## **Data Science Project: Jio Recharge Dataset – Trend Forecasting**

**Case Study Overview:** Jio, as one of India's largest telecom operators, serves millions of customers daily with prepaid and postpaid recharge plans. Understanding trends in recharge patterns is vital for telecom providers to forecast demand, manage network loads, design effective promotions, and optimize inventory for both physical recharge vouchers and digital offers.

**Objective:** The objective of this project is to analyze historical recharge patterns and forecast future demand by recharge amount, plan type, and region. This will help Jio allocate resources efficiently, tailor marketing campaigns, and reduce churn by proactively managing demand fluctuations.

### **Part 1: Data Generation (Python)**

Let's generate a synthetic dataset simulating Jio's recharge data.

import pandas as pd

import numpy as np

import random

from datetime import datetime, timedelta

# Set a seed for reproducibility

np.random.seed(47)

random.seed(47)

# Define parameters for data generation

num\_recharges = 20000

start\_date = datetime(2022, 1, 1)

end\_date = datetime(2024, 6, 30) # Data up to mid-2024 for forecasting

cities = ['Mumbai', 'Delhi', 'Bangalore', 'Chennai', 'Kolkata', 'Pune', 'Hyderabad', 'Ahmedabad', 'Lucknow', 'Jaipur']

recharge\_types = ['Prepaid', 'Postpaid']

plan\_types = ['Monthly', 'Quarterly', 'Annual', 'Data Add-on', 'Roaming Pack']

payment\_modes = ['UPI', 'Credit Card', 'Debit Card', 'Net Banking', 'Wallet', 'Retail Store']

data = []

for i in range(num\_recharges):

recharge\_id = f'JIOREC{i:06d}'

user\_id = f'USER{random.randint(10000, 99999)}'

# Generate recharge date within the range

recharge\_date = start\_date + timedelta(days=random.randint(0, (end\_date - start\_date).days))

city = random.choice(cities)

recharge\_type = random.choice(recharge\_types)

plan\_type = random.choice(plan\_types)

payment\_mode = np.random.choice(payment\_modes, p=[0.4, 0.2, 0.15, 0.1, 0.1, 0.05]) # UPI most common

# Simulate recharge amounts with some variation based on plan type

if plan\_type == 'Monthly':

recharge\_amount = random.choice([199, 239, 299, 399, 479])

elif plan\_type == 'Quarterly':

recharge\_amount = random.choice([666, 719, 849])

elif plan\_type == 'Annual':

recharge\_amount = random.choice([2545, 2879, 2999])

elif plan\_type == 'Data Add-on':

recharge\_amount = random.choice([19, 29, 61, 121])

else: # Roaming Pack

recharge\_amount = random.choice([499, 599, 799])

# Introduce some seasonality/growth over time

year\_factor = (recharge\_date.year - start\_date.year) \* 0.05 # Small growth per year

month\_factor = (recharge\_date.month % 12) / 12 \* 0.02 # Small monthly fluctuation

recharge\_amount = int(recharge\_amount \* (1 + year\_factor + month\_factor \* random.uniform(-0.5, 0.5)))

data.append([

recharge\_id, user\_id, recharge\_date, city, recharge\_type,

plan\_type, payment\_mode, recharge\_amount

])

df = pd.DataFrame(data, columns=[

'recharge\_id', 'user\_id', 'recharge\_date', 'city', 'recharge\_type',

'plan\_type', 'payment\_mode', 'recharge\_amount'

])

# Ensure recharge\_date is datetime

df['recharge\_date'] = pd.to\_datetime(df['recharge\_date'])

# Display basic info and head

print("Generated Data Info:")

print(df.info())

print("\nGenerated Data Head:")

print(df.head())

# Save the dataset to a CSV file

df.to\_csv('jio\_recharge\_data.csv', index=False)

print("\nDataset 'jio\_recharge\_data.csv' generated successfully!")

### **Part 2: Data Science Tasks for Students**

Students will use the generated jio\_recharge\_data.csv file to perform the following tasks using Python (Pandas, NumPy, Matplotlib, Seaborn, and potentially statsmodels or pmdarima for forecasting).

