

Department of Statistics

University of Colombo

IS 4007

Statistics in Practice II

Industry Report

Group 03 – Team Terzo

Agrani Nimshani	-	S14487
Yasiru Pathirana	-	S14565
Malshan Gunasekara	-	S14330
Hasitha Weerasinghe	-	S14383

Abstract

The telecommunication sector is made up of companies that make communication possible on a global scale, whether it is through the phone or the Internet, through airwaves or cables, through wires or wirelessly. These companies created the infrastructure that allows data in words, voice, audio, or video to be sent anywhere in the world. This dataset comprises of data of three telecommunication networks in Bangladesh for 16 regions and their performances are measured using four KPI namely revenue generating base for data usage, average price per data, blended utilization and volume per subscribers. Analysis was carried out using python and necessary charts and graphs were drawn using power bi. We divided the performances of three networks in 16 regions into three categories as well, normal and low performing regions. Out of all regions Barisal is the region with best performances while Rangpur had the worst performance when comparing the business with the competitor. Causes for such behaviors can be found through the implemented dashboard in power bi. Main objective of analyzing the data was to find the current operational gaps in the business and to compare key performance indicators between company and competitor. Implementing an efficient dashboard made it easy to identify patterns and behaviors and was helpful in providing suggestions to enhance business performance.

Table of Contents

Abstract.....	1
Table of Figures	2
1. Introduction	3
1.1 Objectives.....	3
2. Theory and Methodology	4
2.1 Theory	4
2.2 Methodology.....	5
3. Data	6
4. Analysis	7
4.1 Analyzing the Blended Utilization by the Region and Brand	7
4.2 Analysis of APPD by Region and Brand	7
4.3 Analysis of Volume per Subscribers of Com A and Com B.....	8
4.4 Analysis of RGB users	8
4.5 Analysis of KPIs and Extreme APPD values	9
4.6 Analysis of Day wise graphical comparison	10
5. Results of Regional business performance and suspect behaviors	11
6. General Discussion and Conclusions.....	14

Table of Figures

Figure 1 - Blended Utilization by Region and Brand	7
Figure 2 – Average APPD by Region and Brand	7
Figure 3 - VPS by Com A and Com B.....	8
Figure 4 - RGB user map.....	8
Figure 5 - Second page of Dashboard	9
Figure 6 – Page 3 of Dashboard	10
Figure 7 - Dashboard layout for Dhaka South Region.....	11
Figure 8 - Dashboard layout for Chittagong Region.....	122
Figure 9 - Dashboard layout for Comilla Region	13

1. Introduction

Telecommunications is defined as communicating over a distance. The telecommunication sector is made up of companies that make communication possible on a global scale, whether it is through the phone or the Internet, through airwaves or cables, through wires or wirelessly. These companies created the infrastructure that allows data in words, voice, audio, or video to be sent anywhere in the world.

These companies are transmitting or providing services relating to the transmission of, voice, video, or data through owned or leased transmission facilities. Different Telecommunication companies offer telecom services of different types. The services include voice, video, telephone, internet and communication services. The most common form of telecommunications service is phone service, which is done on either a wired or wireless standard.

These services may not be available in all areas or from all companies. The pricing points for the different services vary widely and may be different for residences and businesses. Telecom companies generate revenue via subscription mobile services, fixed landline, and wireless broadband services. Companies can generate revenue from the both revenues generating base data users and voice users. Increase in data traffic will create highly utilized network enabling more people to consume and creating more revenue. The company can increase its revenue by increasing the average data price of the high utilized areas.

Telecommunication is the most competitive and fastest-growing market in the globe. In the past decade, the global telecommunication market has seen a dramatic increase in investment and growth. Dynamic competition in telecom is produced by battle among companies to produce more reliable or more economical commodities. Telecom companies gather different kinds of data to make sure their services answer customers' needs and to develop customer base.

1.1 Objectives

Finding the current operational gaps in order to provide the business with proper suggestions to enhance performance after a thorough region wise comparison of key performance indicators between the company and competitors.

- Identifying regional business performance and suspect behaviors.
- Identifying possible root causes for such behaviors.
- Delivering suggestions to regional managers to improve regional business performances.

2. Theory and Methodology

2.1 Theory

RGB data users

It is a count of users who actually contribute to generate revenue by consuming data facilities of each company after paying for them.

APPD

$$APPD = \frac{Revenue}{Data Volume} \times 100$$

APPD is the average price per data. When average data price is high it will affect the number of users using the network in that specific region.

Utilization

$$Utilization = \frac{Traffic}{Capacity} \times 0.9$$

Network utilization is the proportion of the current network traffic to the maximum amount of traffic that can be handled (capacity). It indicates the bandwidth consumption in the network. While high utilization means more people are consuming the network, generating more revenue and low utilization means few people are consuming the network generating less revenue.

