

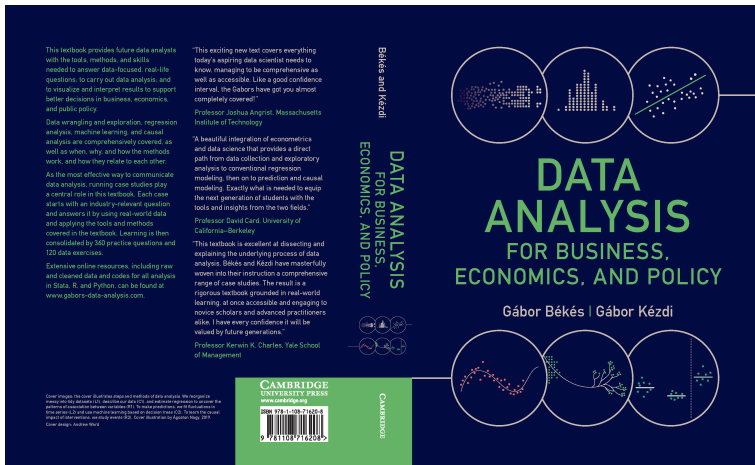
Az adatelemzés oktatása

Békés Gábor (Central European University, KRTK KTI)

MKE – Kézdi Gábor emlékkonferencia

2021 December 20

Békés - Kézdi: Adatok elemzése



Békés - Kézdi: Adatok elemzése

- ▶ Cambridge University Press, 2021
- ▶ cambridge.org/bekeskezdi
- ▶ gabors-data-analysis.com
- ▶ github.com/gabors-data-analysis/da_case_studies
- ▶ Follow us
 - ▶ twitter.com/Gabors_Data
 - ▶ facebook.com/gaborsdata

Tartalmi forradalom

- ▶ Tanítsuk azt, ami gyakran kell
- ▶ Ne tanítsuk, ami nem olyan fontos
- ▶ Életre és ne akadémiaiára készítsünk
- ▶ De akadémiai pályán is hasznos legyen

“This exciting new text covers everything today’s aspiring data scientist needs to know, managing to be comprehensive as well as accessible. Like a good confidence interval, the Gabors have got you almost completely covered!”

Professor Joshua Angrist, Massachusetts Institute of Technology

gabors-data-analysis.com/endorsements

A tudás demokratizálása

- ▶ Legyen didaktikus, magyarázó, de érdekes
- ▶ Alapos és precíz, de befogadható
- ▶ Sok példán keresztül
- ▶ Csak a szükséges matek a főszövegben



gabors-data-analysis.com/whythisbook/

Azt adjuk, ami kell az adat elemzőnek

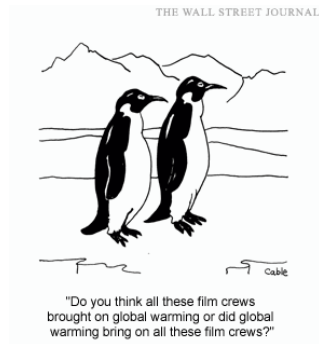
- ▶ Amit használni tud: többféle forrásból
- ▶ Standard ökonometriai készletből nem kell minden
- ▶ Gépi tanulás, előrejelzés kell
- ▶ Data science sok hasznosat ad

"A beautiful integration of econometrics and data science that provides a direct path from data collection and exploratory analysis to conventional regression modeling, then on to prediction and causal modeling. Exactly what is needed to equip the next generation of students with the tools and insights from the two fields."

Professor David Card, University of California–Berkeley

Technikai tudás, eredmények értelmezése és gondolkodás

- ▶ Technikai tudás, mit miért, és hogyan csinálunk
- ▶ Az eredmények értelmezése
- ▶ Statisztika és internal validity
- ▶ Gondolkodás és external validity



All in one - egy könyv, egy teljes program

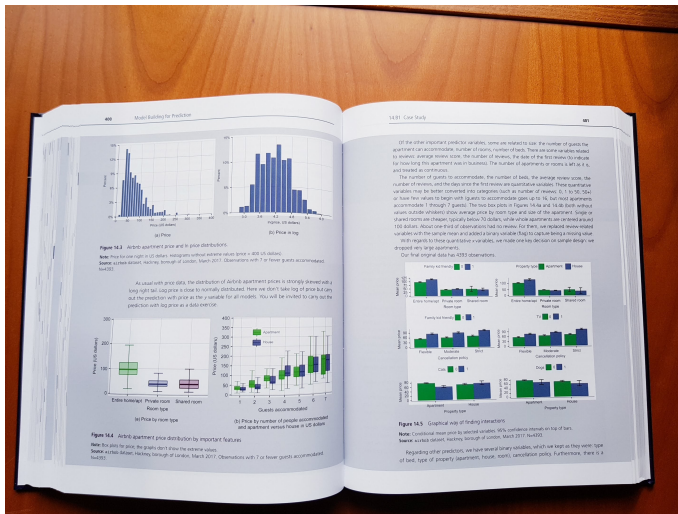
- ▶ Adatok gyűjtése és leírása
- ▶ Regressziók futtatása és értelmezése
- ▶ Előrejelzés és gépi tanulás
- ▶ Oksági elemzés

