Understanding 4G LTE Connections

CSDS 325 Project 5

by *Allard Quek*

Case ID: *axq54*

1. Introduction

This project centers around the theme of understanding 4G LTE connections, from the measurements involved, factors affecting those measurements, to the relationship between different metrics. Mobile communications are known to be inherently stochastic in nature, which means that any guarantees provided network metrics are largely only statistical nature and are highly dependent on the actual physical environment. Nonetheless, analyzing this real world data set is instructive as factors associated with 4G LTE connections can be studied and the relationship between certain factors could be better appreciated or perhaps even explained.

2. Proecdure

2.1 Gathering Data

The dataset used for this project is courtesy of Stefan Farthofer, Matthias Herlich, Christian Maier, Sabrina Pochaba, Julia Lackner, Peter Dorfinger, CRAWDAD dataset srfg/lte-4g-highway-drive-tests-salzburg (v. 2022-01-18), downloaded from https://crawdad.org/srfg/lte-4g-high. This dataset is produced from a 4G LTE long-term mobile communications drive test, containing active 4G LTE measurements via repeated drive tests that covers two years on a typical highway section of 25km length in Austria. There are a total of 267198 data points, where each point includes signal quality metrics, GPS position, GPS time, and instantaneous data rate. The following table contains a description for each of the fields in the dataset and an overview of the measurements taken:

Parameter	Description	Example value	Minimum	Maximum	Mean	Standard deviation
time	ISO 8601 timestamp	`2019-12-24 20:57:32`	2018-01-15	2019-12-20		
lat	latitude coordinate in degrees	`47.822768`	47.842 587	47.856 92	47.851 03	0.004 167
long	longitude coordinate in degrees	`13.040866`	13.080 002	13.333	13.192 63	0.073 307
ele	elevation in meters	`450.7`	398.4	712.7	582.5	38.5
signal	signal strength in "bars"	`5.0`	0	5	3.74	1.40
rssi	received signal strength indicator (RSSI) in dB	`-57.0`	-111	-53	-64.37	8.28
sinr	signal-to-interference-plus-noise ratio (SINR) in dB	`7.0`	-20	30	10.57	9.96
rsrp	reference signal received power (RSRP) in dBm	`-75.0`	-141	-44	-83.03	14.11
rsrq	reference signal receive quality (RSRQ) value in dB	`-7.0`	-20	-3	-11.59	2.25
cell_id	Cell ID	`5123075`				
pci	LTE Physical Cell Identity	`41.0`				
netmode	reported network mode by Huawei HiLink API	`19.0`				
datarate	instantaneous data rate in bit s^{-1}	`2498240.0`	0	153 960 430	40 721 056	31 297 178

Farthofer et al. (2022)

2.2 Analyzing Data

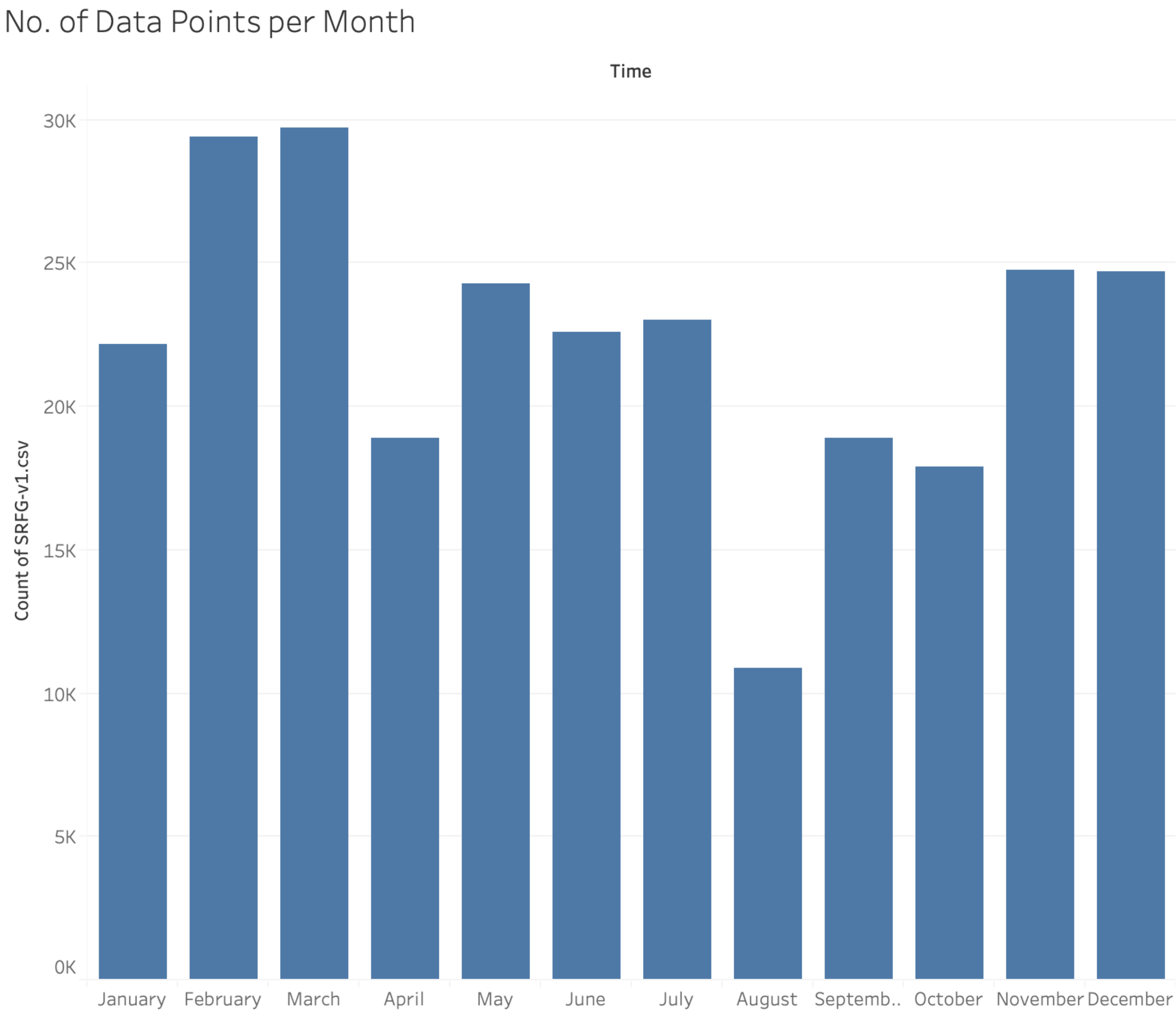
The dataset was analyzed by generating data visualizations, specifically graphs depicting the relationships between the various variables provided in the dataset. For initial data exploration, the typical Python libraries for data analysis were used such as `pandas` and `matplotlib`. For more creating more complex and polished visualizations, the visual analytics platform Tableau was utilized. The Tableau workbook to generate the visualizations shown in this report can be found in this project's submission.

