NEXT HIKES
IT SOLUTIONS

Presentation Focus

TELCO XDR User Behavior & Experience Analytics

Presented by

Deepa Pathak

DataWithDeepahttps://github.com/DataWithDeepa

https://github.com/DataWithDeepa

Company

NeXT hikes IT Solutions

Strategic Business Objectives

Informed Acquisition Decisions

Our primary objective is to conduct an in-depth analysis of TellCo's extensive user data. This analysis provides critical insights to empower investors in making informed acquisition decisions, highlighting the intrinsic value and growth potential embedded within TellCo's user base.

Unveiling User Behavior & Trends

We meticulously identify key patterns related to user growth, engagement, overall experience, and satisfaction. By dissecting these metrics, we paint a comprehensive picture of the current user landscape and forecast future trends, ensuring a clear understanding of TellCo's market position and operational efficiency.

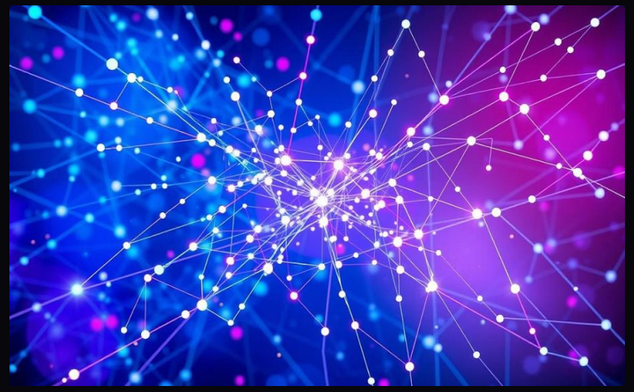
Comprehensive Dataset Overview

Rich Data Source

We leverage one month of comprehensive Telco xDR (eXperience Data Record) data. This detailed dataset captures a wide array of user interactions, providing a granular view of network usage and service consumption.

Diverse App Usage

The dataset includes crucial information on user engagement with popular applications such as YouTube, Netflix, various gaming platforms, and a multitude of social media apps. This allows for in-depth analysis of application-specific behaviors.



Extensive Features

With over 150 features per user, the dataset offers a multi-dimensional perspective on user behavior. These features range from connection details to device specifications and service consumption patterns, enabling robust

Navigating Dataset Challenges

Addressing Missing Values

A common challenge in large datasets is the presence of missing values. We implemented advanced imputation techniques to ensure data integrity and prevent skewed analytical results, maintaining the robustness of our models.

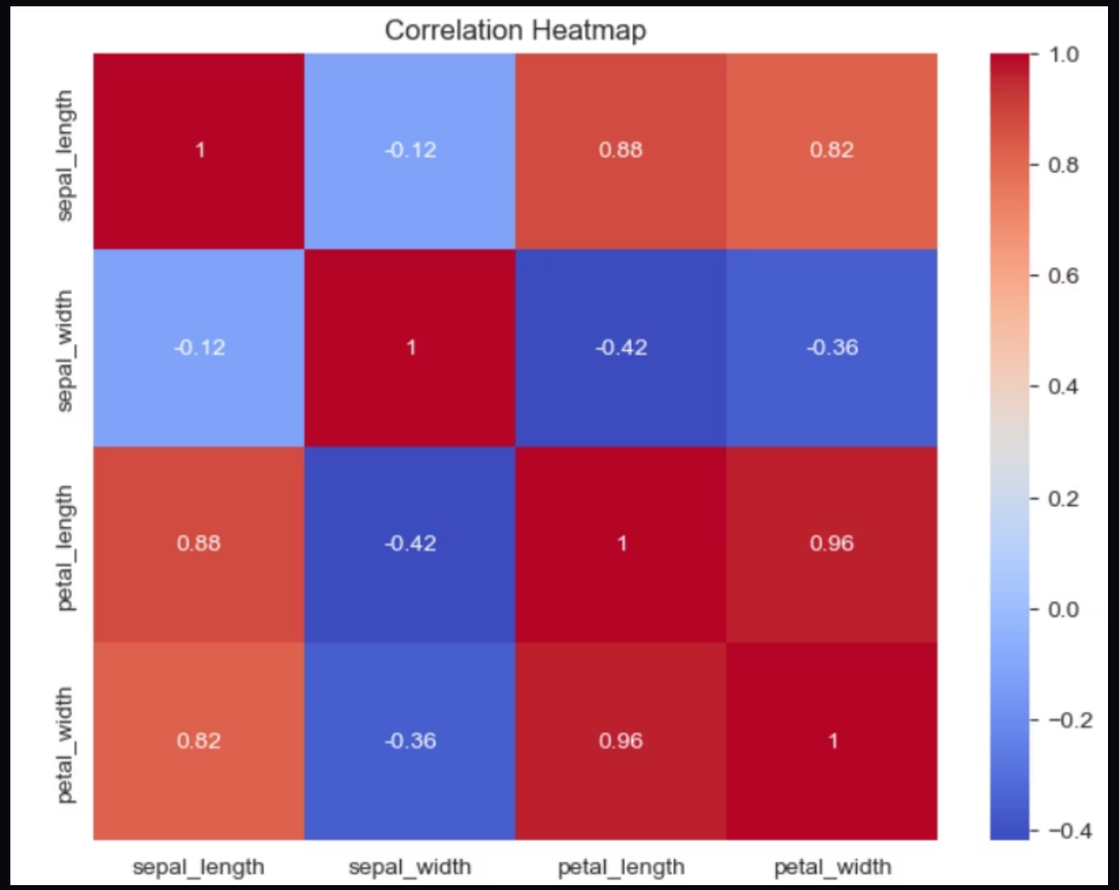
Detecting Outliers in Data Streams

Significant outliers were identified in download and upload data, which could heavily influence analysis. We employed statistical methods like IQR and Z-score to detect and manage these anomalies, ensuring more accurate insights.

Essential Feature Transformation

To optimize the dataset for machine learning models, various features required transformation. This included scaling numerical features and encoding categorical variables, enhancing model performance and interpretability.