gabors-data-analysis.com/chapters/

47 case study, sokféle téma

- ▶ A diákok nagy része nem tudja még, mi érdekli
- ▶ Business: managerek szerepe, árazás, részvényárak, mennyiség előrejelzés
- ▶ Economics: vállalati összeolvadás, otthoni munka hatása,
- ▶ Policy: étkezés, dohányzás szerepe, immunizáció
- ▶ Globális példák: 1/3 USA, 1/3 EU+UK, 1/3 egyéb

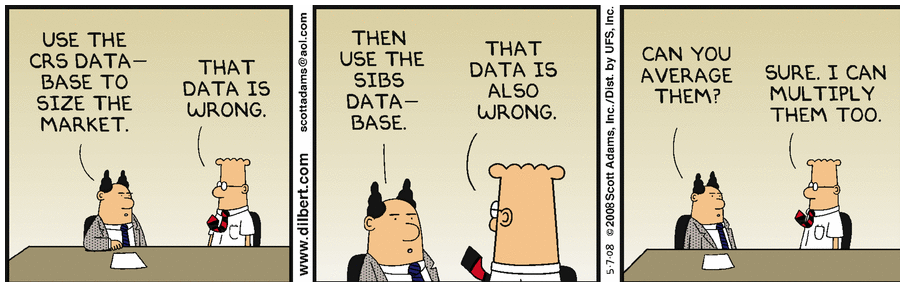
47 case study, sokféle téma



Az adatelemzés egy folyamatának ismétlése

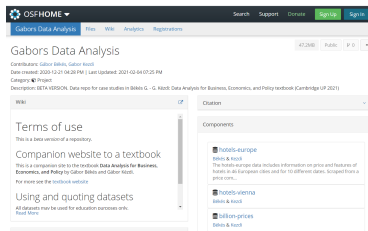
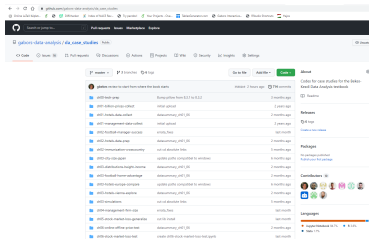
1. Először is kell egy kutatási téma és jól meghatározott **kutatási kérdés**
2. Az **adatok gyűjtése** minden empirikus munka alapja
3. Az **adatok tisztítása és rendszerezése** szükséges és fáradtságos munka
4. A **leíró adatelemzés** két célt szolgál: segíti a tisztítást és előkészíti az elemzést
5. Az **elemzési munka** alapja a modellek becslése és a hipotézisek tesztelése
6. Az eredmények **bemutatását** igazítsuk a hallgatósághoz
7. Végül **válaszoljuk meg** az eredeti kérdést, és tárgyaljuk meg, hogy mennyiben általánosítható.

Valós adatokkal dolgozni nehéz, de fontos tudás



Ecosystem - "Data and code or did not happen"

- ▶ Code: github.com/gabors-data-analysis/da_case_studies,
- ▶ Adatok: osf.io/7epdj/



Egy fontos case study: vállalati exit (ch17) - a példa

CASE STUDY – Predicting Firm Exit: Probability and Classification

Question, data, and feature engineering

In this case study, our aim is to predict corporate exit: a company going out of business, or, as is sometimes said, exiting.

Some business questions need the predicted probability of exit. When a company is thinking about which supplier to select for a long-term relationship, the probability that each candidate will be around in the future is an important factor to include in its decision. Other business questions also need classification. When a bank wants to decide whether to approve a loan application from a firm, it wants to decide whether that firm is going to be in business by the time it needs to repay its loan. Similarly, when the owner of an office building wants to decide whether to lease space to a company for a fixed term, it would like to know whether the company will still be in business at the end of that term.

In this case study, called **Predicting firm exit: probability and classification**, we use a lot of information on a set of firms in a European country to predict the probability of firm exit and to classify firms into prospective exiting firms and prospective staying in business firms, from the **bisnode-firms** dataset. The data was collected, maintained, and cleaned by Bisnode, a major European business information company.

