

Use employment data and compare it to unemployment in more than one indicator

Big Data Department

The presentation is to gain overview on initiatives of Job posts and commons indicators with Labor Force Statistic . Also, to discuss the future of this initiatives.

CONTENTS

- Big Data department Practice
- The Object.
- Target Statistical Population
- Data Collection and process Pipeline
- Web Scraping Challenges
- Data Preparation
- Exploring Data.
- Examples of Indicators.
- Conclusion.

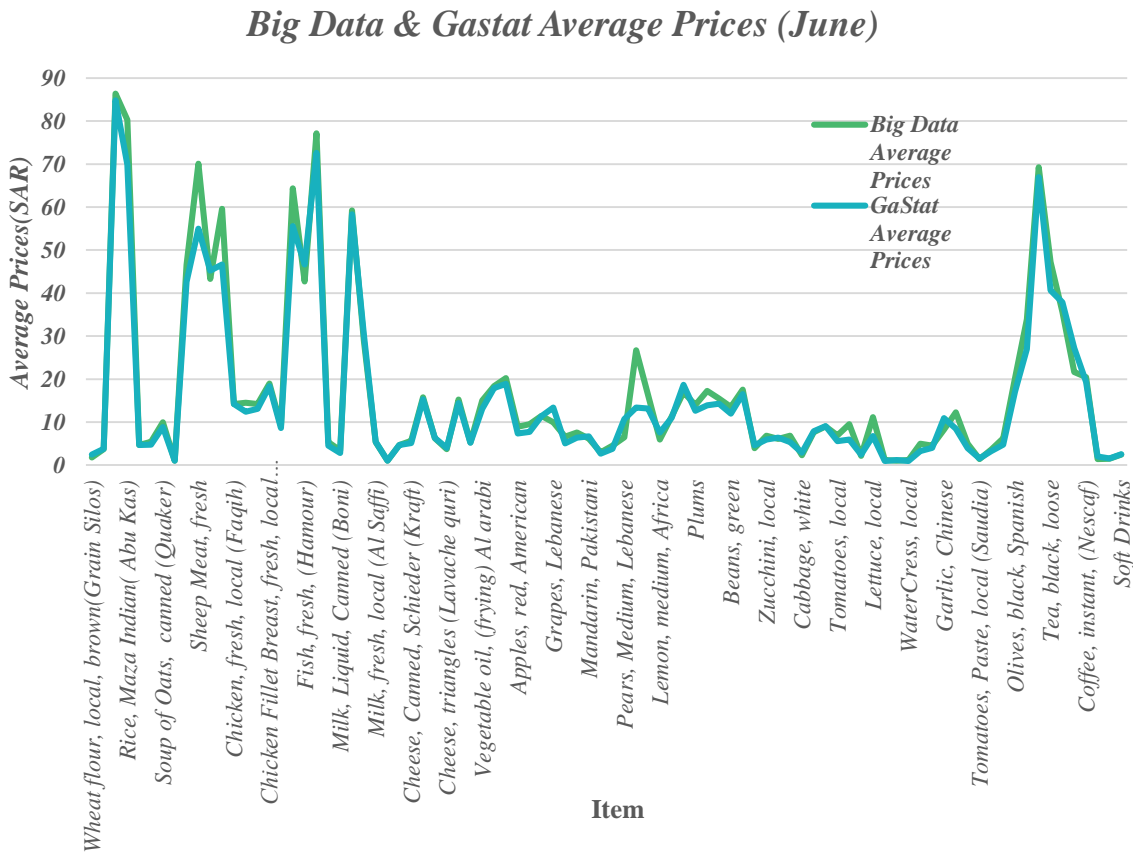
Transportation and Traffic Statistics Using Sensor Data

The object of this project is to use data of road sensors that are installed on the highways to produce official statistics about transport and traffic for the third quarter of 2019.



