



# BIG DATA ANALYSIS

Minh Duy Truong

May 11<sup>th</sup>, 2024

# COMPONENTS

1. Introduction
2. Data explanation
3. Overview
4. Analysis
5. Recommendations

# INTRODUCTION

A live-streaming platform conducted an analysis on customer behavior data from April 1<sup>st</sup> 2022 – April 30<sup>th</sup> 2022.

Purposes:

1. Gain more insights on customers' behavior
2. Improve on customer services for different customer segmentations.

# DATA

**OLTP** was stored in Data Lake as JSON file. After **ETL** process, it became **OLAP** data and stored in Data Warehouse as CSV file.

OLTP data

ETL

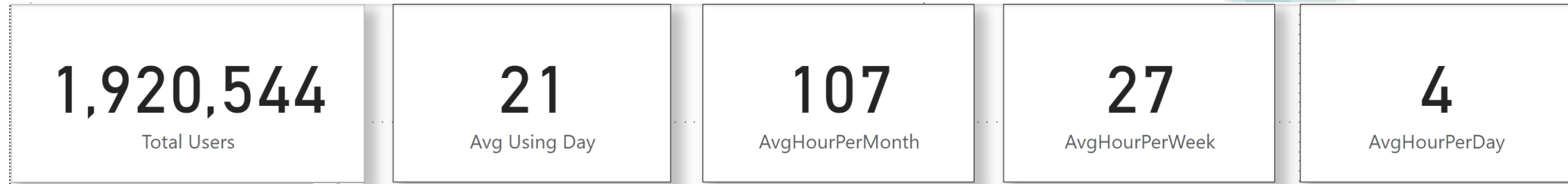
OLAP data

```
_id: "AX_momhfa1FFivsGrn2u"
_index: "history"
_type: "kplus"
_score: 0
▼ _source: Object
  Contract: "DNFD74404"
  Mac: "90324BB44C39"
  TotalDuration: 691
  AppName: "KPLUS"
```

```
_id: "AX_momg9a1FFivsGrnlx"
_index: "history"
_type: "kplus"
_score: 0
▼ _source: Object
  Contract: "HNNH231500"
  Mac: "10394E196F73"
  TotalDuration: 891
  AppName: "KPLUS"
```

	Contract ▾	ChildDuration ▾	MovieDuration ▾	SportDuration ▾	TVDuration ▾	RelaxDuration ▾	MostWatch ▾	Taste ▾	DateCount ▾	Activeness ▾
1	113.182.209.48	0	0	0	63	89 Relax	TV-Relax	1	3.33	
2	14.182.110.125	0	0	0	404	92 TV	TV-Relax	1	3.33	
3	AGAAA0338	0	0	0	278633	0 TV	TV	30	100.0	
4	AGAAA0342	0	0	0	117788	204 TV	TV-Relax	12	40.0	
5	AGAAA0391	0	129	0	158931	373 TV	Movie-TV-Relax	11	36.67	
6	AGAAA0613	0	0	0	9377	26 TV	TV-Relax	24	80.0	
7	AGAAA0638	0	0	0	227016	0 TV	TV	30	100.0	
8	AGAAA0692	0	0	0	107057	0 TV	TV	8	26.67	
9	AGAAA0723	0	0	0	9279	0 TV	TV	19	63.33	
10	AGAAA0729	0	0	0	161781	0 TV	TV	3	10.0	
11	AGAAA0732	0	463	0	99030	0 TV	Movie-TV	23	76.67	
12	AGAAA0750	0	0	0	525708	0 TV	TV	25	83.33	
13	AGAAA0840	0	0	0	12141	0 TV	TV	25	83.33	
14	AGAAA0885	0	0	0	354499	0 TV	TV	20	66.67	
15	AGAAA0886	0	0	0	39163	0 TV	TV	14	46.67	
16	AGAAA0901	0	0	0	796083	0 TV	TV	30	100.0	
17	AGAAA1004	0	0	0	58329	0 TV	TV	25	83.33	
18	AGAAA1047	0	53227	0	305816	0 TV	Movie-TV	30	100.0	
19	AGAAA1147	0	0	0	1299590	103 TV	TV-Relax	23	76.67	
20	AGAAA1218	0	0	0	132794	0 TV	TV	9	30.0	
21	AGAAA1282	427	0	0	168792	0 TV	Child-TV	6	20.0	
22	AGAAA1316	0	0	0	198831	0 TV	TV	29	96.67	
23	AGAAA1345	0	0	0	58622	0 TV	TV	29	96.67	
24	AGAAA1376	0	0	0	259200	0 TV	TV	3	10.0	
25	AGAAA1389	0	0	0	1252129	0 TV	TV	30	100.0	
26	AGAAA1414	0	0	0	106617	0 TV	TV	7	23.33	
27	AGAAA1537	0	257	0	68597	0 TV	Movie-TV	29	96.67	
28	AGAAA1539	0	0	0	172800	0 TV	TV	2	6.67	
29	AGAAA1561	0	0	0	915378	0 TV	TV	24	80.0	
30	AGAAA1587	0	0	0	258299	0 TV	TV	11	36.67	

