

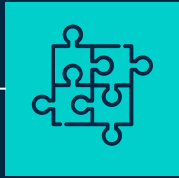
10 Academy  
Inspire the next

# DEEPER INTO TelCo COMPANY

User Analytics in TelCo  
industry

David Elvis  
Junior Data Scientist

# TABLE OF CONTENTS



01

## PROBLEM & DATA OVERVIEW

- Business objective
- Analysis objective
- User overview analysis



02

## OUR PROCESS

- User engagement analysis
- User experience analysis
- User satisfactory analysis



03

## TARGET

- Results
- Recommendation
- Conclusion

# UNDERSTANDING THE PROBLEM

## CAN I BUY TelCo?

What is the current situation  
in TelCo industry?

Are there opportunities  
for growth of TelCo  
industry?



# USER OVERVIEW ANALYSIS

## DESCRIBING OUR DATA

### TYPES OF VARIABLES

#### QUANTITATIVE VARIABLES

##### 1. DISCRETE VARIABLES

- Number of xDR sessions

##### 2. CONTINUOUS VARIABLES

- Session duration

- Total UL & DL

- Total data volume in each session

### DATA TYPES

#### NUMERIC DATA TYPES

- Session duration

- Total ULs % DLs

- Total data volume in each session

- Number of xDR sessions

# Key Insights on Univariate Analysis

1. The Data for each of the variables is symmetrical, the mean and median is similar.
2. Google has the highest mean compared to other application which shows that most of the customers spend more time and data in Google app, hence more advertisements can be done through Google app to increase the number of audience.
3. The standard deviation in Youtube data is large than the Mean which indicates that there is more variability of Customers spending data In Youtube App.
4. Youtube data and Netflix have the highest standard deviation value which indicates greater spread in the data.
5. All the variables have a large range value which indicates the greater dispersion in the data.
6. In most of the applications, most of the customers

according to usage of the data, are in between the median and the upper quartile compared to between median and the lower quartile.

```
#Getting the statistical summary for the quantitative variables of the datamart
summary=df1.describe()
summary1=pd.DataFrame(summary)
summary1.to_csv("summary.csv")
summary1.drop(['Total Uploads','Total Downloads'],axis=1)
```

	MSISDN/Number	no.of xDR sessions	Session duration	Total Uls and DLs	Youtube_total_data	Google_total_data	Email_to
count	1.068560e+05	106856.000000	1.068560e+05	1.068560e+05	1.068560e+05	1.068560e+05	1.068560e+05
mean	4.511474e+10	1.393792	9.662390e+07	9.166177e+08	2.261432e+07	7.807186e+06	2.256186e+06
std	2.889423e+12	0.806022	8.395440e+07	3.445153e+08	9.247766e+06	3.518899e+06	1.070186e+06
min	3.360100e+10	1.000000	7.142988e+06	5.836229e+07	1.296550e+05	4.033000e+04	8.359000e+03
25%	3.365088e+10	1.000000	4.086000e+07	6.691826e+08	1.598535e+07	4.938135e+06	1.356186e+06
50%	3.366365e+10	1.000000	8.639990e+07	9.166781e+08	2.263187e+07	7.815112e+06	2.261860e+06
75%	3.368344e+10	2.000000	1.188538e+08	1.164285e+09	2.925176e+07	1.068372e+07	3.156186e+06
max	8.823971e+14	18.000000	1.823653e+09	1.780674e+09	4.519008e+07	1.552878e+07	4.518600e+06

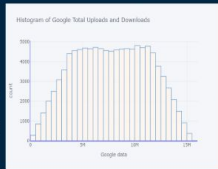
# GRAPHICAL UNIVARIATE ANALYSIS

- Most of the users had one xDR session with the data distribution skewed to the right.
- Most of the users' session lasted between 50k-100k ms with few customers' sessions lasting 400k-450k ms. The data is skewed to the right.
- There is normal distribution in Total ULs % DLs with majority of users using 900-950M data bytes.
- There is normal distribution in usage of data for Google, Youtube, Email, Netflix apps and uniform distribution is exhibited in Social Media and Gaming apps.
- More users spend their data in Gaming apps, where we have a majority spending 740-760M data bytes.