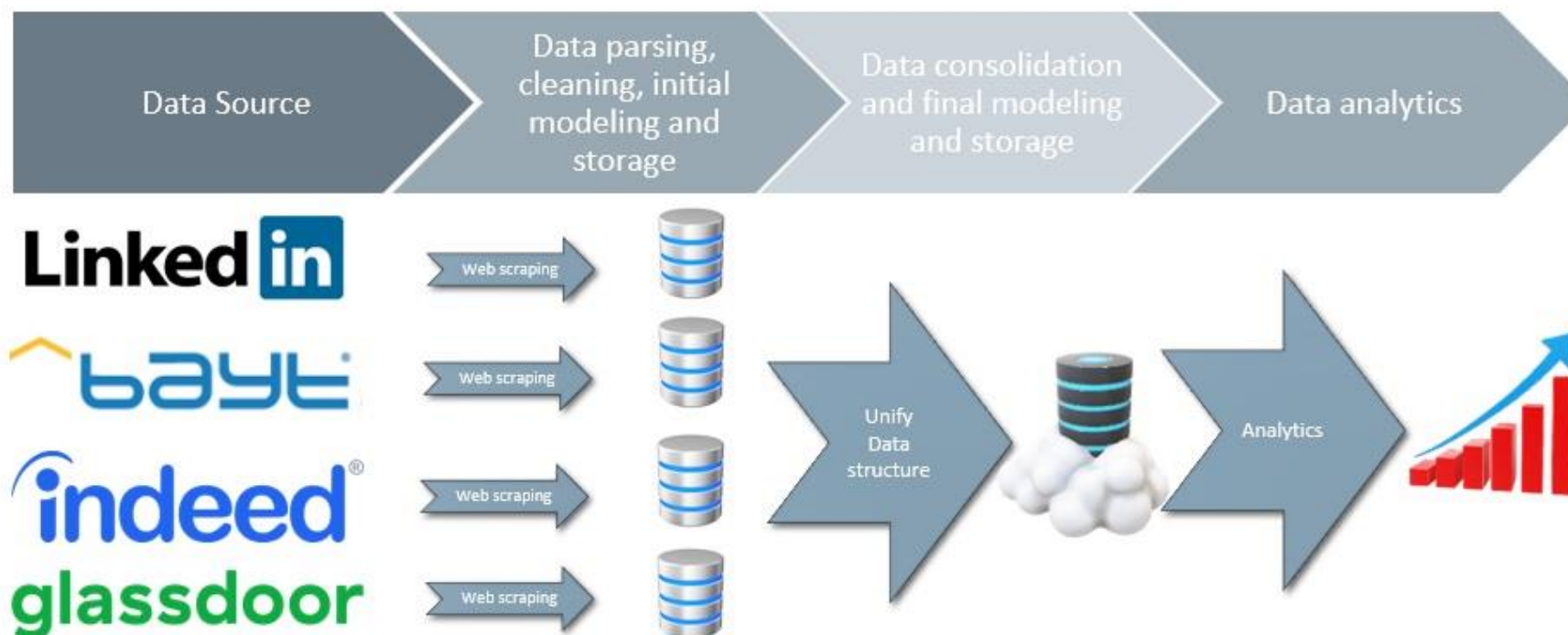
CPI Nontraditional Data Sources

The aim of this pilot study is to compare Gastat’s price averages for foods and non-alcoholic beverages in April ,May and June’s publications with average prices from the web data for the same period to determine how much the web scraping technique might be a reliable and an unbiased source of data that shows consistent results with data collection methods used by the price department (survey).



The purpose of this project is to use employments data to produce statistics and compare it with unemployment data in various indicators. Predict how many new jobs needed each quarter to reduce the unemployment rate to some specific level?

Between two different distribution population. Targeted statistical population is the employment post from career sites for jobs in Saudi Arabia, and LFS UR.



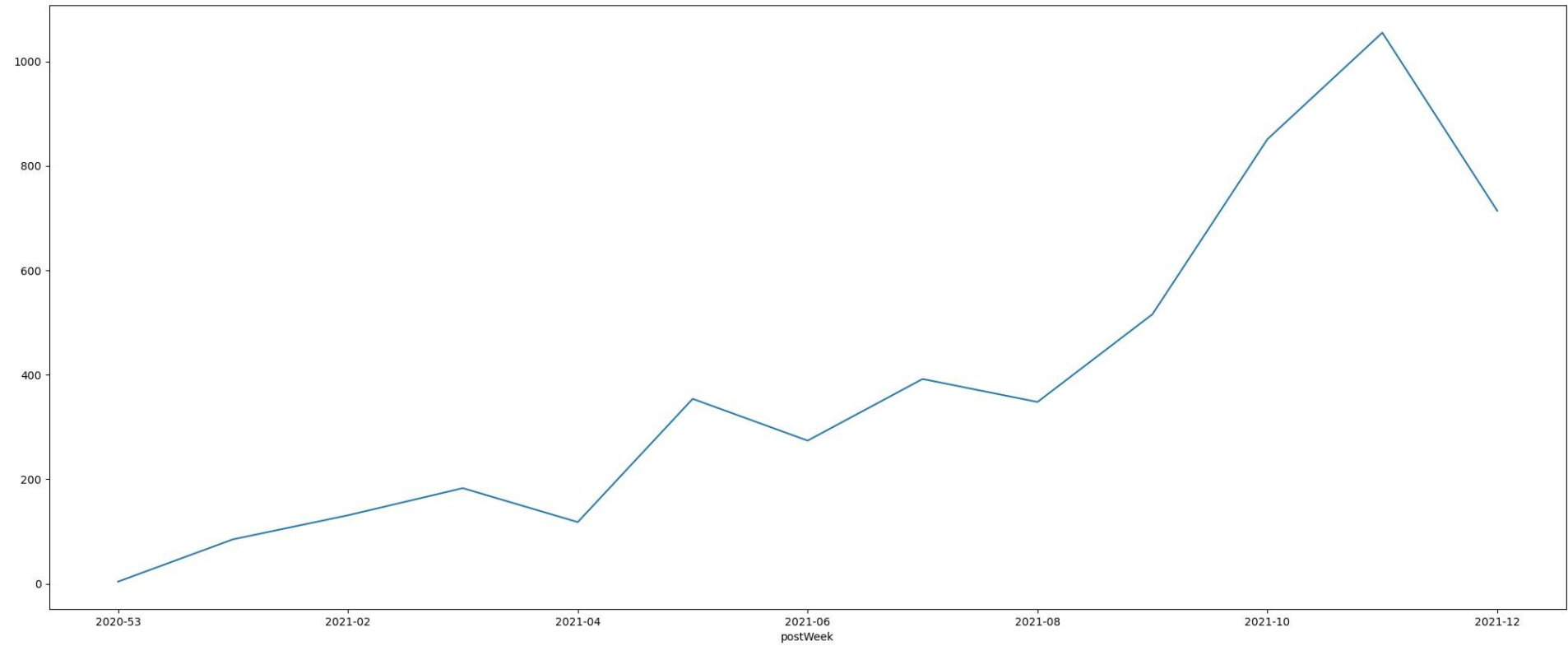
- Different websites structures that need a script to scrape all the needed data.
- Sometimes incomplete data, a second wave of detailed scraping is needed.
- Different website response times, even for the same website.
- Must deal with different timeouts and communication breaks at any collection stages.
- Must deal with websites dynamics as much as possible. For example, closing popup windows.
- Website filtering is not functional and limited to numbers of posts in each search query.
- Ethical Challenges.