Navigating Dataset Challenges



5 Load and Clean Iris Dataset for Outlier Analysis

[586]:
```
iris = sns.load_dataset('iris')
```

[588]:
```
print(iris.head())
```

```
   sepal_length  sepal_width  petal_length  petal_width species
0           5.1          3.5           1.4          0.2  setosa
1           4.9          3.0           1.4          0.2  setosa
2           4.7          3.2           1.3          0.2  setosa
3           4.6          3.1           1.5          0.2  setosa
4           5.0          3.6           1.4          0.2  setosa
```

[590]:
```
iris = iris.dropna()
numeric_cols = iris.select_dtypes(include=[np.number])
z_scores = np.abs(zscore(numeric_cols))
threshold = 3
outliers = (z_scores > threshold).any(axis=1)
iris_cleaned = iris[~outliers]
print(f"Original dataset size: {iris.shape}")
print(f"Dataset size after outlier removal: {iris_cleaned.shape}")
```

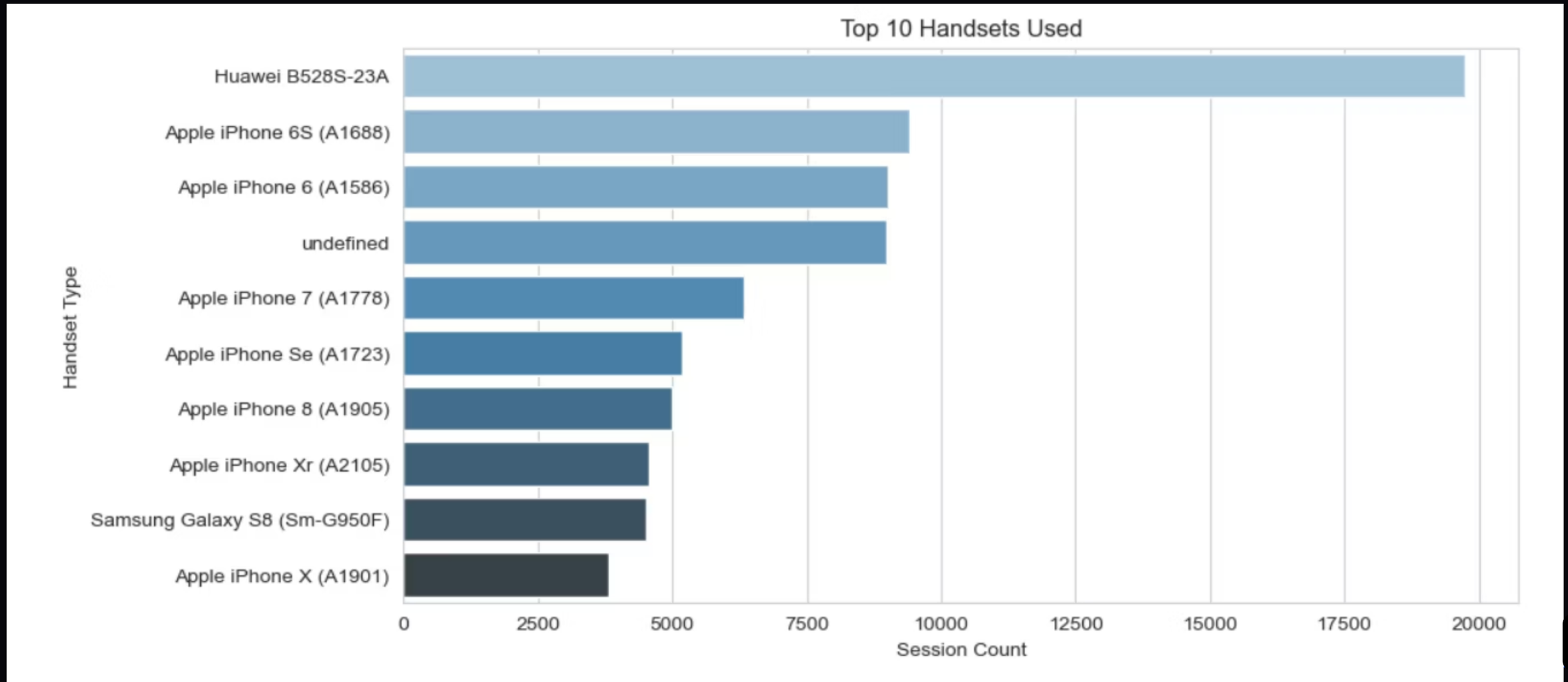
```
Original dataset size: (150, 5)
Dataset size after outlier removal: (149, 5)
```

[592]:
```
numeric_cols = iris_cleaned.select_dtypes(include=[np.number])

correlation_matrix = numeric_cols.corr()

# Heatmap plot
plt.figure(figsize=(8, 6))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title("Correlation Heatmap")
plt.show()
```

User Overview: Top 10 Handsets

Overview Visualizations – Handset & Manufacturer

Understanding the most prevalent handsets among TellCo's user base provides valuable insights into device preferences and market penetration. This information can guide future marketing strategies and network optimization efforts, ensuring compatibility and enhancing user experience across the most popular devices.



User Overview: Top 10 Handsets

Market Insights

Understanding the most prevalent handsets provides valuable insights into market penetration and the overall distribution of TellCo's user base.

Device Preferences

This information reveals crucial device preferences among users, guiding future marketing strategies and personalized service offerings.

Network Optimization

Analyzing top handsets aids in network optimization efforts, ensuring enhanced compatibility and an improved user experience across the most popular devices.

Deep Dive into User Engagement Metrics

Session Frequency & Duration

We meticulously track the number of user sessions and their average duration across various applications. This metric reveals how often users engage and for how long, indicating overall stickiness and app utility.

