



[ANALYSE DES OFFRES DE STAGE DU SITE DE L'APEC]

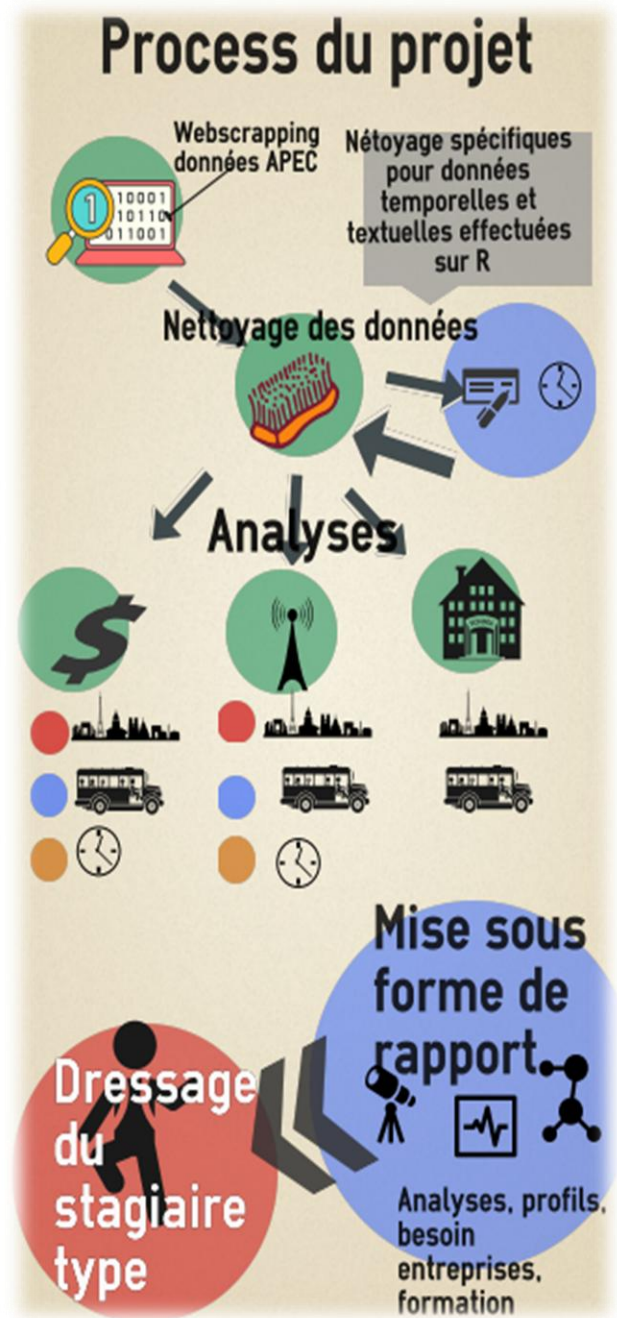
A la recherche d'un stage en entreprise nous avons fait le choix d'utiliser nos outils statistiques, pour connaître l'attente des entreprises concernant les étudiants. Nous avons donc aspiré toutes les offres de stages du site de l'APEC, 5e sur le classement des sites d'emplois en termes de visiteurs. Purement privé et financé par les cotisations des cadres et des entreprises la mission de l'APEC (association pour l'emploi des cadres) est de conseiller ceux-ci ainsi que de faire le lien avec les jeunes issus de l'enseignement supérieur. Cette analyse nous permettra de faire un état des lieux de la demande des entreprises, concernant les nouveaux cadres de ces dernières années. Nous croisons aussi dans ce papier ces données avec celles de l'INSEE, des données afin de resituer géographiquement ces chiffres par région en termes de formation. Ces données très riches et très disparates après un long raffinement nous permettront de répondre à de nombreuses questions que l'on se pose en tant que jeune en formation. Dans quelles régions se trouvent les gros foyers de stage (et donc d'embauche pour le futur) ? A quel moment les entreprises commencent-elles à chercher des stagiaires ? Quels sont les formations qui offrent des stages bien rémunérés (et donc les plus prometteuses en terme économique) ? Tant de questions auxquelles nous essayons de répondre avec comme support le logiciel d'analyse statistique SAS.

Table des matières

Présentation du site de l'APEC	3
A. Les données issues de l'APEC	4
A. Des données issues du site de l'APEC.....	7
1. Analyse portant sur les données temporelles.....	7
a. Analyse chronologique du nombre de publication.....	7
b. Analyse du taux d'évolution des publications	11
c. Analyse sur l'intervalle calculé entre la date de publication et la date de début annoncé...	13
d. A quelle date commence en général les stages proposés ?	14
2. Analyse des gratifications.....	15
3. Analyse Géographique	16
4. La durée du stage.....	18
5. Analyse du niveau d'étude requis.....	19
6. Analyse des champs textuels	20
a. Le titre de l'annonce.....	20
b. Le nom de la société.....	21
c. Poste et mission	21
d. Profil recherché	22
7. Un mot sur les stages à l'étranger.....	23
B. Analyse des données provenant d'autres sources	23
1. Nombre d'inscription par ville.....	23
2. Nombre d'inscription par cursus universitaire.....	25
Conclusion.....	26

I. Introduction

Curieux data-scientistes en devenir, nous sommes étions à l'époque à la recherche d'un stage pour cette fin d'année, et avons fait le choix d'user des dons qu'il nous été permis de développer au fil de nos cursus pour y voir plus clair sur la situation. Pour trouver un stage il y a plusieurs options soit profiter de d'un réseau (faculté, connaissances, amis, famille), soit une rencontre fortuite (ça n'arrive pas tous les jours), il est aussi de possible de faire du porte à porte (quand on n'a pas froid aux oreilles) ... Bien sur celle que nous privilégions tous c'est internet, véritable sources d'annonces professionnelles. Nous avons choisi de nous pencher sur le cas de l'APEC, dont nous avons webscrappé le contenu de toutes les annonces publiés entre 2012 et 2014, échantillon de près de 17 000 annonces exploitables sur 88 000 récupérés à brut. Nous procédons donc à de longs nettoyages avec une certaine perte d'information. Cependant nous avons fait le choix de la qualité et avons pu recueillir 15 variables décrivant les stages proposés dont le salaire, le lieu du stage, la date de publication, le niveau d'étude requis, ainsi que des variables textuelles comme le profil recherché. Une véritable mine d'informations, que nous avons pu croiser avec les données issue de l'INSEE (pour faire des analogies entre bassins de stages), ainsi que celles issues par l'état au sujet des effectifs des établissements de formations supérieurs en France.



Nous développons donc dans ce rapport les résultats issus de cette analyse de-vous conter l'histoire de nos données nous y aborderons aussi les différents procédures SAS utilisé, avec en annexes nos codes et nos rapports SAS.

Let's Talk Data...

Présentation du site de l'APEC

Prenez rendez-vous avec l'avenir.

Accueil Les Services Apec **Mes offres** Mes outils Premier job Observatoire de l'emploi Communauté Réseau Pro Mon compte

Simulateur d'entretien de recrutement
» Essayez-le maintenant

500 offres de stage correspondent à vos critères

Saisissez vos mots clés, fonction, secteur : Rhône-Alpes » Plus de critères

Mes mots-clés
Ajouter un mot clé +

Le flux rss de ces offres

Classer par : Pertinence Affichage : ☐ ☐

Fonction
Achat (7)
Chantier, travaux (3)
Commercial, export (78)
Communication, journalisme (19)
Création, design (12)
[Voir plus de choix](#)

Lieu
Rhône-Alpes (500)

Secteur d'activité
Audit-conseil (29)
Banque et assurance (10)
Commerce, E-commerce (52)

Développeur .net H/F
Durée : >6 mois | Début : 01/09/2015
Lyon
30/04/2015
Teamsquare
Origine : Apec
Encadré par le responsable informatique et au sein de notre équipe de développeur, vous aurez comme mission principale de développer des applications logicielles ou progiciels répondant aux fonctionnalités attendues par les utilisateurs. A ce ...

COMMERCIAL (H/F) STAGE DE 3 MOIS H/F
Durée : 3 à 6 mois | Début : 15/06/2015
lyon
30/04/2015
Sodexo
Origine : Apec
COMMERCIAL (H/F) STAGE DE 3 MOIS Sodexo Services Avantages et Récompenses Nous recherchons un(e) Commercial(e) pour un stage de 3 mois à Lyon Missions Au sein de la Direction des Affiliés, vous êtes rattaché(e) à la Directrice Partenariats ...

H/F Alternance Chef de projet
Durée : >6 mois | Début : 01/09/2015
30/04/2015
Mobilboard Annecy

L'association pour l'emploi des cadres, appelé aussi l'APEC est une association privée, financé par des cadres et des entreprises dont l'objectif est le conseil et le service sur des sujets touchant à l'emploi. Son site est l'un des sites d'emploi les plus visités en France, le choix de cette plateforme n'est pas anodin, projetant de poursuivre notre formation jusqu'à bac +5, nous sommes destinés à devenir cadre en informatique. Nous en profiterons pour vérifier à l'issue de ce rapport si nous avons nos chances face au marché du stage.