## Graphical Univariate Analysis

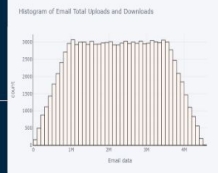
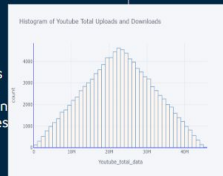
### Google Data

-The distribution is normal with the highest using between 10M-10.5M data bytes in google app



### Youtube Data

-The distribution is normal with most users using between 22M-23M data bytes in Youtube app

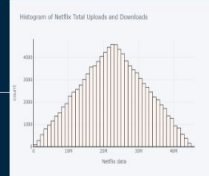


### Email Data

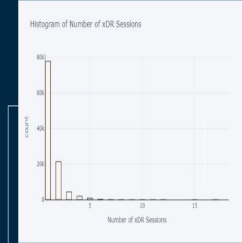
-Most of the Customers used between 1M-1.1M bytes in Emails Concentration of the distribution being around the mean.

### Netflix Data

-The distribution is normal with the most Customers using between 22M-23M data bytes in Netflix app

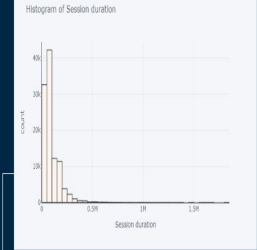


## Graphical Univariate Analysis



### xDR Sessions

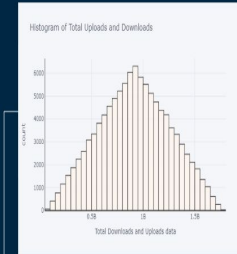
-Most of the Customers had one session with the lowest number having five. The data is skewed to the right.



### Session Duration

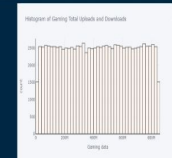
-Most of the customers' session lasted between 50k-100k ms with few customers' sessions lasting 400k-450k ms. The data is skewed to the right.

## Graphical Univariate Analysis



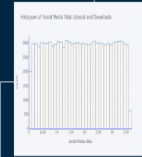
### Total Uploads & Downloads

-The Distribution of Total ULs and DLs is normal with most of the customers using between 900M-950M bytes



