```
import pdfplumber
import re
import pandas as pd

# Path to the PDF file
file_path = 'test3 (1) (1).pdf'

# Initialize data lists
dates, descriptions, credits, debits, balances = [], [], [], [], []

# Open and parse the PDF
with pdfplumber.open(file_path) as pdf:
    for page in pdf.pages:
        text = page.extract_text()
        lines = text.split('\n')

        for line in lines:
            # Match lines with financial entries using regex
            match = re.match(r'(\d{2}-\w{3}-
\d{4})\s+[TC]\s+(.+?)\s+(\d[\d,]*\.\d{2})\s+(\d[\d,]*\.\d{2}Dr)', line)
            if match:
                date, desc, amount, balance = match.groups()

                # Determine credit or debit
                if 'By Cash' in desc or 'IMPS' in desc or 'Cr-' in desc:
                    credit = float(amount.replace(',', ''))
                    debit = None
                else:
                    credit = None
                    debit = float(amount.replace(',', ''))
```

```
            # Clean balance
            balance_clean = float(balance.replace(',', '').replace('Dr', ''))

            # Append to lists
            dates.append(date)
            descriptions.append(desc.strip())
            credits.append(credit)
            debits.append(debit)
            balances.append(balance_clean)

# Create a DataFrame
df = pd.DataFrame({
    'Date': dates,
    'Description': descriptions,
    'Credit': credits,
    'Debit': debits,
    'Balance (Dr)': balances
})

# Show a summary
summary = {
    'Total Credit': df['Credit'].sum(skipna=True),
    'Total Debit': df['Debit'].sum(skipna=True),
    'Opening Balance': df.iloc[0]['Balance (Dr)'],
    'Closing Balance': df.iloc[-1]['Balance (Dr)']
}

print("Transaction Summary:")
for k, v in summary.items():
    print(f"{k}: ₹{v:,.2f}")
```

```
# Optional: Save to Excel
df.to_excel('kamal_bricks_statement.xlsx', index=False)
```