VPS

$$VPS = \frac{Volume}{Total Subscribers}$$

VPS is the volume per subscriber value.

RGB data user percentage

$$RGB data user percentage = \frac{RGB data users}{Total data users} \times 100$$

2.2 Methodology

1. Loading data sets in the python environment.
Company and Competitor sheets of Assignment 1 excel file and Assignment data CSV files were imported to the python environment using the pandas library.
2. Duplicate records of each data file were removed. Records related to an unknown brand in the CSV file were removed.
3. Pivot the CSV file using the Measure Names column and Measure values column taken as values.
4. First 3 characters of the 'site code' column in both company and competitor tables were removed and renamed the columns as 'Site' (This is done to make the site codes in to the same format as in the 'Site' column of the CSV file)
5. 0 values in the 'Data Vol' column in the company table were replaced by Null.
6. APPDs of each site code in the company table were calculated accordingly as that of Competitor and added to a new column named as APPD.
7. Blended Capacity and Blended Traffic columns were created by summing up the capacity and traffic of 3G and 4G users respectively.
8. Competitor table was appended to the company table and competitor was added to the 'Brand' column, records used Competitor. Columns that are not available in the competitor table were dropped before appending the tables. Named that table as 'Company_Competitor'.
9. Unnecessary columns that are not related to the analysis in the CSV dataset were dropped.
10. 'Company_Comprtitor' and CSV data table were merged by left join and the 'Site' column was taken as the reference.
11. Merged table was renamed as the 'Final table' and exported in CSV format.
12. Imported the final table to power bi.
13. For each Blended Utilization, VPS, RGB Data User Percentage, and Average APPD KPI new measures were added.
14. According to the planned layout required charts, cards, and map was created.
15. A conditional formatting was done to add a color code to KPIs APPD and Blended Utilization values as below.

High	-	Green
Mid	-	Blue
Low	-	Red

3. Data

These datasets are of telecommunication data of three networks. The business comprises of two companies as Com A and Com B and they also have a competitor telecommunication company. Data was in multiple excel sheets and we had to model our data to make it useful for our analysis to achieve our objectives. Though there are different types of variables in the data set we have only selected the most important variables which explains our objectives most when forming tables and to conduct our analysis.

Table 01 – company

Qualitative variables

- Brand/ Event.Date/ Region/ Site.Code/ Thana
- Geometry/ Geometry1 – unnecessary for analysis so removed from the dataset

Quantitative variables

- Thana.Code / Capacity.3G1 / Capacity.4G1 / Capacity.Blended / Data.Rev / Data.Vol Traffic.3G1 / Traffic.4G1 / Traffic.Blended

Table 02 – competitor

Qualitative variables

- Site.Code / Event.Date

Quantitative variables

- Appd / Capacity.3G / Capacity.4G / Traffic.3G / Traffic.4G

Table 03 – Company data 2

Qualitative variables

- Event.Date / Date, Brand / Measure.Names1 / Region / Site / Thana

Quantitative variables

- Measure.Values1

Measure.Names1 variable was pivoted to do analysis.

Final table

Final table was created by combining the variables and joining above three tables.

Qualitative variables

- Event.Date / Brand / Region / Site

Quantitative variables

- APPD / Capacity.3G / Capacity.4G / Capacity.Blended / Traffic.3G / Traffic.4G / Traffic.Blended / Rgb.Data.User / Data.User / Data.Vol

