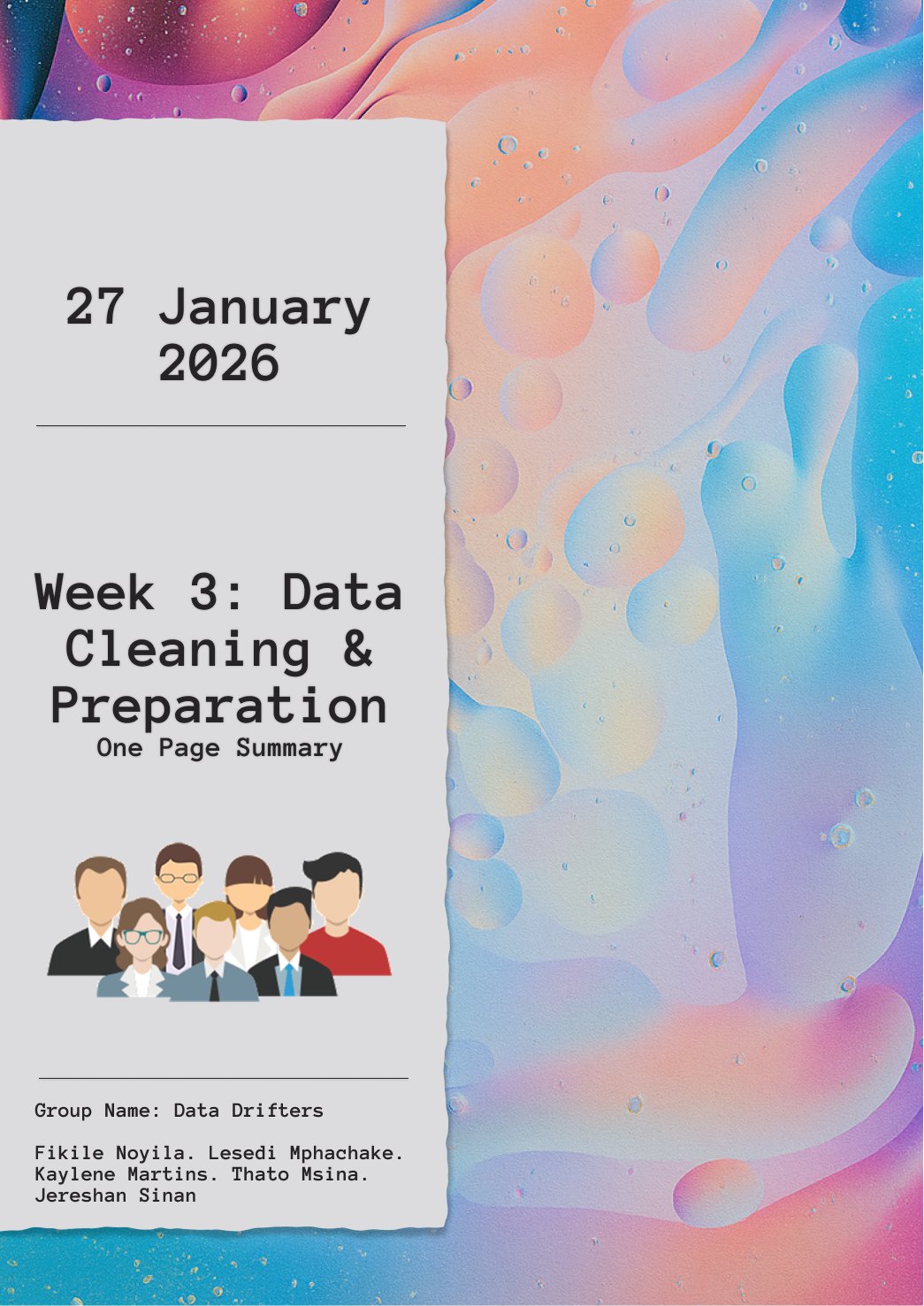
****

# Executive Summary: Employee Data Cleaning & Insights

**Project:** Week 3 Data Cleaning & Preparation

**Tools Used:** Microsoft Excel, Python (pandas, Jupyter), CSV

**Data Overview & Exploration**

The dataset consisted of structured employee records including identifiers, demographics, employment history, and financial data.

* **Type:** **Structured Data** (organized into rows and columns with defined data types).
* **Scope:** Records tracked the lifecycle of employees from hire to exit, including departmental placement and performance metrics.

**Data Cleaning & Transformation (Python)**

Using Python’s panda’s library, we performed rigorous data cleaning to ensure the integrity of HR metrics:

* **Standardization:** Normalized inconsistent department names (e.g.,"HR" vs. "Human Resources") and standardized currency symbols to ISO codes (USD, ZAR, GBP).
* **Data Types:** Converted hire\_date and exit\_date to datetime objects and ensured salary was treated as a numeric float for calculations.
* **Validation:** Trimmed whitespace and applied proper casing to names and emails.
  + Flagged/resolved logical errors, such as emails that were invalid or exit\_date occurring before hire\_date.
* **Missing Values:** Handled strategically; exit\_date remained null for active employees, while missing Performance Review values were filled with ‘Pending’.

**Key Performance Indicators (KPIs)**

After cleaning, the following high-level insights were extracted:

| **Metric** | **Result (Summary)** |
| --- | --- |
| **Current Headcount** | 103 |
| **Manager with Highest Team Size** | 7 |
| **Turnover Rate** | 31.33% (Total Exits / Total Workforce) |
| **Top Department** | Human Resources (Largest by Headcount) |

**Excel vs. Python Validation**

* **Excel Pivot Tables/Charts** were used to generate quick visual summaries of headcount and salary averages per currency.
* **Python** provided deeper statistical analysis, specifically regarding tenure and automated cleaning of thousands of rows that would be prone to human error in Excel.
* **Cross-Verification:** Headcount totals in Excel matched the Python results, confirming a successful data migration and cleaning process.

**Conclusion**

The data is now "Analysis-Ready." Standardizing the categories and fixing date inconsistencies has ensured that future HR reporting particularly regarding turnover and salary benchmarking is accurate and reliable.