Régression Logistique

Par Mohamed Kallel, Simon-Pierre Rodner & Jean-Christophe Rigoni

Description du Projet

Objectif:

Déterminants pour qu'une femme soit le **principal apporteur de ressources** actuel du ménage

Dataset:

Enquête Conditions de Travail 2013

Périmètre:

Femmes vivant en couple dans le même ménage

Méthodologie

1. Exploration & Traitement

- Utilisation du **Dictionnaire**
- Traitement des données

2. Analyse des corrélations

- Calculer les **corrélation**
- **Choisir** les variables

3. Modélisation & Conclusion

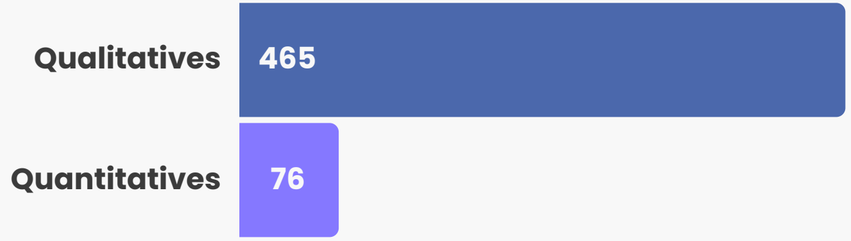
- Régression **Logistique**
- **Optimisation** du modèle
- **Interprétation**

Exploration des données

541 Colonnes
33K Lignes

4.6M ~ 25%
Valeurs Manquantes

Variables: Qualitatives & Quantitatives



Traitement des données

1. Filtrage & Création de variable

- **Femmes** en **couple** vivant **ensemble**
- Création variable **'target'**

2. One Hot Encoding

- One Hot **Encoding** sur les variables **qualitatives**

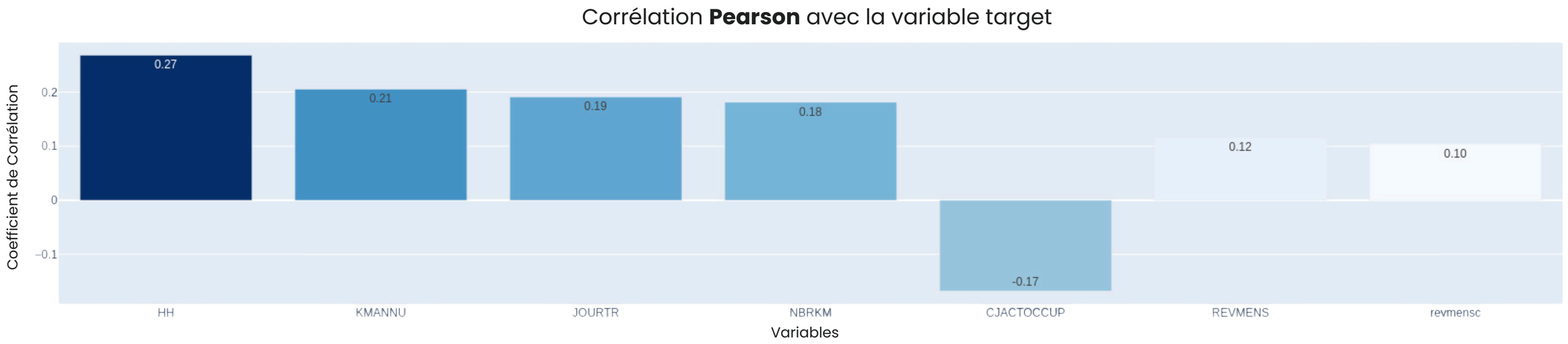
3. Données Manquantes

- Variables **Qualitatives**
- Variables **Quantitatives**

2. Analyse des corrélations

Variables Quantitatives

- Corrélation de **Pearson**
- Choix de **3** variables **quantitatives** avec le **dictionnaire**