Site	Visiteurs Uniques / mois	
1	Pôle emploi	5 576 000
2	Leboncoin.fr Emploi	2 087 000
3	Indeed	1 708 000
4	Meteojob	1 046 000
5	APEC	1 032 000
6	Monster	966 000
7	Cadremploi	918 000
8	Keljob	896 000

Source: Mediametrie//NetRatings

II. Présentation des données

A. Les données issues de l'APEC

Aspect d'une annonce, et mise en valeurs des champs extraits

Data Scientist H/F

[Retourner aux offres](#)

Publiée il y a 18 jours

Référence Apec : 207601
Référence société : DS/LR
Société : AWABOT
Lieu : Lyon
Niveau d'étude : BAC+4/5

Date de publication : 14/04/2015
Durée du stage : 3 à 6 mois
Date de début : dès que possible
Gratification : 508.20€ net/mois

[Postuler à cette offre](#)

[Partager sur un réseau social](#)
[Imprimer](#)

[Envoyer par mail](#)
[Signaler cette offre](#)

L'entreprise

Créée en 2011, AWABot offre une approche innovante dans le développement robotique de service. AWABot s'intéresse plus spécifiquement aux interfaces et aux nouveaux rapports homme-machine et robots. Elle est particulièrement active dans les domaines de la téléprésence, dans l'éducation, la santé et l'assistance à la personne. La société distribue notamment une gamme de robot communicants autonomes ou pilotés à distance. AWABot emploie quinze personnes et collabore avec une vingtaine de partenaires industriels et académiques. www.awabot.com

Poste et mission

Dans le cadre d'un projet, nous visons à améliorer l'échelle de douleur actuellement utilisée dans les cliniques/hôpitaux (échelle papier). ~~À travers un ensemble d'applications sur tablettes, le patient doit lui-même évaluer sa douleur.~~ Nous proposons donc de construire une échelle de douleur numérique qui serait plus précise que les échelles actuelles et qui permettraient de mieux doser les médicaments donnés aux patients.

Les missions seront les suivantes :

- Extraction et mise en forme des données brutes extraites des tablettes.
- Analyse statistique et construction des caractéristiques pertinentes.
- Développement d'algorithmes permettant de mettre en avant les caractéristiques pertinentes.
- Intégration des algorithmes dans les applications.

Profil

- En dernière année d'école d'ingénieur ou équivalent universitaire avec une maîtrise des techniques inhérentes au ~~data mining~~ et à l'ingénierie statistique.
- Qualités d'analyse, de rigueur, de synthèse et d'autonomie.
- Connaissance de la programmation en JAVA.

[Postuler à cette offre](#)

[Partager sur un réseau social](#)
[Imprimer](#)

[Envoyer par mail](#)
[Signaler cette offre](#)



Afin de récupérer dans un premier temps nos données nous extrayons un intervalle [ID ; ID+n] d'annonces (de 111111 à 200000 dans notre cas) en récupérant pour chaque annonce les données sous format txt.

CF. ligne 11-27 Code Programme_APEC.sas

Cette première procédure a duré près de 9h (près de 88 800 annonces extraites) on est vraiment dans du forage de données.

Une fois que les données ont été extraites au format txt. Ces exports correspondent aux pages au format html. L'on va donc pouvoir utiliser les conventions de nommages propres aux formatages des annonces. Nous repérons par exemple <h1> <labels> et <h3> précèdent les informations que nous voulons récupérer, nous utilisons ensuite des marqueurs indiquant les tuples de nos tables étant précédés par ces conventions. Nous nettoyons au passage plusieurs fois certains caractères vides, caractères pouvant poser des problèmes de transformation d'encoding. Une fois extraites nous assemblons toutes ces données en plusieurs fois afin d'obtenir permettant par id, de décrire l'annoncé selon 15 variables. Cette phase de distillation et transformation de données pourraient être améliorée : en marquant la description entière des champs Profil, Présentation entreprise et Poste Mission d'une part et d'une autre en réussissant à mieux formater les données jusqu'à la fusion de celles-ci par ligne (nous rappelons que n'obtenons que 20% d'annonces bien formatés et utilisable pour nos analyses. Ce nettoyage durant près de 24h (en effet nous manipulons près de 88 000 fichiers créons près de 700 000 tables SAS de 88 000 lignes), nous avons dû profiter de la gamme superbe gamme d'ordinateur disponible au département informatique, pour réaliser ce traitement sur 6pc.



CF. ligne 132 à 310 Code Programme_APEC.sas et fichier CF. FINAL_APEC.csv

->L'on obtient alors un fichier csv décrivant 17500 annonces entre 2011 et 2015, sur les quel nous accolons le nom de la région via filtres successifs sur Excel en indexant les villes apparaissant dans notre liste INSEE par, faute de données suffisamment formatées (plusieurs orthographes, fautes...) afin de pouvoir par la suite évaluer nos données par bassins de recrutements, véritable travail d'orfèvre, nous en avons profité pour accolés à ces codes régionaux les coordonnées GPS des chef-lieu de chaque région, afin de pouvoir représenter nos données sur des fonds de cartes.

Nous utiliserons ce fichier ensuite pour traiter nos données textuelles, temporelles à part, et nous analysons nos données en les reformatant au format SAS puis nous supprimons les salaires nul (considérés comme de base peu intéressants), et les salaires extravagants supérieur en outre à 2600€,

Nous réalisons de plus via le package tm de R les quatre DocumentsTermMatrix se référant aux champs textuels, ces matrices décrivent chaque annonces selon les mots les plus caractéristiques de ces champs. Nous passons aussi par de gros nettoyages sur ces données textuelles, nous devons réussir à ne garder que quelques mots représentatifs des champs textuels entrés par les entreprises. Nous passons par une phase de racisation de chaque mots, supprimons espaces et caractère spéciaux y supprimons la liste des 100 mots les plus courant de la langue française, enfin nous ne prenons que les mots dont les mots apparaissent 98% des cas afin de ne pas avoir à traiter trop d'information d'un seul coup.



III. Analyses

A. Des données issues du site de l'APEC

1. Analyse portant sur les données temporelles

Dans un premier temps afin d'analyser « la demande » en stage nous avons besoin de situer celle-ci temporellement. Cette analyse est détaillée selon plusieurs niveaux de finesse : jour/mois/trimestre/année et nous permettra de déterminer les mois durant lesquels les entreprises commencent à chercher des stagiaires, nous abordons ensuite quelques analyse concernant l'évolution du nombre d'annonces publiées au fil des années, enfin nous analyserons les dates de début de stage.

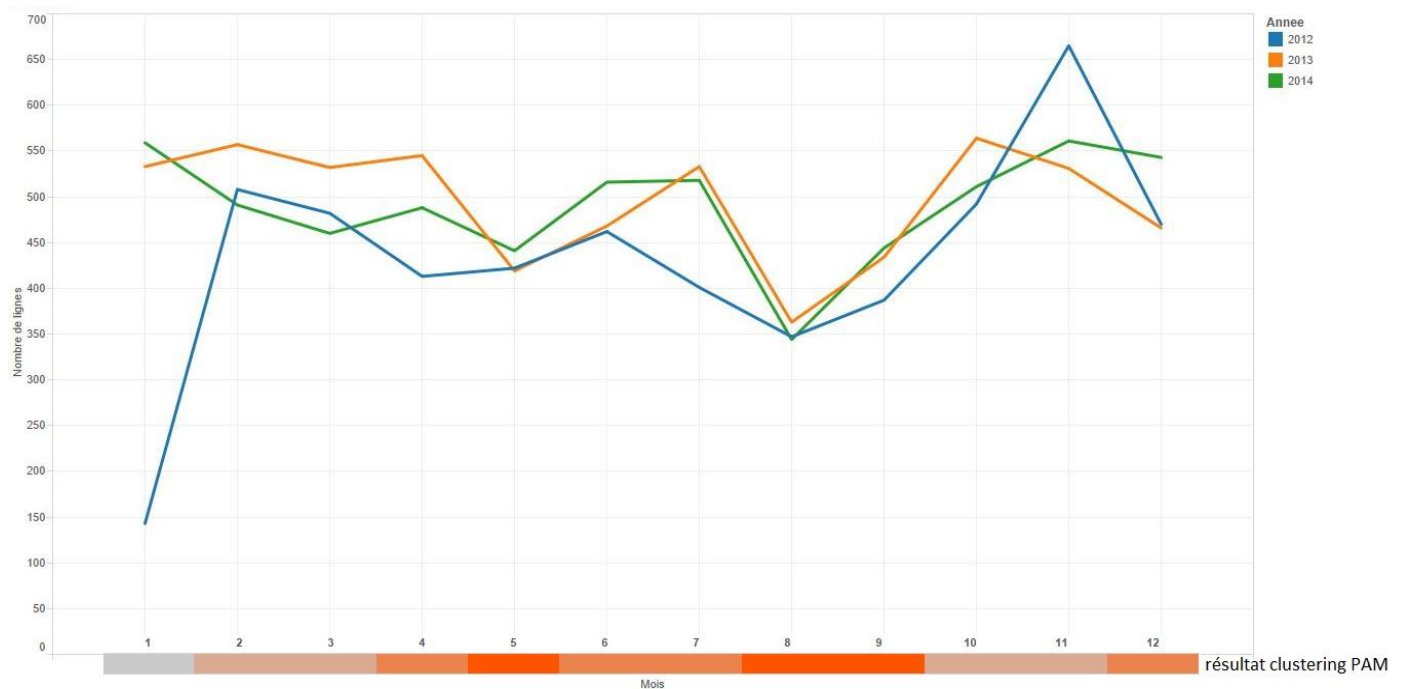
a. Analyse chronologique du nombre de publication

Nous avons pu récolter près de 17000 annonces entre 2012 et 2014 sur le total nous avons donc 15 annonces par jour en moyennes par mois, soit 472 annonces par un ce qui nous donne un peu après une annonce par jour (ce qui est peu au fond). Nous observons sur l'histogramme cumulé dans un premier temps le nombre d'annonces publiée par année, nous remarquons alors que l'année 2013 a eu une plus forte affluence que l'année 2014.