3. Results

3.1 Analysis by Months

To give some context to this data, we first look at the distribution of the data, specifically in terms of the number of data points monthly. Then, the datarate and signal measurements will be analyzed over the months.

3.1.1 Monthly Distribution of Data Points



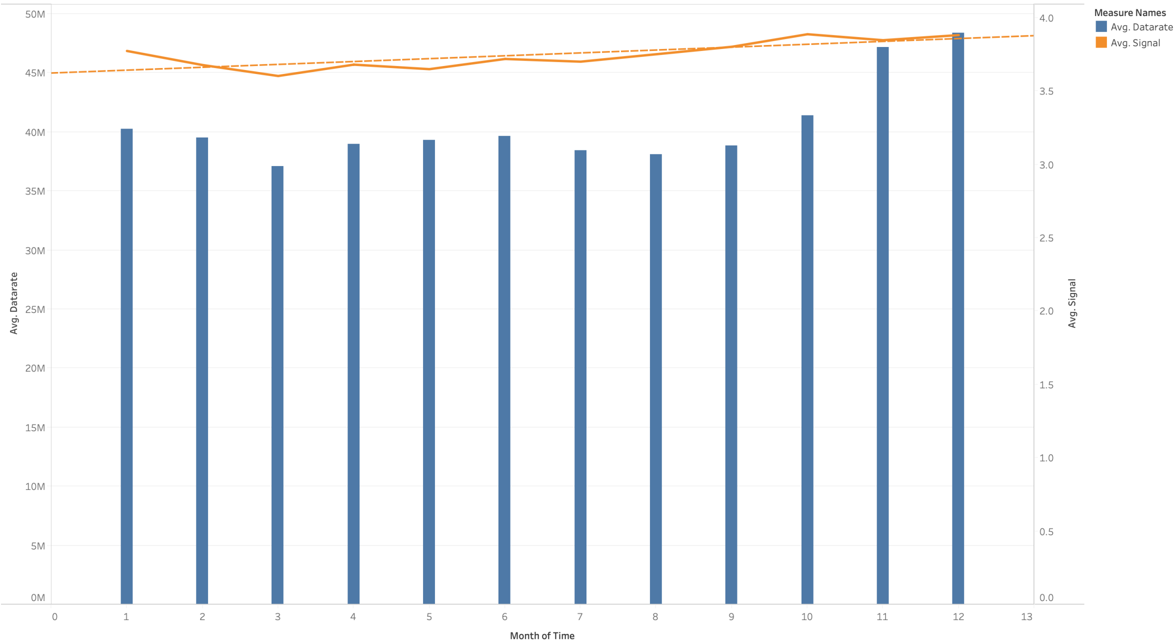
Count of SRFG-v1.csv for each Time Month.

From the above graph, data points are clearly not uniformly distributed over the year, especially during August, one of the summer months. This is most probably due to the absence of **strict measurement scheduling** during the data collection process. Nonetheless, there is at least 10,000 data points for each month across two years, which would make analysis on specific months in the year possible.

Note also that the data collected was from January 2018 to December 2019, which is one recent long-term period when mobile phone usage behavior was **not strongly influenced** by the COVID-19 pandemic. This gives greater confidence in the quality of the dataset because during the pandemic, mobile communications data could show very atypical network traffic patterns owing to significant changes in user movement patterns.

3.1.2 Analysis of Datarate and Signal

Monthly Average Datarate and Average Signal



The trends of Avg. Datarate and Avg. Signal for Time Month. Color shows details about Avg. Datarate and Avg. Signal.