### Gaming Data

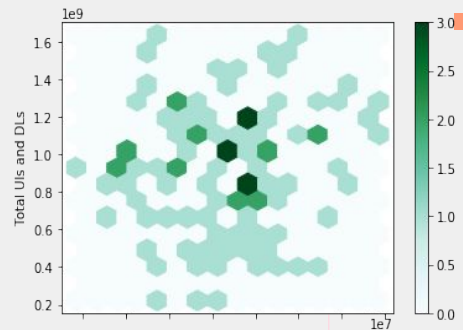
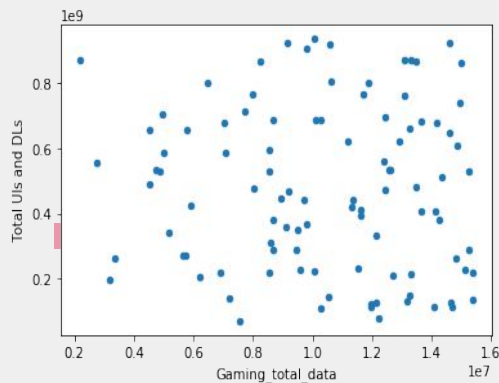
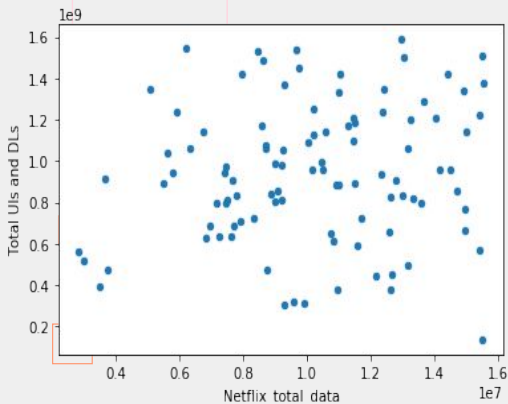
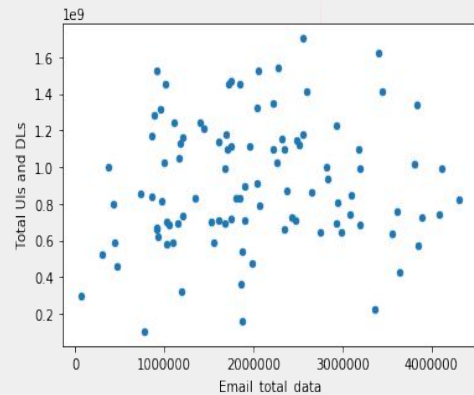
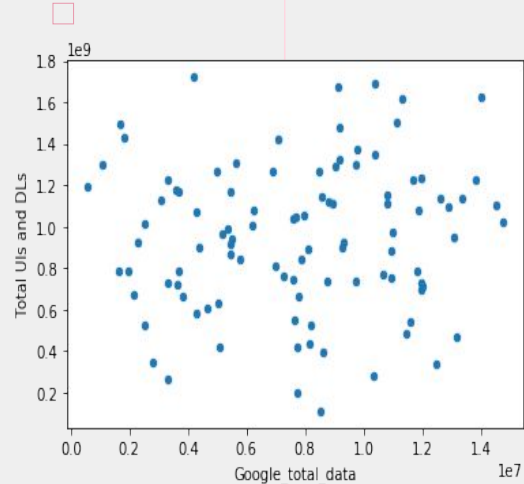
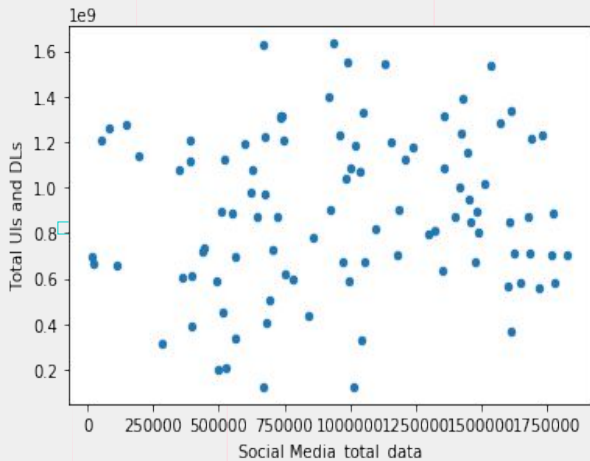
-Many of the Customers are using between 320M-340M data bytes in Gaming apps and the distribution is uniform



### Social Media Data

-The distribution of the data is uniform with many Customers using between 1.3M- 1.4M data bytes in social media

