Big Data in Telecommunications

ISBA 3413

Phase III

Logan Gomer, Tyler Dye,

Caleb McDaniel & Romello Turner

Due: April 12, 2024

Submitted: April 12, 2024

Submitted to: Dr. Richard Segall

Introduction

Big data is an ever present force in the modern business world. One would be very challenged to find an industry or organization that doesn’t utilize some form of big data analytics in order to support their decisions, products, marketing, and structure. The capture, analysis, and application of this data is critical to their success. Understanding how data can reflect changes and trends across the business model is of the utmost importance for the modern company in order to both remain afloat and gain competitive advantage. One industry that is particularly data driven is the field of telecommunications. In 2024, almost every single form of telecommunication (phone calls, text messages, networks, etc.) generates an immense amount of data that companies must be able to analyze and use to gain insight. Different areas of telecommunications will be evaluated throughout this report to help readers understand some of the ways that big data can be used to support companies in the telecommunications industry.

Types of Telecommunications - Logan Gomer

There are several types of telecommunications. There is data, satellite, wireless, optical, etc (*Electronics For You*). We can separate these many types of telecommunications into three simpler categories. These three categories are short message service (SMS), call, and the internet. I'm going to see which of these are used more frequently around the world in the modern times than the others along with if there are other factors that influence these trends. This would be used by telecommunications companies to see which of these types need to be adjusted, updated, or serviced in order to keep certain customers in a certain area happy. This can also be used to see specific trends in which types are being used as we move forward into the future as telecommunications progresses further. Altogether, these three functions are the basis of telecommunications. Also going further into each of the types of telecommunications to see how to maximize each type.

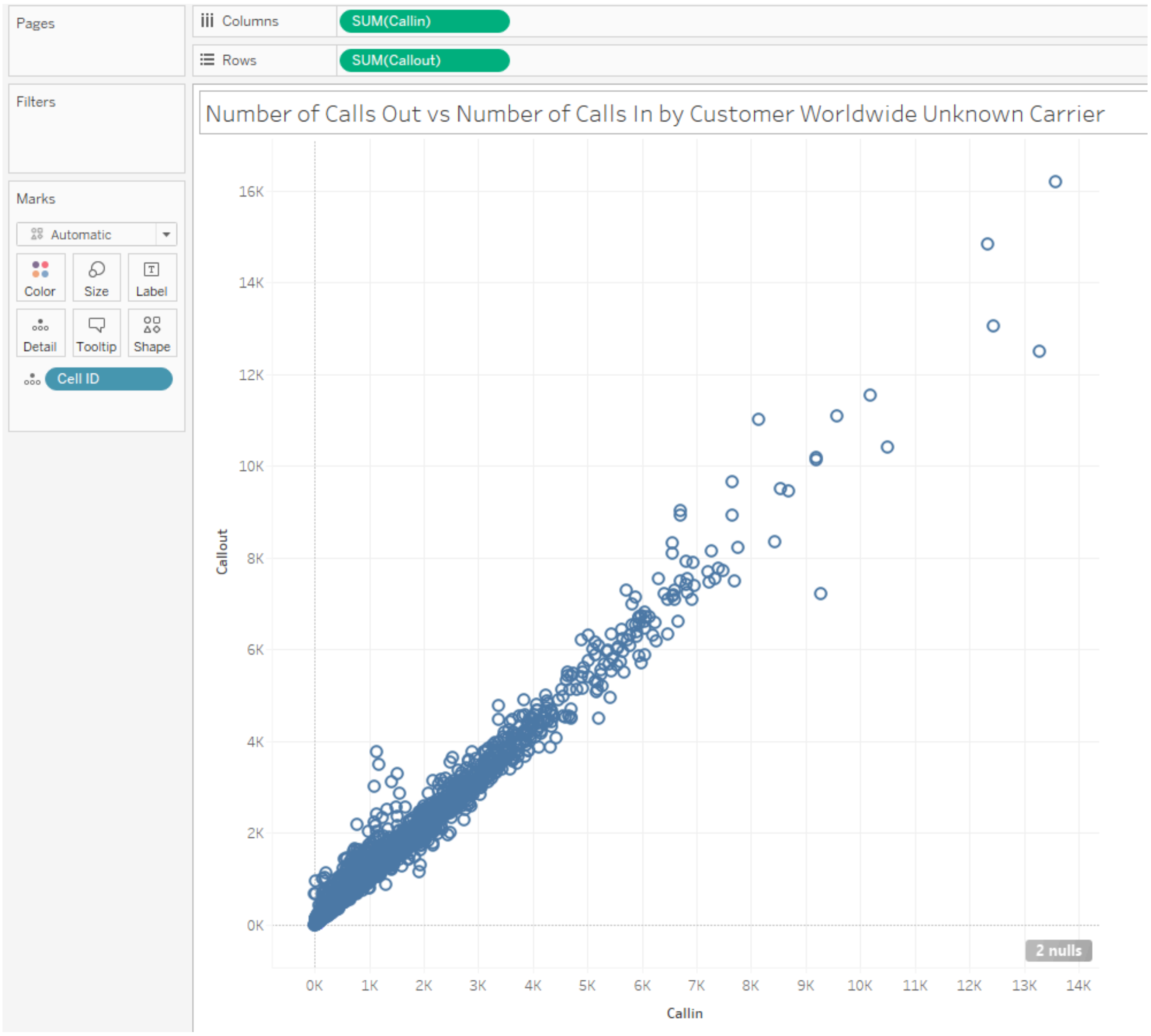


Figure 1.1: Number of Calls Out vs Number of Calls In by Customer July 11, 2013

This graph shows the relationship between how many times a customer calls someone and how many times a person gets calls on July 11, 2013. This graph shows a near one to one linear relationship between the two variables which means calling is a very common thing to do in the area the data was collected. For example if calling was less common you would see a much flatter linear relationship as people would be receiving more calls than they would be making.

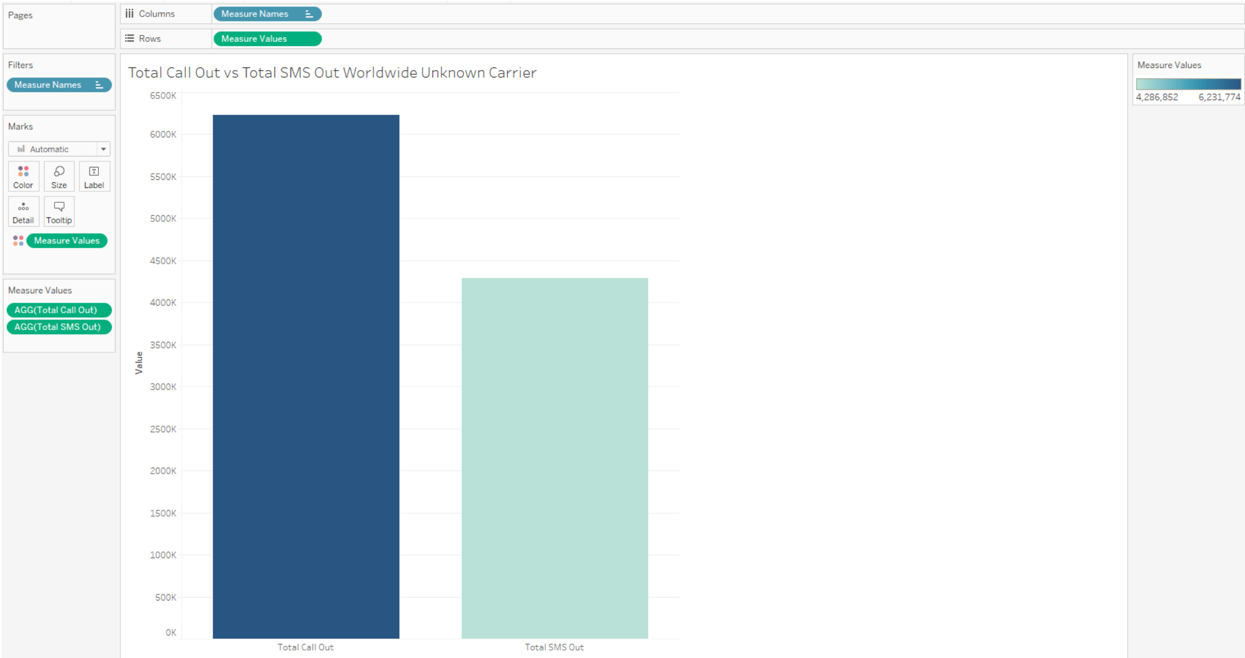
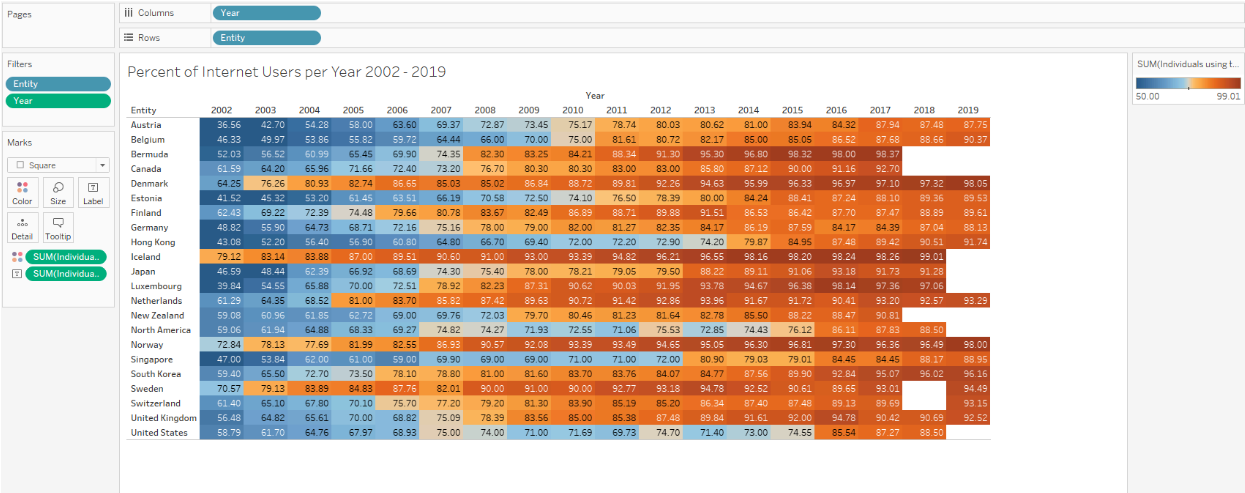


Figure 1.2: Total Calls Out vs Total SMS Out July 11, 2013

This bar chart shows the relationship between the amount of calls made and the amount of SMS messages sent on a particular service on July 11, 2013. As you can see the amount of calls made by customers is much higher than the amount of SMS messages sent by customers. This could show this company that they need to focus on retaining the customers that are more likely to call than to text or that they should upgrade their calling services.



Entity	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Austria	36.56	42.70	54.28	58.00	63.60	69.37	72.87	73.45	75.17	78.74	80.03	80.62	81.00	83.94	84.32	87.94	87.48	87.75
Belgium	46.33	49.97	53.86	55.82	59.72	64.44	66.00	70.00	75.00	81.61	80.72	82.17	85.00	85.05	86.52	87.68	88.66	90.37
Bermuda	52.03	56.52	60.99	65.45	69.90	74.35	82.30	83.25	84.21	88.34	91.30	95.30	96.80	98.32	98.00	98.37		
Canada	61.59	64.20	65.96	71.66	72.40	73.20	76.70	80.30	80.30	83.00	83.00	85.80	87.12	90.00	91.16	92.70		
Denmark	64.25	76.26	80.93	82.74	86.65	85.03	85.02	86.84	88.72	89.81	92.26	94.63	95.99	96.33	96.97	97.10	97.32	98.05
Estonia	41.52	45.32	53.20	61.45	63.51	66.19	70.58	72.50	74.10	76.50	78.39	80.00	84.24	88.41	87.24	88.10	89.36	89.53
Finland	62.43	69.22	72.39	74.48	79.66	80.78	83.67	82.49	86.89	88.71	89.88	91.51	86.53	86.42	87.70	87.47	88.89	89.61
Germany	48.82	55.90	64.73	68.71	72.16	75.16	78.00	79.00	82.00	81.27	82.35	84.17	86.19	87.59	84.17	84.39	87.04	88.13
Hong Kong	43.08	52.20	56.40	56.90	60.80	64.80	66.70	69.40	72.00	72.20	72.90	74.20	79.87	84.95	87.48	89.42	90.51	91.74
Iceland	79.12	83.14	83.88	87.00	89.51	90.60	91.00	93.00	93.39	94.82	96.21	96.55	98.16	98.20	98.24	98.26	99.01	
Japan	46.59	48.44	62.39	66.92	68.69	74.30	75.40	78.00	78.21	79.05	79.50	88.22	89.11	91.06	93.18	91.73	91.28	
Luxembourg	39.84	54.55	65.88	70.00	72.51	78.92	82.23	87.31	90.62	90.03	91.95	93.78	94.67	96.38	98.14	97.36	97.06	
Netherlands	61.29	64.35	68.52	81.00	83.70	85.82	87.42	89.63	90.72	91.42	92.86	93.96	91.67	91.72	90.41	93.20	92.57	93.29
New Zealand	59.08	60.96	61.85	62.72	69.00	69.76	72.03	79.70	80.46	81.23	81.64	82.78	85.50	88.22	88.47	90.81		
North America	59.06	61.94	64.88	68.33	69.27	74.82	74.27	71.93	72.55	71.06	75.53	72.85	74.43	76.12	86.11	87.83	88.50	
Norway	72.84	78.13	77.69	81.99	82.55	86.93	90.57	92.08	93.39	93.49	94.65	95.05	96.30	96.81	97.30	96.36	96.49	98.00
Singapore	47.00	53.84	62.00	61.00	59.00	69.90	69.00	69.00	71.00	71.00	72.00	80.90	79.03	79.01	84.45	84.45	88.17	88.95
South Korea	59.40	65.50	72.70	73.50	78.10	78.80	81.00	81.60	83.70	83.76	84.07	84.77	87.56	89.90	92.84	95.07	96.02	96.16
Sweden	70.57	79.13	83.89	84.83	87.76	82.01	90.00	91.00	90.00	92.77	93.18	94.78	92.52	90.61	89.65	93.01		94.49
Switzerland	61.40	65.10	67.80	70.10	75.70	77.20	79.20	81.30	83.90	85.19	85.20	86.34	87.40	87.48	89.13	89.69		93.15
United Kingdom	56.48	64.82	65.61	70.00	68.82	75.09	78.39	83.56	85.00	85.38	87.48	89.84	91.61	92.00	94.78	90.42	90.69	92.52
United States	58.79	61.70	64.76	67.97	68.93	75.00	74.00	71.00	71.69	69.73	74.70	71.40	73.00	74.55	85.54	87.27	88.50	

Figure 1.3: Percent of Internet Users by Country per Year 2002 - 2019

This heat map shows the percentage of internet users among the years 2002 - 2019 in countries with the highest percentage of users. This graph shows the trend of nearly every single person being on the internet as we move forward into the present. This would show a telecommunications company to start upgrading their internet services to prepare for this wave and be able to bring in eager customers. It also shows that different countries progressed faster so they should move faster in those countries if they have a division there.

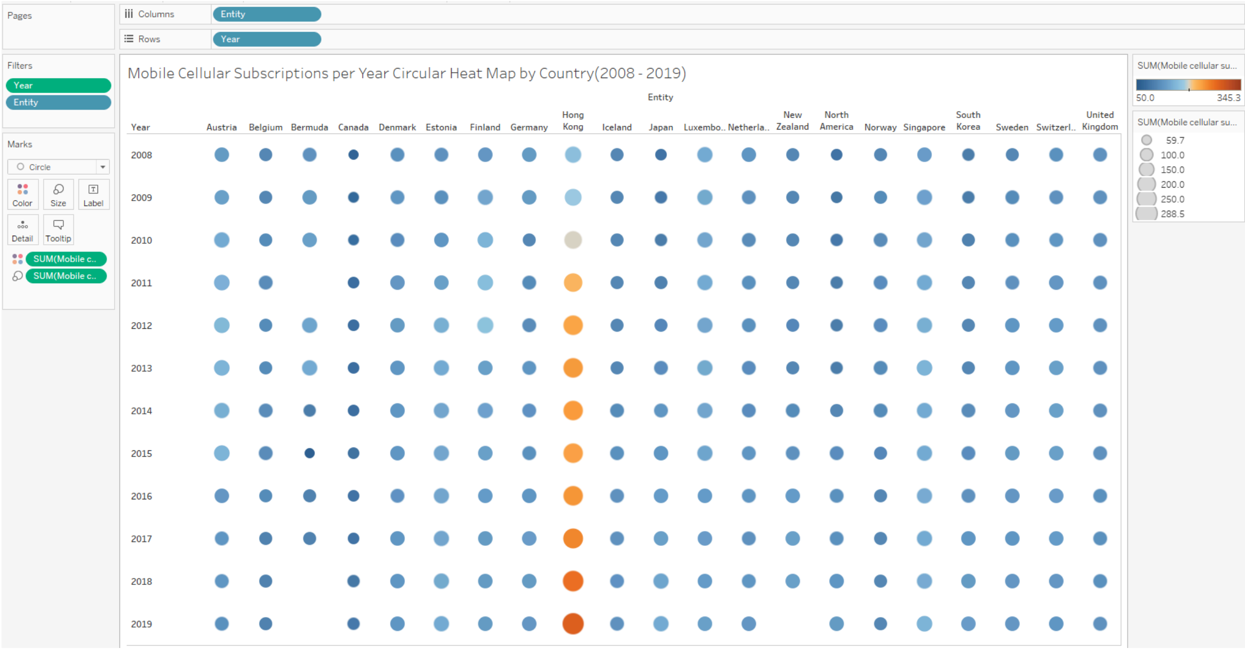


Figure 1.4: Mobile Cellular Subscriptions per Year 2008 - 2019

This circular heat map shows the number of mobile cellular subscribers per year in the same countries as Figure 3. It shows that in most countries this number doesn't change much with the slightest increase being shown in most of the countries and a large increase in Hong Kong specifically. This should show that most people in most countries nowadays already have a phone plan and there isn't an influx of new customers coming in. This could show a company that they need to focus more on making their service different in order to pull customers away from other companies and get a subscription with theirs.

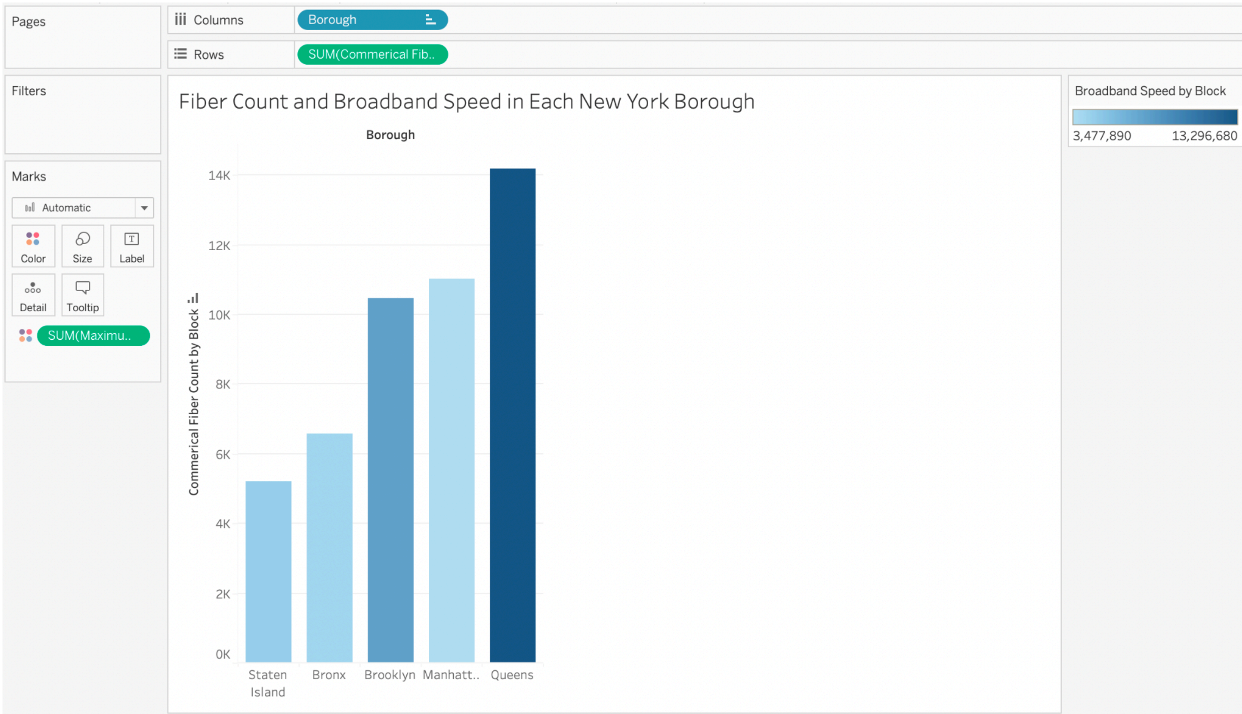


Figure 1.5: Fiber Count and Broadband Speed in Each New York Borough

This graph comes from a dataset that is 14 columns and 37,596 rows. It shows the fiber count and broadband speed in each of the boroughs in New York. The height of the columns shows the fiber count of the fiber optic cables in that borough. The color of each bar shows the broadband speed. The data in this set is from several providers so it wouldn't show anything to any specific company, but what it would show is where companies could expand their network since there is less competition. It could also show where they should possibly upgrade their services to compete more in different boroughs.

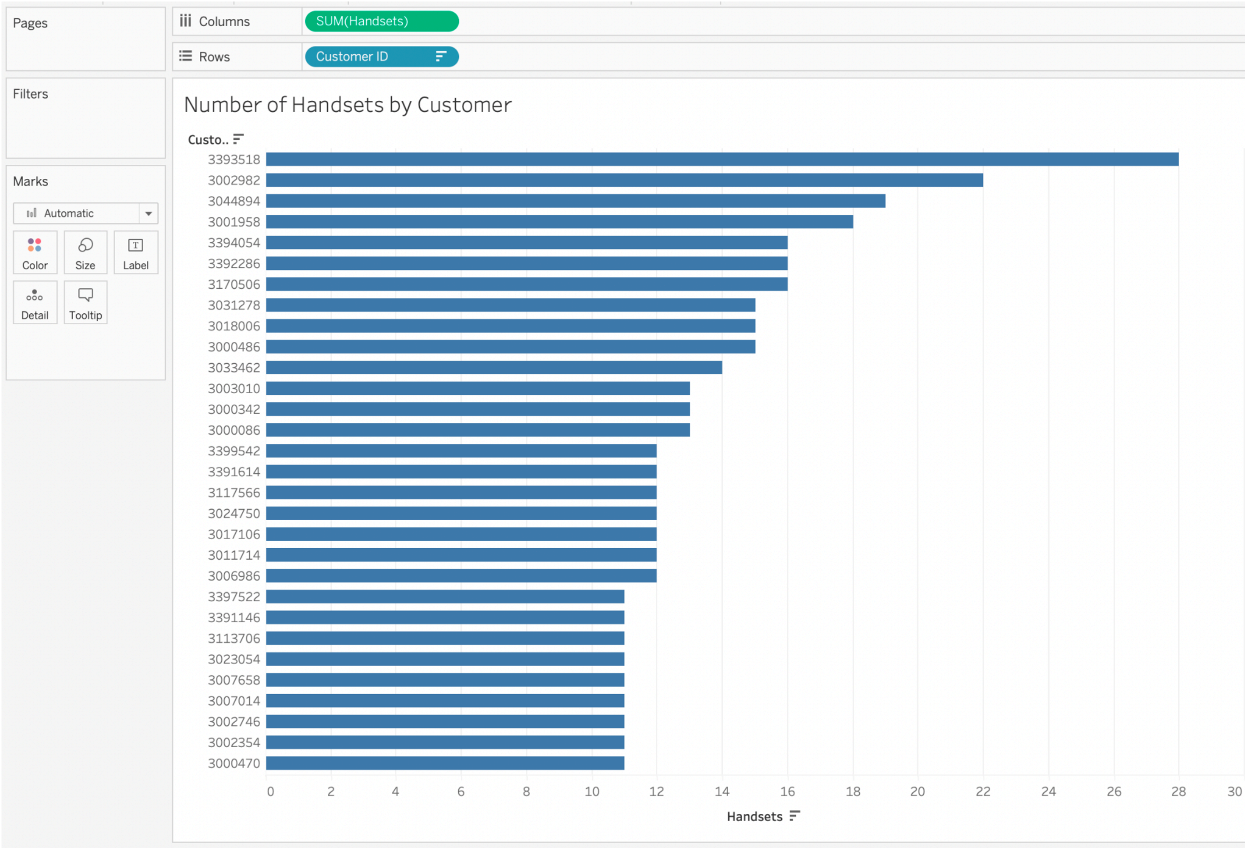


Figure 1.6: Number of Handsets per Customer, Unknown Provider

This chart comes from a dataset with 58 columns and 20,001 rows. It shows the number of handsets that each customer has on their plan. This shows which customers are more valuable because they have more resources tied to them. With this knowledge a company can send out customer specific deals to customers that are more valuable due to their number of handsets.