- Data cleansing
- Identifying spam posts
- Text Analysis
- ML classification (jobs title & Regions)

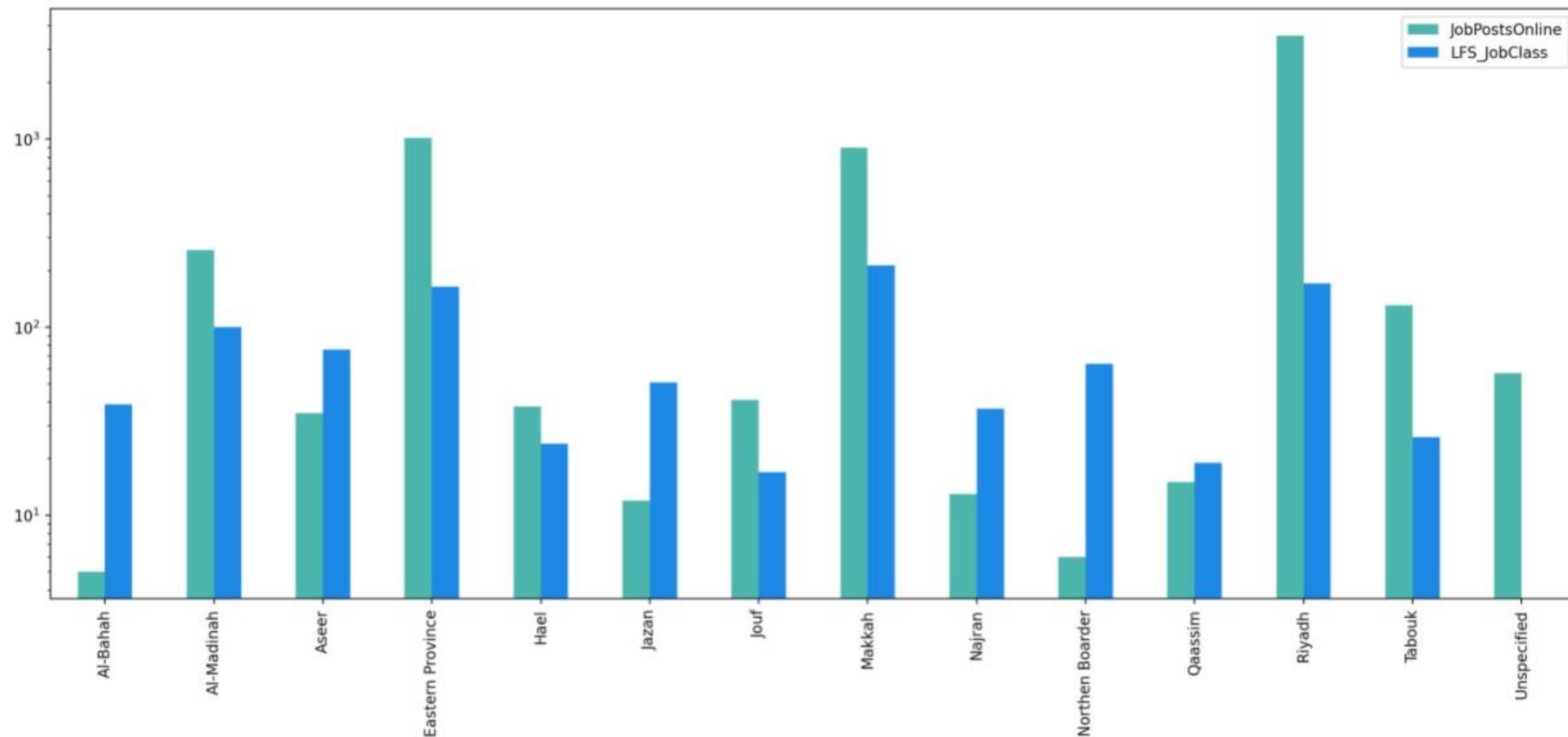
Exploratory Data Analysis & source availability

