this need greatly fluctuates depending on the two major consumption factors, work/school and leisure, and the demographic and socioeconomic makeup of individual households. Thus, we created a flexible mathematical model to calculate a household's yearly internet needs given these factors and familial makeups and utilized this model to determine three households' minimum amounts of bandwidth to cover their internet needs 90% and 99% of the time.

3.1 Local Assumptions

1. *The data collected by the Nielsen Corporation on internet media consumption by different demographics is representative of the people we are testing our model on.*
    a. **Justification:** The Nielsen Corporation is a professional company which has previously published research on the topic[11].
2. *Individuals will require the same amount of bandwidth consumption from week to week.*
    a. **Justification:** For most weeks in the month, individuals' schedules are relatively similar. For other weeks, the possible difference is assumed to be negligible.
3. *All households have their desired level of access to the internet.*
    a. **Justification:** For simplicity of the model, we will assume that demand will be fully met for all applicable households.
4. Cable television does not affect the bandwidth consumption of a household.

   a. **Justification:** Bandwidth is not used for watching traditional television because every channel is broadcast to households in the area, rather than bandwidth being specifically required to watch traditional television.
5. The demographic information of age group and household income are the best predictors of how an individual will interact with the internet.
   a. **Justification 1**: Age is capable of predicting this as different age groups are likely to have different interests and technological capabilities. For example, younger aged individuals are more likely to use the internet for gaming, and this is reflected in the data[12] as well.
   b. **Justification 2**: Household income is capable of predicting this as different levels of wealth have access to different levels of resources. For example, people with higher incomes do not need to rely on cheap entertainment, such as computer usage, as they can go outside for their entertainment more often. This lowers the amount of time they spend on bandwidth requiring activities.

3.2 Model Development

In this model, in order to determine the minimum amount of bandwidth required for different households, we decided to first find the total amount of data each person in the household would use. Different people spend different amounts of time on different activities which require different amounts of bandwidth. The amount of data an individual uses on any given activity can be found this equation:

$$D = T \cdot B$$

Here, the data used for any single activity (D) is found by multiplying the time spent on that activity (T) by the bandwidth required to perform that activity (B). Of course, since the exact time an individual spends on any given activity is not known, it must be approximated. We have assumed that the largest determinants which influence an individual's choice of activities are the age group and household income level they belong to, as explained in Assumption 3.1-5.

In order to determine the relative influence these characteristics have on an individual's choice of activities, we analyzed the data the Nielsen Corporation[12] collected on the amount of average weekly hours spent by different demographic groups on different bandwidth using activities. By performing a linear regression on time spent on each activity for each demographic, we are able to estimate the impact that demographic has on activity preference. That is, the slope of the line of best fit will have a greater magnitude if that demographic group has a larger influence on that type of bandwidth usage. Then, by comparing the magnitude of the slope of the linear regressions, we are able to find the relative weight of the demographics of age and household income compared to each other. The linear regression coefficients and derived relative weight for the demographics age and household income for each activity is listed here:

Analysis – Activity List |	Linear Regression Coefficients for Age Data (M_A)	Linear Regression Coefficients for Household Income Data (M_H)	Relative Weight of that Demographic Characteristic (W_A, W_H)
Watching	8.417142857	-7.479	(0.5295, 0.4705)

Traditional Television			
TV Connected Game Console	-0.7671428571	-1.959	(0.2814, 0.7186)
TV Connected Internet Device	-0.4962857143	-1.625	(0.2340, 0.7660)
Internet on a Computer	-0.469	-1.307	(0.2641, 0.7359)
Total App/Web on a Smartphone	-3.939	-0.813	(0.8289, 0.1711)
Total App/Web on a Tablet	0.834	-0.141	(0.8554, 0.1446)

Table 3.2.1 Calculations Used to Determine Relative Weight to Give to Demographic Factors

The relative weight of demographic characteristics compared to each other was found by adding the linear regression coefficients of both demographics for one activity, then finding the proportion of the linear regression coefficient of one demographic compared to the sum.

$$M_T = M_A + M_H$$

$$W_A = \frac{M_A}{M_T}, W_H = 1 - W_A$$

With these weights, the average amount of time that an individual spends on certain bandwidth using activities can be approximated using their age and household income demographics. By using average demographic data in conjunction with the derived demographic weights, the average amount of time an individual in a given demographic spends on a certain bandwidth using activity can be calculated.

