

Analyse du marché de l'emploi au Canada

Nom des membres de l'équipe :

Hamadoun Dicko

DICH19079502

INF8100 gr. 031

Automne 2023

Remise a :

Nairouz Mrabah

Introduction

Dans un contexte économique en constante évolution, l'analyse du marché de l'emploi devient cruciale pour comprendre les tendances actuelles et futures. Ce rapport se concentre sur le marché de l'emploi canadien, en utilisant les données du site d'emploi **Jobillico**. L'objectif principal est de répondre à une question concrète : "**Quels sont les facteurs déterminants influençant les recherches d'emplois dans les différents secteurs au Canada ?**" Cette analyse nous permettra de mieux comprendre les dynamiques du marché du travail et d'identifier les opportunités et défis pour les chercheurs d'emploi et les employeurs.

Collecte et Description du jeu de Données

Les données ont été récupérées par le biais du ratissage du site (Jobillico), un important portail d'emploi canadien. Le jeu de données comprend les colonnes suivantes :

- **Titres** : Intitulé du poste
- **Compagnies** : Nom de l'entreprise
- **Villes** : Localisation du poste
- **Regions** : Région géographique
- **Heures** : Nombre d'heures travaillées par semaine
- **Type Emplois** : Catégorie d'emploi (Full time, Part time, etc.)
- **Salaires** : Salaire minimum de base offert
- **Salaires Max Possibles** : Plafond salarial potentiel
- **Contrats** : Type de contrat

Les données collectées ont par la suite été stocker dans le fichier csv intitulé « jobs.csv ». Le code du ratissage est élaboré et bien détaillé dans le fichier « ProjectTP3.ipynb » de la partie Ratissage des données.

Partie 1 : Nettoyage et exploration des données

Dans cette section, nous procéderons au nettoyage des données pour éliminer les incohérences et les valeurs manquantes. Les données seront également converties dans des formats appropriés pour l'analyse, en transformant les salaires en valeurs numériques, etc. Cette première phase est essentielle pour garantir que les analyses subséquentes soient basées sur des données fiables et précises. Elle nous permet d'aborder les prochaines étapes de l'analyse avec une base solide, prête pour une exploration plus détaillée dans les sections suivantes du rapport.

Questions traitées :

1. Compter le nombre de valeur manquantes dans toutes les cellules du jeu de données

```
Le nombre de valeurs manquantes : Titres          0
Compagnies          0
Villes              0
Regions             0
Heures              0
Type Emplois        1084
Salaires             7631
Salaires Max Possibles 7808
Contrats             0
dtype: int64
```

2. Convertir les colonnes heures, Salaires et Salaires Max en float

Le résultat de cette question se trouve dans le fichier ProjectTP3.ipynb

3. Supprimer les valeurs aberrantes dans les colonnes Salaires et Salaires Max Possibles

Dans cette partie on supprime les salaires représentant les salaires annuels et on garde les salaires par heures dans notre jeu de données pour ne pas biaiser nos analyses futures. On est passé de 9800 rows × 9 columns à (9054 rows, 9 columns)

4. Supprimer les valeurs nulles et dupliquées du jeu des données

Le résultat de cette question se trouve dans le fichier ProjectTP3.ipynb

Partie 2 : Analyse et visualisation des données

Nous analyserons les données pour dégager des tendances et des motifs. Cette analyse comprendra :

Questions traitées :

1. Faites une analyse descriptive des salaires

Résultat de l'analyse univariée

	Salaires	Salaires Max Possibles
count	9054.000000	9054.000000
mean	3.985020	4.129663
std	9.350014	10.458614
min	0.000000	0.000000
25%	0.000000	0.000000
50%	0.000000	0.000000
75%	0.000000	0.000000
max	35.690000	39.860000

2. Comparer la moyenne des heures de travail par région du Canada

Type Emplois	Regions	
Full time	ON	37.500000
	QC	40.466351
Full time , Part time	QC	40.000000

