Analytical Skills for Business (WS 2025/26)

Business Administration (M. A.)

© Benjamin Gross

September 8, 2025

This document holds the course material for the Analytical Skills for Business course in the Master of Business Administration program. It discusses version control systems such as Git and GitHub for efficient team collaboration, offers an overview of no-code and low-code tools for data analytics including Tableau, Power BI, QlikView, makeML, PyCaret, RapidMiner, and KNIME, and introduces key programming languages such as R, Python, and SQL alongside essential programming concepts like syntax, libraries, variables, functions, objects, conditions, and loops. In addition, it covers working with modern development environments, including Unix-like systems, containers, APIs, Jupyter, and RStudio, and sets expectations for project submissions and evaluation.

Table of contents

1 Introduction

Computer science

1.1 Implementing version control systems

Version control systems are essential tools for managing code, tracking changes, and facilitating collaborative development in modern business analytics projects (Çetinkaya-Rundel & Hardin, 2021). These systems enable teams to work efficiently on shared codebases while maintaining a complete history of all modifications, ensuring reproducibility and accountability in data analysis workflows.

1.1.1 Core Concepts

Version control systems provide systematic approaches to managing changes in documents, programs, and other collections of information:

- **Repository**: A central storage location containing all project files and their complete revision history
- **Commit**: A snapshot of the project at a specific point in time, representing a set of changes
- **Branch**: An independent line of development allowing parallel work on different features
- **Merge**: The process of integrating changes from different branches back together

Figure 1: git

4
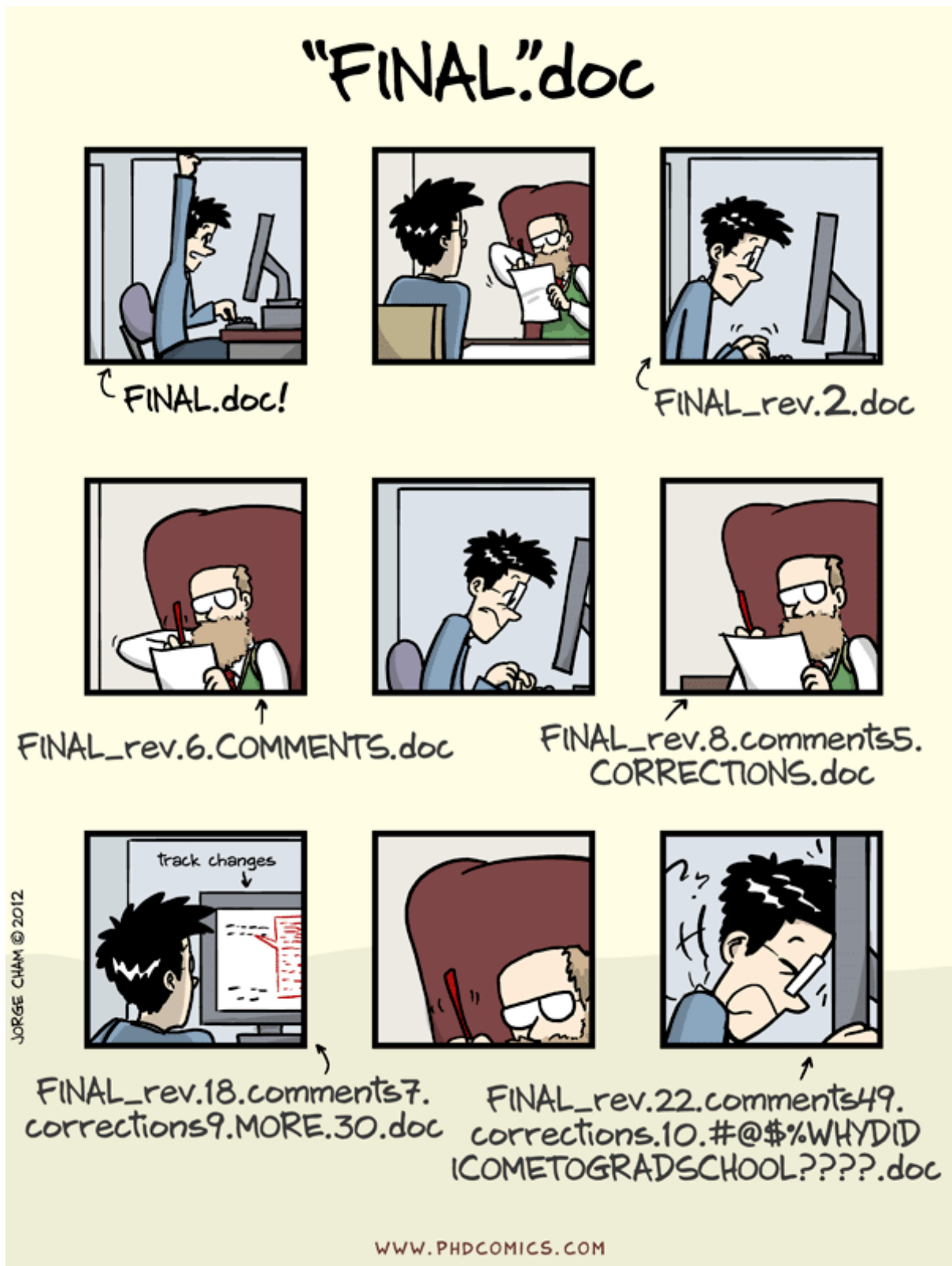
1.1.2 Git: Distributed Version Control