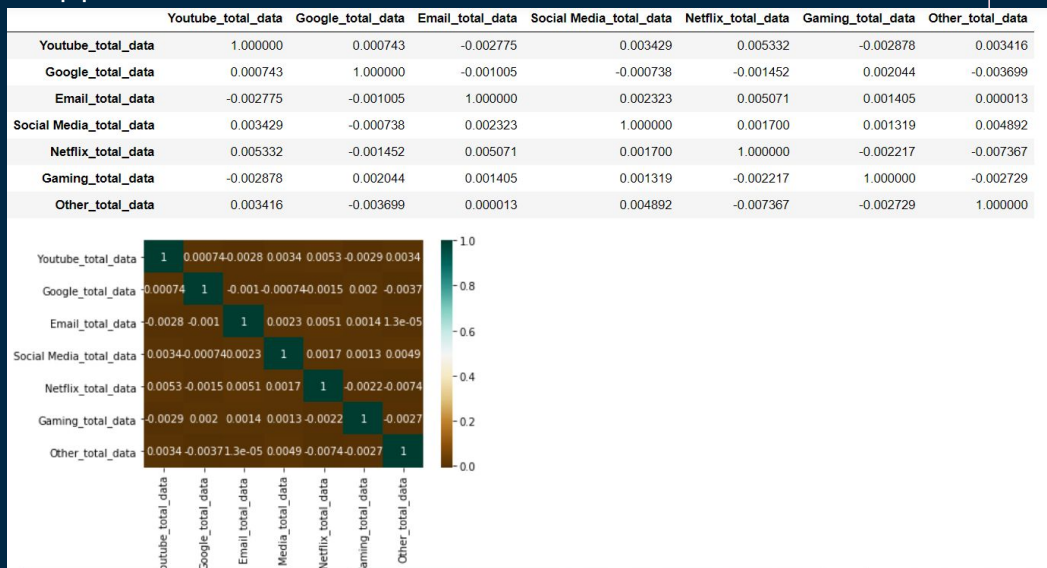
# Bivariate Analysis Plots



**-There is Weak relationship between the data used in the application and the Total Upload, Download data.**

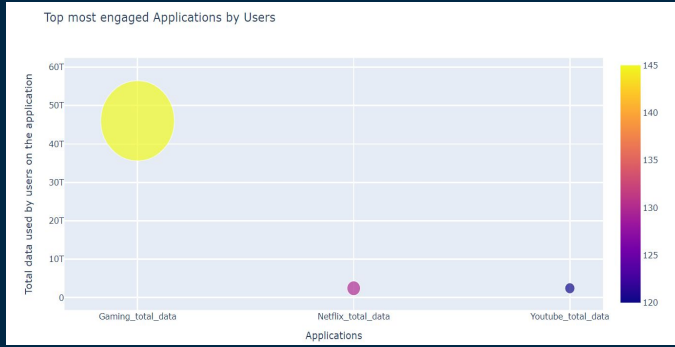
# CORRELATION ANALYSIS

- The correlation between the variables is extremely weak with some of the variables exhibiting positive and other negative relationship.
- Some of the apps exhibiting relatively moderate positive relationship are, Youtube and Netflix, Netflix and Email application.





# MOST USED APPLICATIONS



YOUTUBE APP

NETFLIX APP

GAMING APPS



4.75% of the Total  
data is spent in  
Youtube app



4.76% of the Total  
data is used in  
Netflix app



Gaming apps are the  
most engaged apps  
by users



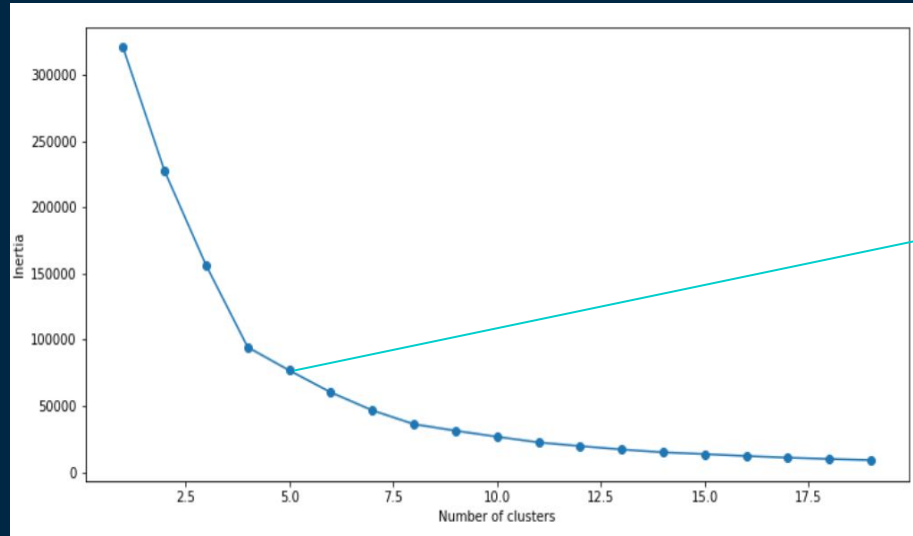
# USER ENGAGEMENT ANALYSIS

-Building & improving the QoS (Quality of Service) to leverage the mobile platforms and to get more users for the business is good but the success of the business would be determined by the user engagement and activity of the customers on available apps

- Most of the users had very few sessions but used much data bytes in the few sessions they had.