Name: Heures, dtype: float64

3. Donner les types de contrats les plus fréquents par région au Canada

Regions	Contrats	
AB	Contract , Permanent , Seasonal job	50
BC	Permanent job	47
ON	Permanent job	46
QC	Permanent job	6336
	Permanent , Telecommuting job	1538
	Seasonal job	700
	Casual , Permanent , Seasonal job	64
	Contract job	51
	Contract , Permanent job	47
	Permanent , Unpaid internship job	47
	Permanent , Seasonal , Student employment job	44
	Contract , Casual job	42
	Contract , Telecommuting job	42

Name: Contrats, dtype: int64

4. Compter le nombre d'emplois par compagnie

Les compagnies avec le plus d' offres d' emplois :

Les Entreprises Michaudville inc.	1400
Village Vacances Valcartier	808
Lainco inc.	700
DSD International inc.	700
Tecno-Métal inc.	700
TBC Constructions inc.	700
Entreprises G.N.P. inc.	700
LES ENTREPRISES GÉNÉREUX	700
Les Services Electro-Techno Inc.	700
Lemieux Nolet, Comptables Professionnels Agréés s.e.n.c.r.l	700

Name: Compagnies, dtype: int64

Les villes avec le plus d' offres d' emplois :

Saint-Gabriel-de-Valcartier	808
Victoriaville	797
Quebec City	790
Terrebonne	700
Thetford-Mines	700
Montreal	700
Mont-Saint-Hilaire	700
Brossard	700
Saint-Jean-de-Matha	700
Sherbrooke	700

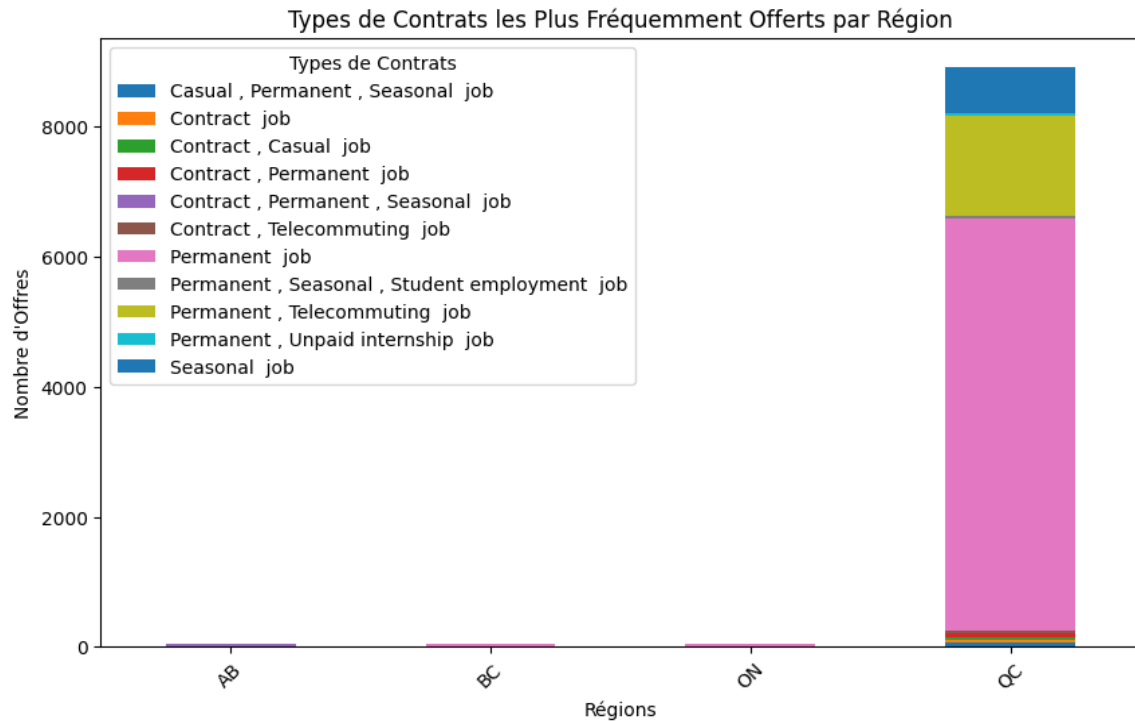
Name: Villes, dtype: int64

5. Calculer le nombre d'heures moyen de travail dans la région du Québec "QC"

Le nombre moyen d'heures de travail dans la région du Québec est de 35.98 heures.

