

Big Data système semi-automatisé d'information sur la demande (emploi et compétences)

Webinar 14 October 2020

Session 3
Brève introduction sur la methodologie; tableau de bord

Alessandro Vaccarino

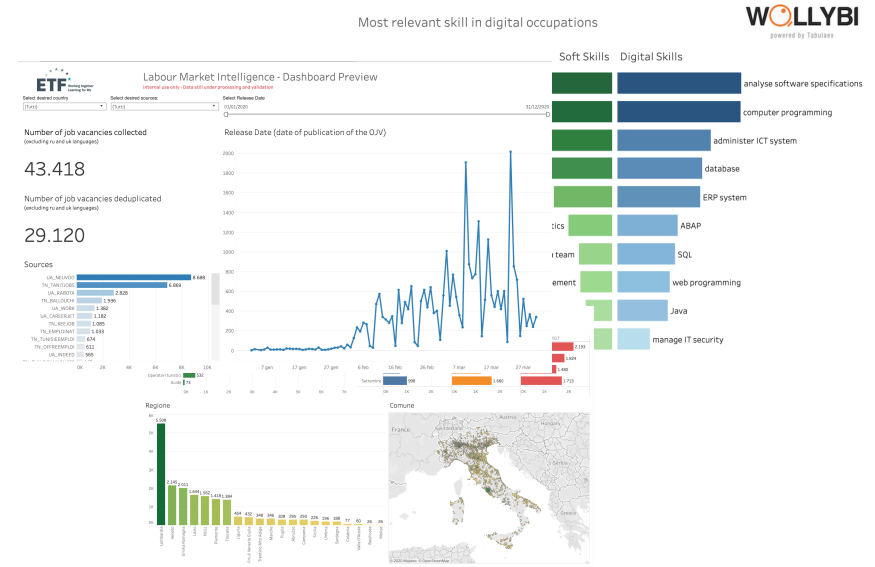
QUEL EST LE BUT?

Transformer ceci...

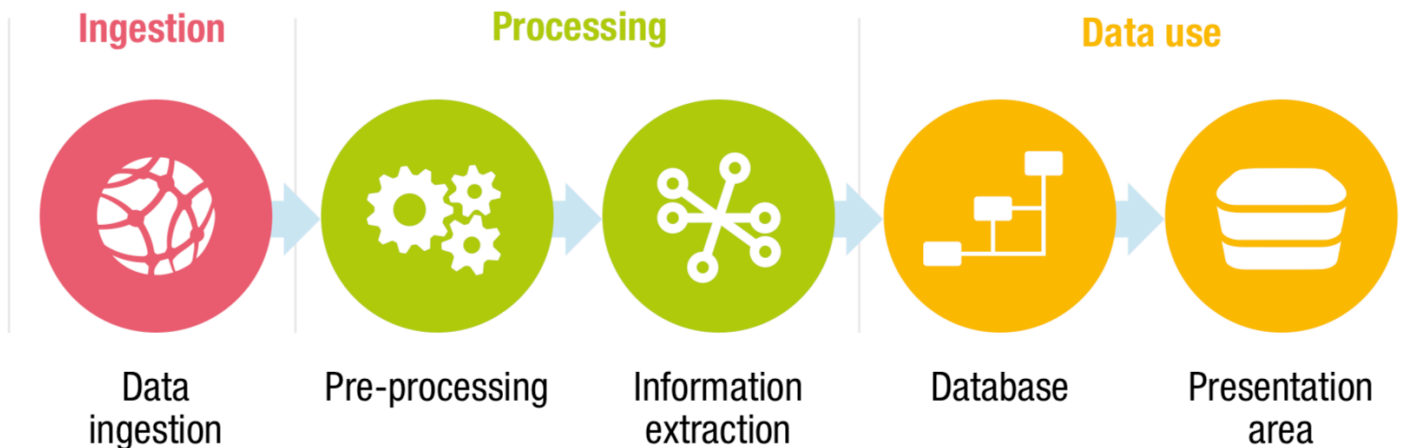
Examples of job ads in the collage:

- JUNIOR SOFTWARE DEVELOPER** (LinkedIn)
- Lead Clinical Data Manager - Office or home based** (Warman O'Brien - Milano, Lombardia, €40.000 - €50.000 all'anno)
- JUNIOR SOFTWARE DEVELOPER** (Indeed)
- SUPPLY CHAIN - MAGAZZINIERE - La Risorsa Umana.it** (MODENA, EMR)

...dans ceci



COMMENT LE FAIRE?



Ingestion de données

Prétraitement

Extraction d'informations

Base de données

Zone de présentation

INGESTION DE DONNÉES – SÉLECTIONNEZ LES BONNES SOURCES

Le marché du travail web peut offrir un large éventail de sources:

- Portails nationaux/internationaux
- Sectoriel/générique
- Sites Web de l'entreprise/Journaux/Agrégateurs/Services publics de l'emploi/...
- Réseaux sociaux
- ...

Comment sélectionner les plus appropriés et les classer?



INGESTION DE DONNÉES – CARTOGRAPHIE E CLASSEMENT DES SOURCES

Avec le soutien d'experts LM dans les pays spécifiques analysés, une liste des portails les plus pertinents a été produite, dans le but de renforcer la fiabilité et de qualité des sources analysées.

