PhonePe Transaction Insights

Project Overview

This project uses real-time digital transaction data from PhonePe Pulse to extract, transform, analyze, and visualize trends in payment categories, user engagement, and geographical usage. It combines ETL, SQL, and an interactive Streamlit dashboard to offer strategic business insights.

Data Extraction & ETL (Python)

Script: scripts/extract_transform_load.py

Purpose:

- Extract JSON files from PhonePe dataset.
- Parse transaction information.
- Load structured data into a SQLite database.

```
# Extract, Transform, Load Transaction Data into SQLite
def load_all_data(conn):
    for year in os.listdir(TRANSACTION_PATH):
        for file in os.listdir(os.path.join(TRANSACTION_PATH, year)):
            file_path = os.path.join(TRANSACTION_PATH, year, file)
            with open(file_path) as f:
                data = json.load(f)
                for txn in data['data']['transactionData']:
                    name = txn['name']
                    count = txn['paymentInstruments'][0]['count']
                    amount = txn['paymentInstruments'][0]['amount']
                    cursor.execute("INSERT INTO aggregated_transaction (...) VALUES
(...) ")
    conn.commit()
```

SQL Queries

Database: dashboard/phonepe.db

```
-- Total transactions by year
SELECT year, SUM(amount) AS total_amount
FROM aggregated_transaction
GROUP BY year
ORDER BY year;


-- Category-wise count
SELECT transaction_type, SUM(count) AS total_count
FROM aggregated_transaction
GROUP BY transaction_type
ORDER BY total_count DESC;
```

Streamlit Dashboard - dashboard/app.py

Features:

- Year selector for filtering
- Bar charts for total transaction value and count by quarter
- Category breakdown

```
selected_year = st.selectbox("Select Year", years["year"])
df = pd.read_sql_query(f"""
   SELECT quarter, transaction_type, SUM(amount) AS total_amount
   FROM aggregated_transaction
   WHERE year = {selected_year}
   GROUP BY quarter, transaction_type
""", conn)
```

Documentation: Process & Insights

Process Flow:

1. **Data Source**: Cloned dataset from PhonePe GitHub Pulse repository.
2. **Directory Setup**: Organized data into structured folders such as aggregated/transaction, aggregated/user, top/map, etc.
3. **Extraction**: Traversed directories by year and state; extracted JSON files for each quarter.
4. **Transformation**: Parsed JSON data into structured rows based on category, type, year, state, and metrics.
5. **Loading**: Loaded cleaned data into a SQLite database with tables like aggregated_transaction, aggregated_user, etc.
6. **SQL Queries**: Formulated to analyze patterns and extract key metrics.
7. **Visualization**: Streamlit dashboard implemented to present visuals and charts based on user selection.
8. **Deployment**: Final app deployed via Streamlit Cloud with shared access.

Insights Extracted:

- ⬆️ **Growth Trend**: Transaction volume and value consistently increase YoY.
- ✉️ **User Behavior**: Users frequently perform recharges and P2P transfers, indicating high mobile usage and interpersonal payments.
- 🌍 **Geographical Variation**: States like Maharashtra, Karnataka, and Andhra Pradesh show higher engagement.
- 🎓 **Quarterly Trends**: Q4 tends to have peak transaction amounts due to festivals and year-end.
- 🔒 **Fraud Potential**: Anomalies in transaction spikes can highlight suspicious activity.

- 🌟 **Category Insights**: Utility payments and insurance purchases show upward trends, indicating potential for value-added services.
- 🌐 **Product/Market Fit**: Certain districts show sharp growth, useful for hyperlocal targeting.