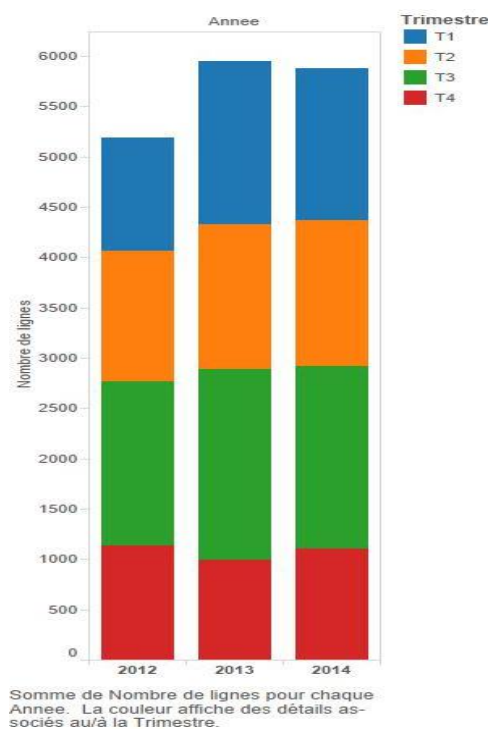
Nous traçons de plus ci-dessus à droite les courbes représentant le nombre d'occurrences par année et par mois. Nous remarquons que la courbe bleu est en en janvier très basse comparé aux autres, cependant dès le mois de février celle-ci remonte au niveau des années suivantes. Nous avons réalisé un cluster PAM avec $k=4$, cette typologie est transposée sur le graphique sous la forme de l'échelle orangée. Nous remarquons donc que le mois de janvier est classé seul (tel un point aberrant) ceci à cause de sa faible popularité en 2012. Dans le cluster saumon clair nous retrouvons les mois durant lesquels le nombre d'annonces est plus ou moins similaires en fonction des années, c'est aussi la période de « pic » en nombre d'annonces, nous repérons donc les mois de février, mars, octobre et novembre. En saumon foncé nous retrouvons les mois ayant une tendance négative, et enfin et orange nous trouvons les mois « creux » que sont aout, septembre (en général réservés aux vacances scolaires et salariale) ainsi que le mois de mai (nombreux jours fériés). Nous réalisons de plus un test de comparaison de la moyenne afin de voir si la hausse enregistré en 2012 était vraiment significative en utilisant un test non paramétrique, notre p-value atteint 0.01, nous pouvons donc dire que l'année 2012 enregistre une hausse significative, puisque nous avons 1

ANALYSE DES OFFRES DE STAGES DU SITE DE L'APEC

chance sur 10 de se tromper en affirmant cela. (cf *Testt_ANNEE.pdf*) Après ces observations nous pouvons dire que la recherche de stage coté entreprise se déroule d'octobre à mars.



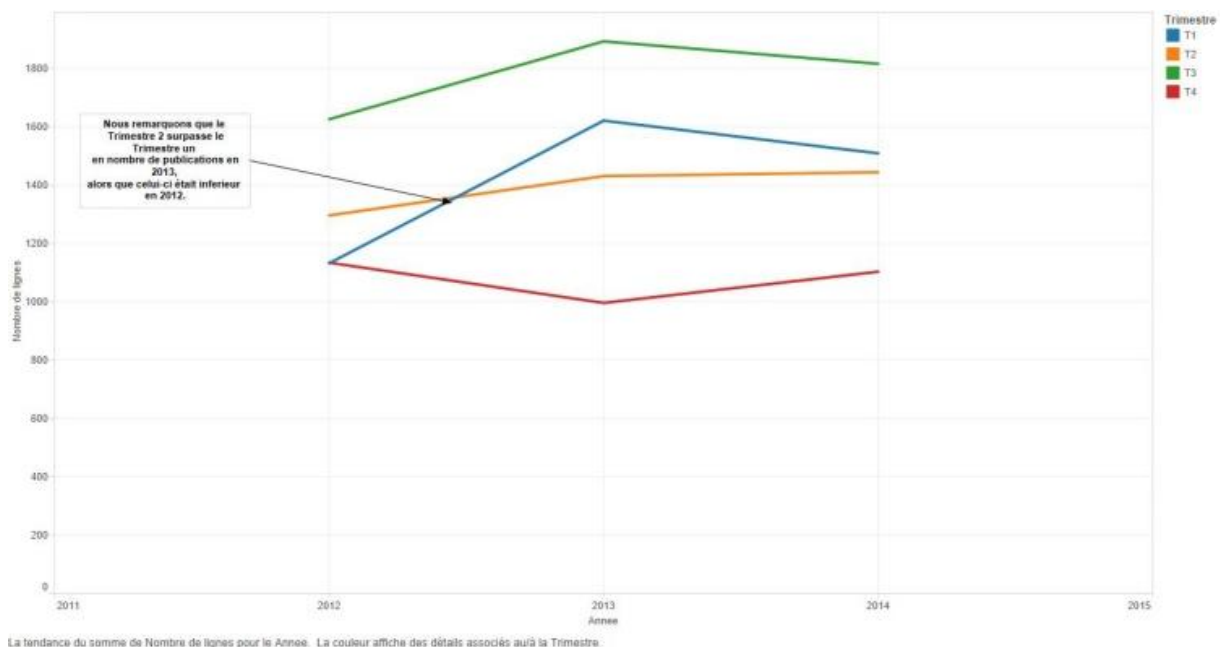
Afin d'aller plus loin dans notre analyse nous allons nous étudier le nombre d'annonces par trimestre et semaine.



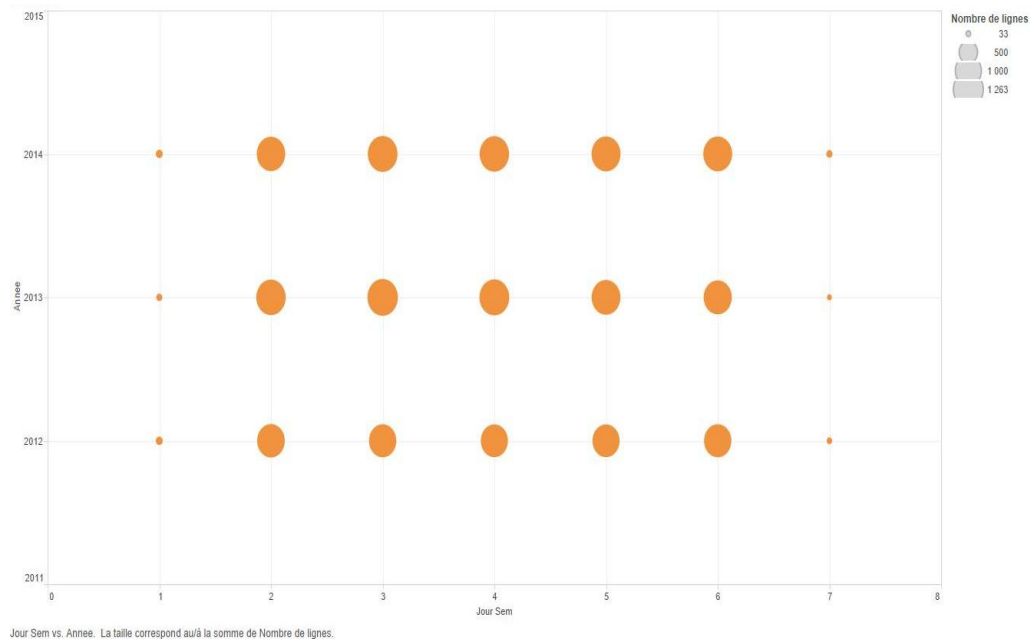
Le graphique ci-

dessous présente le

nombre d'annonces par trimestre, nous remarquons que le troisième trimestre surpasse largement les autres en termes de nombre de publications : cette période est effet celle prévue par la plupart des enseignements de formations. Il est de plus flagrant que les publications du troisième trimestre en 2012 sont en supériorité numérique. Ceci n'est pourtant plus le cas à partir de 2013 où le quatrième trimestre ce qui tend à nous faire passer que cette année-ci les entreprises s'y sont prisent en avance. Nous noterons enfin que les 3 droites ne sont pas du tout parallèle, signe d'un non dépendance entre trimestre. Nous réalisons dans cette optique des tests du khi2 entre le nombre d'annonces par année ainsi que par mois, les résultats de ceux-ci sont disponibles dans les documents *KHIDEUX_NB_ANONCE_ANNEE.pdf* et *KHIDEUX_NBMOIS.pdf* ceux-ci nous indiquant une forte dépendance entre ces deux variables et la date. Ceci signifiant que la pratique de recherche de stagiaire s'est développée depuis 2012.