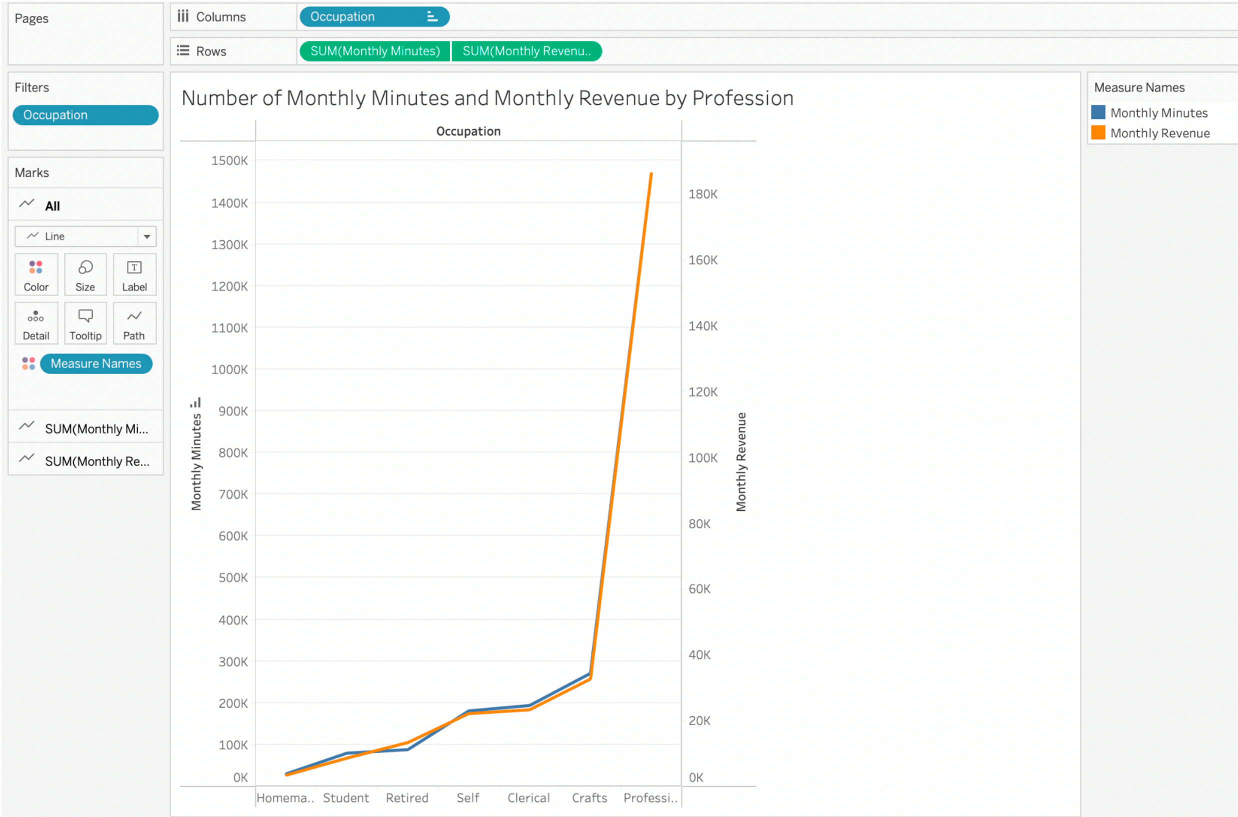


Figure 1.7: Number of Minutes and Revenue per Month by Profession, Unknown Provider

This chart comes from a data set with 58 columns and 20,001 rows. It shows the different professions of a service provider and the cumulative number of minutes they use and their revenue. A service provider can learn multiple things from this. The most important thing to learn from this is that the more someone makes the more minutes they will use per month. Since they are using more minutes then they will have a higher dollar plan so this shows which types of jobs to target when looking for customers. Or if you have a phone plan that's made to be more affordable than you know to target jobs on the left side of this graph.

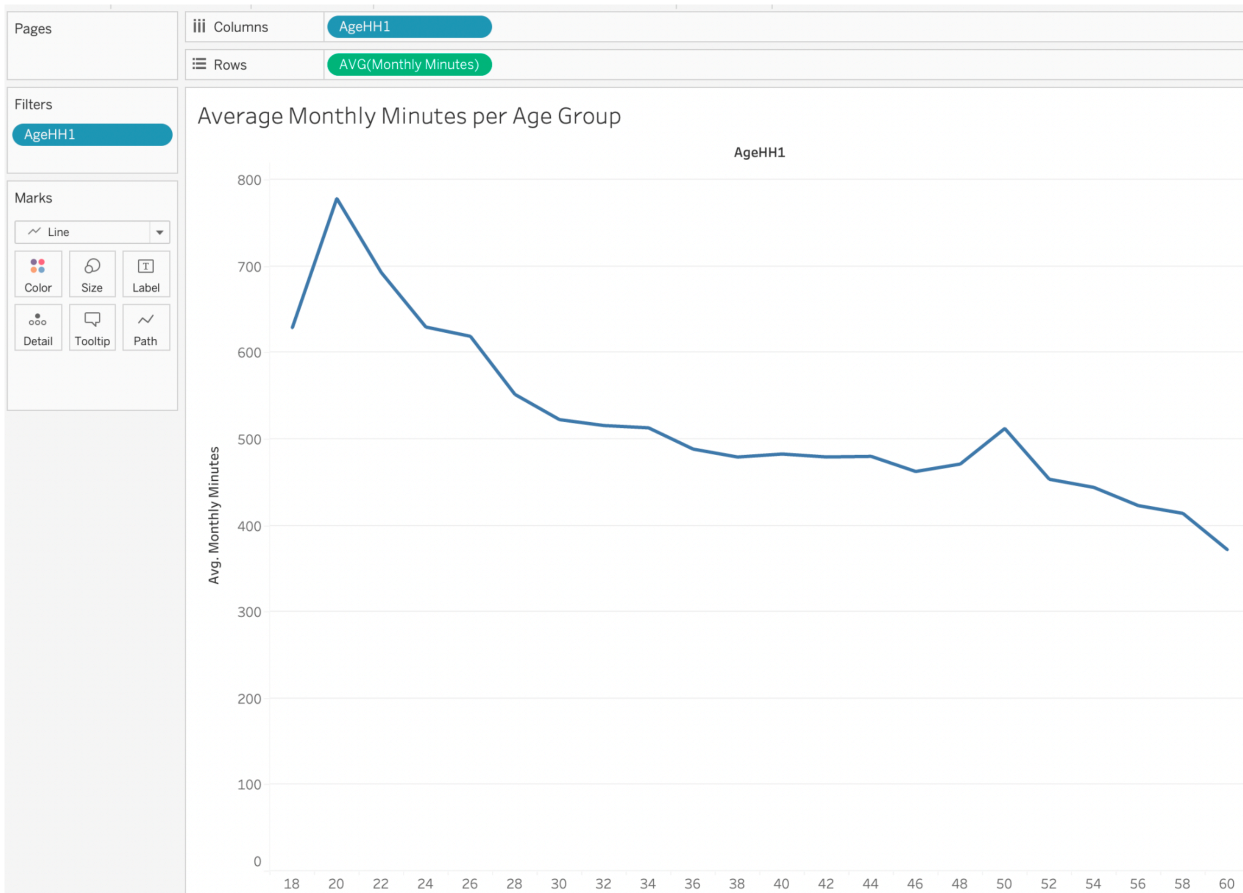


Figure 1.8: Average Monthly Minutes per Age Group, Unknown Provider

This chart comes from a data set with 58 columns and 20,001 rows. This graph shows the average monthly minutes used graph against the age group using the minutes. It shows how as people get older they use less and less minutes. For service providers this shows they should use less resources for people that are older and put more resources targeting the younger population.

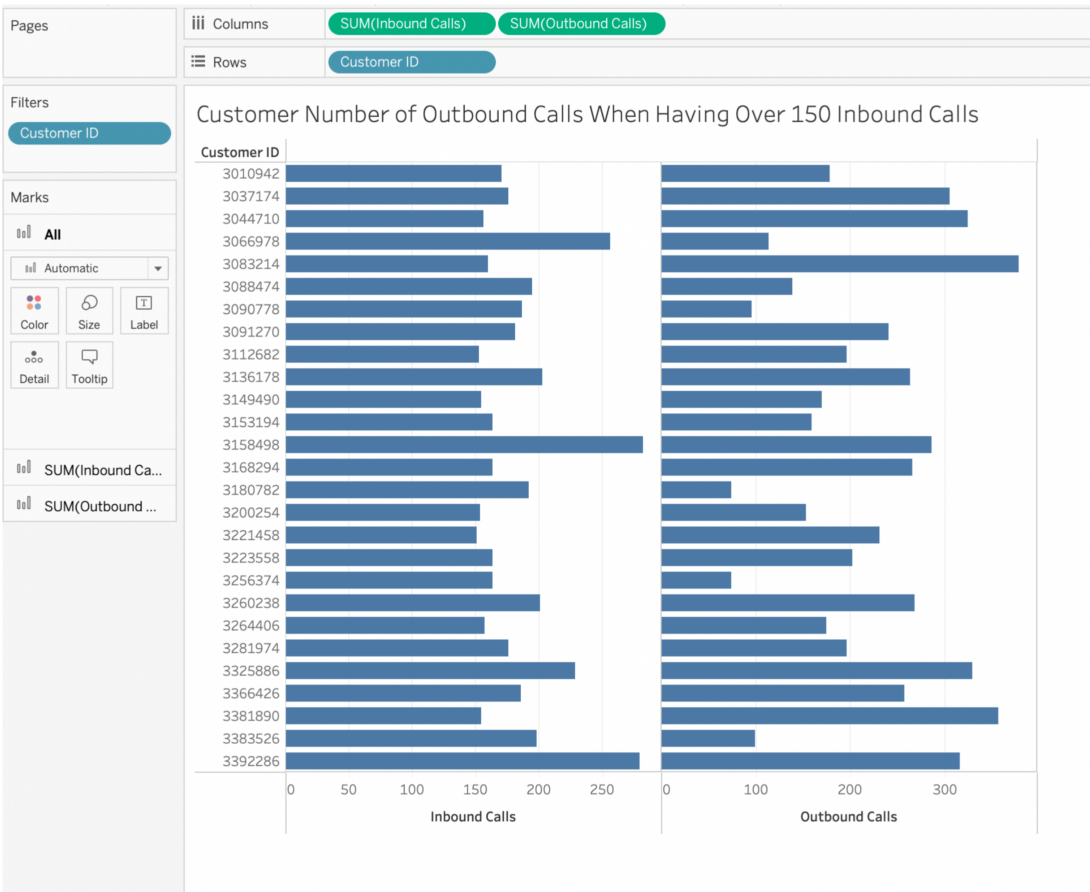


Figure 1.9: Number of Outbound Calls When Having Over 150 Inbound Calls per Customer, Unknown Provider

This chart comes from a data set with 58 columns and 20,001 rows. This shows the spread of people who are receiving the most calls and if they themselves are making calls. Some of these customers are making very few calls compared to the ones they received showing that these people don't enjoy making calls while some are making far more calls than they receive. A provider could use this information to cater to each client more by altering their deal to provide more of what they want to try and keep these customers with their service.

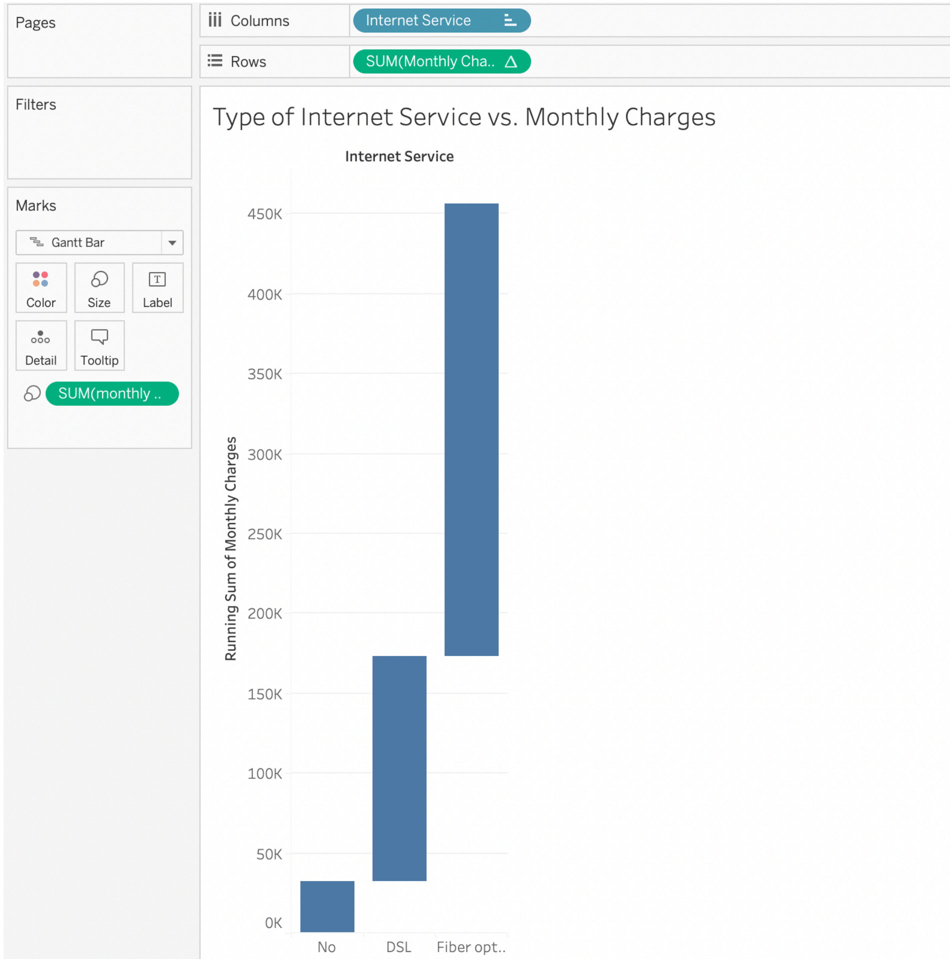


Figure 1.10: Types of Internet Service vs. Monthly Charges, Unknown Provider

This chart comes from a data set with 21 columns and 7,044 rows. This waterfall chart shows the extreme spread of monthly charges as the internet services change. This shows just how much more the company charges for fiber optic versus DSL and then no internet service at all. This can show a provider to try and advertise their fiber optic service more so people will switch to it. It can also show them that they might need to make it more affordable.

Figure 1.11: Number of Internet Service Providers vs. Commercial Fiber Count per Borough

This graph comes from a dataset that is 14 columns and 37,596 rows. This graph shows the number of internet service providers and commercial fiber count in each of the New York boroughs. This shows companies which areas have the largest coverage of providers. It also shows the fiber count of each borough. One thing to gather from this graph is that Manhattan has a very low representation of service providers while having a high fiber count so that means it is an area that a new provider could come into and be able to steal some customers.

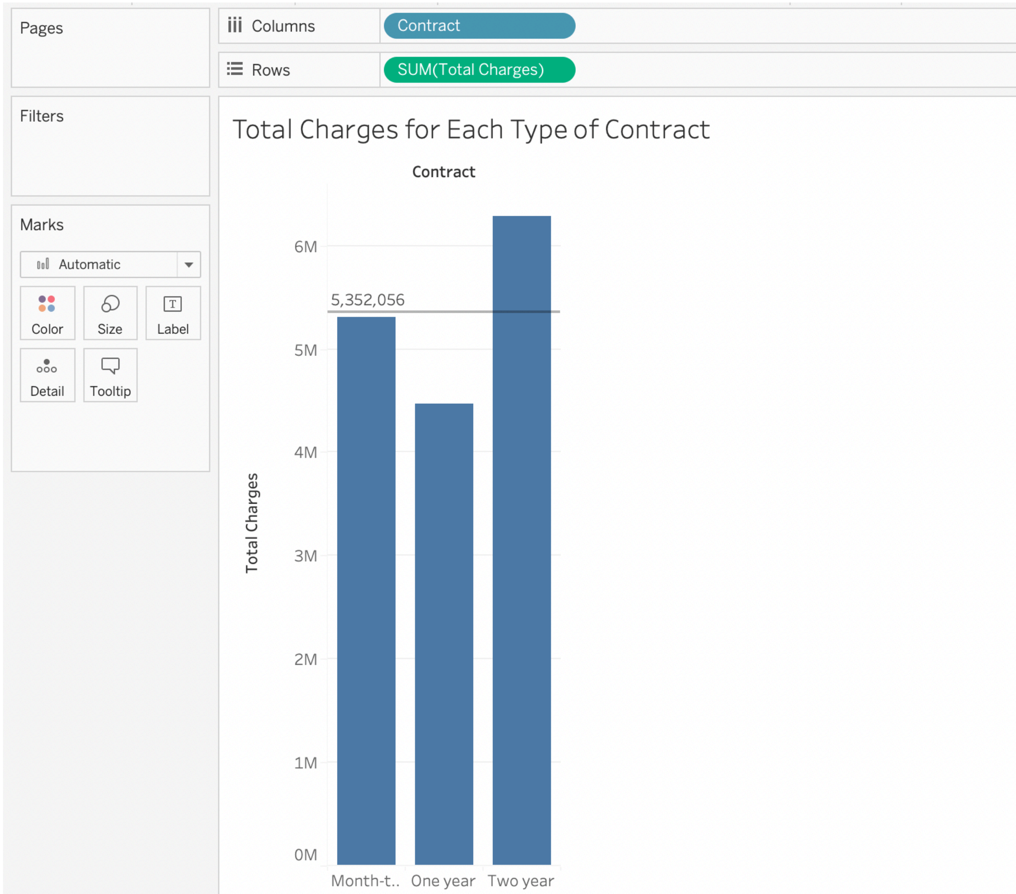


Figure 1.12: Total Monthly Charges for Each Type of Contract, Unknown Provider

This chart comes from a data set with 21 columns and 7,044 rows. This chart details the total monthly charges for each type of contract provided by the provider with an average line showing the average monthly charges. This can show a provider which contract provides the majority of their money. This shows that the one year contract is falling behind the other two and is well below the average so the provider should focus more heavily on the other two contracts or find a way to make the year long contract more enticing to customers.

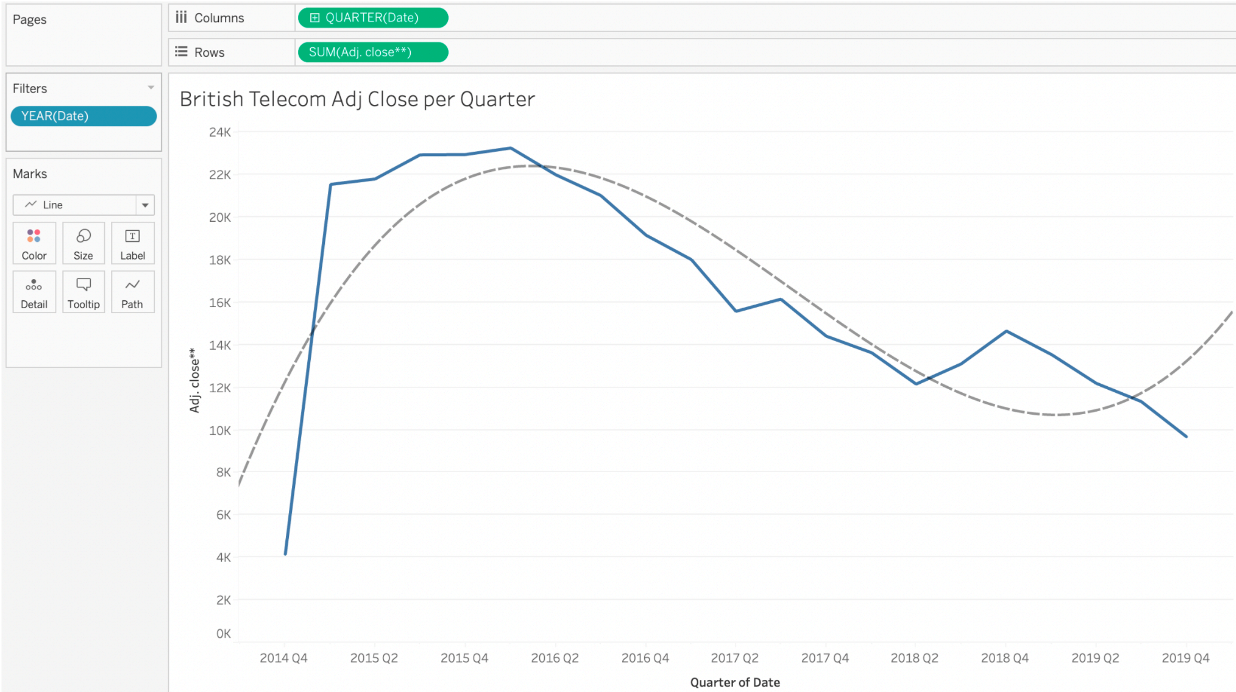


Figure 1.13: British Telecom Adjusted Close per Quarter

This chart comes from a data set with 4 years worth of British Telecom data. This shows the adjusted close from quarter four in 2014 up until quarter 4 in 2019. This shows the spike it had at the beginning of the year in 2015 and the slow fall since. There is a polynomial trend line to show the polynomial trend that the graph has taken. If the stock truly follows this trend then it should trend upwards again.

