DigiTrans 2022

Unlocking the Power of Big Data to Inform Labour Market and Education Policies

Mauro Pelucchi,
Head of Global Data Science

Emsi Burning-Glass

01 March 2022

Mauro Pelucchi is a senior data scientist and big data engineer responsible for the design of the "Real-Time Labour Market Information System on Skill Requirements" for CEDEFOP.

He currently works as Head of Global Data Science @ EMSI Burning-Glass with the goal to develop innovative models, methods and deployments of labour market data and other data to meet customer requirements and prototype new potential solutions.

His main tasks are related to advanced machine learning modelling, labour market analyses, and the design of big data pipelines to process large datasets of online job vacancies.
In collaboration with the University of Milano-Bicocca, he took part in many research projects related to the labour market intelligence systems.

He collaborates with the University of Milano-Bicocca as a lecturer at the Master Business Intelligence and Big Data Analytics and with the University of Bergamo as a lecturer in Computer Engineering.

Topics

- What we have, what we need

- AI-based data systems

- Interoperability with other data systems

- New flights and new dimensions for LMI

Context

- Continuously evolving Labour Market:

- Digitalization of professions

- Relevance of Soft skills

- Internationalisation

- New professions and skills emerging

- Smart and Remote working

- Impact of Covid-19 pandemic

- Green transition

PSET CLOUD

What we have / what we need

We already have **official statistics**, that are:

- *Representative*

- *Strong* in terms of value

But we can benefit of **additional, complementary information** that could be:

- *Fast*, to track what's happening now (e.g. Covid-19 Impact analysis)

- *Granular* and *adherent* to real and current market terms, to capture emerging trends analyzing what companies are actually looking for

How to find a similar, complementary source of information?
Using **Web Labour Market**

By providing data-driven insights into the job market we inform people and businesses to design and implement sustainable employability strategies capable of connecting the dots between education, government, and companies.

PSET CLOUD

Real-time Big Data LMI

Transform Online Job Advertisments…   …in insights and analytics
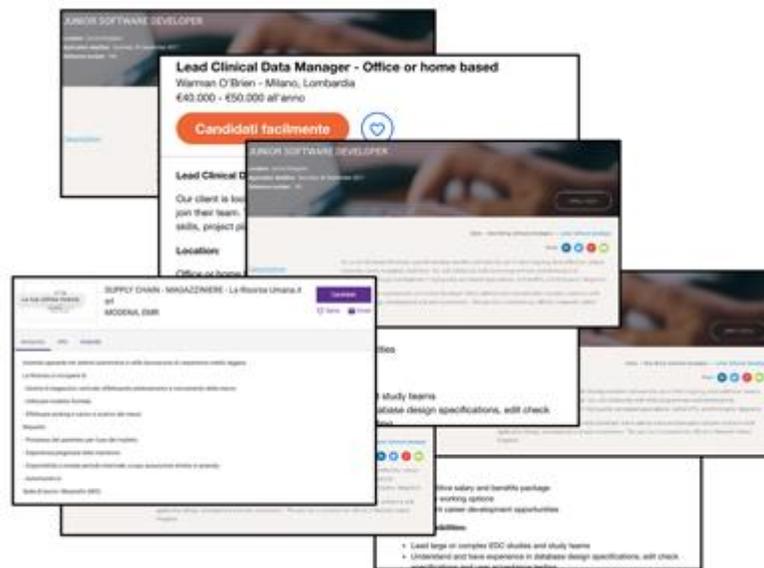
Challenges

PSET CLOUD

- Handle a **huge amount** of near real time data

- Data coming from web → Need to detect and **reduce noise**

- **Multi language** environment

- Need to relate to **classification standards**

- Find a way to **summarize and present** a wide and complex scenario

Collecting Real-Time Labour Market Data

PSET CLOUD

Visit Online Job Sites

Collect & Deduplicate Job Postings

Tagging & Normalising Postings to Generate Detailed Data

- Job Title & Occupation
- Employer & Industry
- Technical Skills
- Foundational Skills
- Certifications
- Educational Requirements
- Experience Levels
- Salaries

Key components

- **Data ingestion**: **collect** raw data from OJV in both structured and unstructured (raw text) formats

- **Data processing**: **classify** data through **machine learning** techniques

- **Data analysis**: **extract information** from data and make it available through **visualization**

Machine learning

Training is the process of making the system able to learn.

Dataset → Sampling dataset and create Training set, (validation set) and test set

Training → Build the model → Model

Test → Validation → Validate Model

Text

Text Preprocessing

Text Transformation
(Feature Generation)