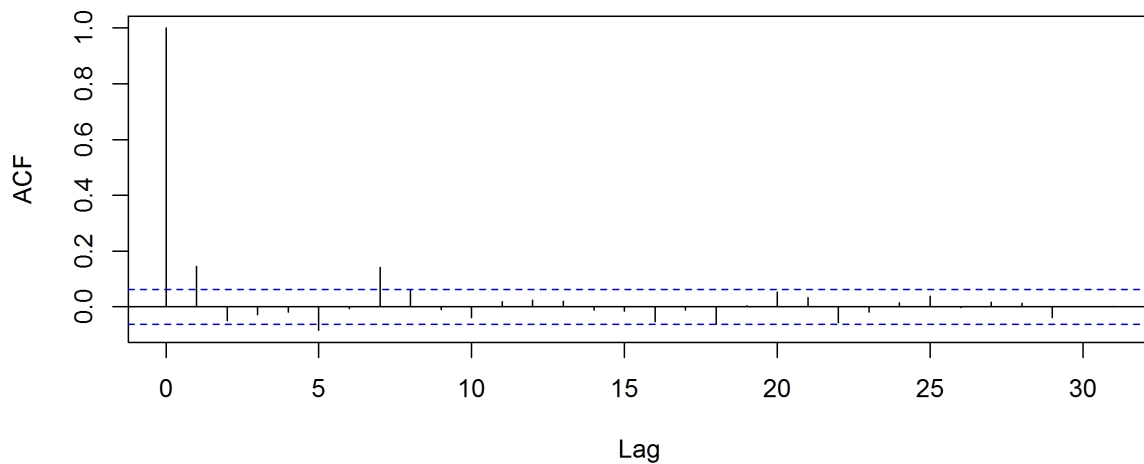


Au niveau jour nous remarquons que le gros des publications d'annonces se réalise du mardi au samedi, l'on remarque avec surprise que certaines entreprises publient leurs annonces le dimanche. Enfin nous noterons le faible poids du lundi, ce qui doit être dû au rythme des entreprises.

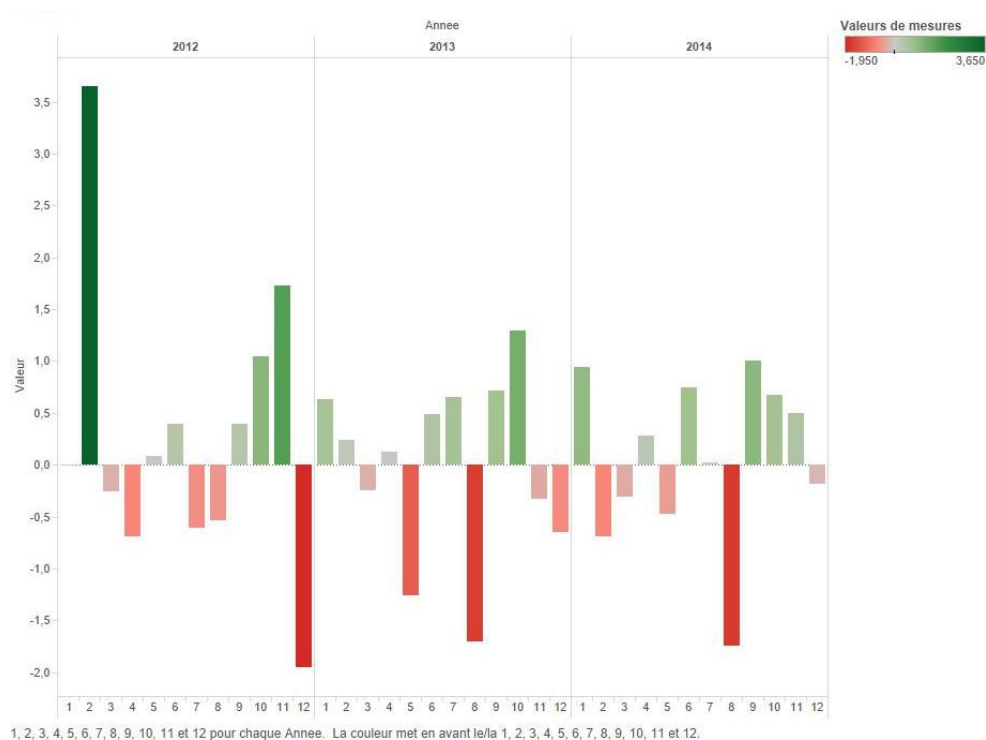


En traçons (via R dans notre cas), l'autocorrélation plot, traduisant de l'effet saisonnier du nombre de publications d'annonces, nous remarquons que le nombre de publications d'un jour dépendent de celui de la veille et de celui de la semaine passée.

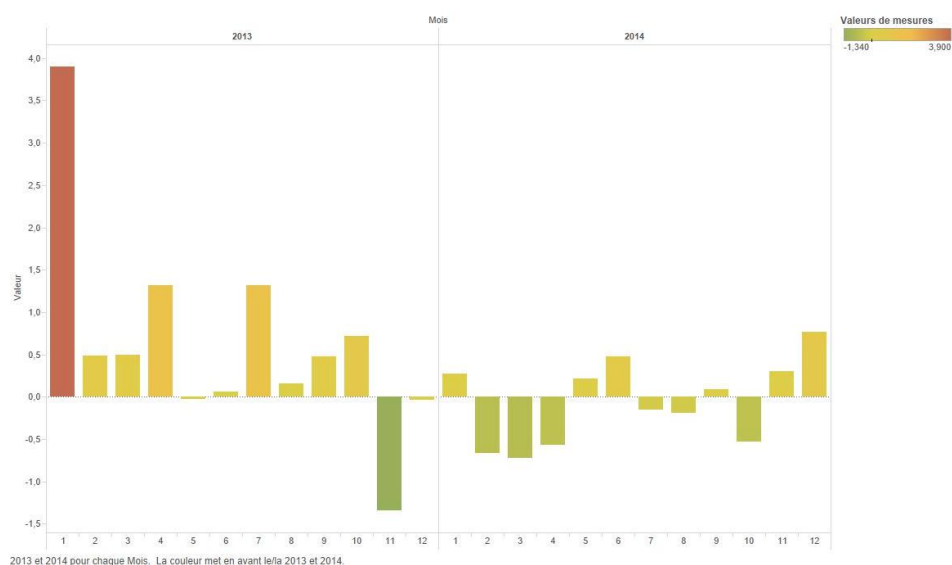
Auto-correlation plot sur les 31 jours



b. Analyse du taux d'évolution des publications

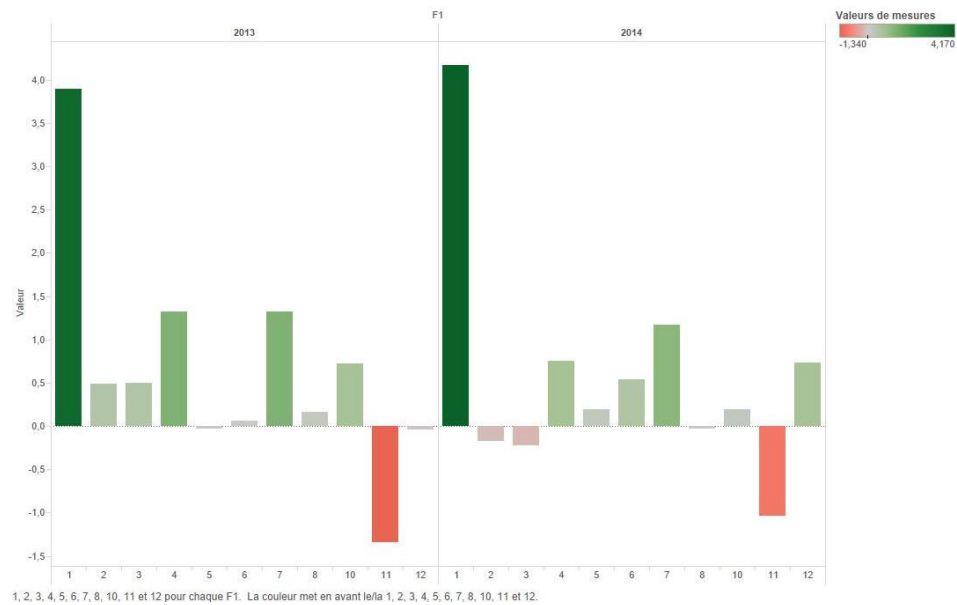
EVOLUTION RELATIVE AU MOIS M-1

Nous remarquons que les mois d'octobre de façon invariable entre année enregistre le plus haut taux d'évolution (obtenu en faisant le rapport $M/(M-1)$), à l'inverse le mois durant lequel l'évolution est toujours négative est celui d'août (mois réserver aux vacances rappelons-le).

EVOLUTION RELATIVE PAR MOIS COMPARE AU MOIS DE L'ANNEE A0 (2012)

Nous remarquons en faisant le taux d'évolution entre le M à l'année A et le mois M à l'année A0 (2012), que le mois de janvier de l'année 2013 connaît la plus grosse hausse comparé à notre année de référence et que le mois de novembre lui encaisse la décroissance la plus élevée. Sur l'année 2014 nous remarquons une suite de décroissance entre février et avril et des taux de croissance inférieur à ceux enregistré durant 2013.