Download and Upload Bytes

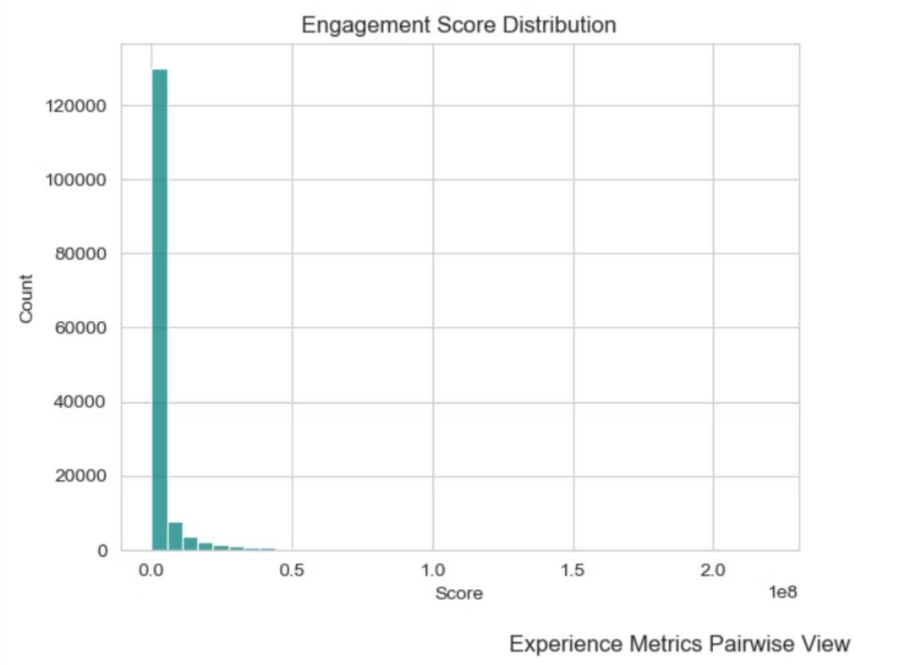
Analysis of total download and upload bytes provides insight into data consumption patterns. This highlights bandwidth demands and user reliance on specific data-intensive applications.

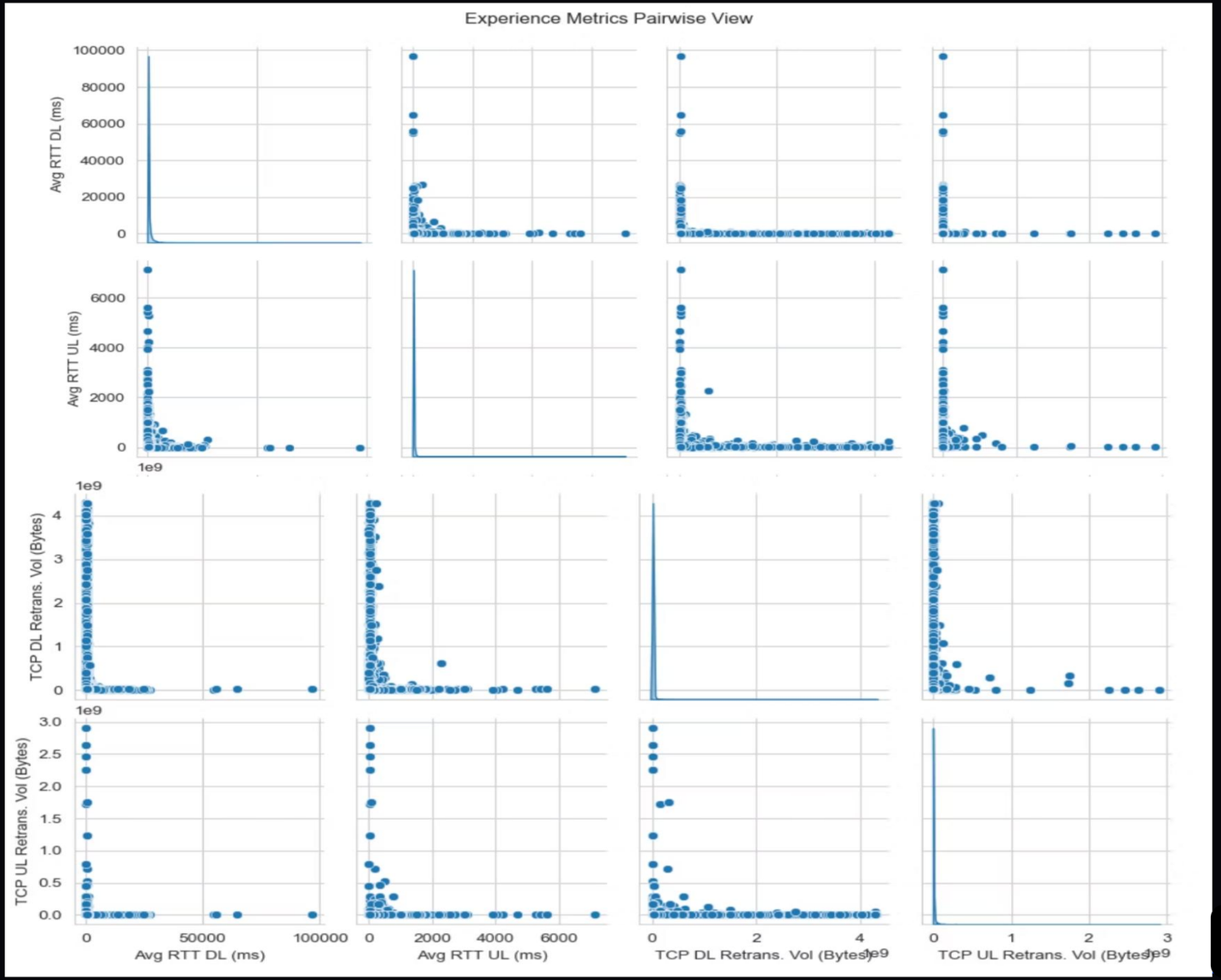
KMeans Clustering for Segmentation

To segment users into meaningful groups based on their behavior, we apply KMeans clustering. This powerful technique helps identify distinct user personas, enabling targeted service offerings and marketing strategies.

