

Les exercices sont indépendants.

Exercice 1 : DataWarehouse

Un sondage ou étude d'opinion permet de connaître les intentions de vote d'un échantillon de la population. Pour cet exercice, nous nous limiterons à des élections en France. Il existe plusieurs types d'élections au suffrage universel direct en France : élections présidentielles, régionales, départementales, européennes, ...

Lors d'une étude d'opinion, les instituts peuvent choisir entre trois types de questionnaires différents : le face à face, par téléphone, par Internet. Les sondages donnent rarement le résultat exact d'un scrutin puisqu'ils ne portent que sur un échantillon considéré comme représentatif par les entreprises de sondages et pas sur la population totale.

Cet échantillon de personnes correspond à un panel représentatif des Français âgés de 18 ans et plus. Pour le respecter, les instituts utilisent la méthode des "quotas" selon des critères de sexe, d'âge (les 18-24 ans, les 25-34 ans, 35-49 ans, 50-64 ans, plus de 65 ans), de région, de catégorie d'agglomération (ville inférieure à 20000 habitants, ou entre 20000 et 100000, ...), de profession, et de catégorie socioprofessionnelle. L'INSEE a défini en 1982 une nomenclature des professions : les 486 professions (PCS) sont regroupées en 31 catégories socioprofessionnelles (CS), elles-mêmes regroupées en 6 groupes socioprofessionnels (GS).

Un sondage est réalisé sur une période, pour un scrutin donné et permet de connaître pour chaque parti/candidat, une répartition des électeurs de l'échantillon. Il est réalisé par un institut de sondage (BVA, CSA, France-Sondages, Ifop, ...). Par exemple, Odoxa a réalisé un sondage pour le scrutin du 2nd tour présidentielle de 2016. Ce sondage a été effectué du 14 au 16 février 2016 auprès d'un échantillon de 1011 personnes.

Un grand hebdomadaire français vous demande de réaliser un entrepôt de données permettant de connaître l'évolution des opinions pour différentes élections. Cette évolution doit pouvoir se faire en fonction du temps, du lieu de l'élection, du type d'élection, de l'organisme de sondage, du candidat, d'un parti, d'un courant (extrême gauche, gauche, centre, droite, extrême droite), du lieu de résidence du sondé, de sa catégorie socioprofessionnelle, ...

Cet entrepôt doit permettre d'obtenir différents rapports, comme les figures 1 et 2 données dans la suite du document.

Question 1 :

Proposez un schéma en étoile pour cet entrepôt. Vous justifierez vos choix et notamment expliquerez les faits. Vous préciserez la(es) mesure(s). Vous apporterez un soin particulier pour détailler la granularité des données et les différents niveaux des hiérarchies au sein des dimensions.

Question 2 :

L'hebdomadaire souhaite avoir les informations sur un temps important pour bien voir les évolutions des opinions politiques de la population française. Il considère donc qu'il lui faut un recul d'au moins 2 générations. Une génération en sociologie est définie comme une période d'environ 25 ans. Nous disposons de 30 instituts de sondage. Chaque institut de sondage fait en moyenne 40 sondages par an. Le 31 décembre 2014, la Commission nationale des comptes de campagne et des financements politiques (CNCCFP) a publié les comptes des partis et groupements politiques français pour l'année 2013. Dans ce document, elle a dénombré 408 formations politiques. Sur ces 408 formations politiques, on distingue 54 partis politiques nationaux, mais également des partis politiques régionaux et locaux. Un sondage portera en moyenne sur 54 partis. Il y a en moyenne 5000 personnes interrogées par sondage.

Estimez la taille de l'entrepôt en prenant soin d'expliquer comment vous êtes arrivé à cette estimation.