Analyses et Visualisation des données

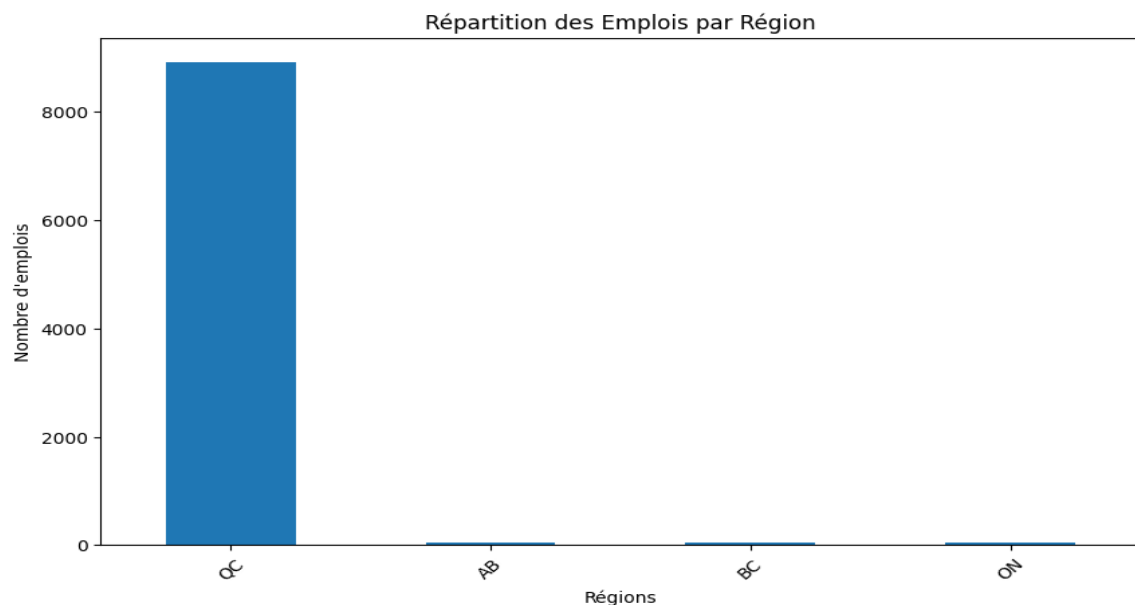
6. Faites une visualisation et une analyse des salaires moyen par titre et par région



Analyse du graphique

Le graphique montre une répartition variée des types de contrats d'emploi par région, avec une prédominance notable dans la région "QC" pour les stages non rémunérés et le télétravail permanents. Les régions "ON", "AB" et "BC" présentent moins d'offres, avec une tendance pour les contrats permanents en "ON". La diversité des contrats est plus marquée dans les régions ayant un plus grand nombre d'offres, ce qui peut refléter une économie plus diversifiée ou un marché de l'emploi plus grand.

8. Visualiser la répartition géographique des emplois



Le graphique montre la répartition des emplois dans quatre régions : Québec (QC), Alberta (AB), Colombie-Britannique (BC), et Ontario (ON). Québec a nettement plus d'emplois que les autres régions, avec plus de 8000 emplois. Les autres régions semblent avoir beaucoup moins d'emplois, mais les chiffres exacts ne sont pas visibles. Cela se produit souvent lorsque la taille de la catégorie dominante éclipse les autres catégories.