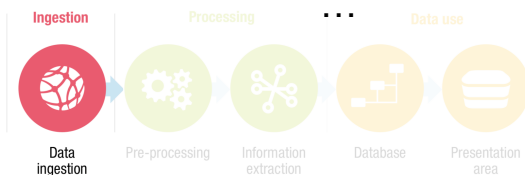
Nombre d'Offres emploi en ligne (OEL) publiés

Type de portail

Données structurées/non structurées

Classement des pages sur Google (proxy d'utilisation)

Final Ranking
2
4
8
7
6
12
13
11
1
3
10
5



INGESTION DE DONNÉES – COLLECTE DE DONNÉES – OBJECTIF...

JUNIOR SOFTWARE DEVELOPER

Location: United Kingdom
Application deadline: Saturday, 30 September 2017
Reference number: 100

APPLY NOW

Home > Now Hiring: Software Developers > Junior Software Developer

Share: [in](#) [t](#) [g+](#) [e](#)

Description

As Junior Software Developer, you will develop excellent software for use in field mapping, data collection, sensor networks, street navigation, and more. You will collaborate with other programmers and developers to autonomously design and implement high-quality web-based applications, restful APIs, and third party integrations.

We're looking for a passionate, committed developer that is able to solve and articulate complex problems with application design, development and user experiences. The position is based in our offices in Harwell, United Kingdom.

Title:

Développeur de logiciels Junior

Area:

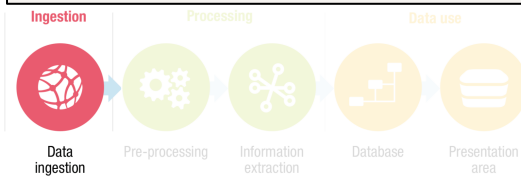
Royaume-Uni

Time:

Samedi 30 septembre 2017

Description:

En tant que développeur de logiciels Junior, vous développerez un excellent logiciel pour une utilisation ...



INGESTION DE DONNÉES – COLLECTE DE DONNÉES – ... STRATÉGIE...



Focus sur les
volumes



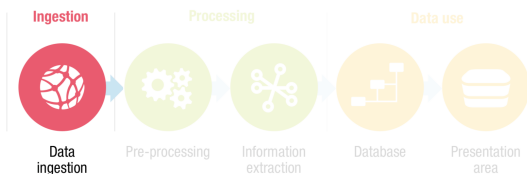
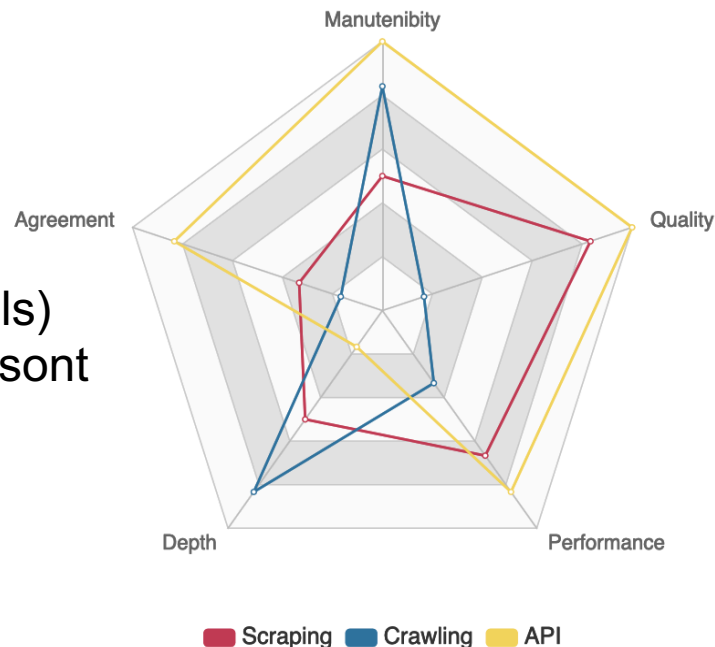
Augmentation de la
couverture et
Maximisation



INGESTION DE DONNÉES – COLLECTE DE DONNÉES – ... TECHNIQUES

Approche sur mesure, visant à maximiser l'efficacité:

- Ingestion polie (pas de surcharge de portails)
- Extraire des attributs spécifiques lorsqu'ils sont disponibles (secteur, lieu,...)
- Monter en échelle en cas de besoin



PRÉTRAITEMENT DES DONNÉES – OBJECTIF...

Après la phase d'ingestion de données, une bonne quantité de pages Web (EOL) est collectée. Mais:

- Combien d'entre eux sont vraiment OELs?
 - Bruit (pages de profil social, pages de stratégie,...)
 - Faux postes vacants (formations, services de conseil,... publié sur les portails des OEL)
 - Combien d'entre eux se réfèrent à un OEL unique?



PRÉTRAITEMENT DES DONNÉES – ... STRATEGY...



Détection de
langage



Réduction
du bruit



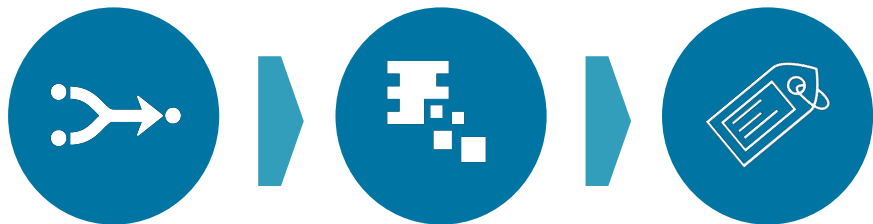
OEL
Déduplication



PRÉTRAITEMENT DES DONNÉES – ... TECHNIQUES

Plusieurs techniques d'intelligence artificielle sont appliquées pour détecter le bruit (tel filtre anti-spam du courrier) et pour détecter les OJV dupliqués.