Feature Selection

Data Mining /
Pattern Discovery

Interpretation /
Evaluation
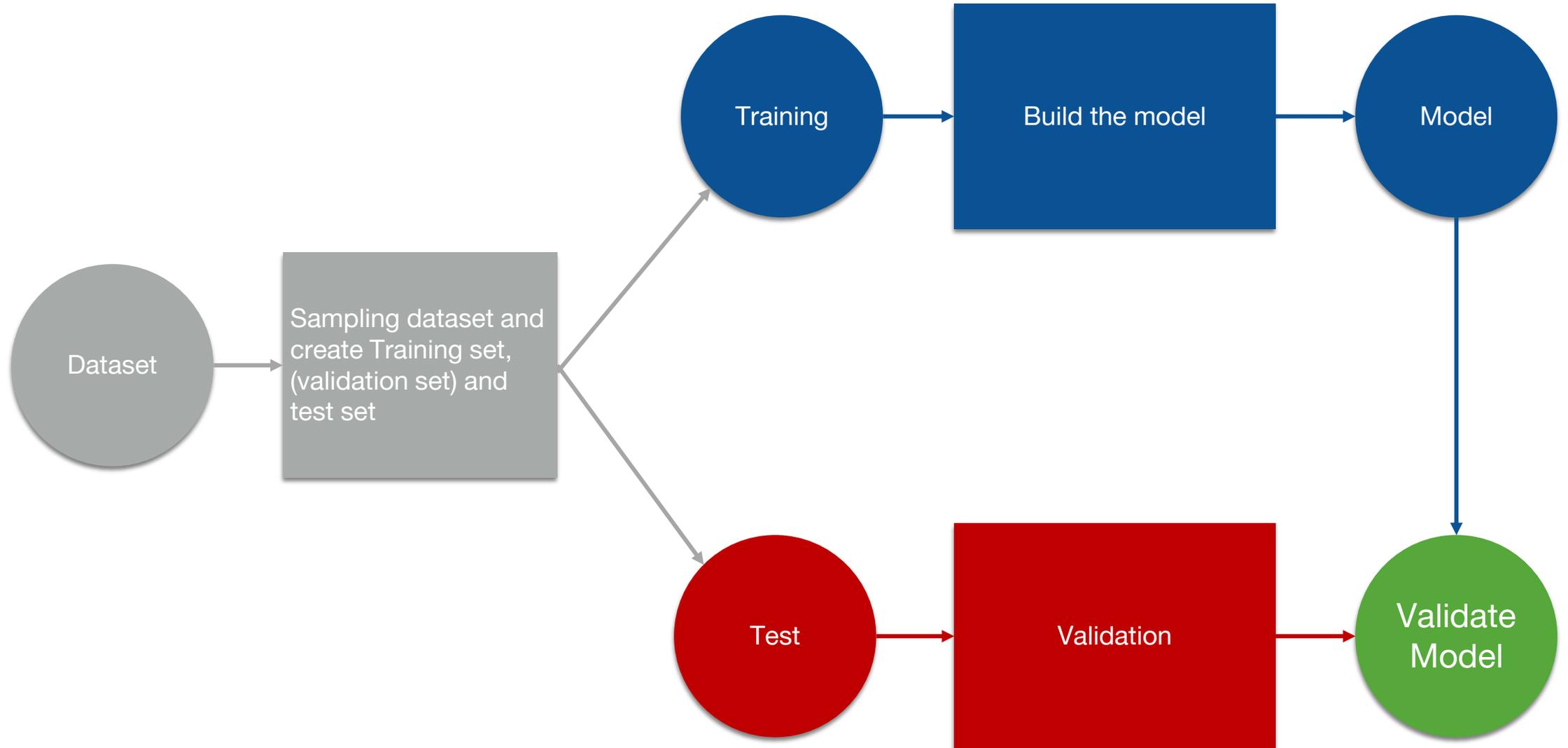