EVOLUTION RELATIVE PAR MOIS COMPARE A L'ANNEE N-1

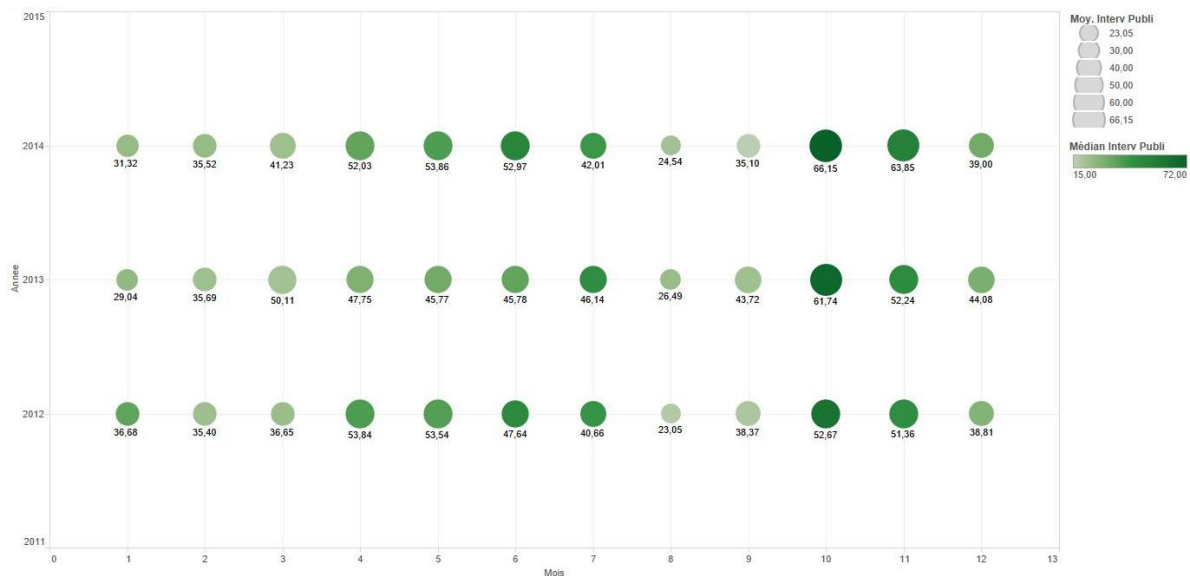


Nous remarquons que le mois de janvier ne cesse d'enregistrer des publications à la hausse durant les années, et que le mois de novembre au fil des années semble de plus en plus « délaissé » par les entreprises.

c. Analyse sur l'intervalle calculé entre la date de publication et la date de début annoncé

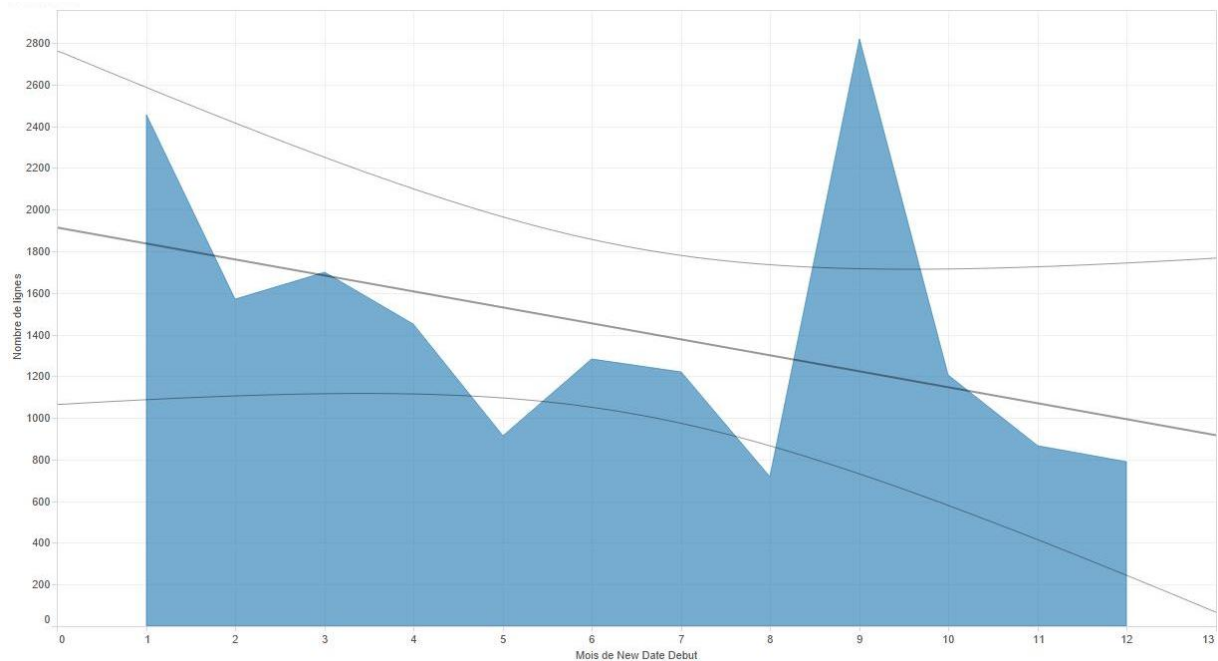
Nous réalisons quelques statistiques sous SAS afin de définir la moyenne de temps entre la date de publications et de début de stage. Nous remarquons que les entreprises s'y prennent sur les 3 années en moyenne entre 26 et 28 jours avant (cf. *MOY_INTERV.pdf*) ci-dessous une vue de l'intervalle moyen par mois et années.

Nous remarquons aussi qu'au fil des années les entreprises tendent à s'y prendre plus en avance quant à la publication de leurs annonces. Ce qui montre que la pratique semble s'être institutionnalisée au fil des années.



d. A quelle date commence en général les stages proposés ?

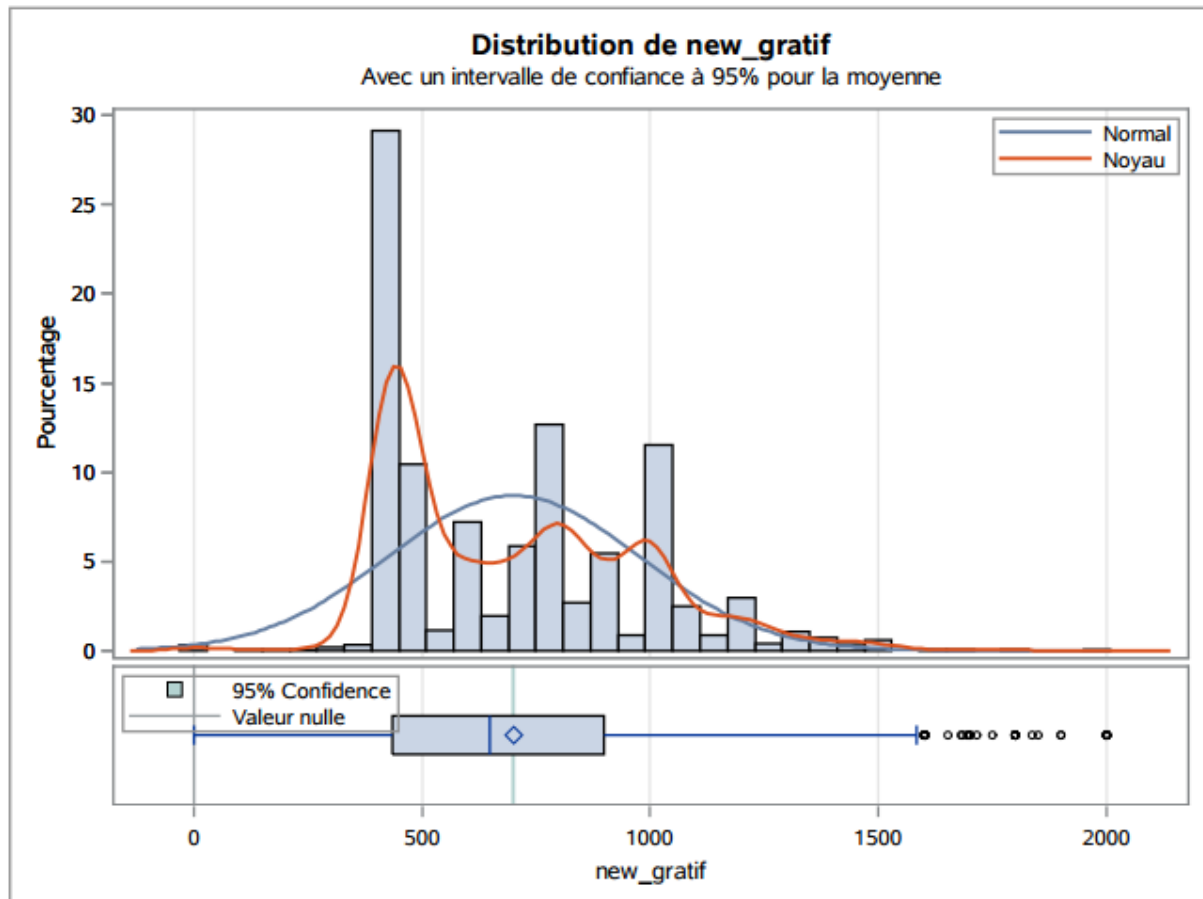
Nous remarquons en traçant ce graphe une première chose : les stages ont tendance à commencer plutôt en début d'année, cependant un pic se forme en septembre, ceci est loin d'être anodin, ce mois est plutôt consacré pour les étudiants à la rentrée scolaire. Cela nous invite à nous demander si ce pic ne représente pas les stages de pré-embauches.