$$T_i = W_{A_i}T_{A_i} + W_{H_i}T_{H_i}$$

For any given activity, the amount of time a person spends on said activity (T_i) is determined by summing the average amount of time people in the same two demographic groups spend on that activity (T_{A_i} for the demographic of age, T_{H_i} for the demographic of household income) times the weight of the influence that demographic has on that activity (T_{A_i} and T_{H_i} respectively).

$$D_i = T_iR_i = \left(W_{A_i}T_{A_i} + W_{H_i}T_{H_i}\right)R_i$$

Then, by multiplying this number (T_i) by the amount of bandwidth that the chosen activity uses (R_i), the total amount of data the individual spends on that activity can be calculated (D_i). To ensure the units match up, this number must be multiplied by 3600 to keep the units consistent (T is in hours, R is in megabits per second, there are 3600 seconds in an hour)

$$D = \sum\left(W_{A_i}T_{A_i} + W_{H_i}T_{H_i}\right)R_i$$

To find the total amount of data used by an individual (D), the sum of bandwidth used for all activities the individual does must be summed up. Finally, because a family consists of all members of a household, finding the sum of data used for all activities for all members of the household and combining them will result in the total amount of data used by a household (D_F).

$$D_F = \sum D$$

Then, the total amount of data used by a household can be divided by their hours of internet use to determine their bandwidth required for average usage,

$$B = \frac{D_F}{T}$$

Then, to determine the minimum amount of required data 90% of the time and 99% of the time, we used a normal distribution function. We assumed above calculated bandwidth was the average bandwidth required 50% of the time. Then, a standard deviation (P_{std}) was 20% of the calculated bandwidth. The bandwidth at 90% and 99% frequencies on the normal distribution function, corresponded with the minimum amount of required bandwidth the household needs 90% of the time and 99% of the time—our answer.

Refer to the below chart to see the different activities and bandwidth required constants used in the above equations. The explanations for how we calculated the bandwidth required for the activity is shown below.

Activity	Bandwidth Required
Watching Traditional Television	5.5 Mbps (Source: D5)
TV Connected Game Console	2 Mbps (Source: D5)
TV Connected Internet Device	6.5 Mbps (Source: D5)
Total Internet on a Computer	3 Mbps (Source: D5)
Total App/Web on a Smartphone	0.43 Mbps (Calculation Explained Below)
Total App/Web on a Tablet	0.688 Mbps (Calculation Explained Below)

Table 3.2.1 List of Activities and Bandwidth Required

We calculated an average of 5.5 Mbps used while watching traditional television, which encompasses watching standard-definition video streaming (3-4 Mbps) and high-definition video streaming (5-8 Mbps). We averaged these values to calculate an average of 5.5 Mbps.

We conducted a similar procedure when calculating the bandwidth used for each remaining TV category. For TV-connected game consoles, online gaming requires 1–3 Mbps, or an average of 2 Mbps. For TV-connected internet devices like Apple TV and Amazon Fire, which offer high-definition video streaming (5–8 Mbps), we calculated an average bandwidth used of 6.5 Mbps.

Meanwhile, in calculating internet and app usage for computers, smartphones, and tablets, we utilized monthly data usage (in gigabytes and megabytes) in order to calculate the average bandwidth required (in Mbps). We cannot assume that the Total App/Web Speed and Total Internet used on a smartphone, tablet, or computer will fall into any one of the categories given in Data Table 5 given that activities on these devices can range from texting (which uses very little amounts of data) to video streaming (which uses large amounts of data). As a result, a more detailed analysis is required.

To calculate Mbps used on a smartphone we used the following methodology: On average, an American spends about 5.4 hours on their phone daily[13]. Also, an average American uses about 251,200 Mb per month (converted from 31.4 GB mentioned in the article)[14]. Therefore, we calculate that the average App/Web consumption of a smartphone in Mbps would be 251200/[(5.4 (hrs/day)*30(average days in a month)*60 min/hr*60 seconds/hr] = 0.43 Mbps

To calculate Mbps on a tablet: According to data, bandwidth usage on a tablet is approximately 160% more than that of a phone. Thus, we can multiply our value from the smartphone by 1.6 to get the bandwidth consumed by a tablet. 0.43 * 1.60 = 0.688[15].