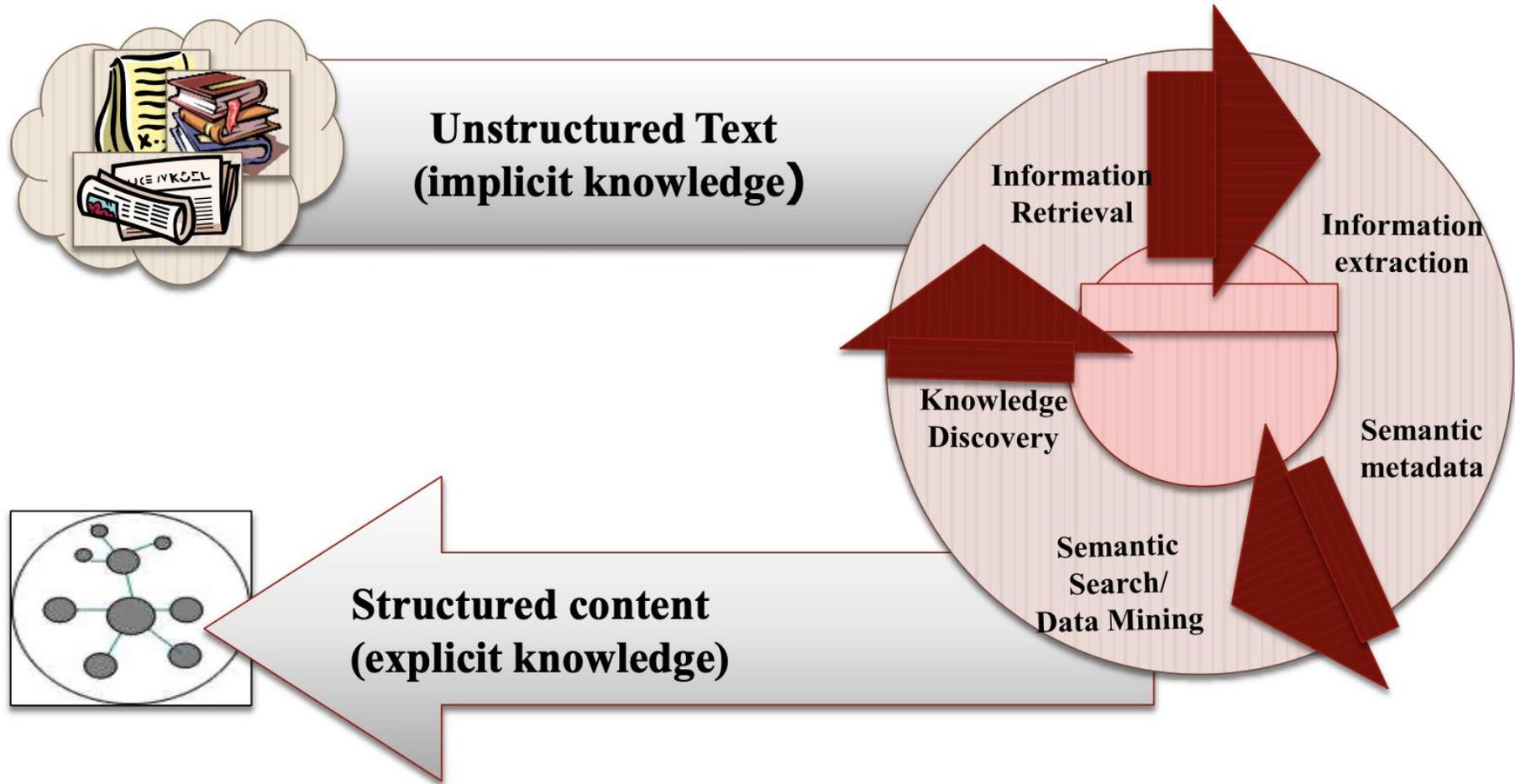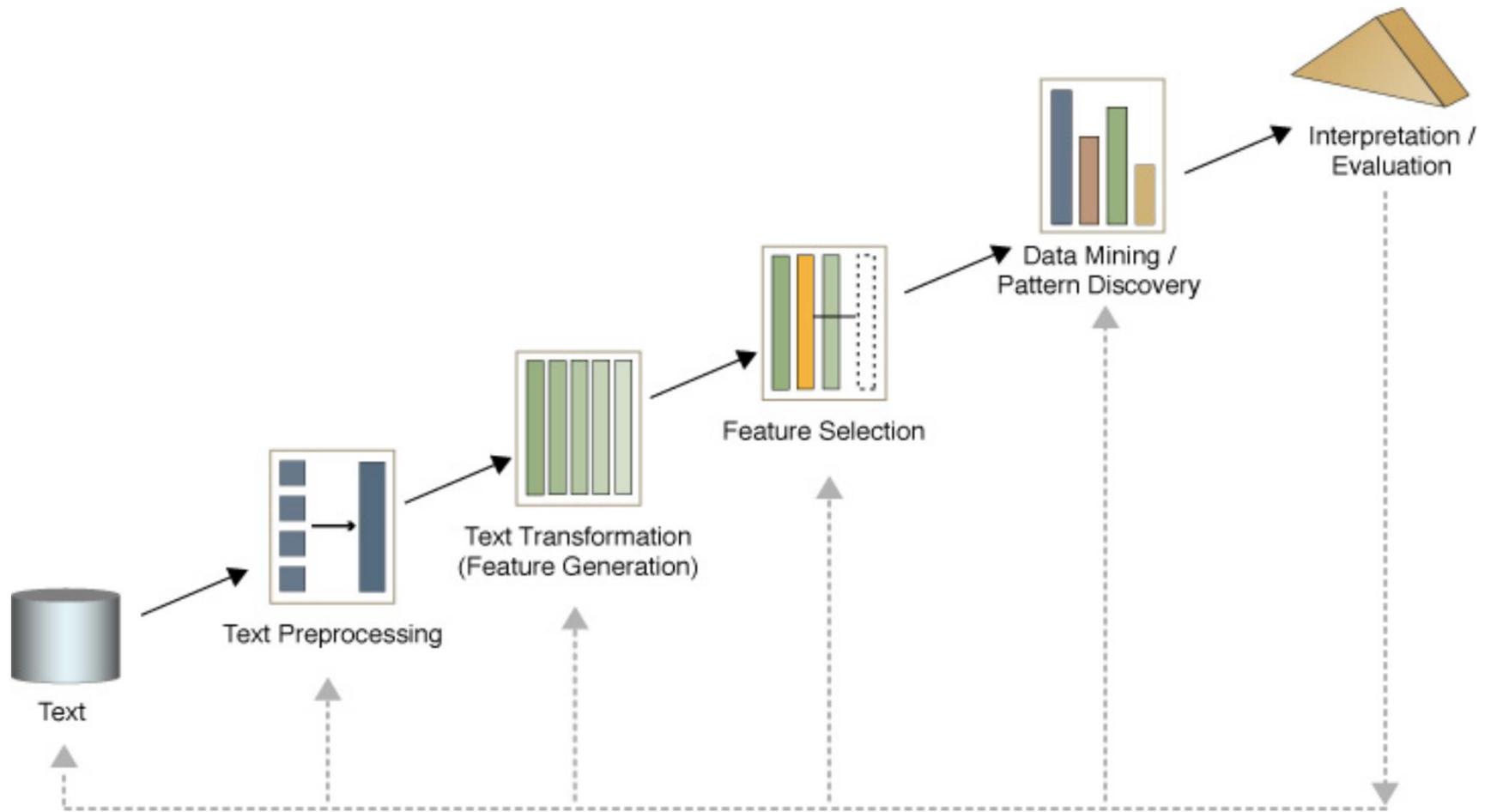
What's we need?
The toolkit

