Table des matières

I. INTRODUCTION

Que pensent les jeunes tunisiens de l'émigration ?

Le but de ce projet est d'apporter quelques éléments de réponse à cette question.

Pour se faire une enquête a été réalisé au cours du quel plusieurs questions ont été posées. Cependant pour ce travail, nous n'allons traiter que les réponses à deux de ces questions :

Quelle est la différence au point de vue social entre la Tunisie et l'étranger ?

Quelle est la différence au point de vue professionnel entre la Tunisie et l'étranger ?

II. LES ETAPES DE L'ANALYSE

1. Mise en place des données

Les données recueillies ont été mise sous format csv avec 258lignes et 16 colonnes. Nos deux corpus sont uniquement les réponses aux questions 5 d'une part et 6 d'autre part. Nous avons traité chaque corpus dans un fichier à part. Nos corpus contenaient chacun 258 documents que nous allions prétraités

2. Retrait des caractères sans aucun intérêt pour l'interprétation

Une brève description des caractères contenus dans chaque corpus, nous n'avons pas remarqué de caractère étranger spécifiques sauf pour le corpus 6 qui contenait des caractères en arabe, mais nous ne les avons pas retirés à cette étape ; cette opération sera faite au niveau de l'identification de la langue. Par conséquent, il a juste été retirés, les caractères de ponctuation.

3. Identification de la langue

Pour chaque corpus, l'algorithme a identifié une multitude de langue mais en majorité le français. Une étude du cas par cas nous a révélé seulement 3 langages étrangers au français que sont l'anglais, l'arabe et le dialecte tunisien.

Alors pour le corpus5 nous avons écarté les documents écrits en anglais et selon le dialecte tunisien pour le corpus6 nous avons en plus écarté ceux écrit en arabe.

4. Tokenisation et Retrait des mots vide

A ce stade, on peut désormais identifier les mots ; et il s'en est suivi une élimination pure et simplement des mots qui n'étaient d'aucune aide à l'analyse et à l'interprétation. Nous avons recherché les mots de taille inférieur ou égale à 2 que nous avons retiré du corpus. Après cela, la taille moyenne des mots était de 8 caractères.

5. Normalisation

Pour le stemming, nous avons préféré garder les mots tel qu'écris sans apporter de modification. Nous avons testé es méthodes de stemming mais que nous n'avons pas utilisé en fin de compte.

6. Représentation des données

Afin de pouvoir extraire des informations, il fallait une représentation numérique du corpus. Ce qui a été fait en utilisant la méthode Bag Of Word

7. Détection des sujets

À cette étape, on analyse les features obtenus à l'étape précédente afin de dégager des topiques. L'objectif étant de déterminer de quelle tendance sont les réponses apportées aux différentes questions.

8. Construction d'un modèle de prédiction

Nous tentons ici de trouver un modèle, après avoir déterminer les termes autour du quelle gravitent chaque réponse, nous cherchons ici les profile de personne (sexe, âge, statut matrimoniale, région, établissement d'étude, domaine d'étude, etc.) qui correspondent à chaque sujet traité afin de pouvoir prédire les réponses à ces questions connaissant uniquement ces informations.

III. DISCUSSION DES CHOIX ET DES RESULTAS DE CHAQUE METHODES EMPLOYEE

Nous allons maintenant expliquer pourquoi nous avons choisi certaine méthode plutôt que d'autres

1. Identification de la langue

Le premier choix que nous avons eu à faire était d'écarter certains langages l'anglais, le dialecte tunisien l'arabe au lieu de tenter une traduction.

Pour le dialecte tunisien et l'arabe, nous les avons retrouvés dans chacun un document et également par manque de vocabulaire adéquat, nous avons préféré les éliminés du corpus.