# OVERVIEW



In April 2022, platform had over **1.9M** Users.

By average, each user used the service for **21** days.

An average user spent:

- **107 hours** on service **per month**,
- **27 hours** per **week**
- **4 hours** per **day**

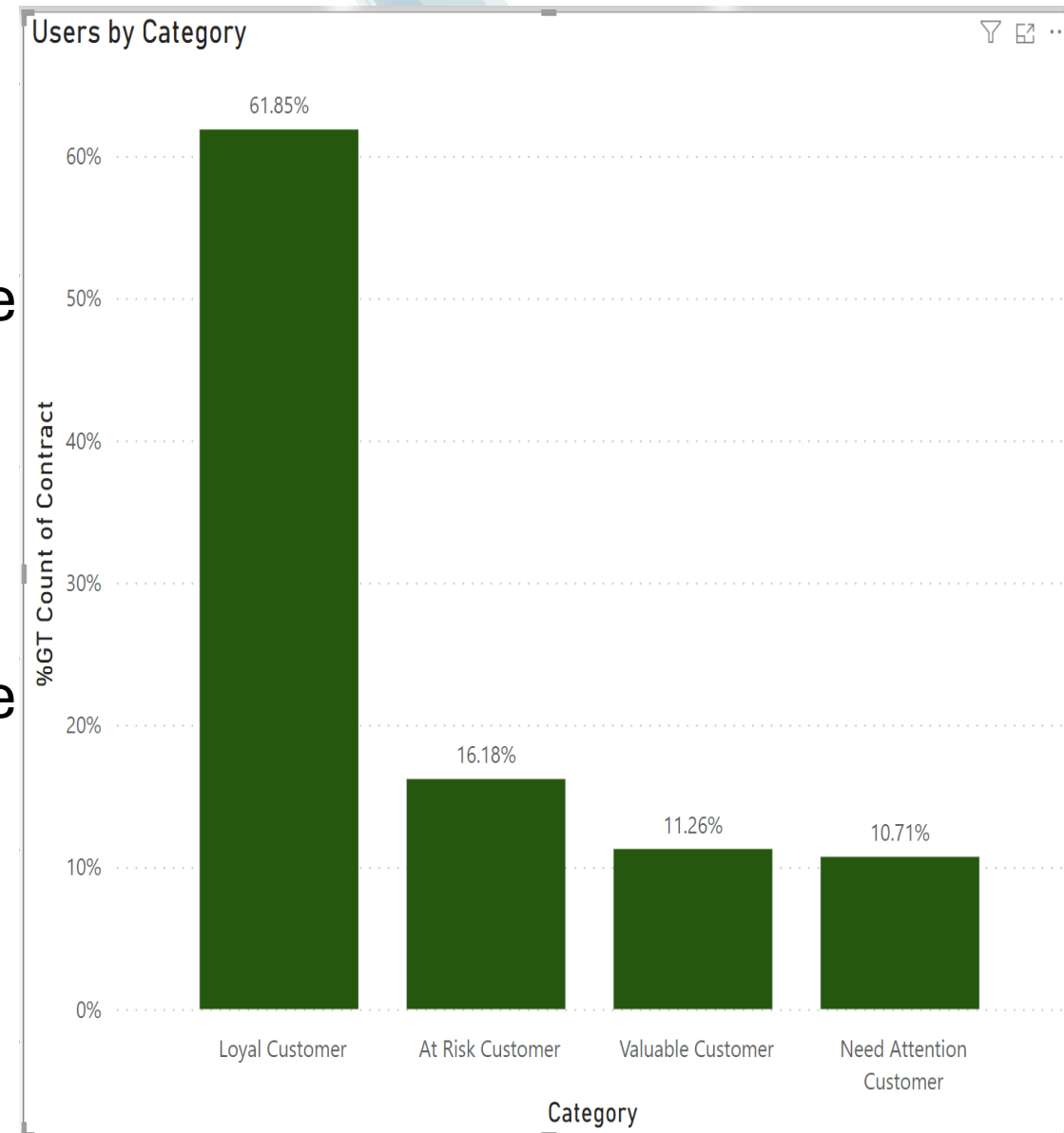
# ANALYSIS

We can split customers into 4 categories:

1. **At risk:** <7 days using service
2. **Need attention:** 7-14 days using service
