EDA on Internet Service Providers

PUBLISHED
July 25, 2024

▸ Code

We are starting our first exploratory data analysis around ISPs in the FCC NBM data set. It should be kept in mind that an ISP can be listed multiple times at the same location (offering multiple service).

Our goal is being able to take FCC data and 1. correctly identify the single ISP that is providing each reported service (i.e., deduplication of ISP names) 2. correctly identify that ISP over time (from the same program and from other FCC products).

How should we define an ISP? How can we define coverage (should a service 0/0 be considered as part of the extent of an ISP's coverage)?

We shifted a bit from exploring to trying to classify the quality of information we have from FCC about ISP.

The query that generated the data set is here:

```
select
    frn,
    provider_id,
    brand_name,
    count(distinct location_id) as cnt_locations,
    count(*) as cnt_locations_services,
    bool_or(case when technology = 10 then true else false end) as has_copperwire,
    bool_or(case when technology = 40 then true else false end) as has_coaxial_cable,
    bool_or(case when technology = 50 then true else false end) as has_fiber,
    bool_or(case when technology >= 70 then true else false end) as has_wireless,
    bool_or(case when technology = 60 or  technology = 61 then true else false end) as has_satel,
    array_agg(distinct state_abbr)
from staging.june23
group by frn, provider_id, brand_name
```

The name of the columns match the FCC's documented descriptions.

We are adding:

- `cnt_locations_services` : count of **services**, in one location you can have multiple services with different providers, technology and speeds provides (sometimes one providers can have multiple technology and/or multiple speeds)
- `cnt_locations` : count of locations covered by this specific set of brand_name, provider_id, state_abbr and technology (here if a provider declare providing different speed in that location it will not be counted)
- a serie of flag (`has`) telling if this "combo" is proving said technology
- an array listing in which states are present our "combo"

We have from FCC:

Source: https://us-fcc.app.box.com/v/bdc-data-downloads-output page 4

- `frn` **F**CC **R**egistration **N**umber; "number of the entity that submited the data". It is supposed to be a string of 10 characters (with padding 0).
- `provider_id` : "unique identifier for the fixed service provider"
- `brand_name` : "Name of the entity or service advertised or offered to consumers."

> **Tip**
>
> Every row is matching a combination of **unique** FRN, Provider ID and brand name.

▸ Code

▸ Code

Show entries Search:

frn	provider_id	brand_name	cnt_locations	cnt_locations_services	has_copperwire	has_coaxial_cable	has_fiber	has_wireless
0000012609	240058	CATSKILL MOUNTAIN CABLEVISION	4060	4060	false	true	false	false
		CATSKILL MOUNTAIN						

0000012609	240058	CABLEVISION CABLEVISION	2005	2005	false	true	true	false
0000012781	240068	Mountain Zone TV	9106	13942	false	true	false	true
0000013722	131087	RAINBOW COMMUNICATIONS LLC	9022	15192	false	true	true	false
0000014936	240058	MID-HUDSON CABLEVISION	29551	36861	false	true	true	false
0000017640	131060	BLUE STREAM	34607	34813	false	true	true	false
0000017756	240025	COMMUNITY ANTENNA SYSTEM INC	667	667	false	true	false	false
0001513068	430166	United Electric Cooperative Services	52429	52429	false	false	true	false
0001514090	131161	Sandwich Isles Communications	521	521	true	false	true	false
0001520980	130254	Hawaiian Telcom	224903	266627	true	false	true	false

Showing 1 to 10 of 3,170 entries Previous 1 2 3 4 5 … 317 Next

Numbers for context:

Raw numbers out of the box:

Number of unique `frn`: 2879

Number of unique `provider_id`: 2184

Number of unique brand name pre cleaning: 2902

▸ Code

Removing all capitalization and change underscore for white space help lower the number of unique brand names to: 2878

▸ Code

Potential sources of errors:

- `FRN` can be wrong or not meaningfull
- `provider_id` can be wrong
- brand names can be different or evolve over time.

