Data Analytics for Management

Week 5: Intro & Excel

Iegor Vyshnevskyi

October 7, 2022

Table of Contents

About me

Hello! My name is *Iegor*.

- Ph.D candidate KDIS
- MSc KDIS (2014)
- MSc KNEU (2009)
- Background: international finance & central banking
- Research interests: banking and central banking

Contact & profile:

- ievysh@kdis.ac.kr
- Google Scholar
- Linkedin
- ResearchGate
- GitHub

About you

- What's your preferred name?
- What are you studying / doing?
- What would you like to do?

General things

- Our Goals: to understand concept and tools to further apply in real-life practice
- See that Data Science is *fun* :-)
- My role: to guide you through the course
- Your suggestions are welcomed (through KSS survey, etc.)!

Our communication

- Feel free to approach me before or after a class
- Office hours: by appointment. You are welcome to discuss course-related issues and questions, carrier plans, etc.
- Emailing policy: email me to submit your home work, set up a one-to-one meeting or in the case of some urgent issues.
- Please indicate the course name / section in the subject line and the issue (e.g., *[Data Analytics for Management]* Meeting request).
- Please write at least two times when you would like to meet and a brief description (1-2 sentence) why you want to meet up with me.

Table of Contents

Logistics & materials

- Learning by doing
- A typical class will be focusing on practical exercises and projects' discussions
- Please see *the updated syllabus*

Course materials:

- Please see GitHub *Repository*[1]

[1] See the instruction about GitHub account there

Broad coverage

- Formal side:
- Concepts
- Understanding
- Real-life cases
- Problem-solving
- Application

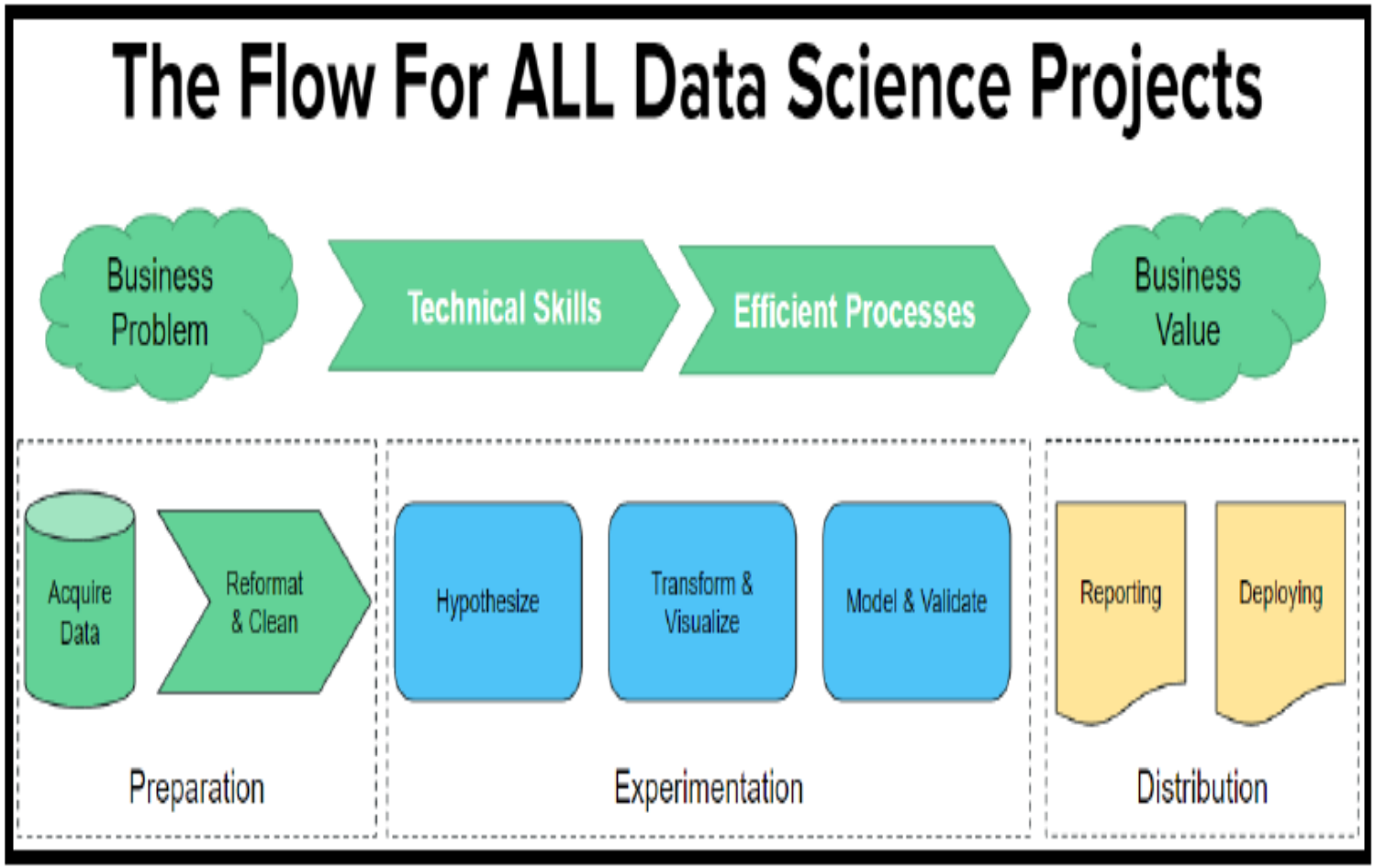
- Hidden side:
- Values
- Communication
- Cooperation
- Discussions
- Standards