4. Analysis

4.1 Analyzing the Blended Utilization by the Region and Brand

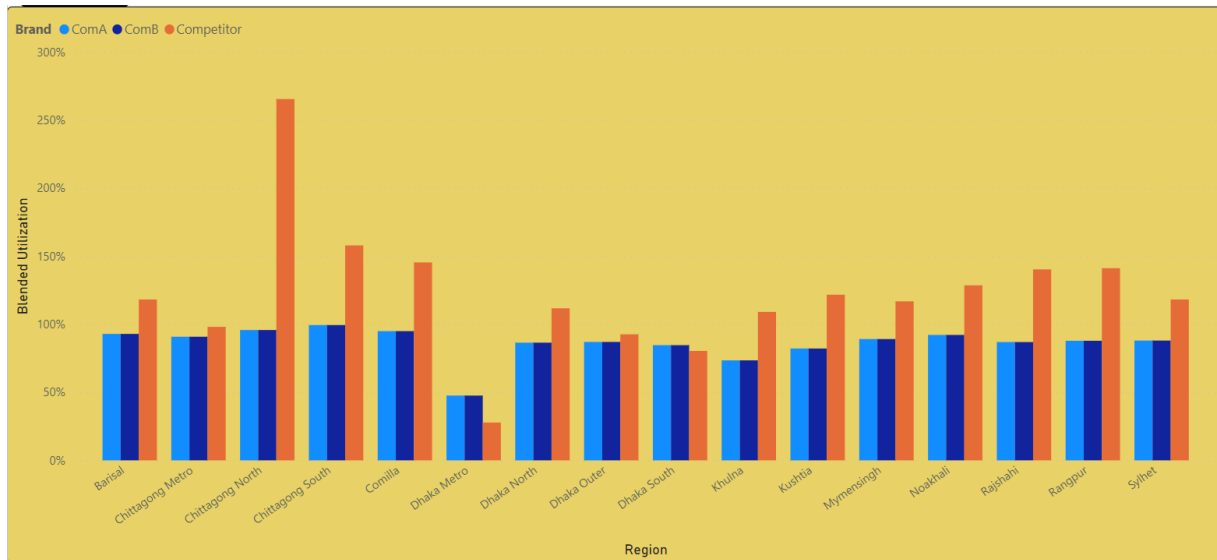


Figure 1 - Blended Utilization by Region and Brand

Through the above multiple bar graph, we can compare the blended utilization of each company region wise. Blended utilization of Com A and Com B are only higher in Dhaka Metro and Dhaka South regions while in all other regions competitor's network is highly utilized than that of Com A and Com B. Since Com A and Com B are two networks of the same company using same resources and same technology, may be the reason for having equal blended utilization in all regions.

4.2 Analysis of APPD by Region and Brand

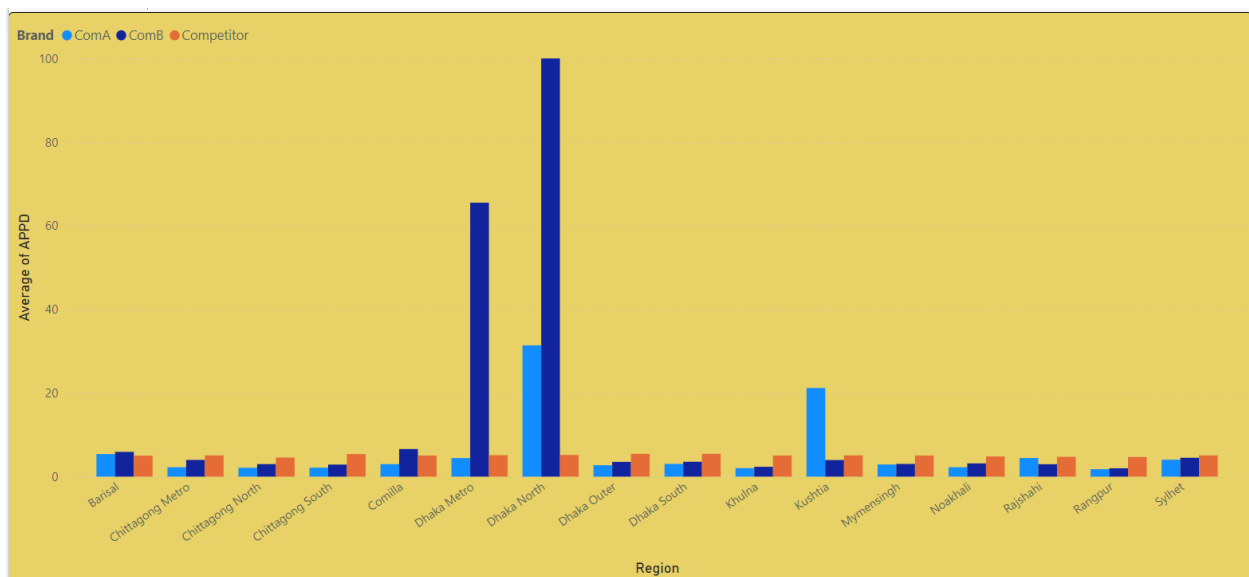


Figure 2 – Average APPD by Region and Brand

Variation of APPD of each company can be observed through the above multiple bar chart. Here we can see that average price per data of Dhaka Metro and Dhaka North of Com B are very high than that of other networks while the competitor network has almost constant average APPD around 5 in every region. Actual average APPD of Dhaka North is 853. Since it is a very high value relative to other average APPDs the range of y axis of above graph is reduced to obtain clear visuals.

