

# USER ANALYTICS IN THE TELECOMMUNICATION INDUSTRY

BY:

LOTOME JOHN

- Analysis was done in four phases as shown below:

1

User Overview  
Analysis

2

User Engagement  
Analysis

3

Experience Analysis

4

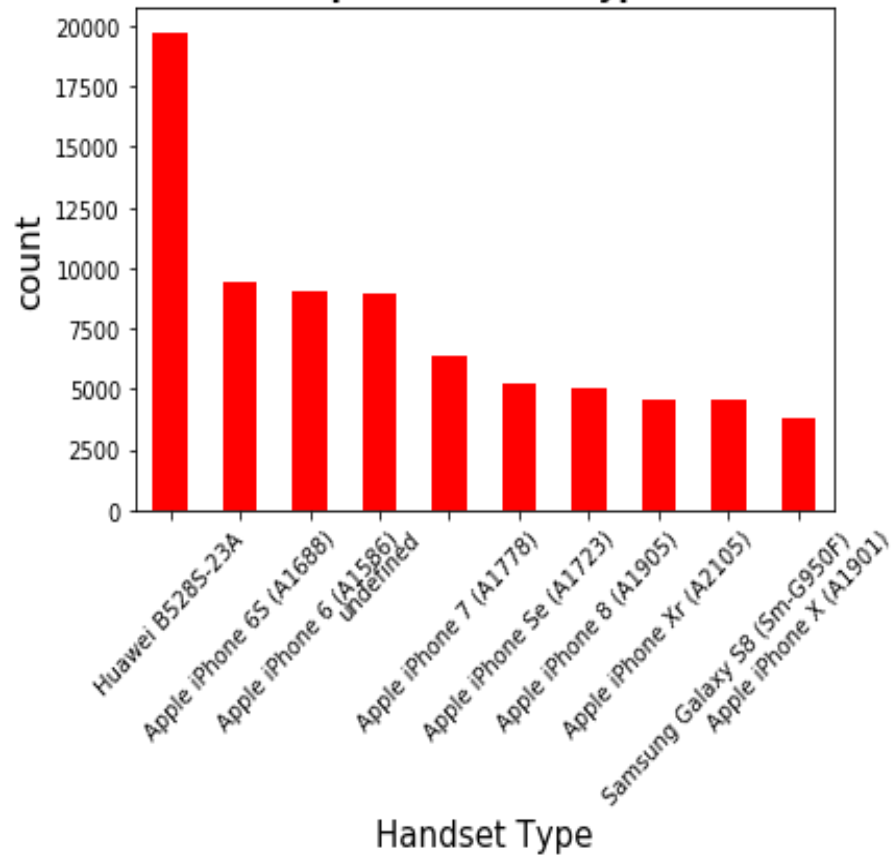
Satisfaction Analysis

# USER OVERVIEW ANALYSIS

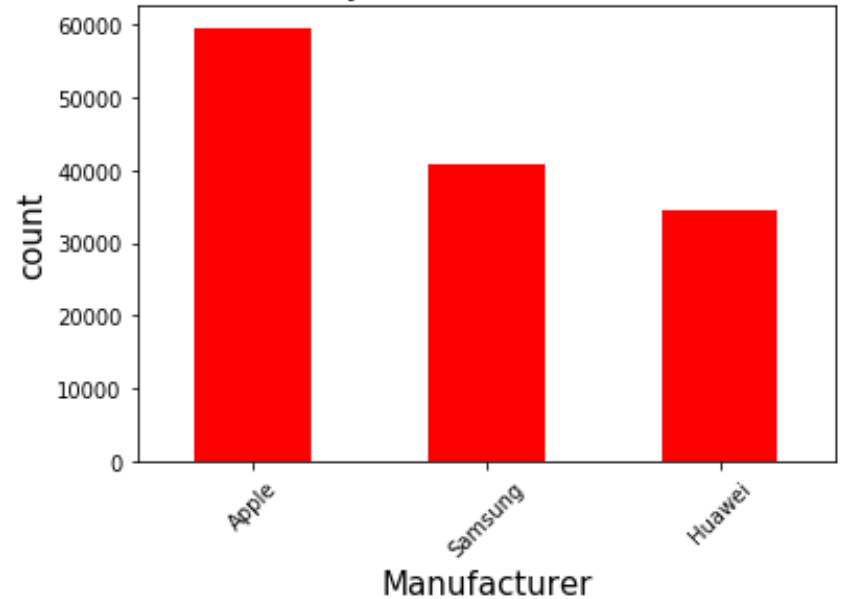
- The metrics used in this phase were as follows:
  - Number of xDR sessions
  - Session Duration
  - Type of Handset
  - Total Downloads and Uploads per Application
  - Total data volume for each session
- xDR, is the data channel. xDR sessions was used to identify each user in measuring the behavior of user based on the time they take for each session and the bandwidth they used in each application.
- Type of Handset made it possible to cluster the behavior of the users based on the Handset Type and Manufacturer.