	LinkedIn	Bayt	Total
Records	9,646	8,520	15,743
Degree / Education	%50	%68	%69
Experience	%59	%58	%76
Industry	%96	%100	%100

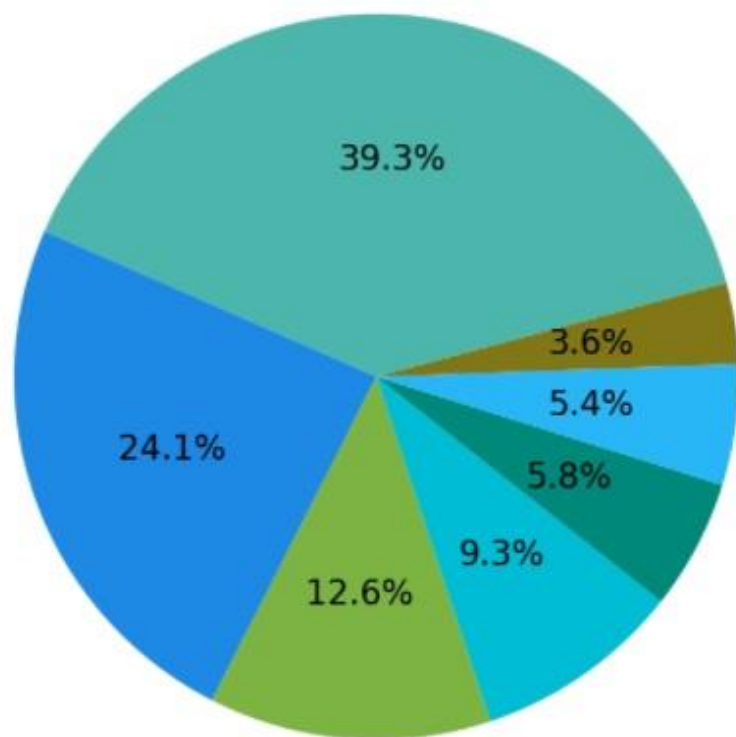
Number of posting Jobs per week in Q1



