

# Projet Indeed prediction de salaire

César Parra  
Jeremy  
Noli Sapunxhiu  
Tim



# Agenda

**01**

## **Trello**

Organisation  
Les tâches de travaux

**02**

## **Schéma fonctionnel**

**03**

## **Développement du projet**

**04**

## **Conclusion**








# 01

## Trello

Trello



Du nouveau !

Espaces de travail

Récent

Favoris

Modèles

Créer

Parcourir

CP

IA Greta  
Gratuit

Tableaux

Membres

Paramètres

Vues de l'espace de travail

Tableur

Calendrier

Vos tableaux

Conduite de projet

Projet è Indeed NTJC

Tableau

Projet è Indeed NTJC

IA Greta

Visible par l'espace de travail

CP NS TA

Inviter

Power-up Calendrier

Automatisation

Filtre

Afficher le menu

Ajoutez une autre liste

Backlog

Data Frame

Preprocessing

Feature engineering

Machine learning et utilisation de pipeline

Evaluation du/des modèles

Interface web - Flask

Schéma de fonctionnement

Google Slide pour la presentation

Ajouter une carte

Livrable

Gitlab avec le code

Diagramme de fonctionnement

Flask dashboard

Ajouter une carte

Doing

Scrapping Indeed

Mongo DB

EDA

Ajouter une carte

Done

Trello

Git Hub (SalairePred\_G4)

