MathWorks Math Modeling Challenge 2021

Julia R Masterman Middle High School

Team # 14665 Philadelphia, Pennsylvania

Coach: Kathryn Smith

Students: Tobias Beidler-Shenk, Tanay Bennur, Hayden Gold, Owen Moss, Ethan Soloway

M3 Challenge FINALIST—$6,500 Team Prize

JUDGES' COMMENTS

Specifically for Team # 14665 —Submitted at the Close of Triage Judging:

COMMENT 1: The executive summary is well written, however it lacks details on the precise results and recommendations the team came up with after their work. In Problem 1 the team defined their parameters and provided their local assumptions. The model was well motivated and provided clear results. A partial sensitivity analysis and strengths and weaknesses of the model are included. In Problem 2 the team invented a survey, which I thought worked really well. They also defined their parameters and provided their local assumptions. The model was well motivated and provided clear results. Strengths and weaknesses of the model are included. In Problem 3, the team provided assumptions (with justification), and variables, and the results. Visuals helped! There is an attempt at providing a common introduction and a common conclusion to the project.

COMMENT 2: Overall, your executive summary was extremely well-written. It provided a convincing argument for why the three questions are worth answering, and you shared a bit about your modeling and methodology, which was good. The only thing missing was some specifics on your findings—a brief summary of your results. Nonetheless, it was one of the best executive summaries I've read.

COMMENT 3: You did a particularly fantastic job of defining terms, sharing global assumptions, and discussing your results throughout your paper. The use of a survey and your other methods deployed to answer question 2 combined for probably the most interesting solution I came across for that particular question. And your work in question 3, including utilization of a weighted k-means algorithm, was creative and demonstrated solid critical thinking skills. All in all, a fantastic job on this paper!

COMMENT 4: You presented good critical and analytical thinking. Good job with assumption statements.

COMMENT 5: There were some very innovative ideas in this project including separating out mobile and broadband, designing a questionnaire and the method of weighting the tower locations. One suggestion is that graphs show a lot of information very quickly.

COMMENT 6: Nice k-mean algorithm. Nice writing and format.

Earth, Water, Air, Fire, and Ethernet

Executive Summary

As the world entered the throes of the coronavirus pandemic in Q1 of 2020, hundreds of millions of Americans and Brits found themselves working and learning from home – that is, over the internet. However, the coronavirus was not so much a cause as it was an exacerbation of our increasing dependence on the internet: over the last decade, internet usage in the United States has more than doubled [4]. Today, the internet is used for purposes far beyond professional and educational pursuits. We use the net to access healthcare, banking, shopping, entertainment, and social networking (from email to Facebook to Zoom) – not to mention to access its original, overarching purpose: information. Despite the fact that internet access has become a necessity to contemporary life, many millions of people do not have adequate speed or quality of service. There is a driving question in today's hyperconnected world: how will internet service providers in the US and the UK overcome economic and infrastructural challenges to provide all of their citizens with sufficient access?

We first worked to model the changing cost of internet service. Key to this section is the fact that changes to the quality of technology often far outpace changes to its price [13]. The internet is generally consumed and provided in two distinct settings and set-ups: broadband (in-home) and mobile (cellular). Thus, we separated our data along this line and ran every calculation twice. Working on the assumption that no new, significantly different internet technology will be deployed in the next decade, but smaller improvements will continue to take place, we modeled the price per unit of bandwidth, or peak data speed, each year from today to 2031.

For our second model, we strove to predict the amount of bandwidth required to satisfy a variety of realistic, hypothetical households. Understanding the vast differences in usage patterns and device ownership from family to family, we devised an eight-question survey about internet use with numerical answers. For the three given households, we answered these questions using data trends for the relevant demographics. We used these numerical answers to create a function of a household's residents, devices, and internet activities that determines the minimum amount of required bandwidth that would satisfy that home 99 percent and 90 percent of the time.

Finally, for the third model, we sought to create a plan for the optimal distribution and location of mobile cellular data towers in a given region. We used a Python script to divide each region into equally spaced nodes, and weighted the nodes according to the region's population density. We then used a weighted k-means algorithm to determine the optimal placement of the towers.

We believe our models provide relevant, current predictions into the rapidly evolving internet (and internet technology) market. We have made both specific and global conclusions; generally, these are resistant to reasonable changes in parameters. Viewed cohesively, our three models give insight into that true question for our era: how will the internet and the way in which we use it change – and how well will the world change with it?

Also included at the end of this paper is an appendix, which contains Python and MATLAB code detailing the algorithms used in model three.

Contents

Introduction

Certain global considerations must be taken into account for the modeling process to be effective. Global assumptions, which will hold for the entirety of this paper, follow.

0.1 Global Assumptions and Justifications

1. *The conversion rate between the US dollar and the UK pound will remain the same in the future.* Although both countries will experience individual volatility (due to factors such as recession and natural disasters), these episodes are not possible to meaningfully predict. Because inflation will affect the two countries at a generally similar rate, we keep this conversion (given in 1.3.1) constant throughout the paper and throughout our predictions.