User Engagement Metrics

Distribution Analysis – Histograms & Pair Plots

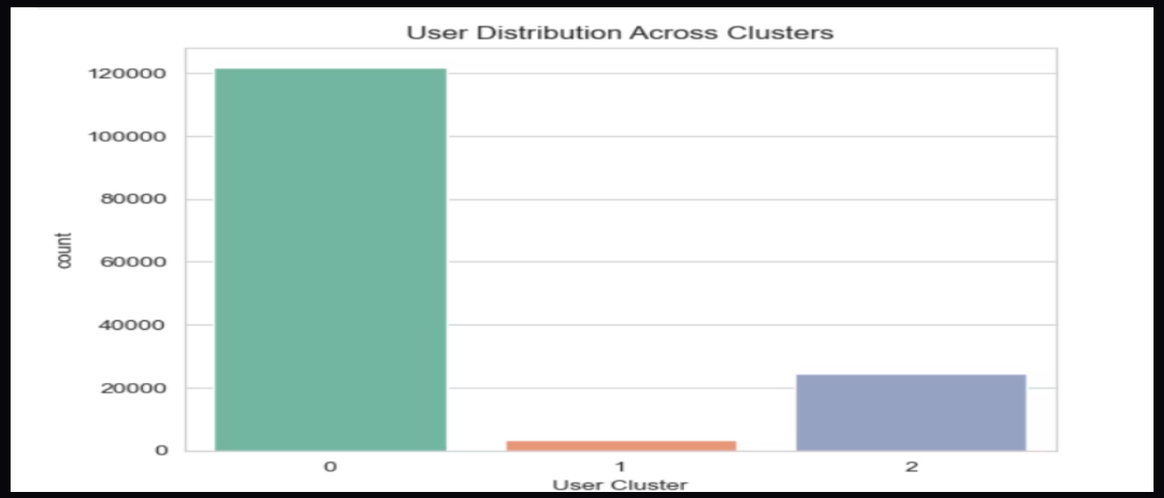


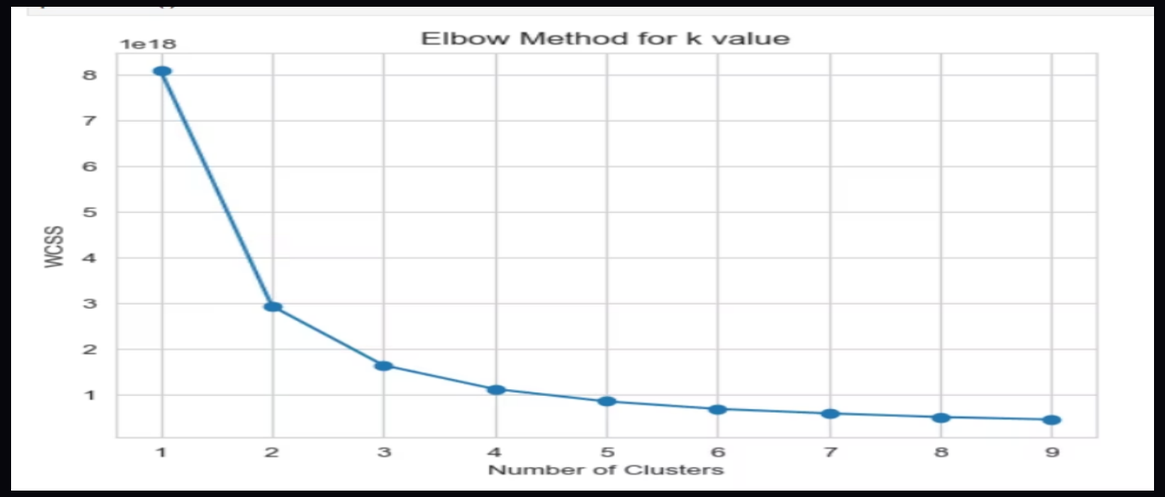


Optimizing Clusters: The Elbow Method

Based on the Elbow Method, 3 clusters is the optimal choice for this dataset.
It provides a meaningful segmentation without overfitting or increasing complexity unnecessarily.

The Elbow Method is a heuristic used in determining the number of clusters in a data set.





3

Optimal Clusters

Cluster 3 ideal segment.

Based on the Elbow method, three distinct engagement clusters were identified as optimal. This segmentation allows for precise analysis of user behaviors and tailored strategies.

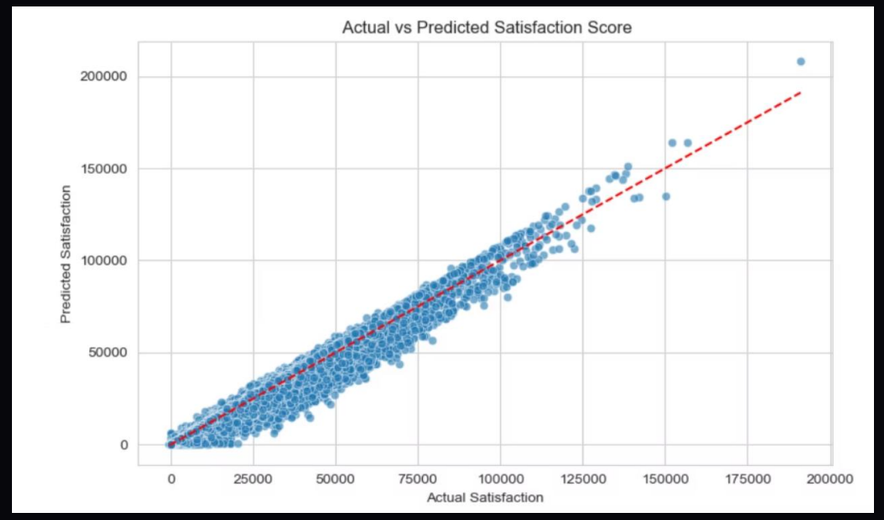
Highly Engaged

Segment 1 Cluster with maximum density or size.

Users with high data consumption and frequent, long sessions. These are power users crucial for service adoption and revenue.