Le tracé du somme de Nombre de lignes pour le New Date Debut Mois. Les données sont filtrées sur New Date Debut Jour, qui conserve 926 sur 942 membres.

2. Analyse des gratifications

Comme nous pouvions l'imaginer les salaires suivent une loi de Pareto : de nombreux stages sont très peu rémunérés et très le sont bien, nous remarquons alors une première chose : 30 % des annonces sont rémunérées moins que le minimum légal, ce qui est curieux c'est que l'on remarque c'est que 93 % de ces stages à basse rémunérations sont des stages d'une durée minimum de 3 mois, ceci allant à plus de 6 mois. Après un nettoyage des rémunérations aberrantes nous traçons la distribution des salaires ci-dessous :

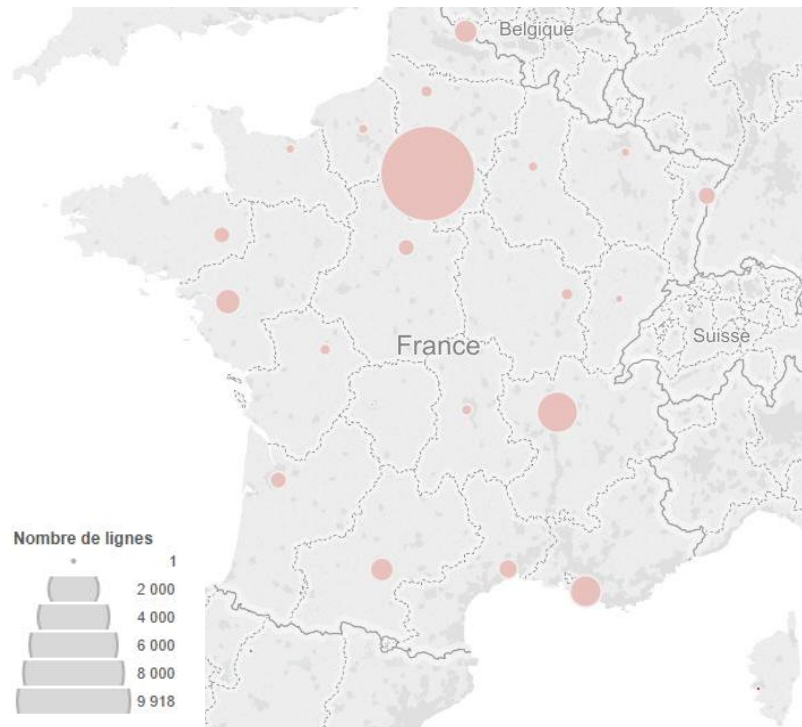


Ainsi nous remarquons une distribution multimodale qui présente trois pics : l'un à 450€ (minimum légal), l'un à 800€ et l'autre à 1000€. La moyenne globale des salaires est de 700€.

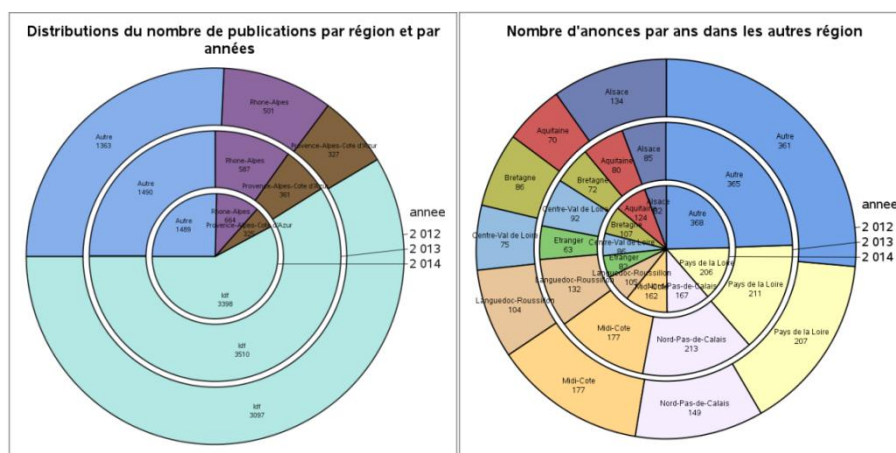
Nous faisons une rapide ANOVA afin de voir si nous avons une différence de salaires proposés en fonction des années, nous en tirons le résultat suivant : les salaires proposés en 2012 sont significativement supérieurs à ceux proposés en 2013 et 2014 (les années 2013 et 2014 n'étant pas significativement différents en terme de salaire). Cf. *ANOVA_ANNEE_SALAIRE.pdf*.

3. Analyse Géographique

Afin de définir les bassins de stages nous réalisons une projection cartographique du nombre de stages proposés par région :



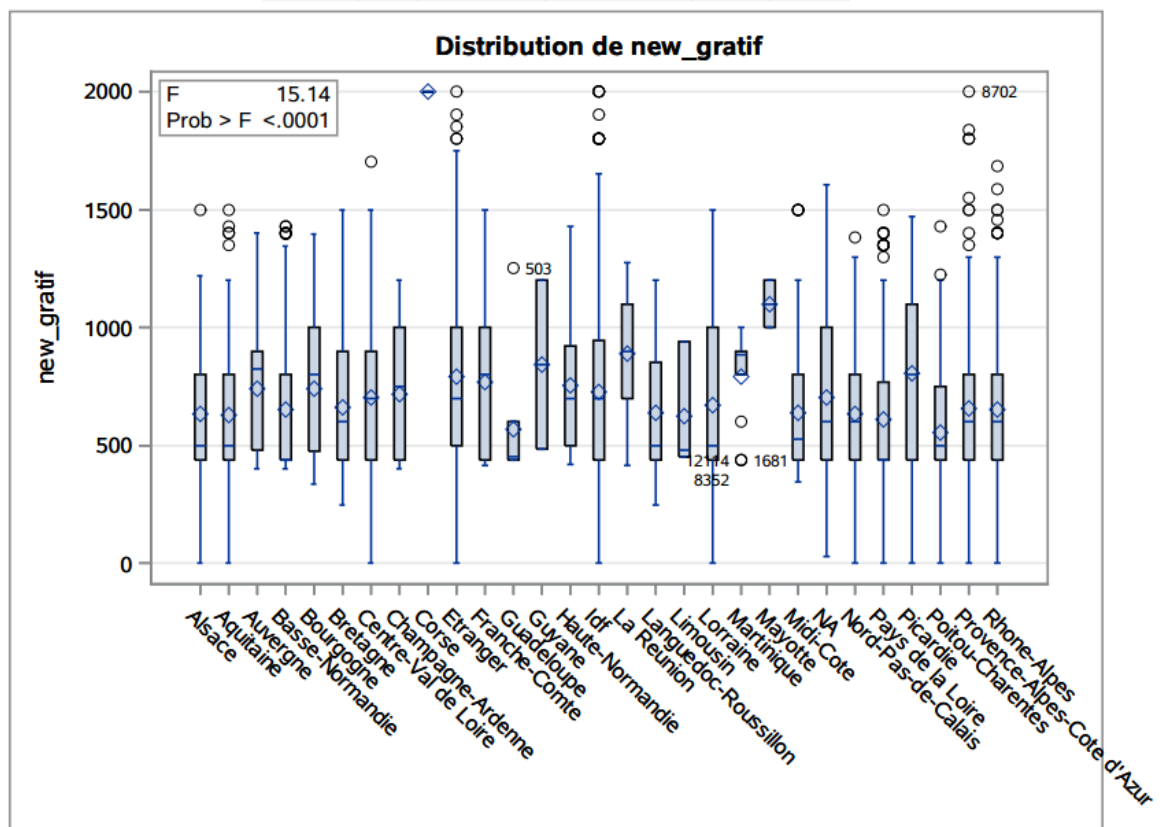
Le bassin parisien semble englober une grande part des stages, semblent suivre derrière la région PACA et Rhône-Alpes. En effet en projetant ci-dessous le nombre stage par année et par région nous obtenons le diagramme circulaire ci-dessous à gauche, à droite nous avons ôté ces trois régions afin de pouvoir observer la distribution des stages dans les autres régions.



Nous notons que la répartition par région n'évolue que très peu selon les années, de plus nous retenons ces quelques chiffres : 58% des stages se situent en régions parisienne, 10% en région Rhône-Alpes, 6% en région PACA, enfin 1% se déroulent à l'étranger.

Ce que nous validons statistiquement la relation entre la région et le nombre de stages dans le rapport sas *KHIDEUX_NB_REGION.pdf*.

Nous réalisons aussi dans le fichier *ANOVA_REGION_SALAIRE.pdf* une comparaison des distributions de salaires par région, dont est tiré ce graphique. Nous y comparons région par région les différences de salaires.



Une lecture détaillée de ce rapport permet de comparer région par région les salaires proposés ainsi que l'écart entre celles-ci.