2. *We only consider download speed.* Consumers generally choose and pay for internet plans based on download speeds, not upload speeds. Additionally, downloading is the primary use of the internet, and upload speed is almost always adequate [3].

3. *Broadband and in-home internet are synonymous.* These terms, meaning the internet hook-up at a physical, fixed location – for our purposes, a home – are generally used interchangeably. For the purposes of our model, these are synonymous; we will generally use simply "broadband."

4. *Cellular data, mobile internet, and data are synonymous.* These terms, meaning the internet accessed through a mobile device on a cellular network, are generally used interchangeably. For the purposes of our model, these are synonymous; we will generally use simply "mobile."

5. *Great Britain and the United Kingdom are the same place.* Though this paper will use "United Kingdom" whenever possible, we sometimes write "Great Britain" to conform to the ISO code for the pound sterling. Both terms refer to England, Wales, Scotland, and Northern Ireland.

1 Part 1: The Cost of Connectivity

Internet bandwidth, or the maximum rate of data transfer from an internet service provider into a consumer home, comes at a cost. When service providers set their prices – and when consumers choose a plan, deciding what to pay – they must take into account factors ranging from improvements to infrastructure, changes to usage patterns (although, as shown with the onset of the pandemic, these are not always possible to predict), and even inflation.

1.1 Restatement of the Problem, Part 1

The problem we are tasked with addressing is as follows:

- Predict the cost per unit of bandwidth, in US dollars per megabits per second of bandwidth (Mbps) over the next ten years for American and British consumers.

1.2 Assumptions and Justifications

1. *We neglect installation fees and equipment rentals.* Though many service providers may charge a first-time activation fee and many consumers rent internet equipment such as routers and modems, these charges are not related to internet speed.

2. *We model the internet speeds that consumers are billed for, not the bandwidth that consumers actually experience.* Some consumers report significantly lower speeds than they should be receiving, but this is often due to factors beyond the control of the service provider, such a non–optimally placed router or outdated equipment [2]. The *peak* speeds we report are assumed to be these billed-for speeds.

3. *There will be no 6G internet in the next 10 years.* The deployment of 6th-generation mobile internet is not scheduled until the mid-2030s [5].

4. *We neglect bundling.* We define bundling as a service in which a provider offers a cellular and broadband in-home internet on the same account for generally a lower price. Because bundling is essentially an artificial discount, is not taken into account.

5. *The rate of inflation is constant.* The changes to the rate of inflation are unpredictable, but though it has fluctuated, this rate of change of inflation has been near constant for the last ten years.

6. *Bandwidth is defined as peak speed.* We will use "peak speed" and "average peak speed" in this section.

1.3 The Model

Because mobile and broadband internet are used, paid for, and infrastructured differently – though, as we will show, follow similar trends in change of speed and price – we chose to separate these in our model.

1.3.1 Parameters

Symbol	Definition	Units	Value
R	The number of US dollars per GB pound [1]	...	1.39
I_{US}	Average inflation rate of the US dollar	...	2.16%
I_{UK}	Average inflation rate of the GB pound	...	2.67%
t	Time in years after 2021	years	...
L	Location: US or UK	...	...
S	Service type: mobile or broadband	...	...
$A_{L,S}$	Current average peak speed of service type S at location L	Mbps	(Multiple)
$M_{L,S}$	Rate of increase in average peak speed of S at L	Mbps/year	(Multiple)
$N_{L,S}$	Current average cost of S at L	...	(Multiple)
$C(L,S,t)$	Cost function per month of S at L at time t	...	...
$D(L,S,t)$	Average peak data speed function of S at L at time t	Mbps	...
$P(L,S,t)$	Cost function per month per unit of speed of S at L	...	...

1.3.1.1 Values of $A_{L,S}, M_{L,S}$, and $N_{L,S}$

Parameter	US Mobile	US Broadband	UK Mobile	UK Broadband	Citation
$A_{L,S}$	67.3	173.7	59.2	81.1	[8]
$M_{L,S}$	13	31	17	19	[8]
$N_{L,S}$	56 [10]	59.99 [6]	47 [7]	34.78 [6]	(Individual)

1.3.2 Developing the Model

In order to predict the 10-year change in the cost per unit of peak speed of data, we separated our model into two constituent functions of time: average cost of an internet service plan, paid monthly, and average peak data speed, always given in megabits per second (Mbps). We further separated these functions into broadband (in-home) and mobile (cellular) internet, because these are generally purchased, billed, provided, and used separately – not to mention the far more rapid emergence of mobile networking in the last decade, relative to the generally staid broadband sector.

Examining the changes in average peak data speed for Americans and Brits over time, an interesting phenomenon emerged. After the release of a new generation of mobile internet technology (3G, 4G, 5G), the rate of change of the average peak data speed would increase at an inflection point.

Accordingly, we modeled the change of the average peak data speed as a piecewise function, particularly after the release of 5G, with the slope remaining constant after each new generation was released. Using linear regression over the past three years since 5G was deployed [8], we found the rate at which average peak data speed has been increasing over that time. Using the slope and intercepts (current average peak data speeds), we were able to predict data speeds up to 10 years in the future. This data speed function is

$$D(L,S,t) = M_{L,S} * t + A_{L,S},$$

which is simply the product of the yearly rate of future change and the desired number of years, added to the current average peak speed. This function was run four times for each combination of location and service type.