One case:

First example different FRN and Provider ID

frn	provider_id	brand_name	cnt_locations	cnt_locations_services	has_copperwire	has_coaxial_cable	has_fiber	has_wireless	has_satel	array_agg
0003738655	130432	“EATEL Corp.”	83537	86548	true	true	true	false	false	{LA}
0009873712	131103	“EATEL Corp.”	34494	34497	false	true	true	false	false	{LA}

Other case:

Acentek here have the same provider ID but different FRN

frn	provider_id	brand_name	cnt_locations	cnt_locations_services	has_copperwire	has_coaxial_cable	has_fiber	has_wireless	has_satel	array_agg
0002626984	130008	Acentek	47	47	true	false	true	false	false	{MN}
0002626984	130008	ACENTEK	1395	1395	true	false	false	false	false	{MN}
0002645927	130008	Acentek	19521	26636	true	false	true	true	false	{IA,MN}

Rules for problems:

Less than 10 locations per row:

Less than 10 locations per frns:

▸ Code

Row with less than 10 locations

Number of locations	Number of cases
1	43
2	9
3	12
4	5
5	2
6	4
7	8
8	5
9	3
10	3

Potential solutions:

- We can decide to not keep those rows
- Merge them with either other rows that is matching `provider_id` or `frn` (when this is an option)

More than one frn for a provider_id

▸ Code

Show entries Search:

provider_id	frn	brand_name	cnt_locations	cnt_locations_services	has_copperwire	has_coaxial_cable	has_fiber	has_wireless
130008	0002626984	1136787599	1	1	true	false	false	false
130008	0002626984	1136787600	1	1	true	false	false	false
130008	0002626984	1136785808	1	1	true	false	false	false
130008	0002626984	1136787577	1	1	true	false	false	false
130008	0002626984	1136787610	1	1	true	false	false	false
130008	0002626984	1136785673	1	1	true	false	false	false
130008	0002626984	1136785814	1	1	true	false	false	false
130008	0002626984	1136787585	1	1	true	false	false	false
130008	0002626984	1136785878	1	1	true	false	false	false
130008	0002626984	1136785846	1	1	true	false	false	false

Showing 1 to 10 of 1,170 entries Previous 1 2 3 4 5 … 117 Next

Unique provider_id + brand_name are kind of “green” (for one time frame):

Number of green isp: 1972

We can have one `provider_id` with multiple `frn` and same or not `brand_name` (see TSC for example / 150266)

It seems:

- Windstream has 37 different `frn`: we can maybe test if it has windstream in it’s name ..
- Otelco/GoNet (18 cases)
- Rally Networks/Oregon Telephone Company, is their frn wrongs ? (17 cases)
- Same Provider for differents `frn` and `brand_name` in Minesota (MN) (16 cases)
- `160127` I do not see any kind of specific pattern for this one
- `131486` seems to be RiverStreet Networks with various `frn` (13 cases) -> will be catch by unique_brand_name_by_provider_id
- `190233` multiple brand name and frn but seems to be in Texas and Oklahoma (13)
- `131226` seems to be Fastwyre Broadband divided by technology and states (12 cases) -> will be catch by unique_brand_name_by_provider_id
- `130804` seems to be Mediacom (+ Bolt) with different states and names indicating their states (11 cases)

- Google Fiber (`240041`) seems to be have `frn` split by states (with a weird 'Webpass, Inc.') (11 cases) -> will be catch by unique_brand_name_by_provider_id (except Webpass, Inc which is weird, technology is 70 that I should correct)
- AT&T Inc (`130077`) multiple frn (10 cases) -> filter by unique_brand_name_by_provider_id
- 130079 = Astound_Broadband (10 cases) -> will be catch by unique_brand_name_by_provider_id
- Verizon -> filter by unique_brand_name_by_provider_id
- long ling (130757) & co are problematics (multiples names / one provider / 3 states)
- `130906` is also hard to fix

Assuming that same name (clean version) + same `provider_id` provide us with a unique ISP, it helps move to greensih:

▸ Code

Show entries Search:

provider_id	frn	brand_name	cnt_locations	cnt_locations_services	has_copperwire	has_coaxial_cable	has_fiber	has_wireless
130077	0002716561	AT&T Inc	898882	1033830	true	false	true	true
130077	0005937974	AT&T Inc	836	836	false	false	true	false
130077	0001551530	AT&T Inc	5939424	6468568	true	false	true	true
130077	0003801651	AT&T Inc	7804193	8826083	true	false	true	true
130077	0001857952	AT&T Inc	12375555	15736990	true	false	true	true
130077	0002776771	AT&T Inc	2276468	2516407	true	false	true	true
130077	0002946986	AT&T Inc	1884758	2136555	true	false	true	true
130077	0001552173	AT&T Inc	164732	181121	true	false	true	true
130077	0002860856	AT&T Inc	2416219	2604957	true	false	true	true
130077	0002904654	AT&T Inc	1178180	1331535	true	false	true	true

Showing 1 to 10 of 194 entries Previous 1 2 3 4 5 … 20 Next

This is removing 194 out of 1170.

