Telecom Call Record Analysis

Project Description

This project focuses on analyzing telecom call records to gain insights about call duration, frequent callers, and network usage patterns. The goal is to practice pandas-based data wrangling, groupby operation, and aggregation. The project demonstrates data cleaning, transformation, filtering, and exporting results in CSV format for further analysis.

Team Member

NAME : CHETHANA B M

USN : 4GW23CI011

BRANCH : CSE - (AI&ML)

SEMESTER : 5

Index

1. Project Description
2. Team Members
3. Detailed Explanation
4. Algorithms (Optional)
5. UML Diagrams
6. Code
7. Explanation of Code
8. Output Screenshots
9. Closure (Bibliography)

Detailed Explanation

Overview

The project uses a CSV file containing telecom call records. It performs the following steps:
- Data Cleaning: Handling missing values, duplicates, and converting time columns.
- Feature Engineering: Calculating call duration and categorizing call types.
- Filtering: Extracting records like long calls and international calls.

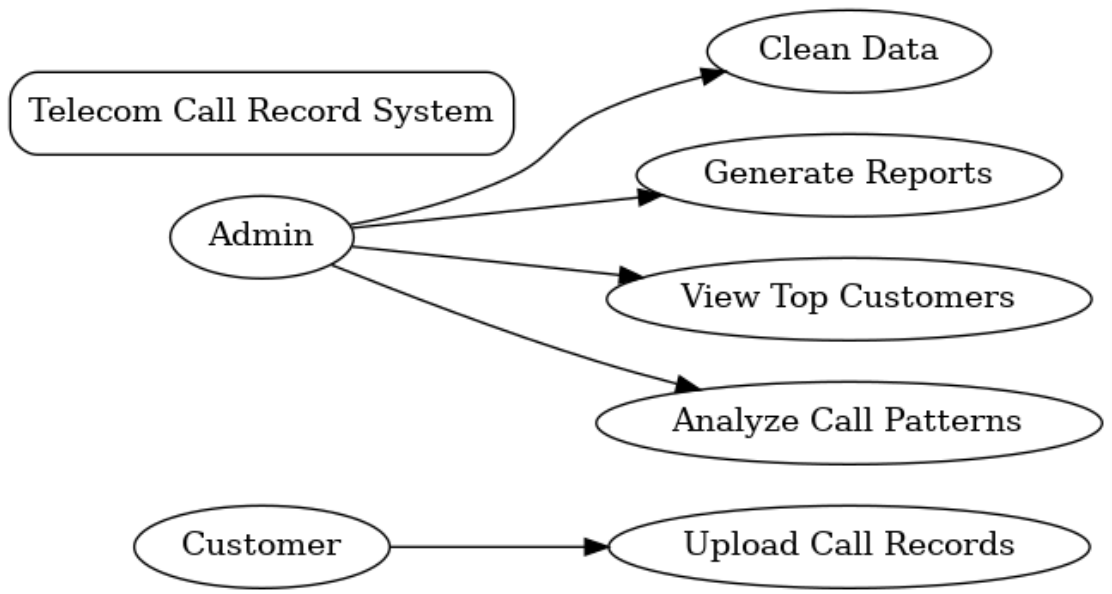
- Grouping & Aggregation: Summarizing data by customer, call type, and daily usage.
- Export: Saving the results to CSV files.