2. Retrait des stopwords

En plus des stopwords téléchargés, nous avons en plus visualisés les mots contenus dans le corpus, détecté des mots, que nous avons manuellement rajouté à la liste de stopwords

3. Normalisation

Les méthodes de stemming étudiées, FrenchStemmer de nltk et celle proposée par le Dr Chiraz réduisent la taille des mots selon leur propre logique afin d'obtenir un mot représentant le plus possible à la racine du mot à réduire. Après l'application de ces méthodes, nous avons remarqué que à plusieurs reprises après avoir appliqué ces méthodes, plusieurs mots perdaient leur sens, et ne ressemblaient plus au mot initial, ce qui rendrait difficile l'extraction d'information de ces termes qui n'avaient pas de sens apparent,

ce qui nous a conduit à garder les termes tel quel, compréhensibles et en moyenne pas trop long.

4. La méthode tfidfvector

Pour l'obtention des futures et de vocabulaire nous avions le choix entre 2 méthodes (tous deux en utilisant le modèle BOW).

Nous avons choisi de travailler à chaque fois avec un vocabulaire de 80 mots. Les deux méthodes nous ont sorti le même vocabulaire cependant, la méthode tfidfvector prenant en compte la fréquence et la rareté du mot dans le corpus représente mieux l'importance des mots.

5. Modèle de prédiction

Nous avons choisi d'utiliser la régression logistique afin d'expliquer les réponses aux questions en fonction des variables dites 'profiles'.

Pour entrainer le modèle, nous avons utilisé dans un premier temps, un data train de 150 individus et un data test de 55 individus pour le corpus6 ce qui nous as donné après le test, une précision de 25% (très faible) et pour le corpus5 nous avons utilisé la totalité des données disponibles, ce qui nous a données une précision de 45% sur un data test de 50individu tiré du data train (ce qui montre que le modèle n'est pas du tout adéquat). Cependant, nous n'avons pas réussi à tester d'autre modèle comme les random Forest.

IV. RESULTATS DES ANALYSES

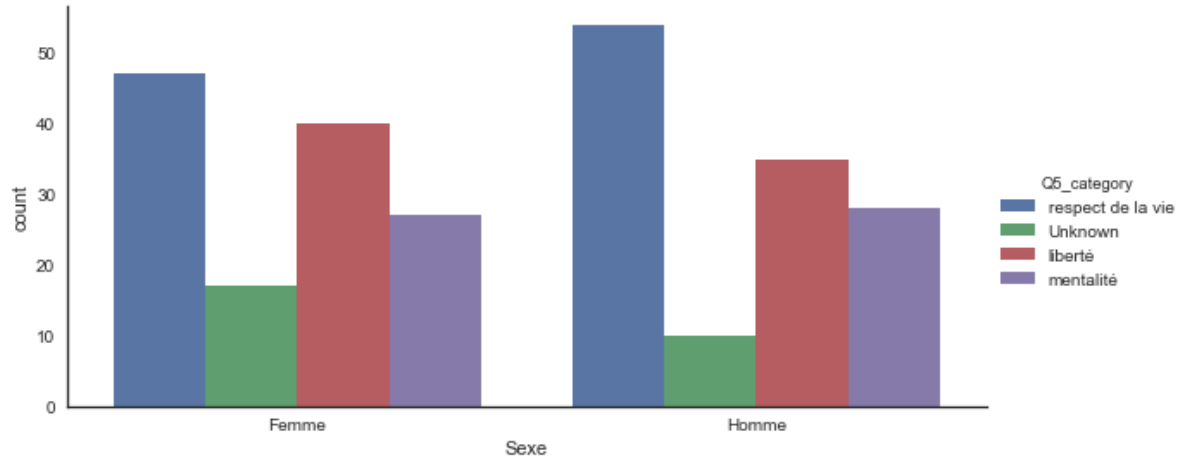
1. Pour la question 5

Quelle est la différence au point de vue social entre la Tunisie et l'étranger ?

A cette question, nous avons relevé que les réponses tournaient autour de 3sujets : le respect de la vie, la mentalité et la liberté.