Egy fontos case study: vállalati exit (ch17) - adatok

Our original dataset covers the entire population of companies between 2005 and 2016 for a few industries in manufacturing (Electronic and optical products, Electrical equipment, Machinery and equipment, Motor vehicles, Other transport equipment, Repair and installation of machinery) and services (Accommodation and Food and beverage service activities). In the data, there are some firms who changed their main classification over time. Large firms are excluded for data protection (above 100 million euros in annual revenues).

The dataset is made up of several data tables, which makes it a little complicated to use. Data tables included are presented in Table 17.1.

Table 17.1 Data tables for firm-level data

Data table	Unit of observations	Important variables
Financial data	firm-year	revenue, material cost, EBITDA, assets,
Managers	person-time Interval-firm	CEO age, gender
Employment	firm-year	average wages
Corporate registry	firm	city of headquarters, year of firm birth
Industry classification	firm	NACE industrial classification

Source: binodo-firms dataset.

Hol lehet tanítani

- ▶ Mesterszakok I.: Economics, Applied Economics, Finance
- ▶ Mesterszakok II.: Business Analytics, Data Science

- ▶ Alapszak I.: Economics
- ▶ Alapszak II.: Quantitative SS, PPE, Business

- ▶ Üzleti Phd
- ▶ MBA/EMBA

gabors-data-analysis.com/teaching-guide/undergraduate-and-graduate-programs

Sok más könyv van, miért a miénk?

- ▶ Diákközpontú: releváns témák, sok példa és gyakorlás
- ▶ Kimenet econometrics és data science irányba
- ▶ Ecosystem
 - ▶ kódok és adatok
 - ▶ gyakorló kérdések, feladatok, MCQ,
 - ▶ ábrák, diasorok elérhetőek (Overleaf)

Sok más könyv van, miért a miénk?

"It is excellent. I plan on using it for my Business Strategy and Analytics module. I find it to be quite exhaustive and very well organized, very consistent with how I organize my module. I think they covered nearly everything that I'd want covered in a book like this." Anil Doshi (UCL)

"I like the book so much that I modified my course to better adhere to the contents of the book.[..] The case studies with code are the biggest bonus, but I also really like the scope of the contents and the intuitive explanations. Excellent book and ecosystem" Peter Fuleky (Hawaii)

"The thing that attracts me about the book is the emphasis on applied work and data management. The value of these skills is recognized. The challenge is that we and our students' employers want those skills in addition to strong applied econometric skills." John Straub (Maryland)

gabors-data-analysis.com/instructor-feedback

Hogyan áll a bevezetés?

- ▶ Econ: UC Merced, Hawaii, Bocconi, Humboldt, Middlesex, Brescia, CEU, Corvinus
- ▶ Policy: Columbia Health, Heidelberg, Rosario (Colombia)
- ▶ Business: Cambridge Judge, ESSEC, Westminster, Essex
- ▶ Computer Sci: Eindhoven, Insper (Brazil)

gabors-data-analysis.com/courses-using/

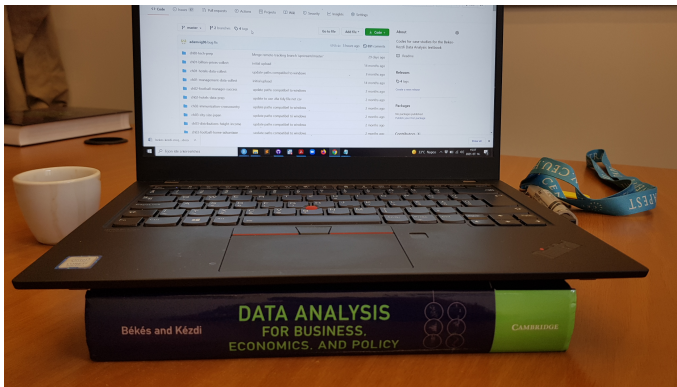
Mit tehetek, ha buzog a tettvágy?

- ▶ Social media: LinkedIn, Twitter (Facebook)
- ▶ Írj 3 ismerősödnek, aki adatot tárgyat tanít
- ▶ Rate + review: Amazon, Goodreads, Facebook
- ▶ Segítség a magyar fordításhoz

Nyitott kérdések

- ▶ Milyen témákat tanítsunk BA és MA diákoknak?
- ▶ Hogyan lehet a case study alapú tanítást még jobban támogatni (labor óra, cloud, online support)?
- ▶ Milyen eszközzel tanítsuk: Excel, BI, Stata, R, Python ?
- ▶ Miért (igen vagy ne) tanítsunk magyarul?

Tartsuk a kapcsolatot



facebook.com/gaborsdata
twitter.com/Gabors_Data