For cost, we found that the price of internet service plans (as a whole) have stayed relatively constant with inflation [11]. Thus, using current costs of plans plus inflation rates for different countries, we were able to predict the future cost of mobile and broadband plans. This cost function is

$$C(L, S, t) = N_{L,S} * (1 + I_L)^t,$$

which is the current average cost of service adjusted for yearly predicted inflation. This function was run four times for each combination of location and service type as well.

1.4 Results

To determine the monthly cost per unit of peak speed, we took the ratio of each cost function output to the corresponding peak speed function output, at discrete intervals on t from 1 to 10; that is,

$$P(L, S, t) = \frac{C(L, S, t)}{D(L, S, t)}.$$

The results of these calculations are in the tables below, where time is in years after 2021, peak speed is in Mbps, and price/unit is always in USD per one Mbps (for the UK data, values are converted to USD using R as given in 1.3.1).

US Broadband

Time	0	1	2	3	4	5	6	7	8	9	10
Cost	59.99	61.29	62.61	63.96	65.34	66.75	68.20	69.67	71.17	72.71	74.28
Peak Speed	173.7	204.7	235.7	266.7	297.7	328.7	359.7	390.7	421.7	452.7	483.7
Price/Unit	0.345	0.299	0.266	0.240	0.219	0.203	0.190	0.178	0.169	0.161	0.154

US Mobile

Time	0	1	2	3	4	5	6	7	8	9	10
Cost	56.00	57.21	58.45	59.71	61.00	62.31	63.66	65.04	66.44	67.88	69.34
Peak Speed	67.3	80.3	93.3	106.3	119.3	132.3	145.3	158.3	171.3	184.3	197.3
Price/Unit	0.832	0.712	0.626	0.562	0.511	0.471	0.438	0.411	0.388	0.368	0.351

UK Broadband

Time	0	1	2	3	4	5	6	7	8	9	10
Cost	34.78	35.71	36.66	37.64	38.65	39.68	40.74	41.82	42.94	44.09	45.27
Peak Speed	59.2	78.2	97.2	116.2	135.2	154.2	173.2	192.2	211.2	23.2	249.2
Price/Unit	0.588	0.457	0.377	0.324	0.286	0.257	0.235	0.218	0.203	0.192	0.182

UK Mobile

Time	0	1	2	3	4	5	6	7	8	9	10
Cost	47.00	48.25	49.54	50.87	52.22	53.62	55.05	56.52	58.03	59.58	61.17
Peak Speed	59.2	76.20	93.2	110.2	127.2	144.2	161.2	178.20	195.2	212.2	229.2
Price/Unit	0.794	0.633	0.532	0.462	0.411	0.372	0.342	0.317	0.297	0.281	0.267

It is clear that since prices will, adjusted for inflation, generally stay constant, while peak speeds will increase, the price per unit of peak speed will decrease in the next ten years. Indeed, we predict that the price per unit of peak speed of broadband in the United States, as one example, will more than halve in the next decade.

1.5 Evaluating the Model

1.5.1 Validation

We believe our predictions to be reasonable and valid because of the state of the broadband and mobile internet markets in both the US and UK, which is essentially an oligopoly: controlled nearly completely by just a few companies. This means, as one service provider rolls out a new technology (networking infrastructure being the primary reason for the price drop in our model), the others will be quick to follow; more generally, as one provider makes changes to their services or or price, the others will be obliged to respond. We believe this market structure actually enables competition, because of the such limited number of providers involved. Considering this, as well as the well-documented historical trend that technology's quality (in this case peak speed of the internet, but in other cases, say, size of a processor or resolution of a display) changes at a far greater rate than its price [13].

1.5.2 Sensitivity Analysis

Good models must be resilient to changes in certain parameters. To test the efficacy of our model over changes in parameters, we test different values of $M_{L,S}$, the rate of increase in average peak speed of service type S at location L. Other than average rate of inflation, which has been been shown not to fluctuate much over time, this parameter is only one that can change our results in a nonlinear manner. Because the United Kingdom showed both the most and least drastic effects due to changes in $M_{L,S}$, we chose this country for our analysis.

A 20 percent change in $M_{L,S}$ resulted in a percent change in the peak speed that was at most 17 percent (for mobile internet in the UK) and at least 6 percent (for broadband internet in the UK). The resulting percent change in the price was at most 15 percent (for mobile internet in the UK) and at least 6 percent (for broadband internet in the UK).

Accordingly, we believe our model is generally resilient.

1.5.3 Strengths and Weaknesses

We chose to separate the internet markets in both countries into the mobile and broadband sectors. We believe this is an advantage of our model, because it allows us to do separate *calculations* – appropriate, because of the fact that mobile and broadband are generally billed for and provided by separately, as well as the fact that mobile and broadband internet have completely different infrastructure – while permitting universal, combined *considerations.* That is to say: mobile and broadband are both getting speedier.

