

Data Science Final Project

# Analisis Pengaruh Network Type terhadap Performance Jaringan Seluler

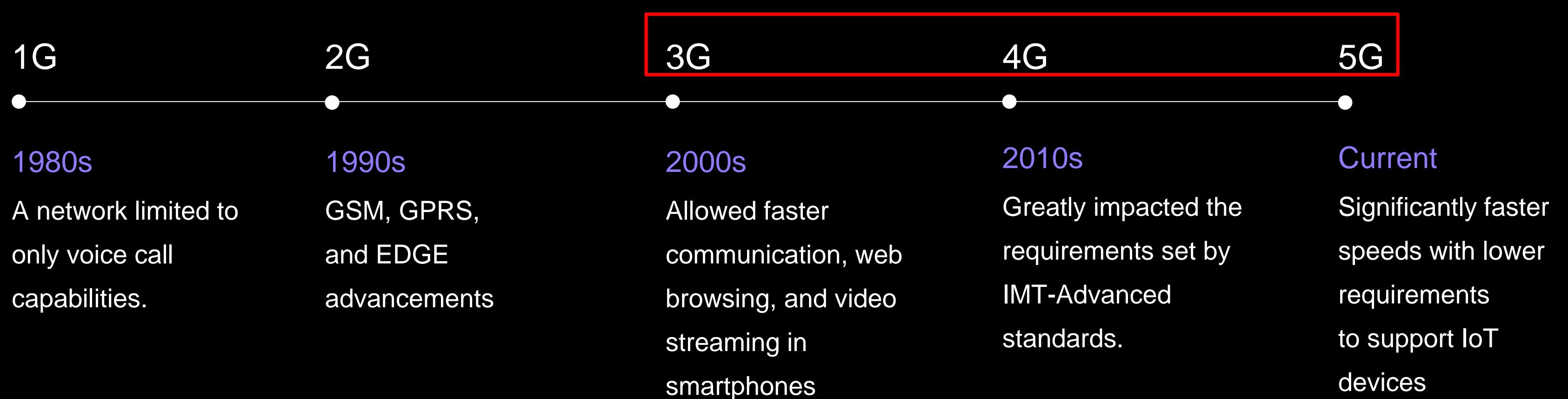
Sumber:

<https://www.kaggle.com/datasets/suraj520/cellular-network-analysis-dataset>

# Pendahuluan

- Nama Dataset adalah "signal\_metrics.csv" dari Kaggle, terdiri atas 16829 baris & 12 kolom.
- Dataset ini memberikan sampel pengukuran sinyal untuk berbagai jenis jaringan di 20 lokasi di Bihar, India yang di ukur oleh 4 device (DragonOS Focal Steam Deck + Spike LTE Analysis Data(BB60C, srsRan, BladeRFxA9) ).
- Dalam final project kali ini, dataset ini akan digunakan untuk menganalisis pengaruh Network type terhadap performance jaringan.

# Mobile Networks and 5G Technology



# Metodologi Analisis

1. Data Understanding & Data Manipulation

---

2. Data Visualisasi

---

3. Multicollinearity Study

---

4. Build Machine Learning & Model Evaluation

---

5. EDA Analysis

---

6. Insight & Recommendation

---

# Data Understanding

RangeIndex: 16829 entries, 0 to 16828

Data columns (total 12 columns):

#	Column	Non-Null Count	Dtype
0	Timestamp	16829	non-null object
1	Locality	16829	non-null object
2	Latitude	16829	non-null float64
3	Longitude	16829	non-null float64
4	Signal Strength (dBm)	16829	non-null float64
5	Signal Quality (%)	16829	non-null int64
6	Data Throughput (Mbps)	16829	non-null float64
7	Latency (ms)	16829	non-null float64
8	Network Type	16829	non-null object
9	BB60C Measurement (dBm)	16829	non-null float64
10	srsRAN Measurement (dBm)	16829	non-null float64
11	BladeRFxA9 Measurement (dBm)	16829	non-null float64
dtypes: float64(8), int64(1), object(3)			

- Terdiri dari 9 Kolom numerik & 3 kolom kategorik
- Tidak ditemukan missing value dan duplikat data,
- Ditemukan Nilai 0 di semua kolom Signal Quality dan di beberapa baris pada kolom BB60C Measurement, srsRAN Measurement, BladeRFxA9 Measurement,
- Terdapat 4 kolom yang memiliki satuan dBm, menggambarkan kekuatan sinyal dari 4 device yang diukur,
- Terdapat masing-masing 1 kolom yaitu Network Type, Data Throughput, dan latency mewakili pengukuran di salah satu device yaitu Valve Steam Deck.

	Latitude	Longitude	Signal Strength (dBm)	Signal Quality (%)	Data Throughput (Mbps)	Latency (ms)	BB60C Measurement (dBm)	srsRAN Measurement (dBm)	BladeRFxA9 Measurement (dBm)
count	16829.000000	16829.000000	16829.000000	16829.0	16829.000000	16829.000000	16829.000000	16829.000000	16829.000000
mean	25.594796	85.137314	-90.072484	0.0	16.182856	101.313624	-68.820150	-74.439562	-68.819930
std	0.089881	0.090095	5.399368	0.0	25.702734	56.010418	40.046739	43.215204	39.996934
min	25.414575	84.957936	-116.942267	0.0	1.000423	10.019527	-115.667514	-124.652054	-119.207545
25%	25.522858	85.064124	-93.615962	0.0	2.001749	50.320775	-94.021959	-101.249987	-93.749032
50%	25.595383	85.138149	-89.665566	0.0	2.997175	100.264318	-89.126942	-96.838442	-89.282746
75%	25.667620	85.209504	-86.145491	0.0	9.956314	149.951112	0.000000	0.000000	0.000000
max	25.773648	85.316994	-74.644848	0.0	99.985831	199.991081	0.000000	0.000000	0.000000

