# **I社データセットの概要**

## 1.データセットについて

本データセットは、[IBM HR Analytics Employee Attrition & Performance](https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset)を最終課題用に変更を加えたものです。

**出典の会社はIBMですが、最終課題の提案先としてはIBMとして考察しても構いませんし、別の会社や架空の企業という設定にしていただいても構いません。**どの方法でも減点・加点はございませんので取り組みやすいように設定してください。

* データセットの内容テクノロジー関連企業I社の過去の人事データで、離職したかどうかなどについて記述されている
  1. 一部カラムについての説明が明らかになっている（「2.カラムの説明」参照）
  2. 約1500人の従業員についてのcsvファイルが提供されている
* データセットの特徴
  1. 初学者に扱いやすいデータセット
     1. カラムの意味が推測しやすいものが多い
     2. 欠損値などが少ないことや数値カラムが多いことから基本的なデータ分析で事業提案まで考察しやすい
     3. データ数も少なく扱いやすい
  2. 他者との差別化が難しいデータセット
     1. 操作可能な変数が少なく、事業提案が他の人と被りがちになり他者との差別化が難しい
* こんな方におすすめ
  1. データサイエンス初学者で、修了まではなんとかたどり着きたい人

特にビジネスについての知識に自信がない方にとってはカラムの意味などから事業を考えやすいのでお勧めです

* 1. 人事データについて触れたい人

## 2.カラムの説明

　I社の担当者にデータについて問い合わせたところ、一部のカラムについて以下の説明が送られてきました。現段階では正式なコンサルティング契約は未締結であり、情報保護規程の厳しいI社からこれ以上の詳しい情報開示を受けることはできません。

　不明な点があるかと思いますが、不明な点があること自体も業務上の与条件として分析を進めてください。

Education

1. 'Below College'
2. 'College'
3. 'Bachelor'
4. 'Master'
5. 'Doctor'

EnvironmentSatisfaction

1. 'Low'
2. 'Medium'
3. 'High'
4. 'Very High'

JobInvolvement

1. 'Low'
2. 'Medium'
3. 'High'
4. 'Very High'

JobSatisfaction

1. 'Low'
2. 'Medium'
3. 'High'
4. 'Very High'

PerformanceRating

1. 'Low'
2. 'Good'
3. 'Excellent'
4. 'Outstanding'

RelationshipSatisfaction

1. 'Low'
2. 'Medium'
3. 'High'
4. 'Very High'

WorkLifeBalance

1. 'Bad'
2. 'Good'
3. 'Better'
4. 'Best'

RemoteWork

1. ‘on-site work’

5. ‘full remote’

Incentive : rewards for performance