**ETL Report for Team 7 - LearnwithWiDa**

1. Title
2. Objective
3. Data Source (Extract Phase)
4. Data Extraction Process
5. Data Transformation Process
6. Data Loading Process
7. Challenges and Solutions
8. Summary and Next Steps
9. Appendix

* Column Descriptions
* Excel Functions Used

# 

# **1. TITLE**

ETL Report on Customer Transaction Data on Moniepoint from 2023-2024

# **2. OBJECTIVE**

The objective of this ETL process is to extract customer transaction data from the source given to us, clean and transform the data, and load it into a SQL database for transaction trend analysis.

# **3. Data Sources (Extract Phase)**

- Moniepoint-Document-2024-11-01T10-47-1.pdf

# **4. Data Extraction Process**

- Used a website called bankstatementconverter.com to convert the file from PDF to a CSV file for use in an Excel sheet.

# **5. Data Transformation Process**

- Two rows were merged into one in the extracted data.  
 - Started from the date column and manually separated it into two rows.  
 - Standardized date format to YYYY-MM-DD.  
 - Used space delimiter to split transaction reference into two columns.  
 - Used space delimiter for settlement credit, settlement debit, balance before, balance after, and charges.  
 - Used space delimiter to split beneficiary, beneficiary institution, and source from double rows to single rows.  
 - Extracted debit and credit from transaction reference.  
 - Dropped unnecessary columns like reversal status, RRN, Terminal ID, and narration.  
 - Used TRIM and CLEAN functions to standardize character lengths.

# 

# **6. Data Loading Process**

- Loaded the transformed data into Excel initially and then into SQL for analysis.

# **7. Challenges and Solutions**

- Issue: Multiple entries in one row.  
 Solution: For the date, due to absence of a delimiter, it was arranged manually; other columns with multiple entries were split using the space delimiter.

# **8. Summary and Next Steps**

Summary:  
 - The ETL process successfully extracted Moniepoint customer transaction data from a PDF file for the period 2023–2024.  
 - Data irregularities such as merged rows and inconsistent formatting were manually corrected and standardized.  
 - Key fields like debit, credit, date, balance, and transaction reference were separated and cleaned.  
 - Non-essential columns were dropped to streamline the dataset.  
 - The cleaned data was loaded into an SQL database for querying and analysis.  
 Next Steps:  
 - Build Looker dashboards to visualize monthly transaction volumes, debit and credit flows, and top beneficiaries.  
 - Create Looker Explores and Looks for dynamic reporting.

# 

# 

# 

# **9. Appendix**

| Column Name | Description |
| --- | --- |
| Date | Transaction date in YYYY-MM-DD format |
| Debit | Amount debited from the account |
| Credit | Amount credited to the account |
| Balance Before | Balance before transaction |
| Balance After | Balance after transaction |
| Transaction Reference | Unique transaction ID |
| Beneficiary | Name of the fund recipient |
| Institution | Financial institution of the beneficiary |
| Charges | Fees or charges associated with the transaction |

Excel Functions Used:

- =TRIM to remove extra spaces

- =CLEAN to clean non-printable characters

- =TEXT("yyyy-mm-dd") to format dates.