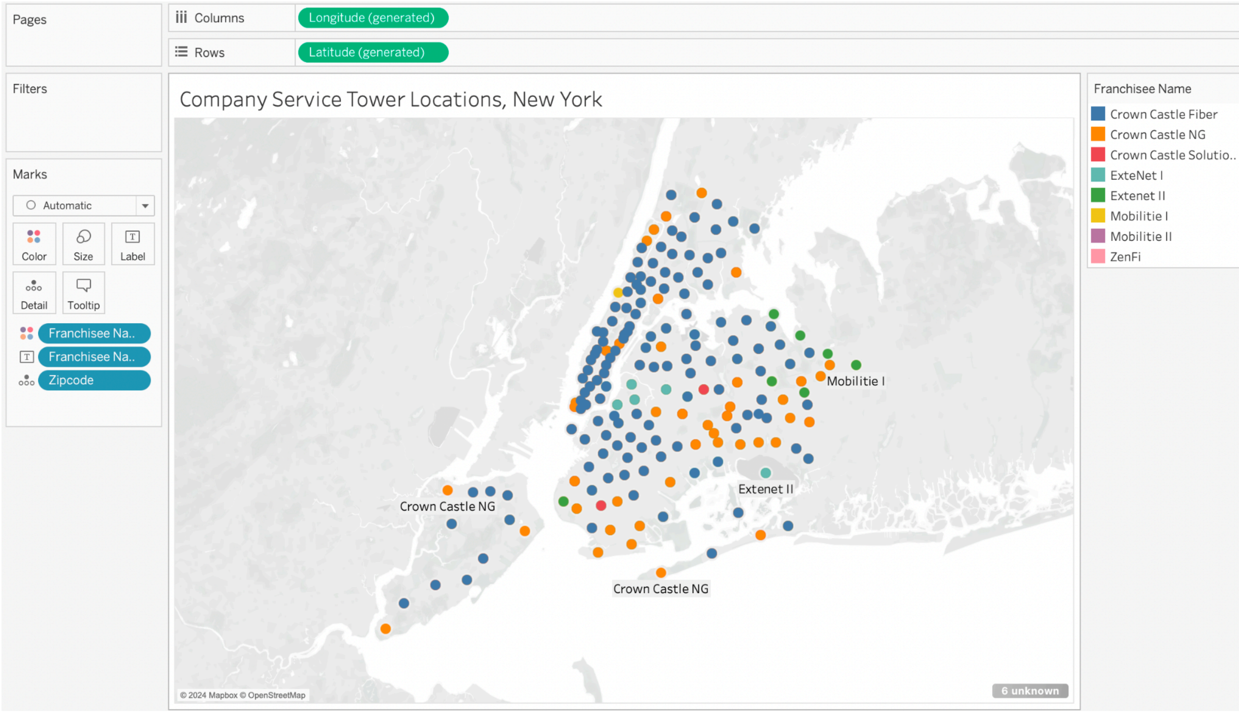


Figure 1.14: Company Service Tower Locations, New York

This chart comes from a data set with 22 columns and 10,970 rows. This map shows the location of several cell towers in the New York areas. This can show providers the areas they are lacking in or the areas where the competition is weak. This map can show companies where they should expand or where they should avoid.

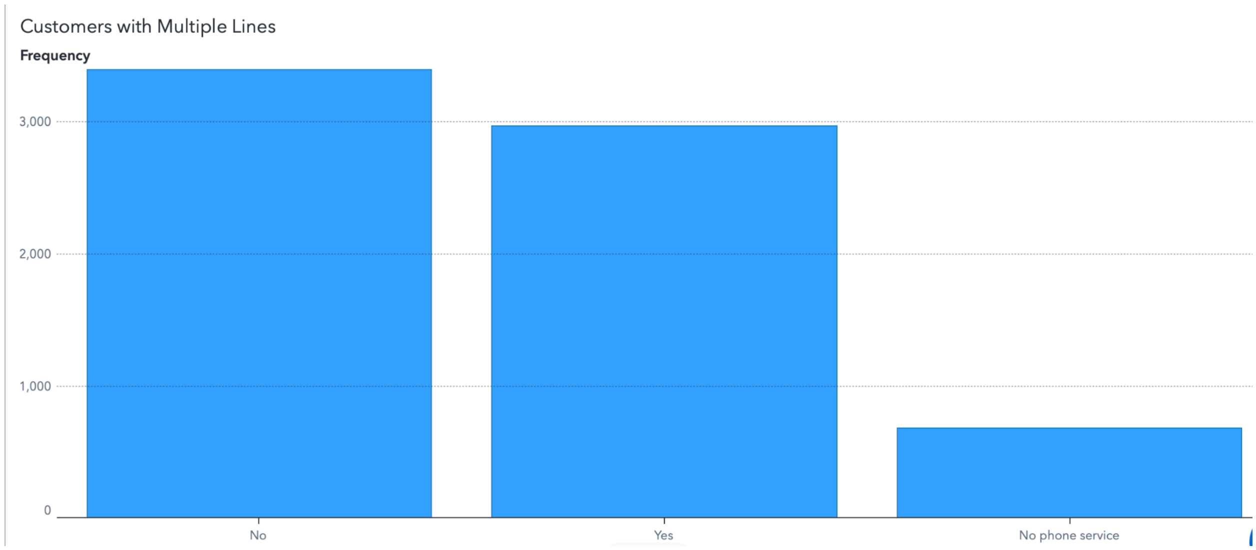


Figure 1.15: Customers with Multiple Lines, Unknown Provider

This chart comes from a data set with 21 columns and 7,044 rows. This can show companies the number of customers that have multiple lines and which ones don't have a phone through them at all. This can show which people they need to advertise their phone plans to and which ones they need to try to get more people on their plan. Made in SAS Viya.

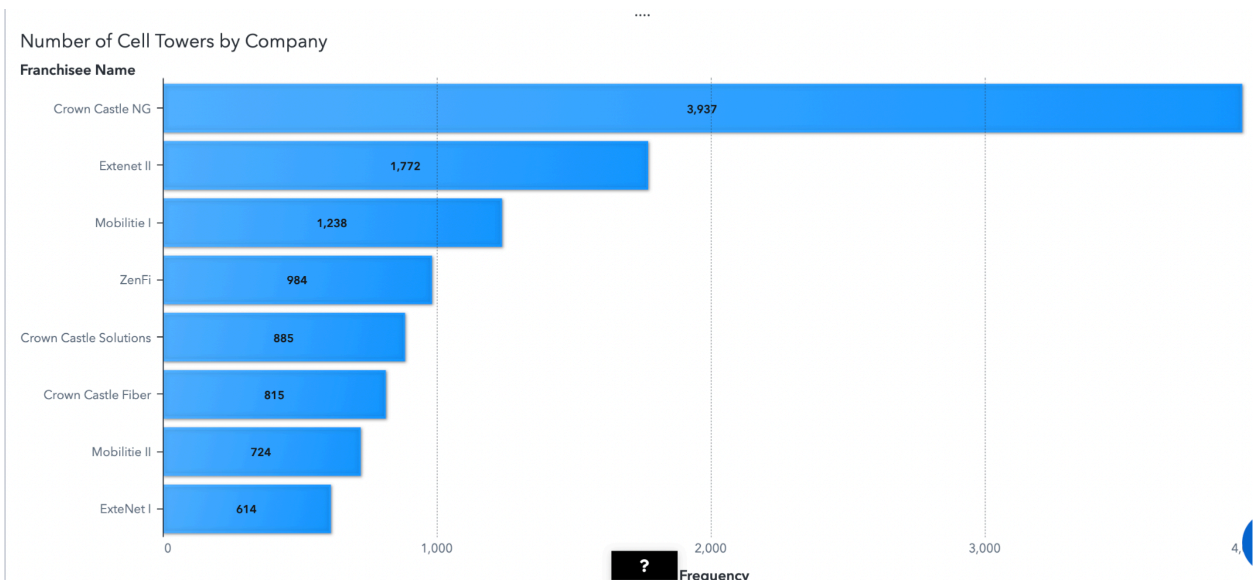


Figure 1.16: Number of Service Towers by Company in New York

This chart comes from a data set with 22 columns and 10,970 rows. This shows which companies have the most towers in New York and therefore are providing the most coverage out of the companies listed. This can show companies which companies they are competing the most with and which ones that are not keeping up. Made in SAS Viya.

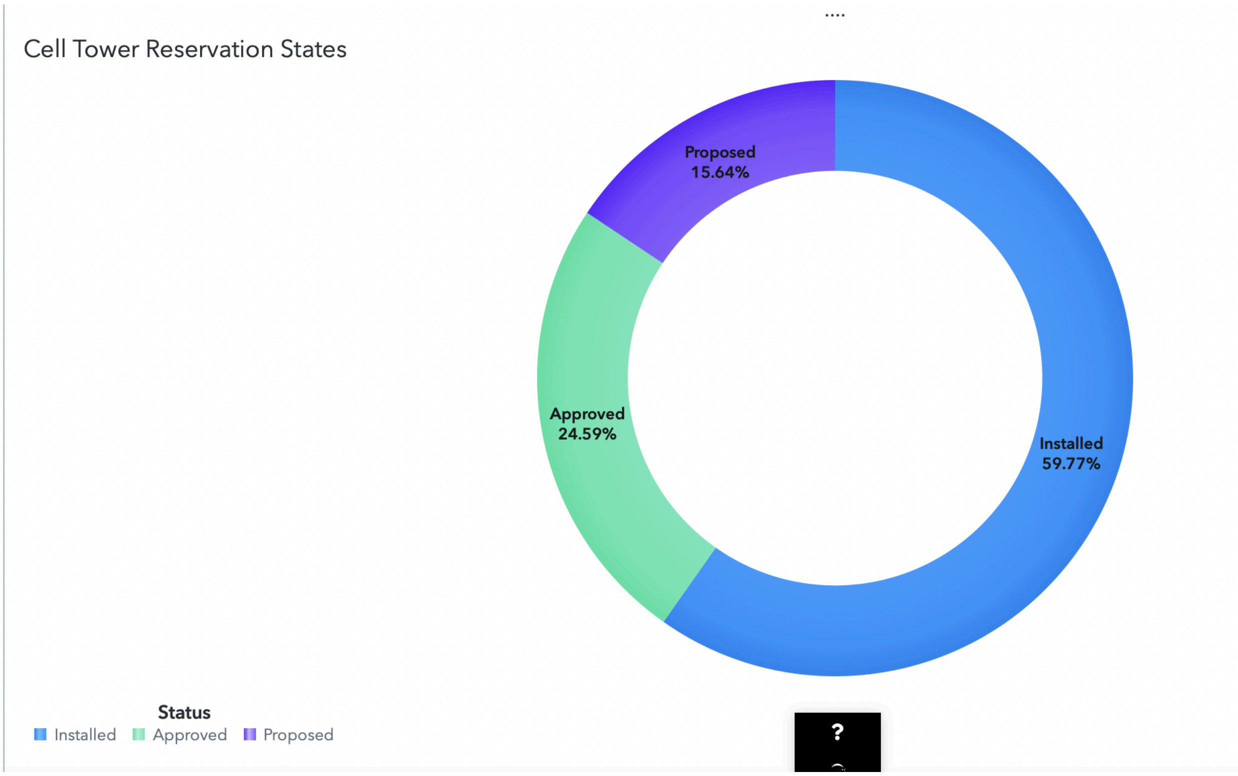


Figure 1.17: Service Tower Reservation State in New York

This chart comes from a data set with 22 columns and 10,970 rows. This shows how many of the cell towers each company has actually installed. It also shows which ones have been approved and how many have just been proposed. This shows companies that there is still plenty of time for the proposal of new towers since there are still 15% that are just proposed. Made in SAS Viya.

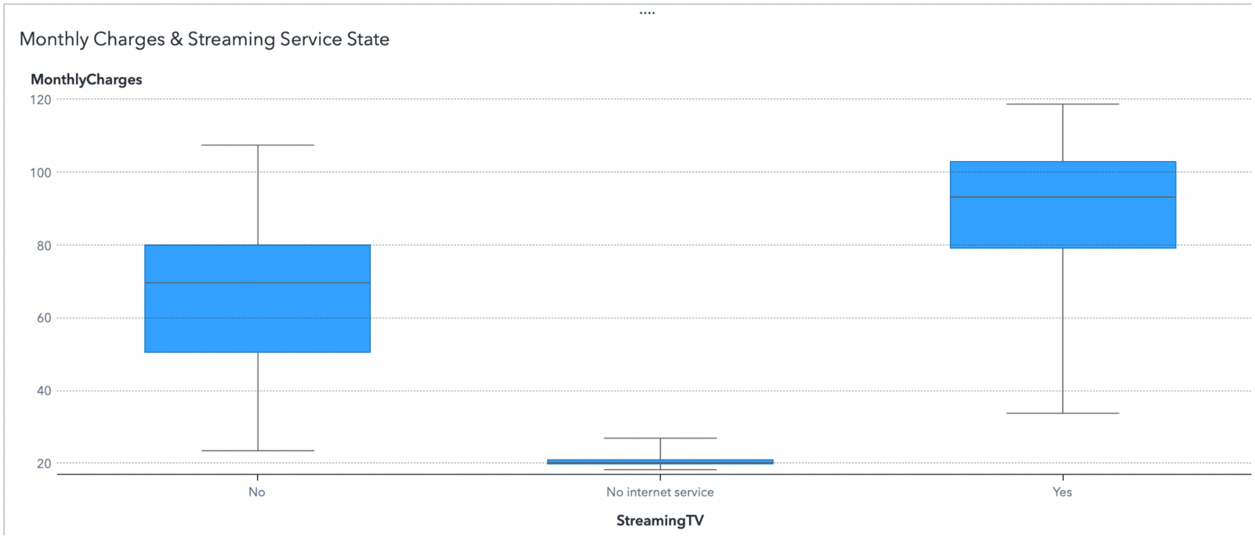


Figure 1.18: Monthly Charges vs. Streaming Service State, Unknown Provider

This chart comes from a data set with 21 columns and 7,044 rows. This shows the monthly charges for people who have streaming services with the provider, people who don't, and people who don't have internet service through the provider at all. This shows that the people who have streaming services have a higher monthly charge rate and the people without streaming services have a large variance in the dataset. This shows the provider that they should try advertising their streaming services more to try and get more people to join that. Made in SAS Viya.

Types of Telecommunications Conclusion

The summary that should be gathered by the charts and graphs shown is that studying the types of telecommunication that customers prefer is a very insightful process that can help telecommunication companies make the right decisions in the furtherment of their companies. Each figure shows their own individual thing but one thing that I believe can be gathered from them cumulatively is that SMS is becoming the least valuable of the three types of telecommunication. Most people are leaning towards the internet and calling over sending a text message. Because of this companies should shift their focus to the other two types to better suit the needs of the consumer. In the Phase II visualizations it shows more on what types of people and areas providers should be focusing their internet coverage and cellular services. It also shows which services are most valuable and which customers are the most valuable.

Figures 1.1 to 1.4 are from Phase I. Figures 1.5 to 1.11 are from Phase II. Figures 1.12 to 1.18 is a part of Phase III. Figures made in SAS Viya are specified, all others are made in Tableau.

Churn in Telecommunications - Tyler

Churn in telecommunications refers to the rate at which users stop doing business with an organization (Vizolution, 2023). This type of data is absolutely critical for businesses to understand what type of strategies they need to adopt to either grow/maintain customer loyalty or to somehow take advantage of the higher or lower churn rates. Businesses need to know why customers are staying or going, and the analysis of all data generated during turnovers can provide useful insights, such as the most common time to leave a provider. This and many other reasons are why it makes sense and is worthwhile for companies to study their churn rates and trends. The descriptive, predictive, and prescriptive statistics provided by these analyses could be a deciding factor in the profitability and prove very important for the very life of the company. In Phase II, relationships with churn will be further evaluated. Churn will be evaluated by factors such as area, reasons for churn, and data type used by customers. The data being evaluated (Telco_Customer_Churn.xlsx) contains over 7,000 rows of data with 33 columns, making it a robust source of information in several areas of concern for the telecommunications firm.

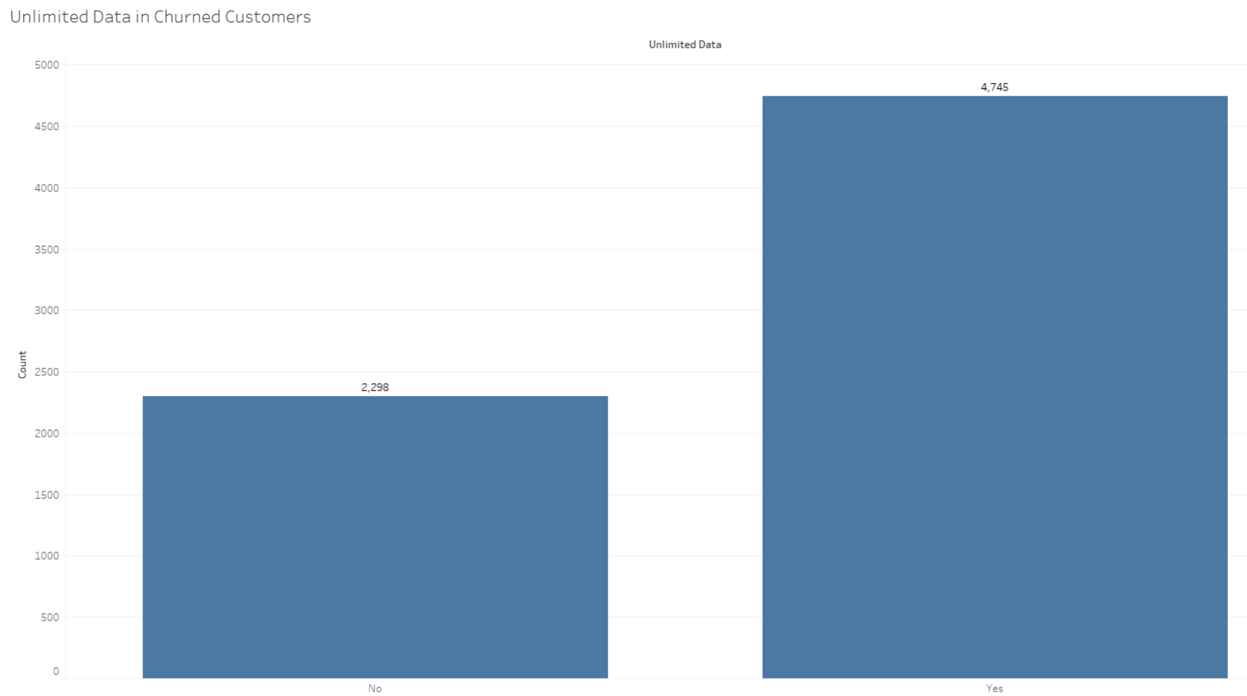


Figure 2.1: Unlimited Data in Churned Customers

Figure 2.1 highlights the difference between churned customers that had unlimited data and those that did not. It helps us conclude that approximately two thirds of customers that left had unlimited data. This is insightful because it tells us that even though unlimited data was offered, it did not have a significant impact on customer turnover. There are likely better options to consider when trying to account for churn.

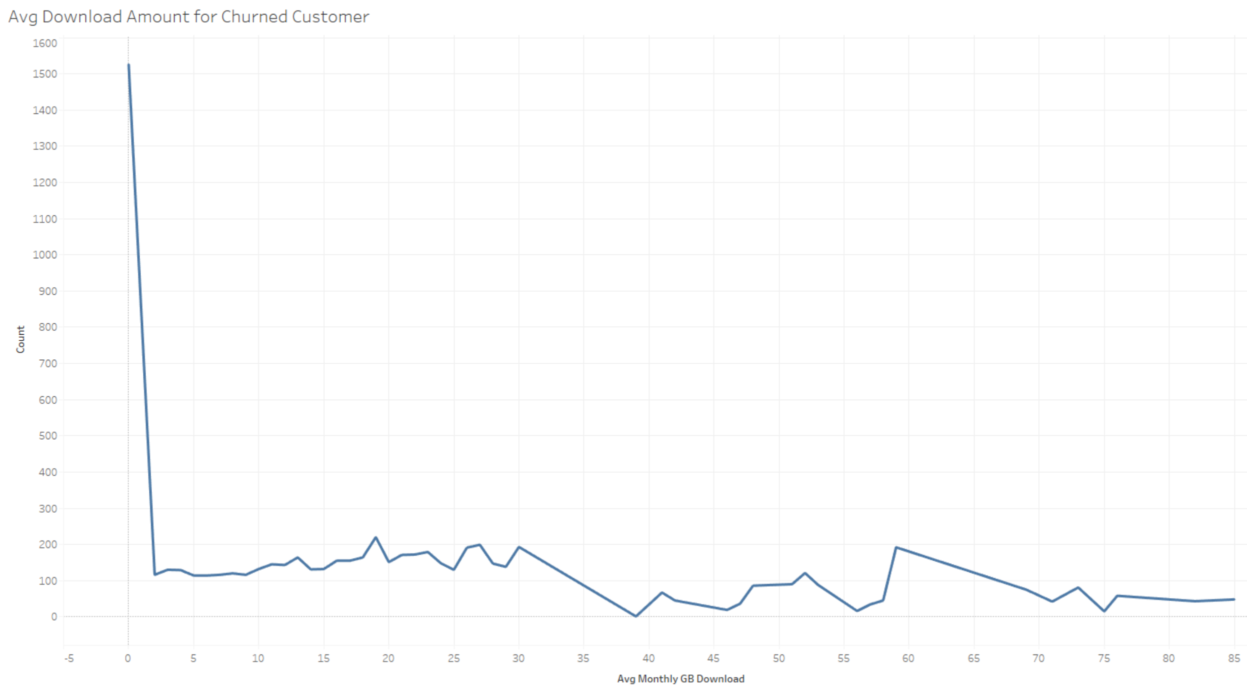


Figure 2.2: Average Download Amount in Churned Customers

Figure 2.2 shows us the average download amount per month that customers who left the provider had. It gives us insight into the traffic activity of these customers. It shows a significantly higher amount at 0 GB, meaning that customers who left were not using very much internet data. This could be due to a number of factors, such as age, device type, and time with the provider, keeping in mind this data was generated for only one quarter. It should be considered that turnover is happening more in this group in particular, but not necessarily more than all other download amounts in total.

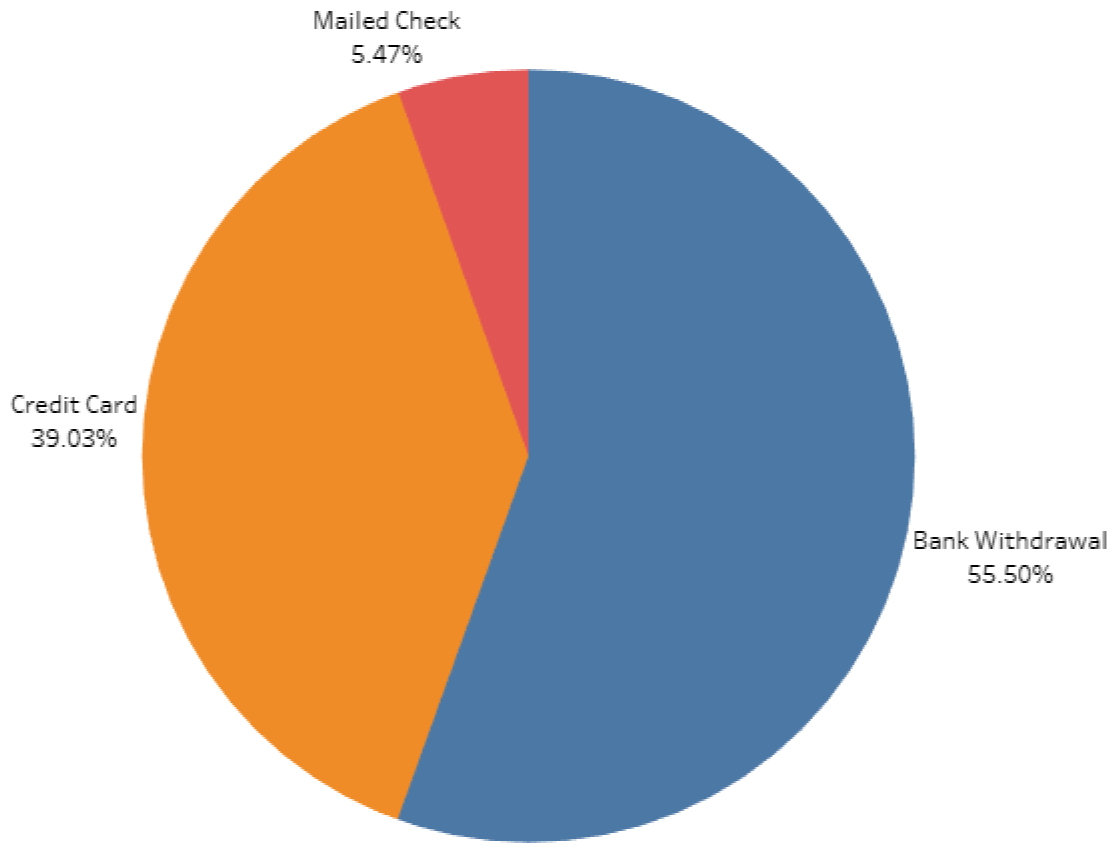


Figure 2.3: Payment Methods of Churned Customers

Figure 2.3 showcases the percentage of customers that left the provider based on their type of payment. This chart is sourced from selected elements in the data source, making it a result of big data analysis. This chart shows us that the payment methods used in churned customers are likely similar to those who stayed with the provider. It tells us that there is no significant correlation between payment type and churn rate, and that we should consider another target type when trying to decrease churn.