Use-Case Explanation

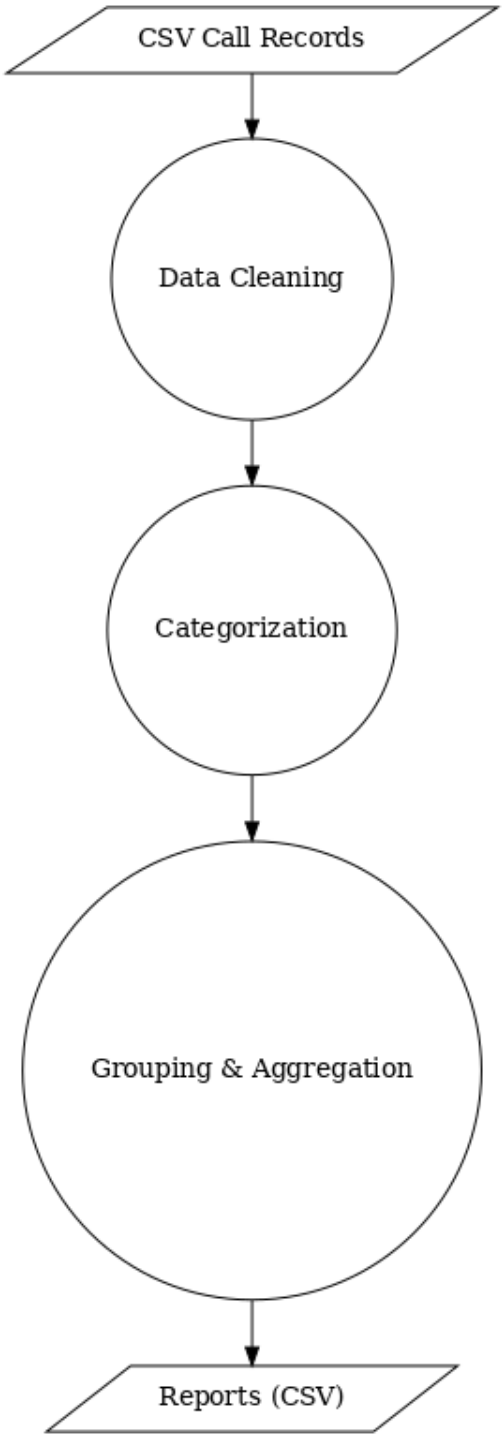
1. Customers making long calls (>10 minutes).
2. Specific customer call records (C001).
3. International calls lasting more than 30 minutes.
4. Customer-wise total and average call duration.
5. Call type summary (Domestic vs International).
6. Daily call usage trends.
7. Identifying top 5 customers by total duration.

