Questionnement :

Quel est l’âge médian d’un conducteur au Québec ?

Nature du dataset:

L’ensemble de données que nous avons choisi pour essayer de répondre à cette question est fourni par la SAAQ (Société d’Assurance Automobile du Québec) et contient un ensemble d’informations sur les permis de conduire en circulation au Québec dont la version publique la plus récente est le jeu de données sur l’année 2017.

Informations sur les potentielles corrélation :

À partir des résultat obtenue sur la partie du jeu de données utilisé, nous ne pouvons pas affirmer mais nous pouvons supposer une corrélation entre l'âge et le sexe des titulaires du permis de conduire en circulation au Québec.

Information sur le traitement et la visualisation :

Nous avons appliqué le traitement suivant sur les données avant leur visualisation :

* Nous avons isolé les colonnes renseignant l’âge et la région des conducteurs
* Nous avons calculé pour chaque âge le nombre de représentant par région
* Nous avons séparé le sous-ensemble des conducteurs masculins de celui des conducteurs féminins.

Nous avons choisi de visualiser ces données de deux façons différentes :

* Un diagramme de dispersion accompagné d’une ligne de tendance non-linéaire par région
* Un histogramme accompagné d’un indicateur de dispersion par région

Le diagramme de dispersion nous a permis d’évaluer rapidement la tendance ainsi que de déceler de potentielles disparités entre les régions.

L’histogramme nous a permis d’obtenir plus précisément les indicateurs de dispersion comme l'âge médian des conducteurs pour chaque région.

Commentaires sur le résultat

Nous avons pu répondre à notre question, la plage d’âge médian d’un conducteur au Québec et de 48 à 52 ans, nous avons pu noter également que dans notre ensemble de données, cet âge médian peut varier assez largement en fonction de la région, cette observation, nous pensons, devrais avoir tendance à s’invalider plus la quantité de données augmente.

Nous n’avons pas pu déceler de corrélation évidente entre le sexe du conducteur et l’âge médian de ce sous-ensemble.

Au niveau du projet, nous avons utilisé plusieurs librairies

* numpy
* plotly

Dans la librairie numpy, nous avons utilisé une méthode :

genfromtxt qui permet de lire le contenue du fichier CSV et le placer dans un tableau.

Dans la librairie plotly nous avons utilisé la classe express qui permet de générer des diagrammes. Dans notre cas, c’est pour afficher un histogramme et un diagramme de dispersion à partir de nos datas.



Pour le diagramme de dispersion nous avons utilisé la fonction « scatter » dont le premier paramètre est le dataframe (notre tableau de données formatées) nous avons renseigné « age » pour le paramètre X et « size » pour le paramètre Y ce qui nous permet de générer un digramme de dispersion représentant la size (nombre de conducteurs avec le même âge en fonction de l’âge), de plus cette fonction nous permet d’assigner une couleur par région, le paramètre size définit la taille des points et nous l’avons fixé à l’attribut « size » et le paramètre « trendline » indique de tracé une courbe de tendance (la valeur « lowess » permet de tracer une courbe de tendance non-linéaire)



Pour l’histogramme nous avons utilisé la fonction « histogram » de la même façon.