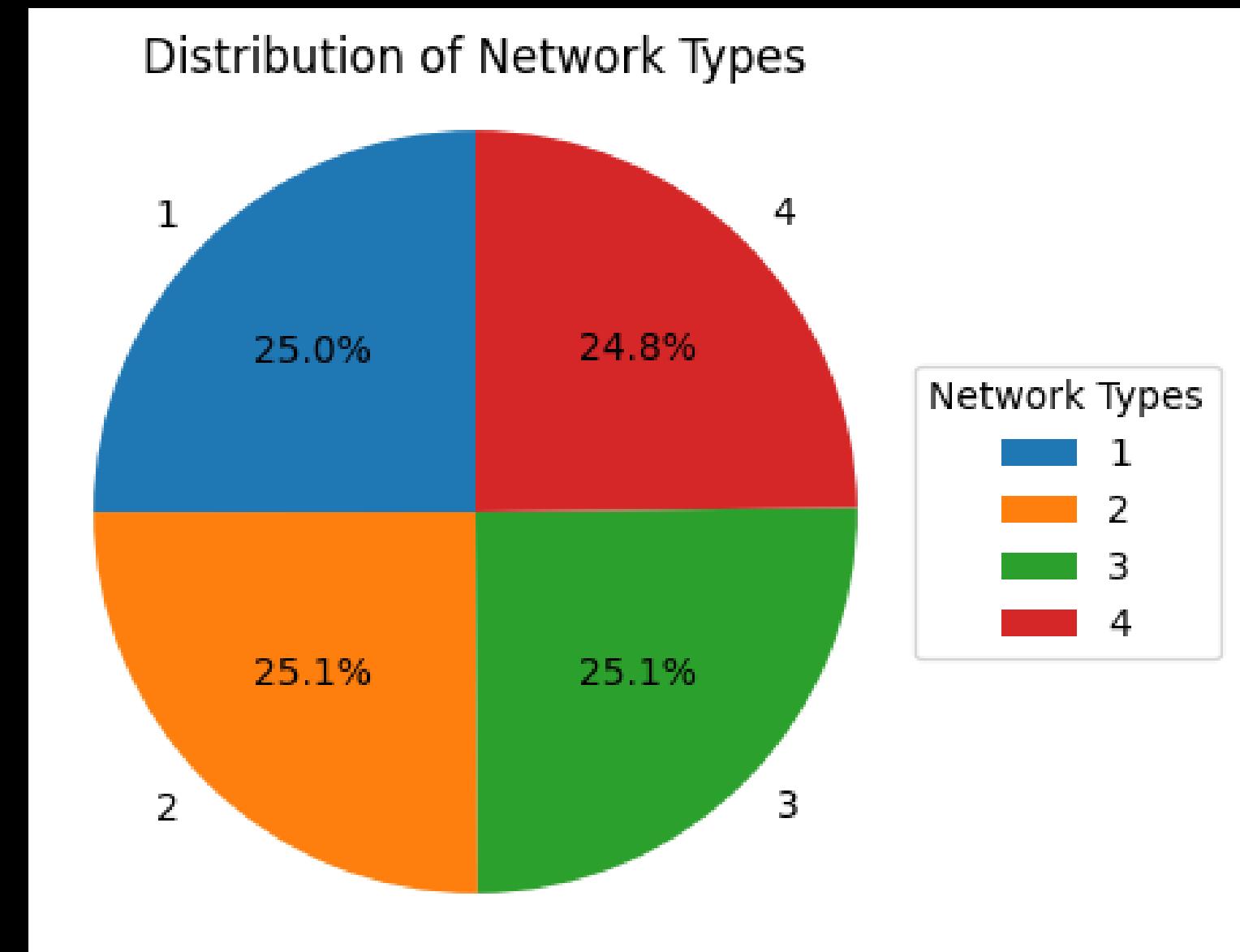
# Data Manipulation

- Drop Kolom Signal Quality (%).
- Replace nilai 0 to median di kolom BB60C Measurement, srsRAN Measurement, BladeRFxA9 Measurement.
- Drop Kolom-kolom yang tidak di gunakan didalam analysis (Timestamp, Locality, Latitude, Longitude).
- Membuat encoding/labeling kolom Network Type berdasarkan urutan Tingkat technology-nya. (3G:1, LTE:2, 4G:3, 5G:4)

	Signal Strength (dBm)	Data Throughput (Mbps)	Latency (ms)	BB60C Measurement (dBm)	srsRAN Measurement (dBm)	BladeRFxA9 Measurement (dBm)	network_encoded
0	-84.274113	1.863890	129.122914	-89.126942	-96.838442	-89.282746	1
1	-97.653121	5.132296	54.883606	-95.810791	-105.452359	-99.920892	3
2	-87.046134	1.176985	119.598286	-91.593861	-95.419482	-87.714070	2
3	-94.143159	68.596932	46.598387	-90.642773	-101.895905	-96.570698	4
4	-94.564765	38.292038	30.342828	-90.489100	-103.318303	-95.102467	4
...	...	...	...	...	...	...	...
16824	-92.024221	3.720856	53.376474	-90.177947	-97.971619	-93.362041	3
16825	-87.579509	2.117080	134.458401	-87.694985	-92.616868	-84.749607	2
16826	-90.451396	72.870842	32.556578	-91.098875	-97.447725	-87.752628	4
16827	-85.661814	2.482843	144.007572	-89.126942	-96.838442	-89.282746	1
16828	-93.228967	1.638291	123.234570	-95.598301	-99.438645	-94.268015	2