Villes

- Anjou
- Boisbriand
- Boucherville
- Brossard
- Bécancour
- Danville
- Dorval
- Edmonton
- La Durantaye
- Langley
- Laval
- Longueuil
- Lévis
- Mont-Saint-Hilaire
- Montreal
- Quebec City
- Saint-Gabriel-de-Valcartier
- Saint-Jean-de-Matha
- Sainte-Croix
- Sainte-Marie
- Sherbrooke
- Terrebonne
- Thetford-Mines
- Toronto
- Trois-Rivières
- Victoriaville
- Warwick

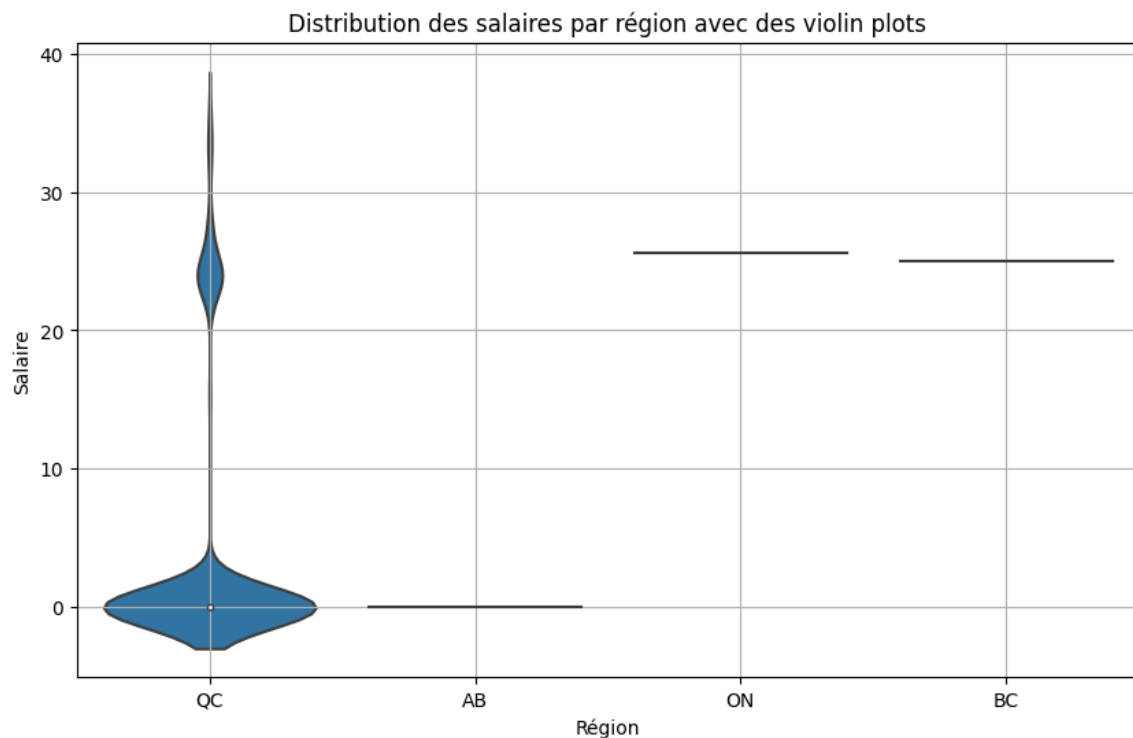
Nombre d'Emplois par Ville et Région

Région	Cities (from bottom to top)	Total Jobs (approx.)
AB	Edmonton	~100
BC	Langley	~100
ON	Toronto	~100
QC	Montreal, Brossard, Longueuil, Laval, Boisbriand, Thetford-Mines, Sherbrooke, Trois-Rivières, Victoriaville, Saint-Gabriel-de-Valcartier, Saint-Jean-de-Matha, Dorval, Boucherville, Danville, Warwick, Anjou	~9000

Analyse du graphique

Le graphique est un diagramme à barres empilées qui montre le nombre d'emplois répartis par ville au sein de quatre régions : Alberta (AB), Colombie-Britannique (BC), Ontario (ON) et Québec (QC). Chaque couleur dans les barres représente une ville différente. Les emplois dans la région de Québec sont les plus nombreux, avec une grande variété de villes contribuant à ce total. Les autres régions ont moins d'emplois et moins de diversité de villes par rapport à Québec. La légende à gauche identifie les villes représentées par chaque couleur dans les barres.

10. A l'aide d'un graphique, analyser la relation entre la valeur du nombre d'heures travaillées et le salaire. Y a-t-il un lien quelconque ? Est-ce que la région y joue un rôle dans cette relation ?



