“ SUMMARY OF NEXTHIKE PROJECT 5”

#Summary of Telecom Data Analysis Project

# Data Preparation and Exploration

Data Loading and Initial Inspection:

- Loaded ‘telcom\_data.csv’ (telecom usage) and ‘field\_data.csv’ (field descriptions).

- Inspected dataset structures and formats using ‘info()’ and previewed initial rows.

#Aggregating User Behavior Metrics:

- Summarized metrics by user using ‘groupby()’ and ‘agg()’.

- Metrics included total call duration, total download/upload bytes, and averages for RTT, throughput, and application-specific data.

# Exploratory Data Analysis (EDA)

Univariate Analysis:

- Computed descriptive statistics (mean, median, etc.) for key variables like call duration and throughput.

- Created histograms to visualize distributions.

#Bivariate Analysis:

- Plotted scatter plots to examine relationships, such as between average bearer throughput and total download bytes.

#Variable Transformations:

- Categorized call duration into segments: Short, Medium, Long, Very Long.

#Correlation and Dimensionality Reduction:

- Computed a correlation matrix to identify relationships between variables.

- Applied PCA to reduce dimensionality and transform features into principal components.

# User Engagement Analysis

Aggregating and Normalizing Metrics:

- Aggregated engagement metrics (e.g., total bytes by application type) and standardized them using ‘StandardScaler’.

#K-Means Clustering:

- Used K-Means to cluster users into three groups based on engagement metrics.

- Analyzed cluster characteristics and determined the optimal number of clusters using the Elbow Method (k from 5 to 8).

#Top Engaged Users:

- Identified users with highest total download and application-specific usage.

# Experience Analytics

Aggregating Experience Metrics:

- Evaluated user experience metrics, including average RTT, throughput, and handset details.

#Top/Bottom/Frequent Values:

- Analyzed RTT values to identify latency issues.

#Throughput Distribution:

- Used box plots to compare throughput distribution across different handsets.

#Handling Missing Values:

- Addressed missing data with mean imputation and dropped incomplete rows before reapplying PCA and K-Means.

# Key Findings

1. User Segmentation:

- Identified different usage patterns and engagement levels through segmentation.

2. Performance Metrics:

- Analyzed how performance metrics, such as throughput and download bytes, affect user behavior.

3. Experience Evaluation:

- Found variations in throughput and latency based on handset type.

# Recommendations

- Targeted Interventions:

- Customize offers and performance improvements based on user cluster characteristics.

- Handset Optimization:

- Address performance issues identified for specific handsets to improve user experience.

- Continuous Monitoring:

- Implement ongoing analysis to monitor changes in user behavior and experience over time.