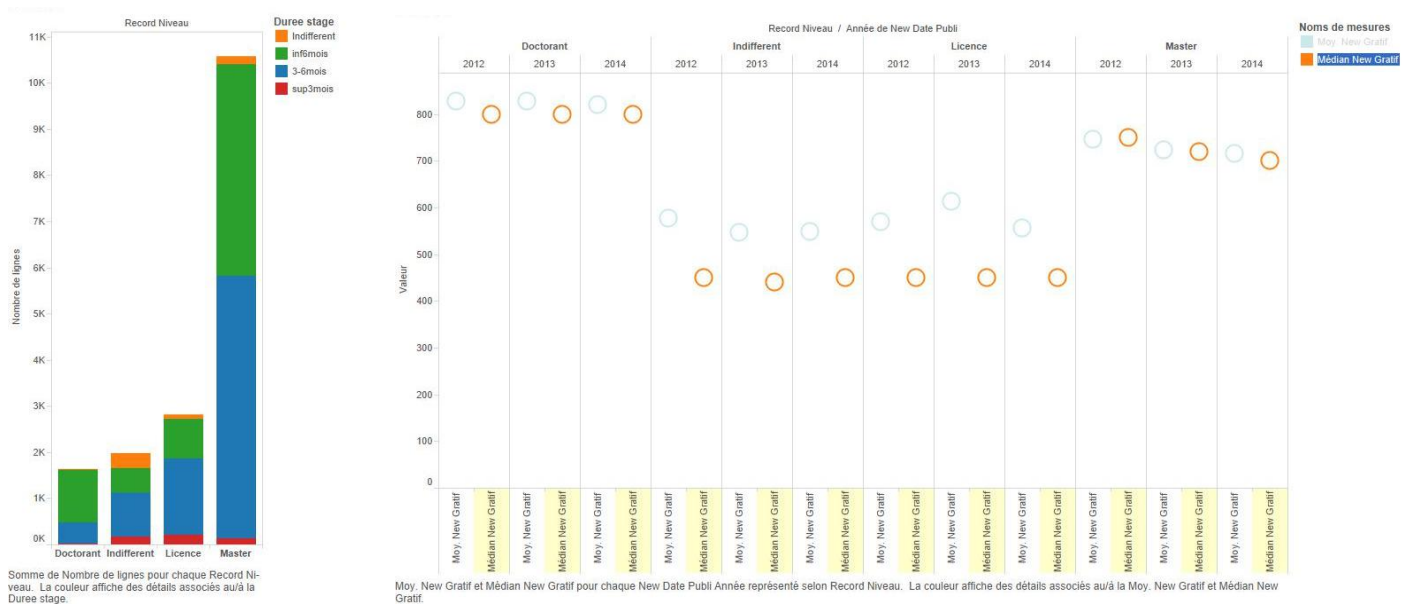
4. La durée du stage

Dans un premier temps nous réalisons un test du Khi2 (cf. *KHI_DEUX_DUREE_STAGE.pdf*) afin de vérifier la dépendance entre le nombre d'annonces en fonction de la durée. Nous notons que le site de l'APEC propose 50% de stage d'une durée de 3 à 6 mois, 40% de stage d'une durée supérieur à 6 mois les 10% restant étant des stages d'une durée inférieure à 3 mois ou d'une durée indifférente. Une anova (cf. *ANOVA_DUREE_SALAIRE.pdf*) nous permet par la suite de mesurer l'impact de la durée du stage sur la gratification nous notons que les stages sont classé par ordre de rémunération de cette manière :



5. Analyse du niveau d'étude requis

En réalisant un test du khi 2 (*KHI2_NB_NIVEAU.pdf*) nous remarquons que le nombre de publication dépend du niveau d'étude en effet 60% des annonces visent des étudiants d'un niveau Master, 15% sont appropriés pour des élèves de Licence, suivent ensuite les annonces ne précisant pas le niveau d'étude requis, enfin nous retrouvons le niveau doctorat (niveau étude supérieur à bac +5) en dernière position.



Ces deux graphiques nous permettent d'illustrer plusieurs phénomènes : la part des annonces par durée et niveau d'étude via le graphique en baton cumulé à gauche, le graphique de droite relate la moyenne et la médiane des annonces par année et niveau d'étude.

Nous remarquons que les doctorant sont en moyenne les mieux rémunérés (près de 800€), les stages qui demandent un niveau licence minimum et ne demandant pas de niveau d'étude spécifiques sont les moins bien rémunérés entre 600€ et 400€ selon les années. Enfin les stages demandant un niveau master sont rémunérés près de 750€, notons que cette rémunération semble légèrement diminuer depuis 2012.

6. Analyse des champs textuels

a. Le titre de l'annonce

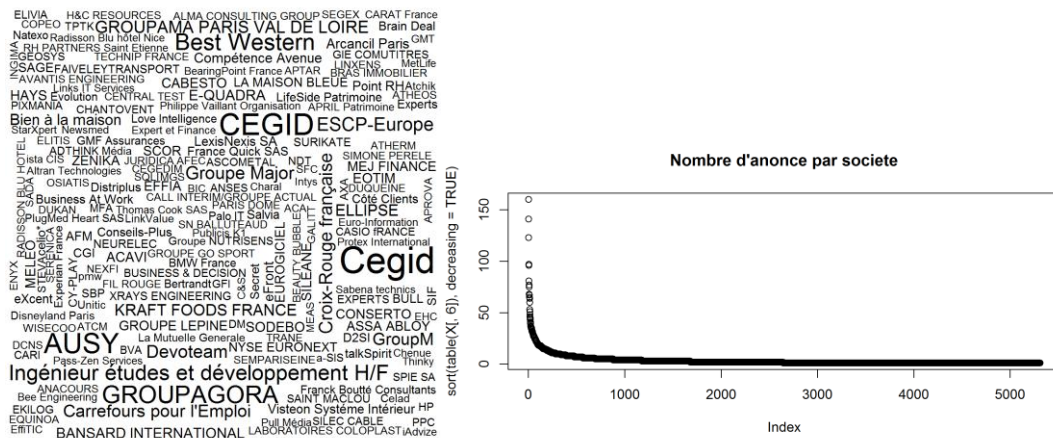


En réalisant une manipulation spécifique nous obtenons le nuage de mot suivant illustrant les principaux mots (racinisés) revenant dans les titres des annonces. Nous réalisons une régression linéaire à titre indicatif afin de connaître l'influence de ces mots sur le salaire. Cette méthode nous permet de savoir quelles sont les variables qui ont le plus d'influence sur celui-ci. Ainsi nous reportons les mots significatifs à 0.05% afin d'observer leurs coefficients :

Mot	Coefficient
Analyst	159
assistante	-50
commercial	-90
consultant	121
développeur	-9
ingénieur	74
junior	60
produit	70
recherche	-146
Recrutement	-96
web	-77

Nous pouvons classer ces mots dans deux catégories ceux ayant une influence positive sur le salaire, et ceux en ayant une négative. Nous remarquons que le mot « analyste » est celui qui donne le plus de chance d'être rémunéré au dessus de la moyenne, et ainsi que le mot « recherche » une fois présent dans le titre de l'annonce promet un salaire inférieur à la moyenne.

b. Le nom de la société



En étudiant de plus près le nom des sociétés nous remarquons les 93 entreprises les plus citées dans les annonces, correspondant donc aux grands fidèle du site, regroupent à eux seuls près de 36% des annonce, ces entreprises suivent la tendance générale en publiant près de 64% d'annonces à l'intention des élèves de master. La plupart de ces entreprises sont de grands groupes, et nombre d'entre elles ont un rapport avec les technologies numériques. Nous observons aussi un fait intéressant : de nombreuses entreprises font appel à de cabinet de recrutement pour chercher des stagiaires.

c. Poste et mission



Nous réalisons comme pour le titre une régression linéaire dont nous tirons les informations présente dans le tableau ci-dessous :

Mot	Coefficient
Depart	49
Entreprise	-71
France	61
Management	73
Mission	-25
Recherche	-35
recrute	-65
Sein	40
suivi	-63

Nous remarquons que le mot « départ » surement assimilé à une recherche de stagiaire à la suite d'un départ en vue d'une embauche, possède un coefficient positif. Le reste des mots retenus, n'est pas très signifiant quant au poste proposé.

d. Profil recherché



Les mots présent dans le profil relatent : du niveau d'étude, des outils informatique, et des qualités attendues du stagiaire.

Mot	Coefficient
Ingenieur	61
Commercial	-65
Communication	-74
Fin	72
Grand	67

Ces quelques mots coefficienté, nous permette de comprendre que les stages techniques sont mieux rémunérés que les stages demandant plus de compétences orale. De plus le mot « fin » n'est pas anodin car encore une fois il nous laisse penser que de nombreux stages proposé sont des stages de fin d'étude.

7. Un mot sur les stages à l'étranger

Nous réalisons dans un premier temps un test de Student afin de comparer les gratifications moyennes des stages à l'étranger à ceux en France, nous obtenons une p-value supérieure à 0.05 nous indiquant que les stages en France sont en moyenne mieux rémunérés. 68% des stages proposés sont des stages longs durés (supérieures à 6mois), et 12% sont à l'initiative de l'entreprise Mosaik (un cabinet de recrutement). Enfin nous noterons que ces stages demandent pour la plupart un bac +5, ou une formation menant à un bac +5.

