MOMO DATA ANALYSIS

Author: Victor Mofiyinfoluwa Akin-Oladiran
Institution: African Leadership University
Date: 06/15/2025

---

1. Introduction

This report details the design and implementation of an SMS parser that reads from an Android-exported XML file and outputs a user-friendly CSV file. The parser is built using Python and is tailored for simple transformation of structured SMS data.

---

2. Approach

- I used Python's `xml.etree.ElementTree` to parse `modified_sms_v2.xml`.
- Extracted necessary fields (`address`, `body`, `date`, and `type`) from each `<sms>` entry.
- Converted timestamps from milliseconds to a human-readable datetime format using `datetime.fromtimestamp()`.
- Wrote each entry as a row in a CSV file using Python's `csv` module.

This approach ensures the output is readable and easy to analyze using tools like Excel.

---

3. Challenges Faced

- Timestamp Conversion: The `date` values were in milliseconds since epoch. I had to divide them by 1000 to convert to standard UNIX time for accurate formatting.
- Encoding Issues: Some message bodies contained special characters. Ensuring the script correctly handles UTF-8 encoding was crucial.
- Tag Mismatch: Understanding the structure of the XML and avoiding parsing errors required careful analysis of the schema.

---

4. Key Decisions

- I focused on parsing only `<sms>` tags to avoid complexity with `<smes>` or `<call>` logs.
- I chose CSV format for the output as it is both human-readable and easy to import into spreadsheets or databases.
- The script was written in a modular way, separating parsing, formatting, and writing logic for maintainability.

---

5. Sample Output

address	date & time	type	body
+1234567890	2023-06-01 14:30	1	Hey! What's up?
+2345678901	2023-06-02 10:12	2	I'll call you later.

---

6. Conclusion

This project was a practical exercise in reading structured XML data, processing it, and transforming it into a flat file format. It reinforced my skills in Python programming, file I/O operations, and data formatting. If extended, the script could include support for MMS, spam detection, or even sentiment analysis using NLP tools.