Same provider_id and same states

Not too sure about this one.

More than one provider_id for a cleaned brand name

This is the case for "EATEL".

▸ Code

Show entries Search:

clean_name	provider_id	frn	brand_name	cnt_locations	cnt_locations_services	has_copperwire	has_coaxial_cable	has_fiber
"eatel corp."	130432	0003738655	"EATEL Corp."	83537	86548	true	true	true
"eatel corp."	131103	0009873712	"EATEL Corp."	34494	34497	false	true	true
bstwifi	400029	0028145480	BSTWIFI	3704	3704	false	false	false
bstwifi	440187	0026140111	BSTWIFI	3704	3704	false	false	false
bstwifi	320051	0023521073	BSTWIFI	3704	3704	false	false	false
btc fiber	130136	0004560645	BTC Fiber	14106	25354	true	false	true
btc fiber	410100	0027930288	BTC Fiber	2098	2098	false	false	true
cambridge telephone company	130192	0003720984	Cambridge Telephone Company	1629	3830	true	false	true
cambridge telephone company	131043	0002393544	Cambridge Telephone Company	1391	1391	true	false	true
citizens			Citizens					

telephone cooperative	130267	0003749264	Telephone Cooperative	3210	4710	true	false	true

Showing 1 to 10 of 63 entries Previous 1 2 3 4 5 6 7 Next

This could also move 63 cases in the greenish spot. -> nop

What are our options/next steps:

- Make a column “ready to go”

▸ Code

Show entries Search:

clean_name	provider_id	frn	brand_name	cnt_locations	cnt_locations_services	has_copperwire	has_coaxial_cable	has_fiber
"eatel corp."	130432	0003738655	"EATEL Corp."	83537	86548	true	true	true
"eatel corp."	131103	0009873712	"EATEL Corp."	34494	34497	false	true	true
"pgb fiber, llc d/b/a pioneer communications"	450101	0030512396	"PGB Fiber, LLC d/b/a Pioneer Communications"	308	308	false	false	true
"pioneer communications"	131044	0002334795	"Pioneer Communications"	16965	17946	true	true	true
1 point communications	270002	0021352968	1 Point Communications	51	52	false	false	true
101netlink	190002	0018247254	101Netlink	64790	64790	false	false	false
1136785653	130008	0002626984	1136785653	1	1	true	false	false
1136785658	130008	0002626984	1136785658	1	1	true	false	false
1136785660	130008	0002626984	1136785660	1	1	true	false	false
1136785662	130008	0002626984	1136785662	1	1	true	false	false

Showing 1 to 10 of 3,170 entries Previous 1 2 3 4 5 … 317 Next

A good example could be 131167 and how we can discriminate Orbitel communications. We can also prob. raise the bar of “few locations”.

A quick summary of where we are:

▸ Code

```
green not green     olive
 1778      1170       222
```

- Make an id <—> provider_id / frn / brand_name table

Typology of ISP

The data was generated from June 23 FCC release and assumed that an FRN = ISP.

Can we guess who is a small ISP?

▸ Code

Show entries Search:

X	frn	cnt_co	cnt_services	cnt_locations	area_sqkm2	states
1	12609	3	6065	4826	452.310958	{36}
2	12781	3	13942	9106	4795.352168	{48}
3	13722	5	15192	9022	339.886684	{20}
4	14936	5	36861	29551	1257.036624	{36}
5	17640	11	34813	34607	326.244992	{12}