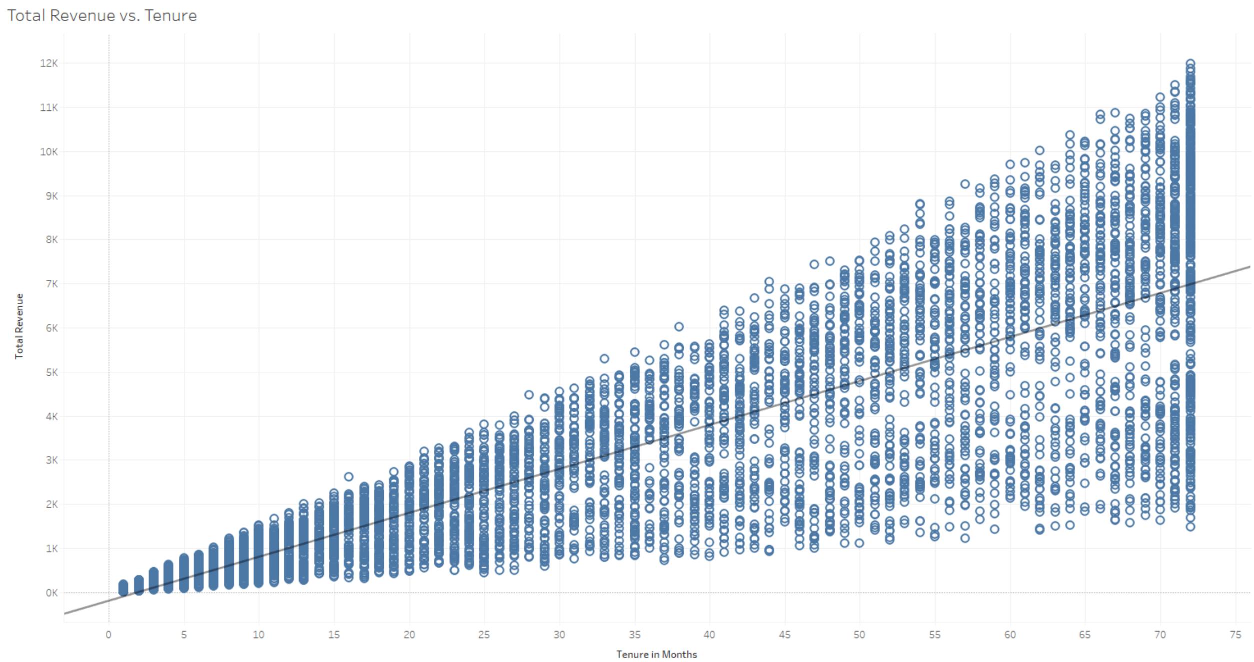


Figure 2.4: Total Revenue vs. Tenure

Figure 2.4 shows total revenue generated by each customer versus their respective tenure. We find that as tenure increases, the revenue generated by these past customers begins to vary much more than in those with lower tenure. This is likely worth studying more, as it seems more logical that customers who were with the company longer should generate the most, but that is not the case. It should be a question of awareness about other plans, and what can the provider do to make customers more aware of buying options.

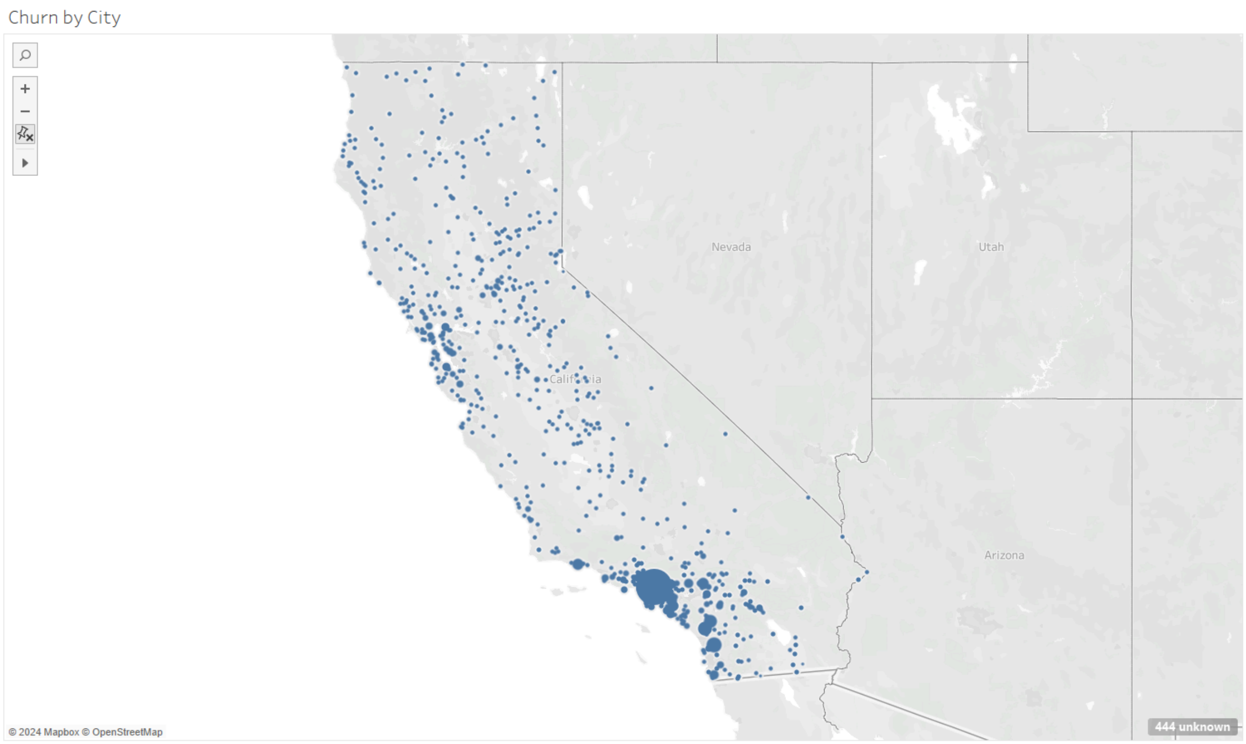


Figure 2.5: Churn by City

Figure 2.5 shows the distribution of curbed customers by city in California. This chart is representative of the 216,000 records in the data source. As expected, the Los Angeles area contributed the most to the pool of churned customers. This is expected because of the disproportionate amount of people in the area. This representation does not provide much insight into why people are switching telecommunications providers.

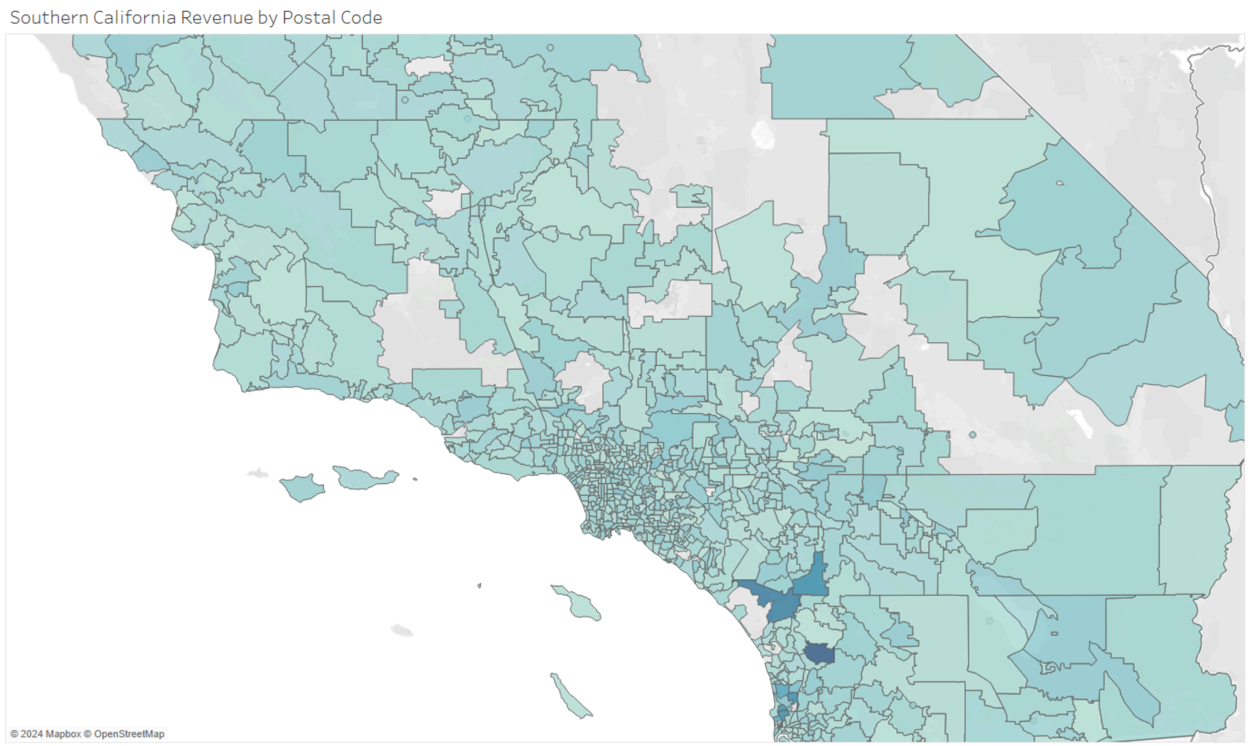


Figure 2.6: Southern California Revenue by Postal Code

Figure 2.6 is useful for determining the revenue generated by each individual postal code. Contrary to the figure before, the greater Los Angeles area does not provide as much revenue as other areas in question. This should create concern about the lack of presence in the L.A. area, as that is a huge market. This could be a significant result of the churn in the area being as high as it is.

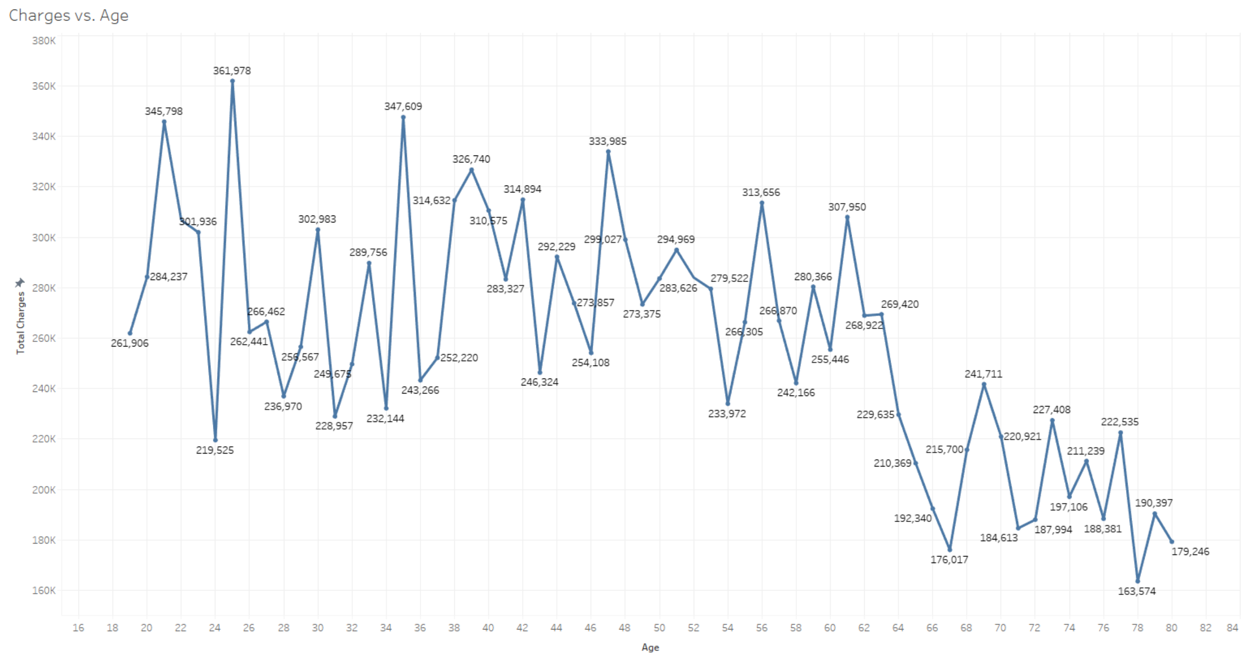


Figure 2.7: Total Charge for Telecommunications Services vs. Age

Figure 2.7 gives insight into what churned age groups were being charged before they left the provider. Big data is deployed here because each year represents all customers in that age group, meaning each point is a sum of all customers that are that age. Younger people, on average, were being charged more. This is likely due to higher data usage or data consumption. It could be worth investigating the younger audiences' reason for leaving, as they provided significant money in response to their charges.

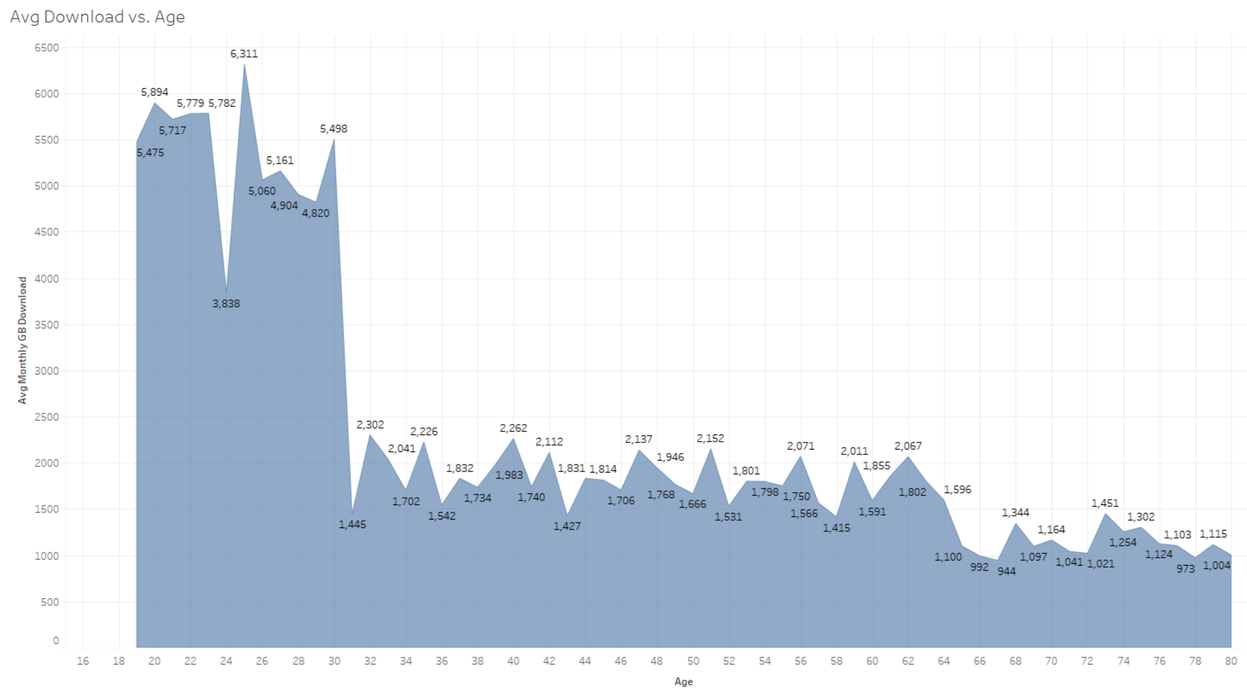


Figure 2.8: Average Monthly Download vs. Age

The takeaway from Figure 2.8 is similar to that of 2.7. This is big data because it is representative of the data source’s 216,000 records. Younger people were consuming more data than other age groups. This should raise flags that younger people do generally consume more data than older groups, and it is worth investigating ways to cater to this audience and prevent as much churn as possible. It is also worth noting that this trend will likely continue into new internet users in the future, so it would be worthwhile to figure out a general solution now and reap the benefits later.

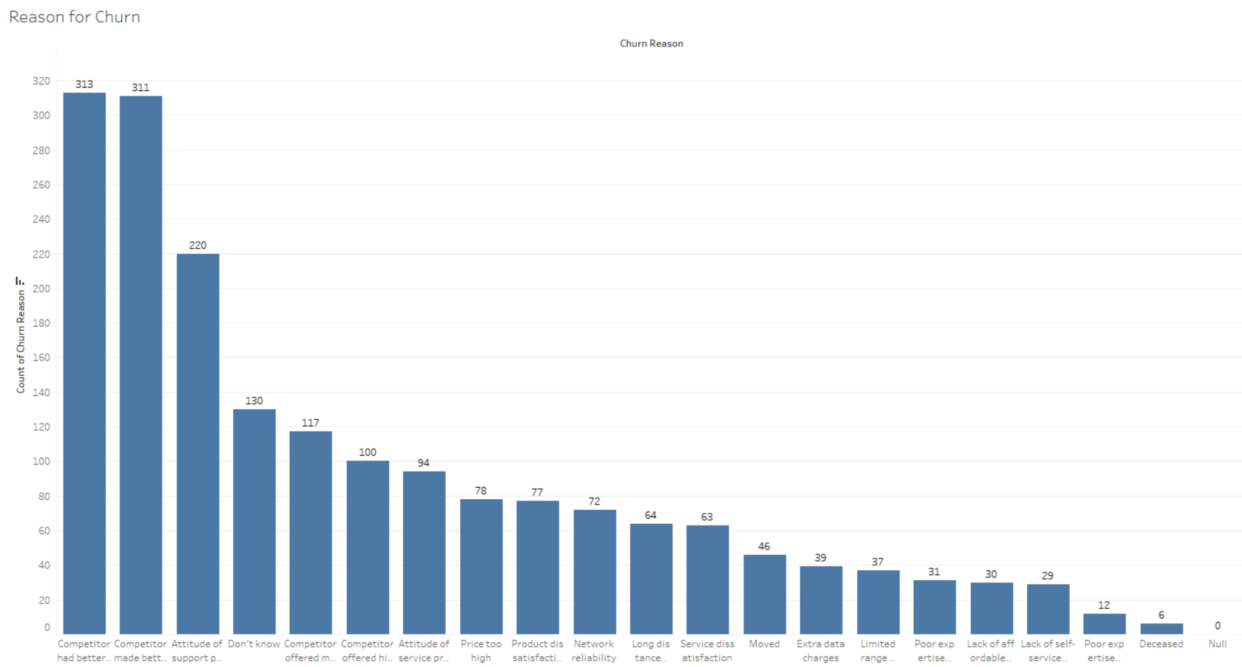


Figure 2.9: Count of Reasons for Churn

As a preface to Figure 2.9, it's important to note that the vast majority of data did not provide a reason for leaving the provider. The big data represented here does not include the portion of the 216,000 records that did not provide a reason for leaving. However, we can examine the respondents that did answer to get a general view of the landscape. A significant portion of churned customers left due to reasons like better offers or devices from competitors. This is important for the provider to understand. Another significant portion of churned customers left due to poor customer service. This is worth knowing as well so the company can evaluate their customer service department or the company they have outsourced this job to.

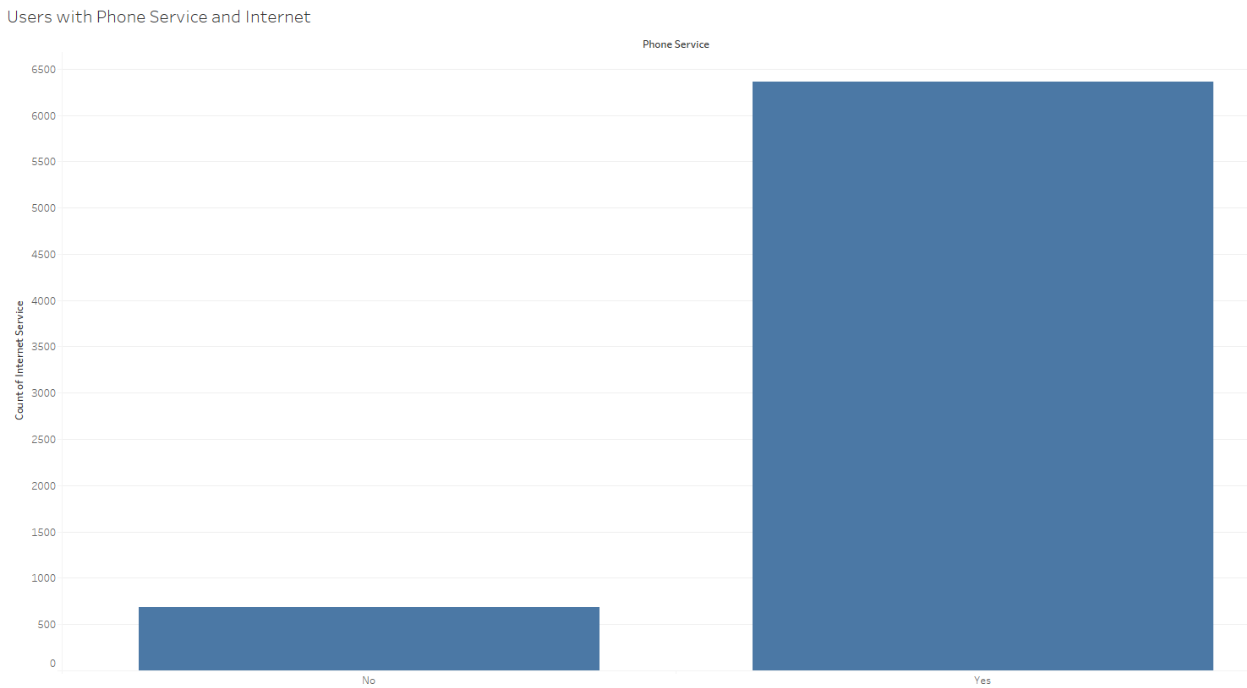


Figure 2.10: Internet and Phone Service

Figure 2.10 highlights the importance of providing suitable internet services. This is from a big data source, however it only examines a single column of the 33 available, making it appear smaller. The vast majority of customers that were churned also purchased internet services. Pair this with 2.9, there seems to be some link between devices, internet, and customer service that is common amongst churned customers. If the provider does not currently match the capabilities of others to provide these services in tandem, it will likely mean more churn.