Statistical Methods	Tools	User Experience Research

Time series analysis

Missing data imputations

Data mining

Multilevel modeling

Classification and clustering

Pattern recognition

AB testing

Principal component and factor analysis

Machine learning

Forecasting

Network analysis

Regression techniques

14
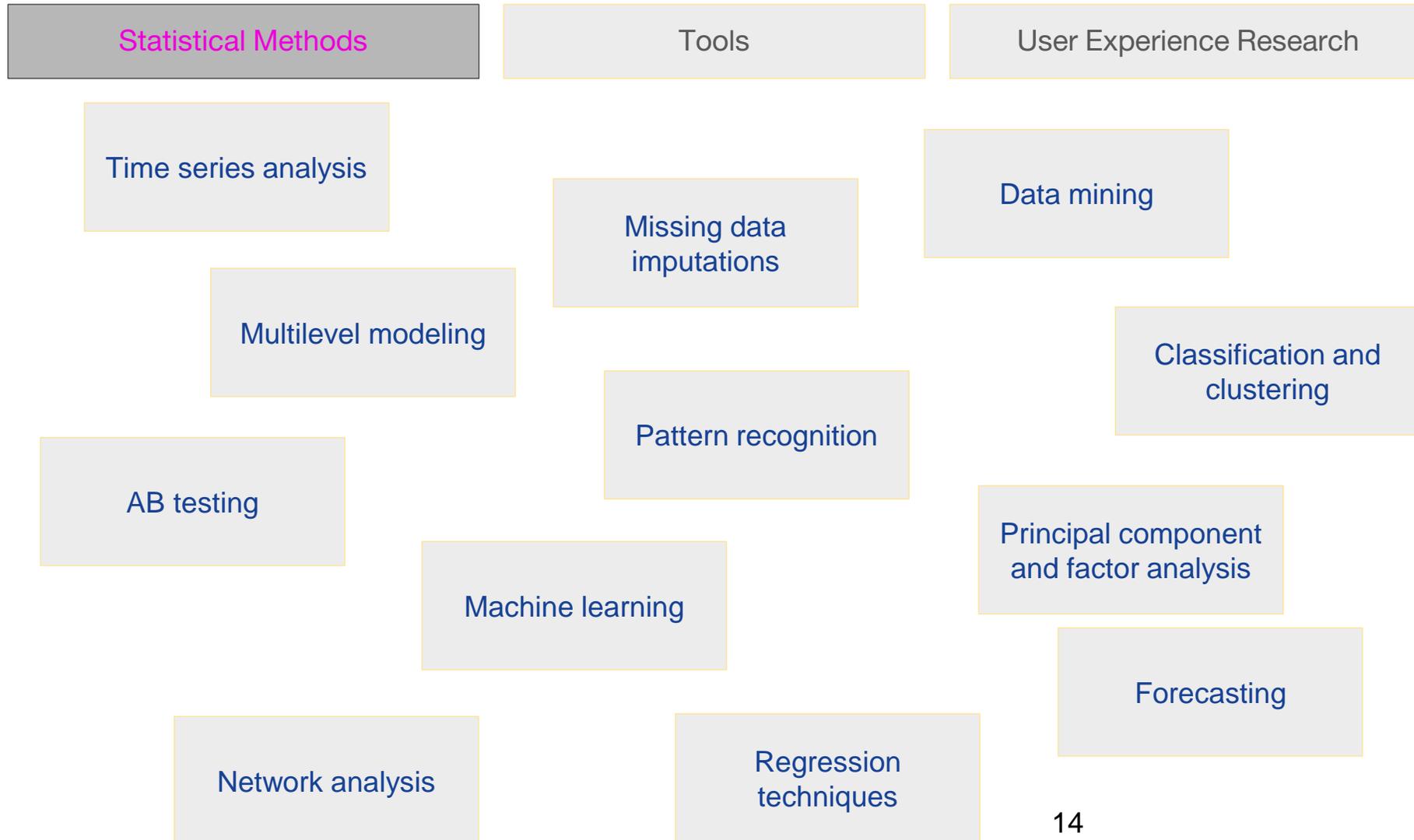
4x4 Traction to explore the labour market

STRUCTURAL LMI

JOB POSTINGS ANALYTICS

PROFILE ANALYTICS

LIBRARIES AND TAXONOMIES

Do job vacancies variations anticipate employment variations by sector? Some preliminary evidence from Italy
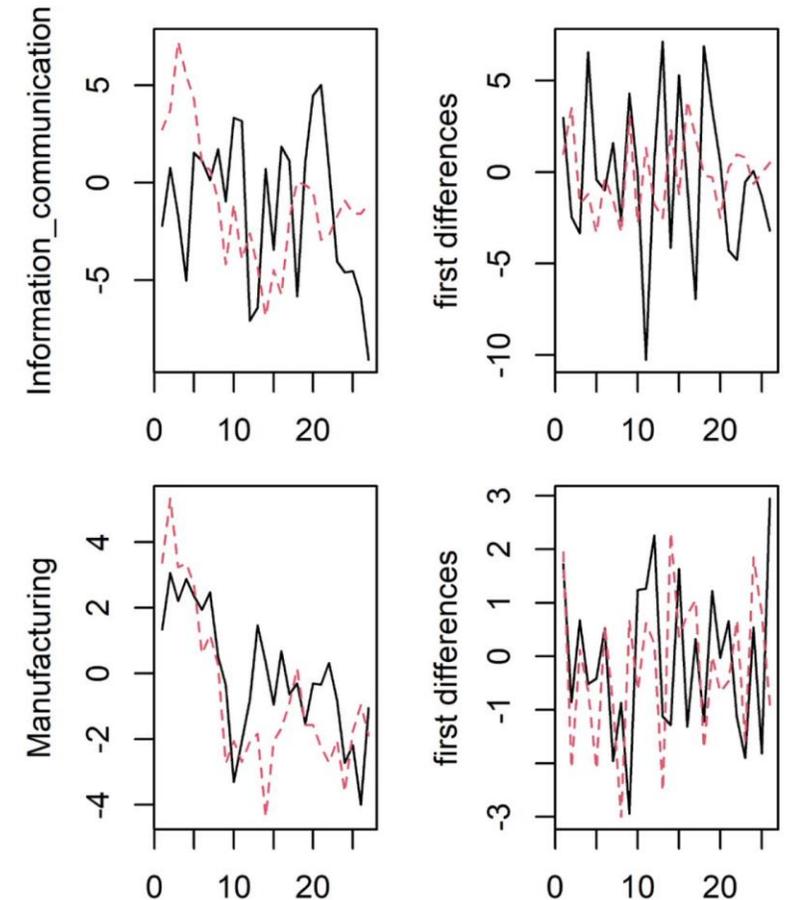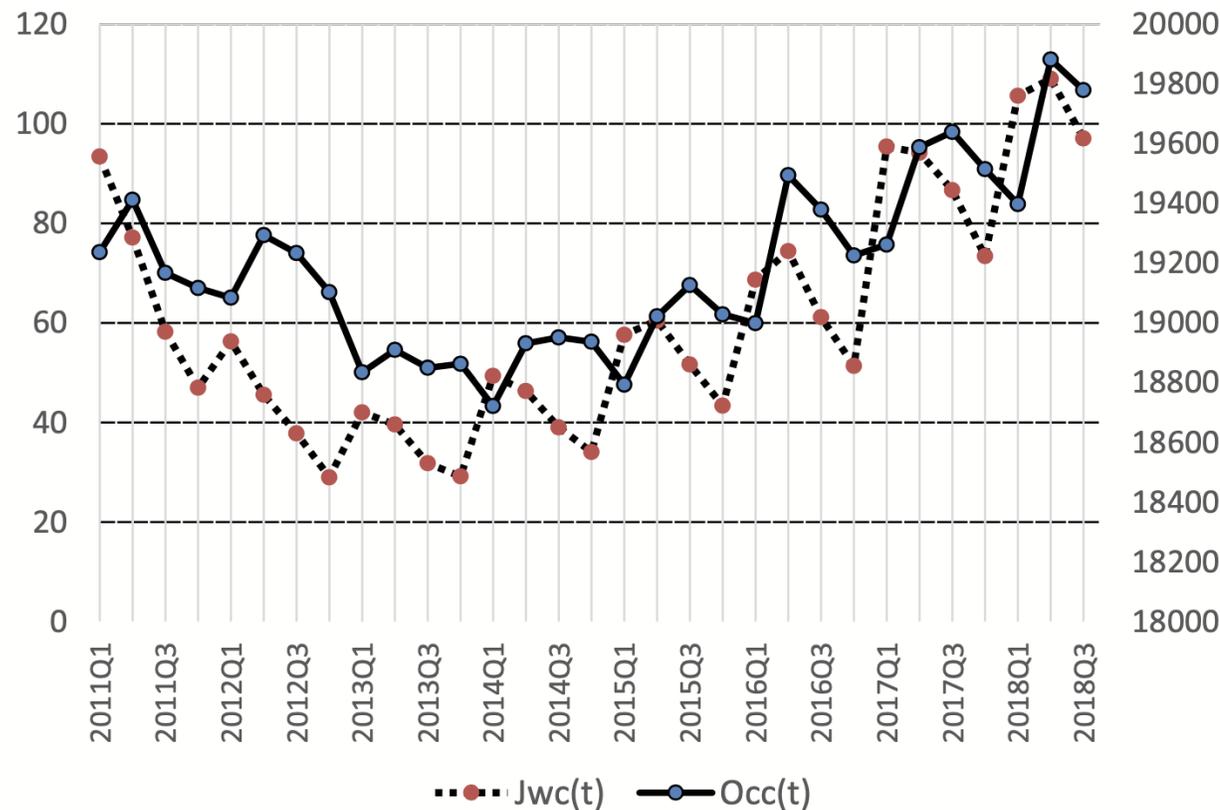
FIGURE 1 Quarterly LFS occupied counts (Occ) and job vacancies counts (Jwc), 2011Q1–2018Q3 (thousands).

