SamKnows Data Analyst Tech Test

Tahmeed Hossain

20/07/2021

Data Cleaning

After loading all the required datasets to the environment, we need to filter the datasets based on the requirements. As we only need data from two cities, Samsville and Databury, we need to identify who lives in other cities so we can remove them from the data. We can see only two people lives in Irrelevantsford.

```
## # A tibble: 2 x 4
##   person_id city            type_of_broadband_connection name_of_isp
##       <dbl> <chr>           <chr>                        <chr>
## 1     51815 Irrelevantsford ADSL                         Fibrelicious
## 2     17777 Irrelevantsford VDSL                         Fibrelicious
```

After removing those, rows with NA values are removed from datasets. Now we can calculate average download and upload speed for each user. First 5 rows are shown below:

Table 1: Average Download speed of each user

person_id	Avg_measured_download_speed_in_Mbps
11218	19.764314
11560	195.918467
11677	8.954182
12309	9.104108
12312	17.027750

Table 2: Average Upload speed of each user

person_id	Avg_measured_upload_speed_in_Mbps
11218	5.1446965
11560	78.1698964
11677	0.5259603
12309	0.5509977
12312	4.4410936

Optional Part

Data Quality

Few issues: 1. Missing Data- Number of missing data is in this case is not really high but in case of analysing, few missing data is enough to make the calculation wrong. Moreover, we are counting average speed, minimum speed and maximum speed etc. So missing data is an issue in this case.

Although, there was no outlier after data was cleaned.

Data Summarization and Presentation

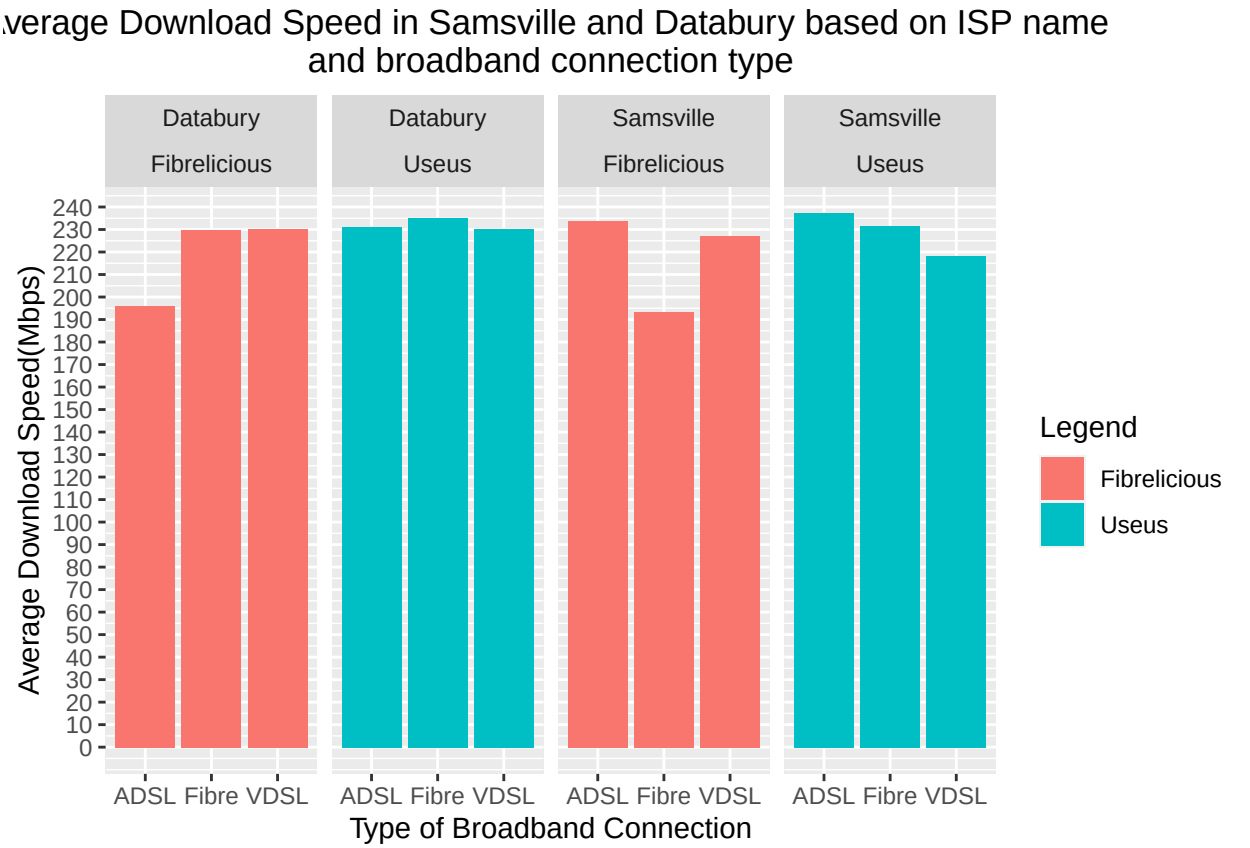
Download and upload speed for this report are in Mbps.

