

1

# The Internet Usage during COVID-19

Jessica de Oliveira Moreira

## Project Objectives

1. Observe changes in the Internet usage during COVID-19.
2. Infer people's behaviour from their Internet usage.

## Changes in Internet usage during COVID-19

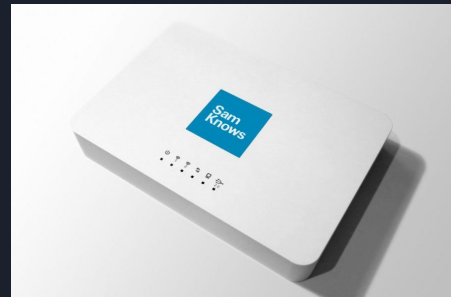
- Analyze Federal Communication Commission's Broadband Measurement Data.
- Did the usage of the Internet change during quarantine?
- Changes in download and upload data consumption and speed, peak periods of usage, etc.

## What can we learn about people's quarantine behaviour from their Internet usage?

- Did people's Internet habits change during quarantine?
- What can we learn about the change from the Internet usage data?
  - Did people work from home?
  - Did people participate in online learning?
  - Did people comply with stay at home orders?

## The Data Set

- Data set gathered by the FCC through the program Measuring Broadband America (MBA).
- 4-5k of volunteers throughout the US who receive whiteboxes and have their Internet consumption measured.
- Data gathered from each whitebox every hour.
  - 20GB of data in total per month.
  - Datasets released every month.



# ISPs and Regions

- 16 different ISPs and 4 different regions represented.

ISP
AT&T
CenturyLink
Charter
Cincinnati Bell
Comcast
Cox
Frontier DSL
Frontier Fiber

ISP
Hawaiian Telcom
Hughes
Mediacom
Optimum
Verizon DSL
Verizon Fiber
Wildblue/ViaSat
Windstream

Census region	total boxes	% total boxes	% total U.S. broadband subscribers
Midwest	1239	28.30%	22.17%
Northeast	772	17.63%	17.80%
South	1321	30.17%	36.93%
West	1046	23.89%	21.96%

Retrieved from Technical Appendix to the Eighth MBA Report

## The Environment to Analyze the Data Set

- Tools used:
  - Virtual Machine on Google Cloud.
  - Jupyter Notebooks (Datalab) for data analysis - Python.
  - BigQuery Database for data storage.
  - Git Repository.

Google Cloud Platform Summer Research

Search products and resources

BigQuery FEATURES & INFO SHORTCUT

Query history

Saved queries

Job history

Transfers

Scheduled queries

Reservations

BI Engine

Resources + ADD DATA

Search for your tables and datasets

summer-research-278923

fcc

census\_block

datausage

dlping

dns

excluded\_units

Query editor

+ COMPOSE NEW QUERY HIDE EDITOR FULL SCREEN

```
1 SELECT FROM `summer-research-278923.fcc.datausage` LIMIT 1000
```

Run Save query Save view Schedule query More

datausage QUERY TABLE SHARE TABLE COPY TABLE DELETE TABLE EXPORT

Schema Details Preview

Row	unit_id	dtime	sk_tx_bytes	sk_rx_bytes	cust_wired_tx_bytes	cust_wired_rx_bytes	cust_wifi_tx_bytes	cust_wifi_rx_bytes	location_id
1	1009168	2019-09-30 00:34:04 UTC	27531531	292127424	10012	171574	276904	6458744	3324928
2	1009168	2019-09-30 01:24:07 UTC	23115400	105221296	17496	50220	662506	1200660	2224020