Comme le montre les graphiques ci-dessous, la plupart des répondants pensent que à l'étranger le respect des droits de l'homme est plus respecté qu'en Tunisie. Ces graphiques montrent la distribution des réponses apporté à cette question en selon certaines caractéristiques sociales.

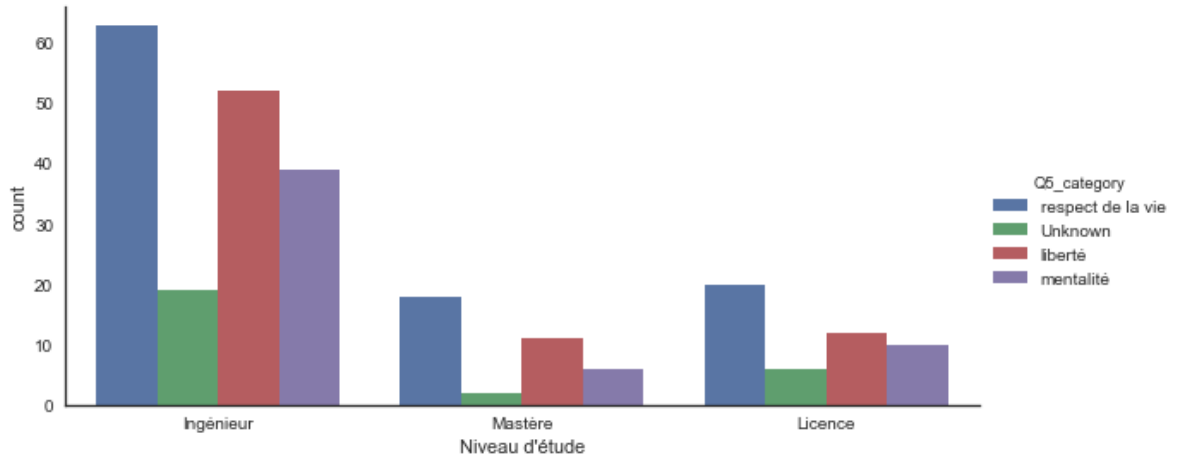
Les réponses selon le sexe des répondants



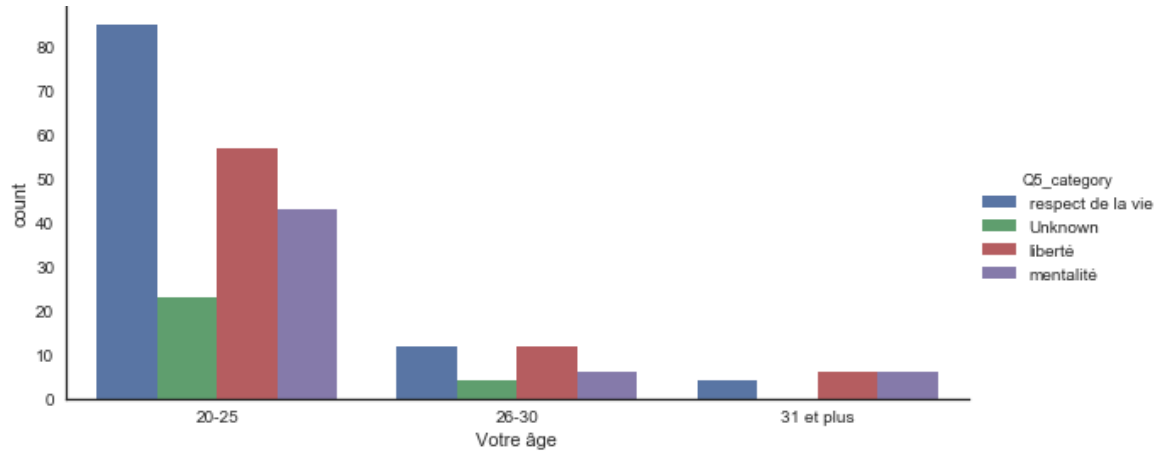
Au vu de ce graphique, en plus du respect de la personne humaine, les hommes et encore plus les femmes affirment qu'il y'a plus de liberté à l'étranger.