# Top used Handset Types and Manufacturer

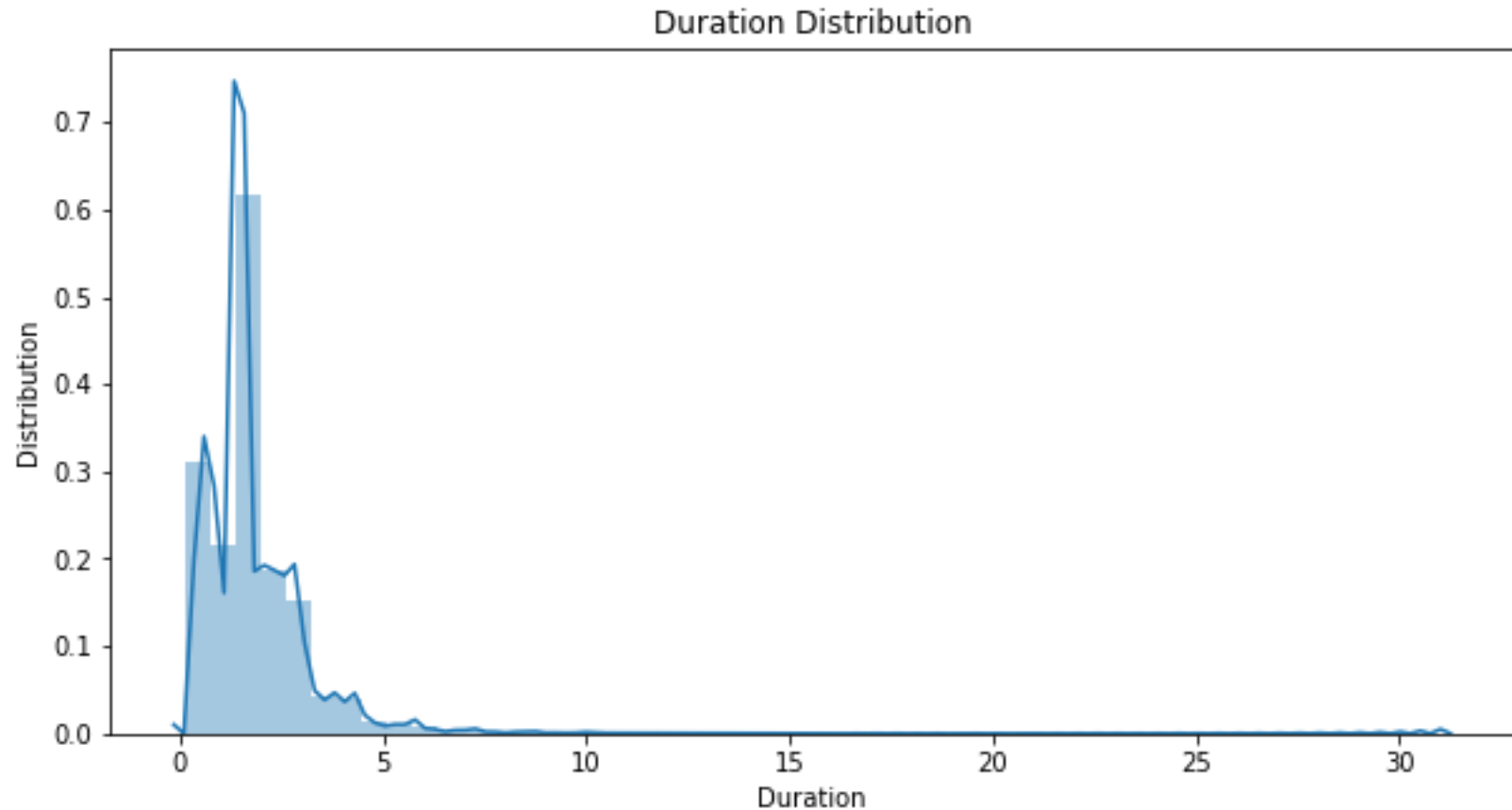
## Top 10 Handset Types



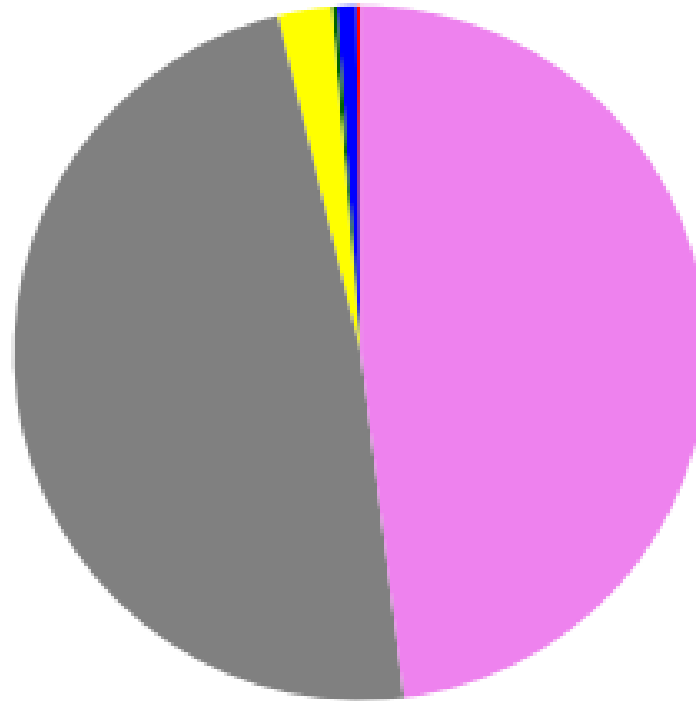
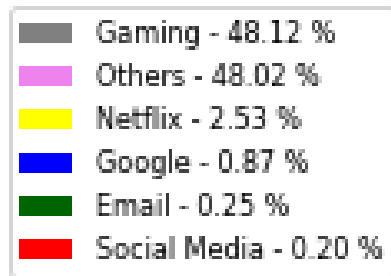
## Top 3 Manufacturers



## A Distribution plot of total duration session per users



- In comparing total xDR data per Application to the total data, Gaming and Others dominated with the highest percentage.
- Below is a pie chart showing the percentage for each.



## Top 5 Decile Class

- The users were segmented into deciles based on their duration for all sessions
- Below is the total data for each decile

Decile Ranks	Total Data(Terabytes)
1	6518.868
2	5608.775
3	5136.263
4	5172.336
5	7159.562

# User Engagement Analysis

- The metrics that were used to track users' engagement :
  - Session frequency
  - Duration of the session
  - Session Total Traffic
- The top 10 customers per engagement metrics is as shown in the proceeding slides:



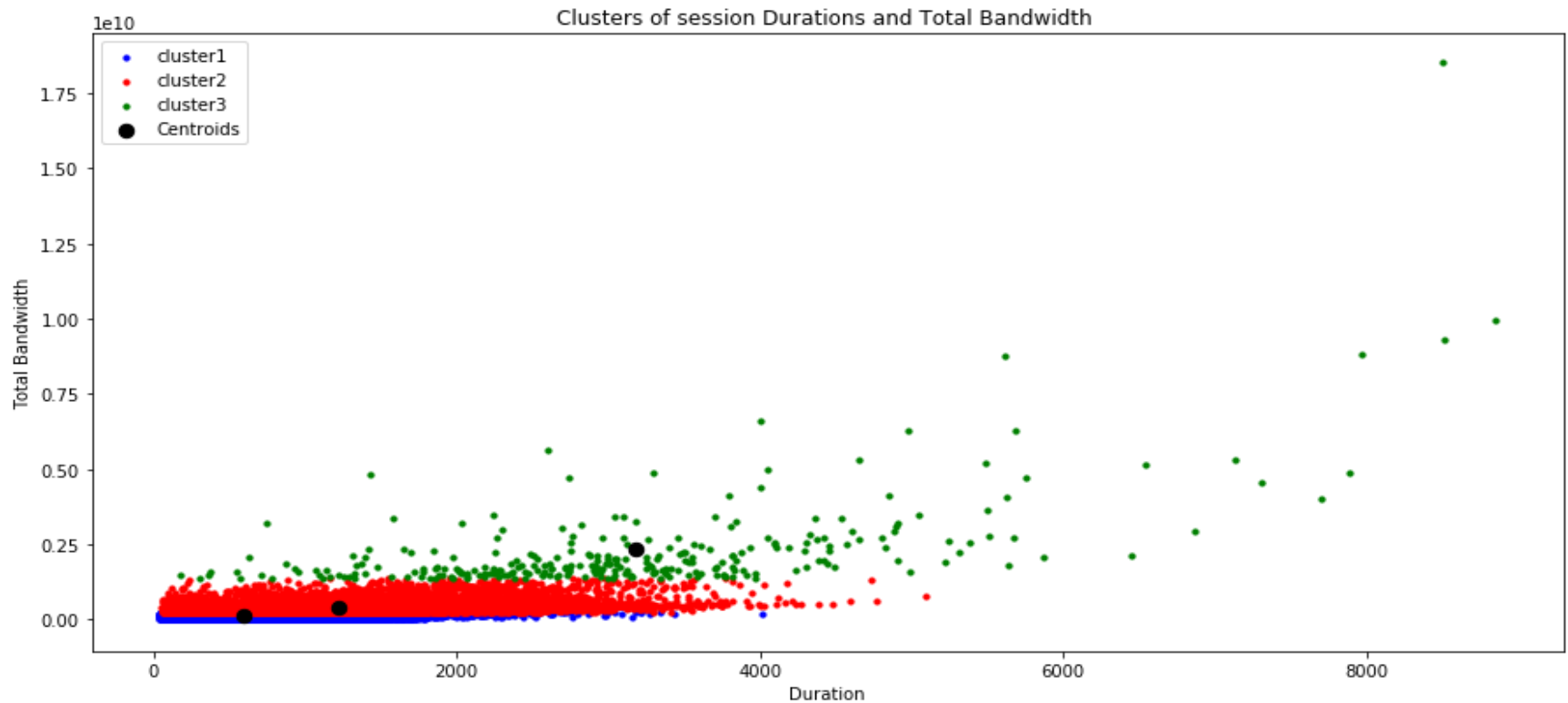
## Top 10 Customers based on session frequency

MSISDN/NUMBER	FREQUENCIES
3.362632e+10	18
3.362578e+10	17
3.361489e+10	17
3.365973e+10	16
3.376054e+10	15
3.367588e+10	15
3.366716e+10	13
3.362708e+10	12
3.360452e+10	12
3.376041e+10	12