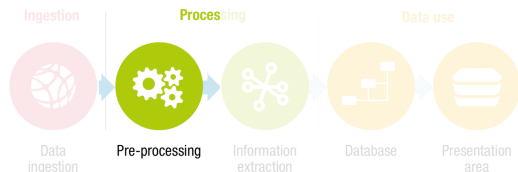
La détection des langues est la première étape, visant à se concentrer sur les pipelines spécifiques à la langue.



Fusion

Nettoyage

Traitement de texte
et résumer



CLASSIFICATION DES DONNÉES – OBJECTIF...

Après la phase de prétraitement des données, un montant pertinent de offres d'emploi en ligne (OEL) est extrait. Maintenant, il est temps d'extraire des informations:

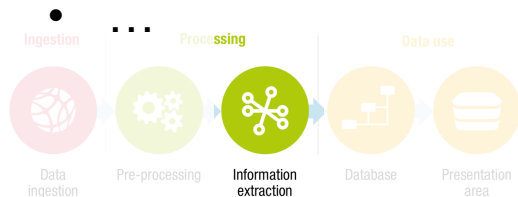
- Quelle profession est demandée (offre)?
- Il y a un endroit où le travail sera fait?
- L'emploi est-il axé sur un secteur spécifique?
- Quels types de compétences sont requises?



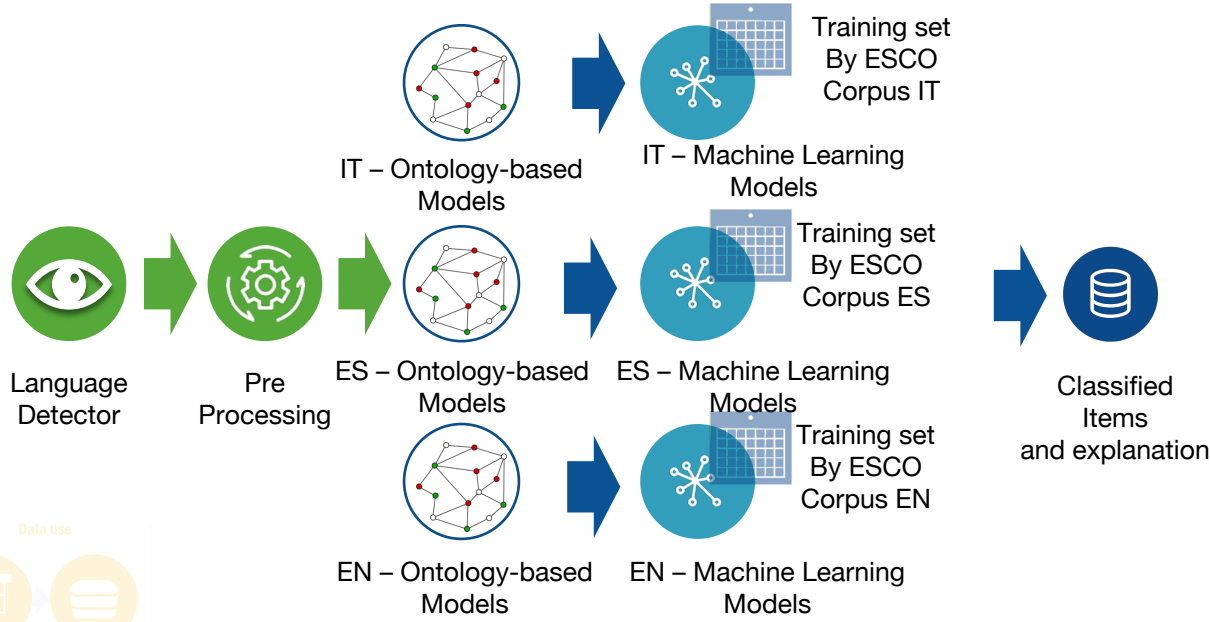
CLASSIFICATION DES DONNÉES – ... STRATÉGIE...

Adoption de taxonomies pour chaque variable analytique, pour définir un domaine de valeurs:

- Métier (ESCO)
- Compétences (ESCO)
- Endroit (NUTS / ISO3166)
- Secteur (NACE)
- Niveau d'éducation (ISCED)



CLASSIFICATION DES DONNÉES – ... TECHNIQUES

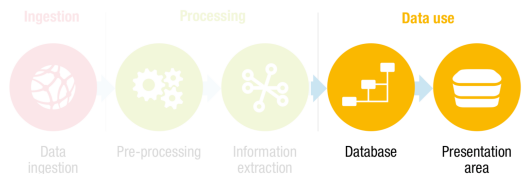


UTILISATION DES DONNÉES ET PRÉSENTATION – OBJECTIF...

À ce stade, nous disposons maintenant d'un ensemble de données complet, structuré et bien défini contenant des OEL uniques, nettoyés et classifiés.

Comment faire face à une telle quantité énorme de données?

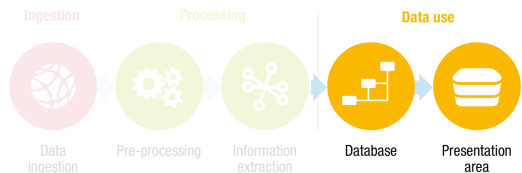
Nous devons fournir des données d'une manière qui garantisse à nos parties prenantes les informations les plus significatives disponibles



UTILISATION DES DONNÉES ET PRÉSENTATION – ... STRATÉGIE...