Issues arose with the billing of mobile internet. For simplicity, we assumed that each mobile phone plan is singular (that is to say, no phone plan would contain multiple lines). This is a weakness, because mobile internet plans with multiple lines do often introduce artificial discounts as an incentive. Additionally, we only took into account the costs of cellular data, not including the price of text messaging and phone calls, which are not internet. Although these are often billed together, we found it difficult to meaningfully separate them.

2 Part 2: Bit by Bit

Different households have vastly different internet needs.

2.1 Restatement of the Problem, Part 2

The problem we are tasked with addressing is as follows:

- Accounting for a global shifts in online education and work, predict a given household's need for internet over one year. Determine the minimum amount of required bandwidth to cover the household's total internet need s 90 percent and 99 percent of the time, based on age of inhabitants, household size, and their typical activities that use the internet. The households are:

    1. A couple in their early 30s; one is a teacher and one is looking for work; they have a 3-year-old child.
    2. A woman in her 70s; she is retired and caring for two school-aged children twice a week.
    3. Three former M3 challenge competitors sharing an apartment while enrolled in college full-time and working part-time.

2.2 Assumptions and Justifications

1. *All internet connections are wireless.* The majority of in-home internet devices, today, are connected by WiFi, not wired internet connections [15].

2. *Multiscreening is defined as the use of two active devices concurrently.* 77 percent of Americans report that they multiscreen, but the chance that any person is in fact multiscreening on a given interval is much smaller – closer to 13 percent [19].

3. *For the 99 percent-of-the-time scenario, everyone in the household is multiscreening and using the highest bandwidth possible.* The chance of this happening is extremely low – 13 percent cubed is about 1 percent [19].

4. *For the 90 percent-of-the-time scenario, everyone in the household is using one device for maximum bandwidth activities* The chance of this happening is 13 percent of the time – but it is yet more unlikely that all people in the household are using the highest bandwidth [19].

5. *No one person is streaming or gaming on more than one device at one time.* We assume this because of the attention that both of these activities require, and the average human will not have the capability to focus on both at once.

6. *No one person is actively using more than two devices at one time.* Again, the attention that more than two devices require to both be actively used at the same time is greater than most humans can handle.

7. *One device is used for one activity at one time.* Active internet use on a given device is generally through one specific activity, such as gaming, information searching, or video streaming.

8. *A device that uses bandwidth is labeled as either active or passive, not both.* An active device is defined as a device that only requires significant bandwidth and can perform its main task using when a person is using it (including browsing social media on a smartphone, streaming TV, playing a multiplayer video game, for examples). A passive device is a device that performs its main task in the background without significant bandwidth (including many Internet of Things devices, such smart appliances like thermostats, light switches, and security systems).

9. *Bandwidth is additive.* The total bandwidth being used by a household is divided among all of its devices, with devices performing more bandwidth-heavy tasks receiving more at any one time, and often slowing other devices on the same network [14]. Therefore, we choose to model total bandwidth as a summation of every device's current bandwidth usage.

10. *This model is best suited for post-pandemic life.* We believe that effective modeling takes a long-term outlook. Though the coronavirus's effects, specifically on internet usage for education and work, have been doubtlessly dramatic, we strongly believe that as the pandemic recedes (beginning in Q2 of 2021 and practically complete by Q1 of 2022 [9]), these shifts in internet usage will stabilize to levels similar to 2019.

2.3 The Model

When considering the adequate internet bandwidth for a given family, many factors must be accounted for, including the number of devices as well as the way these devices are used in the household.

2.3.1 Parameters

Symbol	Definition	Units	Value
N_i	Answer to survey question i, given in 2.3.2.1	...	...
S_{90}	# of people streaming when everyone is using one device	...	...
S_{99}	# of people streaming when everyone is using two devices	...	...
V_{90}	# of people on video calls when when everyone is using one device	...	...
V_{99}	# of people on video calls when everyone is using two devices	...	...
G_{90}	# of people playing video games when when everyone is using one device	...	...
G_{99}	# of people playing video games when everyone is using two devices	...	...
I_{90}	# of people searching the web when when everyone is using one device	...	...
I_{99}	# of people searching the web when everyone is using two devices	...	...

2.3.2 Developing the Model

In order to determine the amount of Mbps that each of the given households needs, we decided to characterize the types of devices used by each household as being active and passive devices. Active devices, such as smartphones, computers and desktops, televisions, and tablets are devices that are used for longer periods of time that require more internet speed. Passive devices, on the other hand, merely require a certain amount of internet speed in the background, such as smart home devices and security cameras.

We then developed a survey with a series of questions about internet usage. We “proposed” these questions to each of our three households. We developed answers to these questions using a variety of real-world aggregated data, scaled down to the individual level for each house.

2.3.2.1 Survey Questions (i.)

1. How many people regularly spend time within your household?
2. How many people regularly use video calls or video conferences?
3. How many people regularly use streaming services?
4. How many people regularly play video games?
5. How many people search the internet in a browser, specifically for access to information, such as through Google or Wikipedia searches?
6. How many laptops, computers, smartphones and tablets - combined - does your household own?
7. How many smart cameras does your house have?
8. How many other smart devices does your house have?