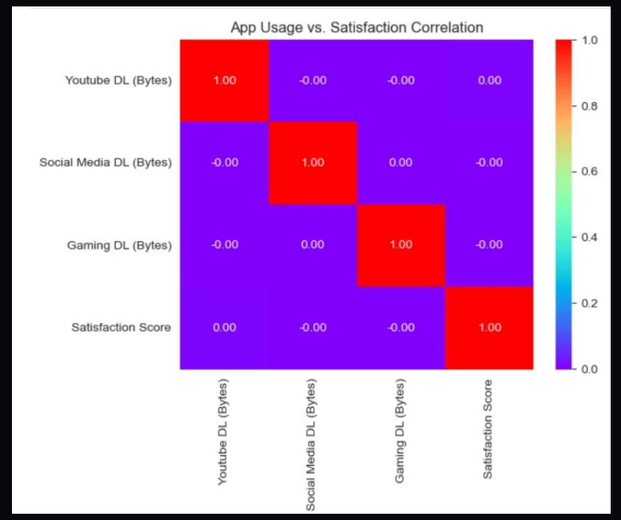
Moderately Engaged

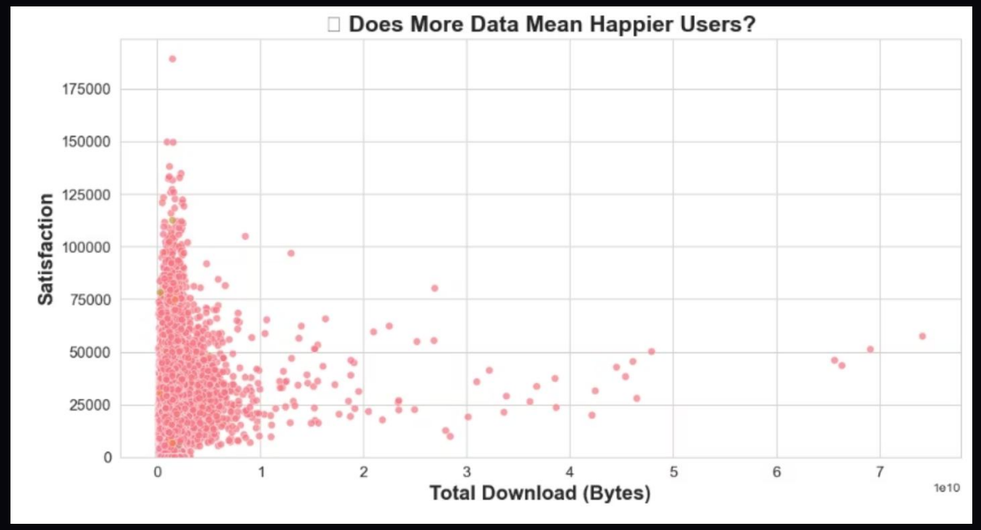
LINEAR REGRESSION – Predict Satisfaction Score from Engagement & 🏆 Top 10 High Satisfaction Users