4.3 Analysis of Volume per Subscribers of Com A and Com B

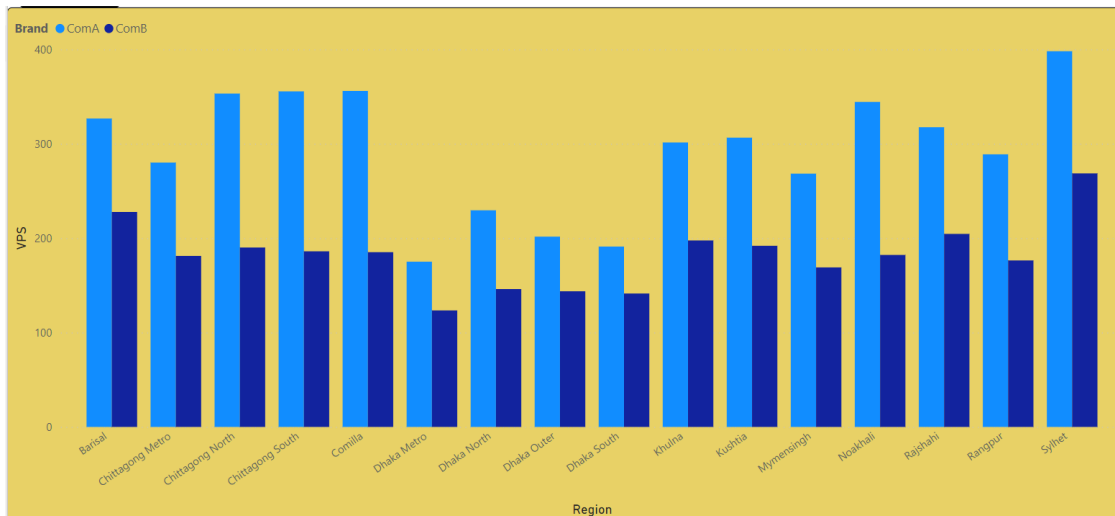


Figure 3 - VPS by Com A and Com B

Variation of Data volume with the number of subscribers in each region is compared above. Here VPS of Com A is always greater than that of Com A in each region. Highest VPS of both networks are from the region Sylhet while the lowest VPS of both networks are from Dhaka Metro.

4.4 Analysis of RGB users

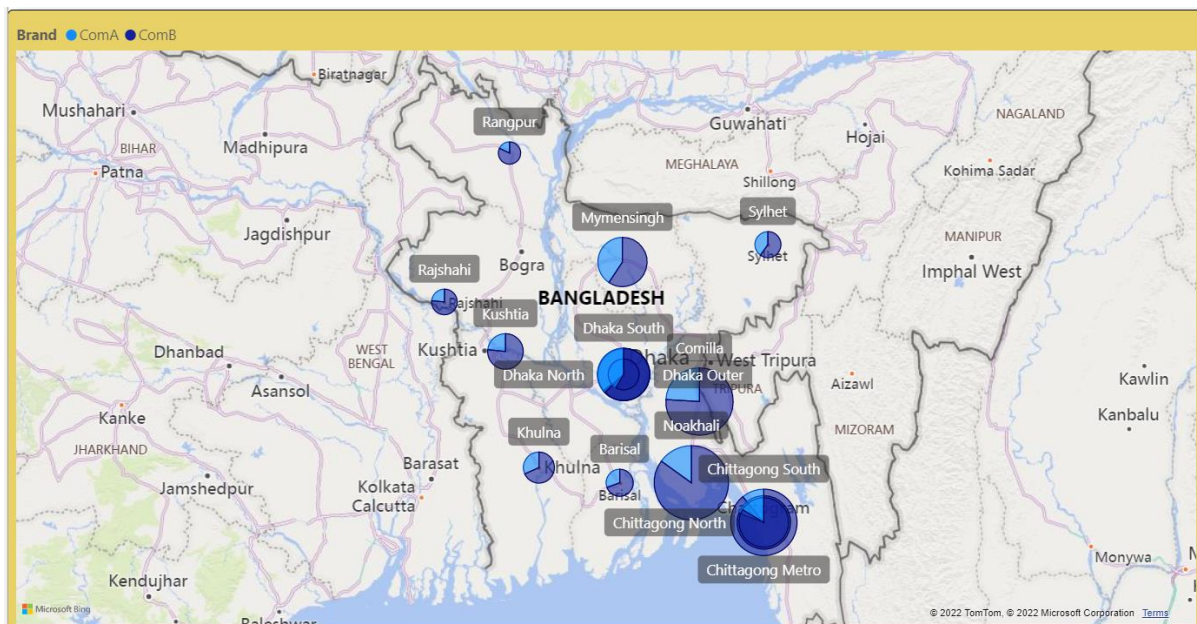


Figure 4 - RGB user map

Revenue generating base by data usage of networks Com A and Com B in regions of Bangladesh are shown above. Each bubble shows the RGB users of that regions indicating a bigger bubble for regions with high data users and vice versa. Here we can observe that RGB users of Com B are highest in Noakhali and lowest in Rangpur while RGB users of Com A are highest in Mymensingh and lowest in Rangpur. Through the above visual we can clearly identify the diversity of network usage of each region.