Understand & Control for Variances

Job Postings vs. Employment Distribution by Occupation – Germany
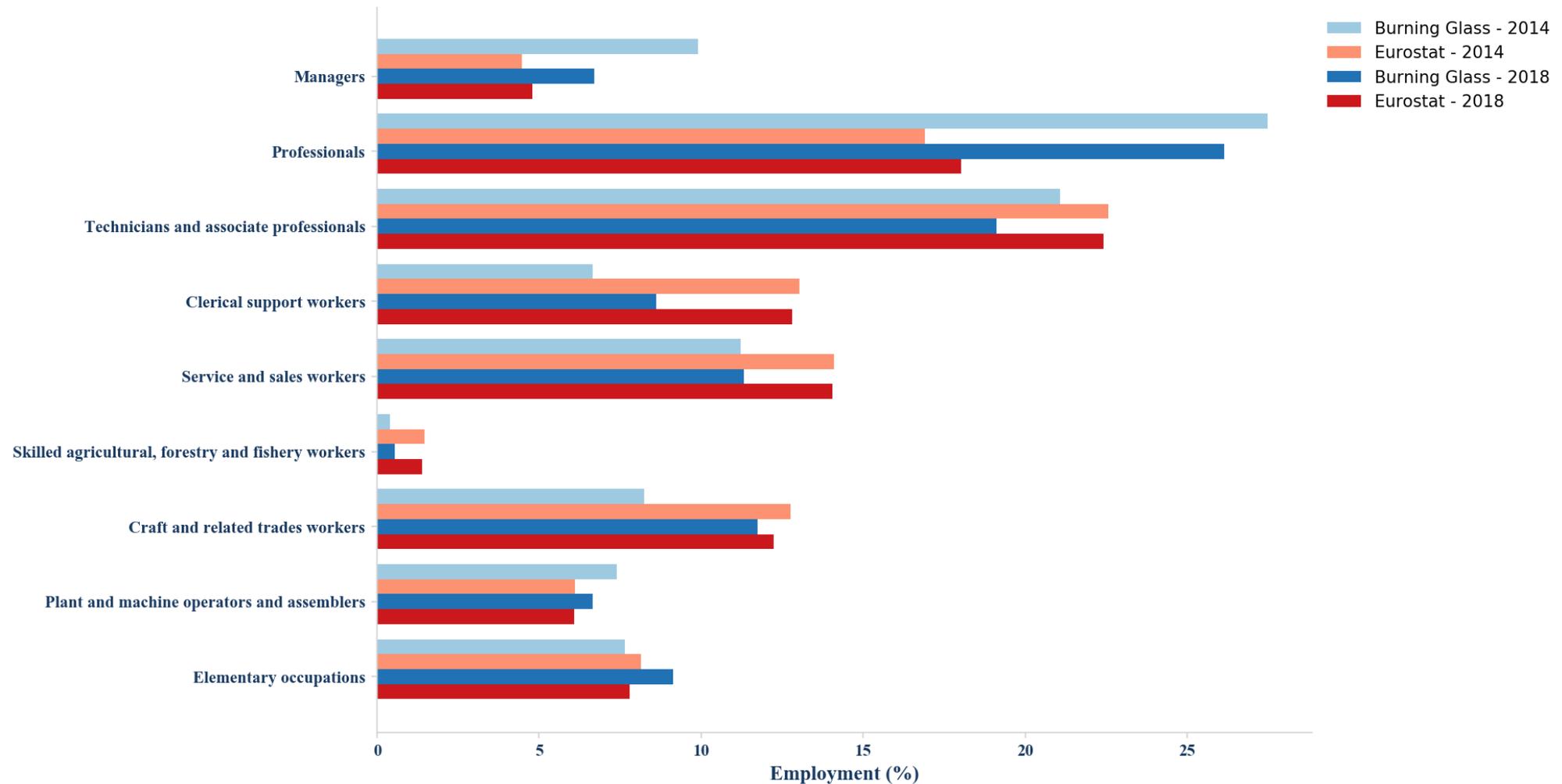
What Is the Best Transition?
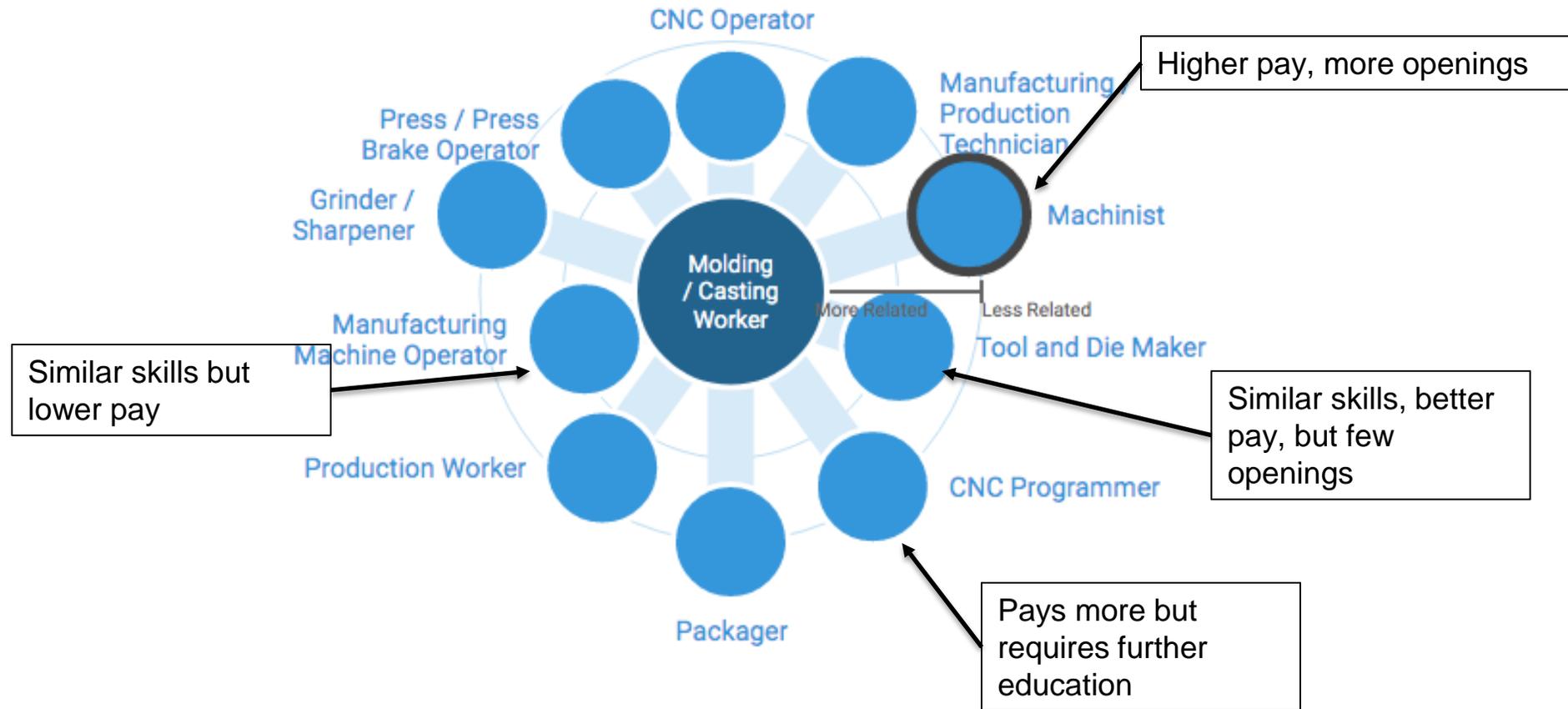
Education