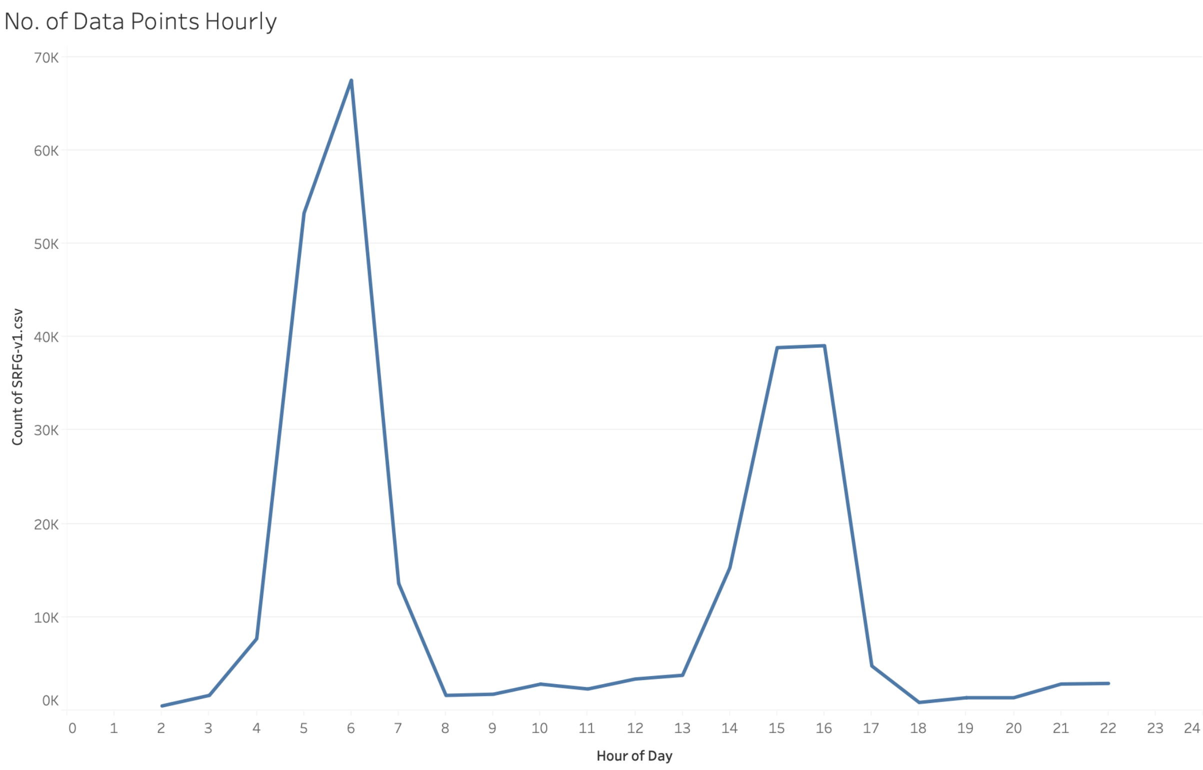
From the above graph, the average datarate was relatively stable from January (month 1) to October (month 10) between 38 to 41MBps, until November and December when the average data rate rose slightly to the region of 47 to 48MBps. As for the average signal recorded, it was relatively constant in the region of 3.6 to 3.8 which does not seem to be a significant difference.

It is difficult to pinpoint the cause of the relatively faster datarate in the final months of the year just based on this dataset alone, but potentially certain environment factors could have come into play here. Moreover, if the highway area had a lower volume of traffic during that period, the network could have experienced a higher bandwith or less noise interfering with the 4G connections, resulting in a slightly higher average datarate during November and December.

That said, the difference in a few megabytes for the datarate could perhaps not be considered not very significant. In the next section, the average datarate and signal measurements will be analyzed on a narrower time interval; hourly.

3.2 Analysis by Hours

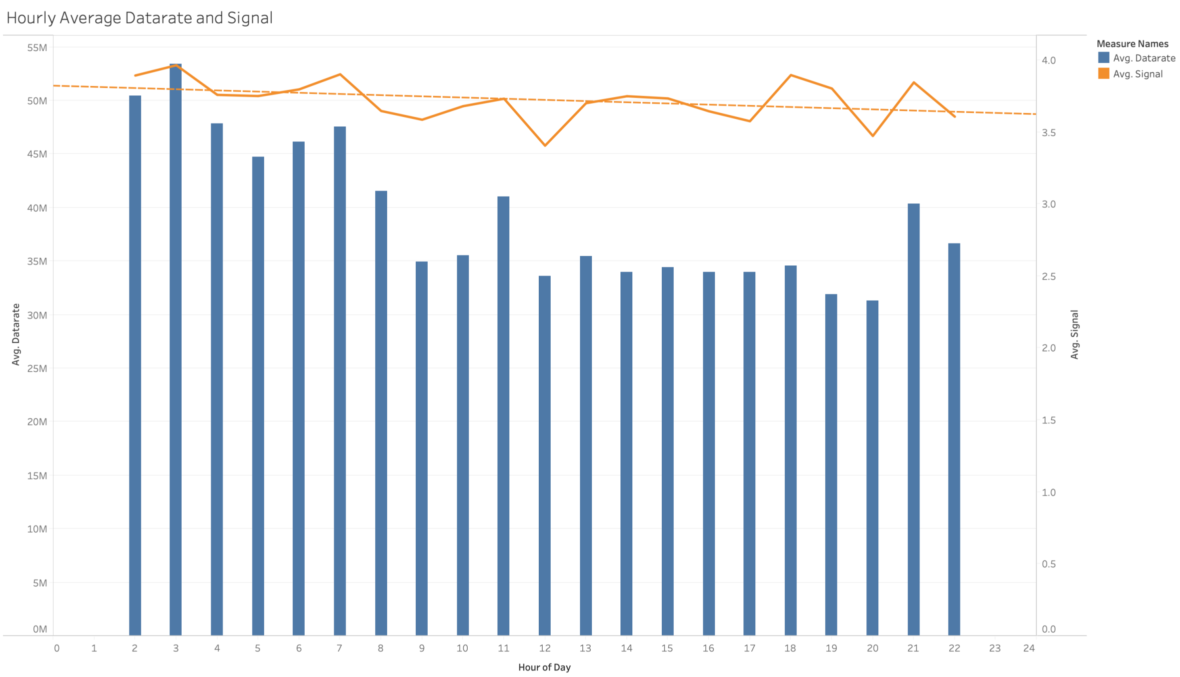
3.2.1 Hourly Distribution of Data Points