4.5 Analysis of KPIs and Extreme APPD values

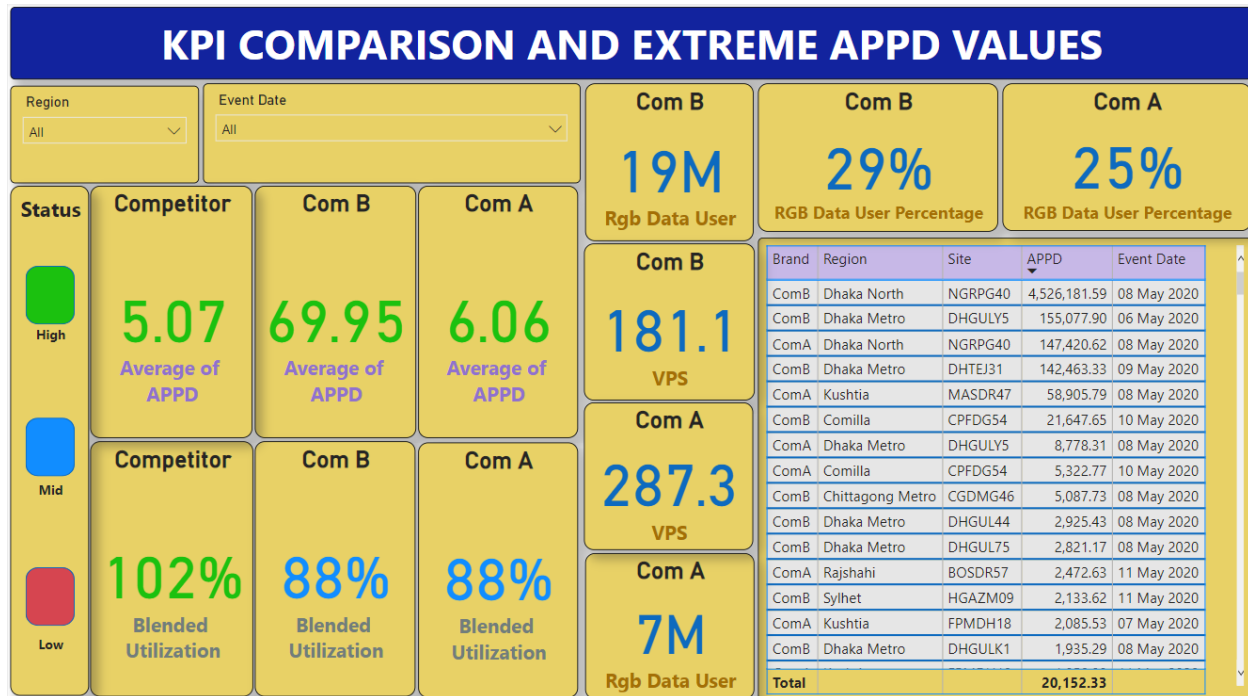


Figure 5 - Second page of Dashboard

This was implemented to compare KPI values of each region of three networks of Com A, Com B and competitor. We can obtain values for each KPI by changing the regions as well as we also can change the date as we only have data of 7 days. Here KPI comparison is obtained for all regions for whole week and we can see that on average APPD of Com B is unusually higher than other 2 networks. This is because there were many extreme values for average APPD for the mentioned site codes as in the above tables.

We also can observe that there are 7M RGB users in Com A and 19M RGB users in Com B networks while from total data users only 25% and 29% contribute to revenue generating in networks of Com A and Com B respectively.