3. **Valuable:** 14-21 days using service
4. **Loyal:** > 21 days using service

We can see **61.85%** customers and revenue come from **Loyal customers**

=> Convert customers into Loyal Customers.



# ANALYSIS

Analyzing Users by Most Watch:

**97%** Users watch **TV** the most

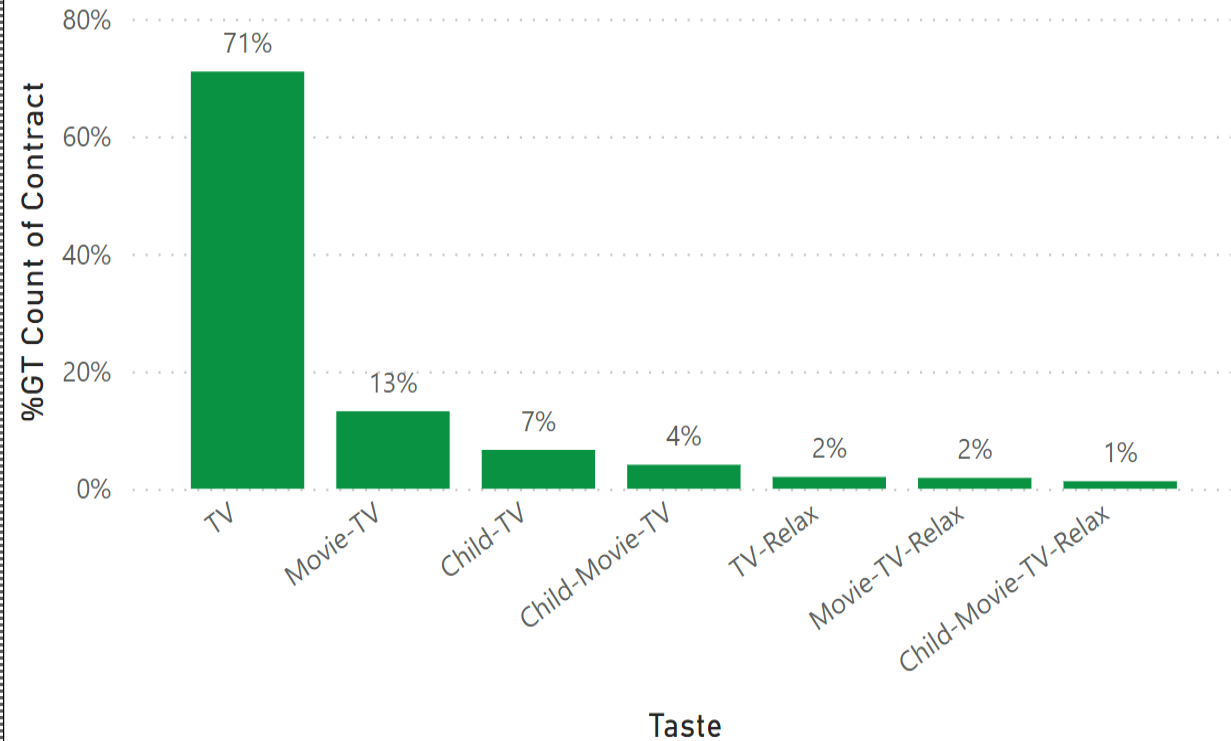
- **71%** Users **ONLY** watch **TV**
- 13% watch TV and Movie
- 7% watch TV and Child
- 4% watch TV, Child, and Movie