16829 rows × 7 columns

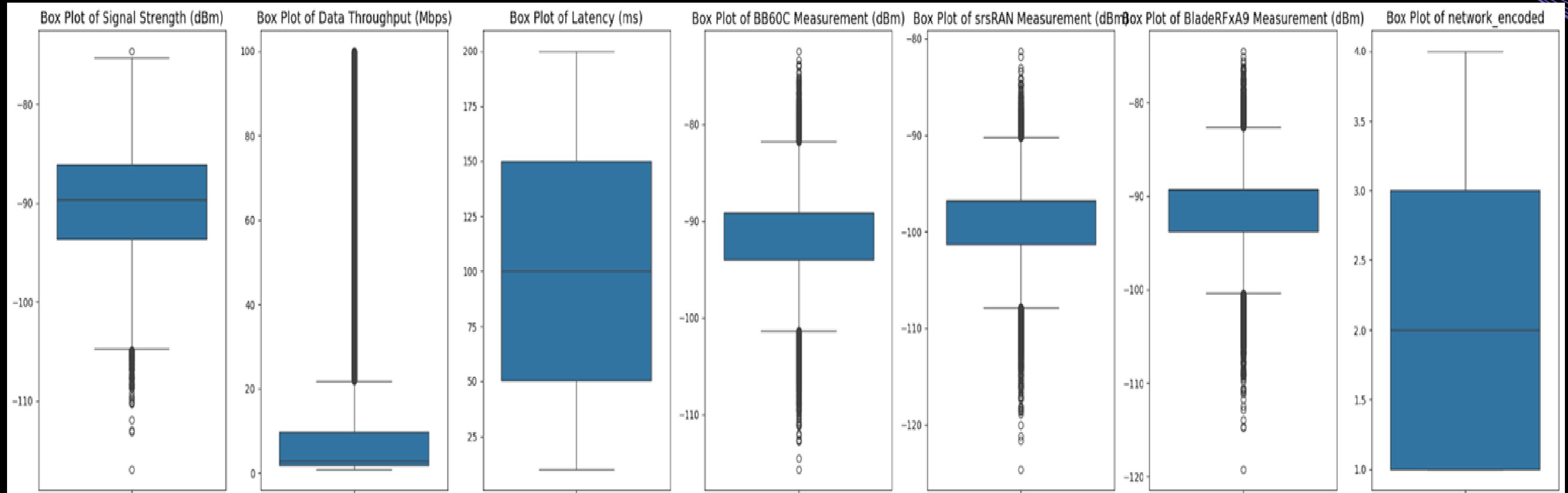
# Data Distribusi Network Type



Keterangan:

Distribusi Network Types masih balance

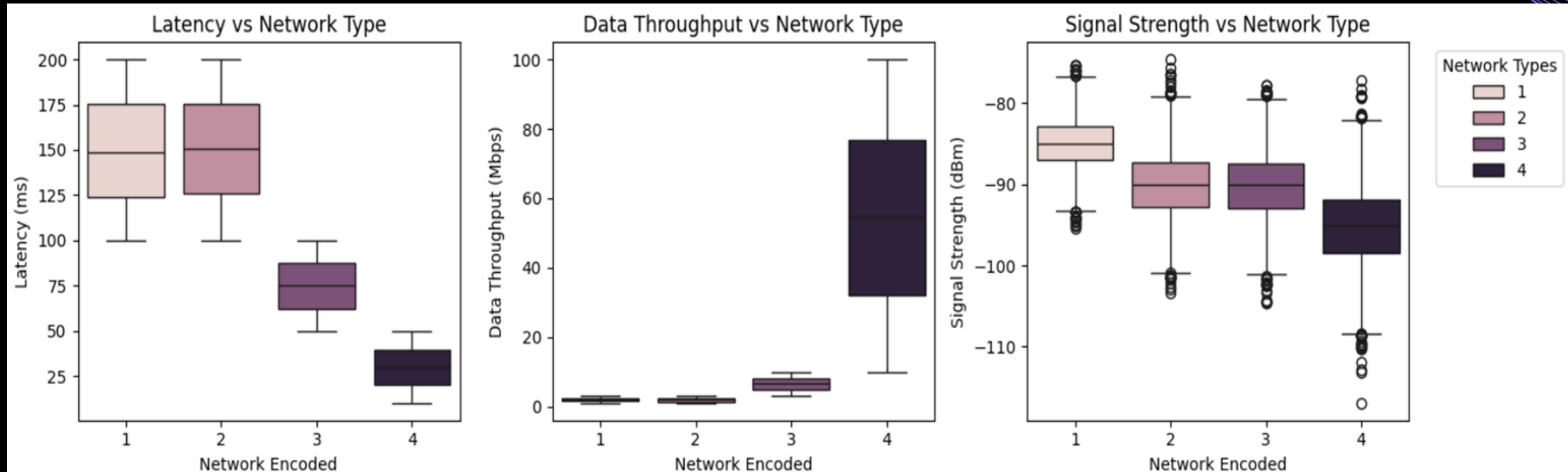
# Data Outlier



Keterangan:

Dari hasil visualisasi dengan boxplot, di temukan bahwa ada outlier di semua kolom, kecuali kolom latency (ms), selanjutnya akan coba di break down masing-masing kelas dari network\_encoded.