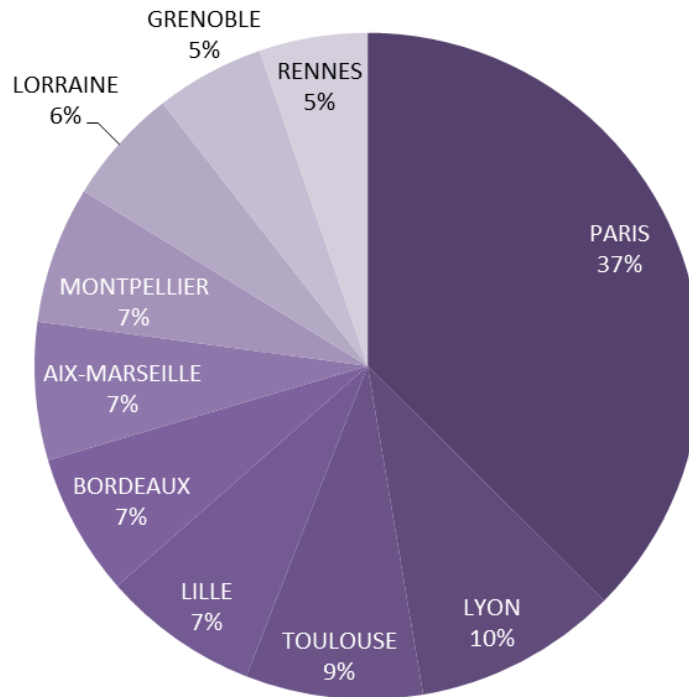
B. Analyse des données provenant d'autres sources

Chaque offre de stage diffusée sur le site de l'APEC comme d'autres, vise un public qui se restreint selon notre étude sur les étudiants inscrit dans les établissements de la république française c'est dans cette mesure que nous avons cherché des données qui complètent ceux de l'APEC :

1. Nombre d'inscriptions par Ville

Après un nettoyage de données qui consistait à établir des agrégations au niveau des inscrits en fonctions des villes, nous avons aboutis à :

10 premières villes qui ont les plus d'inscriptions



Nous remarquerons clairement dans un premier temps que la région parisienne demeure la première en terme de nombre d'inscrits avec un score de 339752 étudiants suivis de la région lyonnaise avec un chiffre de 89821, en revanche MARTINIQUE était la dernière selon le critère choisi, la question qui s'impose : ces étudiants auront-ils des offres de stage dans leurs villes ? ces villes ont-elles la capacité de garantir à leur progéniture des insertions professionnelles durant le cursus académique ?

Toutefois, d'après la section (Analyse Géographique) la région parisienne offre une grande part des stages, semblent suivre Rhône-Alpes. En effet cette grande marge d'offre fournie par ces régions ne peut que provenir du fait que ces dernières englobent la majorité des étudiants français considérées comme des villes de masses, ce qui permet clairement de répondre aux questions : le nombre d'offre de stage est en adéquation avec la demande académique.

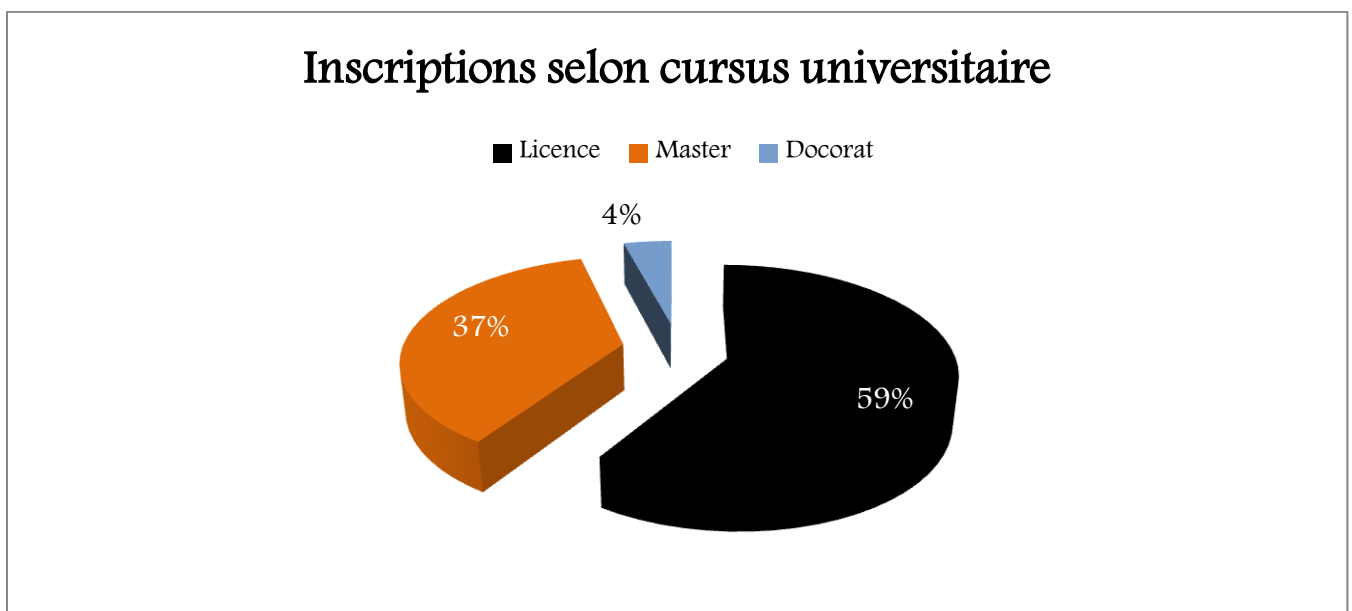
2. Nombre d'inscriptions par Ville et par cursus universitaire

a. Par Licence/Master/Doctorat

Pour raffiner d'avantage notre étude on s'est posé la question suivante : est-ce que le niveau d'étude impacte la faisabilité de l'insertion professionnel dans d'autres termes l'APEC garanti-t-elle toutes les demandes pour n'importe quel niveau d'étude ?

37% d'étudiants inscrit sont au cycle licence 59% sont en master et uniquement 4% ont le souffle afin d'intégrer l'établissement doctorale

D'après la section 5 (Analyse de niveau d'étude) 60% des annonces visent des étudiants d'un niveau Master, 15% sont appropriés pour des élèves de Licence et 7% sont mise pour les doctorants ce qui est conforme avec les taux des demandes en générale.



IV. Conclusions

Analyse des stages de l'APEC :

Nous pouvons à la suite de cette analyse y voir plus clair sur l'offre de stages proposés par le portail, en effet nous avons pu au fil de nos analyses dresser un tableau de ce qu'attendaient les entreprises publiant des annonces sur ce site. Nous pouvons retenir que les entreprises cherchent leurs stages en général sur la période d'octobre à mars, et qu'elles s'y prennent en moyenne à mois à l'avance. Géographiquement nous avons pu repérer trois gros bassins de recrutement : l'île de France, ainsi que les régions Rhône-Alpes et PACA. Au niveau de la gratification de stage nous noterons que celle-ci avoisine les 700€ en moyenne, ce chiffre est assez éloigné du minimum légal, à cause des stages de fin d'études qui tirent celle-ci vers le haut. De plus nous avons pu remarquer que ces annonces étaient plus destinées à des personnes en fin de formation, cherchant un stage de fin d'étude. Nous avons de plus pu dresser un portrait des entreprises recrutant sur le site des stagiaires majoritairement de grandes entreprises ayant un lien avec le numérique ou l'industrie. Cherchant pour la grande part des personnes maniant les outils informatiques, et ayant un fort sens de l'analyse. Enfin nous avons pu remarquer un fait non négligeable : beaucoup d'entreprises délèguent cette mission de recrutement à des cabinets spécialisés.

En nous situant en tant qu'étudiant en informatique nous avons remarqué que les qualités acquises durant notre formation était en général très appréciées des recruteurs, à condition bien sûr de choisir un des trois bassins de recrutement.

Cependant nous noterons que nous avons effectué ces analyses sur un échantillon dont nous n'avons pu mesurer sa représentativité, ces résultats ont donc leurs limites. Une amélioration de la technique de web scraping nous permettrait dans un premier temps de posséder plus de données à analyser (en effet nous perdrons 80% de nos annonces à cause de soucis de formatage), et plus tard de réaliser un échantillonnage de manière plus rigoureuse.

Analyse des données provenant d'autres sources :

Les principaux problèmes, que nous avons rencontrés, concernaient la collecte de données, vu qu'on a construit notre base de données. Ainsi, nous avons touché du doigt la difficulté d'extrapoler le jeu de données disponibles pour obtenir des valeurs cohérentes sur l'ensemble du domaine étudié, ce que pourtant nous serons vraisemblablement amenés à faire dans notre futur métier.

Toutefois, si une telle étude devait être complétée, à plus ou moins longue échéance, par un bureau d'étude, il serait nécessaire de réaliser des jointures entre les deux ressources qu'on n'a pas eu le temps d'achever.

Sur le projet :

Ce projet fut très formateur, car nous avons dû travailler par équipe et réaliser un travail sur la base de données provenant de sources différentes en utilisant divers outils afin d'analyser nos données, en utilisant SAS pour la partie Data management, ainsi que pour de nombreuses analyses, nous avons pu compléter nos analyses afin de combler le manque d'outils d'analyse de textmining sous SAS, par l'utilisation de R dans la phase de création de nos matrices termes documents. Nous avons de plus utilisé Tableau afin de réaliser des graphiques propres afin d'illustrer nos analyse. Et enfin nous avons utilisé Excel comme lien entre chacun de ces logiciels.