=> We'll **exclude** customers who watched **TV only** for further analysis

Users by Most Watch

MostWatch	Users	% Users
TV	1865816	97.15%
Movie	42766	2.23%
Child	10938	0.57%
Relax	893	0.05%
Sport	131	0.01%
Total	1920544	100.00%

Users by Tastes

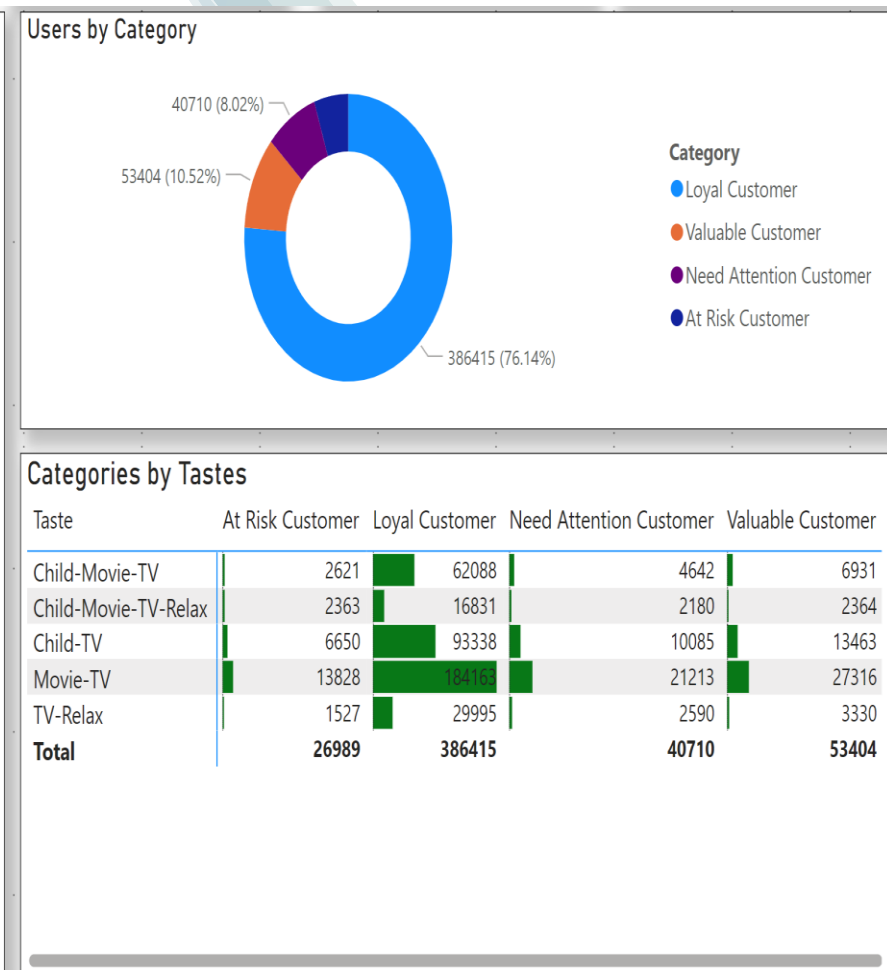
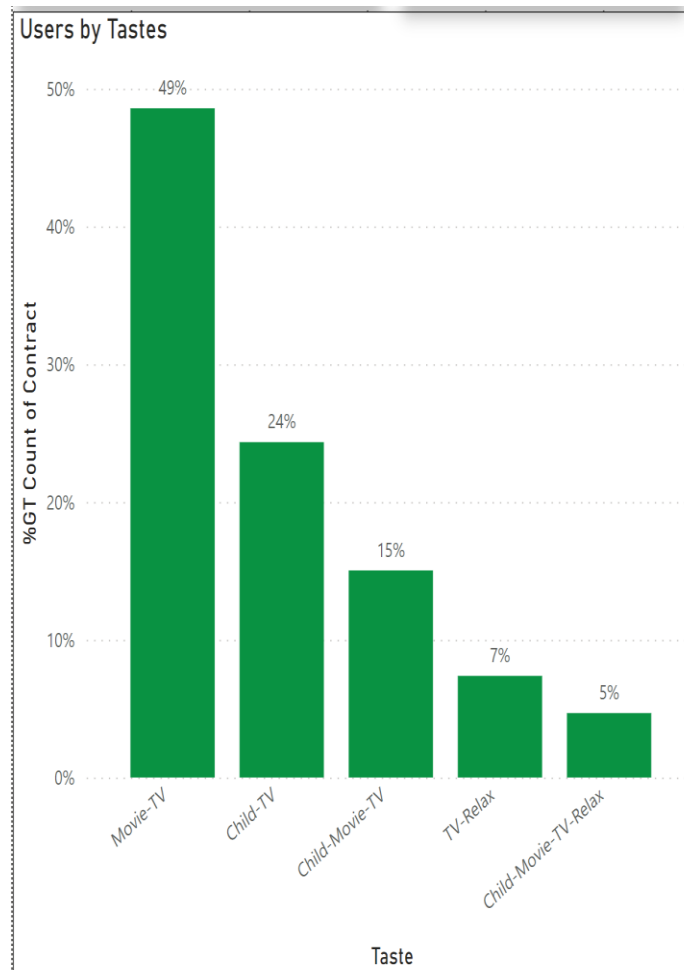