Common uses of the computer include internet related purposes such as general web surfing, standard definition video streaming, and online gaming. These are all low bandwidth uses [D5], so the assumption of 3Mbps is a reasonable one.

3.2.1 Defined Variables and Parameters in the Model

Variable	Definition	Units
D, D_i, D_f	Total Data Used by an Individual, Data Used for a Single Activity, Total Data Used by a Family	Megabits
T, T_i, T_{Ai}, T_{Hi}	Time Spent, Time Spent for a Single Activity, Average Time Spent on an Activity by the Demographic Group of Age, Average Time Spent on an Activity by the Demographic Group of Household Income	Hours
M_A, M_H, M_T	Linear Regression Coefficient for Age Data, Linear Regression Coefficient for Household Income Data, Combined Linear Regression Coefficient	N/A
W_A, W_H	Weight of Age on Chosen Activity Levels, Weight of Household Income on chosen Activity Levels	N/A
R_i	Bandwidth Required for Given Activity	Megabits/Second
B	Bandwidth	Megabits/Second
P_{std}	Constant, Standard Deviation of Mean (Assumed to be 20% of calculated D_f)	0.2

3.3 Results

By using linear regression to determine the influence of the demographic characteristics of age and household income on internet activity, we were able to predict a household's total data usage. Then, we used a normal distribution to determine the probability of how much bandwidth would be used by a household by using the household's total data requirement as a baseline (total data was converted to bandwidth by dividing by hours spent on the internet). This allowed us to determine the minimum bandwidth of the household for 90% and 99% of scenarios by finding the value at 90% and 99% of the normal distribution, respectively.

Situation 1:

	TV Connected Game Console	TV Connected Internet Device	Internet on a Computer	Total App/Web on a Smartphone	Total App/Web on a Tablet
Teacher	(0.2814*3.73)+ (0.7186*5.34) = 4.89	(0.234*5.33)+ (0.766*6.21) = 6.00	(0.2641*4.57)+ (0.7359*9.12) = 7.92	(0.8289*24.67)+ (0.1711*12.05) = 22.51	(0.8554*4.37)+ (0.1446*6.97) = 4.75
Unemployed Spouse	(0.2814*3.73)+ (0.7186*5.34) = 4.89	(0.234*5.33)+ (0.766*6.21) = 6.00	(0.2641*4.57)+ (0.7359*9.12) = 7.92	(0.8289*24.67)+ (0.1711*12.05) = 22.51	(0.8554*4.37)+ (0.1446*6.97) = 4.75
3-Year-Old Child	(0.2814*2.92)+ (0.7186*5.34) = 4.66	(0.234*5.55)+ (0.766*6.21) = 6.06	(0.2641*0)+ (0.7359*9.12) = 6.71	(0.8289*0)+ (0.1711*12.05) = 2.06	(0.8554*0)+ (0.1446*6.97) = 1.01
Total Hours per Week	14.43	18.06	22.55	47.08	10.50
Total Required Mb per Week	103,917	422,691	243,519	72,885	26,006

Total Hours for Household per Week = 181.64175 hours
Total Mb for Household per Week = 869,018.0549 Mb
Mean Mbps = 1.230
90% Effectiveness: 1.230 + (0.2)(1.230)(1.282) = 1.545 Mbps / 0.5 = 3.090 Mbps
99% Effectiveness: 1.230 + (0.2)(1.230)(2.326) = 1.802 Mbps / 0.5 = 3.604 Mbps

The median teacher salary is $64,524, making this household's income fall between $50,000 and $75,000[17].

Situation 2:

	TV Connected Game Console	TV Connected Internet Device	Internet on a Computer	Total App/Web on a Smartphone	Total App/Web on a Tablet
Retired Woman	(0.2814*0.15)+ (0.7186*9.89) = 7.14	(0.234*2.17)+ (0.766*9.14) = 7.51	(0.2641*3)+ (0.7359*12.01) = 9.63	(0.8289*13.37)+ (0.1711*13.14) = 13.33	(0.8554*7.12)+ (0.1446*6.84) = 7.08
Grandchildren	(2)*(2/7)*[(0.2814*4.08)+(0.7186*9.89)] = 4.72	4.46	5.05	1.28	0.57
Total Hours per Week	11.86	11.97	14.68	14.62	7.64
Total Required Mb per Week	85,438	236,974	158,553	22,625	18,934