Table 3: Average download and upload speed for each person

Person_id	city	connection_type	ISP_name	download_speed	upload_speed
50868	Samsville	ADSL	Fibrelicious	19.764314	5.1446965
21044	Samsville	ADSL	Fibrelicious	195.918467	78.1698964
23120	Samsville	ADSL	Fibrelicious	8.954182	0.5259603
69960	Samsville	ADSL	Fibrelicious	9.104108	0.5509977
28676	Samsville	ADSL	Fibrelicious	17.027750	4.4410936

Table 4: Average download and upload speed for each person with 60th percentile

Person_id	city	connection_type	ISP_name	download_speed	upload_speed	60th_percentile
50868	Samsville	ADSL	Fibrelicious	19.764314	5.1446965	26.35735
21044	Samsville	ADSL	Fibrelicious	195.918467	78.1698964	220.36182
23120	Samsville	ADSL	Fibrelicious	8.954182	0.5259603	11.74304
69960	Samsville	ADSL	Fibrelicious	9.104108	0.5509977	11.84145
28676	Samsville	ADSL	Fibrelicious	17.027750	4.4410936	23.31292

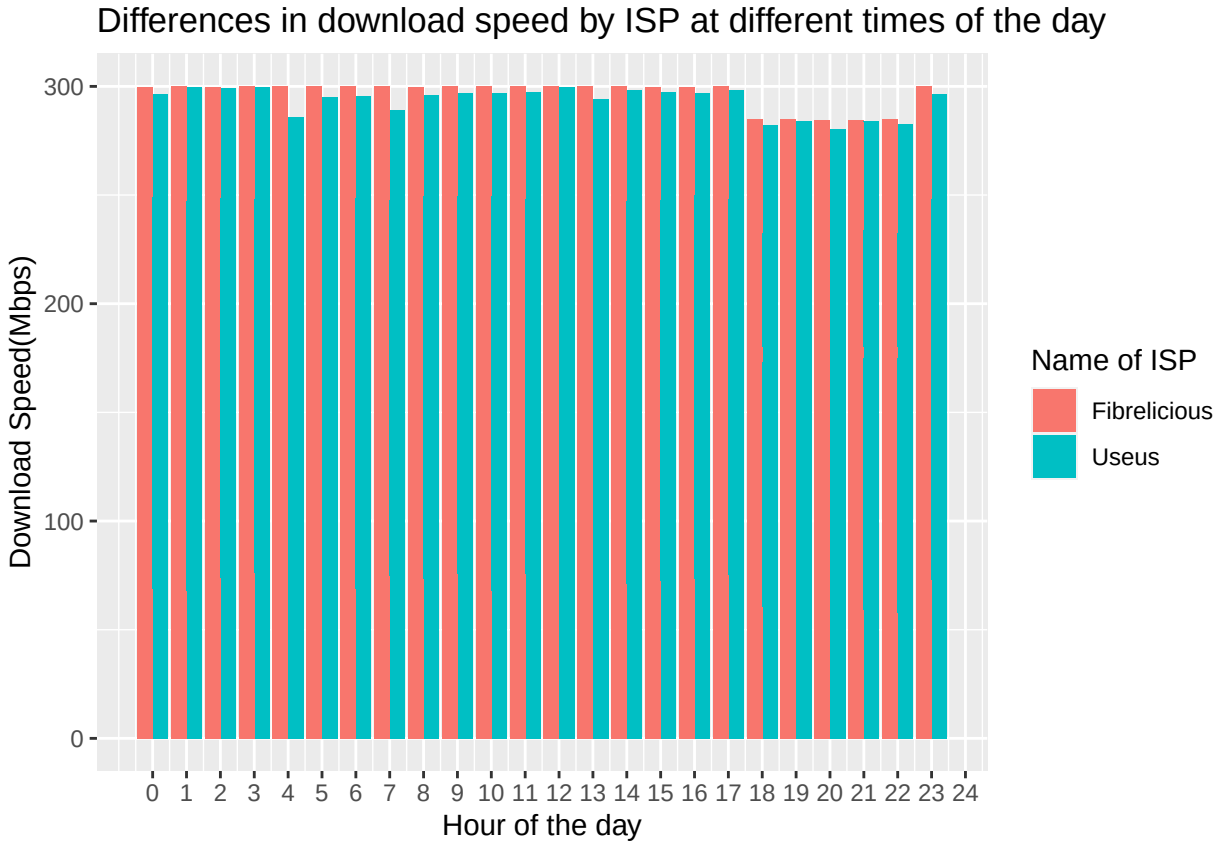


From the plot above, we can see that the people who lives in Samsville and using fibre connection of

Fibrelicious has the lowest download speed. But people who uses ADSl type connection from Useus get the the highest download speed. Moreover, in Samsville, performance of Useus is better than Fibrelicious as two of the three type of connection outperform Fibrelicious.

In case of Databury, scenario is same as Samsville. Useus is outperforming Fibrelicious in every type of connection.

So overall, we can conclude that, Useus has better download speed in both Samsville and Databury. If a consumer who lives in Databury and have fibre connection, he will get a better speed from Useus. He will get 230 mbps if he use Fibrelicious and 235 mbps if he use Useus.



From the above plot, we can see download speed for each provider are is relatively same throughout the day except between 5 pm - 10 pm. There can be multiple reason for this. One main reason can be number of user increases between this time. Another interesting this is, download speed of Fibrelicious doesn't flactuate that much. On the otherhand, download speed of Useus fluctuates quite a few times. As we already know from previous plots that Useus is a better ISP, it indicates that number of users is higher than Fibrelicious.