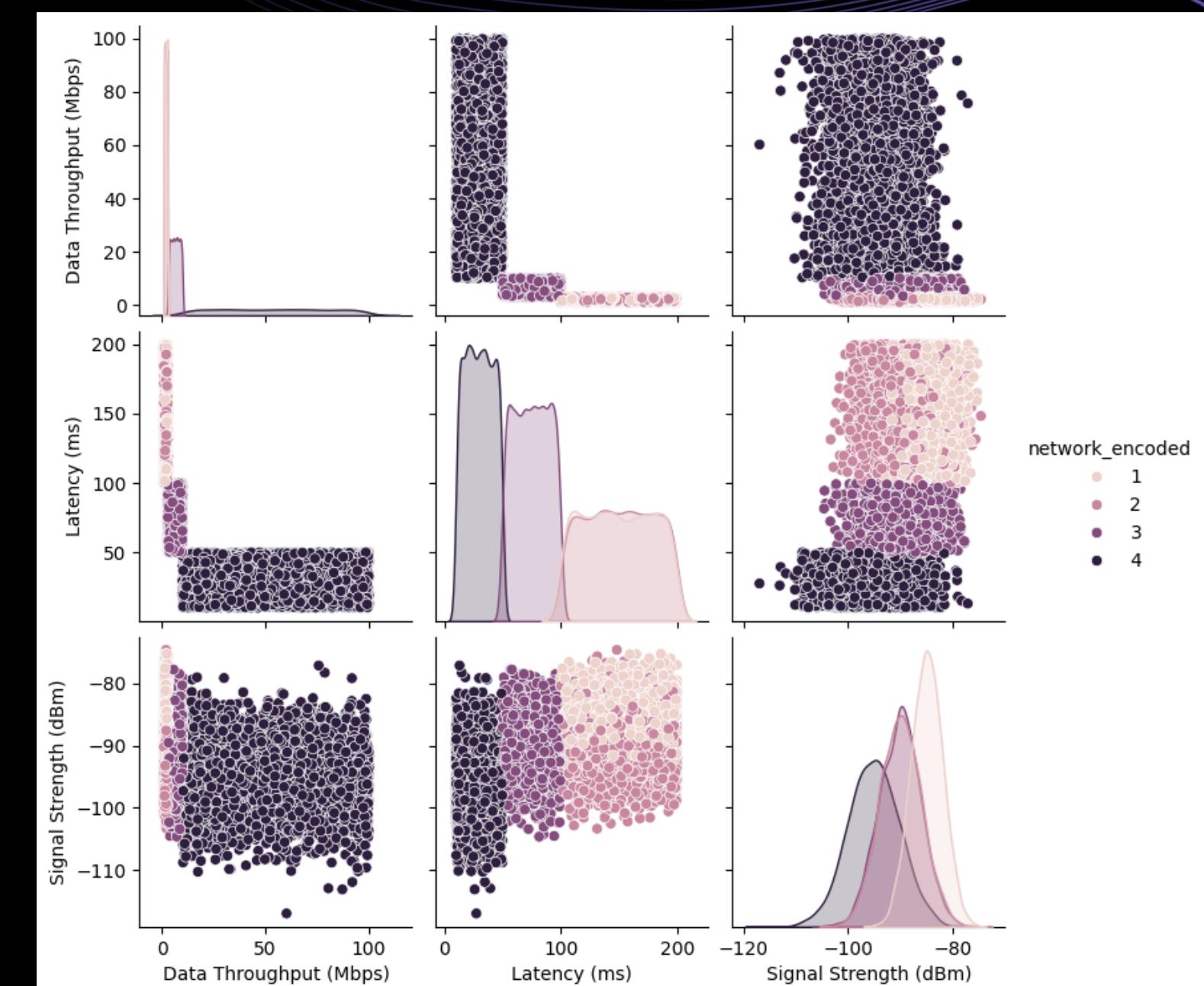
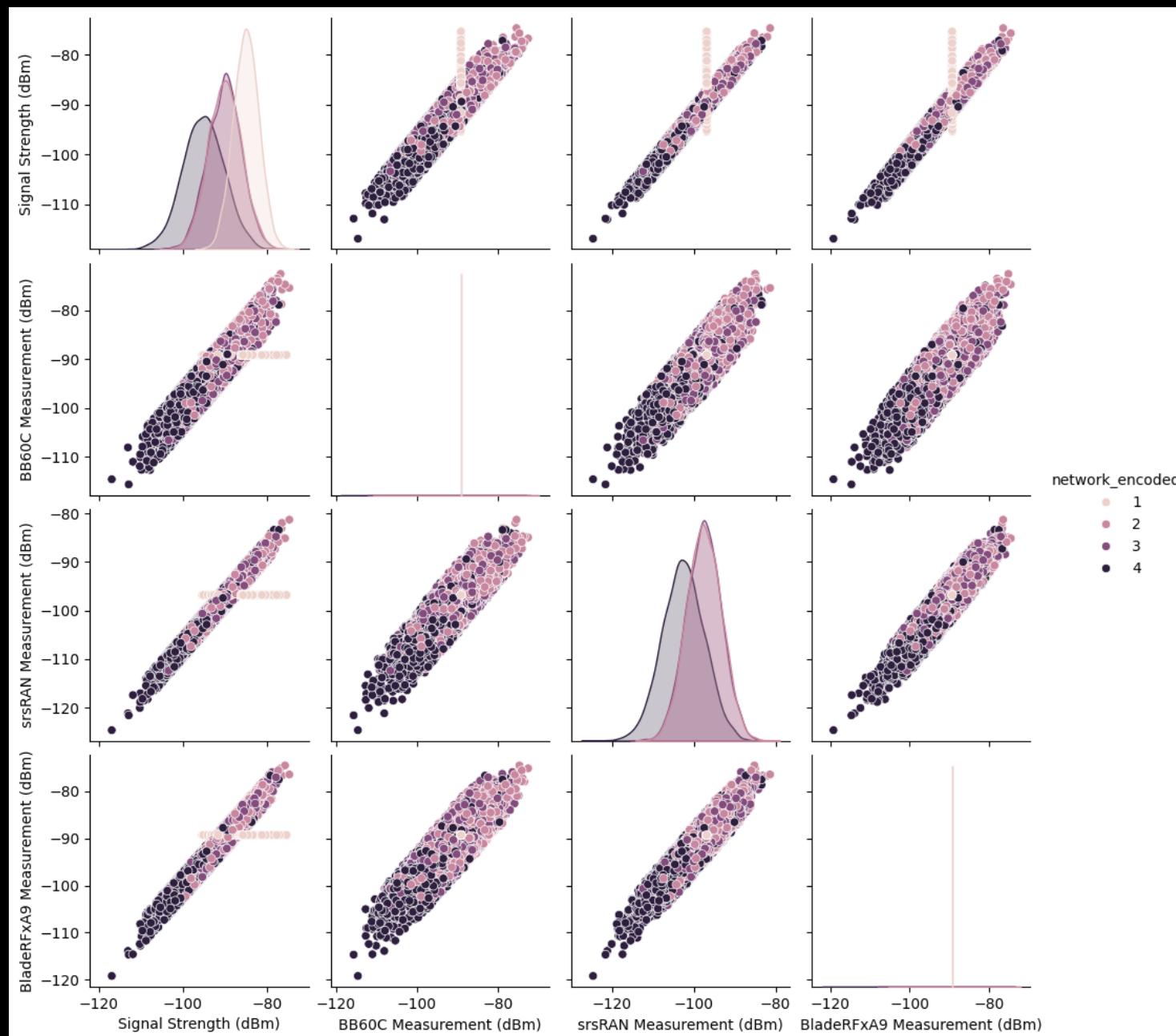
# Latency/ Data Throughput/ Signal Strength vs Network Type



Keterangan:

Setelah di break down berdasarkan network encoded nya, terlihat bahwa masing-masing class dari network\_encoded memiliki profile yang cukup berbeda significant satu dengan yang lain yang dapat menimbulkan nilai yang outlier, sehingga untuk data outlier tetap di biarkan/tidak di imputasi karena di anggap wajar.

