1. **ANALYSING NETWORK TRENDS IN INDIA.**
2. **MEMBERS**

* 20BDA01 – TASHI CHOTSO
* 20BDA09 – LEKSHMI V AJI
* 20BDA42 – M J AISHWARYA
* 20BDA53 – JAYASREE C

1. **INTRODUCTION**

The world has advanced rapidly in the past decade, every bit of technology that is introduced in the world has been adapted by people for their convenience. One of the most game changing trends that technology has bought about is connecting people from across region and around the world. Networks and network providers play a significant role in the same. Different networks come with varying features but with the same purpose, to connect and provide steady network irrespective of the region.

With Network and Network providers come an imminent procedure of people choosing which network they can best use, when in a particular region. Network providers vary in terms of signal strength, download speed, data rate per mbps speed and so forth.

We have come up with few visualization techniques which gives insights into which Network works best according to signal strength, data speed and etcetera region wise in order to give a structure analysis and understanding of how it works region to region across India and also, we derive the comparisons.

1. **PROBLEM STATEMENT:**

ANALYSING THE NETWORK TRENDS WITH RESPECT TO DATA SPEED, SIGNAL STRENGTH ACROSS INDIA FOR THE MONTHS OF JANUARY 2020 – SEPTEMEBER 2020.

1. **METHODOLOGY USED.**

The test sample consisted of data collected from user’s devices through the application “MySpeed” from the Telecom Regulatory Authority of India (TRAI). The main aim of this application is to get a real-time understanding of consumer’s experience across India through an interactive medium.

The data collected was collected for every month from January 2020 to September 2020 and a total of 34,48,203 observations were chosen for analysis after the initial cleaning

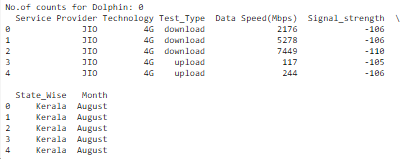
The data cleaning process involved various steps which were performed using packages in Python(Pandas)and RStudio(dplyr and tidyverse).The data was checked for its validity**,**[accuracy, uniformity  and consistency](https://towardsdatascience.com/the-ultimate-guide-to-data-cleaning-3969843991d4) by examining the degree to which the data is close to the true values and by looking for any data-type constraints ,range constraints and mandatory constraints. The initial dataset had recorded several NA under the Data Speed, Signal Strength and the Service Area columns. The NA values in the Signal Strength column corresponds to the device not being able to capture signal and the NA values in Service Area corresponds to the area being unidentifiable. To correct these anomalies the NA values where removed since the data set was huge the removal did not disrupt the accuracy of the result as seen during the univariate analysis. The dataset was also checked for any duplicate values which were removed after checking for any inconsistencies.

The data analysis process of the project was done in different stages in order to get a clear understanding of each variable:





OUTPUT:



1. **Exploratory data analysis:**

At first Univariate analysis was performed on each of the 6 variables and the target variables were decided. The variables were assigned for correct data type and appropriate column names. Box-plots provided a clear information on the outliers in the Data Speed column, these were not removed since the composed of a majority of information on the Service Provider and Area of Service. The distribution was checked for the numerical variables and low variance filter was applied. Time variables (Months) were also checked.

The next step was bivariate analysis here the input and the output variables were compared for correlation using Scatter Plots and pairwise visualizations were studied. The subset of the dataset for each month was analysed for any trends. For accuracy a new column containing the different months were added.

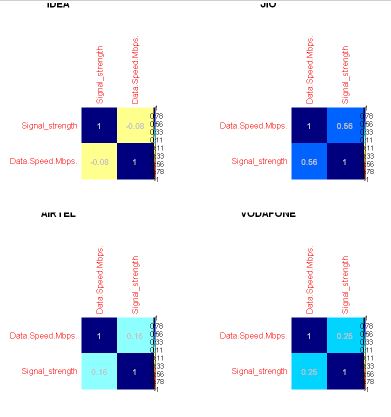
Multi-variate analysis was performing using a set of categorical and numerical variable to further understanding of the data.

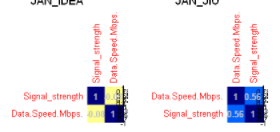
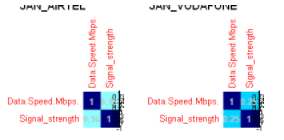
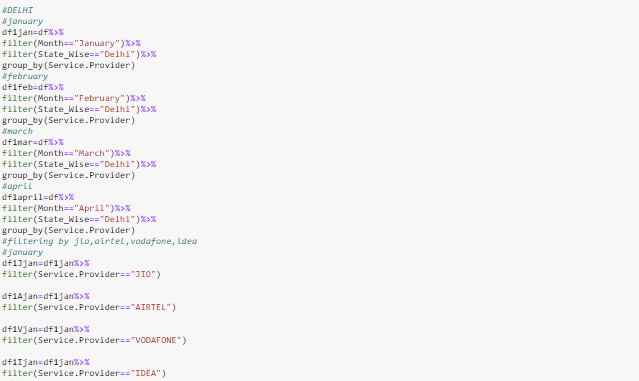
Standardization of the data was done –

For strings, make sure all values are either in lower or upper case.

For numerical values, make sure all values have a certain measurement unit.

Scaling of the Signal Strength values in decibels was checked to fit within a specific scale for classification.



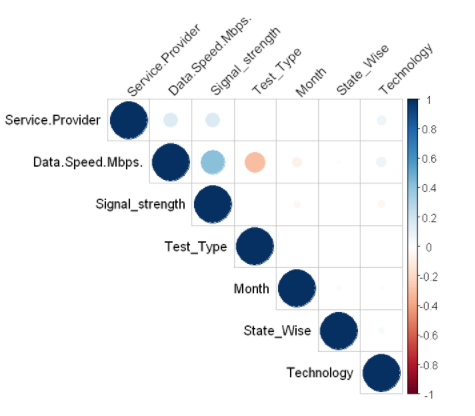


**b. Statistical Methods:**

By analysing and visualizing the data using statistical methods such as mean, standard deviation, range, or quantiles, values that are unexpected and thus erroneous were found.