En effet, quand on n'a pas encore quitté son pays, on pense toujours qu'ailleurs est mieux.

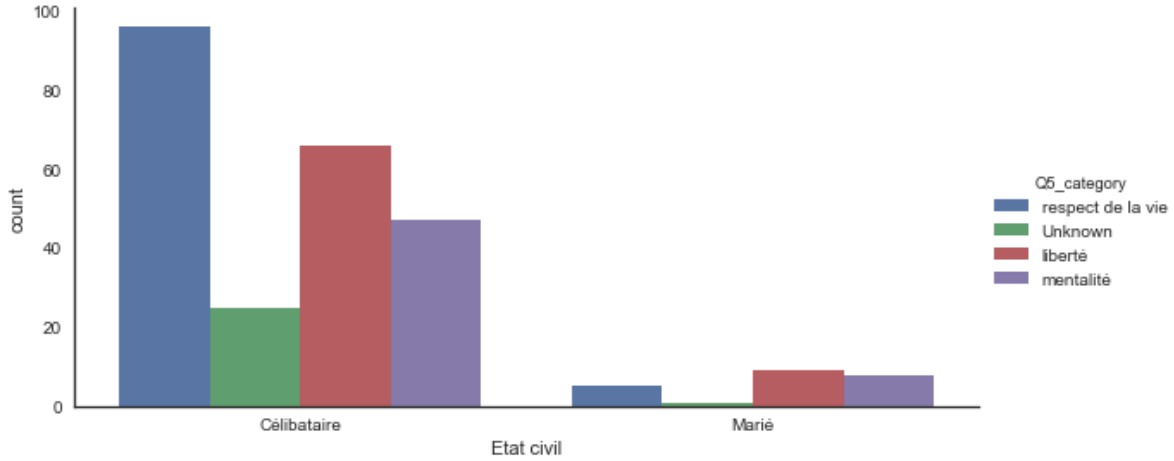
Les réponses selon le niveau d'étude des répondants