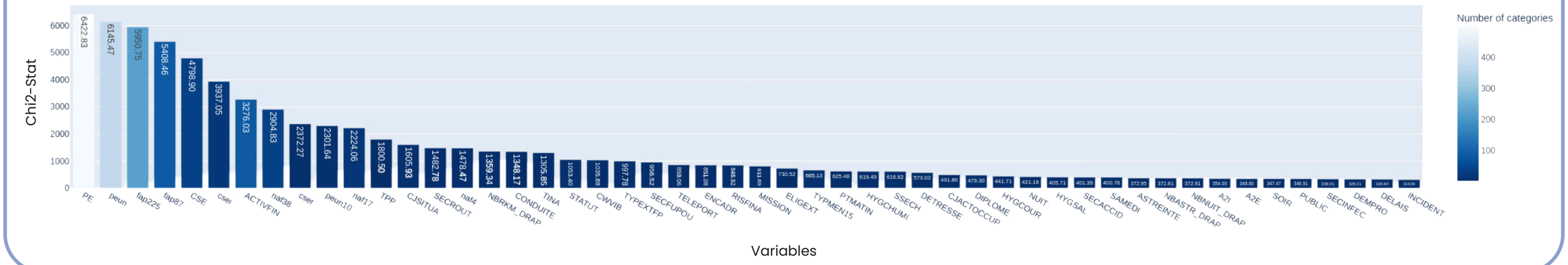


2. Analyse des corrélations

Variables Qualitatives

- Test **Chi2**
- Choix de **50** variables **qualitatives**

Chi2-Stat with qualitative variables



2. Analyse des corrélations

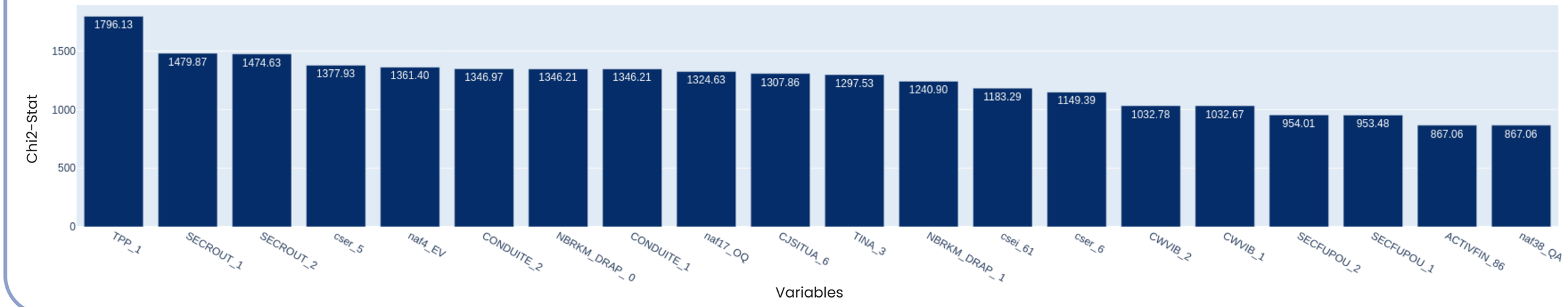
Variables Qualitatives

- Test **Chi2**
- Choix de **50** variables **qualitatives**

Valeurs Qualitatives

- One **Hot Encoding**
- Choix de **20** variables **qualitatives**

Chi2-Stat with qualitative variables

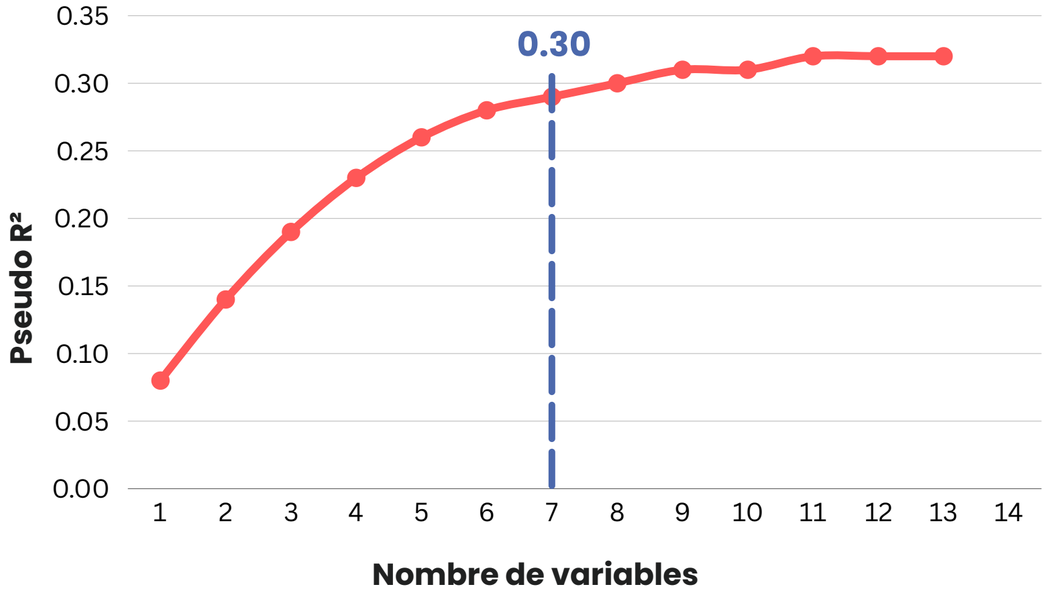


3. Modélisation

Modèle de Régression Logistique

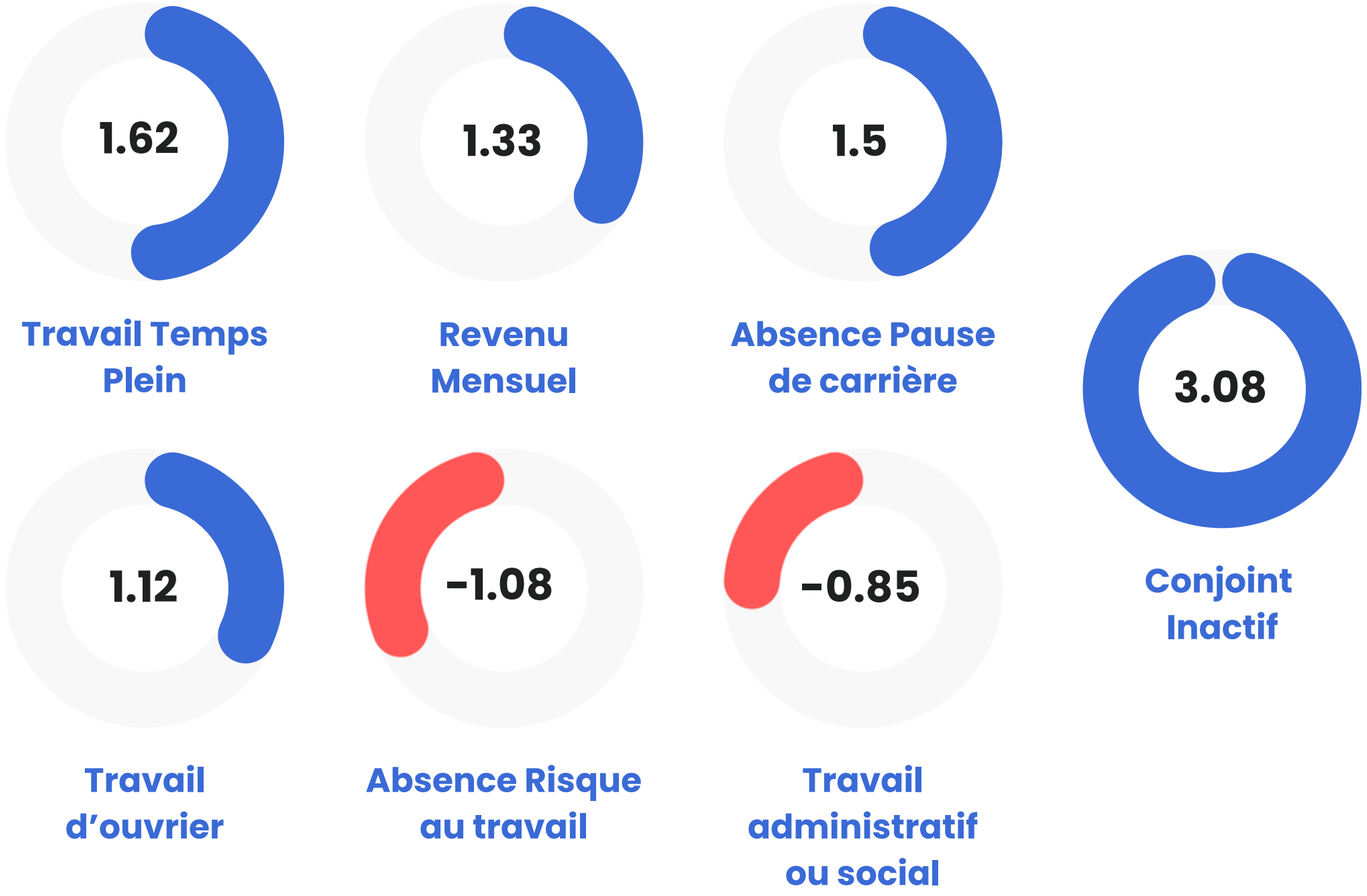
Optimisation du modèle

- **Réduction** de **15 -> 7** variables



Evaluation du modèle

- **Confusion Matrix**
- **Précision, Recall, F1**
- **ROC-AUC: 0.85**



Conclusion

Famille et dynamique de carrière:

- Situation du conjoint (CJSITUA_6 - **Inactif**): **3.08**
- Pauses de carrière (TINA_3 - **Absence**): **1.5**

Caractéristiques du travail:

- Temps de travail (TPP_1 - **Temps Plein**): **1.62**
- Risques de blessure (SECROUT_2 - **Absence**): **-1.08**

Facteurs Professionnels:

- Secteur d'industrie (naf17_OQ - **Administratif/Social**): **-0.85**
- Catégorie Socio-Professionnelle (cser_6-**Ouvrier**): **1.12**

Facteurs Economiques:

- **Revenu Mensuel** (REVMENS): **1.33**