4.6 Analysis of Day wise graphical comparison

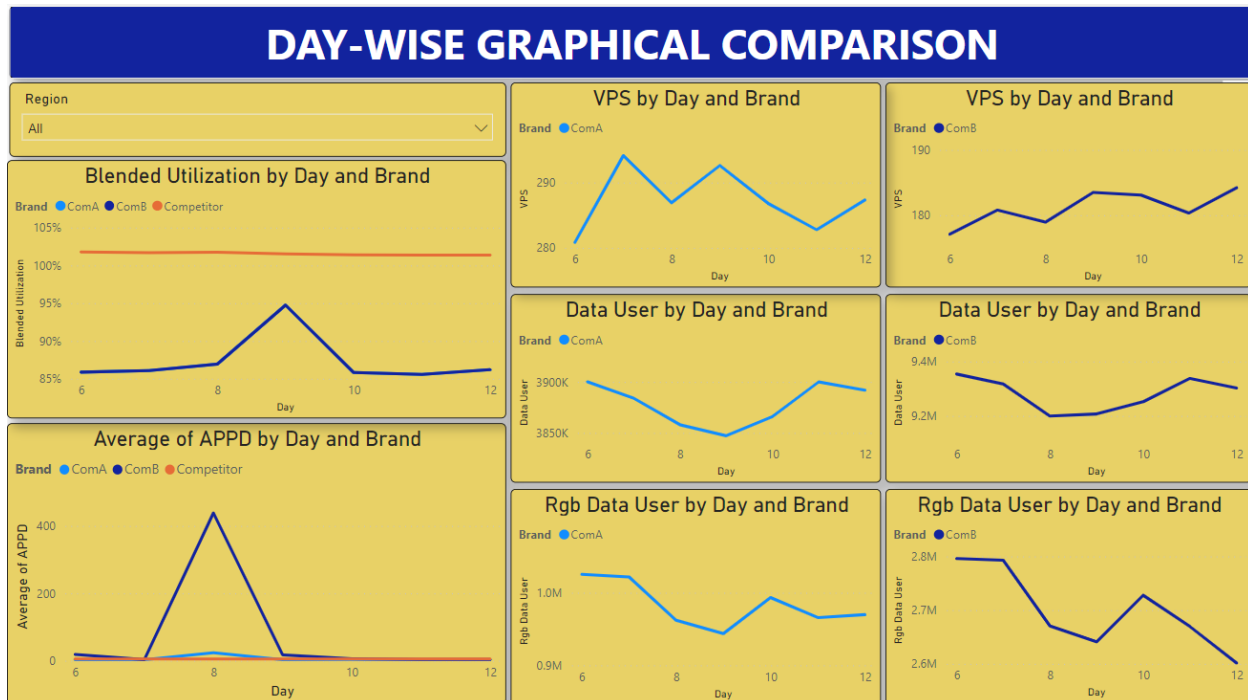


Figure 6 – Page 3 of Dashboard

Here we can observe the trends in behavior of each KPI for each network in all regions for the given 7 days and we can observe the changes in behavior trends for each region separately using the drop down. As an example, from the above interface we expect to identify patterns in each KPI for each network and also, we can observe the variations of all KPIs for a given network.

5. Results of Regional business performance and suspect behaviors

Dhaka South

In this region, the APPD values of all three companies are in the high. And blended utilizations are in the mid-level. When comparing blended utilizations of the companies, both Com A and Com B utilizations are greater than that of Competitor. But APPD of Com A and Com B is lower than the Competitor. When compared with the competitor Com A and Com B teams have not done a great job in price planning.

Overall when considering the whole week VPS is higher in Com A than in Com B. RGB User percentages are equal in both Com A and Com B. There are considerable amounts of extreme APPD values in both Com A and Com B.

In day-wise graphical comparison, from 6th to 12th blended utilization has slightly increased but APPD values have drastically dropped. Although users had used resources (around 85%), company did not gain enough revenue when compared to the competitor. From RGB users' graphs reason behind the APPD dropped can be seen. Number of RGB users of two companies have highly decreased from 6th to 12th and number of Data users has increased at the end part of the graph (although dropped in middle). The non-RGB users might have used a higher amount of data from com B during the days of 10 to 12 (Because RGB user count drastically has dropped from 10 to 12 but VPS was high). Those may be the main reasons for the drop in APPD. The manager of Dhaka South region can arrange promotional campaigns to increase the RGB users, optimize the APPD value, limiting the data usage of non-RGB users to increase the performance of the region.