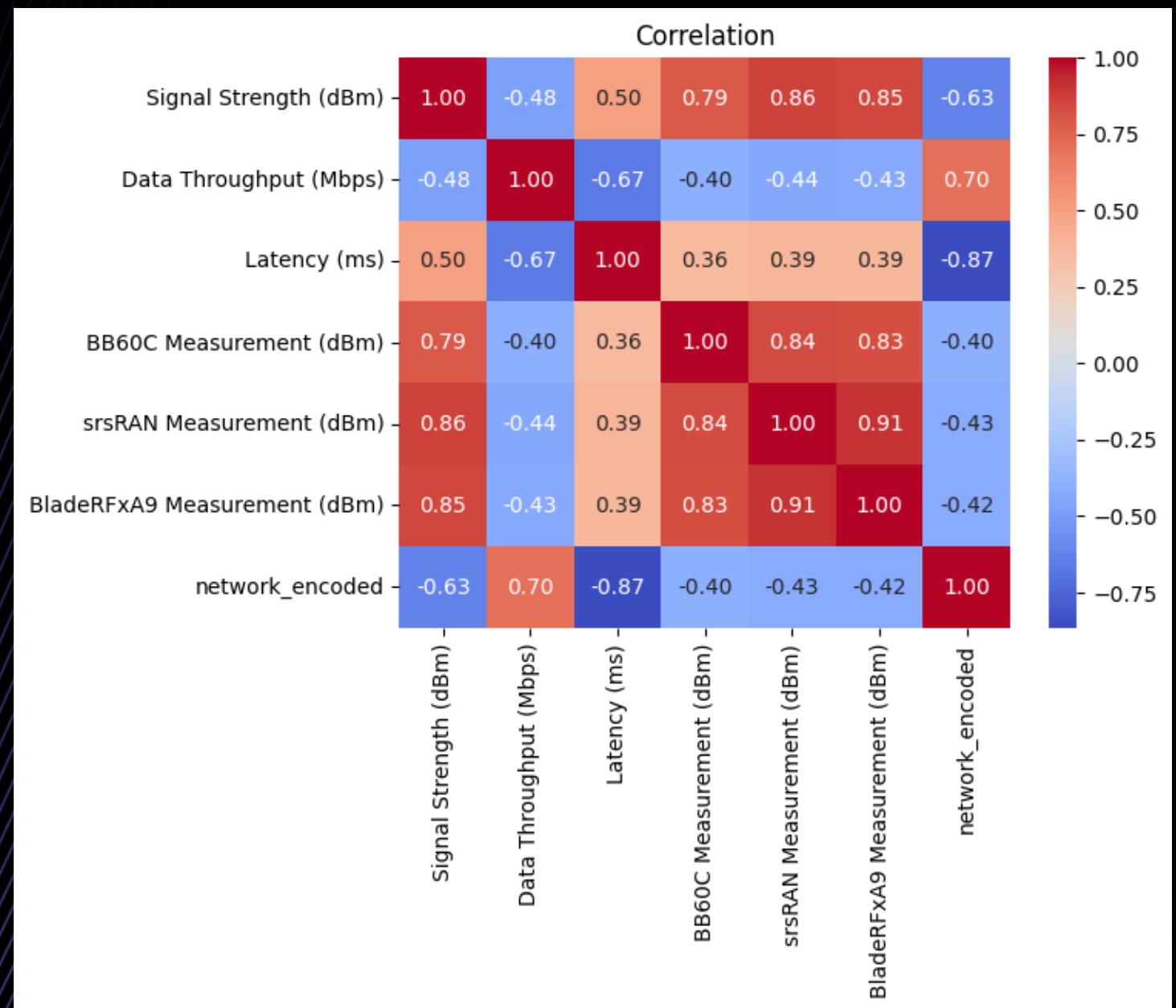
# Korelasi antar Feature



Keterangan:

- Pair plot kiri: terlihat korelasi yang saling berbanding lurus di antara kolom dan memiliki pola distribusi yang terlihat mirip.
- Pair Plot Kanan:
  - Data throughput tetap lebih tinggi di network type 5G, walaupun signal strength nya yang diterima makin minus/ lemah.
  - Latency tetap lebih rendah di network type 5G, walaupun signal strength nya yang diterima makin minus/ lemah.
  - Data Throughput berbanding terbalik dengan Latency

# Multicollinearity Study



	feature	VIF
0	Latency (ms)	2.016844
1	Data Throughput (Mbps)	1.945999
2	Signal Strength (dBm)	5.150204
3	BB60C Measurement (dBm)	3.720075
4	srsRAN Measurement (dBm)	7.607111
5	BladeRFxA9 Measurement (dBm)	6.819045

- Kolom-kolom yang menggambarkan kekuatan sinyal, terlihat memiliki korelasi yang sangat kuat satu sama lain, sehingga akan dilakukan multicollinearity study.
- Network encoded memiliki korelasi yang kuat dengan Latency (ms) dan Data Throughput (Mbps).

Dari VIF Score terlihat ada 3 fitur yang memiliki korelasi yang kuat (VIF Score >4). Dan di putuskan untuk mempertahankan Signal Strength (dBm) dan drop 2 kolom lainnya.