For the modelling the two options explored were classification or regression. Classification could fit the data due to the non-linearity of the data points of the target variables.

**Correlation**: We have found the correlation between data speed and signal strength and as a result we have correlation coefficient of 0.86, as the value of the coefficient is higher, we conclude that there is a good correlation between these two variables across the states.



**6.DATASET**

* How we worked on the dataset?

The test sample consisted of data collected from user’s devices through the application “MySpeed” from the Telecom Regulatory Authority of India (TRAI).

The dataset contains 3,44,203 rows and 7 columns namely

* Service Provider
* Technology
* Test\_Type
* Data.Speed.Mbps
* Signal\_Strength
* State\_Wise
* Month

With the help of this data set we are able to differentiate and segregate that how each network works with respect to region. We see State wise comparison with respect to the data speed and the kind of technology i.e., 3G,4G technology and also how it changes month wise, if there is a difference in the range provided from the previous month or has remained constant. We make use of this dataset for further visualisation in understanding the deeper aspects of how it has succeeded in a particular region. We also use the year wise comparison to see the same.

With the help of these visualisation we want to provide a better perspective for the user with respect to the technology, data speed, signal strength with respect to the region.

**7.RESULT**

* What have we achieved?

We have understood how the network speed, signal strength, data speed differs from region to region.

We have analysed and gained insights from

Frequency distribution – Pie chart for service providers,

Multivariate analysis – Line Graph, Bar Plot, Pie Graph.

Map – for plotting region wise information and to give a complete geographic region.

Insights:

Data Speed.

We understand that Madhya Pradesh, Kolkata and Orissa has good data speed, Jammu and Kashmir, Northeast has comparatively lesser data speed.

January, February, September has relatively good data speed.

April had the weakest data speed, followed May.

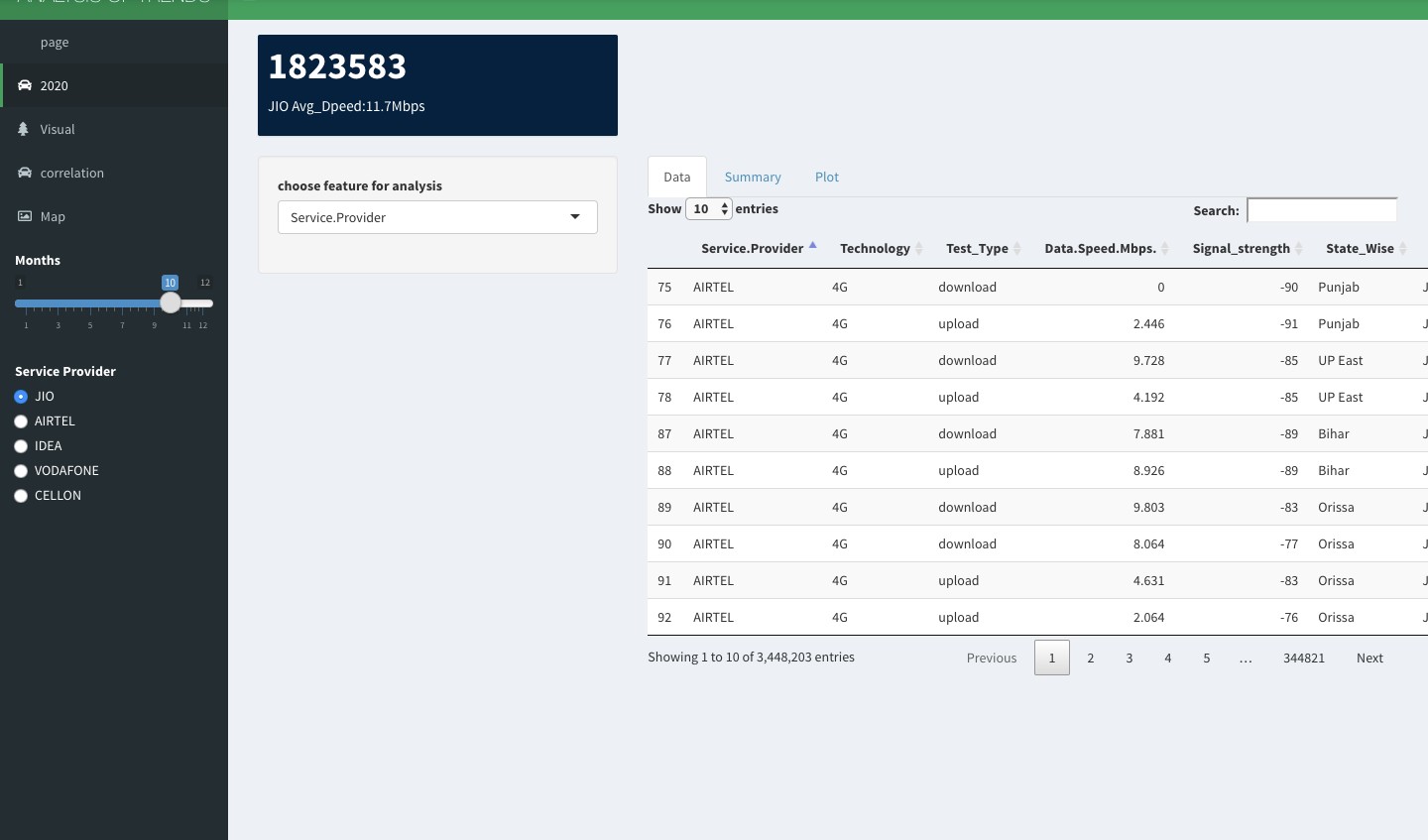
Signal Strength: Madhya Pradesh, Delhi and Orissa have better Signal Strength