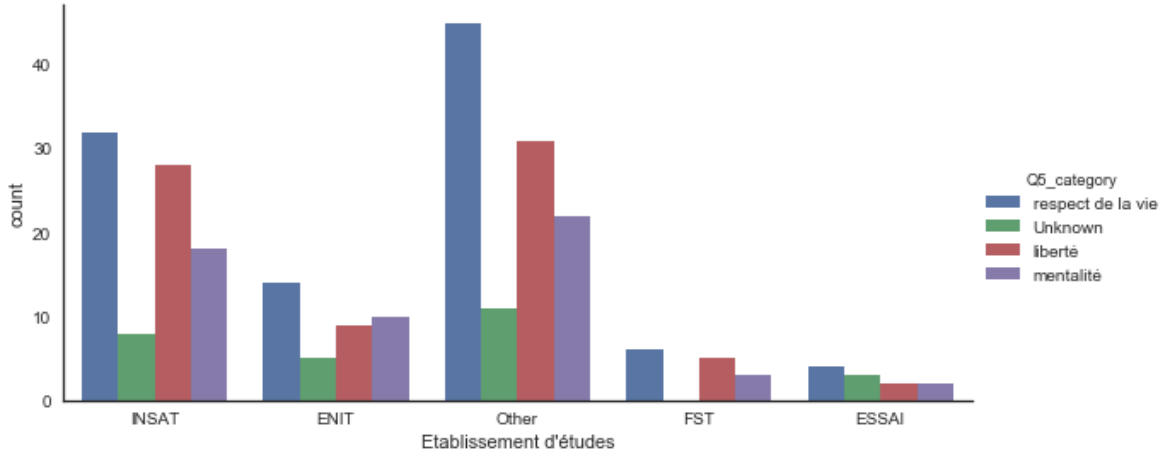
Les réponses selon l'âge des répondants



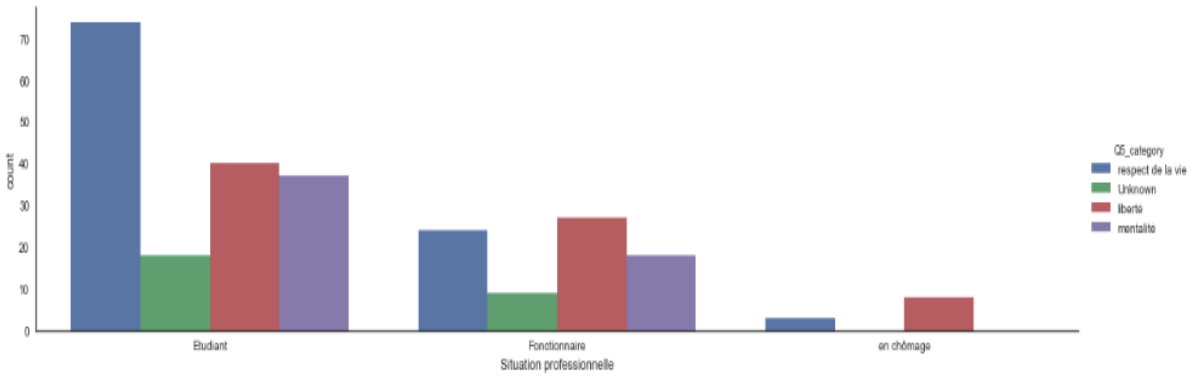
Les réponses selon l'état civil des répondants



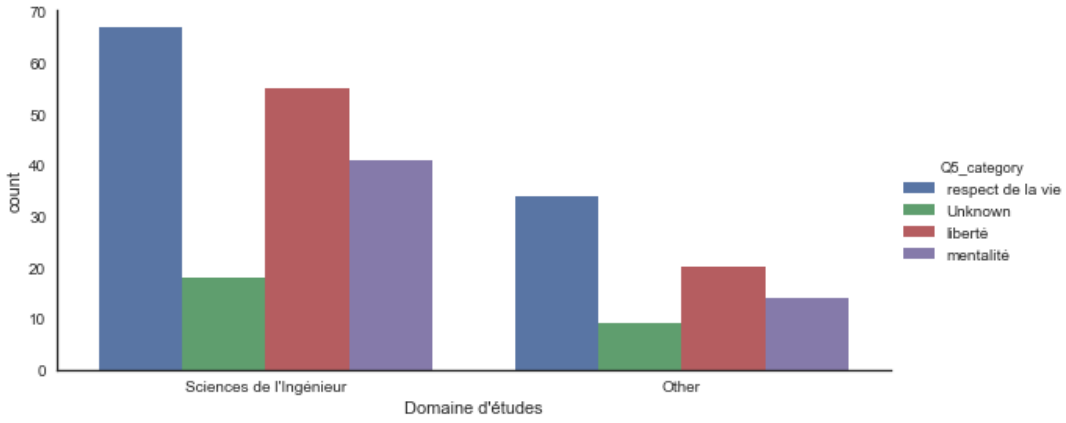
Les réponses selon l'établissement des répondants