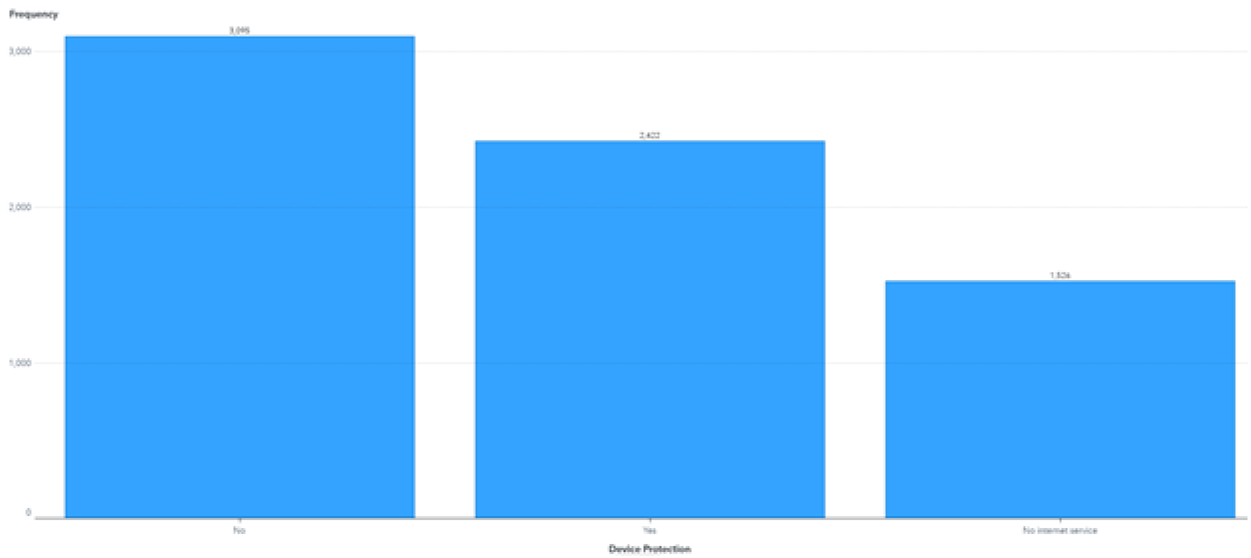


Figure 2.11: Device Protection Rates

Figure 2.11 draws on big data from over 216,000 records. It showcases the amount of churned customers that were subscribed to some type of device protection. While it appears initially that it could be correlated to churn, it is worth noting that some customers likely refused the protection with no basis. This would not be a good place to search for causal relationships to churn.

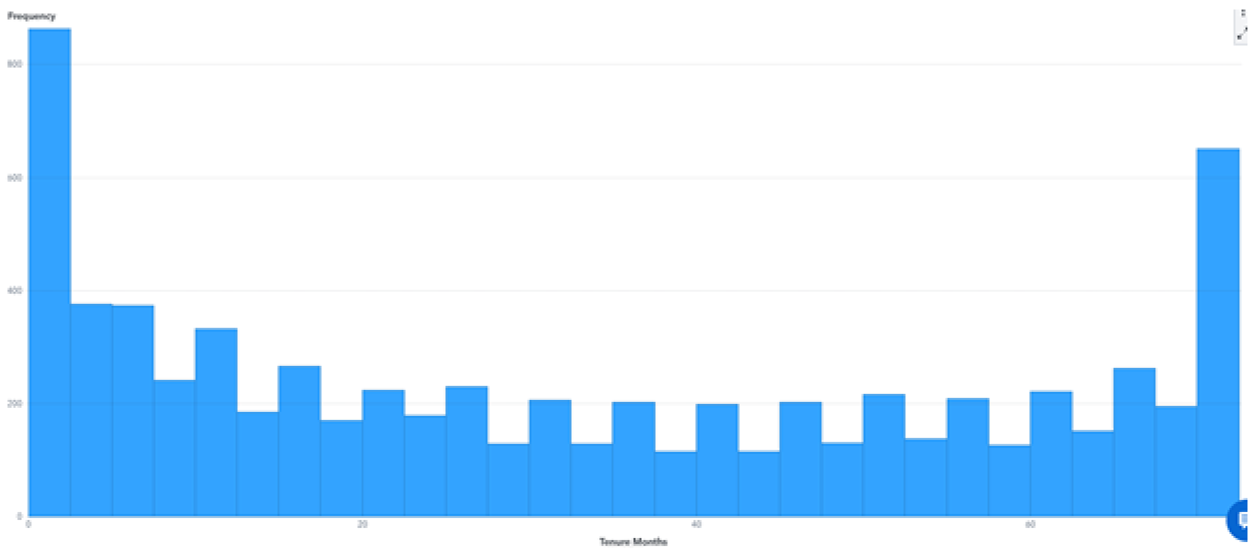


Figure 2.12: Churn Score and Tenure

Figure 2.12 comes from the same 216.000 record data source as previous figures. It can be seen that there is not a strong relationship between the customer's churn scores (score developed to indicate likelihood of churning) and their tenure with the company. The only significant knowledge gained from this chart is that newer customers tend to churn faster and so do longer-tenured customers. It could be worth investigating why longer-tenured customers are leaving more.

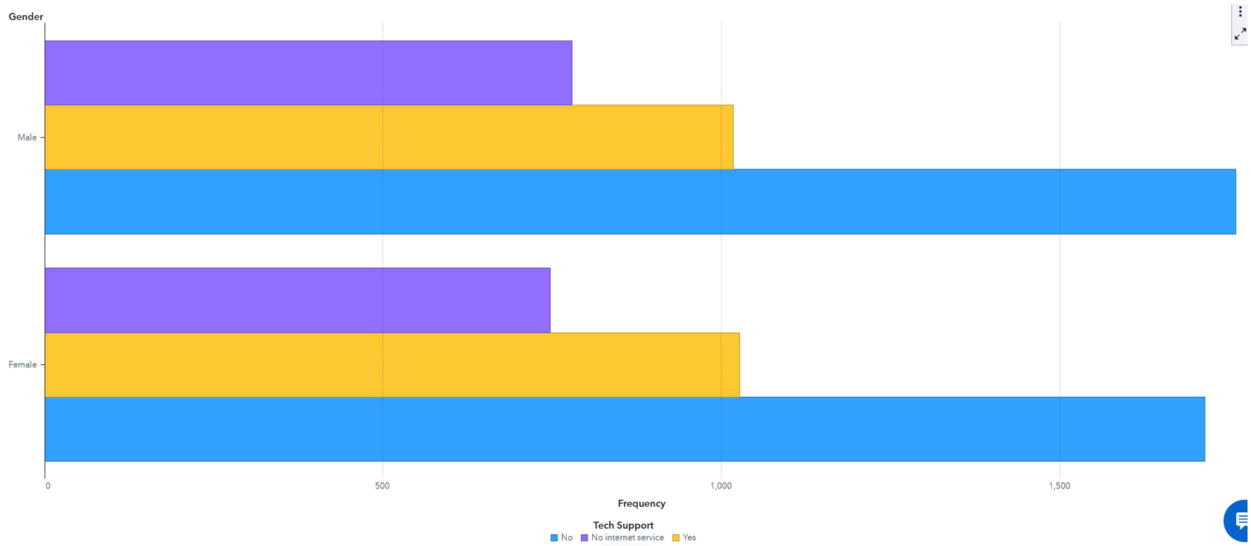


Figure 2.13: Gender and Technical Support

Figure 2.13 comes for the same 216,000 record table as previous figures. It highlights a distinction between the technical support received by men and women. It should be investigated if there is some type of internal bias with the technical support team when it comes to helping men or women. It could also simply be that more men have their names on the accounts, but it is worth investigating.

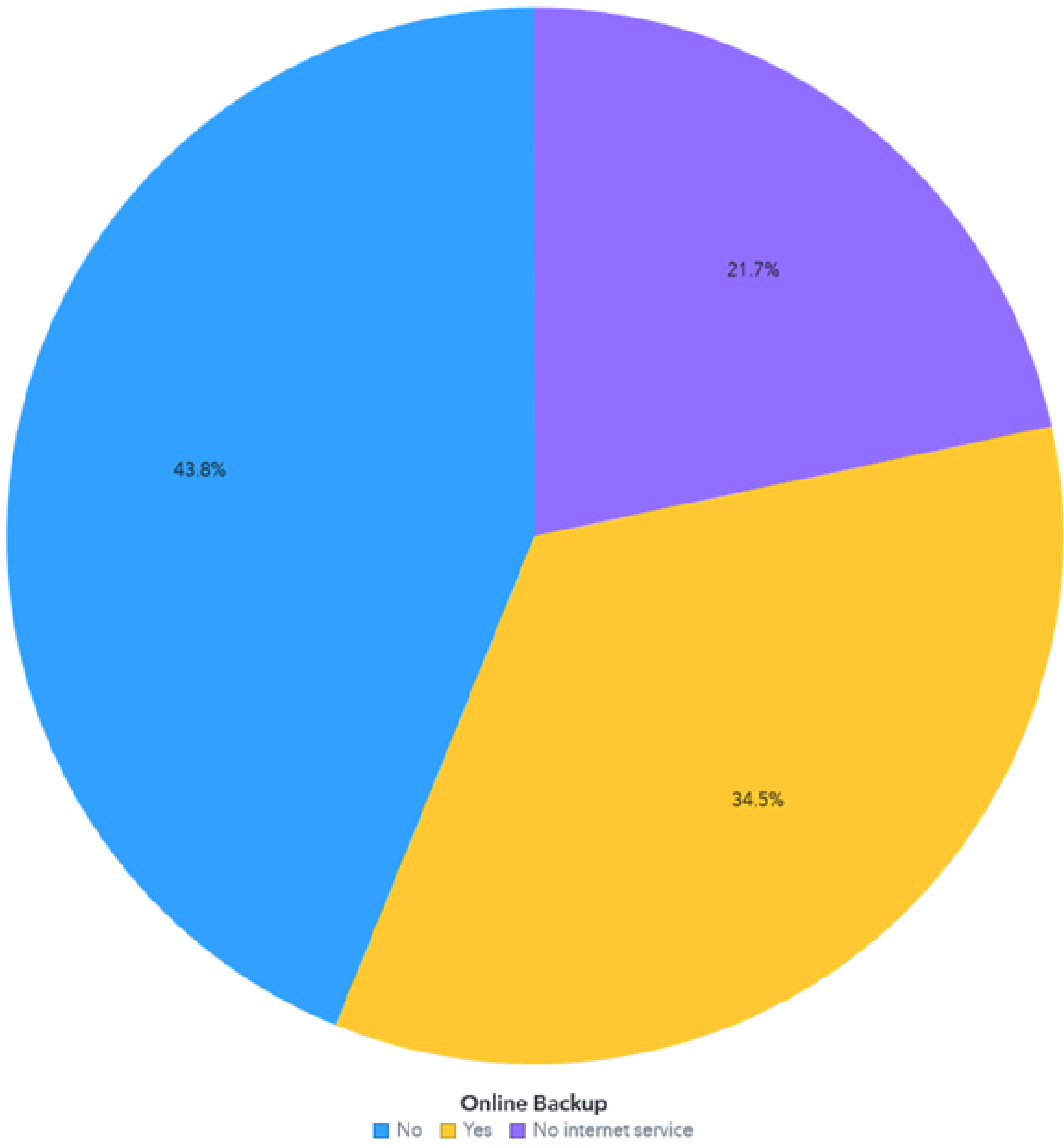


Figure 2.14: Online Backup Usage

Figure 2.14 is based on a source of over 216,000 records. It showcases the percentages of users that took advantage of the online backup service provided by the telecommunications firm. There seems to be no significant indication or anomaly that would suggest that the use of the service could somehow be indicative of churn.

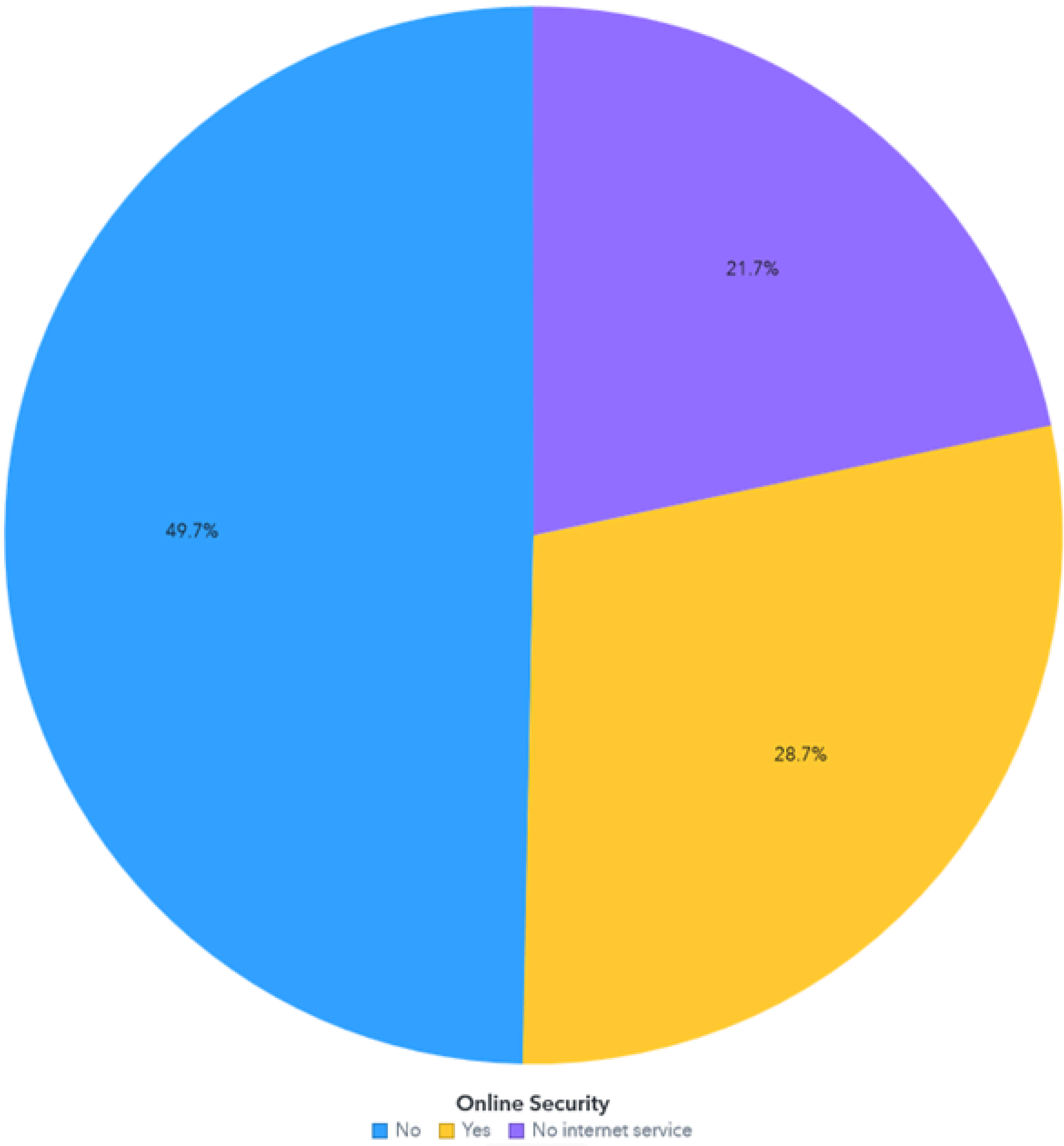


Figure 2.15: Online Security Features Usage

Figure 2.15 is from the same 216,000 record source as the previous figures. It highlights the percentage of users that were using the online security features provided by the firm. It is worth noting that almost half of the churned customers were subscribed to this service before they left. It is possible that a more enticing service caused them to leave or some other type of better bundle or offer.

Churn in Telecommunications Conclusion

Phase II of the analysis of the customer churn data has been more forthcoming. It was discovered that there were some links between the reasons that people were churning, the age groups, and data usage. This is significant because it gives the provider some insight into why people are leaning into other companies to provide certain services. The provider needs to look into how to cater to a larger amount of data consumption amongst users. This is directly correlated to the age of users and what size of them are leaving due to being more digitally literate, connected, and having higher expectations of their hardware. It is also important to know what type of marketing strategies and promotions should be made available in order to minimize churn, particularly in this group as they age and their children become potential customers. Phase III, however, was not as insightful as the previous phase. The discovery of relationships was more difficult than the past two phases. Several relationships were discovered, but they were unlikely to carry any significant meaning. It is apparent that the company needs to work on catering to large data users, respective age groups, and hone in on a more digitally literate generation and customer base.

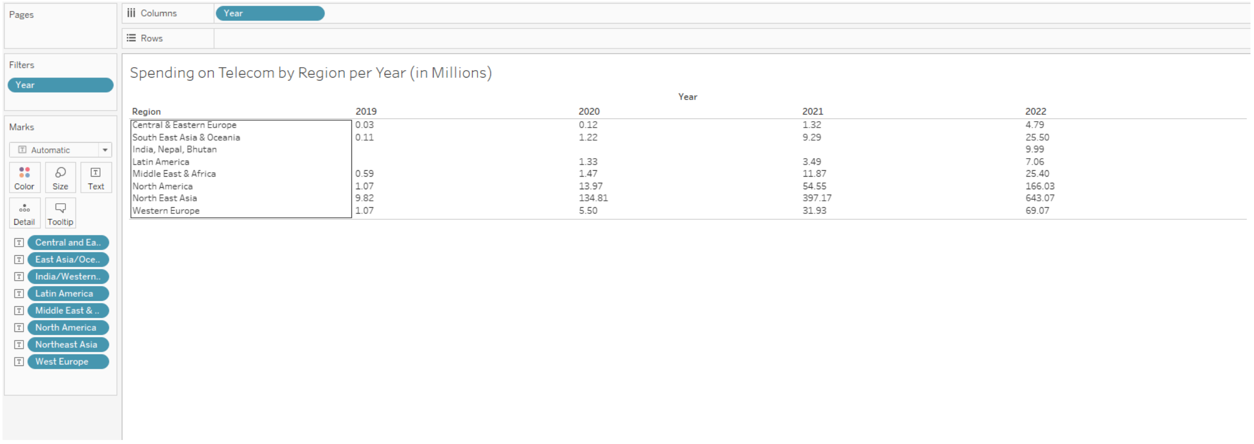
Data as Tested

https://docs.google.com/spreadsheets/d/1I_kQ6bbBb5QAeYt5-hGw6LUH7KrMMDQs/edit?rtpof=true

Telecommunication Data by Region - Caleb

Telecommunication data is available effectively everywhere on Earth save for less developed nations. The spending, amount of data sent or received, number of cell towers, and number of telecommunication customers will vary greatly depending on the region that they live in. Each region has its own providers of telecommunication services, different demographics for the type of people using the telecommunication services, and different geographical locations that the telecommunication company must work around and for. This would have companies work around regulations of countries, bad terrain for cell towers, and knowing the demographics and societal norms of each region to cater to the customer base as best as possible.

In addition to this, regional statistics give an insight into the state of the telecommunications market in specific areas in the world as well as worldwide. This information would be vital to operations in free markets around the world for the purpose of outperforming competition. It would also be useful for predicting upcoming expenditures and budgeting. Another use this Big Data would have is from the bandwidth being used by region in the telecom industry. This data would show the needs for bandwidth of specific areas of the world. This is very useful for decision makers in the industry as they would be able to more accurately tell what upcoming improvements or changes need to be made to the backend systems.



Spending on Telecom by Region per Year (in Millions)

Region	2019	2020	2021	2022
Central & Eastern Europe	0.03	0.12	1.32	4.79
South East Asia & Oceania	0.11	1.22	9.29	25.50
India, Nepal, Bhutan				9.99
Latin America		1.33	3.49	7.06
Middle East & Africa	0.59	1.47	11.87	25.40
North America	1.07	13.97	54.55	166.03
North East Asia	9.82	134.81	397.17	643.07
Western Europe	1.07	5.50	31.93	69.07

Figure 3.1: Spending on Telecommunications Worldwide by Region

(Dataset of 4x4000 Columns/Rows)

Telecommunication companies worldwide over just the past 5 years have increased expenditure sharply. In all regions this remains true with the India, Nepal, Bhutan region missing information until more recently. In 2020 Northeast Asia increased spending more than 10x from 2019 on telecommunications, and in 2021 more than tripled from 2020. North America did similarly, with much smaller figures. This shows massive growth in the market in these regions, while the rest of the world shows similar trends with smaller figures overall.

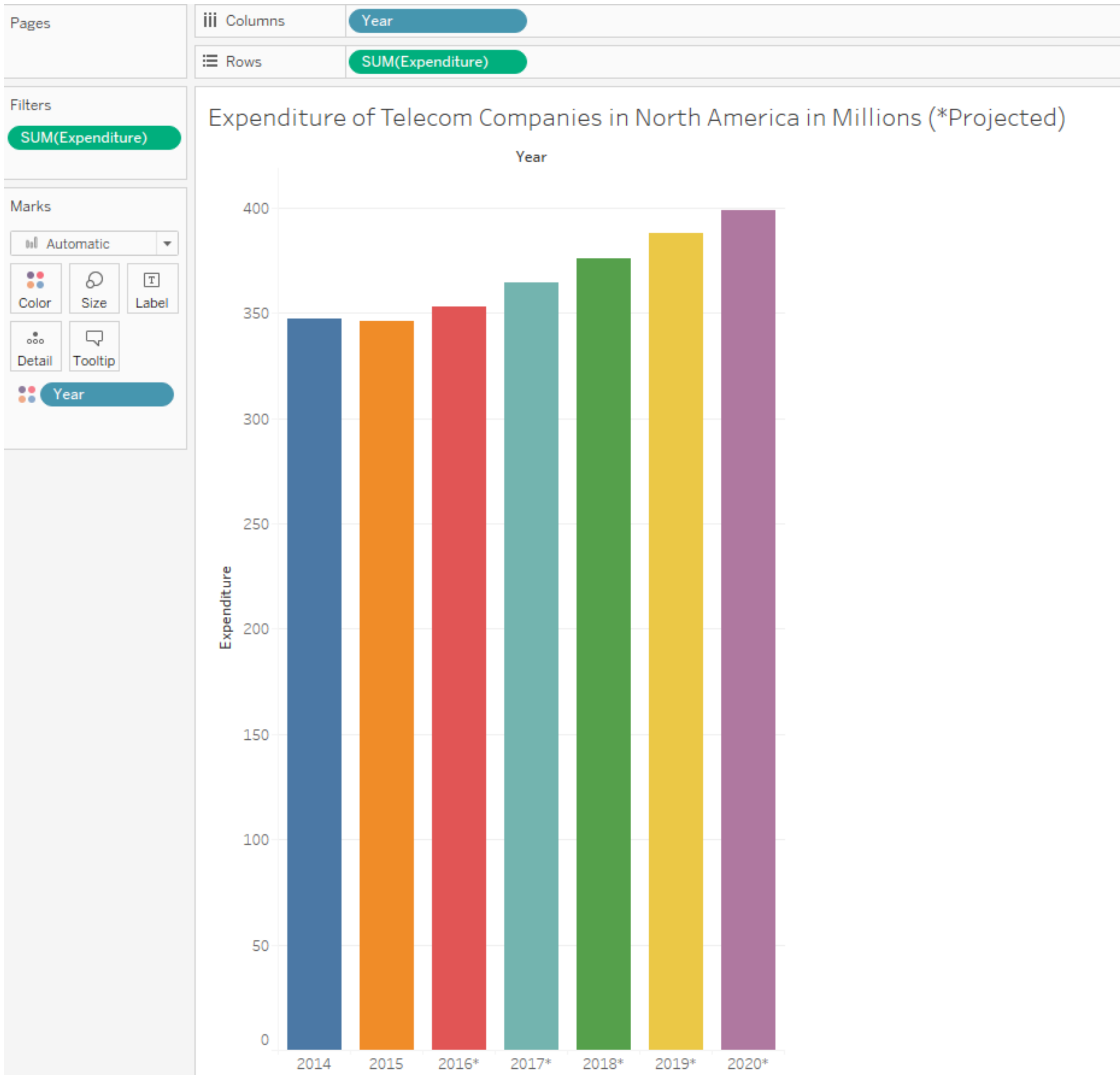


Figure 3.2: Past North American Telecommunication Expenditures

(Dataset of 7x2000 Rows/Columns)

This figure shows a projection of the amount Telecommunication firms in North America expected expenditure to grow by from 2015 to 2020.