2.3.2.2 Survey Answers As per our assumption, these answers are presumed to have been given in Q4 of 2021, which is essentially a post-pandemic era [9], in which activities such as work and school have shifted largely back to in-person.

Question	Household 1	Household 2	Household 3	Justification
N_1	3	3	3	...
N_2	2	0	3	[20]
N_3	3	2	3	[8]
N_4	0	2	3	[8]
N_5	2	3	3	[8]
N_6	5	4	7	[18]
N_7	0	0	0	[16]
N_8	0	1	1	[17]

The values for N_6, N_7, and N_8 are found through expected values with the percentages of certain age groups [18] that have a smartphone, tablet and laptop, respectively. As an example, our calculation for N_6, Household 1 gives

$$0.96 * 2 + 0.57 * 2 + 0.81 * 2 = 4.68 \approx \textbf{5 devices},$$

where the 3-year-old is assumed to own no devices.

***

For the 99 percent-of-the-time values, we declare that everyone in the household is using two active devices concurrently, and that everyone is doing the activities on those devices that use maximum bandwidth. This is a procedure of assignments; we assign devices to activities and assign two people to each activities, so that the total bandwidth is maximized. Note that generally streaming uses more bandwidth than video calls uses more bandwidth than playing video games uses more bandwidth than searching the internet [8].

For S_{99}, we take the minimum of the number of devices and the number of people in the household who stream regularly.

$$S_{99} = min(N_6, N_3).$$

For V_{99}, we take the minimum number of devices left after we assign S_{99} and the number of people in the household who video call regularly.

$$V_{99} = min(N_6 - S_{99}, N_2).$$

For G_{99}, we take the minimum of the number of devices left after both previous assignments and the number of people who play video games regularly, having discounted the number of people who are already streaming. Note that if the number of people who are already streaming is greater than the number of people who game regularly, this value will be zero.

$$G_{99} = min(N_6 - S_{99} - V_{99}, max(N_4 - S_{99}, 0)).$$

For I_{99}, we take the minimum of the number of devices left after the previous three assignments and the number of people who regularly search the internet but have not yet been assigned to two devices. If everyone has already been assigned to two devices or there are no devices left, this value is zero.

$$I_{99} = min(N_6 - S_{99} - V_{99} - G_{99}, min(N_5, max(2 * N_1 - S_{99} - V_{99} - G_{99}, 0))).$$

***

For the 90 percent-of-the-time scenarios, we declare everyone in the household is using one device on a highest-bandwidth activity. This is chosen because the probability that a person will be multiscreening at any given time is around 16 percent [19][22]. The likelihood that all members of a household would each be on one device and in a configuration that utilizes the maximum amount of bandwidth is also very small.

For S_{90}, we take the minimum of the number of devices and the number of people in the household who stream regularly.

$$S_{90} = min(N_6, N_3).$$

For V_{90}, we take the minimum of the number of devices left after we assign S_{90} and the number of people in the household who video call regularly and are not yet assigned to a device.

$$V_{90} = min(N_6 - S_{90}, min(N_2, N_1 - S)).$$

For G_{90}, we take the minimum of the number of devices left after both previous assignments and the number of people who play video games regularly and are not yet assigned to a device.

$$G_{90} = min(N_6 - S_{90} - V_{90}, min(N_4, N_1 - S_{90} - V_{90})).$$

For I_{90}, we take the minimum of the number of devices left after the previous three assignments and the number of people who regularly search the internet and are not yet assigned to a device.

$$I_{90} = min(N_6 - S_{90} - V_{90} - G_{90}, min(N_5, N_1 - S_{90} - V_{90} - G_{90})).$$

***

To find the required minimum bandwidth at a given household, we use

$$\text{Active Bandwidth} = 2 * (8S + 4V + 3G + I),$$

$$\text{Passive Bandwidth} = 2 * (2N_7 + 1.2N_8)$$

$$\text{Total Bandwidth} = ActiveBandwidth + PassiveBandwidth,$$

$$\mathbf{Total\ Bandwidth} = \mathbf{2(8S + 4V + 3G + I + 2N_7 + 1.2N_8)},$$

where the "bandwidth coefficients" on S, V, and G are from [8] and the coefficients on N_7 and N_8 are from [16] and [17], respectively. We run function this using the 99 and 90 percent values. Then, the entire bandwidth equation is multiplied by 2, which is to account for dips in bandwidth due to distance from routers and modems in a household. Dips in internet quality range from 20 percent to 50 percent due to these issues [3].

2.4 Results

Running this bandwidth formula with the appropriate values, we find the adequate bandwidths:

Household 1, 99 percent-of-the-time: 64 Mbps
Household 1, 90 percent-of-the-time: 48 Mbps
Household 2, 99 percent-of-the-time: 46 Mbps
Household 2, 90 percent-of-the-time: 40 Mbps
Household 3, 99 percent-of-the-time: 74 Mbps
Household 3, 90 percent-of-the-time: 50 Mbps.

2.5 Evaluating the Model

2.5.1 Validation

These results make sense. In every case, the 99 percent value is higher than than 90 percent value; additionally, the households with greater usage have greater bandwidth values.