En plus de ces données, nous avons construit un tableau de **bord interactif** sur le Web, axé sur le marché du travail par pays, qui contient plusieurs voies analytiques à la disposition de toutes les parties prenantes.

Grâce à notre expérience professionnelle, ce tableau de bord est **entièrement conforme au projet EU Cedefop Big Data for Labor Market** (par exemple, plateforme OVATE de compétences) : la taxonomie utilisée et la méthodologie appliquée sont les mêmes, afin de s'assurer que les analyses soient comparables.



UTILISATION DES DONNÉES ET PRÉSENTATION – ... TECHNIQUES

Session de tableau de bord interactif à :

<https://public.tableau.com/profile/tabulaex#!/vizhome/ETF-BigDataLMI-Tunisia/Time>





Labour Market Intelligence - Tunisia

Select desired sources: (Tunisi)

Select Release Date: 01/04/2020 31/08/2020

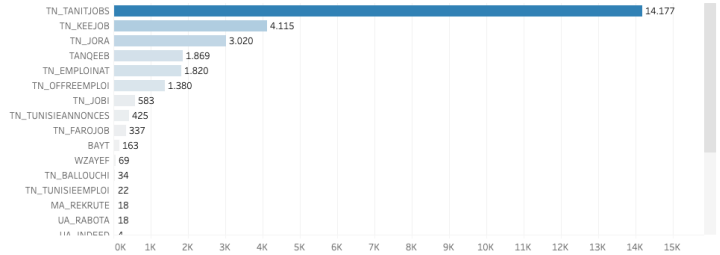
Number of job vacancies collected

138.985

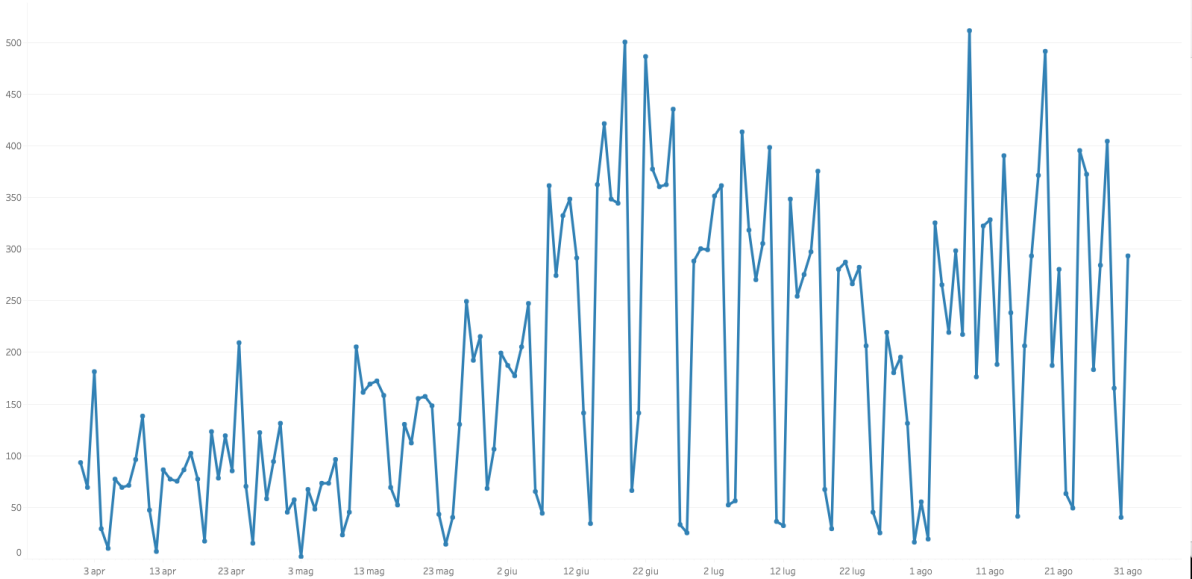
Number of job vacancies deduplicated

27.874

Number of unique Vacancies by Web Source



Distribution by Release Date (date of publication of the OJV)