Les réponses selon la situation professionnelle des répondants



Les réponses selon le domaine d'étude des répondants

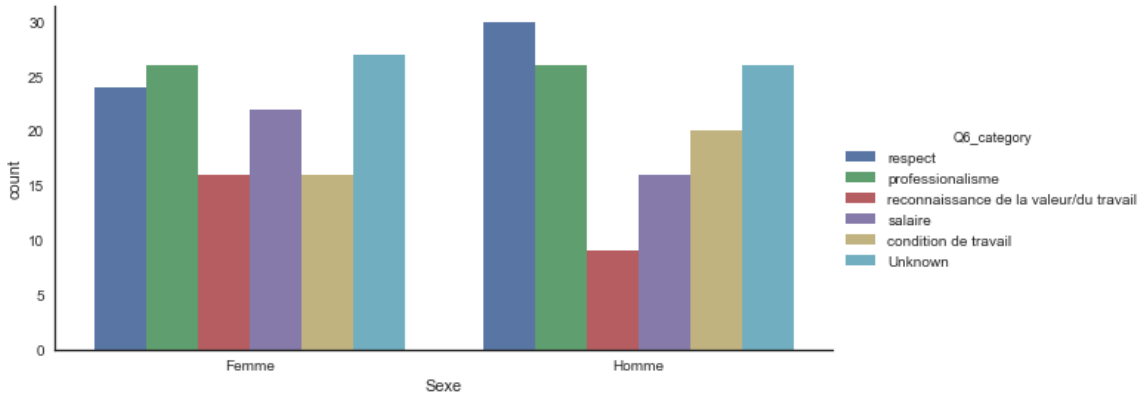


2. Pour la question 6

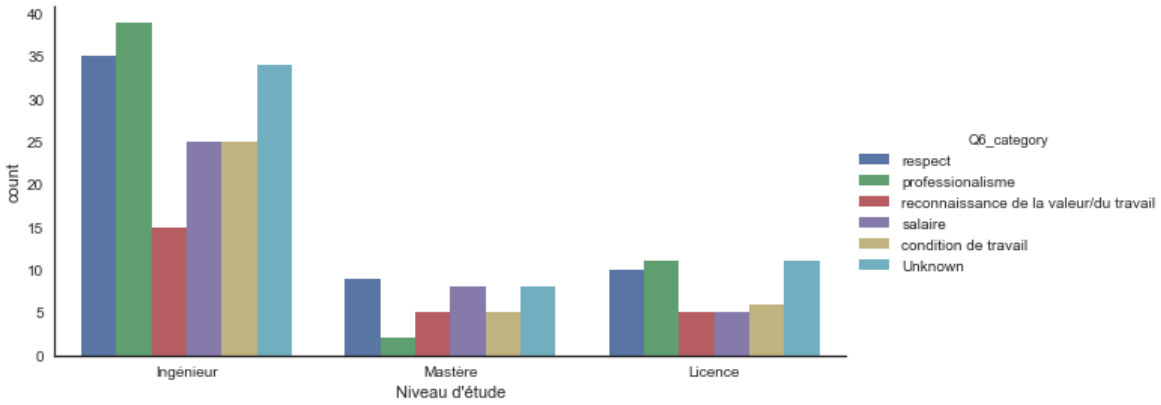
Quelle est la différence au point de vue professionnel entre la Tunisie et l'étranger ?

A cette question, les réponses étaient partagées entre : la condition de travail, la reconnaissance de la valeur du travail, le salaire, le professionnalisme, le respect

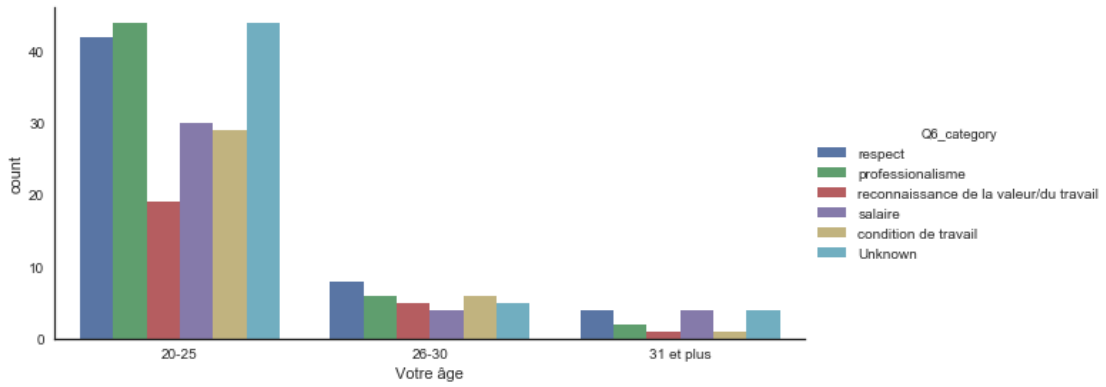
Pour nos répondants, la Tunisie diffère de l'étranger sur différents points :



