DataQuest Solutions
Data Analysis in STATA

Lesson 1. Introduction to STATA
- STATA Interface: Command window, Do-file editor, Data Editor, Results window
- Using menus vs. command syntax
- Installing and managing user-written commands (ssc install, net install)

Lesson 2. Data Management and Preparation
- Importing data (Excel, CSV, text files, database connections)
- Data entry and editing
- Data cleaning:
- Handling missing values
- Identifying and managing outliers
- Data transformation (generate, replace, recode)
- Renaming and labelling variables
- Encoding categorical variables (label define, label values)

Lesson 3. Exploratory Data Analysis (EDA)
- Summary statistics (summarize, tabulate, table)
- Visualizing distributions:
  - Histograms
  - Boxplots
  - Scatter plots
  - Bar charts
- Exploring relationships:
- Correlation matrices
- Cross-tabulations

Lesson 4. Data Manipulation Techniques
- Sorting and filtering (sort, keep, drop, list if)
- Creating new variables (generate, egen)
- Reshaping data (wide to long, long to wide)
- Merging datasets (merge, append, joinby)
- Collapsing data (collapse for summaries)

Lesson 5. Descriptive and Inferential Statistics
- Frequencies and proportions

- Measures of central tendency and dispersion (mean, median, SD, variance)
- Confidence intervals (ci, cii)
- Hypothesis testing:
- t-tests (one-sample, paired, independent)
- ANOVA (one-way, two-way)
- Chi-square tests for independence
- Non-parametric tests (Wilcoxon, Kruskal-Wallis)

Lesson 6. Regression and Modelling
- Linear regression (regress)
- Logistic regression (binary: logit, logistic, probit)
- Multinomial logistic regression (mlogit)
- Ordered logistic regression (ologit)
- Checking model assumptions (heteroskedasticity, multicollinearity, normality of residuals)
- Robust standard errors (robust option)

Lesson 7. Panel Data Analysis
- Setting panel data (xtset)
- Fixed-effects model (xtreg, fe)
- Random-effects model (xtreg, re)
- Hausman test for model choice

Lesson 8. Survey Data Analysis
- Setting survey design (svyset)
- Analyzing survey data (means, regression, proportions using svy: prefix)
- Complex sample designs (weights, stratification, clustering)

Lesson 9. Programming and Automation in STATA
- Writing and running do-files (for reproducible analysis)
- Writing loops and macros
- Creating programs (program define) for repeated analysis
- Automating repetitive tasks

Lesson 10. Best Practices and Reporting
- Creating reproducible reports
- Outputting tables and results to Word, Excel (outreg2, esttab, putexcel)
- Exporting datasets
- Properly documenting analysis steps (annotations in do-files)