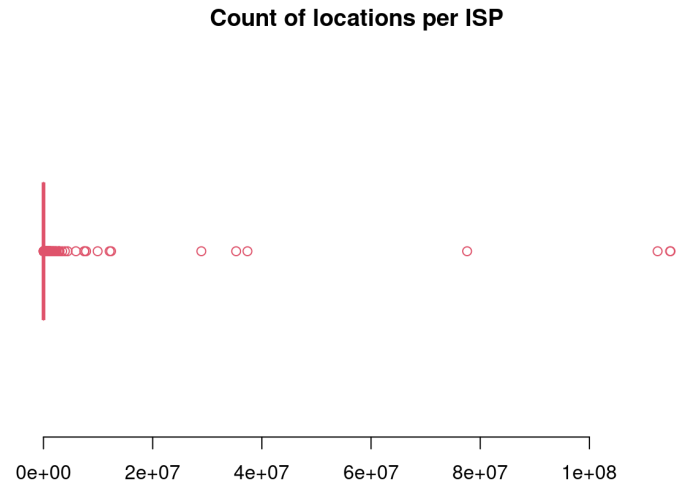
6	17756	2	667	667	23.961069	{55}
7	1513068	9	52429	52429	3269.987131	{48}
8	1514090	4	521	521	81.786318	{15}
9	1520980	4	266627	224903	6380.941177	{15}
10	1543347	3	4610	4610	924.887456	{06}

Showing 1 to 10 of 2,879 entries Previous 1 2 3 4 5 … 288 Next

▸ Code

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      1    1384    5476  253046   19588 114863490
```

▸ Code



Some ISP are declaraing covering a huge number of locations. Some low counts are probably errors.

Count of FRN with a less than 10 locations: 36

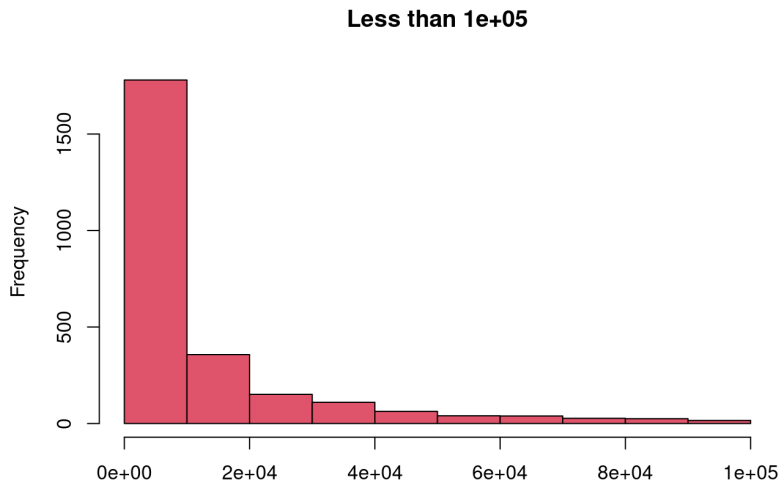
Count of FRN with more than 500 000 locations: 74

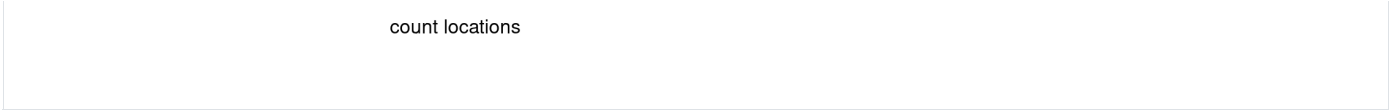
▸ Code

If we filter them out (removing 110 cases):

100 000 10 000

▸ Code





List of ISP that the Broadband team that are good reference of small provider:

Name	FRN	Locations
Newport Utilities, TN	0027152438	15383
SandyNet, OR	0004119376	4439
ECFiber, VT	0027379676	22926
Maple Broadband, VT	0032366692	315
Uplink	0026218602	1611
Black Bear Fiber	0025132648	1583
QCOL	0019663095	5610
Salsgiver	0011167079	29941
All Points Broadband	0023524705	107803
Marquette-Adams Telephone co-op	0003774023	130783
USI fiber	0017096538	71466
Scott county telephone co	0002069862	7829
PANGAEA	0016202236	8410
Blue Mountain Networks	0005450507	310013

Side notes:

- Newport Utilities = NUconnect
- SandyNet, OR = City of Sandy, OR
- USI FIber =
- Blue Mountain Networks = Blue Ridge Mountain Electric Membership Corporation