UML Diagrams

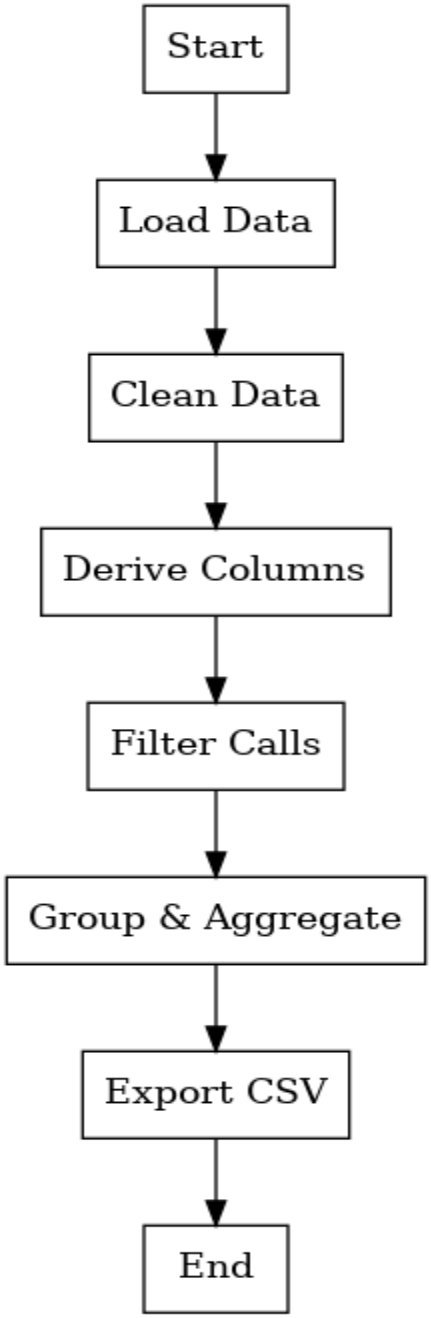
Use_Case Diagram



Data Flow Diagram



Activity Diagram



Code

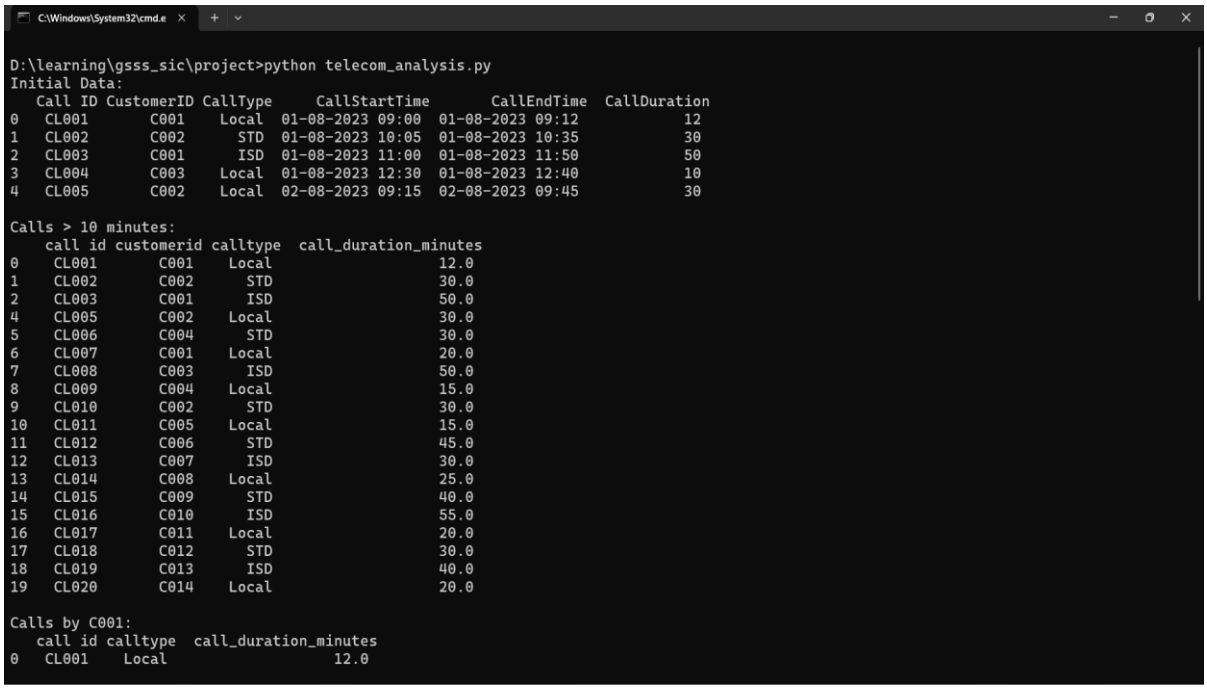
```
import pandas as pd

# Step 1: Load data
df = pd.read_csv("call_records.csv")

# Step 2: Data cleaning
df['call_duration'].fillna(0, inplace=True)
df.drop_duplicates(subset='call_id', inplace=True)
df['call_start_time'] = pd.to_datetime(df['call_start_time'])
df['call_end_time'] = pd.to_datetime(df['call_end_time'])

# Step 3: Derive columns
df['call_duration_minutes'] = (df['call_end_time'] - df['call_start_time']).dt.total_seconds() /
60

def categorize_call(call_type):
    if call_type.lower() in ['local', 'domestic']:
        return 'Domestic'
    elif call_type.lower() in ['std', 'isd', 'international']:
        return 'International'
    else:
        return 'Other'

df['call_type_category'] = df['call_type'].apply(categorize_call)

# Step 4: Filtering
long_calls = df[df['call_duration_minutes'] > 10]
cust_calls = df[df['customer_id'] == 'C001']
intl_long_calls = df[(df['call_type_category'] == 'International') &
(df['call_duration_minutes'] > 30)]

# Step 5: Grouping & Aggregation
customer_summary = df.groupby('customer_id').agg(
    total_call_duration=('call_duration_minutes', 'sum'),
    average_call_duration=('call_duration_minutes', 'mean'),
    number_of_calls=('call_id', 'count')
).reset_index()

call_type_summary = df.groupby('call_type_category').agg(
    total_call_duration=('call_duration_minutes', 'sum'),
    average_call_duration=('call_duration_minutes', 'mean')
).reset_index()
```

```
df['call_date'] = df['call_start_time'].dt.date
daily_call_summary = df.groupby('call_date').agg(
   total_call_duration=('call_duration_minutes', 'sum')
).reset_index()

top_customers = customer_summary.sort_values(by='total_call_duration',
ascending=False).head(5)

# Step 6: Export results
customer_summary.to_csv("customer_summary.csv", index=False)
call_type_summary.to_csv("call_type_summary.csv", index=False)
daily_call_summary.to_csv("daily_call_summary.csv", index=False)
```