Table of Contents

Concepts

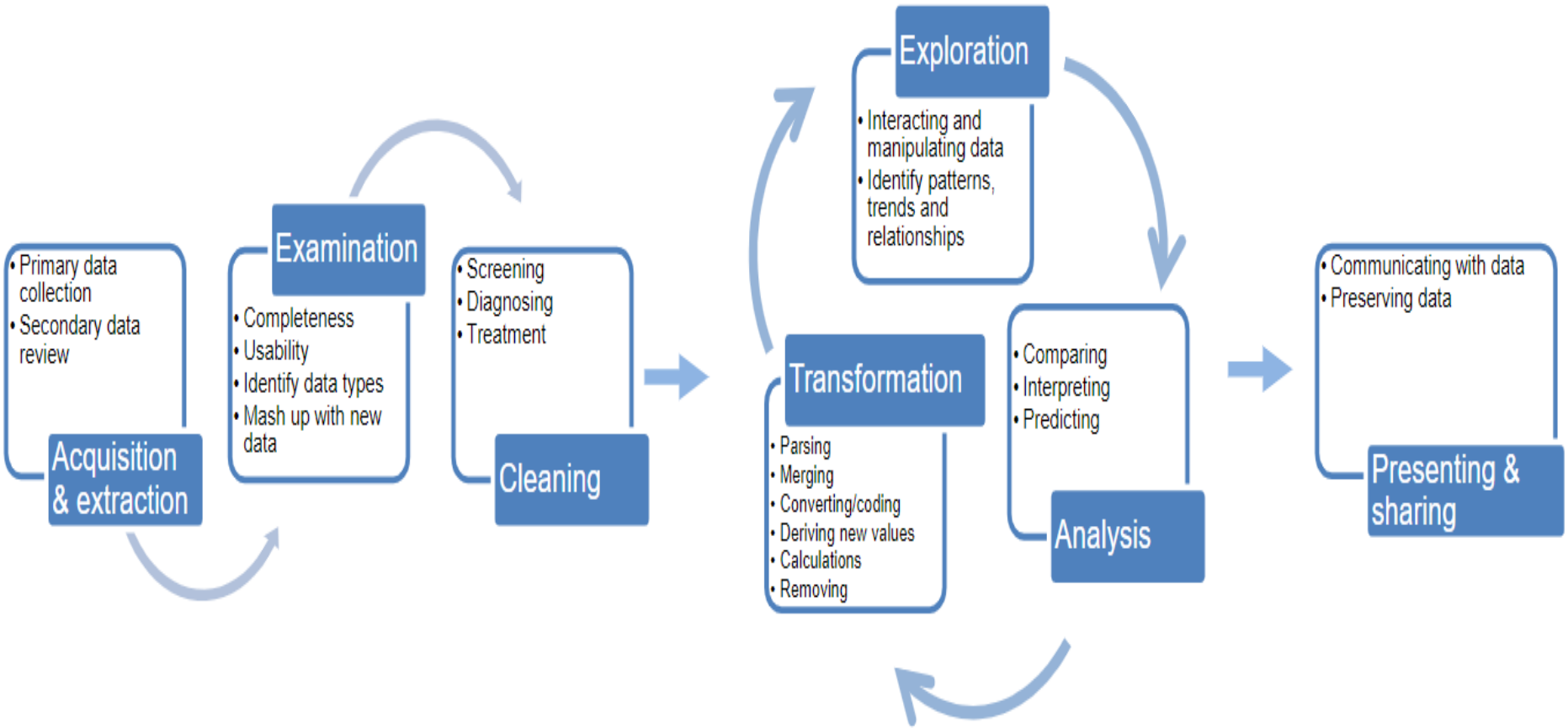
- *Data science* is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract or extrapolate knowledge and insights from noisy, structured and unstructured data,and apply knowledge from data across a broad range of application domains. Data science is related to data mining, machine learning and big data (Wiki).
- *Data analytics* (a subfield of data science) is the science of analyzing raw data to make conclusions about that information (Invest).