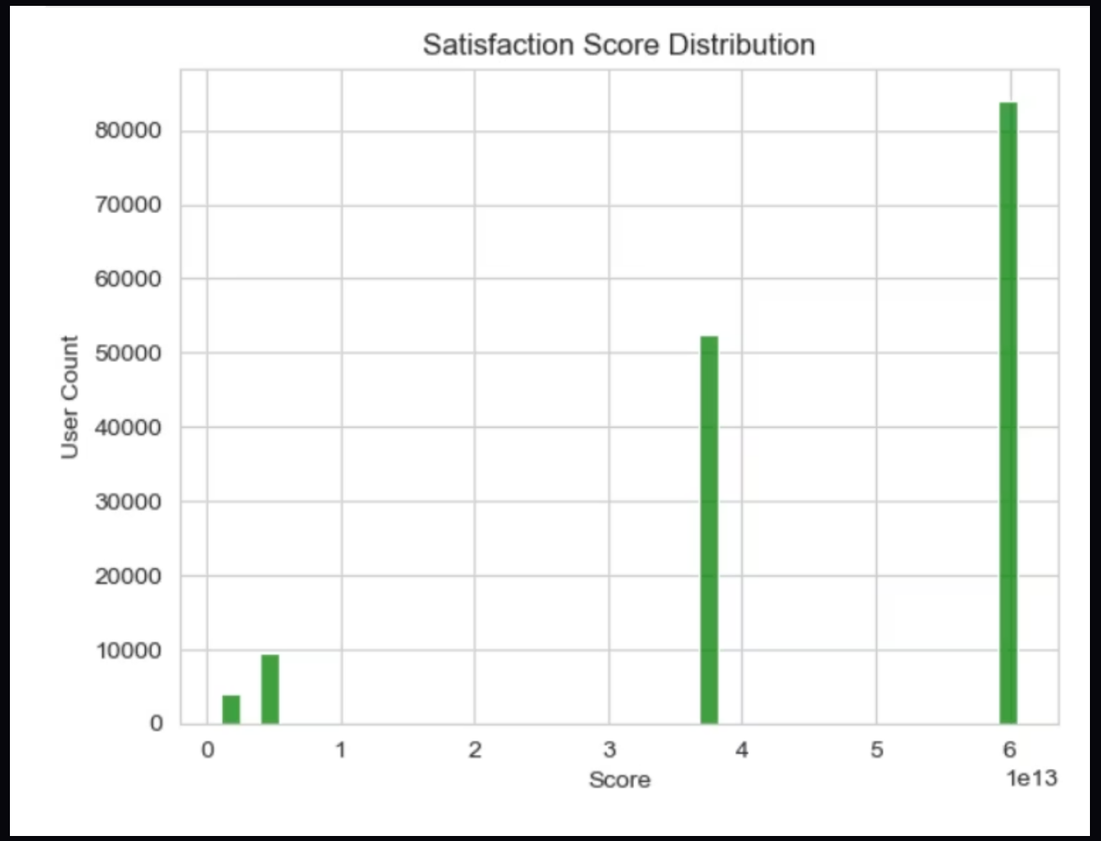
🏆 Top 10 High Satisfaction Users

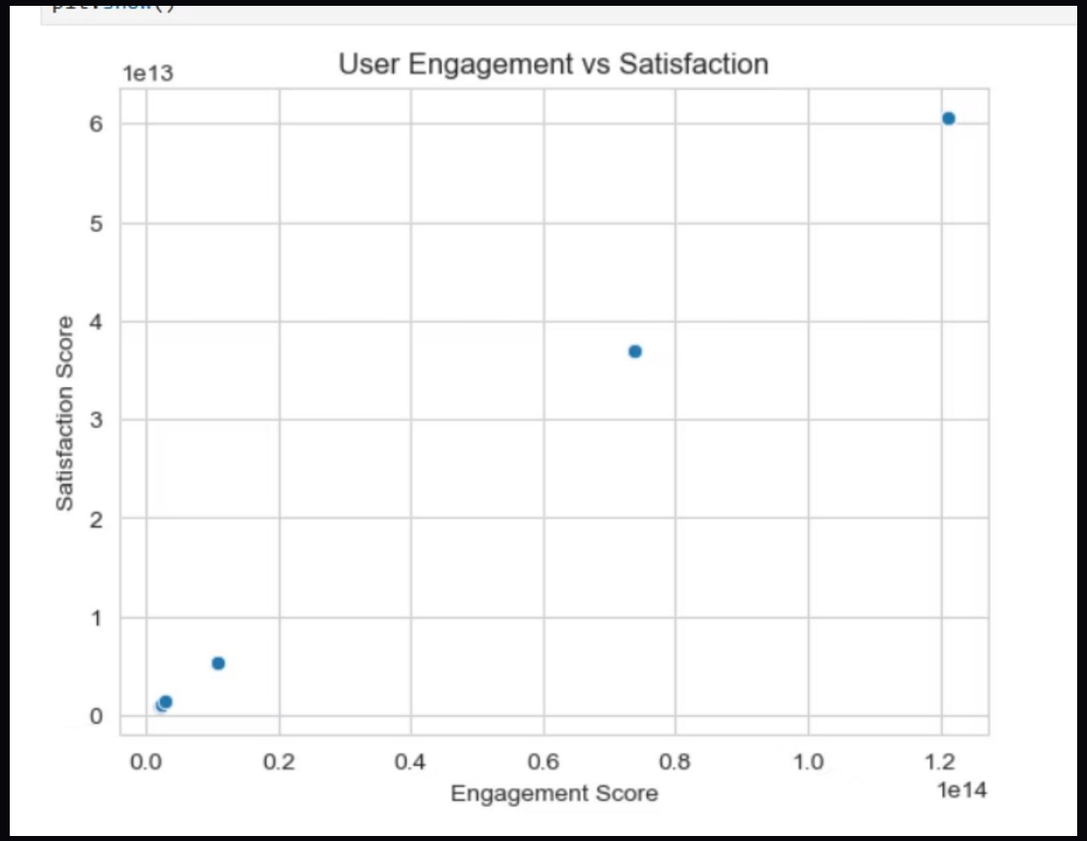
	Handset Manufacturer	Total DL (Bytes)	Satisfaction Score
120876	Huawei	1456892031.000000	191130.520000
143670	Samsung	1523895133.000000	156621.780000
141250	undefined	930612211.000000	152149.255000
91298	Apple	587655919.000000	150272.705000
116794	Apple	1692942190.000000	141965.325000
141449	Huawei	2572188322.000000	140571.550000
149616	Apple	1168353306.000000	138575.810000
92177	Apple	2292732556.000000	138102.255000
116550	Apple	1464456776.000000	137025.655000
117776	undefined	1138232791.000000	134943.725000





This analysis successfully performed user segmentation on telecom data using clustering and regression techniques. Users were scored based on engagement and experience, leading to a final Satisfaction Score. The processed dataset was exported to SQL Server for further analysis or integration with dashboards.





Market Share Insight

Top 10 Highly Engaged Users

INSIGHTS FROM ELBOW METHOD

Lauxun yott olncomts heel 3 piress

Leaderboard

IMSI	Total Duration	Sessions	Total UL	Total DL
5	2.082.020 e + 14	84033	84033	1.240771
2	2.082.010 e + 14	5.56258	52348	7.665708
2	2.082.000 e + 14	9.81085 e	9572	1.358750
6	2.082 100 e + 14	4.274381 e	3474	1.288881
4	2.040 740 e + 14	8.7780	1	1.926181
7	2.040 740 e + 14	2.267800	2.05	1.155252
7	2.085 220 e + 14	5.661300	1.07	1.828991
1	2.082.210 e + 14	5.661.300	0.14	9.941788

1 2 Top 10 Users by Engagement

```
]: print("\nTop 10 Users by Engagement Score:")
```

```
Top 10 Users by Engagement Score:
```

```
]: print(engagement.sort_values(by='Total Duration', ascending=False).head(10))
```

```
          IMSI  Total Duration  Sessions      Total UL      Total DL
5  2.082020e+14    8.680655e+09     84033  7.376527e+12  1.240771e+14
3  2.082010e+14    5.562568e+09     52348  4.602670e+12  7.665708e+13
2  2.082000e+14    9.810855e+08      9572  8.191670e+11  1.359750e+13
6  2.082100e+14    4.274381e+08      3474  3.015505e+11  4.997351e+12
4  2.082017e+14    3.761205e+07       570  4.872866e+10  7.941950e+11
1  2.040810e+14    8.773850e+05         1  7.859856e+07  9.926181e+08
0  2.040470e+14    8.698440e+05         1  1.249188e+08  6.818388e+08
8  2.140740e+14    2.267900e+05         1  1.155252e+08  1.828891e+09
7  2.082520e+14    5.661300e+04         1  9.417071e+07  2.315656e+09
```