2.5.2 Strengths and Weaknesses

Our model considers the dynamics of internet usage in the household, taking into account devices that are actively being used as well as ones that simply run in the background. The model also takes into account how people can use internet at the same time. The survey associated with the model is relatively robust, considering various types of internet usage and who performs them on a regular basis. In order to determine the minimum data speed needed for 90 percent and 99 percent of the time, we utilized occurrences we believed to be rare and whose probabilities, according to our data on multi-screening, were around 10 percent and 1 percent, respectively.

Although our formula for bandwidth attempts to model scenarios when everyone is concurrently on one device or two devices, there are scenarios possible where this distribution is not even. Our model does not take into account economic factors, which may play a role in how much time is even available in a household for internet usage. The model does not focus much on the effects of a transition from a pandemic world to a post-pandemic world and how that affects screen usage.

3 Part 3: Mobilizing Mobile

The success of mobile internet – wireless internet connection broadcast to consumers in the field – relies foremost on the placements of cellular towers that transmit data.

3.1 Restatement of the Problem, Part 3

The problem we are tasked with addressing is as follows:

- Accounting for information on the population and demographic data of a region, create a model that best organizes cellular nodes within that region. Apply this model to the provided regions or regions of your choosing with identical data.

3.2 Assumptions and Justifications

1. *Regions with similar locations and population densities require similar numbers of cell towers per square mile.* We looked at 3 similar areas and the ratios were uniform [21] [23] [24].

2. *Population density can be modeled discretely.* Each area is disjoint, there is no overlap between areas.

3. *Regions require data in proportion to their populations.* Each individual in each region requires approximately the same amount of data, so each region will require data proportional to their inhabitants.

4. *All of the listed regions can be connected with only high-bandwidth towers.* Each region is small, and well within the range of a high-bandwidth tower, and the towers supply more than enough data to support the region's inhabitants.

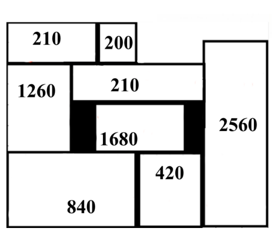
3.3 The Model

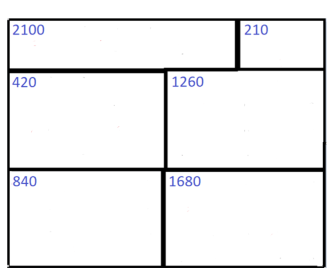
3.3.1 Parameters

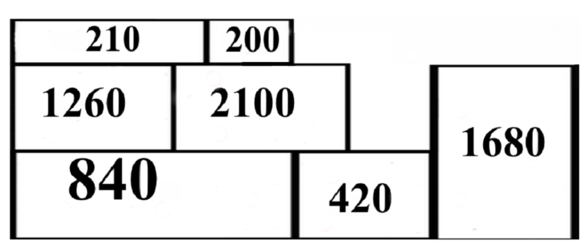
Symbol	**Definition**	**Units**	**Value**
R	Average ratio of cell towers to area for small US towns	towers/mi^2	11.4
T_R	Towers in region R	towers	...
A_R	Area of region R	mi^2	...
P_r	Population of sub-region r	people	...
A_r	Area of sub-region r	mi^2	...
W_r	Weight of nodes in sub-region r	...	...

3.3.2 Developing the Model

In the interest of time, we decided against using the three given regions and chose instead to model regions with similar characteristics but simpler shapes.







Each region has a population and area similar to that of the surveyed towns (population $\sim$ 6000, area $\sim 1\ mi^2$) and contains a variety of sub-regions, each with their own areas and populations.

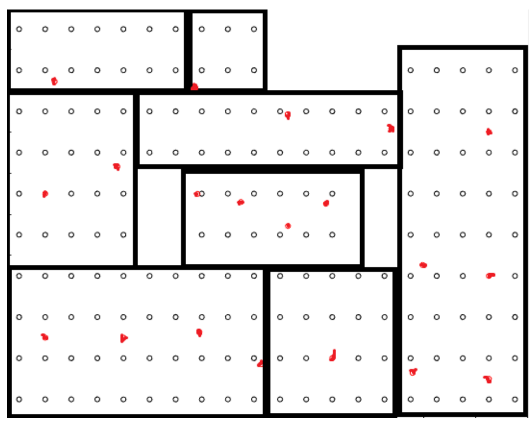
The area of each region was calculated and multiplied by R to determine the number of towers the region enclosed; that is,

$$T_R = R * A_R.$$

A Python script superimposed a grid over each region, creating weighted nodes wherever the grid's lattice points intersect with sections of the region. Each lattice point is weighted according to the population density of the sub-region enclosing it; that is,