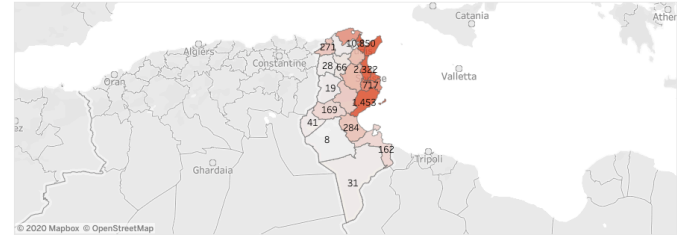
Select Release Date

01/04/2020

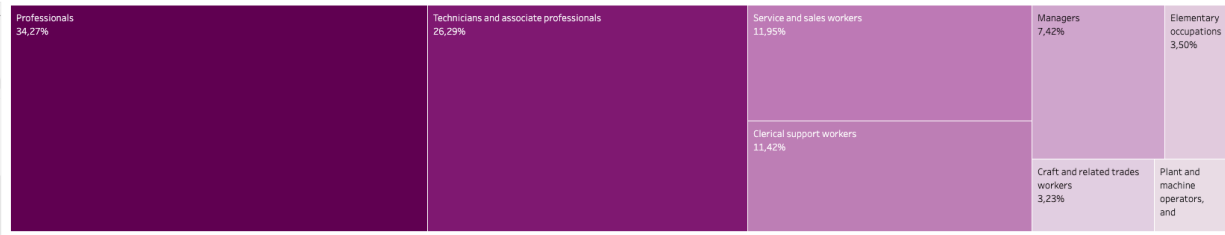
31/08/2020



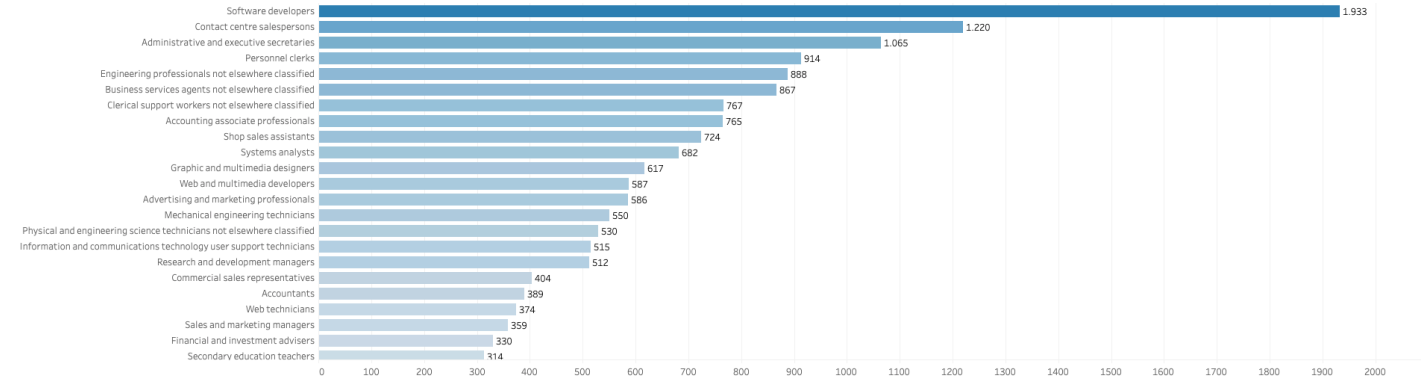
Location



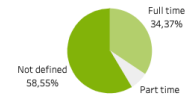
Occupation (level 1)



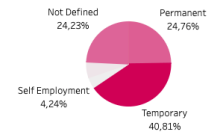
Select the Occupation that you want to Analyze



Working hours



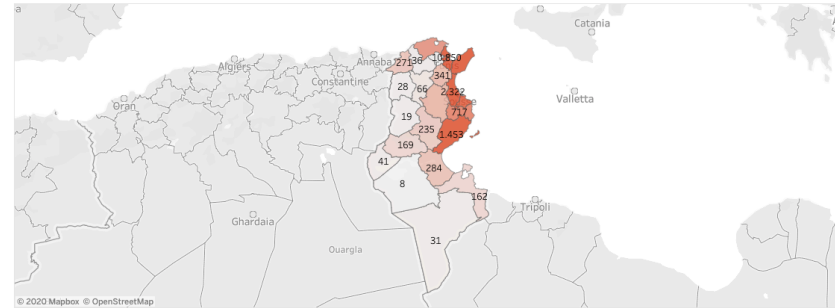
Contract





Select Release Date
01/04/2020

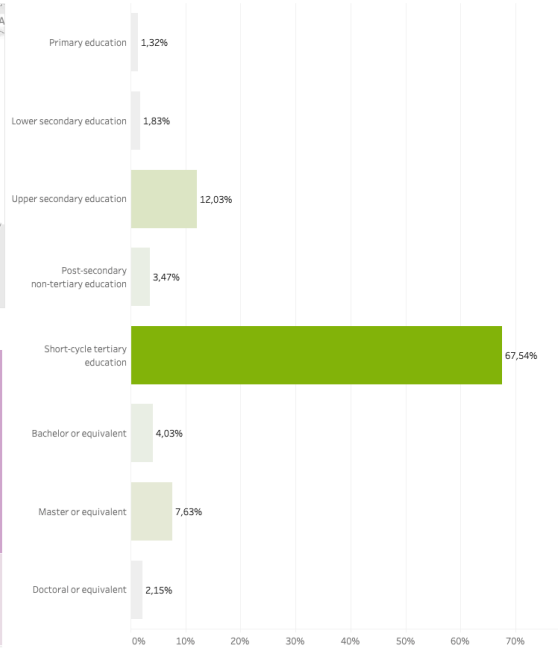
Location



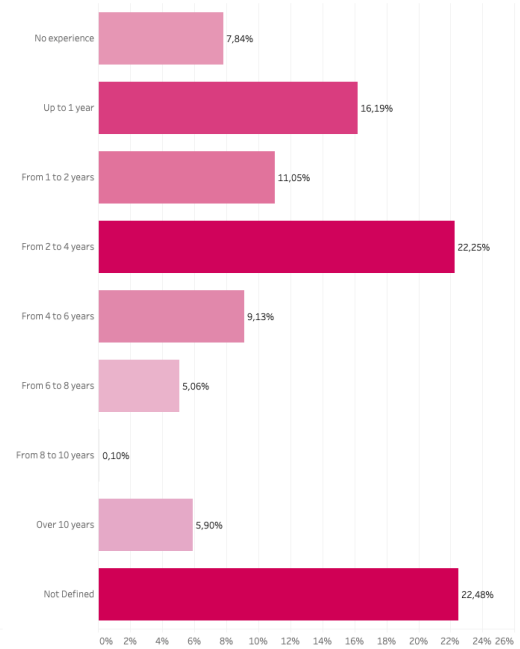
Occupation (level 1)