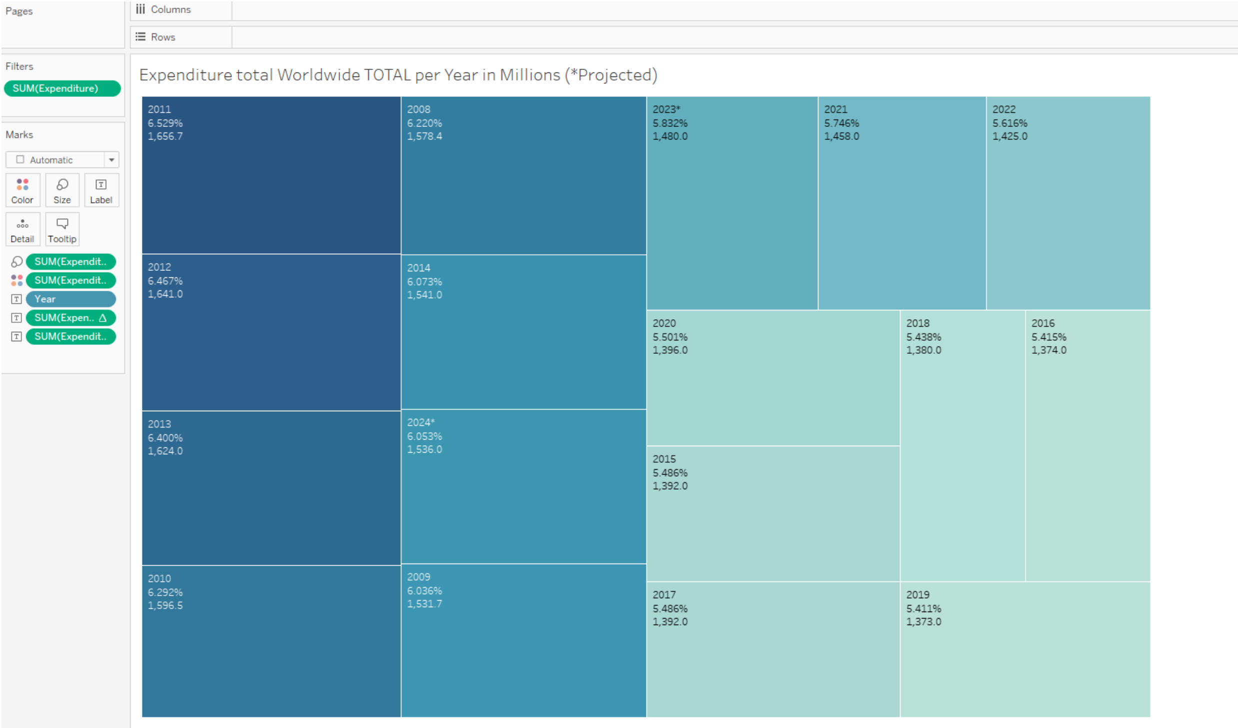


Figure 3.3: Worldwide Expenditure on Telecommunications Growth (Dataset of 4x4000 Columns/Rows)

This figure shows the growth of the telecommunications expenditures worldwide as a percentage. The year with the highest percentage growth was 2011, while 2019 was the lowest. Compared to the last figure and the one before, this shows that regional statistics may not necessarily be influenced by the international markets, or that North America does not determine the growth of the international market.

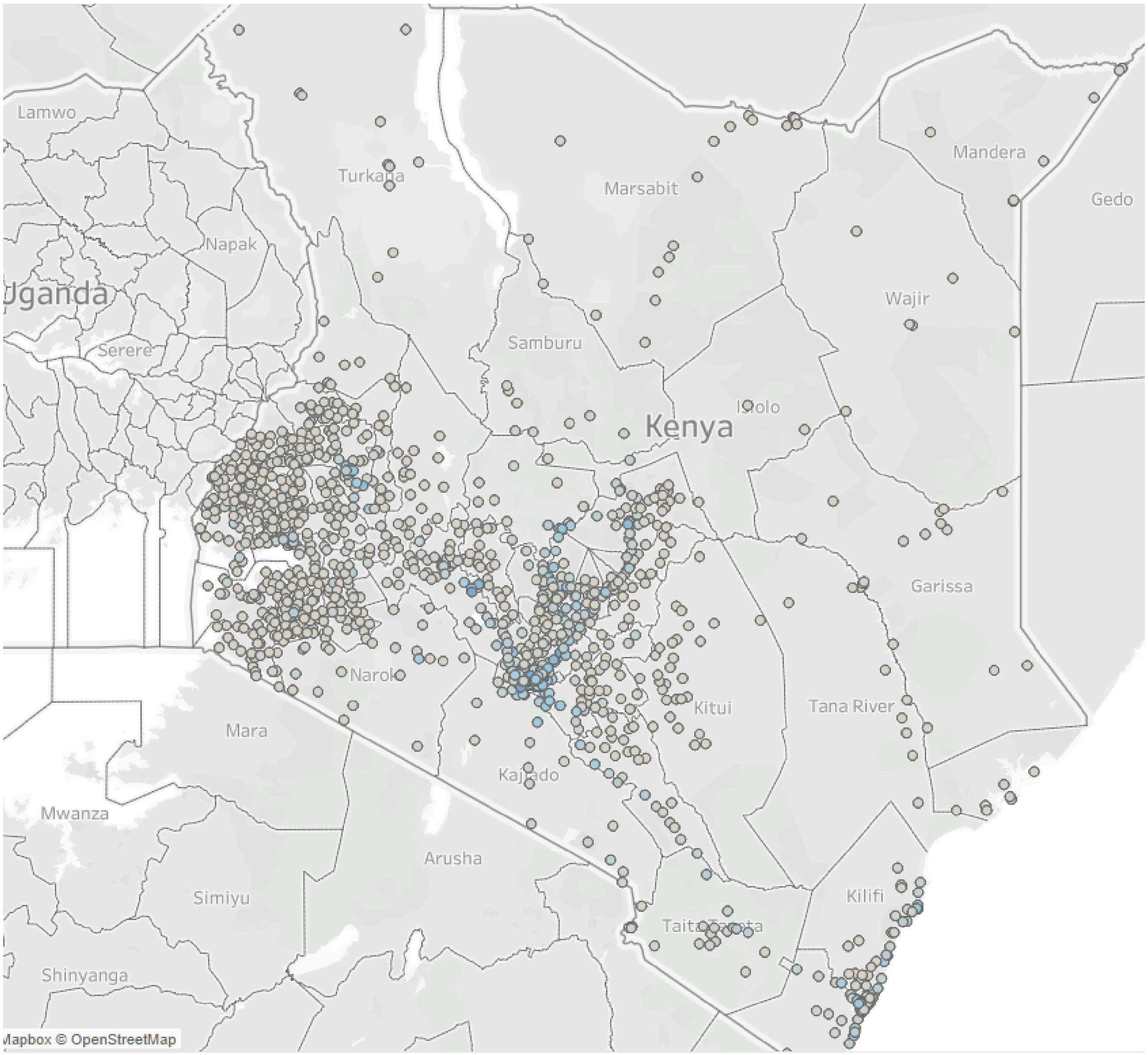


Figure 3.4: New Telecommunication Towers in Kenya

(Dataset of 2x1000 Rows/Columns)

Another region experiencing rapid growth is Kenya, shown in this figure. The blue dots are established cell towers, while the gray dots represent only towers built in the last 5 years.

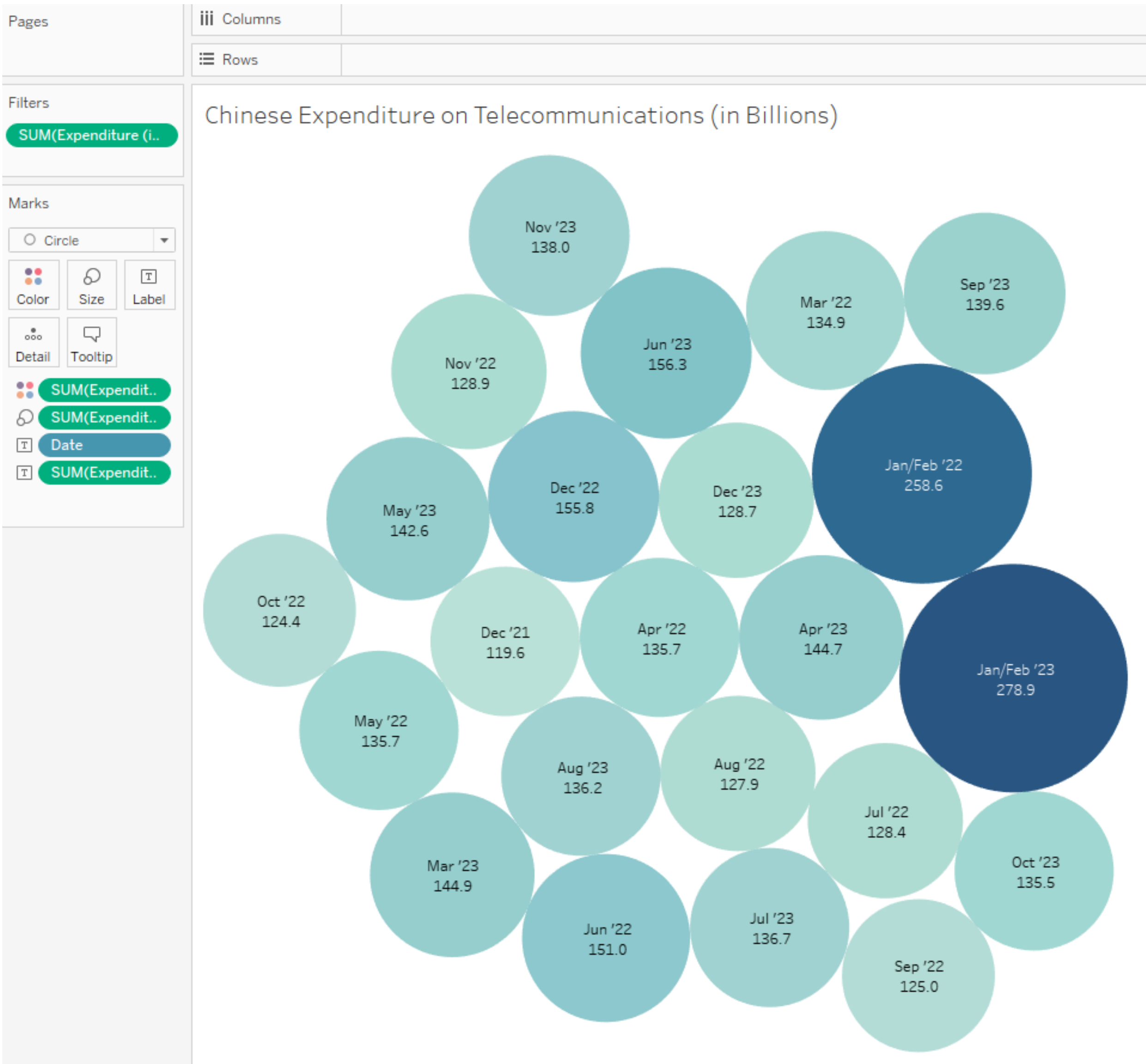


Figure 3.5: Chinese Expenditure on Telecommunications

(Dataset of 7x5000 Columns/Rows)

A region that seems to have slowed in growth (aside from the outliers in Jan/Feb 2022/2023) is China, where over the course of each year the growth either stops or reverses. No growth is shown over 2022, before a sudden massive jump in January of the next year, followed by stagnation again over the rest of 2023.

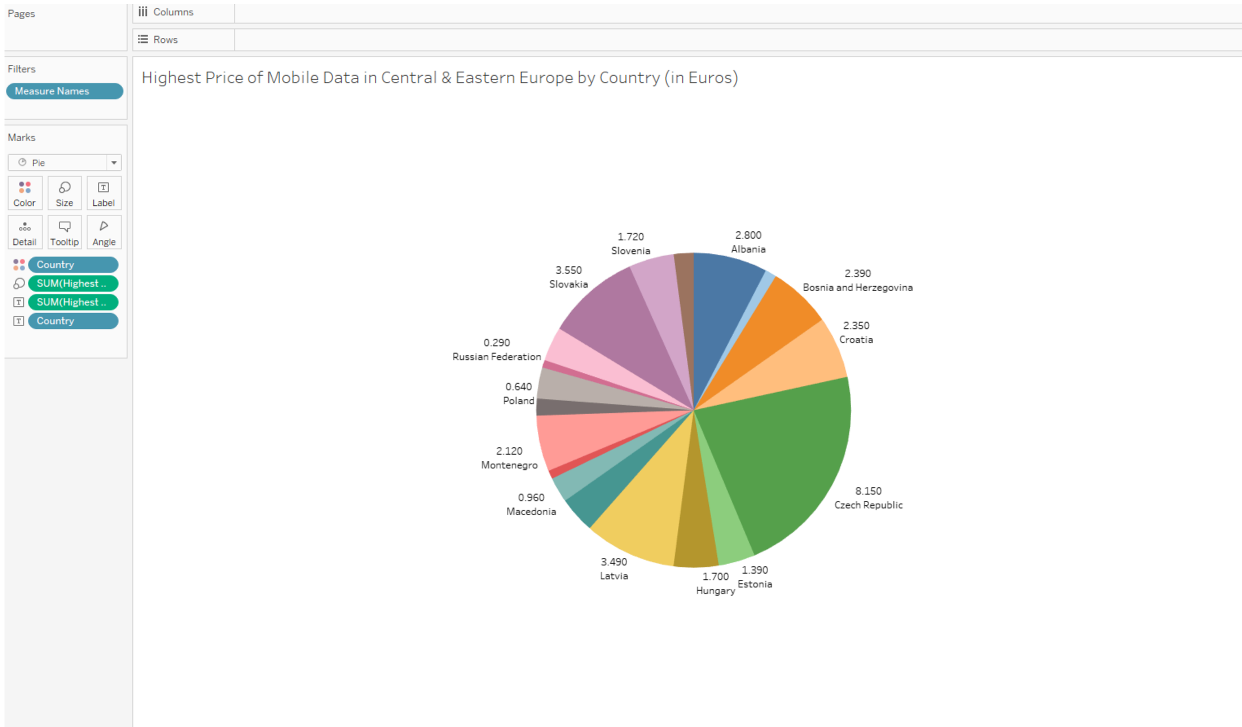


Figure 3.6: Highest Prices of Data in Central & Eastern Europe by Country (Dataset of 4x5000 columns and rows)

This figure shows the highest prices observed for mobile data across Central and Eastern Europe between 2021 and 2023. This could be useful for gauging either the economic state of specific countries (in Central or Eastern Europe) or showing what areas Telecommunication operations could be more costly/less lucrative to serve.

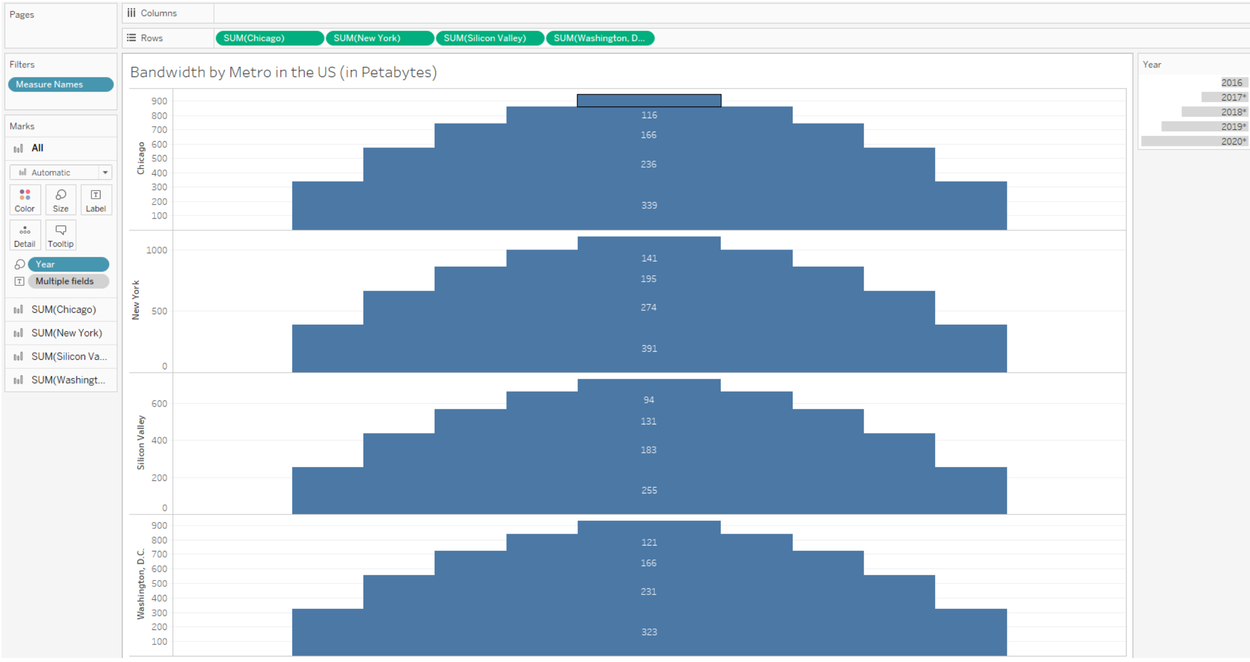


Figure 3.7: US Metro Bandwidth Used between 2016 and 2020

(Dataset of 4x5000 columns and rows)

In the US, bandwidth is always in massive demand, and especially so in metropolitan areas. The measure of bandwidth in high population areas is extremely important for telecommunication companies to predict data needs for the areas. This is an opportunity in Big Data because this is a need by companies for it.

The bars' thickness is what year the data was collected (or projected for).

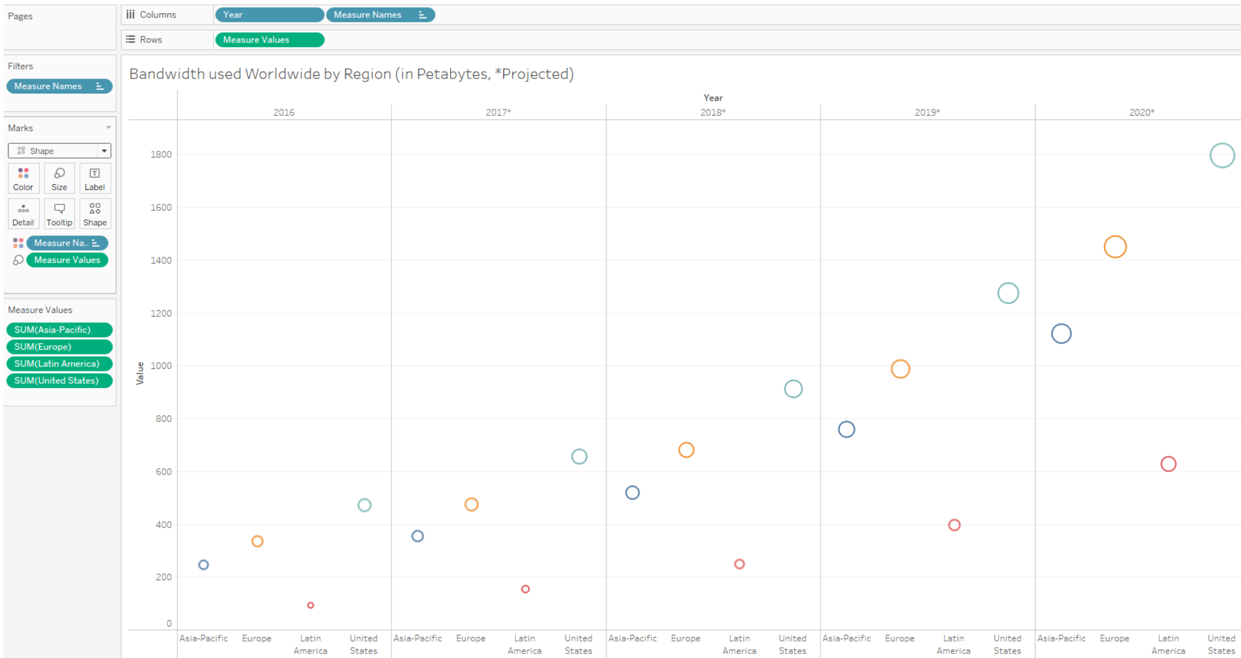


Figure 3.8: Bandwidth used Worldwide by Region

(Dataset of 4x5000 columns and rows)

Side-circles figure of bandwidth broken down into years and regions, worldwide. This figure is a good measure of exactly how much growth is happening in each region per year. Every region here overall is experiencing growth, with all of the regions being projected to have doubled at least.

Figure 3.9: Calls and SMSs in cities across India

(Dataset of 18x1000 columns and rows, SAS Viya)

India is a very large population center with over 1 billion people living there. As such, its impact on telecommunications worldwide is comparable to that of China. This figure shows phone calls and SMSs sent across multiple larger cities in India, which in turn shows the scale of the data moving across India.

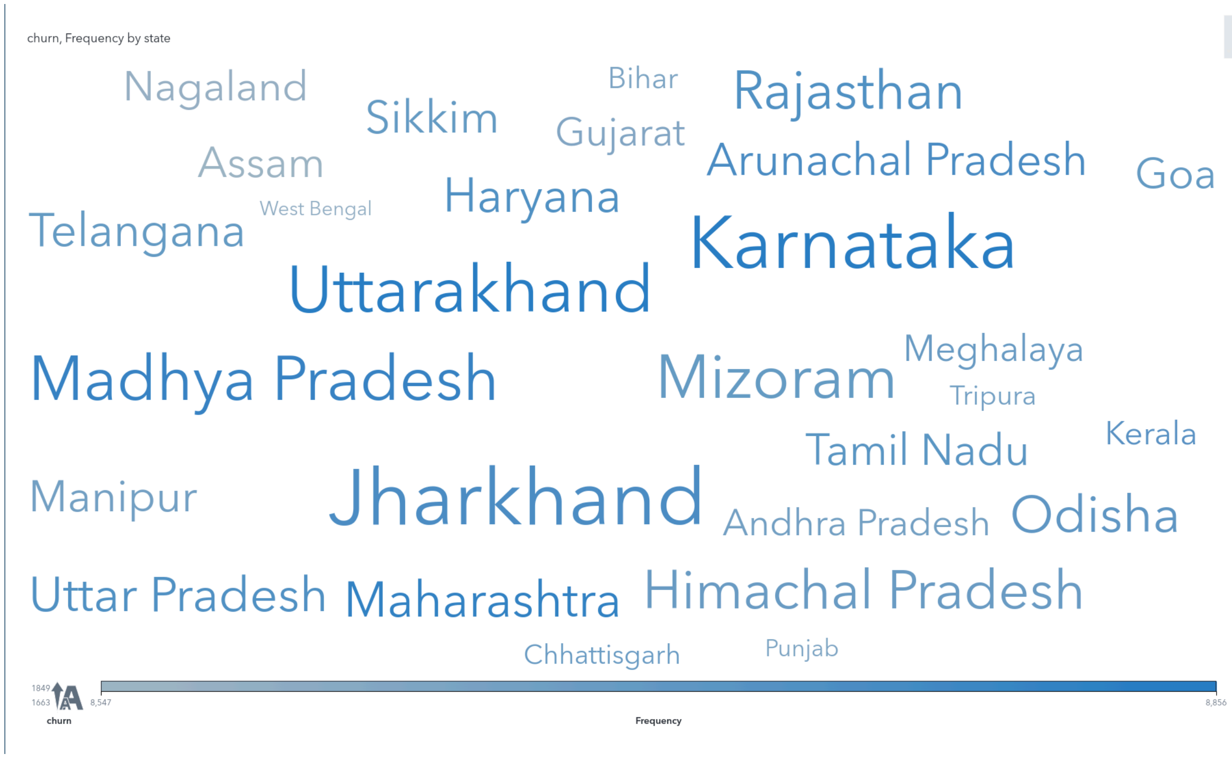


Figure 3.10: Churn of States in India

(Dataset of 18x1000 columns and rows, SAS Viya)

This figure is a word cloud, showing the churn rate of various states of India. The larger and less faded the name of the state, the higher its churn rate is. This helps the telecom industry identify which areas are having a lot of movement between telecommunication providers, which shows opportunities for companies to capitalize on.