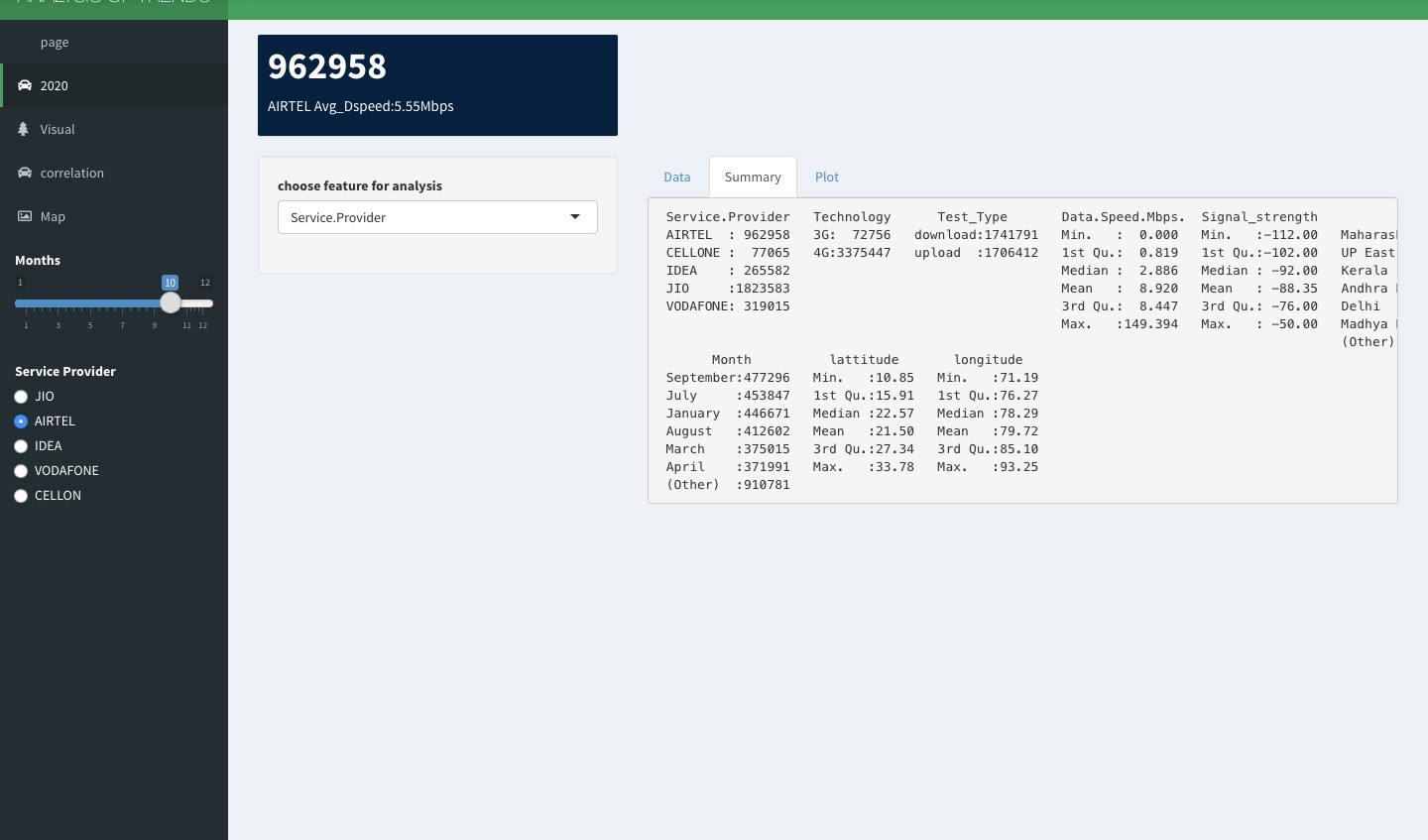
Weak Signal Strength: Tamil Nadu, Kerala and Northeast.