## Top 10 Customers based on session duration

MSISDN/NUMBER	DURATION
3.362578e+10	1.855376e+10
3.361489e+10	9.966906e+09
3.376054e+10	9.279442e+09
3.362632e+10	8.791937e+09
3.366716e+10	8.744922e+09
3.366284e+10	6.614274e+09
3.366469e+10	6.288736e+09
3.360313e+10	6.287767e+09
3.366746e+10	5.649885e+09
3.376041e+10	5.321674e+09

# Users Engagement Clusters



The users were grouped into clusters based on their engagement scores. Three clusters were identified but they are not that distinct.  
The optimized value of k is 3

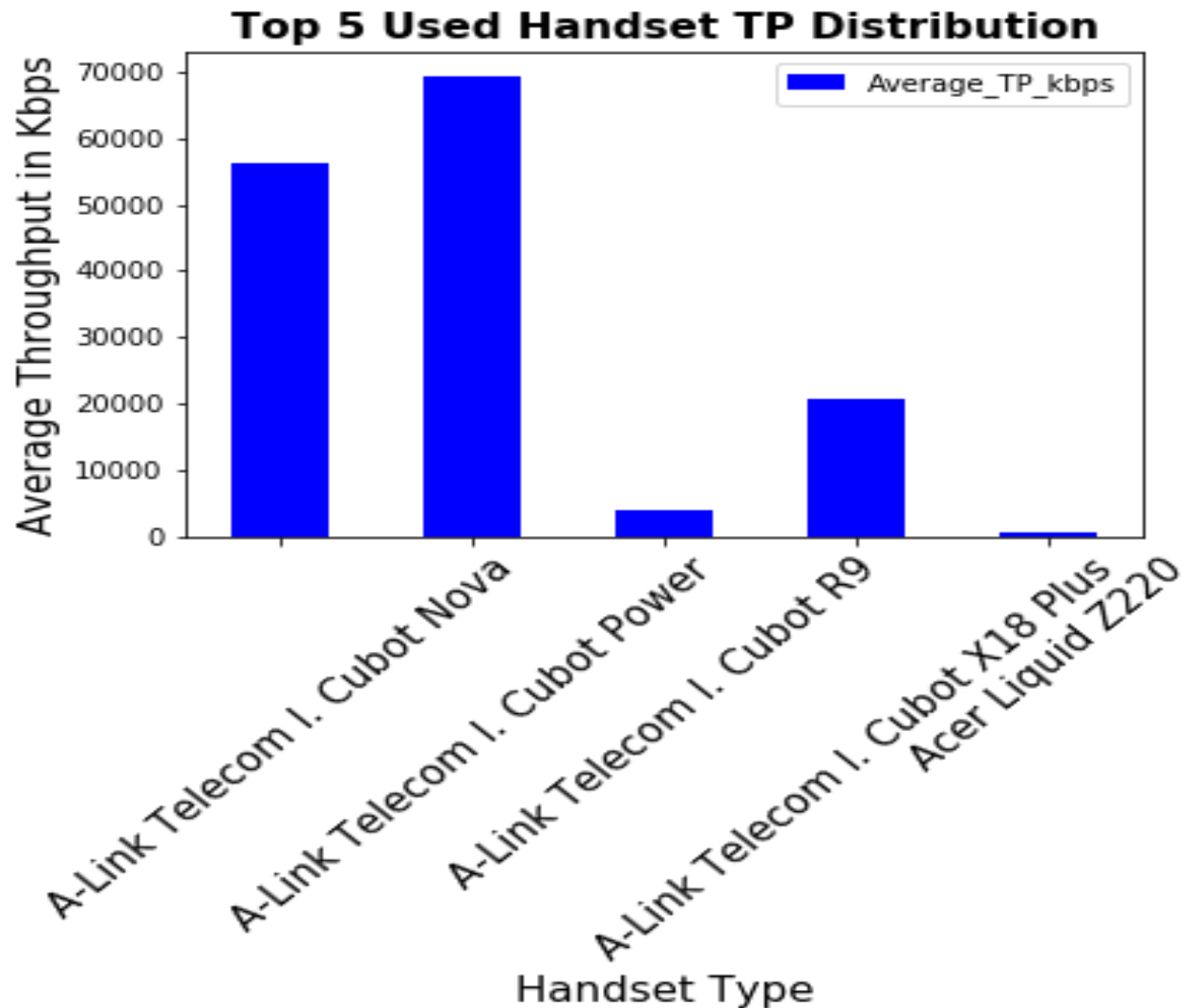
# Experience Analysis

- The following metrics were used in measuring the experience of users:
  - Average TCP retransmission
  - Average Round Trip Time
  - Handset Type
  - Average Throughput