$$W_r = \frac{P_r}{A_r}.$$

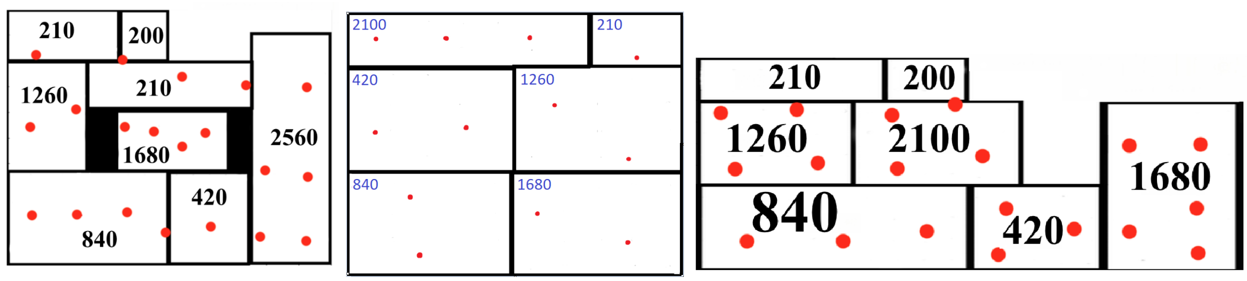
This created a graph of nodes:



We then used a weighted k-means algorithm to determine the optimal positioning of T_r cellular towers (shown in red) along a given route. The k-means algorithm is a two-step algorithm. When the algorithm begins, it assigns k random points along the route to be so-called means and assigns each of the i data points to the closest mean. It then finds the new position of each mean by creating a weighted average of its constituent points, utilizing their distance from starting point as the position and their weight as the bias.

3.4 Results

Our program generated the following tower distributions:



These distributions show that the optimal tower placement requires widespread distribution of towers concentrated in high-population areas.

3.5 Evaluating the Model

3.5.1 Strengths and Weaknesses

This model accounts for differences in population distribution and region geometry - giving it real world applicability - and its design causes it to generate perfect or near-perfect tower arrangements. This model's parameters are derived from data from throughout the it is valid throughout the country.

This model's main weakness is its inability to manage curved or diagonal regions, as these regions occur far more often then the rectangular regions considered by this model. Given more time, we would have designed our model to work for these regions, but time constraints forced us to use regions with simpler bounds.

4 Global Conclusions

In the past two decades, even before the coronavirus pandemic, internet access has transformed from a luxurious novelty into a basic necessity akin to food, water, and shelter. Connecting with colleagues and peers, accessing massive online marketplaces such as Amazon, and simply being able to attain the internet's vast wealth of free information is vitally important in today's economy and society. Our first model dove into the future of internet speeds and their respective costs. We predicted that while the cost of service plans would somewhat increase (partially due to inflation), the peak speeds of the same plans would increase at a far greater rate in the short term, leading to the price per unit of peak speed decreasing in both the US and UK for terms of both mobile and broadband connections. Our second model sought to assign an adequate internet bandwidth to any given household based on its internet usage. Because these usage patterns can differ so widely over a plethora of variables, we devised a survey about a household's consumption and used that survey to find the optimal bandwidth for a given home. In our final model, we employed superimposition and modified k-means algorithms to generate the optimal distribution of towers for our custom regions. Taken together and viewed in the long term, our models give an optimistic view: the world will not cease to expand and evolve to provide affordable, quality internet service to all of its people.

5 Further Studies

We would expand the first model to include upload speeds as well as download speeds, which we opted to not account for due essentially due to time constraints, but also due to the apparent lack of correlation between billed-for download speeds and experiential upload speeds. In addition, we did not incorporate bundling, a package that many companies offer in order that further lowers the price for consumers. Including this would likely lead to an even better price per megabit in the future.

We might extend our second model by devising an even more robust survey. For example, we would ask for economic considerations – we also would like to take into account age to household members – some sense of a probability distribution in assigning *which* members of the household would be using a certain device at a given time.

Our third model could be extended to include slanted and curved regions, as these regions are more prevalent in everyday life. We'd also like to incorporate different types of towers into our calculation, as there might be a better solution incorporating multiple tower types.

6 References

[1] https://www.macrotrends.net/2549/pound-dollar-exchange-rate-historical-chart

[2] https://www.allconnect.com/blog/advertised-vs-actual-internet-speeds

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

[4] https://ourworldindata.org/internet

[5] https://www.cnn.com/2020/02/10/perspectives/6g-future-communications/index.html

[6] https://www.cable.co.uk/broadband/pricing/worldwide-comparison/

[7] https://www.moneysavingpro.com/cell-phone-plans/comparing-us-and-uk-bills/

[8] https://m3challenge.siam.org/node/523

[9] https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/when-will-the-covid-19-pandemic-end

[10] https://www.allconnect.com/blog/the-average-cost-of-a-cellphone-plan-in-the-us-is-113

[11] https://www.in2013dollars.com/Internet-services-and-electronic-information-providers/price-inflation

[12] https://www.alentus.com/articles/top-5-most-popular-isps.asp

[13] https://www.bea.gov/system/files/papers/WP2004-10.pdf

[14] https://www.tanaza.com/tanazaclassic/blog/how-to-calculate-network-bandwidth-requirements/

[15] https://www.howtogeek.com/217463/wi-fi-vs.-ethernet-how-much-better-is-a-wired-connection/