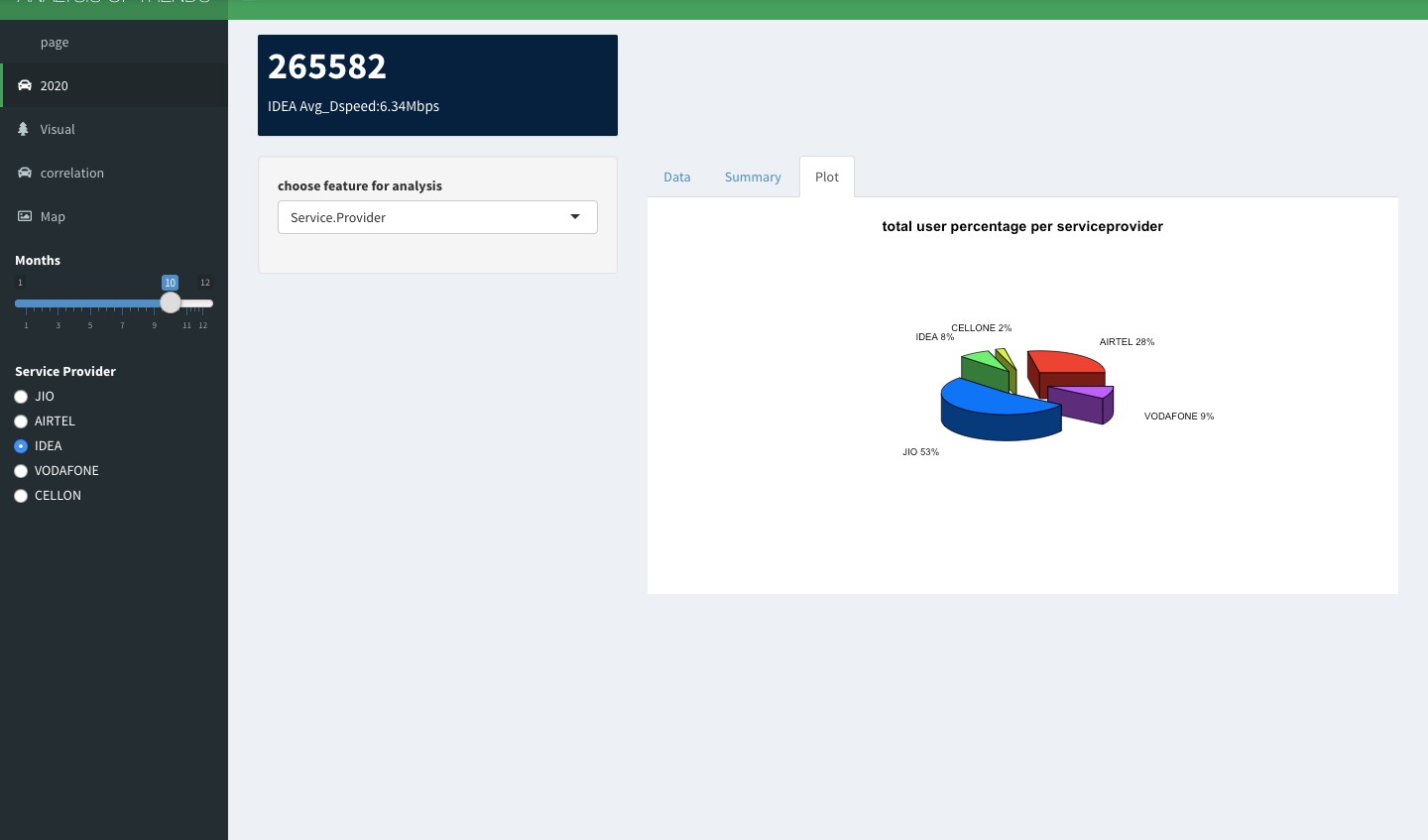
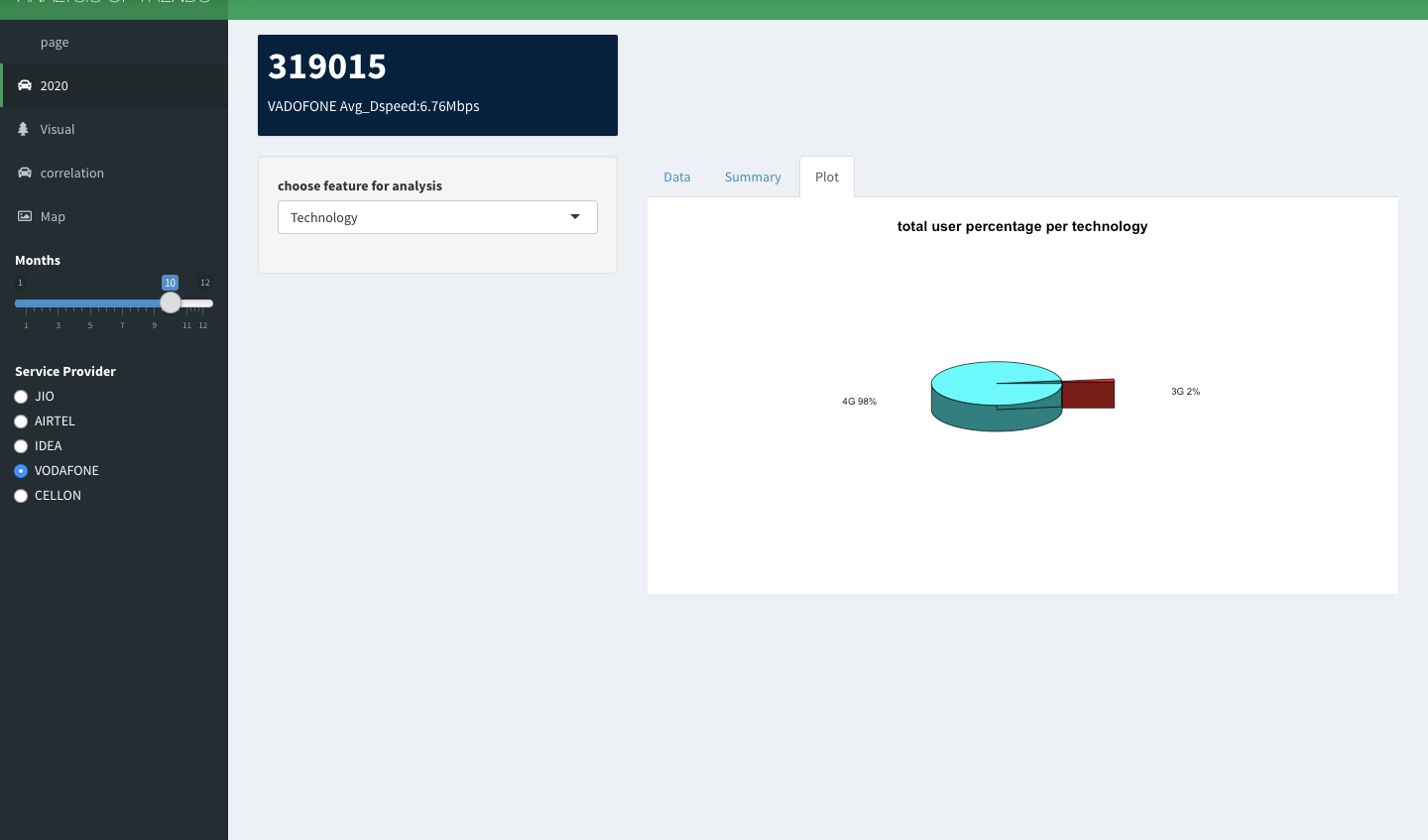
January to September, overall signal strength has been good, April and May were relatively low.