Git is a distributed version control system that tracks changes in files and coordinates work among multiple contributors. Key characteristics include:

Local Repository: Each user maintains a complete copy of the project history, enabling offline work and faster operations.

Staging Area: An intermediate area where changes are prepared before being committed to the repository.

Branching and Merging: Lightweight branching allows for experimental development without affecting the main codebase.

Distributed Workflow: No single point of failure, as every user has a complete backup of the project.

1.1.3 GitHub: Cloud-Based Collaboration Platform

GitHub is a web-based hosting service for Git repositories that adds collaboration features and project management tools:

- **Remote Repositories**: Centralized storage accessible from anywhere with internet connectivity
- **Pull Requests**: Structured code review process for integrating changes
- **Issue Tracking**: Built-in project management for tracking bugs and feature requests
- **Actions and CI/CD**: Automated workflows for testing and deployment
- **Documentation**: Integrated wiki and README support for project documentation

The combination of Git and GitHub creates a powerful ecosystem for collaborative analytics projects, ensuring code quality, facilitating peer review, and maintaining comprehensive project documentation (GeeksforGeeks, 2024a).

1.1.4 Comparison of Git and GitHub

On this image you can see the integration of Git in GitHub:

Figure 2: git-github.png
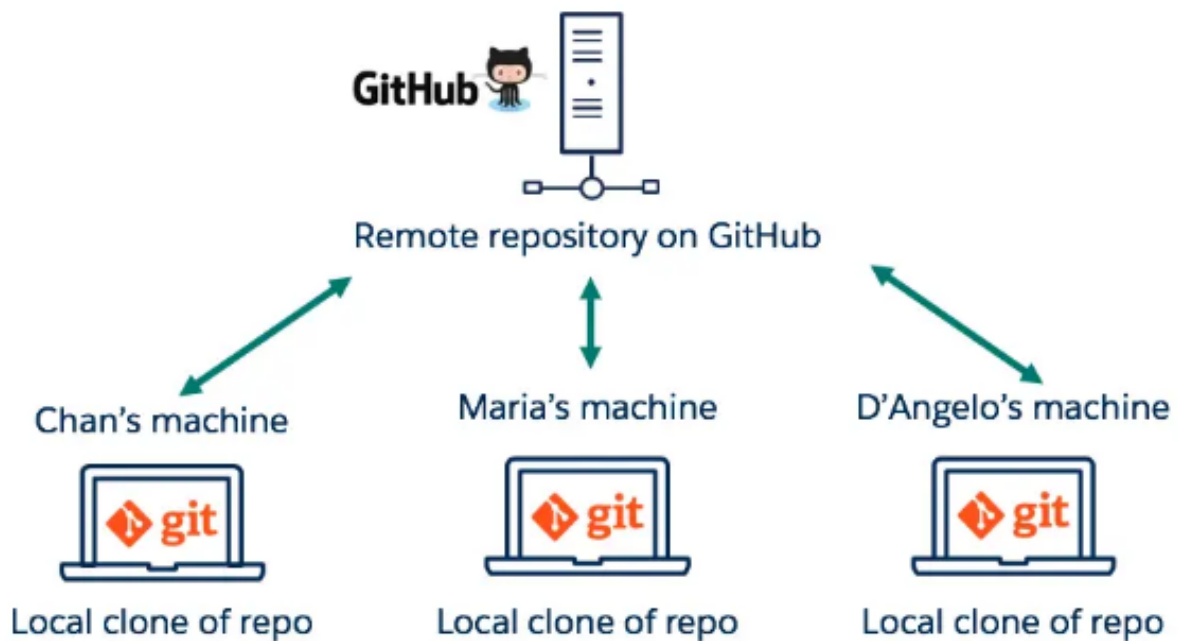
1.1.5 Business Analytics Applications