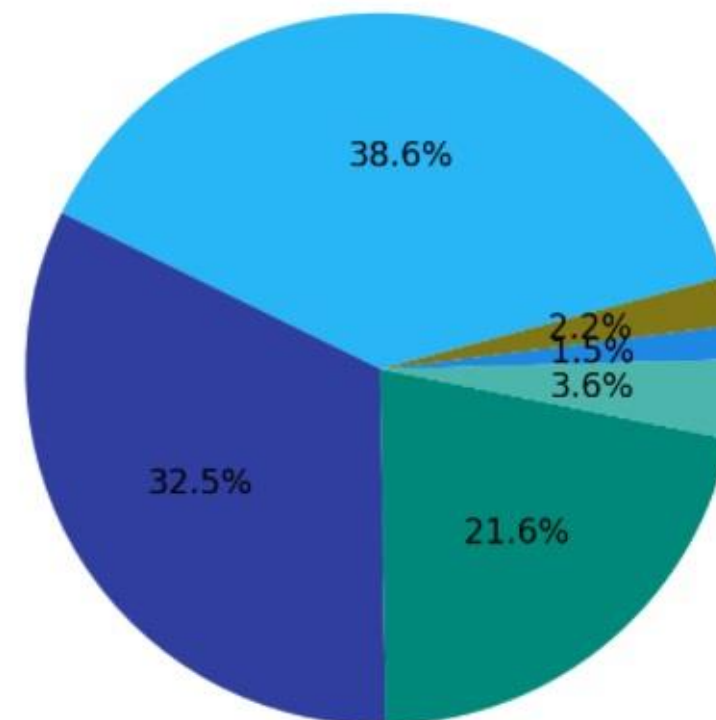
UR & Job post distribution per region



Jobs post per classification

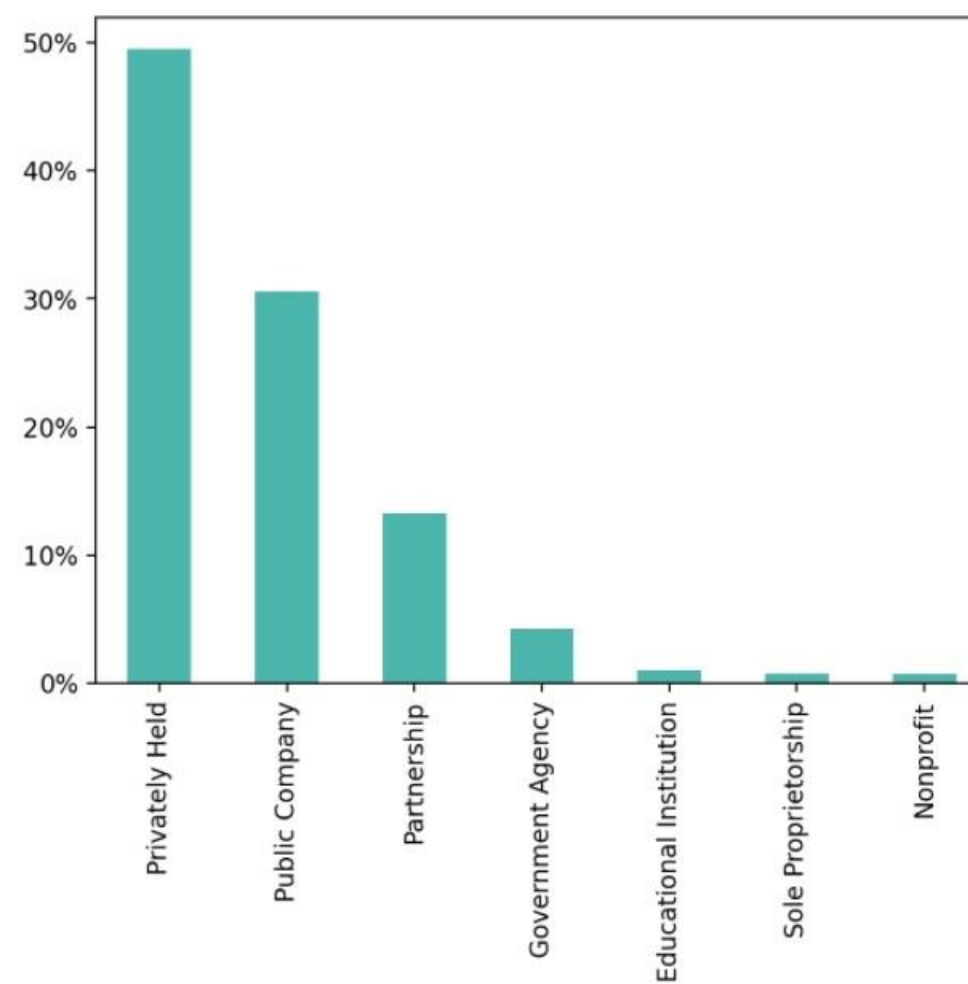
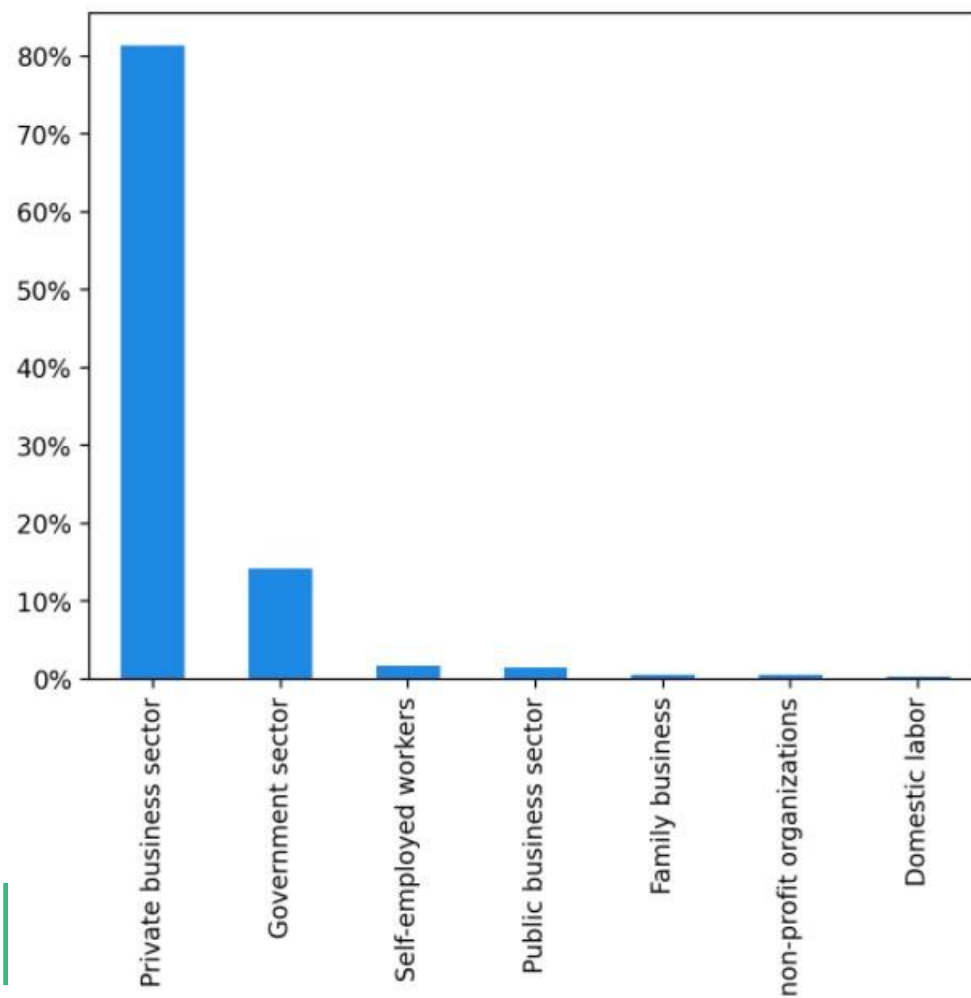