Total Hours for Household per Week = 126.7612187 hours
Total Mb for Household per Week = 522,523.3573 Mb
Mean Mbps = 1.145 Mbps
90% Effectiveness: 1.145 + (0.2)(1.230)(1.282) = 1.439 Mbps / 0.5 = 2.878 Mbps
99% Effectiveness: 1.145 + (0.2)(1.230)(2.326) = 1.678 Mbps / 0.5 = 3.356 Mbps

Situation 3:

	TV Connected Game Console	TV Connected Internet Device	Internet on a Computer	Total App/Web on a Smartphone	Total App/Web on a Tablet
Each Student	(0.2814*3.73)+ (0.7186*7.83) = 6.68	(0.234*5.33)+ (0.766*7.94) = 7.33	(0.2641*4.57)+ (0.7359*9.71) = 8.35	(0.8289*24.67)+ (0.1711*12.05) = 22.51	(0.8554*4.37)+ (0.1446*7.12) = 4.77
Total Hours per Week	20.03	21.99	25.06	67.53	14.30
Total Required Mb per Week	144,207	514,514	104,540	104,540	35,426

Total Hours for Household per Week = 226.979787 hours
Total Mb for Household per Week = 1,069,308.431 Mb
Mean Mbps = 1.309 Mbps
90% Effectiveness: 1.309 + (0.2)(1.230)(1.282) = 1.645 Mbps / 0.5 = 3.290 Mbps
99% Effectiveness: 1.309 + (0.2)(1.230)(2.326) = 1.918 Mbps / 0.5 = 3.836 Mbps

The median wage/salary of a part-time worker is $307 per week, or $15,964 per year[18]. This translates to a household income of $47,892.

For all three situations, the calculated minimum Mbps was divided by 0.5 because the average WiFi speed experienced in one's home will generally be 50% below the maximum download speed due to wireless interference[19]

3.4 Model Evaluation

3.4.1 Validation

Relative to one another, the results of our model make sense in the context of the scenarios provided. For example, it is reasonable that the working and learning college student household in scenario 3 would require a greater bandwidth than the retired woman and her grandchildren in scenario 2. The household in scenario 3 would require greater amounts of internet for their choice of internet usage. Furthermore, numerically, the calculated bandwidths of roughly 3 Mbps would be capable of popular internet activities such as general web surfing, standard video streaming, and video conferencing. The numbers independently seem reasonable as it is a high enough number to be capable of most tasks that require the internet. Furthermore, relative to each other, the numbers seem reasonable, as groups that seem to require more internet are appropriately modeled to have more bandwidth than groups that seem to require less internet.

3.4.2 Sensitivity Analysis

Constant ->	P_{std}	P_{std}
Percentage Change	+5%	-5%
Situation 1 (90%)	+1.04%	-0.96%
Situation 1 (99%)	+1.60%	-1.58%

Situation 2 (90%)	+0.99%	-1.05%
Situation 1 (99%)	+1.57%	-1.61%
Situation 3 (90%)	+1.01%	-1.04%
Situation 3 (99%)	+1.08%	-1.59%

3.4.2 Strengths and Weaknesses

One of the strengths of our model is the fact that it is based on real demographic data. This is most evident in the methodology we used to calculate the weights we assigned to the demographics—those coefficients have a real world basis. Another strength of our model is that it is not sensitive to change. For our constant P_{std}, the output of the model changes by a lesser amount than we changed it by. Since it was difficult to find a specific number to use for P_{std}, a low percentage change is desirable because it indicates that new data (which might significantly change P_{std}) will not disrupt the model overall.

Due to limited data availability, our model displays weaknesses. Better estimates for bandwidth requirement could be found for each activity by breaking down each activity into proportionate sub-sections which could more accurately describe bandwidth requirement. For example, the activity "Total Internet on a Computer" could be broken down into time spent on online gaming, video conferencing, and streaming by using the popularity of each smaller activity. Another weakness our model includes our use of a normal distribution. However, bandwidth usage is more likely skewed to the left. Factors such as high cellphone usage, which require low amounts of bandwidth and distort the mean, support this. This would imply that the minimum required bandwidth should be higher than what the model indicates.