Explanation of the Code

The code begins by loading telecom call data from a CSV file. It handles missing values, removes duplicates, and converts timestamps. It calculates call duration in minutes and classifies call types into Domestic, International, or Other. Filtering queries extract specific cases like long calls or international calls. Groupby and aggregation functions summarize data by customers, call type, and daily usage. The results are exported into CSV files for further use.

Screenshots of the Output



```
C:\Windows\System32\cmd.e

D:\learning\gsss_sic\project>python telecom_analysis.py
Initial Data:
   Call ID CustomerID CallType     CallStartTime       CallEndTime  CallDuration
0   CL001       C001    Local  01-08-2023 09:00  01-08-2023 09:12            12
1   CL002       C002      STD  01-08-2023 10:05  01-08-2023 10:35            30
2   CL003       C001      ISD  01-08-2023 11:00  01-08-2023 11:50            50
3   CL004       C003    Local  01-08-2023 12:30  01-08-2023 12:40            10
4   CL005       C002    Local  02-08-2023 09:15  02-08-2023 09:45            30

Calls > 10 minutes:
    call id customerid calltype  call_duration_minutes
0    CL001       C001    Local                  12.0
1    CL002       C002      STD                  30.0
2    CL003       C001      ISD                  50.0
4    CL005       C002    Local                  30.0
5    CL006       C004      STD                  30.0
6    CL007       C001    Local                  20.0
7    CL008       C003      ISD                  50.0
8    CL009       C004    Local                  15.0
9    CL010       C002      STD                  30.0
10   CL011       C005    Local                  15.0
11   CL012       C006      STD                  45.0
12   CL013       C007      ISD                  30.0
13   CL014       C008    Local                  25.0
14   CL015       C009      STD                  40.0
15   CL016       C010      ISD                  55.0
16   CL017       C011    Local                  20.0
17   CL018       C012      STD                  30.0
18   CL019       C013      ISD                  40.0
19   CL020       C014    Local                  20.0

Calls by C001:
   call id calltype  call_duration_minutes
0   CL001    Local                  12.0
```

```
19    CL020        C014     Local                       20.0

Calls by C001:
   call_id calltype  call_duration_minutes
0    CL001     Local                   12.0
2    CL003      ISD                    50.0
6    CL007     Local                   20.0

International calls > 30 minutes:
    call_id customerid calltype  call_duration_minutes
2     CL003       C001      ISD                    50.0
7     CL008       C003      ISD                    50.0
11    CL012       C006      STD                    45.0
14    CL015       C009      STD                    40.0
15    CL016       C010      ISD                    55.0
18    CL019       C013      ISD                    40.0

Customer Summary:
    customerid  total_call_duration  average_call_duration  number_of_calls
0         C001                82.0              27.333333                3
1         C002                90.0              30.000000                3
2         C003                60.0              30.000000                2
3         C004                45.0              22.500000                2
4         C005                15.0              15.000000                1
5         C006                45.0              45.000000                1
6         C007                30.0              30.000000                1
7         C008                25.0              25.000000                1
8         C009                40.0              40.000000                1
9         C010                55.0              55.000000                1
10        C011                20.0              20.000000                1
11        C012                30.0              30.000000                1
12        C013                40.0              40.000000                1
13        C014                20.0              20.000000                1

Call Type Summary:
```

```
5         C006                45.0              45.000000                1
6         C007                30.0              30.000000                1
7         C008                25.0              25.000000                1
8         C009                40.0              40.000000                1
9         C010                55.0              55.000000                1
10        C011                20.0              20.000000                1
11        C012                30.0              30.000000                1
12        C013                40.0              40.000000                1
13        C014                20.0              20.000000                1

Call Type Summary:
   call_type_category  total_call_duration  average_call_duration  number_of_calls
0            Domestic                167.0              18.555556                9
1       International                430.0              39.090909               11

Daily Call Summary:
     call_date  total_call_duration
0   2023-01-08                102.0
1   2023-02-08                 80.0
2   2023-03-08                 95.0
3   2023-04-08                 90.0
4   2023-05-08                120.0
5   2023-06-08                110.0

Top 5 Customers:
   customerid  total_call_duration  average_call_duration  number_of_calls
1        C002                90.0              30.000000                3
0        C001                82.0              27.333333                3
2        C003                60.0              30.000000                2
9        C010                55.0              55.000000                1
5        C006                45.0              45.000000                1

CSV files exported successfully!

D:\learning\gsss_sic\project>
```