PSET CLOUD

Master BI & Big Data Analytics Università di Milano Bicocca

Target occupations and Alternative Labels

Artificial Intelligence	Big Data	Business Information M..	Cloud Computing
AI system designer	Big Data	Analista dati	Cloud Architect
Artificial Intelligence	Big Data Analista	Business Information	Cloud Computing
artificial intelligence	Big Data Analyst	Manager	Cloud Computing
designer	Big Data Architect	Business Intelligence	Consultant
Artificial Intelligence	Big Data Consultant	Developer	Cloud Computing
Software Engineer	Big Data Developer	Data Analyst	Engineer
	Big Data Engineer	Data mining	Cloud Computing
	Big Data Scientist	Data Scientist	Strategist

Students
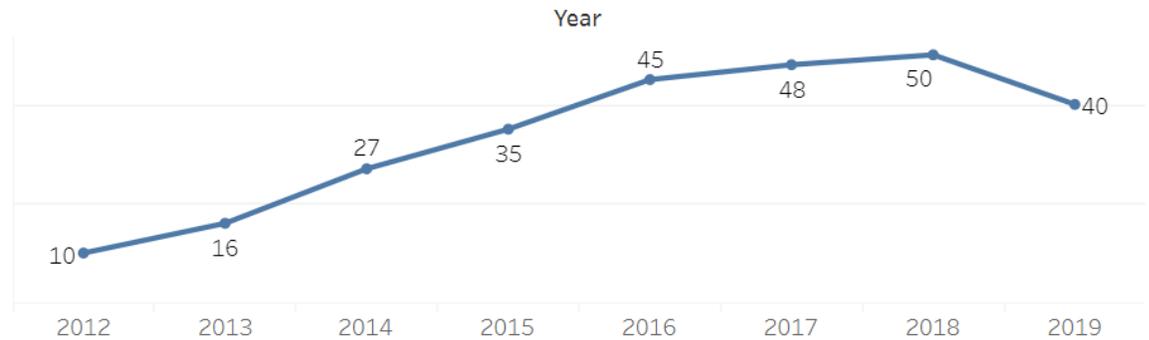
Area
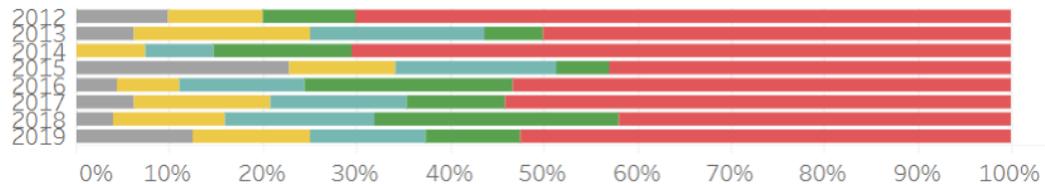
Age
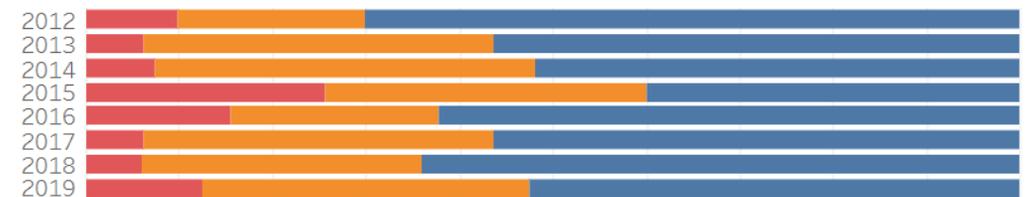
Skills Intelligence