In business analytics contexts, version control systems provide:

- **Reproducible Analysis**: Complete tracking of analytical scripts and data processing steps
- **Collaborative Research**: Multiple analysts can work simultaneously on different aspects of projects
- **Model Versioning**: Systematic management of machine learning models and their evolution
- **Data Governance**: Audit trails for compliance and regulatory requirements
- **Backup and Recovery**: Protection against data loss and accidental modifications

1.1.6 Reference Materials

For comprehensive coverage of version control concepts, implementation strategies, and best practices, consult:

- **Theoretical Foundation**: Introduction to Modern Statistics provides context on reproducible research practices and collaborative analytics (Çetinkaya-Rundel & Hardin, 2021)
- **Git Resources**: Git Cheat Sheet offers quick reference for common Git commands
- **GitHub Documentation**: GitHub Manual contains detailed guidance on platform features
- **Visual Overview**: Git and GitHub Workflow illustrates the relationship between local and remote repositories
- **Online Resources**:
  - GeeksforGeeks Git vs GitHub Guide provides practical comparisons and use cases (GeeksforGeeks, 2024a)

– [Official GitHub Documentation](#) offers authoritative guidance on getting started (GitHub, 2024)

Understanding version control systems is fundamental for modern business analytics, enabling collaborative development, ensuring reproducibility, and maintaining professional standards in data science projects.

1.2 Overview on no-code and low-code tools for data analytics

- Tableau
- Power BI
- QlikView
- makeML
- PyCaret
- Rapidminer
- KNIME

1.3 Overview on Programming languages

- R
- Phyton
- SQL

1.4 Elements of programming languages

- Syntax
- libraries
- variables
- functions
- objects
- conditions
- loops

1.5 Development environments

- Unix-like systems
- containers
- APIs
- Jupyter
- RStudio

2 Descriptive statistics

xy

2.1 Measures of centrality, dispersion, and concentration

xy

2.2 Descriptive analytics

- univariate data
- bivariate data
- multivariate data

2.3 Techniques

- constructing
- interpreting
- evaluating of scores, rankings, metrics, and composite indicators.

2.4 Visualizing and exploration

- categorical
- numerical
- time series data

2.5 Handling messy data

xy

2.6 Association

- measuring the association of variables, including correlation and regression

2.7 Implementing applications

in the programming language R for practical data analysis.

3 Inferential statistics

xy

3.1 Basic concepts of statistical inference

xy

3.2 Quantification of probability through random variables

xy

3.3 Hypothesis testing

Hypothesis testing is a fundamental statistical method used to make inferences about population parameters based on sample data (Illowsky & Dean, 2018). It provides a systematic framework for evaluating claims about populations using sample evidence, enabling data-driven decision making in business contexts.

3.3.1 Core Concepts

A statistical hypothesis test involves formulating two competing hypotheses:

- **Null hypothesis (H)**: The status quo or default position, typically representing no effect or no difference
- **Alternative hypothesis (H or H)**: The research hypothesis representing the effect or difference we seek to detect

The process involves calculating a test statistic from sample data and comparing it to a critical value or determining a p-value to make decisions about rejecting or failing to reject the null hypothesis (GeeksforGeeks, 2024b).

3.3.2 Key Components

Test Statistics: Standardized measures that quantify how far sample data deviates from what would be expected under the null hypothesis.

Significance Level ( ): The probability threshold for rejecting the null hypothesis, commonly set at 0.05 (5%).

P-value: The probability of observing test results at least as extreme as those obtained, assuming the null hypothesis is true.

Critical Region: The range of values for which the null hypothesis is rejected.