# ANALYSIS

Majority is **Loyal**  
**customers 76.14%**

**Movie-TV** is the **second**  
highest taste, follow up is  
**Child-TV**, and **Child-**  
**Movie-TV**

**Top 3 Contents:** TV,  
Movie, and Child.





# ANALYSIS

Customers who watched Movie only, Child only and Child-Movie is less than **0.03%** total customers.

=> Customers **either** watched **TV only**, or **main content is TV** and watch Movie, and Child as an extra.

Users by Tastes			Filter	Share	More
Taste	Users	%GT Count of Contract			
TV	1327403	74.82%			
Movie-TV	246520	13.89%			
Child-TV	123536	6.96%			
Child-Movie-TV	76282	4.30%			
Movie	324	0.02%			
Child	119	0.01%			
Child-Movie	37	0.00%			
<b>Total</b>	<b>1774221</b>	<b>100.00%</b>			

















# ANALYSIS

Only **37** customers watch Child-Movie but not TV

⇒ Child and Movie are **NOT** main products of company.

⇒ Focus on more TV content to convert customers into Loyal Customers.

## Categories by Tastes

Taste ▼	At Risk Customer	Loyal Customer	Need Attention Customer	Valuable Customer
TV	 280618	 731080	 160152	 155553
Movie-TV	 13828	 184163	 21213	 27316
Child-TV	 6650	 93338	 10085	 13463
Child-Movie-TV	 2621	 62088	 4642	 6931
Child-Movie	20	4	10	3

# RECOMMENDATIONS

1. **Convert** customers from into **Loyal Customers** to increase revenue.
2. More **content in TV** to attract more customers.
3. Movie and Child are **NOT** important contents.