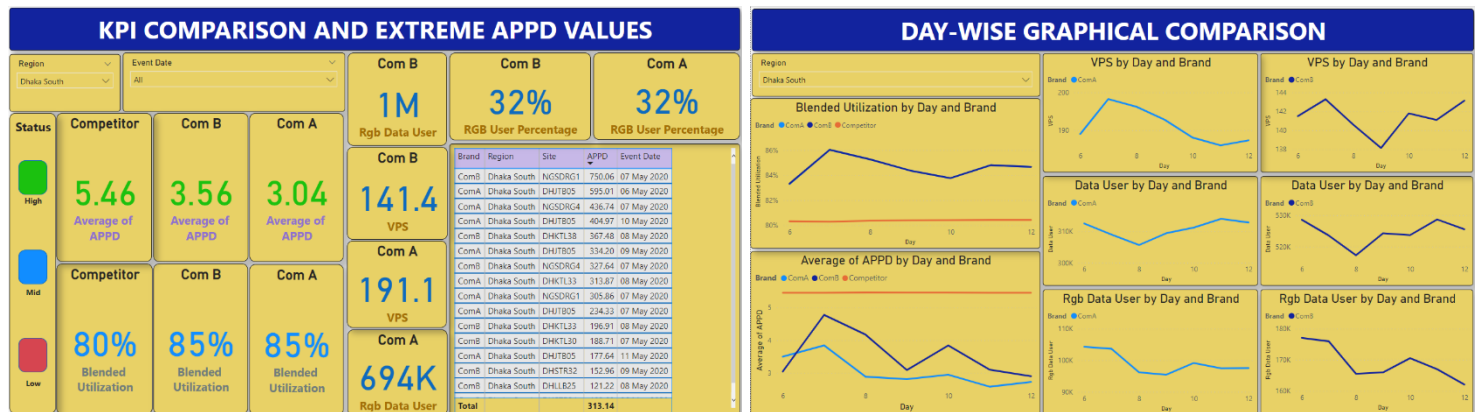


Figure 7 - Dashboard layout for Dhaka South Region

Chittagong Metro

According to the overall weekly data, Competitor and Com B APPD values are high and low in Com A. But all three companies Blended Utilizations are in the mid-range. Com B RGB User Percentage is higher than Com A. 33% and 29% are the values respectively. Com B has five and com A has 3 extreme APPD values in this region. For com B, their team has done a good job in price planning but for com A, the story is different.

According to the day-wise comparison, blended utilization is almost stable. But on 9th it has doubled than the stable level. APPD also has the almost same behavior and on 8th APPD has tripled. Both com A and B's VPS peaked on 8th but only com B's APPD has increased (On 8th non-RGB users of Com A have used more data). In both com A and com B, the RGB User count has drastically dropped from 6th to 12th. The manager of this region should try to find the reasons for dropping of the amount of RGB Users and a solution to increase the performance of the region. He can suggest methods to increase the number of RGB users by introducing new data packages attracting subscribers.

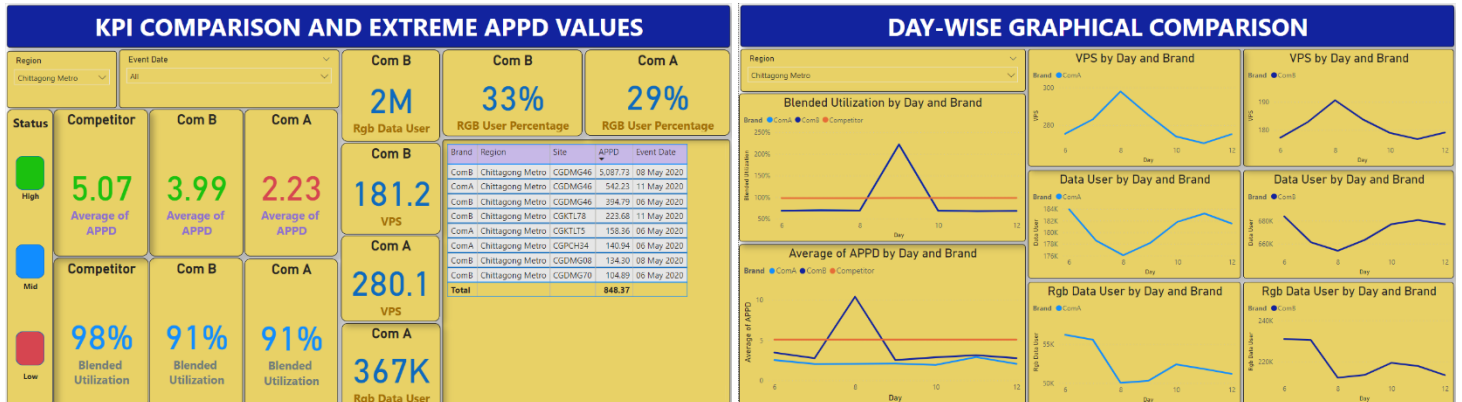


Figure 8 - Dashboard layout for Chittagong Region

Comilla

According to the overall weekly data, Competitor, Com A, and Com B have high APPD values. Blended utilization of competitor is high while both com A and com B blended utilizations are in mid-range. Both com A and com B have three extreme APPD values. Com B RGB User Percentage is a bit higher than com A. Com A VPS is approximately twice of com B. Teams of both com A and com B have done a pretty good job in price planning than the competitor. Though company customers have used a low number of resources than competitor their revenue (APPD) is higher than that of the competitor (Only com B).

According to the day-wise comparison, com A and com B blended utilization are very stable than other regions. And pretty much lower than the competitor. APPD is also stable and close to each other in all companies. But on the 10th APPD of com B drastically increased. That is the cause for the overall week's APPD higher of the com B than the competitor. VPS of both companies has an overall upper trend from 6th to 12th. RGB Users are dropped in both companies on the same range. In the data users also has the

same kind of pattern. The manager of this region can increase the performance by reducing the reducing of RGB users. Because when considering the overall week RGB users have decreased.