3.3.3 Types of Errors

- **Type I Error ( )**: Rejecting a true null hypothesis (false positive)
- **Type II Error ( )**: Failing to reject a false null hypothesis (false negative)
- **Statistical Power (1-)**: The probability of correctly rejecting a false null hypothesis

3.3.4 Reference Materials

For comprehensive coverage of hypothesis testing concepts, methodologies, and applications, consult:

- **Theoretical Foundation**: Introductory Statistics provides detailed explanations of hypothesis testing principles and procedures (Illowsky & Dean, 2018)
- **Visual Guide**: Hypothesis Testing Overview offers a visual representation of key concepts
- **Detailed Methodology**: Hypothesis Testing Documentation contains comprehensive methodological information
- **Online Resource**: Additional perspectives on hypothesis testing applications can be found in the GeeksforGeeks guide (GeeksforGeeks, 2024b)

Understanding hypothesis testing is essential for making informed business decisions based on data analysis, forming the foundation for advanced statistical inference and predictive analytics in business contexts.

3.4 Confidence intervals, p-values, and statistical tests

xy

3.5 Inferential statistics

in the programming language R, translating theoretical knowledge into practical applications.

4 Predictive analytics

xy

4.1 Data mining techniques

xy

4.2 Regression analysis

xy

4.3 Forecasting in predicting future business outcomes

xy

5 Literature

All references for this course.

5.1 Essential Readings

Bruce, P. and A. Bruce (2020). *Practical Statistics for Data Scientists, 2nd Edition.* https://learning.oreilly.com/library/view/practical-statistics-for/9781492072935/preface01.html.

Çetinkaya-Rundel, M. and J. Hardin (2021). *Introduction to Modern Statistics.* https://www.openintro.org/boo https://github.com/DrBenjamin/Analytical-Skills-for-Business/blob/491a9a84dd0227aab44e0a6db7e6330830a literature/Introduction_to_Modern_Statistics_2e.pdf/?raw=true.

Stephenson, P. (2023). *Data Science Practice.* https://datasciencepractice.study/.

5.2 Further Readings

Békés, G. and G. Kézdi (2021). *Resources for Data Analysis for Business, Economics, and Policy.* Instructor resources: https://www-cambridge-org.eux.idm.oclc.org/highereducation/books/data-analysis-for-business-economics-and-policy/D67A1B0B56176D6D6A92E27F3F82AA20/. https://github.com/DrBenjamin/Analytical-Skills-for-Business/blob/c2ec1b2061c7dc36200977cfd58daf6020c1c774/ literature/B%C3%A9k%C3%A9s_Data%20Analysis%20for%20Business%2C%20Economics% 2C%20and%20Policy_2021_First%20Day%20of%20Class%20Slides.pdf/?raw=true.}

Dougherty, J. and I. Ilyankou (2025). *Hands-On Data Visualization.* https://handsondataviz.org/.

Evans, J. R. (2020). "Business Analytics''.

GeeksforGeeks (2024a). *Difference between Git and GitHub.* Comprehensive comparison of Git and GitHub for version control. https://www.geeksforgeeks.org/git/difference-between-git-and-github/.

——— (2024b). *Understanding Hypothesis Testing.* Online resource for software testing and statistical concepts. https://www.geeksforgeeks.org/software-testing/understanding-hypothesis-testing/.

GitHub (2024). *About GitHub and Git.* Official GitHub documentation on Git and GitHub fundamentals. https://docs.github.com/en/get-started/start-your-journey/about-github-and-git.

Illowsky, B. and S. L. Dean (2018). *Introductory Statistics.* OpenStax, Rice University, p. 905. ISBN: 1938168208. https://github.com/DrBenjamin/Analytical-Skills-for-Business/blob/c2ec1b2061c7dc36200977cfd58daf6020c1c774/literature/Introductory%20Statistics.pdf/?raw=true.}

Irizarry, R. A. (2024). "Advanced Data Science: Statistics and Prediction Algorithms Through Case Studies".

http://rafalab.dfci.harvard.edu/dsbook-part-2.

Kumar, U. D. (2017). "Business Analytics: The Science of Data-Driven Decision Making".

Pochiraju, B. and S. Seshadri, ed. (2019). *Essentials of Business Analytics.* Vol. 264. https://link.springer.com/10.1007/978-3-319-68837-4. Cham: Springer International Publishing. ISBN: 978-3-319-68836-7. DOI: 10.1007/978-3-319-68837-4. https://github.com/DrBenjamin/Analytical-Skills-for-Business/blob/c2ec1b2061c7dc36200977cfd58daf6020c1c774/literature/Essentials%20of%20Business%20Analytics.pdf/?raw=true.}

Vaughan, D. (2020). "Analytical Skills for AI and Data Science".

https://learning.oreilly.com/library/view/analytical-skills-for/9781492060932/preface01.html#idm46388898852872.