## Top 10 TP retransmission Values per Customer

MSISDN/NUMBER	AVERAGE TP IN Kbps
3.366871e+10	840424.0
3.365955e+10	746139.0
3.366995e+10	690815.0
3.365808e+10	550352.0
3.366646e+10	524686.0
3.369925e+10	520268.0
3.369894e+10	492358.0
3.366050e+10	478037.0
3.376374e+10	468404.0

## Top 5 Devices with highest Average Throughput



# Satisfaction Analysis

- The satisfaction Analysis was based on Engagement and Experience Scores calculated from the Engagement and Experience Analysis.
- Below is top 10 satisfaction score

MSISDN/Number	SATISFACTION SCORES
3.366871e+10	367509.555904
3.365955e+10	321503.596202
3.366995e+10	292524.932020
3.365808e+10	222405.271503
3.366646e+10	209645.735947
3.369925e+10	208138.382733
3.369894e+10	193363.073153
3.366050e+10	186367.867694
3.376374e+10	181629.777422
3.366911e+10	181352.534346

# Multiple Linear Regression

- A multiple linear Regression was conducted on selected metrics to model the satisfaction score

```

=====
                        OLS Regression Results
=====
Dep. Variable:          Satisfaction_Score      R-squared:                0.073
Model:                  OLS                    Adj. R-squared:           0.073
Method:                 Least Squares          F-statistic:             465.3
Date:                  Sun, 02 Aug 2020        Prob (F-statistic):      0.00
Time:                  08:03:15               Log-Likelihood:         -3.7194e+05
No. Observations:      35288                  AIC:                    7.439e+05
Df Residuals:          35281                  BIC:                    7.440e+05
Df Model:               6
Covariance Type:       nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
const	3.778e+04	109.443	345.167	0.000	3.76e+04	3.8e+04
Handset Type	2531.7246	169.545	14.932	0.000	2199.411	2864.038
Average_TCPR_MB	-13.9354	3.327	-4.189	0.000	-20.456	-7.415
Average_RTT_ms	7.2518	0.499	14.543	0.000	6.274	8.229
Average_TP_kbps	0.0105	0.001	7.924	0.000	0.008	0.013
Duration_Minutes	-149.8630	33.846	-4.428	0.000	-216.202	-83.523
Total	-0.0270	0.178	-0.152	0.879	-0.375	0.322

```

=====
Omnibus:                40766.137      Durbin-Watson:           1.986
Prob(Omnibus):          0.000          Jarque-Bera (JB):       15383379.863
Skew:                   5.584          Prob(JB):               0.00
Kurtosis:               104.675        Cond. No.                2.71e+05
=====

```

## Warnings:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The condition number is large, 2.71e+05. This might indicate that there are strong multicollinearity or other numerical problems.



## RECOMMENDATIONS

- There is high chances for growth on the company, based on the analysis, the Handset type plays a major role in satisfaction, which is more user oriented.
- Packet loss contributed negatively to the satisfaction score. If the Company can improve the infrastructure to reduce packet retransmissions and packet loss rate, then there will be increase in satisfaction score.
- Majority of the data is consumed in Gaming and Others. These applications are resource intensives and requires high QoS, therefore we can purchase the company if we can ensures that there is enough bandwidth for this application.

# LIMITATIONS

- Difficult to identify the different clusters
- Data labeling was not properly data hence consuming time in data wrangling
- Huge dataset that consumes machine power while working with