# OPTIMIZED K-VALUE OF ENGAGEMENT CLUSTERS



5-8

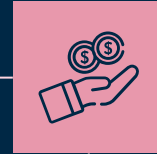
-Using the Elbow method, we can see that the optimized k-value is between 5-8. The elbow is not clear and sharp which tells that our data was not clearly clustered.

# USER EXPERIENCE ANALYTICS

-It is so important to track and evaluate the customer's experience to optimize their products and services so that it meets the evolving user expectations, needs and acceptance.

## MARS

Despite being red,  
Mars is a cold place

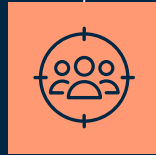


## NEPTUNE

It's the farthest  
planet from the Sun

## JUPITER

It's the biggest planet  
in the Solar System



## SATURN

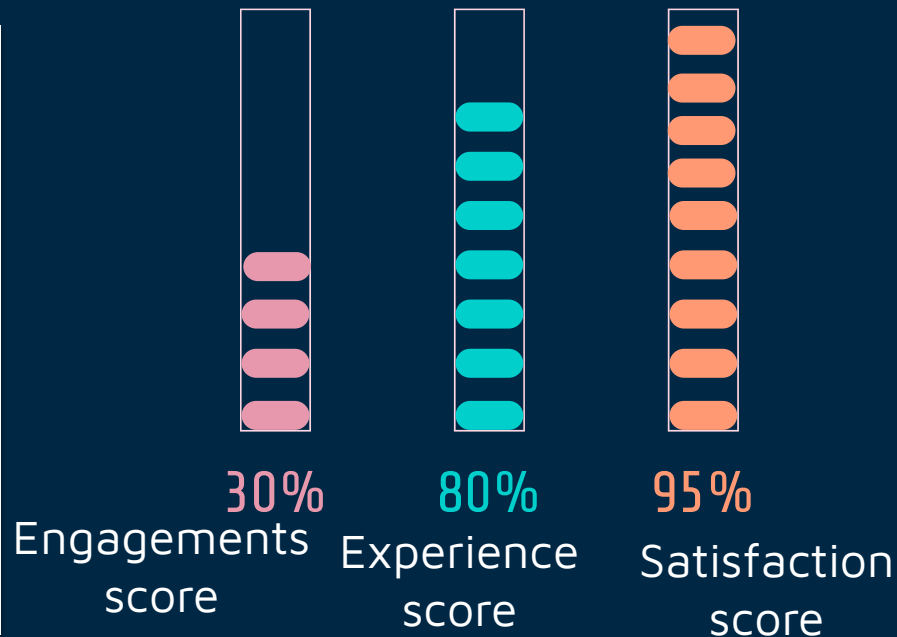
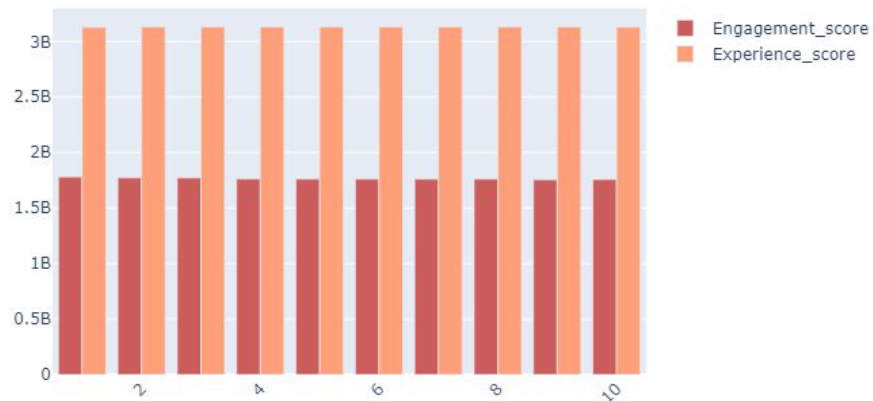
Saturn is the ringed  
one and a gas giant

# USER EXPERIENCE CLUSTERS

	COUNT	DESCRIPTION
CLUSTER 1	149595	This cluster had high numbers meaning that most of the customers were dissatisfied because of the unreliability of the network shown by high TCP retransmission.
CLUSTER 2	244	Few people showed a medium level of satisfaction from the company.
CLUSTER 3	162	Fewer people were fully satisfied according to their experience with TellCo company.

