

Data Science or Data Engineering?

Miklós Koren (@korenmiklos)

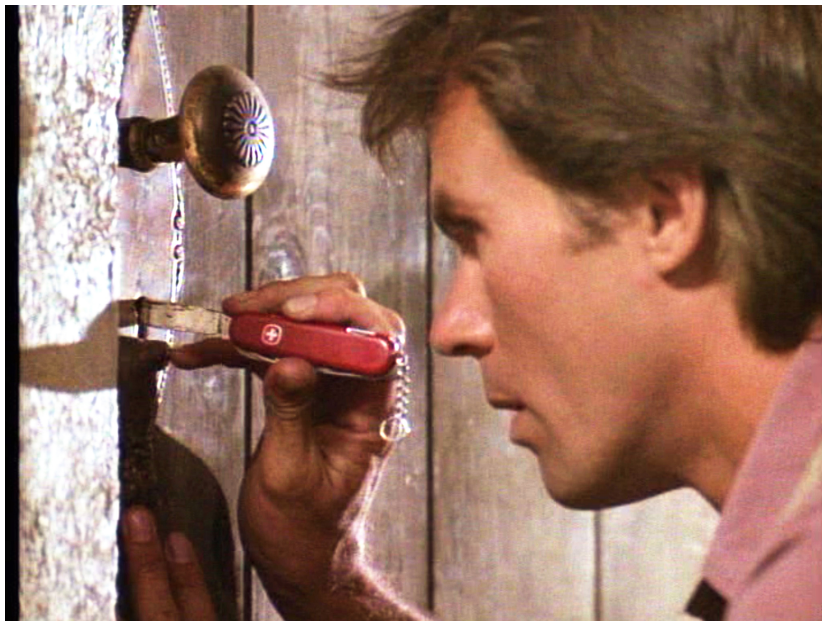
<https://economics.ceu.edu>

Introduction

Do you want to like be like her?



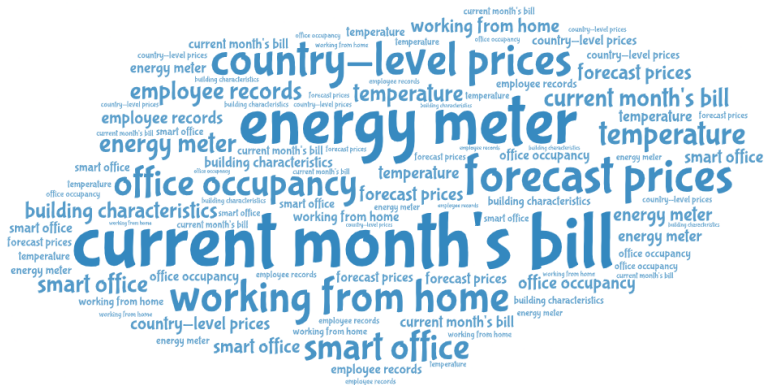
... or like him?



What is the energy bill of your company?

You work for a global company with offices in many locations world wide. You are tasked with forecasting the total energy bill of all offices for next month.

Some factors influencing your forecast



Key concerns in this project

Data Scientist

What model predicts energy bill best?

Key concerns in this project

Data Scientist

What model predicts energy bill best?

Data Engineer

How to gather and deliver timely and harmonized data?

Key concerns in this project

Data Scientist

What model predicts energy bill best?

Data Engineer

How to gather and deliver timely and harmonized data?

Commonalities

Work with **data** to solve business **problems**.

What do data scientists do? (O*NET 2022)

Data Scientists

15-2051.00

 Bright Outlook

 Updated 2022

Develop and implement a set of techniques or analytics applications to transform raw data into meaningful information using data-oriented programming languages and visualization software. Apply data mining, data modeling, natural language processing, and machine learning to extract and analyze information from large structured and unstructured datasets. Visualize, interpret, and report data findings. May create dynamic data reports.

Also see: [Business Intelligence Analysts](#), [Clinical Data Managers](#)

Summary

Details

Custom

 Easy Read

 Veterans

 Español

 Contents

Occupation-Specific Information

Tasks

▼ 5 of 16 displayed

- ➊ Analyze, manipulate, or process large sets of data using statistical software.
- ➋ Apply feature selection algorithms to models predicting outcomes of interest, such as sales, attrition, and healthcare use.
- ➌ Apply sampling techniques to determine groups to be surveyed or use complete enumeration methods.
- ➍ Clean and manipulate raw data using statistical software.
- ➎ Compare models using statistical performance metrics, such as loss functions or proportion of explained variance.

What do data engineers do? (O*NET 2022)

Database Architects

15-1243.00

 Bright Outlook

 Updated 2022

Design strategies for enterprise databases, data warehouse systems, and multidimensional networks. Set standards for database operations, programming, query processes, and security. Model, design, and construct large relational databases or data warehouses. Create and optimize data models for warehouse infrastructure and workflow. Integrate new systems with existing warehouse structure and refine system performance and functionality.

Sample of reported job titles: Data Officer, Database Analyst, Database Architect, Database Consultant, Database Developer, Database Programmer, Information Architect, Information Modeling Engineer Specialist, Information Technology Architect (IT Architect), System Engineer

Also see: [Data Warehousing Specialists](#)

Summary

Details

Custom

 Easy Read

 Veterans

 Español

 Contents ▾

Occupation-Specific Information

Tasks

▾ 5 of 25 displayed

- ➊ Design databases to support business applications, ensuring system scalability, security, performance, and reliability.
- ➋ Develop database architectural strategies at the modeling, design and implementation stages to address business or industry requirements.
- ➌ Collaborate with system architects, software architects, design analysts, and others to understand business or industry requirements.
- ➍ Develop data models for applications, metadata tables, views or related database structures.
- ➎ Set up database clusters, backup, or recovery processes.

Outline

- 1 Mindset
- 2 Toolset
- 3 Job market

Mindset

Mindset

Evaluate alternative solutions to solve complex problems.

Data Scientist

- 1 Explore alternative hypotheses.
- 2 Evaluate them on their performance.

Data Engineer

- 1 Break down complex problems to smaller parts.
- 2 Make sure parts hang together.

Toolset

Survey data from US (O*NET 2022)

- 1 Business Intelligence Analysts
- 2 Database Architects

Toolset

Data Scientist

statistics, machine learning, data visualization, R, Python

Data Engineer

relational database management, SQL, document stores, Unix/Linux, cloud services, task orchestration, Python

Job market

Who is hiring?

- Every large firm has a data warehouse → needs data engineers.
- Every data-drive business needs (a few) data scientists.

Average annual salary in Germany (Payscale 2022)

Job title	Entry level	Experienced
Data Scientist	50k	119k
Data Engineer	46k	70k

Median annual salary in USA (O*NET 2022)

Occupation	Salary
Business Intelligence Analyst	109k
Database Architect	123k

Job market summary

- 1 Data scientist jobs are fewer but have more upside risk. (You can become a superstar!)
- 2 There are more data engineer jobs and they are more stable. But you need a multitude of tools.

You need to be a bit of both!