LinkedIn Network Career Alignment Analysis - Methodology

Objective:

Analyze LinkedIn connections to understand how aligned the user's professional network is with their target field, especially Data Science, using Python for preprocessing and Snowflake for storage, querying, and dashboarding.

Tools Used:

- Python: pandas, tabulate, Jupyter Notebook
- Snowflake: SQL-based data warehouse and visualization (Snowsight)
- Data Source: LinkedIn Connections export (CSV)

Step-by-Step Methodology:

1. Data Collection

- Exported LinkedIn Connections.csv with fields: First Name, Last Name, Email Address, Company, Position, Connected On.

2. Data Cleaning (Python)

- Combined first and last names into full_name
- Standardized text: lowercase, trimmed whitespace
- Handled missing values in job titles and companies

3. Job Role Classification

- Created a rule-based dictionary for 14 job categories:

Data Science, Data Analyst, BI, Software Engineering, QA, Web Dev, DevOps/Cloud, Product/Project,

HR, Cybersecurity, Consulting, Marketing, Student, Other

- Special rules:
  - "business analyst" -> Business Intelligence
  - all other "analyst" -> Data Analyst
  - "student", "graduate", "intern" -> Student
  - "qa", "sdet", "test engineer" -> QA

4. Dataset Export

- Saved final cleaned dataset as cleaned_connections_with_categories.csv

5. Snowflake Setup

- Created database: linkedin_network
- Created schemas: staging (raw upload), analytics (views)
- Uploaded the CSV to staging.connections

6. SQL Analytics (Snowsight)

- Created permanent views:
  - connections_by_category
  - top_companies
  - ds_alignment_summary
  - connection_trends
- Queries included alignment %, top companies, role breakdown, time trends

7. Visualization (Snowsight Dashboard)

- Tiles used:
  - Bar: Job Category Breakdown
  - Scorecard: % aligned to Data Science

- Bar: Trends in Connections Over Time

- Bar: Top Companies

8. Documentation and Sharing

- Generated README.md for GitHub

- Prepared screenshots of the dashboard

- Created .sql script with view logic

- Summarized methodology in this PDF

Conclusion:

This project provides a clear picture of how aligned a professional network is with a user's career direction. It uses structured classification, SQL analytics, and native Snowflake dashboards for interactive insight-all reproducible and fully cloud-native.