# Build Machine Learning & Model Evaluation

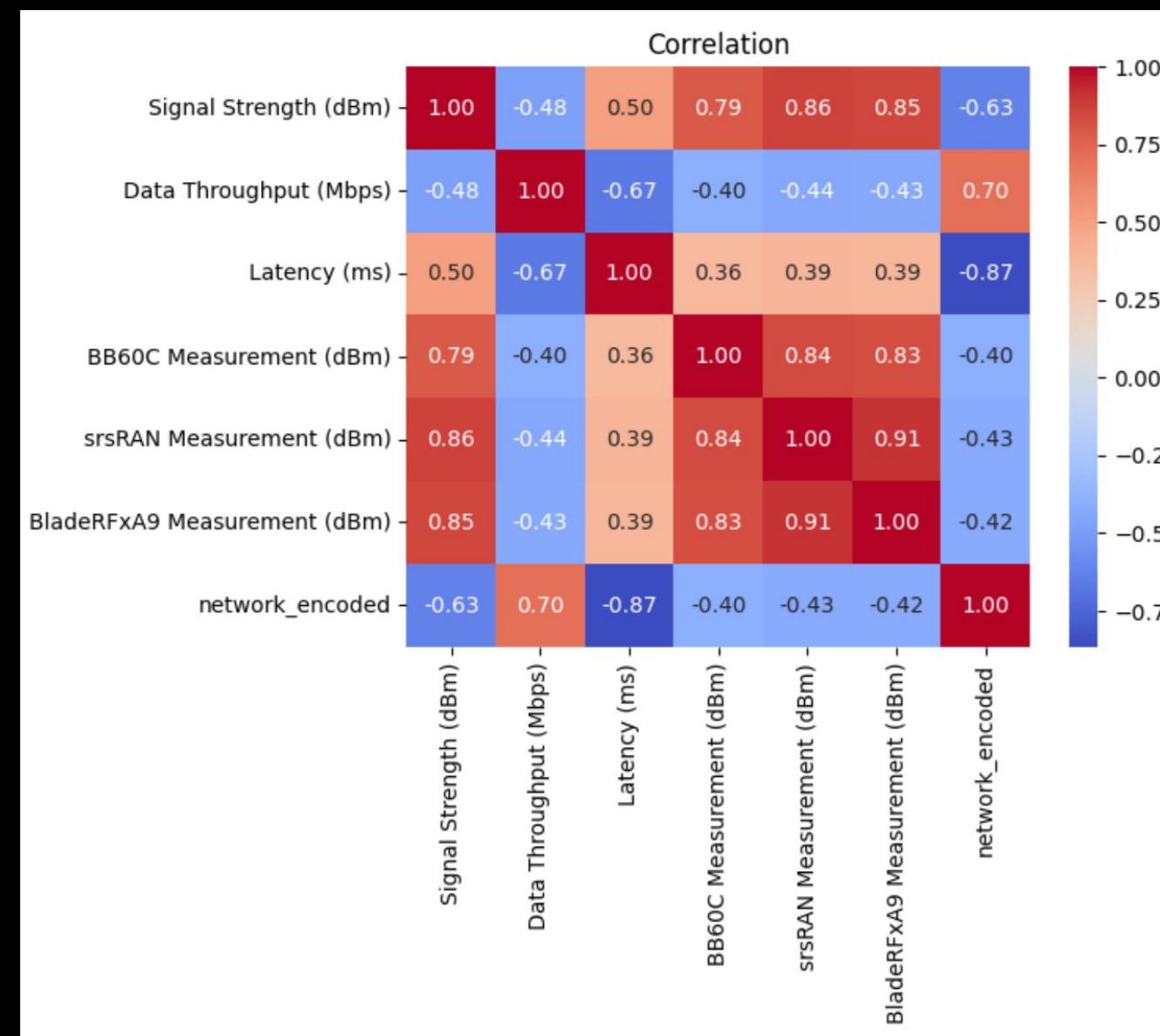
Langkah-Langkah yang sudah dilakukan didalam membangun Machine learning sebagai berikut:

- Split data for features and target variable,
- Dataset Splitting into Training and Testing Sets + Feature Scaling,
- Model Building,
- Model Evaluation

ML Model	Class	Classification Report			
		Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Random Forest	1	99.97	100	100	100
	2		100	100	100
	3		100	100	100
	4		100	100	100
KNN Classifier	1	98%	93	99	96
	2		99	93	96
	3		99	100	99
	4		100	99	99
Logistic Regression	1	89%	79	77	78
	2		78	80	79
	3		99	100	99
	4		100	99	100