[16] https://www.safety.com/home-security-survey/

[17] http://www.parksassociates.com/blog/article/ces-summit2016-pr3

[18] https://www.pewresearch.org/internet/2015/10/29/the-demographics-of-device-ownership/

[19] https://www.statista.com/statistics/319732/daily-time-spent-online-device/

[20] https://www.ipsos.com/en-us/news-polls/coronavirus-prompts-increased-use-of-video-chat

[21] http://www.city-data.com/towers/cell-Norristown-Pennsylvania.html

[22] https://www.tandfonline.com/doi/full/10.1080/02650487.2017.1348042

[23] http://www.city-data.com/towers/cell-Bixby-Oklahoma.html

[24] http://www.city-data.com/towers/cell-Bonney-Lake-Washington.html

A Code Appendix

Below are images of MATLAB code detailing the k-means algorithm and Python code detailing the node creation process. These algorithms are used solely in model three.

```
function centroids = kMeans(Data, weights, numberOfMeans, numberOfIterations)
%%This program implements a weighted kMeans algorithm to determine the %%
%%optimal position of various charging stations (called centroids).    %%


%%This process initializes the centroids and assigns their values to   %%
%%random data points.                                                  %%
centroids = zeros(numberOfMeans, size(Data, 2));
switchedCentroids = randperm(size(Data, 1));
for i = 1:numberOfMeans
    centroids(i, :) = Data(switchedCentroids(i), :);
end

%%This iterative process first assigns the data points to the closest  %%
%%centroid and updates the position of the centroids to the weighted   %%
%%average of their constituent points.                                 %%
for i = 1:numberOfIterations
    idx = findClosestCentroids(Data, centroids);
    centroids = computeCentroids(Data, weights, idx, numberOfMeans);
end

%%This section colors he data points and centroids and displays all of %%
%%the values.                                                          %%
givenColors = zeros((size(Data, 1) + numberOfMeans), 3);
for i = 1:numberOfMeans
    givenColors(size(Data, 1) + i, :) = [1 0 0];
end
scatter([transpose(Data(:, 1)) transpose(centroids(:, 1))], ...
    [transpose(Data(:, 2)) transpose(centroids(:, 2))], 10, givenColors)

%%This function sorts the centroids and returns their values.          %%
centroids = sort(centroids);
return
end
```

```matlab
function idx = findClosestCentroids(Data, centroids)
%%This assigns each point to its closest centroid.                    %%

%%This initializes various useful values.                             %%
numberOfMeans = size(centroids, 1);
idx = zeros(size(Data,1), 1);

%%This process assigns each point to the nearest centroid.            %%
for i = 1:length(Data)
    distance = ones(numberOfMeans, 1);
    for j = 1:numberOfMeans
        distance(j, 1) = norm(centroids(j, :) - Data(i, :));
    end
    [~, id] = min(distance);
    idx(i) = id;
end

return

end
```

```matlab
function centroids = computeCentroids(Data, weights, idx, numberOfMeans)
%%This program uses weighted averages to update the values of each      %%
%%centroid.                                                             %%

%%This initializes various useful values.                               %%
[numberOfDataPoints, numberOfParameters] = size(Data);
centroids = zeros(numberOfMeans, numberOfParameters);
count = zeros(numberOfMeans, 1);

%%This updates the values of each centroid by adding the weighted       %%
%%values of their constituent points.                                   %%
for i = 1:numberOfDataPoints
    for j = 1:numberOfMeans
        if(idx(i, 1) == j)
            for k = 1:numberOfParameters
                centroids(j, k) = centroids(j, k) + weights(i, 1) * Data(i, k);
            end
            count(j, 1) = count(j, 1) + weights(i, 1);
        end
    end
end

%%This descales the value of the centroids.                             %%
for i = 1:numberOfMeans
    for j = 1:numberOfParameters
        centroids(i, j) = centroids(i, j) / count(i, 1);
    end
end

end
```

```python
#This program takes a set of rectangular regions and superimposes them onto a grid.
#Lattice points are used to create a set of nodes, one for each element in the both the grid and the region
#This is formatted for insertion into MATLAB

def make_square_region(margin_list, weighting_list):

    #Creates two lists, one for valid point positions and one for the weight of each valid point
    positions = "["
    weights = "["


    for x in range(0, 20):
        for y in range(0, 10):

            #Determines the position of each lattice point
            x_prime = 0.1 * x + 0.05
            y_prime = 0.1 * y + 0.05


            for z in range(len(margin_list)):

                #Determines if the point falls into any of the given region
                if x_prime >= margin_list[z][0] and y_prime >= margin_list[z][1] and x_prime <= margin_list[z][2] and y_prime <= margin_list[z][3]:

                    #Adds the point to the valid points list
                    positions = positions + str(round(x_prime, 2)) + ", " + str(round(y_prime, 2)) + "; "

                    #Weights the point by its region's population density and adds the weight to the weights list
                    area = (margin_list[z][2] - margin_list[z][0]) * (margin_list[z][3] - margin_list[z][1])
                    weights = weights + str(round(weighting_list[z] / area, 2)) + "; "

    #Finishes formatting lists
    positions = positions + "]"
    weights = weights + "]"
    return [positions, weights]
```