The trend of count of SRFG-v1.csv for Time Hour.

The researchers who created this dataset used defined commercial-off-the shelf LTE devices in a specified scenario (continuous maximum throughput in a moving vehicle) in a well-bounded geographical area to reduce the number of variables and to ensure an evenly and densely populated dataset. According to them, the drive tests were often conducted **during rush hour** when high cell loads could be expected. Moreover, due to practical reasons such as the typical office hours of the drivers there was a temporal bias towards rush-hour. Most of the data points were recorded between 5-6 am and 3-4 pm, which perhaps indicates a slightly earlier rush hour period.

3.2.2 Analysis of Datarate and Signal



The trends of Avg. Datarate and Avg. Signal for Time Hour. Color shows details about Avg. Datarate and Avg. Signal.

Similar to 3.1.2, the average signal was relatively constant throughout the hours of a day. However, when the time interval is narrowed from months to hours, the changes in the average datarate are more stark. Notably, the average datarate is significantly higher in the very early hours of the day at around 50MBps between 2 to 4 a.m, and then decreases significantly to around 34MBps between 2 to 6 p.m.

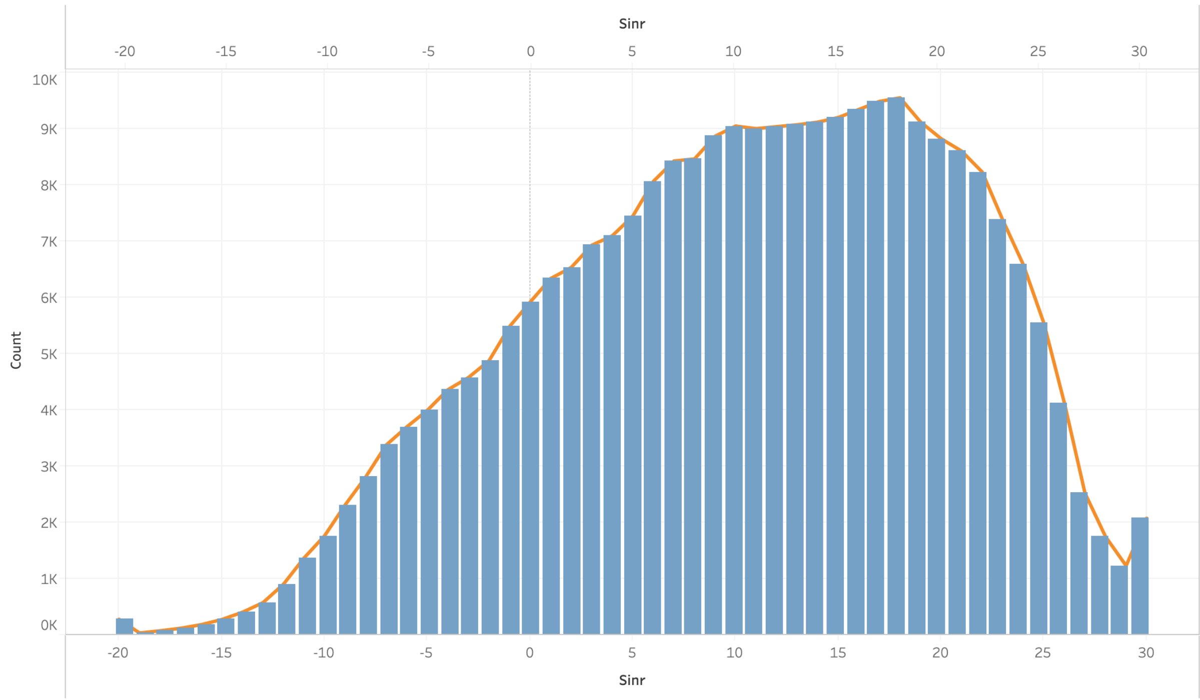
This higher average datarate in the early morning is perhaps understandable, as there would be very little network traffic in the early hours of the day when most people would be asleep. During this time, the 4G connections would be shared across fewer devices and not only result in the network having a greater bandwidth, but fewer devices that might introduce noise to the connection.

