Projet criminalité chicago

Lien données chicago

<https://data.cityofchicago.org/>

Données crimes (bdd principale)

<https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2>

Données socio économiques

<https://data.cityofchicago.org/Health-Human-Services/hardship-index/792q-4jtu>

Données code postal équivalence

<http://robparal.blogspot.com/2013/07/chicago-community-area-and-zip-code.html>

**Études déjà réalisées sur la criminalité**

<https://datascience-enthusiast.com/Python/SparkDataFrames-ExploringChicagoCrimes.html>

spark

<https://github.com/ernest-kiwele/chicago-crime-analysis-apache-spark/blob/master/spark-ml/chicago-crime-data-on-spark.ipynb>

Spark ml - Prédiction type de crime<https://cjango.wordpress.com/portfolio/chicago-crime-data-analysis-python-project/>

<https://github.com/andrewc23/Chicago_Crime_Analysis/blob/master/Andrew_Clark_CIS_545_Final_Project.ipynb>

**Liens sur études réalisées sur Hadoop**

<https://perso.univ-rennes1.fr/pierre.nerzic/Hadoop/tp1.pdf>

<https://docs.microsoft.com/fr-fr/archive/msdn-magazine/2013/september/azure-insider-hadoop-and-hdinsight-big-data-in-windows-azure>

Planification des tâches

**T0 :**

Design de l’architecture technique de la solution :

Données CSV → ETL (Extract AND Transform Data) Python ? → Hadoop → Map Reduce → Spark → PySpark pour ML.

Comparer Yarn et Spark

Librairie Python

Faire schéma + explication globale

Choix de la base de données - Présentation des données et de la base :

Définition Indicateurs, portée et objectifs :

-

-

-

-

-

-

comparer mapreduce yarn et spark

Objectifs prédiction :

Sévérité du crime (crime sévère / Non sévère) ?

Prédiction type de crime (vol, infraction …)

Récupérer données depuis poste local école avec accès internet → Copier/coller dans serveur