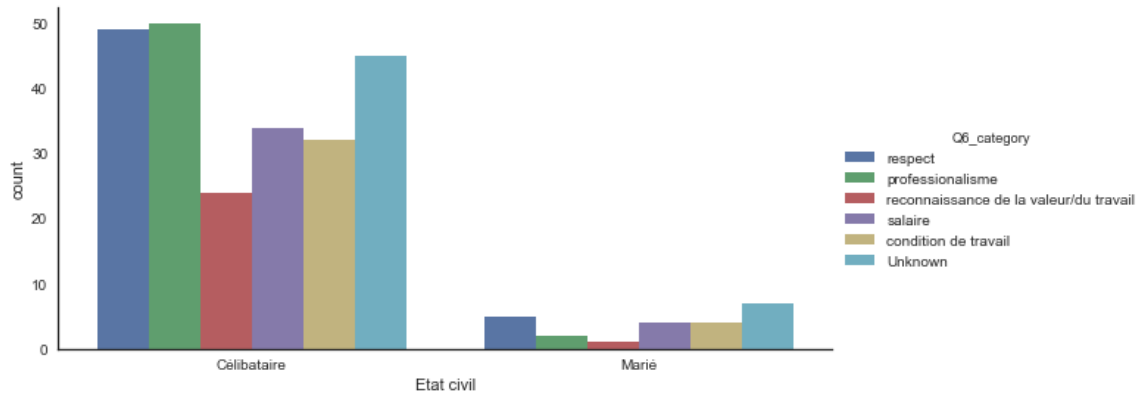
Les hommes et les femmes sont autant d'accord l'un que l'autre sur le fait que le niveau de respect et de professionnalisme dans le travail est bien meilleur à l'étranger.



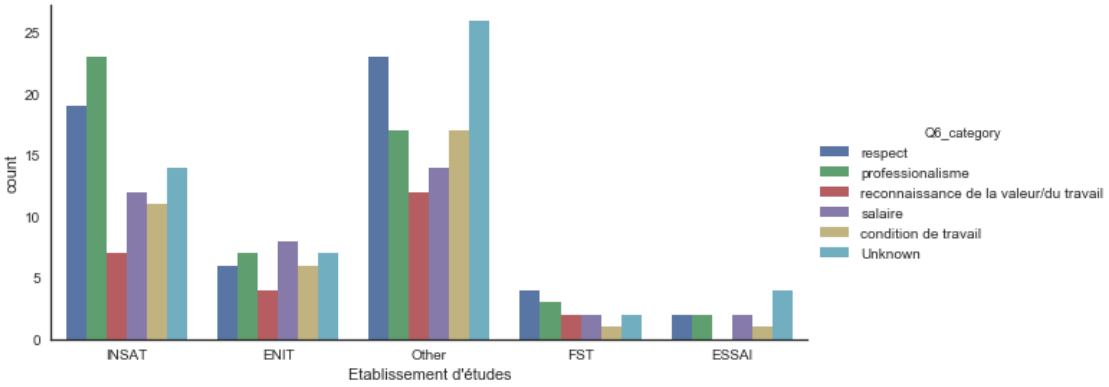
Pendant que les ingénieurs et les étudiants en licence se penchent plus sur le professionnalisme les étudiant en mastère par contre sont préoccupés par le respect et le niveau de salaire qui est nettement plus élevé en France par exemple.