4 Part III: Mobilizing Mobile

Even with an understanding of bandwidth cost and household needs, mobile broadband depends on the effective implementation of cellular nodes across a region to maximize broadband coverage. As such, we developed a model to optimize the distribution of cellular nodes and tested this model with three hypothetical regions.

4.1 Local Assumptions

1. *All nodes in each region will be placed at locations throughout each subregion which will cover the maximum distance possible.*
    1. **Justification:** Our model will not account for overlap of nodes thus we are assuming that each node will be placed at positions to cover the maximum amount of area and disregard any overlap in service.
2. *The population is spread evenly over every subregion.*
    1. **Justification:** There is no data available which describes the distribution of the population within different parts of each subregion.
3. *We assume that the median age and median household will be representative of the entire population.*

1. **Justification:** Given that the population will likely be skewed in a particular direction, the median will be the best value to represent the population (Also, this is the only value provided in the data).

4. *We assume that each band will operate at max capacity in terms of range and Mbps.*
    1. **Justification:** This assumption will aid in simplicity of the model.

4.2 Model Development

Our second model is able to predict the demand of an individual based on his/her age and income bracket. Given this information we will use the median age and median household income to calculate the bandwidth demand of an individual in a particular subregion. After multiplying this value by the total population of the subregion we will be able to approximate the bandwidth requirement for the entire subregion.

The model will then determine the minimum number of nodes required outputting low band for each subregion which will cover the total area of each subregion, returning the value N_A. Next, it will calculate the minimum number of nodes outputting low band required to meet the bandwidth requirement of each subregion returning the value N_B. The model will select whichever value is greater as it will satisfy both of the requirements (bandwidth and area). It will then calculate this for each of the three different bands. Depending on whichever number of nodes is the lowest out of the three different bands we are able to determine the number of nodes required in a particular subregion.

4.2.1 Defined Variables and Parameters in the Model

Variable	Definition	Units
A	Median Age	Years
H	Median Household Income	$
P	Population	# of people
Q	Area	miles^2

4.2.2 Equations Utilized in Model

We reuse the equations we used to calculate bandwidth demand in question 2 to find regional bandwidth demand.

$$D = \sum\left(W_{A_i}T_{A_i} + W_{H_i}T_{H_i}\right)R_i$$

4.3 Results

	Subregion 1(A)	Subregion 2(A)	Subregion 3(A)	Subregion 4(A)	Subregion 5(A)	Subregion 6(A)
Demands in mbps	2312.919	4766.625	4608.282	1080.053	4396.081	4919.509
Optimal Number of Nodes and band	1 node - high band	2 nodes - high band	2 nodes - high band	1 node - high band	1 node - high band	2 nodes - high band

	Subregion 1(B)	Subregion 2(B)	Subregion 3(B)	Subregion 4(B)	Subregion 5(B)	Subregion 6(B)	Subregion 7(B)
Total Mbps	12812.33	6993.358	4678.729	3734.863	9308.915	7932.863	6698.834
Optimal Number of Nodes and band	5 nodes - High band	3 nodes - high band	2 - high band	2 - high band	4 - high band	3 - high band	3 - high band

	Subregion 1(C)	Subregion 2(C)	Subregion 3(C)	Subregion 4(C)	Subregion 5(C)	Subregion 6(C)	Subregion 7(C)
Total Mbps	4856.315	4964.867	3347.814	4284.01	4001.854	3334.581	5918.22
Optimal Number of Nodes and Band	2 nodes - high band	2 nodes - high band	2 nodes - high band	2 nodes - high band	2 nodes - high band	2 nodes- high band	2 nodes - high band

4.4 Model Evaluation

4.4.1 Validation

As can be seen in the results of all of the regions, we can see that the model predicted for the implementation of only high band nodes. This prediction makes sense as the regions in the database provided are all very population dense. In other words, the population is very large per unit of area and in population dense areas it would make sense to use high band nodes. Our model, however, would predict the use of low band nodes in areas where there is a very low population density as area of coverage would become a much larger factor than the Mbps.

4.4.2 Strengths and Weaknesses

Strengths: This model was able to recognize that a high population density would result in the implementation of high band nodes.

Weaknesses: Ideally on a smaller scale one would want a model to consider how a particular node could interface over multiple regions. However, at a scale as small as the regions provided both the mid and low bands could cover the entirety of the regions.