Education



Experience



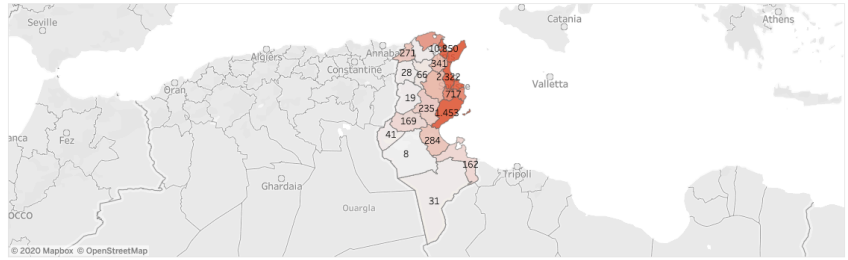


Select Release Date

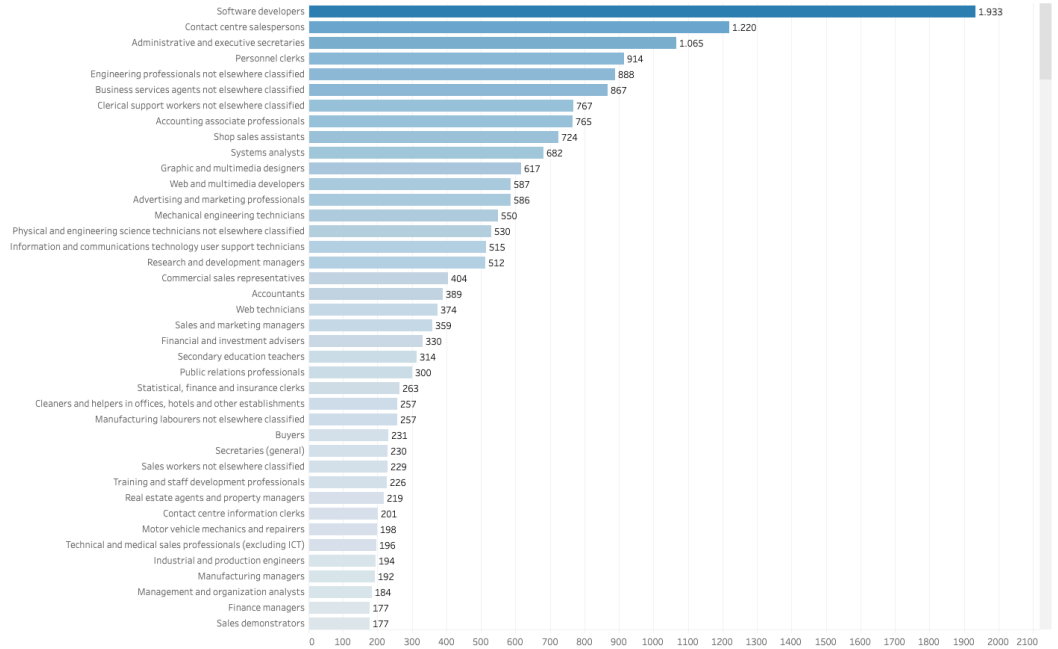
01/04/2020

31/08/2020

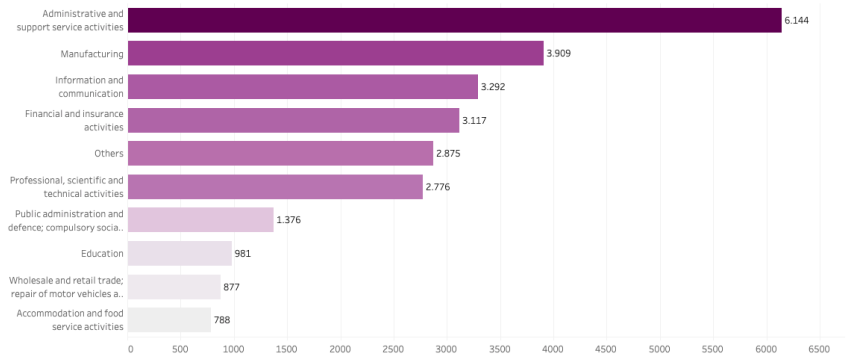
Location



Select the Occupation that you want to Analyze



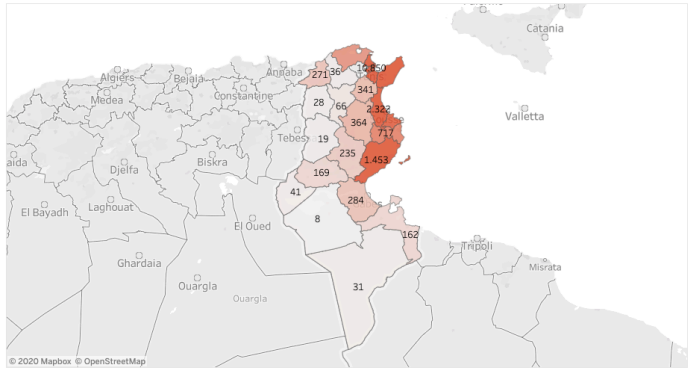
Industry (level 1)