**Task 1: Data Loading & Initial Exploration**

* Load the jio\_recharge\_data.csv into a Pandas DataFrame.
* Display the first 5 rows and check info().
* Identify the number of rows and columns.
* Get describe() statistics for numerical columns (recharge\_amount).
* Check for any missing values.

**Task 2: Data Cleaning & Preparation**

* **Handle Missing Values:** Address any NaN values (e.g., in recharge\_amount or recharge\_date).
* **Data Type Conversion:** Ensure recharge\_date is a datetime object.
* **Data Consistency:**
  + Check for duplicate records or invalid user IDs (if any are evident from user\_id patterns).
  + Handle outliers in recharge\_amount (e.g., extremely high or low values). Discuss your strategy.
* **Feature Engineering:**
  + Extract year, month, day\_of\_week, and week\_of\_year from recharge\_date.
  + Create a is\_weekend (boolean) column.

**Task 3: Trend Analysis & Forecasting**

* **Recharge Volume Trends:**
  + Aggregate recharge\_amount by recharge\_date (daily, weekly, monthly total).
  + Visualize recharge\_amount over time using line plots to detect seasonality or growth trends.
* **Patterns by Recharge Type:**
  + Analyze recharge volume and average recharge amount for Prepaid vs. Postpaid users.
* **Popular Plan Types and Payment Modes:**
  + Visualize the distribution of plan\_type and payment\_mode.
  + Analyze how payment\_mode usage trends over time.
* **City-wise Recharge Frequency:**
  + Analyze and visualize recharge frequency and total recharge amount by city. Identify cities with the highest recharge activity and fastest growth.
* **Time Series Decomposition (Optional but Recommended):**
  + Use statsmodels.tsa.seasonal.seasonal\_decompose to decompose the monthly or weekly recharge volume into trend, seasonality, and residual components. Visualize these components.
* **Simple Forecasting Models (Optional):**
  + Build simple forecasting models (e.g., Moving Average, Exponential Smoothing) on monthly recharge volume.
  + (Advanced Optional) Explore more sophisticated models like ARIMA or Prophet for forecasting recharge demand for the next 1-3 months by key segments (e.g., overall, by plan\_type, or by city). Evaluate model performance.

**Task 4: Key Findings & Business Recommendations**

* Based on your analysis and any forecasting, identify at least **three key insights** for Jio. For each insight:
  + Clearly state what you observed (e.g., "Recharge volumes show a clear seasonal peak during festival months, with a steady growth trend year-over-year.").
  + Explain what it means (e.g., "This implies predictable periods of high demand, requiring proactive network and resource planning.").
  + Propose 2-3 specific business suggestions (e.g., "Implement targeted promotions during low-demand periods to balance network load," "Prepare network infrastructure upgrades in fast-growing cities based on forecasted demand," "Incentivize preferred payment modes to reduce transaction costs," "Design new combo plans based on trending recharge amounts and popular plan types.").
* Summarize the core patterns and forecasts discovered. Explain how this data supports Jio's strategic planning and customer satisfaction.

### **Part 3: Optional Visualization using Power BI**

For students interested in Business Intelligence tools, you can also use Power BI to create an interactive dashboard based on the jio\_recharge\_data.csv dataset.

**Suggested Power BI Tasks:**

* Import the jio\_recharge\_data.csv into Power BI.
* Create calculated columns and measures (using DAX) such as Total Recharge Amount, Recharge Count, Monthly Average Recharge.
* Design an interactive dashboard with the following visualizations:
  + KPIs: Total Recharge Amount, Total Recharge Count, Average Recharge Value.
  + Line chart: Monthly/Quarterly Total Recharge Amount Trend.
  + Bar chart: Recharge Volume by Plan Type.
  + Bar chart: Recharge Volume by Payment Mode.
  + Map/Bar chart: Recharge Volume by City.
  + Forecast visualization (if Power BI's built-in forecasting features are used).
  + Slicers for city, recharge\_type, plan\_type, payment\_mode, and year/month.