Analyse du graphique

D'après le graphique de violon il semble que la région QC ait une distribution de salaire plus variée et un salaire médian inférieur comparé aux autres régions. Les régions AB, ON et BC ont des distributions moins variées, comme indiqué par la minceur de leurs violons, et on ne peut pas distinguer de différence significative dans les médianes de salaire à partir de ce graphique.

L'analyse des coefficients de corrélation indique qu'il existe une tendance générale où le salaire diminue à mesure que le nombre d'heures travaillées augmente, ce qui peut

sembler contre-intuitif dans de nombreux contextes de travail. Il serait intéressant d'examiner plus avant les raisons de cette tendance dans les données. Cela pourrait être dû à la présence de travailleurs à temps partiel qui gagnent des salaires horaires plus élevés ou à d'autres facteurs structurels ou économiques.

L'analyse de variance (ANOVA) montre clairement que la région a un effet significatif sur le salaire. Cela pourrait signifier que des facteurs géographiques, économiques, ou des politiques régionales influencent les niveaux de salaire.

Partie 3 : Modèle de Régression et de classification

1. Existe-t-il une Corrélation entre les nombres d'heures de travail et les salaires

```
{'slope': -0.4890900704124664, 'intercept': 21.397036933515604, 'r_value': -0.7003962401282916, 'p_value': 0.0, 'std_err': 0.005238675102077443}
```

Le résultat de la régression linéaire montre une corrélation négative et statistiquement significative entre les heures de travail et les salaires, indiquant que les salaires diminuent à mesure que le nombre d'heures travaillées augmente. Cependant, cette tendance est contre-intuitive et pourrait nécessiter une investigation plus approfondie pour comprendre les raisons sous-jacentes, comme des facteurs spécifiques à l'industrie ou à la nature des emplois inclus dans votre jeu de données.

2. Dans cette question, on s'intéresse à prédire si le salaire d'un emploi sera supérieur ou inférieur à 20\$ en fonction de la région, du nombre d'heures travailler, du type d'emplois et du contrat. A l'aide d'un algorithme de régression?

```
jobs['Salaire_sup_20'] = (jobs['Salaires'] > 20).astype(float)
0.9602429596907786
```

	precision	recall	f1-score	support
0.0	0.97	0.99	0.98	1504
1.0	0.93	0.83	0.88	307
accuracy			0.96	1811
macro avg	0.95	0.91	0.93	1811
weighted avg	0.96	0.96	0.96	1811

Le modèle fonctionne très bien, avec une précision élevée et des scores F1 élevés pour les deux classes. Cependant, il est légèrement meilleur pour détecter les salaires ≤ 20 (classe 0.0) que pour les salaires > 20 (classe 1.0). Si on veut améliorer davantage les performances, notamment pour la classe 1.0, on pourrait envisager de rééquilibrer

l'ensemble des données, d'ajuster les paramètres du modèle, ou d'utiliser des techniques de validation croisée pour une évaluation plus robuste.

Conclusion et Recommandation

L'analyse du marché de l'emploi canadien, basée sur les données de Jobillico, a révélé des insights précieux sur les dynamiques actuelles du travail au Canada. Notre étude a permis de traverser diverses phases, allant du nettoyage et de la préparation minutieuse des données à des analyses approfondies, et enfin à la modélisation prédictive. A la base nous avons déterminé que la région du Québec par exemple offre le plus d'emplois parmi notre jeu de données et que les salaires y sont compétitifs et le nombre d'offres y publier représente un grand vivier.

Les résultats de cette analyse offrent des implications importantes pour divers acteurs du marché du travail. Pour les chercheurs d'emploi, cette étude fournit des indications sur les régions et villes qui sont les plus lucratives, fournit les types de contrat fréquemment donner et le nombre d'heures à laquelle on pourrait s'attendre de travailler. Pour les employeurs, elle met en lumière les standards salariaux et les tendances du marché, leur permettant d'ajuster leurs offres pour attirer les meilleurs talents.

Annexes

<https://www.jobillico.com/en>

INF8100