Syllabus

Data Analysis 1: Exploration – Business Analytics track

- **Instructor:** Agoston Reguly

  Email: reguly_agoston@phd.ceu.edu
  Office hours: Monday 10:30-12:00 or by appointment

- **Credits:** 1 (2 ECTS)
- **Term:** Fall 2020-2021
- **Course level:** [MA/MS]
- **Prerequisites:** Mathematics and Informatics Pre-session for Business Analytics
- **Course drop:** Course can be dropped free of charge 24 hours after the first session. After this date drop is possible until the course is halfway over (late drop fee applies). No changes are allowed past that date.

1. COURSE DESCRIPTION

About 80% of data science tasks are composed of managing data, from understanding and altering features of the dataset and variables, to combining various datasets. This course introduces the critical tasks of data collection and data wrangling, presentation and understanding of descriptive statistics and basics of visualization.

The course focuses on classic statistics methods and their applications, such as data collection and sampling, generalization from the sample to the population and hypothesis testing.

2. LEARNING OUTCOMES

Key outcomes

By successfully completing the course the students will be able to:

- Understand key issues of data gathering and manipulation
- Successfully formulate research questions that are answerable by empirical analysis;
- Produce meaningful descriptive statistics and informative graphs;
- Test some basic hypothesis

Other outcomes. The course will also help develop skills in the following areas:

Learning Area	Learning Outcome
Critical Thinking	Evaluate data quality, understand simple statistical reports, be able to reveal possible drawbacks of data sets.
Quantitative Reasoning	Understand basic statistical concepts and be able to argue about similarities or differences between different populations.
Technology Skills	Get familiar with data gathering, manipulation and visualization in R
Interpersonal Communication Skills	How to communicate the results of descriptive statistics, hypothesis tests and the messages of the created graphs.
Management Knowledge and Skills	Understand the needs of managers and be able to create such reports that is easily readable.

Cultural Sensitivity and Diversity	While working together on specific tasks students are encouraged to learn and respect cultural sensitivity and diversity.
Ethics and Social Responsibility	Ethics of data gathering and data usage.

3. READING LIST

Required:

- Békés – Kézdi [BK]: Data Analysis for Business, Economics and Policy (forthcoming) - available as handouts; Chapters 1-6

Optional Reading

- Angrist – Pischke (2009): Mostly Harmless Econometrics, Chapter 1
- Wooldridge (2018): Introductory Econometrics, Chapter 1 + Math Refreshment B-C

Databases. The CEU Library boasts a range of databases covering financial and company data, market and industry reports, global news and more. For a full list of databases visit the CEU Library.

- Refinitiv (Thomson Reuters) Eikon for Students + Datastream/Thomson ONE
  - Eikon: Platform used by finance practitioners including market traders to monitor and analyze financial information. Information, analytics and news on all major financial markets including real-time pricing data, financial research, global financial news and commentary, financial estimates, fundamentals analysis, visual analysis through charting. Import/export from Excel.
  - Datastream: Range of economic, securities and company financial data. Excel add-in.
  - Thomson ONE: Global overviews on 55,000 public companies, one million private companies. Reuters News, ownership, deals, private equity, key ratios, company filings, officers and directors. Investext analyst reports, active and historical research from 1,600 independent research firms, brokerages, investment banks.
- Standard & Poor's Capital IQ
  - Web and Excel-based platform combining deep global company information, credit ratings and research, and market research with powerful tools for risk assessments. Real-time and historical information on markets, industries, companies, transactions and people. Tearsheet data.
- Lexis Nexis Academic
  - Global database of news, business, legal and other sources. Full text of 350 newspapers, 300 magazines and journals, 600 newsletters. Wire services including Associated Press, Business Wire and PR Newswire. Company financial information, market research, industry reports.

4. TEACHING METHODS AND LEARNING ACTIVITIES

The course will involve a mix of

- Lectures to present basic ideas in statistical exploration.
- Seminar-type classroom to understand how to formulate proper research questions and how to answer those. Through this we demonstrate the use of the learned theories and concepts.
- Optional homework to practice through extra problems and deepen the student's knowledge.

5. ASSESSMENT (including minimum pass requirement and grading

- Start-of-the-class Quizzes (10%)
- Team assignments (20%)
- Closed book exam (70%)

Grading policy

- Students may not miss more than 2 sessions. Failing to do so will yield an automatic Fail grade.
- To pass, students will need to get at least 50% of the overall grade AND at least 50% of the exam. Failure to do so, will yield a Fail grade

6. TECHNICAL/LAPTOP REQUIREMENT

Having laptops is advised.

7. TOPIC OUTLINE AND SCHEDULE

Session	Topics	Readings
September 28	Origins of data: data table, data quality, survey, scraping, sampling, ethics.	BK Ch01
September 30	Preparing data for analysis: tidy data, source of variation, variable types, missing data, data cleaning.	BK Ch02
October 5	Explanatory data analysis: probability, distributions, extreme values, summary stats, data visualization, missing values.	BK Ch03
October 7	Comparison and correlation: conditional probability, conditional distribution, conditional expectation, visual comparisons, correlation.	BK Ch04
October 12	Generalizing from a dataset: repeated samples, confidence interval, standard error estimation via bootstrap and formula, external validity.	BK Ch05
October 14	Testing hypotheses: null and alternative hypotheses, t-test, false positives / false negatives, p-value, testing multiple hypotheses.	BK Ch06

8. SHORT BIO OF THE INSTRUCTOR (1 para)

Ágoston Reguly is a fifth-year Economics PhD student at the Department of Economics and Business. He has worked more than three years at the Hungarian Government Debt Management Agency as an expert analyst. His research topics are in econometrics, policy evaluation with machine learning techniques and optimal survey designs.