FOCUS COMPETENZE

MECH

TOP 10 PROFESSIONI (ESCO LIVELLO 4)

Professione	Valore
ASSEMBLATORI NON CLASSIFICATI ALTROVE	9,720
DISEGNATORI INDUSTRIALI	7,026
INSTALLATORI E RIPARATORI DI APPARATI ELETTROMECCANICI	6,623
MODELLATORI E TRACCIATORI DI MACCHINE UTENSILI PER LA LAVORAZIO..	6,543
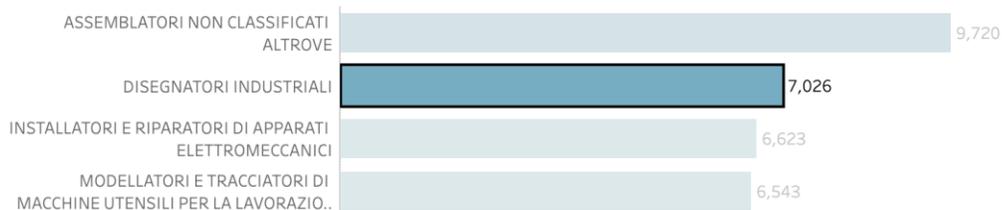
ANNUNCI PUBBLICATI NEGLI ULTIMI 12 MESI

7,026

VARIAZIONE RISPETTO A 1 ANNO PRIMA

41.6%

VARIAZIONE RISPETTO A 5 ANNI PRIMA

24.1%

TOP 10 COMPETENZE PER TIPLOGIA

CONOSCENZE

Conoscenza	%
LINGUE	33.4%
PRODUZIONE E PROCESSI	32.8%
ECONOMIA E CONTABILITÀ	24.2%
INGEGNERIA E TECNOLOG..	21.2%
MECCANICA	21.1%
COMPUTER ED ELETTRON..	17.0%
MEDIA E COMUNICAZIONE	2.6%
EDILIZIA E COSTRUZIONI	0.7%
SERVIZIO CLIENTI E SERVI..	0.5%
AMMINISTRAZIONE D'UFF..	0.3%

COMPETENZE SPECIFICHE

Competenza	%
SOLUZIONE DI PROBLEMI COMPLESSI	21.5%
GESTIONE DEL TEMPO	9.0%
GESTIONE DELLE RISORSE MATERIALI	2.7%
MONITORAGGIO	0.1%
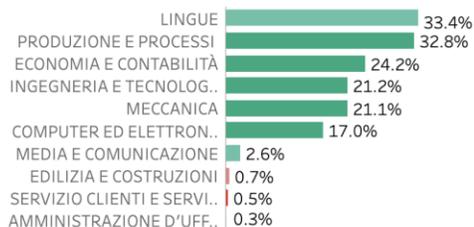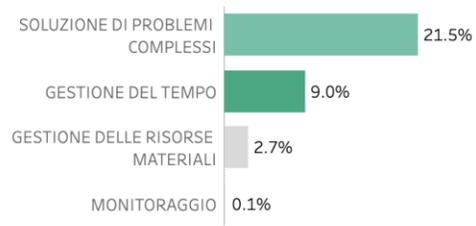
STRUMENTI E TECNOLOGIE

Strumento	%
SOFTWARE ANALITICI O S..	16.4%
SOFTWARE DI GRAFICA O ..	7.8%
SISTEMI INFORMATICI D'..	4.8%
SOFTWARE CAD (COMPUT..	4.6%
SOFTWARE CAM (COMPU..	3.9%
LASER	0.8%
OBJECT OR COMPONENT ..	0.1%

STILI DI LAVORO

Stile	%
ADATTABILITÀ/FLESSIBILI..	48.6%
COOPERAZIONE	29.3%
AFFIDABILITÀ	19.1%
PENSIERO ANALITICO	11.1%
LEADERSHIP	8.8%
PROATTIVITÀ	7.1%
INDIPENDENZA	3.2%
AUTOCONTROLLO	0.7%

ATTIVITÀ LAVORATIVE

Attività	%
PENSARE IN MODO CREAT..	64.3%
INTERAZIONE CON I COM..	50.0%
PROMUOVERE E COSTRUI..	23.5%
VALUTARE LE INFORMAZI..	5.4%
COMUNICARE CON PERSO..	3.9%
GESTIONE E SPOSTAMEN..	3.9%
GIUDICARE LE QUALITÀ DI..	3.6%
DOCUMENTAZIONE / REGI..	3.1%
CONTROLLO DI MACCHINE..	3.0%
STABILIRE E MANTENERE ..	2.4%