3.3 Understanding Trends in Key Fields

3.3.1 SINR Trend

The `SINR` field in this dataset represents the signal-to-interference-plus-noise ratio in DB. SINR (Signal to Interference & Noise Ratio) essentially measures signal quality. Specifically, it is a measurement of the strength of the desired signal compared to the unwanted interference and noise. Mobile network operators seek to maximize SINR at all sites to deliver the best possible customer experience, either by transmitting at a higher power, or by minimizing the interference and noise.

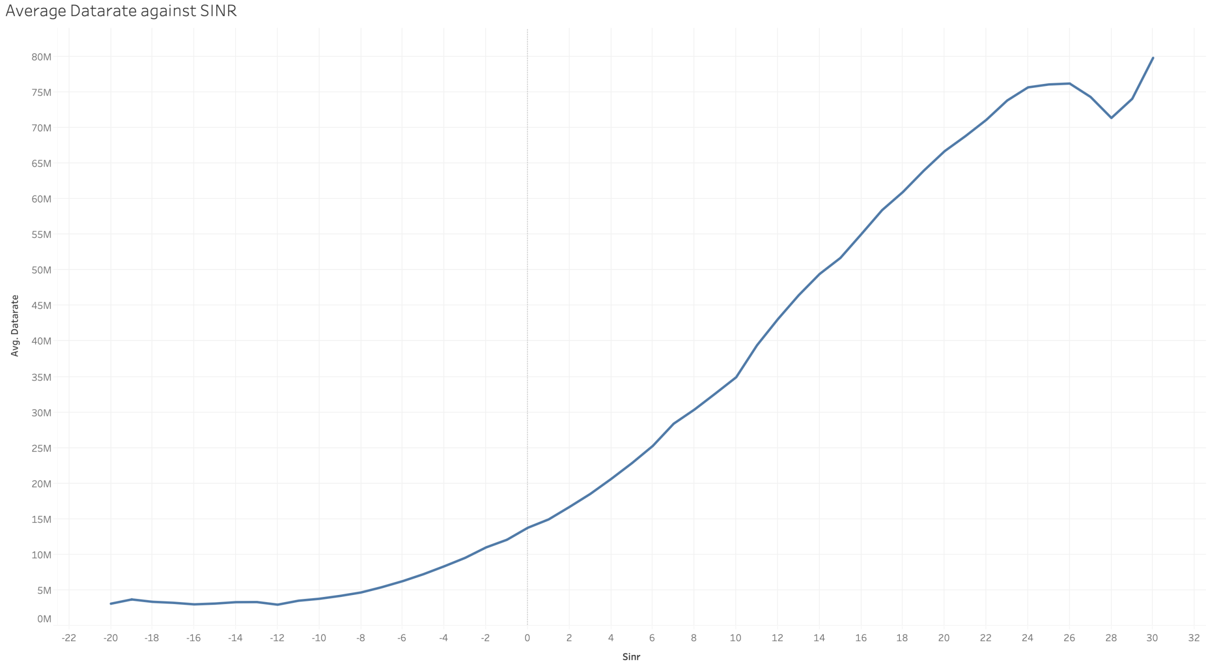
Distribution of SINR



The trends of count of SRFG-v1.csv for Sinr and Sinr.

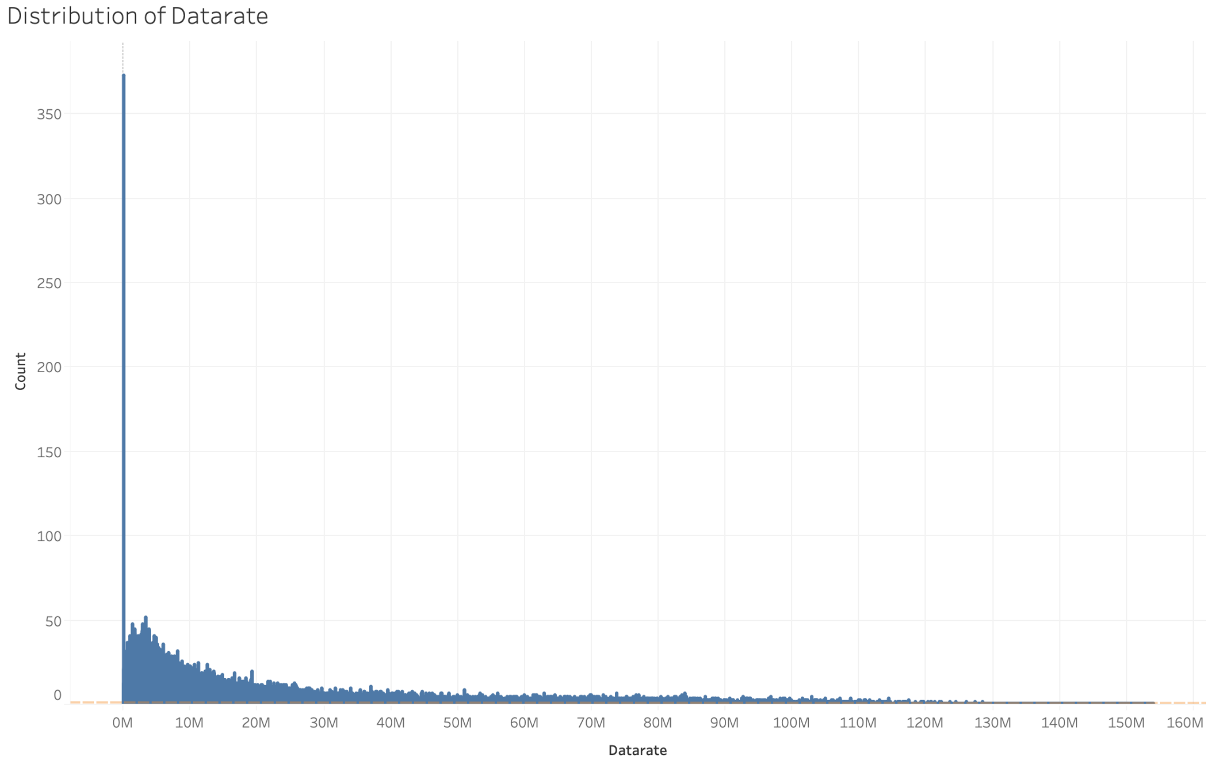
Interestingly from the graph above, the SINR data points follow a Log-normal distribution with mean μ=10.57 and variance σ2=9.96: $SINR \sim N(10.57, 9.96)$. First, the distribution is not symmetrical. Second, it is right skewed. This means that "higher" SINR values are relatively more common, which is something that could be associated with this particular highway location.