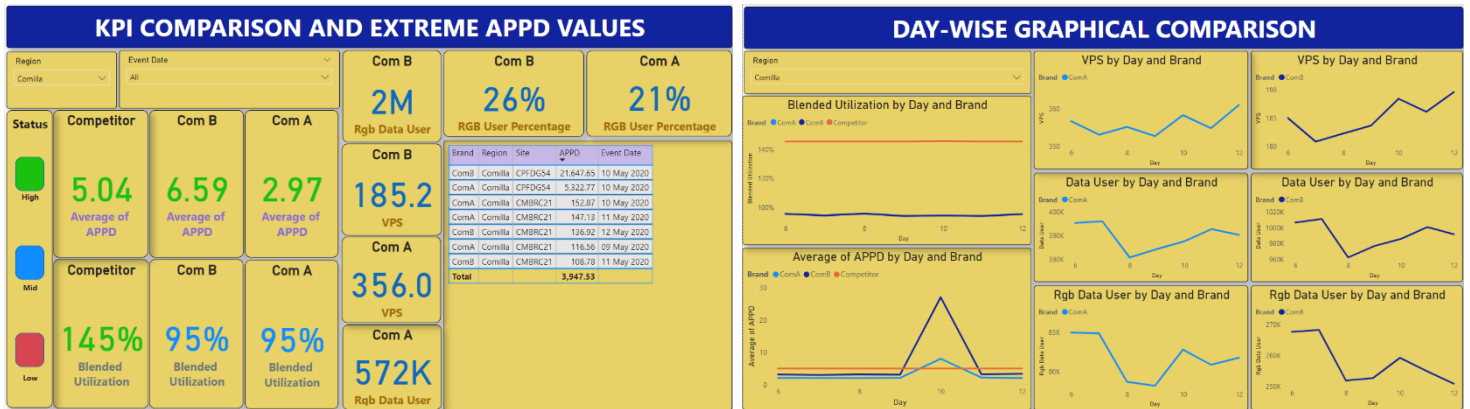


Figure 9 - Dashboard layout for Comilla Region

6. General Discussion and Conclusions

- To the ease of discussing the conclusions and results we have divided the 16 regions into three main categories as well performing, normal performing and low performing regions.
- Overall performance of both Com A and Com B are higher in the region Barisal, Comilla and Sylhet during the whole week than that of the competitor because blended utilization of both Com A and Com B are less than competitor but APPD is greater than that of competitor. So, with the use of minimum resources and with higher APPD value they have managed to increase their revenue in that region.
- Similarly, in regions Rajshahi and Kushtia, APPD for Com A is considerably higher than the competitor's and its blended utilization is lower than competitor. So, again Kushtia and Rajshahi can be taken as a well performed region during the given week.
- Same scenario can be observed in both Com A and Com B of Dhaka North region and it has unusually high APPD of 853.32. This is mainly due to inclusion of very high extreme values for APPD in that region. When we encounter with very high extreme values such as 4526181, it affects to the average APPD making it a bigger value.
- Reasons behind extreme APPDs may be having very low data volume because APPD is obtained by Revenue/Volume. Reasons for very low data volumes may be because of the errors made when entering data. Units of data volume may be incorrectly entered. Another reason may be that the user hasn't used his data though he has paid for it, may be there was a signal drop or was in a no signal area making one unable to use their data even though they have paid for them. We need to further investigate the reason behind these facts.
- We can observe a specialty in the region Dhaka Metro because blended utilization of all Com A, Com B and competitor are less than 50% and even have a higher APPD for Com B than the competitor. This may be again due to presence of extreme values so, we need to further investigate these matters.
- We also can observe normal performances in the regions Chittagon Metro, Chittagon North, Chittagon South, Dhaka Outer, Kulna and Mymensingh because APPDs and utilizations are of same level but lower APPD than that of competitor during the given week.
- Dhaka South, Noakhali and Rangpur can be identified as regions with low performance due to presence of low APPD values and high blended utilizations than the competitor within the given week even in the absence of extreme values or less amount of extreme values.
- According to our analysis we observed that the competitor maintained a constant APPD and constant blended utilization during the given period of time. But those of Com A and Com B were not maintained constant. So, we can suggest the managers to take actions for making their pricing plans and blended utilizations constant.
- Having unstable and extreme APPD values in Com A and Com B will affect the customer base and may increase the churn rate. So, to perform well in future for a long-time, managers should identify the reasons for such extreme APPDs and make necessary actions through our suggestions mentioned above.