ABILITÀ

Abilità	%
ORIGINALITÀ	11.8%
FLUIDITÀ DELLE IDEE	7.2%
ESPRESSIONE ORALE	1.5%

ART-ER
ATTRATTIVITÀ RICERCA TERRITORIO

Demand and supply analysis

Our starting point
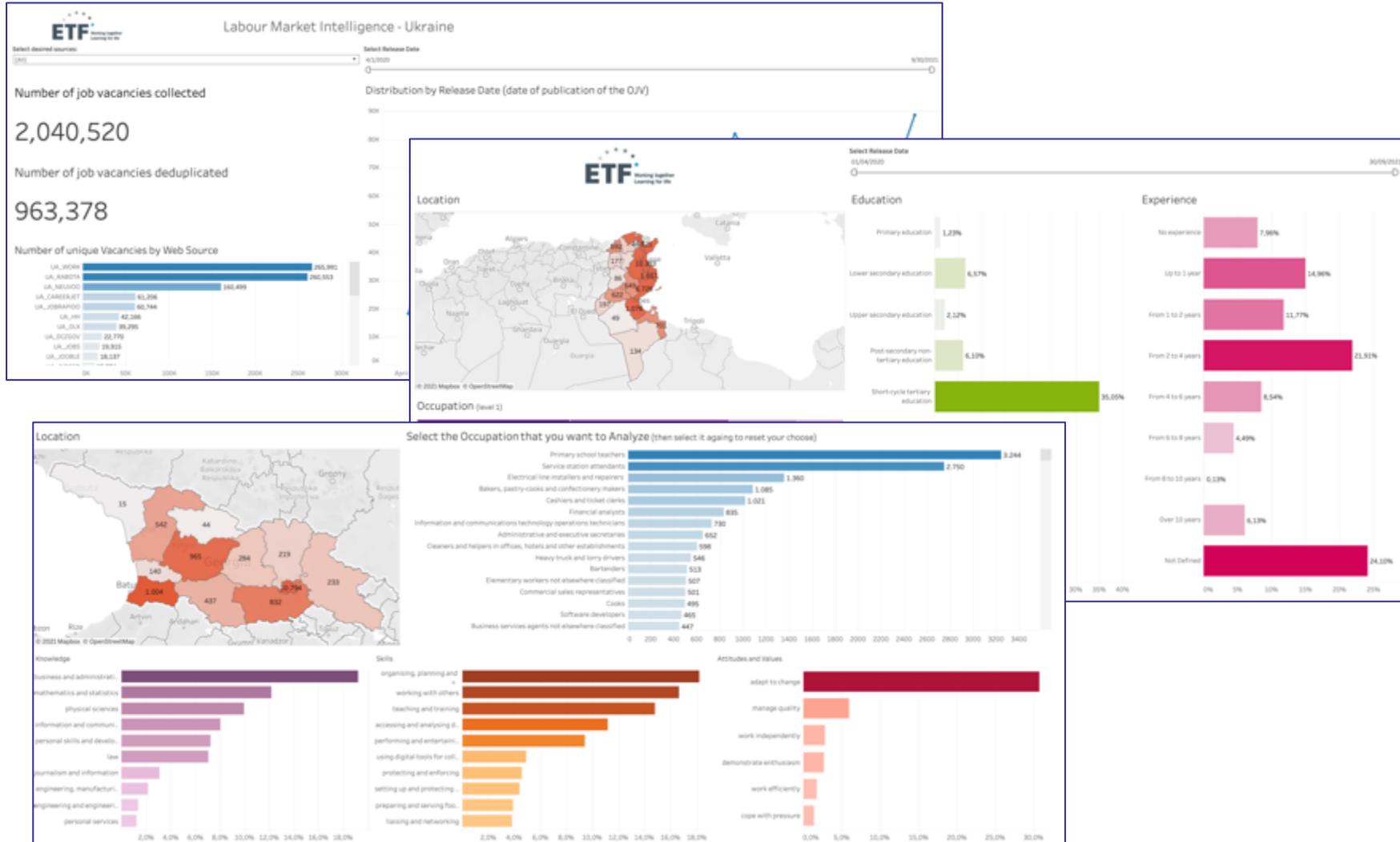
… and now?

- This is a starting point

- Big Data can provide us a gold mine of information

- What we've seen together is just a starting point, the description of what's happening

- One defined a methodology and a system able to collect and classify Big Data, we can extend our view and approach novel analyses

PSET CLOUD

Novel analyses?

- What can we obtain?

- Integration of new data

- Tailored dissemination to different stakeholders

- Re-use of existing components and knowledge

- Definition of new angle of analyses

PSET CLOUD

Why?

- Demand analysis is one of the points of view of Labour Market

- We can obtain additional and complementary information analyzing supply side

- Offered skills

- Matching demand-supply

- Evolution of professional profiles

Supply side data source

PSET CLOUD

- Resumes
  - Detailed
  - Adherent to reality
  - Up-to-date
  - **Unstructured**

- Social profiles (e.g. LinkedIn, Glassdoor, Xing, …)
  - Detailed
  - Adherent to reality
  - Up-to-date
  - **Semi-structured**
  - Private source (data ingestion needed)

Possible data analysis paths - 1

- Supply analysis
  - Most offered professions
  - Top trending skills
  - Geographies with the highest professional availability
  - Careers analysis
  - …

- *See it in action, with an Spain Students Outcomes example*

Spain Students Carrers

Previous Occupation	Current Occupation	Next Occupation

Prev Occupation

Prev Occupation	
Not specified / Student	145.106
Intern/Trainee/Co-op	41.152
Software Developer / Engineer	36.399
Waiter / Waitress	30.904
Computer Programmer	20.583
Office / Administrative Assistant	19.025
Sales Representative	18.624
Business / Management Analyst	18.231
Power Distributor / Plant Operator	16.890
Project Manager	12.373

Day Reoccupation (group)

	<30	<60	<90	<120
Software Developer / Engineer				
Web Developer				
Computer Programmer				
Project Manager				
Business / Management Analy..				
IT Project Manager				
Mobile Applications Developer				

% del totale di ... % del totale di ... % del totale di ... % del totale d

© 2022 Mapbox © OpenStreetMap

Next Occupation

Next Occupation	
Software Developer / Engineer	14.202
Not Specified	5.602
Web Developer	1.354
IT Project Manager	1.226
Project Manager	1.218
Computer Programmer	1.187
Business / Management Analyst	1.160
Computer Systems Engineer / Architect	633
Mobile Applications Developer	610
Chief Information Officer / Director of Information Technology	570

Thank you!

DigiTrans 2022