Métriques d'erreur

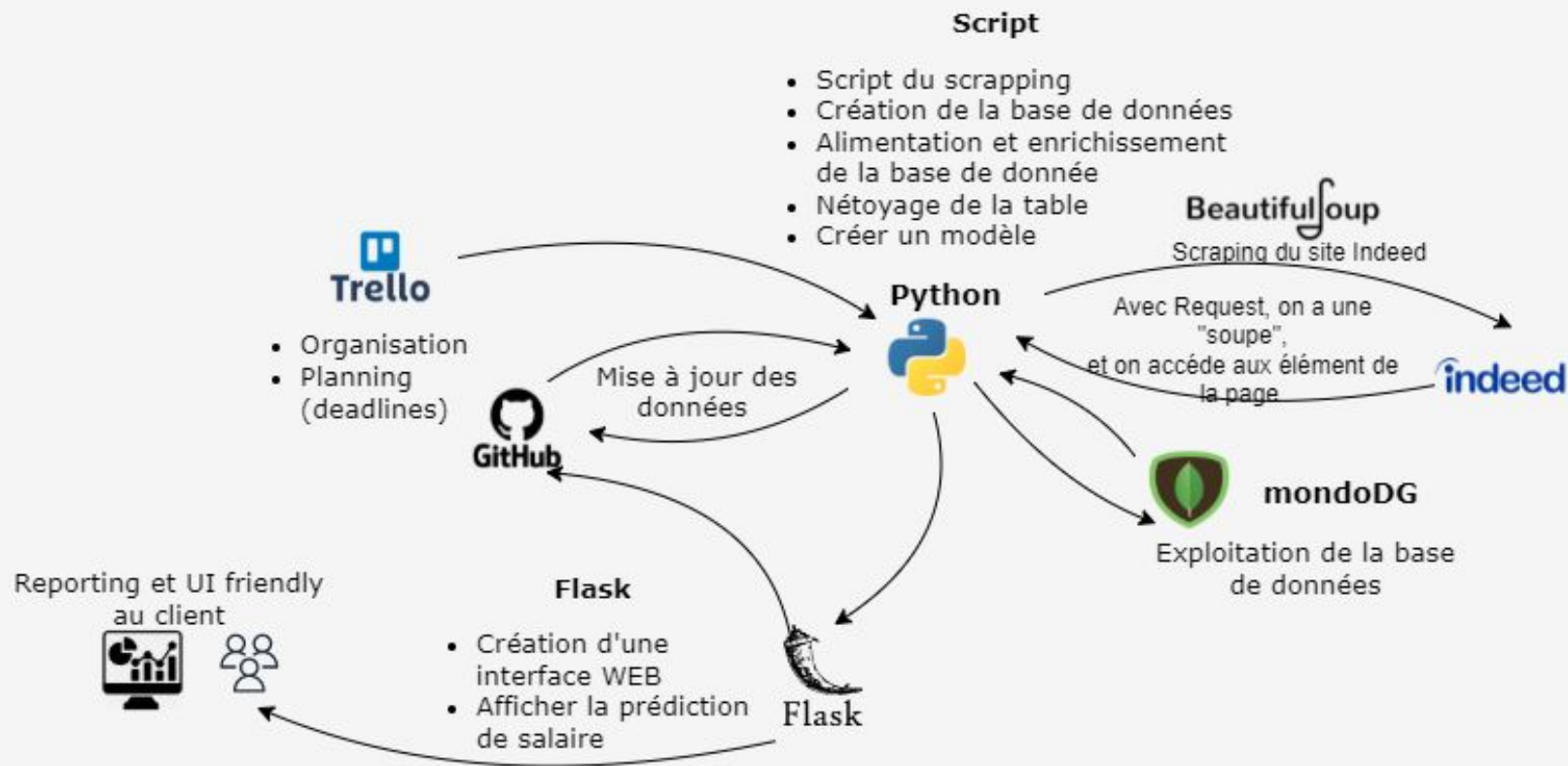
Ajouter une carte

02

# Schéma fonctionnel







# Scraping



# Analyse Exploratoire des Données



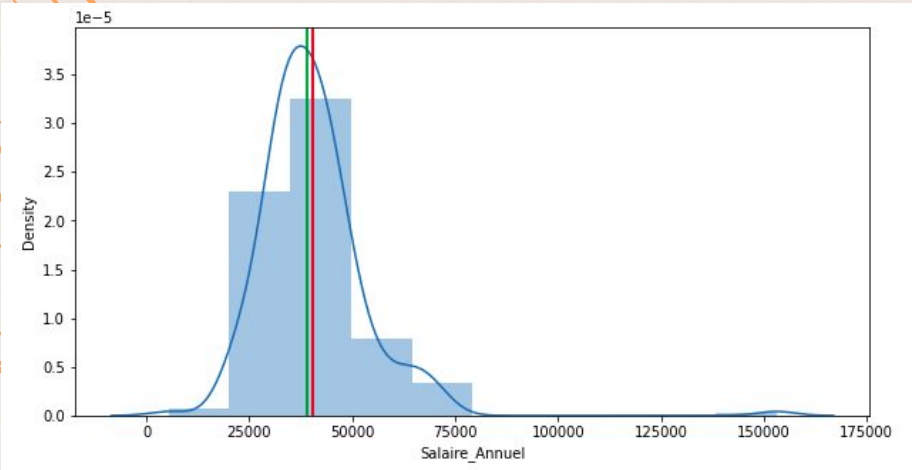


### Dataset statistics

Number of variables	13
Number of observations	206
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	2
Duplicate rows (%)	1.0%
Total size in memory	21.0 KiB
Average record size in memory	104.6 B

### Variable types

Categorical	8
Numeric	5



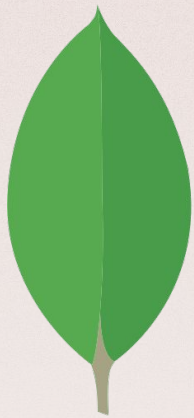
### Salaire\_Annuel

#### Quantile statistics

Minimum	5422.872
5-th percentile	23211.81
Q1	33000
median	39000
Q3	45000
95-th percentile	64750
Maximum	153000
Range	147577.128
Interquartile range (IQR)	12000

#### Descriptive statistics

Standard deviation	13441.3674
Coefficient of variation (CV)	0.3321786967
Kurtosis	23.31025354
Mean	40464.26676
Median Absolute Deviation (MAD)	6000
Skewness	3.131476703
Sum	8335638.953
Variance	180670357.5
Monotonicity	Not monotonic



mongoDB



Connect View Collection Help

Local

14 DBS 28 COLLECTIONS

☆ FAVORITE

HOST

localhost:27017

CLUSTER

Standalone

EDITION

MongoDB 5.0.5 Community

Filter your data

▸ IndeedStersice

▸ Pr6\_SpamC\_Marceline\_Tess

▸ Projet\_6\_Spam\_Celia\_Cesar

▸ admin

▸ config

▸ grp3\_salaire

▸ kalm\_indeed

▸ local

▾ projet\_7\_JNTC

▫ scraping\_indeed

▫ scraping\_indeed\_2 ...

▸ projet\_indeed\_web\_scrapping\_e...

▸ sapunxhiu

▸ smsspamcollection\_eric\_tim

▸ test

▸ testEG

## projet\_7\_JNTC.scraping\_indeed\_2

DOCUMENTS 206

STORAGE SIZE  
28.7KBAVG. SIZE  
286B

INDEXES 1

TOTAL SIZE  
20.5KBAVG. SIZE  
20.5KB

Documents Aggregations Schema Explain Plan Indexes Validation

FILTER { field: 'value' }

► OPTIONS

ANALYZE

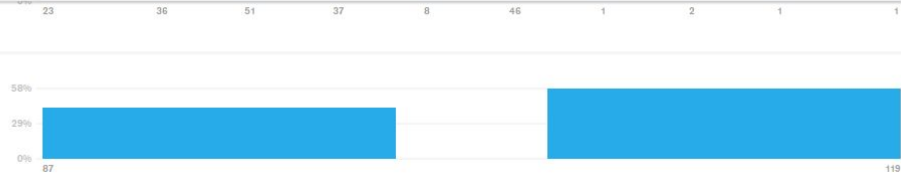
RESET

⌵ ⌵ ⌵

This report is based on a sample of 206 documents. ⓘ

## langages\_programmation

int32



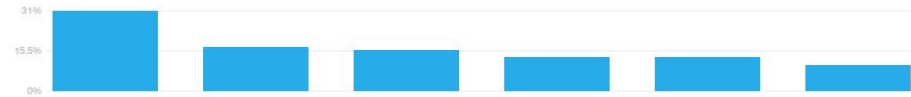
## langages\_Web

int32



## poste

string



## Salaire\_Annuel

double



# Modèle

GradientBoostingRegressor

R2: 0.42

MAE(mean): 6390.52

MAE(median): 5478.97

RMSE: 8463.66



# Flask

IndeedSalairePred



**Thanks!**  
**Any question ???**