Calls and SMSs of Cities in India | Churn of States in India | Data Used (in gigabytes) by City and State India

city ▲	Bangalore	Chennai	Delhi	Hyderabad	Kolkata	Mumbai
state ▲	**data_used**	**data_used**	**data_used**	**data_used**	**data_used**	**data_used**
Andhra Pradesh	7.5M	7.1M	7.1M	7.5M	7.1M	6.9M
Arunachal Pradesh	7.3M	7.6M	7.3M	7.2M	7M	7.4M
Assam	6.9M	7.4M	7.6M	6.9M	7.2M	6.9M
Bihar	7.5M	7.2M	7.1M	7M	7.2M	7.3M
Chhattisgarh	7.2M	7.2M	7.1M	7.2M	7.7M	7.2M
Goa	7.2M	7M	7.1M	7.3M	7.1M	7.2M
Gujarat	6.9M	7.3M	7.1M	7.3M	7.2M	7.1M
Haryana	7.5M	7.4M	7M	6.9M	7.4M	7.6M
Himachal Pradesh	7.4M	7.1M	7.3M	7.2M	7.2M	6.9M
Jharkhand	7.4M	7.4M	7.4M	7.4M	7.2M	7M
Karnataka	7.3M	7.2M	7.5M	7.4M	7.5M	7.1M
Kerala	7.1M	7M	7.1M	7.8M	7.3M	7.6M
Madhya Pradesh	7.2M	7.4M	7.7M	7.3M	7.2M	7.3M
Maharashtra	6.9M	7.7M	7.3M	7.6M	7.4M	7M
Manipur	7.5M	7.3M	7.5M	7.6M	6.7M	7M
Meghalaya	7.3M	7.6M	6.7M	7.6M	7.2M	7.3M
Mizoram	7.1M	7.1M	7.1M	7M	7.3M	7.3M
Nagaland	6.9M	7.2M	6.9M	7M	7.5M	7.2M
Odisha	7M	7.3M	7.3M	7.8M	7.1M	7.3M
Punjab	7.3M	7.1M	7.1M	7.1M	7.1M	7.1M
Rajasthan	7.3M	7.1M	7M	7.3M	7.4M	7.4M
Sikkim	7.5M	7.2M	7.4M	7M	7M	6.8M
Tamil Nadu	7.2M	7M	7.3M	7.3M	7.3M	7.1M
Telangana	7M	7.1M	6.9M	7.6M	7.4M	7.2M
Tripura	7.5M	7.2M	7.4M	7.1M	7.3M	7.2M
Uttar Pradesh	7.2M	7.6M	7.3M	7.1M	7.4M	7.4M
Uttarakhand	7.2M	7.5M	7.3M	7.1M	7.4M	7.4M
West Bengal	7.3M	7.1M	7M	7.1M	7.5M	6.9M

Figure 3.11: Data Used (in gigabytes) by City *and* State in India

(18x1000, SAS Viya)

This figure is a crosstab that shows states and cities in India, and the data used by each one. The data is measured in gigabytes, and across India seems to not vary much between states or cities.

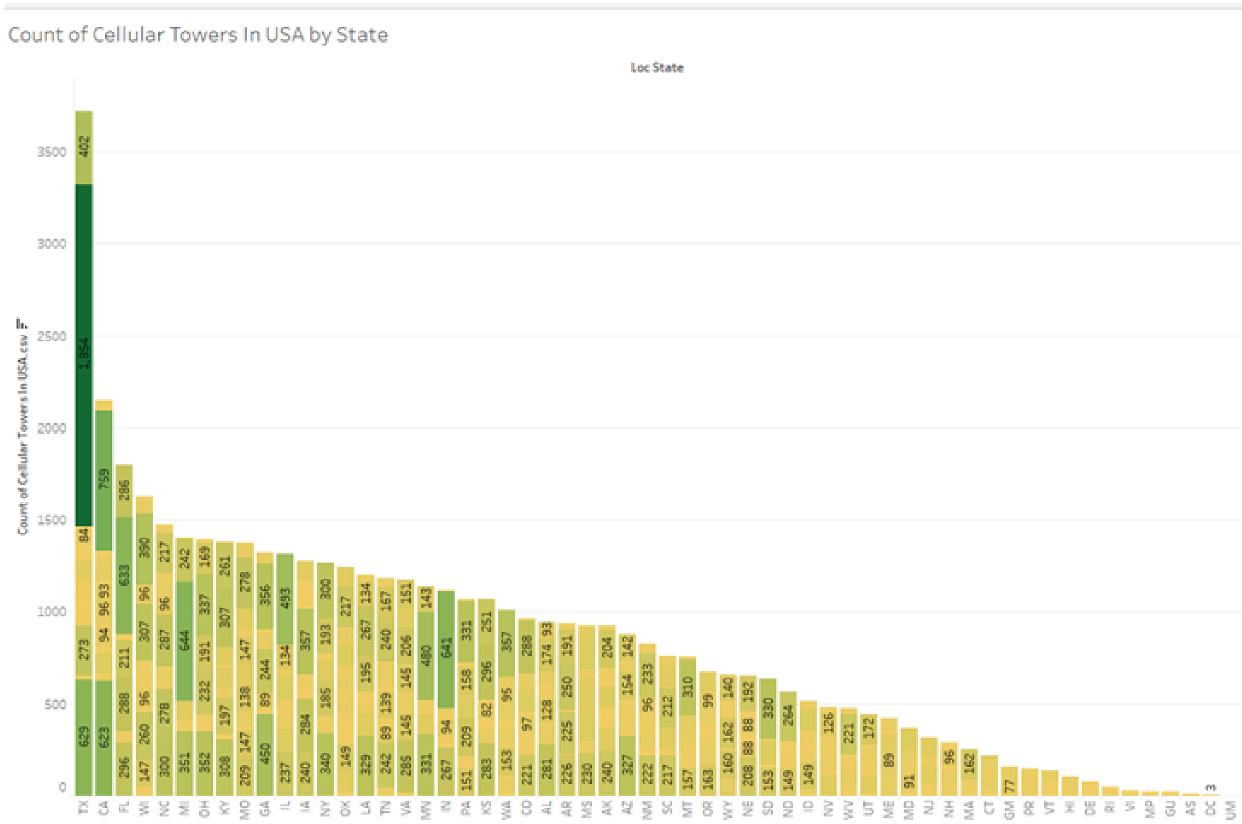
Conclusion for Telecommunication Data by Region:

The conclusions drawn by these charts are straightforward: Telecommunication markets worldwide are rapidly expanding, even in the lowest growth areas, and the data variance from region to region gives good insight into what a company in this market should do. Investors would also find this information useful for predicting which areas will turn more profit than others. High spending in one area means a company is confident in what that area will profit them. All of this will contribute directly to the amount of data being transferred, which would provide much more Big Data for use in the field. The exact nature of the data will be unclear, as this is a *general* look at telecommunication expenditure and bandwidth uses overall, not based on specifically business, academic or governmental communications. While this information *is* vague, it still gives good insight into the data in terms of ‘Big Data’ through the visualizations by showing when or where growth or expected growth in industries is happening; in this case with the Telecommunications industry.

Telecommunication Data by Provider - Romello

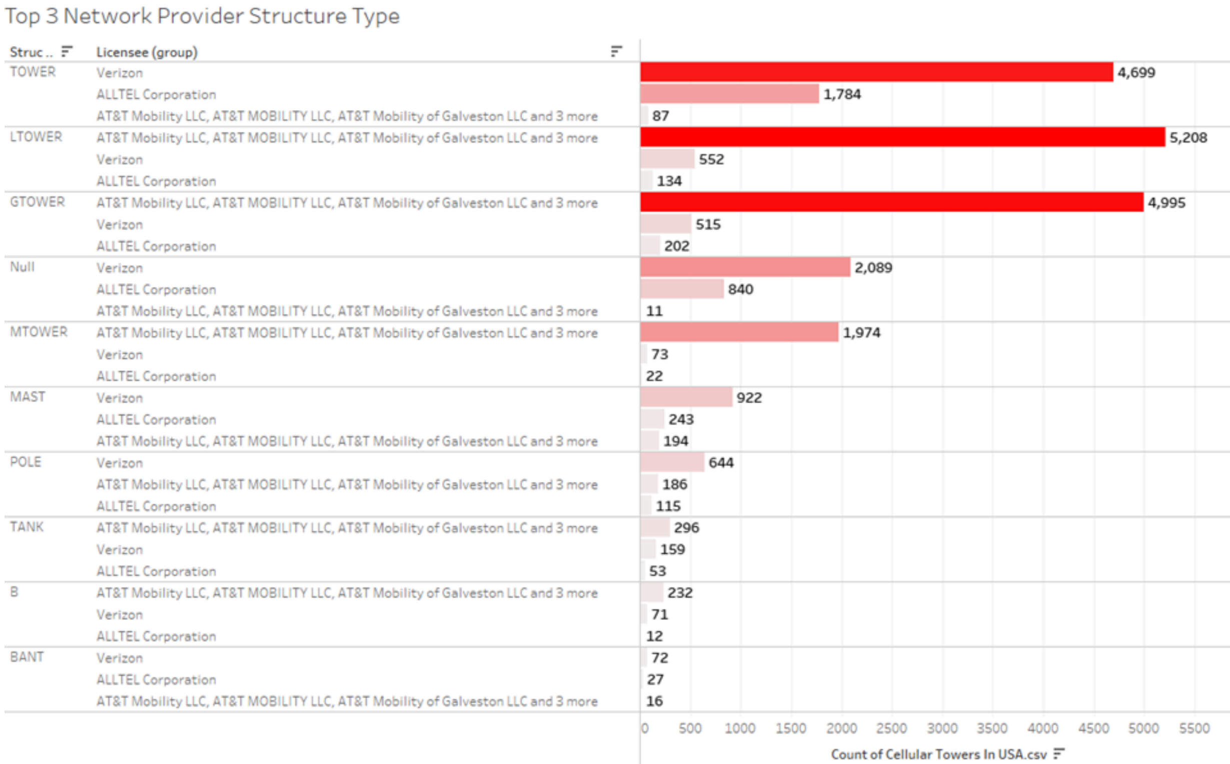
Analyzing telecommunication data by provider is of paramount importance in the telecommunications industry as it offers crucial insights into the performance, market share, and customer experience associated with different service providers. By examining data such as network usage, structure type, license, and location to each provider, telecom companies can gain a comprehensive understanding of their competitive landscape. This analysis facilitates informed decision-making processes, allowing providers to optimize network infrastructure, enhance service quality, and tailor offerings to meet customer demands effectively. Ultimately, telecommunication data by provider serves as a valuable tool for enhancing operational efficiency, driving customer retention, and fostering innovation within the industry.

Figure 4.1: Count of Cellular Towers In USA by State



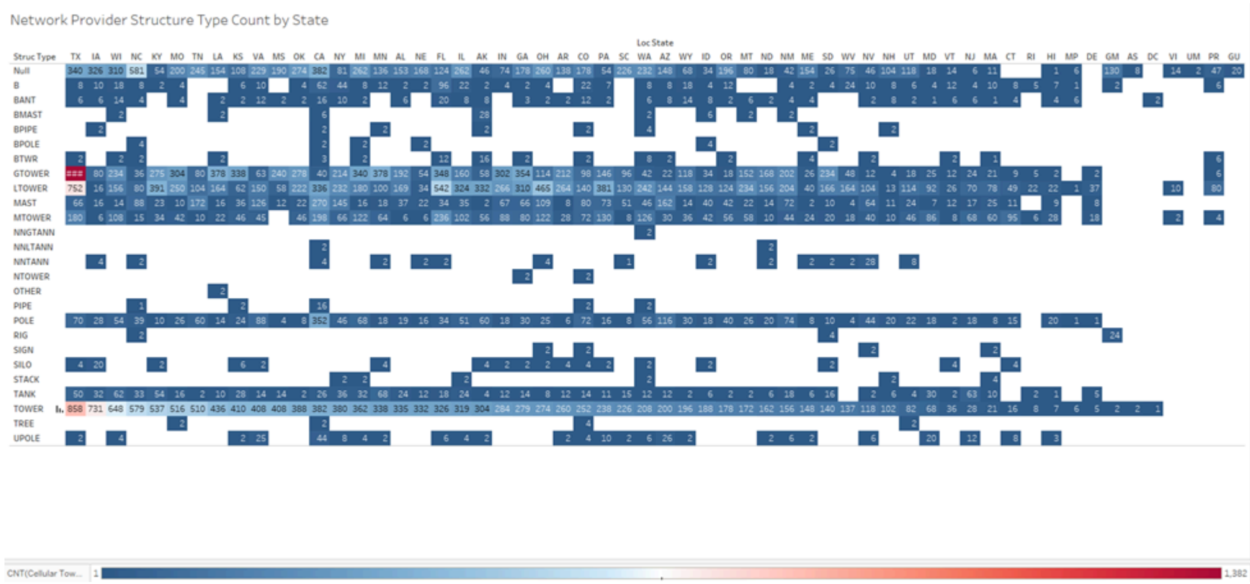
This figure, created in Tableau, illustrates the distribution of cellular towers across the USA states using a vertical bar graph. Each bar represents the total count of cellular towers within a state, with further details provided for each license group. For clarity, only the largest license groups within each state are displayed. The bars are color-coded using a yellow-green gradient to denote different network providers operating within each state, with lighter shades indicating fewer providers and darker shades indicating a greater presence. Notably, Texas stands out with the highest number of network providers, with the largest group consisting of 1,854 locations affiliated with AT&T. This observation aligns with Texas' vast size and the need for extensive telecommunication coverage across its territory.

Figure 4.2: Top 3 Network Provider Structure Type



This figure, created in Tableau, presents the distribution of cellular towers in the USA among the top three network providers, categorized by structure type, displayed in descending order on a horizontal bar graph. The colors range from white to red, with red indicating the largest group within the total count, and white representing smaller counts of structures. Cellular towers, predominantly provided by Verizon, dominate the landscape as the most utilized structure type. Verizon, a major player in the telecommunications industry, competes with other prominent providers such as AT&T and ALLTEL Corporation.

Figure 4.3: Network Provider Type Count by State

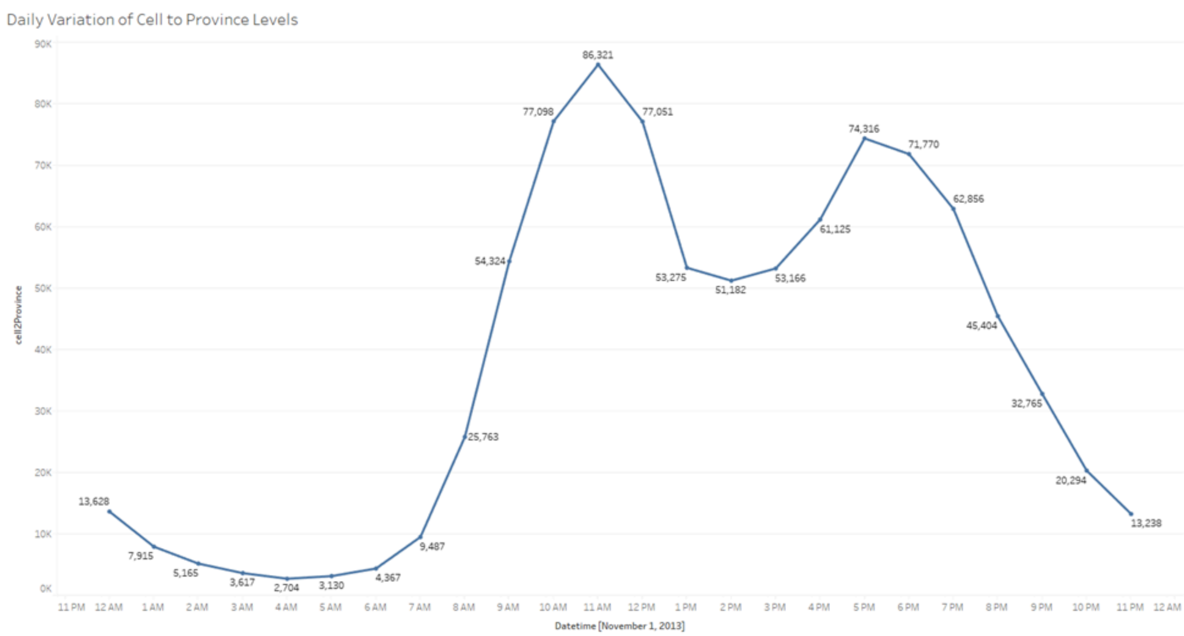


This figure, created in Tableau, shows the Network Providers Structure Type by State in a horizontal bar graph in blocks of a blue-red color spectrum with dark blue being the lowest amount and red being the highest. It's sorted in descending order of the most used structure type being TOWER. There are some blanks shown which tell some states simply don't have the structure type. It can be noted that towers are most commonly used across all states, with Texas at the most due to its large size.

Figure 4.4 Treemap Representation of Cell to Province Groups

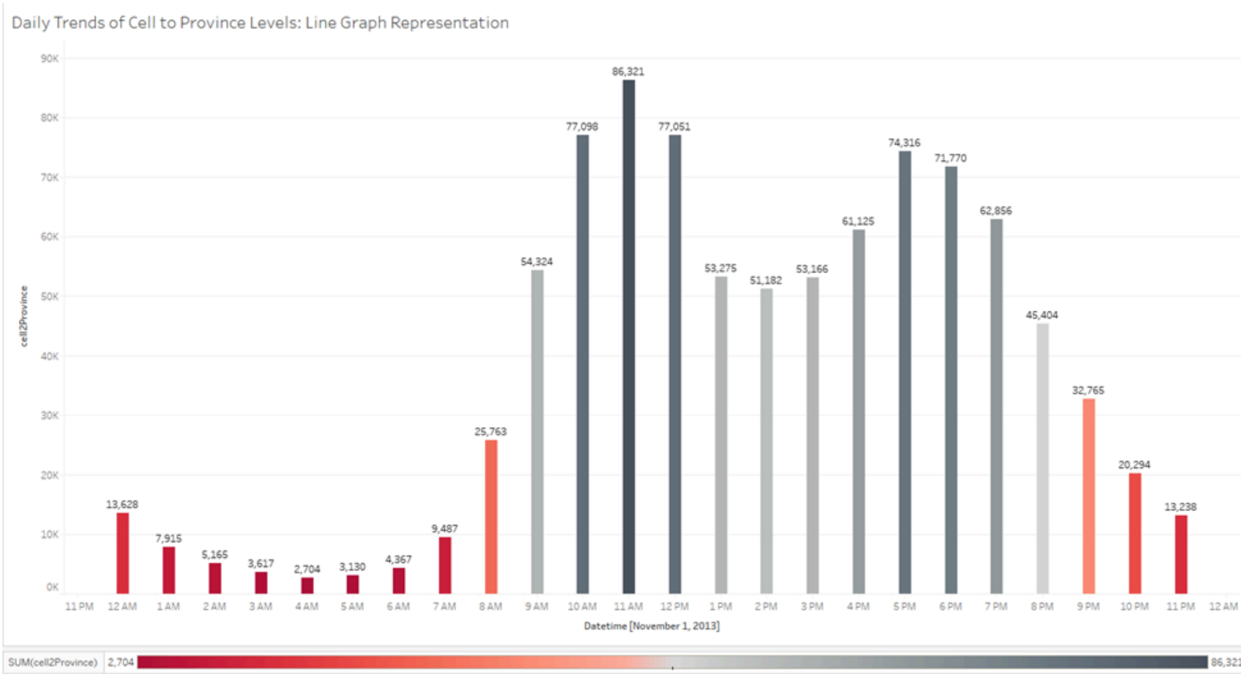
This figure, created in Tableau, illustrates the representation of the sum of cell-to-province groups, derived from a dataset emblematic of Big Data, through a Treemap visualization. The spectrum of colors used ranges from red to green, where red denotes the lowest count (e.g., 66 in Medio Campidano) and green signifies the highest (e.g., 654,074 provinces in Milano). The visualization provides insights into the distribution of cellular communication infrastructure across provincial boundaries. Provinces serve as vital administrative units, sharing political, fiscal, and economic responsibilities with central governments (UNESCO, 2024). Notably, the depicted provinces of Milano, Monza e Della Brianza, and Pavia hail from Italy, showcasing the global reach of the data. This visualization exemplifies the significance of Big Data by offering a comprehensive view of cellular infrastructure deployment at a granular, provincial level, enabling stakeholders to make informed decisions regarding telecommunications strategies, resource allocation, and regional development initiatives.

Figure 4.5 Daily Variation of Cell to Province Levels



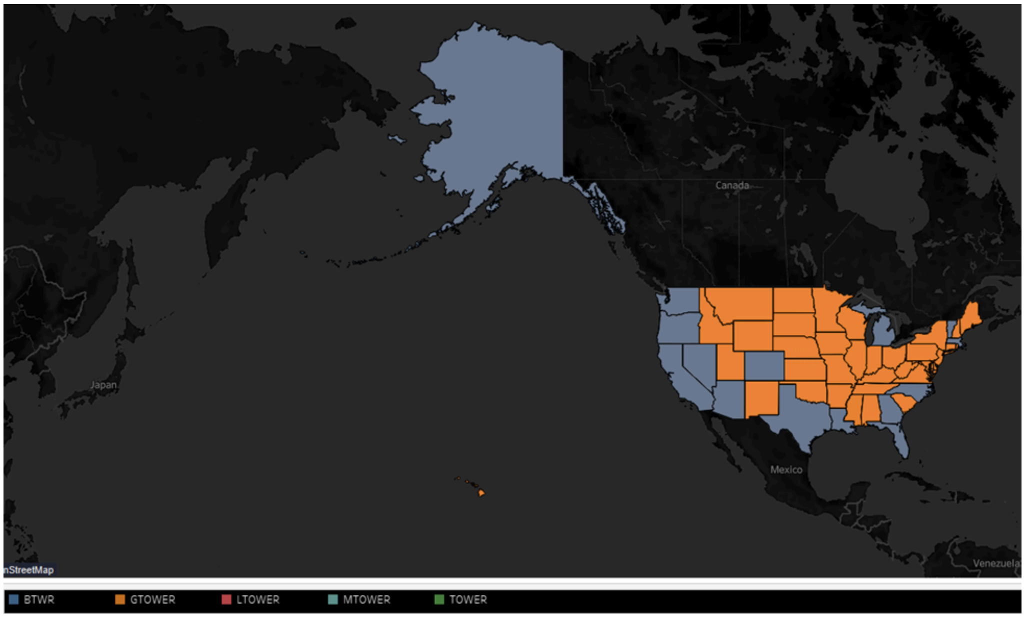
This figure, created in Tableau, illustrates the daily variation of cell to province levels, showcasing the sum of all Big Data recorded on November 1, 2013. The trend is visualized through a line graph, depicting fluctuations in data usage over the course of the day. Notably, data usage ranges from a low point of 2,704 at 4:00 AM to a peak of 86,321 at 11:00 AM. These trends hold significant importance in the realm of Big Data analytics, offering valuable insights into the usage patterns of cellular infrastructure. By understanding how data usage varies throughout the day, stakeholders can gain actionable intelligence to optimize network performance, anticipate future traffic demands, and ensure efficient management of provincial levels. In essence, this figure provides a detailed perspective on the dynamic nature of cell to province levels, enabling deeper analysis of trends and patterns crucial for informed decision-making in telecommunications and infrastructure planning.