# EDA Analysis

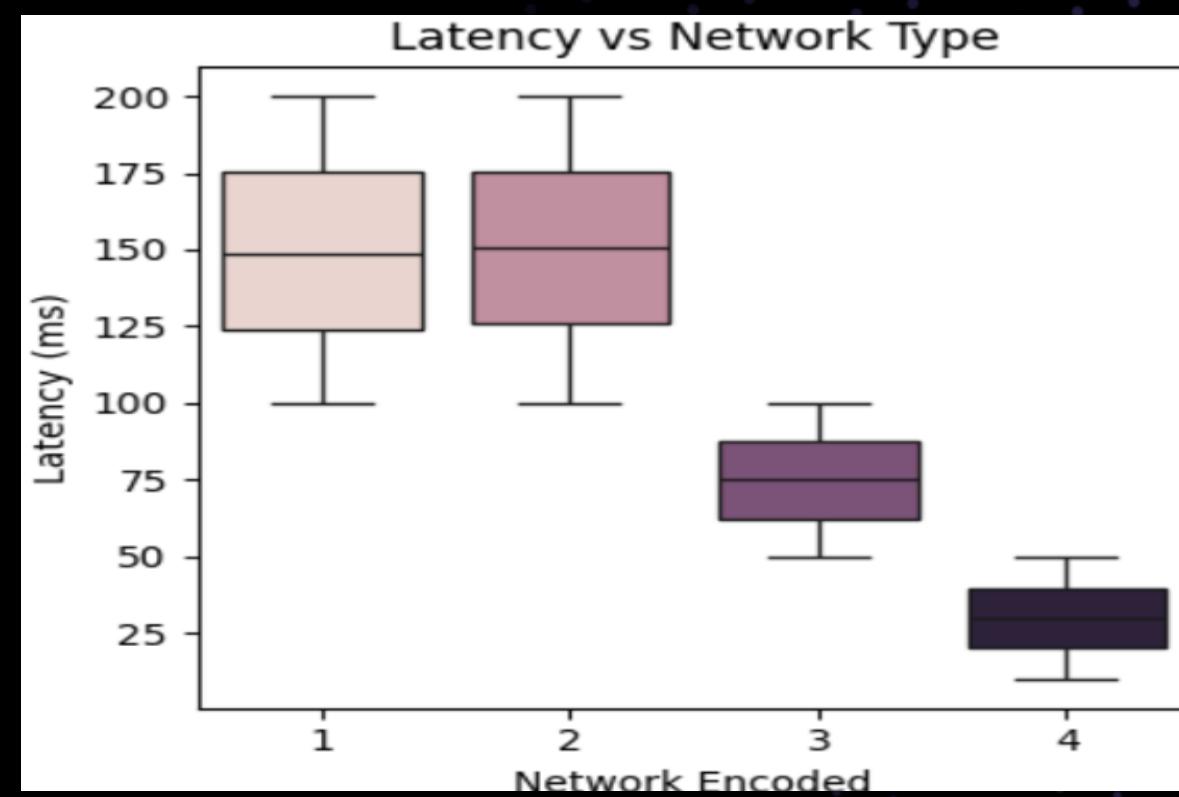
- Fitur yang memiliki korelasi terkuat dengan target adalah Latency (ms), sedangkan fitur yang mempunyai korelasi terendah adalah BB60C Measurement (dBm).



# EDA Analysis

2. Berdasarkan rata-rata latensi untuk setiap jenis jaringan:

- **Jaringan dengan Latensi Terendah:** 5G, dengan rata-rata latensi sekitar 29.86 ms.
- **Jaringan dengan Latensi Tertinggi:** LTE, dengan rata-rata latensi sekitar 150.50 ms, walaupun mirip dengan 3G sekitar 149.4ms



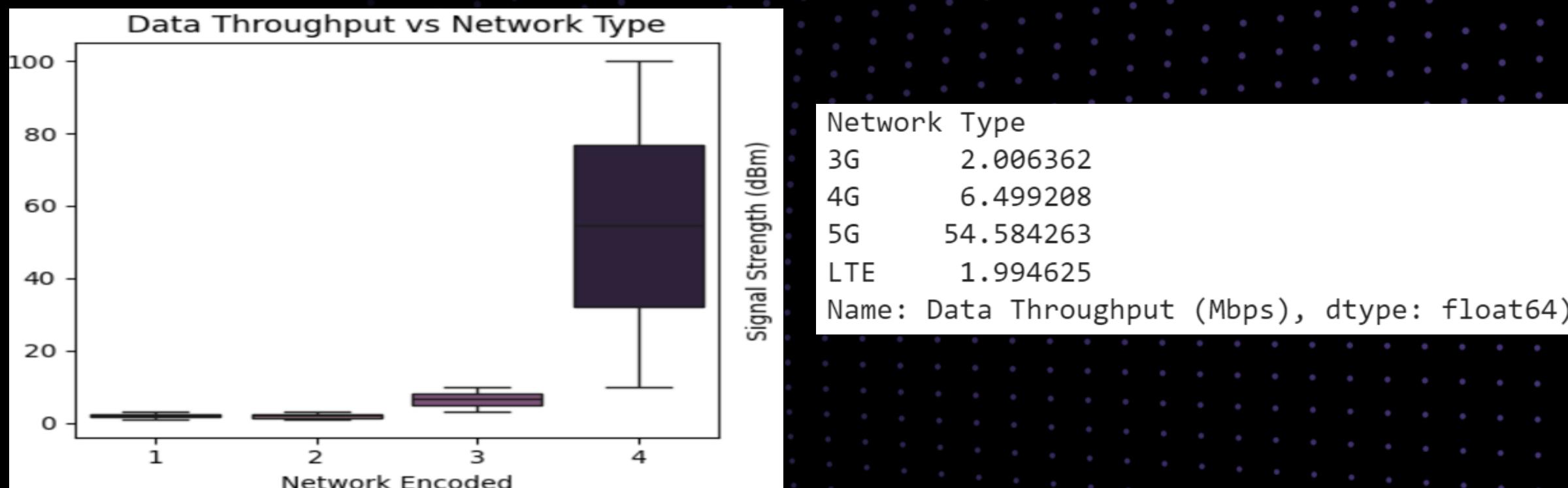
Network Type	Mean Latency (ms)
3G	149.414617
4G	74.846400
5G	29.861568
LTE	150.504659

Name: Latency (ms), dtype: float64)