5 Conclusion

5.1 Summary and Analysis

Overall, our first model centers on determining the cost per unit of bandwidth in dollars per Mbps in both US and UK. It demonstrates that the historical trends in decreasing price plans for both wired and wireless internet connections as well as the rapidly increasing speeds of bandwidth will in fact cause the cost per Mbps of bandwidth to decrease in the next 10 years: from $1.04 to $0.102 in the US and from $0.788 to $0.224. Thus, these quantitative results hint at a positive outlook for the increased digitization of our world, noting that at the very least, the relative cost of connectivity will decrease which will improve access to those in rural and low-income areas.

Next, our second model focuses on calculating the individual internet and bandwidth needs for a year of a randomly chosen household, especially with a newfound awareness for the global shifts in online education and work. Because data amount usage is equal to the time spent using a certain network bandwidth, our model analyzed age and household income levels in relation to time utilization and we performed linear regressions of [D4] data to relatively weigh each factor on their respective impact on time consumption of a bandwidth. From this analysis, we determined that the former challenge students in scenario 3 demanded the most amount of bandwidth while the retired woman and the grandchildren required the least amount of bandwidth, which makes sense considering our new focus on online education and the greater use of technology and the internet among teenagers and young adults.

Finally, the third model optimizes the distribution of mobile broadband transmitting cellular nodes in a region. Once the bandwidth demand of each subregion was calculated, the model can determine the minimum number of nodes outputting low, middle, or high band required to satisfy the bandwidth and area requirements. At the end, whichever number of nodes is the lowest of the three bands is chosen for the particular subregion. In the end, our model predicted that subregions 1, 4, 5 (A) require 1 node of high band while subregions 2, 3, 6 (A) require 2 nodes of high band, subregions 2, 6, 7 (B) require 3 nodes of high band while subregions 3, 4 (B) require 2 nodes of high band, subregion 1(B) needs 5 nodes of high band, and subregion 5 (B) needs 4 nodes of high band, and finally all subregions of C require 2 nodes of high band. The prevalent use of high band is reasonable to maximize broadband coverage to minimize the number of nodes.

5.2 Takeaways

As our world progresses toward a more connected, digitized realm, what was once regarded as a simple commodity now transforms itself into a fundamental need: access to high-speed internet. A disparity further emphasized by the coronavirus pandemic, the versatile nature of high-speed nature has irrefutably entrenched itself in our academic, healthcare, economic, and informational lives, which necessitates the addressment of the technical, logistical, and economic challenges enhancing high-speed internet access entails. Thus, by accounting for both wired and wireless connections to the internet amongst people in all different households and regions as well as balancing bandwidth costs with individual household needs to accurately and optimally provide broadband coverage, we can certainly take a complete leap forward toward shared high-speed internet … together.

6 References

[1] - https://m3challenge.siam.org/node/523

[2] - https://en.wikipedia.org/wiki/Fiber-optic_communication#:~:text=First%20developed%20in%20the%201970s,advent%20of%20the%20Information%20Age

[3] - https://unctad.org/news/coronavirus-reveals-need-bridge-digital-divide

[4] - https://en.wikipedia.org/wiki/Fiber-optic_communication#:~:text=First%20developed%20in%20the%201970s,advent%20of%20the%20Information%20Age

[5] - [D3] Cost for Bandwidth (UK, wired and wireless)

[6] - https://www.statista.com/statistics/217348/us-broadband-internet-susbcribers-by-cable-provider/

[7] - https://www.statista.com/statistics/273412/market-share-of-uk-telecoms-operators-since-2007-by-fixed-broadband-subscribers/

[8] - https://www.forbes.com/sites/niallmccarthy/2019/03/05/the-cost-of-mobile-internet-around-the-world-infographic/?sh=6e1004e9226e

[9] - https://knoema.com/atlas/United-States-of-America/topics/Telecommunication/Telecomm-Services/Fixed-broadband-subscriptions

[10] - https://www.techrepublic.com/article/86-of-us-states-increased-internet-speeds-during-covid-19-lockdown/

[11] - https://www.nielsen.com/wp-content/uploads/sites/3/2019/04/total-audience-report-q3-2015.pdf