Rows per page: 100 1 - 100 of 55255018 First page < > > Last page



```
Google Cloud Datalab D-U Average Consumption 2020 (autosaved)
Notebook + Add Code + Add Markdown X Delete ↑ Move Up ↓ Move Down ▶ Run □ Clear ○ Reset Session ☰ Kernel: python3

#Reading Data
import pandas as pd
jan=pd.read_csv('../Data Usage csv/curr_datausage_jan20.csv')
feb=pd.read_csv('../Data Usage csv/curr_datausage_feb20.csv')
mar=pd.read_csv('../Data Usage csv/curr_datausage_mar20.csv')

#Excluding wrong results for download
jan=jan[jan["cust_wired_rx_bytes"]<3000000000000]
feb=feb[feb["cust_wired_rx_bytes"]<3000000000000]
mar=mar[mar["cust_wired_rx_bytes"]<3000000000000]

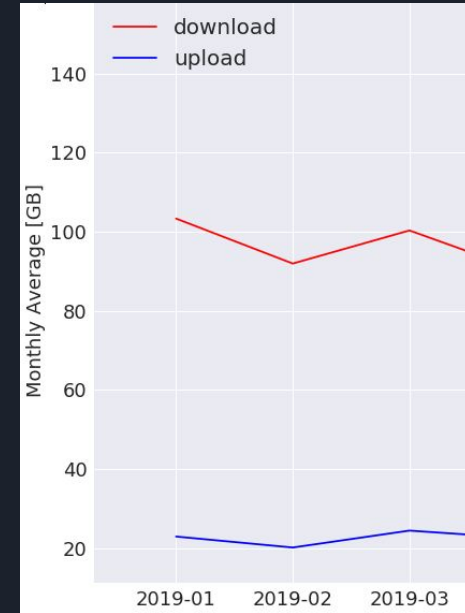
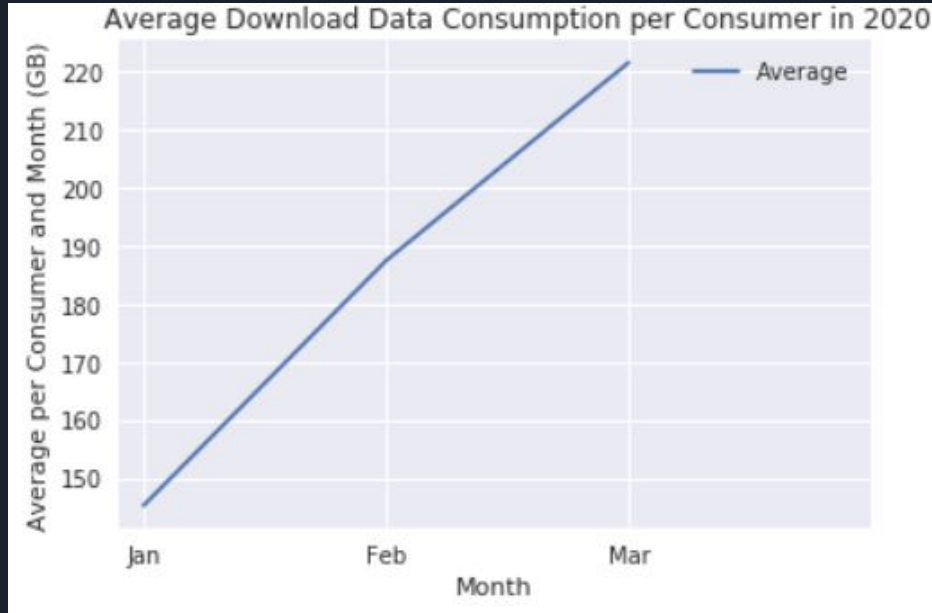
#Excluding wrong results for upload
jan=jan[jan["cust_wired_tx_bytes"]<3000000000000]
feb=feb[feb["cust_wired_tx_bytes"]<3000000000000]
mar=mar[mar["cust_wired_tx_bytes"]<3000000000000]

# Creating bytes_total_download in GB
jan["bytes_total_download"]=(jan["cust_wired_rx_bytes"]+jan["cust_wifi_rx_bytes"])/1000000000
feb["bytes_total_download"]=(feb["cust_wired_rx_bytes"]+feb["cust_wifi_rx_bytes"])/1000000000
mar["bytes_total_download"]=(mar["cust_wired_rx_bytes"]+mar["cust_wifi_rx_bytes"])/1000000000
#Creating total_bytes_upload in GB
jan["bytes_total_upload"]=(jan["cust_wired_tx_bytes"]+jan["cust_wifi_tx_bytes"])/1000000000
feb["bytes_total_upload"]=(feb["cust_wired_tx_bytes"]+feb["cust_wifi_tx_bytes"])/1000000000
mar["bytes_total_upload"]=(mar["cust_wired_tx_bytes"]+mar["cust_wifi_tx_bytes"])/1000000000

# Calculating total download bytes per unit_id
jan1=jan.groupby(["unit_id"])['bytes_total_download'].agg('sum')
feb1=feb.groupby(["unit_id"])['bytes_total_download'].agg('sum')
mar1=mar.groupby(["unit_id"])['bytes_total_download'].agg('sum')
```

