



1

# PréCrime San Francisco

Développer un modèle d'analyse de données pour la prédiction des catégories de crime de données déjà existantes

# L'équipe

- Chef de projet & responsable algorithme
  - ☐ David Dupuis
- Responsable analyse de données
  - ☐ Pierre Comalada
- Responsable intégration des données météo
  - ☐ Nicolas Bonichon
- Responsable visualisation
  - ☐ Jérémie Chevallier

# Sommaire

➤ Présentation de l'équipe	2
➤ Sommaire	3
➤ Le cadre du projet (le concours Kaggle)	4
➤ Les données de crimes	5
➤ Les données météo	6
➤ La visualisation des données	7
➤ L'algorithme de prédiction	8
➤ Nos valorisations	9
➤ Conclusion	10

# Le concours Kaggle

- Kaggle qu'est-ce?
- Le concours Classification des crimes à San Francisco:
  - ❑ Pour apprendre les bases de l'apprentissage machine
  - ❑ 687 équipes internationales
  - ❑ 833 participants
  - ❑ Un système de score
  - ❑ Une évaluation en continue des résultats des participants

# Les données de crimes

## ► Avancement

- ❑ Fréquence des catégories et sous-catégories
- ❑ Les districts avec le plus de crimes et la catégorie la plus dominante
- ❑ Nombre de crimes total par jour de semaine
- ❑ Nombre de crimes par heure pour chaque Catégorie ou District

## ► Difficultés

- ❑ Utilisation de toutes les données
- ❑ Se poser les bonnes questions pour les bonnes analyses

## ► La suite

- ❑ Fréquence des crimes par localisation
- ❑ Différences significatives entre différentes catégories de crime pour un même endroit

# Les données météo

## ► Avancement

- ❑ La librairie weatherData
- ❑ Les données récupérés

## ► Difficultés

- ❑ Des données détaillés
- ❑ Le fonctionnement de la librairie

## ► La suite

- ❑ Ajouter les données météos aux données de crimes

# La visualisation des données

## ➤ Avancement

- ❑ Utilisation du logiciel Tableau pour la visualisation des données d'une base de données

## ➤ Difficultés

- ❑ Découverte du D3.js pour une autre forme de visualisation

## ➤ La suite

- ❑ Fournir une interface interactive sous forme de carte

# L'algorithme de prédiction

## ► Avancement

- ❑ Le deep learning
- ❑ La plateforme H2O

## ► Difficultés

- ❑ Les 68 paramètres de la fonction de deep learning
- ❑ Le temps d'apprentissage
- ❑ Les données de crimes

## ► La suite

- ❑ Utiliser l'analyse des données pour simplifier nos données en entrée, simplifier l'apprentissage et mieux jauger les paramètres



# Nos valorisations

- Le concours
  - ❑ Challenge entre les différentes équipes
- La recherche
  - ❑ Etudier les relations entre les crimes et la météo
- Le savoir
  - ❑ Découverte du deep learning
  - ❑ Découverte des méthodes pour la data science

# Conclusion

- L'analyse des données
  - La récupération des données météo
  - La visualisation des données
  - Le deep learning
- 
- Utiliser le travail de chacune des parties ensemble pour faire avancer globalement le projet.

**MERCI**