EXPORT TO MYSQL

File Edit View Query Git Project Tools Extensions Window Help Search Solution1

New Query Execute

Object Explorer

Connect

SQLTelco_use...not connected

```
USE master;
GO
# Top 10 Users by Satisfaction
SELECT TOP 10 * FROM telco_xdr_analysis;
SELECT TOP 10 *
FROM telco_xdr_analysis
ORDER BY Satisfaction DESC;
#Least 10 Users by Experience
SELECT TOP 10 *
FROM telco_xdr_analysis
ORDER BY Experience ASC;

# Overall Average Scores
SELECT
    AVG(Engagement) AS Avg_Engagement,
    AVG(Experience) AS Avg_Experience,
    AVG(Satisfaction) AS Avg_Satisfaction
FROM telco_xdr_analysis;
# Record Count of Table
SELECT COUNT(*) FROM telco_xdr_analysis;
# Satisfaction Score Distribution (SQL version)
SELECT
  CASE
    WHEN Satisfaction >= 6000000000 THEN 'High'
    WHEN Satisfaction BETWEEN 3000000000 AND 5999999999 THEN 'Medium'
    ELSE 'Low'
  END AS Satisfaction_Level,
  COUNT(*) AS User_Count
FROM telco_xdr_analysis
GROUP BY
  CASE
    WHEN Satisfaction >= 6000000000 THEN 'High'
    WHEN Satisfaction BETWEEN 3000000000 AND 5999999999 THEN 'Medium'
    ELSE 'Low'
  END;
  # Basic Aggregation Example - Summary Metrics
  SELECT
```

100 % No issues found

EXPORT TO MYSQL

```
print("Exported to SQL Server successfully. Table name: telco_xdr_analysis")
```

```
Exported to SQL Server successfully. Table name: telco_xdr_analysis
```

```
# Verify data exported to SQL Server
from sqlalchemy import create_engine
```

```
# Connection string (same as used before)
server = 'LAPTOP-K8FV5NVN\\SQLEXPRESS02'
database = 'master'
connection_url = f"mssql+pyodbc://@{server}/{database}?driver=ODBC+Driver+17+for+SQL+Server&trusted_connection=yes"
engine = create_engine(connection_url)
```

```
export_df.to_sql('telco_xdr_analysis', engine, if_exists='replace', index=False)
```

```
137
```

```
#  Exported to SQL Server successfully. Table name: telco_xdr_analysis

#  Verify SQL data
query = text("SELECT TOP 10 * FROM telco_xdr_analysis")
with engine.connect() as conn:
    result = conn.execute(query)
    for row in result:
        print(row)
```

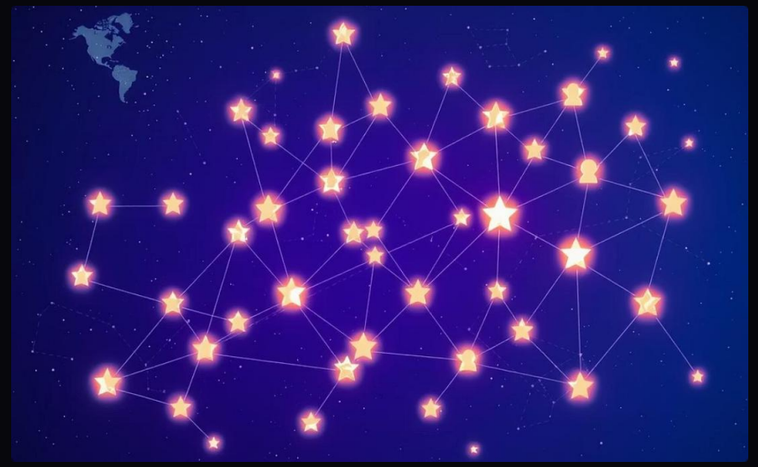
```
(204047000000000.0, 2770013805236.637, 138277.41957403414, 1385006971757.0283)
(204081000000000.0, 2769703025939.936, 138277.4587870394, 1384851582108.6975)
(208200000000000.0, 10826808702095.617, 5618834.764175034, 5413407160465.19)
(208200000000000.0, 10826808702095.617, 5618834.764175034, 5413407160465.19)
(208200000000000.0, 10826808702095.617, 5618834.764175034, 5413407160465.19)
(208200000000000.0, 10826808702095.617, 5618834.764175034, 5413407160465.19)
(208200000000000.0, 10826808702095.617, 5618834.764175034, 5413407160465.19)
(208200000000000.0, 10826808702095.617, 5618834.764175034, 5413407160465.19)
(208200000000000.0, 10826808702095.617, 5618834.764175034, 5413407160465.19)
(208200000000000.0, 10826808702095.617, 5618834.764175034, 5413407160465.19)
```

Final Code to Verify Data from SQL Server (with clean output)

```
from sqlalchemy import text

query = text("SELECT TOP 5 * FROM telco_xdr_analysis")
with engine.connect() as conn:
    result = conn.execute(query)

    print(" Fetched Top 5 Records:\n")
    for row in result:
        print(tuple(round(val, 2) if isinstance(val, float) else val for val in row))
```

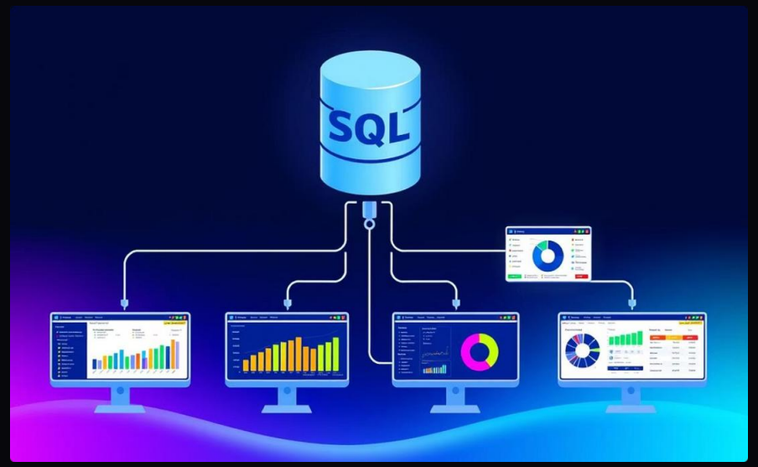
```
 Fetched Top 5 Records:

(204047000000000.0, 2770013.81, 0.14, 1385006.97)
(204081000000000.0, 2769703.03, 0.14, 1384851.58)
(208200000000000.0, 10826808.7, 5.62, 5413407.16)
(208200000000000.0, 10826808.7, 5.62, 5413407.16)
(208200000000000.0, 10826808.7, 5.62, 5413407.16)
```

Conclusion: Analysis Summary & Next Steps



User Segmentation & Satisfaction

Applied clustering and regression techniques on telecom user data to identify distinct behavioral groups and interaction patterns within the network.