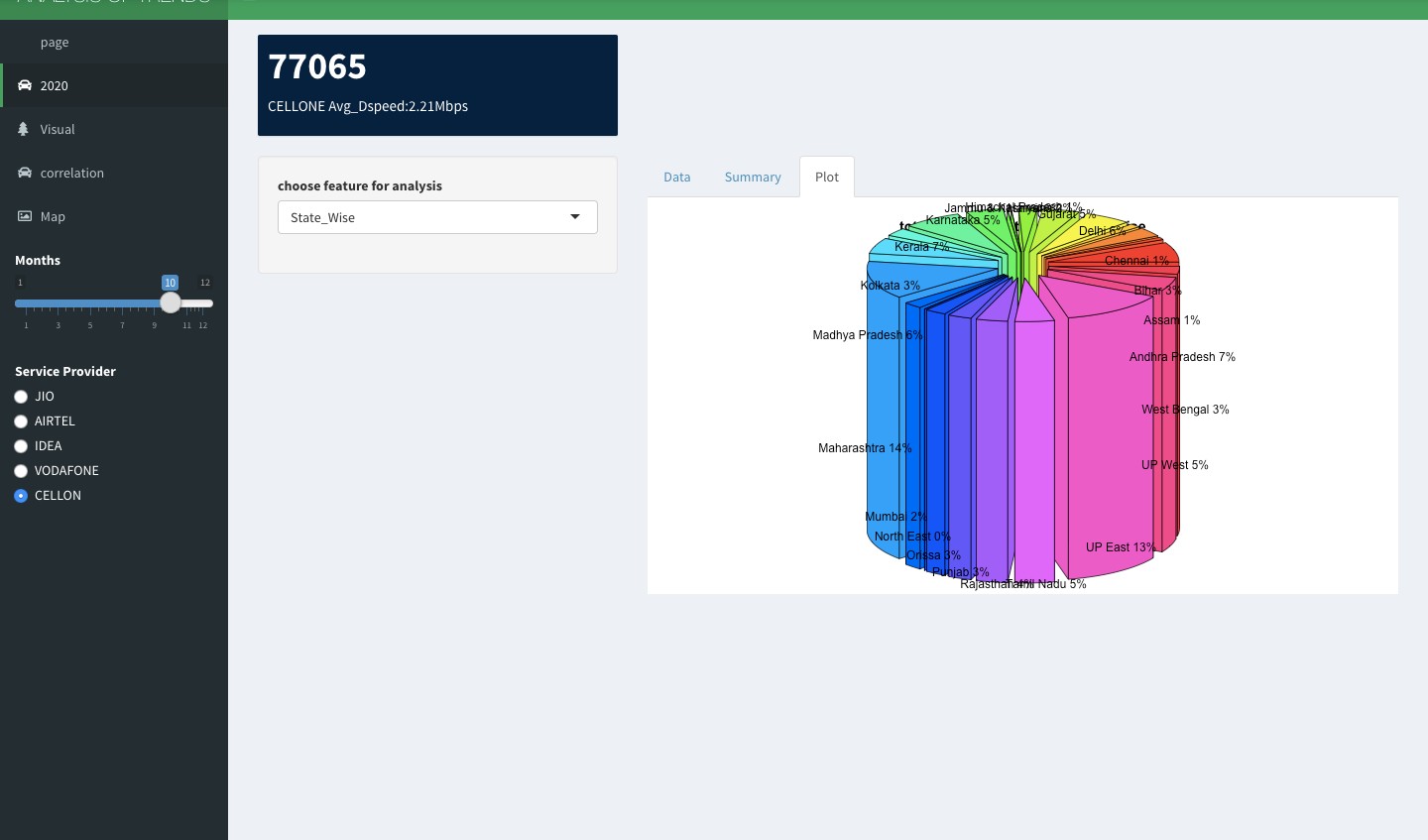
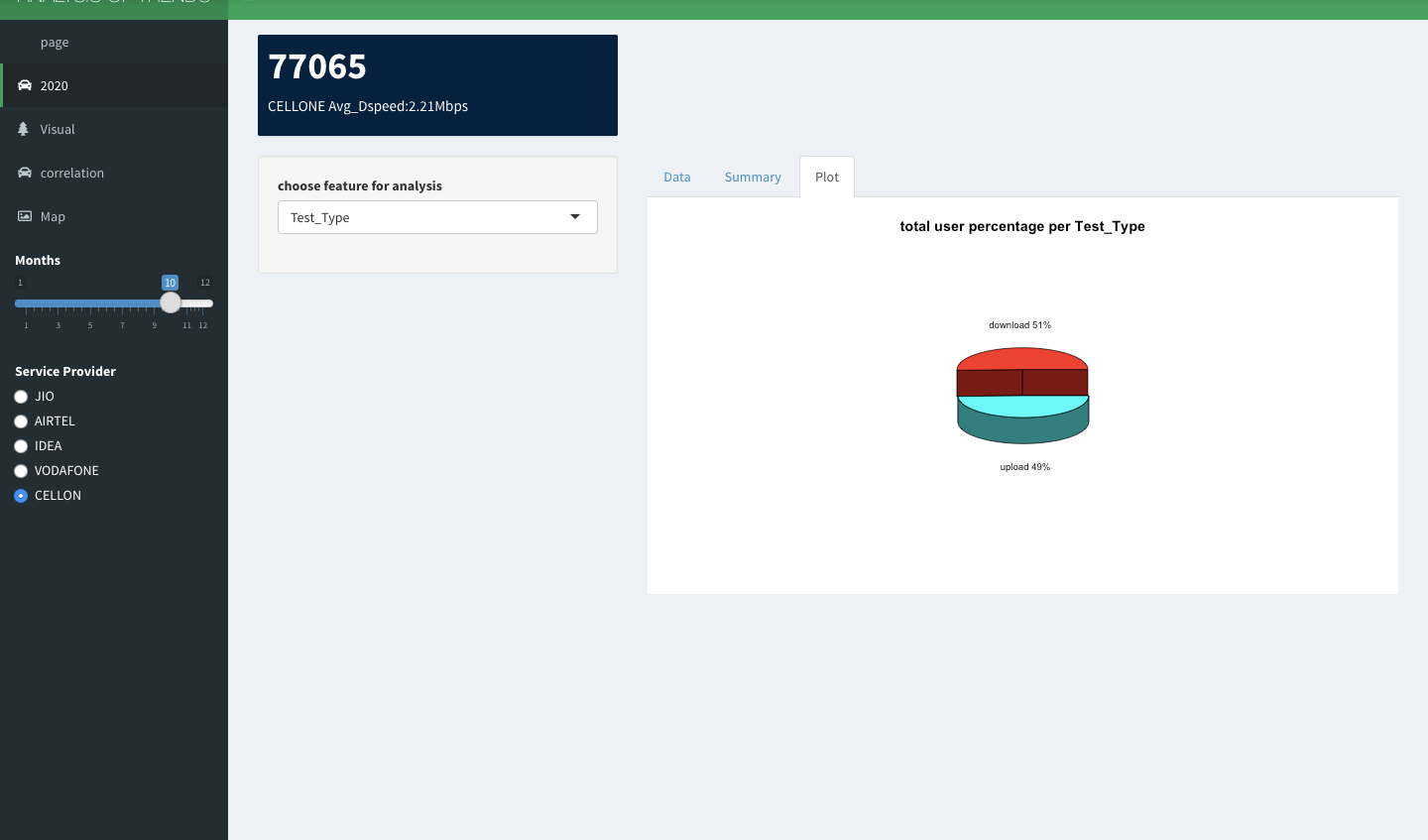
Factors that affect the weak nature: Weather conditions, Less Service providers, or other service providers that we have not considered.