The stronger the signal and the less the noise, the higher the speed of data transmission should be. This is depicted in the following graph as expected:



The trend of average of Datarate for Sinr.

3.2.2 Datarate Trend

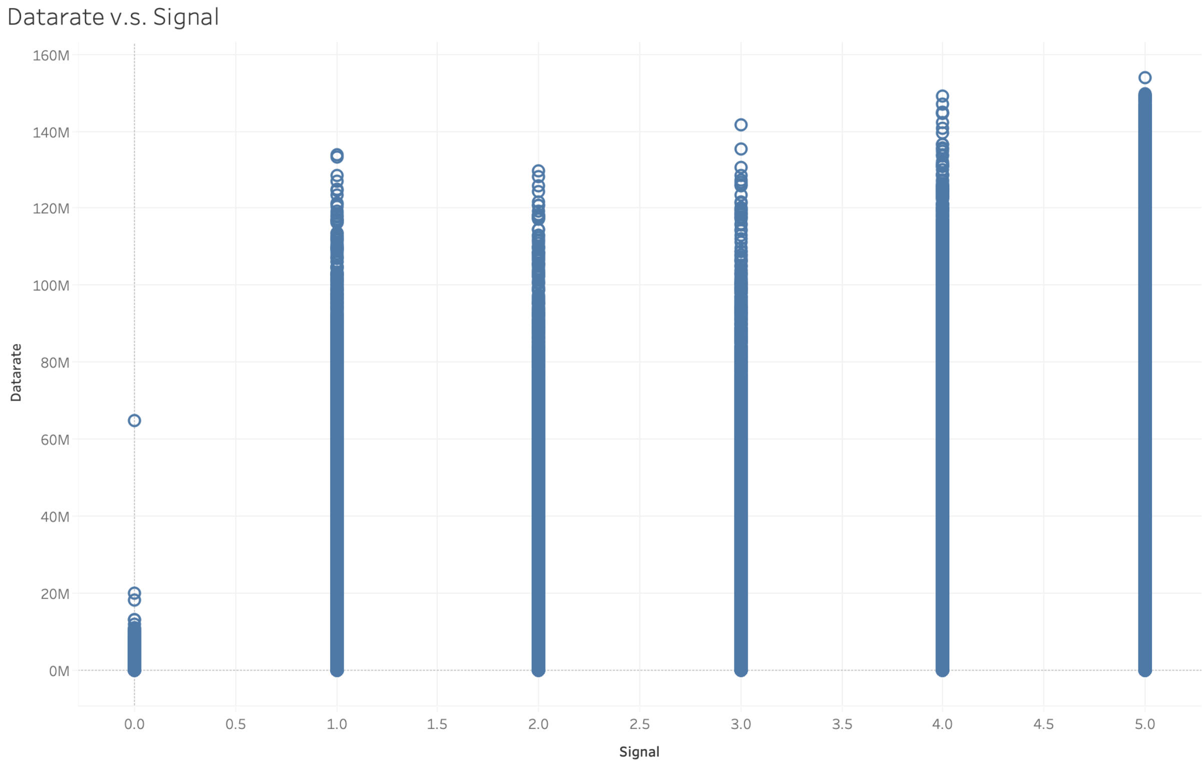


The trend of count of SRFG-v1.csv for Datarate.

The `datarate` field in this dataset is the instantaneous data rate in bit/s. The download data rate is influenced by a variety of factors, such as physical parameters (this includes the received power or noise power), cell load (since the resources are shared between multiple users), and configuration (e.g. tariff limits and network planning). In the data set, the distribution of the download data rate seems to resemble an exponential distribution.