# USER SATISFACTION ANALYSIS

-The Satisfaction of the customers heavily depends on the experience of the user.



# USER SATISFACTION, ENGAGEMENT AND EXPERIENCE TABLE

The screenshot displays the MySQL Workbench interface. The 'Query' tab is active, showing a SQL query that selects data from the 'final\_tellico\_table'. The 'Result Grid' shows the output of the query, which includes columns for engagement distance, MSISDN number, experience distance, and satisfaction score. The 'Output' tab is also visible, showing the execution log of the query, including the creation of the 'final\_tellico\_table' and the execution of the SELECT statement.

Query:

```
1 SELECT user();
2 SELECT * FROM final_tellico_table;
```

Result Grid:

Engagement_Euclidean_Distance	MSISDN/Number	Experience_Euclidean_Distance	satisfaction_score
517373827.4885642	33664962239	3110430259.651428	3627804087.139992
1234089594.4885643	33681854413	3110430259.651985	4344519854.140563
718383561.4885643	33760627129	3110430259.652813	3828813821.1413774
1638392681.4885643	33750343200	3110430259.649714	4748822941.138278
1158390903.4885643	33699755932	3110430259.652813	4268821163.1413774
1439738243.4885643	33668185951	3110430259.6475945	4550168503.136159
922140850.4885643	33665368271	3110430259.6515083	4032571110.140073
1371765179.4885643	33763490140	311320653.615465	4502968833.10403
1078695893.4885643	33698743617	3110430259.6495504	4187326153.138115
937844364.4885643	33659219748	3128761205.2708526	4066605569.7594166
747243932.4885643	3365646348	3110430259.652813	3857674192.1413774

Output:

#	Time	Action	Message	Duration / Fetch
5	13:11:36	SHOW SESSION VARIABLES LIKE 'lower_case_table_names'	OK	0.000 sec
6	13:11:36	SHOW TABLES FROM 'elvis_10acad' like 'final_tellico_table'	OK	0.000 sec
7	13:12:00	CREATE TABLE 'elvis_10acad'. 'final_tellico_table' ('Engagement_Euclidean_Distance' doubl...	OK	0.000 sec
8	13:12:00	PREPARE stmt FROM 'INSERT INTO 'elvis_10acad'. 'final_tellico_table' ('Experience_Euclid...	OK	0.000 sec
9	13:59:13	DEALLOCATE PREPARE stmt	OK	0.000 sec
10	14:48:13	SELECT user() LIMIT 0, 300	1 row(s) returned	0.000 sec / 0.000 sec
11	15:07:41	SELECT * FROM final_tellico_table WHERE ROWNUM <= 3 LIMIT 0, 300	Error Code: 1054. Unknown column 'ROWNUM' in 'where clause'	0.000 sec
12	15:08:09	SELECT * FROM final_tellico_table WHERE ROWNUM <= 3 LIMIT 0, 300	Error Code: 1054. Unknown column 'ROWNUM' in 'where clause'	0.000 sec
13	15:11:09	SELECT * FROM final_tellico_table LIMIT 0, 300	300 row(s) returned	0.016 sec / 0.000 sec

- Final table from the analysis showing the user's satisfaction, engagement and experience table.

# LIMITATIONS TO THE ANALYSIS

- Many missing values in important columns which paralysed our analysis mostly in performing the clustering.
- There was a big number of outliers in our data which even after correcting altered our accuracy in the analysis.
- Some metrics were lacking which could have given us more insight to our questions eg the voice call data of each customer, messaging data.





# RECOMMENDATIONS AND CONCLUSION

Increasing the experience score of the customers, by making better the network services to the customers

Getting closer to customers so as to be able to understand them well.



Create an engaging user experience by building and improving the Quality of Service especially in gaming applications

After Deeper dive into TellCo Company, it is clear that there is a great potential for TellCo to thrive more if the recommendations given are implemented, thus TellCo Company is a worthy business to invest in