Data Science workflow[2]



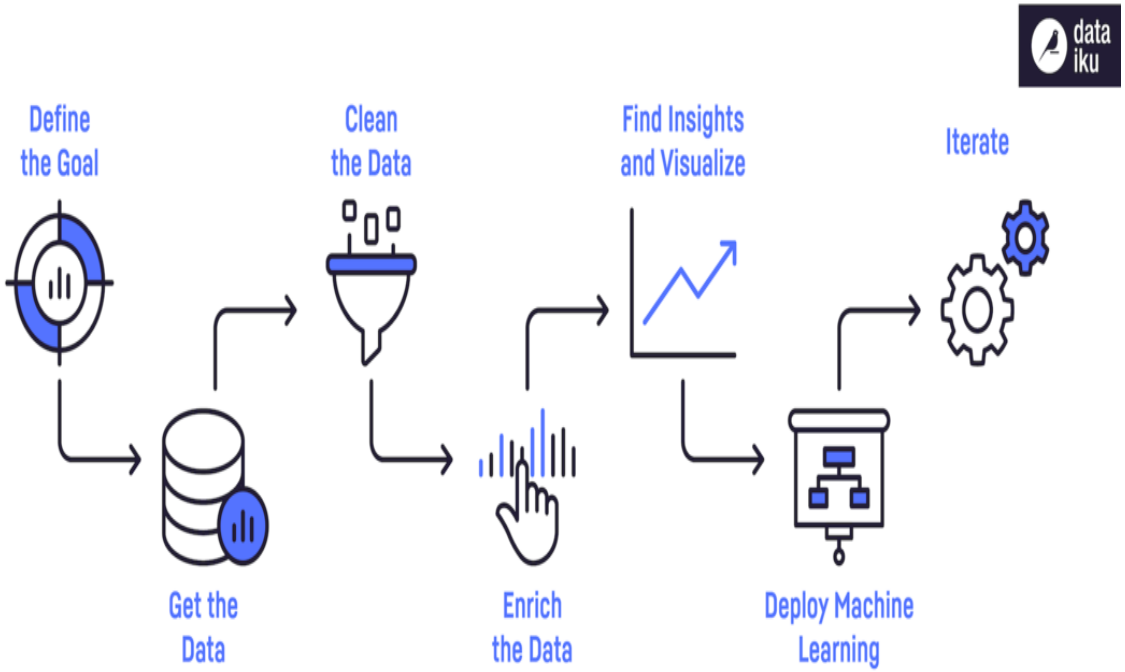
[2] Matt Dancho. Everything You Should Already Know About Data Science

Data Analytics workflow (1)[3]



[3]Kirk 2012, ACAPS 2013

Data Analytics workflow (2)[4]



- Specification of Data Requirements
- Data Gathering
- Data Processing
- Data Cleaning
- Data Analysis
- Data Communication

[4] A Comprehensive Guide on Microsoft Excel for Data Analysis

Analytics Workflow Stages[5]

1. Generate: All the ways data is generated and the systems of record where it is stored or originates from, also referred to as data ingress
2. Collect: All the ways data is collected or ingested
3. Prepare: All the ways data is transformed, including ETL (extract, transform, load), ELT (extract, load, transform), reverse ETL (from a warehouse into business applications), and ML (machine learning)
4. Store: All the ways data is stored, organized, and secured for analytics purposes
5. Analyze: All the ways data is analyzed
6. Deliver: All the ways data is delivered and how it is consumed, also referred to as data egress or data products

[5] Gary Stafford blog. Capturing Data Analytics Workflows and System Requirements

Rules for Data Science[6]

From the board room to the shipping dock, decisions are made every moment of the day using quantifiable, fact-based, trustworthy data (Heine Krog Iversen, Forbes)

1. Usefulness > Complexity
2. Data Quality > Hyperparameter Tuning (i.e., set of optimal values)
3. Simplicity > Novelty
4. Communication > Everything

[6] 4 Powerful Rules for Better Data Science

Rules for Data Analysis

One popular view:

Three rules for data analysis:
PLOT THE DATA, PLOT THE DATA, PLOT THE DATA[7]

[7]Uwe H. Kaufmann

Table of Contents

Rules for working with Excel[8]

1. General:
   - Save original data
   - Work with copied data
   - Save a new version in a new file before updating
   - Have a do-tracking (track changes and to-do things)
   - Add new columns at the far right of your data
   - Check for duplicates

2. Technically specific[a]:
   - One Row of Headings
   - No Empty Columns
   - No Empty Rows
   - All Dates must be in a Single Column
   - Every Unique Data must have its own Column
   - No Totals or Subtotals anywhere in your Table
   - No obstructions around your data

[a] David Brown

[8] based on practical experience

Real-life example

NBU stress test dashboard (link)

Data analysis in Excel

1. Supplementary materials:
   - Video tutorial (link)
   - Excel file: Gapminder (link)
2. What we do:
   - Explore the file. What do we have there (use filter, pivot table)?
   - What can be done?
   - Illustrations: numerical (i.e., mean, SD) and figures
   - Data Communication
   - Conclusions

Data dashboards in Excel

Excel dashboards make it easy to perform quick overviews of data reports rather than going through large volumes of data[9].

1. Supplementary materials:
   - Video tutorial (link)
   - Excel file: Gapminder (link)
2. What we do:
   - Use a previous Gapminder file
   - Build a dashboard (bunch of graphs)
   - Data Communication
   - Conclusions

[9] What are Excel Dashboards?

Computer lab learning by doing activity

Project 1 group discussions

Figure 1: Group Discussion