Unemployment LFS



Online Jobs posts

Distribution per company type



By industry

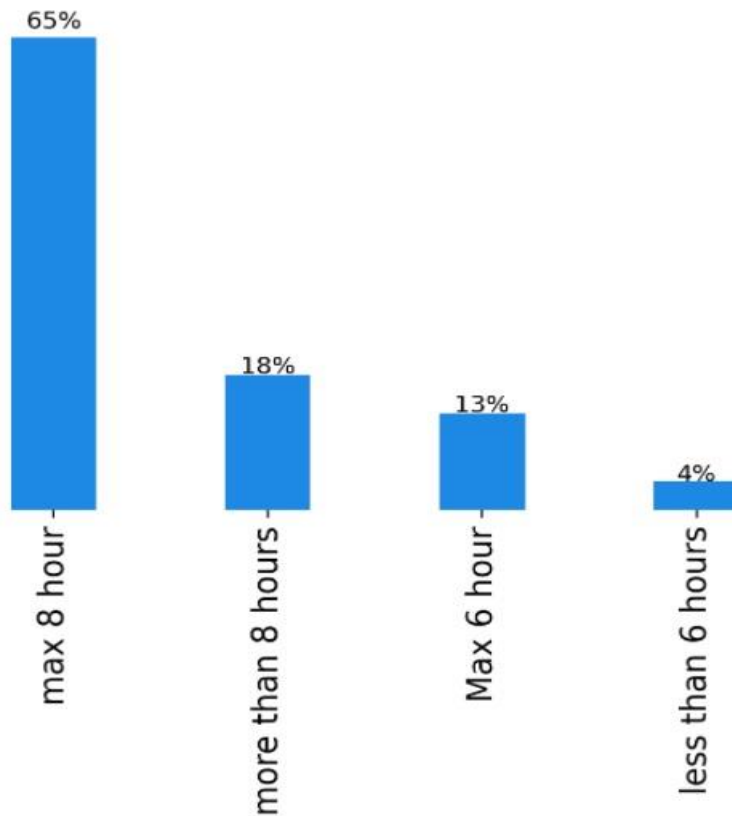
Pervious Industry	%
Education	17.53%
Wholesale and retail trade	17.04%
Manufacturing industry	13.15%
Public administration and defense	11.98%
Construction	9.83%
Accommodation and catering	7.79%
Administrative and support services	4.77%
Other service activates	3.51%
Transportation and storage	32.12%
Professional, scientific and technical activates	2.73%
Other Industry	8.57%

Unemployment pervious industry

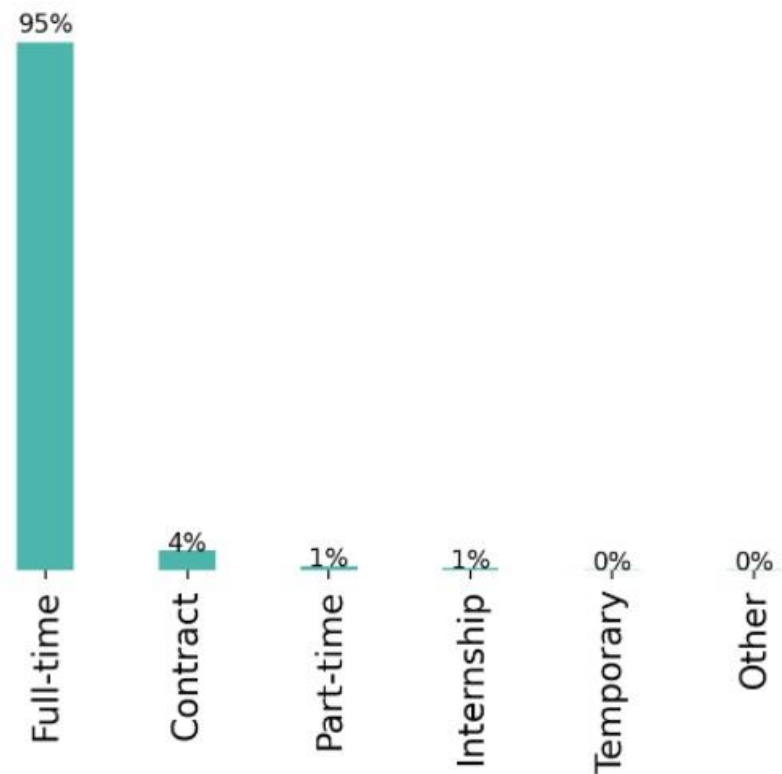
Company Industry	Post %
Staffing & Recruiting	15.34%
Information Technology & Services	14.27%
Hospital & Health Care	11.82%
Oil & Energy	9.82%
Retail	8.96%
Management Consulting	7.06%
Internet	4.72%
Defense & Space	3.15%
Civil Engineering	3.09%
Government Administration	2.31%
Other Industry	19.46%