Figure 4.6 Daily Trends of Cell to Province Levels: Line Graph Representation



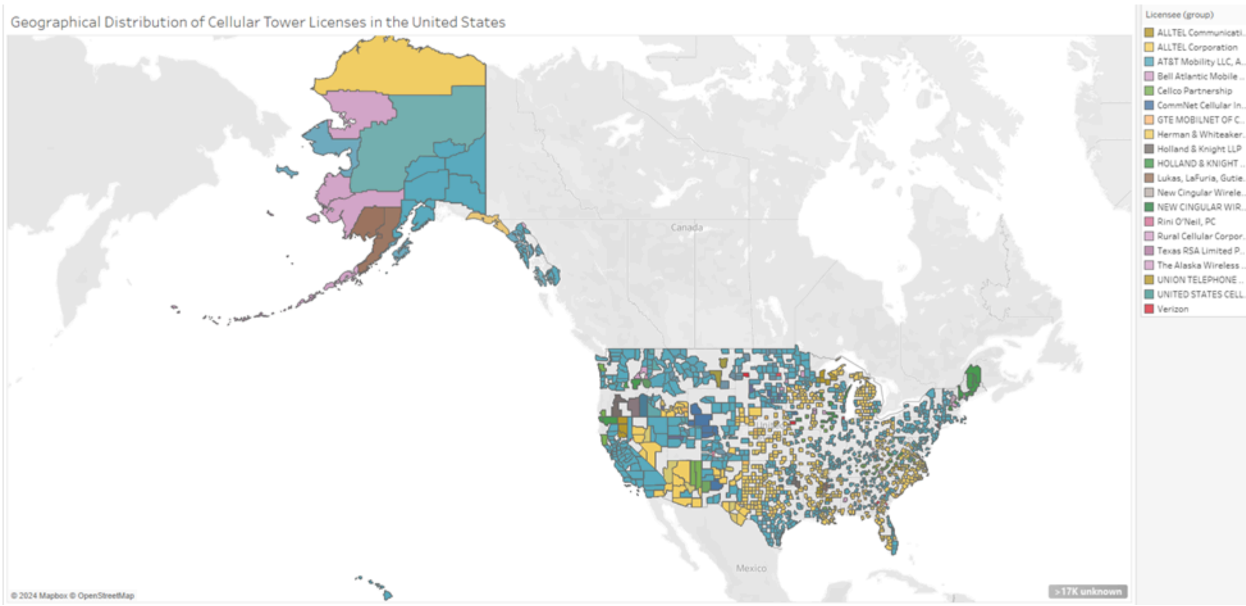
This figure, created in Tableau, portrays the daily variation of cell-to-province levels on November 1, 2013, visualized using a bar graph. The sum of all Big Data is depicted, with the color spectrum ranging from black to red, where red signifies the smallest amount (2,704 at 4:00 AM) and black represents the highest (86,321 at 11:00 AM). This color scheme enhances visibility, facilitating easier identification of peaks and troughs compared to a line graph. The trends observed in this visualization hold critical significance within the domain of Big Data analytics. By examining fluctuations in data usage throughout the day, stakeholders gain actionable insights into cellular infrastructure utilization patterns. This knowledge empowers efficient resource allocation, aids in capacity planning, and enables proactive management of provincial levels. In summary, the bar graph offers a clear and detailed representation of the dynamic nature of cell-to-province levels, allowing for in-depth analysis and informed decision-making in telecommunications and infrastructure management.

Figure 4.7 Geographical Distribution of Cellular Towers in the USA and Structure Type



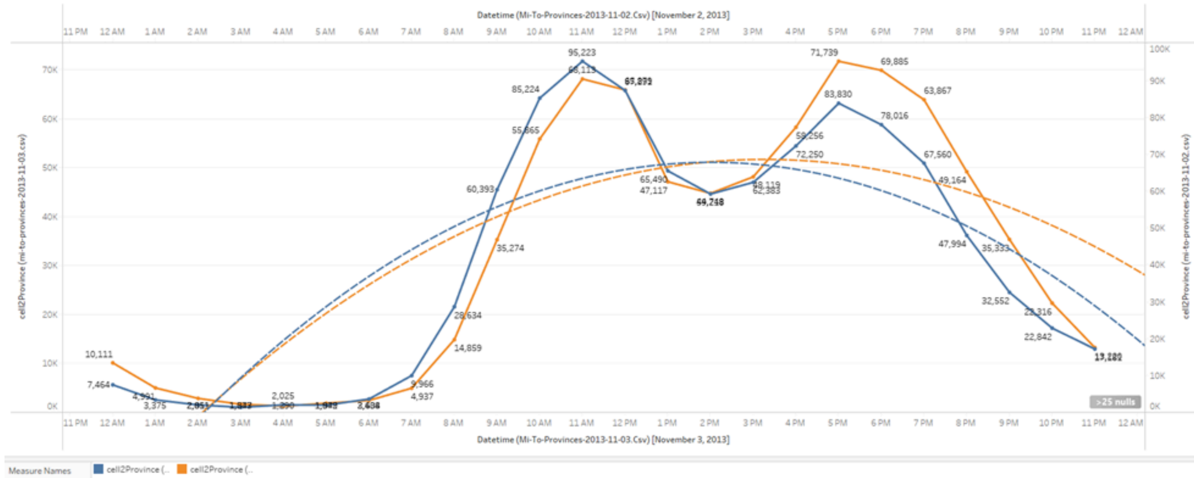
This figure, created in Tableau, depicts the geographical distribution of cellular towers across the United States, organized by state and structure type. This filled map-based visualization with latitude and longitude as axes, employing colors to represent various types of cellular tower structures. Providing insights into spatial distribution and prevalence, this visualization is crucial for telecommunications companies, policymakers, and urban planners in making informed decisions regarding network coverage, capacity planning, and resource allocation. With data spanning all 50 states and representing multiple tower types, this figure qualifies as Big Data, requiring sophisticated computational tools for analysis.The tower types depicted serve different purposes, indicative of varying network requirements. BTWRs are prevalent in rural areas, offering basic cellular coverage, while GTOWERs are commonly found in urban centers and along major highways to meet higher network capacity demands. Additional tower types like LTOWER, MTOWER, and TOWER address specific technical or environmental considerations. This comprehensive depiction enables an understanding of network deployment strategies, essential for optimizing telecommunications operations and ensuring reliable service delivery nationwide.

Figure 4.8 Geographical Distribution of Cellular Tower Licenses in the United States



The figure, created in Tableau, presents a filled map of the United States, with each county delineated and color-coded based on the dominant license group providing coverage in that area. The map is generated based on longitude and latitude coordinates. The color encoding represents details about the licensee group, with each group assigned a distinct color. Notably, the majority of the map appears to be shaded in blue, indicating that AT&T is the predominant licensee group across many counties. The figure provides detailed insights into the geographical distribution of cellular tower coverage and highlights the dominance of certain license groups in specific regions. This visualization is crucial for understanding the landscape of telecommunication infrastructure across the United States and identifying areas where particular licensees have significant presence and influence. Furthermore, this figure qualifies as Big Data due to the vast amount of geographical data being processed and analyzed. It incorporates data from numerous counties, each with its own set of coordinates, and maps out the coverage areas of multiple license groups. The sheer volume and complexity of this geographical data make it a quintessential example of Big Data analytics in the telecommunications industry.

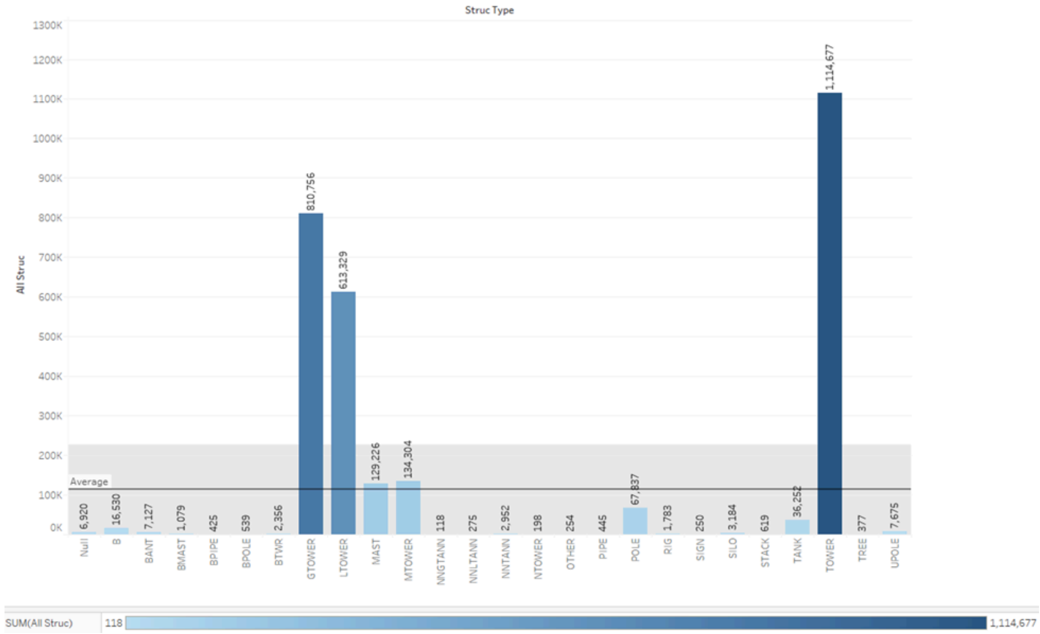
Figure 4.9 Analyzing Daily Variations in Cell to Province Levels: Insights from November 2nd and 3rd, 2013



This dual-axis figure, created in Tableau, presents the daily variation of cell to province levels on November 2nd and November 3rd, 2013, specifically focusing on the data provided by the files mi-to-provinces-2013-11-03.csv and mi-to-provinces-2013-11-02.csv. The visualization utilizes line graphs to display trends in cell to province levels over time, with each line representing a different province.The exponential trend line is to capture and visualize the growth or decline patterns in cell to province levels, aiding in trend analysis and anomaly detection.

Understanding the daily variation of cell to province levels is important for telecom companies and infrastructure planners. It provides insights into usage patterns, peak times, and potential areas for optimization. By analyzing these trends, companies can allocate resources efficiently, improve network performance, and enhance service quality for customers.This figure qualifies as big data due to the sheer volume and complexity of the dataset being analyzed. It encompasses data from multiple provinces over two days, resulting in a large dataset with numerous variables. The analysis involves processing and interpreting vast amounts of information to extract meaningful insights, which aligns with the characteristics of big data analytics. Additionally, the inclusion of historical data and real-time observations further contributes to the complexity and scale of the analysis, reinforcing its classification as big data.

4.10 Distribution of Cellular Tower Structure Types in the USA



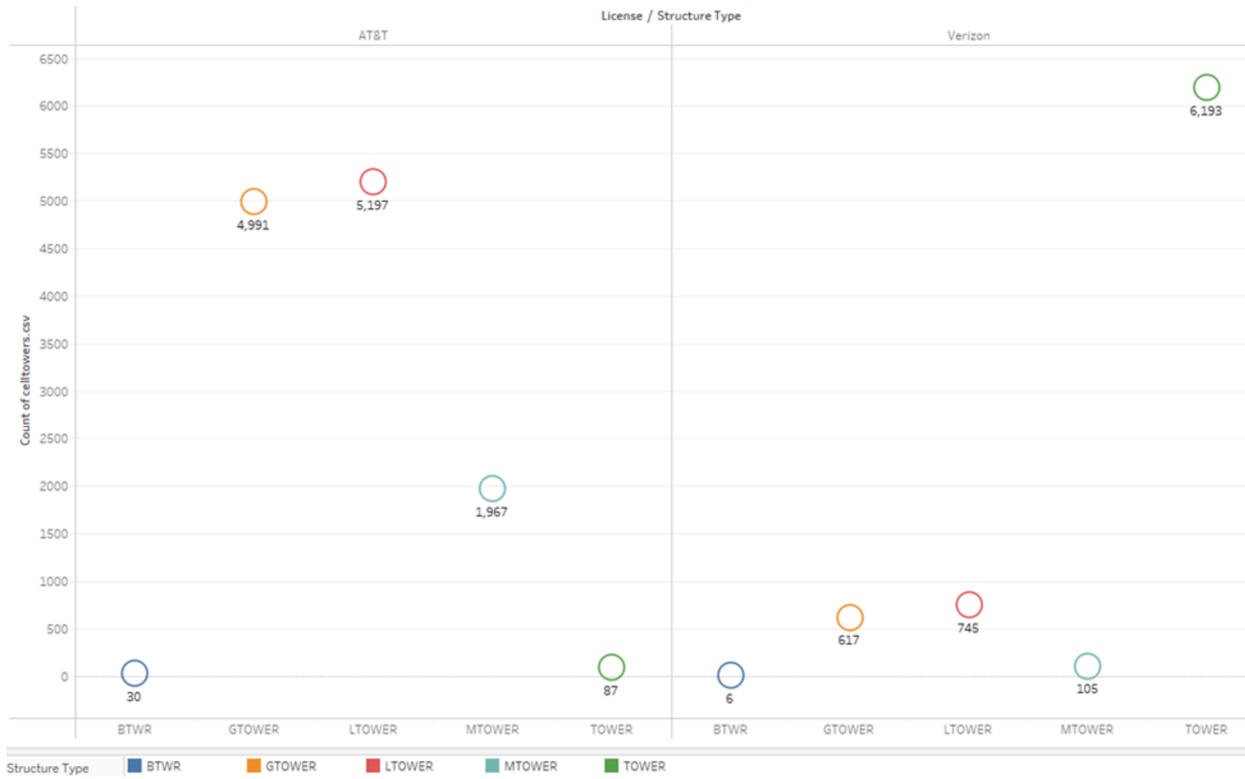
This figure, created in Tableau, visually represents the diversity of structure types utilized for cellular towers across the USA through a bar graph. Each bar corresponds to a specific structure type, with the height of the bar indicating the total count of structures of that type. The color gradient across the bars signifies the varying sums of structures for each type, offering a quick insight into the distribution patterns. This analysis is crucial for telecommunications companies to optimize network coverage and performance efficiently. By understanding which structure types are prevalent in different regions, companies can make informed decisions regarding infrastructure deployment and investment. This figure qualifies as Big Data analytics due to the extensive dataset involved, covering 26 structure types and a wide range of sums, highlighting the complexity and scale of information being processed.

4.11 Provincial Distribution of Province to cell

This figure, created in Tableau, displays a visualization of Province Name data. Each square mark represents a province, with the color and size of the square indicating the sum of Province to Cell. The marks are labeled by Province Name. This provides a view of the distribution and magnitude of Province to Cell across various provinces.

Understanding the distribution of Province to Cell data is crucial for optimizing telecommunication networks at a provincial level. This figure, showcasing Provinceto cell values across 110 provinces, allows stakeholders to identify regions with varying levels of telecommunication infrastructure. Qualifying as Big Data, the dataset spans multiple CSV files and encompasses a wide range of values, from 73 to 656,782. The visualization encompasses 110 provinces, including MILANO and MONZA E DELLA BRIANZA, which are likely to have higher populations compared to other provinces. The significant Province to Cell values associated with these provinces indicate a higher concentration of telecommunication infrastructure and usage, reflecting the importance of these regions in the telecommunications landscape.

Figure 4.12 Cellular Tower Distribution Analysis: Structure Type and License Breakdown



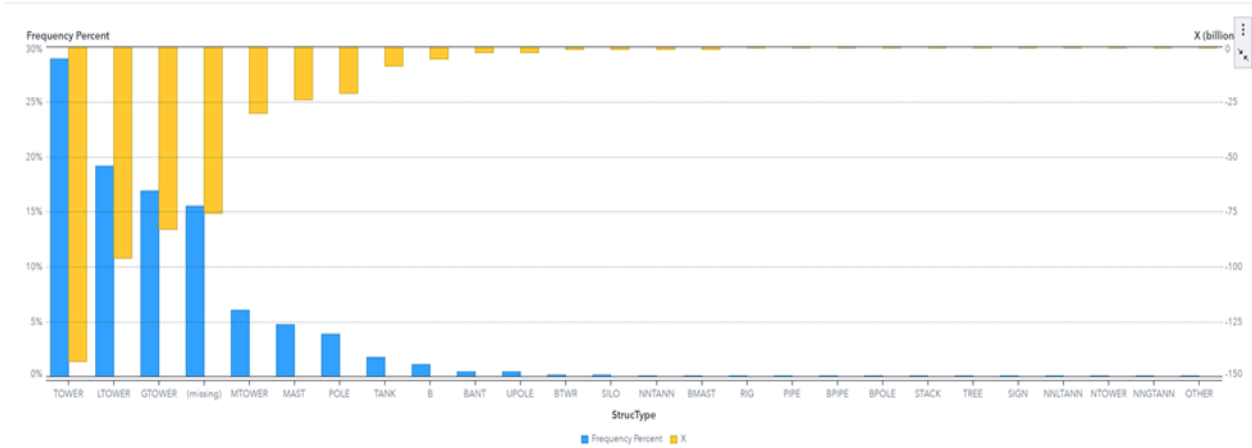
This figure, created in Tableau, offers a comprehensive breakdown of cellular tower distribution in the USA, segmented by Structure Type and License. By visualizing the count of towers for each Structure Type, distinguished by color and shape, the figure provides crucial insights into the coverage patterns of the two primary license groups, AT&T and Verizon. This analysis enables telecom stakeholders to assess network coverage comprehensively, identify areas of strength and improvement, and strategize infrastructure deployment effectively. The figure's ability to handle and present data from multiple sources underscores its qualification as Big Data, emphasizing its capacity to process vast amounts of information and derive actionable insights essential for optimizing telecommunications operations in a data-rich environment.

Figure 4.13 World Cloud of Cellular Towers License

Licensee	Frequency ▾
Verizon Wireless	7,894
AT&T Mobility LLC	6,328
AT&T Mobility Spectrum LLC	4,089
ALLTEL Corporation	3,464
Cellco Partnership	3,305

The presented visualization, created in SAS Viya, shows the landscape of cellular tower licensing across the USA through a word cloud representation. Each word in the cloud corresponds to a licensed company, with the size of the word indicating the magnitude of licensing activity. Dominating the visual are industry giants such as Verizon Wireless, AT&T Mobile, ALLTEL Corporation, and Cellco Partnership, whose prominence underscores their substantial presence within the telecommunications infrastructure. This illustration serves as a testament to the huge scale and complexity inherent in the world of big data, as it aggregates and filters extensive licensing data into an easily understandable and visually compelling format.

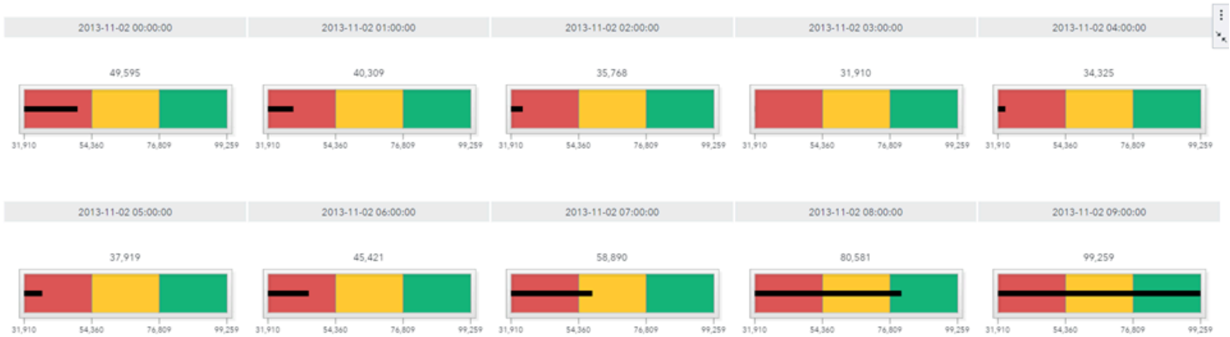
Figure 4.14 Frequency Percent of Cellular Towers Structure Types



StrucType	Frequency Percent	X
TOWER	28.96%	-1.428732E11
LTOWER	19.20%	-95789676055
GTOWER	16.90%	-82761606204
	15.53%	-75553042413
MTOWER	6.07%	-30204544686

Figure 4.14 shows a chart, created in SAS Viya, of different types of cell towers in the USA. Each type is shown as a bar, and the height of the bar represents how many towers there are. The most common type is "TOWER," which is used a lot across the country. Interestingly, the third most common type is "POLE." Even though it's less common, it's still important. This chart helps us understand the variety of cell towers across different areas of the country, which is a big part of managing our cell networks efficiently.

Figure 4.15 Gauge Chart of Call Frequency 11-02-2013



This Gauge chart, created in SAS Viya, maps out the ups and downs of call activity across different provinces in Italy on November 2nd, 2013. It zooms in on the data from the file mi-to-provinces-2013-11-02.csv, focusing on the first ten hours of the day. You can see a clear trend of more calls being made as the morning turns into afternoon, giving us a snapshot of how people use their phones throughout the day. This figure exemplifies the power of big data analysis, helping us uncover patterns in large datasets

Conclusion for Telecommunications Data by Provider

In reviewing the telecommunications data and visualizations presented, it becomes apparent that the analysis extends beyond mere provider statistics to encompass the broader realm of Big Data. The datasets provided, including IS TAT_census_variables_2017 and mi-to-provinces files spanning multiple days, with outside resources of data sets of cellular towers which both represent a substantial volume of data indicative of Big Data analytics. These datasets serve as the foundation for understanding the intricate dynamics of the telecommunications industry, including network usage patterns, infrastructure deployment, and geographical distribution.

By leveraging Big Data analytics, telecom companies gain unparalleled insights into industry trends, customer behavior, and operational efficiency. The visualizations provided, such as the count of cellular towers by state and the distribution of network provider types, offer a glimpse into the vast landscape shaped by the convergence of telecommunications and data analytics.

Furthermore, the inclusion of Big Data in our analysis underscores the importance of harnessing large-scale datasets to drive strategic decision-making and innovation within the telecommunications sector. Through comprehensive analysis of Big Data, companies can optimize network performance, enhance service offerings, and foster customer satisfaction.

Within the third phase, the following figures 4.13, 4.14, and 4.15 were all constructed using SAS Viya. Using SAS Viya has been a new experience since it was my first time engaging on the application. In comparison with Tableau, it’s more complex in visualizing data and with holding data. It’s constructed to hold huge amounts of information within a chart, so that an element of a chart can have its own visualization as well. The complexity of SAS Viya provides scalable, in-memory processing capabilities, advanced analytics, to empower organizations to derive actionable insights and make data-driven decisions..

In conclusion, the utilization of Big Data in telecom analytics represents a paradigm shift in how industry stakeholders approach business intelligence and decision-making. By embracing the power of Big Data, telecom companies can stay ahead of the curve, delivering cutting-edge services and driving sustainable growth in an increasingly data-driven world.

Works Cited

“Telecommunication Definition, Types, Application, Future, & Faqs.” *Electronics For You*, 5 Jan. 2023, www.electronicsforu.com/technology-trends/learn-electronics/telecommunication-basics.

Regional/state/provincial level. UNESCO UIS. (2023, September 12). https://uis.unesco.org/node/3080263

“Telecommunication Definition, Types, Application, Future, & Faqs.” *Electronics For You*, 5 Jan. 2023, www.electronicsforu.com/technology-trends/learn-electronics/telecommunication-basics.

Lightico. (2023, January 4). *A guide to reducing churn in the telecom industry*. Vizolution. https://www.vizolution.com/blog/complete-guide-to-reduce-churn-in-telecom/#:~:text=In%20terms%20of%20telecoms%2C%20the,to%20your%20service%20is%20low.