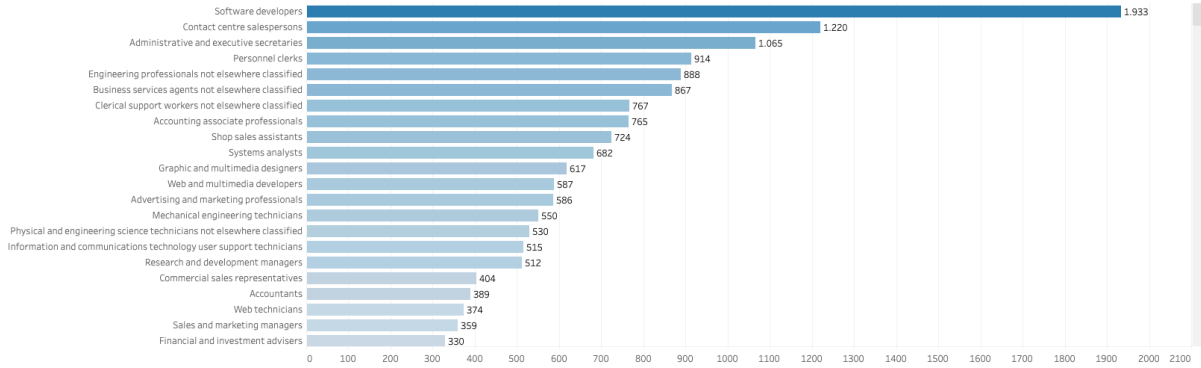


Select Release Date
01/04/2020

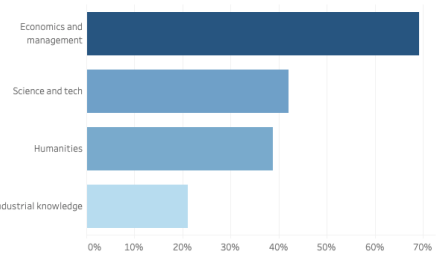
Location



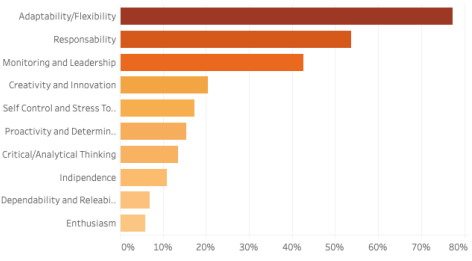
Select the Occupation that you want to Analyze