[12] - [D4] Internet Media Consumption

[13] - https://techjury.net/blog/how-much-time-does-the-average-american-spend-on-their-phone/#gref

[14] - https://www.fiercewireless.com/wireless/alarming-unlimited-data-usage-31-4-gb-per-month-and-rising

[15] - https://pocketnow.com/phones-vs-tablets-how-does-mobile-data-usage-compare

[16] - [D5] Bandwidth requirements for various activities

[17] - https://www.usatoday.com/story/money/2020/09/10/states-where-teachers-are-paid-the-most-and-least/113660742/

[18] - https://www.bls.gov/cps/cpsaat38.htm

[19] - https://broadbandnow.com/guides/how-much-internet-speed-do-i-need

7 Appendix

7.1 Part I: The Cost of Connectivity

```python
import numpy as np
import matplotlib.pyplot as plt
import math

poundtoDollar = 1.39

def func(t):
    x = t-2020
    Propwireless = 0.683*pow(1.03,x)
    Propwired = 0.317*pow(0.95*1.05,x)
    total = Propwired+Propwireless
    usAS =  3.896*pow(math.e, 0.193*(x+11))
    ukAS = 3.274*pow(math.e, 0.299*(x+11))
    usPPwired = 62.2*pow(math.e,-0.009*(x+8))
    usPPwireless = 44.267*pow(math.e,-0.132*(x+5))
    ukPPwired = 27.214*pow(math.e, -0.034*(x+6))
    ukPPwireless = 23.843*pow(math.e, 0.193*(x+5))
    usPrice = (Propwireless*usPPwireless+Propwired*usPPwired)/(total*usAS)
    ukPrice = poundtoDollar*(Propwireless*ukPPwireless+Propwired*ukPPwired)/(total*ukAS)
    return(usPrice, ukPrice)



fig = plt.figure(facecolor='w')
ax = fig.add_subplot(111, axisbelow=True)
x = np.linspace(2020, 2030, 11)
rows = ['%d Year' % y for y in x]
US, UK = func(x)
ax.set_ylabel('Price in Dollars/Mbps')
ax.set_xlabel('Year')
ax.plot(x, US, label = 'US')
ax.plot(x, UK, label = 'UK')
fig.suptitle('Price per Mbps of Bandwidth vs Year in the US and UK')
legend = ax.legend(loc='best')
legend.get_frame().set_alpha(0.5)
plt.show(block=True)
```

7.2 Part II: Bit by Bit

No Code Available

7.3 Part III: Mobilizing Mobile

```
import math

A_subregions_area = [1.21,0.8,0.67,1.65,0.36,2.14]
A_subregions_mbps = [2312.919, 4766.625, 4608.282, 1080.053, 4396.081, 4919.509]

B_subregions_area = [3.48, 4.35, 4.64, 2.32, 7.54, 3.77, 7.54]
B_subregions_mbps = [12812.33, 6993.358, 4678.729, 3734.863, 9308.91 ,7932.863, 6698.834]

C_subregions_area = [0.38, 0.14, 0.1, 0.24, 0.13, 0.31, 0.34]
C_subregions_mbps = [ 4856.315, 4964.867, 3347.814, 4284.01, 4001.854, 3334.581, 5918.22]


#calculate the area in square miles that a particular band can serve
lowRange = math.pi * 400
midRange = math.pi * 9
highRange = math.pi * 1

#mbps that a band can support
lowMbps = 250
midMbps = 900
highMbps = 3000


#calculate minimum number of nodes purely based on area coverage per different band

for i in A_subregions_area:
    print((i)/lowRange)
    print((i)/midRange)
    print((i)/highRange)

for i in B_subregions_area:
    print((i)/lowRange)
    print((i)/midRange)
    print((i)/highRange)

for i in C_subregions_area:
    print((i)/lowRange)
    print((i)/midRange)
    print((i)/highRange)

#calculate minimum number of nodes purely based on mbps coverage per band

for i in A_subregions_mbps:
    print((i)/lowMbps)
    print((i)/midMbps)
    print((i)/highMbps)

for i in B_subregions_mbps:
    print((i)/lowMbps)
    print((i)/midMbps)
    print((i)/highMbps)

for i in C_subregions_mbps:
    print((i)/lowMbps)
    print((i)/midMbps)
    print((i)/highMbps)
```