Description des résultats

Hypothèse générale : le salaire va fluctuer selon divers paramètres socio-démo.

* Hypothèse 1 : Le salaire va augmenter avec le nombre d’années d’études pour les hommes et les femmes.
  + On peut voir à travers la figure 1 et la figure 2, que le salaire augmente bel et bien avec les années d’études. On peut voir que cette tendance est présent tant chez les hommes que les femmes. L’hypothèse nulle est donc rejetée en utilisant un test-t.

Figure 1 et 2:

Chart, bar chart, histogram

Description automatically generated

Chart, bar chart, histogram

Description automatically generated

* Hypothèse 2 : Les hommes ont tendance à gagner plus pour les études sous-graduées alors que cet écart n’est plus significatif une fois atteint les études graduées.
  + En effet, il y a une grande différence quand on compare le salaire des études sous graduées entre les hommes et les femmes. Cependant, cette différence n’est plus significative une fois qu’on atteint les études graduées (research student)

Chart, bar chart

Description automatically generatedChart, bar chart

Description automatically generated

Table

Description automatically generated

* Hypothèse 3 : Les hommes caucasiens ont tendance à gagner plus d’argent que le reste de la population
  + Cette hypothèse est bel et bien confirmée. De plus, nous avons effectué plusieurs test t pour vérifier cela.

Chart, bar chart

Description automatically generated

* Hypothèse 4 : Il y a une relation directe entre le type d’occupation et le salaire gagné et les femmes occupent une plus grande proportion de la catégorie d’occupation gagnant un bas salaire.
  + Nos résultats montrent que les deux catégories qui gagnent le moins d’argent sont les étudiants et les personnes sans emploi. En revanche, ceux qui gagnent le plus sont les gens qui occupent une poste de type manegarial gagnent le plus.

Chart, bar chart

Description automatically generated

* + On peut aussi observer que ce sont les femmes qui occupent la plus grande proportion de salaires bas, par rapport aux hommes. Nos hypothèses sont donc confirmées

Chart, pie chart

Description automatically generated

* Hypothèse 5 : Les personnes qui vivent dans des maisons ont tendance à gagner plus d’argent qu’on soit une homme ou une femme.
  + En observant les résultats, on se rend compte que ce sont les gens qui vivent dans des condominiums qui gagnent le plus. Cependant, en regardant notre test t, cette différence n’est pas significative.

Chart, bar chart

Description automatically generated

**Explications pour la durée nécessaire à la réalisation du projet**

Le projet a pris plus de temps qu’escompté. 12 heures ont été nécessaires pour chaque coéquipière. Le temps prévu était de 15 heures total. La plus grande partie du temps supplémentaire a été accordée au *debugging* du code.

**Comment l’approche personnelle pour ce projet a-t-elle changé ?**

La division des tâches a été modifiée ainsi que la répartition du temps pour travailler sur le projet. D’abord, nous avons décidé de diviser les tâches en fonction de nos forces. Ainsi, Meriem a fait les tests statistiques pour les hypothèses décrites dans le plan et Lisane s’est chargée de faire l’apprentissage machine. Nous avons rédigé ce rapport ensemble. De plus, il n’a pas été possible d’accorder le temps prévu à chaque semaine pour le projet. En effet, lors de nos deux semaines d’examen et de remise, nous avons eu moins de temps pour faire ce projet et ainsi, nous avons dû y mettre plus de temps dans la dernière semaine.

**Description de l’hypothèse – les résultats ont-ils changé l’approche pour l’hypothèse initiale ?**

Non, les résultats n’ont pas changé.

**Description des défis avec le projet**

Le principal défi que nous avons rencontré est la nature catégorielle de toutes nos variables. Ainsi, nous ne pouvions pas tout simplement implémenter le code appris en classe puisque celui-ci s’appliquait à des variables continues. Nous avons donc du faire des recherches sur internet pour apprendre comment prétraiter nos données pour qu’elles puissent être analysées comme il se doit par les algorithmes.