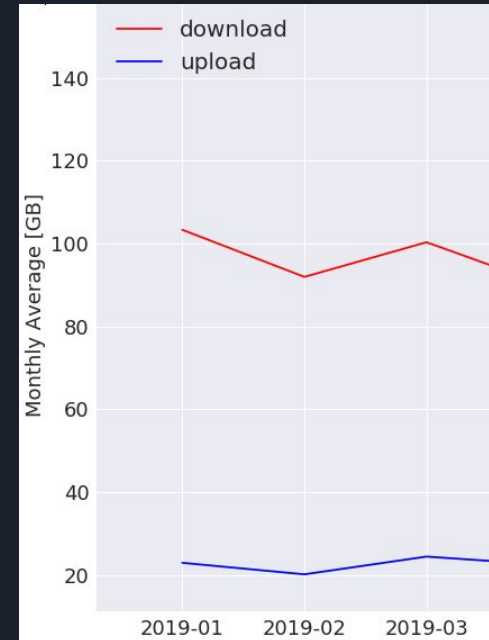
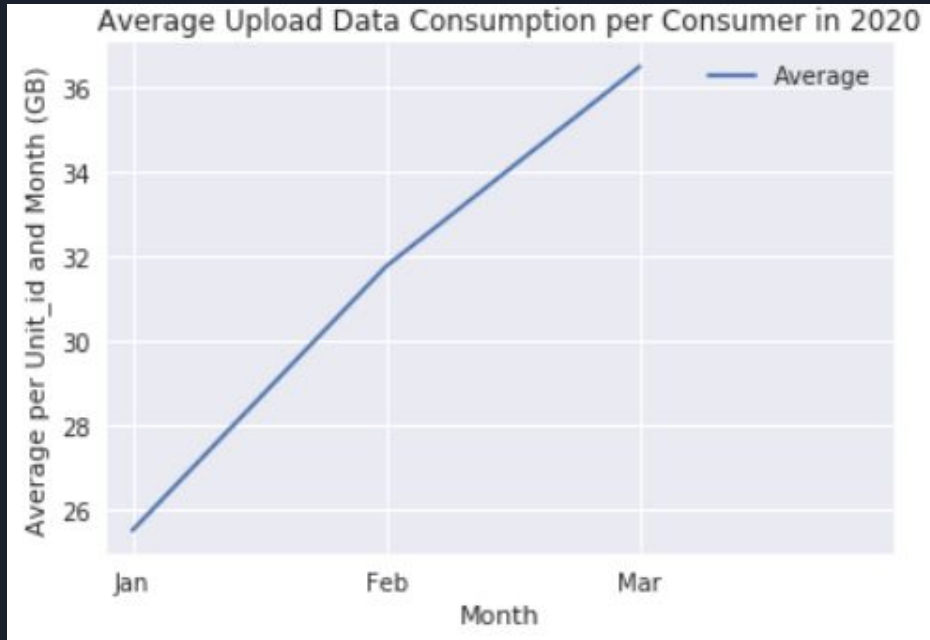
# Average Data Downloaded

- Increase of approximately 53% from January to March in 2020.



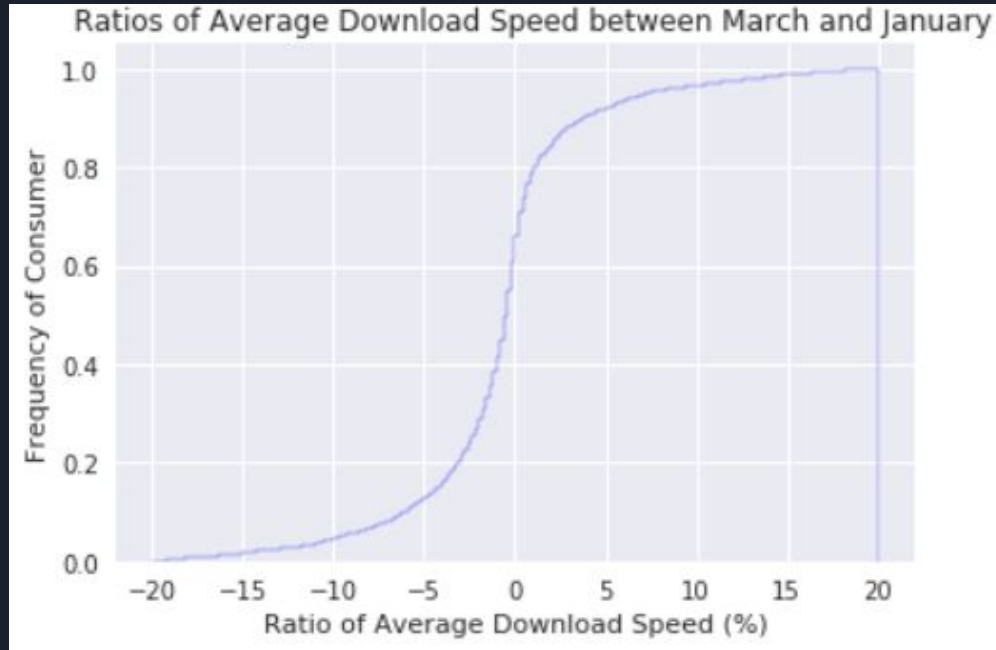
# Average Data Uploaded

- Increase of approximately 38% in the average consumption of upload from January to March.



# Average Download Speed

- Approximately 70% of consumers experienced a decrease on their average download speed.



## Challenges

- Big Datasets: 300GB of data processed in total (2019-20).
- Rate limits to upload to Google Cloud.
- Virtual machine memory.
- Invalid rows in dataset.
- Lack of data description.
- Outdated information about whiteboxes.

## Future Work

- Continue analyzing the changes in Internet consumption, e.g., how consumption changed in business hours.
- Continue analyzing the regional changes in the Internet usage.
- Analyze changes in latency.

## Summary

- Increase in Internet consumption from January to March 2020 - 53% in download and 38% in upload data.
- These numbers potentially show that people tended to indeed stay at home during March.
- It is possible to infer and learn about human behavior through big datasets.

Thank you!

- Jessica de Oliveira Moreira
  - Barnard College' 22
  - [jdm2240@barnard.edu](mailto:jdm2240@barnard.edu)