

Rapport de Stage

– Etape 2: Comparaison des données entre les villes

Yuan ZHANG

May 22, 2018

Contents

1	Introduction	2
2	Complétude des données	2
3	Récupération de variables intéressantes	2
4	Comparaison des variables entre certaines villes	2
5	Conclusion	3

1 Introduction

Le travail dans ce rapport fait partie du projet de maintenance prédictive. Dans ce rapport, je me concentre sur la comparaison des données entre certaines villes (données d'Armoire BT, Points lumineux, le dossier de l'intervention). Les villes qu'on choisit, ce sont BESSANCOURT, BOGOR, CAVB, GOUSSAINVILLE, GUADELOUPE, MONTESSON, NOUMEA. On choisit 15/05/2018 00:00:00 comme l'heure actuelle

Je vais montrer la complétude des données, la sélection des variables et la distribution.

2 Complétude des données

Je calcule le pourcentage de valeurs non nulles pour chaque variable et fixe un seuil. Je trouve le pourcentage de variables dont complétude est supérieure à ce seuil pour les groupes différents. Les 'groupes', ce sont des groupes internes dans BDD.

Pour les BDD Armoires, les données des villes 'CAVB', 'GUADELOUPE', 'MONTESSON' ont une meilleure intégrité des données. Les données des variables de groupe 'MODELE', 'DEPART' sont insuffisantes par rapport à d'autres groupes.

Pour les BDD PL, on peut noter que les données de 'GUADELOUPE' sont les plus complètes. Les données des variables de groupes 'COORDONNEE', 'EQUIPEMENT' ont une meilleure intégrité, mais la complétude des groupes 'CONSOLE', 'LANTERNE' et 'LAMPE' ne sont pas suffisantes.

Pour les BDD Intervention, on a constaté que les données globales sont relativement complètes sauf que 'BOGOR' connaît une grave pénurie de données sur le groupe 'Intervention'.

3 Récupération de variables intéressantes

Je choisit le seuil 0.05 et liste les variables dont le pourcentage de non-NAN pour tous les villes est supérieure au seuil. Après je vérifie les variables manuellement et trouve une liste des variables intéressantes.

- BDD Armoire:

'eq_EtatExploitation', 'arm_NoLampe', ('eq_Commentaire', 'eq_Vetuste')

- BDD PL:

'app_Puissance', 'app_Code', 'eq_EtatExploitation', 'eq_TypeRacine', 'eq_TypeAmont', 'lan_Vetuste', 'pl_Reseau', 'lampe_Puissance', 'lampe_Type', 'pl_NoLanterne'

- BDD Intervention:

'pan_DateSignal', 'pan_MiseSecurite', 'pan_TypeEqt', 'pan_Solde', 'pan_Commentaire', 'pan_HeureSig', 'pan_SourceEqt', 'pan_Default', 'pan_DelaiInt', 'pan_NoPLimp', 'int_NoPLimp', 'int_ElemDefault', 'int_Fin', 'int_Solde', 'int_TypeInt', 'int_Default', 'int_Commentaire', 'int_Debut', 'int_DateIntervention'

4 Comparaison des variables entre certaines villes

On génère les graphs de distribution pour les variables qu'on choisit. On voit que les BDD des villes sont similaires sauf 'BOGOR'.

La plupart des armoires sont en exploitation et le nombre de point lumineux pour une armoire est principalement entre 0 et 50.

Pour les villes ‘BESSANCOURT’ et ‘GOUSAINVILLE’, la plupart des points lumineux sont ‘vétusté’. les matériaux principaux de support, ce sont ‘acier’ et ‘béton’. les principaux types de support sont ‘mât’ et ‘poteau’. Pour tous les villes, la majorité de lanternes sont en bonne ou moyenne état.

Les pannes principales proviennent des points lumineux et le défaut principal, c’est ‘point lumineux éteint’. L’élément défaut principal, c’est ‘lampe’. La majorité des pannes sont soldés rapidement.

5 Conclusion

On peut noter que la distribution des données des villes est généralement similaire et on a sélectionné certaines variables intéressantes et sans beaucoup de NAN.

On peut rejoindre les trois BDD. Pour les données qu’on va générer, une observation est une ligne de BDD Intervention avec des données d’équipement en panne (proviennent de BDD PL, BDD Armoire.)