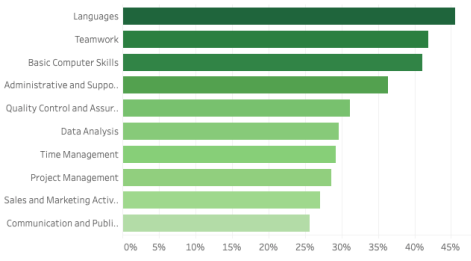
Knowledges



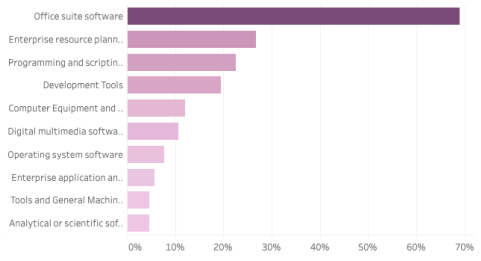
Personal qualities



Skills



Tools and technology



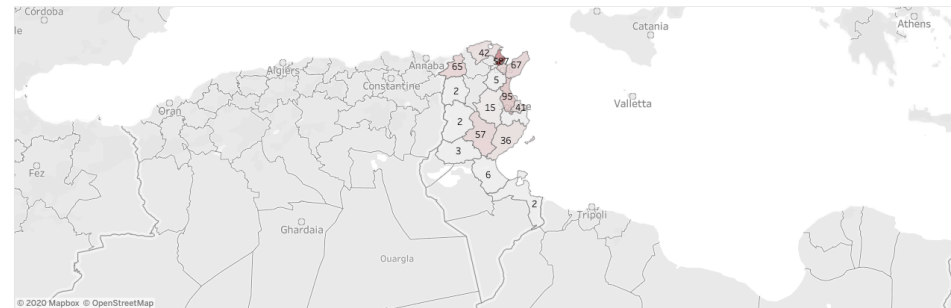


Select Release Date

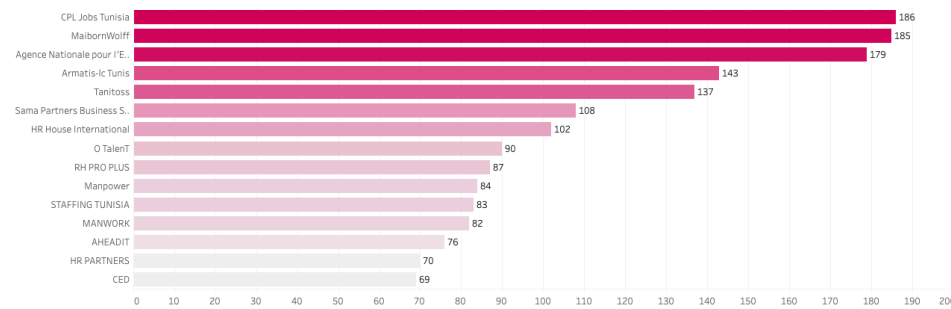
01/04/2020

31/08/2020

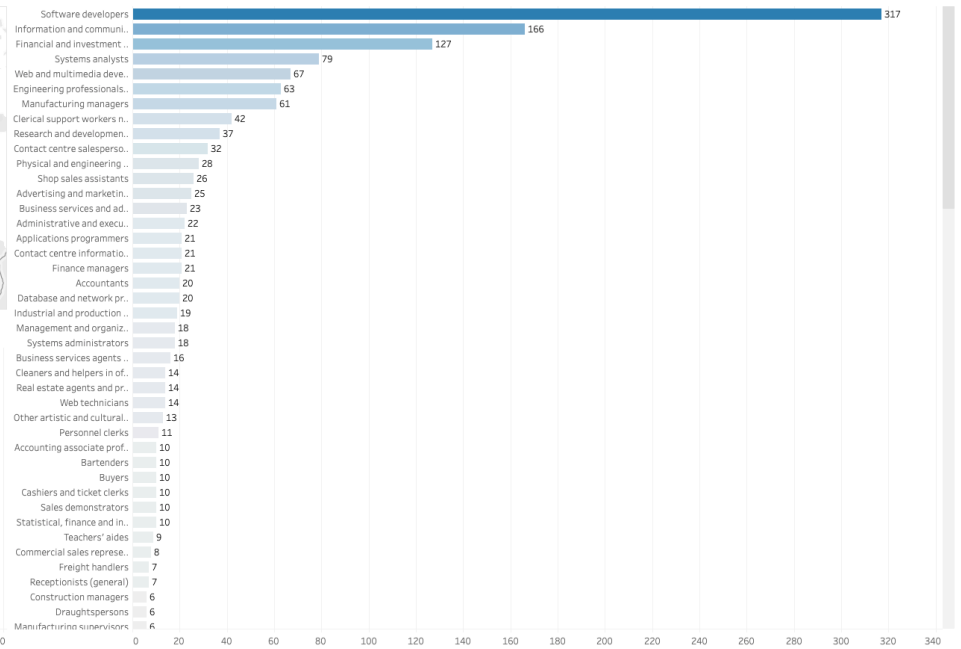
Location



Company Focus



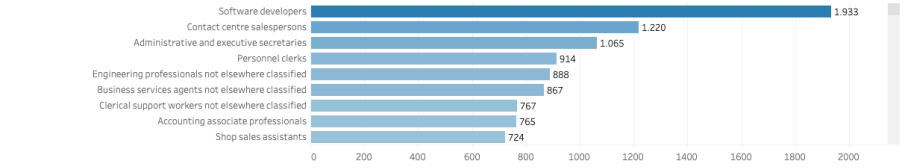
Occupation (level 4)



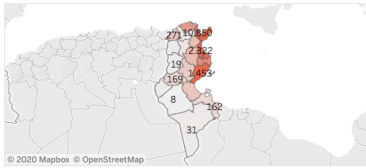


If you want, you can filter by occupation family:

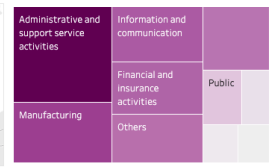
Select the Occupation that you want to Analyze



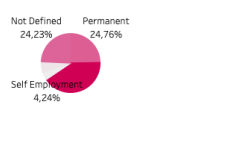
Location



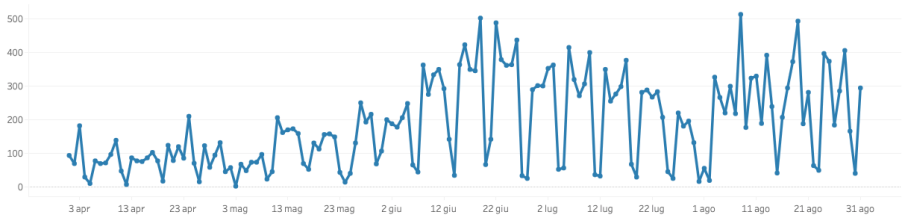
Industry (level 1 NACE)



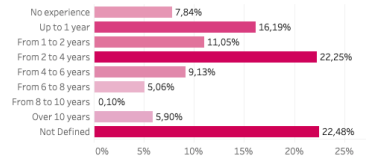
Contract



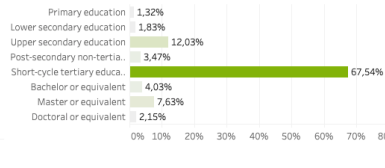
Distribution by Release Date (date of publication of the OJV)



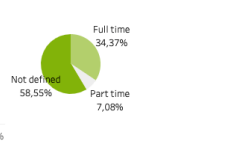
Experience



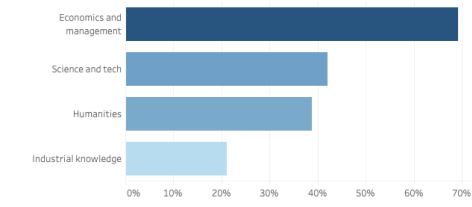
Education



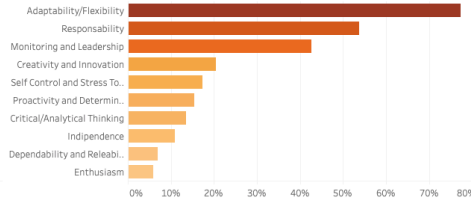
Working hours



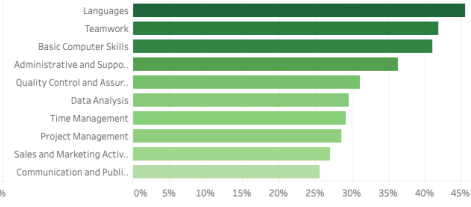
Knowledges



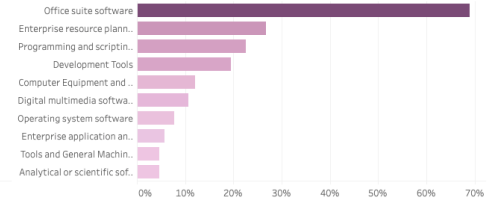
Personal qualities



Skills



Tools and technology



THANK YOU