Les personnes les plus âgés (+26ns), accorde plus d'importance au respect et a la condition de travail et bien entendu le salaire pendant que pour les plus jeunes entre 20 et25ans ils sont plus préoccupés par le niveau de professionnalisme

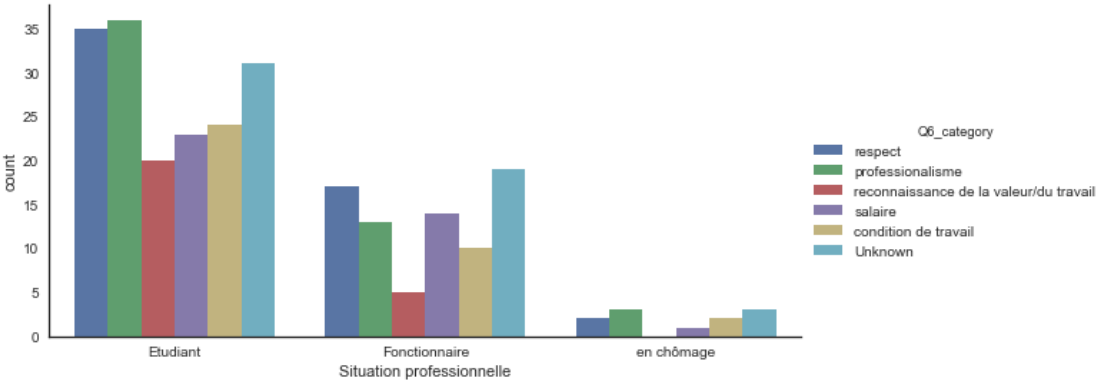


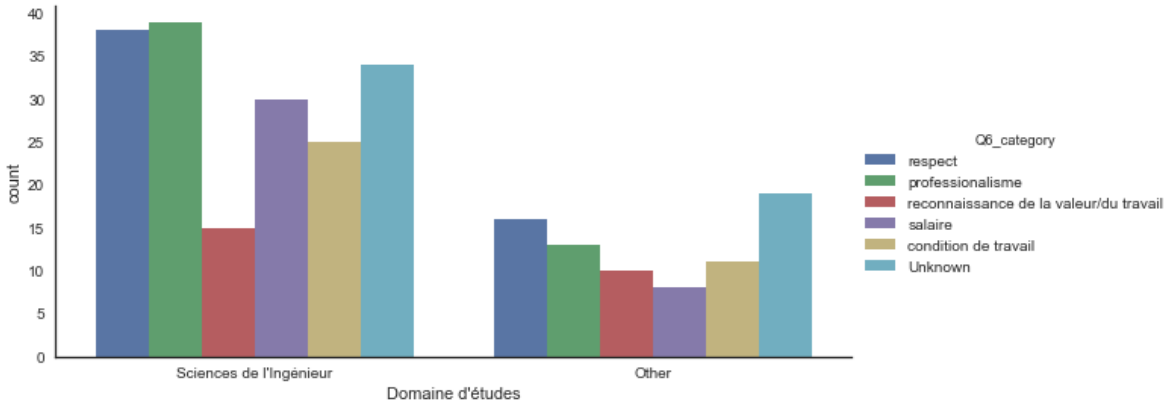
Le même constat est partagé entre célibataires et mariés



Ce qui dérange le plus les ingénieurs de l'enit, c'est le niveau de salaire contrairement à l'autres qui sont entre le respect le professionnalisme même si tous s'inquiètent autant pour les conditions de travail.

On pourrait croire que les ingénieurs de l'enit pensent qu'ils ne sont pas assez bien payés





V. CONCLUSION

Nous avons tenté d'analyser les réponses à deux questions :

Quelle est la différence au point de vue social entre la Tunisie et l'étranger ?

Et

Quelle est la différence au point de vue professionnel entre la Tunisie et l'étranger ?

Ce qui est certains, c'est que pour nos répondants, à l'étranger on est mieux traité tant sur le plan social que professionnel.