From these visualisations we also draw an analysis that Jio (1,823,583) has the highest number of users followed by Airtel (962,958), Vodafone (319,015), Idea (265,582), CellOne (77,065).

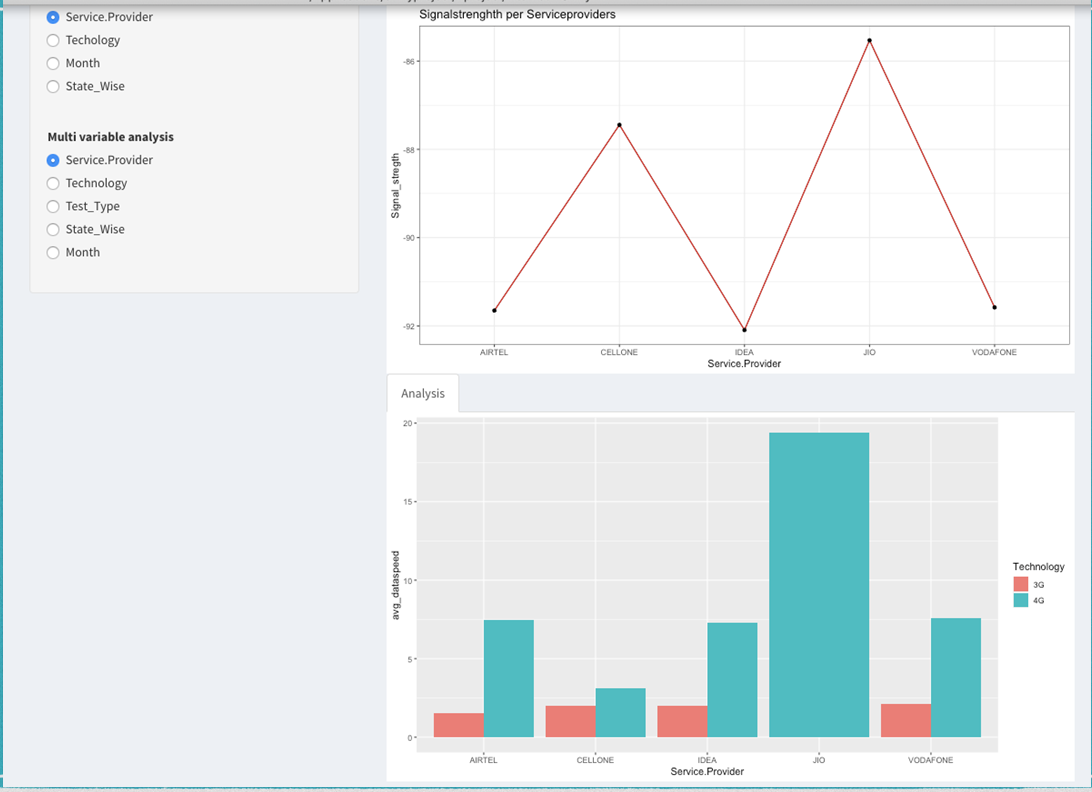
****

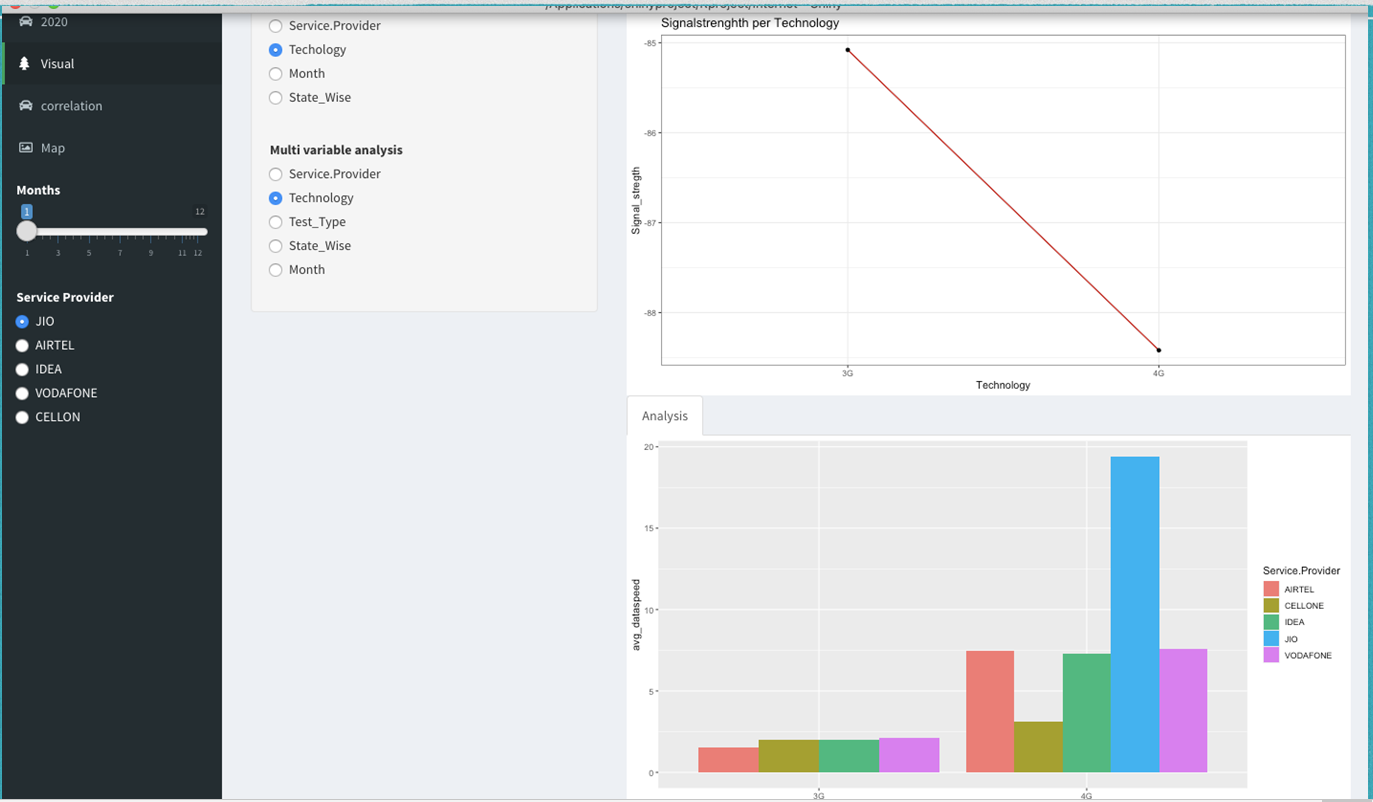
****

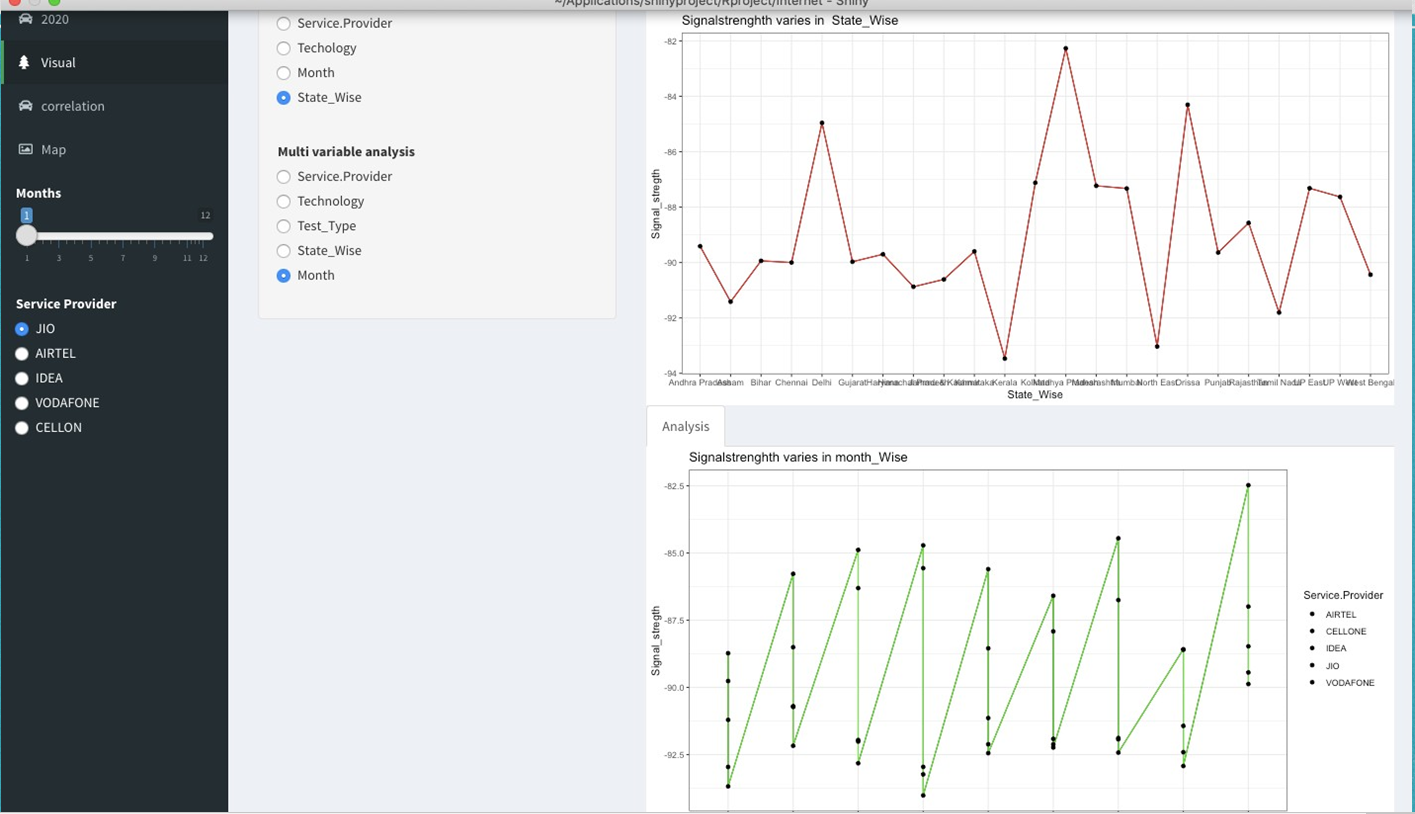
****

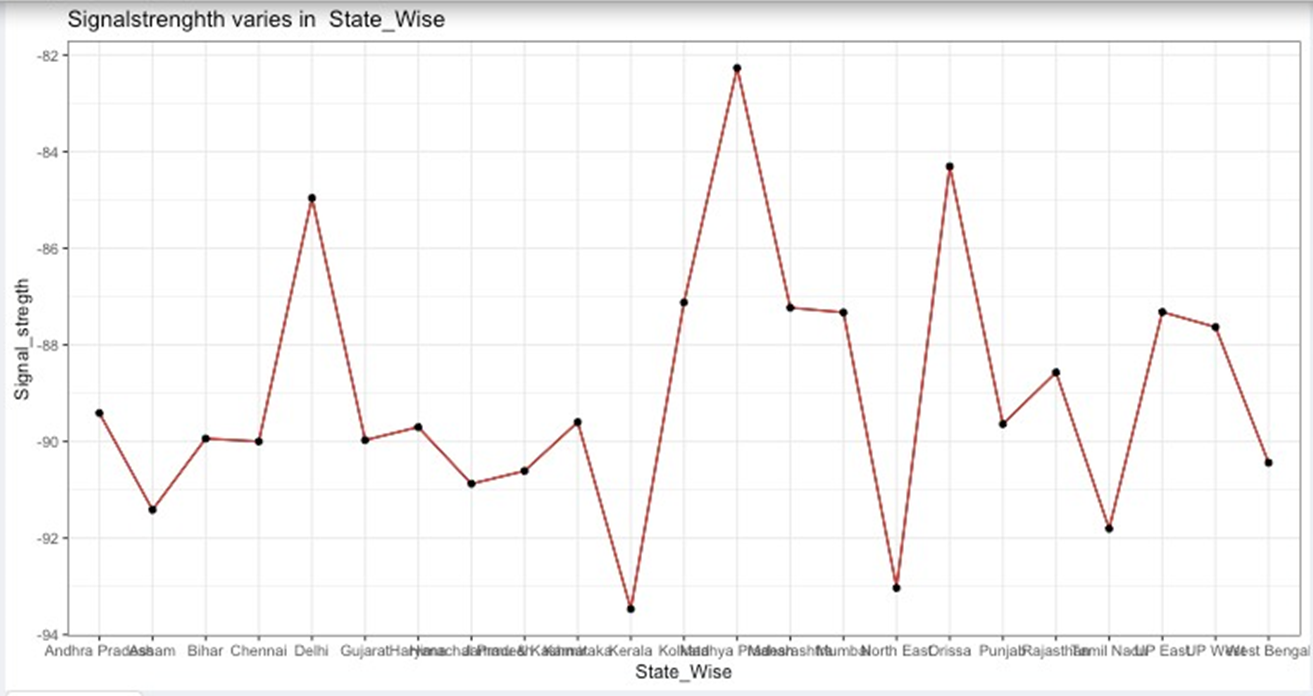
****

**Multivariate Analysis:**

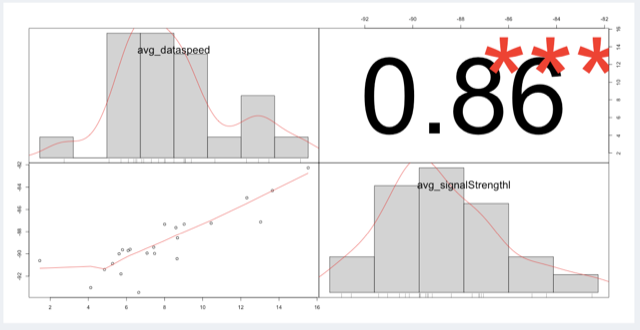




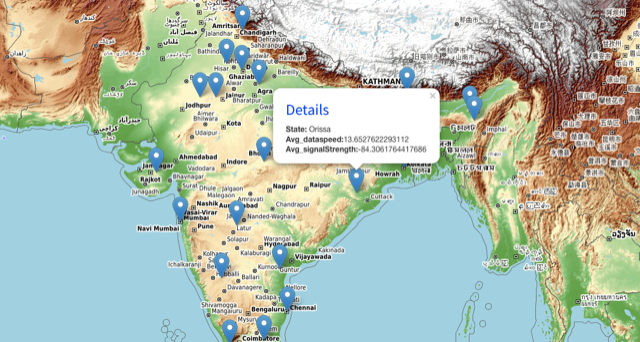
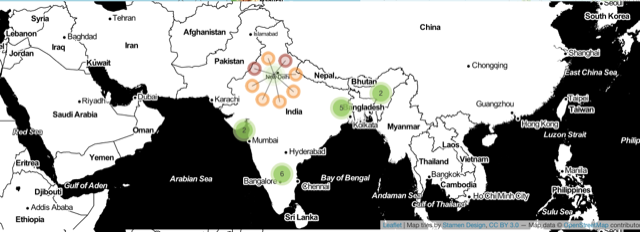




Correlation:



**Maps:**



**8. CONCLUSION AND FUTURE ENHANCEMENT:**

**CONCLUSION:**

We can conclude from the Analysis that-

* Reliance Jio is leading the telecom industry with a very satisfactory consumer experience in terms of Data Speeds for upload and download and Signal Strength. Being a 4G-only network, it is leading the market in terms of both Coverage and Capacity. And this has tremendously increased since the pandemic due on a reliable internet network for communication.