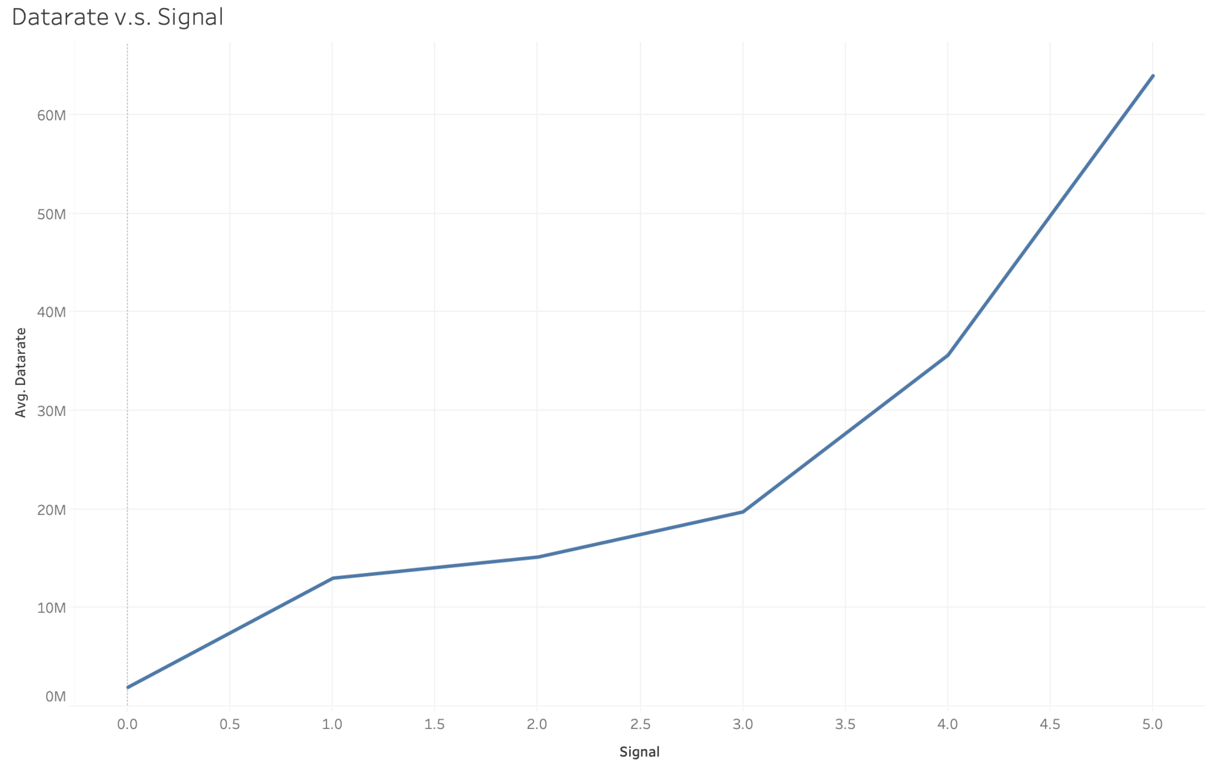
This would correspond with theoretical models of wireless communication, where data rates are often exponentially distributed. Here, as noted by the researchers who created this dataset, the maximum observed download data rate of 150MBps is actually due to a tariff limit and does not change over the measurement time interval of 24 months.

3.3 Datarate v.s. Signal



Signal vs. Datarate.

The above graph depicts the distribution of data points for datarate and signal. Interestingly, there is one significant anomaly where the signal is 0 but the datarate is relatively high at 65MBps. However, besides that, the graph shows that at pretty much any signal level, the datarate could vary from 0MBps to the maximum datarate. Changing the y-axis to show the average datarate instead, we see the following trend:



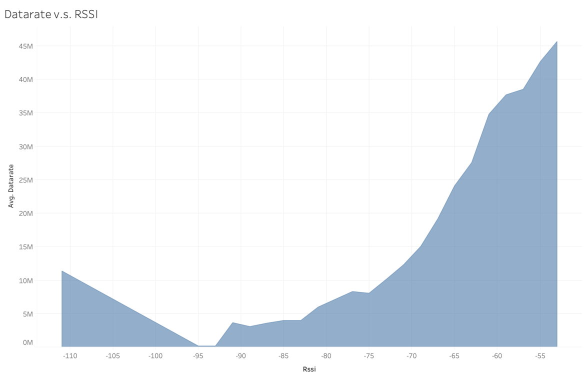
The trend of average of Datarate for Signal.

In general, a higher signal strength is associated with a higher datarate, as shown in the upward trend in the graph above. However, it is important to acknowledge that this may not always be the case as there may be a few exceptions, or assumptions at play such as an optimal physical layer data rate selection algorithm.

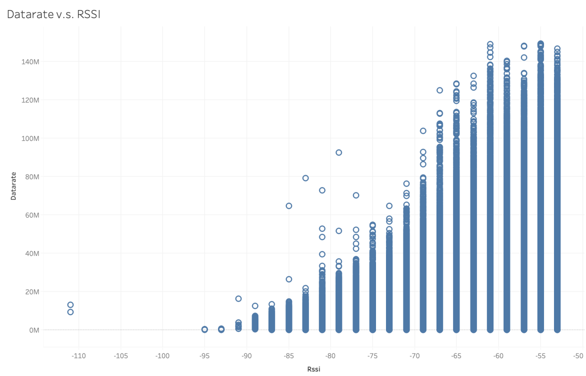
Typically, datarate may increase as signal strength increases due to higher signal strengths enable the use of higher PHY (PHYsical layer data) rates, also known as MCS (Modulation and Coding Scheme). Once there is sufficient signal strength to operate reliably in the maximum supported MCS rate, additional signal strength does not produce additional benefits in terms of the data transmision rate. Factors impacting this relationship between datarate and signal could include the specific capabilities of the transmitting device, the receiving device as well as the environment.