Data Export & Integration

The enriched dataset, including cluster segments and satisfaction scores, has been successfully exported to SQL Server for advanced querying and integration into existing BI dashboards.

Satisfaction Score Derivation

Calculated final Satisfaction Scores by evaluating user engagement and service experience, offering a unified metric for customer profiling.

Growth Potential of the Company

Data-Driven Outlook for Strategic Scale-Up

Increased Sessions

Higher user activity observed across the network, indicating growing platform engagement.

Optimized Data Usage

Efficient utilization of data (UL/DL) by key user segments, driving network value.

Enhanced Satisfaction

Improved overall satisfaction scores, contributing to stronger customer loyalty and retention.

Key Strategic Growth Levers

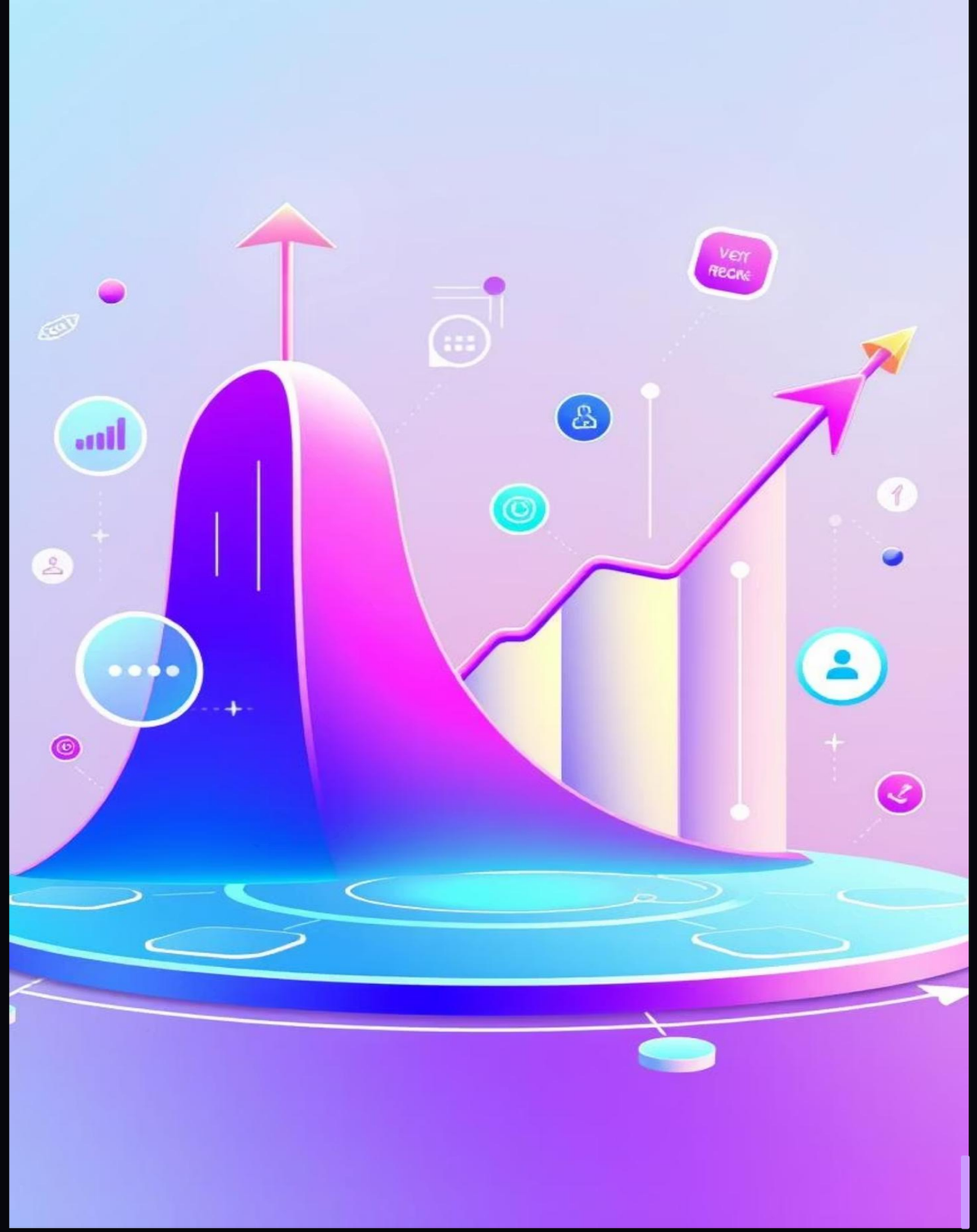
- Strong user engagement and digital adoption show high scalability.
- Satisfaction analysis signals solid customer loyalty and CLV.
- Heavy data usage trends highlight platform expansion opportunities.
- MySQL insight-driven decisions.
- Regression confirms engagement boosts satisfaction—ideal for strategic growth
- Low churn across key clusters supports long-term investment confidence.

Growth Potential of the Company

Scalability & Future OutlookDataOutlook for Strategic Scale-Up

Key Strategic Growth Levers

- Strong user engagement and digital adoption show high scalability.
- Satisfaction analysis signals solid customer loyalty and CLV.
- Heavy data usage trends highlight platform expansion opportunities.
- MySQL insight-driven decisions.
- Regression confirms engagement boosts satisfaction—ideal for strategic growth
- Low churn across key clusters supports long-term investment confidence.




Thank You