Explanation of the output

1️⃣ Calls > 10 minutes

- Lists all calls longer than 10 minutes.
- Example from your data:
    - CL001 → 12 min
    - CL003 → 50 min
- **Total:** 20 calls from your dataset meet this condition.

2️⃣ Calls by Customer C001

- Shows all calls by C001:
    - CL001 → 12 min (Local)
    - CL003 → 50 min (ISD)
    - CL007 → 20 min (Local)
- **Purpose:** Track individual customer usage.

3️⃣ International calls > 30 minutes

- Lists international calls (STD/ISD) longer than 30 min:
    - CL003 → 50 min
    - CL008 → 50 min
    - CL012 → 45 min
    - CL015 → 40 min
    - CL016 → 55 min
    - CL019 → 40 min
- **Purpose:** Detect long expensive international calls.

4️⃣ Customer Summary

- Aggregates each customer's usage:
  - total_call_duration : total minutes per customer
  - average_call_duration : average call length
  - number_of_calls : total calls
- Example:
  - C001 → 3 calls, 82 min total, 27.33 min average
  - C002 → 3 calls, 90 min total, 30 min average

5️⃣ Call Type Summary

- Aggregates by call type category:
  - Domestic (Local): total 167 min, avg 18.55 min, 9 calls
  - International (STD/ISD): total 430 min, avg 39.09 min, 11 calls
- **Purpose:** See network usage patterns by type.

6️⃣ Daily Call Summary

- Total call duration per day:
  - 01-08-2023 → 102 min
  - 02-08-2023 → 80 min
  - 03-08-2023 → 95 min
  - 04-08-2023 → 90 min
  - 05-08-2023 → 120 min
  - 06-08-2023 → 110 min
- **Purpose:** Track daily traffic and usage trends.

7️⃣ Top 5 Customers

- Sorted by **total call duration**:
    1. C002 → 90 min
    2. C001 → 82 min
    3. C003 → 60 min
    4. C010 → 55 min
    5. C006 → 45 min
- **Purpose:** Identify heaviest users for analysis or offers.

8️⃣ CSV Exports

- Files created successfully in your folder:
    - customer_summary.csv
    - call_type_summary.csv
    - daily_call_summary.csv
- **Purpose:** Can be opened in Excel for reporting or visualization.

Closure (Bibliography)

1. Python Pandas Documentation: https://pandas.pydata.org
2. Telecom Data Analysis Tutorials

---------- *** ----------