3.4 Datarate v.s. RSSI

`RSSI`, or “Received Signal Strength Indicator,” is a measurement of how well a device can hear a signal from an access point or router. It’s a value that is useful for determining if a device has enough signal to get a good wireless connection. RSSI is a term used to measure the relative **quality** of a received signal to a client device, but has **no absolute value**. This dataset seems to utilize a more standardized, absolute measure of signal strength: received signal power, which is measured in decibels on a logarithmic scale.



The plot of average of Datarate for Rssi.

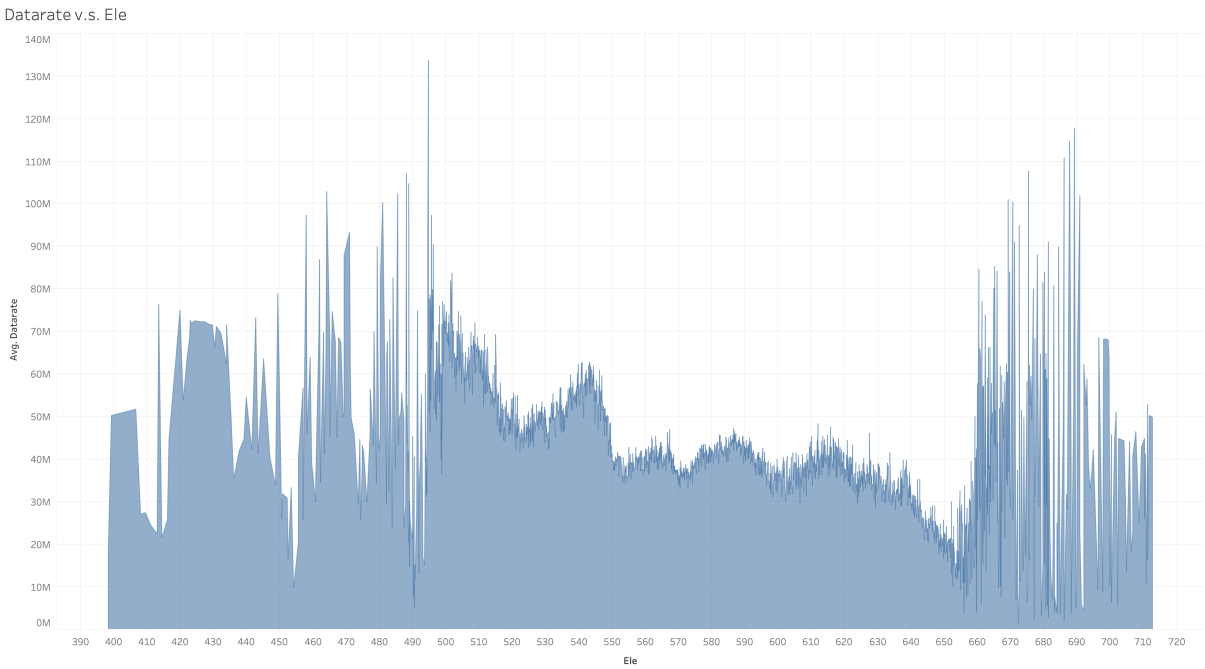


Rssi vs. Datarate.

The left graph above shows the relationship between data rate and RSSI, which is a positive one where datarate generally increases as RSSI increases. This makes intuitive sense, since the better the quality of signal, the faster the transmission of data should be. While it may seem that there is a downward trend in the left graph at the section of very low RSSI measurements, that could be explained by the anomalies seen in the graph on the right above, where there are a couple of data points that seem to skew the graph of average datarate and RSSI. To keep the analysis more focused, those anomalies could be ignored.

3.5 Signal / Datarate v.s. Elevation

The `Ele` field represents the elevation in metres. The following graph shows how average datarate changes as the elevation increases:



The plot of average of Datarate for Ele.

Intiuitively one might expect higher elevations to result in higher signal quality and hence datarate, but from the graph above, this does not seem to hold true and elevation does not appear to play a major role in affecting datarate since there is no obvious positive or negative relationship.

However, this statement needs to be qualified. A crucial factor here is that the elevation measurements taken here are from a highway where elevation might naturally not vary very significantly. Thus, the particular range of elevation of 398-712m in this context, may be insufficient for any relationship between elevation and datarate to be observed.

References

- https://www.salzburgresearch.at/wp-content/uploads/2022/10/An_Open_Mobile_Communications_Drive_Test_Data_Set_and_Its_Use_for_Machine_Learning.p
- https://www.commscope.com/blog/2015/understanding-wi-fi-signal-strength-vs.-wi-fi-speed/
- https://iscointl.com/sinr-optimization/
- https://www.metageek.com/training/resources/understanding-rssi/

Ideas

1. **Factors impacting 4G download speeds**

dataset: https://crawdad.org/srfg/lte-4g-highway-drive-tests-salzburg/2022-01-18/