* Signal Strength is said to have an effect with respect to the location. It is observed that Coastal regions have a fluctuating Data Speed and Signal Strength value during certain months of the year, which can be majorly due the regions erratic weather conditions as compared to the regions in the Central and West of India. The regions North-East of India have also reported a low user experience during the monsoon season.
* The Indian mobile competitive landscape is set for further upheaval with the merger of Idea and Vodafone. The impending merger of Vodafone and Idea.
* The Merged Vodafone Idea network will be India’s largest operator in terms of subscriber share, and the company’s combined customer bases and network and infrastructure assets, will undoubtedly make it a force to be reckoned with.

**FUTURE ENHANCEMENTS:**

1.This analysis provides a broad understanding of the quality of service provided by different Service Providers under different terrains. These can extremely beneficial in improving the locations of operations by focusing more on Tier -2 and Tier-3 cities and improve the chances of development of the service and the economy.

2.Several other metrics such as:

a. **Video Experience** which measures video streams from end-user devices, using an ITU-based approach to quantify factors such as load times, stalling and video resolution over an operator's networks.

b. **Voice App Experience** measures the quality of experience for over-the-top (OTT) voice services - mobile voice apps such as WhatsApp, Skype, Facebook Messenger etc. - using a model derived from the International Telecommunication Union (ITU)-based approach for quantifying overall voice call quality and a series of calibrated technical parameters.

And other metrics such as 4G coverage experience and Gaming experience.

3.This analysis can be broadened to improve the coverage in places for tourist interest.

4.This analysis can be used to efficiently structure the upcoming improvements in 4G/5G technology in a more efficient manner since the pandemic has forced economies to rely on network connection.