Online post by Industry

Distribution of Jobs type

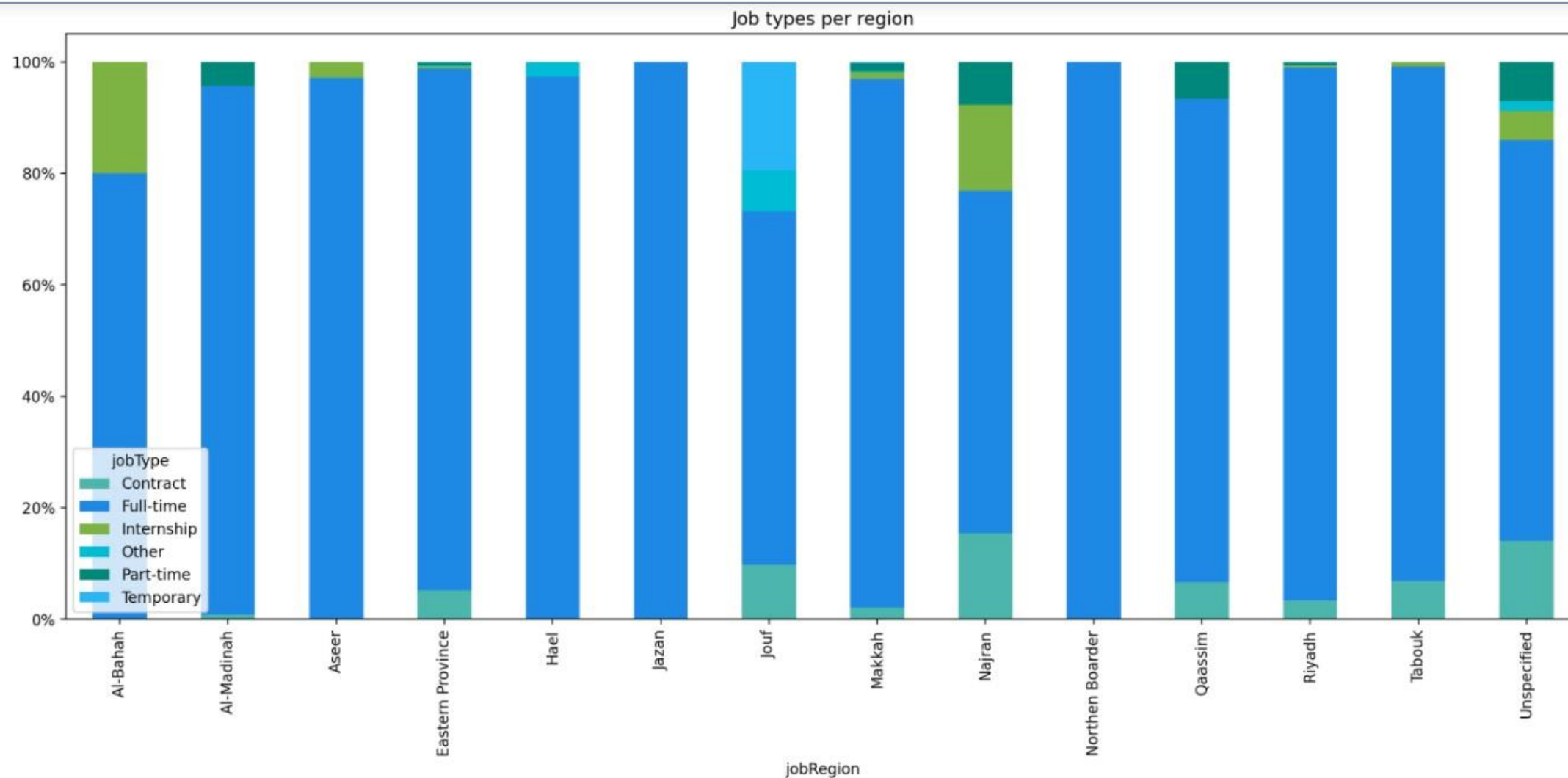


Unemployment LFS



Online Jobs posts

Jobs type per region





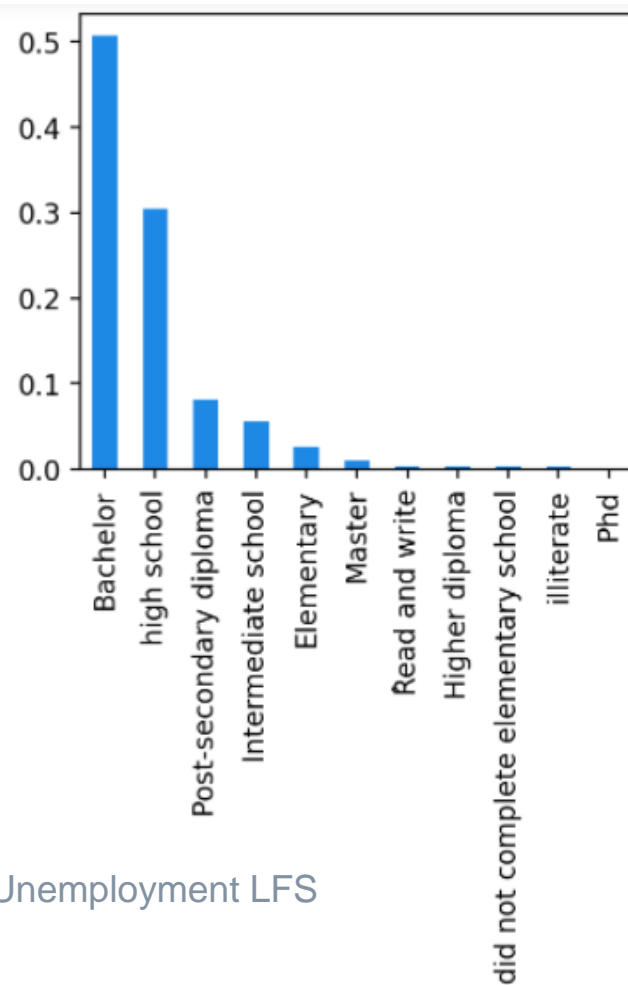
	jobTitle	count	percent
0	Mechanical Engineer	39	0.368063
1	Electrical Technician	31	0.292563
2	HVAC Technician	30	0.283126
3	Accountant	28	0.264251
4	Electrician	23	0.217063
5	Planning Engineer	22	0.207626
6	Civil Engineer	22	0.207626
7	Project Manager	21	0.198188
8	Electrical Engineer	20	0.188750
9	Mechanical Technician	19	0.179313
10	AC Technician	18	0.169875
11	مندوب مبيعات	17	0.160438
12	Graphic Designer	17	0.160438
13	Mechanical Supervisor	15	0.141563
14	Marketing Manager	15	0.141563
15	others	10259	96.819555

Jobs Distribution and Ranking (LFS)

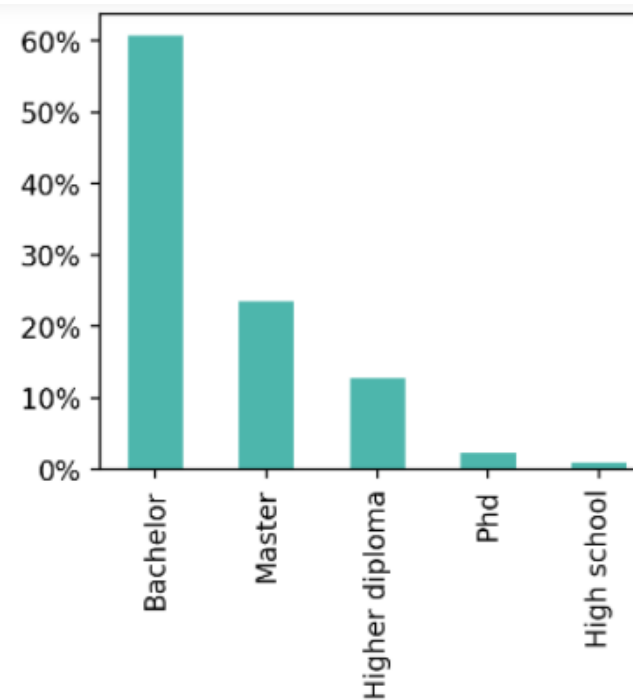
	DD_18_DESC_en	count	percent
0	Private Security Guard	85	6.405426
1	manager assistant	69	5.199699
2	sales representative	55	4.144687
3	guard	53	3.993971
4	seller's Assistant	48	3.617182
5	Public Administrative Clerk	47	3.541824
6	Nanny (teacher)	39	2.938960
7	Internal Security Officer row	29	2.185381
8	Starter Data Entry	25	1.883949
9	Vendor food supplies / year	23	1.733233
10	Defense Warrant Officer	23	1.733233
11	Teacher	22	1.657875
12	Religious education teacher	22	1.657875
13	Readymade seller	21	1.582517
14	Starter reception of my hotel	19	1.431801
15	others	747	56.292389



Distribution of Education levels



Unemployment LFS



Online Jobs posts

Going beyond availability, accessibility, there is a need for more disaggregated data, such as by region, education level, etc.....

These challenges provide an opportunity to use non-traditional data sources as a complement to existing data and approaches.

However, Due to Data size and the difference distribution populations it recommends that this trial to be continued for another six months in order to harmonize data and add more data sources (like official government vacancies data) to reach a conclusion on whether web data is a valid source to present statistics.