# EDA Analysis

3. Berdasarkan rata-rata throughput data untuk setiap jenis jaringan:

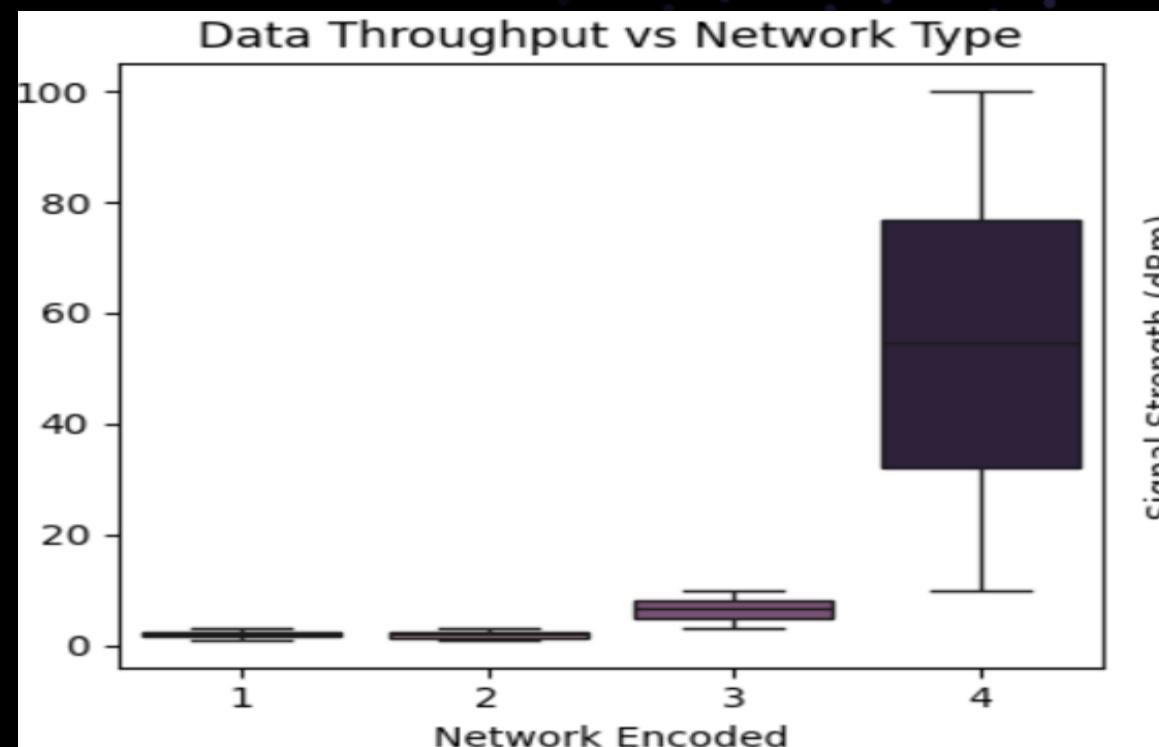
- **Jaringan dengan Throughput Data Terendah:** LTE, dengan rata-rata throughput data sekitar 1.99 Mbps, tidak jauh berbeda dengan 3G, yaitu 2 Mbps.
- **Jaringan dengan Throughput Data Tertinggi:** 5G, dengan rata-rata throughput data sekitar 54.584 Mbps



# EDA Analysis

4. Berdasarkan rata-rata signal strength untuk setiap jenis jaringan:

- **Jaringan dengan Kekuatan Sinyal Terendah:** 5G, dengan rata-rata kekuatan sinyal sekitar -95.15 dBm.
- **Jaringan dengan Kekuatan Sinyal Tertinggi:** 3G, dengan rata-rata kekuatan sinyal sekitar -85.01 dBm.



```
Network Type
3G    -85.010978
4G    -90.169674
5G    -95.150218
LTE   -89.995308
Name: Signal Strength (dBm), dtype: float64)
```

# EDA Analysis

5. Berdasarkan korelasi antara Data Throughput, Latency, dan Signal Strength:

- **Data Throughput dan Latency:** Ada korelasi negatif yang cukup kuat (-0.668) antara throughput data dan latensi. Ini menunjukkan bahwa ketika latency rendah, throughput data cenderung tinggi, dan sebaliknya.
- **Data Throughput dan Signal Strength:** Korelasi antara throughput data dan kekuatan sinyal adalah negatif (-0.484), menunjukkan bahwa kekuatan sinyal yang lebih rendah cenderung berhubungan dengan throughput data yang lebih tinggi.
- **Latency dan Signal Strength:** Ada korelasi positif (0.504) antara latensi dan kekuatan sinyal. Ini berarti bahwa ketika kekuatan sinyal meningkat latensi cenderung meningkat juga.

	Data Throughput (Mbps)	Latency (ms)	Signal Strength (dBm)
Data Throughput (Mbps)	1.000000	-0.668173	-0.483769
Latency (ms)	-0.668173	1.000000	0.503550
Signal Strength (dBm)	-0.483769	0.503550	1.000000

# Insight

Network Type sangat menentukan performance dari jaringan. Semakin tinggi technology network yang di terima device, maka latency akan semakin rendah dan Data Throughput akan semakin tinggi, sehingga 5G adalah network type memiliki performance jaringan paling bagus.

---

LTE dan 4G memiliki perbedaan performance yang jauh significant, sehingga bisa dinggap bahwa LTE yang diukur ini adalah 4G generasi awal, karena memiliki performance yang mirip dengan 3G.

---

Signal strength 3G dan LTE, lebih besar dibandingkan 4G dan 5G, walaupun begitu performance (Data Througput dan Latency) 4G dan 5G lebih baik dibanding 3G dan LTE, sehingga bisa di simpulkan bahwa signal strength yang di terima tidak berpengaruh terhadap performance jaringan dalam case ini, tetapi lebih di tentukan oleh teknologi dari network type-nya.

---

Berdasarkan model machine learning yang coba dibangun yaitu Random Forest, Logistic Regression serta KNN Classifier, yang memiliki hasil evaluation terbaik adalah Random Forest dengan Tingkat akurasi yang hampir sempurna, yaitu 99.97%. Hasil ini bisa didapatkan dikarenakan fitur-fitur yang kuat (sangat prediktif terhadap target) dan model Random forest juga dapat meningkatkan kompleksitas model.

---

# Recommendation

Untuk kepuasan costumer, maka di sarankan untuk lebih memperbanyak Network 4G dan 5G, mungkin juga bisa di lakukan dengan bertahap yaitu dengan mengupgrade Network type 3G dan LTE ke 4G , sedangkan yang 4G bisa diupgrade ke 5G.

Didalam case yang di analysis di dalam project kali ini, tidak diperlukan mambangun machine learning, cukup dengan EDA dikarenakan fitur-fitur yang kuat sangat prediktif terhadap target.

Untuk mengembangkan project selanjutnya, di rekomendasikan untuk melakukan analisis dan prediksi improvement Performance dari masing-masing kota di kolom “locality”.